

MineXpert3 is part of the MsXpertSuite software package

Modelling, simulating and analyzing ionized flying species

MineXpert3 User Manual

Visualization, exploration and analysis of MS^N mass spectrometric data

MineXpert3 11.4.2

MineXpert3 User Manual: Visualization, exploration and analysis of MS^N mass spectrometric data

by Filippo Rusconi

May 07, 2026 , 11.4.2

Copyright 2009-2025 Filippo Rusconi

MsXpertSuite - mass spectrometry software suite

http://www.msxpertsuite.org/

This book is part of the MsXpertSuite project.

The MsXpertSuite project is the successor of the MassXpert project. This project now includes various independent modules:

- MassXpert, a program to model polymer chemistries, perform a wide array of (bio-)chemical reactions and simulate the corresponding mass spectrometric data;
- MineXpert, the successor of MineXpert, a program to visualize and mine MS^n mass spectral data (mass spectrum, drift spectrum, XIC chromatograms) starting from the TIC chromatogram.

This program is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with this program. If not, see http://www.gnu.org (http://www.gnu.org/licenses/).

Revision History

Revision 11.1.0 March 2026 fr

Update the documentation to describe the new centroid extraction and low mass deconvolver features.

Revision 11.0.0 January 2026 fr

Update the documentation to describe the new application preferences feature and the new vertical rule x-axis label in the plot windows.

Revision 10.2.0 November 2025 fr

Update the documentation to describe the new deconvolution processes.

Revision 10.1.1 November 2025 fr

Update the documentation about specifying the format of analysis session discoveries, in particular abour the new feature that let's the user define the number of decimals to be used when printing out float values.

Revision 10.1.0 November 2025 fr

Full port of MineXpert to the new XpertMass project's libraries. Add new chapter on JavaScript features now in MineXpert.

Revision 9.3.1 December 2023 fr

- Added brief documentation of the new MS run probing feature designed to allow the user to configure the loading of timsTOF MS run data from Bruker.

Revision 8.5.8 October 2022 fr

- Entirely reworked the user manual to make it more modern from the coding point of view. The users do not see any difference with the previous version.

Revision 8.4.0 June 2022 fr

- Update the user manual to document the fact that it is now possible to set the bin size even when computing color maps involving m/z as one of their dimension.

Revision 8.3.0 June 2022 fr

- Update the user manual with the description of the new plot ranges history resetting feature.

Revision 8.2.0 31 January 2022 fr

- Update the user manual with the description of the new trace low signal threshold removal filter.

Revision 8.1.0 21 April 2021 fr

- Update the user manual with the description of the new color map filters used to enhance the signal-to-noise ratio.

Revision 8.0.1 24 march 2021 fr

- Update the user manual with a more explicit data set to illustrate the new skewed selection polygon.
- Mention the bibliographic reference to the newly accepted article about MineXpert2 in the Generalities chapter.

Revision 8.0.0 23 march 2021 fr

- Update the user manual with a section pertaining to the new skewed selection polygon. That feature was asked months ago by a user of MineXpert2 and could finally be implemented these last weeks.
- Mention the bibliographic reference to the newly accepted article about MineXpert2 in the Generalities chapter.

Revision 7.4.0 2 january 2021 fr

Happy new year ! With the writing of the new version of the Processing Flow chapter to match the new integration logging mechanisms in MineXpert2.

Revision 7.4.0 december 2020 fr

This revision introduces updates to document the following new features.

- User-definable limitation of the number of threads usable for the combination computations (good request from a reviewer of the article).
- GUI shortcut to the MS level to be targeted for the next integration (good request from a reviewer of the article).

Document the build and installation of the software (another good request from a reviewer).

Revision 7.0.1 october 2020 fr

Update of the user manual to document the new combine+ and combine- combinations features.

Revision 7.0.0 september 2020 fr

Update of the user manual to document new features, in particular the MS^n mass spectrometric data mining and new mass data plotting features.

Revision 6.0.0 january 2020 fr

Full rewrite of the user manual to document the new mineXpert2 software program. The new mineXpert2 version is a full rewrite of the previous version such that it can handle MS^n mass spectrometric data. Major features were included in the new version that are described in detail in this user manual.

Revision 5.8.2 february 2019 fr

Rework the IsoSpec-based isotopic clust calculations to better describe the three available processes. Also, describe the new FWHM calculation feature in the PeakShaperDlg.

Revision 5.8.0 end january 2019 fr

Add section to described the newly developed feature about isotopic cluster calculations based on the IsoSpec library;

Revision 5.7.0 beginning january 2019 fr

Finished porting of the historical LaTeX-based documentation to the DocBook/DAPS/FOP publishing system. Please, see Colophon for details;

Revision not_set september 2018 fr

Update the document to include new features and to make fixes. Insist on the fact that mineXpert reads mzML data files.

Revision not_set march 2018 fr

Update the document to include new m/z~integration features and scripting capabilities.

Revision not_set december 2016 fr

Major rewriting of the document to incorporate all the new features. Large chapter on the scripting of mineXpert.

Revision not_set december 2017 fr

Refactored document. First version documenting almost all the of current features of the software program.

Revision not_set may 2017 fr

Refactored document. First version documenting almost all the of current features of the mineXpert software program.

Revision not_set november 2016 fr

Resume writing, with new program name: mineXpert.

Revision not_set	september 2016	fr

Start of writing.

Dedication

To María Cecilia

To all the admirable people acting in the *"Free Software Movement"* for a better and more ethical computing world

To all involved in the development of the K Desktop Environment (KDE) and of the Plasma desktop

To all the readers who helped me with this manual.

Contents

Preface v

1 Generalities 1

1.1 Citing the MineXpert software. 1

1.2 General concepts and terminologies 1

1.3 Acquiring Mass Data Along Time: To Profile or Not To Profile? 1

1.4 Mass Data Visualisation: To Combine or Not To Combine? 2

1.5 Examples of Various Mass Spectral Data Integrations 4

TIC -> MZ integration 4 • TIC -> DT integration 5 • Using multiple threads during mass data integrations 6

1.6 Installation of the software 6

Installation on MS Windows and macOS systems 6 • Installation on Debian- and Ubuntu-based systems 6 • Installation with an AppImage software bundle 7

1.7 Building the software from source 7

The dependencies required to build MineXpert 7 • Getting the source tarball 8 • Building of the software 9

2 Configuring the application preferences 10

2.1 Setting the general options 10

2.2 Setting the shortcuts 11

2.3 Setting the recording of the discoveries 12

3 The program's graphical user interface 15

3.1 Opening Mass Spectrometry Data Files 15

3.2 The Window Layout 17

3.3 The Main Program Window Menu 18

3.4 The Loaded MS Run Data Sets 20

3.5 Mass Spectral Data Listing in a Table View 21

3.6 The Main Data-Plotting Windows 24

The TIC Chromatogram Window 26 • The Mass Spectrum Window 27 • The Drift Spectrum Window 28 • The Retention Time vs Mass Spectrum Color Map Window 29 • The Drift Time vs Mass Spectrum Color Map Window 30 • The Drift Time vs Retention Time Color Map Window 31 • Exploring data in multiple plot widgets 31

3.7 General Structure of the Plot Widgets 33

The Left Column of Buttons Configures the Reception of a New Plot 33 • The Right Column of Buttons Configures the New Integration to Run 35 • The Plot Widget Main Menu 36

3.8 General Operation of the Plot Widgets 40

4 Mass Data Integrations Featured by MineXpert 42

4.1 General Behaviour of Plot Widgets 42

Processing Flow Entities Document all Integrations 42 • Setting the MS^N Fragmentation Parameters 43 • Setting the m/z Integration Parameters 46 • Step-by-step Determination of the Best m/z Integration Parameters 49 • Removing 0-intensity m/z data points is useful 54 • Savitzky-Golay filtering of any kind of data 56 • Centroid Extraction from Mass Spectra 57

4.2 Various Mass Spectral Data Integrations 58

Considerations on the Diversity of Mass Data Contents 60 • Statistical Analysis of Mass Data 61

4.3 Integrations Originating in Color Map Plots 61

4.4 Noise Reduction in Color Map Plots 64

4.5 Low Intensity Signal Removal in Traces 66

4.6 Chained Integrations 71

5 Mass Spectral Deconvolutions 73

5.1 Deconvolution based on charge state 73

5.2 Deconvolution based on isotopic cluster peaks 76

5.3 Automated deconvolution based on isotopic cluster peaks 77

5.4 Reading the resolving power based on mass spectral data 80

6 Isotopic cluster calculations 82

6.1 Calculating Isotopic Clusters with IsoSpec 82

The IsoSpec Graphical User Interface in mineXpert 83

6.2 Shaping mass peak centroids into well-profiled peaks 91

7 The processing flow in a data exploration session 96

7.1 The processing flow records all the steps in a data exploration session 96

8 Recording data exploration discoveries 99

9 Scripting Interface 103

9.1 Starting the Scripting Console Invocation 103

9.2 The Scripting Console Window 103

9.3 Running Scripts in the Scripting Console 104

9.4 Loading Script Files in the Scripting Console 105

A GNU General Public License version 3 106

Preface

1 Software feature offerings and intended audience

This manual is about the MsXpertSuite mass spectrometric software suite, a software environment that contains two modules:

- *MassXpert* module: Allows users to define brand new polymer chemistries and use these polymer chemistry definitions to model linear polymer sequences. Once modelled, a polymer sequence can undergo chemical reactions (enzymatic or chemical cleavages, gas-phase fragmentations...). The obtained results are a model of what a mass spectrum would look like if the modelled experiment had actually been carried over up to the mass spectrometry analysis;

- *MineXpert* module: Allows users to load mass spectrometry data from mzML files, visualize and mine them throughout all their MS^n depth. The mass data visualization starts either at the TIC chromatogram level or the MS data table view and deepens to the MS^n mass spectra, the drift spectra, the XIC chromatograms, various color maps relating m/z values with either retention times or drift times. A wide array of mass spectral integrations are available to ease the MS^n data finest scrutiny.

This manual is therefore intended for people willing to learn how to use the comprehensive MsXpertSuite software package.

Mass spectrometry has gained popularity across the past twenty years or so. Indeed, developments in polymer mass spectrometry have made this technique appropriate to accurately measure masses of polymers as heavy as many hundreds of kDa, and of any chemical type.

There are a number of utilities—sold by mass spectrometer constructors with their machines, usually as a marketing “plus”—that allow predicting/analyzing mass spectrometric data obtained on polymers. These programs are usually different from a constructor to another. Also, there are as many mass spectrometric data prediction/analysis computer programs as there are different polymer types. You will get a program for oligonucleotides, another one for proteins, maybe there is one program for saccharides, and so on. Thus, the biochemist/massist, for example, who happens to work on different biopolymer types will have to learn to use several different software packages. Also, if the software user does not own a mass spectrometer, chances are he will need to buy all these software packages.

The MsXpertSuite mass spectrometric software is designed to provide *free* solutions to all these problems by providing the following features:

- MassXpert:
    - Model *ex nihilo* polymer chemistry definitions (in the XpertDef module that is part of the MassXpert program);
    - Peform simple yet powerful mass computations to be made in a mass desktop calculator that is both polymer chemistry definition-aware and fully programmable (that's the XpertCalc module also part of the MassXpert program);
    - Edit polymer sequences on a polymer chemistry definition-specific basis, along with chemical reaction simulations, finely configured mass spectrometric computations... (all taking place in the XpertEdit module that is the main module of the MassXpert program);
    - Customize the way each monomer will show up graphically during the program operation (in the XpertEdit module);
    - Edit polymer sequences with immediate visualization of the mass changes elicited by the editing activity (in the XpertEdit module);
    - Open an unlimited number of polymer sequences at any given time and of any given polymer chemistry definition type (in the XpertEdit module);
- MineXpert:
    - Load mass spectometry data files in the mzML format, thanks to the excellent libpwiz library of ProteoWizard[1] fame; Mass data file loading can be performed in *full-memory* mode (all the data read from file are stored in memory) for faster operations or in *streamed* mode when loading files largers than the available memory.
    - Display the data in powerful ways in a unified graphical user interface. The interface was designed to integrate all the most useful characteristics of the various proprietary environments known by the author, thanks to the excellent libqcustomplot[2] library;
    - Configure the way mass spectrometry data integrations are performed from a combination standpoint and optionally apply a Savitzky-Golay smoothing;

1 http://proteowizard.sourceforge.net/.

2 http://qcustomplot.com/.

- Configure the level of MS data (MS^n) for which the integrations are performed. Optionally filter data by precursor ion m/z values or precursor spectrum indices;
- Perform data mining by performing data integrations in various ways;
- Ion mobility mass spectrometry data are supported with a int=f(m/z,dt) color map plot calculation;
- A specific data integration mode allows easy mass spectral intensity determination for any feature (peak in a TIC chromatogram, a mass spectrum, a drift spectrum, any color map region);

2 Feedback from the users

We are always grateful to any constructive feedback from the users.

The MsXpertSuite software team might be contacted *via* the following addresses

```
msxpertsuite@msxpertsuite.org = general mailing list about msXpertSuite
bug-reports@msxpertsuite.org  = report bugs found in msXpertSuite software
```

FIGURE 1: **ADDRESSES TO REPORT FEEDBACK TO**

3 Project History

This is a brief history of MsXpertSuite.

- *1998–2000*

  The name MassXpert comes from a project I started while I was a post-doctoral fellow at the École Polytechnique (Institut Européen de Chimie et Biologie, Université Bordeaux 1, Pessac, France);

  The MassXpert program was published in *Bioinformatics* (Rusconi, F. and Belghazi, M. Desktop prediction/analysis of mass spectrometric data in proteomic projects by using massXpert. Bioinformatics, 2002, 644–655 (https://academic.oup.com/bioinformatics/article/18/4/644/243311) ↗).

  At that time, MS-Windows was at the Windows NT 4.0 version and the next big release was going to be *"you'll see what you'll see"*: MS-Windows 2000.

  When I tried MassXpert on that new version (one colleague had it with a new machine), I discovered that my software would not run normally (the editor was broken). The Microsoft technical staff would advise to *"buy a new version of the compiler environment and rebuild"*. This was a no-go: I did not want to continue paying for using something I had already produced with legitimate software.

- *2001–2006*

During fall 1999, I decided that I would stop using Microsoft products for my development. At the beginning of 2000 I started as a CNRS research staff in a new laboratory and decided to start fresh: I switched to GNU/Linux (I never looked back). After some months of learning, I felt mature to start a new development project that would eventually become an official GNU package: GNU polyxmass.

The GNU polyxmass software, much more powerful than what the initial MassXpert software used to be, was published in *BMC Bioinformatics* in 2006 (Rusconi, F. GNU polyxmass: a software framework for mass spectrometric simulations of linear (bio-)polymeric analytes. BMC Bioinformatics, 225– (https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-7-226)).

Following that publication I got a lot of feedback (very positive, in a way) along the lines: — *"Hey, your software looks very interesting; only it's a pity we cannot use it because it runs on GNU/Linux, and we only use MS-Windows and MacOSX!"*.

- *2007–2016*

  In december 2006, I decided to make a full rewrite of GNU polyxmass. The software of which you are reading the user manual is the result of that rewrite. I decided to "recycle" the MassXpert name because this software is written in C++, as was the first MassXpert software. Also, because the first MS-Windows-based MassXpert project is not developped anymore, taking that name was kind of a "revival" which I enjoyed. However, the toolkit I used this time is not the Microsoft Foundation Classes (first MassXpert version) but the Trolltech Qt framework (see the *About Qt* menu in the *Help* menu in MassXpert).

  Coding with the Qt libraries has one big advantage: it allows the developer to code once and to compile on the three main platforms available today: GNU/Linux, MacOSX, MS-Windows. Another advantage is that the Qt libraries are wonderful software, technically and philosophically (Free Software).

  The rewritten software was published in 2009 (Rusconi, F. massXpert 2: a cross-platform software environment for polymer chemistry modelling and simulation/analysis of mass spectrometric data.Bioinformatics, 2009, 2741–2742 (https://academic.oup.com/bioinformatics/article/25/20/2741/194220)).

- *2016–2019*

  In 2016, I started a new project about visualization of mass spectrometric data. The project developed pretty quickly, as we needed at the mass spectrometry facility a software that would allow to cope efficiently with ion mobility mass spectrometric experimental data. MineXpert was thus started.

  To bundle both MassXpert and MineXpert in a single software suite, I bought the MsXpertSuite website http://msxpertsuite.org and created that new name.

- *2019–*

  The MineXpert software was published in 2019. The reviewers acknowledged that it was a useful piece of software but complained about its not handling the MS/MS data. Reference to cite: Filippo Rusconi. mineXpert: Biological Mass Spectrometry Data Visualization and Mining with Full JavaScript Ability. *J. Proteome Res.* 2019, 18, 5, 2254-2259 https://doi.org/10.1021/acs.jproteome.9b00099.

In response to the reviewers and for my personal project that I had started a little earlier, I decided to fully rewrite MineXpert into MineXpert. The new version of the software had to support MS^n data.

- *2021–*

  The MineXpert software was published in 2019. The main new capability is now the full handling of MS^n data. This new feature required the full rewriting of the software internal data model. See Langella O, Rusconi F. (2021) mineXpert2: Full-Depth Visualization and Exploration of MS^n Mass Spectrometry Data. *J. Am. Soc. Mass Spectrom.*, 4 (32) 1138-1141.

- *2022–*

  A major work is now undertaken: the native handling of the timsTOF data from the Bruker instrument vendor. This kind of data contains ion mobility data.

4 Program and Documentation Availability and License

The programs and all the documentation that are shipped along with the MsXpertSuite software suite are available at http://www.msxpertsuite.org. Most of the time, a new version is published as source, and as binary install packages for MS-Windows (64-bit systems only). No GNU/Linux binary packages are created outside of the autobuilder of the various distributions. As a Debian Developer, the author creates Debian[3] packages that are uploaded on the distribution servers. These packages are available using the system's software management infrastructure (like using the Debian's `apt` command, for example, or the graphical application).

The software and all the documentation are all provided under the Free Software license *GNU General Public License, Version 3, or later, at your option*. For an in-depth study of the *Free Software* philosphy, I kindly urge the reader to visit http://www.gnu.org/philosophy.

3 http://www.debian.org/

1 Generalities

In this chapter, I wish to introduce some general concepts around the MineXpert program and the way data elements are named in this manual and in the program.

A mass spectrometry experiment generally involves monitoring the m/z value of analytes injected in the mass spectrometer along a certain time duration. The m/z value of each detected analyte is recorded along with the corresponding signal intensity $\int$, so that a mass spectrum is nothing but a series of (m/z,i) pairs recorded along the acquisition duration. All along the acquisition, the precise moment at which a given analyte is detected (and its (m/z,i) pair is recorded), is called the retention time of that analyte (rt). This retention time is not to be misunderstood as the drift time of that analyte in an ion mobility mass spectrometry experiment.

1.1 Citing the MineXpert software.

Please, cite the software using the following citation: mineXpert2: Full-Depth Visualization and Exploration of MSn Mass Spectrometry Data. Olivier Langella and Filippo Rusconi. *Journal of the American Society for Mass Spectrometry* (check for pages, as this is a pub ahead of print reference). DOI: 10.1021/jasms.0c00402.

1.2 General concepts and terminologies

Most generally, the mass spectrometer acquires an important number of spectra in, say, one second. But all these spectra are *combined* together, and, on the surface, the massist only sees a "slow" acquisition of 1 spectrum per second. This apparent slow acquisition rate is configurable. At the time of writing, generally 1 spectrum per second is recorded on disk. So, say we record mass spectra for 5 minutes, we would have recorded (5*60) spectra.

1.3 Acquiring Mass Data Along Time: To Profile or Not To Profile?

As a mass spectrometry user, the reader of this manual certainly has used mass spectrometers where mass spectra are acquired and stored in different ways:

- Mass spectra are acquired and summed—the next to the previous—in such a manner that one is left, at the end of the acquisition, with a single spectrum of which the various peak intensities have been increasing all along the acquisition. Indeed, in this mode, each new spectrum is actually *"combined"* to the previously acquired ones. The resulting mass spectrum that is displayed on screen and that gets ultimately stored

on disk is called a *combined spectrum*. This is typically the way MALDI-TOF mass spectrometers are used when manually acquiring data from samples deposited onto sample plates. We refer to this kind of acquisition as an "accumulation" mode acquisition;

- Mass spectra are acquired and stored on disk as a single file containing all the spectra, appended one after the other. There is no combination of the spectra: each time a new spectrum is displayed on screen, that spectrum is appended to the file.[1] This is typically the case when mass spectra are acquired all along a chromatography run and is generally called a *"profile"* mode acquisition. Note that this profile mode acquisition must not be mistaken as the profile mass peak type that negates the centroid mass peak type.

1.4 Mass Data Visualisation: To Combine or Not To Combine?

In the previous section, we mentioned *spectrum combination* a number of times. What does that mean, that spectra are *"combined"* together into a single *"combined spectrum"*? Say we have 200 spectra that need to be combined together into a *single* spectrum that summatively represents the data of these 200 spectra.

First, a new spectrum would be allocated *(result spectrum)*, entirely empty at first. Then, the very first spectrum of the 200 spectra is literally copied into that result spectrum. At this point the combination occurs, according to an iterative process that has the following steps:

- Pick the next spectrum of the 200-spectra dataset;
    1. Pick the first (m/z,i) pair of the currently iterated spectrum;
    2. Look up in the *result spectrum* if a m/z value identical to the m/z value of the current (m/z,i) pair is already present;
    3. If the m/z value is found, increment its intensity by the intensity of the (m/z,i) pair;
    4. Else, if the m/z value is not found, add the current (m/z,i) pair to the result spectrum;
    5. Iterate over all the remaining (m/z,i) pairs of the current spectrum and redo these steps.
- Iterate over all the 198 remaining spectra of the dataset and do the steps above for each single iterated spectrum.

At the end of the two nested loops above, the combined spectrum is still a single spectrum that represents—summatively—all the 200 spectra. This whole process is very computing-intensive, in particular if:

1 Although there certainly is spectrum combination going on in the guts of the software, because the system actually acquires much more spectra than is visible on screen and each newly displayed spectrum is actually the combination of many spectra acquired under the surface.

- The m/z range is large: there are lots of points in each spectrum, which means that for each new (m/z,i) pair we need to iterate in the long list of m/z values that make the result spectrum;

- The resolving power of the mass spectrometer is high: there are many points per m/z range unit.

When a profile mode acquisition is performed, the user gets an innumerable number of distinct spectra, all appended to a single file. These unitary spectra are virtually unusable if an initial processing is not performed. This initial processing of the spectra is called *"total ion current chromatogram calculation"*. What is it? Let's say that the user has performed a profile mode mass spectrometry acquisition on the eluate of a chromatography column. Now, imagine that the spectrometer stores the mass data at a rate of one spectrum per second and that the chromatography gradient develops over 45 min: there would be a total of (45 * 60) spectra in that file. The question is: — *"How can we provide the user with a data representation that might be both meaningful and useful to start exploring the data?"* The conventional way of doing so is to load all the mass spectra and compute the *"total ion current chromatogram"* (the TIC chromatogram). The analogy with chromatography is evident: the TIC chromatogram is the same as the UV chromatogram unless optical density is not the physical property that is measured over time; instead, the amount of ions that are detected in the mass spectrometer is measured over time. That amount is actually the sum of the intensities of all the (m/z,i) pairs detected in each spectrum. When mass data are acquired during a chromatography run, often, the total ion current chromatogram mirrors (mimicks) the UV chromatogram[2]. For each retention time (RT) a TIC value is computed by summing the intensities of all the (m/z,i) pairs detected at that specific RT.

How is this total ion current chromatogram computed? This is an iterative process: from the first spectrum (retention time value 0 s), to the second spectrum (retention time value 1 s) up to the last spectrum (retention time 45 min), the program computes the sum of the intensities of all the spectrum's (m/z,i) pairs. That computation ends up with a map that relates each RT value with the corresponding TIC value. The TIC chromatogram is nothing but a plot of the TIC values as a function of RT values. In that sense, it is indeed a chromatogram.

MineXpert works exactly in this way. When mass spectrometry data are loaded from a file, the TIC chromatogram is computed and displayed. This TIC chromatogram serves as the basis for the mass data exploration, as described in this manual. The TIC chromatogram serves as the basis for spectral combinations that can be performed in various ways, and not all formally *combinations*, which is why I prefer the term *"integrations"*. Some of these integrations are described below:

- Integrating data from the TIC chromatogram to a single mass spectrum;

- Integrating data from the TIC chromatogram to a single drift spectrum;

2 Unless eluted analytes do absorb UV light but do not either desorb/desolvate or ionize, or both.

Note that the reverse actions are possible (and indeed necessary for a thorough data exploration): selecting a region of a mass spectrum and asking that the TIC chromatogram be reconstituted from there; or selecting a region of a drift spectrum and asking that the TIC chromatogram be reconstituted from there also. Finally, integrations may, of course, be performed from a mass spectrum to a drift spectrum, and reverse.

1.5 Examples of Various Mass Spectral Data Integrations

In the sections below, the inner workings of MineXpert are described for some exemplary mass data integrations. For example, when doing ion mobility mass spectrometry data exploration, it is essential to be able to characterize most finely the drift time of each and any analyte. Since each analyte is actually defined as one or more (m/z,i) pairs, it is essential to be able to ask questions like the following:

- What is the drift time of the ions below this mass peak?
- What are all the drift times of all the analytes going through the mobility cell for a given retention time range?
- What are all the ions that are responsible for this shoulder in the drift spectrum?

1.5.1 TIC -> MZ integration

What computation does actually MineXpert do when a mass spectrum is computed starting from a TIC chromatogram region, say between retention time RT minute 7 and RT minute 8.5?

1. List all the mass spectra that were acquired between RT 7 and RT 8.5. In this *spectral set*, there might be many hundreds of spectra that match this criterion, if we think that, in ion mobility mass spectrometry, ≈ 200 spectra are acquired and stored individually every second (I mean it, every 1 s time lapse);
2. Allocate a new empty spectrum—the *"combined spectrum"*—and copy into it without modification the first spectrum of the spectral set;
3. Go to the next spectrum of the spectral set and iterate into each (m/z,i) pair:
    - Check if the m/z value of the iterated pair is already present in the combined spectrum. If so, increment the combined spectrum's (m/z,i) pair's intensity value by the intensity of the iterated (m/z,i) pair's intensity. If not, simply copy the iterated (m/z,i) pair in the combined spectrum;
    - Iterate over all the remaining (m/z,i) pairs and perform the same action as above.
4. Iterate over all the remaining spectra of the spectral set and perform step number 3.

MineXpert then displays the combined spectrum.

1.5.2 TIC -> DT integration

What computation does MineXpert actually do when a drift spectrum is computed starting from a given TIC chromatogram region, say between retention time RT minute 7 and RT minute 8.5?

What is a drift spectrum? A drift spectrum (mobilogram) is a plot where the cumulated ion current of the detected ions is plotted against the drift time at which they were detected. Let's see how that computation is handled in MineXpert:

1. Create a map to store all the (drift time, intensity) pairs that are to be computed below, the (dt,i) map;

2. List all the mass spectra that were acquired between RT 7 and RT 8.5. The obtained list of mass spectra is called the *"spectral set"*;

3. Go to the first spectrum of the spectral set and compute its TIC value (sum of all the intensities of all the (m/z,i) pairs of that spectrum). Get the drift time value at which this mass spectrum was acquired. We thus have a value pair: (dt, i), that is, for drift time *dt*, the intensity of the total ion current is *i*;
   At this point, we need to do a short digression: we saw earlier that, at the time of this writing, one of the commercial instruments on which the author of these lines does his experiments stores 200 spectra each second. These 200 spectra actually correspond to the way the drift cycle is divided into 200 bin (time bins). That means that in the retention time range [7–8.5], there are (1.5*60) complete drift cycles. And thus there are (1.5*60) spectra with drift time x, the same amount of spectra with drift time y, and so on for the reminaing 198 time bins. Of course, a large number of these spectra might be almost empty, but these spectra are there and we need to cope with them.
   The paragraph above must thus lead to one interrogation about the current (dt,i) pair: — *"Has the current dt value be seen before, during the previous iterations in this loop?"*. If not, then create the (dt, i) pair and add it to the (dt,i) map; if yes, get the *dt* element in the map and increment its intensity value by the TIC value computed above;

4. Iterate over all the remaining spectra of the spectral set and perform step number 3.

At the end of the loop above, we get a map in which each item relates a given drift time with a TIC value. This can be understood this way: — *"For each drift time value, what is the accumulated ion current of all the ions having that specific drift time?"*.

At this point, MineXpert displays the drift spectrum (mobilogram).

1.5.3 Using multiple threads during mass data integrations

Whenever the computer has multiple threads available for the integrations to be performed using parallel execution, MineXpert manages to use all these available threads. It is possible to limit the number of threads used for the integration computations as described in Preferences menu, Section 3.3, "The Main Program Window Menu".

1.6 Installation of the software

The installation material is available at http:/msxpertsuite.org.

1.6.1 Installation on MS Windows and macOS systems

The installation of the software is extremely easy on the MS-Windows and macOS platforms. In both cases, the installation programs are standard and require no explanation.

1.6.2 Installation on Debian- and Ubuntu-based systems

The installation on Debian- and Ubuntu-based GNU/Linux platforms is also extremely easy (even more than in the above situations). MineXpert is indeed packaged and released in the official distribution repositories of these distributions and the only command to run to install it is:

```
$ [3] sudo apt install <package_name> RETURN
```

In the command above, the typical *package_name* is in the form `minexpert3` for the program package and `minexpert3-doc` for the user manual package.

Once the package has been installed the program shows up in the *Science* menu. It can also be launched from the shell using the following command:

```
$ minexpert3 RETURN
```



Tip

If the Debian system onto which the program is to be installed is older than *testing*, that is, older than *Buster (Debian 10)*, then using the AppImage program bundle might be a solution. See below for the method to run MineXpert as an AppImage bundle.

3 The prompt character might be % in some shells, like zsh.

1.6.3 Installation with an AppImage software bundle

The *AppImage* software bundle format is a format that allows one to easily run a software program on any GNU/Linux-based distribution. From the http:/appimage.org/:

> The key idea of the AppImage format is one app = one file. Every AppImage contains an app and all the files the app needs to run. In other words, each AppImage has no dependencies other than what is included in the targeted base operating system(s).
>
> —Simon Peter

There are AppImage software bundles for the various MineXpert versions that are available for download. As of writing, the software bundle has been tested on Centos version 8.3.2011 and on Fedora version 22. These are pretty old distribution versions and thus MineXpert should also run on more recent versions of these computing platforms. The AppImage bundle of MineXpert was created on a rather current Debian version: the *testing* Debian 11-to-be distribution.

In order to run the MineXpert software AppImage bundle, download the latest version (like `mineXpert2-0.7.4-x86_64.AppImage`). Once the file has been downloaded to the desired directory, change to that directory and change the permissions to make it executable:

```
$ chmod a+x mineXpert2-0.7.4-x86_64.AppImage RETURN
```

Finally, execute the file that has become a normal program:

```
$ ./mineXpert2-0.7.4-x86_64.AppImage RETURN
```



Tip

If the program complains about a locale not being found, please, modify the command line to read:

```
$ LC_ALL="C" ./mineXpert2-0.7.4-x86_64.AppImage RETURN
```

1.7 Building the software from source

The MineXpert software build is under the control of the CMake build system. There are a number of dependencies to install prior to trying to build the software, as described below.

1.7.1 The dependencies required to build MineXpert

The dependencies to be installed are listed here with package names matching the packages that are in Debian/Ubuntu. In other RPM-based software, most often the package names are similar, albeit with some slight differences.

DEPENDENCIES

The build system

cmake

Conversion of `svg` files to `png` files

graphicsmagick-imagemagick-compat

For the parallel computations

libgomp1

For the isotopic cluster calculations

libisospec++-dev

For all the raw mass calculations like the data model, the mass spectral combinations…

libpappsomspp-dev, libpappsomspp-widget-dev

For all the plotting

libqcustomplot-dev

For the C++ objects (GUI and non-GUI)

qtbase5-dev, libqt5svg5-dev, qttools5-dev-tools, qtchooser

For the man page

docbook-to-man

For the documentation (optional, with `-DMAKE_USER_MANUAL=1` as a flag to the call of cmake, see below.)

daps, libjeuclid-core-java, libjeuclid-fop-java, libjs-jquery, libjs-highlight.js, libjs-mathjax, fonts-mathjax, fonts-mathjax-extras, texlive-fonts-extra, fonts-ebgaramond-extra

1.7.2 Getting the source tarball

In the example below, the version of the software to be installed is `7.3.0`. Replace that version with any latest version of interest, which can be looked for at https://gitlab.com/msxpertsuite/minexpert3/-/releases.

1.7.2.1 Using git

The rather convoluted command below only downloads the branch of interest. The whole git repos is very large...

```
$ git clone https://gitlab.com/msxpertsuite/minexpert3.git --branch master/7.3.0-1 --single-branch minexpert3-7.3.0
```

1.7.2.2 Using wget to download the tarball

```
wget https://gitlab.com/msxpertsuite/minexpert3/-/archive/7.3.0/minexpert3-7.3.0.tar.gz
```

Untar the tarball, which creates the `minexpert3-7.3.0` directory:

```
tar xvzf minexpert3-7.3.0.tar.gz
```

1.7.3 Building of the software

Change directory:

```
$ cd minexpert3-7.3.0
```

Create a build directory:

```
$ mkdir build
```

Change directory:

```
$ cd build
```

Configure the build:

```
$ cmake ../ -DCMAKE_BUILD_TYPE=Release
```

Build the software:

```
$ make
```

2 Configuring the application preferences

A large application like MineXpert has lots of configurable options. The *Application preferences* window described in this chapter is the central place for these customizations. To enter the application preferences window, select *Preferences* › *Edit preferences*.

2.1 Setting the general options

As shown in Figure 2.1, "Application preferences - General page", the general options encompass the following notions:

- *Maximum number of threads:* Set the maximum number of threads when performing resource intensive tasks, like mass spectrum integration.
- *Decimal places for number representation:* Set the numer of decimal digits when representing numbers. Note that not all numbers need the same decimal precision.

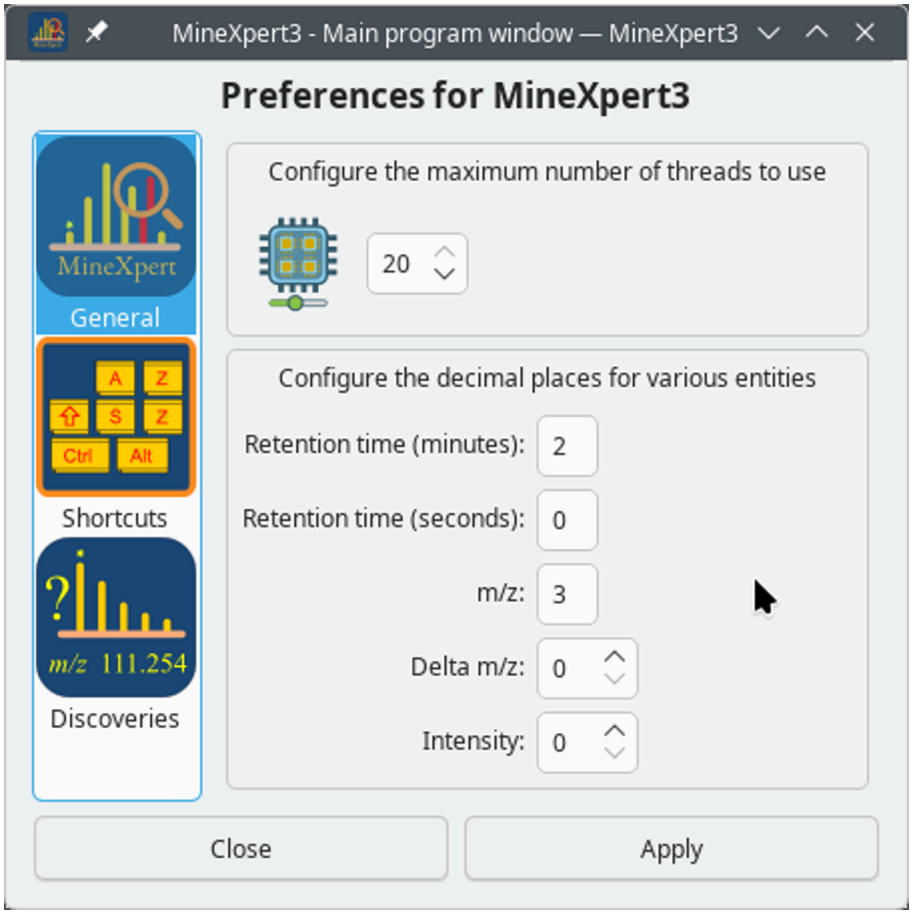


FIGURE 2.1: APPLICATION PREFERENCES - GENERAL PAGE

2.2 Setting the shortcuts

As shown in , it is possible to configure the shortcuts for a large number of tasks throughout the application.



	Context	Category	Unit	Shortcut	Label
1	Global	Integrations	Any	Ctrl+D, T	Duplicate trace
2	Global	Integrations	Any	Ctrl+I, I	Integrate to a TIC intensity value
3	Global	Integrations	Any	Ctrl+I, R	Integrate to a TIC/XIC chromatogram
4	Global	Integrations	Any	Ctrl+I, M	Integrate to a mass spectrum
5	Global	Integrations	Any	Ctrl+I, Ctrl+M	Integrate to a mass spectrum / TIC chromatogram color map
6	Global	Integrations	Any	Ctrl+I, Ctrl+R	Integrate to an ion mobility / TIC chromatogram color map
7	Global	Integrations	Any	Ctrl+I, Ctrl+D	Integrate to an ion mobility / mass spectrum color map
8	Global	Integrations	Any	Ctrl+I, D	Integrate to an ion mobility spectrum
9	Global	Integrations	OpenMzIntegr...	Ctrl+O, I	Open m/z integration parameters dialog
10	Global	Menus	CompositeWid...	Ctrl+P, M	Plot widget menu
11	Global	Menus	PlotWindow	Ctrl+P, Ctrl+A	Pin-down all the widgets
12	Global	Menus	PlotWindow	Ctrl+W, M	Plot window menu
13	Global	Menus	PlotWindow	Ctrl+R, Ctrl+E	Remove all the empty plot widgets
14	Global	Menus	PlotWindow	Ctrl+R, Ctrl+A	Remove all the plot widgets
15	Global	Menus	PlotWindow	Ctrl+Shift+R, ...	Remove all the plot widgets and descendants
16	Global	Menus	PlotWindow	Ctrl+R, Ctrl+P	Remove pinned-down plot widgets
17	Global	Menus	PlotWindow	Ctrl+R, Ctrl+D	Remove pinned-down plot widgets and descendants
18	Global	Menus	PlotWindow	Ctrl+R, S	Reset all the plot widget sizes
19	Global	Menus	PlotWindow	Ctrl+S, Ctrl+P	Save pinned plot(s) to graphics file
20	Global	Menus	PlotWindow	Ctrl+F, P	Set MS fragmentation spec for all pinned-down widgets
21	Global	Menus	PlotWindow	Ctrl+I, P	Set m/z integration params for all pinned-down widgets
22	Global	Menus	PlotWindow	Ctrl+U, Ctrl+A	Un-pin all the widgets
23	MainWindow	File	LoadMSFromC...	Ctrl+L, C	Load mass spectrum from clipboard
24	MainWindow	File	OpenMSFullIn...	Ctrl+O, M	Open mass spectrum file(s) fully in memory
25	MainWindow	File	OpenMSInStre...	Ctrl+O, S	Open mass spectrum file(s) in stream

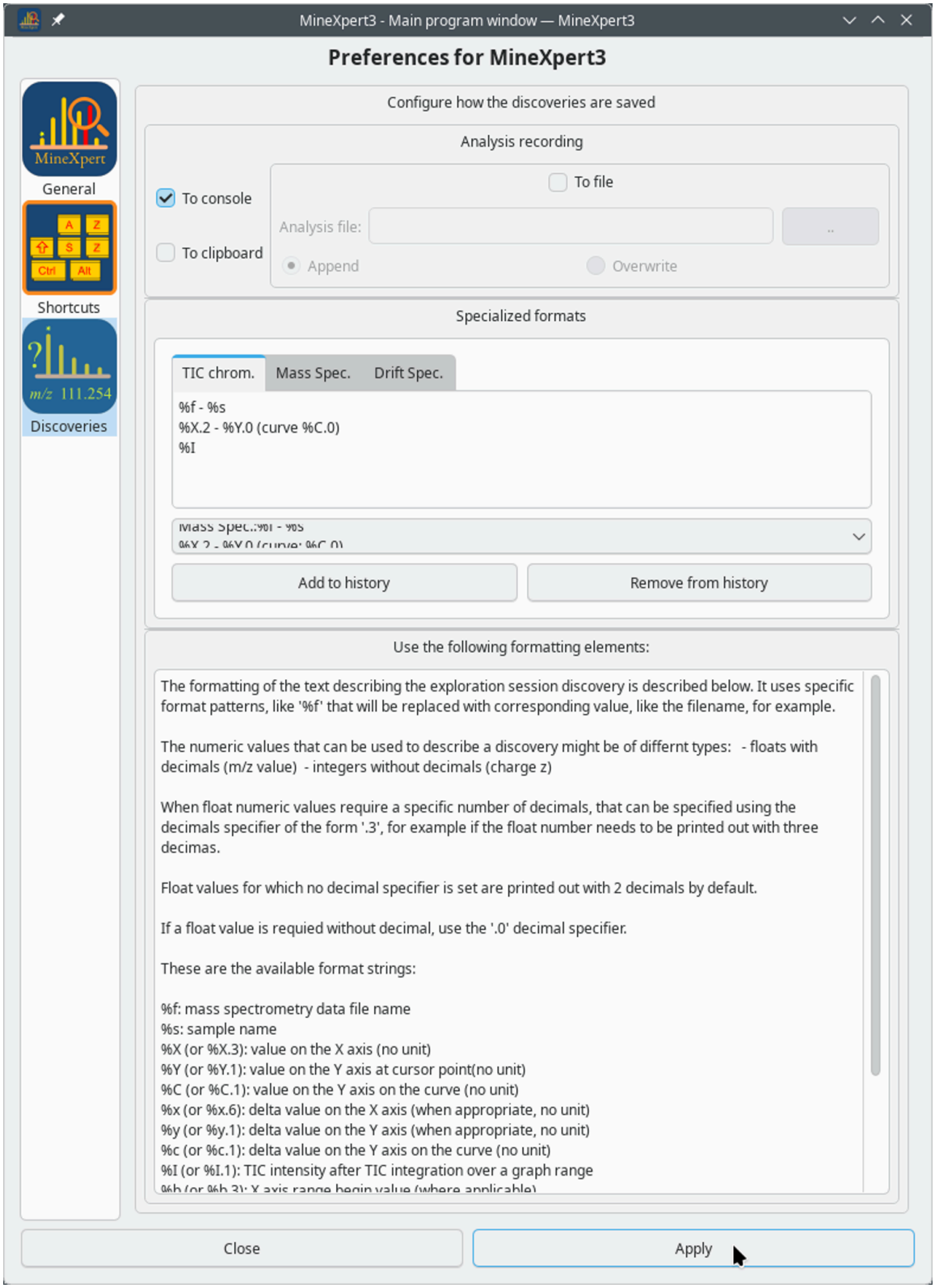
FIGURE 2.2: **APPLICATION PREFERENCES - SHORTCUTS PAGE**

The shortcuts can be configured by double-clicking the item of interest in the cell of the *Shortcut* column and then entering the desired shortcut. Then simply click somewhere else.

When all the shortcuts have been configured, click the button labeled *Apply* and restart the program.

2.3 Setting the recording of the discoveries

As shown in Figure 2.3, "Application preferences - Discoveries page", it is possible to configure the way discoveries are stored.



The text format specification configures the output of the text describing the disovery (see text). That specification is defined for each one of the three main graphs: TIC/XIC chromatogram, mass spectrum and drift spectrum.

FIGURE 2.3: **APPLICATION PREFERENCES - DISCOVERIES PAGE**

The detailed configuration of the discoveries storage is described in Chapter 8, Recording data exploration discoveries .

3 The program's graphical user interface

Data exploration, in mass spectrometry, entails, for a large part, the relentless scrutiny of the mass spectra by an expert eye. Without a powerful mass spectrum viewer, capable of numerous data display modes, the expert eye remains powerless. In this chapter, the graphical user interface is described in detail, starting from the opening of a mass spectrometry data file, to the navigation through all the data depth within a large set of graphical representations of the data.

3.1 Opening Mass Spectrometry Data Files

To start a session, open one or more mass spectra using the menu *File* › *Open mass spectrum file(s) in streamed mode* menu. MineXpert understands the mzML, mzXML and MGF formats. The file loading procedures, for these formats, are delegated to the excellent `libpwiz` library from the ProteoWizard project.[1] Simple txt, asc data where m/z and intensity values are separated by any character that is neither a newline nor a dot nor a digit (loading is handled by a private parser) can be loaded either from file or directly from the clipboard.

The timsTOF successful line of instruments from Bruker produces huge mass spectrometry data files. We provide the user with a mechanism to probe the data and configure the way they should be loaded. When MineXpert detects that such a file is being loaded, it first probes the complexity of the data in it and open the dialog window shown in Figure 3.1, "Probing the specific timsTOF MS runs".

1 Please, see http://proteowizard.sourceforge.net/



The contents of the dialog window are self-explanatory.

FIGURE 3.1: **PROBING THE SPECIFIC TIMSTOF MS RUNS**

There are two actions that can be triggered from the dialog window shown in Figure 3.1, "Probing the specific timsTOF MS runs":

- *Flatten IM scans*: this option will compute one single mass spectrum for each retention time by combining together all the ion mobility spectra acquired at that time (there can be as many as a thousand ion mobility spectra).
- *Keep IM scans*: this option will keep all the ion mobility scans.



Note

The *Keep IM scans* option will force the software to load such a huge amount of spectra for a LC-MS/MS acquisition that this is going to be virtually impossible. The only way to actually load all the IM scans (spectra) is to limit the scope of the data loading using the various parameters configurable in the dialog window shown in Figure 3.1, "Probing the specific timsTOF MS runs".

There are two variants of the mass spectrometry file opening menu, one for which the mass data are read from file and stored *totally* in memory and one for which the mass data are read from file in *streamed mode*, used to compute the TIC chromatogram and discarded. The latter mode is useful when the mass data are so large that they cannot fit in memory. Optionally, a table view can be displayed with all the data in the MS run data set that was loaded from disk. The TIC chromatogram or the table view can then be used to access the mass data in the file according to criteria set by the user (retention time range, for example).

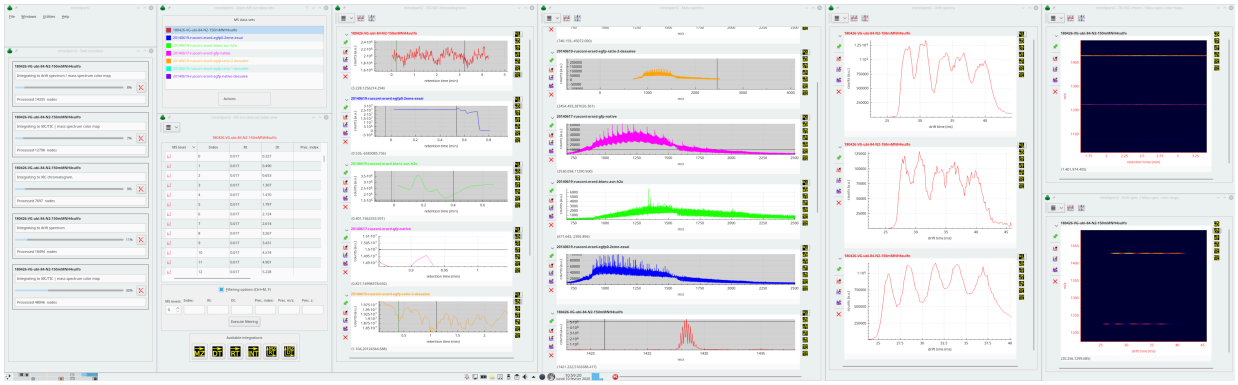
3.2 The Window Layout

The graphical interface of comprises a number of windows where data and informations are displayed. When first started, the program shows the main program window. Upon loading of a MS run data set from file, new windows do show up as described below (see Figure 3.2, "General View of the Graphical User Interface"):

- *mineXpert3* main program window: this is an unintrusive window sporting the main menu;
- The *Task monitors* window, that displays all the ongoing tasks, always providing the user the ability to cancel an ongoing task;
- The *Open MS run data sets* window, that lists all the mass spectrometry data sets that have been loaded from file;
- The *TIC/XIC chromatograms* window[2] where the various TIC/XIC chromatograms are displayed for the various mass spectrometry data files that have been loaded. There is, by definition, a single TIC chromatogram for each MS run data set currently loaded in the program. However, this window will also display ion current chromatograms that are computed as a result of an integration step from the other windows, like from the *Mass spectra* window or from the *Drift spectra* window. In this case, the chromatogram is an extracted ion current chromatogram (XIC chromatogram);
- The *Mass spectra* window, where the various mass spectra are displayed. A given mass spectrum may originate from a TIC chromatogram, from the MS run data set tableview window, from a drift spectrum, or even from a color map;

2 TIC stands for "total ion current" and XIC stands for "extracted ion current".

- The *Drift spectra* window, where the various drift spectra are displayed. Like stated before, drift spectra can originate from a variety of places;
- The *TIC/XIC chrom. /Mass spec. color maps* window, that contains int = f(rt, m/z) color maps, two-dimensional representations of the relation between the retention time and the mass spectral data. The intensities are represented as colors in a gradient map;
- The *Drift spec. /Mass spec. color maps* window, that contains int = f(dt, m/z) color maps, two-dimensional representations of the relation between the drift time and the mass spectral data (in ion mobility mass spectrometry experiments only). The intensities are represented as colors in a gradient map;
- The *Console* window, where the various messages or analysis data elements are displayed for the user to select, copy and paste in an electronic lab-book;



The position and size of all the windows can be stored for a later session

FIGURE 3.2: GENERAL VIEW OF THE GRAPHICAL USER INTERFACE

3.3 The Main Program Window Menu

The menu bar in the main program window displays a number of menu items, reviewed below:

- *File*
  - *File* › *Open mass spectrum file(s) fully in memory*: Select the mass spectrum file(s) to load. The mass spectral data read from disk are stored in memory. This means that if the file is greater than the available system memory (RAM), the user is urged to select the next menu;
  - *File* › *Open mass spectrum file(s) in streamed mode*: Select the mass spectrum file(s) to load. The mass spectral data are read from disk, the TIC chromatogram and the MS run data set statistics are computed and then the data are freed. This file opening mode is compulsory when the mass data file is larger than the available system memory (RAM);

- *File* › *Load mass spectrum from clipboard*: Create a mass spectrum from a textual representation of (m/z,i) pairs in the same format as described above for the txt,asc file format;

- *File* › *Analysis preferences*: Set the preferences for the exploration results storage;

- *Windows*

  The menus are self-explanatory, as they explicitely explain which window is to be shown. The *Save workspace* menu records on disk the position and size of all the windows, so that upon reopening the program, the windows all position themselves at the recorded position and size;

- *Utilities*

  - *Utilities* › *XIC chromatogram extractions*: Open a dialog window to set up XIC extractions for any one of the currently loaded mass spectral data set;

  - *Utilities* › *Isotopic cluster calculations*: Open the window in which the isotopic cluster calculations are configured and run;

  - *Utilities* › *Mass peak shape calculations*: Open the window in which the mass peak shape calculations can be configured and run.

- *Preferences*

  - *Preferences* › *Set max. thread count*: Set the maximum number of execution threads that mass spectral integrations can use. Note that the program starts to be really efficient with a number of threads at least equal to 4 (see Section 1.5.3, "Using multiple threads during mass data integrations").



Tip

Do not mistake "thread" for "core" or "processor". When processors are described, the description usually says "x core". That means that there are x "processors" in the whole central processing unit attached to the motherboard. Each processor most often can branch two different threads. So a computer that sports 2 cores, will most often in fact make 4 threads available for the computations.



Warning

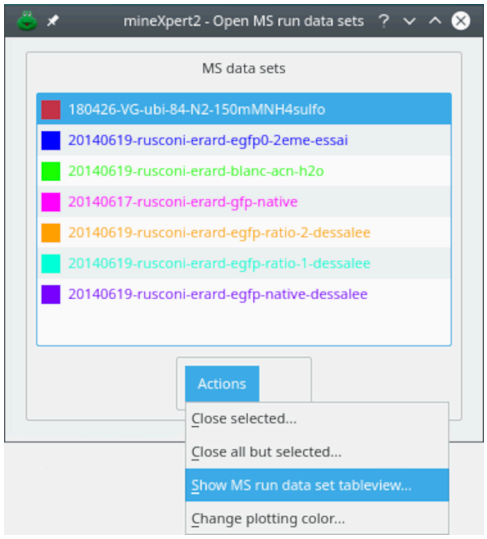
It is of utmost importance to make sure that guest virtualized environments do share at least 2 (better 4) threads with the host system. In VirtualBox, for example, it is possible to allocated a defined number of threads to the guest operating system. Ensure that the number of threads made available to the guest operating system is at least , with 4 being a right low level value.

- *Help*

  This menu's items show help about the program itself and also about the Qt libraries that were used to build it. These informations are essential in case the user wants to make a bug report.

3.4 The Loaded MS Run Data Sets

Each time a new MS run data set is loaded from disk or from clipboard, a corresponding *MS run data set* item is added to the *Open MS run data sets* dialog window (see Figure 3.3, "The open MS run data sets").



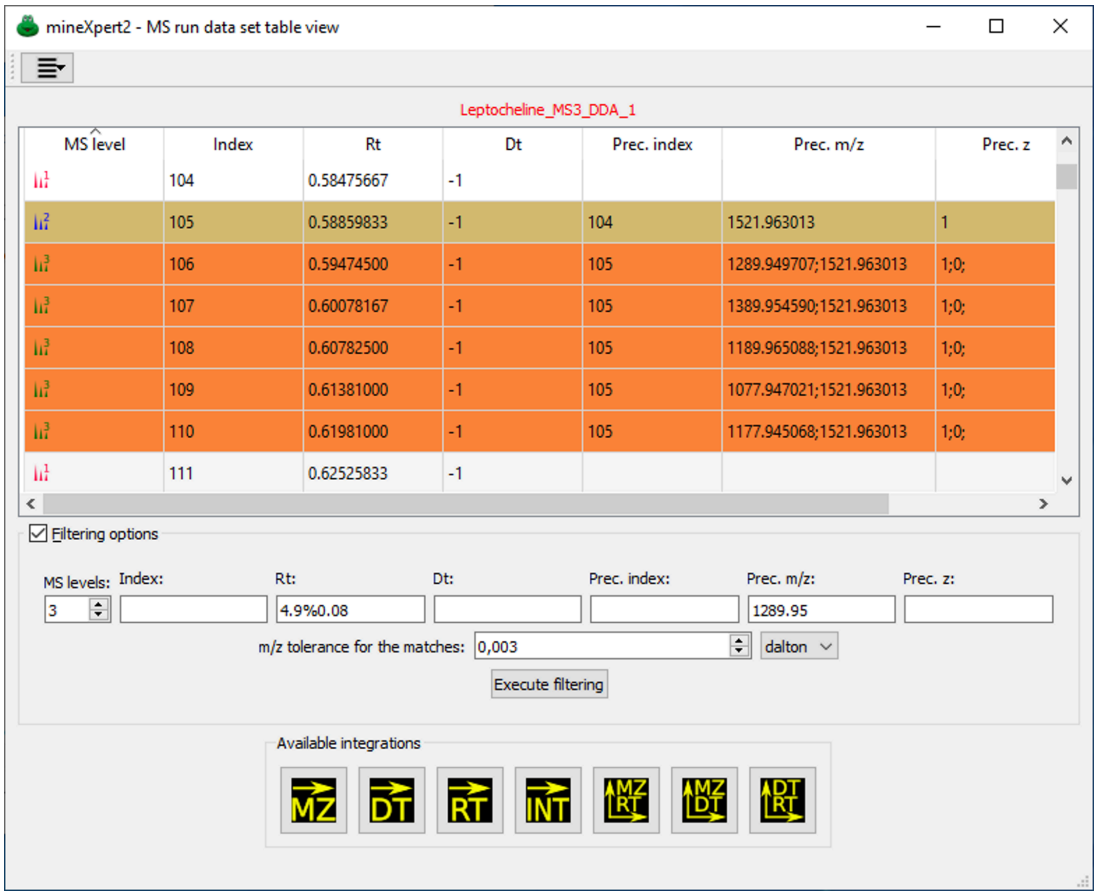
Each time a new MS run data set is loaded, a corresponding item is added to the list widget. The drop-down menu lists a number of actions that can be performed with these items.

FIGURE 3.3: THE OPEN MS RUN DATA SETS

One of the most useful menu items available from the drop-down menu of Figure 3.3, "The open MS run data sets" is the *Show MS run data set table view* item. That menu item triggers the display of a window containing a table view, like in a spreadsheet, that shows the data of each scan contained in the MS run data set loaded from file. This window is described in the following section (see Figure 3.4, "The table view-based scrutiny of one MS run data set").

3.5 Mass Spectral Data Listing in a Table View

From the *Open MS run data sets* dialog window, it is possible to select one list widget item corresponding to the data set of interest and select the *Show MS run data set table view* menu item that opens the *MS run data set table view* window shown in Figure 3.4, "The table view-based scrutiny of one MS run data set".



Each MS run data set might be scrutinized through the use of this table view. Each filtering criterion has a number of data insertion modalities, described in the text. In this example, the user has set a number of values in the *Filtering options* group box but has not yet executed the filtering, so the full mass spectral data set is listed in the table view.

FIGURE 3.4: THE TABLE VIEW-BASED SCRUTINY OF ONE MS RUN DATA SET

The data that belong to each spectrum in the acquired mass spectral data set are shown as values in dedicated columns. Each row represents a mass spectrum of the data set. There might be tens of thousands spectra in an acquisition. Note that two columns may contain ; -separated values. These columns are *Prec. m/z* and *Prec.z*, respectively precursor ion m/z and precursor ion charge. Ions of differing m/z values and of differing charge values might be accumulated in a trap and then fragmented in a single fragmentation step. The mass data file thus contains spectra with a list of selected precursor ions.



Note

When multiple values are listed in precursor m/z and charge cells, these values behave as if they were alone in the cell. When performing filtering of the data (see below), then the program treats each value of the list as if it were alone in the cell. For example, in Figure 3.5, "Mass spectral data set filtering using combined criteria" the user filtered the MS data on the basis of a *Prec. m/z* value of *1289.95*. That filter selected the row that lists, as *Prec. m/z*, the value *1289.952393;1521.973633*.

In the *MS run data set table view* window, the checkable group box labelled *Filtering options* allows one to open a series of widgets in which to set the values for various filtering criteria.

- *MS levels*: Set 0 to retain all the mass spectra, irrespective of their MS level. Enter value 1 to filter out any MS level that is not 1. This fields has no practical limit, allowing one to perform MS^n data exploration.
- *Index*: Filter by the index number of the mass spectra. This may be of interest when the indices are somehow known in advance.
- *Rt*: Filter by retention time, that is, the time at which the mass spectrum was acquired, relative to the start of the mass data acquisition.
- *Dt*: Filter by drift time (ion mobility mass spectrometry generates this kind of data).
- *Prec. index*: Precursor index.
- *Prec. m/z*: Precursor charge.

The general syntax used to fill-in the filtering options is according to the following scheme:

- *simple value*, like 1 for *MS levels*.
- <*value*: keep rows where values are less than the entered value.
- >*value*: keep rows where values are greater than the entered value.
- >*start* <*end*: keep rows where values are in between *start*and *end*. The match is not inclusive.
- *start--end*: like before, but the match is inclusive. Mind the double-dash that delimits the values of the range.
- *value%tolerance*: equivalent to the preceding item, unless *start* is (*value- tolerance*) and *end* is (*value + tolerance*).



Warning: The specific case of precursor m/z values

When inserting a m/z value, there is a specific tolerance widget *m/z tolerance for the matches* that allows one to define the tolerance with which the matches for the precursor m/z values are performed. That mechanism does not override the *value%tolerance* value definition modality described above; instead, it will add another tolerance to the range values.

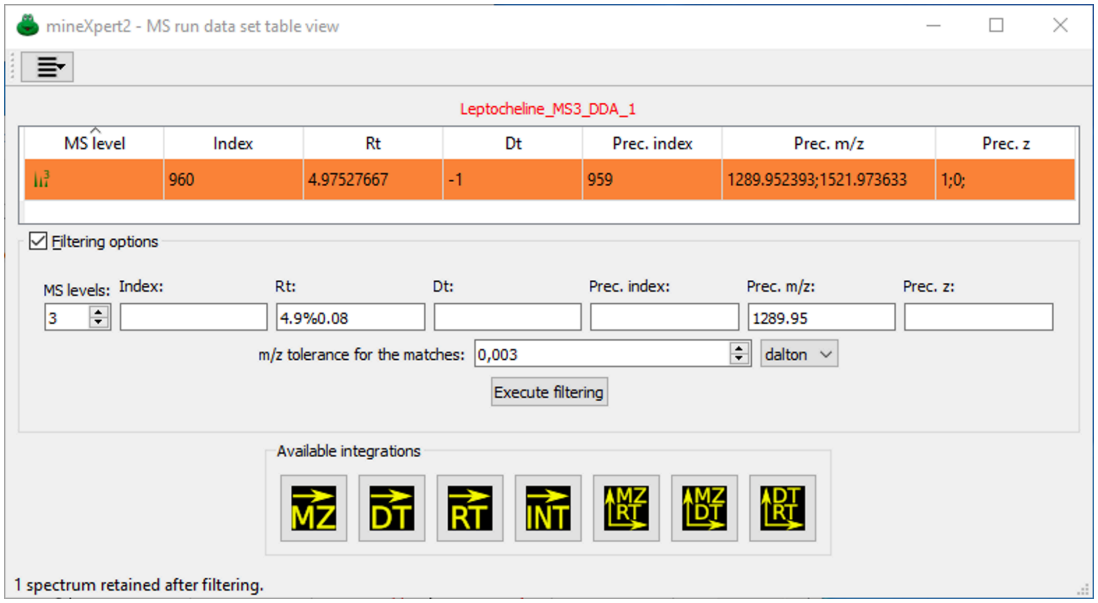
Once the filtering values have been entered and validated by clicking *Execute filtering* or by keying-in Ctrl–ENTER, the items (that is, the table rows) that verified the filter, still present in the table view, might be selected (all of them or only some of them) for their integration to a variety of destinations. The destination of each integration is selected by clicking the proper icon push button from the button row below. The meaning of these push buttons will be described later (see Section 3.7.2, "The Right Column of Buttons Configures the New Integration to Run").



Tip

Make sure to set the m/z integration parameters before performing an integration to a mass spectrum. The setting dialog window opens up using the *Open m/z integration parameters dialog* menu item from the window's main menu, at the top left corner of the window. If the bin size value is too small, the integration might last very much longer than for a "reasonable" size.

The user might enter as many filtering criteria as necessary to pin point the most elusive feature in the mass spectral data set. In Figure 3.5, "Mass spectral data set filtering using combined criteria", the user has filtered the data to retain MS^3 mass spectra obtained by fragmenting ions of a given m/z value using a specified tolerance for the m/z match in Dalton units.



The use of combined filtering criteria allows one to pin point elusive mass spectral features.

FIGURE 3.5: **MASS SPECTRAL DATA SET FILTERING USING COMBINED CRITERIA**

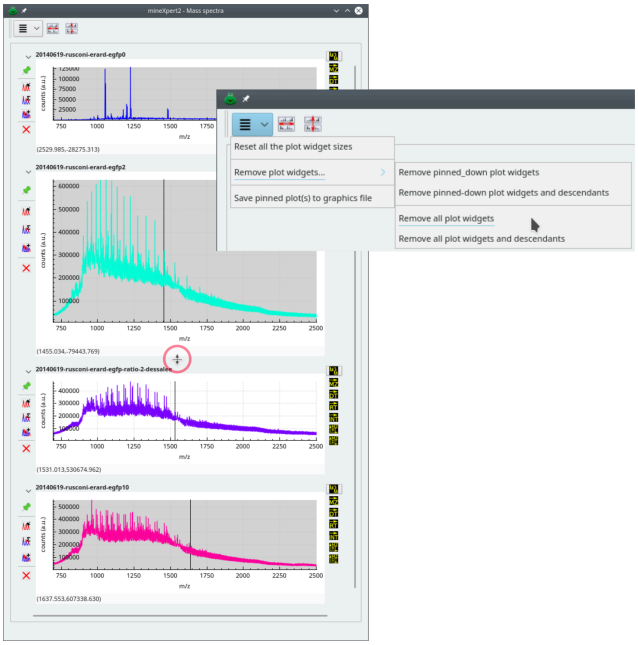
3.6 The Main Data-Plotting Windows

This section will succinctly describe the main data windows of MineXpert. Each window will be described in greater detail when the features of the program will be described.



Tip

Although not visible, the various plot widgets in the various plot windows are glued together with splitters that allow their resizing if the window becomes too crowded, as shown in Figure 3.6, "Resizing the plot widgets with the splitter".



When positioning the mouse cursor between two plot widgets, the cursor switches to the splitter cursor shape (circled in red). Dragging that cursor will widen/shrink the plot widgets. To reset the plot widgets as they were initially created, use the *Reset all the plot widget sizes* from the main window menu shown on the right.

FIGURE 3.6: **RESIZING THE PLOT WIDGETS WITH THE SPLITTER**

As visible in all the figures of the plot widget-containing windows described in the next sections, the general structure of these windows is to have a menu in the toolbar, icon buttons in that same toolbar and then plot widgets stacked below in the order in which they have been created.



Note: The multi-graph plot widget has gone

In the first version of MineXpert the plot window was divided in two parts; the upper part had a plot widget in which all the plots were overlaid, the lower part had as many plot widgets as there had been integrations because each plot widget could only have one plot in it.

This is no more true now, because, as will be described later, it is now extremely easy to duplicate any plot into any plot widget so as to overlay any number of plots into any number of plot widgets.

3.6.1 The TIC Chromatogram Window

Each time a new mass spectrum file is loaded, its corresponding TIC chromatogram is computed and then displayed in a new plot widget in the *TIC/XIC chromatogram window* (Figure 3.7, "The total ion current (TIC) chromatogram window"). That window is indeed specialized in the plotting of TIC or XIC chromatograms. Each new TIC chromatogram plot that is generated as a result of the loading of a mass spectrometry data file is plotted using a new color. That color encodes the filiation of the whole set of plots that are generated while performing various integrations starting from that initial TIC chromatogram plot. For example, a red TIC chromatogram plot that serves as the starting point for a mass spectrum integration triggers the creation of a new mass spectrum trace plot that is plotted in red color. For color map windows, where there is not trace plot, the axis and tick labels of the plot are of the same color as that of the initial TIC chromatogram plot.



As many traces as necessary can be shown in the window. Each plot has its own set of crosshair markers.

FIGURE 3.7: THE TOTAL ION CURRENT (TIC) CHROMATOGRAM WINDOW



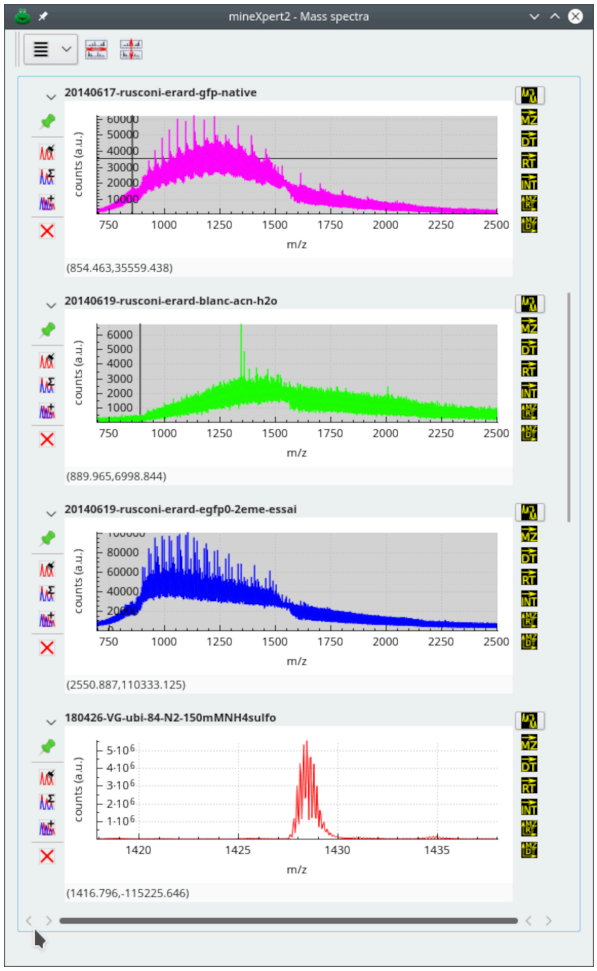
Warning: Unconventional situations while loading mass spectrometry data files

- When the user loads mass spectrometric data from a non-profile acquisition data file or from clipboard data, like when a mass spectrum is opened from a txt,asc,xy text-based format file where the data correspond to a *single* spectrum, *not* a sequence of spectra, the TIC chromatogram really has a single (rt,i) pair denoting the TIC intensity at the single retention time of that very unique spectrum. The TIC chromatogram window thus artificially creates and displays a TIC chromatogram that is a simple line, like shown in the bottom TIC chromatogram plot of Figure 3.7, "The total ion current (TIC) chromatogram window". From there, integration happens like in conventional situations. Make sure that the mouse drag movement encompasses the whole TIC region by first unzooming a bit the trace.

- When loading data files that only contain MS^n data (n > 1), like the proteomics-oriented *Mascot generic files* (MGF), MineXpert cannot compute a TIC chromatogram. Indeed, a TIC chromatogram can only be computed for MS data, not MS^n data. In this case, the TIC chromatogram plot that is created does not contain any plot. The only way the user can start exploring such kind of MS^n (n > 1) data is by using the MS run data set table view window (Section 3.5, "Mass Spectral Data Listing in a Table View").

3.6.2 The Mass Spectrum Window

The mass spectrum window is specialized with the plotting of mass spectra. It thus contains all the plot widgets that display all the mass spectra that were created as a result of mass spectral data integrations. These integration might have originated in any plot widget of any kind (mass spectrum, drift spectrum, color map, for example). For example, the user might select a region in a TIC chromatogram and then ask that a mass spectrum integration be computed. In this case, the resulting mass spectrum is displayed in a new plot widget that is located in the mass spectrum window (Figure 3.8, "The mass spectrum window").

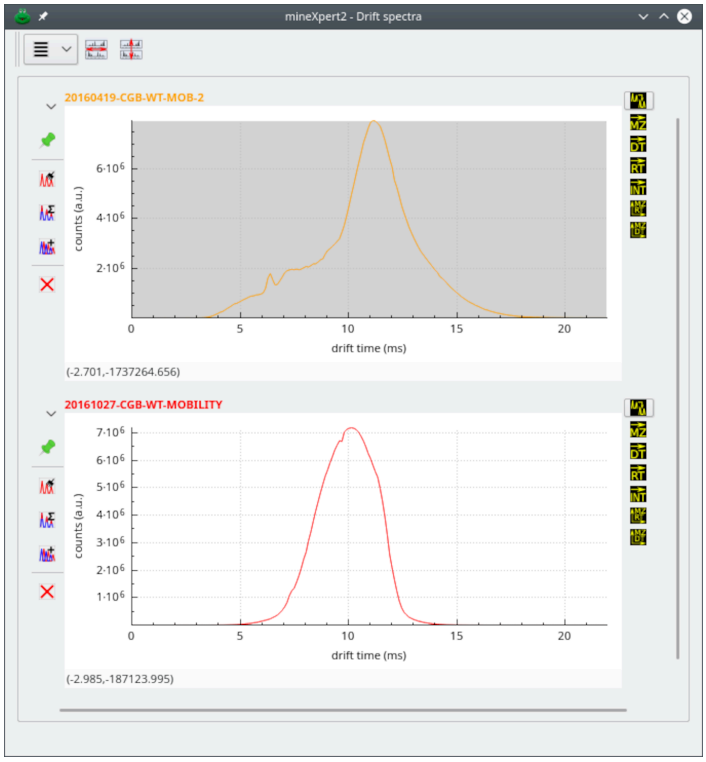


This mass spectrum window shows multiple mass spectra.

FIGURE 3.8: THE MASS SPECTRUM WINDOW

3.6.3 The Drift Spectrum Window

As described for the mass spectrum window, the drift spectrum window contains all the plot widgets that display drift spectra (Figure 3.9, "The drift spectrum window").



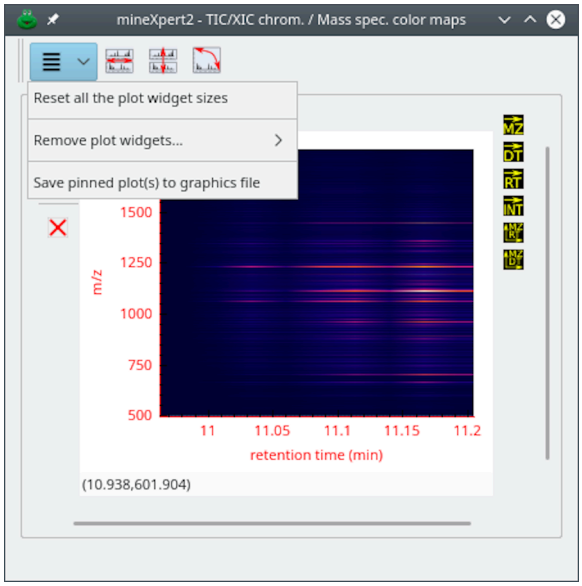
This drift spectrum window shows multiple drift spectra.

FIGURE 3.9: **THE DRIFT SPECTRUM WINDOW**

3.6.4 The Retention Time *vs* Mass Spectrum Color Map Window

The int = f(rt, m/z) color map window displays a heat map relating all the retention time values with the corresponding mass spectra (see Figure 3.10, "The int = f(rt, m/z) color map window"). The window main menu offers menus that are all self-explanatory. The tool bar sports three buttons that:

- Lock the x-axis of all the plot widgets in the window (first button).
- Lock the y-axis of all the plot widgets in the window (first button).
- Transpose the axes of the all the plot widgets that are pinned-down (see Note: The expression "pinned-down widget").

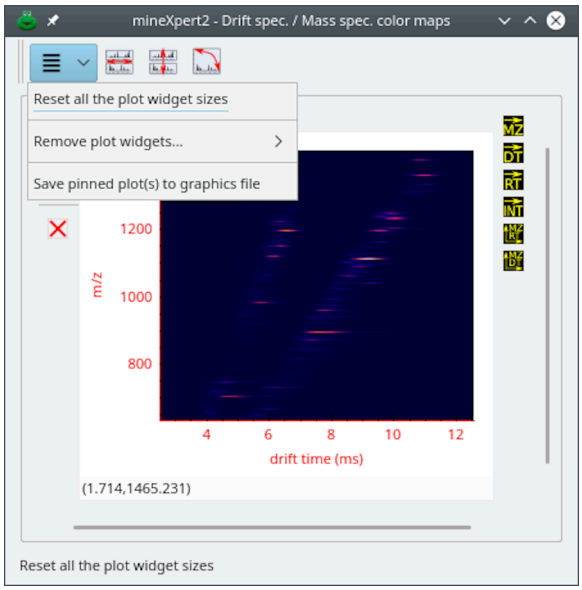


The mass spectrum *vs* retention time color map window relates mass spectra (y-axis) to retention times (x-axis).

FIGURE 3.10: **THE INT = F(RT, M/Z) COLOR MAP WINDOW**

3.6.5 The Drift Time *vs* Mass Spectrum Color Map Window

The int = f(dt, m/z) color map window displays a heat map relating all the drift time values with the corresponding mass spectra (see Figure 3.11, "The int = f(dt, m/z) color map window").

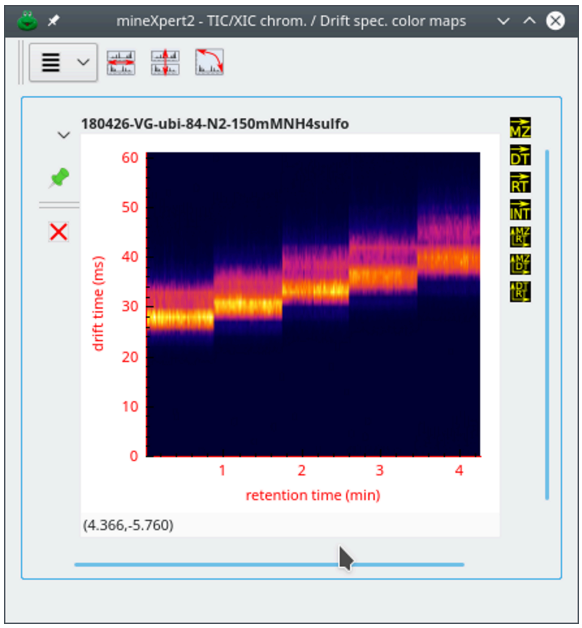


The mass spectrum *vs* drift time color map window relates mass spectra (y-axis) to drift times (x-axis).

FIGURE 3.11: **THE INT = F(DT, M/Z) COLOR MAP WINDOW**

3.6.6 The Drift Time *vs* Retention Time Color Map Window

The int = f(rt, dt;) color map window displays a heat map relating all the drift time values with the corresponding retention time values (see Figure 3.12, "The int = f(rt, dt) color map window").

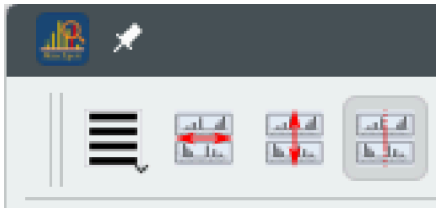


The drift spectrum *vs* TIC chromatogram color map window relates retention times (y-axis) to drift times (x-axis).

FIGURE 3.12: THE INT = F(RT, DT) COLOR MAP WINDOW

3.6.7 Exploring Data in Multiple Plot Widgets

When comparing plots (be them trace plots or color map plots), it is often necessary to align the plots according to the x-axis so as to align features of interest in the various vertically-stacked plot widgets. In such a situation, there is now way to move the cursor through all the plots simultaneously so as to read the x-axis value while marking that position with the crosshair cursor. To enable a vertical mark over all the aligned features in all the plot widgets, click the Figure 3.13, "Activation of the vertical rule" icon.



When this toolbar' third button is checked (the icon shows a dotted vertical red line across multiple widgets), a horizontal slider appears at the top of the window. When that slider is moved, a vertical rule is drawn at its position. The rule is drawn across all the plot widgets in the window.

FIGURE 3.13: ACTIVATION OF THE VERTICAL RULE

If the x-axis range is locked (when first toolbar button in Figure 3.13, "Activation of the vertical rule" with two-arrow-headed red horizontal line is checked) and the vertical rule button is checked also, the vertical slide rule continuously updates the x-axis value near the slider handle. See Figure 3.14, "Using the vertical rule across all the plot widgets" for an illustration.

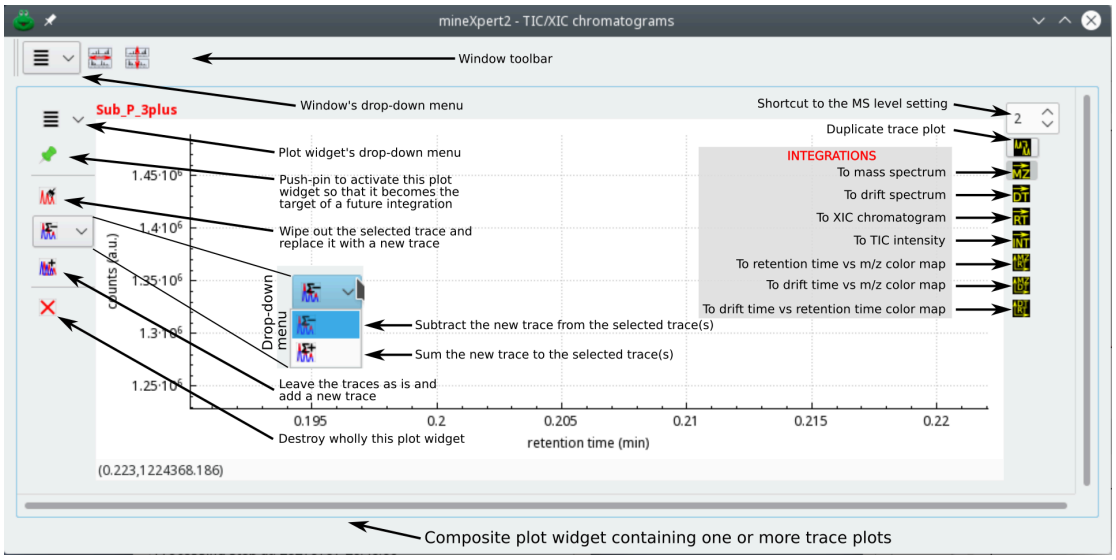


If the x-axis range is locked while having the vertical rule slider moved, a label is shown at the slider handle with the x-axis value at the vertical rule position.

FIGURE 3.14: USING THE VERTICAL RULE ACROSS ALL THE PLOT WIDGETS

3.7 General Structure of the Plot Widgets

The TIC/XIC chromatograms, mass spectra, drift spectra, retention time *vs* mass spectrum and drift time *vs* mass spectrum color map windows (collectively called *plot widget windows*) are all structured in a similar way as described below.



A composite plot widget is a widget that contains a number of other widgets. These sub-widgets are described.

FIGURE 3.15: **DESCRIPTION OF THE VARIOUS WIDGETS THAT MAKE A COMPOSITE PLOT WIDGET**

The window in Figure 3.15, "Description of the various widgets that make a composite plot widget" contains a single composite plot widget. Each composite plot widget's plotting area is sided by two columns of icons, which really are, for most of them, check buttons. The buttons in the left column configure if and how this plot widget is going to be the *receiving container* of a new integration result (result that takes the form of a trace plot). The buttons in the right column configure *what kind of integration* is to be performed using, as a starting point, one of the plots of this plot widget. Indeed, as already mentioned above, integrations in any plot widget might be configured to generate any kind of data: a XIC chromatogram, a mass or drift spectrum... If the data generated by a given integration constitute a XIC chromatogram, that chromatogram will be plotted in the *TIC/XIC chromatograms* window. Both columns of buttons (left and right) are detailed in the next section because these constitute the basis of the operation of MineXpert.

3.7.1 The Left Column of Buttons Configures the Reception of a New Plot

For this description, we will assume that the user is working in the *TIX/XIC chromatograms* window and is performing a new integration to a mass spectrum. The result of that new integration is thus a plot that needs to be installed in the *Mass spectra* window, as depicted in Figure 3.15, "Description of the various widgets that

MAKE A COMPOSITE PLOT WIDGET". Such a window, however, might already contain a number of plot widgets; it is thus necessary to establish where the new plot is going to be created and how it will be created. These two questions are answered in detail below by describing the widgets in the left column of widgets in a composite plot widget.

- The plot widget's drop down menu (downward arrow) will be described in detail later. The menu options are actions that apply to the plot widget that owns the menu;

-  The push-pin button, if it is pushed (therefore colored in red ), indicates that the plot widget "claims" to be the receiving container for any new integration destined to the window containing it.



NOTE: THE EXPRESSION "PINNED-DOWN WIDGET"

A plot widget having its push-pin pushed down is said to be *pinned-down*. This is a terminology that will be used extensively in the rest of this manual.

If the window contains more than one plot widget *and* if more than one plot widget are pinned-down (), then all these plot widgets claim to be the receiving container of any new plot to be installed in the window. If no plot widget in a window is pinned-down, any new plot will be created in its own composite plot widget that will appear at the bottom of the plot widget window.
In the list items below, we assume that the plot widget containing the various check buttons is pinned-down and belongs to a non-color map plot window.[3]

- Checking the  check button indicates that the new plot will *replace* any selected plot in the plot widget.

- Checking the  or the  check button indicates that the new plot will be *subtract-combined* or *add-combined* from or into any selected plot in the plot widget. This feature is particularly useful when trying a step-wise estimation of the mass spectral data out of discrete discontinuous features like TIC chromatogram peaks or drift spectrum peaks. Likewise and reciprocically, it might be useful to extract, from different mass spectral features in a given mass spectrum, the combined TIC chromatogram for them. Typically, when an analyte appears in a mass spectrum under different charge states that are separated by non-baseline regions, being able to assess the XIC chromatogram for them in a series of combined integrations is extremely useful. The subtract-combination is typically used for removing background from a given trace using other blank trace as the background level.

- Checking the  check button indicates that the new plot will be added to the plot widget, that is, it will be overlaid on top of any other plot in the plot widget.

3 Combination of color maps is not yet implemented, although they might be useful in some situations.

3.7.2 The Right Column of Buttons Configures the New Integration to Run

Each plot widget may contain zero or more plots (also called graphs or traces). As mentioned earlier, any plot can serve as the starting point for a new integration to the same or another kind of data. For example, from a plot in the *TIC/XIC chromatograms* window, one might want to integrate a given retention time range to a mass spectrum or to a drift spectrum. How are those destinations determined ? The destination of an integration is selected by pressing one of the buttons in the right column of check buttons described below.

- The top widget of the right hand side column of widgets is a spin box widget where the user can set the MS level of the next integration to be performed. Setting a value in this widget is equivalent to setting the same value in the *MS level* field of the MS fragmentation specification setting dialog window (Figure 4.1, " Setting the MS^N fragmentation parameters " at Section 4.1.2, " Setting the MS^N Fragmentation Parameters ").

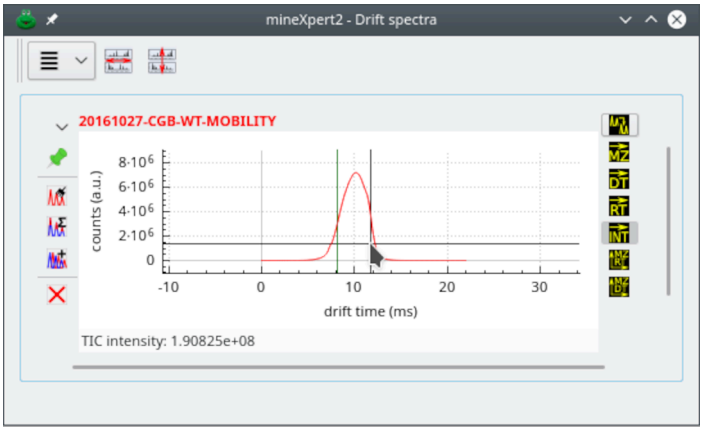
- This button is an exception, because it actually is *not* a check button. It is a true button that, when clicked, elicits the duplication of the currently selected trace(s). There is no integration going on here: the trace(s) are copied "as is" with no processing whatsoever. If none of the plot widgets in the window is pinned-down (that is, all are in the state ; see Note: The expression "pinned-down widget"), the trace(s) are duplicated, each one in its own new plot widget that gets appended to the other widgets at the bottom of the window. If, instead, other plot widget(s) are pinned-down, then the way the duplicated trace(s) will be added depends on the left column buttons. For example, if the is checked, then the duplicated trace(s) will actually be combined *into* any other selected (or the single present) trace in the pinned-down widget.

- This is the check button that, when pressed, directs the new integration to produce a mass spectrum to be added to the *Mass spectra* window. Integration to a mass spectrum implies configuring how the m/z values are to be combined (without or with binning and, if so, the size of the bins). Please, refer to Section 4.1.3, " Setting the m/z Integration Parameters " for a detailed explanation.

- This is the check button that, when pressed, directs the new integration to produce a XIC chromatogram to be added to the *TIX/XIC chromatograms* window. "RT" stands for "retention time".

- This is the check button that, when pressed, directs the new integration to produce a drift spectrum to be added to the *Drift spectra* window. "DT" stands for "drift time".

- This is the check button that, when pressed, directs the new integration to produce a *single TIC intensity value* that is displayed at the bottom of the plot widget, in the widget's status bar.



When using the *single TIC intensity value* function, any feature of a plot will be integrated to a single TIC value that represents the sum of all the TIC values for all the MS data in the MS run data set that match the feature selection. The obtained single TIC intensity value is displayed at the bottom of the plot widget in which the calculation was requested. In this example, the feature that was selected for integration is located between the vertical markers.

FIGURE 3.16: **SINGLE TIC INTENSITY VALUE FOR ANY MASS SPECTRAL DATA FEATURE**

- MZ RT This is the check button that, when pressed, directs the new integration to produce a int = f(m/z,rt) color map. This function is only useful for mass spectrometric data sets that were acquired in "profile" mode; that is, for acquisitions performed along a time frame, like an on-line chromatography-MS experiment.

- MZ DT This is the check button that, when pressed, directs the new integration to produce a int = f(m/z,dt) color map. This function is only useful when the mass spectra were acquired during an ion mobility mass spectrometry experiment.

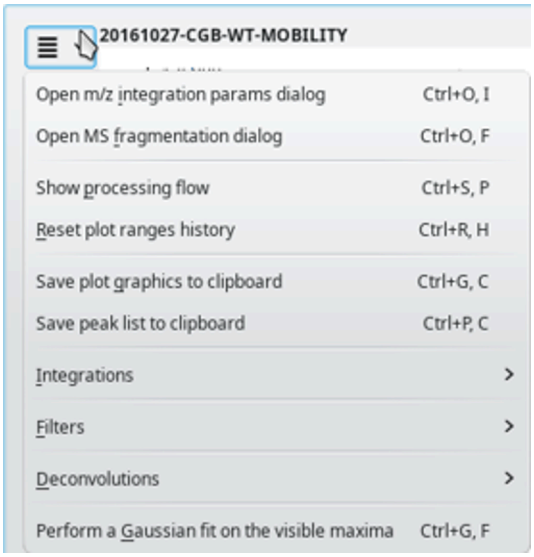
- DT RT This is the check button that, when pressed, directs the new integration to produce a int = f(dt,rt) color map. This function is only useful for mass spectrometric data sets that were acquired in an ion mobility mass spectrometry experiment.

3.7.3 The Plot Widget Main Menu

There are two kinds of plots: the trace plots, like mass spectra, drift spectra, TIC chromatograms, and the color map plots, like the int = f(dt, m/z) color map, for example. In the two cases, the menu is not the same because the features of the plots are not the same.

3.7.3.1 The Trace Plot Widget Main Menu

The trace plot widgets can host any number of traces, be them drift or mass spectra or TIC chromatograms. A trace is a curve-type graph, that is, a two-dimensional graph. The intensity is at the Y axis and the X axis hold either m/z, dt or rt values. The trace plots all share the same main plot widget menu.



The main menu provides actions that may apply only to the plot widget that owns the menu.

FIGURE 3.17: **TRACE PLOT WIDGET MAIN MENU**



Tip

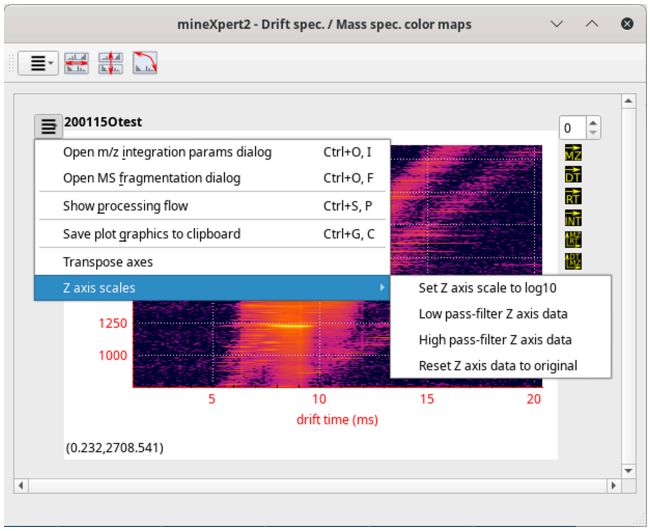
The menu items described below perform actions that apply to individual plots inside the plot widget that owns the menu. MineXpert tries to understand the will of the user: when only one plot is in the widget, the menu action applies to that plot, even if it is not currently selected. As soon as there are more than one plot in the widget, the action requires that the target plot be selected. If no plot is selected, a dialog window will instruct the user to select the specific plot of interest.

- *Open MS fragmentation params dialog*: Open a dialog to set the MS^n requirements for the MS data integration (see Section 4.1.2, " Setting the MS^N Fragmentation Parameters ").
- *Open the m/z integration params dialog*: Open a dialog to set the m/z integration parameters (see Section 4.1.3, " Setting the m/z Integration Parameters ").
- *Show processing flow*: Open a dialog and display the processing flow of the currently selected graph. If there is only one graph in the plot widget, the processing flow for that graph (even if it is not selected) is shown.
- *Reset plot ranges history*: Delete all the plot navigation milestones from the history and start anew with the current plot display. It won't be possible to iterate up in the plot ranges history using the `Backspace` key.

- *Save plot graphics to clipboard*: Save the whole plot widget as a graphics object to the clipboard.
- *Save peak list to clipboard*: Save the currently selected (or unique) trace to the clipboard as numerical data.
- *Integrations* › *Perform single graph point integrations*: integrate the data by moving the cursor over each point of the current trace. Click on the starting trace at the place where point-by-point integration needs to start, then use the F3 to integrate the next plot point.
- *Filter* › *Perform Savitzky-Golay smoothing*: the filter will be run with the configuration settings configured as at Section 4.1.6, " Savitzky-Golay filtering of any kind of data ". The smoothed trace will be added into any pinned-down plot widget of the receiving window. If no plot widget is pinned-down, then a new plot widget will be created.

3.7.3.2 The Color Map Plot Widget Main Menu

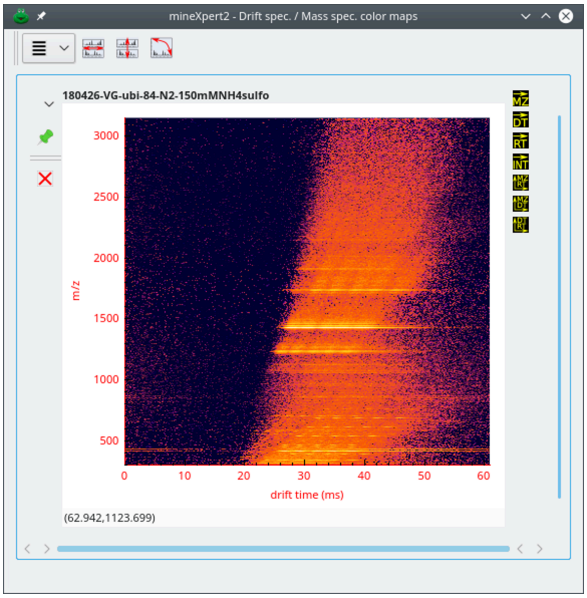
The color map plot widgets can host only one color map since it is not possible to overlay more color maps onto a color map. A color map is a raster-type graph, that is, a bitmap image of which the third dimension is the intensity of the pixel at any given position in the image. That intensity is represented as a color. The color map plots all share the same main plot widget menu that is significantly different than the menu available in trace plot widgets.



The main menu provides actions that may apply only to the plot widget that owns the menu.

FIGURE 3.18: COLOR MAP PLOT WIDGET MAIN MENU

- *Set Z axis scale to log10*: Recompute all the intensities by calculating the new intensity as the log10 of the original intensity.
  When the log10 scale is applied to the intensities of the previous color map plot, where almost no signal was visible, then the signal is made perfectly visible."



The switch to a log10 scale for the intensities might reveal a lot of missing signal when the intensities are used in a linear scale. The drawback is that noise can be hugely enhanced. See below for a way to reduce that noise.

FIGURE 3.19: COLOR MAP PLOT WIDGET MAIN MENU

- *Low pass-filter Z axis data*: The value that is entered in the input box that is shown is a *percentage* of the maximum intensity over the whole map. The function first computes a threshold using the formula threshold = percentage * max_int / 100. All the cells having an intensity value above that threshold have their value set to that threshold value. All other cells are unchanged. This function is used to reduce the dynamic range of the map and potentially enhance features of interest.

- *High pass-filter Z axis data*: The value that is entered in the input box that is shown is a *percentage* of the maximum intensity over the whole map. The function first computes a threshold using the formula threshold = percentage * max_int / 100. All the cells having an intensity value below that threshold have their value set to that threshold value. All other cells are unchanged. This function is typically used for removing noise.

3.8 General Operation of the Plot Widgets

All the plot widget-containing windows, like the *TIC/XIC chromatograms*, *Mass spectra* and *Drift spectra* windows, all contain plot widgets that have a general working scheme as to how the data can be visualized. The main visualization operations are succintly described below. The following convention will be used to describe the mouse buttons:

- 🖱: left mouse button;
- 🖱: middle mouse button;
- 🖱: right mouse button.

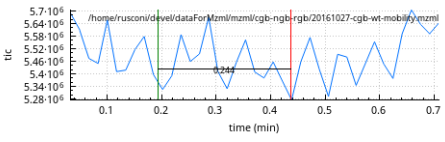


Note

All the visualization operations are performed using the *left* mouse mutton (🖱). Keyboard keys are used to better qualify the zoom-in/-out or panning operations.

The different plot or colormap graph visualization methods are detailed below:

- Zooming-in and zooming-out:
  - Zoom-in: 🖱-click-drag to draw a selection rectangle. When the mouse button is released, the new plot view contains the data contained in the selection rectangle. Note that if the mouse drag over the y-axis spans less than 10 % of the vertical scale, the rectangle does not show up because the software thinks that the mouse drag operation is only horizontal (which has a meaning described later).
  - Zoom-in: `Ctrl` +🖱-click-drag along the x-axis over the region to zoom and release the mouse button. The new zoomed view contains the mouse drag-spanned region. The y-axis spans the region contained between the initial plot bottom y-value and the position on the y-axis of the mouse during the drag operation.



The start and end x locations are visualized with green and red markers, respectively. The x-axis distance between the markers is refreshed all along the mouse movement.

FIGURE 3.20: ZOOMING-IN OPERATION

- Zoom-in/-out: `Ctrl` +🖱-click-drag *onto* either the x- or the y-axis to interactively zoom-in or -out along that selected axis. In this mode, the zoom operates by contracting/expanding the data in such a manner that the left/bottom part of the graph (the origin of the graph) is anchored and does not move. When the drag occurs towards larger values on the clicked axis, the view is zoomed-in along that axis. Conversely, it is possible to zoom-out by dragging the mouse towards lower axis values.

- Zoom-in/-out: The 🖱-wheel-rotation can be used to zoom-in or -out the whole plot on both the x- and y-axis simultaneously. Note that the position of the mouse cursor when the wheel is rolled defines the new view of the plot. Practising a bit allows to make that zooming-in/-out mode very powerful.

- Zoom-out: To reset the zoom along one axis, 🖱-double-click that axis. In this case, only the clicked axis will be full-scale, the other axis remains unchanged. To reset the zoom on both axes in one go, 🖱-double-click one of the axes maintaining the `Ctrl` key pressed;

- Panning:
  🖱-click-drag on one of the axes to pan the plot view along that axis;

- Relative full-scaling along the y-axis: in almost all of the zooming operations described above, it is possible to ensure that the y-axis is full-scaled automatically to the most intense peak visible in the zoomed region by pressing the `Shift` key.

- Recording the history of zoomed views:
  Each time a new zoomed-in/-out view of the plot is triggered, a history element is stored in the plot widget. To back-replay the various steps of the zoom-in/-out operations in sequence, from pre-last to first, hit the `Backspace` key. The exceptions to this mechanics is when the plot view is zoomed using the mouse wheel.



Note: Locking the x and y axes of all the plots

The tool bar located at the top of all the plot widget-containing windows described above contains two buttons that allow to lock the x-axis (the button icon has the horizontal red line) and the y-axis (red line is vertical) range throughout all the plot widgets in the window. This is of great use when the user wants to compare a number of graphs that have been obtained on comparable samples. The movements and zoom-in or zoom-out operations in one graph are then synchronized to all the other graphs.

4 Mass Data Integrations Featured by MineXpert

Analyzing mass spectrometric data (with or without drift data) usually involves performing various data integrations in sequence. We saw earlier that upon loading a mass spectrometry data file, the first data visualization that becomes available is the TIC chromatogram. Then, the user might ask to show the MS run data set table view (see Figure 3.4, "The table view-based scrutiny of one MS run data set"). Thus, the TIC chromatogram and the table view are both starting points for the mass spectrometric data exploration. Once a number of mass data integrations have been performed, new integrations will be available in *any* direction, by selecting the integration type in the plot widget (see Section 3.7.2, "The Right Column of Buttons Configures the New Integration to Run").

Regarding the parallel execution of mass spectral integrations, please see Section 1.5.3, "Using multiple threads during mass data integrations" .

4.1 General Behaviour of Plot Widgets

There are patterns, in the way mass spectral data integrations can be configured, that are common to all plot widgets. These common patterns are described in the next sections. First, however, the *Processing Flow* concept needs to be described, as it serves as the basis for the integration configurations.

4.1.1 Processing Flow Entities Document all Integrations

Each time an integration is performed, MineXpert stores processing information data that permits to characterize the integration for later internal reuse. Any integration, as it is performed, documents itself to the *Processing Flow*. For example, one integration might be from TIC chromatogram to mass spectrum, another integration can be from mass spectrum to drift spectrum. These kinds of integration are documented as *Processing Types* . For each integration, there is thus a source and a destination. But this is not sufficient to document precisely the integration: there must be other informational data along with the integration type. When integrating from TIC chromatogram to mass spectrum, the user might select only a region of interest in the chromatogram. That is another kind of information that is documented: the source range (in this case a retention time range). Likewise, when integrating from a mass spectrum to a drift spectrum, the origin is a m/z range and the destination is a drift spectrum. Other data are stored in the processing information data, like the fragmentation specification—if any has been set— and the m/z integration parameters, in case the integration produces a mass spectrum.

Together, all the processing information data are stored in what are called *Processing Flow* instances. A processing flow instance can contain a default MS^n fragmentation specification and/or default m/z integration parameters. It can contain any number of *Processing Step* instances that in turn can contain any number of *Processing Spec* instances. Like for the processing flow, each processing step can contain a MS^n fragmentation specification and/or m/z integration parameters.

How is the processing flow documented ? When a mass spectrometry data file is loaded from disk, MineXpert iterates in all the mass spectra of the MS run data set and computes a TIC chromatogram. That TIC chromatogram is displayed in the *TIC/XIC chromatograms* window. At this point, no MS^n fragmentation specification is set and no m/z integration parameters are set. When the user starts exploring the data, their first integration will determine how the processing flow will be filled-in with details about that integration.

If an integration is performed from the TIC chromatogram to a mass spectrum, then m/z integration parameters will necessarily be used. Either the user has configured these parameters (Section 4.1.3, " Setting the m/z Integration Parameters ") and they will be used to perform the mass spectral combinations, or default values will be used (these default values are crafted from the statistical analysis of the whole MS run data set). The processing step that is created to document this specific integration will thus document the m/z integration parameters that have been used for the mass spectra combination. That processing step is added to the processing flow that documents the whole integration process. The plot that is created in the *Mass spectra* window will have, associated to it, that processing flow. Think of the *Processing Flow* as a pedigree ID card that each plot has associated to it. That pedigree gets incremented with new processing steps each time an integration is chained to the previous integrations.

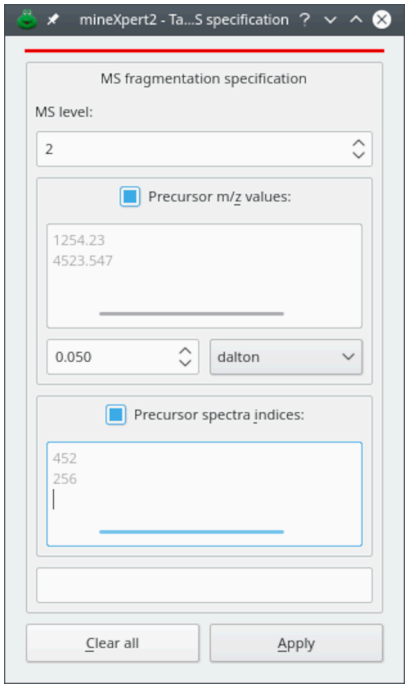


Tip

The original feature of the *Processing Flow* concept is that, because each plot has its own processing flow, when a new integration is performed starting from that plot, then all the preceding steps of the processing flow are replayed and the new step is added to the processing flow. This has the beneficial effect of virtually restricting the scope of the initial data set into a much smaller data set all along the various integrations that occur during a data exploration session. When the scope reduces, the computing times reduce accordingly. However, the initial data set is not reduced in memory, only the scope of it is reduced. This is much different from other software in which mass data are effectively pruned off the initial data set with the drawback that data must be reloaded each time an integration must start from the initial data set.

4.1.2 Setting the MS^N Fragmentation Parameters

At any moment is it possible to set the MS^n integration parameters by selecting the *Open MS fragmentation dialog* menu item from the plot widget main menu (Section 3.7.3, "The Plot Widget Main Menu").



The MS^n fragmentation parameters might be set at any time from the plot widget's main menu. The criteria are of three kinds, as described in detail in the text.

FIGURE 4.1: SETTING THE MS^N FRAGMENTATION PARAMETERS

The dialog window that is displayed is depicted in FIGURE 4.1, " SETTING THE MS^N FRAGMENTATION PARAMETERS " . The available settings are explained below:

- *MS level*: MS level that is targeted by the integration. For example, from a TIC chromatogram, setting *MS level* to 2 would perform an integration only accounting for mass spectra acquired for MS fragmentations of MS level 2, that is, for MS/MS spectra.

- *Precursor m/z values*: SPACE or RETURN separated list of m/z values. The precursor ions' m/z values are taken into account for the integration. When iterating in the mass spectra acquired in the MS run data set, each spectrum will be checked for the existence of an ion selection list. If any of the m/z values entered in this dialog window are found, the spectrum is accounted for; otherwise it is dismissed. A precursor m/z value is the m/z value of an ion that was selected for fragmentation in a precursor spectrum acquired right before the acquisition of the fragmentation spectrum.



Warning

There is a great probability that entering m/z values in the text widget above will fail to provide any result without some value tolerance set. Indeed, the matches are performed strictly using the entered values and the values recorded by the mass spectrometer, which might have 10 decimals! It is thus necessary to enter a tolerance value for the m/z value matches. In the figure, the tolerance is set to *0.05 Dalton*. Other "units" are available: resolution power and ppm.

- *Precursor spectra indices*: SPACE or RETURN separated list of integer values. The precursor spectra indices are taken into account for the integration. When iterating in the mass spectra acquired in the MS run data set, each spectrum will be checked for the existence of precursor spectra indices. If the precursor spectra indices entered in this dialog window are found, the spectrum is accounted for; otherwise it is dismissed. A precursor spectrum index is the number of the spectrum in the MS run data set that contained an ion that was selected for further fragmentation.



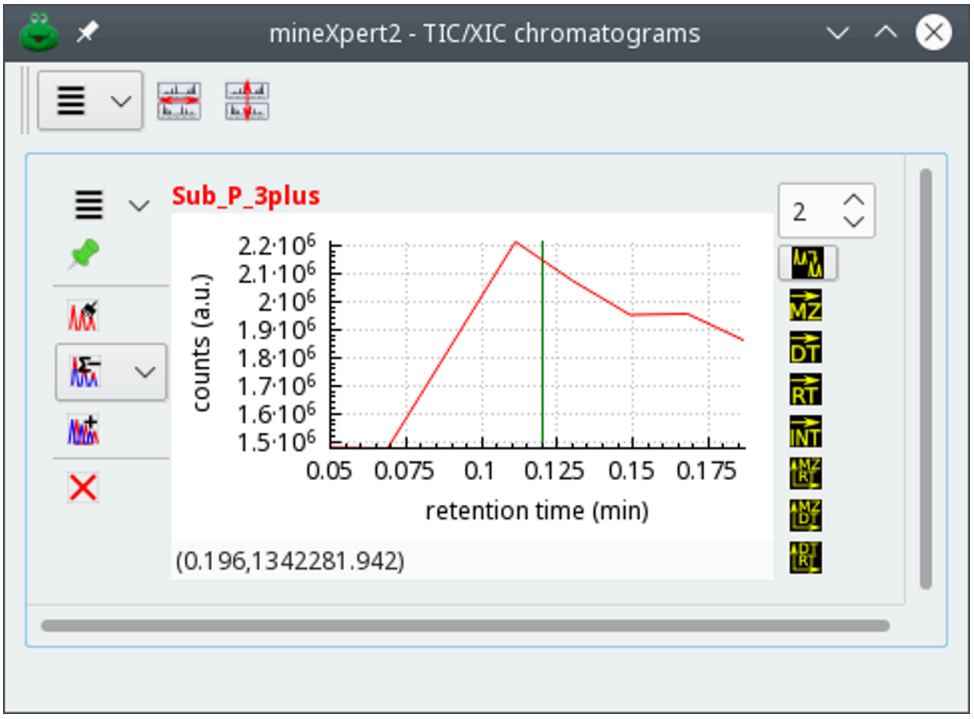
Note

Note that in the dialog window above, none of the three criteria is essential. Setting the *MS level* to 0 disactivates that parameter as a valid criterion for filtering spectra during the integration. In this case, only MS^n fragmentation specifications used in preceding integrations are taken into account (that is, fragmentation specifications present in the *Processing Flow*). If one wants to reset the MS^n fragmentation specification to a lower MS level than that of the last preceding integration, the only way to achieve this is to go back to the previous plot where the processing flow contained processing steps of the right MS^n fragmentation level values. Indeed, it is not possible to reset the MS^n fragmentation level in a plot by setting the *MS level* to 0 because that plot already has a processing flow containing pocessing steps configured with a non modifiable MS level.



Tip

There is a shortcut to the MS^n fragmentation MS level in each composite plot widget, as shown below, where the spin box widget on top of the right hand side column of allows one to set the MS level.



If the only fragmentation setting to be modified is the MS level, then, instead of opening the dialog window described above, it is possible to set the required MS level right into the spin box widget at the top of the right hand side column of buttons in the composite plot widget.

FIGURE 4.2: **SHORTCUT TO THE FRAGMENTATION MS LEVEL SETTING**

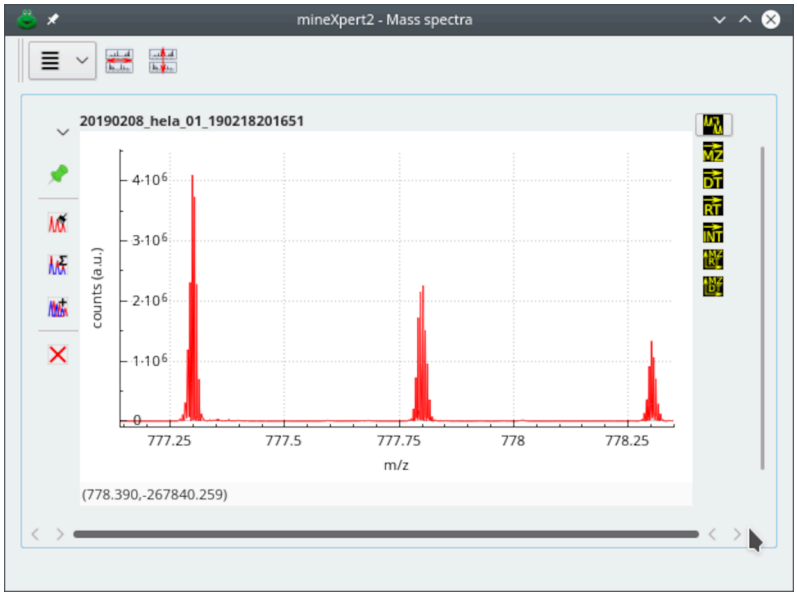
4.1.3 Setting the m/z Integration Parameters

The m/z integration parameters are only compulsorily used in case an integration will produce a mass spectrum or a color map involving m/z data in any one of the two dimensions. Indeed, these parameters are required to configure the way mass spectra are combined together into a single product mass spectrum.

An integration to a mass spectrum occurs when the user -selects a range in a given plot while the MZ checkbutton is checked. Integrations to a mass spectrum can be elicited from any plot.

Very rarely can one perform an integration to a mass spectrum without having to first configure the way the m/z values are binned into the combination spectrum. The determination of the best m/z integration parameters is a process that might be required each time a new data set comes from an instrument from which to data have been obtained before. Indeed, there is a vast number of different mass spectral data acquisition/storage modes that

depend on the vendors or on the mass spectrometer models. In general, this diversity of mass spectral data files [1] creates difficulties in the mass spectra integrations, as shown below in Figure 4.3, “ Unusable combination spectrum without binning ” .

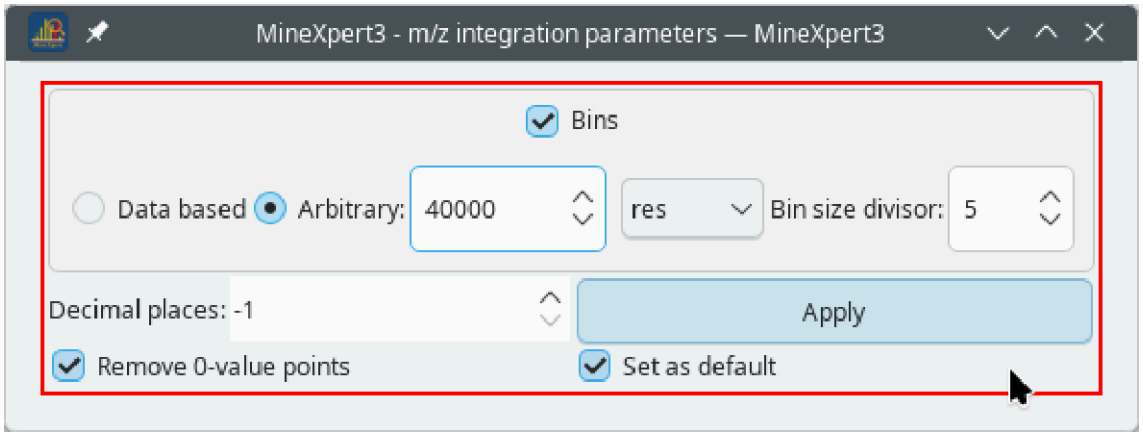


The mass data used to compute this combination spectrum originate from a Lumos Orbitrap analyzer. The visible signal should have been three peaks belonging to a 2-charged ion isotopic cluster. Each peak of the isotopic cluster is artifactually made of numerous data points because the data from the Orbitrap analyzer have 10 decimal digits. Without binning, the mass spectrum is almost unusable.

FIGURE 4.3: **UNUSABLE COMBINATION SPECTRUM WITHOUT BINNING**

In this combination mass spectrum, computed without binning from Lumos Orbitrap-originating data, three peaks that belong to an isotopic cluster appear as made of a number of “sub-peaks”. This is due to the fact that the number of decimals for the m/z values in the data file is so high that a single isotopic peak appears as a set of peaks. The signal in this mass spectrum is totally useless. This is a perfect illustration of the necessity of data binning of the m/z values of the mass spectra to be combined.

1 Even if they are obtained by conversion of raw vendor files to mzML files using ProteoWizard 's msconvert tool.



Integrations to any plot that involves m/z as one of its dimension (be that a mass spectrum or a color map) can be configured to ensure the best results, depending on the kind of mass data. Proper binning configuration is key to getting best results.

FIGURE 4.4: THE M/Z INTEGRATION PARAMETERS WINDOW

MineXpert provides a number of ways to configure mass spectral combinations such that the obtained mass spectrum is usable. The m/z integration parameters that might be set are described in the following sections. The window depicted in Figure 4.4, “ The m/z integration parameters window ” can be displayed from any plot widget menu using the *Open the m/z integration params dialog* menu item (see Section 3.7.3, “The Plot Widget Main Menu”). The following parameters can be set:

- *No binning* : no binning is carried-over during the mass spectral integration. In this case, all the m/z values encountered in all the mass spectra combined into the final combination spectrum are going to be used for the combination and will be encountered in the result spectrum. This is illustrated in Figure 4.3, “ Unusable combination spectrum without binning ” .



Tip

There are instruments that produce perfectly binned mass spectra. In this case, using *No binning* is certainly the best option. For example, the files acquired from the Synapt 2 HDMS mass spectrometer sold by Waters are perfectly binned.

- *Data-based binning*: binning is performed using bins already set in the mass spectrum into which the combination is to be performed. This kind of situation only arises when the user has pushed down the pin in one of the composite plot widgets and also has specified that the combination must be performed in the pinned plot widget.

- *Arbitrary binning*: binning is requested and the size of the bins might be defined using either *dalton*, *res*olving power or *ppm* “units”.

In the case of *res*olution binning, the bins are computed by compounding the m/z value to the inverse of the resolving power. In the case of *ppm* (part-per-million) binning, the bins are computed by compounding the m/z value to the inverse of one million.

- *Decimal places*: when performing the combination, each m/z value needs to be restricted to that number of decimal places. This setting may in some situations improve the signal or accelerate the speed of the combination for very large datasets.



Note

If *Decimal places* is let to the default value -1, that datum will be determined experimentally by looking into the bins size. Typically, if the bins size is, for example, 0.05, then, the decimal places will be determined to be 2 (number of 0 decimals after the decimal point, plus one). If the bins size is 0.003, the decimal places will be determined to be 3.

- *Remove 0-value data points*: remove all the (m/z,i) pairs having a naught *i* value.

When all the parameters have been set, click onto *Apply* or key in CTRL RETURN . At this point the next integrations started from the plot widget in which these m/z integration parameters were set will be performed using the settings. These parameters will propagate to all the descendant widgets. It will be possible, at any moment and in any widget, to modify the parameters again.



Tip

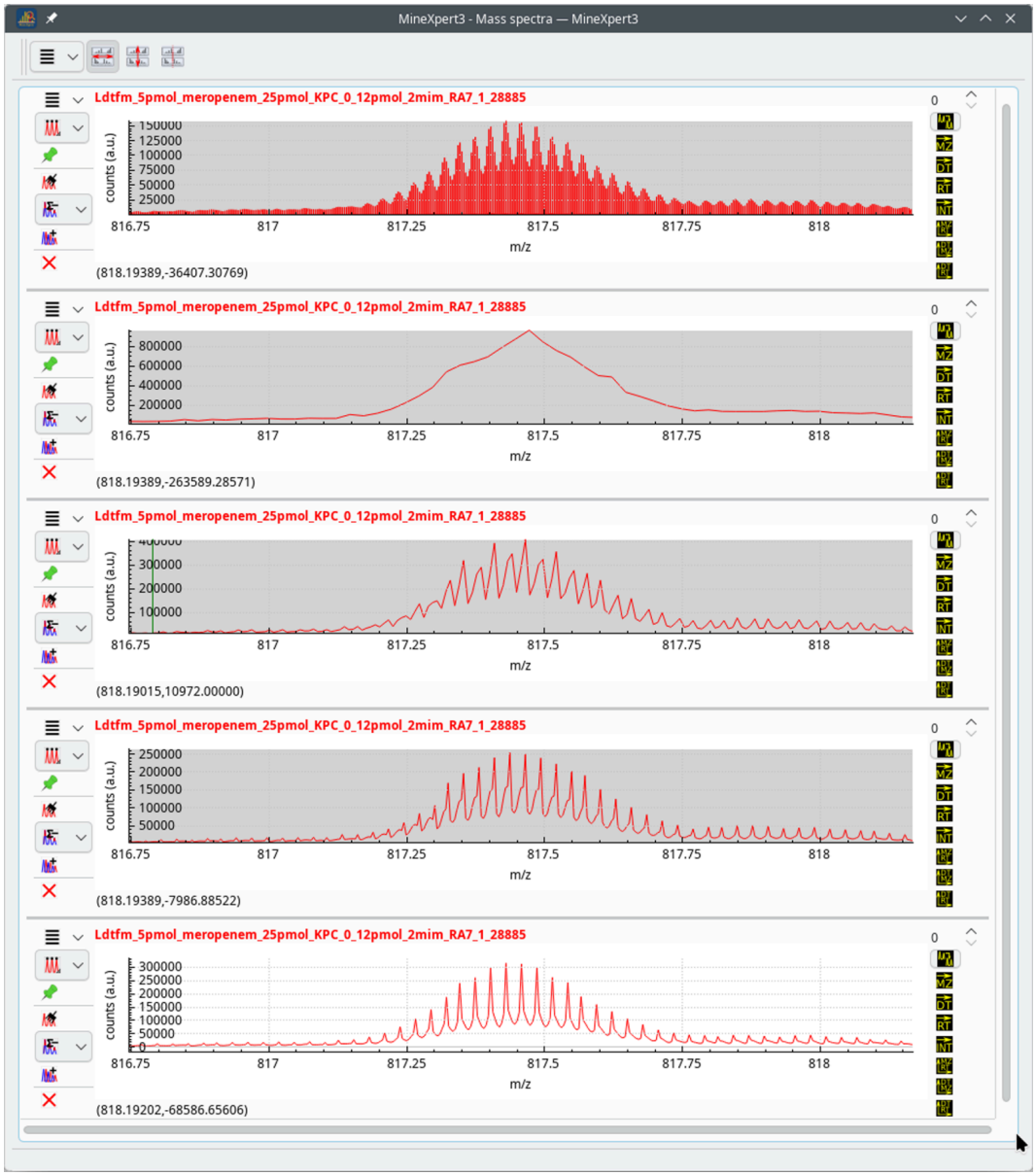
If the check box \guilabel{Set as default} is checked, the system records these parameters and makes them the default ones for any coming integration.

4.1.4 Step-by-step Determination of the Best m/z Integration Parameters

This section provides guidelines on the best way to determine the m/z integration parameters that will result in nicely shaped mass peaks in the combination mass spectrum. There might be situations where there is no need to set the m/z integration parameters, typically if the mass data have been properly binned by the software running the mass spectral data acquisition (Waters and Agilent systems often provide nicely binned mass spectra).

However, in a vast majority of the cases, proper setting of the m/z integration parameters will be essential to achieve a combination of mass spectra resulting in a useful mass spectrum. In particular, it is important to grasp that the binning might need to be dynamically set while combining (m/z,i) pairs along the m/z axis. This dynamic binning is configured by choosing either *res* or *ppm* (for resolving power or part-per-million, respectively; bins have varying sizes throughout of the whole m/z range).

When dealing with a mass spectrometry data set from an instrument that is unknown, the first step is to perform an integration to a mass spectrum without binning. In this situation, *all* the distinct peaks observed in all the mass spectra being combined together are going to be stored in the combination spectrum. When the data come from an instrument that does not produce nicely binned mass spectra, the resulting combination mass spectrum might be highly spiky, with spurious data points interspersed between seemingly correct data points. This is typically visible in the mass spectrum shown in Figure 4.5, “ Determination of the best m/z integration parameters for integration to mass spectra ”.

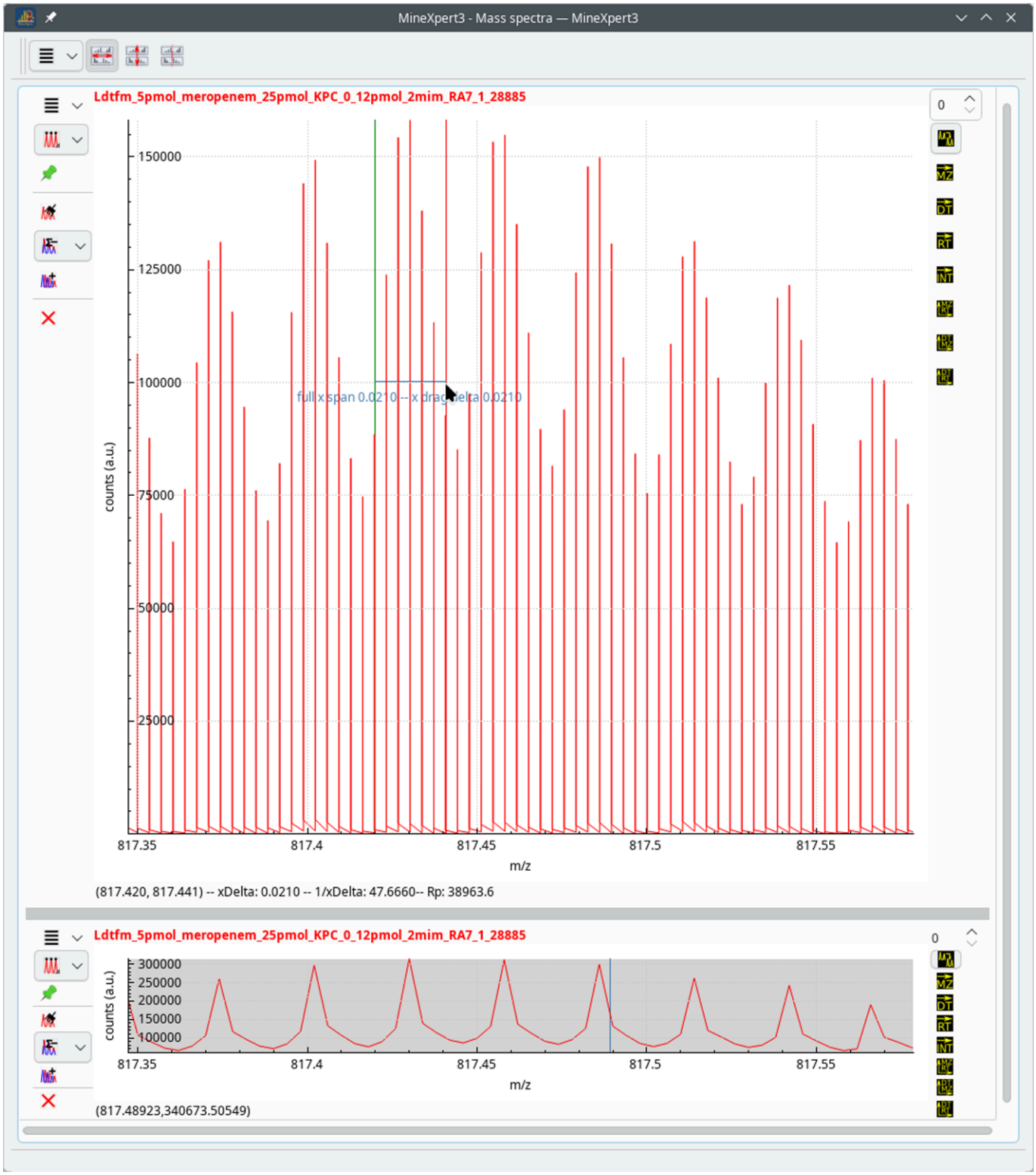


Zoomed in view of a protein mass spectrum.

FIGURE 4.5: **DETERMINATION OF THE BEST M/Z INTEGRATION PARAMETERS FOR INTEGRATION TO MASS SPECTRA**

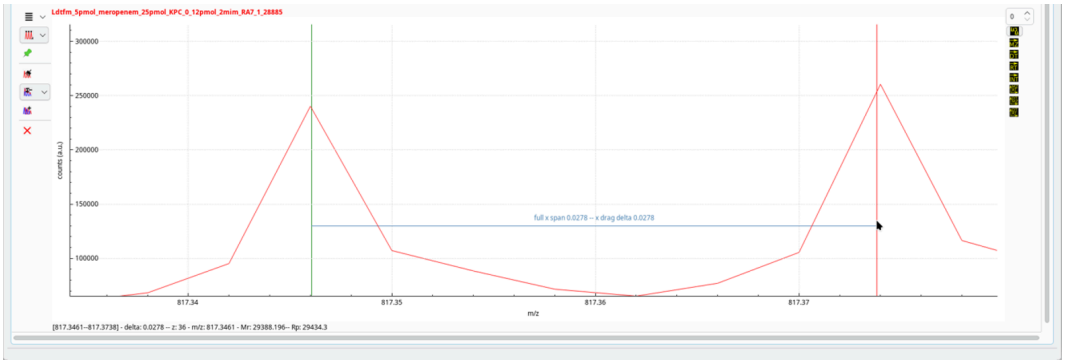
The different plots in the figure show the mass spectral data combined for the same acquisition data and for the same retention time range. From top to bottom:

- *No binning* . Visibly, each peak that is part of the isotopic cluster is made of a number of peaks that appear to be superfluous.
- *Bins res: 30000, bin size divisor: 1*. This resolving power was used as a starting value for exploring the data because the apparent resolving power was found to be around that value in Figure 4.6, " Estimating the resolving power of a mass spectral acquisition " . Remarkably, the bin size divisor value of 1 cannot permit to shape the isotopic cluster peaks properly.
- *Bins res: 30000, bin size divisor: 5*. Dividing the bin size determined above for resolving power of 30000 starts to give meaninful results: the peaks of the isotopologues start to be shaped. They are still a bit rough.
- *Bins res: 30000, bin size divisor: 6*. Using value 6 instead of 5 refines the isotopologue peak shapes. However, there is still an artifactual shoulder on the left of the isotopologue peaks.
- *Bins res: 40000, bin size divisor: 5*. That is the right combination: resolving power of 40000, with bin size divisor of 5. The charge determination by deconvolution of the isotopic cluster yields the right charge of the polymer: +36.



Dragging the mouse cursor between the extremities of an isotopic cluster peak, at half of its height, provides a rough estimation of the resolving power of the instrument.

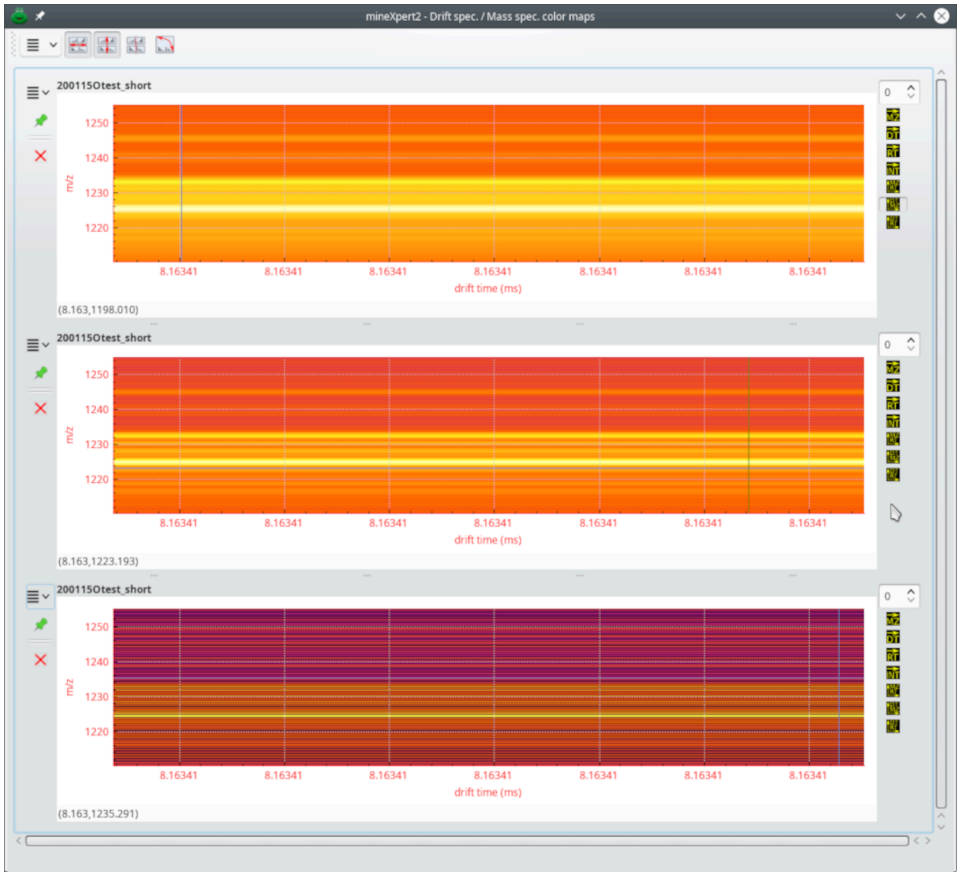
FIGURE 4.6: **ESTIMATING THE RESOLVING POWER OF A MASS SPECTRAL ACQUISITION**



One evidence that the setting for this integration is correct is that it is now possible to use the inter-peak deconvolution method for the determination of the ion charge (36+), which confirms the result of a previous deconvolution based on the charge state envelope.

FIGURE 4.7: **DETERMINATION OF THE ION CHARGE BY USING THE ISOTOPIC CLUSTER**

One important concept to grasp is that the bin size used for the mass spectra combinations will actually determine the final resolution of the mass spectrum (or color map). For example, imagine that the setting is the following: Arbitrary binning with bin size of 1~Da; in that case the mass spectrum or the m/z-based color map will be very rough. Instead, setting the bin size to 0.05 will provide a much better spectrum (or color map). Of course, going down with the size of the bins will increase almost linearly the computation time required to go through them all while adding signal from the various spectra that are being combined. This is exemplified in Figure 4.8, “ Effects of the bin size on the color map involving one m/z dimension ” .



The three different color map regions were computed in the *Arbitrary binning* mode with bin sizes 1~Da, 0.1~Da and 0.01~Da (top to bottom).

FIGURE 4.8: **EFFECTS OF THE BIN SIZE ON THE COLOR MAP INVOLVING ONE M/Z DIMENSION**



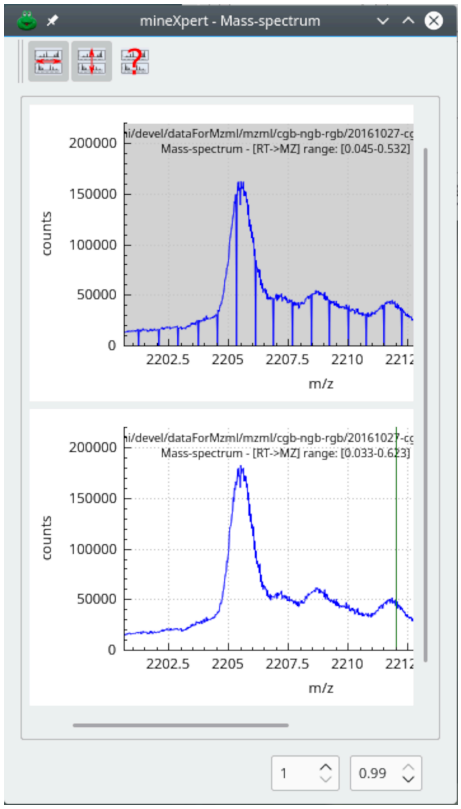
Tip

Because the computation time increases inversely with the size of the bins, it is recommended to start a color map-based data exploration session with an integration over the whole data set but with *large* bins (typically 1~Da) that will provide enough sight on the data. Then, when the feature of interest has been located on the color map to a given region, reduce the bin size to increase the resolution of the color map and integrate from that same region to a new color map.

4.1.5 Removing 0-intensity m/z data points is useful

Mass spectrum data points with an intensity value equal to 0 might arise either from the initial data, as read from the file, or from the binning process, as detailed below.

When bins are prepared, prior to starting a mass spectral combination operation, they all correspond to a m/z value associated to a o-value intensity. Once the bins have been prepared, the combination starts and, while iterating through all the data points of all the spectra, fills-in the bins. If, however, a bin is never filled (it ends up never updated during the combination), that bin will have an intensity of o. Bins having a o-intensity value have a bad effect on the result mass spectrum. Removing them by setting the *Remove o-value data points* parameter proves beneficial, as shown in FIGURE 4.9, " REMOVING O-INTENSITY DATA POINTS " . When the o-intensity data points are not removed (upper spectrum), the signal is deteriorated by spurious inverted spikes. Removal of the o-intensity data points, cleans the trace perfectly.

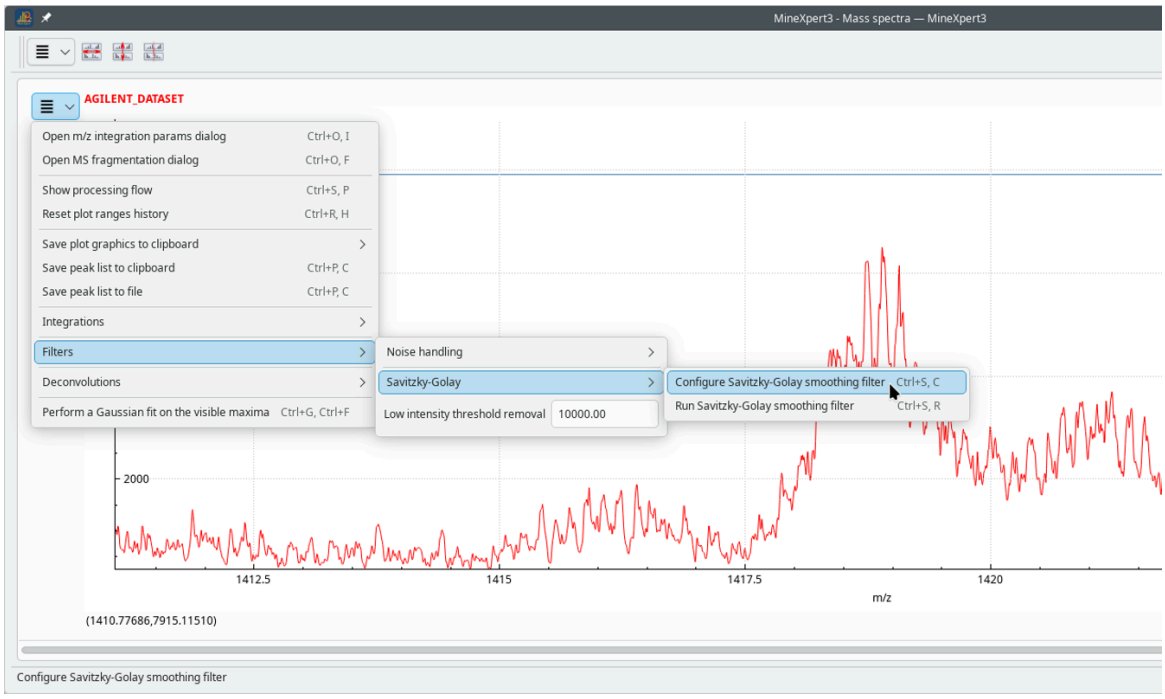


When arbitrary binning is performed, residual o-intensity data points might survive in the combination spectrum, which deteriorates the resulting mass spectrum. Removing these data points from the combined mass spectrum cleans the trace.

FIGURE 4.9: **REMOVING 0-INTENSITY DATA POINTS**

4.1.6 Savitzky-Golay filtering of any kind of data

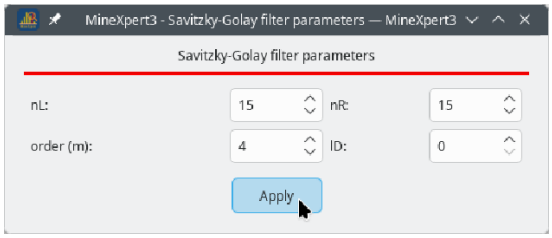
The Savitzky-Golay filtering method is widely known for its effectiveness in removing noise from mass spectral data. It is possible to apply that filter to any plot by selecting the corresponding menus as described at Figure 4.10, "Menus to configure and run the Savitzky-Golay filter to smooth traces". The Savitzky-Golay parameters can be set by selecting the *Configure Savitzky-Golay smoothing filter* menu as shown on Figure 4.10, "Menus to configure and run the Savitzky-Golay filter to smooth traces".



The Savitzky-Golay algorithm is a powerful filter that allows one to smooth any graph. The first menu enables the configuration of the filter and the second menu actually applies the filter to the trace.

FIGURE 4.10: MENUS TO CONFIGURE AND RUN THE SAVITZKY-GOLAY FILTER TO SMOOTH TRACES

The Savitzky-Golay filter requires careful configuration. The configuration provided by default is shown in Figure 4.11, "Configuration of the Savitzky-Golay filter to smooth traces".



The different parameters used to configure the Savitzky-Golay algorithm are described in the text.

FIGURE 4.11: CONFIGURATION OF THE SAVITZKY-GOLAY FILTER TO SMOOTH TRACES

- *nL*: specifies the number of data points to the left of the point being filtered;
- *nR*: specifies the number of data points to the right of the point being filtered. The total number of points in the window that is considered for the regression is thus nL + nR + 1.
- *m*: specifies the order of the polynomial to use in the regression analysis leading to the Savitzky-Golay coefficients (typically between 2 and 6);
- *lD*: specifies the order of the derivative to extract from the Savitzky-Golay smoothing algorithm (for regular smoothing, use value 0);

4.1.7 Centroid Extraction from Mass Spectra

The centroidation of a mass spectrum is also called peak picking. This process is aimed at converting a profile-mode spectrum, that has its peaks properly shaped in Gaussian-like form, into a set of m/z values that recapitulate the initial peaks.

One aspect that is very important in peak picking is the determination of what is considered to be noise and what is considered to be actual signal. The signal-to-noise estimation is an optional task that is only carried out when the *Signal to noise ratio* is set to a value greater than 0. The signal-to-noise estimator needs a number of parameters to be set up (in the *Configuration of the signal to noise estimator* group box widget). One important configuration bit is related to the determination of what intensity value is considered to be the maximum intensity value over the whole mass spectrum. There are multiple ways to define that maximum intensity value, as described in .

- *Manual setting*: this is the simplest method to define that maximum intensity value, by having the user setting it up.
- *Standard deviation factor*: the software first makes a statistical analysis of the intensity values throughout the mass spectrum and any data point that has an intensity over the standard deviation compounded by this factor is discarded.
- *Max percentile*: any data point that has an intensity over the intensity of the data point representing the top 5 percent of the intensity value space of the mass spectrum is discarded.

Other configuration bits are the *Rolling window size* which specifies the m/z range considered when computing the local signal-to-noise ratio. The rolling window is moved data point by data point over the whole mass spectrum. Each mass spectrum data point has thus a corresponding signal-to-noise ratio value. The *Mininum required elements* configuration bit is the number of data points that are required to be found inside the rolling window. The *Bin count* configuration bit is the number of bins in the histogram of the intensities in the mass spectrum. This parameter defines the granularity with which the signal-to-noise ratio is computed for each mass spectrum data point.

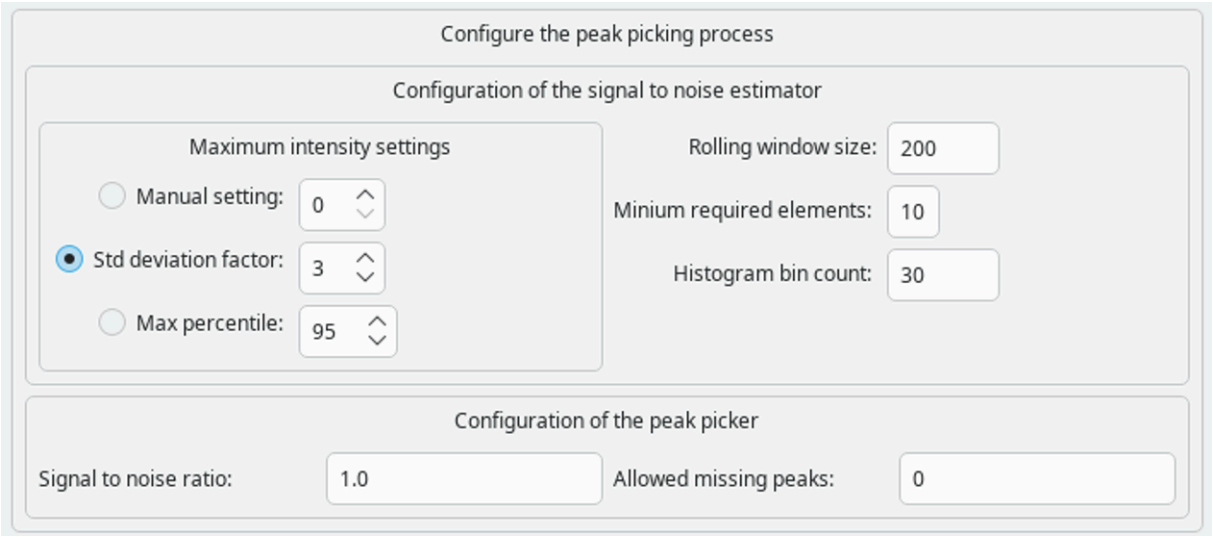


FIGURE 4.12: **CONFIGURATION OF THE CENTROIDATION OF MASS SPECTRA**

Finally, one last bit of configuration for the centroid extraction is the *Allowed missing peaks* value that defines the number of m/z values that can be found to be absent from the rolling window. This parameters provides some leeway when dealing with sparse mass spectra.

When the centroid extraction process is terminated, the centroids are plotted as a bar graph. These centroid data points can be exported to a file or a clipboard exactly the same way as for the normal trace graphs.

4.2 Various Mass Spectral Data Integrations

Now that the configuration patterns common to all plot widgets have been described, the various mass spectral data integrations can be reviewed.



Note

All the integration operations are performed using the *right* mouse mutton (). This setting allows to easily distinguish all these integration operations from all the -based visualization operations (Section 3.8, "General Operation of the Plot Widgets").

- *Integrations to a mass spectrum* This kind of operation is triggered upon -click-dragging the mouse over the region of interest after having selected the MZ button on the right button column of the widget. MineXpert integrates all the spectra that have been acquired between the start marker and the end marker as set during the mouse drag movement. A new mass spectrum is then plotted in the *Mass spectra* window.



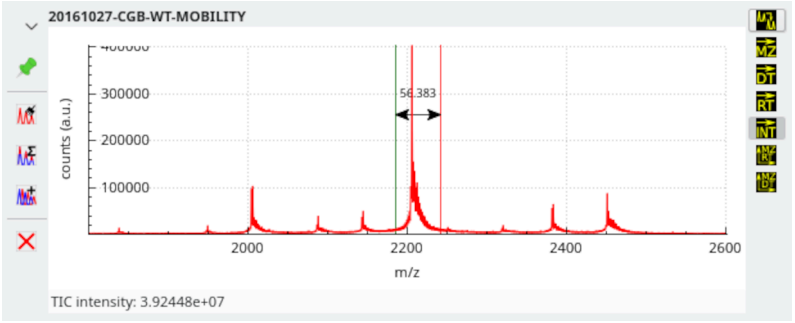
The integration is defined to be a *[TIC chromatogram to mass spectrum]* integration because the plot widget is in the *TIC/XIC chromatograms* window and the *MZ* push button is checked.

FIGURE 4.13: **INTEGRATING MASS DATA FROM A TIC CHROMATOGRAM TO A MASS SPECTRUM**

As seen on the figure above, the region defined by the ⌥-click-dragging operation is delimited by arrows, a green marker at the start and a red marker at the end. The arrows at the ends of the horizontal line allow one to make the difference between an integration selection and a simple distance-measuring selection (detailed later).

- *Integrations to a drift spectrum* This kind of operation is similar to the one described above, unless for the DT button that is now checked. The new plot is added to the *Drift spectra* window.

- *Integrations to TIC intensity* This operation involves ⌥-selecting a plot region of interest encompassing the mass spectral feature for which the intensity is to be integrated while the INT is checked. The obtained numerical result is displayed in the status bar below the plot widget and in the *Console window*.



The integration is defined to be a *Mass spectrum to TIC intensity* integration because the plot widget is in the *Mass spectra* window and the *INT* push button is checked. The TIC intensity value is printed in the status bar.

FIGURE 4.14: **INTEGRATING MASS DATA FROM A MASS SPECTRUM TO A SINGLE TIC INTENSITY VALUE**

- The same mechanics is at work in the other plot widget windows. For example, to trigger the integration of any kind of plot to an intensity = f(dt,m/z) color map, simply check the MZ DT button and drag the mouse over the plot region of interest.

4.2.1 Considerations on the Diversity of Mass Data Contents

Loading data from mass data files in mzML format does not guarantee that the data will be of the same kind when they originate from different mass spectrometers. For example, data from Orbitrap mass spectrometers have the following characteristics:

- All spectra do not start at the same m/z value;
- All spectra do not have the same number of data points (they do not have the same size);
- A large number of data points might have 0 values (intensity at a given m/z value is 0);
- The m/z delta between two consecutive m/z values is not constant, and this is the major difficulty for data integration to a mass spectrum.

This is the output of the statistical analysis of the data loaded from a Lumos Orbitrap-originating file:

```
Spectral data set statistics:
Total number of spectra: 6203
Average of spectrum size: 391.311946
StdDev of spectrum size: 168.062934
Mininum m/z value: 400.007111
Average of first m/z value: 401.448935
StdDev of first m/z value: 1.590049
Maximum m/z value: 1999.928589
Average of last m/z value: 1901.852315
StdDev of last m/z value: 45.864131
Minimum m/z shift: -0.344452
Maximum m/z shift: 0.000000
Average of m/z shift: 1.097372
StdDev of m/z shift: 1.590049
Smallest Delta of m/z (step): 0.006195
Average of smallest Delta of m/z (step): 0.023757
StdDev of smallest Delta of m/z (step): 0.013179
Greatest Delta of m/z (step): 405.356934
Average of greatest Delta of m/z (step): 163.112057
StdDev of greatest Delta of m/z (step): 75.947334
```

As mentioned earlier, the most interesting bit of information is in the line reproduced below:

```
Smallest Delta of m/z (step): 0.006195
```

That 0.0062 value somehow gives an indication of the "definition" of the spectrum, that is, the smallest distance possible between two consecutive points in the m/z-axis.

In general, the fact that the spectra of an acquisition do not all have the same m/z vector as the m/z-axis is a great difficulty for mass spectral integration because it requires setting up binning prior to performing the mass spectral combination. That binning is nothing else than crafting a m/z value vector able to receive the intensities of all the m/z data points in the spectra to be combined. These concepts are developed in the next paragraph.

4.2.2 Statistical Analysis of Mass Data

At the end of the data file loading, MineXpert performs a rudimentary statistical analysis of the data. The main datum of interest is the smallest m/z step that is observed in the whole set of mass data loaded from disk (the mass spectrum list, that can hold mass spectra in the thousands). For each mass spectrum in the list, the smallest m/z delta between any two consecutive data points is recorded. Then, the smallest ever m/z delta value is sought amidst all the recorded values. Intuitively, that smallest m/z delta value provides an idea of the resolution power of the instrument that generated the mass spectra. Interestingly, this is not the proper value to configure binning. The best value is the median value of the smallest m/z delta values encountered over all the mass spectra of the data file. It is the value that is suggested by default to arbitrarily construct the bins during an integration to a mass spectrum, as described in Figure 4.4, " The m/z integration parameters window " (*Arbitrary binning* value with bin size unit *dalton*).

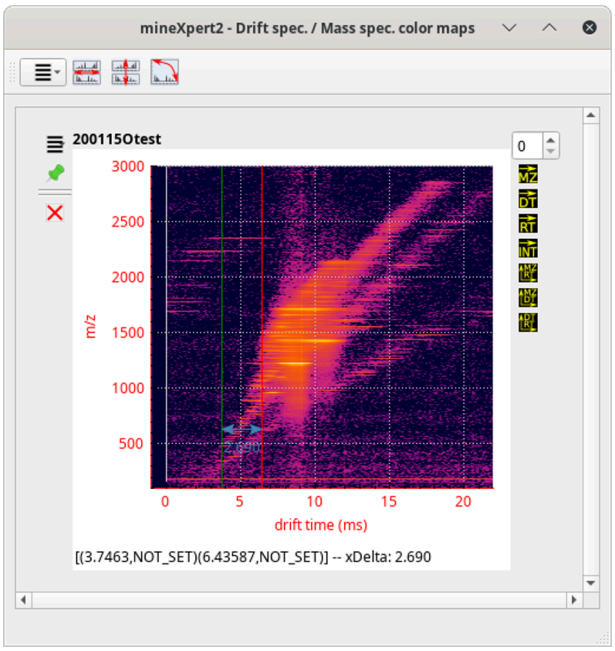
4.3 Integrations Originating in Color Map Plots

There are two different kinds of plots: the "trace plots" and the "color map plots". TIC chromatograms, mass spectra and drift spectra are trace plots. Color map plots plot intensity values as colors for pairs of data, like mass spectra as a function of retention times or drift times, or retention times as a function of drift times.

When an integration starts from a trace plot, the integration region selection (with a right-button mouse drag operation) of a given plot range (over the X axis) can be performed without drawing a rectangle over the region of interest. Indeed, only the X axis is the criterion for the selection of the region of interest. When an integration starts from a color map plot, the second dimension (the Y axis) has a meaning for the selection of the region of interest. In color map plots, the selection region has to be bi-dimensional, that is, not a horizontal line, but a real rectangle over both the X and the Y axes.

By default, the rectangle that is drawn upon mouse dragging is a real rectangle, that is, a four-square-corner polygon. There are a number of situations, however, where the user might want to select a non-rectangle region, like if the rectangle were skewed. This is illustrated in Figure 4.16, “ Integrations from a color map plot using a skewed selection rectangle (last step) ”

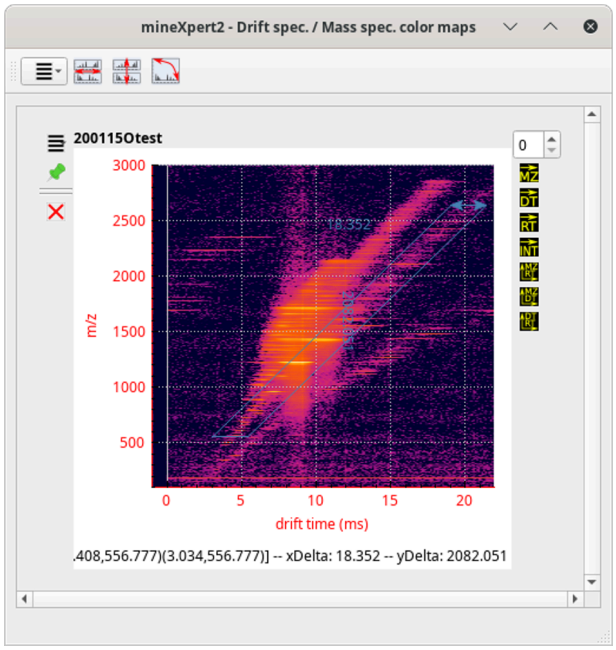
Drawing a skewed selection rectangle is a two-step process. One first draws the base of the rectangle, as usual, with a horizontal-only mouse drag movement (see Figure 4.15, “ Integrations from a color map plot using a skewed selection rectangle (first step) ”). When the base of the rectangle has the right width, the user presses the S keyboard key. Pressing that keyboard key triggers the definition of the width of the skewed selection rectangle, that gets stored in memory.



The first step in the definition of a skewed selection rectangle is to right-button-drag the mouse along the X axis in order to draw a selection line that has the final skewed rectangle width. Once the width is properly defined, the user presses the S keyboard key, which stores the rectangle width in memory.

FIGURE 4.15: **INTEGRATIONS FROM A COLOR MAP PLOT USING A SKEWED SELECTION RECTANGLE (FIRST STEP)**

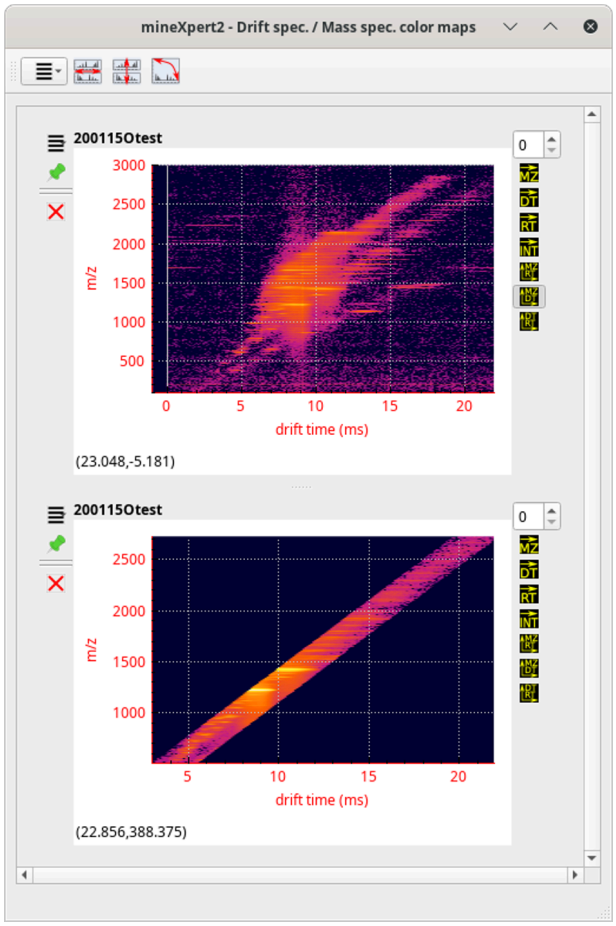
When the mouse drag operation draws the actual two-dimensional polygon (by dragging the mouse cursor also along the Y axis), the user maintains the Alt keyboard key pressed. Maintaing that key pressed has the effect of switching the drawing of a conventional four-square-corner selection rectangle into a skewed selection rectangle (see Figure 4.16, “ Integrations from a color map plot using a skewed selection rectangle (last step) ”).



The last step in the definition of a skewed selection rectangle is to drag the right mouse button over a bi-dimensional space (that is, also along the Y axis) while maintaining the `Alt` modifier key pressed. The rectangle that is drawn is now skewed, in contrast to a normal square-corner rectangle.

FIGURE 4.16: INTEGRATIONS FROM A COLOR MAP PLOT USING A SKEWED SELECTION RECTANGLE (LAST STEP)

When the mouse right-button is released, the integration computation is started that involves checking if the points that are processed are indeed contained inside the skewed selection rectangle. The result is shown in Figure 4.17, “ Integrations from a color map plot using a skewed selection rectangle (first step) ” .



The result of a skewed selection rectangle-based integration is inherently of the shape of the selection.

FIGURE 4.17: **INTEGRATIONS FROM A COLOR MAP PLOT USING A SKEWED SELECTION RECTANGLE (FIRST STEP)**

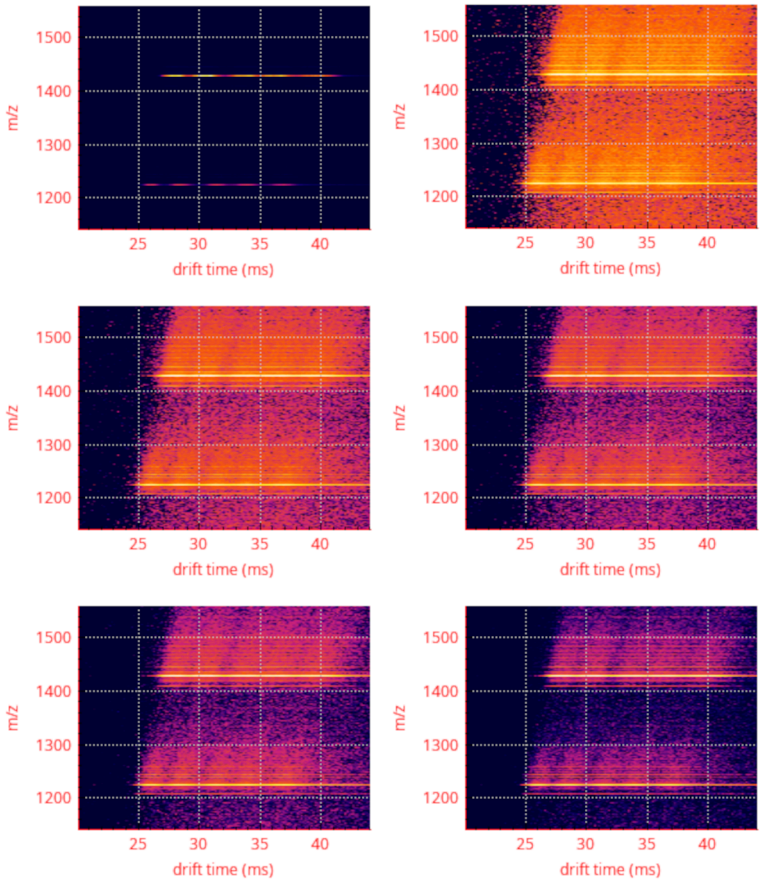
4.4 Noise Reduction in Color Map Plots

The signal obtained in ion mobility–mass spectrometry experiments may be of a very wide dynamics, which hamper proper visualization of details. One solution is to run a Z axis scale (the axis holding intensity values) conversion from linear scale to log10 scale (see Figure 3.19, "Color map plot widget main menu"). While this has the benefit of highlighting details, it also has the drawback of potently enhancing noise. As of version 8.1.0, MineXpert provides means to signal-to-noise ratio tweaking as detailed below.

There are two different approaches to signal-to-noise ratio enhancement:

- *Low pass-filter Z axis data*: The value that is entered in the spin box widget is a *percentage* of the maximum intensity over the whole map. The function first computes a threshold using the formula threshold = percentage * max_int / 100. All the cells having an intensity value above that threshold have their value set to that threshold value. All other cells are unchanged. This function is used to reduce the dynamic range of the map and potentially enhance features of interest.

- *High pass-filter Z axis data*: The value that is entered in the spin box widget is a *percentage* of the maximum intensity over the whole map. The function first computes a threshold using the formula threshold = percentage * max_int / 100. All the cells having an intensity value below that threshold have their value set to that threshold value. All other cells are unchanged.
  This function is typically used for removing noise. Indeed, by setting the percentage value to 5, for example, all the cells having a pretty low intensity value (that is, below 5 % of the maximum intensity value in the whole map data) will be set to that value. An example is provided in Figure 4.18, " Example of signal-to-noise ratio enhancement ".



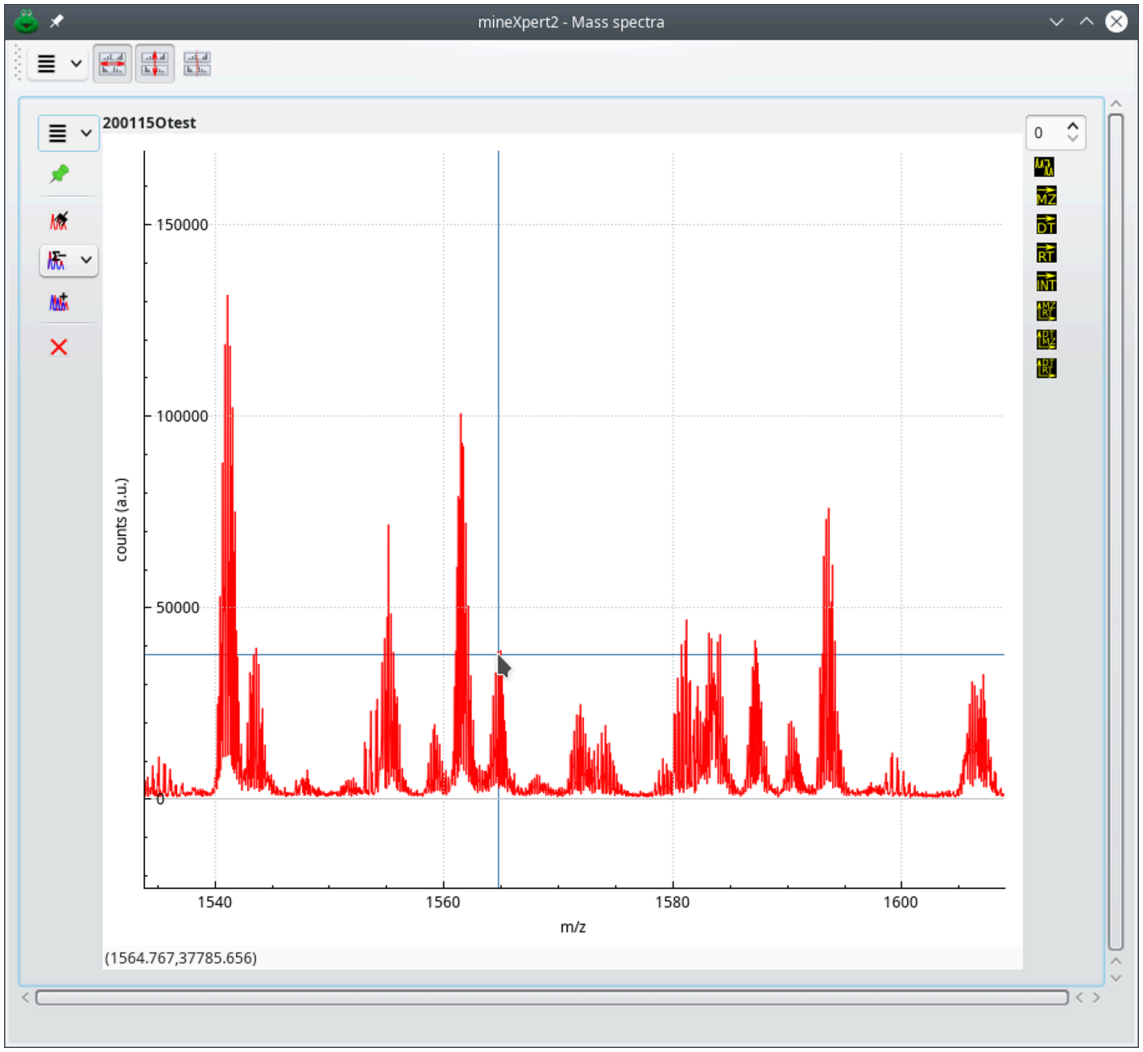
The top left vignette shows the colormap as it is computed. The top right vignette show that colormap after a log10 recomputation has been performed. The remaining vignettes correspond to the application of a high pass filter with percentage values increasing from 10 to 40 in 10 % increments.

FIGURE 4.18: **EXAMPLE OF SIGNAL-TO-NOISE RATIO ENHANCEMENT**

4.5 Low Intensity Signal Removal in Traces

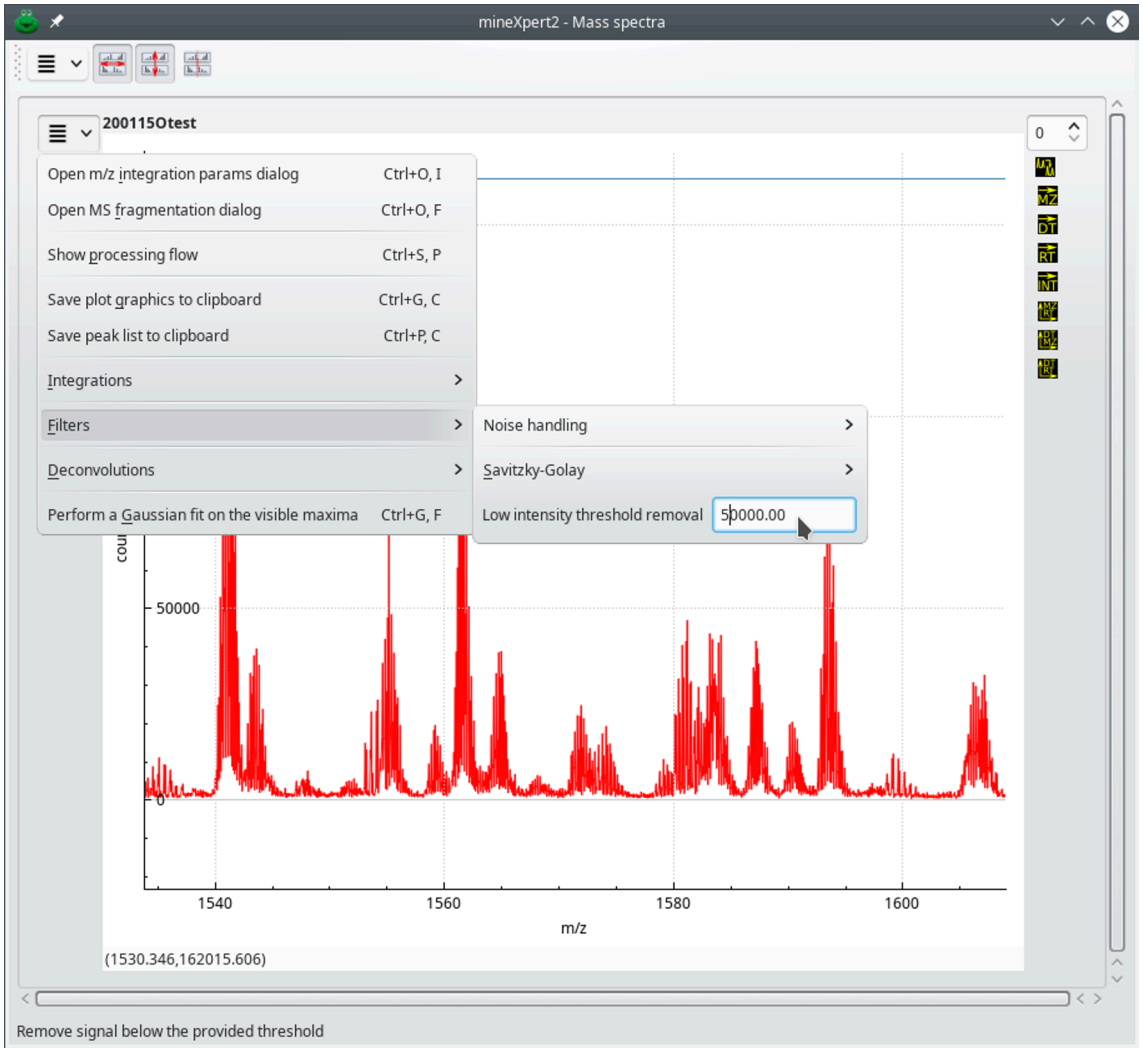
Any trace (that is, any plot that is not a color map) can be noisy. That is particularly true for mass spectra coming from non-Orbitrap analyzers. If one can tell what is the *level* of the noise, it is possible to remove all the peaks that are below that level. The process is complex because it needs to take into account ever-more frequent high resolution mass spectrometry data that shows resolved isotopic cluster for which certains peaks might be below the threshold while others are above threshold. Of course, if any peak of an isotopic cluster is above threshold, then the whole isotopic cluster, no matter how it might comprise below-threshold peaks, should be retained.

At the time of this writing, the low intensity signal removal filter is working best for samples containing large analytes, where the isotopic cluster is most approaching an overall Gaussian shape. It is considered that this feature is still experimental as it still has a number of rough edges. The step are described in the following figures (see Figure 4.20, “ Example of low intensity signal removal (configuration) ”).



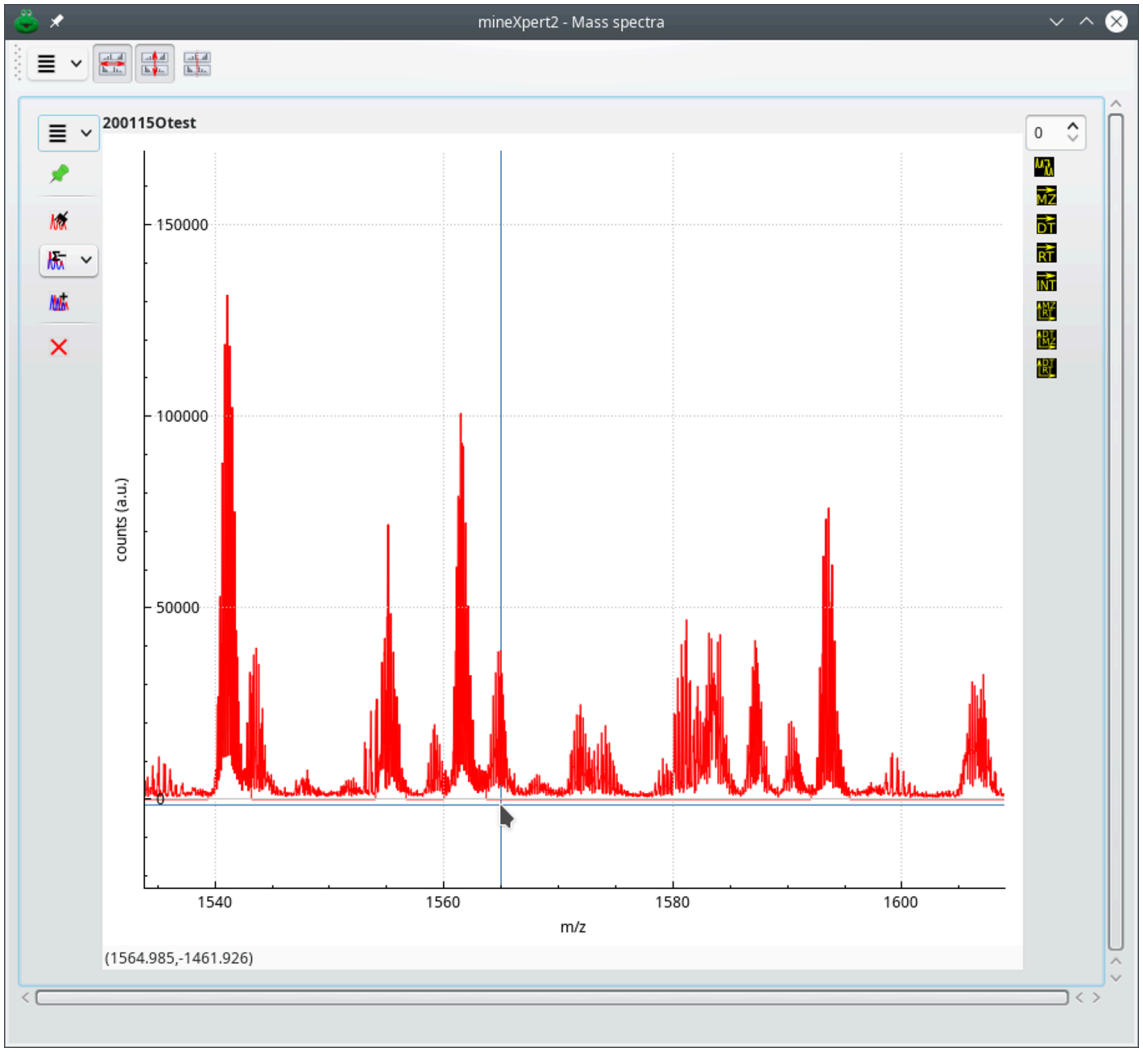
This is a segment of the initial mass spectrum with some isotopic clusters having part of their peaks above the 50000 threshold and some others that are entirely below that threshold.

FIGURE 4.19: **EXAMPLE OF LOW INTENSITY SIGNAL REMOVAL (STARTING TRACE)**



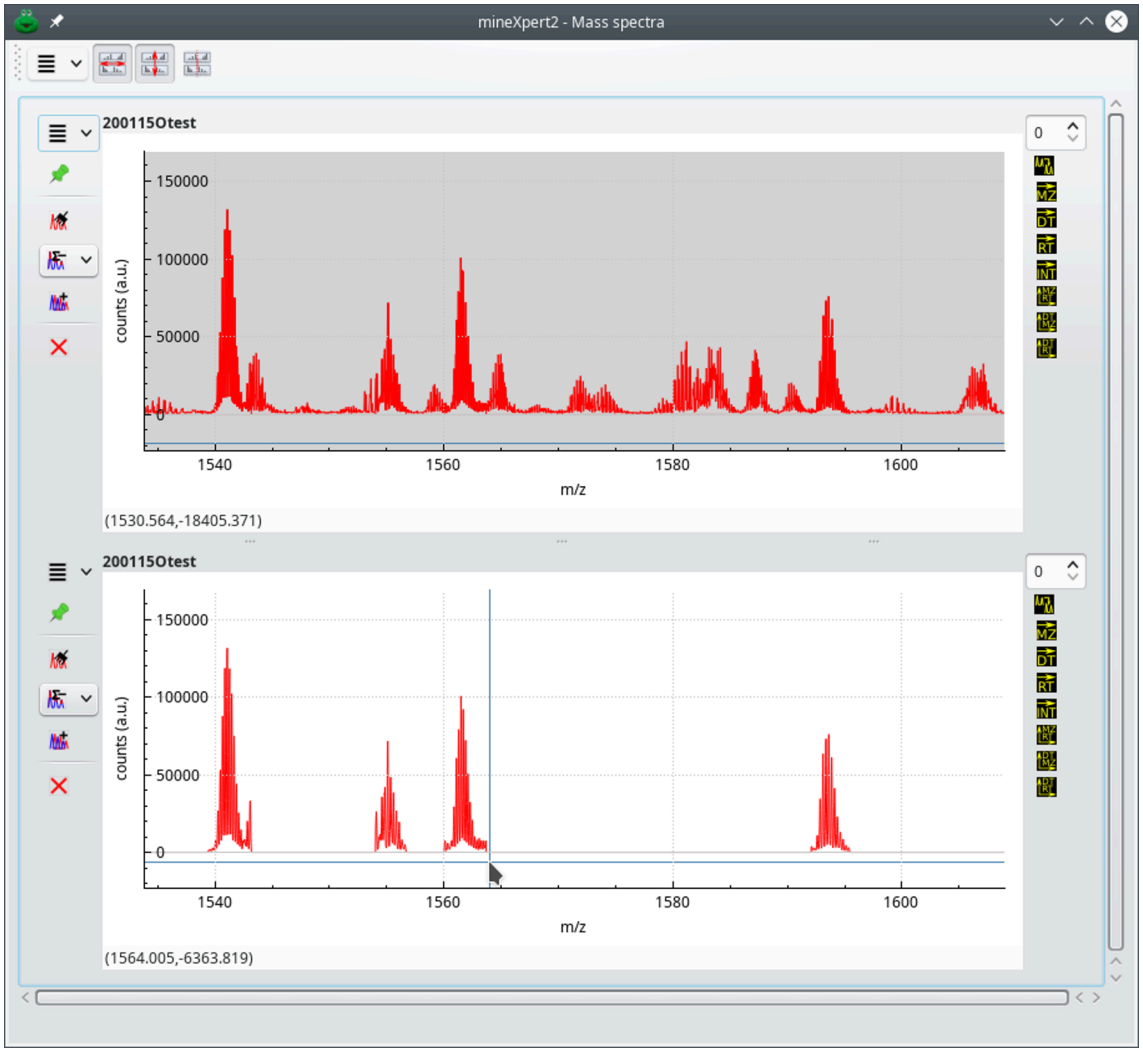
The menu item shows the *Low intensity threshold removal* menu item. The line edit widget allows one to enter the threshold below which signal is to be removed from the trace. Once the proper value has been entered, the user clicks the menu text. The program runs the filter and produces a new trace that is displayed along with the original trace.

FIGURE 4.20: **EXAMPLE OF LOW INTENSITY SIGNAL REMOVAL (CONFIGURATION)**



The new trace is displayed in the same composite plot widget. The user needs to zoom in enough to find it. Here, it is clearly visible at the location of the cursor near the baseline. It appears as a straight line under the mass peaks that were filtered out. In order to better visualize the new trace, select it and duplicate it using the duplication icon button at the top of the functional button on the right hand side of the composite plot widget.

FIGURE 4.21: **EXAMPLE OF LOW INTENSITY SIGNAL REMOVAL (CREATION OF A NEW TRACE)**



As is visible in the image, a great amount of signal present in the original mass spectrum (top plot) was removed from the trace (bottom plot). Note that the line connecting the isotopic cluster is faint where mass peaks were removed, but it is actually there. Note some remaining artifacts right after the very first isotopic cluster in the filtered trace. This is experimental work on progress.

FIGURE 4.22: **EXAMPLE OF LOW INTENSITY SIGNAL REMOVAL (DISPLAY OF THE NEW TRACE)**

4.6 Chained Integrations

The user, in the process of exploring the data, will inevitably chain integrations to pinpoint a specific feature of interest. For example, let's say that the user is exploring ion mobility mass spectrometry data. [2] After having loaded the data file, the TIC chromatogram is computed and displayed. From there, it is possible to perform a TIC chromatogram to an intensity = f(dt,m/z) color map integration (see Figure 3.11, "The int = f(dt, m/z) color map window" .)

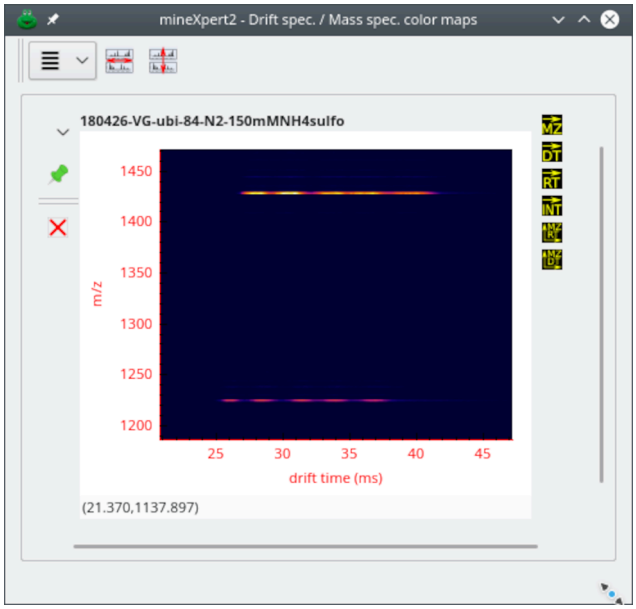


FIGURE 4.23: EXAMPLE OF CHAINED INTEGRATIONS, COLOR MAP WHERE DATA EXPLORATION MAY START

There starts the exploration. The user sees that there are a number of species having discrete drift times at the m/z ratio around 1220 (lower region of the colormap). They thus integrate to a single drift spectrum that horizontal lower region of the colormap. The obtained drift spectrum is shown at the left hand side of Figure 4.24, " Multiple chained integrations " .

Because there are five drift peaks in the drift spectrum, the user performs as many individual mass data integrations to a mass spectrum, from left to right. The obtained mass spectra are all shown in the window next to the *Drift spectra* window of the same figure.

2 Thanks to Dr Valérie Gabelica (IECB, Bordeaux, France) for permission to use her data set.



FIGURE 4.24: **MULTIPLE CHAINED INTEGRATIONS**

Most interestingly, the various drift regions are integrated to almost identical m/z values in their respective mass spectrum. In order to know when the various molecular species eluted in the chromatogram, the user performs for each mass spectrum an integration to a XIC chromatogram. The XIC chromatograms are shown on the window next to the *Mass spectra* window. Visibly, each molecular species was eluting from the chromatography column at discrete retention times (this was clearly not a *true* chromatography but instead an infusion during which instrument parameters were changed to modify the mobility properties of the ubiquitin molecule).

At this point, it is interesting to confirm that each XIC chromatogram contains exactly the molecular species that were responsible for the appearance of the purple "bands" on the int = f(dt, mz) initial color map. The different XIC chromatograms are thus integrated to a color map and the results are shown below the initial color map (less one color map that did not fit in the window). Visibly, the extracted color maps reconstitute the initial pattern visible at the top of the *Drift spec / mass spec color maps* window.

Note

The dataset used for Section 4.6, " Chained Integrations " was kind courtesy of Dr . Valérie Gabelica and correspond to a work entitled *Optimizing Native Ion Mobility Q-TOF in Helium and Nitrogen for Very Fragile Noncovalent Structures* published in *JASMS* with DOI: 10.1007/s13361-018-2029-4.

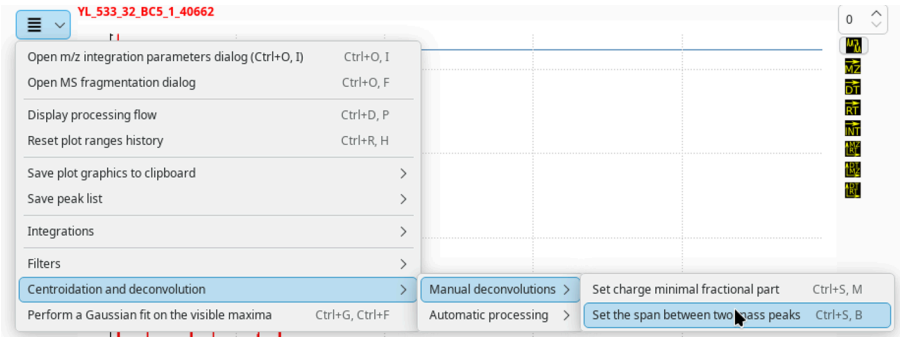
5 Mass Spectral Deconvolutions

When analysing a mass spectrum, there are two available deconvolution modes:

- Manual deconvolutions:
  - Isotopic cluster-based deconvolutions (Section 5.2, "Deconvolution based on isotopic cluster peaks");
  - Charge state envelope-based deconvolutions (Section 5.1, "Deconvolution based on charge state");
- Automated deconvolutions (Section 5.3, "Automated deconvolution based on isotopic cluster peaks")

Deconvolutions are performed in order to get back to the M_r mass of the analyte while reading m/z values. In the following sections, all the avaiable deconvolution processes are described.

Before delving into the deconvolutions, it is necessary to present two menu options that are found in the plot widgets contained in the *Mass spectra* window: the menu items under the *Centroidation and deconvolutions* menu (Figure 5.1, "Mass spectrum plot widget-specific deconvolution menu").



The two menu items are needed to configure the mouse-based deconvolution of mass spectra.

FIGURE 5.1: **MASS SPECTRUM PLOT WIDGET-SPECIFIC DECONVOLUTION MENU**

These two menus allow one to set parameters for the manual deconvolutions (see text for details).

5.1 Deconvolution based on charge state

In this kind of deconvolution, at the present time, the software assumes that the ionization agent is the proton and that the ionization is positive.

The deconvolution is based on the determination of the distance between two peaks —consecutive or not— of a given charge state envelope. When the user ⇧-click-drags the cursor from one peak to another, the program tries to calculate if the distance between two peaks matches one or more charge difference(s). If so, it computes the molecular (M_r [^{12}C]-relative molecular mass) mass of the analyte.



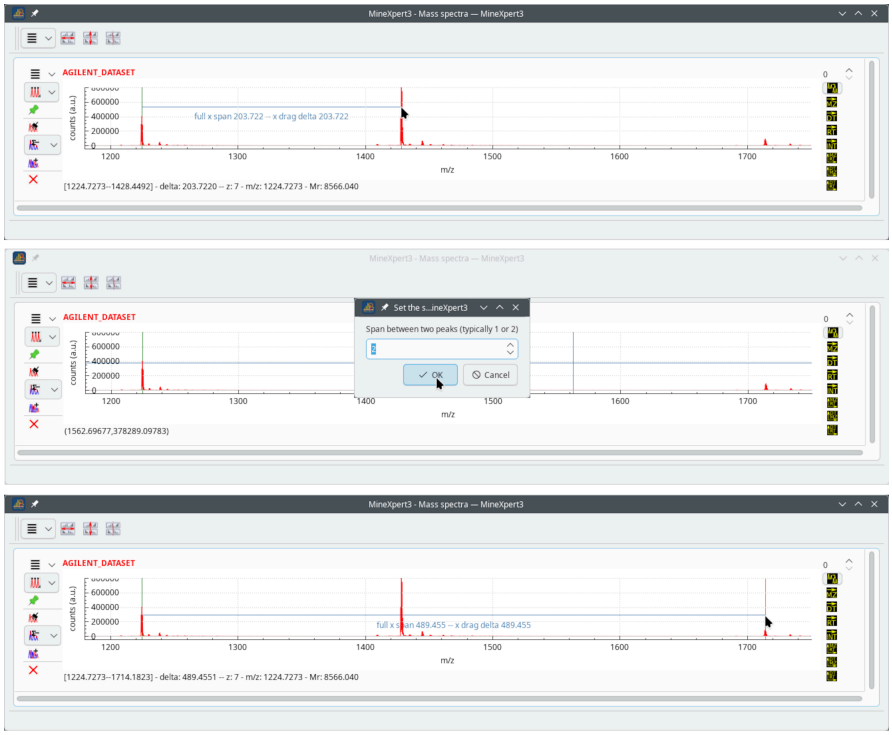
Note: The mouse drag movement is significant

The M_r value that is computed is for the analyte below the mass peak at which the mouse drag moment operation *started*.

Top panel of Figure 5.2, "Charge state-based mass deconvolution" shows the charge state envelope-based deconvolution process for a protein of $M_r \asymp 8566$ Da. In the top panel, the deconvolution has involved two consecutive peaks (*span*> is 1). The mouse drag movement occurred from left to right. Thus, the m/z value chosen for the computation is that under the left peak (start point of the movement). The status line at the bottom of the plot indicates the selection range, the delta movement on the x-axis, the computed charge, the M_r value chosen for the calculation (the M_r value at the *start* of drag movement) and finally the calculated M_r value.

This "two consecutive mass peaks" method is the default method. However, it might happen that no clearly visible mass peak is available around one nice peak that might be chosen as the start of the mouse drag operation. In this case it is possible to define a different *span*> between two peaks elected for the deconvolution (see Figure 5.1, "Mass spectrum plot widget-specific deconvolution menu"). In the figure, than span has been set to 2, which means that the mouse drag movement encompasses three mass peaks: the movement start peak, one peak in the middle and finally the movement stop peak (the span is thus of two intervals between the extreme peaks).

The bottom panel of the figure now displays the same M_r value for the protein even if the span is now of two intervals.



Deconvolution approach using two peaks belonging to the same charge state envelope. The top deconvolution involves two consecutive mass peaks (peak span value is 1). The bottom deconvolution involves two non-consecutive peaks (peak span value is 2). The M_r value, expectedly, did not change whatever the configured span.

FIGURE 5.2: **CHARGE STATE-BASED MASS DECONVOLUTION**



Note

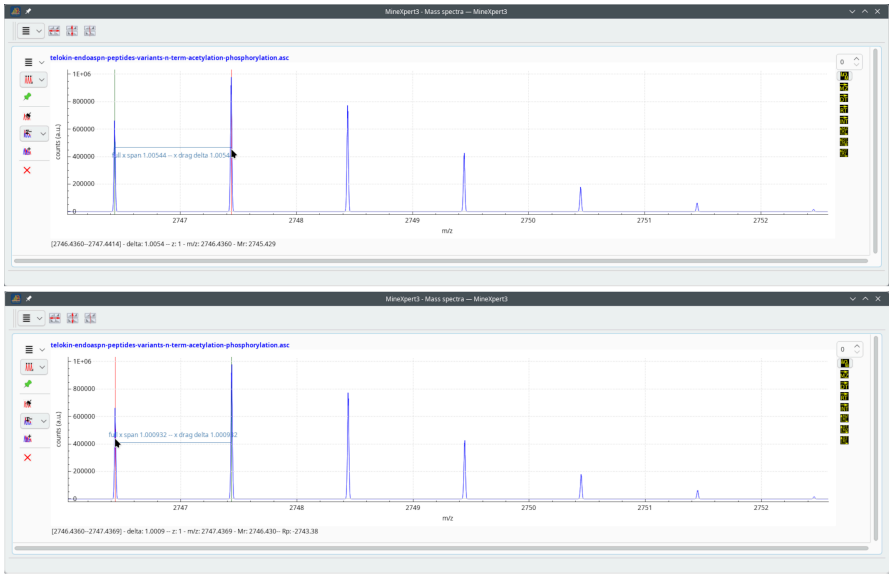
The charge calculation, which is at the heart of the deconvolution, almost never produces an integer value with no fractional part (say, charge z=15.0) because it is almost impossible to drag the mouse cursor the exact number of pixels that would match a m/z range leading to such an integral charge value. Almost always, the charge that is calculated looks like 14.995 or 15.001, for example. This is due to the fact that the mouse moves at discrete positions on the screen and these positions might be more or less far apart, depending on the mouse capabilities and on the current zoom factor over the mass spectrum region of interest.

It is advised to zoom-in as much as possible over the peaks at hand so as to minimize the difficulties above. It may happen, however, that even zoomed-in peaks are not sufficiently distant to allow a charge calculation. In this case, reduce the stringency over the fractional part that is allowed in the charge (see menu item *Set charge minimal fractional part* at FIGURE 5.1, "MASS SPECTRUM PLOT WIDGET-SPECIFIC DECONVOLUTION MENU"). By default, the stringency is set at 0.99, that is, any calculated value that has a fractional part either superior or equal to 0.99 or inferior or equal to 0.01 would lead to a successful round-up/round-down to the nearest integer value. Outside of the [0.01-0.99] interval, no charge calculation

is performed and thus no deconvolution is performed. When the stringency is too high, reducing it will allow the deconvolution to be carried-over. General experience is that setting that value to 0.997 is fine for most situations and provides very reliable results.

5.2 Deconvolution based on isotopic cluster peaks

In this kind of deconvolution, the user ⌥-click-drags the cursor between the first two peaks (when possible) of the isotopic cluster. The charge state of the ion is the inverse of the distance between the two consecutive peaks (that is, the m/z delta value). Figure 5.3, "Isotopic cluster-based mass deconvolution" shows that deconvolution process at work.



Top panel: deconvolution involving a left-to-right mouse drag movement. Bottom panel: deconvolution involving a right-to-left mouse drag movement. The calculated M_r values differ.

FIGURE 5.3: ISOTOPIC CLUSTER-BASED MASS DECONVOLUTION

Note: The mouse drag movement start position is significant

The ⌥-click-dragging direction (left→right or right→left) determines the final M_r that is computed because that value is calculated for the peak under the mouse when the mouse drag movement is initiated. This is visible in the two panels of Figure 5.3, "Isotopic cluster-based mass deconvolution", where the top panel shows the M_r computed for the left peak and the bottom panel shows the M_r computed for the right peak. Since the ion is monocharged, the difference is 1 Da.

This is a significant departure from the previous versions, where the postulate was that the single real peak of interest in an isotopic cluster was the left-most *monoisotopic* peak. Since this software has been used by scientists in research projects using almost 100 % labelled bacteria (with [^{13}C] and [^{15}N]), that concept has become moot. Indeed, analytes from these bacteria have their monoisotopic peak at the far right end of the isotoopic cluster.

The new behaviour allows the scientists to compute the M_r value of the peak of interest in an isotopic cluster, be that for a non-labelled or for a labelled analyte. See the following articles as examples of heavy isotope almost full labelling of bacteria.

Heavy isotope labeling and mass spectrometry reveal unexpected remodeling of bacterial cell wall expansion in response to drugs. Atze H, Liang Y, Hugonnet JE, Gutierrez A, Rusconi F, Arthur M. *Elife*, 2022, doi: 10.7554/eLife.72863, PMID: 35678393 (https://pubmed.ncbi.nlm.nih.gov/35678393/) .

Peptidoglycan-tethered and free forms of the Braun lipoprotein are in dynamic equilibrium in Escherichia coli. Liang Y, Hugonnet JE, Rusconi F, Arthur M. *Elife*, 2024, doi: 10.7554/eLife.91598, PMID:39360705 (https://pubmed.ncbi.nlm.nih.gov/39360705/) .

(p)ppGpp modifies RNAP function to confer β-lactam resistance in a peptidoglycan-independent manner. Voedts H, Anoyatis-Pelé C, Langella O, Rusconi F, Hugonnet JE, Arthur M. *Nat Microbiol*, 2024, PMID:38443580 (https://pubmed.ncbi.nlm.nih.gov/38443580/) .

5.3 Automated deconvolution based on isotopic cluster peaks

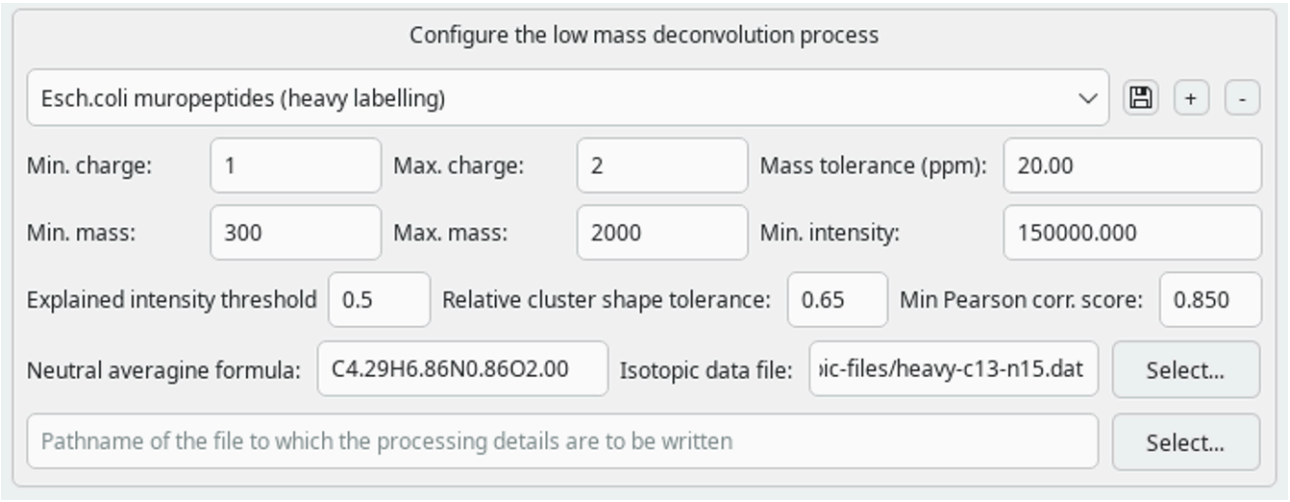
The automated deconvolution that is based on the isotopic cluster peaks is designed to be reliable for low-mass analytes (typically below 5 kDa). The process involves two separate steps: first, the centroids are extracted from the mass spectrum (see Section 4.1.7, " Centroid Extraction from Mass Spectra "); second, the extracted centroids are used for the deconvolution proper. The deconvolution is a highly complex process that would require extensive knowldedge to understand. The main concepts are the following:

- Iterate in each *observed* centroid (from smallest (m/z,i) values to greatest) of the input centroided mass spectrum and, for each centroid, assume that is has a given charge (in the charges range provided in the configuration; see below). With that assumed charge, compute a neutral analyte mass.

- The neutral analyte mass is then used to construct a *theoretical* neutral averagine-based isotopic cluster. Each centroid of the isotopic is then taken in sequence and subject to a matching process.

- The matching operation takes place by first converting back the theoretical neutral analyte cluster centroids to a charged analyte m/z value (same charge as that initially used to craft the neutral mass); second, the m/z value is searched for in the input mass spectrum centroids. When an input spectrum m/z value matches that of the theoretical isotopic cluster, a match is stored.
- When all the centroids of the theoretical isotopic cluster have been iterated into, the set of found matches is checked for reliability (typically, there must have been at least three matches in the cluster, albeit not necessarily contiguous). Other checks are performed also. If the checks are all positive, then the match is stored for the next step.
- At then end of the processing of all the input centroids, all the found matches are processed in order to consolidate them to a neutral mass. Each neutral mass is listed along with all the m/z value that was actually matched or the set of m/z values corresponding to different charges of the same analyte molecule.

The low mass deconvolution algorithm can be tightly configured using a large set of parameter settings that are detailed below and that are accessible via the application preferences window shown in FIGURE 5.4, "" and detailed below.



Interface to the configuration of the low mass deconvolution process.

FIGURE 5.4:

- *Min. charge (Max. charge)*: the charge range that is explored.
- *Mass tolerance (ppm)*: the tolerance with which all the mass matches are performed. A typical value is 20.
- *Min. mass (Max. mass)*: the neutral mass range that is explored.
- *Min. intensity*: the intensity value below which input mass spectrum centroids are discarded.
- *Explained intensity threshold*: Once the matches between the centroids in the theoretical isotopic cluster and those in the input centroided mass spectrum have been performed, the reliability of these matches is checked like so: select the monoisotopic centroid and its contiguous centroid in both the theoretical

isotopic cluster and in the matched centroided mass spectrum. Compute the ratio between the intensities of these centroids. That ratio is used to normalize the theoretical centroids intensities so they can be compared to the matched input mass spectrum centroid intensities. Compute the sum of the matched centroids intensities for all the matches (input mass spectrum vs theoretical). Compute the [input / theoretical] summed intensities ratio, which reflects the 'explained intensity'. The greater the ratio, the better the intensity correspondence between the theoretical and experimental isotopic clusters. That ratio value is required to be greater than the value set for this parameter for the matches to be considered reliable. Typical values: 0.5-0.7.

- *Relative cluster shape tolerance*: Once the matches between the centroids in the theoretical isotopic cluster and those in the input centroided mass spectrum have been performed, the reliability of these matches is checked like so: get the monoisotopic (highest probability centroid, isotopologue 0) match. Get the isotopologue 1 match. For each one of these two matches compute the ratio of the intensities for the theoretical isotopic cluster data and for the input centroided mass spectrum (observed data) This yields the following intensity ratios: [theor_ratio = theor. intensity iso 1 / theor. intensity iso 0] and [obser_ratio -> obser. intensity iso 1 / obser. intensity iso 0]. Compute the ratio [abs(obser_ratio - theor_ratio) / theor_ratio] which amounts to a 'normalized difference'. If the ratio above is smaller than this parameter setting, then the matches are considered to be reliable. Typical values: 0.6-0.7.

- *Min Pearson corr. score*: Once the matches between the centroids in the theoretical isotopic cluster and those in the input centroided mass spectrum have been performed, the reliability of these matches is checked like so: A Pearson correlation factor is computed to check that the observed and theoretical isotopic cluster centroids are correlated by their intensity. If the obtained correlation factor is greater or equal to this setting, then the matches are considered to be reliable. Typical values: 0.7-0.90.

- *Neutral averagine formula*: This the averagine formula to be used to model the isotopic clusters starting from a neutral analyte mass.

- *Isotopic data file*: The path name of the file holding the isotopic data. This is a file having the following format (removed data for clarity):

```
1,hydrogen,H,1,1.007825,1,0,0.999884,-0.000115,0
1,hydrogen,H,1,2.014101,2,1,0.0001157,-9.0644,0
2,helium,He,2,3.0160293,3,0,0.00000,-13.5206046 0 [
... ] 6,carbon,C,6,12.00000,12,0,0.9892119,-0.0108466,0
6,carbon,C,6,13.003354,13,1,0.0107,-4.529315,0
7,nitrogen,N,7,14.0030,-0.003648,0
7,nitrogen,N,7,15.00010,15,1,0.0036419,-5.615226,0
8,oxygen,O,8,15.99491,16,0,0.9975676 -0.00243,0
8,oxygen,O,8,16.999131,17,1,0.00038093,-7.872715,0
8,oxygen,O,8,17.999159,18,2,0.0020513,-6.189236,0
```



Note

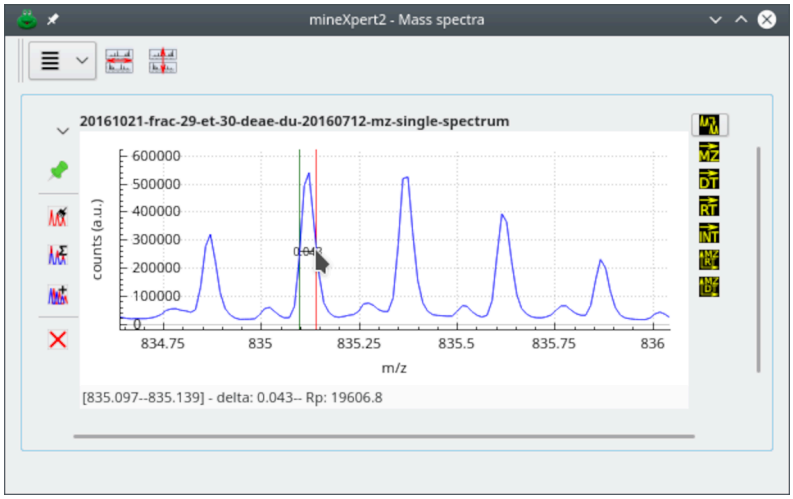
It is possible to use an isotopic data file that lists heavy isotopes as the most probable isotopes for any chemical element. In this manual author's research, bacteria are fully labelled with the heavy stable isotopes of carbon and nitrogen, which produces peptidoglycan structural elements that display as inverted isotopic clusters (with respect to natural isotopic abundance analytes). The deconvolver works also in this case by only detecting the input mass spectrum isotopic cluster having the proper inverted shape, and reports the analytes with the diagnostic ions at monoisotopic m/z values corresponding to the highest value in the isotopic cluster that was correctly recapitulated by the averagine-based theoretical isotopic cluster. This means that whatever the isotopic abundance configuration, the reported ions are correct: their monoisotopic m/z value is reported.

- The last field may be used to indicate the path name to a file where all the processing details are stored. Perusing these data might be interesting when the deconvolution results do not correspond to expected results.

When the deconvolution process has ended, the results are shown in the console window.

5.4 Reading the resolving power based on mass spectral data

When ⇧-click-dragging the mouse cursor between two mass spectrum locations of interest, the program computes the apparent resolving power. This process is shown on Figure 5.5, "Calculation of the resolving power" , where the resolving power is calculated by dragging the mouse cursor from one edge of a peak to the other at half maximum height (this is called *full width at half maximum* [FWHM] resolution).



Click-dragging the mouse cursor will trigger the calculation of the resolving power of the instrument. That value is printed in the status bar.

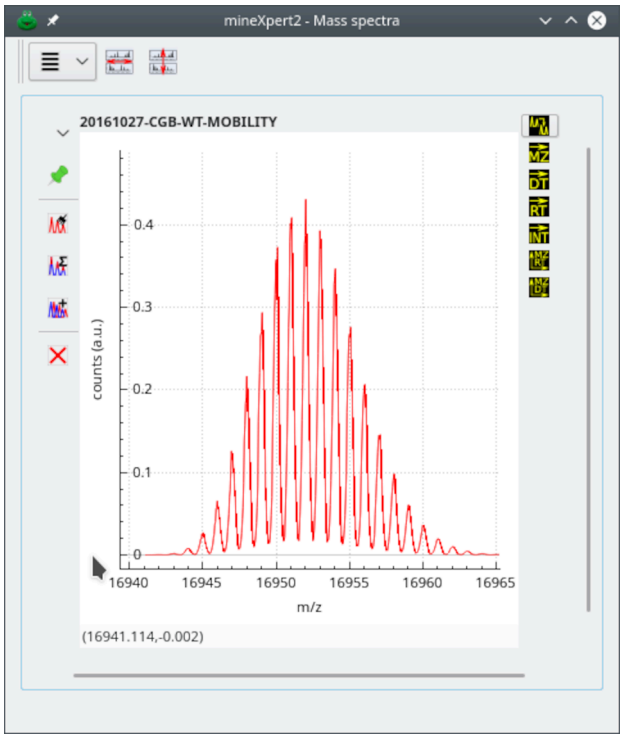
FIGURE 5.5: CALCULATION OF THE RESOLVING POWER

6 Isotopic cluster calculations

6.1 Calculating Isotopic Clusters with IsoSpec

When the resolution of the mass spectrometer is good, zooming-in on a mass peak may reveal that a given ion has given rise not to one peak but to a set of peaks. This set of peaks is called an "*isotopic cluster*".

It is possible to predict how a given ion (of known chemical formula) is supposed to be revealed in a mass spectrum, in the form of such an isotopic cluster. One such cluster is shown in Figure 6.1, "Calculation of the isotopic cluster of an analyte", for the horse apomyoglobin protein in its monoprotonated form $[M+H^+]^+$, of elemental composition $C_{769}H_{1213}N_{210}O_{218}S_2$ (this formula is typeset like this intentionally, to show how the formulæ need to be entered in the IsoSpec module).



Calculated isotopic cluster for the monoprotonated apomyoglobin protein.

FIGURE 6.1: CALCULATION OF THE ISOTOPIC CLUSTER OF AN ANALYTE

MineXpert provides an interface to the `libIsoSpec++` library.

> IsoSpec: Hyperfast Fine Structure Calculator
>
> Mateusz K. Łącki, Michał Startek, Dirk Valkenborg, and Anna Gambin
>
> *Analytical Chemistry*, 2017, 89, 3272–3277
>
> DOI: 10.1021/acs.analchem.6b01459

This library performs high-resolution isotopic cluster calculations. In order to run the calculations, it is necessary to have the following items ready:

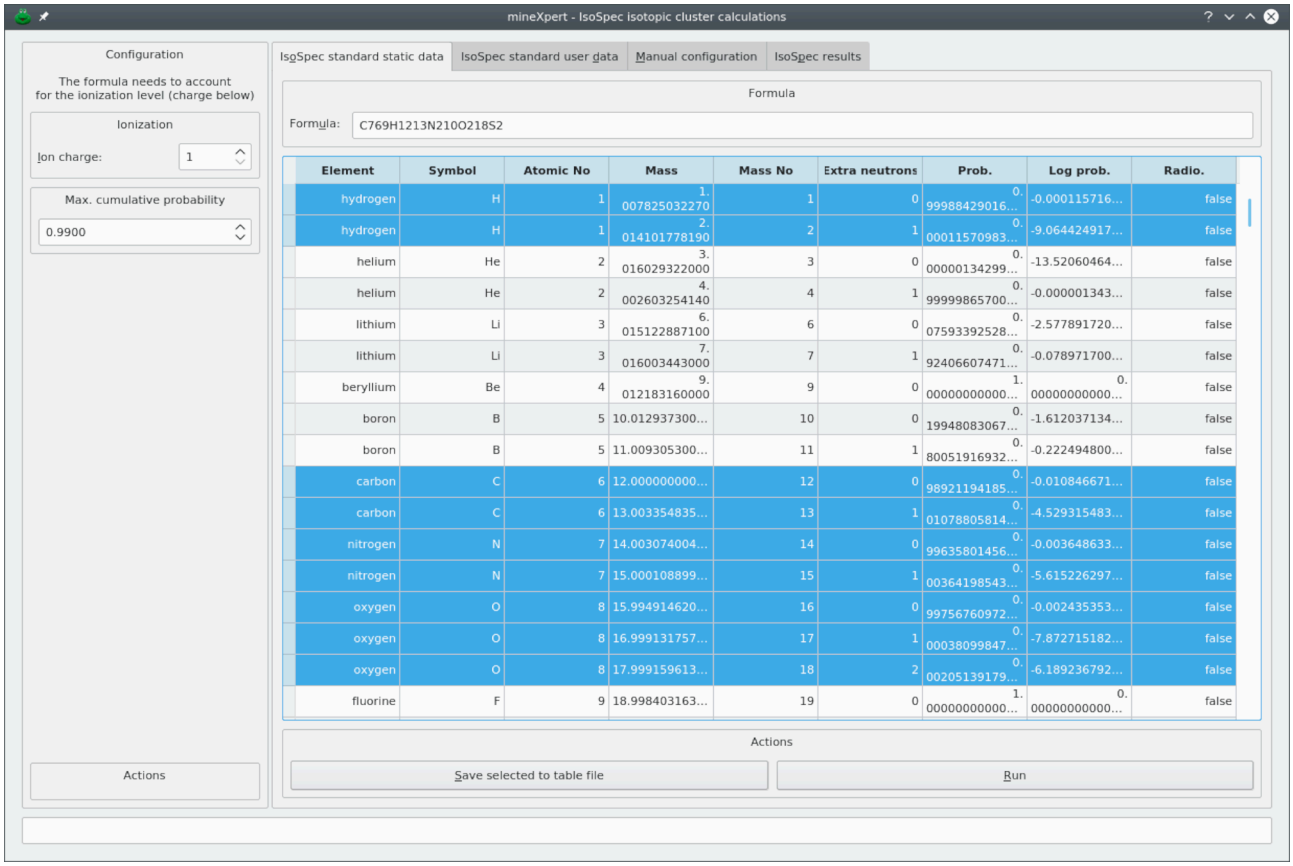
- An elemental composition formula of the analyte (for example, H_2O_1). This formula needs to account for the ionization agent that is involved in the ionization of the analyte prior to its detection in the mass spectrometer.

  > **Important**
  >
  > The IsoSpec software requires that all the chemical elements of a chemical formula be indexed. This means that, for water, for example, the formula should be H_2O_1 (notice the index 1 after the O element symbol).

- A detailed isotopic configuration of all the chemical elements that are used in the elemental composition formula. MineXpert provides two interfaces to define the isotopic characteristics of the chemical elements. These will be described in the following sections.

6.1.1 The IsoSpec Graphical User Interface in mineXpert

Generating iosotopic clusters using the IsoSpec software package is not easily carried over, in particular because this remarkable library is designed to be highly performant. The authors rightfully put their energy into optimizations for accuracy and speed instead of investing in a graphical user interface. MineXpert provides that graphical user interface, shown in Figure 6.2, "Isotopic cluster calculation dialog window", that shows up upon selection of the program's main window's *Utilities* › *Isotopic cluster calculations* menu.



The dialog window contains two panels. The left hand side panel configures the charge for which the calculation is to be carried over and the maximum cumulative isotopic presence probability that IsoSpec must reach during the calculation. The right hand side panel contains a tab widget that contains the configuration tabs and the results tab.

FIGURE 6.2: **ISOTOPIC CLUSTER CALCULATION DIALOG WINDOW**

An isotopic cluster calculation is most probably performed with the aim of simulating an expected isotopic cluster for an analyte that is being analyzed by mass spectrometry. It is thus logical that the analyte be in an ionized form. The way that the analyte has been ionized needs to be taken into account in the chemical formula that describes the ion for which the isotopic cluster is being calculated. For example, when determining the chemical formula of a protein in the positive ion mode, the number of protons used to ionize the protein need to be included in the analyte elemental composition formula.



Warning

The IsoSpec software is "charge-agnostic" in the sense that it does not know what element in the chemical formula is responsible for the ionization of the analyte. Therefore, IsoSpec does not know of (and does not care about) the charge of the analyte. The ionization level of the analyte can be handled by MineXpert if that information is set to the *Ion charge* spin box widget. By default, the charge state of the analyte is 1.

The *Max. cumulative probability* spin box widget serves to configure the extent to which IsoSpec simulates the theoretically expected isotopic cluster. A value of *0.99* tells the software to simulate enough combinations of the analyte isotopes to represent 99 % of the theoretically expected combinations.



Tip

For large biopolymers, it might be prudent to start with a relatively low value for *Max. cumulative probability*, because setting this value too high, that is, near 1, would increase notably the calculation duration.

To perform isotopic cluster calculations, the simulation software needs to be aware of all the isotopes of all the chemical elements that enter in the composition of the ionized analyte. An isotope is defined by its mass and by the probability that it is found in nature. Carbon has two major isotopes that can be found in nature: the [^{12}C] most abundant isotope and the [^{13}C] least abundant isotope (the [^{14}C] isotope is irrelevant for conventional mass spectrometry unless molecules have been artificially enriched in that isotope).

There are three ways that the user might use the IsoSpec software package. Two of them involve a configuration preparation on the part of the user. The third one, that we'll describe first, does not involve any chemical element configuration. The other ways are reviewed in the next sections.

6.1.1.1 Static Standard Element Tables Shipped within the IsoSpec Library

In order to document all the chemical elements' isotopes' characteristics, the IsoSpec library has, in its own headers, a number of arrays that MineXpert automatically loads up when opening the Figure 6.2, "Isotopic cluster calculation dialog window" dialog window.[1] These data are displayed in the *IsoSpec standard static data* tab. The table view widget is not editable, hence the *static* qualifier in the tab name.

In this tab, all the user has to do is enter the formula for which the isotopic cluster calculation is to be performed. The formula needs to be pasted in the *Formula* line edit widget.

Once the formula has been set to its line edit widget, configure the charge of the ion (*Ion charge* spin box widget) and the *Max. cumulative probability*. Now, click *Run*.

1 That process is performed at build time and is thus not configurable at run time.

6.1.1.2 Modified Standard Element Tables Shipped within the IsoSpec Library

While the previous section showed how to use the static element tables from the IsoSpec library, this section shows a way to configure these elemental data. For this, start from the standard IsoSpec-based data and modify them according to one's needs. To proceed, select all the element rows of interest from the *IsoSpec standard static data* tab and click *Save selected to table file*.



Tip

The order with which the rows are selected is respected in the export, so make sure to select the rows in the order that makes sense for you. Also, be sure to select rows and not only some cells; to achieve, this click in the left margin of the table view widget.

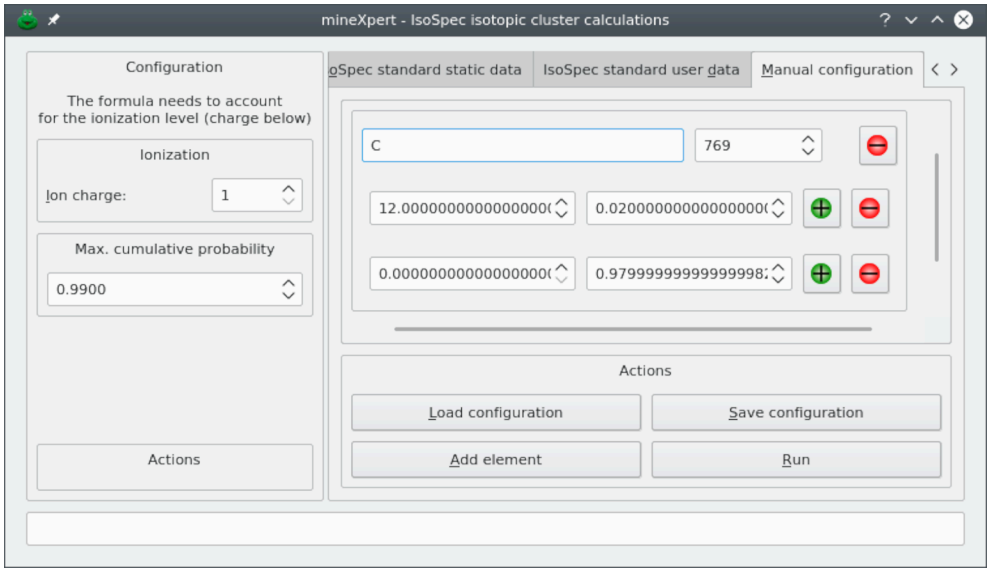
The export is performed as a text TSV (tab-separated value format) file in a layout that closely mimicks the data visible in the table view. Once the file has been saved, open it up in LibreOffice and modify the data to suit your needs. For example, in case of a labelling with [^{13}C], with an efficiency of almost 98 %, lets's change the [^{12}C] abundance (probability) to 0.02 and the [^{13}C] abundance to 0.98.

Now that the modifications were performed, save the file under the same format and load it in the *IsoSpec standard user data* tab of the dialog window by clicking *Load table from file*. The table view now shows the new carbon abundance values. From now on, use this tab exactly like you did for the standard isospec *static* data work described above.

Here also, it is possible to select determinate rows and then save them to a file for later use.

6.1.1.3 User hand-made configuration of the element data within mineXpert

The last way the user might configure the chemical elements to be used during an isotopic cluster calculation is based on the fully manual description of the elements and of the isotopes. That configuration is performed in the *Manual configuration* tab of the dialog window. This method is slightly more involved than the previous one but provides also for a much greater flexibility: it allows one to create "new chemical elements" that might be required in specific labelling experiments. The manual configuration is carried over in the *Manual configuration* tab of the dialog, as shown in Figure 6.3, "Hand-made user configuration of the chemical elements and formula".



When the dialog is created, the tab is empty. To start creating element definitions, click *Add element*.

FIGURE 6.3: **HAND-MADE USER CONFIGURATION OF THE CHEMICAL ELEMENTS AND FORMULA**

Upon creation of the dialog window, the *Manual configuration* tab is empty, with only two rows of buttons at the bottom of the tab. To start configuring chemical elements, click *Add element* to create an "element group box" that contains a number of widgets organized in two rows:

- Top row, a line edit widget to receive the chemical element symbol, *C* in the example;
- A spin box widget in which to set the number of such atoms in the formula for which the isotopic cluster is being calculated. In the example, we set this value to *769*;
- A button with a "minus" image that removes all the "element group box" in one go;
- The bottom row contains an "isotope frame widget" with two spin boxes for the mass of the isotope being configured (left) and its corresponding abundance (right);
- In addition to the spin boxes, two buttons, with a "plus" or a "minus" figure, allow one to respectively add or remove isotope frames.



Note

It is not possible to remove all the isotope frames from an element group box, otherwise that group box would become useless.

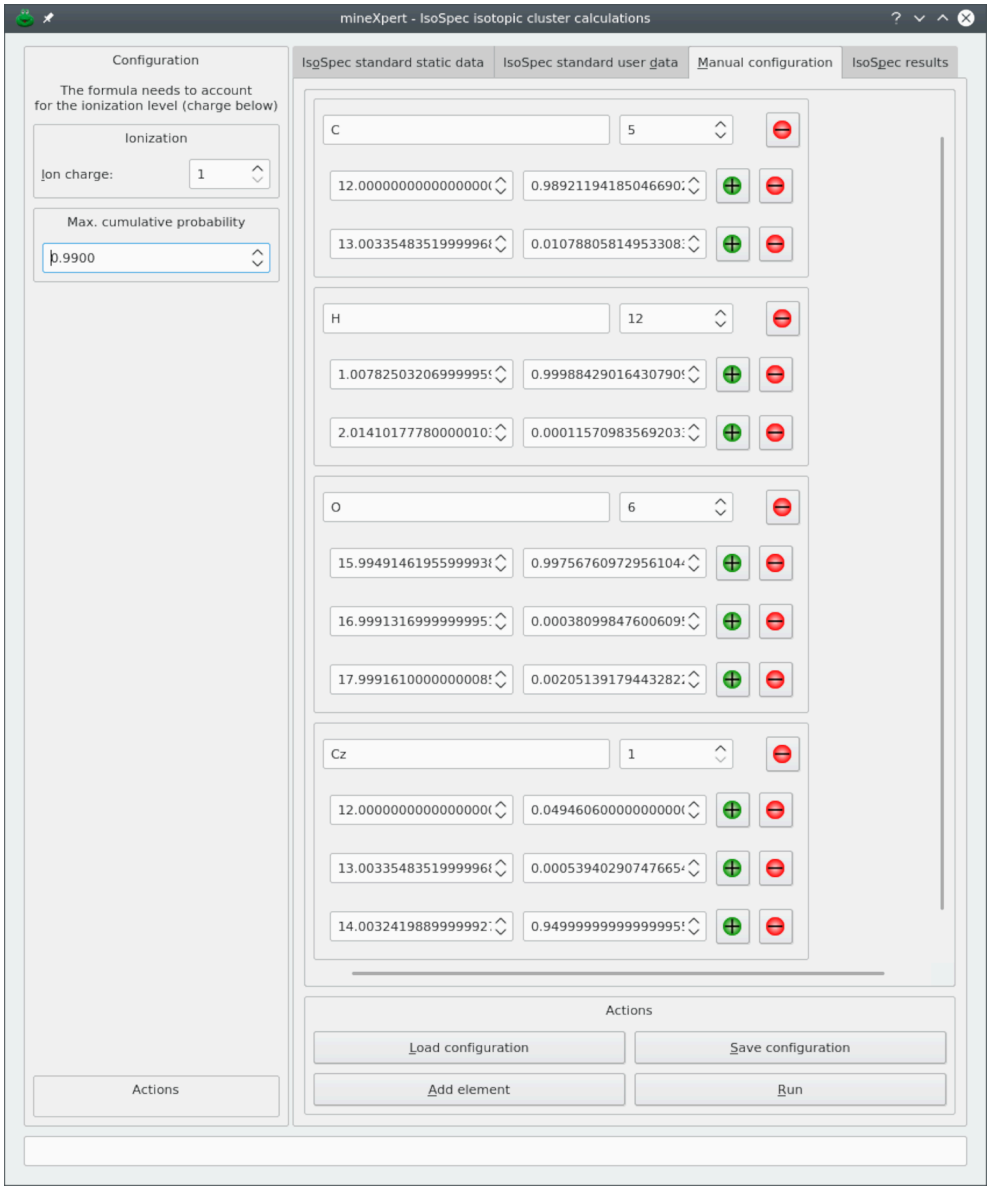
Once an isotope frame has been filled-up, a new line might be required. To create a new isotope frame widget, click any "plus"-labelled button in any of the isotope frames. Once a new frame is created, the spin box widgets that it contains are set to *0.00000*. Fill-in these spin boxes with mass and abundance and go on along this path to create as many isotopes as required.

Once all the isotopes for a given chemical element have been defined, a new element might be needed. For this, click *Add element* and start the configuration of the new element as described above.

The manual isotopic configuration of the chemical elements required to perform an isotopic cluster calculation for a given formula is tedious. The user may want to save a given configuration to a file (click *Save configuration*) so that it is easier to recreate automatically all the widgets upon loading of that saved configuration (click *Load configuration*).

The final configuration is shown in Figure 6.4, "Typical manual configuration of the isotopic characteristics of the chemical elements". The experiment that was configured above is a labelling of a glucose molecule with *Cz*, an imaginary chemical element that is like carbon but that has a [^{14}C]. The glucose molecule (normal formula: C6H12O6) is labelled on one single carbon atom with an efficieny of 95 %. This means that, when the labelling fails (in 5 % of the cases) the carbon atom has its isotopes with usual probabilities (compounded by the fact that the normal atom is found at that position only in 5 % of the cases). The isotopic abundances for the Cz element are thus:

- For [^{12}C]: 0.05 * normal [^{12}C] abundance;
- For [^{13}C]: 0.05 * normal [^{13}C] abundance;
- For [^{14}C]: 0.95;



The user has configured a labelling experiment where the glucose molecule is labelled at a single carbon position with a [^{14}C] atom (the efficiency of the labelling is 95 %).

FIGURE 6.4: TYPICAL MANUAL CONFIGURATION OF THE ISOTOPIC CHARACTERISTICS OF THE CHEMICAL ELEMENTS

Note

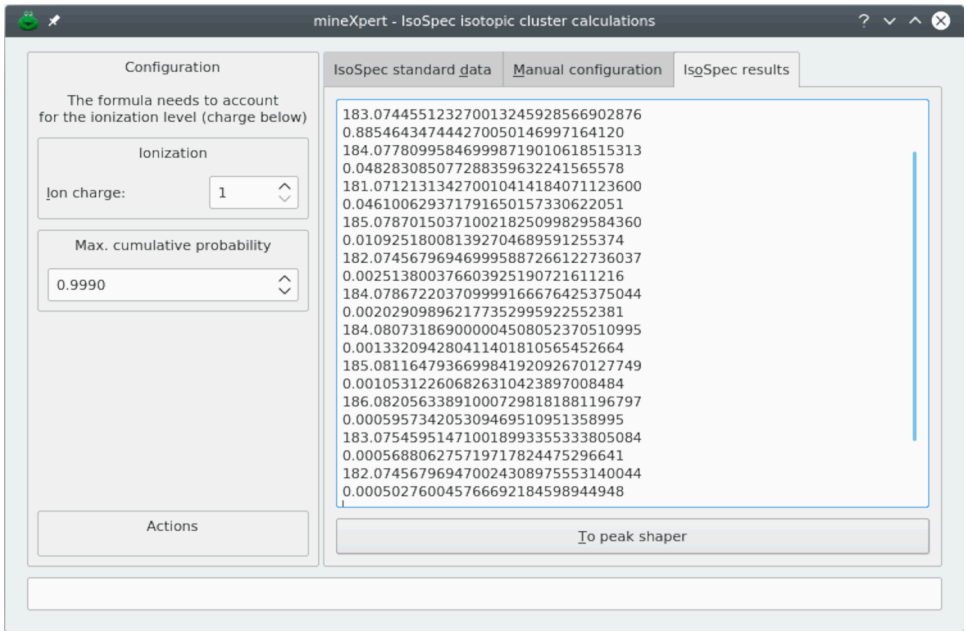
Note that the "normal" carbon count is 5 (and not 6), that the hydrogen count is 13 (and not 12, because the glucose is protonated) and the labelled carbon is present only once.

6.1.1.4 The IsoSpec results are not shaped mass peaks

Once the configurations have been terminated, the calculations can finally be performed by the IsoSpec library. In the manual configuration setting, the formula is automatically handled since each chemical element that is defined goes along with the count of the correponding atoms. In the case of the standard IsoSpec configuration (either modified or not), the user has to enter the chemical formula of the analyte in the *Formula* line edit widget.

Click *Run*. If the configuration was correct and IsoSpec could run the calculation properly, then the dialog window switches to the *IsoSpec results* tab (Figure 6.5, "Results from the isotopic cluster calculation"). That tab contains a text edit widget in which the results are displayed.

Note that the m/z values calculated by IsoSpec are "corrected" for the charge level that was specified in the left panel of the dialog window prior to their display in the results tab (Figure 6.2, "Isotopic cluster calculation dialog window").



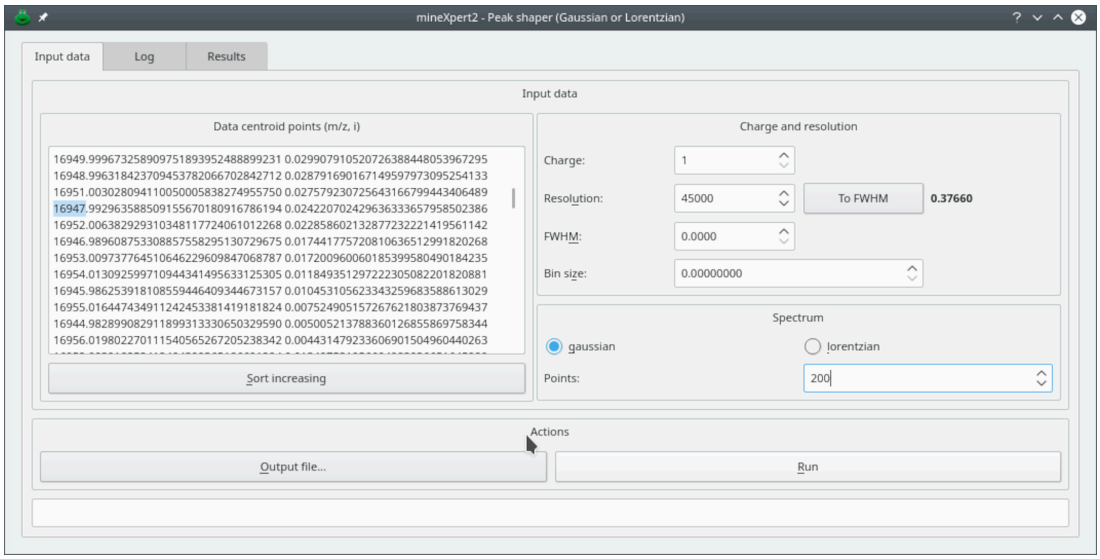
The IsoSpec library computes the probability of the various combinations of all the isotopes that make the chemical formula submitted to it. The results are in the form of peak centroid values along with corresponding probabilities. The sum of the probabilities corresponds to the *Max. cumulative probability* value that was set by the user.

FIGURE 6.5: RESULTS FROM THE ISOTOPIC CLUSTER CALCULATION

The results that are produced by IsoSpec represent the peak centroids of the isotopic cluster. The results are thus a set of (m/z,i) pairs that have not the characteristic shape (the profile) that is found in mass spectra. MineXpert features the ability to give a shape to the centroid peaks. For that, click *To peak shaper* to open the *Peak Shaper* dialog window preloaded with the IsoSpec-generated peak centroids. The workings of this peak shaping feature is described in Section 6.2, "Shaping mass peak centroids into well-profiled peaks".

6.2 Shaping mass peak centroids into well-profiled peaks

The shape of mass peaks is typically Gaussian or Lorentzian (or a mix thereof). There are some data simulation or analysis processes that lead to having mass peaks characterized by a single centroid m/z value and a corresponding intensity. Plotted to a graph, a centroid mass peak yields a bar. In order to convert mass peak centroids into something that resembles a real "profile" mass peaks, a mathematical formula can be applied, with some parameters to configure the shapes generated. MineXpert includes that feature, accessible *via* the *Utilities* › *Mass peak shape calculations* menu item. The window that opens up is shown in Figure 6.6, "Setting-up of the centroid mass peak shaping process"



This dialog window allows one to configure the shaping of mass centroid peaks. Setting a spectrum name in the *Mass spectrum name* line edit widget will help recognize the result mass spectrum once displayed in the *Mass spectra* window (see below).

FIGURE 6.6: SETTING-UP OF THE CENTROID MASS PEAK SHAPING PROCESS

The mass centroid peaks are listed in the *Data centroid points (m/z,i)* text edit widget. These values are pasted there by the user or copied automatically from the isotopic cluster calculation dialog window (see Section 6.1.1.4, "The IsoSpec results are not shaped mass peaks"). The width of the "profile" mass peak is determined either by setting the resolution of the instrument (in the example, that is set to 45000) or by setting the width of the peak at half maximum of its height (FWHM).

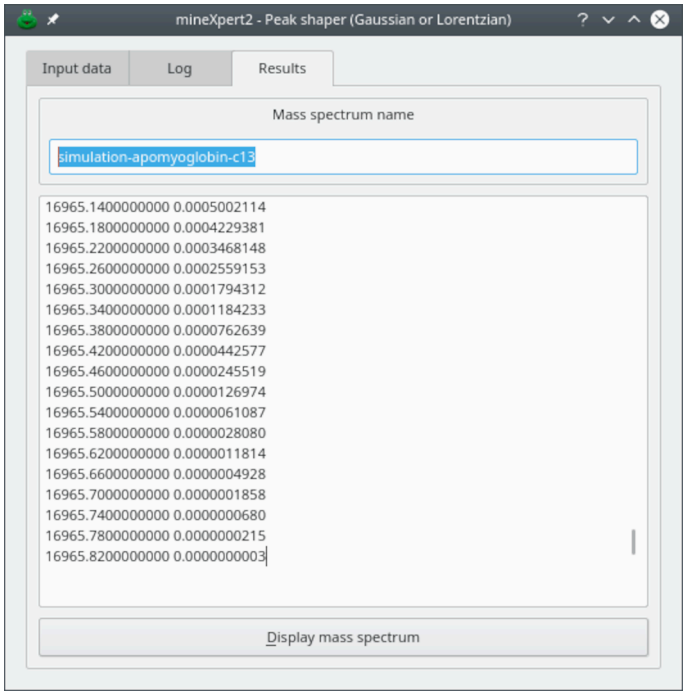
The spectrum that is generated can be of a Gaussian or a Lorentzian shape. That parameter is configured by selecting the corresponding radio button widget. The number of points used to actually craft the shape of the peak is configurable. In the example, that parameter is set to 200.



Tip

When defining the parameters for the peak shaping process, it might be useful to have and idea of the FWHM value at a given m/z value for a given resolution. To compute that FWHMvalue, double-click-select a single mass peak centroid m/z value from the text edit widget, set the resolving power of your instrument and then click *To FWHM*. The computed FWHM value will be displayed next to the button. The text is selectable so as to copy it to the clipboard and then in the *FWHM* spin box widget.

Once the peak shaping parameters have been set, click *Run*. The dialog window shifts tab to *Results*, as shown in Figure 6.7, "Shaped peaks mass spectral data".



The mass spectral data corresponding to the combination of each individual "peak-shaped" centroid peak are displayed in this tab. Plotting these data will produce a nicely shaped isotopic cluster matching what would be seen by recording a mass spectrum on an instrument.

FIGURE 6.7: **SHAPED PEAKS MASS SPECTRAL DATA**



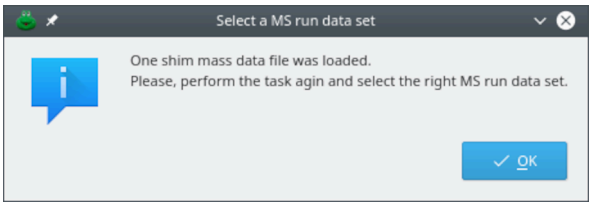
Tip

To ensure that various isotopic cluster calculation-based mass spectra can later be recognized, the user is advised, for the different simulations, to insert distinct names in the *Mass spectrum name* line edit widget. This name will be used later when creating the mass spectrum if the user asks that the mass

spectrum be displayed by clicking *Display mass spectrum*. In the eventuality that the mass spectrum name is not changed from a simulation to the other, a safeguard process ensures that names are absolutely unambiguous by appending to the mass spectrum name the time at which the mass spectrum is displayed.

It is possible to automatically plot the result mass spectrum by clicking *Display mass spectrum*. The process gets a little involved here because the mass spectrum is not created from a file, and MineXpert's handling of plots is based on the fact that a MS run data set has been created, typically at data file loading time. Two situations may arise:

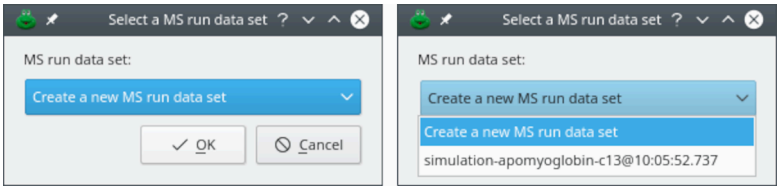
- If not a single MS run data set is currently available (that is, no single mass spectrometry data file was loaded during the current session), the program will inform the user that a shim MS run data set was created (Figure 6.8, "Creation of a shim MS run data set"). At this time, the dialog window listing all the open MS run data sets will show up (Figure 3.3, "The open MS run data sets"). The information message advises to perform the task again, that is, click *Display mass spectrum*.



This information dialog window informs the user that a new shim MS run data set was created. Shim MS run data sets need to be created when new plots are created for data that were not loaded from a disk file. The user is advised to repeat the last action. In this case, click *Display mass spectrum* again. The next step is described in the text below.

FIGURE 6.8: **CREATION OF A SHIM MS RUN DATA SET**

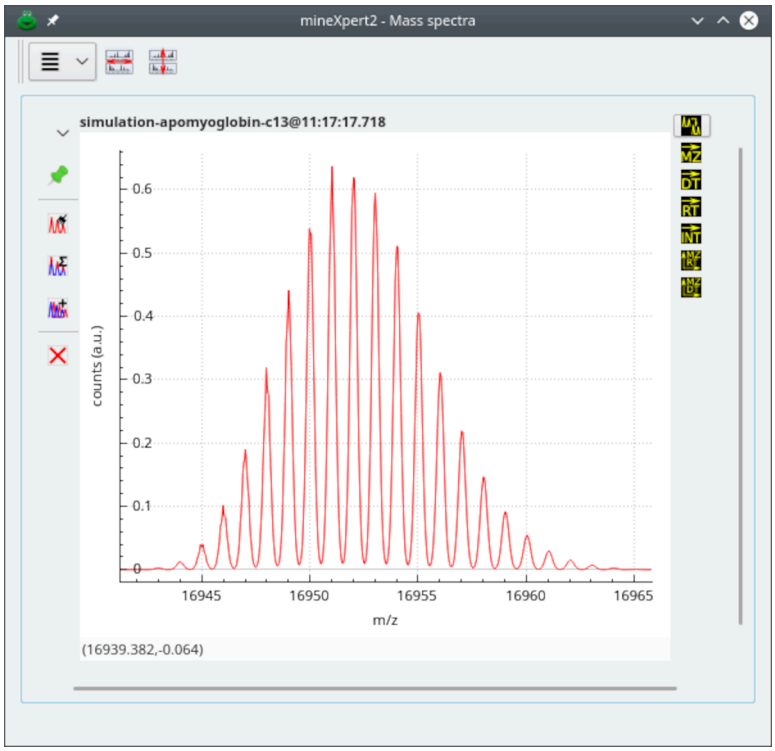
- If at least one MS run data set was loaded already in the program, then an input dialog window shows up letting the user select the MS run data set to which the isotopic cluster calculation mass spectrum is to be anchored to (Figure 6.9, "Selection of a MS run data set"). It is always possible to ask that a new MS run data set be created *ex nihilo* for the new mass spectrum by selecting *Create a new MS run data set*. In this case, the process goes back to the previous item. This process is useful if each mass spectrum must have a distinct color.



If one or more MS run data set(s) preexisted in the program, this input dialog shows up (left). Upon clicking the combo box widget, the user can select the MS run data set to anchor the new mass spectrum to (right).

FIGURE 6.9: **SELECTION OF A MS RUN DATA SET**

The isotopic cluster calculation-based mass spectrum shows up in the *Mass spectra* window, as shown in Figure 6.10, "Spectrum created using the peak shaping feature".



The spectrum corresponds to a combination of each individual spectrum obtained by shaping each individual mass peak centroid in the input data list.

FIGURE 6.10: **SPECTRUM CREATED USING THE PEAK SHAPING FEATURE**

Note that if the resolution asked is very high, the resulting shaped mass peaks might appear a bit "hairy". By tweaking the *Bin size* value, the binning of the spectra might improve the situation. Otherwise, using the plot widget main menu to apply a Savitzky-Golay filter, described at Section 4.1.6, " Savitzky-Golay filtering of any kind of data " (page 56), will certainly improve things.



Tip

If the peak centroids were not "corrected" for their charge in the previous generation step (as in the case of the isotopic cluster calculation, it is still time to apply this "correction" by setting the charge in the *Charge* spin box widget. If the charge was already accounted for, as described in Section 6.1.1.4, "The IsoSpec results are not shaped mass peaks" (page 90), then leave the charge to *1* and the results will be correct.

7 The processing flow in a data exploration session

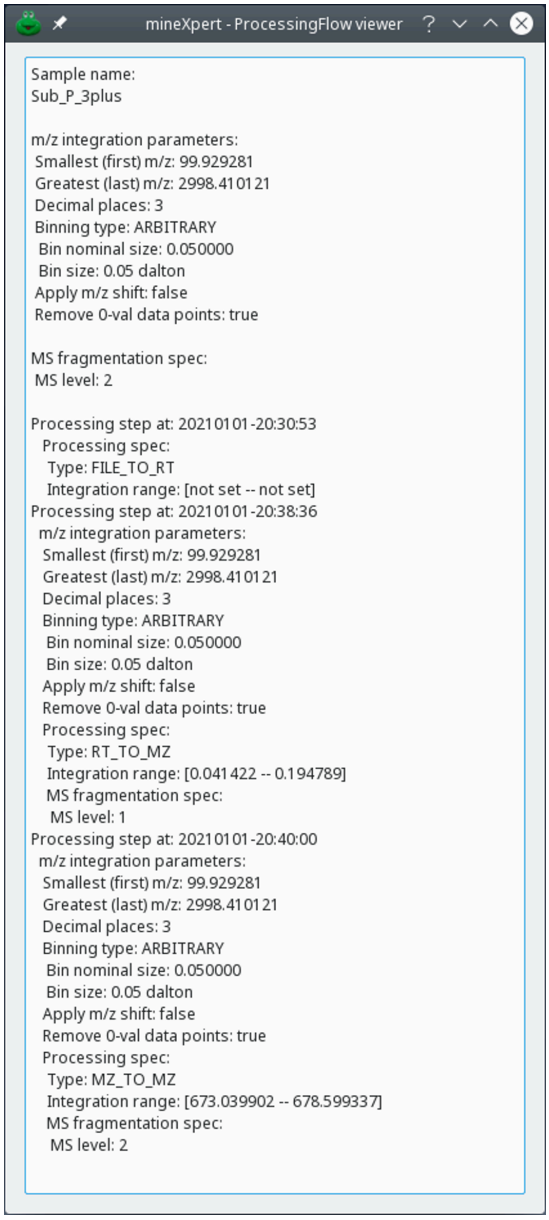
The mass spectral data sets that are open in a giving data exploration session are identifiable by the color of the traces. All the trace plots or color map plots obtained for a given mass spectral data set are distinguished from the other by the color used for the plotting (trace plots) or for the axes and labels (color map plots). The color is thus a means to maintain a visual "filiation" between plots. But for the various steps of mass spectral data exploration session involving the same mass spectra data set, there needs to be a way to record all these steps. That means is the subject of this chapter and is called *Processing flow*.

7.1 The processing flow records all the steps in a data exploration session

As said above, the "filiation" of the plots is maintained using identifying colors. However, color is not enough to unambiguously identify the "filiation" of any given plot. Indeed, the same TIC chromatogram or color map plot can be used multiple times as the starting point to perform multiple integrations. For example, from a given TIC chromatogram, one would be interested in a number of different chromatogram peaks, between various time retention ranges. By performing integrations for each peak of interest, the user will generate a number of plots that actually originate from the very same TIC chromatogram of the very same mass data set. The newly created plots will have the same color as the one of the originating plot, but it will not be possible to distinguish between all the "child" plots. This is why whenever a new plot is created, a processing flow is attached to it that corresponds to the same processing flow of the originating plot plus a new step with the last integration details. This way, each plot will carry with itself the whole integration "history" that led to its creation. This processing flow is shown upon selection of the *Show processing flow* menu item from the main menu belonging to the composite plot widget that contains the plot for which the "history" is enquired about.

A processing flow is a collection of as many steps are necessary to describe thoroughly all the processing steps that were required to produce a given trace or color map plot. A processing step is described using as many processing specifications as necessary to thoroughly describe that step.

A typical processing flow of a currently selected trace (or the only trace plotted in the composite plot widget) is shown like depicted in Figure 7.1, "The processing flow dialog window".



Each trace or color map plot has a *Processing flow* record that shows all the steps that led to it through the exploration session.

FIGURE 7.1: **THE PROCESSING FLOW DIALOG WINDOW**

This figure shows the processing flow for a MS/MS spectrum (MS level 2). The text details all the steps, from the loading of the mass data file to the obtention of the MS/MS spectrum. In this example, the processing flow lists the following informative elements:

- The *Sample name*: name of the sample inside of the mass spectral data file (a file might contain more than one *run*, each run having a sample name).

- The *m/z integration parameters*: the values listed in this section are relative to the whole mass spectral data set, like *Smallest (first) m/z* or *Greatest (last) m/z*, for example. The other values, like *Decimal places*, or the *Binning type* are the last values used during the course of the integrations. These values describe the configuration of the most recent integration step. In the present case, the last integration was a MS level 2 integration and the *MS level* is thus described to be of value 2.

- *Processing step*: processing step containing processing specifications (*Processing spec*). The processing steps are recorded with a time tag so that it is possible to sort them out in increasing time order. In the example, the first processing specification is of type *FILE_TO_RT* thus describing the data file loading of the mass spectral data set. There is of course no integration range for that specification as the loading of a mass spectral data set is not formally an integration.

- Another processing step was performed later in the data exploration session and was recorded in another processing step. The specification is of type *RT_TO_MZ*, describing the integration from the TIC chromatogram to a mass spectrum[1].That integration only accounted for mass spectra acquired between the values listed for the *Integration range*. The integration only took into account MS 1 level spectra.

- A third processing step describes the second integration, and this step started at a mass spectrum (the mass spectrum generated at the previous integration), as described with the information bit *Type: MZ_TO_MZ*. Indeed, the integration began in a mass spectrum and yielded a mass spectrum. The MS level is set to 2 because that integration dealt with MS/MS spectra recorded for the precursor ions contained in the m/z range mentioned for information bit *Integration range*.

1 *RT* stands for "retention time", that is, TIC or XIC chromatogram; *MZ* stands for mass spectrum; *DT* stands for drift spectrum.

8 Recording data exploration discoveries

When doing mass analysis work it is often desirable to store the painstakingly manually picked m/z or M_r values for later use. MineXpert provides a flexible process allowing one to record the data discoveries during the exploration of a mass data set to any combination of the following destinations: file, clipboard, console.

In order to record the innumerous analysis steps that result from a data exploration session, the configuration of the discoveries storage must have been done by accessing *Preferences* › *Edit preferences...* and then selecting the *Discoveries* section, as described in Figure 2.3, "Application preferences - Discoveries page".

In that window, the user can select the destination of the data analysis recording system: console, clipboard, file or any combination of the three. When selecting file recording, the user might specify if the recording would overwrite any preexisting file or, instead, append to that file. Depending on the kind of plot where data exploration occurs, the format of the data to be recorded needs to change. Indeed, it would make no sense to record the charge *z* when exploring data in the *Drift spectra* window. This is why the text format of the data export needs to be defined for each one of the three kinds of plots: TIC/XIC chromatogram, mass spectrum or drift spectrum (*Specialized formats* group box).

Before delving into the configuration intricacies, let us tell immediately how to trigger the recording of the exploration discoveries: using the SPACE bar in the composite plot widget containing the graph being analyzed.

The text format specification can contain particular tokens as described below.



Note

Almost all the format string elements below are numeric (representing either integer or float values). For float values, it is possible to define the number of decimals to be used when printing the value. The decimals specifier is of the form ".<number>", like ".3".

When used with a format string, that gives "%X.3", which specifies that when printing out the x-axis value, 3 decimals should be added after the decimal point. To prevent decimal at all, do not use any decimal specifier at all ("%X").

- *%f* : mass spectrometry data file name.
- *%s* : sample name.
- *%X* : value on the x-axis of the graph (no unit). For a drift spectrum, that would be ion mobility times, for a mass spectrum, that would be m/z values, for a TIC/XIC chromatogram, that would be retention times.
- *%Y* : value on the y-axis of the graph (no unit). This value corresponds to the value at the cursor point (not at a point on the curve). In all the graph plots, that would be intensities in any unit provided by the mass spectrometer (typically, counts).

- *%C* : value on the y-axis of the graph (no unit). This value corresponds to the value at the curve of the selected trace plot (not at the cursor point). In all the graph plots, that would be intensities in any unit provided by the mass spectrometer (typically, counts).
- *%x* : delta value on the x-axis (when appropriate, no unit).
- *%y* : delta value on the y-axis (when the mouse has been dragged enough to draw a rectangle). The delta value is computed as the difference between the y-axis value at the mouse drag start point and the y-axis value at the mouse drag stop point.
- *%c* : delta value on the y-axis (when the mouse has been dragged enough to draw a rectangle). The delta value is computed as the difference between the y-axis value *on the curve* at the mouse drag start point and the y-axis value *on the curve* at the mouse drag stop point.
- *%I* : TIC intensity after TIC integration over a graph range
- *%b* : x-axis range *begin* point for the computation (where applicable, for example for the TIC integration to a single value).
- *%e* : x-axis range *end* point for the computation (where applicable).
- *For mass spec plot widgets:*
  - *%z* : charge
  - *%M* : M_r as computed during deconvolution (see Chapter 5, Mass Spectral Deconvolutions).



Note

It is important to keep in mind that the *%z* and *%M* format strings can only work if the user is actually analyzing a mass spectrum and if the user has *effectively* performed a deconvolution operation that has allowed computing these two values. If the values are not available, the program shows the format string untouched ("%M", for example) in the textual output upon hitting the SPACE bar (see below). Likewise, the *%I* format string will only hold meaningful data if an integration to a single TIC intensity value has been determined (see Figure 4.14, " Integrating mass data from a mass spectrum to a single TIC intensity value ") for that specific data analysis record.

The recording of data analysis steps works in any trace plot widget (not yet in the color map plot widgets) even if there are more than one trace in a given plot widget. In the case the plot widget has more than one trace and none or more than one is selected, the program will ask that a single trace be selected. If data are to be scrutinized for more traces, simply select each trace in turn and trigger the analysis stanza to be output using the SPACE bar each time.

Once configured, the format strings might be stored in a drop down box for later use. To that end, click onto the *Add to history* button while having the format text displayed in the text editor and it will be appended to the drop-down list. The list gets stored when the dialog window is closed and will be filled-up again when the program is restarted.

As an example, if the user defined the following format string for a mass spectrum:

```
Mass spec. :
mz = (%X, %Y) z = %z
filename = %f
date = 20161021
session = 20161021
mslevel = 1 msion = esi msanal = tof
chrom = DEAE fraction = 25
seq =  pos =    oxlevel = 0 pos =
intensity =
comment =
```

Then, a resulting data exploration stanza that would be recorded will look like this:

```
Mass Spec. :
mz = (1051.8, 50863) z = 1
filename = 20161017-rusconi-frac-25-deae-20160712.mzml
date = 20161021
session = 20161021
mslevel = 1 msion = esi msanal = tof
chrom = DEAE fraction = 25
seq =  pos =    oxlevel = 0 pos =
intensity =
comment =
```

Interestingly, the user can define any kind of format, leaving fields available for later filling-in. This feature is of immense value when the analysis file is used later to fill-in a database for easy storage and interrogation of the exploration discoveries. In this case, it would be useful to have the file opened in an editor and at each new stanza, edit the *comment* field if something needs to be commented, like the shape/intensity of a mass peak, for example.

Note that the program closes the file each time a new stanza has been written. This makes it possible to edit that file safely in between each stanza record. Remember to force the editor to reload the file from disk after each exploration discovery recording.

When the recording involves sending the analysis data to the console, the data are sent to it as text that is colored the same as the spectrum that was under scrutiny.

When the mouse cursor has been placed at the proper location on the graph (with or without Ctrl-click-dragging, depending on the situation), the user hits the SPACE bar and the data analysis stanza is recorded to the selected destination(s): console, clipboard, file.

9 Scripting Interface

The interface to scripting has been conceived as a console allowing the user to automate a number of tasks. It is, however, currently in a preliminary development state.

9.1 Starting the Scripting Console Invocation

The scripting console window is easily called by pulling down the *Scripting* menu item from the MineXpertprogram's *Utilities*menu, as represented in Figure 9.1, “Starting the scripting console window” .

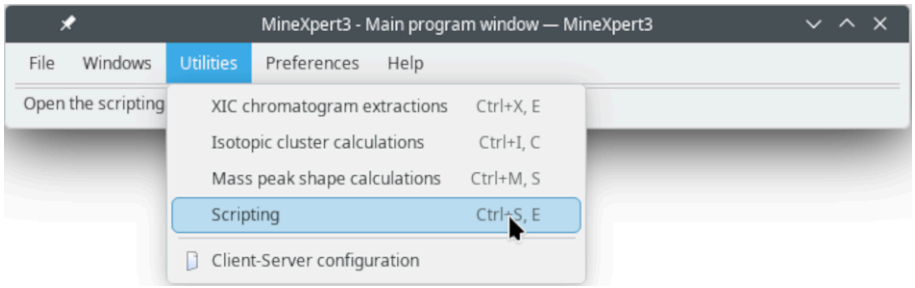
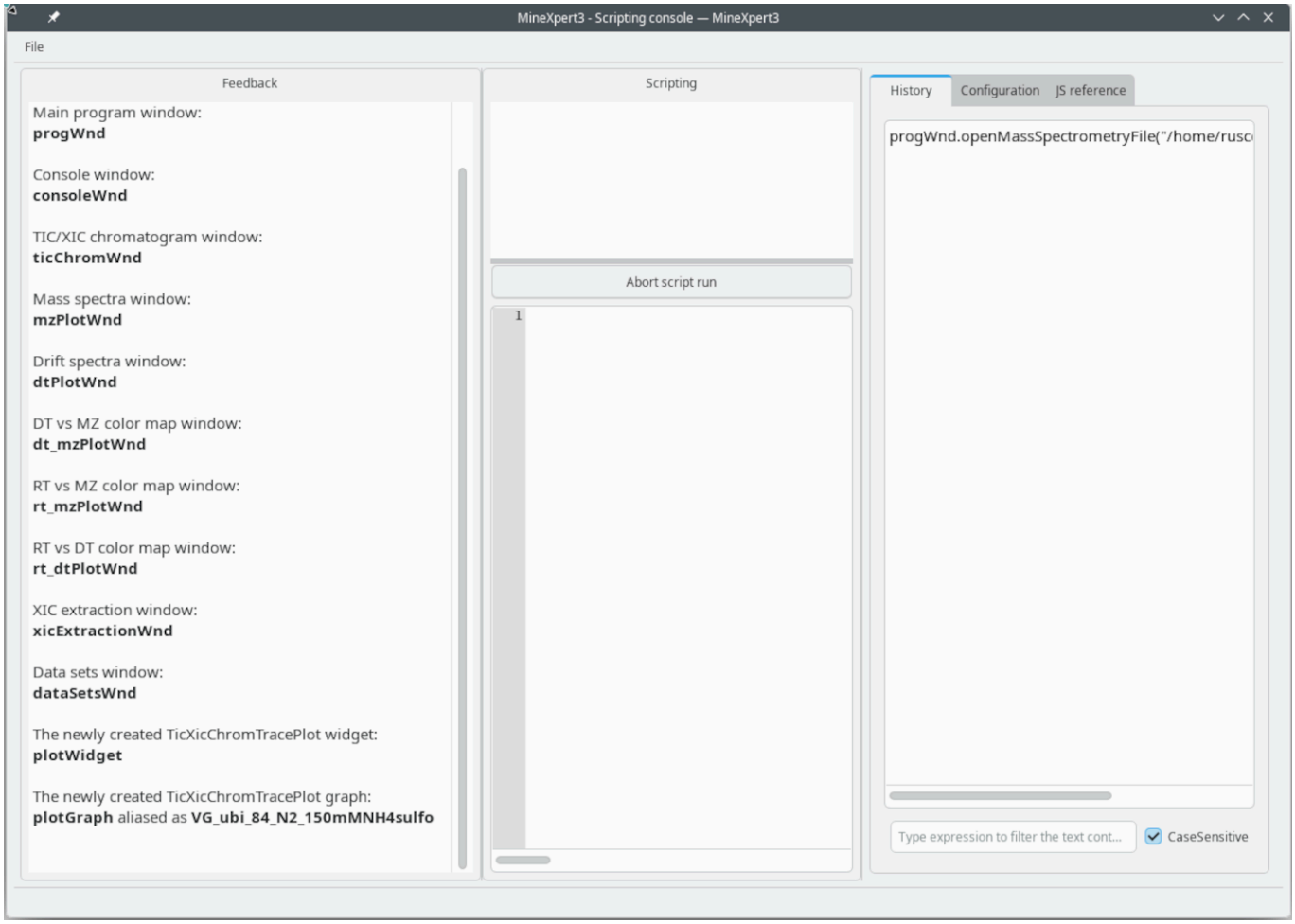


FIGURE 9.1: **STARTING THE SCRIPTING CONSOLE WINDOW**

9.2 The Scripting Console Window

The scripting console window is described at Figure 9.2, “The Scripting Console Window” .



The window is divided into three vertical panes. The left hand side pane collects all the program's feedback to the user, so they are informed of any new variable that is created in the JavaScript environment and made available to them. The middle pane has two regions. The upper region will display all the return values after any JavaScript command has been run. The lower region is the actual script input region. The right hand side region contains mainly the history of the commands issued during the previous and current scripting sessions.

FIGURE 9.2: THE SCRIPTING CONSOLE WINDOW

9.3 Running Scripts in the Scripting Console

The lower part of the middle panel of the scripting window is a text editor that has syntax colouring for JavaScript code. After having entered the script, hitting `Enter` while maintaining `Ctrl` pressed will run the script.

Upon running of the script, the output of the script is printed in the upper part of the same middle panel. Most often the output will be *undefined* (in grey colour), which is just fine. If the output is some actual value, it will be printed in green. If there was an error, the ouptut will be printed in red.

Note that each time a script is run, it is appended in the right hand side panel of the window, in the the *History* tab. The coloring of the items added to the *History*> tab follows the same logic as that described above.

9.4 Loading Script Files in the Scripting Console

It is possible to craft ECMAscript-compliant scripts outside of the JavaScript console window, store them in files and open them for running in the JavaScript console window. To do so, select *File→ Load a JavaScript file in the editor* .

Once the script has been inserted into the script editor widget, the user can run it by hitting `Enter` while maintaining `Ctrl` pressed.

A GNU General Public License version 3

Version 3, 29 June 2007

Copyright © 2007 Free Software Foundation, Inc. https://fsf.org/

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

Preamble

The GNU General Public License is a free, copyleft license for software and other kinds of works.

The licenses for most software and other practical works are designed to take away your freedom to share and change the works. By contrast, the GNU General Public License is intended to guarantee your freedom to share and change all versions of a program—to make sure it remains free software for all its users. We, the Free Software Foundation, use the GNU General Public License for most of our software; it applies also to any other work released this way by its authors. You can apply it to your programs, too.

When we speak of free software, we are referring to freedom, not price. Our General Public Licenses are designed to make sure that you have the freedom to distribute copies of free software (and charge for them if you wish), that you receive source code or can get it if you want it, that you can change the software or use pieces of it in new free programs, and that you know you can do these things.

To protect your rights, we need to prevent others from denying you these rights or asking you to surrender the rights. Therefore, you have certain responsibilities if you distribute copies of the software, or if you modify it: responsibilities to respect the freedom of others.

For example, if you distribute copies of such a program, whether gratis or for a fee, you must pass on to the recipients the same freedoms that you received. You must make sure that they, too, receive or can get the source code. And you must show them these terms so they know their rights.

Developers that use the GNU GPL protect your rights with two steps: (1) assert copyright on the software, and (2) offer you this License giving you legal permission to copy, distribute and/or modify it.

For the developers' and authors' protection, the GPL clearly explains that there is no warranty for this free software. For both users' and authors' sake, the GPL requires that modified versions be marked as changed, so that their problems will not be attributed erroneously to authors of previous versions.

Some devices are designed to deny users access to install or run modified versions of the software inside them, although the manufacturer can do so. This is fundamentally incompatible with the aim of protecting users' freedom to change the software. The systematic pattern of such abuse occurs in the area of products for individuals to use, which is precisely where it is most unacceptable. Therefore, we have designed this version of

the GPL to prohibit the practice for those products. If such problems arise substantially in other domains, we stand ready to extend this provision to those domains in future versions of the GPL, as needed to protect the freedom of users.

Finally, every program is threatened constantly by software patents. States should not allow patents to restrict development and use of software on general-purpose computers, but in those that do, we wish to avoid the special danger that patents applied to a free program could make it effectively proprietary. To prevent this, the GPL assures that patents cannot be used to render the program non-free.

The precise terms and conditions for copying, distribution and modification follow.

TERMS AND CONDITIONS

0. Definitions.

"This License" refers to version 3 of the GNU General Public License.

"Copyright" also means copyright-like laws that apply to other kinds of works, such as semiconductor masks.

"The Program" refers to any copyrightable work licensed under this License. Each licensee is addressed as "you". "Licensees" and "recipients" may be individuals or organizations.

To "modify" a work means to copy from or adapt all or part of the work in a fashion requiring copyright permission, other than the making of an exact copy. The resulting work is called a "modified version" of the earlier work or a work "based on" the earlier work.

A "covered work" means either the unmodified Program or a work based on the Program.

To "propagate" a work means to do anything with it that, without permission, would make you directly or secondarily liable for infringement under applicable copyright law, except executing it on a computer or modifying a private copy. Propagation includes copying, distribution (with or without modification), making available to the public, and in some countries other activities as well.

To "convey" a work means any kind of propagation that enables other parties to make or receive copies. Mere interaction with a user through a computer network, with no transfer of a copy, is not conveying.

An interactive user interface displays "Appropriate Legal Notices" to the extent that it includes a convenient and prominently visible feature that (1) displays an appropriate copyright notice, and (2) tells the user that there is no warranty for the work (except to the extent that warranties are provided), that licensees may convey the work under this License, and how to view a copy of this License. If the interface presents a list of user commands or options, such as a menu, a prominent item in the list meets this criterion.

1. Source Code.

The “source code” for a work means the preferred form of the work for making modifications to it. “Object code” means any non-source form of a work.

A “Standard Interface” means an interface that either is an official standard defined by a recognized standards body, or, in the case of interfaces specified for a particular programming language, one that is widely used among developers working in that language.

The “System Libraries” of an executable work include anything, other than the work as a whole, that (a) is included in the normal form of packaging a Major Component, but which is not part of that Major Component, and (b) serves only to enable use of the work with that Major Component, or to implement a Standard Interface for which an implementation is available to the public in source code form. A “Major Component”, in this context, means a major essential component (kernel, window system, and so on) of the specific operating system (if any) on which the executable work runs, or a compiler used to produce the work, or an object code interpreter used to run it.

The “Corresponding Source” for a work in object code form means all the source code needed to generate, install, and (for an executable work) run the object code and to modify the work, including scripts to control those activities. However, it does not include the work's System Libraries, or general-purpose tools or generally available free programs which are used unmodified in performing those activities but which are not part of the work. For example, Corresponding Source includes interface definition files associated with source files for the work, and the source code for shared libraries and dynamically linked subprograms that the work is specifically designed to require, such as by intimate data communication or control flow between those subprograms and other parts of the work.

The Corresponding Source need not include anything that users can regenerate automatically from other parts of the Corresponding Source.

The Corresponding Source for a work in source code form is that same work.

2. Basic Permissions.

All rights granted under this License are granted for the term of copyright on the Program, and are irrevocable provided the stated conditions are met. This License explicitly affirms your unlimited permission to run the unmodified Program. The output from running a covered work is covered by this License only if the output, given its content, constitutes a covered work. This License acknowledges your rights of fair use or other equivalent, as provided by copyright law.

You may make, run and propagate covered works that you do not convey, without conditions so long as your license otherwise remains in force. You may convey covered works to others for the sole purpose of having them make modifications exclusively for you, or provide you with facilities for running those works, provided that you

comply with the terms of this License in conveying all material for which you do not control copyright. Those thus making or running the covered works for you must do so exclusively on your behalf, under your direction and control, on terms that prohibit them from making any copies of your copyrighted material outside their relationship with you.

Conveying under any other circumstances is permitted solely under the conditions stated below. Sublicensing is not allowed; section 10 makes it unnecessary.

3. Protecting Users' Legal Rights From Anti-Circumvention Law.

No covered work shall be deemed part of an effective technological measure under any applicable law fulfilling obligations under article 11 of the WIPO copyright treaty adopted on 20 December 1996, or similar laws prohibiting or restricting circumvention of such measures.

When you convey a covered work, you waive any legal power to forbid circumvention of technological measures to the extent such circumvention is effected by exercising rights under this License with respect to the covered work, and you disclaim any intention to limit operation or modification of the work as a means of enforcing, against the work's users, your or third parties' legal rights to forbid circumvention of technological measures.

4. Conveying Verbatim Copies.

You may convey verbatim copies of the Program's source code as you receive it, in any medium, provided that you conspicuously and appropriately publish on each copy an appropriate copyright notice; keep intact all notices stating that this License and any non-permissive terms added in accord with section 7 apply to the code; keep intact all notices of the absence of any warranty; and give all recipients a copy of this License along with the Program.

You may charge any price or no price for each copy that you convey, and you may offer support or warranty protection for a fee.

5. Conveying Modified Source Versions.

You may convey a work based on the Program, or the modifications to produce it from the Program, in the form of source code under the terms of section 4, provided that you also meet all of these conditions:

a. The work must carry prominent notices stating that you modified it, and giving a relevant date.

b. The work must carry prominent notices stating that it is released under this License and any conditions added under section 7. This requirement modifies the requirement in section 4 to "keep intact all notices".

c. You must license the entire work, as a whole, under this License to anyone who comes into possession of a copy. This License will therefore apply, along with any applicable section 7 additional terms, to the whole of the work, and all its parts, regardless of how they are packaged. This License gives no permission to license the work in any other way, but it does not invalidate such permission if you have separately received it.

d. If the work has interactive user interfaces, each must display Appropriate Legal Notices; however, if the Program has interactive interfaces that do not display Appropriate Legal Notices, your work need not make them do so.

A compilation of a covered work with other separate and independent works, which are not by their nature extensions of the covered work, and which are not combined with it such as to form a larger program, in or on a volume of a storage or distribution medium, is called an "aggregate" if the compilation and its resulting copyright are not used to limit the access or legal rights of the compilation's users beyond what the individual works permit. Inclusion of a covered work in an aggregate does not cause this License to apply to the other parts of the aggregate.

6. Conveying Non-Source Forms.

You may convey a covered work in object code form under the terms of sections 4 and 5, provided that you also convey the machine-readable Corresponding Source under the terms of this License, in one of these ways:

a. Convey the object code in, or embodied in, a physical product (including a physical distribution medium), accompanied by the Corresponding Source fixed on a durable physical medium customarily used for software interchange.

b. Convey the object code in, or embodied in, a physical product (including a physical distribution medium), accompanied by a written offer, valid for at least three years and valid for as long as you offer spare parts or customer support for that product model, to give anyone who possesses the object code either (1) a copy of the Corresponding Source for all the software in the product that is covered by this License, on a durable

physical medium customarily used for software interchange, for a price no more than your reasonable cost of physically performing this conveying of source, or (2) access to copy the Corresponding Source from a network server at no charge.

c. Convey individual copies of the object code with a copy of the written offer to provide the Corresponding Source. This alternative is allowed only occasionally and noncommercially, and only if you received the object code with such an offer, in accord with subsection 6b.

d. Convey the object code by offering access from a designated place (gratis or for a charge), and offer equivalent access to the Corresponding Source in the same way through the same place at no further charge. You need not require recipients to copy the Corresponding Source along with the object code. If the place to copy the object code is a network server, the Corresponding Source may be on a different server (operated by you or a third party) that supports equivalent copying facilities, provided you maintain clear directions next to the object code saying where to find the Corresponding Source. Regardless of what server hosts the Corresponding Source, you remain obligated to ensure that it is available for as long as needed to satisfy these requirements.

e. Convey the object code using peer-to-peer transmission, provided you inform other peers where the object code and Corresponding Source of the work are being offered to the general public at no charge under subsection 6d.

A separable portion of the object code, whose source code is excluded from the Corresponding Source as a System Library, need not be included in conveying the object code work.

A “User Product” is either (1) a “consumer product”, which means any tangible personal property which is normally used for personal, family, or household purposes, or (2) anything designed or sold for incorporation into a dwelling. In determining whether a product is a consumer product, doubtful cases shall be resolved in favor of coverage. For a particular product received by a particular user, “normally used” refers to a typical or common use of that class of product, regardless of the status of the particular user or of the way in which the particular user actually uses, or expects or is expected to use, the product. A product is a consumer product regardless of whether the product has substantial commercial, industrial or non-consumer uses, unless such uses represent the only significant mode of use of the product.

“Installation Information” for a User Product means any methods, procedures, authorization keys, or other information required to install and execute modified versions of a covered work in that User Product from a modified version of its Corresponding Source. The information must suffice to ensure that the continued functioning of the modified object code is in no case prevented or interfered with solely because modification has been made.

If you convey an object code work under this section in, or with, or specifically for use in, a User Product, and the conveying occurs as part of a transaction in which the right of possession and use of the User Product is transferred to the recipient in perpetuity or for a fixed term (regardless of how the transaction is characterized),

the Corresponding Source conveyed under this section must be accompanied by the Installation Information. But this requirement does not apply if neither you nor any third party retains the ability to install modified object code on the User Product (for example, the work has been installed in ROM).

The requirement to provide Installation Information does not include a requirement to continue to provide support service, warranty, or updates for a work that has been modified or installed by the recipient, or for the User Product in which it has been modified or installed. Access to a network may be denied when the modification itself materially and adversely affects the operation of the network or violates the rules and protocols for communication across the network.

Corresponding Source conveyed, and Installation Information provided, in accord with this section must be in a format that is publicly documented (and with an implementation available to the public in source code form), and must require no special password or key for unpacking, reading or copying.

7. Additional Terms.

"Additional permissions" are terms that supplement the terms of this License by making exceptions from one or more of its conditions. Additional permissions that are applicable to the entire Program shall be treated as though they were included in this License, to the extent that they are valid under applicable law. If additional permissions apply only to part of the Program, that part may be used separately under those permissions, but the entire Program remains governed by this License without regard to the additional permissions.

When you convey a copy of a covered work, you may at your option remove any additional permissions from that copy, or from any part of it. (Additional permissions may be written to require their own removal in certain cases when you modify the work.) You may place additional permissions on material, added by you to a covered work, for which you have or can give appropriate copyright permission.

Notwithstanding any other provision of this License, for material you add to a covered work, you may (if authorized by the copyright holders of that material) supplement the terms of this License with terms:

a. Disclaiming warranty or limiting liability differently from the terms of sections 15 and 16 of this License; or

b. Requiring preservation of specified reasonable legal notices or author attributions in that material or in the Appropriate Legal Notices displayed by works containing it; or

c. Prohibiting misrepresentation of the origin of that material, or requiring that modified versions of such material be marked in reasonable ways as different from the original version; or

d. Limiting the use for publicity purposes of names of licensors or authors of the material; or

e. Declining to grant rights under trademark law for use of some trade names, trademarks, or service marks; or

f. Requiring indemnification of licensors and authors of that material by anyone who conveys the material (or modified versions of it) with contractual assumptions of liability to the recipient, for any liability that these contractual assumptions directly impose on those licensors and authors.

All other non-permissive additional terms are considered "further restrictions" within the meaning of section 10. If the Program as you received it, or any part of it, contains a notice stating that it is governed by this License along with a term that is a further restriction, you may remove that term. If a license document contains a further restriction but permits relicensing or conveying under this License, you may add to a covered work material governed by the terms of that license document, provided that the further restriction does not survive such relicensing or conveying.

If you add terms to a covered work in accord with this section, you must place, in the relevant source files, a statement of the additional terms that apply to those files, or a notice indicating where to find the applicable terms.

Additional terms, permissive or non-permissive, may be stated in the form of a separately written license, or stated as exceptions; the above requirements apply either way.

8. Termination.

You may not propagate or modify a covered work except as expressly provided under this License. Any attempt otherwise to propagate or modify it is void, and will automatically terminate your rights under this License (including any patent licenses granted under the third paragraph of section 11).

However, if you cease all violation of this License, then your license from a particular copyright holder is reinstated (a) provisionally, unless and until the copyright holder explicitly and finally terminates your license, and (b) permanently, if the copyright holder fails to notify you of the violation by some reasonable means prior to 60 days after the cessation.

Moreover, your license from a particular copyright holder is reinstated permanently if the copyright holder notifies you of the violation by some reasonable means, this is the first time you have received notice of violation of this License (for any work) from that copyright holder, and you cure the violation prior to 30 days after your receipt of the notice.

Termination of your rights under this section does not terminate the licenses of parties who have received copies or rights from you under this License. If your rights have been terminated and not permanently reinstated, you do not qualify to receive new licenses for the same material under section 10.

9. Acceptance Not Required for Having Copies.

You are not required to accept this License in order to receive or run a copy of the Program. Ancillary propagation of a covered work occurring solely as a consequence of using peer-to-peer transmission to receive a copy likewise does not require acceptance. However, nothing other than this License grants you permission to propagate or modify any covered work. These actions infringe copyright if you do not accept this License. Therefore, by modifying or propagating a covered work, you indicate your acceptance of this License to do so.

10. Automatic Licensing of Downstream Recipients.

Each time you convey a covered work, the recipient automatically receives a license from the original licensors, to run, modify and propagate that work, subject to this License. You are not responsible for enforcing compliance by third parties with this License.

An “entity transaction” is a transaction transferring control of an organization, or substantially all assets of one, or subdividing an organization, or merging organizations. If propagation of a covered work results from an entity transaction, each party to that transaction who receives a copy of the work also receives whatever licenses to the work the party's predecessor in interest had or could give under the previous paragraph, plus a right to possession of the Corresponding Source of the work from the predecessor in interest, if the predecessor has it or can get it with reasonable efforts.

You may not impose any further restrictions on the exercise of the rights granted or affirmed under this License. For example, you may not impose a license fee, royalty, or other charge for exercise of rights granted under this License, and you may not initiate litigation (including a cross-claim or counterclaim in a lawsuit) alleging that any patent claim is infringed by making, using, selling, offering for sale, or importing the Program or any portion of it.

11. Patents.

A “contributor” is a copyright holder who authorizes use under this License of the Program or a work on which the Program is based. The work thus licensed is called the contributor's “contributor version”.

A contributor's “essential patent claims” are all patent claims owned or controlled by the contributor, whether already acquired or hereafter acquired, that would be infringed by some manner, permitted by this License, of making, using, or selling its contributor version, but do not include claims that would be infringed only as a consequence of further modification of the contributor version. For purposes of this definition, “control” includes the right to grant patent sublicenses in a manner consistent with the requirements of this License.

Each contributor grants you a non-exclusive, worldwide, royalty-free patent license under the contributor's essential patent claims, to make, use, sell, offer for sale, import and otherwise run, modify and propagate the contents of its contributor version.

In the following three paragraphs, a “patent license” is any express agreement or commitment, however denominated, not to enforce a patent (such as an express permission to practice a patent or covenant not to sue for patent infringement). To “grant” such a patent license to a party means to make such an agreement or commitment not to enforce a patent against the party.

If you convey a covered work, knowingly relying on a patent license, and the Corresponding Source of the work is not available for anyone to copy, free of charge and under the terms of this License, through a publicly available network server or other readily accessible means, then you must either (1) cause the Corresponding Source to be so available, or (2) arrange to deprive yourself of the benefit of the patent license for this particular work, or (3) arrange, in a manner consistent with the requirements of this License, to extend the patent license to downstream recipients. “Knowingly relying” means you have actual knowledge that, but for the patent license, your conveying the covered work in a country, or your recipient's use of the covered work in a country, would infringe one or more identifiable patents in that country that you have reason to believe are valid.

If, pursuant to or in connection with a single transaction or arrangement, you convey, or propagate by procuring conveyance of, a covered work, and grant a patent license to some of the parties receiving the covered work authorizing them to use, propagate, modify or convey a specific copy of the covered work, then the patent license you grant is automatically extended to all recipients of the covered work and works based on it.

A patent license is “discriminatory” if it does not include within the scope of its coverage, prohibits the exercise of, or is conditioned on the non-exercise of one or more of the rights that are specifically granted under this License. You may not convey a covered work if you are a party to an arrangement with a third party that is in the business of distributing software, under which you make payment to the third party based on the extent of your activity of conveying the work, and under which the third party grants, to any of the parties who would receive the covered work from you, a discriminatory patent license (a) in connection with copies of the covered work conveyed by you (or copies made from those copies), or (b) primarily for and in connection with specific products or compilations that contain the covered work, unless you entered into that arrangement, or that patent license was granted, prior to 28 March 2007.

Nothing in this License shall be construed as excluding or limiting any implied license or other defenses to infringement that may otherwise be available to you under applicable patent law.

12. No Surrender of Others' Freedom.

If conditions are imposed on you (whether by court order, agreement or otherwise) that contradict the conditions of this License, they do not excuse you from the conditions of this License. If you cannot convey a covered work so as to satisfy simultaneously your obligations under this License and any other pertinent obligations, then as a consequence you may not convey it at all. For example, if you agree to terms that obligate you to collect a royalty for further conveying from those to whom you convey the Program, the only way you could satisfy both those terms and this License would be to refrain entirely from conveying the Program.

13. Use with the GNU Affero General Public License.

Notwithstanding any other provision of this License, you have permission to link or combine any covered work with a work licensed under version 3 of the GNU Affero General Public License into a single combined work, and to convey the resulting work. The terms of this License will continue to apply to the part which is the covered work, but the special requirements of the GNU Affero General Public License, section 13, concerning interaction through a network will apply to the combination as such.

14. Revised Versions of this License.

The Free Software Foundation may publish revised and/or new versions of the GNU General Public License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns.

Each version is given a distinguishing version number. If the Program specifies that a certain numbered version of the GNU General Public License “or any later version” applies to it, you have the option of following the terms and conditions either of that numbered version or of any later version published by the Free Software Foundation. If the Program does not specify a version number of the GNU General Public License, you may choose any version ever published by the Free Software Foundation.

If the Program specifies that a proxy can decide which future versions of the GNU General Public License can be used, that proxy's public statement of acceptance of a version permanently authorizes you to choose that version for the Program.

Later license versions may give you additional or different permissions. However, no additional obligations are imposed on any author or copyright holder as a result of your choosing to follow a later version.

15. Disclaimer of Warranty.

THERE IS NO WARRANTY FOR THE PROGRAM, TO THE EXTENT PERMITTED BY APPLICABLE LAW. EXCEPT WHEN OTHERWISE STATED IN WRITING THE COPYRIGHT HOLDERS AND/OR OTHER PARTIES PROVIDE THE PROGRAM “AS IS” WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE OF THE PROGRAM IS WITH YOU. SHOULD THE PROGRAM PROVE DEFECTIVE, YOU ASSUME THE COST OF ALL NECESSARY SERVICING, REPAIR OR CORRECTION.

16. Limitation of Liability.

IN NO EVENT UNLESS REQUIRED BY APPLICABLE LAW OR AGREED TO IN WRITING WILL ANY COPYRIGHT HOLDER, OR ANY OTHER PARTY WHO MODIFIES AND/OR CONVEYS THE PROGRAM AS PERMITTED ABOVE, BE LIABLE TO YOU FOR DAMAGES, INCLUDING ANY GENERAL, SPECIAL, INCIDENTAL OR CONSEQUENTIAL DAMAGES ARISING OUT OF THE USE OR INABILITY TO USE THE PROGRAM (INCLUDING BUT NOT LIMITED TO LOSS OF DATA OR DATA BEING RENDERED INACCURATE OR LOSSES SUSTAINED BY YOU OR THIRD PARTIES OR A FAILURE OF THE PROGRAM TO OPERATE WITH ANY OTHER PROGRAMS), EVEN IF SUCH HOLDER OR OTHER PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

17. Interpretation of Sections 15 and 16.

If the disclaimer of warranty and limitation of liability provided above cannot be given local legal effect according to their terms, reviewing courts shall apply local law that most closely approximates an absolute waiver of all civil liability in connection with the Program, unless a warranty or assumption of liability accompanies a copy of the Program in return for a fee.

END OF TERMS AND CONDITIONS

How to Apply These Terms to Your New Programs

If you develop a new program, and you want it to be of the greatest possible use to the public, the best way to achieve this is to make it free software which everyone can redistribute and change under these terms.

To do so, attach the following notices to the program. It is safest to attach them to the start of each source file to most effectively state the exclusion of warranty; and each file should have at least the “copyright” line and a pointer to where the full notice is found.

```
one line to give the program's name and a brief idea of what it does.
Copyright (C) year name of author

This program is free software: you can redistribute it and/or modify
it under the terms of the GNU General Public License as published by
the Free Software Foundation, either version 3 of the License, or
(at your option) any later version.
```

```
This program is distributed in the hope that it will be useful,
but WITHOUT ANY WARRANTY; without even the implied warranty of
MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.  See the
GNU General Public License for more details.

You should have received a copy of the GNU General Public License
along with this program.  If not, see https://www.gnu.org/licenses/.
```

Also add information on how to contact you by electronic and paper mail.

If the program does terminal interaction, make it output a short notice like this when it starts in an interactive mode:

```
program Copyright (C) year name of author
This program comes with ABSOLUTELY NO WARRANTY; for details type ‘show w’.
This is free software, and you are welcome to redistribute it
under certain conditions; type ‘show c’ for details.
```

The hypothetical commands ‘`show w`’ and ‘`show c`’ should show the appropriate parts of the General Public License. Of course, your program's commands might be different; for a GUI interface, you would use an “about box”.

You should also get your employer (if you work as a programmer) or school, if any, to sign a “copyright disclaimer” for the program, if necessary. For more information on this, and how to apply and follow the GNU GPL, see https://www.gnu.org/licenses/.

The GNU General Public License does not permit incorporating your program into proprietary programs. If your program is a subroutine library, you may consider it more useful to permit linking proprietary applications with the library. If this is what you want to do, use the GNU Lesser General Public License instead of this License. But first, please read https://www.gnu.org/licenses/why-not-lgpl.html.

Colophon

About the author. Filippo Rusconi is a senior research scientist at the French national research council (*Centre national de la Recherche scientifique*, CNRS). Filippo has a background in biochemistry and organic chemistry and was trained during his Ph.D. as a bioanalytical chemist. He has extensive knowledge of analytical techniques involved in the study of biopolymers.

Filippo Rusconi is the author of a handbook about mass spectrometry for biochemists (French). The book was published by the French sci/tech publisher Lavoisier (https://e.lavoisier.fr/produit/23400/9782743017705/manuel-de-spectrometrie-de-masse-a-l-usage-des-biochimistes).

Colophon. The look of this book (PDF file) is the result of me having read many books from the O'Reilly publisher.

The typesetting of the book has been done on a Debian GNU/Linux computer using only Free Software. Use of the DocBook Authoring and Publishing Suite (DAPS (https://github.com/openSUSE/daps)) from SUSE was key in the process.

The layout adopted for this book is an adaptation of the SUSE stylesheets. I would like to thank Frank Sundermeyer <fsundermeyer@opensuse.org>, Stefan Knorr <sknorr@suse.de> and Tom Schraitle (see here (https://github.com/tomschr)) for being helpful with all my questions.

The main font used was EBGaramond (https://github.com/georgd/EB-Garamond) and the symbol/mathematical font was from the STIX project (https://www.stixfonts.org/) (font: STIX2Math).

The screen shots were taken with Spectacle, the screen capture program shipped along with my KDE (https://www.kde.org/) desktop environment and resampled using The GNU image manipulation program The Gimp (https://www.gimp.org/). Illustrations were done in Inkscape (https://inkscape.org/), a vectorial drawing software.