CP-SEEKER Documentation

User Guide NOT-SCI-004 version 003 for CP-Seeker software

TABLE OF CONTENTS

1. GENERAL CONSIDERATIONS

1.1. PURPOSE

CP-Seeker is an open-source software in *R*, developed with the *Shiny* framework and the *Chromium Portable* v16.0.3153.0 web browser (64-bit). It is dedicated to the **post-acquisition processing**, visualization and analysis of ion signals from **polychlorinated alkanes (PCAs)** and related chemical families within chromatography–high resolution mass spectrometry (HRMS) data sets. Those related chemical families are polychlorinated mono-, di- and triolefins (PCOs, PCdiOs, PCtriOs), polybrominated alkanes (PBAs), various putative phase I and phase II PCA metabolites, and polyhalogenated (mixed Cl and Br) alkanes (PXAs). *CP-Seeker* is freely available upon request at contact.cpseeker@oniris-nantes.fr, under the CC-BY 4.0 license. *CP-Seeker 2.0* is packed as a zip folder of about 820 MB. Once unzipped, it runs out of the box without need for computer skills or administrative rights, when installed in a folder in which the user has full read/write access (*e.g.*, C:\Users\<your_username>).

The present document aims at describing *CP-Seeker* workflow and functionalities.

1.2. CP-SEEKER TEAM

Contributor	Position and/or skills	Contribution
Ronan CARIOU	Researcher; Analytical chemist	Project initiator and leader
Sébastien HUTINET	Senior bioinformatician	Code lines of the original version 0.1 and upgrades up to v1.1
Yann GUITTON	Metabolomics platform manager	Advisor on the selection of adequate bioinformatics tools
Mélanie MADELIN	Bioinformatics internship (6M)	v1.0 upgrade
Akissi KOUAMÉ	Bioinformatics internship (2M)	v2.0 upgrade
Julien SAIN-VANNE	Senior bioinformatician	Upgrade code lines from v2.0

Non-exhaustive list of beta-testers:

Ronan CARIOU, Marie MÉZIÈRE, Cherine AMOURA, Thomas MCGRATH

1.3. CITING CP-SEEKER

Whenever citing *CP-Seeker*, please use the following reference that describes version 2.1.2:

- McGrath TJ, Saint-Vanne J, Hutinet S, Vetter W, Poma G, Fujii Y, Dodson R, Johnson-Restrepo B, Muenhor D, Dervilly G, Covaci A, Cariou R. Provisional title: Detection of bromochloro alkanes in indoor dust using a novel CP-Seeker data integration tool. JOURNAL, YEAR. IN PREP.

1.4. LIST OF PACKAGES AND APPLICATIONS EMBEDDED IN CP-SEEKER

bsplus 0.1.2

stringr 1.4.0

shiny 1.5.0

shinyWidgets 0.5.3

shinydashboard 0.7.1

fst 0.9.4

msConvert 3.0.23056-80d75d6 from **ProteoWizard**

MSnbase 2.12.0

blob 1.2.1

RSQLite 2.2.1

plotly 4.9.2.1

Jsonlite 1.7.1

openxlsx 4.2.2

xcms 3.8.2

pracma 2.3.3

dbscan 1.1-6

shinyFeedback 0.3.0

shinyFiles 0.8.0

DT 0.15

shinycssloaders 1.0.0

shinyjs 2.0.0

enviPat 2.4

1.5. EXAMPLES OF REFERENCES TAKING ADVANTAGE OF CP-SEEKER

With our research team

Previous (beta and v1) versions of *CP-Seeker* have been used in a few peer-reviewed articles.

- Amoura C, Larvor F, Marchand P, Le Bizec B, Cariou R, Bichon E. Quantification of chlorinated paraffins by chromatography coupled to high-resolution mass spectrometry - Part A: influence of gas chromatography and ionisation source parameters. *Submitted*.
- Amoura C, Larvor F, Marchand P, Le Bizec B, Cariou R, Bichon E. Quantification of chlorinated paraffins by chromatography coupled to high-resolution mass spectrometry - Part B: influence of liquid chromatography separation. *Submitted*.
- McGrath TJ, Poma G, Hutinet S, Fujii Y, Dodson R, Johnson-Restrepo B, Muenhor D, Dervilly G, Cariou R, Covaci A. An international investigation of chlorinated paraffin concentrations and homologue distributions in indoor dust. *Environmental Pollution, 2023, 333, 121994*. https://doi.org/10.1016/j.envpol.2023.121994.
- Mézière M, Marchand P, Hutinet S, Larvor F, Baéza E, Le Bizec B, Dervilly G, Cariou R. Accumulation of short-, medium-, and long-chains chlorinated paraffins in tissues of laying hens after dietary exposure. *Food Chemistry 2021, 351, 129289*. https://doi.org/10.3390/ijerph18020690
- Mézière M, Marchand P, Hutinet S, Larvor F, Baéza E, Le Bizec B, Dervilly G, Cariou R. Transfer of short-, medium-, and long-chain chlorinated paraffins to eggs of laying hens after dietary exposure. *Food Chemistry 2021, 343, 128491*. https://doi.org/10.1016/j.foodchem.2020.128491
- Mézière M, Krätschmer K, Pērkons I, Zacs D, Marchand P, Dervilly G, Le Bizec B, Schächtele A, Cariou R, Vetter W. Addressing main challenges regarding short- and medium-chain chlorinated paraffin analysis using GC/ECNI-MS and LC/ESI-MS methods. *Journal of the American Society for Mass Spectrometry 2020, 31, 1885–1895*. https://doi.org/10.1021/jasms.0c00155
- Mézière M, Cariou R, Larvor F, Bichon E, Guitton Y, Marchand P, Dervilly G, Le Bizec B. Optimized characterization of short-, medium, and long-chain chlorinated paraffins in liquid chromatography-high resolution mass spectrometry. *Journal of Chromatography A 2020, 1619, 460927*. https://doi.org/10.1016/j.chroma.2020.460927

1.6. CITED REFERENCES

- Darnerud PO, Bergman Å. Critical review on disposition of chlorinated paraffins in animals and humans. Environment International 2022, 163, 107195. https://doi.org/10.1016/j.envint.2022.107195.
- Tautenhahn R, Böttcher C, Neumann S. Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinformatics 2008, 9, 504*. https://doi.org/10.1186/1471-2105-9-504.

1.7. DOCUMENT VERSION

The present version of the document NOT-SCI-004 is dedicated to *CP-Seeker* application. It replaces the previous version. It is expected to evolve frequently according to LABERCA's internal quality management system.

Table 1 reports the successive modifications of both the application and the present document, its user guide. Each new version cancels and replaces the previous one.

Table 1: Record of the successive modifications in *CP-Seeker* application and its user guide.

CP-Seeker application version	Modifications	Entry into force	NOT-SCI-004 documentation version	Modifications	Entry into force
0.1	• Original version (PCAs only, ESI adducts only, HBCDD standards only)	October 22nd, 2019			
1.0	• Addition of families (COs) and some standards • Improvement of ergonomics ((dynamic matrices, grey cells when out of range)	September 6th, 2021			
1.1	• Addition of families (PBAs, metabolites and PXAs) and ENCI adducts.	May 18th, 2022			
2.0.0	• Addition of some standards • Improvement of ergonomics (PXAs, filters, *Excel* export)	June 22nd, 2023	001	• Original version	July 17th, 2023
2.1.0	• Sequence name limited to 50 characters • *msConvert Peak Picking* parameter set to *Vendor* rather than *CWT* • Specification of related standard for each retention time in the deconvolution parameter setting • Correction of the normalised area calculation that was formerly the sum of maximum abundances for each observed isotopomer • Correction of the weighted average deviation formula which was given the opposite sign • Addition of time point within the *error.log* file for deconvolution and export functions • Export functions moved to another place within the menu, so that export can be launched without visualising the data matrices • Export function working event of no standards deconvoluted • Correction of target analyte family names in the *Excel* export file • Minor typo and debug	August 30th, 2023	002	• Description of evolutions	August 30th, 2023
2.1.1	• Correction of the normalised area at the baseline for the standards • Finalisation of the CSV export function	September 9th, 2023			

2.1.2	• Debug on CSV export	September 22nd, 2023	003	• Description of evolutions since v2.1.0	September 22nd, 2023

2. GETTING STARTED

2.1. INSTALLATION

CP-Seeker 2.0 is packed as a zip folder of about 820 MB. All needed instructions are embedded in the folder. If you have any trouble installing *CP-Seeker*, please let us know at contact.cpseeker@oniris-nantes.fr. This software is available on Windows with a 64-bit operating system with at least 4 Go RAM.

To install *CP-Seeker*, unzip the folder to a location **in which the user has full read/write access**. Once unzipped in such a location, it runs out of the box without need for computer skills or administrative rights. We **strongly recommend** that you work within "**C:\Users\<your_username>**" and avoid "C:\Program files".

2.2. LAUNCHING

To launch *CP-Seeker*, click on the file *CP-Seeker.bat* at folder root. It is advised to install a shortcut to this file elsewhere for convenience.

The *CP-Seeker.bat* command file launches the interface application, which is a portable version of the *Chromium Portable* v16.0.3153.0 web browser (64-bit) to avoid browser compatibility problems.

3. PRESENTATION OF THE INTERFACE AND WORKFLOW OVERVIEW

3.1. INTERFACE ARRANGEMENT

The application arrangement is composed of three parts (**Figure 1**):

- Title bar (1): select your username, sequence and see associated information.
- Menu (2): access to the different tabs.
- Parameters (3): parameters and data visualization for the selected tab.

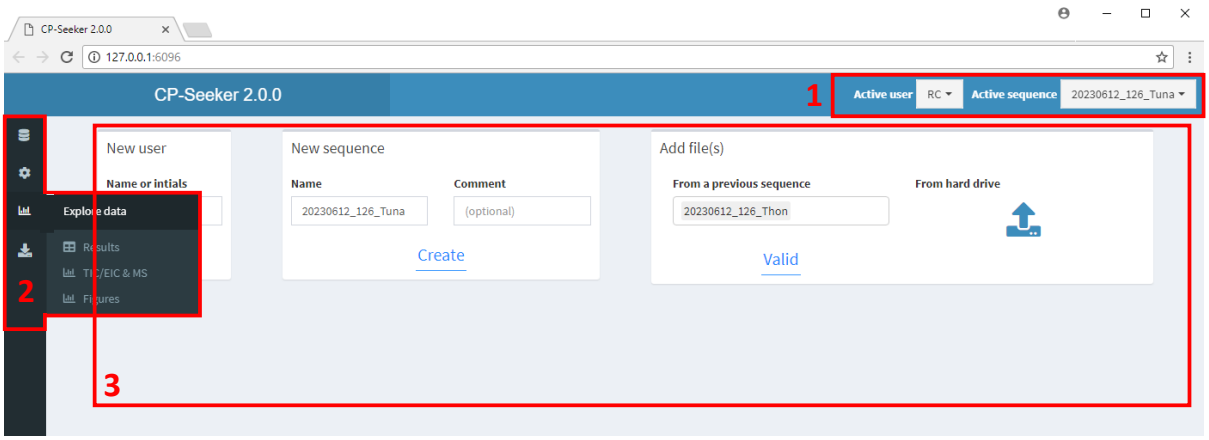


Figure 1: Arrangement of *CP-Seeker* interface.

3.2. TITLE BAR

The title bar (**Figure 2**) is composed of two picklists used to select the active user and the active sequence among previously defined users and sequences, respectively.

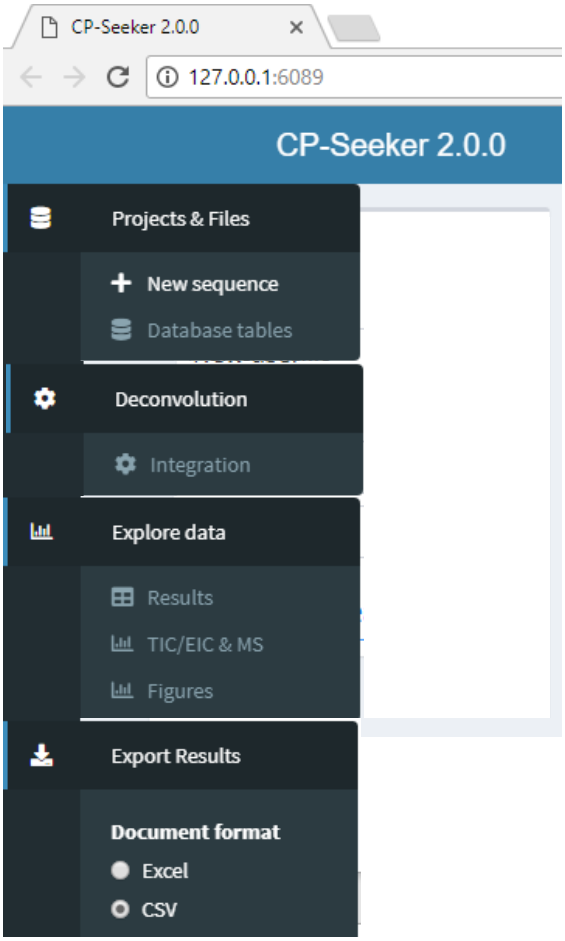
Figure 2: Title bar.

The button opens the user manual in a new Chrome tab. Functionality to be added.

3.3. MENU

The menu is composed of four parts (**Figure 3**):

- Sequences & Files: Create a new sequence and manage database tables.
- Deconvolution: Select parameters and launch integration of the signals.
- Explore data: Visualize deconvolution results in interactive tables, total and extracted ion chromatograms, and visualize the results in Figures.
- Export results: Export results in formatted *Excel* files or in *CSV* files.



3.4. PLOT INTERACTIVITY

For the ion chromatograms and mass spectra, zoom-in is possible using the left button of the mouse. Axes can be reset when double-clicking on the left button of the mouse.

With the mouse over the legend, all traces turn grey except for the one pointed out, in order to highlight it.

In addition, all plots in the application exhibit a toolbar on their upper right corner (**Figure 4**), which appears when moving the mouse over the plot. Icons are available, as hereafter described from the left to the right.

- Download plot as *png* file.
- Zoom-in.
- Autoscale.
- Reset legend. Restores the legend at its initial state. A click on an item on the legend hides it from the plot, a double click isolates it, other items being hidden.

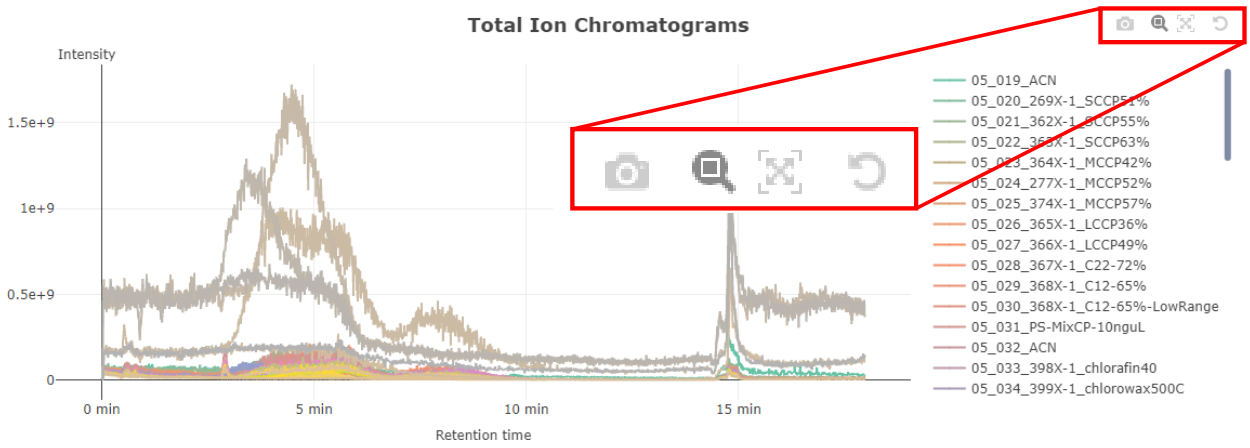


Figure 4: Interactive plot toolbars.

3.5. WORKFLOW OVERVIEW

An overview of the workflow is shown **Figure 5**.

Firstly, in the "Sequences & Files/New sequence" sub-tab, the user creates a sequence name. Then, in the same sub-tab, LC/HRMS or GC/HRMS data files are selected, either as proprietary format files or as universal format (*mzXML*). During import, proprietary format data files are automatically converted to *mzXML* files, and relevant parts are stored within the application database.

Secondly, in the "Deconvolution/Integration" sub-tab, the user selects appropriate parameters. Then, during the deconvolution step, the application identifies so-called "*Regions of Interest*" containing chromatographic peaks, integrates signal areas of those belonging to isotopic patterns of selected chemical families. Once the isotopic patterns are reconstituted, pattern scores and mass deviations are calculated.

Thirdly, results (normalised intensities, pattern scores and mass deviations) are displayed in interactive tables according to chemical formulas of the parent homologue groups.

Eventually, parameters and results can be exported in formatted *Excel* files or in a *CSV* file.

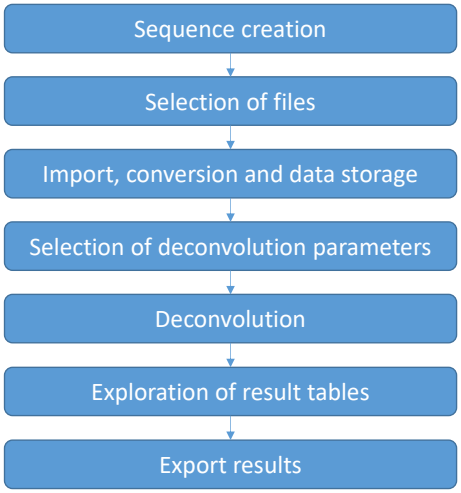


Figure 5: Workflow overview.

4. TAB SEQUENCES & FILES

4.1. SUB-TAB NEW SEQUENCE

4.1.1. Create a new user

A new user identification can be created using the dedicated sub-tab "*New Sequence*" (**Figure 6**). Do not forget to click the "*Create*" button. Once created, it is added to the *ad hoc* Active user list of the Title bar. Thus, do not forget to select it as Active user. The active user is recorded for each processing.

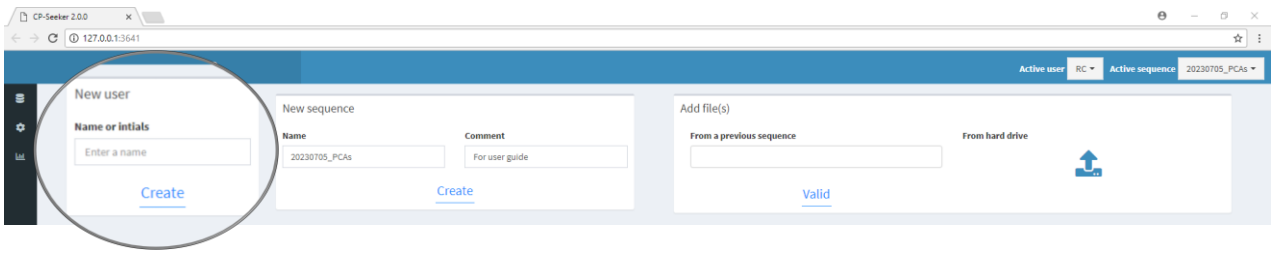


Figure 6: Creation of a new user.

4.1.2. Create a new sequence

A new sequence name can be created using the dedicated sub-tab "*New Sequence*" (**Figure 7**). No more than 50 characters are allowed. A comment can be provided as well. Do not forget to click the "*Create*" button. Once created, it is added to the *ad hoc* Active sequence list of the Title bar. Thus, do not forget to select it as Active sequence.

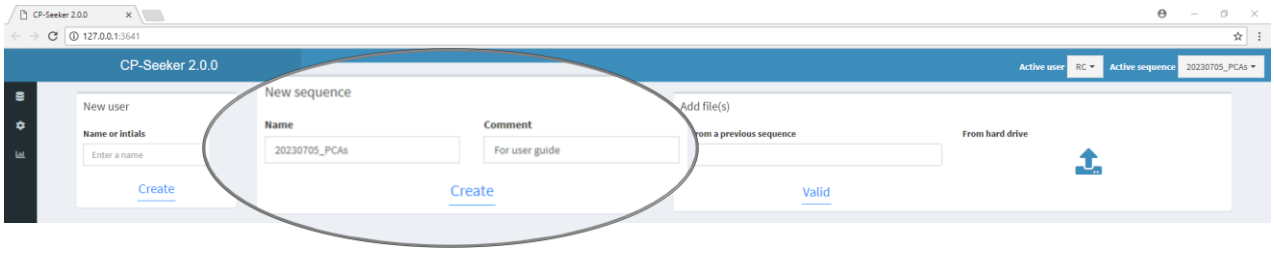


Figure 7: Creation of a new sequence.

4.1.3. Import files to a sequence

File(s) can be imported to the Active sequence either from the hard drive and/or amongst other sequences stored in the application database (**Figure 8**).

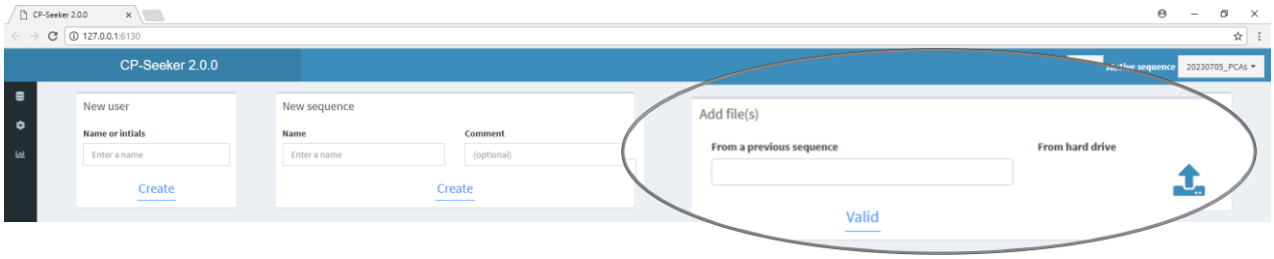


Figure 8: Import of HRMS files into a sequence.

4.1.3.1. From hard drive

Click on the button. It will display a modal allowing to navigate through the file system (**Figure 9**). Some upper buttons facilitate the navigation by changing the layout or sorting out files by their name, size and creation date. Major manufacturer and open formats are handled in the application (**.raw*, **.RAW*, **.d*, **.YEP*, **.BAF*, **.FID*, **.WIFF*, *.CDF, *.MGF, non-centroided **.mzXML*, **.mzML*, **.MGF*). Waters and Bruker raw directories can also be imported (they are considered as files by *CP-Seeker* if they have extension-like names). Once files are selected (coloured after a click), do not forget to click the "*Select*" button.

File names must be unique. Only data files containing data acquired in the **negative mode** are allowed.

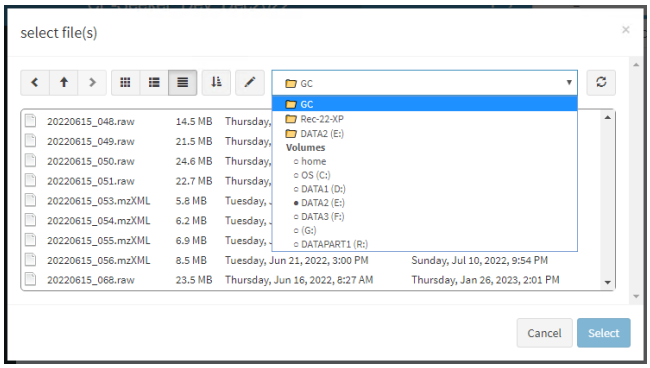


Figure 9: Selection of HRMS files to be imported into a sequence from a hard drive.

Once the files selected, the application displays a modal allowing associating data files with so-called *Labels* (sample ID) for convenience (**Figure 10**). The default *Label* is <filename>. *Labels* are used for selecting files when reviewing the data and results. *Labels* are limited to 30 characters (excluding special ones ":", "?", "/", "\", "*", "[", "]", """) and **must be unique** as well, in order to avoid conflicts in worksheets names of the export *Excel* file. Do not forget to click on the "*Valid*" button.

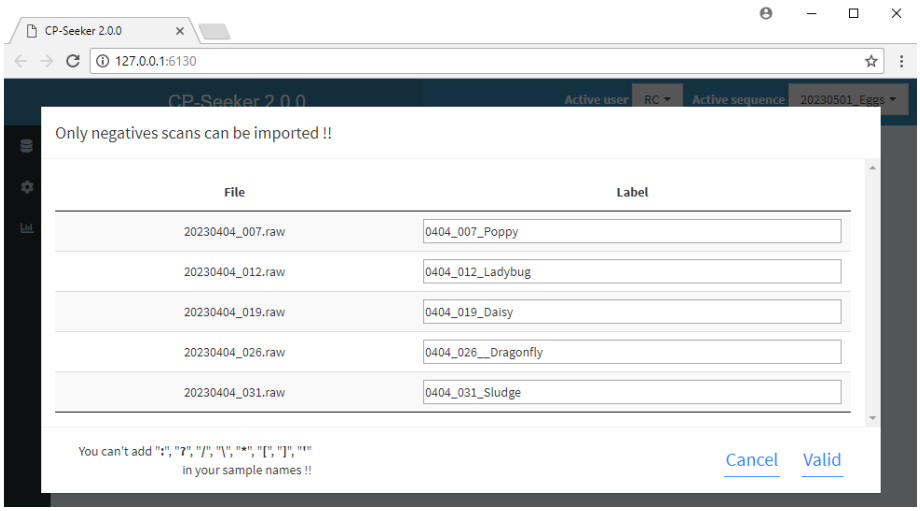


Figure 10: Labelling of imported HRMS data files.

Once validated, the files are automatically converted in an *mzXML* format in the negative polarity, using *msConvert* version 3.0.23056-80d75d6, an application of the software produced by *ProteoWizard*. **Some Microsoft files are mandatory to make it work. In case you miss them, please refer to the FAQ section**. In case of dual Pos/Neg files, CP-Seeker considers the trace corresponding to the negative polarity. Only one-dimensional data are considered (vendor msLevel=1). The peak Picking parameter was set to *Vendor* rather than *CWT*, so that **.raw* files from *e.g. Thermo* or *Agilent* are processed appropriately. If you intend to work with **.raw* file directories from *Waters*, it is recommended to switch to *CWT* directly in the code, because there are no *Vendor* setting for this manufacturer in *msConvert*.

The data are also automatically centroided, in order to "lighten" data. The user is also allowed to import already centroided file(s) in the most common format (**.mzXML*, **.mzML*, **.CDF*). These files are checked for centroided and readable data. Import of non-centroided *mzXML* files does not work. Why centroid MS data? Because a peak in profile mode contains too much information, that the software cannot use for the moment (and it will be too hard to work with). Thus, each peak will be displayed as discrete *m/z* with zero line widths (**Figure 11**). For that, *msConvert* includes most of the manufacturers' algorithms, except for Waters Corporation. In this case, *ProteoWizard* developed their own algorithm and, depending on the instrument, the script might fail. In this case, it is advised that the user converts such files independently and then imports converted files into *CP-Seeker*. Do not hesitate to contact us for tips.

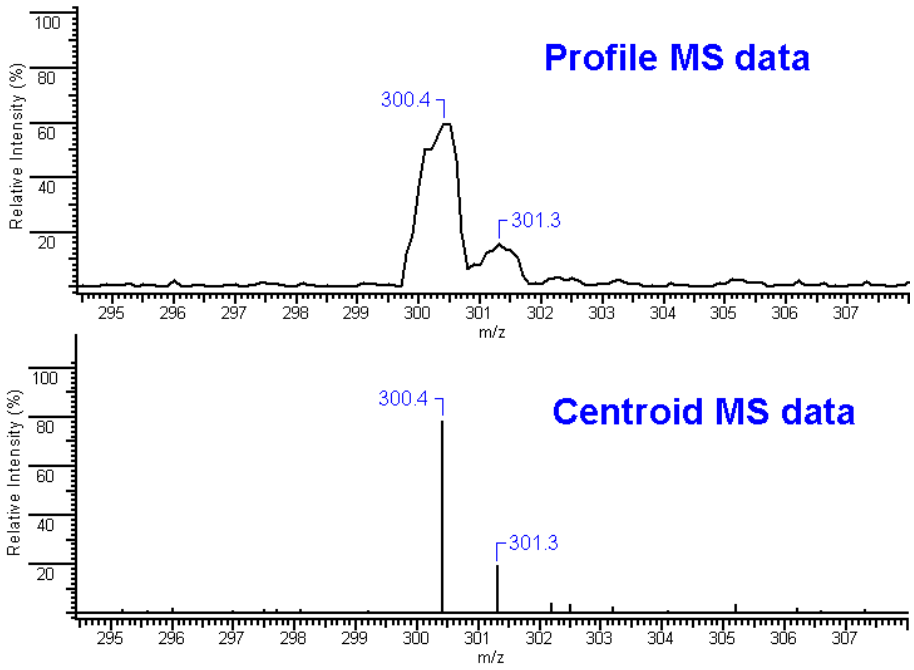


Figure 11: Illustration of profile *versus* centroid MS data.

4.1.3.2. From previous sequences

All data files imported within the application can be re-used in multiple sequences. Using the picklist showing files sorted by sequence (**Figure 12**), it is possible to select previously imported, converted and centroided HRMS data files. Do not forget to click on the "*Valid*" button. Then, the modal allowing to provide a *Label* (sample ID) different from the file name appears (see §4.1.3.1). Do not forget to click on the "*Valid*" button.

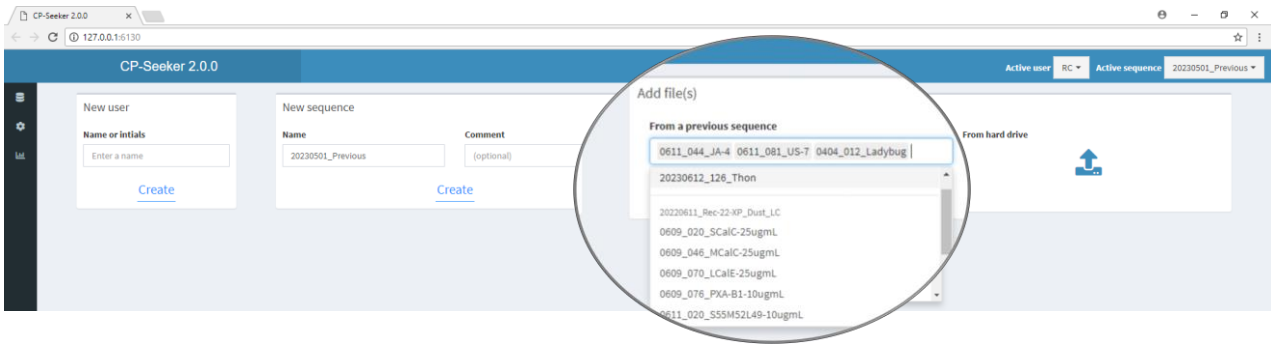


Figure 12: Selection of HRMS files to be imported into a sequence from previous sequences.

4.1.3.3. Success of import

Once files are imported from the hard drive, converted and stored in the database, a pop-up window specifies whether everything went well (**Figure 13**).

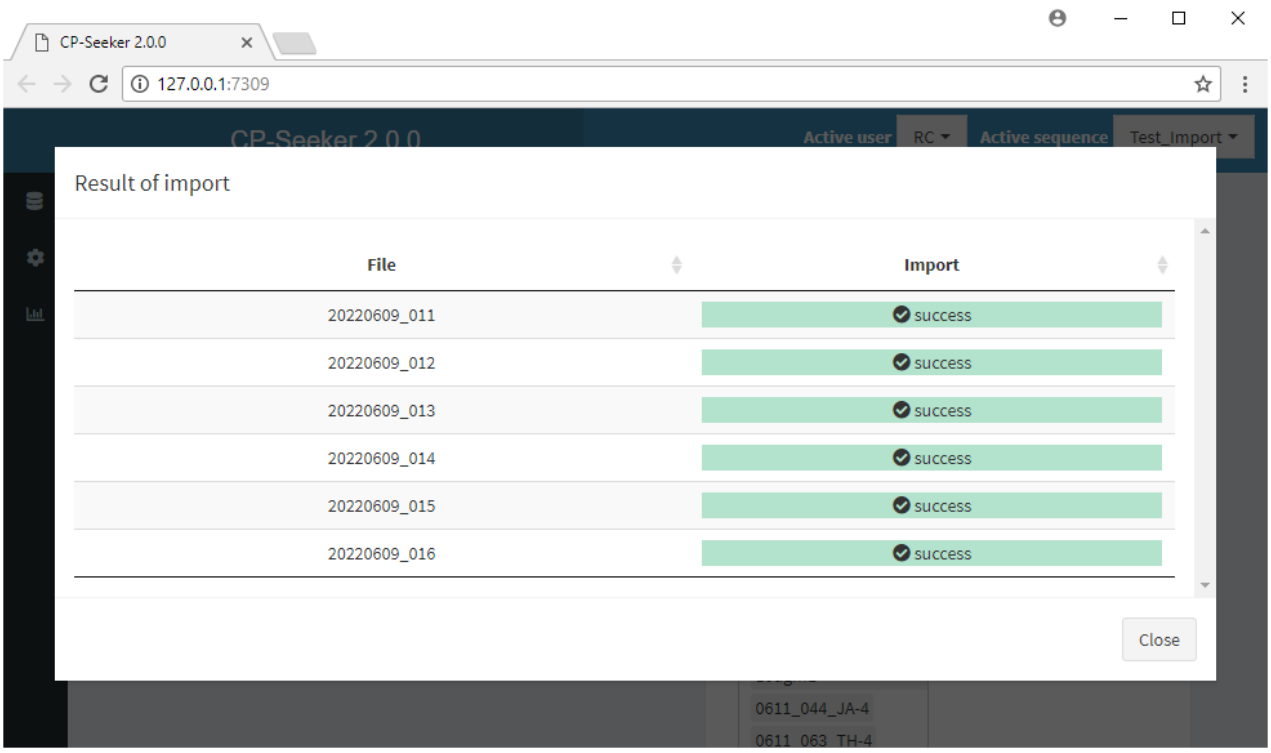


Figure 13: Import HRMS files from previous sequences.

4.2. SUB-TAB DATABASE TABLES

The "*Database tables*" sub-tab of the "*Sequences & Files*" tab (**Figure 14**) allows managing entries stored in the *CP-Seeker* database (*SQLite* format). The head area (1) allows to select one of the 2 accessible tables (*Sequence* or *Sample*), which are presented on the bottom area (2). It is possible to delete entries by selecting one or multiple rows in the table (3), and then by clicking on the button "*Delete entry(ies)*".

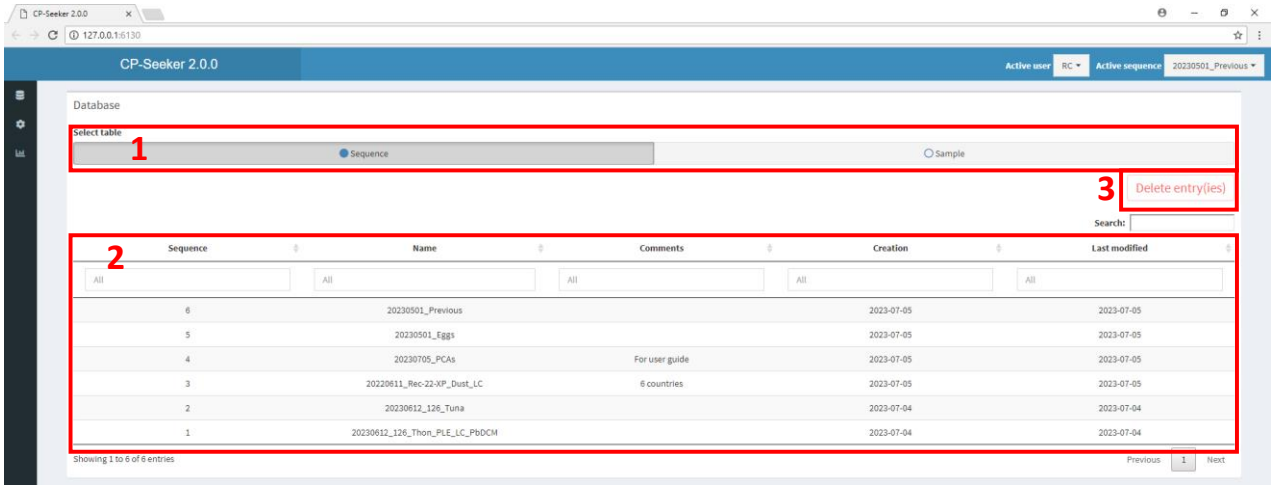


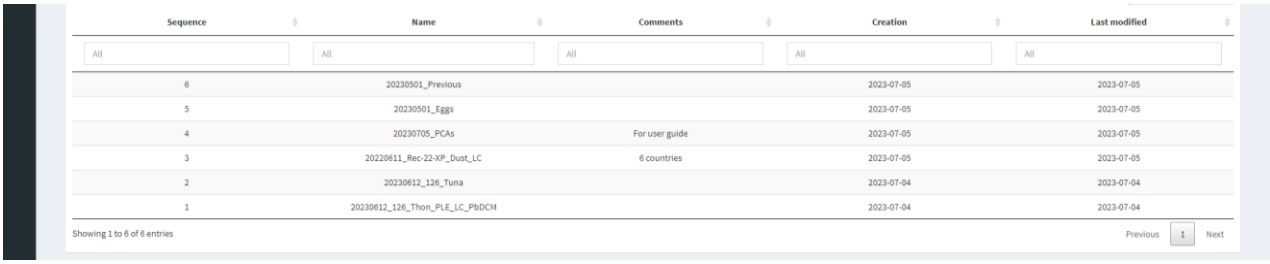
Figure 14: "*Database tables*" sub-tab.

The *SQLite* database (database.sqlite file) is available at the root directory of the installed software. It is possible (and advised) to **archive it elsewhere** once a sequence is over.

4.2.1. Sequence

Each line represents a sequence stored in the application database (**Figure 15**).

- Sequence increment.
- Name: name of the sequence (§ 4.1.2).
- Comments: information about the sequence (optional area, § 4.1.2).
- Creation: creation date of the sequence.
- Last modified: date of the last modification operated on the sequence.



Sequence	Name	Comments	Creation	Last modified
All	All	All	All	All
6	20230501_Previous		2023-07-05	2023-07-05
5	20230501_Eggs		2023-07-05	2023-07-05
4	20230705_PCAs	For user guide	2023-07-05	2023-07-05
3	20220611_Rec-22-XP_Dust_LC	6 countries	2023-07-05	2023-07-05
2	20230612_126_Tuna		2023-07-04	2023-07-04
1	20230612_126_Thon_PLE_LC_PbDCM		2023-07-04	2023-07-04

Showing 1 to 6 of 6 entries Previous 1 Next

Figure 15: Database "*Sequence*" table.

4.2.2. Sample

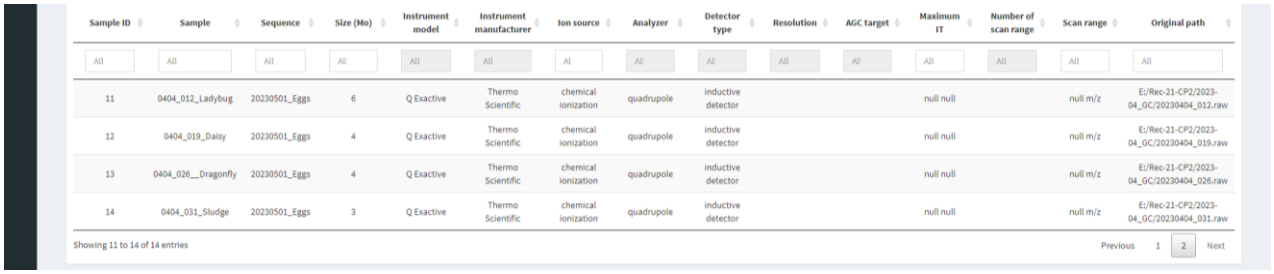
Each line represents a file imported in a sequence (**Figure 16**). The same file can be present in more than one row if it associated to more than one sequence.

- Sample file increment
- Sample: *Label* of the file imported (**Figure 10**).
- Sequence: name of the sequence associated to the file (§ 4.1.2).
- Size (Mo): actual size in Mo of the compressed & stored *mzXML* file in the database.

In addition, a range of metadata can be retrieved from some Thermo Scientific raw files using the *ThermoRawMetaDump* tool.

- Instrument model.
- Instrument manufacturer.
- Ion source.
- Analyser.
- Detector type.
- Resolution.
- AGC target.
- Maximum IT.
- Number of scan range.
- Scan range.

Lastly, the original path and mane of the imported file is given.



Sample ID	Sample	Sequence	Size (Mo)	Instrument model	Instrument manufacturer	Ion source	Analyzer	Detector type	Resolution	AGC target	Maximum IT	Number of scan range	Scan range	Original path
All	All	All	All	All	All	Al	All	All	All	All	All	All	All	All
11	0404_012_Ladybug	20230501_Eggs	6	Q Exactive	Thermo Scientific	chemical ionization	quadrupole	inductive detector			null null		null m/z	E:/Rec-21-CP2/2023-04_GC/20230404_012.raw
12	0404_019_Daisy	20230501_Eggs	4	Q Exactive	Thermo Scientific	chemical ionization	quadrupole	inductive detector			null null		null m/z	E:/Rec-21-CP2/2023-04_GC/20230404_019.raw
13	0404_026__Dragonfly	20230501_Eggs	4	Q Exactive	Thermo Scientific	chemical ionization	quadrupole	inductive detector			null null		null m/z	E:/Rec-21-CP2/2023-04_GC/20230404_026.raw
14	0404_031_Sludge	20230501_Eggs	3	Q Exactive	Thermo Scientific	chemical ionization	quadrupole	inductive detector			null null		null m/z	E:/Rec-21-CP2/2023-04_GC/20230404_031.raw

Showing 11 to 14 of 14 entries Previous 1 2 Next

Figure 16: Database "*Sample*" table.

5. TAB DECONVOLUTION

5.1. SUB-TAB INTEGRATION

The "*Integration*" sub-tab of the "*Deconvolution*" tab (**Figure 17**) gathers the parameters for the pattern reconstitution scripts of the workflow in the left area (**1**), which are sub-divided as "General", "Target analyte" and "Standard" parameters.

The second area (**2**) shows overlapped Total Ion Currents (TICs) of all samples in the sequence. In order to minimize time to display the chromatograms, data are smoothed to a maximum of 120 points per minute. Still, depending on the number of files, **it takes a few dozens of seconds to be displayed**, so be patient. A click on the TICs displays the mass spectra (**3**) corresponding to the selected retention time (t_R) coordinate. Report to §3.4 for plot interactivity tips.

Once the parameters are defined, do not forget to click on the "*Launch deconvolution process*" button. The selected integration steps will then be processed at once, including the pattern reconstitution for each selected chemical family (and standard) and for each selected adduct. This processing might take a while (**a few dozens of minutes or hours**), so be patient. It is advised that you start with a limited number of files before going for larger sequences, in order to apprehend the speed of the process. The screen refreshing might not be working at the end of the process, so do not hesitate to move the mouse across the screen window to make it happen. Once finished, the sub-tab "*Results*" is automatically shown (§ 5.2).

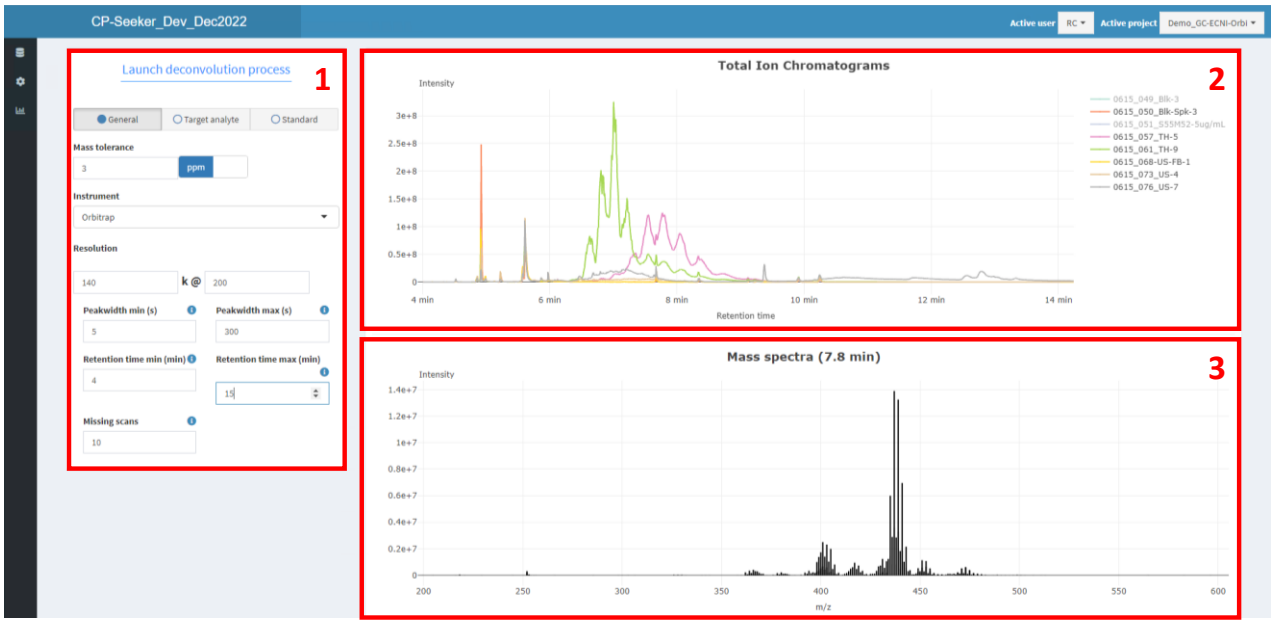


Figure 17: "*Integration*" sub-tab.

5.1.1. General parameters

The following general parameters are applied to the generation of extracted ion chromatograms (EICs), to their integration and to the pattern reconstruction (**Figure 18**):

- Mass tolerance (±) set either in ppm or in mDa.
- Instrument: Selection of the HRMS instrument. This information is useful to compute theoretical isotopic profile. For available ToF instruments, cubic interpolations (12 points) were done using the experimental data available from enviPat (**Figure 19**). In case on an Orbitrap mass analyser, the nominal resolution and the reference *m/z* for his nominal resolution are considered for calculating theoretical resolution at any *m/z*.
- Resolution: Proposed only if Orbitrap is selected as mass analyser.
- Peakwidth min and max: indicative range of peak elution times (in seconds).
- Retention time min and max: retention time range to be considered for the EICs.
- Missing scans: maximum number of empty scan between two ROIs to merge them (see deconvolution description §5.1.4 for further details).

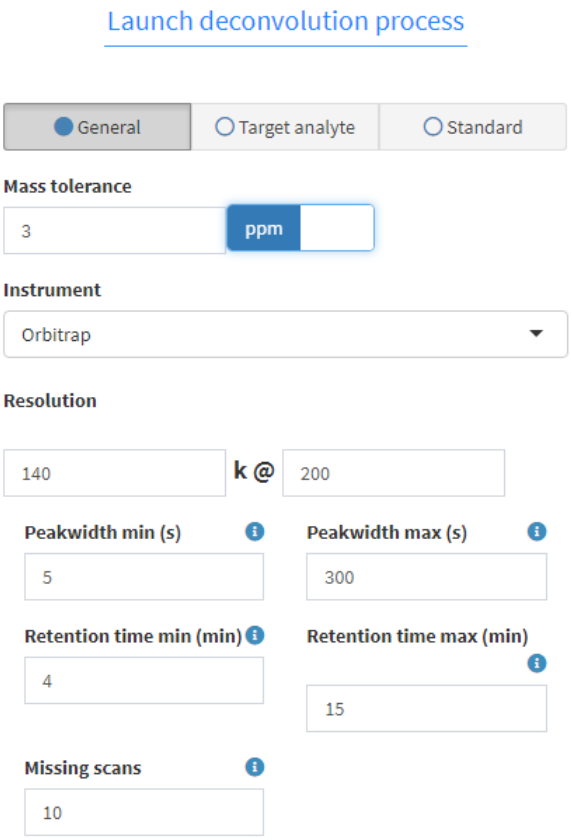


Figure 18: General parameters tab of deconvolution process.

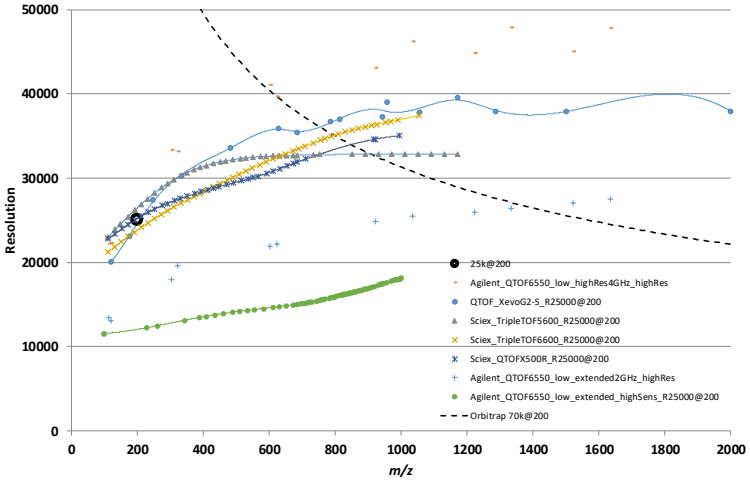


Figure 19: Modelling of resolution according to *m*/*z* for a selection of ToF mass analysers and theoretical resolution reached by an Orbitrap working at a resolution of 70,000 at *m*/*z* 200.

5.1.2. Target analyte parameters

Two picklists are available in the Target analyte parameters tab, for the selection of targeted chemical families and for the selection of ion adduct types, (**Figure 20**). Adding new possibilities to these lists will require modifications in the database structure (contact us).

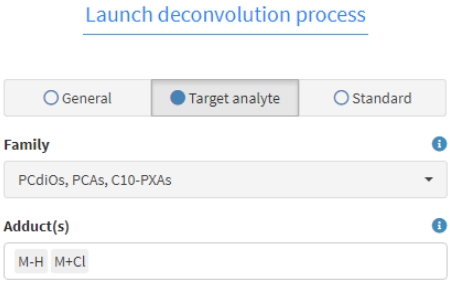


Figure 20: Target analyte parameters tab of deconvolution process.

Various chemical families can be selected (**Table 2**), among them:

- Polybrominated alkanes (PBAs);
- Polychlorinated mono-, di- and triolefins (PCOs, PCdiOs, PCtriOs);
- Polychlorinated alkanes (PCAs);
- Polychlorinated polybrominated alkanes (mixed Cl and Br alkanes; PXAs);
- Putative phase I and phase II PCA metabolites, considered according to Darnerud and Bergman (2022).

Within *CP-Seeker*, we set the range of **n, the number of carbon atoms of the related alkane backbone, to [6, 36]**, based on considerations found in the literature. We also set the **minimum and maximum numbers of halogen atoms $x+y$ to 3 and 30**, respectively. Last, considering that is it unlikely that a carbon atom bears more than one halogen atom, we considered **$x+y \leq n+3$**.

Table 2: List of chemical families available in *CP-Seeker*.

Type	Family	Description		Chemical formula
Polybrominated paraffins	PBAs	Polybrominated alkanes		$C_nH_{2n+2-y}Br_y$
Polybrominated olefins	PCOs	Polychlorinated monoolefins		$C_nH_{2n-x}Cl_x$
	PCdiOs	Polychlorinated diolefins		$C_nH_{2n-2-x}Cl_x$
	PCtriOs	Polychlorinated triolefins		$C_nH_{2n-4-x}Cl_x$
Polychlorinated paraffins	PCAs	Polychlorinated alkanes		$C_nH_{2n+2-x}Cl_x$
Mixed paraffins	C6-PXAs	Polychlorinated polybrominated alkanes	C_6-chain length	$C_6H_{14-x-y}Cl_xBr_y$
	C7-PXAs		C_7-chain length	$C_7H_{16-x-y}Cl_xBr_y$
	C8-PXAs		C_8-chain length	$C_8H_{18-x-y}Cl_xBr_y$
	C9-PXAs		C_9-chain length	$C_9H_{20-x-y}Cl_xBr_y$
	C10-PXAs		C_{10}-chain length	$C_{10}H_{22-x-y}Cl_xBr_y$
	C11-PXAs		C_{11}-chain length	$C_{11}H_{24-x-y}Cl_xBr_y$
	C12-PXAs		C_{12}-chain length	$C_{12}H_{26-x-y}Cl_xBr_y$
	C13-PXAs		C_{13}-chain length	$C_{13}H_{28-x-y}Cl_xBr_y$
	C14-PXAs		C_{14}-chain length	$C_{14}H_{30-x-y}Cl_xBr_y$
	C15-PXAs		C_{15}-chain length	$C_{15}H_{32-x-y}Cl_xBr_y$
	C16-PXAs		C_{16}-chain length	$C_{16}H_{34-x-y}Cl_xBr_y$
	C17-PXAs		C_{17}-chain length	$C_{17}H_{36-x-y}Cl_xBr_y$
	C18-PXAs		C_{18}-chain length	$C_{18}H_{38-x-y}Cl_xBr_y$
	C19-PXAs		C_{19}-chain length	$C_{19}H_{40-x-y}Cl_xBr_y$
	C20-PXAs		C_{20}-chain length	$C_{20}H_{42-x-y}Cl_xBr_y$
	C21-PXAs		C_{21}-chain length	$C_{21}H_{44-x-y}Cl_xBr_y$
	C22-PXAs		C_{22}-chain length	$C_{22}H_{46-x-y}Cl_xBr_y$
	C23-PXAs		C_{23}-chain length	$C_{23}H_{48-x-y}Cl_xBr_y$
	C24-PXAs		C_{24}-chain length	$C_{24}H_{50-x-y}Cl_xBr_y$
	C25-PXAs		C_{25}-chain length	$C_{25}H_{52-x-y}Cl_xBr_y$
	C26-PXAs		C_{26}-chain length	$C_{26}H_{54-x-y}Cl_xBr_y$
	C27-PXAs		C_{27}-chain length	$C_{27}H_{56-x-y}Cl_xBr_y$
	C28-PXAs		C_{28}-chain length	$C_{28}H_{58-x-y}Cl_xBr_y$
	C29-PXAs		C_{29}-chain length	$C_{29}H_{60-x-y}Cl_xBr_y$
	C30-PXAs		C_{30}-chain length	$C_{30}H_{62-x-y}Cl_xBr_y$
	C31-PXAs		C_{31}-chain length	$C_{31}H_{64-x-y}Cl_xBr_y$
	C32-PXAs		C_{32}-chain length	$C_{32}H_{66-x-y}Cl_xBr_y$
	C33-PXAs		C_{33}-chain length	$C_{33}H_{68-x-y}Cl_xBr_y$
	C34-PXAs		C_{34}-chain length	$C_{34}H_{70-x-y}Cl_xBr_y$
	C35-PXAs		C_{35}-chain length	$C_{35}H_{72-x-y}Cl_xBr_y$
	C36-PXAs		C_{36}-chain length	$C_{36}H_{74-x-y}Cl_xBr_y$
Phase I metabolites	OH-PCAs	Hydroxypolychlorinated alkanes		$C_nH_{2n+2-x}Cl_xO$
	COOH-PCAs	Carboxypolychlorinated alkanes		$C_nH_{2n-x}Cl_xO_2$
	oxo-PCAs	Oxopolychlorinated alkanes		$C_nH_{2n-x}Cl_xO$
Phase II metabolites	GSH-OH-PCAs	Glutathione S-conjugates of polychlorinated alkanes		$C_{n+10}H_{2n+17-x}Cl_xN_3O_7S$
	SCys-OH-PCAs	Cysteine S-conjugates of polychlorinated alkanes		$C_{n+3}H_{2n+7-x}Cl_xNO_3S$
	Mercapturic-OH-PCAs	N-acetylcysteine S-conjugates of polychlorinated alkanes (mercapturic acids)		$C_{n+3}H_{2n+9-x}Cl_xNO_4S$

Ion adducts can be selected among lists related to ECNI or ESI/APCI ionisation processes (**Table 3**). **For ECNI, be aware that a given ion adduct can arise indistinctly from the loss of a Cl or a Br atom**. Thus, the ion adduct type would rather be $[M - X]^-$ or $[M - HX]^-$, and it appears impractical to determine the molecular ion formula. In addition, **series from $[M - H]^-$ pseudo-molecular ions of PCAs and $[M + Cl]^-$ adduct ion of PCOs exhibit the same ion formulas** ($[C_nH_{2n+1-x}Cl_x]^-$ and $[C_nH_{2n+2-z}Cl_{z+1}]^-$, respectively, when z = x - 1), so that a confusion is possible between the two families with a shift in the homologue pattern. It is advised to favour only one ion type in the ionisation process to avoid confusion, or at least to check all possible ions before concluding on the correct annotation of the signals. Similarly, series from $[M - H]^-$ pseudo-molecular ions of PCOs and $[M + Cl]^-$ adduct ion of PCdiOs exhibit the same ion formulas; and series from $[M - H]^-$ pseudo-molecular ions of PCdiOs and $[M + Cl]^-$ adduct ion of PCtriOs exhibit the same ion formulas; and series from $[M - H]^-$ pseudo-molecular ions of OH-PCAs and $[M + Cl]^-$ adduct ion of oxo-PCAs exhibit the same ion formulas.

Table 3: List of ion adducts available in *CP-Seeker*.

Ionisation source	Coupled chromatography	Adduct ion
Electron capture ionization (ECNI)	Gas chromatography (GC)	$[M - Br]^-$
		$[M - Cl]^-$
		$[M - HBr]^-$
		$[M - HCl]^-$
Electrospray or Atmospheric Pressure Chemical Ionization (ESI or APCI)	Liquid chromatography (LC)	$[M + Hac - H]^-$ (acetate adduct)
		$[M - H]^-$
		$[M + Cl]^-$
		$[M + Br]^-$

5.1.3. Standard parameters

CP-Seeker offers the possibility to integrate internal and recovery standards as well (**Figure 21**). When this option is activated by the user, it is possible to select standards among a picklist (**Table 4**). This list is not exhaustive of what has been used in the literature. Available ion adducts are the same as those proposed for the target chemical families (**Table 3**), except that [M – H]- could also be $[M - D]^-$ (or $[M - {}^2H]^-$) in the case of $^2H_{18}$-HBCDD. For each selected standards, a retention time must be provided (±2 min), so that the application maximises chances to identify and integrate it properly.

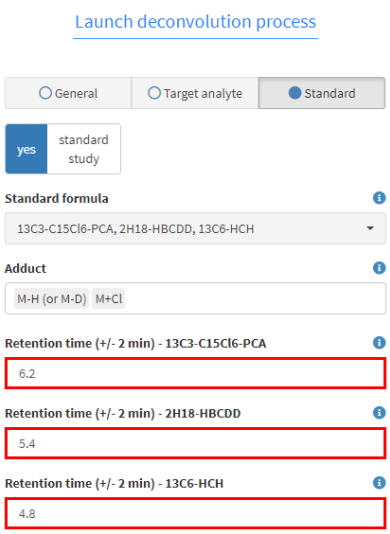


Figure 21: Standard parameters tab of deconvolution process.

Table 4: List of standards available in *CP-Seeker*.

Standard name	Chemical formula	Comment
$^{13}C_3$-$C_{11}Cl_6$-PCA	$^{13}C_3C_8H_{18}Cl_6$	Convenient for the 3rd generation Chloffin standards (Chiron)
$^{13}C_3$-$C_{12}Cl_6$-PCA	$^{13}C_3C_9H_{20}Cl_6$	
$^{13}C_3$-$C_{13}Cl_6$-PCA	$^{13}C3C_{10}H_{22}Cl_6$	
$^{13}C_3$-$C_{14}Cl_6$-PCA	$^{13}C3C_{11}H_{24}Cl_6$	
$^{13}C_3$-$C_{15}Cl_6$-PCA	$^{13}C3C_{12}H_{26}Cl_6$	
$^{13}C_3$-$C_{16}Cl_6$-PCA	$^{13}C3C_{13}H_{28}Cl_6$	
$^{13}C_3$-$C_{21}Cl_8$-PCA	$^{13}C_3C_{18}H_{36}Cl_8$	
$^{13}C_{10}$-$C_{10}Cl_6$-PCA	$^{13}C_{10}H_{16}Cl_6$	
$^{2}H_{18}$-HBCDD	$C_{12}{}^{2}H_{18}Br_6$	Labelled hexabromocyclododecane
$^{13}C_{12}$-HBCDD	$^{13}C_{12}H_{18}Br_6$	
$^{13}C_6$-HCH	$^{13}C_6H_6Cl_6$	Hexachlorocyclohexane
$^{13}C_{12}$-pentaCB	$^{13}C_{12}H_5Cl_5$	Pentachlorinated biphenyl
$^{13}C_6$-HCB	$^{13}C_6Cl_6$	Hexachlorobenzene
$^{13}C_8$-Mirex	$^{13}C_8C_2Cl_{12}$	
$^{13}C_{10}$-Chlordane	$^{13}C_{10}H_6Cl_8$	
$^{13}C_{10}$-Dechlorane_Plus	$^{13}C_{10}C_8H_{12}Cl_{12}$	
$^{13}C_{10}$-Dechlorane_602	$^{13}C_{10}C_4H_4Cl_{12}O_1$	
$^{13}C_{12}$-Br_6-PBDE	$^{13}C_{12}H_4Br_6$	Hexabromodiphenyl ether
MeO-Br_6-PBDE	$C_{13}H_6Br_6O_2$	Methoxylated polybrominodiphenyl ether
MeO-Br_5-PBDE	$C_{13}H_7Br_5O_2$	

5.1.4. Deconvolution process

Once the parameters are set, click on the "*Launch deconvolution process*" button. The selected integration steps will then be processed at once, as described below. A double progression bar allows to monitor which sample is processed and where do we stand within this sample. This processing might take a while (**a few dozens of minutes or hours**), so be patient. It is advised that you start with a limited number of files before going for larger sequences, in order to apprehend the speed of the process. For each chemical family and adduct type and homologue group, the resulting data include four pieces of information:

- The normalised (cumulative) area of the considered homologue group;
- A pattern score between the observed and theoretical isotopic patterns;
- The weighted average deviation between paired observed and theoretical isotopomers;
- Whether the theoretical isotopic pattern fully falls inside or outside the *m/z* range of the data, or straddles the *m/z* range borders. Straddling isotopic patterns for which the base peak A and/or the A-2 or A+2 isotopic peak fall outside the *m/z* range are considered outside. That way, straddling isotopic patterns contain at least [A and A-2] or [A and A+2] isotopomers, so that pattern scores can reach high levels.

Firstly, the application determines all the combinations between the requested chemical families the adduct ion formulas. Then, the application compiles all the centroided theoretical isotopic patterns using *enviPat*, based on the modelled or theoretical resolution set by the user. All and only the peaks above 1% of the base peak abundance are kept, and sorted in the decreasing order of abundance.

Secondly, a double loop considers each sample file of the sequence stored in the database and each centroided theoretical isotopic pattern for a search in the data, if not outside the *m/z* range (see fourth output piece of information above). For a given sample file and homologue group, the EIC corresponding to the *m/z*

of base peak isotopomers is loaded (±ppm/mDa tolerance set by the user; within the retention time range set by the user). Then, the application searches for so-called "*Regions Of Interest*" (ROIs), thanks to the *xcms* package. ROIs correspond to a *m/z* detected along several scans, with increased intensity and decreased *m/z* variation (**Figure 22**). It is the sign of a chromatographic peak. Gaussian fit is not applied, so that chromatographic humps that are typical from PCAs can be considered. It is possible that several ROIs are found within the same EIC. In that case, those ROIs that differ from less than the number of missing scans set by the user are merged. That way, noisy and minor isotopomers can be completed. If several ROIs are still observed after the merging step, only the most intense is kept and the total area (from zero) is integrated using the *trapz* function of the *pracma* package. If a ROI was found for the base peak, the same process is applied to the other isotopomer of the isotopic pattern, in the decreasing order of theoretical abundance, until no ROI is found.

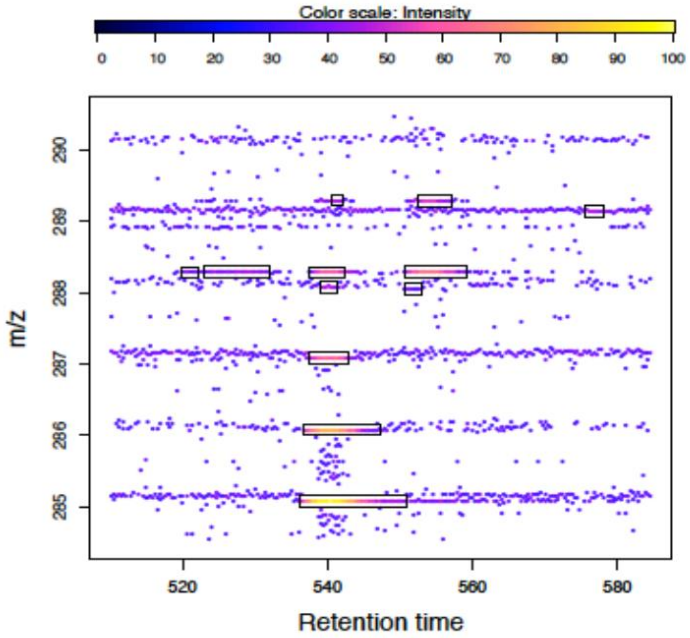


Figure 22: Region Of Interest (ROI) detection (Tautenhahn et al., 2008).

Thirdly, the three firsts pieces of information (see above) are calculated.

- The normalised (cumulative) area is the sum area of integrated isotopomers.
- The pattern score is determined according to **Equation 1**. It considers paired isotopomers between observed areas and theoretical abundances. Relative abundance deviations are weighted according to expected abundances.

$$Score = 100 \times \left(1 - \sum_{i=1}^{i=N} \left|Abd_i^{theo} - Abd_i^{obs}\right| \times \frac{Pat_i^{theo}}{Abd_i^{theo}}\right) \qquad \textbf{Equation 1}$$

where N is the number of centroided signals constitutive of the theoretical isotopic pattern above 1% of the base peak abundance determined according to the resolution specifications, i the increment of the paired theoretical and observed signal, Abd_i^{theo} the theoretical abundance of the signal i, $Area_i^{obs}$ the observed area

of the signal i, Pat_i^{theo} the theoretical abundance of the signal i relative to the cumulative abundances of the theoretical isotopic pattern. Abd_i^{theo} and $\sum_{i=1}^{i=N} Abd_i^{theo}$ are determined by a rule of three according to the observed plus missing isotopic signals and the theoretical isotopic pattern.

- The spectrometric deviation is calculated as a weighted average (on the basis of theoretical abundances) of the absolute *m/z* differences between paired observed and theoretical isotopomers.

For the standards, the application proceeds the same way as for chemical families, pending slight differences. Retention time ranges of EICs are restricted to the retention time set by the user (±2 min). Besides the integration at the baseline, the application proposes the integration from zero. The comparison between these two values is convenient in the case of an Orbitrap mass analyser due to the usual absence of chromatographic noise.

Once finished, the sub-tab "*Results*" of the tab "*Explore data*" is automatically activated (§ 5.2).

5.2. SUB-TAB LINEAR REGRESSION

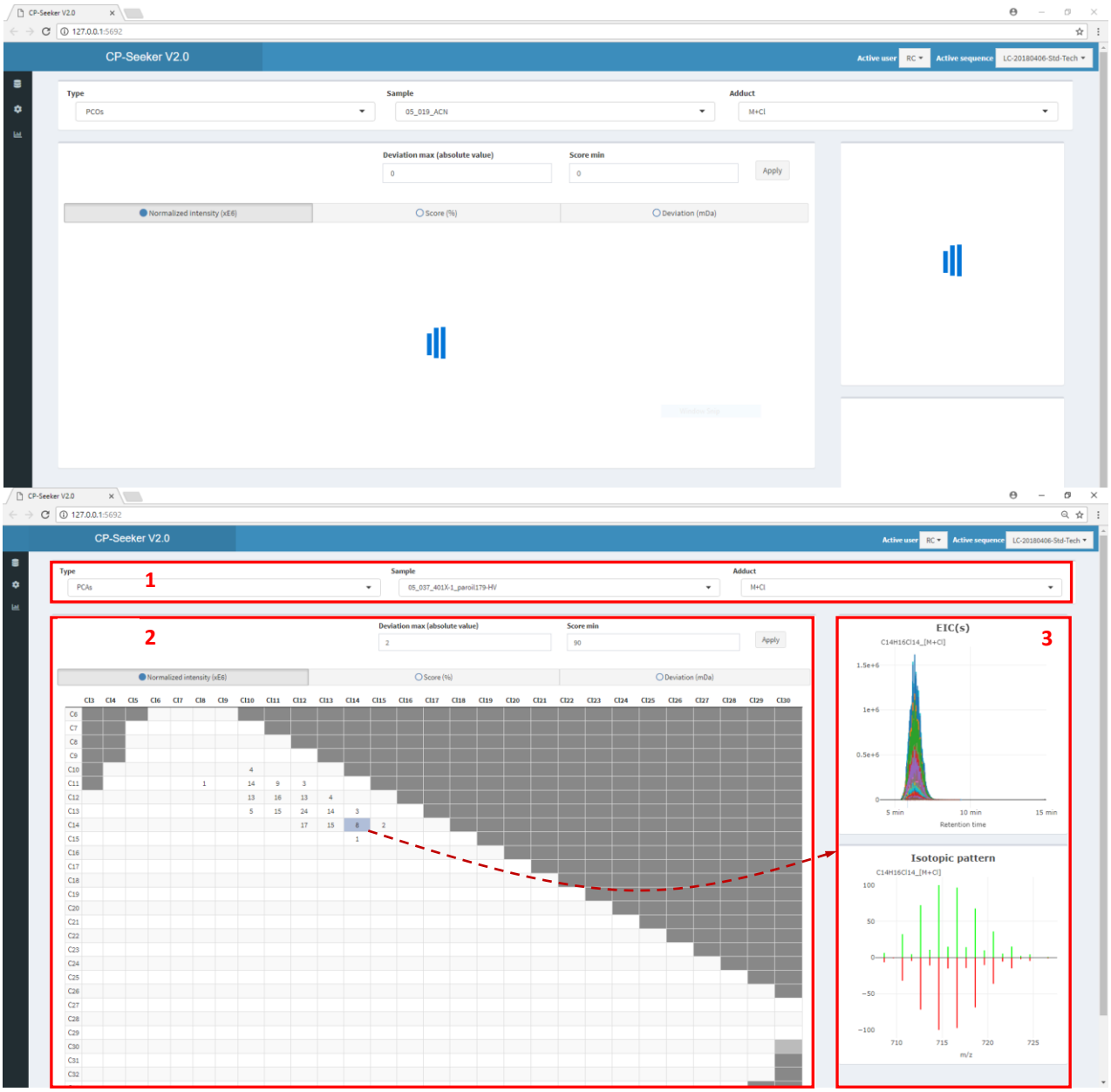
A sub-tab "Linear regression" is hidden in the current version of CP-Seeker. This beta version is not functional. Someday, we ambition to enrich CP-Seeker with the determination of best combinations of standard samples to explain unknown sample patterns.

6. TAB EXPLORE DATA

6.1. SUB-TAB RESULTS

6.1.1. Overview

Once the deconvolution is done, the application loads the results into display tables at once for a considered sequence (**Figure 23**). The processing might take a while (**up to a few dozens of minutes**), so be patient. The process starts again when changing the active sequence but not when coming back to the sub-tab within a same project. We intend to make improvements to minimise the waste of time here.



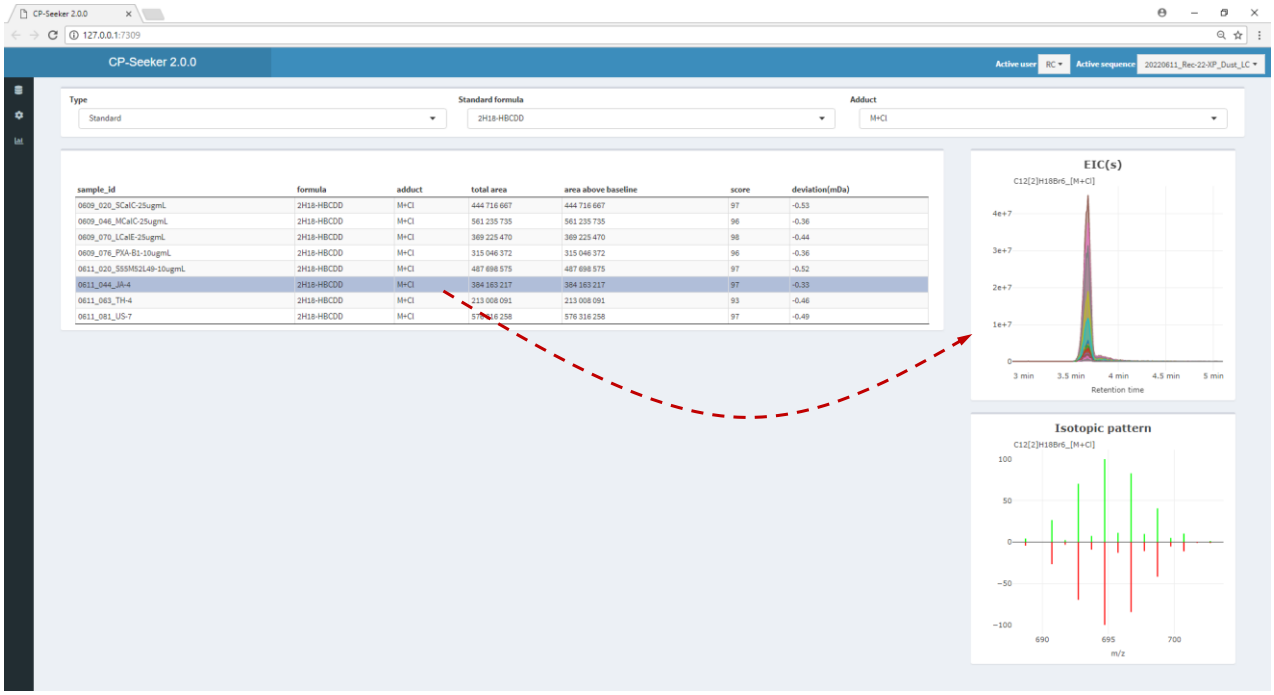


Figure 25: Displayed "*Results*" for $[M + Cl]^-$ adduct ions of $^2H_{18}$-HBCDD standard in a selected sample.

6.1.3. Target chemical families

For the target chemical families, the tables generally display the carbon number in lines and the chlorine number in columns (**Figure 26**). For PBAs, the bromine number is displayed in columns. For PXAs, the carbon number being set for the selected family, the bromine number is displayed in lines and the chlorine number in columns.

The user can select the targeted family, the sample and the adduct the top-area picklists. The user can also switch between the tab of normalised intensities, pattern scores and mass deviations. Filters on the maximum absolute mass deviation and on the minimum pattern score are also available. Do not forget to click on the "*Apply*" button. In that case, the application updates all the displayed tables of the sequence.

The displayed normalised intensities are rounded (no digit after the coma) and divided by 1,000,000 for convenience when working with an Orbitrap analyser.

The displayed pattern scores between the observed and theoretical isotopic patterns are rounded as well (%, no digit after the coma). The user must define its own threshold criteria above which he considers the homologue group identified. In addition, pay attention to potential spectrometric interferences.

The displayed weighted average mass deviations between paired observed and theoretical isotopomers are rounded (one digit after the coma).

The dark grey cell represent those isotopic patterns that fall outside considered range of halogens ($3 \leq x+y \leq n+3$; see §5.1.2), or fall outside the m/z range of the sample file, or that were not requested by the user. The light grey cells represent those straddling isotopic patterns as defined §5.1.4.

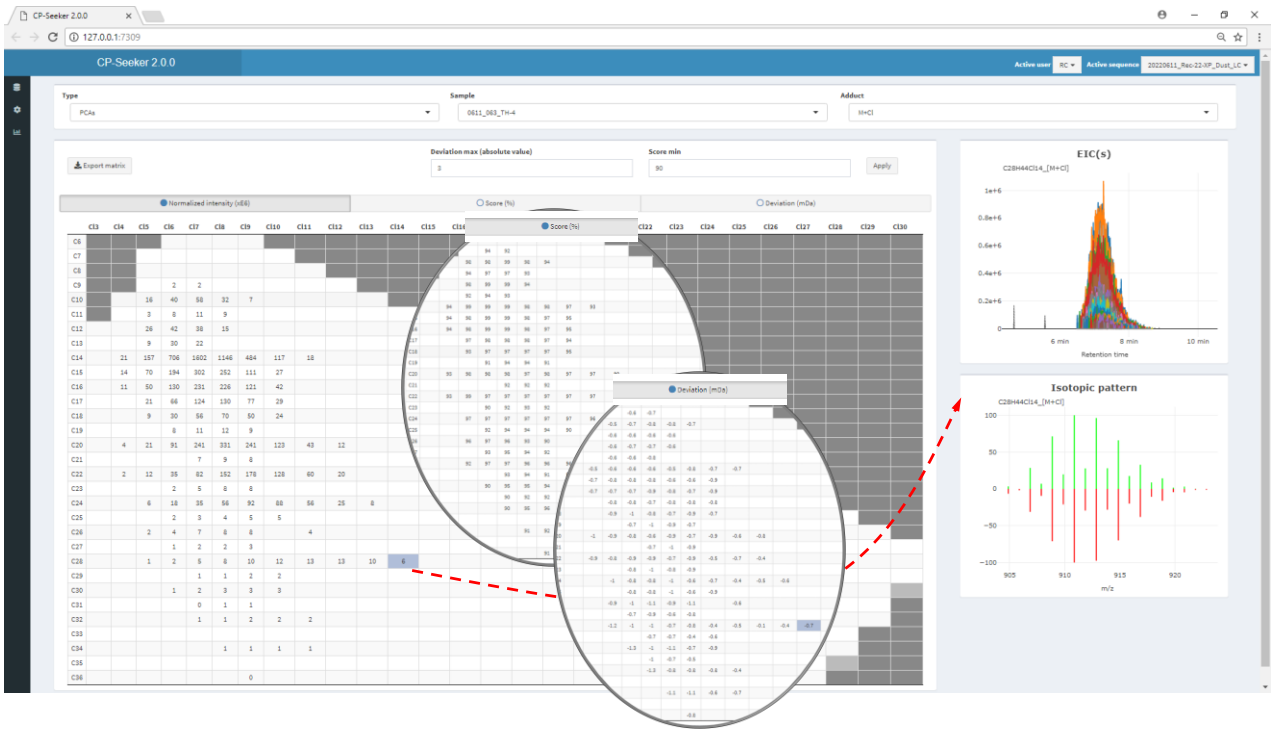


Figure 26: Displayed "*Results*" for $[M + Cl]^-$ adduct ions of PCAs in a selected sample.

6.1.4. Quantification

Someday, we ambition to enrich CP-Seeker with the semi-quantification scripts.

6.2. SUB-TAB TIC/EIC & MS

The "*TIC/EIC & MS*" sub-tab (**Figure 27**) offers limited possibilities to explore the sample data. After selecting the samples from the sequence in the picklist (1), a click on the "TIC" button will display the total ion chromatograms in (2). After providing a precise *m/z* with and associated tolerance (mDa) (1), a click on the "EIC" button will display the extracted ion chromatograms (2). A click on the TICs or EICs displays the mass spectra (3) corresponding to the selected retention time (t_R) coordinate. Report to §3.4 for plot interactivity tips.

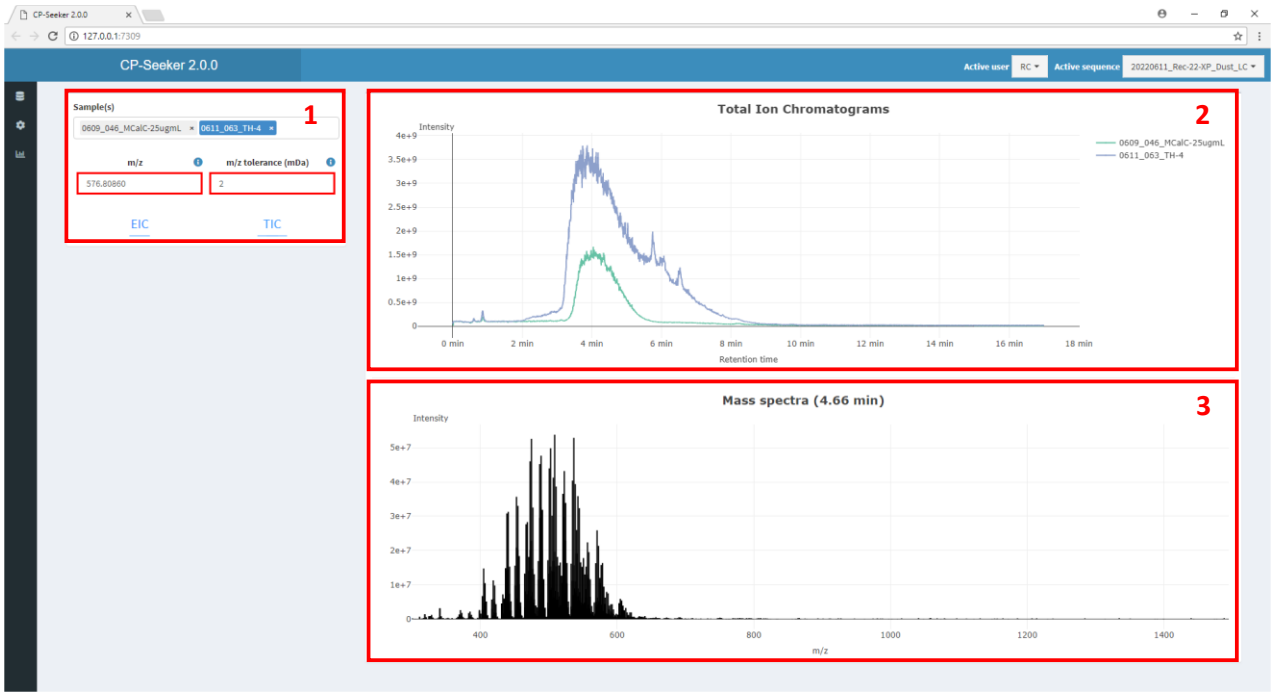


Figure 27: Overview of the sub-tab "*TIC/EIC & MS*".

6.3. SUB-TAB FIGURES

The application proposes four different graphical representations of the output data (normalised intensities, pattern scores or mass deviations) (**Figure 28**). Report to §3.4 for plot interactivity tips. Other interactive functions are also proposed here, such as rotation.

This part of CP-Seeker deserves to be enriched in a close future. **Currently, it is not working with brominated families**.

Figure 28: Overview of the sub-tab "*Figures*".

7. TAB EXPORT RESULTS

Selection of *Excel* document format and a click on the "*Launch Report*" button prepares export of the sequence data into formatted *Excel* files. Again, this processing might take a while (**a few dozens of minutes or hours**), so be patient. We plan to find a way to decrease this waste of time.

For each file, the entire sequence is exported. Numbers show many digits. Three templates are available, depending on the chemical family:

- The single family template for PCAs and for PBAs;
- The triplet family template for polychlorinated olefins (PCOs, PCdiOs, PCtriOs), for putative phase I (OH-PCAs, COOH-PCAs, oxo-PCAs) and phase II (GSH-OH-PCAs, SCys-OH-PCAs, Mercapturic-OH-PCAs) PCA metabolites;
- The PXA template for the C_6- to C_{36}-PXAs.

For each ion adduct a separated *Excel* file is prepared. All *Excel* files are exported within the folder "*C:\Users\<username>\Documents\.CP-Seeker 2.1.2*".

The *Excel* file names are composed of the sequence name, the creation day of the sequence, the adduct type and the family type.

The templates comprise a several worksheets:*

- Worksheet "*Sequence*" providing the sequence details;
- Worksheet "*Parameters*" gathering deconvolution parameters;
- One worksheet per standard;
- One worksheet per sample for the considered chemical family group and adduct ion.

Selection of *CSV* document format and a click on the "*Launch Report*" prepares export of the sequence data into a CSV file (comma-separated values). Headers are "*filename*" with extension, "*file_label*" as defined by the user, "*chemical_type*" (*e.g.* PCdiOs), "*homologue*" as the combination of C, Br and Cl atoms of the corresponding homologue group, "*adduct*" ion, normalised "*area*", isotopic pattern "*score*" and "*deviation*" in mDa.

8. FREQUENTLY ASKED QUESTIONS (FAQ)

The shortcut to launch CP-Seeker is not working anymore

Sometimes, antiviruses like Avast can mark the software *R* as a virus. If the application does not launch, check if your antivirus has been assigned to quarantine.

How to install the Microsoft update to make MSConvert work?

MSConvert requires updated Microsoft.NET framework 3.5 SP1 and 4.0 that we cannot embed in CP-Seeker. Indeed, full admin rights are needed to update it. They can be downloaded at:

https://www.microsoft.com/en-US/download/details.aspx?id=22

https://www.microsoft.com/en-US/download/details.aspx?id=17851

I face difficulties with my wiff files

In the case of a wiff file, please check the existence of its corresponding wiff.scan in the same directory.

I face difficulties with my Waters files

Depending on the MS method, converting Waters raw data files to *mzXML* is a bit tricky but manageable with msConvert. First, process the files with Masslynx\Tools\Accurate mass measure and save it (**Figure 29**). That step will avoid one of the known ppm deviation regularly seen during Waters file conversion to *mzXML*. In addition, if the method contains lockmass, then remove it before using msConvert by renaming the FUNC002.dat to FUNC002.old (here we assume that FUNC002.dat contains lockmass data but it is not always the case depending on the MS method). In addition, select the *CWT* value for the *Peak Picking* parameter, because there are no *Vendor* setting for this manufacturer in *msConvert*.

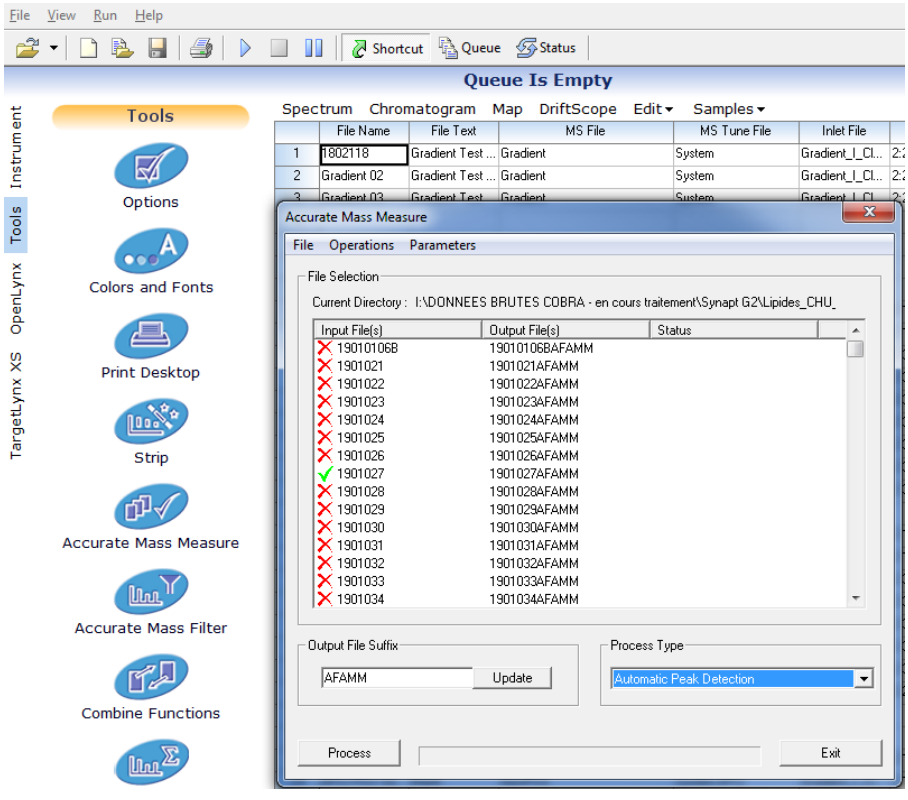


Figure 29: Accurate mass measure window of MassLynx to adapt Waters data files.

The *Chromium Portable* browser user interface (UI) turns grey and does not respond

The UI turning grey indicates that something went wrong and that *R* disconnected from the interface (and possibly even closed) for various reasons. Possibly, it happens when the computer is automatically paused (e.g. for **updates that you can temporarily desactivate if needed**). You can restart the application, possibly after forcing the **shutdown of the *R* application in the task manager** (ctrl+alt+suppr). Alternatively, contact us by e-mail, attaching the file *error.log* to help us improve the software. The file *error.log* can be found in the directory *C:\Users\<username>\.CP-Seeker 2.1.2*.

A weird and incomprehensible error message appears

This means that an error occurred during a process from a non-expected action. Contact us by e-mail, attaching the file *error.log* to help us improve the software. The file *error.log* can be found in the directory *C:\Users\<username>\.CP-Seeker 2.1.2*.