

Open tools for FTICR-MS data processing

EU-FTICR-MS European infrastructure
End User School - 13 Dec 2022 Lille

M-A Delsuc - IGBMC - Université de Strasbourg
CC 4.0 BY

Overview

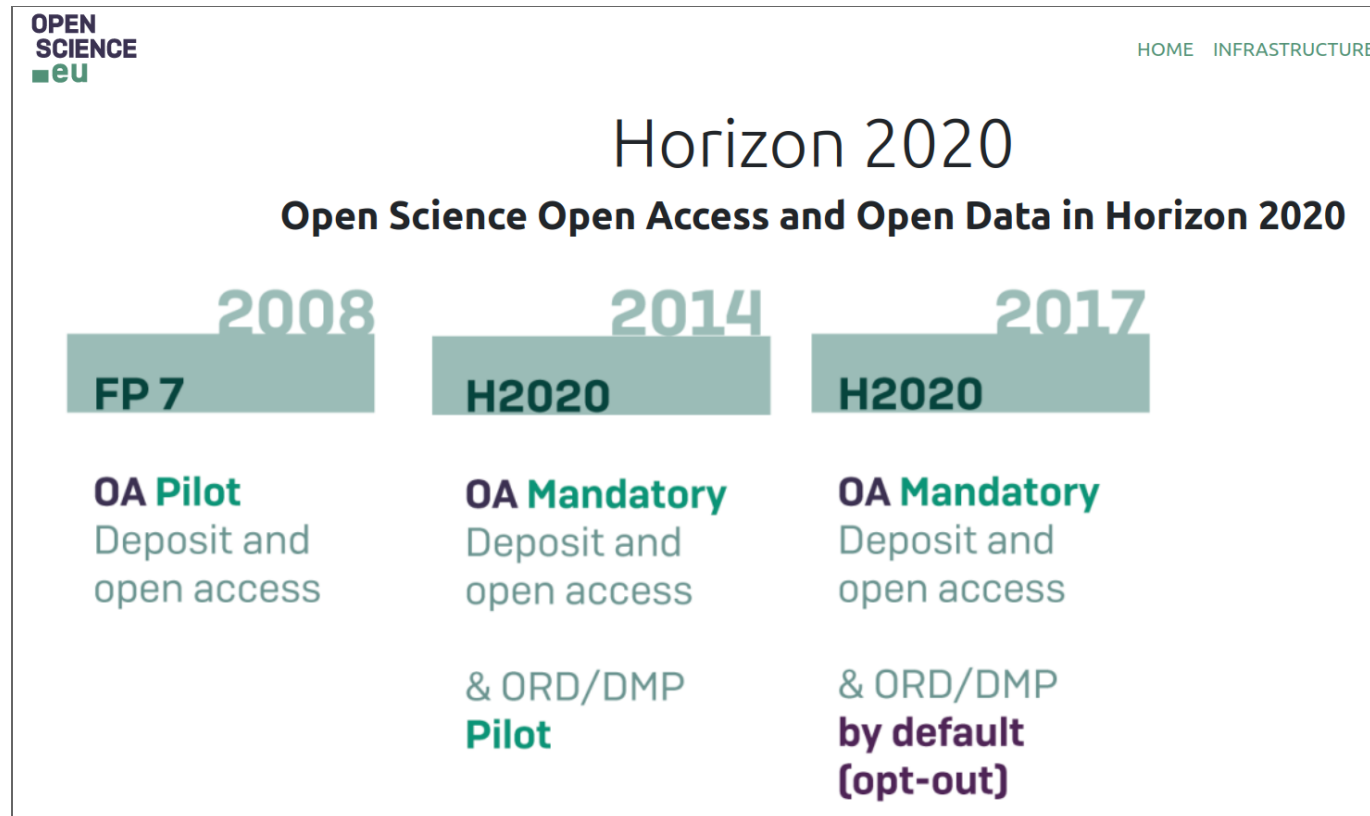
- Open Science
- Open Data
- Open Source Software
- Data Mining
- Interactive Program
- Larger datasets

Open Science



Open Science

Required by EU:



source: openscience.eu

Open Science

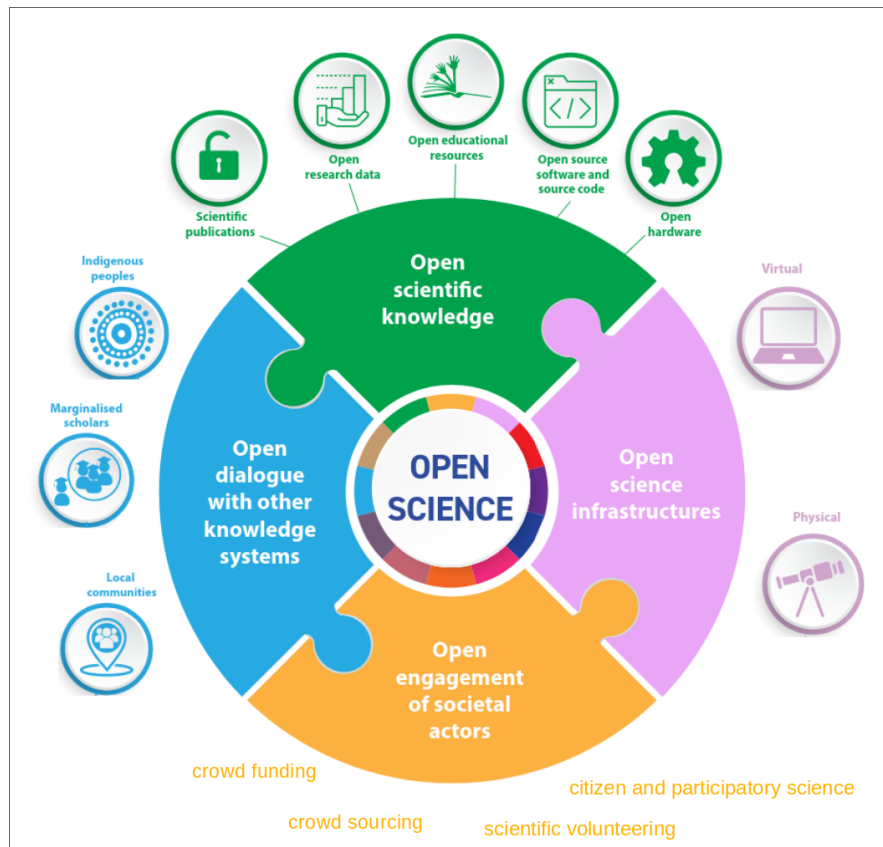
and by **UNESCO**:

Recommendation on Open Science adopted by the 41st session of UNESCO General Conference in Nov. 2021

Building on the essential principles of academic freedom, research integrity and scientific excellence, open science sets a new paradigm that integrates into the scientific enterprise practices for reproducibility, transparency, sharing and collaboration resulting from the increased opening of scientific contents, tools and processes.

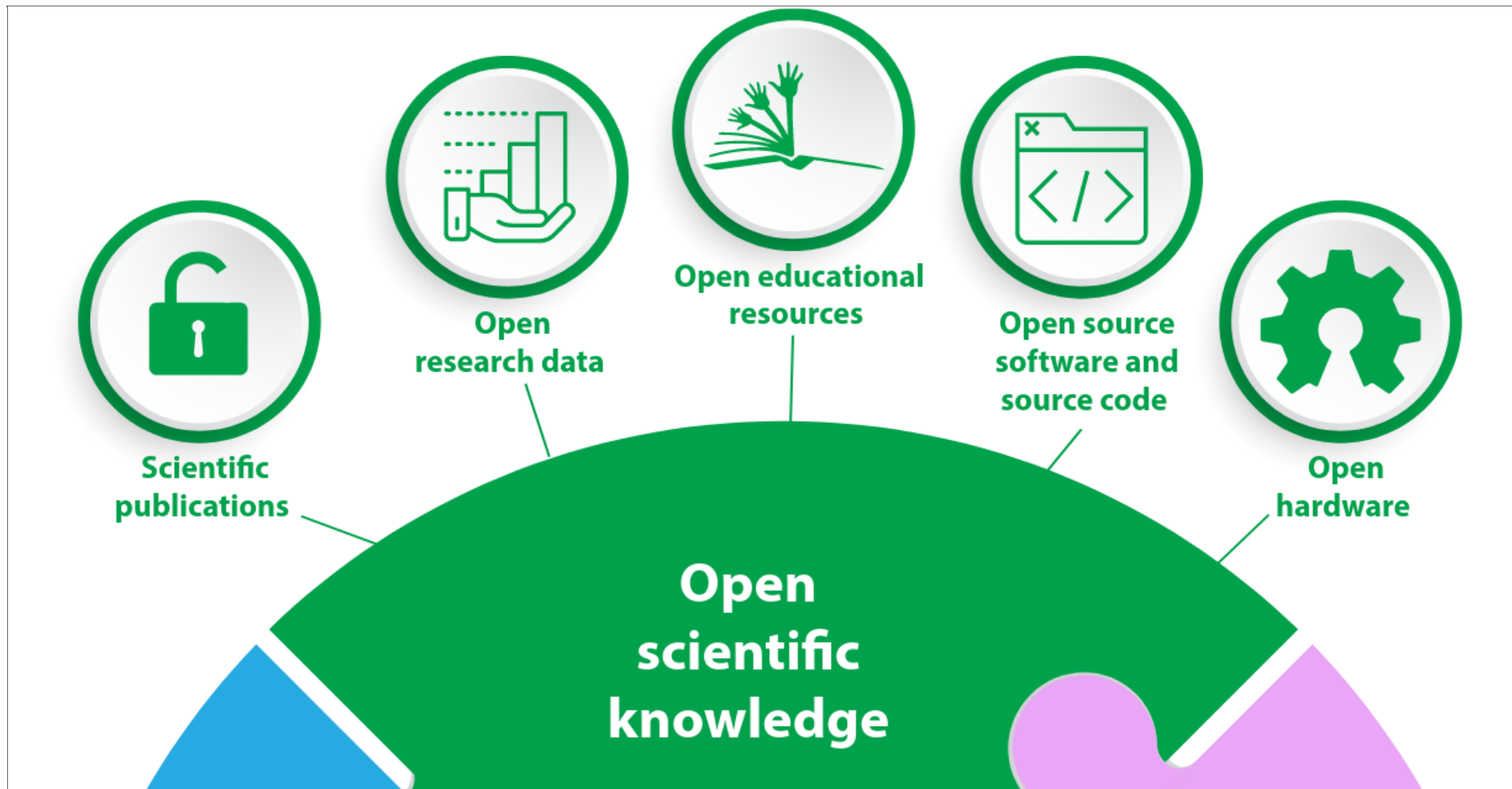
source: www.unesco.org/en/natural-sciences/open-science

Open Science Wheel

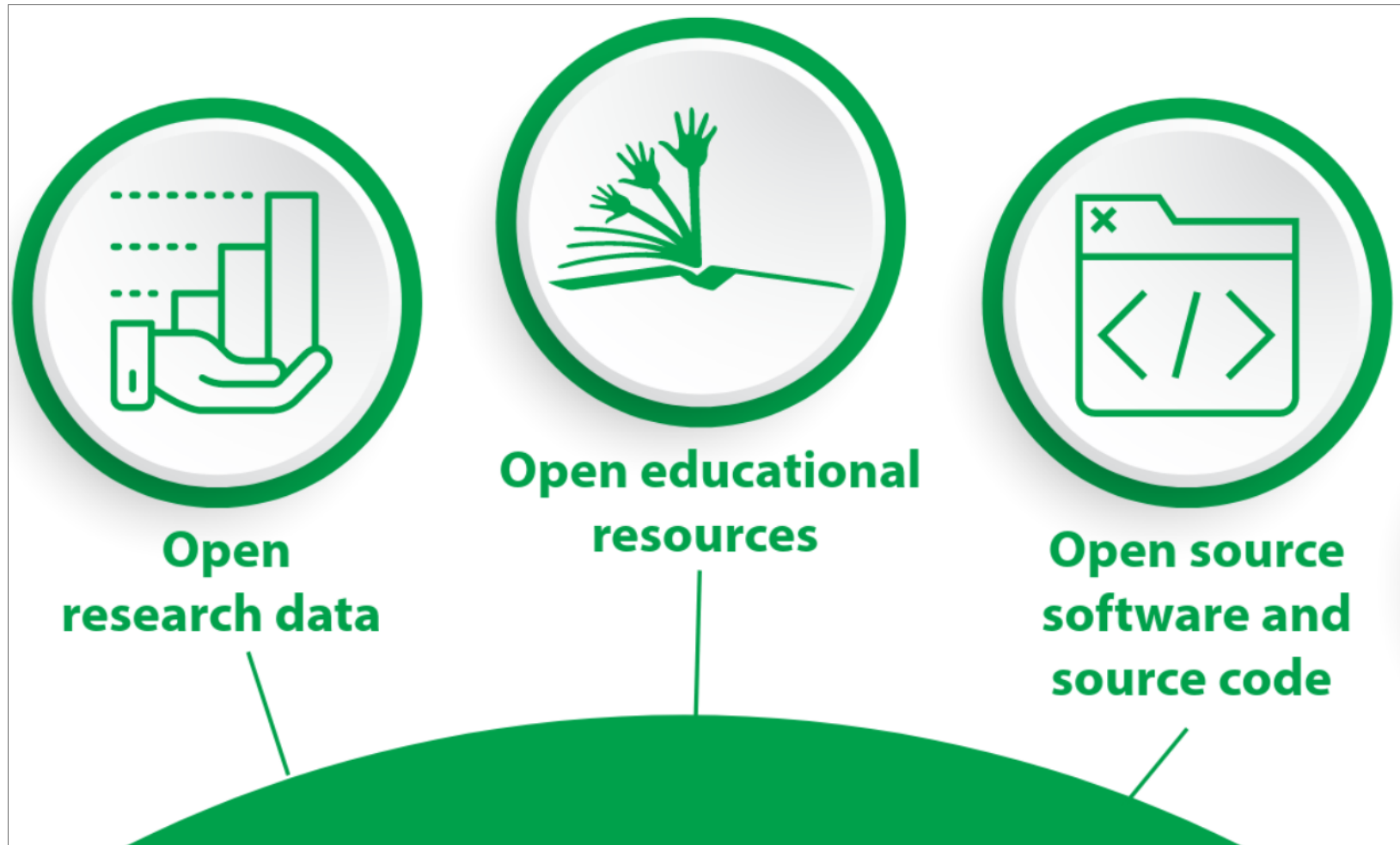


[source: www.unesco.org/en/natural-sciences/open-science](http://www.unesco.org/en/natural-sciences/open-science)

Open Science Wheel



Open Science Wheel



Open Data



EU-FTICR-MS Data repository

EU-FTICR-MS Public Data Repository

https://data.eu-fticr-ms.eu/#

Welcome on the public data repository for the EU FTICR MS project!


The European Network of Fourier-Transform Ion-Cyclotron-Resonance Mass Spectrometry Centers is a **H2020** project.

This network aims at:

- Providing the EU academic, SME and industrial communities with **access to world-class FT-ICR MS centers**
- Building an **EU community** of end-users and FT-ICR MS scientists
- Offering **open access** to data and **open source** software to the EU FT-ICR MS network
- Strengthening the FT-ICR MS application fields by promoting **innovative and cooperative research** between European FT-ICR MS academic scientists and private companies (instrumentation and software)

[Learn more](#)


Information
May, 2021



How to search for data?


- View data tree by clicking "Go to Data Collection" on top of the page
- Once on data page, type data description to filter through data

Information
May, 2021




What can be find on this repository?

- 14 projects available
- 6.5GB of data



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731077. This website reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Powered by **CASCADE**

Unless noted otherwise, data on this website are available under the licence [CC BY-NC 4.0](#) 

[Contact](#) [Terms and Conditions](#)

EU-FTICR-MS Data repository

Part of file name or content. Matching files are shown in red.

- DATA
 - MOSCOW_Skoltech
 - ROMA_Sapienza
 - Dopfer_September2020
 - Giampà_September2019
 - Niloufar_February2020
 - Rajeev_June2020
 - Salpin_January2020
 - Sarandrea_October2020
 - Data
 - ExemplarySpectrum.png
 - Sarandrea_October2020_v0.meta
 - Spezia_July2019
 - Tripodo_January2020
 - Usharan_November2018
 - ROSTOCK_UNI
 - ROUEN_CNRS
 - WARWICK_UNI
 - Welcome
 - WelcomeOnEUFTPPublicData.pdf

Preview Pane

Download File

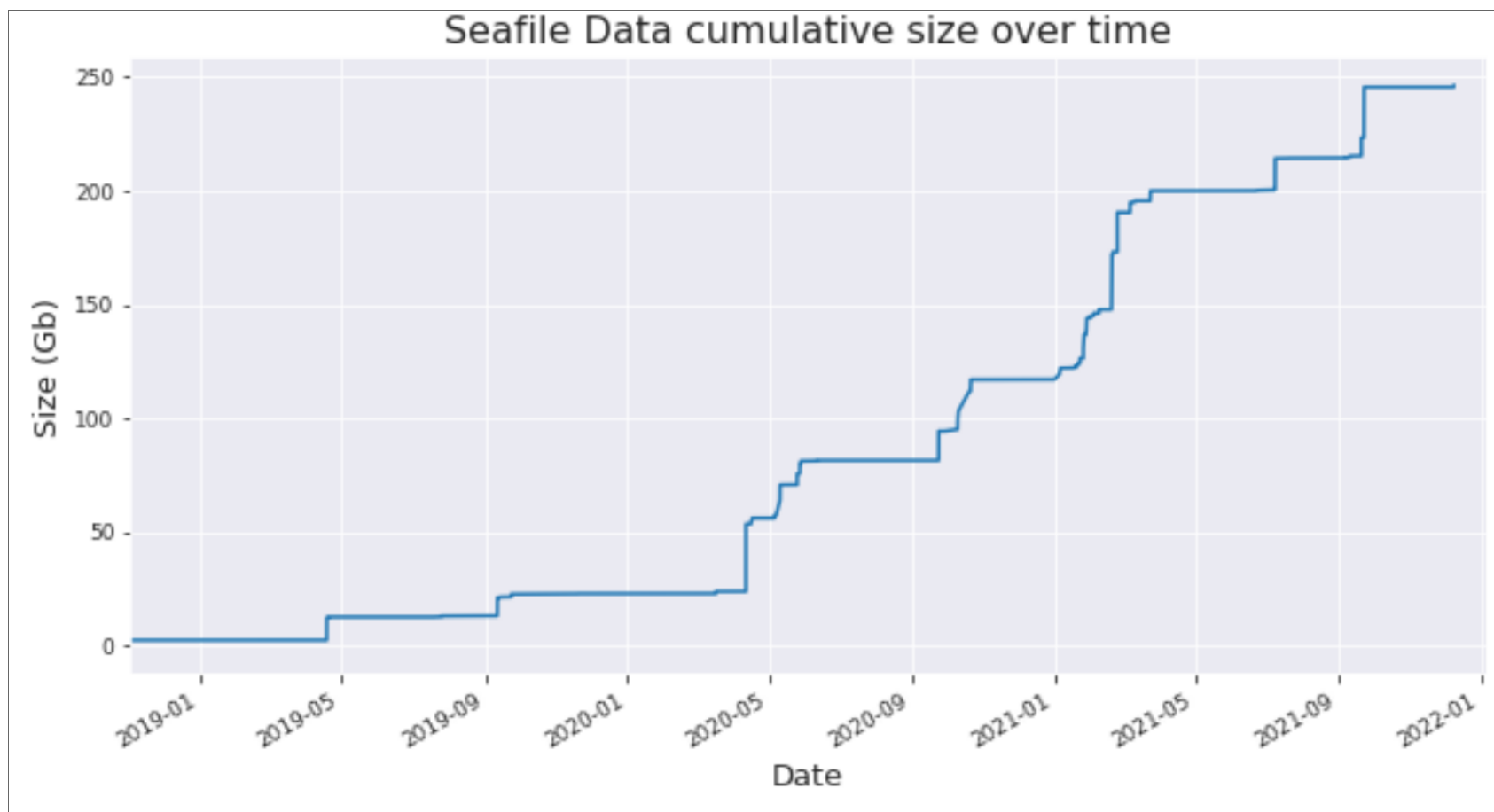
My Libraries/Repository/Sarandrea_October2020/1232/fid

a.u.

m/z

data.eu-fticr-ms.eu

EU-FTICR-MS Data repository



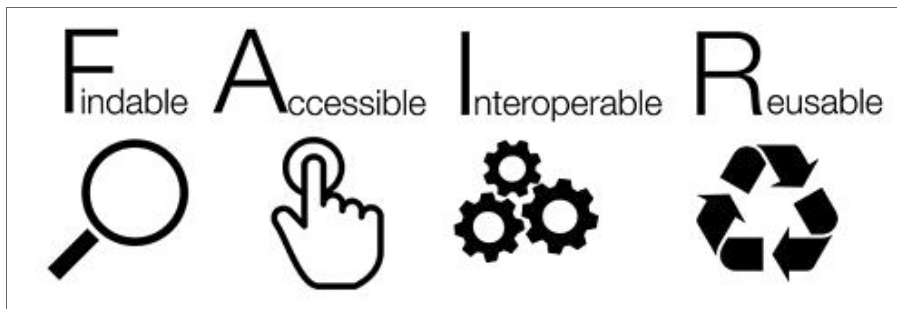
data.eu-fticr-ms.eu

EU-FTICR-MS Data repository

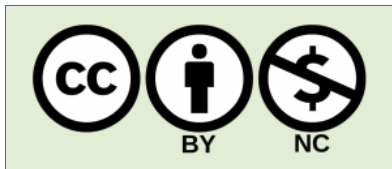
- **Data Management Plan**
describes repository policy

- **FAIR**

- Findable
- Accessible
- Interoperable
- Reusable



- **Open:** Data in free access

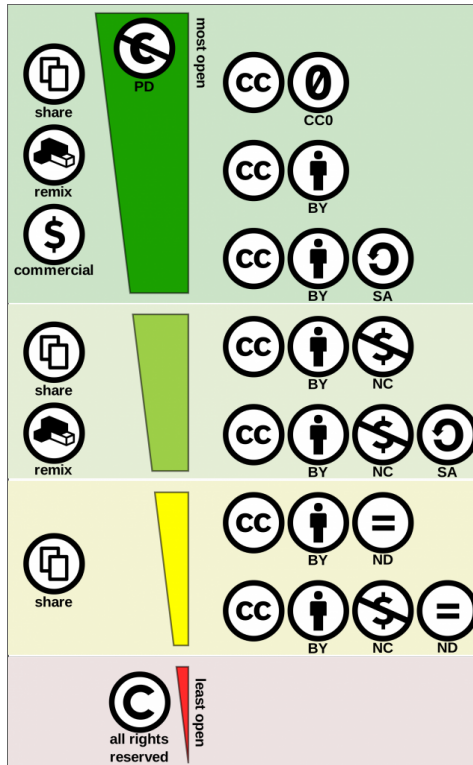


CC 4.0 BY NC

- Open \neq Free

Creative Common licences

An generic flexible open licence for texts



source : creativecommons.org

Public Domain

|

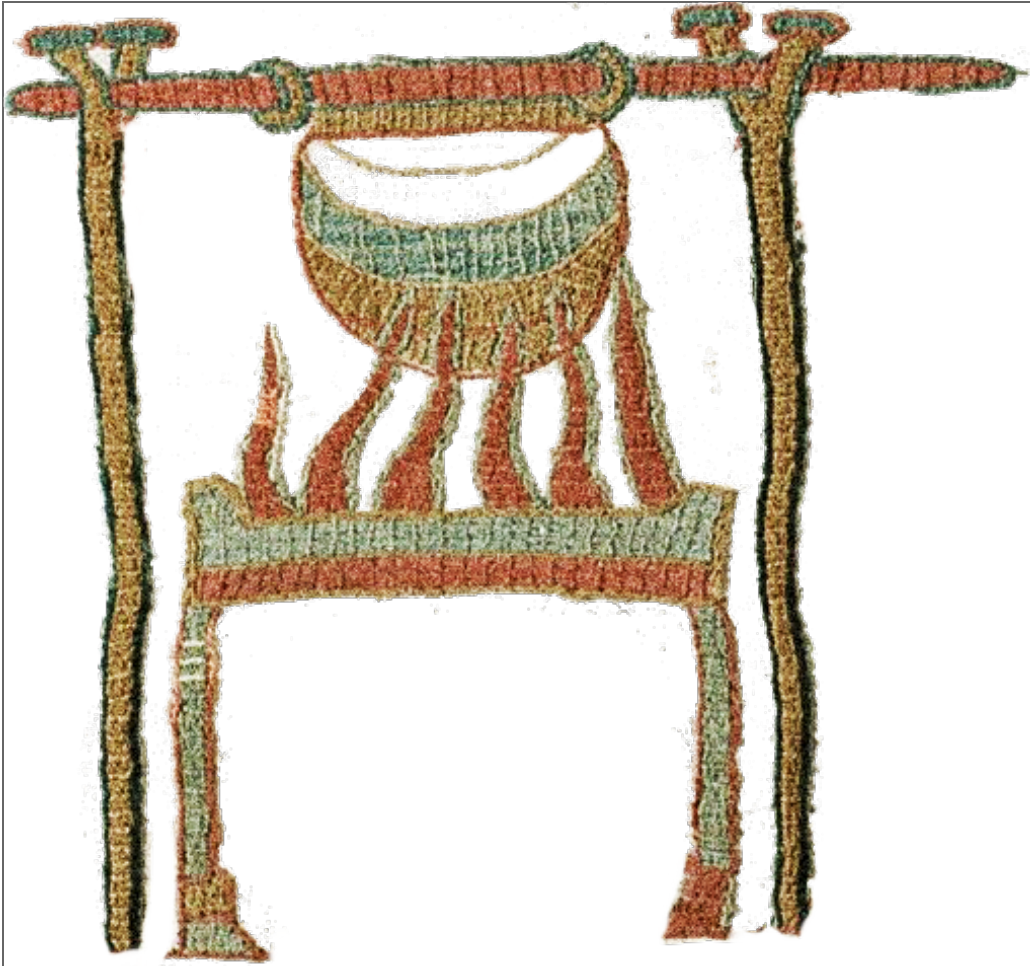
|

|

|

Copyrighted

Open Source Software



UNESCO Recommendation on Open Science

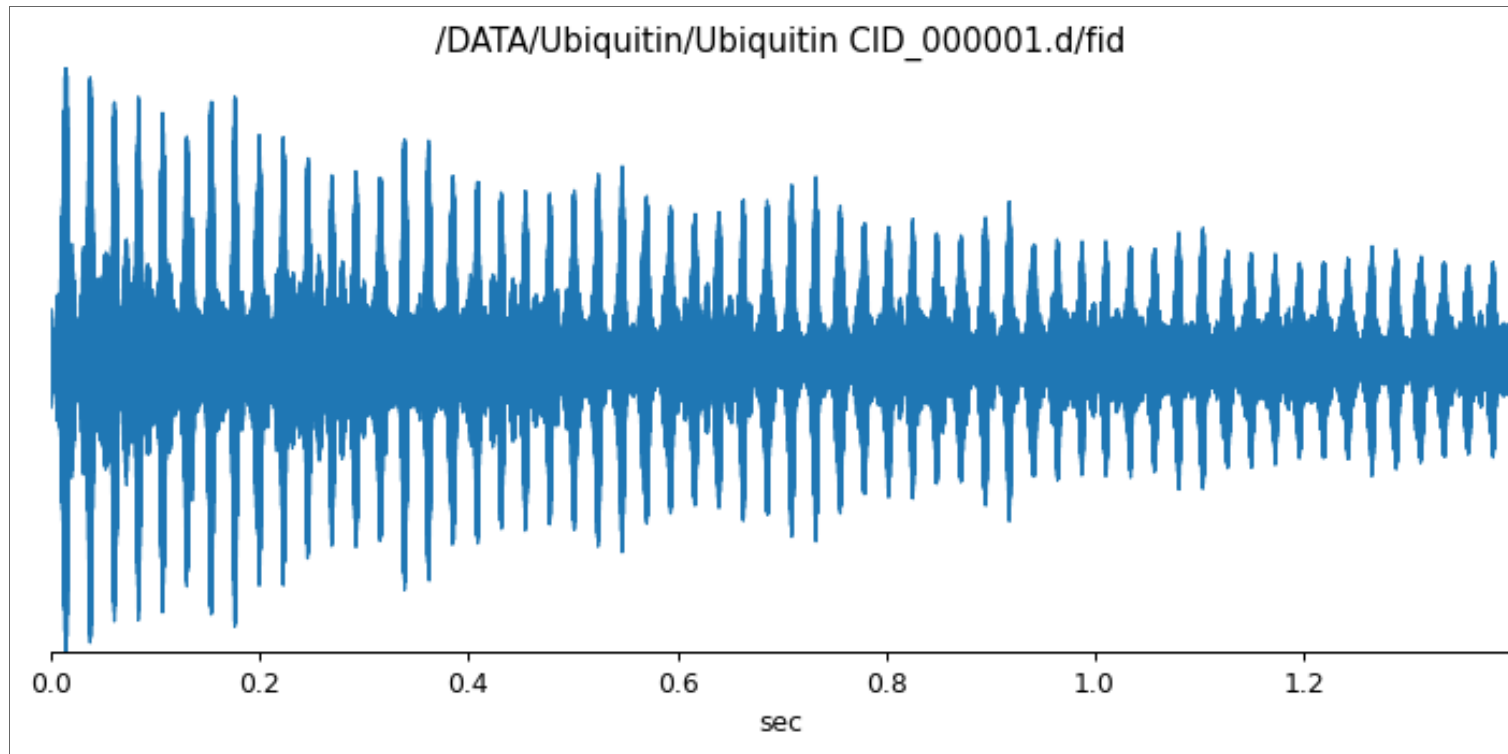
Building on the essential principles of academic freedom, research integrity and scientific excellence, open science sets a new paradigm that integrates into the scientific enterprise practices for reproducibility, transparency, sharing and collaboration resulting from the increased opening of scientific contents, tools and processes.

reproducibility...

Ion Cyclotron Resonance-MS is special !

FT-ICR does not measure a Mass Spectrum !

It measures a **Transient** 😞

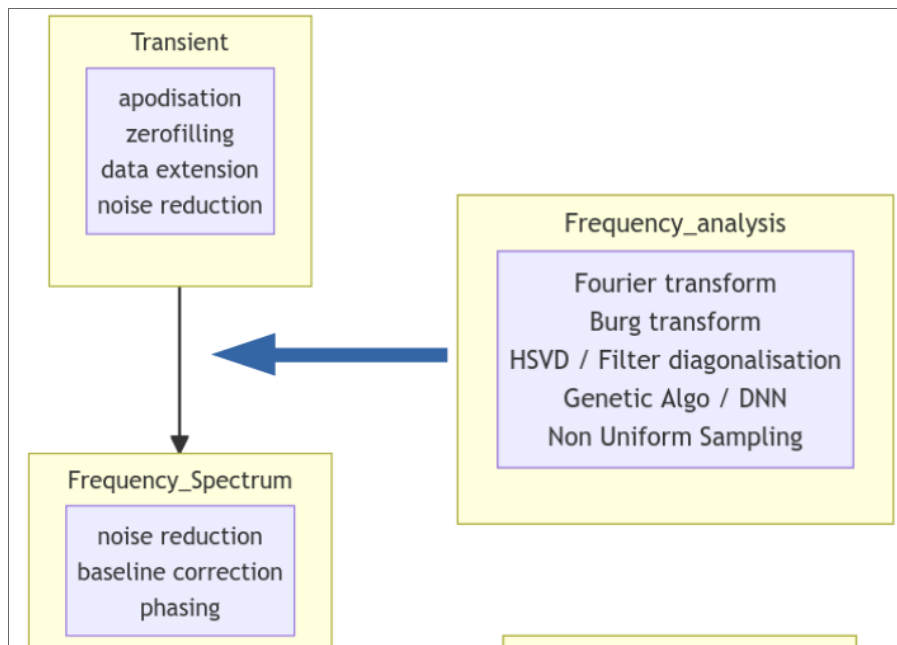


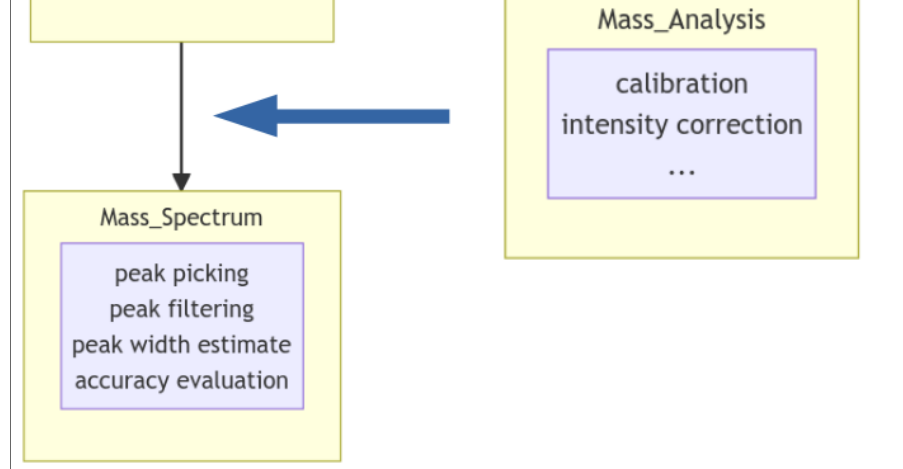
Processing

- **Transient**
- → Fourier Transformation
- Frequency Spectrum
- → to Mass Spectrum
- **Mass Spectrum**

Data Size

- Peak list \ll Transient
- Transient $<$ Spectrum

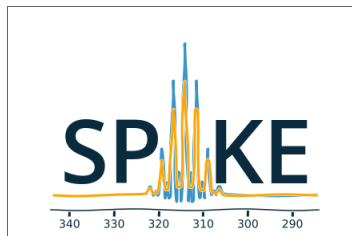




SPIKE

A data-processing library devoted to Fourier spectrometries.

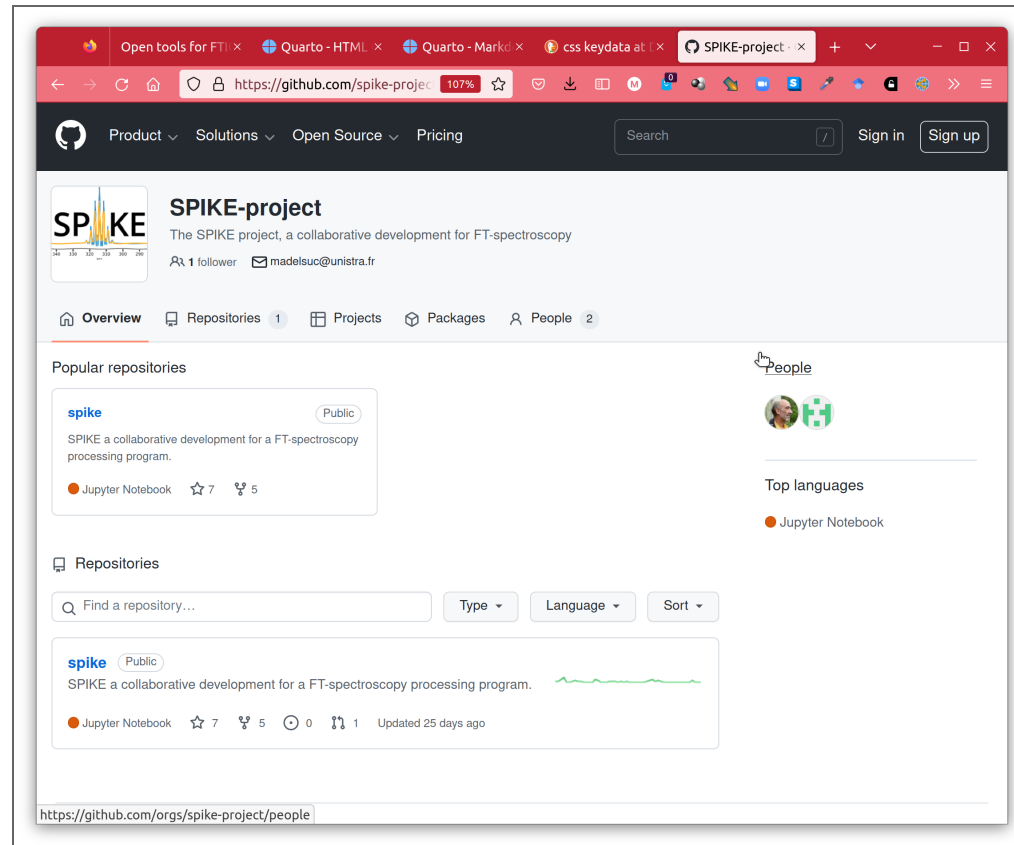
- Nuclear Magnetic Resonance
- FT-ICR Mass Spectrometry
- Orbitrap Mass Spectrometry
- ...



SPIKE a processing software dedicated to Fourier spectroscopies

Chiron L., Coutouly M-A., Starck J-P., Rolando C., Delsuc M-A. (2016) *arXiv*
1608.06777]

SPIKE



SPIKE is open-source, under GPL licence, distributed on GitHub

SPIKE in details

- written in python
 - multi platform (Windows, MacOS, Linux)
 - based on standard library ([numpy](#), [scipy](#), ...)
 - multiprocessing
 - GPU capabilities ([numba](#))
 - large files and off-memory processing ([HDF5](#))
 - internal compression
- specific methods
 - urQRd noise reduction,
 - Hypercomplex arithmetics
- complete access
 - metadata
 - batch processing
 - some interactive modules

MS & NMR spectrometries !

SPIKE Features

- Transient
 - apodisation
 - zerofilling
 - data extension
 - noise reduction
- Frequency_Spectrum
 - noise reduction
 - baseline correction
 - phasing
- Mass_Spectrum
 - peak picking
 - peak filtering
 - peak width estimate
 - accuracy evaluation
- Frequency_analysis
 - Fourier transform
 - Burg transform

- *HSVD / Filter diagonalisation*
- *Genetic Algorithm / DNN*
- Non Uniform Sampling
- Mass_Analysis
 - calibration
 - *intensity correction*
- More
 - MS series / LC-MS
 - 2D FTICR-MS

But...

user interface: 🤔

```
print('Reading file ',FC.selected)
d1 = BrukerMS.Import_1D(FC.selected)      # Import_1D creates a SPIKE FTICRData object, from which everything is available
d1.set_unit('sec')                        # it can be acted upon
d1.filename = FC.selected                 # and be extended at will
print(d1)                                # print() of the dataset shows a summary of the parameters

I.Show1D(d1, title=FC.selected, yratio=1) # and display (yratio=1 to have it centered)
# alternatively you can use the low-level tool below:
# d1.display() # with many options, plus access to matplotlib details
```

Reading file /DATA/Ubiquitin/Ubiquitin CID_000001.d/fid

FTICR data-set

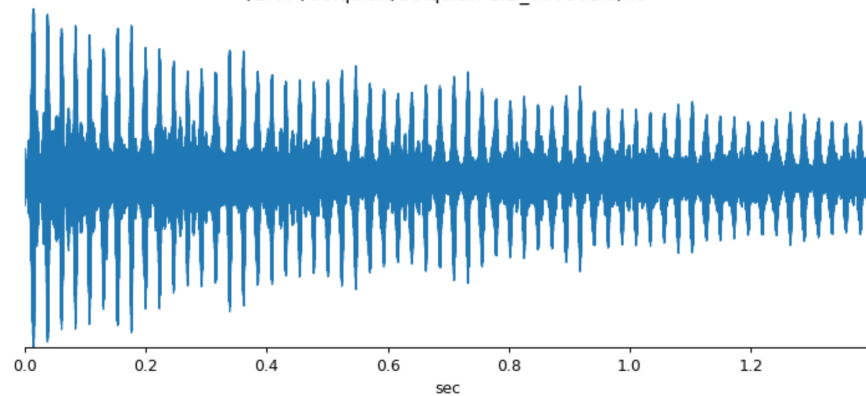
Bo: 15.00

Single Spectrum data-set

FT-ICR axis at 1500.000000 kHz, 4194304 real points, from physical m/z = 153.524 to m/z = 3000.000 R max (M=400) = 1609819

Show1D(children=(VBox(children=(Button(button_style='success', description='Reset', layout=Layout(width='80px'...

/DATA/Ubiquitin/Ubiquitin CID_000001.d/fid



In [5]:

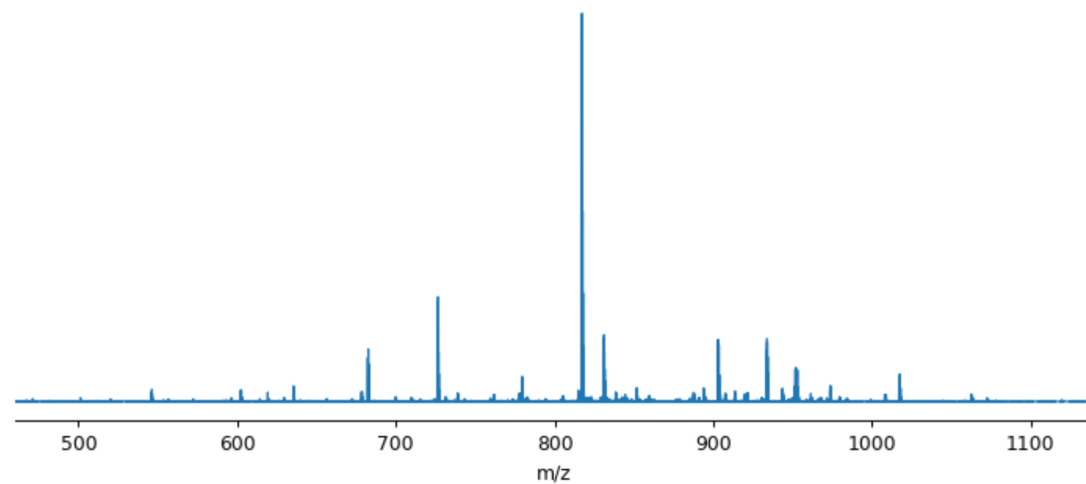
```
D1 = d1.copy() # copy the imported data-set to another object for processing
D1.center().kaiser(4).zf(4) # chaining centering - apodisation - zerofill
# kaiser(4) is an apodisation well adapted to FTICR, slightly more resolution than hamming - try varying th
if D1.axis1.itype == 0: # means data is real (common case)
    D1.rfft().modulus() # chaining real FT - modulus
else: # data is complex, in Narrow-band acquisition
    D1.fft().modulus() # chaining complex FT - modulus
D1.bcorr(xpoints=50) # flatten the baseline
D1.set_unit('m/z')

FI.Show1D(D1, title=FC.nmrname) # and display

# D1.display(title=FC.selected_path) # alternative, lower level display method with more options
```

Show1D(children=(VBox(children=(Button(button_style='success', description='Reset', layout=Layout(width='80px'...

Ubiquitin/Ubiquitin CID_000001.d

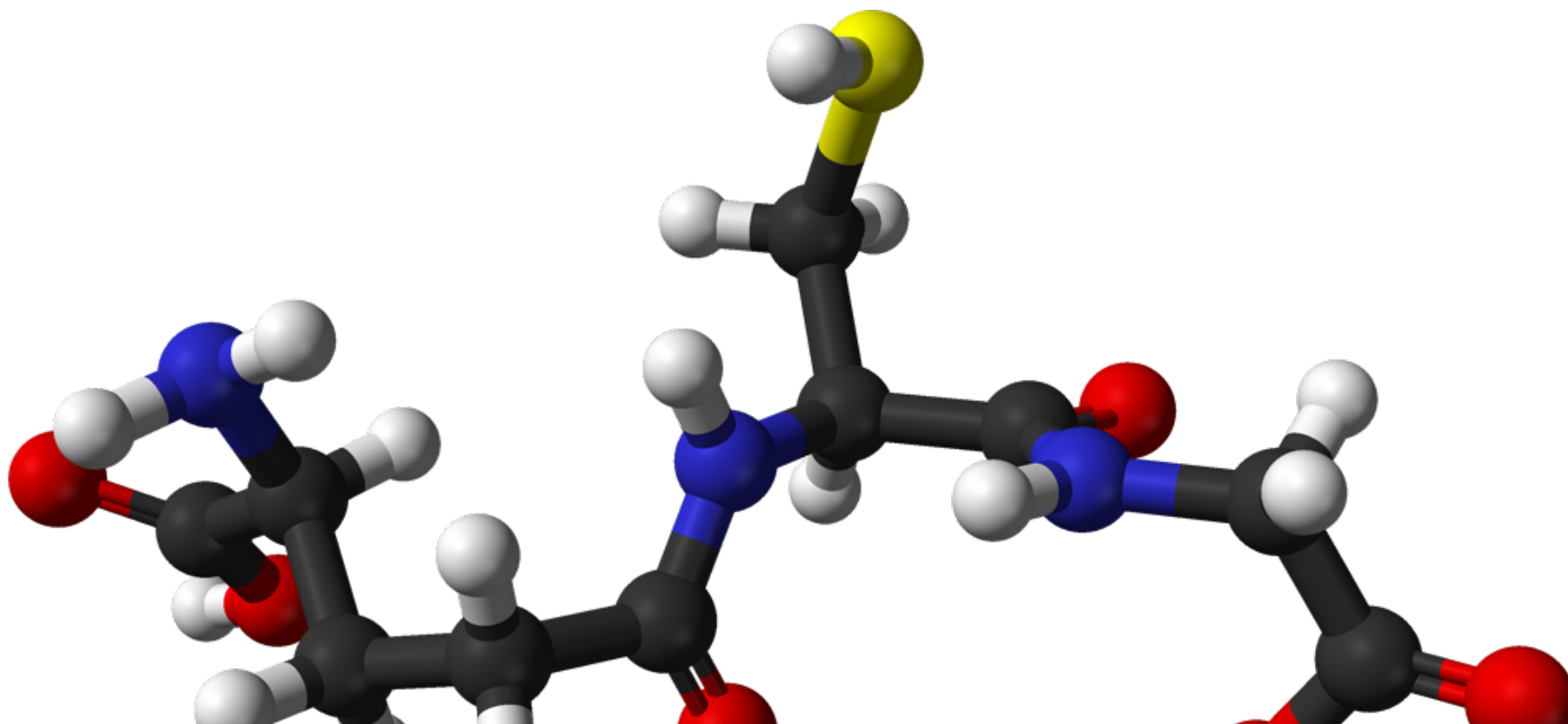
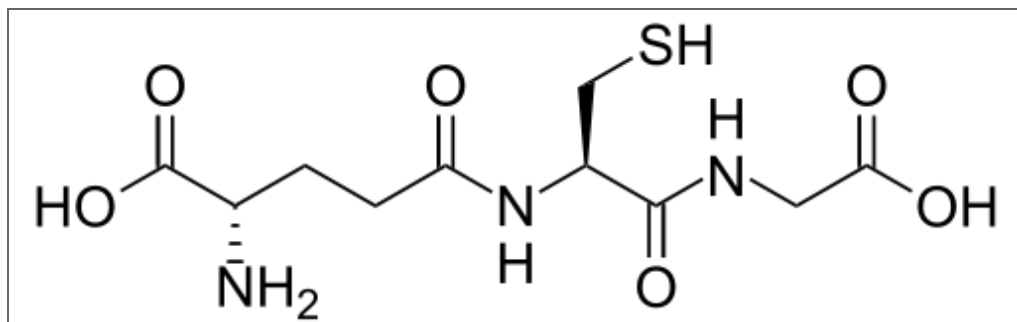


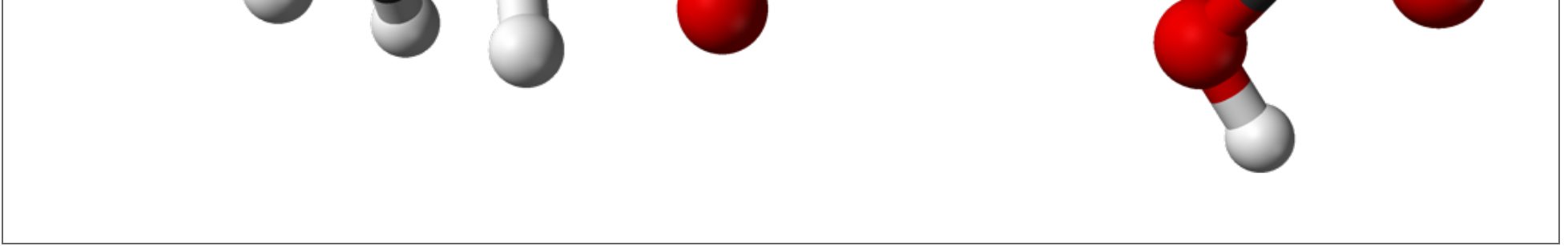
Data Mining



Data Mining

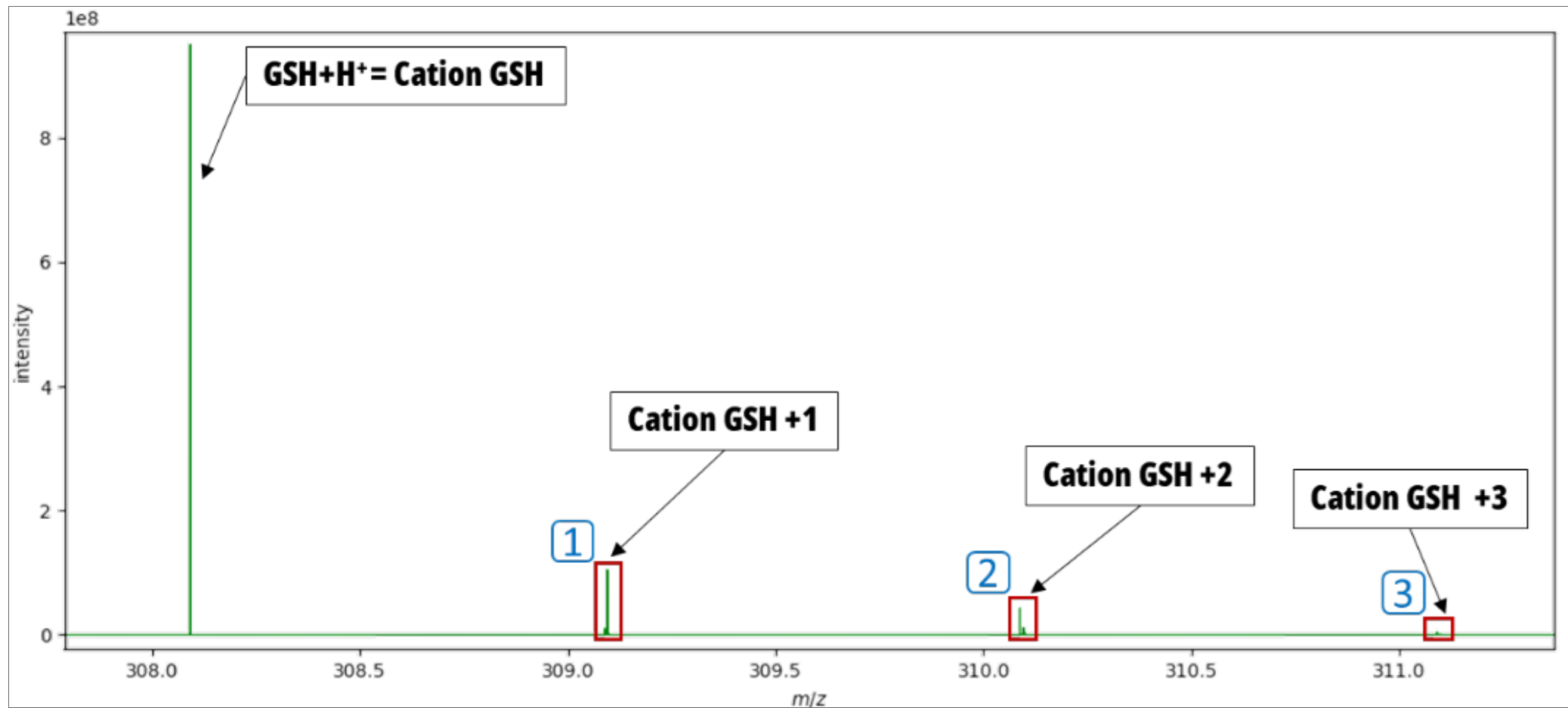
A Round Robin test performed over all Instruments in the Network.



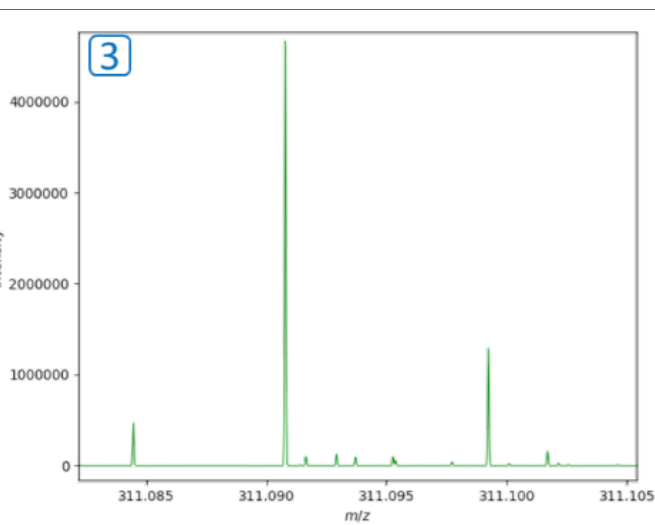
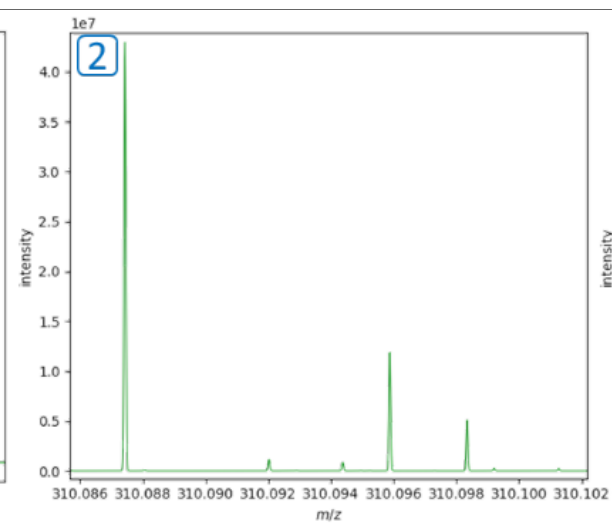
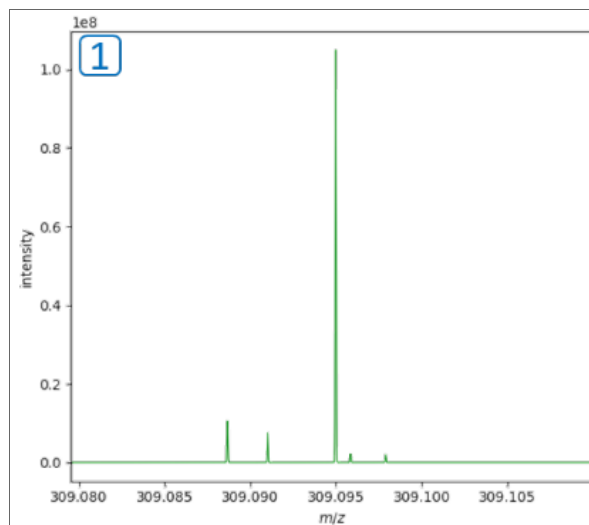
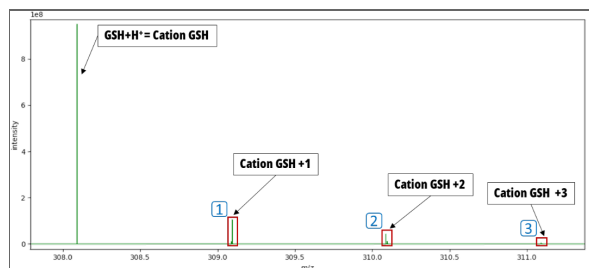


Glutathione molecule

chosen for its fine isotopic pattern



fine isotopic pattern of Glutathione



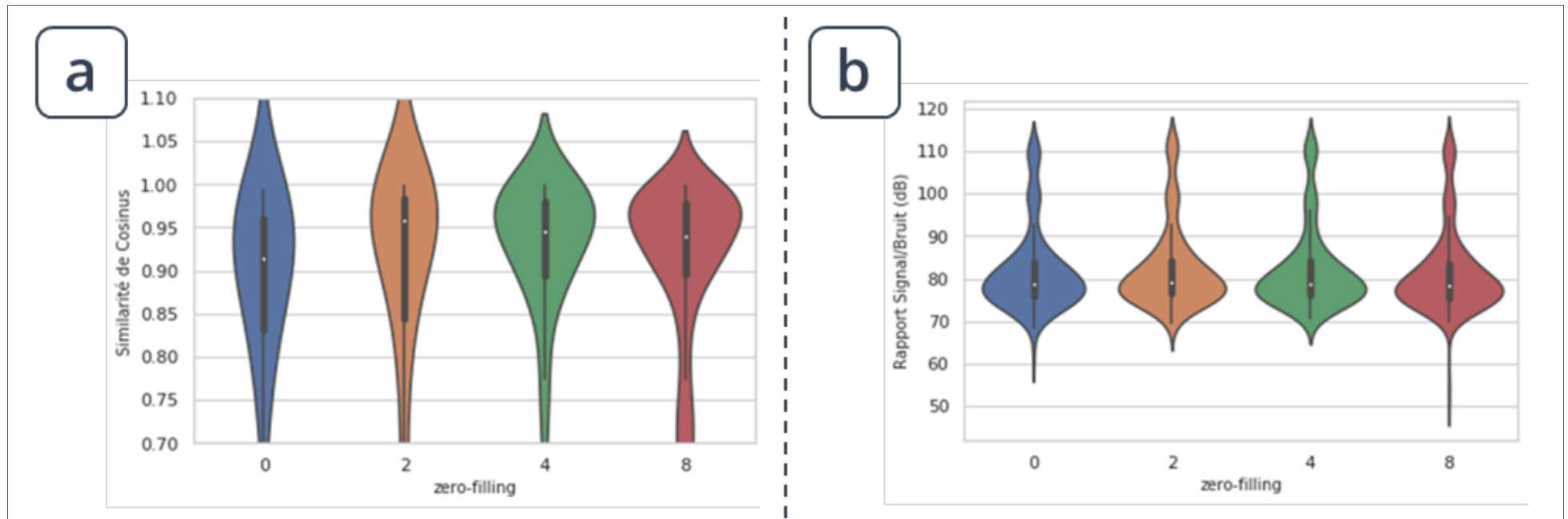
fine isotopic pattern of Glutathione

DATA Mining

Laura Duciel

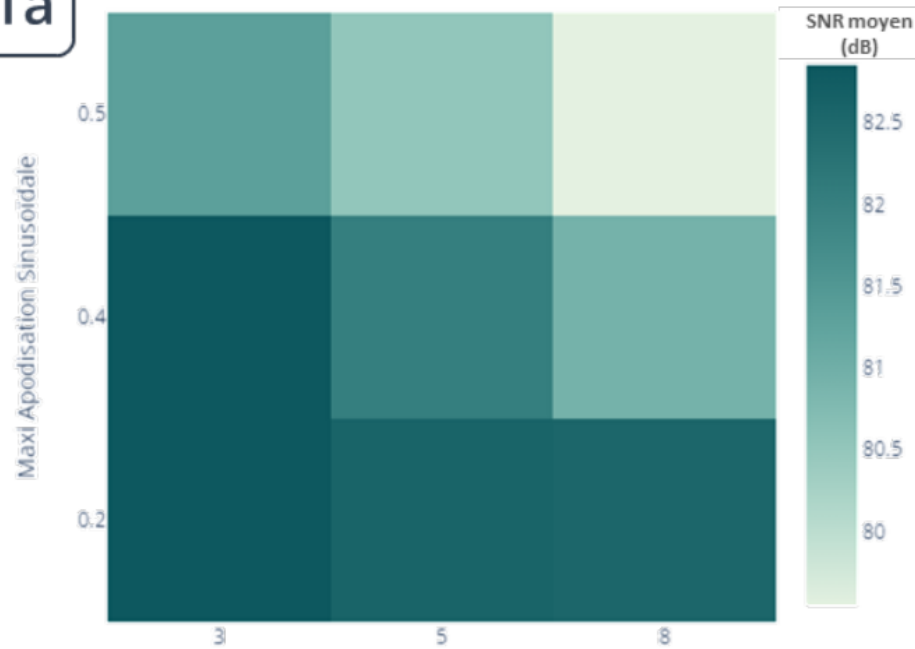
Comparison between the various processing parameters

Batch processing of several *different* MS experiment

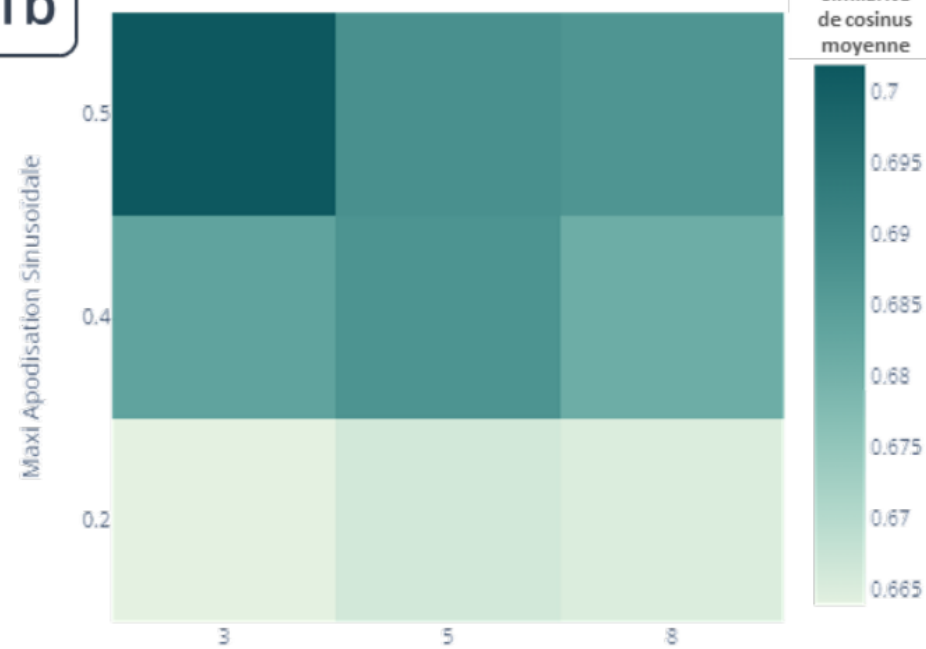


comparing zero-filling
with 2 methods:
(left) cosine difference
(right) SNR

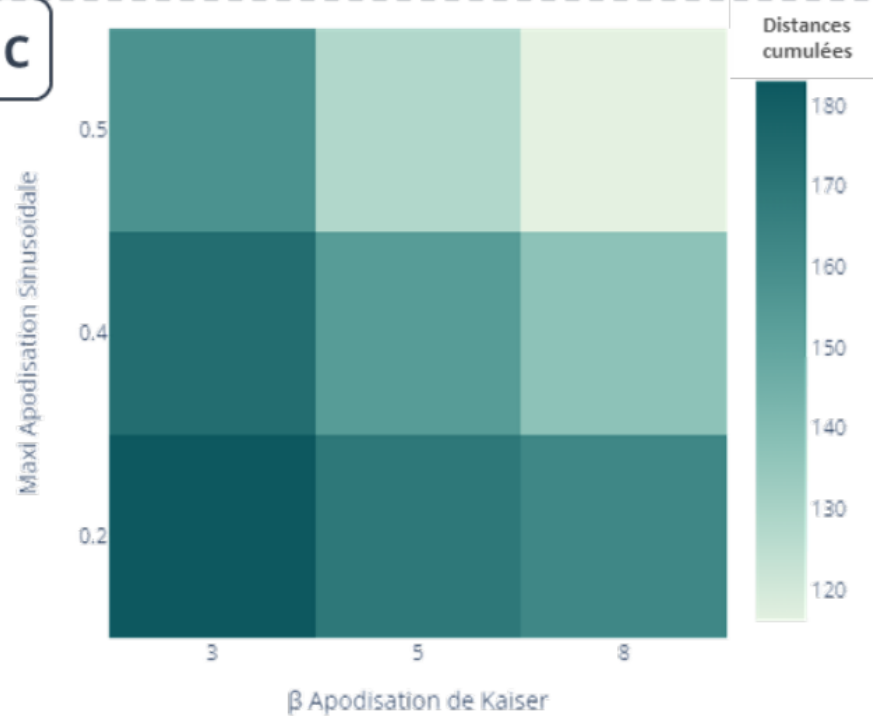
1a



1b



1c

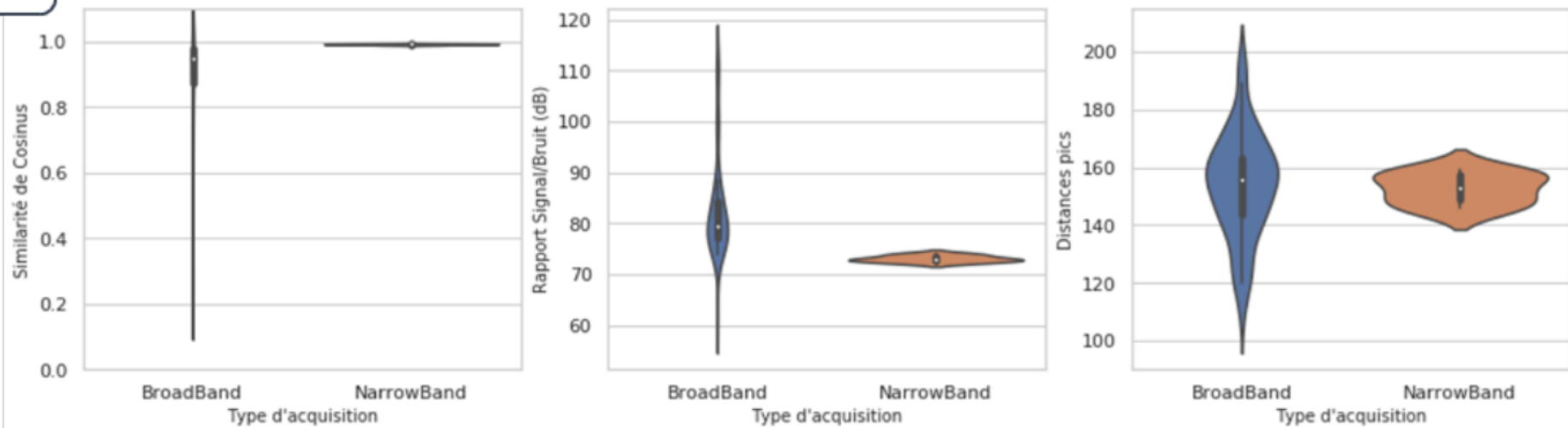


comparing apodisation

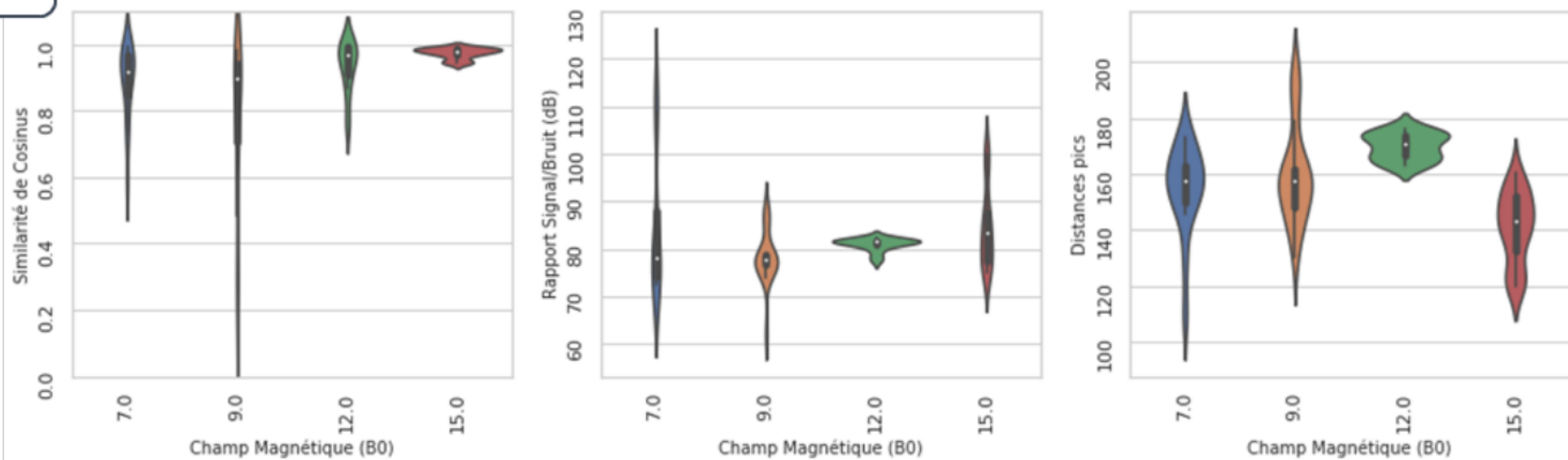
DATA Mining

Comparison between the various acquisition optimisation

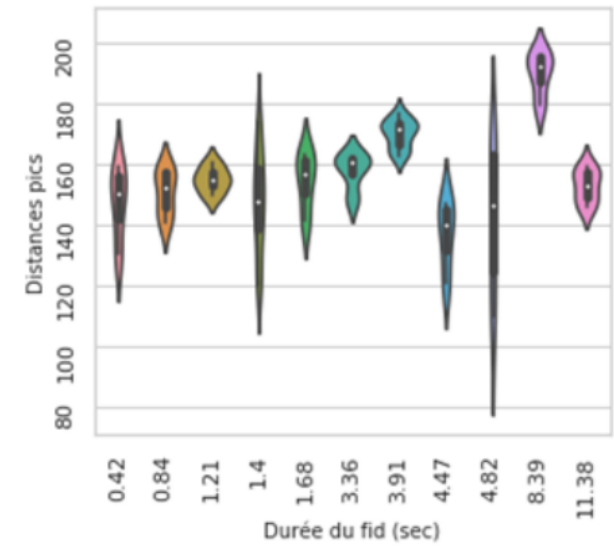
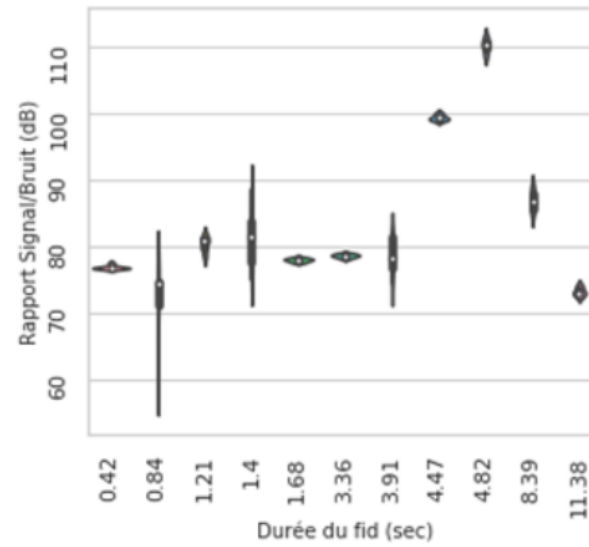
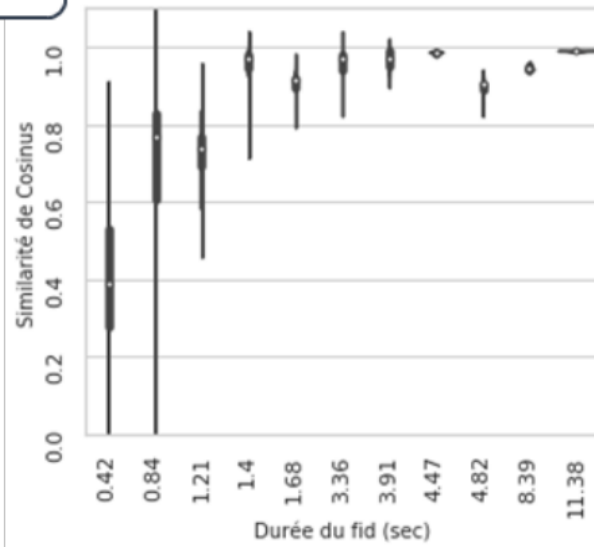
1



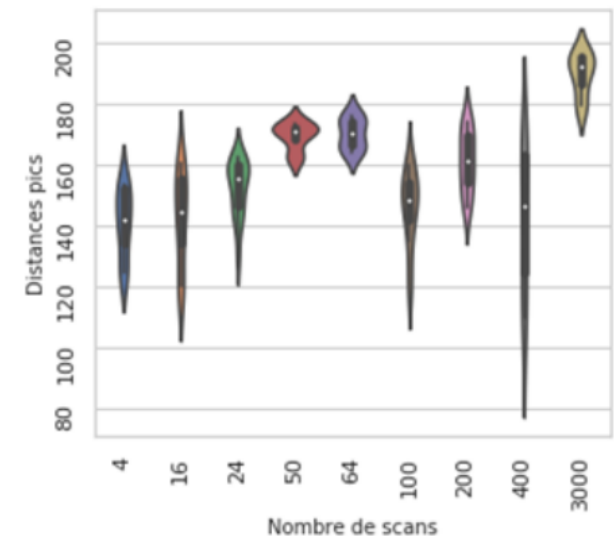
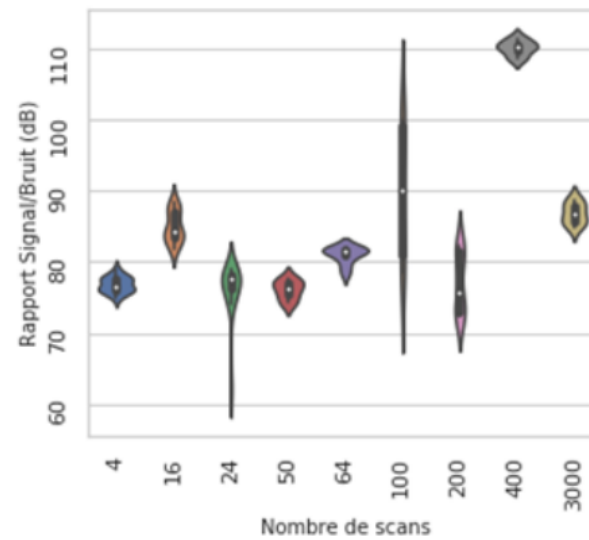
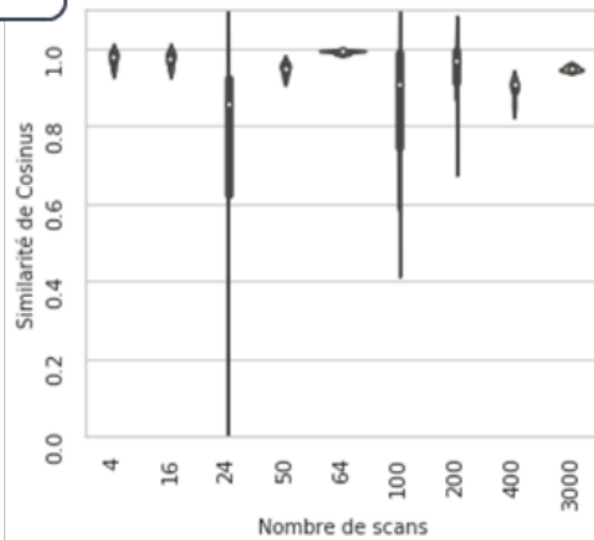
2



3

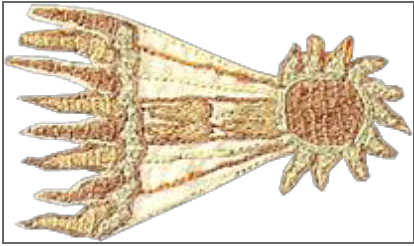


4



publication in progress

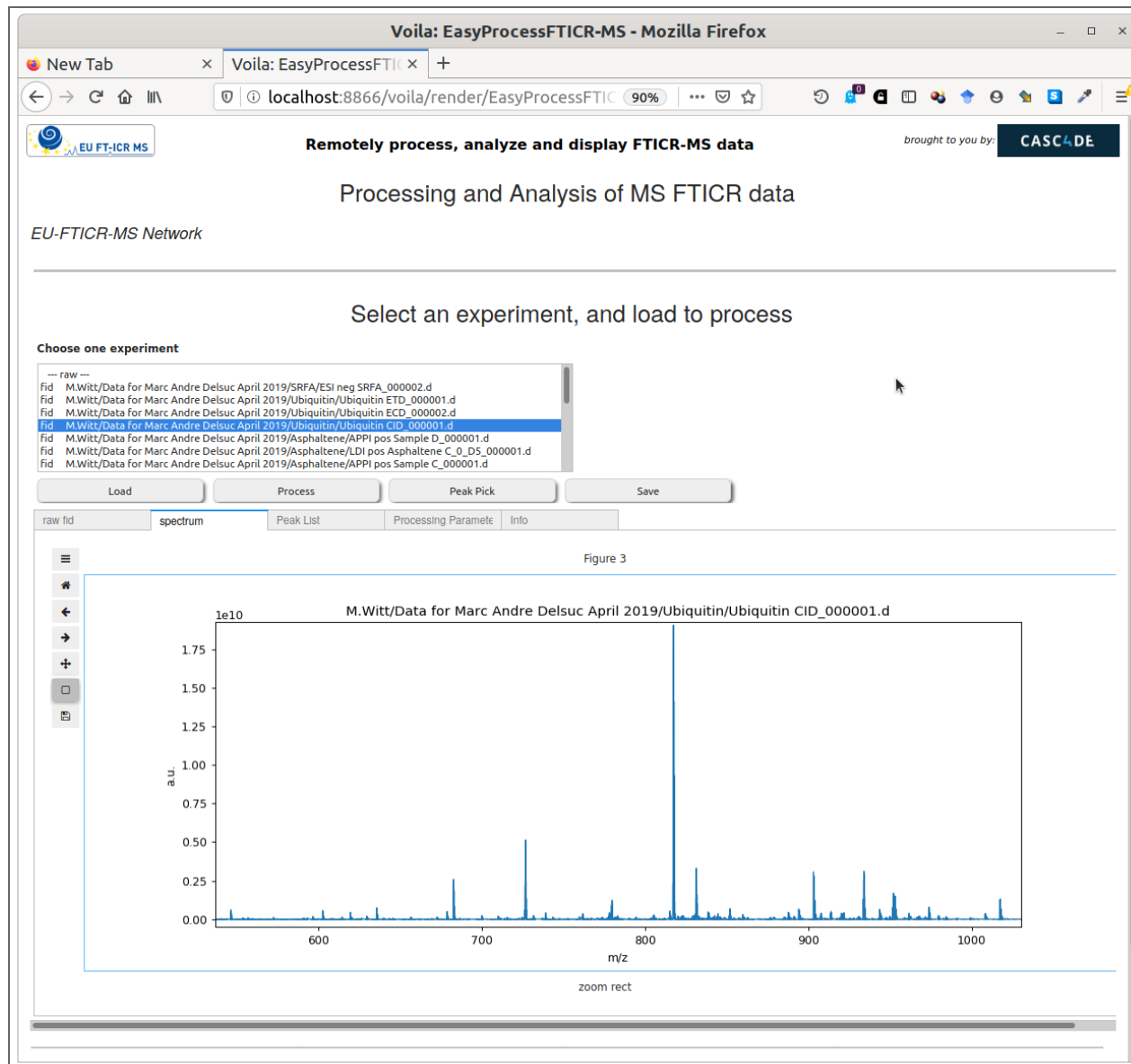
Interactive Program



developped by



presentation



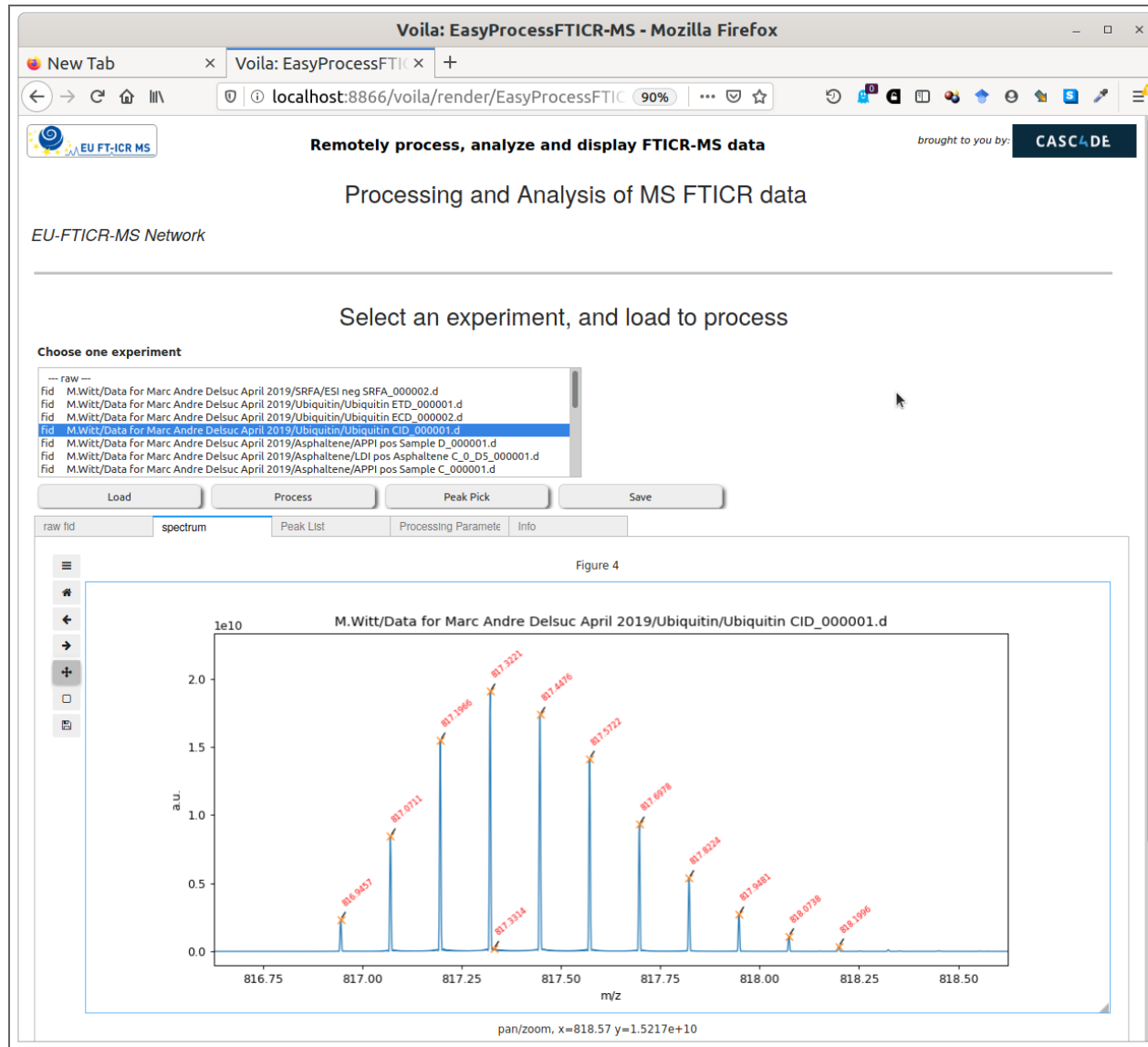
On-line tool

developed for
EU project
used internally
in the network so far

all data-sets accessible

integrated environment

presentation



On-line tool
developed for

EU project
used internally
in the network so far

all data-sets accessible

integrated environment

detailed parameters

MS cho.d/MS_Spectrum_extracted_from_4.00_to_4.00_minute.msh5
MS 2319_1/Processed_01.msh5
MS 2319_1/Processed_1.msh5
MS 2319_1/Processed.msh5
MS 2319_1/Processed_02.msh5
MS 2319_1/Processed_03.msh5

Load

Process

Peak Pick

Save

raw fid

Spectrum

Peak List

Calibration

Simulation

Processing Parameters

Info

Processing

center fid ☒ Yes
☐ No

apod todo

hamming

try **kaiser3.5** for best possible resolution and SNR at the price of wiggles at the feet of large peaks

try **kaiser8** for very low wiggles, for resolution and SNR close to Hanning

zf level 2

zf=2 means final size is doubled

baseline todo

Offset correction

grass noise todo

Only when storing output file (smaller files)

grass noise level 3.00

grass noise consists in setting to zero datapoints below the *grass noise level*

Peak Picking

peakpicking todo

Manually on-demand

peakpicking noise level 10.0

peaks above this level will be detected

centroid ☒ Yes
☐ No

max peak displayed

200

used to reduce display burden

interface kept
simple
optimised
data processing
auto adapt
to dataset type

various possibilities

raw fid	Spectrum	Peak List	Calibration	Simulation	Processing Paramete	Info
---------	----------	-----------	-------------	------------	---------------------	------

Enter formula: ? isotopic

Draw isotopic patt... scaling charge RP k Draw

C₁₁₇H₂₀₀N₄₂O₃₁ monoisotopic mass: 2689.536468 average mass: 2691.107

2689.536468	68.7904542228
2690.539209	100.0000000000
2691.541874	76.5765037769
2692.544486	40.8797547432
2693.547055	17.0202532342
2694.549592	5.8694815952
2695.552102	1.7402145423
2696.554589	0.4549160745
2697.557058	0.1067642088
2698.559508	0.0227945309
2699.561935	0.0044670225
2700.564261	0.0008021174
2701.566610	0.0001338547

peak assignment

raw fid	Spectrum	Peak List	Calibration	Simulation	Processing Paramete	Info
---------	----------	-----------	-------------	------------	---------------------	------

Calibration parameters

The equation used for calibration is unique, it is as follows:

$$freq = A / (m/z) - B + C / (m/z)^2$$

so there is the following correspondance between A, B and C and the Boker parameters ML1 ML2 ML3 :

- A = ML1
- B = ML2 if ML3 is null
- B = -ML2 if ML3 is not null (SPIKE does not follow the Bruker convention to inverse the role of B depending on the calibration equation used)
- C = ML3

A B C values hold the calibration parameters, you can change them manually, or use the a reference peak (right)

Then use the "Update" button to propagate the values to the dataset in memory.

A	<input type="text" value="108327023.2844505"/>	Hz/Th	reference peak	
B	<input type="text" value="19.836843680681607"/>	Hz	Observed	<input type="text" value="539.1213"/> m/z
C	<input type="text" value="7489.927504388354"/>	Hz/Th²	Theoretical	<input type="text" value="539.1104"/> m/z

recalibration

audit-trail

All actions to the data are logged

raw fid	Spectrum	Peak List	Calibration	Simulation	Processing Par	Info
<h2>Raw Dataset</h2> <p>FTICR_DATA/MvA/Phase/histonepeptide_ms2_000002.d</p> <p>FTICR data-set Bo: 7.05 Single Spectrum data-set FT-ICR axis at 535.714286 kHz, 524288 real points, from physical mz = 202.203 to m/z = 1450.000 R max (M=400) = 265036</p> <h2>Audit-Trail</h2> <hr/> <h2>MS processing</h2> <ul style="list-style-type: none">date: Mon, 12 Dec 2022 08:45:45 CET <h2>Processing conditions</h2> <ul style="list-style-type: none">working directory: /home/mad/Documents/ mad/CASC4DE/CODESSPIKE kernel version: 0.99.30 rev 563 dated 22-03-2022 <h2>FTICR data-set</h2> <p>Bo: 7.05 Single Spectrum data-set FT-ICR axis at 535.714286 kHz, 524288 real points, from physical mz = 202.203 to m/z = 1450.000 R max (M=400) = 265036</p> <h2>Pre processing phase</h2> <h3>stat analysis of FID</h3>						

- offset: -6.301655572297774
- noise: 47039.92696468851

vertical shift of FID

- center_fid: Yes
- offset: 6.301655572297774

Spectral Analysis phase

FID apodisation before FT

- apod_todo: kaiser, beta=3.5

Fourier Transform

- initial size: 524288
- zf_level: 2
- Fourier_algo: rfft
- final size: 1048576

Post processing phase

Baseline correction

- mean offset: 1077806.4713130128

final

- modulus: applied
 - unit: m/z
 - calibration - A: 108327023.2844505
 - calibration - B: 19.836843680681607
 - calibration - C: 7489.927504388354
 - spectrum final size: 524288
- SHA256: 0558b4161e58c31d2c928f2e9bfec418aa5868cc938272aa24bc22bfd2580d6

MS post-processing

- date: **Mon, 12 Dec 2022 08:45:57 CET**

Processing conditions

- working directory: **/home/mad/Documents/ mad/CASC4DE/CODES**
- SPIKE kernel version: **0.99.30** rev **563** dated 22-03-2022

FTICR data-set

Bo: 7.05 Single Spectrum data-set FT-ICR axis at 535.714286 kHz, 524288 real points, from physical m/z = 202.203 to m/z = 1450.000 R max (M=400) = 265036

MS post-processing

- date: **Mon, 12 Dec 2022 08:46:07 CET**

Processing conditions

- working directory: **/home/mad/Documents/ mad/CASC4DE/CODES**
- SPIKE kernel version: **0.99.30** rev **563** dated 22-03-2022

FTICR data-set

Bo: 7.05 Single Spectrum data-set FT-ICR axis at 535.714286 kHz, 524288 real points, from physical m/z = 202.203 to m/z = 1450.000 R max (M=400) = 265036

Peak-Picking phase

Peak-Picking

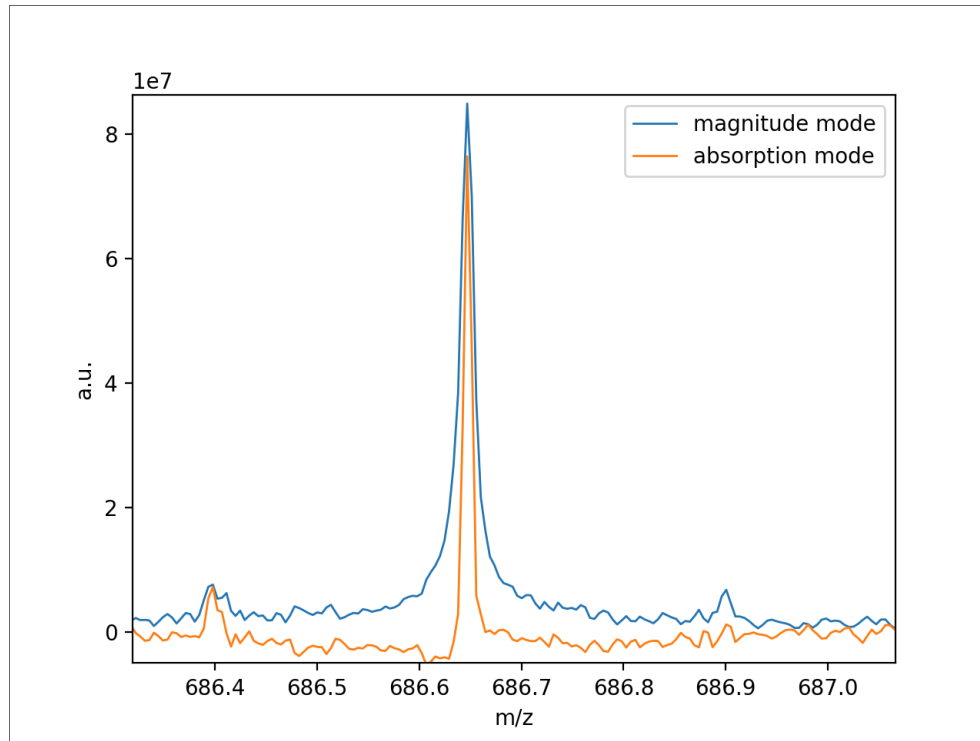
- threshold above noise level: 50.0
- zoom: None
- number of detected peaks: 890
- output file:: histonepeptide_ms2_000002.d/peaklist_04.csv
SHA256: 3fa94ee05bbcc02f2387de0b27ad7dd087220f254a3dd4f5e78580c4b054f5c7

The log file is signed, and can be exported.

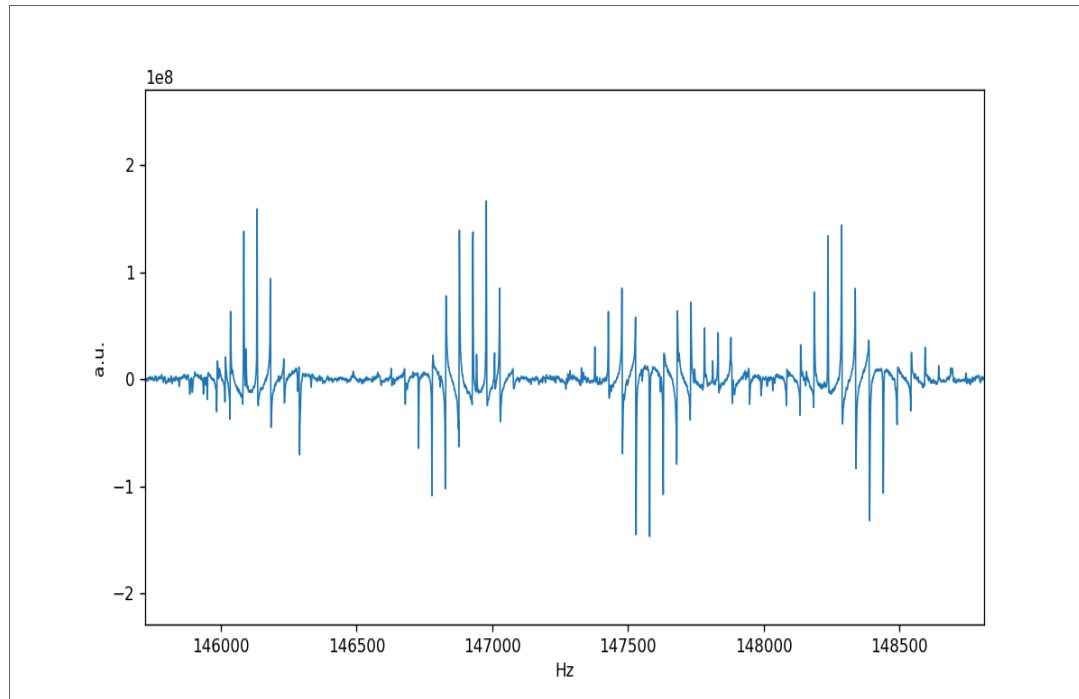
phase sensitive processing

Spectra, presented in “Absorption” mode bring:

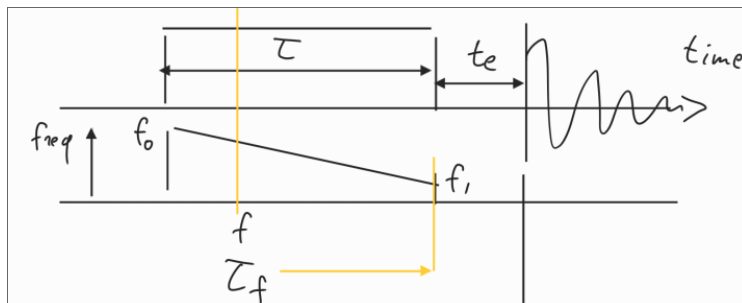
- better precision
- better Signal-to-Noise



However, they need to be “Phased”

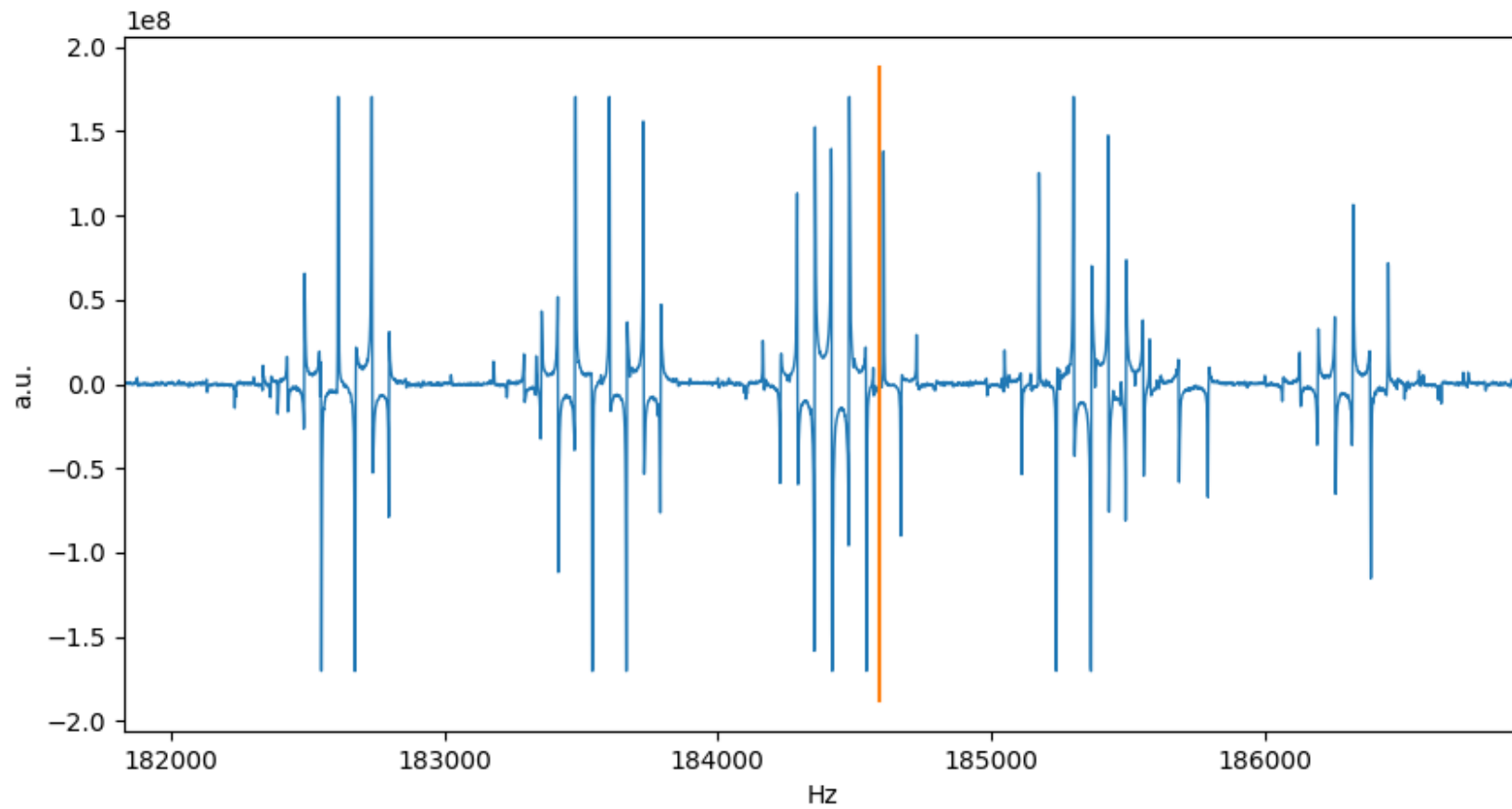


zoom on a small region
Note display in frequency

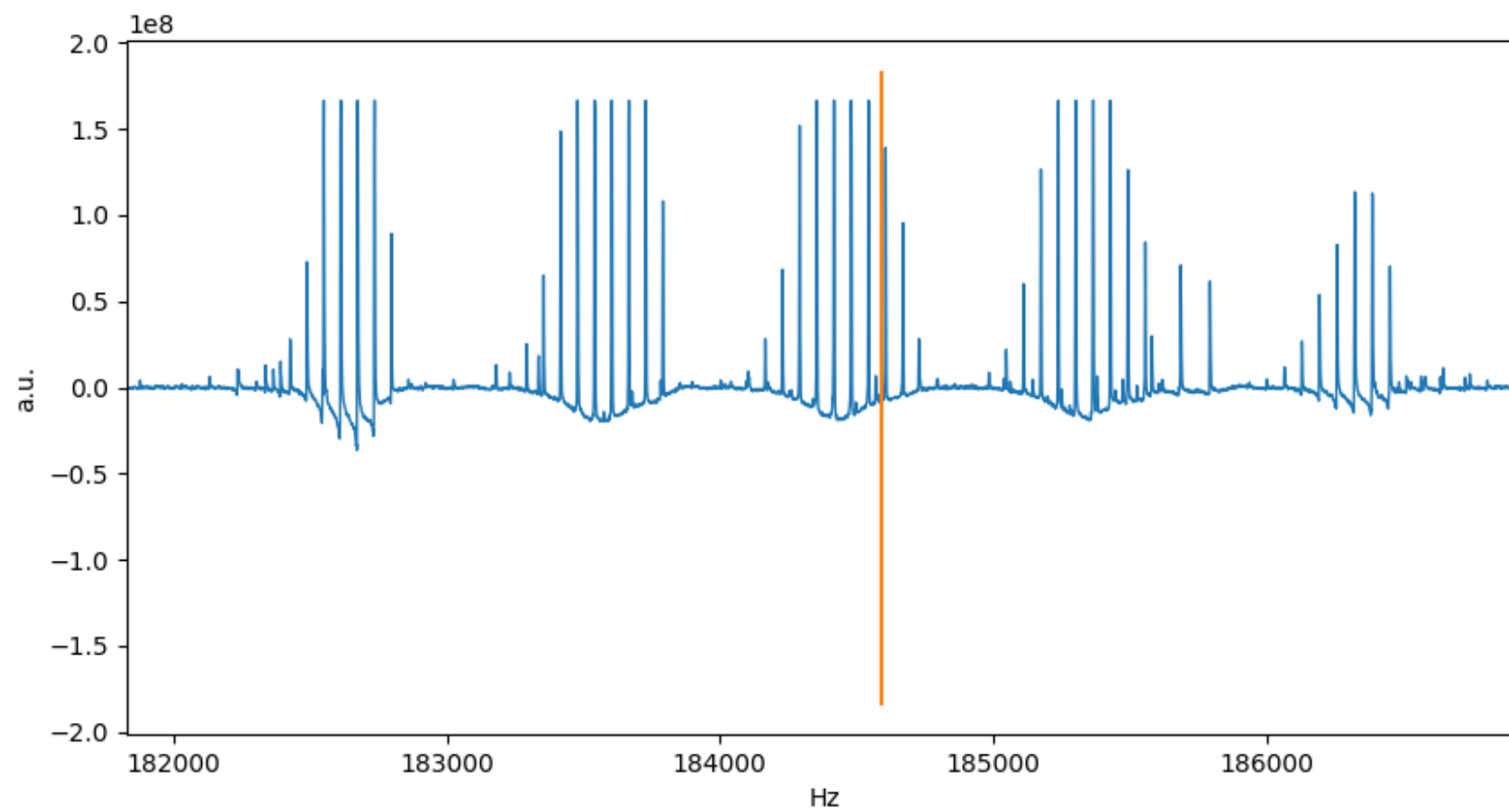


Schematic of the excitation pulse

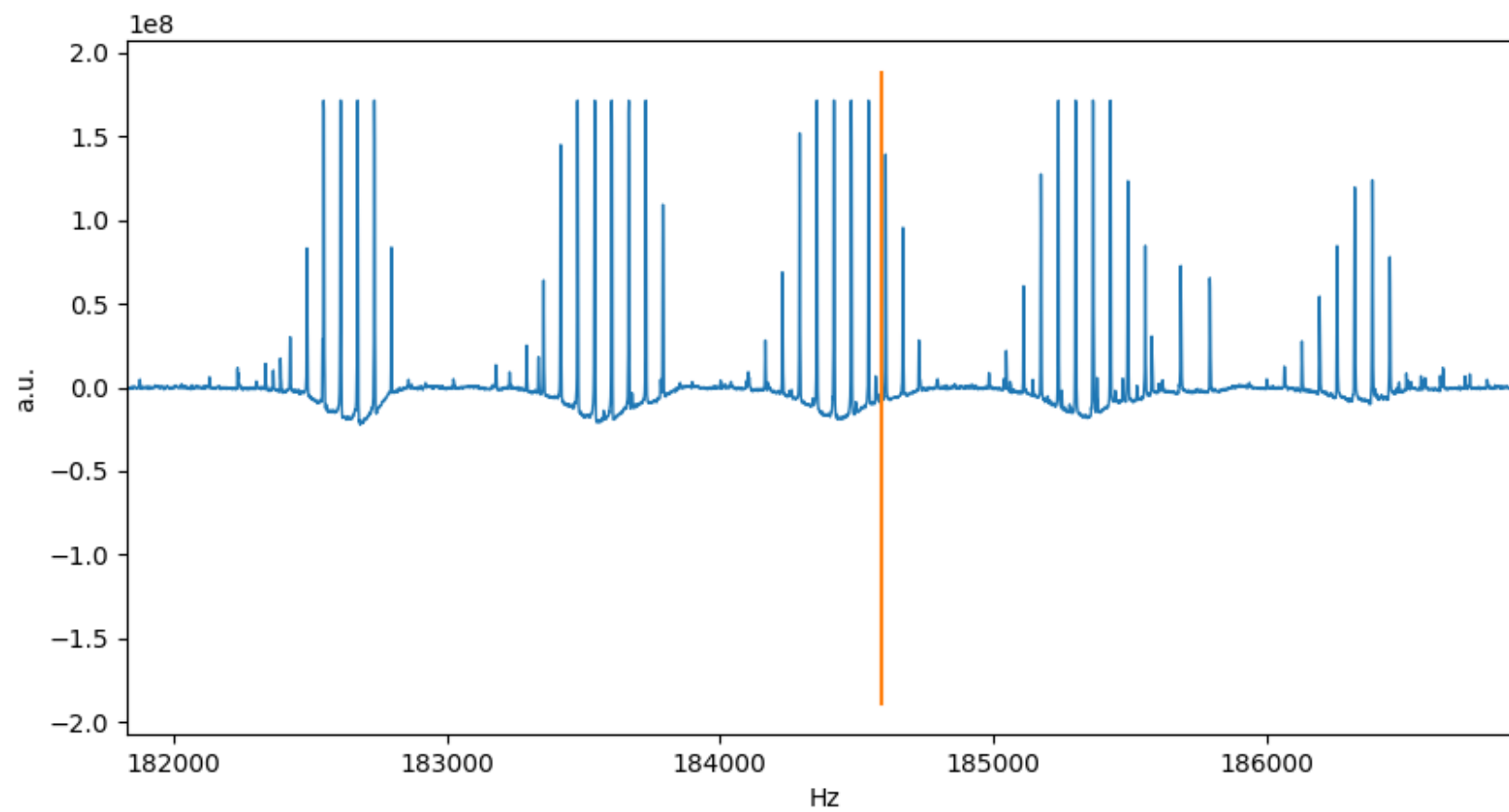
Phasing procedure



1/ first select a zone (here around 184kHz : m/z 588)

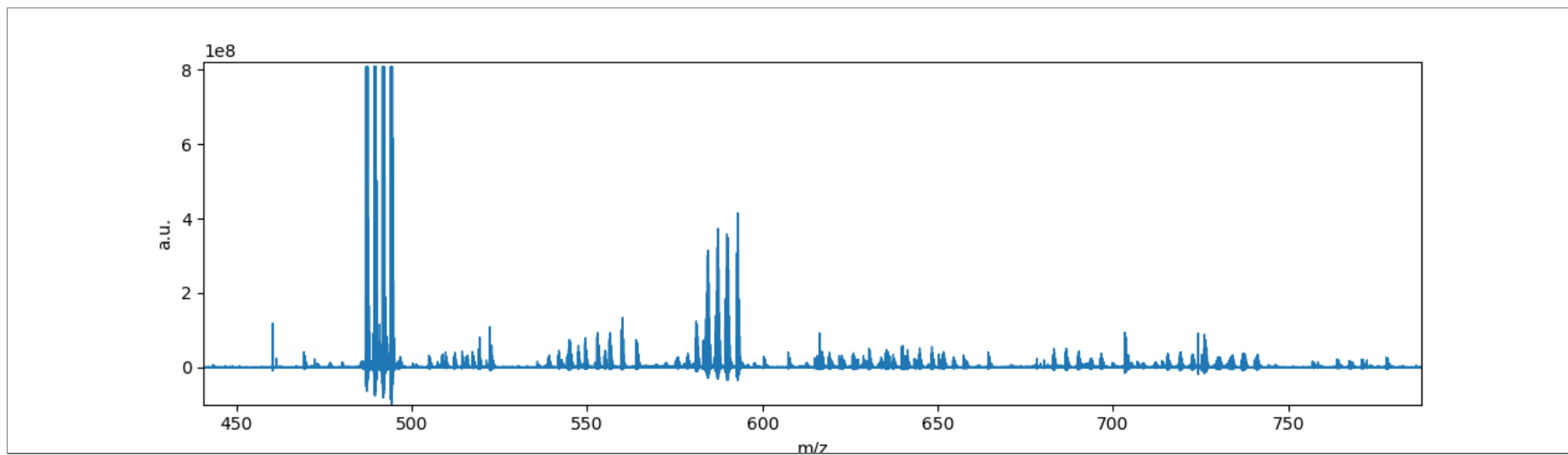


2/ unroll linear dependence



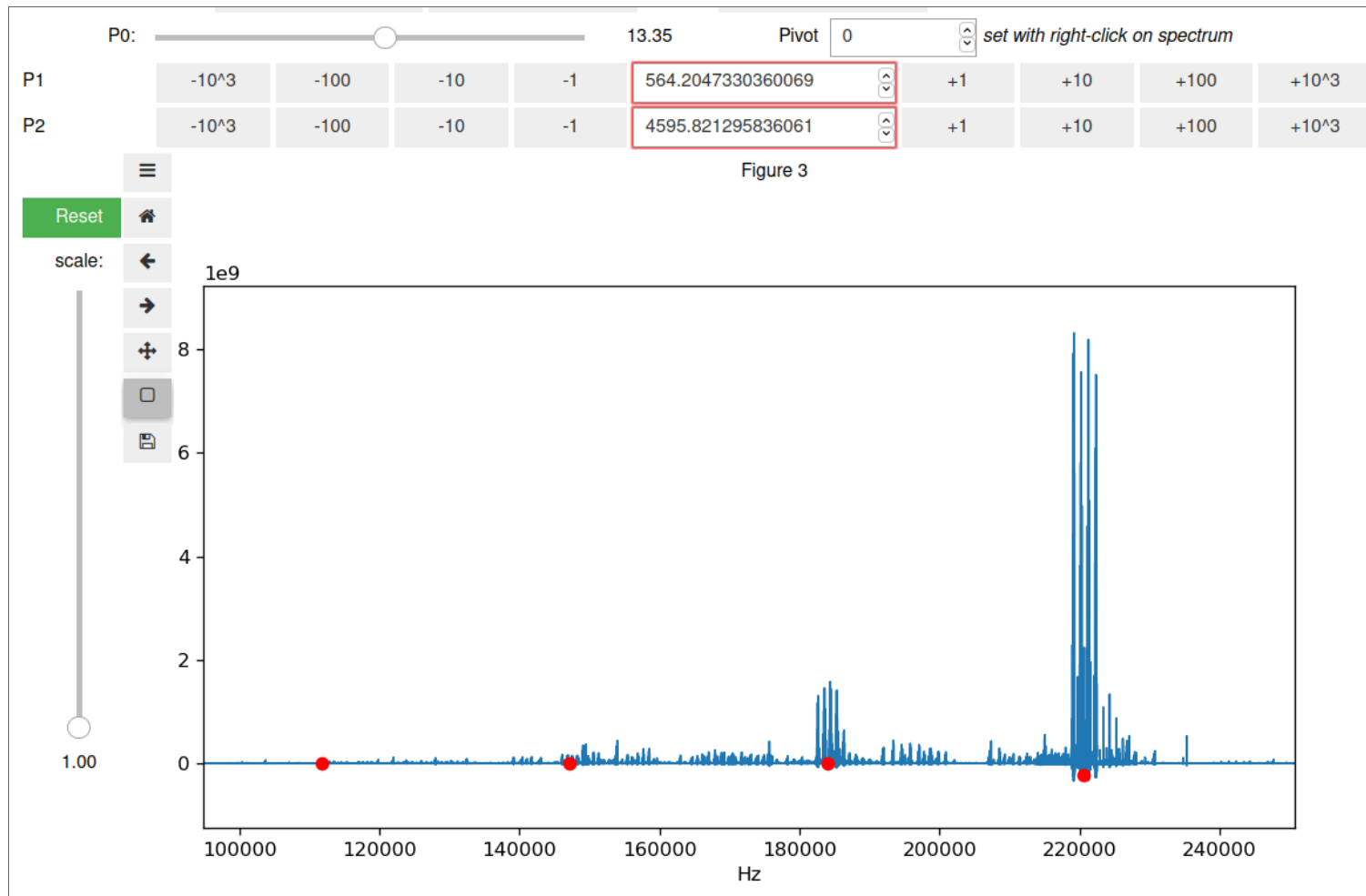
3/ unroll quadratic dependence

Phased Broadband spectrum

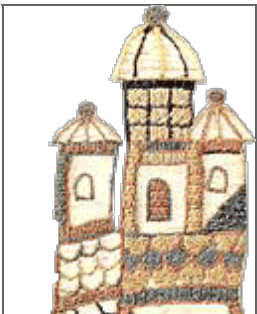
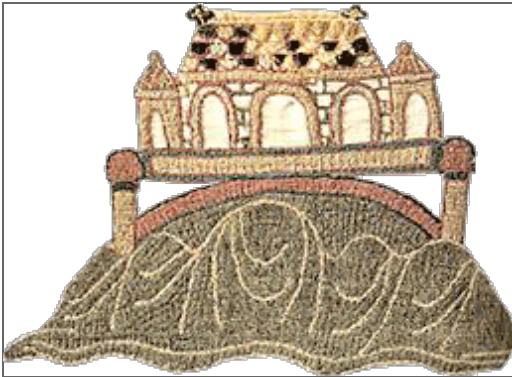
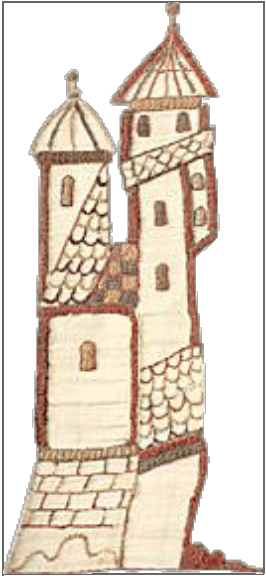


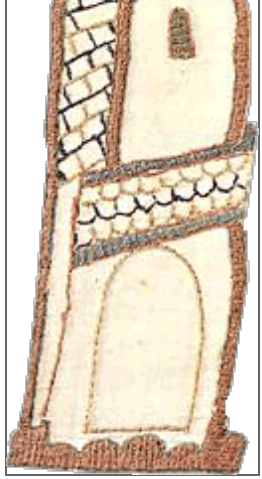
Phased Broadband spectrum

Phasing helper tool

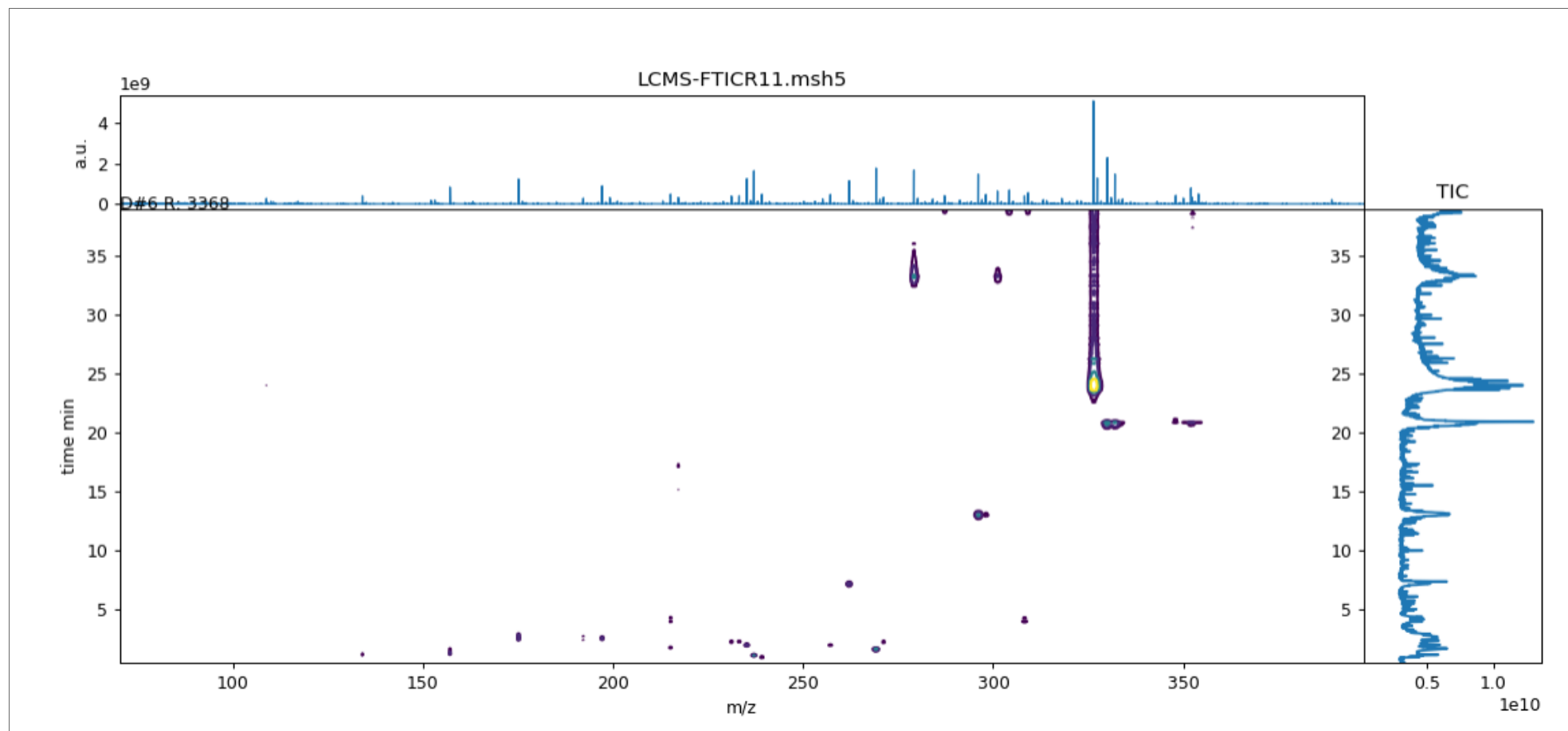


Larger datasets

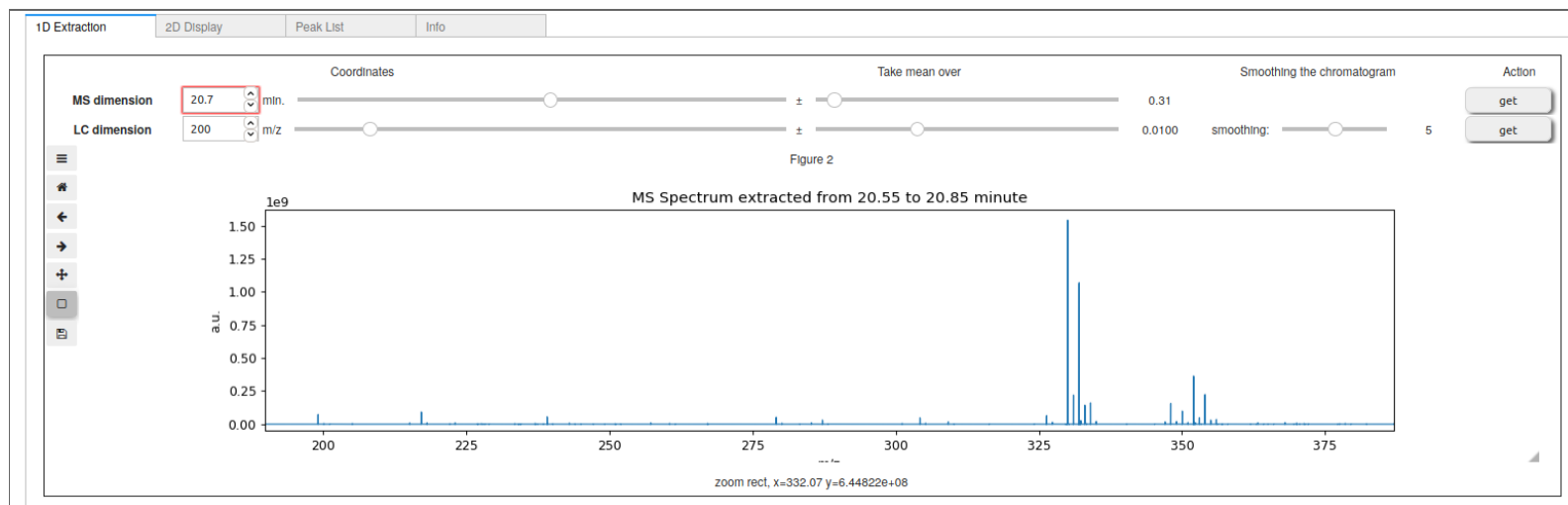




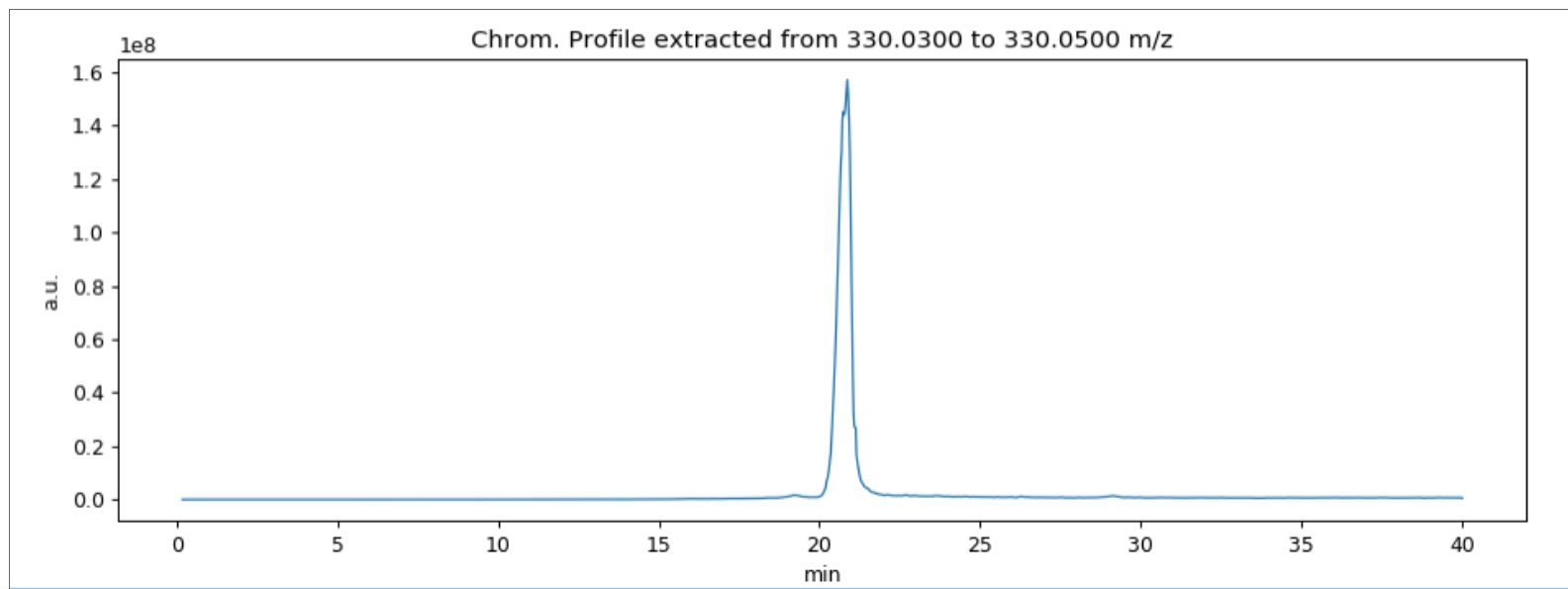
LC-MS



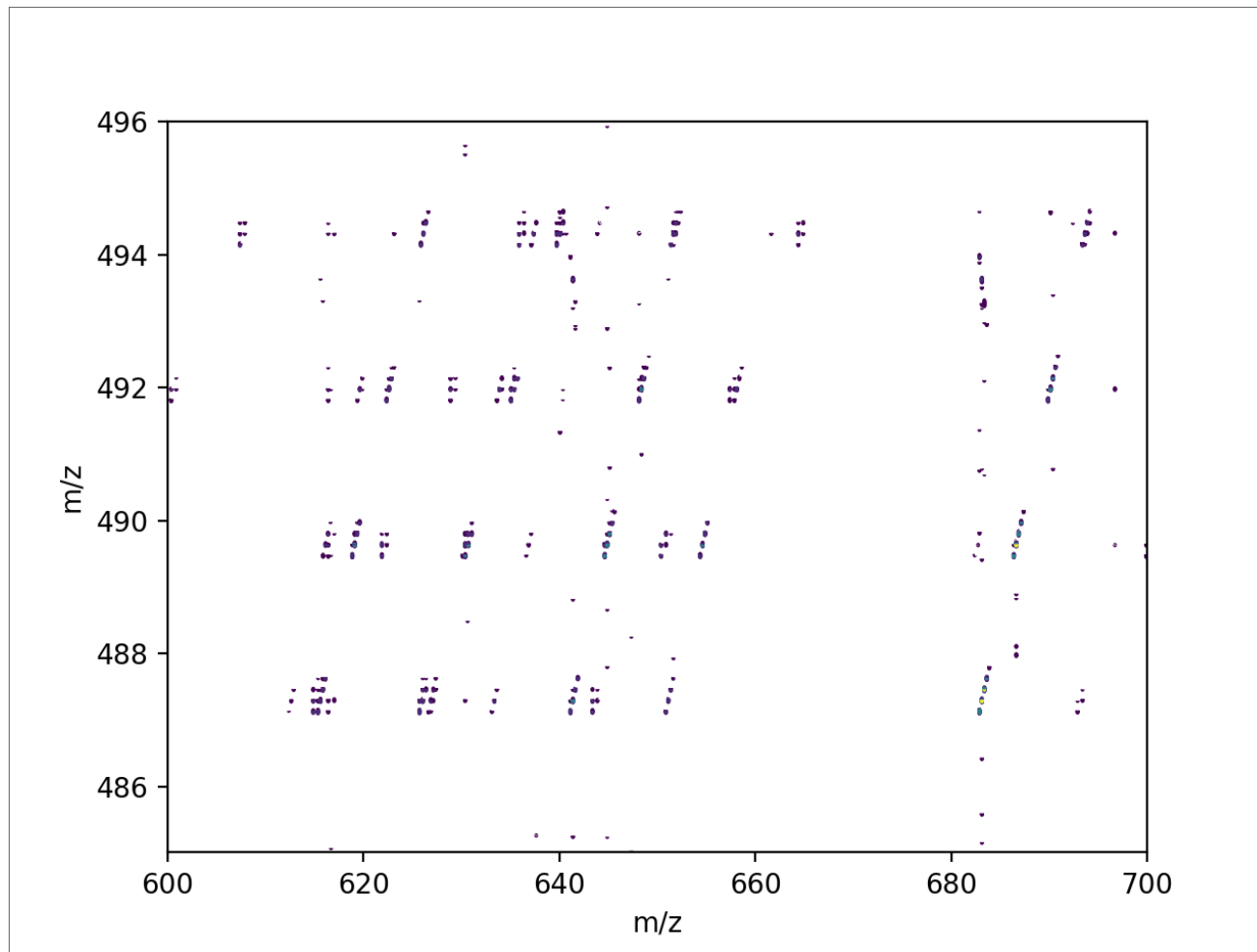
LC-FTICR-MS



extracting slices

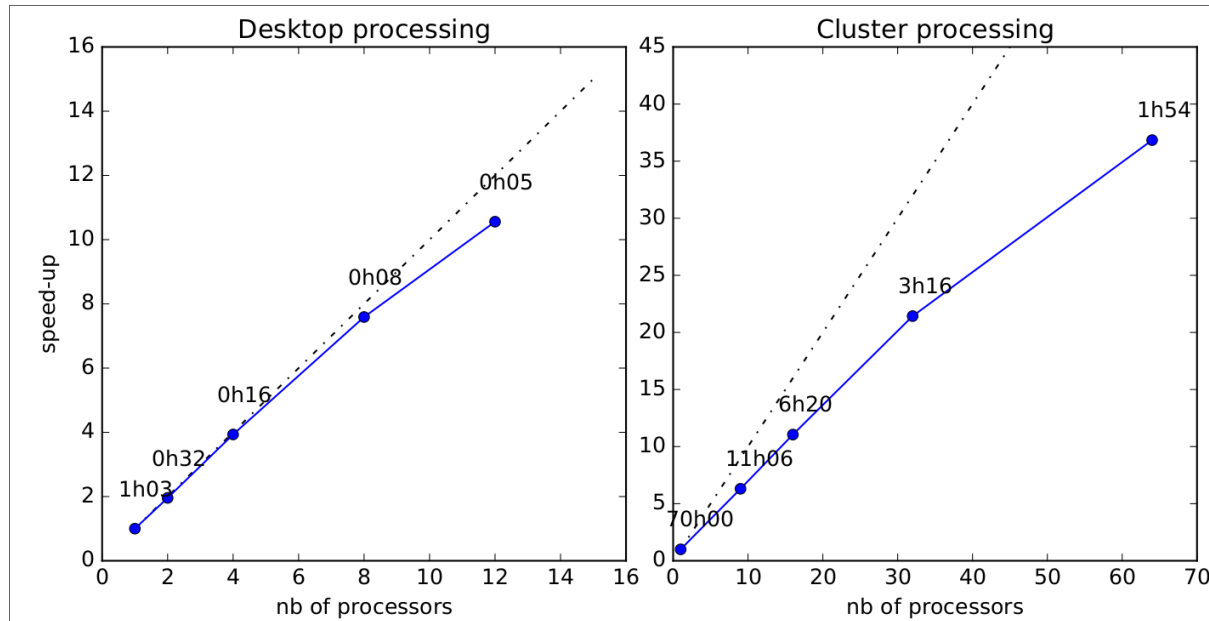


2D FTICR-MS



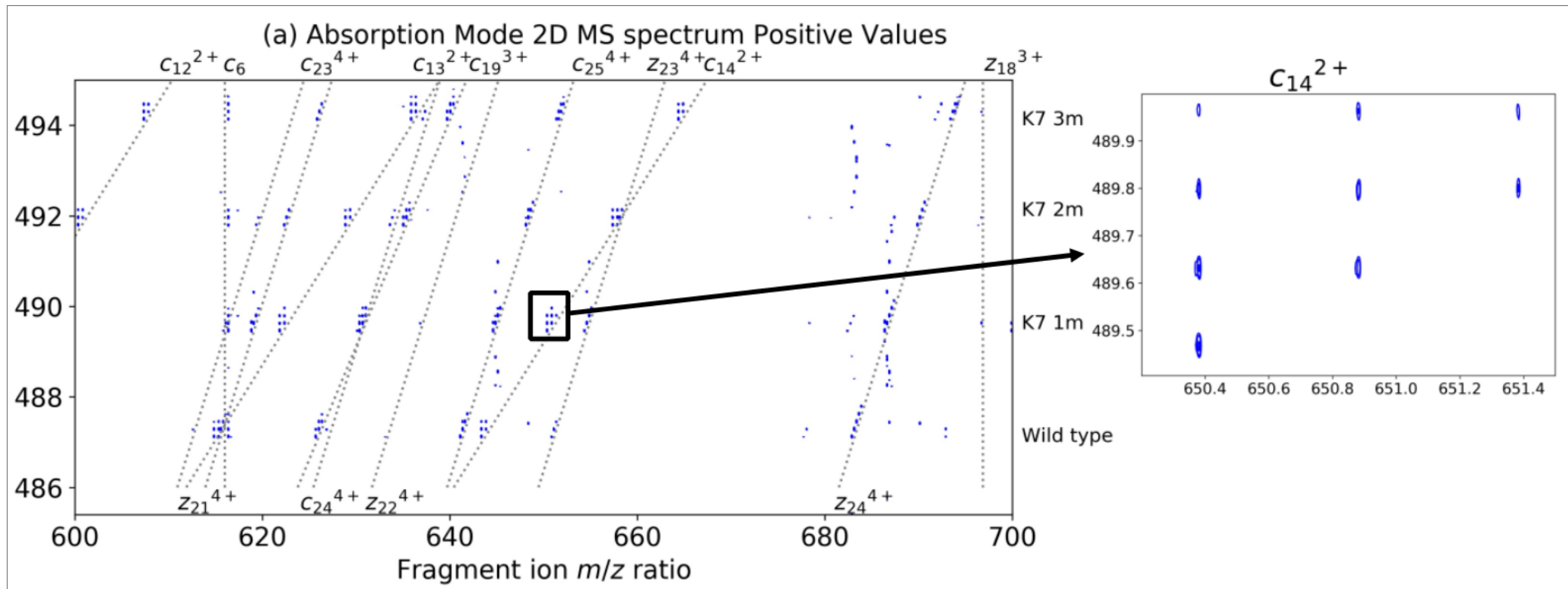
Several Challenges

- large File (10 .. 100GB)
- Large Processing



- DeNoising is usually required
 - urQRd : fast and efficient

Phasing 2D FTICR-MS



4MS

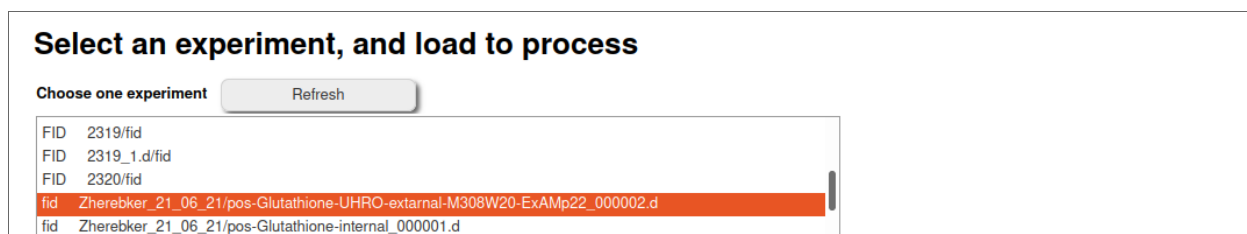
New Development

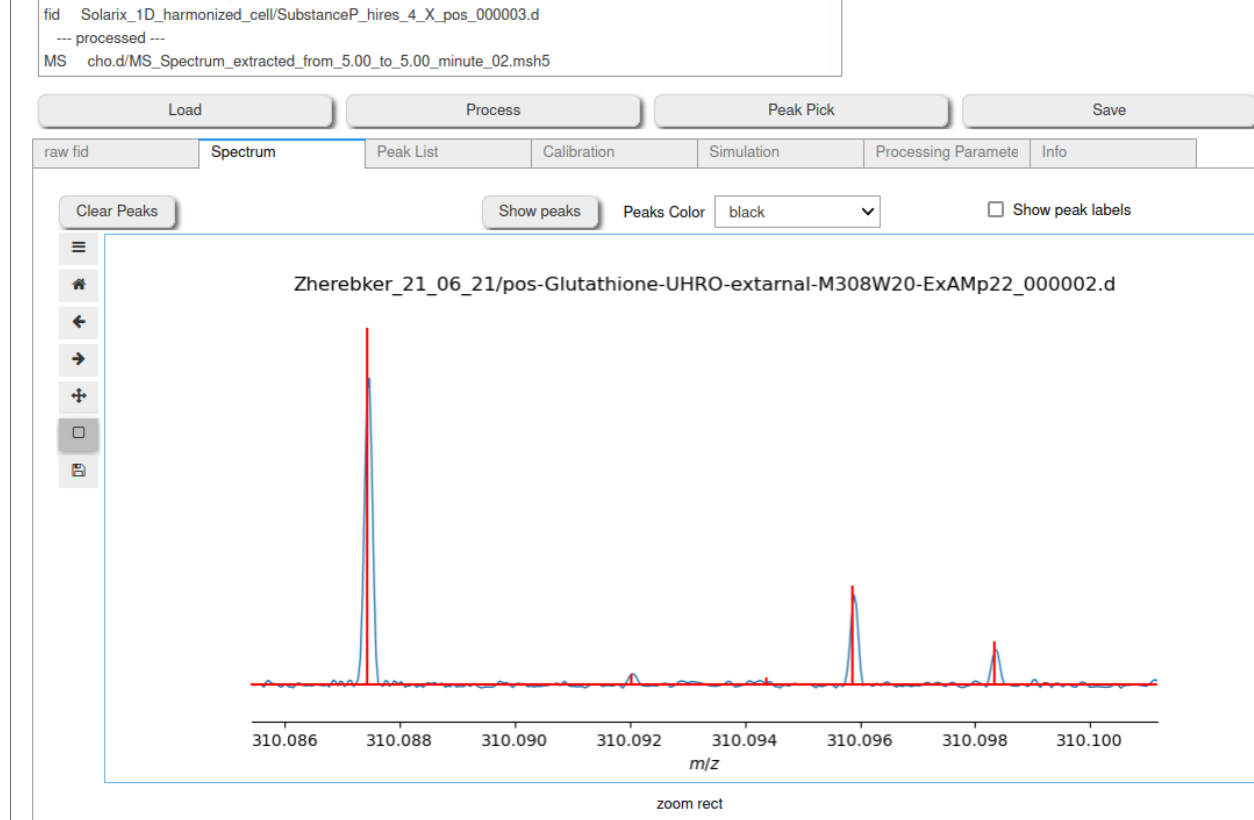
- by



- open-source scientific lib.
 - SPIKE
 - isotopes
 - neutronstar
 - rocospin
- technical lib.
 - tornado
 - ipywidget
 - voilà

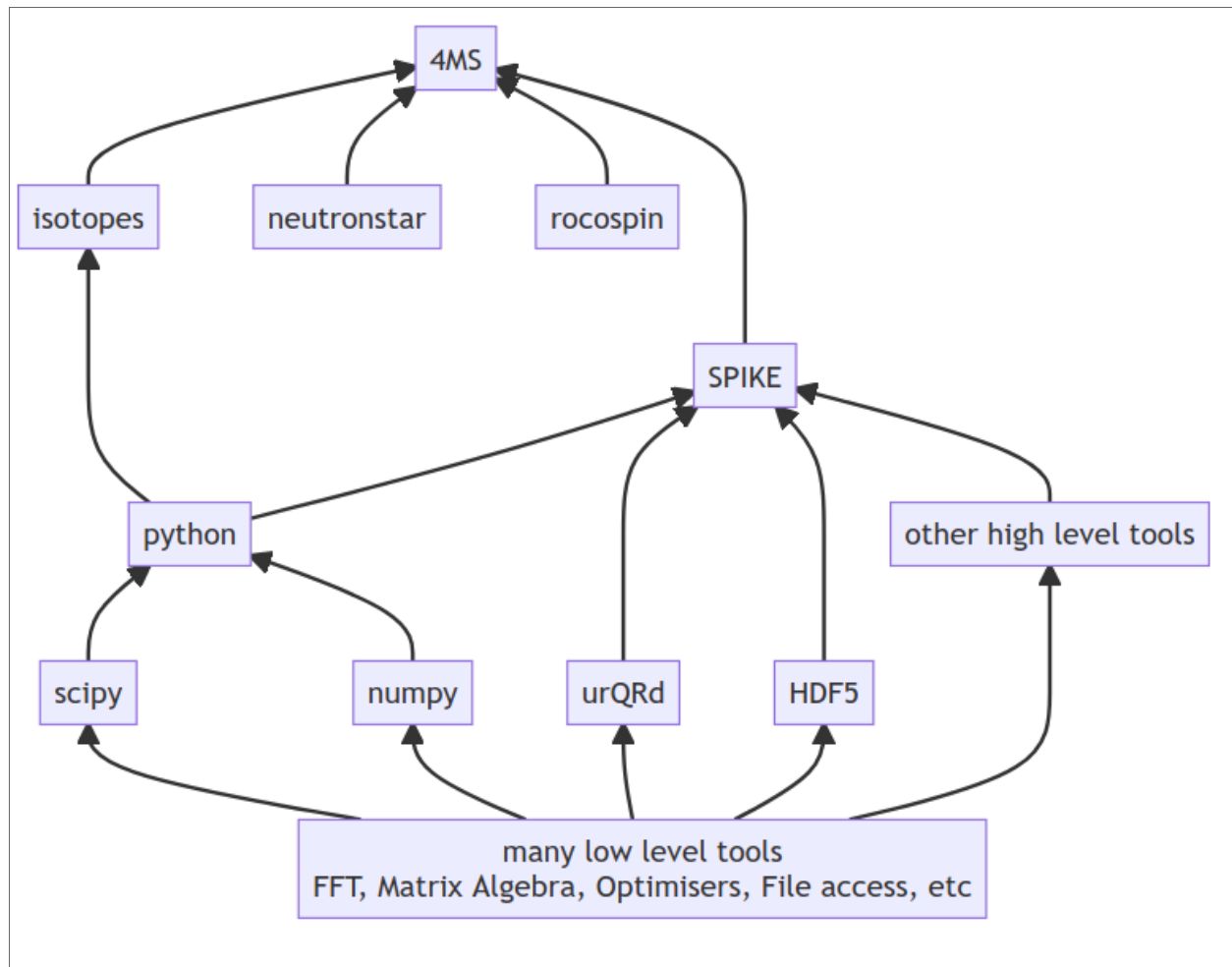
⇒ Available 2023





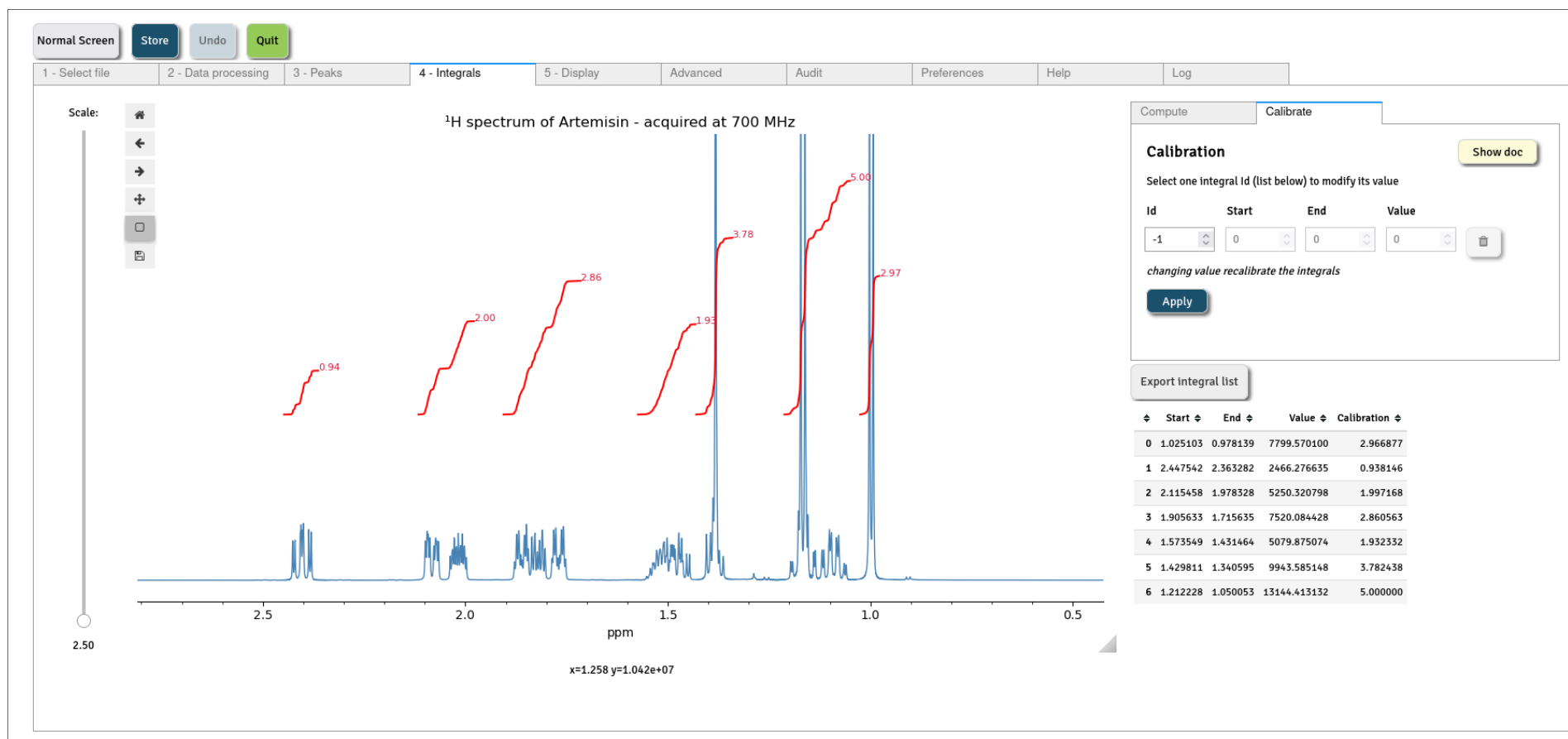
4MS with fine isotopic pattern analysis

stack of Open-Source software



4NMR

Same development, available soon !



Acknowledgments



work of many

DEVELOPMENTS

- B.Kieffer
- *all open-source authors !*

2D FTICR-MS

- C.Rolando
- G.Bodenhausen
- P.O'Connor
- M.van Agthoven
- F.Bray
- ...

CASC4DE

- Early
 - M-A.Coutouly
 - J-P Starck
 - L.Chiron
 - J.Asencio
 - C.Marin
- Data Mining / Software
 - L.Duciel
 - L.Baptista
 - A.Briot-Dietsch



source: **"Historic Tale Construction Kit"** github.com/htck
presentation done with **Quarto / Revealjs**