What CcpNmr Analysis can do for and with small molecules

Vicky Higman

University of Leicester

UNIVERSITY OF LEICESTER

LISCB

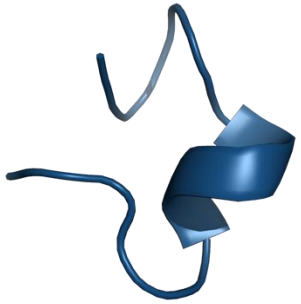
Medical Research Council

Ralph:

I get the feeling that most of our NMR managers either use [CCPN's software], are scared to try it, or don't know anything about it and are too busy or set in their ways to try it.

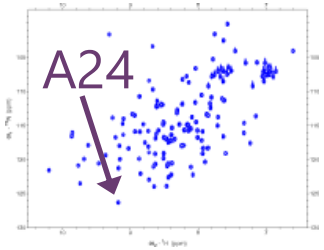
Biomolecular NMR

Proteins



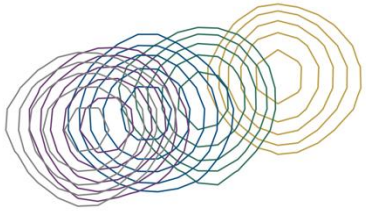
Peptides

RNA/DNA

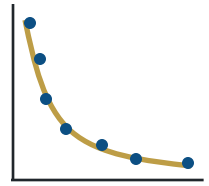


Spectrum Visualisation and Assignment

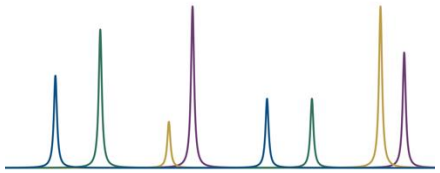
Structure Calculation



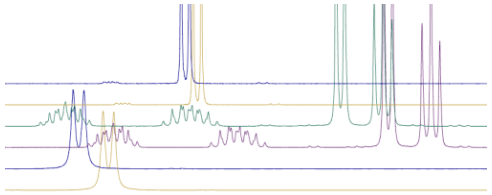
Binding Studies



Dynamics



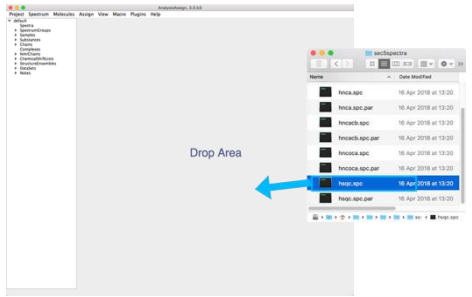
Screening



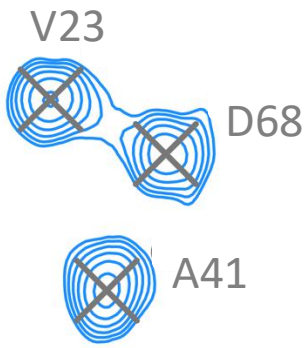
Metabolomics

Our Approach

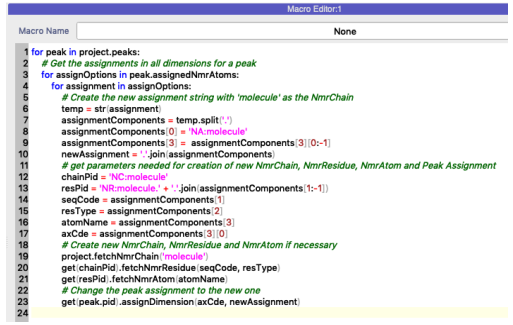
Modern



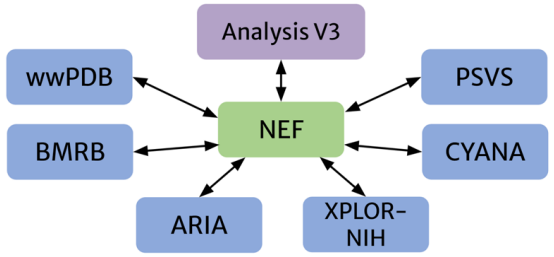
Flexible



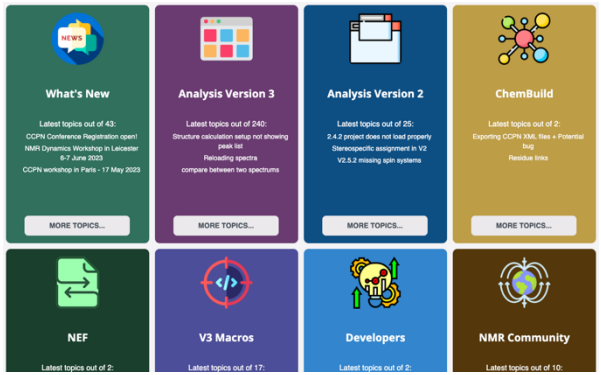
Customisable



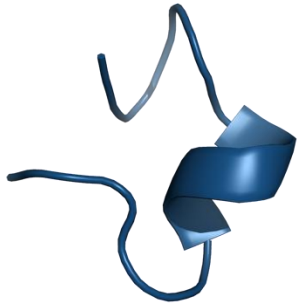
Integrated



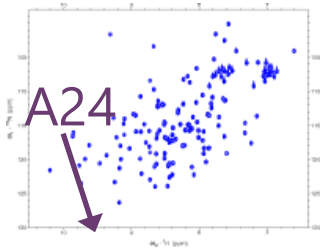
Supported



Small Molecule NMR



Peptides



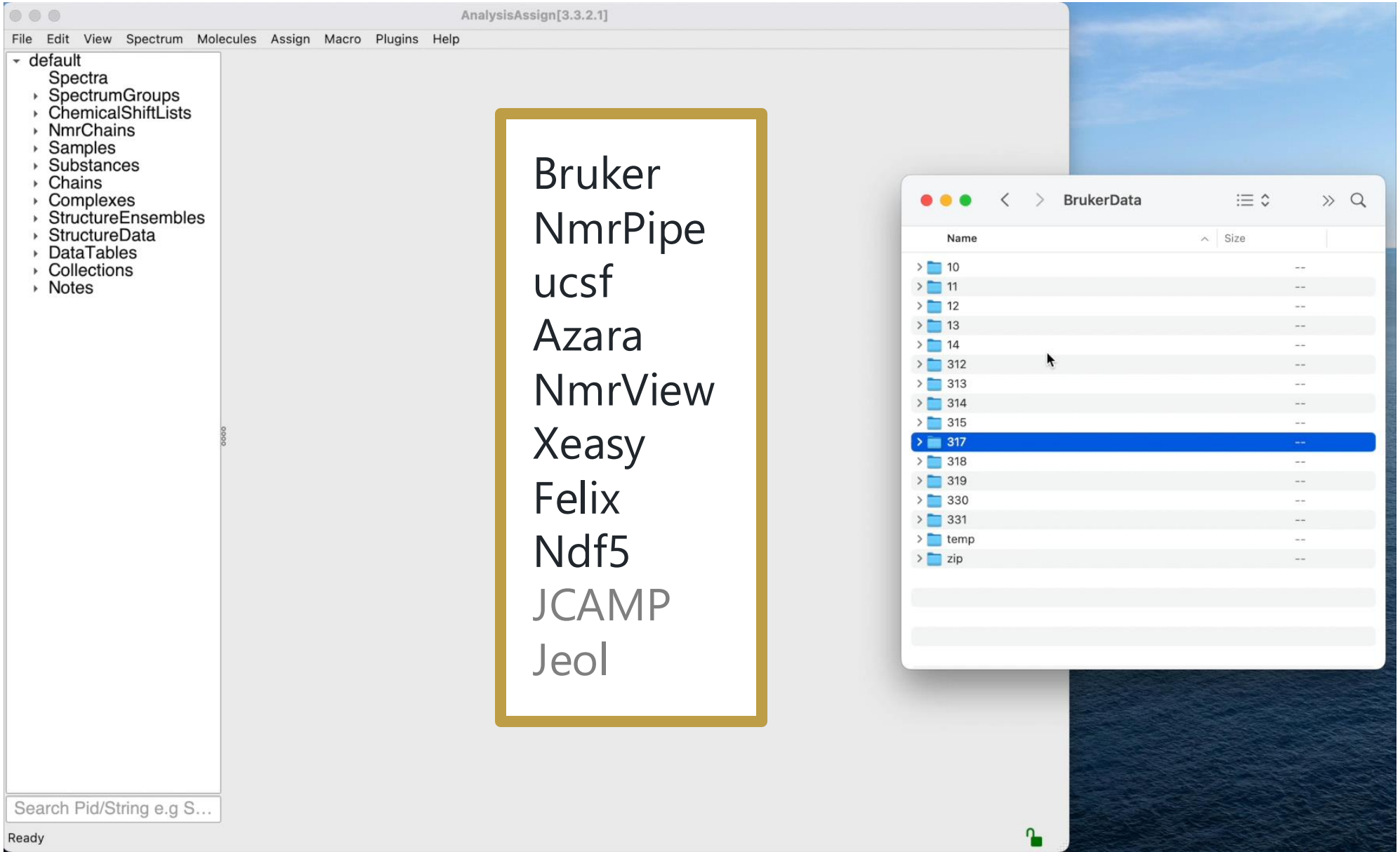
Spectrum Visualisation and Navigation

Screening

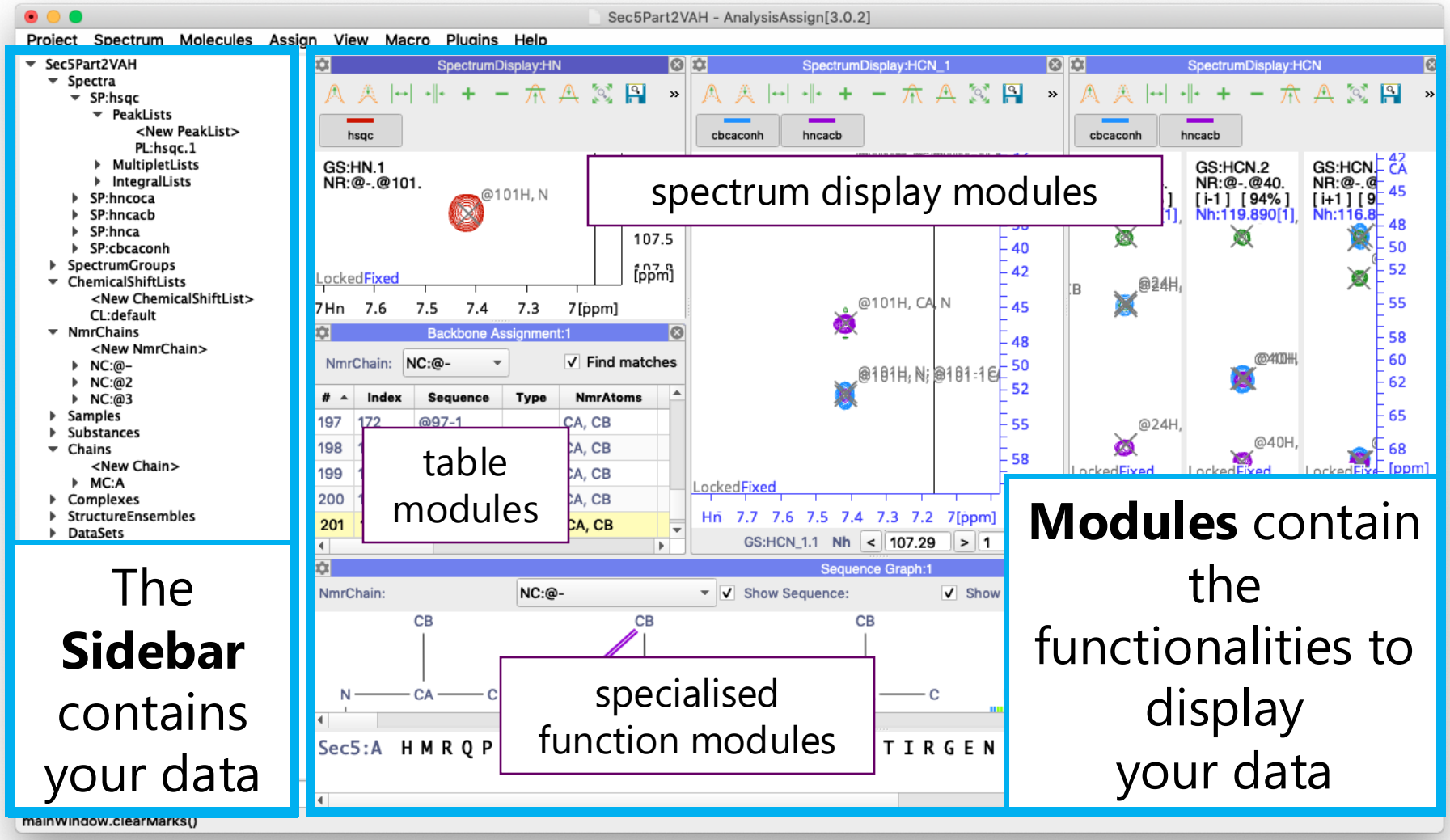
Metabolomics

Macro Writing

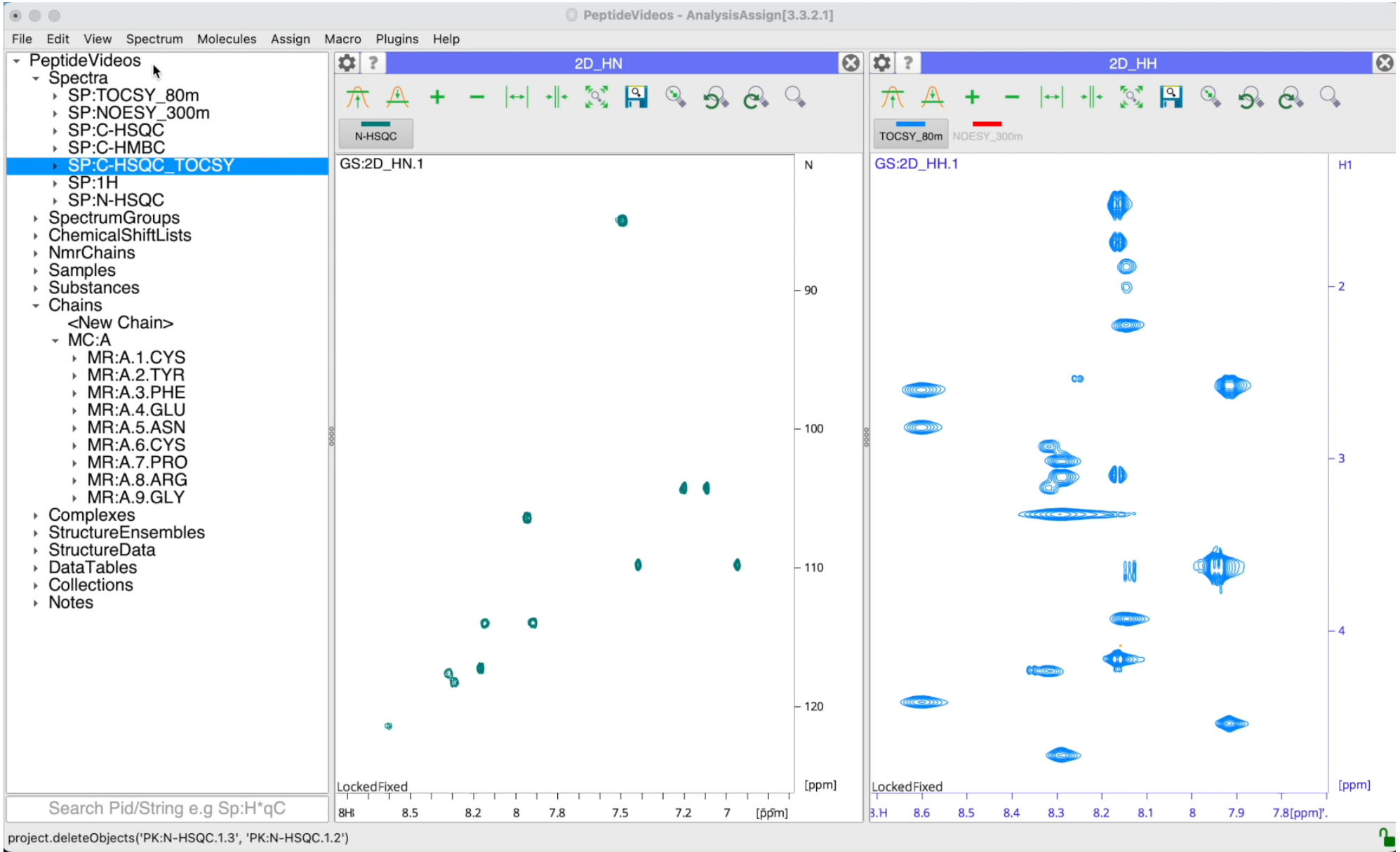
Opening and Overlaying Spectra



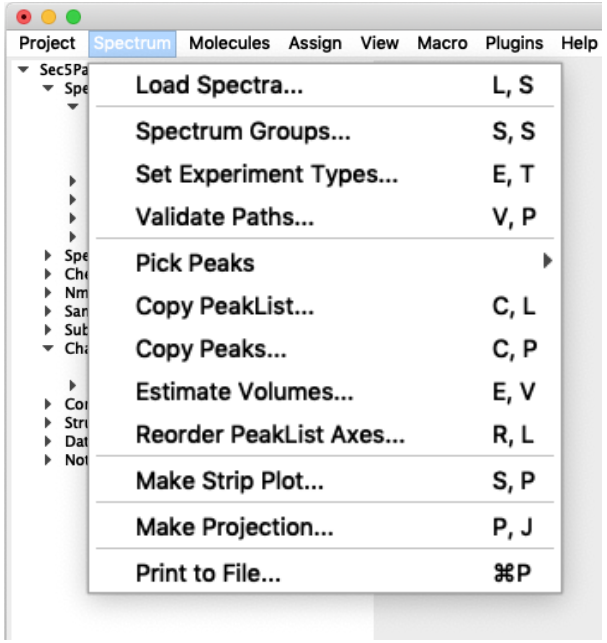
Programme Layout



Assignment: Marking and Peak Picking



Keyboard Shortcuts



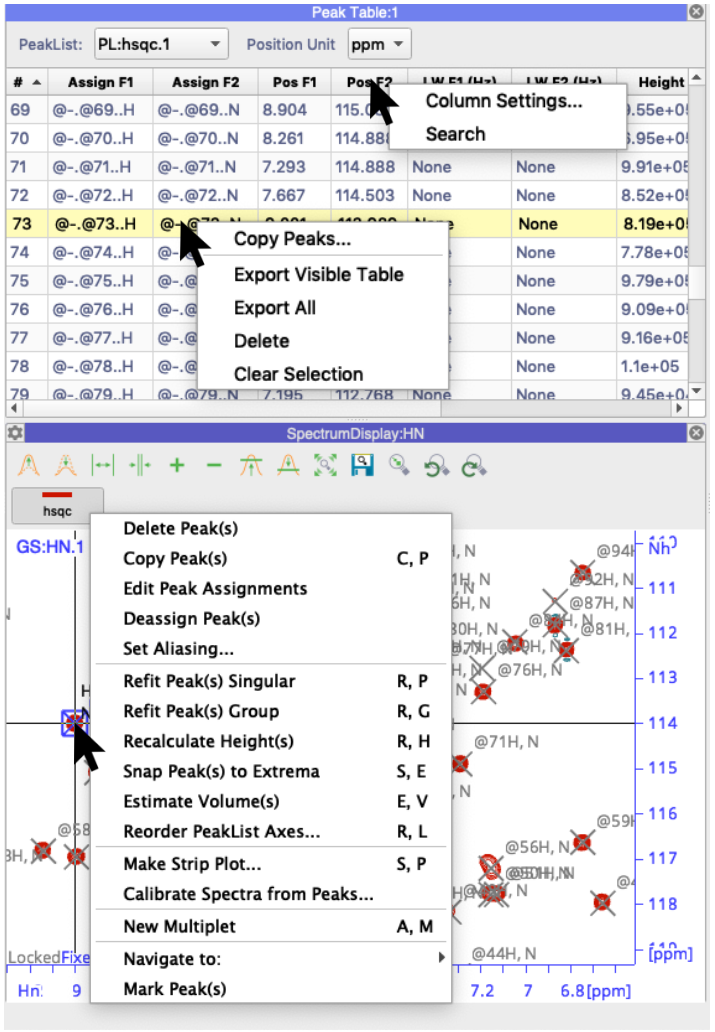
Two-key shortcuts are shown in the menus

They are case insensitive (e.g. type ‘cp’)

Press **Esc** to reset

A full list is available from **Help / Show Shortcuts**

Mouse Menus



Mouse Menus are context-dependent

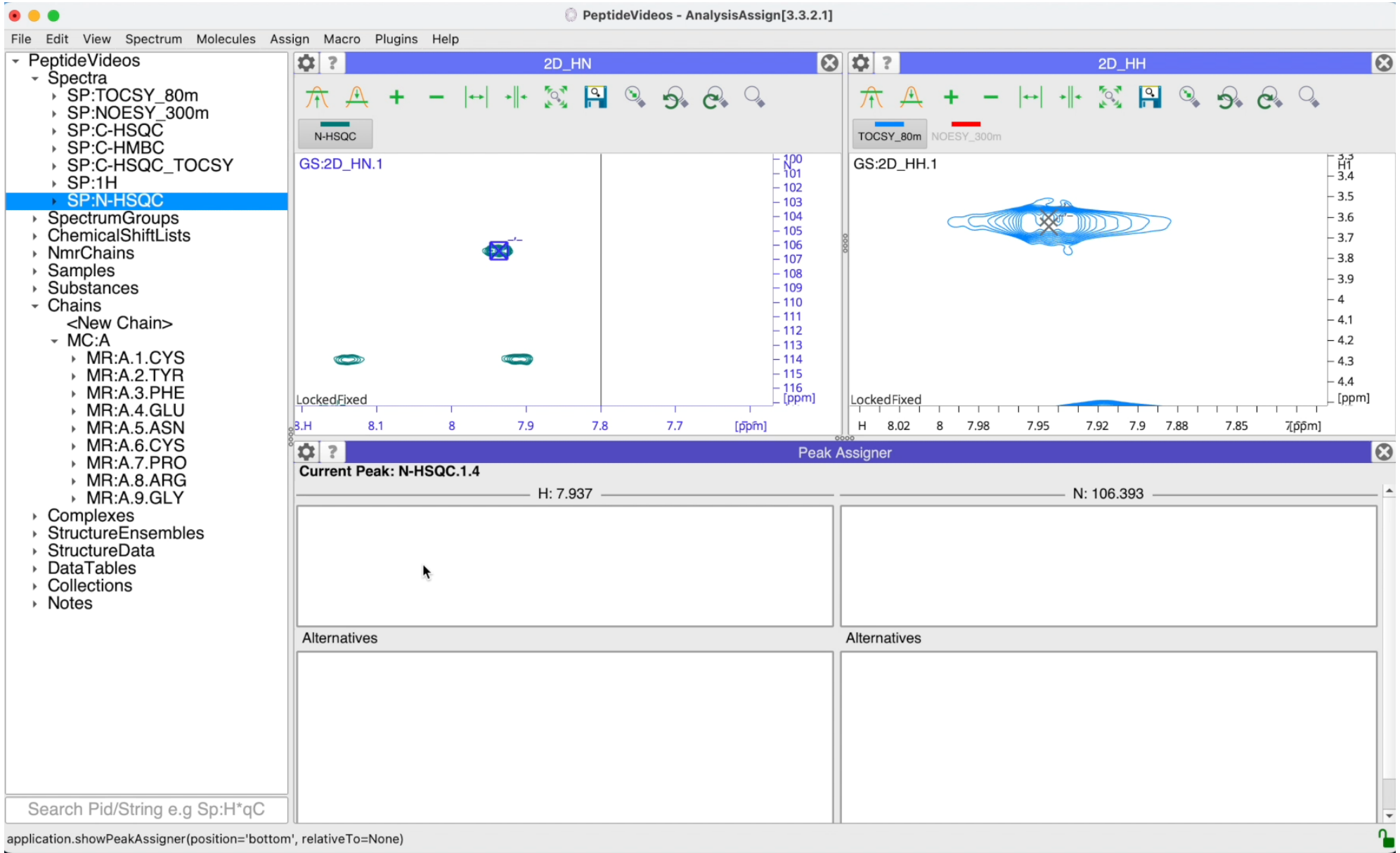
Peak Labels / NmrAtoms

Peak Labels are divided into four parts:

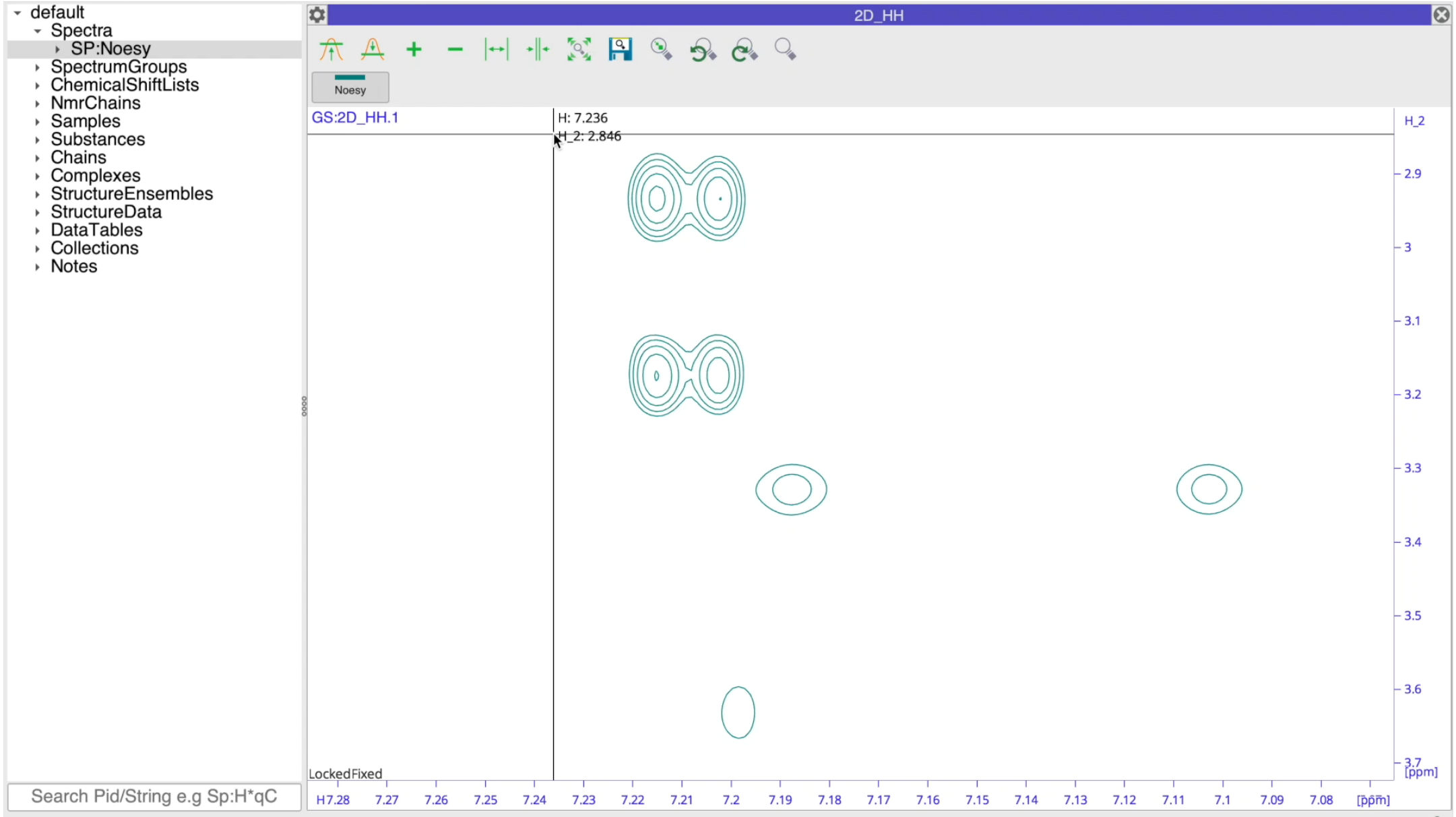
NmrChain	NmrResidue	NmrAtom
A	A.57.TRP	A.57.TRP.H

- NmrChains
  - NC:A
    - NR:A.57.TRP
      - NA:A.57.TRP.H
      - NA:A.57.TRP.N
      - NA:A.57.TRP.CA
      - NA:A.57.TRP.CB

Assignment: Peak Labelling



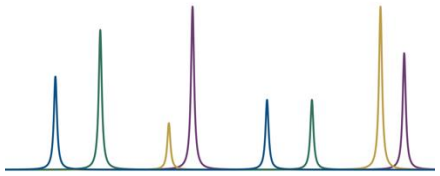
Multiplets


default
Spectra
SP:Noesy
SpectrumGroups
ChemicalShiftLists
NmrChains
Samples
Substances
Chains
Complexes
StructureEnsembles
StructureData
DataTables
Collections
Notes
Search Pid/String e.g Sp:H*qC
2D_HH
Noesy
GS:2D_HH.1
H: 7.236
H_2: 2.846
H_2
2.9
3
3.1
3.2
3.3
3.4
3.5
3.6
3.7
[ppm]
LockedFixed
H7.28
7.27
7.26
7.25
7.24
7.23
7.22
7.21
7.2
7.19
7.18
7.17
7.16
7.15
7.14
7.13
7.12
7.11
7.1
7.09
7.08
[ppm]

Small Molecule NMR

Peptides

Spectrum Visualisation and Navigation

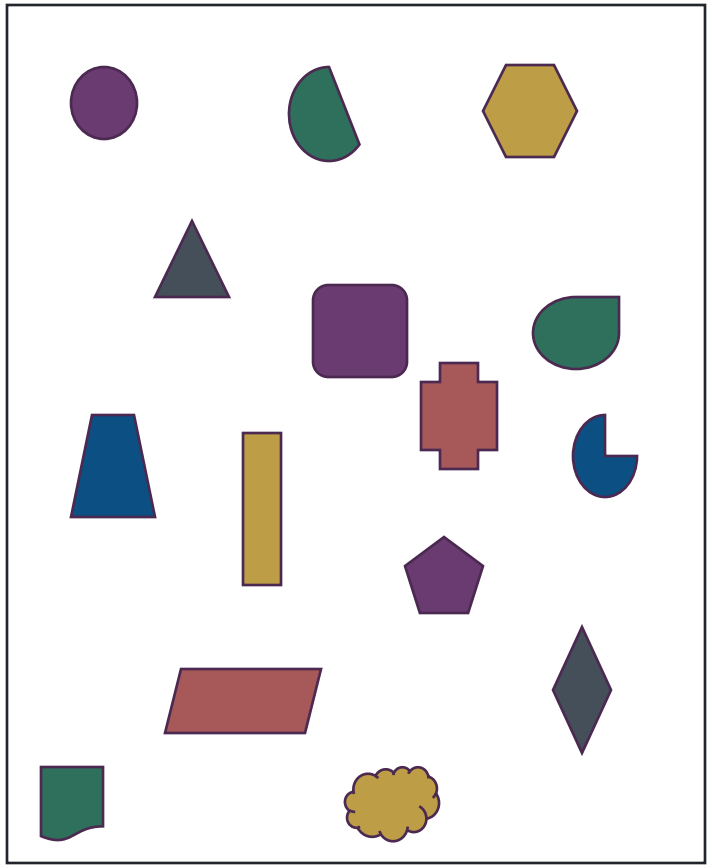


Screening

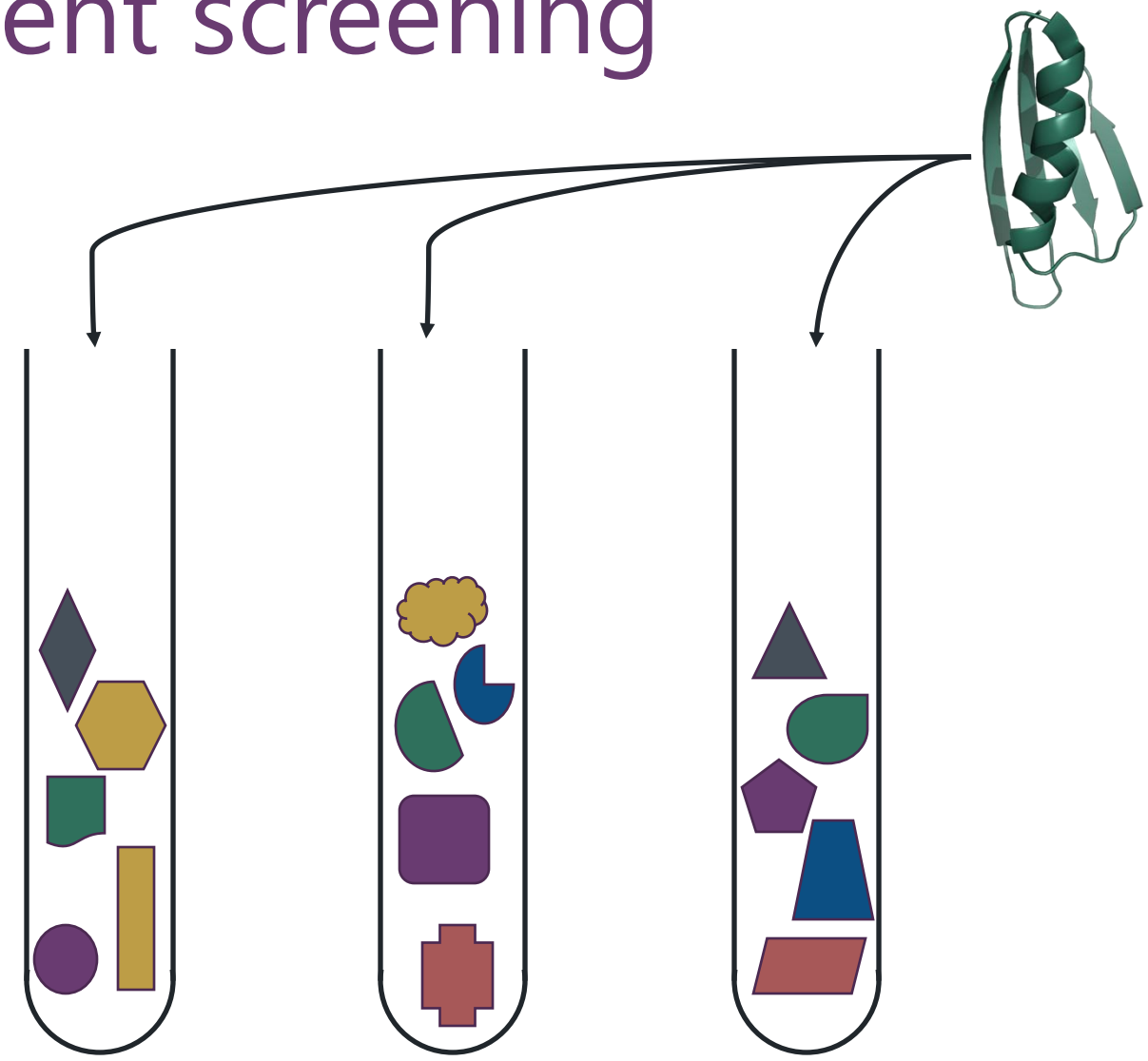
Metabolomics

Macro Writing

NMR-based fragment screening



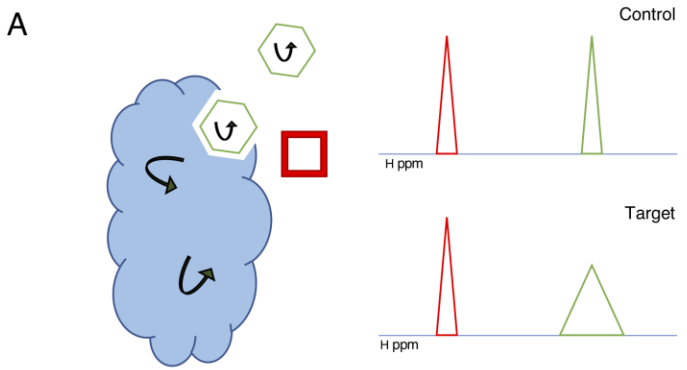
Library
(500-5000 compounds)

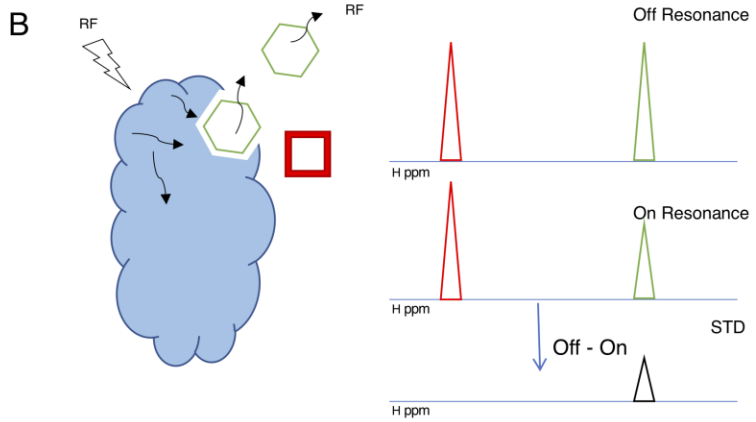


Mixtures

NMR-based fragment screening

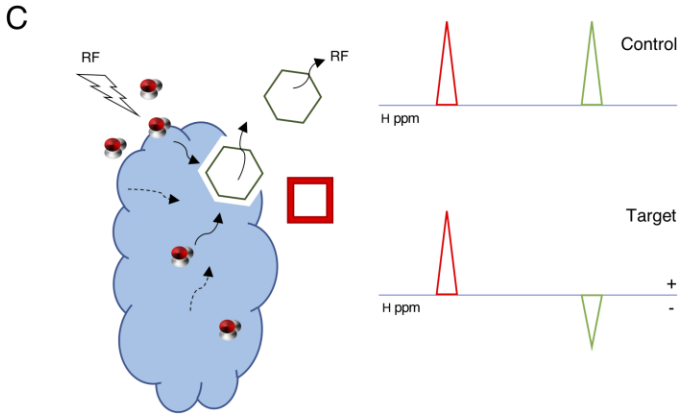
$^{1}H/^{19}F$ Relaxation-edited

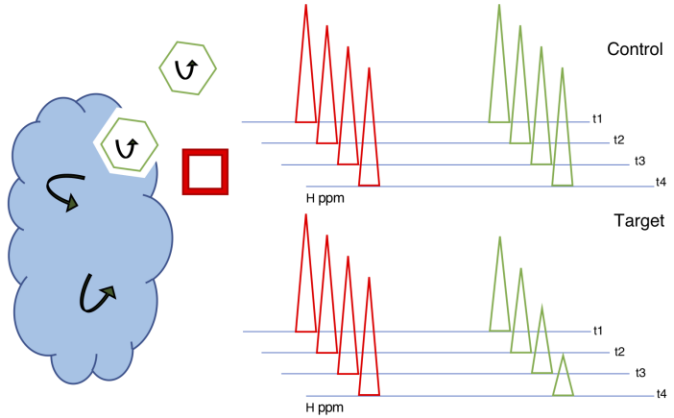




Saturation Transfer Difference (STD)

WaterLOGSY



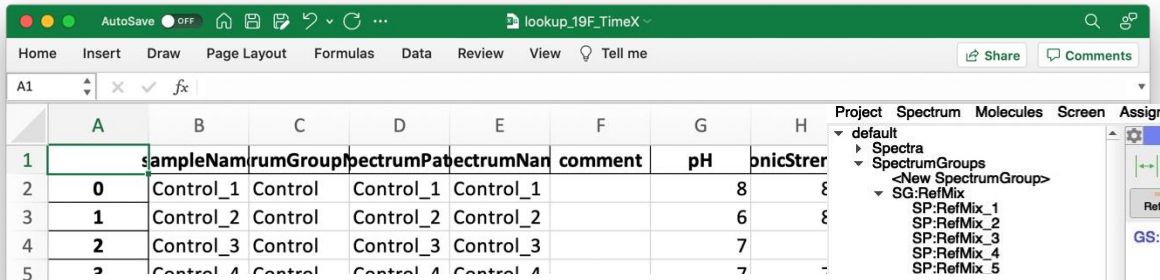


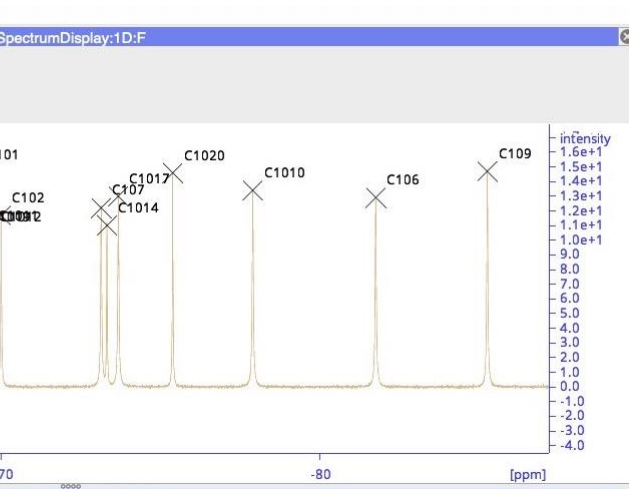
T_{1Q} series

Mureddu et al, J. Biomol. NMR, 2020; https://doi.org/10.1007/s10858-020-00321-1

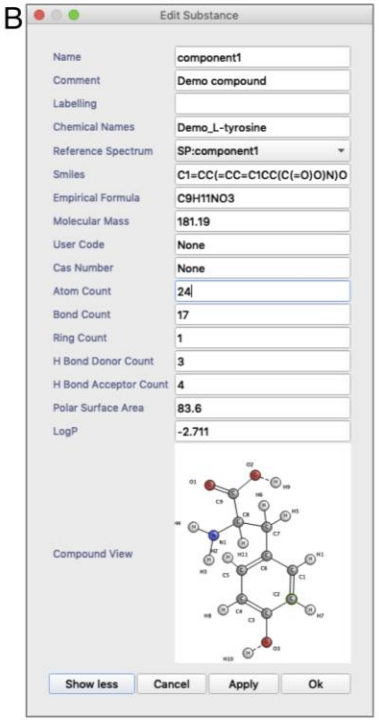
Loading data

Excel/CSV/NEF
- reference spectra
- compound information
- metadata

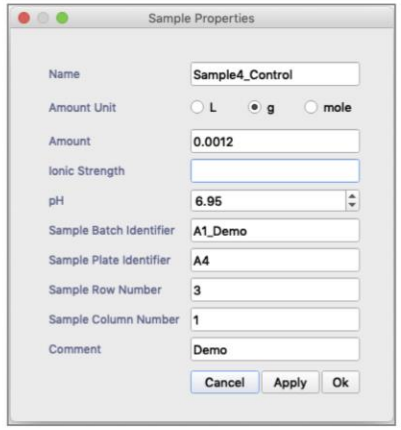




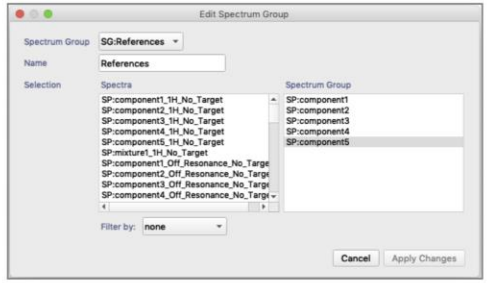
B



C

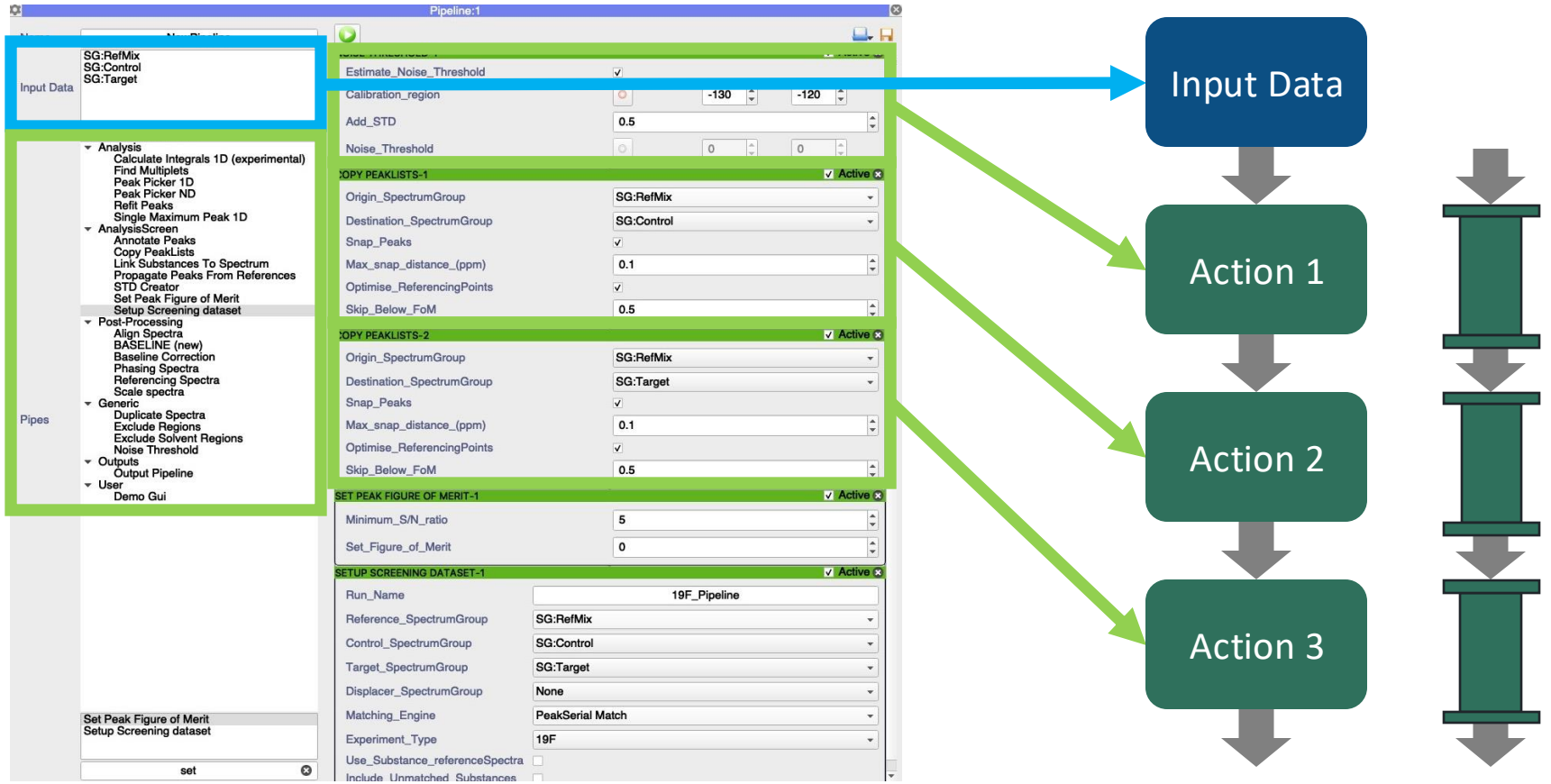


D



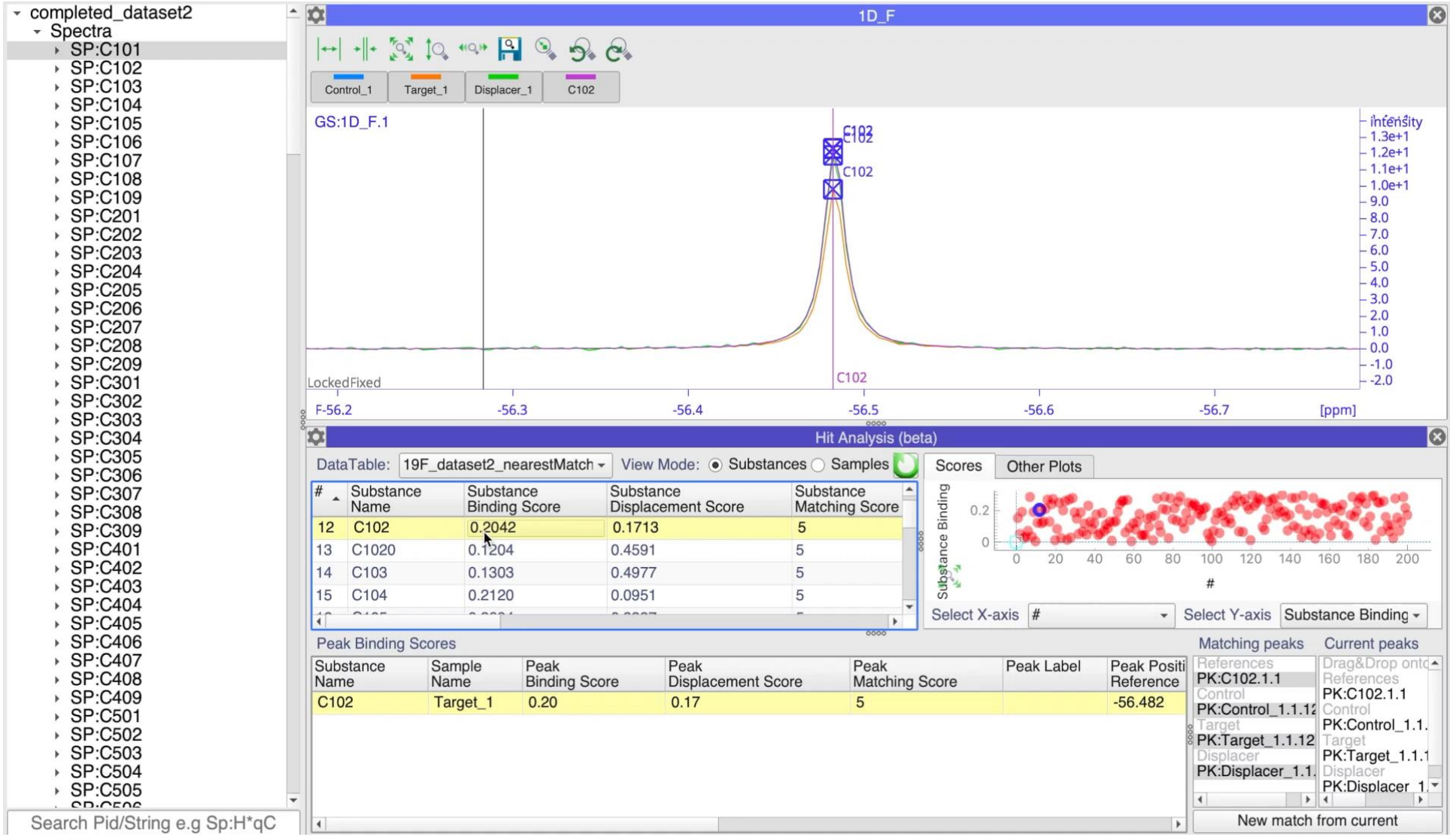
Mureddu et al, J. Biomol. NMR, 2020; https://doi.org/10.1007/s10858-020-00321-1

Automation: pipelines



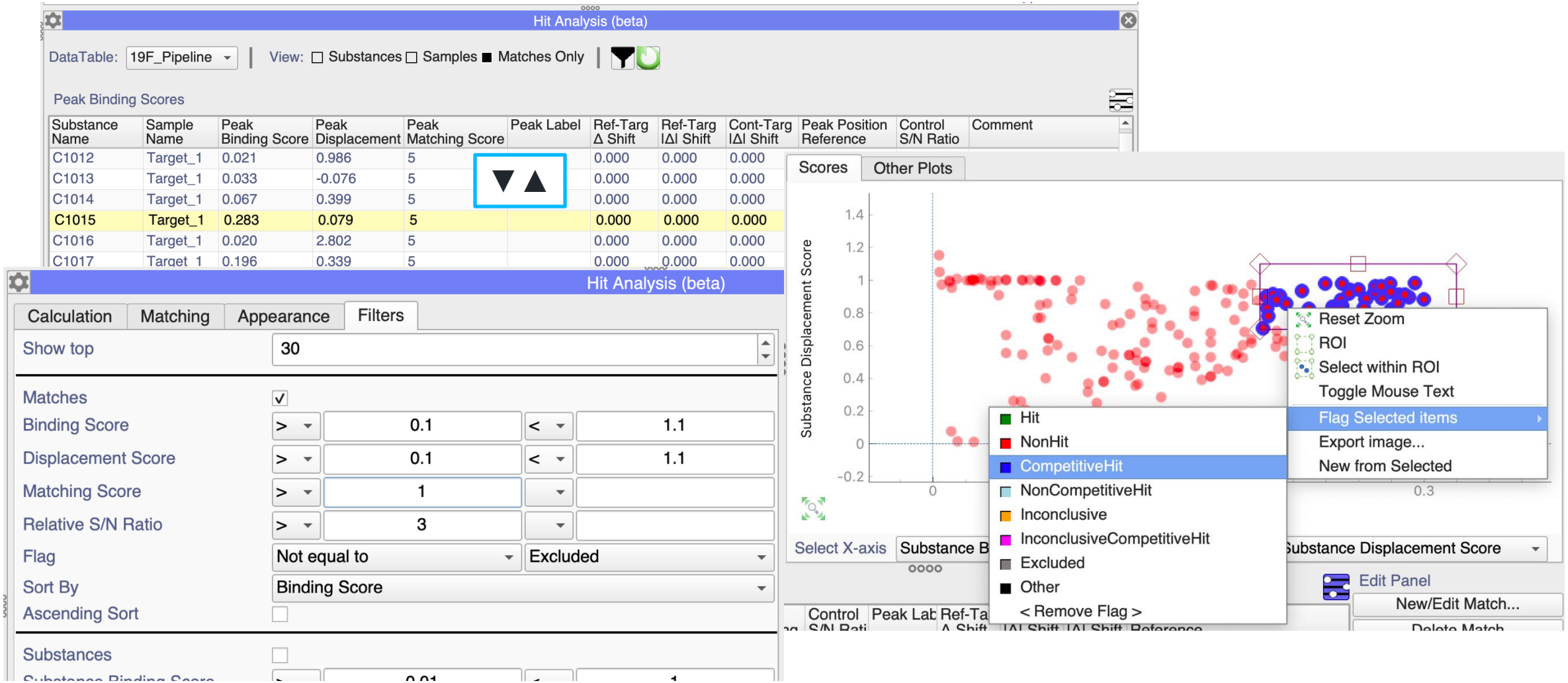
Mureddu et al, J. Biomol. NMR, 2020; https://doi.org/10.1007/s10858-020-00321-1

User-friendly Hit Analysis



Mureddu et al, J. Biomol. NMR, 2020; https://doi.org/10.1007/s10858-020-00321-1

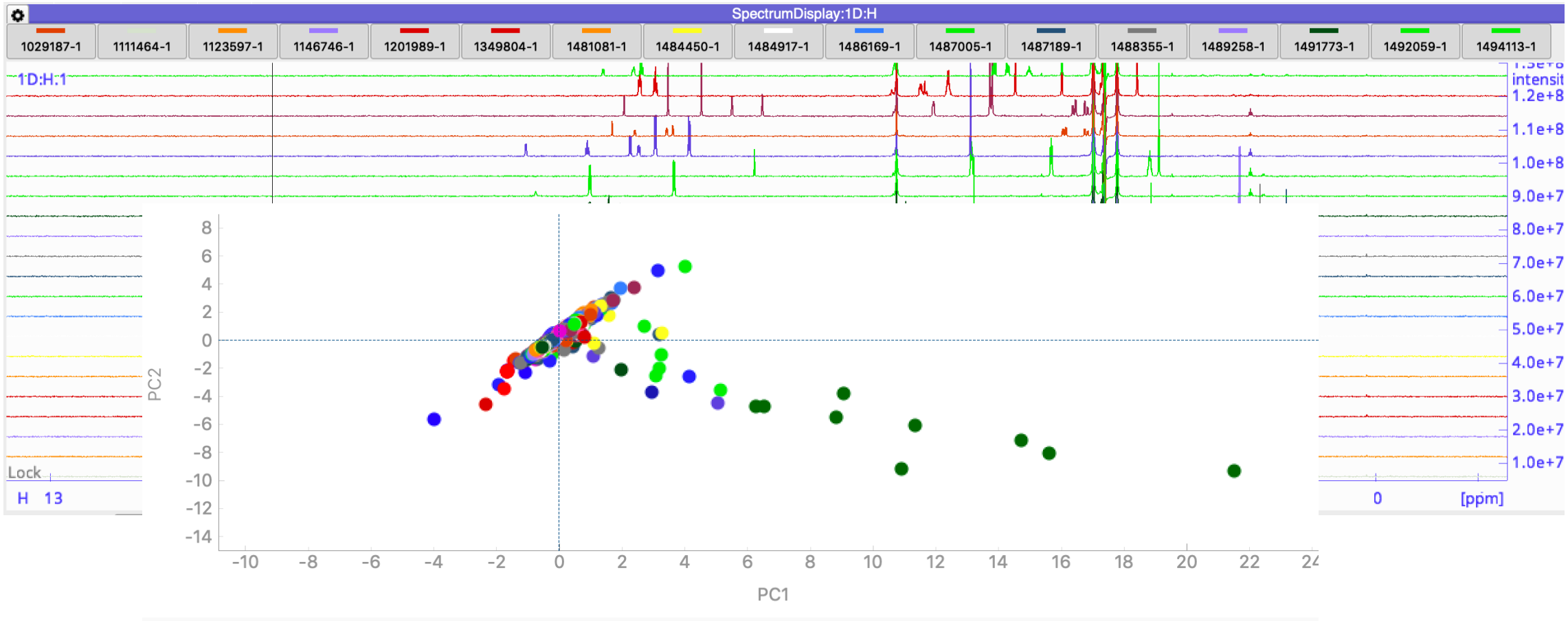
User-friendly Hit Analysis



Mureddu et al, J. Biomol. NMR, 2020; https://doi.org/10.1007/s10858-020-00321-1

Quality control

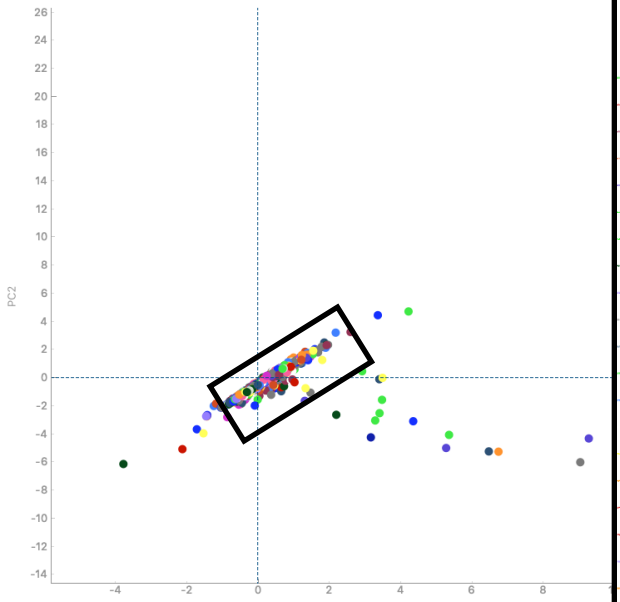
Reference data Principal Component Analysis (PCA)

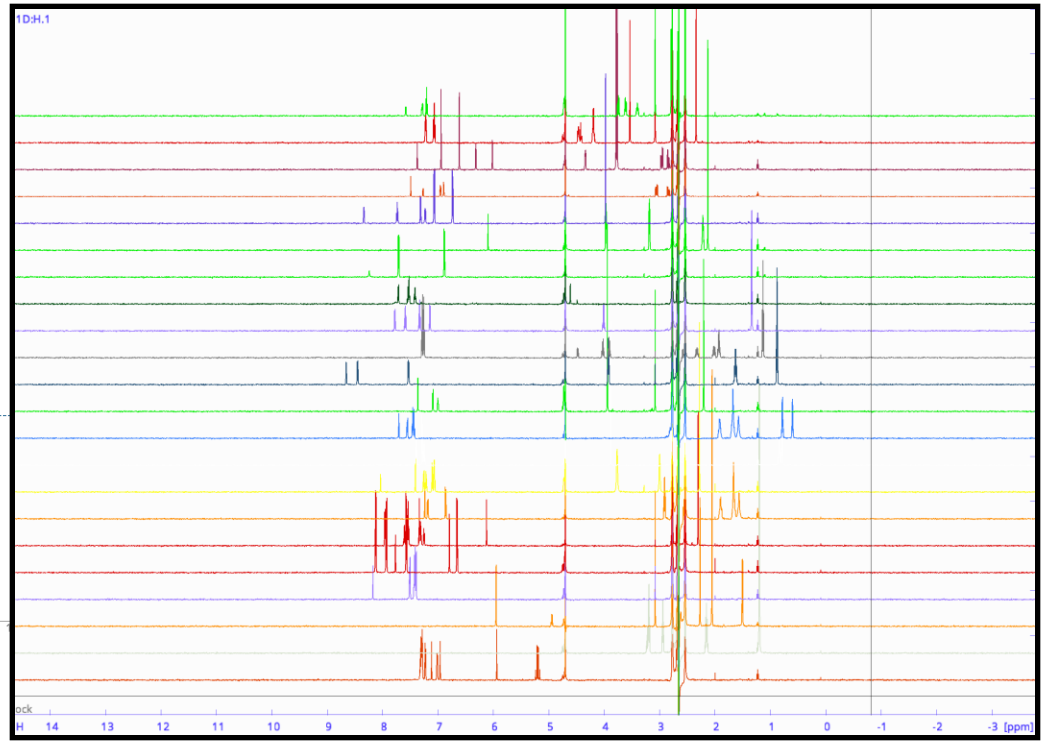


Mureddu et al, J. Biomol. NMR, 2020; https://doi.org/10.1007/s10858-020-00321-1

Quality control

Reference data PCA

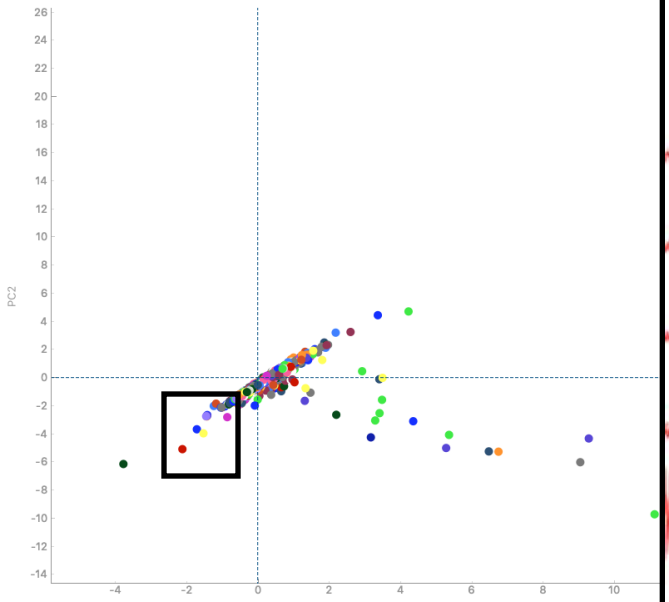


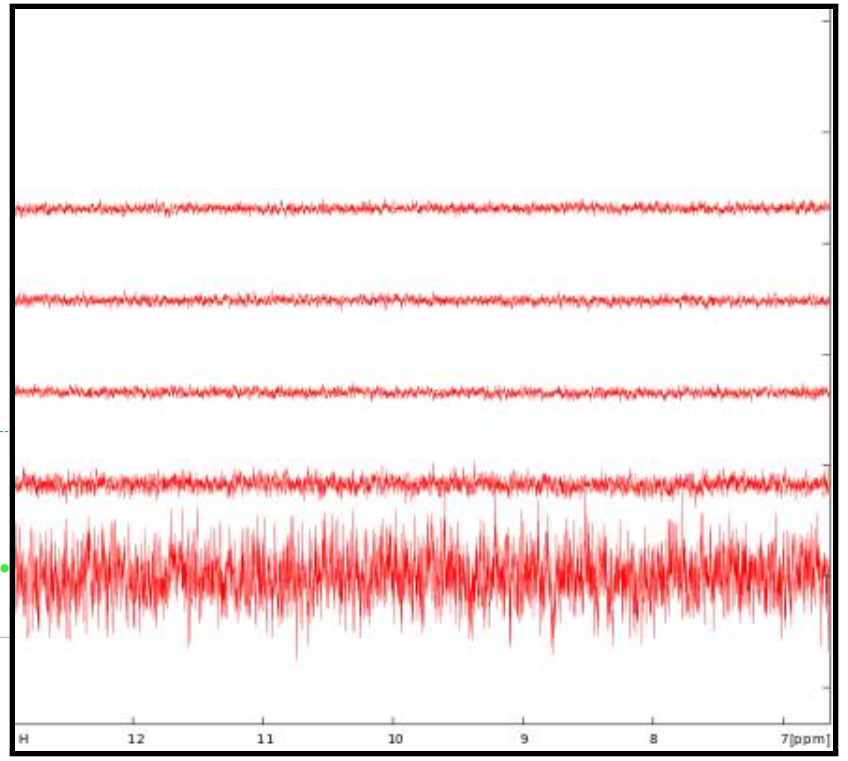


Mureddu et al, J. Biomol. NMR, 2020; https://doi.org/10.1007/s10858-020-00321-1

Quality control

Reference data PCA

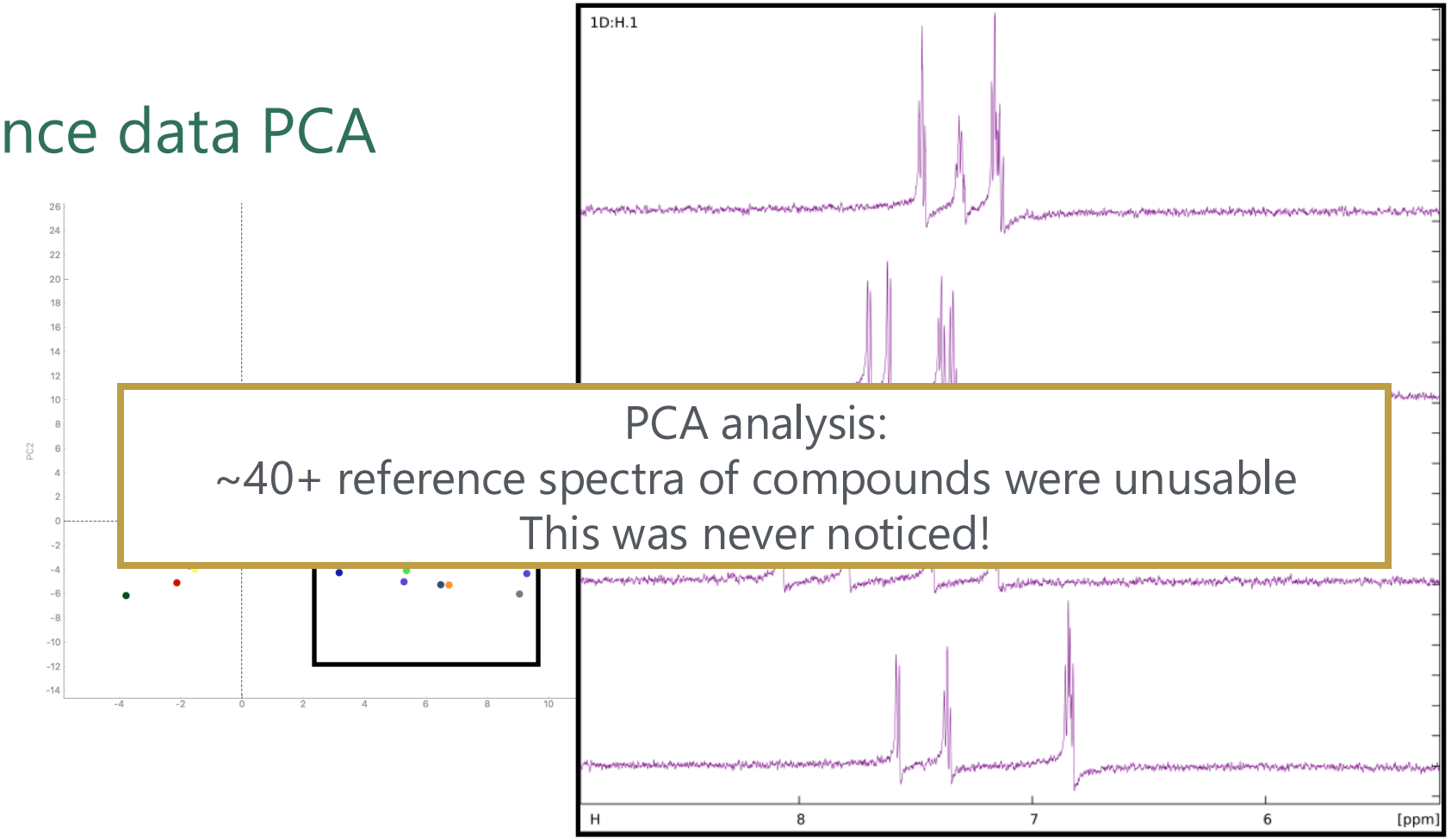




Mureddu et al, J. Biomol. NMR, 2020; https://doi.org/10.1007/s10858-020-00321-1

Quality control

Reference data PCA



PCA analysis:
~40+ reference spectra of compounds were unusable
This was never noticed!

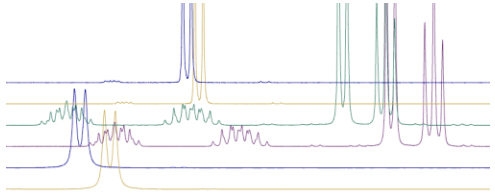
Mureddu et al, J. Biomol. NMR, 2020; https://doi.org/10.1007/s10858-020-00321-1

Small Molecule NMR

Peptides

Spectrum Visualisation and Navigation

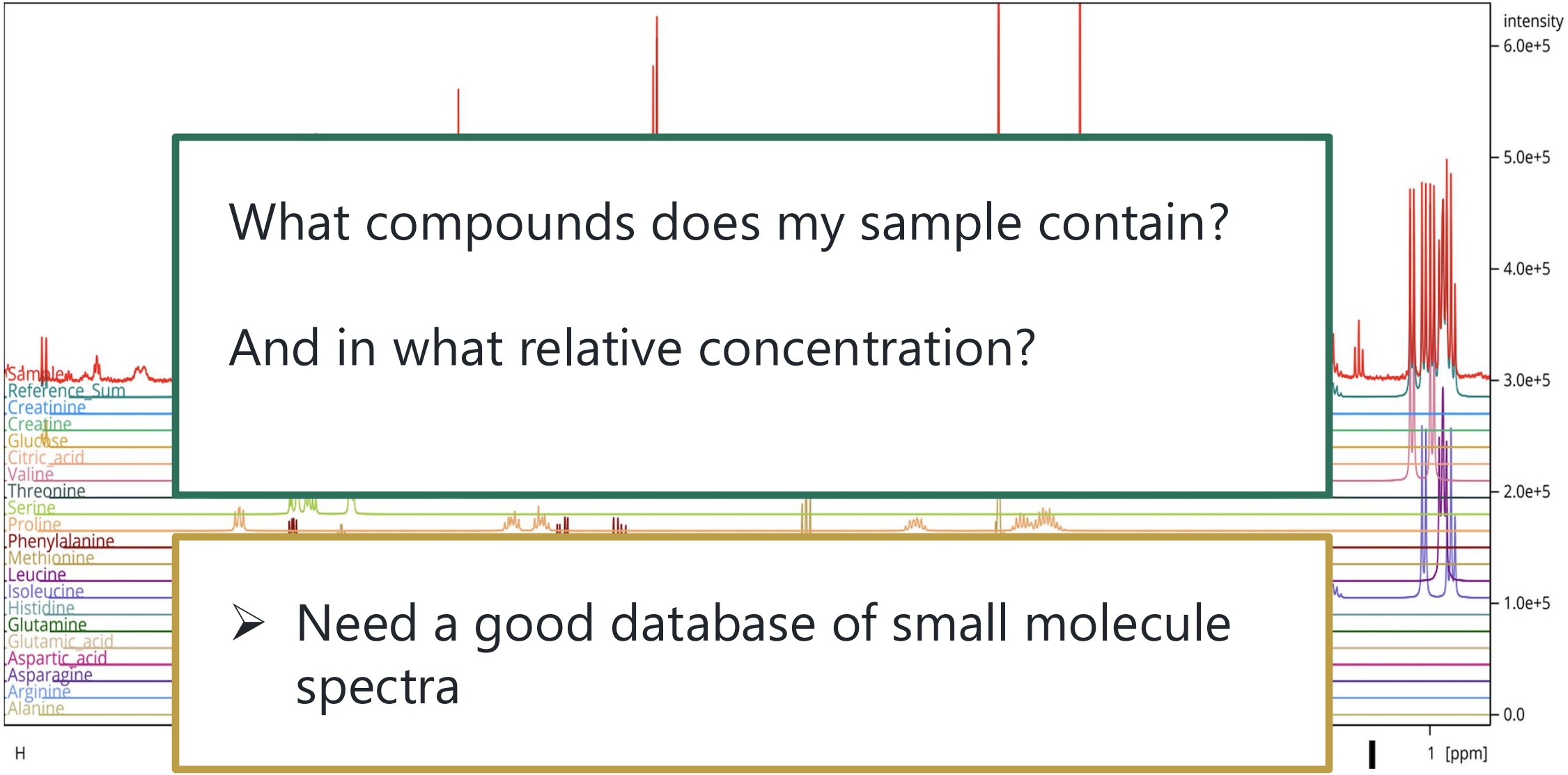
Screening



Metabolomics

Macro Writing

Metabolomics



What compounds does my sample contain?

And in what relative concentration?

- Need a good database of small molecule spectra

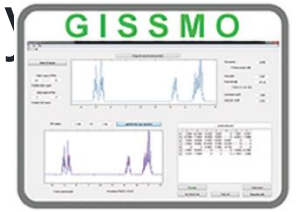
Metabolomics Databases

Commercial (integrated into software):

- Chenomx (336)
- KnowItAll (~160,000, including many pharmaceuticals)

Free:

- Human Metabolics Data Base (HMDB)
- Biological Magnetic Resonance Bank (BMRB) (>2700)
- Guided Ideographic Spin System Model Optimization library (GISSMO)



Metabolomics Databases

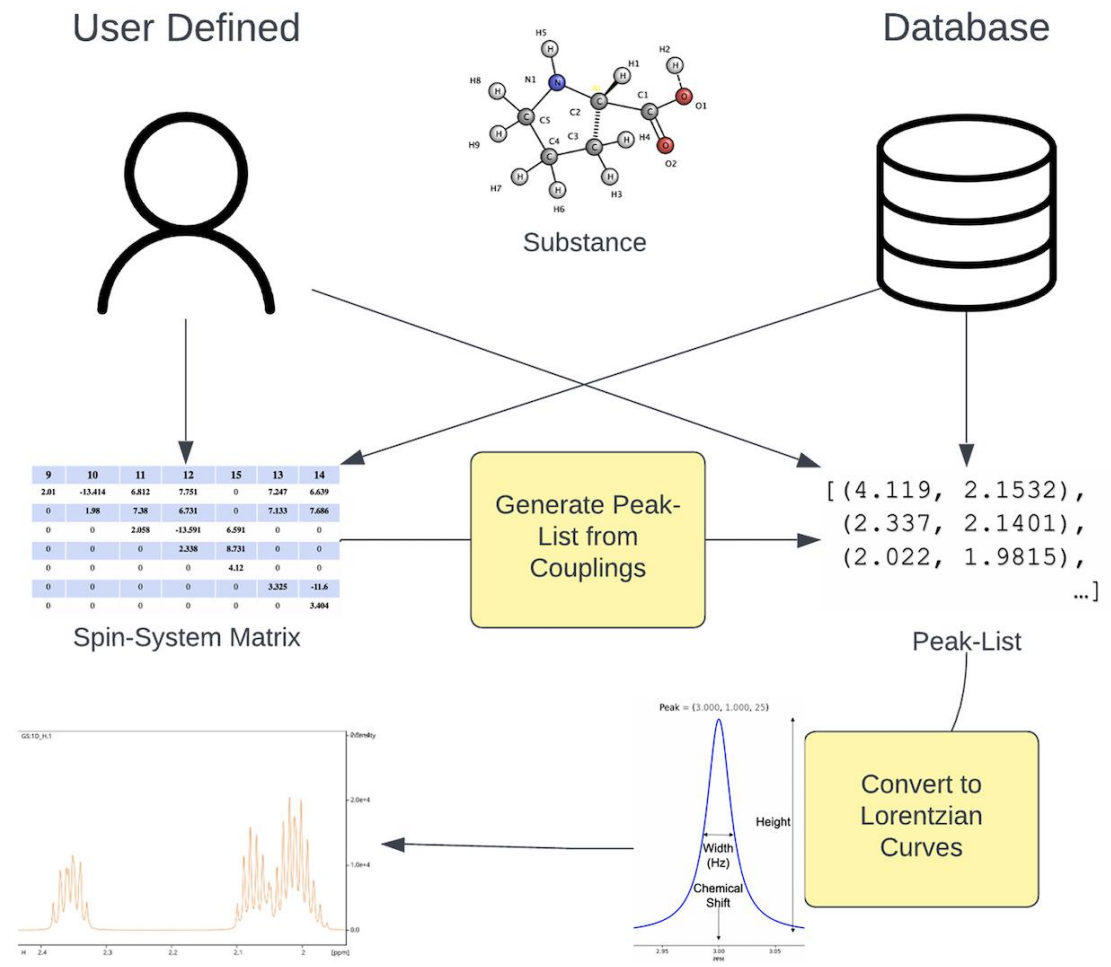
Depositied metabolite NMR spectra will vary depending on:

- temperature
- pH
- solvent
- field strength of the spectrometer

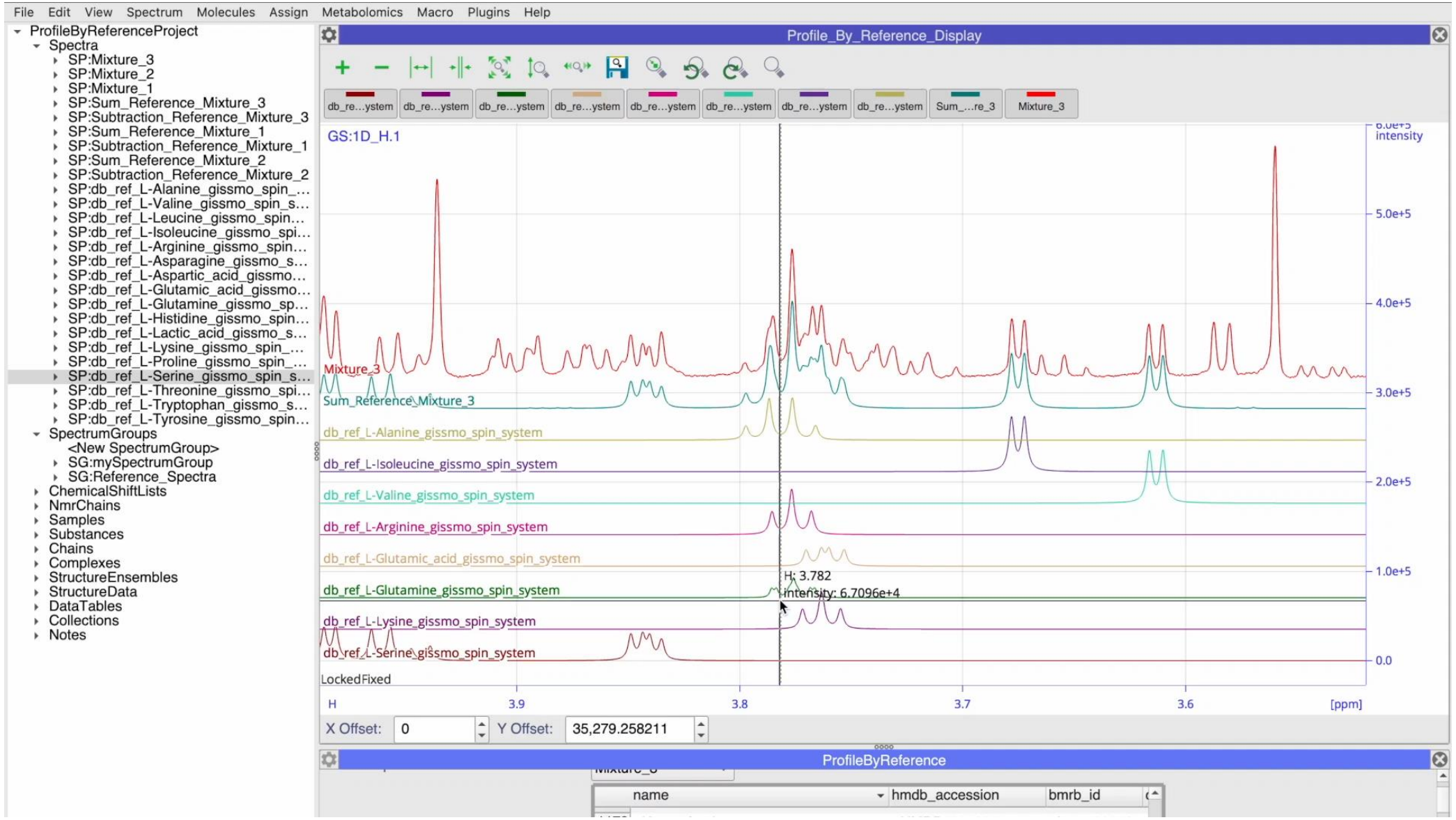
➢ideally want completely annotated spectrum with its respective metabolite chemical data

➢use simulation to extrapolate to other conditions or spectrometer field strengths

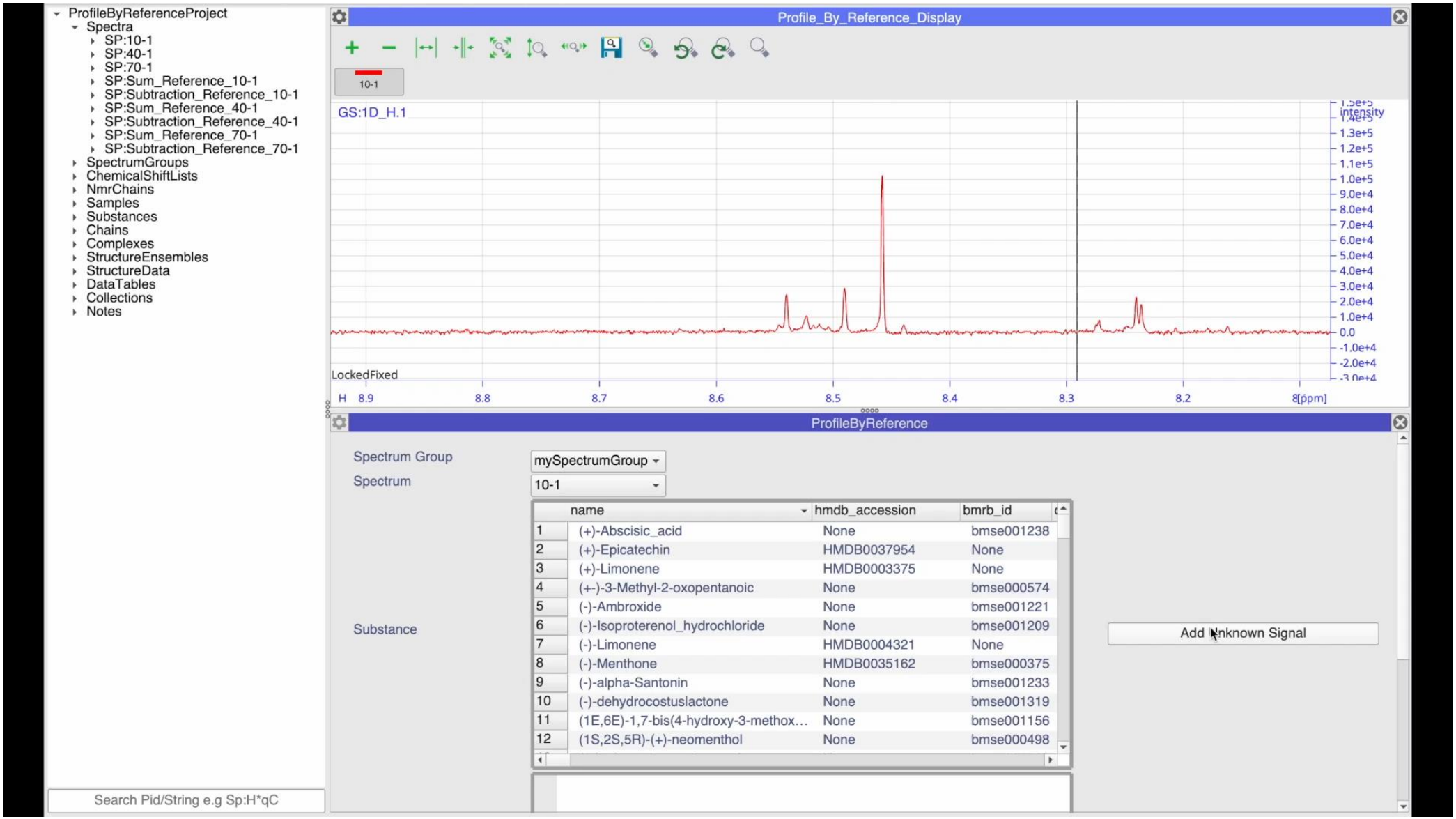
CcpNmr Analysis Simulated Metabolite DataBase (CASMDB)



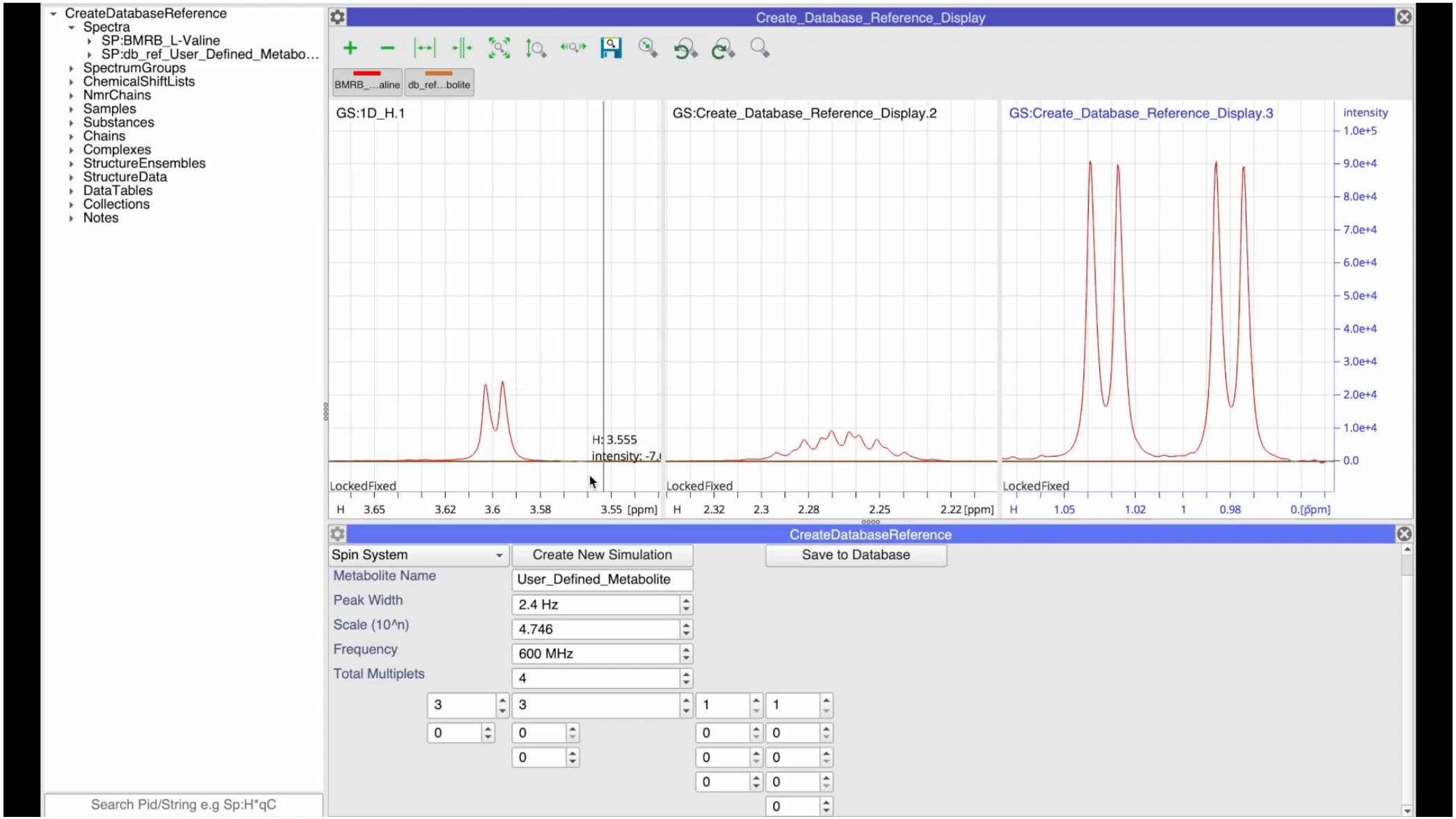
AnalysisMetabolomics



AnalysisMetabolomics



AnalysisMetabolomics



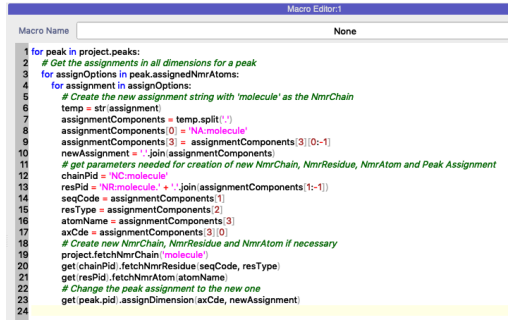
Small Molecule NMR

Peptides

Spectrum Visualisation and Navigation

Screening

Metabolomics



Macro Writing

Analysis V3 as a development platform

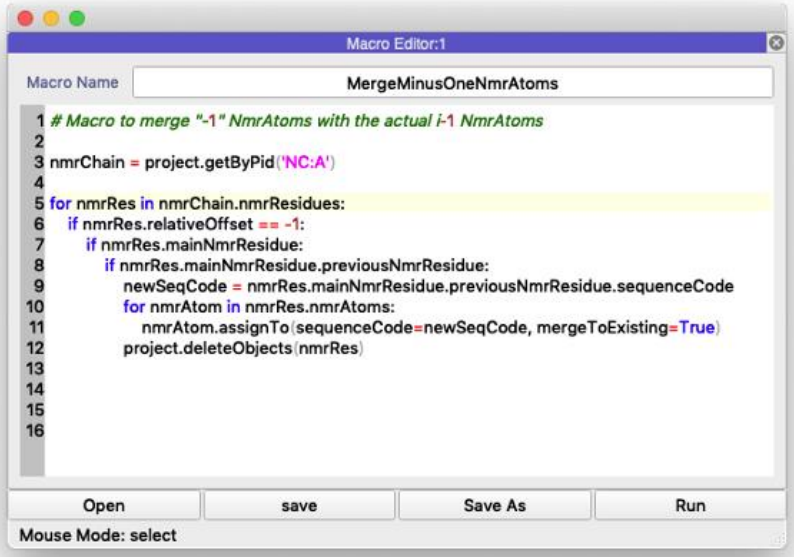
LMFIT

Plus: all the CcpNmr Analysis V3 routines to read, store, display and manipulate NMR data

Macro Writing Tutorials

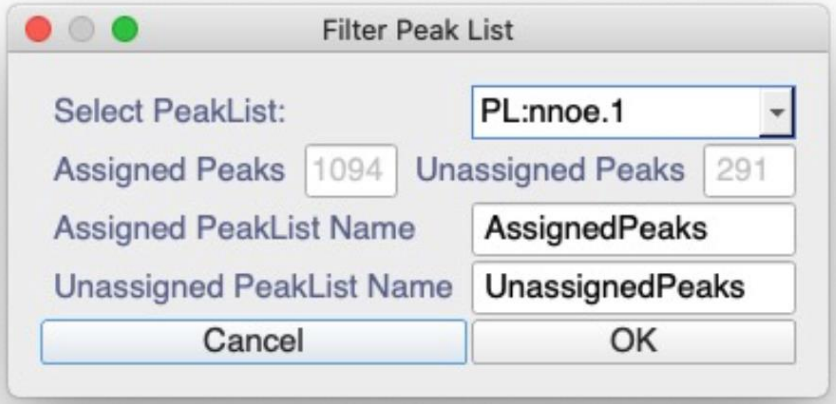
CCPN CcpNmr Analysis Version 3

Macro Writing Tutorial

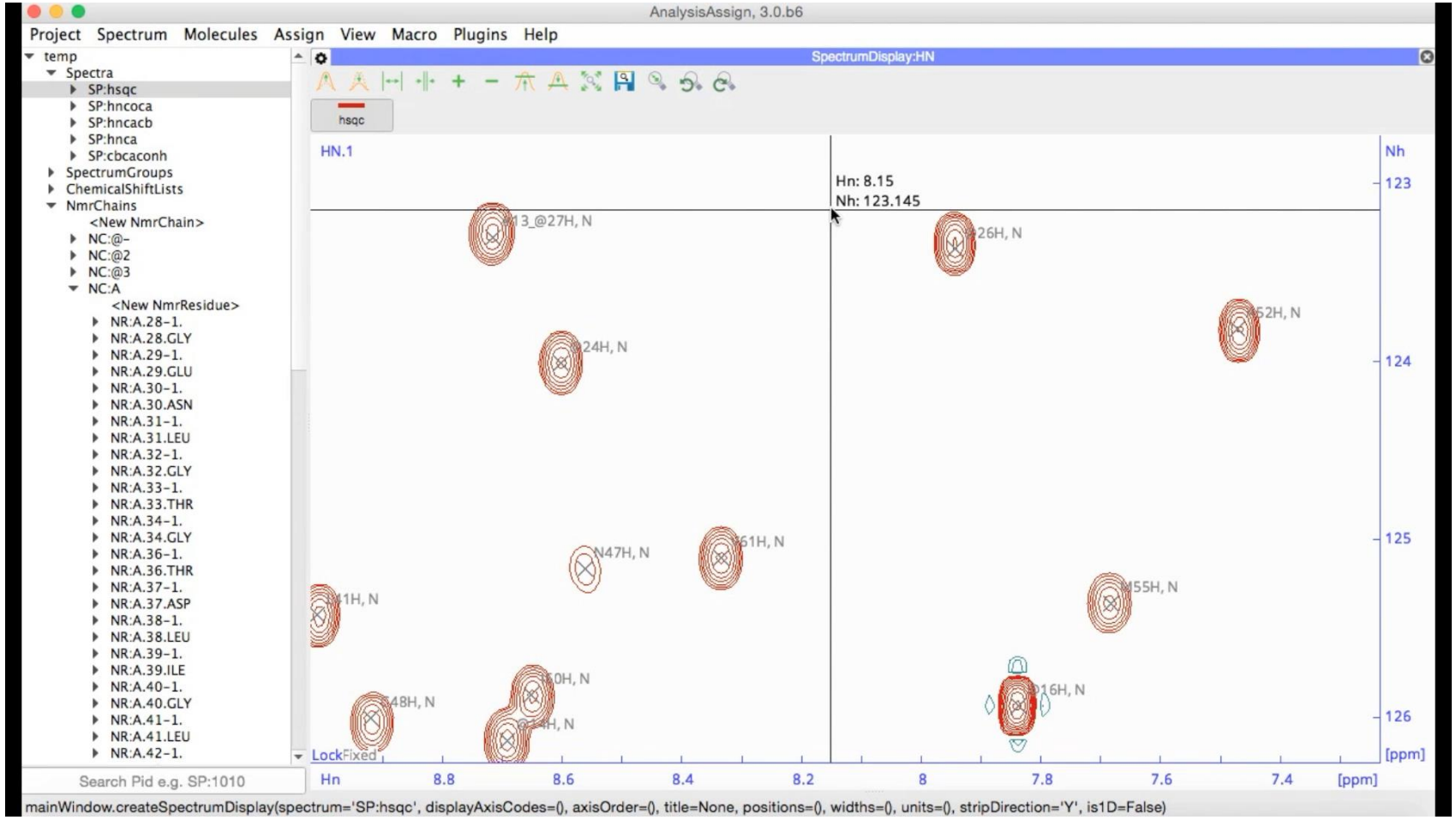


CCPN CcpNmr Analysis Version 3

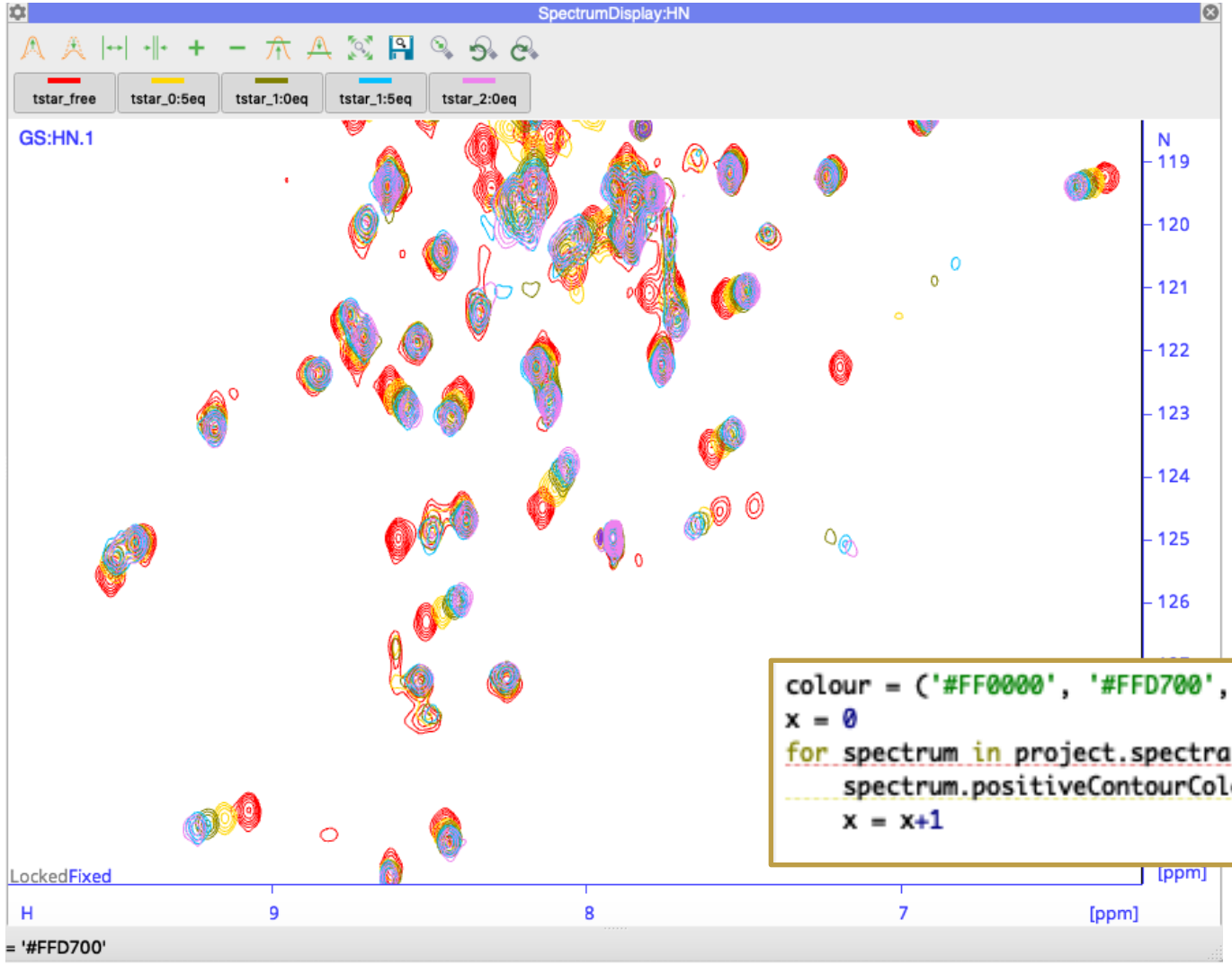
GUI Macro Writing Tutorial



GUI command echo & python console

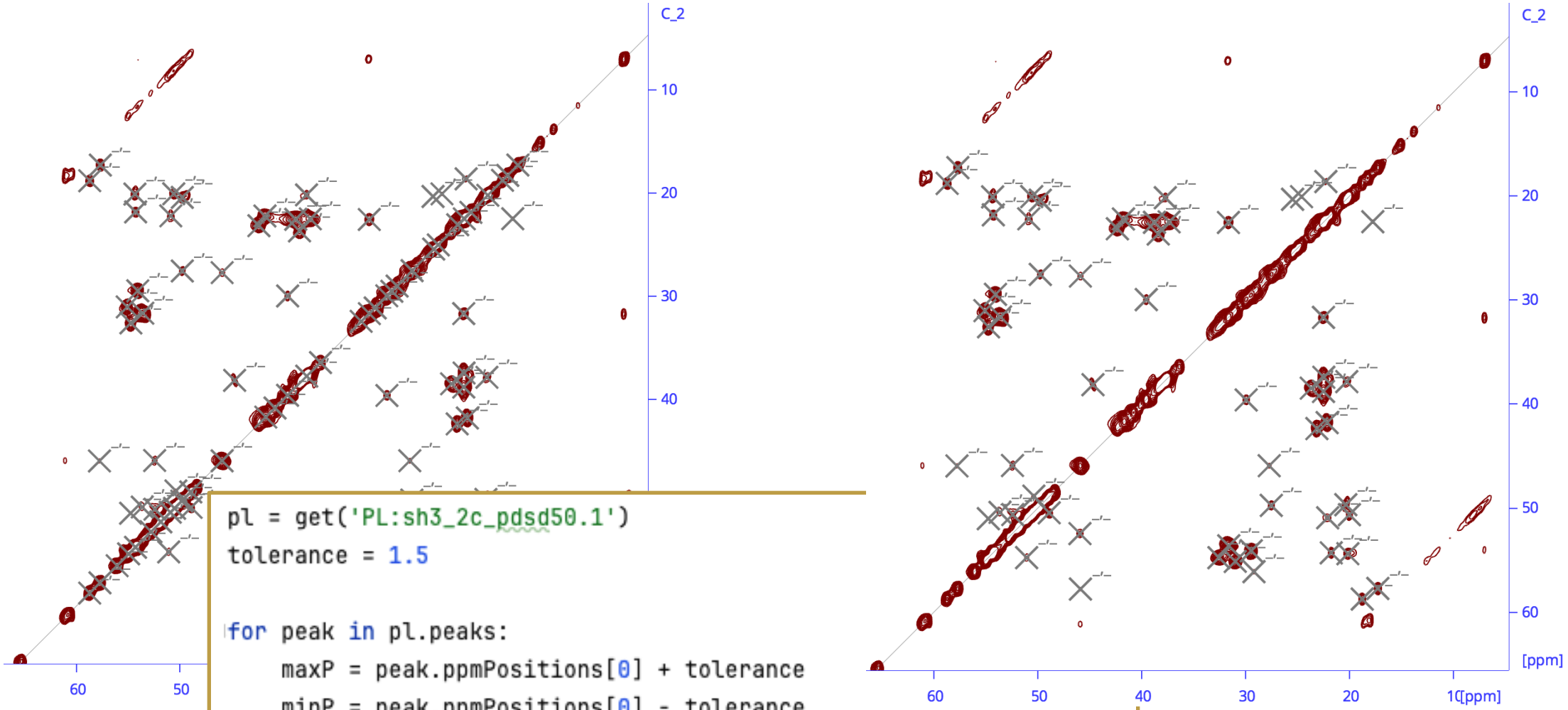


Macro: Colour a series in rainbow colours



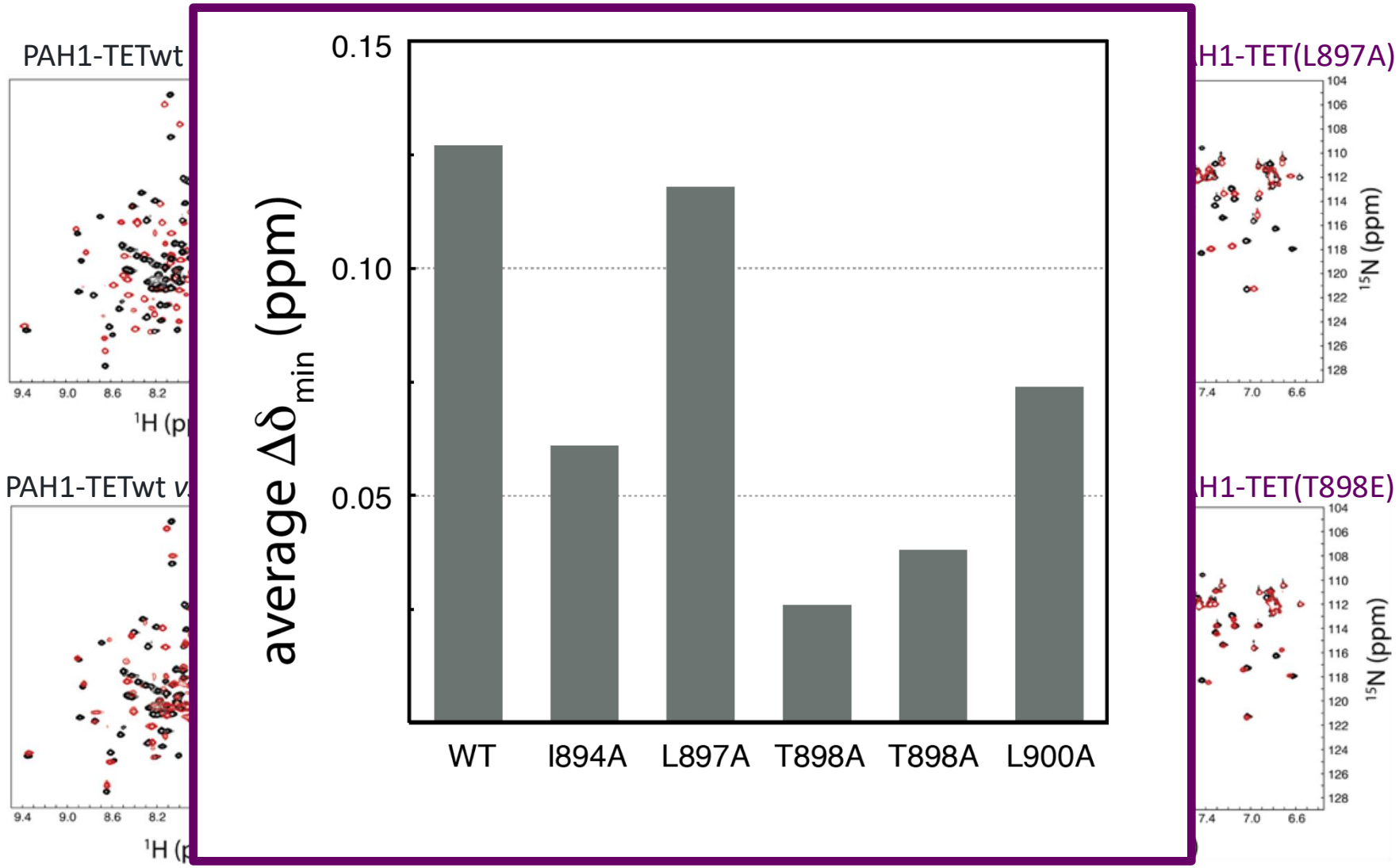

```
colour = ('#FF0000', '#FFD700', '#808000', '#00BFFF', '#EE82EE')
x = 0
for spectrum in project.spectra:
    spectrum.positiveContourColour = colour[x]
    x = x+1
```

Macro: Delete diagonal peaks



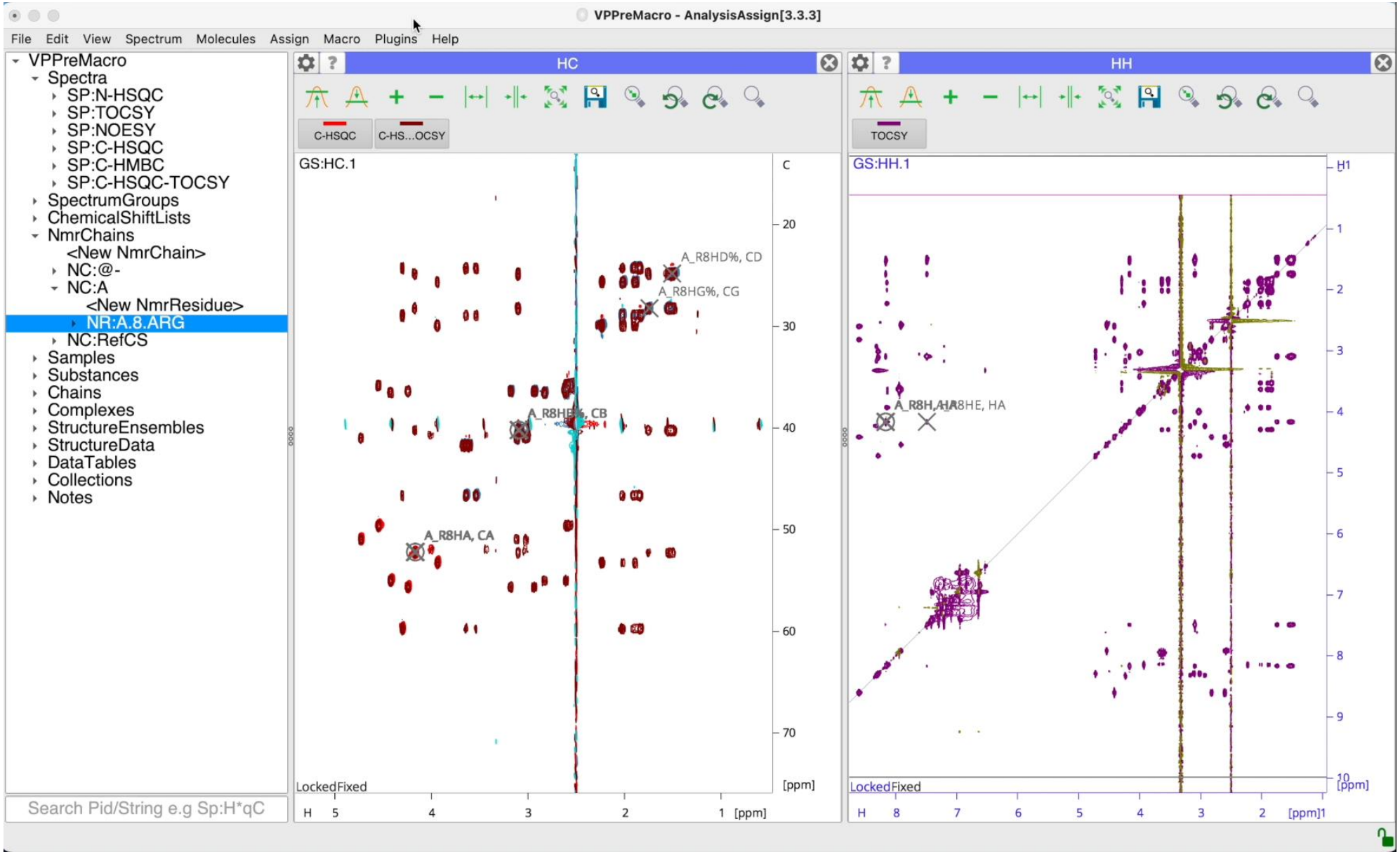

```
pl = get('PL:sh3_2c_pdsd50.1')
tolerance = 1.5

for peak in pl.peaks:
    maxP = peak.ppmPositions[0] + tolerance
    minP = peak.ppmPositions[0] - tolerance
    if peak.ppmPositions[1] < maxP and peak.ppmPositions[1] > minP:
        peak.delete()
```

Macro: Data Analysis

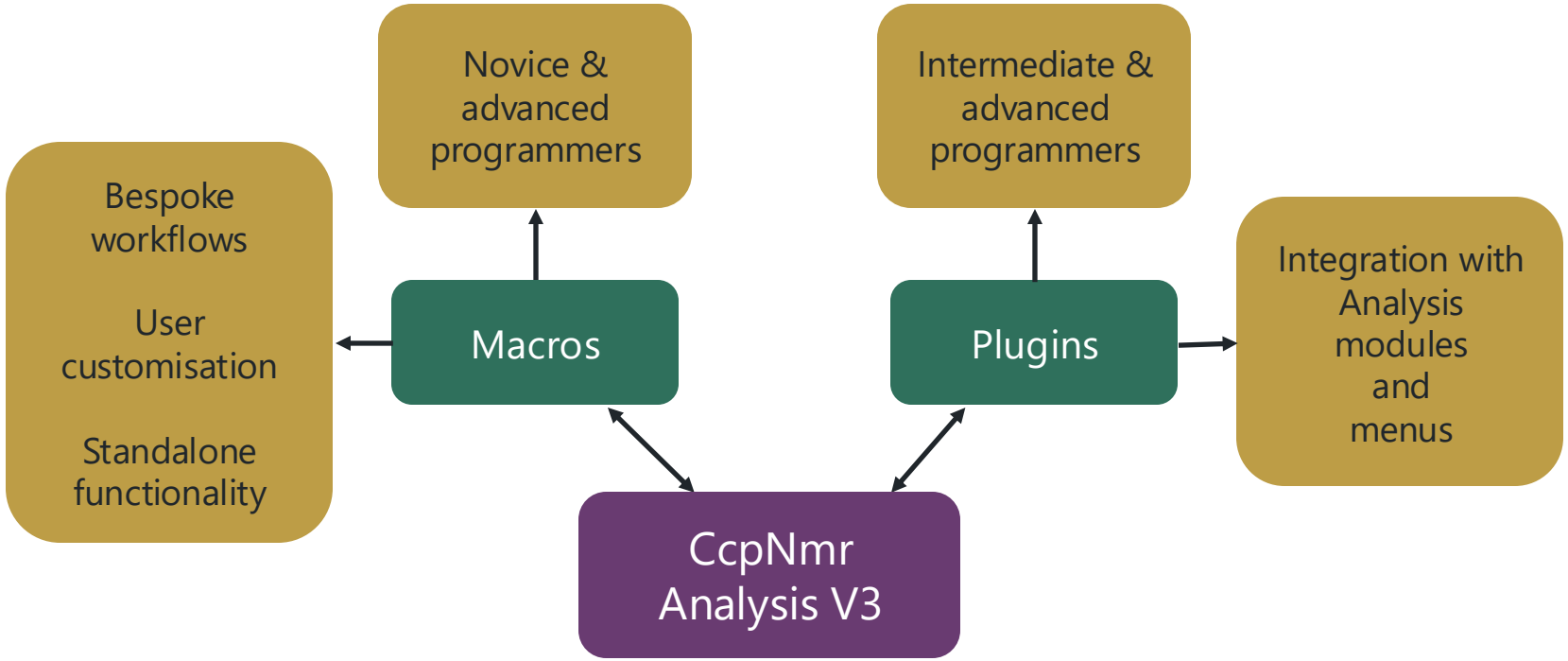


Chandru et al. (2018) *Sci Rep* 8:14689

Macro: Peptides



Analysis V3 as a development platform



Plugins

SCIENCE ADVANCES | RESEARCH ARTICLE

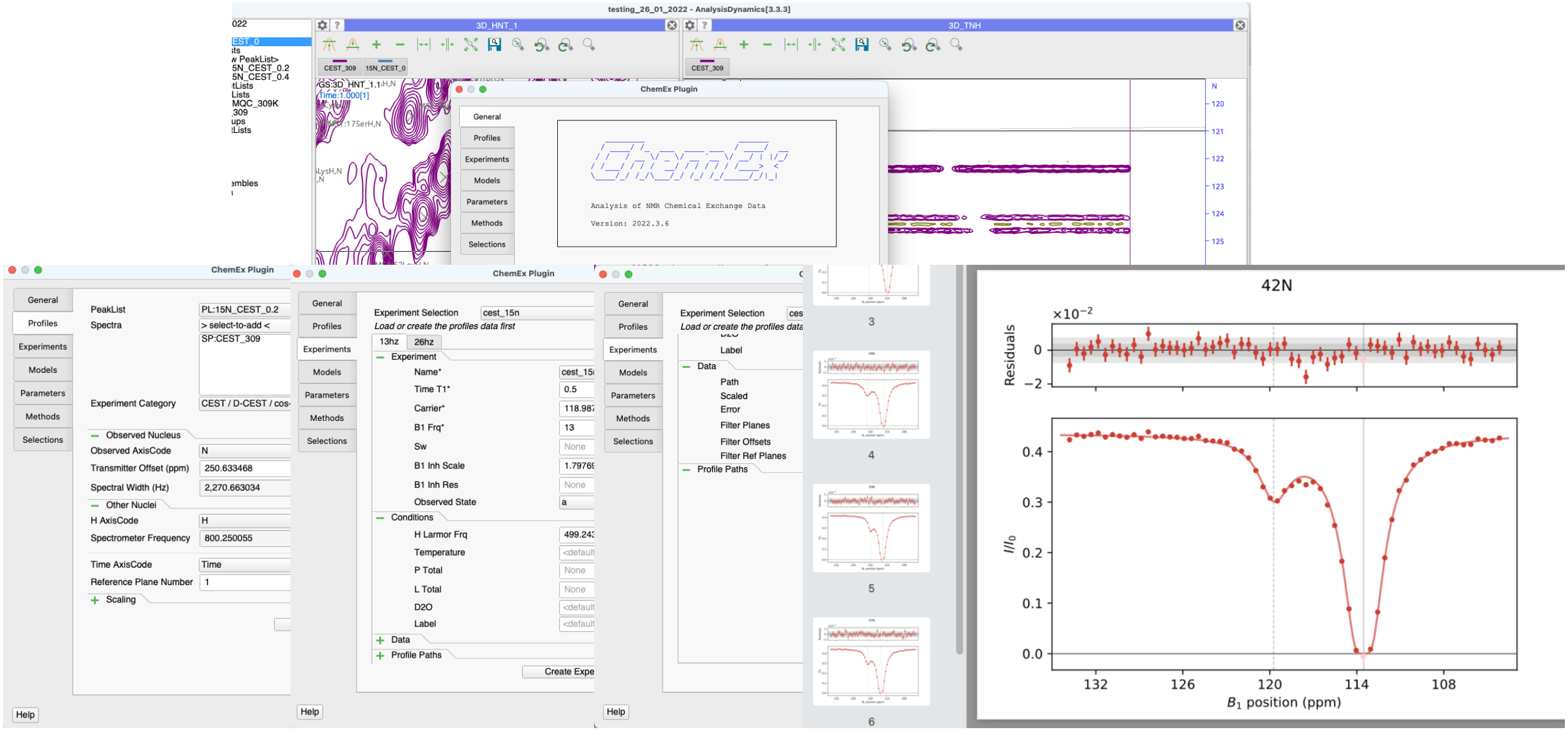
CHEMICAL PHYSICS

Protein NMR assignment by isotope pattern recognition

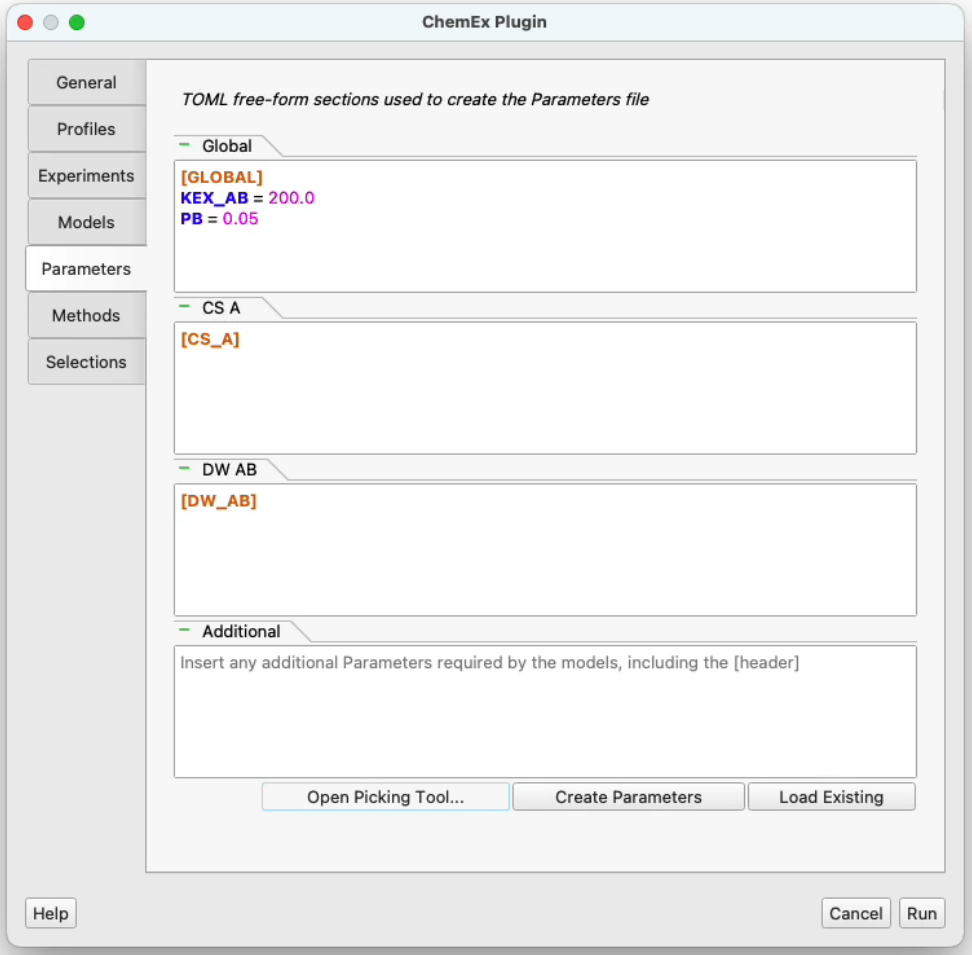
Uluk Rasulov[1], Harrison K. Wang[2,3], Thibault Viennet[4], Maxim A. Droemer[5], Srđan Matosin[2,3], Sebastian Schindler[2,3], Zhen-Yu J. Sun[2,3], Luca Mureddu[6], Geerten W. Vuister[6], Scott A. Robson[2,3], Haribabu Arthanari[2,3]*, Ilya Kuprov[1]*

The current standard method for amino acid signal identification in protein NMR spectra is sequential assignment using triple-resonance experiments. Good software and elaborate heuristics exist, but the process remains laboriously manual. Machine learning does help, but its training databases need millions of samples that cover all relevant physics and every kind of instrumental artifact. In this communication, we offer a solution to this problem. We propose polyadic decompositions to store millions of simulated three-dimensional NMR spectra, on-the-fly generation of artifacts during training, a probabilistic way to incorporate prior and posterior information, and integration with the industry standard CcpNmr software framework. The resulting neural nets take [^{1}H,^{13}C] slices of mixed pyruvate–labeled HNCA spectra (different CA signal shapes for different residue types) and return an amino acid probability table. In combination with primary sequence information, backbones of common proteins (GB1, MBP, and INMT) are rapidly assigned from just the HNCA spectrum.

Plugins - ChemEx

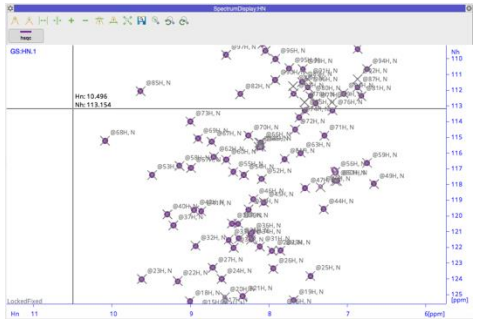


Plugins - ChemEx



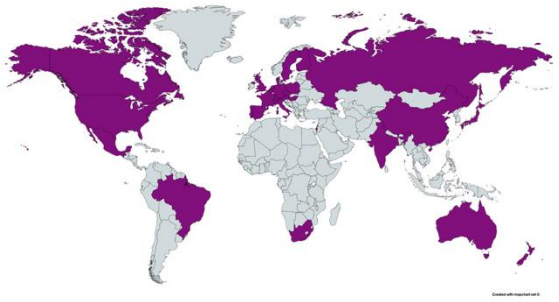
Advantages to using the V3 environment

Python packages



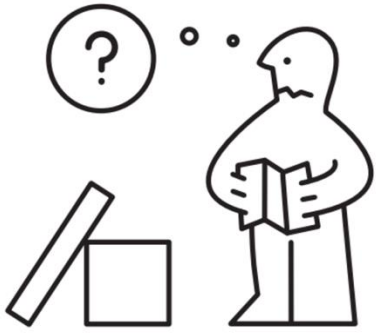
Analysis Routines

Developer Acess



Global user base

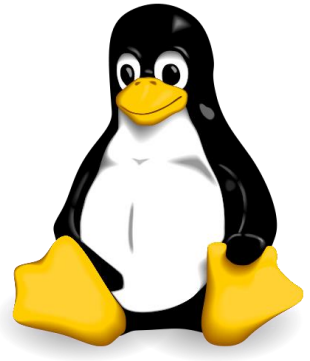
Publicity



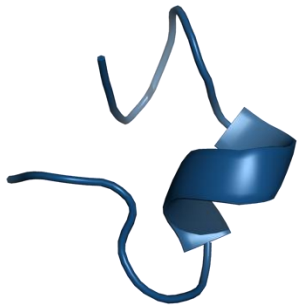
Documentation & Workshops

Give it a try!

ccpn.ac.uk

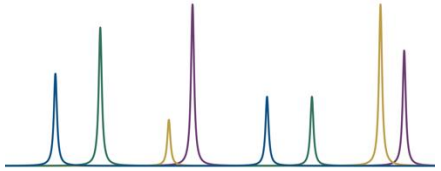


Summary



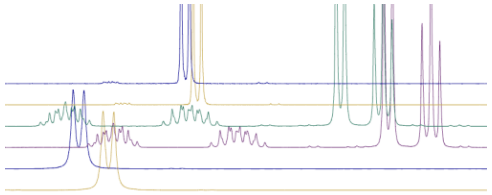
Peptides

Modern and flexible spectrum display, assignment and navigation tools



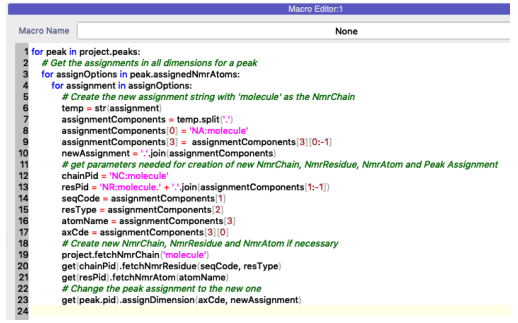
Screening

Easy analysis of large datasets & inclusion of small molecule metadata



Metabolomics

Free access to high quality database & availability of spectrum simulation tools



Macro Writing

Easy implementation of bespoke workflows and analysis needs with easy dissemination available

Thanks

CCPN
Luca Mureddu
Ed Brooksbank
Vicky Higman
Eliza Płoskoń
Morgan Heyward
Dan Thompson
Geerten Vuister

University of Kent
Gary Thompson

University of Glasgow
Brian Smith

CCPN Partners

Amber Analytical Services
Mike Bernstein

Novartis Basel
Anke Blechschmidt
Christian Schleberger
Andrea Wiget
Ulrich Hommel
Andreas Lingel

University of Liverpool
Marie Phelan
Rudi Grossman
Alex Hill

Harvard Medical School
Harrison Wang
Haribabu Arthanari

University of Southampton
Uluk Rasolov
Ilya Kuprov

ENS/CNRS Paris
Guillaume Bouvignies