



What CcpNmr Analysis can do for and with small molecules

Vicky Higman
University of Leicester



LISCB



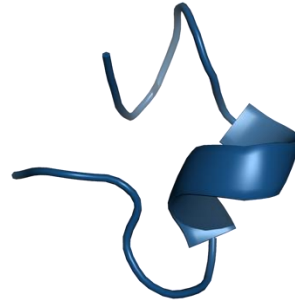
Ralph:

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



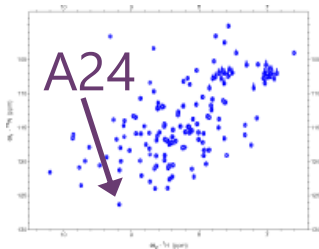
Proteins



Peptides



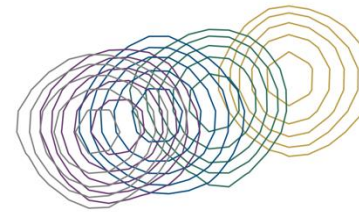
RNA/DNA



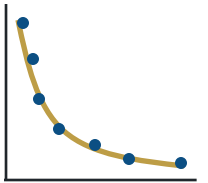
Spectrum Visualisation
and Assignment



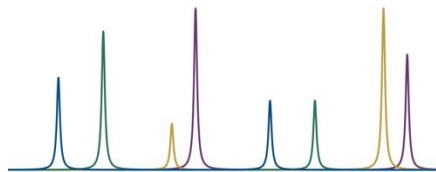
Structure
Calculation



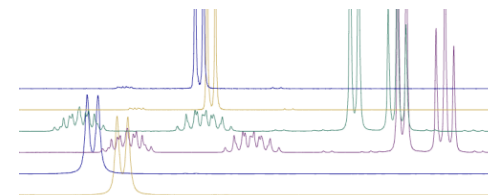
Binding Studies



Dynamics



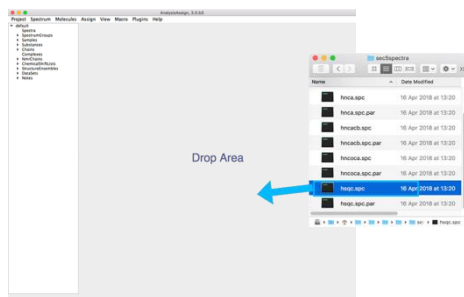
Screening



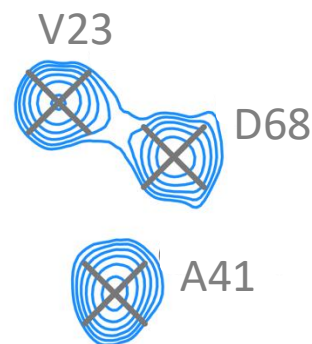
Metabolomics

Our Approach

Modern



Flexible

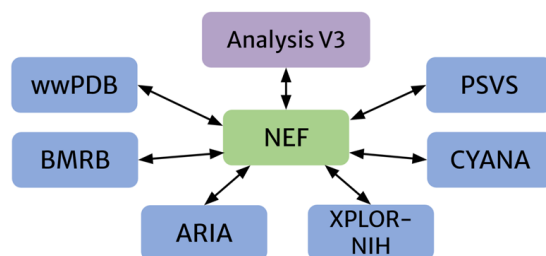


Customisable

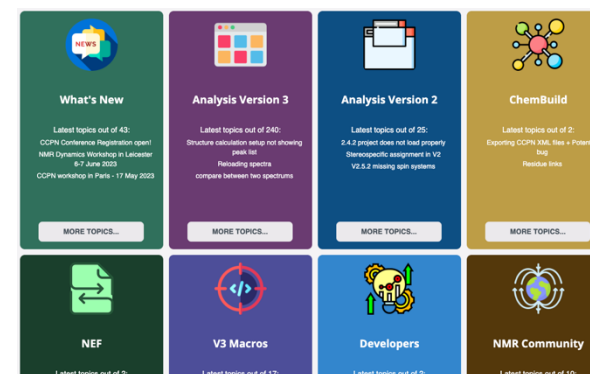
```
Macro Name: None

1 for peak in project.peak:
2   # Get the assignments in all dimensions for a peak
3   for assignOptions in peak.assignedNmrAtoms:
4     # Assignment in assignOptions:
5     # Create the new assignment string with 'molecule' as the NmrChain
6     temp = str(assignOptions)
7     assignmentComponents = temp.split(',')
8     assignmentComponents[0] = 'NA.molecule'
9     assignmentComponents[3] = assignmentComponents[3][0:-1]
10    newAssignment = ','.join(assignmentComponents)
11    # get parameters needed for creation of new NmrChain, NmrResidue, NmrAtom and Peak Assignment
12    chainPid = 'NC.molecule'
13    resPid = 'NR.molecule.' + ','.join(assignmentComponents[1:-1])
14    seqCode = assignmentComponents[1]
15    resType = assignmentComponents[2]
16    atomName = assignmentComponents[3]
17    axCde = assignmentComponents[3][0]
18    # Create new NmrChain, NmrResidue and NmrAtom if necessary
19    project.fetchNmrChain('molecule')
20    getChainPid, fetchNmrResidue = seqCode, resType
21    getResPid, fetchNmrAtom = resPid, atomName
22    # Change the peak assignment to the new one
23    getPeakPid, assignDimension = axCde, newAssignment
24
```

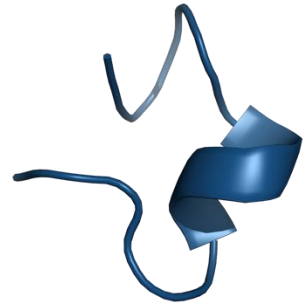
Integrated



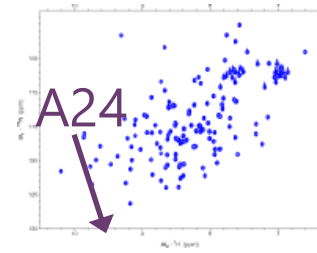
Supported



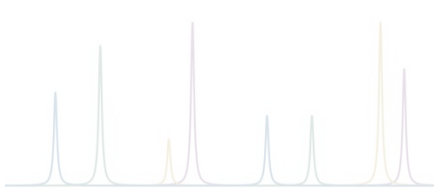
Small Molecule NMR



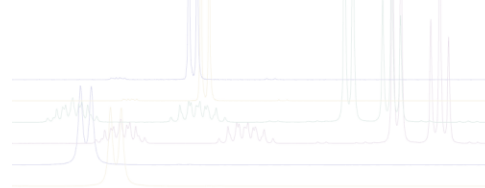
Peptides



Spectrum Visualisation
and Navigation



Screening

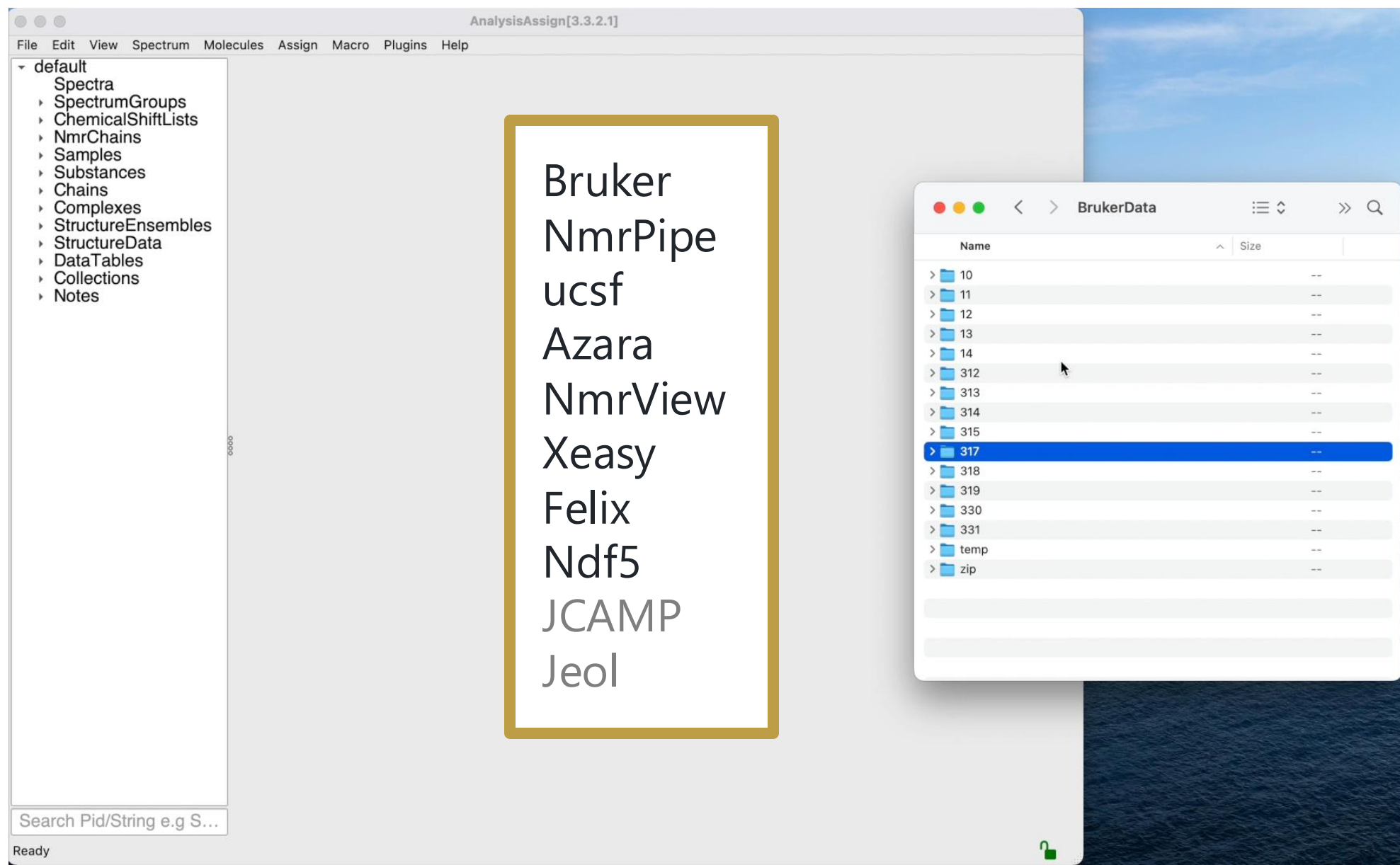


Metabolomics

```
Macro Name: None
1 for peak in project.peaks:
2 # Get the assignments in all dimensions for a peak
3 for assignOptions in peak.assignNewAtoms:
4 for assignment in assignOptions:
5 # Create the new assignment string with 'molecule' as the NmrChain
6 temp = str assignment
7 assignmentComponents = temp.split(':')
8 assignmentComponents[0] = 'NA_molecule'
9 assignmentComponents[3] = assignmentComponents[3][0:-1]
10 newAssignment = ':'.join assignmentComponents
11 # Get parameters needed for creation of new NmrChain, NmrResidue, NmrAtom and Peak Assignment
12 chainPid = 'NA_molecule'
13 resPid = 'NA_molecule' + ':'.join assignmentComponents[1:-1]
14 seqCode = assignmentComponents[1]
15 resType = assignmentComponents[2]
16 atomName = assignmentComponents[3]
17 axCde = assignmentComponents[3][0]
18 # Create new NmrChain, NmrResidue and NmrAtom if necessary
19 project.fetchNmrChain('molecule')
20 getChainPid().fetchNmrResidue(seqCode, resType)
21 getResPid().fetchNmrAtom(atomName)
22 # Change the peak assignment to the new one
23 getPeakPid().assignDimension(axCde, newAssignment)
24
```

Macro Writing

Opening and Overlaying Spectra



Programme Layout

The screenshot displays the software interface for 'Sec5Part2VAH - AnalysisAssign[3.0.2]'. The interface is divided into several sections:

- Sidebar (Left):** Contains a tree view of data and analysis modules. The 'Spectra' section is expanded, showing 'SP:hsqc' and 'PeakLists'. The 'NmrChains' section is also expanded, showing 'NC:@-' and 'NC:@2'.
- Spectrum Display Modules (Top):** Three panels showing NMR spectra. The first panel is labeled 'SpectrumDisplay:HN' and shows a 1D spectrum. The second panel is labeled 'SpectrumDisplay:HCN_1' and shows a 2D spectrum. The third panel is labeled 'SpectrumDisplay:HCN' and shows a 2D spectrum.
- Table Modules (Middle):** A table titled 'Backbone Assignment:1' with columns for '#', 'Index', 'Sequence', 'Type', and 'NmrAtoms'. The table contains data for residues 197, 198, 199, 200, and 201.
- Sequence Graph (Bottom):** A graph showing the sequence of residues. The sequence is 'Sec5:A H M R Q P' followed by 'T I R G E N'.

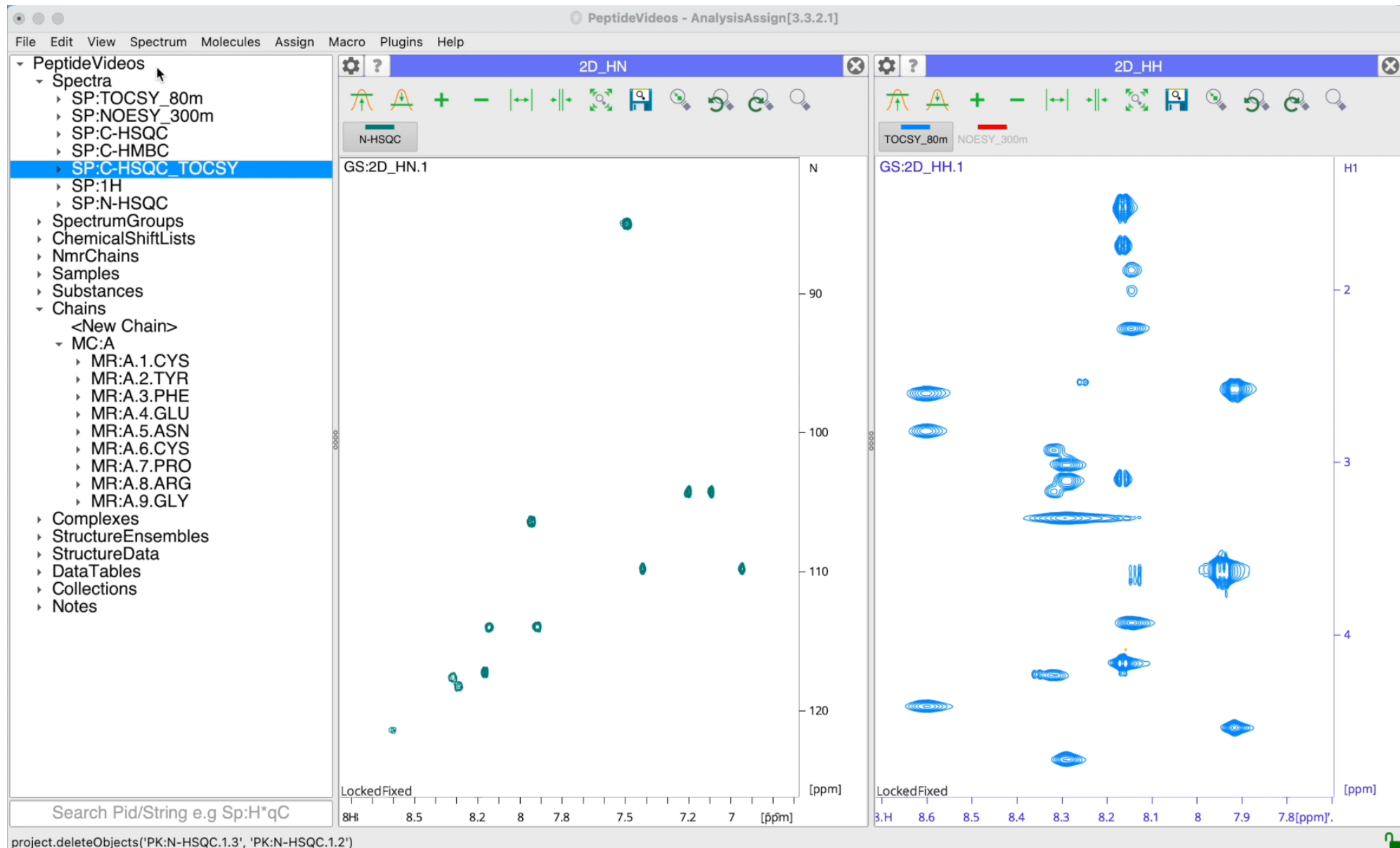
Annotations on the screenshot:

- A box labeled 'spectrum display modules' points to the three spectrum panels.
- A box labeled 'table modules' points to the 'Backbone Assignment:1' table.
- A box labeled 'specialised function modules' points to the 'Sequence Graph:1' section.

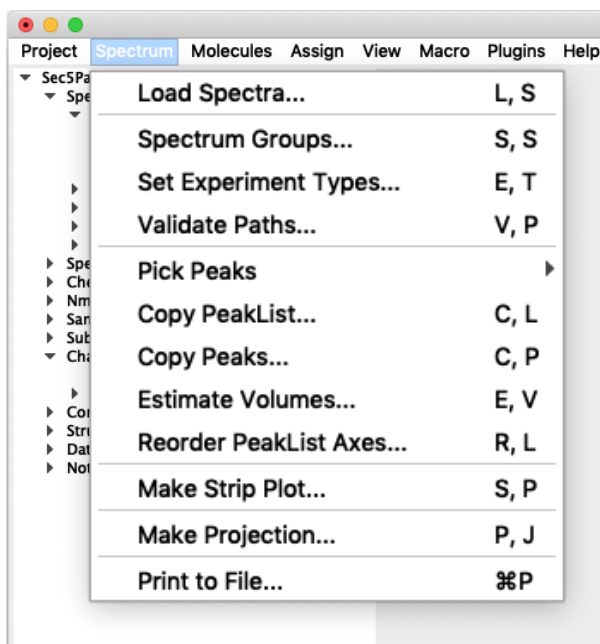
The Sidebar contains your data

Modules contain the functionalities to display your data

Assignment: Marking and Peak Picking



Keyboard Shortcuts



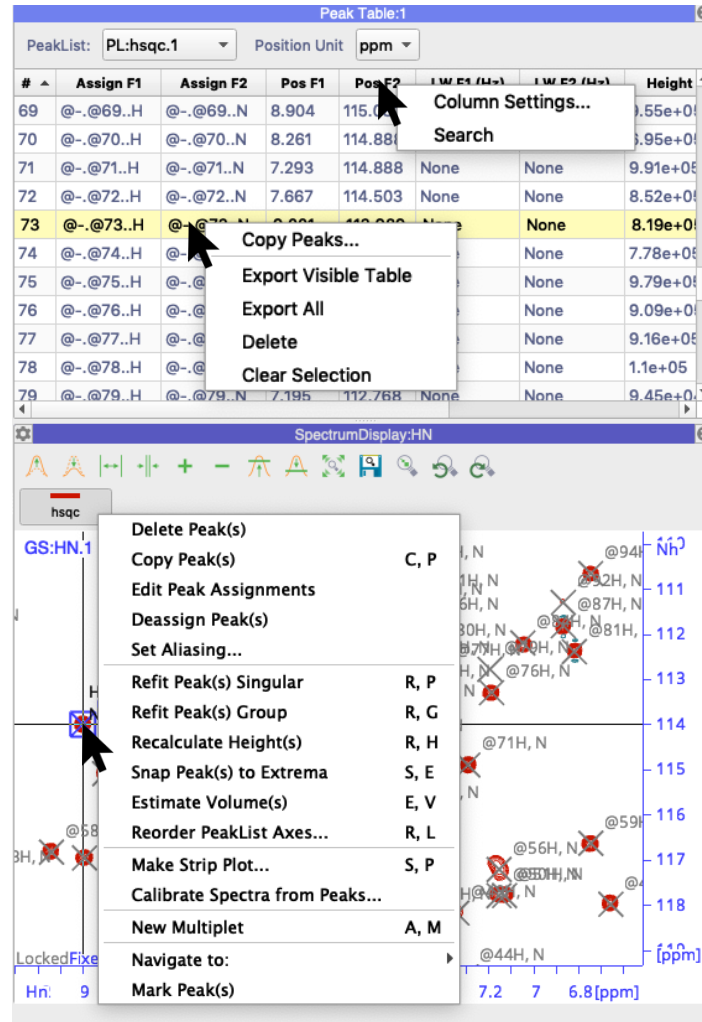
Two-key shortcuts are shown
in the menus

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:

Chain **Seq. Code** **AA type** **Atom**

A	▼	57	▼	TRP	▼	H	▼
---	---	----	---	-----	---	---	---

NmrChain
A

NmrResidue
A.57.TRP

NmrAtom
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

PeptideVideos - AnalysisAssign[3.3.2.1]

File Edit View Spectrum Molecules Assign Macro Plugins Help

PeptideVideos

- ▼ Spectra
 - SP:TOCSY_80m
 - SP:NOESY_300m
 - SP:C-HSQC
 - SP:C-HMBC
 - SP:C-HSQC_TOCSY
 - SP:1H
 - SP:N-HSQC**
- ▼ SpectrumGroups
- ▼ ChemicalShiftLists
- ▼ NmrChains
- ▼ Samples
- ▼ Substances
- ▼ Chains
 - <New Chain>
 - ▼ MC:A
 - MR:A.1.CYS
 - MR:A.2.TYR
 - MR:A.3.PHE
 - MR:A.4.GLU
 - MR:A.5.ASN
 - MR:A.6.CYS
 - MR:A.7.PRO
 - MR:A.8.ARG
 - MR:A.9.GLY
- ▼ Complexes
- ▼ StructureEnsembles
- ▼ StructureData
- ▼ DataTables
- ▼ Collections
- ▼ Notes

Search Pid/String e.g Sp:H*qC

2D_HN

N-HSQC

GS:2D_HN.1

LockedFixed

3.H 8.1 8 7.9 7.8 7.7 [ppm]

100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116 [ppm]

2D_HH

TOCSY_80m NOESY_300m

GS:2D_HH.1

LockedFixed

H 8.02 8 7.98 7.95 7.92 7.9 7.88 [ppm]

3.3
3.4
3.5
3.6
3.7
3.8
3.9
4
4.1
4.2
4.3
4.4 [ppm]

Peak Assigner

Current Peak: N-HSQC.1.4

H: 7.937

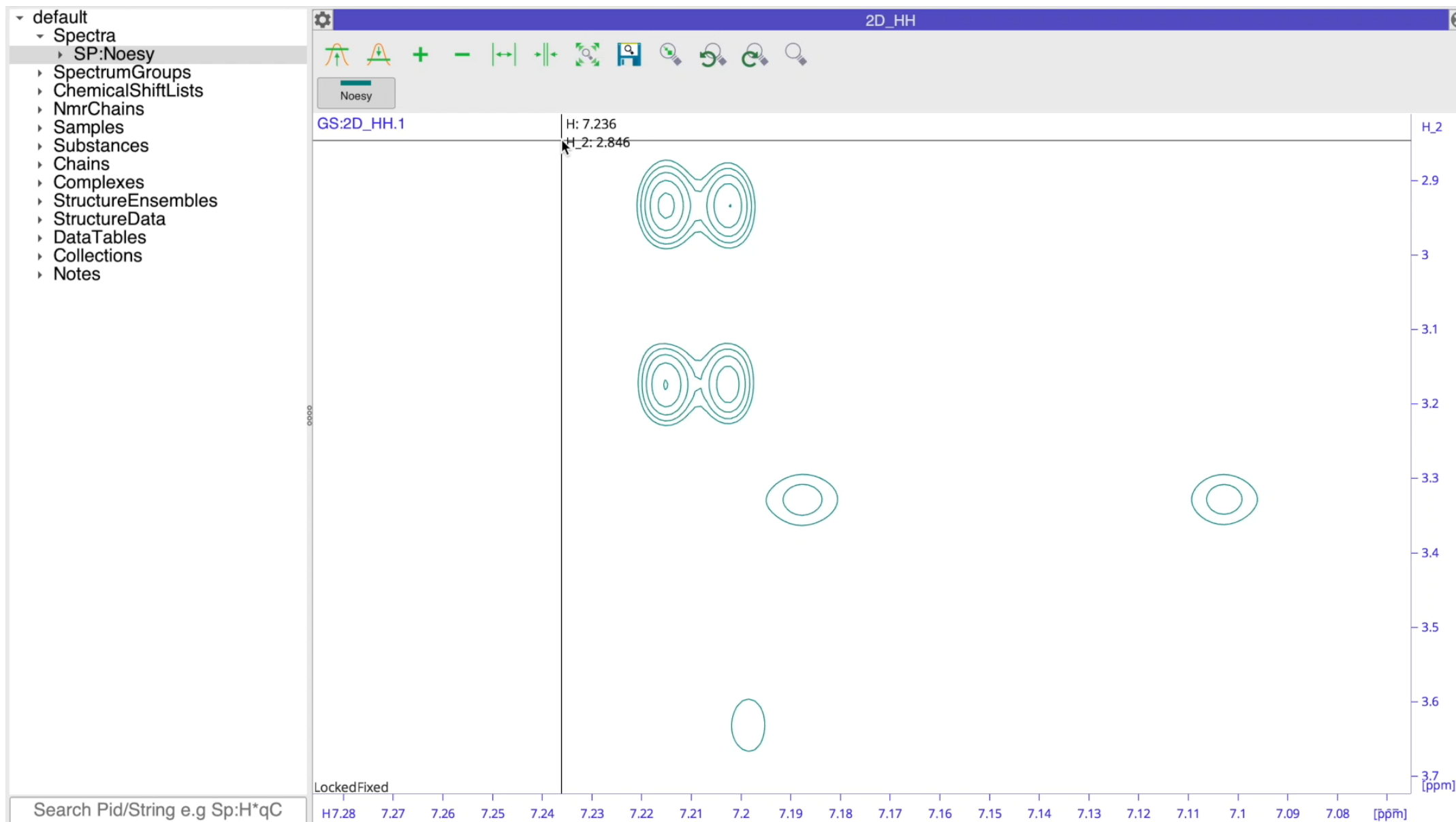
N: 106.393

Alternatives

Alternatives

application.showPeakAssigner(position='bottom', relativeTo=None)

Multiplets



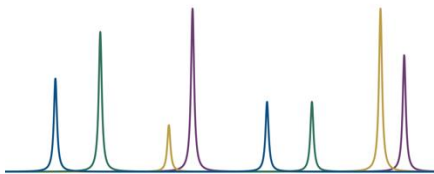
Small Molecule NMR



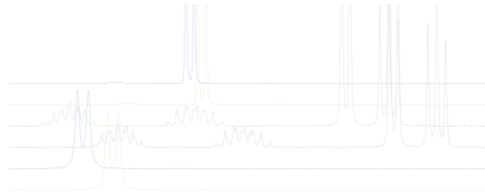
Peptides



Spectrum Visualisation
and Navigation



Screening

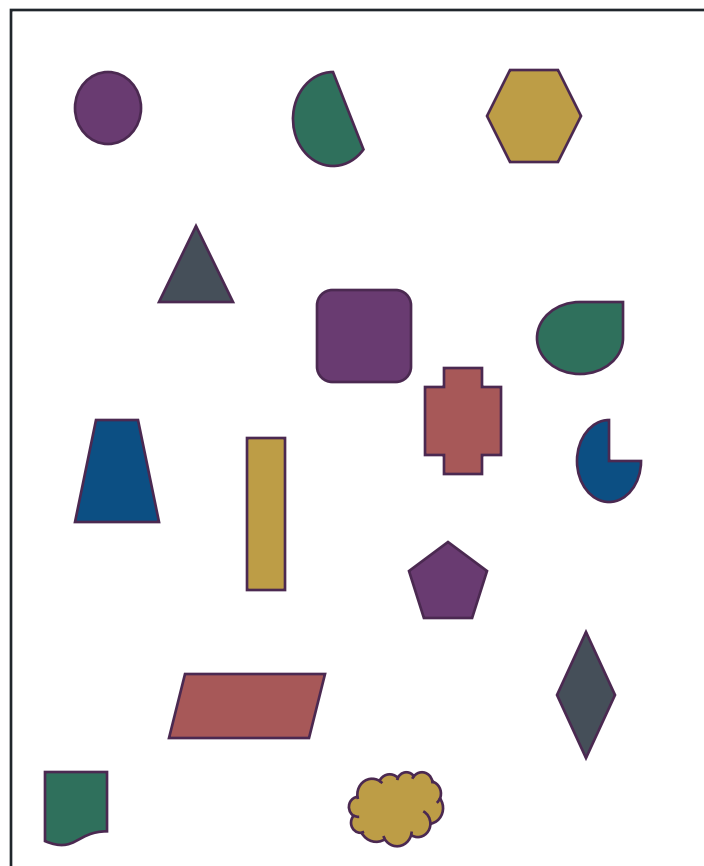


Metabolomics

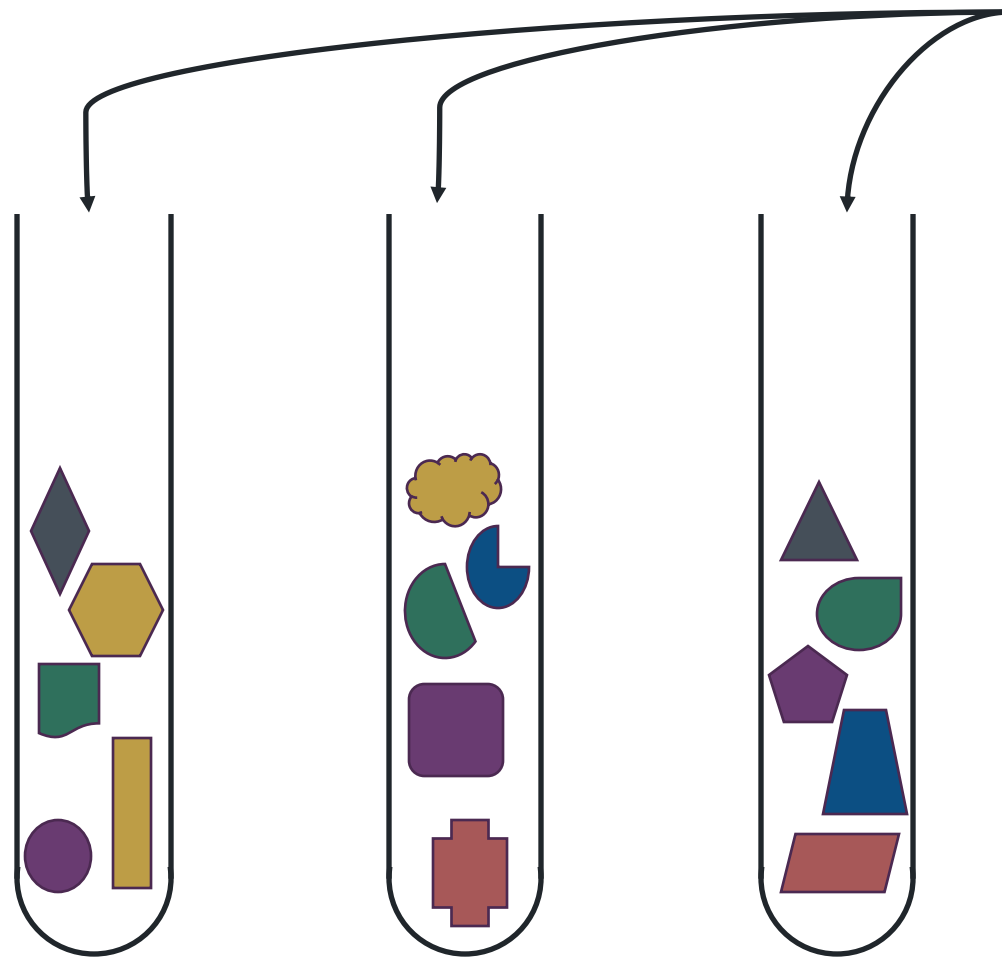
```
Macro Name: None
1 for peak in project.peaks:
2     # Get the assignments in all dimensions for a peak
3     for assignOptions in peak.assignOptions:
4         for assignment in assignOptions:
5             # Create the new assignment string with 'molecule' as the NmrChain
6             temp = str(assignment)
7             assignmentComponents = temp.split(':')
8             assignmentComponents[0] = 'NA_molecule'
9             assignmentComponents[3] = assignmentComponents[3][0:-1]
10            newAssignment = ':'.join(assignmentComponents)
11            # Get parameters needed for creation of new NmrChain, NmrResidue, NmrAtom and Peak Assignment
12            chainPid = 'NA_molecule'
13            resPid = 'NA_molecule' + ':'.join(assignmentComponents[1:-1])
14            seqCode = assignmentComponents[1]
15            resType = assignmentComponents[2]
16            atomName = assignmentComponents[3]
17            axCde = assignmentComponents[3][0]
18            # Create new NmrChain, NmrResidue and NmrAtom if necessary
19            project.fetchNmrChain('molecule')
20            getChainPid().fetchNmrResidue(seqCode, resType)
21            getResPid().fetchNmrAtom(atomName)
22            # Change the peak assignment to the new one
23            getPeakPid().assignDimension(axCde, newAssignment)
24
```

Macro Writing

NMR-based fragment screening



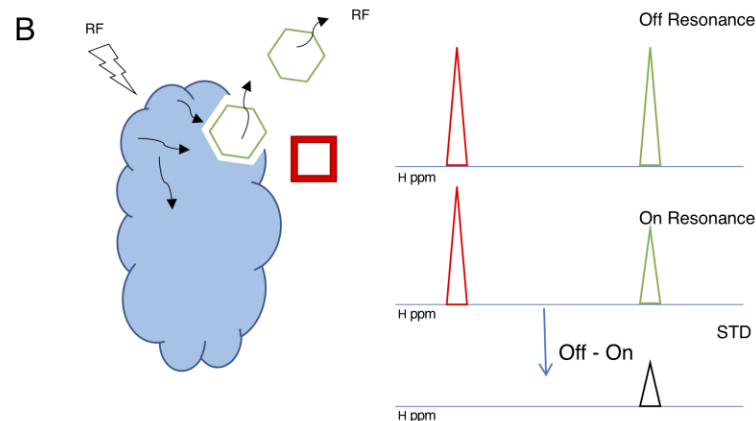
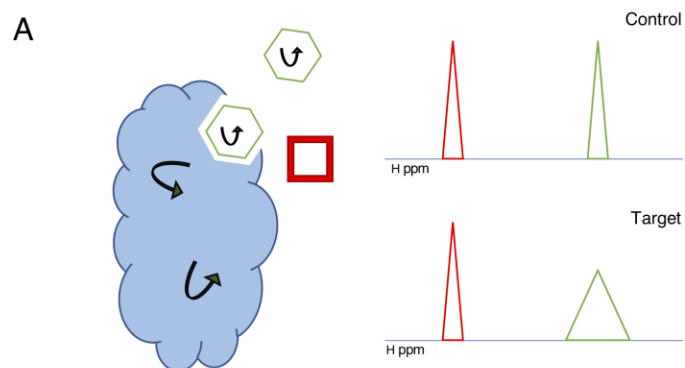
Library
(500-5000 compounds)



Mixtures

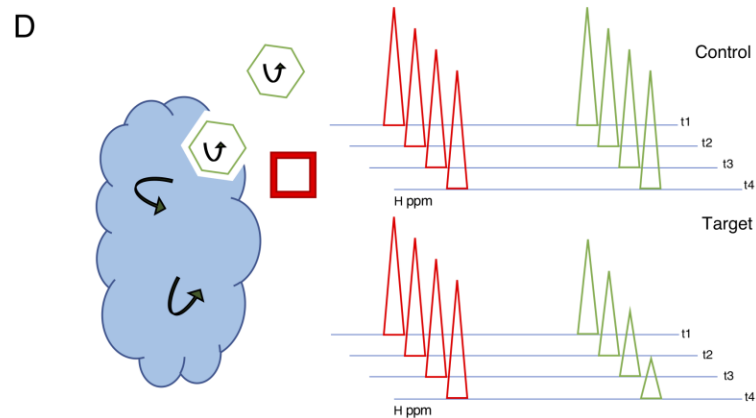
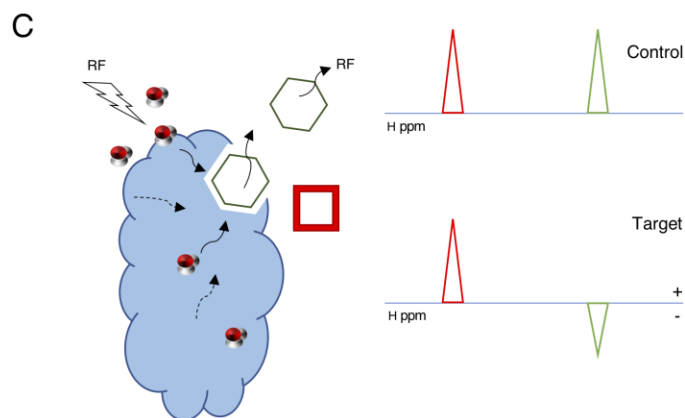
NMR-based fragment screening

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



Saturation Transfer Difference (STD)

WaterLOGSY

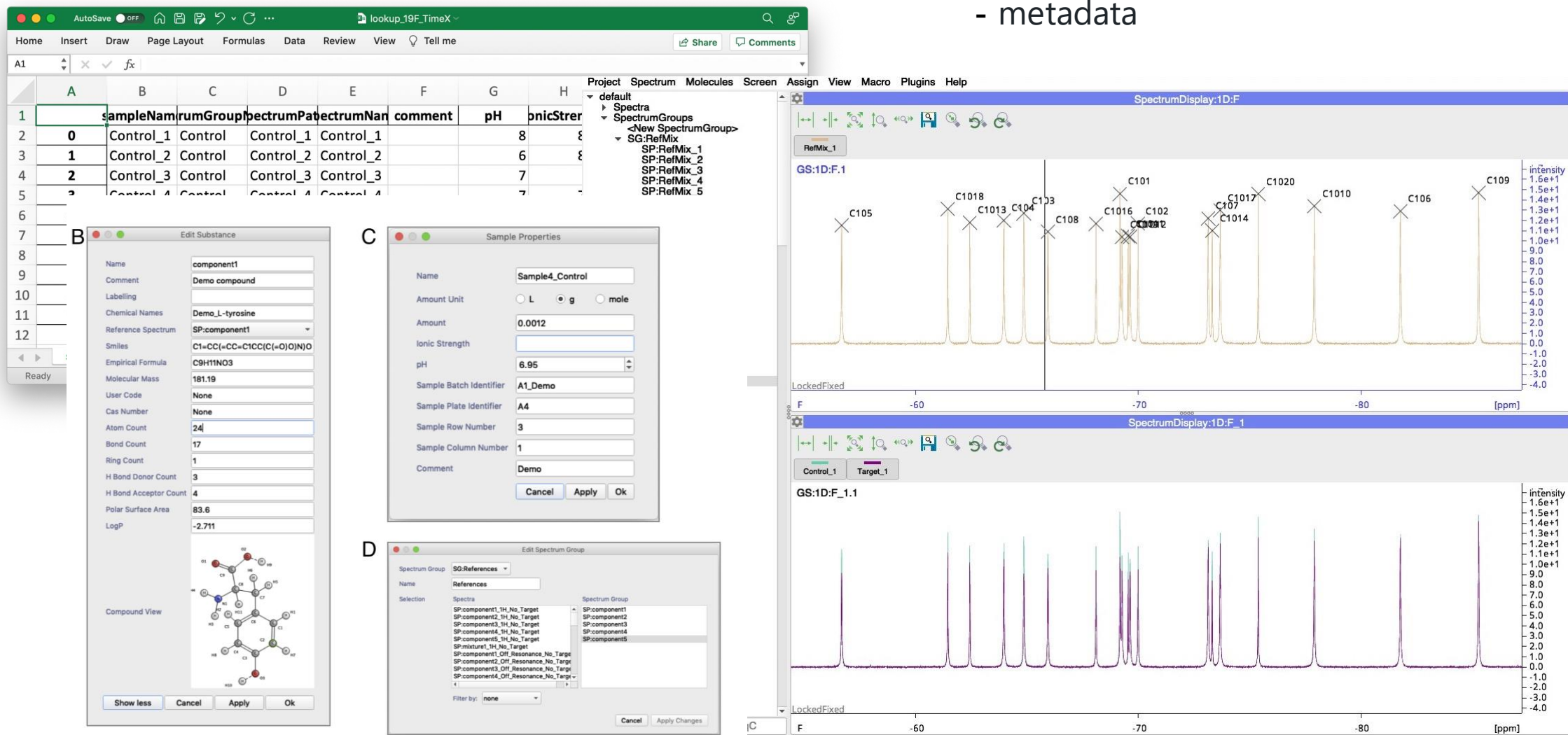


T_{1Q} series

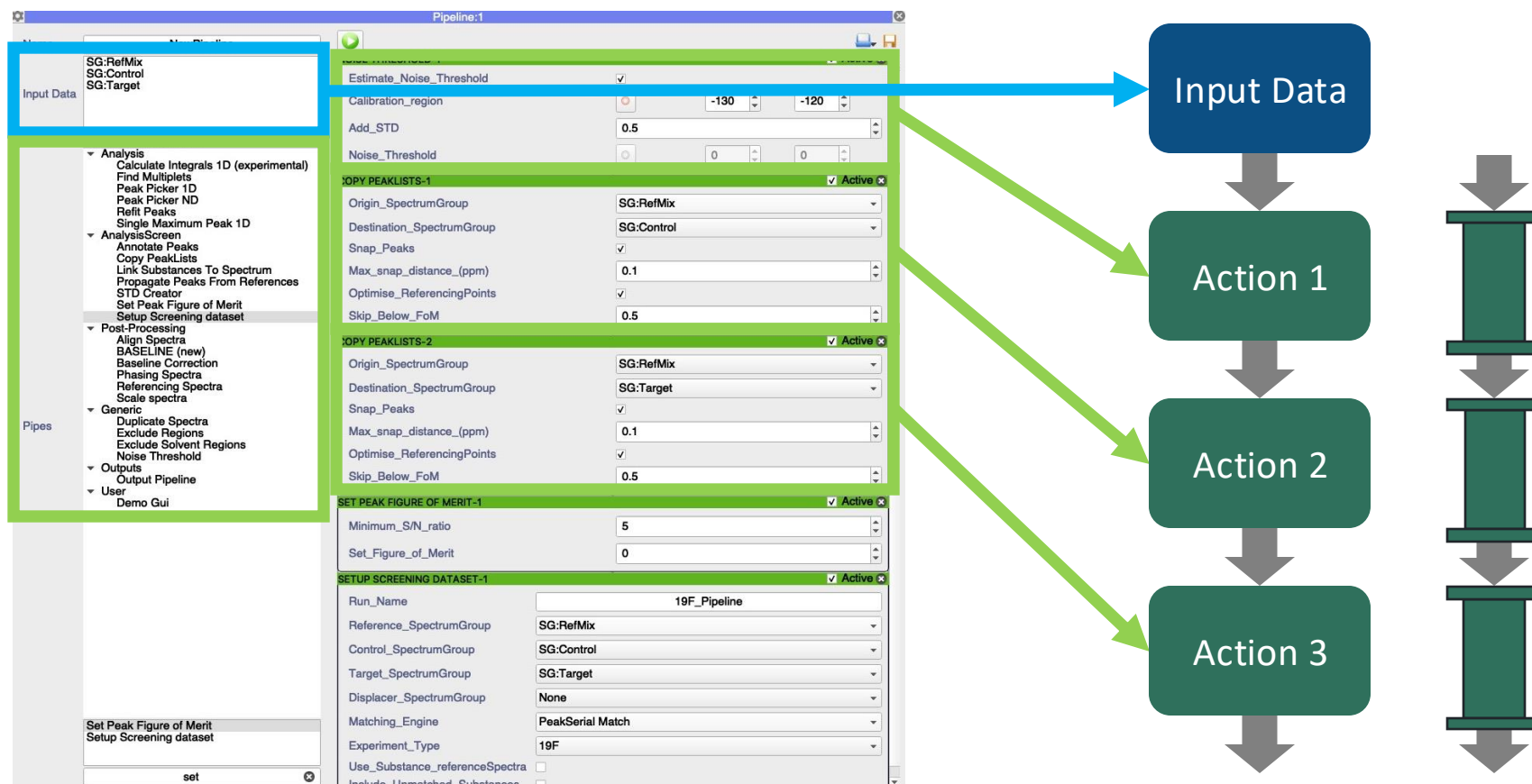
Loading data

Excel/CSV/NEF

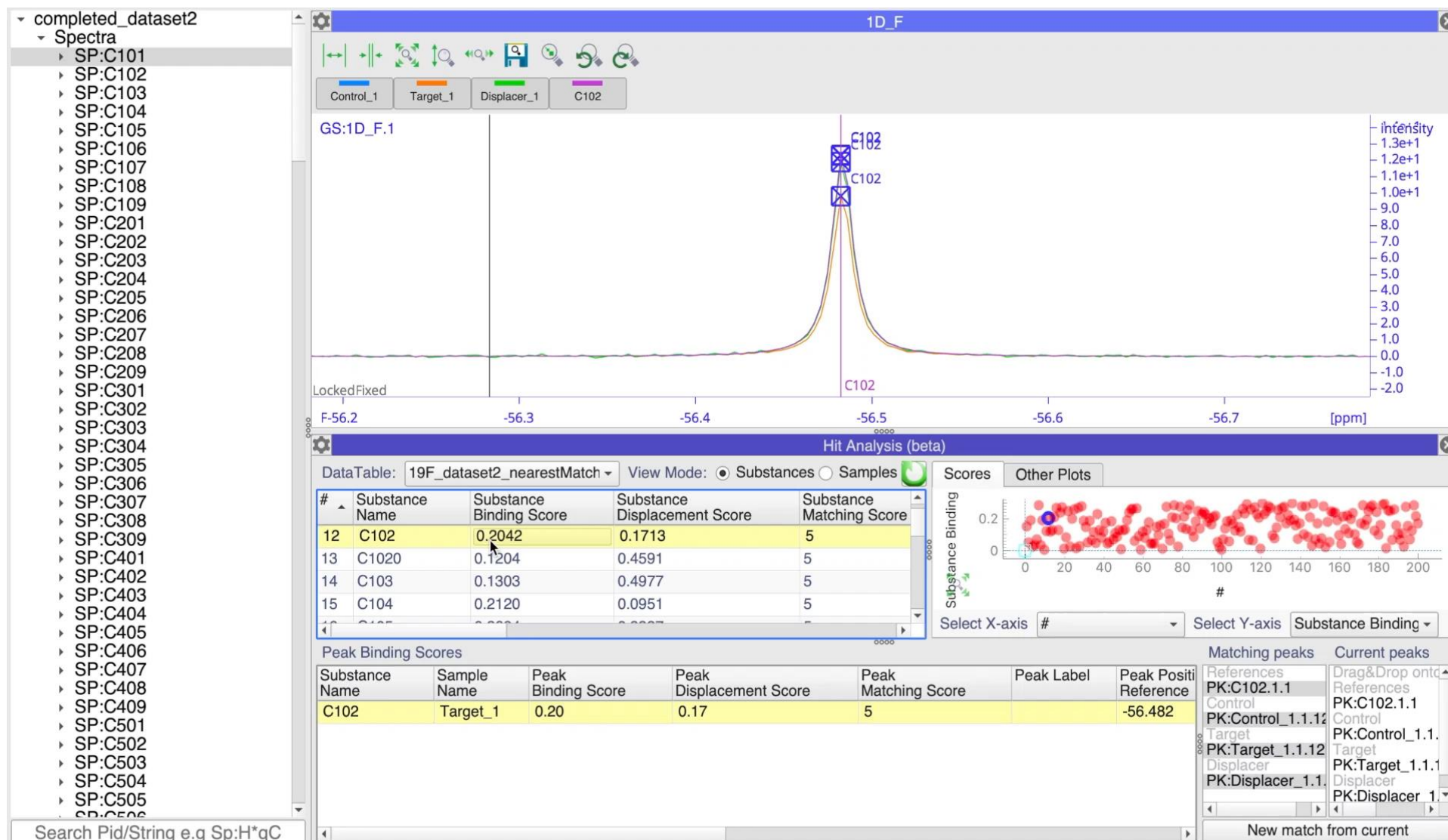
- reference spectra
- compound information
- metadata



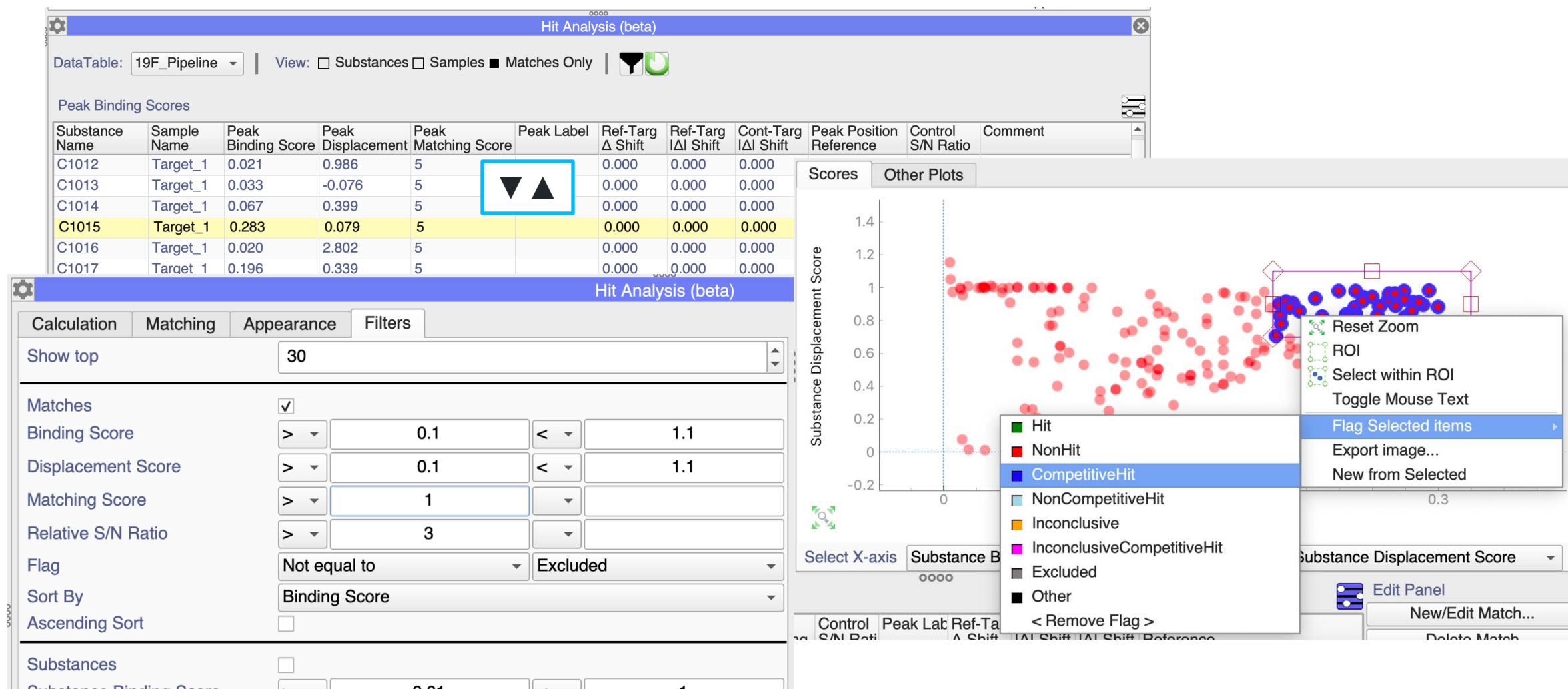
Automation: pipelines



User-friendly Hit Analysis

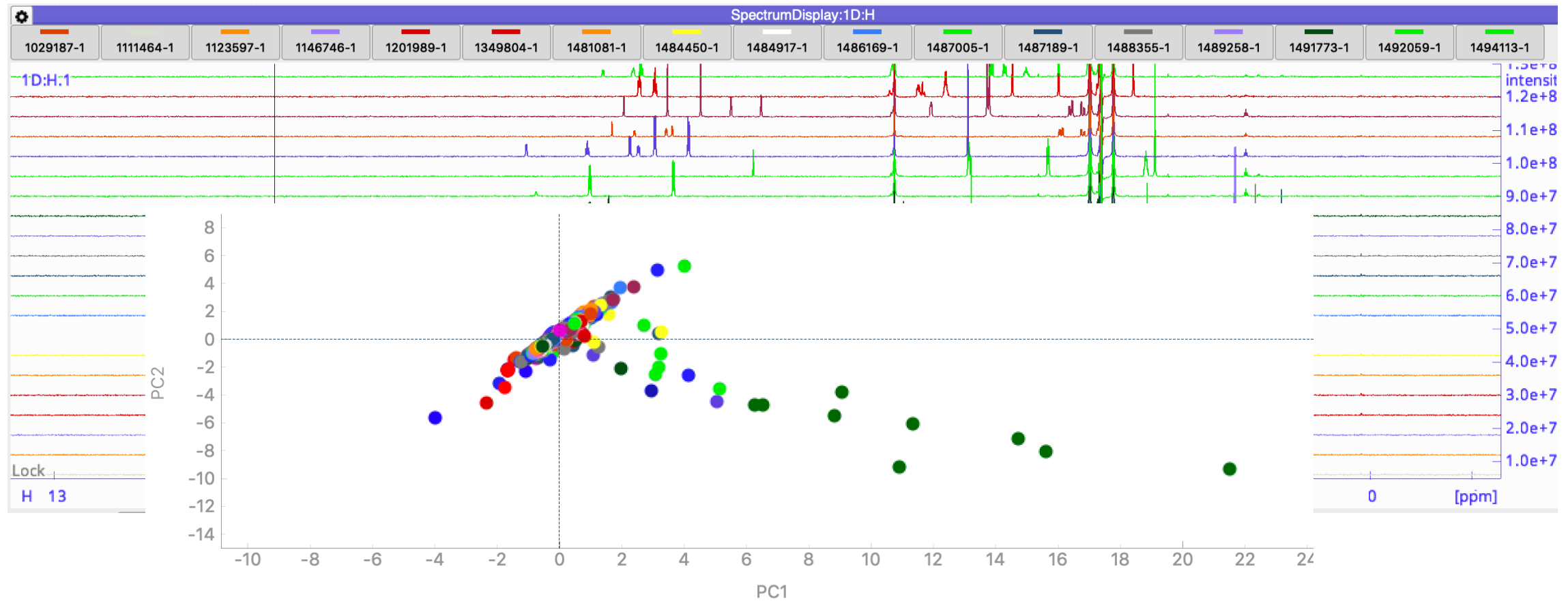


User-friendly Hit Analysis



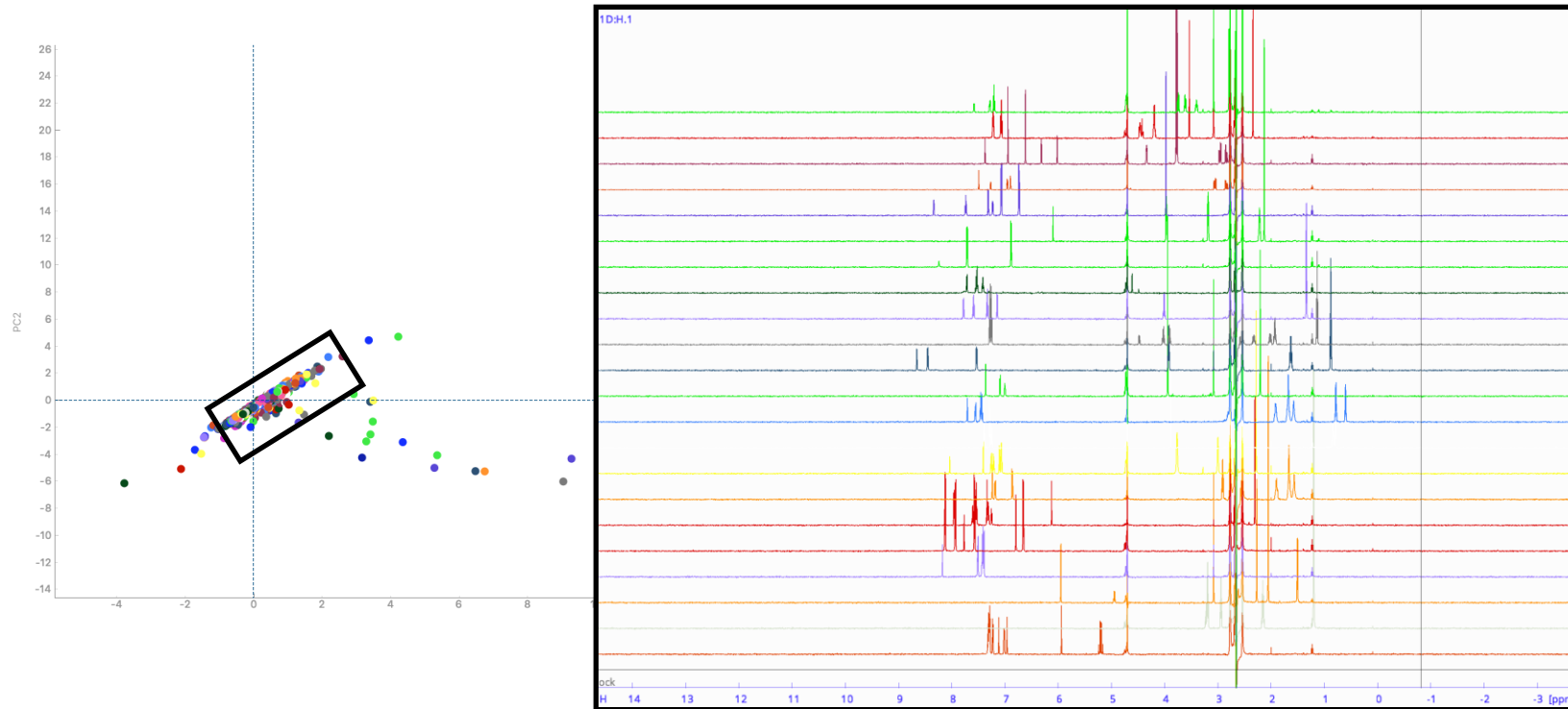
Quality control

Reference data Principal Component Analysis (PCA)



Quality control

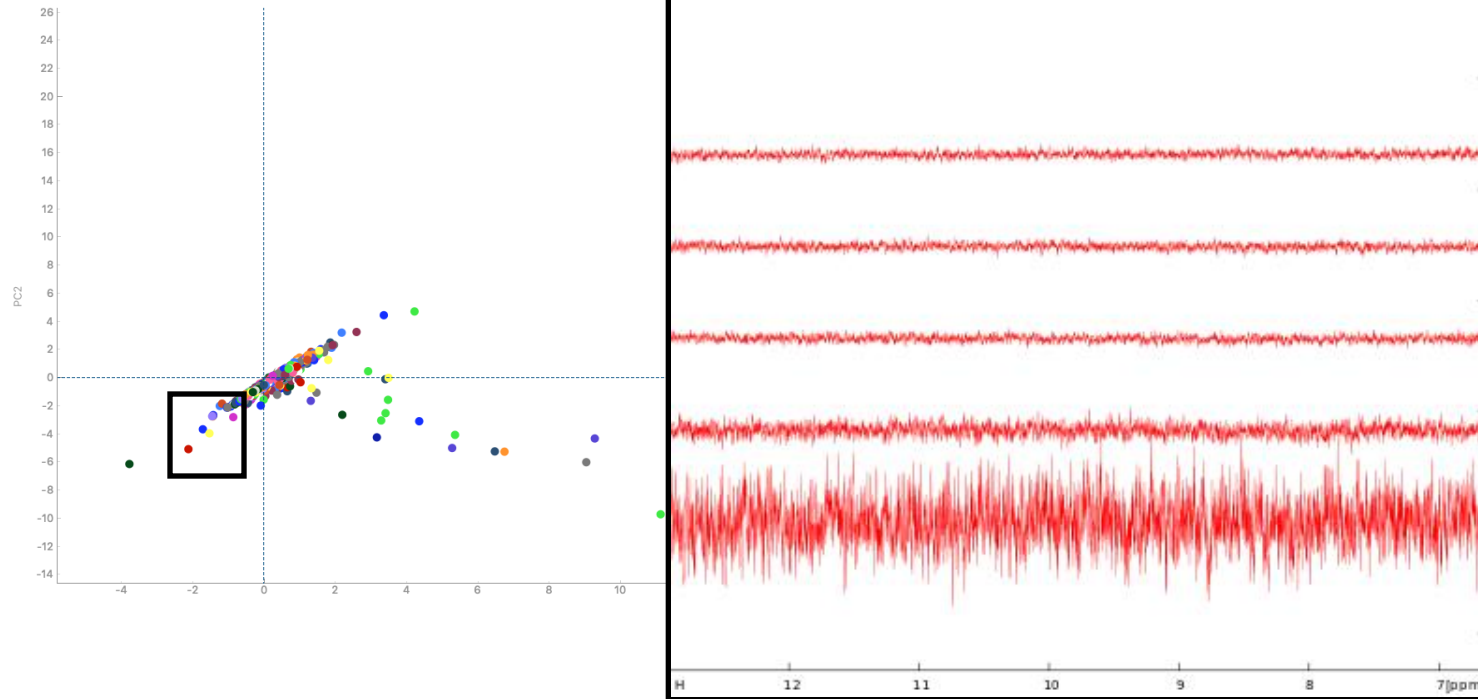
Reference data PCA



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

Quality control

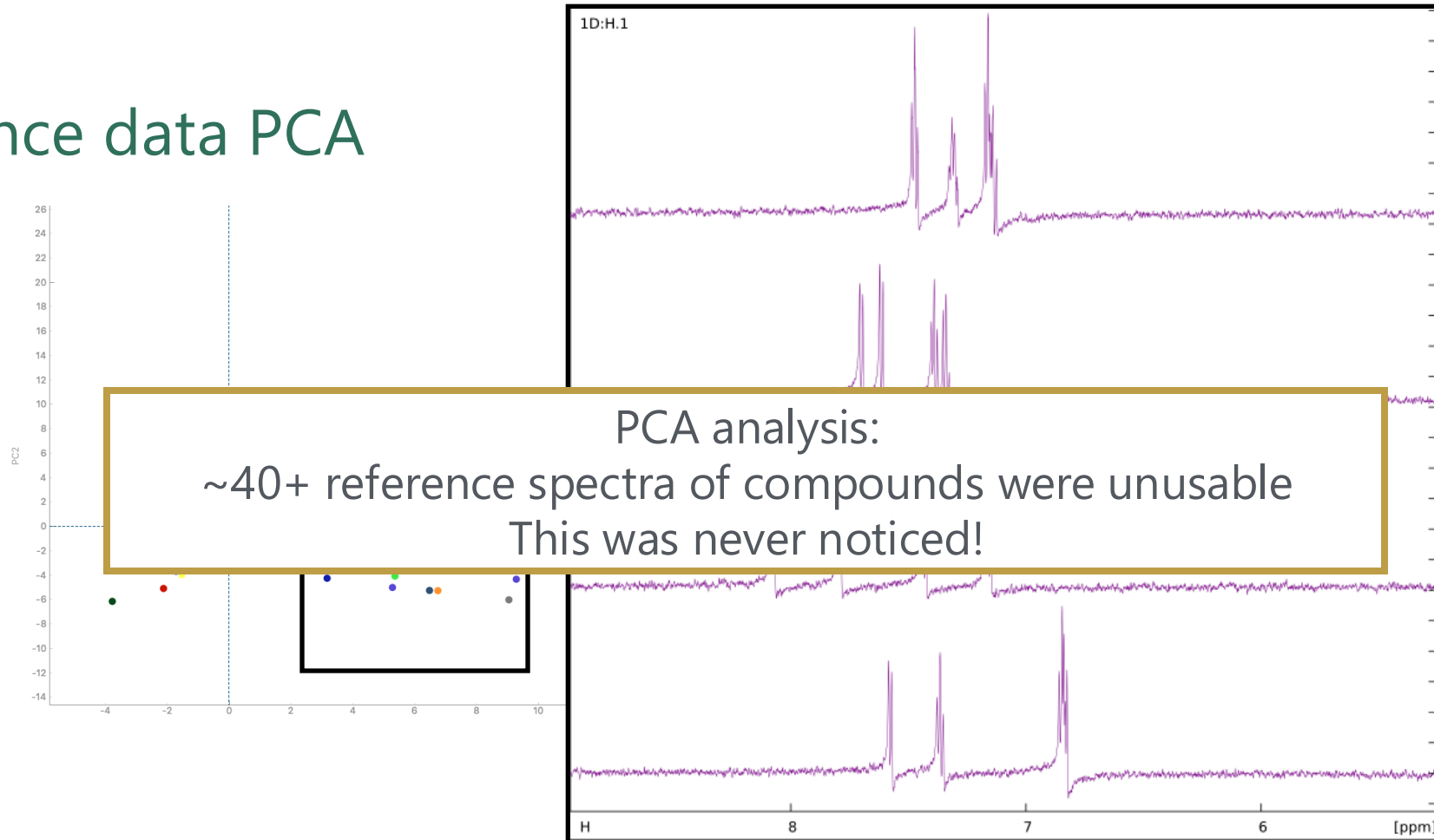
Reference data PCA



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

Quality control

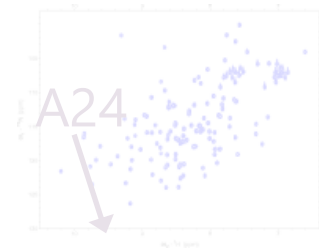
Reference data PCA



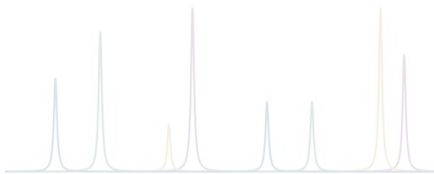
Small Molecule NMR



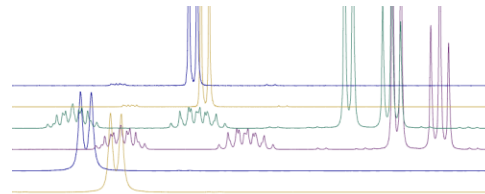
Peptides



Spectrum Visualisation
and Navigation



Screening

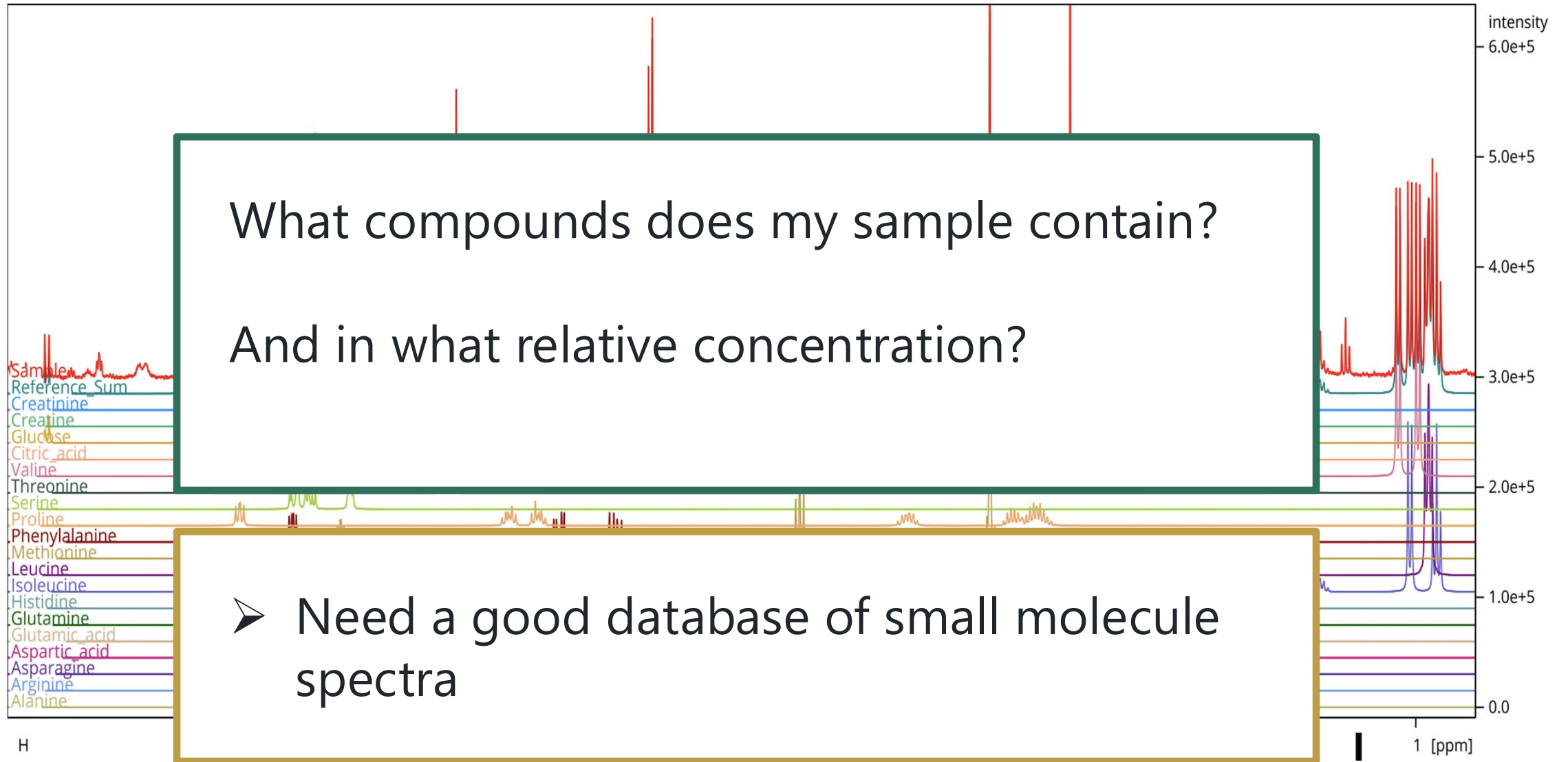


Metabolomics

```
Macro Editor 1
Macro Name: None
1 for peak in project.peaks:
2 # Get the assignments in all dimensions for a peak
3 for assignOptions in peak.assignNewAtoms:
4 for assignment in assignOptions:
5 # Create the new assignment string with 'molecule' as the NmrChain
6 temp = str assignment
7 assignmentComponents = temp.split(':')
8 assignmentComponents[0] = 'NA_molecule'
9 assignmentComponents[3] = assignmentComponents[3][0:-1]
10 newAssignment = ':'.join assignmentComponents
11 # Get parameters needed for creation of new NmrChain, NmrResidue, NmrAtom and Peak Assignment
12 chainPid = 'NA_molecule' + ':'.join assignmentComponents[1:-1]
13 resPid = 'NA_molecule' + ':'.join assignmentComponents[1:-1]
14 seqCode = assignmentComponents[1]
15 resType = assignmentComponents[2]
16 atomName = assignmentComponents[3]
17 axCde = assignmentComponents[3][0]
18 # Create new NmrChain, NmrResidue and NmrAtom if necessary
19 project.fetchNmrChain('molecule')
20 getChainPid().fetchNmrResidue(seqCode, resType)
21 getResPid().fetchNmrAtom(atomName)
22 # Change the peak assignment to the new one
23 getPeakPid().assignDimension(axCde, newAssignment)
24
```

Macro Writing

Metabolomics



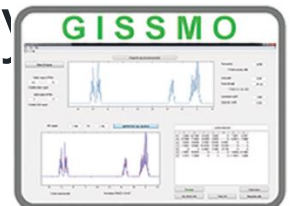
Metabolomics Databases

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)

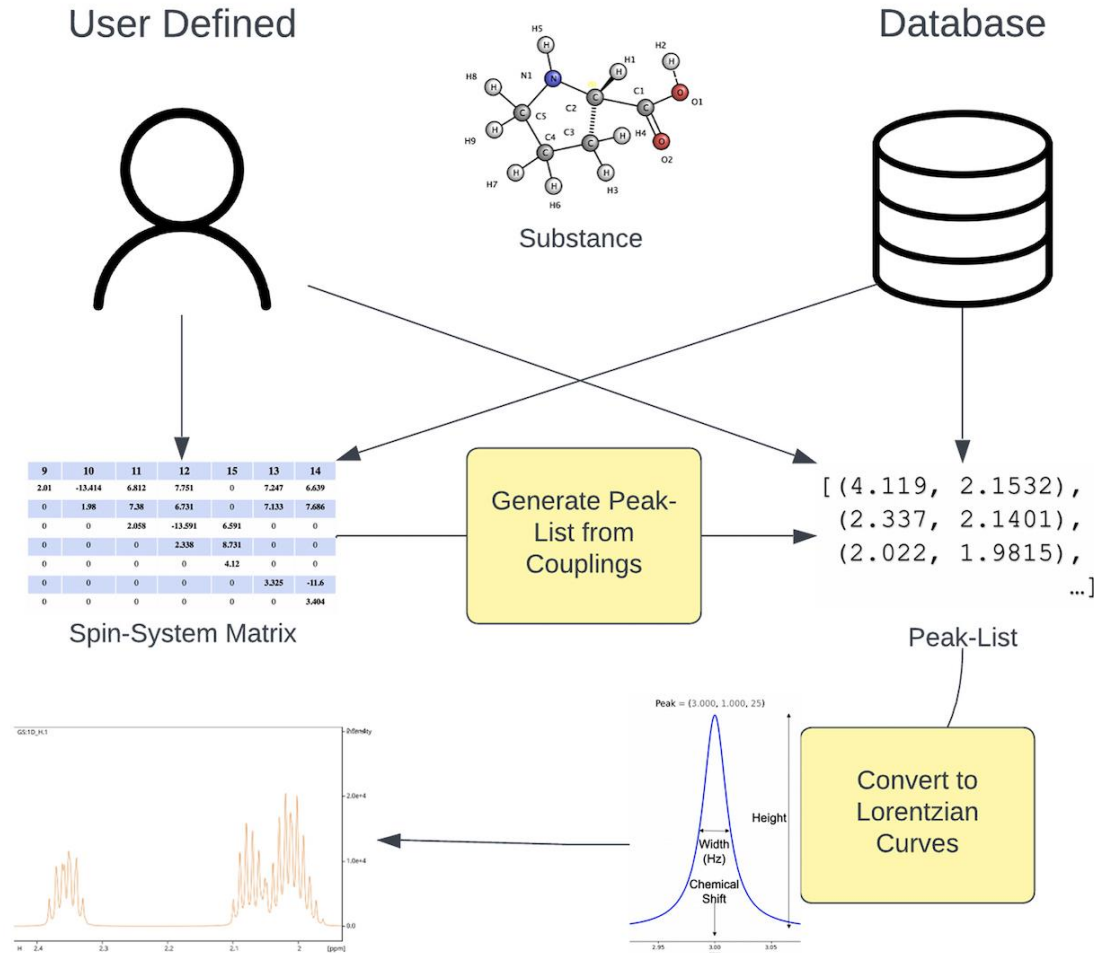


Metabolomics Databases

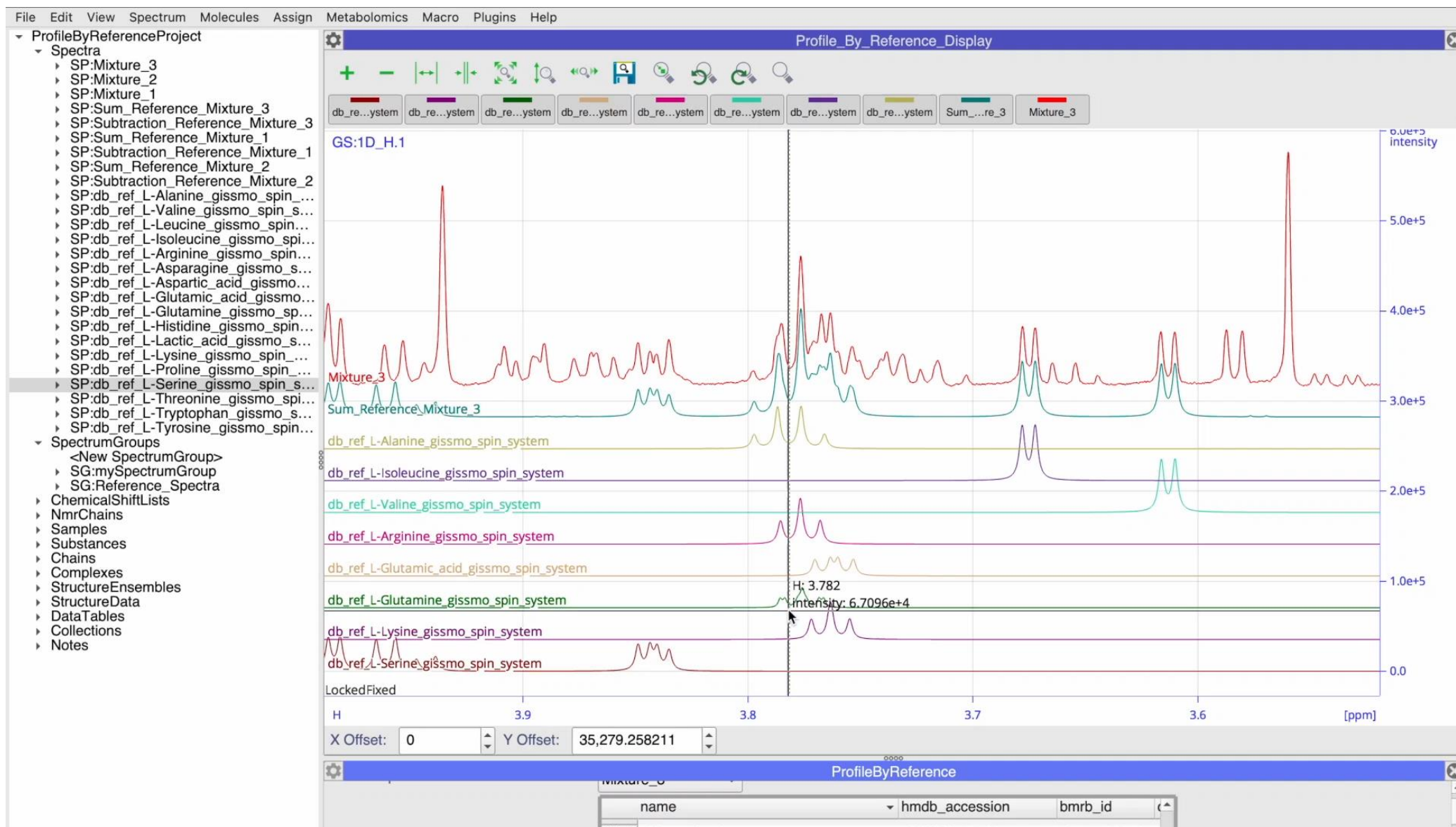
Deposited metabolite NMR spectra will vary depending on:

- 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

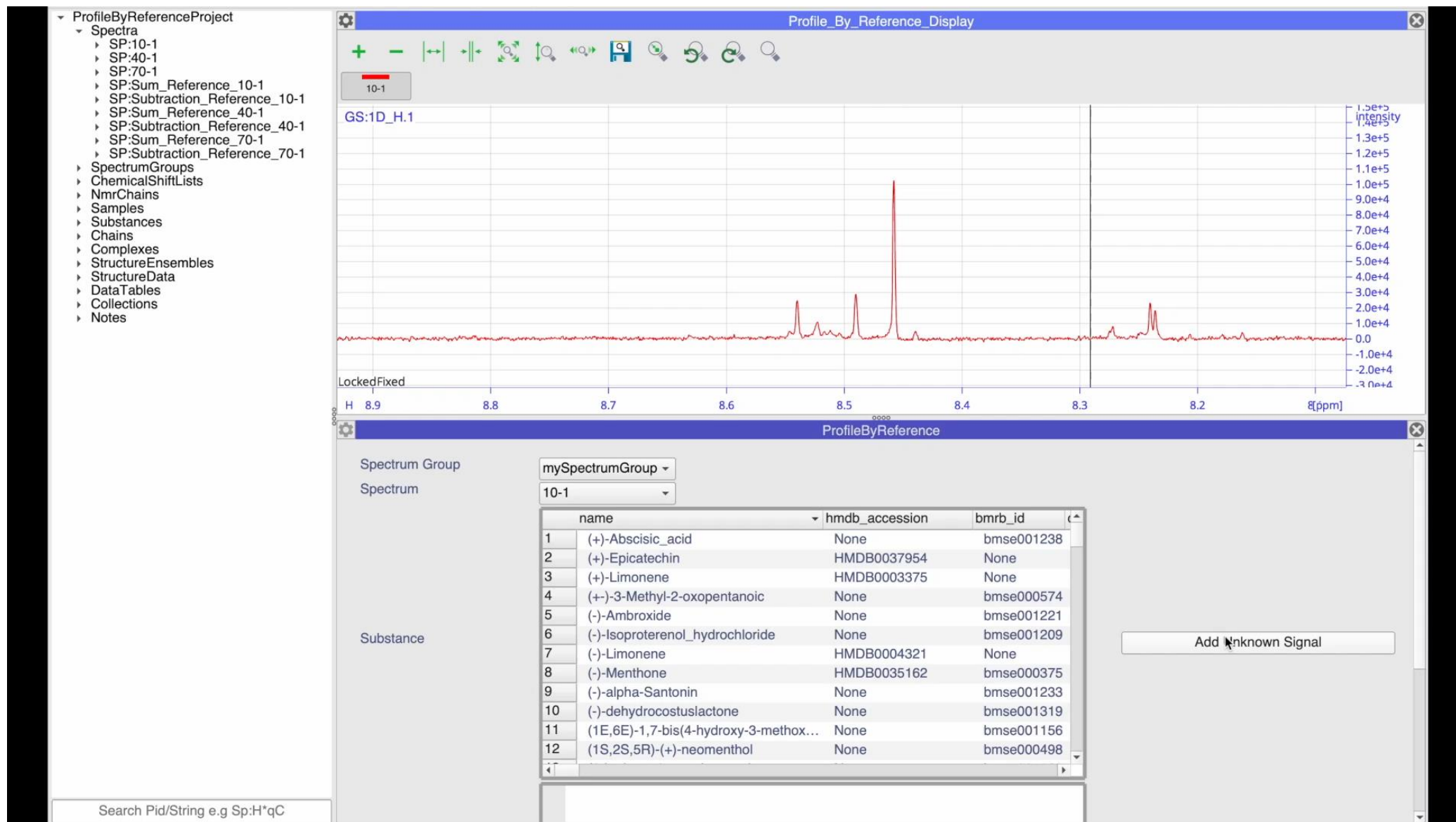
CcpNmr Analysis Simulated Metabolite DataBase (CASMDB)



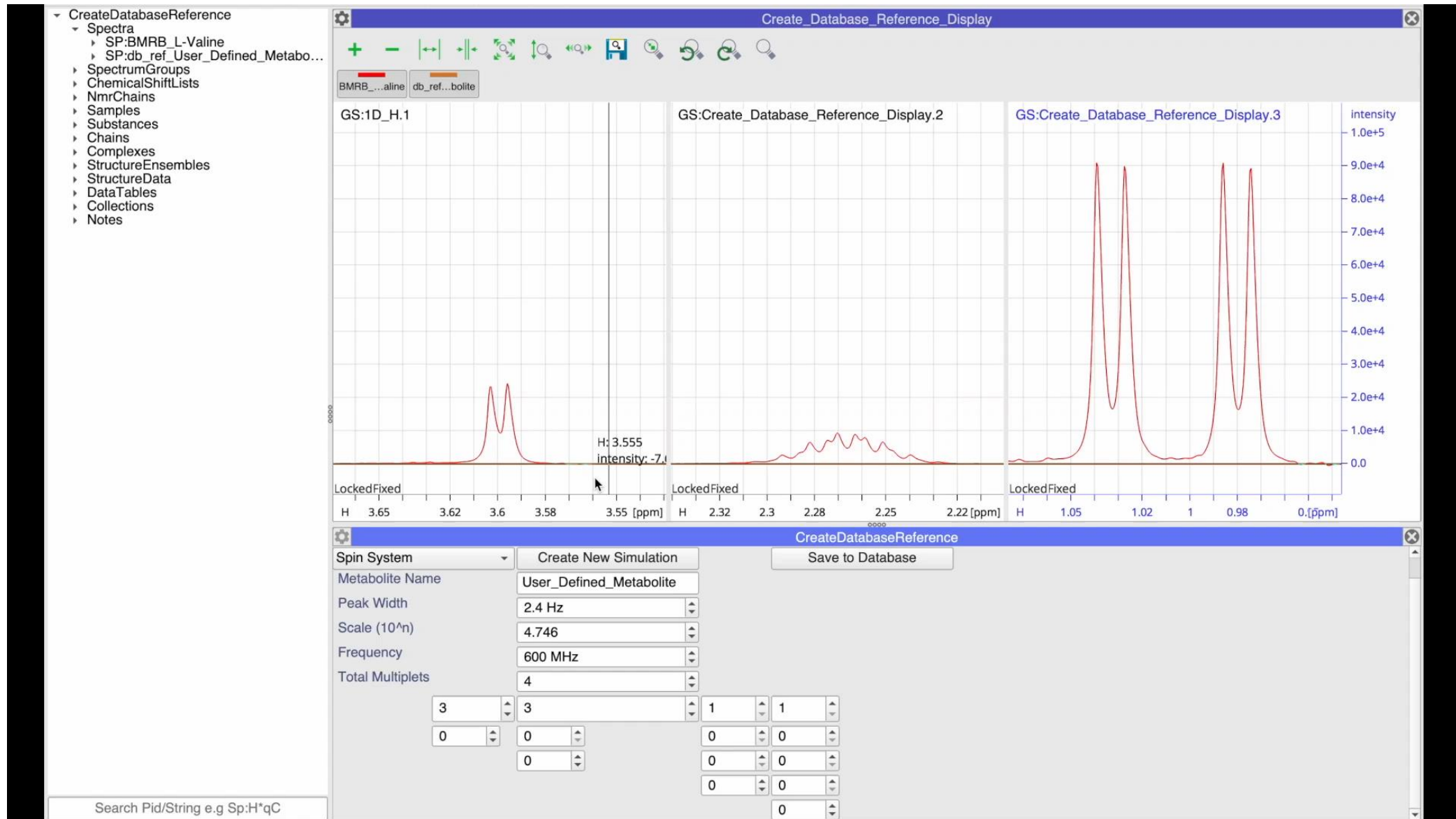
AnalysisMetabolomics



AnalysisMetabolomics



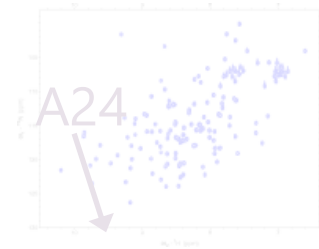
AnalysisMetabolomics



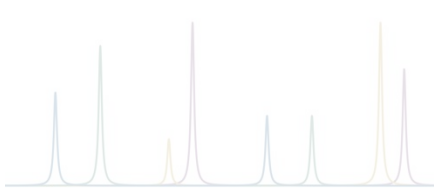
Small Molecule NMR



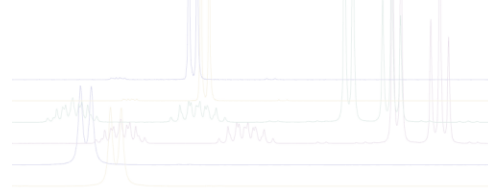
Peptides



Spectrum Visualisation
and Navigation



Screening



Metabolomics

```
Macro Editor:1
Macro Name: None
1 for peak in project.peaks:
2 # Get the assignments in all dimensions for a peak
3 for assignOptions in peak.assignedNmrAtoms:
4 for assignment in assignOptions:
5 # Create the new assignment string with 'molecule' as the NmrChain
6 temp = str assignment
7 assignmentComponents = temp.split('.')
8 assignmentComponents[0] = 'NA.molecule'
9 assignmentComponents[3] = assignmentComponents[3][0:-1]
10 newAssignment = '.'.join assignmentComponents
11 # get parameters needed for creation of new NmrChain, NmrResidue, NmrAtom and Peak Assignment
12 chainPid = 'NC.molecule' + '.'.join assignmentComponents[1:-1]
13 resPid = 'NB.molecule' + '.'.join assignmentComponents[1:-1]
14 seqCode = assignmentComponents[1]
15 resType = assignmentComponents[2]
16 atomName = assignmentComponents[3]
17 axCde = assignmentComponents[3][0]
18 # Create new NmrChain, NmrResidue and NmrAtom if necessary
19 project.fetchNmrChain('molecule')
20 get(chainPid).fetchNmrResidue(seqCode, resType)
21 get(resPid).fetchNmrAtom(atomName)
22 # Change the peak assignment to the new one
23 get(peak.pid).assignDimension(axCde, newAssignment)
24
```

Macro Writing

Analysis V3 as a development platform



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



LMFIT



Open-Source Cheminformatics
and Machine Learning



Macro Writing Tutorials

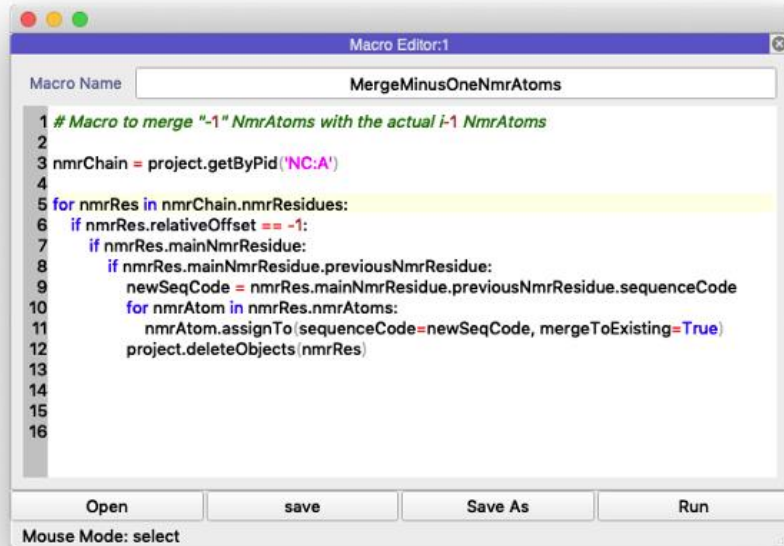
CCPN

CcpNmr Analysis Version 3

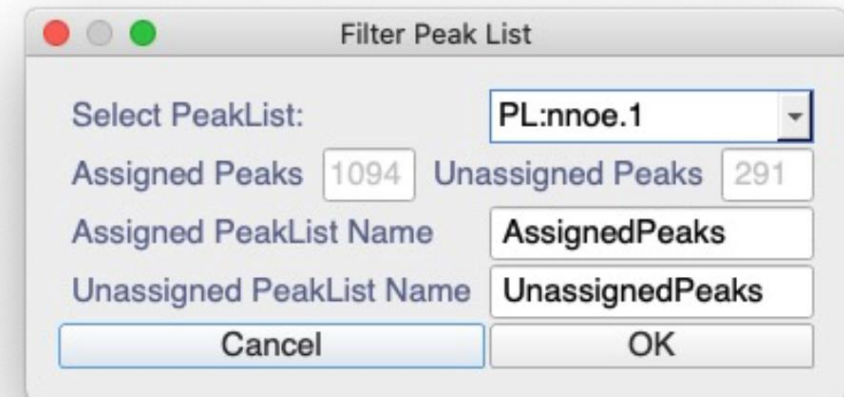
CCPN

CcpNmr Analysis Version 3

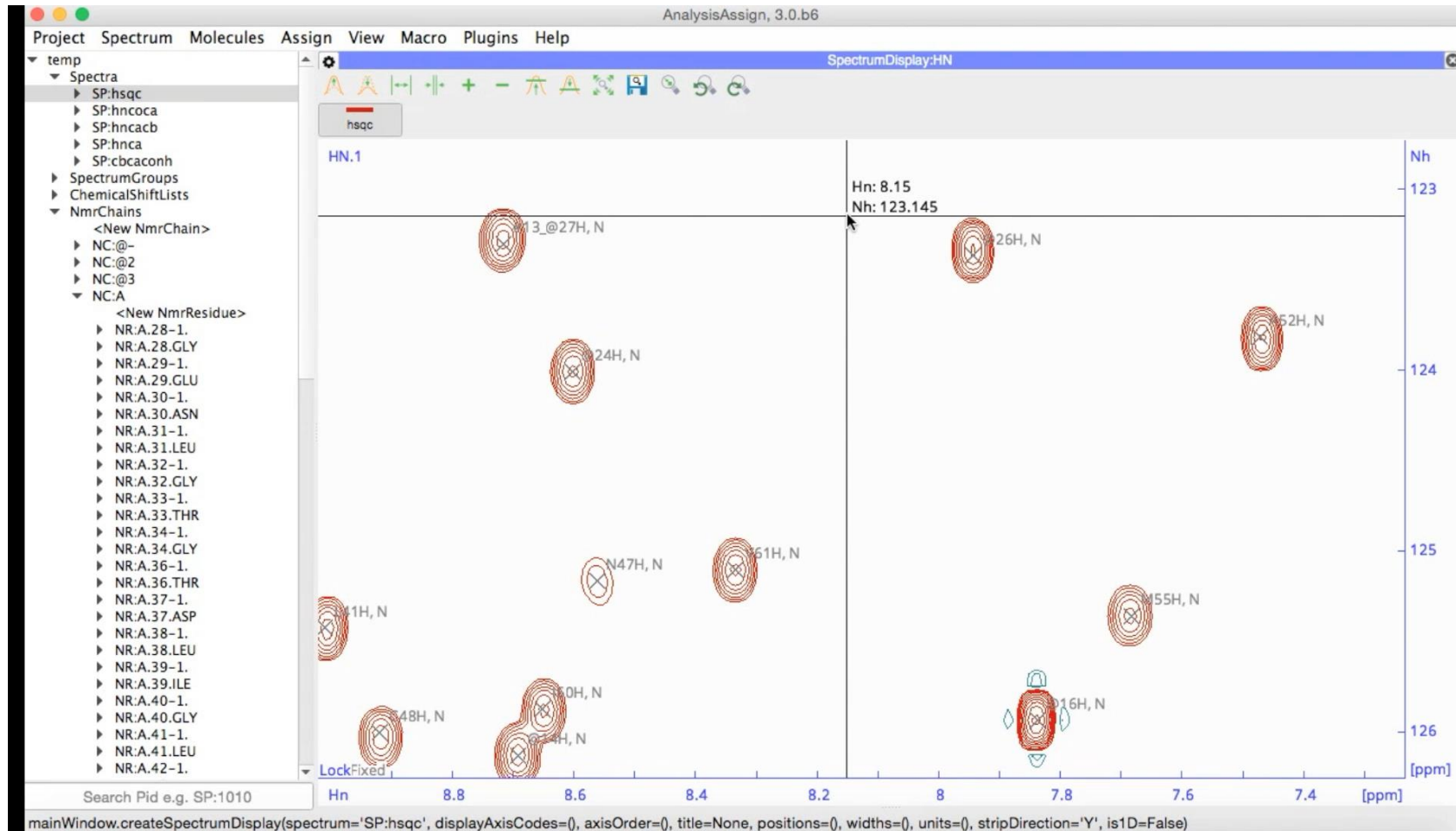
Macro Writing Tutorial



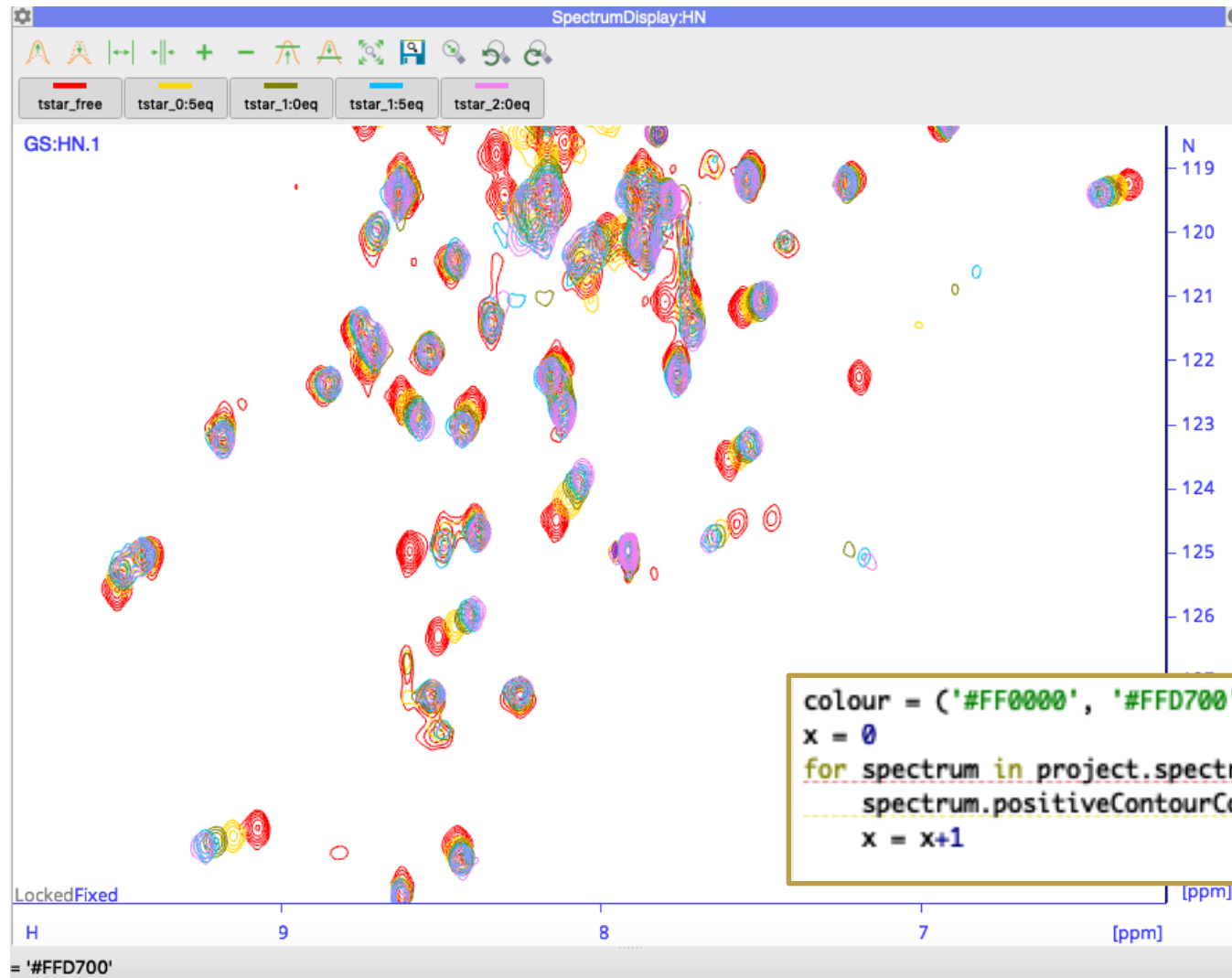
GUI Macro Writing Tutorial



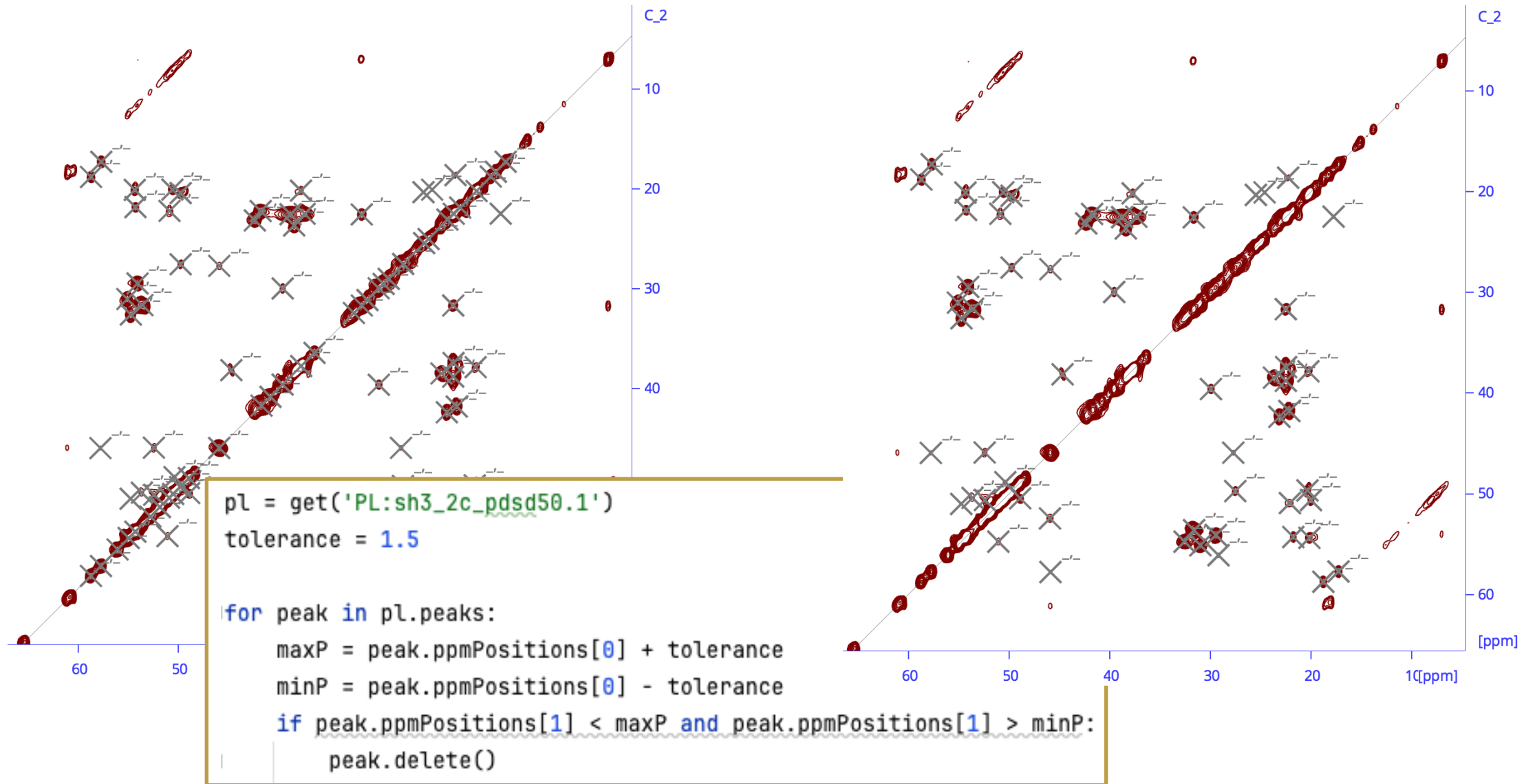
GUI command echo & python console



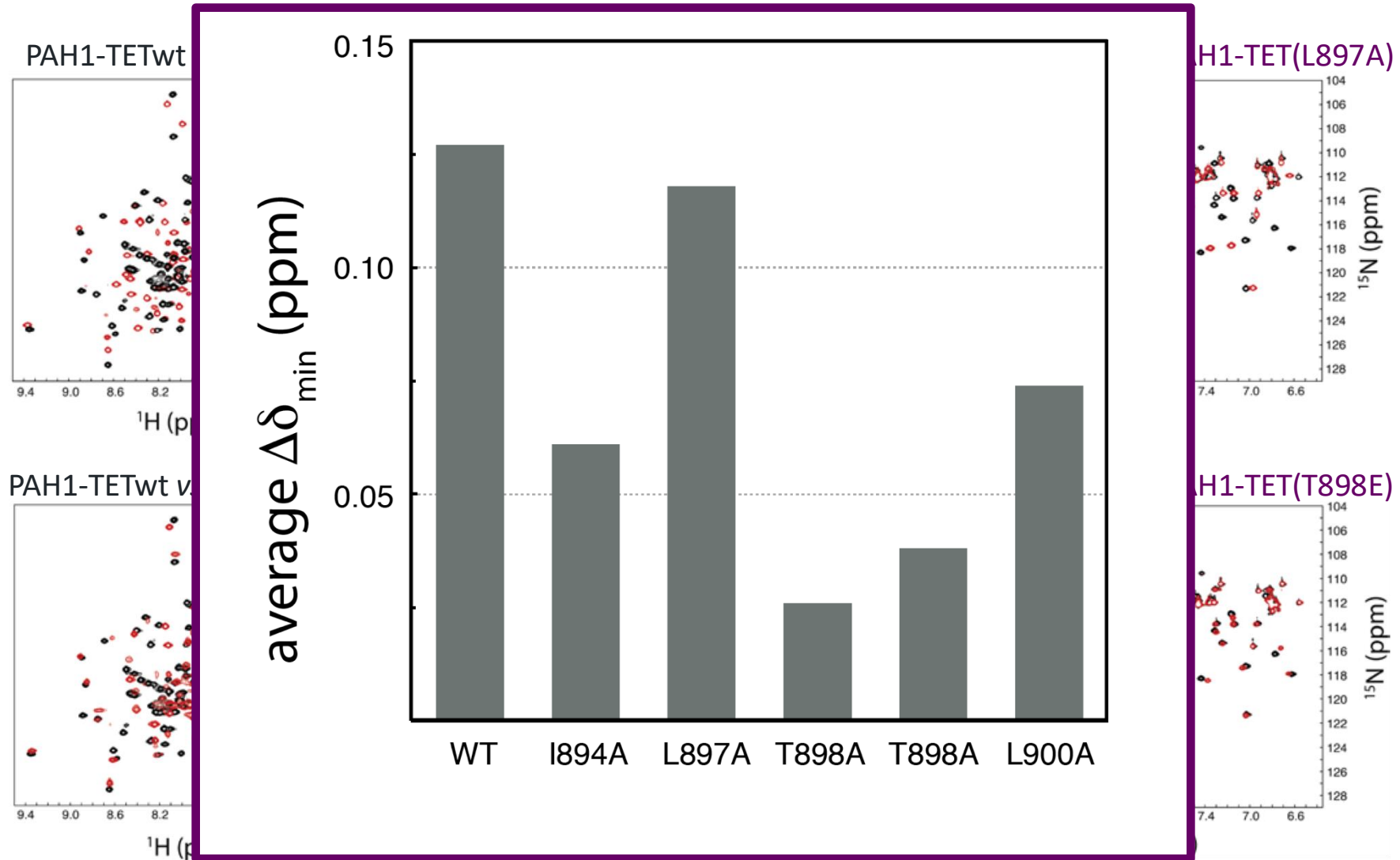
Macro: Colour a series in rainbow colours



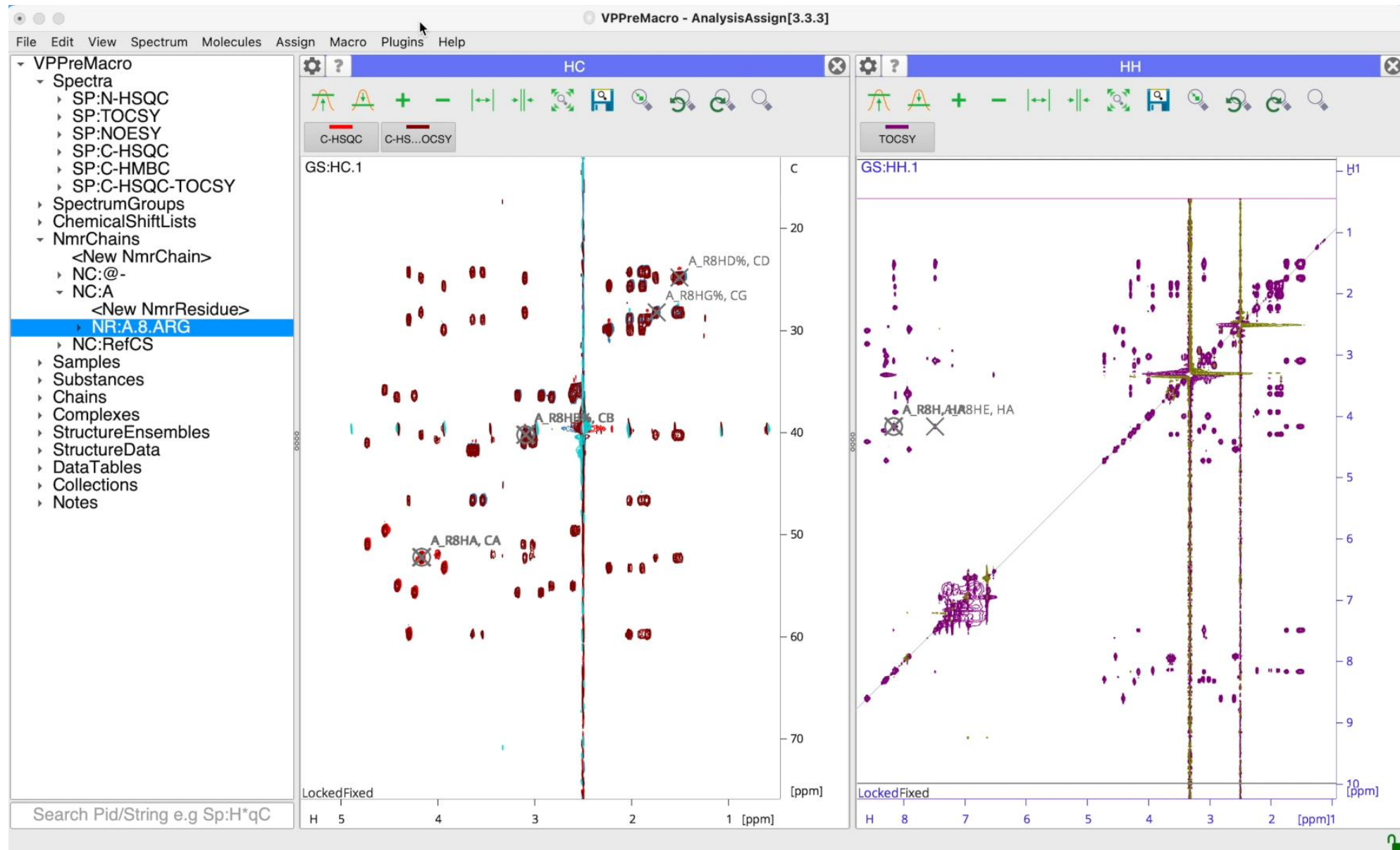
Macro: Delete diagonal peaks



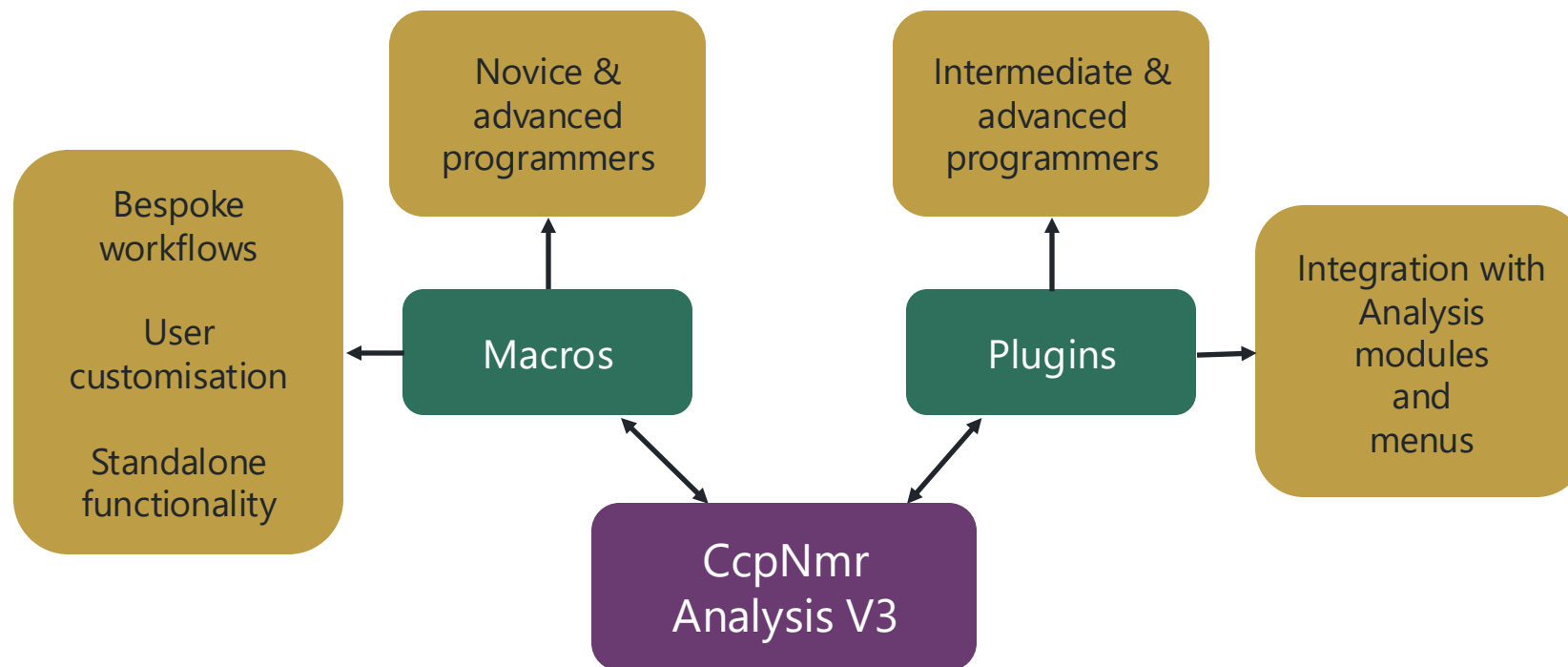
Macro: Data Analysis



Macro: Peptides



Analysis V3 as a development platform



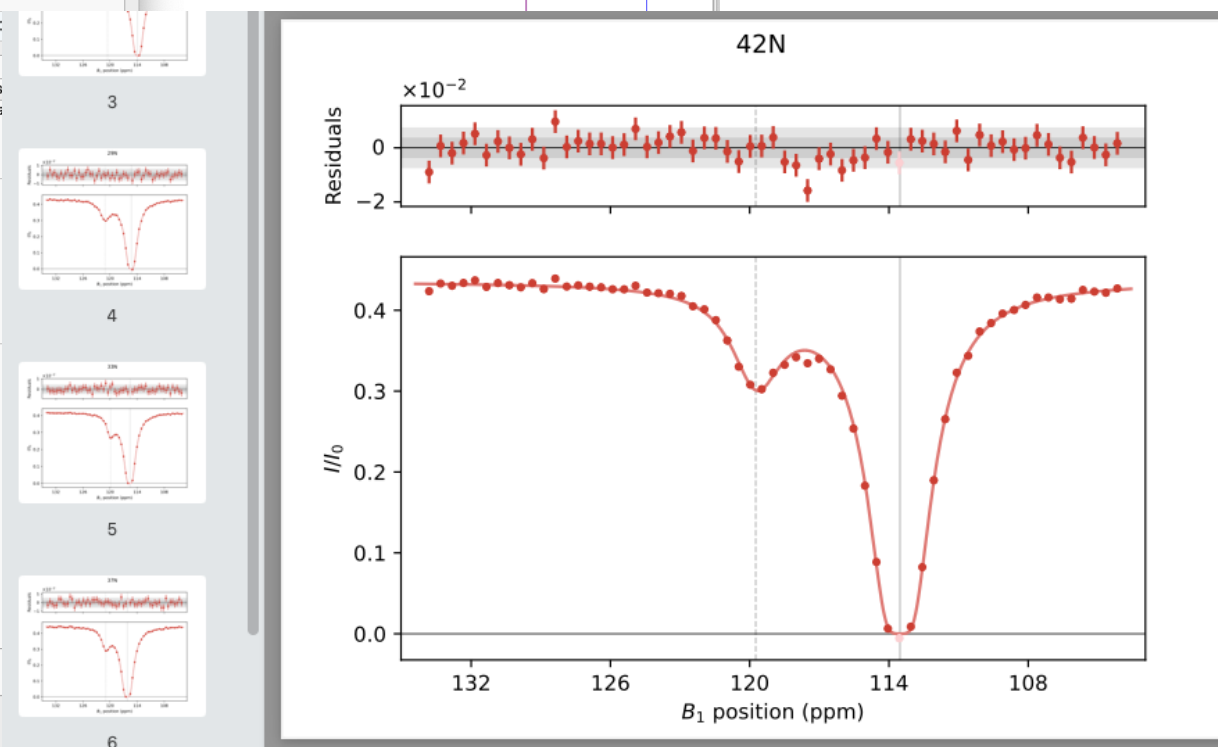
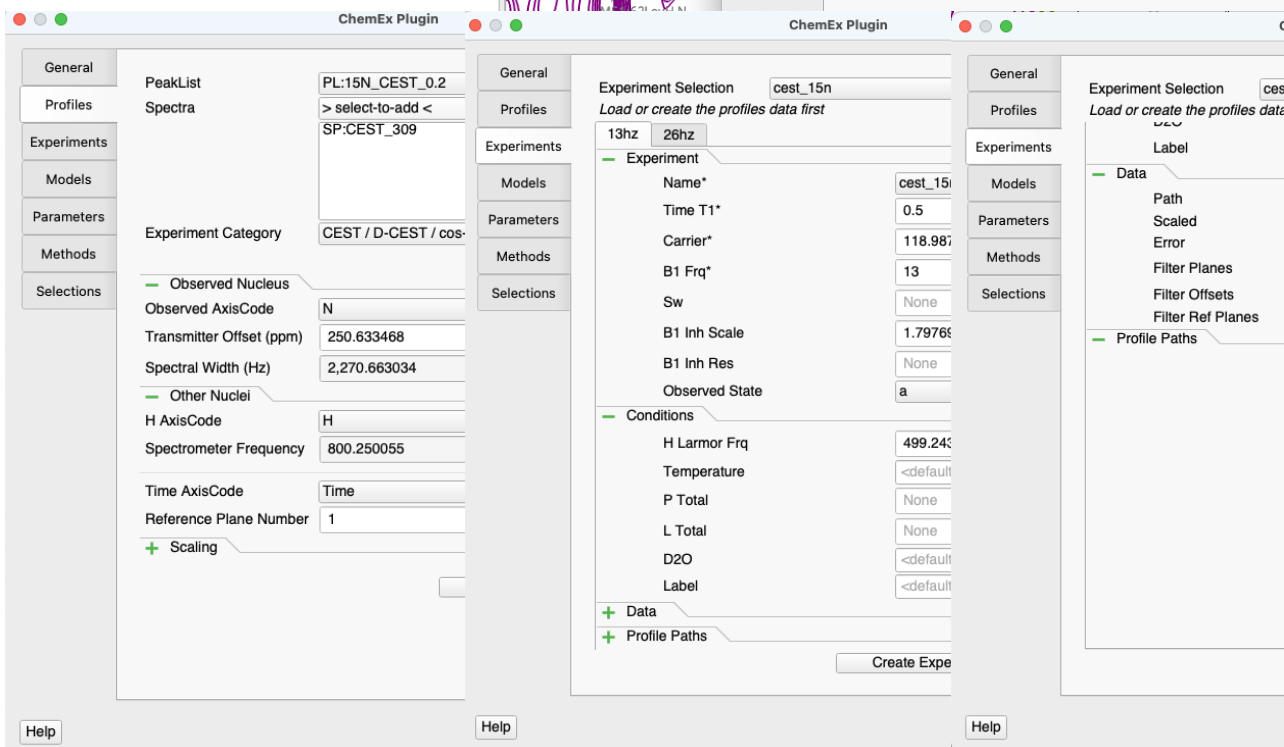
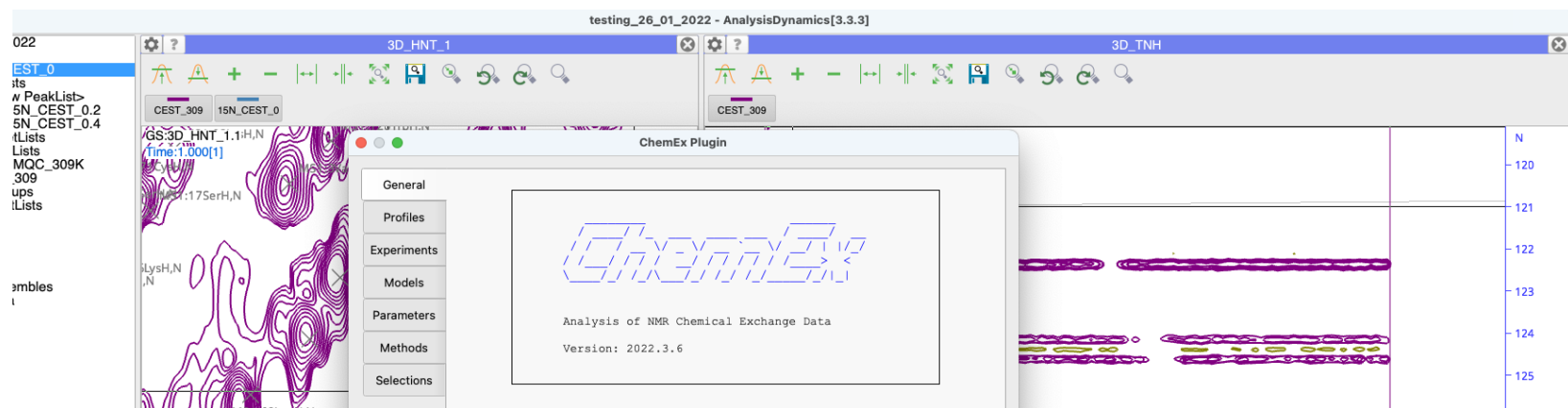
CHEMICAL PHYSICS

Protein NMR assignment by isotope pattern recognition

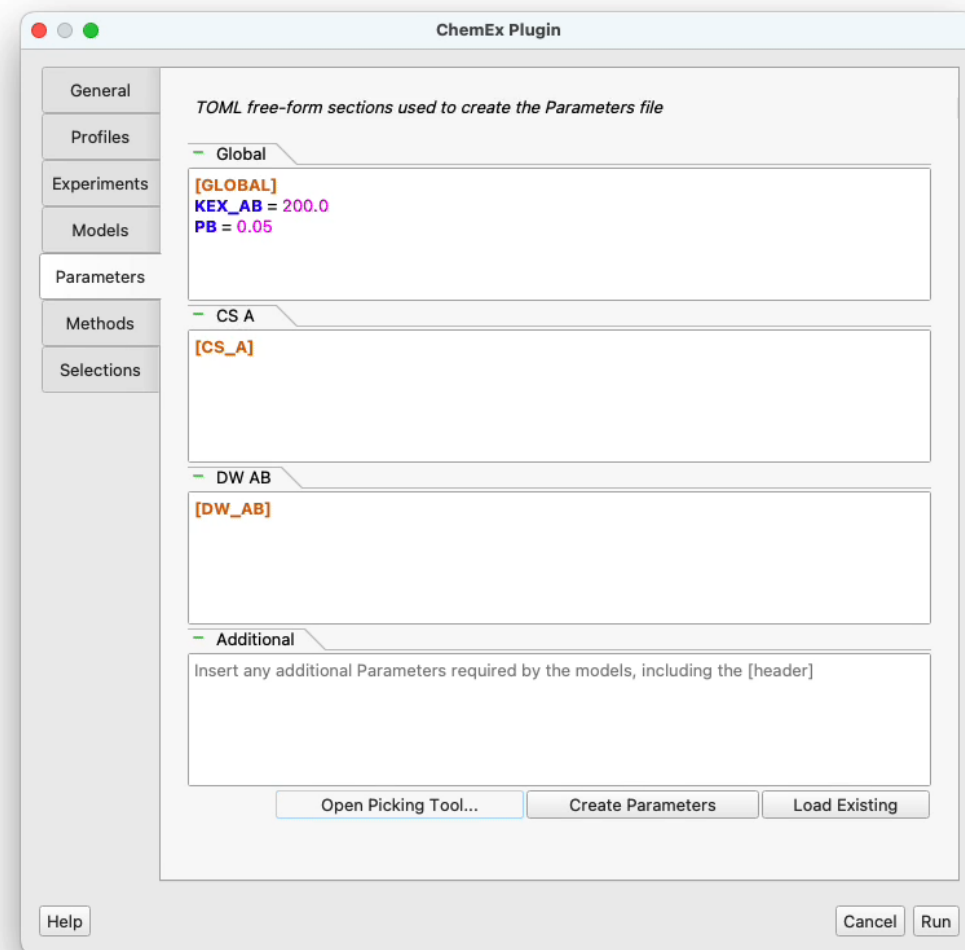
Uluk Rasulov¹, Harrison K. Wang^{2,3}, Thibault Viennet⁴, Maxim A. Droemer⁵, Srđan Matosin^{2,3}, Sebastian Schindler^{2,3}, Zhen-Yu J. Sun^{2,3}, Luca Mureddu⁶, Geerten W. Vuister⁶, Scott A. Robson^{2,3}, Haribabu Arthanari^{2,3*}, Ilya Kuprov^{1*}

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 [¹H, ¹³C] 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



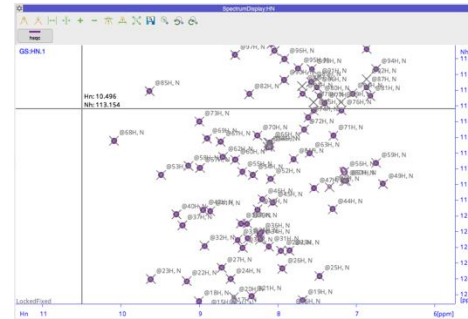
Plugins - ChemEx



Advantages to using the V3 environment

alabaster 0.7.12	gizmo 0.9.8.1	ilcbox 9.0.1	openblas 0.3.6	python3.8.1	sphinxcontrib-
asproptre 0.10	gn 2.25.0	ilcbox41 4.0.1	openblas-devel 2.8	python3-dataviz	jmathplot 0.1.0
astropy 4.0.1	glib 2.68.0	ilcbox41 4.0.1	openblas-opencl 0.1	python3-diff	jsmath 1.0.0
babel 2.8.0	graphviz 2.38.0	1.3.20181029	openpyxl 3.0.0	pytz 2019.3	jsmath-gif 1.0.0
backcall 0.10.0	gyp 2.50.0	haskell 2.0.0	openpyxl 1.1.e	python3.8	sphinxcontrib-
backports 0.2.0	haskell 2.0.0	haskell 2.0.0	packagelint 0.1.0	openpyxl 1.1.e	haskell 2.0.0
backports-shutil.get- terminal_size 1.0.0	haskell 2.0.0	ilcbox 1.5.0	pandas 1.0.3	et 5.9.7	haskell 2.0.0
blis 0.7.1	haskell 2.0.0	ilcbox 1.5.0	par 1.42.4	openoffice 4.7.2	haskell 2.0.0
blis 0.7.1	haskell 2.0.0	ilcbox 1.5.0	par 1.42.4	openoffice 4.7.2	haskell 2.0.0
ca-certificates 2020.11.11	haskell 2.0.0	ilcbox 1.5.0	par 1.42.4	openoffice 4.7.2	haskell 2.0.0
certifi 2019.11.28	haskell 2.0.0	ilcbox 1.5.0	par 1.42.4	openoffice 4.7.2	haskell 2.0.0
cffi 1.14.0	haskell 2.0.0	ilcbox 1.5.0	par 1.42.4	openoffice 4.7.2	haskell 2.0.0
chardet 3.0.4	haskell 2.0.0	ilcbox 1.5.0	par 1.42.4	openoffice 4.7.2	haskell 2.0.0
click 7.0.0	haskell 2.0.0	ilcbox 1.5.0	par 1.42.4	openoffice 4.7.2	haskell 2.0.0
click 7.0.0	haskell 2.0.0	ilcbox 1.5.0	par 1.42.4	openoffice 4.7.2	haskell 2.0.0
cycler 0.10.0	haskell 2.0.0	ilcbox 1.5.0	par 1.42.4	openoffice 4.7.2	haskell 2.0.0
cython 0.29.15	haskell 2.0.0	ilcbox 1.5.0	par 1.42.4	openoffice 4.7.2	haskell 2.0.0
dataclasses 0.8	haskell 2.0.0	ilcbox 1.5.0	par 1.42.4	openoffice 4.7.2	haskell 2.0.0
decorator 4.4.2	haskell 2.0.0	ilcbox 1.5.0	par 1.42.4	openoffice 4.7.2	haskell 2.0.0
defusedxml 0.6.0	haskell 2.0.0	ilcbox 1.5.0	par 1.42.4	openoffice 4.7.2	haskell 2.0.0
et_xmlfile 1.0.1	haskell 2.0.0	ilcbox 1.5.0	par 1.42.4	openoffice 4.7.2	haskell 2.0.0
expat 2.2.9	haskell 2.0.0	ilcbox 1.5.0	par 1.42.4	openoffice 4.7.2	haskell 2.0.0
fontconfig 2.13.1	haskell 2.0.0	ilcbox 1.5.0	par 1.42.4	openoffice 4.7.2	haskell 2.0.0
fonttools 4.23.1	haskell 2.0.0	ilcbox 1.5.0	par 1.42.4	openoffice 4.7.2	haskell 2.0.0
freeze 0.10.0	haskell 2.0.0	ilcbox 1.5.0	par 1.42.4	openoffice 4.7.2	haskell 2.0.0
h2o 3.11.0	haskell 2.0.0	ilcbox 1.5.0	par 1.42.4	openoffice 4.7.2	haskell 2.0.0
get_terminal_size 1.0.0	haskell 2.0.0	ilcbox 1.5.0	par 1.42.4	openoffice 4.7.2	haskell 2.0.0

Python packages

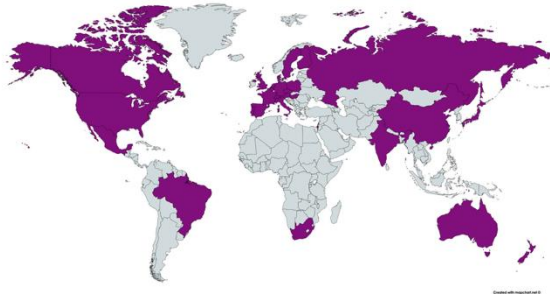


Analysis Routines



GitHub

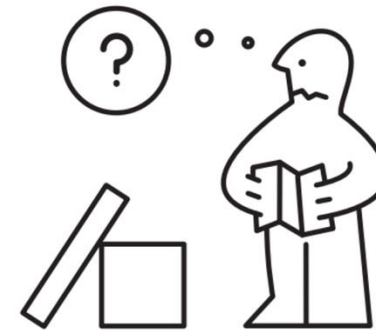
Developer Access



Global user base



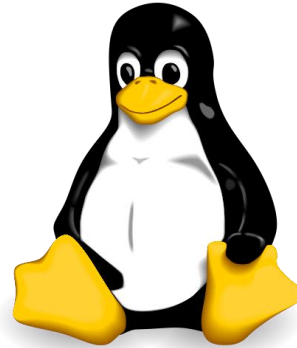
Publicity



Documentation & Workshops

Give it a try!

ccpn.ac.uk



CCPN

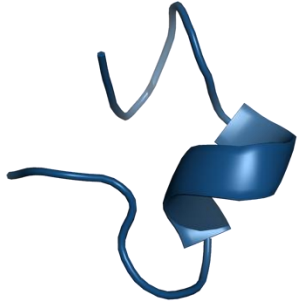
[Software](#) ▾ [Support](#) ▾ [Outreach](#) ▾ [About](#) ▾ [Search](#)

Collaborative Computational Project for NMR

CCPN is a public non-profit project, funded by the Medical Research Council

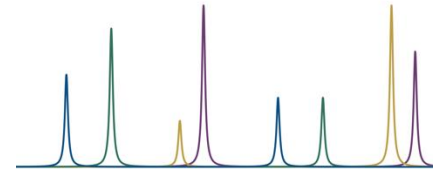
[DOWNLOAD OUR SOFTWARE](#)

Summary



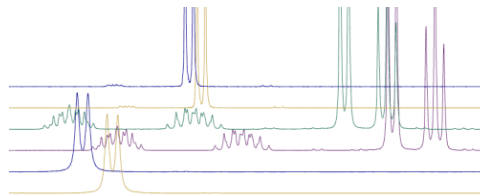
Peptides

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 Editor: 1
Macro Name: None
1 for peak in project.peaks:
2 # Get the assignments in all dimensions for a peak
3 for assignOptions in peak.assignedNmrAtoms:
4 for assignment in assignOptions:
5 # Create the new assignment string with 'molecule' as the NmrChain
6 temp = str(assignment)
7 assignmentComponents = temp.split(':')
8 assignmentComponents[0] = 'NA:molecule'
9 assignmentComponents[3] = assignmentComponents[3][0:-1]
10 newAssignment = ':'.join(assignmentComponents)
11 # get parameters needed for creation of new NmrChain, NmrResidue, NmrAtom and Peak Assignment
12 chainPid = 'NC:molecule'
13 resPid = 'NR:molecule' + ':'.join(assignmentComponents[1:-1])
14 seqCode = assignmentComponents[1]
15 resType = assignmentComponents[2]
16 atomName = assignmentComponents[3]
17 axCde = assignmentComponents[3][0]
18 # Create new NmrChain, NmrResidue and NmrAtom if necessary
19 project.fetchNmrChain('molecule')
20 get(chainPid).fetchNmrResidue(seqCode, resType)
21 get(resPid).fetchNmrAtom(atomName)
22 # Change the peak assignment to the new one
23 get(peak.pid).assignDimension axCde, newAssignment
24
```

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