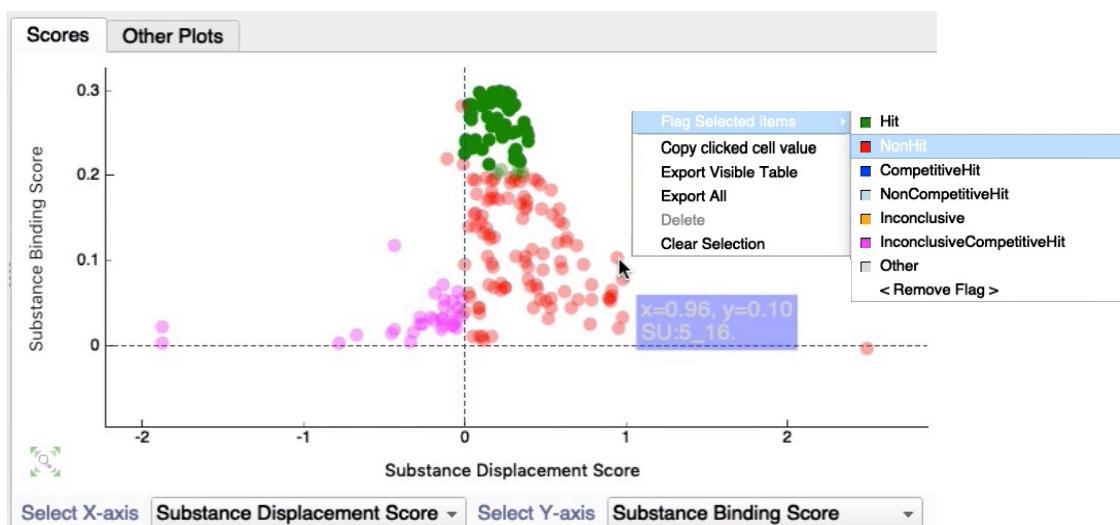


## AnalysisScreen Hit Analysis Tutorial



# Introduction

This tutorial will show the general usage of the Hit-Analysis module in CcpNmr AnalysisScreen Version 3.0.4 in multiple independent parts.

It is assumed that you have some basic familiarity with the program, e.g., from having completed our [Beginners Tutorial](#).

You will need to use the data located in the `/data/AnalysisScreen_Tutorial` directory of the CcpNmr V3 examples data which you can download from:

<https://www.ccpn.ac.uk/v3-software/tutorials/tutorial-data-and-examples>.

## Contents

### Part 1: Manual And Semi-Automatic Analysis

- Data importing and preparation *(Section 1)*
- Hit Analysis Module and Peak matching *(Section 2)*
- Binding scores and calculation engines *(Section 2)*

### Part 2: Automatic Analysis with Pipelines

- Pipelines *(Section 3)*
- Binding Plots *(Section 4)*
- Filtering and Exporting Datasets *(Section 4)*

### Part 3: Recurring Analyses

- Create and Export NEF files *(Section 5)*
- Screen From NEF files *(Section 6)*
- Pipelines *(Section 6)*

## Start CcpNmr Analysis V3

- Apple users by running Screen on the Launcher
- Unix users by using the terminal command: `bin/screen`
- Windows users by double-clicking on the `screen.bat` file

## Disclaimer

Datasets used for this tutorial are randomly generated and don't have any biological significance.  
All spectra shown are synthetic and for demonstration purposes only.

All compound names are randomly chosen and might have incorrect chemical properties or not be represented by the linked spectra.

Please note that the images shown are only representative and you may encounter minor differences in your setup.

# Introduction

## Getting started, basic operations

### Sidebar

All data contained in a project, such as spectra and peak lists are located in the sidebar. **Double-clicking** on an item will open its properties popup.

### Display

A display can contain multiple overlaid spectra which share the same axes. To show/hide a single spectrum, click on its toolbar button. If you close a display, you can open a spectrum by **dragging and dropping** it into the drop area from the sidebar or by **right-clicking** on a sidebar item and selecting **Open as module**. You can also add additional spectra to a spectrum display module or drag several spectra into the drop area together to open them simultaneously.

### Mouse

- Pan → **Left-drag** in display
- Zoom in/out → **Scroll wheel** in display
- Context menu → **Right-click**
- Select a peak → **Left-click** on a peak symbol “X”
- Move a peak → select first, then **middle-click and drag**

### Two-Letter Shortcuts

Press the first letter on your keyboard e.g., M, followed by the second letter, e.g., K (case insensitive). Press Esc to cancel the first letter.

Common in this tutorial:

- |                    |                                       |
|--------------------|---------------------------------------|
| SE                 | → Snap to Extremum the selected peaks |
| HA                 | → Open the Hit Analysis GUI Module    |
| PI                 | → Open the Pipeline GUI Module        |
| MC                 | → Clear all marks                     |
| <b>Space-Space</b> | → Open the Python console GUI Module  |

### For more commands and operations

**Main Menu → Help → Tutorials → Beginners Tutorial**

OR

**Main Menu → Help → Show Shortcuts**

# Introduction

## CcpNmr AnalysisScreen Nomenclatures

### Sample

A CcpNmr object containing information about the NMR physical sample, e.g., pH, ionic strength etc  
CcpNmr links: Sample component, Spectrum (e.g.: the spectrum Control, Target...)

### Sample component

A CcpNmr object containing information about the Substance in the NMR physical sample, (e.g., concentration)

CcpNmr links: Substance

### Substance

A CcpNmr object containing information about a biological molecule, (e.g., the small molecule and its general properties such as: SMILES, MW etc.)

CcpNmr links: Sample component, Spectrum (e.g.: the Singleton Spectrum)

### SpectrumGroup

A CcpNmr object containing a collection of spectra.

CcpNmr links: Spectra

### Control (spectrum)

The spectrum recorded at time X for a sample containing one or multiple substances prior the addition of a biological target

CcpNmr links: Sample

### Target (spectrum)

The spectrum recorded at time X for a sample containing one or multiple substances plus a biological target

CcpNmr links: Sample

### Displacer (spectrum)

The spectrum recorded at time X for a sample containing one multiple substances plus a biological target and a known binder. Also named as “competitor”

CcpNmr links: Sample

### Reference Mixture

The spectrum recorded for a sample containing multiple substances. Its peaks and their annotations are used as a template and to identify substances in future screening analyses

CcpNmr links: Substances

### Reference Singleton

The spectrum recorded for only one substance

CcpNmr links: Substance

### Binding Substance

The substance linked to a spectrum (reference) whose peaks have been matched to the spectral peaks (Control – Target) denoting a binding event

### Peak Match

The virtual linkage between a Reference – Control – Target ( – Displacer) peaks in a screen Dataset

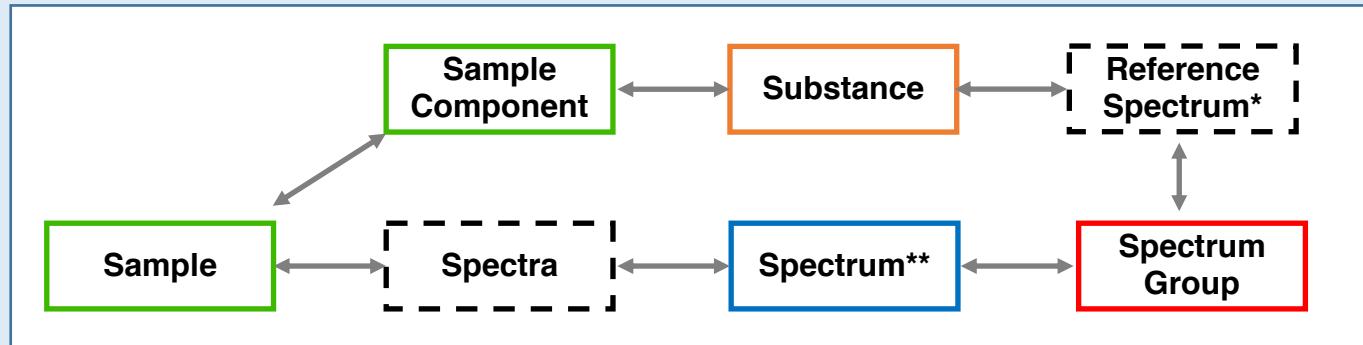


Figure 1. CcpNmr AnalysisScreen object links

The schematic diagram shows how objects are linked in CcpNmr AnalysisScreen.

\* Reference Singleton; \*\* Control, Target, Displacer, Reference Mixture

# Introduction

## Schematic representation of the Screen analysis workflow

In this tutorial you will analyse few  $^{19}\text{F}$  datasets following the workflow shown in Fig. 2. These steps and their tools in AnalysisScreen can also be applied to other experiment types analyses, such as:  $^1\text{H}$  Relaxation-edited, WaterLogsy and STD.

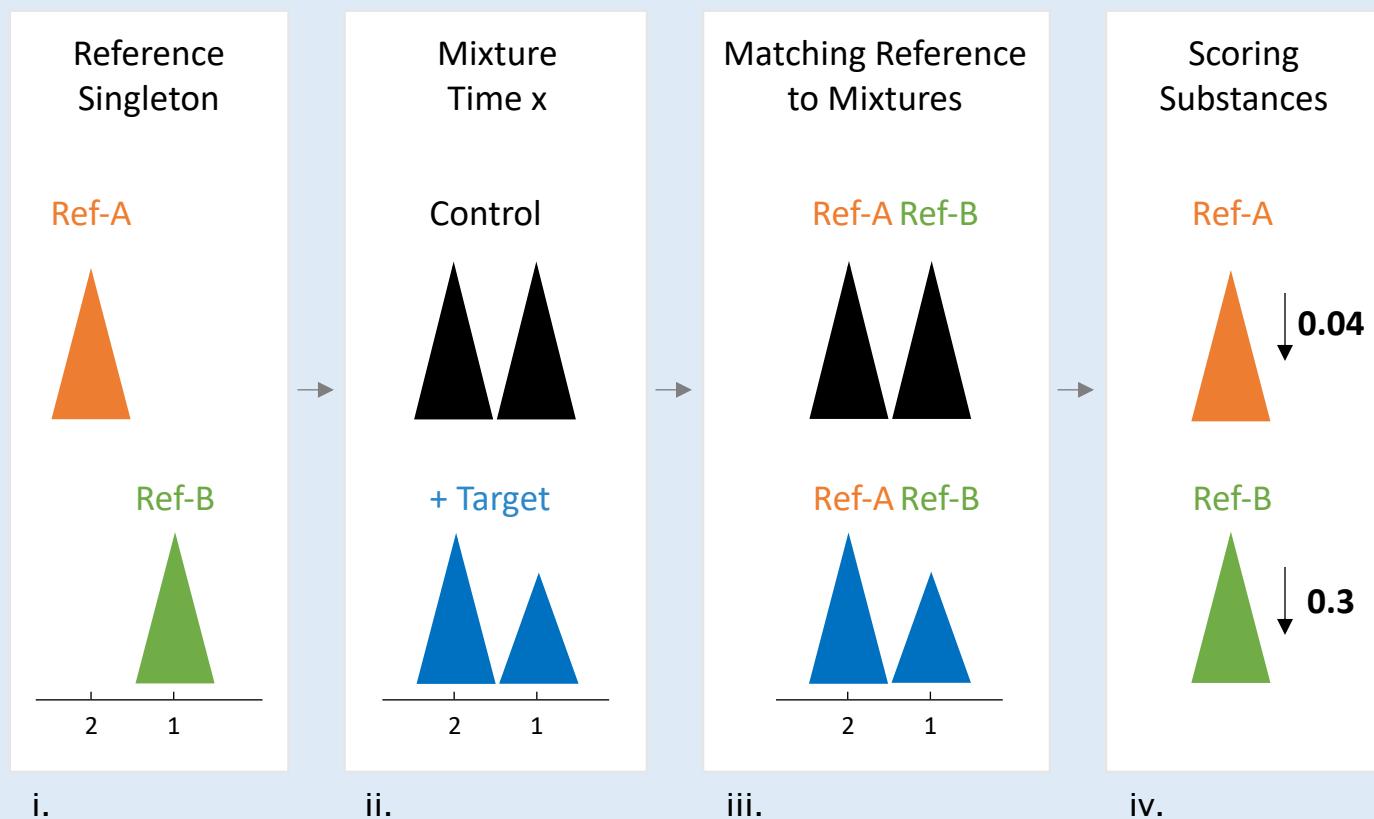


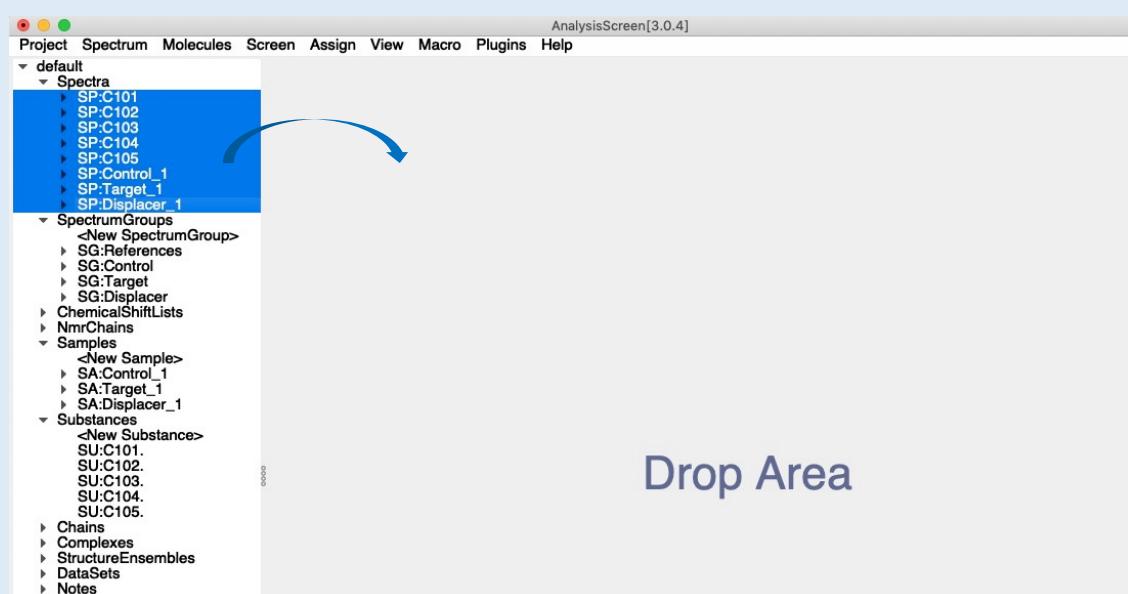
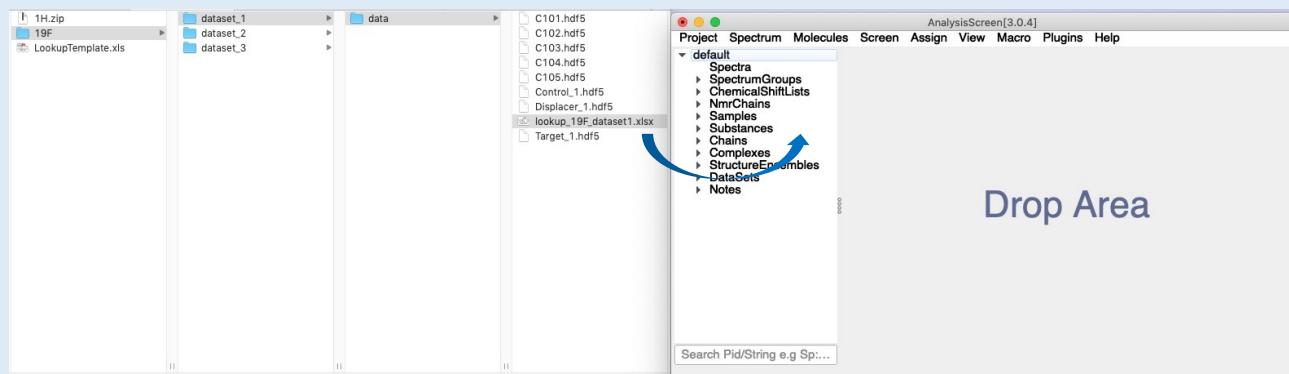
Figure 2. CcpNmr AnalysisScreen common workflow

- i. The *Ref-A* and *Ref-B* cartoons represent the two Reference Singleton spectra recorded for the **Substances** Ref-A and Ref-B
- ii. The *Control* and *+Target* cartoons represent spectra recorded at a time X for **Samples** containing only substances (*Control*) and plus a biological target (*+Target*)
- iii. The Reference Singleton signals are matched to the Control and Target spectral signals based on chemical shift position. This creates a **Peak Match**
- iv. Spectral differences between the Control and Target are scored using different calculation methods. Scores are then transferred to the matching Substances as indication of their binding activity.

In the Hit Analysis module, each Peak Match will have a **Peak Binding Score**, e.g., 0.3 for the Ref-B (Fig.3 iv), corresponding to a 30% intensity reduction for a single peak. The combination of all Peak Binding Scores for a substance will define the **Substance Binding Score**, in this example, only one peak is recorded for Ref-B, so the Substance Binding Score is 0.3. Whereas the combination of all Substances Binding Scores within a sample will define the **Sample Binding Score**.

If we set a Substance Binding Score threshold at 0.2 (arbitrary), the Ref-B Substance could be considered as a **Binding Hit** for this dataset.

# Part 1: Manual And Semi-Automatic Analysis



By loading an Excel file, correctly formatted, all the necessary links are automatically established ensuring an optimal functioning of the screening tools.  
See the “HowTos\_SidebarObjects” and the “HowTos\_ImportDataFromExcel” manuals for more information.

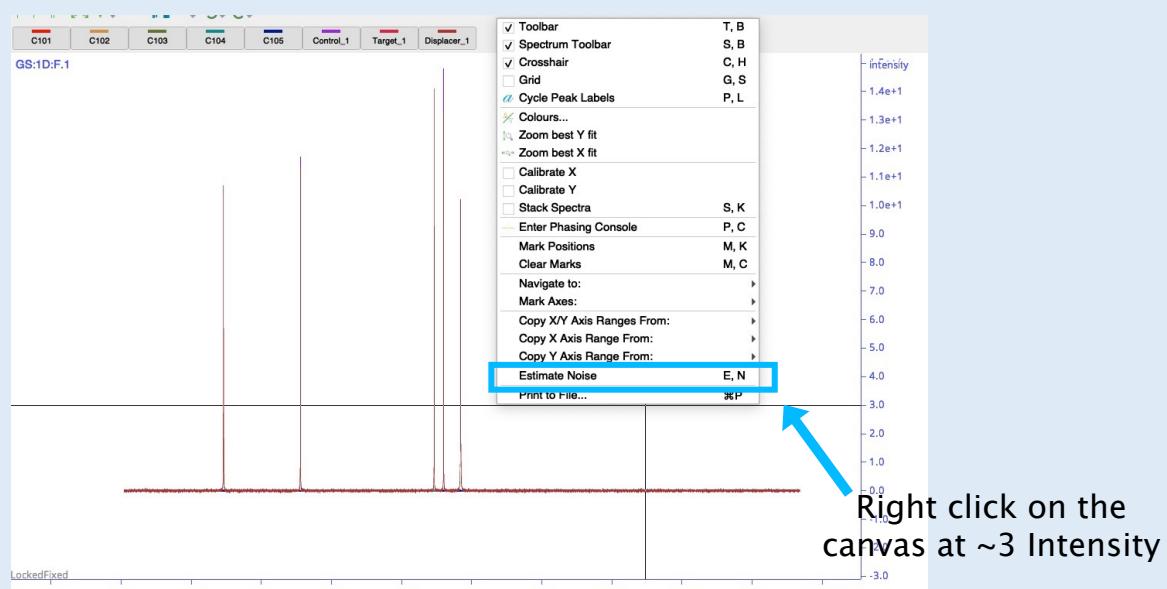
## 1A Drag & drop the file **lookup\_19F.xlsx** into the program

- Locate the demo dataset in the AnalysisScreen tutorial folder in `AnalysisV3/data/testProjects/AnalysisScreen_Tutorial/19F/dataset_1`
- Find the Excel file **lookup\_19F\_dataset1.xlsx** in the data directory and drop it into the software

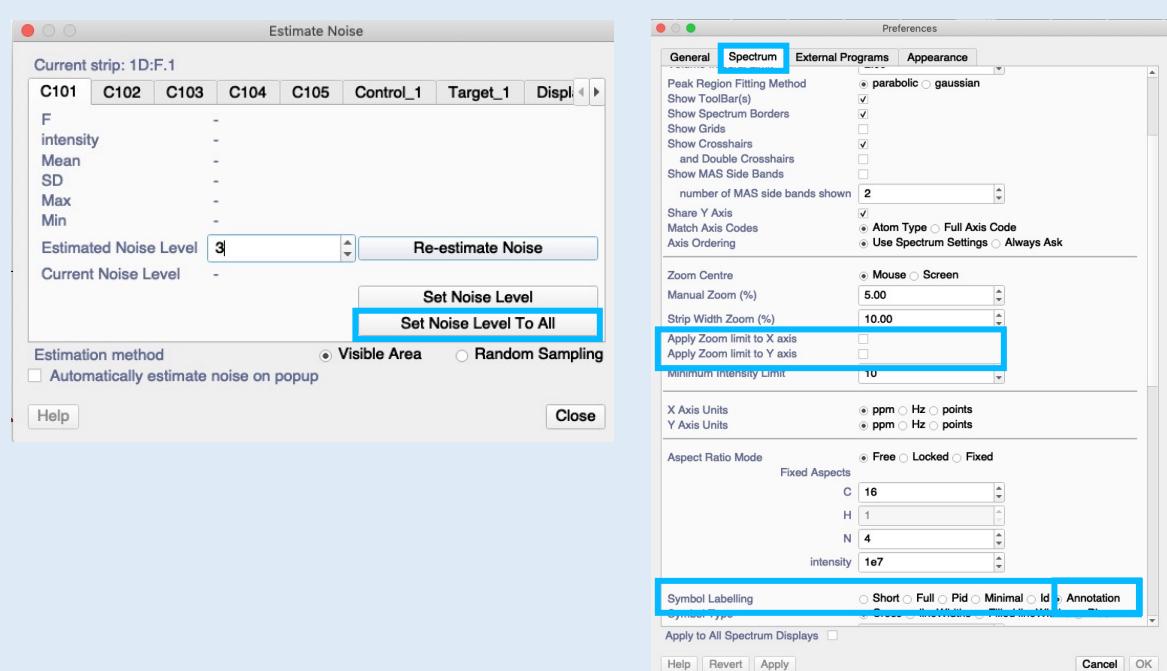
## 1B Open all spectra

- On the sidebar, expand the branch Spectra
- select all spectra, drag and drop them onto the Drop Area.

# Prepare Data



Right click on the canvas at ~3 Intensity



## 1c Set noise level

On display:

- place the mouse cursor at a position where you want to set the noise level, for example, at ~ 3 in the Intensity axis.
- right click -> **Estimate Noise** -> **Set Noise Level To All** (displayed spectra)

Go to Main Menu -> Project -> Preferences -> Spectrum Tab

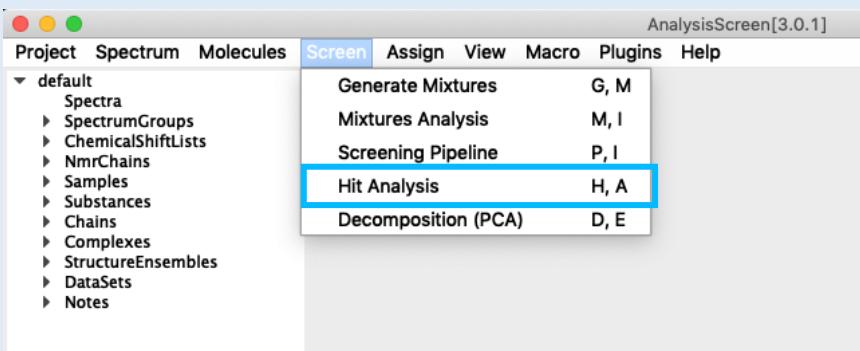
- set unchecked **Apply zoom limit** both for X and Y axes
- Set Annotation for Symbol Labelling**

## 1d Stack Spectra

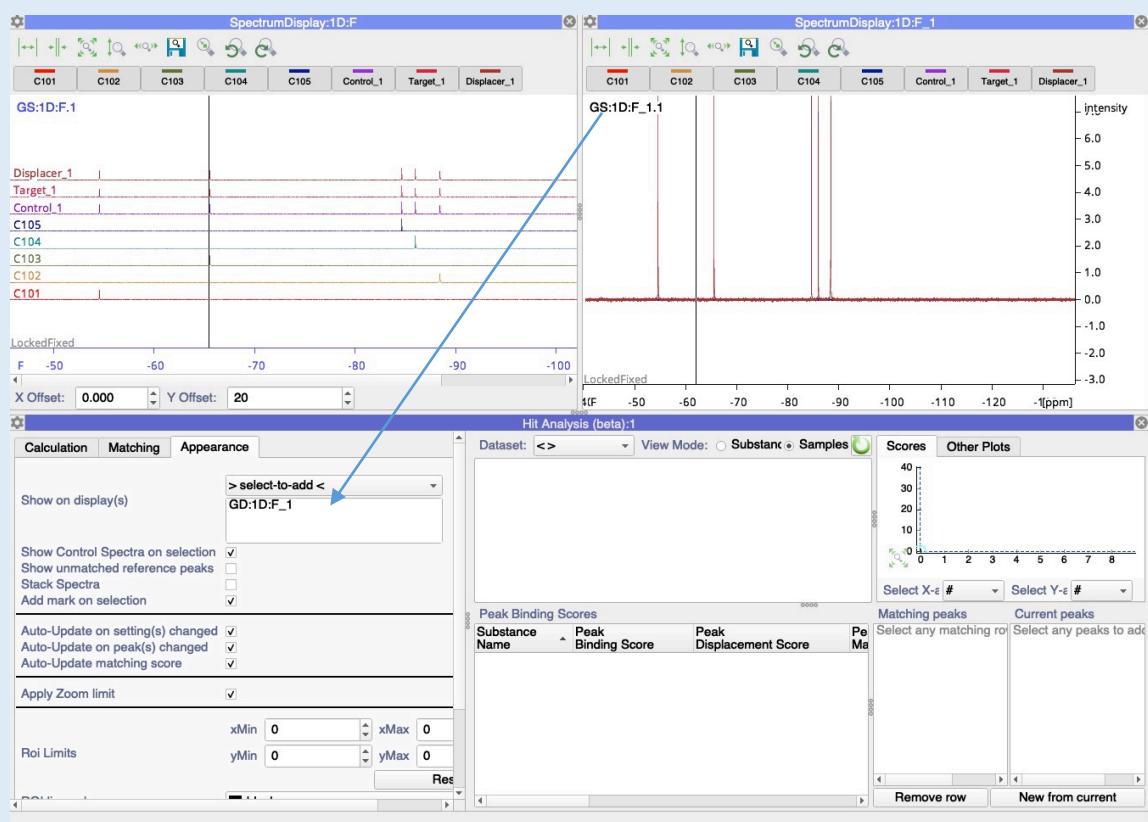
On display:

- right click -> **Stack Spectra** (shortcut **SK**)
- Y Offset:** 20
- Zoom in-out as required to show all spectra in the display
- Scroll mouse-wheel over the Intensity axis to adjust the Y-range

# Hit Analysis Module setup



Shortcut  
"HA"



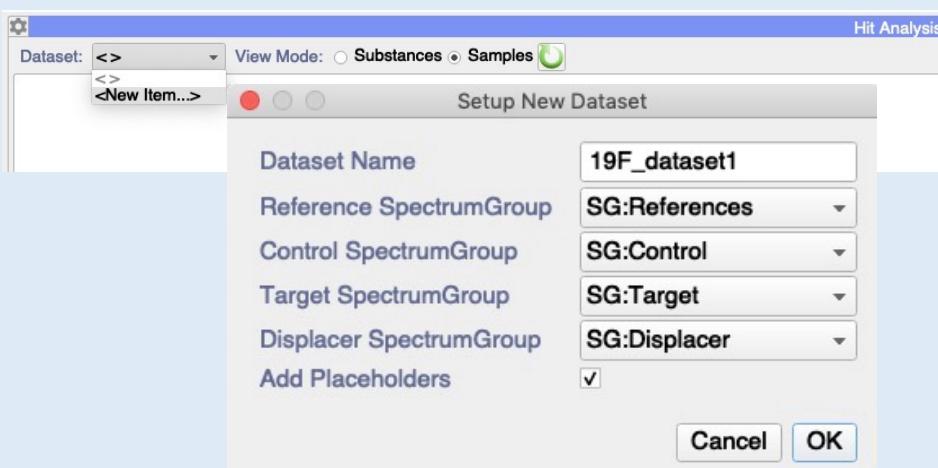
## 2A Open the Hit Analysis Module

- Go to Main Menu → Screen → Hit Analysis (shortcut HA)

## 2B Open a second Spectrum Display and re-arrange layout

- Open any spectrum next to the already opened stacked display
- Click on the Hit Analysis module **Settings** gear icon :
- **Appearance tab**
- **Show on display(S) -> right click -> remove all**
- Add the new unstacked **Spectrum Display** (e.g. GD:1D:F\_1) to the list  
The module will perform soon a series of dynamic actions on this display.
- Close the settings
- re-arrange the Hit Analysis Module below the Spectrum Displays so to show all its widgets on the screen.

# Hit Analysis Module setup



Hit Analysis (beta):1								
#	Substance Name	Substance Binding Score	Substance Displacement Score	Substance Matching Score	Substance Label	Relative S/N Ratio	Sample Name	Target Name
1	C101	0.0000	0.0000				Target_1	
2	C102	0.0000	0.0000				Target_1	
3	C103	0.0000	0.0000				Target_1	
4	C104	0.0000	0.0000				Target_1	
5	C105	0.0000	0.0000				Target_1	

The **Hit Analysis** module contains two views: by Substance and by Sample. This selection determines the behaviour of the two main tables.

The top table contains a list of the substances or samples depending on the selected view. The lower table contains a list of all the peaks for the single substance reference spectrum or all the substances present for a selected sample.

Add empty placeholders to organise automatically Substances and Samples on Tables based on your project data. This option will allow the Hit Analysis module to display all spectra in the correct order, facilitating the manual/semi-automatic analysis.

## 2C Create new Screening Dataset

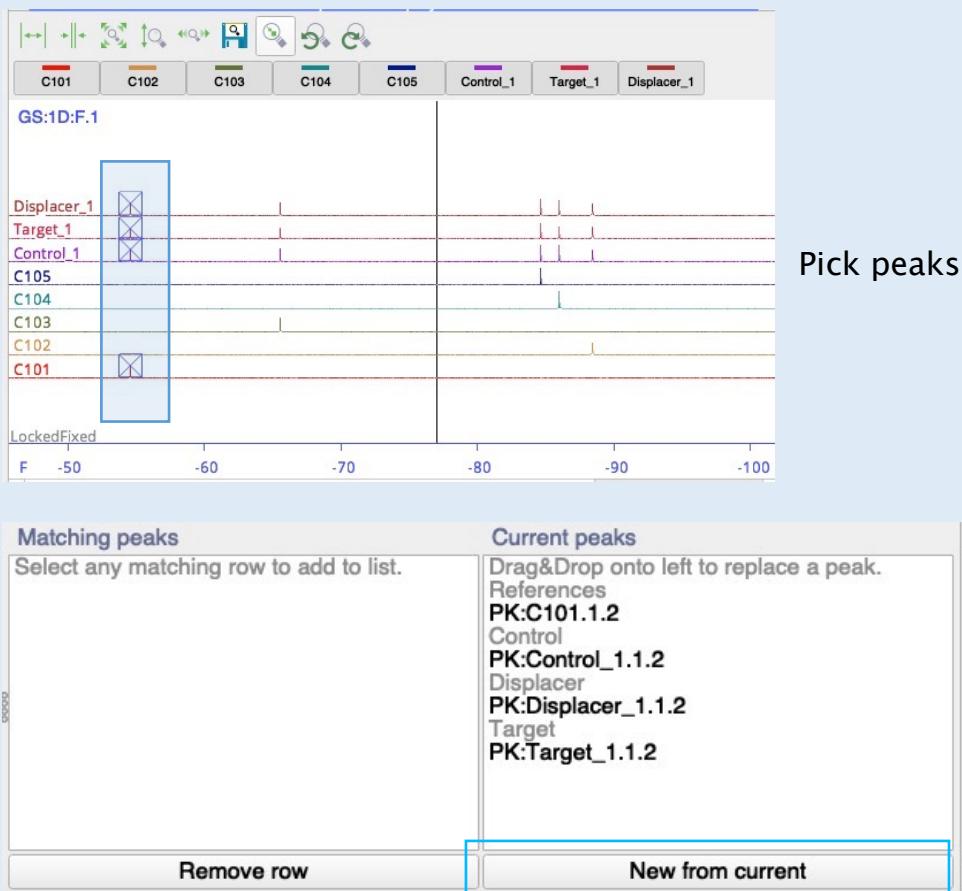
On the Hit Analysis module:

- Select **<New Item...>** on the **Dataset** pulldown
- Change name to **19F\_dataset1** or keep default
- Set checked the Add Placeholders
- Press Ok to proceed and close the popup

All substances and samples in the project are now organised in the tables and ready to create *Peak Matches* and scores

- On the View Mode option, select Substances
- On the Substance Table, select the first entry, e.g., C101  
the Spectrum Display F\_1.1 will update to show the Control-Target-Displacer and C101 spectra

# Hit Analysis Module



## 2D Pick and match

- Pick peaks at ~ ppm -54.59
- use CTRL (or CMD for Mac) + SHIFT + Left-drag to create a (blue) picking region so to include all spectral signals around -54 ppm.
- This signal correspond to the Substance C101

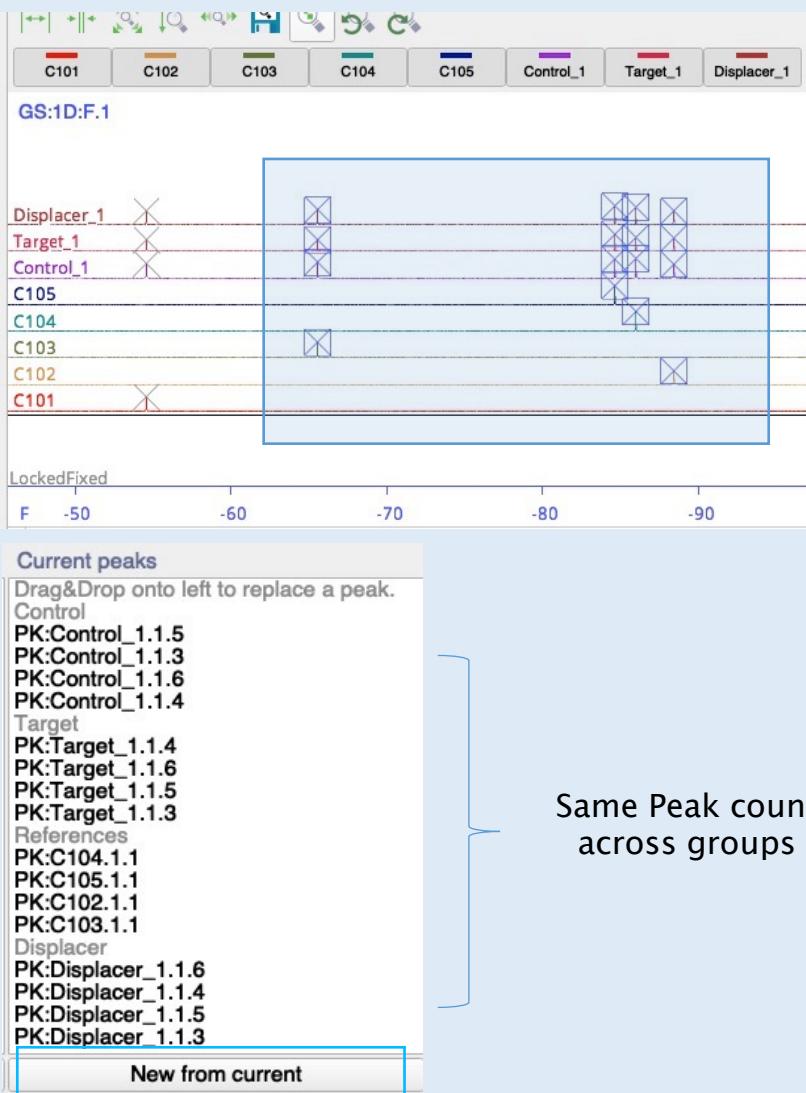
On the Hit Module:

- the **Current peaks** will appear at the bottom-right corner
- Press **New from current** to create a new *Peak Match*

A new Peak Binding Score will appear for the Substance C101 in the Peak Matches Binding Scores table. Select the row to navigate to the relative peaks in the spectrum display

The figure shows a screenshot of the Hit Analysis (beta):1 interface. At the top, there is a header bar with 'Dataset: 19F\_dataset1' and 'View Mode: Substances'. Below the header is a 'Substance Table' with columns: #, Substance Name, Substance Binding Score, Substance Displacement Score, Substance Matching Score, Substance Label, Relative S/N Ratio, Sample Name, Target Name, and Comment. The table has 5 rows, numbered 1 to 5. Row 1 (C101) is highlighted in yellow. A blue arrow points from the text 'Substance Table' to this row. To the right of the table is a 'Peak Binding Scores' table with columns: Substance Name, Peak Binding Score, Peak Displacement Score, Peak Position Reference, Peak Position Control, Peak Position Target, Peak Position Displacer, Reference-Target Δ Shift, Control-Target Δ Shift, and Comment. The table has 2 rows, both highlighted in yellow. The first row corresponds to the highlighted row in the Substance Table. A blue arrow points from the text 'Peak Matches Table' to the second row of the Peak Binding Scores table. To the right of the Peak Binding Scores table is a 'Matching peaks' panel listing substances: PK:C101.1.2, Control, PK:Control\_1.1.2, Target, PK:Target\_1.1.2, Displacer, and PK:Displacer\_1.1.2. A 'Remove row' button is located at the bottom right of the interface.

### Multi-picking and semi-automatic peak matching



### 2E Semi-automatic picking and matching

- Pick the remaining spectral signals by creating a larger picking region with the shortcut CTRL (or CMD for Mac) + SHIFT + Left-drag
- Select all peaks, shortcut CTRL (or CMD for Mac) + A

On the Hit Module:

- the **Current peaks** will appear at the bottom-right corner, make sure there is an even number for each category, e.g.: 4 Reference peaks – 4 Control Peaks etc... in any order.
- Press **New from current**, tables will update accordingly

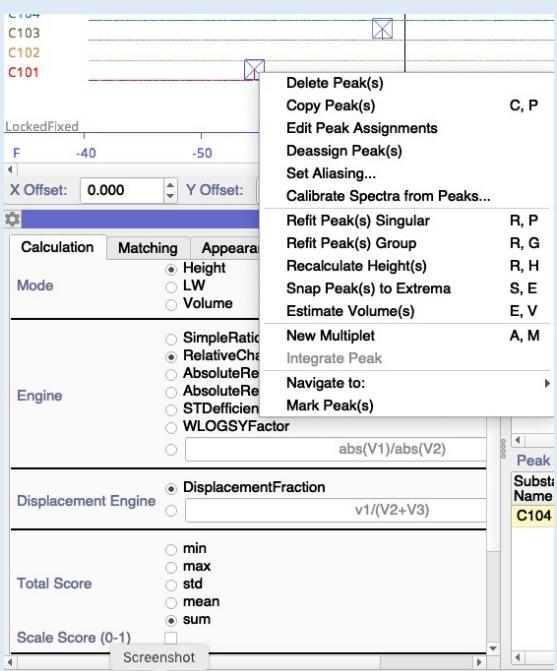
Dataset: 19F\_dataset1 View Mode:  Substances  Samples

#	Substance Name	Substance Binding Score	Substance Displacement Score
4	C104	0.2952	0.3392
3	C103	0.2595	0.1343
1	C101	0.1950	0.2053
5	C105	0.0951	0.0002
2	C102	0.0301	-0.0000

Peak Binding Scores

Substance Name	Peak Binding Score	Peak Displacement Score
C104	0.2952	0.3392



Select all peaks

Right click on a selected peak

(from the first spectrum if in stack mode)

- Refit Singular → shortcut RP
- Estimate volumes → shortcut EV

## 2F Scoring Engines

On the Hit Module:

- Select Substances in the View Mode and sort by Substance Binding Score

The **Substance Binding Score** gives an indication of the spectral changes between the Control-Target spectrum for the peaks matched to the Substance reference spectrum, therefore, it can be used to assess the substance binding quality (see Introduction Fig. 2).

Peaks are compared by one of following properties: height, linewidth or volume. As default, linewidths and volumes are not calculated. To calculate:

- Select all peaks from display 1D:F.1, press shortcut RP to refit all linewidths, and EV to estimate volumes.

Open Hit Module settings from the gear icon, select the Calculation tab.

The binding score is given by the calculation engine.

- Change the default by selection one of the options. By hovering over the labels for each option, a tip-text window will show the equation used
  - You can also define your equation in the free entry box:  
use V1 and V2 to define the variables for the calculation matrix
    - V1 represent each Ligand signal in the presence of the target (Target)
    - V2 represent each Ligand signal in the absence of the target (Control)
- The following arithmetic operations are supported:

``` + ```, ``` - ```, ``` * ```, ``` / ```, ``` ** ```, ``` % ```, ``` // ```

Same rule applies for the **Displacement Engine** which gives the Substance Displacement score. When building your own equation, define the Displacer signal with the variable V3

# Hit Analysis Module

**Calculation**

**Matching**

**Appearance**

**Mode**

- Height
- LW
- Volume

**Engine**

- SimpleRatio
- RelativeChange
- AbsoluteRelativeChange
- AbsoluteRelativeDifference
- STDeficiency
- WLOGSYFactor
- 

**Displacement Engine**

- DisplacementFraction
- 

**Total Score**

- min
- max
- std
- mean
- sum

**Scale Score (0-1)**

**Dataset:** Matching **View Mode:** # Substance Name Substance Binding

**Hit Analysis (beta):1**

**Dataset:** 19F\_dataset1 **View Mode:**  Substances  Samples

## Scoring Engines (continued)

Furthermore, this score is the result of all the single peak-matches contributions.

In this 19F demo dataset, only one observation is recorded per reference spectrum.

However, if multiple peaks per substance are present, the total Substance Binding score will be given by one of :

*Min, Max, std, mean, or sum*

of all the single peak binding scores.

- Select this option from the Settings **Total Score**

Whereas, if **Samples** is selected in the View Mode, the Sample Binding Score is given by all the Substances binding scores.

**By changing settings or modifying/fitting peaks, the whole module will update and recalculate automatically all scores.** For larger datasets this can be time-expensive, disable this feature from:

- Settings -> Appearance tab
- set uncheck Auto-Updates on settings(s) changed
- set uncheck Auto-Updates on peaks(s) changed

The refresh button will turn red the first time and every changes detected while working on the dataset, click it to update all.

# Hit Analysis Module

## Table Selections

This module has multiple dynamic selections

Selecting a row on the **Substances table**, it will:

1. List all the contributing matches in the Peak Matches Binding Scores table
2. Display all the spectra associated to the binding match
3. Select a relative item in the Scores Scatter plot

Selecting a row on the **Samples table**, it will:

1. List all the contributing matches in the Peak Matches Binding Scores table for all the substances present in the sample
2. Display all the spectra associated to the binding match, including all references spectra.

Selecting a row on the **Peak Binding Scores table**, it will:

1. Select all peaks included in the match
2. Navigate to Peak Position
3. Populate the Matching/Current Peaks lists widgets

Double-clicks on tables will re-execute the single selection.

The screenshot shows the Hit Analysis (beta):1 interface. At the top, there is a navigation bar with 'Dataset: 19F\_dataset1' and 'View Mode: Substances (Samples)'. Below this is the 'Substances Table', which contains four rows of data: C101, C102, C103, and C104. A blue arrow points from the text 'Substances Table' to the second row of the table. Below the table is the 'Peak Binding Scores' section, which includes a table with two rows: C101 and C102. A blue arrow points from the text 'Peak Matches Table' to the second row of this table.

#	Substance Name	Substance Binding Score	Substance Displacement Score	Substance Matching Score	Substance Label	Relative S/N Ratio	Sample Name
1	C101	0.0010	0.5321				Target_1
2	C102						Target_1
3	C103						Target_1
4	C104						Target_1

Peak Binding Scores				Peak Position Target	Peak Position Displacer
Substance Name	Peak Binding Score	Peak Displa	Sample Displa	Peak Position Target	Peak Position Displacer
C101	0.0010	0.532		-54.5931	-54.5931
C102	-0.0082	0.303		-88.4319	-88.4319

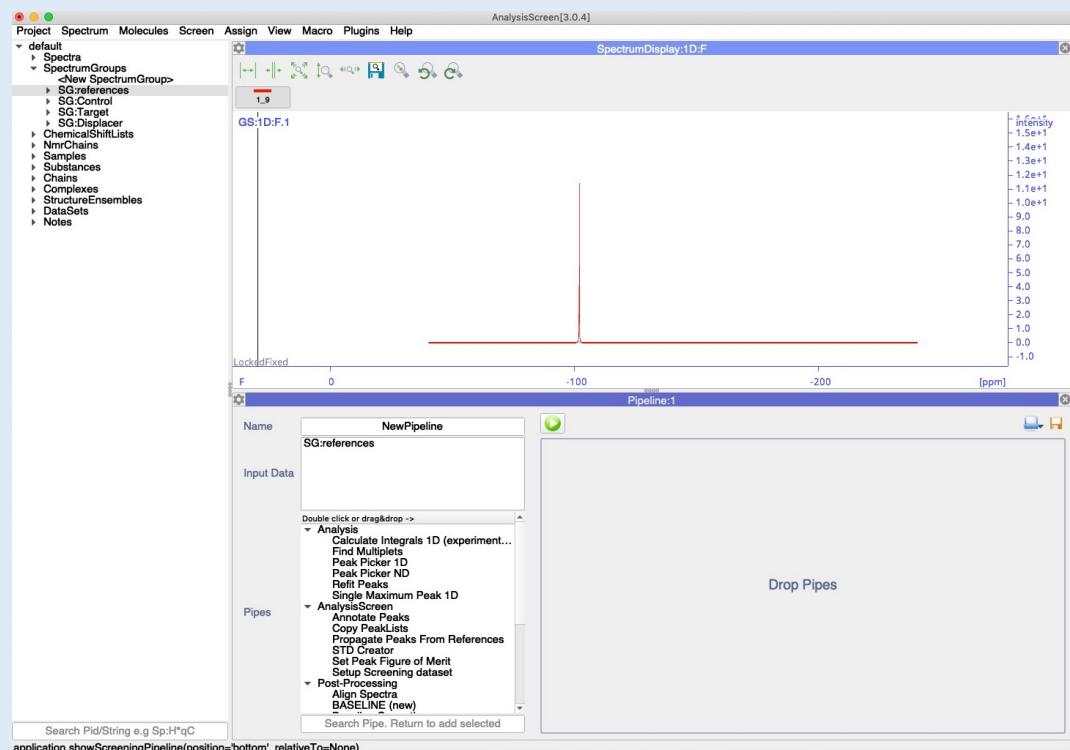
The screenshot shows the Hit Analysis (beta):1 interface. At the top, there is a navigation bar with 'Dataset: 19F\_dataset1' and 'View Mode: Substances (Samples)'. Below this is the 'Samples Table', which contains one row of data: Target\_1. A blue arrow points from the text 'Samples Table' to this row. Below the table is the 'Peak Binding Scores' section, which includes a table with five rows: C101, C103, C105, C104, and C102. A blue arrow points from the text 'Peak Matches Table' to the last row of this table.

#	Sample Name	Target Name	Sample Binding Score	Sample Displacement Score	Sample Matching Score	Comment
1	Target_1	Target_1				

Peak Binding Scores							
Substance Name	Peak Binding Score	Peak Displacement Score	Peak Position Reference	Peak Position Control	Peak Position Target	Peak Position Displacer	
C101		0.2053	-54.5932	-54.5931	-54.5931	-54.5931	
C103		0.1343			728	-65.5728	
C105		0.0001			519	-84.6519	
C104		0.3391			021	-86.0021	
C102		0.0003	-88.4319	-88.4319	-88.4319	-88.4319	

# Part 2: Automatic Analysis with Pipelines



## Automatic peak matching and binding analysis

In this section, a pipeline is built to inspect spectral changes between simulated samples recorded for mixtures of substances with and without a target, and with target and displacer.

A completed project is provided in the dataset2 folder.

### 3A Open a new project, load lookup\_19F\_dataset2.xlsx into the program

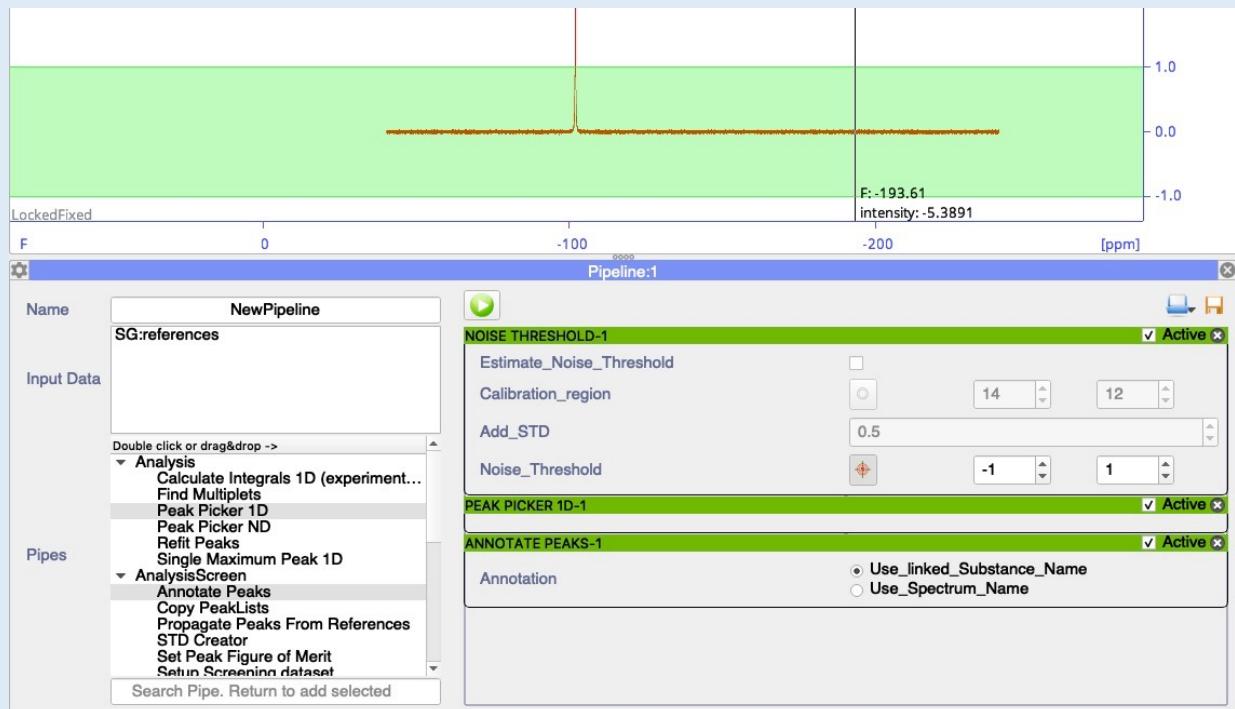
- Open a new project
- Locate the data of dataset\_2 inside the folder:  
AnalysisV3/data/testProjects/AnalysisScreen\_Tutorial/19F/dataset\_2/data
- Drag & drop the file lookup\_19F\_dataset2.xlsx into the sidebar or drop area.  
See the “HowTos\_ImportDataFromExcel” for creating excel files

### 3B Open the first reference spectrum

- select the first spectrum on Sidebar, drag and drop it onto the Drop Area.

### 3c Open the pipeline module, shortcut PI

- Open the pipeline module from the main menu:  
Menu > Screen > Pipeline  
or use the shortcut PI
- Expand the Sidebar branch for SpectrumGroups
- Select SG:References, drag & drop it into the Input Data of the Pipeline



Multiple Pipes can be added and re-order by holding and dragging the green top bar. See “HowTos\_Pipelines” for more information

### 3D Pick peaks on SG:References

- On the Pipes list widget search and add to the pipeline area by double clicking the pipe name or via drag and drop:

#### 1. Generic > Noise Threshold

- set uncheck **Estimate Noise threshold**
- Click on the Target button
- Insert the **Noise Threshold** values -1, 1 either by dragging the green lines that appears on displays or inserting the values in the entries

#### 2. Analysis > Peak Picker 1D

#### 3. AnalysisScreen > Annotate Peaks

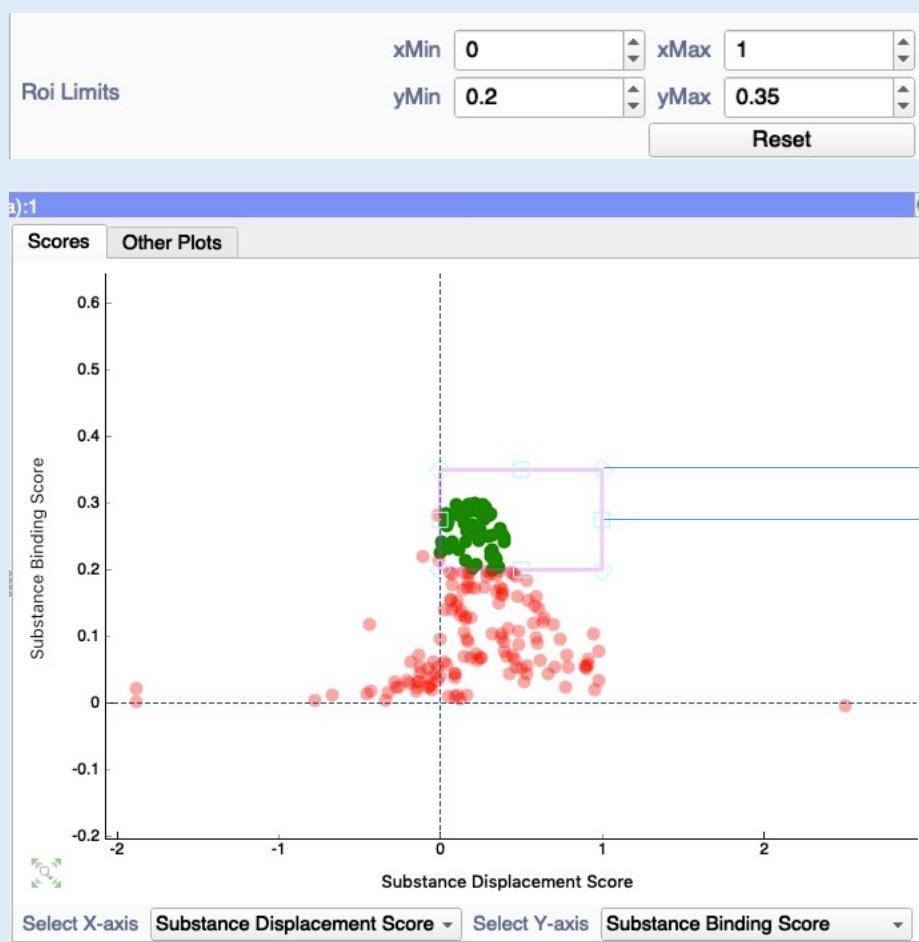
- Check **Use linked Substance Name**

- Run the pipeline using the green play button (click once only!). A popup will appear when completed.



### 3E Setup screening pipeline

- Clear the input data (right click, Clear all)
- On sidebar, multiselect 'SG:Control', 'SG:Target', 'SG:Displacer'
- drag and drop on the pipeline **Input Data**
- **close all pipes** (right click on any pipe header -> close all)
- On the Pipes list widget search and add to the pipeline area:
  - 1. Generic > Noise Threshold**
    - Calibration region: -120, -130 ppm
  - 2. AnalysisScreen > Propagate Peaks from References**
    - Propagate to: SG:Control
  - 3. AnalysisScreen > Copy PeakLists (1)**
    - Origin SpectrumGroup: SG:Control
    - Destination SpectrumGroup: SG:Target
  - 4. AnalysisScreen > Copy PeakLists (2)**
    - Origin SpectrumGroup: SG Control
    - Destination SpectrumGroup: SG:Displacer
  - 5. AnalysisScreen > Setup Screendataset**
    - Run name: 19F\_Pipeline
    - Reference SpectrumGroup: **None**
    - SG:Control, SG:Target, SG:Displacer on their respective entries
    - Matching engine: **Nearest Match**
    - set checked Use Substance ReferenceSpectra
- Run the pipeline using the green play button. Calculations will take less than a minute



Before flagging Substances, always inspect the matches with the table selections. Snap peaks with the shortcut SE, or correct matches from the two list widgets at the bottom right corner.

#### 4A Define Binding Hits Thresholds

- Main Menu → Screen → Hit Analysis (shortcut HA)
- Select Dataset: 19F\_Pipeline
- View mode: Substances
- Scatter Plot:

X-axis: Substance Displacement Score

Y-axis: Substance Binding Score

Open settings:

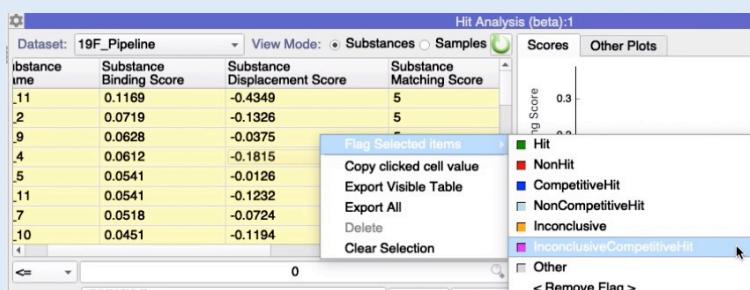
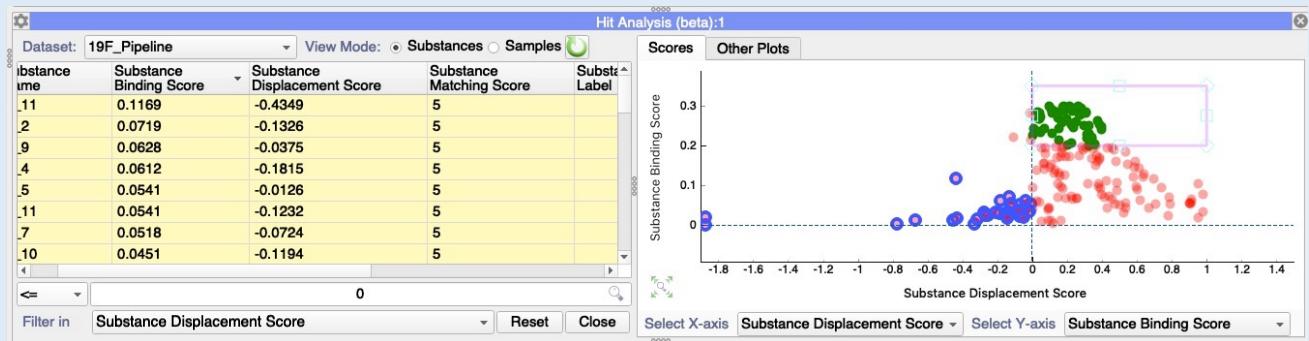
- Calculation tab
  - Engine: AbsoluteRelativeChange
- Appearance tab
  - Roi Limits: click reset
  - xMin: 0 xMax: 1
  - yMin: 0.2 , yMax: 0.35

this will create a Region of interest on the scatter plot

On the scatter plot:

- right click on top of a scatter item -> Select within ROI
- right click on top of a scatter item -> Flag Selected Items -> Hit

# Hit Analysis Module – Filters



The Hit analysis module has several filters for defining hits based on dynamically set threshold limits.

## 4B Filter Tables

Another way of defining hits is by applying one or multiple filter from the substance table:

- Right click on the Substance Table header → **filter**  
first filter:
  - select: “ $\leq$ ” 0.2 (less than)
  - filter in: Substance Binding Score
  - press the search button
 second filter:
  - select: “ $\leq$ ” 0
  - filter in: Substance Displacement Score
  - press the search button
- Select all rows
- if a display is opened, a warning will pop up: Click **No** to don't add all the selected spectra on the current Spectrum Display
- right click on a row: Flag Selected Items → Inconclusive Competitive Hit



#### 4C Extract and export

- Sort the substance table by Substance Label or filter by **Equal “Hit”** in Visible Table or Substance Label (**reset** any previous filter first)
- Select all the rows for Substances flagged as Hit
- Move to the scatter plot, right click -> **New from selected**  
This will create a new dataset containing only this subsets, named from the time-stamp. You can rename the dataset from the sidebar.
- Select the newly created dataset on the Hit Analysis module,
- Either inspect the substances as shown in Section 2 or export the table
- Right click on a Substance Table row -> **Export All**

#### 4D Export raw data

To get the raw data which is used to build the Hit Analysis Module

- Right click on a Substance Table row -> **Export Raw Data**
- On the file dialog, Save as: type a name plus the extension, e.g.: .xlsx, .csv, .tsv or .json

This will create a table which contains all peak metadata, including peak ppm position, height etc for further inspection or macros.

## Part 3: Recurring Analyses

### Schematic representation of a recurring Screen analysis workflow

When screening same mixtures of compounds against different targets, information about the reference spectra can be transferred across projects using NEF files.

Annotated peaks from singleton reference spectra can be transferred to a Reference Mixture. These will be used for matching future screening without the need of loading and peak picking again large reference libraries, therefore reducing the analysis time and potential errors.

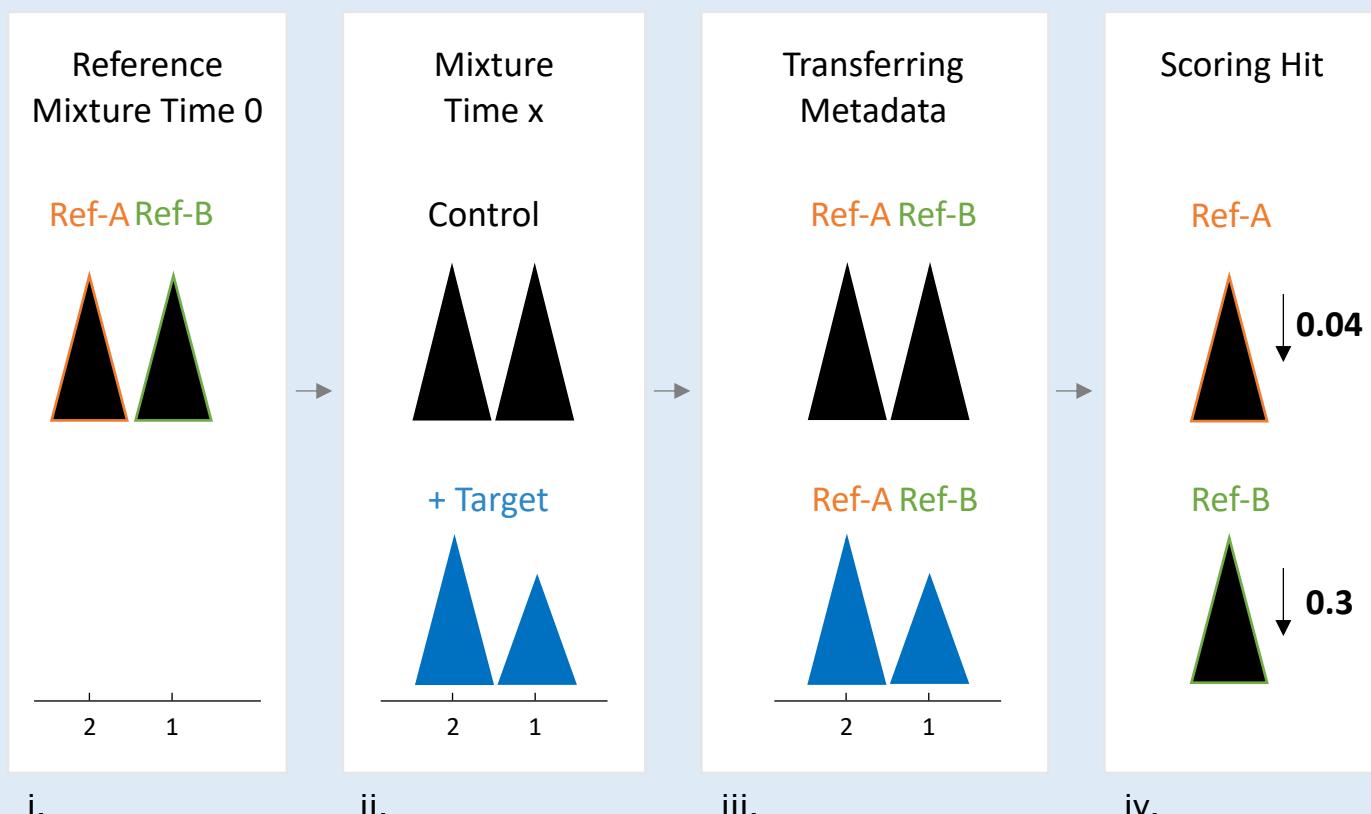


Figure 3. CcpNmr AnalysisScreen Recurring Analyses workflow

- The *Reference Mixture Time 0* represents a spectrum recorded at a time 0 for a sample containing only the components *Ref-A* and *Ref-B*. These are labeled in the peak annotation and contain links to the original Substance objects (see-next slides)
- The *Control* and *+Target* cartoons represent spectra recorded at a time X for samples containing only the components (*Control*) and plus a biological target (*+Target*)
- The *Reference Mixture Time 0* peaks and labels are transferred to the *Control* and *Target* spectral signals and peaks are refitted to the new spectra
- Spectral differences between the *Control* and *Target* are scored using different calculation methods. Scores are then transferred to the linked Substances as indication of their binding activity. (See Introduction Fig. 2 form more details)

In a NEF files are stored all the information regarding Samples, Substances, Spectra and Spectrum Groups

A NEF file can be created from an existing screening dataset with few modifications. These include **renaming** Samples and Spectrum Group from “Control” to “ReferenceMixture” (or similar), to avoid name-clashes; **deleting** the singleton reference spectra and finally **linking** the Substances to the Reference mixtures .

The pipe **LinkSubstancesToSpectrum** connects all substances presents in a mixtures to a Reference Mixture Spectrum

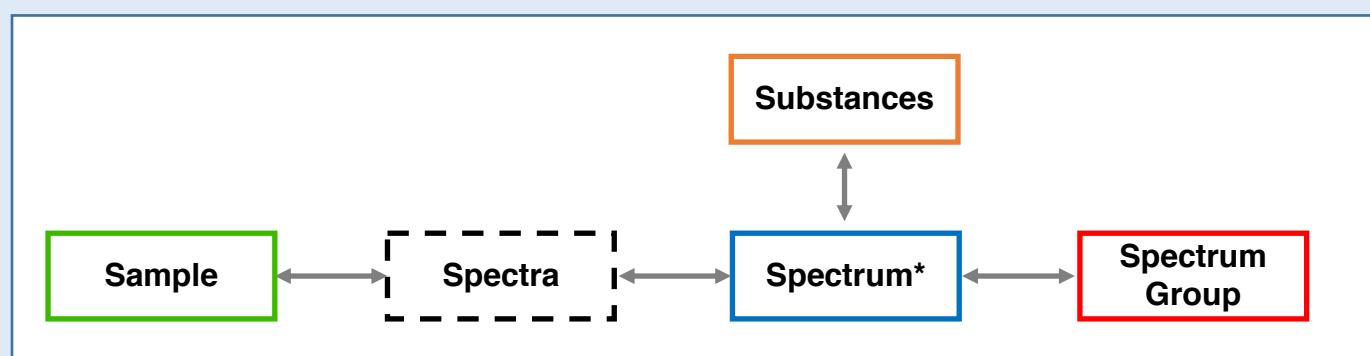


Figure 4. CcpNmr Screen object links for Recurring Analyses

\* Reference Mixture Spectrum at Time 0

## 5A Load and prepare Data

- Load the project 19F\_referencesPicked.ccpn from the dataset\_3 directory,  
.../AnalysisScreen\_Tutorial/19F/dataset\_3/19F\_referencesPicked.ccpn
- Open the Pipeline module
- Add 'SG:RefMix' on **Input Data**
- Add the pipe **Propagate Peaks from References**
  - select Propagate to: 'SG:RefMix' and keep the rest as default
- Add the pipe **LinkSubstancesToSpectrum**
- **Run**

## 5B Correct peaks (optional for this tutorial)

Because the NEF file will function as a template for future screening analyses, it is wise to inspect all Reference Mixtures, ensuring all peaks are correctly annotated to the respective Reference Substance

- Open the first sample, 'SA:Control\_1', right click -> Open Linked spectra or drag and drop it to the dropping area.
- inspect peaks on the Control\_1 are correctly positioned from the references, or use the shortcut “SE” to re-snap to extremum the selected peak(s)  
(You can change snapping limits from Preferences, Spectrum Tab, **1d Search Box Widths**)
- on sidebar, use the shortcut Ctrl (or Cmd)+up/down directional keys to visualise the next/previous sample and associated spectra.

## Peak annotation

When using NEF files as template for screening calculations, Substances are tracked in the Reference Mixture signals through the peak annotations.

**Peak annotation** name can be made by three parts:

Prefix	substance name	The exact substance name	Mandatory
Separator	_	Underscore	Mandatory if Suffix
Suffix	Any	Any single word tag that can help identify the signal; e.g.: impurity, TFA, Salt, a serial number, an atom name etc.	Optional

Peak annotation examples: “Compound1\_CF3, Compound2\_Salt, Unknown”

Peaks with figure of merit of 0 are excluded from screening calculations

## 5C Add extra annotations for impurities, solvents etc and exclude them from calculations

- Close all GUI Modules
- Drag and drop the ‘SG:RefMix’ on the dropping area
- Pick>Select all peaks in the region -99.9, -100.1 ppm
- open the Python Console from Menu > View or shortcut “Space-Space”
- Run the command:

```
tag = 'Impurity'

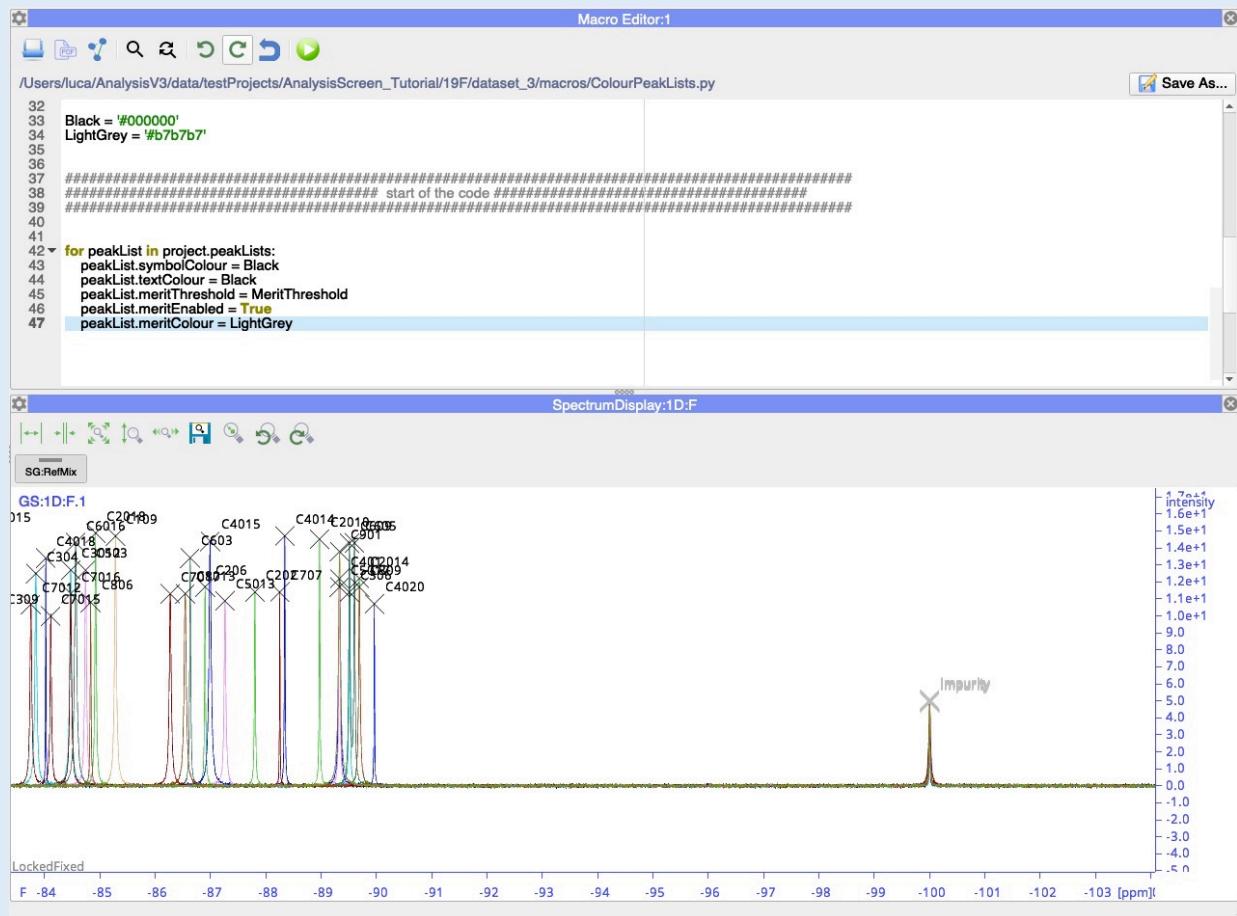
for peak in current.peaks:
    peak.annotation = '_'.join(filter(None, set([peak.annotation, tag])))
    peak.figureOfMerit = 0
```

Find the code as a Macro in the tutorial directory:

.../19F/dataset\_3/macros/AnnotateCurrentPeaks.py

5

# Handle spurious signals



## 5D Change colours for peaks

Change symbol/text colours so that real signal peaks (black) are graphically distinguishable from excluded (light grey)

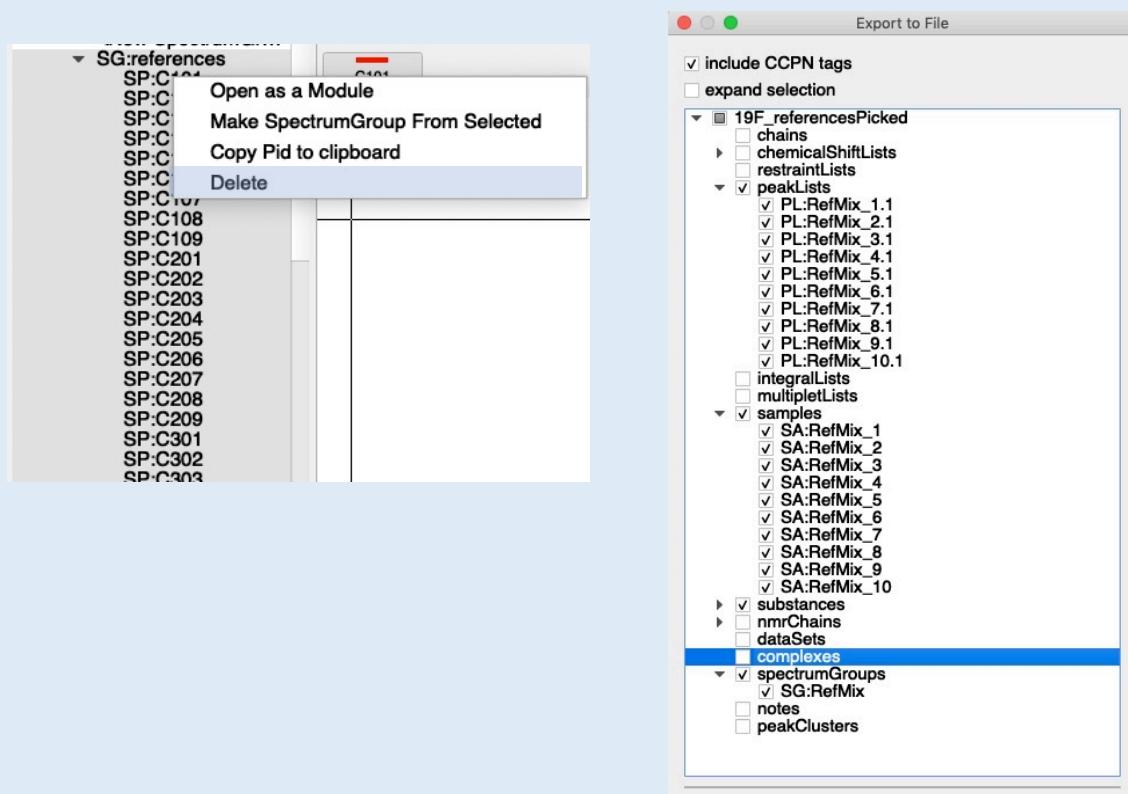
- on the Python run the command:

```
black = '#000000'  
lightGrey = '#b7b7b7'  
  
for peakList in project.peakLists:  
    peakList.symbolColour = black  
    peakList.textColour = black  
    peakList.meritThreshold = 0.5  
    peakList.meritEnabled = True  
    peakList.meritColour = lightGrey
```

Find the code as a Macro in the tutorial directory:

.../19F/dataset\_3/macros/ColourPeakLists.py

# Export to NEF



## 5E Delete Reference Spectra

On Sidebar:

- Expand the SpectrumGroup tree for References
- Select all spectra in the 'SG:references', including the SpectrumGroup
- 'SG:references' : right click -> **Delete**

Or run the command:

```
sg = get('SG:references')
project.deleteObjects(*list(sg.spectra)+[sg])
```

## 5F Export to NEF

Export metadata related to the Reference Mixtures

- Main Menu -> Project -> Export -> NEF File
- On the NEF Dialog Set checked:

General:

include CCPN Tags,

Project tree:

PeakLists

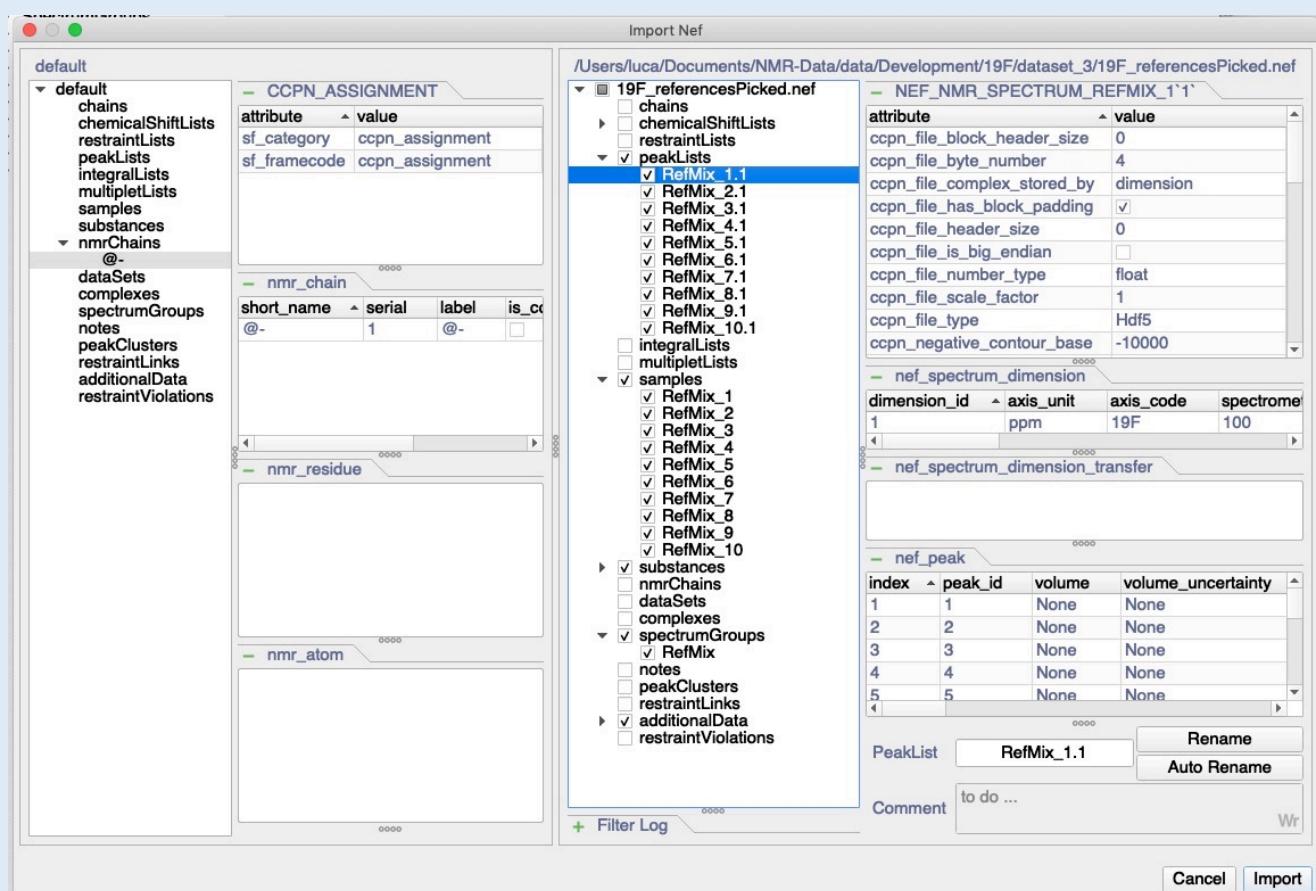
Samples

Substances

SpectrumGroups

- Save to your local disk

# Screen From NEF



## 6A Import From NEF

- Start a new project
- Main Menu → Project → Import → NEF File
- Select the lately created NEF file or the file provided in the tutorial data directory
- On the NEF Dialog Set checked:
  - Project tree:
    - PeakLists
    - Samples
    - Substances
    - SpectrumGroups
    - AdditionalData
  - Import

Make sure there are not name clashes

If there are empty peak lists, run on PythonConsole:

```
toDelObjs = [pl for pl in project.peakLists if not pl.peaks]
project.deleteObjects(*toDelObjs)
```

# Screen From NEF



Sometimes your latest dataset might be slightly offset compared to the Reference Mixtures in the NEF files. In that case and before copying peaks, you could re-reference the X-axes using the pipe **Align Spectra** and the Y-axes using the pipe **Scale spectra**.

## 6B Import Data from Excel

Import the latest screening data from the excel file

- Load the excel file from the tutorial data folder:

`.../19F/dataset_3/Data_Time_x/lookup_19F_TimeX.xlsx`

Note how this lookup only contains the Sample sheets without the SampleComponents field and the Substances sheet. This information has been carried over from the NEF file.

## 6C Setup screening pipeline

- Open the Pipeline (Menu -> Screen -> Pipeline or shortcut PI)
- On sidebar, multiselect 'SG:Control', 'SG:Target', 'SG:RefMix'
- drag and drop on the pipeline **Input Data**
- On the Pipes list widget search and add to the pipeline area:

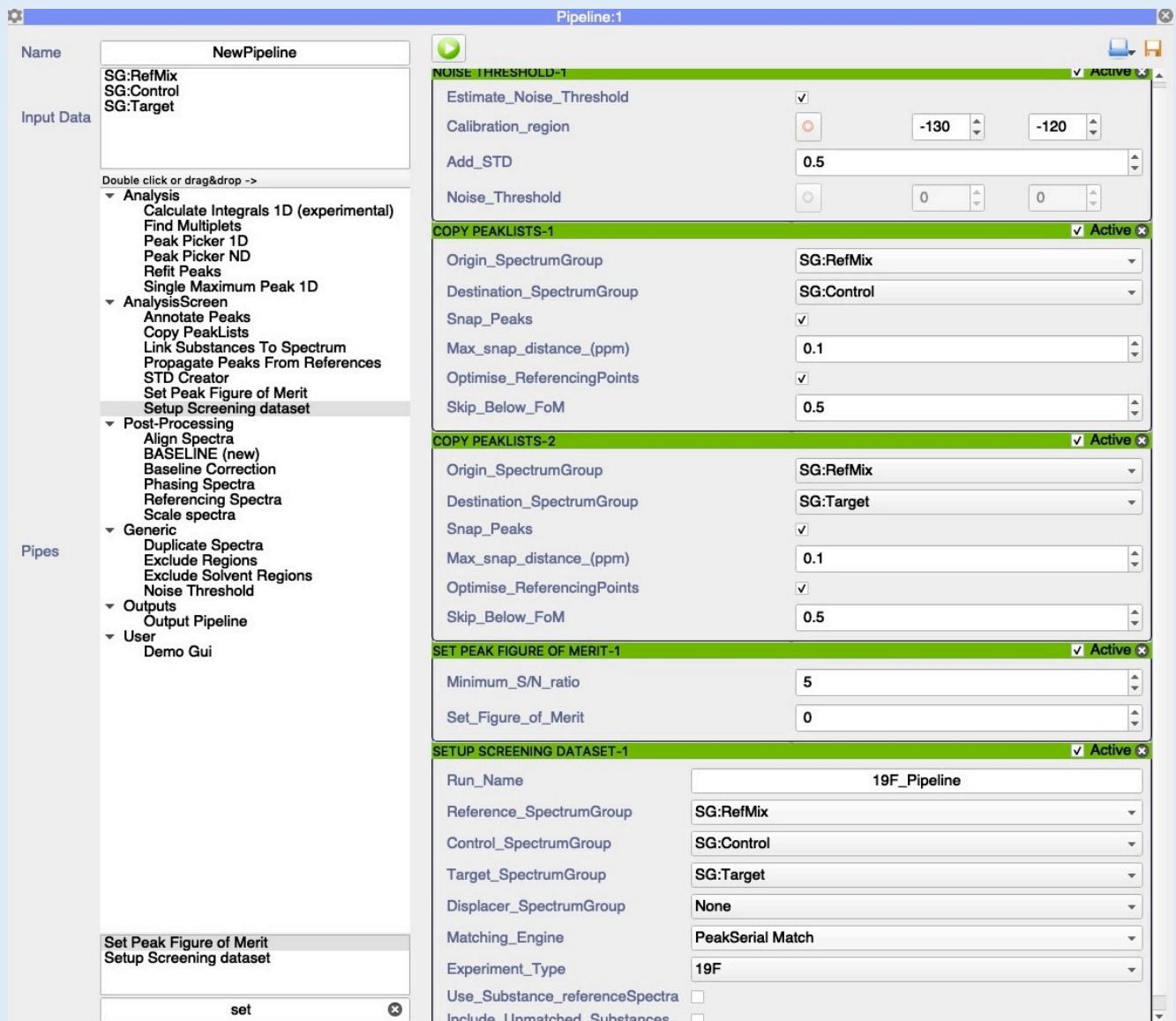
### 1. Noise Threshold

- set Checked Estimate Noise Threshold
- Calibration region -130, -120

### 2. Copy PeakLists (1)

- Origin SpectrumGroup: SG:RefMix
- Destination SpectrumGroup: SG:Control

*... Continues Next*



## 6c ...Continued

add the following pipes

### 3. Copy PeakLists (2)

- Origin SpectrumGroup: SG:RefMix
- Destination SpectrumGroup: SG:Target

### 4. Set Peak Figure of Merit

- Minimum S/N ratio: 5
- Merit 0

### 5. Setup ScreenDataset

- Run name: 19F\_Pipeline
- Reference SpectrumGroup: SG:RefMix
- SG:Control, SG:Target on their respective entries
- Matching engine: **Peak Serial Match**
- Uncheck Use\_Substance\_ReferenceSpectra

- Run the pipeline
- Check results on the Hit Analysis Module as in Section 4

# Screen From NEF



## 6D Completed Project

An example of a completed project is available in the dataset\_3 directory:

.../19F/dataset\_3/19F\_NEF\_Completed.ccpn

### Hit Analysis Module

On selection of Substance or Sample items from the main tables, you may notice how the Singleton Reference Spectra are replaced by the Reference Mixture spectra. The peak annotations will now provide a visual reference to the matching Substance.

## Contact Us

**Website:**

[www ccpn.ac.uk](http://www ccpn.ac.uk)

**Suggestions and comments:**

[support@ccpn.ac.uk](mailto:support@ccpn.ac.uk)

**Issues and bug reports:**

<https://www ccpn.ac.uk/forums>

## Cite Us

Mureddu, L. et al. CcpNmr AnalysisScreen, a new software programme with dedicated automated analysis tools for fragment-based drug discovery by NMR. *J. Biomol. NMR* (2020)

Skinner, S. P. et al. CcpNmr AnalysisAssign: a flexible platform for integrated NMR analysis. *J. Biomol. NMR* 66, (2016)