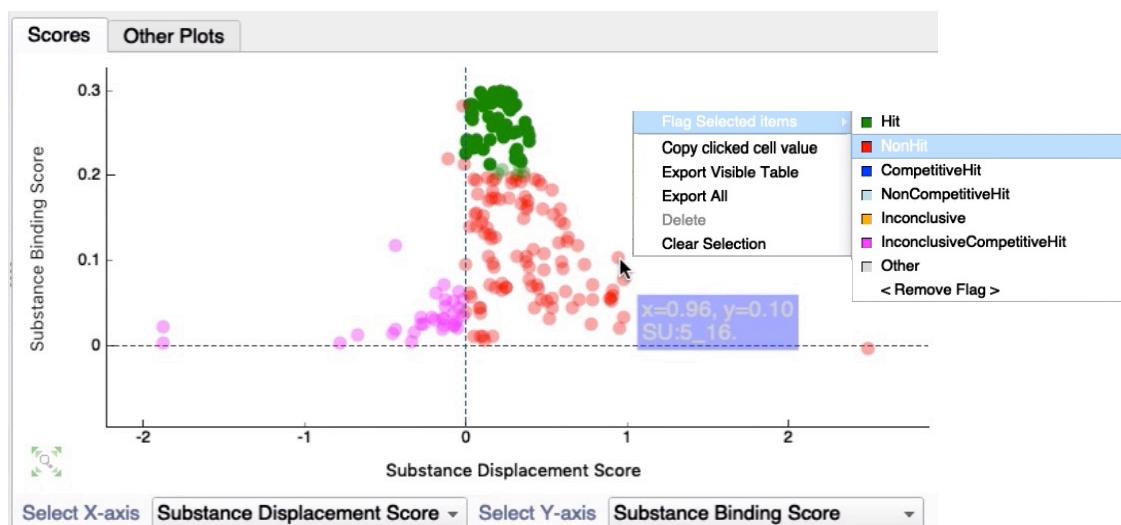


AnalysisScreen Hit Analysis Tutorial



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 |
| Space-Space | → 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 physical NMR 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 physical NMR sample (e.g., concentration).

CcpNmr links: Substance

Substance

A CcpNmr object containing information about a molecule, (e.g., a 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 or multiple substances plus a biological target and a known binder. Also referred to as a “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 screening Dataset.

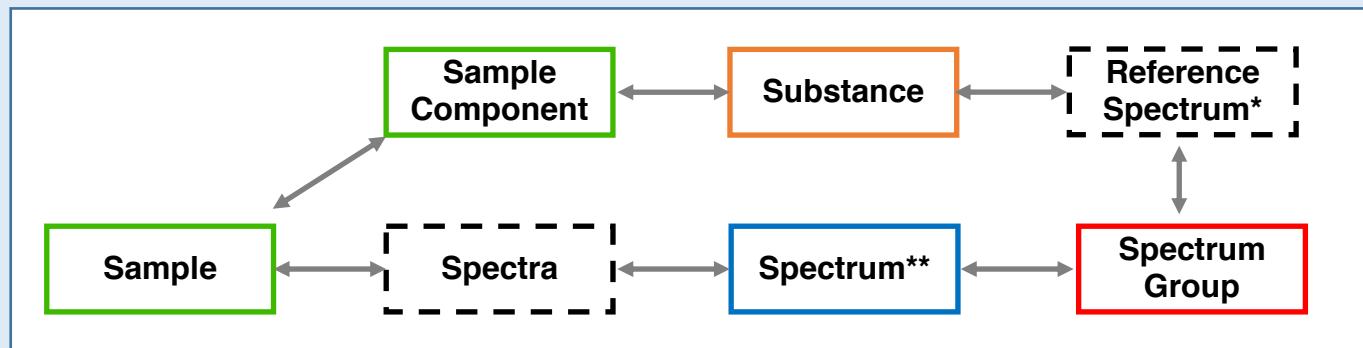


Figure 1. CcpNmr AnalysisScreen object links

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 several ^{19}F datasets following the workflow shown in Fig. 2. These steps and their tools in AnalysisScreen can also be applied to other experiment type analyses, such as: ^1H Relaxation-edited, WaterLOGSY and STD.

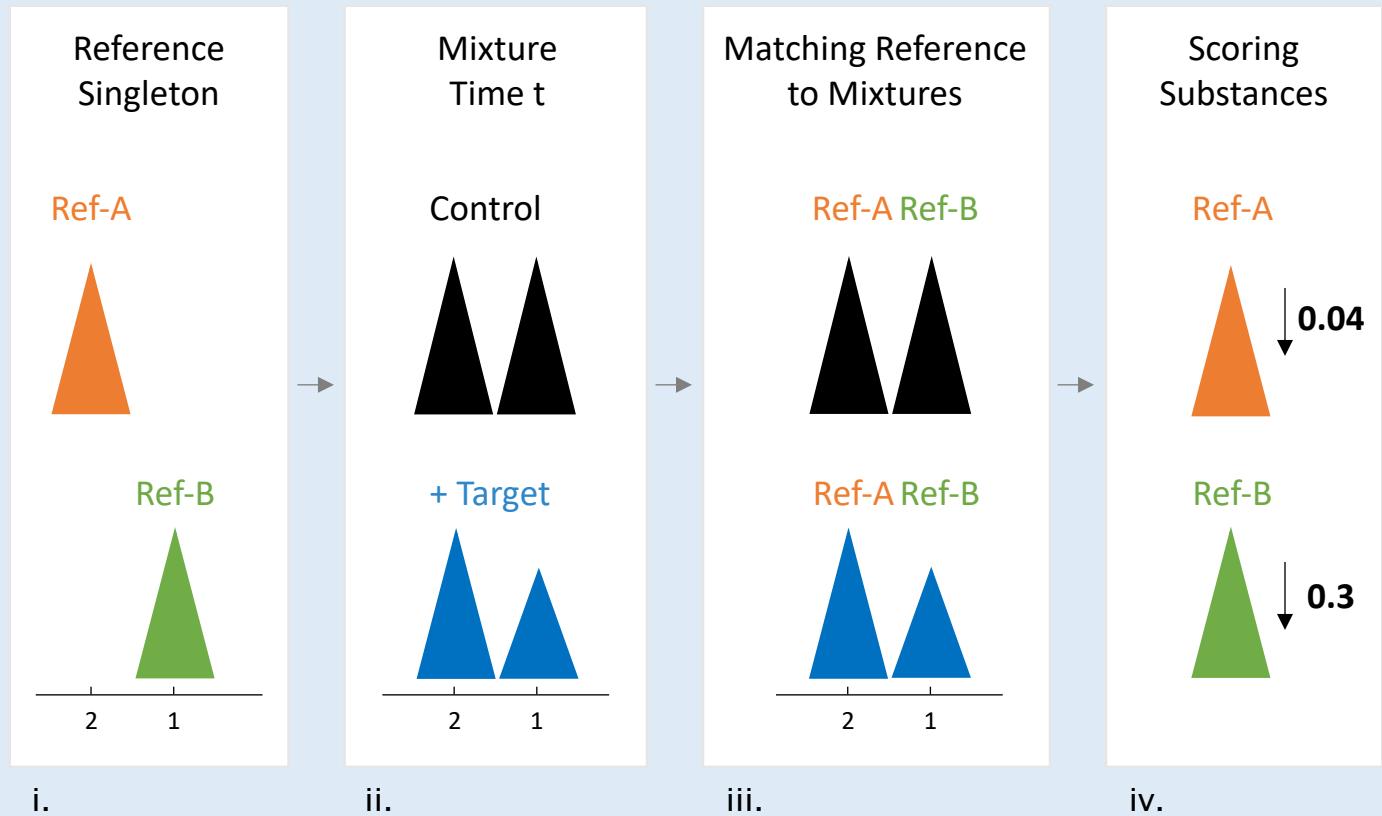


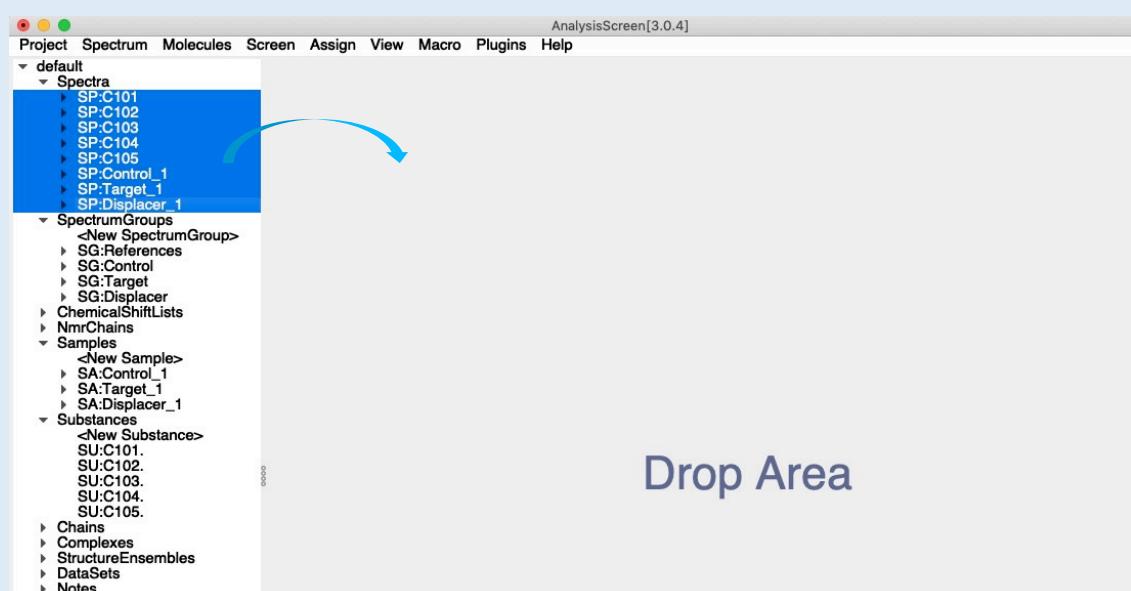
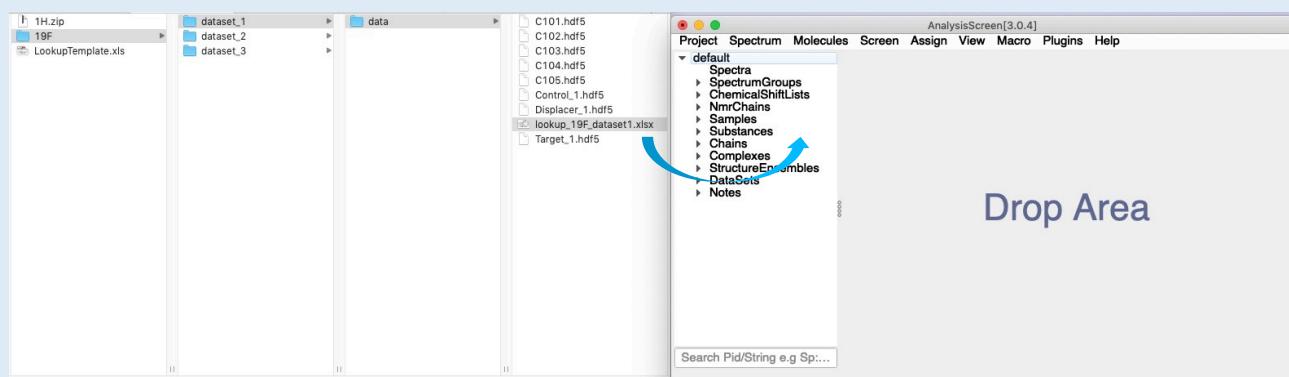
Figure 2. CcpNmr AnalysisScreen common workflow

- 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 time t 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 an 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 a correctly formatted Excel file, all the necessary links are automatically established ensuring optimal functioning of the screening tools.

See the “HowTos_SidebarObjects” and the “HowTos_ImportDataFromExcel” manuals for more information.

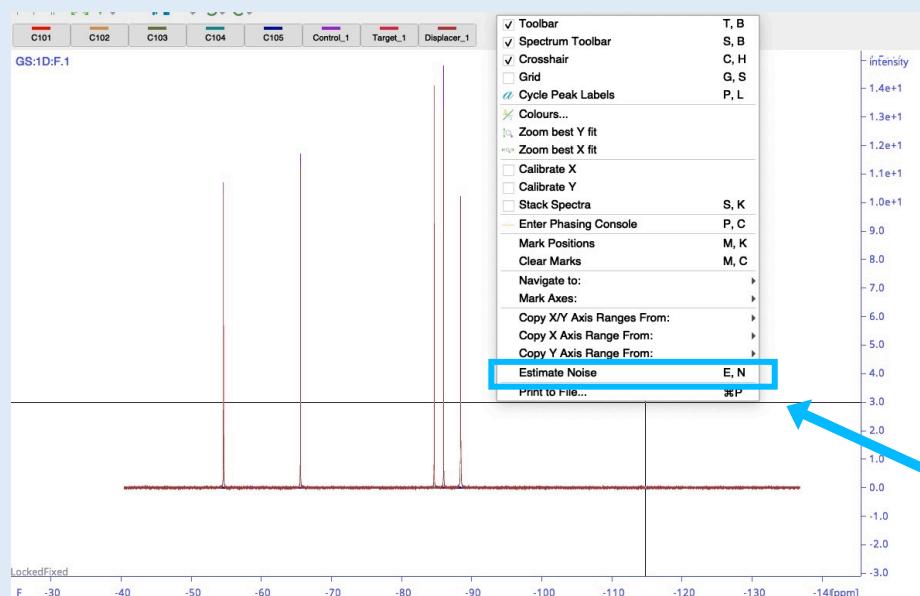
1A Drag & drop the file **lookup_19F.xlsx** into the program

- 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

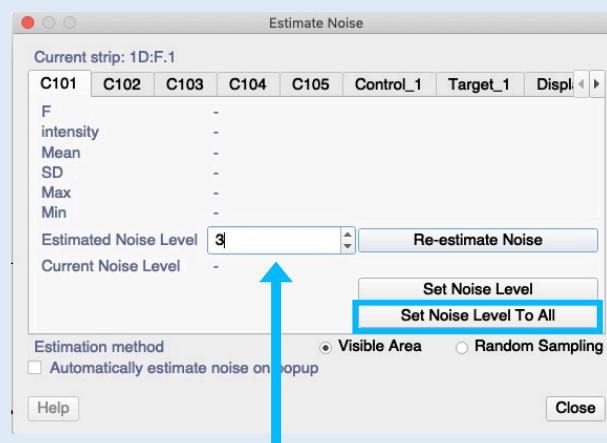
1B Open all spectra

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

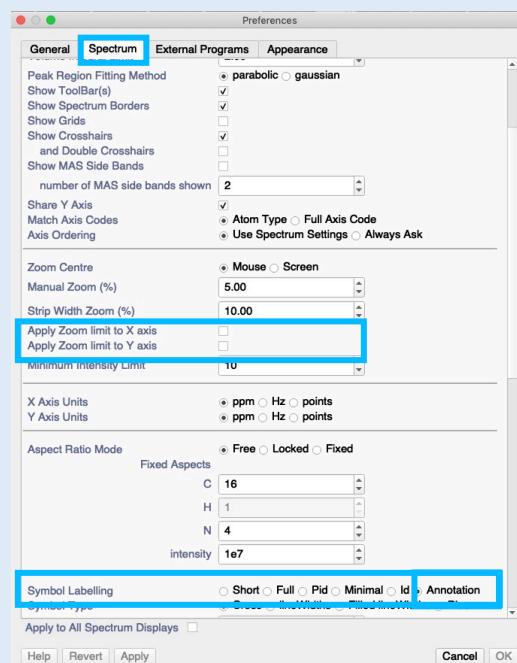
Prepare Data



Right click on the canvas at intensity ~3



When picking 1D peaks, only peaks above this intensity level will be picked.



1c Set noise level

In the Spectrum 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

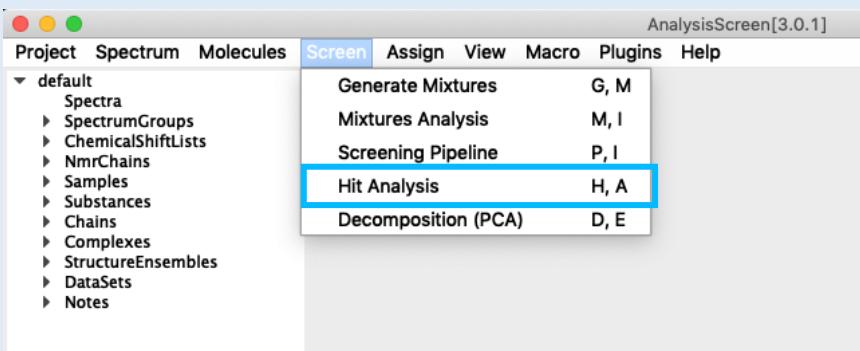
- untick Apply zoom limit both for X and Y axes
- Set Annotation for Symbol Labelling

1d Stack Spectra

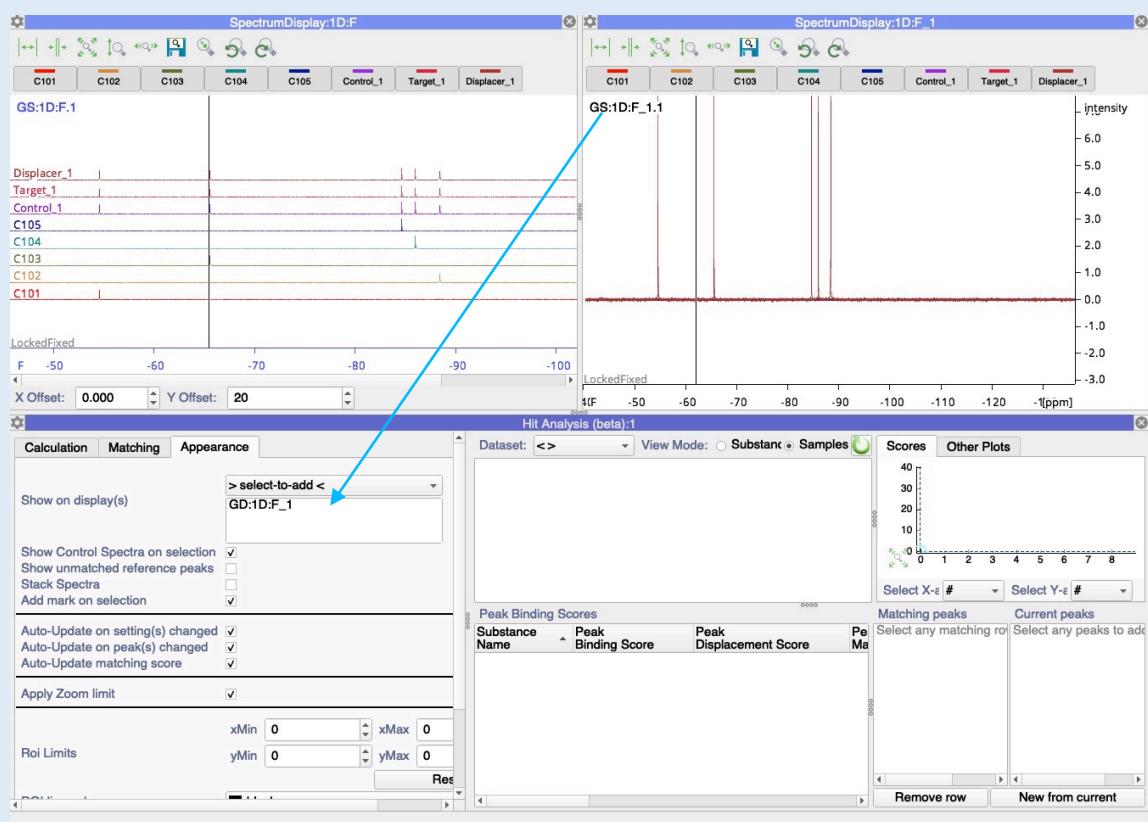
In the Spectrum Display :

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

Hit Analysis Module setup



Shortcut
"HA"



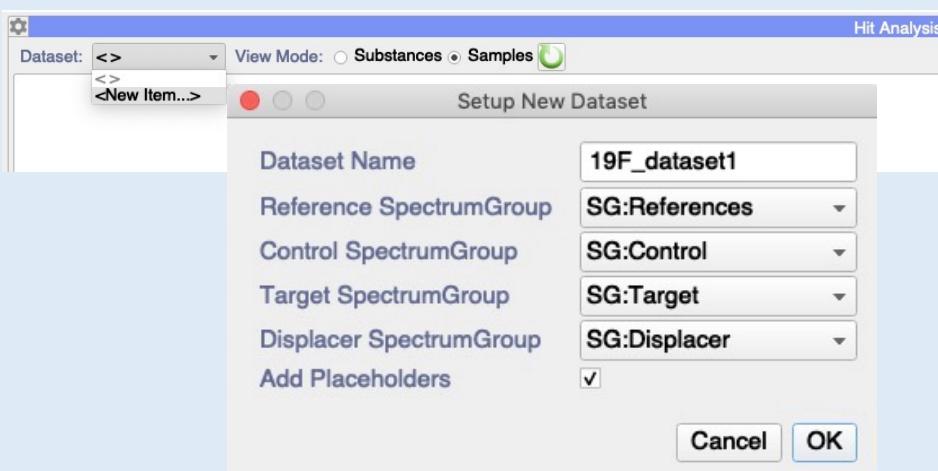
2A Open the Hit Analysis Module

- 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 soon perform a series of dynamic actions in this display.
- Close the settings panel.
- Re-arrange the Hit Analysis Module below the Spectrum Displays 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.

Pre-populate tables with blanks for Substances and Samples present in 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

In the Hit Analysis module:

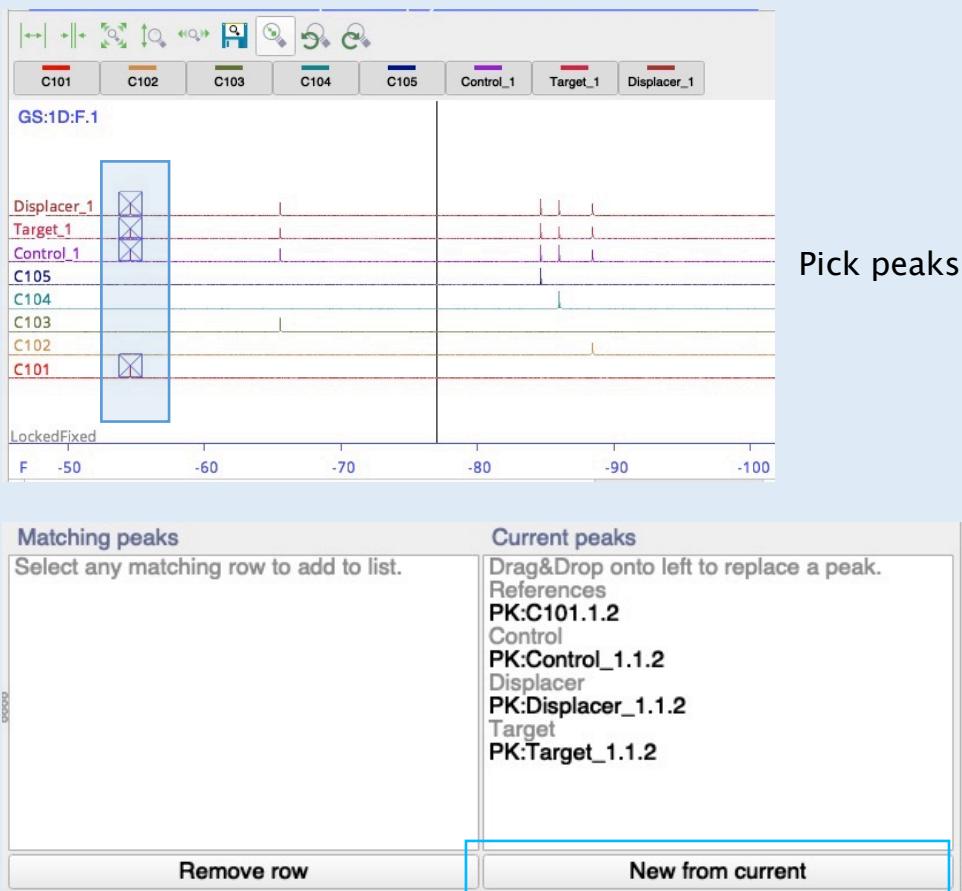
- Select **<New Item...>** in the **Dataset** pulldown,
- Change the name to **19F_dataset1** or keep the default.
- Tick **Pre-populate Tables**.
- Press **Ok** to proceed and close the popup.

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

- Set View Mode as **Substances**.
- In 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 the peaks at ~ -54.59 ppm in the stacked Spectrum Display:

- Use **CTRL** (or **CMD** for Mac) + **SHIFT** + **Left-drag** to create a (blue) picking region to include all spectral signals around -54 ppm.
- This signal correspond to the Substance **C101**.

In the Hit Module:

- the **Current peaks** will appear in 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 Module. 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 labeled 1 to 5, corresponding to substances C101, C102, C103, C104, and C105. The "C101" row is highlighted. An arrow points from the text "Substance Table" to this row. To the right of the substance table is a "Peak Matches Table" with columns: Substance Name, Peak Binding Score, Peak Displacement Score, Peak Position Reference, Peak Position Control, Peak Position Target, Peak Position Displacer, Peak Position Displacer, Reference-Target Δ Shift, Control-Target Δ Shift, and Comment. The "C101" row is highlighted. To the right of the peak matches table is a "Matching peaks" table with columns: References, Control, Target, and Displacer. The "C101" row is highlighted. At the bottom right of the interface is a "Remove row" button.

Multi-picking and semi-automatic peak matching



2E Semi-automatic picking and matching

In the Spectrum Display:

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

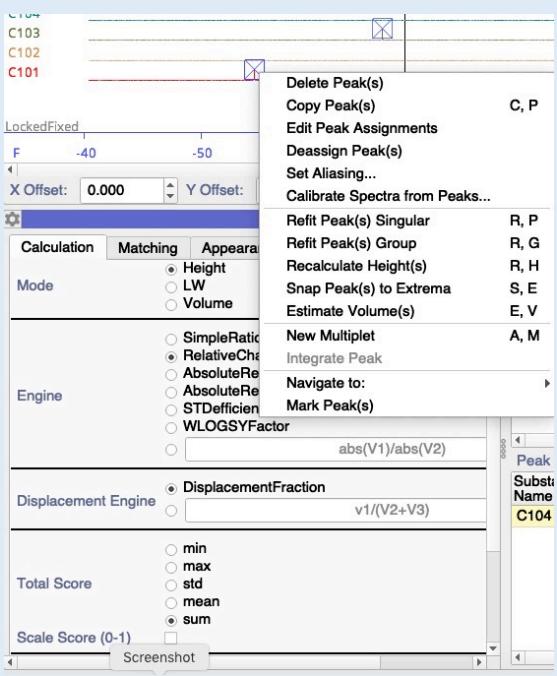
In the Hit Analysis Module:

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

The tables will update:
for each substance you
select in the upper table, a
peak binding score is visible
in the lower table.

Dataset: 19F_dataset1		View Mode: <input checked="" type="radio"/> Substances <input type="radio"/> 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

In the Hit Analysis Module:

- Select **Substances** in the **View Mode** and sort by **Substance Binding Score** (click on the column header).

The **Substance Binding Score** gives an indication of the spectral changes between the Control and Target spectra for 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 the following properties: height, linewidth or volume. As default, linewidths and volumes are not calculated. To calculate:

- Select all peaks from display **1D:F.1** (e.g. with CTRL/CMD+A), type shortcut **RP** to refit all linewidths, and **EV** to estimate volumes.

Open the Hit Analysis Module settings from the gear icon and select the **Calculation** tab.

The binding score is given by the calculation **Engine**.

- Change the default by selecting one of the options. When hovering over the labels for each option, a Tooltip window will show the equation used.
- You can also define your own 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:

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

The same applies to the **Displacement Engine** which gives the **Substance Displacement Score**. When writing your own equation, define the Displacer signal with the variable **V3**.

# Hit Analysis Module

The screenshot shows the Hit Analysis Module interface. On the left, there are several configuration sections:

- Calculation:** Mode (Height, LW, Volume), Engine (SimpleRatio, RelativeChange, **AbsoluteRelativeChange**, AbsoluteRelativeDifference, STDeficiency, WLOGSYFactor).
- Displacement Engine:** DisplacementFraction.
- Total Score:** sum (min, max, std, mean).
- Scale Score (0-1):** .

A tooltip for the **AbsoluteRelativeChange** engine is displayed, explaining the formula  $|V2-V1|/|V2|$  where  $V1$  is the target ligand signal and  $V2$  is the control ligand signal. It notes that a higher value indicates a greater value change, such as a larger intensity drop.

On the right, there is a dataset view with columns for #, Substance Name, Substan, Binding, Substance Name, and Peak Position Reference.

At the bottom, the title "Hit Analysis (beta):1" is shown, along with "Dataset: 19F\_dataset1" and "View Mode:  Substances  Samples" with a refresh button.

## Scoring Engines (continued)

Note that the **Substance Binding Score** is derived from all the peaks matched to that substance.

In this  $^{19}\text{F}$  demo dataset, only one observation is recorded per reference spectrum (as is typical for  $^{19}\text{F}$  data).

However, if multiple peaks per substance are present (as might be the case for  $^1\text{H}$  data), 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 your preferred option from the **Total Score** Settings.

If the **View Mode** is set to **Samples**, a **Sample Binding Score** is shown and derived from the **Substance Binding Scores** of all substances in that sample.

**As soon as you change settings or modify/fit peaks, the whole module will update and recalculate all scores automatically.** For larger datasets this can be time-expensive. You can disable this feature:

- In **Settings -> Appearance tab**
- untick **Auto-Updates on setting(s) changed**
- untick **Auto-Updates on peak(s) changed**

The refresh button  will turn red whenever changes are detected while working on the dataset. Click the refresh button to update all scores after which it will turn green.

# Hit Analysis Module

## Table Selections

The Hit Analysis module has multiple dynamic selections

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

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

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

1. List all contributing peak 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 reference spectra

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

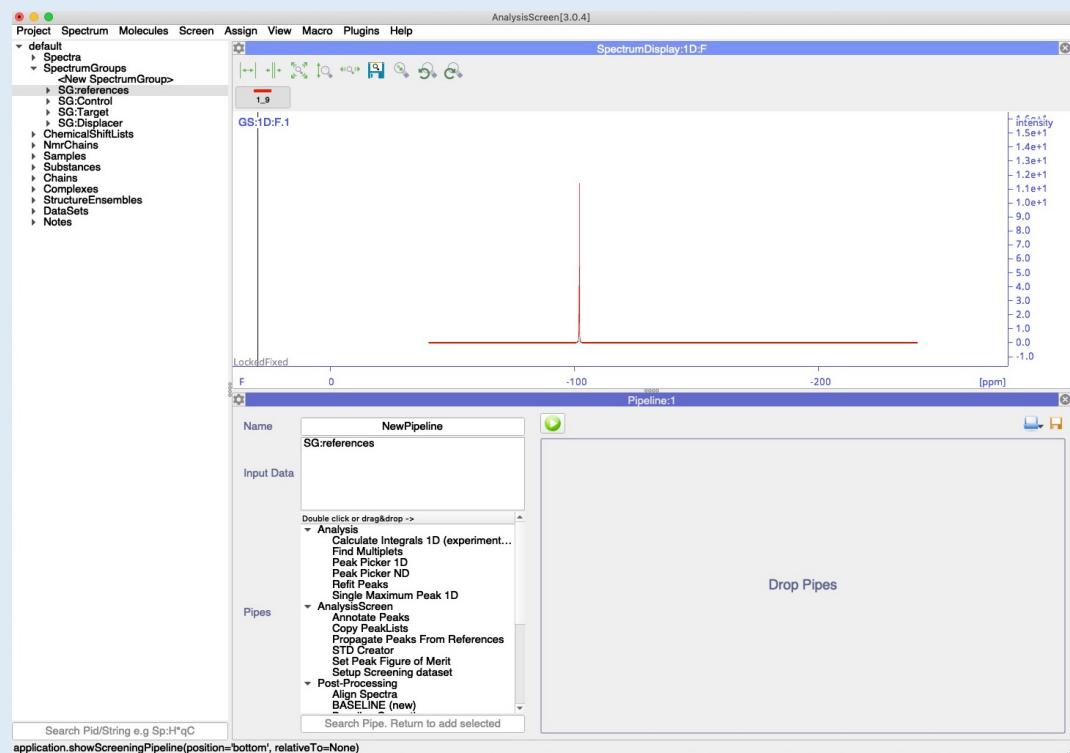
1. Select all peaks included in the match
2. Navigate to the Peak Position
3. Populate the **Matching/Current Peaks** lists widgets

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

The screenshot shows the Hit Analysis (beta):1 interface. At the top, there's a navigation bar with 'Dataset: 19F\_dataset1' and 'View Mode: Substances'. Below it is the 'Substances Table' with columns: #, Substance Name, Substance Binding Score, Substance Displacement Score, Substance Matching Score, Substance Label, Relative S/N Ratio, and Sample Name. Row 1 (C101) is highlighted. A large orange box surrounds the first four columns of the table, with the text 'Substances Table' overlaid. Below the table is the 'Peak Binding Scores' section, which includes a 'Peak Matches Table' with columns: Substance Name, Peak Binding Score, Peak Displa, Peak Position Target, and Peak Position Displacer. Rows for C101 and C102 are listed.

The screenshot shows the Hit Analysis (beta):1 interface. At the top, there's a navigation bar with 'Dataset: 19F\_dataset1' and 'View Mode: Samples'. Below it is the 'Samples Table' with columns: #, Sample Name, Target Name, Sample Binding Score, Sample Displacement Score, Sample Matching Score, and Comment. Row 1 (Target\_1) is highlighted. A large blue box surrounds the first four columns of the table, with the text 'Samples Table' overlaid. Below the table is the 'Peak Binding Scores' section, which includes a 'Peak Matches Table' with columns: Substance Name, Peak Binding Score, Peak Displacement Score, Peak Position Reference, Peak Position Control, Peak Position Target, and Peak Position Displacer. Rows for C101, C103, C105, C104, and C102 are listed.

# Part 2: Automatic Analysis with Pipelines



## Automatic peak matching and binding analysis

In this section, a pipeline is built to inspect changes between spectra 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

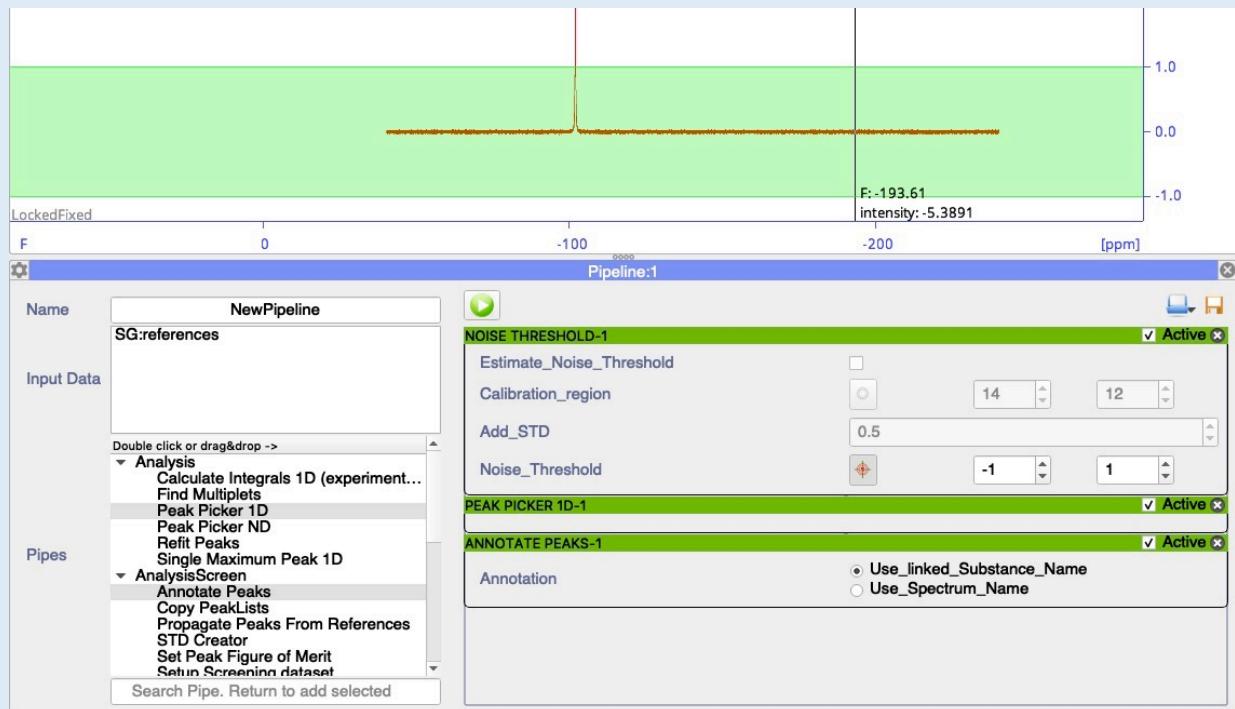
- Open a new project
  - Locate the data of dataset\_2 inside the folder:  
AnalysisV3/data/testProjects/AnalysisScreen\_Tutorial/19F/dataset\_2/data
  - Drag & drop the file **lookup\_19F\_dataset2.xlsx** into the sidebar or drop area.
- See the **HowTos\_ImportDataFromExcel** for creating excel files.

### 3B Open the first reference spectrum

- select the first spectrum in the 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-ordered by holding and dragging the green top bar. See **HowTos\_Pipelines** for more information.

### 3D Pick peaks on SG:References

- In the Pipes list widget search for and add the following pipes to the pipeline area by double clicking the pipe name or via drag & drop:

#### 1. Generic > Noise Threshold

- untick **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

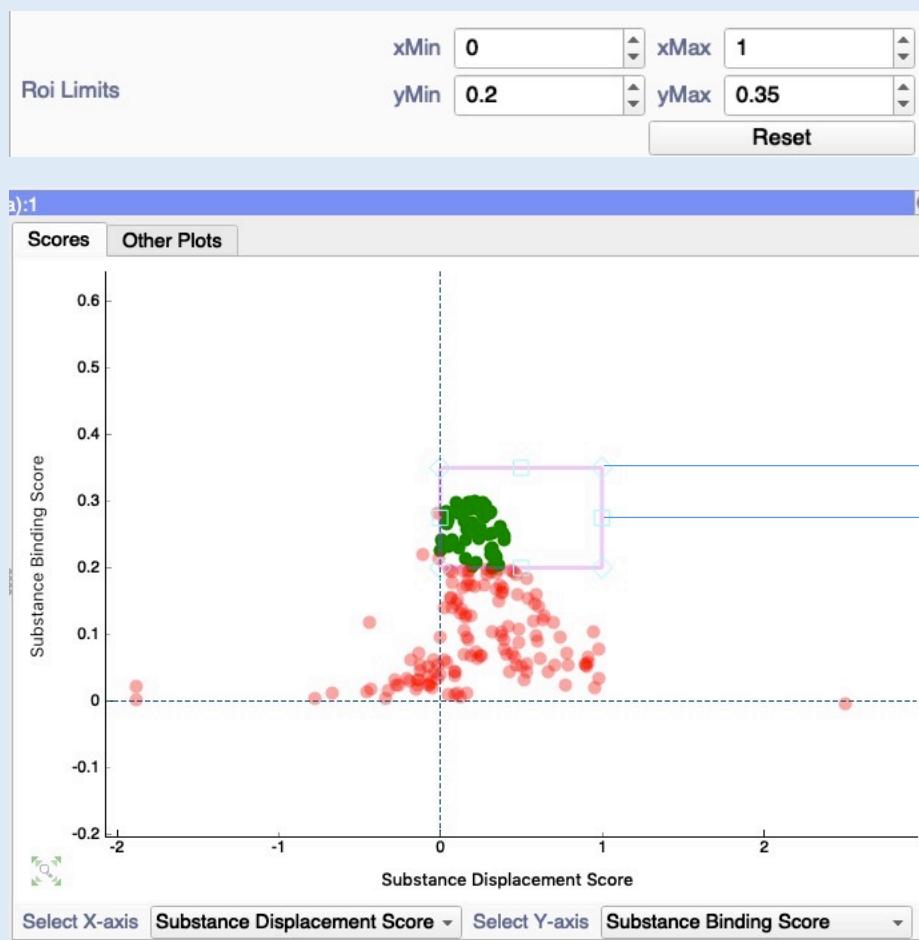
- Tick **Use linked Substance Name**

- 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' and then drag & drop into the pipeline **Input Data** area.
- Close all pipes (**right-click** on any pipe header → **Close All**)
- In the list of **Pipes** search for and add these pipes 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 Screening dataset**
    - **Run name:** 19F\_Pipeline
    - **Reference SpectrumGroup:** None
    - select **SG:Control**, **SG:Target**, **SG:Displacer** for their respective entries
    - **Matching Engine:** Nearest Match
    - tick **Use Substance ReferenceSpectra**
- Run the pipeline using the green play button. The calculations should 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:

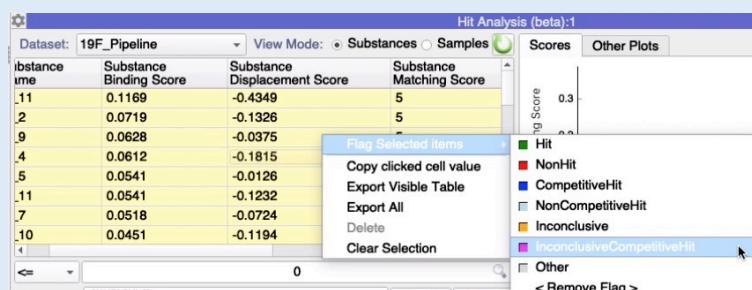
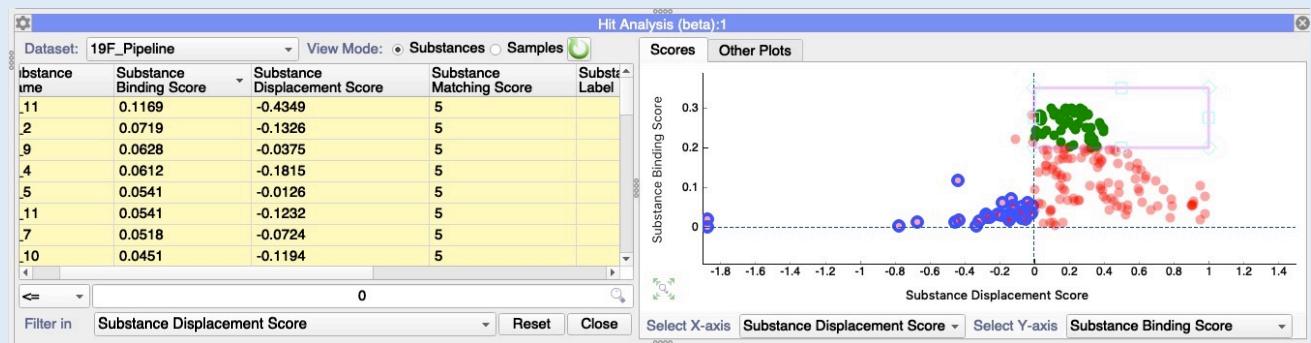
- Calculation tab
  - Engine: AbsoluteRelativeChange
- Appearance tab
  - **Roi Limits:** click **Reset**
  - xMin: 0 xMax: 1
  - yMin: 0.2 , yMax: 0.35

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

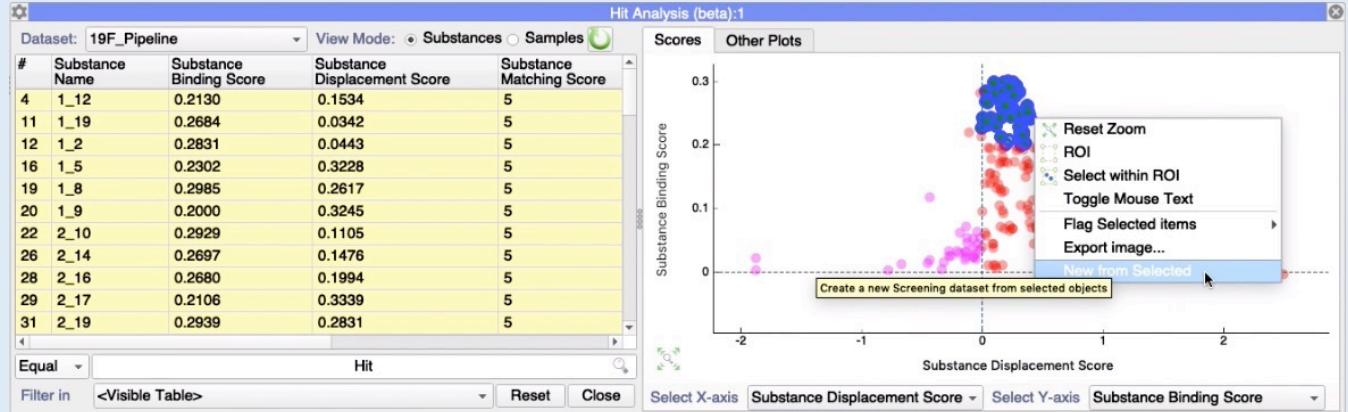


The Hit analysis module has several filters for defining hits based on dynamically set threshold limits.

## 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 open, a warning will pop up: Click **No** to not 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** (click on the column header) 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 subset of substances, named from the time-stamp. You can rename the dataset in the sidebar.
- Select the newly created dataset on the Hit Analysis module, by selecting it from the **Dataset** dropdown menu.

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 export the raw data which is used to build the Hit Analysis Module:

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

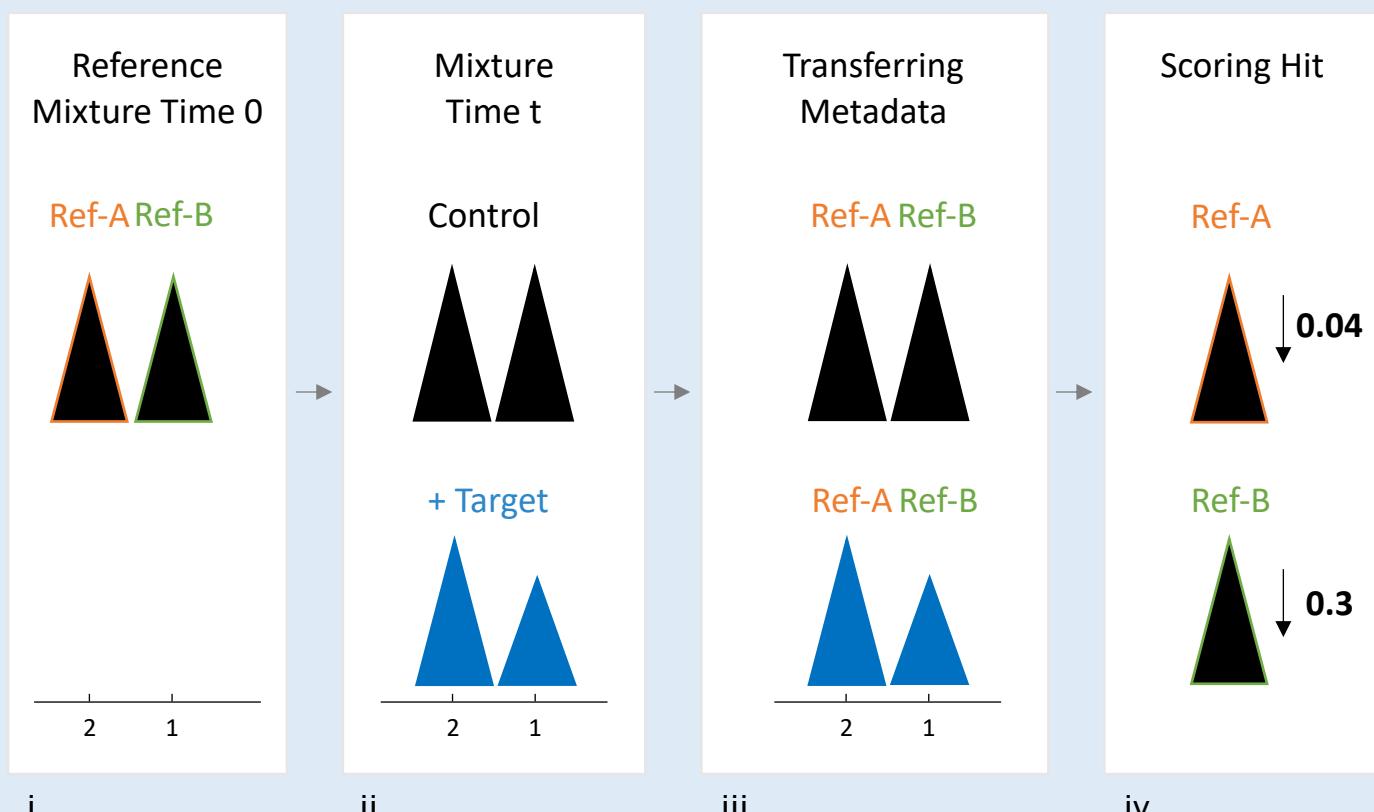
This will create a table which contains all peak metadata, including peak ppm position, height etc. for further inspection or macros.

## Part 3: Recurring Analyses

### Schematic representation of a recurring Screen analysis workflow

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

Annotated peaks from singleton reference spectra can be transferred to a Reference Mixture. These will be used to match future screening data without the need to load and peak large reference libraries again, thus reducing the analysis time and potential errors. This is particular helpful if peaks in the control mixtures have moved relative to the reference singleton spectra due to changes in sample conditions such as pH.



**Figure 3. CcpNmr AnalysisScreen Recurring Analyses workflow**

- The *Reference Mixture Time 0* represents a spectrum recorded at 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 time t 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 an indication of their binding activity. (See Introduction Fig. 2 for more details.)

# Create NEF File

All the information regarding Samples, Substances, Spectra and Spectrum Groups are contained within a NEF file.

A NEF file can be created from an existing screening dataset with only a few modifications. These include **renaming** Samples and Spectrum Groups 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 mixture to a Reference Mixture Spectrum.

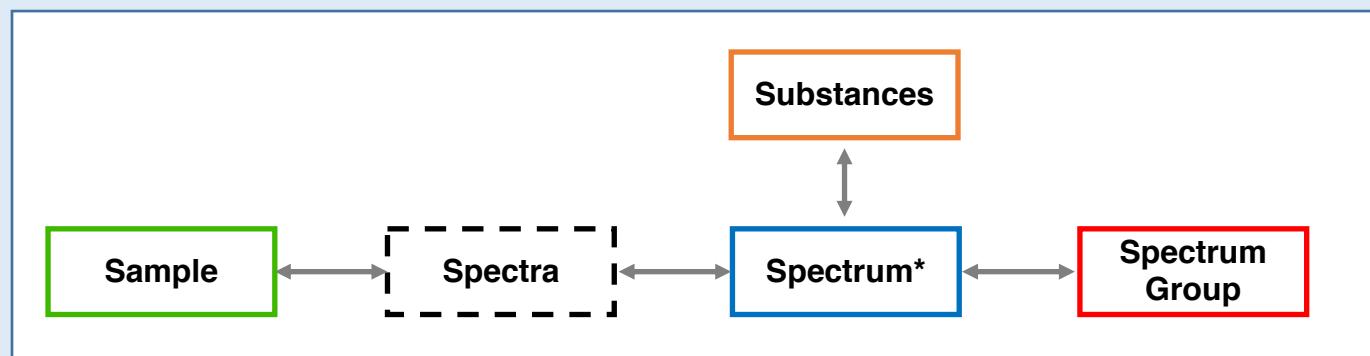


Figure 4. CcpNmr Screen object links for Recurring Analyses

\* 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 with **PI** shortcut or from main menu
- Add the **SG:RefMix** Spectrum Group to **Input Data**
- Add the pipe **Propagate Peaks from References**
  - select **Propagate to:** **SG:RefMix** and keep the other defaults
- 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 with the respective Reference Substance:

- In the sidebar, go to the first sample **SA:RefMix\_1** and **right-click → Open Linked spectra** or drag & drop it into the Drop Area.
- Make sure the **RefMix\_1** peaks are correctly positioned compared to the references, or use **SE** to re-snap the selected peak(s) to their extremum. (You can change snapping limits in Preferences, Spectrum Tab, **1d Search Box Widths**)
- In the 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 a template for screening calculations, Substances are tracked in the Reference Mixture signals through the peak annotations.

**Peak annotation** names 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 a 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** Spectrum Group from Sidebar into the Drop Area
- Pick and select all peaks in the region -99.9, -100.1 ppm
- open the **Macro Editor** from **Macro → New**
- Drag and drop over the text editor the file **AnnotateCurrentPeaks** located in the macros tutorial directory:  
`.../19F/dataset_3/macros/AnnotateCurrentPeaks.py`
- Run with the play button (while the peaks are selected)

A code snippet is shown below

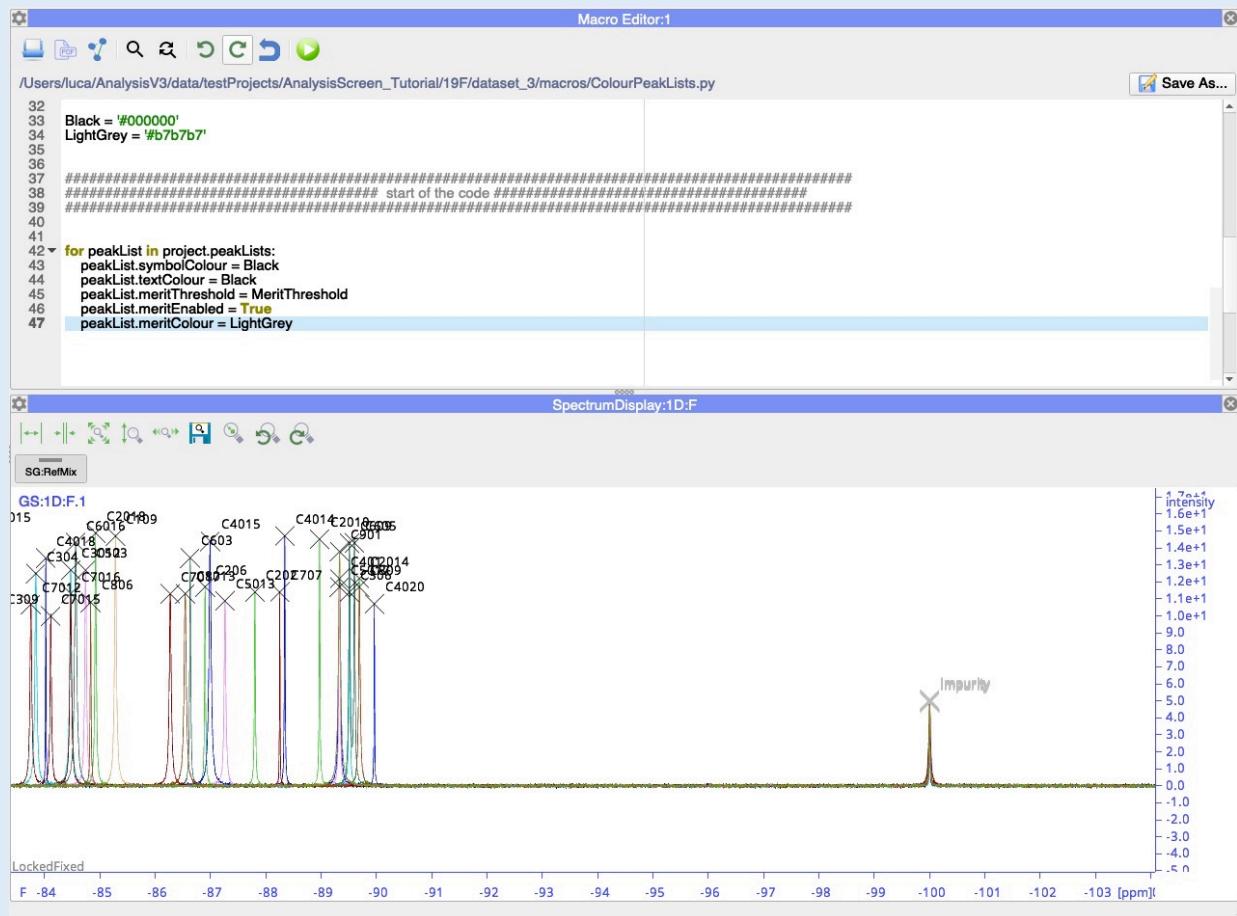
```
tag = 'Impurity'

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

**Warning** Copy&Paste code from PDF might lose the original indentation causing syntax errors

5

# Handle spurious signals



## 5D Change peak annotation colours

You can change the peak symbol/text colours so that real signal peaks (black) are graphically distinguishable from excluded (light grey) ones:

- Find the code in the Macro directory:  
.../19F/dataset\_3/macros/ColourPeakLists.py
  - Run from the macro editor

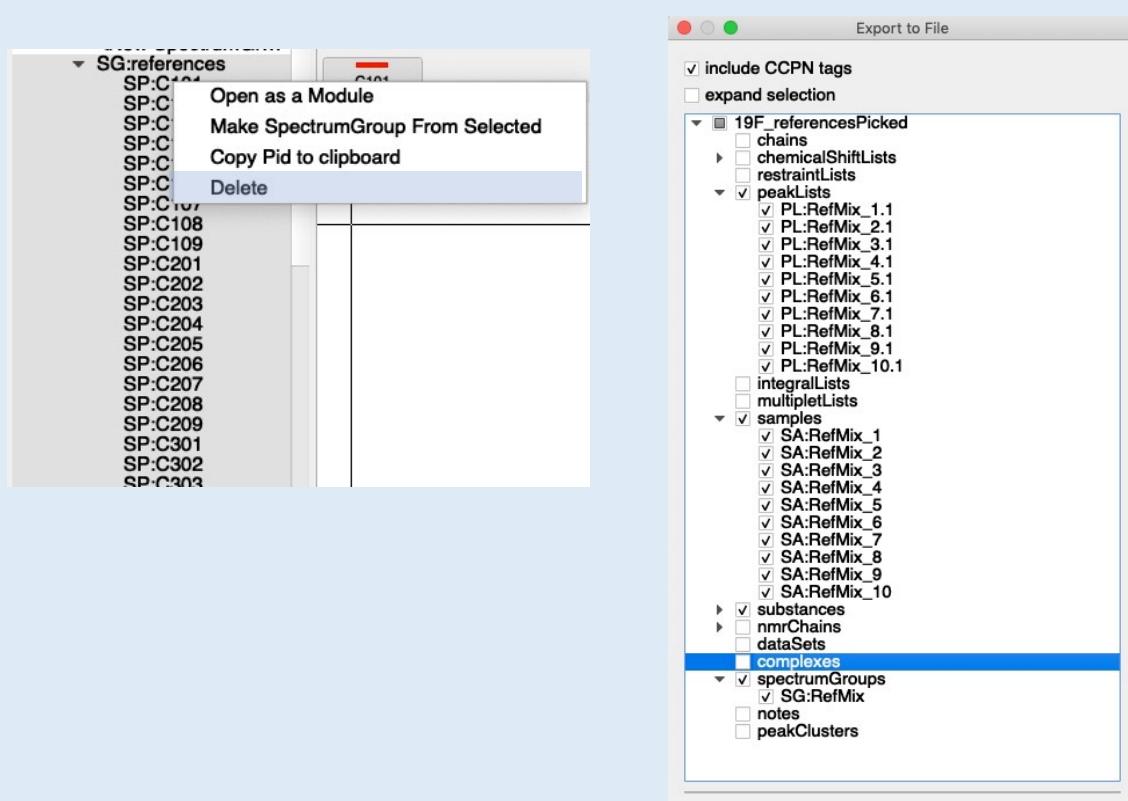
A code snippet is shown below

```
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
```

**Warning** Copy&Paste code from PDF might lose the original indentation causing syntax errors



## 5E Delete Reference Spectra

In the Sidebar:

- Expand the **SG:References** SpectrumGroup tree
- Select all spectra in the **SG:references** Spectrum Group, including the SpectrumGroup **SG:references** itself: **right-click → Delete**

Or in the Python Console, 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** (or use shortcut **EX**)
- In the NEF Dialog untick all first, and tick only the following:

*General:*

include CCPN Tags,

*Project tree:*

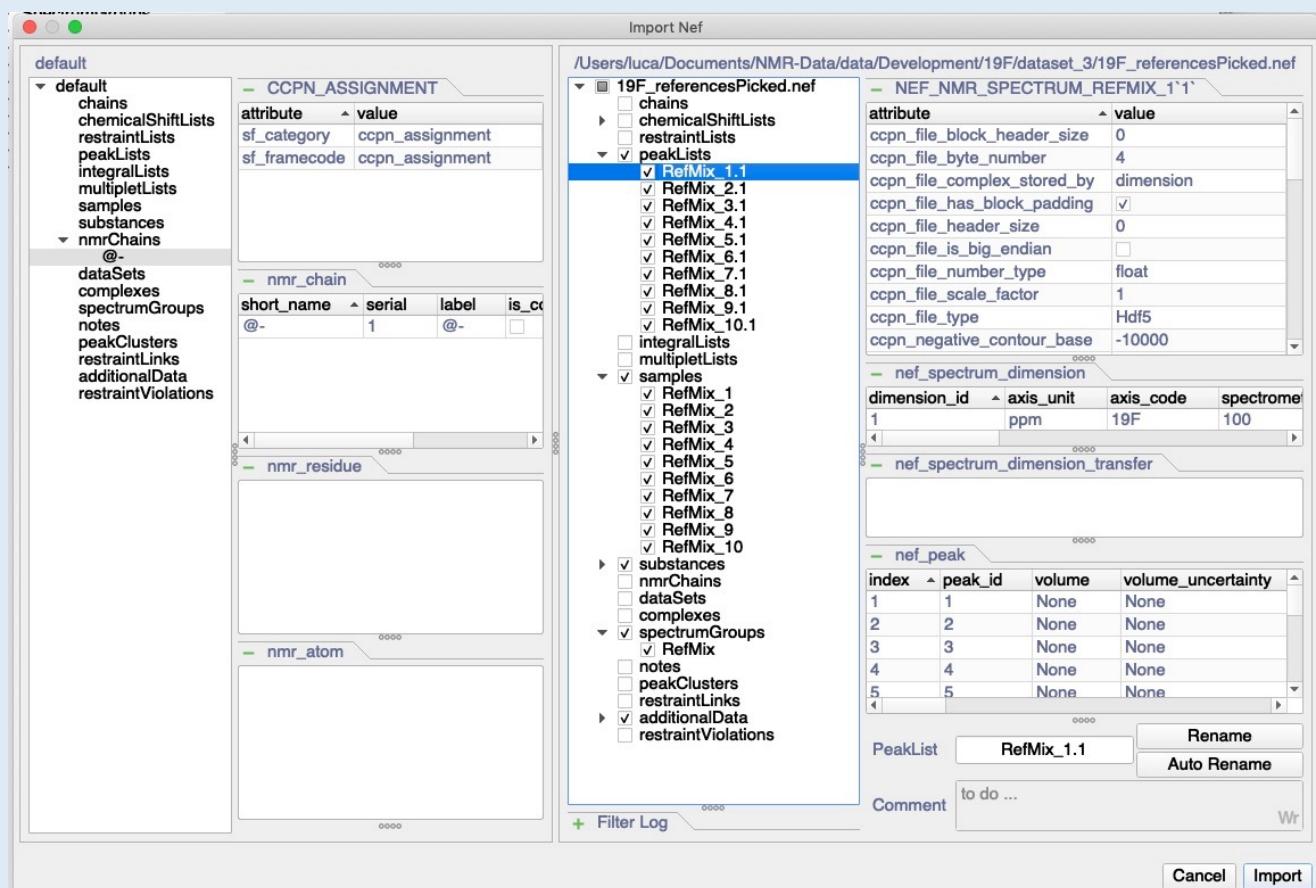
PeakLists

Samples

Substances

SpectrumGroups

- Save to your local disk



## 6A Import From NEF

- Start a new project
- Main Menu → Project → Import → NEF File (or shortcut IN)
- Select the NEF file you have just created or the file provided in the tutorial data directory.
- In the NEF Dialog tick the following on the right hand side:

Project tree:

- PeakLists
- Samples
- Substances
- SpectrumGroups
- AdditionalData

- Resolve any name clashes (highlighted in red) if relevant.
- Click Import

Go to Sidebar and check if spectra have two listed peak lists.

If there are empty peak lists (opening on a new module will appear as an empty table), run the following script in the PythonConsole:

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

6

# Screen From NEF



Sometimes your latest dataset might be slightly offset compared to the Reference Mixtures in the NEF file. In that case, 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 **dataset\_3** tutorial data folder:  
.../19F/dataset\_3/Data\_Time\_x/lookup\_19F\_TimeX.xlsx

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 module (**Menu → Screen → Pipeline** or shortcut **PI**)
  - In the sidebar, multiselect SpectrumGroups **SG:Control**, **SG:Target**, **SG:RefMix**
  - drag and drop them into the pipeline **Input Data** box
  - In the list of **Pipes** search for and add to the pipeline area:

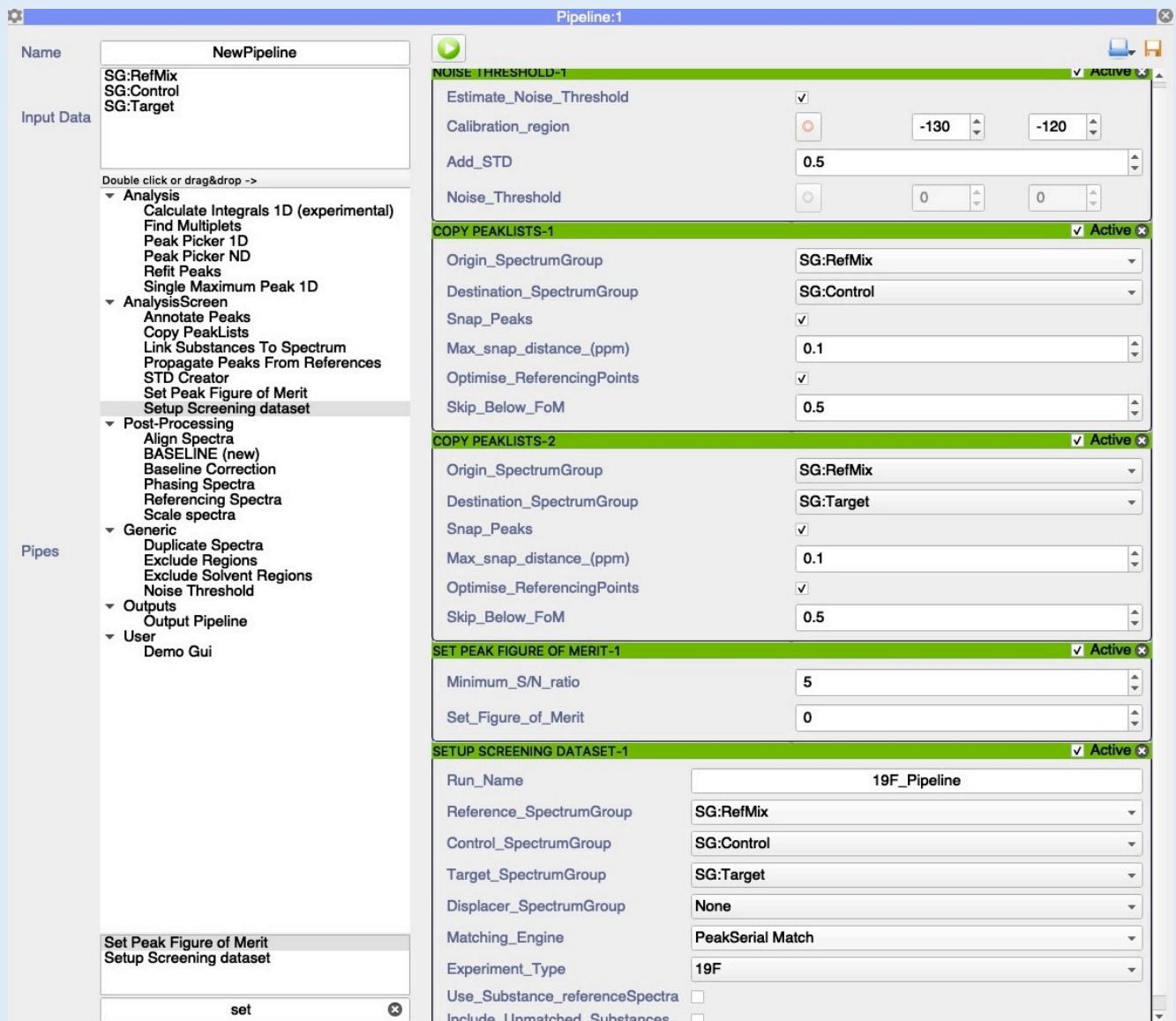
## 1. Noise Threshold

- tick **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
- Select **SG:Control** and **SG:Target** for their respective entries
- **Matching engine:** Peak Serial Match
- untick **Use\_Substance\_ReferenceSpectra**

- Run the pipeline.
- Inspect the results in the **Hit Analysis Module** as shown 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

When selecting Substance or Sample items in the main tables, you may notice how the Singleton Reference Spectra are replaced by the Reference Mixture spectra. The peak annotations 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)