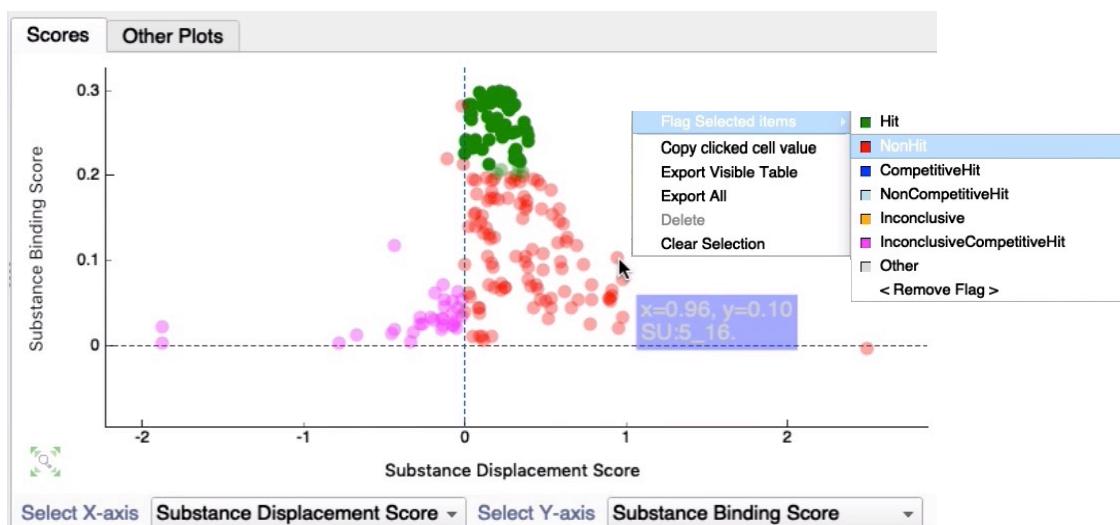


AnalysisScreen Hit Analysis Tutorial



Introduction

This tutorial will show the usage of the Hit-Analysis module in CcpNmr AnalysisScreen Version 3.1.

We strongly recommend that you read the Introduction to familiarise yourself with our terminology and workflows. Not all users may wish or need to work through all sections and using the projects provided it is easy to skip between different sections.

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 **CcpnTutorialDataScreening Jan2023** directory of the CcpNmr V3 Screening data which you can download from the CCPN website (<https://ccpn.ac.uk/support/tutorials>).

Contents

Introduction

Part I: Data Processing using Reference Singletons

- Import from Excel Files *(Section 1)*
- Manual Peak Matching from Reference Singletons *(Section 2)*
- Automated Peak Matching from Reference Singletons *(Section 3)*

Part II: Hit Analysis

- Binding Scores and Calculation Engines *(Section 4)*
- Plots and Filtering *(Section 5)*
- Exporting Data *(Section 6)*

Part III: Working with Reference Mixtures

- Create/update a NEF file *(Section 7)*
- Import from Excel and NEF Files *(Section 8)*
- Automated Peak Matching from Reference Mixtures *(Section 9)*

Part IV: Compare Screens

- Using data with multiple CPMG delays *(Section 10)*

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
- NMRbox users by using the terminal command *analaysisscreen*

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.

Spectrum Display

A Spectrum Display can contain multiple overlaid spectra which share the same axes. To show/hide a single spectrum, click on its spectrum 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

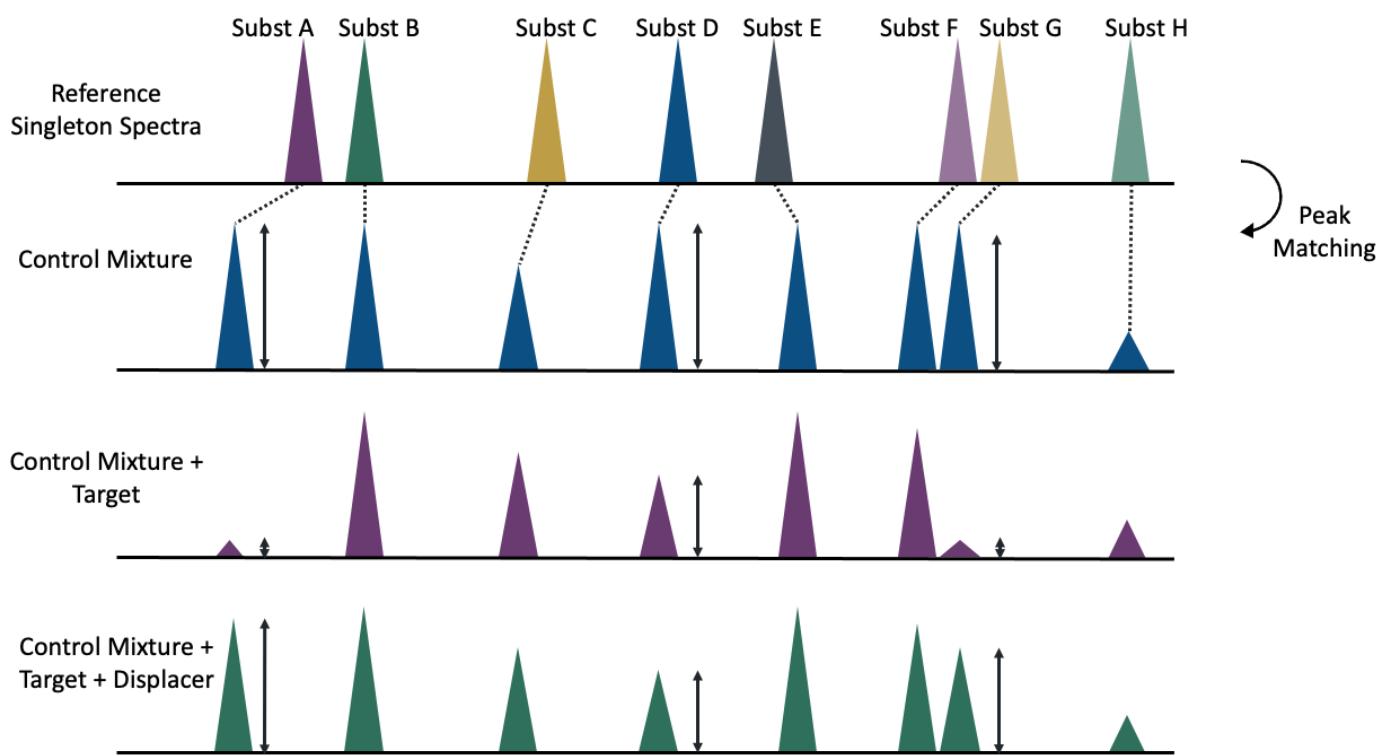
OR

Main Menu → Help → Show Shortcuts

Introduction

Schematic representation of the Screen analysis workflows

In this tutorial you will analyse several ^{19}F datasets but similar steps and tools in AnalysisScreen can also be applied to other experiment type analyses, such as ^1H relaxation-edited, WaterLOGSY and STD experiments.



We assume that you start with a set of individual spectra for each Substance (or compound) in your library. We refer to these as **Reference Singleton Spectra**.

Each screening experiment will start with the recording of a set of **Control Mixture Spectra**, a control spectrum of only the **Substance Mixtures**. Typically some of the peaks in the Control Mixtures will move relative to the Reference Singleton Spectra. The first task, therefore, is to match the peaks from the Reference Singleton Spectra to those in the Control spectra so that we know which peak in the Control Mixtures belongs to which Substance. This **Peak Matching** step can be done manually (Section 3) or automatically (Section 4). If done automatically, a **Peak Matching Score** will help you find matches that might need to be checked/corrected manually before you proceed.

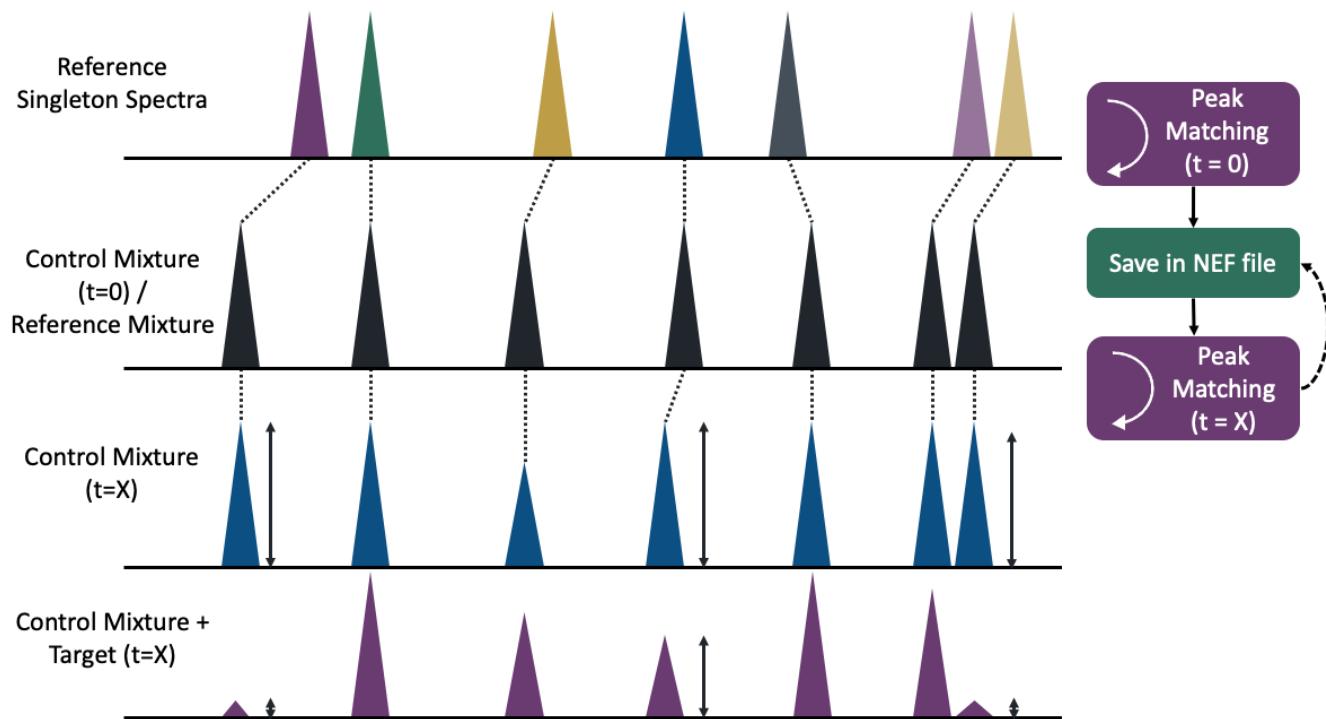
In addition to your Control Mixture Spectra you will have spectra where you have added a **Target** molecule and perhaps also a **Displacer** (also referred to as a **Competitor**). These spectra are in fact also included in the Peak Matching step, so that you end up creating a **Screening Dataset** in which the peaks are matched across all spectra and to a particular Substance.

Now you can use the **Hit Analysis module** to calculate the **Peak Binding Score** from the Control and Target spectra (exactly how is experiment type dependent, Section 6). Then inspect, classify and flag your Hits (Section 7) and export them (Section 8). Note that Binding Scores are also provided per Substance and per Sample.

Introduction

Recurring Screen analysis workflow

If you are repeating your screens with the same library of Substances and different Targets, then you can make the automated Peak Matching step faster, more accurate and more reliable by using previous Control Mixtures as **Reference Mixtures**.



Typically, there will be more differences between the Reference Singleton Spectra and the first Control Mixture spectrum you record, than between the Control Mixtures recorded at different time points.

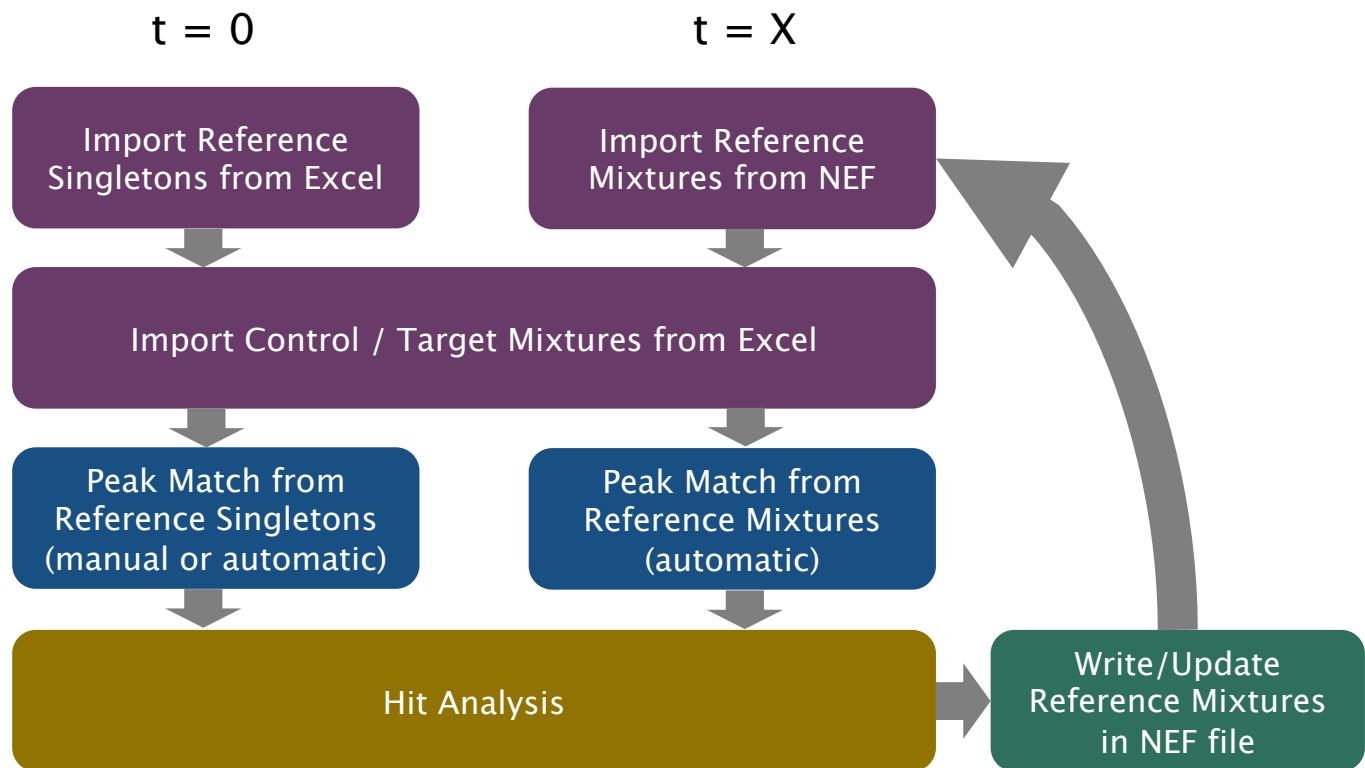
Therefore, we recommend that after your first Peak Matching (at time = 0), you save your Control Mixture data in an NMR Exchange Format (NEF) file and use this as a set of **Reference Mixtures** for your future screens. If your Control Mixtures continue to change over time, you can save each Control Mixture as the Reference Mixture for the next screen or you could create a new Reference Mixture NEF file once every 6 months or year, depending on how stable your Substance Mixtures are.

See the next page for an outline of the resulting workflow.

Introduction

... continued

The following diagram shows the steps conducted during the first analysis of your screening data at time = 0 and then of later analyses if you use previous Control spectra as Reference Mixtures, saved in an NMR Exchange Format (NEF) file.



Introduction

CcpNmr AnalysisScreen Nomenclature

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 “displacer”.

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) peak in a Screening Dataset.

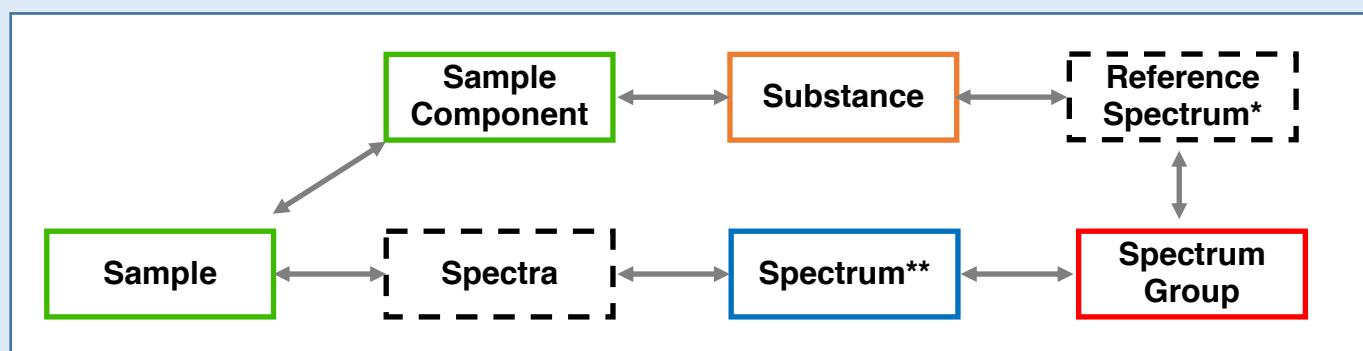


Figure showing CcpNmr AnalysisScreen object links

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

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

Import Data from Excel

In Section 1 we show you how to create an Excel file for data import. In practice, if you are importing the data for a library containing 100s or even 1000s of compounds, doing everything manually with copy/paste may not be viable. We are very happy to help create a script for you to import the data from your current system into your first the AnalysisScreen compatible Excel file, simply contact us at support@ccpn.ac.uk for help.

The program can read .xls or .xlxs files with multiple sheets that include the words **Sample** or **Substance** in the sheet name.



You can create files that contain either the Substance or Sample page or both.

1A Overview

Open a new Excel file or find a template in:

CcpnTutorialDataScreeningJan2023/LookupTemplate.xls

Mandatory Sheet names:

Title must start with “Substance” or “Sample”

Mandatory columns:

Substance sheet:

SubstanceName

Sample sheet:

SampleName

| | A | B | C | D | E | F | G | H |
|---|---------------|--------------|-------------------|----------------|-------------------|---------|---|--|
| 1 | substanceName | spectrumPath | spectrumGroupName | experimentType | spectrumHexColour | comment | | smiles |
| 2 | 0 | C101 | C101 | References | 19F | #BA55D3 | | OCCOC[C@@]1(CO1)N=[N]=N |
| 3 | 1 | C102 | C102 | References | 19F | #BA55D3 | | COCCN1C[C@@H](CC1=O)[NH2]C[NH3] |
| 4 | 2 | C103 | C103 | References | 19F | #BA55D3 | | N=[N]=N[C@]1(C#N)CCO[C@@H]1OC(=O)C |
| 5 | 3 | C104 | C104 | References | 19F | #BA55D3 | | N=[N]=NC(=O)[C@@H](n1nnn(c1=S)C(=O)[C@@H]1 |
| 6 | 4 | C105 | C105 | References | 19F | #BA55D3 | | N=[N]=N[C@@H](C@H)(n1nnn(c1)c1cccc1N1C(=O) |

1B Create the Substance Sheet

The first sheet, **Substance**, can contain metadata associated with small molecules whose spectra, for example, have been used as references in a screen.

- Place the Lookup file template from the **ScreenTutorial** directory into the directory containing your spectra (**CcpnTutorialDataScreening Jan2023/19F/dataset1/data** if using the tutorial data).
- Open the template and fill in the **substanceName** column. This is the only mandatory column to fill in.
- Copy and paste the following:
C101
C102
C103
C104
C105

| A | B | C | D | E | F |
|---|---------------|--------------|-------------------|----------------|-------------------|
| 1 | substanceName | spectrumPath | spectrumGroupName | experimentType | spectrumHexColour |
| 0 | C101 | C101 | References | 19F | #BA55D3 |
| 1 | C102 | C102 | References | 19F | #BA55D3 |
| 2 | C103 | C103 | References | 19F | #BA55D3 |
| 3 | C104 | C104 | References | 19F | #BA55D3 |
| 4 | C105 | C105 | References | 19F | #BA55D3 |

1c Add reference spectrum information

To include the Substance reference spectra, you need to insert the **spectrumPath** (AnalysisScreen will recognise any spectrum format and you do not need to include any file extensions such as .ucsf, ndf5, .ft etc.).

You have three options:

1. If all the spectra files are located in the same directory as the lookup file, insert only the file names as above.
2. If the spectra are located in a subdirectory, insert the directory name first followed by a slash and the filename (the relative path starting from the Excel file), e.g. **references/C101**
3. If the spectra files are located in a completely different location, insert the full path, e.g. **/Users/username/Desktop/data3/MySpectra/C101**

For Bruker files, you can insert the path to the **1r** file, e.g.:

~CcpnTutorialDataScreeningJan2023/19F/dataset_1/data/pdata/1/1r

For clarity, we recommend keeping all the files in the same directory together with the Excel lookup file.

- Insert the **spectrumGroupName**; e.g. **References**. This will create a **Spectrum Group** with that name and place the spectra into it.
- Insert the **experimentType**. For these 1-dimensional 19F spectra, simply type **19F** into the cell.
- Insert the **spectrumHexColour** of your choice. We will use **#BA55D3** (mediumorchid) for reference spectra throughout this tutorial.

| | A | B | C | D | E | F | G | H | I | J |
|---|---------------|--------------|-------------------|----------------|-------------------|--|--------|---------|----------|---------------|
| 1 | substanceName | spectrumPath | spectrumGroupName | experimentType | spectrumHexColour | comment | smiles | | synonyms | molecularMass |
| 2 | 0 C101 | C101 | References | 19F | #BA55D3 | OCCOC[C@@]1(CO1)N=[N]=N | () | 160.151 | | |
| 3 | 1 C102 | C102 | References | 19F | #BA55D3 | COCCN1C[C@@H](CC1=O)[NH2]C[NH3] | () | 189.255 | | |
| 4 | 2 C103 | C103 | References | 19F | #BA55D3 | N=[N]=N[C@]1(C#N)CCO[C@@H]1OC(=O)C | () | 197.171 | | |
| 5 | 3 C104 | C104 | References | 19F | #BA55D3 | N=[N]=NC(=O)[C@@H](n1nnn(c1=S)C(=O)[C@@H]1C[C@@H]1Br)C | () | 347.172 | | |
| 6 | 4 C105 | C105 | References | 19F | #BA55D3 | N=[N]=N[C@H]([C@H](n1nnn(c1)c1cccc1N1C(=O)C=CC1=O)C)C | () | 338.344 | | |
| 7 | | | | | | | | | | |
| 8 | | | | | | | | | | |
| 9 | | | | | | | | | | |

1D Add Substance metadata

- The **comment** column will store any textual information about the substance. Avoid using letters with accents etc. (e.g. é, ä, œ, ſ, ß, ç, ...) as these may cause problems when trying to save your CCPN project.
- If you enter the **SMILES** for your substances, the program will automatically generate the structures inside the software. For the tutorial, copy and paste these SMILES:

OCCOC[C@@]1(CO1)N=[N]=N
 COCCN1C[C@@H](CC1=O)[NH2]C[NH3]
 N=[N]=N[C@]1(C#N)CCO[C@@H]1OC(=O)C
 N=[N]=NC(=O)[C@@H](n1nnn(c1=S)C(=O)[C@@H]1C[C@@H]1Br)C
 N=[N]=N[C@H]([C@H](n1nnn(c1)c1cccc1N1C(=O)C=CC1=O)C)C
- In the **synonyms** column you can insert the chemical name of the substance and again select **Match Destination** formatting
- All the following columns contain the substance chemical properties. Fill them in if you want to display them within the software.
- Save the file.

A fully completed lookup file is provided at

[CcpnTutorialDataScreeningJan2023/19F/dataset1/data/lookup_19F_data_set1.xls](#).

| | A | B | C | D | E | F | G | H | I | J |
|---|------------|-------------------|--------------|--------------|----------------|-------------------|---------|--------------------------|-----|-----|
| 1 | sampleName | spectrumGroupName | spectrumPath | spectrumName | experimentType | spectrumHexColour | comment | sampleComponents | pH | ion |
| 2 | 0 | Control_1 | Control | Control_1 | Control_1 | #1E90FF | | C101,C102,C103,C104,C105 | 9.5 | |
| 3 | 1 | Target_1 | Target | Target_1 | Target_1 | #FF8C00 | | C101,C102,C103,C104,C105 | 9.5 | |
| 4 | 2 | Displacer_1 | Displacer | Displacer_1 | Displacer_1 | #32CD32 | | C101,C102,C103,C104,C105 | 9 | |
| 5 | | | | | | | | | | |

1E Create the Samples Sheet

The next sheet in the template is **Samples**. This can contain metadata associated with particular samples, e.g. in a screening trial the sample could contain lots of spectra recorded with different experimental conditions. The only mandatory column is the **sampleName** column.

- Insert the **sampleName** in the first column, e.g. **Control_1**

The next four columns are specific to the spectra recorded for this sample:

- Insert the **spectrumGroupName**, e.g. **Control**, if you want the spectrum to be included in a Spectrum Group
- Insert the **spectrumPath**, e.g. **Control_1** (see the section **1C** for how to insert the spectrum path)
- Insert the **spectrumName** if you don't want to use the program's default spectrum names (not very user friendly for Bruker data, for example)
- Insert the spectrum **experimentType**, e.g. **19F**
- Fill in the **sampleComponents** column. Insert the names of the components (**Substances**) that are present in the sample. In the case of a mixture containing components 1 to 5, insert them as a comma-separated list without spaces:

C101,C102,C103,C104,C105

- The other columns record a sample's chemical properties and other information. Fill them in if you want to display them within the software.

To add extra spectra for the same sample, repeat points 1 to 3 as shown in the figure. There is no need to duplicate the sample properties (yellow columns) as long as the sample name is the same. If you add the same information twice, only the first entry will be used.

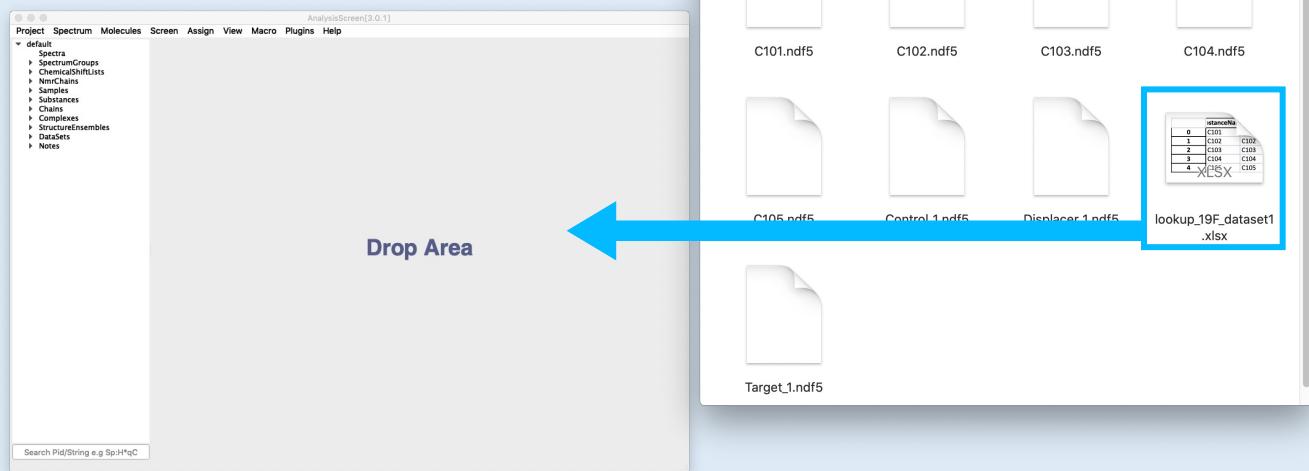
To add an additional sample, simply fill in further rows as shown above.

Note that different samples can also be listed on different sheets, as long as each sheet names starts with the word **Sample**.

Import Data from Excel

Don't forget to contact us at support@ccpn.ac.uk

if you want help setting up your first
AnalysisScreen compatible Excel file.



1F Import Excel Lookup File into AnalysisScreen

- Drag and Drop either your newly created Excel file or the **lookup_19F_dataset1.xls** file located in the **CcpnTutorialDataScreening Jan2023/19F/dataset1/data/** directory of the tutorial data from your file browser into the **sidebar** or **drop area** of AnalysisScreen.

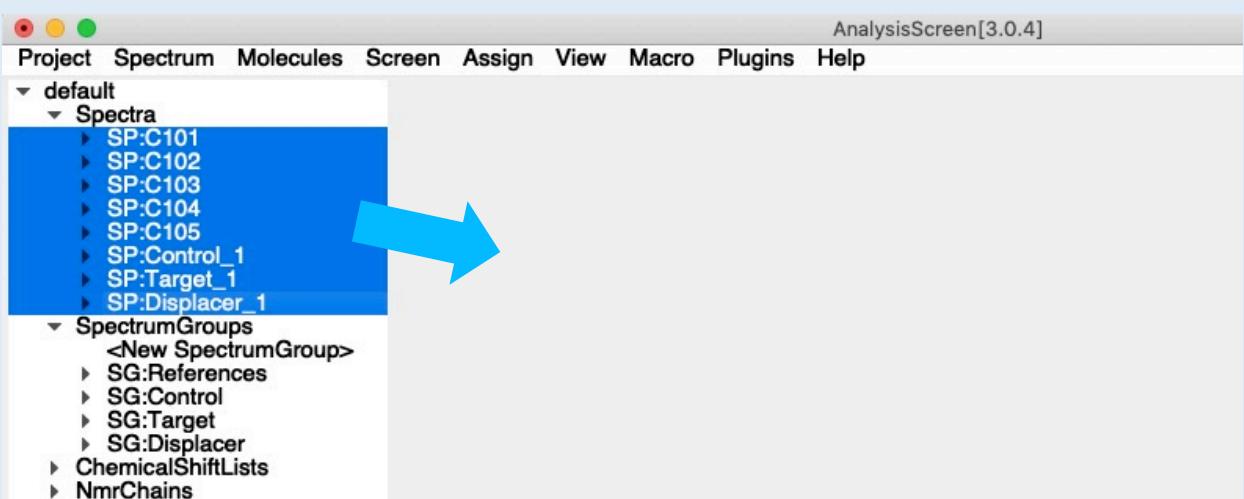
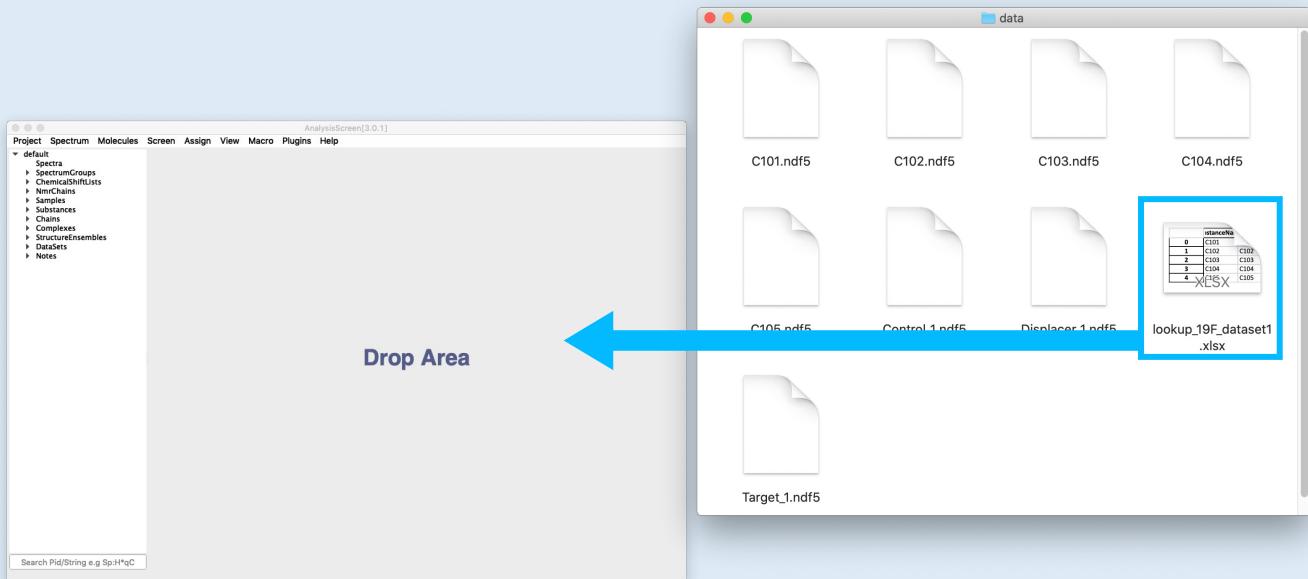
You will now be able to see all the imported data in the sidebar:

- ▼ default
 - ▼ Spectra
 - SP:C101
 - SP:C102
 - SP:C103
 - SP:C104
 - SP:C105
 - SP:Control_1
 - SP:Target_1
 - SP:Displacer_1
 - ▼ SpectrumGroups
 - <New SpectrumGroup>
 - SG:References
 - SG:Control
 - SG:Target
 - SG:Displacer
 - ChemicalShiftLists
 - NmrChains
- ▼ Samples
 - <New Sample>
 - SA:Control_1
 - SA:Target_1
 - SA:Displacer_1
- ▼ Substances
 - <New Substance>
 - SU:C101.
 - SU:C102.
 - SU:C103.
 - SU:C104.
 - SU:C105.
 - Chains
 - Complexes
 - StructureEnsembles
 - StructureData
 - DataTables
 - Collections
 - Notes

Please note that you cannot drop the same lookup file containing the same values into the same project twice. This is because the project cannot create new objects with pre-existing names. When dropping the same file onto a project twice, only the first entries will be used.

You can now proceed to match your peaks manually in **Section 2** or automatically in **Section 3**.

2 Manual Peak Matching from Reference Singletons



2A Add your data to the program with an Excel file

If you haven't already added the data from **dataset_1** to your project in **Section 1**, then:

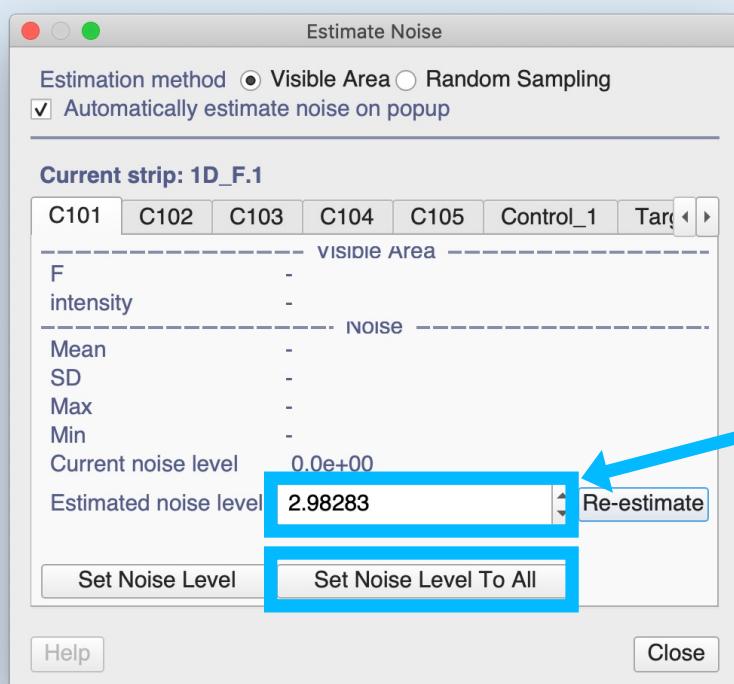
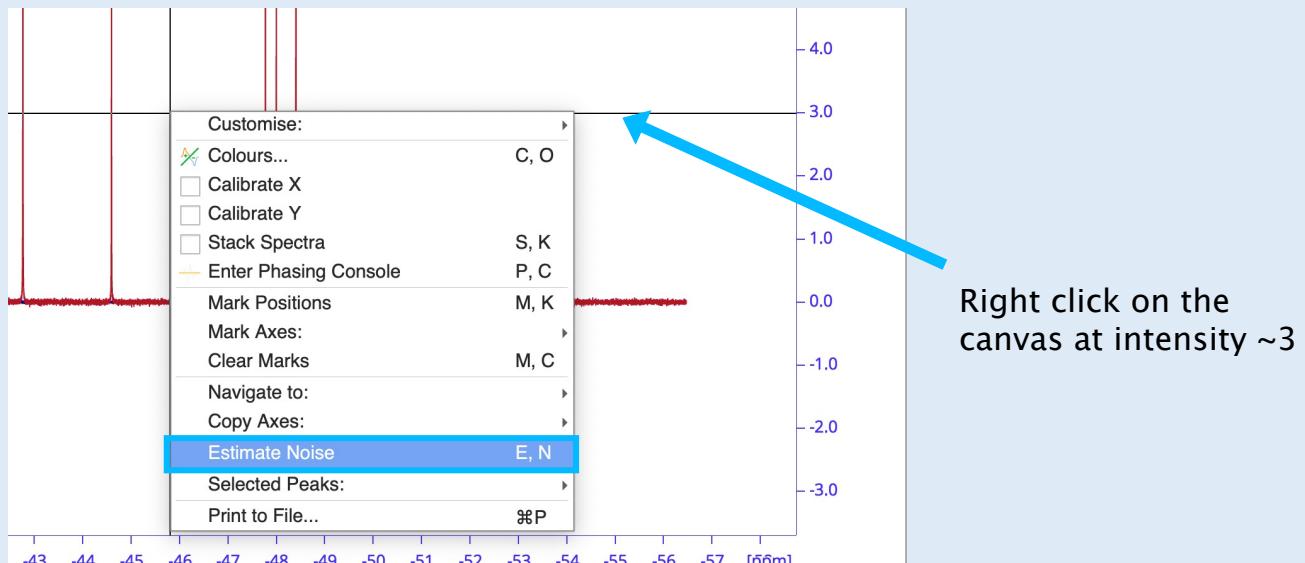
- Drag and drop the Excel file **lookup_19F_dataset1.xlsx** in the **CcpnTutorialDataScreeningJan2023/19F/dataset_1/data/** directory into the program.

By loading a correctly formatted Excel file, all the necessary links are automatically established ensuring optimal functioning of the screening tools. See our **HowTo: Sidebar Objects** and the **HowTo: Import Data From Excel** tutorials for more information.

2B Open all spectra

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

Manual Peak Matching from Reference Singlets



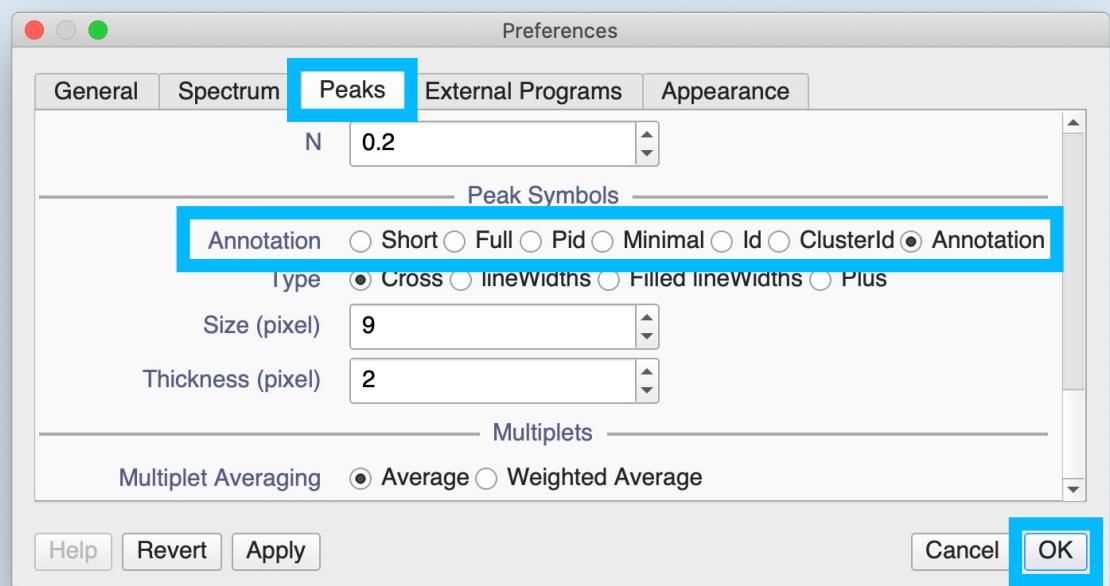
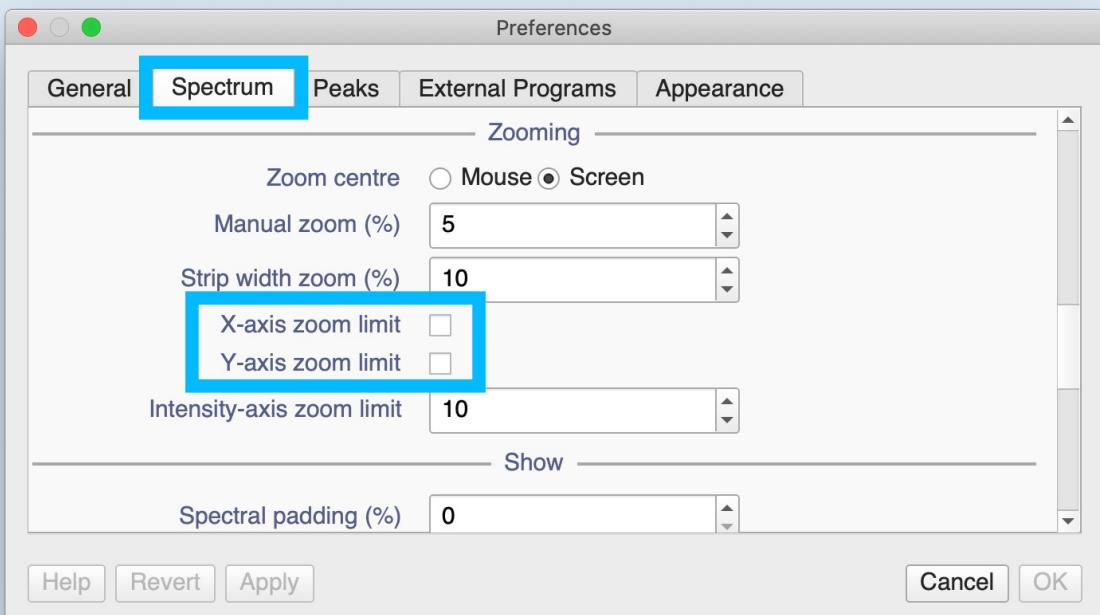
When picking 1D peaks, only peaks above this intensity level will be picked.

2c 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 on the Intensity axis.
- right click -> **Estimate Noise**
- in the pop-up click on **Set Noise Level To All**

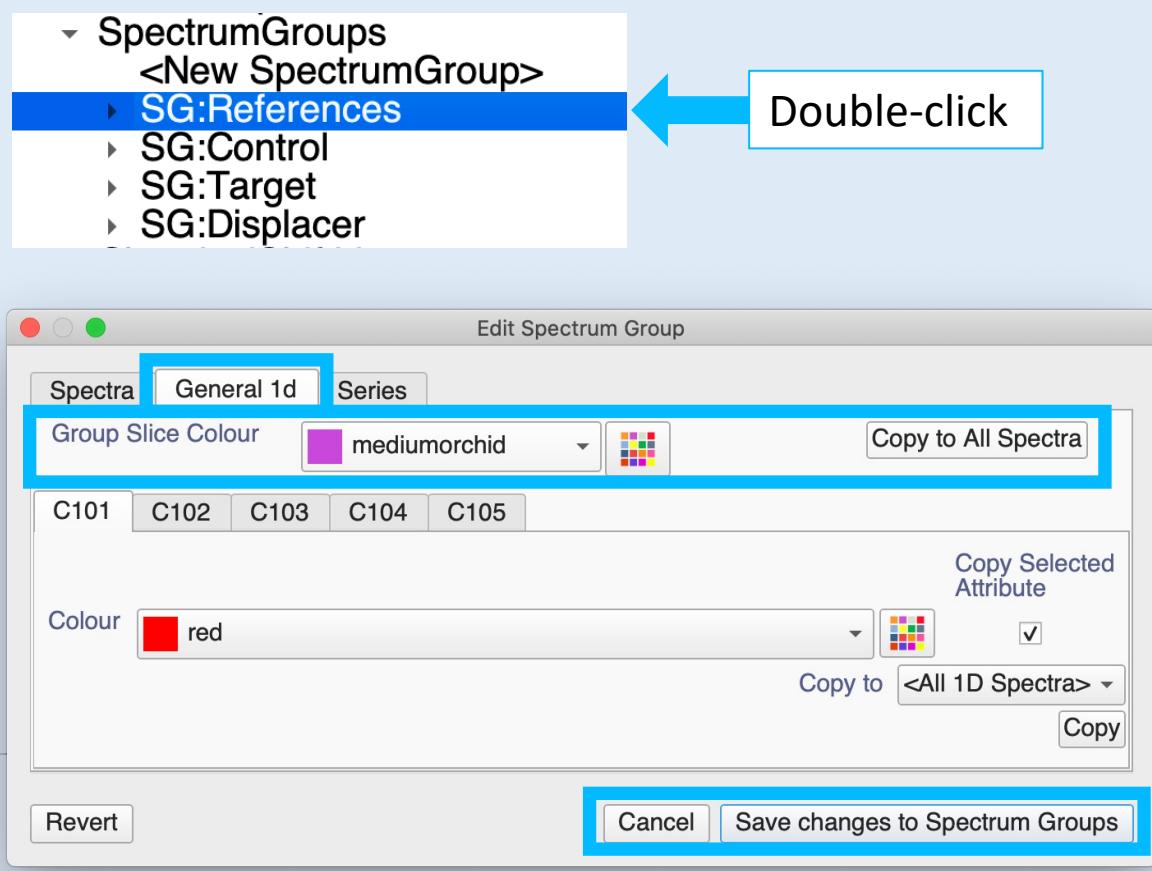
Manual Peak Matching from Reference Singletons



2D Adjust Preferences

- Go to Main Menu -> File -> Preferences
- In the **Spectrum** tab, untick **Apply zoom limit** both for X and Y axes
- In the **Peaks** tab, set **Label** to **Annotation**.

Manual Peak Matching from Reference Singletons



2E Colour spectra by Spectrum Group (optional)

If you didn't set the spectrum colours in your Excel file, you can do this now. It is possible to set all the spectra in one Spectrum Group to be displayed in the same colour in one go. It can be helpful to colour your Reference / Control / Target / Displacer spectra in the same colours in each project so that you can quickly recognise which group a spectrum belongs to.

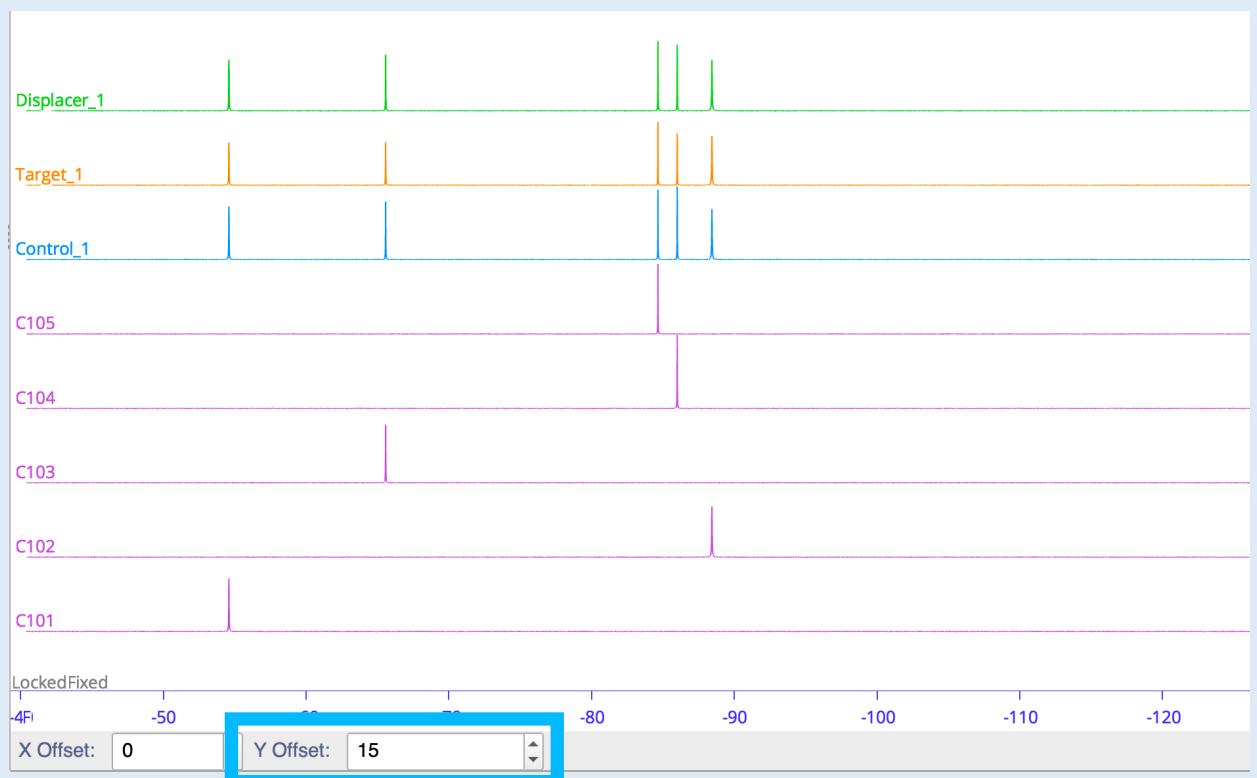
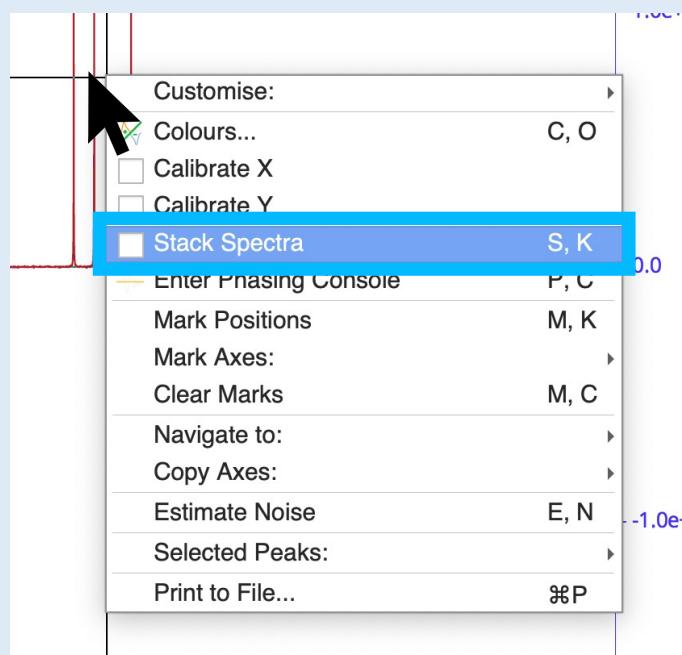
- Double-click on the **SG:References** Spectrum Group in the sidebar and go to the **General 1d** tab.
- Select a **Group Slice Colour** and click on **Copy to All Spectra**.
- Click on **Cancel** or **Save Changes to Spectrum Groups** to close the pop-up.

Throughout this tutorial we will use the following colour coding:

| | |
|------------|------------------------------|
| References | pink (mediumorchid, #BA55D3) |
| Control | blue (dodgerblue, #1E90FF) |
| Target | orange (darkorange, #FF8C00) |
| Displacer | green (limegreen, #32CD32) |

Change the colouring of your spectrum groups to reflect this colouring if you like.

Manual Peak Matching from Reference Singletions

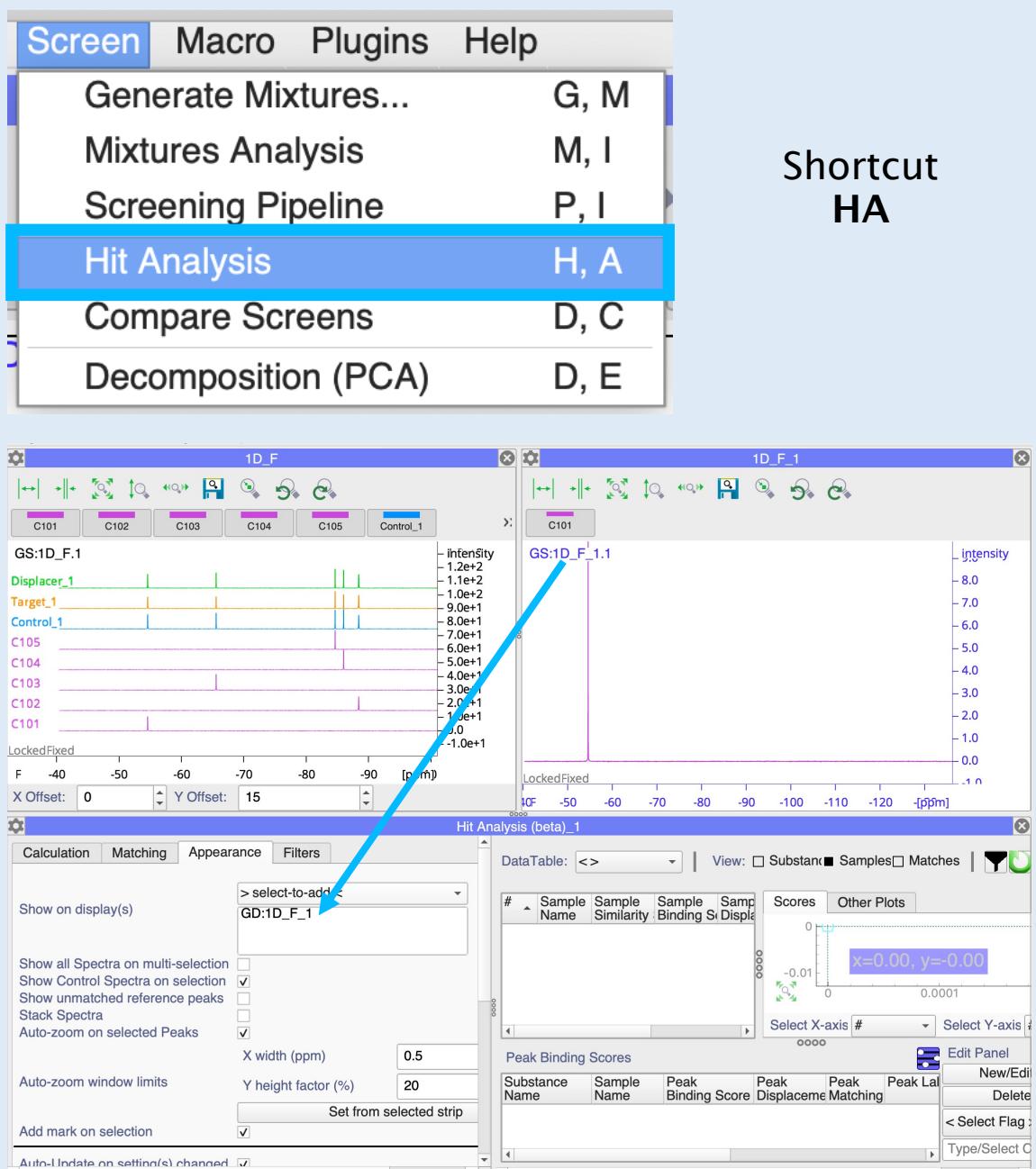


2F Stack Spectra

In the Spectrum Display :

- right click -> **Stack Spectra** (shortcut **SK**)
- **Y Offset: 15**
- 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

Manual Peak Matching from Reference Singletons



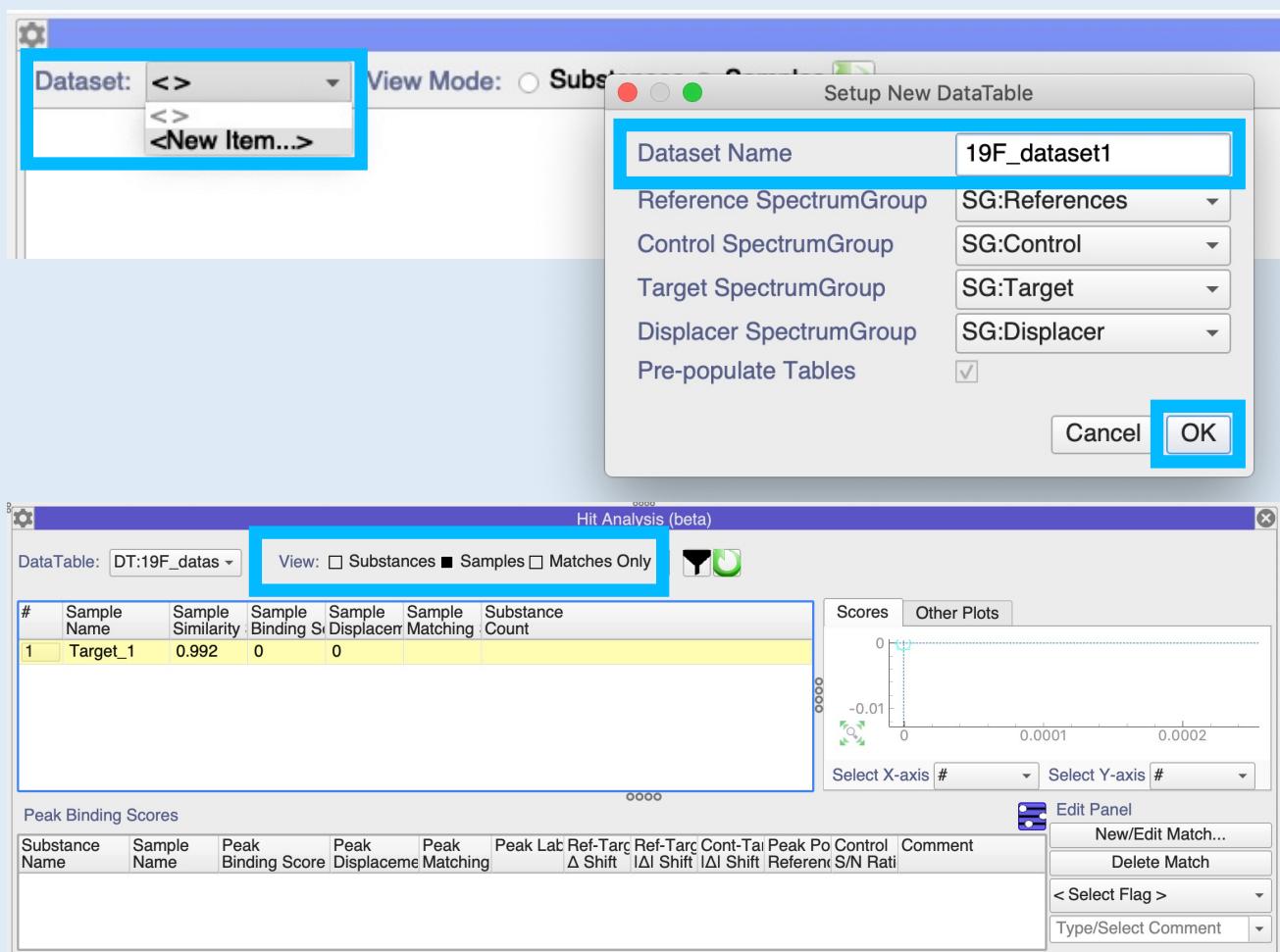
2G Open the Hit Analysis Module

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

2H 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.

Manual Peak Matching from Reference Singletons



The **Hit Analysis** module contains three view modes: by **Substance**, by **Sample** and by **Matches Only** which determine the behaviour of the two main tables.

The top table contains a list of the substances or samples for the **Substances** and **Samples** views, respectively. The lower table lists all the peak matches associated with the Substance/Sample selected. The **Matches Only** view shows only the lower table and includes all peak matches.

2 Create a new Screening Dataset

In the Hit Analysis module:

- Select **<New Item...>** in the **Dataset** pulldown,
- Change the name, e.g. to **19F_dataset1** or keep the default.
- Press **OK** to proceed and close the popup.

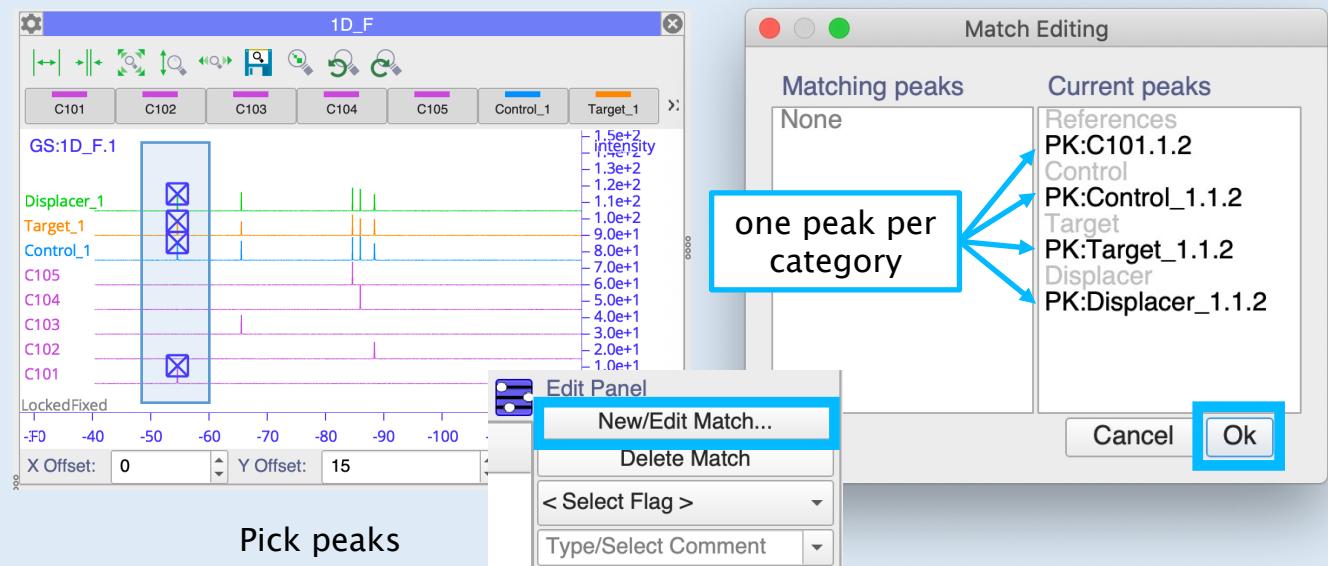
A new **DataTable** called **19F_dataset1** has now been created and can be seen in the sidebar. This is where all the data relating to this screening dataset will be stored.

▼ **DataTables**
<New DataTable>
DT:19F_dataset1

If you set the **View** to **Samples**, you can see that the **Target_1** sample is already given a **Sample Similarity** score of 0.992. This score reflects the similarity between the Target and Control spectra. A high value reflects few hits, a low value may indicate many hits or some other unexpected differences between the spectra such as shifting peaks.

Now we are ready to create **Peak Matches** and generate the remaining scores.

Manual Peak Matching from Reference Singletons



2K Pick and match peaks

Pick the peaks at ~ -54.59 ppm in the stacked Spectrum Display:

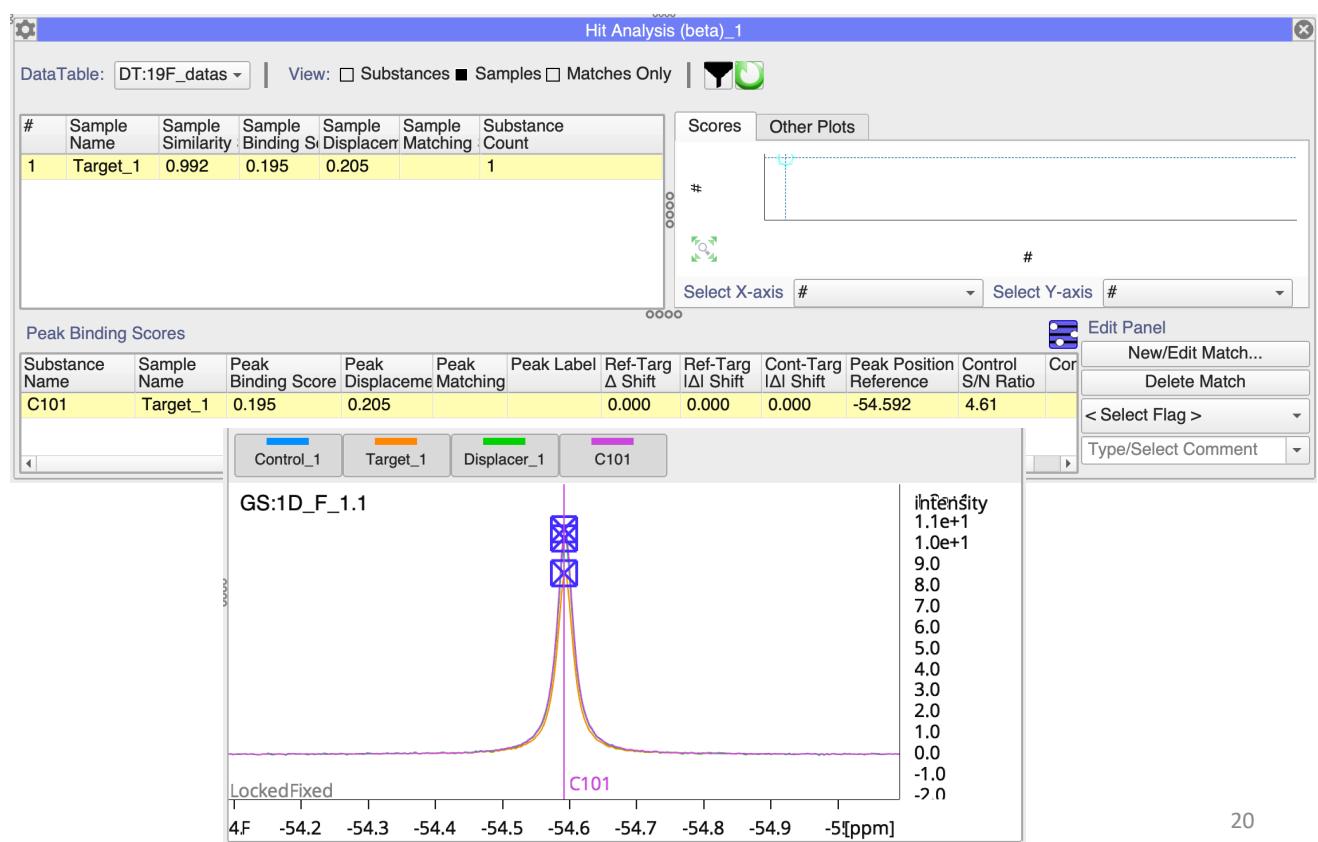
- Use **CTRL** (or **CMD** for Mac) + **SHIFT** + **Left-drag** to create a (blue) picking region to include all spectral signals around -54 ppm.

These signals correspond to the Substance **C101**.

In the Hit Module **Edit Panel** (if necessary open by clicking the icon):

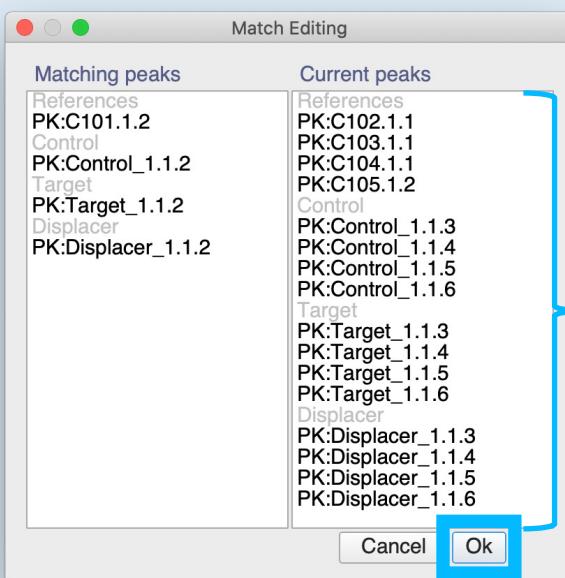
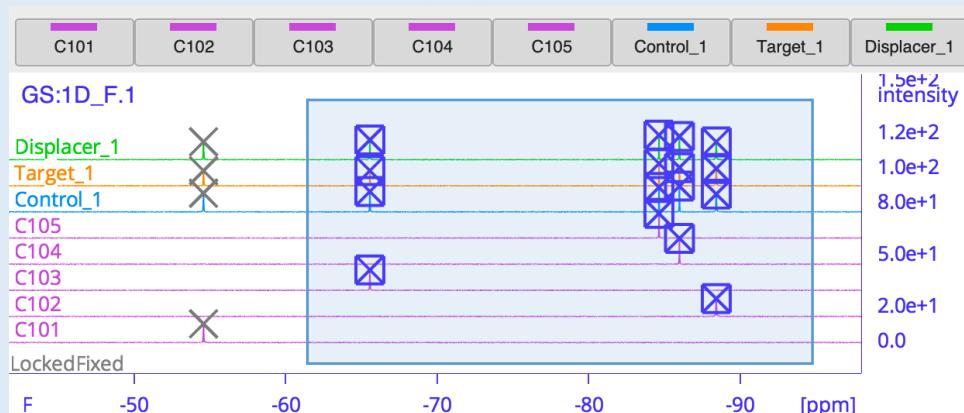
- Press **New/Edit Match** and there is only one peak per Category
- Press **Ok** to create a new **Peak Match**
- Press the orange **Update icon** to update the tables.

A new row for the Substance **C101** will appear in the **Peak Binding Scores** table which corresponds to that peak match. Select the row to navigate to the corresponding peaks in the spectrum display. There will be more information on the scores provided in **Section 4: Binding Scores and Calculation Engines**.



2 Manual Peak Matching from Reference Singletons

Multi-picking and semi-automatic peak matching



2L Semi-automatic peak 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:

- Press **New/Edit Match...** in the **Edit Panel**. Make sure there is an equal number of peaks in each category, e.g.: 4 Reference peaks, 4 Control Peaks etc... in any order.
- Press **Ok**.
- Now press the orange update icon again to update the table.

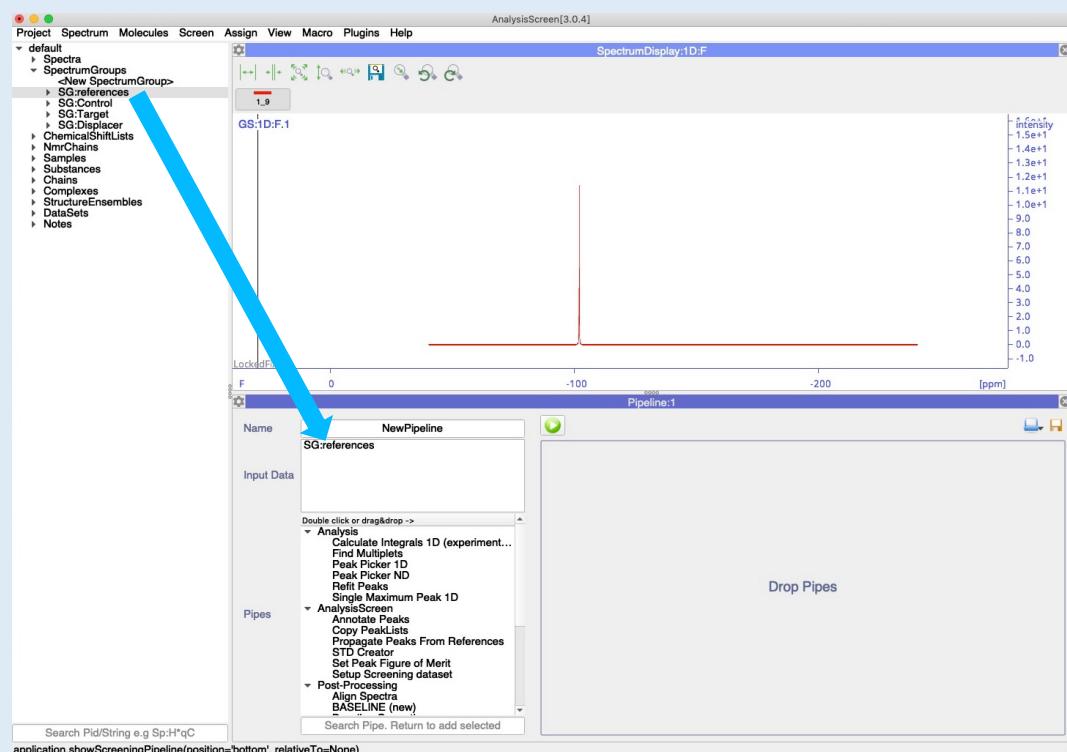
Additional Peak Matches have now been added to the Peak Binding Scores table.

| Peak Binding Scores | | | | | | | | | | |
|---------------------|-------------|--------------------|-------------------|---------------|------------|------------------|--------------------|-------------------|-------------------------|-------------------|
| Substance Name | Sample Name | Peak Binding Score | Peak Displacement | Peak Matching | Peak Label | Ref-Targ Δ Shift | Ref-Targ Δ Shift | Cont-Targ Δ Shift | Peak Position Reference | Control S/N Ratio |
| C101 | Target_1 | 0.195 | 0.205 | | | 0.000 | 0.000 | 0.000 | -54.592 | 4.61 |
| C102 | Target_1 | 0.030 | -0.000 | | | 0.000 | 0.000 | 0.000 | -88.432 | 4.39 |
| C103 | Target_1 | 0.260 | 0.134 | | | 0.000 | 0.000 | 0.000 | -65.574 | 5.04 |
| C104 | Target_1 | 0.005 | 0.000 | | | 0.000 | 0.000 | 0.000 | 0.000 | 0.00 |

- If you like, move between the **Substances**, **Samples** and **Matches Only** views to explore how the tables change.

More information on these tables and scores is provided in **Section 4: Binding Scores and Calculation Engines**.

Automatic Peak Matching from Reference Singletons



Automatic peak matching

In this section, a pipeline is built to create a screening dataset from several samples, each containing a mixture of substances. Control (mixtures only), Target (mixtures+target protein) and Displacer (mixtures + target + displacer) are used in addition to the Reference Singleton spectra.

A completed project is provided in the **dataset_2** folder.

3A Open a new dataset

- Open a new project with **Main Menu** → **File** → **New Project**
- Drag & drop the **lookup_19F_dataset2.xlsx** file from the **CcpnTutorialData ScreeningJan2023/19F/dataset_2/Data** folder into the sidebar or drop area.

See the **Section 1** for more information about setting up Excel files.

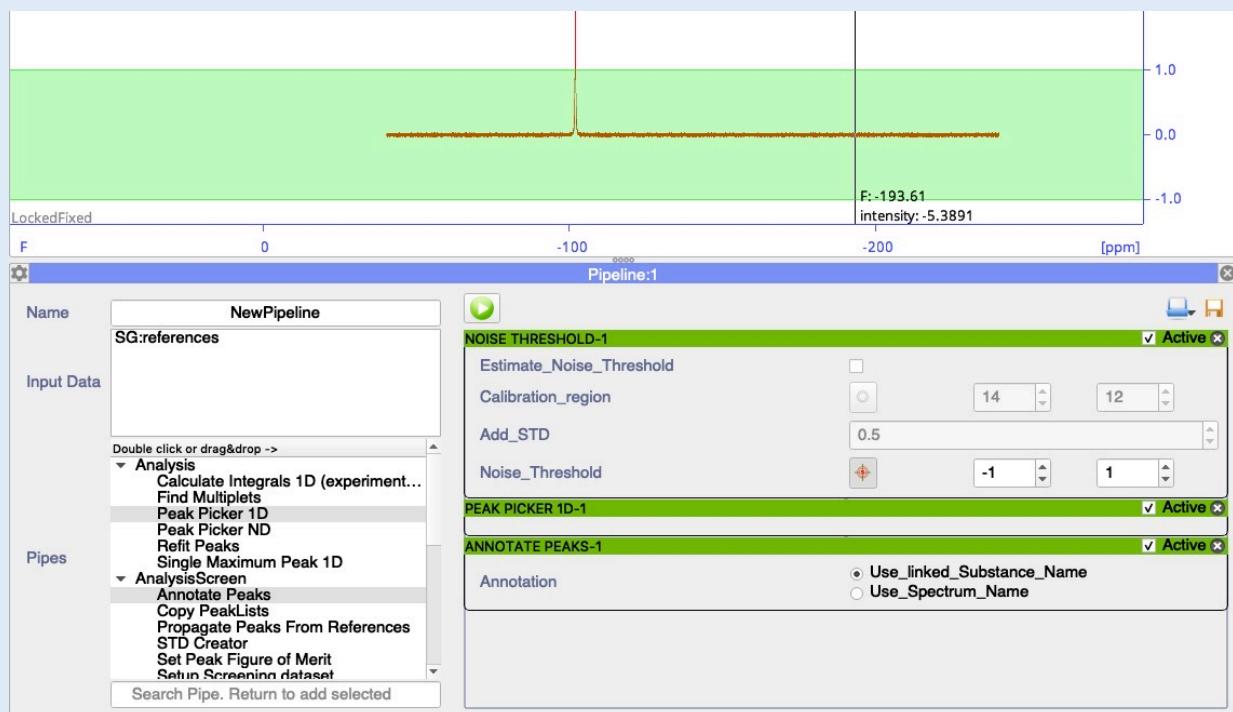
3B Open the first reference spectrum

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

3C Open the pipeline module

- Open the pipeline module from the main menu:
Menu → **Screen** → **Screening 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

Automatic Peak Matching from Reference Singletons



Multiple Pipes can be added and re-ordered by holding and dragging the green top bar. See **HowTo: Pipelines** tutorial for more information.

3D Picking and annotating peaks in the Reference Singleton spectra

- 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
- Enter the **Noise Threshold** values -1, 1 either by dragging the green lines that appear in the SpectrumDisplay or inserting the values in the entry boxes

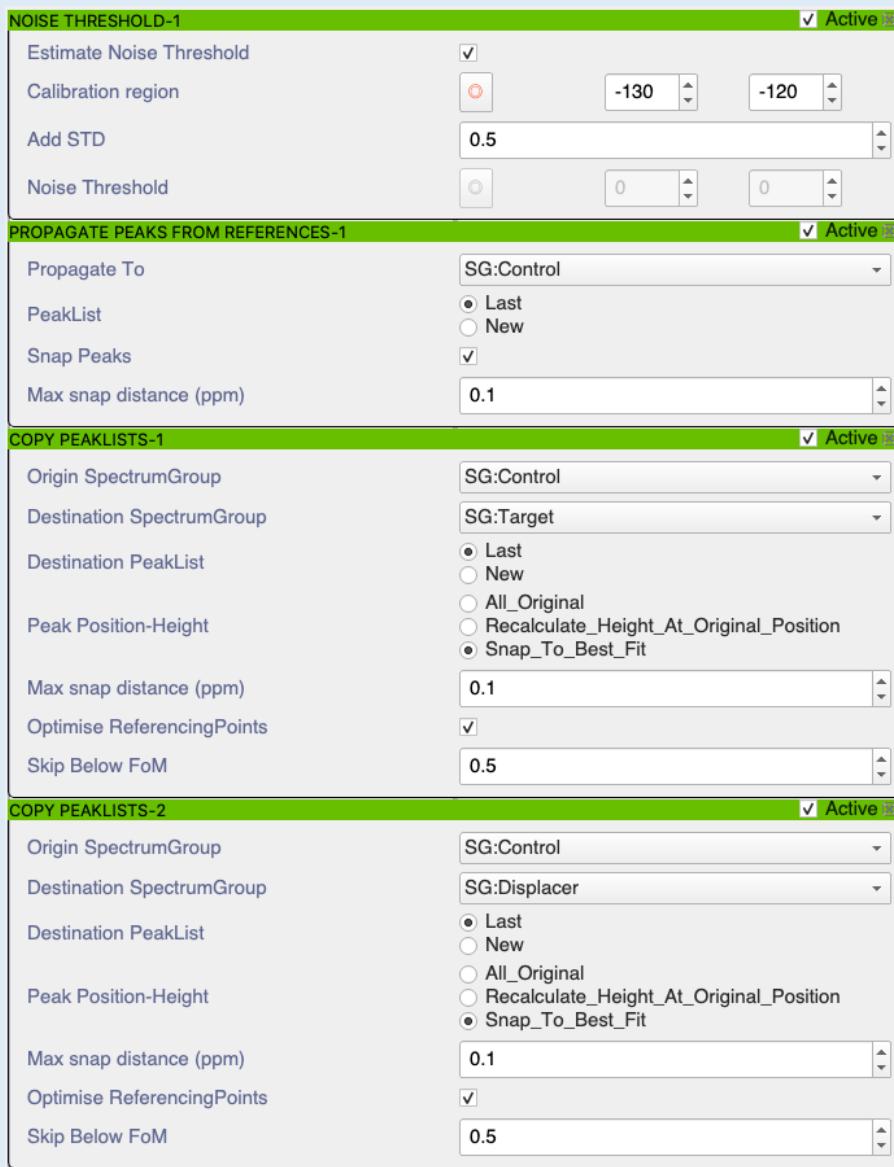
2. Analysis > Peak Picker 1D

3. AnalysisScreen > Annotate Peaks

- Select **Use linked Substance Name**

- Run the pipeline using the green play button (click once only!). A popup will appear when completed.
- If you wish, you can save this or any other pipeline you create by clicking on the icon. You can open a pipeline with – but check your parameters are set correctly, as not all are saved.

Automatic Peak Matching from Reference Singletons



Propagate Peaks will copy peaks from several (Reference Singleton) peak lists and collate them into a smaller number of (Control mixture) peak lists.

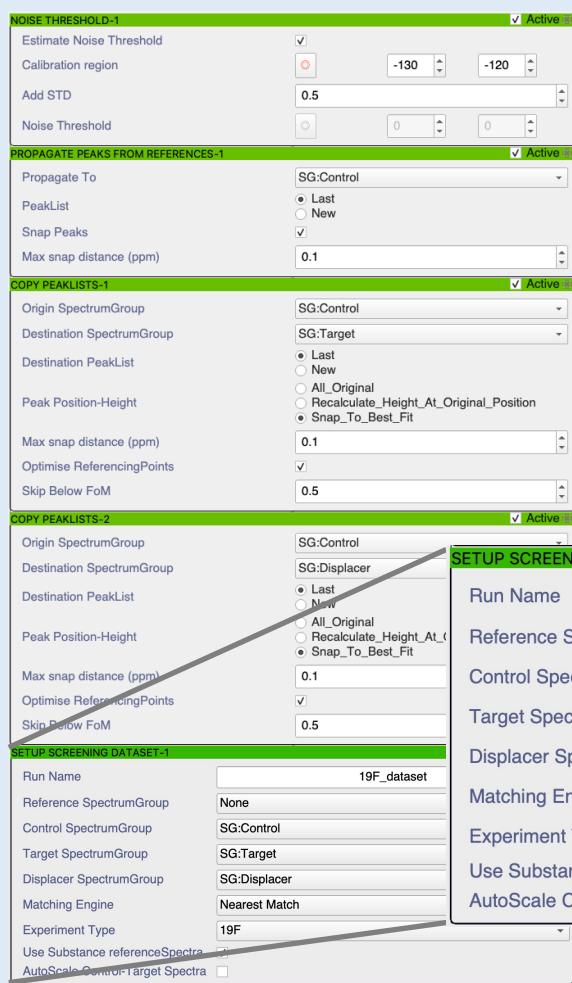
Copy Peaks will copy the peak lists one to one (from the Control mixtures to the Target and Displacer mixtures).

3E Setup screening pipeline

- Clear the input data (**right-click** → **Clear all**)
- On sidebar, multiselect **SG:Control**, **SG:Target** and **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
(note that using a Calibration region allows for the fact that different samples may have different signal to noise levels)
 - 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

continued....

Automatic Peak Matching from Reference Singletons



If you wish, you may choose to **AutoScale Control-Target Spectra**.

This will automatically rescale your spectra to take account of sample dilution and other experimental effects (e.g. changes in shimming) that might affect the relative intensities of your spectra. Auto-scaling provides a more reproducible and robust way of accounting for differences in spectrum intensities than doing this manually.

...continued

5. AnalysisScreen > Setup Screening dataset

- **Run name:** 19F_dataset
- **Reference SpectrumGroup:** None
- select **SG:Control**, **SG:Target**, **SG:Displacer** for their respective entries
- **Matching Engine:** Nearest Match
- tick **Use Substance ReferenceSpectra**
(this option ensures the Reference Singleton spectra are used as the Reference Spectra)

- Run the pipeline using the green play button.

The calculations should take less than a minute. When it is finished you should see a new entry under DataTables in your sidebar:

- **DataTables**
 - <New DataTable>
 - DT:19F_dataset
- **Collections**
- **Notes**

This is the new Screening Dataset that was set up in the final pipe. It is a DataTable containing all the peak positions, heights, volumes, binding scores, etc.

- Save the pipeline if you wish.

4 Binding Scores and Calculation Engines

The screenshot shows the CCPN software interface. At the top, a menu bar includes 'Screen' (which is highlighted in blue), 'Macro', 'Plugins', and 'Help'. Below the menu is a list of options: 'Generate Mixtures...' (G, M), 'Mixtures Analysis' (M, I), 'Screening Pipeline' (P, I), 'Hit Analysis' (H, A, currently selected), 'Compare Screens' (D, C), and 'Decomposition (PCA)' (D, E). The main workspace is titled 'Hit Analysis (beta)'. It features a 'DataTable' dropdown menu with options like '<>', '<>', and '<New Item...>'. The current selection is '19F_Pipeline'. To the right of the dropdown is a 'View' section with checkboxes for 'Substances', 'Samples', 'Matches Only', and a filter icon. Below this is a table with columns: '#', 'Sample Name', 'Binding Score', 'Sample Displacement', 'Sample Matching', and 'Substance Count'. A chart titled 'Scores' shows a distribution centered at 0 with a dashed vertical line. A 'Select X-axis' button is below the chart. At the bottom, there's another table titled 'Peak Binding Scores' with columns: 'Substance Name', 'Sample Name', 'Peak Binding Score', 'Peak Displacement', 'Peak Matching', 'Peak Lab', 'Ref-Targ Δ Shift', 'Ref-Targ |Δ| Shift', 'Cont-Targ |Δ| Shift', and 'Cont-Targ Δ Shift'. The 'Sample Name' column is currently sorted.

4A Open data in program

You can now either continue with your project from **Section 3** or load a project which has already completed those steps:

- Drag and drop the **CcpnTutorialDataScreeningJan2023/19F/dataset_2/dataset2_completed.ccpn** folder in the sidebar or Drop Area.
- Open a spectrum in a Spectrum Display module.

Open the Hit Analysis module:

- Go to **Main Menu** -> **Screen** -> **Hit Analysis** or use shortcut **HA**.
- From the **DataTable** drop-down menu select **19F_dataset** and wait a few moments for the tables to be updated.

4 Binding Scores and Calculation Engines

Hit Analysis Module Table Selections

The Hit Analysis module has multiple dynamic selections

Substance View: selecting a row on the **Substances** table will:

1. List all contributing peak matches in the **Peak Binding Scores** table
2. Display all the spectra associated with the substance
3. Select the relevant items in the Scores scatter plot on the right

Sample View: selecting a row on the **Samples** table will:

1. List all contributing peak matches in the **Peak Binding Scores** table for all the substances present in the sample
2. Display all the spectra associated with the sample, including all reference spectra

All Views: 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 **Match Editing** pop-up (click on **New/Edit Match...** to show this).

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

| # | Substance Name | Sample Name | Substance Binding Score | Substance Displacement Score | Substance Matching Score | Flag | Relative S/N Ratio |
|----|----------------|-------------|-------------------------|------------------------------|--------------------------|------|--------------------|
| 1 | C101 | Target_1 | 0.142 | 0.286 | 5 | | 67.85 |
| 12 | C102 | Target_1 | 0.202 | 0.162 | 5 | | 81.07 |
| 14 | C103 | Target_1 | 0.124 | | | | |
| 15 | C104 | Target_1 | 0.194 | | | | |
| 16 | C105 | Target_1 | 0.267 | | | | |
| 17 | C106 | Target_1 | 0.126 | | | | |

Substances View

Peak Binding Scores Table

| # | Sample Name | Sample Similarity Score | Sample Binding Score | Sample Displacement Score | Sample Matching Score | Substance Count |
|---|-------------|-------------------------|----------------------|---------------------------|-----------------------|-----------------|
| 1 | Target_1 | 0.995 | 2.69 | 13.2 | 5 | 20 |
| 2 | Target_2 | 0.996 | 2.52 | 9.33 | 5 | 20 |
| 3 | Target_3 | 0.995 | 2.56 | | | |
| 4 | Target_4 | 0.996 | 2.67 | | | |
| 5 | Target_5 | 0.995 | 3.42 | | | |
| 6 | Target_6 | 0.995 | 3.23 | 5.15 | 5 | 20 |

Samples View

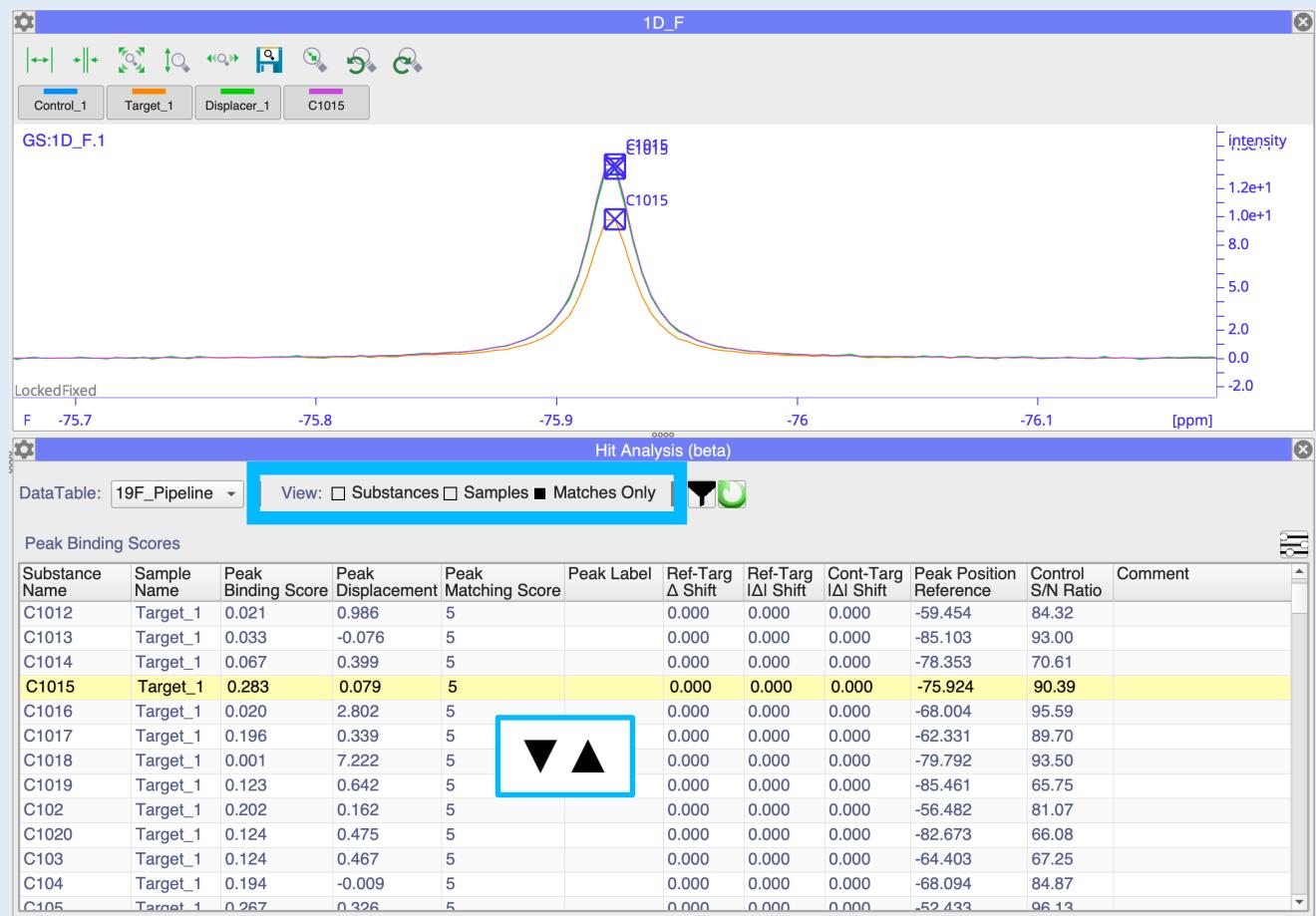
Peak Binding Scores Table

| Substance Name | Sample Name | Peak Binding Score | Peak Displacement | Peak Matching | Ref-Targ Δ Shift | Ref-Targ IΔ Shift | Cont-Targ Δ Shift | Cont-Targ IΔ Shift | Peak Position Reference | Control S/N Ratio | Comment |
|----------------|-------------|--------------------|-------------------|---------------|------------------|-------------------|-------------------|--------------------|-------------------------|-------------------|---------|
| C101 | Target_1 | 0.142 | 0.286 | 5 | 0.000 | 0.000 | 0.000 | 0.000 | 58.554 | 67.85 | |
| C1010 | Target_1 | 0.052 | 0.912 | 5 | 0.000 | 0.000 | 0.000 | 0.000 | | | |
| C1011 | Target_1 | 0.191 | 0.003 | 5 | 0.000 | 0.000 | 0.000 | 0.000 | | | |
| C1012 | Target_1 | 0.021 | 0.986 | 5 | 0.000 | 0.000 | 0.000 | 0.000 | | | |

Matches Only View

Peak Binding Scores Table

4 Binding Scores and Calculation Engines



4B Looking through peaks manually

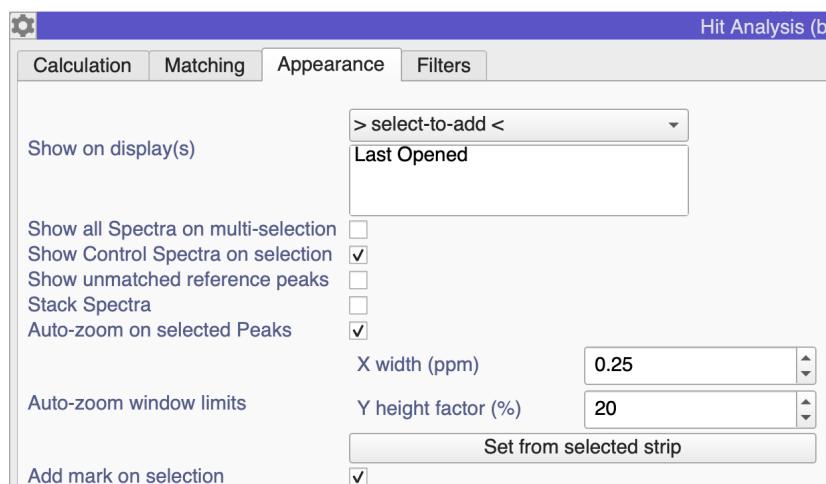
You can scan through all your peaks by hand fairly quickly to see if you spot anything interesting or unusual:

- Set the **View Mode** to **Matches Only**, so that you only see the **Peak Binding Scores** table.
- Use your **up/down arrow keys** to move down the list. Each time the Spectrum Display will automatically focus on the peak in question.

If you prefer to look at your peaks by Sample, simply:

- Set the **View Mode** to **Samples**
- Select the sample of interest in the upper table
- Use the **up/down arrow keys** in the **Peak Binding Scores Table** below to move through and see the peaks for that sample.

You can change the auto-zoom options in the **Appearance** tab of the Hit Analysis module **Settings** if you like.



4 Binding Scores and Calculation Engines

Scores and Scoring Engines

Peak Binding Score

This is a measure of how well a substance binds to the target. It is shown in the **Peak Binding Scores Table** and can be calculated using peak height, linewidth or volume, depending on the **Mode** set in the **Calculation** tab of the Hit Analysis module **Settings**. (Note that linewidths and volumes are not calculated by default when peaks are picked and have to be determined with **Spectrum → Estimate Volumes or shortcut EV.**)

The binding score is calculated according to the **Peak Binding Score Engine** specified in the **Calculation** tab of the Hit Analysis module **Settings**. When hovering over one of the options, a Tooltip will appear showing the actual equation.

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** represents each Ligand signal in the presence of the target (Target)
- **V2** represents each Ligand signal in the absence of the target (Control)

The following arithmetic operations are supported:

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

The screenshot shows the 'Hit Analysis (beta)' software window. On the left, there's a navigation bar with icons for Settings, File, and Help. Below it, a toolbar has buttons for Calculation, Matching, Appearance, and Filters. The 'Calculation' tab is selected. Under 'Mode', 'Height' is chosen. Under 'Engine', 'AbsoluteRelativeChange' is selected, with a tooltip explaining the formula  $\frac{|V2| - |V1|}{|V2|}$  and stating that V1 is the target signal and V2 is the control signal. A higher value indicates a greater change. Under 'Displacement Engine', 'DisplacementFraction' is selected, with a formula  $v1/(V2+V3)$  displayed. To the right, a 'DataTable' titled '19F\_Pipeline' shows a list of substances and their sample names. The first few rows are: #, Substance Name, Sample Name; 134, C703, Target\_7; 135, C704, Target\_7; 136, C705, Target\_7. At the bottom, there are more rows: 140, C709, Target\_7; 141, C801, Target\_8; 142, C8010, Target\_8; 143, C8011, Target\_8.

### Peak Displacement Score

This is a measure of how well a substance binds the target in the presence of a Displacer/Competitor (which is usually expected to bind more strongly than the fragments used in the screen).

It is calculated using the equation selected in **Displacement Engine**. Again, it is possible to enter your own equation, defining the Displacer signal with the variable **V3**.

**continued...**

# 4 Binding Scores and Calculation Engines

## ... Scores and Scoring Engines continued

### Substance and Sample Binding Scores

These scores are shown in the **Substances** and **Samples Tables**, respectively, and are derived from all the peaks matched to that Substance or Sample. Therefore, they can be used to assess the Substance or Sample binding quality. Note that in the case of  $^{19}\text{F}$  screens (and this demo set), most Substances only have one peak, so the Substance Binding Score will usually be the same as the Peak Binding Score.

The Substance and Sample Binding Scores are given by the **Total Substance/Sample Binding Score** selected in the Calculation tab of the Hit Analysis Settings:

**min**: the minimum peak binding score in that substance/sample

**max**: the maximum peak binding score in that substance/sample

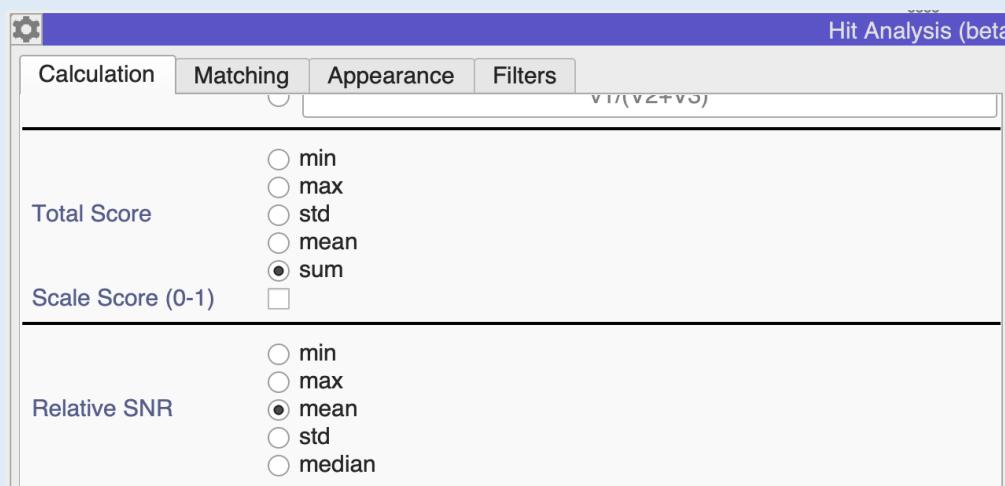
**stDev**: the standard deviation of all peak binding scores in that substance/sample

**mean**: the mean peak binding score for that substance/sample

**sum**: the sum of all peak binding scores for that substance/sample

If you wish, you can select **Scale Scores (0-1)** to get a quick sense of their relative size.

You can select similar options for the **Relative Signal to Noise Ratio (SNR)** shown in the tables.



### Peak Matching Scores

When matching peaks automatically (see Section 9) the Peak Matching Scores give an indication for how reliably the peaks were matched: 0 is unreliable, 5 is most reliable.

# 4 Binding Scores and Calculation Engines

The screenshot shows the 'Hit Analysis (beta)' settings panel with the 'Calculation' tab selected. The interface includes tabs for 'Calculation', 'Matching', 'Appearance', and 'Filters'. In the 'Calculation' tab, there are several configuration options:

- Show on display(s):** A dropdown menu set to '> select-to-add <' with a 'Last Opened' option.
- Show all Spectra on multi-selection:** An unchecked checkbox.
- Show Control Spectra on selection:** A checked checkbox.
- Show unmatched reference peaks:** An unchecked checkbox.
- Stack Spectra:** An unchecked checkbox.
- Auto-zoom on selected Peaks:** A checked checkbox.
- X width (ppm):** A numeric input field set to 0.25.
- Y height factor (%):** A numeric input field set to 20.
- Set from selected strip:** A button.
- Add mark on selection:** A checked checkbox.
- Auto-Update on setting(s) changed:** A checked checkbox.
- Auto-Update on peak(s) changed:** An unchecked checkbox.
- Auto-Update matching score:** A checked checkbox.

## 4C Updating the tables

- Open the Hit Analysis **Settings** panel with the gear icon
- In the **Calculation** tab, try changing the **Peak Binding Score Engine** or **Total Score** calculation method.

As soon as you change any settings or modify / re-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**

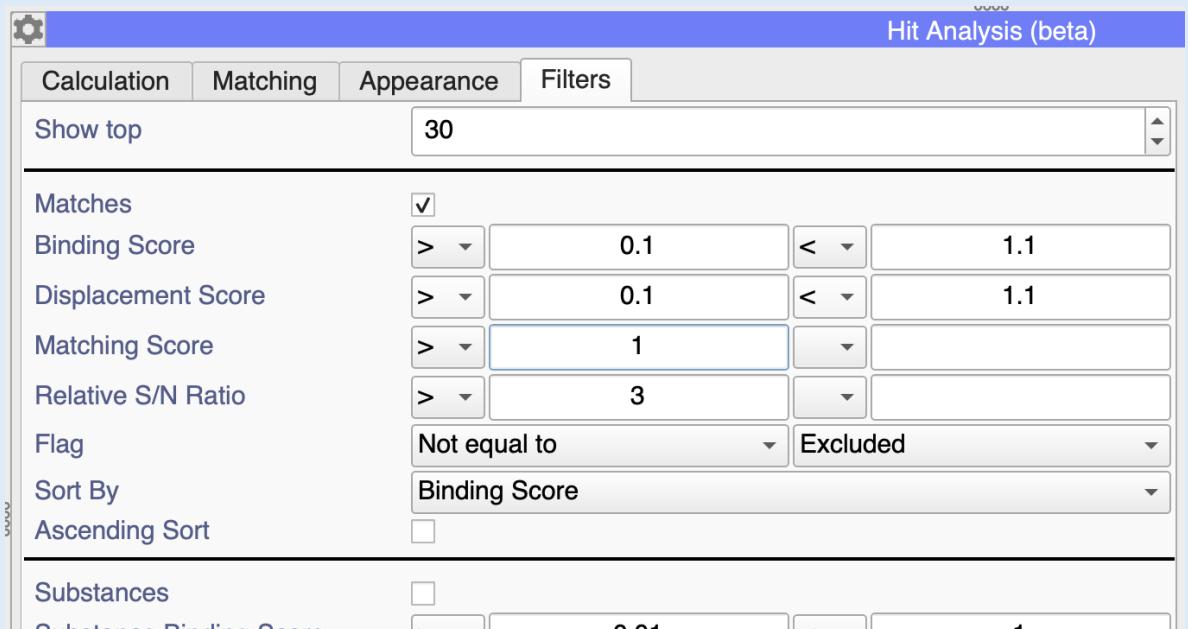
The refresh button will turn orange whenever changes are detected while working on the dataset. Click the refresh button to update all scores, after which it will turn green.

After trying out the effect of changes make sure you select

- Engine: **AbsoluteRelativeChange**

in the **Calculation** tab.

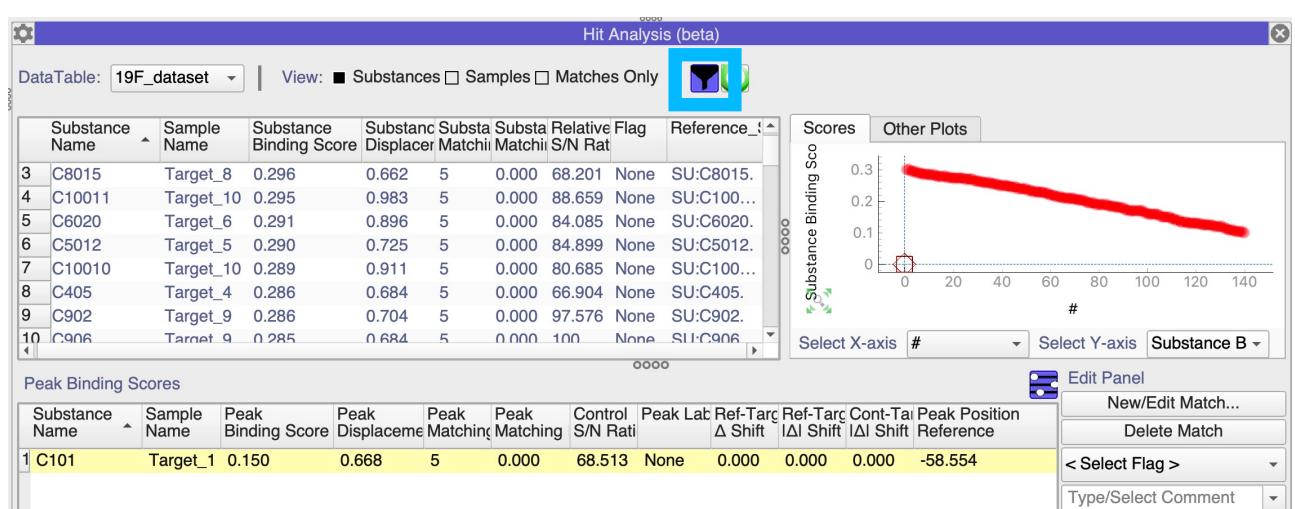
# Filtering and Plots

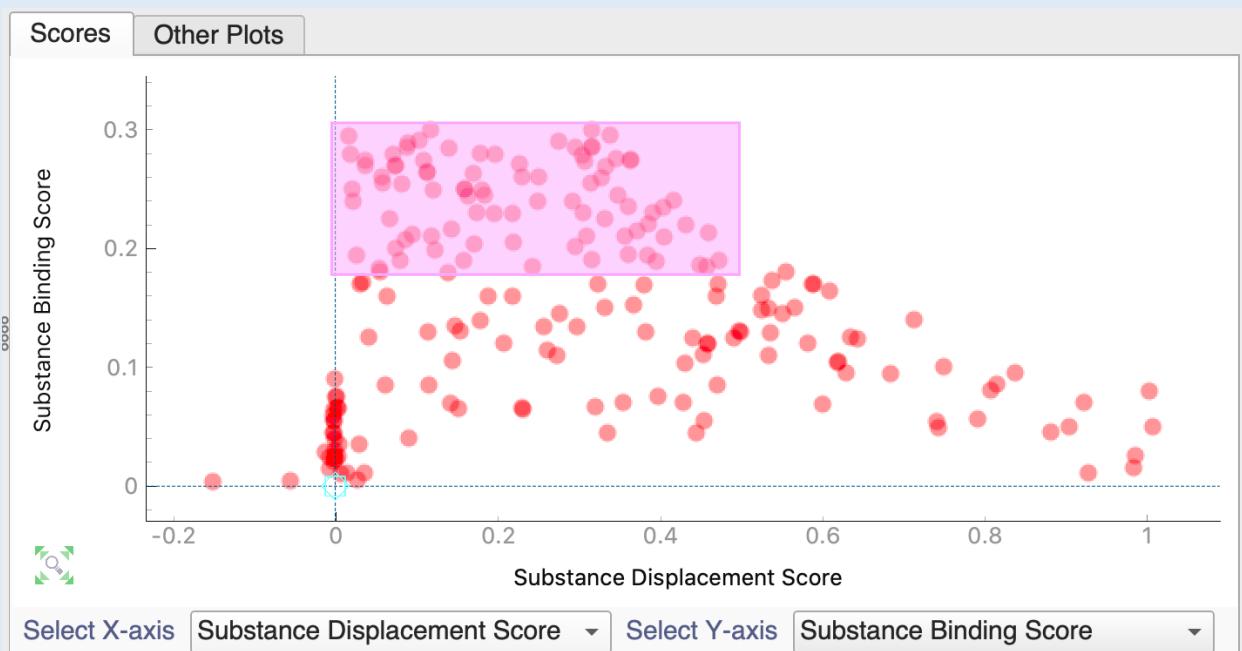


## 5A Filtering

Filters are a useful way to find to find Hits quickly and they also reduce the size of the tables which can speed up the program.

- Select the **Substances** View mode
  - Go to Hit Analysis Settings Panel → Filters
  - Select the **Matches** filters and set them up as shown in the image above
  - Close the Settings and click on the filtering icon to activiate the filters.
- The filtering icon will go blue and the Substances table and Scores Plot will update based on the filtering criterea.





## 5B Plots

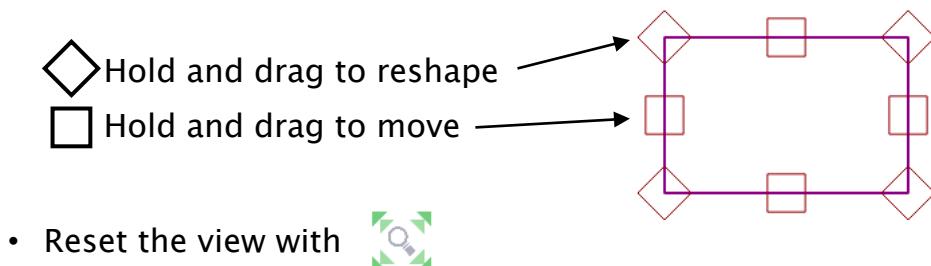
The Hit Analysis module has a Plots window in which you can plot any two variables of your choice against one another. If you have done automatic peak matching, you could, for example, plot the **Substance Matching Score** along the y-axis in order to find outliers quickly.

- Selecting an item in a table will select it in the plot and vice versa.

You can manipulate the plot and its points like a spectrum and peaks:

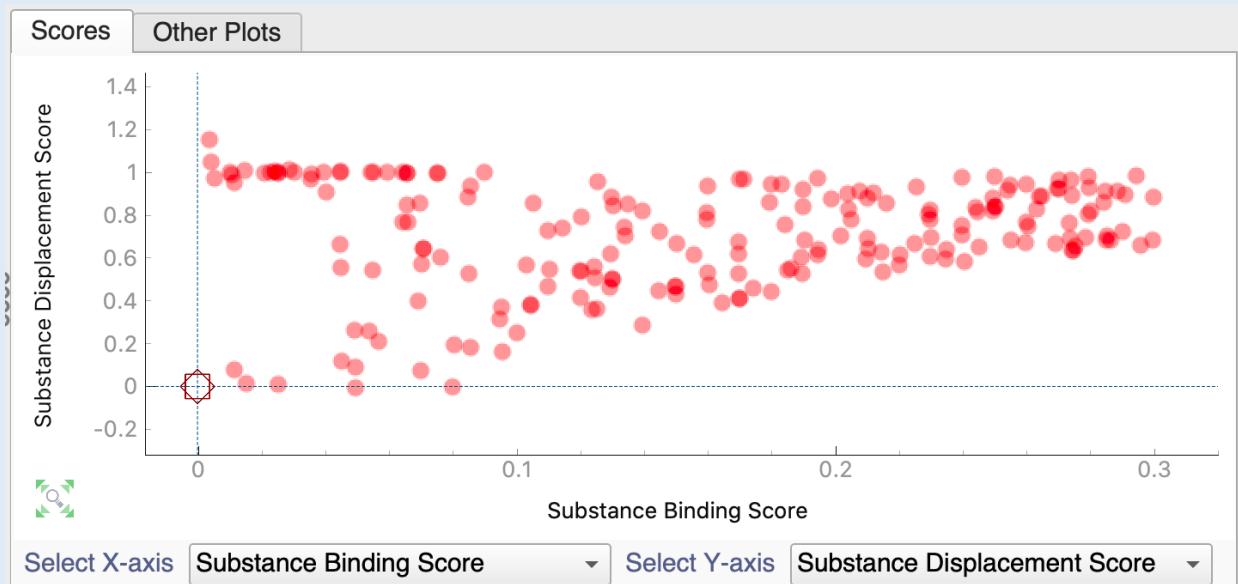
- Zoom with the **mouse wheel**, either on the plot or selectively on a single axis
- Move the plot around with **left-drag**
- (Multi-)select items with **Ctrl/Cmd + left-click**
- Select items in an area with **Ctrl/Cmd + left-drag**

This will draw a Region of Interest (ROI) box which you can change as follows:



Additional options are available in the **Appearance** tab of the **Settings**:

|                                                                 |                                         |        |        |        |
|-----------------------------------------------------------------|-----------------------------------------|--------|--------|--------|
| Roi Limits                                                      | xMin 0                                  | xMax 0 | yMin 0 | yMax 0 |
| <input checked="" type="checkbox"/> Link Roi with selection box |                                         |        |        |        |
| ROI line colour <input checked="" type="color"/> purple         |                                         |        |        |        |
| Scatter point colour                                            | <input checked="" type="color"/> red    |        |        |        |
| Scatter point size                                              | 10                                      |        |        |        |
| Scatter point Symbol                                            | <input checked="" type="radio"/> circle |        |        |        |



## 5c Useful Plots for analysing Screening Data

### # vs Substance Matching Score

This can be a quick way to assess how many matches may need manual checking. Sorting the table by the Matching score will enable you to move through these quickly and assess them, making changes if necessary.

### Substance Binding Score vs. Relative Signal to Noise

This can be a helpful way to choose thresholds for the Signal to Noise and Binding Scores.

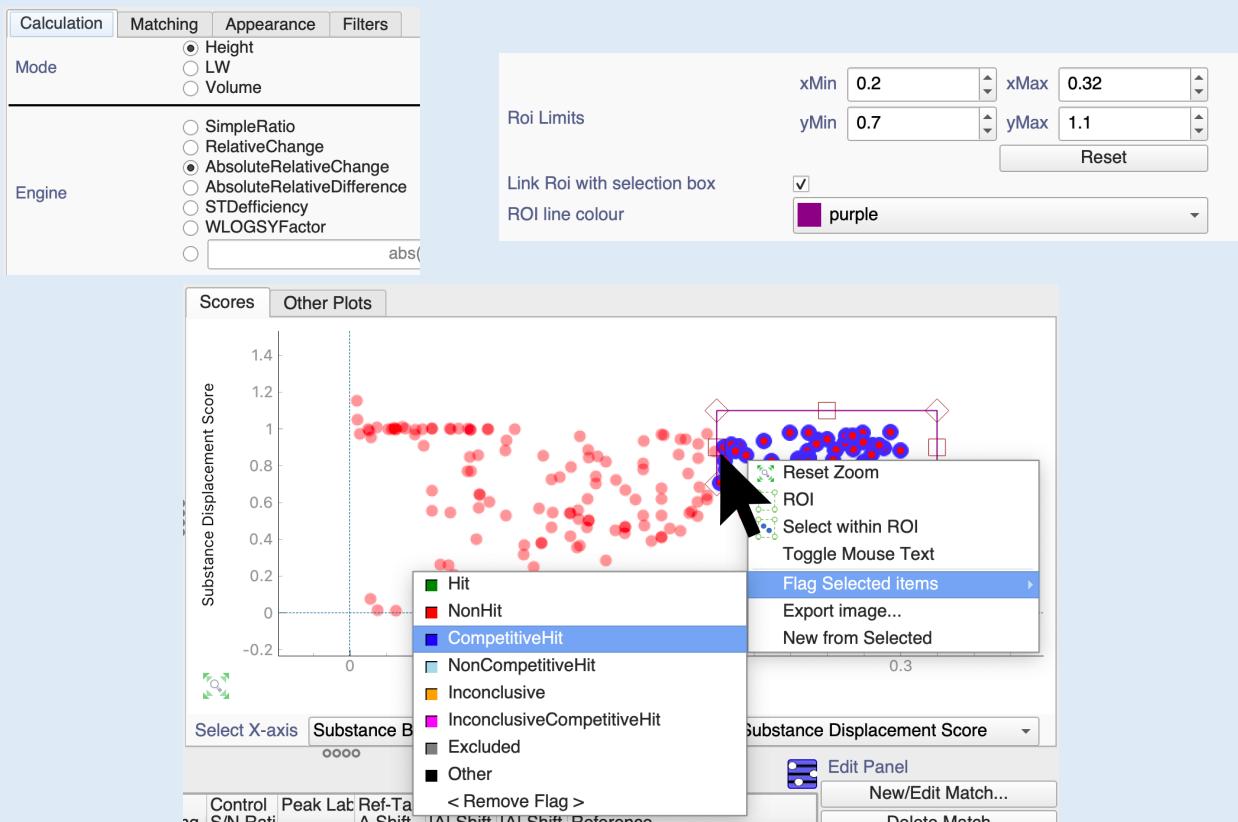
### # vs Substance Binding Score

When filtering with **Sort By Binding Score**, this can be a helpful way to choose a suitable threshold Binding Score above which to mark Hits.

### Substance Binding Score vs Substance Displacement Score

If using a Displacer, this is a useful plot on which to identify Competitive Hits.

Further types of plots, including the molecular structure of the substances (if SMILES were entered into the proproject), are available in the **Other Plots** tab. More information on these is available from our online documentation on the Hit Analysis Plots at <https://www.ccpn.ac.uk/manual/v3/ScreenHAPlots.html>.



Before flagging Substances, always inspect the matches with the table selections. Snap peaks with the shortcut **SE**, or correct matches using the **New/Edit Match...** button in the **Edit Panel**.

## 5D Setting Flags in the Scatter Plot

- View mode: **Substances**
- Scatter Plot: X-axis: **Substance Binding Score**  
Y-axis: **Substance Displacement Score**

Open Settings:

- **Calculation** tab
  - Engine: **AbsoluteRelativeChange**
- **Appearance** tab
  - **Roi Limits**: click **Reset**
  - xMin: 0.2      xMax: 0.32
  - yMin: 0.7      yMax: 1.1

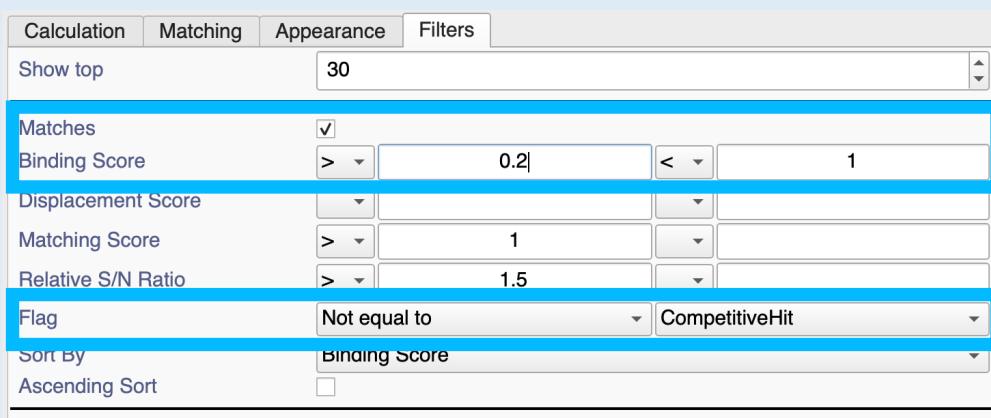
This will create a **Region of Interest** on the scatter plot.

On the scatter plot:

- **right-click** on an item in the plot → **Select within ROI**
- **right-click** on an item in the plot → **Flag Selected Items** → **CompetitiveHit**

Note: when using Reference Mixtures imported from a NEF file, any peaks that have a **Merit** value of 0, have their matches automatically flagged as **Excluded**. Matches with this flag are not included in aggregated (total) Substance/Sample scores.

# Filtering and Plots



## 5E Setting Flags in the Peak Binding Scores Table

We will now mark all the non-competitive Hits:

- View mode: **Matches Only**

Open **Settings**:

- **Filters tab**
  - Matches: **ticked**
  - Binding Score: **> 0.2**      **< 1**
  - Flag: **Not equal to**      **CompetitiveHit**

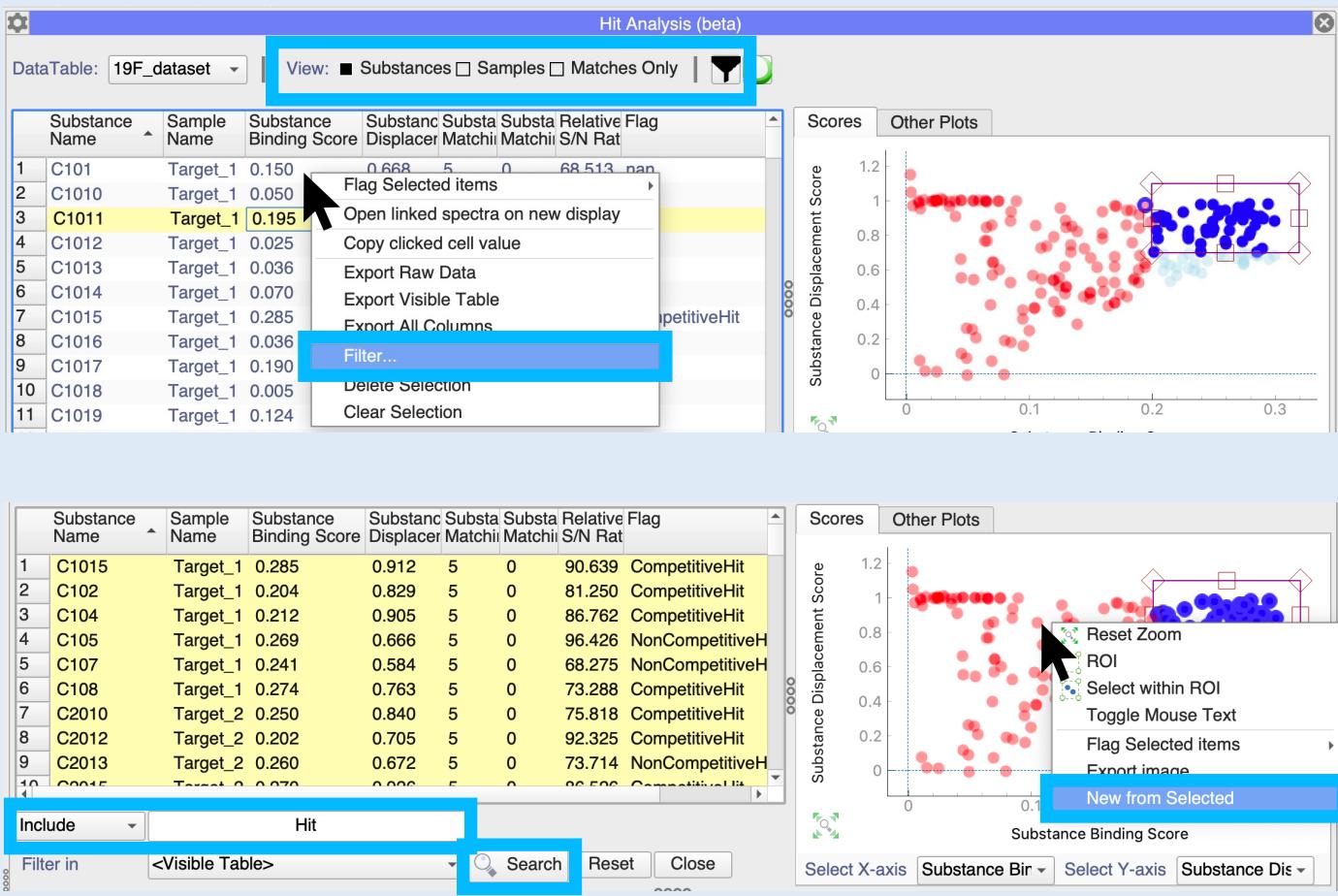
In the main Hit Analysis Module:

- Make sure Filtering is on ( ) and you click the Update button if orange
- Type **Cmd/Ctrl+A** to select all rows in the table

In the Edit Panel:

- click on **<Select Flag>**
- select **NonCompetitiveHit**

# Exporting Data



## 6A Extracting data

- Switch the **Filtering** off →
- Set the View mode to **Substances**

Use the normal Table Filtering function:

- **right-click** and select **Filter**
  - Select **Include** from the drop-down menu
  - Type **Hit** into the search
  - Press **Enter** or the **Search** button
- Select all the remaining rows with **Ctrl/Cmd+A**

Move to the scatter plot:

- **right-click** → **New from Selected**

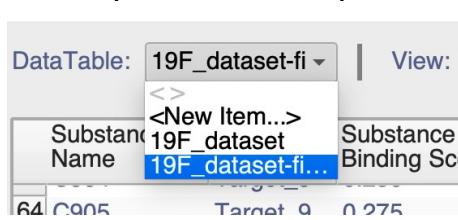
This will create a new dataset containing only this subset of Substances. You can rename the dataset by double-clicking on it in the sidebar under

**DataTables**.

▼ **DataTables**  
 <New DataTable>  
 DT:19F\_dataset  
 DT:19F\_dataset-filtered

- Select the newly created dataset on the Hit Analysis module, by selecting it from the **DataTable** dropdown menu.

Either continue to inspect the data as shown in Sections 4 and 5 or export the table as shown in Section 6B.



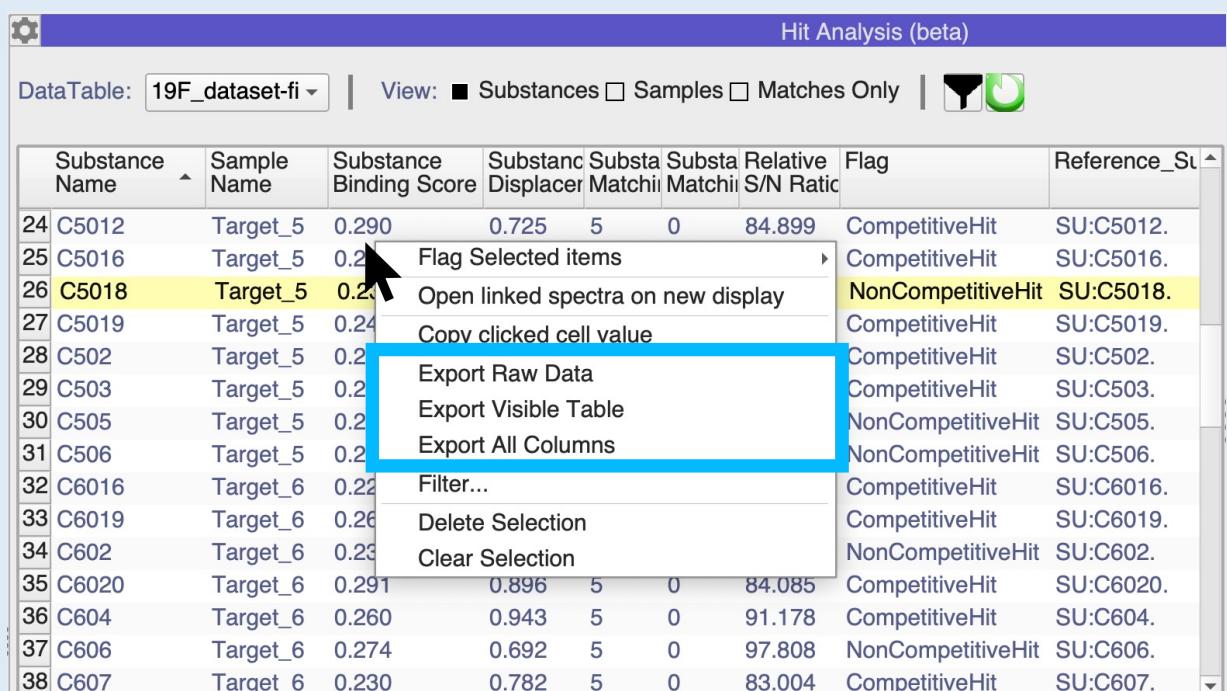
# Exporting Data

There are three options for exporting data from a Hit Analysis module table:

**Export Visible Table** only exports those columns and rows currently in the table (e.g. after any Table Filtering has been applied or columns have been added / removed after right-clicking on the column heading)

**Export All Columns** will export all the columns in the table regardless of which ones are currently being displayed. However, only those rows will be exported that are currently present, based on filtering.

**Export Raw Data** will export the underlying DataTable (accessible from **DataTables** in the sidebar) which is used to build the Hit Analysis module. This includes peak metadata, peak ppm positions, heights etc.



The screenshot shows a table titled "Substance Name" with various columns including Sample Name, Substance Binding Score, Substanc Displacer, Substa Matchii, Substa Matchii, Relative S/N Ratio, Flag, and Reference\_Su. A context menu is open over the row for substance C5018. The menu items are: Flag Selected items, Open linked spectra on new display, Copy clicked cell value, Export Raw Data, Export Visible Table, and Export All Columns. The last three items are highlighted with a blue box.

|    | Substance Name | Sample Name | Substance Binding Score | Substanc Displacer                 | Substa Matchii | Substa Matchii | Relative S/N Ratio | Flag              | Reference_Su |
|----|----------------|-------------|-------------------------|------------------------------------|----------------|----------------|--------------------|-------------------|--------------|
| 24 | C5012          | Target_5    | 0.290                   | 0.725                              | 5              | 0              | 84.899             | CompetitiveHit    | SU:C5012.    |
| 25 | C5016          | Target_5    | 0.2                     | Flag Selected items                |                |                |                    | CompetitiveHit    | SU:C5016.    |
| 26 | C5018          | Target_5    | 0.2                     | Open linked spectra on new display |                |                |                    | NonCompetitiveHit | SU:C5018.    |
| 27 | C5019          | Target_5    | 0.24                    | Copy clicked cell value            |                |                |                    | CompetitiveHit    | SU:C5019.    |
| 28 | C502           | Target_5    | 0.2                     | Export Raw Data                    |                |                |                    | CompetitiveHit    | SU:C502.     |
| 29 | C503           | Target_5    | 0.2                     | Export Visible Table               |                |                |                    | CompetitiveHit    | SU:C503.     |
| 30 | C505           | Target_5    | 0.2                     | Export All Columns                 |                |                |                    | NonCompetitiveHit | SU:C505.     |
| 31 | C506           | Target_5    | 0.2                     |                                    |                |                |                    | NonCompetitiveHit | SU:C506.     |
| 32 | C6016          | Target_6    | 0.22                    | Filter...                          |                |                |                    | CompetitiveHit    | SU:C6016.    |
| 33 | C6019          | Target_6    | 0.26                    | Delete Selection                   |                |                |                    | CompetitiveHit    | SU:C6019.    |
| 34 | C602           | Target_6    | 0.23                    | Clear Selection                    |                |                |                    | NonCompetitiveHit | SU:C602.     |
| 35 | C6020          | Target_6    | 0.291                   | 0.896                              | 5              | 0              | 84.085             | CompetitiveHit    | SU:C6020.    |
| 36 | C604           | Target_6    | 0.260                   | 0.943                              | 5              | 0              | 91.178             | CompetitiveHit    | SU:C604.     |
| 37 | C606           | Target_6    | 0.274                   | 0.692                              | 5              | 0              | 97.808             | NonCompetitiveHit | SU:C606.     |
| 38 | C607           | Target_6    | 0.230                   | 0.782                              | 5              | 0              | 83.004             | CompetitiveHit    | SU:C607.     |

## 6B Export data

- Right-click on a Substance Table row and select one of these options:
  - Export Raw Data
  - Export Visible Data
  - Export All Columns

In the Save Table file dialog:

- Type a name plus the extension OR select the extension from the dropdown menu.

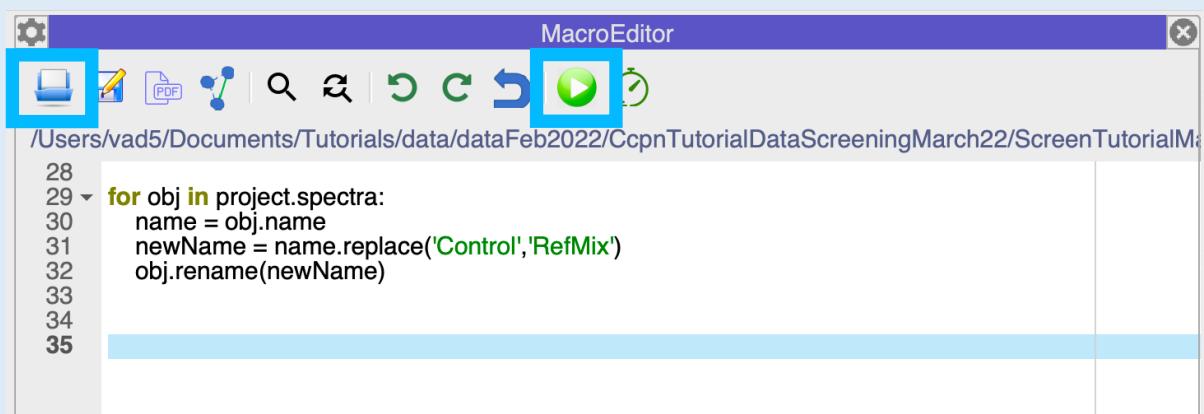
Possible file types and their extensions are:

|                 |       |
|-----------------|-------|
| Excel           | .xlsx |
| Comma-separated | .csv  |
| Tab-separated   | .tsv  |
| JSON            | .json |

# Create a NEF file

All the information regarding Samples, Substances, Spectra and Spectrum Groups can be contained within an NMR Exchange Format (NEF) file.

A NEF file can be created from an existing screening dataset with a few steps. These include **renaming** Samples, Spectra and Spectrum Groups from **Control** to **ReferenceMixtures** (or similar); **deleting** all other spectra and samples and finally **linking** the Substances to the Reference mixtures.



## 7A Load project

- Continue with your Hit Analysis project or load the **dataset2\_completed.ccpn** project from the **CcpnTutorialDataScreeningJan2023/19F/dataset\_2/** directory.

## 7B Renaming Control Samples, Spectra and Spectrum Group

- Double-click on the **SG:Controls** Spectrum Group in the sidebar and rename it to **RefMix**

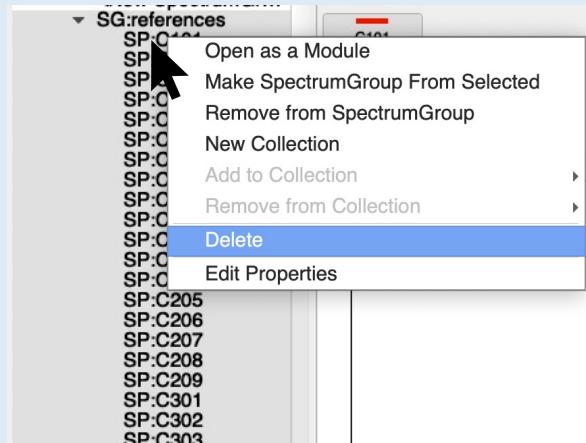
To rename the Control Spectra and Samples we will use a macro.

- Go to **Main Menu** → **Macro** → **New Macro Editor**
- Click on the **Open File** icon and select the **CcpnTutorialDataScreeningJan2023/19F/dataset\_3/macros/RenameSpectra.py** file (or drag and drop the file onto the Macro Editor).
- Click on the **Play** button to run the macro. This may take a few moments.
- Now open the **CcpnTutorialDataScreeningJan2023/19F/dataset\_3/macros/RenameSamples.py** file and run this macro, too.

You should see in the sidebar that all Control Spectra and Samples have been renamed.

You can set your preferred Macro directory path in **File / Preferences / General**.

# Create a NEF file



```
Python Console
SP:C208_1 , SP:C0020_1 , SP:C308_1 , SP:C10015_1 , SP:C10
project.deleteObjects('SP:Target_12', 'SP:Target_20', 'SP:Target_11', 'SP:Target_13', 'SP:Target_15', 'SP:Target_17'
project.deleteObjects('SP:Displacer_20', 'SP:Displacer_18', 'SP:Displacer_12', 'SP:Displacer_19', 'SP:Displacer_17', 'SP:Displacer_15', 'SG:Displacer')

In [6]: sg = get("SG:References")
In [7]: project.deleteObjects(*list(sg.spectra)+[sg])
In [8]: sg = get("SG:Target")
In [9]: project.deleteObjects(*list(sg.spectra)+[sg])
In [10]: sg = get('SG:Displacer')
In [11]: project.deleteObjects(*list(sg.spectra)+[sg])
In [12]:
```

## 7C Delete Reference, Target and Displacer Data

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

- If it isn't open already, open the **Python Console** by pressing the **Spacebar** twice.
- Run the commands:

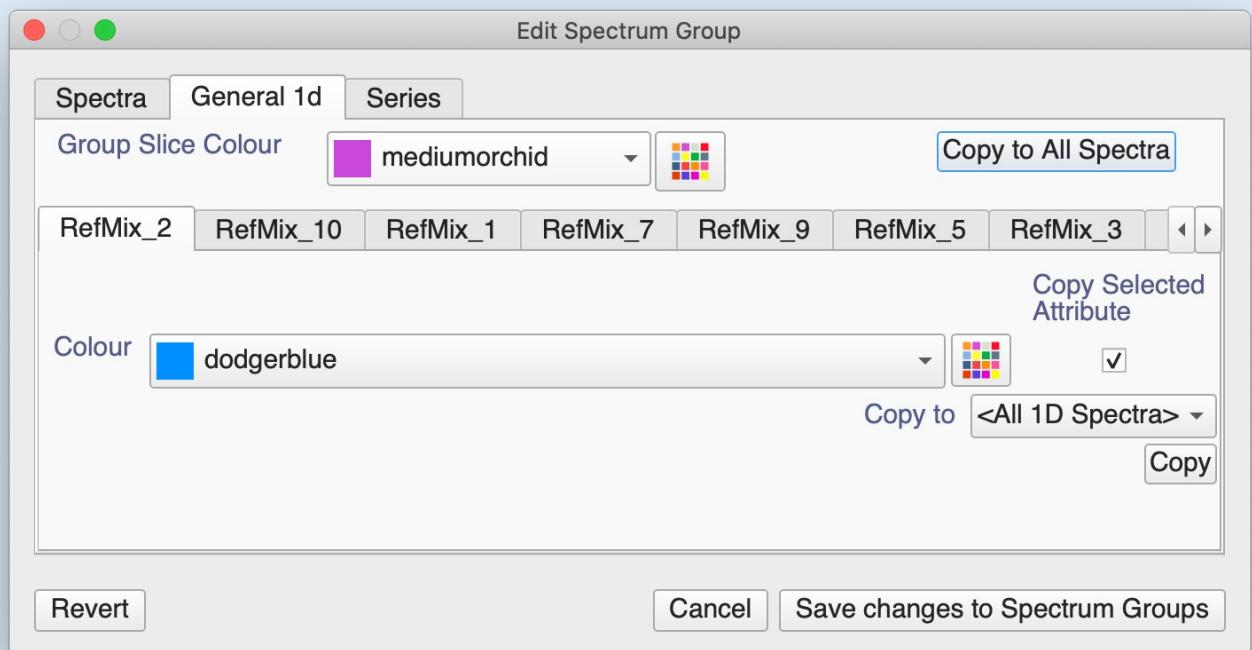
```
sg = get('SG:References')
project.deleteObjects(*list(sg.spectra)+[sg])
```

- Repeat for the **SG:Target** and **SG:Displacer** Spectrum Groups ammending the first line of code if you are using it.
- Expand the **Samples** branch in the sidebar and select all **Target** and **Displacer** samples and delete with **right-click → Delete**

## 7D Link Substances to Spectra

- Go to **Main Menu -> Macro -> New Macro Editor**
- Click on the **Open File** icon and select the **CcpnTutorialDataScreeningJan 2023/19F/dataset\_3/macros/LinkSubstancesToSpectra.py** file.
- Click on the **Play** button to run the macro.

# Create a NEF file



## 7E Recolour Reference Mixture Spectra (optional)

- If you wish, change the colour of the spectra in the new **SG:RefMix** Spectrum Group as shown in **Section 2E** by double-clicking on **SG:RefMix** in the sidebar, going to the **General 1d** tab, selecting a new **Group Slice Colour** (e.g. **mediumorchid**) and pressing **Copy to All Spectra** followed by either **Cancel** or **Save changes to Spectrum Groups**.

## 7F 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 in 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 will be flagged as Excluded in the resulting Screening Dataset.

## 7G Add extra annotations for impurities, solvents etc and exclude them from calculations

- Close all Modules in the DropArea.
- Drag and drop the **SG:RefMix** Spectrum Group from the 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 Macro Editor**
- Open the **CcpnTutorialDataScreeningJan2023/19F/dataset\_3/macros /AnnotateCurrentPeaks.py** file or drag and drop it onto the Macro Editor.
- Run with the play button (while the peaks are selected).

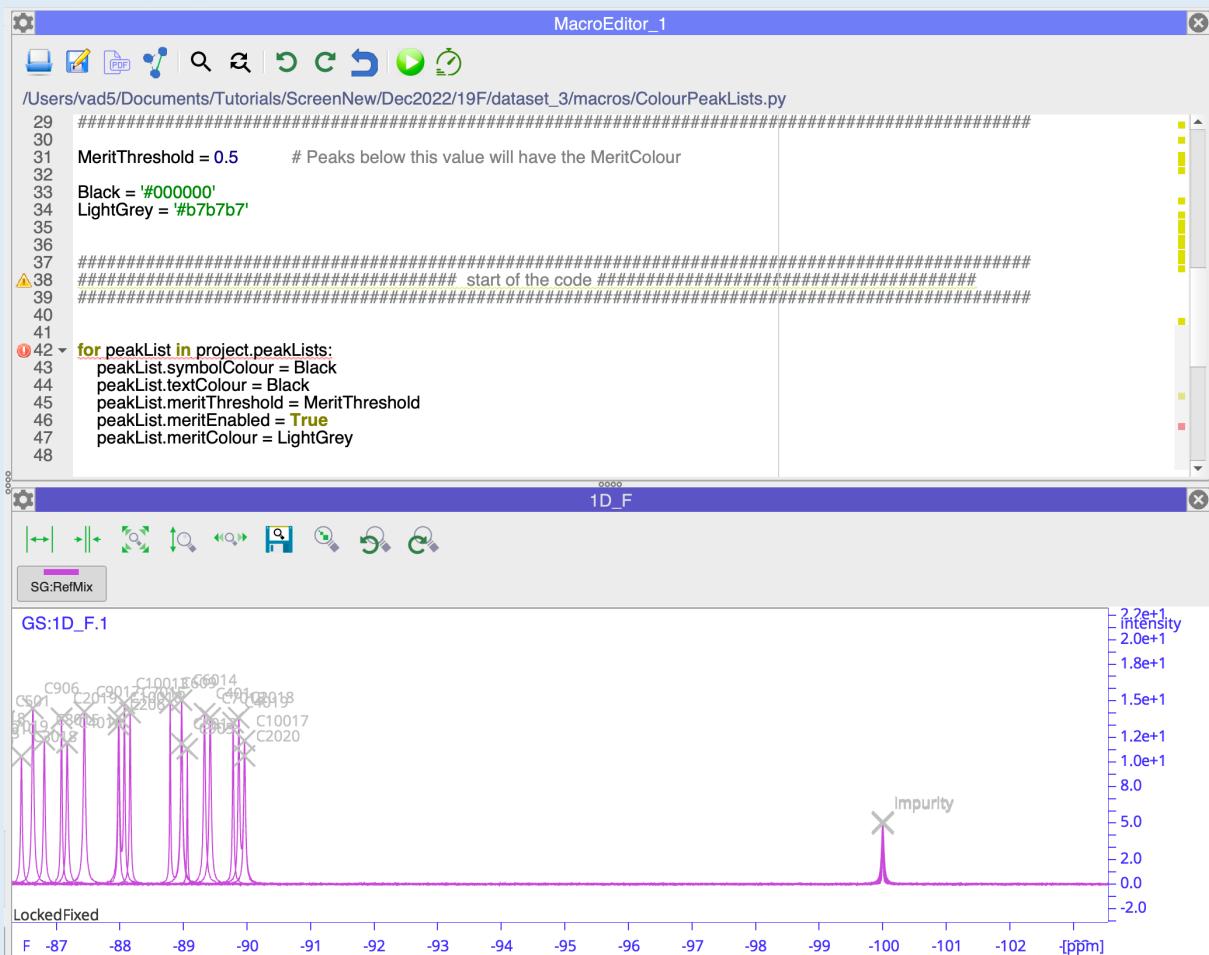
A code snippet is shown below

```
tag = 'Impurity'

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

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

# Create a NEF file



## 7H 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:

- Open the **CcpnTutorialDataScreeningJan2023/19F/dataset\_3/macros/ColourPeakLists.py** macro in the Macro Editor and run.

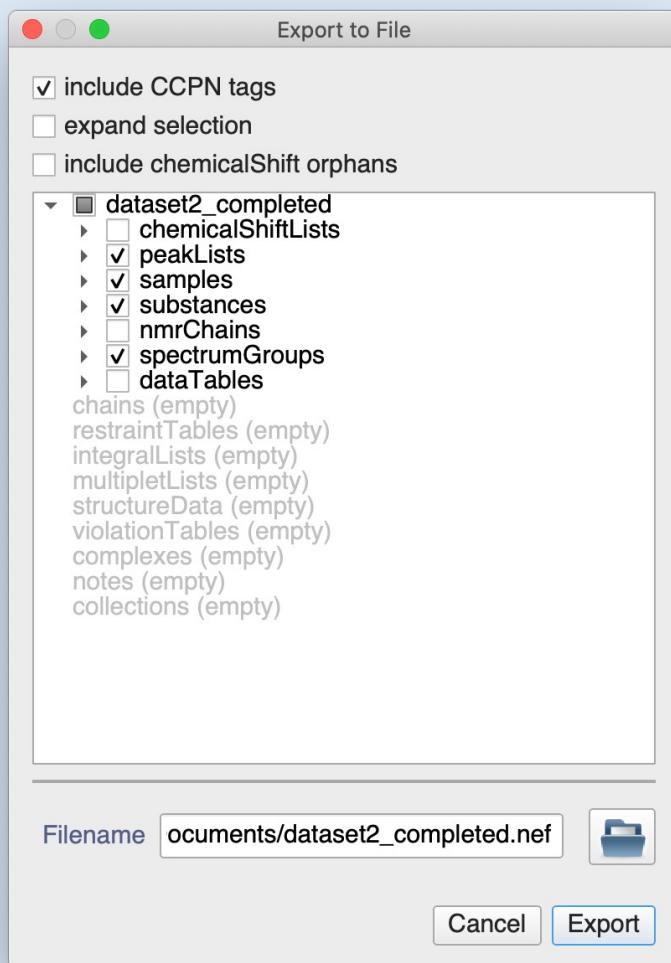
A code snippet is shown below

```
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&Pasting code from a PDF might lose the original indentation causing syntax errors

# Create a NEF file



## 7J Export to NEF

Export all the data 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,

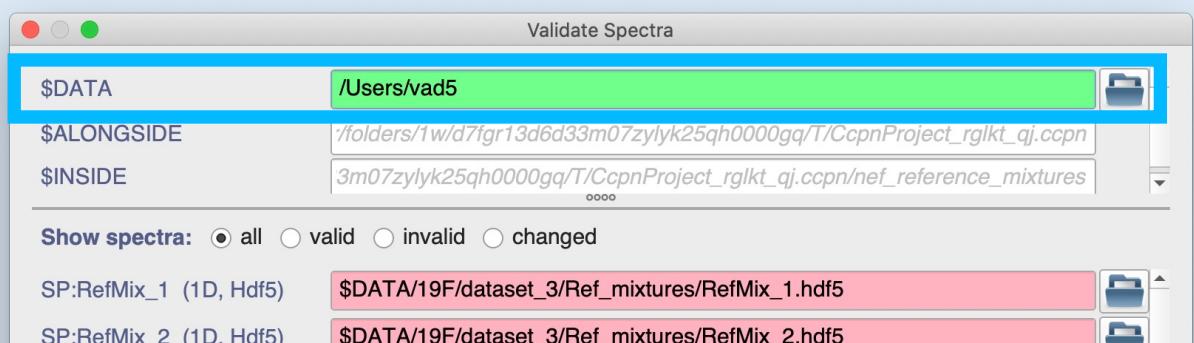
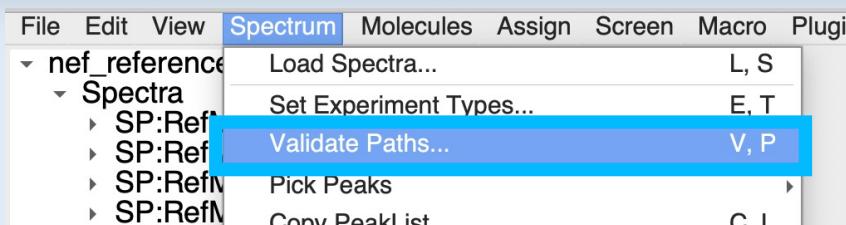
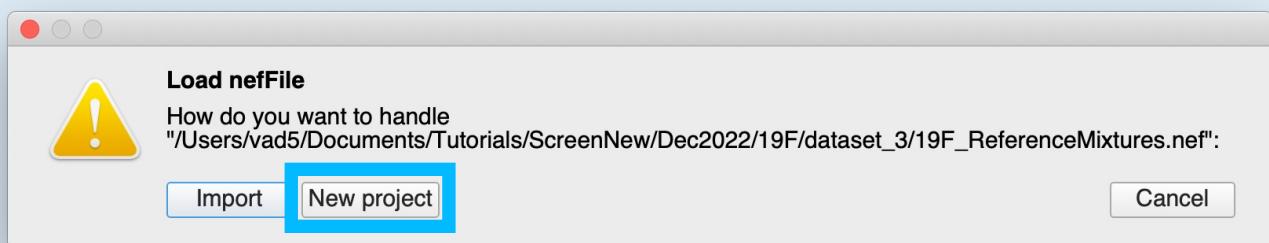
*Project tree:*

- PeakLists
- Samples
- Substances
- SpectrumGroups

- Save to your local disk

Note that you can join the elements in **Sections 7B-E** and **7H** into a single macro, as has been done in **ScreenTutorial/19F/dataset\_3/macros/PrepareProjectForNEFExport.py**; leaving you only with having to check and annotate your peaks before doing the NEF export.

# Import Data from Excel and NEF



## 8A Import From NEF

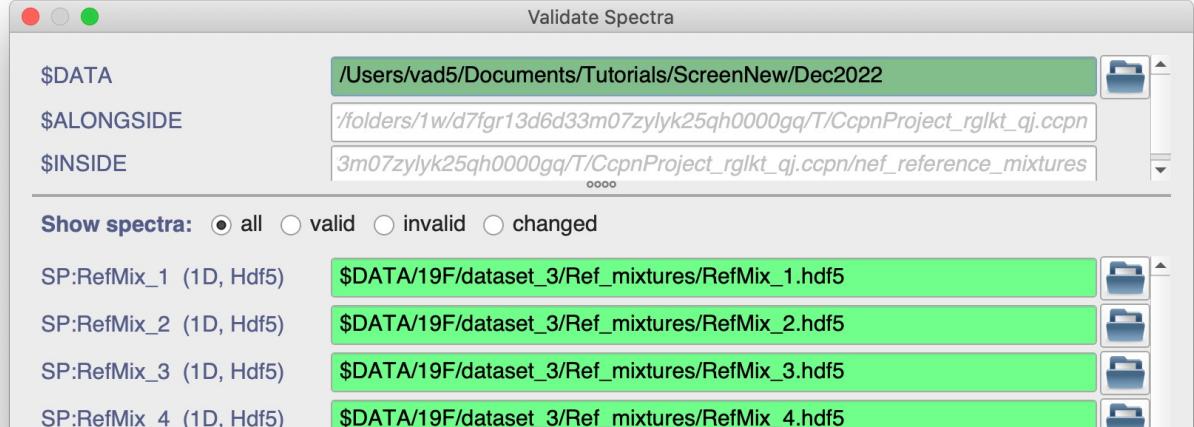
The NEF file contains information about your reference mixtures, including peak lists, samples and substances as well as some formatting.

- Select the **19F\_ReferenceMixtures.nef** located in the **CcpnTutorialData ScreeningJan2023/19F/dataset\_3/** directory and drag it onto the sidebar.
- When prompted, select to open as **New Project**.

Because the NEF file originated from a different computer, the spectrum file paths will be incorrect. To correct them:

- Go to **Main Menu → Spectrum → Validate Paths...**
- Change the **\$DATA** path the Tutorial Data Path on your computer (either by typing or by clicking on the Folder icon ).

Once the paths are correct, they should turn green:



# Import Data from Excel and NEF

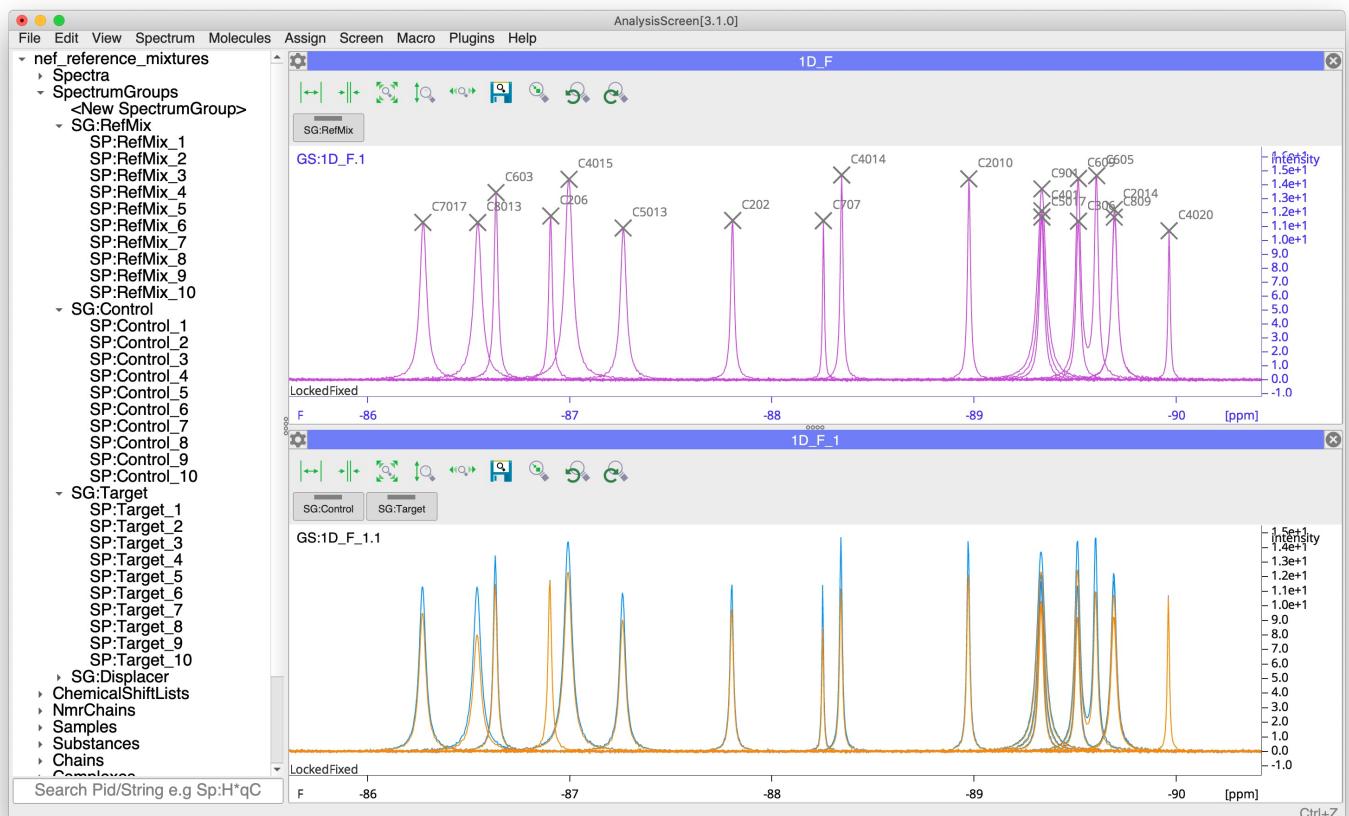
| A1   | sampleName | spectrumGroupName | spectrumPath | spectrumName | experimentType | spectrumHexColour | comment | sampleComponents | pH  | ion |
|------|------------|-------------------|--------------|--------------|----------------|-------------------|---------|------------------|-----|-----|
| 2 0  | Control_1  | Control           | Control_1    | Control_1    | 19F            | #1E90FF           |         |                  | 8   |     |
| 3 1  | Control_2  | Control           | Control_2    | Control_2    | 19F            | #1E90FF           |         |                  | 6   |     |
| 4 2  | Control_3  | Control           | Control_3    | Control_3    | 19F            | #1E90FF           |         |                  | 7   |     |
| 5 3  | Control_4  | Control           | Control_4    | Control_4    | 19F            | #1E90FF           |         |                  | 7   |     |
| 6 4  | Control_5  | Control           | Control_5    | Control_5    | 19F            | #1E90FF           |         |                  | 9   |     |
| 7 5  | Control_6  | Control           | Control_6    | Control_6    | 19F            | #1E90FF           |         |                  | 6   |     |
| 8 6  | Control_7  | Control           | Control_7    | Control_7    | 19F            | #1E90FF           |         |                  | 7   |     |
| 9 7  | Control_8  | Control           | Control_8    | Control_8    | 19F            | #1E90FF           |         |                  | 7   |     |
| 10 8 | Control_9  | Control           | Control_9    | Control_9    | 19F            | #1E90FF           |         |                  | 7.5 |     |
| 11 9 | Control_10 | Control           | Control_10   | Control_10   | 19F            | #1E90FF           |         |                  | 9   |     |
| 12   |            |                   |              |              |                |                   |         |                  |     |     |
| 13   |            |                   |              |              |                |                   |         |                  |     |     |

## 8B Import Data from Excel

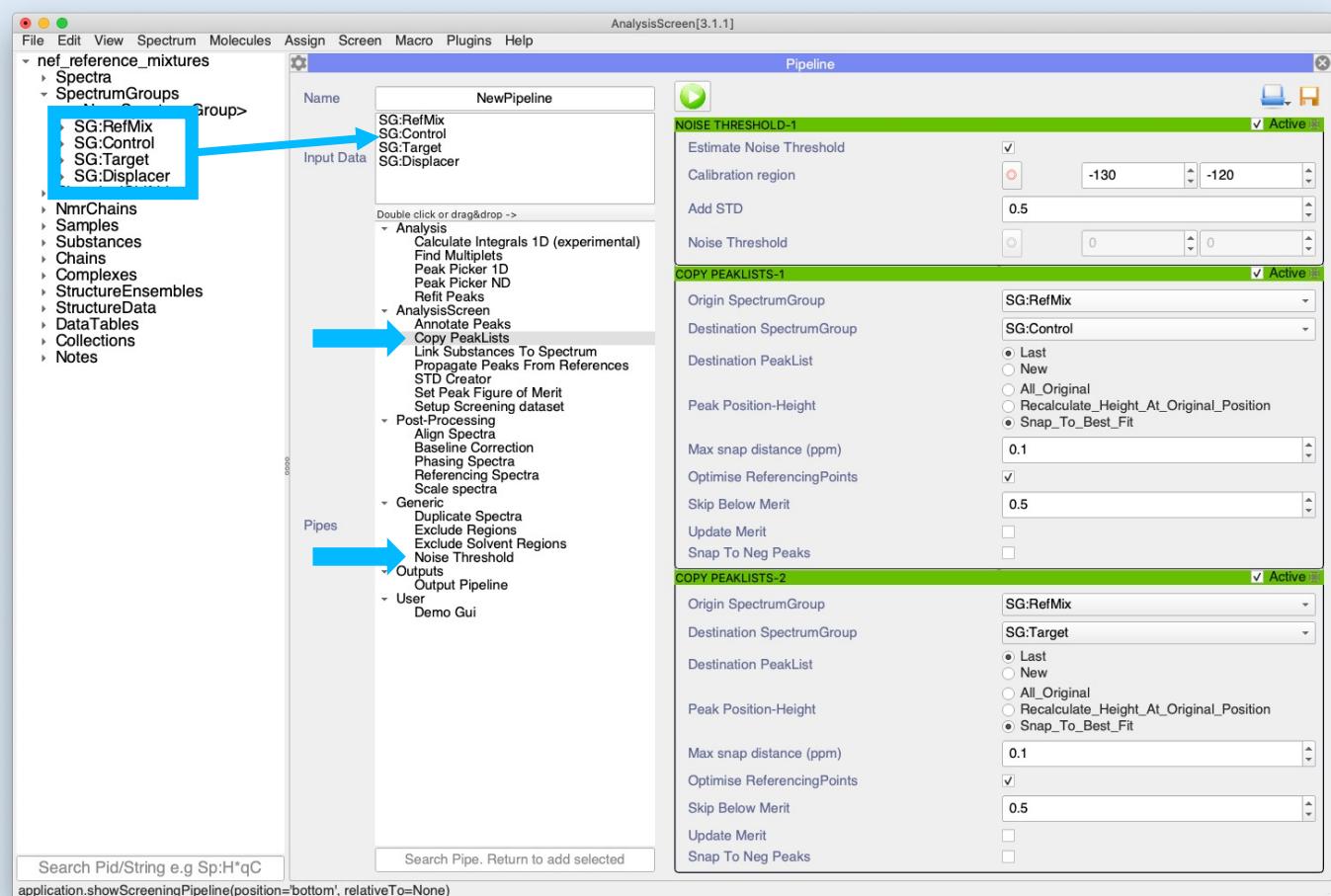
Import the new screening data from the Excel file

- Load the Excel file from the dataset\_3 tutorial data folder: CcpnTutorialData ScreeningJan2023/19F/dataset\_3/Data\_TimeX/lookup\_19F\_TimeX.xlsx

This lookup only contains the **Sample** sheets for the Control, Target and Displacer data without the **SampleComponents** field and the **Substances** sheet. This information has already been imported into the project from the NEF file.



# Automated Peak Matching from Reference Mixtures



## 9A Load Data

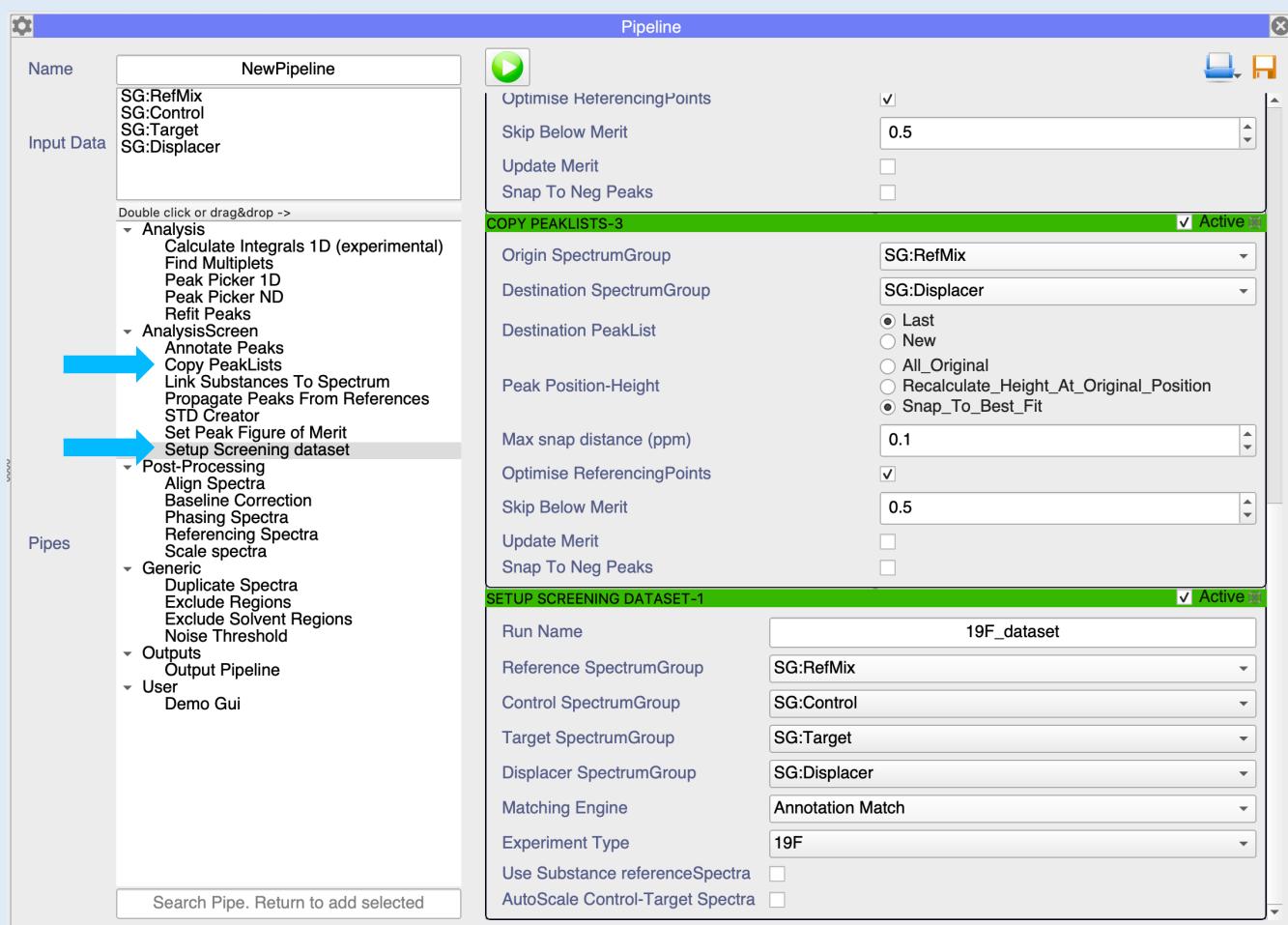
Import your data as shown in **Section 8**.

## 9B Setup screening pipeline

- Open the Pipeline module (**Menu → Screen → Pipeline** or shortcut PI)
- In the sidebar, multiselect SpectrumGroups **SG:RefMix**, **SG:Control**, **SG:Target** and **SG:Displacer**
- 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**
  - 3. Copy PeakLists (2)**
    - Origin SpectrumGroup: **SG:RefMix**
    - Destination SpectrumGroup: **SG:Target**

*... continued*

# Automated Peak Matching from Reference Mixtures



*...continued*

add the following pipes

#### 4. Copy PeakLists (3)

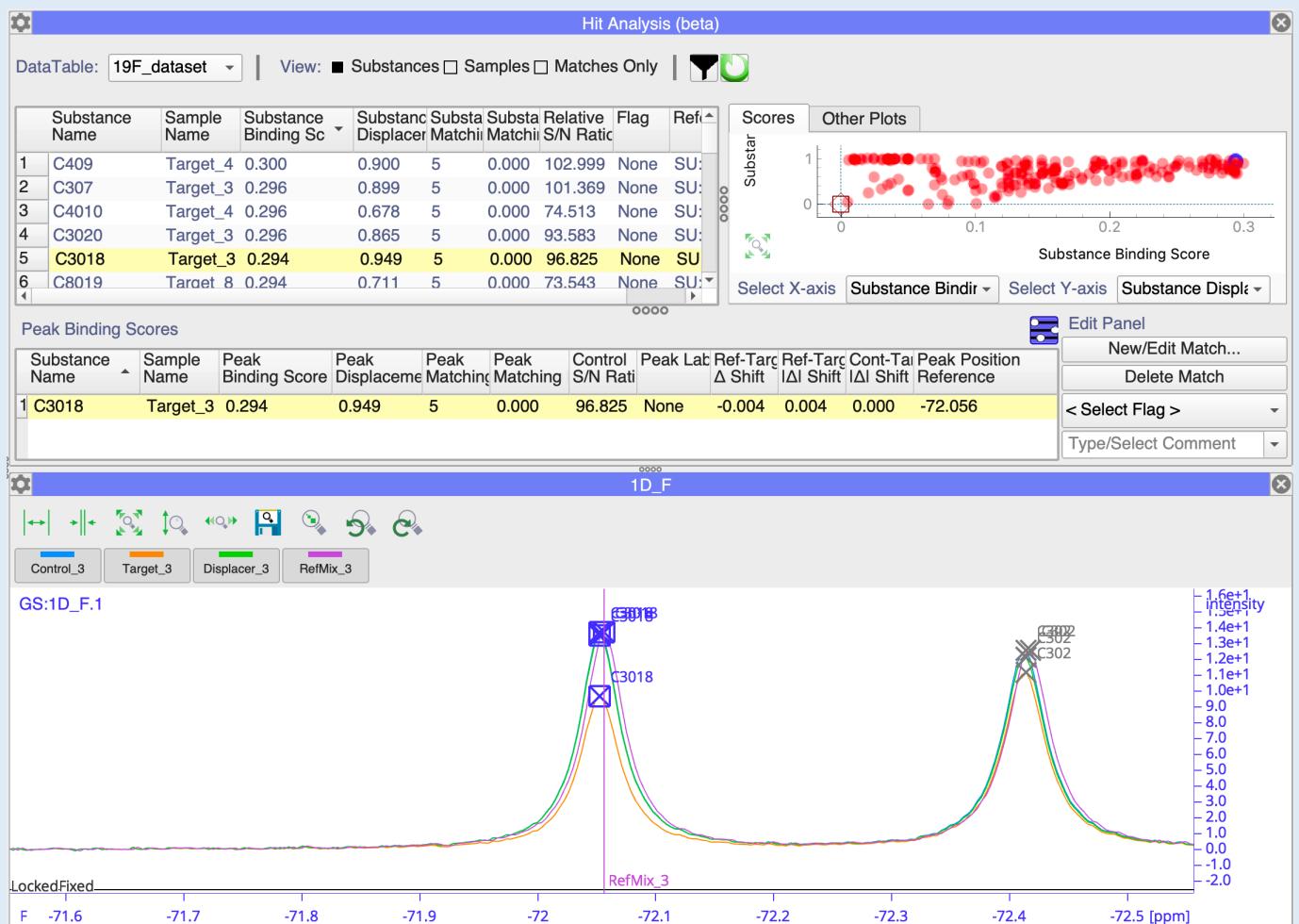
- Origin SpectrumGroup: **SG:RefMix**
- Destination SpectrumGroup: **SG:Displacer**

#### 5. Setup ScreenDataset

- Run name: **19F\_dataset**
- Reference SpectrumGroup: **SG:RefMix**
- Select **SG:Control**, **SG:Target** and **SG:Displacer** for their respective entries
- Matching engine: **Annotation Match**
- untick **Use\_Substance\_ReferenceSpectra**

- Run the pipeline.
- Inspect the results in the Hit Analysis Module as shown in Sections 4–6.

# Automated Peak Matching from Reference Mixtures

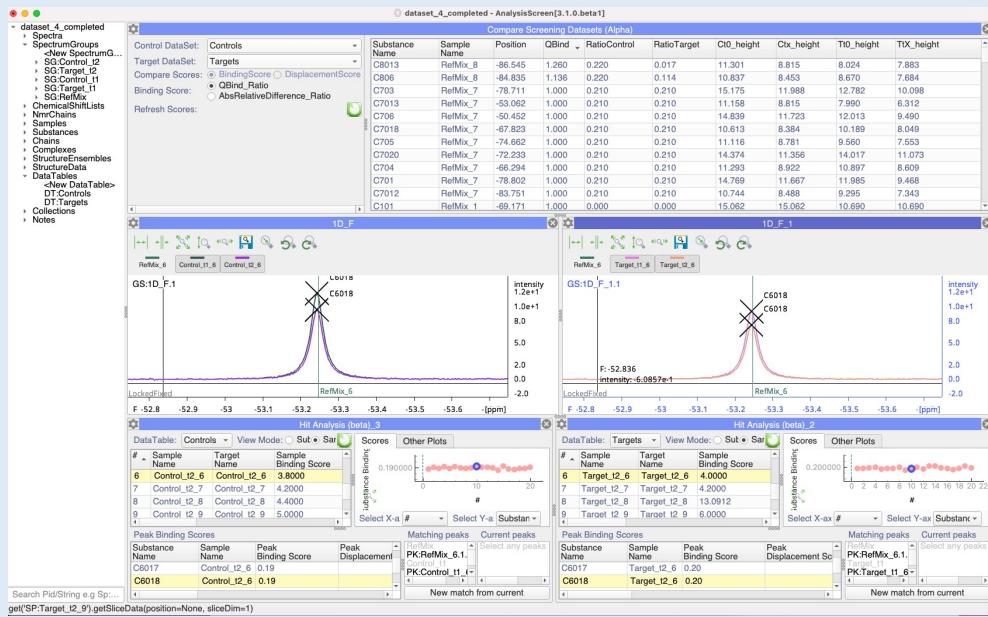


## 9C Completed Project

An example of a completed project is available in the dataset\_3 directory:  
`.../19F/dataset_3/dataset3_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.



## 10A Open data in program

In this section you will compare two screens performed on the same library at different CPMG times.

You can now either recreate the project in **Sections 10B-D** or load a project which has already two completed screening datasets and skip to **Section 10E**:

- Drag and drop the **CcpnTutorialDataScreeningJan2023/19F/ dataset\_4/dataset\_4\_completed.ccpn** folder in the sidebar or Drop Area

## 10B Setup two parallel screens (optional)

From the **dataset\_4** folder:

- Load the NEF file, as a new project.
- Load the Excel files in the directory **Data\_Time\_1** and **Data\_Time\_2**.

## 10C Copy peakLists (optional – see figure on next page)

- Open the pipeline module (**PI**), add all Spectrum Groups as input data.

Use the **Copy Peak Lists** pipe:

- Origin SpectrumGroup: **SG:RefMix**
- Destination SpectrumGroup: **SG:Control\_t1**
- Keep all settings as default.

Repeat this pipe for the other spectrumGroups:

- **SG:RefMix** → **SG:Target\_t1**
- **SG:RefMix** → **SG:Control\_t2**
- **SG:RefMix** → **SG: Target \_t2** (see image next page)

Run the pipeline, once completed close all pipes and keep the pipeline opened.

# 10 Using data with multiple CPMG delays

**COPY PEAKLISTS-1** ✓ Active

|                            |                                                                                                                                                                                                                                |
|----------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Origin SpectrumGroup       | SG:RefMix                                                                                                                                                                                                                      |
| Destination SpectrumGroup  | SG:Control_t1                                                                                                                                                                                                                  |
| Destination PeakList       | <input checked="" type="radio"/> Last<br><input type="radio"/> New<br><input type="radio"/> All_Original<br><input type="radio"/> Recalculate_Height_At_Original_Position<br><input checked="" type="radio"/> Snap_To_Best_Fit |
| Peak Position-Height       | 0.1                                                                                                                                                                                                                            |
| Max snap distance (ppm)    | 0.1                                                                                                                                                                                                                            |
| Optimise ReferencingPoints | <input checked="" type="checkbox"/>                                                                                                                                                                                            |
| Skip Below FoM             | 0.5                                                                                                                                                                                                                            |

**COPY PEAKLISTS-2** ✓ Active

|                            |                                                                                                                                                                                                                                |
|----------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Origin SpectrumGroup       | SG:RefMix                                                                                                                                                                                                                      |
| Destination SpectrumGroup  | SG:Target_t1                                                                                                                                                                                                                   |
| Destination PeakList       | <input checked="" type="radio"/> Last<br><input type="radio"/> New<br><input type="radio"/> All_Original<br><input type="radio"/> Recalculate_Height_At_Original_Position<br><input checked="" type="radio"/> Snap_To_Best_Fit |
| Peak Position-Height       | 0.1                                                                                                                                                                                                                            |
| Max snap distance (ppm)    | 0.1                                                                                                                                                                                                                            |
| Optimise ReferencingPoints | <input checked="" type="checkbox"/>                                                                                                                                                                                            |
| Skip Below FoM             | 0.5                                                                                                                                                                                                                            |

**COPY PEAKLISTS-3** ✓ Active

|                            |                                                                                                                                                                                                                                |
|----------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Origin SpectrumGroup       | SG:RefMix                                                                                                                                                                                                                      |
| Destination SpectrumGroup  | SG:Control_t2                                                                                                                                                                                                                  |
| Destination PeakList       | <input checked="" type="radio"/> Last<br><input type="radio"/> New<br><input type="radio"/> All_Original<br><input type="radio"/> Recalculate_Height_At_Original_Position<br><input checked="" type="radio"/> Snap_To_Best_Fit |
| Peak Position-Height       | 0.1                                                                                                                                                                                                                            |
| Max snap distance (ppm)    | 0.1                                                                                                                                                                                                                            |
| Optimise ReferencingPoints | <input checked="" type="checkbox"/>                                                                                                                                                                                            |
| Skip Below FoM             | 0.5                                                                                                                                                                                                                            |

**COPY PEAKLISTS-4** ✓ Active

|                            |                                                                                                                                                                                                                                |
|----------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Origin SpectrumGroup       | SG:RefMix                                                                                                                                                                                                                      |
| Destination SpectrumGroup  | SG:Target_t2                                                                                                                                                                                                                   |
| Destination PeakList       | <input checked="" type="radio"/> Last<br><input type="radio"/> New<br><input type="radio"/> All_Original<br><input type="radio"/> Recalculate_Height_At_Original_Position<br><input checked="" type="radio"/> Snap_To_Best_Fit |
| Peak Position-Height       | 0.1                                                                                                                                                                                                                            |
| Max snap distance (ppm)    | 0.1                                                                                                                                                                                                                            |
| Optimise ReferencingPoints | <input checked="" type="checkbox"/>                                                                                                                                                                                            |
| Skip Below FoM             | 0.5                                                                                                                                                                                                                            |

**SETUP SCREENING DATASET-1** ✓ Active

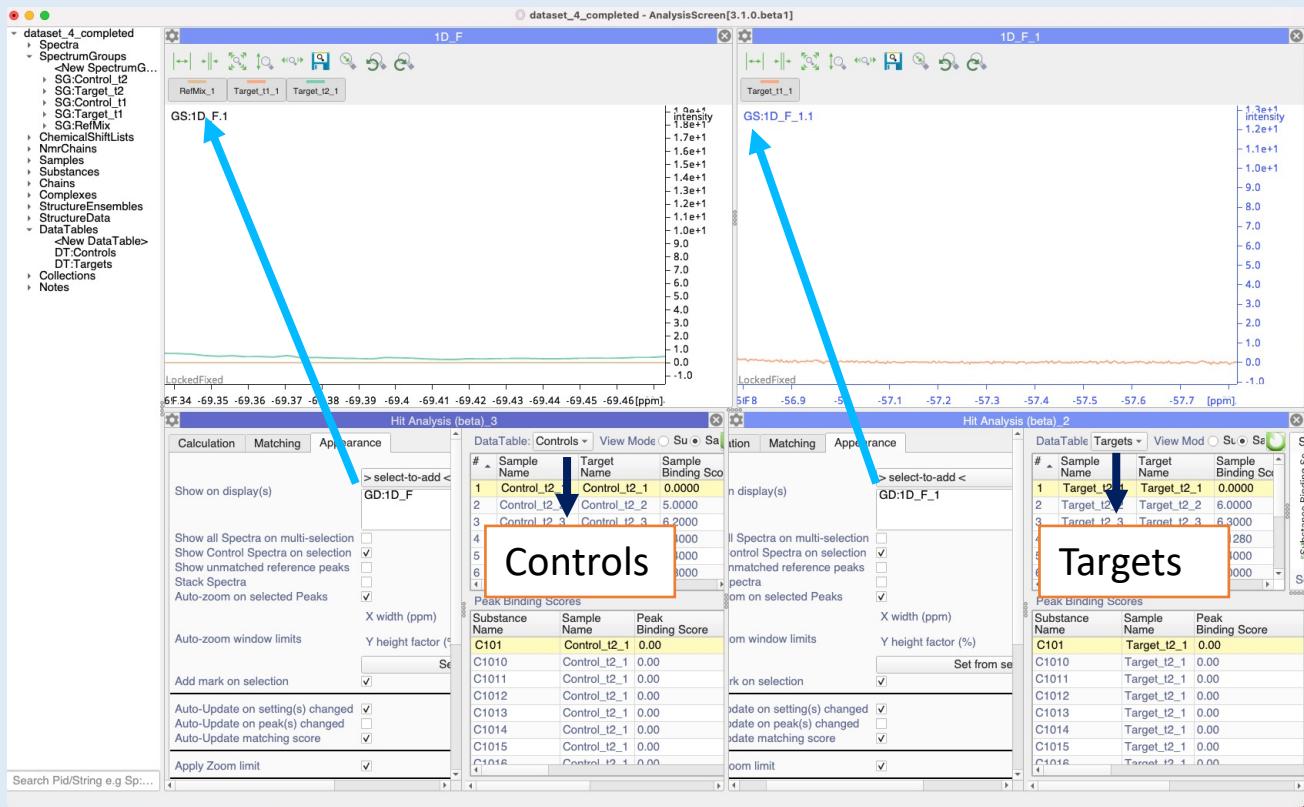
|                                  |                          |
|----------------------------------|--------------------------|
| Run Name                         | Controls                 |
| Reference SpectrumGroup          | SG:RefMix                |
| Control SpectrumGroup            | SG:Control_t1            |
| Target SpectrumGroup             | SG:Control_t2            |
| Displacer SpectrumGroup          | None                     |
| Matching Engine                  | Annotation Match         |
| Experiment Type                  | 19F                      |
| Use Substance referenceSpectra   | <input type="checkbox"/> |
| AutoScale Control-Target Spectra | <input type="checkbox"/> |

**SETUP SCREENING DATASET-2** ✓ Active

|                                  |                          |
|----------------------------------|--------------------------|
| Run Name                         | Targets                  |
| Reference SpectrumGroup          | SG:RefMix                |
| Control SpectrumGroup            | SG:Target_t1             |
| Target SpectrumGroup             | SG:Target_t2             |
| Displacer SpectrumGroup          | None                     |
| Matching Engine                  | Annotation Match         |
| Experiment Type                  | 19F                      |
| Use Substance referenceSpectra   | <input type="checkbox"/> |
| AutoScale Control-Target Spectra | <input type="checkbox"/> |

## 10D Setup Screening Datasets (optional)

- Add the **Set up Screening dataset** pipe twice.
- Set the first pipe up for the Control data (as shown above):
  - Run Name: **Controls**
  - Reference SpectrumGroup: **SG:RefMix**
  - Control SpectrumGroup: **SG:Control\_t1**
  - Target SpectrumGroup: **SG: Control\_t2** (not the target!)
  - Displacer SpectrumGroup: **None**
  - Matching Engine: **Annotation Match**
  - Experiment type: **19F**
  - Untick the last two checkboxes.
- Repeat for the Target data with **Targets**, **SG:Target \_t1** and **SG:Target \_t2** for the Run Name, Control SpectrumGroup and Target SpectrumGroup, respectively.
- Run the pipeline and then close the module.



## 10E

## Set up displays and HitAnalysis Modules

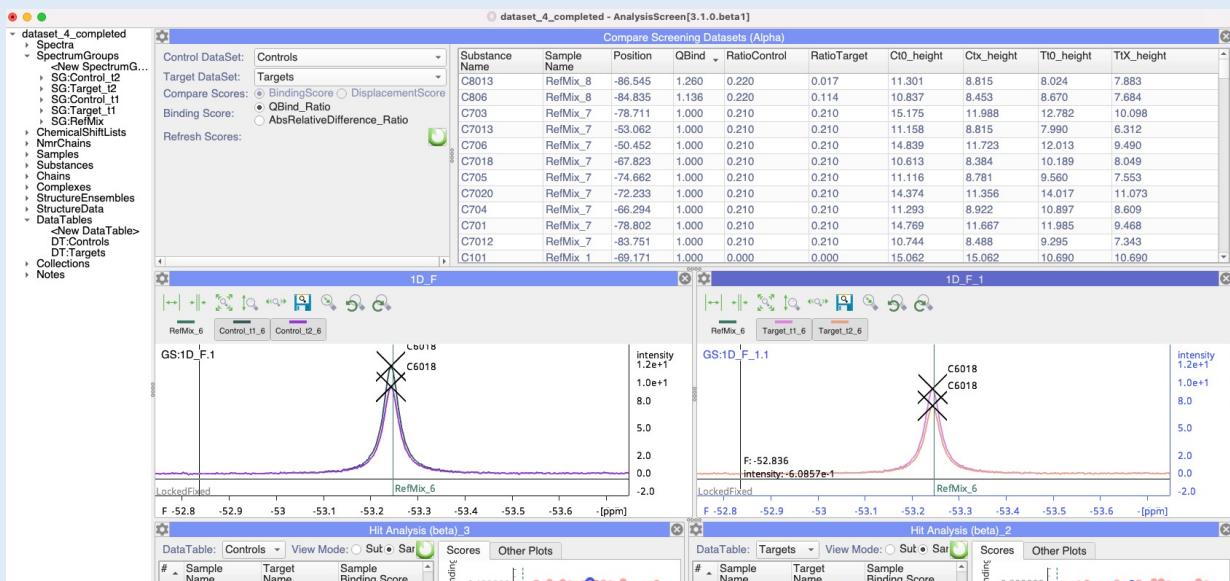
- Open two SpectrumDisplays and two HitAnalysis Modules as shown above.

In the Hit Analysis modules:

- Set the **DataTable** to **Controls** on the left and to **Targets** on the right.
- Open the **Settings** panels: Set **Show on display(s)** to **GD:1D\_F** on the left and to **GD:1D\_F\_1** on the right.

You will not need to apply any other settings in the Hit Analysis modules. You might like to reduce their size to a minimum height.

# 10 Using data with multiple CPMG delays



## $Q^{\text{bind}}$ score

One way of comparing multiple CPMG times is by using the  $Q^{\text{bind}}$  score\*

$$Q^{\text{bind}} = \frac{\text{Intensity Ratio} + ^{\text{Target}}}{\text{Intensity Ratio} ^{\text{Control}}}$$

with:

$$\text{Intensity Ratio} = \frac{\text{Peak Intensity}^{\text{time } x}}{\text{Peak Intensity}^{\text{time } 0}}$$

A lower value such as  $Q^{\text{bind}} < 0.32$  may reflect a strong binding event; values between 0.33–0.66 a medium binding, whereas  $Q^{\text{bind}} > 0.67$  may indicate no interaction between the target and the molecule.

\* 19F NMR-Based Fragment Screening for 14 Different Biologically Active RNAs and 10 DNA and Protein Counter-Screens. Binas et al. *ChemBioChem* 2021, 22, 423 – 433. doi.org/10.1002/cbic.202000476

## 10F Open Compare Screening Datasets

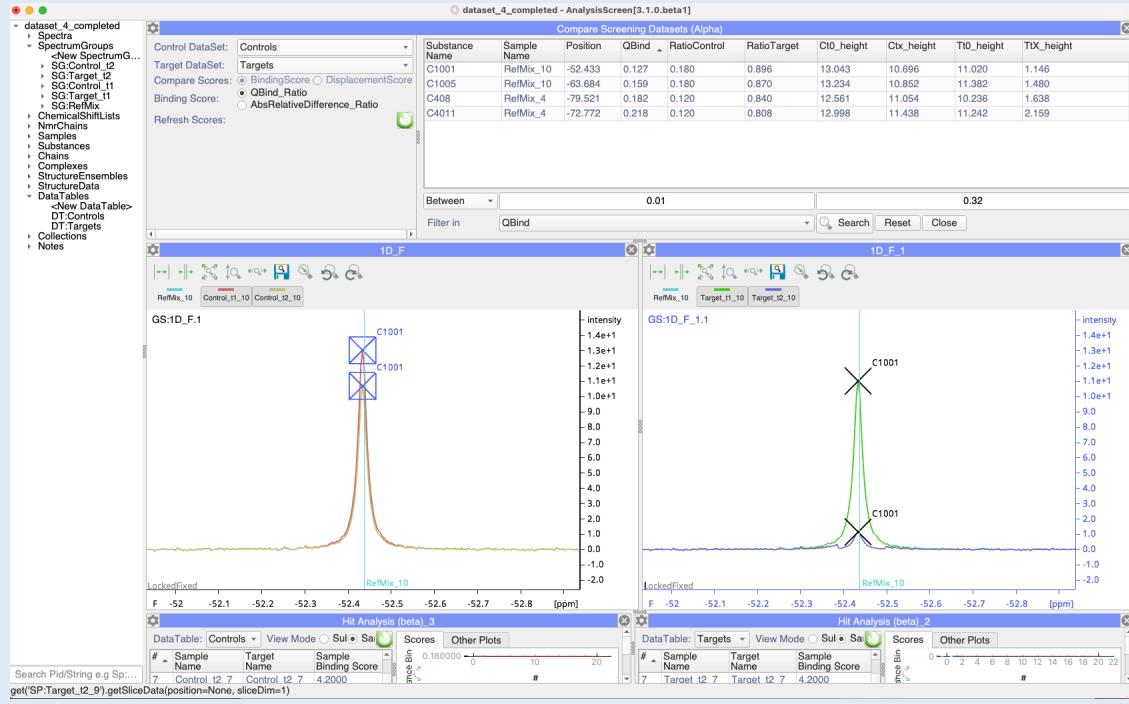
If not already present, open the module **Compare Screens** from the main menu:

- Main Menu → Screen → Compare Screens.
- Open the Settings panel, select the **Control** and **Target** datasets and select  **$Q^{\text{bind}}$  Ratio** as the **Binding Score**.
- Click the refresh button to populate the table.
- On the table, filter by the **QBind** score to identify potential strong binders:
- Right click the header and click **Filter...** (or use shortcut **FT**).

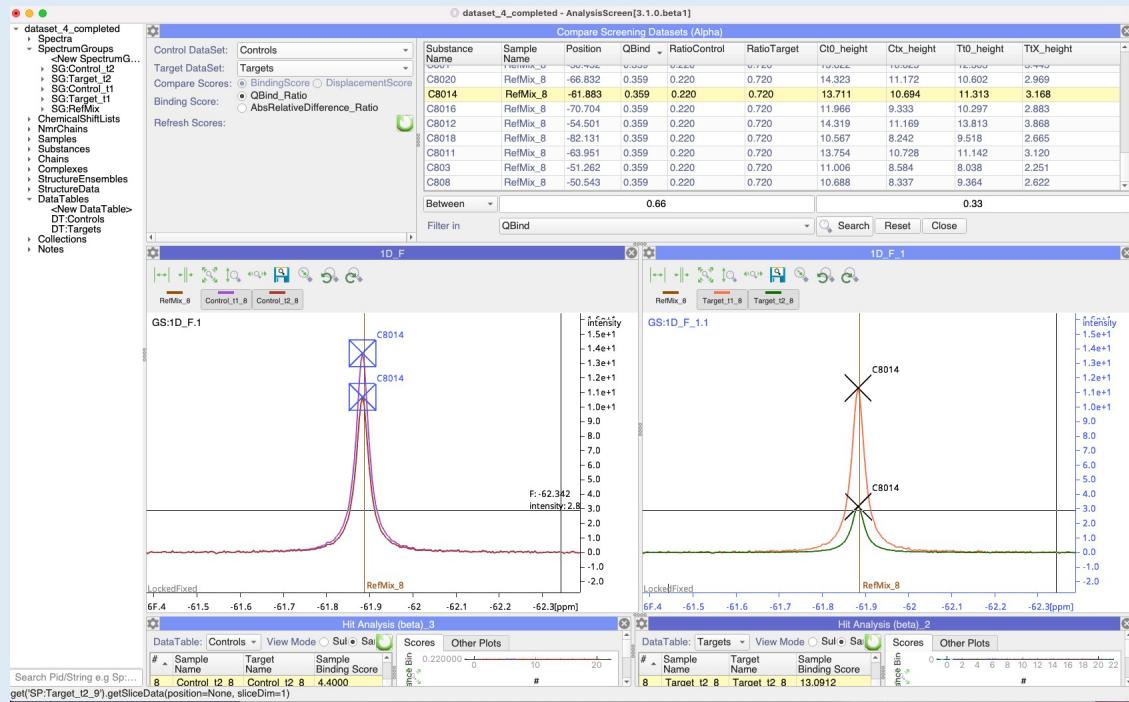
Filter between 0.01 and 0.32 in **Qbind**.

- Click **Search**.

The table will now display only a few entries. Inspect the spectra by selecting the row. The displays will navigate accordingly to the peaks.



**Strong Binders**  
QBind < 0.32



**Medium Binders**  
QBind 0.33-0.66



**Non Binders**  
QBind > 0.67

## 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://forum.ccpn.ac.uk/>

## 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)