Mestrelab Research

chemistry software solutions

Mnova Software Tools for Fragment-Based Drug Discovery

Chen Peng, PhD,
VP of Business Development, US & China
Mestrelab Research SL
San Diego, CA
chen.peng@mestrelab.com 858.736.4563

Mestrelab Research
chemistry software solutions

Agenda

- Brief intro on fragment-based drug discovery (FBDD)
- The relevant Mnova software tools for
  - QC and solubility test of library compounds and building the reference spectra database.
  - Pooling of compounds with least peak overlap.
  - Batch analysis of 1D ligand-observed screening spectra.
  - Analysis of 2D chemical shift perturbation spectra.
- Demo
- Questions

Introduction: Fragment-based lead discovery using NMR

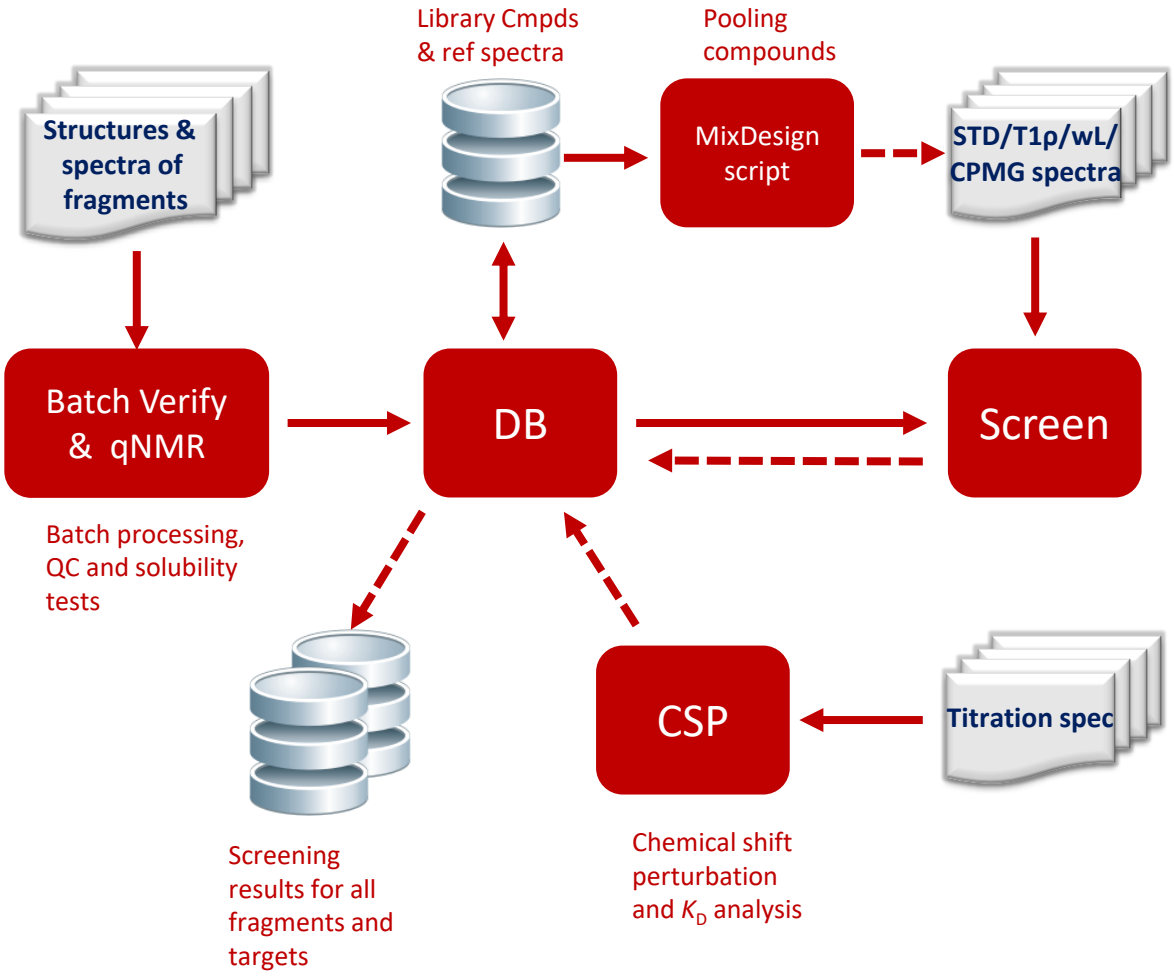
- NMR has been widely used for high-throughput or detailed hit finding and hit validation since mid-1990s
  - Ideally suited for detecting ligand-protein bindings with K_d in µmol-mmol range.
  - "In-built" quality control: structure consistency check, concentration measurement, and binding assessment all from the same sample.
- Ligand-observed NMR binding spectra: commonly used for primary fragment screening, no labeling needed, no size restriction by receptor, ^{1}H or ^{19}F
  - STD (Saturation transfer difference exp.)
  - T1ρ (Relaxation-edited exp.)
  - CPMG (Relaxation-edited exp.)
  - WaterLOGSY (Water-ligand observed via gradient spectroscopy)
- Protein-observed chemical shift perturbation spectra: Residue-specific info, mapping to binding site on protein, K_d measurement, SAR-by-NMR etc.
  - ^{15}N or ^{13}C labeled HSQC spectra of protein.

Mestrelab Research
chemistry software solutions

Large amount of NMR data to process

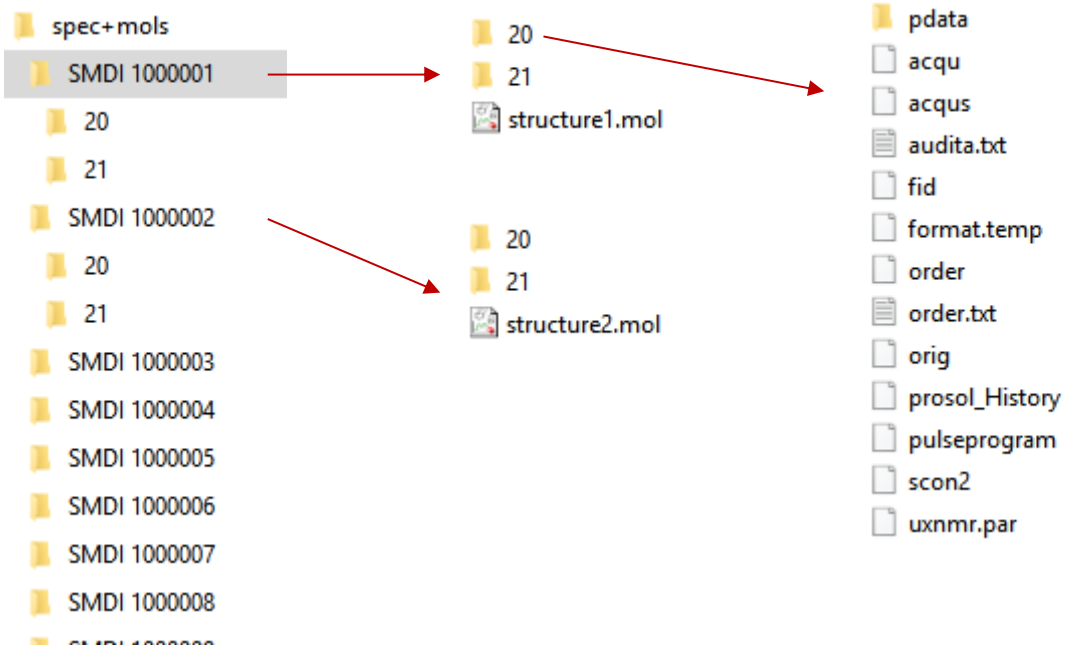
- A typical mid-size compound library: 500-2000 compounds
- ^{1}H detected experiments:
  - Primary screening: 6-12 fragments with min. peak overlap per sample => 50-300 samples per library => primary hits in a few days.
  - Confirmation of hits: single compound samples.
- ^{19}F detected experiments: high sensitivity (low-µm concen.), simple spectra, large $\delta^{19}F$ range
  - Mixtures of 10-30 cmpds per sample => 20-50 samples per library.
- ^{1}H-^{15}N or ^{1}H-^{13}C HSQC of target protein
  - One spectrum for each ligand to compare with the reference spectrum.
  - Or 6-10 titration points per ligand for titration analysis.
- **Hundreds or thousands of spectra to process and analyze: a bottleneck.**

Mnova tools for FBDD



Quality control and databasing of library compounds

- The example here has a dataset organized shown below. Note your data does not have to be exactly like this.



Multiple datasets located under a parent directory "spec+mols"

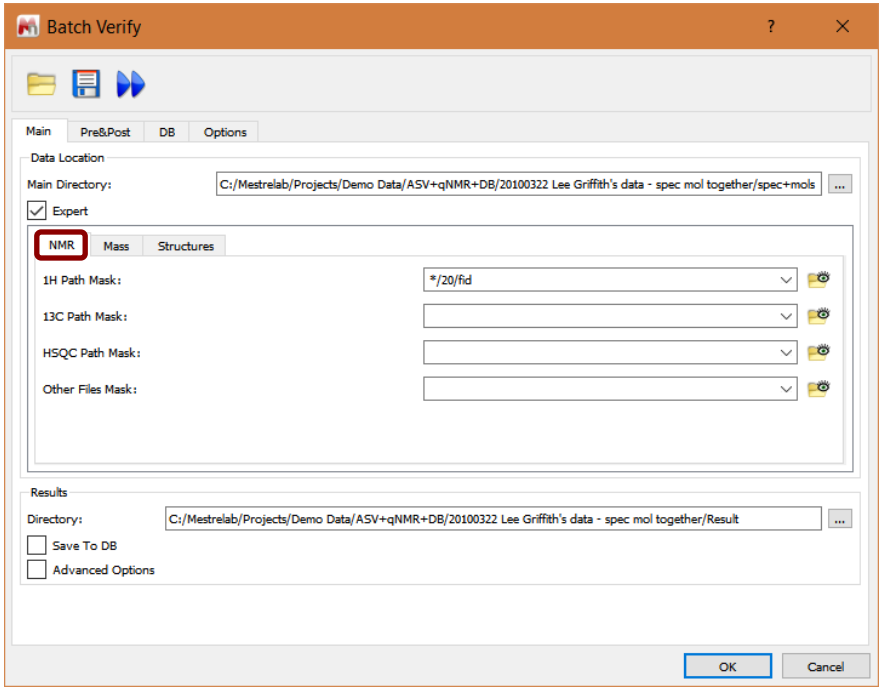
Each dataset has a H-1 (20), HSQC (21), and a molecule file .mol

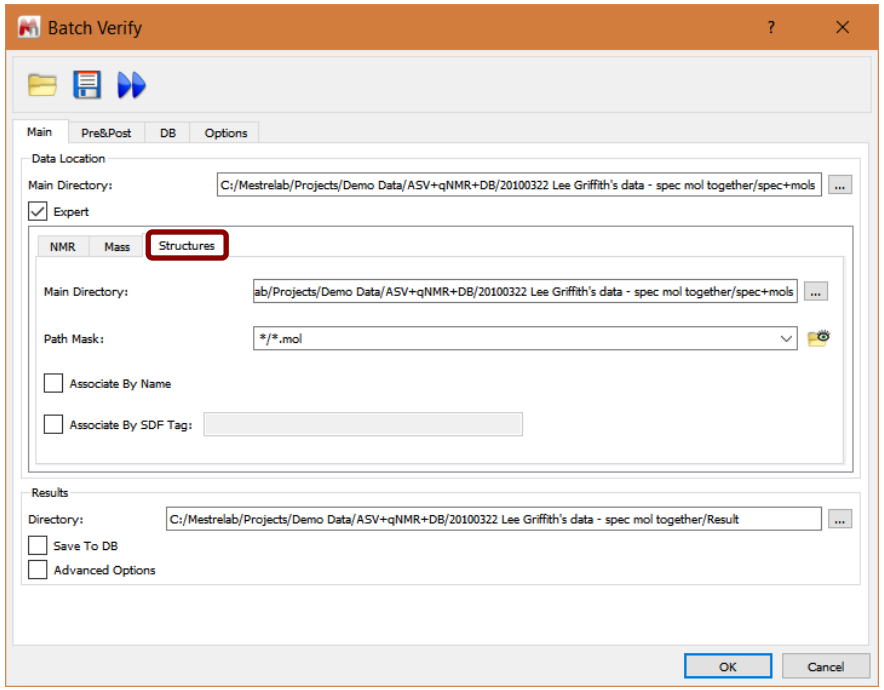
Each H-1 has the typical Bruker files. We will reprocess using the *fid* files

Setup Batch Verify

Once you are ready, start Batch Verify by choosing Analysis | Verification | Batch Verify. The sample data mentioned previously is used as an example here.

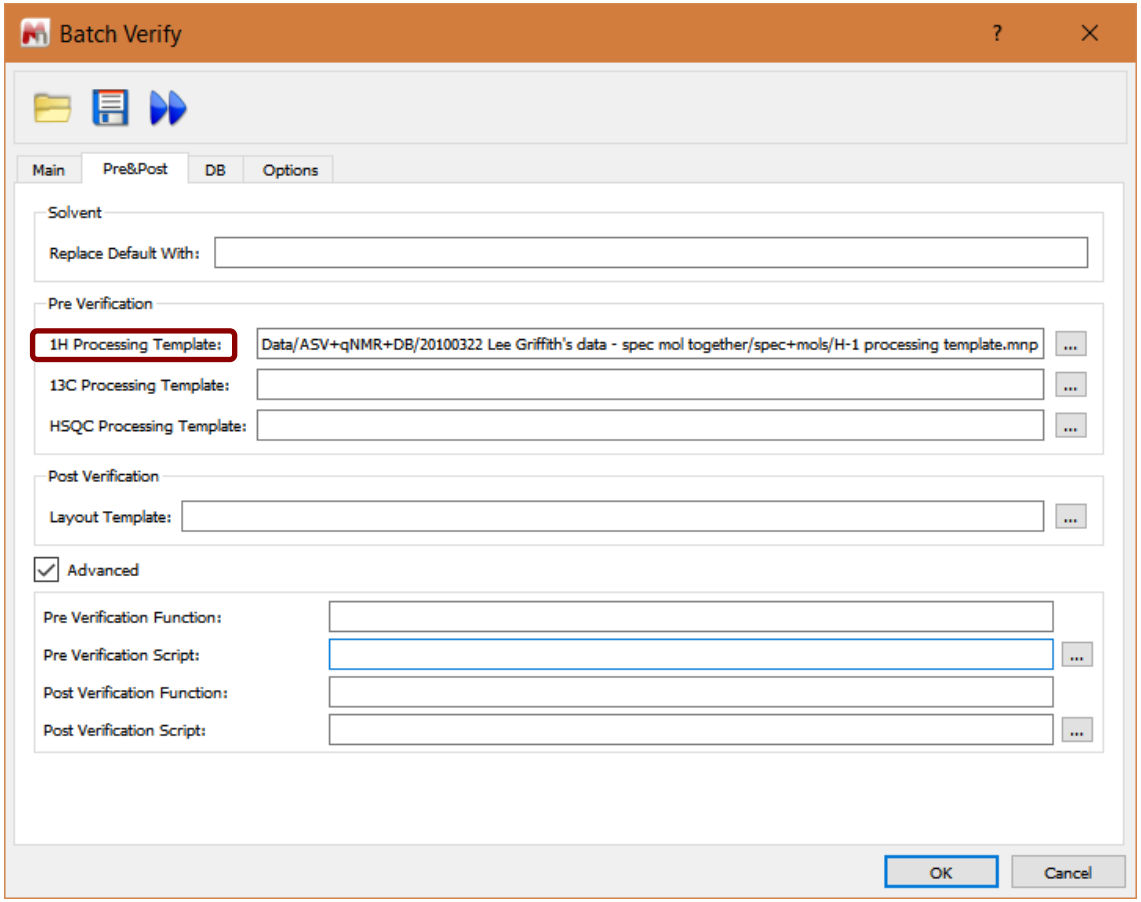
In the Main tab, setup the NMR and structure files to use and the Results folder etc.



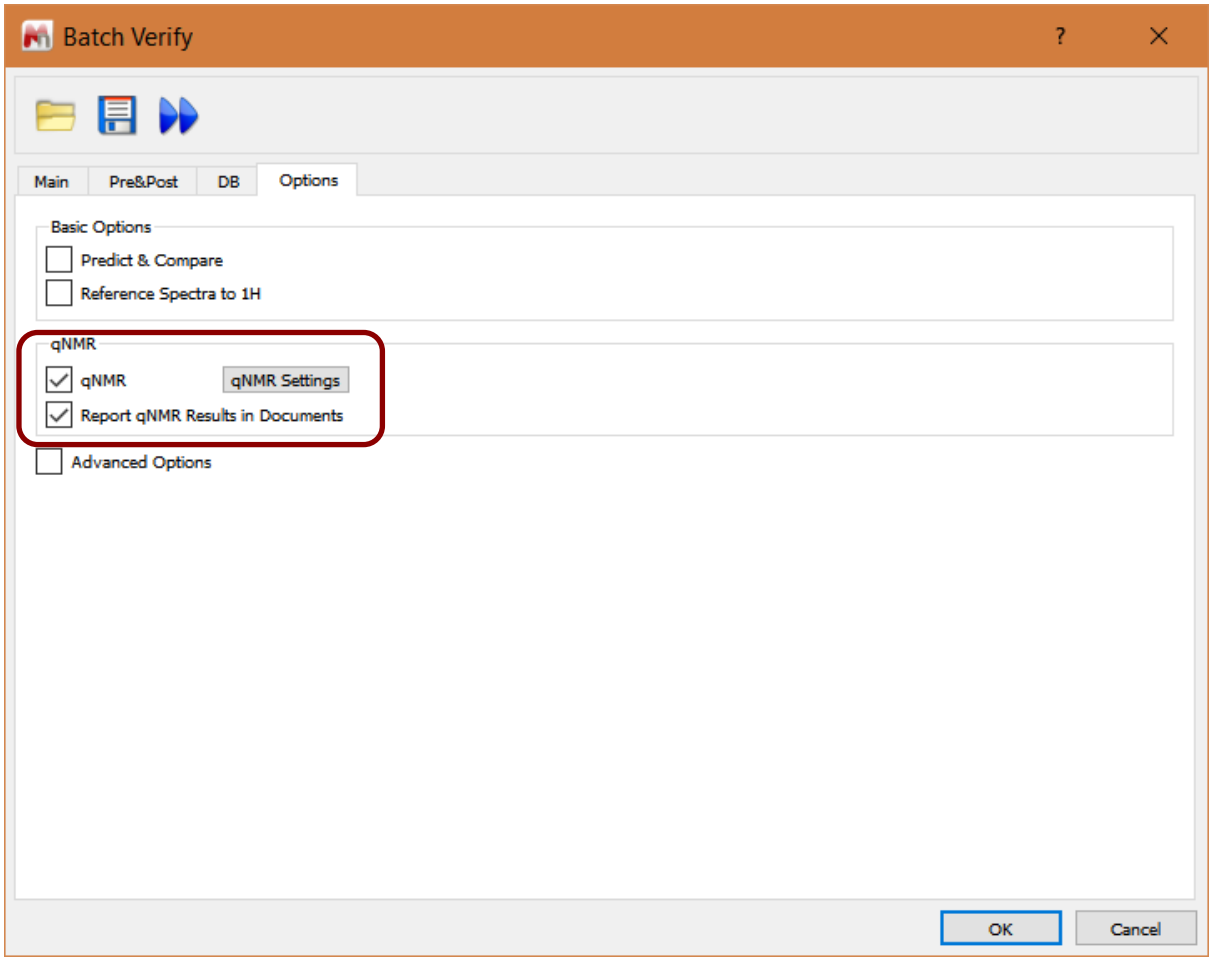


Setup Batch Verify

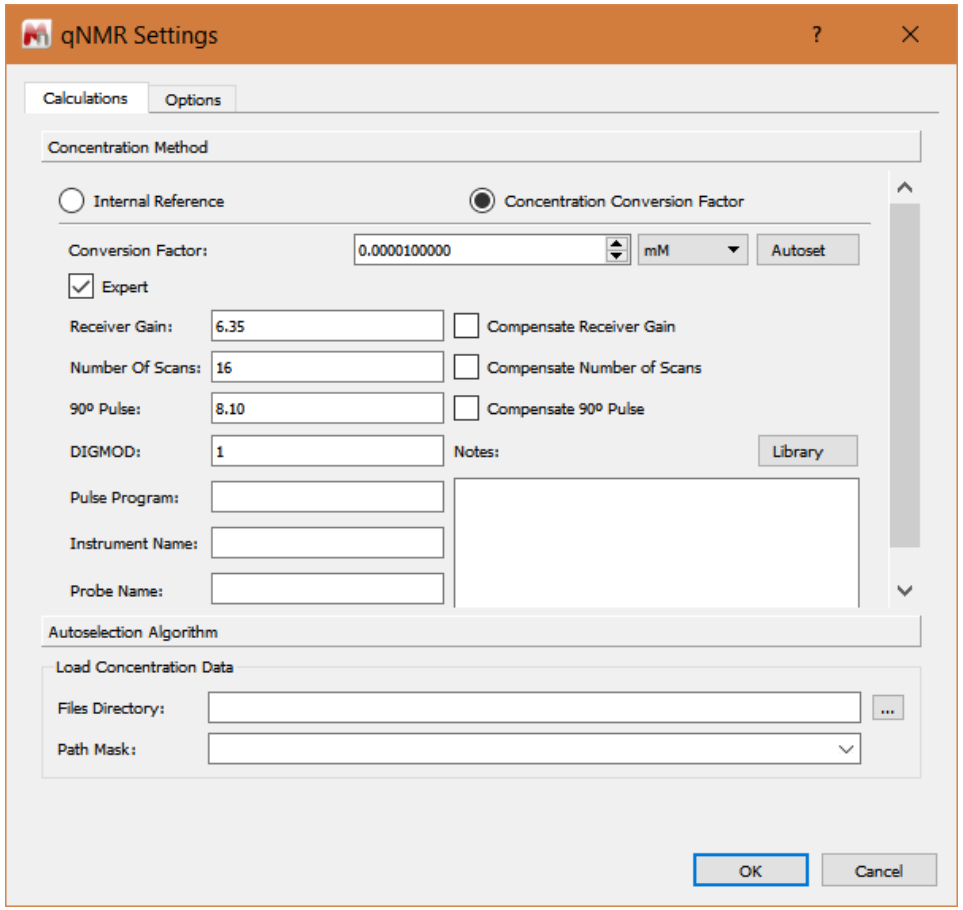
In the Pre&Post Tab, specify the processing template to use for all the H-1 NMR processing:



In the Options Tab, choose to do quantitation (determination of molar concentration using external reference info in this case).

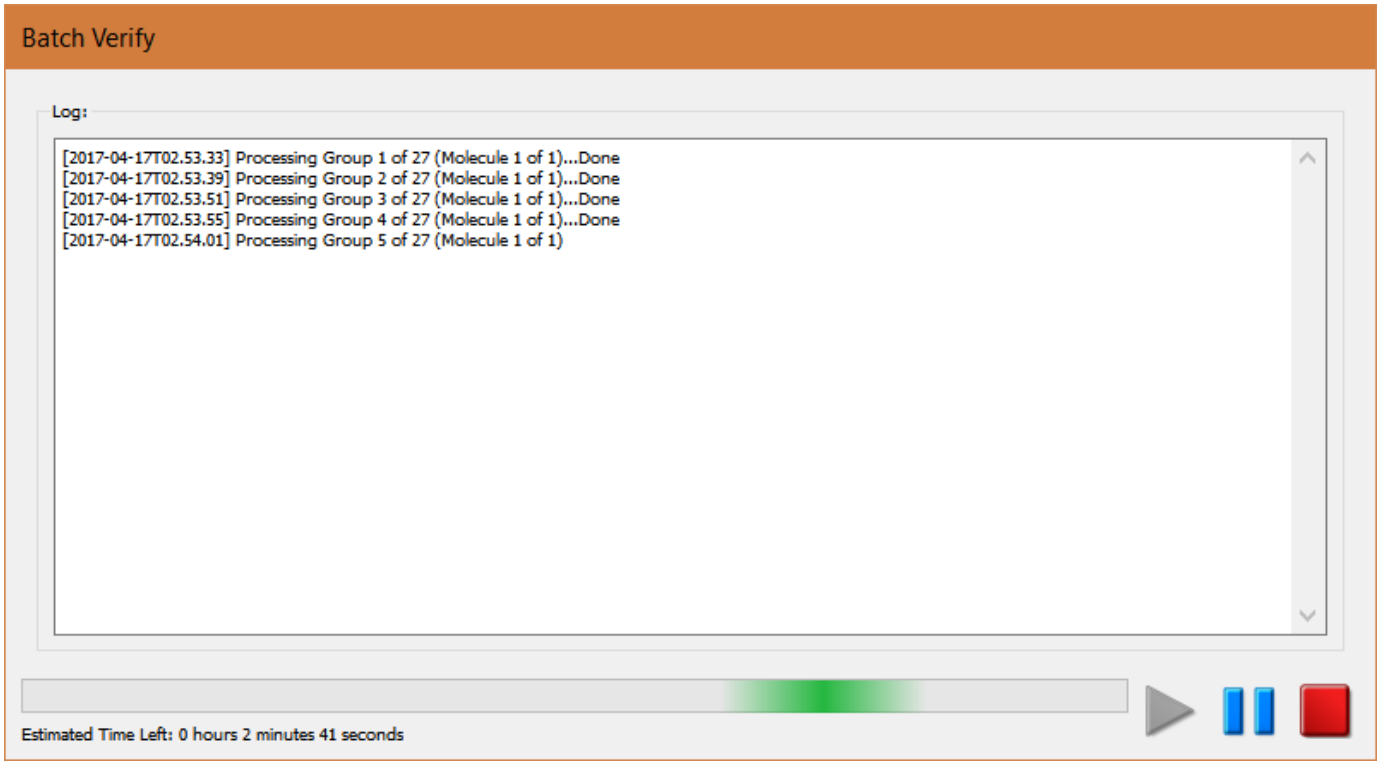


In the Options Tab, click qNMR Settings button to define the details for molar concentration determination.



Run Batch Verify

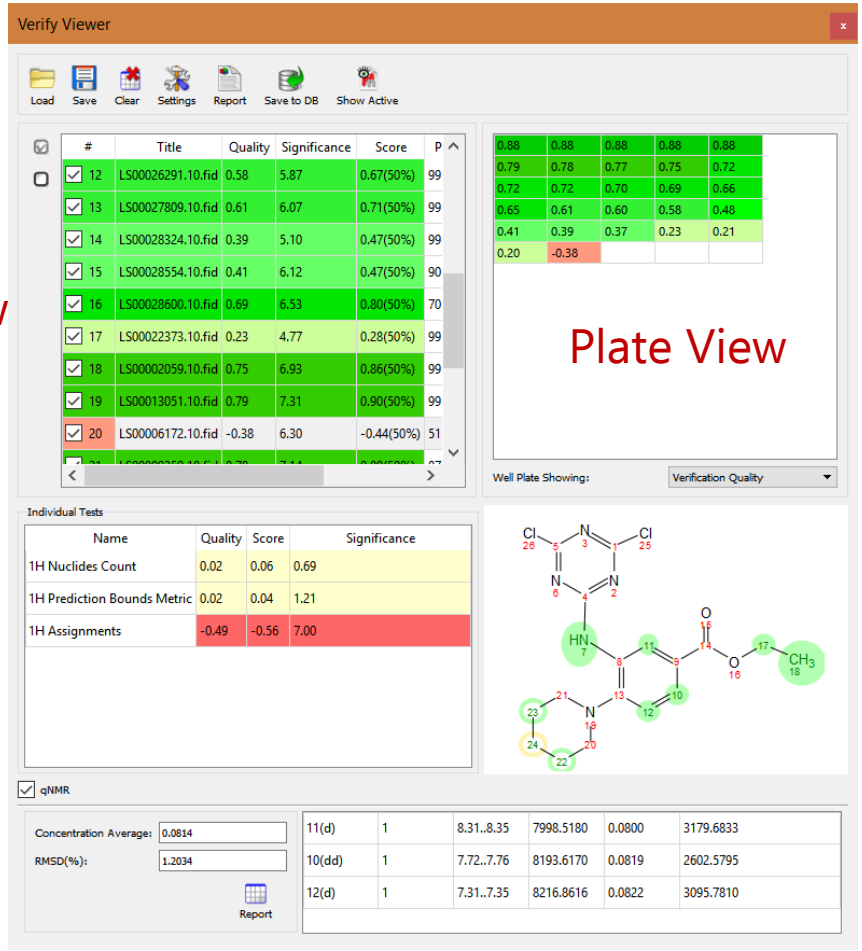
Once the settings are done, click OK to start the batch processing. It process all the spectra, does structure verification and quantitation.



View results from batch processing

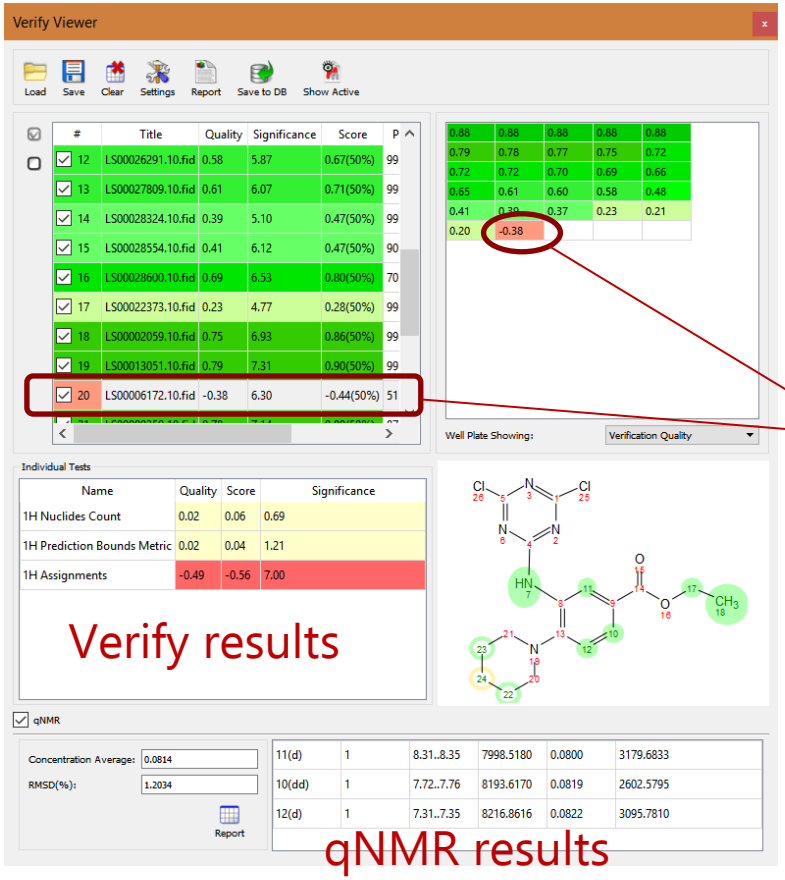
Upon completion, the verification results are written to under the specified Results directory. Choose Analysis | Verification | Verify Viewer, and click the Load button to open the results.dat file. All the results are loaded for visualization:

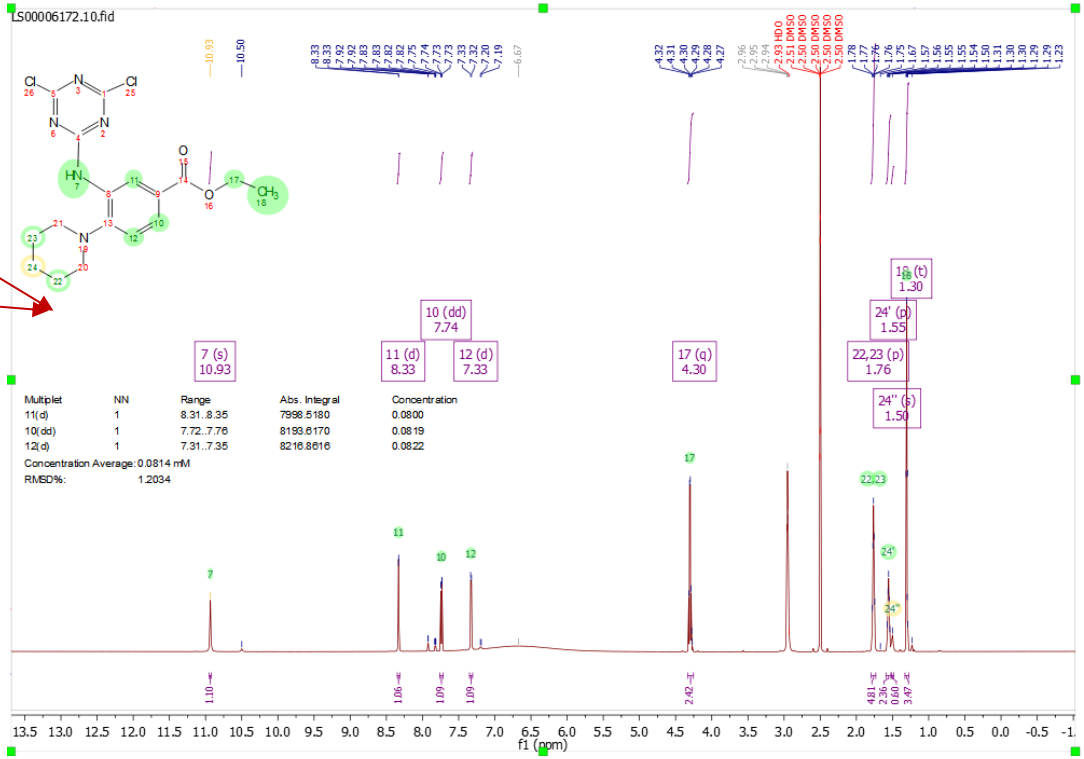
Table View



View results

In the Verify Viewer, click on any items in the Table or Well-plate View to see the details of the spectrum/molecule. Pay attention to the ones with red/yellow flags. You can re-analyze the results (peak picking, multiplet analysis) and apply Verify or qNMR to revise the results for the current spectrum.

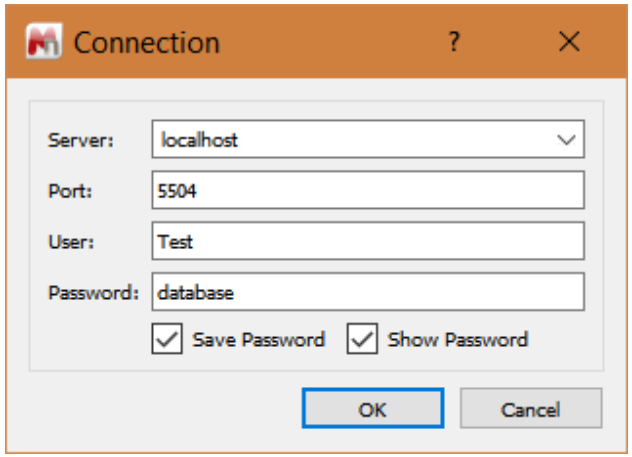




Build a Ref DB

Once you are ready, create a new database to save the spectra and molecules.

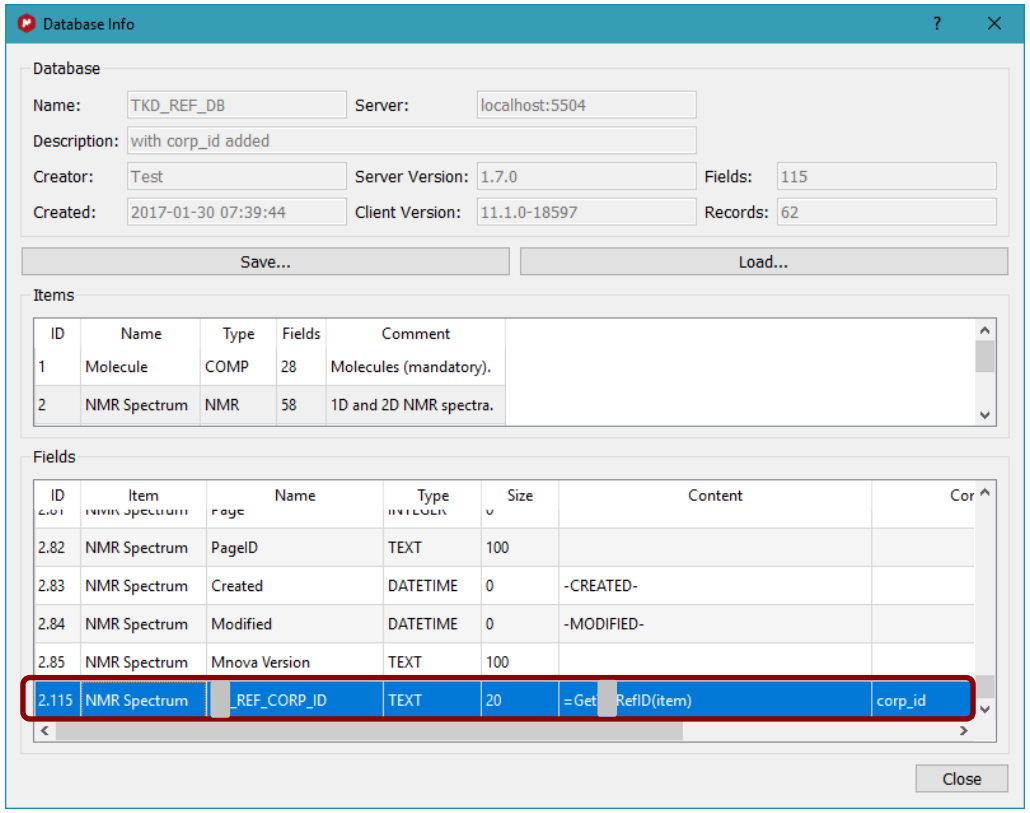
Choose Database | Connect to connect to the DB Server.



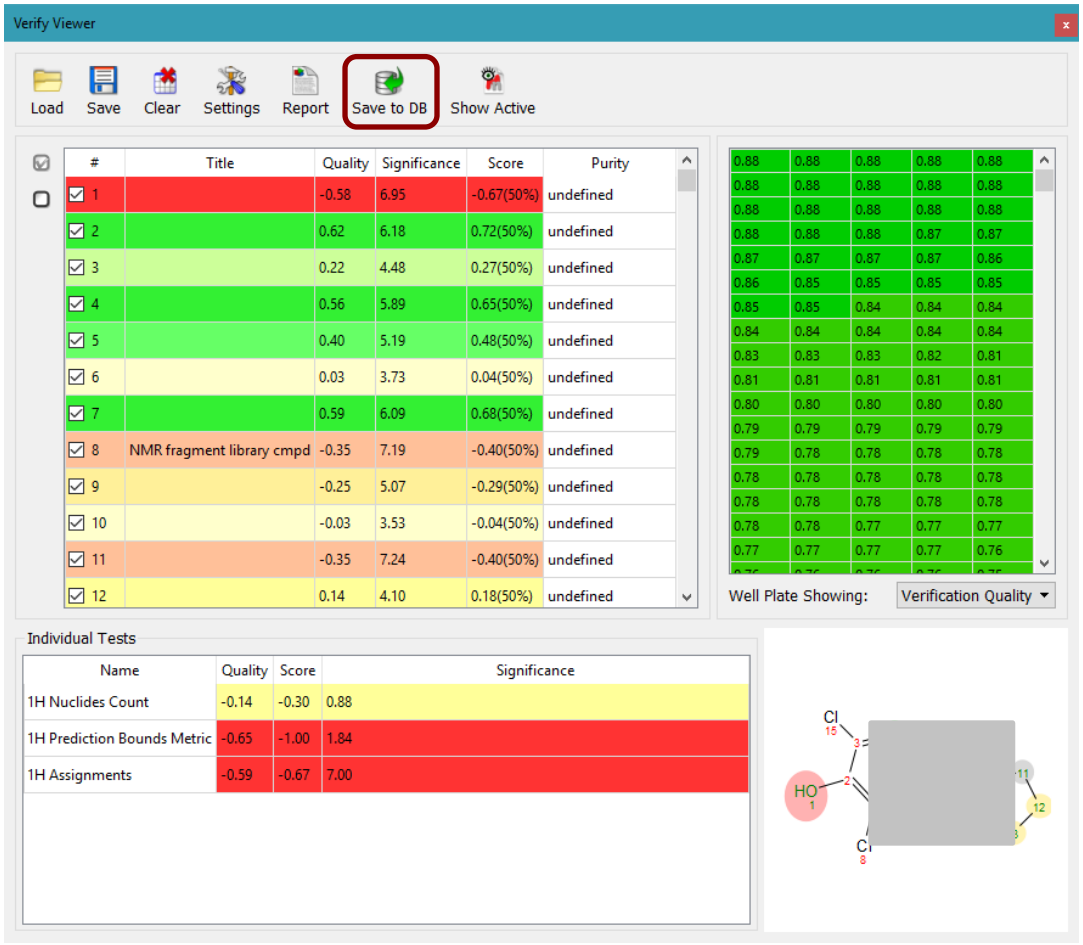
Build Ref Database

Typically we add a custom field to hold the compound IDs of each spectrum.

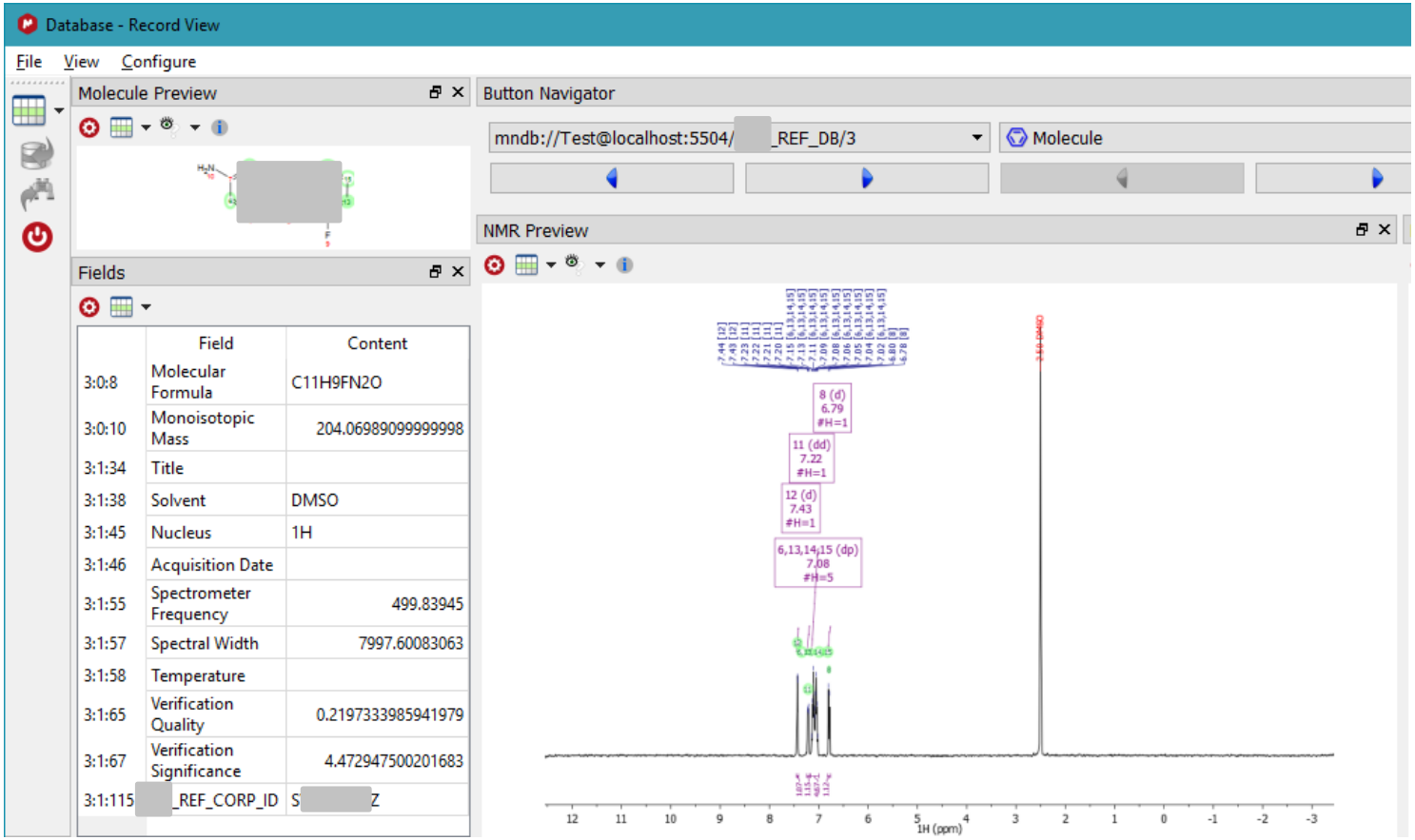
Usually a short script is necessary to extract the compound IDs from the NMR filename, Title, or Comment fields and save them to the relevant field(s) in the database.



Click the Save to DB button. This will save all the spectra/molecules to the database.

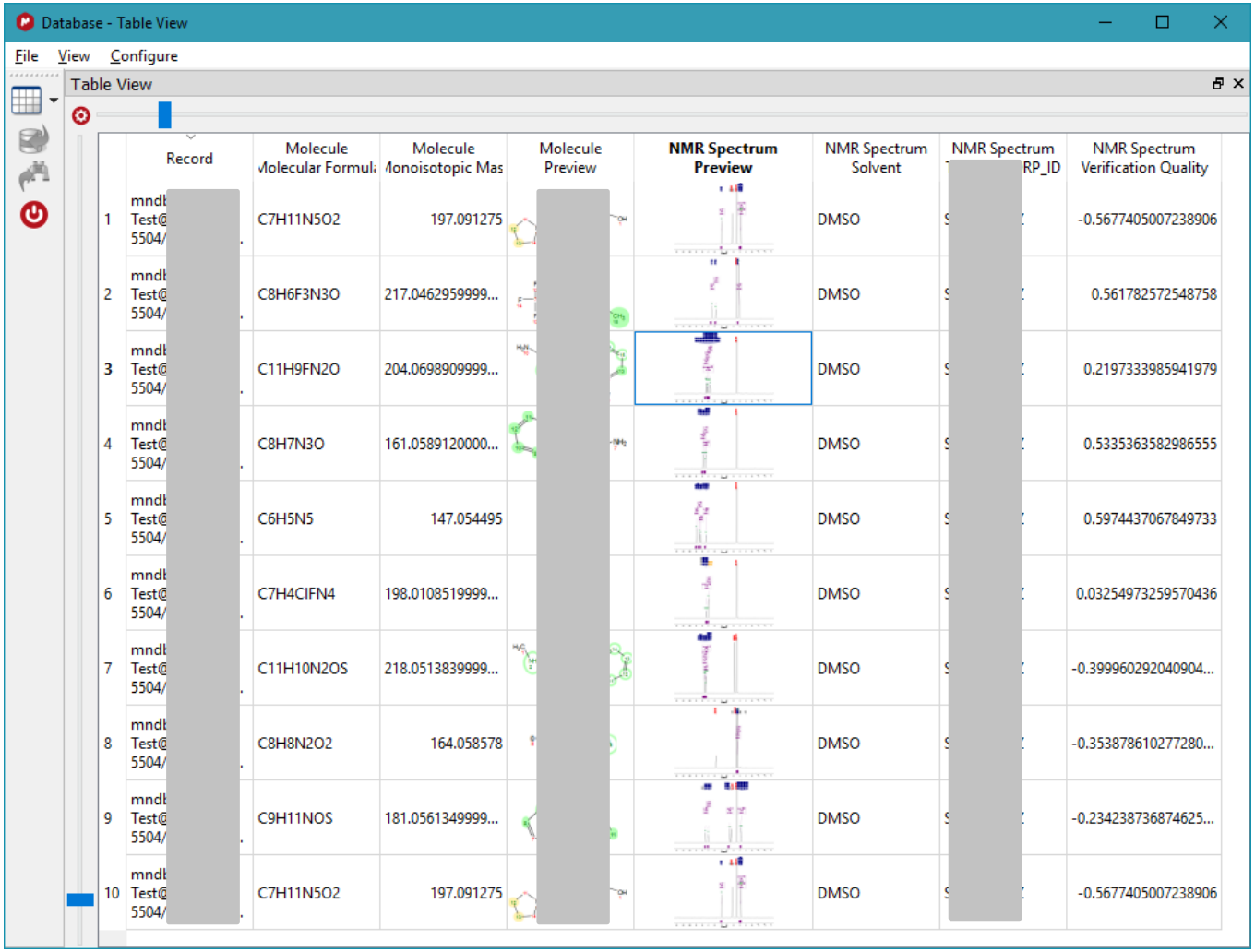


Record view of a record in the database:



	Field	Content
3:0:8	Molecular Formula	C11H9FN2O
3:0:10	Monoisotopic Mass	204.06989099999998
3:1:34	Title	
3:1:38	Solvent	DMSO
3:1:45	Nucleus	1H
3:1:46	Acquisition Date	
3:1:55	Spectrometer Frequency	499.83945
3:1:57	Spectral Width	7997.60083063
3:1:58	Temperature	
3:1:65	Verification Quality	0.2197333985941979
3:1:67	Verification Significance	4.472947500201683
3:1:115	_REF_CORP_ID	S Z

Table view of multiple records in the database:



	Record	Molecule Molecular Formula	Molecule Monoisotopic Mas	Molecule Preview	NMR Spectrum Preview	NMR Spectrum Solvent	NMR Spectrum [illegible]RP_ID	NMR Spectrum Verification Quality
1	mndb Test@ 5504/	C7H11N5O2	197.091275			DMSO		-0.5677405007238906
2	mndb Test@ 5504/	C8H6F3N3O	217.0462959999...			DMSO		0.561782572548758
3	mndb Test@ 5504/	C11H9FN2O	204.0698909999...			DMSO		0.2197333985941979
4	mndb Test@ 5504/	C8H7N3O	161.0589120000...			DMSO		0.5335363582986555
5	mndb Test@ 5504/	C6H5N5	147.054495			DMSO		0.5974437067849733
6	mndb Test@ 5504/	C7H4ClFN4	198.0108519999...			DMSO		0.03254973259570436
7	mndb Test@ 5504/	C11H10N2OS	218.0513839999...			DMSO		-0.399960292040904...
8	mndb Test@ 5504/	C8H8N2O2	164.058578			DMSO		-0.353878610277280...
9	mndb Test@ 5504/	C9H11NOS	181.0561349999...			DMSO		-0.234238736874625...
10	mndb Test@ 5504/	C7H11N5O2	197.091275			DMSO		-0.5677405007238906

Our experience

- We have done this for several libraries of 1-3K compounds.
- After some trial and error in setup, each run usually takes 3-4 hours to complete.
- The tools make it very convenient to browse through the results, focus on the ones with possible problems and make changes as necessary.
- There are typically ~10-20% of compounds with red flags and the problems are mostly real problems with the compound itself, low solubility, or low-quality spectra. Sometimes Mnova makes mistakes too.
- The database tools makes much more efficient to manage the data.

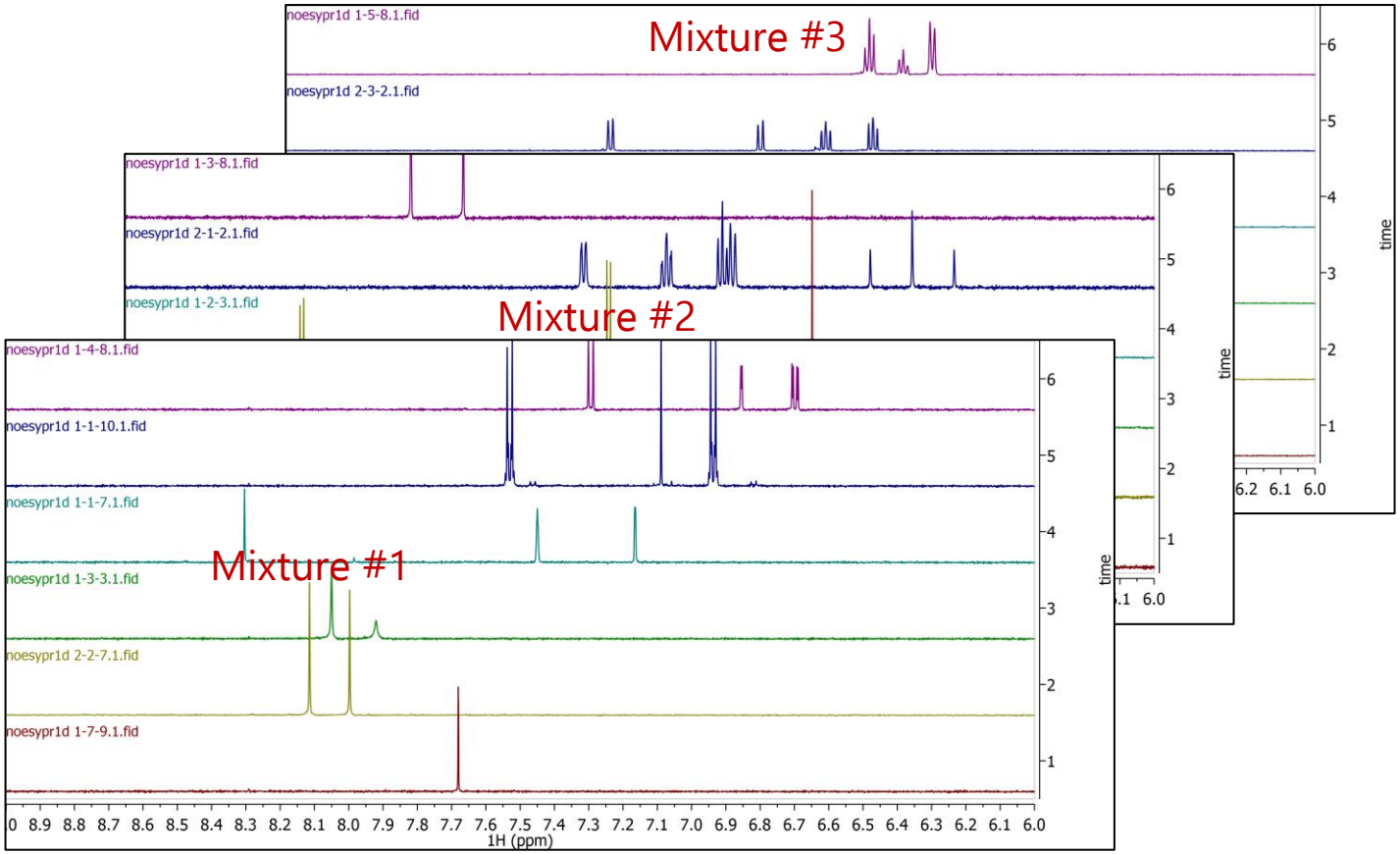
Mestrelab Research
chemistry software solutions

MixDesign for pooling compounds

Pool compounds

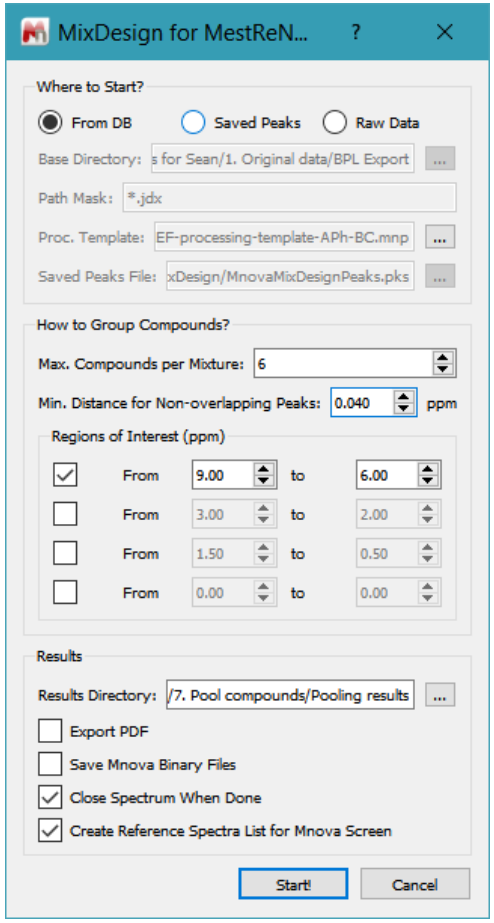
Goal: To optimize the combinations of 6-10 compounds per mixture, so that peaks within region of interest don't overlap, or at least one non-overlapping peak for each spectrum

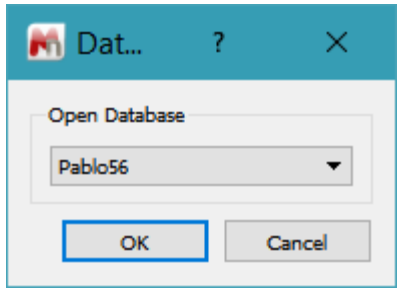
Why? More reliable for subsequent analysis to "deconvolute" the compounds in the mixtures.

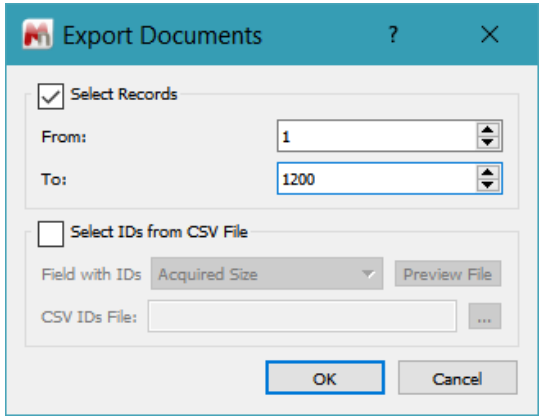


Pool compounds

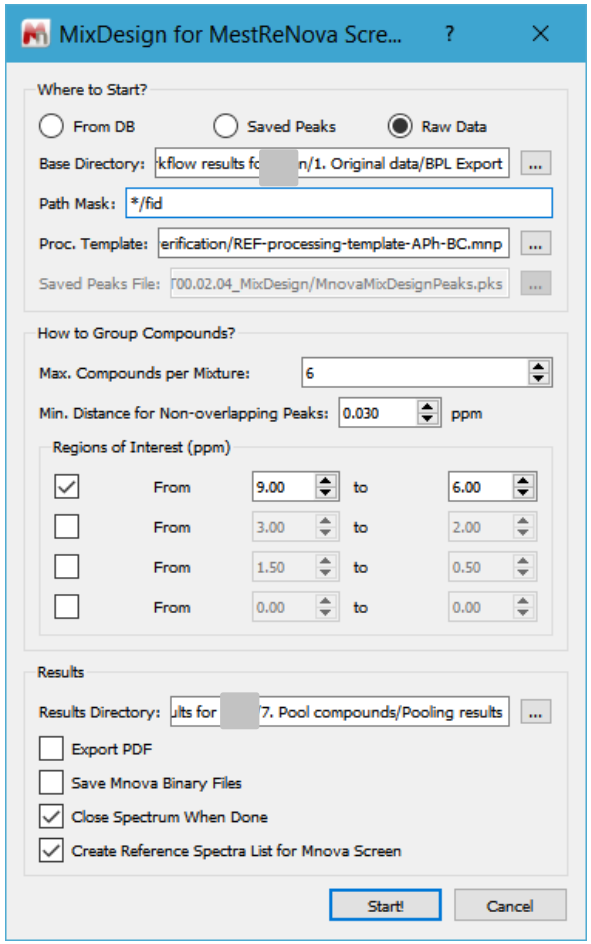
The MixDesign.qs script can pool compounds using the ref spec either saved in a DB or using the raw data. Choose Scripts | Run Script to run it.







If to start with the raw data, a processing template is used to process all the spectra:

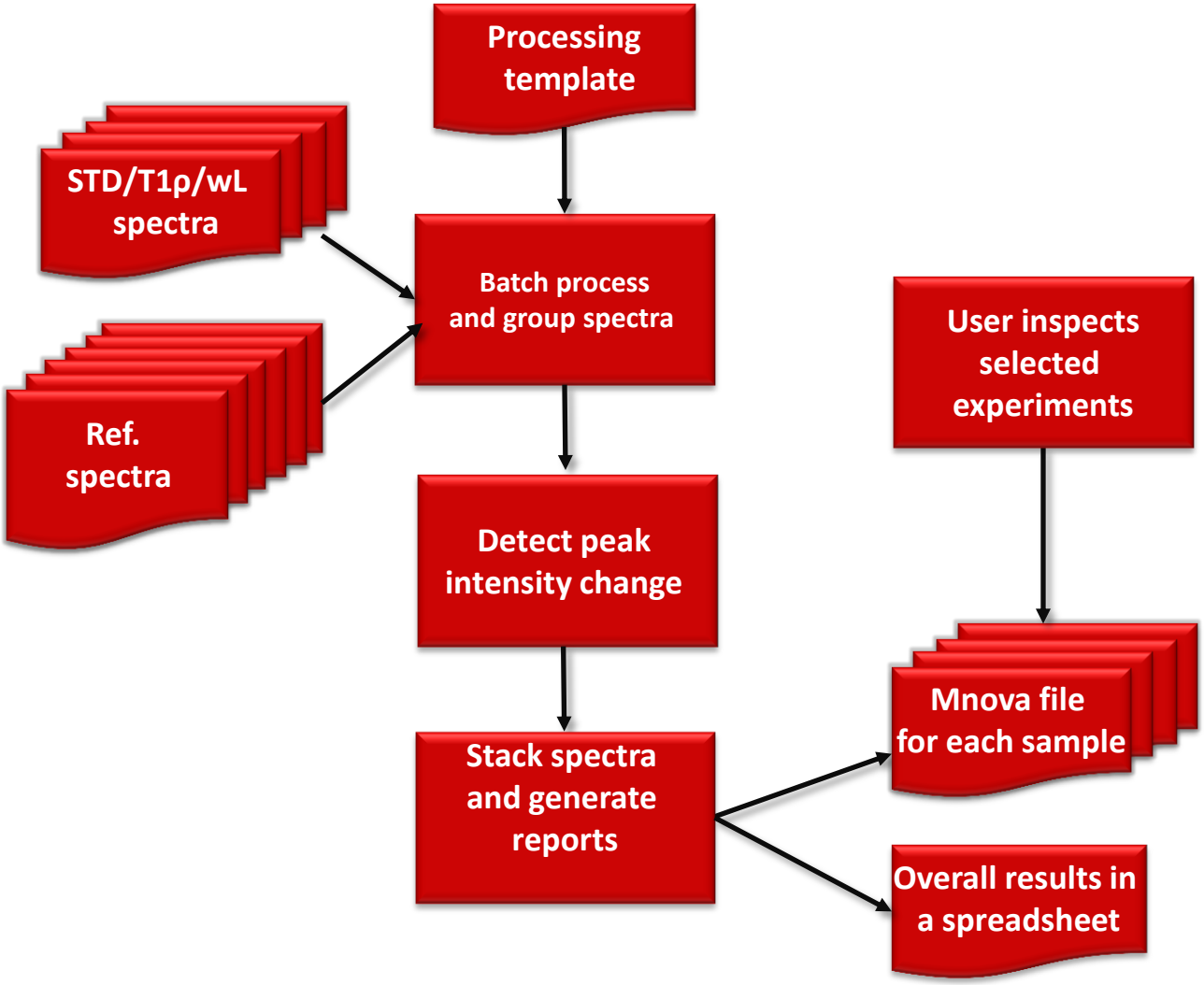


- Example:
  - 1578 ^{1}H spectra/compounds.
  - 6 compounds per mixture.
  - ROI: 9-6 ppm.
  - 0.04 ppm as minimum distance for non-overlapping spectra.
  - Got 263 mixtures in ~40 minutes.
  - 13 of them have one spectrum completely overlapping.
  - The output spreadsheets can be used by Mnova Screen for associating mixture spectra and reference spectra.
- We continue to improve the script based on users' feedback

Mestrelab Research
chemistry software solutions

Mnova Screen for ligand binding spectral analysis

Screen: From data to hits

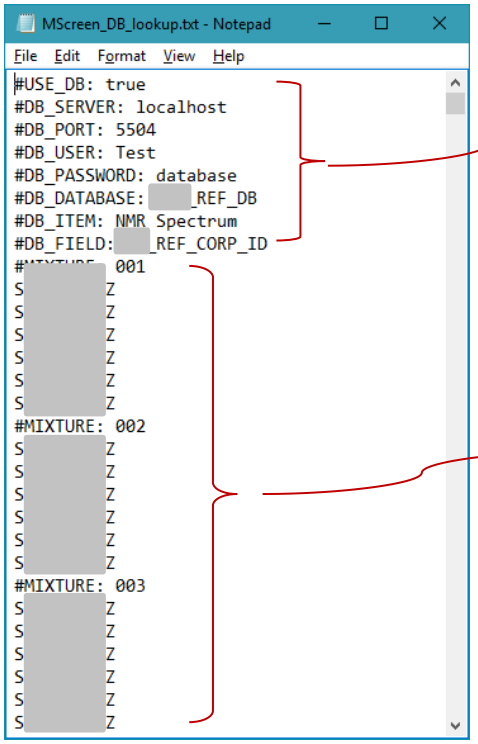


Data it handles

- H-1 or F-19
- Single compounds or mixtures
- With or without reference spectra
- Single or multiple types of spectra (STD, T1rho, WaterLogsy, CPMG)
- Use of Blank, w/ Protein, & w/ Protein+ Inhibitors
- Mnova Screen can handle all of them

Example: Run Screen using a Ref DB

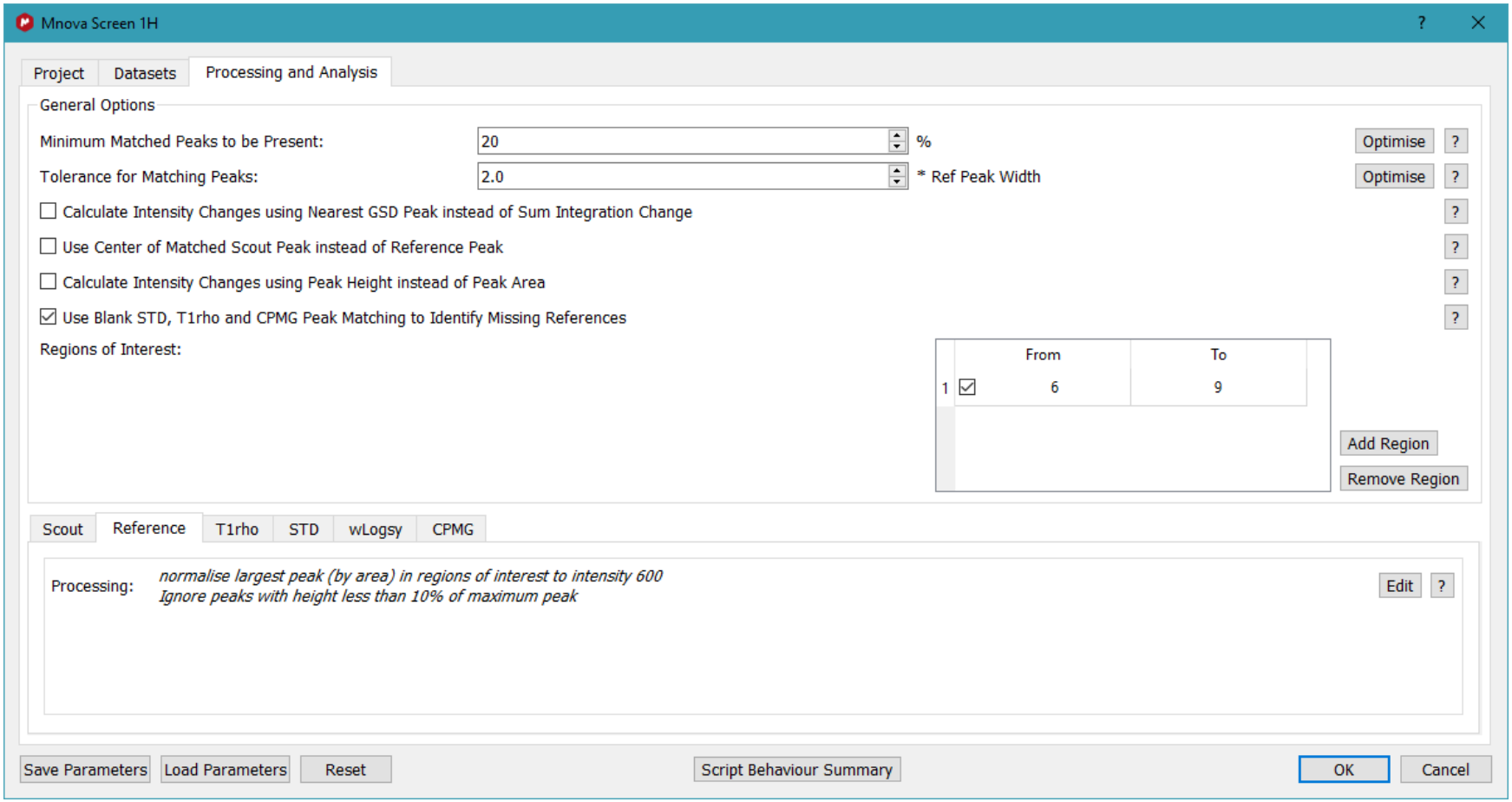
- ❑ 263 STD on/off resonance spectra for a total of 1,578 compounds.
- ❑ Each mixture has 6 compounds.
- ❑ A lookup table for Screen to find the reference spectra for each sample.



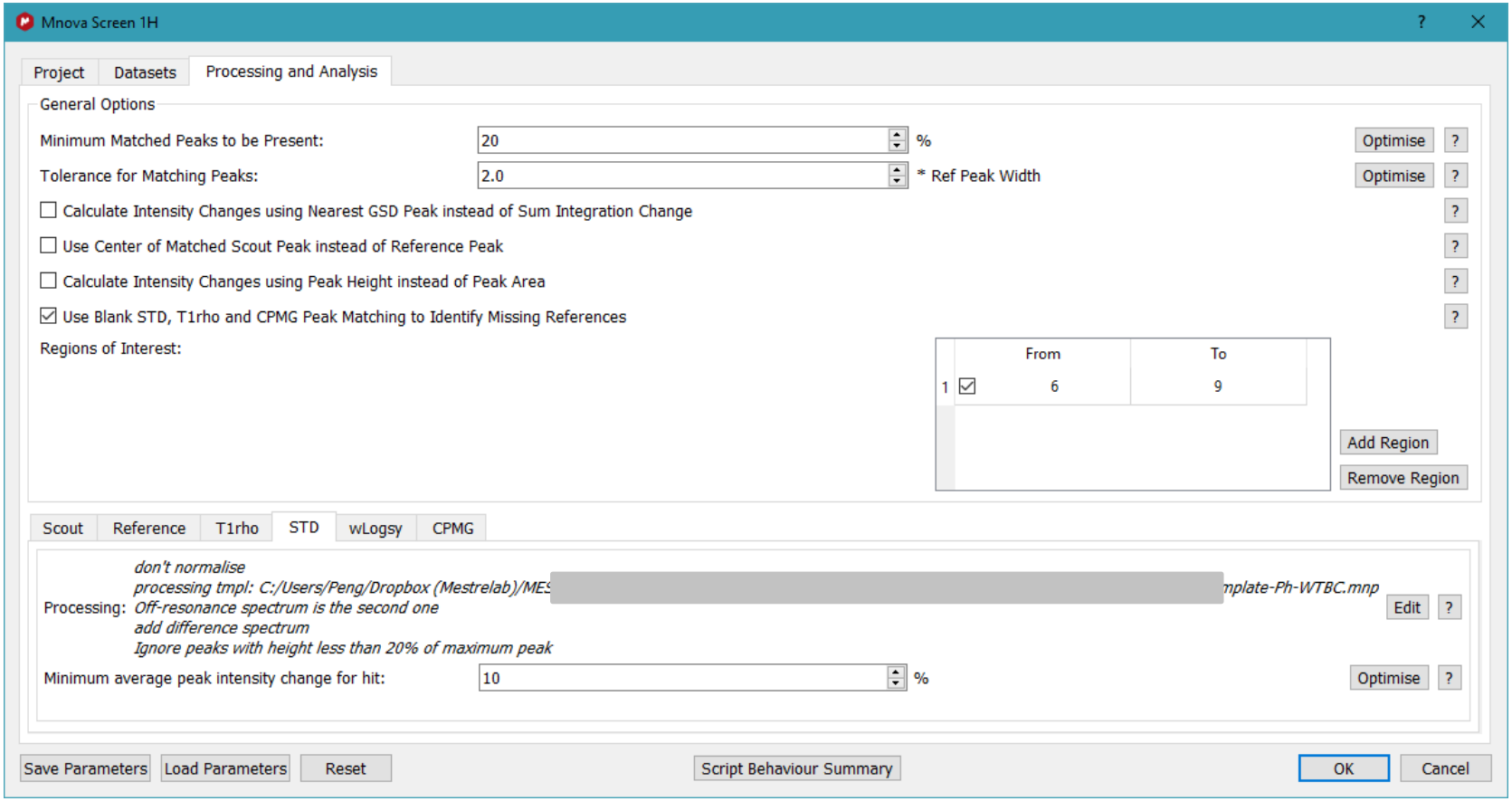
Database related information. Enter info about your database server, login account, database name, item and field (where to find the compound IDs).

IDs of the fragments for each sample/mixture.

- The Processing and Analysis tab: setup for Reference spectra.

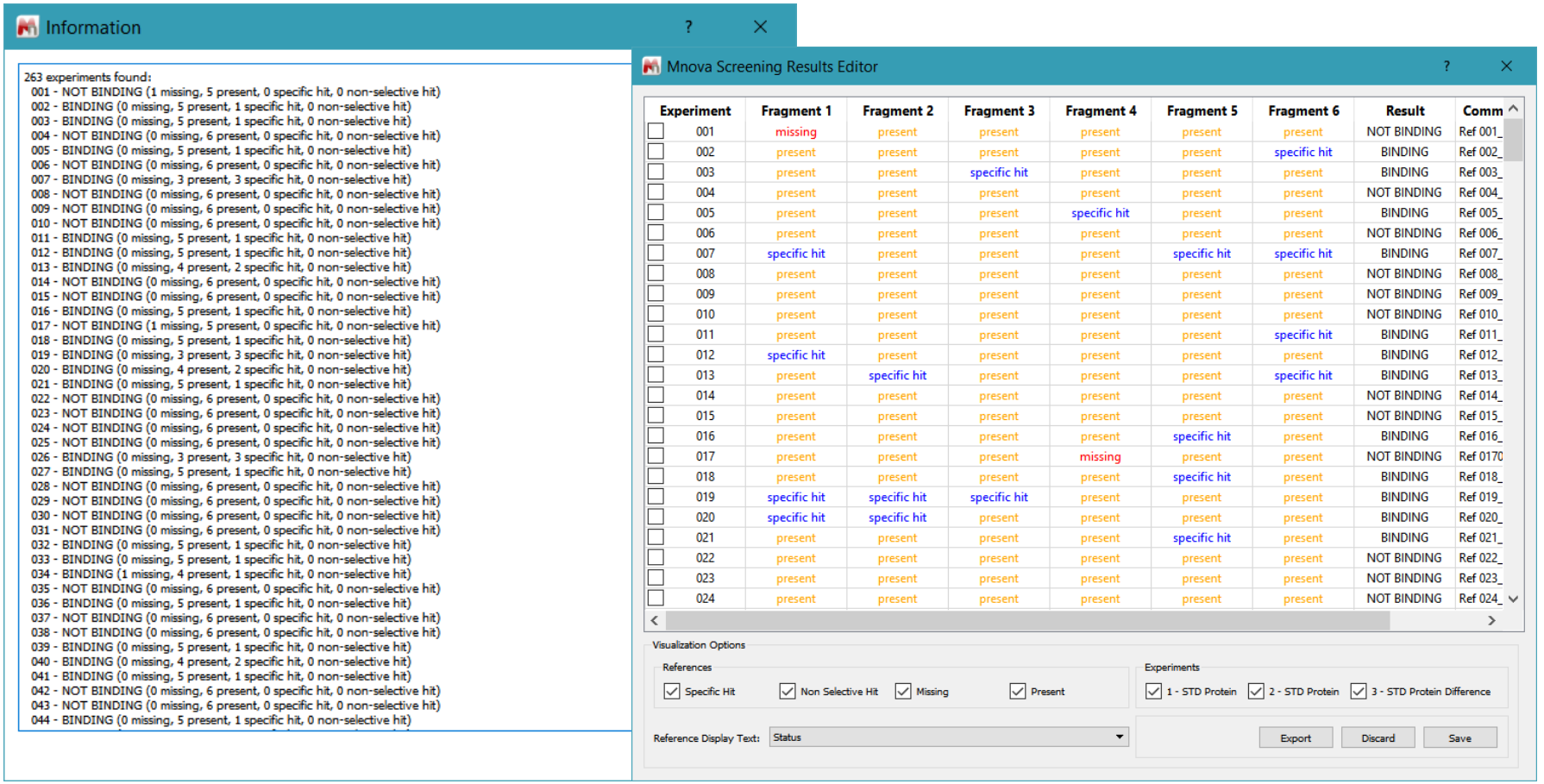


- The Processing and Analysis tab: setup for STD spectra.



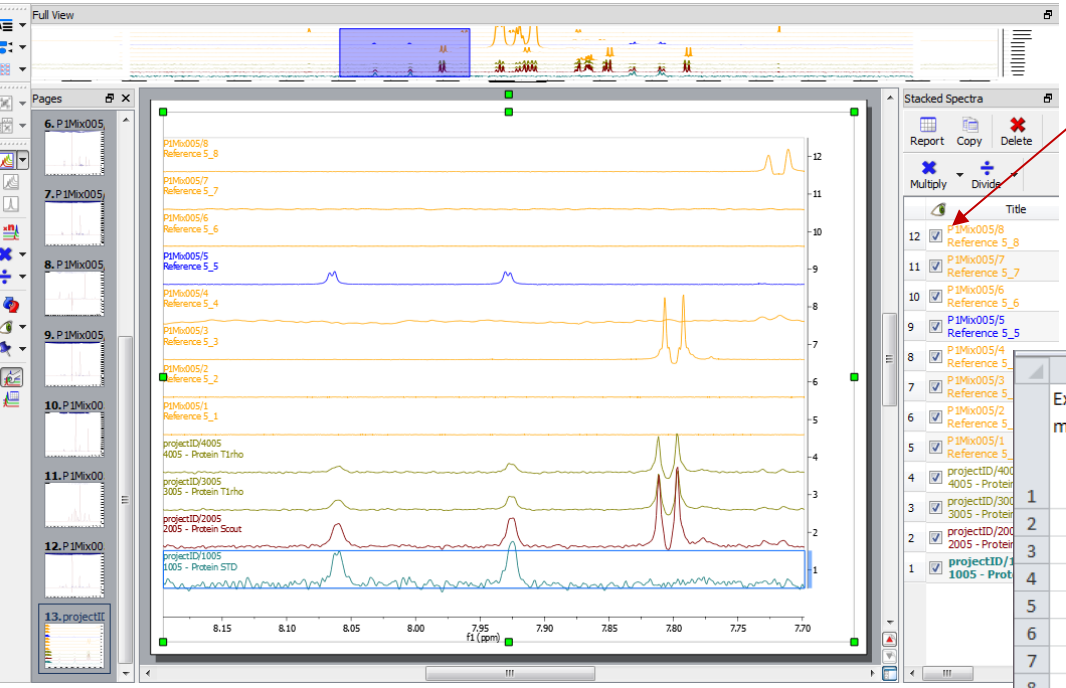
Results of Screen

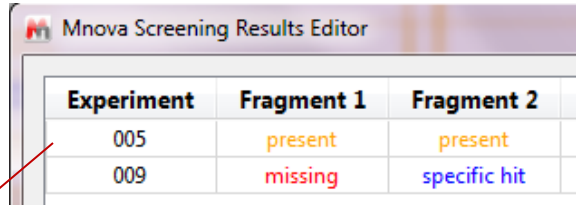
- Batch processing finished in about 2 hours.



Mestrelab Research
chemistry software solutions

Verify the screening results





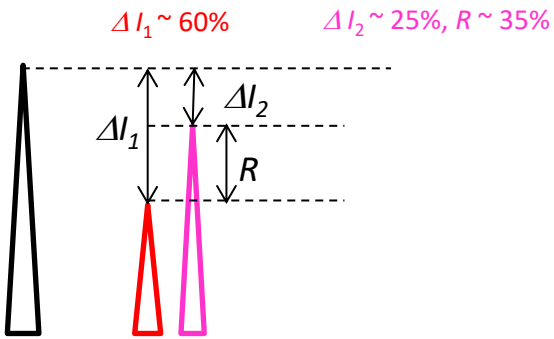
	A	B	C	D	E	F	G	H	I
1	Experiment	Reference	Status	Label	Ref Peak Count	Average Scout/Blank Match Proportion	STD Protein Match Proportion	T1rho Protein Average Intensity Change	T1rho Protein Max Intensity Change
2	5	5_1	present	5_1	19	42.10%	0.00%	14.50%	43.40%
3	5	5_2	present	5_2	12	33.30%	0.00%	26.20%	81.80%
4	5	5_3	present	5_3	9	44.40%	0.00%	17.20%	33.60%
5	5	5_4	present	5_4	9	88.90%	0.00%	7.70%	19.50%
6	5	5_5	specific hit	5_5	8	50.00%	87.50%	45.40%	51.90%
7	5	5_6	present	5_6	7	57.10%	0.00%	17.70%	45.10%
8	5	5_7	missing	5_7	0	-	-	0.00%	0.00%
9	5	5_8	present	5_8	11	27.30%	0.00%	22.20%	52.20%
10	9	9_1	missing	9_1	1	0.00%	0.00%	18.30%	18.30%
11	9	9_2	specific hit	9_2	7	71.40%	28.60%	13.80%	24.90%
12	9	9_3	present	9_3	12	66.70%	0.00%	12.30%	31.80%
13	9	9_4	specific hit	9_4	22	59.10%	27.30%	18.20%	28.40%
14	9	9_5	present	9_5	19	63.20%	0.00%	7.90%	30.90%
15	9	9_6	present	9_6	8	50.00%	0.00%	4.00%	6.60%
16	9	9_7	present	9_7	9	77.80%	0.00%	5.70%	22.40%
17	9	9_8	specific hit	9_8	32	62.50%	31.30%	12.10%	43.10%

Determining specific hits from competition experiments

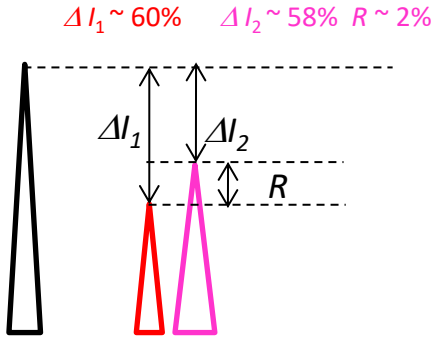
- Two thresholds are defined by the user:
  - T_1: minimum intensity decrease for a hit.
  - T_2: minimum intensity recovery rate for a specific hit.
- If $\Delta I_1 > T_1$ and $R > T_2$: specific hit (see Ex.1 below)
- If $\Delta I_1 > T_1$ and $R \leq T_2$: non-specific hit (see Ex.1 below)

Example 1

T_2 = 5%



Example 2



Demo of Mnova Screen

- Two mixtures with 8 compounds each.
- Using STD difference spectra, and T1rho (short/long-spin lock).

- Typically it takes about 3-4 hours to complete a screening batch of 2-300 mixtures.
- Using database of reference spectra is usually faster.
- The automated results are comparable with careful manual analysis results but much faster. See results comparison and discussion at
  C. Peng, A. Frommlet, M. Perez, C. Cobas, A. Blechschmidt, S. Dominguez, and A. Lingel., *J. Med. Chem.* 2016, **59**, 3303−3310.
- The tools that allow you to easily browse through the results, and verify and correct them manually is very convenient.
- Mnova Screen has been used routinely by > 10 companies.

Mestrelab Research
chemistry software solutions

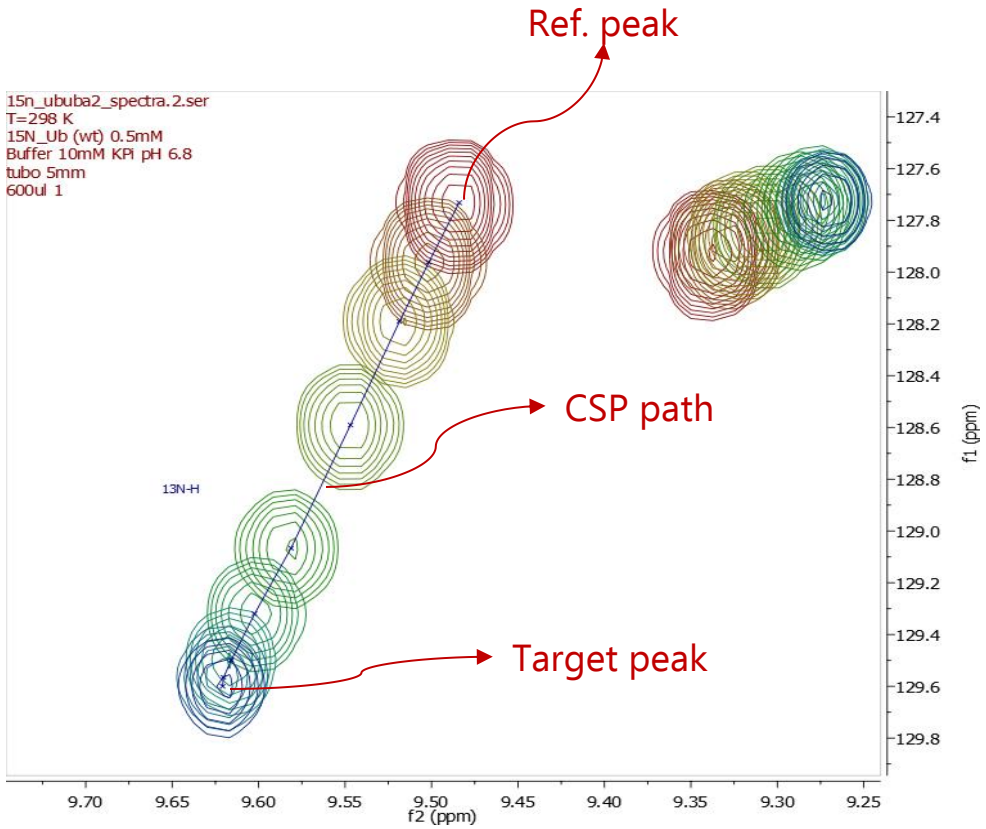
Mnova CSP – Chemical shift perturbation analysis

General workflows

- Mnova CSP allows you to process and analyze a series CSP spectra fully automatically, or interactively, or both
- Full automatic processing and analysis starting from 2D raw data to K_D's
  - Prepare 3 information files: Titration file, Ligand file, Peaks file; and enter them to CSP.
  - CSP processes all HSQC spectra, stacks them, tracks the peak movements, calculates the CSPs and K_D for each peak, and does statistics of all the K_Ds.
- Manual analysis
  - You open and stack multiple HSQC spectra interactive.
  - You pick the peaks and let CSP monitor automatically track their shift path across all the spectra.
  - You manually correct the peak tracking as needed.
  - CSP calculates the CSPs and K_D for each peak, and does statistics of all the K_Ds in real time.

Chemical shift perturbation (CSP)

- A reference peak, usually assigned to an amino acid residual in a protein, shifts its location in $^{1}H/^{15}N$ (or $^{1}H/^{13}C$) HSQC spectra as the ligand is added.
  - [P]: concentration of protein
  - [L]: concentration of ligand
  - [L]/[P]: ratio of ligand/protein – Column "Lt/Pt" in CSP Panel



	Spectrum	Lt/Pt
1	**15n_ububa2_spectra.2.ser**	**0**
2	15n_ububa2_spectra.24.ser	0.25
3	15n_ububa2_spectra.26.ser	0.5
4	15n_ububa2_spectra.28.ser	1
5	15n_ububa2_spectra.30.ser	2
6	15n_ububa2_spectra.32.ser	3
7	15n_ububa2_spectra.34.ser	5
8	15n_ububa2_spectra.36.ser	7
9	15n_ububa2_spectra.38.ser	9

Chemical shift perturbation (CSP)

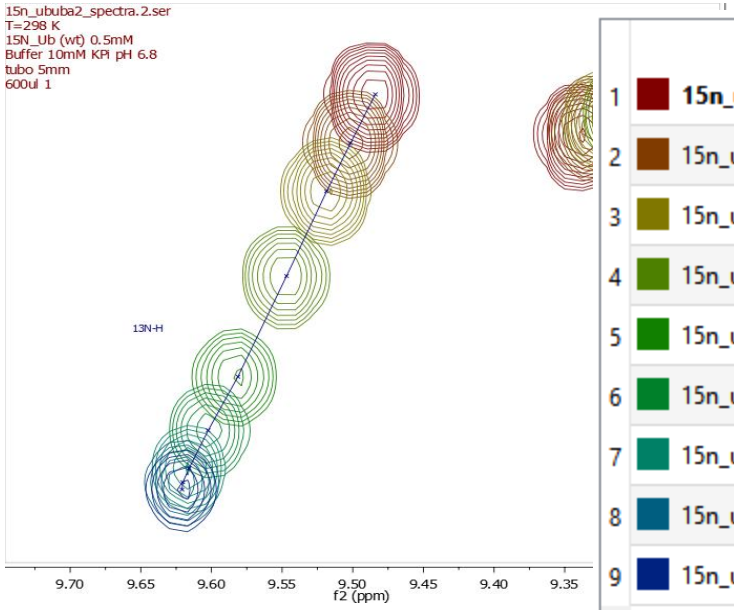
- The chemical shift changes along the path from the ref. peak to target peak is measured and normalized:

$$CSP\ (\text{ppm}) = \sqrt{(F(H)\cdot\Delta\delta_H)^2 + (F(N)\cdot\Delta\delta_N)^2}$$

or

$$CSP\ (\text{ppm}) = \sqrt{(F(H)\cdot\Delta\delta_H)^2 + (F(C)\cdot\Delta\delta_C)^2}$$

By default: *F(H)* = 1; *F(N)* = 0.156; *F(C)* = 0.185. You can change the values in Settings.

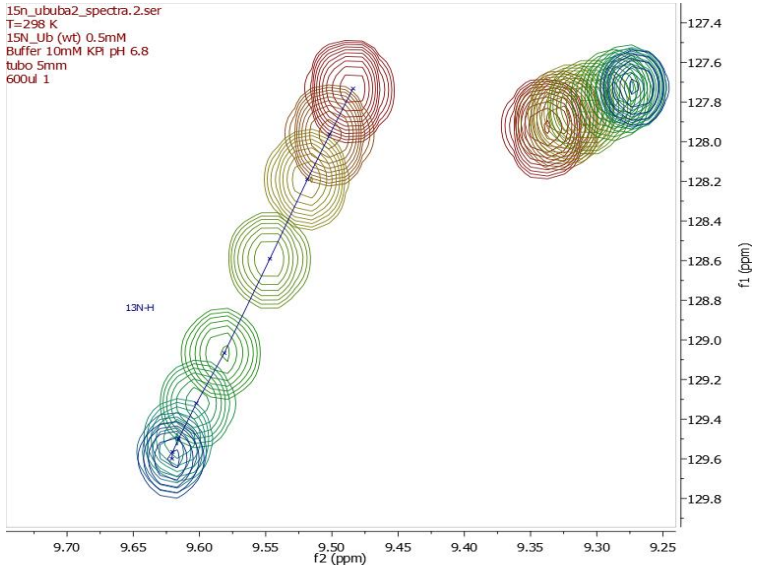


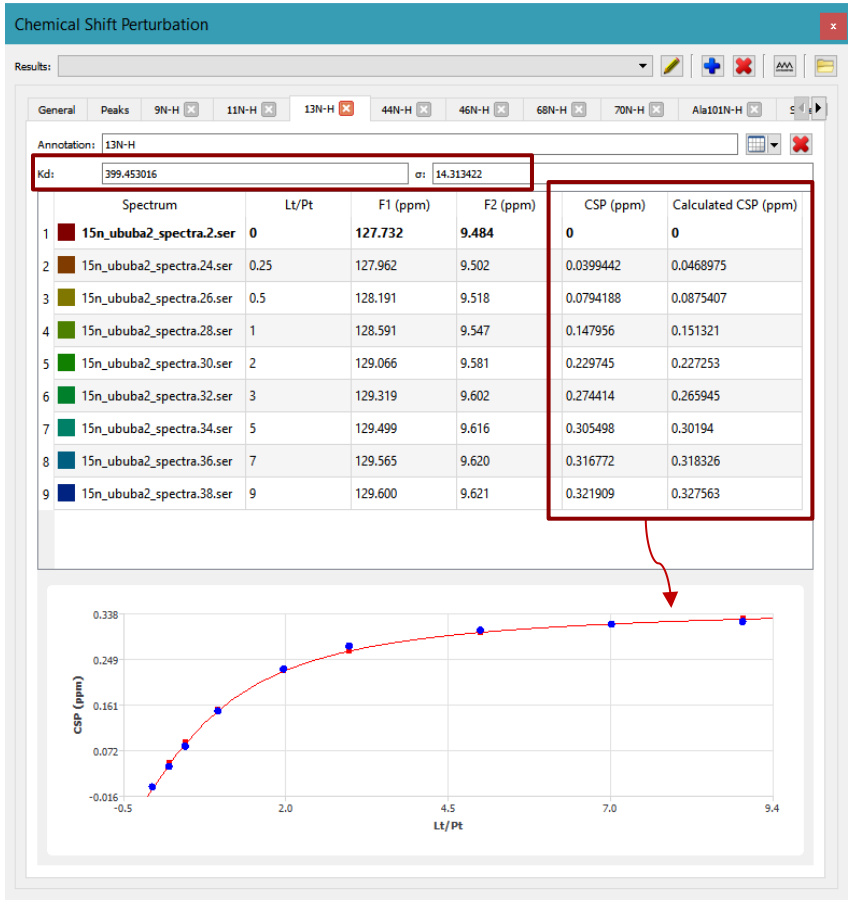
	Spectrum	Lt/Pt	F1 (ppm)	F2 (ppm)	CSP (ppm)
			$\Delta\delta_N$	$\Delta\delta_H$	*CSP*
1	**15n_ububa2_spectra.2.ser**	**0**	**127.732**	**9.484**	**0**
2	15n_ububa2_spectra.24.ser	0.25	127.962	9.502	0.0399442
3	15n_ububa2_spectra.26.ser	0.5	128.191	9.518	0.0794188
4	15n_ububa2_spectra.28.ser	1	128.591	9.547	0.147956
5	15n_ububa2_spectra.30.ser	2	129.066	9.581	0.229745
6	15n_ububa2_spectra.32.ser	3	129.319	9.602	0.274414
7	15n_ububa2_spectra.34.ser	5	129.499	9.616	0.305498
8	15n_ububa2_spectra.36.ser	7	129.565	9.620	0.316772
9	15n_ububa2_spectra.38.ser	9	129.600	9.621	0.321909

Chemical shift perturbation (CSP)

- The CPS values are plotted against the ratios of ligand/protein concentrations and fit to a titration curve to determine the dissociation constant, K_d and the fitting error (σ) according to

$$CSP = \frac{\Delta CSP_{max}}{2}\left\{\left(1+\left(\frac{L_T}{P_T}\right)+\frac{K_d}{P_T}\right)-\sqrt{\left(1+\left(\frac{L_T}{P_T}\right)+\frac{K_d}{P_T}\right)^2-4\left(\frac{L_T}{P_T}\right)}\right\}$$



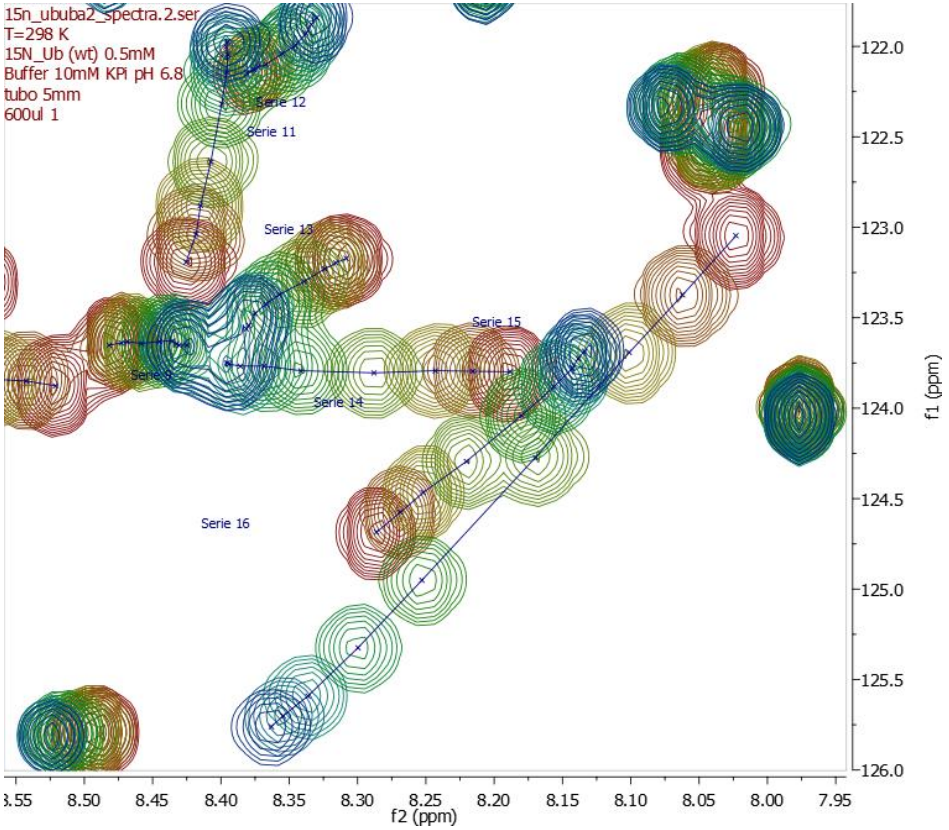


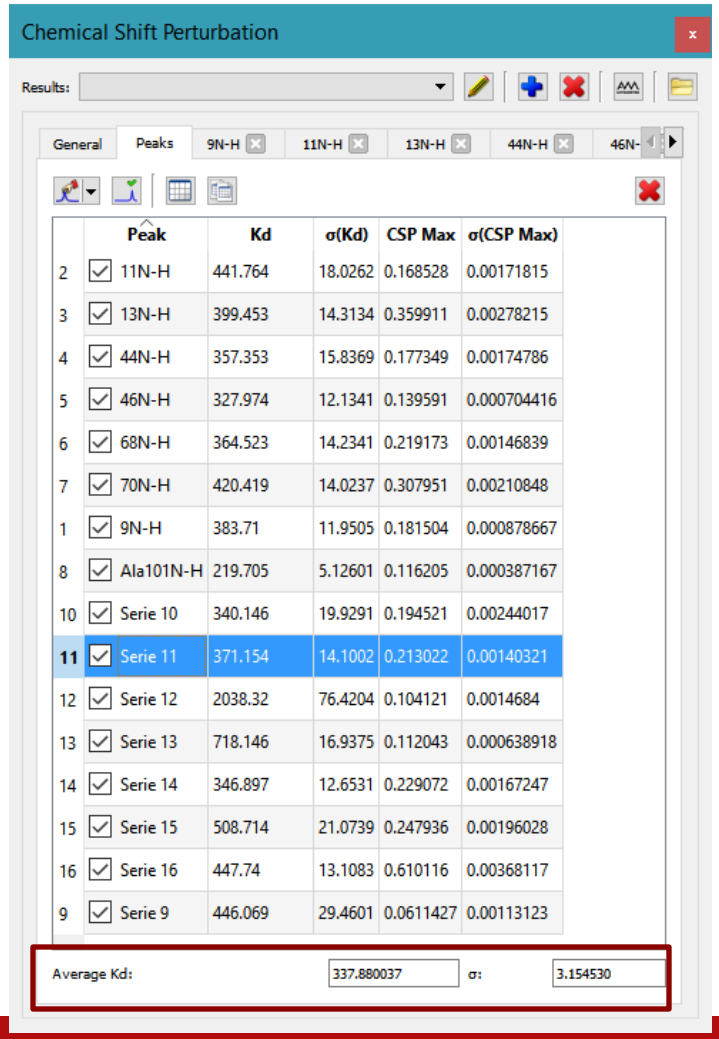
	Spectrum	Lt/Pt	F1 (ppm)	F2 (ppm)	CSP (ppm)	Calculated CSP (ppm)
1	**15n_ububa2_spectra.2.ser**	**0**	**127.732**	**9.484**	**0**	**0**
2	15n_ububa2_spectra.24.ser	0.25	127.962	9.502	0.0399442	0.0468975
3	15n_ububa2_spectra.26.ser	0.5	128.191	9.518	0.0794188	0.0875407
4	15n_ububa2_spectra.28.ser	1	128.591	9.547	0.147956	0.151321
5	15n_ububa2_spectra.30.ser	2	129.066	9.581	0.229745	0.227253
6	15n_ububa2_spectra.32.ser	3	129.319	9.602	0.274414	0.265945
7	15n_ububa2_spectra.34.ser	5	129.499	9.616	0.305498	0.30194
8	15n_ububa2_spectra.36.ser	7	129.565	9.620	0.316772	0.318326
9	15n_ububa2_spectra.38.ser	9	129.600	9.621	0.321909	0.327563

Mestrelab Research
chemistry software solutions

Statistics of multiple K_d

- Multiple reference peaks can be tracked and K_d calculated for each of them.
- The average K_d and standard deviation are automatically computed for them.

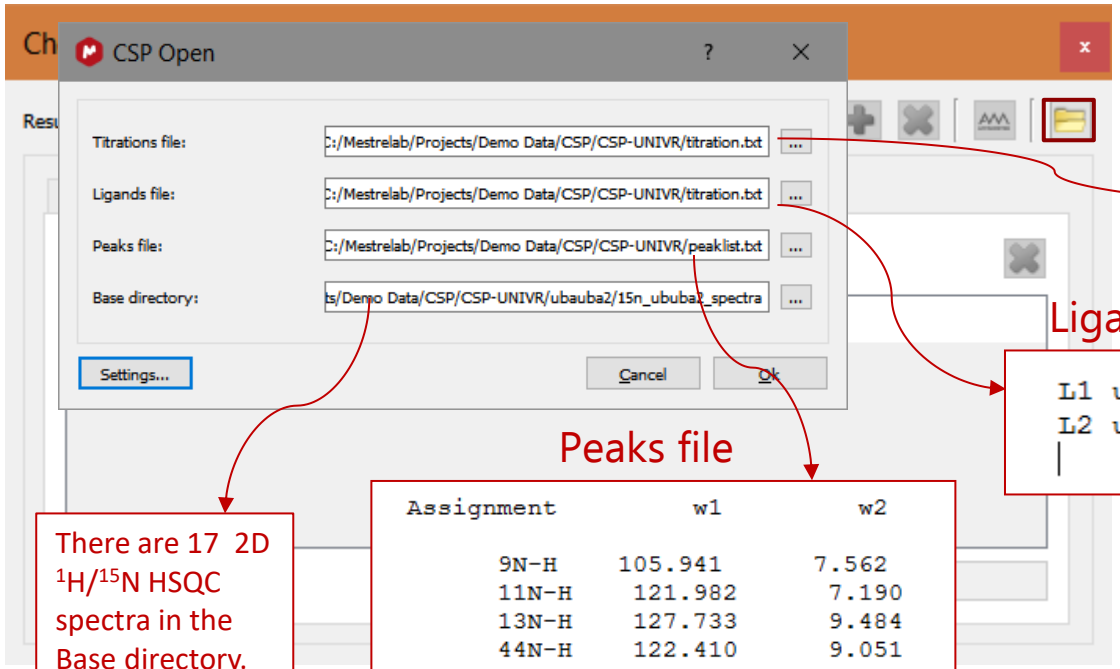




Tip: Un-check the peaks that you don't want to be used for the statistics analysis. The results will be automatically updated.

Fully automatic processing and analysis

- ❑ Choose Advanced > Chemical Shift Perturbation to open the CSP Panel.
- ❑ Click Open and enter the relevant info files and Base directory (where the 2D HSQC spectra are located). Click OK to start the auto processing.



There are 17 2D $^{1}H/^{15}N$ HSQC spectra in the Base directory. They are used based on the info in the Titration file

Peaks file

Assignment	w1	w2
9N-H	105.941	7.562
11N-H	121.982	7.190
13N-H	127.733	9.484
44N-H	122.410	9.051
46N-H	133.093	8.945
68N-H	119.736	9.148
70N-H	126.581	9.097

Ligand file

```
L1 ububa
L2 ububa2
```

Titration file

```
[P]uM          500

EXPNO          [L]/[P]

REF
2                    0.00

L1
4                    125
6                    250
8                    500
10                   1000
12                   1500
14                   2500
16                   3500
18                   4500

L2
24                    125
26                    250
28                    500
30                   1000
32                   1500
34                   2500
36                   3500
38                   4500
```

Mestrelab Research
chemistry software solutions

Batch processing for multiple ligands

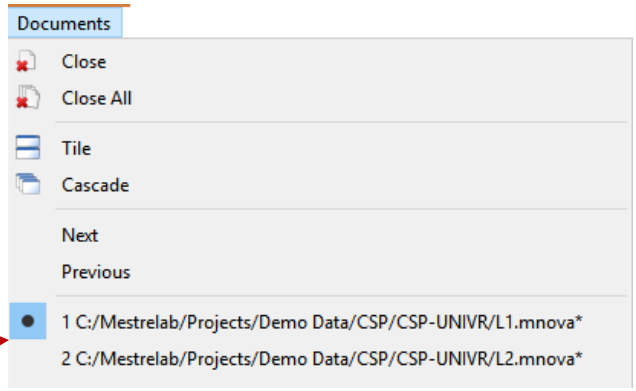
- If you put multiple ligands in the Titration file, then they will be processed and saved as multiple Mnova documents.
- Use the Document Menu to switch between the documents for details.

Titration file

```
[P]uM          500

EXPNO        [L]/[P]

REF
2                    0.00

L1
4                     125
6                     250
8                     500
10                   1000
12                   1500
14                   2500
16                   3500
18                   4500

L2
24                    125
26                    250
28                    500
30                   1000
32                   1500
34                   2500
36                   3500
38                   4500
```

Spectra and concentrations for Ligand #1

Spectra and concentrations for Ligand #2



Ligand file

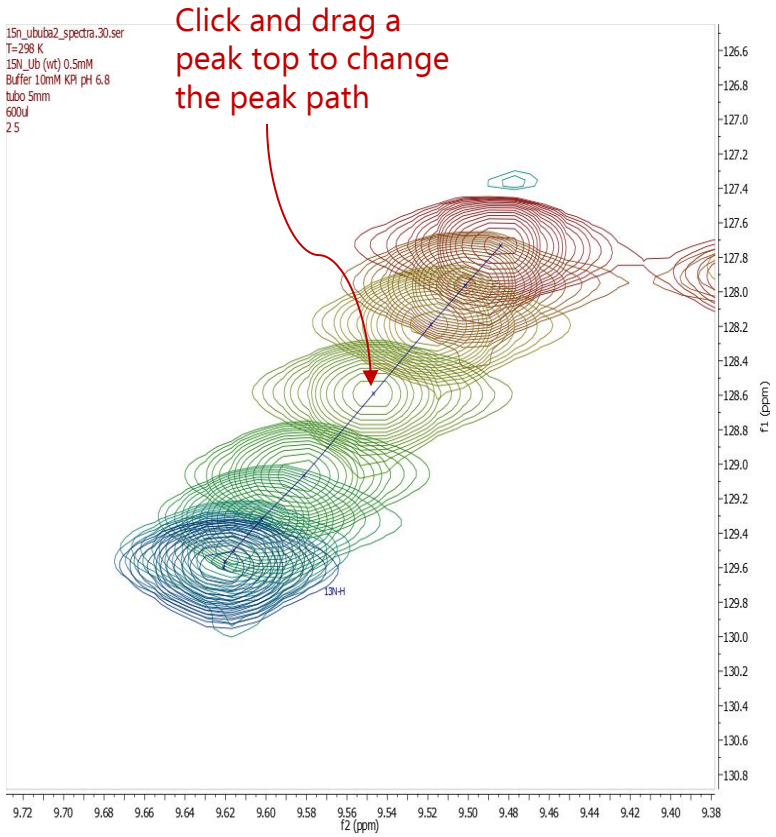
```
L1 ububa
L2 ububa2
```

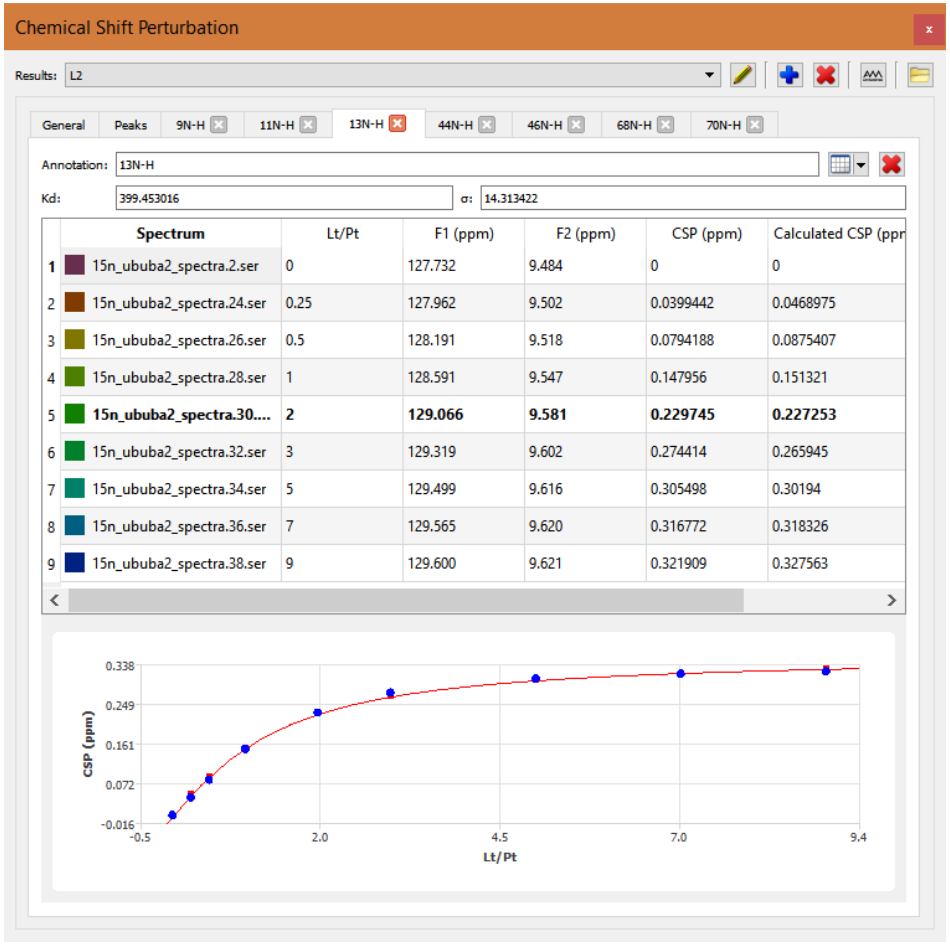
Labels for all ligands

Mestrelab Research
chemistry software solutions

Manual editing a peak path

- From the Peaks Tab in CSP Panel, double click on a row to switch to display its details and zoom to that peak path in the spectra.
- Click on any peak top and drag to change the peak path. The CSP and Kd results are updated automatically.

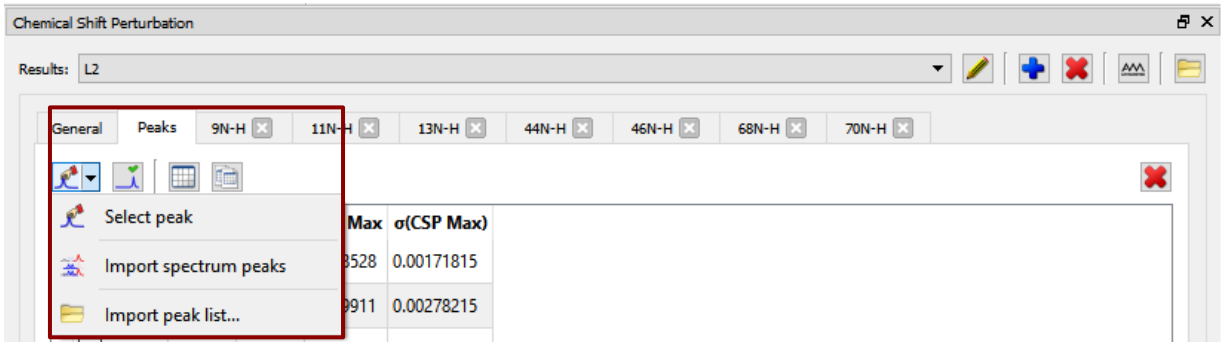




Mestrelab Research
chemistry software solutions

Which reference peaks to track?

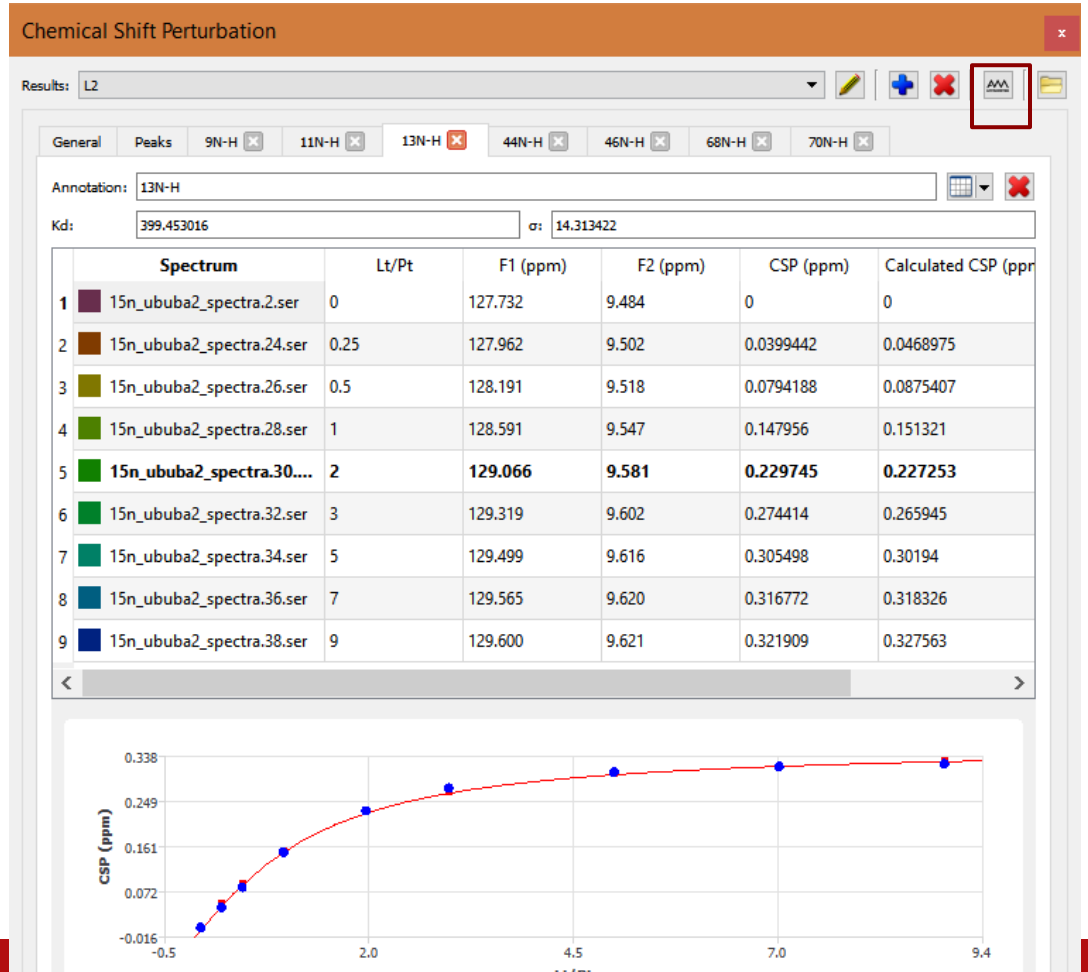
- You can enter the peaks in 3 ways:
  - **Select peak**: click to select reference peaks in the stacked plot.
  - **Import spectrum peaks**: do auto or manual peak picking in the reference spectrum first, and use those as the reference peaks.
  - **Import peak list**: Use peaks in a peak assignment table as reference peaks.
- Mnova CSP automatically track peaks across the titration spectra, and you can manual correct the peak paths if necessary.



Mestrelab Research
chemistry software solutions

More complex binding models?

- We collaborate with AffiniMeter Inc. for ligand-protein binding studies.
- The CSP results can be sent to the AffiniMeter for further analysis.
- See https://www.affinimeter.com/ for more details.



	Spectrum	Lt/Pt	F1 (ppm)	F2 (ppm)	CSP (ppm)	Calculated CSP (ppr
1	15n_ububa2_spectra.2.ser	0	127.732	9.484	0	0
2	15n_ububa2_spectra.24.ser	0.25	127.962	9.502	0.0399442	0.0468975
3	15n_ububa2_spectra.26.ser	0.5	128.191	9.518	0.0794188	0.0875407
4	15n_ububa2_spectra.28.ser	1	128.591	9.547	0.147956	0.151321
5	**15n_ububa2_spectra.30....**	**2**	**129.066**	**9.581**	**0.229745**	**0.227253**
6	15n_ububa2_spectra.32.ser	3	129.319	9.602	0.274414	0.265945
7	15n_ububa2_spectra.34.ser	5	129.499	9.616	0.305498	0.30194
8	15n_ububa2_spectra.36.ser	7	129.565	9.620	0.316772	0.318326
9	15n_ububa2_spectra.38.ser	9	129.600	9.621	0.321909	0.327563

AFFINIMETER

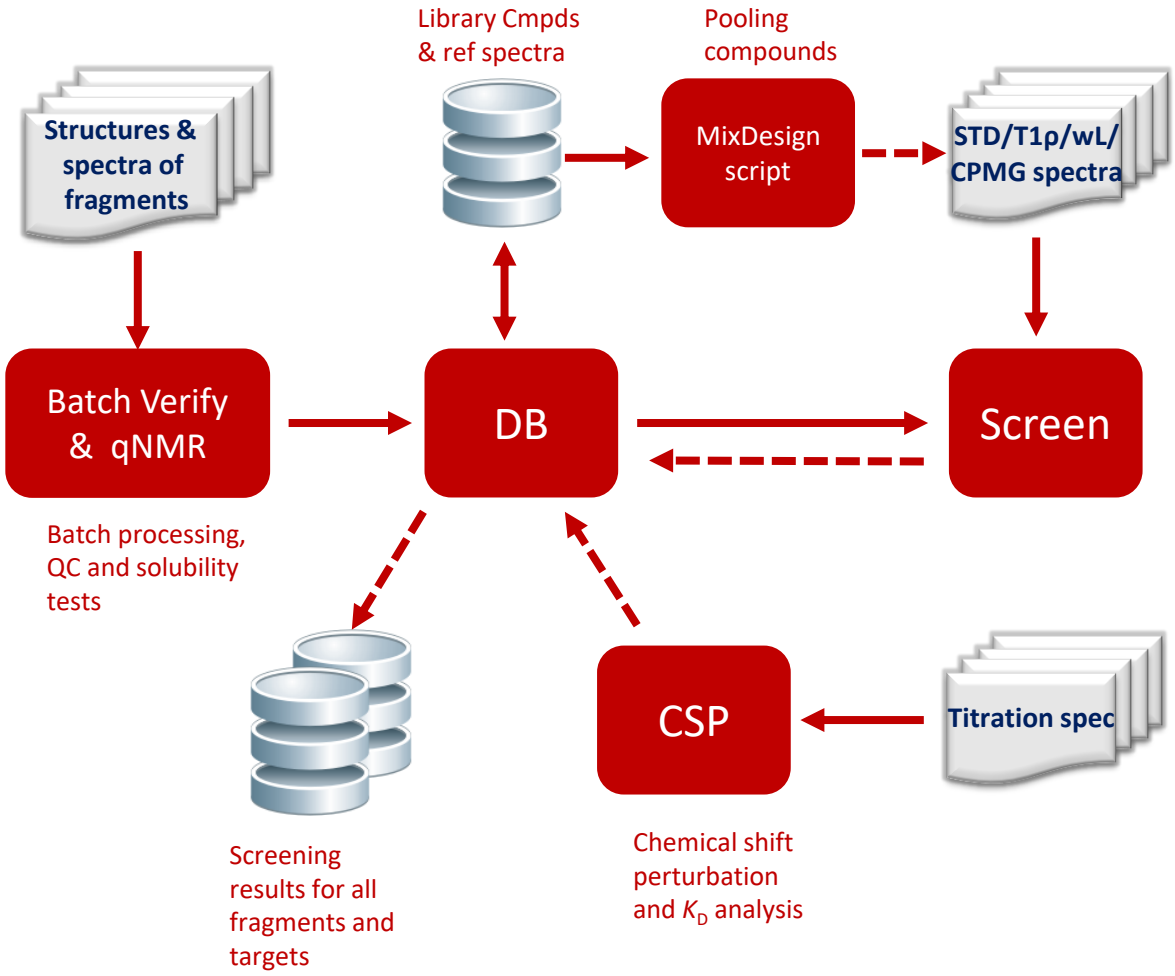
Demo of Mnova CSP

- H-N HSQC titration spectra, 8 points for each ligand.
- Two ligands.

Conclusions

Mestrelab Research
chemistry software solutions

Powerful software tools seamlessly integrated by Mnova DB

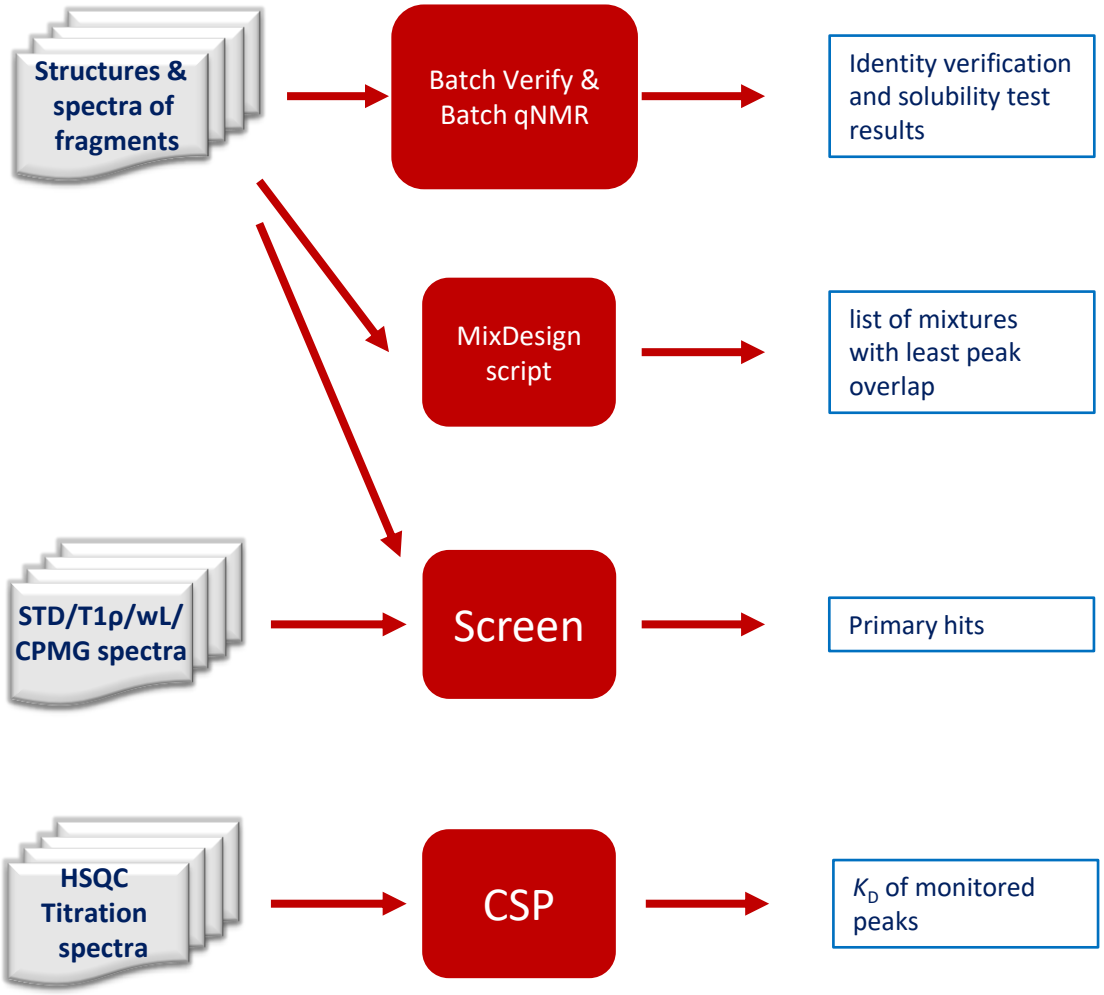


Efficient software tools for processing, analysis and management of NMR data from almost every stage of FBDD:

- library compounds QC, solubility test, and database management.
- Pooling fragments.
- Data processing for and screening results.
- Titration spectra analysis.
- Datamining of screening results in progress.

Mestrelab Research
chemistry software solutions

They can also work independently & without a database



- Flexible: use tools when you need.
- Processing and peak picking reference spectra at each run - Not an efficient way to manage and reuse your reference spectra.
- Results saved as flat files - not efficient for info management and data mining.

Mestrelab Research
chemistry software solutions

Acknowledgement

- Our collaborators
  - Pfizer (La Jolla): Jiangli Yan and Wei Wang
  - Novartis (Cambridge): Xiaolu Zhang and Jasna Fejzo
  - Novartis (Emeryville): Andreas Lingel and Alexandra Frommlet
  - Abbvie: Andrew Petros and Andrew Namanja
- Developers and product managers at Mestrelab Research
  - Chen Peng, Manuel Perez, and Silvia Mari
  - Agustin Barba and Jose Garcia
  - Carlos Cobas, Stan Sykora, Santi Dominguez

- Thank you for attending the webinar!
  Questions?