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
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
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.
A typical mid-size compound library: 500-2000 compounds

1H 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.

19F detected experiments: high sensitivity (low-μm concen.), simple spectra, large δ19F range

- Mixtures of 10-30 cmpds per sample => 20-50 samples per library.

1H-15N or 1H-13C 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

- **Batch Verify & qNMR**
  - Structures & spectra of fragments
  - Batch processing, QC and solubility tests

- **DB**
  - Library Cmpds & ref spectra
  - MixDesign script
  - Pooling compounds
  - STD/T1p/wL/CPMG spectra

- **Screen**
  - CSP
  - Titration spec

- **CSP**
  - Chemical shift perturbation and $K_D$ analysis

- **Screening results for all fragments and targets**
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
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.
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.
Once the settings are done, click OK to start the batch processing. It processes all the spectra, does structure verification and quantitation.
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:
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.
Once you are ready, create a new database to save the spectra and molecules. Choose Database | Connect to connect to the DB Server.
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:
Table view of multiple records in the database:
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.
MixDesign for pooling 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.
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:

![MixDesign for MestReNova Scre...](image)
Our experience

• 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
Mnova Screen for ligand binding spectral analysis
Screen: From data to hits

- STD/T1ρ/wL spectra
- Ref. spectra

Processing template

Batch process and group spectra

Detect peak intensity change

Stack spectra and generate reports

Mnova file for each sample

Overall results in a spreadsheet

User inspects selected experiments
- 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.

- Minimum Matched Peaks to be Present: 20
- Tolerance for Matching Peaks: 2.0
- Calculate Intensity Changes using Nearest GSD Peak instead of Sum Integration Change
- Use Center of Matched Scout Peak instead of Reference Peak
- Calculate Intensity Changes using Peak Height Instead of Peak Area
- Use Blank STD, T1rho and CPMG Peak Matching to Identify Missing References

Regions of Interest:

Processing:
- Normalize largest peak (by area) in regions of interest to intensity 600
- Ignore peaks with height less than 10% of maximum peak
- The Processing and Analysis tab: setup for STD spectra.
Batch processing finished in about 2 hours.
Verify the screening results
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**  
\[ T_2 = 5\% \]

\[ \Delta I_1 \sim 60\% \quad \Delta I_2 \sim 25\%, \ R \sim 35\% \]

**Example 2**

\[ \Delta I_1 \sim 60\% \quad \Delta I_2 \sim 58\% \quad R \sim 2\% \]
• 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.
Mnova CSP – Chemical shift perturbation analysis
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 $K_D$s.

- 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 $K_D$s in real time.
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
The chemical shift changes along the path from the ref. peak to target peak is measured and normalized:

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

or

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

By default: \( F(H) = 1; F(N) = 0.156; F(C) = 0.185 \). You can change the values in Settings.
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 ($\sigma$) according to

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

<table>
<thead>
<tr>
<th>Assignment</th>
<th>w1</th>
<th>w2</th>
</tr>
</thead>
<tbody>
<tr>
<td>9N-H</td>
<td>105.941</td>
<td>7.562</td>
</tr>
<tr>
<td>11N-H</td>
<td>121.982</td>
<td>7.190</td>
</tr>
<tr>
<td>13N-H</td>
<td>127.733</td>
<td>9.484</td>
</tr>
<tr>
<td>44N-H</td>
<td>122.410</td>
<td>9.051</td>
</tr>
<tr>
<td>46N-H</td>
<td>133.093</td>
<td>8.945</td>
</tr>
<tr>
<td>68N-H</td>
<td>119.736</td>
<td>9.148</td>
</tr>
<tr>
<td>70N-H</td>
<td>126.581</td>
<td>9.097</td>
</tr>
</tbody>
</table>

**Titration file**

<table>
<thead>
<tr>
<th>[P] μM</th>
<th>500</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXPNO</td>
<td>[L]/[P]</td>
</tr>
<tr>
<td>REF</td>
<td>0.00</td>
</tr>
<tr>
<td>2</td>
<td>125</td>
</tr>
<tr>
<td>4</td>
<td>250</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>L2</td>
<td>125</td>
</tr>
<tr>
<td>24</td>
<td>250</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>36</td>
<td>3500</td>
</tr>
<tr>
<td>38</td>
<td>4500</td>
</tr>
</tbody>
</table>
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.
- 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.

Click and drag a peak top to change the peak path.
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.
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/](https://www.affinimeter.com/) for more details.
• H-N HSQC titration spectra, 8 points for each ligand.
• Two ligands.
Conclusions
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.
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.

- Structures & spectra of fragments
- Batch Verify & Batch qNMR
  - Identity verification and solubility test results
- MixDesign script
  - List of mixtures with least peak overlap
- Screen
  - Primary hits
- CSP
  - $K_d$ of monitored peaks

STD/T1ρ/wL/CPMG spectra
HSQC Titration spectra

MixDesign script list of mixtures with least peak overlap

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.
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?