Build Curated and Annotated HRAM MS^n Spectral Libraries To Aid In Unknown Structure Elucidation

Caroline Ding[1], Kate Comstock[1], Seema Sharma[1], Mark Sanders[1], Michal Raab[2],
[1]Thermo Fisher Scientific, San Jose, CA, USA 95134, [2]HighChem, Bratislava, Slovakia

ABSTRACT

Purpose: Validate new library searching algorithms and structure ranking algorithm mzLogic™ for local proprietary MS^n spectral libraries to aid in structure elucidation of unknown drug metabolites, impurities or degradants.

Methods: Ten Sildenafil drug standards were infused onto the Thermo Scientific™ Orbitrap ID-X™ Tribrid™ mass spectrometer to obtain MS^n spectral trees for building a spectral library. LCMS data were collected with a mixture of four sildenanfil standards to test library searching, mzLogic and unknown structure elucidation. Mass Frontier™ 8.0 software was used for the complete workflow including library building, data processing, identification and structure elucidation.

Results: Curated HRAM MS^n spectral library which contains substructure annotations helps to elucidate the structures of true unknowns. mzLogic utilizes the spectra and structure information in the local library for unknown structure ranking.

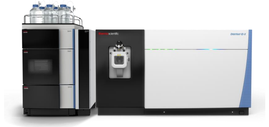
INTRODUCTION

Small molecule structure elucidation is a very challenging and time consuming task. A mass spectral library with extensive MS^n spectral tree and substructure information is a valuable tool for rapid identification of small molecule unknowns and unknown unknown structure characterization. The objective of this work is to demonstrate a complete workflow from building local version of "mzCloud" - HRAM MS^n spectral library using the Orbitrap ID-X Tribrid mass spectrometer for in-house proprietary compounds with automated curation and structure annotations in Thermo Scientific™ Mass Frontier™ structure identification software 8.0. The MS^n spectral library with structure annotations not only enables library searching to quickly identify unknown compounds in the library, but also enables substructure matching. The new mzLogic structure ranking algorithm is able to use the spectral knowledge in the library, rank possible structure candidates based on a combination of spectral similarity and common substructure overlapping, which results in quick and confident structure identification of unknown drug metabolites, impurities or degradants.

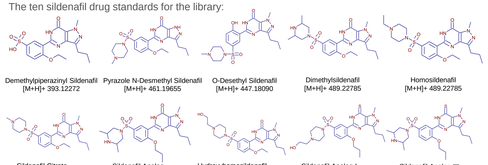
MATERIALS AND METHODS

Data Acquisition

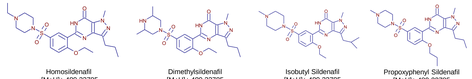
An Orbitrap ID-X Tribrid mass spectrometer coupled with a Thermo Scientific™ Vanquish™ UHPLC system were used for data acquisition. As proof of concept, a library was created for ten sildenafil drug analog standards. Compounds were infused by electrospray ionization and MS^n data collected with varying collision energies for multiple fragmentation activation types. Four sildenafil standards with same mass (two of them were in the library, and two of them were not in the library) were mixed and analyzed by LC/MS^n.



The ten sildenafil drug standards for the library:



The four sildenafil compound standards in the LCMS mixture:



Methods

The library compounds were acquired on the Thermo Fisher Scientific Orbitrap Tribrid ID-X mass spectrometer with Tune 3.1 software which includes library builder templates for infusion and LCMS.

The library template automatically acquires multiple collision energies, multiple spectra for each MS^n stages in both HCD and CID, and goes up to MS^4. It uses assisted collision energy (15,30,45,60,75) for the branch leading to CID MS^3. The purpose is to capture fragmentation data at multiple collision energy levels and fragmentation modes in the library to allow confident matching with LCMS fragmentation data.

The four LCMS standard compounds were first run individually to obtain the retention times. They were then mixed into one vial and run with the same LCMS condition. Fragmentation data for the four 489 compounds included up to MS^4 : MS^2 at HCD stepped collision energies 40,60,90; MS^3 at HCD 40 and MS^4 at CID 40.

Data Analysis

Mass Frontier 8.0 software was used to build the curated and annotated MS^n spectral library. The DICD (direct infusion component detection) algorithm automatically compiles HCD, CID spectra for each precursor ions at each MS^n stages into a spectral tree. The MS^n spectral trees were curated in the Curator module in an automated fashion for each compounds and saved into a local library.

Figure 1. The Curator module in Mass Frontier 8.0

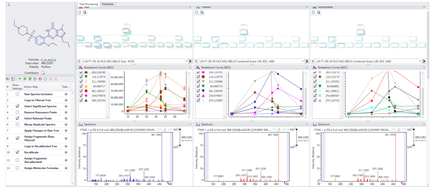
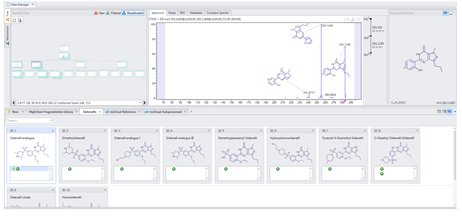


Figure 2. The local MSn spectral library in Data Manager with precursor structures and fragment annotations at each MSn stages.



LCMS data of the 4 standards was processed in Mass Frontier 8.0 by JCD component detection and spectral tree deconvolution, followed by library-searching including identity, tree search, similarity forward, identity substructure, subtree search, and mzLogic structure ranking.

Figure 3. JCD detected four components with m/z of 489.2285 which corresponded to the four sildenafil compounds in the LCMS mixture.

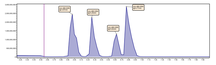


Figure 4. Deconvoluted spectral tree in Mass Frontier 8.0 for each of the 489 compounds with up to MS^4 fragmentation spectra from the Orbitrap ID-X.

RESULTS

Compound Identification with Library Searching

The four components with spectral trees were searched against the local Sildenafil compound library in Chromatogram Processor module. Mass Frontier 8.0 includes new library searching types, new and improved matching algorithms that are trained on the extensive mzCloud database with real fragmentation data.

Below is a summary of the library search types, used stages and constraints and use in Mass Frontier 8.0.

Table 1. Library Search Types in Mass Frontier 8.0

Search Types	Used Stages and Constraints	Use
Identity	• Compares the MS^2 library spectra against the MS^2 query spectra; uses best confidence match score calculation. • The MS^2 precursor ions must match.	Compound Identification
Identity Substructure	• Compares any MS^n library spectra against any MS^n query spectra; uses best confidence match score calculation. • The precursor ions at any MS^n stage must match.	Substructure Identification
Similarity (Forward and Reverse)	• Compares the MS^2 library spectra against the MS^2 query spectra, uses best confidence match score calculation. • The MS^2 precursor ions do not have to match.	Identify structurally similar compounds
Tree Search	• Compares any MS^n library spectra against any MS^n query spectra; takes into account of the whole MS^n hierarchy; uses aggregated tree match score calculation. • The MS^2 precursors for the query spectrum and the library spectrum must match.	Compound Identification **with increased specificity**
Subtree Search	• Compares any MS^n library spectra against any MS^n query spectra; takes into account of the MS^n subtree hierarchy; uses aggregated subtree match score calculation. • The precursor ions at any MS^n stage must match.	Substructure Identification **with increased sensitivity**

Compound Identification

For the two 489 compounds (RT 5.97min and 6.25min) that are in the local Sildenafil library, Mass Frontier 8.0 returned the right compound with 99 confidence match scores for both Identity search and Tree Search.

Figure 5. Compound matches with Identity search.

Name	Scan No.	Precursor m/z	Match	MSn	t_R (min)	Abundance	Annotation Sources
Components							
Component 1	1569	489.2285	99 Homosildenafil	4	5.966	2,205,419,520	Identity
Component 2	1732	489.2284	99 Dimethylsildenafil	4	6.254	1,925,246,848	Identity
Component 4	1937	489.2285	57 Dimethylsildenafil	4	6.620	794,834,816	Identity
Component 6	2024	489.2284	58 Dimethylsildenafil	4	6.766	1,999,293,184	Identity

Figure 6. Compound matches with tree search.

Name	Scan No.	Precursor m/z	Match	MSn	t_R (min)	Abundance	Annotation Sources
Components							
Component 1	1569	489.2285	99 Homosildenafil	4	5.966	2,205,419,520	Tree Search
Component 2	1732	489.2284	99 Dimethylsildenafil	4	6.254	1,925,246,848	Tree Search
Component 4	1937	489.2285	57 Dimethylsildenafil	4	6.620	794,834,816	Tree Search
Component 6	2024	489.2284	58 Dimethylsildenafil	4	6.766	1,999,293,184	Tree Search

MS^n Tree search was able to separate out the two library compound hits more in match scores compared to identity search which is MS^2 vs MS^2 only. Comparison of identity search vs tree search for component #1 are shown below.

Figure 7. The two library hits and match scores for component #1 with identity search (MS^2 vs MS^2 only).

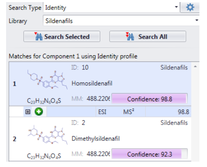
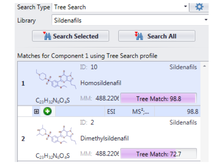


Figure 8. The two library hits and tree match scores for component #1 with MS^n tree search (MS^n vs MS^n).



Similarity and Substructure Matching

For the other two 489 compounds (RT 6.62min and 6.77min) that are not in the Sildenafil library, identity or tree search gave 57, 58 confidence scores due to no exact compound matches in the Sildenafil library which is expected. The next step is to try the similarity, identity substructure and subtree search to identify substructure matches in the library.

With similarity search, we got confidence score of 38 for both of these two 489 compounds. Since similarity search compares query MS^2 vs library MS^2 only, it did not give good substructure matches from the library.

Figure 9. Similarity forward search results for the two 489 compounds at RT 6.62min and 6.77min.

Name	Scan No.	Precursor m/z	Match	MSn	t_R (min)	Abundance	Annotation Sources
Components							
Component 1	1569	489.2285	99 Homosildenafil	4	5.966	2,205,419,520	Tree Search
Component 2	1732	489.2284	99 Dimethylsildenafil	4	6.254	1,925,246,848	Tree Search
Component 4	1937	489.2285	38 Dimethylsildenafil	4	6.620	794,834,816	Similarity Forward
Component 6	2024	489.2284	38 Dimethylsildenafil	4	6.766	1,999,293,184	Similarity Forward

On the other hand, identity substructure and subtree searches which match MS^n vs MS^n both gave good substructure matches for these two 489 compounds, matching substructure of O-Desethyl Sildenafil in the library. The substructure search results are consistent with the common substructure overlap between these two 489 compounds (**Isobutyl Sildenafil** @RT6.62min and **Propoxyphenyl Sildenafil** @6.77min) and **O-Desethyl Sildenafil** in the library.

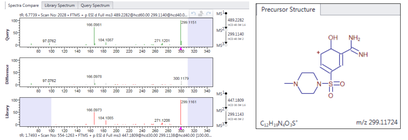
Figure 10. Identity substructure search results for the two 489 compounds at RT 6.62min and 6.77min.

Name	Scan No.	Precursor m/z	Match	MSn	t_R (min)	Abundance	Annotation Sources
Components							
Component 1	1569	489.2285	99 Homosildenafil	4	5.966	2,205,419,520	Tree Search
Component 2	1732	489.2284	99 Dimethylsildenafil	4	6.254	1,925,246,848	Tree Search
Component 4	1937	489.2285	90 O Desethyl Sildenaf	4	6.620	794,834,816	Identity Substructure
Component 6	2024	489.2284	90 O Desethyl Sildenaf	4	6.766	1,999,293,184	Identity Substructure

Figure 11. Subtree search results for the two 489 compounds at RT 6.62min and 6.77min.

Name	Scan No.	Precursor m/z	Match	MSn	t_R (min)	Abundance	Annotation Sources
Components							
Component 1	1569	489.2285	99 Homosildenafil	4	5.966	2,205,419,520	Tree Search
Component 2	1732	489.2284	99 Dimethylsildenafil	4	6.254	1,925,246,848	Tree Search
Component 4	1937	489.2285	90 O Desethyl Sildenaf	4	6.620	794,834,816	Subtree Search
Component 6	2024	489.2284	89 O Desethyl Sildenaf	4	6.766	1,999,293,184	Subtree Search

Figure 12. Example of substructure match on the MS3 level, indicating these two 489 compounds share this common substructure from the library..



mzLogic Structure Ranking

mzLogic search algorithm in Mass Frontier 8.0 utilizes the fragmentation data and structures in the spectral library to rank the possible structure candidates for a true unknown compound that is not in the library. The ranking is based on spectral similarity and maximum common substructure overlap. For this study, we proposed nine structure isomers for mzLogic ranking for the 489 compounds @ 6.62min and @6.77min. The correct structures were ranked among the top scored candidates.

Figure 13. For the 489 compound @6.62min, the correct structure was ranked #2 and had same mzLogic score as rank #1.

Figure 14. For the 489 compound @6.77min, the correct structure was ranked #2 and had same mzLogic score as rank #1.

CONCLUSIONS

- Small molecule unknown structure elucidation is a very time and resource consuming task. If you have a lot of in-house standard compounds, building a curated and annotated HRAM MS^n spectral library can be a very valuable tool for compound identification and identifying true unknowns that are similar or share common substructures to the compounds in the library.
- mzLogic can utilize the fragmentation spectra knowledge in the spectral library to quickly narrow down the list of possible structure candidates for further validation. It is an innovative approach for unknown structure characterization utilizing the latest library searching technology.

ACKNOWLEDGEMENTS

We would like to thank Health Canada for supplying the sildenafil standard compounds.

TRADEMARKS/LICENSING

PO65519-EN0519S