High Resolution Mass Spectrometry (HRMS) Library spectra collection, review, and approval

Summary

The Division of Laboratory Sciences (DLS) of the National Center for Environmental Health has created a library to support identification of exposure to chemicals of public health concern. The DLS HRMS Library includes toxins, drugs of abuse, and other environmental chemicals, as well as select corresponding biomarkers. It is posted for public access at High-Resolution Mass Spectral Libraries for Drug and Toxin Analysis | CTTL | CDC. DLS collected the library spectra on three instrument platforms—an Agilent 6545/6546 QTOF, a Sciex 6500+ TripleTOF, and a Thermo Exploris 480 Orbitrap—to enable broad use by laboratories with different resources. Additionally, the spectra were curated and evaluated as described below to ensure high quality and confirm reproducibility of each spectrum.

HRMS library spectral matching is a tool to identify compounds. To perform library matching, the reference material for the compound of interest is fragmented to generate a fragmentation spectrum. This spectrum is generally reproducible for a given compound at a given energy level; therefore, it can serve as a "molecular fingerprint" for the compound under those conditions. These libraries can be built in-house or procured from outside sources. Experimental fragmentation spectra can be compared against these library spectra and scored based on the degree that fragment masses and abundances match between the experimental and library data.

Fragmentation Spectra Collection, Selection, and Curation

Prior to data collection, DLS calibrated each HRMS instrument per the manufacturer's recommendations. For all instruments, this included a daily external calibration. For the Agilent and Thermo HRMS platforms, an additional internal calibrant was co-infused into the mass spectrometer along with the sample solution for internal calibration. We employed liquid chromatography, as opposed to direct infusion, to collect the spectra to best simulate expected conditions and to collect internal retention time (RT) information on a variety of columns and methods. Depending on class of compound, several different columns and chromatography were used (See Appendix). Chromatographic performance (RT and peak shape) was confirmed by a function check, consisting of known analytes appropriate for the chromatography selected, that DLS analyzed before spectra data collection. DLS verified that this function check passed identification parameters before analysis of the compounds.

We diluted all reference compounds to 100 ng/mL in a solvent appropriate for both the compound and chromatography (typically water or methanol) and injected in each instrument at least four separate times: one injection to collect fragmentation spectra to add to the library and at least three subsequent injections for evaluation of library match scores. On all instruments fragmentation was performed on all compounds at a wide range of discrete collision energy levels, typically but not exclusively 20, 40, 60, 80, and 100 eV. On the Thermo Exploris stepped collision data was also collected, typically at 20, 40, 60 eV and 40, 60, 80 eV. On the Sciex 6600+, ramped collision data was occasionally collected.

We extracted fragmentation spectra closest to the apex of the chromatographic peak and added them to the library. To account for experimental variability in the spectra, DLS curated where possible based on the tools provided by each vendor.

- **Agilent:** ions in the fragmentation spectra were annotated based on the precursor-ion's known molecular formula, and only product-ions that could be assigned a molecular formula were included in the library spectra. In addition, the mass of the fragments was adjusted to match the exact mass of the calculated molecular formula.
- **Thermo:** the fragmentation spectra were recalibrated based on the precursor-ion's molecular formula using Thermo MZVault's default recalibration parameters. An automatic threshold was then applied to remove low level ions that may be noise. On occasion, manual thresholds were applied if the automatic threshold was deemed to be too lenient.
- **Sciex**: no curation is available for Sciex data.
- **NIST:** data is based off library collected on Thermo instrumentation and all curation applied to that data.

For each compound in the library the CAS (where available), InCHI key, SMILES, and IUPAC name were added in the designated columns; if no designated column existed, we entered the data into additional notes associated with the entry. While RT is not included in any library, it is available upon request.

Fragmentation Spectra Review

For all replicates, identification parameters in the data processing software were set to only match experimental data with library data collected at the same collision energy, so that the reproducibility of the fragmentation spectra can be evaluated. A library score cutoff of 75 is used for the library. Library scores of 73 have previously been found to provide sensitivity of 93% and specificity of 91% in a 169-compound drug screen test, but a slightly higher cutoff was chosen to take into account variability in instruments and matching algorithm.[1] To determine if library scores close to the cutoff were statistically above the cutoff, a one sample, two-sided, t-test was employed. The *t* value is calculated by:

$$t = \frac{x - u}{s/\sqrt{n}}$$

where *x* is the library score average, *u* is the cutoff value (75), *s* is the library score standard deviation, and *n* is the number of replicates. After statistical analysis, spectra were accepted or rejected based on the following criteria:

- If $\bar{x} < 70$, reject
- If $p > 0.05$, accept
- If $p \leq 0.05$ or p = error
    - If $\bar{x} \geq 75$ accept,
    - Else reject

If library spectra did not pass the t-test, new spectra from the replicates were added and data reprocessed. If after reprocessing the replicates pass the t-test than the new spectra were retained in the library for future use. In some cases, no spectra were collected at select energy levels that allowed

for consistent identification that passed the t-test. In those situations, the spectra were removed from the library with the compound having no representative spectra at that collision energy.

Appendix: Chromatography and Column Information

Table 1. Chromatography information for drug compounds.

Column information:	50x2.1 mm, 1.7 µm C18 column
Injection volume:	15 µL
Flow rate:	0.40 mL/min
Mobile phase A:	10 mM ammonium formate in water
Mobile phase B:	1% formic acid in acetonitrile

Gradient

Time	%A	%B
0.0	95	5
1.0	95	5
10.0	5	95
13.0	5	95
13.4	95	5
15.5	95	5

Table 2. Chromatography information for polar toxin compounds.

Column information:	150x2.1 mm, 1.7 µm BEH Amide
Injection volume:	7.5 µL
Flow rate:	0.30 mL/min
Mobile phase A:	10 mM ammonium formate and 1% formic acid in water
Mobile phase B:	1% formic acid and 2 mM ammonium formate in acetonitrile

Gradient

Time	%A	%B
0.00	15	85
15.00	95	5
15.01	15	85
20.00	15	85

Table 3. Chromatography information for natural toxin compounds.

Column information:	100x3 mm, 2.6 µm Biphenyl column
Injection volume	15 µL
Flow rate:	0.75 mL/min
Mobile phase A:	95:5 water:methanol with 10 mM ammonium formate
Mobile phase B:	5:95 water:methanol with 10 mM ammonium formate

Gradient

Time	%A	%B
0.00	100	0
10.00	0	100
15.01	100	0
18.00	100	0

Table 4. Chromatography information for microcystin compounds.

Column information:	50x2.1 mm, 1.7 µm C18 column
Injection volume:	15 µL
Flow rate:	0.40 mL/min
Mobile phase A:	1% formic acid in water
Mobile phase B:	1% formic acid in acetonitrile

Gradient

Time	%A	%B
0.0	85	15
4.5	25	75
4.6	5	95
5.0	5	95
5.5	85	15
6.5	85	15

References

1. Jennifer M. Colby, Katie L. Thoren, Kara L. Lynch, Optimization and Validation of High-Resolution Mass Spectrometry Data Analysis Parameters, *Journal of Analytical Toxicology*, Volume 41, Issue 1, January-February 2017, Pages 01–05, https://doi.org/10.1093/jat/bkw112

Disclaimers

Results obtained from the use of these libraries may vary based upon instrument type, instrument settings, and specific method procedures. The library spectra by themselves are not intended for screening, monitoring, or diagnostic purposes.

All libraries and the information contained therein provided and covered under this agreement are being provided on an 'as is' basis. Except as expressly set forth herein, the CDC makes no representations, of any kind, either express or implied, with respect to the libraries and expressly disclaims any and all representations of any kind with respect thereto, including any representations of quality or fitness for a particular purpose. Unless otherwise stated, all libraries and related materials are considered to satisfy the quality standards relative to the purpose for which the libraries were collected. However, neither the author nor any part of the federal government can assure the accuracy, reliability, or suitability of the libraries or the information contained therein for a particular purpose. The act of distribution shall not constitute any such warranty, and no responsibility is assumed for a user's application of the libraries or related materials.

The findings and conclusions in this report are those of the authors and do not represent the official position of the Centers for Disease Control and Prevention. Use of trade names is for identification only and does not imply endorsement by the Centers for Disease Control and Prevention, or the US Department of Health and Human Services.