Journal of The American Society for Mass Spectrometry, Volume 33, Issue 4, April 2022, pp. 740-743
DOI: 10.1021/jasms.2c00016

Creating a Searchable Chromatographic Database with the NIST Mass Spectral Search Program

Robert B. Cody[a]*, O. David Sparkman[b], Hannah Moore[c]

[a] JEOL USA, Inc., 11 Dearborn Rd., Peabody MA 01960 USA

[b] University of the Pacific, Stockton, CA 95211 USA

[c]Cranfield Forensic Institute, Cranfield University, Defence Academy of the United Kingdom, Shrivenham, Wiltshire, SN7 8LA, UK

*Corresponding author

[a]E-mail: cody@jeol.com

Abstract:

The NIST Mass Spectral Search Program can be used to create a searchable database of chromatograms. This approach was tested for a small database of chromatograms for gin volatiles and for a database of insect cuticular hydrocarbons with retention times and reconstructed total ion current chromatographic peak areas substituted for *m/z* values and abundance. The *In-source HiRes* Identity search permitted matching of randomly selected chromatograms against the database with good results. This approach is not intended to replace commercial software for chromatographic database management as it does not address the problems of chromatographic alignment or chromatographic deconvolution, but it does provide a method to manage a simple chromatographic database if other options are not available.

Introduction

Commercial software provided with gas chromatography-mass spectrometry (GC-MS) instrumentation provides functions for compound identification using both GC and MS information. However, it is sometimes useful to maintain and search databases of chromatograms alone. Gas chromatograms can be used to identify unknown substances by comparison with chromatograms of known substances or materials ("chromatographic fingerprinting"). Published and on-line databases of gas chromatograms are available [1-4] and application-specific chromatograms are generally available on the websites of vendors of chromatographic supplies. Open-source[5] and commercial software[6] has been developed to analyze data from analytical pyrolysis (Py-GC/MS). A commercial add-on is available for the open-source software[7]. The NIST 20 GC Method / Retention Index Library[8] contains 447,289

Published by the American Chemical Society. This is the Author Accepted Manuscript issued with:
Creative Commons Attribution Non-Commercial License (CC:BY:NC 4.0).
The final published version (version of record) is available online at DOI:10.1021/jasms.2c00016. Please refer to any applicable publisher terms of use.

chromatographic methods and retention index data for 112,253 compounds and provides functions for searching compounds by retention indices. Here, we show that the tools provided with the NIST MS Search software for creation of custom databases of mass spectra can also be used to create searchable chromatographic databases containing information about reconstructed total ion current (RTIC) chromatographic peak-areas and retention times (RT) or retention indices (RI).

Experimental

GC-MS analysis of gin volatiles

For proof of concept, volatile compounds from seven brands of gin purchased at local liquor stores were analyzed with an Agilent 7890B gas chromatograph coupled with a JEOL JMS-T200GC AccuTOF™-GCx-Plus time-of-flight mass spectrometer. The measurement conditions are given in Table S1. The retention times and RTIC chromatographic peak areas for the total ion current chromatogram were exported and saved as text files. A linear alkane retention index standard (Qualitative Retention Index Standard, Catalog number 31080, Restek, Bellefonte PA USA) was measured using the same GC conditions. Although mass spectra were measured for each component, only the RTIC chromatograms were used for this test.

Importing chromatographic data into the NIST Mass Spectral Search Program

Version 2.4 of the NIST Mass Spectral Search Program (Build March 25, 2020) was used for searching chromatograms against the databases created for this work. Text files saved in suitable formats[9] can be imported directly into the NIST Search Program's Librarian tab or the Lib2NIST NIST Library Conversion Tool[10]. To ensure that the fractional part is retained for imported retention data, the NIST Spectrum Import options should specify "Accurate *m/z*" (Figure S1). For this work, the batch processing function in *Mass Mountaineer* software version 6.5.0.0 [11] was used to convert the individual chromatograms into NIST format in a NIST *.MSP file that were imported into the NIST Mass Spectral Search Program (V. 2.4). The NIST Librarian functions were used to create a custom database that was reindexed by using the Lib2NIST NIST Library Conversion Tool. A replicate measurement for one of the gins was measured to be used as a test sample but was not included in the chromatogram database. All of the software referenced herein can be obtained from Diablo Analytical Inc. (Antioch CA, USA).

It is very important to note that custom databases created by using the NIST Librarian functions are not indexed for accurate-*m/z* value searches, and by extension, exact-retention time/index searches. To ensure that a custom database can be searched in the *In-source Hires* search using the Default Presearch, an indexed database must be created by using the Lib2NIST NIST Library Conversion Tool. This should be done in two stages. First, the custom database must be converted to an intermediate *.SDF file by using Lib2NIST. Then the Lib2NIST program must be used in a command line with the appropriate switches to convert the SDF file to a searchable NIST database. (See the example batch file in the Supporting Information) It is strongly recommended to read the Lib2NIST.txt file located in the C:\NISTxx\MSSEARCH folder for details on the use of this tool.

Insect cuticular hydrocarbon chromatograms

A second test database was created by using chromatographic data from a previously published study. Classification of *Sarcophaga* insects based on the chromatographic analysis of cuticular hydrocarbons was described in a separate publication[12]. Retention-time aligned chromatograms from this study were exported into a single Excel spreadsheet with retention time and peak area information for each component. A utility program was created in Microsoft Visual Basic to separate the 195 chromatograms into individual text files for use with *Mass Mountaineer* software. Each text file was saved by the utility program with the name of the species and a numeric identifier. The chromatograms were converted into a NIST-formatted database using *Mass Mountaineer* and the NIST Librarian functions.

NIST MS Search Program screenshots and parameter settings

Figure 1 shows the NIST Search Program Librarian tab with imported chromatograms for the seven gin samples. Retention time was substituted for exact *m/z* and peak area was used in place of relative intensity. The process was repeated by using Kovats retention indices calculated by using a function in *Mass Mountaineer*. A separate database was created containing the 195 chromatograms from Reference 12 for the 11 *Sarcophaga* species.

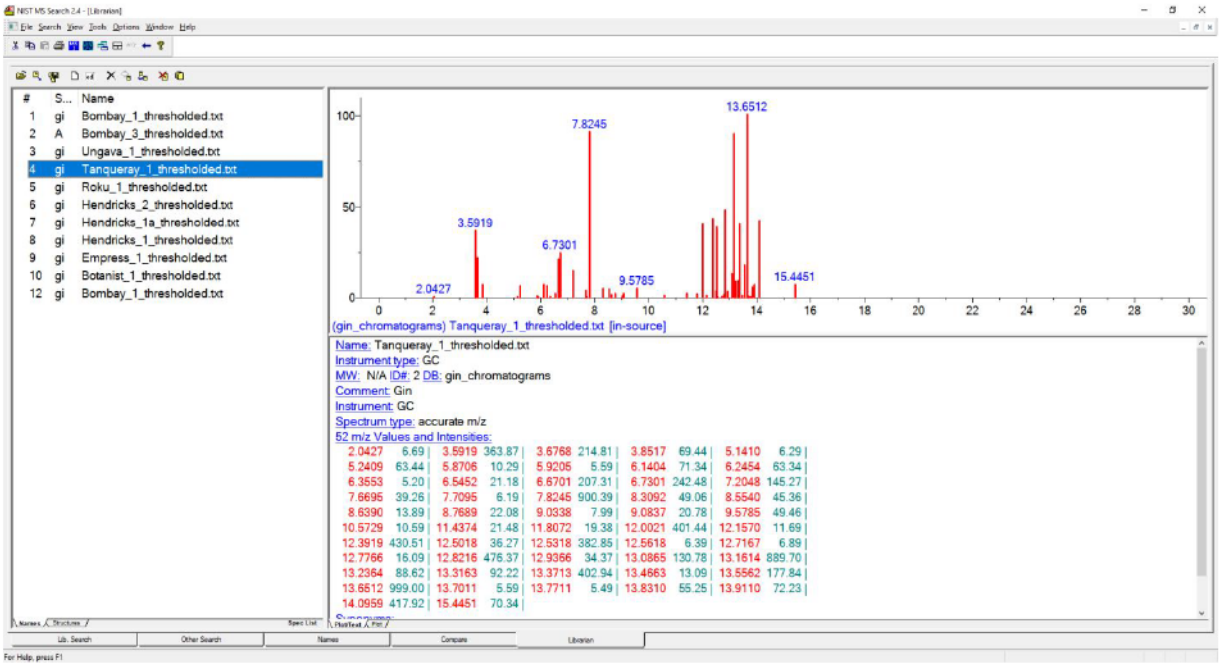


Figure 1. A gin volatiles chromatogram imported into the NIST Mass Spectral Search Program's Librarian tab.

The parameter settings for searching chromatograms are shown in Figure 2.

Of these, the critical parameters are as follows:

- Search type: *In-source HiRes*
- Precursor ion *m/z*: In spectrum
- Spectrum Search Options: Off
- Presearch: Default
- MS/MS Search *m/z* Tolerance: 0.02 (variable, depending on RT/RI precision)

The *In-source HiRes* search can be used for matching of chromatographic retention times. The tolerances on the MS/MS tab for precursor ion *m/z* and product ion *m/z* should be adjusted to match the variability in retention times by substituting retention time for *m/z*. The Default Presearch can only be used for a database that has been indexed by the Lib2NIST utility. If Presearch is used, the MS/MS Search *m/z* Tolerance must be set to a value that reflects the precision of the retention time or retention index measurements.

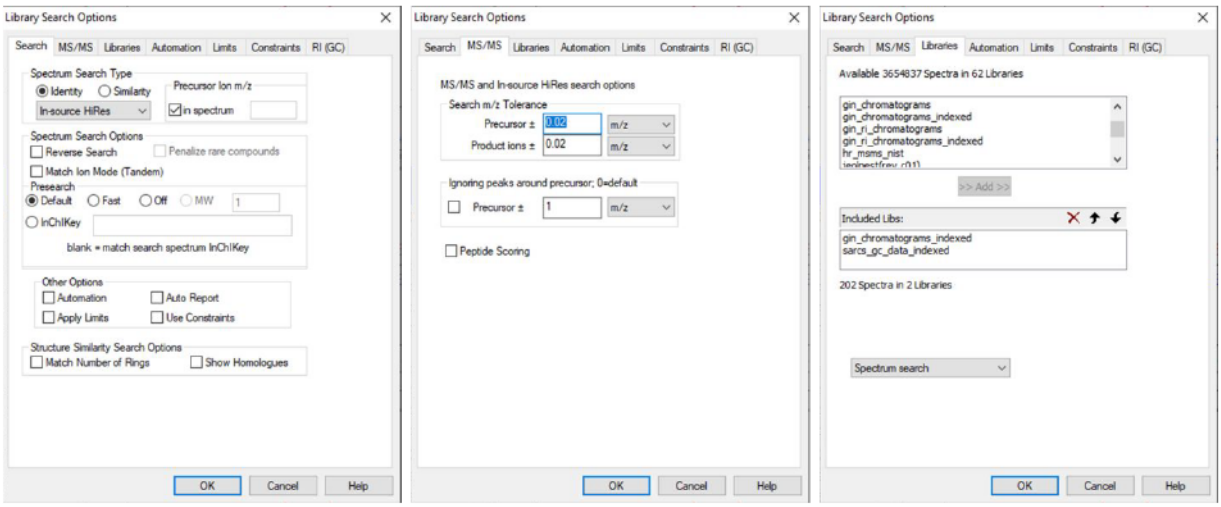


Figure 2. NIST Search parameter settings for matching chromatograms using retention times. The "Search *m/z* Tolerance" values must be adjusted to reflect the retention time or retention index reproducibility.

Results and Discussion

Searching the replicate measurement of the gin volatiles for Bombay Sapphire Gin gave the correct assignment as the best match (Table S1) with Empress gin as the second-best match. The next-best matches had similar, but not identical volatile profiles. The procedure was repeated by creating a database using retention indices with the same results.

Next, individual chromatograms from the *Sarcophaga* chromatogram database were selected on the Names tab from the new database and searched against the other chromatograms in the database. The top hit was the identity match. In all cases, the next best matches were the correct species and gender (for samples having replicates for both male and female). The Search parameters are shown in Figure 2.

Figure 3 shows the match results for a the chromatogram for a randomly selected *S.argyrostoma* sample. The highest-scoring 8 matches are for *S.argyrostoma* with the top match being the identity match.

Figure 3. Chromatographic matches for a randomly selected *S. argyrostoma* sample.

In all cases, searching the database for matching chromatograms showed the highest-scoring matches for the correct species. The three samples (*S. carnaria*, *S. melanura* and *S. pumila*) that were misclassified by Support Vector Machine analysis in the full dataset of males and females (Reference 12, Table S6) gave correct matches when searched by this method. The highest-scoring 20 matches for those samples are given in Tables S3, S4, and S5. These results suggest that chromatographic database matching is a viable approach to insect identification by cuticular hydrocarbon profiling that can complement and/or validate the results of chemometric analysis of the same data.

The NIST Mass Spectral Search software provides a broad set of tools for database creation, management, and searching. The wide availability of this software in mass spectrometry laboratories makes it a convenient approach to managing chromatographic databases in the absence of dedicated chromatography database software. The procedure described here for indexing chromatographic databases to enable *In-Source Hires* searching with the Default Presearch is the same procedure required to index mass spectral data for *In-Source Hires* searches. An example batch file for indexing a database is included in the Supporting Information.

Conclusion

The NIST Mass Spectral Search program can be used to create and search databases for gas chromatographic data. The method can potentially be applied to liquid chromatograms or any spectral data that can be represented as x,y data pairs. Possible applications might include rpLC-MS/MS of proteomics data, chemical liability mapping of biopharmaceuticals, rpLC/MS/MS of metabolites, if the appropriate databases are/were available. This approach does not address the problem of chromatographic alignment, which must be managed by the GC software or third-party software. Further, it does not consider differences in peak width or peak shape. This method also does not

compensate for intra- and interlaboratory variations in retention time unless they fall within a specified error tolerance, or if the databases are created by using retention index data. Regardless, this illustrates a potentially useful extension of the NIST Mass Spectral Search program to other types of data.

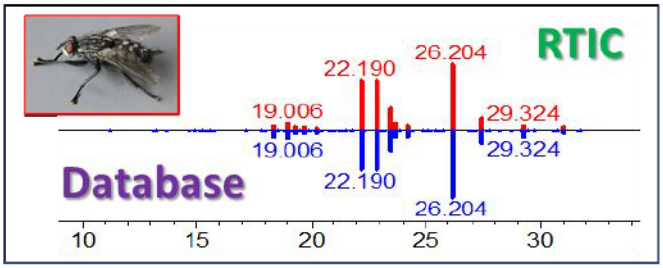
Supporting Information Available

Measurement conditions for the gin chromatograms, spectrum import options for importing accurate m/z data into the NIST Search Program, gin brands used for feasibility test, search results for Bombay Blue Sapphire Gin volatiles, highest-scoring 20 chromatographic match results for an S. vagans sample, highest-scoring 20 chromatographic match results for a female S. variegata sample, highest-scoring 20 chromatographic match results for a second female S. variegata sample, example Windows *.bat file for indexing a database with Lib2NIST.

Conflict of Interest Statement

Robert B. Cody is employed by JEOL USA, Inc., the supplier of the AccuTOF GC-Alpha mass spectrometer, and is the author of the *Mass Mountaineer* software. O. David Sparkman is a contractor to NIST responsible for end-user and distributor support.

For Table of Contents Use Only



For "Creating a Searchable Chromatographic Database with the NIST Mass Spectral Search Program"

The figure shows a database match for a chromatogram of the cuticular hydrocarbons of *Sarcophaga variegata*.

References

1. RTI International, Web Accessible Mass Spectral Database: DART-TOF, Mass Spectrometry, FTIR, and Chromatography Spectra / Chromatograms. *https://www.forensicdb.org* (accessed)
2. Lennard, C.J., Tristan Rochaix, V., Margot, P., Huber, K.: A GC–MS database of target compound chromatograms for the identification of arson accelerants. Science & Justice. **35**, 19-30 (1995)
3. Restek Searchable Chromatogram Library. *https://www.restek.com/en/pages/chromatogram-library/* (accessed: Jan. 3, 2022)
4. ACD/Chromatography Applications Database. *https://www.acdlabs.com/products/adh/chrom/chrom_apps_db/index.php* (accessed: Jan. 3, 2022)
5. OpenChrom Data Analysis System. *https://lablicate.com/platform/openchrom* (accessed: Jan. 3, 2022)
6. Mass spectrum search software for rapid identification of polymers and additives. *https://www.frontier-lab.com/products/multi-functional-pyrolysis-system/17821/* (accessed: Jan. 3, 2022)
7. ChromIdent. Chromatogram Matching and Fingerprinting. *https://lablicate.com/marketplace/extensions/chromident* (accessed: Jan. 3, 2022)
8. NIST 20 GC Method / Retention Index Library. *https://chemdata.nist.gov/dokuwiki/doku.php?id=chemdata:ridatabase* (accessed: Jan. 3, 2022)
9. NIST/EPA/NIH Mass Spectral Database (NIST 11) and NIST Mass Spectral Search Program (Version 2.0g) User's Manual. *https://chemdata.nist.gov/mass-spc/ms-search/docs/Ver20Man_11.pdf* (accessed: Jan. 3, 2022)
10. Library Conversion Tool. *https://chemdata.nist.gov/mass-spc/ms-search/Library_conversion_tool.html* (accessed: Jan. 3, 2022)
11. massmountaineer.com, Mass Mountaineer. Spectral Inspection, Interpretation, & Reporting Tools. *massmountaineer.com* (accessed: Jan. 3, 2021)
12. Moore, H.E., Hall, M.J.R., Drijfhout, F.P., Cody, R.B., Whitmore, D.: Cuticular hydrocarbons for identifying Sarcophagidae (Diptera). Scientific Reports. **11**, 7732 (2021)

https://dspace.org/

Creating a searchable chromatographic database with the NIST mass spectral search program

Cody, Robert B.; Sparkman, O. David; Moore, Hannah

2022-04-06
American Chemical Society (ACS)
https://doi.org/10.1021/jasms.2c00016
Cody RB, Sparkman OD, Moore H. (2022) Creating a searchable chromatographic database with the NIST mass spectral search program, Journal of The American Society for Mass Spectrometry, Volume 33, Issue 4, April 2022, pp. 740-743

Downloaded from a DSpace repository. https://dspace.org/