Poster Reprint

ASMS 2023
Poster number ThP 104

Generating MassBank-ready files from accurate mass library spectra: a proof-of-concept study

Andrew D. McEachran[1], Tristan Chutka[1], Alex Chao[2], Gregory Janesch[3], Elin M. Ulrich[2], Jon Sobus[2], Antony Williams[2]

[1]Agilent Technologies, Inc. Santa Clara, CA

[2]Chemical Characterization and Exposure Division (CCED), Office of Research and Development, U.S. EPA, RTP, NC

[3]ORAU Student Services Contractor to Center for Computational Toxicology and Exposure, Office of Research and Development, U.S. EPA, RTP, NC

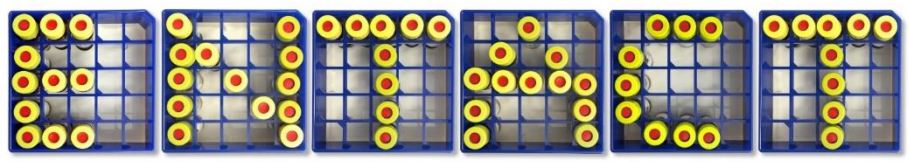
Introduction

Despite advances in structure elucidation tools, empirical high-resolution MS/MS spectra remain the standard to confidently identify unknowns via HRMS targeted screening and non-targeted analysis workflows. Commercial libraries make up a large percentage of available spectra and boost identification confidence due to their high curation level but are often locked and restricted to commercial software solutions. The introduction of open spectral repositories, such as MassBank, MassBank of North America (MoNA), and GNPS in accessible formats has substantially increased the available spectra for library matching and identification workflows. To see continued growth and usage of openly available spectra, contribution must be encouraged and simple. We introduce a new workflow to generate MassBank files from accurate mass library spectra within a single software application.

Experimental

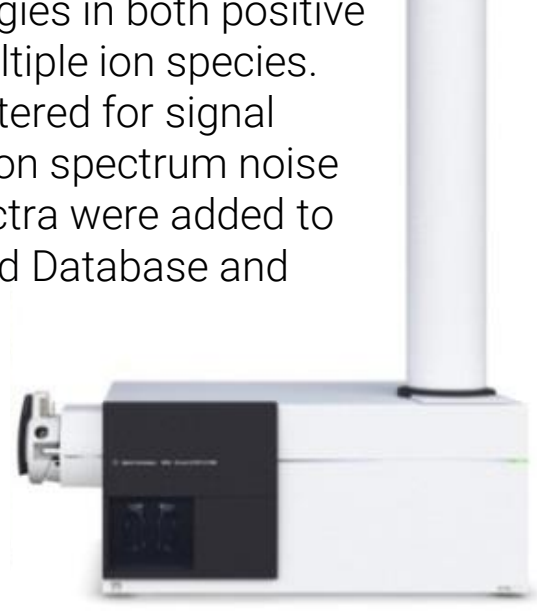
Data acquisition

As part of EPA'S Non-targeted Analysis Collaborative Trial (ENTACT)[1], individual chemical samples from the EPA ToxCast Screening Library were provided to Agilent Technologies.



EPA's Non-Targeted Analysis Collaborative Trial

Spectra were acquired via Agilent 6545 Q-TOF coupled with an Agilent 1290 UHPLC system. Acquisition was optimized to ensure quality spectral collection and spectra were acquired at 10, 20, and 40 V collision energies in both positive and negative ion mode for multiple ion species. Acquired spectra were then filtered for signal intensity and reviewed based on spectrum noise and impurities. Reviewed spectra were added to an Agilent Personal Compound Database and Library (PCDL).



Experimental

Software Application Details

Agilent ChemVista is a new software application that can support a variety of spectral library and compound database formats. It is a server-client application package built on the extensible .NET Framework, and utilizes an advanced relational database system to support concurrent read-write access to large numbers of compounds.

MassBank .txt file creation

The PCDL containing a subset of the acquired ToxCast library spectra was imported into ChemVista. Once imported, chemical structure identifiers (InChI String, InChIKey, and SMILES) were generated (Figure 1) and data standardized prior to storage alongside compound metadata and spectral information to support downstream data needs (Figure 2).

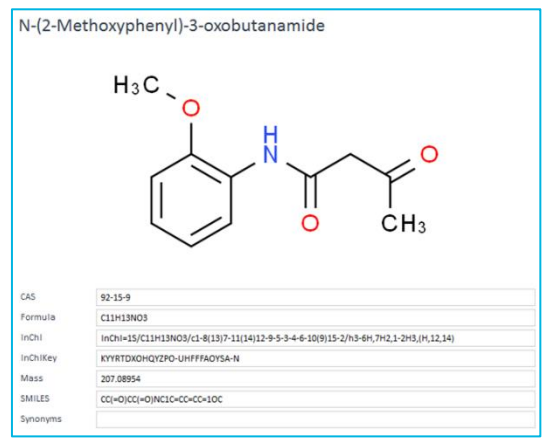


Figure 1. Structural information generated on import.

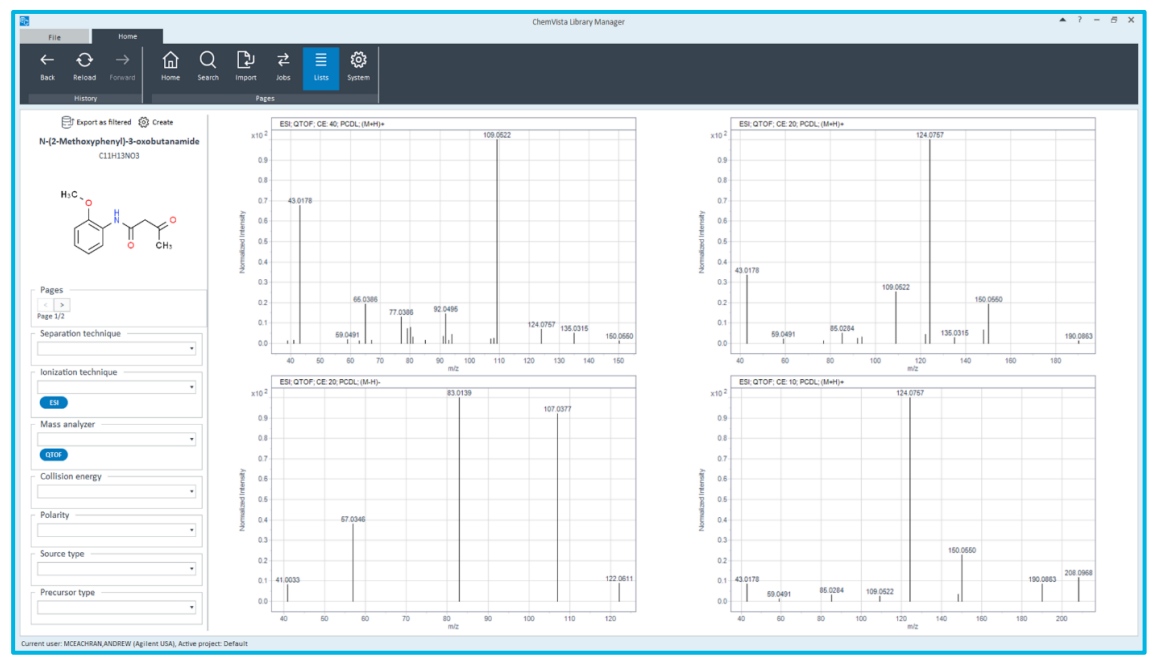


Figure 2. Spectral view within the ChemVista application. Spectra can be filtered and/or exported directly from this view

MassBank .txt file creation

330 compounds containing ~1000 spectra were selected for export. On the export page, the MassBank (.txt) format was selected and Contributor ID and Accession ID details were added to satisfy the MassBank record format specifications (Figure 3).[2] Upon completion of the export, all spectra in individual .txt files were included in a zip file for download from the application server. Individual .txt files included expected metadata, structural identifiers and spectral information (Figure 4).

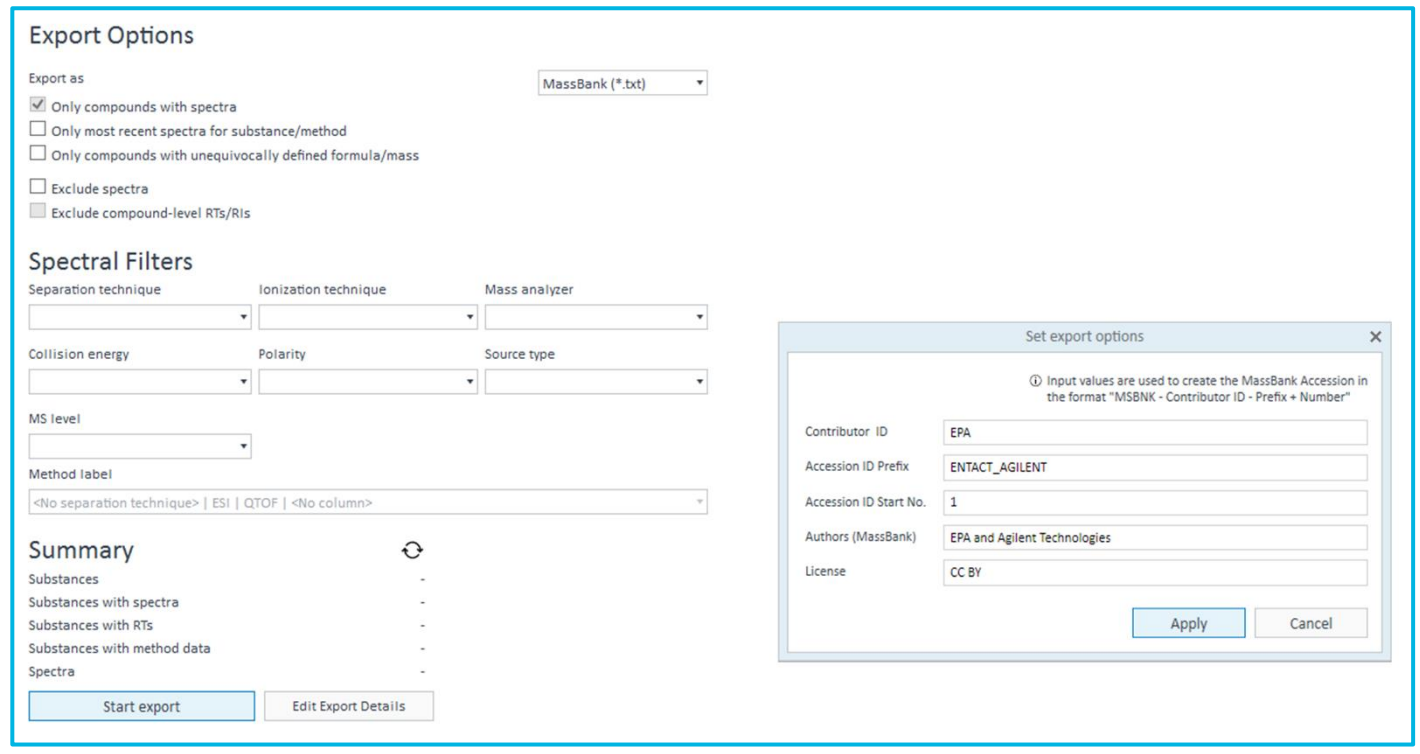


Figure 3. Defining MassBank-specific export options from the export page to satisfy the MassBank Record Format specifications.

```
ACCESSION: MSBNK-EPA-ENTACT_AGILENT000006
RECORD_TITLE: N-(2-Methoxyphenyl)-3-oxobutanamide; ESI-QTOF; MS2; CE: 10; [M+H]+
DATE: 2023.05.01
AUTHORS: EPA and Agilent Technologies
LICENSE: CC BY
CH$NAME: N-(2-Methoxyphenyl)-3-oxobutanamide
CH$COMPOUND_CLASS: N/A
CH$FORMULA: C11H13NO3
CH$EXACT_MASS: 207.089543289
CH$SMILES: CC(=O)CC(=O)NC1C=CC=CC=1OC
CH$IUPAC: InChI=1S/C11H13NO3/c1-8(13)7-11(14)12-9-5-3-4-6-10(9)15-2/h3-6H,7H2,1-2H3,(H,12,14)
CH$LINK: CAS 92-15-9
CH$LINK: INCHIKEY KYYRTDXOHQYZPO-UHFFFAOYSA-N
CH$LINK: PUBCHEM CID:7078
AC$INSTRUMENT: N/A
AC$INSTRUMENT_TYPE: ESI-QTOF
AC$MASS_SPECTROMETRY: MS_TYPE MS2
AC$MASS_SPECTROMETRY: ION_MODE POSITIVE
AC$MASS_SPECTROMETRY: IONIZATION ESI
AC$MASS_SPECTROMETRY: COLLISION_ENERGY 10
MS$FOCUSED_ION: PRECURSOR_M/Z 208.0968197407
MS$FOCUSED_ION: PRECURSOR_TYPE [M+H]+
MS$DATA_PROCESSING: WHOLE Agilent ChemVista Version 1.0
PK$SPLASH: splash10-00di-0900000000-0234d150f062b273d24e
PK$NUM_PEAK: 9
PK$PEAK: m/z int. rel.int.
  43.017841 8.613896 86
  59.049141 1.293349 12
  85.028406 3.141906 31
  109.052215 2.625904 26
  124.07569 100.000005 999
  148.07569 3.295823 32
  150.054955 22.7995 227
  190.086255 8.353603 83
  208.09682 11.657475 116
//
```

Figure 4. Completed MassBank files exported from ChemVista (example shown).

MassBank .txt file deposition

Once files are created, the deposition process follows the MassBank standard procedures, including validation of the files and committing data via a pull request to the MassBank GitHub repository. Prior to this, users must register their contributor and institutional information with the MassBank team so that identifying information can be validated.[3]

Results and Discussion

All data were successfully imported from the PCDL file and structural data were generated according to defined structure parameters. ~1000 .txt files were created successfully and run through a local instance of the MassBank validator to determine validity. The MassBank validator was cloned from the MassBank project and run using the provided shell script with default parameters. All files contained the expected structural, method, and spectral information to satisfy the requirements of the MassBank record format. The pull request was accepted by the MassBank development team, and all data appears in the MassBank development branch ready for merging to the master branch. At this time, all data can be accessed via GitHub (Figure 5). Additionally, these spectra as well as additional spectra from MassBank will be used by EPA to support internal non-targeted analysis workflows and software in development.

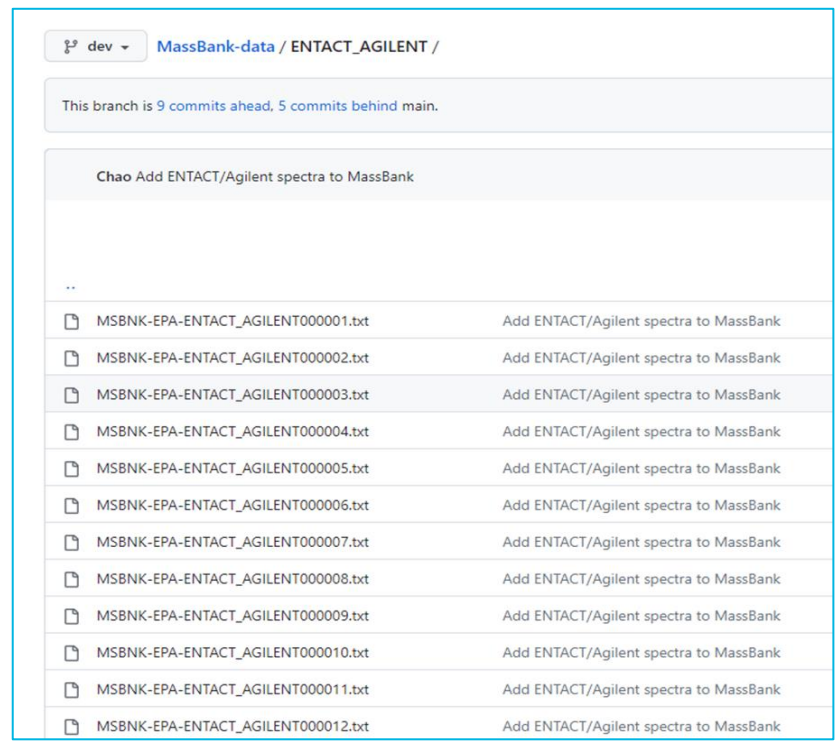


Figure 5. MassBank files merged into the MassBank GitHub (https://github.com/MassBank/MassBank-data)

MassBank provides support, documentation, and access to an R script to generate .txt files from spectra. This process has worked tremendously well for decades, now providing the community with access to >90,000 spectra. With this new software application, users are provided another means to generate .txt files that can be deposited to MassBank. And in this fashion, users can take their data from the Agilent PCDL format and simply and efficiently generate files that are ready to be deposited to MassBank (Figure 6).

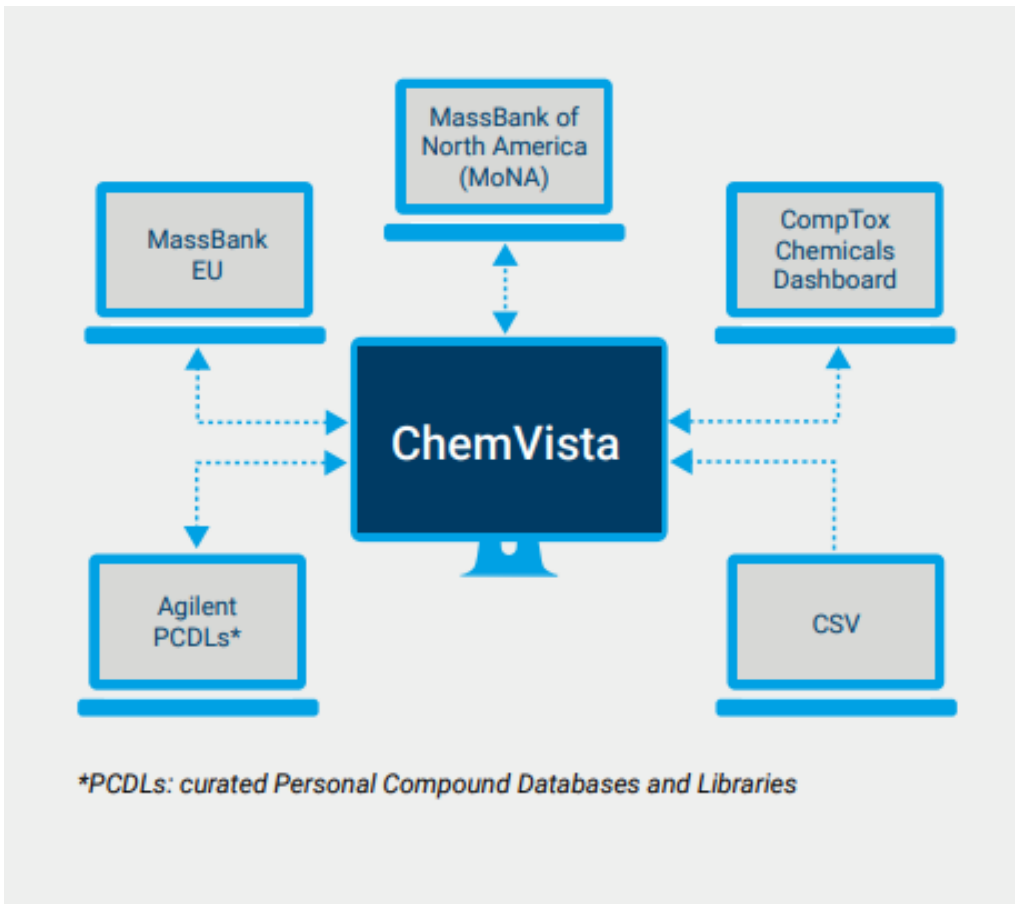


Figure 6. Overview depiction of Agilent ChemVista and the types of supported data.

Conclusions

- Agilent ChemVista presents a new option for mass spectrometry users to create MassBank .txt files from their library spectra
- Spectra were imported into Agilent ChemVista from the PCDL format and chemical structure data was generated to support downstream needs
- Spectra were exported into validated MassBank .txt files prior to merging into the MassBank repository where they are now publicly available

Acknowledgements

The authors would like to acknowledge Tobias Schulze, Emma Schymanski, Rene Maier, and the MassBank team for their ongoing support and assistance with file validation. The authors would also like to acknowledge Tarun Anumol and Craig Marvin from Agilent Technologies for their efforts in the ENTACT data analysis. This poster may not reflect Agency policy.

References

[1]Ulrich, et al. 2019. Anal Bioanal Chem. doi: 10.1007/s00216-018-1435-6

[2]https://github.com/MassBank/MassBank-web/blob/dev/Documentation/MassBankRecordFormat.md

[3]https://massbank.github.io/MassBank-documentation/).

https://www.agilent.com/en/promotions/asms

DE29771757

Published in USA, May 31,2023

Agilent
Trusted Answers