Technical Overview

Agilent
Trusted Answers

Agilent ChemVista Library Manager

An innovative tool for confident identification

Introduction

Despite advances in structure elucidation tools, empirical high-resolution MS/MS (and full scan) spectra remain the standard to confidently identify unknowns via targeted screening and nontargeted analysis workflows. Generating spectral libraries, however, is a costly and time-consuming process. Even when complete, these libraries are often siloed and not in easily accessible formats, limiting the scope and identifiable chemical space of screening workflows. Integrating spectra from multiple sources into an identification workflow is necessary for expanding the capabilities of confident identification. Agilent ChemVista software is a standalone software application that can integrate compound, retention time, and spectral information from multiple sources to facilitate identification workflows within Agilent MassHunter data analysis applications and beyond. Efficient library management is accomplished through list organization, intuitive searching, and filtering options throughout a streamlined user interface. Optionally included with the software are extensive, curated libraries for the metabolomics, food safety, environmental, material, and forensic toxicology markets. This technical overview describes the use of ChemVista software to improve the efficiency and productivity of high-resolution mass spectrometry (HRMS) spectral management.

Extending data access with ChemVista

Import

Integrating compound and spectral data from internal and third-party sources is critical to expanding identification capabilities. As a standalone application, ChemVista supports the import of data from the Agilent Personal Compound Database and Library (PCDL) format, MassBank EU, MassBank of North America (MoNA), and the U.S. EPA CompTox Chemicals Dashboard. Details and access links are presented in Table 1.

Table 1. Formats supported by Agilent ChemVista.

Format	Supported File Type	Access Links
Agilent PCDL	.cdb	
CSV	.csv	
MassBank EU	.txt	https://massbank.eu/MassBank/ https://github.com/MassBank/MassBank-data
MassBank of North America (MoNA)	.sdf	https://massbank.us/ https://mona.fiehnlab.ucdavis.edu/downloads
EPA CompTox Chemicals Dashboard	.sdf	https://comptox.epa.gov/dashboard/ https://comptox.epa.gov/dashboard/chemical-lists

Consolidation

On import to ChemVista, data is parsed according to source format specifications and stored in an accessible and flexible manner. Consolidation of data occurs first at the chemical structure and compound level, with a chemical structure identifier generation (e.g., SMILES, InChI, MOL) and merging protocol. The result of this background process is a compound-centric organizational structure, where identifiers are used to consolidate data corresponding to the same compound and structure under one chemical substance record within the system. If properly matched, a compound will only exist once within the system, but will include all stored information from imports. This allows for a consistent organizational structure and representation despite disparate organizational structures in third-party repositories (for example, MassBank and MoNA are spectral-centric repositories). Certain settings are configurable to support varied data use cases, such as the classification process entirely and whether stereochemistry errors should be ignored in incoming data.

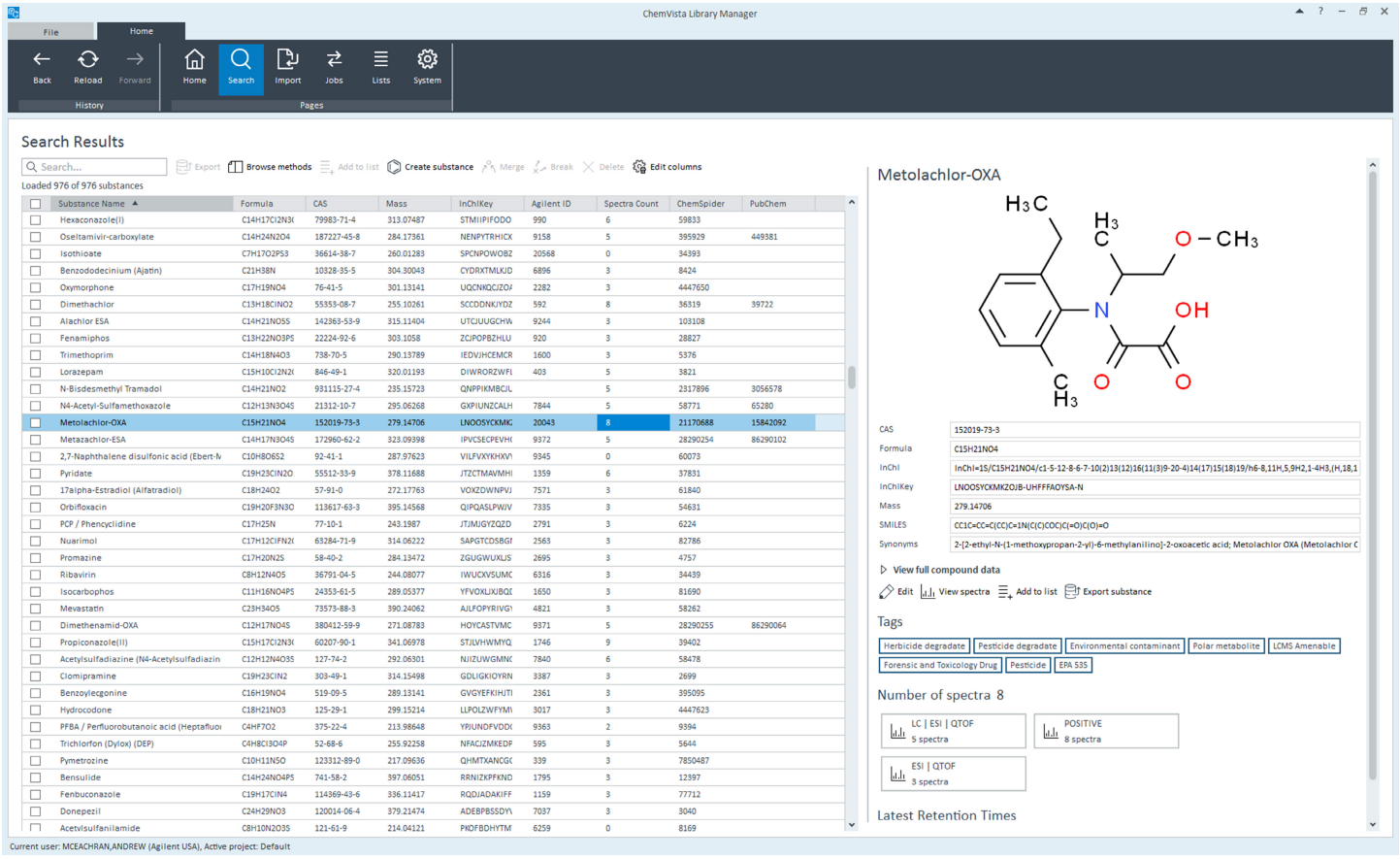

Figure 1. Spectral and compound data consolidated from multiple sources to compile a Pesticides for Targeted Screening list. Using the default classification and merging protocols results in streamlined data instead of many entries for the same compound. Data here contains Agilent IDs, spectra from MassBank and PCDLs, and synonyms from multiple sources.

Using ChemVista for library management

Organize, manage, and edit compounds

Establishing a flexible means to organize data is critical when managing library data from multiple sources without reliance on single files on disk. Lists are introduced in ChemVista to keep compounds organized. Lists can be (A) created on import, (B) created or appended from any compound view, (C) merged, and (D) removed. Lists are organizational tools and do not affect the underlying data. A compound can exist in multiple lists but is still just a single record, and a compound can be removed from a list but will not be deleted from the application.

Tailoring compound and chemical structure data may be necessary to achieve the desired set of targets for screening. Editing compound and chemical structure information is achievable when viewing compounds in a list view. Compound metadata is not overwritten by default; rather, when new data is added, the prior value is moved to the background of that data field. On commit of chemical structure information for a compound, new identifiers and a new structural image are generated automatically using the embedded EPAM Indigo toolkit[1] when possible. If needed, edits can be reverted to restore data to the state before the changes.

Enhanced cheminformatics features with the Indigo toolkit result in fully populated structure-identifier sets to support downstream workflows, but also enable support for structural skeleton searching using the first block of the InChIKey and quick identification of missing structures using the empty InChI designations (empty InChI: InChI=1S//, empty InChIKey: MOSFIJXAXDLOML-UHFFFAOYSA-N).

View, edit, and create spectra

In line with the compound-centric data model in ChemVista, spectral and method data are organized underneath chemical substance records, regardless of the data structure from the source format (i.e., when spectral-centric records are imported from MassBank, they are parsed and stored according to the ChemVista data model). Spectral and method data are processed in a manner that allows for consistent display and filter of pertinent information, including display of the separation type, ionization type, mass analyzer type, and collision energy. Further, SPLASH identifiers are generated for all spectra in the system and then deduplicated using this SPLASH, to prevent identical spectra from being stored under the same chemical substance.[2]

Browsing spectra is helpful for understanding data availability and fragmentation patterns, and further interrogating individual spectral data can elucidate method parameters for method development. Spectra can be created manually within ChemVista, and spectra can be edited via the Clone operation. Cloning spectral records copies all existing spectral details to a new record that can then be edited, preserving the original spectral record while allowing a user-manipulated spectral record to co-exist. Again, if needed, changes can be reverted to restore data to the state before the changes.

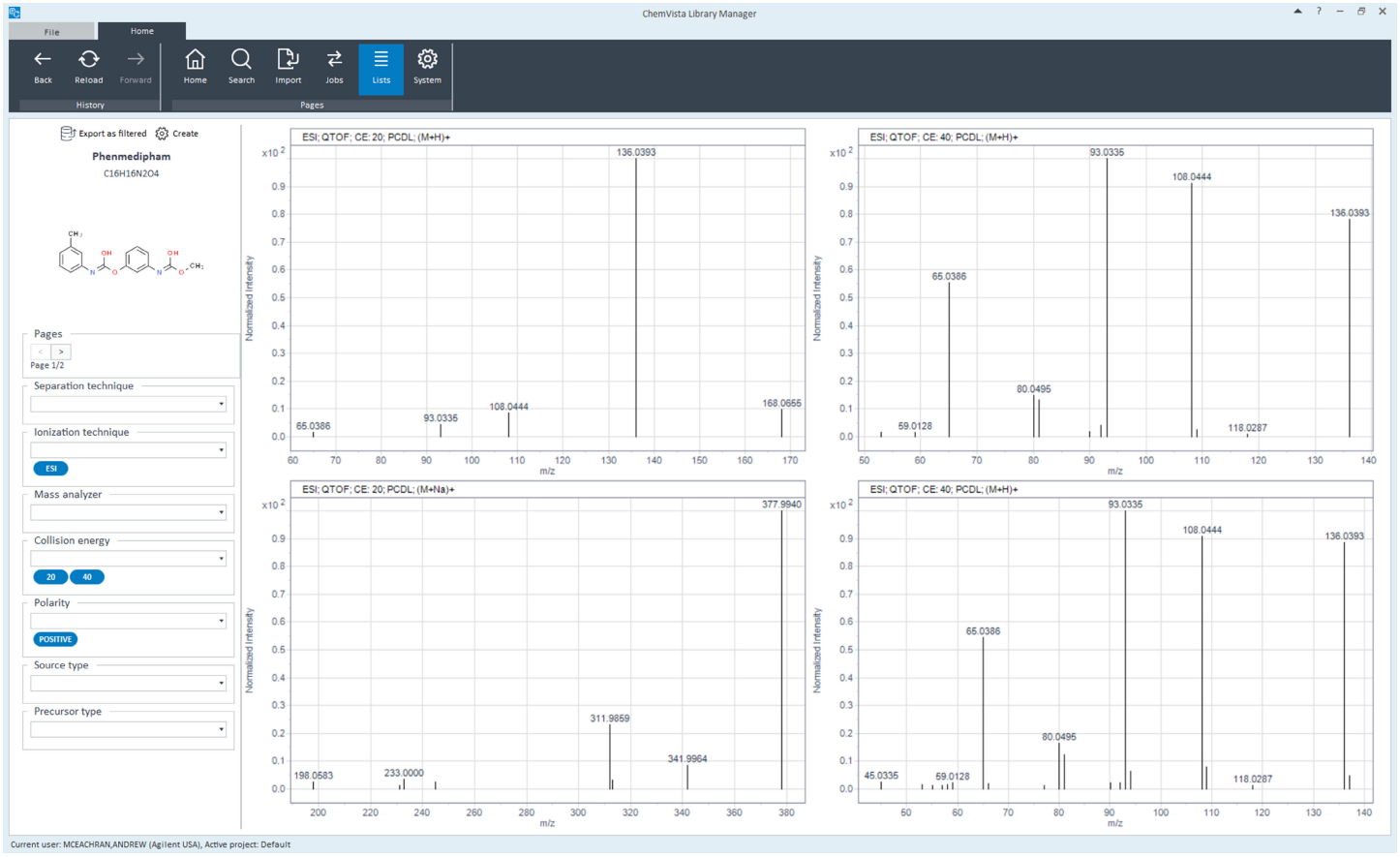

Figure 2. Spectral view with filter options on the left. Clicking a spectral plot opens extra details.

Organize, manage, and edit retention times and indices

Identification confidence is bolstered by inclusion of retention times (RTs) and indices (RIs). ChemVista supports organization of RTs and RIs by method information and customizable method labels, allowing many RTs and RIs to be stored per compound. Interactive browsing of RT/RI information by method and method label supports user interrogation and ultimate selection of RT/RI sets for export and inclusion in data analysis and identification workflows.

Editing RTs/RIs and creation of new method labels is achievable in the Browse Methods page. Further, a complete matrix of all method data (including RTs, RIs) for the compounds in view can be downloaded as a Microsoft Excel-compatible XML file. This organizes pertinent RT/RI and chemical structure data for more in-depth analysis outside ChemVista (for RT prediction, for example).

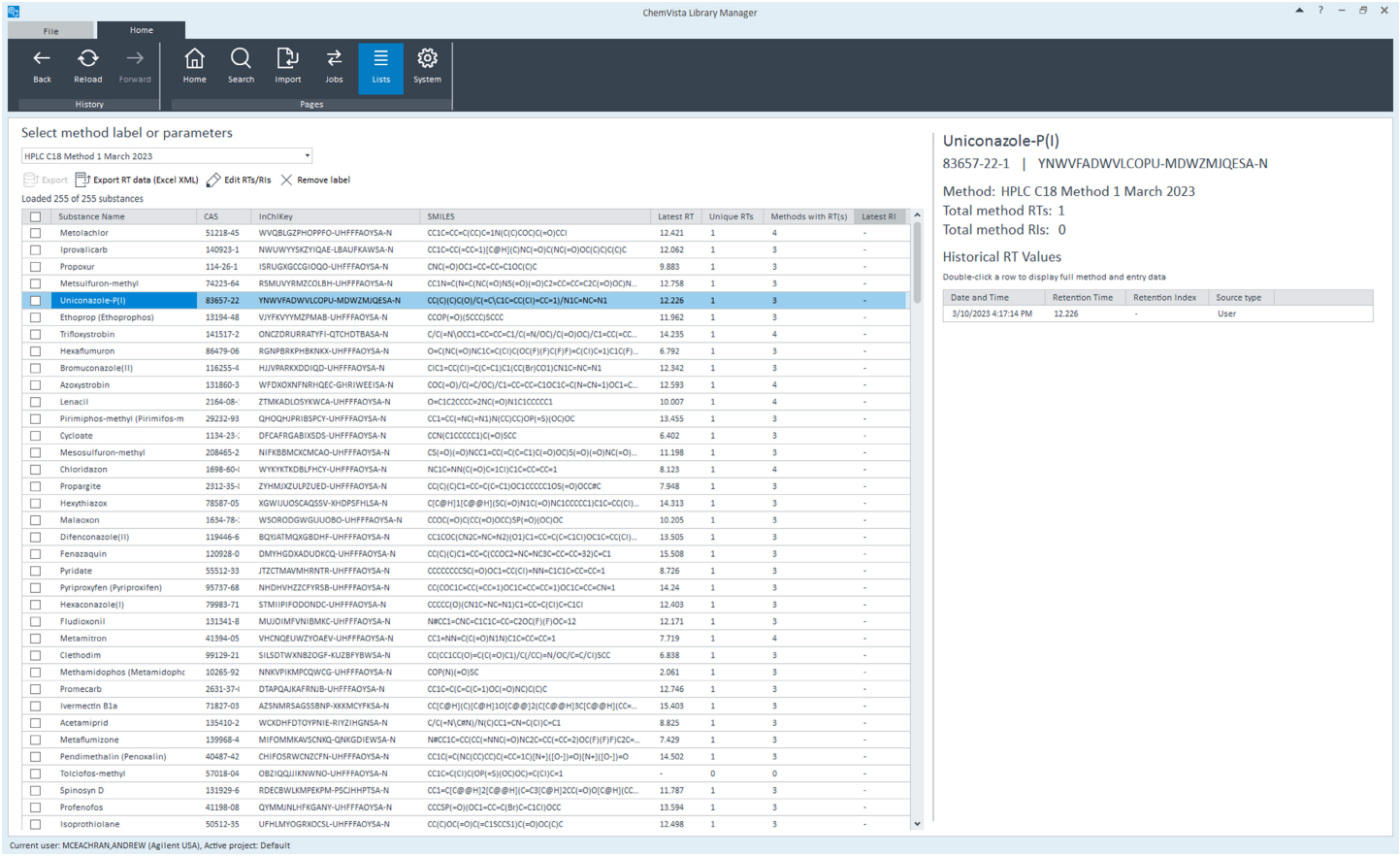

Figure 3. The Browse Methods page allows for RT and RI organization, and supports many RTs/RIs per compound. In this page, the user can toggle between existing methods to determine the best method to include RTs from an export for downstream analysis.

Search underlying data

Searching the underlying data in the application enhances list creation and target-screening subsetting. Data can be searched from the ChemVista home page in a single-entry manner using a name/synonym, CAS, or formula. Advanced and batch searches can be conducted via the Search page. Single-field searches can be conducted by entering search terms in the user-entry field. Identifiers (name, CAS, InChIKey, and SMILES) and formulas can be searched in batch mode by entering many search terms in new lines, and can be searched in part or as an exact match. Identifier searches can also be conducted against nonprimary data, meaning those identifiers that exist in the background set of data stored for a chemical substance (for example, an alternative name/synonym or alternate CAS). Other query options include searching by mass (range or mass ± error), tag, and presence in defined lists.

Combined searches built using the Search Builder tool are initiated from the basic controls when search terms are added as sets to the builder. If more than one set is added, only chemical substances that are common between sets will be returned. Searches built in the Search Builder tool can be saved and reloaded later to conduct the same query.

Complex searches against any data field in the system can be generated using regular expression queries via the Pattern Search. Here, the field of interest (i.e., AgilentID) is selected first and then the selected field is searched against using a regular expression (i.e., ^(?:[1-9][0-9]{3}|[1-9][0-9]{2}|[1-9][0-9]|[1-9])$ to conduct a search for all Agilent IDs between 1-9999). Pattern Searches can also be added to search sets to construct complex search combinations.

Multiuser configurations

When desirable, ChemVista can be set up for multiuser access in a client/server configuration. The backend server can be installed on a local-networked PC, and multiple clients can be installed on other PCs connected to the same network. In this fashion, multiple users can access the same backend data store from multiple PCs to utilize the same library data.

Exporting data to support workflows

Data can be exported from ChemVista into multiple file formats to support downstream workflows (see Table 2). Exports can be initiated from the List, Search Results, Browse Methods, and Spectral views. On initiation to the export page, filters and export options can be selected to fine-tune the make up of the export. For example, users can select **Only compounds with spectra**, or the spectra could be filtered to only include spectra with a positive polarity (Figure 4). To produce deposition-ready MassBank .txt files, export details can be edited to provide all required information.[3]

Table 2. ChemVista export formats.

Export Format	File Type	Downstream Support
PCDL	.cdb	Agilent MassHunter Qualitative Analysis software, MassHunter Quantitative Analysis software, MassHunter Mass Profiler software, etc.
MassBank	.txt	MassBank EU contribution
SDF	.sdf	Agilent MassHunter Molecular Structure Correlator software NIST Library format (via Lib2NIST conversion)

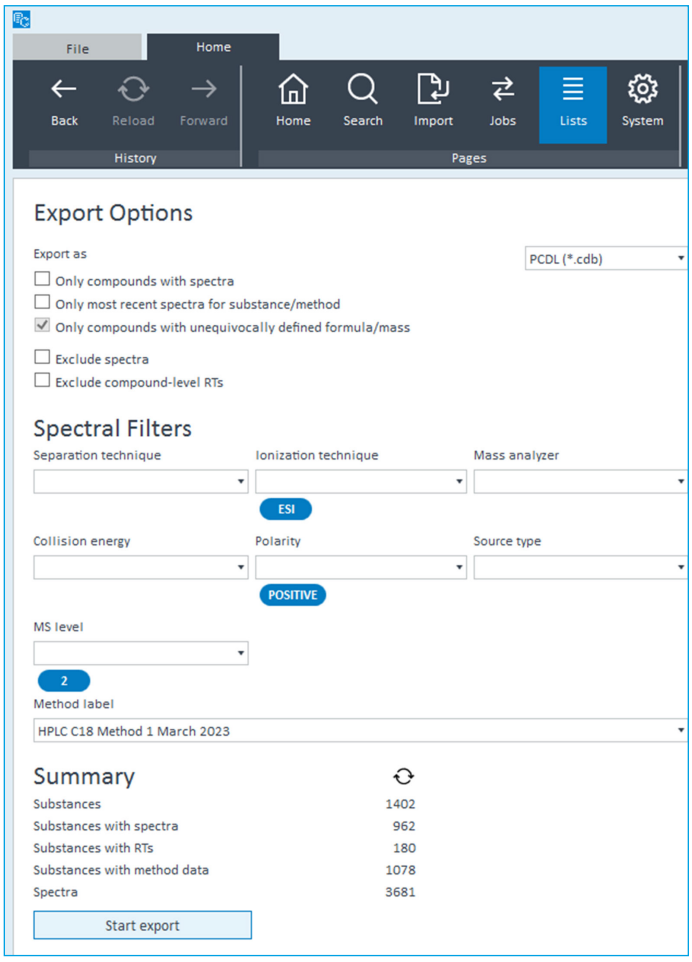

Figure 4. Export data to the desired format to fit your downstream workflow. Filter the spectra that are included by making selections from the Spectral Filters options.

Conclusion

Agilent ChemVista software enables users to merge compound, spectral, and retention time data from multiple sources into one location. In doing so, ChemVista software enables users to screen against spectra that have been previously inaccessible, enhancing the compound identification possibilities.

References

1. EPAM Life Sciences. https://lifescience.opensource.epam.com/indigo/
2. Wohlgemuth, G. *et al.* SPLASH, A Hashed Identifier for Mass Spectra. *Nat. Biotechnol.*, **2016**. *34*, 1099–1101. DOI: https://doi.org/10.1038/nbt.3689
3. MassBank Record Format. https://github.com/MassBank/MassBank-web/blob/main/Documentation/MassBankRecordFormat.md

www.agilent.com

DE24642158

Printed in the USA, March 21, 2023
5994-5924EN

Agilent
Trusted Answers