Poster Reprint

ASMS 2023
Poster number MP 373

Enhancing compound identification workflows with a novel library manager software application

Elizabeth Almasi, Emma E. Rennie, James S. Pyke, Tristan Chutka, Andrew D. McEachran

Agilent Technologies, Santa Clara, CA

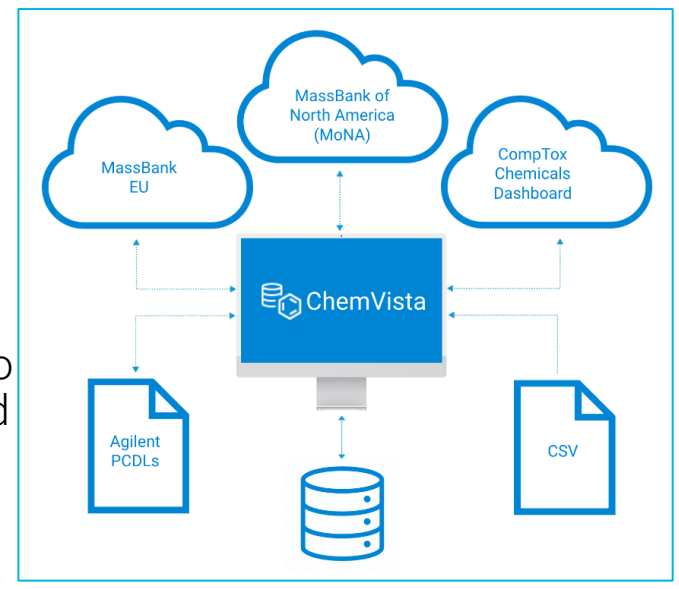
Introduction

As targeted screening and non-targeted identification workflows advance, empirical high-resolution MS/MS spectra remain the benchmark for confident compound identifications/annotations. MS/MS spectra exist in many formats, both openly available (e.g., MassBank) and commercial (various MS vendors) in nature. Incorporating spectra from multiple sources into a single screening analysis is crucial to expand the identifiable chemical space of screening workflows. However, integrating these disparate formats into a single consolidated structure is cumbersome and error-prone, if not impossible due to commercial format restrictions. To expand the identifiable chemical space of screening workflows, we present a novel standalone software application that was built to centrally manage library data where data can be interconverted between formats without reliance on a single format.

Experimental

Software architecture

Agilent ChemVista software is implemented as a server-client application incorporating a lightweight desktop client and standalone server. Data processing tasks are offloaded to a single high-performance server, while multiple users can concurrently use clients to manage and curate imported data. Multiple source modules can be installed to the server to facilitate the import and export of data in a variety of formats and depending on project needs.



An extensible data model was established allowing for full translation of both standard and non-standard data fields. Once imported, physical data management and optimization is handled by an advanced relational database system.

Results and Discussion

Chemical structure generation and standardization

Chemical structure information is generated and standardized during the import process using the EPAM Indigo Toolkit[1]. Structure information can be generated using a SMILES string, or an InChI String, or MOL Text and then is used to populate all other identifiers for a given chemical substance (Figure 1).

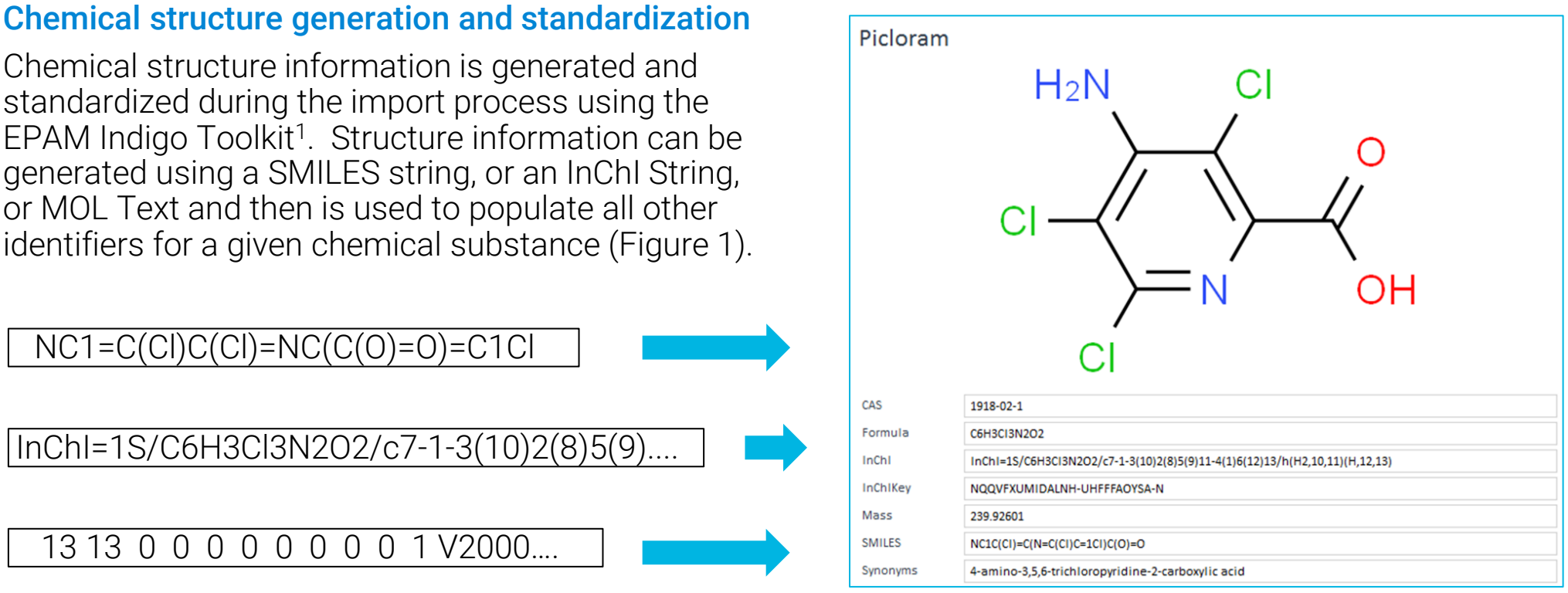
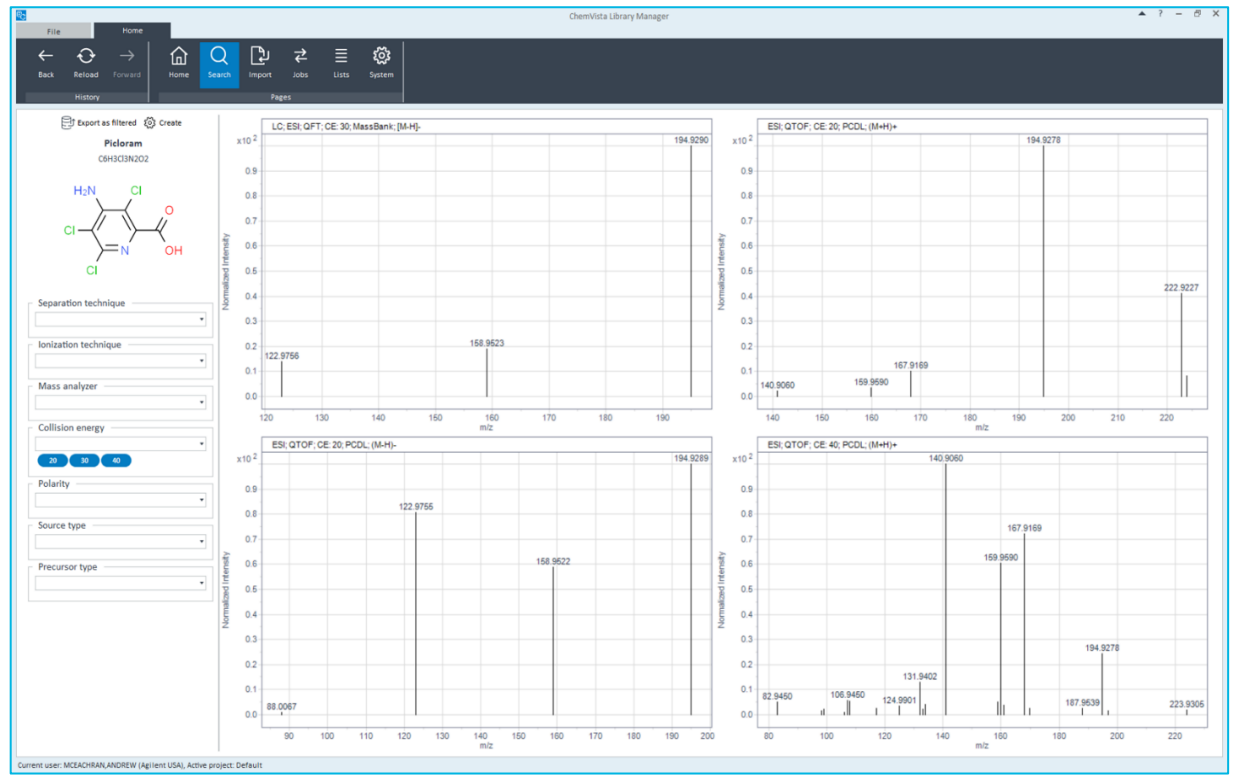


Figure 1. Chemical structure details (MOL, InChI String, and SMILES) provided during data import or data addition generate chemical structure information within the software application.

Compound-centric organizational structure

Default processes identify when data being imported match an already existing compound and associate the incoming data with the same compound and chemical structure. Spectra and other compound metadata from many sources are then stored and viewable underneath the same compound (Figure 2).

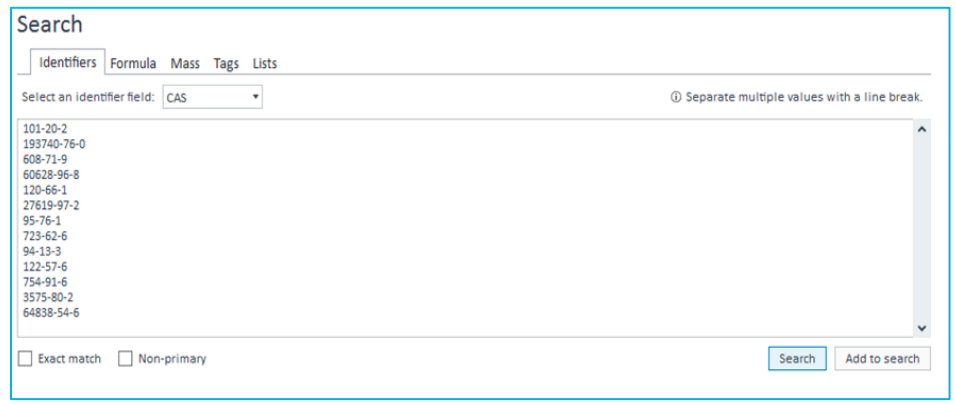
Figure 2. Spectral view showing spectra from both MassBank and an Agilent PCDL underneath the same compound record Picloram.

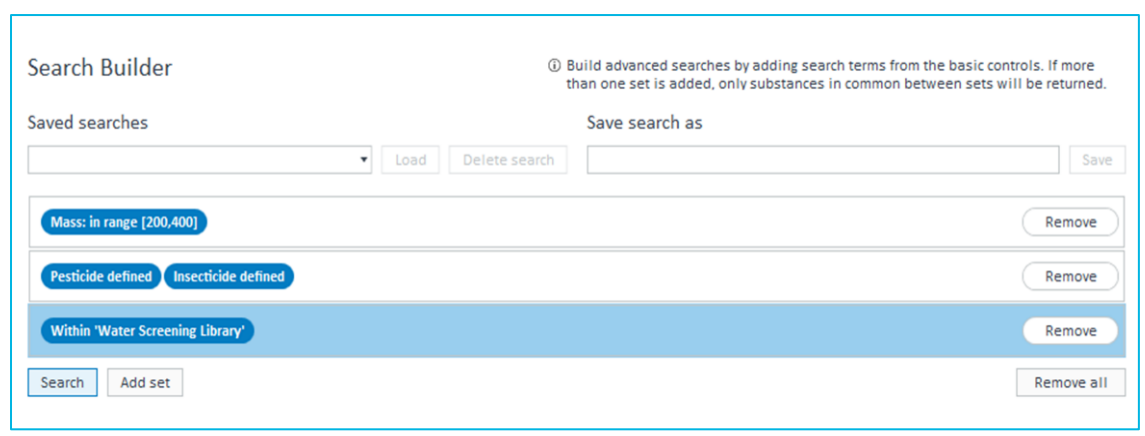


Search and subset data

Tailoring screening lists for project-specific needs can be accomplished using single search terms (Figure 3A) or by constructing an advanced search using multiple search sets with the Search Builder function (Figure 3B). Searching can be conducted using identifers, formulas, masses, etc. Search results can be saved as a List in ChemVista to further interrogate or use downstream.

Figure 3A (top). Single search terms can be searched by batch or one at a time. Figure 3B (bottom). Advanced searches combining multiple search sets can be constructed in the Search Builder.

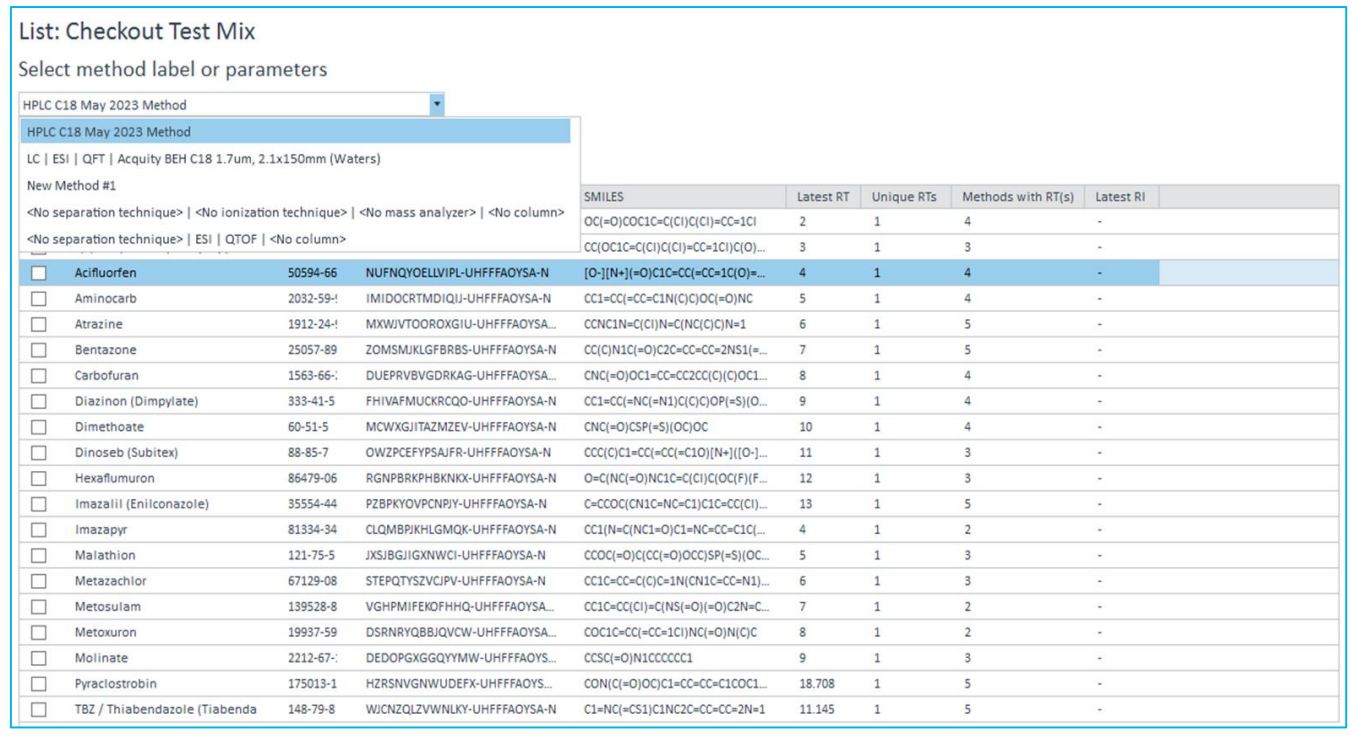




Method/RT organization

Retention times (RTs) are organized according to method parameters or a user-defined method label. Many method labels can exist for the same compound or sets of compounds, allowing for many sets of RTs to be stored and viewed in the Browse Methods view (Figure 4). These method labels can be used to select which set of RTs to use in downstream data analysis.

Figure 4. Browse Methods page view, showing multiple method labels available for selection for the given set of compounds.



Results and Discussion

Export for data analysis and identification

Compound and spectral data can be exported to multiple formats to facilitate a variety of identification workflows. When exporting to the PCDL format, compound and spectral data can be used directly in Agilent MassHunter identification workflows, like MassHunter Qualitative Analysis (Figure 5)..

As proof-of-concept, ~5500 spectral files from MassBank[2] were imported on top of an existing Agilent Water Screening PCDL. All new, incoming spectra were added underneath existing compound records as expected based on the merging and classification protocol. After exporting the same starting PCDL list with the new spectra, the available spectra for identification in MassHunter Qualitative Analysis increased by 50%.

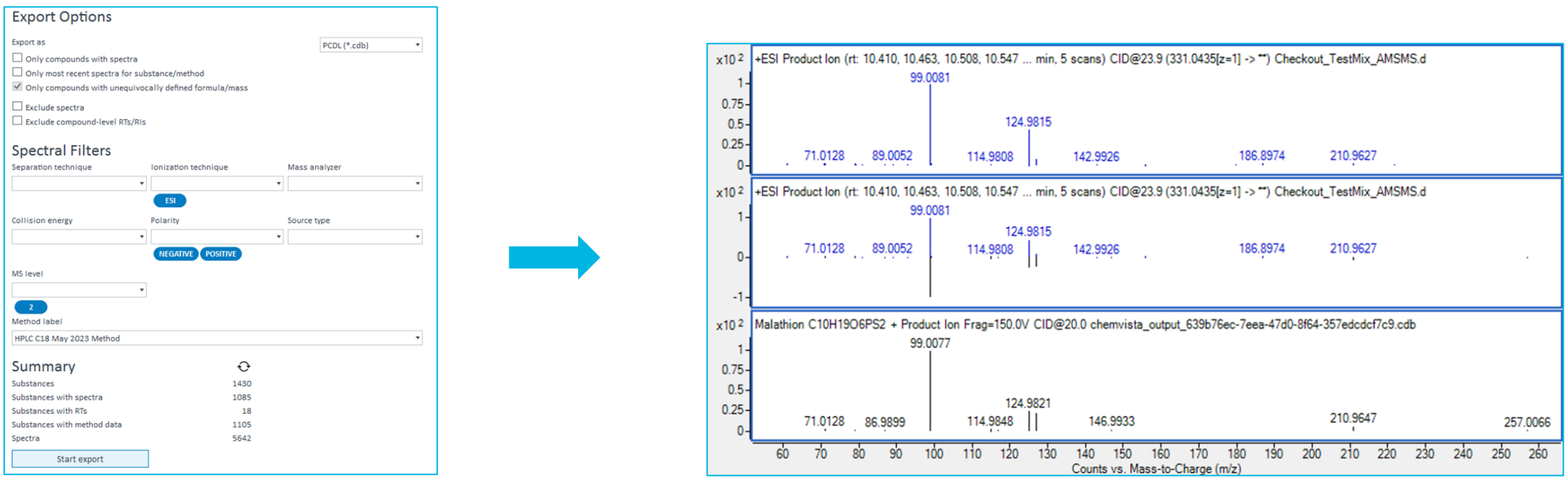


Figure 5. Export options page view (left), showing available filters for spectra included in an export. When exported to the Agilent PCDL format, spectra can be readily included in a library search in MassHunter Qualitative Analysis (right).

List creation for managing data

Lists are the organizational focal point of ChemVista and can be created from multiple points throughout the software (Figure 6). Flexible list creation and management enables customization for screening lists that can be tailored based on the needs of a project or when new data is added.

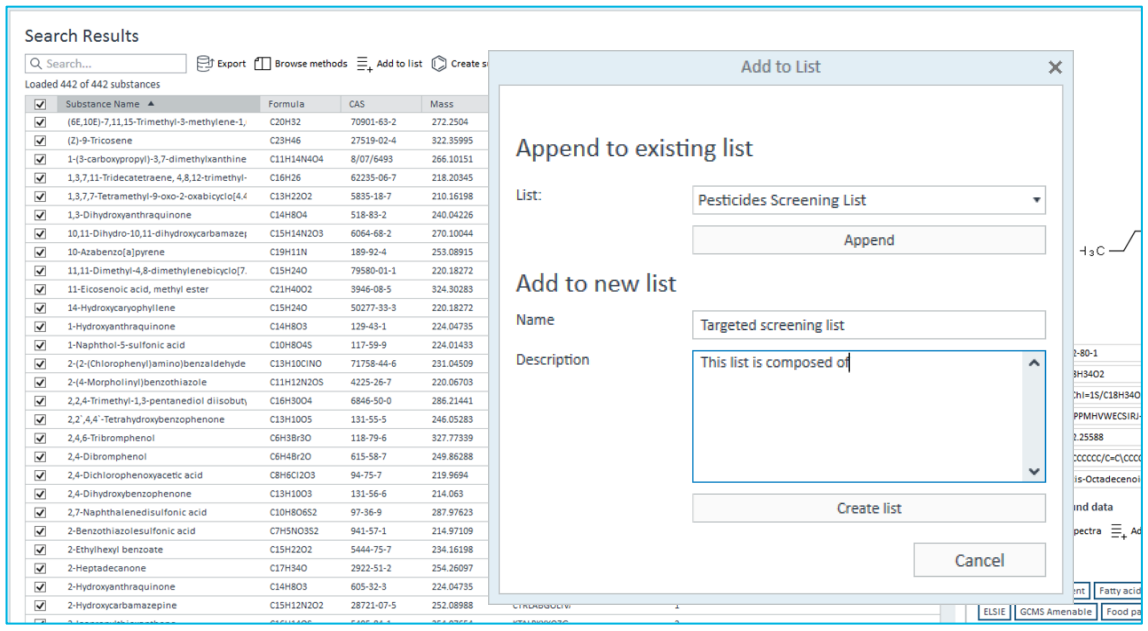


Figure 6. Creating a new list or adding to an existing list from the search results page view in ChemVista. Lists can also be created during the import process.

Conclusions

- Agilent ChemVista is a novel library manager software application
- ChemVista supports compound and spectral data from the Agilent PCDL format and multiple open-source formats
- Features within the software are designed to tailor screening lists for project-specific needs
- Exporting data to compatible file formats can increase available spectral coverage for identification workflows

References

[1]EPAM Indigo Toolkit. https://lifescience.opensource.epam.com/indigo/

[2]MassBank High Quality Mass Spectral Database. https://massbank.eu/MassBank/

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

DE06381796

Published in USA, May 31,2023

Agilent

Trusted Answers