

Poster Reprint

ASMS 2023 Poster number MP 373

Enhancing compound identification workflows with a novel library manager software application

<u>Elizabeth Almasi,</u> 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 highresolution 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 errorprone, 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 serverclient 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 Toolkit¹. 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).

NC1=C(CI)C(CI)=NC(C(O)=O)=C1CI







InChIKey	NQQVFXUMIDALNH-UHFFFAOYSA-N
Mass	239.92601
SMILES	NC1C(CI)=C(N=C(CI)C=1CI)C(O)=O
Synonyms	4-amino-3,5,6-trichloropyridine-2-carboxylic acid

2

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.

Results and Discussion

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.

[2) ్లు LC; ESI; QFT; CE: 30; Ma 0.2 120 0.9 0.2 106.9450 124.9901 Search Identifiers Formula Mass Tags Lists Select an identifier field: CAS (i) Separate multiple values with a line break

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.

101-20-2		^
193740-76-0		
508-71-9		
50528-95-8		
120-66-1		
2/013-3/-2		
722.62.6		
94-13-3		
122-57-6		
754-91-6		
3575-80-2		
64838-54-6		
		¥
Exact match Non-primary		Search Add to search
		than one set is added, only substances in common between sets will be re-
Saved searches		Save search as
-	Load Delete search	
Mass: in range [200,400]		Re
Mass: in range [200,400]		Re
Mass: in range [200,400] Pesticide defined Insecticide defined		Re
Mass: in range [200,400] Pesticide defined Insecticide defined Within 'Water Screening Library'		Re
Mass: in range [200,400] Pesticide defined Insecticide defined Within 'Water Screening Library'		Re

3

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.

List: Checkout Test Mix Select method label or parameters					
HPLC C18 May 2023 Method					
HPLC C18 May 2023 Method					
LC ESI QFT Acquity BEH C18 1.7um, 2.1x150mm (Waters)					
New Method #1	010050	Internet DT	Unious DT-	March a da unitabilitzation	Lobort DI
<no separation="" technique=""> <no ionization="" technique=""> <no analyzer="" mass=""> <no column=""></no></no></no></no>	SMILES	Latest RI	Unique Ris	Methods with RI(S)	Latest KI

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

No separation technique> ESI QTOF <no column=""></no>			CC(OC1C=C(CI)C(CI)=CC=1CI)C(O)	3	1	3			
	Acifluorfen	50594-66	NUFNQYOELLVIPL-UHFFFAOYSA-N	[O-][N+](=O)C1C=CC(=CC=1C(O)=	4	1	4	-	
	Aminocarb	2032-59-!	IMIDOCRTMDIQIJ-UHFFFAOYSA-N	CC1=CC(=CC=C1N(C)C)OC(=O)NC	5	1	4	(=)	
	Atrazine	1912-24-!	MXWJVTOOROXGIU-UHFFFAOYSA	CCNC1N=C(CI)N=C(NC(C)C)N=1	6	1	5		
	Bentazone	25057-89	ZOMSMJKLGFBRBS-UHFFFAOYSA-N	CC(C)N1C(=O)C2C=CC=2NS1(=	7	1	5		
	Carbofuran	1563-66-:	DUEPRVBVGDRKAG-UHFFFAOYSA	CNC(=O)OC1=CC=CC2CC(C)(C)OC1	8	1	4	-	
	Diazinon (Dimpylate)	333-41-5	FHIVAFMUCKRCQO-UHFFFAOYSA-N	CC1=CC(=NC(=N1)C(C)C)OP(=S)(O	9	1	4	1.00	
	Dimethoate	60-51-5	MCWXGJITAZMZEV-UHFFFAOYSA-N	CNC(=O)CSP(=S)(OC)OC	10	1	4		
	Dinoseb (Subitex)	88-85-7	OWZPCEFYPSAJFR-UHFFFAOYSA-N	CCC(C)C1=CC(=CC(=C1O)[N+]([O-]	11	1	3	(*)	
	Hexaflumuron	86479-06	RGNPBRKPHBKNKX-UHFFFAOYSA-N	O=C(NC(=O)NC1C=C(CI)C(OC(F)(F	12	1	3		
	Imazalil (Enilconazole)	35554-44	PZBPKYOVPCNPJY-UHFFFAOYSA-N	C=CCOC(CN1C=NC=C1)C1C=CC(CI)	13	1	5		
	Imazapyr	81334-34	CLQMBPJKHLGMQK-UHFFFAOYSA-N	CC1(N=C(NC1=O)C1=NC=CC=C1C(4	1	2		
	Malathion	121-75-5	JXSJBGJIGXNWCI-UHFFFAOYSA-N	CCOC(=O)C(CC(=O)OCC)SP(=S)(OC	5	1	3		
	Metazachlor	67129-08	STEPQTYSZVCJPV-UHFFFAOYSA-N	CC1C=CC=C(C)C=1N(CN1C=CC=N1)	6	1	3	-	
	Metosulam	139528-8	VGHPMIFEKOFHHQ-UHFFFAOYSA	CC1C=CC(CI)=C(NS(=O)(=O)C2N=C	7	1	2		
	Metoxuron	19937-59	DSRNRYQBBJQVCW-UHFFFAOYSA	COC1C=CC(=CC=1CI)NC(=O)N(C)C	8	1	2		
	Molinate	2212-67-:	DEDOPGXGGQYYMW-UHFFFAOYS	CCSC(=O)N1CCCCCC1	9	1	3		
	Pyraclostrobin	175013-1	HZRSNVGNWUDEFX-UHFFFAOYS	CON(C(=O)OC)C1=CC=CC=C1COC1	18.708	1	5		
	TBZ / Thiabendazole (Tiabenda	148-79-8	WJCNZQLZVWNLKY-UHFFFAOYSA-N	C1=NC(=CS1)C1NC2C=CC=CC=2N=1	11.145	1	5		

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 MassBank² 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.

Q Se	arch Export (Browse methods	\equiv_+ Add to list	Create s		Add to List	×	2	
baded	442 of 442 substances			-					
✓	Substance Name 🔺	Formula	CAS	Mass					
✓	(6E,10E)-7,11,15-Trimethyl-3-methylene-1,	C20H32	70901-63-2	272.2504					
~	(Z)-9-Tricosene	C23H46	27519-02-4	322.35995	Appond to ovicti	ng list			
~	1-(3-carboxypropyl)-3,7-dimethylxanthine	C11H14N4O4	8/07/6493	266.10151	Append to existi	lig list			
√	1,3,7,11-Tridecatetraene, 4,8,12-trimethyl-	C16H26	62235-06-7	218.20345					
√	1,3,7,7-Tetramethyl-9-oxo-2-oxabicyclo[4.4	C13H22O2	5835-18-7	210.16198	List:	List: Pesticides Screening List			
~	1,3-Dihydroxyanthraquinone	C14H8O4	518-83-2	240.04226					
√	10,11-Dihydro-10,11-dihydroxycarbamaze;	C15H14N2O3	6064-68-2	270.10044		Annand			
~	10-Azabenzo[a]pyrene	C19H11N	189-92-4	253.08915		Append		H3C	
√	11,11-Dimethyl-4,8-dimethylenebicyclo[7.	C15H24O	79580-01-1	220.18272	A 1 1				
1	11-Eicosenoic acid, methyl ester	C21H40O2	3946-08-5	324.30283	Add to new list				
1	14-Hydroxycaryophyllene	C15H24O	50277-33-3	220.18272					
~	1-Hydroxyanthraquinone	C14H8O3	129-43-1	224.04735	Name	Targeted screening list			
~	1-Naphthol-5-sulfonic acid	C10H8O4S	117-59-9	224.01433				1.00.1	
v	2-(2-(Chlorophenyl)amino)benzaldehyde	C13H10CINO	71758-44-6	231.04509	Description	This list is composed of	^	1-00-1	
~	2-(4-Morpholinyl)benzothiazole	C11H12N2OS	4225-26-7	220.06703				SH3402	
√	2,2,4-Trimethyl-1,3-pentanediol diisobuty	C16H30O4	6846-50-0	285.21441				hl=15/	
√	2,2',4,4'-Tetrahydroxybenzophenone	C13H10O5	131-55-5	246.05283				PPMHV	
✓	2,4,6-Tribromphenol	C6H3Br3O	118-79-6	327.77339				2.25588	
√	2,4-Dibromphenol	C6H4Br2O	615-58-7	249.86288				ccccc,	
√	2,4-Dichlorophenoxyacetic acid	C8H6CI2O3	94-75-7	219.9694			~	is-Octa	
✓	2,4-Dihydroxybenzophenone	C13H10O3	131-56-6	214.063					
✓	2,7-Naphthalenedisulfonic acid	C10H8O6S2	97-36-9	287.97623		Create list		ind da	
~	2-Benzothiazolesulfonic acid	C7H5NO3S2	941-57-1	214.97109				pectra	
~	2-Ethylhexyl benzoate	C15H22O2	5444-75-7	234.16198			Canaal		
~	2-Heptadecanone	C17H34O	2922-51-2	254.26097			cancer		
	2-Hydroxyanthraquinone	C14H8O3	605-32-3	224 04735					

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

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.

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

This information is subject to change without notice.

DE06381796

© Agilent Technologies, Inc. 2023 Published in USA, May 31,2023

References

¹EPAM Indigo Toolkit. <u>https://lifescience.opensource.epam.com/indigo/</u>

²MassBank High Quality Mass Spectral Database. <u>https://massbank.eu/MassBank/</u>

