White Paper

Improving Confidence in Compound Identification Using Agilent Curated Databases and Libraries

Authors

Emma E. Rennie,
Robert H. Williams,
Ruben Garnica, and
Maria VanDamme

Introduction

The level of confidence in target and unknown compound identification is determined by the authenticity and integrity of the data contained within the databases and spectral libraries used for compound identification. Much of the database and library information that is publicly available on the internet, and from many vendors, is often crowd-sourced with little or no review. Poor quality libraries can negatively impact the quality of a laboratory's results.

A museum curator or art collector undertakes extensive research to evaluate the authenticity and integrity of each and every exhibit in their collection of art or artifacts. In a similar manner, Agilent uses an extremely detailed curation process to ensure the quality of each data entry and accurate mass spectrum in its compound databases and libraries.

This White Paper explains how Agilent accurate mass libraries and databases are developed, the nature of the curation process, and why this curation process will improve the quality of the results produced by a laboratory.

What are Agilent Accurate Mass Personal Compound Databases and Libraries?

An Agilent accurate mass Personal Compound Database and Library (PCDL) is a database containing detailed compound information along with a library of accurate, high-resolution mass spectra acquired using a Quadrupole Time-of-Flight mass spectrometer (Q-TOF MS). PCDLs are available for both gas chromatography and liquid chromatography applications. Using a PCDL, Agilent MassHunter Data Analysis Software, and a TOF or Q-TOF MS, compounds can be identified with up to four degrees of confirmation: accurate mass, isotopic distribution, accurate mass fragment spectrum, and retention time (RT). When using Agilent Ion Mobility Quadrupole TOF (IM-Q-TOF) instruments, an additional dimension, collision cross section (CCS), is available for further confirmation. The end user can target specific compounds of interest or screen for several thousand targets without the need for individual reference standards. Using the Agilent MassHunter PCDL Manager Software package provided with each PCDL, PCDLs can be customized by the end user to suit their particular analysis or system configuration by the addition of RTs, new compounds, new spectra and, for the IM-Q-TOF MS, ion mobility information.

Full isotopic MS information, in combination with accurate mass spectra, can provide the end user with highly confident compound identification. Adding RTs specific to the chromatographic method being used provides an additional level of confidence. RTs are particularly useful when analyzing isomeric compounds. Retrospective data analysis of historic samples can easily be accomplished by adding compounds to the target PCDL. New curated LC/MS/MS spectra can easily be added through the same import settings Agilent chemists use when engaging the automatic curation workflow integrated into Agilent MassHunter Qualitative Analysis Workflows Software.

Agilent accurate mass databases and libraries can be used across the Agilent MassHunter Software portfolio. MassHunter PCDL Manager interfaces directly with the Agilent MassHunter Software Suite including the data analysis software packages, Qualitative Analysis and Quantitative Analysis, as well as the statistical analysis and visualization software packages: Mass Profiler, Mass Profiler Professional, and Profinder.

All Agilent MassHunter PCDL products include a user guide with data examples and familiarization exercises as well as upgrade protection to ensure end users have up-to-date information after the initial purchase. Some PCDLs such as the *MassHunter Pesticides PCDL and Workflow for GC/Q-TOF* also include electronic methods, recommended configurations, and a workflow guide to allow the PCDL and compound identification workflows to be rapidly deployed in an analytical laboratory.

Development of Agilent Accurate Mass Personal Compound Databases and Libraries

Each PCDL is created following a detailed standard procedure consisting of four main steps:

1. Identify the target compound list in collaboration with leading experts.
2. Generate the compound database, and verify the information for each target entry.
3. Collect fragment spectral data using high purity reference standards.
4. Create the library with fragment spectra curated per the rigorous Agilent quality control curation process.

Step 1: Create Compound Lists in Collaboration With Market Field Experts

The first step in the database and library creation process is the identification of compounds to be included in or added to a PCDL. Agilent establishes partnerships with leading experts in various disciplines to prioritize these target lists, ensuring the resulting products are comprehensive, relevant, and current to a specific area of study.

Once the contents of each PCDL product are carefully selected (see Figure 1 for an example contents list), the compounds are prioritized for fragment spectral collection.

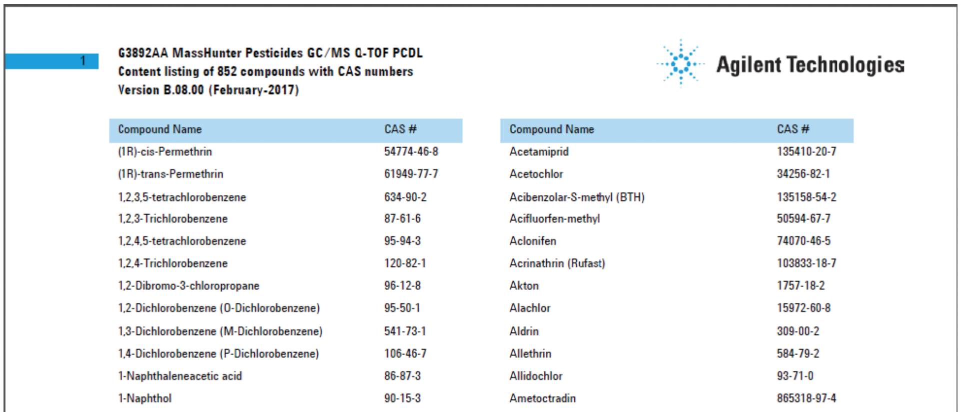
Step 2: Create the Database Entries for Each Compound

A database entry is created for each target compound included in a PCDL. Each entry includes a variety of information about the compound, and each piece of information is manually verified. Molecular formulas, names, structures, CAS registry numbers, and compound class tags are cross-checked across multiple reputable sources (for example, SciFinder, ChemSpider, PubChem, ChemIDplus, or Chemistry Dashboard) by an experienced Agilent chemist. In addition, each individual chemical identifier is manually verified at its source site. This activity ensures a high degree of accuracy for the compound information included in each entry.

The specific types of chemical identifiers present in the database depends on the application area of the PCDL, and include such entries as:

- Common name
- IUPAC name
- Synonyms
- Molecular formula
- CAS registry number
- ChemSpider ID
- PubChem ID
- InChI
- SMILES
- Vendor product codes

Metabolite-based identifiers such as KEGG, METLIN ID, HMDB ID, LMP ID, ChEBI ID, and BioCyc ID are included for metabolites.



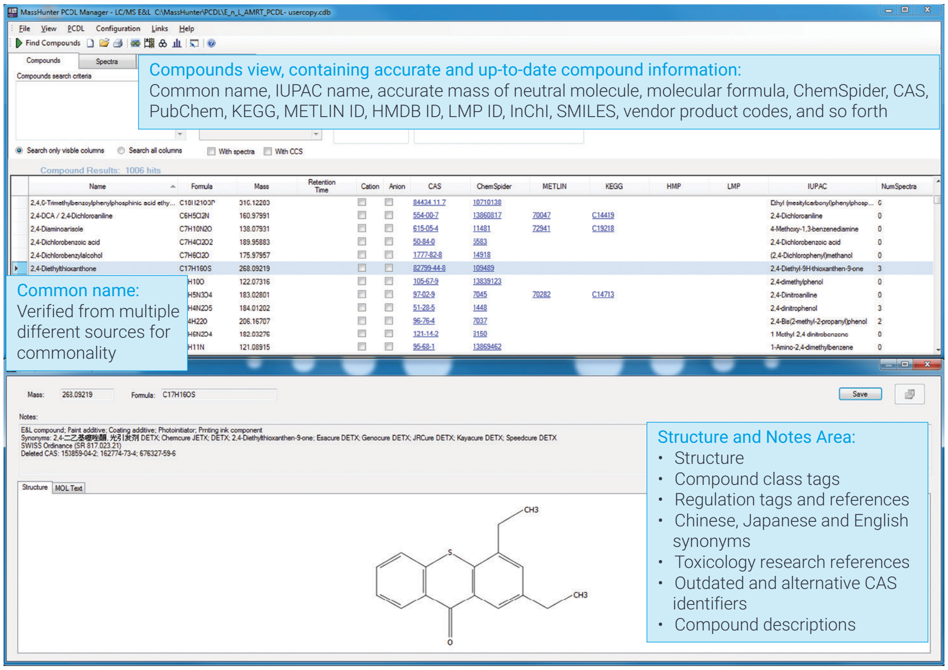
1

G3892AA MassHunter Pesticides GC/MS Q-TOF PCDL
Content listing of 852 compounds with CAS numbers
Version B.08.00 (February-2017)

Agilent Technologies

Compound Name	CAS #	Compound Name	CAS #
(1R)-cis-Permethrin	54774-46-8	Acetamiprid	135410-20-7
(1R)-trans-Permethrin	61949-77-7	Acetochlor	34256-82-1
1,2,3,5-tetrachlorobenzene	634-90-2	Acibenzolar-S-methyl (BTH)	135158-54-2
1,2,3-Trichlorobenzene	87-61-6	Acifluorfen-methyl	50594-67-7
1,2,4,5-tetrachlorobenzene	95-94-3	Aclonifen	74070-46-5
1,2,4-Trichlorobenzene	120-82-1	Acrinathrin (Rufast)	103833-18-7
1,2-Dibromo-3-chloropropane	96-12-8	Akton	1757-18-2
1,2-Dichlorobenzene (O-Dichlorobenzene)	95-50-1	Alachlor	15972-60-8
1,3-Dichlorobenzene (M-Dichlorobenzene)	541-73-1	Aldrin	309-00-2
1,4-Dichlorobenzene (P-Dichlorobenzene)	106-46-7	Allethrin	584-79-2
1-Naphthaleneacetic acid	86-87-3	Allidochlor	93-71-0
1-Naphthol	90-15-3	Ametoctradin	865318-97-4

Figure 1. Each PCDL includes a content listing of all the included compounds with their Chemical Abstracts Service (CAS) registry number.



MassHunter PCDL Manager - LC/MS E&L C:\MassHunter\PCDL\E_n_L_AMRT_PCDL- usercopy.cdb
File
View
PCDL
Configuration
Links
Help
Find Compounds
Compounds
Spectra
Compounds search criteria
Search only visible columns
Search all columns
With spectra
With CCS
Compound Results: 1006 hits
Save
Structure
MOL Text
S
O
CH3
CH3

Compounds view, containing accurate and up-to-date compound information:
Common name, IUPAC name, accurate mass of neutral molecule, molecular formula, ChemSpider, CAS, PubChem, KEGG, METLIN ID, HMDB ID, LMP ID, InChI, SMILES, vendor product codes, and so forth

Name	Formula	Mass	Retention Time	Cation	Anion	CAS	ChemSpider	METLIN	KEGG	HMP	LMP	IUPAC	NumSpectra
2,4,6-Trimethylbenzoylphenylphosphinic acid ethy...	C18H21O3P	316.12283		□	□	84434-11-7	10710138					Ethyl (mesitylcarbonyl)phenylphosp...	6
2,4-DCA / 2,4-Dichloroaniline	C6H5Cl2N	160.97991		□	□	554-00-7	13860817	70047	C14419			2,4-Dichloroaniline	0
2,4-Diaminoanisole	C7H10N2O	138.07931		□	□	615-05-4	11481	72941	C19218			4-Methoxy-1,3-benzenediamine	0
2,4-Dichlorobenzoic acid	C7H4Cl2O2	189.95883		□	□	50-84-0	5583					2,4-Dichlorobenzoic acid	0
2,4-Dichlorobenzylalcohol	C7H6Cl2O	175.97957		□	□	1777-82-8	14918					(2,4-Dichlorophenyl)methanol	0
2,4-Diethylthioxanthone	C17H16OS	268.09219		□	□	82799-44-8	109489					2,4-Diethyl-9H-thioxanthen-9-one	3
	H10O	122.07316		□	□	105-67-9	13839123					2,4-dimethylphenol	0
	H5N3O4	183.02801		□	□	97-02-9	7045	70282	C14713			2,4-Dinitroaniline	0
	H4N2O5	184.01202		□	□	51-28-5	1448					2,4-dinitrophenol	3
	4H22O	206.16707		□	□	96-76-4	7037					2,4-Bis(2-methyl-2-propanyl)phenol	2
	H6N2O4	182.03276		□	□	121-14-2	3150					1-Methyl-2,4-dinitrobenzene	0
	H11N	121.08915		□	□	95-68-1	13869462					1-Amino-2,4-dimethylbenzene	0

Common name:
Verified from multiple different sources for commonality

Mass: 263.09219 Formula: C17H16OS

Notes:
E&L compound; Paint additive; Coating additive; Photoinitiator; Prnting ink component
Synonyms: 2,4-二乙基噻吨酮, 光引发剂 DETX; Chemcure JETX; DETX; 2,4-Diethylthioxanthen-9-one; Esacure DETX; Genocure DETX; JRCure DETX; Kayacure DETX; Speedcure DETX
SWISS Ordinance (SR 817.023.21)
Deleted CAS: 153859-04-2; 162774-73-4; 676327-59-6

Structure and Notes Area:
- Structure
- Compound class tags
- Regulation tags and references
- Chinese, Japanese and English synonyms
- Toxicology research references
- Outdated and alternative CAS identifiers
- Compound descriptions

Figure 2. Agilent MassHunter PCDL Manager showing typical compound database entries.

Each database entry also includes a chemical structure and notes for each compound. The notes are compound-specific, and can include:

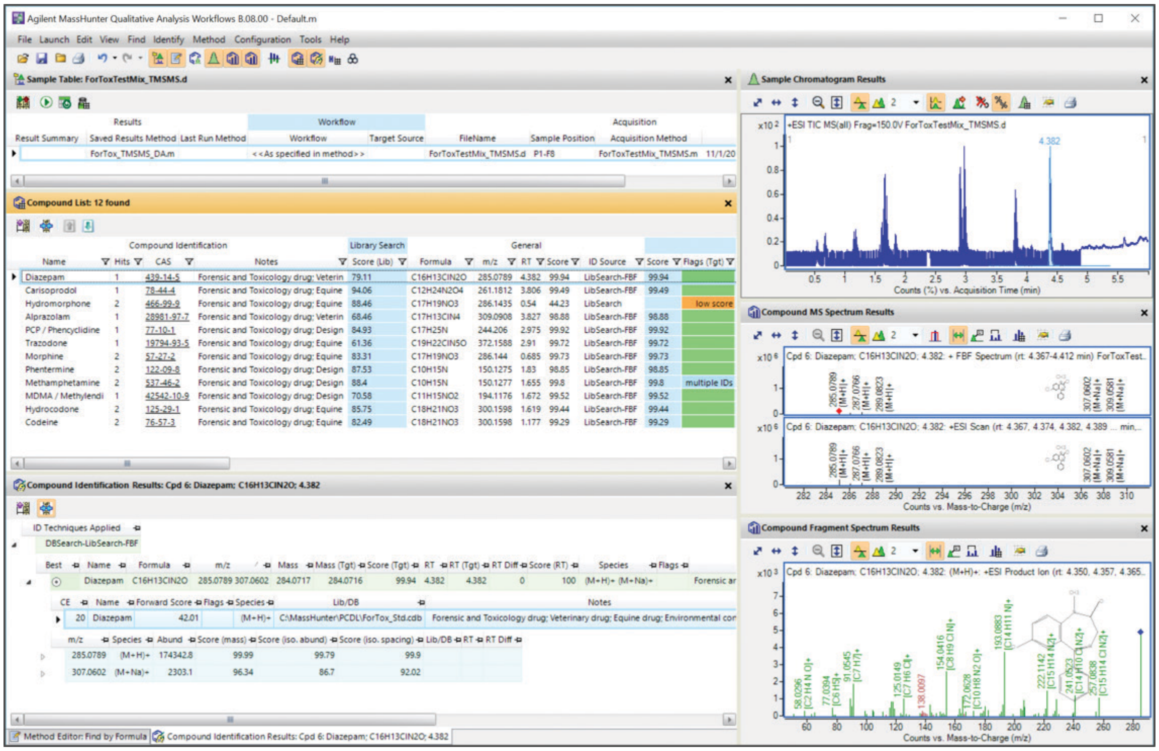
- Compound, regulation, and country-specific class tags
- Chinese, Japanese, and English synonyms
- Toxicology research references
- Deleted (or outdated) and alternative CAS registry numbers
- Compound descriptions

All the above information for each compound entry in the database can be searched in PCDL Manager, using simple or advanced search functionalities (see Figure 2).

These entries can be used to filter the database to create smaller editable subsets of PCDLs from each PCDL product. For example, an end user could establish their own custom PCDL to create a screening workflow for a particular compound class or suspect list.

Finally, each compound entry is checked against the Agilent Master PCDL to eliminate the possibility of duplicate compound entries, which could drastically increase the time needed for results review in data analysis.

Compound identification information, such as common name, CAS, structure, and notes, is available through MassHunter Data Analysis software, as well as MassHunter PCDL Manager (see Figure 3).


Agilent MassHunter Qualitative Analysis Workflows B.08.00 - Default.m
File Launch Edit View Find Identify Method Configuration Tools Help
Sample Chromatogram Results
+ESI TIC MS(all) Frag=150.0V ForToxTestMix_TMSMS.d
4.382
x10 2
0 0.2 0.4 0.6 0.8 1
0.5 1 1.5 2 2.5 3 3.5 4 4.5 5 5.5
Counts (%) vs. Acquisition Time (min)
Compound MS Spectrum Results
Cpd 6: Diazepam; C16H13ClN2O; 4.382: + FBF Spectrum (rt: 4.367-4.412 min) ForToxTest...
Cpd 6: Diazepam; C16H13ClN2O; 4.382: +ESI Scan (rt: 4.367, 4.374, 4.382, 4.389 ... min,...
x10 6
285.0789 [M+H]+
287.0766 [M+H]+
289.0823 [M+H]+
307.0602 [M+Na]+
309.0581 [M+Na]+
282 284 286 288 290 292 294 296 298 300 302 304 306 308 310
Counts vs. Mass-to-Charge (m/z)
Compound Fragment Spectrum Results
Cpd 6: Diazepam; C16H13ClN2O; 4.382: (M+H)+: +ESI Product Ion (rt: 4.350, 4.357, 4.365...
x10 3
58.0296 [C2 H4 N O]+
77.0394 [C6 H5]+
91.0545 [C7 H7]+
125.0149 [C7 H6 Cl]+
138.0097
154.0416 [C8 H9 Cl N]+
172.0628 [C10 H8 N2 O]+
193.0883 [C14 H11 N]+
222.1142 [C15 H14 N2]+
241.0523 [C14 H10 Cl N2]+
257.0838 [C15 H14 Cl N2]+
60 80 100 120 140 160 180 200 220 240 260 280
Counts vs. Mass-to-Charge (m/z)
Method Editor: Find by Formula
Compound Identification Results: Cpd 6: Diazepam; C16H13ClN2O; 4.382

Sample Table: ForToxTestMix_TMSMS.d

Results			Workflow				Acquisition	
Result Summary	Saved Results Method	Last Run Method	Workflow	Target Source	FileName	Sample Position	Acquisition Method	
	ForTox_TMSMS_DA.m		<<As specified in method>>		ForToxTestMix_TMSMS.d	P1-F8	ForToxTestMix_TMSMS.m	11/1/20

Compound List: 12 found

Compound Identification				Library Search	General						
Name	Hits	CAS	Notes	Score (Lib)	Formula	m/z	RT	Score	ID Source	Score	Flags (Tgt)
Diazepam	1	439-14-5	Forensic and Toxicology drug; Veterin	79.11	C16H13ClN2O	285.0789	4.382	99.94	LibSearch-FBF	99.94	
Carisoprodol	1	78-44-4	Forensic and Toxicology drug; Equine	94.06	C12H24N2O4	261.1812	3.806	99.49	LibSearch-FBF	99.49	
Hydromorphone	2	466-99-9	Forensic and Toxicology drug; Equine	88.46	C17H19NO3	286.1435	0.54	44.23	LibSearch		low score
Alprazolam	1	28981-97-7	Forensic and Toxicology drug; Veterin	68.46	C17H13ClN4	309.0908	3.827	98.88	LibSearch-FBF	98.88	
PCP / Phencyclidine	1	77-10-1	Forensic and Toxicology drug; Design	84.93	C17H25N	244.206	2.975	99.92	LibSearch-FBF	99.92	
Trazodone	1	19794-93-5	Forensic and Toxicology drug; Equine	61.36	C19H22ClN5O	372.1588	2.91	99.72	LibSearch-FBF	99.72	
Morphine	2	57-27-2	Forensic and Toxicology drug; Equine	83.31	C17H19NO3	286.144	0.685	99.73	LibSearch-FBF	99.73	
Phentermine	2	122-09-8	Forensic and Toxicology drug; Design	87.53	C10H15N	150.1275	1.83	98.85	LibSearch-FBF	98.85	
Methamphetamine	2	537-46-2	Forensic and Toxicology drug; Design	88.4	C10H15N	150.1277	1.655	99.8	LibSearch-FBF	99.8	multiple IDs
MDMA / Methylendi	1	42542-10-9	Forensic and Toxicology drug; Design	70.58	C11H15NO2	194.1176	1.672	99.52	LibSearch-FBF	99.52	
Hydrocodone	2	125-29-1	Forensic and Toxicology drug; Equine	85.75	C18H21NO3	300.1598	1.619	99.44	LibSearch-FBF	99.44	
Codeine	2	76-57-3	Forensic and Toxicology drug; Equine	82.49	C18H21NO3	300.1598	1.177	99.29	LibSearch-FBF	99.29	

Compound Identification Results: Cpd 6: Diazepam; C16H13ClN2O; 4.382

ID Techniques Applied

DBSearch-LibSearch-FBF

Best	Name	Formula	m/z	Mass	Mass (Tgt)	Score (Tgt)	RT	RT (Tgt)	RT Diff	Score (RT)	Species	Flags	
⊙	Diazepam	C16H13ClN2O	285.0789 307.0602	284.0717	284.0716	99.94	4.382	4.382	0	100	(M+H)+ (M+Na)+		Forensic ar

CE	Name	Forward Score	Flags	Species	Lib/DB	Notes
20	Diazepam	42.01		(M+H)+	C:\MassHunter\PCDL\ForTox_Std.cdb	Forensic and Toxicology drug; Veterinary drug; Equine drug; Environmental con

m/z	Species	Abund	Score (mass)	Score (iso. abund)	Score (iso. spacing)	Lib/DB	RT	RT Diff
285.0789	(M+H)+	174342.8	99.99	99.79	99.9			
307.0602	(M+Na)+	2303.1	96.34	86.7	92.02			

Figure 3. Agilent MassHunter Qualitative Analysis Workflows, showing compound identifications for an LC/MS/MS analysis including structures, CAS, and notes.

Step 3: Collect Mass Spectral Data Using High Purity Reference Standards

Accurate mass spectra are acquired using high purity single compound reference standards. All the data are acquired using a strict Standard Operating Procedure (SOP), which includes instrument tuning and calibration as well as acquisition of quality control samples, ensuring that all data are acquired under the same experimental conditions. Compounds not available as single standards, such as fungal or bacterial metabolites, are synthesized by our leading-edge expert scientific collaborators.

GC/MS accurate mass spectra are collected in EI full scan mode, producing a rich fragmentation pattern and isotopic cluster information (see Figure 4).

LC/MS/MS accurate mass spectra are collected for the lowest *m/z* isotopic molecular ion in both positive and negative polarities, for multiple adduct ion species and collision energies (see Figure 5). To provide rich fragmentation spectra for compounds of varying stability, each compound is acquired at 10, 20, and 40 V collision energies. If the 10 V spectra are still overfragmented (very labile compounds), or the 40 V spectra are still under-fragmented (for very stabile compounds), then spectra at additional collision energies are acquired, as appropriate.

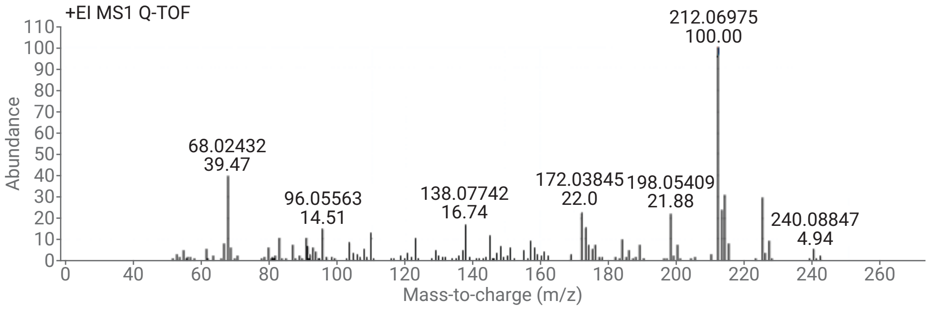


Figure 4. GC/MS EI accurate mass spectrum for the pesticide cyanazine.

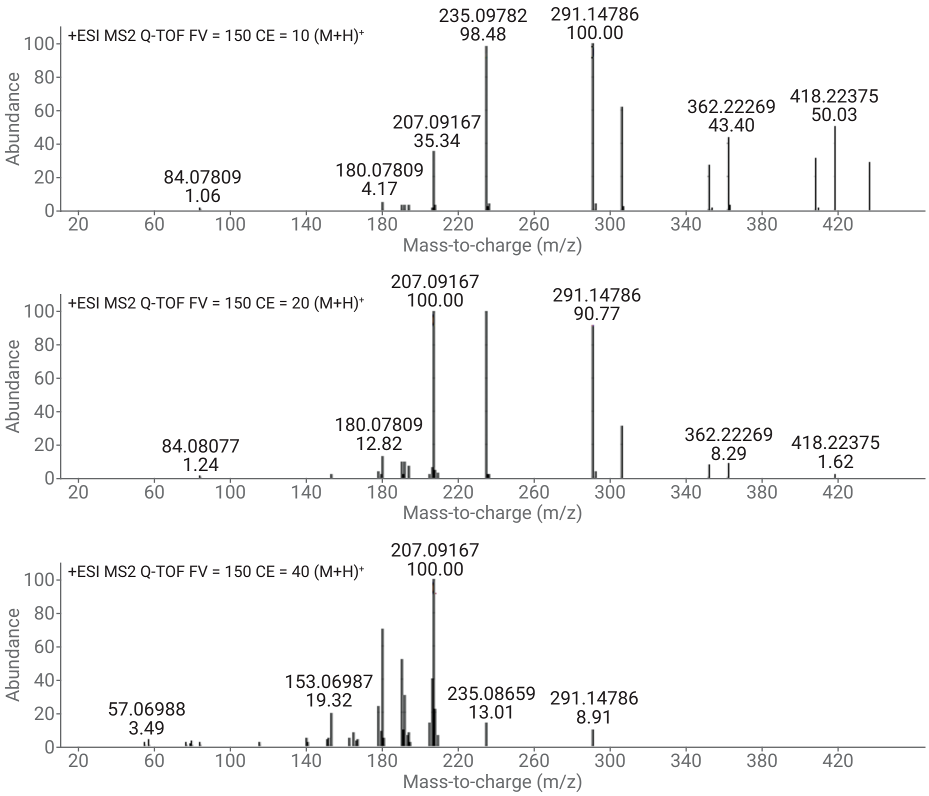


Figure 5. LC/MS/MS ESI accurate mass spectra for the drug valsartan.

Step 4: Create the Library Spectra Entries for the Prioritized Compound List

Mass spectral information is acquired by Agilent applications chemists or one of the Agilent expert collaborators using the appropriate SOP. The choice of SOP depends on the chromatographic method being used.

LC/MS/MS Spectral Curation

The accurate LC/MS/MS mass spectra are extracted and analyzed using algorithms present in MassHunter Qualitative Data Analysis Software. First, the spectra are extracted using the *Find by Formula* algorithm. *Molecular Formula Generation* (MFG) assigns a formula to each *m/z* fragment ion peak in the mass spectrum, then each peak is validated to determine if it is a valid subformula of the parent ion formula. For most fragment ion peaks, a number of possible formula assignments exist within the MFG mass tolerance. In these cases, the ion peak is assigned to the fragment formula with the closest mass. The MFG mass tolerance used in this process is varied, depending on the mass, and is approximately three times lower than the library search default mass tolerance. When the spectra are imported into the PCDL, all ion *m/z* peaks are corrected to their theoretical accurate mass (see Figures 6A and 6B). Fragment ion peaks that cannot be assigned to a subformula of the parent ion are rejected as chemical noise or as an impurity. Fragment ion peaks at or below 1 % of the base peak are always rejected as noise.

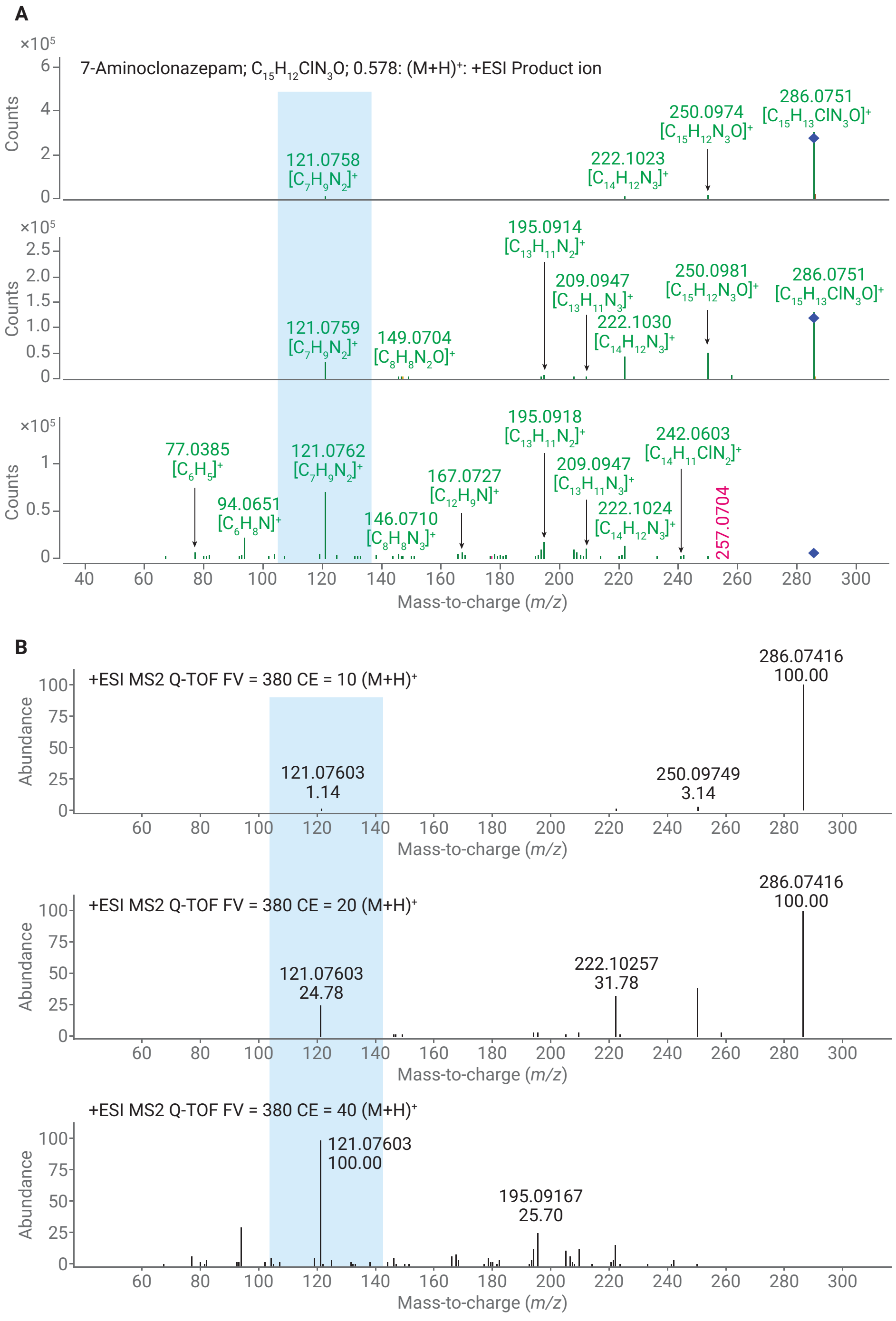


Figure 6. A) LC/MS/MS ESI experimental accurate mass spectra for 7-aminoclonazepam and B) LC/MS/MS ESI spectra for 7-aminoclonazepam corrected to theoretical accurate mass. The blue highlighted areas show examples of the mass correction.

GC/MS Spectral Curation

The accurate mass spectra are extracted and analyzed using algorithms present in MassHunter Qualitative Data Analysis Software. First, the spectra are manually extracted based on the extracted ion chromatogram (EIC) of the characteristic ion. MFG is then used to assign a formula to each *m/z* fragment ion peak in the mass spectrum based on the known formula of the compound, taking into account isotope ratios and spacing. For most fragment ion peaks, a number of possible formula assignments exist within the MFG mass tolerance. In these cases, the ion peak is assigned to the fragment formula with the closest mass with the correct the isotopic ratio and spacing. The MFG mass tolerance used in this process is varied, depending on the mass, and is approximately three times lower than the library search default mass tolerance. When the spectra are imported into the PCDL, all ion *m/z* peaks are corrected to their theoretical accurate mass. Fragment ion peaks that cannot be assigned to a subformula of the parent ion are rejected as chemical noise or as an impurity. Fragment ion peaks at or below 100 ion counts are always rejected as noise.

To ensure high quality searchable spectra, data sets are inspected carefully by Agilent scientists, and are automatically rejected for the following reasons:

- Incorrectly set instrumentation parameters
- The base peak is <1,000 ion counts
- Fragment ion peaks that have been rejected from the spectrum as noise, or as an impurity, lower the library search score
- Indication of an impurity, degradant, or transformation product, such as multiple compounds or components being found in the LC/MS/MS chromatogram for a single compound reference standard

As the final step in the curation process for all PCDL spectra, each spectrum imported into the PCDL is visually qualified in MassHunter Qualitative Data Analysis by a specialized and highly skilled mass spectrometrist.

Each entry undergoes multiple levels of review to ensure data quality. The curation process ensures that all data are collected and curated equally, eliminating errors from inconsistently acquired data. Inconsistently acquired data are one of the major drawbacks of open-access or user-contributed libraries, and can result in poor quality search results and low confidence in the results obtained using these databases and libraries.

For many of the forensic toxicology drug compounds, the LC/MS/MS spectrum compound information contains content about the single chemical standard used for each spectrum, providing traceability.

How to Use the Agilent Accurate Mass Personal Compound Databases and Libraries

Agilent PCDLs interface directly with the Agilent MassHunter Software Suite, and provide molecular formulas and reference library spectra along with compound information, ensuring unique compound identifiers for a variety of compound data mining and identification workflows.

Creating smaller targeted PCDLs specific to an analytical process such as metabolomic pathway, compound class, regulation, or country-specific list can significantly improve data analysis efficiency and accuracy by removing compounds that are not of interest from the PCDL and the compound data mining and identification pipeline.

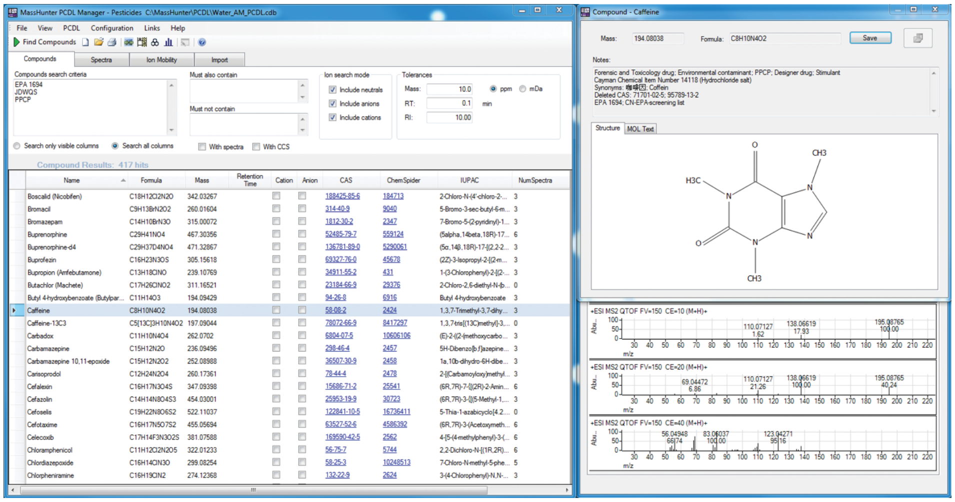
Smaller targeted PCDLs can easily be created by:

1. Searching relevant compound class, regulation, or country-specific tags in PCDL Manager (see Figure 7 for an example)
2. Highlighting the desired compound results
3. Selecting **Create Subset PCDL** or **Append to PCDL** with a right mouse button click

Metabolomics workflows can use the MassHunter Pathways to PCDL software to create PCDLs for pathway directed data analysis (see Figure 8).

When new analytes emerge, the MassHunter PCDL Manager import functionality easily allows:

- Import of new compounds with chemical identifiers
- Addition or update of RTs for compounds or a subset of compounds
- Import of CCS values as well as other detailed Ion Mobility information (IM-Q-TOF LC/MS)


MassHunter PCDL Manager - Pesticides C:\MassHunter\PCDL\Water_AM_PCDL.cdb
File View PCDL Configuration Links Help
Find Compounds
Compounds Spectra Ion Mobility Import

Compounds search criteria: EPA 1694, JDWQS, PPCP

Must also contain:

Must not contain:

Ion search mode: ☑ Include neutrals ☑ Include anions ☑ Include cations

Tolerances: Mass: 10.0 ● ppm ○ mDa; RT: 0.1 min; RI: 10.00

○ Search only visible columns ● Search all columns □ With spectra □ With CCS

Compound Results: 417 hits

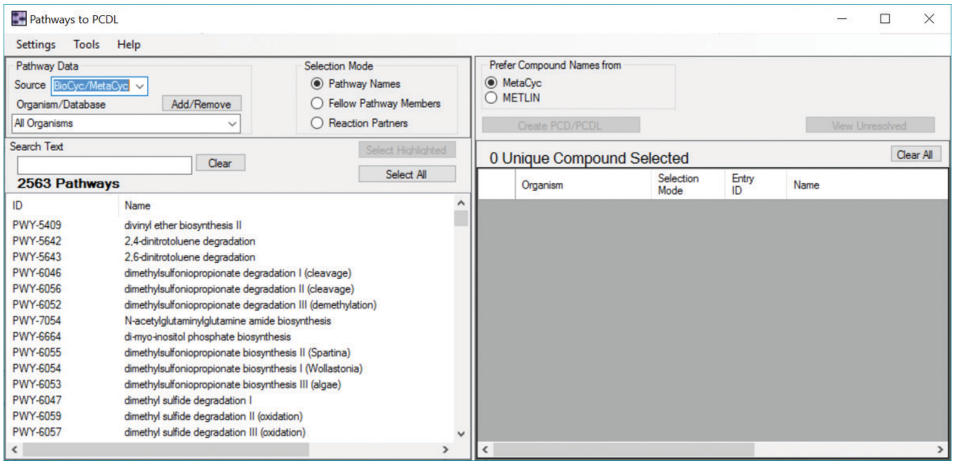
Name	Formula	Mass	Retention Time	Cation	Anion	CAS	ChemSpider	IUPAC	NumSpectra
Boscalid (Nicobifen)	C18H12Cl2N2O	342.03267		□	□	188425-85-6	184713	2-Chloro-N-(4'-chloro-2-...	3
Bromacil	C9H13BrN2O2	260.01604		□	□	314-40-9	9040	5-Bromo-3-sec-butyl-6-m...	3
Bromazepam	C14H10BrN3O	315.00072		□	□	1812-30-2	2347	7-Bromo-5-(2-pyridinyl)-1...	3
Buprenorphine	C29H41NO4	467.30356		□	□	52485-79-7	559124	(5alpha,14beta,18R)-17...	6
Buprenorphine-d4	C29H37D4NO4	471.32867		□	□	136781-89-0	5290061	(5α,14β,18R)-17-[(2,2-2...	3
Buprofezin	C16H23N3OS	305.15618		□	□	69327-76-0	45678	(2Z)-3-Isopropyl-2-[(2-m...	3
Bupropion (Amfebutamone)	C13H18ClNO	239.10769		□	□	34911-55-2	431	1-(3-Chlorophenyl)-2-[(2-...	3
Butachlor (Machete)	C17H26ClNO2	311.16521		□	□	23184-66-9	29376	2-Chloro-2,6-diethyl-N-(b...	0
Butyl 4-hydroxybenzoate (Butylpar...	C11H14O3	194.09429		□	□	94-26-8	6916	Butyl 4-hydroxybenzoate	3
Caffeine	C8H10N4O2	194.08038		□	□	58-08-2	2424	1,3,7-Trimethyl-3,7-dihy...	3
Caffeine-13C3	C5[13C]3H10N4O2	197.09044		□	□	78072-66-9	8417297	1,3,7-tris[(13C)methyl]-3,...	0
Carbadox	C11H10N4O4	262.0702		□	□	6804-07-5	10606106	(E)-2-((2-(methoxycarbo...	3
Carbamazepine	C15H12N2O	236.09496		□	□	298-46-4	2457	5H-Dibenzo[b,f]azepine...	3
Carbamazepine 10,11-epoxide	C15H12N2O2	252.08988		□	□	36507-30-9	2458	1a,10b-dihydro-6H-dibe...	3
Carisoprodol	C12H24N2O4	260.17361		□	□	78-44-4	2478	2-[(Carbamoyloxy)methyl...	3
Cefalexin	C16H17N3O4S	347.09398		□	□	15686-71-2	25541	(6R,7R)-7-{[(2R)-2-Amin...	6
Cefazolin	C14H14N8O4S3	454.03001		□	□	25953-19-9	30723	(6R,7R)-3-{[(5-Methyl-1,...	3
Cefoselis	C19H22N8O6S2	522.11037		□	□	122841-10-5	16736411	5-Thia-1-azabicyclo[4.2....	0
Cefotaxime	C16H17N5O7S2	455.05694		□	□	63527-52-6	4586392	(6R,7R)-3-(Acetoxymeth...	6
Celecoxib	C17H14F3N3O2S	381.07588		□	□	169590-42-5	2562	4-[5-(4-methylphenyl)-3-(...	6
Chloramphenicol	C11H12Cl2N2O5	322.01233		□	□	56-75-7	5744	2,2-Dichloro-N-[(1R,2R)...	6
Chlordiazepoxide	C16H14ClN3O	299.08254		□	□	58-25-3	10248513	7-Chloro-N-methyl-5-phe...	5
Chlorpheniramine	C16H19ClN2	274.12368		□	□	132-22-9	2624	3-(4-Chlorophenyl)-N,N-...	3

Compound - Caffeine

Mass: 194.08038 Formula: C8H10N4O2 Save

Notes:
Forensic and Toxicology drug; Environmental contaminant; PPCP; Designer drug; Stimulant
Cayman Chemical Item Number 14118 (Hydrochloride salt)
Synonyms: 咖啡因; Coffein
Deleted CAS: 71701-02-5; 95789-13-2
EPA 1694; CN-EPA-screening list

Figure 7. Agilent MassHunter PCDL Manager B.08.00 showing compound search results for US and Japanese governmental regulation tags (EPA 1694 and JDWQS) as well as the compound class tag for Pharmaceuticals and Personal Care Products (PPCP).



Pathway Data: Source: BioCyc/MetaCyc; Organism/Database: All Organisms

Selection Mode: ● Pathway Names ○ Fellow Pathway Members ○ Reaction Partners

Prefer Compound Names from: ● MetaCyc ○ METLIN

Search Text:

2563 Pathways

ID	Name
PWY-5409	divinyl ether biosynthesis II
PWY-5642	2,4-dinitrotoluene degradation
PWY-5643	2,6-dinitrotoluene degradation
PWY-6046	dimethylsulfoniopropionate degradation I (cleavage)
PWY-6056	dimethylsulfoniopropionate degradation II (cleavage)
PWY-6052	dimethylsulfoniopropionate degradation III (demethylation)
PWY-7054	N-acetylglutaminylglutamine amide biosynthesis
PWY-6664	di-myo-inositol phosphate biosynthesis
PWY-6055	dimethylsulfoniopropionate biosynthesis II (Spartina)
PWY-6054	dimethylsulfoniopropionate biosynthesis I (Wollastonia)
PWY-6053	dimethylsulfoniopropionate biosynthesis III (algae)
PWY-6047	dimethyl sulfide degradation I
PWY-6059	dimethyl sulfide degradation II (oxidation)
PWY-6057	dimethyl sulfide degradation III (oxidation)

0 Unique Compound Selected

	Organism	Selection Mode	Entry ID	Name

Figure 8. Agilent MassHunter Pathways to PCDL software for pathway directed analysis.

New spectra can be added from MassHunter Qualitative Analysis, copied from another PCDL, or added from a journal reference.

Curated LC/MS/MS spectra can easily be added using the same import settings used by Agilent R&D, with the **Send Spectra to PCDL** auto-curation workflow integrated into MassHunter Qualitative Analysis Workflows (see Figure 6).

End users using the same import parameters when adding new spectra ensures that the entire spectral library will produce the same accurate compound identification results with consistent score rankings for all compounds in the spectral library.

MassHunter Qualitative Data Analysis allows multiple PCDLs to be daisy-chained together and queried sequentially, reducing data analysis time while performing a comprehensive search. One of the most common sequential DB/Library searching workflows starts with a small targeted compound PCDL, then moves on to a suspect compound PCDL, finally searching through a large general PCDL for all possible compounds. The end user can choose to report identifications from all PCDLs for each compound or save data processing time by moving on to the next compound once a compound has been successfully identified.

Conclusion: Agilent Curated Accurate Mass Personal Compound Databases and Libraries

High quality information in mass spectrometry databases and libraries is a necessary critical factor for fast, confident compound identification. The Agilent data collection and curation process ensures the quality of each compound entry and accurate mass spectrum in every Agilent MassHunter PCDL.

As part of the Agilent MassHunter software suite, PCDLs interface directly with data analysis, statistical analysis, and visualization software packages, providing confident compound identifications with consistent score rankings for a variety of compound data mining and identification workflows.

The MassHunter PCDL Manager Software package provided with each PCDL allows PCDLs to be customized by the end user to suit their workflow by the addition or removal of RTs, compounds, spectra and, for the IM-Q-TOF MS, ion mobility information.

Appendix: Agilent Accurate Mass Personal Compound Databases and Libraries products

Food Safety

Accurate Mass LC/MS/MS PCDL for LC/TOF and Q-TOF	Compounds	Compounds with Accurate Mass LC/MS/MS spectra	Total number of spectra	Compounds with RTs
Pesticides	1,750	>825	>2,700	0
Veterinary drugs	>2,150	>1,525	>5,200	104
Mycotoxins	>450	>300	>1,350	0
Accurate Mass GC/MS PCDL for GC/ Q-TOF	**Compounds**	**Compounds with Accurate Mass GC/MS spectra**	**Total number of spectra**	**Compounds with RTs**
Pesticides	>850	>850	>850	>840 (15 × 15, 20 and 40-minute methods) ~750 (5 × 15, 20-minute method)

Environmental

Accurate Mass LC/MS/MS PCDL for LC/TOF and LC/Q-TOF	Compounds	Compounds with Accurate Mass LC/MS/MS spectra	Total number of spectra	Compounds with RTs
Pesticides	1,750	>825	>2,700	0
Water contaminants	>1,400	>1,050	~3,900	268
Accurate Mass GC/MS PCDL for GC/ Q-TOF	**Compounds**	**Compounds with Accurate Mass GC/MS spectra**	**Total number of spectra**	**Compounds with RTs**
Pesticides	>850	>850	>850	>840 (15 × 15, 20 and 40-minute methods) ~750 (5 × 15, 20-minute method)

Forensic Toxicology

Accurate Mass LC/MS/MS PCDL for LC/TOF and LC/Q-TOF	Compounds	Compounds with Accurate Mass LC/MS/MS spectra	Total number of spectra	Compounds with RTs
Broecker, Herre, and Pragst Forensic Toxicology	>9,200	>3,900	>13,500	0

For Forensic Use.

Metabolomics

Accurate Mass LC/MS/MS PCDL for LC/TOF and LC/Q-TOF	Compounds	Compounds with Accurate Mass LC/MS/MS spectra	Total number of spectra	Compounds with RTs
Metabolites*	>249,450	>11,000	>37,260	>680
NIST 2014 MS/MS#	9,345	9,345	>234,000	0

*For Research Use Only. Not for use in diagnostic procedures.

# The NIST 2014 MS/MS PCDL consists of both accurate and unit mass LC/MS/MS spectra.

Pharma

Accurate Mass LC/MS/MS PCDL for LC/TOF and LC/Q-TOF	Compounds	Compounds with Accurate Mass LC/MS/MS spectra	Total number of spectra	Compounds with RTs
Extractables and Leachables	>1,000	>350	>1,300	129

www.agilent.com/chem

This information is subject to change without notice.

Printed in the USA, October 25, 2017
5991-8580EN

Agilent
Trusted Answers