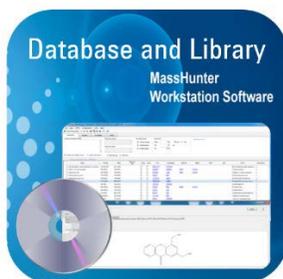


使用安捷伦优化的数据库和谱库提高化合物鉴定的可信度



作者

Emma E. Rennie、Robert H. Williams、Ruben Garnica 和 Maria VanDamme

前言

目标化合物与未知化合物的鉴定可靠性，取决于化合物鉴定所使用的数据库与谱库中所包含数据的真实性和完整性。大部分能够在网络上公开获取的数据库和谱库信息来源于众多供应商，通常以众包形式构建并且缺少必要的审核。质量差的谱库会对实验室结果的质量带来负面影响。

博物馆策展人或艺术收藏家会通过大量研究来评估他们收藏的艺术品或古器物中每一件展品的真实性和完整性。同样，安捷伦采用非常详细的优化过程，确保化合物数据库和谱库中每个数据条目及其精确质量谱图的质量。

本白皮书介绍了安捷伦如何开发精确质量数谱库和数据库，其优化过程的本质，以及优化过程为何能改善实验室产出结果的质量。

什么是安捷伦精确质量数个人化合物数据库与谱库？

安捷伦精确质量数个人化合物数据库与谱库 (PCDL) 是一个数据库，其中包含详细的化合物信息，以及使用四极杆飞行时间质谱仪 (Q-TOF MS) 测得的高分辨率精确质量谱图。PCDL 可用于气相色谱和液相色谱应用。使用 PCDL、Agilent MassHunter 数据分析软件以及 TOF 或 Q-TOF MS，可以对化合物进行多达四个维度的鉴定和确认，即精确质量数、同位素分布、精确质量数碎片谱图和保留时间 (RT)。使用安捷伦离子淌度四极杆 TOF (IM-Q-TOF) 仪器时，可获得另一维度的信息（即碰撞截面 (CCS)）用于进一步确认。最终用户无需采用化合物特异的参比标样，即可轻松鉴定特定的目标化合物或筛查数千个目标。使用每个 PCDL 附带的 Agilent MassHunter PCDL Manager 软件包，最终用户可通过添加 RT、新化合物、新谱图以及用于 IM-Q-TOF MS 的离子淌度信息来定制 PCDL，以满足特定分析或系统配置的需要。

完整的同位素质谱信息结合精确质量谱图，能够为最终用户提供高度可靠的化合物鉴定。针对采用的色谱方法添加 RT 信息，可以提供更高的可靠性。RT 尤其适用于分析同分异构化合物。将化合物添加到目标 PCDL，可以轻松实现历史样品的回顾性数据分析。采用与安捷伦化学家相同的导入设置，将自动优化工作流程集成到 Agilent MassHunter 定性分析工作流程软件中，可以轻松添加经过优化的全新 LC/MS/MS 谱图。

安捷伦精确质量数数据库和谱库可用于所有 Agilent MassHunter 系列软件。MassHunter PCDL Manager 可直接与 Agilent MassHunter 软件套装联用，包括数据分析软件包、定性分析和定量分析以及统计分析和可视化软件包：Mass Profiler、Mass Profiler Professional 和 Profinder。

所有 Agilent MassHunter PCDL 产品中都包括带数据示例和熟悉练习的用户指南，配备的升级保护还可确保用户在首次购买后能够持续获得最新信息。例如用于 GC/Q-TOF 的 MassHunter 农药 PCDL 和工作流程等 PCDL 中也包括电子方法、推荐配置和工作流程向导，有助于将 PCDL 和化合物鉴定工作流程快速部署到分析实验室。

安捷伦精确质量数个人化合物数据库与谱库的开发

每个 PCDL 均按照详细的标准流程创建，包括以下 4 个主要步骤：

1. 与顶尖专家共同确定目标化合物列表
2. 生成化合物数据库，并验证每个目标物条目的信息
3. 使用高纯度参比标样收集碎片谱图数据
4. 按照安捷伦严格的质量控制优化过程，用优化的碎片谱图创建谱库

第 1 步：与市场领域专家共同创建化合物列表

数据库和谱库创建过程的第一步是确定 PCDL 内需要包含或需要添加到 PCDL 的化合物。安捷伦与各学科的一流专家建立合作关系，对目标化合物列表进行优先级筛选，确保得到的产品对某一特定研究领域是全面、相关且最新的。

在精心挑选出每个 PCDL 产品的目录后（目录列表示例见图 1），对化合物进行优先级排序以进行碎片谱图采集。

第 2 步：为每个化合物创建数据库条目

为 PCDL 中包含的每个化合物创建一个数据库条目。每个条目中包括关于化合物的各种信息，每条信息都要经过人工验证。由经验丰富的安捷伦化学家根据各种权威来源（例如 SciFinder、ChemSpider、PubChem、ChemIDplus 或 Chemistry Dashboard）对分子式、名称、结构、CAS 登记号和化合物类别标签进行反复核对。另外，每个化学标识符都要对其来源位置进行人工验证。这样可以确保每个条目中化合物信息的高度准确性。

数据库中显示的特定化学标识符取决于 PCDL 的应用领域，包括的条目如：

- 通用名
- IUPAC 名称
- 异名
- 分子式
- CAS 登记号
- ChemSpider ID
- PubChem ID
- InChI
- SMILES
- 供应商产品编码

还包括 KEGG、METLIN ID、HMDB ID、LMP ID、ChEBI ID 和 BioCyc ID 等代谢物基于代谢物的标识符。

G3892AA MassHunter Pesticides GC/MS Q-TOF PCDL Content listing of 852 compounds with CAS numbers Version B.08.00 (February-2017)			
Compound Name	CAS #	Compound Name	CAS #
(1R)-cis-Permethrin	54774-46-8	Acetamidrid	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	Acclonifen	74070-46-5
1,2,4-Trichlorobenzene	120-82-1	Acinathrin (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	15972-80-8
1,3-Dichlorobenzene (M-Dichlorobenzene)	541-73-1	Aldrin	309-00-2
1,4-Dichlorobenzene (P-Dichlorobenzene)	106-46-7	Allsethrin	584-79-2
1-Naphthaleneacetic acid	86-87-3	Allidochlor	93-71-0
1-Naphthol	90-15-3	Ametoctradin	865318-97-4

图 1. 每个 PCDL 中包括一个目录列表，列出了所有化合物及其化学文摘社 (CAS) 登记号

化合物查看，包括最新的准确化合物信息：
通用名、IUPAC 名称、中性分子的精确质量数、分子式、ChemSpider、CAS、PubChem、KEGG、METLIN ID、HMDB ID、LMP ID、InChI、SMILES 和供应商产品编码等

Name	Formula	Mass	Retention Time	Cation	Anion	CAS	ChemSpider	METLIN	KEGG	HMP	LMP	IUPAC	NumSpectra
2,4-D-Terminylbenzoylphenylacetic acid ethyl ester	C19H21O2	316.12903				8464.112	1021038					Diethyl 2-terminylbenzoylphenylacetate	0
2,4-DCCA, 2,4-Dichloroacetate	ClH2O2	169.97991				56650-2	1368022	2047	C14419			2,4-Dichloroacetate	0
2,4-Daminocacetic acid	C7H10NO2	138.07931				615-05-4	13481	2261	C18218			4-Methyl-1,3-benzenediamine	0
2,4-Dichlorobenzoic acid	C7H4Cl2O2	189.95883				50-84-0	2583					2,4-Dichlorobenzoic acid	0
2,4-Dichlorobenzalcohol	C7H6Cl2O	175.97957				1777-82-8	14918					G,4-Dichlorophenylmethanol	0
2,4-Diethylthioanthrone	C17H18O2	268.09219				82799-44-8	109489					2,4-Diethyl-9H-thioanthrone-9-one	3
H140	C12H10N2	122.07316				195-67-8	1363124					2,4-dimethylphenol	0
HN204	H8N2O4	189.02951				87-62-9	3045	20262	C14713			2,4-Dinitrophenol	0
HN205	H8N2O5	184.01202				51-38-5	1449					2,4-dinitrophenol	3
HN206	H8N2O6	206.16707				96-26-4	2627					2,4-Bis(2-methyl-2-propenyl)phenol	2
HN204	H8N2O4	182.02276				121-11-2	3350					1-Methyl-2,4-dinitrobenzene	0
H11N	C12H11N	121.08915				95-68-1	1389562					1-Amino-2,4-dimethylbenzene	0

通用名：
从多个不同来源验证
通用性

结构和说明区域：
• 结构
• 化合物类别标签
• 法规标签和参考文献
• 中文、日语和英语异名
• 毒理学研究参考文献
• 已过时和替代的 CAS 号
• 化合物描述

图 2. Agilent MassHunter PCDL Manager，显示典型的化合物数据库条目

每个数据库条目中还包括每个化合物的化学结构与说明。说明的内容针对特定化合物，可包括：

- 化合物、法规和国家/地区特定的类别标签
- 中文、日语和英语异名
- 毒理学研究参考文献

- 已删除（或已过时）的替代 CAS 登记号
- 化合物描述

数据库中每个化合物条目的以上所有信息都能够在 PCDL Manager 中使用简单或高级搜索功能检索到（见图 2）。

这些条目可用于过滤数据库，由每个 PCDL 产品创建更小的可编辑 PCDL 子集。例如，最终用户可以建立定制 PCDL，为特定化合物类别或可疑物列表创建筛选工作流程。

最终，对照 Agilent Master PCDL 检查每个化合物条目，排除重复的化合物条目，以免数据分析中结果查看所需的时间大幅增加。

化合物鉴定信息（例如常用名称、CAS、结构和说明）通过 MassHunter 数据分析软件以及 MassHunter PCDL Manager 获得（见图 3）。

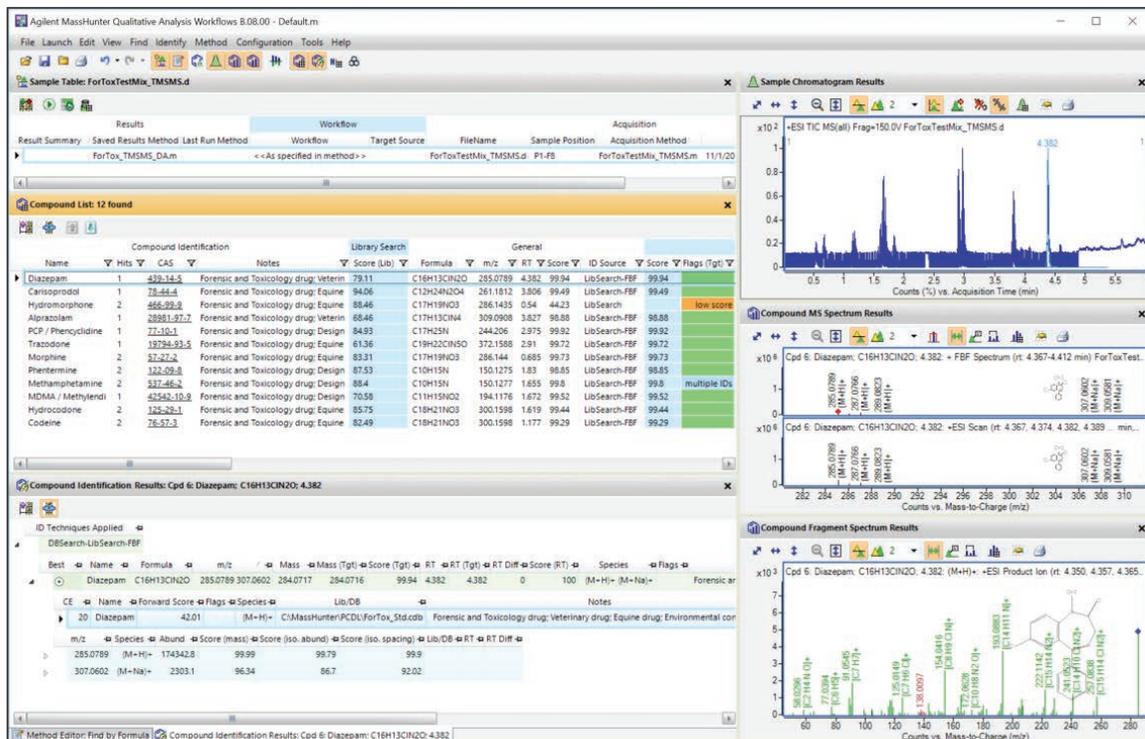


图 3. Agilent MassHunter 定性分析工作流程，显示 LC/MS/MS 分析的化合物鉴定，包括结构、CAS 和说明

第 3 步：使用高纯度参比标样收集质谱数据

使用单化合物参比标样获得精确质量谱图。所有数据均采用严格的标准操作规程 (SOP) 获得，包括仪器调谐和校准以及质量控制样品采集，确保所有数据都在相同的实验条件下获得。由专业领先的科学合作伙伴合成无市售单物质标样的化合物 (例如真菌或细菌代谢物)。

在 EI 全扫描模式下采集 GC/MS 精确质量谱图，产生大量碎裂模式和同位素离子簇信息 (见图 4)。

以正离子和负离子两种模式采集 m/z 最低的同位素分子离子 LC/MS/MS 精确质量谱图，能够获得多种加合离子形态和碰撞能量 (见图 5)。为了能够对不同稳定性的化合物提供丰富的碎片谱图，每个化合物分别在 10、20、40 V 的碰撞能量下采集。如果 10 V 谱图的碎片化程度仍过高 (对于极不稳定的化合物)，或 40 V 谱图的碎片化程度仍不足 (对于非常稳定的化合物)，那么根据情况在其他碰撞能量下采集谱图。

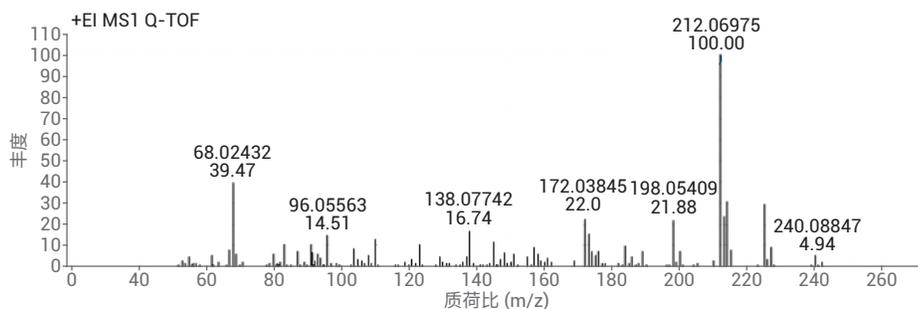


图 4. 农药氟草津的 GC/MS EI 精确质量谱图

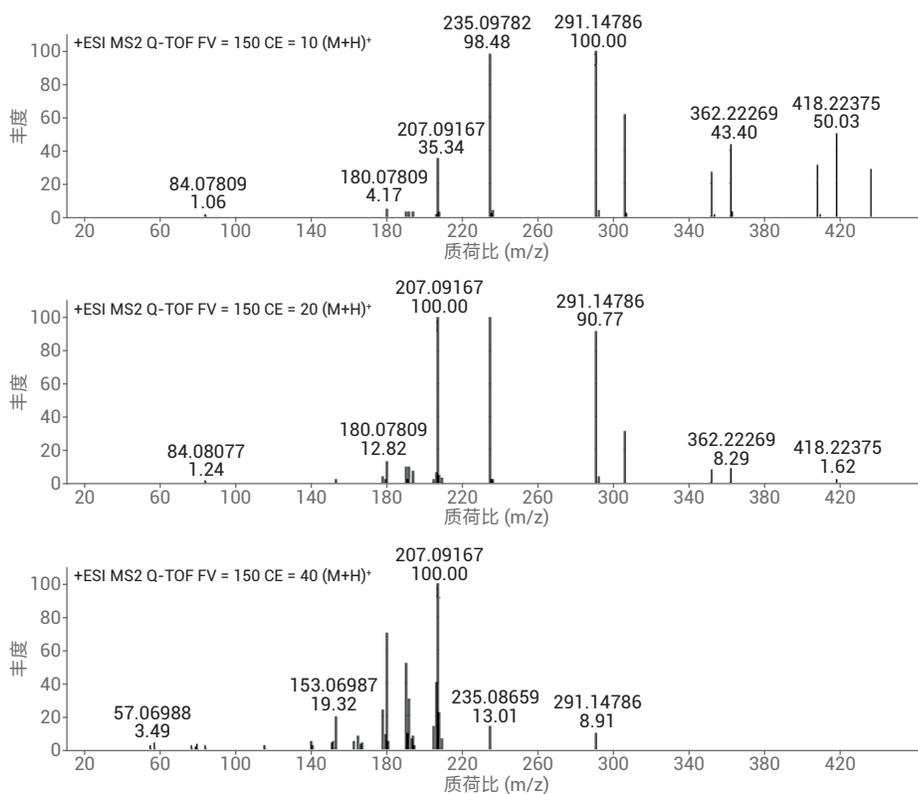


图 5. 药物缬沙坦的 LC/MS/MS ESI 精确质量谱图

第 4 步：为优先化合物列表创建谱库谱图条目

由安捷伦应用化学家或安捷伦合作专家之一按照适当的 SOP 采集质谱信息。SOP 的选择取决于所使用的色谱方法。

LC/MS/MS 谱图优化

精确质量 LC/MS/MS 谱图使用 MassHunter 定性分析软件中的算法进行提取与分析。首先，通过分子式查找算法提取谱图。分子式生成 (MFG) 为质谱中的每个 m/z 碎片离子峰分配一个分子式，然后对每个峰进行验证，以确定其是否是母离子分子式的有效子式。对于大多数碎片离子峰，在 MFG 质量数容差范围内存在许多可能的分子式分配。这些情况下，会将离子峰分配到质量数最接近的碎片分子式。这一过程中使用的 MFG 质量数容差是根据质量数变化的，是谱库搜索默认质量数容差的 1/3。谱图导入到 PCDL 后，将所有离子 m/z 峰校正到理论精确质量数（见图 6A 和图 6B）。无法分配到母离子的子式的碎片离子峰会被当做化学噪音或杂质去除。基峰 1% 或低于 1% 的碎片离子峰全部当作噪音去除。

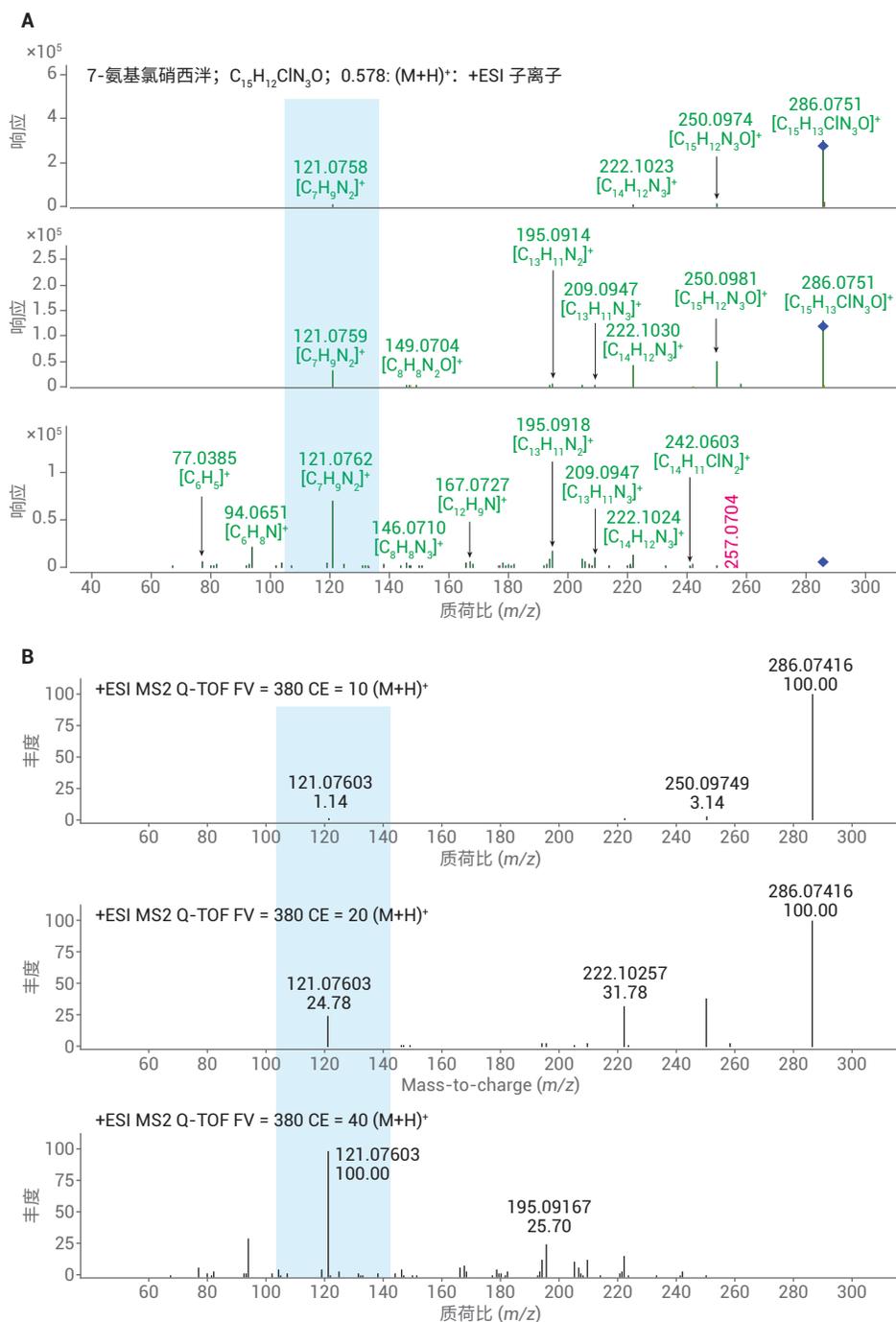


图 6. A) 7-氨基氯硝西洋的 LC/MS/MS ESI 实验精确质量谱图 B) 修正至理论精确质量数的 7-氨基氯硝西洋的 LC/MS/MS ESI 谱图。蓝色突出显示区域为质量数修正示例

GC/MS 谱图优化

精确质量谱图使用 MassHunter 定性分析软件中的算法进行提取与分析。首先，根据特征离子的提取离子色谱图 (EIC) 手动提取谱图。然后根据已知的化合物分子式，使用 MFG 将分子式分配到质谱图中的每个 m/z 碎片离子峰，充分考虑同位素比和间隔。对于大多数碎片离子峰，在 MFG 质量数容差范围内存在许多可能的分子式分配。这些情况下，会将离子峰以正确的同位素比和间隔分配到质量数最接近的碎片分子式。这一过程中使用的 MFG 质量数容差是根据质量数变化的，是谱库搜索默认质量数容差的 1/3。谱图导入到 PCDL 后，将所有离子 m/z 峰校

正到理论精确质量数。无法分配到母离子的子式的碎片离子峰会被当做化学噪音或杂质去除。离子计数为 100 或低于 100 的碎片离子峰全部当作噪音去除。

为了确保谱图质量高且便于搜索，由安捷伦科学家对数据集进行仔细检查，并根据以下原因自动去除：

- 错误设置的仪器参数
- 基峰离子计数小于 1000
- 作为噪音或杂质从谱图去除的碎片离子峰，降低谱库搜索分值
- 显示杂质、降解或转化产物，例如在单化合物参比标样的 LC/MS/MS 色谱图中发现的多个化合物或组分

作为所有 PCDL 谱图优化过程的最后一步，导入 PCDL 的每幅谱图都要经技术娴熟的专业质谱学家在 MassHunter 定性分析软件中直观确认。

每个条目都经过多个级别的审查，以确保数据的质量。优化过程确保所有数据以相同的条件进行采集和优化，避免数据采集不一致导致的错误。数据采集不一致是开放式或用户贡献谱库的主要缺点，可能会导致搜索结果质量差，以及使用这些数据库和谱库获得的结果可信度低。

对于法医毒理学药物化合物，LC/MS/MS 谱图化合物信息中包含每个谱图所用单化学品标样的内容，具有可追溯性。

如何使用安捷伦精确质量数个人化合物数据库与谱库

安捷伦 PCDL 可与 Agilent MassHunter 软件套装直接联用，能够提供分子式和参比谱库谱图以及化合物信息，确保为各种化合物数据挖掘和鉴定工作流程提供独特的化合物标识符。

从 PCDL、化合物数据挖掘和鉴定流程中移除非目标化合物，为特定分析过程（例如代谢组学通路、化合物类别、法规或特定国家/地区列表）创建小型靶向 PCDL 能够显著提高数据分析的效率和准确性。

可通过以下方式轻松创建小型靶向 PCDL:

1. 在 PCDL Manager 中搜索相关的化合物类别、法规或国家/地区特定标签（示例见图 7）
2. 选中所需化合物结果
3. 选中**创建 PCDL 子集**或**添加到 PCDL**并单击鼠标右键

代谢组学工作流程可以使用 MassHunter Pathways to PCDL 软件为通路导向的数据分析创建 PCDL（见图 8）。

出现新的分析物时，MassHunter PCDL Manager 导入功能可以轻松为以下操作提供支持：

- 导入新的化合物以及化学标识符
- 添加或更新化合物或化合物子集的 RT
- 导入 CCS 值以及其他详细的离子淌度信息 (IM-Q-TOF LC/MS)

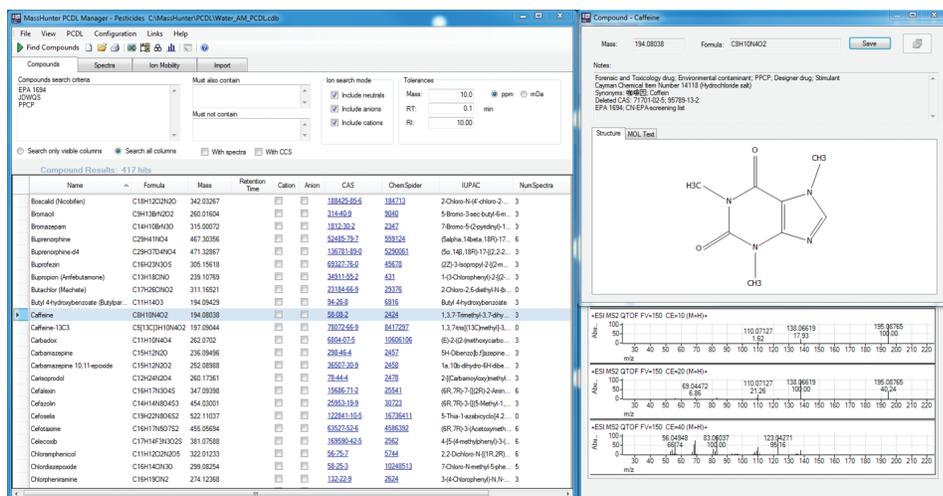


图 7. Agilent MassHunter PCDL Manager B.08.00，显示了美国和日本政府法规标签（EPA 1694 和 JDWQS）以及药物和个人护理产品 (PCPP) 化合物类别标签的搜索结果

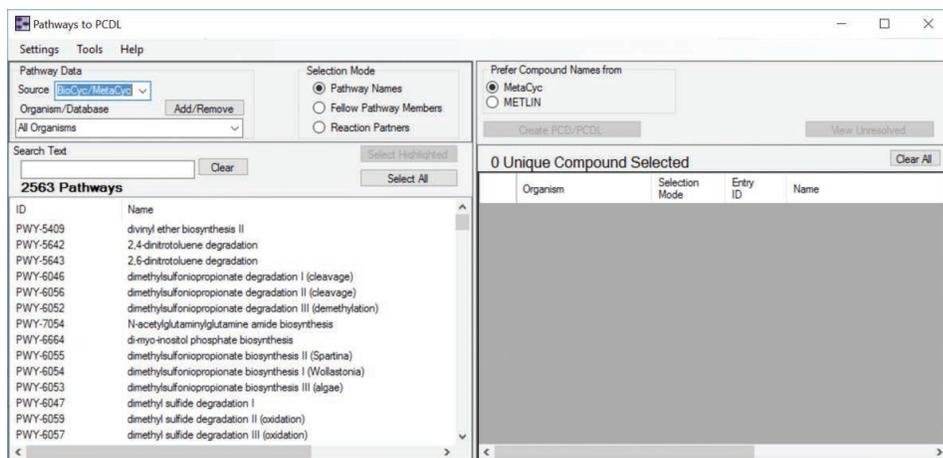


图 8. 用于通路导向分析的 Agilent MassHunter Pathways to PCDL 软件

新的谱图可通过 MassHunter 定性分析软件添加、从另一 PCDL 复制或从参考期刊添加。

使用与安捷伦研发团队相同的导入设置，**将谱图发送至 PCDL** 自动优化流程集成到 MassHunter 定性分析工作流程，可以轻松地完成优化 LC/MS/MS 谱图的添加（见图 6）。

最终用户在添加新谱图时，使用相同的导入参数可以确保整个谱库能生成同样的准确化合物鉴定结果，并保证谱库中所有化合物具有一致的评分。

MassHunter 定性分析软件支持多个 PCDL 串联并依次查询，从而能够执行全面搜索并缩短数据分析时间。一种最常见的序列是数据库/谱库搜索工作流程以小

型靶向化合物 PCDL 开始，然后转向可疑化合物 PCDL，最后在大型综合 PCDL 中搜索所有可能的化合物。最终用户可以选择通过所有 PCDL 报告每个化合物的鉴定结果，也可以选择成功鉴定一个化合物后转向下一个化合物，以节省数据处理时间。

结论：安捷伦优化精确质量数个人化合物数据库与谱库

质谱数据库与谱库中的高质量信息对于实现快速、可靠的化合物鉴定而言非常关键。安捷伦数据采集和优化流程确保了每个化合物条目和每个 Agilent MassHunter PCDL 中精确质量谱图的质量。

作为 Agilent MassHunter 软件套装的一部分，PCDL 能够直接与数据分析、统计分析、和可视化软件包联用，为各种化合物数据挖掘和鉴定工作流程提供一致的评分和可靠的化合物鉴定。

每个 PCDL 附带的 MassHunter PCDL Manager 软件包允许最终用户通过添加或移除 RT、化合物、谱图以及用于 IM-Q-TOF MS 的离子淌度信息来定制 PCDL，以满足工作流程的需求。

附录：安捷伦精确质量数个人化合物数据库与谱库产品

食品安全

用于 LC/TOF 和 Q-TOF 的精确质量数 LC/MS/MS PCDL	化合物	具有精确质量数 LC/MS/MS 谱图的化合物	总谱图数	具有 RT 的化合物
农药	1750	> 825	> 2700	0
兽药	> 2150	> 1525	> 5200	104
真菌毒素	> 450	> 300	> 1350	0
用于 GC/ Q-TOF 的精确质量数 GC/MS PCDL	化合物	具有精确质量数 GC/MS 谱图的化合物	总谱图数	具有 RT 的化合物
农药	> 850	> 850	> 850	> 840 (15 × 15、20、40 分钟方法) 约 750 (5 × 15、20 分钟方法)

环境

用于 LC/TOF 和 LC/Q-TOF 的精确质量数 LC/MS/MS PCDL	化合物	具有精确质量数 LC/MS/MS 谱图的化合物	总谱图数	具有 RT 的化合物
农药	1750	> 825	> 2700	0
水污染物	> 1400	> 1050	约 3900	268
用于 GC/ Q-TOF 的精确质量数 GC/MS PCDL	化合物	具有精确质量数 GC/MS 谱图的化合物	总谱图数	具有 RT 的化合物
农药	> 850	> 850	> 850	> 840 (15 × 15、20、40 分钟方法) 约 750 (5 × 15、20 分钟方法)

法医毒理学

用于 LC/TOF 和 LC/Q-TOF 的精确质量数 LC/MS/MS PCDL	化合物	具有精确质量数 LC/MS/MS 谱图的化合物	总谱图数	具有 RT 的化合物
Broecker、Herre 和 Pragst 法医毒理学	> 9200	> 3900	> 13500	0

用于法医鉴定。

代谢组学

用于 LC/TOF 和 LC/Q-TOF 的精确质量数 LC/MS/MS PCDL	化合物	具有精确质量数 LC/MS/MS 谱图的化合物	总谱图数	具有 RT 的化合物
代谢物*	> 249450	> 11000	> 37260	> 680
NIST 2014 MS/MS#	9345	9345	> 234000	0

* 仅限研究使用。不可用于诊断目的

NIST 2014 MS/MS PCDL 包括精确质量数和单位质量数 LC/MS/MS 谱图

制药

用于 LC/TOF 和 LC/Q-TOF 的精确质量数 LC/MS/MS PCDL	化合物	具有精确质量数 LC/MS/MS 谱图的化合物	总谱图数	具有 RT 的化合物
可萃取物与可浸出物	> 1000	> 350	> 1300	129

www.agilent.com

本文中的信息、说明和指标如有变更，恕不另行通知。