Application Note

Environmental Analysis

Agilent

Trusted Answers

Accelerating Confident Unknown Identification Using Agilent Revident LC/Q-TOF and MassHunter Explorer 2.0

Analysis of water samples

Authors

Udi Jumhawan,
Sureshbabu C. V., and Eric Tan
Agilent Technologies, Inc.

Abstract

Environmental monitoring of water samples requires comprehensive screening approaches to detect emerging and established pollutants beyond targeted compound analysis. This study demonstrates an integrated workflow combining high-resolution Agilent Revident Quadrupole Time-of-Flight LC/MS with MassHunter Explorer 2.0 software for unknown compound identification in a fortified water sample. A streamlined analysis pipeline incorporating base peak chromatogram comparison, chemometric analysis, and library matching to personal compound databases successfully detected and identified 20 pesticides with excellent mass accuracy (less than 1 ppm) and reduced data complexity by greater than 90% (from 528 to 51 features). Integration of machine learning-based SIRIUS with CSI: FingerID validated putative identifications and eliminated false positives. The automated workflow significantly reduced manual peak picking effort and processing time while maintaining high confidence in compound identification. Overall, this workflow provides a powerful solution for unknown compound identification, with broad applicability to environmental quality control, food safety, and biomarker discovery settings.

Introduction

The assessment of water quality and environmental contamination represents a critical component of public health protection and regulatory compliance. Traditional approaches to water quality monitoring have relied on targeted analysis of a predefined set of known contaminants, enabling quantification of specific compounds of regulatory concern. However, this methodology presents a significant limitation: it provides only a narrow view of the actual pollutant burden present in environmental samples.[1] By focusing exclusively on known analytes, targeted screening approaches may substantially underestimate both the exposure to, and the risk posed by, emerging contaminants, previously unidentified pollutants, and transformation products that escape conventional detection schemes.

The shift toward non-targeted analysis using high-resolution mass spectrometry addresses this critical knowledge gap by enabling comprehensive, unbiased screening of complex environmental matrices. High-resolution liquid chromatography coupled with quadrupole time-of-flight mass spectrometry (LC/Q-TOF) provides the analytical sensitivity and mass accuracy necessary to detect and characterize unknown compounds across a broad chemical space. However, the application of such powerful analytical platforms generates substantial data complexity; non-targeted datasets typically contain hundreds to thousands of detected features, the vast majority of which represent background noise, matrix components, or artifacts rather than true pollutants of interest.

The bottleneck in non-targeted analysis has traditionally resided not in data acquisition but in data interpretation. Manual peak picking, feature alignment, and compound identification are labor-intensive processes prone to operator bias and inconsistency. Furthermore, distinguishing true positive identifications from false positives—compounds present in both sample and control matrices—requires sophisticated statistical and chemometric approaches integrated seamlessly with spectral library matching and structure elucidation tools.

This study demonstrates a streamlined, integrated workflow for unknown compound identification in environmental water samples that combines high-resolution LC/Q-TOF analysis with advanced data processing and chemometric analysis. By leveraging automated feature extraction, fold-change statistical analysis, multi-database spectral matching, and machine learning-based structure prediction, this approach substantially reduces analytical workload while improving confidence in compound identification and minimizing the risk of false positive assignments. The methodology is presented as a generalizable solution applicable to diverse fields, including environmental monitoring, food safety, and biomarker discovery.

Experimental

Chemicals and solvents

Agilent InfinityLab LC/MS-grade methanol (part number 5191-5111) was obtained from Agilent Technologies. LC/MS-grade formic acid was purchased from Fisher Scientific (Massachusetts, USA). LC/MS-grade ammonium formate was obtained from Sigma-Aldrich (Missouri, USA). A fortified water sample and control were provided by collaborator and were ready for direct injection to LC/MS without further preparation.

Instrumentation

Liquid Chromatography System	
Module	**Part Number**
Agilent 1290 Infinity III High-Speed Pump	G7120A
Agilent 1290 Infinity III Multisampler	G7167B
Agilent 1290 Infinity III Multicolumn Thermostat	G7116B
Agilent InfinityLab Assist Hub	G7180A
Mass Spectrometry System	
Module	**Part Number**
Agilent Revident Quadrupole Time-of-Flight LC/MS	G6575A
Agilent Dual Spray Jet Stream Technology Ion Source (AJS)	G1959A

LC/MS methods

Experimental methods for LC conditions, Q-TOF settings, and acquisition parameters are shown in Tables 1 and 2.

Table 1. Agilent 1290 Infinity III LC parameters.

Parameter	Value
Analytical Column	Altura ZORBAX Eclipse Plus C18 2.1 × 100 mm, 1.8 µm (part number 204210-308)
Column Temperature	45 °C
Autosampler Temperature	4 °C
Mobile Phase	A) 10 mM ammonium formate pH 5 B) 10 mM ammonium formate pH 5 in methanol
Flow Rate	0.45 mL/min
Gradient Program	Time (min) %B 0 2 0.25 2 12.25 99 15.00 99 15.01 2 17 2
Injection Volume	20 µL

Table 2. Agilent Revident Q-TOF instrument parameters.

Parameter	Value
Source Parameters	
Instrument Mode	1,700 stable
Ion Source	Agilent dual Jet Stream (dual AJS)
Polarity	Positive
Gas Temperature	325°C
Drying Gas Flow	8 L/min
Nebulizer	45 psi
Sheath Gas Temperature	325°C
Sheath Gas Flow	8 L/min
Capillary Voltage	4,000 V
Nozzle Voltage	1,000 V
Fragmentor	65 V
Skimmer	45 V
Octopole RF	750 V
Reference Mass	*m/z* 121.050873
Reference Nebulizer	8 psi
MS Method Parameters	
MS Range	*m/z* 50–500
MS Acquisition Rate	3 spectra/s
Directed MS/MS Method Parameters	
MS Range	*m/z* 50–500
MS/MS Range	*m/z* 50–500
MS Acquisition Rate	8 spectra/s
MS/MS Acquisition Rate	6 spectra/s
Isolation Width	Narrow (~*m/z* 1.3)
Delta	20 ppm
RT Window	0.01 min
Collision Energy	Fixed: 40 eV
Maximum Precursor per Cycle	1
Variable Acquisition Rate	No
Precursor Threshold	3,000 counts and 0.001%
Active Exclusion	Disabled
Purity	Stringency 100%, cutoff 30%
Isotope Model	Common organic molecules
Iterative MS/MS Mass Error Tolerance	20 ppm
Iterative MS/MS RT Exclusion Tolerance	±0.02 min

Software and data analysis

Agilent MassHunter Acquisition software for LC/MS (Q-TOF), version 12.1, was used for data collection. Agilent MassHunter Qualitative Analysis, version 12.0 update 1, was employed for base peak chromatogram (BPC) comparison. Agilent MassHunter Explorer software, version 2.0, was used for feature extraction, statistical analysis, and compound identification.

Peak and alignment settings in MassHunter Explorer 2.0 were applied using the default environmental method with the following modifications: height filter at 5,000 counts; gap filling disabled; multi-pass exhaustive grouping disabled; isotope model at common organic; retention time tolerance at ±0.02 minutes.

Feature extraction was performed on both the fortified water sample and the control. Fold change (FC) analysis (FC >10) was applied to identify significant features with elevated abundance in the fortified sample relative to the control. Additionally, MassHunter Explorer 2.0 provides other means of statistical analysis such as volcano plot, unique features, hierarchical clustering analysis, and two-way ANOVA.

Compound identification

The ions of interest were subjected to library matching against four Agilent Personal Compound and Database Library (PCDL) databases: pesticides, EPA1699 water quality screening, extractable and leachable (E&L), and MassBank of North America (MoNA). The four PCDL databases were stored in Agilent ChemVista software. Agilent ChemVista manages spectral libraries created by LC/Q-TOF or GC/Q-TOF mass spectrometry. It integrates compound details, spectral information, and retention time, when available, from multiple sources.

The MS and MS/MS mass tolerance were set at 5.0 and 15.0 ppm, respectively. Putative identifications from PCDL matching were further validated using SIRIUS with CSI: FingerID software (BrightGiant GmbH, Jena, Germany), which facilitates prediction of the most accurate elemental composition from high-resolution isotope patterns and complex fragmentation patterns. Subsequently, it ranks candidate structures within molecular structure databases through machine learning and molecular fingerprinting approaches.

Results and discussion

Identification workflow and data processing

The study developed a streamlined, integrated workflow to identify unknown organic pollutants in fortified water using high-resolution mass spectrometry combined with chemometric analysis (Figure 1). The workflow integrated base peak chromatogram (BPC) comparison, feature extraction, statistical analysis, library matching, and structure-based confirmation. Analysis of a 20 µL fortified water sample by LC/Q-TOF mass spectrometry in full scan mode revealed 20 distinctive peaks in the BPC when compared to a water control (Figure 2).

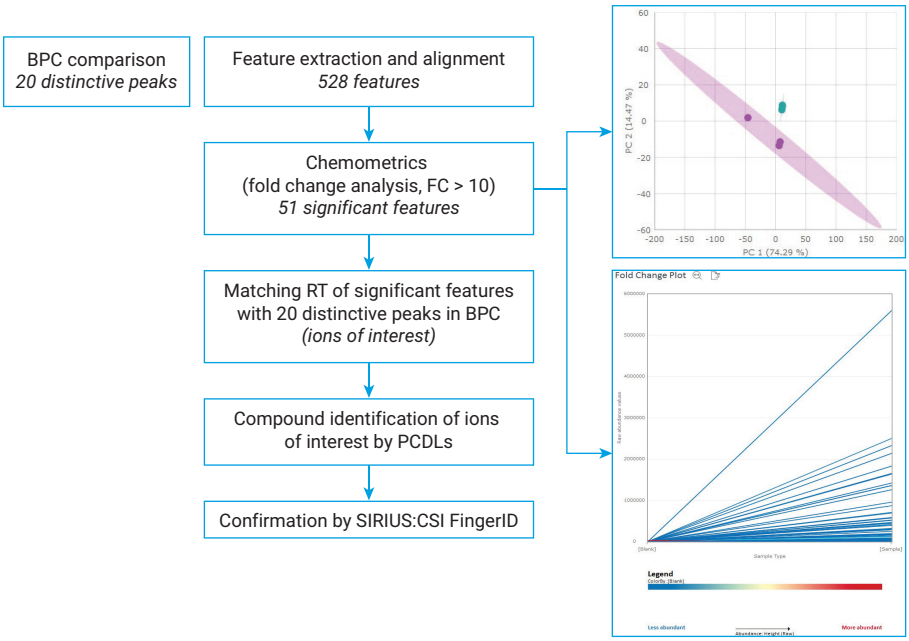


Figure 1. Workflow schematic for unknown compound identification. The integrated approach combines BPC comparison, feature extraction and alignment (528 features), chemometric analysis including PCA plot generation and fold change analysis (FC > 10, resulting in 51 significant features) in Agilent MassHunter Explorer 2.0, achieving greater than 90 percent data reduction, retention time matching to ions of interest, library matching to Agilent PCDLs, and validation by SIRIUS: CSI FingerID.

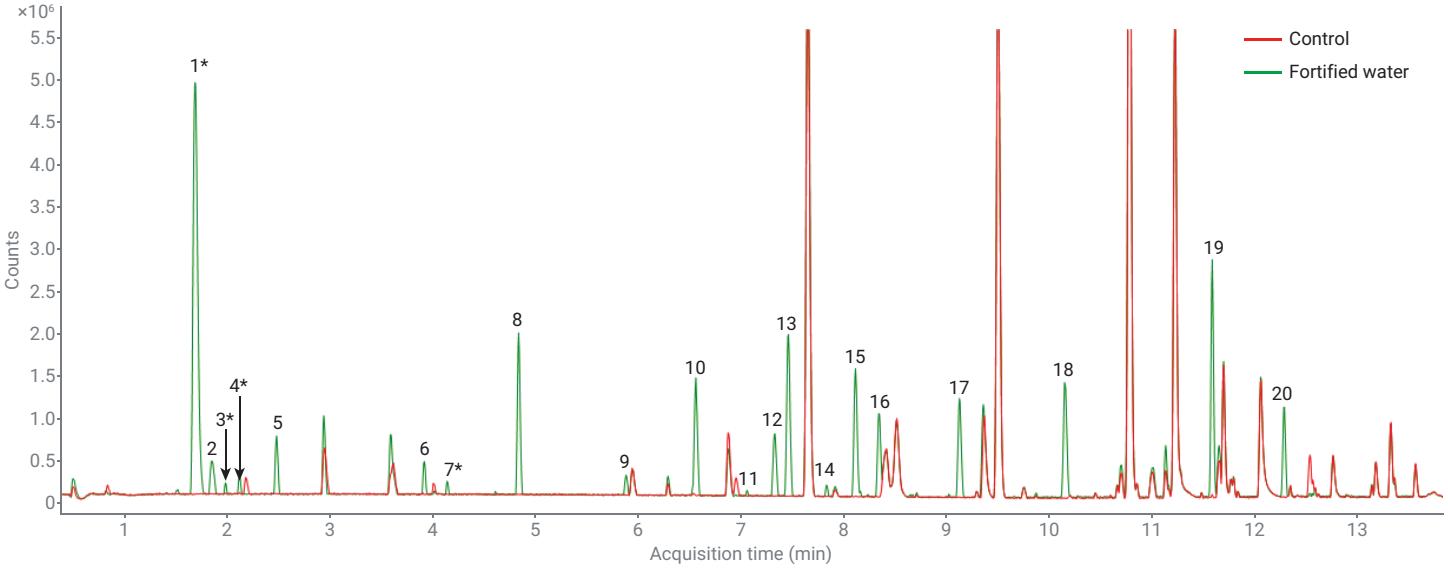


Figure 2. BPC comparison of fortified water sample and control. The fortified water sample shows 20 distinctive peaks (numbered 1–20) compared to the control. Peaks marked with asterisks (1*, 3*, 4*, 7*) represent false positives identified through chemometric analysis and MS/MS confirmation.

Feature extraction using MassHunter Explorer 2.0 generated 528 total features. Application of fold change analysis (FC > 10) to identify features with significantly higher abundance in the fortified sample compared to the control narrowing the feature list to 51 candidates, representing > 90% data reduction. This chemometric filtering effectively eliminated features with equal abundance in both samples, thereby removing the majority of potential false positives and substantially reducing analysis time.

Compound identification by PCDLs, validation with SIRIUS: CSI FingerID, and false positive removal

The 51 features were exported for directed MS/MS data acquisition, and the resulting MS/MS spectra were analyzed within the same MassHunter Explorer project.[2] Significant features with RT matching BPC peaks are initially considered as the ions of interest; these were subjected to library matching against four Agilent personal compound database and library (PCDL) resources: pesticides, EPA1699 water quality screening, extractable and leachable (E&L), and MoNA. Mass spectrometry and MS/MS tolerances were set at 5.0 and 15.0 ppm, respectively. Putative identifications from PCDL matching were verified using SIRIUS with CSI: FingerID, which predicts molecular formulas from accurate MS/MS spectra and ranks candidate structures using machine learning and molecular fingerprinting against structure databases containing more than 100 million structures.[3,4]

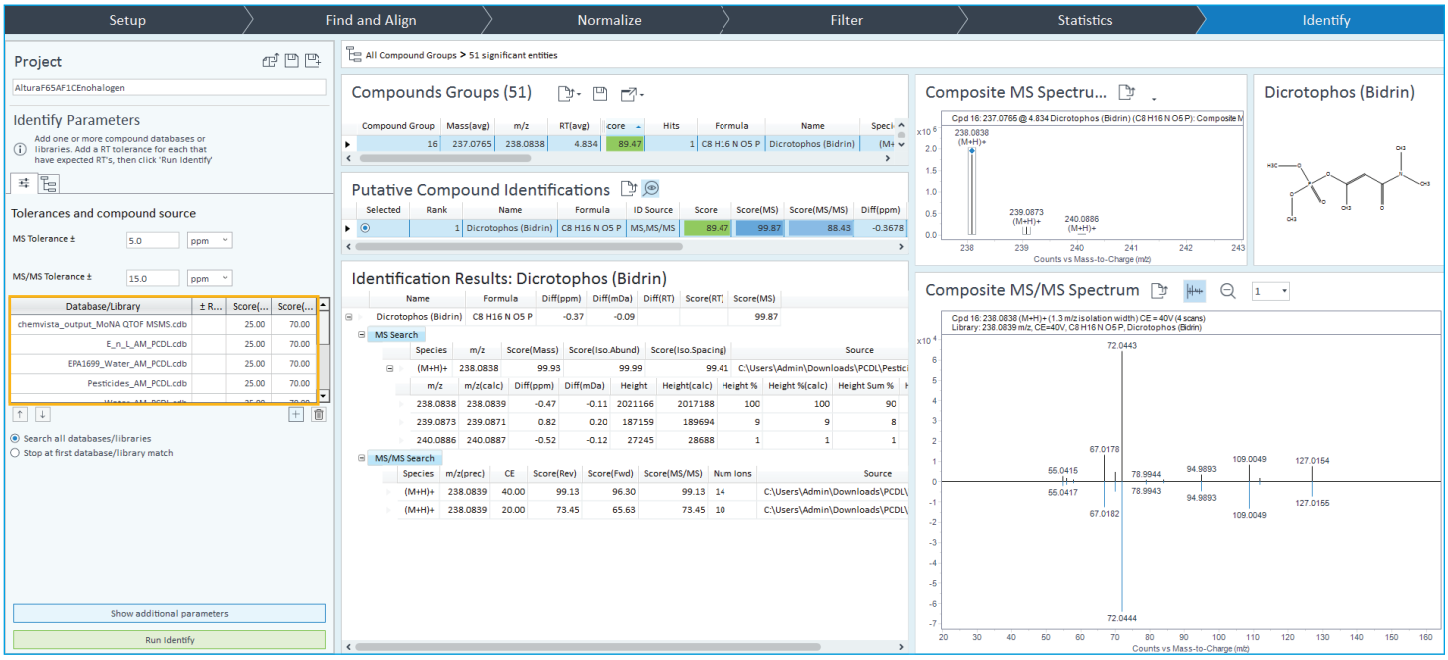
Out of the 20 peaks showing BPC differences between fortified and control samples, 17 were putatively identified as pesticides by matching MS/MS spectra to Agilent PCDLs. All 17 pesticides demonstrated excellent mass accuracy (less than 1 ppm) and relatively good MS and MS/MS identification scores (above 81)(Table 3). An example identification was dicrotophos, with formula $C_8H_{16}NO_5P$, identified at *m/z* 238.0838 $(M+H)^+$ with a mass accuracy of −0.37 ppm and MS/MS score of 88.43 (Figure 3).

Table 3. Putative compound identification by PCDLs and SIRIUS in fortified water.

No.	Peak No. in BPC	Putative Identification	RT (min)	*m/z*	Adduct	Mass Accuracy (ppm)	MS/MS Similarity Score	EIC Area
1	2	Methamidophos	1.86	142.0086	$(M+H)^+$	0.16	90.38	994,628
2	5	Acephate	2.48	184.0193	$(M+H)^+$	0.69	83.71	922,812
3	6	Methomyl	3.92	163.0536	$(M+H)^+$	0.41	99.21	380,218
4	8	Dicrotophos	4.83	238.0838	$(M+H)^+$	−0.37	88.43	2,335,428
5	9	Atrazine-desethyl	5.88	188.0698	$(M+H)^+$	0.74	99.51	551,687
6	10	Metoxuron	6.56	229.0738	$(M+H)^+$	−0.13	97.89	2,017,534
7	11	Cyanazine	7.06	241.0963	$(M+H)^+$	0.09	95.87	226,093
8	12	Simazine	7.32	202.0855	$(M+H)^+$	0.54	99.05	1,326,981
9	13	Hexazinone	7.46	253.1659	$(M+H)^+$	0.40	99.62	2,856,973
10	14	Monolinuron	7.83	215.0582	$(M+H)^+$	0.09	95.51	315,527
11	15	Chlorotoluron	8.11	213.0790	$(M+H)^+$	0.17	97.98	2,286,958
12	*	Metobromuron	8.16	259.0077	$(M+H)^+$	0.31	99.16	213,924
13	16	Atrazine	8.34	216.1012	$(M+H)^+$	0.69	98.12	1,664,839
14	*	Diuron	8.49	233.0243	$(M+H)^+$	−0.23	95.83	887,229
15	*	Linuron	9.09	249.0192	$(M+H)^+$	−0.33	89.20	87,682
16	17	Sebuthylazine	9.12	230.1167	$(M+H)^+$	0.39	99.76	1,838,875
17	*	Terbuthylazine	9.35	230.1167	$(M+H)^+$	0.36	99.65	1,745,850
18	18	Metolachlor	10.15	284.1411	$(M+H)^+$	−0.54	93.77	2,106,579
19	19	Buprofezin	11.58	306.1632	$(M+H)^+$	−0.79	83.06	2,791,088
20	20	Fenpropimorph	12.28	304.2635	$(M+H)^+$	0.01	81.62	1,925,226

*No difference in BPC

A MassHunterExplorer 2.0 putative identification result interface



B SIRIUS CSI: FingerID result

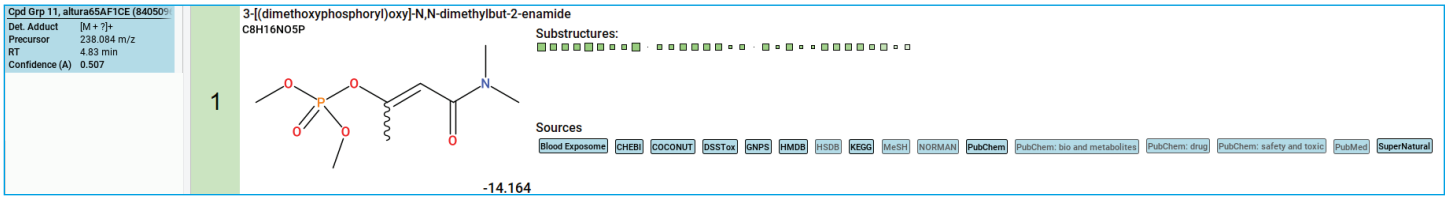


Figure 3. Putative identification of dicrotophos in MassHunter Explorer 2.0. (A) The interface of putative identification shows composite MS and MS/MS spectrum (mirror plot against library) and identification results table. The left panel shows a list of Databases/Libraries (indicated with yellow rectangle) used for database matching. (B) Confirmation of compound identification by SIRIUS CSI: FingerID; dichrotophos as the first hit for 238.084 *m/z* on structure annotation.

Four peaks (peaks 1, 3, 4, and 7) showed differences in BPC between fortified and control samples but were considered as potential false positivesThe four peaks showed no unique features when comparing the two samples; only different in peak abundance. These potential false positives were confirmed as non-matches by SIRIUS (no identification match by MS/MS) and removed from further analysis. This result demonstrated that evaluating results solely based on BPC differences could lead to misidentification and that integration of chemometric analysis and MS/MS confirmation was essential for accurate compound identification.

Three additional compounds (diuron, linuron, and metobromuron) were identified from the remaining significant features generated by chemometric analysis, despite showing identical or minimal BPC differences between fortified water and control. As an example, diuron ($C_9H_{10}Cl_2N_2O$) was putatively identified at *m/z* 233.0243 with excellent matching score (greater than 95) and displayed the characteristic isotopic pattern expected for a halogenated compound (Figure 4). This demonstrates the prowess of chemometric filtering for assisting selection of true analytes.

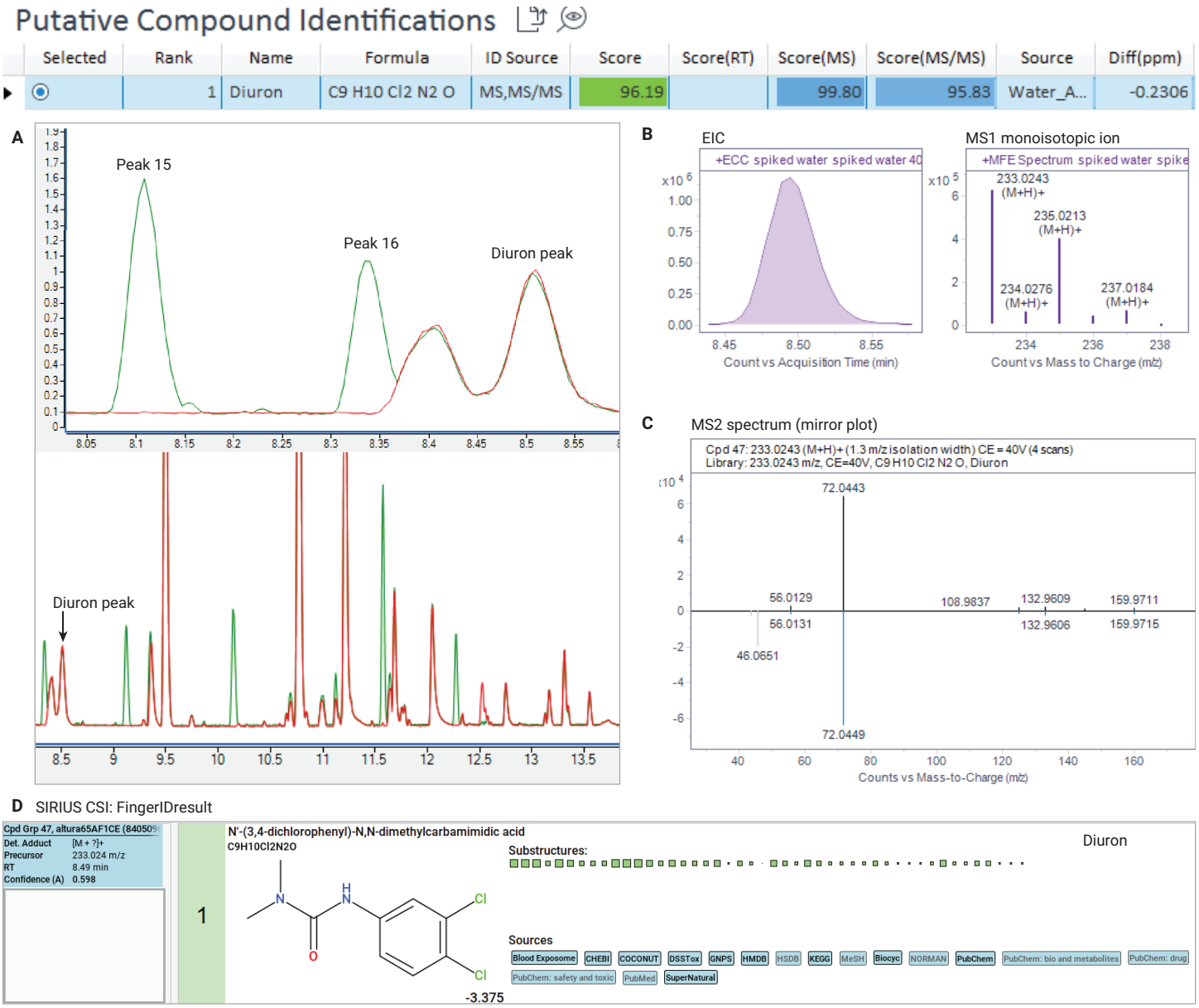


Figure 4. Identification of diuron despite identical BPC in fortified water and control. (A) BPC showing identical diuron signal in fortified water and control (inset). (B) Extracted ion chromatogram (EIC) showing diuron peak at 8.49 min with *m/z* 233.0243 $(M+H)^+$ and MS1 spectrum displaying characteristic isotopic pattern for a dichlorinated compound. (C) MS2 spectrum mirror plot matching the sample to the library spectrum with fragment ions at *m/z* 56.0129 and 72.0443. (D) Confirmation of compound identification by SIRIUS CSI: FingerID; diuron as the first hit for 233.024 *m/z* on structure annotation.

Two structural isomers, sebuthylazine and terbuthylazine, both with formula $C_9H_{16}ClN_5$ and *m/z* 230.1167 $(M+H)^+$, were successfully differentiated by retention time (Figure 5). Sebuthylazine was identified at 9.12 minutes and terbuthylazine at 9.35 minutes. Although both compounds produced identical MS/MS spectra, retention time served as valuable information for putative identification and isomer differentiation. Confirmation by LC/triple quadrupole pesticide database analysis verified the identifications at the respective retention times.[5]

In total, 20 pesticides were detected with mass accuracy less than 1 ppm and sensitivity ranging from 87,000 to 2,800,000 counts using a 20 µL injection volume in positive ion mode (Table 3). Five pesticides (chlorotoluron, cyanazine, diuron, linuron, metoxuron) were also detected in negative ion mode.

Advantages over manual workflow

The integrated MassHunter Explorer 2.0 workflow provided significant advantages over traditional manual data processing. The project-based, guided workflow streamlined peak picking, statistical analysis, library matching, and confirmation steps. This workflow allows users to edit and incorporate additional data into the same project, demonstrating the software's flexibility and offering a user-friendly option. Involvement of chemometrics successfully eliminated false positives while reducing processing time and user effort for manual peak picking. The combination of accurate mass determination, MS/MS spectral matching, and structure-based confirmation using SIRIUS provided high confidence in compound identification. This approach is applicable to diverse research areas including environmental monitoring, food safety contamination assessment, and clinical analysis, offering a powerful solution for unknown compound identification in broad research and development settings.

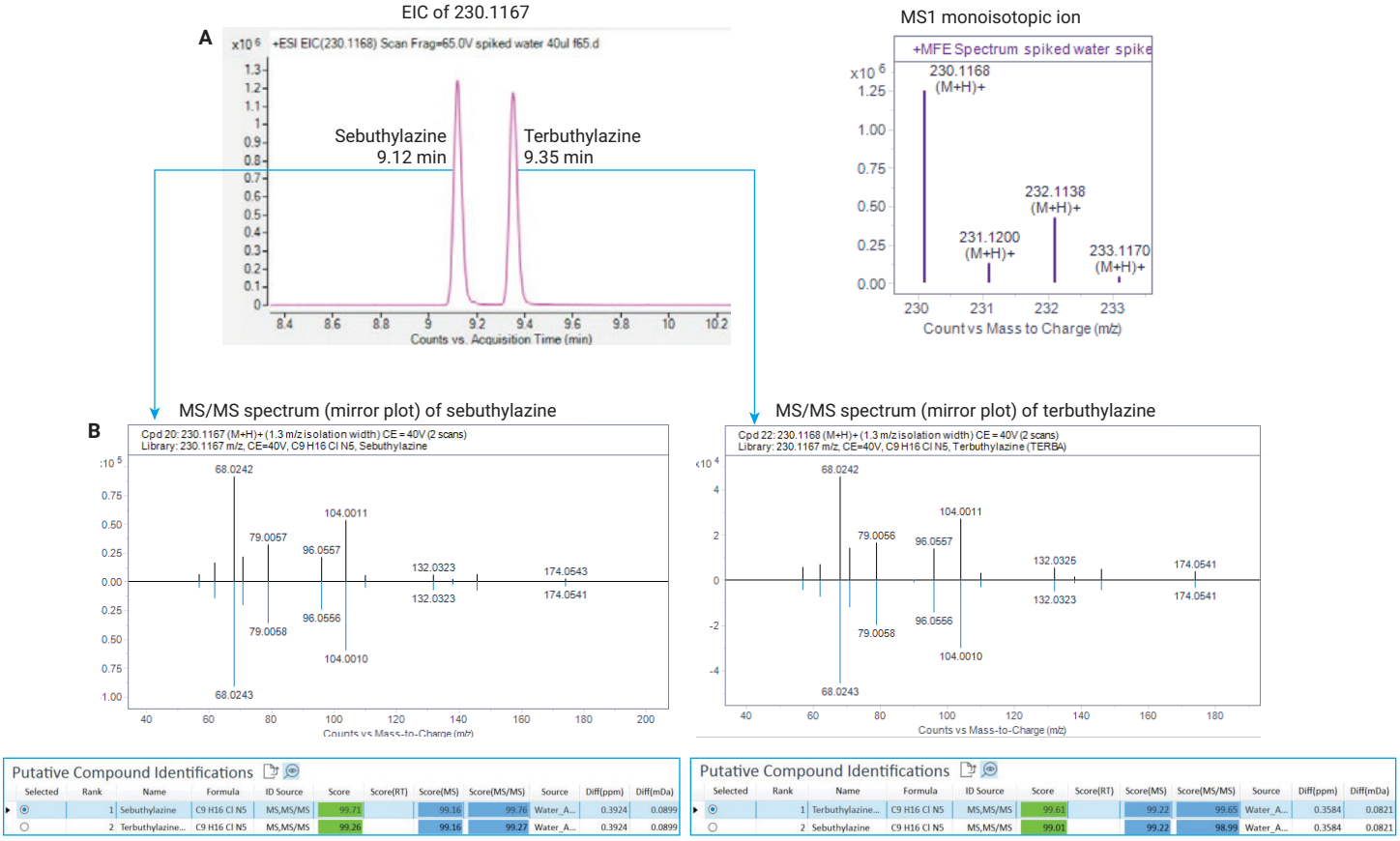


Figure 5. Differentiation of structural isomers sebuthylazine and terbuthylazine. (A) Extracted ion chromatogram (EIC) of *m/z* 230.1167 showing baseline resolution of the two isomers at 9.12 minutes (sebuthylazine) and 9.35 minutes (terbuthylazine). (B) MS/MS spectra (mirror plots) show identical fragmentation patterns, demonstrating that retention time is critical for isomer differentiation.

Conclusion

Unknown compound identification in a fortified water sample was successfully demonstrated using an Agilent Infinity III UHPLC coupled to an Agilent Revident LC/Q-TOF. The integrated workflow combined base peak chromatogram comparison, chemometric analysis (fold change), and library matching to Agilent personal compound databases with SIRIUS validation using MassHunter Explorer 2.0. Incorporation of fold change analysis (> 90% data reduction from 528 to 51 features) effectively eliminated false positives while reducing manual peak picking effort and analysis bias. Twenty pesticides were detected in fortified water and putatively identified with excellent mass accuracy (less than 1 ppm) and confirmed via SIRIUS molecular formula prediction and structure database searching. This streamlined workflow offers a powerful and scalable solution for non-targeted compound identification applicable to environmental monitoring, food safety, and biomarker discovery settings.

References

1. Nieto, S.; Chen, K.; Alaimo, C.; Young, T. Comprehensive Profiling of Environmental Contaminants in Surface Water Using High-Resolution GC/Q-TOF. *Agilent Technologies application note*, publication number 5994-1371EN, **2019**.
2. Sartain, M.; Bertram, L.; McEachran, A.; Pyke, J.; Hoffmann, M. A. Drug Metabolite Identification with a Streamlined Software Workflow: Combining Agilent Revident Q-TOF LC/MS and MassHunter Explorer 2.0. *Agilent Technologies application note*, publication number 5994-8867EN, **2026**.
3. Dührkop, K.; Fleischauer, M.; Ludwig, M.; Aksenov, A. A.; Melnik, A. V.; Meusel, M.; Dorrestein, P. C.; Rousu, J.; Böcker, S. SIRIUS 4: A Rapid Tool for Turning Tandem Mass Spectra into Metabolite Structure Information. *Nat. Methods* **2019**, *16(4)*, 299–302. DOI: 10.1038/s41592-019-0344-8.
4. Dührkop, K.; Shen, H.; Meusel, M.; Rousu, J.; Böcker, S. Searching Molecular Structure Databases with Tandem Mass Spectra Using CSI: FingerID. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112(41)*, 12580–12585. DOI: 10.1073/pnas.1509788112.
5. Kornas, P.; Chadha, M. Quantitation of 764 Pesticide Residues in Tomato by LC/MS According to SANTE 11312/2021 Guidelines. *Agilent Technologies application note*, publication number 5994-5847EN, **2023**.

www.agilent.com

DE- 016665

Printed in the USA, August 7, 2026
5994-9398EN

Agilent
Trusted Answers