Application note | 001605

Environmental

Differential analysis of soil using the Orbitrap Exploris GC 240 mass spectrometer and Compound Discoverer software

Authors

Łukasz Rajski[1], Nicholas Warner[1], Daniel Kutscher[1], and Dominic Roberts[2]

[1]Thermo Fisher Scientific, Bremen, Germany

[2]Thermo Fisher Scientific, Hemel Hempstead, UK

Keywords

Orbitrap Exploris GC 240 mass spectrometer, mass resolving power, sensitivity, mass accuracy, Orbitrap technology, gas chromatography, soil, Compound Discoverer software

Goal

To demonstrate the performance of the Thermo Scientific™ Orbitrap Exploris™ GC 240 mass spectrometer for the differential analysis of soil samples. To show the power of Thermo Scientific™ Compound Discoverer™ software for the processing of GC HRMS data and present statistical differences between soils from different locations.

Introduction

Environmental samples, (e.g., soil, sediments, or surface water) can contain a broad spectrum of volatile or semi-volatile contaminants, including polychlorinated biphenyls (PCBs), polychlorinated dibenzodioxins (PCDDs), polyaromatic hydrocarbons (PAHs), brominated flame retardants (BFRs), and pesticides. The combination of gas chromatography (GC) with quadrupole-based mass spectrometers is a common analytical setup for detection of these contaminants. GC-MS quadrupole-based instruments are well known for their robustness and ease of use. However, they have some important limitations for this application. Single quadrupole instruments can perform full scan analysis, but their sensitivity and selectivity in this mode are limited. Triple quadrupole (GC-MS/MS) systems have greater sensitivity and selectivity; however, their advantages are limited to targeted acquisition within a specified compound list.

thermo scientific

Creation of targeted methods requires time-consuming optimization and use of analytical standards, which can be very expensive and, on occasion, unavailable. However, the biggest drawback is that only targeted compounds included in the method can be detected, whereas other contaminants present will be overlooked.

High-resolution accurate mass (HRAM) mass spectrometry provides a very sensitive and selective non-target acquisition and surpasses quadrupole instruments in all non-targeted applications. Orbitrap™ MS-based instruments offer unmatched resolving power (up to 240 000 at *m/z* 200), mass accuracy greater than 1 ppm, wide dynamic range, and high sensitivity. However, to fully realize the benefits of a HRAM system, powerful software is essential to convert high quality data into scientific discovery.[1] Thermo Scientific™ Compound Discoverer™ software is designed to process large non-targeted data sets acquired using high-resolution mass spectrometry instruments, like the Orbitrap Exploris GC 240 mass spectrometer. The software contains a wide range of tools for unknown compound identification and statistical analysis.

In this study, GC-Orbitrap technology and Compound Discoverer software were used to assess the chemical profile of soil sample extracts taken from three locations near Bremen, Germany. Data were acquired in full-scan with electron ionization (EI) mode. Positive chemical ionization (PCI) and negative chemical ionization (NCI) were used to confirm the elemental composition of the molecular ions using accurate mass information, isotopic match (measured versus theoretical), and presence of specific adducts.

Experimental

Sample preparation

Three soil samples were taken from various locations in and near to Bremen in Germany. The samples received the following letter codes: D, L, and M. They were collected in proximity of a motorway junction, close to an airport, and in a stand-alone house neighborhood, respectively. The samples were extracted without any pretreatment. A 2 g portion of soil was weighed in a polypropylene tube followed by the addition of 4 mL of acetonitrile and vortexed for 5 minutes. Acetonitrile is a water miscible solvent, facilitating the extraction of organic contaminants in humid soils. Next, 4 mL of hexane were added and vortexed again in the same manner. Organic contaminants were transfered to the hexane phase through liquid-liquid partitioning between hexane and acetonitrile solvent layers. Subsequently, the tube was centrifuged for 5 minutes at 4,000 rpm with the hexane layer transferred to a GC vial and injected.

In the EI mode, the extracts were analyzed in triplicate in a random order. After every sixth injection, a pooled sample was injected. The pooled sample contained equal volume of the three soil extracts. The confirmatory injections in the PCI and NCI modes were performed without repetitions.

Acquisition method

The samples were analyzed with an Orbitrap Exploris GC 240 mass spectrometer coupled to a Thermo Scientific™ TRACE™ 1610 gas chromatograph and a Thermo Scientific™ TriPlus™ RSH SMART autosampler. All the instrumental parameters are shown in Tables 1–4.

Table 1. Parameters of the TRACE 1610 GC

TRACE 1610 GC	
Injector	
Injection volume (µL)	1
Liner	Single gooseneck with glass wool Thermo Scientific™ LinerGOLD™ (P/N 453A1925-UI)
Inlet temperature (˚C)	300
Inlet module and mode	SSL, Splitless
Splitless time (min)	1
Septum purge flow (mL/min)	5
Oven and column	
Carrier gas, flow rate (mL/min)	He, 1.2
Column	Thermo Scientific™ TraceGOLD™ TG-5SilMS 30 m × 0.25 mm i.d. × 0.25 µm (P/N 26096-1420)
Oven temperature program	
Temperature 1 (˚C)	40
Hold time (min)	2
Temperature 2 (˚C)	300
Rate (˚C/min)	10
Hold time (min)	7
Total GC run time (min)	35

Table 2. Parameters of the Orbitrap Exploris GC 240 mass spectrometer in electron ionization mode

Orbitrap Exploris GC 240 mass spectrometer in EI mode	
Transfer line (˚C)	300
Ion source (ionization type)	Thermo Scientific™ ExtractaBrite™ (EI) source
Ion source (˚C)	280
Electron energy (eV)	70
Emission current (µA)	50
Acquisition mode	Full scan (FS)
Mass range (m/z)	50–550
Resolving power	120,000
AGC target	Standard
Maximum injection time	Auto
Lock masses	133.01356; 207.03235; 225.04292; 281.05114; 299.06171; 355.06993

Table 3. Parameters of the Orbitrap Exploris GC 240 mass spectrometer in positive chemical ionization mode

Orbitrap Exploris GC 240 mass spectrometer in PCI mode	
Transfer line (˚C)	300
Ion source (ionization type)	ExtractaBrite (CI)
Ion source (˚C)	250
CI gas	Methane
CI gas flow (mL/min)	1.0
Acquisition mode	Full scan/ddMS^2
Mass range (*m/z*)	50–550
Resolving power	120,000
AGC target	Standard
Maximum injection time	Auto
ddMS^2 Scans	5
ddMS^2 Filters	Dynamic exclusion, Apex detection
ddMS^2 Isolation window (*m/z*)	1.2
ddMS^2 HCD collision energies (V)	20; 40; 60
ddMS^2 Resolving power	15,000
ddMS^2 Scan range	Auto
ddMS^2 AGC target	Standard
ddMS^2 Maximum injection time	Auto

Table 4. Parameters of the Orbitrap Exploris GC 240 mass spectrometer in negative chemical ionization mode

Orbitrap Exploris GC 240 mass spectrometer in NCI mode	
Transfer line (˚C)	300
Ion source (ionization type)	ExtractaBrite (CI)
Ion source (˚C)	250
CI gas	Methane
CI gas flow (mL/min)	1
Acquisition mode	Full scan
Mass range (*m/z*)	50–550
Resolving power	120,000
AGC target	Standard
Maximum injection time	Auto

Compound Discoverer software workflows

Compound Discoverer software contains template workflows for GC EI, as well as GC PCI data. In this study, the EI data were used for statistical analysis and compound identification, whereas the PCI data were used for the confirmation purposes. PCI is an alternative and complimentary form of ionization, which is considered a softer ionization that often gives molecular ion information through mass adduct patterns and lower fragmentation. It is often seen as an important option when an unknown compound is suspected. Although not mandatory, combination of EI and PCI data is advisable as it increases the confidence of identification due to molecular ion confirmation.

The main features of the EI workflows in Compound Discoverer software are spectral deconvolution, compound identification, and statistical analysis. Compound identification is based on the library search using both high-resolution and nominal mass spectral libraries. The EI workflow (Figure 1) used here was one of the default workflows available in the software. The workflow contained spectral deconvolution and statistical data evaluation (Descriptive Statistics and Differential Analysis visible in the lower part of Figure 1). The template workflows in Compound Discoverer software come with optimized parameters. Thus, the user only needs to select the spectral libraries to be used for identification. The deconvoluted spectra were identified against the NIST™ 2020 (nominal mass) and GC-Orbitrap Contaminants Library (HRAM library).

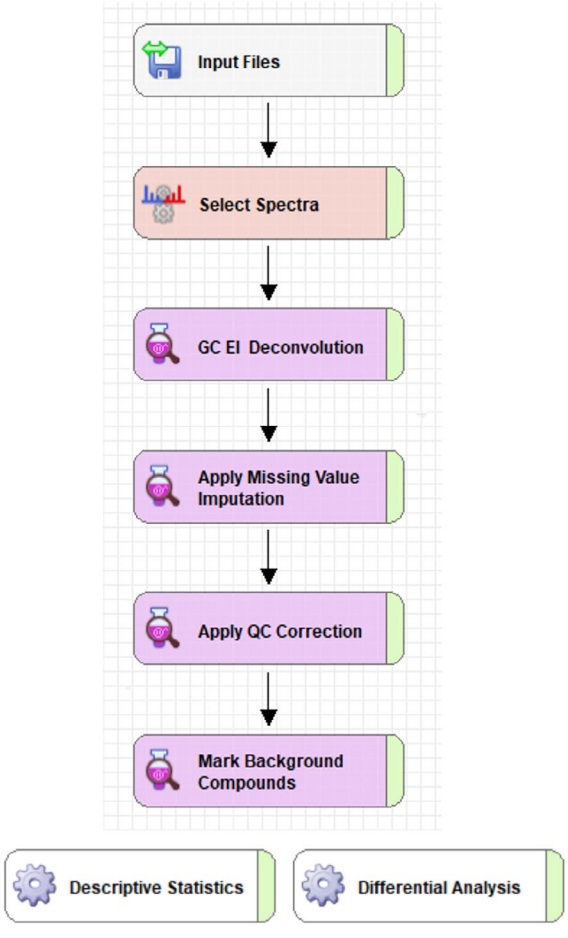

Figure 1. Electron ionization (EI) workflow used in Compound Discoverer software. The node-based structure enables a flexible approach to research data processing.

For various statistical analyses, zero values within the sample set can lead to erroneous results. To avoid this type of error, Compound Discoverer software provides methods for imputing missing chromatographic peak areas for detected compounds across the set of input files. This is the role of the "Missing Value Imputation" node. Additionally, two extra nodes—"Apply QC Correction" and "Mark Background Compounds"—were aggregated to the default nodes present in the workflow. The "Apply QC Correction" node is useful when a long sequence of samples is acquired and compensates for time-dependent batch effects. To use this node, a QC sample is required. To create the QC sample, a small aliquot from each sample must be pooled in one vial. The pooled sample is injected at regular intervals along the sequence, in this case, after every six injections. The "Mark Background Compounds" node is applied to flag compounds that are found not only in the sample but also in the instrumental or matrix blanks. A compound that has

$$\frac{\textit{peak area in sample}}{\textit{peak area in blank}}$$

below a desired threshold (5 by default) is marked as a background compound and can be hidden in the results table.

Positive chemical ionization

Figure 2 shows the workflow applied for the PCI data processing. The PCI workflows are strongly related to the presence of molecular ion. In this study, the following identification nodes were applied:

- **Predict Composition:** predicts the chemical formulas of the unknown compounds
- **Search ChemSpider™:** enables search in ChemSpider of elemental composition proposed
- **Search mzCloud™:** performs a search in the mzCloud library, which is an exact mass library that contains both MS as well as MS^2 data
- **Search Mass List:** serves to a databases search (this node is also available for EI workflows)

The task of the "Assign Compound Annotations GC CI" node is to assign and prioritize the annotation coming from the nodes (Predict Composition, Search ChemSpider, Search mzCloud, and Search Mass List). In the software, there is also a node that enables search in Thermo Scientific™ mzVault™ libraries, however in this study it was not employed.

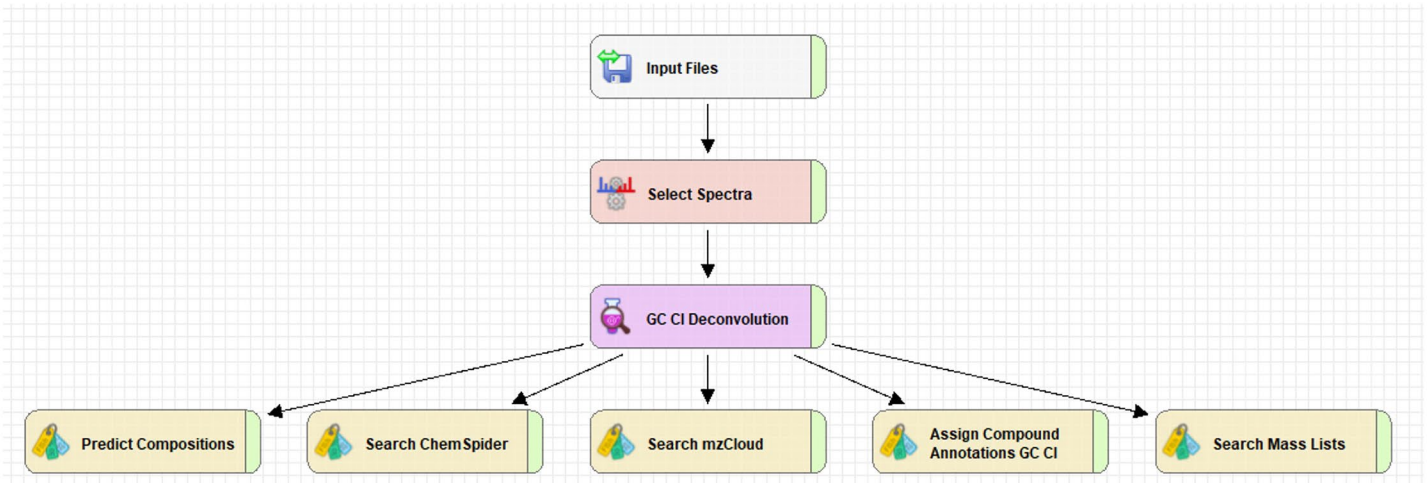

Figure 2. Positive chemical ionization (PCI) workflow used in Compound Discoverer software. Detection of the molecular ion is the focus in this workflow to confirm identification of compounds or to propose an elemental composition of an unknown.

Data evaluation - statistical tools and unknown compound identification

The first objective was to identify if there was any significant difference between the three soil samples at locations D, L, and M. This was achieved through a PCA plot of the replicate injections of each sample. Figure 3 shows the PCA plot that demonstrates that there are clear differences between the samples and good agreement of the replicate injections. The following steps then help identify which peaks contribute to the differences so that a compound identification can be proposed. As an example, Figure 4 shows a volcano plot for the samples D and L. The volcano plot is a type of scatter plot for replicate data where the x axis represents the log2 of the fold change between two sample groups (generated ratio), and the y axis represents the negative log10 of the p-value (test of significance) of the fold change. In other words, when a point (compound) is more on the right (positive values on x axis), the peak area of that compound is much higher in the sample D than in the sample L. Whereas, points that are higher on the graph are statistically more significant.

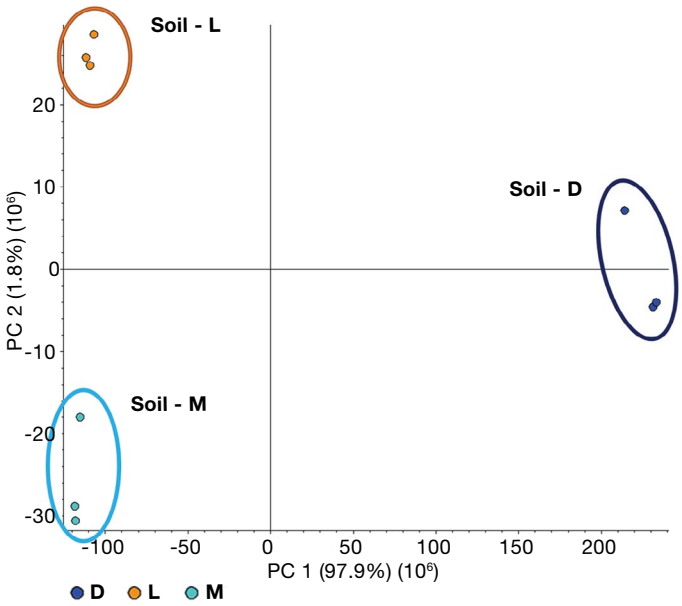

Figure 3. PCA score plot of the volatile compounds that differentiate the three soil samples from different locations. A complete separation between the sample groups was observed and good agreement between replicate injections.

Figure 4. V-plot scatterplot showing the statistical significance (P value) versus magnitude of change (fold change) when comparing the soil sample D (right) versus the soil sample L (left). The main chemical components that are responsible for sample diversity between two sample groups are located in the upper right and left sides of the plot.

In Figure 4, it can be easily recognized that sample D contained more characteristic compounds than sample L, possibly due to a higher content of the organic fraction. Double-clicking on any of the points (features) takes the user to the results table, where the compound details can be checked (Figure 5). The table contains library search results (name, formula, search index, reverse search index, etc.) as well as information about peak areas. Moreover, each of the rows contain related tables where the user can check the details of NIST identification.

In the discussed samples, the majority of detected compounds were related with the presence of the soil organic matter. Nevertheless, a deeper insight revealed some typical contaminants. Pyrene is an interesting example. Figures 6a and 6b show pyrene identification details. This compound was present in all three analyzed soils; however, its signal was the highest in the sample M. In the EI workflow, pyrene was identified, achieving a search index of 897 and reversed search index of 932. Pyrene and other polyaromatic hydrocarbons produce a stable and intense molecular ion, which is not very common in EI. The molecular ion presence is mentioned in the results table. In Figure 6a, a comparison between the deconvoluted spectrum and spectrum present in the library can be seen.

The identification of pyrene was confirmed in the CI mode, where all the annotation sources suggested its presence. As discussed, the CI identification is based on the presence of the molecular ion. In contrary to the electron ionization, the chemical ionization is a softer ionization process, allowing for easy identification of the molecular ion due to reduced fragmentation. Information on the molecular ion allowed the Compound Discoverer software to predict the molecular formula for our unknown compound (Figure 6b). That prediction was verified by the evaluation of the isotopic profile and search for alternative adducts (Figure 6b). The obtained molecular formula ($C_{16}H_{10}$) was assigned to pyrene by the following nodes: "ChemSpider Search", "MassList Search", and "mzCloud Search". During the data acquisition, full scan MS was combined with a data dependent MS^2 ($ddMS^2$). This means that after each full scan MS, there was a series a MS^2 scans. The five most abundant ions from each MS spectrum were fragmented, each in a separate fragmentation event. In this manner, high quality MS^2 spectra were obtained. Thus, the "mzCloud Search" involved the MS^2 data as well.

Pyrene was not the only polyaromatic hydrocarbon detected. The samples also contained fluoranthene and perylene. The peak areas of these compounds were the greatest in the sample M. That sample was collected in a stand-alone house neighborhood, and many of the houses situated there have a fireplace. This could be the reason of the elevated amount of PAHs.

In addition to the Volcano Plot, Compound Discoverer software offers other tools for data visualization. Box Whisker Charts can be used to easily visualize the pyrene peak areas in the investigated samples. As mentioned above, the peak areas in the sample M were higher than in the samples L and D. This is clearly visible in Figure 7. Moreover, the figure reveals the relations between the samples D, M, and L were the same in EI and CI. This is a strong suggestion that the peaks belong to the same compound.

Another interesting contaminant discovered in the investigated samples was a PCB containing six chlorine atoms. Sample D showed the highest levels of contamination; however, the analyte was also present in the two remaining soil samples. Thanks to the presence of six chlorine atoms, the isotopic pattern is very characteristic and can aid in the confirmation of the molecular formula. Figure 8a shows a comparison between the experimental spectrum and the theoretical isotopic pattern of an ion with elemental composition corresponding to a hexachlorinated PCB. As can be seen, the match is nearly perfect. Unfortunately, polychlorinated biphenyls do not ionize easily in positive chemical ionization mode. However, PCBs with six or more chlorines show excellent response with negative chemical ionization. Presence of an ion corresponding to $C_{12}H_4Cl_6$ was also confirmed in the NCI data. Figure 8b shows the acquired spectrum and theoretical isotopic pattern of $C_{12}H_4Cl_6$ in the negative ionization mode. To specify which PCB congener was present in the sample, the knowledge about the retention index would be very helpful.

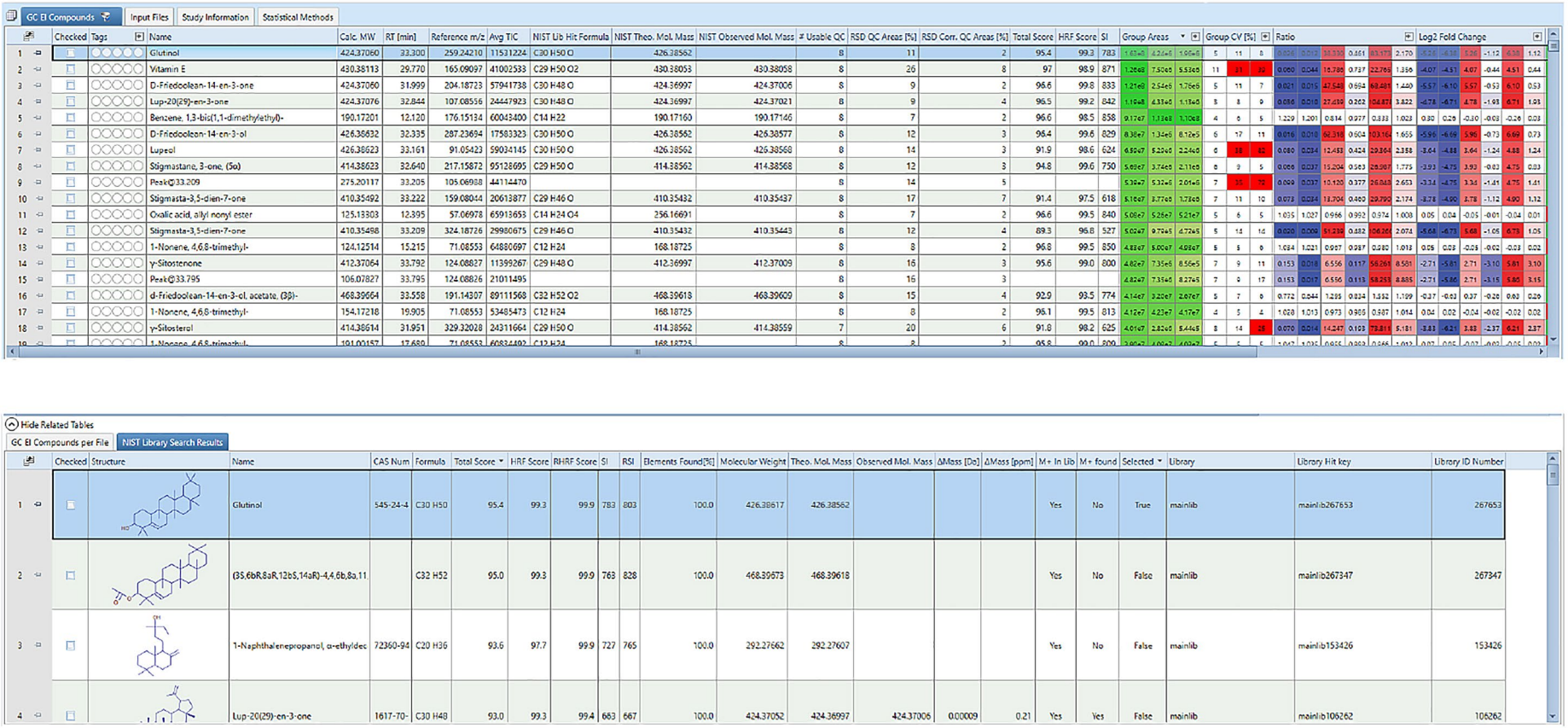

Figure 5. Results table in Compound Discoverer software showing a list of detected compounds with information on identification criteria

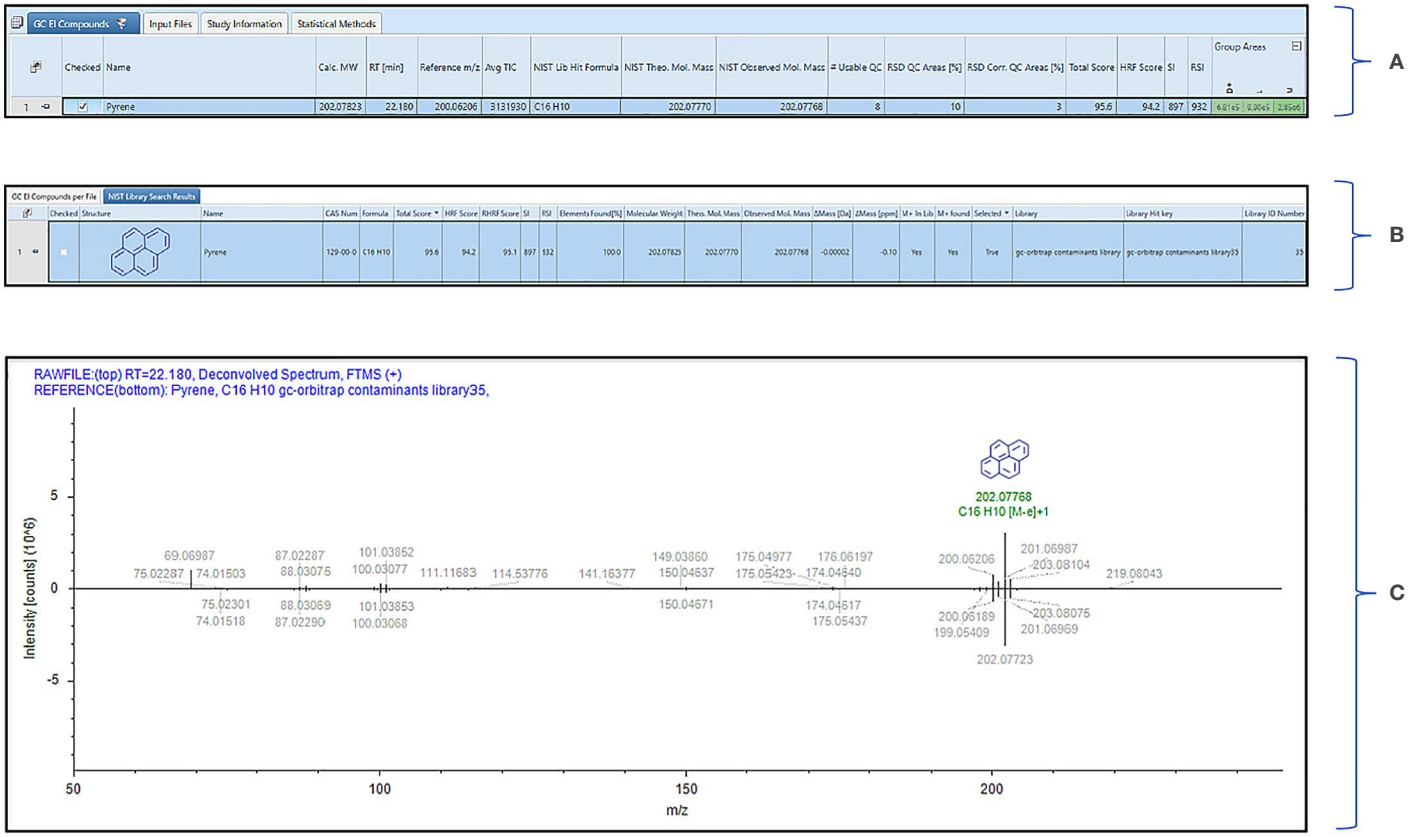

Figure 6a. Identification of the peak at 22.18 minutes as pyrene. Table data (A) showing search index (SI) at 897 and reverse SI (932). The compound molecular mass was detected with excellent mass accuracy of 0.1 ppm (B). The spectrum mirror plot (C) comparison to the library is displayed in Compound Discoverer software to support the proposed identification.

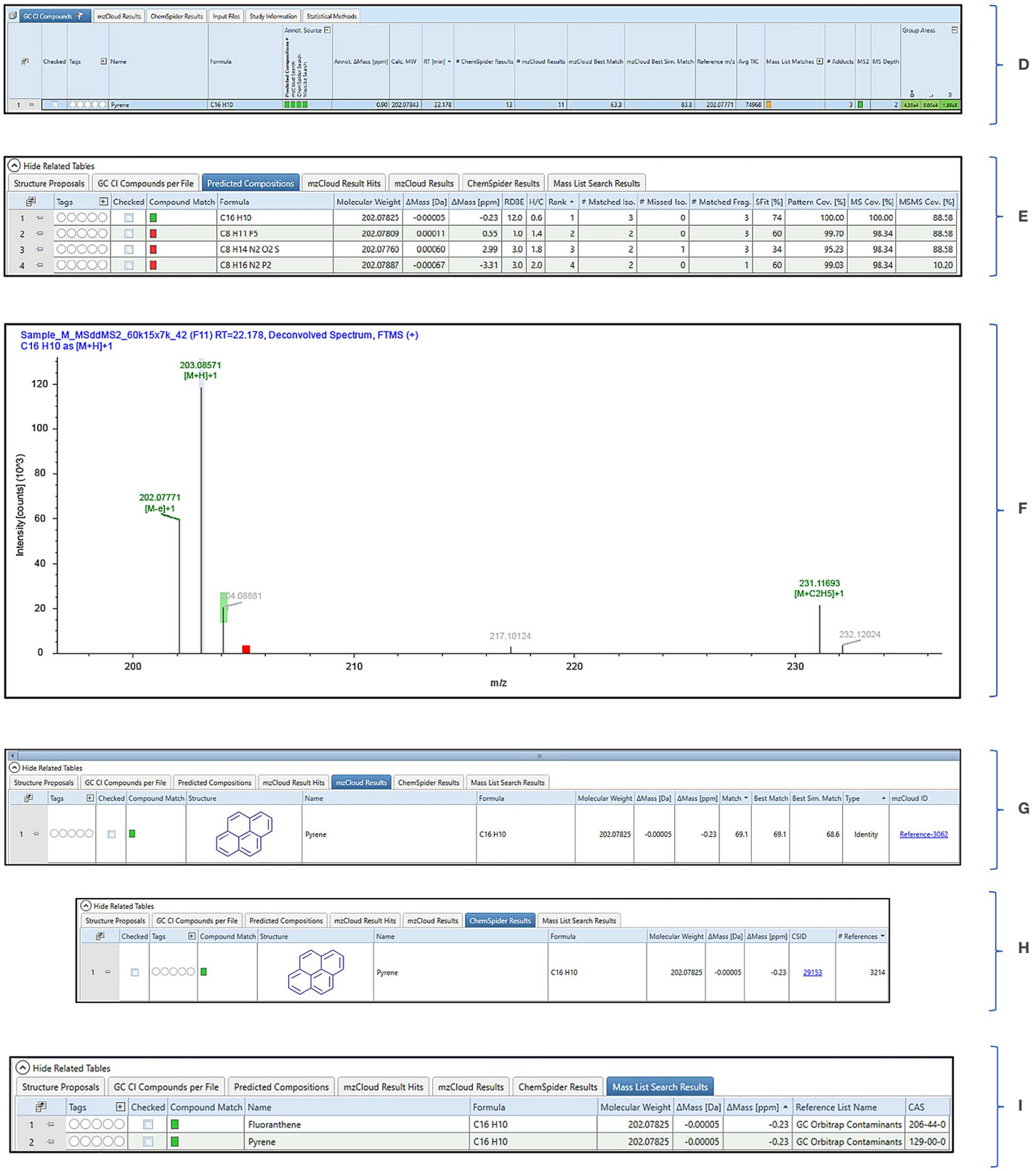

Figure 6b. Identification and confirmation of pyrene through positive chemical ionization (PCI) section (D). Elemental composition proposal for molecular ion 0.23 ppm mass accuracy (E) with annotated spectrum (F). Library search results for the elemental composition proposed in mzCloud (G) and Chemspider (H). Mass List (I). All results combine to provide high confidence in compound identification.

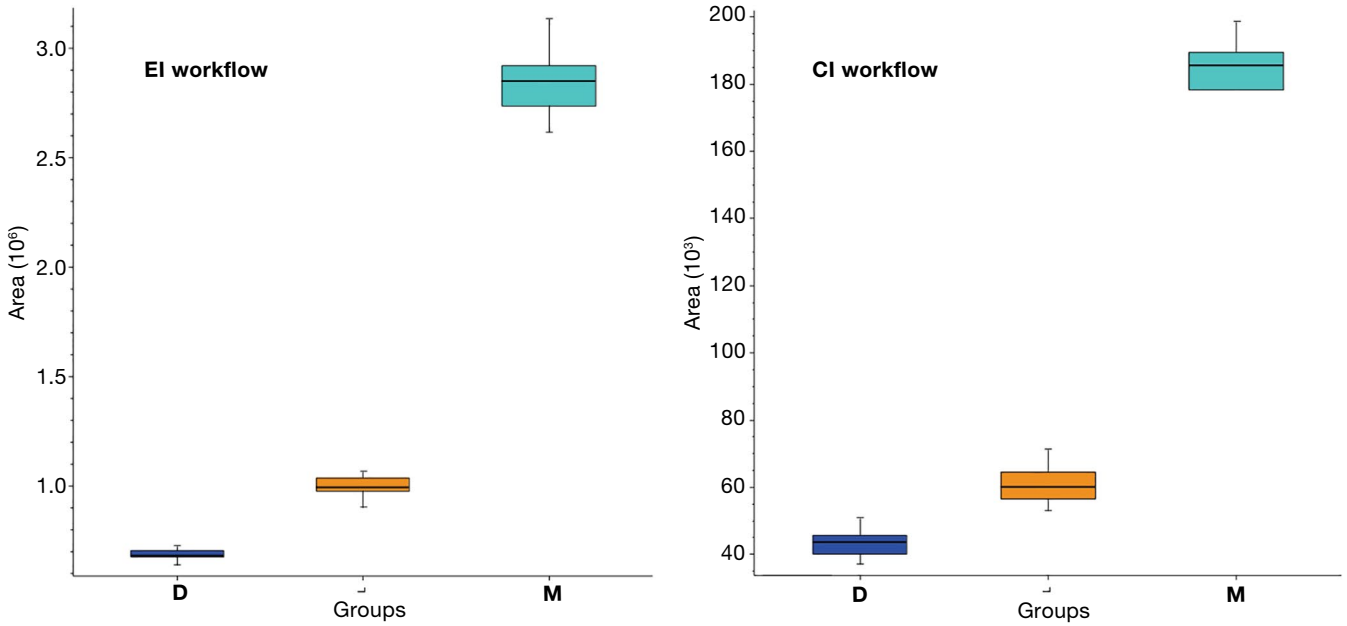

Figure 7. Box plot view from Compound Discoverer software showing how pyrene peak area varies across the three soil samples. The response was significantly higher in sample M.

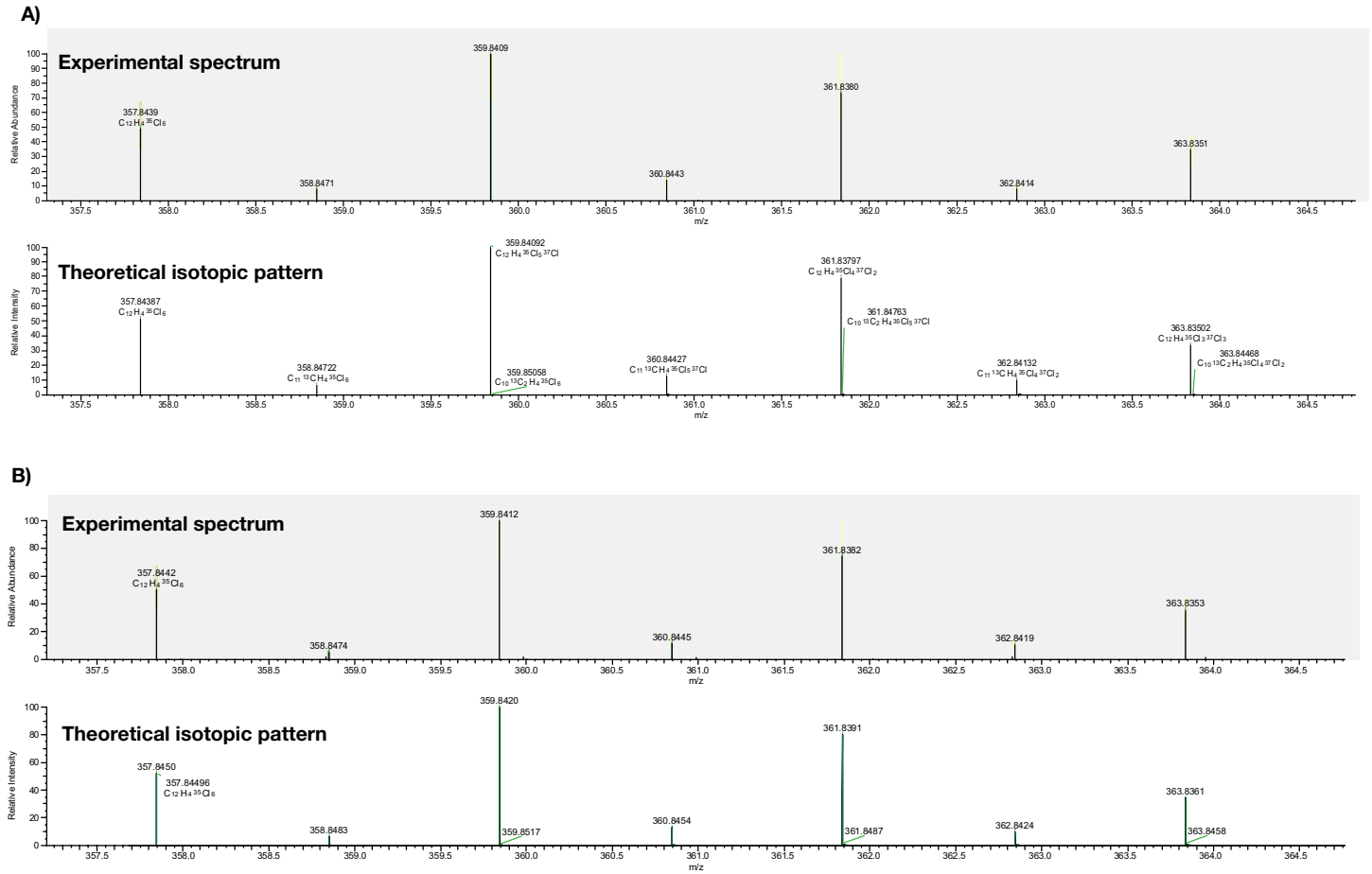

Figure 8. Comparison between experimental and theoretical isotopic pattern for $C_{12}H_4Cl_6$ in EI (A) and in NCI (B)

Conclusion

The Orbitrap Exploris GC 240 mass spectrometer in combination with Compound Discoverer software is an excellent tool for environmental sample analysis. Even in such a complex matrix as soil, the software detected and identified untargeted contaminants. The identification in EI was confirmed in CI. Moreover, the statistical analysis and graphical visualization tools facilitated the results interpetation. Differencial analysis (Volcano Plot) was useful in the global comparison between two selected samples, whereas Whisker Charts allowed presentation of a particular compound peak area across all samples. The high quality data obtained from the Orbitrap Exploris GC 240 mass spectrometer enabled the sensitive detection and confident identification of compounds in this study.

Reference

1. Thermo Fisher Scientific Technical Note 10730: Mass resolving power of 240,000: for confident compound identification, 2021. https://assets.thermofisher.com/TFS-Assets/CMD/Technical-Notes/tn-10730-gc-ms-power-confident-compound-identification-tn10730-en.pdf

Learn more at **thermofisher.com/OrbitrapExplorisGC240**

AN001605-EN 1222S