Combining Chromatographic Deconvolution, Spectral Library Search, and Principal Component Analysis to Detect and Identify Important Flavor and Fragrance Components with High Resolution GC/MS

Xin Zheng[1], John Voss[1], Jason Cole[1], [1]Thermo Fisher Scientific, Austin

ABSTRACT

Chromatographic deconvolution has frequently been used in untargeted gas chromatography/mass spectrometry (GC/MS) software to group individual mass peaks together into compound spectra. Following deconvolution, identification is typically accomplished by searching each spectrum against a spectral library database. With the growing popularity of GC accurate mass instrumentation for identification of unknown compounds in various sample types, there has been growing interest to couple these traditional compound identification tools to statistical methods, especially for omics analysis, that can be used to distinguish sample types and highlight marker compounds by fold changes. In this study, a new prototype node named "GC Deconvolution" has been integrated into Thermo Scientific™ Compound Discoverer™ (CD) to fully process high resolution accurate mass (HRAM) GC/MS electron ionization (EI) data with statistical algorithms. Lavender oils were selected as an example and applied for this classical untargeted GC/MS workflow to demonstrate the utilities of this new prototype node.

INTRODUCTION

Deconvolution is a computational process of separating coeluted components and creating clean spectra for each component. This technique has been widely used for GC/MS data processing and is especially useful for untargeted GC/MS analysis where sample matrices are high. GC/MS deconvolution can resolve compounds that prove challenging to completely chromatographically resolved, or eliminate matrix interference ions that improves library searching results, further increasing the number of compounds being identified and thus increasing confidence in the identification of unknown compounds. The deconvolution process includes peak detecting all the features through the entire chromatogram, binning the features that share the same retention time and peak shape together, and then constructing the spectrum of a compound based on this collection.

This is the first integration of GC deconvolution algorithms into Compound Discoverer software, which is a comprehensive mass spectrometry analysis platform for small molecule analysis including identification, relative quantification, and statistical analysis. This new prototype node is dedicated to processing HRAM GC/MS data including deconvolution, peak grouping, gap filling, and library searching. Statistical analysis including differential analysis, principal component analysis (PCA), and partial least squares-discriminant (PLS-DA) are also performed after GC deconvolution.

MATERIALS AND METHODS

Sample Preparation

All the lavender oils were diluted 100/1 by hexane before being injected on a Thermo Scientific™ Q Exactive™ GC. All the solvents were purchased from Sigma-Aldrich.

Instrument Setup and Data Acquisition

All the samples were analyzed on a Thermo Scientific™ Q Exactive™ GC Orbitrap™ GC-MS/MS system. The GC/MS parameters are listed in Table 1.

Data Analysis

Data was acquired using Thermo Scientific™ TraceFinder™ 4.1 software and processed through Compound Discoverer. This includes peak integration and calculation of fold changes of compounds across the sample batch, as well as data review and reporting. Library searching was performed by either the NIST 17 library or an HRAM GC-Orbitrap library.

Compound Discoverer Workflow

The GC deconvolution workflow is shown in Figure 1 including "Input Files", "Select Spectra," and "GC Deconvolution" nodes. Post-processing nodes are "Descriptive Statistics" and "Differential Analysis," which performs statistical analysis after the "GC Deconvolution" node.

The "GC Deconvolution" node parameters are listed in Figure 2. These include peak detection criteria such as setting mass tolerance, signal/on-noise threshold, TIC threshold, and ion overlap window. In retention indexing (RI) settings, it allows importing of n-alkane reference retention times for the retention indexing calculation. Group compounds settings group all the detected peaks across the batch based on the retention time window and dot product threshold. Composition threshold helps to eliminate certain compound groups that are either formed by spurious peaks or formed as a separated group that should belong to a larger group due to its retention time shift beyond the grouping retention window. Gap fill is performed after grouping using a reference mass for each group to fill in low concentration compounds that were not detected during the initial deconvolution step. Library search is also included in this node in which both high resolution accurate mass and unit resolution libraries can be searched against with observed spectrum. Library search indexing score (SI) and high resolution filtering score (HRF) are calculated as well.

RESULTS

Lavender oil is the most commonly used essential oil for food, pharmaceutical, therapeutic, fragrance, and beauty products. There are several varieties and species of lavender as essential oils. The two main lavender plant types are lavender and lavandin, which have similar aromas but a vastly different chemical composition. Since the essential oil industry is still young, there are many novel methods that essential oil brands use to adulterate oils, especially for lavender oils. One major difference between lavender and lavandin is camphor percentage. Lavandin oil contains much higher camphor (6-10%), whereas real lavender oil contains much less (0-0.6%). That's why lavandin oils have much stronger and more pungent scents, whereas lavender oils are more sweet and floral. One of the best ways to check purity and chemical composition of lavender oils from sample to sample and batch to batch is to analyze them on a GC/MS and do aroma profiling with statistical analysis. In this study, four subtypes of lavender oil and two subtypes of lavandin oil were selected and analyzed. Three replicate injections were made for each subtype oil. The data files were processed through Compound Discoverer using GC Deconvolution workflow (Figure 1). Statistical analysis was performed based on study factors created with the workflow.

The data processing results are shown in the default layout in Figure 3. On the upper left is a chromatographic peak overlaid on the current selected compound throughout the entire batch; the upper right panel shows the selected peak spectra including raw and deconvoluted spectra. If the molecular ion of the selected compound was found, the molecular ion and its isotopes are zoomed and highlighted in green (Figure 4). In the deconvoluted spectrum, the entire spectrum is displayed with the molecular ion cluster highlighted in green (Figure 5). The table in Figure 3 lists all the grouped deconvoluted compounds with formula, retention time (RT), reference mass, calculated RI, molecular mass, SI, HRF, total score, grouped peak area, fold change, group CV%, and P values. In the "Library search results" tab, it shows all the potential library hits for the current selected compound (Figure 6). More library searching information can be found in this tab such as RI column type, library name, and library ID number. A mirror plot (Figure 7) will be displayed on the mass spectrum area on the upper right of the layout when a library hit is selected in the "Library search results" tab. The deconvoluted spectrum is on the top, and the library hit spectrum is at the bottom of the mirror plot.

Different statistical results and plots can also be generated and reviewed in the same layout such as PCA plot, Box Whisker chart, S plot, volcano plot, and so on. Aroma profiling of different lavender types was used to detect key aroma compounds in discriminating flavor especially for adulteration tests. In the PCA plot (Figure 8), it shows clear differentiation between lavender and lavandin oils. Bulgarian and Population lavender are all grown at higher altitude, thus contain more esters, so they have similar aroma compounds (under the blue cycle). Lavender 40-42 is blended of different natural lavender oils to produce an oil containing 40% linalool and 42% linalyl acetate, which are the two primary aromas occurring in lavender. The purpose is to create a standardized lavender aroma from lot to lot and season to season. That is the reason lavender 40-42 is commonly used as an affordable lavender oil in the market. English lavender is one of the oils that were blended in lavender 40-42, so these two oils can be grouped together and share similar aromas (under the yellow cycle).

A volcano plot is also generated to target aroma markers that have high negative log P values and fold changes. One example is the spot under the red circle in Figure 8 identified as camphor, which is the main aroma distinguishing lavandin from lavender. The box whisker chart is also generated around camphor as shown in Figure 9. Another key aroma in true high altitude lavender oil is linalyl acetate, which is an essential ester in lavender oil. In the box whisker chart, a clear trend demonstrates the difference between both Bulgarian and Population and other oils.

Table 1. Gas chromatograph and mass spectrometer analytical parameters.

Thermo Scientific™ TRACE™ 1310 GC Parameters		Q Exactive GC MS Parameters	
Injection Volume (µL):	1.0	Transfer line (°C):	250
Column	TG-5SilMS 30mx 0.25mmx0.25um	Ionization type:	EI
Liner	Single taper without glass wool	Ion source(°C):	250
Inlet (°C):	250	Electron energy (eV):	70
Inlet Module and Mode:	SSL, Split 20:1	Acquisition Mode:	Full scan
Carrier Gas, (mL/min)	He, 1.2	Mass resolution (FWHM):	60,000
Oven Temperature Program:		Mass range (*m/z*):	30-500
Temperature 1 (°C):	40	Lock masses (*m/z*):	73.04680; 133.01356; 207.03235; 281.05114; 355.06990
Hold Time (min):	1		
Temperature 2 (°C):	280		
Rate (°C/min)	10		
Hold Time (min):	2		

Figure 1. GC Deconvolution workflow on Compound Discoverer 3.1

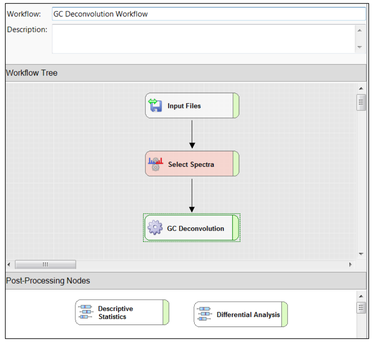


Figure 2 . Parameters of GC Deconvolution node.

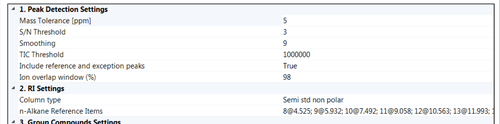


Figure 4. Raw spectrum of a selected compound with molecular ion cluster highlighted in green

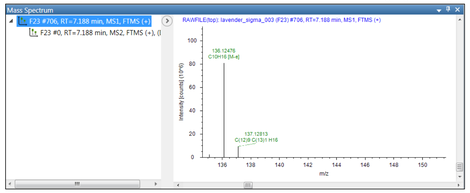


Figure 5. Deconvoluted spectrum that shows all the fragment ions with their formula and molecular ion cluster highlighted in green.

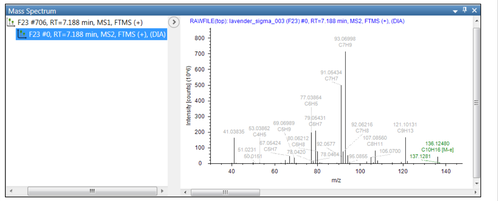


Figure 6. "Library search results" tab shows all the potential library hits of the current selected compound in the "Grouped deconvolved compound" tab. In this tab, it lists library search scores, HRF scores, RI information, library name, and library ID number.

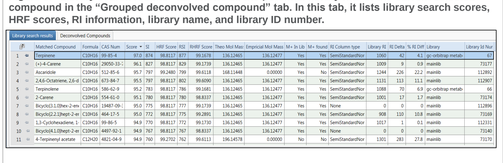


Figure 7. Mirror plot shows the comparison between library spectrum and deconvoluted spectrum of the current selected library hit with the molecular ion highlighted in green.

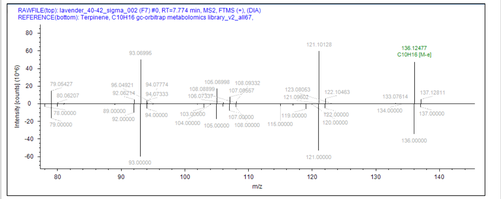


Figure 8. PCA plot (on the left) showing the excellent separation among lavender oil (in black circle) and lavandin oil (under the red circle). In lavender group, Bulgarian and Population lavender oils have more similarity whereas English and 40-42 shared similar aromas. Volcano plot shows the significant compounds for the comparison of lavender and lavandin. The red circle is identified as camphor and exhibits a high fold change in the samples compared.

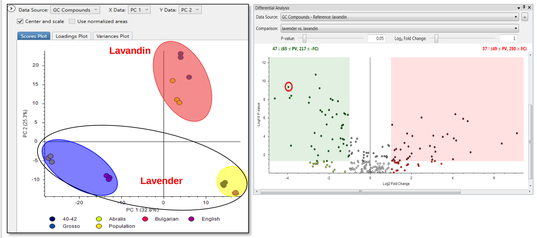
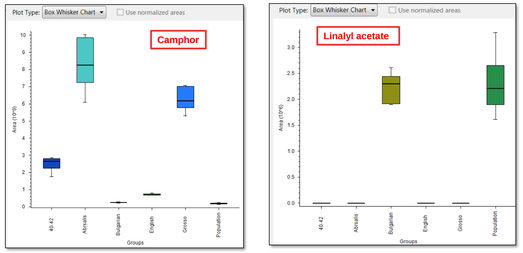


Figure 9. Box whisker chart shows the significant trend difference on key aroma compounds. Camphor is a key aroma marker to differentiate lavandin and lavender oils. Linalyl acetate is an essential ester in high altitude Bulgarian and Population lavender oils.



CONCLUSIONS

This is the first instance of integrating a GC deconvolution algorithm into Compound Discoverer software to allow it to fully process GC/MS data with statistical analysis. This comprehensive GC Deconvolution node can do peak detection, deconvolution, grouping, gap filling, and library searching. Statistical analysis can also be performed throughout the whole workflow. This new prototype node will dramatically help users who are focused on omics analyses such as global metabolomics, foodomics, aroma profiling, and adulteration analysis.

TRADEMARKS/LICENSING

PO65524-EN0519S