Olive Oil Characterization using Agilent GC/Q-TOF MS and Mass Profiler Professional Software

Application Note

Food Testing & Agriculture

Authors

Stephan Baumann,
Sofia Aronova
Agilent Technologies, Inc.
Santa Clara, CA USA

Abstract

A model was constructed that predicts whether an olive oil will pass the extra virgin sensory test. Using the Agilent 7890A GC system coupled to the Agilent 7200 series accurate-mass Q-TOF MS in both electron ionization (EI) and positive chemical ionization (PCI) modes, a large number of compounds was found in olive oil. Mass Profiler Professional software was used to perform statistical analysis and construct a classification model that uses the presence of five specific compounds to accurately predict whether an olive oil would fail the sensory test.

Agilent Technologies

Introduction

The demand for olive oil is growing rapidly in the United States, driven by an increased interest in Mediterranean foods as well as the health benefits associated with olive oil. The US market is expected to surpass $1.8 billion by 2013 [1]. Olive oil is thought to be responsible for the longevity of southern European populations and their low rates of heart disease. In fact, the Food and Drug Administration (FDA) approved a health claim for monounsaturated fat from olive oil and reduced risk of coronary heart disease. Recent studies have attributed the anti-inflammatory benefits of olive oil primarily to the extra virgin olive oil (EVOO) obtained from the first pressing of the oil, because few foods are as rich in antioxidants and anti-inflammatory compounds.

The International Olive Council (IOC) and USDA have established standards for the classification of EVOO, including a sensory test conducted by a tasting panel and chemical tests. However, recent studies [2] have stated that imported olive oils, which account for 99% of the EVOO on the US market, often fail the sensory test for EVOO classification. Moreover, the sensory tests are expensive and subjective.

Given the growing demand and the size of the EVOO market in the US, there is value in developing a chemical screen that could predict whether an olive oil would pass the sensory test. This would allow producers to submit only those olive oils for sensory testing that have a high probability of passing. Such a chemical screen could also reduce certification costs and time, while increasing the quality of the EVOO available in the marketplace.

This application note demonstrates the feasibility of developing a model that can predict whether an olive oil will pass the sensory test. It uses a nontargeted compound analysis approach similar to that recently used for wine classification [3]. The data were obtained in both electron ionization (EI) and positive chemical ionization (PCI) modes, using the Agilent 7890A GC system coupled to the Agilent 7200 series accurate-mass Q-TOF MS. Chromatographic deconvolution was performed using Agilent MassHunter software, while further statistical analysis and construction of the classification model were accomplished with Mass Profiler Professional (MPP). The accumulation of five specific compounds in an olive oil sample correlated with a failed sensory test.

Experimental

Reagents and Standards

Cyclohexane, spectrophotometric grade, Sigma-Aldrich.

Samples

In total, 10 olive oil samples were obtained from the UC Davis Olive Center. All of these samples had been subjected to IOC sensory test using a panel sanctioned by the IOC to determine if they passed or failed the criteria for EVOO. They were stored in the dark at room temperature. The samples were diluted 1:10 in cyclohexane, injected into the GC with a 1:10 split, and analyzed in random order.

Instruments

This study was performed on an Agilent 7890A GC system coupled to an Agilent 7200 series GC/Q-TOF system. The instrument conditions are listed in Table 1.

Table 1. GC and MS Conditions

GC run conditions	
Column	DB-5 MS, 30 meter, 0.25 mm id, 0.25 µm film (p/n 122-5532)
Injection volume	1 µL
MMI Injector	50 °C for 0.01 minute 600 °C/min to 300 °C
Purge to split vent	60 mL/min at 1 minute
Oven temperature program	45 °C for 4.25 minutes 5 °C/min to 75 °C, 0 minute hold 10 °C/min to 320 °C, 10 minute hold
Carrier gas	Helium at 1.3 mL/min constant flow
Transfer line temperature	290 °C
MS conditions	
Ionization mode	EI, positive CI (20% methane flow)
Source temperature	230 °C
Quadrupole temperature	150 °C
m/z range	40 to 800 *m/z*
Spectra acquisition rate	5 Hz, collecting both in centroid and profile modes

Data Processing and Statistical Analysis

MassHunter Qualitative Analysis (version B.05 SP1) was used for data processing. Peaks were found by the chromatographic deconvolution tool in MassHunter. The mass and compound filters were adjusted so that one to three hundred components were identified by the software. In order to import deconvoluted data into MPP, CEF files were generated using the MassHunter export tool.

Mass Profiler Professional (version 2.1.5) was used for statistical analysis. The data processing steps were as follows:

1. Setting the importation filters and alignment parameters
2. Selecting normalization criteria
3. Defining the sample groups
4. Setting data filters
5. Evaluating data clustering with a PCA plot

Once these steps were completed, the data were evaluated through statistical tools such as fold-change and significance analysis. The final analysis steps were to construct and test a classification model. Further data processing included identification of the compounds used in the model by mass spectral library search and assignment of molecular formula estimation.

Results and Discussion

Data Acquisition and Processing

The analysis of EVOO samples was performed to survey the compounds that could be detected by GC/Q-TOF. Typically, approximately 150 peaks were identified in MassHunter by chromatographic deconvolution (Figure 1). A cold split injection was used, and the inlet temperature was ramped from 50 to 300 °C to minimize thermal decomposition. The initial oven temperature ramp was 5 °C per minute in order to better separate the early eluting peaks.

Peak detection was performed in MassHunter using deconvolution. The MassHunter export tool was then used to generate CEF files that could be imported into Mass Profiler Professional (MPP) software.

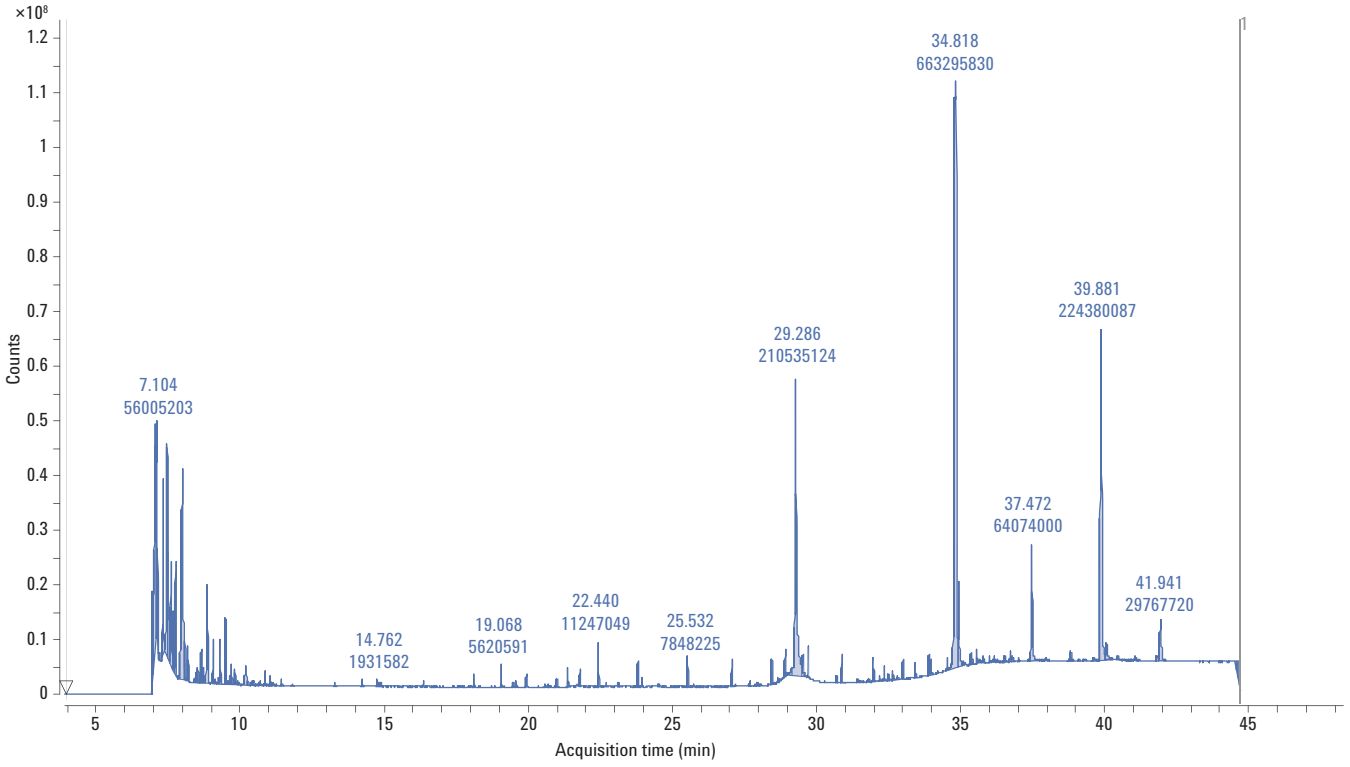

Figure 1. Typically, approximately 150 peaks are identified by chromatographic deconvolution with a relative area filter of 0.1% of largest peak.

The Extracted Ion Chromatograms (EICs) were aligned in MPP. The alignment was based on spectral pattern and retention time. The extracted spectra needed to have a cross correlation factor of 0.6 and a retention time match of 0.05 minutes to be considered the same component. The mass versus retention time plot in Figure 2 indicates that 442 unique entities were identified in the olive oil samples. Most of them occur only once or twice and were removed during the data filtering step.

Once the sample groups were defined, data filters were set in MPP. Entity filtering permits the creation of a higher quality data set, so that subsequent multivariate analysis is more meaningful. The first filter determined which entities (compounds) were in at least one group 100% of the time (frequency analysis). This frequency filter reduced the number of tentative markers from 442 entities to 91.

Statistical Analysis

Statistical analysis for marker discovery is often tedious and time intensive when using complicated statistical software written to handle ASCI or text type results. MPP is ideal for the sophisticated data management, filtering, statistical analysis, interpretation, model creation, and prediction required to efficiently utilize complex and "noisy" data. It provides an easy-to-follow guided workflow that helps the user decide how best to evaluate the data, and expert users can go directly to the data processing they wish to use (see the Mass Profiler Professional brochure 5990-4164EN for further details).

Principle Component Analysis (PCA) is a frequently employed unsupervised multivariate analysis technique enabling data dimensionality reduction, while retaining the discriminating power in the data.

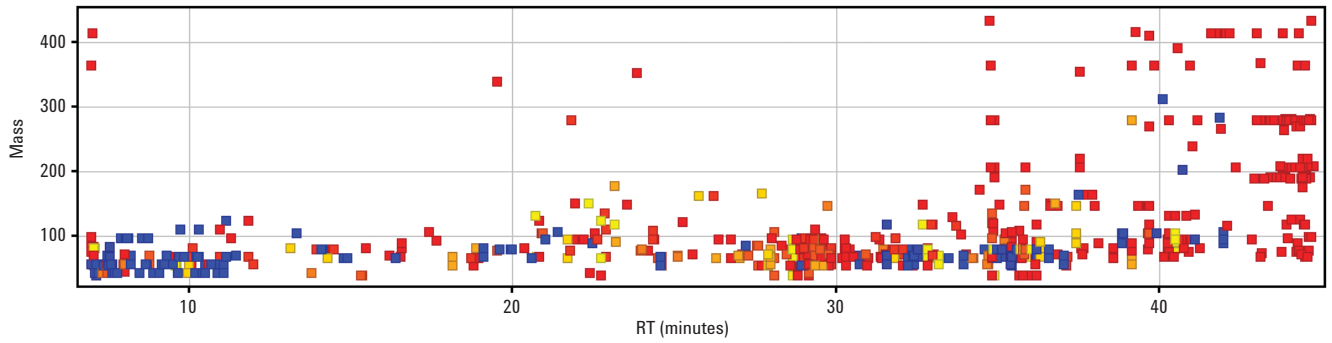

Figure 2. This mass versus retention time plot shows that 442 unique compounds were distinguished by chromatographic deconvolution, most of which occur only once or twice and are filtered out by MPP. The low frequency components are shown in red while the higher frequency components are shown in blue.

It is performed through the transformation of measured variables into uncorrelated principal components, each being a linear combination of the original variables. It is used as a quality control tool to provide an idea of how the data clusters and identify sample outliers. PCA of the entities that varied in amount between the sensory test pass and fail samples confirmed distinctive grouping of the data (Figure 3).

Fold Change and Statistical Significance

Once two classes of data were identified corresponding to olive oil samples that passed and failed the sensory test, statistical analysis was performed. The amount of fold-change (increase) in the concentration of any given compound was determined first. This analysis identified entities with large abundance differences between the selected data classes, that is, those that differed in concentration by two fold, three fold, four fold, and so forth between pass and fail EVOO samples.

Next, Analysis of Variance (ANOVA) was used to determine if the differences between those compounds that met the fold change criteria were statistically significant. Using a probability p value of 0.01, the 91 entities from the frequency filter were reduced to five significant compounds. The results of fold change analysis and ANOVA are displayed as a Volcano Plot (Figure 4). The five compounds with the lowest p values and highest fold-changes were selected for construction of the classification model.

Classification Model

The goal of classification is to produce general hypotheses based on a training set of examples that are described by several variables and identified by known labels corresponding to the class information. The task is to learn the mapping from the former to the latter. Numerous techniques, based either on statistics or on artificial intelligence, have been developed for that purpose [4]. In this case, the goal was to predict which olive oil samples would fail the sensory test, based on the five compounds that were shown to be associated with failure to pass the test.

Partial Least Square (PLS) analysis is particularly adapted to situations where there are fewer observations (that is, number of samples) than measured variables (for example detected entities, *m/z*). Its use has become very popular due to its ability to deal with many correlated and noisy variables. Partial Least Square Discrimination Analysis (PLSDA) is used to sharpen the partition between groups of observations, such that a maximum separation among classes is obtained, and has become a potent tool for the classification of metabolomics data [4]. Therefore, PLSDA was used to construct the olive oil classification model.

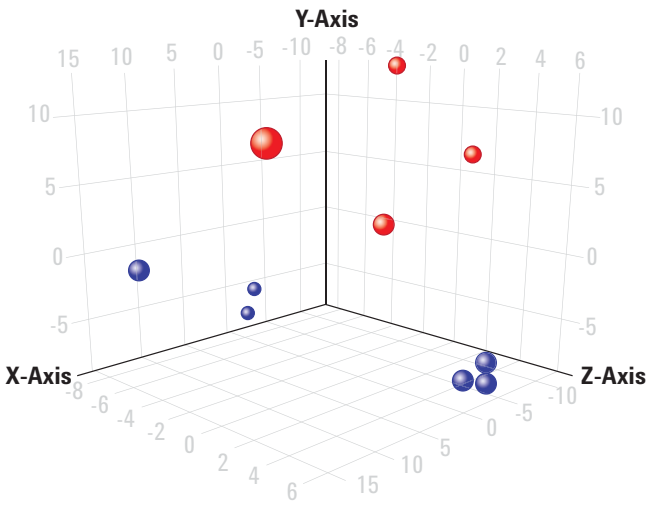

Figure 3. Principal Component Analysis (PCA) shows how data clusters. The samples that failed the sensory test are marked in red and the ones that passed are blue.

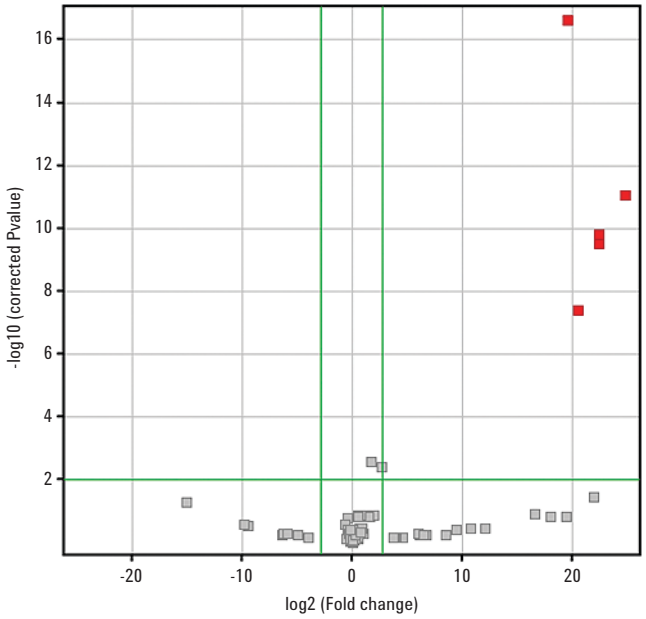

Figure 4. The Volcano Plot shows fold-change for each entity on the x-axis and significance on the y-axis. These five compounds are accumulated in the samples that failed the sensory test.

Having established two data classes with the five compounds that were selected through processing with MPP, the next step was to create a model that could predict whether an olive oil sample would pass the sensory test. The first step in building the classification model was to train the model with the data (Figure 5).

To test the model, the same training data as well as samples not included in the construction of the model were used. Although redundant, this is a valid statistical procedure. The same class prediction model was used for the validation of the trained model. The accuracy of the model for this limited number of samples, including those not used to construct the model, was 100% (Figure 6). These results demonstrate the feasibility of developing a model to accurately predict whether an EVOO would pass the sensory test.

Identifier	Training	Predicted(Training)	Confidence
CSC1-EI-1: Ig2	[F, Training]	[F, Training]	1.000
FSW2-EI-1: Ig2	[F, Training]	[F, Training]	1.000
ESC1-EI-1: Ig2	[P, Training]	[P, Training]	1.000
ESC2-EI-1: Ig2	[P, Training]	[P, Training]	1.000
RSA1-EI-1: Ig2	[P, Training]	[P, Training]	1.000

Figure 5. PLSD training set which contains representatives from each of the three clusters found in the PCA plot.

Prediction Results

Identifier	Grade	Training	Predicted(Class Pre...	Confidence
PAC1-EI-1: Ig2	F	None	[F, Training]	1.000
ESC2-EI-1: Ig2	P	Training	[P, Training]	1.000
ESC1-EI-1: Ig2	P	Training	[P, Training]	1.000
SAC1-EI-1: Ig2	F	None	[F, Training]	1.000
RFC2-EI-1: Ig2	P	None	[P, Training]	1.000
RSA2-EI-1: Ig2	P	None	[P, Training]	1.000
CSC1-EI-1: Ig2	F	Training	[F, Training]	1.000
RSA1-EI-1: Ig2	P	Training	[P, Training]	1.000
EFC1-EI-1: Ig2	P	None	[P, Training]	1.000
FSW2-EI-1: Ig2	F	Training	[F, Training]	1.000

Figure 6. The model correctly predicted the pass or fail status of all samples, including those not used to construct the model. The samples that were not used for building the prediction model are listed with the Training variable set as 'None'.

Identification of Compounds

The advantage of an instrument such as the Agilent 7200 series GC/Q-TOF is that it can collect data in EI, CI, and Product Ion Scan modes. These orthogonal modes of operation aid confirmation. EI spectra allow library searching and provide fragmentation data; CI provides information about the empirical formula, and Product Ion Scan MS/MS generates data for an accurate mass substructure search that can be applied to EI or CI generated ions.

While it is not necessary to know the identity of the compounds used in the classification model, identification could lead to an understanding of the mechanism by which those chemical components might, directly or indirectly, adversely affect the sensory qualities of olive oil. The Agilent 7200 series GC/Q-TOF offers a strong advantage for compound identification (ID) by providing accurate mass structural information.

Agilent MassHunter Qualitative Software was used to perform accurate mass chromatographic deconvolution and extract clean spectra from interfering peaks. These EI spectra were then searched against the NIST database (Figure 7). EI spectra of all except the last compound were found to have a corresponding match. Although the fragmentation patterns were consistent, the match factors were lower than those that would be obtained with a quadrupole-based instrument, since most of the data in the NIST library comes from quadrupole mass spectrometers. The two types of mass spectrometers exhibit optimal performance at different mass ranges, as the response of the quadrupole-based instrument is optimal at a lower mass range than that of the time-of-flight.

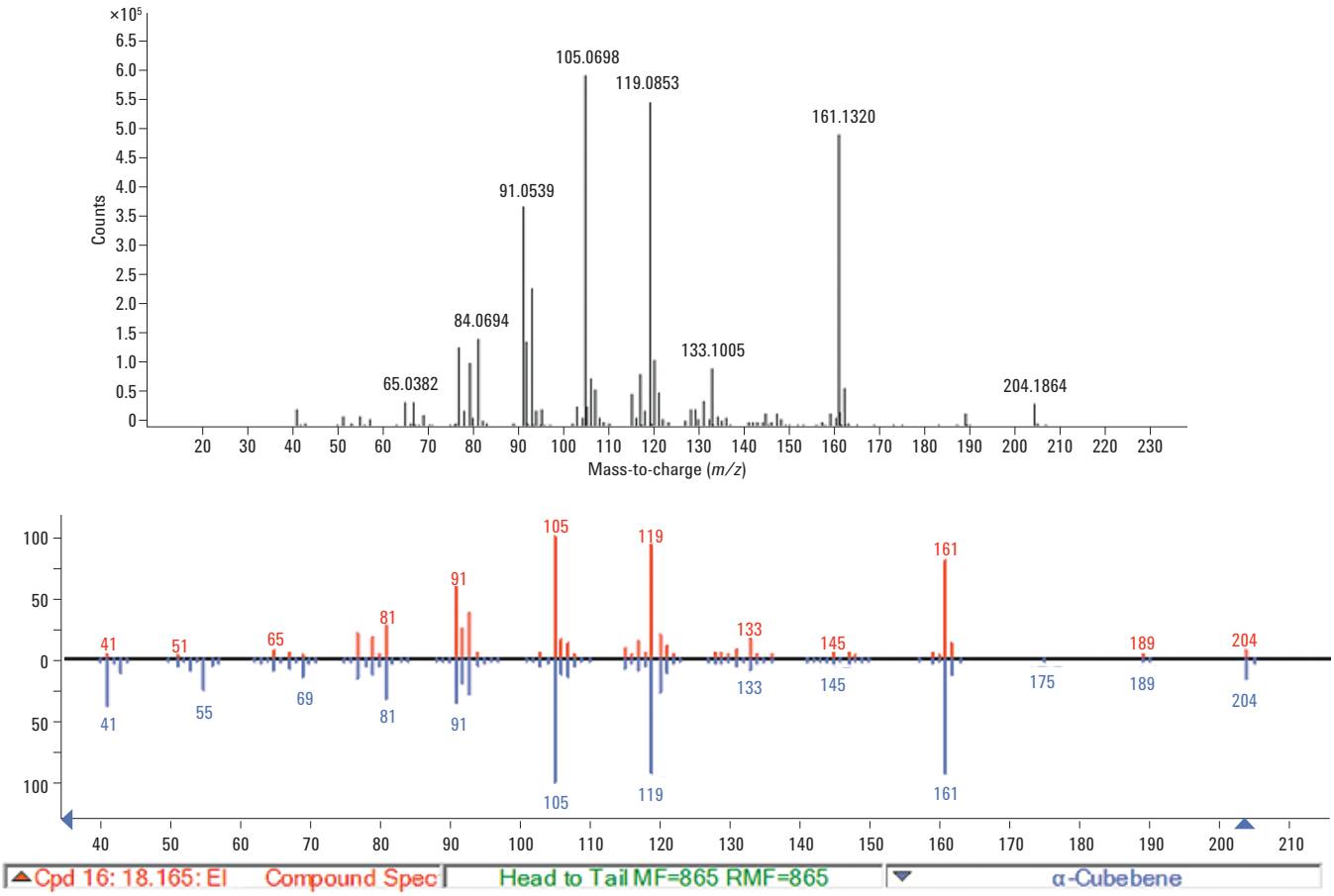

Figure 7. Commercial unit mass EI spectral libraries like Wiley and NIST can be searched using accurate mass EI GC/Q-TOF data to identify compounds.

Accurate mass information available for EI ions helped to confirm tentative identifications for accumulated compounds, with a mass accuracy below 5 ppm for all but one of the compounds, which lacked a prominent molecular ion (Figure 8).

Using the positive CI accurate mass data, a molecular formula for each marker compound was generated (Figure 8). This data confirmed the molecular formula of the fifth compound (bottom row), due to the fragmentation pattern observed. In addition to the expected peak at 227 *m/z* and its $(M+C_3H_5)+$ PCI adduct, a 209 *m/z* fragment, which is consistent with the loss of H_2O, was observed. The presence of a 191 *m/z* fragment indicates the loss of a second water molecule. This data leads to the hypothesis that the compound is a diol with an empirical formula of $C_{14}H_{26}O_2$. The mass accuracy for this compound is just above 8 ppm, consistent with the low signal intensity. A search for the properties of the four compounds identified by the NIST database search determined that they may all have scents that contribute to the flavor of the olive oil and result in its failure to pass the sensory test (Figure 9).

					EI, M*+			PCI, [M+H]+		
MPP ID	Tentative NIST ID	NIST Match	Formula	CAS	Calculated	Measured	Mass Error (ppm)	Calculated	Measured	Mass Error (ppm)
55.0@27.546	n-Hexadecanoic acid	789	$C_{16}H_{32}O_2$	57-10-3	256.2397	256.2385	4.7	257.2475	257.2470	1.9
73.0@29.750	Octadecanoic acid, ethyl ester	703	$C_{20}H_{40}O_2$	111-61-5	312.3023	312.3008	4.8	313.3101	313.3091	3.2
81.0@35.731	Squalene	831	$C_{30}H_{50}$	111-02-4	410.3907	410.3904	0.7	411.3985	411.3987	0.5
105.0@20.906	α-Cubebene	880	$C_{15}H_{24}$	17699-14-8	204.1873	204.1883	4.9	205.1951	205.1945	2.9
71.0@27.260	Not in NIST Database	N/A	$C_{14}H_{26}O_2$	N/A	226.1927	ND	ND	227.2006	227.1987	8.4

Figure 8. PCI spectral data provided accurate mass information for molecular ions of the accumulated compounds in olive oils that fail the sensory test, including the case where the EI spectrum showed no prominent molecular ion (last row).

MPP ID	Tentative NIST ID	NIST Match	Formula	CAS	Odor	Source
55.0@27.546	n-Hexadecanoic acid	789	$C_{16}H_{32}O_2$	57-10-3	Faint Oily	Bedoukian Research
73.0@29.750	Octadecanoic acid, ethyl ester	703	$C_{20}H_{40}O_2$	111-61-5	Waxy	The Good Scents Company
81.0@35.731	Squalene	831	$C_{30}H_{50}$	111-02-4	Floral	The Good Scents Company
105.0@20.906	α-Cubebene	880	$C_{15}H_{24}$	17699-14-8	Herbal	The Good Scents Company

Figure 9. A list of the odor characteristics of four of the identified compounds associated with failure of the sensory test.

Structure Confirmation using Molecular Structure Correlator

Q-TOF product ion spectra can help verify that all the fragment ions generated can be correlated to the proposed structural isomer. The Molecular Structure Correlator performs a substructure search of the ChemSpider database and correlates the results to all the possible structural isomers. Each individual fragment ion is ranked based on mass error corresponding to the proposed formula, along with a penalty based on how many bonds needed to be broken to generate that proposed formula. An isomer's individual Compatibility Score is a weighted average of the fragment ion scores, taking into account the intensity and the mass of each fragment ion (Figure 10).

Note that this tool is complementary to the EI library search which clearly identifies the peak at 29.75 minutes as being an ethyl ester. The best match for the compound at 29.75 minutes is ethyl octadecanoate. The Molecular Structure Correlator shows that the product ions of the 312 *m/z* precursor correlate well to ethyl octadecanoate, with a compatibility score greater than 98 (Figure 10). Moreover, the accurate masses of the fragments correlate well to the expected masses, with all fragments being within mass error of 5 ppm. This provides additional supporting information for compound identification.

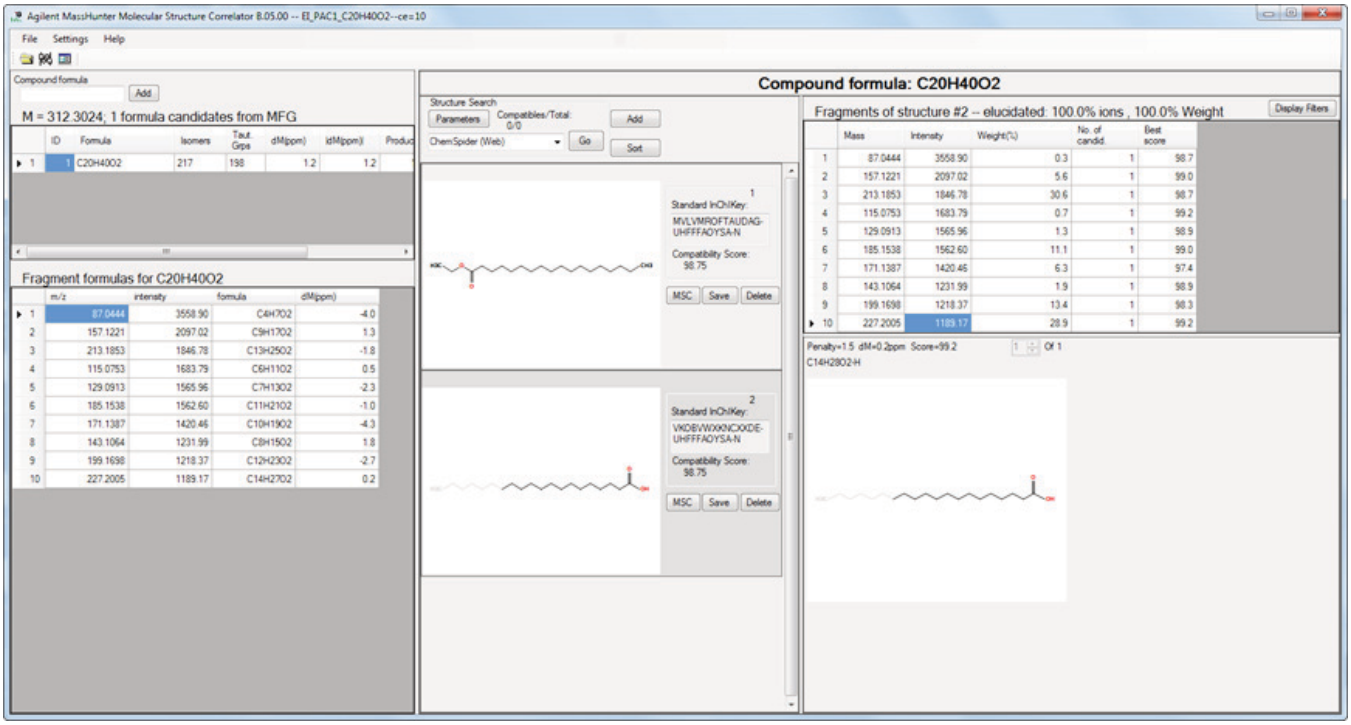

Figure 10. Molecular Structure Correlator compares the Q-TOF product ion spectra to the structural isomers of the empirical formula. It determines which product ions correlate to fragments of the isomers and generates a compatibility score.

Conclusions

Using accurate mass EI and positive CI scan data generated by the Agilent GC/Q-TOF system, a model was constructed that accurately predicted whether an olive oil would pass the sensory test. Although it was constructed using a very small sample set, it demonstrates the feasibility of this approach. A predictive model constructed using a significantly larger sample size would give olive oil producers an inexpensive, quick test to determine whether their oil would pass the sensory test, thus avoiding the costly and time-consuming sensory testing of inferior oils. The identification of compounds accumulated in EVOOs that fail the sensory test as scents lends credibility to their negative contribution to the flavor of olive oil. This approach could also conceivably be used to construct a model that could predict whether an olive oil has been adulterated with other less expensive oils.

References

1. Olive Oil in the U.S., 3rd Edition, Apr 1, 2009, a marketing report from Packaged Facts (http://www.packagedfacts.com /Olive-Oil-Edition-2071654/)
2. E. N. Frankel, R. J. Mailer, C. F. Shoemaker, S. C. Wang, J. D. Flynn, "Tests indicate that imported "extra virgin"olive oil often fails international and USDA standards" UC Davis Olive Center, July 2010 (http://olivecenter.ucdavis.edu/news-events/news/files/olive%20oil%20final%20071410% 20.pdf)
3. L. Vaclavik, O. Lacina, J. Hajslova, J. Zweigenbaum. "The use of high performance liquid chromatography-quadrupole time-of-flight mass spectrometry coupled to advanced data mining and chemometric tools for discrimination and classification of red wines according to their variety.", *Anal Chim Acta*. **685**, 45-51 (2011)
4. J. Boccard, J. L. Veuthey, S. Rudaz. "Knowledge discovery in metabolomics: an overview of MS data handling." *J Sep Sci*. **33**, 290-304 (2010)

For More Information

These data represent typical results. For more information on our products and services, visit our Web site at www.agilent.com/chem.

www.agilent.com/chem

Printed in the USA
March 23, 2012
5991-0106EN

Agilent Technologies