Classification of Chamomile Flowers, Essential Oils, and Commercial Products Using Chemometrics and the Agilent 5975 GC/MSD

Application Note

Food Testing & Agriculture

Authors

Mei Wang, Bharathi Avula, Yan-Hong Wang, Jianping Zhao, Amar G. Chittiboyina, Cristina Avonto, Jon F. Parcher, Vijayasankar Raman, and Ikhlas A. Khan
National Center for Natural Products Research,
University of Mississippi
University, MS
USA

Jerry A. Zweigenbaum and Philip L. Wylie
Agilent Technologies, Inc.
Wilmington, DE
USA

Abstract

A highly accurate statistical model has been developed to determine the exact type of chamomile used in the production of commercial herbal products. The model was developed from accurate GC/MS data obtained with an Agilent 7890 GC and Agilent 5975 GC/MSD. Quality control of the samples was performed by Principal Component Analysis (PCA) and a sample class prediction model based on Partial Least Squares Discriminant Analysis (PLS-DA) was constructed. The model demonstrated 100% accuracy for both recognition and prediction abilities. In addition, 35 commercial products and 11 essential oils purported to contain chamomile were subsequently predicted by the model based on PLS-DA.

Agilent Technologies

Introduction

As one of the most widely used medicinal plants in the world, chamomile has been reported to be beneficial for relief of sleeping disorders, diarrhea, colic, wounds, mucositis, and eczema [1,2]. Additional medicinal properties have been attributed to chamomile, including anti-allergic, antibacterial, anti-inflammatory, and antispasmodic attributes [3]. There have, however, been some reports of allergic reactions, including skin rashes, throat swelling, shortness of breath, and anaphylaxis.

Unfortunately, there is no universally accepted characterization of chamomile, making it difficult to determine and control the quality, safety, and efficacy of herbal medicines containing such a poorly defined natural product. In fact, one of the most serious obstacles to the promotion of herbal chamomile products is adulteration.

This application note describes a recently published study [4] of the chemical composition of the three most common types of chamomile found in commercial products: German chamomile (*Matricaria chamomilla* L. syn: *M. recutita* L.); Roman chamomile (*Chamaemelum nobile* (L). All. syn: *Anthemis nobilis* L.) and Juhua (*Chrysanthemum morifolium* Ramat.). A gas chromatography/mass spectrometry (GC/MS) method was developed and applied for the nontargeted, volatile, apolar compound analysis of a collection of chamomile samples that included authenticated plants, commercial products, and essential oils.

The method used an Agilent 7890 GC and an Agilent 5975 GC/MSD to generate the data. Control of input variety, alignment of retention time, and data reduction was accomplished using an automatic data processing procedure and data filters employing various criteria. PLS-DA was used to construct a model that classifies and discriminates samples of interest. The model was used to assess commercial samples that claimed to contain chamomile. Identification of the major marker compounds correlated with each type of chamomile was also achieved.

Experimental

Samples

The investigated samples included 27 authenticated plants, 35 solid commercial products, and 11 essential oils. Specimens of all samples are deposited at the botanical repository of the National Center for Natural Products Research (NCNPR), University of Mississippi (documented with NCNPR accession code) [4].

Reagents and standards

Chemicals and standards were obtained and used as described [4].

Instruments

This study was performed on an Agilent 7890 GC equipped with an Agilent 7693A Automatic Liquid Sampler and coupled to an Agilent 5975 GC/MSD system. The instrument conditions are listed in Table 1.

Table 1. GC and Mass Spectrometer Conditions

GC conditions	
Precolumn	None
Analytical column	Agilent J&W HP-5MS 30 m × 0.25 mm, 0.25 µm 5% phenyl methyl silicone (p/n 19091S-133)
Injection temperature	250 °C
Injection mode	Split ratio set to 25:1
Oven program	2 minutes at 45 °C 1.5 °C/min to 100 °C 2 °C/min to 200 °C
Column flow	1 mL/min constant flow
Carrier gas	Helium
Transfer line temperature	280 °C
GC run time	90 minutes
MS conditions	
Ionization mode	Electron impact at 70 eV
Ion source temperature	230 °C
Solvent delay time	5 minutes
Acquisition mode	Scan (40–550 amu)

Sample preparation

Solid samples were ground and homogenized to obtain a uniform matrix. Approximately 1 g of the fine powder was accurately weighed, dispersed in 4 mL of *n*-hexane, and sonicated for 1 hour. The supernatant was filtered with a Millex GV (0.22 µm) filter prior to GC/MS analysis. For the essential oils, 10 µL samples were diluted in 1 mL of *n*-hexane. The selected internal standard tridecane (n-$C_{13}H_{28}$) was added to each sample solution to a final concentration of 90 µg/mL.

Data processing and statistical analysis

Agilent MSD Productivity ChemStation software (E.02.02) was used to acquire the GC/MS data. NIST Automated Mass Spectral Deconvolution and Identification Software (AMDIS) was used to extract the GC/MS data. Entities were identified as ions with identical elution profile and similar spectral data, and characterized by retention time (t_R), peak intensity, and *m/z*.

The ELU file for each sample (created by AMDIS) was imported into Mass Profiler Professional (MPP) software, which included several Sample Class Prediction (SCP) algorithms. A minimum abundance setting of 5,000 counts was used to select entities for further analysis in a 5 to 90-minute retention time window. A tolerance retention time window of 0.15 minutes and similarity of spectral pattern were used to align entities across the entire sample set. Normalization of the peak intensity, using the internal standard, was performed to account for the difference in the abundances of each compound.

Stepwise reduction of entity dimensionality was performed based on their presence across samples and parameter values (filter by flags), frequency of occurrence (filter by frequency), abundance of the respective entities in classes (filter by sample variability), and results of one-way analysis of variance (ANOVA). PCA was used for quality control of the sample data, and a sample class prediction model based on PLS-DA was constructed.

A cross-validation procedure was carried out to validate the model, using a series of 12 samples that included authenticated plant samples used in the previous model training as well as commercial samples with known labels that were not included in the model training.

Results and Discussion

Data acquisition

Since no specific group of target analytes had been defined *a priori*, nontargeted analysis in the scan mode was performed to maximize the information gained. A large number of compounds were detected in the GC/MS analysis of the authenticated chamomile plant samples. There were some slight variations in the concentrations of components in the plant samples of a given type of chamomile, but the chromatographic patterns from the same type of chamomile were consistent. Different types of chamomile showed distinct differences in their chemical profiles (Figure 1).

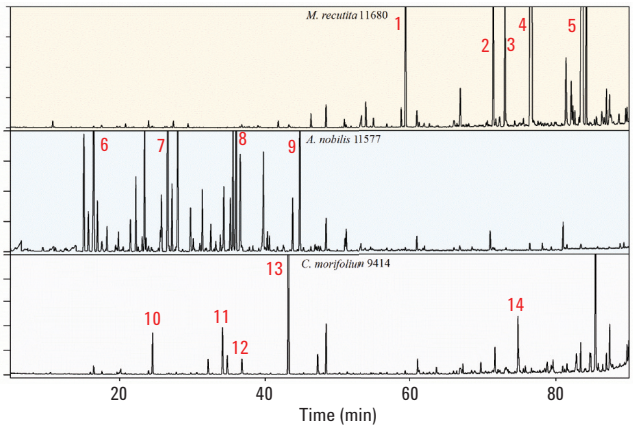

Figure 1. Typical chromatograms of German chamomile (top panel), Roman Chamomile (middle panel), and Juhua (lower panel). Major compounds identified in different types of chamomile were: (1) Farnesene; (2) Bisabolol oxide B; (3) a-Bisabolol; (4) Bisabolol oxide A; (5) cis-Enyne-dicycloether; (6) a-Pinene; (7) 2-Butenoic acid, 3-methyl-, butyl ester; (8) 2-Butenoic acid, 3-methyl-, 3-methylbutyl ester; (9) 2-Butenoic acid, 3-methyl-, hexadecyl ester; (10) Eucalyptol; (11) Trimethylcyclohexane aldehyde; (12) Borneol; (13) Pinene acetate; (14) Lanceol.

Data mining

Using an intensity threshold of 5,000 counts, a total of 2,560 entities were obtained by the MPP software. A stepwise filtering procedure was used to identify the most characteristic marker compounds representing different types of chamomile, and to reduce the dimensionality of the data prior to PCA and PLS-DA. The first step in this process was to ''filter by flags''. Flags are attributes that denote the quality of entities within a sample and indicate whether the entities were detected in each sample as 'present' or 'marginal'. The entities present in all samples were removed from further analysis, and only those entities unique to each sample were retained.

In the second filter, a ''filter by frequency'' step, entities that were not present in at least 100% of the samples in at least one sample group (for example, all of the Roman chamomile samples) were removed. The third filter was a ''filter by sample variability'' step, in which entities were filtered based on a Coefficient of Variation (CV) in their abundance level of less than 25%.

The final filter step selected the most reproducible data based on p-values calculated for each entity using one-way ANOVA. A p-value cut-off of 0.05 was used to ensure that only entities that differed in the respective varieties with 95% statistical significance were passed.

Although the initial number of entities before the filtering procedure was applied was 2,560, this number was reduced to 50 after stepwise filtering. This filtering process assured that only the most discriminant entities were used to construct the prediction model.

Chemometric analysis

PCA is a mathematical method that enables data dimensionality reduction while retaining the discriminating power in the data. It uses an unsupervised approach (without using the conditions or groups) to find differences between samples, determine group associations, and weigh the relative contributions of compounds to the separation of the groups. After PCA was applied, 74% of the variability in the data was explained by Principal Component 1 (PC1), providing good separation of Roman, German, and Juhua types of chamomile (Figure 2A). An additional 22% of the variation was found in PC2, and further separated Juhua from German. The Juhua samples vary significantly across PC3, but that principal component only accounted for 1.5% of the total variation.

Thus, PCA was used to provide a visual representation of how the data clusters and to identify outliers, as a quality control tool. This set of data that was processed by filtering and PCA quality control analysis was then used to create the sample prediction model.

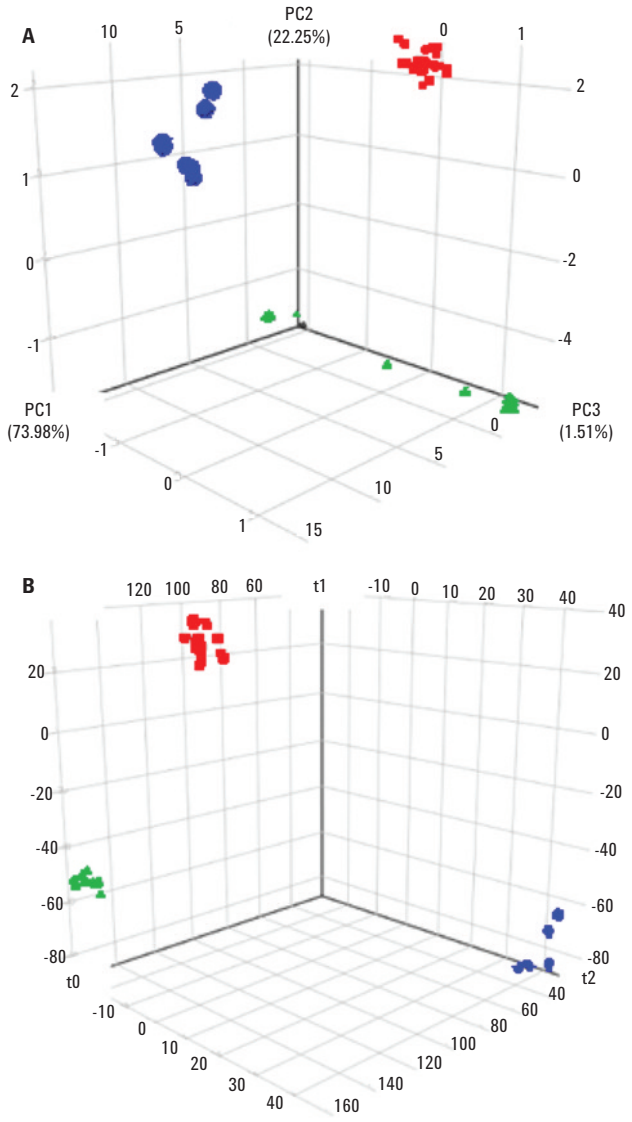

Figure 2. Scores plots of (■) German Chamomile, (●) Roman Chamomile, and (▲) Juhua. (A) PCA. (B) PLS-DA.

Class prediction models

Several techniques have been developed to construct sample prediction models. Five algorithms are provided by the MPP software for building sample prediction models: PLS-DA, Support Vector Machines (SVM), Naive Bayes (NB), Decision Tree (DT), and Neutral Network (NN). PLS-DA is a well-established regression-based method especially adapted to situations involving fewer samples than measured variables. It is often used to sharpen the partition between groups of observations and maximize the separation among classes. In this study, it was found to be the best algorithm for constructing a statistical model for chamomile classification and differentiation.

The first step in building the prediction model was to train it with the spectral data from authenticated plant samples, including 15 German, eight Juhua, and four Roman chamomile samples. Six authenticated samples used for the model training and six commercial samples not included in the construction of the model were used for validation. This is a legitimate statistical procedure (k-fold cross validation), in spite of the fact that it is redundant [5].

The results of sample classification using this model are summarized in Table 2. The percentages of the samples correctly classified during model training and validation are represented by the recognition and prediction abilities, respectively. The validation procedure is used to select the most appropriate model generated from the five algorithms, and a difference between training and validation results can indicate overfitting. Our results indicated that no samples were misclassified during model training and validation. The three types of chamomile were well separated.

Table 2. Chamomile Classification Model Training and Validation Results

	German	Roman	Juhua	Accuracy (%)
Model training				
German	15	0	0	100.0
Roman	0	4	0	100.0
Juhua	0	0	8	100.0
Recognition ability (%)	–	–	–	100.0
Model validation				
German	4	0	0	100.0
Roman	0	4	0	100.0
Juhua	0	0	4	100.0
Prediction ability (%)	–	–	–	100.0

The PLS-DA t-score plot, shown in Figure 2B, is supervised, unlike PCA, and it uses the conditions to 'fit' the model to the data. The t-scores plot gives a visual representation of how well the samples in each group were separated. Although the scores plot of both PCA and PLS-DA look similar, PCA reveals the structure of the data and PLS-DA shows how the model fits the data. Figure 2B shows excellent separation (even Juhua is grouped tightly together) and suggests that the model can predict the three types of chamomile, if the model is not overfit.

Classification of chamomile samples

Thirty-five solid samples and 11 chamomile essential oils were classified and differentiated using the validated PLS-DA model. The solid samples included chamomile flowers, extracts, teas, flower and leaf, dietary supplements containing herbal chamomile, and fruit teas. The prediction results were expressed in the form of a 'confidence measure' (Table 3). Confidence measures in the range > 0.7 indicate a high degree of certainty that the samples belong to the indicated chamomile type, while confidence measures of 0.5–0.7 indicate problematic sample classifications. Confidence measures < 0.5 suggest probable misclassification, mishandling, adulteration, or impurity of the samples. All samples with < 0.6 confidence measures were chosen for further investigation. Several of the essential oils samples were classified as deriving from Roman chamomile, but surprisingly, none of the commercial solid samples were classified by the model as containing Roman chamomile.

The model prediction results were consistent with the labels, except for four outliers, samples 2061, 3998, 9384, and 9425. The results suggested that German chamomile is the major type of chamomile used in the U.S. market, because all the unknown labeled chamomile teas or extracts from the U.S. were identified as German chamomile. In contrast, all the chamomile samples purchased from China were identified as Juhua chamomile. In spite of the fact that the essential oils were obtained by steam distillation whereas the solid samples were extracted with hexane, the PLS-DA model showed good prediction results for all the chamomile essential oil samples (Table 3).

Table 3. Chamomile Type Prediction Results from the PLS-DA Sample Class Prediction Model for Commercial Products and Essential Oils

No.	NCPR accession code	Product information from the label	Predicted	Confidence measure
Commercial samples in solid form purchased from food markets, retail pharmacies and online				
1	2061	Roman chamomile	German	0.47
2	3670	Chamomile flower	German	0.92
3	3998	Chamomile extracts	German	0.53
4	4903	Chamomile powder	German	0.90
5	5770	Chamomile powder	German	0.93
6	7359	Chamomile powder	German	0.81
7	9357	Chamomile flowers	German	0.82
8	9359	Chamomile flowers	German	0.84
9	9361	Chamomile Flower and Leaf Dietary Supplement	German	0.76
10	9362	Chamomile flowers	German	0.84
11	9364	Chamomile flowers	German	0.92
12	9365	Bulk Chamomile Flowers, German	German	0.65
13	9367	Chamomile Flowers, Herbal Dietary Supplement	German	0.68
14	9382	Chamomile Organic Tea (Leaves and flowers)	German	0.94
15	9383	Herbal Chamomile & Fruit Tea (Rosehips, chamomile, orange peel, lemon peel & lemon myrtle)	German	0.72
16	9384	Chamomile Herb Tea	German	0.58
17	9385	Organic Tea	German	0.81
18	9386	Chamomile Tea	German	0.75
19	9387	Chamomile Herbal Tea	German	0.91
20	9388	Chamomile Herb Dietary Supplement	German	0.89
21	9389	Chamomile Herbal Tea	German	0.61
22	9390	Chamomile Herbal Tea	German	0.92
23	9391	Chamomile Herbal Tea	German	0.77
24	9393	Whole German Chamomile Flowers	German	0.87
25	9422	Chamomile Herbal Dietary Supplement	Juhua	0.80
26	9423	Chamomile Herbal Dietary Supplement	Juhua	0.83
27	9424	Chamomile Herbal Dietary Supplement	Juhua	0.84
28	9425	Chamomile Herbal Dietary Supplement	Juhua	0.60
29	9426	Chamomile Herbal Dietary Supplement	Juhua	0.86
30	9427	Chamomile Herbal Dietary Supplement	Juhua	0.78
31	9428	Chamomile Herbal Dietary Supplement	Juhua	0.82
32	9429	Chamomile Herbal Dietary Supplement	Juhua	0.81
33	9430	Chamomile Herbal Dietary Supplement	Juhua	0.77
34	9431	Chamomile Herbal Dietary Supplement	Juhua	0.72
35	9432	Chamomile Herbal Dietary Supplement	Juhua	0.99

Continued next page

Table 3. Chamomile Type Prediction Results from the PLS-DA Sample Class Prediction Model for Commercial Products and Essential Oils (contuinued)

No.	NCPR accession code	Product information from the label	Predicted	Confidence measure
Chamomile essential oils obtained by steam distillation from plant samples or purchased from difference commercial sources				
1	9254E	Chamomile Oil (Anthemis nobilis), Steam Distillation from Plant (9254)	Roman	0.72
2	9359E	Chamomile Oil (Matricaria recutita), Steam Distillation from Plant (9359)	German	0.76
3	9362E	Chamomile Oil (Matricaria recutita), Steam Distillation from Plant (9362)	German	0.71
4	11577E	Chamomile Oil (Anthemis nobilis), Steam Distillation from Plant (11577)	Roman	0.70
5	11680E	Chamomile Oil (Matricaria recutita), Steam Distillation from Plant (11680)	German	0.77
6	11681E	Chamomile Oil (Matricaria recutita), Steam Distillation from Plant (11681)	German	0.89
7	9368	Chamomile Essential Oil (Anthemis nobilis)	Roman	0.70
8	9369	Chamomile Oil, German	German	0.91
9	9370	Chamomile Essential Oil (Anthemis nobilis)	Roman	0.76
10	9380	Chamomile Essential Oil (Chamaemelum nobilis)	Roman	0.69
11	9381	Roman Chamomile Essential Oil	Roman	0.73

The model used three groups of entities that were primarily specific to each of the three types of chamomile. The Venn diagram illustrating these three groups and the minimal overlap between them is shown in Figure 3. The major compounds corresponding to these entities in each type of chamomile were identified and are given in Table 4. These compounds were consistent with those previously reported in the literature as components of chamomile.

Since this type of model enables assignment of new samples into previously determined groups in an unbiased fashion, it would be very useful for the quality control of many natural products and dietary supplements. The batch samples can be automatically acquired, processed, and class predicted.

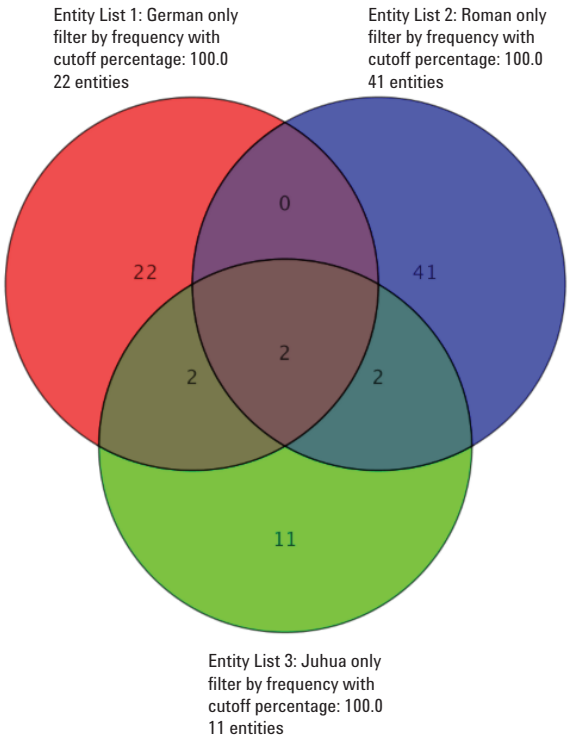

Figure 3. Venn diagram of chamomile samples.

Table 4. The Major Marker Compounds Tentatively Identified in the Construction of the Class Prediction Model for Chamomile

Entities	*m/z*	t_R (min)	Tentative compound identification *	Molecular weight	CAS number
Roman chamomile					
1	71.0	15.10	Isobutyric acid, isobutyl ester [a]	144	97-85-8
2	93.0	16.43	1R-α-Pinene [a,b]	136	7785-70-8
3	71.0	23.42	Isobutyric acid, 2-methylbutyl ester [a]	158	2445-69-4
4	55.0, 83.0	26.64	2-Butenoic acid, 3-methyl-, butyl ester [a]	156	54056-51-8
5	70.0	34.33	Trans-(–)-Pinocarveol [a]	152	547-61-5
6	55.0, 83.0, 100.0	36.01	2-Butenoic acid, 3-methyl-, 3-methylbutyl ester [a]	170	56922-73-7
7	81.0	36.58	Pinocarvone	150	30460-92-5
8	83.0	39.75	3-Methyl-2-butenoic acid, 3-methylbut-2-enyl ester	168	299309
9	100.0	44.75	2-Butenoic acid, 3-methyl-, hexadecyl ester [a]	324	60129-26-2
German chamomile					
1	205.0	66.94	Spathuleno [a,b]	220	77171-55-2
2	143.0	71.43	α-Bisabolol oxide B [a,b]	238	26184-88-3
3	93.0, 141.0	73.04	α-Bisabolol [a,b]	222	515-69-5
4	176.0	75.07	Coumarin, 7-methoxy-	176	531-59-9
5	143.0	76.07	Bisabolol oxide A [a,b]	238	22567-36-8
6	143.0	81.36	α-Bisabolol oxide A derivative [c]		
7	143.0	82.30	α-Bisabolol oxide A derivative [c]		
8	128.0	83.70	E-1,6-Dioxaspiro[4.4]non-3-ene, 2-(2,4-hexadiynylidene)- [b]	200	50257-98-2
9	200.0	84.10	Z-1,6-Dioxaspiro[4.4]non-3-ene, 2-(2,4-hexadiynylidene)- [b]	200	4575-53-5
Juhua					
1	95.0	36.82	Borneol [b]	154	10385-78-1
2	132.0	61.06	α-Curcumene [a]	202	644-30-4
3	91.0	67.27	Caryophyllene oxide [a]	220	1139-30-6
4	105.0, 121.0	69.75	Alloaromadendrene oxide [a]	220	156128
5	204.0	71.69	Eudesm-7(11)-en-4-ol [a]	222	473-04-1
6	69.0	79.61	Isoaromadendrene epoxide	220	159366
7	109.0	85.38	Cyclopropanemethanol, α,2-dimethyl-2-(4-methyl-3-pentenyl)-,[1α(R'),2α]- [c]	182	121959-70-4

* Identified by data base search.

[a] Identified by comparison of relative retention index to literature.

[b] Identified by reference standards.

[c] Identified with low database match probability.

Conclusion

Untargeted GC/MS analysis using the Agilent 7890 GC and Agilent 5975C GC/MSD can provide the information-rich data required to classify and differentiate natural products such as chamomile. Agilent MSD Productivity ChemStation and Agilent Mass Profiler Professional software enabled the automatic mining and processing of the data to find the most characteristic marker compounds and construct a highly accurate model for predicting which of the three major types of chamomile was the source for a commercial product. This tool can provide clear definition of chamomile-derived products, thus improving their quality, safety, and efficacy, as well as detecting adulteration and substitutions.

References

1. D.L. McKay and J.B. Blumberg "A review of the bioactivity and potential health benefits of chamomile tea (*Matricaria recutita* L.)" *Phytotherapy Research* **20**, 519–530 (2006).
2. S. Petronilho, *et al.* "*In vitro* and *in vivo* studies of natural products: A challenge for their valuation. The case study of chamomile (*Matricaria recutita* L.)" *Industrial Crops and Products* **40**, 1–12 (2012).
3. G.E. Buono-Core, *et al.* "Structural elucidation of bioactive principles in floral extracts of German chamomile (*Matricaria recutita* L.)" *Journal of the Chilean Chemical Society* **56**, 549–553 (2011).
4. M. Wang, *et al.* "An integrated approach utilising chemometrics and GC/MS for classification of chamomile flowers, essential oils and commercial products" *Food Chem* **152**, 391–398 (2014).
5. L.A. Berrueta, R.M. Alonso-Salces, K. Heberger "Supervised pattern recognition in food analysis" *J. Chromatogr A* **1158**, 196–214 (2007).

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
February 17, 2014
5991-3969EN

Agilent Technologies