Journal of Food Composition and Analysis 64 (2017) 85–92

Contents lists available at ScienceDirect

Journal of Food Composition and Analysis

journal homepage: www.elsevier.com/locate/jfca

Original research article

Botanical supplements: Detecting the transition from ingredient to product☆

James Harnly*, Yinjiang Lu[1], Jianghao Sun, Pei Chen

Food Composition and Methods Development Lab, Beltsville Human Nutrition Research Center, Agricultural Research Service, U.S. Department of Agriculture, Beltsville, MD 20705, USA

ARTICLE INFO

Keywords:
Botanical ingredients
Botanical supplements
Echinacea purpurea
Flow injection mass spectrometry
One class modeling
Principal component analysis
Analysis of variance
Spectral correlation

ABSTRACT

Preparation of commercial supplements from botanical ingredients results in a change in the chemical composition that is easily followed with flow injection mass spectrometry. Spectral fingerprints (counts vs ion with no chromatographic separation) were acquired by flow injection mass spectrometry (FIMS) for aerial parts of 16 *Echinacea purpurea* botanical ingredients (BIs) and 18 commercial botanical supplements (BSs) whose labels claimed or suggested *E. purpurea* aerial content. One class modeling of the BI spectral fingerprints showed them to be different from the BIs at the 95% confidence limit. The spectral components (ions) that provided discrimination between the BSs and BSs were identified using analysis of variance to obtain an F-value for each ion and produce an F-test spectrum. Ions found in both BI and BS spectra were identified by multiplication of the spectra to yield a correlation spectrum. The correlation spectra obtained from multiplication of the average BI spectrum and 6 single-ingredient BS spectra showed only one supplement to be significantly different. These correlation spectra verified the presence of the BI in the BSs and indicated that the analytical extraction and the BS preparation procedures used by 5 of the manufacturers were similar. Correlation spectra for 13 mixed or unknown ingredient BSs showed that 6 could be verified to contain *E. purpurea* aerial material.

1. Introduction

Sale of botanical supplements (BSs) is a billion dollar, world-wide industry that has grown at an incredible rate in the last two decades. Unfortunately, economically motivated adulteration of BSs has grown at an equal pace. Consequently, there is an increased demand for analytical methods to verify (i.e. authenticate) the content of BSs. BSs are the finished commercial products composed of one or more botanical ingredients (BIs). BIs may be whole plants or plant parts with purported health-promoting properties. In general, extracts of BIs are used either directly (liquid form) or added to an excipient and encapsulated or pressed into a tablet (solid form). Thus, BSs will consist of a subset of the components of the BIs and will differ in composition from BIs depending on the extraction process. Development of analytical and statistical methods that can relate the composition of the BSs to the original BIs are an analytical challenge.

Fig. 1 provides an overview of the transition of BIs to BSs and the analysis of both materials. BSs are prepared through what is often a complex preparation process that is inevitably proprietary (Upton, 2016). Compounds may be lost (through extraction) and gained (through addition of other ingredients). Analytically, the standard procedure is to prepare both BIs and BSs in the same manner, usually employing an aqueous alcohol solvent optimized for the extraction of small molecules, leaving proteins, nucleic acids, and complex carbohydrates behind. If the BS preparation process matches the analytical extraction, then the fingerprints of the BI and BS will match; i.e. both the analytical extraction and the BS preparation will select the same compounds with similar efficiencies.

Every manufacturer will expect some variation of BSs from lot to lot because of the natural variation of the BIs. However, there can be significant variation between manufacturers because they either start with different BIs or employ a different preparation process. Since the commercial processes are proprietary, it is impossible to determine the source of the differences. This problem is further complicated because many manufacturers acquire extracted BIs from brokers (Upton, 2016). Thus, it is inevitable that the fingerprints of the BIs cannot be used as reference samples to authenticate BSs using non-targeted analysis. In principle, only historical BSs from a manufacturer can be used as reference samples for their newest products.

The identity of a complex BI material is best confirmed using non-

Abbreviations: EPA, *Echinacea purpurea* aerial; EPAS, *E. purpurea* aerial solid supplement; EPAL, *E. purpurea* aerial liquid supplement

☆ This paper was originally given as an oral presentation at the 39th National Nutrient Databank Conference (NNDC) held from May 16–18, Alexandria, VA, USA.

* Corresponding author at: Food Composition and Methods Development Lab, Beltsville Human Nutrition Research Center, Agricultural Research Service, U.S. Department of Agriculture, 10300 Baltimore Ave Building 161 BARC-East, Beltsville, MD 20705, USA.

E-mail address: james.harnly@ars.usda.gov (J. Harnly).

[1] Current address: Food Quality Lab, Beltsville Agricultural Research Center, Agricultural Research Service, U.S. Department of Agriculture, Beltsville, MD, USA.

http://dx.doi.org/10.1016/j.jfca.2017.06.010
Received 19 October 2016; Received in revised form 10 January 2017; Accepted 30 June 2017
Available online 16 July 2017
0889-1575/ Published by Elsevier Inc.

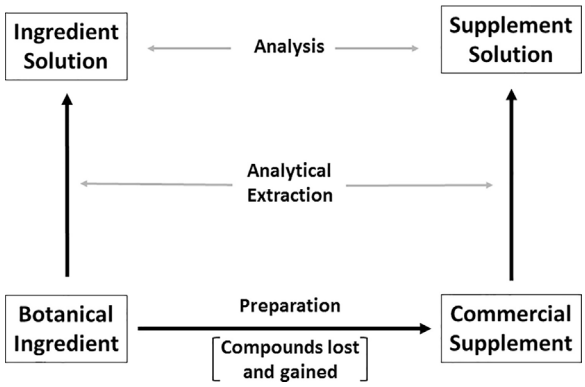

Fig. 1. Schematic for preparation of commercial botanical supplements from botanical ingredients and their analysis.

targeted fingerprinting methods coupled with chemometric analysis. The approach is conceptually simple. Reference samples (for the BI of interest) are collected and analyzed, a model is constructed, statistical limits are established, and the unknown BI is determined to be the same as (within the statistical limits) or different from (outside the limits) the reference samples. Regardless which BI is of interest, the approach is the same logistically, analytically, and chemometrically.

Logistically, the collection of BI reference samples should account for as much of the natural variation of the plant or plant part as possible. This includes variables associated with genetics (species, cultivar, hybrid), environment (geographic location, climate, season, growing conditions), and management (harvesting, processing, storage) (Walthall et al., 2012). Certified botanical reference materials from metrological organizations, such as the National Institutes of Standards and Technology, have limited value. They are intended for quantitative validation. While they can validate the general composition of the collection of reference samples (falling within the range of the reference samples), they cannot provide insight into the natural variability of a botanical material.

A large collection of authentic materials is ideal, but it will often reveal a lack of homogeneity. For example, systematic differences have been observed for American ginseng (*Panax quinquefolius*) grown in the US, Canada, and China (Sun et al., 2013) and black cohosh (*Actaea racemosa*) grown at 22 sites in the eastern US (Harnly et al., 2016). Thus, a conscientious effort to characterize variability may, in fact, produce unanswered questions due to incomplete metadata, i.e. full background information for the samples.

Analytically, the chosen method must be non-targeted and as comprehensive as possible. While targeted methods have been used to characterize different species (e.g. ginsenoside ratios for *Panax* species, Chan et al., 2000), genetic and metabolite fingerprinting methods are much more robust. The cost of genetic testing is decreasing, but contamination, low quality of the DNA, and inability to differentiate between plant parts (a common form of adulteration) can be problematic. Second generation sequencing can provide considerably more information but is still fairly expensive. In general, however, chemical and genetic methods are complimentary.

Metabolite fingerprinting is commonly used as a rapid method for high throughput qualitative screening, particularly when the primary aim is sample comparison and there is no initial interest in identifying the metabolites (Hall, 2006). Both chromatographic (LC, UHPLC, CE) and spectral (MS, NIR, Raman, UV) fingerprints have been used for this purpose. Flow injection mass spectrometry (FIMS) has been found to be particularly useful. This method provides counts vs ions with no chromatographic separation, avoids the necessity of retention time alignment for chromatograms, provides analytical results in a matter of minutes (following sample extraction), and offers the possibility of identifying discriminating masses, whether that is the main goal or not.

Chemometrically, the objective is to determine whether the fingerprint of an unknown sample matches those of the reference samples. There are numerous statistical and chemometric methods for analyzing the data patterns of the fingerprints. For authentication (or adulteration), there are two common approaches: one-class modeling (soft modeling) and classification (hard modeling) (Brerton, 2009). The chemometric methods most commonly associated with these two approaches are principal component analysis (PCA) and partial least squares-discriminant analysis (PLS-DA), respectively.

One-class modeling applies PCA to a specified class of samples. If more than one class is identified, a PCA model is applied independently to each class. Thus, there is a model for each class that is based only on the features (variables) of the samples in that class. For cases of more than one class, soft independent modeling of class analogy (SIMCA) compares the model parameters of each class and determines whether they are statistically different. An unknown sample may be found to belong to a single class, to more than one class, or to none of the classes. This approach is ideal for authentication (adulteration) because a PCA model is only needed for the authentic, or reference, samples. A one-class model applied to authentic *Echinacea purpurea* samples will allow the analyst to determine if the unknown sample is *E. purpurea* (within limits) or not (outside limits). Thus, an unknown sample of black cohosh (*Actaea racemosa*) will be classified as not *E. purpurea*.

Classification models are applied to two or more classes. The classes are identified in advance and reference samples are required for each class. The model is then constructed using all the classes, incorporating all the features (variables) of all the reference samples. An unknown sample will be assigned to one of the classes used to construct the model. It is not possible for the unknown sample to belong to more than one class or to none of them. This makes the classification approach unsuited for non-targeted analysis since it is impossible to anticipate and collect reference samples for the infinite number of possible classes. A classification model constructed using fingerprints for *E. purpurea* and *E. angustifolia* will assign an unknown sample of black cohosh to one of the two classes. Correct classification can occur only if the model was constructed including black cohosh as one of the classes.

Confidence limits can be constructed around each class of samples to help statistically evaluate their separation from other samples. However, an influence plot, the Q statistic plotted as a function of the Hotelling T^2 statistic, provides a more systematic and meaningful approach to determining the relationship between the samples (Brerton, 2009). This approach also fits well with the challenge of authentication. A PCA model is fit to the reference samples and all the samples are evaluated with respect to that model. The Hotelling T^2 statistic accounts for the variance associated with the model, the Q statistic accounts for the variance outside the model. Samples lying outside the 95% confidence limit for the Hotelling T^2 statistic have common features with the rest of the samples, but are more extreme. Samples outside the 95% limit for the Q statistic generally contain features missing from the reference samples. In general, the Q statistic takes precedence over the Hotelling T^2 statistic.

The purpose of this study was to develop a method for verifying the presence of BIs in BSs and to apply the method to a series of commercially available BSs. In this study, 16 *Echinacea purpurea* (L.) Moench aerial (EPA) samples comprised the BIs and 6 EPA single-ingredient solid and liquid supplements (EPAS and EPAL) and 13 mixed or unidentified ingredient supplements comprised the BSs (Table 1) that were analyzed by flow injection mass spectrometry (FIMS). PCA score plots were used to compare the BI and BS samples, PCA loadings and F-test spectra were used to identify discriminating ions, correlation spectra were used to identify common ions, and correlation spectra were used to verify the presence of EPA in the single-ingredient and mixed or unidentified ingredient BSs.

Table 1
Botanical ingredients and supplements.[a]

Sample Code	Genus, Species	Plant Part	Form of Ingredient or Supplement	Number of Samples
Botanical Ingredients (BI)				
EPA	*E. purpurea*	aerial	powder	16
Single Ingredient Botanical Supplements (BS)				
EPAS	*E. purpurea*	aerial	capsule	5
EPAL	*E. purpurea*	aerial	liquid	1
Mixed Ingredient/Unknown Botanical Supplements (mBS)				
EXAS	Not listed	aerial	capsule	3
EXRS	Not listed	root	capsule	3
EM1RS	*E. purpurea* & *E. angustifolia*	root	capsule	1
EM2RS	*E. purpurea* & *E. pallida*	root	capsule	1
EPML	*E. purpurea*	aerial & root	liquid	1
EM1ML	*E. purpurea* & *E. angustifolia*	aerial & root	liquid	2
EM1XL	*E. purpurea* & *E. angustifolia*	not listed	liquid	2

[a] A 4-part code was assigned to each BI and BS; genus, species, plant part, and form. All genus *Echinacea* (E). Species were either *E. angustifolia* (A), *E. purpurea* (P), *E. purpurea* and *E. angustifolia* (M1), *E. purpurea* and *E. pallida* (M2), or unknown (X). Plant parts were aerial (A), root/rhizome (R), aerial and root/rhizome (M), or unknown (X). Supplement form was liquid (L) and solid (S). Form for all ingredients were powder and left blank in the code.

2. Materials and methods

2.1. Solvents and plant material

2.1.1. Solvents

Water, acetonitrile, and methanol were of optimal grade (Fisher Scientific, Pittsburgh, PA, USA). Formic acid was of mass spectrometry grade (Sigma-Aldrich, St. Louis, MO, USA).

2.1.2. Echinacea ingredients

Sixteen *Echinacea purpurea* (L.) Moench aerial samples (coded EPA1-EPA16) were obtained from and authenticated by the Missouri Botanical Gardens (St. Louis, MO, USA) (Table 1). The vouchers for all standards are resident at the Food Composition and Method Development Laboratory, Beltsville Human Nutrition Research Center, Agricultural Research Service, United States Department of Agriculture (Beltsville, MD, USA).

2.1.3. Echinacea supplements

Thirteen liquid and solid (tablets and capsules) supplements were purchased from local drug stores in Beltsville, MD, USA (Table 1). Six were single-ingredient supplements and were labeled as containing only *E. purpurea* aerial material. The other seven were mixed or unknown *Echinacea* supplements; labeled to contain mixtures of *E. purpurea* with *Echinacea angustifolia* DC or *Echinacea pallida* (Nutt.) Nutt, mixtures of aerial and root parts, or with no information regarding the species or plant part (Table 1). A 4-part code was assigned to each BI and BS; genus, species, plant part, and form. All genus were *Echinacea* (E). Species were either *E. angustifolia* (A), *E. purpurea* (P), *E. purpurea* and *E. angustifolia* (M1), *E. purpurea* and *E. pallida* (M2), or unknown (X). Plant parts were aerial (A), root/rhizome (R), aerial and root/rhizome (M), or unknown (X). BS form was either liquid (L) or solid (S) and was left blank for BIs (powdered). Thus, EPA was an *E. Purpurea* aerial ingredient, EPAS was an *E. Purpurea* aerial solid supplement, EPAL was an *E. Purpurea* aerial liquid supplement, EM1ML was a mixed *E. Purpurea* and *E. angustifolia*, mixed aerial and root, liquid supplement. There were 6 single-ingredient supplements and 13 mixed or unknown supplements.

2.2. Sample preparation, solid samples

Roots were dried and ground, tablets were ground, and capsules were emptied and ground if necessary. One hundred mg of each dried ground sample was mixed with 5 mL of methanol: water (6:4, v:v) in a 15 mL centrifuge tube. All samples were sonicated for 20 min at room temperature. The sample extract was centrifuged at 5000*g* for 20 min (IEC Clinical Centrifuge, Danon/IEC Division, Needham, MA, USA). The supernatant was filtered through a 17 mm (0.45 µm) PVDF syringe filter (VWR Scientific, Seattle, WA, USA).

2.3. Sample preparation, liquid samples

50 µL of sample was mixed with 5 mL of methanol: water (6:4, v:v) in a 15 mL centrifuge tube. All samples were sonicated for 20 min at room temperature. The sample extract was centrifuged at 5000*g* for 20 min. The supernatant was filtered through a 17 mm (0.45 µm) PVDF syringe filter. Sample extracts were stored at 4 °C immediately and analysis was finished within 24 h of the extraction to avoid potential degradation. Each sample was analyzed three times for the FIMS experiment and all samples were run in randomized order. The injection volume was 5 µL for all samples.

2.4. Flow injection mass spectrometry

The FIMS system consisted of a LCQ DecaXP ion-trap mass spectrometer (Thermo Fisher Scientific Inc., San Jose, CA, USA) with an Agilent (Santa Clara, CA, USA) 1200 HPLC system (a binary pump with a vacuum degasser, a thermostated column compartment, an autosampler, and a diode array detector (DAD). Samples were injected through a guard column (Adsorbosphere All-Guard Cartridge, C18, 5 µm, 4.6 × 7.5 mm, Alltech Associates, Inc., Deerfield, IL, USA) to minimize potential contamination of the MS system. The mobile phases consisted of 0.1% formic acid in H_2O (phase A) and 0.1% formic acid in acetonitrile (phase B) with isocratic elution at 60:40 (v v-1) at a flow rate of 0.5 mL min-1. Electrospray ionization (ESI) was performed in the negative ion mode from *m/z* 150 to 1500 to obtain the FIMS fingerprints. The following conditions were used for the DecaXP mass spectrometer: sheath gas flow rate, 80 (arbitrary units); auxiliary gas flow rate, 10 (arbitrary units); spray voltage, 4.50 kV; heated capillary temperature, 220 °C; capillary voltage, 4.0 V; tube lens offset, 25 V. Spectra were collected over the 1.0 min interval that contained the sample bolus. Five repeat analyses of the 29 different samples provided 145 fingerprints.

2.5. Data processing for FIMS fingerprint

The mass spectrum for each sample consisted of a one-dimensional matrix (ion counts versus mass-to-charge ratio (*m/z*) from 150 to 1500. The 145 spectra acquired were exported to Excel (Microsoft, Inc., Bellevue, WA, USA) from Xcalibur for data pre-processing. The pre-processing in Excel involved combining the 145 spectra, sorting the data by sample names, deleting unnecessary and redundant information (headers, sample information, instrument information, etc.), and aligning the masses (each spectrum was a different length because not all *m/z* have ion counts in each spectrum) using a MS Excel macro written in house. The resulting two-dimensional 145 samples by 1351 *m/z* matrix was then exported to Solo (Eigenvector Research, Inc., Wehnatchee, WA, USA) for chemometric analysis. PCA (Wold et al., 1987) and SIMCA (Wold and Sjostrom, 1977) were used to analyze the data. Pre-processing consisted on normalization of the 145 spectra (sum of counts^2 = 1) and mean centering of the 1351 variables. PCA provided an initial look at sample similarity. SIMCA, used as a one-class classifier, provided a statistical evaluation of their similarity. The Q statistic was used as the primary statistic to characterize the statistical variation of the samples (Brerton, 2009).

2.6. Analysis of variance (ANOVA)

This method, as applied to full matrices, has been previously described (Harnly and Harrington, 2013). In brief, classic ANOVA is performed for every ion in the MS spectra and the resulting F-values were used to construct an F-test spectrum.

2.7. Spectral correlation

Correlation coefficients are a classic statistical method of judging the similarity of two sets of data. Perhaps the best known is Pearson's correlation coefficient (Lane, 2016). Mathematically, this is expressed as:

$$(1)\ r_{XY} = n\ (x_i\text{-}x_{ave})\ (y_i\text{-}y_{ave})/(\sqrt{S(x_i\text{-}x_{ave})^2}\ \sqrt{\Sigma(y_i\text{-}y_{ave})^2})$$

where, r_{XY} is the correlation coefficient between spectra X and Y, n is the number of spectral variables, and x and y are variables in the X and Y spectra, respectively. The numerator is the product of the 2 spectra, for mass spectrometry, a mass-by-mass multiplication. The denominator is the normalization factor that assures that the value of r_{XY} approaches 1 as the spectra approach perfect, positive correlation or -1 as they approach perfect, negative correlation. Thus, the correlation of a reference spectrum with spectra for a series of test samples will give a single value for r_{xy} varying from 1 to -1. An alternative approach is to use PCA of the denominator arrays, $n(x_i\text{-}x_{ave})(y_i\text{-}y_{ave})$, computed for each sample and the reference standard. Normalization is not necessary. Samples with similar arrays will cluster together.

3. Results and discussion

Fig. 2A shows the 2 dimensional PCA score plot (PC2 versus PC1) for the FIMS fingerprints for the 16 *Echinacea purpurea* aerial BIs, 5 *E. Purpurea* aerial solid BSs, and 1 *E. Purpurea* aerial liquid BS listed as EPA, EPAS, and EPAL in Table 1. Each sample was analyzed 5 times. Each of the 3 classes were assigned a separate symbol to facilitate recognition. It can be seen that the EPA samples are well separated from the EPAS and EPAL samples. Fig. 2B shows the same score plot in 3 dimensions to include PC3. In this case, the separation of the clusters in Fig. 2A and B are similar. However, significant differences are frequently found with higher principal components; clusters that appear to be overlapping in 2 dimensions may be found to separate in the 3rd or 4th dimension (PC3 or PC4). Higher PCs are not commonly used for fear of over fitting the data.

PCA also allows identification of ions that are primarily responsible for the separation of the clusters. Fig. 2C shows the loadings for the ions for PC1 and PC2. The further an ion is from the origin (in either the positive or negative direction), the more influence it has on the separation of the clusters. Thus *m/z* 307, 320, and 473 and *m/z* 307, 487, and 785 had the strongest influence on the horizontal and vertical dispersion of the data, respectively. The other ions had influences proportional to their distance from the origin. An alternative to the PCA loadings is to use ANOVA for each ion to compute an F-test spectrum (Harnly and Harrington, 2013). This approach is based on the signal-to-noise ratio rather than just variance, as is the case for PCA. In addition, the noise is based on the total variance of the data rather than parsing it between the PCs. Fig. 3 (negative going plot) shows the F-test spectrum for the EPA and EPAS and EPAL samples. The most influential ions were *m/z* 377, 201, 215, and 539, in descending order.

A one-class model fit to just the fingerprints of the EPA samples allows an evaluation of the statistical significance of the separation of the clusters. Fig. 2D presents the influence plot for the EPA samples shown in the PCA score plot of Fig. 2A. All the EPAS and EPAL samples fell within the 95% confidence limit for the Hotelling T^2 statistic, but were outside (above) the 95% confidence limit for the Q statistic. Thus, it is not possible to verify the label claim that *E. purpurea* aerial material was present in the EPAS and EPAL supplements, since the BSs were statistically different in chemical composition from the BIs.

If it is assumed that the labels are correct and *E. purpurea* BIs were used in the preparation of the supplements, how is it possible to verify the presence of components from the ingredients? One approach is targeted analysis of specific components that the literature has documented as being present in *E. purpurea* aerial parts. Obviously, the more components that can be targeted, the more robust the method, and the more confidence the analyst will have in the presence of *E purpurea* aerial parts. The logical extrapolation of this approach would be a non-targeted comparison of all the components in the BSs that should be extracted from the BI.

The simplest means of establishing common components in the BIs and BSs is multiplication of their spectra or chromatograms. In a correlation spectrum, common peaks will show a relative increase and peaks with no match will show a relative decrease or disappear. Fig. 3 shows the correlation spectrum for the average EPA BI fingerprints and the average BS (EPAS and EPAL) fingerprints. The strongest common ions are *m/z* 473, 311, 387, 487, and 785.

Fig. 3 compares the F-test and correlation spectra for the averaged BI fingerprints (EPA1-EPA16) and the averaged BS fingerprints (EPAS1-EPAS5 and EPAL1). Since the values in the two spectra were computed in dramatically different ways, the spectra in Fig. 3 have been normalized (the sum of squares for each point in the spectrum is equal to 1.0) for comparison. It can be seen that the correlation spectrum is less complex than the F-test spectrum with relatively fewer ions but higher counts. Conversely the F-test spectrum is more complex with more ions at lower counts. The significant ions for each spectrum (those ions providing 25% of the total counts) are listed in Table 2. More ions are necessary for the F-test spectrum to provide 25% of the total counts. Of the 60 ions listed in Table 2, only eight were found in both spectra (shaded ions). The fact that they were useful in discriminating between the BIs and BSs indicates that they were present at consistently different levels.

Fig. 4 shows the individual correlation spectrum for each of the 6 BSs, i.e. the average fingerprint for each single-ingredient supplement (EPAS1-EPAS5 and EPAL) was multiplied by the average EPA fingerprint. In addition, the average EPA fingerprint was correlated with itself (autocorrelated). It can be seen that the correlation spectrum for the EPA fingerprint is similar to the spectra for 5 of the supplements (EPAS1-EPAS4 and EPAL). Only the correlation spectrum for EPAS5 was significantly different. However, it is visually difficult to compare all the correlation spectra in detail. A PCA score plot (data not shown) of the correlation spectra showed that EPA, EPAS1-EPAS4, and EPAL fell in a single cluster while EPAS5 was considerably distant.

These results provide some interesting conclusions. First, since the correlation spectra for EPAS1-EPAS4, and EPAL were similar, the supplements contained similar extracted components. This means that the starting ingredients and the processes used by the manufacturers were similar. Second, since the autocorrelation spectrum of EPA was similar to the correlation spectra of EPA with EPAS1-EPAS4 and EPAL, the BIs in this study were similar to the BIs used by the manufacturers. Finally, since the extracted compounds were the same, the analytical extraction process used in this study was similar to the preparation method employed by the manufacturers. These data also indicate that supplement EPAS5 was produced using either a different starting ingredient or a different extraction process.

The last phase of the project was to examine the 13 BSs with mixed or unidentified BIs. According to the labels on these BSs, they contained mixtures of *E. purpurea*, *E. angustifolia*, and *E. pallida* and mixtures of aerial and root parts (Table 1). However, in 3 cases the species for the aerial parts were not listed, in 3 cases the species for the root parts were not listed, and in 2 cases the plant parts were not listed. Still, there was the possibility that each of these BSs contained *E. purpurea* aerial ingredients. PCA of the correlation spectra was used to verify the presence of EPA.

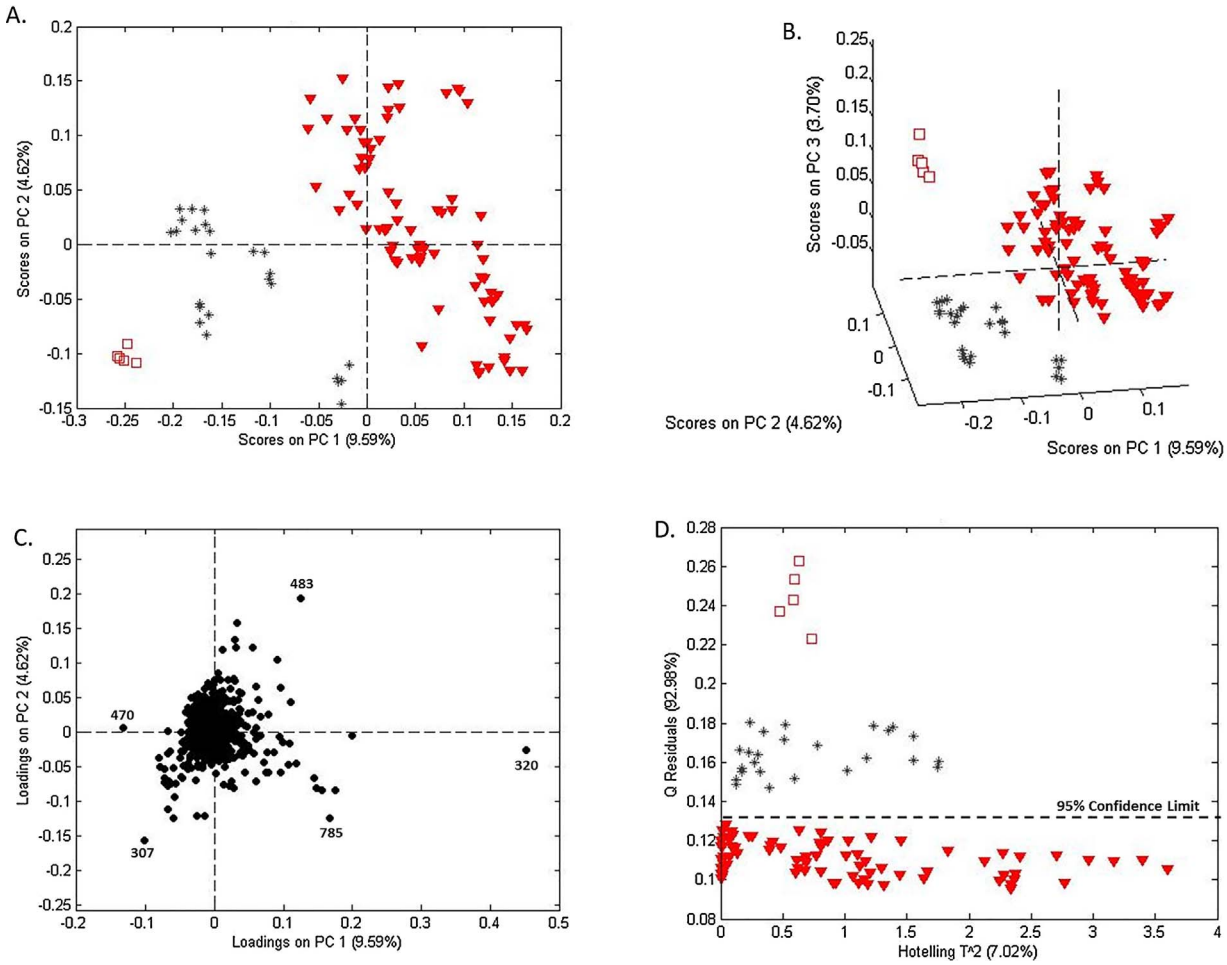

Fig. 2. Analysis of FIMS fingerprints for *E. purpurea* aerial botanical ingredients (EPA, ▼), *E. pupurea* aerial single-ingredient solid botanical supplements (EPAS, *), and *E. pupurea* aerial single-ingredient liquid botanical supplements (EPAL, □): A) 2 dimensional PCA score plot for PC1 and PC2, B) 3 dimensional PCA score plot for PC1, PC2, and PC3, C) PCA loadings for PC1 and PC2, and D) influence plot for one-class modeling of *E. purpurea* aerial botanical ingredients.

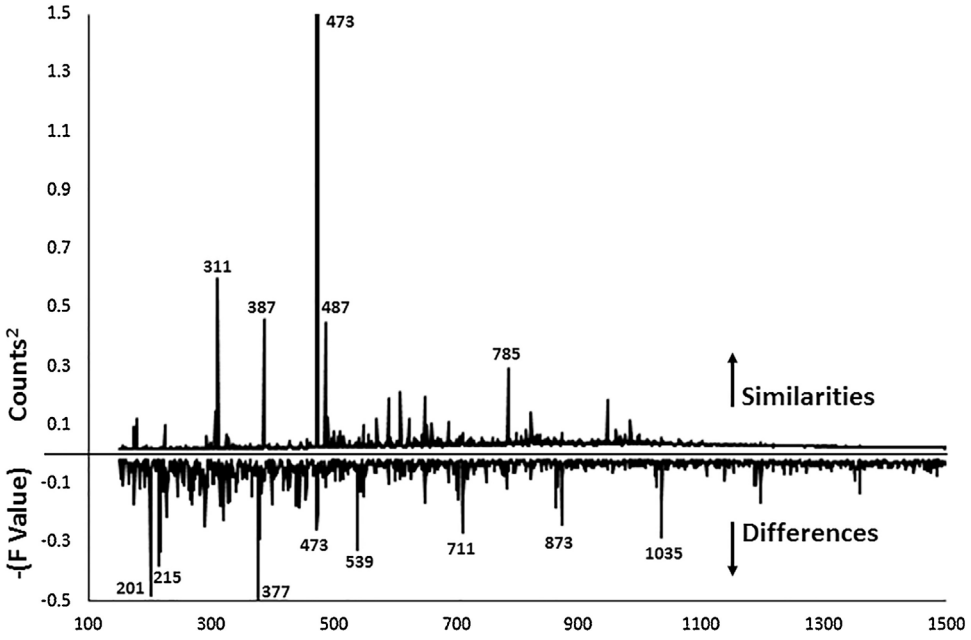

Fig. 3. F-test spectrum (lower trace, negative direction) and correlation spectrum (upper trace, positive direction) obtained using average fingerprints for *E. purpurea* aerial botanical ingredients (EPA) and average fingerprints for *E. purpurea* aerial single-ingredient botanical supplements (EPAS and EPAL).

Table 2
Comparison of Botanical Ingredient and Supplement Spectra.[a]

m/z	D	S	m/z	X	X	m/z	X	X
175	X	-	329	X	X	623	-	X
179	-	X	331	X	-	649	X	X
201	X	-	367	X	-	653	-	X
215	X	-	377	X	-	661	-	X
217	X	-	379	X	-	687	-	X
225	-	X	387	-	X	703	X	-
227	X	-	399	X	-	711	X	X
229	X	-	439	X	-	712	X	-
245	X	-	445	X	-	785	-	X
267	X	-	455	X	-	815	-	X
269	X	-	473	X	X	823	-	X
283	X	-	474	X	X	863	X	X
289	X	-	475	-	X	873	X	X
290	X	-	487	-	X	947	-	X
293	X	-	491	-	X	977	-	X
307	-	X	539	X	-	985	-	X
311	-	X	549	X	X	1035	X	X
315	X	-	571	-	X	1036	X	-
320	X	-	590	-	X	1197	X	-
321	X	-	609	-	X	1359	X	-

[a] *m/z* – mass to charge ratio, D- different as determined by an F-test, S – similar as determined by a correlation spectra.

Fig. 5A shows the PCA score plot for all the correlation spectra for the BSs, with the exception of EPAS5. The central cluster is too dense to permit labeling so only the outliers have been labeled. This problem was remedied in Fig. 5B where all the data points in the influence plot have been labeled. Fig. 5A shows that 5 of the mixed or unidentified BSs appear to not contain EPA. However, the Q statistic in Fig. 5B shows that 7 of the BSs fall above the 95% confidence limit. Thus, 7 of the BSs do not have EPA present at sufficient levels to produce a qualifying correlation spectrum. Of these 7 BSs, the labels for 2 did not specify the species and the label for a third BS did not specify the plant part. The other 4 contained either mixtures of species or plant parts.

The test described above positively verifies the presence of EPA in the 5 single-ingredient BSs and 7 mixed or unidentified BSs lying within the 95% confidence limit in Fig. 5B but does not prove the absence of EPA in the BSs lying outside the 95% confidence limit. Variations in the correlation spectra are determined by the ions present and by the counts for those ions. If the differences in the correlation spectra arise from missing ions, the probability is high that EPA is not present. In other words, key ions present in the BI fingerprint are not found in the BS fingerprint. If the difference in the correlation spectra arise from differences in the count level, EPA may still be present.

Differences in counts for individual ions in the fingerprint can occur in two different ways. First, the concentration of the BI in the BS may be low. Thus, the counts for the ions in the normalized fingerprints will be lower and will produce a correlation spectrum with EPA that is different from EPA fingerprint correlated with itself. Second, the relative counts for the ions in the fingerprint may vary. This can occur with different sources of BIs or as a result of a difference in extraction conditions. Either condition would result in a correlation spectrum that is statistically different from the autocorrelated EPA fingerprint. The solution to the latter condition with different relative counts can be minimized using binary spectra. This is done by normalizing the fingerprints, establishing an arbitrary threshold, and determining if an ion is present (1) or not (0). This is an area for future work and would address the current limitation of the method that is associated with relative ion counts.

The general principles of this method are applicable to all botanical supplements. Spectral multiplication is the easiest way to identify common components. Application of PCA to the correlated spectra eliminates the need for normalization, provides easy visual inspection of the results, allows statistical evaluation of class separation, and permits identification of key components that are the basis for discrimination.

4. Conclusions

Flow injection mass spectrometry coupled with chemometric one-

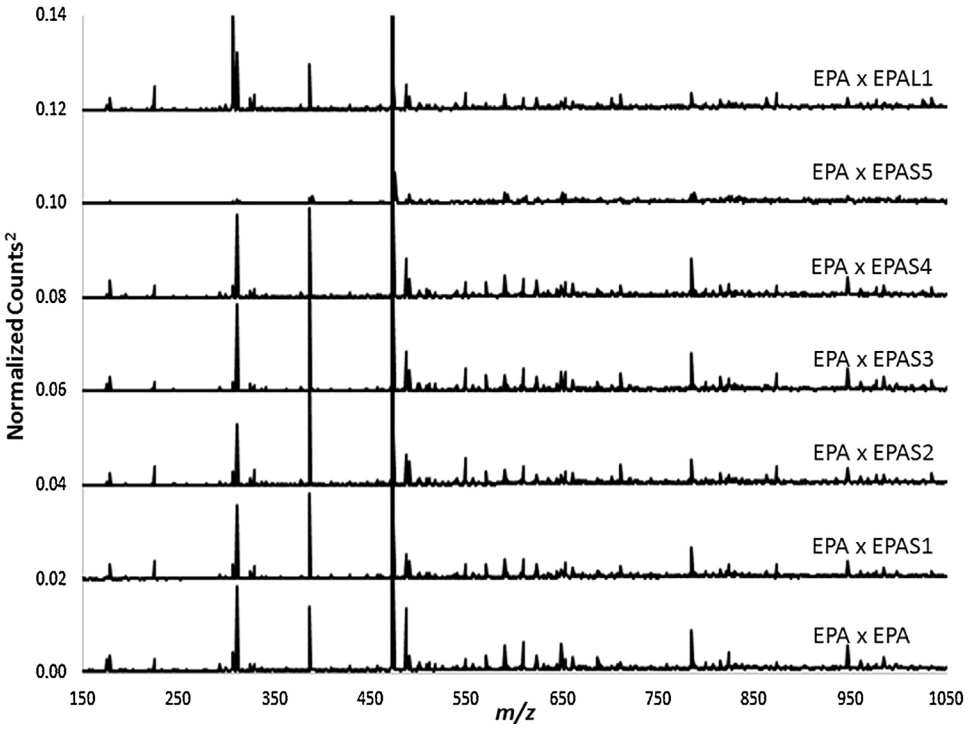

Fig. 4. Correlation spectra obtained using average fingerprints for *E. purpurea* aerial botanical ingredients (EPA) and individual fingerprints for *E. purpurea* aerial single-ingredient botanical supplements (EPAS1, EPAS2, EPAS3, EPAS4, EPAS5, and EPAL1).

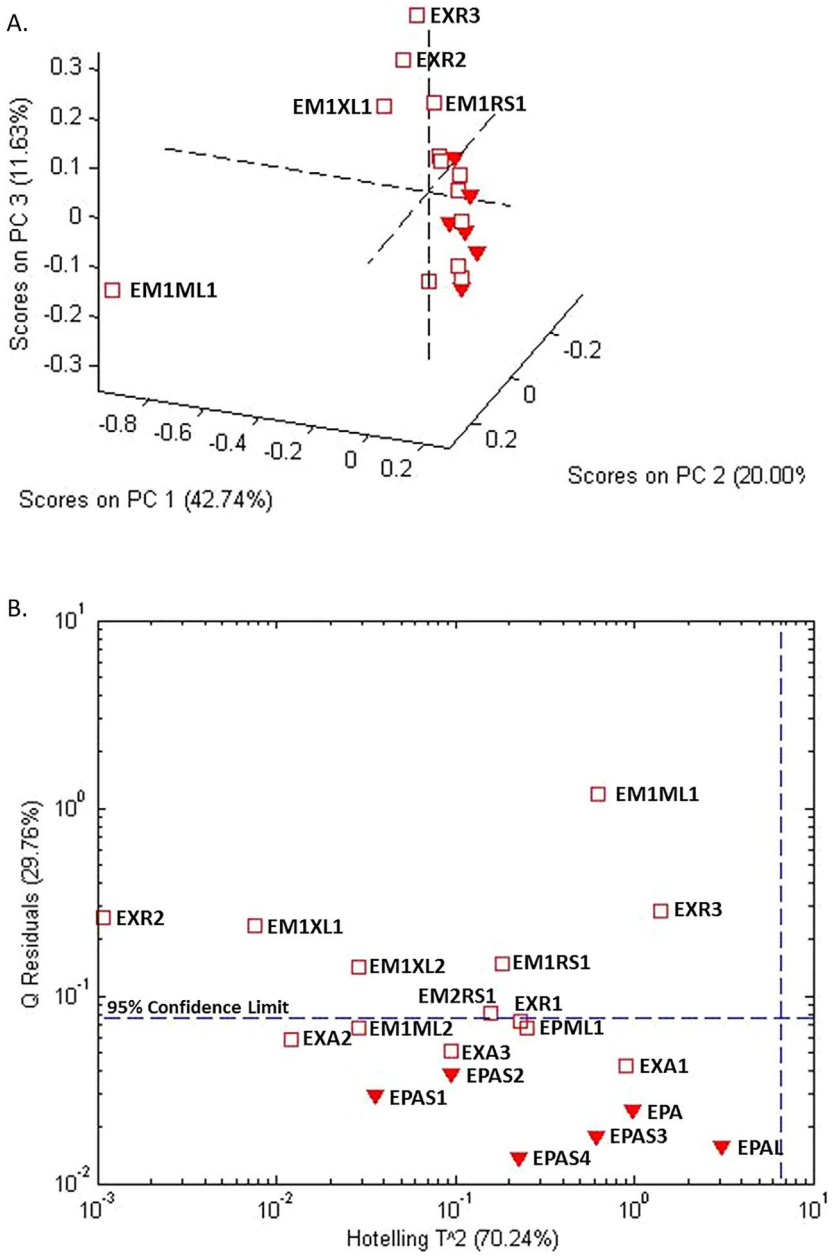

Fig. 5. Chemometric analysis of correlation spectra for single-ingredient *E. purpurea* aerial botanical supplements from Fig. 4 (▼) and correlation spectra for mixed or unknown *E. purpurea* botanical supplements (□): A) PCA score plots of correlation spectra and B) influence plot for a one-class model of the correlation spectra of single-ingredient botanical supplements materials as extracts (liquid form) or extracts added to an excipient and, the compounds of interest may be unknown, specific marker compounds may or may not exist, the concentrations of the secondary metabolites (the major reason for intake) are relatively low, and validated methods generally do not exist., differing with respect to components and relative concentrations depending on the extraction process. In addition, other ingredients may be added to the BS and components may be contributed by the excipient.

class modeling is a powerful method for comparison of botanical ingredients and commercial botanical supplements. This study demonstrated that spectral fingerprints for 16 botanical ingredients of aerial *E. Purpurea* could be used to positively verify its presence in 5 of the 6 single ingredient botanical supplements and 7 of the 13 mixed or unidentified ingredient botanical supplements. The key to the success of this method was the use of correlation spectra acquired by multiplication of fingerprints to identify ions found in both the ingredients and supplements. PCA of correlation spectra and one-class modeling were used to statistically verify the presence of compounds from the botanical ingredients in the commercial botanical supplements.

Funding

This work was supported by the Agricultural Research Service, United States Department of Agriculture [Project Number 8040-52000-063-00D] and by the Office of Dietary Supplements, National Institutes of Health, United States Department of Health and Human Services [Interagency Agreement Number AOD12026-001-01004].

References

Brerton, R.G., 2009. Chemometrics for Pattern Recognition. John Wiley and Sons, Ltd., New York, NY, USA.

Chan, T.W.D., But, P.P.H., Cheng, S.W., Kwok, I.M.Y., Lau, F.W., Xu, H.X., 2000. Differentiation and authentication of panax ginseng, panax quinquefolius, and ginseng products by using HPLC/MS. Anal. Chem. 72, 1281–1287.

Hall, R.D., 2006. Plant metabolomics: from holistic hope to hype, to hot topic. New Phytol. 169, 453–468.

Harnly, J.M., Harrington, P.B., 2013. Adulteration of American with Asian ginseng: spectral addition and experimental verification of probability of identification. J. AOAC Int. 96, 1258–1265.

Harnly, J.M., Chen, P., Sun, J., Huang, H., Colson, K.L., Yuk, J., McCoy, J.H., Harbaugh-Reynaud, D.T., Harrington, P.B., Fletcher, E.J., 2016. Comparison of flow injection MS, NMR, and DNA sequencing: methods for identification and authentication of Black Cohosh (Actaea racemosa). Planta Med. 82, 250–262.

Lane, D.M., 2016. Describing bivariate data, Computing Pearson's r. (Retrieved on August 8 from: http://onlinestatbook.com/2/describing_bivariate_data/calculation.html).

Sun, J., Chen, P., Cook, S., Jackson, G., Harnly, J.M., Harrington, P., 2013. Classification of cultivation locations of *Panax quinquefolius* L. samples using high performance liquid chromatography-electrospray ionization mass spectrometry and chemometric analysis. Anal. Chem. 84, 3628–3634.

Upton, R., 2016. Pers. Commun.

Walthall, C.L., Hatfield, J., Backlund, P., Lengnick, L., Marshall, E., Walsh, M., Adkins, S.,

Aillery, M., Ainsworth, E.A., Ammann, C., Anderson, C.J., Bartomeus, I., Baumgard, L.H., Booker, F., Bradley, B., Blumenthal, D.M., Bunce, J., Burkey, K., Dabney, S.M., Delgado, J.A., Dukes, J., Funk, A., Garrett, K., Glenn, M., Grantz, D.A., Goodrich, D., Hu, S., Izaurralde, R.C., Jones, R.A.C., Kim, S.-H., Leaky, A.D.B., Lewers, K., Mader, T.L., McClung, A., Morgan, J., Muth, D.J., Nearing, M., Oosterhuis, D.M., Ort, D., Parmesan, C., Pettigrew, W.T., Polley, W., Rader, R., Rice, C., Rivington, M., Rosskopf, E., Salas, W.A., Sollenberger, L.E., Srygley, R., Stöckle, C., Takle, E.S., Timlin, D., White, J.W., Winfree, R., Wright-Morton, L., Ziska, L.H., 2012. Climate Change and Agriculture in the United States: Effects and Adaptation. USDA Technical Bulletin 1935. National Agricultural Library, Beltsville, MD, USA.

Wold, S., Sjostrom, M., 1977. SIMCA: a method for analyzing chemical data in terms of similarity and analogy. In: Kowalski, B.R. (Ed.), Chemometrics Theory and Application, American Chemical Society Symposium Series 52. American Chemical Society, Washington D.C, USA, pp. 243–282.

Wold, S., Esbensen, K., Geladi, P., 1987. Principal component analysis. Chemom. Intell. Lab. Syst. 2, 37–52.