JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

J. Agric. Food Chem. **2007,** *55,* 4443–4452 **4443**

Midinfrared Spectroscopy for Juice Authentication—Rapid Differentiation of Commercial Juices

JIAN HE, LUIS E. RODRIGUEZ-SAONA, AND M. MONICA GIUSTI*

Department of Food Science, The Ohio State University, 2015 Fyffe Road, Columbus, Ohio 43210

The determination of food authenticity is a crucial issue for food quality and safety. Midinfrared spectroscopy provides rapid chemical profiling of agricultural products and could become an effective tool for authentication when coupled to chemometrics. This study developed a simple protocol for classifying commercial juices using attenuated total reflectance infrared spectroscopy. Spectra from a total of 52 juices together with their extracted sugar-rich and phenol-rich fractions were obtained to construct multivariate models [hierarchical cluster analysis (HCA) and soft independent modeling of class analogy (SIMCA)] for pattern recognition analysis and prediction. Spectra of the sugar-rich fraction, comprised primarily of sugars and simple acids, almost superimposed the whole juice spectra. Solid-phase extraction enriched phenol compounds and provided signature-like spectral information that substantially improved the SIMCA modeling power over the whole juice or sugar-rich fraction models and allowed for the differentiation of juices with different origins. Zero percent misclassification was achieved by the phenol-rich fraction model. HCA successfully recognized the natural grouping of juices based on ingredients similarity. The infrared technique assisted by a simple fractionation and chemometrics provided a promising analytical method for the assurance of juice quality and authenticity.

KEYWORDS: Juice; authentication; classification, phenols; anthocyanins; sugar; FTIR; solid-phase extraction; SIMCA; HCA; PCA; chemometrics

INTRODUCTION

The fruit juice industry is one of the fastest growing sectors worldwide in the beverage industry. The increasing demands from free trade, globalization, and changing technology on the agrifood industry further the drive to determine the authenticity of foods, one of the most crucial issues in food quality control and safety (*1*). The source of raw fruit material is essential to the steadily high quality of the finished product and to the compliance with labeling. For example, to be called Concord grape juice, over 95% of the juice should come from the Concord grapes and only less than 5% is allowed to come from other grapes. Similarly, in the wine industry, often times a certain wine needs to be made from a specific variety of grape with a specific geographical origin. However, partial replacement of high-cost ingredients with lower grade or cheaper substitutes like sugar solutions can be very attractive and lucrative for a fruit supplier. Raw fruit materials with different varieties and geographical origins could have greatly different prices; yet, it is hard to differentiate the source. The juice and wine manufacturers are increasingly concerned with such types of fraudulence, which not only causes economic loss but also affects the quality of the finished product and may have major health implications to consumers. These factors have underlined the need for rapid, reliable, easy-to-use, and cost-effective techniques for the industry and regulatory agencies to effectively check the authenticity of the incoming fruit material and to monitor the quality of the finished product.

Traditional methods for varietal differentiation, adulteration detection, or quality monitoring commonly rely on sensory panels or chromatographic methods to analyze for marker compounds. Sensory panels are hard to operate and very costly. Chromatography methods generate large amounts of organic solvent waste, require time-consuming analysis by trained technicians, and could be restricted by the dependence on reference compounds such as sorbitol (*2*) and certain anthocyanins (*3*, *4*), which may not be widely present in foods. Yet, the consideration of only one analytical parameter is not sufficient, and the use of combined parameters is encouraged. Physical tests such as pH and °Brix (*5*), and more sophisticated methods like stable carbon isotope ratio analysis, have also been studied (*6*). Although many methods have been investigated to examine fruit authenticity, not many are practical for routine analyses (*7*).

Currently, there is an evolving trend for the use of profiling methods combined with chemometrics for the determination of authenticity. The advantage of the profiling methods would be the evaluation of the overall components in a sample rather than looking for a single marker compound. Contemporary Fourier transform infrared (FT-IR) spectroscopy has gained the capabil-

* To whom correspondence should be addressed. Tel: 614-247-8016. Fax: 614-292-0218. E-mail: giusti.6@osu.edu.

10.1021/jf062715c CCC: $37.00
Published on Web 05/08/2007

Table 1. Name Tag and Composition As Indicated in Their Label of All of the Juices Involved in This Study

		ingredient[a]				
experiment stage	juice name tag	first abundant	second abundant	third abundant	fourth abundant	manufacturer[b]
training set	apple	apple				1
	blueberry_M	blueberry				1
	blueberry_M2	blueberry				2
	blueberry_M3	blueberry				3
	blueberry_blend	apple	blueberry			5
	cranberry_M1	cranberry				1
	cranberry_M2	cranberry				2
	cranberry_M3	cranberry				3
	cranberry_blend	cranberry	apple	grape		5
	Concord grape_M1	Concord grape				1
	Concord grape_M2	Concord grape				2
	Concord grape_M4	Concord grape				4
	Concord grape_M5	Concord grape				5
	plum_blend	plum nectar	apple			6
validation set	blueberry_M1	blueberry				1
	Concord grape_M5	Concord grape				5
	grape_blend (15%)	grape	HFCS[c]	sugar	cranberry	7
	grape_blend (100%)	white grape	cranberry	blueberry		7

[a] All of the juices in the present study, except grape_blend (15%), contain 100% juice without added sweeteners. [b] The manufacturer names are coded. [c] HFCS, high fructose corn syrup.

ity of rapidly obtaining reproducible biochemical patterns that would allow for the composition-based statistical classification. Infrared (IR) spectroscopy is ideal for rapid screening and characterization of chemical composition variation. Distinct and reproducible components exist in different commodities of fruits (*8*−*10*); thus, biochemical fingerprints may be produced by FT-IR to allow for the differentiation of subtle differences. Variation in the chemical composition attributed to variety, geographical origin, or alien ingredients might be elucidated through chemometric analysis of spectral-based grouping (*11*). Advances on instrument design and auxiliary optics as well as the development of powerful supervised chemometric software have made FT-IR spectroscopy a suitable tool for the assessment of quality and authenticity in various foods (*12*, *13*). This technique requires low sample volume and is environmental friendly. It does not require a large amount of hazardous organic solvents as liquid chromatography. Minimal or no sample preparation is required, which greatly speeds up sample analysis. Nearly real-time measurement was made possible by immediate software prediction, and once the instrument is purchased, there is minimal operational cost involved in performing the technique (*12*).

Because of the already established use of IR techniques in some food companies for quality and process control applications, the industry is familiar with the technology and the potential exists to extend its application to monitor juice quality and authenticity. IR spectroscopy is finding increased applications in the food industry (*13*), and recently, the list of investigations for authentication (*11*, *14*), classification (*15*−*17*), and product quality control purpose (*18*) was expanded to various foods, such as meat, honey (*19*), tea, apple juice (*18*), berry purees (*11*, *14*, *15*), jams (*20*), and wines (*12*), among others. However, in the beverage industry, IR technology has not been broadly applied. Insufficient sensitivity was probably a major hurdle. Kemsley and co-workers (*11*) stated that raspberry juices adulterated with 10% apple juice could be wrongly accepted as pure raspberry, and obviously, an improvement of sensitivity is necessary. Anthocyanins and other phenols in grapes (*3*, *10*, *21*, *22*), berries (*23*, *24*), and wines (*3*) had been reported to be fingerprint compounds for taxonomic purposes, but according to our preliminary study, their IR signal could be masked by other dominant compounds in the intact juices. In this study, we evaluated the intact fruit juices as well as their sugar-rich and phenol-rich fractions, in search of unique spectral information to differentiate juices or to identify potential adulterants.

Our objective was to develop a simple protocol for the rapid, high-throughput, reproducible, and sensitive analysis of fruit juices based on their unique IR spectral information. This objective was fulfilled by evaluating intact fruit juices as well as their sugar-rich and phenol-rich fractions, seeking spectral information that would allow differentiation of juices or identification of potential adulterants. The ultimate goal of our research is to provide the food industry and regulatory agencies with a protocol for the rapid and specific analysis of quality and authenticity of fruit juices of economic importance.

MATERIALS AND METHODS

Juice Samples. Twelve juices were obtained from two grocery stores, comprising three cranberry juices, two Concord grape juices, three blueberry juices, one cranberry juice blend, one blueberry juice blend, one plum nectar juice blend, and one apple juice (**Table 1**), to examine the feasibility of a FT-IR technique for juice differentiation. Additionally, three of those fruit juice commodities (cranberry, blueberry, and Concord grape), each manufactured by three companies and four different batches from each company (a total of 36 samples), were obtained from four local grocery stores (**Table 1**) to evaluate differences caused by origin/manufacturer and processing conditions. All fruit juices were labeled as "100% pure juice", except for one grape blend, which was only 15% juice. We assumed that the manufacturers had obtained genuine fruit material. For validating the calibration model, other four juices including two blends (**Table 1**) were used. The juice color, pH, and °Brix were recorded (**Table 2**). Juice color was measured on a CIE *L*C*h* color scale by a ColorQuest XE spectrophotometer (Hunter Associates Laboratory, Inc., Reston, VA), using total transmittance mode, 2 mm path length, 10° observer angel, and D65 light source. All juices were refrigerated after purchasing and analyzed promptly to minimize decomposition.

Sample Preparation. Juice samples were acidified with 0.1% HCl followed by centrifugation at 3000 rpm for 10 min to remove nonsoluble solids. Each juice was prepared as whole juice fraction, sugar-rich fraction, and phenol-rich fraction with three replications. The whole juice fraction was prepared by mixing 1 volume of supernatant juice with 2 volumes of double-distilled (DD) water containing 0.1% HCl (**Figure 1A**). The sugar-rich and phenol-rich fractions were prepared using a modified solid-phase extraction (SPE) method described by

Table 2. Sugar Content, Acidity, and Color Characteristics of 40 Juices Involved in the Batch Variance Evaluation

Training Set

		manufacturer[a]														
		1					2					3				
juice	batch	°Brix	pH	*L**[b]	*C**	*h*°	°Brix	pH	*L**	*C**	*h*°	°Brix	pH	*L**	*C**	*h*°
blueberry	batch 1	10.0[c]	3.57	49.1	46.8	15.7	10.1	3.56	67.4	28.3	33.2	9.9	3.31	72.0	27.0	46.5
	batch 2	10.0	3.67	43.4	53.7	12.2	10.1	3.84	65.1	30.5	28.5	10.0	3.35	71.7	27.0	45.9
	batch 3[d]	9.8	3.74	36.9	58.7	11.5	10.0	3.71	67.7	27.4	33.2	10.0	3.25	72.2	27.9	50.5
	batch 4	10.0	3.68	43.1	52.9	10.6	10.1	3.61	68.1	27.5	34.3	10.0	3.28	73.0	27.0	50.5
		1					2					3				
cranberry	batch 1	7.4	3.07	44.1	69.4	19.7	7.6	2.93	78.1	28.6	26.3	7.4	3.02	56.9	52.7	14.2
	batch 2	7.5	2.94	51.3	61.1	16.0	7.7	3.01	79.4	25.4	25.3	7.3	3.02	52.5	53.8	14.6
	batch 3	7.4	2.97	43.8	69.7	19.6	7.6	2.79	80.1	25.2	31.1	7.6	2.85	58.7	49.3	13.2
	batch 4	7.5	3.09	56.9	51.1	13.3	7.9	2.82	78.8	27.4	27.1	7.6	3.05	81.3	21.1	32.0
		1					2					4				
Concord grape	batch 1	14.2	4.12	65.2	31.5	38.3	16.0	3.36	72.2	29.5	46.9	16.0	3.75	51.3	36.9	21.1
	batch 2	15.4	3.86	60.6	37.2	16.2	16.0	3.59	71.2	29.8	44.9	16.0	3.83	39.8	39.3	17.2
	batch 3	15.0	4.08	64.8	32.0	34.1	16.0	3.45	71.4	30.2	46.6	16.3	3.74	41.7	38.0	20.8
	batch 4	15.2	3.98	63.4	33.6	23.4	16.0	3.40	71.8	29.7	46.3	16.2	3.91	42.6	36.8	27.0

Validation Set

		manufacturer[a]														
		4					5					7				
juice	batch	°Brix	pH	*L**	*C**	*h*°	°Brix	pH	*L**	*C**	*h*°	°Brix	pH	*L**	*C**	*h*°
Concord grape		16.1	3.76	48.5	37.4	22.7										
Concord grape							15.4	3.97	47.2	40.2	19.8					
grape_blend (15%)[e]												14.0	3.18	77.2	32.7	15.2
grape_blend (100%)[f]												14.6	3.55	85.3	13.2	27.6

[a] The manufacturer names are coded. [b] *L**, lightness; *C**, chroma; *h*°, hue angle. [c] All of the values in the table are the means of three measurements. [d] This juice appears to be an outlier as determined by PCA. Its °Brix and color characteristics are also deviated from the mean of the other three batches. [e] Grape_blend (15%) is a grape blend containing 15% fruit juice as indicated in the label, with grape juice, HFCS, sugar, cranberry juice, and synthetic dyes listed as ingredients. [f] Grape_blend (100%) is a grape blend made of 100% fruit juice as indicated in the label containing white grape juice, cranberry juice, blueberry juice, and natural flavors.

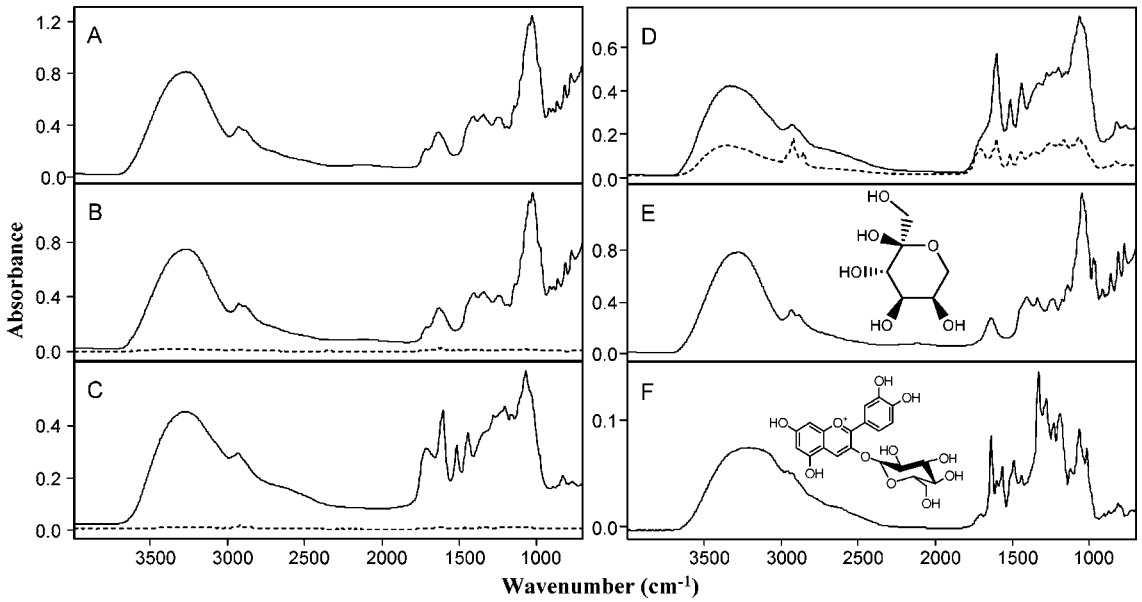

Figure 1. Mid-IR spectra of a grape whole juice (**A**), the sugar-rich fraction (**B**), the phenol-rich fraction (**C**), the anthocyanin (solid line) and nonanthocyanin (dotted line) fractions from the phenol-rich fraction (**D**), fructose (**E**), and a single anthocyanin cyanidin-3-glucoside (**F**). The dotted lines in **B** and **C** are the spectra of residues left on the C_{18} cartridge after the elution of interested compounds.

Giusti and co-workers (*25*). Briefly, 5 mL of juice was loaded onto a Sep-Pak Vac (6 cm^3, 1 g) C_{18} cartridge (Waters, Milford, MA) that had been preconditioned with 10 mL of acidified methanol (0.1% HCl) and 10 mL of acidified DD water (0.1% HCl) and then washed with 5 mL of acidified DD water. This aqueous eluate was collected and constituted the sugar-rich fraction containing polar compounds such as sugars and simple acids (**Figure 1B**). The column was further cleaned by passing 5 mL of DD water. Complete recovery of polar compounds

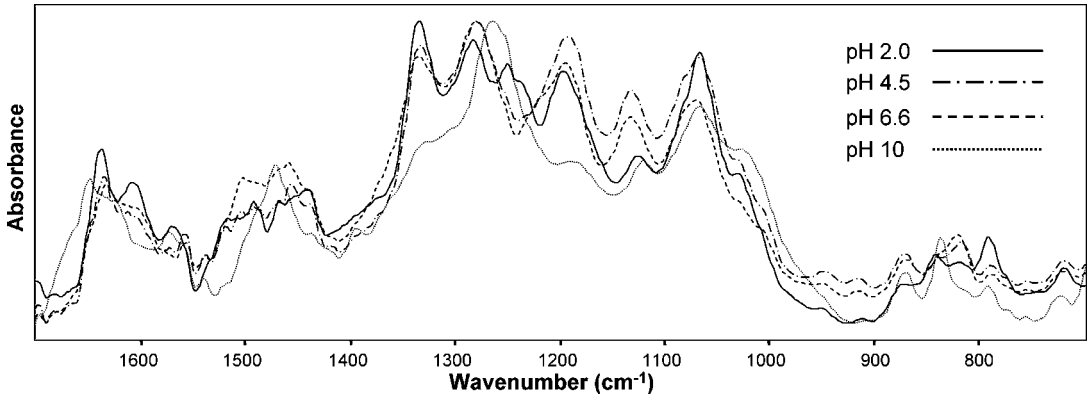

Figure 2. Effect of pH on the IR profile of a purified berry anthocyanin mixture. The absorbance has been normalized to the most intense band for each spectrum.

was confirmed by monitoring the IR signal of the succeedingly eluted water fraction (**Figure 1B**). The phenol-rich fraction, which contained primarily anthocyanins, other flavonoids, and cinnamic acids, was finally eluted from the column with 5 mL of acidified methanol (**Figure 1C**). Complete recovery was confirmed by monitoring the IR signal of the succeedingly eluted methanol fraction (**Figure 1C**). An aliquot of the phenol-rich fraction was dried under nitrogen flow and reconstituted by adding 500 μL of 30% methanol containing 0.1% HCl. Standards (all reagent grade) of sucrose, glucose (Acros, NJ), and fructose (Sigma, St. Louis, MO) were dissolved in DD water (0.10 g/mL with 0.1% HCl) to approximately 10 °Brix, which is similar to the °Brix level of many juices. Phenol standards including cyanidin-3-glucoside (Polyphenols, Norway) and gallic acid (MP Biomedicals, Aurora, OH) were dissolved in 0.1% HCl of acidified methanol (70% v/v) to attain various concentrations ranging from 0.25 to 1 mg/mL. The phenol-rich fraction was further separated into an anthocyanin fraction and nonanthocyanin fraction using the modified method from Giusti and co-workers (*25*) to evaluate the role of anthocyanins. A purified chokeberry anthocyanin fraction (with <5% impurity determined by chromatography peak area under the 254−700 nm max plot) was dissolved in aqueous methanol (50% v/v) and titrated with HCl or KOH to pH 2, 4.5, 6.6, and 10 to study the pH effect on IR profile. All solutions were stored at refrigerated conditions before analysis.

Spectra Acquisition. A Digilab Excalibur 3500 FT-IR spectrometer (Digilab, Randolph, MA), equipped with a potassium bromide beam splitter, a deuterated triglycine sulfate detector, and a horizontal attenuated total reflectance (ATR) accessory (Pike Technologies, Madison, WI) was used to collect all spectra, operating at 4 cm^{-1} resolution to improve the signal-to-noise ratio. Ten microliters of whole juice/sugar-rich sample or 15 μL of phenol-rich sample was deposited onto the amorphous material transmitting infrared radiation (AMTIR) crystal (Pike Technologies) and allowed to evaporate in a vacuum chamber for 5 min until a desolvated homogeneous film was obtained. Spectra were collected over the frequency region from 4000 to 700 cm^{-1}. Interferograms of 128 scans were coadded followed by Beer−Norton apodization. The absorbance spectrum was ratioed against the blank AMTIR crystal spectrum (background). Spectra were displayed in terms of absorbance as calculated from the transmittance spectra using the Win infrared Pro software (version 3.42, 2003, Digilab, Inc., Randolph, MA). To minimize the interference absorbance from moisture and CO_2, the instrument was continuously purged with CO_2-free dry air from a CO2RP140 CO_2 removal purifier (Dominick Hunter Ltd., Charlotte, NC). Each juice was fractionated three times to account for sample preparation variation, and each fraction was measured in triplicate to account for instrument reading variation.

Chemometrics. Multivariate statistical software collected the information from over hundreds of spectra and further conducted data reduction analysis to construct statistic models.

Data Preprocessing. The spectra were imported as .spc files into multivariate statistics program Pirouette 3.11 (Infometrix Inc., Woodinville, WA). Second derivative transformation and normalization were performed on mean-centered data using a five-point polynomial-fit Savitzky−Golay function. Normalization and second derivatives were used to mitigate the influence of signal intensity variation.

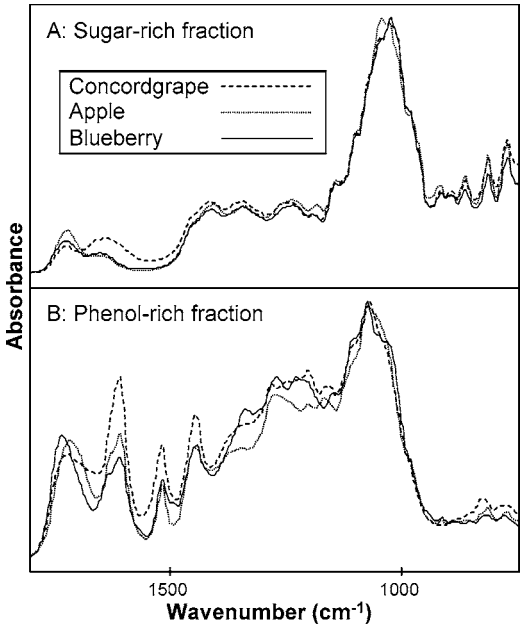

Figure 3. Mid-IR spectra (1800 to 750 cm^{-1}) of the sugar-rich fraction (**A**) and the phenol-rich fraction (**B**) of blueberry, Concord grape, and apple juices. The absorbance has been normalized to the most intense band for each spectrum.

Supervised Clustering. The data reduction technique, known as principal component analysis (PCA), extracted important information from high-dimensional IR spectra data in the training set by generating principal components (PCs). A PCA-based pattern recognition method, soft independent modeling of class analogy (SIMCA), was used to construct calibration models for predicting the identity of future unknown samples (*26*). An important characteristic of SIMCA was the unlimited number of measurement variables allowed. For spectral data, often times, the variables far exceeded the number of samples; thus, the collinearity problem prevented many standard discrimination techniques from working properly (*27*). A global PCA using NIPALS (nonlinear iterative partial least-squares) algorithm was conducted prior to SIMCA to identify outliers by means of Mahalanobis distance and sample residual, as well as to optimize the selection of PCs (*26*). SIMCA was subsequently performed on mean-centered data of all juice samples, each fraction separately, and "leave-one-out" cross-validation was used to determine the dimension of local models (*28*). Predetermined (supervised) classes and calculated 95% confidence intervals for all classes were projected onto a three-dimensional (3D) PCA scores plot using the first three PCs as axes, so the separation of classes could be

easily perceived with respect to the spectral variance captured. Identities of unknown observations were predicted on the basis of their best fit to the model.

Unsupervised Clustering. The agglomerative type hierarchical cluster analysis (HCA) provided an easy visualization, known as the dendrogram, of the relationships among different juices with respect to the spectra (*29*). The analysis involved the progressive pairwise comparison of relative similarity and step-by-step fusing of clusters (*30*). HCA was performed on mean-centered data of all juice samples, each fraction separately, using the centroid linkage method.

RESULTS AND DISCUSSION

Fractionation Procedure and ATR-FTIR Measurement Optimization. The IR spectra of intact juices were largely attributed to their sugar composition (**Figure 1A**,**B**). A SPE fractionation procedure involving the use of C_{18} cartridges offered good separation of phenols from more hydrophilic compounds such as sugars and simple acids, making it possible to obtain reproducible IR spectra of the juice phenols from a concentrated phenol-rich fraction.

Anthocyanins at pH below 2 exist primarily as the stable flavylium cation form. However, at higher pH, protonation is lost and other forms such as the quinoidal base and the carbinol pseudobase forms predominate. Spectra of anthocyanin standards at various pH conditions revealed that the IR spectral profile changed substantially (**Figure 2**). Deprotonated organic acids have also been shown to yield different IR spectra as compared to the protonated molecules (*31*). To ensure that anthocyanins are in uniform and stable form and that other phenols are in the protonated form, 0.1% HCl was present in the phenol extract to keep the pH below 2. The use of a pH-resistant AMTIR crystal and a three-reflection ATR accessory, which allows greater IR absorbance as compared to the single reflection crystal, enabled the collection of high-quality spectra (*32*). The high IR signals combined with the low noise level attained on the AMTIR crystal were advantageous for the measurement of low concentration phenol compounds in the juices.

The absorbance of methanol and water was strong in the mid-IR range; thus, the solvents were removed before the FT-IR measurements by drying the samples under vacuum. The use of a water/methanol mixture facilitated the generation of a homogeneous film on the AMTIR crystal within 5 min. Varying the percentage of water in a range of 25−100% only slightly affected the anthocyanin or sugar spectra, ensuring reproducibility to the models.

Mid-IR Spectra. Mid-IR spectroscopy (4000 to 700 cm^{-1}) has much to offer the analyst because specific bands may be assigned to specific chemical entities and provide fairly narrow bands arising from functional group vibrations with known assignment in most cases. **Figure 1A**−**C** shows the IR spectra of a Concord grape juice and its sugar and phenol fractions. A predominant band in the region of 1000−1160 cm^{-1} was assigned to the C−O stretching vibration of the sugars in the sugar-rich fraction (**Figure 1B**) and whole juice (**Figure 1A**) (*15*, *18*). In the phenol-rich fraction, a similar band was assigned to the sugar moiety of glycosides and probably to some extent to the aromatic C−O stretching (**Figure 1C**). In all samples, the region at 2500−3600 cm^{-1} was dominated by one broad band, which represented the H-bonded O−H stretching of carbohydrate, carboxylic acids, and residual water. The doublets near 2900 cm^{-1} were caused by the C−H asymmetric and symmetric stretching of the methyl C−H group (*33*). In the whole juice (**Figure 1A**) and sugar-rich fraction (**Figure 1B**) spectra, the band at 1630 cm^{-1} had slight variation across replications, and the height was positively associated with the height of the 2500−3600 cm^{-1} band, indicating the involvement of H−O−H bending of crystal/residual water. The H-bonded OH group of carbohydrates may also contribute to this band. A band at 1710 cm^{-1} was assigned to the C═O stretching of aldehyde/ketone groups in carbohydrates. In the phenol-rich

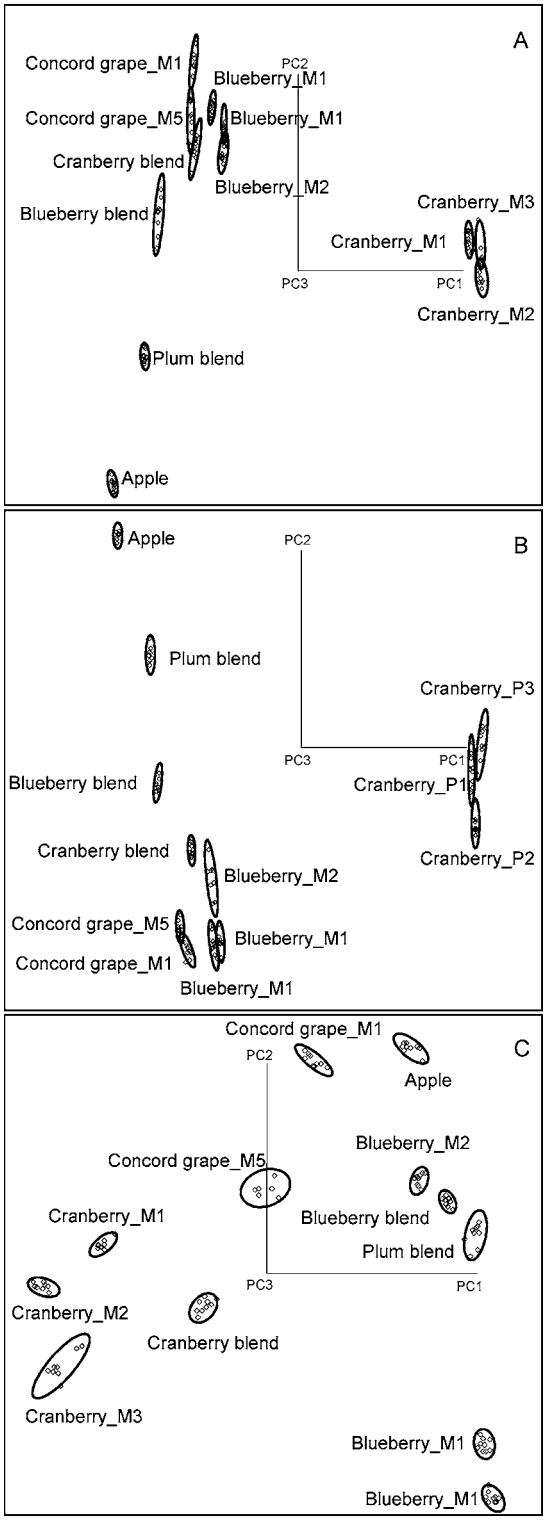

Figure 4. SIMCA 3D class projections plots based on the 1800 to 750 cm^{-1} region of the whole juice (**A**), the sugar-rich fraction (**B**), and the phenol-rich fraction (**C**) spectral data to examine the spectra reproducibility. Three axes represent the first three PCs in the corresponding group. The ovals represent the 95% confidence intervals of each cluster. Only one angle is shown to illustrate the clusters. Abbreviations are described in **Table 1**.

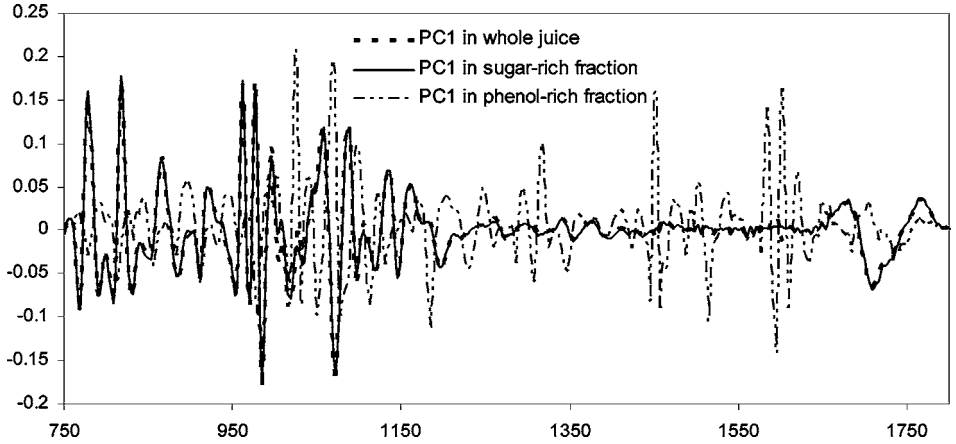

Figure 5. Loadings plot of PC1 for the whole juice, sugar-rich fraction, and phenol-rich fraction models.

fraction (**Figure 1C**), the band at 1720 cm^{-1} was assigned to the carbonyl C═O stretching of protonated carboxylic acids (*34*, *35*). Because berries usually contain monomeric anthocyanins that have no carbonyl or ketone groups, the band ~1720 cm^{-1} usually had limited absorbance. This was illustrated by the spectrum of cyanidin-3-glucoside, an anthocyanin widely distributed in berries (**Figure 1F**). The three bands between 1445 and 1610 cm^{-1} were attributed to aromatic ring vibrations (*33*, *36*, *37*). Multiple bands between 1150 and 1400 cm^{-1} comprised intricate absorption of the C−O stretch and C−O−H bending of phenols, carboxylic acids, and carbohydrates (*35*).

The IR spectrum of the whole juice was overwhelmed by the signal from sugars (**Figure 1B**). The sugar-rich fractions from all juices except cranberry juices were visually similar to fructose (**Figure 1E**) or glucose with respect to their major bands. As compared to the huge amount of sugars, organic acids such as tartaric, malic, citric, and quinic acids provided visually little contribution to the IR absorbance except in cranberry where the high acid content constituted a major band in the whole juice and sugar-rich fraction profiles at ~1720 cm^{-1} (C═O stretching). The signature-like phenol compounds showed more pronounced spectral differences among juices (**Figure 3B**), paving the foundation for sensitive classification. Anthocyanins, the most prevalent phenols in berries and grapes, only represent up to 500 mg/100 g (*38*), whereas grape juices generally contain ~15 °Brix sugars (15% sugar), 300−900 mg tartaric acid/100 g, and 150−500 mg malic acid/100 g (*39*). Thus, extraction was necessary for successful modeling based on phenols. Further separation of the phenols into anthocyanin fraction and nonanthocyanin fraction revealed that anthocyanins were the major components contributing to the phenol-rich fraction spectra (**Figure 1C**,**D**).

Supervised Clustering of Juices. Spectra (*108*) from each fraction (whole juice, sugar-rich, and phenol-rich) were collected from 12 juices. At this stage, replication was restricted at the sample preparation and data acquisition levels to challenge the capability of the FT-IR method in detecting any difference of juice components. SIMCA using the optimized 1800 to 750 cm^{-1} spectral range generated a 3D PCA scores plot (**Figure 4**) that helped to visualize class separation among samples. The boundary ellipse around each cluster visualizes the 95% confidence interval, and each data point in the cluster represents one juice spectrum. Excellent cluster separation among all juice samples was achieved in our data set. The fractionation step enabled resolving the phenolic signal that was initially masked in the whole juices (**Figure 4C**). The 3D plots based on whole juice and sugar-rich fraction samples showed comparable clustering patterns (**Figure 4A**,**B**). Combining their visual

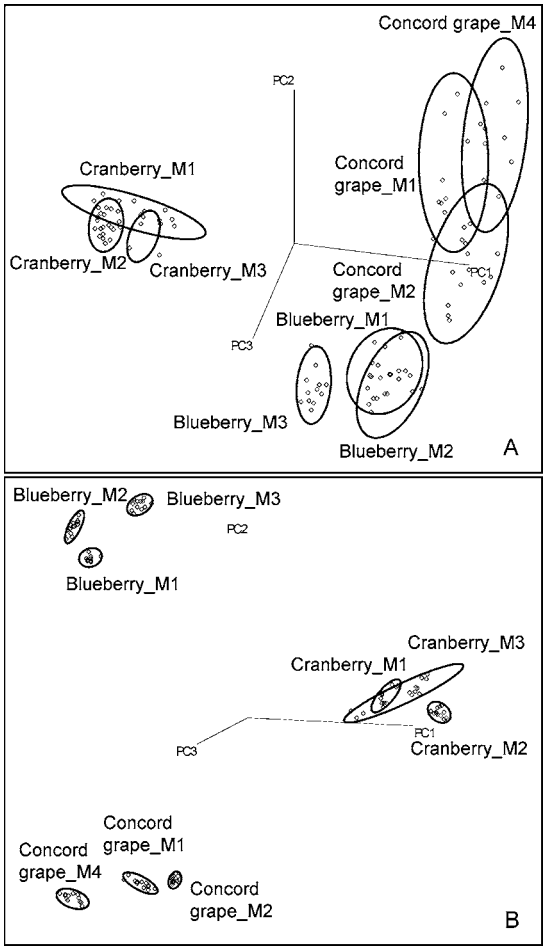

Figure 6. SIMCA 3D class projections plots based on the 1800 to 700 cm^{-1} region of the whole juice (**A**) and the phenol-rich fraction (**B**) spectral data to examine the batch to batch variation. Three axes represent the first three PCs in the corresponding group. The ovals represent the 95% confidence intervals of each cluster. Only one angle is shown to illustrate the clusters. Abbreviations are described in **Table 1**.

spectral similarity (**Figure 1A**,**B**) and similar loadings plots of PCs (**Figure 5**), it was apparent that the spectra from the sugar-rich fraction contained nearly identical discriminating information to that of the whole juices spectra. Therefore, in the later study, we only built calibration models based on the whole juice and the phenol-rich fraction.

The results showed that our sample preparation and data acquisition methods generated reproducible spectra for the

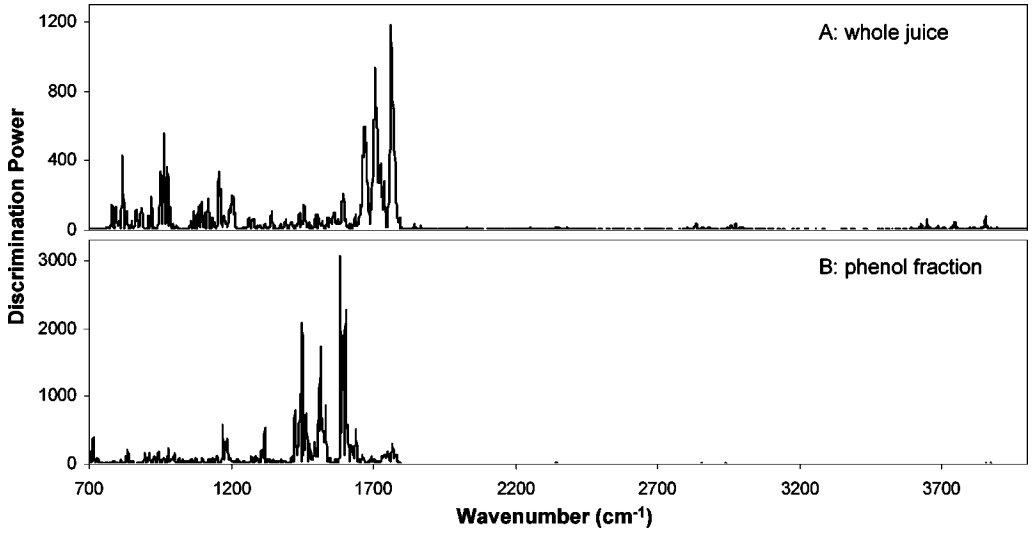

Figure 7. Discriminating power contributed by the independent variables in the whole juice spectra (**A**) and the phenol-rich fraction spectra (**B**).

Table 3. Interclass Distances in the Whole Juice SIMCA Model and Phenol-Rich Fraction SIMCA Model Built on the Training Set Spectra

			blueberry			cranberry			Concord grape		
class	no. of PCs retained	fraction	_M2	_M3	_M1	_M3	_M1	_M2	_M1	_M2	_M4
blueberry_M2	4	whole	0								
	4	phenol	0								
blueberry_M3	4	whole	1.06	0							
	5	phenol	2.26	0							
blueberry_M1	4	whole	0.33	0.97	0						
	2	phenol	3.17	1.89	0						
cranberry_M3	4	whole	5.58	4.15	5.14	0					
	3	phenol	7.08	5.23	6.10	0					
cranberry_M1	3	whole	5.70	4.38	5.23	0.25	0				
	5	phenol	7.98	5.88	6.52	1.29	0				
cranberry_M2	4	whole	6.47	4.87	6.07	0.67	0.66	0			
	5	phenol	9.57	6.68	8.25	3.79	4.12	0			
Concord grape_M1	4	whole	1.10	2.06	1.34	6.27	6.40	7.11	0		
	5	phenol	6.90	4.90	5.14	6.45	7.55	9.14	0		
Concord grape_M2	4	whole	0.91	2.19	1.20	6.61	6.65	7.51	0.83	0	
	4	phenol	7.30	5.31	6.54	7.36	8.88	9.97	2.39	0	
Concord grape_M4	3	whole	1.67	2.80	1.87	7.13	7.27	8.02	0.91	0.87	0
	5	phenol	6.14	5.64	5.85	6.87	7.98	10.42	3.52	4.26	0

detection of compositional difference among different juices; however, the variance due to juice compositional deviation had not been determined. To incorporate batch-to-batch variance and to increase the robustness of the models, 36 additional juices (made from three different commodities) were evaluated. Because the variation due to spectral data acquisition was determined to be negligible, spectral reading replication was excluded. A global PCA determined three obvious outlier spectra beyond the 95% Mahalanobis distance threshold. Those three spectra were from the same blueberry juice (manufacturer 1, batch 3) whose °Brix, pH, and color characteristics were apparently different than the other batches (**Table 2**). These differences were attributed to improper handling or abused juice storage. Therefore, the SIMCA model for each fraction was built on 105 spectra (**Figure 6**). Overall, 90% of the total variance (sum of eigenvalues/total variance) was explained by four PCs for the whole juice or five PCs for the phenol-rich fraction. To avoid overfitting, four or five maximum factors were used in SIMCA for whole juice or phenol-rich fractions, respectively (*40*).

Class projections of the spectra collected from the whole juice (**Figure 6A**) showed clusters that came very close or overlapped with others while improved cluster separation was observed for the phenol fraction (**Figure 6B**), suggesting that the phenol-rich fraction incorporated sufficient signature differences among samples. It was notable that across batches the anthocyanin level could vary substantially as demonstrated by the variability on color characteristics (**Table 2**), but the batch variation was insignificant in the IR spectral model. On the contrary, the sugar level was fairly constant across batches; yet, the batch variation was remarkable in the IR spectral model. We speculate that the signature information contained in the phenols contributed to large interclass variance (modeled variance) and levered down the importance of intraclass variance (residual variance). In the matrix of class distance (**Table 3**), which is a measure of interclass to intraclass distance ratio for the samples, the class distance was smaller in the whole juice model, showing the lack of modeling power. As a rule of thumb, a class distance over three indicates well-separated classes (*27*). The phenol-rich fraction model showed larger class distances between almost any pair of classes, suggesting an enhanced potential to discriminate subtle differences. The cross-validation process determined 0% misclassification in both the whole juice and the phenol-rich fraction models; yet, the latter performed much better for predicting the external validation data set, as will be discussed later.

Table 4. Best Prediction and the Next Best Predictions for the Identities of Validation Set Samples in the SIMCA Models[a]

	prediction based on different fractions[c]						
	whole juice						phenol fraction
		next best					
juices[b]	best	1	2	3	4		best
Concord grape_M4	Concord grape_M2	Concord grape_M1	blueberry_M2	Concord grape_M4	blueberry_M1		Concord grape_M4
	Concord grape_M2	Concord grape_M1	blueberry_M2	Concord grape_M4	blueberry_M1		Concord grape_M4
	Concord grape_M2	Concord grape_M1	blueberry_M2	Concord grape_M4	blueberry_M1		Concord grape_M4
Concord grape_M5	Concord grape_M1	Concord grape_M4	Concord grape_M2	blueberry_M2	N/A[d]		Concord grape_M1
	Concord grape_M1	Concord grape_M4	Concord grape_M2	blueberry_M2	blueberry_M1		Concord grape_M1
	Concord grape_M1	Concord grape_M4	Concord grape_M2	blueberry_M2	N/A		Concord grape_M1
grape_blend (15%)	blueberry_M1	N/A	N/A	N/A	N/A		N/A
	N/A	N/A	N/A	N/A	N/A		N/A
	blueberry_M1	N/A	N/A	N/A	N/A		N/A
grape_blend (100%)	Concord grape_M2	Concord grape_M4	Concord grape_M1	blueberry_M2	N/A		N/A
	Concord grape_M2	Concord grape_M1	Concord grape_M4	N/A	N/A		N/A
	Concord grape_M2	Concord grape_M1	blueberry_M2	Concord grape_M4	N/A		N/A

[a] The "next best" prediction usually occurs when the boundaries of clusters approximate each other or even overlap. [b] Refer to the compositional details in **Table 1**. [c] Each juice in the validation set was predicted with three replicated spectra. [d] N/A means no matching to established category.

The discrimination power revealed that most of the variance among juice samples was explained in the 1800 to 700 cm^{-1} range for both the whole juice (**Figure 7A**) and the phenol-rich fraction spectra (**Figure 7B**). This region contained unique information from the functional groups of organic substances (*33*). A primary discriminating factor in the sugar-rich profile (**Figure 7A**) came from the C═O stretching band of aldehyde/ketone groups at around 1705 cm^{-1} and to a less extent from the region around 1670 cm^{-1} associated with quinone or conjugated ketone, as well as around 962 cm^{-1} associated with pyranose ring vibration. The discriminating factors in phenol profile (**Figure 7B**) were attributed mostly to the aromatic structure (1584 and 1450 cm^{-1}) and second to the C−O−H structure (1317 and 1171 cm^{-1}) of phenolic −OH groups and sugar moieties. The importance of such regions agreed with the fact that different anthocyanins have varied numbers of −OH and methyl groups, and organic acids acylation (*p*-coumaric acid as in grape anthocyanins) besides the common aromatic backbone (**Figure 1F**).

Unsupervised Clustering. Unlike SIMCA, which is a supervised pattern recognition approach, HCA is an unsupervised clustering method used to visualize the natural grouping of the juice spectra (1800 to 700 cm^{-1}), based on the Euclidean distance calculated from the original variables (*41*). The largest Euclidean distance among any two spectra in the multivariate space is equivalent to a similarity level of 0, whereas a value of 1 indicates identical spectra (*42*). Similar spectra would usually be clustered close together with high similarity level. The sugar-rich fraction dendrogram (**Figure 8A**) shows that 36 samples were grouped into three distinct classes: blueberry, Concord grape, and cranberry. The cranberry cluster possessed the lowest similarity with other clusters, probably due to the exceptionally high quinic and other acids content in cranberry, which turned out to produce a uniquely strong C═O stretching band at ∼1720 cm^{-1} in the whole juice and sugar-rich fraction spectra. However, within each type of juice, there was a lack of sensitivity to differentiate the manufacturer, due to the similar sugar profile within commodity (*43*). In the phenol-rich fraction dendrogram, the similarity values at batch level were generally above 0.7, the similarity values at the manufacturer level were generally above 0.4, and the similarity values among different types of juices were around 0.1 (**Figure 8B**). With the distinct similarity levels, juices were grouped into naturally related clusters and even the manufacturer of juice was distinguished, again indicating the potential to differentiate varietal or processing condition differences.

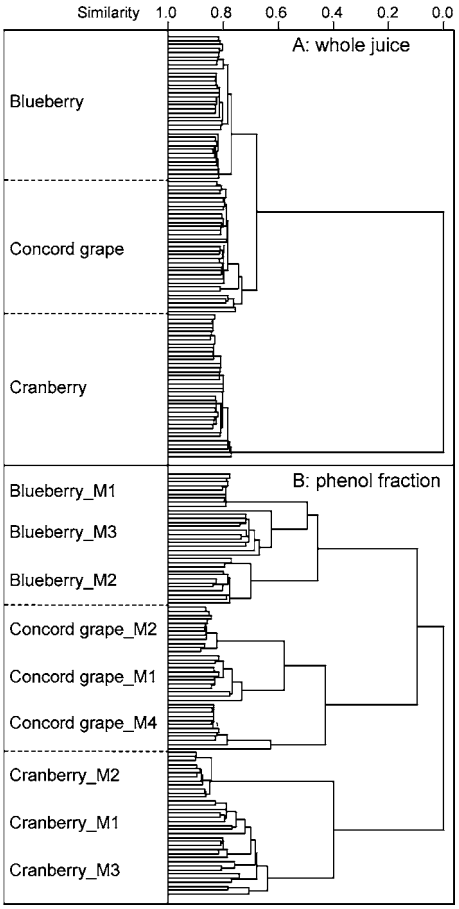

Figure 8. HCA dendrograms based on the 1800 to 700 cm^{-1} region of the whole juice fraction (**A**) and phenol-rich fraction (**B**). Each horizontal stick represents a spectrum. Each vertical stick indicates the similarity level as projected onto the *X*-axis. Misclassification at the manufacturer level frequently occurred in the whole juice fraction model; thus, clusters could only be labeled for each type of fruit juice.

The phenol-rich fraction also contained useful information to associate the juice blends to their ingredients (**Table 1**). When juice blends were clustered with the training set data (data not shown), the cranberry juice blend was grouped close to cranberry

juices, the Grape_blend (100%) containing mainly grape juice was clustered close to Concord grape juices, and the blueberry juice blend was positioned between the apple and the pure blueberry juices, which were the first and second abundant ingredients in this juice blend, respectively. These observations pointed out the potential of using multivariate regression models to predict the concentration of juice ingredients.

Prediction of Unknown Samples. Four juices were used as an external validation set (**Table 2**) to evaluate the prediction performance of the SIMCA models built on the training set (*28*). An unknown data was only assigned to the established class for which it had a high probability (*27*). Any new point falling outside of the 95% confidence interval boundary of an established class would have a standard deviation exceeding the upper limit for the residual standard deviation (RSD), which was calculated by multiplying the mean RSD with the square root of a critical *F* value, and would be rejected to the identity of that class (*44*). With respect to prediction, an important advantage of SIMCA over hard modeling is the ability not only to determine whether a sample does belong to any of the predefined categories but also to determine if it does not belong to any class. Class prediction in SIMCA could fall into three possible outcomes: (i) the sample was properly classified into one of the predefined categories, (ii) the sample did not fit any of the categories, and (iii) the sample properly fit into more than one category (*26*). On the basis of the phenol-rich fraction model (**Table 4**), 100% correct classification was achieved at the commodity level and neither of the juice blends was assigned as any of the pure juices in the model, supporting the excellent prediction power of the phenol-rich fraction model. It should be noted that no blends were used in developing the calibration model. On the basis of the whole juice model, 100% correct classification was achieved for pure juices at the commodity level. However, five out of the six juice blends were incorrectly assigned the identities of pure juices. It is evident that the phenol-rich fraction model effectively improved the prediction power for unknown samples. We expect that the improvement of modeling power would enhance the detection limit for adulteration or quality deviation.

Because phenols are almost universally present in any plant material and are easily extractable, authentication of fruit material using the phenol-rich fraction is expected to be valuable for various applications. The sample preparation procedure could be simplified to shorten the fractionation procedure by directly analyzing the SPE eluate without losing useful information or reproducibility, giving even higher efficiency to the method. Caution must be taken for possible adulteration by phenol-free material, such as sugar solution. Therefore, we recommend that both the whole juice and the phenol-rich fraction be analyzed and their data combined for more robust modeling and improved screening for juice authentication. Ideally, potential adulterants should be included in constructing the model for increased prediction power. Our constructed statistical models demonstrated the potential to provide the juice/wine industry with a rapid and reliable tool for checking the authenticity of incoming materials and monitoring the finished product quality. Once a broad range of genuine samples are collected, especially to involve the juice of interest and the potential adulterants, the model could become robust and able to identify deviated samples.

ABBREVIATIONS USED

IR, infrared; FT-IR, Fourier transform infrared; NIPALS, nonlinear iterative partial least-squares; ATR, attenuated total reflectance; AMTIR, amorphous material transmitting infrared radiation; SPE, solid-phase extraction; DD, double distilled; HCA, hierarchical cluster analysis; PC, principal components; PCA, principal component analysis; SIMCA, soft independent modeling of class analogy; RSD, residual standard deviation.

LITERATURE CITED

(1) Anderson, K. A.; Smith, B. W. Chemical profiling to differentiate geographic growing origins of coffee. *J. Agric. Food Chem.* **2002**, *50*, 2068−2075.

(2) Wrolstad, R. E.; Shallenberger, R. S. Free sugars and sorbitol in fruits—A complication from the literature. *J. Assoc. Off. Anal. Chem.* **1981**, *64*, 91−103.

(3) Revilla, E.; Garcia-Beneytez, E.; Cabello, F.; Marti-Ortega, G.; Ryan, J. M. Value of high-performance liquid chromatographic analysis of anthocyanins in the differentiation of red grape cultivars and red wines made from them. *J. Chromatogr. A* **2001**, *915* (1−2), 53−60.

(4) Chaovanalikit, A.; Thompson, M. M.; Wrolstad, R. E. Characterization and quantification of anthocyanins and polyphenolics in blue honeysuckle (*Lonicera caerulea* L.). *J. Agric. Food Chem.* **2004**, *52* (4), 848−852.

(5) Robards, K.; Antolovich, M. Methods for assessing the authenticity of orange juice. A review. *Analyst* **1995**, *120*, 1−28.

(6) Antolovich, M.; Li, X.; Robards, K. Detection of adulteration in Australian orange juices by stable carbon isotope ratio analysis (SCIRA). *J. Agric. Food Chem.* **2001**, *49*, 2623−2626.

(7) Widmer, W. W.; Cancalon, P. F.; Nagy, S. Methods for determining the adulteration of citrus juices. *Trends Food Sci. Technol.* **1992**, *3* (11), 278−286.

(8) Lee, H. S.; Wrolstad, R. E. Apple juice composition: Sugar, nonvolatile acid, and phenolic profiles. *J. Assoc. Off. Anal. Chem.* **1988**, *71*, 789−794.

(9) Taruscio, T. G.; Barney, D. L.; Exon, J. Content and profile of flavanoid and phenolic acid compounds in conjunction with the antioxidant capacity for a variety of northwest Vaccinium berries. *J. Agric. Food Chem.* **2004**, *52*, 3169−3176.

(10) Gonzalez-Sanjose, M. I.; Diez, C. Varietal characterization based on anthocyanin composition in grapes: Discriminant analysis. *Agrochimica* **1993**, *37* (1−2), 86−92.

(11) Kemsley, E. K.; Holland, J. K.; Defernez, M.; Wilson, R. H. Detection of adulteration of raspberry purees using infrared spectroscopy and chemometrics. *J. Agric. Food Chem.* **1996**, *44*, 3864−3870.

(12) Edelmann, A.; Diewok, J.; Schuster, K. C.; Lendl, B. Rapid method for the discrimination of red wine cultivars based on mid-infrared spectroscopy of phenolic wine extracts. *J. Agric. Food Chem.* **2001**, *49*, 1139−1145.

(13) Downey, G. Food and food ingredient authentication by mid-infrared spectroscopy and chemometrics. *Trends Anal. Chem.* **1998**, *17*, 418−424.

(14) Holland, J. K.; Kemsley, E. K.; Wilson, R. H. Use of Fourier transform infrared spectroscopy and partial least squares regression for the detection of adulteration of strawberry purees. *J. Sci. Food Agric.* **1998**, *76*, 263−269.

(15) Defernez, M.; Kemsley, E. K.; Wilson, R. H. Use of infrared spectroscopy and chemometrics for the authentication of fruit purees. *J. Agric. Food Chem.* **1995**, *43*, 109−113.

(16) Alonso-Salces, R. M.; Guyot, S.; Herrero, C.; Berrueta, L. A.; Drilleau, J. F.; Gallo, B.; Vicente, F. Chemometric characterisation of Basque and French ciders according to their polyphenolic profiles. *Anal. Bioanal. Chem.* **2004**, *379*, 464−475.

(17) Andrade, J. M.; Gomez-Carracedo, M. P.; Fernandez, E.; Elbergali, A.; Kubista, M.; Prada, D. Classification of commercial apple beverages using a minimum set of mid-IR wavenumbers selected by Procrustes rotation. *Analyst* **2003**, *128*, 1193−1199.

(18) Irudayaraj, J.; Tewari, J. Simultaneous monitoring of organic acids and sugars in fresh and processed apple juice by Fourier transform infrared-attenuated total reflection spectroscopy. *Appl. Spectrosc.* **2003**, *57*, 1599−1604.

(19) Irudayaraj, J.; Xu, F.; Tewari, J. Rapid determination of invert cane sugar adulteration in honey using FTIR spectroscopy and multivariate analysis. *J. Food Sci.* **2003**, *68*, 2040−2045.
(20) Defernez, M.; Wilson, R. H. Mid-infrared spectroscopy and chemometrics for determining the type of fruit used in jam. *J. Sci. Food Agric.* **1995**, *67*, 461−467.
(21) Cantos, E.; Espin, J. C.; Tomas-Barberan, F. A. Varietal differences among the polyphenol profiles of seven table grape cultivars studied by LC-DAD-MS-MS. *J. Agric. Food Chem.* **2002**, *50*, 5691−5696.
(22) Mazzuca, P.; Ferranti, P.; Picariello, G.; Chianese, L.; Addeo, F. Mass spectrometry in the study of anthocyanins and their derivatives: Differentiation of Vitis vinifera and hybrid grapes by liquid chromatography/electrospray ionization mass spectrometry and tandem mass spectrometry. *J. Mass. Spectrom.* **2005**, *40*, 83−90.
(23) Ichiyanagi, T.; Tateyama, C.; Oikawa, K.; Konishi, T. Comparison of anthocyanin distribution in different blueberry sources by capillary zone electrophoresis. *Biol. Pharm. Bull.* **2000**, *23*, 492−497.
(24) Prior, R. L.; Lazarus, S. A.; Cao, G.; Muccitelli, H.; Hammerstone, J. F. Identification of procyanidins and anthocyanins in blueberries and cranberries (*Vaccinium* spp.) using high-performance liquid chromatography/mass spectrometry. *J. Agric. Food Chem.* **2001**, *49*, 1270−1276.
(25) Giusti, M. M.; Rodriguez-Saona, L. E.; Griffin, D.; Wrolstad, R. E. Electrospray and tandem mass spectroscopy as tools for anthocyanin characterization. *J. Agric. Food Chem.* **1999**, *47*, 4657−4664.
(26) De Maesschalck, R.; Candolfi, A.; Massart, D. L.; Heuerding, S. Decision criteria for soft independent modeling of class analogy applied to near infrared data. *Chemom. Intell. Lab. Syst.* **1999**, *47*, 65−77.
(27) Kvalheim, O. M.; Karstang, T. V. SIMCA−Classification by means of disjoint cross validated principal components models. In *Multivariate Pattern Recognition in Chemometrics: Illustrated by Case Studies*; Brereton, R. G., Ed.; Elsevier: Amsterdam, Netherlands, 1992; pp 209−248.
(28) Vanden, B. K.; Hubert, M. Robust classification in high dimensions based on the SIMCA Method. *Chemom. Intell. Lab. Syst.* **2005**, *79*, 10−21.
(29) Alonso-Salces, R. M.; Herrero, C.; Barranco, A.; Berrueta, L. A.; Gallo, B.; Vicente, F. Technological classification of basque cider apple cultivars according to their polyphenolic profiles by pattern recognition analysis. *J. Agric. Food Chem.* **2004**, *52*, 8006−8016.
(30) Guess, M. J.; Wilson, S. B. Introduction to hierarchical clustering. *J. Clin. Neurophysiol.* **2002**, *19*, 144−151.
(31) Schindler, R.; Vonach, R.; Lendl, B.; Kellner, R. A rapid automated method for wine analysis based upon sequential injection (SI)-FTIR spectrometry. *Fresenius J. Anal. Chem.* **1998**, *362*, 130−136.
(32) *Performance and Application Comparisons for Single and Three Reflection Diamond Crystals for the MIRacle ATR Accessory*; http://www.piketech.com/technical/application-pdfs/Diamond-Crystal-Plates.pdf.
(33) Coates, J. Interpretation of infrared spectra, a practical approach. In *Encyclopedia of Analytical Chemistry*; Meyers, R. A., Ed.; John Wiley & Sons, Ltd.: Chichester, England, 2000; pp 10815−10837.
(34) Cabaniss, S. E.; McVey, I. F. Aqueous infrared carboxylate absorbances: aliphatic monocarboxylates. *Spectrochim. Acta, Part A* **1995**, *51*, 2385−2395.
(35) Cabaniss, S. E.; Leenheer, J. A.; McVey, I. F. Aqueous infrared carboxylate absorbances: Aliphatic di-acids. *Spectrochim. Acta, Part A* **1998**, *54*, 449−458.
(36) Mohammed-Ziegler, I.; Billes, F. Vibrational spectroscopic calculations on pyrogallol and gallic acid. *THEOCHEM* **2002**, *618*, 259−265.
(37) Araujo, P. Z.; Morando, P. J.; Blesa, M. A. Interaction of catechol and gallic acid with titanium dioxide in aqueous suspensions. 1. Equilibrium studies. *Langmuir* **2005**, *21*, 3470−3474.
(38) Eder, R. Pigments. In *Food Analysis by HPLC*; Nollet, L. M. L., Ed.; Marcel Dekker, Inc.: Monticello, NY, 2000; pp 845−880.
(39) Ruhl, E. H.; Clingelefer, P. R.; Nicholas, P. R.; Cirami, R. M.; McCarthy, M. G.; Whiting, J. R. Effect of rootstocks on berry weight and pH, mineral content and organic acid concentrations of grape juice of some wine varieties. *Aust. J. Exp. Agric.* **1988**, *28*, 119−125.
(40) Tominaga, Y. Comparative study of class data analysis with PCA-LDA, SIMCA, PLS, ANNs, and k-NN. *Chemom. Intell. Lab. Syst.* **1999**, *49*, 105−115.
(41) Bratchell, N. Cluster analysis. In *Multivariate Pattern Recognition in Chemometrics: Illustrated by Case Studies*; Brereton, R. G., Ed.; Elsevier: Amsterdam, Netherlands, 1992; pp 179−208.
(42) Ruiz-Jimenez, J.; Priego-Capote, F.; Garcia-Olmo, J.; Luque de Castro, M. D. Use of chemometrics and mid infrared spectroscopy for the selection of extraction alternatives to reference analytical methods for total fat isolation. *Anal. Chim. Acta* **2004**, *525*, 159−169.
(43) Wrolstad, R. E.; Culbertson, J. D.; Cornwell, C. J.; Mattick, L. R. Detection of adulteration in blackberry juice concentrates and wines. *J. Assoc. Off. Anal. Chem.* **1982**, *65*, 1417−1423.
(44) De Braekeleer, K.; De Maesschalck, R.; Hailey, P. A.; Sharp, D. C. A.; Massart, D. L. Online application of the orthogonal projection approach (OPA) and the soft independent modeling of class analogy approach (SIMCA) for the detection of the end point of a polymorph conversion reaction by near infrared spectroscopy (NIR). *Chemom. Intell. Lab. Syst.* **1999**, *46*, 103−116.

Received for review September 22, 2006. Revised manuscript received March 2, 2007. Accepted March 5, 2007.

JF062715C