JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Rapid and Simultaneous Determination of Lycopene and β-Carotene Contents in Tomato Juice by Infrared Spectroscopy

THAIS DE NARDO,[†] CECILIA SHIROMA-KIAN,[†] YUWANA HALIM,[†] DAVID FRANCIS,[§] AND LUIS E. RODRIGUEZ-SAONA*,[†]

Department of Food Science and Technology, 110 Parker Food Science and Technology Building, The Ohio State University, 2015 Fyffe Road, Columbus, Ohio 43210, and Department of Horticulture and Crop Science, The Ohio State University, 210 Williams OARDC—Wooster, Wooster, Ohio 44691

The rapid quantification of lycopene and β-carotene in tomato juices by attenuated total reflectance (ATR) infrared spectroscopy combined with multivariate analysis was evaluated. Two sample preparation methods were compared: a direct measurement of the tomato paste and an extraction method using hexane to isolate carotenoids. HPLC was used as the reference method. Cross-validated (leave-one-out) partial least-squares regression (PLSR) was used to create calibration models to predict these phytonutrient concentrations in blind test samples. The infrared spectra showed unique marker bands at 957 and 968 cm^{-1} for lycopene and β-carotene, respectively. Multivariate analysis of the infrared spectral data gave correlation coefficients (*r* values) of >0.9 between the ATR-IR predicted and HPLC reference values, and standard errors of cross-validation (SECV) of 0.5 and 0.04 mg/100 g of juice for lycopene and β-carotene, respectively. ATR-IR could provide the tomato industry with a simple, rapid, and high-throughput technique for the determination of tomato quality.

KEYWORDS: Lipid extract; ATR-IR spectroscopy; multivariate analysis; HPLC

INTRODUCTION

The United States is one of the world's largest producers of processed tomato products, second only to China. Over the past 20 years, average U.S. production has drastically increased, and the farm value of the tomato crop is now around $800 million (*1*). Research has shown that processed tomato products have a higher content of lycopene and improved bioavailability as compared to their fresh counterparts (*2*, *3*). Processing may add value in terms of human health, which could drive consumer interest and demand for tomato products as functional foods.

Lycopene ($C_{40}H_{56}$) is the major carotenoid found in tomatoes, accounting for up to 90% of total carotenoids. Lycopene is the most unsaturated acyclic carotenoid, with 11 conjugated double bonds that act as effective free radical scavengers (*3*). The ability of lycopene to act as a potent antioxidant is thought to be responsible for protecting cells against oxidative damage and, therefore, decreasing the risk of the development of chronic diseases, such as atherosclerosis, coronary heart disease, and certain types of cancer (*4*). β-Carotene ($C_{40}H_{56}$) is another important carotenoid found in tomatoes with potential health benefits. As a provitamin A carotenoid, it can be enzymatically converted in the intestinal mucosa to retinal and, ultimately, retinol (vitamin A). Retinol is required for vision, maintenance of differentiated epithelia, mucus secretion, and reproduction (*5*).

The health benefits offered by carotenoids have generated interest in developing breeding lines with diverse carotenoid profiles. Carotenoid concentration in fruits depends on genetics, maturity, environmental factors, and cultivation techniques (*6*). Thus, the development of tomato varieties with increased carotenoid levels and altered carotenoid profiles requires efficient selection and the ability to measure the metabolite content in thousands of samples. Currently, separation and quantification of carotenoids rely heavily on analysis by high-performance liquid chromatography (HPLC) (*7*, *8*). Although reliable and accurate, this method is time-consuming and requires extensive sample preparation and the use and disposal of hazardous organic solvents. Conventional UV−vis spectrophotometric assays that utilize reduced volumes of organic solvents have been shown to be simple, rapid, and inexpensive methods for measuring lycopene content in tomato and tomato products (*9*), but absorbance interference from lycopene provides poor accuracy and overestimation of β-carotene levels (*10*). Novel chemical techniques that have been evaluated for direct determination of lycopene and other nutritionally important carotenoids include the application of optothermal methods (*11*),

* Author to whom correspondence should be addressed [telephone (614) 292-3339; fax (614) 292-0218; e-mail rodriguez-saona.1@osu.edu].

[†] Department of Food Science and Technology, The Ohio State University, Columbus.

[§] Department of Horticulture and Crop Science, The Ohio State University, Wooster.

10.1021/jf802920z CCC: $40.75
Published on Web 01/29/2009

resonance Raman spectroscopy (*12*), and infrared (*13*) and near-infrared spectroscopy (*14, 15*).

Fourier transform infrared (FT-IR) in combination with multivariate analysis offers a powerful and rapid technique for the analysis of agricultural and food products (*16*). Spectral bands arising from functional group vibrations of organic molecules in the mid-infrared (4000−700 cm^{-1}) region may be associated with specific chemical entities with known assignment in most cases. In addition, the mid-IR spectra allows for the chemically based discrimination of organic constituents, producing distinct and reproducible biochemical fingerprints that reflect the total biochemical composition of the sample, with bands due to major constituents such as lipids, proteins, nucleic acids, polysaccharides, and phosphate-carrying compounds (*17*).

Attenuated total reflection Fourier transform infrared spectroscopy (ATR-IR) is one of the most powerful and widely used sampling tools (*18*) for collecting IR spectra of biological materials. The most important feature of the ATR accessory is the evanescent field, which develops during the reflection of radiation at the interface of a material with a high refractive index (ATR crystal) and a material with a low refractive index (sample). Attenuation of this electric field by functional groups in the lower refractive index material results in a spectrum analogous to an absorbance spectrum (*19*). The depth of penetration of the evanescent field protrudes only a few micrometers (0.5−5 μm) beyond the crystal surface and into the sample. A number of different crystal materials are available such as diamond, zinc selenide, silicon, AMTIR, KRS-5, and germanium. The selection of the ATR crystal characteristics should include the depth of penetration of the IR beam, hardness, and resistance to acid or caustic to prevent crystal damage (*18*). ATR allows qualitative or quantitative analysis of samples with little or no sample preparation, which greatly speeds data collection. In addition, developments of multivariate techniques for classification and regression have resulted in reliable, accurate, robust, and simple methods for routine analysis of spectroscopic data (*20*).

Genetic resources that include naturally occurring variants of genes affecting both the structure and regulation of key enzymes in the carotenoid biosynthesis pathway make tomatoes an excellent model food for nutritional studies. The development of improved analytical techniques that allow the efficient measurement of diverse carotenoids in very large genetically characterized populations is greatly needed. Thus, the objective of this research was to develop a methodology for the simple, rapid, and accurate determination of lycopene and β-carotene contents in tomato juices by ATR-IR spectroscopy combined with multivariate techniques. The performance of multivariate models using direct measurements of the tomato paste or a hexane extraction procedure for the determination of lycopene and β-carotene content was evaluated. In addition, we explored the feasibility of using infrared spectroscopy as a chemotaxonomic tool for the classification of tomato juice samples obtained from a diverse genetic pool of tomato varieties and breeding lines.

MATERIALS AND METHODS

Plant Material. A total of 20 different tomato varieties were processed into juice at The Ohio State University food industry center pilot plant (Columbus, OH). The tomatoes were grown and harvested at the North Central Agricultural Research Center at Fremont, OH. Juice samples were obtained from varieties and breeding lines that encompassed a range of pigment levels. Plant materials included experimental hybrids and commercial hybrids used in the tomato-processing industry. Red tomatoes used included TSH04 (Tomato Solutions), Ohio OX52, TSH8 (Tomato Solutions), Progress (Seminis), TSH16 (Tomato Solutions), GEM31 (RedGold), FG01-107 (Ohio State University), GEM94 (RedGold), TC680 (HarrisMoran), FGH04-120 (Ohio State University), FGH04-142 (Ohio State University), GEM111 (RedGold), FGH04-133 (Ohio State University), HyPeel 696 (Seminis), FGH04-472 (Ohio State University), FGH01-163 (Ohio State University), FGH01-158 (Ohio State University), FGH04-179 (Ohio State University), FGH04-478 (Ohio State University), and OH8245. The genetic differences were conferred by the og^c allele for the fruit-specific β-cyclase, as well as quantitative trait loci that affect fruit color and lycopene content (*21*). Also included in the set of varieties was a low-carotenoid variety, FGH04-138, which is homozygous for the yellow flesh gene, *r*, containing a mutation in the phytoene synthase enzyme, which results in a very low fruit carotenoid content. Thus, the collection of varieties spans a range of variation from very low to very high lycopene and β-carotene contents. Tomatoes were grown in a randomized complete block design with two blocks. The tomatoes were mechanically harvested and sorted with only ripe fruit used for juice preparation. For the red varieties, ripeness was assessed on the basis of uniform red color, and for FGH04-138, ripe tomatoes appear uniformly yellow.

Processing. Tomato juice processing operations at The Ohio State University pilot plant (Columbus, OH) included washing, sorting, and chopping (Fitzpatrick mill with a $^3/_4$ in. screen). Tomatoes were pumped through a heat exchanger to reach a hot-break temperature of 93 °C, at which they were heated for approximately 2 min to inactivate enzymes that affect flavor and viscosity (*22*). Juice was extracted using a screw-type extractor with a 0.2 screen (Chisholm-Ryder Co., model CLE-360-D28), filled into 300 × 406 cans and processed at 104 °C for 30 min. Canned tomato juices were stored at room temperature until analysis. Standards of *all-trans*-lycopene and β-carotene were purchased from Sigma-Aldrich Chemical Co. (St. Louis, MO). All solvents were of certified HPLC or ACS grade (Fisher Scientific Co., Fair Lawn, NJ).

Direct Determination of Lycopene and β-Carotene. Tomato juice aliquots (5 mL) were mixed with methanol (5-mL) and centrifuged for 5 min at a speed of 3000 rpm. The supernatant was discarded, and the pellet (juice solids) was placed onto an AP15 glass fiber filter disk (Millipore, Billerica, MA); samples were pressed onto the ATR-IR diamond crystal using a high-pressure clamp (Pike Technologies, Madison, WI) for infrared spectra acquisition. Each tomato juice sample was analyzed in duplicate (two cans), and four independent measurements of the paste were collected per sample, giving a total of eight spectra per juice sample.

Extraction of the Tomato's Lipid Phase. Extraction of lycopene and β-carotene from tomato juices was done as reported by Nguyen and Schwartz (*8*) with some modifications. A controlled light environment was used to prevent isomerization and photodegradation. Tomato juice aliquots (5 mL) were mixed with methanol (5 mL) and centrifuged for 5 min at a speed of 3000 rpm. The supernatant was discarded, and 0.2 g of the pellet was resuspended with a mixture of a 1:1:2 acetone/ethanol/hexane solution (20 mL volume) and continuously shaken for 15 min (Rotamix model RK DYNAL, Dynal Biotech Inc., Lake Success, NY) at 180 rpm. Deionized distilled water (3 mL) was later added and shaken continuously for 5 min at 180 rpm to induce phase separation. The hexane (top) layer was collected and brought up to 10 mL in a volumetric flask. An aliquot (3 mL each) was dried under nitrogen, resuspended in chloroform (60 μL), and applied onto the ATR-IR ZnSe crystal for infrared spectral acquisition. The remaining aliquots (2 × 3 mL) were dried under nitrogen and immediately analyzed by HPLC. Two cans for each tomato variety sample were opened, and duplicate or triplicate carotenoid extractions were performed with two independent measurements made by ATR-IR spectroscopy, giving a total of 8−12 spectra per juice sample.

Infrared Spectroscopy. A FTS 3500GX Fourier transform infrared spectrometer (Varian Inc., Palo Alto, CA) was used in combination with a potassium bromide beam splitter and deuterated triglycine sulfate (DTGS) detector for all readings, operating at 8 cm^{-1} resolution. The direct method analysis used an ATR accessory with a three-reflection diamond crystal plate (MIRacle, Pike Technologies, Madison WI), providing a 3-fold increase in sample response compared to the standard

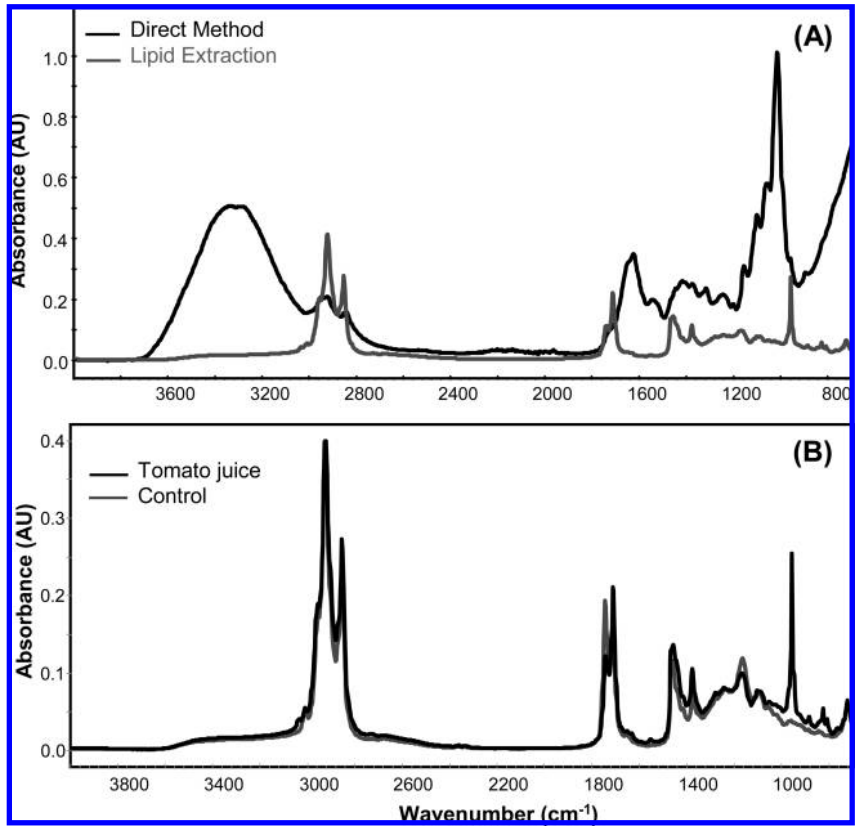

Figure 1. (**A**) Comparison of attenuated total reflectance (ATR) infrared absorption spectrum of a tomato juice sample by the direct method using a diamond crystal plate and of a tomato lipid extract using a ZnSe crystal plate. (**B**) Infrared absorption spectrum of the lipid extract of a high-lycopene tomato juice sample and a low-carotenoid control using an ATR-infrared accessory equipped with a ZnSe crystal plate.

single-reflection crystal plate (*23*). The lipid extraction analysis used an ATR ZnSe crystal plate with a refractive index of 2.5 that permitted a triple reflection within the sample at an incidence angle of 45°, for highest infrared sample throughput (Pike Technologies). Spectra were collected over the frequency region from 4000 to 700 cm^{-1}, and interferogram of 128 scans was co-added followed by Beer−Norton apodization. Multiple lipid-phase tomato juice samples and juice solids samples from each breeding selection were measured to evaluate the reproducibility of the technique. The absorbance spectrum was obtained by rationing the sample single-beam spectrum against that of a blank optical path (reference spectrum). Spectra were displayed in terms of absorbance and viewed by using Win-IR Pro software (Varian). The instrument was continuously purged with CO_2-free dry air from a CO_2RP140 dryer (Dominick Hunter, Charlotte, NC).

Multivariate Analysis. The multivariate analysis was performed as reported by Halim et al. (*13*). The spectra were exported as GRAMS.spc files and imported into the multivariate statistics program Pirouette, for Windows Comprehensive Chemometrics Modeling Software, version 3.11 (Infometrix, Inc., Bothell, WA). The spectra were derivatized using a five-point polynomial-fit Savitzky−Golay function and normalized using the maximum normalization function. Spectral data were then analyzed by partial least squares regression (PLSR) that was cross-validated (leave-one-out approach) to generate calibration models. PLSR is a bilinear regression analysis method that extracts a small number of orthogonal factors that are linear combinations of the spectral (*X*) variables and uses these factors as regressors for the analyte's concentration (*Y* variable). These orthogonal factors (latent variables) explain as much as possible of the covariance of the *X* and *Y* variables. PLSR has the potential to estimate the component concentration and chemical and physical properties (loading vectors, vector of final calibration regression coefficients, and spectral residuals) from the spectra (*24*). PLSR has been particularly successful in developing multivariate calibration models for the spectroscopy field because it uses the concentration information (*Y* variable) actively in determining how the regression factors are computed from the spectral data matrix (*X*), reducing the impact of irrelevant *X* variations in the calibration

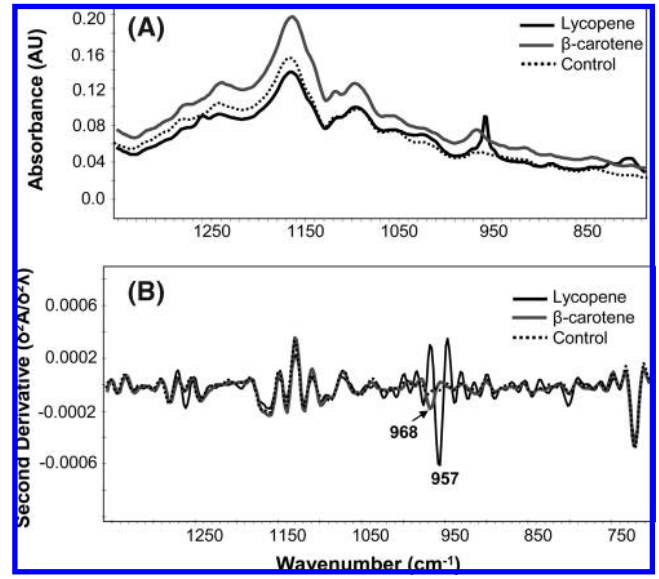

Figure 2. Attenuated total reflectance (ATR) infrared spectra (**A**) and second-derivative spectral transformation (**B**) of a low-carotenoid tomato sample (control), FGH04-138, supplemented with lycopene and β-carotene standard solutions to yield levels of 5 mg/100 g of tomato juice.

model (*25*). This capability provides a more information-rich data set of reduced dimensionality and eliminates data noise, which results in more accurate and reproducible calibration models (*25*). PLSR models were evaluated in terms of loading vectors, standard error of calibration (SEC), standard error of cross-validation (SECV), correlation coefficient (*r* value), and outlier diagnostics.

Spectral data were also analyzed by soft independent modeling of class analogy (SIMCA) to generate clustering groups. SIMCA is a multivariate technique based on principal component analysis (PCA),

Table 1. PLSR Models for Direct[a] (1500−800 cm^{-1}) and Extraction[b] (1300−800 cm^{-1}) Methods for Determination of Lycopene and β-Carotene in Tomato Juice

carotenoid	carotenoid range[c] (mg/100 g of juice)	PLSR calibration	no. of samples	latent variables[d]	R-Val[e]	SECV[f] (mg/100 g of juice)	R-Cal[g]	SEC[h] (mg/100 g of juice)
lycopene	9.25−15.22	direct	160	10	0.97	0.40	0.99	0.18
		extraction	200	6	0.96	0.52	0.97	0.43
β-carotene	0.18−0.66	direct	160	9	0.91	0.054	0.97	0.031
		extraction	200	12	0.96	0.036	0.98	0.025

[a] Application of the tomato paste directly onto an ATR diamond crystal for spectral collection. [b] Extraction of the lipid phase from the tomato paste and application onto an ATR ZnSe crystal for spectral collection. [c] Carotenoid range determined on the basis of an HPLC reference method. [d] Latent variables: orthogonal factors that provide maximum correlation with dependent variable. [e] R-Val, coefficient of correlation for cross-validated model. [f] SECV, standard error of cross-validation. [g] R-Cal, coefficient of correlation for calibration model. [h] SEC, standard error of calibration.

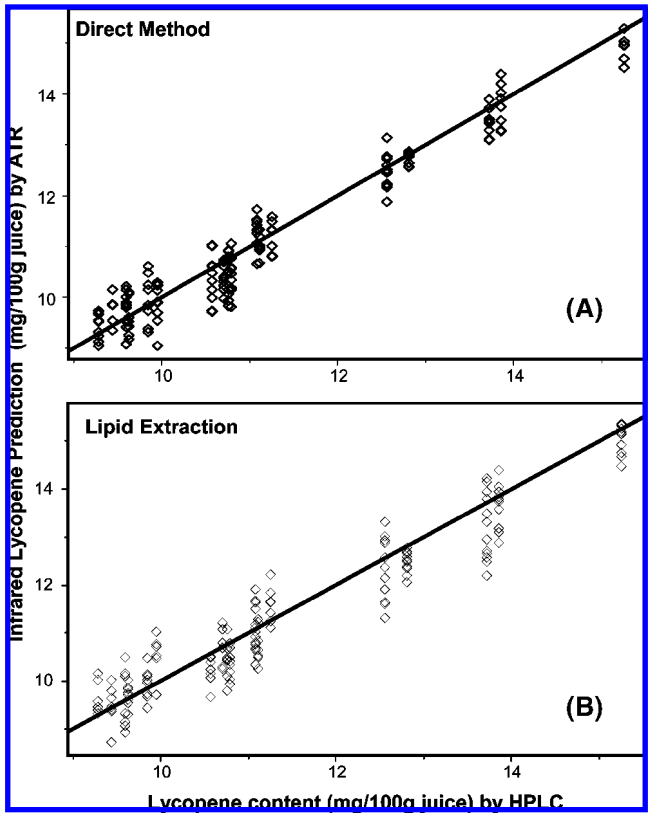

Figure 3. Cross-validated (leave-one-out) PLSR plots for lycopene content in tomato juice using the direct method (**A**) and the lipid extraction method (**B**).

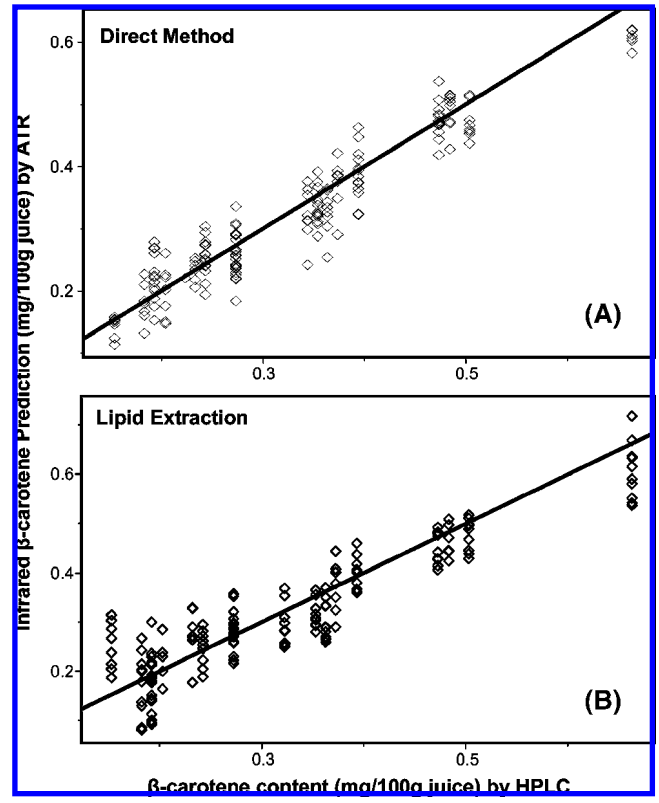

Figure 4. Cross-validated (leave-one-out) PLSR plots for β-carotene content in tomato juice using the direct method (**A**) and the lipid extraction method (**B**).

which was used in this study to evaluate the ability of the ATR-IR spectral data to discriminate among tomato varieties (*26*). In SIMCA, training sets are assigned to classes and a principle component model is generated for each class with distinct confidence regions within them (*27*). The performance of this method depends not only on the difference between classes but also strongly on the training (calibration) set for each class (*27*). The scores plot allows the visualization of clustering among samples (sample patterns, groupings, or outliers). Sample residual and Mahalanobis distance (*28*) were used for outlier diagnostics. The discriminating power in SIMCA provides a diagnostic tool that indicates on a relative scale how important a wavenumber in the spectrum (*x* variable) is to differentiating classes. This is accomplished by comparing the class residual variance when the class samples are modeled within the classes not containing them to the class residual variances when the class samples are modeled within their own classes. SIMCA analysis assesses itself by predicting each sample included in the training set comparing that prediction to its assigned class; this assessment is referred to as misclassifications. Zero misclassifications typify a model in which all samples were correctly predicted to the preassigned class.

Reference High-Performance Liquid Chromatography (HPLC) Analysis. HPLC was used as reference method because of its sensitivity and selectivity, which provide reliable analysis of individual carotenoids (*10*). Lycopene and β-carotene concentrations in each tomato juice were determined by an HP1050 reverse-phase HPLC system equipped with a photodiode array detector (Agilent Techologies, Palo Alto, CA). The conditions reported by Halim et al. (*13*) were used to analyze the extracted carotenoid fraction. Dried carotenoid-containing extracts (see extraction section) were redissolved in a 1.5 mL mixture of 2:1 methanol/MTBE solvent, filtered using a nylon membrane filter (Fisherbrand, 13 mm diameter, 0.45 μm pore size, Fisher Scientific Co.), and 50 μL of sample was injected into the HPLC system for carotenoid analysis. The column consisted of a C_{18} Vydac (5 μm particle size, 256 × 4.6 mm) (Grace, Deerfield, IL), and a guard column packed with Novapack C_{18} stationary phase (Milford, MA) was used in-line for all separations. Elution solvents used in this experiment were (A) 1% (v/v) methanol with ammonium acetate and (B) MTBE. Separations were carried with a 20 min linear gradient from 0 to 20% B, followed by a 5 min hold at 20% B, and a final 10 min linear gradient from 20 to 0% B. The column was then washed with solvent A for at least 5 min before the next analysis. The flow rate was 1.0 mL/min with detection at 450 and 471 nm for β-carotene and *trans*-lycopene,

Table 2. Reported Partial Least-Squares Regression (PLSR) Performance Statistics for Determination of Lycopene and β-Carotene in Tomato Products by Infrared (Mid- and Near-) and Raman Spectroscopic Techniques

method	lycopene range (mg/100 g)	β-carotene range (mg/100 g)	RMSEP,[a] lycopene	R value,[b] lycopene	RMSEP, β-carotene	R value, β-carotene	source
NIR	14.4−92.1	0.51−1.75	2.16	0.99	0.75	0.99	*14*
ATR-IR	2.6−629.0	0.23−2.83	33.20	0.98	0.16	0.97	*15*
NIR	2.6−629.0	0.23−2.83	91.19	0.85	0.41	0.80	*15*
FT-Raman	2.6−629.0	0.23−2.83	74.34	0.91	0.34	0.89	*15*
ATR-IR	5.8−13		0.69	0.93			*13*

[a] RMSEP, root mean standard error of prediction. [b] *R* value, correlation coefficient.

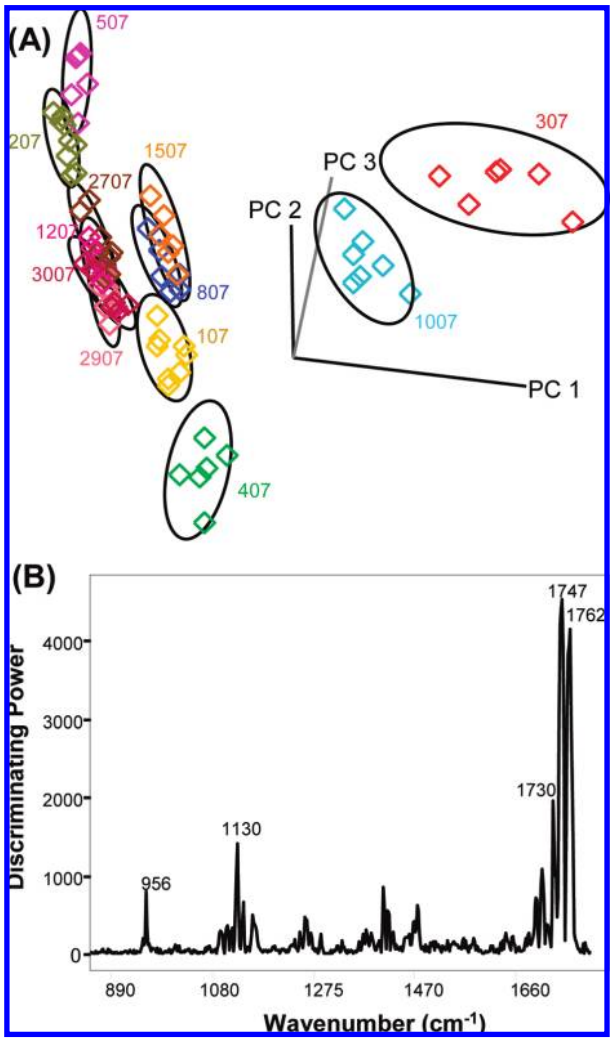

Figure 5. SIMCA class projections of infrared spectra (**A**) and discriminating power (**B**) for different red tomato juice varieties. Data were collected in the 4000−700 cm^{-1} spectral region. Varieties included the following: 107, TSH04 (Tomato Solutions); 207, Ohio OX52; 307, TSH08 (Tomato Solutions); 407, Progress (Seminis); 507, TSH16 (Tomato Solutions); 807, GEM31 (RedGold); 1007, GEM94 (RedGold); 1207, TC680 (HarrisMoran); 1507, FG04-142 (Ohio State University); 2707, FGH04-472 (Ohio State Universtiy); 2909, FGH01-163 (Ohio State University); 3007, FGH01-158 (Ohio State University).

respectively. The identification of *all-trans*-carotenoids was carried out by comparing the retention times and absorption spectra with the reference standard, with typical retention times of ∼19 min for lycopene and ∼15.5 min for β-carotene. Lycopene and β-carotene standards were used to build calibration curves for quantitation based on peak area. Triplicate analyses were performed, and the mean value was determined. Because the concentrations of lycopene and β-carotene may differ slightly, either due to raw fruit concentrations or differences in the processing, the moisture content (ranging from 65 to 72% with an overall average value of 68%) for all samples was determined and used for determination of lycopene and β-carotene levels in the tomato juices.

Predicting Carotenoid Content in Tomato Juice Samples. An independent set of randomly selected canned tomato juice samples (TSH04, TC680, HyPeel696, FGH04-472, FGH01-158, and OH8245) that included tomato varieties representative of those used for PLSR modeling and commercial tomato juices purchased from a local store (Columbus, OH) were analyzed to estimate lycopene and β-carotene levels. The carotenoid levels were determined by the HPLC reference method. Infrared spectra were collected for samples using the direct and lipid extraction methods as described previously and were used to test the predictive ability of the PLSR calibration model for lycopene and β-carotene content in tomato juices. Predictions were performed using Pirouette software (Infometrix, Inc., Woodvile, WA).

RESULTS AND DISCUSSION

Spectral Analyses. A simple, fast, and reliable direct determination of lycopene content in tomatoes by ATR-IR spectroscopy was previously reported by Halim et al. (*13*). In this study we evaluated ATR-IR spectroscopy for the simultaneous determination of the major carotenoid pigments in tomato juice products, lycopene and β-carotene. In addition, we compared the performance of calibration models obtained from spectra generated by either directly measuring the tomato sample or using a lipid extraction procedure. The extraction protocol separated lipid-soluble components, allowing for the removal of overlapping signals, especially the broad absorption signals of water, carbohydrate, and protein bands that may influence the spectral resolution of the target phytochemical groups (**Figure 1**). The formation of a homogeneous film onto the ZnSe crystal was critical to yield reproducible spectra and required controlled delivery of sample to provide adequate evaporation of the lipid extract solvent (chloroform). The typical ATR-IR spectrum for the lipid phase analysis of a tomato juice sample (**Figure 1B**) showed characteristic bands associated with lipid groups at 1380−1370 and 1485−1440 cm^{-1} (C−H bending), 1725−1700 and 1750−1725 cm^{-1} (C═O esters), and 3000−2800 cm^{-1} (C−H stretch) (*17*). The infrared region between 1200 and 700 cm^{-1} is characteristic of the "fingerprint region", which provides complex but unique and reproducible spectral information (ester, ethers, alcohols) with a significant contribution for substance identification (*29*). Visual analysis of the ATR-IR spectrum of the lipid phase for the high *all-trans*-lycopene-containing sample and the control (low carotenoid containing sample) showed a unique IR spectral band at 957 cm^{-1} absorbance band corresponding to the presence of *trans*-HC═CH− bending out of plane vibrations (*30*). Furthermore, comparison of the ATR-IR spectra of the low-carotenoid breeding line variety, FGH04-138, supplemented (5 mg/100 g) with lycopene and β-carotene standards (Sigma-Aldrich Chemical Co.) allowed corroboration of the unique spectral bands for lycopene (957 cm^{-1}) and β-carotene (968 cm^{-1}) (**Figure 2**).

Table 3. PLSR Models Predicted Concentration of Lycopene and β-Carotene in Tomato Juice Samples Using ATR-IR for the Direct and Extraction Methods

analyte	sampling method	sample[a]	reference value[b] (mg/100 g of juice)	SD[c]	CV[d] (%)	IR predicted value[e] (mg/100 g of juice)	SD[c]	CV[d] (%)
lycopene	direct	107	13.68	1.00	7.31	13.74	0.21	1.56
		2507	9.81	0.29	2.95	10.34	0.17	1.69
		3007	9.26	0.98	10.63	9.65	0.11	1.15
		3507	10.75	0.08	0.73	10.12	0.51	5.08
	extraction	1207	9.56	0.85	8.89	9.94	0.34	3.41
		2707	9.59	0.88	9.18	9.62	0.39	4.07
		3507	10.75	0.08	0.73	9.63	0.14	1.49
		juice A	8.21	1.02	12.40	8.30	0.20	2.40
		juice B	9.30	0.72	7.77	8.84	0.49	5.52
β-carotene	direct	107	0.36	0.02	6.44	0.38	0.02	5.90
		2507	0.36	0.04	10.18	0.35	0.01	1.63
		3007	0.24	0.02	9.36	0.21	0.01	2.93
		3507	0.18	0.04	21.40	0.16	0.03	18.43
	extraction	1207	0.55	0.05	8.29	0.55	0.02	4.56
		2707	0.17	0.04	23.36	0.16	0.03	20.04
		3507	0.18	0.04	21.40	0.18	0.02	9.18
		juice A	0.16	0.02	12.69	0.19	0.01	7.65
		juice B	0.52	0.06	11.87	0.52	0.01	1.65

[a] Sample numbers are codes given by the North Central Agricultural Research Center at Fremont (Ohio) for genetically diverse varieties of tomatoes: 107, TSH04; 1207, TC680; 2507, Hypeel 696 (or PS696); 2707, FGH04-472; 3007, FGH01-158; 3505, OH8245. Commercial juices (A and B) were purchased from a local store (Columbus, OH). Tomato juice B indicated a lycopene content of 8.5 mg/100 mL of juice. [b] Reference value reports the average carotenoid (lycopene or β-carotene) level determined by HPLC, based on triplicate extractions. [c] SD, standard deviation. [d] CV, coefficient of variability. [e] IR predicted values report the estimated average carotenoid (lycopene or β-carotene) levels determined by ATR-IR spectroscopy, based on triplicate (lipid fraction) or duplicate (direct) independent measurements.

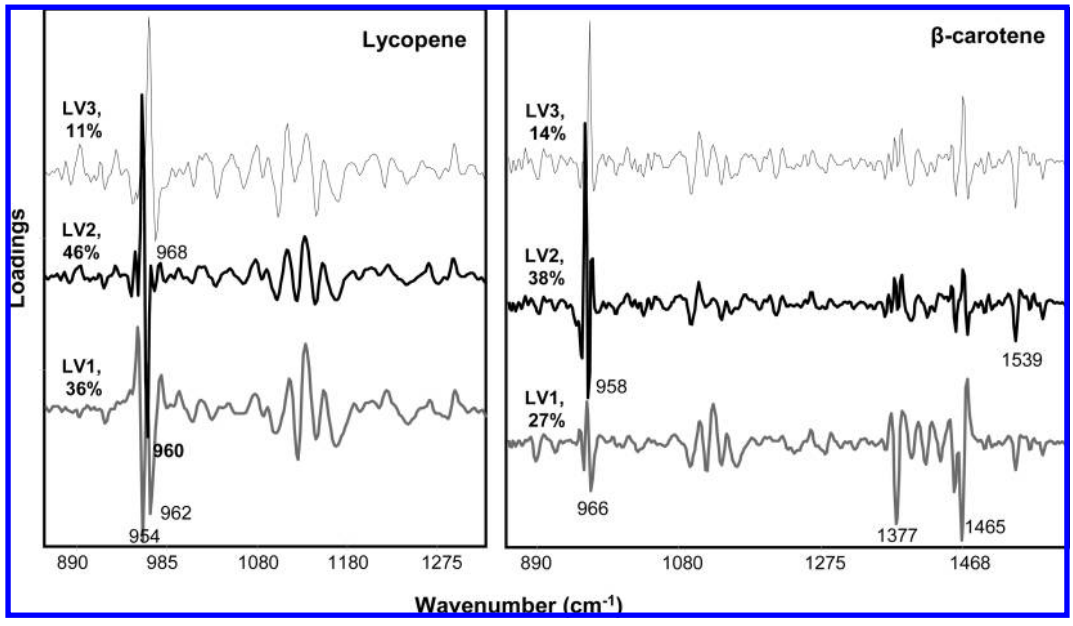

Figure 6. Partial least-squares loadings plot for cross-validated models for the determination of lycopene and β-carotene using the lipid extraction method.

Quantitative Analysis of Tomato Juice Carotenoids. The lycopene and β-carotene contents of the tomato juices ranged from 9.2 to 15.2 mg/100 g of juice and from 0.18 to 0.66 mg/100 g of juice, respectively, based on the HPLC analysis. A Pearson correlation of 0.59 ($p = 0.006$) between lycopene and β-carotene levels provided evidence of a positive relationship between the levels of these carotenoids for the different tomato breeding selections used for juice processing. Levels of carotenoids in tomato juice have been reported to range from 5.8 to 12.17 mg/100 g for lycopene and from 0.48 to 0.60 mg/100 g of juice for β-carotene (*3*). Average levels of 10.8 and 0.27 mg/100 g of lycopene and β-carotene, respectively, have been reported in tomato juices (*31*). The carotenoid levels (lycopene and β-carotene) determined in this study are close to those reported in the literature, taking into consideration that carotenoid levels may be affected by factors such as variety, maturity, temperature, altitude, and soil nutrients, among others. Our precisions (% CV) for replicated HPLC assays were <14 and <18% for the lycopene and β-carotene samples, respectively.

The cross-validated (leave-one-out) PLSR model results are shown in **Table 1**. Second derivative (5 point window) transformation of the spectral measurements improved the quantitative analysis by resolving overlapped bands and limiting variations in spectral baselines. The estimated contents of lycopene (**Figure 3**) and β-carotene (**Figure 4**) measured by ATR-IR spectroscopy showed good correlation with the refer-

ence HPLC analysis. The PLSR models based on mid-infrared spectra collected by the direct determination of tomato paste (r value > 0.91, and SECV of 0.40 mg of lycopene/100 g of juice and 0.05 mg of β-carotene/100 g of juice) gave performance statistics similar to those of models generated from spectra collected from the lipid fraction of tomato juices (r value > 0.95, and SECV 0.52 mg of lycopene/100 g of juice and 0.04 mg of β-carotene/100 g of juice). Removal of interfering signals (i.e., carbohydrates, proteins, moisture) from the spectra by separation of the lipid fraction only slightly improved the performance of the PLSR models. Nevertheless, the lipid extraction method proved to be best for qualitative analysis, as shown by the SIMCA model (**Figure 5A**). Direct analysis of the tomato juice solids limits the error introduced by the lipid fractionation steps (weighing, extraction efficiency, carotenoid degradation), and the unique infrared fingerprint of carotenoids (*13*) in the tomato matrix allowed PLSR to minimize the impact of irrelevant spectral variations in the calibration model (*24*). The reported performance of PLSR models for the determination of lycopene and β-carotene content in different tomato products by infrared (mid- and near-) and Raman spectroscopic techniques are shown in **Table 2**. Overall, our results showed superior performance statistics, and the direct analysis of tomato juice solids allowed the estimation of carotenoid content in a few minutes (~2 min). The PLSR models were used to predict the concentrations of lycopene and β-carotene in an independent set of juice samples. ATR-IR predictions were comparable to the HPLC reference values, as shown in **Table 3**. Improved reproducibility among measurements (coefficient of variability, CV) was obtained with the ATR-IR method for both the lipid extract and direct analysis as compared to HPLC analysis. In general, the CV for β-carotene was higher than that for lycopene, probably because of the low β-carotene concentrations in tomato juices and the signal-to-noise ratio might have considerable effects on the reproducibility of the methods.

The PLSR loading vectors provide interpretative tools by identifying frequencies associated with the highest variation in the calibration set, reflecting inter-relationships among different variables (infrared spectra) that can be related to chemical components. The first three and five latent variables (factors) explained most of the variance (93%) in the PLSR models for determination of lycopene and β-carotene, respectively. The partial least-squares loading plots for the cross-validated PLSR models (**Figure 6**) of both lycopene and β-carotene using the mid-infrared spectra from the tomato juice lipid fraction showed bands in the region ranging from 970 to 950 cm^{-1} that correlated with unique spectral signal attributed to the deformation vibrations of *trans*-C−H out-of-plane bend at 957 cm^{-1} (lycopene) and 968 cm^{-1} (β-carotene), respectively (**Figure 6**). In addition, aliphatic functional group [1300−1600 cm^{-1} (C−H bending)] vibrations of lipid moieties influenced the spectral variation in estimating β-carotene in tomato juices.

Chemometric Classification of Tomato Samples. The feasibility of the application of a pattern recognition method (SIMCA) for the classification of different red tomato juice samples based on the infrared spectral patterns of their lipophilic components (lipid extraction) was evaluated. Twelve tomato juice samples were shown instead of the complete set of juices (20) because they provided simple visualization of clustering patterns and included a mixture of experimental, commercial, and elite hybrids with wide genetic variation. The SIMCA class projection (**Figure 5A**) for selected tomato juice samples exhibited well-separated and tight clusters with interclass distances [based on Mahalanobis distance measurements computed between the centroids of classes (*27*)] ranging from 2.9 to 24.5. Clusters with interclass distance values of >3 are considered to be different from one another (*32*). There were zero misclassifications in validating the predictive ability of the SIMCA model. Spectra obtained from the direct method did not generate well-defined clusters, probably due to interferences of the components in the juice solids. **Figure 5B** shows the discriminating power for the SIMCA model, identifying the frequencies in the spectrum that have a prevailing and important effect for sample classification (*32*), which then can be correlated with known chemical functional groups. The bands associated with most of the discrimination among tomato juice samples corresponded to the region between 1770 and 1720 cm^{-1} associated with carbonyl ester (−C═O) stretching vibrations of triglycerides (*17, 29*) (**Figure 5B**). The *trans*-lycopene band (957 cm^{-1}) contributed only slightly to the discrimination among tomato juices, showing a weak intensity in the discriminating power plot. Thus, our results showed that the presence of lipid components other than carotenoids can yield specific spectral profiles that discriminate among tomato juice samples obtained from a diverse genetic pool. Lipid profiling reflects the species taxonomy and cellular biochemistry (*33*), being associated as chemical descriptors to differentiate among plant varieties (*34*).

The ATR-IR technique used with two different methods of sample preparation, a direct method and lipid-phase extraction method, allowed the development of PLSR and SIMCA models for the quantitative and qualitative analysis of carotenoids in tomato juice samples. Our PLSR models for quantification of lycopene and β-carotene showed superior performance statistics compared to those reported in literature for various tomato products. PLSR models generated from mathematically transformed (normalized, second derivative) infrared spectral data gave correlation coefficients (r value) of >0.95 between the ATR-IR predicted and HPLC reference values and SECV of 0.5 and 0.06 mg/100 g of juice for lycopene and β-carotene, respectively. Both sample preparation methods were accurate for the analysis of lycopene and β-carotene content in processed tomato juices, based on the predictions of an independent validation set of samples. However, given the greater simplicity and speed of the "direct" method over the "lipid extraction" method and the similar predictive abilities of the chemometric models, r value and SECV, we recommend the direct method for quantitation of lycopene and β-carotene in tomato juice. In contrast, extraction of the lipid phase of tomato juices for ATR-IR analysis removed interfering signals, allowing for the clustering of tomato juice samples probably based on variety. This technique provides for the fast analysis of food carotenoids with minimal personnel training, simple data acquisition, and immediate predictions.

ABBREVIATIONS USED

Tomato, *Solanum lycopersicum* L.; ATR-IR, attenuated total reflectance infrared; HPLC, high-performance liquid chromatography; PLSR, partial least-squares regression; SECV, standard error of cross-validation; r, correlation coefficient; SIMCA, soft independent modeling of class analogy; PCA, principal component analysis; CV, coefficient of variability.

LITERATURE CITED

(1) Lucier, G.; Dettmann, R.; Jerardo, A. Vegetables and Melons Outlook, 2007; www.ers.usda.gov/Publications/VGS/2007/12Dec/VGS324.pdf.

(2) Gartner, C.; Stahl, W.; Sies, H. Lycopene is more bioavailable from tomato paste than from fresh tomatoes. *Am. J. Clin. Nutr.* **1997**, *66*, 116–122.

(3) Shi, J.; Le Maguer, M. Lycopene in tomatoes: chemical and physical properties affected by food processing. *Crit. Rev. Biotechnol.* **2000**, *20*, 293–334.
(4) Rao, A. V.; Agarwal, S. Role of antioxidant lycopene in cancer and heart disease. *J. Am. Coll. Nutr.* **2000**, *19*, 563–569.
(5) International Agency for Research on Cancer, World Health Organization (IARC). In *Carotenoids*; IARC Press: Lyon, France, 1998; Vol. 2.
(6) Periago, M. J.; Rincon, F.; Aguera, M. D.; Ros, G. Mixture approach for optimizing lycopene extraction from tomato and tomato products. *J. Agric. Food Chem.* **2004**, *52*, 5796–5802.
(7) Emenhiser, C.; Simonovic, N.; Sander, L. C.; Schwartz, S. J. Separation of geometrical carotenoid isomers in biological extracts using a polymeric C_{30} column in reversed-phase liquid chromatography. *J. Agric. Food Chem.* **1996**, *44*, 3887–3893.
(8) Nguyen, M. L.; Schwartz, S. J. Lycopene stability during food processing. *Proc. Soc. Exp. Biol. Med.* **1998**, *218*, 101–105.
(9) Davis, A. R.; Fish, W. W.; Perkins-Veazie, P. A rapid spectrophotometric method for analyzing lycopene content in tomato and tomato products. *Postharvest Biol. Technol.* **2003**, *28*, 425–430.
(10) Olives Barba, A. I.; Camara Hurtado, M.; Sanchez Mata, M. C.; Fernandez Ruiz, V.; Lopez Saenz de Tejada, M. Application of a UV−vis detection−HPLC method for a rapid determination of lycopene and β-carotene in vegetables. *Anal. Nutr. Clin. Methods* **2006**, *95*, 328–336.
(11) Bicanic, D.; Fogliano, V.; Luterotti, S.; Swarts, J.; Piani, G.; Graziani, G. Quantification of lycopene in tomato products: comparing the performances of a newly proposed direct photothermal method and high-performance liquid chromatography. *J. Sci. Food Agric.* **2005**, *85*, 1149–1153.
(12) Bhosale, P.; Ermakov, I. V.; Ermakova, M. R.; Gellermann, W.; Bernstein, P. S. Resonance Raman quantification of nutritionally important carotenoids in fruits, vegetables, and their juices in comparison to high-pressure liquid chromatography analysis. *J. Agric. Food Chem.* **2004**, *52*, 3281–3285.
(13) Halim, Y.; Schwartz, S. J.; Francis, D.; Baldauf, N. A.; Rodriguez-Saona, L. E. Direct determination of lycopene content in tomatoes (*Lycopersion esculentum*) by attenuated total reflectance infrared spectroscopy and multivariate analysis. *J. AOAC Int.* **2006**, *89*, 1257–1262.
(14) Pedro, A. M. K.; Ferreira, M. M. C. Nondestructive determination of solids and carotenoids in tomato products by near-infrared spectroscopy and multivariate calibration. *Anal. Chem.* **2005**, *77*, 2505–2511.
(15) Baranska, M.; Schutze, W.; Schulz, H. Determination of lycopene and β-carotene content in tomato fruit and related products: comparison of FT-Raman, ATR-IR, and NIR spectroscopy. *Anal. Chem.* **2006**, *78*, 8456–8461.
(16) Jha, S. N.; Matsuoka, T. Nondestructive techniques for quality evaluation of intact fruits and vegetables. *Food Sci. Technol. Res.* **2000**, *6*, 248–251.
(17) Coates, J. Interpretation of infrared spectra: a practical approach. In *Encyclopedia of Analytical Chemistry*; Meyers, R. A., Ed.; Wiley: Chichester, U.K., 2000; pp 10815−10837.
(18) Pike Technologies. ATR−Theory and Applications. Application Note, 2005; http://www.piketech.com/technical/application-pdfs/ATR_Theory& Appl.pdf.
(19) Harrick, N. J. *Internal Reflection Spectroscopy*; Harrick Scientific; New York, 1987.
(20) Udelhoven, T.; Naumann, D.; Schmitt, J. Development of a hierarchical classification system with artificial neural networks and FT-IR spectra for the identification of bacteria. *Appl. Spectrosc.* **2000**, *54*, 1471–1479.
(21) Kabelka, E.; Yang, W.; Francis, D. M. Improved tomato fruit color within inbred backcross line derived from *Lycopersicon esculentum* and *L. hirsutum* Involves the interaction of loci. *J. Am. Soc. Hortic. Sci.* **2004**, *129*, 250–257.
(22) Goodman, C. L.; Fawcett, S.; Barringer, S. A. Flavor, viscosity, and color analyses of hot and cold break tomato juices. *J. Food Sci* **2002**, *67*, 404–408.
(23) Pike Technologies. Application Note 0402, 2005; http://www.piketech.com/technical/application-pdfs/Diamond-Crystal-Plates.pdf.
(24) Bjorsvik, H. R; Martens, H. Data analysis: calibration of NIR instruments by PLS regression. In *Handbook of Near-Infrared Analysis*; Burns, D., Ciurczak, E., Eds.; Dekker: New York, 1992; pp 159−180.
(25) Martens, H.; Naes, T. *Multivariate Calibration*; Wiley, Chichester, U.K., 1989; pp 419.
(26) Lavine, B. K. Clustering and classification of analytical data. In *Encyclopaedia of Analytical Chemistry*; Meyers, R. A., Ed.; Wiley: Chichester, U.K., 2000; pp 9689−9710.
(27) De Maesschalck, R.; Candofi, A.; Masart, 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.
(28) De Maesschalck, R.; Jouan-Rimbaud, D.; Massart, D. L. The Mahalanobis distance. *Chemom. Intell. Lab Syst.* **2000**, *50*, 1–18.
(29) Gunzler, H.; Gremlich, H. U. Qualitative spectral interpretation. In *IR Spectroscopy: An Introduction*; Gunzler, H., Gremlich, H. U., Eds.; Wiley-VCH: Weinheim, Germany, 2002; pp 171−274.
(30) Guillen, M. D.; Cabo, N. Infrared spectroscopy in the study of edible oils and fats. *J. Sci. Food Agric.* **1997**, *75*, 1–11.
(31) Tonicci, L. H.; Holden, J. M.; Beecher, G. R.; Khachik, F.; Davis, C. S.; Mulokozi, G. Carotenoid content of thermally processed tomato-based food products. *J. Agric. Food Chem.* **1995**, *43*, 579–586.
(32) Wold, S. Pattern recognition by means of disjoint principal component models. *Pattern Recognit.* **1976**, *8*, 127–139.
(33) Welti, R.; Wang, X. Lipid species profiling: a high-throughput approach to identify lipid compositional changes and determine the function of genes involved in lipid metabolism and signaling. *Curr. Opin. Plant Biol.* **2004**, *7*, 337–344.
(34) Bertrand, B.; Villarreal, D.; Laffargue, A.; Posada, H.; Lashermes, P.; Dussert, S. Comparison of the effectiveness of fatty acids, chlorogenic acids, and elements for the chemometric discrimination of coffee (*Coffea arabica* L.) varieties and growing origins. *J. Agric. Food Chem.* **2008**, *56*, 2273–2280.

Received for review September 18, 2008. Revised manuscript received December 13, 2008. Accepted December 16, 2008. The project was supported by the National Research Initiative of the USDA Cooperative State Research, Education and Extension Service, Grant 2006-35503-17571.

JF802920Z