JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

J. Agric. Food Chem. **2007,** *55,* 231–236 **231**

Prediction of Japanese Green Tea Ranking by Gas Chromatography/Mass Spectrometry-Based Hydrophilic Metabolite Fingerprinting

WIPAWEE PONGSUWAN,[†,‡] EIICHIRO FUKUSAKI,[*,†] TAKESHI BAMBA,[†] TSUTOMU YONETANI,[§] TOSHIYAKI YAMAHARA,[§] AND AKIO KOBAYASHI[†]

Department of Biotechnology, Graduate School of Engineering, Osaka University, Japan, and Tea branch, Nara Prefecture Agricultural Experiment Station, Japan

An innovative technique for green tea's quality determination was developed by means of metabolomics. Gas-chromatography coupled with time-of-flight mass spectrometry and multivariate data analysis was employed to evaluate the quality of green tea. Alteration of green tea varieties and manufacturing processes effects a variation in green tea metabolites, which leads to a classification of the green tea's grade. Therefore, metabolic fingerprinting of green tea samples of different qualities was studied. A set of ranked green tea samples from a Japanese commercial tea contest was analyzed with the aim of creating a reliable quality-prediction model. Several multivariate algorithms were performed. Among those, the partial least-squares projections to latent structures (PLS) analysis with the spectral filtering technique, orthogonal signal correction (OCS), was found to be the most practical approach. In addition, metabolites that play an important role in green tea's grade classification were identified.

KEYWORDS: Metabolomics; metabolic fingerprinting; GC-TOF/MS; quality evaluation; PLS

INTRODUCTION

Tea is the most popular beverage, and it is made from the leaves of the *Camellia sinensis* plant. There are three main types of tea on the basis of length of oxidation reaction and how the leaves are processed; they are black tea, oolong tea, and green tea. Black tea is the most processed and has the greatest oxidation. Oolong tea is partially oxidized and dried. Green tea undergoes very little processing without oxidation. In green tea, the enzymes are inactivated at an onset temperature to prevent oxidation of the leaf polyphenols. Green tea is very popular in Southeast Asian countries, not only because of its effective pharmaceutical activity (*1*) but also because of its excellent taste. The market price ranges from $1 dollar to over $100 per 100 g (*2*). Green tea taste is determined by the kind of tea tree, plucking time, and the cultivation method (*3*). The brothy taste of the tea brew originates in amino acids, especially the unique theanine which accounts for about 60−70% of the amino acids in tea leaves (*4*). Its astringent taste is attributed to catechin (tannin) while its bitter taste is attributed to caffeine. Volatile compounds are the main contributors to differences in odors of teas of different grades (*5*−*6*). The cultivar differences, environmental effects, and methods of processing all suggest that the quasi-steady-state amounts of intermediate pathway metabolites and the end accumulation of terminal metabolites should also vary. To follow these changes, metabolites must be monitored both spatially and temporally (*7*−*9*). Sensory evaluation of tea quality has traditionally been assessed by highly trained specialists who evaluate product quality on the basis of leaves' appearance, aroma, color, and taste of the brew (*3*). Because it takes years of experience to acquire these skills, it would be advantageous to determine product quality by some form of nonhuman measurement. Chemical analysis seems to be the most reliable method for estimating the quality of green tea (*10*). From the combination of instrumental analytical methods with powerful computer-driven pattern recognition techniques, new capabilities for quality control and characterization of complex materials have been realized (*11*−*12*). The basic concept of this new approach is fast analysis by using chemical fingerprints instead of the characterization on a limited number of individual compounds. In chromatography, fingerprinting based on pattern recognition methods has been used in various analytical areas, for example, food and nutritional fields as process monitoring and control (grading of raw materials, routine online quality checks, or determining the process by which a product was made) (*13*) and geographical origin (determining sources of ingredients by chemical composition and tracing origin of finished products by flavor and aromatic components) (*3*).

This research is aimed at developing a fast and reliable model to determine quality/grade of green tea samples. A set of known-rank samples from the Kansai tea contest, Japan, was utilized to create a model with the aim of using this model to predict the quality of unknown samples. A combination of gas chromatography and mass spectrometry (GC/MS) allows the iden-

* Author to whom correspondence should be addressed. Tel/Fax: +81-6-6879-7424. E-mail: fukusaki@bio.eng.osaka-u.ac.jp.

[†] Osaka University.

[‡] The study represents a portion of the dissertation submitted by W.P. to Osaka University in partial fulfillment of the requirements for her Ph.D. degree.

[§] Nara Prefecture Agricultural Experiment Station.

10.1021/jf062330u CCC: $37.00
Published on Web 12/20/2006

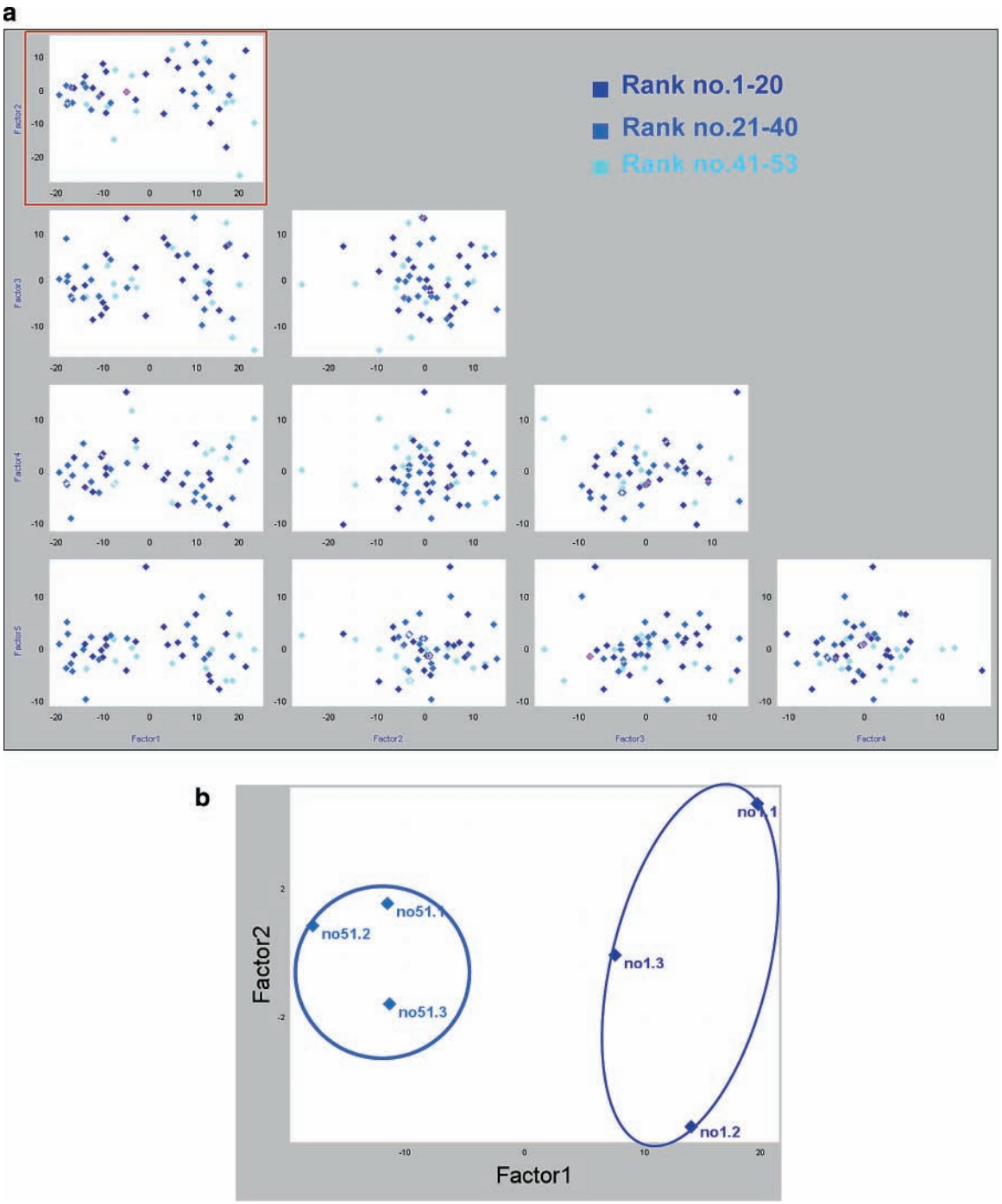

Figure 1. PCA analysis. (**a**) Score plot of all 53 ranked samples. Data were preprocessed with mean-center without transformation. Color spots are graded from dark blue to lighter blue in parallel with their ranks varying from highly ranked to lowly ranked samples. (**b**) Score plot of only the best and worst ranked samples. Once again, data were preprocessed with mean-center without transformation. The first-ranked samples are represented by dark blue while light blue represents the lowest ranked samples.

tification and robust quantification of several hundred metabolites within a single extract. Hydrophilic primary metabolites were primarily studied since there are stable protocols for machine setup and maintenance, sample preparation and analysis, and chromatogram evaluation and interpretation for these (*14*−*15*). In MS-based metabolomics studies, the identification of differences between samples has normally involved a variety of multivariate tools, such as principal component analysis (PCA), hierarchical cluster analysis (HCA) (*16*), projection to latent structures (PLS) (*17*), and discriminate analysis (*18*). PCA, which is probably the oldest and best known technique used for exploratory multivariate analysis, was first selected as the data visualization tool. Subsequently, partial least-squares projection to latent structures (PLS) was employed to verify the relationship between two groups of related variables, in this case, correlation between green tea's metabolites and its quality.

MATERIALS AND METHODS

Materials. The dried leaves after processing of 53 ranked first-crop tea samples (spring-harvested, called "Ichi-ban-cha" in Japanese) from the 2005 contest were analyzed. These tea samples enrolled in a commercial tea contest among the Kansai area were obtained from

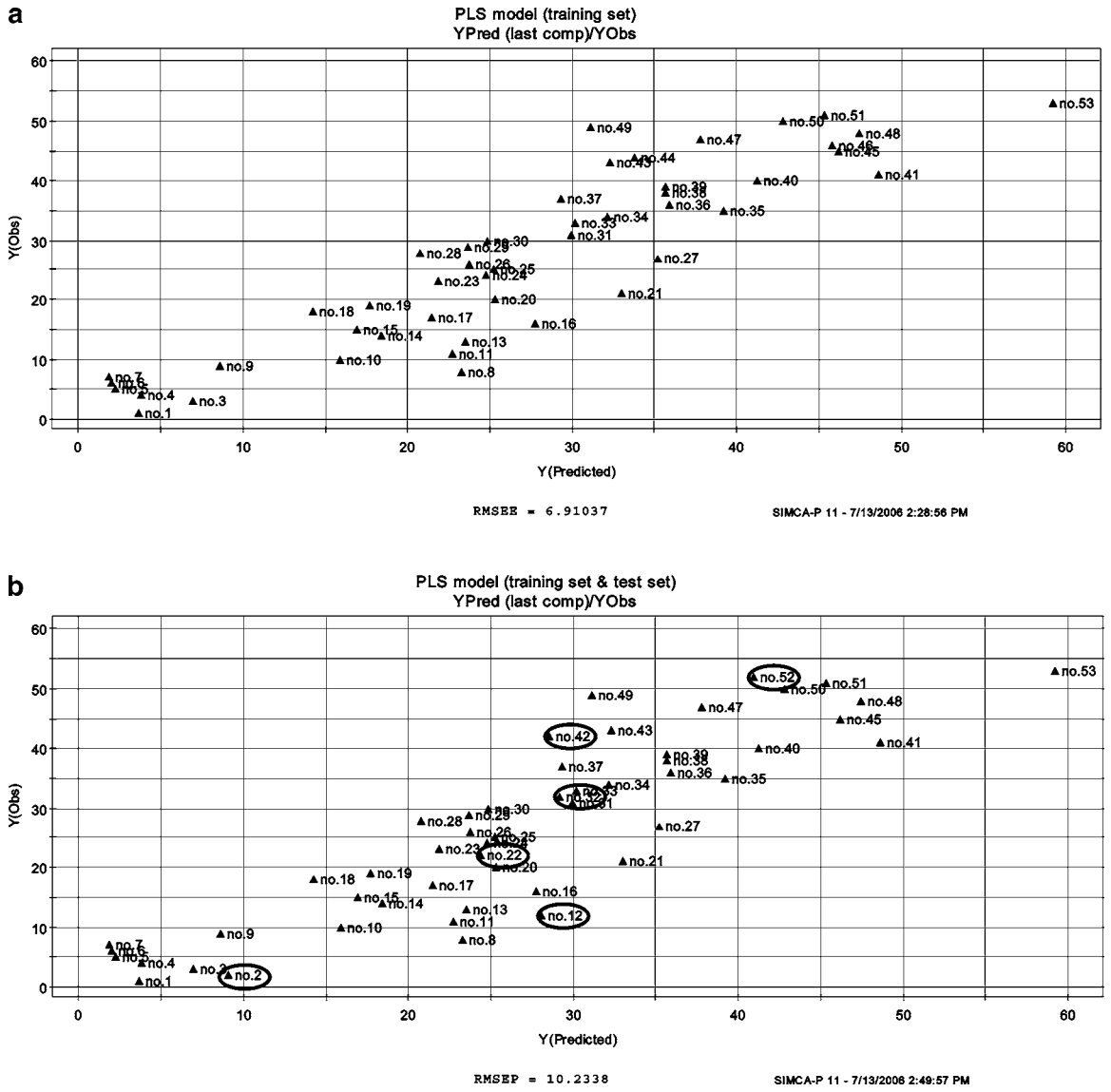

Figure 2. Relationship between measured and predicted green tea quality (ranking) of PLS model (**a**) for 46 green tea samples as a training set and (**b**) for all 53 ranks of both testing (marked by circles) and training sets.

the Tea branch of the Nara Prefecture Agricultural Experiment Station. Ranking of teas was determined by the total scores of the sensory tests, which are leaf appearance, smell, and color of the brew and its taste, judged by professional tea tasters.

Reagents. All chemical used in this study were analytical grade. Methanol and chloroform used as extraction solvents, ribitol (diluted with deionized water to a concentration of 0.2 mg/mL), and pyridine used as a solvent were purchased from Wako (Osaka, Japan). Methoxyamine hydrochloride was purchased from Sigma (WO, United States). *N*-Methyl-*N*-(trimethylsilyl)trifluoroacetamide (MSTFA) was purchased from GL Science, Inc. (Tokyo, Japan).

Sample Preparation for GC/MS Analysis. Dried tea leaves (30 mg) in 2-mL Eppendorf tubes were freeze-dried and ground with a Retsch ball mill (20 Hz, 1 min). Hydrophilic primary green tea metabolites were extracted using a single-phase solvent mixture of MeOH, H_2O, and $CHCl_3$ in ratio of 2.5/1/1 (v/v/v), respectively. The mixture was shaken with 60 μL ribitol used as an internal standard for 5 min and was centrifuged at 16 000*g*, 4 °C, for 3 min. Subsequently, 900 μL of the supernatant was transferred to a 1.5-mL Eppendorf tube. By adding 400 μL of water purified using a Millipore Milli-Q system (Berdford, MA) vortex and centrifuge, 400 μL of polar phase was then transferred to another 1.5-mL Eppendorf tube capped with pierced cap. The extract was dried in a vacuum centrifuge dryer until dryness (overnight).

For derivatization, 50 μL of methoxyamine hydrochloride in pyridine (20 mg/mL) was added as a first derivatizing agent. The mixture was incubated at 30 °C for 90 min. A second derivatizing agent, 100 μL of *N*-methyl-*N*-(trimethylsilyl)trifluoroacetamide (MSTFA), was added and incubated at 37 °C for 30 min. A 1 μL of sample was injected in split mode (25:1, v/v)

GC/MS Analysis. For the metabolites' identification and tea's grade classification, the gas chromatograph used in this study was a 689 CN (Agilent Co., Palo Alto, CA) equipped with a 30 m × 0.25 mm i.d. fused silica capillary column coated with 0.25-μm CP-SIL 8 CB low bleed (Varian Inc., Palo Alto, CA) coupled with a Pegasus III TOF mass spectrometer (LEGO, St. Joseph, MI) and a 7683B series injector (Agilent Co., Palo Alto, CA) as an autosampler. The injection temperature was 230 °C. The helium gas flow rate through the column was 1 mL/min. The column temperature was held at 80 °C for 2 min isothermally and then was raised by 15 °C/min to 330 °C and was held there for 6 min isothermally. The transfer line and the ion source temperatures were 250 °C and 200 °C, respectively. Ions were generated by a 70 kV electron impact (EI), and 20 scans per second were recorded over the mass range 85−650 m/z. The acceleration voltage was turned on after a solvent delay of 250 s.

Data Preprocessing. For data preprocessing, raw chromatographic data (Pegasus file, *.peg) were converted into ANDI files (Analytical Data Interchange protocol, *.cdf). With the ANDI format, the conver-

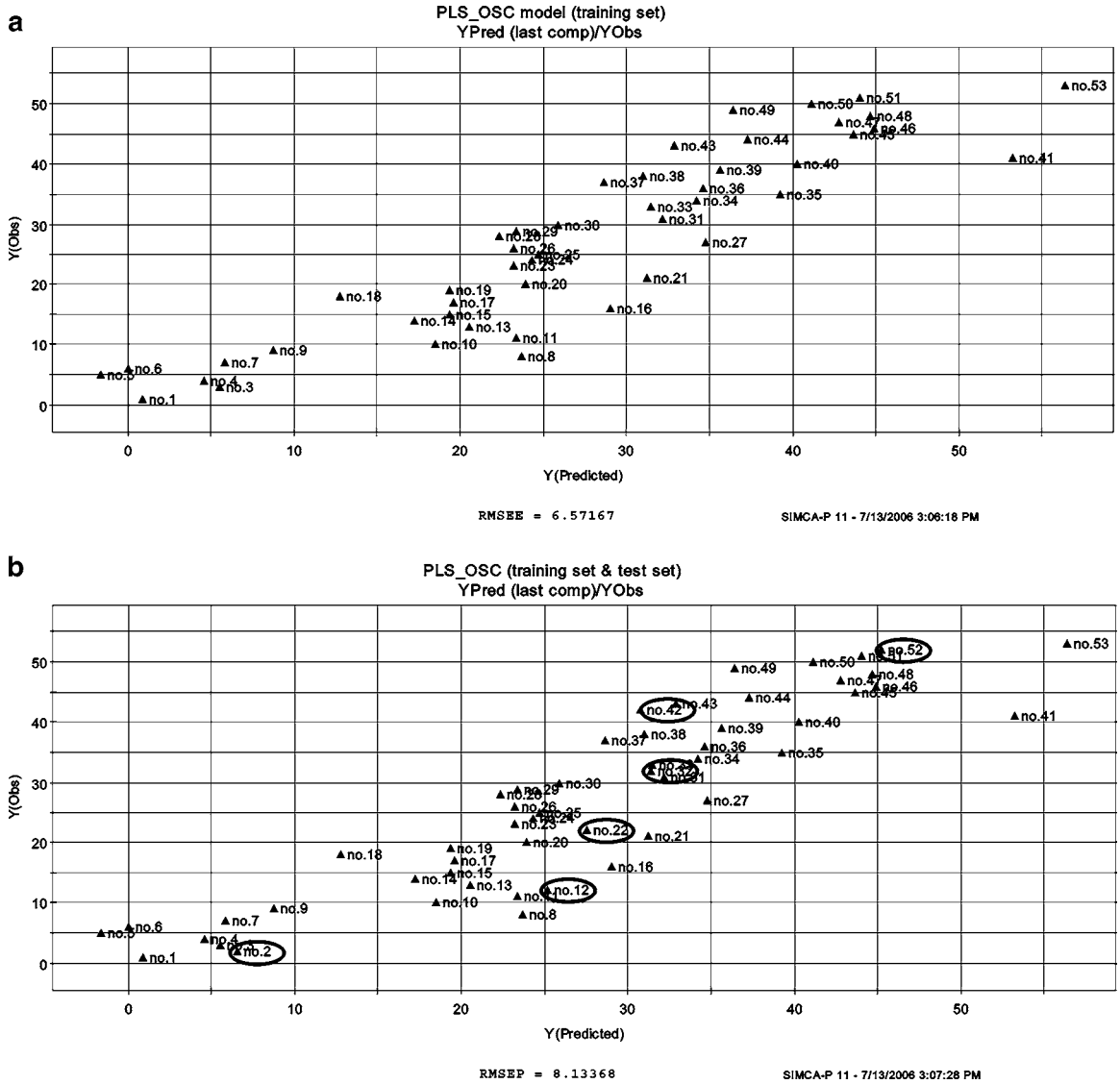

Figure 3. Relationship between measured and predicted green tea quality (ranking) of PLS model with orthogonal signal correction method (**a**) for 46 green tea samples as a training set and (**b**) for all 53 ranks of both testing (marked by circles) and training sets.

sion and transfer of data between different mass spectral data systems could be done. The converted files (ANDI) were subjected to a data preprocessing procedure: the data points were adjusted in detail. In addition, data transformation could also be done to achieve the best chromatographic data. The total ion chromatographic data were then extracted and saved as ANDI format files without fragment data. These files were imported to commercially available software, LineUp (Informatrix, Inc, WA), for multiple alignment of the retention times. The misaligned peaks were aligned using a correlation optimized warping algorithm (*19*).

Multivariate Analysis of GC/MS. Principle component analysis (PCA) was initially selected to comprehend the relationships expressed in terms of similarity or dissimilarity among groups of multivariate data. Commercially available software, Pirouette (Informatrix, Inc.), was applied for this purpose. Projections to latent structures by means of partial least square, PLS (SIMCA-P version 11.0, Umetrics, Umeå, Sweden), was then chosen to create a prediction model. PLS finds a relation between two sets of variables: observations and responses. Further, orthogonal signal correction (OSC) was used to remove unassociated data (*20, 21*−*23*).

Significant compounds were identified by comparing their mass spectra with those in libraries (the NIST library and an in-house library prepared from authentic standard chemicals). Moreover, the library provided by the Max-Planck Institute of Molecular Plant Physiology, Germany was also used for this purpose (http://www.mpimp-golm-.mpg.de/mms-library/index-e.html).

RESULTS AND DISCUSSION

Metabolite Fingerprinting with PCA. The 53 ranked samples were divided into three groups ranked according to their grades. It was found that the classification of these three groups was not distinct (**Figure 1a**). Highly ranked samples did not show an obvious cluster separate from the lowly ranked samples since the variation in each group of samples was too high as there were about 20 samples in each defined group. However, an additional analysis was done with only the best and the worst ranked samples. It was found that some discrimination did exist between these two groups (**Figure 1b**). The first ranked sample consisted of a greater amount of amino acids, quinic acid, phosphoric acid, ribose, and arabinopyranose while the key metabolites for the lowest ranked sample were mainly sugars (fructose, glucose, and mannose). This illustrated that a metabolomics analysis could be useful in the quality determination of green tea.

Quality-Predictive Model. *Metabolite Fingerprinting with Projection to Latent Structure PLS.* PLS was considered as an algorithm to create a quality prediction model for green tea. A relationship between sample's metabolite profiling (matrix *X*) and its quality (matrix *Y*) was observed. Samples were divided into groups of training and testing sets. The samples ranked

Table 1. Primary Metabolites Detected by GC-TOF/MS from Japanese Green Tea

name	retention time[a]	mass fragment[b,c]	derivatized[d]
		Organic Acids	
oxalic acid	5.68	**116**, 147, 190, 218	TMS(×2)
phosphoric acid	7.11	133, 211, **299**, 314	TMS(×3)
succinic acid	7.56	**147**, 172, 247	TMS(×2)
malic acid	9.1	**147**, 189, 233, 245, 307, 335	TMS(×3)
shikimic acid	11.53	**117**, 147, 292	TMS(×1)
citric acid	11.63	**147**, 273, 347, 375, 465	TMS(×4)
quinic acid	11.88	**147**, 255, 345, 436	TMS(×1)
		Amino Acids	
serine	7.96	100, 147, 188, **204**, 218, 278, 306	TMS(×3)
threonine	8.20	**117**, 147, 219, 291, 320	TMS(×3)
aspartic acid	9.38	100, 147, 218, **232**, 306, 334	TMS(×3)
pyroglutamic acid	9.48	**156**, 230, 258	TMS(×2)
threonic acid	9.65	**147**, 220, 292, 319	TMS(×4)
glutamic acid	10.18	128, 147, 156, **246**, 348	TMS(×3)
glutamine1	11.00	**147**, 203, 227, 317	TMS(×4)
theanine	11.28	**159**, 183, 273, 285	
glutamine2	11.36	**156**, 203, 245, 347	TMS(×3)
		Sugars	
fructose1	11.95	**103**, 147, 217, 307, 364	TMS(×5)
fructose2	12.01	**103**, 217, 307, 364	TMS(×5)
glucose MEOX2	12.13	**147**, 160, 205, 217, 319	TMS(×5)
mannose	12.28	103, **147**, 205, 217, 319	TMS(×5)
sucrose	16.36	**217**, 271, 319, 438	TMS(×8)
arabinopyranose	16.60	147, **217**, 343, 434, 479	TMS(×4)
		Others	
caffeine	12.19	109, **194**	
inositol	13.43	**147**, 191, 217, 305, 318	TMS(×6)
silane	17.50	179, 355, **368**, 488	TMS(×5)

[a] Retention time (min). [b] Ions in boldface indicate the most intense product ion. [c] Lists of first five ions with the highest intensity. [d] Number of hydrogen atoms derivatized.

2nd, 12th, 22nd, 32nd, 42nd, and 52nd were excluded as a testing set for model validation. None of the variables were transformed; all of them were centered and scaled to Pareto variance to decrease chromatographic data noise effects (*24*). The model complexity, that is, the number of latent factors in the PLS model, can be determined by cross-validation. The optimum number could be found at the balance between fit (to the model) and predictive ability. Additionally, the PLS model is validated with a test set in which the root mean squared error of prediction (RMSEP) is computed. Two significant components were extracted, describing 80.7% of the variation in Y ($R2Y = 0.807$) and predicting 31.3% of the variation in Y ($Q2Y = 0.313$) according to cross-validation. The test set was subsequently predicted into the PLS model resulting in a predictive accuracy for the test samples (RMSEP = 10.23) over the model estimations on the basis of training set samples (RMSEP = 6.91) (**Figure 2**). The predictive ability of this PLS model was rather poor. This might be due to interference by a set of X variables that are not associated in the prediction of Y-values. Such variables cause imprecise predictions and also affect the robustness of the model (*21*).

Metabolite Fingerprinting with Projection to Latent Structure Assisted by Orthogonal Signal Correction (PLS_OSC). To enhance the predictive power of the multivariate calibration model, a spectral filtering technique was applied. OSC is a PLS-based solution that removes X-data variation that is unrelated to the modeling of Y. With OSC, one component is removed at a time from matrix X using the nonlinear iterative partial least-squares (NIPALS) algorithm (*21*−*22*). By removing two OSC components from the prior PLS model, $Q2Y$ was increased to 0.500. The remaining sum of squares was 41.40%. Hence, 58.60% of the variation in X was not related to Y and was removed. Prediction after subjecting a test set into the model now made the predictive accuracy for the test samples (RMSEP = 8.13) to be in a good agreement with the model estimations on the basis of the training samples (RMSEE = 6.57) (**Figure 3**). Furthermore, variables with high relevance for explaining Y were also identified from VIP (variable importance in the projection) values. Large VIP values, more than 1, are the most

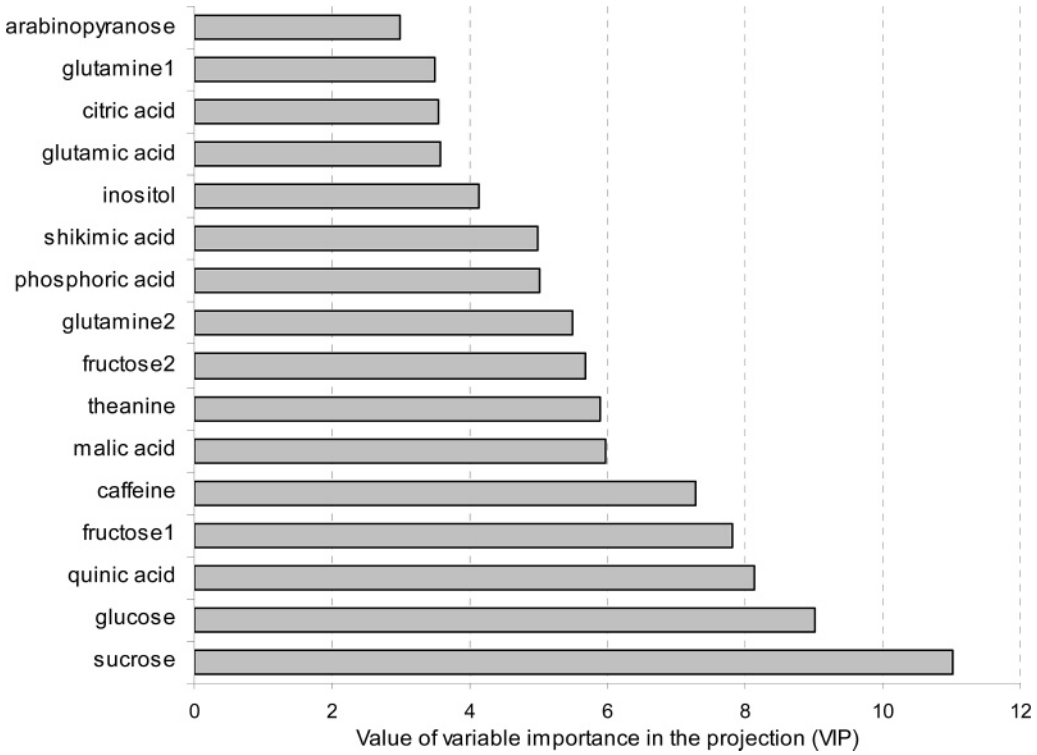

Figure 4. Bar chart showing influence of variables used to create a quality predictor for green tea (*Y*-axis is value of variable importance in the projection, VIP).

relevant for explaining *Y*. Quinic acid, amino acids (especially theanine), and groups of sugars were found to be significant in creating a quality prediction model for green tea (**Figure 4**).

Green Tea Metabolic Profiling. Green tea samples were analyzed to acquire information about the main metabolites of green tea dried leaves. Yamatocha, from the Tea branch of the Nara Prefecture Agricultural Experiment Station, which is accepted as a high-grade tea, was used. Analytical materials and methods were the same as above. Since resolution of chromatogram is satisfactory, peak areas of each metabolite were calculated by total ion count (TIC). Subsequently, every peak of identified metabolite was normalized to the internal standard, ribitol. From the analysis, it was found that dried leaves of green tea comprised many compounds, mainly organic acids, amino acids, and sugars (both monosaccharides and disaccharides), as shown in **Table 1**. Compounds detected in the highest amounts were sugars, especially sucrose and fructose. Theanine, which is the key amino acid contributing to the unique taste of tea, was also detected. Several other amino acids could be identified in the extract but in very small amounts. Many compounds still remain unidentified (secondary metabolites). About 23 identified metabolites were identified from the chromatogram.

In summary, this study demonstrates that metabolomics analysis using GC/MS combined with chemometrics provides some useful information in the study of green tea. It provides information on all interested metabolites in a single-run analysis whereas earlier analysis focused on specific groups of compounds that were believed to be significant for certain functions such as antioxidant compounds for medical purposes and catechins and other phenolic compounds for taste and aroma. Since characteristics of green tea might not be derived from single metabolites but from combinations, metabolomics is an excellent solution. This study also illustrates that green tea can be added to the list of products for which chromatography and spectrometry constitute a consistent quick and informative screening technique. In addition, metabolite fingerprinting also allowed the assessment of the quality of tea without standard samples. Nevertheless there may still be some problems to be overcome, since for green tea fingerprinting, an important factor in discriminating tea grades is the quantity of metabolites. Therefore, we need to ensure that these differences derive from variation of the samples themselves. Experimental repeatability must be robustly ensured. Homogeneity of samples is also important since there is a wide variation in components among leaf parts.

LITERATURE CITED

(1) Kuroda, Y.; Hara, Y. Antimutagenic and anticarcinogenic activity of tea polyphenols. *Mutat. Res.* **1999**, *436*, 69−97.

(2) Horie, H.; Fukatsu, S.; Mukai, T.; Goto, T.; Kawanaka, M.; Shimohara, T. Quality evaluation on green tea. *Sens. Actuators, B* **1993**, *13*−*14*, 451−454.

(3) Gall, G. L.; Colquhoun, I. J.; Defernez, M. Metabolite profiling using ^{1}H NMR spectroscopy for quality assessment of green tea, *camellia sinensis* (L.). *J. Agric. Food Chem.* **2004**, *52*, 692−700.

(4) Shu, K.; Kenji, K.; Hideki, M.; Andrea, H.; Thomas, H. Molecular and sensory studies on the umami teast of Japanese green tea. *J. Agric. Food Chem.* **2006**, *54*, 2688−2694.

(5) Shimoda, M.; Shigemetsu, H.; Shiratsuchi, H.; Osajima, Y. Comparison of volatiles compounds among different grades of green tea and their relations to odor attributes. *J. Agric. Food Chem.* **1995**, *43*, 1621−1625.

(6) Kato, M.; Shibamoto, T. Variation of major volatile constituents in various green teas from Southeast Asia. *J. Agric. Food Chem.* **2001**, *49*, 1394−1396.

(7) Goodacre, R.; Vaidyanathan, S.; Dunn, W. B.; Harrigan, G. G.; Kell, D. B. Metabolomics by numbers: acquiring and understanding global metabolite data. *Trends Biotechnol.* **2004**, *22*, 245−252.

(8) Fiehn, O. Metabolomics − the link between genotypes and phenotypes. *Plant Mol. Biol.* **2002**, *48*, 155−171.

(9) Scholz, M.; Gatzek, S.; Sterling, A.; Fiehn, O.; Selbig. J. Metabolite fingerprinting: detecting biological features by independent component analysis. *Bioinformatics* **2002**, *18*, 241−248.

(10) Horie, H.; Kohata, K. Application of capillary electrophoresis to tea quality estimation. *J. Chromatogr., A* **1998**, *802*, 219−223.

(11) Whitfield, P. D.; German, A. J.; Noble, P-J. M. Metabolomics: and emerging post-genomic tool for nutrition. *Br. J. Nutr.* **2004**, *92*, 549−555(7).

(12) Kuiper, H. A.; Kleter, G. A.; Noteborn, H. P.; Kok, E. J. Assessment of the food safety issues related to genetically modified foods. *Plant J.* **2001**, *27*, 503−28.

(13) Hu, Y.; Meng, Q.; Liu, Y.; Jiang, S. Methods for quality control of fingerprint chromatograms in traditional Chinese medicine. *Sepu* **2004**, *22*, 361−5.

(14) Fernie, A. R.; Trethewey, R. N.; Krotzky, A. J.; Willmitzer, L. Metabolite profiling: from diagnostics to systems biology. *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 1−7.

(15) Krishnan, P.; Kruger, N. J.; Ratcliffe, R. G. Metabolite fingerprinting and profiling in plants using NMR. *J. Exp. Bot.* **2004**, *56*, 255−265.

(16) Sumner, L. W.; Mendes, P.; Dixon, R. A. Plant metabolomics: large-scale phytochemistry in the functional genomics era. *Phytochemistry* **2003**, *62*, 817−836.

(17) Jonsson, P.; Gullberg, J.; Nordstrom, A.; Kusano, M.; Kowalczyk, M.; Sjostrom, M.; Moritz, T. A Strategy for identifying differences in large series of metabolomic samples analyzed by GC/MS. *Anal. Chem.* **2004**, *76*, 1738−1745.

(18) Allen, J.; Davey, H. M.; Broadhurst, D.; Heald, J. K.; Rowland, J. J.; Oliver, S. G.; Kell, D. B. High-throughput classification of yeast mutants for functional genomics using metabolomic footprinting. *Nat. Biotechnol.* **2003**, *21*, 692−696.

(19) Neilsen, N. P. V.; Carstensen, J. M.; Smedsgaard, J. Aligning of Single and Multiple Wavelength Chromatographic Profiles for Chemometric Data Analysis using Correlation Optimised Warping. *J. Chromatogr., A* **1998**, *805*, 17−35.

(20) Wold, S.; Antti, H.; Lindgren, F.; Ohman, J. Orthogonal signal correction of near-infrared spectra. *Chemom. Intell. Lab. Syst.* **1998**, *44*, 175−185.

(21) Eriksson, L.; Trygg, J.; Johansson, E.; Bro, R.; Wold, S. Orthogonal signal correction, wavelet analysis, and multivariate calibration of complicated process fluorescence data. *Anal. Chim. Acta* **2000**, *420*, 181−195.

(22) Wold, S.; Trygg, J.; Berglund, A.; Antti, H. Some recent developments in PLS modeling. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 131−150.

(23) Trygg, J.; Wold, S. Orthogonal projection to latent structures (O-PLS). *J. Chemom.* **2002**, *16*, 119−128.

(24) van den Berg, R. A.; Hoefsloot, H. C. J.; Westerhuis, J. A.; Smilde, A. K.; van der Werf, M. J. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* **2006**, *7*, 142.

Received for review August 12, 2006. Revised manuscript received October 21, 2006. Accepted November 12, 2006.

JF062330U