JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Quality Prediction of Japanese Green Tea Using Pyrolyzer Coupled GC/MS Based Metabolic Fingerprinting

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

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

A couple between pyrolyzer and gas chromatography/mass spectrometry (GC/MS) has allowed a fast, simple, and low-cost approach to evaluate a quality of Japanese green tea without any sample preparation or derivatization techniques. Using our method, errors from sample preparation could be avoided since raw samples were directly extracted through the extreme heat of the pyrolyzer. In addition, undesired reactions from expensive derivatizing agents, which are commonly needed to treat the samples before analyzing with GC/MS, could be omitted. In order to illustrate the efficiency of this technique, a set of green tea samples from the Tea contest in 2005 in the Kansai area were used. Projection to latent structure by means of partial least squares (PLS) along with orthogonal signal correction (OSC) was selected to explain the relation between green tea's metabolite profiling and its quality. The quality of the model was validated by testing and comparing the predictive ability to the respective model.

KEYWORDS: Metabolic fingerprinting; PY-GC/MS; quality of green tea; prediction model

INTRODUCTION

Metabolomic research has been proven to be a helpful tool offering valuable information of use in many fields (*1–3*). It reveals that there is additional information in system integration. By aiming at identification and quantification of multiple targets, an overview of compound classes could be achieved. In the medicinal research field, metabolomics has been used to search for biomarkers that function as indicators of alteration from disease, disorder of system, or toxicology (*4*). In plant metabolomics, a large number of compounds were investigated, in which their levels can be regarded as the response of biological systems to genetic or environmental changes (*5, 6*).

Our research employs techniques in metabolomics to be a powerful approach in green tea's quality prediction. In Japan, a tea contest is held annually in order to appraise the price of green tea since product quality could be varied for each cultivation year from environmental changes, processing methods, and postharvested treatment. Green tea quality has been evaluated mainly by human sensory investigation based on taste, aroma, and appearance of dried leaves and brew of the tea. Apart from personal proficiency, it takes several years of specialized training to become a professional tea taster. Recently, many instrumental measurements and analysis have been introduced to evaluate and determine the quality of tea by focusing on a specific group compound, such as amino acids (*7, 8*), polyphenols (*9, 10*), and volatile flavor compounds (*11, 12*). By those techniques, the relationship between the content of targeted compounds and tea's quality has been examined. We believe that the characteristic of tea results from the whole metabolites composed in tea leaves rather than a single compound. Therefore, techniques in metabolomics are in our consideration. Instead of investigating a certain profile (metabolite identification and quantification), metabolic fingerprinting, which is a pattern comparison and classification, is a selected approach.

In our previous work, an effective and reliable method for Japanese green tea's quality determination based on its primary hydrophilic metabolite profile using gas chromatography (GC)/mass spectrometry (MS) has been reported (*13*). Although the great sensitivity and high selectivity have claimed GC/MS to be one of the most commonly used techniques for plant metabolomics, there are some difficulties in the operation. Since a plant consists of vast chemical diversity, sample extraction and derivatization are required to prepare those compounds for analysis by GC (*14, 15*). Additionally, the sample preparation step could lead to errors including some loss in metabolites. Further, derivatizing agents are expensive and cause uncontrollable side reactions (*16*).

Consequently, a new methodology has been developed; a technique of pyrolyzer connected with GC/MS was employed. Hyphenated PY-GC/MS permits an analysis of samples, which were previously unsuitable for analysis without lengthy extractions or derivatizations. Pyrolysis works by applying heat greater

* Corresponding author. Tel/Fax: +81-6-6879-7424. E-mail: fukusaki@bio.eng.osaka-u.ac.jp.
[†] Osaka University.
[‡] Nara Prefecture Agricultural Experiment Station.

10.1021/jf072791v CCC: $40.75
Published on Web 01/09/2008

Table 1. Compounds Detected by PY-GC/MS from Japanese Green Tea (Dried Leaves)

no.[a]	RT (min)[b]	compound	mass fragments[c,d]	no.[a]	RT (min)[b]	compound	mass fragments[c,d]
1	6.28	toluene	51, 65, **91**	15	23.22	phenol	55, 66, **94**
2	8.15	1-ethyl-1*H*-pyrrole	53, 67, 80, **95**	16	24.15	benzenepropanenitrile	65, **91**, 107, 131
3	9.43	styrene	51, 63, 78, **104**	17	24.50	4-methylphenol	51, 77, 90, **107**
4	12.71	acetic acid	**60**	18	26.08	2-ethylphenol	51, 77, **107**, 122
5	13.25	furfural	**95**	19	26.51	4-ethyl-2-methoxyphenol	51, 77, 107, **135**, 150
6	14.17	pyrrole	**67**, 95	20	27.51	2,6-dimethoxyphenol	51, 65, 93, 139, **154**
7	14.56	propanoic acid	57, **74**	21	29.29	2,3-dihydrobenzofuran	51, 65, 91, **120**
8	14.68	3-methyl-2-cyclopenten-1-one	53, 67, 81, **96**	22	29.76	3-pyridinol	67, **95**
9	14.92	3-methyl-1*H*-pyrrole	53,**80**	23	30.27	indole	63, 90, **117**
10	16.95	2-furanmethanol	53, 69, 81, **98**	24	30.49	2,5-pyrrolidinedione	56, **99**
11	19.15	1,2-cyclopentanedione	55, 69, **98**	25	37.56	hexadecanoic acid	55, 57, 60, 71, **73**, 129, 213, 256
12	20.25	3-methyl-2-cyclopentenanedione	55, 69, 83, 97, **112**	26	40.02	hydroquinone	55, 81, **110**
13	20.86	2-methoxyphenol	53, 81, **109**, 124	27	45.01	caffeine	55, 67, 82, 109, **194**
14	22.19	benzyl nitrile	60, 73, 90, **117**				

[a] Number of compounds listed with respect to the peak order of the pyrogram from **Figure 1**. [b] RT = retention time. [c] Ions in boldface type indicate the most intense product ion. [d] Lists of the first five ions with the highest intensity.

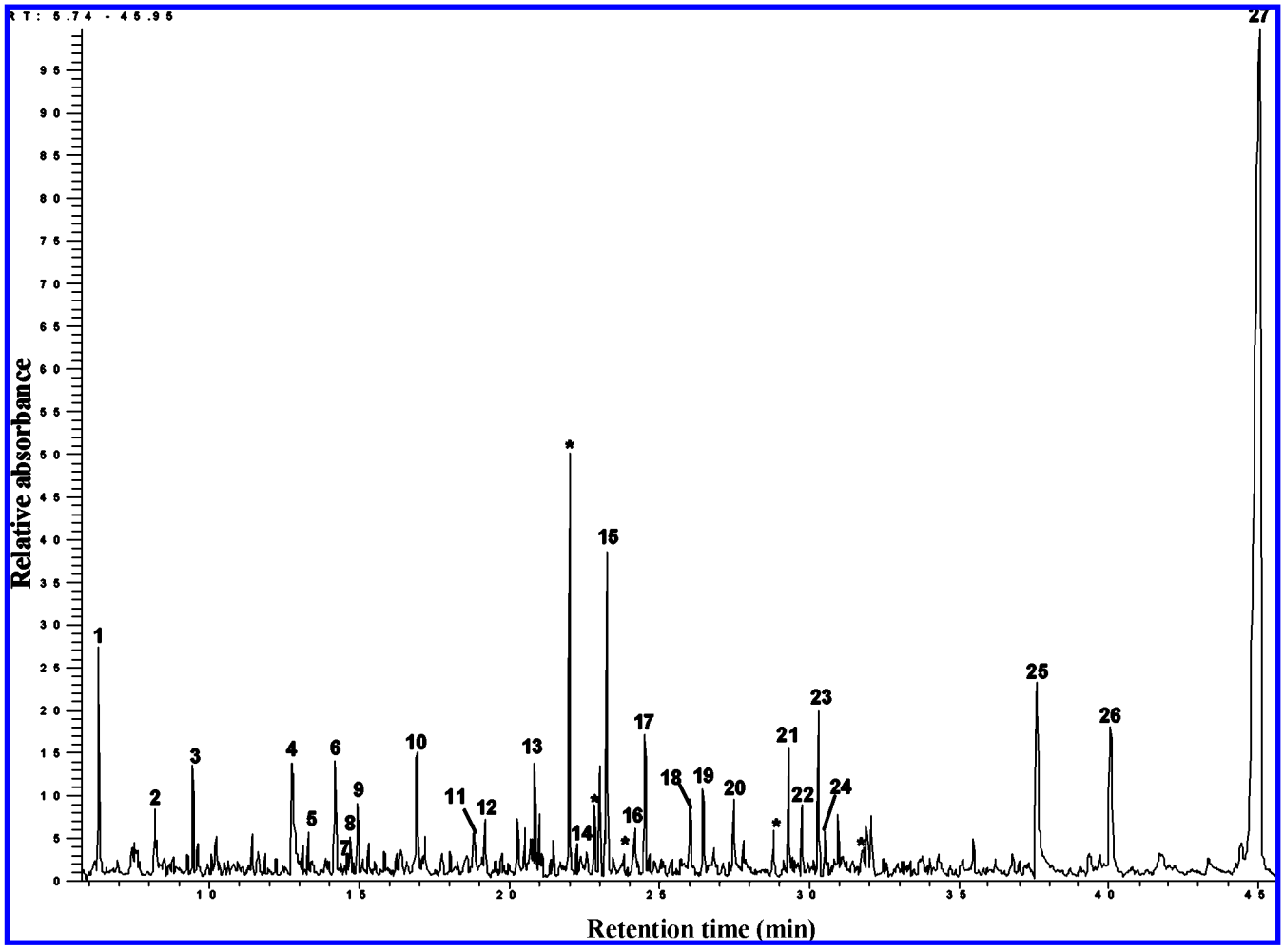

Figure 1. Pyrogram of Japanese green tea dried leaves. The peaks identified were numbered orderly with respect to retention time, whereas unidentified peaks were marked with an asterisk.

than the energy of specific bonds so that the molecule will fragment in a reproducible way. The fragments are then separated by the capillary column of the GC to produce the chromatogram (pyrogram), which contains both qualitative and quantitative information. Pyrolysis is also useful in application to macromolecular and other nonvolatile components. In this paper, a simple, fast, and reliable green tea quality estimation model constructed by metabolic fingerprinting was illustrated. Projection to latent structure (PLS) was a multivariate analysis employed to estimate a relationship between green tea's metabolite profile and its quality. Moreover, orthogonal signal correction (OSC) was introduced with the intention of removing chromatographic data that are not related in the sample's quality prediction (*17, 18*). The model was proven to be consistent by the model validation of the test set samples.

MATERIALS AND METHODS

Materials. According to our previous report (*13*), the same batch of samples were studied to compare the results. All 53 ranked green

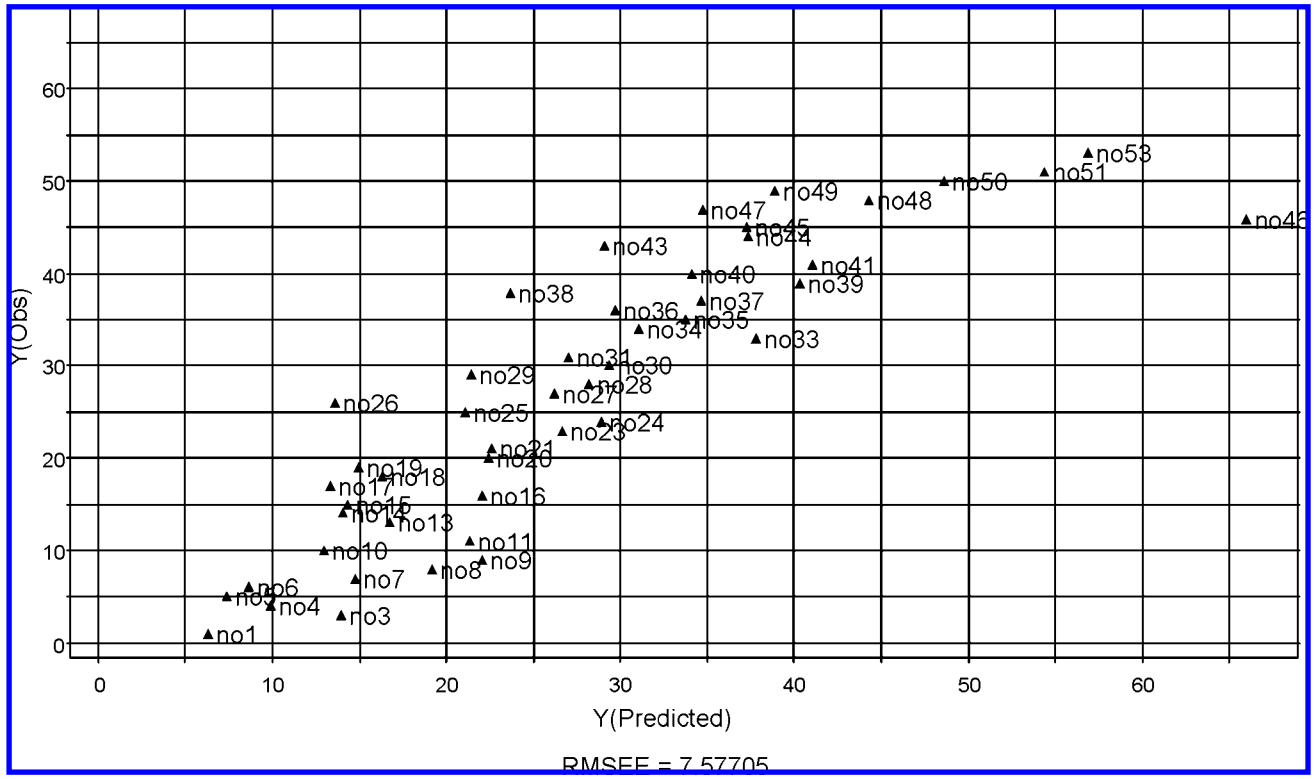

Figure 2. Relationship between measured and predicted green tea quality (ranking) of the PLS model for 47 green tea samples as a training set, with an overestimated prediction value of sample 46.

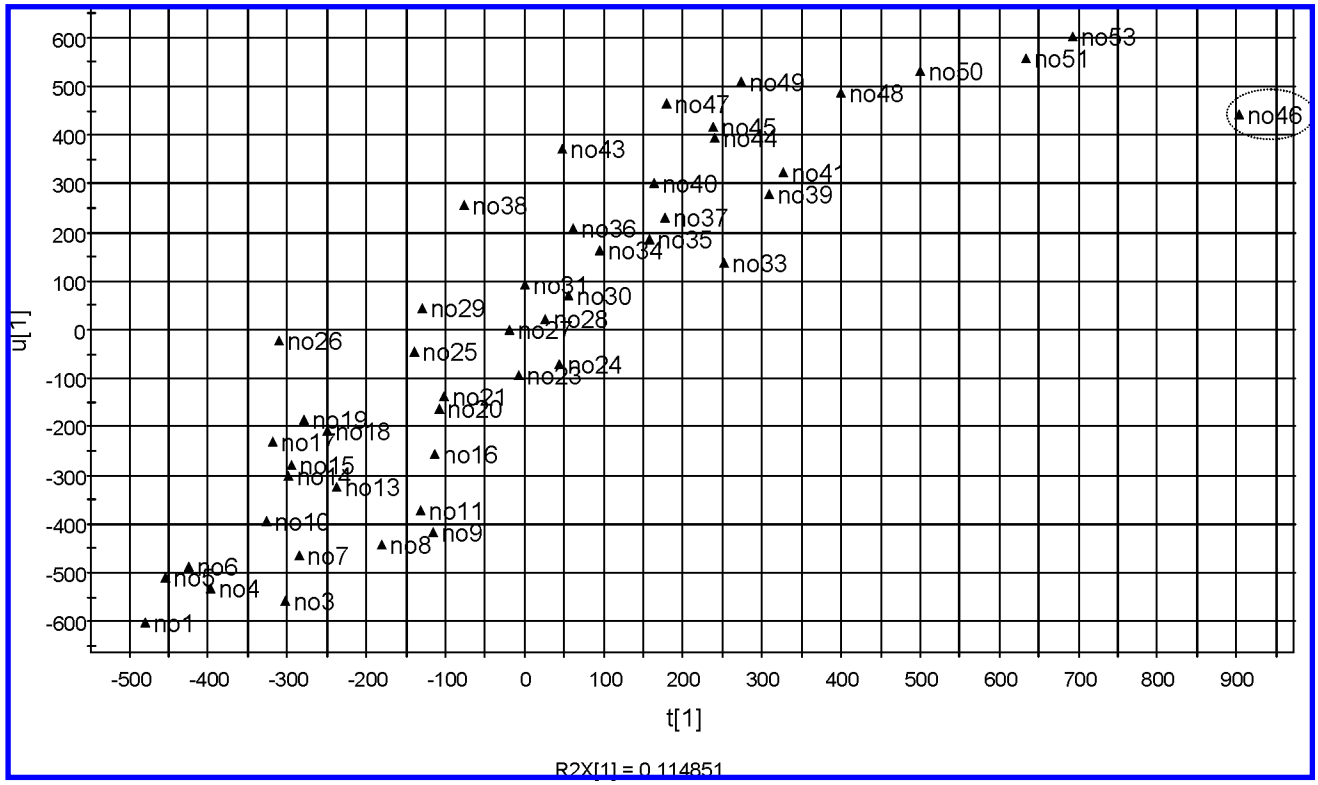

Figure 3. Score plot between t_1/u_1 of the model indicating a correlation between the predictors (*X*-data) and the responses (*Y*-data) showing an outlier of sample 46.

tea samples from the Tea contest of year 2005 of the Kansai area were obtained from the Tea Branch of Nara Prefecture Agricultural Experiment Station. Ranking of these samples was determined by the total scores of leaf appearance, smell, color of the brew, and its taste, judged by professional tea tasters.

Sample Preparation and PY-GC/MS Analysis. All 53 ranked green tea samples were ground into powder with a blender for the homogeneity of leaves and stalks of samples. The samples were kept constant in a −20 °C freezer until analysis. Samples were analyzed in random order to avoid any bias.

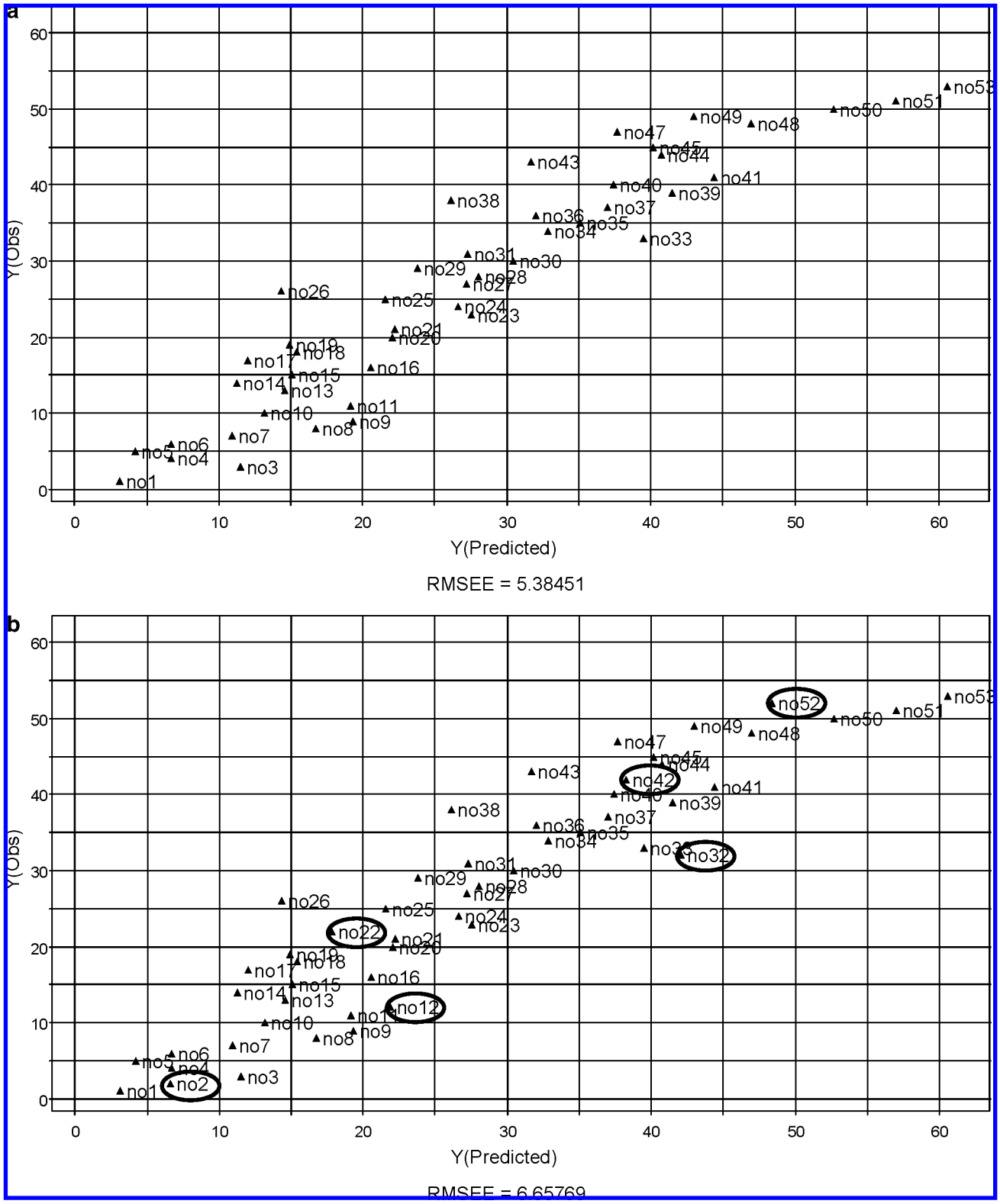

Figure 4. Relationship between measured and predicted green tea quality (ranking) of the PLS model (**a**) for 46 green tea samples as a training set and (**b**) for all ranks of both testing (marked by a circle) and training sets.

Pyrolyzer Operation. Ground green tea (1 ± 0.005 mg) was weighed and put into an Eco-cup S (Frontier Laboratory, Japan) before being introduced to the PY-2020iD pyrolyzer (Frontier Laboratory, Japan). The oven temperature of the pyrolyzer was set at 500 °C and the interface at 250 °C. The pyrolysis time was set to be 5 min for every sample.

GC/MS Analysis. The gas chromatograph was connected directly to the pyrolyzer through the inlet. The TRACE GC gas chromatograph (Thermo Electron Co., Waltham, MA) equipped with a THERMO TR-WaxMS column (60 m × 0.25 mm i.d. × 0.25 μm) (Thermo Electron Co., Waltham, MA) coupled with a TRACE DSQ mass spectrometer (Thermo Electron Co., Waltham, MA) was the selected system for the study. The injection temperature was set to 230 °C. High-purity helium was used as the carrier gas at a flow rate of 1 mL/min. The split ratio of the GC inlet was 20:1. The analysis was performed at the following oven temperature program: 70 °C for 2 min isothermally, then raised to 260 °C by 7 °C/min ramp, and finally kept at 260 °C for 20 min. Both transfer line and the ion source temperatures were set to 250 °C.

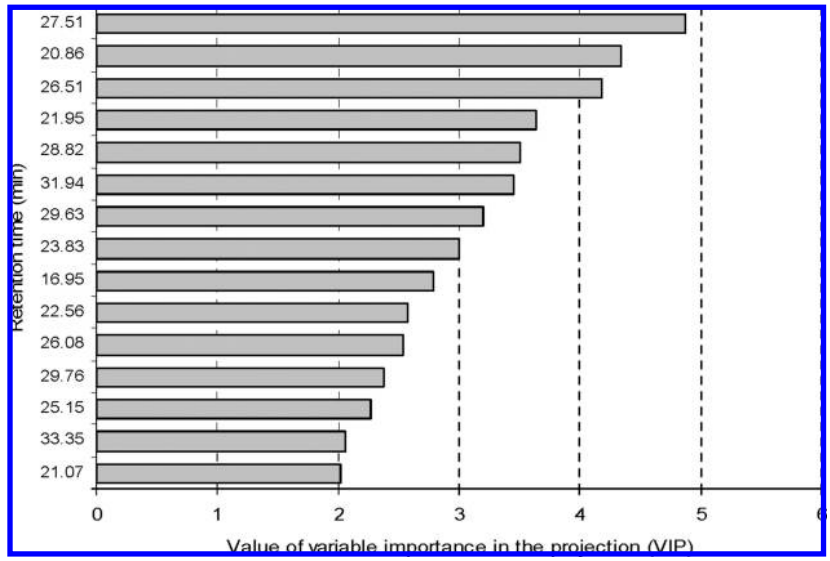

Figure 5. Bar chart showing the influence of variables used to create a quality predictor for green tea. (*Y*-axis is the value of variable importance in the prediction, VIP. *X*-axis shows significant metabolites used in the model creation marked by their retention time.)

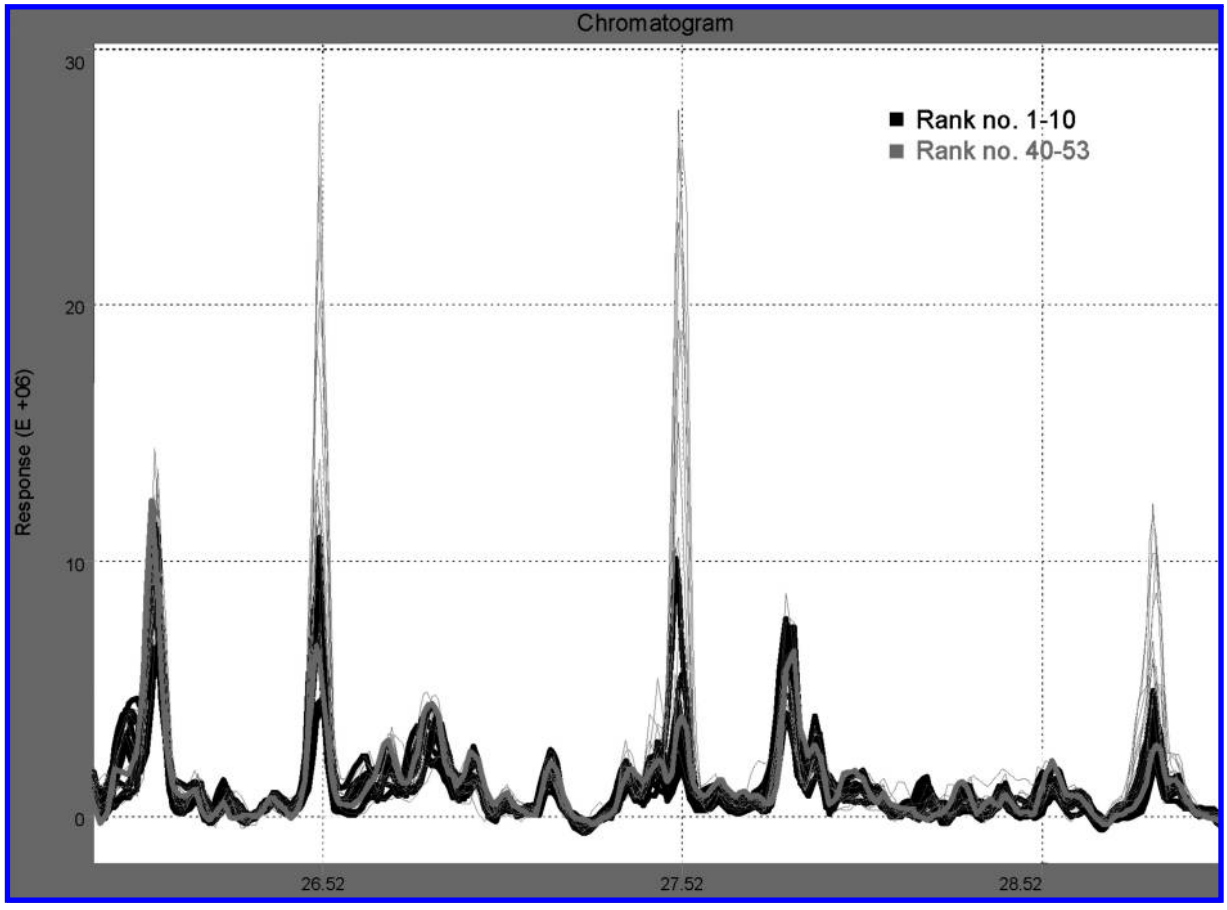

Figure 6. Differences in chromatogram intensities of significant metabolites of high-ranked green tea (colored in black) resulting in variation in green tea's quality. (Only a considerable range of the chromatogram was selected and shown here.)

Ions were generated by a 70 kV electron impact (EI), and 20 scans per second with a scanning range of *m*/*z* 50–650 were recorded. The acceleration voltage was turned on for all GC run time.

For instrumental validation, repeatability of the system was verified by analysis of high- and low-quality green tea samples (samples 1 and 53, $n = 3$) on three separate days. The area ratio of 27 identified peaks was calculated and revealed a maximum at about 10% of the relative standard deviation (% RSD) for all tested samples.

Data Pretreatment. Nonprocessed chromatographic data (Xcalibur file type, *.raw) were converted to ANDI files (Analytical Data Interchange protocol, *.cdf). Data points of the chromatograms were automatically adjusted to scan intensity for every 0.01 min before saving as AIA files without fragment data. Then the data were imported to the commercially available software LineUp (Informatrix, Inc.) for chromatographic retention time alignment using the correlation optimized warping algorithm (*19*).

Multivariate Analysis. Projections to latent structures by partial least squares (PLS) (SIMCA-P version 11.0) (Umetrics, Umeå, Sweden) with the spectral filtering technique, orthogonal signal correction (OSC), was used to create a regression model to examine a correlation between green tea's metabolite profile (*X* variables) and its quality ranking (*Y* variables). The efficiency and reliability of the PLS regression model were verified by percent variation explained by the model (R2Y), predictivity parameter (Q2), root mean square error of estimation of the training set (RMSEE), and root mean square error of prediction of the test set (RMSEP). The efficiency and reliability of the PLS regression model were verified by a training set and validated with a test set.

RESULTS AND DISCUSSION

Metabolite Identification. Significant compounds were identified by comparing their mass spectra with references from libraries. The NIST library and the in-house library from standard chemicals were utilized for the compound identification. In case that deconvolution of coeluting peaks was required for getting pure compounds, the AMDIS software was used.

In general, compounds detected from dried tea leaves were subcomponents of polyphenol, polysaccharides, and N-containing compounds (**Table 1**). The major compound observed from the pyrogram was caffeine, which was claimed to account for up to 2–3% of tea leaves (*20*). Fragments of the phenolic and flavonoid (catechin) group were also detected in significant amount in various ranges of retention time. Hydroquinone, which is one essential subcomponent of phenolic compounds, was found at retention time 40.02 min. Overall, 27 peaks were identified and left 17 peaks as unidentified compounds due to the complexity of mass fragments at that retention time (**Figure 1**).

Quality Predictive Model. All 53 GC-MS data of ranked green tea samples were divided into 47 training set samples and 6 test set samples by excluding the second sample in every tenth ranked. All variables of the *X* matrix (metabolite profile) were nontransformed and scaled to Pareto scale to diminish artifacts and noises from chromatographic data. The orthogonal signal correction (OSC) algorithm was computed in order to improve the accuracy of the model.

Two OSC components were removed, leaving 52.11% of related *X* variables to *Y* variable in the model. After that, a PLS model was constructed resulting in one extracted component reflecting 77.9% of variation in *Y* (R2Y = 0.779) and predicting 66.9% (Q2 = 0.669) of variation in *Y* according to cross-validation. The model created by this training set yielded a RMSEE value of 7.57 (**Figure 2**). By the plot between the first summary of *X* (t_1) and *Y* (u_1), a good correlation was observed with the exception of sample 46, in which the predicted value was miscalculated tremendously from the observed value (**Figure 3**). Therefore, sample 46 was removed. The PLS model was refitted again and yielded a better model with R2Y = 0.876 and Q2 improved to 0.778. Moreover, an excellent relationship between *X* and *Y* was acquired with no outlier (**Figure 4a**). Comparing to the first model, excluding the outlier lowered the RMSEE value to 5.38. Afterward, a model validation was confirmed by the test set samples. The root mean square error of prediction calculated from the test set (RMSEP) was in an excellent conformity with the error calculated on the basis of the training set (**Figure 4**). This result proved the consistency of our model to be used in green tea's quality evaluation.

Further, the variables, which influenced the model construction, were studied. The VIP values were computed from the influence on *Y* of every term (*X*'s). Fragments of metabolites found significantly different between high and low grade green tea were marked (**Figure 5**). The content of these compounds was found to be greater in low grade green tea, while some were not detected in high grade green tea. Most were phenolic-derived compounds and also some unidentified compounds (marked by an asterisk, **Figure 1**), which originated from a long-chain hydrocarbon. The significance of variables was found to be complemented with those peaks showing an obvious difference in intensity between high-ranked green tea and low-ranked green tea (**Figure 6**). This emphasized that only significant metabolites have been used to build this quality-predictive model.

In summary, the Japanese green tea quality prediction model was successfully constructed by metabolic fingerprinting utilizing spectroscopic data along with multivariate analysis. The robust calibration model was achieved with a satisfactory lower level of prediction error and better predictivity compared to the previously reported technique (*13*). In addition, since examining by GC has a limitation of compound's classes, analysis through the pyrolyzer prior to GC extends the range of detectable components that could not be analyzed by the preceding technique. However, each method does have its own good points. The previous method (*13*) provided precise and valuable information of significant metabolites conducing to wide range of tea grade classification, though there were some complications in sample preparation and expensive derivatizing agents that could lead to undesired side reactions. On the contrary, advantages and strengths of the present method were its simple operation, time savings, lower expenses, and lower level of random error, while significant compounds were defined as fragments of the compound's subcomponents. Nevertheless, method selection depends on a purpose of a study. If the objective is only to create a quality prediction model, in order to judge ranking of a green tea sample, the current method is adequate, while utilizing the solvent extraction system prior to GC/MS will be more suitable for compound identification aiming to improve product quality. For pyrolysis work, data preprocessing is very important to manage chromatograms to be in order and get rid of undesirable noise signals that could lead to weakening of model predictive power. In addition, the objective of metabolomics is to identify and quantify the complete set of metabolites; hence it is necessary to draw a range of analytical methods. Although MS has established itself as the method of choice, complementary information from other techniques is also valuable to extend the coverage of the metabolome.

ACKNOWLEDGMENT

The authors thank Minako Hada, Thermo Electron Co., Japan, for guidance and valuable suggestions in pyrolyzer and GC analyses.

LITERATURE CITED

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

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

(3) Dixon, R. A.; Gang, D. R.; Charlton, A. J.; Fiehn, O.; Kuiper, H. A.; Reynolds, T. L.; Tjeerdema, R. S.; Jeffery, E. H.; German, J. B.; Ridley, W. P.; Seiber, J. N. Applications of metabolomics in agriculture. *J. Agric. Food Chem.* **2006**, *54*, 8984–8994.

(4) Kell, D. B. System biology, metabolic modeling and metabolomics in drug discovery and development. *Drug Discov. Today* **2006**, *11*, 1085–1092.

(5) Feihn, O.; Kopka, J.; Dormann, P.; Altmann, T.; Trethewey, R. N.; Willmitzer, L. Metabolite profiling for plant functional genomics. *Nat. Biotechnol.* **2000**, *18*, 1157–1160.
(6) Scholz, M.; Gatzek, S.; Sterling, A.; Fiehn, O.; Selbig, J. Metabolite fingerprinting: detecting biological features by independent component analysis. *Bioinformatics* **2002**, *18*, 241–248.
(7) Kaneko, S.; Kumazawa, K.; Masuda, H.; Henze, A.; Hofmann, T. Molecular and sensory studies on the umami taste of Japanese green tea. *J. Agric. Food Chem.* **2006**, *54*, 2688–2694.
(8) Horie, H.; Fukatsu, S.; Mukai, T.; Goto, T. Quality evaluation on green tea. *Sens. Actuators, B* **1993**, *13–14*, 451–454.
(9) Horie, H.; Kohata, K. Application of capillary electrophoresis to tea quality estimation. *J. Chromatogr. A* **1998**, *802*, 219–223.
(10) Goto, T.; Yoshida, Y.; Kiso, M.; Nagashima, H. Simultaneous analysis of individual catechins and caffeine in green tea. *J. Chromatogr. A* **1996**, *749*, 295–299.
(11) Shimoda, M.; Shigematsu, H.; Shiratsuchi, H.; Osajima, Y. Comparison of volatile compounds among different grades of green tea and their relations to odor attributes. *J. Agric. Food Chem.* **1995**, *43*, 1621–1625.
(12) Yu, H.; Wang, J. Discrimination of LongJing green-tea grade by electronic nose. *Sens. Actuators, B* **2007**, *122*, 134–140.
(13) Pongsuwan, W.; Fukusaki, E.; Bamba, T.; Yonetani, T.; Kobayashi, A. Prediction of Japanese green tea ranking by GC/MS-based hydrophilic metabolite fingerprinting. *J. Agric. Food Chem.* **2007**, *55*, 231–236.
(14) Halket, J. M.; Waterman, D.; Przyborowska, A. M.; Patel, R. K. P.; Fraser, P. D.; Bramley, P. M. Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. *J. Exp. Bot.* **2004**, *56*, 219–243.
(15) Gullberg, J.; Jonsson, P.; Nordstrom, A.; Sjostrom, M.; Moritz, T. Design of experiments: an efficient strategy to identify factors influencing extraction and derivatization of *Arabidopsis thaliana* samples in metabolomic studies with gas chromatography/mass spectrometry. *Anal. Chem.* **2004**, *331*, 283–295.
(16) Little, J. L. Artifacts in trimethylsilyl derivatization reactions and ways to avoid them. *J. Chromatogr. A* **1999**, *844*, 1–22.
(17) Wold, S.; Antti, H.; Lindgren, F.; Ohman, J. Orthogonal signal correction of near-infrared spectra. *Chemom. Intell. Lab. Syst.* **1998**, *44*, 175–185.
(18) 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.
(19) Neilsen, N. P. V.; Carstensen, J. M.; Smedsgaard, J. Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimized warping. *J. Chromatogr. A* **1998**, *805*, 17–35.
(20) Baumann, T. W.; Morath, P.; Schuthess, B. H. Caffeine biosysthesis starts with metabolically channeled formation of 7-methyl XMP-a new hypothesis. *Phytochemistry* **1996**, *41*, 179–175.

Received for review September 19, 2007. Revised manuscript received November 23, 2007. Accepted November 26, 2007. This study was supported, in part, by the Collaboration of Regional Entities for the Advancement of Technological Excellence from Japan Science and Technology Corporation (JST-CREATE).

JF072791V