Journal of Pharmaceutical and Biomedical Analysis 49 (2009) 567–571

Contents lists available at ScienceDirect

Journal of Pharmaceutical and Biomedical Analysis

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

Short communication

Classification and prediction of free-radical scavenging activities of *dangyuja* (*Citrus grandis* Osbeck) fruit extracts using ^{1}H NMR spectroscopy and multivariate statistical analysis

Somi K. Cho[a,1], Seung-Ok Yang[b,1], So-Hyun Kim[b], Hana Kim[a], Jeong Soon Ko[a], Key Zung Riu[a], Hyo-Yeon Lee[c], Hyung-Kyoon Choi[b,*]

[a] *Faculty of Biotechnology, College of Applied Life Sciences, Cheju National University, Jeju 690-756, Republic of Korea*
[b] *College of Pharmacy, Chung-Ang University, Seoul 156-756, Republic of Korea*
[c] *The Research Institute for Subtropical Agriculture and Biotechnology, Cheju National University, Jeju 690-756, Republic of Korea*

ARTICLE INFO

Article history:
Received 19 September 2008
Received in revised form
24 November 2008
Accepted 28 November 2008
Available online 3 December 2008

Keywords:
Classification
Prediction
Dangyuja (*Citrus grandis* Osbeck)
^{1}H NMR
PLS

ABSTRACT

Different parts of *dangyuja* (*Citrus grandis* Osbeck) fruits at different maturation stages were classified using a ^{1}H NMR-based metabolomic technique. Principal components analysis allowed the clear separation of fractions extracted with 50% methanol of different parts of *dangyuja* fruits at different maturation stages by combining principal components PC1 and PC2, which together accounted for 80.4% of the variance. A loading-plot analysis revealed that sucrose, glucose, oxaloacetic acid and citric acid were dominant in mature flesh, while naringin, tyramine, proline and alanine were dominant in immature fruit samples. Projections to latent structures using a partial least squares (PLS) model were used to predict the free-radical scavenging activities (FRSA) of *dangyuja* fruit extracts based on their ^{1}H NMR spectra. The present study suggests the usefulness of combining ^{1}H NMR spectroscopy with multivariate statistical analysis for discriminating *dangyuja* fruit samples, and predicting the FRSA of different parts of *dangyuja* fruit samples at different stages of maturation.

1. Introduction

Citrus grandis Osbeck of the family Rutaceae is widely cultivated in subtropical areas of Asia. The flesh is usually separated from the skin in segments and eaten with or without sugar, and the leaves are used to flavor food. *Dangyuja* is the local name for *C. grandis* Osbeck on Jeju Island in the southern region of the Republic of Korea, where the fruit is used as a folk remedy for hangovers and the dried leaves are brewed in water as a drink. Previous studies of *C. grandis* Osbeck focused on changes in the limonoid content during fruit growth and on the level of antioxidants in the fruits of the species [1,2]. It was reported that HepG2 cells are protected from oxidative stress by the water extract of young leaves of *dangyuja* [3] and by the water-extracted and butanol-extracted fractions of young *dangyuja* fruit [4]. Immature and mature *dangyuja* fruits have generally been used for traditional medicine and in foods, respectively. The degree of maturation is determined by the color of the fruit peel. However, no metabolomic profiles of *dangyuja* fruits have been reported to date, and there are no published chemical criteria for determining the degree of maturation from the viewpoint of metabolomic profiling. The metabolome refers to the observable chemical profile or fingerprint of the metabolites present in cells, tissues, and biofluids [5,6]. Chemometric techniques combining spectrometric methods and multivariate statistical analysis such as principal components analysis (PCA), and projections to latent structures by means of partial least squares (PLS) have been applied for the metabolomic profiling and characterization of various types of plants and foods [7–10]. PCA is an unsupervised clustering method that reduces the dimensionality of multivariate data while preserving most of the variance therein [11]. PLS is a multivariate calibration method by which two sets of data, comprising the independent variable and the dependent variable are related using regression [12]. The use of PLS makes it possible to estimate the specific activities from multivariate data sets. This paper describes a comprehensive method for the metabolic profiling of different parts of *dangyuja* at different maturation stages using ^{1}H NMR spectroscopy followed by PCA, and represents the use of a PLS model to predict the free-radical scavenging activities (FRSA) of *dangyuja* extracts based on ^{1}H NMR data and measured FRSA of *dangyuja* extracts.

* Corresponding author. Tel.: +82 2 820 5605 fax: +82 2 816 7338.
E-mail address: hykychoi@cau.ac.kr (H.-K. Choi).
[1] These authors equally contributed to this work.

0731-7085/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpba.2008.11.037

2. Experimental

2.1. Dangyuja (C. grandis Osbeck) fruits

Dangyuja (*C. grandis* Osbeck) fruits were obtained from the National Institute of Subtropical Agriculture in Jeju Province, Republic of Korea. Botanical samples were taxonomically identified previously [13], and a voucher specimen (number SKC.070531) was deposited in the laboratory of Dr. S.K. Cho at the College of Applied Life Sciences, Cheju National University. We collected 10 immature (about 1 in. in diameter) and mature fruits (about 5 in. diameter) from each five tree at June, 2006 (1 month later after flowering) and November, 2006 (6 months later after flowering), respectively. The 10 fruits from a tree were pooled to reduce the variability between fruit samples. Therefore, the pooled samples from 4 trees (totally 40 fruits) in this study were used as four replicated samples for PCA and PLS as a training set. In addition, the pooled sample of 10 fruits from extra 1 tree was used as a test set for external validation. The peel and flesh could be separated in the mature fruits but not in the immature fruits, and hence the complete immature fruits were used for experiments.

2.2. Solvents and chemicals

First-grade chloroform, methanol, ethanol, D_2O (99.9%, containing 0.01% 3-(trimethylsilyl)-propionic-2,2,3,3-*d*4 acid, sodium salt (TSP), dimethyl sulfoxide, and 1,1-diphenyl-2-picrylhydrazyl (DPPH) were purchased from Sigma (St. Louis, MO, USA). NaOD were obtained from Cambridge Isotope Laboratories (Miami, FL, USA).

2.3. Extraction of dangyuja fruit samples

For NMR analysis, 100 mg of freeze-dried *dangyuja* samples was transferred into a centrifuge tube. Five milliliters of a 50% water–methanol mixture and 5 ml of chloroform were added to the *dangyuja* samples in the tube, vortexed for 30 s, and sonicated for 1 min. The materials were then centrifuged at 500 × *g* force for 20 min. The extraction was performed twice. The aqueous fraction was transferred to a 50-ml round-bottomed flask and then dried in a rotary vacuum evaporator. KH_2PO_4 was added to D_2O to make 0.1 M concentration as a buffering agent and the pH of the D_2O was adjusted to 6.0 by careful addition of 1 N NaOD solution. The pH adjusted D_2O of 1 ml was added to the 50-ml round-bottomed flask and subjected to sonication for 10 min to dissolve the dried aqueous extract. The dissolved solutions were transferred to NMR tube (Norell Inc., NJ, USA) for NMR measurements.

FRSA were measured by pulverizing freeze-dried *dangyuja* fruits in a milling machine followed by extraction with 50% methanol by stirring for 24 h at room temperature. The extract was then filtered, concentrated with a rotary vacuum evaporator under reduced pressure, and lyophilized. The extracted powders were dissolved in dimethyl sulfoxide and diluted with ethanol to the desired final concentration.

2.4. NMR measurements

All spectra were obtained by a NMR spectrometer (Avance 600 FT-NMR, Bruker, Germany) operating at a proton NMR frequency of 600.13 MHz. For each sample, 128 scans were recorded with the following parameters: 0.155 Hz/point, pulse width of 4.0 μs (30°), and relaxation delay of 1.0 s. The spectra were referenced to TSP at 0.00 ppm for aqueous fractions. TSP (0.01%, w/v) was used as an internal standard for D_2O. The peak intensities in 0.04 ppm bins in the ^{1}H NMR spectra for δ = 0.52–10.00 were used as variables.

2.5. Data analysis

The spectral region δ = 0.52–10.00 was segmented into regions of 0.04 ppm width giving a total of 237 integrated regions per NMR spectrum. Spectral intensities were scaled to total intensity for aqueous extracts. The region from 4.60 to 4.90 was excluded from the analysis because of the residual signal of water in aqueous extracts. All spectral data were mean centered and scaled to unit variance, then analyzed by PCA based on the correlation matrix. PCA and PLS were performed with SIMCA-P software (version 12.0, Umetrics, Umeå, Sweden).

2.6. Free-radical scavenging activity (FRSA)

1,1-Diphenyl-2-picrylhydrazyl radical scavenging activity was measured using the method described by Nanjo et al. [14]. An ethanol solution of 60 μl sample (2 mg/ml or ethanol itself as control) was added to 60 μl of DPPH (60 μmol/l) in ethanol solution. After mixing vigorously for 10 s, the solutions were then transferred into a 100 μl teflon capillary tube and fitted into the cavity of the ESR spectrometer. The spin adduct was measured on an ESR spectrometer exactly 2 min later. Measurement conditions: central field 3475 G, modulation frequency 100 kHz, modulation amplitude 2 G, microwave power 5 mW, gain 6.3×10^5, temperature 298 K.

3. Results and discussion

3.1. Identification of chemical constituents in dangyuja samples

The representative ^{1}H NMR spectra of aqueous fraction of IMFR, MFL, MP, and MFR extracts were shown in Fig. 1. The peaks in the aromatic region (6–9 ppm) showed relatively higher levels in IMFR samples compared to the other samples. In MFL samples, there were unique peaks in aliphatic region (0–3 ppm). The patterns of NMR peaks of MP and MFR were similar to each other by visual inspection.

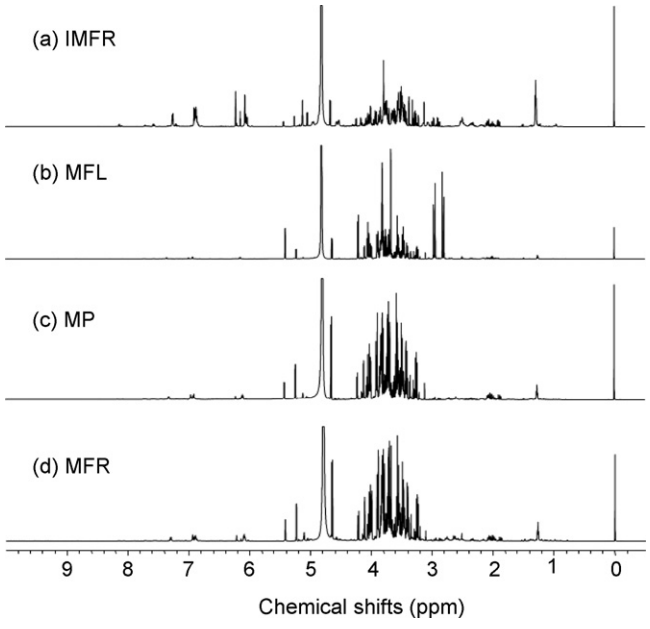

Fig. 1. ^{1}H NMR spectra of the aqueous fraction of *dangyuja* fruits extracts: (a) IMFR, complete immature fruit including flesh and peel; (b) MFL, mature fruit flesh; (c) MP, mature peel; (d) MFR; complete mature fruit including flesh and peel.

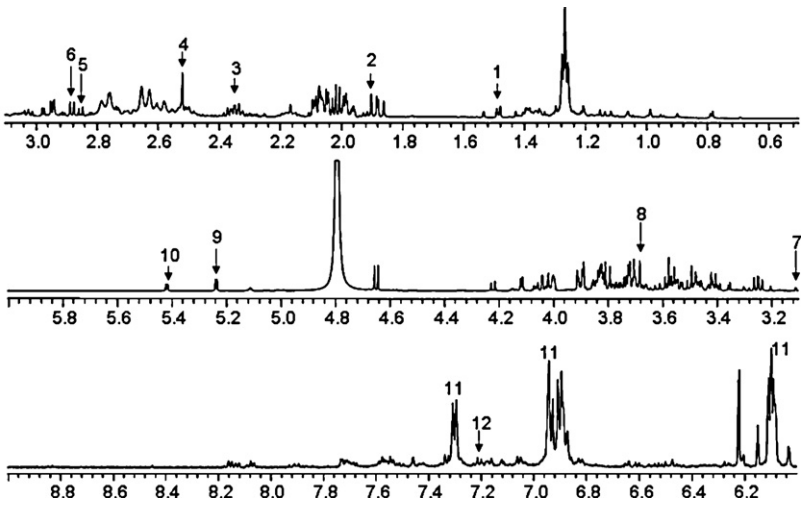

Fig. 2. Representative total ^{1}H NMR spectrum of the aqueous fraction of complete mature *dangyuja* fruits extracts: 1, alanine; 2, acetic acid; 3, proline; 4, succinic acid; 5, citric acid; 6, asparagine; 7, creatinine; 8, oxaloacetic acid; 9, glucose; 10, sucrose; 11, naringin; 12, tyramine. Values on the *X*-axis are the chemical shifts (in ppm) relative to TSP.

Fig. 2 shows a representative NMR spectrum of the aqueous fraction of complete (i.e., including the flesh and peel) mature *dangyuja* fruits. Peaks were assigned before performing the PCA. The following signals were assigned based on comparisons with the chemical shifts of standard compounds using the Chenomx NMR software suite (version 4.6, Chenomx, USA): alanine at $\delta = 1.46$ (*d*, $J = 7.2$ Hz), acetic acid at $\delta = 1.90$ (*s*), proline at $\delta = 2.34$ (*m*), succinic acid at $\delta = 2.52$ (*s*), citric acid at $\delta = 2.82$ (*d*, $J = 15.8$ Hz), asparagines at $\delta = 2.85$ (*d*, $J = 8.4$ Hz), creatinine at $\delta = 3.11$ (*s*), oxaloacetic acid at $\delta = 3.68$ (*s*), glucose at $\delta = 5.22$ (*d*, $J = 3.7$ Hz), sucrose at $\delta = 5.40$ (*d*, $J = 3.9$ Hz), and tyramine at $\delta = 7.20$ (*d*, $J = 8.5$ Hz). Naringin peaks were assigned based on previously published data [15,16] and spiking of the standard compound as $\delta = 7.28$ (*d*, $J = 8.4$ Hz).

3.2. PCA

The present study applied the correlation method of PCA rather than a covariance method since PCA produces a better separation. Fig. 3 (a) illustrates that PCA allowed the discrimination of samples according to the degree of maturation and parts of *dangyuja* samples. The two principal components PC1 and PC2 together accounted for 80.4% of the total variance, but only 7.9% of the total variance was explained by PC3. Therefore, it was assumed that the plots of each sample could be explained by just using PC1 and PC2. The different samples of *dangyuja* fruits were separated in score plots by combining PC1 and PC2. The distance between the score plots of samples of mature and immature *dangyuja* fruits indicated that the metabolomic changes had occurred during maturation period. Score plots for matured peel and matured fruit were close, which implies that their metabolomes were similar. These observations imply that the metabolomic characteristics of matured peel influenced on those of mature fruit rather than those of mature flesh. A loading-plot analysis of PC1 revealed that sucrose, glucose, oxaloacetic acid and citric acid were dominant in mature flesh, while naringin, tyramine, proline and alanine were dominant in immature fruit samples. It could be suggested that the relative levels of sugars, oxaloacetic acid and citric acid in MFR were lower than the MFL, and the levels of those metabolites in MFR were similar to MP (Fig. 3 (b)). In the loading-plot analysis of PC2, the peaks of succinic acid, asparagine and oxaloacetic acid were located in the positive position (Fig. 3 (c)). It meant that relatively higher levels of those compounds were existed in the IMFR and MFL compared to the MP and MFR.

3.3. FRSA-predictive model

PLS is a multivariate calibration method by which two sets of data, comprising the independent variable (e.g., data from NMR spectra) and the dependent variable (e.g., FRSA), are related using regression [17,18]. Values of unknown independent variables are

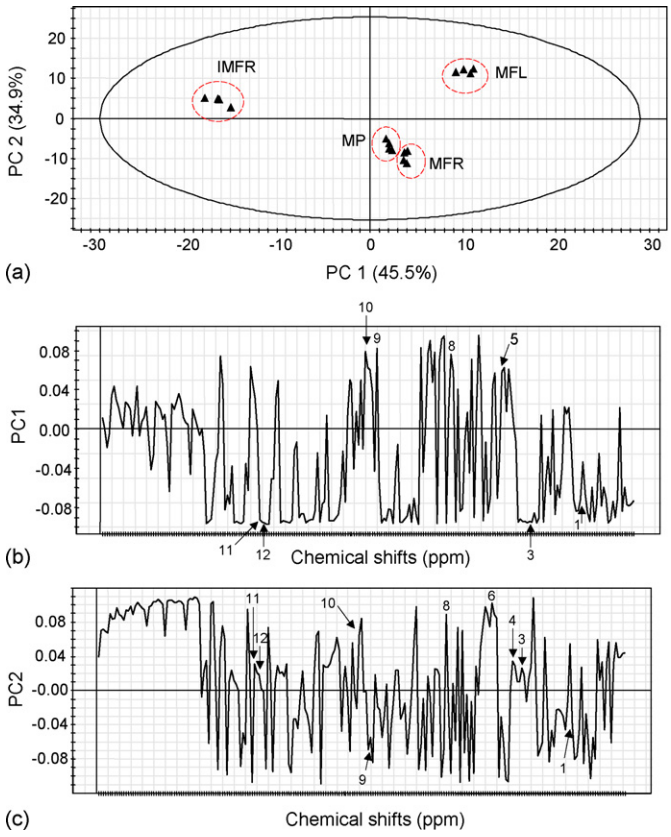

Fig. 3. PCA-derived score plot (a) of PC1 and PC2, loading plot of PC1 (b) and loading plot of PC2 (c) of the aqueous fraction of *dangyuja* fruit extracts. The solid ellipse represents the 95% confidence region for Hotelling's T^2 statistic. The dashed ellipses in (a) are labeled with the *dangyuja* fruit samples: MFL, mature fruit flesh; MFR, complete mature fruit including flesh and peel; MP, mature peel; and IMFR, complete immature fruit including flesh and peel. The numbers in the loading plots (b and c) were same as in Fig. 2 legend.

then predicted by using the resulting training model [11]. PLS regression was applied to the ^{1}H NMR spectrum data for predicting the FRSA of *dangyuja* fruit samples in this study.

The FRSA obtained by ESR method were used as a dependent variable. The FRSA of fractions extracted with 50% methanol fractions were tested by measuring scavenging of the stable free-radical DPPH. The FRSA was highest in samples of immature *dangyuja* fruit, and progressively decreased in samples of matured fruit, matured peel, and matured flesh (data not shown). These results suggest that samples of immature *dangyuja* fruits would be a good source of the antioxidants that have been implicated in inhibiting aging and cancer.

The PLS model was constructed by dividing the entire data set into two parts: (1) a training set that was used to create a model, and (2) a test set that was not used in the regression model and used to verify the model's predictive ability. The data were centered and scaled to unit variance before analysis without applying any transformation. The PLS results are presented as the number of PLS components, *R2*, *Q2*, root mean square error of the estimation (RMSEE), and root mean square error of the prediction (RMSEP, test set validation).

The goodness of fit was quantified by *R2*, while the predictive ability was indicated by *Q2* [11]. Generally, *R2* – which describes how well the data in the training set are mathematically reproduced – varies between 0 and 1, where 1 indicates a model with a perfect fit. If the *Q2* is larger than 0.5, it is considered that the model has good predictability and if the *Q2* is larger than 0.9 and less than 1, it is considered that the model has excellent predictability. As shown in Table 1, *R2* values were 0.994, 0.997, and 0.999 for PLS component 5, 6, and 7, respectively. It suggested the training set data was mathematically reproduced excellently in PLS model when more than five PLS components were included in the model. The cumulative *Q2* values were 0.970, 0.975, and 0.982 for PLS component 5, 6, and 7, respectively, which showed excellent predictive abilities.

Cross-validation was performed by dividing the data of 16 samples (4 replicated pooled samples of peel and flesh of mature fruit, total mature and immature fruits) into 7 groups, and then developing parallel models from reduced data with one of the groups deleted. The omitted data were used as a test set in the reduced model, and the differences between actual and predicted FRSA were calculated for these data points. The PLS regression model of *dangyuja* fruit extract of the training set showed that RMSEE values decreased to 0.92 when the number of PLS components included in the model increased to seven, which indicated excellent prediction ability of the PLS model (Table 1). The cumulative values of R2 and Q2 started to be plateau at the five PLS components, so five PLS components were used to make the prediction model in this study. Fig 4 showed the score plots (a) and loading plots (b) derived from the PLS model. PLS score plot, u scores, were windows in the Y space (FRSA in this study), displaying the observations as situated on the projection plane. It was observed that samples of 11–14

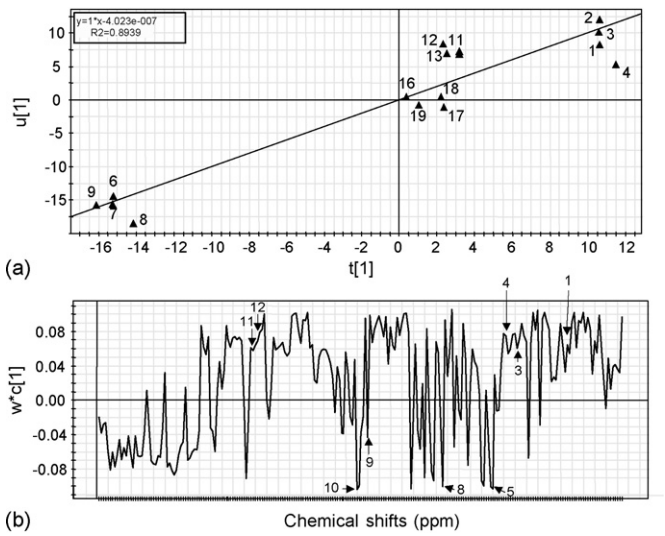

Fig. 4. PLS-derived score plot (a) and loading plot (b) of the aqueous fraction of *dangyuja* fruit extracts using the total ^{1}H NMR spectra sets. Each plot in (a), corresponds to one observation: 1–4, complete immature fruit including flesh and peel; 6–9, mature fruit flesh; 11–14, mature peel; and 16–19, complete mature fruit including flesh and peel. The numbers in the loading plots (b) were same as in Fig. 2 legend.

are somewhat different from the others, but the overall correlation was shown to be strong (Fig. 4 (a)). From the results of loading-plot analysis derived from PLS model, alanine, proline, succinic acid, naringin, and tyramine might contribute directly and indirectly for the enhanced FRSA (Fig. 4 (b)).

VIP (variable importance in the projection) value is a weighted sum of squares of the PLS weights, taking into account the amount of explained Y-variance in each dimension. It has been indicated that a factors which have more than 0.7 of VIP values could be regarded influential for separation of each sample in a PLS model [11]. As shown in Table 2, the VIP values of the major contributing compounds for the separation in the score plots derived from PLS were like follows; alanine: 1.04, proline: 0.91, citric acid: 1.45, oxaloacetic acid: 1.46, glucose: 0.69, sucrose: 1.50, tyramine: 0.94, naringin: 0.87. Among those compounds, citric acid, oxaloacetic acid and sucrose were the most influential for the separation of each sample in the PLS model in this study.

The ability of the model to predict the FRSA of *dangyuja* fruits was tested by subsequently using the external test set in the resulting PLS regression. The PLS relationships between measured and predicted values of FRSA of *dangyuja* samples in training and test sets were presented in Fig. 5 (a) and (b). Internal validation result of RMSEE = 2.29 and external prediction result of RMSEP = 2.41 were achieved using five PLS components, which suggested the good agreement between measured and predicted FRSA in the *dangyuja* samples of training and test set. These result proved the consistency of the model in this study to be used in evaluation of FRSA of *dangyuja* samples.

Table 1
PLS-derived summary of fits, RMSEE, and RMSEP of the PLS model using ^{1}H NMR data and free-radical scavenging activities of the aqueous fraction of *dangyuja* fruits extracts.

Number of PLS components	*R2* values (cumulative)	*Q2* values (cumulative)	RMSEE	RMSEP
1	0.897	0.874	7.93	6.53
2	0.938	0.909	6.41	4.08
3	0.971	0.943	4.34	6.63
4	0.988	0.951	2.94	6.72
5	0.994	0.970	2.29	2.41
6	0.997	0.975	1.64	2.18
7	0.999	0.982	0.92	4.09

Table 2
The VIP (variable importance in the projection) values of the major contributing compounds for the separation in the score plots derived from PLS model.

Chemical shift (ppm)	Compounds	VIP values
1.46	Alanine	1.04
2.34	Proline	0.91
2.82	Citric acid	1.45
3.68	Oxaloacetic acid	1.46
5.22	Glucose	0.69
5.40	Sucrose	1.50
7.20	Tyramine	0.94
7.28	Naringin	0.87

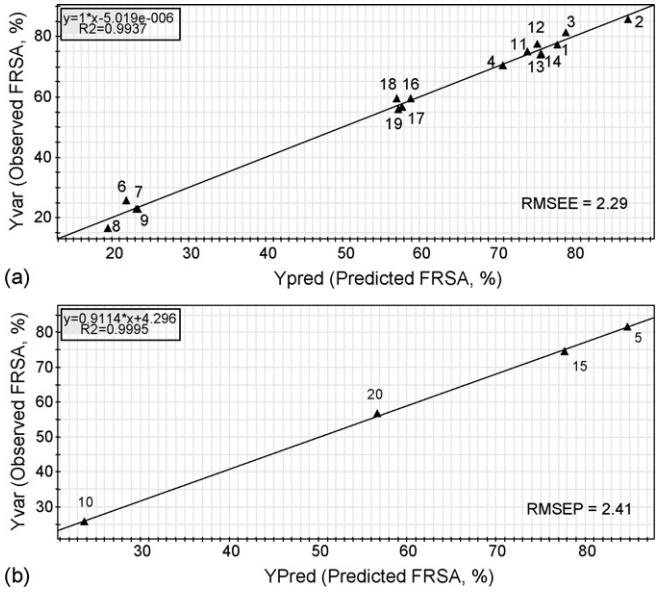

Fig. 5. PLS-derived relationship between observed and estimated FRSA (%) of the aqueous fraction of *dangyuja* fruit extracts using the total ^{1}H NMR spectra of the training (a) and test (b) sets. Each plot corresponds to one observation: 1–5, complete immature fruit including flesh and peel; 6–10, mature fruit flesh; 11–15, mature peel; and 16–20, complete mature fruit including flesh and peel.

4. Conclusions

This study used a metabolomic approach based on ^{1}H NMR spectroscopy to reveal how the characteristic metabolic profile of *dangyuja* fruits varies with the part of the fruit and the maturation stage. Combining NMR spectroscopy and PCA is a relatively simple and efficient technique for the metabolomic profiling of *dangyuja* fruits, and revealed that the FRSA could be predicted using a PLS model of ^{1}H NMR data and free-radical scavenging data sets. ESR method has been known to be sensitive method for measuring free-radical scavenging activities of various samples. However, the cost of the ESR spectrometer is expensive, and additional pre-preparation steps are necessary to analyze free-radical scavenging activities using ESR method. It was suggested that free-radical scavenging activities of *dangyuja* fruits could be predicted by using their ^{1}H NMR spectra data and multivariate statistical analysis. The information obtained in this study can be applied for the prediction of various bioactivities, such as anti-oxidative, anti-inflammatory, or anti-cancer activities of medicinal plants using their ^{1}H NMR spectra data and multivariate statistical analysis.

Acknowledgement

This work was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD) (KRF-2007-412-J05502).

References

[1] C.D. Sun, K.S. Chen, Y. Chen, Q.J. Chen, Food Chem. 93 (2005) 599–605.
[2] M.S. Mokbel, F. Hashinaga, Food Chem. 94 (2006) 529–534.
[3] H.K. Lim, E.S. Yoo, J.Y. Moon, Y.J. Jeon, S.K. Cho, Food Sci. Biotechnol. 15 (2006) 312–316.
[4] H.J. Lee, G.J. Kang, W.J. Yoon, H.K. Kang, Y.S. Kim, S.M. Kim, E.S. Yoo, Kor. J. Pharmacogn. 37 (2006) 74–80.
[5] H. Tweeddale, L. Notley-McRobb, T. Ferenci, Bacteriology 180 (1998) 5109–5116.
[6] K.H. Ott, N. Aranibar, B. Singh, Q.W. Stockton, Phytochemistry 62 (2003) 971–985.
[7] L. Tarachiwin, K. Ute, A. Kobayashi, E. Fukusaki, J. Agric. Food Chem. 55 (2007) 9330–9336.
[8] R. Consonni, L.R. Cagliani, F. Benevelli, M. Spraul, E. Humpfer, M. Stocchero, Anal. Chim. Acta 611 (2008) 31–40.
[9] J. Moros, I. Llorca, M.L. Cervera, A. Pastor, S. Garrigues, M. de la Guardia, Anal. Chim. Acta 613 (2008) 196–206.
[10] W. Pongsuwan, T. Bamba, T. Yonetani, A. Kobayashi, E. Fukusaki, J. Agric. Food Chem. 56 (2008) 744–750.
[11] L. Eriksson, E. Johansson, N. Kettaneh-Wold, J. Trygg, C. Wikstrom, S. Wold, Multi- and Megavariate Data Analysis, Umetrics Academy, Umeå, 2006.
[12] S. Wold, H. Martens, H. Wold, Lect. Notes Math. 973 (1983) 286–293.
[13] H.Y. Kim, Distribution, taxonomy, horticultural characters of the local *Citrus* spp. in Cheju and the genetic markers among them, Ph.D thesis, Chonnam National University, 1988.
[14] F. Nanjo, K. Goto, R. Seto, M. Suzuki, M. Sakai, Y. Hara, Free Radic. Biol. Med. 21 (1996) 895–902.
[15] J.A. del Rio, A. Ortuno, F.R. Marin, D. Garcia Puig, F. Sabater, Plant Cell Rep. 11 (1992) 592–596.
[16] M. Cuny, E. Vigneau, G. Le Gall, I. Colquhoun, M. Lees, D.N. Rutledge, Anal. Bioanal. Chem. 390 (2008) 419–427.
[17] S. Wold, P. Geladi, K. Esbensen, J. Ohman, J. Chemometr. 1 (1987) 41–56.
[18] S. Wold, M. Sjostrom, L. Eriksson, J. Chemometr. 58 (2001) 109–130.