Plant Prod. Sci. 11(4) : 481—486 (2008)

【Short Report】

Use of Near-infrared Reflectance Spectroscopy for the Estimation of the Isoflavone Contents of Soybean Seeds

Tetsuo Sato, Kentaro Eguchi, Tetsuya Hatano and Yoichi Nishiba

(National Agricultural Research Center for Kyushu Okinawa Region (KONARC), National Agriculture and Food Research Organization (NARO), 2421 Suya, Koshi, Kumamoto 861-1192, Japan)

Key words : Analysis, *Glycine max* L., Isoflavone, Near-infrared, Nondestructive, Soybean, Spectroscopy

The consumers' needs for agricultural products are now widely diversified with attention focused not only on the major constituents, but also on the physiological functions. In Japan, soybean (*Glycine max* L.) is used to produce various foods. Increase of nutritional functions is important to promote the consumption of soybean. However, the conventional method for estimation of elements of physiological functions such as isoflavone is labor intensive. Isoflavones may help prevent osteoporosis (Yamori, 2001). A simple and rapid method for estimating their contents is necessary for screening soybean varieties for breeding.

Near infrared spectroscopy (NIRS) has been used as one of the most powerful analytical tools in the agricultural and food fields (Williams, 2006). NIRS is an official method for analyzing the major constituents of soybean in the trade based on the quality in USA. Then, if other criteria can be estimated by the NIRS method, this method will become more useful. In this study, the feasibility of NIRS for the estimation of the isoflavone contents of soybean seeds was examined.

Materials and Methods

1. Samples and chemical measurements

The soybean samples were the same as those used in a previous study (Nishiba et al., 2007). Forty-eight samples were collected from various areas of Japan for this analysis in 2003. The fundamental statistics of isoflavone and its components of the samples were described in Table 1. The cultivated areas were from north (Hokkaido) to south (Kumamoto) in Japan. These samples were sent to our research center and were milled by a ultra-centrifugal mill ZM1000 (Retsch Co., Germany) through a screen (ϕ=1.0 mm).

The isoflavone content was determined by HPLC method (Nishiba et al, 2007). The respective components such as glucosides (daizdin, glycitin, and genistin), malonyl gulucosides, acetyl glucosides, and aglycons (daidzein, glycitein, and genistein) were also determined in this process. The unit is mg (100 g DW)$^{-1}$. By the way, the amounts of acetyl glycitin and glycitein were negligible, and then, NIRS analyses were not carried out.

2. Near infrared spectroscopic measurements

An InfraAlyzer 500 (Bran+Luebbe (B+L) GmbH, Norderstedt, Germany) was used to measure the NIR reflectance spectra in the wavelength range from 1100 to 2500 nm at 2-nm intervals. Samples were packed in a standard cup on a standard drawer for soybean powder (about 3 g), or packed in a whole grain cell on a moving drawer for intact plural soybean seeds (about 60 g). The samples were divided into two sets: a calibration set (n=36) and a validation set (n=12).

3. Statistical analysis.

Multiple linear regression (MLR) analysis of the NIRS data with the chemical data was carried out using IDAS software (B+L) on the calibration set. When the first- and second-derivative NIR spectra were calculated, the default parameters were used. The validations of the calibration equations obtained, or the validation process, were carried out using the validation set. IDAS is an accessory software to control IA500, to manage data, and to analyze data. The Unscrambler (version 9.6; Camo Co., Sweden), which was a software for the data-analysis and is sold separately, was also used on the IA500 data for partial least square regression (PLSR) or principal component regression (PCR) analysis. The authors analyzed the data not only on the original spectra, but also on the derivative spectral data, i.e., pretreated spectral data. In this case, the conditions to obtain the derivatives were as follows: gap 11 (=22 nm), segment 10 (=20 nm) for the first derivative (abbreviated as d1); and gap 10, segment 11 for the second derivative (d2). The gap and segment are the parameters in the Gap-Segment derivatives. Gap is the length of the interval that separates the two segments that are being averaged, and segment is an interval over which data values are averaged.

Received 25 February 2008. Accepted 26 April 2008. Corresponding author: T. Sato (sato@affrc.go.jp, fax: +81-96-249-1002).

Table 1 The fundamental statistics of the contents of isoflavone and its components of the samples to be analyzed. [mg (100 g DW)$^{-1}$]

		Calibration set (n=36)				Validation set (n=12)			
		Max	Min	Mean	SD	Max	Min	Mean	SD
Glycoside	Daidzin	56.71	8.41	24.55	11.68	43.59	11.57	21.34	10.63
	Glycitin	11.81	2.31	6.37	2.50	12.81	2.27	6.10	3.36
	Genistin	87.49	14.40	36.24	16.26	57.65	17.72	32.70	13.56
	Total	154.01	26.28	67.16	28.77	109.93	34.31	60.14	25.25
Malonyl glycoside	Malonyl daidzin	192.27	31.82	95.77	40.54	162.15	39.36	84.90	39.08
	Malonyl glycitin	22.22	4.77	12.60	4.49	25.98	4.70	11.81	6.02
	Malonyl genistin	264.06	51.79	133.72	50.15	200.18	65.12	125.77	42.95
	Total	473.30	102.96	242.08	89.84	368.42	119.91	222.48	82.47
Acetyl glycoside	Acetyl daidzin	1.64	0.04	0.69	0.35	1.53	0.06	0.71	0.47
	Acetyl glycitin	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Acetyl genistin	1.82	0.21	0.88	0.40	1.76	0.00	0.87	0.48
	Total	3.46	0.26	1.56	0.73	3.25	0.10	1.59	0.92
Aglycone	Daidzein	2.66	0.25	1.07	0.59	1.72	0.26	0.78	0.45
	Glycitein	0.23	0.00	0.01	0.04	0.00	0.00	0.00	0.00
	Genistein	2.82	0.35	1.37	0.65	1.87	0.32	1.08	0.50
	Total	5.48	0.60	2.45	1.20	3.58	0.58	1.86	0.93
Total isoflavone	Total	633.42	133.44	313.26	116.83	482.24	156.96	286.07	107.79

Max: maximum; Min: minimum; SD: standard deviation.

Results and Discussion

1. Data analysis with IDAS

Table 2 describes the calibration process (left side) and the validation results (right side) developed for powdered soybean analysis: the calibration equations, the correlation coefficient (R), the standard error of calibration (SEC), the standard error of prediction (SEP), mean-corrected SEP (MC-SEP), and bias. These calibrations provided the best validation in the validation process. In the tables, when the contribution ratio, R^2, exceeded 0.5 in the validation, the results are described in bold letters. For the total isoflavone content described in the bottom column, SEP was adequate for the estimation. The selected wavelengths were mainly due to C-H bonds (Osborne et al., 1993). The numbers of wavelengths selected were also adequate, i.e., not so many wavelengths. Furthermore, especially as for powdered soybean, the respective components, such as glucosides and malonyl glucosides, also could be estimated separately, as described in Table 2, where the contribution ratios of the respective components, glycosides and malonyl glycosides, exceeded 0.5. On the other hand, acetyl glucosides, and aglycons contents were poorly estimated because of their small range fluctuations.

Table 3 describes the calibration process and the validation results developed for intact plural seeds analysis. SEP for the total isoflavone content described in the bottom column, was also small enough for the estimation. The selected wavelengths were mainly due to C-H bonds. Some of the respective components in intact plural soybean seeds still could be estimated. However, the acetyl glucosides and aglycons contents were poorly estimated.

The number of wavelengths selected was adequate, i.e., not too many. In general, one wavelength can be selected for each 5 to 15 samples in MLR analysis (Hruschka, 2001), i.e., three to seven wavelengths can be selected in this case, because 36 samples were used for developing the calibration equations. Further, in the validation process, different samples from the calibration set were used to check the overfitting. Both calibration equations for the estimation of the total isoflavone content obeyed this rule. Further, the calibrations for some of the respective components of isoflavone in the powder were also adequate. On the other hand, for intact plural soybean seeds, the contribution ratio (R^2) was low, even when many wavelengths were selected for the variables for development of respective components.

Fig. 1 shows the validation results of total isoflavone analyses for the powder (Fig. 1a), and intact plural seeds (Fig. 1b), the correlation between the chemical method and NIR method (R), and standard error of prediction (SEP). The SEP value was one third to one half of SD described in Table 1. Judging from the SEPs in Fig. 1 in comparison with the standard deviation of

Table 2. The calibration and the validation results (powdered soybean). [mg (100 g DW)$^{-1}$] (The bold letters mean that the contribution ratio, R^2, exceeded 0.5 in the validation.)

		Calibration			Validation			
		Calibration equations	R	SEC	R	SEP	MC-SEP	Bias
Glycoside	**Daidzin**	**6.333+3711.588*L(1680)−2157.619*L(2236)**	**0.74**	**8.17**	**0.83**	**6.00**	**6.26**	**−0.24**
	Glycitin	**64.550+12043.211*d1L(1747)+4971.409*d1L(1943)+4965.430* d1L(2147)+3507.567*d1L(2383)−7121.926*d1L(2391)+7089.626*d1L(2471)**	**0.90**	**1.22**	**0.80**	**2.22**	**2.10**	**0.94**
	Genistin	**505.186−3698.519*L(1188)+10295.895* L(1692)+14410.360*L(2184)−20650.920*L(2236)**	**0.89**	**8.10**	**0.93**	**6.63**	**6.92**	**−0.09**
	Total	**790.541−5087.836*L(1188)+17185.570* L(1692)+24393.557*L(2184)−35110.328* L(2236)**	**0.88**	**14.68**	**0.90**	**12.99**	**13.56**	**0.10**
Malonyl glycoside	**Malonyl daidzin**	**603.122 −209987.141*d1L(1115)+45497.367*d1L(1131)+105438.195* d1L(2195)+33559.066*d1L(2275)+19353.848*d1L(2311)+68175.016*d1L(2347)−35739.527*d1L(2383)**	**0.96**	**12.30**	**0.95**	**15.43**	**14.98**	**5.69**
	Malonyl glycitin	**75.083+7364.281*d1L(1851)+8141.267*d1L(1939)−5801.995*d1L(2307)−15994.678*d1L(2419)+6439.025*d1L(2463)+13697.766*d1L(2475)**	**0.91**	**2.03**	**0.74**	**4.30**	**4.33**	**1.12**
	Malonyl genistin	**285.521+32238.084*L(1700)−20059.330* L(1724)−7438.677* L(2220)**	**0.85**	**27.84**	**0.89**	**21.03**	**20.61**	**−7.26**
	Total	**299.066+32153.232*L(1680)−18727.914*L(2244)**	**0.79**	**57.01**	**0.84**	**159.45**	**51.81**	**151.53**
Acetyl glycoside	**Acetyl daidzin**	**8.823−1013.054*d1L(1115)−1259.124*d1L(1651)+637.541*d1L(1743)+1699.458*d1L(2171)+1042.941*d1L(2235)**	**0.83**	**0.22**	**0.79**	**0.33**	**0.32**	**−0.11**
	Acetyl glycitin							
	Acetyl genistin	**3.151+ 249.333*L(1912)−241.899*L(1936)−774.832*L(2396)+783.164*L(2400)**	**0.75**	**0.29**	**0.75**	**0.34**	**0.35**	**−0.04**
	Total	4.372 −11028.082*d1L(1271)+9765.335*d1L(1751)−13480.412*d1L(1755)−3364.715*d1L(2107)+1197.979*d1L(2191)+2073.840*d1L(2347)−1159.196*d1L(2363)	0.93	0.30	0.66	0.69	0.71	0.05
Aglycone	Daidzein	−0.393+325.013*d2L(1354)+321.968*d2L(1746)	0.59	0.50	0.57	0.53	0.39	0.38
	Glycitein							
	Genistein	1.397+2028.833*L(2188)−2034.049*L(2192)	0.54	0.57	0.39	0.53	0.51	0.20
	Total	−3.722−639.700*L(1556)+758.846*L(1668)−797.737*L(2296)−1044.998*L(2364)+3996.521*L(2376)−2180.532*L(2380)	0.82	0.76	0.02	1.36	1.26	0.63
Total isoflavone	**Total**	**3187.184−32504.389*L(1188)+77647.656*L(1688)+72715.414*L(2184)−115129.492*L(2236)**	**0.92**	**48.88**	**0.95**	**38.51**	**39.94**	**−4.54**

R: Correlation coefficient between chemical method and NIR method.
SEC: Standard error of calibration.
SEP: Standard error of prediction.
MC–SEP: Mean–corrected SEP.
L(1680): raw spectral data at 1680 nm.
d1L(1747): first derivative spectral data at 1747 nm.
d2L(1354): second derivative spectral data at 1354 nm.
The amounts of Acetyl glycitin and Glycitein were almost none, and then, their NIRS analyses were not carried out.

$SEP = \sqrt{(\Sigma (p-a)^2 / (n-1))}$
$MC\text{–}SEP = \sqrt{(\Sigma (p-a-bias)^2 / (n-1))}$
$bias = \Sigma (p-a) / n$
p = predicted value by NIR
a = actual value by HPLC

Table 3. The calibration and the validation results (intact plural soybean seeds). [mg (100 g DW)$^{-1}$] (The bold letters mean that the contribution ratio, R^2, exceeded 0.5 in the validation.)

		Calibration			Validation			
		Calibrationequations	R	SEC	R	SEP	MC-SEP	Bias
Glycoside	**Daidzin**	**15.836+8495.836*d2L(1630)+19661.049*d2L(1746)−7798.740*d2L(2294)**	**0.79**	**7.62**	**0.80**	**6.50**	**6.73**	**0.88**
	Glycitin	2.216−2035.867*d1L(1203)−871.379*d1L(1947)−2424.079*d1L(2335)	0.81	1.55	0.49	2.92	3.05	−0.08
	Genistin	−221.719−2591.441*L(2236)+5188.63*L(2280)−9483.96*L(2336)+6995.796*L(2360)	0.80	10.50	0.50	12.84	13.41	0.18
	Total	−317.899+49439.781*L(1712)−49635.672*L(1732)−3795.33*L(2236)+10268.384*L(2296)−18372.227*L(2320)+12424.359*L(2368)	0.89	14.39	0.59	22.37	22.42	−6.30
Malonyl glycoside	**Malonyl daidzin**	**271.898+14656.766*d2L(1186)+16708.812*d2L(1230)+32634.916*d2L(1322)+3833.288*d2L(1370)+59551.992*d2L(1742)−19410.926*d2L(2290)−20096.738*d2L(2378)**	**0.95**	**14.03**	**0.82**	**25.68**	**26.68**	**−2.65**
	Malonyl glycitin	−40.402−3555.739*d1L(1199)+1830.240*d1L(1443)+5130.403*d1L(1739)−2498.594*d1L(1959)−4061.035*d1L(2335)−4874.354*d1L(2415)	0.89	2.25	0.58	5.14	5.30	0.84
	Malonyl genistin	**−88.362−3214.610*d1L(1395)+112284.875*d1L(1723)−22645.461*d1L(2099)−30608.988*d1L(2255)−16372.336*d1L(2291)+38457.582*dlL(2327)−74740.891*dlL(2351)**	**0.94**	**19.54**	**0.87**	**21.75**	**22.62**	**−6.30**
	Total	−664.818+52651.801*L(2192)−93205.727*L(2208)+75700.070*L(2232)−49904.352*L(2236)+32285.248*L(2272)−54251.477*L(2336)+36967.672*L(2356)	0.92	38.97	0.44	151.29	73.54	133.91
Acetyl glycoside	Acetyl daidzin	5.798−1114.726*L(1112)+1542.183*L(1124)+633.824*L(1132)−1533.524*L(1136)+577.389*L(1144)−110.439*L(1304)	0.79	0.24	0.66	0.35	0.36	−0.07
	Acetyl glycitin							
	Acetyl genistin	2.639−407.296*L(1648)+402.968*L(1660)−130.999*L(2268)+425.173*L(2280)−294.664*L(2288)	0.83	0.24	0.57	0.39	0.41	−0.03
	Total	−3.182+68.181*d1L(1395)−534.575*d1L(1655)−222.068*d1L(2255)	0.68	0.56	0.69	0.70	0.73	−0.08
Aglycone	**Daidzein**	**1.749−1492.639*L(1256)+3776.976*L(1260)−2284.478*L(1268)**	**0.66**	**0.47**	**0.85**	**0.39**	**0.24**	**0.31**
	Glycitein							
	Genistein	0.551+788.167*d1L(1263)−800.418*d1L(1643)−221.268*d1L(1927)+476.425*d1L(2287)	0.66	0.52	0.40	0.49	0.49	0.16
	Total	0.749−1805.018*L(2140)+2272.727*L(2148)−979.042*L(2184)+859.504*L(2212)−900.250*L(2236)+886.456*L(2256)−337.247*L(2264)	0.76	0.88	0.56	1.00	0.84	0.60
Total isoflavone	**Total**	**281.104+67746.070*d2L(1630)+169499.172*d2L(1746)−83235.156*d2L(2294)**	**0.85**	**65.89**	**0.82**	**63.43**	**65.63**	**8.67**

see footnotes in Table 2.

the samples (about 110 mg (100 g DW)$^{-1}$ as described in Table 1), the contents of the constituents could be estimated.

2. Data analysis with the Unscrambler

Table 4 describes the results of PLSR/PCR analysis obtained using the Unscrambler, the calibration process (left side) and the validation results (right side) developed for powdered soybean. The treatment on the original spectra, the number of factors, the correlation coefficient (R), the standard error of calibration (SEC), root mean squared error of

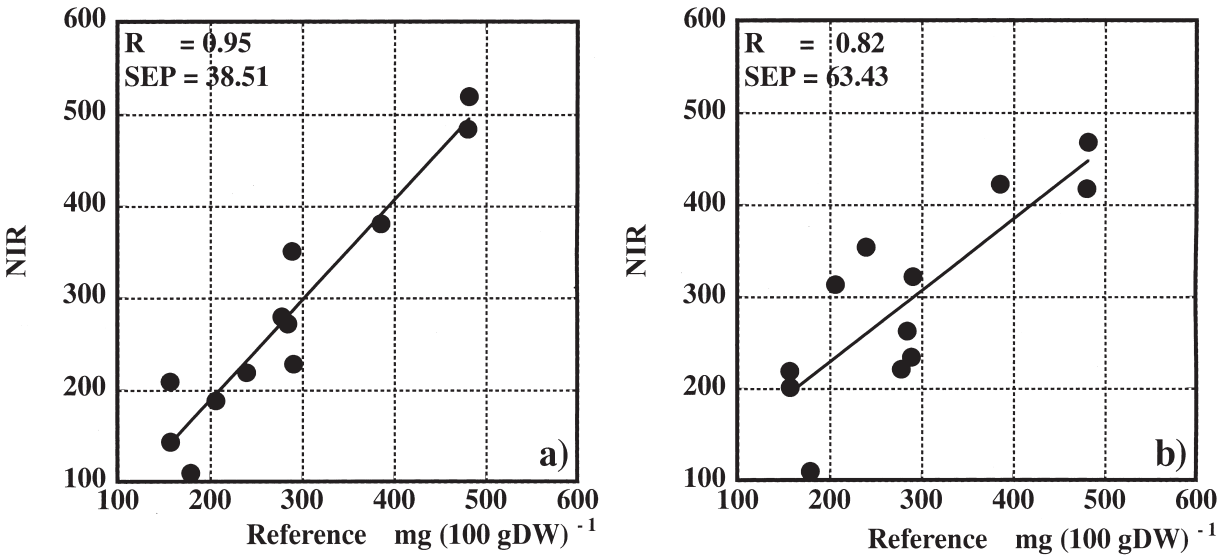

Fig. 1. The prediction results of the total isoflavone content analysis by NIR method: a) in powdered soybean; b) in intact plural soybean seeds

Table 4. The calibration and the validation results (powdered soybean) with the Unscrambler. [mg (100 g DW)$^{-1}$] (The bold letters mean that the contribution ratio, R^2, exceeded 0.5 in the validation.)

		Calibration				Validation			
		Treatment	Factors	R	SEC	R	RMSEP	SEP	Bias
Glycoside	**Daidzin**	**d1**	**pls-6**	**0.78**	**7.45**	**0.80**	**6.88**	**6.79**	**2.27**
	Glycitin	**d1**	**pls-9**	**0.89**	**1.16**	**0.77**	**2.40**	**2.32**	**0.92**
	Genistin	**raw**	**pls-8**	**0.85**	**8.64**	**0.90**	**6.63**	**6.49**	**2.31**
	Total	**raw**	**pls-8**	**0.85**	**15.22**	**0.90**	**12.21**	**11.99**	**4.14**
Malonyl glycoside	**Malonyl daidzin**	**d1**	**pls-7**	**0.91**	**17.25**	**0.91**	**17.68**	**17.95**	**4.16**
	Malonyl glycitin	**d2**	**pls-12**	**0.93**	**1.68**	**0.74**	**4.25**	**4.37**	**0.76**
	Malonyl genistin	**d2**	**pls-7**	**0.90**	**22.31**	**0.90**	**21.78**	**22.71**	**–1.25**
	Total	**d1**	**pls-7**	**0.92**	**36.59**	**0.93**	**34.31**	**35.68**	**3.13**
Acetyl glycoside	Acetyl daidzin	raw	pls-6	0.71	0.25	0.51	0.40	0.41	–0.08
	Acetyl glycitin								
	Acetyl genistin	d1	pls-1	0.42	0.37	0.69	0.39	0.40	0.02
	Total	d1	pls-1	0.42	0.67	0.61	0.77	0.80	0.01
Aglycone	Daidzein	d2	pls-3	0.60	0.48	0.37	0.54	0.47	0.30
	Glycitein								
	Genistein	d1	pls-8	0.74	0.44	0.45	0.61	0.55	0.31
	Total	d1	pls-9	0.77	0.77	0.50	1.23	1.02	0.74
Total isoflavone	**Total**	**raw**	**pls-8**	**0.91**	**48.83**	**0.95**	**40.01**	**38.88**	**14.66**

see footnotes in Table 2.

prediction (RMSEP), the standard error of prediction (SEP), and bias were as described. The better cases were for PLSR and PCR. For the total isoflavone content, SEP was adequate for the estimation. Further, especially as for powdered soybean, the respective component, glucoside, malonyl glucoside, also could be estimated separately, as described in Table 4, where the contribution ratios of the respective components glycosides and malonyl glycosides exceeded 0.5. On the other hand, acetyl glucoside and aglycon contents were poorly estimated as in MLR analysis.

Table 5 describes the calibration process and the validation results developed for intact plural seeds. As for the total isoflavone content, SEP was fair enough for the estimation. As for intact plural soybean seeds, the results were improved: some of the respective

Table 5. The calibration and the validation results (intact plural soybean seeds) with the Unscrambler. [mg (100 g DW)$^{-1}$] (The bold letters mean that the contribution ratio, R^2, exceeded 0.5 in the validation.)

		Calibration				Validation			
		Treatment	Factors	R	SEC	R	RMSEP	SEP	Bias
Glycoside	**Daidzin**	**d1**	**pls-11**	**0.96**	**3.34**	**0.81**	**6.98**	**7.28**	**0.33**
	Glycitin	d2	pls-20	0.99	0.26	0.65	2.56	2.67	–0.13
	Genistin	**d2**	**pls-12**	**0.97**	**4.10**	**0.84**	**8.33**	**7.84**	**–3.63**
	Total	**d2**	**pls-12**	**0.97**	**6.96**	**0.84**	**16.19**	**15.43**	**–6.62**
Malonyl glycoside	**Malonyl daidzin**	**d2**	**pls-13**	**0.98**	**7.44**	**0.94**	**14.83**	**14.87**	**–4.15**
	Malonyl glycitin	**raw**	**pls-12**	**0.99**	**0.75**	**0.73**	**4.16**	**4.33**	**–0.40**
	Malonyl genistin	**d2**	**pls-11**	**0.96**	**13.43**	**0.93**	**15.91**	**16.36**	**–2.81**
	Total	**d2**	**pls-12**	**0.98**	**19.86**	**0.95**	**26.64**	**26.52**	**–8.07**
Acetyl glycoside	Acetyl daidzin	raw	pls-6	0.69	0.26	0.63	0.36	0.38	–0.05
	Acetyl glycitin								
	Acetyl genistin	raw	pcr-6	0.68	0.30	0.69	0.36	0.37	–0.02
	Total	raw	pcr-6	0.67	0.55	0.69	0.68	0.71	–0.07
Aglycone	Daidzein	d2	pls-7	0.81	0.35	0.36	0.55	0.49	0.29
	Glycitein								
	Genistein	raw	pls-8	0.83	0.36	0.40	0.63	0.52	0.39
	Total	d2	pls-6	0.75	0.80	0.36	1.14	0.98	0.66
Total isoflavone	**Total**	**d2**	**pls-12**	**0.97**	**26.57**	**0.96**	**32.80**	**30.28**	**–15.34**

see footnotes in Table 2.

component also could be estimated. However, acetyl glucoside and aglycon contents were poorly estimated.

The total isoflavone content could be estimated not only with powdered soybean but also with intact plural soybean seeds. However, the content of the respective isoflavone component could be estimated only using powdered soybean. The present findings suggest that the total isoflavone content of the soybean seeds could be estimated for simple, rapid, and nondestructive breeding selection by the NIRS method. The respective elements in the powder could be estimated. PLSR and PCR analyses were also tried, and the results were similar to those obtained by MLR. Further, for total malonyl glycoside, the bias was drastically improved by PLSR analysis. NIRS analysis of major constituents and the deterioration indices in soybean (Sato et al. 1994), and fatty acid composition in soybean (Sato et al. 2002) have been reported. The present findings showed that the isoflavone content could be estimated by the NIRS method, and increase the value of this method in soybean analysis.

References

Hruschka W.R. 2001. In P. Williams, and K. Norris eds. Near Infrared Technology in the Agricultural and Food Industries (Second Edition), American Association of Cereal Chemists, Inc. Minnesota. 49.

Nishiba, Y. et al. 2007. Nippon Shokuhin Kagaku Kogaku Kaishi. 54 : 295-303. (in Japanese)

Osborne, B.G. et al. 1993. In B.G. Osborne et al. eds., Practical NIR Spectroscopy with Applications in Food and Beverage Analysis. Longman Science and Technical, 2nd. London. 13-25.

Sato, T. et al. 1994. J. Am. Oil Chem. Soc. 71 : 1049-1055.

Sato, T. et al. 2002. J. Am. Oil Chem. Soc. 79 : 535-537.

Williams, P.C. 2006. In Ozaki et al. eds. Near-infrared Spectroscopy in Food Science and Technology, Wiley-Interscience, NJ. 165-217.

Yamori, Y. 2001. In Y. Yamori, S. Ohta, and S. Watanabe eds., Soybean Isoflavones (in Japanese), Saiwai Shobou Co., Tokyo. 79-87.