Oil and Protein Analysis of Whole Rapeseed Kernels by Near Infrared Reflectance Spectroscopy[1]

R. TKACHUK, Grain Research Laboratory, Canadian Grain Commission, 1404-303 Main St., Winnipeg, Manitoba, Canada R3C 3G9

ABSTRACT

Oil, protein, chlorophyll and glucosinolate content were analyzed in whole rapeseed kernels by use of a near infrared reflectance technique. Oil and protein content could be estimated with high correlation and good accuracy when predicted results for 89 samples were compared to standard laboratory results. For oil, a multiple correlation coefficient (R) of 0.954 and a standard error of estimate (Sy) value of 0.83 were obtained when reflectance was measured at eight wavelengths. For protein, R = 0.964 and Sy = 0.88 were obtained when reflectance was measured at six wavelengths. Significantly lower correlations were obtained for prediction of chlorophyll (R = 0.506) and glucosinolate (R = 0.707) content, and presently near infrared data cannot be used to measure these two constituents. For the prediction of oil and protein, the levels of accuracy obtained are sufficient for many analytical purposes, and if needed, the accuracy can be improved by repeated measurements. The method is rapid, involves no sample preparation, and leaves intact, viable seed available for other purposes.

INTRODUCTION

Rapid and nondestructive analytical methods have a great potential for analyzing cereals and oilseeds, since current methods involve seed grinding followed by slow, tedious and expensive wet chemical analytical tests. This present study describes a near infrared reflectance (NIR) procedure for analyzing whole rapeseed kernels. A reflectance spectrophotometer was used to obtain NIR data which was correlated with the oil, protein, chlorophyll and glucosinolate content of rapeseed samples. Only the NIR data which were highly correlated with the constituent contents were then analyzed by a stepwise multiple linear regression procedure to select the wavelengths giving data which would best predict the constituents. Finally, the NIR analytical procedure was tested by predicting oil, protein, chlorophyll and glucosinolate contents for a new set of rapeseed samples. The analytical method is a variation of a method described for the analysis of protein in whole wheat kernels (1) and is based on an NIR method developed by K.H. Norris for the analysis of protein and moisture in ground wheat (2).

MATERIALS AND METHODS

A total of 178 samples of rapeseed (*Brassica napus* and *Brassica campestris*) grown during 1977 and 1979 in Western Canada were used in the study. Of this lot, 138 samples graded No. 1 Canada Western (1 CW), 34 graded 2 CW and 6 graded 3 CW. The color of the rapeseed varied from light brown to black (163 samples); 15 samples contained different admixtures of yellow and brown rapeseed. Laboratory analyses of the samples showed the following ranges: moisture content, 6-7%; oil content, 35.4-47.9%; protein content, 15.3-29.6% (whole seed basis) and 27.9-48.7% (oil-free meal basis); chlorophyll content, 1-67 ppm (whole seed basis); glucosinolate, 0.2-8.9 mg/g whole seed. Approximately two-thirds of the samples contained less than 1% of erucic acid (whole seed basis). Standard laboratory analyses were used to determine oil (3), protein (4), chlorophyll (5), and glucosinolate (6) content of rapeseed samples. For reasons of consistency, all results were reported on 8.5% moisture basis. One-half of the samples (89) were used for calibration and the other half for prediction purposes.

Reflectance data were obtained with a Cary 17I spectrophotometer controlled by a PDP 11/34 minicomputer. The spectrophotometer was modified to increase the signal-to-noise ratio in the 1-3 μ near infrared region by using a gold-covered integrating sphere, Suprasil optics, and a PbS detector cooled to 0 C ± 0.1. The minicomputer was equipped with 128K words of memory, two one-megaword discs and a hard copy and CRT terminals. The RSX-11M operating system was used with a Fortran IV compiler and Macro assembler. Reflectance was recorded for ca. 15 g whole rapeseed held in a cylindrical sample holder (1 cm deep and 6.5 cm in diameter) faced with an Infrasil cover. The data were collected with the spectrophotometer in a doublebeam mode using a sulfur pellet as a reference reflectance standard (7). For each rapeseed sample, an average of five readings was recorded at each 2.0 nm interval from 1000 to 2400 nm for a total of 700 values. Individual readings were rejected if they varied by more than 0.004 (reflectance) when compared to the average of the previous four readings, a procedure which accounted for noise-spikes or any other spurious readings. The reflectance values (700 for each sample) were smoothed with a 9-point quartic convoluting function (8) using a computer program (9) which was adapted for the PDP 11 computer. These smoothed values were converted to apparent absorbance values (A'), where $A' = \log 1/\text{reflectance}$. Next, for a set of 89 rapeseed samples, simple correlation coefficients (r) were computed for a sample constituent, determined by a standard chemical technique, vs various algebraic relationships of A'. The highest correlations were obtained at wavelengths λ_X and λ_Y using the relation $(A'\lambda_X - A'\lambda_Y)/A'\lambda_Y$, henceforth termed as the independent variable.

To obtain a calibration equation relating the independent variable to a rapeseed constituent, a subset of 70 independent variables selected (from a total of 244,650) for having the highest r values was entered into a stepwise regression computation (10). A subset of 70 variables is close to the maximal capability of a PDP 11/34 minicomputer. In the regression program, values of 4.0 for F and 0.01 for Tolerance were chosen as criteria for entering and removing these variables. With this limitation, the program would usually select only several variables from the subset of 70. To obtain the highest multiple correlation coefficient (R) and the lowest standard error of estimate (Sy), subsequent sets of 70 variables were entered into the regression computation. These subsets were chosen by selecting different variables concentrated more closely around the variables which were already selected by the program in the previous computation. The highest R and the lowest Sy values were used as criteria for evaluating the best predicting calibration equation. This equation was tested by using it to predict the same constituent in another set of 89 samples. This procedure was followed by estab-

[1] Contribution no. 468 of the Grain Research Laboratory. Presented in part at the 65th Annual Meeting of the American Association of Cereal Chemists, September 1980, Chicago.

lishing the best predicting calibration equation for each of the constituents. Results obtained were then compared with those determined by the standard laboratory methods.

RESULTS AND DISCUSSION

Reflectance spectra for whole and ground rapeseed (Fig. 1) are quite similar, indicating that both spectra probably contain similar compositional information. Some of the peaks in the ground sample are better resolved, as is illustrated by the height of the 2310 and 2340 nm absorption peaks.

The optimal wavelengths, as selected by the stepwise regression program, for the prediction of oil, protein, chlorophyll and glucosinolate content of whole rapeseed, are shown in Table I. In this table, each independent variable (x) is represented by two wavelengths, λ_X and λ_Y. At each step of the regression analysis, a value for R and Sy is computed. For the calibration procedure represented by one set of 89 samples, it is seen that as the program enters additional variables into the regression equation, the values for R progressively increase and those for Sy progressively decrease. To test the validity of the calibration, a prediction procedure was also carried out for another set of 89 samples. In this prediction procedure, the seed constituent content was calculated using the calibration equation computed during the calibration procedure. During the prediction, values for R and Sy were computed also. It is seen (Table I) that during the prediction, the R and Sy values also increase and decrease, respectively, as additional variables are entered; however, the changes are less significant than those observed during the calibration procedure. This comparison between calibration and prediction is shown in Table I for each of the four constituents examined.

For oil content, the program selected four pairs of wavelengths for an optimal calibration equation. The wavelengths represent a wide range of the spectral region, i.e., from 1218 to 2200 nm. The 1734 nm and the 2200 nm points appear to have the greatest effect on R in the computation of the calibration equation and also in the prediction of oil content. In examining the spectra of rapeseed

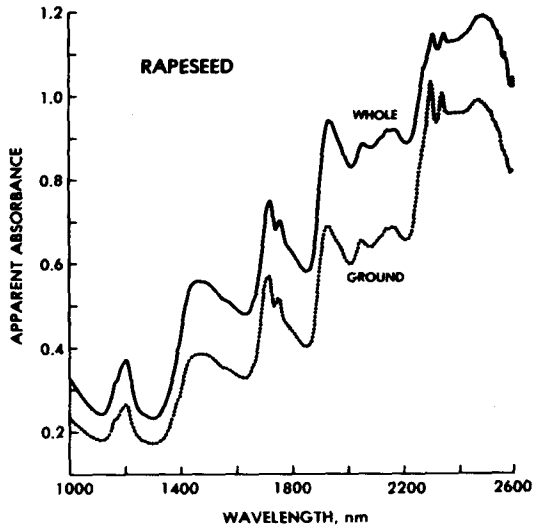

FIG. 1. NIR reflectance spectra for whole and ground rapeseed (Tower cultivar from 1979 crop) recorded vs a sulfur pellet reflectance standard.

shown in Figure 1, several prominent peaks were noted which appeared to be of interest for predicting oil content in rapeseed. These peaks were similar to the ones found for fatty acids by Holman and Edmondson (11), e.g., at 1210 nm (absorption due to C-H second overtone of the CH_2 groups), at 1720 and 1760 nm (due to C-H first overtone of the CH_2 groups) and at 2310 and 2340 nm (due to C-H combination of the CH_2 groups). However, in our calibration equation, the only two wavelengths which were selected close to these were at 1218 and 1734 nm. Some others selected in our equation, at 1410 nm wavelength (due to O-H first overtone of the OH groups [11]) and at 1390 nm (possibly due to C-H combination of the CH_2 groups [12, 13]), were found to be on the shoulders of the peaks.

Figure 2 shows a plot of NMR oil content vs reflectance oil content for 89 samples of whole rapeseed used for obtaining a calibration equation. The points for both 1977 and 1979 crop years are well distributed in the plot and

TABLE I

Optimal Wavelength Selection for Oil, Protein, Chlorophyll and Glucosinolate Prediction in Rapeseed

	Step	Variable entered	Wavelength selected (nm)		Calibration		Prediction	
			λ_X	λ_Y	R	Sy	R	Sy
Oil	1	X_1	2200	1734	0.943	0.95	0.888	1.21
	2	X_2	1530	1392	0.971	0.69	0.913	1.09
	3	X_3	1358	1218	0.974	0.65	0.929	1.00
	4	X_4	1494	1410	0.979	0.59	0.954	0.83
Protein	1	X_1	2164	2148	0.978	0.72	0.959	0.93
	2	X_2	2158	2140	0.980	0.69	0.963	0.89
	3	X_3	2202	2130	0.983	0.65	0.964	0.88
Chlorophyll	1	X_1	1630	1486	0.728	8.68	0.366	10.47
	2	X_2	2338	2294	0.756	8.34	0.378	10.55
	3	X_3	2230	2074	0.774	8.11	0.419	10.58
	4	X_4	2014	1958	0.802	7.69	0.506	9.83
Glucosinolate	1	X_1	1642	1634	0.665	1.54	0.587	1.66
	2	X_2	1634	1620	0.761	1.35	0.707	1.47

$X = (A'\lambda_X - A'\lambda_Y)/A'\lambda_Y$, where A' = apparent absorbance at λ_X and λ_Y. Multiple correlation coefficient (R) and standard error of estimate (Sy) were computed using adaptions of the PDP-11 version (10). During the stepwise regression computations, values of 4.0 for F and 0.01 for Tolerance were used as criteria for entering and removing X-variables.

contribute to a high multiple correlation (R = 0.979) and a low standard error of estimate (Sy = 0.59). In Figure 3, which represents another set of 89 samples for the prediction of oil content, the R value (0.954) is not as high and the Sy value (0.83) is not as low as for the calibration. This is to be expected since the calibration is invariably a better plot than the prediction. In both plots, the calculated regression line is almost at the 45° line and the intercept is reasonably close to zero, which shows that there is a good agreement between the NMR and reflectance methods. The standard error of estimate values (0.59 and 0.83) indicate the degree of scatter of the points about the regression line; a wider scatter is shown for prediction than for calibration. In comparing the individual years from the calibration plot in Figure 2, the calculated slope and intercept for 1977 (0.97, 1.3) are similar to the ones for 1979 (0.95, 2.0). Also, for the prediction plot in Figure 3, the slope and intercept for 1977 (0.92, 3.2) are similar to the ones for 1979 (0.88, 4.8). This similarity in slope and intercept indicates that the same set of wavelengths and the same equation can be used for prediction of oil content for both years.

The plots for the calibration and prediction of protein content of whole rapeseed are shown in Figures 4 and 5. Again, as for oil content, the R value (0.983) is higher and the Sy value (0.65) is lower for the calibration plot than for the prediction (R = 0.964; Sy = 0.88). In comparing the individual years for the calibration plot in Figure 4, the calculated slope and intercept for 1977 (0.96, 0.92) are similar to the ones for 1979 (0.97, 0.62). Also, for the

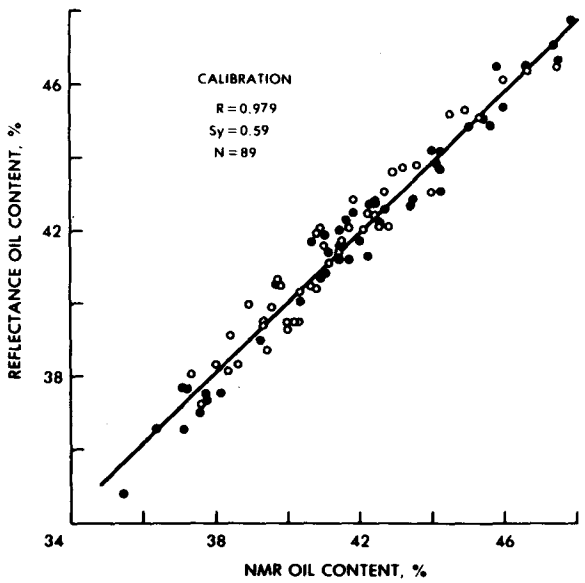

FIG. 2. Calibration plot for oil content, NIR reflectance vs NMR method (8.5% moisture basis), for 89 samples of whole rapeseed from the 1977 (• — •) and 1979 (○ — ○) crop years.

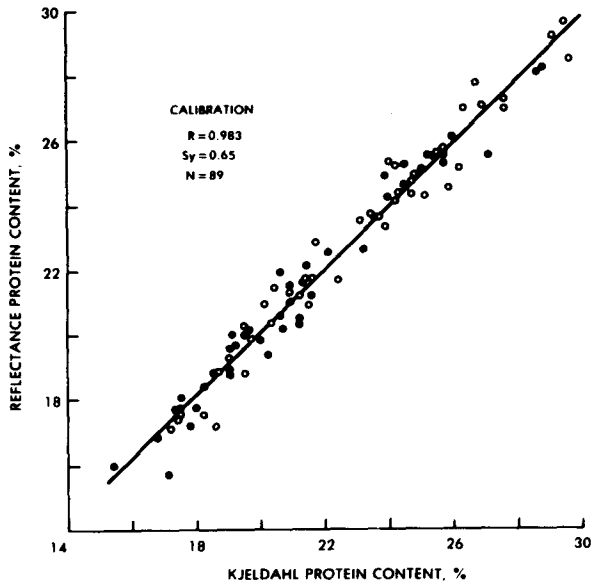

FIG. 4. Calibration plot for protein content, NIR reflectance vs Kjeldahl method (N X 6.25, calculated on 8.5% moisture basis and on whole seed basis), for 89 samples of whole rapeseed from the 1977 (• — •) and 1979 (○ — ○) crop years.

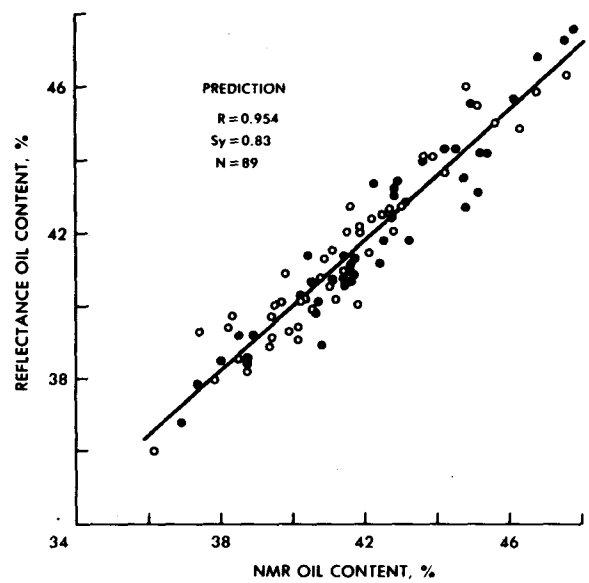

FIG. 3. Prediction plot for oil content, NIR reflectance vs NMR method (8.5% moisture basis), for another set of 89 samples of whole rapeseed from the 1977 (• — •) and 1979 (○ — ○) crop years.

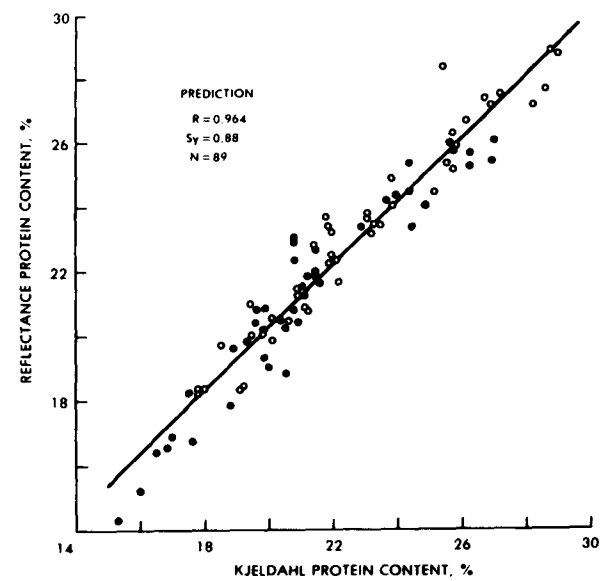

FIG. 5. Prediction plot for protein content, NIR reflectance vs Kjeldahl method (N X 6.25, calculated on 8.5% moisture basis and on whole seed basis), for another set of 89 samples of whole rapeseed from the 1977 (• — •) and 1979 (○ — ○) crop years.

prediction plot in Figure 5, the slope and intercept for 1977 (0.96, 0.88) are similar to the ones for 1979 (0.94, 1.66). This similarity in slope and intercept indicates that the same set of wavelengths and the same equation can be used for prediction of protein content for both years. The wavelengths selected by the computer program for calibration and prediction of protein content in rapeseed (Table I) are shown to be in a fairly narrow range as compared to those selected for oil content. Only three pairs were selected ranging from 2130 to 2202 nm. The wavelength, 2164 nm, is similar to 2167 nm, a wavelength used in the equation for measuring protein by one of the commercial manufacturers of NIR reflectance instruments (14). Some of the supporting wavelengths are fairly close to 2164 nm (e.g., 2158, 2148 and 2140 nm); however, these were included because they were significantly important for a more accurate prediction of protein (as shown by the increase in R and decrease in Sy in Table I).

In addition to the determination of oil and protein content on whole rapeseed, the method may eventually be used to predict the contents of other constituents such as chlorophyll and glucosinolate. The optimal wavelength selection for these two other constituents is also shown in Table I. For chlorophyll, the wavelength range of interest was shown to be between 1486 and 2338 nm. A lower degree of accuracy with our NIR study, as compared to the results obtained in the visible region by Daun (5), indicates that further investigation is necessary, with the possibility of combining the wavelengths of both regions. Also, some additional work is required for a more accurate determination of glucosinolate content in rapeseed. Wetter and Youngs (6) described an effective method for determining total glucosinolate content based on a specific UV absorbance of thioureas and oxazolidine-2-thiones. However, the method involves grinding samples followed by lengthy chemical procedures.

For each rapeseed constituent, a similarity between the standard deviation (SD) values for the 15 samples containing yellow and brown admixtures of rapeseed and the SD values for the total number of samples (178) showed that color of rapeseed was not a factor in the determination of oil, protein, chlorophyll, or glucosinolate content.

This study did not involve the measurement of moisture content of whole rapeseed. However, it should be relatively simple to carry out such an analysis by NIR, since water forms a very prominent peak at 1930 nm in whole rapeseed (Fig. 1). Accordingly, a reflectance value at 1930 nm minus a value at 2100 nm (for background purposes) should allow one to measure moisture content accurately, as suggested in the methods for whole corn and sorghum (15), or ground wheat (16).

These results indicate that it is possible to adapt several NIR instruments, now commercially available for analysis of ground wheat, to analyze whole rapeseed for oil, protein and moisture contents quite accurately. To use wavelengths for oil and protein outlined in Table I (including 1930 and 2100 nm for moisture), such instruments would require 16 specific wavelength filters or ca. six variable (tilting) filters; each variable filter can be used within a maximal range of ca. 100 nm. Only instruments with reprogramming capability can be modified, since to carry out such analyses new and additional calibrations are required.

The relationship between constituents (oil and protein content) determined by the NIR method and by the accepted analytical procedures is excellent. The differences between the methods can be attributed to errors in both methods rather than in the NIR only. One of the major problems which contributes to some of the error in the analysis of oilseeds is in the preparation of the sample. Hymowitz et al. (17) indicated that sample grinding time affected the results of protein and oil content of soybeans as determined by NIR. The method we used for determination of constituents by NIR on whole seeds not only eliminates error due to grinding but also saves time. The relative accuracy of the NIR reflectance method, in addition to the advantages gained in time by analyzing for two or more constituents simultaneously, should make the method useful to cereal researchers and plant breeders. Another advantage of this nondestructive method is that more tests can be done with the intact and viable seeds.

ACKNOWLEDGMENTS

The author thanks F.D. Kuzina and Y.F. Lewanski for technical assistance and J.K. Daun and L.D. Davidson for collecting and analyzing some of the rapeseed samples.

REFERENCES

1. Tkachuk, R., Cereal Foods World (in press).
2. Norris, K.H., in "Cereals '78: Better Nutrition for the World's Millions," AACC, St. Paul, MN, 1978, pp. 245-251.
3. Robertson, J.A., and W.H. Morrison, JAOCS 56:961 (1979).
4. American Association of Cereal Chemists, Approved Methods of the AACC, Vol. 2, The Association, St. Paul, MN, 1962 (revised to 1979).
5. Daun, J.K., JAOCS 53:767 (1976).
6. Wetter, L.R., and C.G. Youngs, Ibid. 53:162 (1976).
7. Tkachuk, R., and F.D. Kuzina, Appl. Opt. 17(17):2817 (1978).
8. Savitzky, A., and M.J.E. Golay, Anal. Chem. 36:1627 (1964).
9. Jones, R.N., T.E. Bach, H. Fuhrer, V.B. Kartha, J. Pitha, K.S. Sashadri, R. Venkataraghavan and R.P. Young, in "Computer Programs for Infrared Spectrophotometry," Vol. I, 2nd Edn., NRCC Bulletin No. 11, Ottawa, 1976, pp. 65-97.
10. PDP-11 version, (Software Development Inc., PO Box 500, Middlebury College, Middlebury, VE 05753) of the BMDP-77 Statistical Program Package, Health Science Computing Facility, Center for Health Sciences, University of California, 1977.
11. Holman, R.T., and P.R. Edmondson, Anal. Chem. 28:1533 (1956).
12. Wheeler, O.H., Chem. Rev. 59:629 (1959).
13. Goddu, R.F., and D.A. Delker, Anal. Chem. 32:140 (1960).
14. Watson, C.A., Ibid. 49:835A (1977).
15. Stermer, R.A., Y. Pomeranz and R.J. McGinty, Cereal Chem. 54:345 (1977).
16. Law, D.P., and R. Tkachuk, Ibid. 54:874 (1977).
17. Hymowitz, T., J.W. Dudley, F.I. Collins and C.M. Brown, Crop Sci. 14:713 (1974).

[Received February 2, 1981]