♣The Calibration of NIR Reflectance Spectrometer for the Determination of Diverse Compositional Parameters

F. E. Barton II[a] and **G.C. Cavanagh**[b]

[a]Richard B. Russell Research Center, P.O. Box 5677, Athens, GA 30613, and [b]Cavanagh Associates, Fresno, CA

Near Infrared Reflectance Spectroscopy (NIRS) has been used to analyze a variety of agricultural commodities. Many times the success or lack of it has been due to inadequate computational capabilities or asking more of the instrument than its capabilities. This study showed that for the development of multiple analytical methods by Near Infrared Reflectance Spectroscopy (NIRS) a sophisticated instrument computer software system can perform multiple analyses and that different data treatments were required for each constituent. In this case a scanning monochromator with a digital mini computer and the USDA/Pennsylvania State University software system could easily be calibrated to determine simultaneously the percentages of moisture, protein and residual oil in cottonseed meal. High RSQ (.98–.99 for all constituents) and low standard errors of prediction (SEP) (0.18 moisture, 0.1 oil and 0.34 protein) were obtained. Success was due to the acquisition of sufficient data for all three analyses (via the monochromator) and the use of sufficient computing power to obtain an optimized calibration where each constituent required a different mathematical data treatment.

Near infrared reflectance spectroscopy (NIRS) has been used to measure protein and moisture in wheat (1–4), various compositional parameters in forages (5–12) and oil and moisture in sunflowers (13, 14). NIRS measurements have been made on fixed filter (15), tilting filter (7) and monochromator instruments (6, 7, 12, 14). Computations and data transformations have been made with everything from calculators to large mini-computers. The capabilities of the spectrometers in the above studies ranged from acquiring data from as few as six data points to as many as 2,000. The question to be answered is, "What is needed to acquire a particular analysis or group of analyses in terms of instrumentation and computational capabilities?"

In this paper cottonseed meal was used as an example to show that the optimal mathematical data treatment required to obtain a good calibration was different for each constituent (residual oil, moisture and protein), and that some form of scanning instrument and at least a microcomputer system to handle data computations was necessary.

MATERIALS AND METHODS

The samples used in this study were provided by Ranchers Cotton Oil, Fresno, California. They consisted of a set of nine and a set of 53 samples of commercially prepared cottonseed meal. The samples were ground through a UDY cyclone mill with a 1-mm screen and packed into sample cups for scanning. Data for the oil, moisture and crude protein contents of the samples were provided by Ranchers Cotton Oil and obtained in their laboratory by standard AOAC procedures (16).

The NIRS data were collected with a Pacific Scientific Model 6350 monochromator and data stored in a Digital Equipment Corporation PDP 11/34 mini-computer. The USDA/Pennsylvania State University software system was used for spectrometer control, data transformation and statistical analysis (17) and used as discussed in ARS Handbook 643 (18). The approximation of a 19-filter instrument and production of a scatter plot were accomplished with a new program called REG 70 (Westerhaus, M.O., and J.S. Shenk, personal communication). The Fourier self-deconvolution was accomplished with the University of Georgia Scientific Applications Software Package on a Digital Equipment Corporation Micro Vax II computer (Carreira, L., personal communication).

The first set of nine samples was scanned and plots of the spectra examined to see if spectral differences were present. The set of nine was added to the set of 53 and the resulting 62 samples used for calibration and validation. Of the 62 samples, 47 were used for calibration and 15 for validation. The software reserved every fourth sample for validation.

RESULTS AND DISCUSSION

The data in Table 1 are for the first set of nine samples which were used to see if sufficient information was present in the spectra to conduct all three analyses, residual oil (oil) moisture (H_2O) and protein (CP). It was noted that sample 9 was the most "different" sample in the set. The spectral difference of sample 9 around 1900-2000 nm probably is due to its high moisture, CP, and oil content compared to the other samples (Table 1). Crossovers usually occur as a result of particle size and moisture difference which exhibit different scatter characteristics (19). In this case the spectrum of sample 9 appears to exhibit increased scatter due to being the finest ground sample, as expected from its composition. The more finely ground the higher

TABLE 1

Percentage Composition of Nine Cottonseed Meal Samples

Sample #	H_2O	CP	Oil
1	9.7	41.50	3.1
2	9.6	40.55	3.0
3	9.6	43.20	3.2
4	9.7	43.60	3.4
5	9.5	42.52	3.2
6	9.6	44.12	3.5
7	9.5	43.90	3.8
8	9.5	44.32	3.6
9	9.7	48.65	4.0
mean	9.6	43.60	3.4
STD DEV	0.08	2.3	0.3

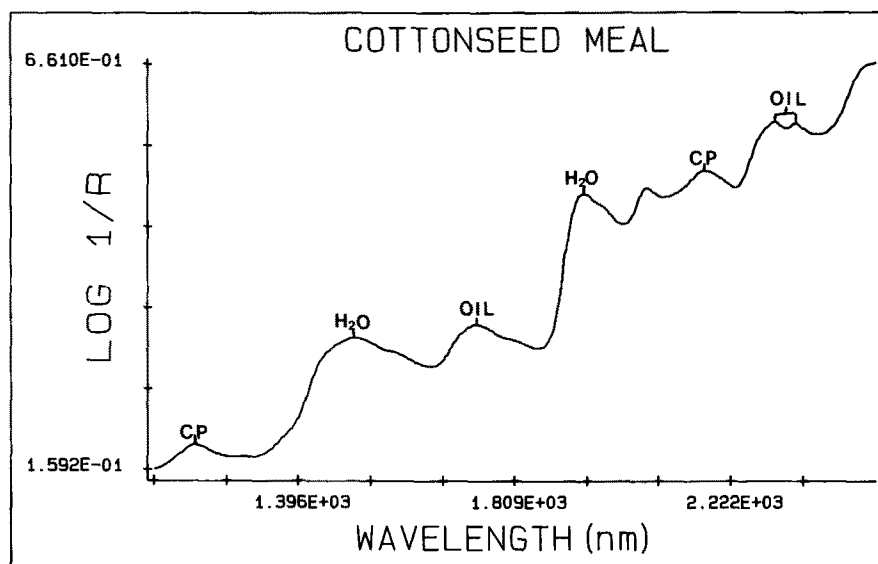FIG. 1. A plot of Log 1/R vs wavelength (nm) of the mathematical average spectrum of samples with prominent wavelengths that correlate to oil, water (H_2O) and protein (CP) noted on the spectrum.

the scatter the lower the initial absorbance (20,21). Eventually the significantly higher oil and CP content contribute to an overall more intense absorbance spectrum for the sample. The wavelength regions at which the constituents usually absorb have been indicated on Figure 1, the mean (average) spectrum of the nine-sample set. These wavelengths agree with what has been found previously (4,14). The initial calibration with nine samples indicated that the three constituents required different mathematical data treatments and different derivative widths to optimize the equations. Moisture and oil required a second derivative data treatment (Table 2), but oil required a much narrower derivative width (8 nm vs 16 nm/segment). Protein required an even wider width (24 nm/segment), but a reasonable fit was obtained with unsmoothed data consisting of the logarithm of reciprocal reflectance (log 1/R).

Nine samples can give an indication of the potential for the measurement of a given constituent by NIRS, but are too few for the development of analytical equations (22). The results from Birth (24) showed that with small sample sets (<20), high RSQ could be obtained with random noise.

A second set of 53 samples was obtained and added to the original nine samples. The range, mean and standard derivation of the resulting data set for H_2O, oil and CP were 4.6–11.1, 1.5–3.8 and 35.9–53.7; 8.1, 2.7 and 45.1; and 2.2, 0.7 and 5.0 percentage units. Two stepwise multiple regression analysis programs, "CAL" and "BEST", were used to develop equations. Both programs allow one to reserve every "I"th sample in the set for validation of the equations, in this case I=3. Table 3 contains the data on errors of calibration and prediction obtained from both regression programs.

Essentially, the results are the same as before; small standard error of calibration (SEC) and large RSQ values were found. The standard error of performance (SEP) for the three constituents also was quite small. The

TABLE 2

Mathematical Data Treatments for Expressed Cottonseed Meal

Variable	SEC	R^2	# Terms[a]	Math treatment (N,D,S,S)[b]
Moisture	0.026	.90	2	2, 16, 16, 2
Oil	0.055	.97	2	2, 8, 8, 2
Protein	0.122	.99	2	2, 24, 24, 2
Protein	0.239	.99	2	0, 0, 2, 2

[a]# Terms, the number of terms in the regression equation.
[b]Where N = derivative number, D = derivative segment width in nm and S = nm for first and second smoothing function (2 nm = no smoothing).

slope of actual vs predicted values showed the equation to be valid in that a slope of 1.0 would be perfect. There was no bias in the equations from program "BEST" and essentially none from "CAL."

The equation developed from the larger sample set required wider derivative widths than for the small set. The important thing is that all three constituents yielded improved results compared to those obtained from the logarithm of reciprocal reflectance (log 1/R).

Maddams in 1980 (23) reviewed curve fitting and derivative techniques to show their utility and limitations for enhancing resolution of spectra. Later, Bowley et al. (24) and Compton and Maddams (25) looked at Fourier Self-Deconvolution (FSDC) and compared it to derivative, as developed initially by Kauppiner et al. (26,27). According to these authors, derivative and FSDC can be used to resolve individual peaks in complex bands. In both resolution techniques spectral resolution is obtained at the expense of the signal-to-noise ratio (SNR). The FSDC retains the normal absorbance spectra while the lines narrow and the even derivatives

F.E. BARTON AND G.C. CAVANAGH

TABLE 3

Standard Error of Calibration and Prediction for Cottonseed Meal

Variable/ regression program	Sec[a]	RSQ	(SEP(c)[b]	R	Terms	Math treatment	Slope[c]	Bias
"CAL"								
Moisture	.199	.991	.184	.992	1	1, 16, 8, 4	1.007	−0.035
Oil	.099	.979	.088	.987	2	2, 24, 24, 4	1.019	−0.0004
Protein	.446	.991	.353	.996	1	2, 24, 16, 4	1.029	0.182
"BEST"								
Moisture	.176	.993	.172	.993	3	2, 16, 16, 2	1.019	0.000
Oil	.090	.983	.088	.984	3	2, 16, 4, 4	0.975	0.000
Protein	.352	.994	.343	.995	3	2, 24, 24, 2	0.980	0.000

[a]Standard error of calibration.
[b]Standard error of performance (corrected for bias).
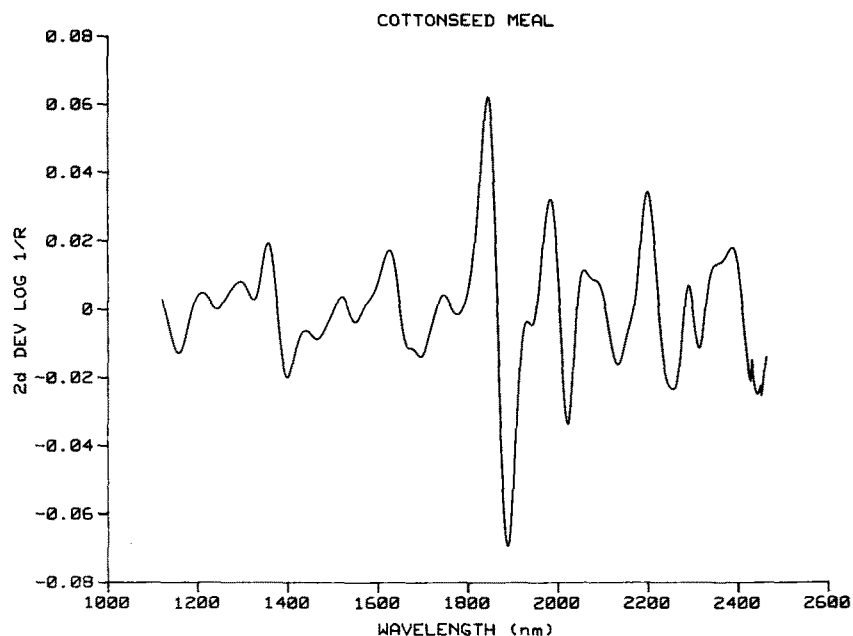[c]Slope of laboratory calibration data vs NIRS data.

FIG. 2. A plot of the second derivative of Log 1/R vs wavelength (nm) of the average spectrum in Fig. 1. The wavelength minimum can be compared to the maximum in Fig. 1 to show where the correlation with composition exists.

either give negative peaks (second derivative) or normal absorption peaks (fourth derivative) with side lobes.

Figure 2 is the second derivative of the average of the small cottonseed meal spectral file with a medium 24-nm derivative gap. Figure 3 shows the log 1/R cottonseed-meal spectral (3A) and the log 1/R spectra with the baseline removed by rational interpolation (3B). This baseline slope removal is necessary to properly take the inverse Fourier transform. Figure 4A and 4B are the log 1/R (baseline removed) and the superimposed FSDC of the cottonseed meal. The second derivative minima and the FSDC peaks should correspond to the wavelengths chosen by the statistics for the calibration equations. Table 4 shows that indeed they do. The peak

at 2298 nm is very likely the CH stretch of both the oil fraction and any cellulose in the sample.

Another way of assessing the utility of the more complex data treatments is to approximate a 19-filter instrument. This can be done by using a standard set of wavelengths (28) with the calibration data and performing a stepwise multiple regression as we would with the monochromator data. In this case we have only 19 wavelength intensities instead of 700 and use only the logarithm of reciprocal reflectance as a data treatment. The results are in Table 5. Basically, the SEC's are almost twice as large, the RSQ lower, and five terms are required as compared with one or two as is the case in Table 3. The "scatter plot" (Fig. 5) shows
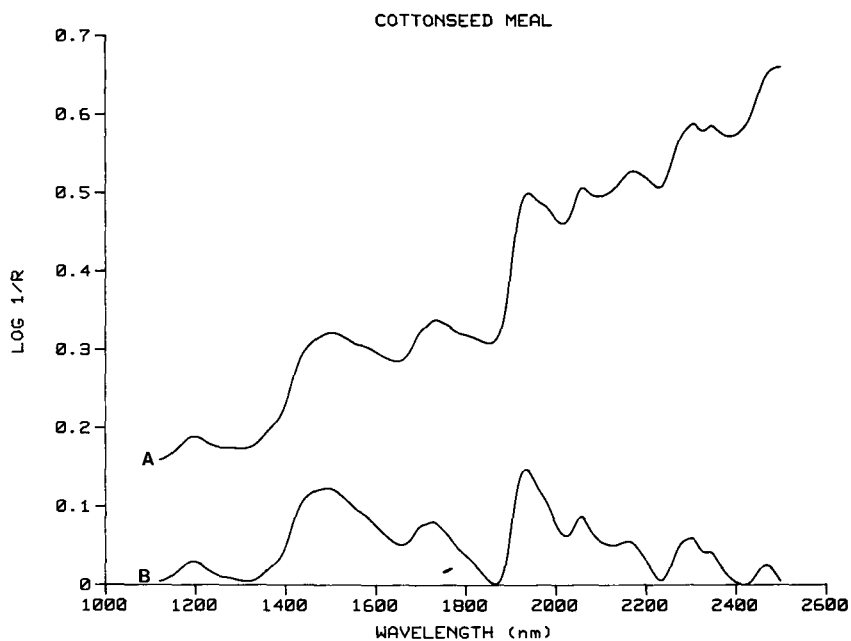
COTTONSEED MEAL

FIG. 3. Figure 3A is the same as Fig. 1 (i.e., Log 1/R vs wavelength for the average of 9 samples), and is shown to compare it with Fig. 3B which has had the sloping baseline removed by rational interpolation.
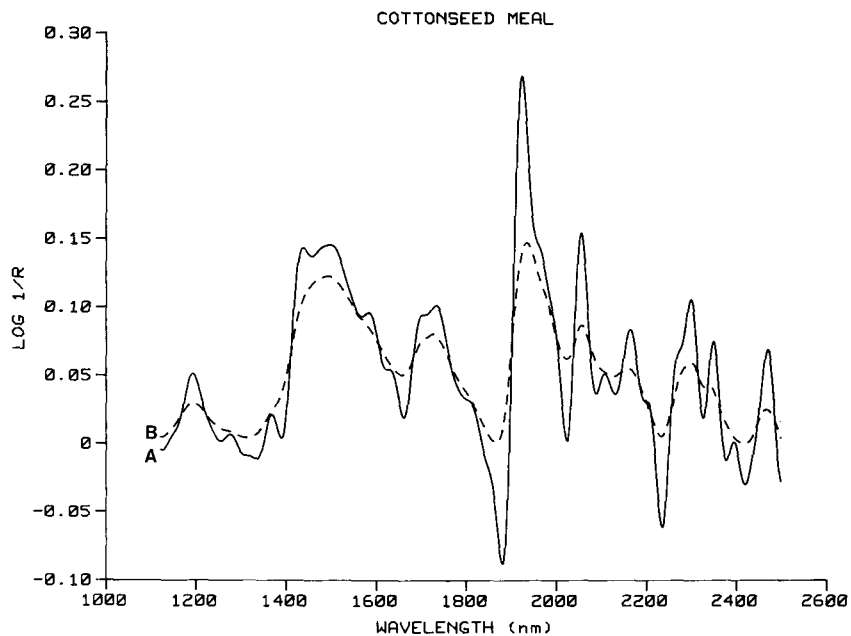
COTTONSEED MEAL

FIG. 4. Fig. 4A is a plot of Log 1/R vs wavelength (nm) with resolution enhancement via Fourier Self-Deconvolution. Fig. 4B is Fig. 3B repeated for comparison.

TABLE 4

Comparison of Wavelengths (nm) Chosen by Statistical Programs and Peak Assignments in the Spectral

Log 1/R[a]	FSDC[b]	Second derivative[c]	Cal/BEST[d]	Vibration[e]
	1274	1275		-CH aromatic
	1368	1362	1362	-O-H 1st overtime
	1438	1430	1436	N-H 1st overtime
1492	1496	1502		
		1685		
1706	1706	1705	1714	Oil CH 1st overtime
	1808	1805		
1938	1922	1920	1956	H-OH bend
2058	2055	2062	2076	C bend
2162	2164	2170	2166	N-H bend
2302	2298	2292	2286	CH-stretch comb
2342	2350	2350	2326	2nd overtime C-H
2466	2470	2480	2486	Ar-CH stretch

[a]Logarithm of reciprocal reflectance, Log 1/R.
[b]Fourier self-deconvolution.
[c]The second derivative with respect to wavelength of Log 1/R.
[d]Cal and Best are two step-wise multiple linear regression programs described in USDA Handbook 643 (18).
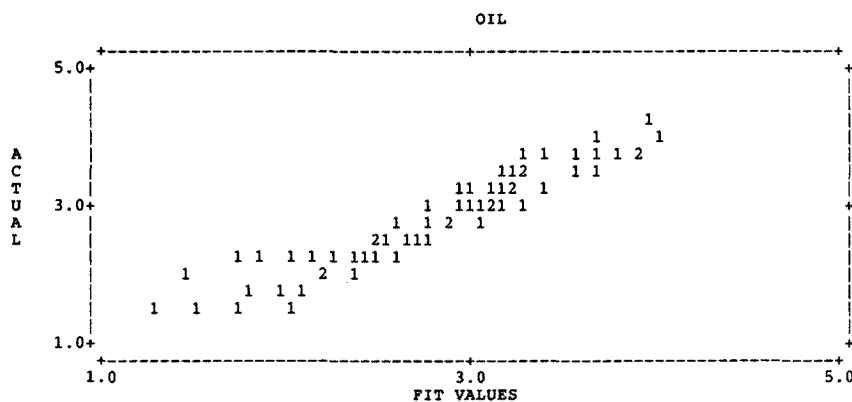[e]Type of vibrational motion typically found at this wavelength.

FIG. 5. Fig. 5 is a scatter plot of laboratory- versus NIRS-determined values for oil.

the problem to be one of precision. At the lower residual oil levels (<2%), the precision is almost one-half to equal the amount of oil (±0.6%). This raises the error of calibration to 75-100% at the lower levels averaged with a much tighter fit at >4% oil.

The conclusion that results from this study is that optimal evaluation of all three constituents on one instrument requires that the instrument be capable of yielding sufficient spectral information (e.g., monochromator or tilting filter) and have a large enough microcomputer to perform the appropriate mathematical data treatments and regression analyses. A fixed filter instrument could be and has been used to analyze for any one of these constituents or moisture and protein together, but it can give a loss of precision and

accuracy which could cause someone not thoroughly familiar with NIRS to draw an erroneous conclusion as to the suitability of NIRS for a particular analysis.

TABLE 5

Approximation of Fixed Filter Calibration

Variable	SEC[a]	RSQ[b]	# Terms[c]
Moisture	0.24	.99	5
Oil	0.19	.92	5
Protein	0.74	.98	5

[a]SEC, Standard error of calibration.
[b]RSQ, Squared correlation coefficient.
[c]# Terms, number of terms in the resulting regression equation.

REFERENCES

1. Hymowitz, T., J.W. Dudley, F.I. Collins and C.M. Brown, *Crop Sci. 14*:713 (1974).
2. Williams, P.C., *Cereal Chem. 52*:561 (1975).
3. Kaffka, K.J., K.H. Norris, J. Peredi and A. Balogh, *Acta Alimentaria 11*(3):253 (1982).
4. Rinne, R.W., S. Gibsons, J. Bradley, R. Seif and C.A. Brim, Soybeans Protein and Oil Percentages Determined by Infrared Analysis, Agric. Res. Service, USDA, ARS, NC-26, July 1975.
5. Norris, K.H., and R.F Barnes, in *First International Symposium, Feed Composition, Animal Nutrient Requirements, and Computerization of Diets*, edited by P.V. Fonnesbeck, L.E. Harris and L.C. Kearl, Utah Agr. Exp. Sta., Utah State University, Logan, Utah, 1976, pp. 237-241.
6. Norris, K.H., R.F Barnes, J.E. Moore and J.S. Shenk, *J. Anim. Sci. 43*:889 (1976).
7. Shenk, J.S., W.N. Mason, M.L. Risius, K.H. Norris and R.F Barnes, in *First International Symposium on Feed Composition, Animal Nutrient Requirements and Computerization of Diets*, edited by P.V. Fonnesbeck, L.E. Harris and L.C. Kearl, Utah State University, Logan, UT, 1976, pp. 242-247.
8. Shenk, J.S., and M.R. Hoover, in *Technicon International Congress, Vol. 2, Industrial Symposia*, New York, NY, 1976, pp. 122-125.
9. Shenk, J.S., K.H. Norris, R.F Barnes and G.W. Fissel, in *Proceedings of the 13th International Grasslands Congress*, Leipzig, German Democratic Republic, 1977, pp. 454-463.
10. Shenk, J.L., M.O. Westerhaus and M.R. Hoover, in *Proceedings International Grain and Forage Harvesting Conference*, American Society of Agricultural Engineers, St. Joseph, MI, 1978, pp. 242-244.
11. Shenk, J.S., I. Lada, M.R. Hoover and M.O. Westerhaus, *Crop Sci. 21*:355 (1981).
12. Norris, K.H., *Agric. Eng. 45*:370 (1964).
13. Robertson, J.A., and W.R. Windham, *J. Am. Oil Chem. Soc. 60*:1773 (1983).
14. Robertson, J.A., and F.E. Barton II, *Ibid. 61*:543 (1984).
15. Issac, R.A., and W.C. Johnson, *J. Assoc. Off. Anal. Chem. 67*:506 (1984).
16. *Association of Official Analytical Chemists Methods of Analysis*, 14th edn., AOAC, Arlington, VA, 1985, Methods 7.015, 28.002 and 28.033.
17. Shenk, J.S., in *Near Infrared Reflectance Spectroscopy (NIRS): Analysis of Forage Quality*, edited by G.C. Marten, J.S. Shenk and F.E. Barton II, U.S. Department of Agriculture, Agriculture Handbook No. 643, 1985, p. 96.
18. *Near Infrared Reflectance Spectroscopy (NIRS): Analysis of Forage Quality*, edited by G.C. Marten, J.S. Shenk and F.E. Barton II, USDA, Agriculture Handbook No. 643, 1985, p. 96.
19. Petersen, J.C., W.R. Windham, L.G. Brown and C.S. Hoveland, in *Proc. Amer. Forage and Grassland Council*, American Forage and Grassland Council, Lexington, KY, 1987, pp. 151-155.
20. Birth, G.C., in *Quality Detection in Foods*, edited by J.J. Gaffney, Am. Soc. Agric. Eng., 1976, pp. 6-11.
21. Westerhaus, M.O., in *Near Infrared Reflectance Spectroscopy (NIRS): Analysis of Forage Quality*, edited by G.C. Marten, J.S. Shenk and F.E. Barton II, U.S. Department of Agriculture, Agriculture Handbook No. 643, 1985, p. 960.
22. Birth, G.S., *J. Appl. Spec. 39*:729 (1985).
23. Maddams, W.F., *J. Appl. Spec. 34*:245 (1980).
24. Bowley, H.J., S.M.H. Collin, D.L. Gerrard, D.I. James, W.F. Maddams, P.B. Tooke and I.D. Wyatt, *Ibid 39*:1004 (1985).
25. Compton, D.A., and W.F. Maddams, *Ibid. 40*:239 (1986).
26. Kauppinen, J.K., D.J. Moffat, H.H. Mantsch and D.G. Cameron, *Anal. Chem. 53*:1454 (1981).
27. Kauppinen, J.K., D.J. Moffat, D.G. Cameron and H.H. Mantsch, *Appl. Optics 20*:1866 (1981).
28. Williams, P.C., K.R. Preston, K.H. Norris and P.M. Starkey, *J. Food Sci. 49*:17 (1984).