

Application of Cluster Analysis (CLA) in Feed Chemical Imaging To Accurately Reveal Structural–Chemical Features of Feeds and Plants within Cellular Dimension

PEIQIANG YU[†]

College of Agriculture, 6D10 Agriculture Building, University of Saskatchewan, 51 Campus Drive, Saskatoon, Saskatchewan, Canada S7N 5A8

Synchrotron Fourier transform infrared (FTIR) microspectroscopy can explore molecular chemical features of the microstructure of feeds. The most straightforward method of data analysis is the mapping of specific functional group intensities and frequencies by peak heights and/or areas. However, this univariate statistical method does not always accurately identify functional group locations and concentrations, because the so-called “unique” bands for the peak area mapping have more or less inference with other nonunique bands in feed and plant tissues. The objective of this study was to use a multivariate analysis method—called agglomerative hierarchical cluster analysis (CLA)—to analyze infrared spectra for chemical imaging. The results show the CLA method gave satisfactory results and was conclusive in showing that it can discriminate and classify functional group differences existing in different structure regions. This approach (CLA for chemical imaging) places synchrotron FTIR microspectroscopy at the forefront of those techniques that could potentially be used in the rapid characterization of feed microstructure.

KEYWORDS: Synchrotron; infrared spectroscopic characteristics; cluster analysis; imaging; feeds; plant

INTRODUCTION

Synchrotron Fourier transform infrared (FTIR) microspectroscopy can explore molecular chemical features of the microstructure of biological samples (1–4), which include plants and cereal grains designated for animal feeds (5–7). The microstructural–chemical features of feeds could be related to feed digestive behaviors in animals (7–9).

The most straightforward method of spectral data analysis is the mapping of specific functional group intensities and frequencies by peak heights and/or areas (3). However, this univariate statistical approach does not always accurately identify functional group locations and concentrations even though individual components are uniquely identified, particularly when different functional group spectra are in the same infrared region but show different band patterns. For example, structural carbohydrates and nonstructural carbohydrates in feed and plant tissues both show peaks at the ~ 1180 – 950 cm^{-1} region. Himmelsbach et al. (10) reported that the band at 1515 cm^{-1} shows no inference with any other bands, and thus it is an ideal diagnostic band for aromatics. However, our study showed that the lignin band under the ~ 1515 cm^{-1} peak has more or less inference with part of amide I band.

Cluster analysis (CLA) is a multivariate analysis; function performs an (agglomerative hierarchical) cluster analysis of an infrared spectral data set and displays the results of CLA as images and dendrograms. First, it calculates the distance matrix, which contains information on the similarity of spectra. Then, in hierarchical clustering, the algorithm searches within the distance matrix for the two most similar IR spectra (minimal distance). These spectra are combined into a new object (called a “cluster”). The spectral distances between all remaining spectra and the new cluster are recalculated (11). It is a technique that clusters infrared spectra in a map based on similarity with other spectra in the same map (3).

The objective of this study was to use CLA to analyze infrared spectra for chemical imaging of feed and plant microstructure to reveal structural–chemical features at a cellular level and to illustrate the importance of CLA in feed microstructural imaging. The structural–chemical features of feeds can be related to feed digestive behaviors in animals.

MATERIALS AND METHODS

Selected Feed Example for Imaging. Grain corn (cv. Pioneer 39P78) at harvest maturity was selected for chemical imaging to illustrate the importance of CLA in the chemical imaging of feeds. The corn samples were obtained from Henry Penner (Morden, MB, Canada), arranged by the Prairie Feed Resource Centre (Canada) (Director, Vern Racz).

[†] Telephone (306) 966-4132; fax (306) 966-4151; e-mail yupe@sask.usask.ca.

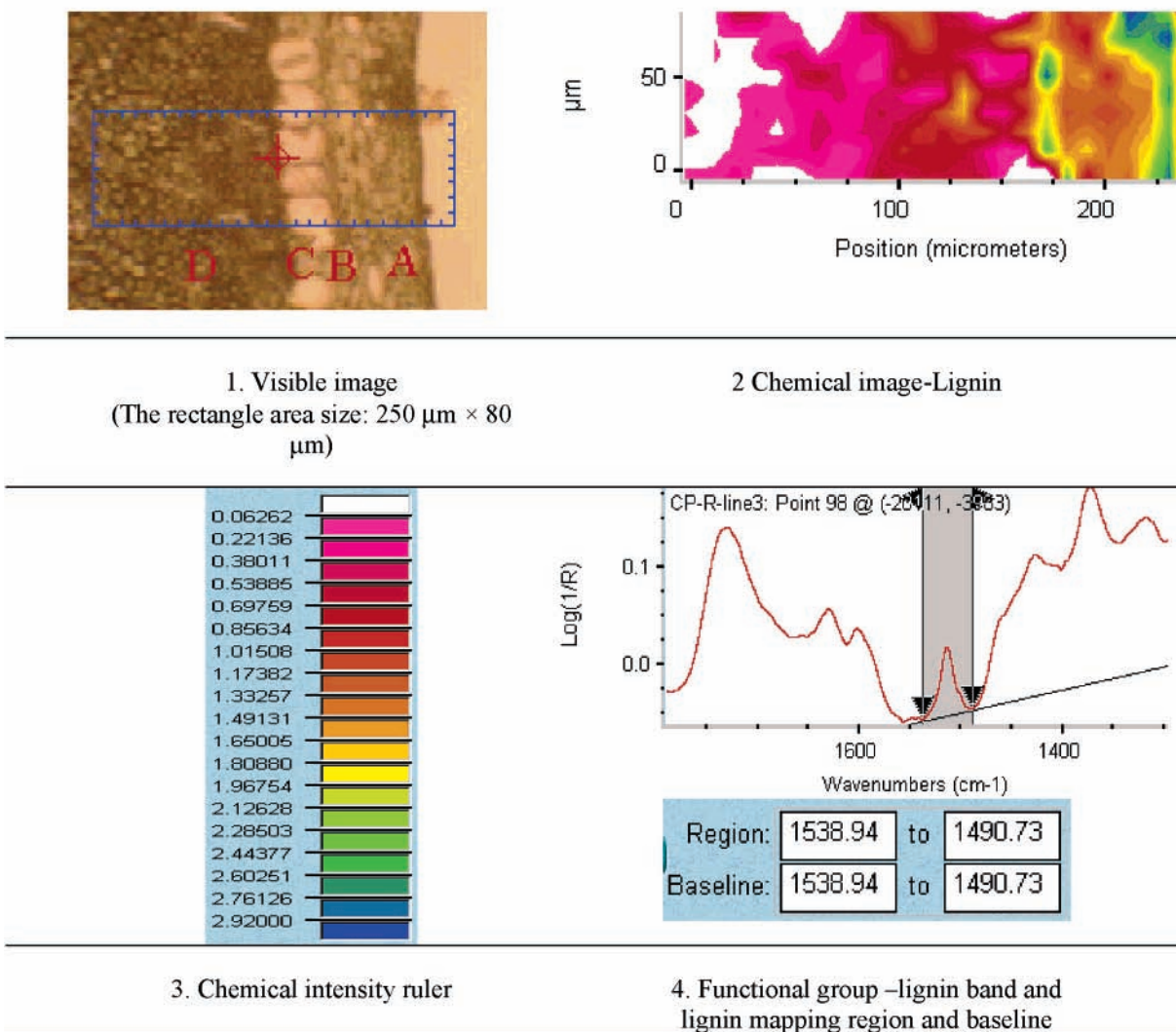


Figure 1. Lignin image of the Pioneer corn tissue from the pericarp (A, in visible image), seed coat (B), aleurone (C), and endosperm (D): area under 1515 cm^{-1} peak (lignin aromatic compound) with mapping region and baseline 1538.94–1490.73 cm^{-1} .

Synchrotron FTIR Slide Preparation. The corn was cut into thin cross sections ($\sim 6 \mu\text{m}$ thick) at the Western College of Veterinary Medicine, University of Saskatchewan. The unstained cross sections were mounted onto low-E IR microscope slides (Kevley Technologies, Chesterland, OH) for synchrotron FTIR microspectroscopy in a reflectance mode.

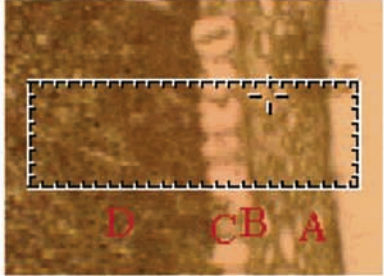
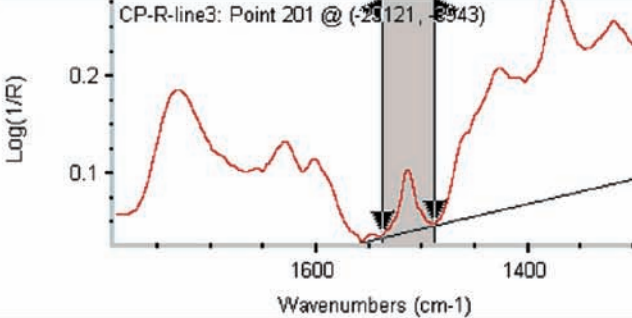
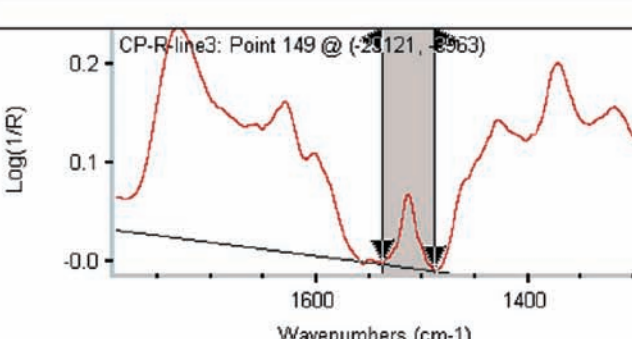
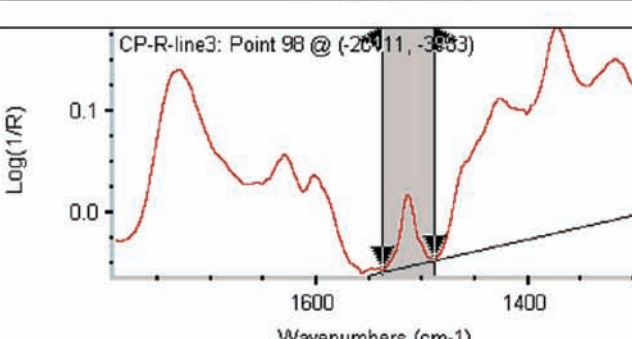
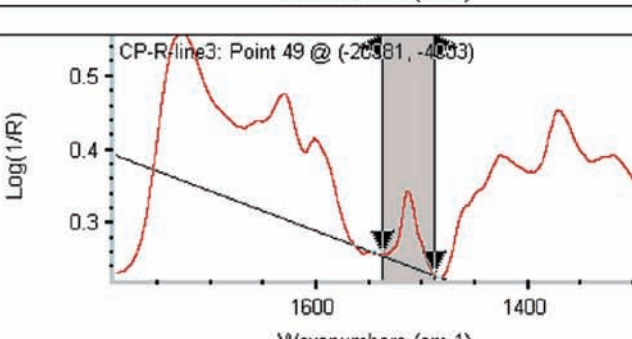
Synchrotron Light Source and Synchrotron FTIR Microspectroscopy. The experiment was carried out on the U2B beamline at the Albert Einstein Synchrotron Bioscience Center in National Synchrotron Light Source at the Brookhaven National Laboratory (NSLS-BNL, Upton, NY). The beamline was equipped with a FTIR spectrometer (Nicolet Magna 860) with a KBr beam splitter and a mercury cadmium telluride (MCT-A) detector coupled with an infrared microscope (Nic-Plan, Nicolet Instruments, Madison, WI). The infrared spectra were collected in the mid-infrared range, 4000–800 cm^{-1} , at a resolution of 4 cm^{-1} with 64 scans co-added and an aperture setting of ca. 10 $\mu\text{m} \times 10 \mu\text{m}$. To minimize infrared absorption by CO_2 and water vapor in ambient air, the optics were purged using dry N_2 . A background spectroscopic image file was collected from an area free of sample. The mapping steps were equal to aperture size. Stage control, data collection, and basic processing were performed using OMNIC 6.0 (Thermo-Nicolet, Madison, WI). Scanned visible images were obtained

using a charge-coupled device (CCD) camera linked to the infrared images (objective $\times 35$).

Chemical Imaging. The spectral data of the Pioneer corn tissues were collected, and the background spectrum was corrected, displayed in the absorbance mode, and analyzed using OMNIC software 6.0. The data can be displayed either as a series of spectroscopic images collected at individual wavelengths or as a collection of infrared spectra obtained at each pixel position in the image. The structural chemistry of functional groups from the pericarp, seed coat, and aleurone layer to endosperm was imaged under peaks at ca. 1515, 1650, 1736, and 1180–950 cm^{-1} using the OMNIC software 6.0 by peak area mapping. False color maps were used, which were derived from the area under particular spectral features.

Hierarchical Cluster Analysis. Multivariate analysis, CLA, was used to perform an agglomerative hierarchical cluster analysis of infrared spectra data sets, and the results of CLA were displayed as images and dendrograms. For CLA, Ward's algorithm was used for clustering. Spectral distances were computed as D values. The detailed method was as follows: (1) First, a distance matrix, which contains information on the similarity of spectra, was calculated. (2) Next, in hierarchical clustering the algorithm searched within the distance matrix for the two most similar IR spectra (minimal distance). (3) These spectra

(a)

The spot samples in pericarp region (A, in visible image)	Spectra selected from corresponding pixels (Left column) from the visible image in corn pericarp region (A)
	
Pixel 201	
Pixel 149	
Pixel 98	
Pixel 49	<div style="border: 1px solid black; padding: 5px; width: fit-content; margin: auto;"> <p>Region: 1538.94 to 1490.73</p> <p>Baseline: 1538.94 to 1490.73</p> <p style="text-align: center;">Lignin mapping region and baseline</p> </div>
Pericarp region	

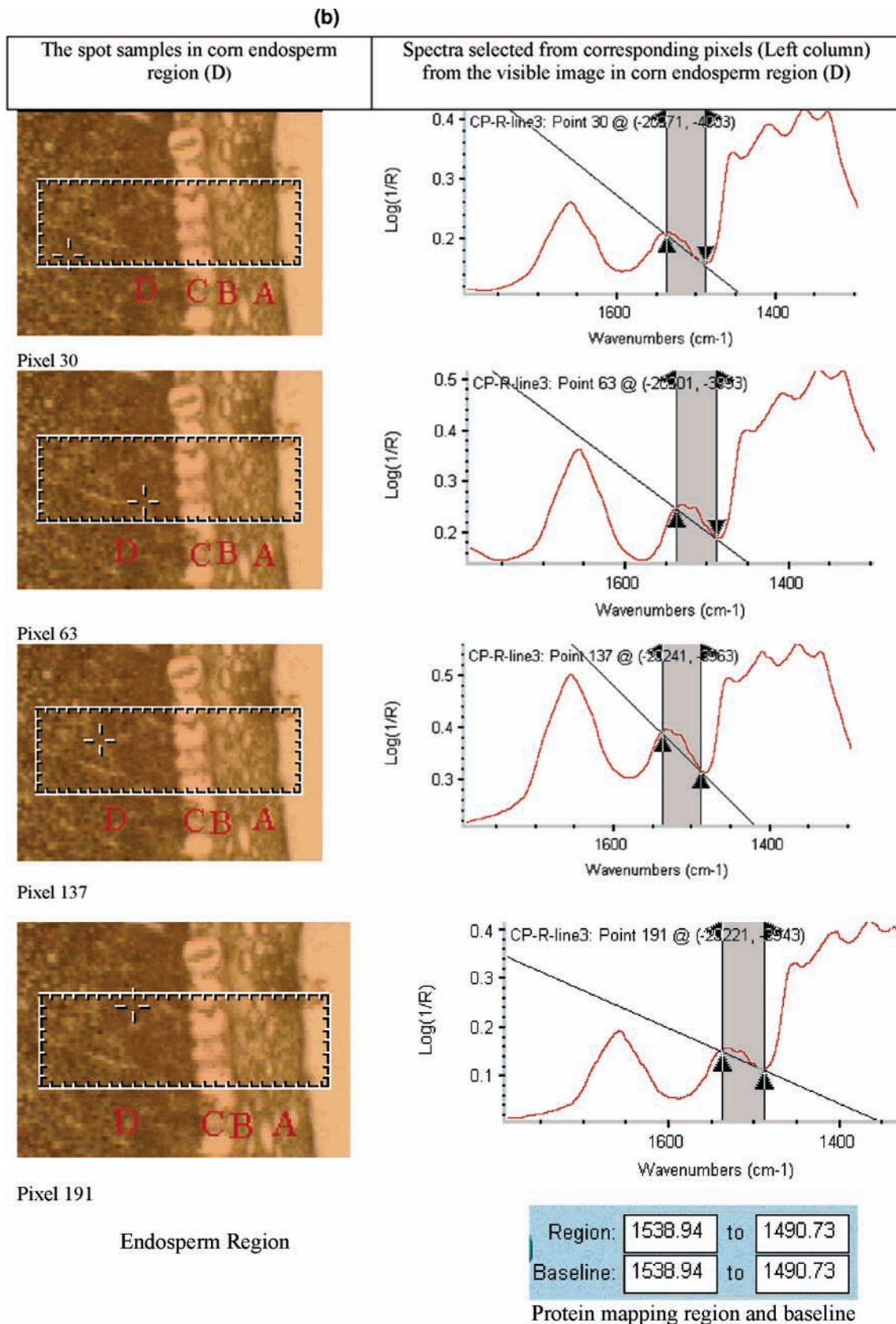


Figure 2. Spectra in pericarp and endosperm regions of grain corn tissue selected from corresponding pixels from the visible images, showing the area under the 1515 cm^{-1} peak (aromatic compound) in the pericarp for lignin imaging (see **Figure 1a**) covering part of amide II band in the endosperm (see **Figure 1b**), indicating lignin imaging under the 1515 cm^{-1} peak having inference with part of the amide I band and thus not accurately representing lignin intensity and distribution in the tissue.

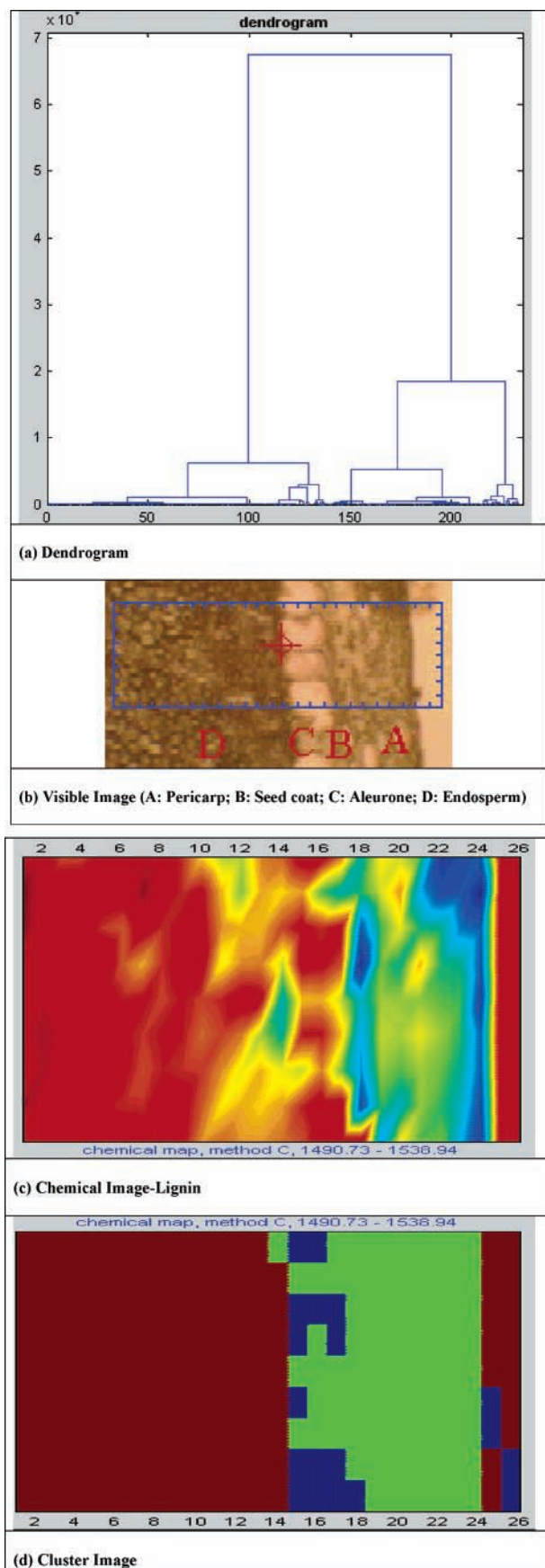


Figure 3. CLA cluster image of lignin clearly showing that the clusters in the pericarp region (green) and endosperm region (brown) are different [CLA analysis: (1) select spectral region, 1538.94–1490.73 cm^{-1} ; (2) distance method: *D* values; (3) cluster method, Ward's algorithm; (4) number of clusters, three in the cluster image].

were combined into a new object called a “cluster”. (4) The spectral distances between all remaining spectra and the new cluster were recalculated. (5) The program performed a new search for the two most similar objects (objects may be spectra or cluster). These objects were merged and, again, the distance values of the newly formed cluster were calculated. The displayed average spectra were encoded by the same color utilized for displaying the cluster in the CLA image. The dendrograms and CLA images were reported. In a dendrogram, spectral distances are given on the Y-axis. Horizontal lines illustrate the merging process, for example, the combination of two clusters to a new cluster of spectra.

RESULTS AND DISCUSSION

Univariate and Multivariate Statistical Approaches for Chemical Mapping. Statistical approaches to analyze spectral data collected under synchrotron FTIR microspectroscopy usually include uni- and multivariate statistical methods. The univariate methods of analysis consist of various mapping displays of spectral data. Usually researchers may select band intensities, integrated intensities, band frequencies, band intensity ratios, etc., to construct color maps of the spectral data. For example, feed chemical mapping (5, 6). The multivariate methods of data analysis create spectral corrections and maps by including not just one intensity or frequency point of a spectrum but by utilizing the entire spectral information. The methods include CLA (3).

Chemistry Mapping To Reveal Structural Features of the Feeds. In animal feeds, lignin, protein, lipid, and carbohydrates are four major components that are significantly related to animal digestive behavior and nutrient utilization. Lignin is the most indigestible component. Lipids and carbohydrates provide the energy sources, and protein provides a protein source for animals. Therefore, these four important components have been chosen for chemical image study.

The molecular chemistry of functional groups from the pericarp, seed coat, aleurone layer, and part of the endosperm was imaged under peaks at ca. 1515 cm^{-1} (vibrations of aromatic character of lignin) (**Figure 1**), 1650 cm^{-1} (amide I C=O) (**Figure 4**), 1736 cm^{-1} (lipid carbonyl C=O ester), and 1180–950 cm^{-1} (carbohydrate C–O stretching) (figures not shown). These functional group band assignments were made according to prior publications (1, 10, 12). They are considered to be “unique” bands for chemical mapping of lignins, proteins, lipids, and carbohydrates.

The absorption band of lignin is found at $\sim 1515 \text{ cm}^{-1}$ in the mid-infrared region. This is considered to be indicative of the aromatic character of the lignin. Aromatic compounds give two major bands ca. 1600 and 1500 cm^{-1} regions, referred to as quadrant and semicircle ring stretch (10, 13). These are well exemplified in the lignin spectrum (softwood) that shows bands at ca. 1595 and 1515 cm^{-1} . It was reported (10) that “the first of these bands is possible interferent with the other bands and the second of these bands, at 1515 cm^{-1} , shows no inference with any other bands and thus it is an ideal diagnostic band for aromatics”.

However, our spectral data analysis results (**Figure 2**) show that the band at $\sim 1515 \text{ cm}^{-1}$ does have inference with part of the amide I band. The spectra in the pericarp and endosperm

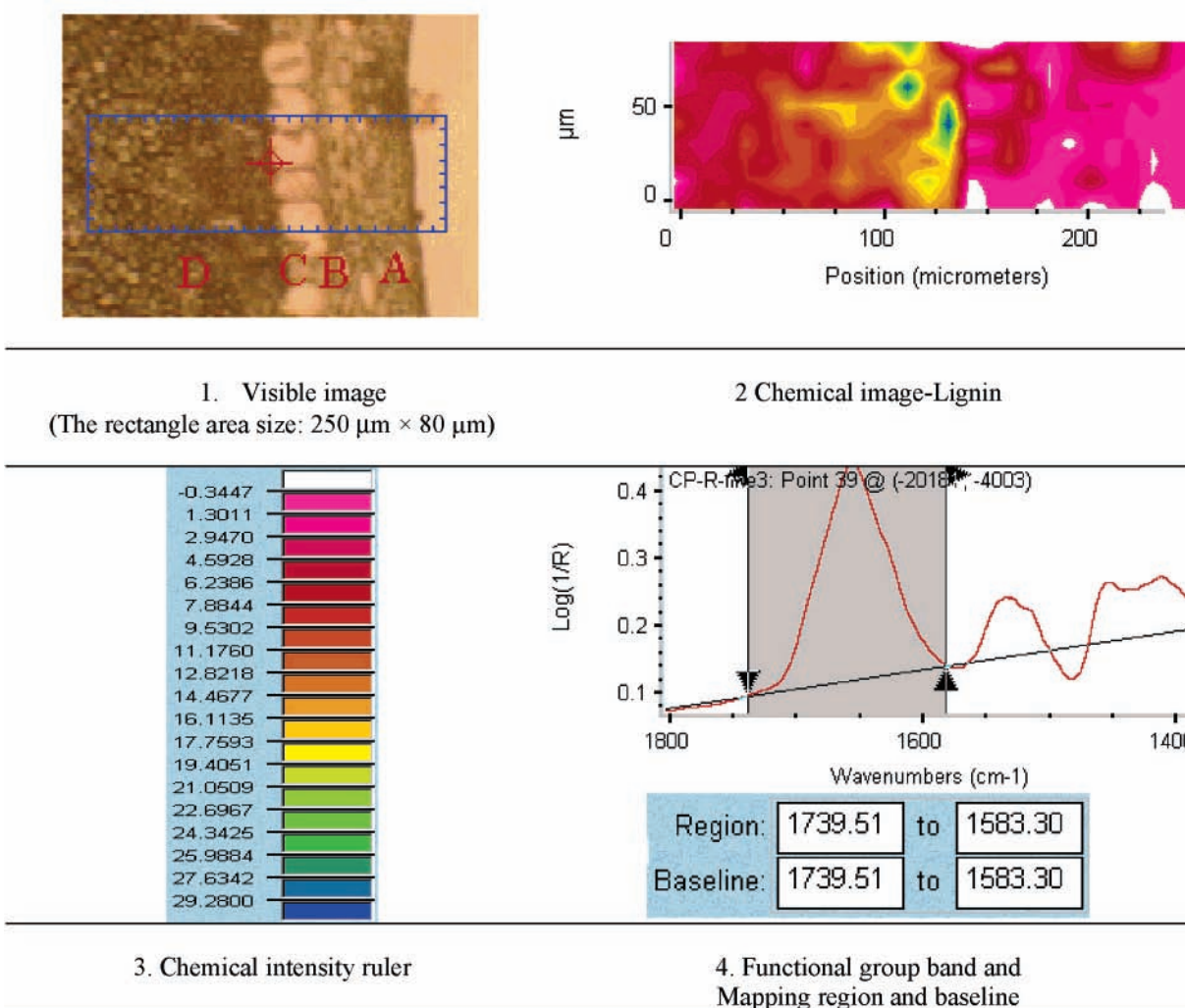


Figure 4. Protein image of the Pioneer corn tissue from the pericarp (A, in visible image), seed coat (B), aleurone (C) and endosperm (D): area under the 1650 cm^{-1} peak (protein amide I) with mapping region and baseline 1739.51–1583.30 cm^{-1} .

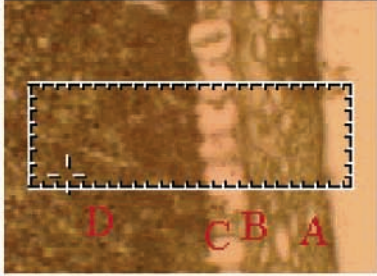
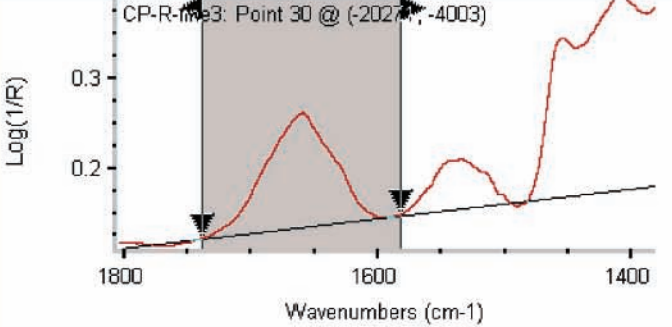
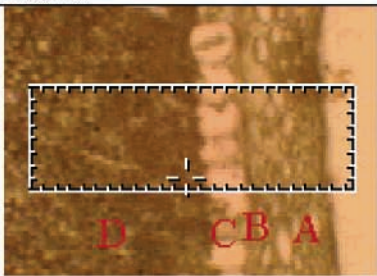
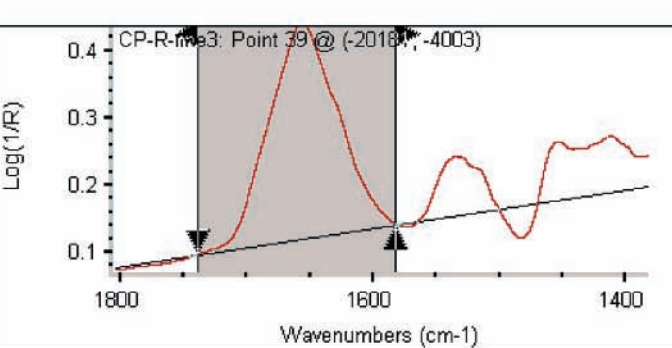
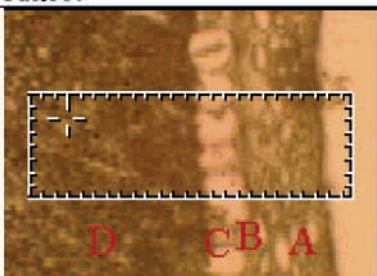
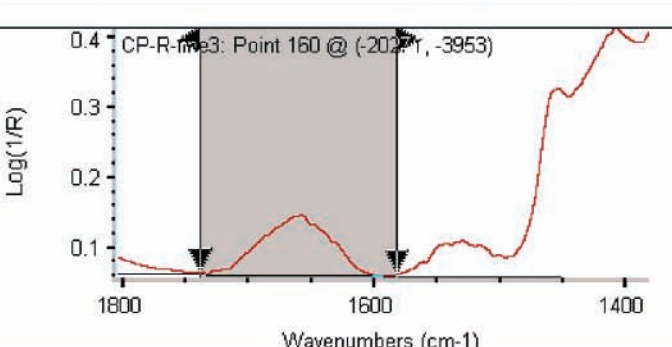
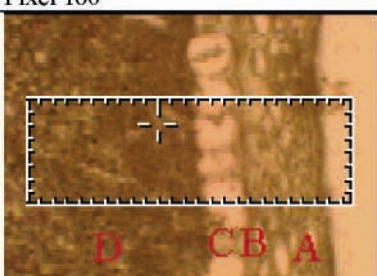
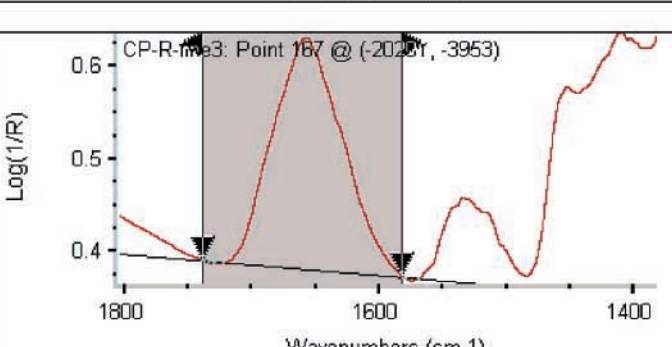
regions of the corn tissue selected from the corresponding pixels from the visible images showed that the area under the 1515 cm^{-1} peak (aromatic compound) in the pericarp for lignin imaging (Figure 2a) covers part of the amide II band in the endosperm (Figure 2b). This indicates that lignin imaging under the $\sim 1515 \text{ cm}^{-1}$ peak has inference with other bands and thus does not accurately represent lignin intensity and distribution in the whole region. Our CLA (Figure 3) further proved that the clusters in the pericarp and endosperm regions were completely different after the lignin spectral region from 1538.94 to 1490.73 cm^{-1} was selected for CLA.

The protein infrared spectrum has two primary features, the amide I ($\sim 1600\text{--}1700 \text{ cm}^{-1}$) and amide II ($\sim 1500\text{--}1560 \text{ cm}^{-1}$) bands, which arise from specific stretching and bending vibrations of the protein backbone. The amide I band arises predominantly from the C=O stretching vibration of the amide C=O group. The frequency of the amide I band is particularly sensitive to protein secondary structure (2, 4) and can be used to predict protein secondary structure. The amide II (predominantly an N–H bending vibration coupled to C–N stretching) is also used to assess protein conformation. However, as it arises

from complex vibrations involving multiple functional groups, it is less useful for protein structure prediction than the amide I (14). Figure 4 is a chemical image under the area at $\sim 1650 \text{ cm}^{-1}$ with the baseline 1739.51–1583.30 cm^{-1} , showing the area under the 1650 cm^{-1} peak attributed to protein absorption. However, the spectral results (Figure 5) show that the unique amide I band at $\sim 1650 \text{ cm}^{-1}$ in the endosperm region has overlapped with non-amide I bands in the pericarp region. The CLA (Figure 6) shows that the spectral clusters in the pericarp and endosperm regions differed. Our results (figures not shown) also show that the unique band for lipid at $\sim 1736 \text{ cm}^{-1}$ (which arises from the stretching vibration of the carbonyl ester C=O groups) has inference with non-carbonyl C=O ester bands in the endosperm region. The carbohydrate image under the peak at 1180–950 cm^{-1} also could not distinguish structural and nonstructural carbohydrates. With CLA, spectral clusters in the pericarp region and endosperm region were distinguished.

In conclusion, the CLA for chemical imaging gave satisfactory imaging results and was conclusive in showing that it can discriminate and classify functional groups existing in different structures in chemical maps of feeds.

(a)

The spot samples in corn pericarp region (A)	Spectra selected from corresponding pixels (Left column) from the visible image in corn pericarp region (A)								
									
<p>Pixel 30</p> 									
<p>Pixel 39</p> 									
<p>Pixel 160</p> 									
<p>Pixel 167</p> <p style="text-align: center;">Endosperm Region</p>	<table border="1" style="margin-left: auto; margin-right: auto;"> <tr> <td style="padding: 2px;">Region:</td> <td style="padding: 2px;">1739.51</td> <td style="padding: 2px;">to</td> <td style="padding: 2px;">1583.30</td> </tr> <tr> <td style="padding: 2px;">Baseline:</td> <td style="padding: 2px;">1739.51</td> <td style="padding: 2px;">to</td> <td style="padding: 2px;">1583.30</td> </tr> </table> <p style="text-align: center;">Lignin mapping region and baseline</p>	Region:	1739.51	to	1583.30	Baseline:	1739.51	to	1583.30
Region:	1739.51	to	1583.30						
Baseline:	1739.51	to	1583.30						

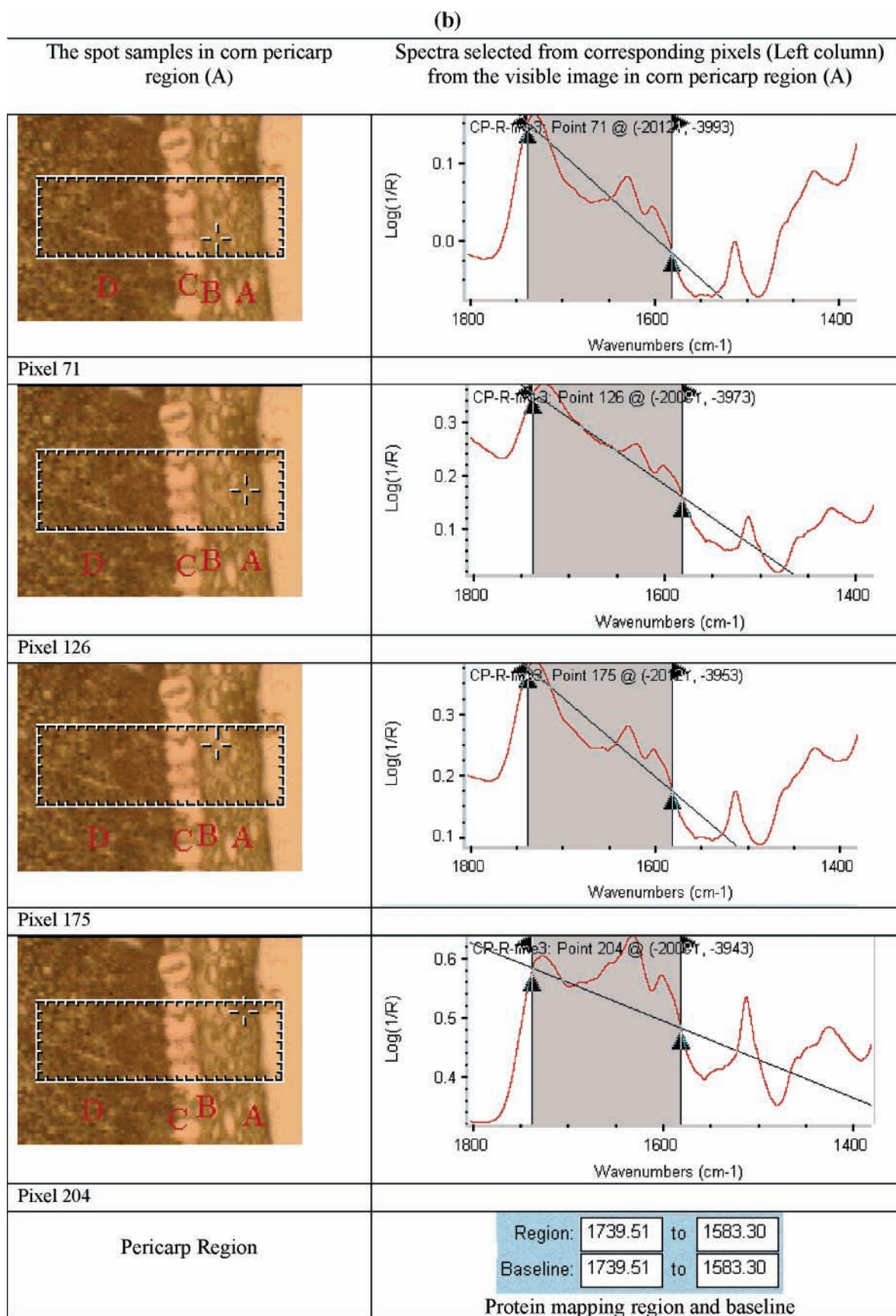


Figure 5. Spectra in the pericarp and endosperm regions of the corn tissue selected from corresponding pixels from the visible images, showing the area under the 1650 cm^{-1} peak (amide I band) in the endosperm (see **Figure 5a**) for protein imaging covering several non-amide I bands in the pericarp (**Figure 5b**), indicating protein imaging under the 1650 cm^{-1} peak does not accurately represent protein intensity and distribution.

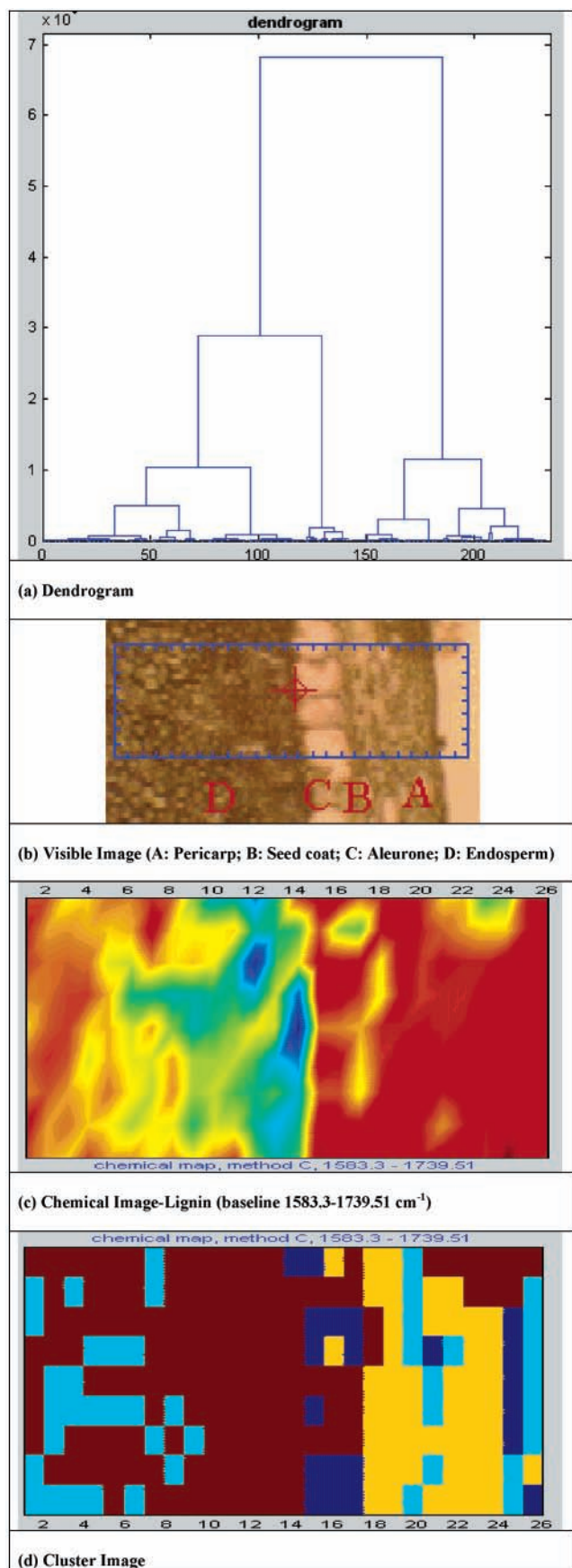


Figure 6. CLA cluster image of protein clearly showing that the clusters in the pericarp region and endosperm region are different [CLA analysis: (1) select spectral region, 1739.51–1583.30 cm^{-1} ; (2) distance method, D values; (3) cluster method, Ward's algorithm; (4) number of clusters, four in the cluster image].

ACKNOWLEDGMENT

We are grateful to Drs. N. S. Marinkovic and L. M. Miller (NSLS-BNL, Upton, NY) for helpful discussion and data collection at the U2B station in NSLS and to Vern J. Racz (Director, Prairie Feed Resource Centre) for arranging the feed sample. The work was done in full support by Drs. J. J. McKinnon and D. A. Christensen (University of Saskatchewan) and C. R. Christensen (Canadian Light Source).

LITERATURE CITED

- (1) Wetzel, D. L.; Eilert, A. J.; Pietrzak, L. N.; Miller, S. S.; Sweat, J. A. Ultraspatially resolved synchrotron infrared microspectroscopy of plant tissue in situ. *Cell. Mol. Biol.* **1998**, *44*, 145–167.
- (2) Miller, L. M. The impact of infrared synchrotron radiation on biology: past, present, and future. *Synchrotron Radiat. News* **2000**, *13*, 31–37.
- (3) Miller, L. M. Infrared microspectroscopy and imaging. Retrieved in October 2002 from <http://nslsweb.nsls.bnl.gov/nsls/pubs/nslspubs/imaging0502/irxrayworkshopintroduction.ht>.
- (4) Marinkovic, N. S.; Huang, R.; Bromberg, P.; Sullivan, M.; Toomey, J.; Miller, L. M.; Sperber, E.; Moshe, S.; Jones, K. W.; Chouparova, E.; Lappi, S.; Franzen, S.; Chance, M. R. Center for Synchrotron Biosciences' U2B beamline: an international resource for biological infrared spectroscopy. *J. Synchrotron Radiat.* **2002**, *9*, 189–197.
- (5) Yu, P.; McKinnon, J. J.; Christensen, C. R.; Christensen, D. A. Mapping plant composition with synchrotron infrared microspectroscopy and relation to animal nutrient utilization. Presented at the Canadian Society of Animal Science Annual Conference, University of Saskatchewan, Saskatoon, SK, Canada, June 10–13, 2003 (invited article and conference speech).
- (6) Yu, P.; McKinnon, J. J.; Christensen, C. R.; Christensen, D. A.; Marinkovic, N. S.; Miller, L. M. Chemical imaging of microstructures of plant tissues within cellular dimension using synchrotron infrared microspectroscopy. *J. Agric. Food Chem.* **2003**, *51*, 6062–6067.
- (7) Yu, P. Application of advanced synchrotron-based Fourier transform infrared microspectroscopy (SR-FTIR) to animal nutrition and feed science: a novel approach. *Br. J. Nutr.* **2004**, *92*, 869–885.
- (8) Yu, P.; McKinnon, J. J.; Christensen, C. R.; Christensen, D. A. Using synchrotron transmission FTIR microspectroscopy as a rapid, direct and nondestructive analytical technique to reveal molecular microstructural–chemical features within tissue in grain barley. *J. Agric. Food Chem.* **2004**, *52*, 1484–1494.
- (9) Yu, P.; Christensen, D. A.; Christensen, C. R.; Drew, M. D.; Rosnagel, B. G.; McKinnon, J. J. Use of synchrotron Fourier transform infrared microspectroscopy to identify chemical differences in barley endosperm tissue in relation to rumen degradation characteristics. *Can. J. Anim. Sci.* **2004**, *84*, 523–527.
- (10) Himmelsbach, D. S.; Khalili, S.; Akin, D. E. FT-IR microscopic imaging of flax (*Linum usitatissimum* L.) stems. *Cell. Mol. Biol.* **1998**, *44*, 99–108.
- (11) Cytospec. Software for infrared spectral imaging, v. 1.1.01, 2004.
- (12) Wetzel, D. L. When molecular causes of wheat quality are known, molecular methods will supersede traditional methods. Presented at the International Wheat Quality Conference II, Manhattan, KS, May 2001; pp 1–20.
- (13) Colthup, N. B.; Daly, L. H.; Wiberley, S. E. *Introduction to Infrared and Raman Spectroscopy*, 3rd ed.; Academic Press: Boston, MA, 1990; p 547.
- (14) Jackson, M.; Mantsch, H. H. Infrared spectroscopy ex vivo tissue analysis. In: *Biomedical Spectroscopy*. **2000**.

Received for review December 3, 2004. Accepted February 24, 2005. This research has been supported by grants from the Saskatchewan Agricultural Development Fund (ADF). The National Synchrotron Light Source in Brookhaven National Laboratory (NSLS-BNL, Upton, NY) is supported by the U.S. Department of Energy, Contract DE-AC02-98CH10886.