DOI:10.1021/jf9001523

JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY
ARTICLE

Molecular Classification of Barley (*Hordeum vulgare* L.) Mutants Using Derivative NIR Spectroscopy

Paul R. Wiley,[†] Greg J. Tanner,[†,§] Peter M. Chandler,*[,†,§] and Robert S. Anderssen[#]

[†]CSIRO—Plant Industry, G.P.O. Box 1600, Canberra, ACT 2601, Australia, [§]CSIRO—Food Futures National Research Flagship, G.P.O. Box 1600, Canberra, ACT 2601, Australia, and [#]CSIRO—Mathematical and Information Sciences, G.P.O. Box 664, Canberra, ACT 2601, Australia

Near-infrared reflectance (NIR) spectroscopy was used in the characterization of grain morphology mutants of barley (*Hordeum vulgare* L.) in relation to grain nitrogen (N) content and protein composition. Derivative spectroscopy provided spectra with enhanced resolution, allowing wavelengths to be identified with clear differences in contribution from associated chemical bonds. Comparisons of fourth-derivative spectra of wholemeal flour from high-N grains with flour from low-N grains identified wavelengths at which there were statistically significant differences between the groups. Their importance was independently confirmed by step-up regression using these wavelengths to generate an equation predicting N content (R^2 = 0.98). Fourth-derivative spectral comparisons also allowed novel biochemical differences to be predicted. Visual assessment of the spectra of all mutants revealed a variable region (1470−1520 nm, corresponding to N−H stretch vibrations) that allowed two extreme sets to be defined. The protein extracted from these two sets differed markedly in hordein content.

KEYWORDS: Derivative spectroscopy; near-infrared reflectance; nitrogen; hordein; resolution enhancement; *Hordeum vulgare* L.

INTRODUCTION

There have been two broad applications of NIR spectroscopic data relevant to barley. The first involved the determination of particular components of barley grains, for instance, β-glucans (*1*, *2*) and dietary fiber (*3*). The second application involved comparisons between grains from normal and mutant barleys. There is a large range of mutations affecting endosperm composition (e.g., contents of particular amino acids, starch quality, and quantity) that have been selected for use in breeding studies. Munck et al. (*4*) compared NIR spectra from normal and *lys3a* lines and, using principal component analysis (PCA), they were able to differentiate between these genotypes. Chemical analysis of these samples revealed differences between normal and *lys3a* lines in amino acid composition, as well as in their contents of β-glucan, starch, fat, and insoluble fiber. Visual inspection of a specific region (2270−2360 nm) of representative NIR spectra showed clear differences between genotypes, and in this region there were spectral assignments for starch (O—H and C—C, 2276 nm), amino acids (N—H and C=O, 2294 nm), cellulose (C—H, 2336 nm), and unsaturated fat (HC=$CHCH_2$, 2347 nm). Thus, the *lys3a* mutant, initially selected as having a high lysine phenotype, also has pleiotropic effects on other grain components which contribute to the physical and chemical phenotype that is revealed by the NIR spectra.

*Corresponding author (telephone +61 2 6246 5251; fax +61 2 6246 5255; e-mail Peter.Chandler@csiro.au).

Derivative spectroscopy is widely utilized in the analysis of chemical and pharmaceutical compounds for the qualitative and quantitative detection of the presence of specific components (*5*, *6*). It is also widely applied to enhance the resolution of spectral data, and it is this latter aspect that is the focus of this paper. In the analysis of NIR data, numerical second-order differentiation is often utilized to remove the linear trend associated with the changing particle size distribution from one sample to the next of the same type of biological material. In reality, when applied, it not only removes the linear trend but also performs a resolution enhancement. Consequently, when similar biological samples (e.g., different wheat varieties grown at different locations, or mutant versus wild type grains) are compared, the second-order differentiation plays an implicit derivative spectroscopy role.

The focus of this paper is the potential for higher order (e.g., fourth) derivative spectroscopy to assist with the analysis and interpretation of NIR spectra of a new collection of barley grain mutants, by virtue of its ability to enhance resolution. We are aware of only a few earlier studies using this methodology to compare similar biological samples, such as cereal grains, other plant material, and foods. Gutierrez (*7*) used second derivatives to quantify α- and β-acids in hops, and Scholz et al. (*8*) detected changes in the size and distribution of polymeric proteins in wheat using second derivatives coupled with PCA. Wesley et al. (*9*) used a combination of second derivatives and curve fitting to predict gliadin and glutenin composition of wheat. Derivative spectroscopy has also been used for resolution enhancement of other

 Published on Web 4/15/2009

spectroscopic techniques in the analysis of food products. Robert et al. (*10*) used second derivatives of Fourier transform infrared spectra to analyze cell wall components of wheat grains, whereas Lüthi-Peng and Puhan (*11*) successfully used fourth-derivative UV spectrophotometry to determine protein (and specifically casein) content in milk.

In this paper, the emphasis is on the utility of derivative spectroscopy in the extraction of molecular insight from very similar NIR spectra (mutants with a common genetic background and grown in the same controlled environment). When the NIR spectra are not that similar (for instance, mutants with a range of genetic backgrounds and grown in different field environments), other methodologies, such as PCA, have been utilized. A detailed and informative discussion of such methodologies can be found in Munck (*12*, *13*). Many of the calculations discussed below were obtained using WinISI II software, which is based on the types of methodologies investigated by Munck.

DERIVATIVE SPECTROSCOPY METHODOLOGY

From a data analysis perspective, modern computer-controlled instruments, such as various spectroscopic devices, record observational data on a very fine (time and/or spatial) grid after averaging a reasonably large number of replications to produce a set of measurements with (very) small errors. This is true for all types of spectroscopic data and NIR in particular, where 25 or more scans have been automatically performed and then averaged by the instrument, in order to give a highly accurate measurement of the spectrum being scanned. This is why NIR spectra of the same sample, after scatter correction, overlay each other to graphical accuracy and better and are often treated as being "exact" curves. This technology is being extended to hyperspectral imaging applications, including its use in the classification of cereal grains and foods (*14*−*16*).

It was the need to work with data defined on coarse grids that spawned much of the earlier methodologies developed for the statistical analysis and numerical differentiation of observational data. Consequently, as a result of the mentioned major improvements in computer-controlled spectrometers, the possibility of obtaining accurate higher derivatives of the data has become a reality.

A brief discussion about the different ways in which differentiation can be performed computationally is given in the Supporting Information.

From the perspective of this paper, as outlined in the Introduction, the key differentiation methodology is "derivative spectroscopy" (*5*). The advantages of derivative spectroscopy include the following: for the first derivative, zero values precisely determine the wavelengths at which peak maxima and minima occur in the function being differentiated and, because the derivative of a constant is zero, it removes the background constant; for the second derivative, it performs a mild resolution enhancement, zero values precisely determine the wavelengths at which peak maxima and minima occur in the first derivative, and it removes the background linear trend. There are similar advantages for the third, fourth, and higher derivatives, each providing increasingly enhanced resolution.

This enhancement is illustrated for the function

$$f(x) = \sin x + \varepsilon \cos \omega x \quad 0 \le x \le 4\pi, \omega > 1 \tag{1}$$

where ε and ω are indicative and illustrative constants for the type of situation that occurs in analytical chemistry and NIR. For this function, the first, second, and fourth derivatives are respectively given by

$$f'(x) = \frac{\mathrm{d}f(x)}{\mathrm{d}x} = \cos x - \varepsilon\omega \sin \omega x \tag{2}$$

$$f''(x) = \frac{\mathrm{d}^2 f(x)}{\mathrm{d}x^2} = -\sin x - \varepsilon\omega^2 \cos \omega x \tag{3}$$

and

$$f''''(x) = \frac{\mathrm{d}^4 f(x)}{\mathrm{d}x^4} = \sin x + \varepsilon\omega^4 \cos \omega x \tag{4}$$

In **Figure 1a**, the functions $f(x)$, $f'(x)$, and $f''(x)$ are plotted for $\varepsilon = 0.0005$ and $\omega = 13$. The structure of these plots is directly reflected in the algebra of $f(x)$ and the corresponding derivatives. The plot of $f(x)$ agrees to graphical accuracy with $\sin x$ because of the very small size of ε (0.0005). Although derivative spectroscopic enhancement is visible in $f''(x)$, there is no such enhancement in $f'(x)$ because the size of $\varepsilon\omega$ is only $0.0005 \times 13 = 0.0065$ as opposed to the larger value of $0.0005 \times 13^2 = 0.0845$ for $\varepsilon\omega^2$.

The functions $f(x)$, $f''(x)$, and $f''''(x)$ are plotted in **Figure 1b**, with ε now given the value of 0.00001 and $\omega = 13$ as above. For this much smaller value of ε, spectroscopic enhancement in $f''(x)$ is now not visible, whereas it is clearly visible in $f''''(x)$ because the size of $\varepsilon\omega^4$ has become 0.00001×13^4. This illustrates why, in some situations, it is necessary to go to the fourth derivative before the utility of derivative spectroscopic enhancement is realized. Clearly, if a signal contains no hidden higher frequency component (i.e., if $\omega = 1$ in eq 1), then no level of differentiation will produce an enhancement of the type illustrated in **Figure 1**.

The order up to which an accurate numerical derivative can be recovered is limited by the fineness of the grid, the accuracy with which the data has been recorded, and the algorithm performing the differentiation. From the perspective of this paper, it will be shown that derivative spectroscopy using the fourth derivative can be used to obtain a much clearer discrimination between subtly different spectra.

Algorithms for Numerical Differentiation. For accurate data on fine grids, suitably designed moving average techniques appear to be the appropriate choice for numerical differentiation (*17*). In WinISI II software, the numerical differentiation is performed using this technique, where the parameters D, G, and $S1$ (with $S2 = 1$) correspond, respectively, to the order of the differentiation, the size of the gap (footprint) of the moving average, and the subset of data points in the gap that are averaged. Although the details about the weights that define the moving average are not given, parsimony implies that an equally weighted central difference formula is the likely choice. The structure of such formulas and their numerical performance are discussed in Anderssen et al. (*17*). All of the numerical differentiation results given below were obtained after experimentation with different values of the parameters D, G, and $S1$.

MATERIALS AND METHODS

Plant Material. The barley (*Hordeum vulgare* L.) mutants were isolated in the naked grain (huskless) variety 'Himalaya' following mutagenesis with sodium azide as previously described (*18*). The treated grains were sown in the field, bulk harvested, and then sown for one further generation. A single well-formed head from each M_2 plant was harvested and threshed. The grains from each head were examined using

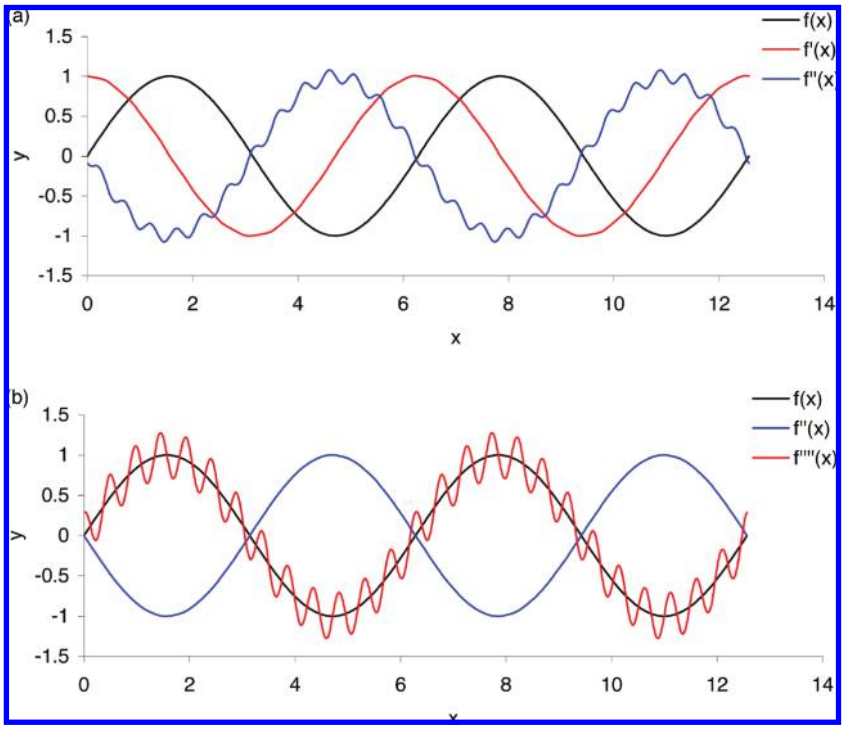

Figure 1. Plots of a function of x demonstrating the resolution-enhancing capacity of numerical differentiation. The function $f(x) = \sin x + \varepsilon \cos \omega x$, $\varepsilon = 0.0005$ and $\omega = 13$, is plotted in (**a**) along with its first and second derivatives. The same function, but with $\varepsilon = 0.00001$, is plotted in (**b**) along with its second and fourth derivatives.

a dissecting microscope, and an assessment was made as to whether (i) all grains were abnormal in phenotype or (ii) there were at least several grains within the head that showed a distinctive abnormal phenotype. In either case, abnormal grains were sown for a further two successive glasshouse generations, and any lines that failed to produce grains that were different from wild type (WT) on a visual screen were discarded. Approximately 150 *grm* (*g*rain *m*orphology) lines were selected from 3500 M_2 heads. The grains of the *grm* mutants differ from the WT in any of numerous criteria, such as grain shape, grain size, degree of grain filling, and surface characteristics (**Figure 2**). M_6 grains were sown in a soil/compost mix in a naturally lit glasshouse at temperatures of 17 °C (day) and 9 °C (night), with a 16 h photoperiod. All lines were grown in the same glasshouse at the same time to minimize the effects of environmental differences. The majority of mutants in the collection differed from wild type only in their grain morphology, although many also had a reduced seed set. A minority of mutants was also affected in one or more vegetative traits, such as vigor, rate of maturity, plant height, and tiller number.

Mature grains harvested from four plants of each line were ground using a cyclone sample mill (Udy Analyzer Co.) to pass through a 1 mm mesh.

NIR Spectroscopy. Wholemeal flour samples of 114 *grm* mutants were analyzed by NIR spectroscopy using an NIRSystems 5000 (Foss NIRSystems, Inc., Silver Spring, MD). Twenty-five scans were automatically performed on each sample by rotation and the readings averaged. Each mutant was reanalyzed once (25 scans) following repacking of the sample, and the resulting spectra were averaged. Readings were taken in the range of 1100−2500 at 2 nm intervals, resulting in 700 data points per sample. The spectra were subjected to a number of mathematical treatments using WinISI II software (Foss NIRSystems, Inc.). Raw data were corrected for scatter resulting from particle size differences using multiplicative scatter correction (MSC), which removes the linear trend associated with particle size by comparing all spectra to the mean "ideal" spectrum and then correcting each spectrum so that they have the same scatter as the ideal (*19*). This results in spectra that have the same overall appearance as the raw data. For derivative spectroscopy, raw data were transformed with WinISI II software using various values for the order of the differentiation, gap width, and smoothing factor.

Determination of Nitrogen (N) and Protein Extraction. Total N content of wholemeal flour from *grm* mutants was determined according to the Dumas dry combustion method (*20*). Total protein was extracted from wholemeal flour using SDS buffer according to the method of Shewry et al. (*21*). SDS-PAGE was carried out on 10% Bis-Tris precast gels (Bio-Rad, Sydney, Australia). Hordeins were extracted from wholemeal flour by shaking 20 mg of wholemeal flour with a stainless steel ball bearing at 30/s for 3.5 min in a 96-well Vibration Mill (Retsch GmbH, Rheinische, Germany) in 0.5 mL of Milli-Q water, followed by centrifugation for 10 min at 5000*g*. The supernatant was removed and discarded. Alcohol-soluble proteins (including the hordeins) were extracted by shaking the pellet (as above) in 0.5 mL of fresh 50% isopropyl alcohol and 1% DTT. Protein concentration was determined according to the Bradford assay (*22*).

Protein Sequencing by Mass Spectroscopy. Plugs of 1 mm^3 from protein bands of interest were taken from Coomassie-stained gels using a Pasteur pipet. Proteins were reduced with DTT, alkylated with acrylamide, and digested with sequencing grade trypsin (Promega, Sydney, Australia). Peptides were extracted with trifluoroacetic acid and cleaned up using C18 Zip Tips (Millipore, Sydney, Australia). Sequencing was performed by MALDI-TOF-TOF mass spectrometry at the John Curtin School of Medical Research, Australian National University, Canberra, Australia.

RESULTS AND DISCUSSION

A forward and a reverse approach to spectral analysis are used in this paper. First, spectra were divided into groups based on known biochemical composition (N content), and the average spectra of the groups were compared using derivative spectroscopy. Second, the spectra were divided into groups based on differences in the visual appearance of individual fourth-derivative spectra, and these differences were used as a basis for investigating the biological variation between the groups.

High and Low N Content. From the full set of NIR spectra for the 114 barley mutants, subsets were selected on the basis of N content. This allows averaged NIR spectra for the high N subset to be compared with the low N subset. The 30 highest lines had

Figure 2. Examples of some of the barley grain phenotypes observed in the mutant collection. Each pair of grains shows a dorsal (left) and ventral (right) view of a different mutant. The WT Himalaya pair is top left.

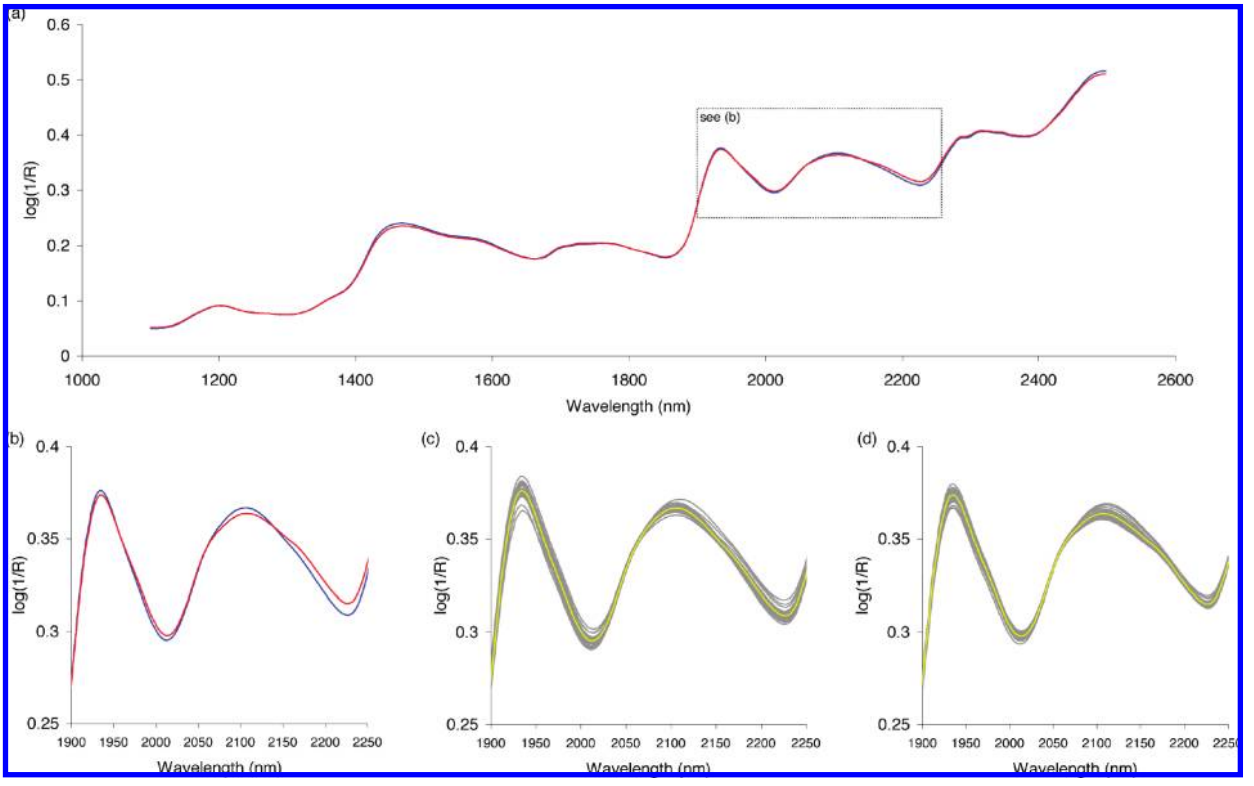

Figure 3. MSC-treated NIR spectra, log(1/*R*), of wholemeal flour from barley mutants varying in N content: (**a**) average full spectra for high (red) and low (blue) N subsets; (**b**) enlargement of boxed area in (**a**) of average spectra for 1900−2250 nm range of high (red) and low (blue) N subsets; (**c**, **d**) spectra for all mutants (gray) and average spectra (yellow) for the high (**c**) and low (**d**) N subsets over the same range.

N contents from 3.12 to 3.71%, and the 30 lowest ranged from 1.75 to 2.7%. The averages of the MSC-treated log 1/*R* NIR spectra are plotted in **Figure 3a**. It can be seen from this figure that, over the wavelength ranges of 2054−2235 and 1400−1500 nm, there appear to be differences between the averages. This first region is highlighted in **Figure 3b**−**d**, which show, for the wavelength range of 1900−2250 nm, the averages of the two subsets of NIR spectra (**Figure 3b**) and the spectra for the high N subset and the low N subset (**Figures 3c**,**d**). Although the difference in the averages in **Figure 3b** is more pronounced than in **Figure 3a** (from the different scaling), it remains difficult to conclude that there is a clear difference in their biological content and structure. This is especially so given the scatter in peaks and troughs within each subset. To confirm the differences between the spectra of the two subsets, derivative spectroscopy, for the reasons outlined in the methodology, was utilized.

Figure 4 shows plots of the second-derivative spectra for the high and low N subsets, with **Figure 4a** showing the averages for the full NIR spectral range and **Figure 4b**−**d** showing the second-derivative counterparts of **Figure 3b**−**d**. Although the second derivative has provided some resolution enhancement and there are differences between the averages over the specific wavelength range of interest (**Figure 4b**), there is still a lot of variation around the peaks and troughs (**Figures 4c**,**d**). The motivation for showing the second derivative is two-fold. First, it provides an alternate method to MSC for the removal of the

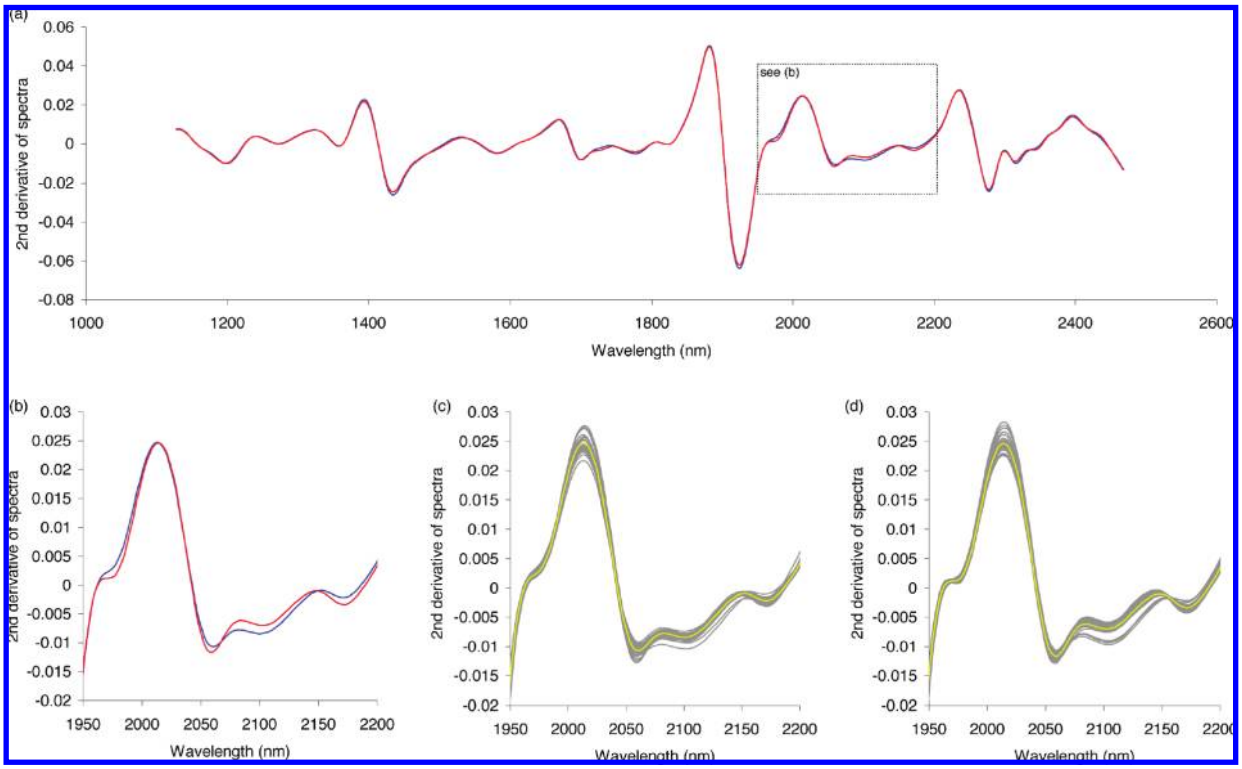

Figure 4. Second derivatives (gap width, 10; smoothing, 10) of NIR spectra of high and low N subsets: (**a**) average full spectra for high (red) and low (blue) N subsets; (**b**) enlargement of boxed area in (**a**) of average spectra for 1950−2200 nm range of high (red) and low (blue) N subsets; (**c**, **d**) spectra for all mutants (gray) and average spectra (yellow) for the high (**c**) and low (**d**) N subsets over the same range.

linear trend associated with light scatter from particle size differences between samples. Second, it plays a derivative spectroscopic role. A comparison of second-derivative spectra within the high and low N subsets (**Figures 4c**,**d**) shows that some spectra have a small offset which appears to be linear and, hence, must correspond to a quadratic offset in the corresponding original spectra, which have been differentiated twice. This conclusion is consistent with the results of Martens et al. (*23*), who concluded that the light scattering effect contains a quadratic function of wavelength in addition to a linear component. As this offset is not observed uniformly across all wavelengths, this may mean that the coefficient of the quadratic term is itself a function of wavelength.

On an inductive basis, having established the derivative spectroscopic enhancement of second differentiation, it follows that the fourth differentiation represents an analogous enhancement of second differentiation and, thereby, an even greater enhancement of the higher frequency components in the original spectrum.

As shown in **Figure 5**, which displays the fourth-derivative counterparts of the second derivatives of **Figure 4**, the degree of variability between the spectra of each subset is reduced compared to the second derivative, resulting from the removal of any cubic and quadratic scattering effects. It can be seen that there is a clear difference in the fourth derivatives of the averaged high and low N subsets. (That this represents a statistically significant difference is demonstrated below.) The importance of this difference is that it highlights the wavelengths at which a significant biochemical difference occurs. The increased number of peaks and troughs in the fourth derivative (**Figure 5a**), compared with the MSC-treated spectra (**Figure 3a**), is indicative of the resolution enhancement that derivative spectroscopy can provide. To understand the biological differences that are being observed, these wavelength positions were related to the known chemical bonds that vibrate in response to an NIR stimulus.

The wavelengths in the spectral region identified in **Figures 3**−**5** are within the N−H combination region of the NIR spectrum, which runs from 1960 to 2294 nm (*24*). Within this region there are three bands associated with proteins, the major nitrogen sink in the barley grain, located at 1980, 2050, and 2180 nm (*25*). These correspond to the amide II band combining with N−H asymmetric and symmetric stretches (1980 and 2050 nm, respectively), and the combination of two amide I bands and an amide III band (2180 nm). Fox et al. (*26*) identified the same region as being strongly correlated to grain protein in barley.

Figure 5a also shows that there are additional differences between high and low N subsets at around 1400−1500 nm. This region corresponds to the first overtone of the same fundamental N−H bond stretching vibrations that contributed to the combination bands observed at 1980 and 2050 nm. The observation of spectral variation at different, harmonically related overtones and combinations of N−H vibrations serves as confirmation that the differences seen in the spectra are due to differences in protein. Slight differences in the original spectra, and in the second derivative, are also visible in this region (**Figures 3a** and **4a**).

Developing an equation to predict nitrogen content confirmed the relationship between the nitrogen content of the mutants and the wavelengths of the peaks and troughs identified by derivative spectroscopy. The equation was developed using the step-up regression facility in the WinISI II software to generate coefficients for user-defined wavelengths using the spectra from all mutants in the collection and their associated N values. The wavelengths selected were 1982, 2018, 2054, 2082, and 2176 nm, corresponding to the peaks and troughs in **Figure 5b**, where the difference between high and low nitrogen was the greatest. This gave an equation that predicted N content

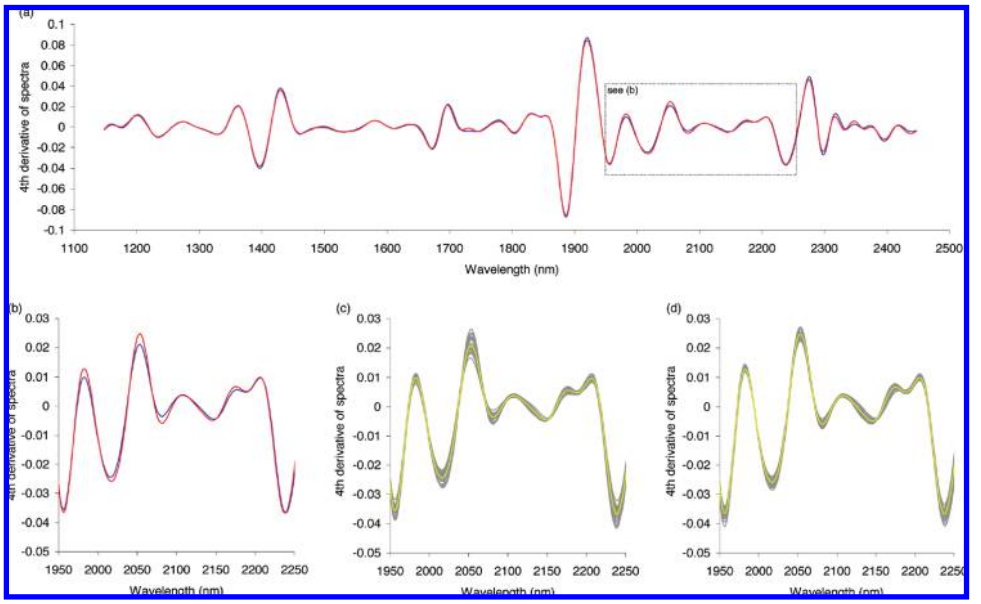

Figure 5. Fourth derivatives (gap width, 10; smoothing, 10) of NIR spectra of high and low N subsets: (**a**) average full spectra for high (red) and low (blue) N subsets; (**b**) enlargement of boxed area in (**a**) of average spectra for 1950−2250 nm range of high (red) and low (blue) N subsets; (**c**, **d**) spectra for all mutants (gray) and average spectra (yellow) for the high (**c**) and low (**d**) N subsets over the same range.

with an R^2 of 0.98 (SEP = 0.069) when tested against 24 independent mutants. The statistical significance of the differences at the selected wavelengths is shown in **Figure 6**, where the fourth derivative of **Figure 5b** is replotted on a larger scale, along with standard error bars. At the wavelengths where clear differences are observed, there is a large gap between the standard error bars of the two curves, indicating a significant difference between the curves. The fourth derivative is identifying statistically significant biochemical differences between the high and low N subsets.

Derivative Spectroscopy and Hordein Content. In the previous example, we demonstrated that derivative spectroscopy enhances spectra to a degree whereby classification into different spectral subsets is possible using previously obtained biochemical data as a starting point. In the following example, we demonstrate that, without biochemical data, derivative spectroscopy can be used to group mutants purely on the basis of differences in their spectra. These groupings can be used as a basis for further investigation.

A visual inspection of the overlaid individual fourth-derivative spectra of all mutants revealed a region of the spectrum that was variable between mutants which was not apparent in the MSC-treated raw spectra (**Figures 7a**,**b**). The wavelength range of 1470−1520 nm contains two spectral features: a broad, asymmetric peak at 1482 nm and a symmetric, Gaussian-like peak at 1498 nm. Mutants could be separated into three broad classes on the basis of these peaks: those with a peak at 1482 nm and a contribution from a peak at 1498 nm (shown in blue in **Figure 7d**); those with a peak at 1498 nm and virtually no contribution from 1482 nm (shown in red in **Figure 7d**); and an intermediate situation with contributions from both wavelengths (the spectrum of the wild-type Himalaya is a member of the intermediate class; shown in black in **Figure 7d**). This visual separation into classes was only possible by using derivative spectroscopy, as demonstrated in **Figure 7**. Panels **a** and **b** of **Figure 7** show the raw data treated with MSC and the fourth derivative, respectively. There are no clear features that could be used to separate the mutants in **Figure 7a**; however, in **Figure 7b** the peaks at 1482 and 1498 nm identified above can be used to separate the mutants into discrete classes. Examples of the two outermost spectral phenotypes (colored in red and blue) are replotted for MSC-treated raw spectra and for the fourth derivative in panels **c** and **d**, respectively, of **Figure 7**, demonstrating the clear separation when using derivative spectroscopy. Differences in the slopes of the curves between the two classes are now visible in the raw spectra; however, this becomes apparent only following the separation in classes using the fourth-derivative spectra. Both peaks are annotated in the WinISI II software as being associated with N−H bonds, with 1482 nm resulting from the first overtone of the N−H stretch vibration in a peptide bond. The peak at 1498 nm is simply annotated as resulting from the first overtone of an N−H stretch vibration, with no further specificity as to the environment of the N−H bond. Between these two peaks, there are contributions from other N−H vibrations from amides, aromatic amines, and urea.

The association of this region with N−H stretching vibrations suggests that the differences observed in the spectra result from differences in the protein content or composition of the mutants. As there was no significant difference between the spectra of the high and low N mutants in this region (**Figure 5a**), it is concluded that the differences observed in **Figure 7d** arose from altered protein composition rather than content. To assess this, proteins were extracted from the wholemeal flour of mutants from the two classes shown in **Figure 7b** (i.e., those with the most pronounced spectral phenotypes in the wavelength range of 1470−1520 nm) and were separated by SDS-PAGE (**Figure 8**). The two classes of mutants had clearly distinguishable protein profiles, with mutants within each of the two classes showing a high degree of similarity to each other. Major differences appear at regions commonly associated with the hordein proteins, with lower levels of B-, C- and D-hordeins [approximately 30−50, 55−75, and 100 kDa, respectively (*27*)] in the mutants with a reduced peak height at 1498 nm than compared with the other group of mutants. The reduction of hordein content was confirmed by extracting and quantifying the alcohol-soluble proteins from these lines. All lines selected with a reduced peak at 1498 nm had hordein contents that were $<50\%$ of WT, with some as low as 25% of WT. These results are consistent with the annotation of the peak as being associated with vibrations of the

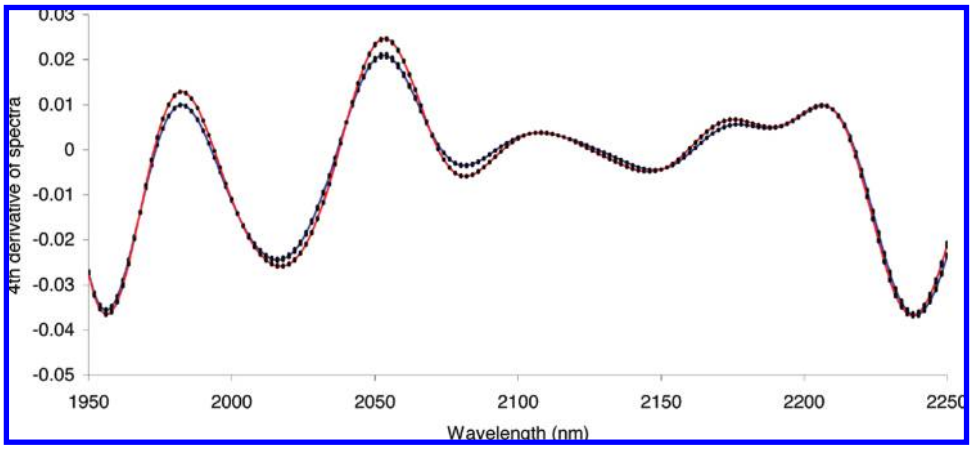

Figure 6. Standard errors (black) of the fourth derivative of NIR spectra of high (red) and low (blue) N subsets for the wavelengths 1950−2250 nm. Error bars are very small, frequently only slightly larger than the thickness of the curves.

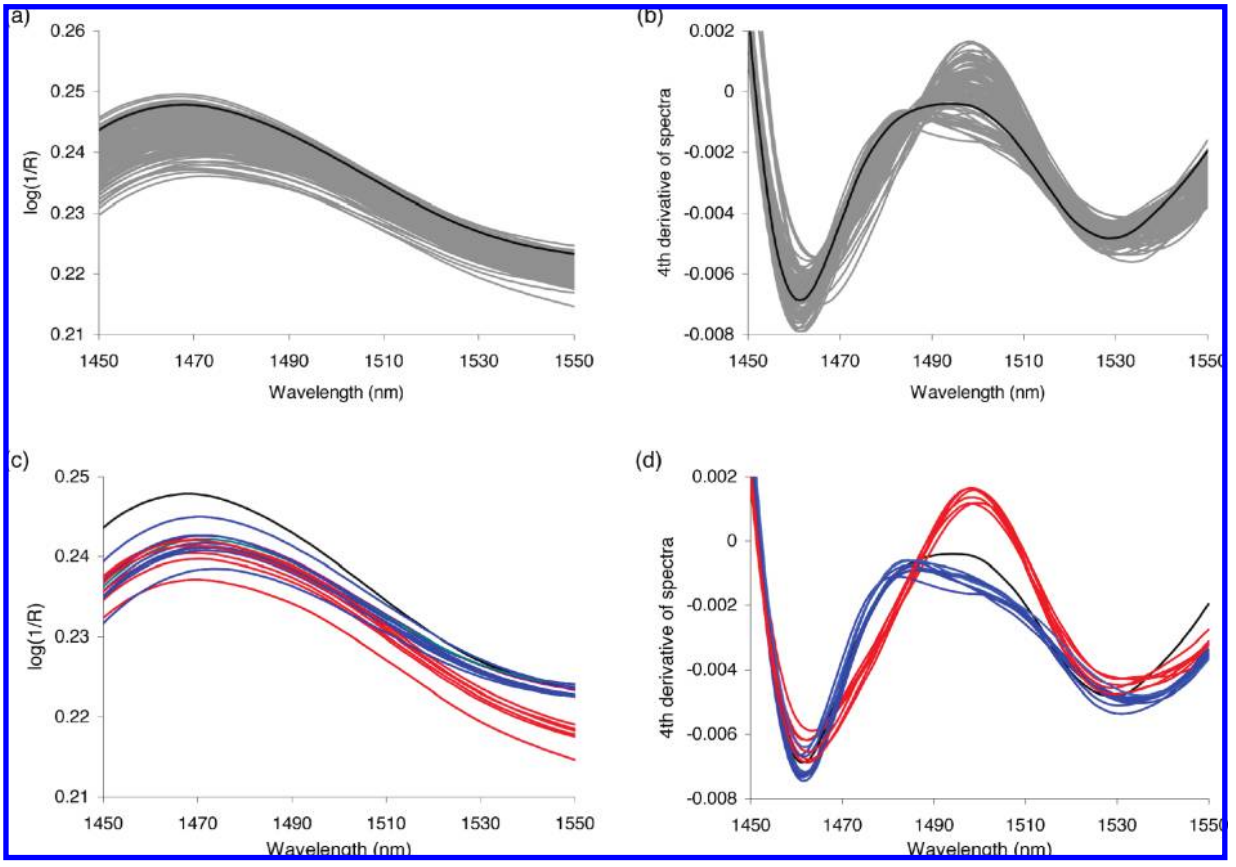

Figure 7. Separation of mutants into classes using derivative spectroscopy: overlaid MSC-treated (**a**, **c**) and fourth-derivative (**b**, **d**) NIR spectra of all mutants (**a**, **b**) and mutants selected on the basis of features at 1482 and 1496 nm in the fourth-derivative spectra (**c**, **d**). Spectra for members of the two classes with the most pronounced spectral features are shown in red and blue, and the spectrum for WT Himalaya is shown in black.

N−H bond, as the hordeins belong to the prolamin group of proteins, which has a greater than average contribution of nitrogen-rich glutamine residues. As the hordeins account for between 35 and 50% of total grain protein, a reduction of these proteins by at least 50% could be expected to have a resultant effect on the NIR spectrum (see Note added in Proof).

The relationship between the hordein content of the mutants and the spectral region identified by derivative spectroscopy was confirmed using stepwise regression in the WinISI II software. In contrast to the step-up regression used previously for N content, here the software assesses all data points in the spectra and selects wavelengths with the highest correlation to a laboratory-measured property. The software selected 1496 nm, only 2 nm from the peak maximum in **Figure 7d**, as the wavelength that correlates most strongly with measured hordein content (F value of 41.7).

Mutants within the low hordein group often had increased amounts of a polypeptide at 73 kDa. Sequencing of trypsin digests of this polypeptide by MALDI-TOF mass spectrometry identified the protein as barley embryo globulin 1 (BEG1), which, despite its name, is also found in aleurone (*28*). It is clear from two-dimensional gels (data not shown) that there are many other non-hordein proteins that are also present at higher contents in this group of mutants, indicating that there is some

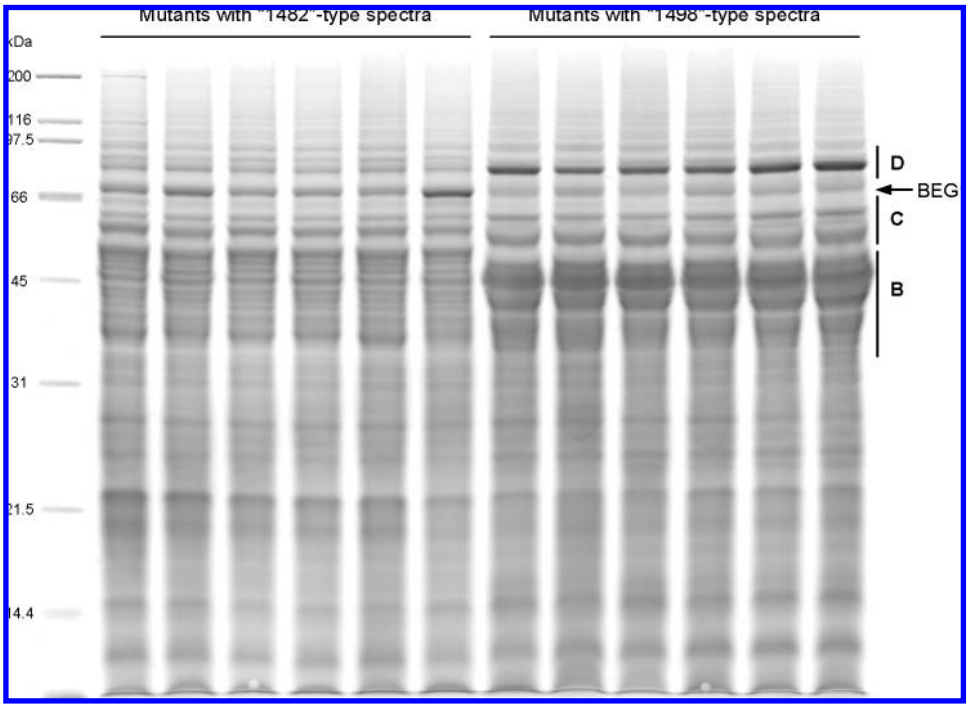

Figure 8. SDS-PAGE of total proteins extracted from wholemeal flour of mutants identified by NIR derivative spectroscopy. The six lanes on the left are from mutants with a stronger contribution from 1482 nm in their fourth-derivative spectra, and the six lanes on the right are from mutants with a stronger contribution from 1498 nm (colored blue and red, respectively, in **Figure 7d**). B, C, and D indicate the approximate locations of the B-, C-, and D-hordeins, respectively; BEG1 indicates the location of barley embryo globulin 1.

compensation in protein composition that is occurring associated with hordein deficiency. Several different classes of transcription factor (bZIP, DOF, and R2R3MYB) have been implicated in the synthesis of hordeins in developing endosperm (*29*), so there may be a range of different genes and physiological mechanisms responsible for the low and high hordein phenotypes. Further characterization of the mutants is required to define the mechanisms involved.

We have used derivative spectroscopy to compare grain mutants with each other and with the wild type. The essence of this methodology is the identification of wavelengths at which nontrivial changes have occurred, from which known molecular bonds can be identified. The resulting information, along with independent biological information, forms the starting point for drawing conclusions about grain properties of the mutants. In particular, use of the fourth derivative has allowed barley mutants to be spectrally distinguished on the basis of either grain nitrogen content or high or low hordein content. These differences will form the basis for NIR diagnostics to provide an additional criterion for phenotypically grouping mutants and to ultimately predict the grain composition of individual mutants with unknown biochemical makeup.

ABBREVIATIONS USED

DTT, dithiothreitol; *grm*, grain morphology; MALDI-TOF, matrix-assisted laser desorption/ionization time-of-flight; MSC, multiplicative scatter correction; N, nitrogen; NIR, near-infrared reflectance; PCA, principal component analysis; R, reflectance; SDS-PAGE, sodium dodecyl sulfate−polyacrylamide gel electrophoresis; SEP, standard error of prediction; UV, ultraviolet; WT, wild type.

ACKNOWLEDGMENT

We gratefully acknowledge Carol Harding, Russell Heywood, and Peter Milburn for technical assistance, and Alec Zwart for statistical advice. R.S.A. acknowledges his debt to Markus Hegland, who first drew his attention to the concept of "derivative spectroscopy", and support from the Radon Institute of Computation and Applied Mathematics, Austria.

Supporting Information Available: Differentiation of data. This material is available free of charge via the Internet at http://pubs.acs.org.

NOTE ADDED IN PROOF

Confirmation that the spectral differences observed at 1498 nm were due to altered hordein content was provided by spiking the flour of a low-hordein mutant with purified hordein. This increased the intensity of the peak at 1498 nm in the fourth derivative spectrum so that it resembled the spectra of the high-hordein mutants (data not shown).

LITERATURE CITED

(1) Czuchajowska, Z.; Szczodrak, J.; Pomeranz, Y. Characterization and estimation of barley polysaccharides by near-infrared spectroscopy. I. Barleys, starches, and β-D-glucans. *Cereal Chem.* **1992**, *69*, 413–418.

(2) Henry, R. J. Near-infrared reflectance analysis of carbohydrates and its application to the determination of (1→3), (1→4)-β-D-glucan in barley. *Carbohydr. Res.* **1985**, *141*, 13–19.

(3) Kays, S. E.; Shimizu, N.; Barton, F. E.II; Ohtsubo, K. Near-infrared transmission and reflectance spectroscopy for the determination of dietary fiber in barley cultivars. *Crop Sci.* **2005**, *45*, 2307–2311.

(4) Munck, L.; Pram Nielsen, J.; Møller, B.; Jacobsen, S.; Søndergaard, I.; Engelsen, S. B.; Nørgaard, L.; Bro, R. Exploring the phenotypic expression of a regulatory proteome-altering gene by spectroscopy and chemometrics. *Anal. Chim. Acta* **2001**, *446*, 171–186.

(5) Karpinska, J. Derivative spectrophotometry—recent applications and directions of development. *Talanta* **2004**, *64*, 801–822.

(6) Lee, A. R. Derivative spectroscopy and its applications in drug analysis. *Chin. Pharm. J.* **1992**, *44*, 87–96.

(7) Gutierrez, M. C. Derivative spectroscopy applied to the determination of α- and β-acids in hops. *J. Inst. Brew*. **1992**, *98*, 277–281.
(8) Scholz, E.; Prieto-Linde, M. L.; Gergely, S.; Salgó, A.; Johansson, E. Possibilities of using near infrared reflectance/transmittance spectroscopy for determination of polymeric protein in wheat. *J. Sci. Food Agric*. **2007**, *87*, 1523–1532.
(9) Wesley, I. J.; Uthayakumaran, S.; Anderssen, R. S.; Cornish, G. B.; Bekes, F.; Osbourne, B. G.; Skerritt, J. H. A curve-fitting approach to the near infrared reflectance measurement of wheat flour proteins which influence dough quality. *J. Near Infrared Spectrosc*. **1999**, *7*, 229–240.
(10) Robert, P.; Marquis, M.; Barron, C.; Guillon, F.; Saulnier, L. FT-IR investigation of cell wall polysaccharides from cereal grains. Arabinoxylan infrared assignment. *J. Agric. Food Chem*. **2005**, *53*, 7014–7018.
(11) Lüthi-Peng, Q.-Q.; Puhan, Z. Determination of protein and casein in milk by fourth derivative UV spectrophotometry. *Anal. Chim. Acta* **1999**, *393*, 227–234.
(12) Munck, L. Conceptual validation of self-organisation studied by spectroscopy in an endosperm gene model as a data-driven logistic strategy in chemometrics. *Chemom. Intell. Lab*. **2006**, *84*, 26–32.
(13) Munck, L. A new holistic exploratory approach to Systems Biology by near infrared spectroscopy evaluated by chemometrics and data inspection. *J. Chemom*. **2007**, *21*, 406–426.
(14) Choudhary, R.; Mahesh, S.; Paliwal, J.; Jayas, D. S. Identification of wheat classes using wavelet features from near infrared hyperspectral images of bulk samples. *Biosyst. Eng*. **2009**, *102*, 115–127.
(15) Mahesh, S.; Manickavasagan, A.; Jayas, D. S.; Paliwal, J.; White, N. D. G. Feasibility of near-infrared hyperspectral imaging to differentiate Canadian wheat classes. *Biosyst. Eng*. **2008**, *101*, 50–57.
(16) Nansen, C.; Kolomiets, M.; Gao, X. Considerations regarding the use of hyperspectral imaging data in classifications of food products, exemplified by analysis of maize kernels. *J. Agric. Food Chem*. **2008**, *56*, 2933–2938.
(17) Anderssen, B.; de Hoog, F.; Hegland, M. A stable finite difference ansatz for higher order differentiation of non-exact data. *Bull. Aust. Math. Soc*. **1998**, *58*, 223–232.
(18) Zwar, J. A.; Chandler, P. M. α-Amylase production and leaf protein synthesis in a gibberellin-responsive dwarf mutant of "Himalaya" barley (*Hordeum vulgare* L.). *Planta* **1995**, *197*, 39–48.
(19) Isaksson, T.; Næs, T. The effect of multiplicative scatter correction (MSC) and linearity improvement in NIR spectroscopy. *Appl. Spectrosc*. **1988**, *42*, 1273–1284.
(20) AACC. Method 46-30. In *Approved Methods of the AACC*, 10th ed.; American Association of Cereal Chemists: St. Paul, MN, 2000
(21) Shewry, P. R.; Tatham, A. S.; Fido, R. J. Separation of plant proteins by electrophoresis. In *Methods in Molecular Biology: Plant Gene Transfer and Expression Protocols*; Jones, H., Ed.; Humana Press: Totowa, NJ, 1995; Vol. *49*, pp 399−422.
(22) Bradford, M. M. Rapid and sensitive method for quantitation of microgram quantities of protein utilizing principle of protein−dye binding. *Anal. Biochem*. **1976**, *72*, 248–254.
(23) Martens, H.; Nielsen, J. P.; Engelsen, S. B. Light scattering and light absorbance separated by extended multiplicative signal correction. Application to near-infrared transmission analysis of powder mixtures. *Anal. Chem*. **2003**, *75*, 394–404.
(24) Osborne, B. G.; Fearn, T.; Hindle P. H. *Practical NIR Spectroscopy with Applications in Food and Beverage Analysis*, 2nd ed.; Longman Scientific and Technical: New York, 1993; pp 227.
(25) Law, D. P.; Tkachuk, R. Near infrared diffuse reflectance spectra of wheat and wheat components. . *Cereal Chem*. **1977**, *54*, 256–265.
(26) Fox, G. P.; Onley-Watson, K.; Osman, A. Multiple linear regression calibrations for barley and malt protein based on the spectra of hordein. *J. Inst. Brew*. **2002**, *108*, 155–159.
(27) Shewry, P. R.; Kreis, M.; Parmar, S.; Lew, E. J.-L.; Kasarda, D. D. Identification of γ-type hordeins in barley. *FEBS Lett*. **1985**, *190*, 61–64.
(28) Heck, G. R.; Chamberlain, A. K.; Ho, T.-H. D. Barley embryo globulin 1 gene, *Beg1*: characterization of cDNA, chromosome mapping and regulation of expression. *Mol. Gen. Genet*. **1993**, *239*, 209–218.
(29) Diaz, I.; Martinez, M.; Isabel-LaMoneda, I; Rubio-Somoza, I.; Carbonero, P. The DOF protein, SAD, interacts with GAMYB in plant nuclei and activates transcription of endosperm-specific genes during barley seed development. *Plant J*. **2005**, *42*, 652–662.

Received for Review January 14, 2009. Revised manuscript received February 24, 2009. Accepted March 17, 2009.