Noninvasive, On-Line Monitoring of the Biotransformation by Yeast of Glucose to Ethanol Using Dispersive Raman Spectroscopy and Chemometrics

ADRIAN D. SHAW, NAHEED KADERBHAI, ALUN JONES, ANDREW M. WOODWARD, ROYSTON GOODACRE, JEM J. ROWLAND, and DOUGLAS B. KELL*

Institute of Biological Sciences, Cledwyn Building, (A.D.S., N.K., A.J., A.M.W., R.G., D.B.K.) and Department of Computer Science (J.J.R.), University of Wales, Aberystwyth SY23 3DB, U.K.

We describe the first application of dispersive Raman spectroscopy using a diode laser exciting at 780 nm and a charge-coupled device (CCD) detector to the noninvasive, on-line determination of the biotransformation by yeast of glucose to ethanol. Software was developed which automatically removed the effects of cosmic rays and other noise, normalized the spectra to an invariant peak, then removed the "baseline" arising from interference by fluorescent impurities, to obtain the "true" Raman spectra. Variable selection was automatically performed on the parameters of relevant Raman peaks (height, width, position of top and center, area and skewness), and a small subset used as the input to cross-validated models based on partial least-squares (PLS) regression. The multivariate calibration models so formed were sufficiently robust to be able to predict the concentration of glucose and ethanol *in a completely different fermentation* with a precision better than 5%. Dispersive Raman spectroscopy, when coupled with the appropriate chemometrics, is a very useful approach to the noninvasive, on-line determination of the progress of microbial fermentations.

Index Headings: **Raman spectroscopy; On-line monitoring; Chemometric methods; Bioinformatics; Fermentation; Variable selection.**

INTRODUCTION

There is a continuing need for on-line methods for the characterization of bioprocesses.[1–3] The ideal method[4] would be rapid, noninvasive, reagentless, precise, and cheap; although to date, with the possible exception of near-infrared (NIR) spectroscopy (see below), almost no such method has been found.

During the last few years there has been a renaissance in Raman instrumentation suitable for the analysis of biological systems, initially with the development of Fourier transform (FT)-Raman instruments in which the wavelength of the exciting laser is in the near-infrared (usually a Nd:YAG at 1064 nm) rather than in the visible region, an arrangement which therefore avoids the background fluorescence typical of biological samples illuminated in the visible.[5–23] In addition, and at least as importantly, exceptional Rayleigh light rejection has come from the development of holographic notch filters,[24–27] and a recent innovation is the use of Hadamard transform-based spectrometers.[28,29]

Although the FT approach to both infrared and Raman spectroscopy possesses well-known advantages of optical throughput,[8,30] there are still problems for FT-Raman with many aqueous biological samples since water may absorb both the exciting laser radiation at 1064 nm and the Raman scattered light. In addition, it is often necessary to co-add many hundreds of spectra to produce high-quality data from biological systems, and acquisition times are frequently 15–60 min. More recently, therefore, it has been recognized that charge-coupled device (CCD) array detectors are ideal elements for use in *dispersive* (non-FT) Raman spectroscopy. However, they normally have very low quantum efficiency at 1064 nm photons. Thus holographic notch filters and CCD array detectors have been combined with a dispersive instrument, using diode laser excitation at 780 nm (a wavelength which again suppresses fluorescence from most samples but which penetrates water well). The cooled CCD is a multichannel device which has exceptional sensitivity and very low intrinsic noise (dark current), so that the signal-to-noise ratio is improved by at least two orders of magnitude (compared with an uncooled CCD), and data acquisition is correspondingly fast.[13] These and other[31,32] major technical advances now make Raman a very promising tool for the rapid, noninvasive, and multiparameter analysis of aqueous biological systems, and its use for the estimation of metabolite concentrations in ocular tissue has been reported.[33,34]

In 1987, Shope et al.[35] used attenuated total reflectance (ATR) Raman spectroscopy for the on-line monitoring of the fermentation by yeast of sucrose to ethanol, using the argon-ion laser line at 514.5 nm. Gomy et al.[36,37] monitored their alcoholic fermentation using the same laser with a fiber-optic probe attached to a Raman spectrometer but analyzed the ethanol levels only at higher wavenumbers (2600–3800 cm^{-1}). This was because the Raman monitoring of these processes using 514.5 nm excitation gave significant fluorescence in the lower wavenumber region, as can be observed in the spectra shown in these papers. NIR spectroscopy continues to be applied to on-line fermentation and biotransformation monitoring, for example, of ethanol and biomass in rich medium in a yeast fermentation,[38,39] lactic acid production,[40,41] bioconversion of glycerol to 1,3-dihydroxyacetone,[42] and nutrient and product concentrations in commercial antibiotic fermentations.[43,44] Hall, Macaloney, and colleagues[45,46] reported NIR spectroscopic monitoring of industrial fed-batch *Escherichia coli* fermentation of varying levels of acetate, ammonium, glycerol, and biomass which they had previously studied in shake flasks,[47] while Yano et al.[48] used NIR spectroscopy to determine the concentra-

Received 9 February 1999; accepted 20 May 1999.
* Author to whom correspondence should be sent.

0003-7028/99/5311-1419$2.00/0

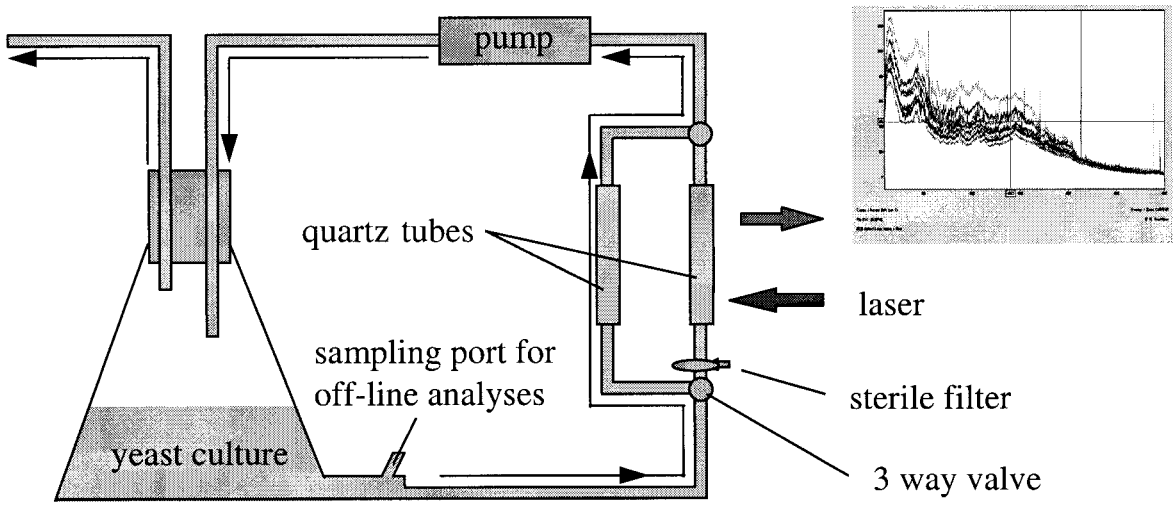

FIG. 1. Schematic of the experimental setup for monitoring yeast ethanol biotransformation using a dispersive Raman spectrometer.

tions of ethanol and acetate in rice vinegar fermentations with good precision. Due to spectral interference from the biomass, Fayolle et al.[49,50] monitored ethanol fermentation using FT-mid-IR spectroscopy and collected spectra of filtered and nonfiltered culture media.

Although fluorescence has been a major hindrance for the use of Raman spectroscopy in biology, Shope et al.[35] clearly showed that the narrow Raman peaks were distinct from the broad features of fluorescence and proposed the use of full widths at half-height of the peaks for chemical quantitation from Raman spectra. Additionally, data analysis using multiple linear regression (MLR)[45] for acetate and glycerol and the more sophisticated partial least-squares (PLS) was necessary to model ammonium and biomass for NIR spectral data to allow simultaneous immediate analysis for the *E. coli* fermentation process. Similarly, Shope et al.[35] used a least-squares fit to analyze the Raman spectra for quantitation of the production of ethanol during the yeast fermentation process. Finally, Spiegelman et al.[51] have recently shown that the amount of glucose in aqueous solution can be measured by using Raman spectroscopy.

Here we describe the on-line monitoring of the glucose fermentation by yeast to ethanol using a dispersive Raman spectrometer with a 780 nm diode laser, after on-line filtration of the biomass. The NIR diode laser at 780 nm suppresses the bulk of the fluorescence from most biological samples and has a significantly greater sensitivity than is the case with excitation at 1064 nm as a consequence of the inverse fourth power dependence of Raman scattering efficiency on wavelength. Therefore, the high sensitivity and increased efficiency mean that data acquisition times and low laser powers may be used. Although chemometrics has been applied for analysis of Raman spectra,[52–57] recently Vickers et al.[58] showed that quantitative use of the on-line and at-line spectra obtained from a compact dispersive Raman spectrometer require different data treatments from those described for FT-Raman instruments, and Spiegelman et al.[51] have exploited such using Raman spectroscopy. Therefore, our aim was to develop and apply specific chemometric methods for the rapid analysis of the data obtained by using a compact dispersive Raman spectrometer.

MATERIALS AND METHODS

On-line Setup of the Yeast Biotransformation. All the chemicals were obtained from Sigma and were of analytical grade, unless stated otherwise. Two types of biotransformations were performed as described below.

Allinson's baking yeast (Westmill Foods Ltd., Berkshire, U.K.) was added as a 1% inoculum to 130 mL of 500 mM D-glucose in distilled water and monitored on-line by using the setup shown in Fig. 1. The culture was stirred and recycled continuously unless stated otherwise (with the use of a modified 100 mL conical flask with a glass side arm fixed to the bottom) via a 90 cm long silicone rubber tubing fixed to a Pharmacia Peristaltic P-1 pump (set to pump the culture at a flow rate of 1 mL min^{-1}) and connected back to the flask with a Sterican 1.10 × 50 mm/19 g × 2 in. needle (Sterican, B. Braun, Melsungen, AG) inserted into a 41 size Subaseal stopper. The silicone tubing was fitted with two 3 cm long 3 mm bore quartz tubes halfway along its length (Comer Instruments, 70, Cambridge, U.K.) and connected to two by-passes with two 3-way taps, with one by-pass for normal culture cycling and the other fitted with a sterile 0.2 μm, 10 mm filter for removal of cells prior to spectral analysis. This second tube was fixed onto the Macropoint stage of the microscope of the Raman spectroscope (see below). A Pasteur pipette was also inserted into the flask stopper for the escape of CO_2 produced as the second product of the glucose biotransformation. At the end of the biotransformation, the nonfiltered sample in the second quartz tube was analyzed spectrally as for the filtered sample.

For biotransformation 1, samples were withdrawn at 1 h intervals on day 1 and 0.5 h intervals on day 2 and day 3 with stirring and circulation stopped overnight and the culture maintained at ambient temperature. Biotransformation 2 was carried with sampling at 1 h intervals on day 1 and subsequent sampling at 0.5 h intervals for the following days, but the culture was maintained at 4 °C overnight in order to slow down the rate of the transformation and obtain a larger number of data points.

Microbial and Biochemical Analysis. Samples (0.5 mL) were withdrawn for biochemical analysis at times

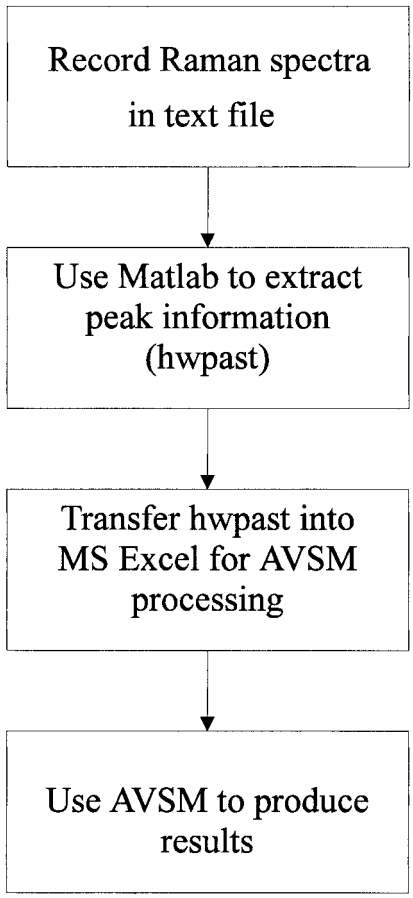

FIG. 2. The steps performed in processing the data.

similar to those for spectral analysis and were stored frozen at −20 °C. They were thawed on ice before microcentrifugation, and after appropriate dilutions of each of the samples the supernatants were used for ethanol and glucose estimation with the Sigma diagnostic kits, 332-B (alcohol dehydrogenase assay) and 510-DA (glucose oxidase and peroxidase assay), and were monitored at 340 and 450 nm, respectively.

Dispersive Raman Spectroscopy. Spectra were collected with a Renishaw Model 2000 dispersive Raman spectrometer[31,32] with a low-power (15 mW) near-infrared diode laser emitting at 780 nm with incident power on the sample of 2–3 mW in a 2 μm diameter spot. The instrument grating was calibrated by using neon lines[59] and was routinely checked with a silicon wafer focused under the 50× microscope objective and collected as a static spectrum centered at 520 cm^{-1} for 10 s.

Data for the on-line monitoring of the filtrate (with the setup described above), in a 3 mm quartz tube clamped to an adapted Macropoint stage, were obtained by focusing with a 10× microscope objective fixed to a Macropoint assembly fitted to the standard objective aperture. Spectra were collected as 3 scans of 60 s each.

Data Analysis. The GRAMS WiRE software package (Renishaw) running under Windows 95 was used for instrument control and data capture. Spectra were collected over 100–3000 wavenumbers with 3382 data points. The spectral resolution of the Renishaw system is around 1.5 cm^{-1}. The data were displayed as intensity of Raman photon counts against Raman shift in wavenumbers (cm^{-1}).

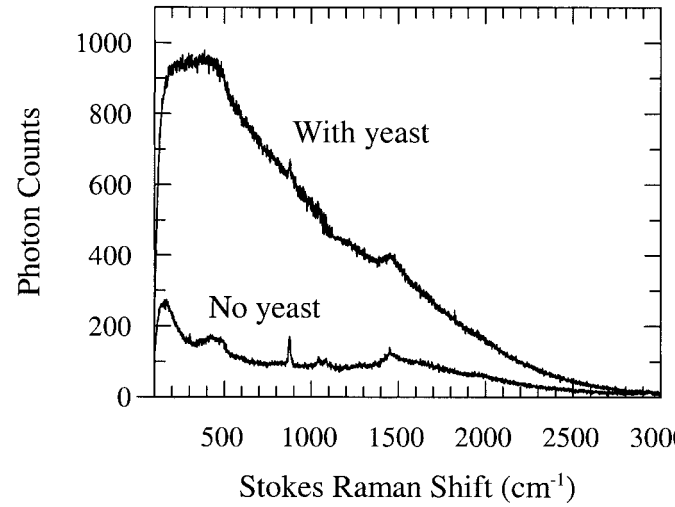

FIG. 3. Spectra of glucose biotransformation 2 after 60 h with and without yeast cells. Upper trace with cells, lower trace without cells.

The spectral data were extracted into text files and imported into Matlab Version 5.0. With the use of Matlab routines written for the purpose, the spectra were normalized, and the baselines and any background fluorescence removed. As described in detail in the Results section, the height, width, area, position of top, position of center, and offset of top from center of 17 identifiable peaks likely to be important for ethanol and glucose were calculated (these dimensions are referred to as the *hwpast*, for height, width, position, area, skewness, and top). The resulting 102 variables (17 × 6) were then imported into Microsoft Excel, where they were processed by using a suite of Excel macros (known as AVSM), which have been produced in-house[60,61] for carrying out variable selection, creating statistical models, and producing graphs of the results in a largely automated fashion. With the use of statistical programs written in C++ by Dr. A. Jones, these macros process the data so as to select the best variables (using characteristicity and Fisher or ANOVA for classification problems, Product Moment Correlation to the Y data for quantification problems such as this), carry out statistical analysis on the data [predictions using PLS, neural nets, principal component regression (PCR), MLR, and principal component analysis (PCA)], establish the best variables for a given training and/or query set, and apply the model formed to an independent validation set. The steps carried out in processing the data are shown in Fig. 2.

RESULTS AND DISCUSSION

The Raman spectral analysis of this biotransformation was complicated by interference from spectral features due to cellular components (Fig. 3). In addition, the Raman scattered light is detected at 180° to the incident excitation (back-scatter) but the absolute extent of photons that are potentially able to be back-scattered even by Rayleigh scattering is a complex function of the total cell mass and other factors in heterogeneous media,[62–64] which can vary significantly during the biotransforma-

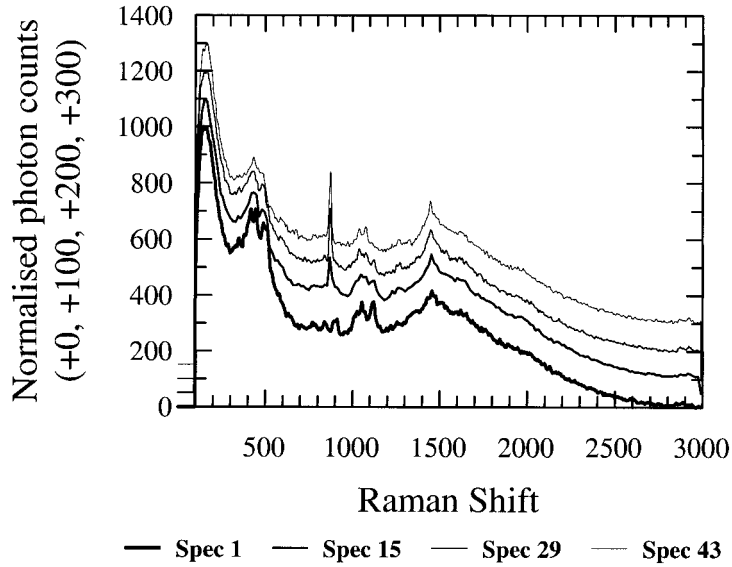

FIG. 4. Comparison of smoothed, normalized spectra from biotransformation 2, taken at equally spaced intervals through the experiment, showing the change in the spectrum over time. Spectra are artificially displaced by 100 photon counts for clarity.

tion. Therefore, a by-pass with a filter to remove the cells from the interrogation zone was fitted into the setup (Fig. 1), which also caused a substantial decrease in the baseline, as seen in the lower trace in Fig. 3 as the yeast cells showed negligible fluorescence with excitation at 780 nm with the use of a spectrofluorimeter (data not shown).

The spectra of the filtrates at varying time intervals (Fig. 4) clearly show the increase in the sharp peak at 877 cm^{-1} which is characteristic of the ethanol symmetric C–C–O stretch[35] as the biotransformation progresses. The utilization of glucose was monitored across the spectrum with a number of peaks which are specifically assigned to the glucose molecule.[19]

Figure 5 shows the ethanol production as glucose depleted during the biotransformations. The apparent stoichiometric efficiency of biotransformation 1 was 83.2% and of biotransformation 2 was 76.5%.

For the present work, two independent biotransformations were performed (see above). During the first biotransformation 33 spectra and samples were obtained and for biotransformation 2 there were 45 spectra and samples. In the second, the last five points were removed before chemometric analysis as the biotransformation had reached completion. The "true" concentrations for the glucose and the ethanol used in the PLS modeling were obtained by using the enzymatic assays described above.

The procedure for extracting the data from the raw spectra is next described, with the example of a spectrum (Fig. 6A) taken from the midpoint of biotransformation 2. For the sake of clarity, the abscissa in these plots shows the actual wavenumbers; the analysis is, however, carried out with the use of data points. One wavenumber (cm^{-1}) is equivalent to 1.17 points, and the first point represents a wavenumber of 100 cm^{-1}.

The three obvious "spikes" in the spectrum at ~250, 600, and 2700 cm^{-1} are due to the interaction of cosmic rays with the sensitive CCD detector. These cosmic rays must first be removed. This is simply a process of finding large peaks in the spectra which have a width less than or equal to a predetermined maximum (seven data points is usually sufficient). A straight line is drawn between the start and end points (Fig. 6B).

Following the removal of cosmic rays, the spectra are smoothed. Of the most common available methods,[65,66] we have here used a Fourier transform method and a moving average method. With the Fourier transform method (in which spectra are Fourier transformed, high-frequency "noise" bins removed, and the residue inverse Fourier transformed), it is often found that some "ringing" occurs, causing extraneous peaks to be added to the spectrum. Through experimentation, we have found a moving average generally to be better for denoising these spectra. The moving average is carried out on the first derivative of the spectrum, in an attempt to reduce the flattening off of peaks, which is otherwise the main problem with a moving average. A width of 15 has been found generally to be best. This is applied by weighting the points used to calculate the average by their distance from the center point, using a sine function. Therefore, points nearest the center point are given a greater weighting. The weighting therefore for a moving average of width 15 is as given in Table I. The resulting spectrum is shown in Fig. 6C.

To ensure that all the spectra are of the same magnitude, and to correct for any variations of excitation intensity for example, it is necessary to normalize the spectra. This is done by setting the lowest point to 0 and the highest to 1000. It is possible to use the highest point as a standard in this case. The result of the normalization is shown in Fig. 6D.

The next step is to remove the baseline. Because of the nature of the Raman baseline, it is possible to "scoop" it out, and indeed this has proved to be the most effective baseline removal method of all those tried (which included using a Fourier transform method). Note that this method is unable reliably to pick the peak at 1580 data points (1450 wavenumbers), which is due to the C–O–H bond vibration of ethanol;[67] an alternative method would have to be used if it was desired to include this peak in the analysis.

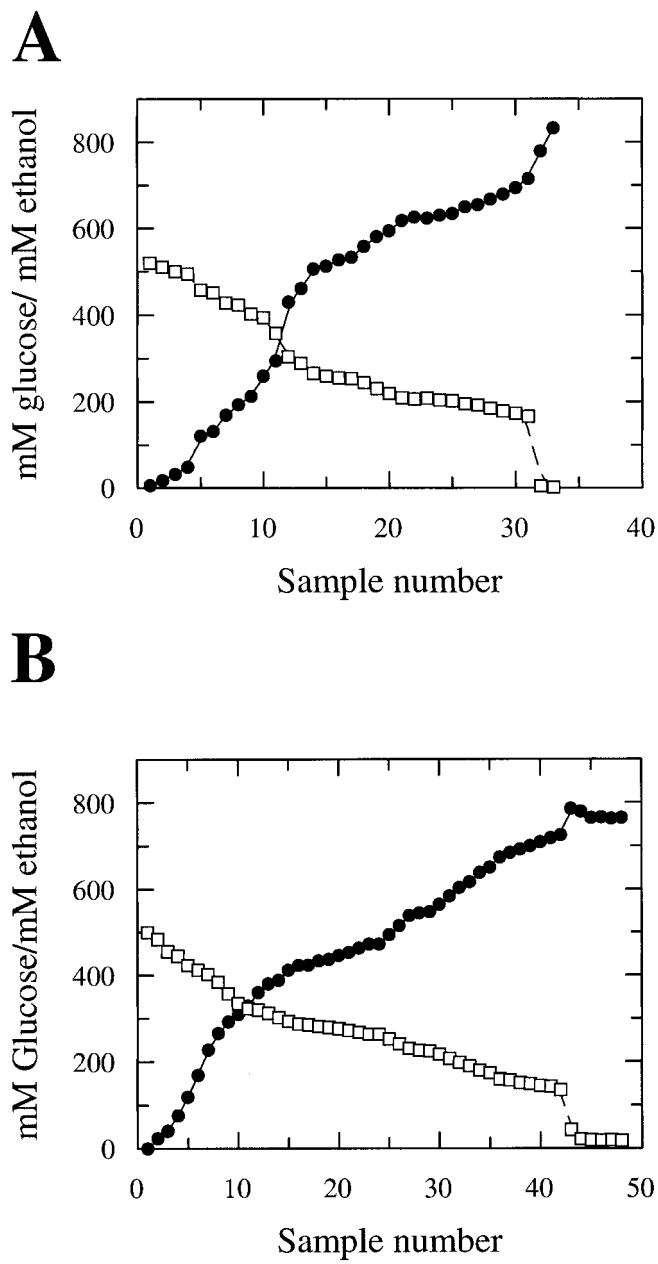

FIG. 5. Time dependence of the biotransformation of glucose (□) to ethanol (●) by yeast. **(A)** Biotransformation 1; **(B)** biotransformation 2. Measurements were performed enzymatically.

The scoop method, devised herein for this analysis, works by looking ahead a predetermined number of points and finding the lowest point within that range relative to the distance from the point. It is therefore analogous to placing a ruler vertically so that it pivots on the point, and turning it anticlockwise until a point further along the ruler (to the right) first encounters a point on the spectrum within the specified range. A straight line is then drawn between the pivotal point and the point of contact. This then forms the baseline between these points. The number of points to look ahead that has been used here is 275 for points 700 to 1400 (wavenumber 697 to 1296), and 80 elsewhere. The first 100 points (wavenumbers 100 to 185) are ignored (equated to the baseline), since they are not important, and the scoop method would not work due to the negative curve in this region.

We note that, although a curve fit could be used instead of a straight line, it is unnecessarily complicated and introduces more problems—principally how to determine the shape of the curve in such a way as to be consistent for all spectra.

The spectrum of Fig. 6D is shown here in Fig. 6E as a dotted line, along with the baseline. The residual, "true" Raman spectrum is therefore simply a subtraction of the baseline from the smoothed, normalized spectrum, as shown in Fig. 6F.

Now we are in a position to extract the data from the "true", smoothed Raman spectrum which has been corrected for baseline and fluorescence. From the litera-

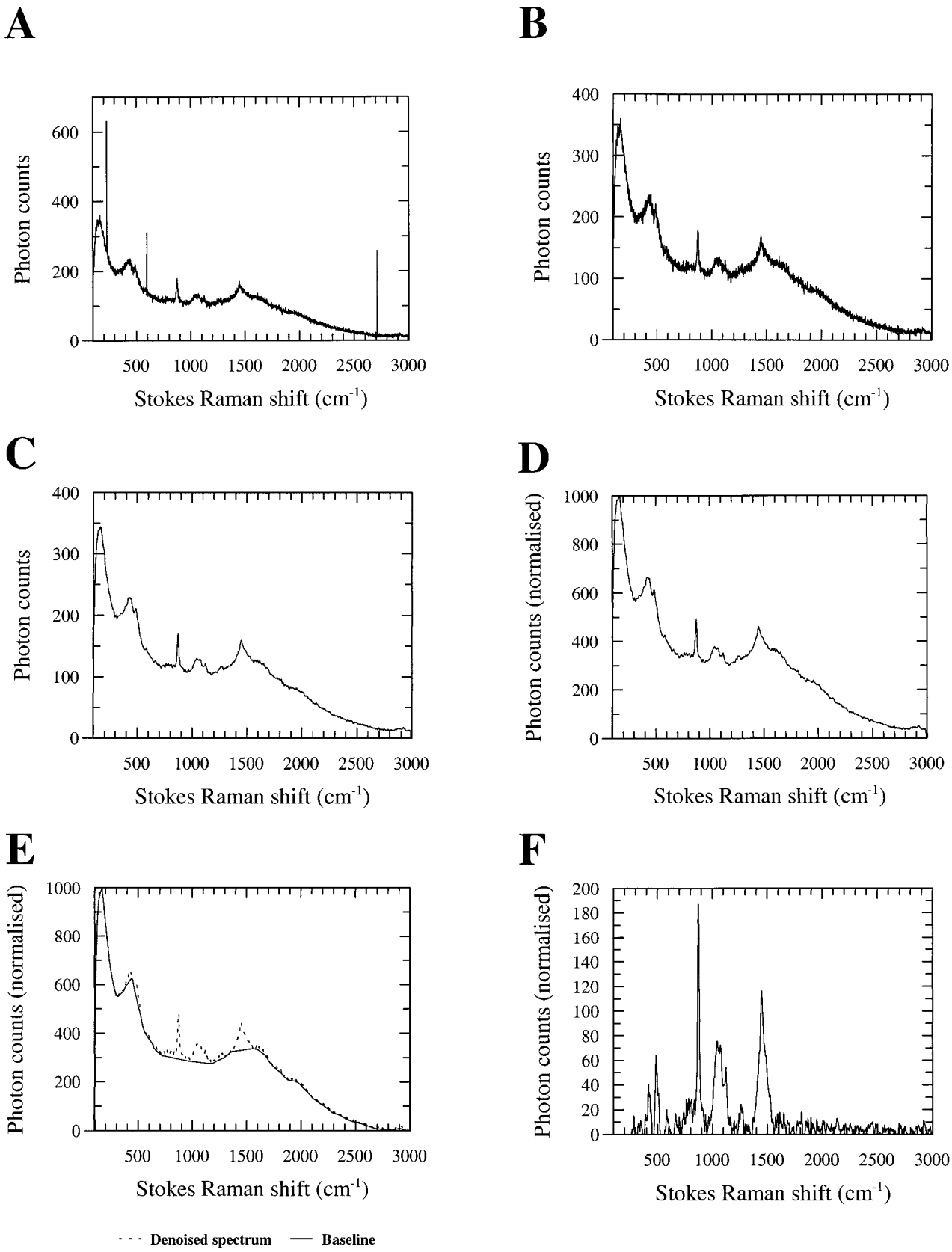

FIG. 6. Processing of the Raman data, illustrated using the spectrum of fermentation 2 at approximately half way, containing 494 mM ethanol and 252 mM glucose. **(A)** Raw data; **(B)** raw data with cosmic rays removed; **(C)** reconstructed spectrum of **B** after denoising using a moving average filter of width 15 on the first derivative; **(D)** spectrum of **C** normalized by setting the quartz peak (highest point) to 1000 and scaling the remainder of the spectrum accordingly; **(E)** finding the baseline in the spectrum, using the scoop method; **(F)** final Raman spectrum with the baseline removed.

TABLE I. Weighting for a moving average with width of 15 in the denoising process.

From center	7	6	5	4	3	2	1	0
Weighting	0.195	0.383	0.556	0.707	0.832	0.924	0.981	1

ture,[19,35] the peaks corresponding to ethanol and glucose in the spectra are known, and through examining the spectra, it was possible to establish the important peaks. The following peaks were chosen (in wavenumbers):

Glucose peaks: 420, 437, 509, 536, 840, 850, 910, 1056, 1118, 1347 cm^{-1}.
Ethanol peaks: 488, 874, 1080, 2925 cm^{-1}.
Unknown peaks: 1038, 2873, 2971 cm^{-1}.

A peak is said to be present if there is a maximum within 10 points of these values. The height h of the peak is the height of this maximum point. The peak start and end are the minima on either side of this peak. Where the spectrum has a small peak but becomes lower than the first minimum found within seven points, this is ignored and included within the peak. This approach helps to overcome any remaining noise in the data. The width w is the difference between these two points. The position p of the peak is half way between the start and end of the peak. The area a is the sum of the height of all the points between the start and the end, inclusive. The top t is the position of the top of the peak (where the height was measured) and the skewness s is $t-p$. These values are henceforth collectively referred to as the *hwpast* of the peak. Where there is no peak within 10 points of the specified point, h is set to the height at the specified point; w, a, and s are set to 0; and p and t are set to the value of the specified point. This is necessary so as not to skew the statistical methods to be applied to the data unduly.

The *hwpast* values of the 17 peaks for all the spectra were then imported into Microsoft Excel, to be processed by the AVSM suite of Excel macros. These have been previously described in Ref. 60, where they were used for the classification of olive oil variety and region, but have since been updated for quantification data. They perform variable selection, using (in this case) either Mutual Information[68,69] or simple (product moment) correlation[70] on each variable against the measured concentration from the chemical assay. Mutual Information is particularly useful where the data are nonlinear, and nonlinear methods of analysis are to be used (e.g., artificial neural networks). Since we deal here with PLS (which is linear), and do not suspect a great degree of nonlinearity in the data in any case, the results shown here have used product moment correlation for variable selection. With the use of statistical programs written in-house (see Refs. 61 and 62), which include PCR, MLR, and standard backpropagation neural networks as well as the PLS used here, the Excel macros establish the best number of variables to use for a prediction.

The data are divided into three sets: a *training set*, which is used to form the model; a *query set*, which in conjunction with the training set is used to establish the optimum number of factors for the model; and an independent *validation set* (sometimes called the *test set*), unseen by the model creation program at any time, which is used to create a prediction. An independent validation set enables us to help ensure that the model created is valid; that is, that it is not modeling some chance trend in the data. The data are divided into their data sets simply by placing the first sample (that is, the first measurement of the fermentation) in the training set, the second in the query set, the third in the validation set, and so on, ensuring that the last point is in the training set (to avoid extrapolation of models).

Although in practice the validation set has been used to determine the order of the variables, this arrangement does not appear to make a substantial difference to the order of selection.

The data having been divided into appropriate sets, the AVSM macros are run to produce the results. A PLS prediction[71] of the concentration of ethanol, using three data sets as described earlier, gave a best prediction of the query set using just three variables; these were the height of peaks 11 (874 cm^{-1}, ethanol), 9 (1118 cm^{-1}, glucose), and 8 (1056 cm^{-1}, glucose) in that order. The model used one factor (Fig. 7A). Similarly (Fig. 7B), with the same variables and one factor, a successful prediction of the amount of glucose was obtained. The percentage root mean squared error (RMSE) is expressed as a percentage of the total range of the data, so, in Figs. 7 and 8 the actual RMSE can be obtained thus:

$$\frac{(\max(ethanol) - \min(ethanol)) \times \%\mathrm{RMSE}}{100}$$

$$= \frac{(785.9 - 0) \times 3.892}{100} = 30.587.$$

Having shown that it is possible to produce very good predictions of the stage of a fermentation using data from the same fermentation for creating a model, the next step is to try to predict a second fermentation[72] using the first biotransformation for the training and query sets.

The macros, using PLS, carried out on the query data, showed that eight variables (correlation selected) were best as judged by the query set: $h11$ (906, ethanol), $h9$ (1192, glucose), $a9$, $a11$, $h3$ (479, glucose), $h8$ (1120, glucose), $h7$ (949, glucose), and $h3$, although the results (Fig. 8) are inevitably not quite as good as was a prediction of ethanol given training, query and, validation data from the same biotransformation. Nonetheless, it is clear from these results that, with sufficient data, predictions of the progress of a biotransformation are possible using a model formed only on data from a previous biotransformation. This is testament to the reproducibility both of our methods and of the chemical analysis of the dispersive Raman spectroscope.

For comparison, PLS was also carried out on the spectra for biotransformation 2 on the concentration of ethanol at different stages in the processing (see Table II).

It may be noted that Table II shows an *increase* in the error at the stage of baseline removal. This is to be expected, because it is almost impossible to remove a baseline without some information being lost. However, without baseline removal of some form, the subsequent steps of calculating peak dimensions would not be possible. It is also known (from previous unpublished observations) that variations in the baseline are often due to effects such as changes in the optical density (turbidity) of the sample or of fluorescence, which may be related to the stage of

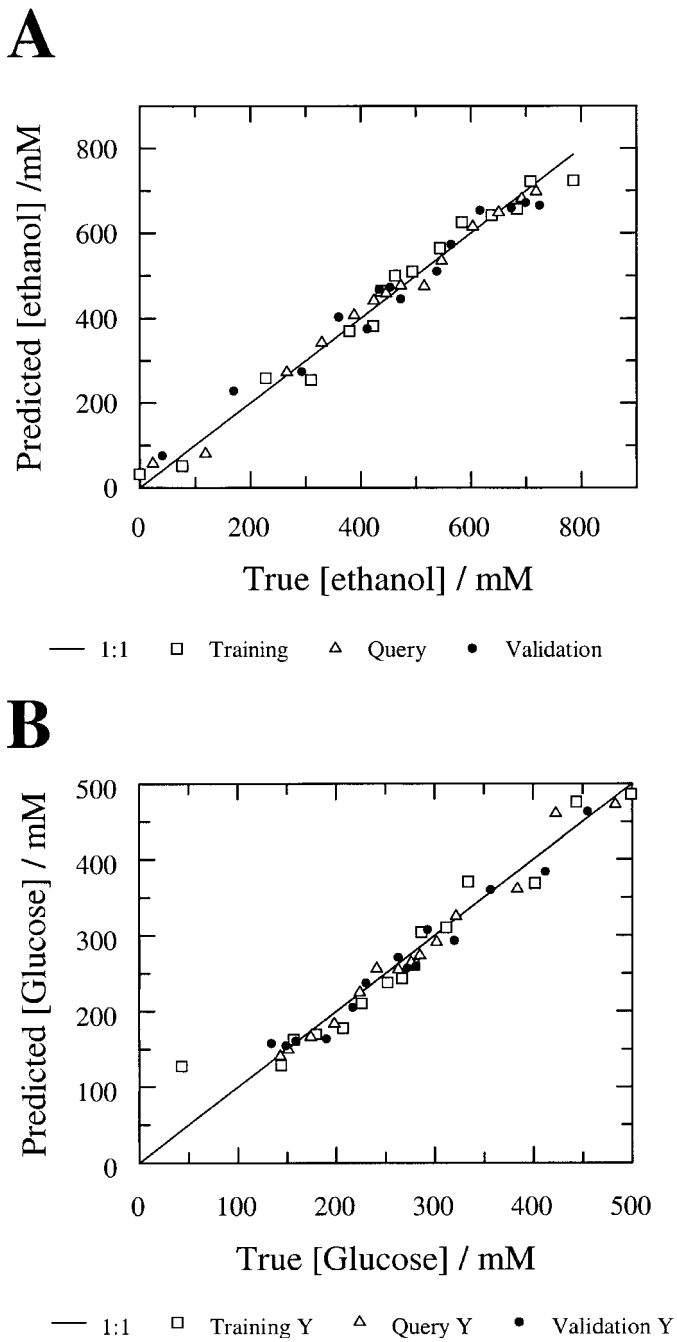

FIG. 7. PLS prediction of concentrations of ethanol and glucose during biotransformation 2. Best three variables (from 102), selected by correlation. **(A)** Ethanol; percentage RMSE 3.89 (validation set 4.48; query set 2.64). **(B)** Glucose; percentage RMSE 4.80 (validation set 3.72, query set 3.21).

the biotransformation but are not therefore reflected in the peak dimensions. Such variations should be ignored by a robust modeling system, as in the present case.

It has been suggested that the accuracy of the peak dimension measurements may yield results inferior to that which may be obtained through simply using regions of the spectra around the peaks. With 11 bins taken from around each individual peak (from 5 below to 5 above) it has been shown that there is little difference. However, in spectra where the position of the top of the peak is not consistent (this is apparently not the case here), such a method would be severely flawed. It must also be noted that many more variables were required to obtain the best results by using bins alone, which must count considerably against it. Table II shows the results obtained by using actual spectral values, both before and after baseline removal.

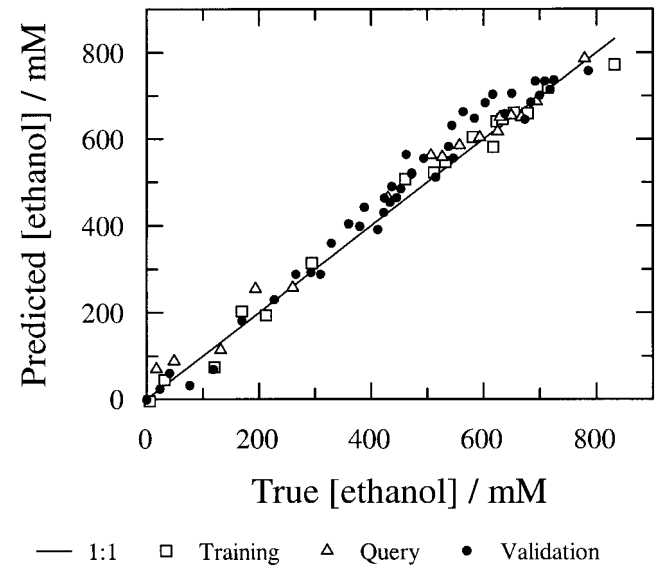

FIG. 8. Prediction of concentration of ethanol in biotransformation 2 from the model formed on biotransformation 1. Best eight variables, correlation selected. Percentage RMSE 4.58 (validation set 5.24, query set 3.72).

CONCLUSION

The present work has demonstrated that Raman spectroscopy has properties that make it an ideal method for following chemical changes occurring during biotransformations in a nondestructive and noninvasive manner. The dispersive Raman spectrometer used in the present work is stable and robust, as observed from the highly reproducible spectra and data. The 780 nm diode laser has also proved to be an effective wavelength for suppressing the overwhelming influence of fluorescence from the yeast cells. It is possible that excitation at 830 nm (presently under investigation) will yield even further reduced fluorescence. The collection times for the spectra were relatively long, and thus the use of a higher intensity laser with shorter collection times to eliminate any sample heating effects is also presently under investigation.

TABLE II. The error in the predictions of the data sets at various stages of the data processing, showing that the chemometric methods applied are essential for optimal results.

	All data	Query set	Validation set
Raw spectra	9.972	12.473	12.243
Cosmic and smoothed	7.162	8.492	9.183
Normalized	6.114	6.688	8.364
Eleven bins from each peak, best 14 variables	3.573	4.003	2.896
Baseline removed	7.972	9.626	10.124
Eleven bins, baseline removed, best 62 variables	3.206	3.153	4.39
All *hwpast* variables	5.663	6.817	7.123
Best 3 *hwpast* variables	3.892	2.641	4.478

The amount of data produced by Raman spectroscopy in terms of data points is huge. To analyze all these points would mean that most of the data under analysis are not relevant. The distinguishing features in Raman spectra are the positions and size of peaks. It is therefore necessary to measure only the parameters for the peaks that are known, or observed, to be of significance for the system under examination. Not only does this approach speed up the data analysis, but it means that the models are not unduly affected by noise, which leads to more parsimonious models[73] and therefore better predictive ability. If there were no noise in the data, one peak would be able to predict perfectly; the ethanol peak, being the largest and less subject to measurement errors, would be chosen. However, because of noise and errors, a number of peaks are required. The methods described here have yielded results that could not have been achieved without the use of chemometric techniques.

If the number of photons reaching the CCD per unit of metabolite concentration varies greatly with time, then errors will inevitably be introduced into the predictions. Thus, as with many spectroscopic methods, normalization to a peak of reproducible height is important, and has been found to improve our results, although great variations in photon numbers have not been experienced in the experiments presented here.

While others[33,34,51] have shown that the amount of glucose in aqueous solution can be measured by using Raman spectroscopy, this is, to our knowledge, the first online measurement of such a biotransformation. Indeed, the residual errors in the results presented here are sufficiently close to the expected experimental error to suggest that significant further improvement in precision cannot easily be expected.

ACKNOWLEDGMENTS

We thank the UK BBSRC, Renishaw plc Transducers Systems Division and Zeneca Life Science Molecules for financial support, and Drs. Ken Williams, Ray Chaney, and Chris Dyer for many useful discussions.

1. M.-N. Pons, *Bioprocess Monitoring and Control* (Hanser, Munich, 1991).
2. D. B. Kell and B. Sonnleitner, Trends Biotechnol. **13,** 481 (1995).
3. G. A. Montague, *Monitoring and Control of Fermenters* (Institute of Chemical Engineers, London, 1997).
4. D. B. Kell, G. H. Markx, C. L. Davey, and R. W. Todd, Trends Anal. Chem. **9,** 190 (1990).
5. T. Hirschfeld and B. Chase, Appl. Spectrosc. **40,** 133 (1986).
6. J. Góral and V. Zichy, Spectrochim. Acta **46A,** 253 (1990).
7. B. Schrader, in *Practical Fourier Transform IR Spectroscopy (FTIR)*, J.R. Ferraro and K. Krishnan, Eds. (Academic Press, New York/London, 1990), p. 167.
8. B. Schrader, *Infrared and Raman Spectroscopy: Methods and Applications* (Verlag Chemie, Weinheim, 1995).
9. J. G. Graselli and B. J. Bulkin, *Analytical Raman Spectroscopy* (John Wiley, New York, 1991).
10. P. Hendra, C. Jones, and G. Warnes, *Fourier Transform Raman Spectroscopy* (Ellis Horwood, Chichester, 1991).
11. G. J. Puppels, W. Colier, J. H. F. Olminkhof, C. Otto, F. F. M. Demul, and J. Greve, J. Raman Spectrosc. **22,** 217 (1991).
12. G. J. Puppels and J. Greve, Adv. Spectrosc. **20A,** 231 (1993).
13. B. Chase, Appl. Spectrosc. **48,** 14A (1994).
14. D. L. Gerrard, Anal. Chem. **66,** R 547 (1994).
15. S. Keller, B. Schrader, A. Hoffmann, W. Schrader, K. Metz, A. Rehlaender, J. Pahnke, M. Ruwe, and W. Budach, J. Raman Spectrosc. **25,** 663 (1994).
16. S. F. Parker, Spectrochim. Acta **50A,** 1841 (1994).
17. B. Schrader, G. Baranovic, S. Keller, and J. Sawatzki, Fresenius' J. Anal. Chem. **349,** 4 (1994).
18. P. J. Treado and M. D. Morris, Appl. Spectrosc. Rev. **29,** 1 (1994).
19. J. Twardowski and P. Anzenbacher, *Raman and Infrared Spectroscopy in Biology and Biochemistry* (Ellis Horwood, Chichester, 1994).
20. D. Naumann, S. Keller, D. Helm, C. Schultz, and B. Schrader, J. Mol. Struct. **347,** 399 (1995).
21. P. J. Hendra, H. M. M. Wilson, P. J. Wallen, I. J. Wesley, P. A. Bentley, M. Arruebarrena Baez, J. A. Haigh, P. A. Evans, C. D. Dyer, R. Lehnert, and M. V. Pellow-Jarman, Analyst **120,** 985 (1995).
22. G. J. Puppels, T. C. B. Schut, N. M. Sijtsema, M. Grond, F. Maraboeuf, C. G. Degrauw, C. G. Figdor, and J. Greve, J. Mol. Struct. **347,** 477 (1995).
23. F. Adar, R. Geiger, and J. Noonan, Appl. Spectrosc. Rev. **32,** 45 (1997).
24. M. M. Carrabba, K. M. Spencer, C. Rich, and D. Rauh, Appl. Spectrosc. **44,** 1558 (1990).
25. G. J. Puppels, A. Huizinga, H. W. Krabbe, H. A. Deboer, G. Gijsbers, and F. F. M. Demul, Rev. Sci. Instrum. **61,** 3709 (1990).
26. M. Kim, H. Owen, and P. R. Carey, Appl. Spectrosc. **47,** 1780 (1993).
27. J. M. Tedesco, H. Owen, D. M. Pallister, and M. D. Morris, Anal. Chem. **65,** A 441 (1993).
28. P. J. Treado and M. D. Morris, Spectrochim. Acta Rev. **13,** 355 (1990).
29. J. F. Turner and P. J. Treado, Appl. Spectrosc. **50,** 277 (1996).
30. P. R. Griffiths and J. A. de Haseth, *Fourier Transform Infrared Spectrometry* (John Wiley, New York, 1986).
31. K. P. J. Williams, G. D. Pitt, D. N. Batchelder, and B. J. Kip, Appl. Spectrosc. **48,** 232 (1994).
32. K. P. J. Williams, G. D. Pitt, B. J. E. Smith, A. Whitley, D. N. Batchelder, and I. P. Hayward, J. Raman Spectrosc. **25,** 131 (1994).
33. J. P. Wicksted, R. J. Erckens, M. Motamedi, and W. F. March, Appl. Spectrosc. **49,** 987 (1995).
34. R. J. Erckens, M. Motamedi, W. F. March, and J. P. Wicksted, J. Raman Spectrosc. **28,** 293 (1997).
35. T. B. Shope, T. J. Vickers, and C. K. Mann, Appl. Spectrosc. **41,** 908 (1987).
36. C. Gomy, M. Jouan, and N. Q. Dao, Comptes Rendus de l'Academie des Sciences Serie Ii-Mecanique Physique Chimie Sciences de l'Univers Sciences de la Terre **306,** 417 (1988).
37. C. Gomy, M. Jouan, and N. Q. Dao, Anal. Chim. Acta **215,** 211 (1988).
38. A. G. Cavinato, D. M. Mayes, Z. H. Ge, and J. B. Callis, Anal. Chem. **62,** 1977 (1990).
39. Z. H. Ge, A. G. Cavinato, and J. B. Callis, Anal. Chem. **66,** 1354 (1994).
40. G. Vaccari, E. Dosi, A. L. Campi, and G. Mantovani, Zuckerindustrie **118,** 266 (1993).
41. G. Vaccari, E. Dosi, A. L. Campi, A. Gonzalezvara, D. Matteuzzi, and G. Mantovani, Biotechnol. Bioeng. **43,** 913 (1994).
42. M. Varadi, A. Toth, and J. Rezessy, *Application of NIR in a Fermentation Process* (VCH Publishers, New York, 1992).
43. S. V. Hammond and I. K. Brookes, in *Harnessing Biotechnology for the 21st Century*, A. Bose, Ed. (American Chemical Society Washington, D.C., 1992), pp. 325–333.
44. S. V. Hammond, in *Making Light Work: Advances in Near Infrared Spectroscopy*, I. Murray and I. A. Cowe, Eds. (VCH, New York, 1992), pp. 584–589.
45. J. W. Hall, B. McNeil, M. J. Rollins, I. Draoer, B. Thompson, and G. Macaloney, Appl. Spectrosc. **50,** 102 (1996).
46. G. Macaloney, I. Draper, J. Preston, K. B. Anderson, M. J. Rollins, B. G. Thompson, J. W. Hall, and B. McNeil, Food Bioprod. Process. **74,** 212 (1996).
47. G. Macaloney, J. W. Hall, M. J. Rollins, I. Draper, B. G. Thompson, and B. McNeil, Biotechnol. Tech. **8,** 281 (1994).
48. T. Yano, T. Aimi, Y. Nakano, and M. Tamai, J. Ferment. Bioeng. **84,** 461 (1997).
49. P. Fayolle, D. Picque, B. Perret, E. Latrille, and G. Corrieu, Appl. Spectrosc. **50,** 1325 (1996).
50. P. Fayolle, D. Picque, and G. Corrieu, Vib. Spectrosc. **14,** 247 (1997).
51. C. H. Spiegelman, M. J. McShane, M. J. Goetz, M. Motamedi, Q. L. Yue, and G. L. Cote, Anal. Chem. **70,** 35 (1998).

52. D. M. Haaland, K. L. Higgins, and D. R. Tallant, Vibr. Spectrosc. **1,** 35 (1990).
53. T. J. Vickers and C. K. Mann, in *Analytical Raman Spectroscopy*, J. G. Graselli and B. J. Bulkin, Eds. (John Wiley, New York, 1991), p. 107.
54. N. Q. Huy, V. T. Bich, M. Jouan, and N. Q. Dao, Analysis **20,** 141 (1992).
55. N. Q. Huy, M. Jouan, and N. Q. Dao, Appl. Spectrosc. **47,** 2013 (1993).
56. R. Manoharan, J. J. Baraga, M. S. Feld, and R. P. Rava, J. Photochem. Photobiol. B–Biology **16,** 211 (1992).
57. Y. Liu, B. R. Upadhyaya, and M. Naghedolfeizi, Appl. Spectrosc. **47,** 12 (1993).
58. T. J. Vickers, C. A. Rosen, and C. K. Mann, Appl. Spectrosc. **50,** 1074 (1996).
59. C. H. Tseng, J. F. Ford, C. K. Mann, and T. J. Vickers, Appl. Spectrosc. **47,** 1808 (1993).
60. A. D. Shaw, A. DiCamillo, G. Vlahov, A. Jones, G. Bianchi, J. Rowland, and D. B. Kell, Anal. Chim. Acta **348,** 357 (1997).
61. A. Jones, A. D. Shaw, G. J. Salter, G. Bianchi, and D. B. Kell, in *Lipid Analysis of Oils and Fats*, R. J. Hamilton, Ed. (Chapman and Hall, London, 1998), p. 317.
62. P. J. Wyatt, Meth. Microbiol. **8,** 183 (1973).
63. H. M. Davey and D. B. Kell, Microbiol. Rev. **60,** 641 (1996).
64. A. Jones, D. Young, J. Taylor, D. B. Kell, and J. J. Rowland, Biotechnol. Bioeng. **59,** 131 (1998).
65. B. K. Alsberg, A. M. Woodward, M. K. Winson, J. Rowland, and D. B. Kell, Analyst **122,** 645 (1997).
66. B. K. Alsberg, A. M. Woodward, and D. B. Kell, Chemom. Intell. Lab. Syst. **37,** 215 (1997).
67. B. Schrader, *Raman/Infrared Atlas of Organic Compounds* (VCH Publishers, New York, 1989), 2nd ed.
68. C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (University of Illinois Press, Urbana, Illinois, 1949).
69. R. Battiti, IEEE Trans. Neural Networks **5,** 537 (1994).
70. R. J. Gilbert, R. Goodacre, A. M. Woodward, and D. B. Kell, Anal. Chem. **69,** 4381 (1997).
71. H. Martens and T. Næs, *Multivariate Calibration* (John Wiley, Chichester, 1989).
72. A. M. Woodward, A. Jones, X. Zhang, J. Rowland, and D. B. Kell, Bioelectrochem. Bioenerg. **40,** 99 (1996).
73. M. B. Seasholtz and B. Kowalski, Anal. Chim. Acta **277,** 165 (1993).