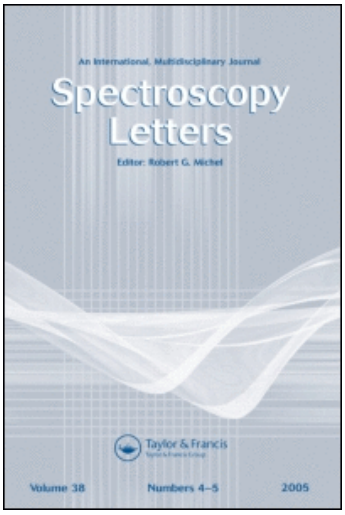

Spectroscopy Letters
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597299

Sequential Injection/Mid-Infrared Spectroscopic Analysis of an Acetone-Butanol-Ethanol Fermentation: Analyte Cross-Correlation Effects

Mustafa Kansiz[ab]; K. Christian Schuster[a]; Don McNaughton[b]; Bernhard Lendl[a]
[a] Institute of Analytical Chemistry, Vienna University of Technology, Vienna, Austria [b] School of Chemistry, Centre for Green Chemistry, Monash University, Melbourne, Victoria, Australia

To cite this Article Kansiz, Mustafa , Schuster, K. Christian , McNaughton, Don and Lendl, Bernhard(2005) 'Sequential Injection/Mid-Infrared Spectroscopic Analysis of an Acetone-Butanol-Ethanol Fermentation: Analyte Cross-Correlation Effects', Spectroscopy Letters, 38: 6, 677 — 702
To link to this Article: DOI: 10.1080/00387010500315801
URL: http://dx.doi.org/10.1080/00387010500315801

Spectroscopy Letters, 38: 677–702, 2005

ISSN 0038-7010 print/1532-2289 online
DOI: 10.1080/00387010500315801

Taylor & Francis
Taylor & Francis Group

Sequential Injection/Mid-Infrared Spectroscopic Analysis of an Acetone-Butanol-Ethanol Fermentation: Analyte Cross-Correlation Effects

Mustafa Kansiz
Institute of Analytical Chemistry, Vienna University of Technology, Vienna, Austria and School of Chemistry, Centre for Green Chemistry, Monash University, Melbourne, Victoria, Australia

K. Christian Schuster
Institute of Analytical Chemistry, Vienna University of Technology, Vienna, Austria

Don McNaughton
School of Chemistry, Centre for Green Chemistry, Monash University, Melbourne, Victoria, Australia

Bernhard Lendl
Institute of Analytical Chemistry, Vienna University of Technology, Vienna, Austria

Abstract: Mid-infrared spectroscopy together with sequential injection analysis (SIA) and partial least squares (PLS) regression analysis was used to monitor acetone-butanol-ethanol (ABE) fermentations under different fermentation conditions. Five analytes were simultaneously predicted (acetone, acetate, butyrate, *n*-butanol, and glucose). In order to compare the overall model prediction ability, a relative average

Received 4 September 2004, Accepted 28 July 2005
This paper was by special invitation as a contribution to a special issue of the journal entitled "Quantitative Vibrational Spectrometry in the 21st Century." This special issue was organized by Professor Miguel de la Guardia, Professor of Analytical Chemistry at Valencia University, Spain.
Address correspondence to B. Lendl, Institute of Analytical Chemistry, Vienna University of Technology, Getreidemarkt 9/151, Vienna A-1060, Austria. E-mail: blendl@mail.zserv.tuwien.ac.at

of the root mean square error of prediction (RMSEP) across all five analytes was employed. To form a PLS model devoid of any cross-correlations between analytes, a synthetic calibration data set was created by the SIA system. As a test of their robustness, PLS models from synthetic samples and those from real fermentation samples were compared and used to predict samples from the opposite data set and from independent "acid-crash" fermentations. The PLS model developed from the synthetic samples proved to be far more robust and accurate and used fewer factors than PLS models from the real fermentations, which were found to contain analyte cross-correlations. The use of synthetic data enabled more accurate selection of factors and showed the importance of investigating spectral regression coefficients plots to aid and confirm appropriate factor selection. In addition, an alternative method of factor selection was proposed, using a "similarity measure" between the regression coefficient plots of factors for certain analytes and their standard spectra. Predictions using this method of factor selection over the common "minimum from an error vs. factor" plot proved to be more accurate and used far fewer factors.

Keywords: ABE fermentation, analyte cross-correlations, mid-infrared spectroscopy (FTIR), PLS

INTRODUCTION

Mid-infrared spectroscopy in the analysis of fermentations was introduced in the early 1980s for the analysis of antibiotic fermentation samples,[1] analysis of model fermentations,[2] feasibility studies for real-time process control of fermentations,[3] and the investigation of pyruvic acid assimilation by *Escherichia coli*.[4]

To date, near-infrared spectroscopy has been more widely applied to fermentation systems[5–9] than mid-infrared spectroscopy due to the ease of sample handling, which allows aqueous solutions to be sampled in flow-through cells with pathlengths in the millimeter and centimeter ranges, as opposed to the few micrometers traditionally required in the mid-infrared. In addition, the ready availability of fiberoptic materials in the near-infrared has made *in situ* measurements practical.[5] Although near-infrared spectroscopy can have sample handling advantages, it is lacking in qualitative spectral information because spectra generally consist of broad overlapping spectral features with little or no qualitative information. Mid-infrared spectroscopy on the other hand is rich in both quantitative and qualitative spectral information, making the interpretation of the spectral changes over the course of a fermentation much easier because individual bands can be assigned to particular chemical moieties of the analytes.

With recent improvements in instrumentation and in particular in attenuated total reflection (ATR) probes for *in situ* measurement of chemical reactions, the number of publications concerned with mid-infrared spectroscopy for fermentations monitoring has increased.[3,10–17] The use of an ATR sampling probe is most advantageous when dealing with aqueous

and particulate containing solutions, because the very small and constant effective pathlength of the ATR eliminates the need for transmission cells. ATR accessories typically have a depth of penetration of $1-2\,\mu m$, providing for very small effective pathlengths, therefore making for the easy subtraction or ratioing of highly absorbing matrices such as water. One important practical problem associated with the use of *in situ* ATR probes is the potential for the development of biofilms on the probe surface, which can interfere with the intended monitoring of the liquid samples.

For evaluation of the Fourier transform infrared (FTIR) spectroscopy recorded from fermentations, multivariate statistical techniques, most commonly partial least squares (PLS) regression, have been employed. These data evaluation methods enable the simultaneous determination of several analytes from within a complex media background. However, the PLS method developed needs to be carefully prepared and analyzed to ensure that a robust calibration model covering all possible combinations of analyte concentrations is created. In this context, the analysis of possible analyte cross-correlations is particularly useful. Cross-correlations between various analytes within fermentations are well-known and are to be expected due to the stoichiometry of the physiological processes of a microbiological cell. For example, for the acetone-butanol-ethanol (ABE) fermentation[18–20] used in this study and previously.[17] essentially glucose is consumed, as acetone and *n*-butanol are produced. Therefore, the glucose concentration is negatively correlated with acetone and *n*-butanol. Also, because acetone and *n*-butanol are produced at the same time and from the same source, they are positively correlated with each other, with approximately twice as much *n*-butanol as acetone being produced. Correlations with or among acetate and butyrate, which are the other two analytes that can be produced or consumed during glucose consumption, are much less pronounced. These two components are present in relatively lower concentrations, with, in general, their concentrations starting off low and slowly decreasing over the course of the fermentation.

For the sake of model robustness, it is important to carefully examine the PLS model for evidence that only the spectral features of the analyte of interest are being used in prediction calculations. The presence of analyte cross-correlations in the calibration can significantly degrade the robustness of the developed model. This is of particular importance when fermentations are analyzed that proceed differently to those fermentations on which the calibration has been based.

A possible solution to this problem is the use of synthetic samples devoid of analyte cross-correlations for the construction of the PLS model. In this paper, the impact of analyte cross-correlations on method robustness is demonstrated. Furthermore, it is shown how the spectral regression coefficients plots can efficiently be used to detect analyte cross-correlations present in the calibration model. This paper follows on from a previous study on the same fermentation system.[17]

EXPERIMENTAL

Organism

Clostridium beijerinckii NRRL B592 (capable of producing a mixture of acetone, *n*-butanol, and ethanol) was obtained from the laboratory of J. R. Gapes (Institute of Chemical Engineering, Fuel and Environmental Technology, Vienna University of Technology, Vienna, Austria). A previous high-solvent-producing and well-sporulated batch fermentation on semisynthetic medium was used as spore stock.

Medium

A modified semisynthetic medium according to Ref.[21] was used for both inoculum and batch cultures and contained per liter of distilled water the following: glucose 60.0 g, yeast extract 5.0 g, ammonium acetate 3.0 g, KH_2PO_4 1.0 g, K_2HPO_4 0.8 g, $MgSO_4 \cdot 7H_2O$ 1.0 g, cysteine hydrochloride 0.5 g, $FeSO_4 \cdot 7H_2O$, and *p*-aminobenzoic acid 10 mg.

Inoculum Preparation

The medium for inoculation was made anaerobic by bubbling oxygen-free sterile nitrogen for 10 min before autoclaving at 10^5 Pa (121°C) for 16 min. The inoculum precultures were prepared by injecting 10 mL of a spore suspension into 330 mL of media in a 500 mL screw-cap bottle plugged with a rubber stopper into which was placed a needle connected to a bicycle valve via a short length of tubing. This acted as a pressure release valve to release the buildup of carbon dioxide and hydrogen during fermentation. This inoculum preculture was then heat shocked at 85°C for 10 min and cooled rapidly in cold water before being placed in a 35°C incubator.

Batch Cultures in the Bioreactor

Fermentations No. 1, No. 2, and No. 3 "Normal"

The bioreactor was a three-necked 2-L glass vessel with a working volume of 1700 mL. All contents of the bioreactor including the pH probe and polytetrafluoroethylene (PTFE) sampling tubing (for mid-infrared spectroscopic measurements) were autoclaved after the medium was made anaerobic by bubbling oxygen-free sterile nitrogen for 20 min. After autoclaving at 10^5 Pa (121°C) for 25 min and until inoculation, oxygen-free sterile nitrogen was bubbled through the medium to maintain anaerobic conditions while sitting the bioreactor in a

cold-water bath. The batch culture was brought up to 35°C in a thermostated water bath before being inoculated with 100 mL of a 20–24 hr inoculum preculture. The pH was continuously monitored on-line, with optical density (615 nm, 1 cm plastic cuvette) measurements being taken for each sample. Samples were collected (both spectra and for reference analysis) at regular intervals over the course of the fermentation, up to 50 hr. Each individual fermentation had 10–12 complete samples (spectra and reference analysis). The data from fermentation no. 3 has been previously reported in Ref.[17].

Fermentations No. 4 and No. 5 "Acid Crash"

These fermentation were performed exactly as for fermentations 1–3, except that the pH was controlled at pH 7 via the automated addition of 2 M KOH. This was done in an attempt to induce an "acid-crash" fermentation, as it is believed that the concentration of free acids play an important role in the onset of an "acid crash."[22] Spectra and samples for reference analysis were collected at regular intervals throughout the course of the fermentation. Between 7 and 8 complete samples (spectra and reference analysis) were collected for each fermentation.

A summary of the data sets used is present in Table 1.

Preparation and Measurement of Synthetic Calibration Standards No. 6

A full factorial design employing five variables (the five analytes: acetone, acetate, butyrate, *n*-butanol, and glucose) at two concentration levels (high and low concentrations, where the lower concentrations were one-fifth of the higher concentrations) was calculated giving a total of 32 combinations (2^5). This full factorial design calculates all statistically possible combinations of the high and low levels for all five analytes, hence giving a sample set in which all cross-correlations between the analyte concentrations have been removed.

All analyte stock solutions for the automated preparation of the synthetic calibration set were made up in volumetric flasks using distilled water. The media background preparation method was modified from that described

Table 1. Summary of fermentations/experiments used to compile each data set

Data set	Number of samples	Fermentation/experiments from which data were collected	Fermentation type
A	35	1, 2, & 3	Normal fermentation
B	15	4 & 5	Acid-crash fermentation
C	50	1, 2, 3, 4, & 5	Normal & acid-crash fermentation
Syn.	32	6	Synthetic mixtures

earlier with the acetate omitted, because this was added from an acetate stock solution, and only 12 g/L of glucose monohydrate was added to act as the "low" glucose sample as opposed to the normal 60 g/L. To reduce the complexity, combinations of high and low glucose and *n*-butanol were pre-prepared in four (hiGhiB, hiGLoB, loGhiB, loGloB) 320-mL volumes in 500-mL Schott bottles. The *n*-butanol was added aseptically to the media background after autoclaving. An additional 15.36 g of glucose monohydrate was added to the bottles that required a "high" glucose level. The standard solutions and all other reagents were of analytical grade quality. The acetate and butyrate solutions were prepared from their acids (acetic acid and *n*-butyric acid, respectively) and were neutralized by the addition of NaOH pellets. The final glucose concentrations were determined by the reference enzymatic assay method, because the exact glucose content of glucose monohydrate is not known. The acetone, acetate, and butyrate solutions were added automatically by the sequential injection analysis (SIA) system to the media background (containing the glucose and *n*-butanol) in order to obtain the desired final concentrations of the solutions to be measured. The final concentrations of the analytes and the stock solutions is presented in Table 2.

For analysis of the spectral regression coefficients plots, FTIR spectra of pure aqueous standard solutions of the analytes were also required. Standards were prepared with the following concentrations: acetone, 8 g/L; *n*-butanol, 14 g/L; glucose, 60 g/L; acetic acid, 4 g/L; and *n*-butyric acid, 4 g/L. The pH of the acid solutions were increased to approximately 12, by the addition of NaOH pellets, to deprotonate all of the acids.

The spectra of all of pure aqueous standard solutions and samples for the synthetic calibration set were collected in the exact same manner as the spectra for real fermentation samples.

Sequential Injection Analysis Manifold Coupled to the FTIR Spectrometer

The SIA system was set up with a Cavro XP 3000 syringe pump (syringe size 2500 μL), 4-mL holding coil, and a Valco 14 port selection valve equipped

Table 2. Summary of synthetic calibration sample concentrations

	High concentration (g/L)	Low concentration (g/L)
Acetone	6.00	1.20
Acetate	4.00	0.80
Butyrate	4.00	0.80
n-Butanol	15.00	3.00
Glucose	49.0 (60.0)[a]	9.4 (12.0)[a]

[a]Figures in parentheses indicate the actual amount of glucose monohydrate weighed out.

with an electric microactuator. PTFE tubing (i.d. 0.75 mm) was used except for the sampling tubing from the fermenter, which had an i.d. of 0.97 mm. A Bruker IFS-55 spectrometer (Bruker Optik, Ettlingen, Germany) equipped with liquid nitrogen cooled mercury-cadmium-telluride (MCT) detector was used for spectral acquisition. A diamond ATR cell (Spectroscopy Central, Warrington, England) was employed with an in-house built flow-through accessory. This enables the pumping through of solutions over the ATR diamond surface with minimal dead-volume. All tubings and fittings were obtained from Global FIA (Gig Harbour, WA, USA). The SIA system was controlled by a personal computer, using the AnalySIA software package from the Turku Centre for Biotechnology (Åbo, Akademi University and University of Turku). This software was also used to trigger the FTIR spectrometer for the exact timing of the collection of spectra. A schematic diagram showing the experimental system is shown in Fig. 1.

FTIR On-line Fermentation Monitoring

A sample measurement cycle involves the aspiration of 1.5 mL of sample directly from the fermenter into the holding coil, followed by a plug of 250 μL of 1 M NaOH and 250 μL of distilled water. The addition of NaOH is necessary to bring the pH of the solution to be analyzed to approximately 10. This is to ensure that all of the acids (acetate and butyrate) are in a deprotonated state, well above their pKa's of 4.75 and 4.82, respectively. After pH adjustment, 2000 μL of this solution is then transferred into the

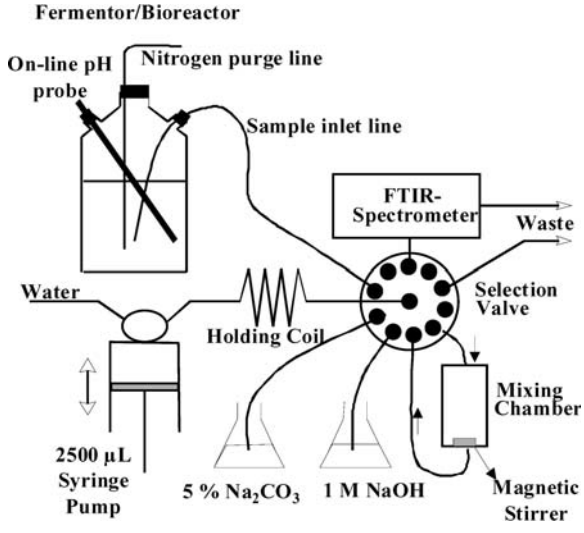

Figure 1. Schematic of SIA system.

mixing chamber where it is stirred by a small magnetic bead for 2 min. This allows sufficient time for degassing (CO_2) and complete mixing. Subsequently, 1.7 mL of this solution is pumped into the ATR flow-through cell at a rate of 720 μL/min, during which time the spectrometer is automatically triggered to collect the required spectra. Spectra were collected between 4000 and 700 cm^{-1} at 8 cm^{-1} resolution with 128 scans. A macro was written that enabled the recording of spectra upon receiving a trigger from the AnalySIA software. New background spectra were collected with distilled water in the cell, before the pumping through of each sample. Each sample was injected three times with three spectra recorded from each sample injection, giving in total nine spectra for each sampling. The triplicate spectra for each injection were averaged and used in the subsequent analysis. After spectral acquisition, all tubing, mixing chamber, and the ATR cell were washed, using the SIA setup with 2.5 mL of an aqueous 5% Na_2CO_3 solution to remove any biofilms. Subsequently, the flow system is washed again twice with 2.5 mL of distilled water. All these steps are performed automatically by the computer-controlled SIA setup.

Reference Measurements

The solvents (acetone and *n*-butanol) together with the acids (acetic and butyric) were determined simultaneously by gas chromatography based on a modified method of Thompson et al.[23] To 1 mL of clarified sample (centrifuged at 25,000 × g for 5 min) is added from a Pasteur pipette 1 drop of concentrated sulfuric acid and 100 μL of a 40 g/L *n*-propanol internal standard solution. A model GC-9A (Shimadzu Corp., Kyoto, Japan) GC was used, equipped with a flame ionization detector and a glass column (3.2 mm by 2.6 m) packed with Chromosorb 101 (Supelco, Inc., Bellefonte, PA, USA) with nitrogen gas (80 mL/min) as the carrier gas. A temperature program: 150°C (11.5 min), rising at 30°C/min to 190°C, for 7.2 min was used. The injector port temperature was set to 220°C and the detector temperature to 170°C. Glucose was determined using the enzymatic UV method (340 nm) test kit (Boehringer Mannheim cat. no. 716251, R-Biopharm GmbH, Darmstadt, Germany). All reference analyses were performed in triplicate. The precision of the reference methods compared with the FTIR predicted results have been reported previously.[17]

Data Analysis

The multivariative statistical method of partial least squares (mean-centered) was employed to create calibration models to relate the spectra to the known and established reference measurements. A PLS-1 method was used to create models for data sets where only one variable is being predicted. This is the

case where individual spectral window per analyte is to be employed. A PLS-2 method was employed where spectral region 1 (3000–2800, 1800–800 cm^{-1}) was used for all five analytes. This method enabled one model to be calculated for more than one analyte at a time.

All PLS models were developed with a leave-out-one cross-validation process (leaves one sample at a time out of the model-forming calculations). This sample is then predicted using this newly created model, and its prediction values are compared to the reference values. This process is repeated until all the samples have been left out once. It is by this method that the predicted versus measured values plot is constructed and prediction error versus factor plot are determined. It is from the minimum in this latter plot that the "optimum" number of factors for predictions are commonly chosen.

With large data sets and a number of measures of predictive model quality (as determined from a predicted vs. measure plot), such as root mean square error (RMSEP), R^2, slope, and offset, it is difficult to compare the performance of the various models predicting for the various data sets. Consequently, in order to more easily compare the results from each model, a measure of model accuracy, the relative RMSEP, was used and averaged across all five analytes to produce the average relative RMSEP. The measure will be referred to herein as the average relative model error (ARME). A mathematical description is presented below:

$$ARME = 100/m \sum_{s=1}^{m} \left[\frac{\sum_{i=1}^{p} \sqrt{(x_{pi} - x_{pr})^2}}{np} \right]_s$$

where n = the average reference value for the respective analyte, p = number of samples predicted per analyte, x_{pi} = predicted analyte value, x_{pr} = reference analyte value, and m = number of analytes predicted per model.

The ARME can be described as the prediction residual (predicted value − reference value) averaged across all samples for each analyte in a data set, then divided by the average reference value for the respective analyte. This percentage figure is averaged across all five analytes to provide a model average, the ARME. Expressing the results in a relative manner prevents analytes, such as glucose with its relatively higher concentrations, from disproportionately influencing the average. The ARME is thus able to give a good overall measure of the "goodness-of-fit" of a model by expressing the average relative prediction residual across all five analytes, allowing for a relatively easy comparison between models.

One of the many diagnostic tools available upon the formation of a PLS model, and one that is central to this study, is the spectral regression coefficients plot. In a regression model equation, regression coefficients are the numerical coefficients that express the link between variation in the predictors (absorbance values for each wavenumber in the infrared spectrum) and the variation in the response (analyte concentration). The spectral regression coefficients plot

summarizes the relationship between all predictors (absorbance values) for a given response (analyte concentration). The spectral regression coefficients can be computed for any number of factors. The spectral regression coefficients for 5 factors, for example, summarize the relationship between the predictors and the response as it is approximated by a model with 5 factors.[24]

Derivative spectra were used in preprocessing the data prior to PLS regression. Spectral derivatives can remove baseline artefacts, such as spectral offsets and sloping baselines. Spectral derivatives can also be considered as a pseudo resolution enhancement technique, because they are able to highlight slight variations in the slope and contours of bands. All these statistical methods were performed using the computer software package, The Unscrambler 6.11 (Camo Process, Oslo, Norway).

RESULTS AND DISCUSSION

FTIR Spectroscopy

Spectra from each of the triplicate injections per sample from the synthetic calibration sample were averaged to give a total of 32 averaged spectra Raw spectra and second derivative spectra are shown in Fig. 2, with the major spectral features of interest highlighted and labeled for band assignments as shown in Table 4. Despite the complex chemical background of

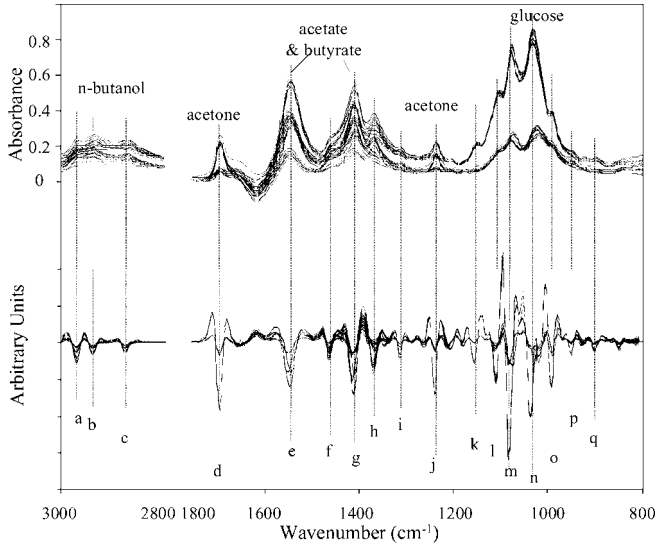

Figure 2. Assignment of major spectral features to analytes and spectral labeling (see Table 4).

Table 3. Summary of spectral regions and windows for each analyte

Analyte	Spectral region 1 (SR1) (cm^{-1})	Spectral windows (SW) (cm^{-1})
Acetone	3000–2800, 1800–800	1750–1650, 1400–1330, 1280–1200
Acetate	3000–2800, 1800–800	1640–1280
Butanol	3000–2800, 1800–800	2990–2850, 1550–1350, 1130–900
Butyrate	3000–2800, 1800–800	2990–2850, 1640–1280
Glucose	3000–2800, 1800–800	1230–960

the media, the spectral signals originating from the added chemical analytes can still be clearly discerned.

The removal of baseline features can be observed in the second-derivative spectra. This improves quantification by removing some of the differences in absorbance that are not related to any chemical component. The improvements to spectral precision can be seen quite clearly in Fig. 2 with the underivatized spectra showing a degree of "vertical spread" for a given absorbance reading at a certain wavenumber, whereas the derivatized spectra exhibit spectral features whose absorbances almost overlap. The power of spectral derivatives has been demonstrated in our previous publications.[17,25,26] It is also important to note the reversal of peak direction upon performing a second derivative. In order to assess the impact of using selected spectral "windows" compared with using a broader nonanalyte-specific region, spectral region 1 (SR1) (3000–2800, 1800–800 cm^{-1}) and analyte-specific spectral "windows" were defined, which are shown graphically as "boxed" areas in Fig. 3 and are tabulated in Table 3.

Analyte Cross-Correlations and Erroneous Spectral Regression Coefficients Plots

A number of research groups including ourselves[14,15,25,26] have successfully employed mid-IR spectroscopy and multivariative statistics to tackle various analytical problems, but most have not directly considered the possible cross-correlations that may exist between various analytes in the sample. Spectral regression coefficients plots from PLS models have seldom been used to investigate whether or not the spectral features of the analytes in question are in fact being used by the model for the prediction.

An example of a spectral regression coefficients plot for glucose from a PLS-2 model of data set A using SR1 at factor 3 where a minimum prediction error is reached for glucose (as determined by an error vs. factor plot) is presented in Fig. 4. The pure component spectra of the five analytes are also shown offset in Fig. 4 for comparison. This PLS model was calculated from data set A together with reference measurements.

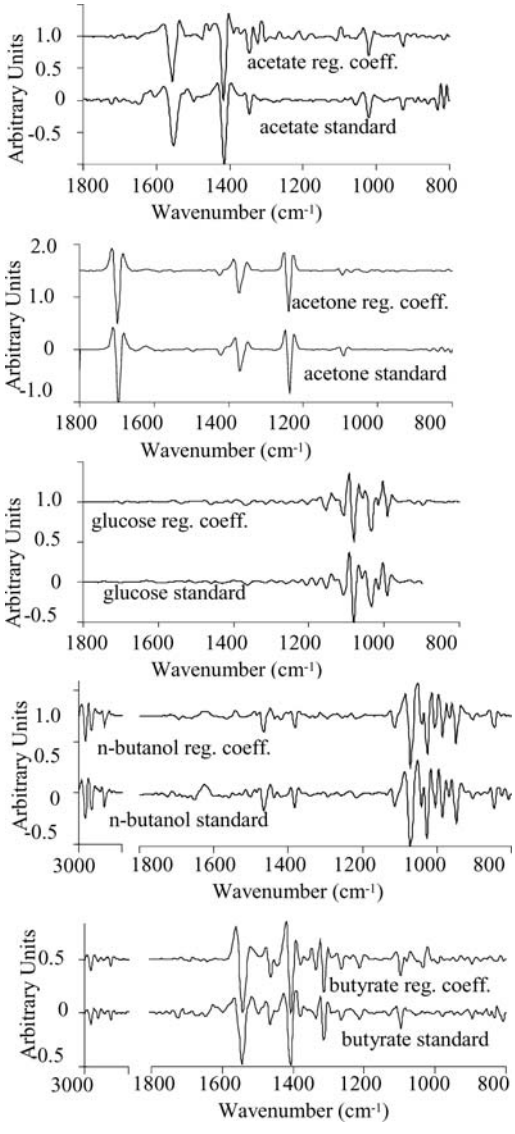

Figure 3. Comparison spectral regression coefficients plots from data set "syn" for all analytes at five factors, to spectral standards. Boxed areas indicate the analyte-specific spectral window.

From Fig. 4, it can be seen that despite selecting a factor for which there is a minimum in the error versus factor plot from leave-out-one cross-validated PLS model, only a few spectral features of glucose are discernable in the factor 3 regressions coefficients plot. Features from all of the other four

Table 4. Band assignments

Spectral label[a]	Band position (cm^{-1})	Assignments	Comments
a	2969	Asymmetric C–H stretching vibration of CH_3 groups	Mostly *n*-butanol, with some butyrate contributions.
b	2940	Asymmetric C–H stretching vibration of CH_2 groups	The 3000–2800 cm^{-1} is characteristic of general C–H stretching vibrations.
c	2879	Symmetric C–H stretching vibration of CH_3 groups	
d	1697	C═O (carbonyl) stretching vibration	Characteristic of acetone. 1800–1600 cm^{-1} is characteristic of C═O stretching vibrations.
e	1549	Asymmetric COO^- (carboxylate) stretching vibration	Contributions from acetate and butyrate. Their individual absorbances are at 1553 and 1542 cm^{-1}, respectively.
f	1466	Asymmetric C–C–H bending vibrations	Most contributions from *n*-butanol, with some from butyrate.
g	1410	Symmetric COO^- (carboxylate) stretching vibration	Contributions from acetate and butyrate. Their individual absorbances are at 1414 and 1409 cm^{-1}, respectively.
h	1370	Symmetric C–C–H bending vibrations	Acetone
i	1314	Symmetric C–C–H bending vibrations	Butyrate
j	1240	C–C stretching vibration	Acetone
k, l, m, n, o	1154, 1107, 1080, 1033, 991	C–O stretching vibrations	Mainly glucose. Characteristic region for strong carbohydrate (sugars) absorptions. Some *n*-butanol contributions as well.
p	947	CH_3 rocking vibration	*n*-Butanol

[a]The letter codes in column 1 refer to the letters in Fig. 2.

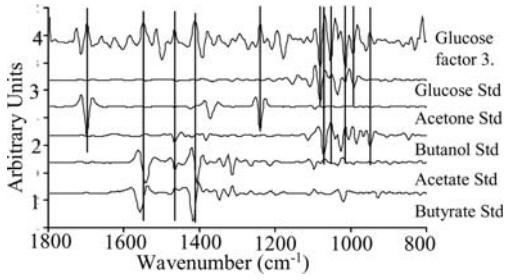

Figure 4. Spectral regression coefficient plot of glucose from data set "A" at three factors (where a minimum in the error vs. factor plot is reached) compared to all five analyte spectral standards.

analytes can be seen and are indicated by the vertical lines in Fig. 4, connecting the particular spectral feature of the other analytes to the features in the regression coefficients plot. The fact that there are clearly spectral features of analytes other than glucose indicates that the predicting model is using these spectral features in addition to glucose features for prediction. Such a situation diminishes the robustness of the predictive model as changes in the concentration of analytes other than glucose become predictive for glucose. This situation indicates that systematic variations in the concentration of these other analytes with respect to glucose have been modeled, with the source of this systematic variation being the inherent analyte cross-correlation of the fermentation system. It should be noted that the spectral features seen are of opposite sign to the glucose features, indicating that the correlations are negative in nature.

To further illustrate the existence of analyte cross-correlation, the PLS model from data set A was used to predict for the synthetic data set (which contains no analyte cross-correlations). The predicted versus measured plot is shown in Fig. 5. The prediction for glucose at factor 3 showed quite a poor prediction, with a very large vertical spread of predicted values for the "high" and "low" glucose concentrations (see Table 2) of the synthetic data set. A close examination of the spread of the predictions results reveals a symmetric pattern around the center of the spread. Cross-referencing of these samples to their actual analyte concentration revealed that samples at the extremes of the spread of predicted results were a function of the acetone and butanol concentrations. For the prediction of both low and high glucose concentrations, those synthetic samples that were low in acetone and butanol were predicted the highest. Conversely, samples high in acetone and butanol were predicted the lowest. This is easily rationalized when reference is made to Fig. 3, which contained the glucose regression coefficient plot for factor 3 of data set A, whose model was used to predict for the synthetic data set. The spectral features of acetone and butanol

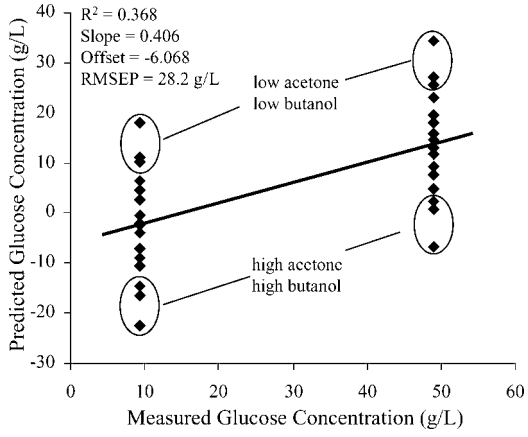

Figure 5. Prediction results of data set "A": glucose at three factors, predicting for data set "syn."

among others were identified, in addition to those of glucose. Therefore, the concentration of the analytes whose spectral features have been modeled as being predictive for glucose have caused systemic interferences in the prediction of glucose for the synthetic data set. Some of the other "structure" seen in the vertical spread is due to various combinations of the other analyte concentrations (acetate and butyrate), which also showed some representation in the glucose regression coefficients factor 3 plot. The effect of these other analytes is less pronounced as their correlation with glucose is much less than either acetone or butanol.

Despite the obvious presence of other analyte features, when a traditional leave-out-one cross-validation is performed for the PLS model of data set A, at three factors for glucose, a plot of predicted versus measured produces relatively very good results ($R^2 = 0.972$, slope $= 0.974$, offset $= 1.02$, and RMSEP $= 3.25$). Therefore, if the common method of choosing the "optimal" number of factors at or just before the point of minimum RMSEP is followed exclusively as was done in this example, then nonoptimal results are possible. What may appear to be very good results on the surface based solely on a simple leave-out-one cross-validation on the one data set could in fact be due to the use of spectral features other than those of the analyte of interest.

The presence of analyte cross-correlation (negative and/or positive) in the PLS models leads to changes in a certain analyte being predictive for changes in other analytes, such as decrease in acetone or butanol absorbances being predictive for an increase in glucose, or conversely, an increase in glucose absorbances being predictive for a decrease in acetone or glucose. This in fact would hold true if the fermentation were to behave in exactly

the same manner each time and the ratios between the analytes remained constant. However due to the unpredictable nature of fermentations, it is unlikely that the conditions and performance of the fermentation would always be identical.

Such problems reduce the robustness of a model, because unusual changes in the concentration of a particular analyte, such as acetone or glucose, which might not necessarily mean a proportional change in glucose, would be interpreted as a change in the concentration of this component and hence give erroneous values.

Synthetic Calibration

An appropriate way to remove these cross-correlations among analytes is to calibrate the PLS model using data with no cross-correlations. This is achieved by creating a synthetic sample set. A PLS model on this data set (data set Syn) not only produced excellent results from the standard leave-out-one cross-validation predictions, but the spectral regression coefficient plots for each analyte (at the minimum prediction error) resembled very closely those of the pure component spectra for each analyte, and most importantly they did not contain features that could be attributed to the other analytes.

Figure 3 shows the comparisons between the spectral regression coefficients and the pure component spectra for each analyte. Boxed areas highlight those spectral regions defined in Table 3 as analyte-specific spectral windows (SW). These spectral regression coefficients plots show that the PLS regression model for a given analyte has been able to find, from a complex mixture of analytes and media background, the spectral regions where each analyte of interest has its specific absorptions and has been able to incorporate these into the predictive model for that, and only that, analyte. This then allows for the formation of much more robust models, models that are independent of other analyte correlations and are hence applicable to a wider set of conditions. In addition to the excellent results from the leave-out-one full cross-validation, the prediction errors drop off dramatically after five factors, before rising very slightly in the following factors as would be expected for a system in which there are only five varying analytes.

Predictions

Various combinations of "real" (from both "normal," "acid crash," and "normal + acid crash," see Table 1) and synthetic data were used to predict for each other and vice versa leading to a total of eight predictions, the

results of which are given in Table 5. The existence of analyte cross-correlations in PLS regression models and the subsequent reduction in model robustness can be further illustrated by examining Table 5.

Three sets of prediction results are presented, two of which are different from methods of PLS model factor selection and the third set being the actual optimum results or "best-case scenario," determined "after the fact" by analysis of the actual prediction results at each factor to locate the "actual" prediction error minimum. The two different factor selection methods consist of one already discussed, the "minimum error from error vs. factor plot," which is a commonly used method, and the second, a new and alternative approach termed the "similarity measure" (SM), is a method where the similarity of the spectral regression coefficients plots at each factor for each analyte is compared and calculated to that of the analyte standard spectra. The similarity measure was calculated by normalizing both regression coefficients spectra and standard spectra to the most intense band and determining the sum of the squared differences at each absorbance value. This was calculated for all factors, with the factor having the lowest sum of squared difference being chosen as "optimal." This method ensures that factors with spectral regressions coefficient plots that most closely matched those standard spectra are chosen, which in theory should lead to more accurate and robust predictions. Table 6 lists the similarity measures for all models.

The data in Table 6 reinforce the notion that by selecting spectral regions specific to the analyte of interest, significant improvements can be made. We can see that the overall average and in fact for each analyte of each data set, improvements in the "similarity measure" are observed in the use of SW over SR1. This effect is most pronounced for analytes that have strong and relatively uniquely positioned absorbances, such as acetone and glucose. Together with such improvements has come significant reductions in the number of factors required to obtain these minima. This indicates that the removal of the spectral regions deemed to have no information has in fact led to a data set containing more relevant information, which therefore requires fewer factors to model.

A comparison within Table 6 was also made with the "similarity measures" calculated for factors determined to be the minimum from an "error vs. factor" plot. It can be clearly seen that for factors deemed to be the minimum from an "error vs. factor" plot, they are far less closely matched to the analyte standard spectra than for factors chosen by the "similarity measure." In terms of the effect on the end prediction quality, it can be seen in Table 5 that some improvement can be seen for the results from factor determination via the "similarity measure" over the factor determination via "error vs. factor" plot. Additionally, the fewer factors required reduce the risk of "over-fitting."

A graphical comparison in made in Fig. 6, showing the regression coefficients plot for glucose at factor 3, determined to be the "optimum" via the minimum in the "error vs. factor" plot and at factor 1, determined to be the

Table 5. Summary of prediction results

		Minimum error from factor vs. error plot		Similarity measure		Actual minima	
Prediction number	Predictor–predictee	Average factors	ARME (%)	Average factors	ARME (%)	Average factors	ARME (%)
1	C–Syn						
	SW	6	32	3	32	6	25
	SR1	13	59	9	52	14	27
2	Syn–C						
	SW	6	14	2	15	7	9
	SR1	5	13	5	13	6	10
3	A–Syn						
	SW	10	44	2	37	7	29
	SR1	13	72	7	59	13	46
4	Syn–A						
	SW	6	12	2	14	7	7
	SR1	5	13	5	13	8	8

5	B–Syn						
		SW	4	47	3	56	4
40							
	SR1	13	93	5	102	7	78
6	Syn–B						
	SW	6	17	2	17	7	13
	SR1	5	16	5	16	9	14
7	B–A						
	SW	4	16	3	16	4	15
	SR1	13	24	5	29	8	20
8	A–B						
	SW	10	28	2	25	8	15
	SR1	13	36	7	40	11	25

SW, spectral window; SR1, spectral region 1; ARME, average relative model error.

Table 6. Summary of similarity measure (SM) between spectral regression coefficient plots and standard spectra and comparison of SM at the factor minimum from a "error vs. factor" plot

Data set	Average (SM) (factor)		Acetone (factor)		Acetate (factor)		Butanol (factor)		Butyrate (factor)		Glucose (factor)	
	Error vs. factor	SM	Error vs. factor	SM	Error vs. factor	SM	Error vs. factor	SM	Error vs. factor	SM	Error vs. factor	SM
A SR1	98 (13)	41 (7)	49 (15)	24 (8)	161 (13)	71 (6)	111 (15)	31 (6)	87 (19)	55 (13)	82 (3)	23 (1)
A SW	40 (5)	16 (2)	90 (12)	5 (1)	33 (2)	30 (1)	22 (2)	18 (3)	25 (5)	25 (3)	28 (2)	3 (1)
B SR1	99 (13)	45 (5)	86 (13)	42 (9)	120 (13)	49 (3)	102 (13)	56 (8)	58 (13)	57 (5)	131 (13)	23 (2)
B SW	17 (4)	15 (3)	4 (3)	4 (4)	16 (2)	16 (2)	29 (6)	28 (5)	28 (5)	22 (4)	10 (2)	3 (1)
C SR1	85 (14)	27 (9)	58 (16)	19 (9)	47 (14)	26 (9)	63 (10)	34 (9)	79 (20)	32 (17)	177 (8)	22 (2)
C SW	28 (7)	11 (3)	13 (3)	4 (1)	60 (15)	9 (4)	21 (5)	21 (5)	30 (7)	17 (5)	15 (3)	3 (1)
Syn SR1	5 (5)	5 (5)	2 (5)	2 (5)	4 (5)	4 (5)	3 (5)	3 (5)	5 (5)	5 (5)	11 (5)	11 (5)
Syn SW	33 (7)	4 (2)	3 (5)	1 (1)	2 (5)	2 (2)	64 (8)	2 (3)	6 (5)	3 (2)	90 (9)	10 (1)

Note: Lower values for the similarity measure denote a better match. SR1, spectral region 1; SW, spectral window.

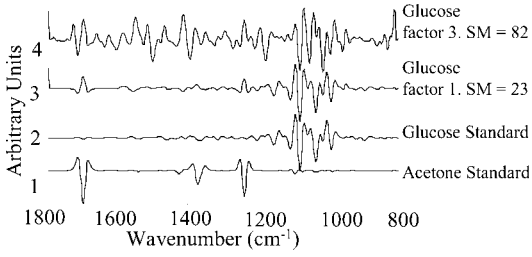

Figure 6. Spectral regression coefficient plot of glucose from data set "A" at three factors (where a minimum in the error vs. factor plot is reached) and at one factor (which has the best "similarity measure") compared to glucose and acetone spectral standards.

minimum from the newly defined "similarity measure." Included in Fig. 6 are the standard spectra of both glucose and acetone. As was seen in Fig. 3, there are many spectral features for the regression coefficients plot of glucose at factor 3, most of which are not due to glucose itself but to other analytes. However, when examining the regression coefficients plot for glucose at factor 1, the "optimum" factor as determined by the "similarity measure," very distinct glucose features are clearly visible. In addition, distinct acetone features are also apparent. As previously mentioned, this is due to the strong correlation that exists between these two analytes. The "similarity measure" is included within Fig. 6 with factor 3 at SM = 82 and factor 1 at SM = 23, giving a numerical interpretation of the visual comparison. The net effect of putting the two different methods of factor selection to the test is displayed in Fig. 7.

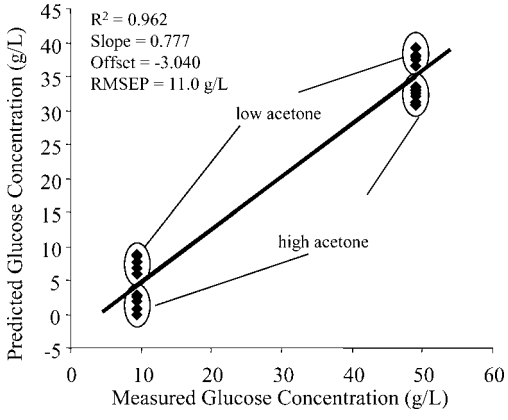

Figure 7. Prediction results of data set "A": glucose at one factor, predicting for data set "syn."

It can be seen that when the common method of factor selection (the minimum in an "error vs. factor" plot of cross-validated PLS model) was used and factor 3 was selected as being the "optimum," then rather poor prediction results for the synthetic data would be obtained with the R^2, slope, offset, and RMSEP at 0.368, 0.406, −6.068, and 28.2 g/L, respectively (see Fig. 5). On the other hand, if the factor selection from the cross-validated PLS model was performed using the "similarity measure" between all the regression coefficient plots for all factors and the standard spectrum for glucose, then the "optimum" would have been determined as being factor 1. The prediction results for glucose now using factor 1 are significantly improved with the R^2, slope, offset, and RMSEP now at 0.962, 0.777, −3.040, and 11.0 g/L. Splitting of the results into "high" and "low" groups is again observed, as discussed previously, but this time it is only a function of the acetone levels within the synthetic samples.

Examining prediction no. 3 (A−Syn), where data set A is predicting for the synthetic data set, we see that when the spectral region 1 (SR1) is used, we obtain a relatively poor result for the ARME of 72% with an average number of factors across all five analytes at 13. Taking the same prediction, but using spectral windows as defined in Table 3, we find an improvement in ARME to 44% and a reduction in the average number of factors to 10. Despite this improvement, this is still a relatively poor prediction. Data set A comprises data from three separate "normal" fermentations (see Table 1). Although being from three separate fermentations, because all are "normal," the same approximate correlations (ratio) between the various analytes still exists. The same prediction conducted with spectral window selection greatly improved the results, by not only removing spectral regions contributing noise but by also removing spectral regions containing analyte cross-correlations that lead to inaccurate predictions. A similar trend is observed for prediction no. 5 (B−Syn) with relatively high ARME and significant improvement with the use of spectral windows. The predictor data set in prediction no. 5, data set B, comprises data from two separate "acid-crash" fermentations, where correlations, though now different to those of "normal" fermentations as in data set A, still exist. When both data sets are combined to form data set C and are used to predict for the synthetic data set, as in prediction no. 1, we see improvements in the ARME for the spectral region 1 and spectral windows to 59% and 32% respectively. These improvements, albeit small, show that the combination of the data sets from the "normal" and "acid-crash" fermentations has created more robust models, both by increasing the number of samples, broadening the range of analyte concentration, and, most importantly, by reducing the effect of analyte cross-correlations. The observation of a reduction in the analyte cross-correlation effect is best observed by examining the actual minima "after the fact" data as opposed to the "minimum error from factor vs. error plot." Here the ARME is practically the same with either SR1 or SW. This is in contrast to predictions no. 3 and no. 5 where we see significant differences

in the ARME between SR1 and SW for both actual minima and "minimum error from factor vs. error plot." The fact that the ARMEs for both SR1 and SW are similar indicates that a reduction in the analyte cross-correlations has been achieved in the larger spectral range of SR1, hence using targeted spectral window (SW), which would have otherwise removed the effects of any analyte cross-correlations, have made very little difference. This effect is achieved by using a PLS model derived from data sets that, although individually containing their own analyte cross-correlations, when combined the net effect of the differing analyte cross-correlations is nullified.

The actual minima are a measurement of best possible predictions, the true optimum. Differences between the ARMEs of "actual minima" versus those of factor selection via "minimum error from factor vs. error plot," or "similarity measure," highlight the shortcomings of these methods. ARMEs calculated from the actual minima are generally significantly lower than for those calculated from the other methods of factor selection. It is also interesting to note that, although in general the ARMEs from either factor selection, "minimum error from factor vs. error plot" or "similarity measure" are not appreciably different from one another, the average number of factors for the overall model is much lower for the latter. The use of fewer factors is preferred, as this reduces the likelihood of "over-fitting," which could further reduce accuracy and robustness by potentially including in the model noise and other spectral contributions not related to the analyte of interest.

When the synthetic data are used to predict for the real data sets A, B, and C in predictions no. 4, 6, and 2, respectively, we see that the ARMEs are dramatically decreased to 12–17% from as high as 32–93%. This dramatic increase in accuracy when predicting for three very different data sets can be directly attributed to the fact that the PLS model derived from the synthetic data is much more robust than those derived from the "real" fermentation data. The primary reason being the lack of analyte cross-correlations, and the effect of this can be observed in the spectral regression coefficient plots, where all are matched very closely to the standard spectra (see Fig. 3). As for prediction no. 1 with data set C, there is also very little difference between ARMEs from SR1 and SW indicating again that there are no influences from analyte cross-correlations. Additionally, for the synthetic PLS models, the differences between the ARMEs when calculated via the two factor selection methods "minimum error from factor vs. error plot" and "actual minima" are significantly less, indicating that both methods produce results close to the "best-case scenario." One difference between the factor selection methods is as noted above, that the "similarity measure" requires significantly fewer factors.

In order to ensure appropriate factor selection, one should not simply rely on the "minimum error from the error vs. factor plot" of cross-validated PLS model. Some form of a "similarity measure" between the regression coefficient plots and the standard spectra of the analyte(s) under investigation,

such as that described here, should be conducted. In its most simple form, a basic visual comparison between the regression coefficient plots and standard analyte spectra will aid in determining the presence of spectral features other than those of interest.

CONCLUSIONS

This study has demonstrated the usefulness of FTIR spectroscopy coupled to SIA with PLS regression analysis for the simultaneous on-line measurement of acetone, acetate, *n*-butanol, butyrate, and glucose in an ABE fermentation under various conditions.

Employing a PLS regression model built from a synthetic calibration set, analyte cross-correlations that are present in PLS regression models built using only the reference values have been removed. Synthetic calibration set models had lower prediction residuals than those predictions using "real" data and required fewer factors. By careful examination of the spectral regression coefficients plots, the presence of analyte cross-correlations were recognized. Furthermore, by studying the spectral regression coefficient plots and identifying analyte cross-correlations by the presence of features other than those belonging to the analyte of interest, the optimal selection of factors is possible and over-fitting is avoided.

The importance of selecting the factors whose regression coefficients plots most resemble those of the analyte standard spectra have been demonstrated and two different methods of "optimum" factor selection compared. The results of this comparison have shown that by simply following the common methods of factor selection, using the minimum in an "error vs. factor" plot of cross-validated PLS model, that the cross-correlation of analytes is incorporated into the PLS model. This is evident in the appearance of spectral features of other analytes, analytes that are correlated to analyte of interest. We have demonstrated that improvements in the prediction results and a reduction of number of "optimum" factors can be achieved when factors are chosen more appropriately such that the regression coefficients plots at these factors more closely resemble those of the analytes of interest. The use of the "similarity measure" has been able to facilitate this more accurate factor selection.

In the instance where the formation of synthetic data sets is not possible, we have shown that by ensuring the careful selection of factors via appropriate factor selection methods such as the "similarity measure" together with the use of spectral windows, that model robustness through the incorporation of truly relevant spectral features has been improved. By taking this approach, we have built more robust PLS regression models that are applicable to a broader range of fermentation conditions and synthetic samples and are not only valid for a certain set of ideal process conditions.

This method has the added advantages of being rapid, with up to 12 fermentation samples being analyzed per hour, (which is equivalent to

60 analyte determinations), computer control, ease of automation, and minimal reagent and sample consumption (1.5 mL), therefore generating minimal waste.

The higher rates of sampling compared to conventional methods will enable closer monitoring of the fermentation, which facilitates and allows for the optimization of the fermentation so that it may be kept operational.

The adaptability and flexibility of the SIA system and the ability of the ATR sampling technique to handle difficult and varied samples makes this technique suitable for monitoring other fermentations and industrial processes where rapid on-line measurement of one or more analytes is required.

ACKNOWLEDGMENTS

The authors wish to acknowledge the Australian Research Council, the Monash Graduate Scholarship, and the Austrian Science Fund for support of this work within project no. 13686.

REFERENCES

1. Wong, J. S.; Rein, A. J.; Wilks, D. Infrared spectroscopy of aqueous antibiotic solutions. *Appl. Spectrosc.* **1984**, *38* (1), 32–35.
2. Keuhl, D.; Crocombe, R. The quantitative analysis of a model fermentation broth. *Appl. Spectrosc.* **1984**, *38* (6), 907–909.
3. Alberti, J. C.; Phillips, J. A. Off-line monitoring of fermentation samples by FTIR/ATR: a feasibility study for real-time process control. *Biotechnol. Bioeng.* **1985**, *15*, 689–722.
4. White, R. L.; Roberts, D. E.; Attridge, M. C. Fourier transform infrared detection of pyruvic acid assimilation by E. coli. *Anal. Chem.* **1985**, *57*, 2487–2491.
5. Hagman, A.; Sivertsson, P. The use of NIR spectroscopy in monitoring and controlling bioprocesses. *Process Control Quality* **1998**, *11* (2), 125–128.
6. McShane, M. J.; Cote, G. L. Near-infrared spectroscopy for determination of glucose, lactate, and ammonia in cell culture Media. *Appl. Spectrosc.* **1998**, *52* (8), 1073–1078.
7. Riley, M. R.; Arnold, M. A.; Murhammer, D. W.; Walls, E. L.; DelaCruz, N. Adaptive calibration scheme for quantification of nutrients and byproducts in insect cell bioreactors by near-infrared spectroscopy. *Biotechnol. Progr.* **1998**, *14*, 527–533.
8. Riley, M. R.; Crider, H. M.; Nite, M. E.; Garcia, R. A.; Woo, J.; Wegge, R. M. Simultaneous measurement of 19 components in serum-containing animal cell culture media by Fourier transform near-infrared spectroscopy. *Biotechnol. Progr.* **2001**, *17*, 376–378.
9. Gallignani, M.; Garrigues, S.; Delaguardia, M. Direct determination of ethanol in all types of alcoholic beverages by near-infrared derivative spectrometry. *Analyst* **1993**, *118* (9), 1167–1173.

10. Hammouri, M. K.; Ereifej, K. I.; Shibli, R. A.; Al-Karaki, G. N. Quantitative analysis of fructose fate in a plant fermentation system. *J. Agric. Food Chem.* **1998**, *46* (4), 1428–1432.
11. Bellon, V. Fermentation control using ATR and an FT-IR spectrometer. *Sensors Actuators B* **1993**, *12*, 57–64.
12. Doak, D. L.; Phillips, J. A. In situ monitoring of an Escherichia coli fermentation using a diamond composition ATR probe and mid-infrared spectroscopy. *Biotechnol. Progr.* **1999**, *15* (3), 529–539.
13. Fairbrother, P.; George, W. O.; Williams, J. M. Whey fermentation: on-line analysis of lactose and lactic acid by FTIR spectroscopy. *Appl. Microbiol. Biotechnol.* **1991**, *35*, 301–305.
14. Fayolle, P.; Picque, D.; Perret, B.; Latrille, E.; Corrieu, G. Determination of major compounds of alcoholic fermentation by middle-infrared spectroscopy—study of temperature effects and calibration methods. *Appl. Spectrosc.* **1996**, *50* (10), 1325–1330.
15. Fayolle, P.; Picque, D.; Corrieu, G. Monitoring of fermentation processes producing lactic acid bacteria by mid-infrared spectroscopy. *Vibrat. Spectrosc.* **1997**, *14* (2), 247–252.
16. Gallignani, M.; Garrigues, S.; Delaguardia, M. Derivative Fourier transform infrared spectrometric determination of ethanol in alcoholic beverages. *Anal. Chim. Acta* **1994**, *287* (3), 275–283.
17. Kansiz, M.; Schuster, K. C.; Gapes, J. R.; McNaughton, D.; Lendl, B. Mid-infrared spectroscopy coupled to sequential injection analysis for the on-line monitoring of the acetone-butanol-ethanol fermentation Process. *Anal. Chim. Acta* **2001**, *438*, 175–186.
18. Durre, P. New insights and novel developments in Clostridial acetone/butanol/isopropanol fermentation. *Appl. Microbiol. Biotechnol.* **1998**, *49*, 639–648.
19. Jones, D. T.; Wood, D. R. Acetone-Butanol Fermentation Revisited. *Microbiol. Rev.* **1986**, *50* (4), 484–524.
20. Schuster, K. C. Applied acetone-butanol fermentation. *J. Mol. Microbiol. Biotechnol.* **2000**, *2*, 3–4.
21. Schuster, K. C.; van den Heuvel, R.; Gutierrez, N. A.; Maddox, I. S. Development of markers for product formation and cell cycle in batch cultivation of Clostridium acetobutylicum ATCC 824. *Appl. Microbiol. Biotechnol.* **1998**, *49*, 669.
22. Maddox, I. S.; Steiner, E.; Hirsh, S.; Wessner, S.; Gutierrez, N. A.; Gapes, J. R.; Schuster, K. C. The cause of "acid crash" and "acidogenic fermentations" during the batch acetone-butanol-ethanol (ABE) fermentation process. *J. Mol. Microbiol. Biotechnol.* **2000**, *2* (1), 95–100.
23. Thompson, R. N. Quantification of the end-products of the acetone-butanol-ethanol fermentation by gas chromatography. *Enzyme Microbial Technol.* **1991**, *13*, 722–726.
24. *The Unscrambler User Manual*; version 6.11, CAMO Process, Oslo, Norway, 1998.
25. Kansiz, M.; Billman-Jacobe, H.; McNaughton, D. Quantitative determination of the biodegradable polymer poly(beta-hydroxybutyrate) in a recombinant Escherichia coli strain by use of mid-infrared spectroscopy and multivariative statistics. *Appl. Environ. Microbiol.* **2000**, *66* (8), 3415–3420.
26. Kansiz, M.; Heraud, P.; Wood, B.; Burden, F.; Beardall, J.; McNaughton, D. Fourier transform infrared microspectroscopy and chemometrics as a tool for the discrimination of cyanobacterial strains. *Phytochemistry* **1999**, *52*, 407–417.