

# Nondestructive Near-Infrared Spectroscopic Measurement of Multiple Analytes in Undiluted Samples of Serum-Based Cell Culture Media

Martin Rhiel,<sup>1\*</sup> Michael B. Cohen,<sup>2</sup> David W. Murhammer,<sup>1</sup> Mark A. Arnold<sup>3</sup>

<sup>1</sup>Department of Chemical and Biochemical Engineering, University of Iowa, Iowa City, Iowa

<sup>2</sup>Department of Pathology, University of Iowa, Iowa City, Iowa

<sup>3</sup>Department of Chemistry, University of Iowa, Iowa City, Iowa 52242; telephone: 319-335-1368; fax: 319-353-1115; e-mail: mark-arnold@uiowa.edu

Received 14 January 2001; accepted 9 August 2001

**Abstract:** An adaptive calibration procedure is used to build selective multivariate calibration models for the measurement of glucose, lactate, glutamine, and ammonia in undiluted serum-based cell culture media. This adaptive procedure removes metabolism-induced covariance between these analytes in a series of calibration samples collected during the cultivation of PC-3 human prostate cancer cells. Partial least-squares calibration models are generated from single-beam near-infrared (NIR) spectra collected over the 4800- to 4200-cm<sup>-1</sup> combination spectral range. Calibration models were generated with both the full spectral range and optimized spectral ranges. In both cases, the number of model factors was optimized and model validity was determined by comparing analyte concentrations predicted from a series of independent and unaltered samples that were obtained during a subsequent cultivation of the PC-3 cells. Similar analytical performance was achieved with fewer model factors when the optimized spectral range was used. The lowest standard errors of prediction were 0.82, 0.94, 0.55, and 0.76 mM for glucose, lactate, glutamine, and ammonia, respectively. Different spectral ranges were optimal for each analyte and the optimized spectral range coincided with the distinguishing spectral features of the analyte. The results of this study demonstrate that NIR spectroscopy can be used effectively in the off-line measurement of important nutrients (glucose and glutamine) and byproducts (lactate and ammonia) in a serum-based animal cell culture medium. © 2002 John Wiley & Sons, Inc. *Biotechnol Bioeng* 77: 73–82, 2002.

**Keywords:** near-infrared spectroscopy (NIR); bioreactor-monitoring; mammalian cell culture

## INTRODUCTION

Mammalian cell cultures are widely used to produce valuable pharmaceuticals and to study *in vivo* behavior.

Correspondence to: M. A. Arnold

\*Current address: Process Development, Cytos Biotechnology AG, Zurich-Schlieren, Switzerland

Contract grant sponsor: National Aeronautics and Space Administration

Contract grant number: NAG 9-824

The chemical environment must be controlled to maximize productivity and mimic *in vivo* conditions, respectively. The measurement of key analytes can be used to assess the state of the cultivation and this information can be used to initiate the appropriate control actions. An example is the control of feed levels of glucose and glutamine in order to reduce the generation of toxic waste products (Glacken et al., 1986). Several types of on-line sensors and on-line sensing schemes have been reported for these analytes, and have been reviewed elsewhere (Schügerl, 1991). The widespread application of *on-line* sensors for critical analytes, such as glucose, glutamine, and ammonia, is limited, however, due to problems with long-term stability and complications associated with sterilization.

Near-infrared (NIR) spectroscopy offers a way to overcome the aforementioned limitations of conventional chemical sensing technology. With NIR spectroscopy, analytical information can be collected without directly exposing the sample to the active surface of a sensing element. The chemical information is obtained by analyzing a spectrum that is collected by transmitting a band of NIR radiation through the sample of interest. This measurement scheme is noninvasive and, as a result, issues of sterility are irrelevant. Without a dependency on chemical reagents, NIR spectroscopy calibration models can be stable for extended periods, although periodic model verification is required to compensate for both sample and instrumentation variations. In addition, multiple components can be monitored simultaneously from a single NIR spectrum, thereby eliminating the need for multiple sensors.

The analytical utility of NIR spectroscopy is established for measuring millimolar levels of various biological species in simulated matrices (Arnold et al., 1998; Burmeister et al., 1998; Burmeister and Arnold, 1999) and natural matrices from clinical chemistry (Hazen et al., 1998), fermentation (Cavinato et al., 1990; Ge

et al., 1994; Hall et al., 1996; Yano et al., 1998), and cell culture (Chung et al., 1995; Lewis et al., 2000; McShane and Coté, 1998; Riley et al., 1997, 1998a, 1998b; Zhou et al., 1995) systems. The chemical information originates from overtone and combination transitions associated with fundamental C–H, O–H, and N–H vibrations. The challenge in applying NIR spectroscopy for bioreactor monitoring is to accurately extract analyte specific information from complex spectra composed of highly overlapping and weak spectral features. This task is particularly difficult when the analytes of interest are contained in mixtures with similar chemical species present at similar concentrations. Additional complexity is introduced if some components in the mixture are not defined, as is typically the case for serum-based cell culture media. Multivariate calibration methods, such as principal component analysis (PCA) and partial least-squares (PLS) regression, can selectively extract analytical information from such overlapping spectra (Martens and Naes, 1989; Sjöstrom et al., 1983; Wold et al., 1990).

Successful multivariate calibration models require that all matrix and analyte variation to be encountered in subsequent samples must be represented within the training data set. In addition, proper calibration models demand that analyte concentrations are not correlated with concentrations of any other species within the chemical matrix (ASTM Practice E 1655-94). Samples collected during cell cultivation will naturally possess a metabolism-induced concentration correlation between cellular substrates and metabolic waste products (Riley et al., 1998a). Indeed, the well-known relationship between glucose consumption and lactate production generates a concentration correlation that is unacceptable for building PLS calibration models. The implications of cell metabolism on method calibration have been examined in detail elsewhere (Lewis et al., 2000; McShane and Coté, 1998; Riley et al., 1997, 1998a, 1998b). Briefly, the calibration samples are best collected at various timepoints during cultivation in order to account for the global chemical changes in the samples. In addition, cellular metabolism creates a concentration correlation between the nutrients and metabolic waste products. For the PC-3 human prostate cancer cells used in this work, glucose and glutamine are consumed, whereas lactate and ammonia are produced. Metabolism-induced concentration correlations are created within the samples where glucose and lactate are inversely correlated, glutamine and ammonia are inversely correlated, glucose and glutamine are directly correlated, and lactate and ammonia are directly correlated. The existence of such correlations within the training data set prohibits the generation of an analyte specific calibration model (Arnold et al., 1998; Lewis et al., 2000). With concentration correlations in the training data set, accurate analyte measurements cannot be performed independently, but depend on the concentrations of these other components.

An adaptive calibration procedure has been proposed to destroy concentration correlations within a set of samples (Riley et al., 1997, 1998a, 1998b). This procedure involves spiking samples with known and random amounts of standard materials, thereby adapting the collected samples so they may be used to generate NIR spectra suitable for training multivariate calibration models.

The objective of this study was to establish the utility of coupling the adaptive calibration method with NIR spectroscopy for measuring multiple components in a serum-based cell culture medium used in the growth of human prostate cancer cells. Samples collected during typical cell cultivation conditions were adapted for calibration purposes. A second set of samples was collected from a series of subsequent cell cultivations in order to assess the validity of the resulting PLS calibration models. Analytical performance was judged by the accuracy of predictions for glucose, lactate, glutamine, and ammonia concentrations in this independent set of validation samples.

## EXPERIMENTAL

### Cell Line and Culture Medium

PC-3 human prostate cancer cells (ATCC CRL 1435) were used in all experiments. Cells were grown in RPMI-1640 medium (Gibco, Grand Island, NY) supplemented with 10% fetal bovine serum (FBS; Gibco), 2 mM glutamine, 0.1 mM 2-mercaptoethanol, 100 mg/mL streptomycin, and 100 U/mL penicillin (Rokhlin and Cohen, 1995). Cell cultivation was performed in either 75- or 150-cm<sup>2</sup> tissue-culture flasks with 10- and 15-mL volumes of growth medium, respectively.

### Sample Collection and Preparation of Adaptive Standards for Calibration

Growth medium samples collected from the 150-cm<sup>2</sup> flasks were adapted for calibration purposes, whereas those obtained from the 75-cm<sup>2</sup> flasks were assayed directly for prediction and model validation. Initially, a 75-mL volume of fresh medium was taken and stored at 0° to 4°C. Subsequently, fresh medium and cells were added to 15 separate 150-cm<sup>2</sup> flasks and the cells were allowed to grow under normal conditions (37°C, 90% relative humidity, and 20% oxygen). After 24 h, the medium supernatant was collected from five flasks and the resulting ca. 75 mL of cell-free medium was stored at 0° to 4°C. After an additional 24 h, a second 75-mL volume of growth medium was collected from another five flasks and stored in the same manner. The remaining five flasks generated a 75-mL volume of medium after another 48 h. This process provided four batches of growth medium corresponding to 0, 24, 48, and 96 h of

cultivation. The concentrations of glucose, lactate, glutamine, and ammonia were determined in each solution by the reference methods described in what follows. For the 0-, 24-, 48-, and 96-h samples, concentrations were approximately 9.8, 8.2, 5.1, and 2.0 mM for glucose; 1.7, 5.1, 9.8, and 12.9 mM for lactate; 3.5, 2.4, 1.6, and 0.4 mM for glutamine; and 0.7, 2.3, 3.8, and 4.5 mM for ammonia, respectively.

Calibration standards were generated from the four batches of growth medium described earlier. Each standard was prepared by adding a known and random amount of glucose, sodium lactate, glutamine (HCl salt), and ammonium chloride to a designated vial. A 4-mL sample from one of the four batches of growth medium was then added to the vial and the mixture was stirred to dissolve the added solutes. Once prepared, these standard solutions were stored at  $-20^{\circ}\text{C}$ . Approximately 17 standards were obtained from each batch of medium. Overall, 66 standard solutions were prepared. Analyte concentrations in each standard were calculated by adding the spiked amount to the endogenous concentration for that particular batch of medium. The addition of these solid standard materials did not alter the solution pH. As a result, no precautions were taken to account for or minimize pH variations between samples.

Samples for prediction and model validation were collected from 34 individual cell-cultivation runs. Fresh medium and cells were added to 34 individual 75-cm<sup>2</sup> tissue-culture flasks and the cells were permitted to grow under the same conditions as detailed for the standard solutions. Samples were collected by removing the medium supernatant, and these cell-free samples were stored at  $-20^{\circ}\text{C}$  until analysis. This procedure resulted in 34 samples taken at various timepoints during the cell growth cycle. Analyte concentrations were measured by the reference methods described in what follows.

## Reference Analyses

Glucose and lactate concentrations in cell culture samples were measured with a glucose/lactate analyzer (Model 2700, YSI, Yellow Springs, OH). The pooled standard deviation for all spiked tissue culture samples was 0.28 and 0.19 mM for glucose and lactate, respectively. This value for glucose is similar to the standard error previously reported for this method (Burmeister and Arnold, 1995).

Glutamine levels were determined by a high-performance liquid chromatography (HPLC) procedure adapted from the method of Jones and Gilligan (1983). Briefly, 50  $\mu\text{L}$  of 100-fold diluted aliquots of culture medium supernatant was mixed with 75  $\mu\text{L}$  of an *o*-phthalaldehyde/ $\beta$ -mercaptoethanol solution in borate buffer. A 10- $\mu\text{L}$  aliquot of the mixture was injected onto a Supelcosil 5- $\mu\text{m}$  particle size C<sub>18</sub> reverse-phase column (Supelco, Bellefonte, PA) equilibrated with a mobile phase composed of 85% of solution A and 15% meth-

anol. Solution A was composed of 0.1 M sodium acetate buffer, pH 7.2, with 0.5% THF. Separation was achieved by linearly increasing the methanol concentration to 43% (0 to 31 min), from 43% to 55% (31 to 32 min), from 55% to 73% (32 to 48 min), from 73% to 100% (48 to 55 min), and holding at 100% methanol for another 13 min. The flow rate was 1 mL/min at all times. The Shimadzu HPLC system (Shimadzu, Columbia, MD) was equipped with two LC10AS pumps, an on-line degasser, an SCL10A controller, an SIL10A autosampler, and EZCHROM software. Fluorescence of the amino acid derivatives was measured with a Spectra-Physics SP8410 fluorescence detector (Spectra-Physics, San Jose, CA) set for excitation at 370 nm and emission at 418 to 700 nm. Glutamine values were quantified using a linear calibration curve established with glutamine standards of known concentration. The standard error of calibration was 0.18 mM for these reference glutamine measurements.

Ammonia was measured with a solid-state diffuse reflectance-based optical sensor as described by Spear et al. (1998). Briefly, samples were injected as 10- $\mu\text{L}$  aliquots into a phosphate buffer (pH 7.2) carrier stream. The carrier stream was mixed with 0.1 M sodium hydroxide to increase the solution pH just before entering the sensing unit. Within the sensing unit, gaseous ammonia diffused into the pores of a microporous polypropylene membrane onto which chlorophenol red was adsorbed. Interaction between the basic ammonia and acidic indicator caused a change in the absorption properties of the chlorophenol red. Changes in absorbance were linearly related to the dissolved ammonia concentration by a calibration curve. The standard error of calibration was 0.16 mM for these reference ammonia measurements.

## Sample Spectra

Spectra were stored as single beams where the radiant power of the transmitted light was recorded as a function of spectral wavenumber. These single-beam spectra were collected with a Fourier transform infrared spectrometer (Model 550 Magna, Nicolet, Madison, WI) equipped with a 20-W tungsten-halogen source, a CaF<sub>2</sub> beam-splitter, a liquid-nitrogen-cooled InSb detector, and a multilayer optical interference filter (Barr Associates, Westford, MA) to isolate the range 5000 to 4000  $\text{cm}^{-1}$ . Spectra were collected through a thermostated transmission cell (Wilmad, Buena, NJ) adjusted to 1.5-mm pathlength. Samples were placed into the transmission cell and allowed to thermally equilibrate at  $37^{\circ}\text{C}$  prior to collecting spectra. Spectra collection was automated by placing the spectrometer software commands of OMNIC 3.1a (Nicolet) under the control of a VISUAL BASIC 4.0 (Microsoft, Seattle, WA) program by using the MACRO/PRO package (Nicolet). Spectra were collected as 256 coadded 8k interferograms, which provided  $4\text{-cm}^{-1}$

point spacing. Spectra were collected in triplicate without moving the sample between spectra.

In total, 231 spectra were collected from 66 unique calibration samples. Triplicate spectra were collected once for 55 of these samples and twice for 11 of these samples. Similarly, 123 spectra were collected from the 34 validation samples. Triplicate spectra were collected once for 32 samples, twice for one of the samples, and three times for the remaining sample. In all cases, the repeated spectral collections were performed on different days. In addition, all replicate spectra obtained from a given sample were moved together into either the calibration or prediction data sets.

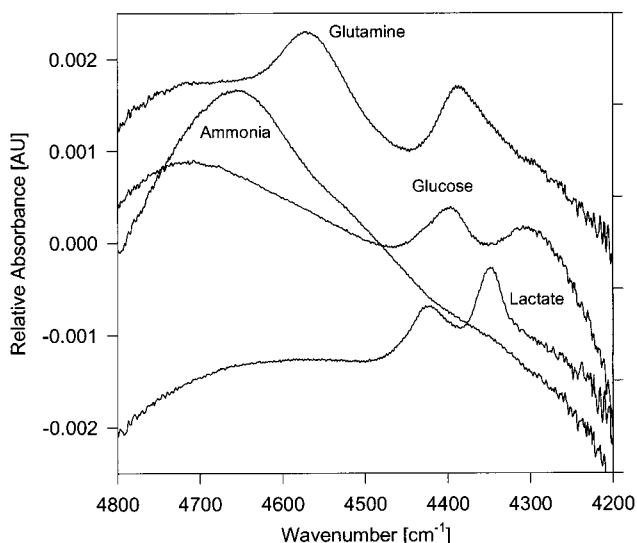
## Data Analysis

All calibration models were generated with single-beam spectra. Spectral processing was performed on a Silicon Graphics Iris Indigo computer with software supplied from Professor Gary W. Small, Center of Intelligent Chemical Instrumentation, Department of Chemistry, Ohio University (Athens, OH). All code for spectral processing was implemented in FORTRAN-77, with subroutines for PLS from an IMSL software package (IMSL, Inc., Houston, TX). The spectral range optimization algorithm was implemented in the Unix C-shell script (Topham, 1990).

## RESULTS AND DISCUSSION

### NIR Spectra

Unique spectral features are the basis for a selective, quantitative spectrum analysis. The spectral features of each analyte are presented in Figure 1 for comparison. Each plot in Figure 1 is the NIR absorbance spectrum of the pure component at a concentration of 10 mM in a



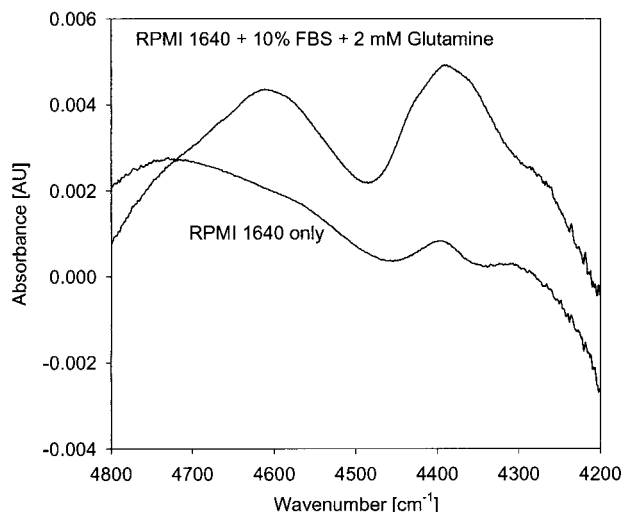
**Figure 1.** Superimposed absorbance spectra of pure components in phosphate/carbonate buffer at pH 7.4.

phosphate/carbonate buffer (pH 7.4) solution. The analyte specific information is contained within the shape and position of each absorption band. Glucose, for example, has three absorption bands centered at 4707, 4400, and 4309  $\text{cm}^{-1}$ . While these bands are unique, they nevertheless overlap with bands from several of the other analytes. In particular, the 4650- $\text{cm}^{-1}$  ammonia band overlaps with the 4707- $\text{cm}^{-1}$  glucose band, and the 4384- $\text{cm}^{-1}$  glutamine band strongly overlaps with the 4400- $\text{cm}^{-1}$  glucose band. Several distinct absorption bands are also evident in Figure 1 where overlap is minimal. Examples include the 4567  $\text{cm}^{-1}$  band for glutamine and the 4345  $\text{cm}^{-1}$  band for lactate. Despite the high degree of overlap, previous work has demonstrated the effectiveness of multivariate analysis techniques for selectively extracting analyte specific information from these types of spectra (Chung et al., 1996; Hazen et al., 1998; McShane and Coté, 1998; Riley et al., 1998b).

Actual culture media, however, are mixtures of many different components at different concentrations. Figure 2 shows the absorbance spectra of the basal cell culture medium RPMI-1640 and the complete cell culture medium RPMI-1640 after the addition of 10% fetal bovine serum (FBS), 2 mM glutamine, and the additives listed earlier (see Experimental section). The spectrum of RPMI-1640 alone appeared similar to the spectrum of glucose in buffer, which is reasonable because RPMI-1640 is composed of 10 mM glucose and submillimolar concentrations of most other constituents. When serum and additional glutamine were added, however, the spectrum changed dramatically as the absorbance features of the serum components masked those of the original components.

### Sample Matrix and Analyte Correlations

The PLS algorithm can selectively extract analyte specific information from highly overlapping NIR spectra.



**Figure 2.** Absorbance spectra of cell culture media before and after adding 10% fetal bovine serum and 2 mM glutamine.

Robustness of multivariate calibration models based on PLS regression demands that all sources of spectral variance (both of instrumental and sample origin) must be represented within the training data set. A principal source of sample variance for bioreactor monitoring is the chemical composition of the sample matrix. This matrix changes continuously during cell growth as nutrients are consumed and waste is generated. Metabolic processes create concentration correlations between glucose, lactate, glutamine, and ammonia. The extent of correlation is given by  $r^2$ -values between each component. The following  $r^2$ -values were obtained for the medium samples collected at 0, 24, 48, and 96 h of cultivation: 0.9839, 0.9816, 0.9449, 0.9812, 0.9845, and 0.9560 for the glucose/lactate, glucose/glutamine, glucose/ammonia, lactate/glutamine, lactate/ammonia, and glutamine/ammonia concentration correlations, respectively.

Concentration correlations can be destroyed within the calibration samples by spiking these samples with known and random amounts of the principal compounds. This spiking procedure was used to provide a nonsystematic relationship between the concentrations of glucose and each of the other analytes in the 66 prepared calibration solutions. Regression analysis revealed  $r^2$ -values of 0.0362, 0.0920, 0.0555, 0.0230, 0.0004, and 0.1049 for the glucose/lactate, glucose/glutamine, glucose/ammonia, lactate/glutamine, lactate/ammonia, and glutamine/ammonia concentration correlations, respectively. The effectiveness of this procedure was gauged by comparing  $r^2$ -values before and after spiking. In every case, the initial strong correlation was completely destroyed by the spiking process.

Although the random spiking procedure destroyed concentration correlations between sample analytes, this procedure resulted in a slightly distorted analyte distribution toward higher concentrations. This nonideal distribution was created by the fact that the concentration spike always adds to the basal analyte concentration. A potential consequence of this process is illustrated by the values presented in Table I, where the analyte statistics are tabulated for both the calibration and prediction data sets. By virtue of adding compo-

nents to the calibration samples, the mean concentrations were higher and the standard deviations were wider compared with the unaltered prediction samples. More importantly, the minimum concentration in the calibration samples can be higher than the minimum concentration in the prediction data set, which was true in this case for glucose, glutamine, and lactate. The inclusion of concentrations in the prediction data set that fall outside the concentration range used for model training is potentially unstable, because this condition requires that the model to extrapolate outside the trained environment. Accurate extrapolation is possible with PLS calibration models, but requires a stable and well-defined relationship between the multivariate parameters and analyte concentration.

Complications associated with this type of nonideal concentration distribution were less severe for the metabolic byproducts in comparison to the nutrients. The low concentration of byproducts during initial stages of the cultivation allowed the inclusion of low concentrations in the calibration data set. Nonideal concentration distribution is still a factor, however, because the concentrations of byproducts are both nonzero and variable in serum-based media. Our results for lactate demonstrate how concentration distribution issues are possible with metabolic byproducts.

## PLS Calibration Model Development

Model rank (or number of factors used in the PLS calibration model) and spectral range are important input parameters for the PLS analysis. Each PLS factor accounts for variation within the spectral data set and care must be taken to avoid overmodeling the data with too many factors. An accepted rule-of-thumb is that the calibration data set must contain five or six independent observations (e.g., spectra from unique calibration samples) for each PLS factor used in the final model (ASTM Practice E 1655-94). With 66 samples in the calibration data set, the PLS calibration models for this work were limited to 11 to 13 factors. Of course, fewer factors are desirable to enhance the robustness of model performance (Martens and Naes, 1989).

The optimum model rank was taken as the number of factors that produced the lowest statistically significant pooled standard error of monitoring (pSEM), as determined from an  $F$ -test comparison of pSEM values. For this analysis, the full calibration set of 231 spectra from 66 samples was randomly rearranged into an optimization training set (48 samples) and a monitoring set (18 samples). All spectra for a given sample were kept together and placed into either the training or monitoring data set. For a given number of factors and spectral range, the PLS calibration model was established with the optimization training data set and this model was then used to predict analyte levels in the respective

**Table I.** Analyte statistics for calibration and prediction data sets.<sup>a</sup>

Analyte	Data Set	Count	Min	Max	Mean	SD
Glucose	Calibration	66	0.51	20.30	10.50	4.00
	Prediction	34	0.07	10.07	4.67	3.34
Glutamine	Calibration	66	0.54	12.42	4.79	2.42
	Prediction	34	0.20	4.00	1.41	1.06
Lactate	Calibration	66	1.73	30.49	14.75	6.95
	Prediction	34	1.68	19.07	10.67	5.54
Ammonia	Calibration	66	0.33	13.14	6.29	2.95
	Prediction	34	0.58	5.10	3.32	1.35

<sup>a</sup>Minimum concentration (Min), maximum concentration (Max), mean concentration (Mean), and standard deviation of concentrations (SD) are given in millimolar units.

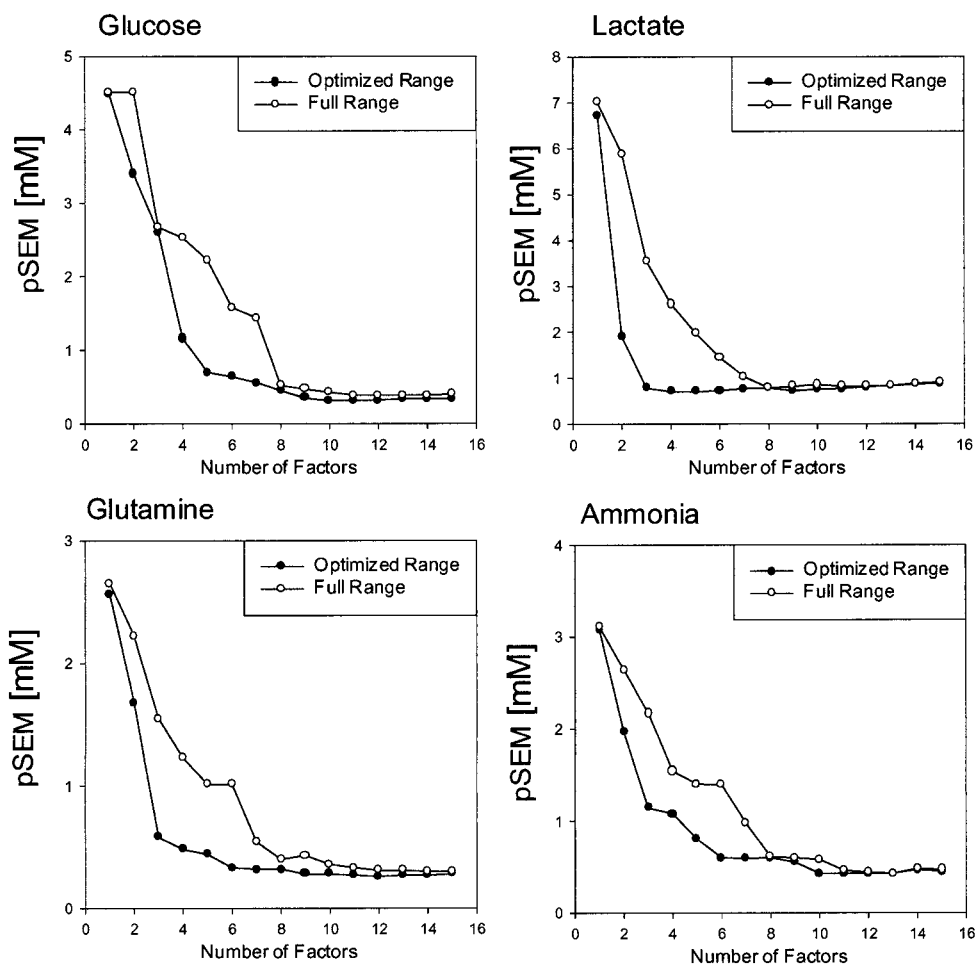
monitoring data set. Once the optimized conditions of spectral range and model rank were determined, the standard error of calibration (SEC) was computed on the basis of the full calibration set (i.e., the recombination of the optimization training and monitoring data sets). The predictive performance of each model is expressed as the standard error of prediction (SEP).

Model performance can be enhanced by limiting the PLS analysis to spectral regions where the analyte signal-to-noise ratio is highest. Selected spectral regions can enhance measurement selectivity by the exclusion of spectral variations that are not associated with the analyte of interest. A modified grid search was used to optimize the spectral range. Details of this method have been presented elsewhere (Rhiel, 1998). Briefly, PLS calibration models were initially evaluated for a series of broad sections of the full range (4800 to 4200  $\text{cm}^{-1}$ ) into smaller ranges with widths of 100, 200, 300, 400, and 500  $\text{cm}^{-1}$ . This procedure was repeated for each of the four splittings of the calibration data set (as described earlier) and performance was judged on the basis of the prediction error (pSEM). The “best” broad spectral range

was then fine-tuned by stepwise expanding and contracting this range in a systematic fashion with sequential wave-number steps of 50, 30, 20, and 10  $\text{cm}^{-1}$ . This entire procedure was executed independently for PLS model ranks from 1 to 15 factors. Other, more sophisticated, optimization algorithms are available to identify the ideal wavelengths for PLS analysis (Ding et al., 1998; Lucasius et al., 1994; Spiegelman et al., 1998).

### Calibration Model Performance

Performance of the PLS calibration models for glucose, lactate, glutamine, and ammonia are indicated in the plots presented in Figure 3. These models originate solely from an analysis of the spiked calibration samples. For each analyte, the pSEM is plotted as a function of model rank for calibration models based on the full spectral range and for models where the spectral range was optimized for each number of PLS factors. Relevant model parameters are provided in Tables II and III for the full spectral range and optimized spectral range analyses, respectively.



**Figure 3.** Effect of number of factors on the pooled prediction errors in the monitoring data sets (pSEM) for PLS calibration models generated with the full (○) and optimum (●) spectral ranges.

**Table II.** PLS calibration models with the full spectral range.<sup>a</sup>

Analyte	Spectral range (cm <sup>-1</sup> )	PLS factors	SEC (mM)	SEP (mM)
Glucose	4800 to 4200	8	0.40	1.43
Lactate	4800 to 4200	8	0.74	0.94
Glutamine	4800 to 4200	8	0.47	0.55
Ammonia	4800 to 4200	8	0.57	0.98

<sup>a</sup>Spiked standards used for calibration and unmodified samples used for prediction.

For both the full spectral range and the optimized spectral range models, the results in Figure 3 indicate a dramatic drop in the pSEM with the addition of the first few factors. The point where the impact of more factors no longer significantly lowers the pSEM represents the optimum number of factors. As noted earlier, a conventional *F*-test was used to compare pSEM values. For each analyte, an eight-factor model is optimal when the full spectral range is used. Fewer factors are necessary with the optimized spectral range. The need for fewer factors is consistent with a reduction in spectral variance for the narrower spectral ranges. For lactate and glutamine, in particular, the model rank drops to three factors when the optimum spectral range is used.

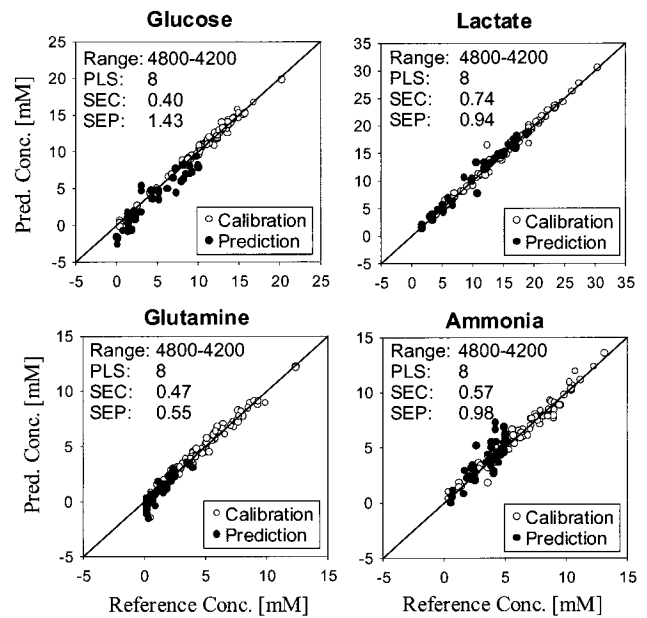
Each of the analyte specific calibration models listed in Tables II and III were subsequently validated with a set of independently collected, unmodified samples collected from a series of cell-cultivation flasks. The results of these predictions are presented in Figures 4 and 5 as concentration correlation plots. Plots for the full spectral range models are presented in Figure 4 and plots for the optimized spectral ranges are shown in Figure 5. In addition, the corresponding SEP values are listed in Tables II and III. In general, these concentration-correlation plots indicate that the predicted analyte values match the results from the reference method, thereby validating the spectroscopic method.

The optimized spectral range model has been deemed superior for glucose measurements. Prediction errors are significantly larger for the full spectral range model. As noted earlier, models with fewer factors are generally more robust (Martens and Naes, 1989) because too many factors increase the specificity of the model to a particular data set, thereby degrading the ability of the model to predict with spectra outside this data set. This

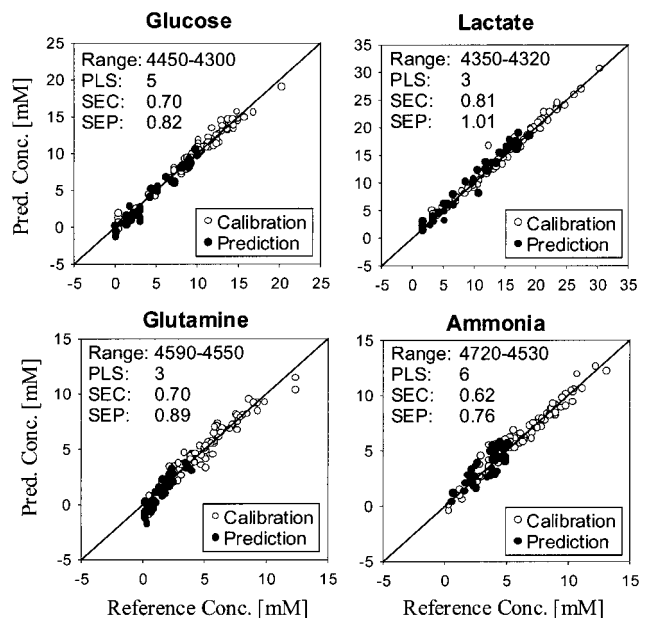
**Table III.** PLS calibration models with optimal spectral range.<sup>a</sup>

Analyte	Spectral range (cm <sup>-1</sup> )	PLS factors	SEC (mM)	SEP (mM)
Glucose	4450 to 4300	5	0.70	0.82
Lactate	4350 to 4320	3	0.81	1.01
Glutamine	4590 to 4550	3	0.70	0.89
Ammonia	4720 to 4530	6	0.62	0.76

<sup>a</sup>Spiked standards used for calibration and unmodified samples used for prediction.

**Figure 4.** Concentration correlation plots for calibration models generated with the full spectral range, showing predictions for the calibration (spiked) data (○) and the prediction (unmodified) data (●).

phenomenon was apparent for the eight-factor, full spectral range model for glucose where the SEP was nearly fourfold greater than the SEC (see Table II). Inspection of the glucose plot in Figure 4 reveals that the full spectral range model tended to underpredict the reference glucose concentrations over the entire concentration range of the prediction data. The glucose plot in Figure 5 reveals considerably less negative bias in

**Figure 5.** Concentration correlation plots for calibration model generated with the optimum spectral ranges, showing predictions for the calibration (spiked) data (○) and the prediction (unmodified) data (●).

predictions with the optimized spectral range. Nevertheless, a slight negative bias in glucose predictions is apparent in Figure 5 for concentrations below 5 mM. This negative bias may have been caused by interference from lactate, which would be present at high concentrations for all low glucose levels. Another issue is the limited number of low glucose concentrations in the calibration data set, which increased the overall measurement uncertainty in this concentration range.

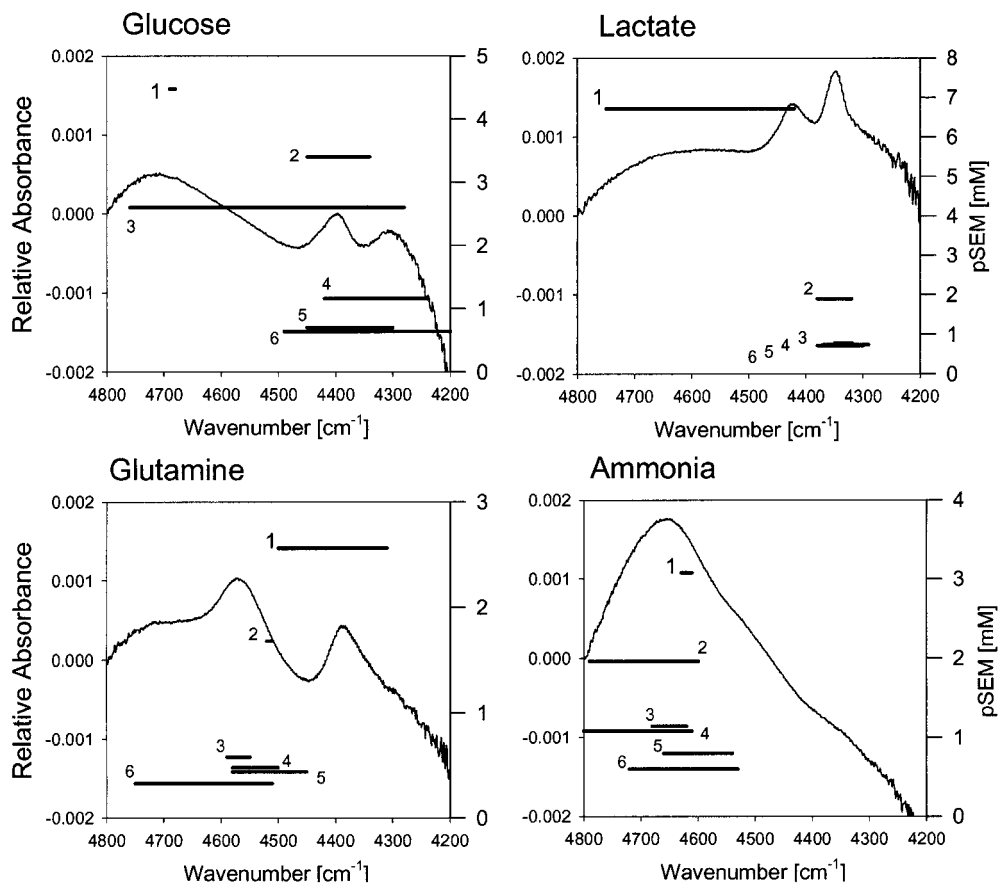
For lactate, models with the full spectral range and the optimized spectral range display approximately the same calibration and prediction ability as determined by the values of SEC and SEP. Inspection of the correlation plots in Figures 4 and 5 indicate no measurement bias in the prediction data set. In this case, the optimized spectral range model is preferred because significantly fewer PLS factors are required.

SEC and SEP values were lower for glutamine measurements with the full spectral range compared to models with the optimized spectral range. The corresponding concentration correlation plots in Figures 4 and 5 indicate a systematic bias at low glutamine concentrations for both the full and optimized spectral range models. This bias may have been caused by interference from ammonia because low glutamine levels

are always associated with high ammonia levels. In addition, the high-frequency spectral features of glutamine overlapped those of ammonia (see Fig. 1). As with glucose, the small number of low concentration standards in the glutamine calibration data set likely contributed to this measurement inaccuracy. Nevertheless, the finding of lower prediction errors with the full spectral range calls into question the effectiveness of the modified grid search used here to optimize the spectral range for analysis.

Little actual difference was noted for the SEC and SEP values for ammonia calibration models with the full or optimized spectral ranges. Likewise, the corresponding concentration correlation plots indicated similar performance between models. A slight positive bias was apparent in the prediction values with the full spectra range model. No such bias was evident in the optimized spectral range model. Again, the optimized spectral range model was preferred on the basis of fewer PLS factors.

It is of interest to examine the relationship between the optimized spectral ranges and the absorbance features for each molecule. Pure component absorbance spectra are plotted separately in Figure 6 for glucose, lactate, glutamine, and ammonia. In addition, the



**Figure 6.** Optimum spectral range and the corresponding pSEM values for glucose, lactate, glutamine, and ammonia superimposed on the respective absorbance spectrum.

optimized spectral range and the corresponding pSEM value are indicated in each plot for the first six PLS factors. As expected, the pSEM decreased as the number of factors increased to the optimal level. The plots in Figure 6 demonstrate how the optimum spectral range varied with the number of factors. As the number of factors increased, there was a definite migration of the optimum spectral range to the region of the spectrum that best distinguished that particular analyte. For lactate, for example, the optimum spectral range with only one factor covered a wide range. The optimum range rapidly converged to a narrow region specific for the distinguishing absorption feature centered at  $4386\text{ cm}^{-1}$ . For glutamine, on the other hand, the optimum spectral range converged around the distinguishing feature at  $4567\text{ cm}^{-1}$ . Likewise, the optimum spectral range converged on the distinguishing absorption features for glucose ( $4400\text{ cm}^{-1}$ ) and ammonia ( $4650\text{ cm}^{-1}$ ).

A noteworthy observation is that, in some cases, only a fraction of the distinguishing analyte absorption band was used. An example is lactate where the best calibration model was realized with three factors and a spectral range of  $4350\text{ to }4320\text{ cm}^{-1}$ . This spectral range incorporates only half of the distinguishing  $4345\text{ cm}^{-1}$  absorption feature of lactate. The use of this particular half of this lactate band effectively reduces the impact of both glucose and glutamine, which possess overlapping spectral features around  $4400\text{ cm}^{-1}$ .

Finally, the plot for glucose in Figure 6 underscores the fact that selective glucose measurements demand the widest spectral range compared with these other analytes. This fact is related to the significant degree of spectral overlap between glucose and these other components. Both glucose and glutamine possess absorption bands around  $4400\text{ cm}^{-1}$ . The  $4707\text{-cm}^{-1}$  band of glucose overlaps with the principal band of ammonia and there is a significant degree of overlap between the  $4309\text{-cm}^{-1}$  band of glucose and the  $4345\text{ cm}^{-1}$  band of lactate. Although these overlapping features can be distinguished by the PLS algorithm on the basis of subtle differences in bandwidth and location (Zhou et al., 1995), more spectral information and more factors are required.

## CONCLUSIONS

The results presented here demonstrate that NIR spectroscopy coupled with PLS regression and adaptive calibration procedures provide sufficient selectivity for the accurate measurement of glucose, glutamine, lactate, and ammonia in a serum-based cell culture medium. The complexities of serum-based media and the metabolic activity of cultured cells generated a challenging environment for measurement accuracy. The adaptive calibration procedure produced robust, analyte-specific calibration models. Model validation indicated predic-

tion errors on the order of  $0.6\text{ to }1.0\text{ mM}$ , depending on the analyte. Spectral analysis illustrated how the spectral information used in the PLS regression correlated strongly with the distinguishing absorption features for each analyte. Such a spectral analysis permits rationalization of the molecular basis of selectivity for these calibration models.

These results are an important advance toward the development of on-line noninvasive spectroscopic monitors. The aim is to provide continuous concentration information for all relevant endogenous chemical species. This information could be used for control purposes to maintain ideal growth conditions and maximize cellular productivity. Extrapolation of the PLS calibration models at low analyte concentrations is one potential limitation of the proposed adaptive calibration procedure for on-line monitoring. Under feedback-controlled conditions, however, high levels of nutrients and low concentrations of byproducts can be maintained, thereby lessening the impact of extrapolations in the measurement of cell nutrients.

## References

- Arnold MA, Burmeister JJ, Small GW. 1998. Determination of physiological levels of glucose in an aqueous matrix with digitally filtered Phantom glucose calibration models from simulated noninvasive human near-infrared spectra. *Anal Chem* 70:1773–1781.
- Burmeister JJ, Arnold MA. 1995. Accuracy of the YSI Stat Plus Analyzer for glucose and lactate. *Anal Lett* 28:581–592.
- Burmeister JJ, Arnold MA. 1999. Evaluation of measurement sites for noninvasive blood glucose sensing with near-infrared transmission spectroscopy. *Clin Chem* 45:1621–1627.
- Burmeister JJ, Chung H, Arnold MA. 1998. Phantoms for noninvasive blood glucose sensing with near-infrared transmission spectroscopy. *Photochem Photobiol* 67:50–55.
- Cavinato AG, Mayes DM, Ge Z, Callis JB. 1990. Noninvasive method for monitoring ethanol in fermentation processes using fiber-optic near-infrared spectroscopy. *Anal Chem* 62:1977–1982.
- Chung H, Arnold MA, Rhiel M, Murhammer DW. 1995. Simultaneous measurement of glucose and glutamine in aqueous solutions by near infrared spectroscopy. *Appl Biochem Biotechnol* 50:109–126.
- Chung H, Arnold MA, Rhiel M, Murhammer DW. 1996. Simultaneous measurements of glucose, glutamine, ammonia, lactate and glutamine in aqueous solutions by near-infrared spectroscopy. *Appl Spectrosc* 50:270–276.
- Ding Q, Small GW, Arnold MA. 1998. Genetic algorithm-based wavelength selection for near-infrared determination of glucose in biological matrixes: Initialization strategies and effects of spectral resolution. *Anal Chem* 70:4472–4479.
- Ge Z, Cavinato AG, Callis JB. 1994. Noninvasive spectroscopy for monitoring cell density in a fermentation process. *Anal Chem* 66:1354–1362.
- Glacken MW, Fleischaker RJ, Sinskey AJ. 1986. Reduction of waste product excretion via nutrient control: Possible strategies for maximizing product and cell yields on serum in cultures and mammalian cells. *Biotechnol Bioeng* 28:1376–1389.
- Hall JW, McNeil B, Rollins MJ, Draper I, Thompson BG, Macaloney G. 1996. Near-infrared spectroscopic determination of acetate, ammonium, biomass, and glycerol in an industrial *Escherichia coli* fermentation. *Appl Spectrosc* 50:102–108.

- Hazen KH, Arnold MA, Small GW. 1998. Measurement of glucose in water with first overtone near infrared spectra. *Anal Chim Acta* 371:255–267.
- Jones BN, Gilligan JP. 1983. *O*-phthalaldehyde precolumn derivatization and reversed-phase high-performance liquid chromatography of polypeptide hydrolysates and physiological fluids. *J Chromatogr* 266:471–482.
- Lewis CB, McNichols RJ, Gowda A, Côté GL. 2000. Investigation of near-infrared spectroscopy for periodic determination of glucose in cell culture media *in situ*. *Appl Spectrosc* 54:1453–1457.
- Lucasius CB, Beckers MLM, Kateman G. 1994. Genetic algorithms in wavelength selection—a comparative study. *Anal Chim Acta* 286:135–153.
- Martens H, Naes T. 1989. *Multivariate calibration*. New York: John Wiley & Sons.
- McShane MJ, Côté GL. 1998. Near-infrared spectroscopy for determination of glucose, lactate, and ammonia in cell culture media. *Appl Spectrosc* 52:1073–1078.
- Rhiel M. 1998. Application of near-infrared spectroscopy to bioreactor monitoring. PhD dissertation, University of Iowa, Iowa City, IA, USA.
- Riley MR, Rhiel M, Zhou X, Arnold MA, Murhammer DW. 1997. Simultaneous measurement of glucose and glutamine in insect cell culture media by near infrared spectroscopy. *Biotechnol Bioeng* 55:11–15.
- Riley MR, Arnold MA, Murhammer DW, Walls EL, De la Cruz N. 1998a. Adaptive calibration scheme for quantification of nutrients and byproducts in insect cell bioreactors by near infrared spectroscopy. *Biotechnol Progr* 14:527–533.
- Riley MR, Arnold MA, Murhammer DW. 1998b. Matrix enhanced buffer calibration procedure for multivariate calibration models with near infrared spectra. *Appl Spectrosc* 52:1339–1347.
- Rokhlin OW, Cohen MB. 1995. Expression of cellular adhesion molecules on human prostate tumor cell lines. *The Prostate* 26:205–212.
- Schügerl K. 1991. On-line analysis of broth. In: Reed HJ, Puhler G, Stadler P, editors. *Biotechnology*. Vol. 4. Weinheim, Germany: VCH.
- Sjöström M, Wold S, Lindberg W, Persson J-A, Martens H. 1983. A multivariate calibration problem in analytical chemistry solved by partial least-squares models in latent variables. *Anal Chim Acta* 150:61–70.
- Spear SK, Rhiel M, Murhammer DW, Arnold MA. 1998. Ammonia measurements in mammalian cell bioreactors with a diffuse reflectance-based fiberoptic ammonia sensor. *Appl Biochem Biotechnol* 75:175–186.
- Spiegelman CH, McShane MJ, Goetz MJ. 1998. Theoretical justification of wavelength selection in PLS calibration: Development of a new algorithm. *Anal Chem* 70:35–44.
- Topham DW. 1990. *A System V guide to UNIX and XENIX*. New York: Springer.
- Wold S, Albano C, Dunn WJ III, Edlund U, Esbensen K, Geladi P, Hellberg, Johansson E, Lindberg W, Sjöström M. 1990. In: Kowalski D, editor. *Chemometrics. mathematics and statistics in chemistry*. Berlin: Reidel. p 17–95.
- Yano T, Aimi T, Nakano Y, Tamai M. 1998. Prediction of the concentration of ethanol and acetic acid in the culture broth of a rice vinegar fermentation using near infrared spectroscopy. *J Ferment Bioeng* 85:461–465.
- Zhou X, Chung H, Arnold MA, Rhiel M, Murhammer DW. 1995. Selective measurement of glutamine and asparagine in aqueous media by near-infrared spectroscopy. In: Rogers KR, Mulchandani A, Zhou W, editors. *Biosensor and chemical sensor technology: Process monitoring and control*. ACS symposium series 613. Washington, DC: ACS. p 116–132.