Adaptive Calibration Scheme for Quantification of Nutrients and Byproducts in Insect Cell Bioreactors by Near-Infrared Spectroscopy

Mark R. Riley,[*,†,‡] **Mark A. Arnold,**[§] **David W. Murhammer,**[†] **Ed L. Walls,**[¶,Δ] **and Neslihan DelaCruz**[¶]

Department of Chemical and Biochemical Engineering and Department of Chemistry, University of Iowa, Iowa City, Iowa 52242, and American Cyanamid Corporation, San Leandro, California 94577

Spectroscopic methods are gaining in popularity in biotechnology because of their ability to deliver rapid, noninvasive measurements of the concentrations of multiple chemical species. Such measurements are particularly necessary for the implementation of control schemes for cell culture bioreactors. One of the major challenges to the development of spectroscopic methods for bioreactor monitoring is the generation of accurate and robust calibration models, particularly because of the inherent variability of biological processes. We have evaluated several methods of building calibration models, including synthetic calibrations and medium spiking methods. The approach that consistently produced reliable models incorporated samples removed from a bioreactor that were subsequently altered so as to increase the sample variation. Several large volume samples were removed from a bioreactor at varying time points and divided into multiple aliquots to which were added random, known amounts of the analytes of interest. Near-infrared spectra of these samples were collected and used to build calibration models. Such models were used to quantify analyte concentrations from independent samples removed from a second bioreactor. Prediction errors for alanine, glucose, glutamine, and leucine were 1.4, 1.0, 1.1, and 0.31 mM, respectively. This adaptive calibration method produces models with less error and less bias than observed with other calibration methods. Somewhat more accurate measurements could be attained with calibrations consisting of a combination of synthetic samples and spiked medium samples, but with an increase in calibration development time.

Introduction

The ability to monitor the concentrations of nutrients and byproducts in cell culture bioreactors is necessary for the development of intelligent bioreactor control schemes. Flow injection systems and off-line analyses provide accurate evaluation of the medium composition, however, they also require that samples be removed from the reactor prior to analysis (Christensen et al., 1996; Ozturk et al., 1997). This removal step can create concerns for bioreactor contamination, can restrict the number of samples taken from a small bioreactor, and can complicate automation schemes. *In situ* biosensor probes provide an attractive alternative because they may be placed directly in the bioreactor. Unfortunately, such enzymatic biosensors require frequent *in situ* calibrations because of enzyme degradation and consequent baseline drift (White et al., 1995). Because of these difficulties, a new noninvasive and nondestructive method to monitor the composition of bioreactor growth media is desirable.

Spectroscopic methods provide a promising alternative for the monitoring of bioreactors. Such methods may be used to rapidly quantify the concentration of numerous species in aqueous systems in a noninvasive and nondestructive manner. Spectroscopic quantifications are based on introducing a beam of near-infrared (NIR) radiation to a sample, collecting transmitted or reflected light, and correlating the amount and frequencies of light absorbed to the composition of the sample. This quantification process is rapid, requires no sample preparation, and can be used to quantify the concentration of multiple species simultaneously.

Spectroscopic methods have recently been applied to monitor a number of biological reactors. For yeast cell fermentations, the concentrations of ethanol (Cavinato et al., 1990); fructose, glycerol, glucose, and ethanol (Fayolle et al., 1996); and the yeast cell density (Ge et al., 1994) have been quantified by NIR spectroscopic methods. In a number of bacterial fermentations, the concentrations of lactate, glucose, and biomass (Vaccari et al., 1994); the concentrations of acetate and ammonia, and the cell density (Hall et al., 1996); and the concentration of NH_4OH (Macaloney et al., 1996) have been similarly quantified by NIR spectroscopy. Yano and

* To whom correspondence should be addressed.
† Department of Chemical and Biochemical Engineering, University of Iowa.
§ Department of Chemistry, University of Iowa.
¶ American Cyanamid Corp.
‡ Current address: Department of Agricultural and Biosystems Engineering, University of Arizona, Tucson, AZ 85721-0038.
Δ Current address: Biochemical Manufacturing Department, Genentech, Inc., 460 Point San Bruno Blvd., San Francisco, CA 94080-4990.

S8756-7938(98)00022-8 CCC: $15.00 © 1998 American Chemical Society and American Institute of Chemical Engineers
Published on Web 05/07/1998

Harata (1994) quantified the concentrations of glucose, glutamine, ammonia, and lactate in hybridoma cell cultivations. Previously, we have applied NIR spectroscopy to measure glucose and glutamine concentrations in insect cell culture samples of Sf-900 II growth medium (Riley et al., 1997).

The methods used to generate calibration models are fairly sophisticated because of the need for calibration samples to have uncorrelated chemical species concentrations, required for partial least-squares (PLS) regression analysis, and because of the large number of chemical species present in growth media. If correlations exist between similar analytes present in the calibration samples, then calibration models applied to samples that do not contain these same correlations will yield incorrect measurements.

Although many papers have been published in this area, there has been little evaluation of the merits of methods for producing spectroscopic calibration models. Most of the studies mentioned above used a sample removal approach in which a bioreactor is operated and spectra of the cell growth media are collected at varying times along with samples removed for the determination of the medium composition by standard techniques. If a single bioreactor run is used to generate samples for calibration, then such calibrations will contain unwanted correlations in the analyte concentrations because of cell metabolism. This problem can be diminished if calibration samples are drawn from more than one bioreactor. Fayolle and co-workers (1996) operated four calibration fermentations and Hall and co-workers (1996) ran five calibration fermentations. The use of multiple runs does not guarantee that species concentrations will be uncorrelated. Purely synthetic approaches have been used to generate calibration samples by dissolving known amounts of the analytes of interest in a suitable buffer used to represent the complex growth medium (Cavinato et al., 1990; Chung et al., 1996). Such methods simplify the attainment of uncorrelated compositions; however, producing a substantial sample set requires a prohibitive amount of time. Hybrid methods incorporate samples of culture medium to which are added random and known amounts of the desired analytes to eliminate analyte correlations (Riley et al., 1997). Our previous approach using only spent medium produced calibration models that could not be adequately applied to additional bioreactor runs.

Several methods for building calibration models for spectroscopic bioreactor monitoring have been evaluated and through this evaluation trends in the measurement error have become apparent. Calibration models constructed from samples collected from the end of a batch bioreactor run yield good predictions in other bioreactors only at the end of their operation. By expanding the range and composition of the background of medium samples utilized for the construction of calibrations, the applicability of spectroscopic measurements can be greatly enhanced. Such an adaptive calibration scheme yields models with low error and minimal bias when compared with models produced with previous calibration techniques. Calibration schemes that build on the adaptive calibration method yield small improvements in measurements but with increases in calibration development time.

Materials and Methods

Spodoptera frugiperda (Sf-9) insect cells were cultivated in a 3-L bioreactor with American Cyanamid Growth Medium at 27 °C for 6 days in two separate, independent bioreactor runs. Run 1 had an initial cell density of 6.3×10^5 cells/mL. Twenty samples of 20 mL each were removed from this bioreactor at varying times. These samples are termed the small volume (SV) samples. Run 2 had an initial cell density of 7.6×10^5 cells/mL and six samples of 100 mL each, termed the large volume (LV) samples, were removed at varying times (~1 sample/day). The spent medium from the end of run 2 was collected after 140 h of operation. Cells were removed from all samples by centrifugation, and 0.044% (w/v) 5-fluorouracil was added to each sample to retard microbial growth.

The concentrations of chemical species in the growth medium were determined by conventional reference methods. Glucose and lactate concentrations were measured using a YSI glucose and lactate analyzer (Yellow Springs Instruments, Yellow Springs, OH). This method has an accuracy of ~0.28 mM over the concentration range used (Burmeister and Arnold, 1995). Amino acid concentrations were determined by an HPLC method employing a C-18 reversed-phase column (Supelco, Bellefonte, PA) in conjunction with an OPA derivatization-based fluorescence detection scheme (Burbach et al., 1982). This method has an accuracy of ~0.8 mM (data not shown).

The collected growth medium samples were further divided into sets of samples for the development of spectroscopic calibration models. The pooled spent medium from run 2 was divided into 50 small aliquots to which were added random and known amounts of alanine, glucose, glutamine, and leucine. These samples are termed the spiked spent medium samples. The six LV samples were each divided into 10 small aliquots to which were added random and known amounts of alanine, glucose, glutamine, and leucine. These samples are termed the spiked LV samples. The spiked spent medium samples and the spiked LV samples were used to generate calibration models. The SV samples, from run 1, were used as an independent validation or prediction set to evaluate the ability of the calibration models to be applied to a separate bioreactor run. The errors associated with these predictions are calculated as deviations between the reference measurements (either HPLC or YSI) and spectroscopic measurements. NIR spectra were collected for the original SV samples, the spiked LV samples, and the spiked spent medium.

Also used for the development of calibration models were two sets of purely synthetic samples with a buffer of 1.05 g of $NaHCO_3$ and 3.039 g of NaH_2PO_4 in 3 L of deionized water, with the pH adjusted to 6.35. These buffer conditions are similar to the buffer capacity of the commercially available Sf-900 II growth medium for Sf-9 insect cells. Seventy samples with independent analyte concentrations were generated for each synthetic set. The first set of synthetic samples is termed the 6-component set because each sample contained alanine, glucose, glutamate, glutamine, lactate, and leucine at random and known concentrations dissolved in the buffer. Concentration ranges for each analyte were the following: alanine, 0−25 mM; glucose, 0−60 mM; glutamate, 0−40 mM; glutamine, 0−40 mM; lactate, 0−25 mM; and leucine, 0−25 mM. These ranges are wider than those normally encountered in insect cell media. The second set of synthetic samples is termed the 10-component set and contains the aforementioned six analytes plus aspartate, cysteine, isoleuceine, and serine at random, known concentrations dissolved in the same buffer. Concentration ranges for each analyte were the follow-

ing: alanine, 0−25 mM; aspartate, 0−25 mM; cysteine, 0−25 mM; glucose, 0−60 mM; glutamate, 0−40 mM; glutamine, 0−40 mM; isoleucine, 0−25 mM; lactate, 0−25 mM; leucine, 0−25 mM; and serine, 0−25 mM.

Spectra were collected on a Nicolet 550 Magna Fourier transform infrared (FTIR) spectrometer (Nicolet Analytical Instruments, Madison, WI) equipped with a 50 W Tungsten source, calcium fluoride beam splitter, and liquid nitrogen-cooled InSb detector. Samples were placed in a thermostated cell maintained at 27 °C with a path length of 1.5 mm. Each spectra consisted of 128 co-added triangularly apodized scans collected from 5000 to 4000 cm^{-1}. An interference filter (K filter from Barr and Associates, Westford, MA) was used to isolate this spectral region. A background spectrum of buffer was collected after every fourth sample. Collected spectra were transferred to a Silicon Graphics Indy workstation and analyzed using PLS regression analysis software developed by Professor Gary Small from the Center for Intelligent Chemical Instrumentation in the Department of Chemistry at Ohio University.

To select a calibration model parameter including the spectral range and number of PLS factors to be applied for each calibration set, one-fourth of the calibration samples were temporarily removed from the set and employed as an internal validation, or monitoring set. Optimal conditions were determined for the remaining calibration samples used to predict concentrations in the monitoring set. Once such conditions were determined, the monitoring samples were returned to the calibration set, and the same calibration parameters were used to establish PLS models. By employing a monitoring set, model parameters are independent of the prediction data set, which should produce more robust calibration models.

To thoroughly investigate the many possible combinations of calibration parameters, a C-shell computer script was written to systematically develop calibration models with varying numbers of PLS factors from 1 to 20 and with varying spectral ranges within the 5000−4000 cm^{-1} region. The script follows a modified grid search that permits the unbiased evaluation of many calibration data sets. For each model size (as defined by the total number of PLS factors), the script searches for spectral ranges that contain significant analyte information. Standard error of monitoring (SEM) values initially are calculated for 100 cm^{-1} wide regions at 100 cm^{-1} intervals beginning with 4100−4000 cm^{-1}. The region with the lowest SEM is assumed to contain significant analyte information and becomes the focus of further evaluation. The upper and lower values of this range are increased and subsequently decreased by a predetermined amount, and the corresponding SEM values are calculated. The process of modifying the maximum and minimum frequencies and evaluating ranges is repeated four times; each ensuing iteration has a more narrow step change in the spectral range. The number of PLS factors is incremented and another spectral range search is implemented. The combination of spectral range and number of PLS factors that yields the minimum SEM is consistently used to predict analyte concentrations in the SV samples.

Results and Discussion

Most components present in the American Cyanamid insect cell growth medium maintain a fairly constant level during each bioreactor run (data not shown). The concentrations of alanine, glucose, glutamine, and leucine change by the largest percentage and, therefore have

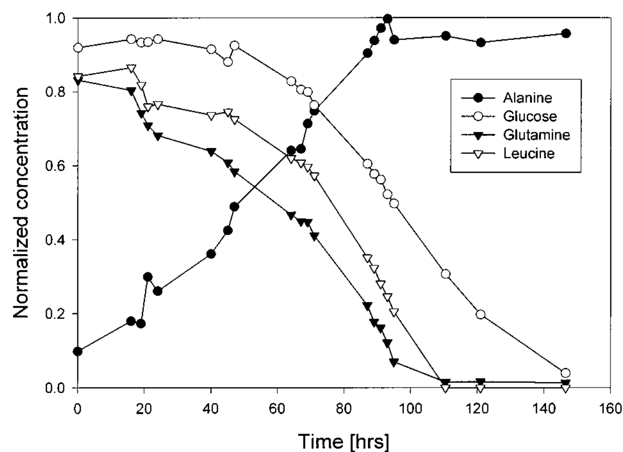

Figure 1. Representative profile of normalized concentrations of alanine, glucose, glutamine, and leucine present in batch cultivations of Sf-9 cells.

been selected as the analytes to be followed in this study. All other constituents are considered to be part of the background matrix. Figure 1 displays normalized concentration profiles of the four analytes of interest in a representative batch cultivation. Concentrations are normalized because of the proprietary nature of the growth medium. These four analytes follow similar trends in the two bioreactor runs, but the actual concentrations have measurable differences. These differences in medium composition support the need to develop robust monitoring schemes. Absorbance spectra of the four analytes followed in this study are presented in Figure 2. Each analyte has spectral features in this region that can be used to discriminate one analyte from another.

A variety of calibration models were generated with the 6- and 10-component synthetic samples, the spiked spent medium samples, and the spiked LV samples. Table 1 displays the standard error of calibration (SEC), standard error of monitoring (SEM), standard error of prediction (SEP), and mean percent error (MPE) along with the optimum calibration model parameters for each calibration set tested. In all cases, the SV samples were used as a prediction data set to evaluate analyte measurements.

Synthetic Samples for Calibration. Purely synthetic samples were used to generate calibration models for alanine, glucose, glutamine, and leucine in the SV samples. The synthetic samples alone yield very poor calibration models (Table 1) with SEPs of 6.5−49 mM for MPEs >2000%. Clearly, the synthetic samples do not adequately model the varying analyte and background concentrations in the complex growth medium.

Spent Medium for Calibration. The spiked spent medium samples were used to build calibration models. Analyte measurements by the references methods and the NIR predictions using the spiked spent medium are presented in Figure 3 for glucose, glutamine, leucine, and alanine, respectively. Curves are polynomial fits to the reference measurements. Compared with the synthetic sample calibrations, the spiked spent medium yields significantly better models with errors of 3.9 mM for alanine (MPE of 38%), 3.5 mM for glucose (MPE of 9.7%), 2.6 mM for glutamine (MPE of 27%), and 0.52 mM for leucine (MPE of 32%). Although these measurements are significant improvements over the purely synthetic calibrations, the errors are too high for implementation of a good control scheme and present significant measurement bias.

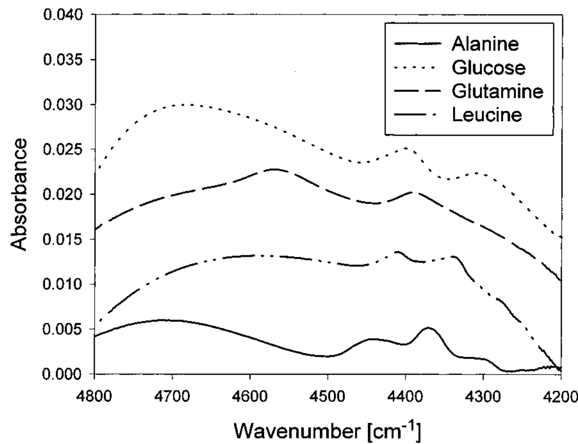

Figure 2. Absorbance spectra of alanine, glucose, glutamine, and leucine dissolved in a buffer.

Table 1. Parameters and Results with Calibration Models for Glucose, Glutamine, Leucine, and Alanine[a]

calibration	range, cm^{-1}	# factors	SEC, mM	SEM, mM	SEP, mM	MPE, %
Glucose Calibration Models						
6-component	4800−4200	12	0.928	1.110	6.459	26.4
10-component	4710−4240	20	0.831	1.570	49.30	201
spent	4720−4260	11	0.383	0.506	3.474	9.70
LV	4680−4280	10	1.350	1.802	1.008	2.96
LV + spent	4790−4400	18	0.843	1.339	1.941	6.58
6 + 10 + LV	4630−4300	12	2.744	1.698	1.662	5.30
6 + 10 + LV + spent	4720−4260	15	1.832	1.589	4.076	15.5
Glutamine Calibration Models						
6-component	4750−4270	18	0.141	0.506	16.46	290
10-component	4800−4270	19	0.321	0.969	10.34	204
spent	4670−4390	12	0.146	0.285	2.594	27.4
LV	4700−4300	11	0.585	0.389	1.095	14.7
LV + spent	4630−4300	11	0.541	0.380	0.955	12.6
6 + 10 + LV	4720−4490	9	1.807	0.652	1.499	30.4
6 + 10 + LV + spent	4780−4200	15	1.165	0.704	1.528	27.2
Leucine Calibration Models						
6-component	4640−4300	11	0.179	0.238	28.65	2010
10-component	4700−4250	17	0.393	0.787	7.175	492
spent	4520−4250	14	0.142	0.242	0.521	32.2
LV	4570−4300	11	0.241	0.283	0.309	15.9
LV + spent	4510−4290	11	0.272	0.298	0.235	12.3
6 + 10 + LV	4420−4230	8	1.552	0.685	0.835	53.0
6 + 10 + LV + spent	4620−4290	12	0.649	0.301	0.222	10.9
Alanine Calibration Models						
6-component	4750−4250	14	0.235	0.396	44.22	634
10-component	4800−4250	17	0.411	0.541	39.26	559
spent	4650−4400	9	0.481	0.256	5.038	38.1
LV	4510−4250	9	0.743	0.491	1.377	11.8
LV + spent	4620−4270	13	0.530	0.527	1.431	15.3
6 + 10 + LV	4770−4300	15	1.904	0.731	2.259	21.8
6 + 10 + LV + spent	4570−4300	14	1.887	1.067	1.309	11.2

[a] In all cases, a separate monitoring set was removed from the calibration set and used to optimize the spectral range and number of PLS factors using a modified grid-search method. The same prediction set of SV samples was used for each model.

Predictions from the spent medium calibrations for measurements of glucose and glutamine at the beginning of the bioreactor run show significant bias. Glucose concentrations are overpredicted (Figure 3a), whereas glutamine concentrations are underpredicted (Figure 3b). Errors for leucine are reasonably close to the target level of error, but spectroscopic measurement overpredicts at the beginning of the bioreactor run (Figure 3c). Alanine measurements are limited because high alanine concen-

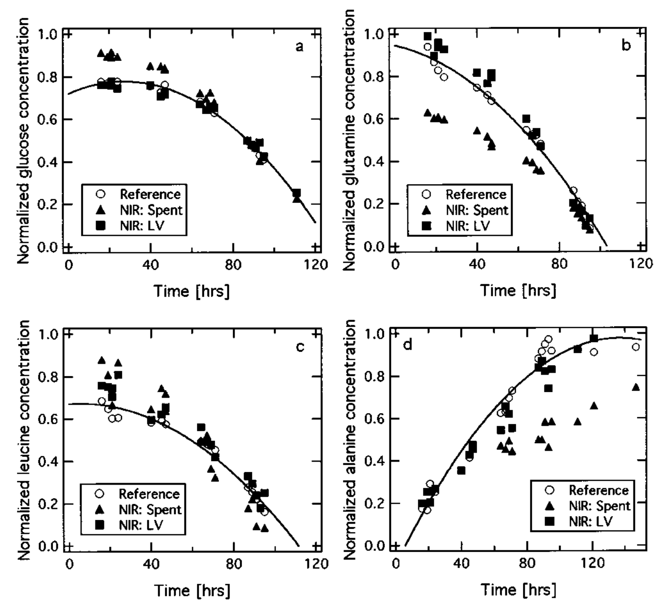

Figure 3. Concentrations of analytes in the SV samples as measured by reference methods (○) and spectroscopically using spent medium calibrations (▲) and the large volume calibrations (■) for glucose (a), glutamine (b), leucine (c), and alanine (d). Curves represent a polynomial fit to the reference measurements.

trations are present in the spent medium which establishes the lowest concentration in the calibration model. Samples out of the valid calibration range (time < 60 h) have been temporarily removed from the prediction data set and only the remaining 12 samples are displayed in Figure 3d. Prediction is poor for these remaining samples, with a significant negative bias relative to the reference measurements. The likely cause for these prediction biases is that the background constituents of the calibration and prediction samples differ significantly.

Trends in the apparent prediction bias using the spiked spent medium calibrations suggest an alternate method for calibration model development. The analyte predictions using the spiked spent medium are most accurate for SV samples collected at the end of the run because the background material of the calibration samples is most similar to the background of the prediction samples at the end of the run. These spent medium samples do not reflect the media composition present at the beginning and middle of the bioreactor run. By including samples from the beginning and middle of the bioreactor run to the calibration data set, a more accurate representation of the changing composition of the background materials may be obtained, and thus should improve the spectroscopic measurements.

LV Samples for Calibration. Analyte predictions of the SV samples are greatly improved by using the spiked LV samples as a calibration data set, compared with the spiked spent medium calibrations. Spectroscopic measurements are consistent with the reference measurements throughout the bioreactor run, errors are substantially decreased (Table 1), and bias is greatly reduced (Figure 3). The LV calibration contains a substantial diversity in the background material, which is representative of changes encountered during the course of a bioreactor run, and so analyte predictions are more accurate over a greater time span.

Measurements of each of the analytes using the LV calibrations represent substantial improvements compared with the spent medium calibrations. For glucose

Table 2. Glucose Prediction Using the LV Calibration Models with One or Two Sets of LV Samples Removed[a]

model	samples removed from calibration	spectral range, cm^{-1}	# factors	SEC, mM	SEM, mM	SEP, mM	MPE, %
0	none (all 6LV used)	4680–4280	10	1.380	1.802	1.008	2.96
1	LV1	4630–4300	15	0.747	0.648	1.433	5.34
2	LV2	4750–4300	18	0.428	0.328	1.239	4.65
3	LV3	4650–4380	14	0.857	0.705	1.520	5.39
4	LV4	4790–4300	19	0.450	0.533	1.377	4.45
5	LV5	4600–4390	13	0.748	1.171	3.350	13.16
6	LV6	4800–4390	12	0.877	1.442	1.206	4.17
7	LV1 and LV6	4590–4300	16	0.465	1.415	1.683	5.85
8	LV1 and LV2	4680–4300	17	0.373	0.401	1.214	4.80
9	LV2 and LV4	4680–4350	16	0.443	0.543	1.413	5.42
10	LV3 and LV4	4700–4370	17	0.524	0.439	2.422	8.96
11	LV3 and LV5	4620–4390	13	0.689	1.101	3.495	13.75
12	LV5 and LV6	4450–4300	16	0.588	1.671	1.830	6.13

[a] In all cases, a separate monitoring set was removed from the calibration set and used to optimize the spectral range and number of PLS factors using a modified grid-search method. The same prediction set of SV samples was used for each model.

(Figure 3a), the SEP is reduced from 3.5 to 1.0 mM, MPE is reduced from 9.7 to 2.9%, and the correlation coefficient between predicted and known concentrations increased from 0.9428 to 0.9889. Glutamine measurements display a reduction in the SEP from 2.6 to 1.1 mM and a small increase in the correlation coefficient from 0.9573 to 0.9743. Predictions of glutamine for samples taken before 60 h have a small positive bias, although the magnitude of these errors is lower than those from the spent medium calibrations. Leucine measurements are improved by using the LV calibrations, with diminishing values of the SEP and MPE, lower overall bias (Figure 3c), and an increase in the correlation coefficient from 0.7581 to 0.9371. The spent medium calibrations lead to an overprediction of leucine at high concentrations and underprediction at low concentrations, however, LV calibrations yield fairly consistent measurement over the entire concentration range. Alanine measurements using the LV calibration are also improved, with reductions in the SEP and MPE (Figure 3d) and increase in the correlation coefficient from 0.7626 to 0.9628. The dynamic calibration range for alanine is broadened for the LV calibration models compared with the spent media samples.

The success of the LV calibrations compared with the synthetic sample calibrations and the spent medium calibrations is due to the greater variations in the composition of the background matrix. These variations are apparent in that the SEC is greater for the LV samples than for the spent medium samples because of these background differences. By including greater matrix variation in the calibration data set, more chemical information is incorporated into the PLS models, thereby improving prediction quality and enhancing model robustness. The spent medium calibrations, which contain no changes in the background matrix, can provide accurate analyte measurements only for samples taken from similar time points and bioreactor compositions. Insensitivity to such alterations in the background matrices is critical for bioreactor monitoring, and so the LV approach provides a more feasible scheme for calibration model development. The LV calibrations also provide a greater valid concentration range because of the incorporation of samples from the beginning of bioreactor operation.

Combination Calibrations. Because the amount of background variations incorporated into the calibration set had a substantial effect on the applicability of these calibration models, the effect of increasing sample variations was further evaluated. Combinations of the 6-component, 10-component, spent medium, and LV samples were used to build calibration data sets. The performance characteristics for these hybrid models are summarized in Table 1.

Calibration models based on both spent medium and LV samples result in slightly lower errors for glutamine and leucine measurement, but greater errors for alanine and glucose. Errors for glucose are substantially higher than obtained for the LV calibration models.

Calibrations using the 6- and 10-component samples and the LV samples yield higher errors for all analytes. Errors for alanine and leucine nearly double, and those for glucose and glutamine increase by ~50%. Clearly, the spectral variation added by these synthetic samples is not useful. In fact, the information provided by these samples degrades performance, which suggests that variations in spectral features from the sample matrix are dominating relative to spectral variations caused by the spectrometer and data collection parameters.

Models were built using all of the 6- and 10-component samples, spent medium samples, and LV samples. Compared with the previous set of models, adding the spent medium samples to this calibration set increases the error in predictions for glucose and glutamine, but decreases error for alanine and leucine. In fact, models generated from all these data provide the lowest prediction errors for alanine and leucine.

Although some of these combination calibration models yield the lowest errors for certain analytes, the predictions for the other analytes using the same approach often are quite poor. This inconsistency brings into question the robustness of this combination approach. The one method that produces low errors consistently for all analytes is the use of LV samples alone for calibration, which also has the advantage of minimal preparation time. The additional time invested in generating samples and collecting spectra for the synthetic samples is substantial, but yields little improvement in performance. Adding the spent media samples to the LV calibrations does in some cases improve measurements, however, these improvements are small and do not necessarily merit the doubling in preparation time that is required.

Partial LV Calibrations. The number of LV samples originally used to generate calibrations was based on the number of days of bioreactor operation, resulting in a total of six LV samples. To evaluate whether all these samples are needed to construct reliable calibrations, calibration models were constructed for glucose quantification using only a portion of the LV samples. The first

set of models (Models 1–6), summarized in Table 2, incorporate only five sets of LV samples; the second set of models (Models 7–12), also summarized in Table 2, incorporate only four sets of LV samples. Glucose models using all six sets of LV samples previously produced an SEP of 1.01 mM.

Removal of one set of LV samples increases prediction errors by 20–40% for most samples compared with the full LV calibration set (Table 2, Models 1–6). Removal of the LV5 samples yields a model with more than three times the error of the full calibration set. Apparently, the LV5 samples contain spectral information that is particularly relevant to applying the LV calibration to predict glucose in the SV samples.

Removal of two sets of LV samples leads to an even greater increase in the glucose prediction error (Table 2, Models 7–12). Withdrawal of the first two samples (LV1 and LV2) collected from the beginning of the bioreactor run yields the smallest increase in prediction error (Model 8). Conversely, models constructed without the last two samples (LV5 and LV6) yield a somewhat higher level of error (Model 12) with an SEP of 1.8 mM. One might anticipate that the LV1 and LV6 samples would be the most critical for producing robust calibration models as these samples span the range of medium composition. However, Model 7 (without LV1 and LV6 samples) yields only a moderate level of error with an SEP of 1.7 mM. Removal of samples from the middle of the bioreactor run appear to have the greatest impact on prediction errors. For example, calibration models without LV3 and LV5 samples (Model 11) yield an SEP of 3.5 mM, whereas models without LV3 and LV4 samples (Model 10) have an SEP of 2.4 mM. Models for all other conditions yield SEPs <2 mM.

Models constructed with less than complete calibration sets yield errors greater than those obtained with using all six LV samples. Each set of LV samples contributes some unique spectral information, either analyte dependent or analyte independent, to the calibration models. Clearly, greater amounts of relevant spectral information provide more robust calibration models. The challenge remains in determining which information is useful for PLS regression analysis and which is confounding.

Conclusions

NIR spectroscopic methods have been developed to noninvasively measure the concentration of alanine, glucose, glutamine, and leucine in samples removed from an Sf-9 insect cell culture bioreactor. Attempts at applying a purely synthetic calibration method to predict analyte concentrations in the bioreactor samples yielded poor results. Calibration models developed from spiked spent medium produced better measurements than did the synthetic calibrations, however, these predictions had a substantial amount of error and significant prediction bias. An adaptive calibration scheme using six large-volume samples removed from a bioreactor and divided into multiple aliquots provided the most reliable and accurate calibration models. These levels of error should be appropriate for applying on-line control of insect cell bioreactors, specifically to maintain nutrient and waste levels within fairly tight tolerances. Future studies should evaluate the ability to apply these calibration models to predict analyte concentrations in samples removed from a greater number of bioreactor runs and possibly to other types of growth media.

Acknowledgment

The authors thank Mr. Martin Rhiel for development of the computer script for calibration model evaluation, Dr. John Wiencek for use of his HPLC, Dr. Gary Small for the spectral processing software, and Dr. Ron Chiarello for development of the defined culture medium. Financial support for this project was provided by American Cyanamid Corp. and from NASA (No. NAG 9-824).

Notation

LV	large volume samples
LV#	large volume samples collected on day #
MPE	mean percent error
PLS	partial least-squares
SEC	standard error of calibration
SEM	standard error of monitoring
SEP	standard error of prediction
SV	small volume samples

References and Notes

Burbach, J. P. H.; Prins, A.; Lebouille, J. L. M.; Verhoef, J.; Witler, A. Sensitive and rapid amino acid analysis of peptide hydrolysates by high performance liquid chromatography of *o*-phthaldialdehyde derivatives. *J. Chromatogr.* **1982**, *237*, 339–343.

Burmeister, J. J.; Arnold, M. A. Accuracy of the YSI stat plus analyzer for glucose and lactate. *Anal. Lett.* **1995**, *28*, 581–592.

Cavinato, A. G.; Mayes, D. M.; Ge, Z.; Callis, J. B. Noninvasive method for monitoring ethanol in fermentation processes using fiber-optic near-infrared spectroscopy. *Anal. Chem.* **1990**, *62*, 1977–1982.

Christensen, L. H.; Marcher, J.; Schulze, U.; Carlsen, M.; Min, R. W.; Nielsen, J.; Villadsen, J. Semi-on-line analysis for fast and precise monitoring of bioreaction processes. *Biotechnol. Bioeng.* **1996**, *52*, 237–247.

Chung, H.; Arnold, M. A.; Rhiel, M.; Murhammer, D. W. Simultaneous measurements of glucose, glutamine, ammonia, lactate, and glutamate in aqueous solutions by near-infrared spectroscopy. *Appl. Spectrosc.* **1996**, *50*, 270–276.

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*, 1325–1330.

Ge, Z.; Cavinato, A. G.; Callis, J. B. Noninvasive spectroscopy for monitoring cell density in a fermentation process. *Anal. Chem.* **1994**, *66*, 1354–1362.

Hall, J. W.; McNeil, B.; Rollins, M. J.; Draper, I.; Thompson, B. G.; Macaloney, G. Near-infrared spectroscopic determination of acetate, ammonium, biomass, and glycerol in an industrial *Escherichia coli* fermentation. *Appl. Spectrosc.***1996**, *50*, 102–108.

Macaloney, G.; Draper, I.; Preston, J.; Anderson, K. B.; Rollins, M. J.; Thompson, B. G.; Hall, J. W.; McNeil, B. At-line controls and fault analysis in an industrial high cell density *Escherichia coli* fermentation, using NIR spectroscopy. *Food Bioprod. Test.* **1996**, *74*, 212–220.

Naes, T.; Isaksson, T. Selection of samples for calibration in near-infrared spectroscopy. Part I: General principles illustrated by example. *Appl. Spectrosc.* **1989**, *43*, 328–335.

Ozturk, S. S.; Thrift, J. C.; Blackie, J. D.; Naveh, D. Real-time monitoring and control of glucose and lactate concentrations in a mammalian cell perfusion reactor. *Biotechnol. Bioeng.* **1997**, *53*, 372–378.

Riley, M. R.; Rhiel, M.; Zhou, X.; Arnold, M. A.; Murhammer, D. W. Simultaneous monitoring of glucose and glutamine in insect cell cultures by NIR spectroscopy. *Biotechnol. Bioeng.* **1997**, *55*, 11–15.

Vaccari, G.; Dosi, E.; Campi, A. L.; Gonzales-Vara y R., A.; Matteuzzi, D.; Mantovani, G. A near-infrared spectroscopy technique for the control of fermentation processes: An application to lactic acid fermentation. *Biotechnol. Bioeng.* **1994**, *43*, 913–917.

White, S. F.; Turner, A. P. F.; Bilitewski, U.; Bradley, J.; Schmid, R. D.; On-line monitoring of glucose, glutamate, and glutamine during mammalian cell cultivations. *Biosens. Bioelectron.* **1995**, *10*, 543–551.

Yano, T.; Harata, M.; Prediction of the concentration of several constituents in a mouse-mouse hybridoma culture by near-infrared spectroscopy. *J. Ferment. Bioeng.* **1994**, *77*, 659–662.

Accepted March 20, 1998.

BP980022D