

Multivariate Calibration Models Based on the Direct Analysis of Near-Infrared Single-Beam Spectra

GENG LU, XIANGJI ZHOU, MARK A. ARNOLD,* and GARY W. SMALL

Department of Chemistry, University of Iowa, Iowa City, Iowa 52242 (G.L., X.Z., M.A.A.); and Center for Intelligent Chemical Instrumentation, Department of Chemistry, Clipping Laboratories, Ohio University, Athens, Ohio 45701-2979 (G.W.S.)

Multivariate calibration models are generated for glucose, glutamine, and asparagine on the basis of partial least-squares regression analysis of near-infrared single-beam spectra covering the 5000–4000-cm⁻¹ (2000–2500-nm) spectral range. Models are constructed with both raw and digitally Fourier-filtered single-beam spectra. Model performance is evaluated and compared with that of analogous models constructed from the corresponding computed absorbance spectra. Five unique data sets are examined corresponding to the measurement of (A) glucose in phosphate buffer with different temperatures, (B) glucose with variable albumin protein levels, (C) glucose with variable triacetin levels, (D) glucose and glutamine in a set of binary mixtures, and (E) glutamine and asparagine in a set of binary mixtures. In all cases, models based on single-beam spectra perform as well as those based on computed absorbance spectra.

Index Headings: Near-infrared analysis; Single-beam spectral analysis; Spectroscopic glucose measurements.

INTRODUCTION

The possibility of performing noninvasive blood glucose measurements with near-infrared (near-IR) spectroscopy has received considerable attention in recent years.^{1–7} Most reported approaches to date are based on extracting analytical information from transmission measurements based on passing a band of near-IR radiation through some region of the body. This type of measurement is complicated by the strongly absorbing and highly scattering characteristics of the human body.

In attempting to establish the feasibility of noninvasive glucose measurements, the philosophy of our research program is to increase the complexity of the sample matrix systematically while validating our ability to measure clinically relevant levels of glucose accurately at each step. These studies have been performed with a conventional Fourier transform spectrometer and have consisted of transmission measurements performed with samples held in standard liquid cells. Single-beam spectra of the glucose samples are obtained through this procedure.

To construct quantitative calibration models with spectra of this type, it is standard practice to convert the spectra to absorbance units by ratioing the collected single-beam spectra of the samples to a similarly collected background single-beam spectrum and converting the resulting transmittance values to absorbance. If the instrument response function of the spectrometer has remained stable between the background and sample single-beam measurements, the computed absorbance spectrum approximates the spectrum that would be obtained in a true double-beam spectral measurement.

The conversion of the single-beam spectra to absorbance units has two purposes. First, since the Beer-Lambert law describes a linear relationship between absorbance and concentration, it is generally held that the use of spectra in absorbance units will facilitate the development of linear calibration models for use in predicting analyte concentrations. Second, if a background spectrum can be collected that provides a good match to the sample matrix, the resulting ratioed spectrum will be simplified and the analyte spectral features will be enhanced. Stated differently, the ratioing procedure will remove a principal source of variation from the spectrum (i.e., the instrument response function) that is unrelated to the spectral response of the analyte.

In our work, as the matrix complexity has increased from simple aqueous buffer solutions to more demanding matrices such as undiluted plasma, human serum, and whole blood, a fundamental limitation of using ratioed spectra has surfaced. This limitation is based on the difficulty of identifying a representative background spectrum for use in computing the ratioed absorbance spectrum. For example, in the analysis of blood, a glucose-depleted sample of the blood for use in collecting a background spectrum is not typically available. In this situation, we have ratioed single-beam spectra from complex samples to a reference spectrum obtained from a simple aqueous buffer solution.^{6,7} As the sample matrix has become more complicated, however, the mismatch between sample and background spectra has become more severe. Our findings indicate that a severely mismatched background spectrum degrades analytical performance by introducing spectral variation that is unrelated to the analyte.⁴

Two additional issues cast further doubt on the wisdom of converting the sample spectra to absorbance units in the glucose analysis. First, the near-IR spectral response of the aqueous sample matrix is so temperature-sensitive that, even under strict experimental control, there will invariably be some degree of temperature variation between the sample and background spectra. This variation leads to the introduction of baseline shifts and curvature that can interfere with the construction of calibration models.⁵ Second, the near-IR spectra of biological samples consist of a series of broad, heavily overlapped spectral bands. The glucose bands represent a tiny component of the overall absorbance and are heavily overlapped with the bands of water, proteins, fats, and other matrix constituents.⁷ A spectral response of this type violates the assumptions upon which the Beer-Lambert law is based. Thus, there is no guarantee of a linear relationship between absorbance and concentration. For this reason, cal-

Received 25 September 1996; accepted 15 March 1997.

* Author to whom correspondence should be sent.

ibration models constructed from near-IR spectra tend to be multivariate in nature, often requiring a large number of model terms.⁸

Given these problems, an alternative strategy would be to construct calibration models directly from the single-beam spectra of aqueous samples. The major complications of using single-beam spectra directly are created by an inability to remove transient spectral artifacts caused by uncontrollable instrument and environment variations. Nevertheless, the complications caused by poorly matched background spectra and the long-range goal of noninvasive clinical measurements, which precludes collection of a suitable background spectrum, motivate an investigation to assess the utility of calibration models based on single-beam spectra.

In this work, five independent spectral data sets are used to compare the analytical performance of calibration models based on single-beam and ratioed absorbance spectra. Each data set and the corresponding absorbance-based calibration models have been described previously.^{5,6,8,9} These data sets correspond to the measurement of (A) glucose in phosphate buffer at different temperatures; (B) glucose in a variable protein matrix; (C) glucose in a matrix with different levels of triacetin; (D) binary mixtures of glucose and glutamine in phosphate buffer; and (E) glutamine and asparagine in phosphate buffer.

EXPERIMENTAL

Experimental details have been published elsewhere for each data set, and the relevant features for each are summarized in Table I.^{5,6,8,9} Samples were prepared with reagent-grade chemicals obtained from common suppliers. For all data sets, the solvent was a 0.1 M phosphate buffer (pH 7.2–7.4). Individual samples were prepared by mixing appropriate amounts of stock solutions and diluting with the buffer. In all cases, near-IR spectra were collected over the spectral range from 5000 to 4000 cm^{-1} (2000–2500 nm) with a Nicolet 740 Fourier transform spectrometer (Nicolet Analytical Instruments, Madison, WI). This spectral range corresponds to the combination spectral region and was isolated by placing a multilayer interference filter (Barr Associates, Westford, MA) in front of the sample cell. The sample pathlength was 1 mm throughout, and samples were contained in a temperature-controlled sample cell with Infracil quartz windows. Double-sided interferograms consisting of 16,384 points were collected on the basis of 256 coadded scans. The interferograms were sampled at every zero-crossing of the HeNe reference laser, producing spectra with a point spacing of 1.9 cm^{-1} and a maximum frequency of 15,798.57 cm^{-1} . Single-beam spectra were obtained by Fourier processing the interferograms with triangular apodization and Mertz phase correction. These calculations were performed with software resident on the Nicolet 620 computer controlling the spectrometer. In general, three replicate spectra were collected for each sample. The order in which the samples were measured was randomized with respect to concentration to ensure that no correlations were introduced between concentration and time. Periodically throughout the data collection, spectra of the phosphate buffer were acquired for subsequent use in computing absorbance values for the sam-

TABLE I. Summary of sample matrices and data sets.

Data set	Matrix	Analyte(s)	Number of samples	Number of spectra	Data collection temperature (°C)	Comments	Reference
A	Glucose buffer	Glucose	58	173	37	1.25–19.66 mM glucose	5
			6	178	32–41	3.09, 5.80, 9.01, 12.08, 15.03, and 17.98 mM glucose	
B	Glucose/protein buffer	Glucose	110	322	37	1.25–20.00 mM glucose	6
						47.5–90.25 and 0 g/L protein (BSA)	
C	Glucose/triacetin buffer	Glucose	86	253	37	2–20 mM glucose	6
						7.17, 8.87, 10.64, 14.19, and 17.02 mM triacetin	
D	Glucose/glutamine buffer	Glucose, glutamine	70	209	27	1.10–58.91 mM glucose	8
						1.41–30.65 mM glutamine	
E	Glutamine/asparagine buffer	Glutamine, asparagine	66	197	27	1.0–13.6 mM glutamine	9
						1.0–11.7 mM asparagine	

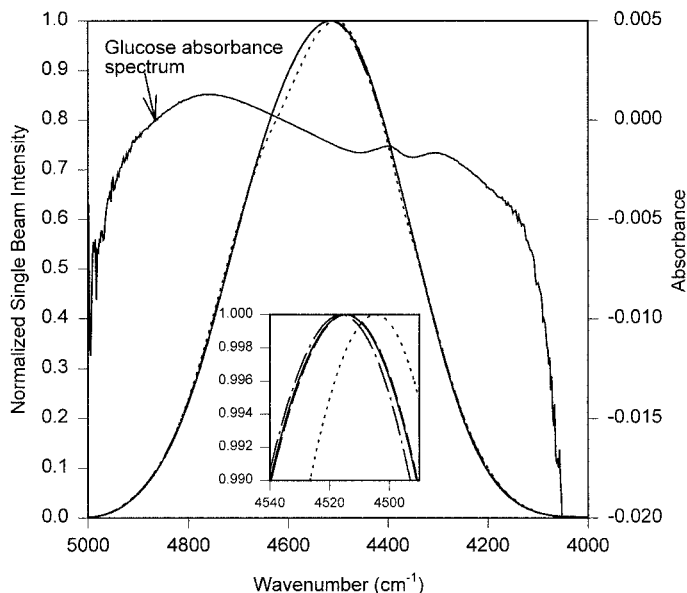


FIG. 1. Near-IR single-beam spectra of buffer (solid line), 10 mM glucose in buffer (short dashed line), 10 mM glucose and 66.5 g/L BSA in buffer (dotted line), and 10 mM glucose and 10.66 mM triacetin in buffer (dash-dot line). Spectra for buffer and 10 mM glucose in buffer are essentially identical. The inset amplifies single-beam spectra to high-light differences. An absorbance spectrum corresponding to 10 mM glucose in phosphate buffer is superimposed on these single-beam spectra.

ple spectra. Other relevant information about the five data sets is provided in Table I.

The computed single-beam spectra were transferred to a Silicon Graphics Iris Indigo R4000 computer (Silicon Graphics, Inc., Mountain View, CA) operating under Irix (Version 5.2). The remaining calculations reported here were performed on the Silicon Graphics computer with original software written in FORTRAN 77. Multiple linear regression and Fourier transform calculations used in the development of the calibration models were performed with subroutines from the IMSL library (IMSL, Inc., Houston, TX).

Calibration models were generated by application of partial least-squares (PLS) regression¹⁰ to a subset of spectra termed the calibration set. The spectra corresponding to individual samples were assigned randomly to the calibration set. Across the five data sets, the calibration set typically contained 75–90% of the available samples. All replicate spectra of a selected sample were carried together into the calibration set, with the remaining spectra assigned to a prediction set for use in evaluating the computed models. The performance of the models was judged on the basis of values for the standard error of calibration (SEC), standard error of prediction (SEP), and mean percent error of prediction (MPEP). SEC and SEP denote the root-mean-squared error in predicted concentrations for the spectra in the calibration and prediction sets, respectively. The degrees of freedom used in the SEC calculation are adjusted for the number of terms in the calibration model. The value of MPEP is computed as

$$\text{MPEP} = \frac{\sum_{i=1}^n \frac{C_i - \hat{C}_i}{C_i}}{n} \times 100\% \quad (1)$$

where C_i is the actual analyte concentration associated with spectrum i , \hat{C}_i is the corresponding concentration predicted by the model, and n denotes the number of spectra in the prediction set.

Fourier filtering was evaluated as a preprocessing tool for use in eliminating unwanted variation from both the single-beam and absorbance spectra. This procedure uses the Fourier transform to model the spectrum as a sum of sine and cosine waveforms. The filter attenuates the amplitudes of specified sine and cosine frequencies, thereby serving to suppress certain features in the spectrum. For example, baseline variation tends to be modeled by low-frequency sines and cosines, while spectral noise is modeled primarily by high frequencies. In the work performed here, the frequency response of the filter was specified as a Gaussian function defined by a bandpass position and width. These position and width parameters were expressed in units of digital frequency (f) as the mean and standard deviation, respectively, of the corresponding Gaussian. The digital frequency scale has a range of 0–0.5, where 0.5 designates the maximum sine/cosine frequency of the data. The filter is applied by multiplying the frequency response function by the result of the Fourier transform. After the application of the filter, the inverse Fourier transform returns the filtered data to the original domain (i.e., to the single-beam or absorbance spectrum). In previous work with computed absorbance spectra, a protocol was developed for coupling Fourier filtering with PLS regression.¹¹

Two procedures were used to identify the optimal position and width of the filter bandpass. The procedure used for data sets A–C matched that described before.¹¹ The samples in the calibration set were subdivided into a calibration subset and a monitoring set, and a grid-search procedure was used to build calibration models with spectra filtered with various combinations of filter positions and widths. For each filter tested, the model was computed with the calibration subset and tested by predicting the concentrations corresponding to the spectra in the monitoring set. For data sets D and E, an alternative strategy was implemented in which samples corresponding to a specified fraction of the calibration set were randomly extracted and used to form the monitoring set. This procedure was repeated six times at each combination of filter position and filter width, thus forming six random splittings of the data. The decision regarding the effectiveness of a given set of filter parameters was based on pooling the calibration and prediction results from each calibration/monitoring subset. The latter procedure has proven slightly more robust as the influence of any potentially outlying samples in either the calibration or monitoring sets is diluted. The data sets used in this work were large and consistent enough, however, to ensure that both procedures were able to provide analogous results.

RESULTS AND DISCUSSION

Characteristics of Single-Beam Spectra. Water is the predominant component in the matrices examined in this work. The high concentration and strong absorption features of water largely dictate the structural features of the corresponding single-beam spectra. Representative single-beam spectra are presented in Fig. 1 for comparison.

The plotted spectra correspond to phosphate buffer, phosphate buffer with 10 mM glucose, phosphate buffer with 10 mM glucose and 66.5 g/L albumin protein, and phosphate buffer with 10 mM glucose and 10.66 mM triacetin. The raw spectra have different intensity levels because of differences in the instrument response function of the spectrometer when the data sets were collected. To enhance comparisons, we normalized each spectrum by scaling the maximum intensity to unity. A computed absorbance spectrum of 10 mM glucose in buffer is also included.

The basic shape is the same for all of the single-beam spectra. A maximum intensity is observed at 4514 cm^{-1} (2215 nm), which corresponds to the minimum water absorbance in this spectral region. The intensity decreases at both lower and higher wavenumbers corresponding to the strong water absorption bands centered at 3800 (2632 nm) and 5200 cm^{-1} (1923 nm), respectively. Even at 4514 cm^{-1} , the influence of water is significant. The absorbance at 4514 cm^{-1} is approximately 1.0 for a 1-mm-thick sample of water, which corresponds to the absorption of 90% of the incident radiation.

No differences can be visualized when comparing single-beam spectra of phosphate buffer with and without glucose (i.e., these spectra overlap in Fig. 1). Absorption of light by glucose is simply too small to alter the spectral appearance significantly. Some differences can be observed, however, in the presence of albumin protein and triacetin. Slightly lower intensities are evident at the characteristic N–H and C–H combination bands of protein at 4603 cm^{-1} (2172 nm) and 4368 cm^{-1} (2289 nm), respectively. A similar decrease in intensity is observed at the characteristic C–H combination band of triglycerides (4445 cm^{-1} , 2250 nm). Similarly, analyte absorption features are not readily apparent in single-beam spectra from the binary mixtures of glucose and glutamine or glutamine and asparagine. The effects of unavoidable instrumental variations, such as minor alignment changes and source intensity fluctuations, are generally eliminated in absorbance spectra by using appropriately collected background spectra. With single-beam spectra, however, such variations are retained within the data set. Intensity variation can be significant even under ideal conditions of spectrometer operation. For example, we have measured the intensity at 4514 cm^{-1} for 18 individual single-beam spectra of phosphate buffer. These spectra were collected over a 14-day period as part of a routine data collection session. The mean intensity was 88.25 Nicolet single-beam units (NSBUs), and the standard deviation was 2.28 NSBUs, which corresponds to a 2.6% variation. A variation of this magnitude is larger than the intensity changes due to absorptions from the highest concentrations of either glucose, glutamine, or asparagine.

Glucose Measurements in Buffer. *Constant Temperature.* The feasibility of building functioning PLS calibration models with single-beam spectra was initially established by analyzing spectra of glucose in phosphate buffer. In this experiment, 191 spectra corresponding to 64 samples were extracted from data set A listed in Table I. These spectra were collected at $37\text{ }^{\circ}\text{C}$, and the glucose concentration ranged from 1 to 20 mM. The spectra were divided into a calibration set (58 samples and 173 spectra) and a prediction set (6 samples and 18 spectra). PLS

calibration models were constructed for glucose by using 1 to 20 PLS factors in conjunction with the $4457\text{--}4354\text{-cm}^{-1}$ ($2244\text{--}2297\text{-nm}$) spectral range. This spectral range isolates the 4400-cm^{-1} (2273-nm) glucose absorption band, which was found to be ideal during the analysis of computed absorbance spectra.⁵ The best calibration model was based on four PLS factors and exhibited SEC, SEP, and MPEP values of 0.33 mM, 0.43 mM, and 4.84%, respectively. The PLS procedure was not significantly affected by variation in single-beam spectral intensities due to changes in the instrument response function over the course of the data collection. No spectral normalization calculations were needed.

For comparison, PLS models were constructed after converting these single-beam spectra to absorbance units by use of a representative buffer spectrum as the background. Essentially equivalent model performance was found with SEC, SEP, and MPEP values of 0.29 mM, 0.34 mM, and 4.1% for a four-factor PLS model. SEC and SEP values for both spectral types track each other when plotted as a function of the number of PLS factors. These results demonstrate that the PLS algorithm is able to extract sufficiently selective glucose information from the single-beam spectra to allow valid calibration models to be computed.

Variable Temperature. Temperature strongly influences single-beam spectra because of the temperature sensitivity of the underlying water absorption bands. As noted before,⁵ the 5200- and 3800-cm^{-1} water absorption bands shift toward higher frequencies with increasing solution temperature. The corresponding changes in the single-beam intensity are large on the scale of the glucose absorbances.⁵ Systematic variations in solution temperature were used to examine the effects of large spectral variations on the accuracy of PLS calibration models based on single-beam spectra. Our experimental protocol was analogous to that reported for absorbance spectra,⁵ where calibration models were generated from spectra collected at a single temperature ($37\text{ }^{\circ}\text{C}$), while predictions were made from spectra collected at different solution temperatures ($32\text{--}41\text{ }^{\circ}\text{C}$). The calibration data set for this experiment consisted of 173 single-beam spectra from 58 glucose solutions maintained at $37\text{ }^{\circ}\text{C}$. For the prediction data set, 178 single-beam spectra were collected from six different glucose solutions maintained at 10 different temperatures ranging from 32 to $41\text{ }^{\circ}\text{C}$ at $1\text{ }^{\circ}\text{C}$ increments. PLS calibration models were constructed with four unique spectral regions, and the number of factors was varied from 1 to 20 for each region. These spectral ranges, which correspond to those used in the analysis of absorbance spectra, are listed in Table II along with results from models based on single-beam and absorbance spectra. Again, the optimal number of PLS factors corresponds to the minimum SEP which, in this case, included prediction errors across all temperatures. The effect of temperature is demonstrated by comparing values in Table II for the $4457\text{--}4354\text{-cm}^{-1}$ spectral range with those discussed above for measurements at $37\text{ }^{\circ}\text{C}$. More factors are required to reach a minimum prediction error, and the resulting SEP is higher for the multiple temperature experiment (10 factors with $\text{SEP} = 0.66\text{ mM}$ compared with four factors with $\text{SEP} = 0.34\text{ mM}$).

Figure 2 summarizes the prediction data as a function

TABLE II. Glucose calibration models with variable sample temperature.

Spectra type ^a	Spectra range (cm ⁻¹)	Mean position (<i>f</i>) ^b	Standard deviation (<i>f</i>) ^b	PLS factors	SEC (mM)	SEP (mM)	MPEP (%)
S	4850–4220	N/F	N/F	12	0.08	0.45	4.48
S		0.0200	0.0015	8	0.11	0.11	1.04
A		N/F	N/F	9	0.09	0.19	2.00
A		0.0264	0.0049	6	0.12	0.15	1.60
S	4811–4457	N/F	N/F	9	0.43	1.86	21.69
S		0.0185	0.0015	4	0.14	0.14	1.56
A		N/F	N/F	9	0.49	3.07	37.09
A		0.0200	0.0022	4	0.18	0.19	2.01
S	4457–4354	N/F	N/F	10	0.22	0.66	6.81
S		0.0200	0.0020	4	0.13	0.17	1.50
A		N/F	N/F	10	0.21	0.73	7.55
A		0.0286	0.0051	1	0.18	0.17	1.88
S	4354–4227	N/F	N/F	6	0.61	5.08	58.85
S		0.0185	0.0015	7	0.11	0.15	1.35
A		N/F	N/F	12	0.54	3.68	43.31
A		0.0254	0.0034	10	0.16	0.17	1.75

^a S: single beam; A: absorbance.

^b N/F: no filtering.

of temperature for models based on the 4457–4354-cm⁻¹ spectral range. The distribution of the SEP values as a function of sample temperature is plotted in Fig. 2A for the optimal models based on unfiltered single-beam spectra (open bars), unfiltered absorbance spectra (slashed bars), and Fourier-filtered single-beam spectra (cross-hatched bars). A temperature-sensitive bias is clearly evident in the model constructed with the unfiltered single-beam spectra. The SEP values decrease as the sample temperature approaches 37 °C due to the presence of only 37 °C data in the calibration set. For this model, Fig. 2B shows a correlation plot of predicted vs. actual glucose concentrations, with the temperatures denoted by differ-

ent symbols. Least-squares lines are also plotted on the basis of the predicted vs. actual data at each temperature. These lines clearly indicate the offsets in the predicted values according to the deviation of the temperature from 37 °C. This systematic variation is consistent with the observed temperature-induced spectral shifts of the water absorption bands.

Similar results were obtained with the models based on the computed absorbance spectra. Again, the calibration model was generated from spectra collected at 37 °C, while prediction spectra were collected over the 32–41 °C temperature range. In this experiment, however, absorbance spectra were computed by use of a background spectrum collected at 37 °C for all spectra regardless of temperature. Pertinent values for this calibration model are provided in Table II for the four spectral ranges. The distribution of the SEP values in Fig. 2A is analogous to that obtained from models based on the single-beam spectra.

The magnitude of the temperature-induced errors depends on the spectral range. The 4850–4220-cm⁻¹ (2062–2370-nm) range is least sensitive to temperature, closely followed by the 4457–4354-cm⁻¹ range. Neither the 4811–4457-cm⁻¹ (2079–2244-nm) nor the 4354–4227-cm⁻¹ (2297–2366-nm) range performed well individually, producing values of MPEP greater than 20 and 40%, respectively. Poor performance is explained by the fact that the corresponding glucose absorption bands are located in spectral regions most affected by variations in the water bands (see Fig. 1).

Our previous work with absorbance spectra demonstrated the effectiveness of using Fourier filtering to eliminate the temperature-induced calibration errors described above. As discussed previously, these filters can remove

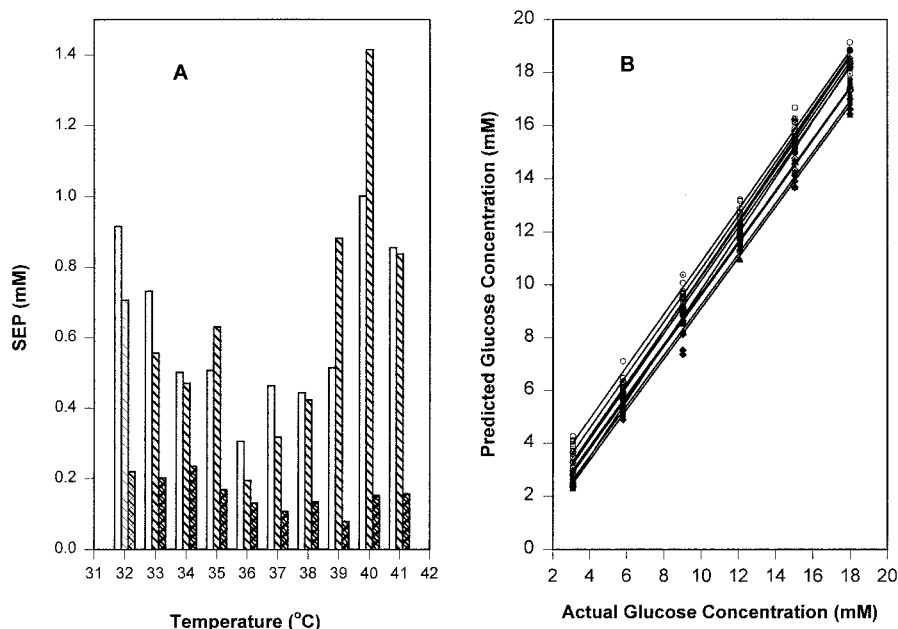


Fig. 2. Effect of sample temperature on glucose prediction error. (A) SEP values for models computed from the 4457–4354-cm⁻¹ spectral range and based on single-beam spectra without Fourier filtering (open bars), absorbance spectra without Fourier filtering (upward slash), and single-beam spectra with Fourier filtering (cross hatch). (B) For the model computed from the 4457–4354-cm⁻¹ range and based on unfiltered single-beam spectra, concentration correlation plots are provided for sample temperatures of 32 (circle), 33 (square), 34 (diamond), 35 (dotted circle), 36 (dotted square), 37 (dotted diamond), 38 (solid circle), 39 (solid square), 40 (solid diamond), and 41 (solid up triangle) °C. Additional model parameters are provided in Table II.

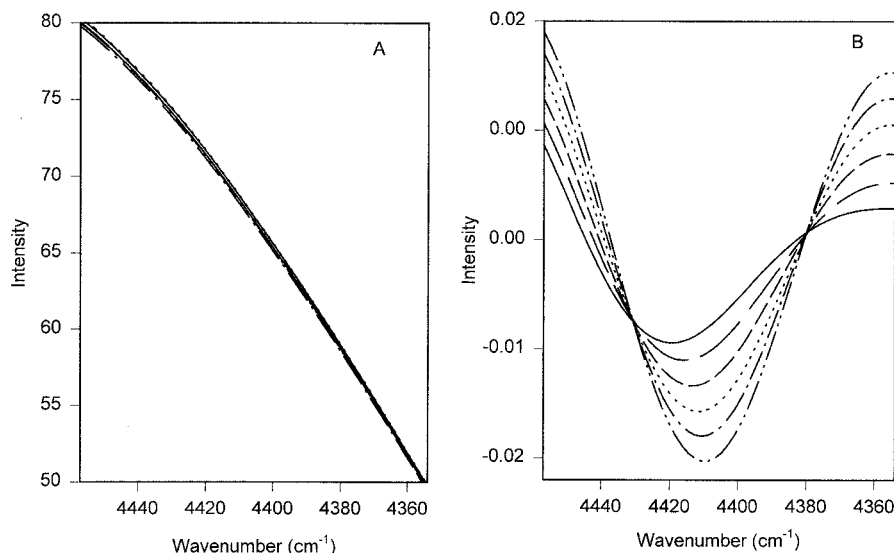


FIG. 3. Raw (A) and Fourier-filtered (B) single-beam spectra for glucose concentrations of 3.09 mM (solid line), 5.80 mM (long dash line), 9.01 mM (short dash line), 12.08 mM (dotted line), 15.03 mM (dash-dot line), and 17.98 mM (dash-dot-dot line). All spectra were collected at 37 °C and Fourier filters used values of 0.02 and 0.002 f for the position and width, respectively.

low-frequency baseline variations and high-frequency spectral noise. This past success motivated our attempt to apply Fourier filtering to single-beam spectra. The effectiveness of Fourier filtering in isolating glucose-dependent information from single-beam spectra is demonstrated in Figs. 3 and 4. Figure 3 shows raw and filtered single-beam spectra collected at 37 °C. The filter position and width used were 0.02 and 0.002 f , respectively. No glucose information is apparent in the raw spectra presented in Fig. 3A. However, a clear relationship between spectral intensity and glucose concentration can be easily identified after Fourier filtering (see Fig. 3B). The action of the filter can be understood by an inspection of Fig. 4. Figures 4A and 4B plot groups of amplitude spectra over the range of 0.01–0.03 f . The amplitude spectra represent the combined magnitudes of the sine and cosine components obtained by applying the Fourier transform to the selected single-beam spectra. In Figs. 4A and 4B, the optimized Gaussian filter frequency response function used in the generation of Fig. 3B is plotted as a dashed line superimposed on the amplitude spectra. The area of each amplitude spectrum underneath the Gaussian is passed by the filter to an extent that is weighted by the Gaussian shape, while the area outside the Gaussian (i.e., approximately 0–0.01 and 0.03–0.5 f) is completely suppressed.

Figure 4A plots two groups of seven amplitude spectra corresponding to 17.98 (solid line) and 5.80 (dotted line) mM glucose collected at 32, 33, 35, 36, 37, 39, and 41 °C. All spectra were acquired during the same data collection session, thus ensuring that the overall single-beam spectral intensities were consistent. This figure illustrates that the filter bandpass identifies a region of the amplitude spectrum in which the variation due to a change in glucose concentration exceeds that due to changes in temperature. Outside the filter bandpass, little variation is noted due to the change in glucose concentration (i.e., the solid and dotted lines are virtually superimposed).

Figure 4B provides further illustration that the filter isolates information related to a change in glucose con-

centration. Amplitude spectra are plotted corresponding to 19.19 (solid line), 15.03 (dotted line), 11.39 (dash-dot line), and 3.09 (dashed line) mM glucose collected at 37 °C. All spectra were acquired during the same data collection session, again ensuring that the single-beam spectral intensities were consistent. An inspection of Fig. 4B confirms that the region of the amplitude spectra isolated by the filter is optimal in terms of extracting information that relates to variation in glucose concentration. The plots in Figs. 3 and 4 demonstrate that even though the dominant sources of variation in the single-beam spectra are related to the spectrometer characteristics and the absorbance of the sample matrix, analyte-dependent information can be isolated directly without the use of a background spectral measurement.

Results of PLS calibration models based on Fourier filtered spectra are summarized in Table II for the same four spectral ranges examined before. Inspection of calibration and prediction errors reveals significant improvement for all spectral ranges relative to models based on unfiltered single-beam spectra. Models based on filtered spectra produce lower errors with fewer factors. The need for fewer factors suggests the presence of less spectral variation in the filtered spectra, which is consistent with the removal of temperature-dependent spectral variations and with the data in Fig. 4A. The true effectiveness of the Fourier filtering step is illustrated in Fig. 2A, where prediction errors are plotted as a function of sample temperature. No temperature-dependent bias is apparent in results based on Fourier-filtered single-beam spectra. The optimal filter parameters are also provided in Table II for each of the spectral ranges. The parameters for the filter position and width are similar regardless of spectral range. These values are also similar to those obtained with absorbance spectra.

Glucose Measurements with Variable Protein. Protein is a major component of most biological fluids and can represent a significant interference for glucose measurements because of strong near-IR absorption features and high concentrations. PLS regression can selectively discrim-

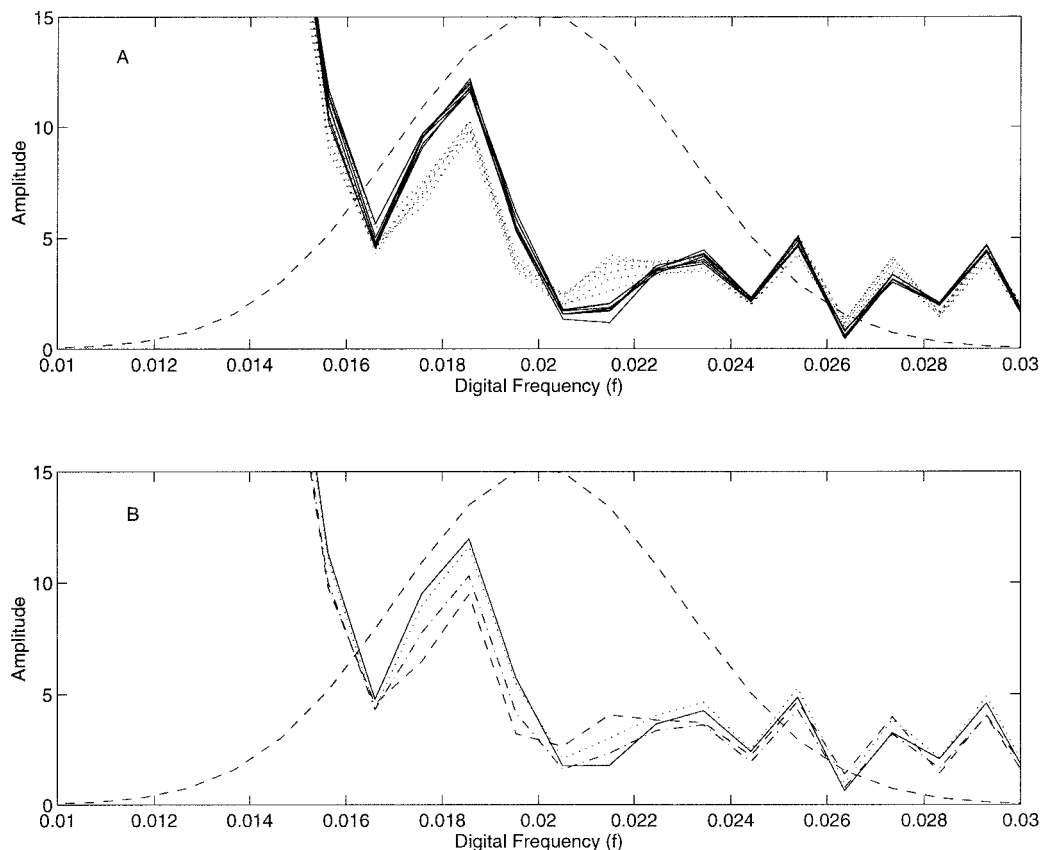


FIG. 4. Amplitude spectra are plotted over the digital frequency range of 0.01–0.03 f . The dashed line superimposed on the amplitude spectra is a Gaussian filter frequency response function characterized by position and width parameters of 0.02 and 0.002 f , respectively. (A) Two groups of seven amplitude spectra are plotted corresponding to 17.98 (solid line) and 5.80 (dotted line) mM. The spectra in each group were collected at 32, 33, 35, 36, 37, 39, and 41 °C. (B) Amplitude spectra are plotted corresponding to 19.19 (solid line), 15.03 (dotted line), 11.39 (dash-dot line), and 3.09 (dashed line) mM. All spectra were collected at 37 °C.

inate glucose and protein from absorbance spectra computed with a protein-free background spectrum.⁶ The ability to measure glucose selectively from single-beam spectra is demonstrated below. Our protein data set (listed as data set **B** in Table I) contains spectra collected from solutions consisting of bovine serum albumin (BSA) and glucose dissolved in a 0.1 M/pH 7.2 phosphate buffer. A total of 100

standard solutions were prepared with 10 levels of BSA varying from 47.5 to 90.25 g/L and 10 levels of glucose ranging from 1 to 20 mM. An additional 10 glucose solutions were prepared without protein. As was done with models based on absorbance spectra,⁶ all spectra associated with 0.0, 52.5, and 85.5 g/L BSA were set aside and were not part of either the calibration, monitoring, or prediction data sets. As such, 233 spectra from 80 standard solutions were divided into the different data sets. Specifically, 186 spectra from 64 standards were used in the calibration set with the remaining 47 spectra from 16 samples in the prediction set. For the Fourier filter optimization process, the calibration set was reduced further by placing 46 spectra from 16 samples into the monitoring set. All these data sets were the same as those described before for models based on absorbance spectra.⁶

Results are summarized in Table III for all models generated from single-beam spectra of protein-containing solutions. For models computed without Fourier filtering, the indicated number of factors corresponds to the optimal values as judged by the minimum SEP. For models that employed Fourier filtering, five factors were used to determine the optimal values for the filter position and width parameters, after which the number of factors giving the lowest SEP was established. The tested spectral ranges match those described above for the measurement of glucose in buffer. These ranges are similar to those reported for the corresponding models based on absorbance spectra,⁶ but they do

TABLE III. Glucose calibration models with variable protein.

Spectra type ^a	Spectral range (cm ⁻¹)	Mean position (f) ^b	Standard deviation (f) ^b	PLS factors	SEC (mM)	SEP (mM)	MPEP (%)
S	4850–4220	N/F	N/F	10	0.31	0.49	2.84
S		0.0140	0.0010	6	0.36	0.19	2.92
A		N/F	N/F	7	0.30	0.24	4.31
A		0.0225	0.0055	6	0.35	0.16	3.28
S	4811–4457	N/F	N/F	6	0.52	0.42	6.41
S		0.0170	0.0020	7	0.35	0.19	2.42
A		N/F	N/F	8	0.41	0.44	7.00
A		0.0190	0.0025	10	0.35	0.17	2.84
S	4457–4354	N/F	N/F	6	0.50	0.52	10.90
S		0.0140	0.0010	7	0.35	0.17	2.35
A		N/F	N/F	4	0.54	0.34	5.04
A		0.0185	0.0015	5	0.36	0.17	2.54
S	4354–4227	N/F	N/F	7	0.59	0.51	9.26
S		0.0170	0.0010	7	0.35	0.19	2.57
A		N/F	N/F	9	0.42	0.51	6.66
A		0.0175	0.0010	9	0.37	0.21	3.43

^a S: single beam; A: absorbance.

^b N/F: no filtering.

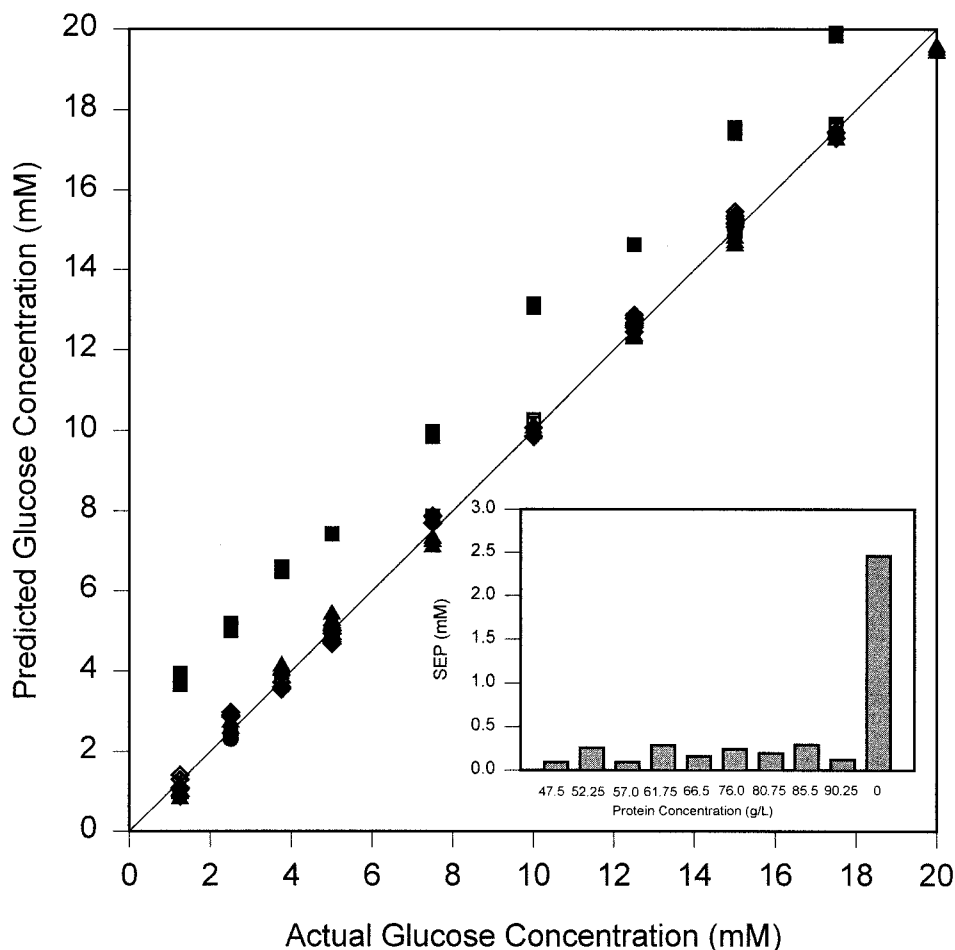


FIG. 5. Concentration correlation plots for glucose predictions with protein levels of 0.0 g/L (solid square), 47.5 g/L (circle), 52.25 g/L (solid diamond), 57.0 g/L (square), 61.75 g/L (diamond), 66.5 g/L (dotted circle), 76.0 g/L (dotted square), 80.75 g/L (dotted diamond), 85.5 g/L (solid up angle), and 90.25 g/L (solid circle). The inset shows SEP values as a function of protein level. All results correspond to the 4457–4354-cm⁻¹ spectral range with Fourier-filtered single-beam spectra (see Table II for further details).

not match exactly. For comparison purposes, new models were computed by using these same spectral ranges with both raw and Fourier-filtered absorbance spectra, and the results are included in the table. Models based on unfiltered single-beam spectra compare favorably with those generated with unfiltered absorbance spectra. In both cases, the lowest prediction errors are obtained with the widest spectral range, which incorporates all three glucose absorption bands and both protein absorption bands. More factors are required for the single-beam model to achieve equivalent prediction performance. Fewer factors are needed and inferior performance is indicated for both spectral types with the narrower spectral ranges. Similar model performance and optimal filter parameters are also obtained with Fourier-filtered single-beam and absorbance spectra. Prediction errors are approximately 0.2 mM regardless of spectral type or spectral range. This insensitivity to spectral range for Fourier-filtered spectra is consistent with our earlier findings. In addition, optimal values for the filter parameters are essentially the same for both spectral types over all spectral ranges tested. The optimal filter position and width values are on the order of 0.02 and 0.002 f , respectively.

As noted before,⁶ the prediction errors in Table III may be overly optimistic, because all the protein levels in the prediction data set are represented exactly in the calibra-

tion set. A more rigorous evaluation was performed by judging the ability of these models to predict glucose levels from spectra where the BSA levels were 0.0, 52.25, and 85.5 g/L. The resulting predictions are presented as concentration correlation plots in Fig. 5. Glucose predictions are coded according to protein level. No difference in prediction error is evident when comparing values for 52.25 and 85.5 g/L BSA to any other protein level. The inset shows SEP values for glucose prediction at the various BSA levels tested. Again, no significant differences are apparent with any of the BSA levels used. Prediction errors are significantly larger, however, when the BSA level is 0.0. Predictions in this latter case require an extrapolation outside the boundaries of the calibration model, which result in the offset shown in Fig. 5.

Glucose Measurements with Variable Triacetin. Triglycerides are a class of biological compounds that present a major challenge for the near-IR measurement of glucose because of their relatively high concentrations in biological matrices and the existence of a strong near-IR absorption band at 4450 cm⁻¹. This absorption band significantly overlaps with the key glucose absorption band centered at 4400 cm⁻¹. Triacetin has been used as a model triglyceride to demonstrate the accuracy of PLS models for glucose with absorbance spectra.⁶ This same set of spectra is evaluated

TABLE IV. Glucose calibration models with variable triacetin.

Spectral type ^a	Spectra range (cm ⁻¹)	Mean position (<i>f</i>) ^b	Standard deviation (<i>f</i>) ^b	PLS factors	SEC (mM)	SEP (mM)	MPEP (%)
S	4850–4220	N/F	N/F	6	0.55	0.54	4.99
S		0.0315	0.0055	5	0.41	0.51	6.01
A		N/F	N/F	5	0.68	0.56	5.49
A	4811–4457	0.0215	0.003	9	0.42	0.50	5.77
S		N/F	N/F	9	0.55	0.88	9.46
S		0.0265	0.003	7	0.44	0.55	6.68
A	4457–4354	N/F	N/F	12	0.46	0.98	11.71
A		0.0200	0.0015	10	0.43	0.50	6.25
S		N/F	N/F	7	0.47	0.58	6.44
S	4354–4227	0.0205	0.0005	7	0.42	0.51	5.87
A		N/F	N/F	5	0.53	0.68	6.83
A		0.0195	0.0015	7	0.45	0.50	5.26
S	4354–4227	N/F	N/F	10	0.57	0.88	10.62
S		0.026	0.0005	5	0.48	0.55	7.31
A		N/F	N/F	12	0.46	0.70	9.03
A		0.016	0.0050	9	0.46	0.61	6.83

^a S: single beam; A: absorbance.

^b N/F: no filtering.

here to assess models based on single-beam spectra. Our triacetin data set is composed of 253 single-beam spectra generated from 86 standard solutions. Within this data set, triacetin values range from 7.17 to 17.02 mM and glucose levels range from 1 to 20 mM. The calibration set contained 208 spectra from 71 solutions, and the prediction set contained 45 spectra from 15 solutions. For filter optimization, the calibration set was reduced to 157 spectra (53 solutions) after 51 spectra (18 solutions) were transferred to the monitoring set. Six PLS factors were used to identify the best Fourier filter parameters. Comparisons between models based on single-beam and absorbance spectra reveal essentially the same results as noted above for measurements of glucose in buffer with variable temperature and glucose in buffer with variable levels of BSA. Optimization results and model performance are summarized in Table IV for this variable triacetin data set. Again, prediction errors are essentially the same for all spectral ranges regardless of the spectral type. With this data set, slightly fewer PLS factors were needed for the single-beam models. Fourier filtering of either single-beam or absorbance spectra results in models with lower prediction errors and less sensitivity to spectral range. Prediction errors for all spectral ranges are either equivalent or slightly lower for models based on filtered single-beam spectra compared with filtered absorbance spectra. Errors in glucose prediction are essentially constant across all levels of triacetin, which indicates no prediction bias caused by variations in the levels of this triglyceride.

Measurement of Glucose and Glutamine in Binary Mixtures. As a first step toward developing near-IR spectroscopic methods for monitoring bioreactors non-invasively, we have established the ability to measure both glucose and glutamine in a set of binary mixtures. Accurate measurements were demonstrated for each analyte from individual PLS models generated from both raw and Fourier-filtered absorbance spectra.⁸ The corresponding models based on single-beam spectra are evaluated here. The glucose/glutamine data set is composed of 209 spectra collected from 70 standard solutions. Glucose and glutamine concentrations were selected randomly for each solution, and, as noted before, no correlation exists between these concentrations across the entire data

TABLE V. Glucose calibration models in glucose/glutamine binary mixtures.

Spectral type ^a	Spectra range (cm ⁻¹)	Mean position (<i>f</i>) ^b	Standard deviation (<i>f</i>) ^b	PLS factors	SEC (mM)	SEP (mM)	MPEP (%)
S	4800–4250	N/F	N/F	10	0.65	0.50	3.19
S		0.025	0.004	6	0.69	0.47	3.21
A		N/F	N/F	7	0.57	0.43	2.48
A	4800–4470	0.025	0.004	5	0.64	0.36	1.89
S		N/F	N/F	7	1.01	0.91	5.00
S		0.02	0.003	6	0.76	0.47	3.28
A	4470–4250	N/F	N/F	6	0.88	0.48	2.91
A		0.02	0.003	6	0.66	0.42	3.50
S		N/F	N/F	7	0.74	0.50	3.46
S	4350–4250	0.025	0.004	8	0.70	0.48	3.02
A		N/F	N/F	6	0.64	0.41	2.21
A		0.023	0.003	6	0.65	0.38	2.14
S	4350–4250	N/F	N/F	8	0.77	0.96	7.02
S		0.030	0.004	7	0.71	0.52	3.18
A		N/F	N/F	5	0.91	0.55	4.02
A		0.025	0.004	6	0.72	0.62	3.71

^a S: single beam; A: absorbance.

^b N/F: no filtering.

set.⁸ Glucose and glutamine values ranged from 1.10 to 58.91 and 1.41 to 30.65 mM, respectively. Distribution of the spectra into calibration, prediction, and monitoring sets is described elsewhere.⁸ Results are summarized in Tables V and VI for glucose and glutamine models based on raw and Fourier-filtered spectra. The spectral ranges used for these models incorporate different combinations of absorbance bands for glucose and glutamine, respectively. Comparison of values in Table V for glucose models based on unfiltered spectra reveals a trend where absorbance spectra typically outperform single-beam spectra. This trend is not as clear when comparing SEP values for glutamine models based on unfiltered spectra (Table VI). Glutamine models based on single-beam spectra actually possess lower prediction errors for several of the tested spectral ranges. For models based on filtered spectra, however, prediction errors are slightly higher for both glucose and glutamine when models are constructed from single-beam spectra as opposed to absorbance spectra.

TABLE VI. Glutamine calibration models in glucose/glutamine binary mixtures.

Spectral type ^a	Spectra range (cm ⁻¹)	Mean position (<i>f</i>) ^b	Standard deviation (<i>f</i>) ^b	PLS factors	SEC (mM)	SEP (mM)	MPEP (%)
S	4800–4250	N/F	N/F	13	0.43	0.83	6.80
S		0.045	0.0013	6	0.68	0.97	7.56
A		N/F	N/F	8	0.70	0.86	7.60
A	4700–4450	0.024	0.005	7	0.87	0.88	7.95
S		N/F	N/F	9	0.55	0.91	7.77
S		0.045	0.0013	5	0.70	1.03	7.69
A	4650–4320	N/F	N/F	6	0.78	1.01	8.71
A		0.018	0.004	4	0.93	0.90	6.73
S		N/F ^b	N/F ^b	10	0.59	0.87	6.95
S	4450–4320	0.045	0.0013	6	0.69	0.97	7.03
A		N/F	N/F	8	0.61	0.96	9.90
A		0.019	0.004	5	0.87	0.86	8.86
S	4450–4320	N/F	N/F	8	0.68	1.10	8.73
S		0.005	0.0013	6	0.74	1.03	7.23
A		N/F	N/F	7	0.74	1.06	13.38
A		0.019	0.005	7	0.86	1.18	13.12

^a S: single beam; A: absorbance.

^b N/F: no filtering.

TABLE VII. Asparagine calibration models in asparagine/glutamine binary mixtures.

Spectra type ^a	Spectra range (cm ⁻¹)	Mean position (<i>f</i>) ^b	Standard deviation (<i>f</i>) ^b	PLS factors	SEC (mM)	SEP (mM)	MPEP (%)
S	4800–4250	N/F	N/F	12	0.11	0.14	2.11
S		0.025	0.004	9	0.15	0.14	2.36
A		N/F	N/F	12	0.11	0.20	3.48
A		0.02	0.005	10	0.14	0.18	2.50
S	4700–4450	N/F	N/F	9	0.93	1.23	18.82
S		0.02	0.001	5	0.17	0.19	3.29
A		N/F	N/F	7	1.11	1.32	20.00
A		0.018	0.002	4	0.19	0.22	3.42
S	4650–4320	N/F	N/F	9	0.18	0.18	2.80
S		0.02	0.001	7	0.17	0.18	2.94
A		N/F	N/F	10	0.14	0.22	2.90
A		0.02	0.002	8	0.18	0.24	3.92
S	4450–4320	N/F	N/F	8	0.29	0.40	6.26
S		0.02	0.001	6	0.21	0.22	3.69
A		N/F	N/F	7	0.33	0.39	4.93
A		0.018	0.001	9	0.22	0.27	3.84

^a S: single beam; A: absorbance.

^b N/F: no filtering.

This trend is evident from the tabulated SEP values. Mean differences in SEP values from filtered spectra (single-beam minus absorbance) across all tested spectral ranges were 0.13 and 0.12 mM for glucose and glutamine models, respectively. Again, filtered spectra (single-beam and absorbance) yield superior model performance with less sensitivity to spectral range.

Measurement of Glutamine and Asparagine in Binary Mixtures. The last sample matrix tested corresponds to our recent effort to evaluate the selectivity of near-IR spectroscopy by attempting to differentiate glutamine and asparagine in aqueous solutions. These amino acids differ by only one methylene unit, which results in only minor spectral differences in the combination region. An analysis of PLS models based on absorbance spectra has been published for this data set.⁹ This analysis demonstrates sufficient selectivity to resolve these structurally similar compounds. The glutamine/asparagine data set is composed of 197 spectra collected from 66 unique standard solutions. Glutamine and asparagine concentrations were prepared randomly in each solution. A regression analysis shows no correlation between the concentrations of these species throughout the data set.⁹ Spectra associated with 16 standard solutions (48 spectra) were selected randomly for the prediction set. The remaining 149 spectra from 50 solutions were used in the calibration set. Description and performance of the best calibration models for asparagine and glutamine are listed in Tables VII and VIII, respectively. Prediction errors are essentially identical for models based on single-beam and absorbance spectra. SEP values are consistently lower when single-beam spectra are used, but these differences are small. In addition, no significant differences in prediction errors are indicated from models with filtered vs. unfiltered spectra.

CONCLUSION

The results presented in this paper demonstrate that valid calibration models can be generated from single-beam

TABLE VIII. Glutamine calibration models in asparagine/glutamine binary mixtures.

Spectra type ^a	Spectra range (cm ⁻¹)	Mean position (<i>f</i>) ^b	Standard deviation (<i>f</i>) ^b	PLS factors	SEC (mM)	SEP (mM)	MPEP (%)
S	4800–4250	N/F	N/F	11	0.09	0.08	1.71
S		0.035	0.008	9	0.11	0.09	2.39
A		N/F	N/F	9	0.12	0.10	2.19
A		0.02	0.003	8	0.13	0.10	2.25
S	4700–4450	N/F	N/F	9	0.89	1.14	22.89
S		0.02	0.003	7	0.15	0.12	3.22
A		N/F	N/F	9	0.70	1.27	31.58
A		0.02	0.002	8	0.16	0.12	2.49
S	4650–4320	N/F	N/F	10	0.11	0.09	1.79
S		0.03	0.005	9	0.12	0.09	1.81
A		N/F	N/F	8	0.15	0.12	2.42
A		0.02	0.003	8	0.13	0.10	2.00
S	4450–4320	N/F	N/F	7	0.18	0.22	4.46
S		0.02	0.001	7	0.13	0.10	2.27
A		N/F	N/F	6	0.21	0.22	4.83
A		0.018	0.002	6	0.15	0.12	2.51

^a S: single beam; A: absorbance.

^b N/F: no filtering.

near-IR spectra. Even though the analyte signals employed in this work represented only a small fraction of the variation present in the single-beam spectra, the digital filtering and multivariate calibration methods employed were still able to extract the analyte-dependent information and use it to build successful calibration models. These models were observed to predict analyte concentrations accurately under a variety of matrix conditions. This finding encourages further development of near IR spectroscopic methods for non-invasive *in vivo* monitoring where relevant reference spectra are not possible due to mismatches between the sample and reference matrices.

ACKNOWLEDGMENT

The financial support from the National Institutes of Health (Grant number DK 45126) is acknowledged.

1. M. A. Arnold and G. W. Small, *Anal. Chem.* **62**, 1457 (1990).
2. I. Aldhous, *Science* **258**, 892 (1992).
3. R. Marbach, TH. Koschinsky, F. A. Gries, and H. M. Heise, *Appl. Spectrosc.* **47**, 875 (1993).
4. R. E. Shaffer, G. W. Small, and M. A. Arnold, *Anal. Chem.* **68**, 2663 (1996).
5. K. H. Hazen, M. A. Arnold, and G. W. Small, *Appl. Spectrosc.* **48**, 477 (1994).
6. S. Pan, H. Chung, M. A. Arnold, and G. W. Small, *Anal. Chem.* **68**, 1124 (1996).
7. L. A. Marquardt, M. A. Arnold, and G. W. Small, *Anal. Chem.* **65**, 3271 (1993).
8. H. Chung, M. A. Arnold, M. Rhiel, and D. W. Murhammer, *Appl. Biochem. Biotech.* **50**, 109 (1995).
9. X. Zhou, H. Chung, M. A. Arnold, M. Rhiel, and D. W. Murhammer, in *Biosensor and Chemical Sensor Technology: Process Monitoring and Control*, K. R. Rogers, A. Mulchandani, and W. Zhou, Eds. (ACS Symposium Series 613 (ACS, Washington, D.C., 1995), Chap. 12, pp. 116–132).
10. H. Martens and T. Næs, *Multivariate Calibration* (Wiley, New York, 1989), Chap. 3.
11. G. W. Small, M. A. Arnold, and L. A. Marquardt, *Anal. Chem.* **65**, 3279 (1993).