Anal. Chem. **1996,** *68,* 1124−1135

Near-Infrared Spectroscopic Measurement of Physiological Glucose Levels in Variable Matrices of Protein and Triglycerides

Shengtian Pan,[†] Hoeil Chung,[‡] and Mark A. Arnold*

Department of Chemistry, University of Iowa, Iowa City, Iowa 52242

Gary W. Small

Center for Intelligent Chemical Instrumentation, Department of Chemistry, Clippinger Laboratories, Ohio University, Athens, Ohio 45701-2979

Selective calibration models are generated for glucose over the 1−20 nM concentration range by use of partial least-squares regression analysis of near-infrared spectra from 5000 to 4000 cm^{-1}. Two spectral data sets are used to simulate triglyceride and protein variations in clinical samples. Triacetin is used in one data set to simulate variations in triglyceride levels, and bovine serum albumin (BSA) is used in the second data set to simulate variations in blood protein levels. Although these matrix components possess strong absorption bands that overlap and overshadow the absorption bands of glucose, successful calibration models can be generated with no evidence of prediction bias caused by the different levels of the matrix components. Furthermore, the benefits of using digital Fourier filtering as a preprocessing step are evaluated in terms of calibration performance. The resulting calibration models provide standard errors of prediction of 0.5 and 0.2 mM in triacetin and BSA matrices, respectively. Accurate glucose predictions are demonstrated from spectra that correspond to protein concentrations not present in the calibration data set. Lastly, digital Fourier filtering alone is shown to have only limited ability to isolate glucose signals from those of BSA and triacetin due to similarities in the widths of the absorption bands of the three species.

The use of near-infrared (near-IR) spectroscopy for noninvasive clinical chemistry measurements is a fascinating possibility[1−7] and, if successfully implemented, would represent a major advancement in the field of analytical chemistry. This measurement approach is based on extracting analytical information from near-IR spectra collected directly from the body tissue of an individual. In principle, if the analyte of interest absorbs characteristic frequencies of near-IR light, the degree of absorbance can be related to analyte concentration through the Beer−Lambert relationship. For essentially all clinically relevant analytes except hemoglobin, near-IR absorption spectroscopy is based on relatively weak and broad overtone and combination bands associated with C−H, N−H, and O−H vibrational transitions. The development of a reliable clinical analysis based on measurements of these transitions is difficult, however, due to the small absorptivities associated with the bands, a strong and temperature-sensitive background absorbance of water, the highly scattering optical properties of human tissue, and the multitude of potential spectral interferences present.

To establish the foundation for actual noninvasive clinical measurements with human subjects, the analytical merits of clinical near-IR absorption spectroscopy must be evaluated under well-controlled laboratory conditions. The fundamental attributes and limitations of the method can only be established by identifying and controlling the key physical and chemical parameters that affect measurement accuracy and precision.

Toward this goal, our strategy has been to demonstrate the ability to measure a selected clinical analyte at physiologically relevant concentrations in relatively simple aqueous solutions and then increase the matrix complexity in a systematic manner. With glucose as the analyte of interest, our results have shown that near-IR spectra over the 5000−4000 cm^{-1} range contain sufficient information to permit accurate glucose measurements in the physiological range of 1−20 mM in phosphate buffer,[8] phosphate buffer with temperature variations from 32 to 41 °C,[9] and phosphate buffer with 60.8 g/dL bovine serum albumin (BSA).[10] In these previous studies, the chemical composition of the matrix was constant for each calibration model, and the ability to extract glucose information in the presence of reasonable levels of potential interferences was demonstrated. Furthermore, the

[†] Present address: College of Engineering, University of California, Riverside, CA 92521.

[‡] Present address: Yukong, Ltd., Ulsan Research Center, Chemical Process Research Laboratory, Ulsan, Korea.

(1) Amato, I. *Science* **1992**, *258*, 892−893.
(2) Sodickson, L. A.; Block, M. J. *Clin. Chem.* **1994**, *40*, 1838−1844.
(3) Robinson, M. R. Eaton, R. P.; Haaland, D. M. Koepp, G. Q.; Thomas, E. V.; Stallard, B. R.; Robinson, P. L. *Clin. Chem.* **1992**, *38*, 1618−1622.
(4) Marbach, R.; Koschinsky, Th.; Gries, F. A.; Heise, H. M. *Appl. Spectrosc.* **1993**, *47*, 875−881.
(5) Heise, H. M.; Marbach, R.; Koschinsky, Th.; Gries, F. A. *Artif. Organs* **1994**, *18*, 439−447.
(6) Jagemann, K. U.; Fischbacher, C.; Danzer, K.; Müller, U. A.; Mertes, B. *Z. Phys. Chem.* **1995**, *191*, 179−190.
(7) Arnold, M. A. New Developments and Clinical Impact of Noninvasive Monitoring. In *Handbook of Clinical Laboratory Automation, Robotics, and Knowledge Optimization*; Kost, G. R., Ed.; John Wiley & Sons: New York, 1996; Chapter 12, pp 631−647.
(8) Arnold, M. A.; Small, G. W. *Anal. Chem.* **1990**, *62*, 1457−1464.
(9) Hazen, K. H.; Arnold, M. A.; Small, G. W. *Appl. Spectrosc.* **1994**, *48*, 477−483.
(10) Marquardt, L. A.; Arnold, M. A.; Small, G. W. *Anal. Chem.* **1993**, *65*, 3271−3278.

0003-2700/96/0368-1124$12.00/0

ability to measure glucose in undiluted bovine plasma has been demonstrated.[11] In this work, three unique plasma matrices were used to prepare 69 glucose samples. Although the extent of matrix variation was small, no matrix-dependent systematic errors were detected in the analysis.

In a related series of experiments, we have demonstrated that glucose can be measured accurately in samples with matrix variation. In the first case, a series of binary mixtures of glucose and glutamine was examined, where the levels of glucose and glutamine varied independently from 1.7 to 59.9 and from 1.1 to 30.6 mM, respectively.[12] A five-component mixture was used in a subsequent experiment, where the levels of glucose, glutamine, glutamate, lactate, and ammonia−nitrogen were independently varied over a concentration range from 1 to 20 mM.[13] In both cases, calibration models could be generated for glucose with standard errors of prediction below 0.6 mM. In fact, acceptable calibration models were established for each of the varied species in these experiments.

In the matrix variation experiments summarized above, each of the varied species contributed significantly to the near-IR spectrum because concentrations and absorptivities were similar for each matrix component. Clinical samples present a more demanding situation, however, because matrix variations include endogenous species with much greater near-IR absorbances than those of glucose. Besides water, protein and triglycerides are the key endogenous components within clinical samples that exhibit larger absorbances than glucose in the near-IR region. The objective of the work presented here is to assess the ability to measure clinical levels of glucose under conditions where matrix variations due to model protein and model triglycerides cause large spectral changes relative to those from glucose. Two independent matrices have been studied. The first is composed of a phosphate buffer with five unique levels of triacetin, and the second is a phosphate buffer with 10 different BSA concentrations. Triacetin is used to simulate total triglyceride blood levels. Absorbance features for both triacetin and BSA are significantly greater than those of glucose, and there is considerable spectral overlap between the spectral bands.

EXPERIMENTAL SECTION

Apparatus and Reagents. Near-IR spectra were collected with a Nicolet 740 FT-IR spectrometer (Nicolet Analytical Instruments, Madison, WI), equipped with a 150 W tungsten−halogen source, a calcium fluoride beam splitter, and a cooled indium antimonide detector. The spectral range from 5000 to 4000 cm^{-1} was isolated with a multilayer optical interference filter (Barr Associates, Westford, MA).

Reagent grade glucose and potassium phosphate salts were purchased from common suppliers. Triacetin, BSA (fraction V, part no. A-4503), and 5-fluorouracil were used as received from Sigma Chemical Co. (St. Louis, MO). All solutions were prepared with reagent grade, type I water generated from a Milli-Q three-house purification unit.

Procedures. *Standard Solutions.* Phosphate buffers were prepared for both the triacetin and protein matrices. These buffer solutions contained 0.1 M phosphate, to which 0.044% 5-fluorouracil was added as a preservative.

For the triacetin matrix, the phosphate buffer was adjusted to pH 7.4, and five solutions were prepared by dissolving the appropriate amount of triacetin to achieve concentrations of 1564, 1933, 2320, 3093, and 3712 mg/L. These concentrations were selected to match typical plasma values of triglycerides for normal subjects.[14] Glucose standard solutions were prepared by dissolving dried glucose powder in each of these triacetin solutions. Glucose concentrations ranged from 2 to 20 mM at each triacetin level.

For the protein matrix, the buffer was adjusted to pH 7.2, and BSA was added to give a 95.0 g/L stock protein solution. A separate 0.5 M stock glucose solution was prepared in 0.1 M phosphate buffer. Glucose standards were then prepared by adding the required volumes of each stock solution to a 10 mL volumetric flask and then diluting with the 0.1 M phosphate buffer. Protein levels ranged from 47.5 to 90.25 g/L and glucose levels from 1.25 to 20.0 mM in order to cover typical physiological ranges. In all, 110 samples were prepared with 10 unique glucose concentrations for each of 10 different protein levels and 10 additional glucose solutions with zero protein.

Data Collection and Spectral Processing. Spectra were collected from solutions placed in a 1 mm path length sample cell composed of Infrasil quartz (Wilmad Glass Co., Buena, NJ). The standard solution was heated to 37.0 ± 0.2 °C before being placed in the sample cell. A thermostated glass jacket was placed around the cell to maintain the solution temperature at 37 °C. A thermocouple probe was submerged in the sample during data collection to ensure the sample temperature was constant. Spectra were generally collected in triplicate by leaving each sample in the spectrometer while three consecutive spectra were collected. In a few instances, only two spectra were obtained for a given sample.

Spectral information was collected as 256 coadded, double-sided interferograms with 16 384 points. The corresponding single-beam spectra were computed by use of the Nicolet SX-FTIR software package (version 4.4), and the resulting spectra possessed a point spacing of 1.9 cm^{-1}. Further spectral processing was accomplished with either a Vax 6400 computer located in the Gerard P. Weeg Computer Center at the University of Iowa or an Iris Indigo computer (Silicon Graphics, Inc., Mountain View, CA). All software was implemented in Fortran 77. Some calculations made use of subroutines from the IMSL software package (IMSL, Inc., Houston, TX).

For the triacetin data set, spectra were collected for all glucose concentrations at a given level of triacetin. The order of triacetin levels was 1933, 3093, 1564, 2320, and 3712 mg/L. Samples were selected randomly within a given triacetin level. Background spectra corresponded to 0.1 M phosphate buffer with no triacetin present. A single background spectrum was collected at the beginning of each new triacetin level. Ratioed spectra were subsequently computed by ratioing all sample single-beam spectra with a given triacetin level by the background spectrum collected immediately before these sample spectra were collected.

For the protein data set, samples were selected randomly with respect to both glucose and protein. Again, background spectra were collected from 0.1 M phosphate buffer (no protein present).

(11) Small, G. W.; Arnold, M. A.; Marquardt, L. A. *Anal. Chem.* **1993**, *65*, 3279−3289.

(12) Chung, H.; Arnold, M. A.; Rhiel, M.; Murhammer, D. W. *Appl. Biochem. Biotechnol.* **1995**, *50*, 109−125.

(13) Chung, H.; Arnold, M. A.; Rhiel, M.; Murhammer, D. W. *Appl. Spectrosc.* **1996**, *50*, 270−276.

(14) Naito, H. K. In *Methods in Clinical Chemistry*; Pesce, A. J., Kaplan, L. A., Eds.; C. V. Mosby Co.: St. Louis, MO; 1987; Chapter 153, pp 1215−1227.

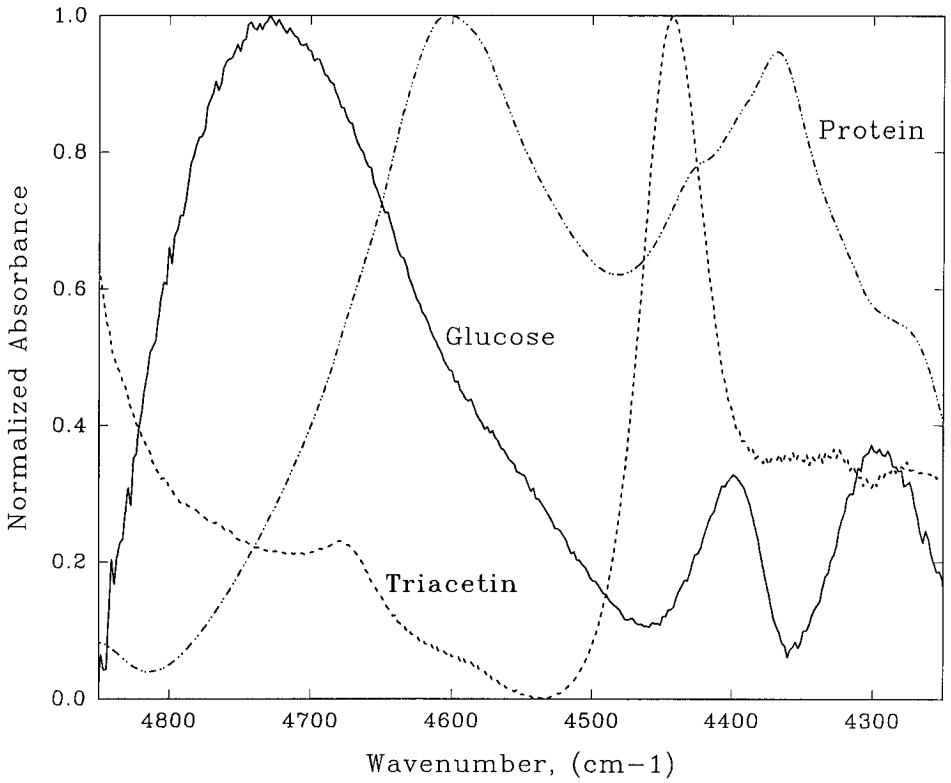

Figure 1. Near-IR absorbance spectra over the combination band region for glucose (10 mM), bovine serum albumin (66.5 g/L), and triacetin (2320 mg/L).

A background spectrum was collected at the beginning of the experiment and then after every fourth sample. This schedule effectively accounted for minor variations in interferometer alignment throughout the course of the data collection period. No studies were performed to determine the minimum frequency of background collection needed to reduce errors. The background spectrum collected before a set of four sample spectra was used to compute ratioed spectra.

For both data sets, replicate spectra corresponding to a given sample were kept together as the whole data set was split into calibration, monitoring, and prediction data sets. In all cases, the chemical distribution in the prediction data set tracked that in the calibration data set. Details of how these data sets were subdivided are provided below.

RESULTS AND DISCUSSION

Spectral Features. Triglycerides and total protein are examples of matrix components with large near-IR spectral features that overlap with those of glucose. Successful differentiation of glucose from triglycerides or albumin protein requires differences in their spectral absorption features. Differences in the location of these absorption bands can best be visualized by comparing normalized spectra. Figure 1 presents normalized absorbance spectra for 10 mM glucose, 2320 mg/L triacetin, and 66.5 g/L BSA, where normalization was achieved by scaling the maximum absorbance in each spectrum to unity. Glucose possesses three characteristic absorption bands centered at 4750, 4400, and 4300 cm^{-1}. The spectrum of triacetin is dominated by a strong band centered at 4450 cm^{-1}, with a smaller second band at 4680 cm^{-1}. This 4450 cm^{-1} triacetin band overlaps with the 4400 cm^{-1} glucose band, and, for equal molar concentrations of glucose and triacetin, the 4450 cm^{-1} triacetin band is 5.6 times larger than the 4400 cm^{-1} glucose band. This overlap and greater absorptivity are significant because the 4400 cm^{-1} glucose band provides the most reliable analytical information in aqueous solutions.[5−9] These findings suggest that information from other glucose bands will be needed for successful glucose measurements. The spectral features of BSA also overlap significantly with those of glucose. Furthermore, absorbances for the protein bands centered at 4370 and 4600 cm^{-1} are ∼100 times greater than that for the 4400 cm^{-1} glucose band. Absorbances of 58 and 60 mAU were measured for these two protein bands, respectively, for a 66.5 g/L solution of BSA. In comparison, the absorbance at 4400 cm^{-1} is only 0.48 mAU for a 10 mM glucose solution. Overall, the presence of glucose is difficult to identify in spectra collected from glucose/triacetin mixtures and virtually impossible to detect in spectra collected from glucose/protein mixtures.

The severe overlap and large absorptivities of these matrix components relative to glucose precludes the use of simple univariate calibration methods for extracting the glucose information from near-IR spectra. Indeed, multivariate calibration methods have been required in previous studies when the matrix was either constant[5−7] or varied only slightly.[8−10] In the work reported here, partial least-squares (PLS) regression analysis was used to correlate variations within the spectral data set with different glucose concentrations. In addition, the merits of using a digital Fourier filter before the PLS analysis were assessed for both data sets.

Glucose Measurements with Triacetin. *Constant Triacetin.* Before attempting to build valid calibration models for glucose with a variable triacetin matrix, studies were first undertaken to verify that glucose could be measured in the presence of a constant level of triacetin. The highest triacetin level (3712 mg/

Table 1. Results from PLS Glucose Calibrations with Variable Triacetin Levels

spectral range (cm^{-1})	glucose bands (cm^{-1})	PLS[a] factors	SEC (mM)	MPEP (%)	SEP (mM)	CV-SEP (mM)
4850−4250	4750, 4400, 4300	5	0.67	5.13	0.57	0.71
4850−4350	4750, 4400	5	0.73	6.84	0.61	0.77
4470−4250	4400, 4300	5	0.58	8.99	0.70	0.66
4850−4470	4750	8	0.60	7.38	0.68	0.70
4470−4350	4400	6	0.56	7.54	0.69	0.63
4350−4250	4300	9	0.58	9.14	0.92	0.76

[a] Optimum number of factors (lowest SEP).

L) was selected to maximize any adverse effects. The data set corresponding to 3712 mg/L triacetin consisted of 57 spectra from 19 samples. Spectra from 16 of these samples (48 spectra) were used to build calibration models, and spectra from the remaining three samples (nine spectra) were used for prediction to test the validity of the computed models. The absorbance spectra used throughout these studies were constructed by use of the single-beam spectra of the glucose samples and a background single-beam spectrum collected from 0.1 M phosphate buffer (no triacetin present).

Two spectral ranges were evaluated. The first (4850−4250 cm^{-1}) included all three glucose absorption bands, while the second (4470−4250 cm^{-1}) included only the 4400 and 4300 cm^{-1} bands. A series of calibration models were generated and tested for each spectral range by varying the number of PLS factors from 1 to 10. The optimum number of factors was identified as the number that gave the lowest standard error of prediction (SEP), as described in detail previously.[9−11]

The best calibration model with the 4850−4250 cm^{-1} spectral range was obtained with six PLS factors. This model possessed a standard error of calibration (SEC) of 0.41 mM, mean percent error of prediction (MPEP) of 3.0%, and a SEP of 0.29 mM. The best model for the narrower spectral range was obtained with seven PLS factors, and the error statistics were essentially the same as those from the wider range (SEC = 0.21 mM, MPEP = 1.8%, and SEP = 0.29 mM). In both cases, the SEC fell continuously as the number of PLS factors increased, and the SEP reached a minimum before increasing slightly as the system was overmodeled. This type of behavior is consistent with our earlier findings for other spectral data sets.[9−11] Overall, these results demonstrate that valid models can be prepared for glucose in a matrix with constant triacetin levels.

Variable Triacetin with PLS Alone. The entire data set, encompassing all five triacetin levels, consisted of 253 spectra collected from 86 samples. These spectra were divided randomly into calibration (208 spectra from 71 samples) and prediction (45 spectra from 15 samples) data sets. All replicate spectra from each sample were placed together in either the calibration or prediction sets. Six spectral ranges were examined, and the number of PLS factors was varied from one to 20 for each spectral range. Each spectral range was selected to incorporate different combinations of the three glucose bands (see Table 1).

Valid glucose calibrations were obtained with each of the six spectral ranges tested. Table 1 summarizes the calibration statistics for each range. Calibration performance was essentially the same for the first three entries in Table 1, which correspond to models incorporating multiple glucose bands. For each of these models, the SEP values ranged from 0.6 to 0.7 mM, and five PLS factors was found to be optimal. A comparison of measurement errors for the individual glucose bands indicates that the most reliable information is obtained from the 4400 cm^{-1} band, in spite of the significant overlap with the triacetin band. Calibration performance with only the 4400 cm^{-1} band was essentially the same as those from multiple-band models, but one additional PLS factor was required. Models based on either the 4750 or 4300 cm^{-1} band alone required even more factors to achieve nearly equivalent calibration performance.

Concentration correlation plots are presented in Figure 2 to illustrate the calibration and prediction data sets for the model based on the 4850−4250 cm^{-1} spectral range and five PLS factors. Glucose concentrations obtained from this model are plotted against known glucose concentrations for the standard solutions. In addition, the ideal unity line is shown for comparison. Results from the model correlate well with known values, which is particularly significant for the prediction data. Points in this figure are coded according to the triacetin level. Inspection of these coded points shows that different triacetin levels do not adversely affect the accuracy of the measurement. Figure 3 highlights this point by plotting percent error in glucose measurements from the prediction data set as functions of glucose and triacetin concentrations. Although a slight increase in percent error of prediction is evident as the concentration of glucose decreases, the increase in percent error does not track with the concentration of triacetin. Variable tiacetin levels do have an impact, however, relative to prediction errors from models with constant triacetin levels. The SEP increases from 0.29 to 0.57 mM when the sample matrix is changed from constant to variable triacetin. Nevertheless, a prediction ability of 0.57 mM is clinically acceptable.

As indicated in Table 1, for the first two spectral ranges, values of SEP are observed that are lower than the corresponding values for SEC. This is most likely an artifact of the availability of only 15 prediction samples. To provide a further validation test of the calibration models, a leave-one-out-cross-validation procedure was performed. In this study, the calibration and prediction sets were recombined to the full set of 86 samples (253 spectra). The replicate spectra of each sample were withheld from this set in turn, with a calibration model being constructed on the basis of the remaining spectra. The number of PLS factors used for each spectral range was set at the optimal value found previoulsy and indicated in Table 1. The resulting model was used to predict the concentrations of the spectra withheld. A pooled cross-validation SEP (CV-SEP) was computed on the basis of the results of the 86 sets of prediction spectra. The computed CV-SEP values are indicated in the last column of Table 1. As expected, these values exceed SEC in each case, although for four of the six spectral ranges, the increase is less than 15%. The cross-validation

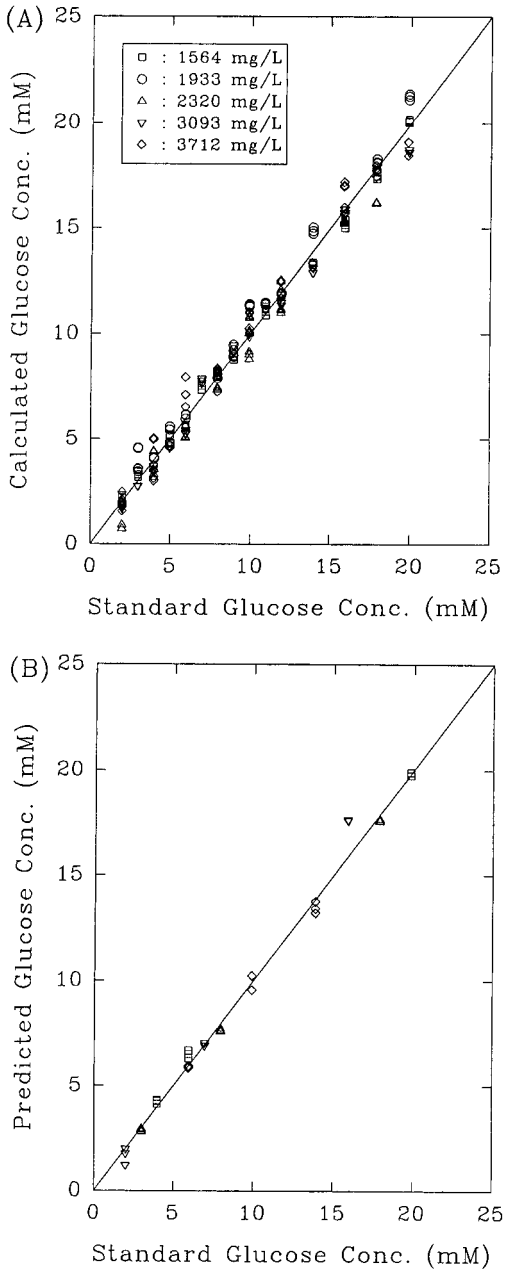

Figure 2. Concentration correlation plots for glucose showing (A) calibration and (B) prediction data for the five-factor PLS model (4850−4250 cm^{-1}) with triacetin levels of 1564 (□), 1933 (○), 2320 (△), 3093 (▽), and 3712 mg/L (◇).

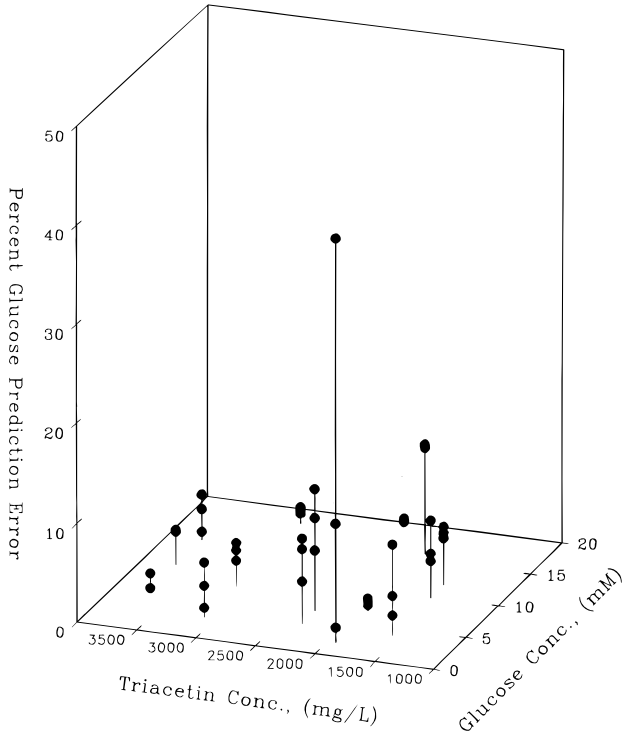

Figure 3. Interrelationship between glucose measurement errors and concentrations of glucose and triacetin.

results provide further evidence that the calibration model parameters found previously are valid and that the models yield accurate predictions.

Variable Triacetin with Digital Filtering and PLS. Significant analytical benefits have been demonstrated by incorporating a digital filtering step before the PLS analysis.[8−11] Digital Fourier filtering effectively reduces noise and baseline variation within the spectral data set, thereby enhancing spectral quality and improving analytical performance. Details for coupling this filtering concept with PLS regression analysis have been discussed.[11]

As in previous work, the bandpass Fourier filters used here were Gaussian shaped, with the specific filter bandpass function defined by the mean position and standard deviation of the Gaussian.[8−11] The optimum combination of these parameters was determined by evaluating PLS calibration models built after passing spectra through the appropriate filter. For this procedure, the spectra were divided randomly into three data sets, where one was used to establish the calibration model (calibration set), one was used to test the predictive ability of the model during the filter optimization procedure (monitoring set), and the last was used to evaluate the prediction ability following the completed process (prediction set). As before, all replicate spectra of a given sample were kept together in either the calibration, monitoring, or prediction set. Filter performance was evaluated by comparing calibration models in terms of the inverse of the sum of the mean square errors of calibration (MSEC) and prediction (MSEP), where the MSEP was based on the prediction errors for the monitoring set. For the measurement of glucose in a variable triacetin matrix, the calibration, monitoring, and prediction data sets consisted of 115, 93, and 45 spectra, respectively. The prediction set used here was the same as that employed in the analysis based on the use of PLS alone. Thus, the calibration and monitoring sets were subsets of the original calibration set used with PLS alone.

The best combination of mean position and standard deviation (SD) width was identified for the three spectral ranges corresponding to multiple analyte absorption bands (4850−4250, 4850−4350, and 4470−4250 cm^{-1}) and for the 4700−4200 cm^{-1} spectral range, which incorporates both low-frequency bands and part of the high-frequency band. For each spectral range, calibration models were evaluated with PLS factors ranging from 3 to 8, and the mean position and SD width were varied from 0.0 to 0.2*f* and from 0.0 to 0.02*f*, respectively, in a grid search procedure based on step sizes of 0.001*f*. These filter bandpass specifications are based on a linear scale of digital frequency (*f*) that varies between 0 and a maximum frequency of 0.5*f*. This optimization procedure resulted in 4221 (201 × 21) calibration model evaluations for each number of PLS factors and a total of 25 326 calibration evaluations

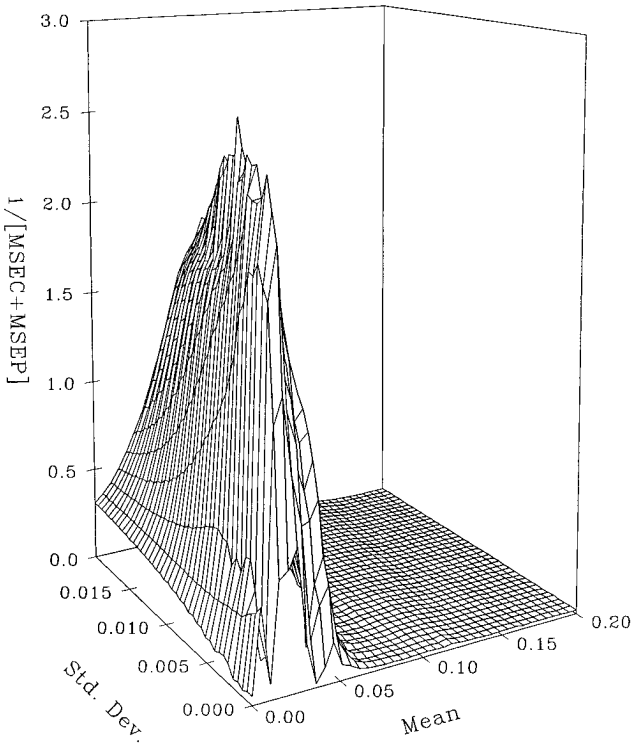

Figure 4. Glucose in triacetin Fourier filter optimization surface map using a five-factor PLS model over the 4700−4200 cm^{-1} spectral range.

for each spectral range. A typical result is shown graphically in the three-dimensional surface plot presented in Figure 4 for the 4700−4200 cm^{-1} spectral range with five PLS factors. The surface reveals a peak superimposed on a ridge of high values. The peak position gives the optimum filter parameters, which correspond to a mean position of 0.028*f* and a SD width of 0.006*f*. Similar optimum filter parameters were obtained for all spectral ranges tested. The best filter parameters and number of PLS factors are tabulated in Table 2 for each spectral range.

All calibration models based on Fourier filtering coupled with PLS regression gave approximately the same calibration performance regardless of the spectral range. Again, no indication of prediction bias caused by different levels of triacetin was noted upon examining concentration correlation plots for each calibration model listed in Table 2.

Based on the optimal spectral ranges, numbers of PLS factors, and filter parameters, the cross-validation procedure described previously was performed. The computed CV-SEP values are also listed in Table 2. For each spectral range, the value of CV-SEP exceeds the corresponding values of SEC by no more than 16.5%, with two of the four values under 10%. This result provides further validation for the optimized model parameters.

Comparison of values in Tables 1 and 2 reveals that Fourier filtering enhances calibration performance by providing lower values for SEC, SEP, MPEP, and CV-SEP. Effective glucose calibration models can be generated in a variable triacetin matrix, however, with or without Fourier filtering.

Glucose Measurements with Protein. Glucose measurements with constant protein have been established previously with 60.8 g/L BSA as the sample matrix.[10] In this previous work, glucose could be measured with an overall SEP of 0.24 mM over the clinically relevant concentration range. Such measurements were possible despite the strong absorptivity of BSA at 4600 and 4370 cm^{-1} and the severe overlap between the 4400 and 4370 cm^{-1} bands for glucose and protein, respectively.

For assessing the ability to measure glucose in a variable protein matrix, standard solutions were prepared with glucose concentrations of 1.25, 2.50, 3.75, 5.00, 7.50, 10.00, 12.50, 15.00, 17.50, and 20.00 mM in each of 11 buffered protein solutions. Buffered protein solutions consisted of 0, 47.5, 52.25, 57.0, 61.75, 66.5, 71.25, 76.0, 80.75, 85.5, and 90.25 g/L BSA. All spectra corresponding to the 0, 52.25, and 85.5 g/L protein matrices were removed from the data set and used subsequently to test the validity of the computed calibration models. Calibration, monitoring, and prediction data sets were assembled from spectra corresponding to the remaining 80 samples. For processing without Fourier filtering, 186 spectra from 64 samples were used as the calibration set, with the remaining 47 spectra from 16 samples serving as the prediction set. For the filter optimization studies, 46 spectra corresponding to 16 samples were removed from the calibration set and used as the monitoring set. The remaining 140 spectra from 48 samples were used as the calibration set.

Variable Protein with PLS Alone. Calibration models were constructed over several spectral ranges, and the number of PLS factors was varied from 1 to 20 for each spectral range. The four spectral ranges tested corresponded to (a) the full spectral range (5000−4000 cm^{-1}); (b) a spectral range that incorporated all three glucose bands but excluded the noisy spectral regions, where little light is transmitted (4900−4200 cm^{-1}); (c) a range that included part of the 4750 cm^{-1} glucose band but retained the two low-frequency bands for glucose and both protein bands (4600−4200 cm^{-1}); and (d) a range that contained only the 4400 and 4300 cm^{-1} glucose bands and one of the protein bands (4500−4200 cm^{-1}).

PLS calibration models computed over the 4900−4200, 4600−4200, and 4500−4200 cm^{-1} spectral ranges produced the characteristic patterns observed previously for SEC and SEP as the number of PLS factors was increased sequentially from 1 to 20. SEC values were observed to decrease continuously as more of the spectral variation was modeled, and SEP values reached a minimum and then increased as the system was overmodeled. Both SEC and SEP were observed to decrease rapidly over the first 4−6 factors before asymptotically approaching minimum values. Typically, SEC and SEP values were ∼6 mM with a one-factor model and <0.5 mM with models based on four factors. Calibration performance values are summarized in Table 3 for the best model with each spectral range. Also listed in Table 3 are the results for the same cross-validation experiment described previously for the triacetin data. The CV-SEP results are highly consistent with the corresponding SEC values, except for the 5000−4000 cm^{-1} spectral range.

For the 5000−4000 cm^{-1} range, the SEP continued to decrease as the number of factors was increased to 20. This spectral range includes considerable amounts of noise over the regions from 5000 to 4900 and from 4200 to 4000 cm^{-1} because of the strong absorbance of water. This inability to generate functioning calibration models when noisy regions are included within the spectral range is consistent with our earlier findings.[8,10]

Fourier Filtering and PLS Combined. Optimum Fourier filters were established for the following five spectral ranges: 4900−4200, 4600−4200, 4850−4460, 4460−4355, and 4355−4255 cm^{-1}. The 4900−4200 cm^{-1} range contains all three glucose bands, the

Table 2. Optimum Fourier Filter Parameters for Glucose in Triacetin

spectral range (cm^{-1})	PLS factors	mean (f)	SD (f)	SEC (mM)	MPEP (%)	SEP (mM)	CV-SEP (mM)
4850−4250	6	0.029	0.005	0.44	4.73	0.49	0.46
4700−4200	4	0.028	0.006	0.46	4.00	0.50	0.52
4850−4350	5	0.021	0.003	0.47	5.00	0.53	0.51
4470−4250	8	0.021	0.002	0.43	5.73	0.49	0.50

Table 3. PLS Models for Glucose in Variable Protein Matrix

spectral range (cm^{-1})	glucose bands (cm^{-1})	PLS[a] factors	SEC (mM)	MPEP (%)	SEP (mM)	CV-SEP (mM)
5000−4000	4750, 4400, 4300	>20	0.37[b]	12.1[b]	0.63[b]	0.65[b]
4900−4200	4750, 4400, 4300	6	0.43	3.6	0.23	0.42
4600−4200	1/4 4750, 4400, 4300	6	0.49	3.4	0.25	0.49
4500−4200	4400, 4300	5	0.54	5.2	0.32	0.55

[a] Optimum number of factors (lowest SEP). [b] Values obtained with 20 PLS factors.

Table 4. Best Filter Parameters and Calibration Performance with Variable Protein

spectral range (cm^{-1})	mean[a] (f)	SD[a] (f)	1/[MSEC + MSEP][a,b]	PLS[c] factors	SEC (mM)	SEP (mM)	CV-SEP (mM)
4900−4200	0.026	0.005	4.78	6	0.38	0.23	0.39
4600−4200	0.021	0.003	5.15	4	0.41	0.22	0.39
4850−4460	0.019	0.002	4.81	11	0.36	0.23	0.37
4460−4355	0.019	0.001	4.45	8	0.39	0.23	0.38
4355−4255	0.017	0.001	3.02	6	0.40	0.25	0.39

[a] Computed with four PLS factors. [b] Maximum value. [c] Optimum number of factors (lowest SEP) with given filter parameters.

4600−4200 cm^{-1} range contains primarily the 4400 and 4300 cm^{-1} glucose bands, and the last three ranges correspond to the single glucose bands at 4750, 4400, and 4300 cm^{-1}, respectively. The best combination of mean position and SD width was established as described above for the variable triacetin matrix. Four PLS factors were used in each evaluation, and the mean and standard deviation were systematically varied over a range from 0.0 to 0.05f with 0.001f step sizes. This corresponds to 2601 (51 × 51) evaluations for each spectral range. Results of this experiment are presented in Table 4.

A more detailed investigation was performed to find the best filter parameters for the 4600−4200 cm^{-1} spectral range. In the first experiment, greater resolution was used to refine the optimal parameter settings. The tested ranges were 0.01−0.03f for the mean position and 0.0−0.006f for the SD width, with step sizes of 0.0005f and 0.0001f, respectively. Optimum parameters were 0.0205f and 0.0029f for the mean and standard deviation, respectively, which match those found when the wider ranges and larger step sizes were used.

In addition, the number of PLS factors used during the filter optimization procedure was varied for the 4600−4200 cm^{-1} spectral range. Optimum filter parameters were established with 1, 2, 3, 4, 5, 6, 7, 8, 11, and 15 PLS factors. Surface maps for the results with 3, 4, 5, and 11 factors are presented in Figure 5. The optimum filter parameters are essentially identical when 1−8 factors are used. As the number of factors increases beyond 8, however, the region behind the normal ridge structure begins to dominate, and different filter parameters are obtained (0.013f, 0.003f for 11 factors and 0.006f, 0.004f for 15 factors). This analysis confirms our previously published results with the constant protein matrix[10] and demonstrates the need to limit the number of factors used in the filter optimization procedure. Values for the response function (1/[MSEC + MSEP]) increase as the number of factors increases from 1 to 4. This value levels off after four factors and then increases again as the number of factors approaches 11. This pattern of values for the response function indicates that the first four PLS factors provide most of the glucose-dependent analytical information, and spectral noise is modeled after the eleventh factor.

The best calibration model obtained with a combination of Fourier filtering and PLS regression corresponds to the 4600−4200 cm^{-1} spectral range with four PLS factors and a Fourier filter defined by a mean position of 0.0205f and SD width of 0.0029f (see Table 4). Concentration correlation plots for both calibration and prediction data sets are provided in Figure 6 for this best model. Protein levels have been denoted in the figure by coding data points corresponding to low (47−57 g/L), medium (61−71 g/L), and high (76−90 g/L) protein concentrations. These symbols overlap at all glucose levels, which indicates no bias in accuracy caused by BSA protein. Analysis of glucose residuals confirms this conclusion. Plots of glucose residuals as functions of glucose and protein concentrations reveal that these residuals are randomly distributed along the zero deviation line. Residuals are independent of glucose and protein concentrations, which indicates no interference by protein. The mean glucose residual is 0.17 mM across all protein concentrations for the prediction data set.

Cross-validation experiments were also performed for the calibration models that incorporated digital filtering. The resulting CV-SEP values are listed in Table 4. The model parameters appear quite valid, as the maximum difference between corresponding values of CV-SEP and SEC is <5%.

Comparison of Models from PLS Alone and PLS with Fourier Filtering. Values in Tables 3 and 4 reveal that there is little difference in the calibration performance with and without the additional Fourier filtering step. The best models are obtained

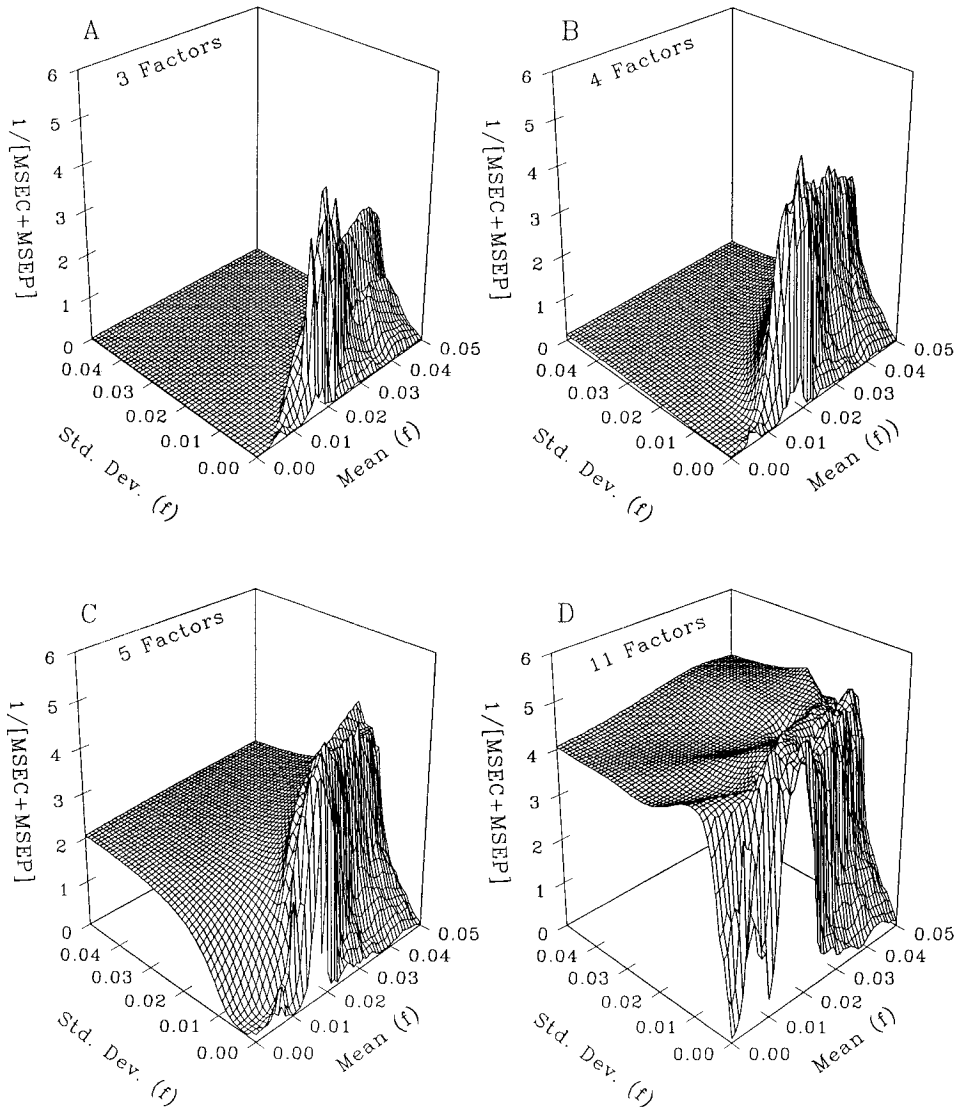

Figure 5. Fourier filter optimization surface maps for glucose with protein using (A) 3, (B) 4, (C), 5, and (D) 11 PLS factors with the 4600−4200 cm^{-1} spectral range.

with spectral ranges of 4600−4200 and 4900−4200 cm^{-1} for algorithms with and without Fourier filtering, respectively. Without filtering, the 4600−4200 cm^{-1} spectral range provides an essentially equivalent model, and noticeably worse models are obtained with both the 5000−4000 and 4500−4200 cm^{-1} ranges. With filtering, however, model performance is essentially independent of spectral range. The same values for SEC and SEP are obtained even when ranges incorporating single glucose bands are used. Furthermore, a Fourier filter/PLS model based on the entire spectral range (5000−4000 cm^{-1}), a Fourier filter defined by a mean of 0.0205*f* and standard deviation of 0.0029*f*, and 10 PLS factors results in SEC and SEP values of 0.39 and 0.25 mM, which are essentially equivalent to the values listed in Table 4. This performance is strikingly different compared to the model obtained over this spectral range without the use of Fourier filtering. The additional filtering step effectively eliminates noise associated with low light levels at the outer regions of the spectrum, thereby enhancing calibration performance.

A potential limitation in the analysis of the calibration models described above is the fact that all protein levels in the prediction data set are represented in both the calibration and monitoring data sets. A more rigorous test of protein interference is to predict glucose concentrations from spectra corresponding to protein concentrations not used in the calibration procedure. Such a test was used to evaluate the accuracy of models computed with and without Fourier filtering. Spectra were collected for this purpose with 10 glucose levels that ranged from 1.25 to 20 mM and protein levels of 0, 52.25, and 85.5 g/L. Glucose values were predicted from these spectra with calibration models based on both PLS alone and PLS coupled with Fourier filtering. In both cases, the 4600−4200 cm^{-1} spectral range was used with the parameters specified in Tables 3 and 4.

Accurate glucose predictions were achieved from both calibration models at each of the tested protein levels. The corresponding correlation plots are presented in Figure 7. Computed SEP values across all 10 glucose concentrations in matrices with 0, 52.25, and 85.5 g/L BSA were 0.70, 0.30, and 0.53 mM, respectively, for the model based on PLS alone and 0.90, 0.24, and 0.38 mM, respectively, for the model based on PLS with Fourier filtering. Prediction errors with 52.25 and 85.5 g/L BSA are similar to the SEC and SEP values listed in Tables 3 and 4. Prediction errors from spectra without protein are slightly larger than the others, which is not surprising because zero protein

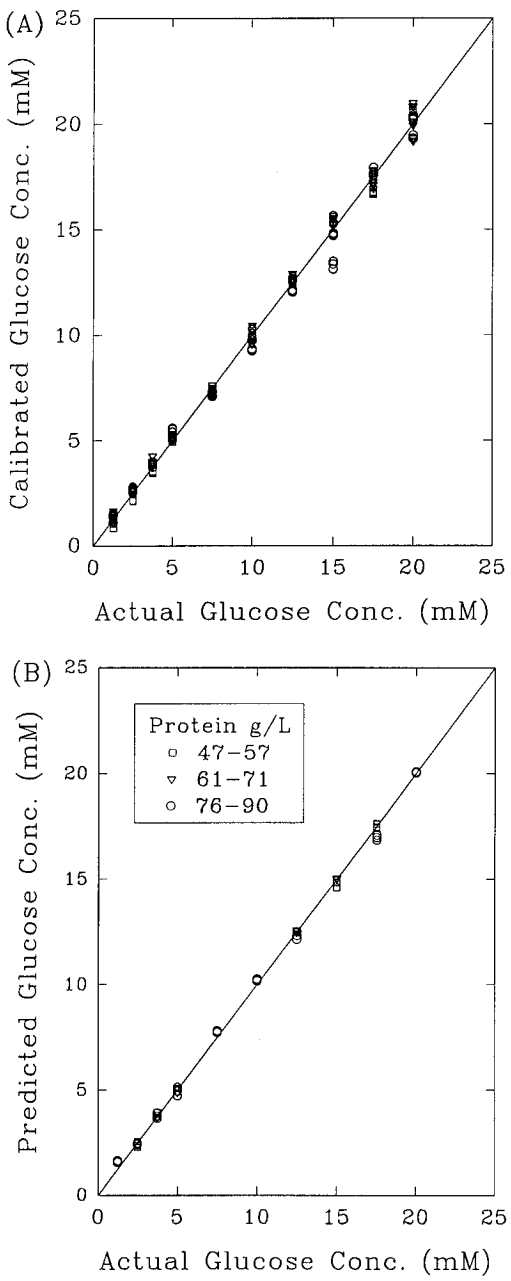

Figure 6. Concentration correlation plots for glucose showing (A) calibration and (B) prediction data for a four-factor PLS model (4600−4200 cm^{-1}) coupled with Fourier filtering (mean = 0.021*f*, SD = 0.003*f*) with low (□), medium (▽), and high (○) protein levels.

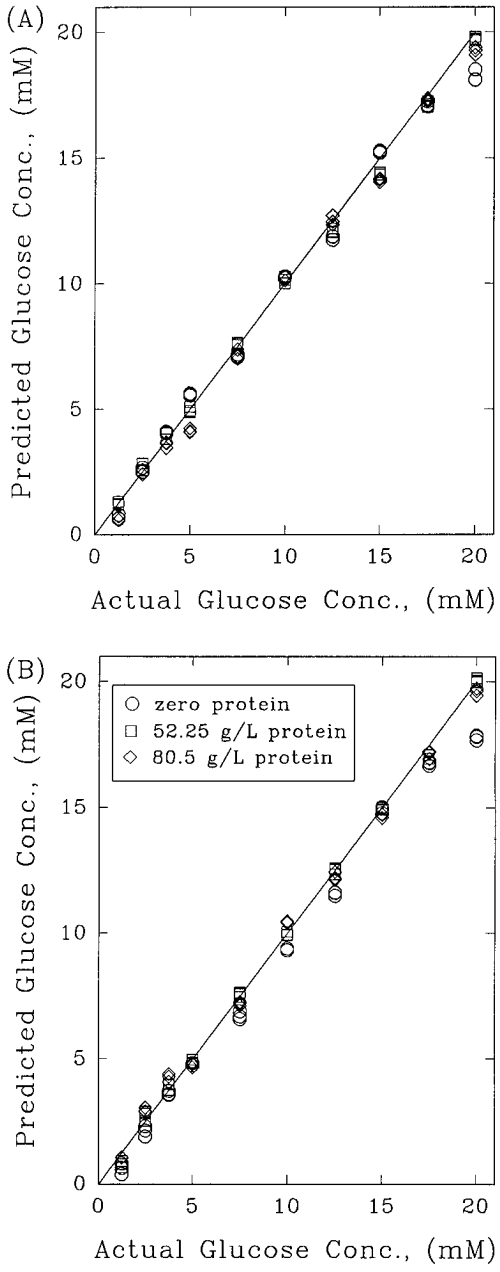

Figure 7. Concentration correlation plots for glucose predictions from spectra corresponding to samples with protein levels different from those used in the calibration data set. Graph A corresponds to predictions from the four-factor PLS model (4600−4200 cm^{-1}) without Fourier filtering, and graph B corresponds to predictions from the four-factor PLS model (4600−4200 cm^{-1}) with Fourier filtering with protein levels of 0 (○), 52.25 (□), and 80.5 g/L (◇).

requires extrapolation to a condition not encompassed within the calibration data. Prediction ability in this experiment is essentially equivalent for the two calibration models.

Chemical Differentiation by Fourier Filtering. In our work with measuring clinically relevant levels of glucose in different aqueous matrices, similar Fourier filters have been obtained. Optimum values for filter parameters with the variable protein matrix are similar for each spectral range tested (see Table 4). In addition, these values are essentially the same as those for the variable triacetin matrix (see Table 2). Similar values are published for matrices of fixed protein[10] and for phosphate buffer with different temperatures.[9] In all cases, the mean position is ~0.02*f*, while the SD width is much smaller, ranging from 0.007*f* to 0.001*f*.

The similarity of these filter bandpass parameters raises the question of whether the parameters are specific for glucose in aqueous solutions. Futhermore, can Fourier filtering selectively distinguish different chemical species on the basis of different bandwiths? Fourier-filtered near-IR spectra have been analyzed to begin addressing these fundamental questions.

The presence of glucose-dependent information in near-IR spectra collected from buffered protein solutions is evident when the effect of protein is removed by use of a background spectrum collected from a glucose-free buffered protein solution. Figure 8A shows a series of five near-IR spectra that have been filtered with a Gaussian-shaped filter bandpass defined by a mean position of 0.0205*f* and standard deviation of 0.0029*f*. In this case, the original absorbance spectra correspond to different glucose concentrations with 47.5 g/L BSA in both the sample and

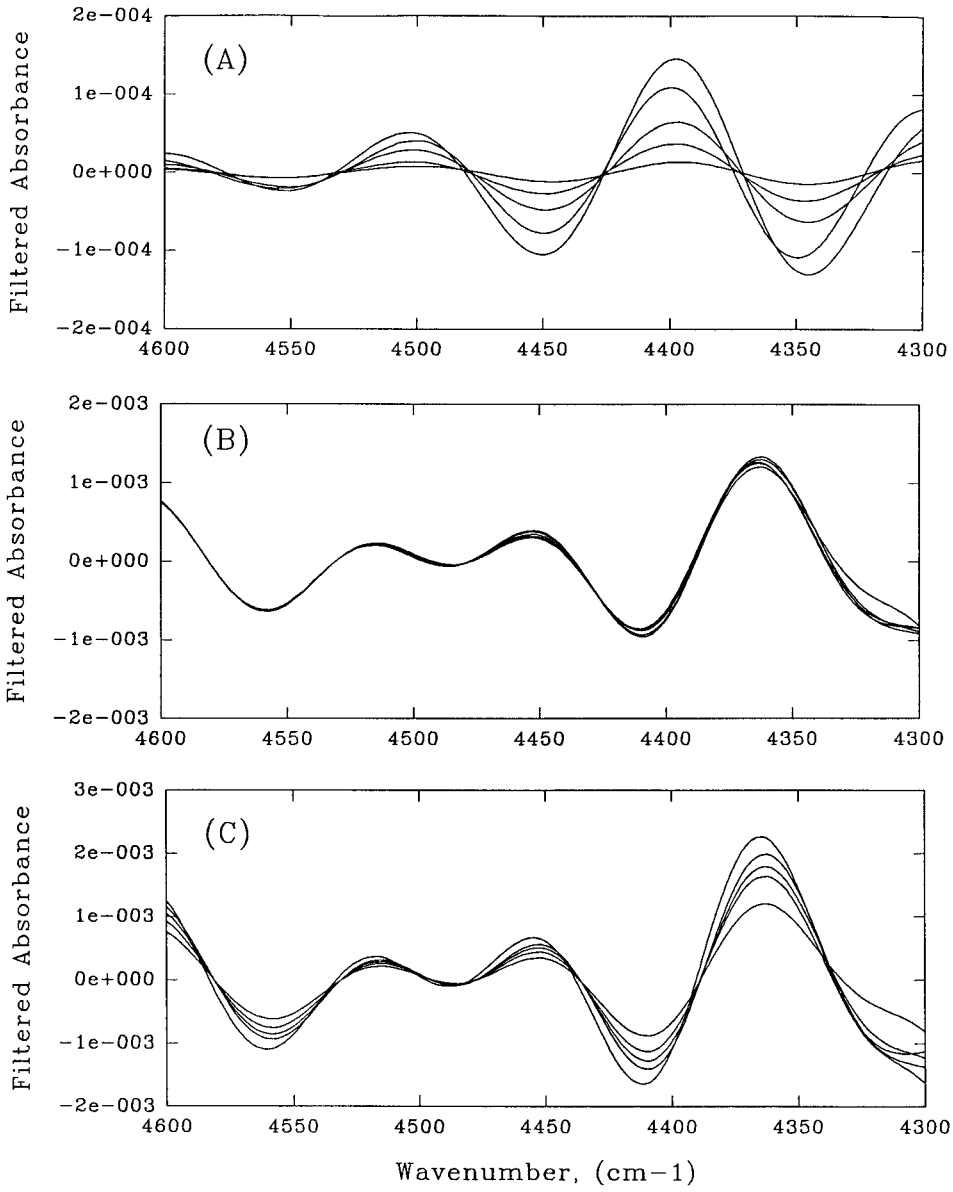

Figure 8. Effect of Fourier filtering (mean = 0.0205*f*; SD = 0.0029*f*) on spectra with (A) different glucose levels, constant protein (47.5 g/L), and protein in the reference solution; (B) different glucose levels, constant protein (47.5 g/L), and 0.1 M phosphate buffer as the reference solution; and (C) constant glucose (10 mM), different protein levels, and 0.1 M phosphate buffer as the reference solution. Order of spectra from bottom to top at 4400 cm^{-1} for A and 4370 cm^{-1} for B and C corresponds to glucose concentrations of 2.5, 5.0, 10, 15, and 20 mM for A; glucose concentrations of 10, 2.5, 20, 5.0, and 15 mM for B; and BSA concentrations of 47.5, 57.0, 66.5, 76.0, and 85.5 g/L for C.

background solutions. As noted before, these filtered spectra resemble derivative spectra.[11] Fourier-filtered spectra for glucose are characterized by a dominant feature centered at 4400 cm^{-1}, which corresponds to the middle glucose absorption band in Figure 1. As shown in Figure 8A, the magnitude of this feature is on the order of tenths of milliabsorbance units for millimolar levels of glucose, and, as expected, this magnitude correlates with glucose concentration.

Fourier-filtered spectra are significantly different when protein is not included in the background spectrum. Figure 8B shows filtered spectra generated from the same sample single-beam spectra used in Figure 8A but with a protein-free phosphate buffer as the background spectrum (as used in the calibration work described above). The resulting features are considerably different compared to Figure 8A. The predominant feature has shifted from 4400 to 4370 cm^{-1}. In addition, the magnitude of this main feature is ~10 times larger than before, and it does not correlate with glucose concentration. Figure 8C shows Fourier-filtered spectra where the glucose level has been held constant at 10 mM and the protein level varies. Clearly, the spectral features in Figure 8B and C correspond to protein and not glucose.

Analysis of spectra collected from mixtures of glucose and triacetin reveals similar results.[15] The predominant features in glucose/triacetin spectra after Fourier filtering correspond to triacetin. The largest feature is centered at 4450 cm^{-1}, and the magnitude of this feature depends on triacetin levels and is independent of glucose concentration.

The principal conclusion from the above spectral analysis is that the Fourier filters used in this work, and identified as optimal

(15) Chung, H. Ph.D. Dissertation, University of Iowa, Iowa City; IA, 1994.

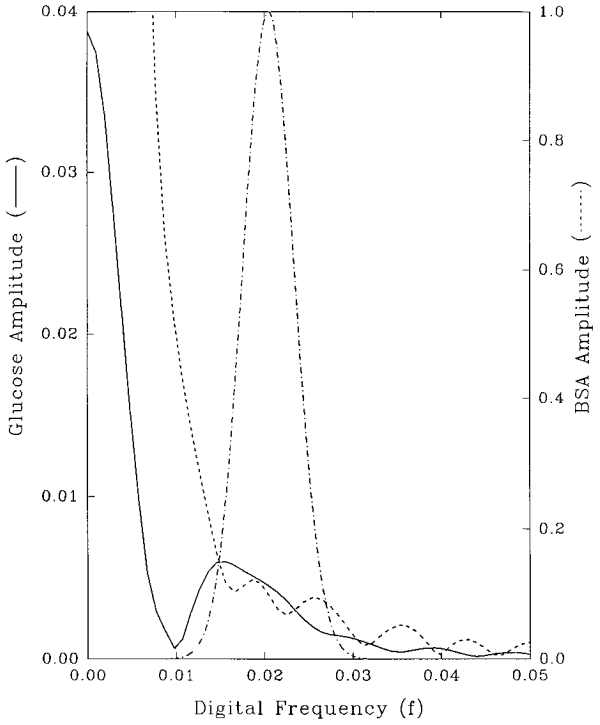

Figure 9. Fourier domain amplitude spectra of glucose (solid line) and BSA (dashed line), with the optimal Gaussian-shaped filter bandpass superimposed (dash-dot line). The glucose and BSA absorbance spectra displayed in Figure 1 were used to generate these amplitude spectra.

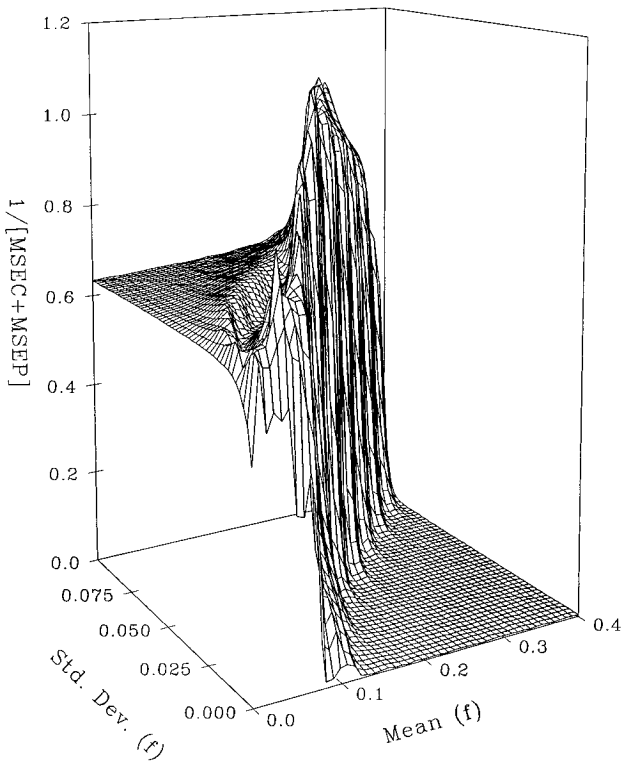

Figure 10. Fourier filter optimization surface map for two-factor PLS model (4800−4200 cm^{-1}) for protein measurements.

for glucose, cannot selectively distinguish glucose from chemical interferences such as BSA and triacetin. As illustrated by Figure 8, significant amounts of signal due to the interfering species pass through such filters. This places the burden of selectivity on the ability of the PLS algorithm to separate overlapping spectral features. Figure 9 illustrates this point further by providing a comparison of the Fourier domain signals arising from the glucose and BSA spectra plotted in Figure 1. To generate Figure 9, the ranges of 4600−4200 cm^{-1} in the glucose and BSA absorbance spectra were windowed with a Hamming function, zero-filled to 1024 points, and Fourier transformed. Over the range of 0.0−0.05*f*, Figure 9 plots the amplitude (magnitude) spectra resulting from this operation for glucose (solid line) and BSA (dashed line). The dash-dot line in the figure plots the same Gaussian-shaped filter bandpass used in generating Figure 8. Through the application of the filter, the signals at digital frequencies outside this bandpass are eliminated from the absorbance spectrum. Due to the presence of much greater absorbance values in the BSA spectrum and only slightly different spectral bandwidths between glucose and BSA, the Fourier domain spectrum of BSA severely overlaps and swamps the glucose spectrum. While only ∼3% of the total area of the Fourier domain spectrum of BSA passes through the filter (versus 15% of the glucose spectrum), the large BSA absorbance still dominates the filtered spectrum.

Although these Fourier filters do not selectively pass individual compounds, different compounds with different spectral bandwidths do yield different filter bandpass parameters in order to obtain an optimal response. For example, the spectral data set described above for measurement of glucose in a variable protein matrix has been used to identify optimal Fourier filter parameters for the measurement of protein. The ideal mean position and SD width for protein measurements are 0.17*f* and 0.0046*f*, respectively. The corresponding surface map for protein is presented in Figure 10. The surface map for protein shows a ridge-shaped feature similar to that for glucose, but the best filtering parameters are considerably larger than those identified for glucose. The Fourier filtering step is of little value for the measurement of protein because of the strong absorbances involved. Indeed, a two-factor PLS model constructed without the use of filtering provides the same calibration performance as a two-factor PLS model with filtering. The values of SEP for both models were 1.1 g/L, and the MPEP values were 1.2 and 1.3%, respectively.[16]

CONCLUSIONS

The results presented above demonstrate the ability to measure glucose with acceptable clinical accuracy in matrices with variable levels of triacetin and albumin protein. Glucose prediction errors of 0.5 and 0.2 mM are common for calibration models with variable triacetin and protein, respectively. Such measurements are possible in spite of the strong overlap between the key glucose absorption features and those from the matrix components. These results continue to lend confidence to the feasibility of using near-IR spectroscopy for both routine and noninvasive clinical measurements. The matrix variations used in this work represent those of normal plasma values. Additional experiments are needed to assess performance under abnormal ranges of proteins and triglycerides.

The primary benefits of digital Fourier filtering on the analysis are reductions of both high-frequency noise and baseline variations. As implemented in our work, Fourier filtering is not capable of selectively extracting information pertaining to one analyte over another when the spectral bandwidths are similar, such as for

(16) Pan, S. Ph.D. Dissertation, University of Iowa, Iowa City, IA, 1995.

glucose, BSA, and triacetin. In general, Fourier filtering does reduce the dependency of the calibration model on spectral range and the number of PLS factors required to achieve the best calibration model.

ACKNOWLEDGMENT

This work was supported by the National Institutes of Health under Grant DK45126. Partial funding for the purchase of the Nicolet 740 spectrometer was provided by the National Institutes of Health under Grant RR04583.

Received for review July 25, 1995. Accepted January 19, 1996.[⊗]

AC950751X

[⊗] Abstract published in *Advance ACS Abstracts,* March 1, 1996.