

Determination of Glucose in a Biological Matrix by Multivariate Analysis of Multiple Band-Pass-Filtered Fourier Transform Near-Infrared Interferograms

Mutua J. Mattu[†] and Gary W. Small*

Center for Intelligent Chemical Instrumentation, Department of Chemistry, Ohio University, Athens, Ohio 45701

Mark A. Arnold

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

A multivariate calibration method is described in which Fourier transform near-infrared interferogram data are used to determine clinically relevant levels of glucose in an aqueous matrix of bovine serum albumin (BSA) and triacetin. BSA and triacetin are used to model the protein and triglycerides in blood, respectively, and are present in levels spanning the normal human physiological range. A full factorial experimental design is constructed for the data collection, with glucose at 10 levels, BSA at 4 levels, and triacetin at 4 levels. Gaussian-shaped band-pass digital filters are applied to the interferogram data to extract frequencies associated with an absorption band of interest. Separate filters of various widths are positioned on the glucose band at 4400 cm⁻¹, the BSA band at 4606 cm⁻¹, and the triacetin band at 4446 cm⁻¹. Each filter is applied to the raw interferogram, producing one, two, or three filtered interferograms, depending on the number of filters used. Segments of these filtered interferograms are used together in a partial least-squares regression analysis to build glucose calibration models. The optimal calibration model is realized by use of separate segments of interferograms filtered with three filters centered on the glucose, BSA, and triacetin bands. Over the physiological range of 1–20 mM glucose, this 17-term model exhibits values of R², standard error of calibration, and standard error of prediction of 98.85%, 0.631 mM, and 0.677 mM, respectively. These results are comparable to those obtained in a conventional analysis of spectral data. The interferogram-based method operates without the use of a separate background measurement and employs only a short section of the interferogram.

Near-infrared (near-IR) spectroscopy is currently being evaluated as a viable measurement option for the determination of clinically relevant analytes such as blood or tissue glucose.^{1–5} Our

laboratories have demonstrated the successful hyphenation of spectral preprocessing techniques such as digital filtering and multivariate calibration methods for the selective determination of glucose in a variety of matrices.^{6–12} A drawback to most of the previously reported work is that the data analysis protocols require the use of a background spectrum to remove spectral components of the sample matrix. In many of the desired applications of near-IR spectroscopy to glucose measurements, however, collection of a representative background spectrum of the sample matrix is difficult.

One approach to eliminating the need for a background measurement is to use the single-beam spectra of the samples directly in building calibration models. Several successful studies that used this approach have been reported.^{13–15} The work presented here is motivated similarly by the need to eliminate the collection of a background or reference spectrum for use in extracting glucose information and by a desire to simplify the instrumentation required for the glucose measurement.

The current work is based on the construction of glucose calibration models by direct analysis of the interferogram data collected with a Fourier transform spectrometer. In addition to eliminating the need for a background measurement, by implementing the analysis with a short segment of the interferogram, a simpler (i.e., lower resolution) interferometer design can also be implemented. This results in a decrease in both the data acquisition and data processing requirements for the measurement, resulting in a potentially faster, more rugged, and less

[†] Present address: Instrumentation Metrics, Inc., 2085 Technology Circle, Suite 102, Tempe, AZ 85284.

(1) Arnold, M. A. In *Clinical Laboratory Automation, Robotics and Knowledge Optimization*; Kost, G. J., Ed.; Wiley: New York, 1995.
(2) Robinson, M. R.; Eaton, R. P.; Haaland, D. M.; Koeppe, G. W.; Thomas, E. V.; Stallard, B. R.; Robinson, P. L. *Clin. Chem.* **1992**, *38*, 1918–1922.
(3) Marbach, R.; Koschinsky, Th.; Gries, F. A.; Heise, H. M. *Appl. Spectrosc.* **1993**, *47*, 875–881.

(4) Heise, H. M. In *Biosensors in the Body: Continuous in vivo Monitoring*; Fraser, D. M., Ed.; Wiley: Chichester, U.K., 1997; Chapter 3.
(5) Müller U. A.; Mertes B.; Fischbacher C.; Jageman K. U.; Danzer K. *Int. J. Artif. Organs* **1997**, *20*, 285–290.
(6) Arnold, M. A.; Small, G. W. *Anal. Chem.* **1990**, *62*, 1457–1464.
(7) Hazen, K. H.; Arnold, M. A.; Small, G. W. *Appl. Spectrosc.* **1994**, *48*, 477–483.
(8) Pan, S.; Chung, H.; Arnold, M. A.; Small, G. W. *Anal. Chem.* **1996**, *68*, 1124–1135.
(9) Marquardt, L. A.; Arnold, M. A.; Small, G. W. *Anal. Chem.* **1993**, *65*, 3271–3278.
(10) Small, G. W.; Arnold, M. A.; Marquardt, L. A. *Anal. Chem.* **1993**, *65*, 3279–3289.
(11) Bangalore, A. S.; Shaffer, R. E.; Small, G. W.; Arnold, M. A. *Anal. Chem.* **1996**, *68*, 4200–4212.
(12) Shaffer, R. E.; Small, G. W.; Arnold, M. A. *Anal. Chem.* **1996**, *68*, 2663–2675.
(13) Bittner, A.; Marbach, R.; Heise, H. M. *J. Mol. Struct.* **1995**, *349*, 341–344.
(14) Heise, H. M.; Bittner, A. *J. Mol. Struct.* **1995**, *348*, 127–130.
(15) Lu, G.; Zhou, X.; Arnold, M. A.; Small, G. W. *Appl. Spectrosc.* **1997**, *51*, 1330–1339.

expensive instrument. In this paper, an interferogram-based approach is developed for application to the measurement of glucose in a simulated biological matrix of bovine serum albumin (BSA), triacetin, and phosphate buffer.

EXPERIMENTAL SECTION

Instrumentation and Apparatus. Interferograms used in this work were collected with a Digilab FTS-60A Fourier transform spectrometer (Bio-Rad, Cambridge, MA). The spectrometer was operated with a 100-W tungsten-halogen lamp, CaF₂ beam splitter, and InSb detector. The wavelengths of radiation striking the detector were restricted to the range of 5000–4000 cm⁻¹ by use of a K-band interference filter (Barr Associates, Westford, MA). Samples were placed in a 2-mm path length rectangular cell composed of Infrasil quartz, and sample temperatures were maintained between 37 and 38 °C through the use of a water-jacketed cell holder and refrigerated temperature bath.

Reagents. Reagent-grade glucose, sodium phosphate salts, triacetin, and 5-fluorouracil were purchased from common suppliers. The BSA used was a Cohn fraction V powder obtained from Sigma Chemical Co. (St. Louis, MO, product no. A 4503). All solutions were prepared with reagent-grade water obtained from a Milli-Q Plus water purification system (Millipore, Inc., Bedford, MA).

Procedures. A factorial design approach was adopted for the collection of the data set employed in this work. The full data set consisted of standards prepared from a combination of glucose at 10 levels spanning the range 1–20 (1, 3, 5, 7, 9, 11, 13, 15, 17, 19) mM, triacetin at 4 different levels spanning the range 1.4–3.5 (1.4, 2.1, 2.8, 3.5) g/L, and BSA at 4 levels spanning the range 50–95 (50, 65, 80, 95) g/L. These solutions were prepared in a pH 7.4 buffer solution that contained 0.1 M phosphate and 0.05% (w/w) 5-fluorouracil added as a preservative. A total of 160 (10 × 4 × 4) standards were prepared. The preparation and data collection of the various standards were randomized to eliminate the possibility of chance correlations between glucose concentration and time-dependent data artifacts.

Three replicate interferograms were collected sequentially without removing a given sample from the spectrometer. Single-sided interferograms of 16 384 points were collected and Fourier processed with triangular apodization and Mertz phase correction to yield single-beam spectra with a nominal point spacing of 2 cm⁻¹. The interferograms were sampled at each zero crossing of the HeNe reference laser, defining a maximum digitized frequency of 15 800.8 cm⁻¹. Periodically, reference interferograms of pure phosphate buffer were collected for use in computing spectra in absorbance units. All Fourier processing was performed with software resident on the Bio-Rad SPC-3200 computer used to control the spectrometer. Each replicate interferogram consisted of 256 coadded scans. This data collection procedure produced a total of 160 × 3 = 480 interferograms/spectra for analysis.

The interferograms and computed spectra were transferred to a Silicon Graphics 4D/460 computer operating under Irix (version 4.0.5, Silicon Graphics, Inc., Mountain View, CA). All computer software used in processing the interferogram and spectral data on the Silicon Graphics system was written in FORTRAN 77. Fourier filtering and multiple linear regression computations were performed with subroutines from the IMSL software package (IMSL, Inc., Houston, TX).

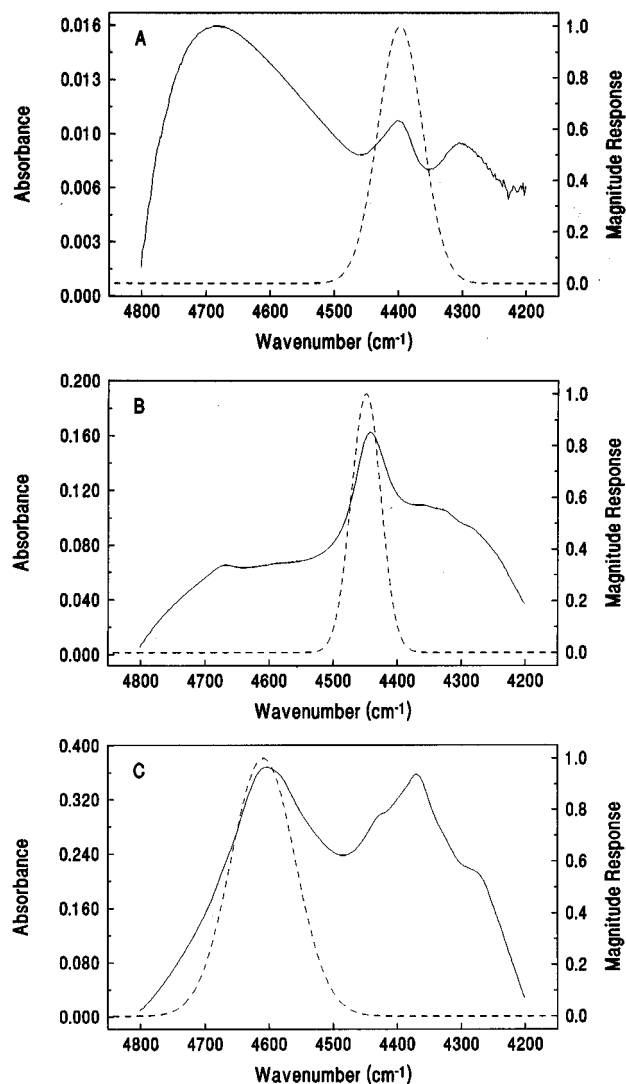


Figure 1. Absorbance spectra of (A) 50 mM glucose, (B) 35 g/L triacetin, and (C) 190 g/L BSA (solid lines). The dashed line superimposed on each spectrum is the frequency response function of the optimized digital filter that was applied to the interferogram data.

RESULTS AND DISCUSSION

Absorption Band Positions. The goal of this study was to collect data with a matrix that matched closely that of blood plasma. Glucose, water, protein, and triglycerides are blood constituents that have significant absorptivities in the near-IR region of the spectrum. BSA and triacetin were used to model the total protein and triglycerides present in blood, respectively. The spectra of these four matrix components show significant overlap in the region of the spectrum that is dominated by the first combination bands (5000–4000 cm⁻¹) of the fundamental vibrational absorption transitions of the dipolar C–H, O–H, and N–H bonds.

The absorbance spectra of pure solutions of 50 mM glucose (A), 35 g/L triacetin (B), and 190 g/L BSA (C) are presented in Figure 1 over the spectral range of 4800–4200 cm⁻¹. These concentrations formed the stock solutions from which the calibration standards were made and are used here to illustrate band locations and widths. Absorption bands of glucose are centered around 4700, 4400, and 4300 cm⁻¹. These are combination bands arising from C–H vibrational modes. Triacetin has a strong C–H combination band around 4446 cm⁻¹, while BSA has C–H and

N–H combination bands centered around 4373 and 4606 cm^{-1} , respectively. There is significant overlap between all of the glucose bands and one or more of the absorption bands of triacetin and BSA.

The above spectral features are located between two dominant absorption bands of water centered at 5200 and 3800 cm^{-1} . A successful determination of glucose in this sample matrix requires that the glucose-dependent information be isolated from the background absorbance due to the interferences. Previous work in our laboratories has focused on the glucose band at 4400 cm^{-1} due to its presence near the minimum in the water absorbance.^{6–7}

Overview of Interferogram-Based Analyses. In our initial work directed to the direct quantitative analysis of interferogram data,¹⁶ an approximate linear relationship between analyte concentration and interferogram intensity was derived from the Beer–Lambert law by use of several assumptions. One of these assumptions was that the frequency content of the interferogram must be reduced to contain only the frequencies associated with a single analyte spectral band. This was accomplished by applying a band-pass digital filter to the interferogram data. A successful univariate calibration procedure was developed with digitally filtered interferogram data of benzene and nitrobenzene of varying concentrations.¹⁶

In more complex samples, the requirement to isolate the frequencies of a single analyte spectral band becomes increasingly difficult due to spectral overlap. This is especially true in the near-IR spectral region because of the occurrence of broad bands. In a second study, multivariate calibration techniques based on partial least-squares (PLS) regression in conjunction with digital filtering were shown to overcome this problem while still allowing a quantitative analysis to be performed with short interferogram segments.¹⁷ In this study, PLS regression was used with filtered near-IR interferogram data to build calibration models for glucose in an aqueous matrix under conditions of varying sample temperatures. A key to the success of this study was the differences in widths of the overlapping spectral bands. A digital filter was used to isolate the frequencies of the 4400- cm^{-1} glucose band, and the underlying tail of the 3800- cm^{-1} water band was suppressed by choosing an interferogram segment displaced from the center burst in which the narrower glucose feature was more prominent than the wider water feature. This procedure takes advantage of a familiar property of Fourier transform pairs: broad features such as the water spectral band at 3800 cm^{-1} have a narrow representation in the interferogram. Thus, interferogram segments displaced from the center burst are automatically enriched in information pertaining to narrower spectral bands.

In the work reported here, the sample matrix employed provides a significant further challenge to this interferogram-based calibration methodology. The presence of the strong BSA and triacetin bands in the region of the glucose bands provides additional interference. Furthermore, the similarity in widths of the glucose, BSA, and triacetin bands dictates that greater overlap of information will occur in the filtered interferograms. To help address this problem of information overlap, a novel calibration methodology based on multiple filters and interferogram segments will be presented.

Implementation of Digital Filtering. A digital filter is defined by its frequency response function or band-pass, a profile of the degree to which each frequency is attenuated by the filter.

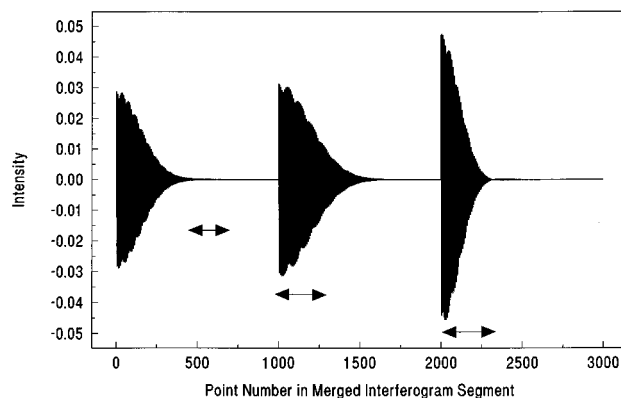


Figure 2. Merged interferogram produced by the application of filters placed on the glucose, BSA, and triacetin bands at 4400, 4606, and 4446 cm^{-1} , respectively. The arrows denote the shorter segments used as independent variables in the PLS calculation.

When the filter is applied to a near-IR interferogram, the frequency response defines which near-IR spectral frequencies will be passed by the filter. The filter can be applied through a Fourier transform procedure (Fourier filtering)^{18–19} or directly to an interferogram segment through a convolution approach (time domain filtering).^{20–21} In the work described here, Fourier filtering was used in the optimization of the methodology on the basis of its ease of application. Time domain filtering was then used to demonstrate that similar results could be obtained with this filtering method.

It is important to contrast this approach with previously reported procedures in which absorbance spectra were Fourier transformed and calibration models were constructed from the resulting Fourier domain data.²² The use of absorbance spectra requires a separate spectral background measurement, while the direct use of interferogram data isolates the analyte information without the use of a background measurement.

Multiple Band-Pass Digital Filters. To implement a calibration strategy based on the use of multiple filters, separate filters of various band-pass positions and widths were centered on either of the following absorption bands: (1) the glucose band at 4400 cm^{-1} alone; (2) the glucose band at 4400 cm^{-1} and the BSA band at 4606 cm^{-1} ; (3) the glucose band at 4400 cm^{-1} and the triacetin band at 4446 cm^{-1} ; and (4) the glucose, triacetin, and BSA bands at 4400, 4446, and 4606 cm^{-1} , respectively. The frequency responses of the optimized glucose, triacetin, and BSA filters are plotted as the dashed lines in Figure 1. Each filter was applied to separate copies of each raw interferogram, producing one, two, or three filtered interferograms, depending on the number of filters used. Segments of these filtered interferograms were concatenated and used together in the PLS analysis. Figure 2 displays an example of the data combined from the use of three filters. Segments of 1000 points from each of three filtered interferograms are plotted. The segments obtained by use of the glucose, BSA, and triacetin filters are plotted as points 1–1000, 1001–2000, and 2001–3000, respectively. As denoted by the arrows in the figure, various shorter segments of each of these 1000-point segments were investigated.

(17) Mattu, M. J.; Small, G. W.; Arnold, M. A. *Appl. Spectrosc.* **1997**, *51*, 1369–1376.

(18) Horlick, G. *Anal. Chem.* **1972**, *44*, 943–947.

(19) Betty, K. R.; Horlick, G. *Appl. Spectrosc.* **1976**, *30*, 23–27.

(20) Parks, T. W.; Burrus, C. S. *Digital Filter Design*; Wiley: New York, 1987.

(21) Small, G. W.; Harms, A. C.; Kroutil, R. T.; Dittilo, J. T.; Loerop, W. R. *Anal. Chem.* **1990**, *62*, 1768–1777.

(22) Heise, H. M.; Marbach, R. *Mikrochim. Acta* **1990**, *3*, 79–85.

(16) Mattu, M. J.; Small, G. W. *Anal. Chem.* **1995**, *67*, 2269–2278.

Data Partitioning. The full data set of 160 samples (480 interferograms) was randomly assigned into a major calibration set and a prediction set. The calibration set comprised 80% (128 samples, 384 interferograms) of the total while the remaining 20% (32 samples, 96 interferograms) formed the prediction set. The calibration set was further partitioned randomly into another minor calibration set and another prediction set which we term the monitoring set. The minor calibration set had 80% (102 samples, 306 interferograms) of the total interferograms in the major calibration set, and the remaining 20% (26 samples, 78 interferograms) were placed in the monitoring set. The composition of the calibration sets was representative of the variation present in the whole data set. The assignment of interferograms to the respective sets was performed on the basis of the individual samples; i.e., the replicate interferograms associated with a given solution were carried together into the calibration, monitoring, or prediction sets.

In a separate study to test the stability of computed calibration models, the major calibration set of 127 samples (381 interferograms) was composed of data collected during the first eight weeks of the data collection, while the prediction set of 33 samples (99 interferograms) comprised data collected during the final week of data collection. The glucose concentrations of the samples in both data sets spanned the variation found in the whole data set. The major calibration set was randomly assigned to a minor calibration set (105 samples, 315 interferograms) and monitoring set (22 samples, 66 interferograms).

During the optimization of experimental variables, the minor calibration set was used to build the calibration models, which were tested by predicting the glucose concentrations associated with the interferograms in the monitoring set. The choice of the best calibration model was based on the model with the lowest standard error of monitoring (SEM). The values of SEM were computed as

$$\text{SEM} = \sqrt{\frac{\sum_{i=1}^{n_m} [c_{m,i} - \hat{c}_{m,i}]^2}{n_m}} \quad (1)$$

where n_m is the number of interferograms in the monitoring set, $c_{m,i}$ is the actual glucose concentration associated with the i th interferogram in the monitoring set and $\hat{c}_{m,i}$ is the corresponding glucose concentration predicted by the model.

The optimal parameters corresponding to the best model were then used to build a single model with the major calibration set. The predictive capability of this model was assessed by use of the prediction set that had been set aside during the optimization process.

Optimization of Experimental Variables. In previous studies, it was established that the best filter position is located centered on the analyte spectral band when simple data with no overlapping spectral bands were analyzed. However, this location may not be optimal if other species with overlapping spectral bands are present in the sample matrix. This is the case with the data set used in this work, and therefore the filter position was a parameter to be optimized for each of the filtering cases. The optimal location in the filtered interferogram segment where analyte-dependent information is located is also governed by the width of the band-pass filter. Typically, a filter wider than the

spectral band of interest has been found to be optimal. Therefore, for each of the filter positions optimized, the widths were also optimized. The starting point and stopping points of the filtered interferogram segments that were used as the input independent variables in the PLS regression calculation formed another set of parameters that were optimized. Finally, the number of PLS factors used to construct the calibration models was optimized.

For each of the filtering strategies to be discussed below, the 306 interferograms in the minor calibration set were filtered, and interferogram segments of different lengths and locations were used in the PLS calculation. The SEM value for the monitoring set was used as the measure of goodness of fit of the calibration model. The smallest SEM value indicated the best fit obtainable.

For the single-filter approach, the initial filter position studied was 4397 cm^{-1} , coinciding with the glucose C–H combination band. For this filter position, the initial full width at half-height (fwhh) values of the Gaussian-shaped filters studied were 72.68, 81.76, and 90.86 cm^{-1} . The choice of this filter as the starting point of the optimization was based on the previous study of glucose in water.¹⁷ The filtered interferogram segments studied were 100, 200, 300, 400, and 500 points, located between interferogram points 300 and 1000, relative to the center burst. The starting and stopping points of this window of interferogram segments were moved across the specified range in steps of 50 points until the entire range was covered. Previous studies have demonstrated that points close to the center burst are dominated by unwanted information and, when included, do not improve the performance of the calibration model. For each set of parameters tested, calibration models based on 8, 10, 12, and 14 PLS factors were evaluated. On the basis of the parameters defining the best performing models, different filter positions and widths, interferogram segments, and calibration model sizes were selected in further pursuit of the globally optimal parameters. Other filter positions studied included 4378, 4387, 4401, 4407, and 4416 cm^{-1} . For each of these positions, different filter widths were investigated. The rest of the study was similar to that described above.

Figure 3A is a plot of SEM vs the starting points of the filtered interferogram segments used for various size calibration models generated with the best filter centered at 4397 cm^{-1} and with a fwhh value of 81.76 cm^{-1} . A filtered interferogram segment of 200 points was sufficient to model the glucose information. The interferogram points noted are relative to the center burst. From this plot, a 12-term model built using filtered interferogram points 551–750 was found to perform best, based on the lowest SEM value. The increase in SEM for models larger than 12 factors is a sign of overmodeling.

The minor calibration set was used to construct these models, and prediction was performed with the monitoring set. Using these optimal parameters, a model was next built with the major calibration set and prediction was performed with the separate prediction set. Table 1 lists the value of R^2 (%) associated with the calibration model, along with the computed standard error of calibration (SEC) and standard error of prediction (SEP). The computation of SEC and SEP employs the calibration and prediction samples, respectively, and is analogous to eq 1.

The two-filter experiments were based on the combination of a glucose filter with either a BSA or triacetin filter. For this two-filter approach, optimization of nine parameters was required: (1) the filter position of the glucose filter, (2) the width of the glucose filter, (3) the filter position of the BSA or triacetin filter, (4) the

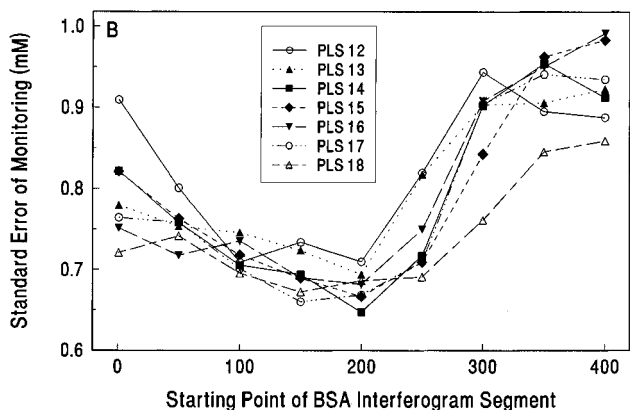
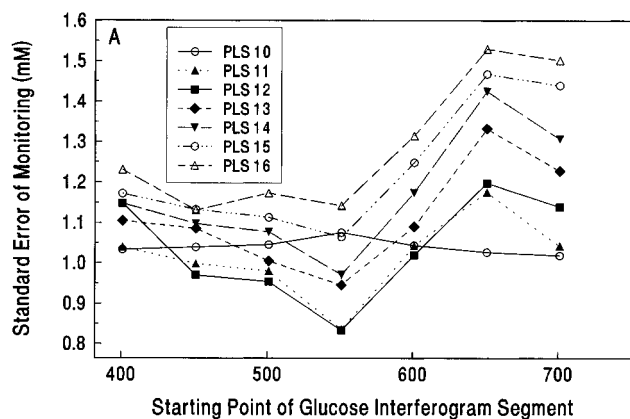


Figure 3. Standard error of monitoring (mM) vs starting points of the filtered interferogram segments for a single glucose filter (A) and joint use of glucose and BSA filters (B). The individual curves correspond to models based on different numbers of PLS factors.

width of the BSA or triacetin filter, (5) the starting and (6) stopping points of the interferogram filtered with the glucose filter, (7) the starting and (8) stopping points of the interferogram filtered with the BSA or triacetin filter, and (9) the size of the calibration models.

A full factorial optimization of these nine parameters was infeasible given the available computing resources. As an alternative, the BSA or triacetin filter position was first optimized with a limited design of three filter positions in conjunction with three filter widths, three glucose filter positions, and three glucose filter widths. A limited study of the interferogram segment and the model size was also performed with this step. Once the BSA or triacetin filter position was selected, a further optimization was performed.

In the second optimization, the previously optimized parameters pertaining to the glucose-filtered interferograms were fixed, and then the parameters were optimized for the BSA-filtered or triacetin-filtered interferograms. Through this approach, the total number of variables to be optimized was reduced to four: (1) the width of the filter centered on the BSA or triacetin band, (2) the starting and (3) stopping points for the BSA or triacetin filtered interferogram segments, and (4) the size of the calibration model.

The BSA filter was positioned at 4609 cm^{-1} on the basis of the first optimization, and the fwhh values were varied between 95.39 and 131.73 cm^{-1} . For each of these filter specifications, calibration models were built using interferogram segments of 100, 200, and 300 points from the BSA-filtered segment of the interferograms, in addition to the 200 points from the glucose-filtered portion of the interferograms. The range of the BSA-filtered interferograms

Table 1. Results from Calibration Models Based on Multiple Fourier Filters

filter type	R^2 (%)	SEC (mM)	SEP (mM)	PLS factors	interferogram segment ^a
1 filter (glucose)	97.72	0.881	0.792	12	551–750
2 filters (glucose, BSA)	98.58	0.698	0.719	14	551–750 (glucose) 201–400 (BSA)
2 filters (glucose, triacetin)	98.17	0.795	0.765	17	551–750 (glucose) 251–450 (triacetin)
3 filters (glucose, BSA, triacetin)	98.85	0.631	0.677	17	551–750 (glucose) 201–400 (BSA) 1–200 (triacetin)
unfiltered	98.17	0.797	1.254	19	101–600

^a Point numbers relative to the interferogram center burst.

investigated was located between points 1 and 550, relative to the center burst. The model sizes studied were 13, 14, 15, 16, and 17 PLS factors.

Figure 3B is a plot of SEM vs the starting point of the BSA-filtered interferogram segments for different sizes of calibration models constructed with the best combination of a glucose filter and a BSA filter. The filter positions and fwhh values, respectively, were 4397 and 81.76 cm^{-1} for the glucose filter, and 4609 and 118.11 cm^{-1} for the BSA filter. Only the BSA-filtered segment starting points are displayed in Figure 3B. The same calibration, monitoring, and prediction sets described previously were used in this study. From this plot, a 14-term model based on glucose-filtered interferogram points 551–750 and BSA-filtered interferograms points 201–400 was selected as optimal. As in Figure 3A, the increase in SEM for models larger than 14 terms is a sign of overmodeling. Table 1 lists the calibration and prediction statistics for this model. Comparison of the one-filter and two-filter results reveals a 9.2% improvement in SEP for the two-filter model.

A similar optimization was performed for a two-filter model based on glucose and triacetin filters. The best experimental parameters were a glucose-band filter at position 4397 cm^{-1} and fwhh 81.76 cm^{-1} , triacetin-band filter at position 4449 cm^{-1} and fwhh 72.71 cm^{-1} , glucose-filtered interferogram points 551–750, triacetin-filtered interferogram points 251–450, and a model consisting of 17 PLS terms. The calibration and prediction statistics for this model are listed in Table 1. A 3.4% improvement in SEP is noted vs the one-filter model. The two-filter model based on glucose and triacetin filters does not perform as well as the glucose–BSA model because the amount of spectral variation induced by the different protein levels is much greater than the corresponding variation produced by the different triacetin concentrations.

The total number of parameters to be optimized increases to 13 when three filters are used. As in the two-filter case, a stepwise optimization approach was implemented. The optimal filter specifications and interferogram segments obtained from the two-filter study involving the glucose and BSA absorption bands were kept constant, and the optimal triacetin filter position of 4449 cm^{-1} was used. The number of variables to be optimized was thus reduced to four. These were the width of the triacetin filter, the starting and stopping points for the triacetin-filtered interferogram segments, and the size of the calibration model. The fwhh values studied for the triacetin filter were 54.51 , 63.61 , 72.71 , and 81.76 cm^{-1} . The lengths of the triacetin-filtered interferogram segments studied were 100, 200, and 300 points, located between points 1

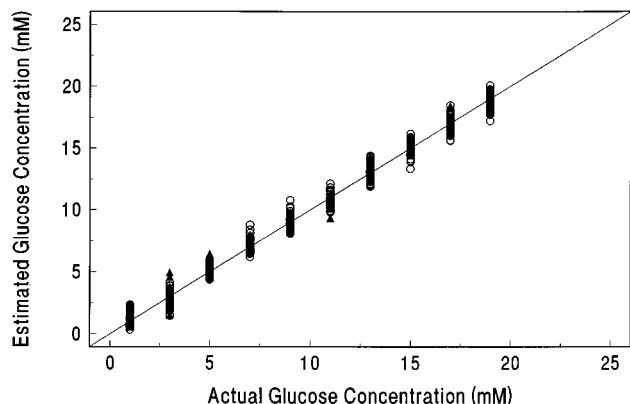


Figure 4. Correlation plot of estimated vs actual glucose concentrations corresponding to the combined glucose, BSA, and triacetin filtering case. Interferograms in the calibration set are denoted by open circles, while interferograms in the prediction set are indicated by closed triangles.

and 600, relative to the center burst. Models based on 13, 14, 15, 16, and 17 PLS factors were evaluated.

In the best three-filter model, the glucose-band filter was centered at 4397 cm^{-1} and had a fwhh value of 81.76 cm^{-1} . The BSA-band filter was positioned at 4609 cm^{-1} and had a fwhh value of 118.11 cm^{-1} , while the triacetin-band filter was centered at 4449 cm^{-1} and had a fwhh value of 54.51 cm^{-1} . The dashed lines in Figure 1 correspond to the frequency responses of these three filters. A total of 600 interferogram points corresponding to glucose (551–750), BSA (201–400), and triacetin (1–200) were used in the 17-term model. The calibration and prediction statistics for this model are listed in Table 1. A 14.5% improvement in SEP is noted vs the model based on one filter.

Figure 4 displays a correlation plot of estimated vs actual glucose concentrations produced by the three-filter model. Calibration samples are indicated by open circles, while prediction samples are denoted by closed triangles. The prediction data fall within the span of variation of the calibration data, and a strong correlation exists between estimated and actual concentrations. Inspections of residual plots revealed no relationships between residuals and BSA/triacetin concentrations. The relatively constant spread in the estimated values at each glucose level indicates that the BSA and triacetin interferences are still influencing the model, however.

Calibration Model Validity. Multivariate calibration models should be able to extract signals of interest from the overlapping information in the interferogram without prior preprocessing. This raises the question of whether the digital filtering step is necessary. To address this question, an optimization study was performed to identify the best segment in the unfiltered interferogram for use in building a multivariate calibration model. The lengths of the unfiltered interferogram segments studied were 100, 200, 300, 400, 500, and 600 points, located between interferogram points 1 and 800, relative to the center burst. The starting and stopping points of each of these unfiltered interferogram segments were moved in steps of 50 points across the range of points 1–800. The calibration models studied included 8–20 PLS factors. The results from this study are tabulated in the last row of Table 1. The best model found was based on interferogram points 101–600 (500 points) and included 19 PLS factors. The SEP (1.25 mM) increased by a factor of ~ 2 compared to the best case (0.677 mM) obtained when the filtering step was used. This

Table 2. Results from Calibration Models Based on Time Domain Filtering

filter type	R^2 (%)	SEC (mM)	SEP (mM)	PLS factors	interferogram segment ^a
1 filter (glucose)	96.12	1.1416	1.135	7	500–999
2 filters (glucose, BSA)	98.66	0.678	0.902	15	500–999 (glucose) 251–450 (BSA)
2 filters (glucose, triacetin)	98.87	0.626	1.027	19	500–999 (glucose) 101–300 (triacetin)
3 filters (glucose, BSA, triacetin)	98.84	0.634	0.915	17	500–999 (glucose) 251–450 (BSA) 301–400 (triacetin)

^a Point numbers relative to the interferogram center burst.

model based on the unfiltered data is also larger by two PLS factors. This result confirms that the filtering step is beneficial in obtaining good calibration models from the interferogram data.

Time Domain Filtering. The work described above used digital filters implemented through the Fourier filtering approach. This method entails Fourier transforming the entire interferogram, multiplying the resulting single-beam spectrum by the filter frequency response function, and inverse Fourier transforming the filtered spectrum back to the interferogram domain. The result is an interferogram containing only the frequencies passed by the filter. The drawback to this method is that the entire interferogram is required for the Fourier transform to estimate the spectrum accurately, thus eliminating the possible instrumental and data collection advantages of using only a small interferogram segment for the entire calculation.

Time domain filters are digital filters designed to be applied directly to segments of time-varying signals such as interferograms. They employ a convolution procedure that approximates the Fourier filtering computation described above. The time domain filtering strategy used in this work was developed in our laboratory specifically for application to interferogram signals.²¹ Using the above optimal filter specifications obtained with Fourier filtering, time domain filters were constructed and applied to the calibration, monitoring, and prediction interferograms used previously. Procedures for optimization of the filtered interferogram segment lengths, locations, and sizes of the calibration models were similar to those described previously. No further optimization of the filter band-pass position and width was performed, however. The results of this study are displayed in Table 2.

The prediction results are somewhat degraded relative to those obtained with Fourier filtering. The principal reason for this degradation is the difficulty in designing a time domain filter that matches the specifications of the Fourier filter exactly. Design of the time domain filter involves the calculation of a finite time domain function that approximates the Fourier transform pair of the frequency response function. In this design process, a tradeoff exists between the width of the filter band-pass and the degree of attenuation of frequencies outside the band-pass. It is anticipated that an independent optimization of the band-pass positions and widths (rather than simply using the Fourier filter specifications) would yield improved results. This optimization was not performed here due to the magnitude of the computation. For the current work, the significant achievement was the demonstration that the analysis can indeed be performed with multiple time domain filters applied directly to short interferogram segments.

Comparison with Spectral Domain Analyses. Comparisons were made between the results obtained with the interferogram-

Table 3. Results from Analysis of Spectral Data

spectral type (point spacing)	R^2 (%)	SEC (mM)	SEP (mM)	PLS factors	spectral range ^a (cm^{-1})
absorbance (2 cm^{-1})	99.17	0.531	0.764	12	4630–4260
absorbance (16 cm^{-1})	98.28	0.764	0.905	9	4630–4285
single-beam (2 cm^{-1})	98.36	0.747	1.111	11	4745–4265
$\log(1/I)$ (2 cm^{-1})	98.96	0.594	0.774	9	4715–4290

^a Segment of spectrum used to construct the calibration model.

based calibration models and the corresponding results from conventional analyses of spectral data. PLS calibration models were optimized for four different types of spectral data: (1) absorbance spectra with 2 cm^{-1} point spacing, (2) absorbance spectra with 16 cm^{-1} point spacing, (3) single-beam spectra with 2 cm^{-1} point spacing, and (4) single-beam spectra at 2 cm^{-1} point spacing that were transformed as $\log(1/I_i)$, where I_i is the single-beam intensity at point i . The absorbance spectra at 16 cm^{-1} point spacing were computed by truncating the 16 384-point interferograms to 2048 points before Fourier processing. Analyses based on single-beam spectra were included since, as in the interferogram-based analysis, no background or reference measurement is required. The log transform was used to help linearize the relationship between single-beam spectral intensity and glucose concentration.

An optimization procedure reported previously was used to obtain calibration models for each of the four spectral data sets.¹² This procedure employs a genetic algorithm²³ to couple digital filtering, PLS regression, and selection of the optimal spectral segment for use in constructing the calibration model. The same calibration/monitoring/prediction sets used with the interferogram data were employed in the analysis of the spectral data. Background spectra of pure phosphate buffer acquired on the same day as the spectra of the glucose samples were used in computing the absorbance spectra.

Table 3 reports calibration and prediction results for each of the four spectral-based analyses. The absorbance spectra at 2 cm^{-1} point spacing and the transformed single-beam spectra produced the best results, roughly equivalent to the best results obtained through the interferogram-based analysis. These two spectral-based models featured slightly lower values of SEC and slightly higher SEP values than the interferogram-based model that employed three filters. The absorbance spectra at reduced resolution and the untransformed single-beam spectra both produced results significantly worse than those obtained in the interferogram-based analyses.

Study of Calibration Model Stability. For the calibration model stability experiment, a full factorial grid-search optimization strategy was utilized. As noted previously, the data set was reorganized for this study. The calibration and monitoring sets included samples whose interferograms were collected over the first eight weeks of the experiment, while the prediction set contained samples whose interferograms were collected during the ninth week. For this study, only the two-filter approach involving the glucose and BSA bands was evaluated. The glucose band filter positions studied were 4387, 4397, and 4401 cm^{-1} . For each of these centering positions, the fwhh values studied were

72.68, 81.76, and 90.86 cm^{-1} . The filtered interferogram segments used were 100, 200, and 300 points, located between points 400 and 900, relative to the center burst. The BSA band filters were centered at 4590, 4600, and 4609 cm^{-1} . For each of these positions, the fwhh values investigated were 104.45, 113.56, and 122.66 cm^{-1} . The filtered interferogram segments used were 100, 200, and 300 points, located between points 1 and 500, relative to the center burst. The starting and stopping points of each of these interferogram segment windows were moved in steps of 50 points, until the entire range was covered for each specific filter specification and model size. The model sizes studied were 13, 14, 15, 16, and 17 PLS terms. The best results from this study were based on a 14-term model, glucose-filtered interferogram points 551–650, and BSA-filtered interferogram points 151–250, to give a total of 200 points. The R^2 , SEC, and SEP values were 98.58%, 0.710 mM, and 0.807 mM, respectively. These results are only slightly worse than the results displayed in Table 1 for the original two-filter glucose–BSA model. This suggests that the interferogram-based modeling scheme has potential for overcoming calibration instability brought about as a result of changes in the instrumental response function over time. This is very encouraging since no separate background measurement is being employed at any step in the data processing.

CONCLUSIONS

This study demonstrated that a direct interferogram analysis for glucose could be implemented despite the presence of strong, overlapping absorption bands of BSA and triacetin. Through the use of multiple digital filters, sufficient spectral selectivity could be incorporated into the interferogram-based analysis to allow the variation of the matrix constituents to be encoded in the calibration model. These results were achieved by direct analysis of short interferogram segments and without the use of a background or reference spectrum for extracting glucose information. This is significant given that, in any implementation of a clinical glucose monitor based on near-IR spectroscopy, the collection of a representative background spectrum (i.e., a measurement of the sample matrix without glucose) will be impractical. Performing the analysis directly using short interferogram segments possesses the additional potential advantage of decreasing the data acquisition and data processing requirements for the spectrometer, resulting in a more rugged, reliable, and inexpensive instrument design.

This study involved the optimization of many parameters associated with the filters and interferogram segments. The optimizations performed in this work were by no means exhaustive, and for a thorough investigation of the optimization, a more formalized optimization technique is needed. Current efforts in our laboratories are devoted to evaluating the use of genetic algorithms for this purpose.

Finally, the relatively large model sizes found to be optimal (e.g., 17 PLS terms for the three-filter case) raise concerns regarding the presence of nonlinearities in the interferogram-based procedure. The PLS algorithm may be utilizing additional model terms in an attempt to model these nonlinearities in a piecewise linear fashion. In previous work, we demonstrated that an approximate linear relationship exists between interferogram intensity and concentration for cases in which the band-pass filter passes information arising from a single absorbing species.¹⁶ This requirement is clearly violated in the present work owing to the presence of the overlapping water, glucose, BSA, and triacetin

(23) Davis, L., Ed. *Handbook of Genetic Algorithms*; Van Nostrand Reinhold: New York, 1991.

spectral bands. The optimal solution to this problem may lie in the use of either nonlinear transformations of the interferogram data or formal nonlinear modeling techniques such as artificial neural networks or nonlinear PLS algorithms. This is also an area of current work in our laboratories.

ACKNOWLEDGMENT

The total costs of this research were supported by the National Institutes of Health under Grant 1-R01-DK45126-01A1. Ndumiso

Cingo and Qing Ding are thanked for assisting in the data collection. The Department of the Army is acknowledged for providing the Silicon Graphics 4D/460 computer system.

Received for review May 30, 1997. Accepted August 27, 1997.[⊗]

AC9705529

[⊗] Abstract published in *Advance ACS Abstracts*, October 1, 1997.