

Scattering and Absorption Effects in the Determination of Glucose in Whole Blood by Near-Infrared Spectroscopy

Airat K. Amerov, Jun Chen,[†] Gary W. Small, and Mark A. Arnold*

Department of Chemistry & Optical Science and Technology Center, University of Iowa, Iowa City, Iowa 52242

Optical properties of whole bovine blood are examined under conditions of different glucose loadings. A strong dependency is established between the scattering properties of the whole blood matrix and the concentration of glucose. This dependency is explained in terms of variations in the refractive index mismatch between the scattering bodies (predominately red blood cells) and the surrounding plasma. Measurements in the presence of a well-known glucose transport inhibitor indicate that variations in refractive index mismatch are related to the penetration of glucose into the red blood cells and demonstrate that increased scattering involves the uptake of glucose by red blood cells. Finally, multivariate calibration models are presented for the measurement of glucose in a whole blood matrix. These models are based on near-infrared spectral data collected from 80 different samples prepared from a single whole blood matrix. Calibration studies are performed over the combination, first-overtone, and short-wavelength spectral regions. The best calibration model is generated from combination region spectra, providing a standard error of prediction (SEP) of less than 1 mM over the concentration range of 3–30 mM. The model based on the first-overtone region is slightly degraded but still provides acceptable performance (SEP = 1.20 mM). The model based on the short-wavelength region is further degraded (SEP = 2.53 mM). To rationalize these results, an analysis of the selectivity of the calibration models is performed by computing the glucose net analyte signal. It is established that the models based on the combination and first-overtone regions are dominated by glucose absorption information, while the model computed from the short-wavelength region is based primarily on scattering information. This result provides evidence that absorption information is needed in order to obtain a glucose calibration model with acceptable performance.

Near-infrared spectroscopy is capable of measuring glucose in complex biological matrixes such as undiluted serum-based cell

growth media and human serum.¹ In these measurements, a selected band of near-infrared light is passed through the sample and glucose information is obtained from a subsequent analysis of the resulting spectrum. Such measurements are reagentless and nondestructive, thereby providing attractive advantages over existing methods for measuring glucose in such samples.

Whole blood is a particularly complex sample, mostly composed of plasma and red blood cells. The ability to measure glucose spectroscopically in whole blood depends greatly on both the scattering and absorption properties of this complex matrix. Light scattering is induced by formal elements of the blood (erythrocytes, leukocytes, etc.) with different refractive indices compared to the surrounding plasma. The exact nature of light scattering depends on many diverse parameters such as size, shape, orientation, refractive index, aggregation properties, and volume percent of the scattering bodies, as well as the wavelength of light.^{2–4} In addition, a glucose dependency has been reported for light-scattering processes involving tissue simulating phantoms,⁵ living dermal tissue^{6,7} and whole blood.^{8,9} This glucose-dependent phenomenon is related to changes in the magnitude of the refractive index mismatch between extracellular fluid and cellular membranes. Furthermore, glucose is reported to alter the aggregation properties of erythrocytes within the blood matrix, thereby affecting scattering properties.^{8,9}

Findings of correlations between glucose concentrations and light-scattering properties in actual or simulated biological matrixes are examples of indirect glucose measurements based on the impact of glucose on a secondary process such as a change in refractive index. Selectivity is a principal concern in all such

- (1) Burns, D. A.; Ciurczak, E. W. *Handbook of Near-Infrared Analysis*, 2nd ed.; Marcel Dekker: New York; 2001; pp 633–647.
- (2) Khalil, O. S. *Clin. Chem.* **2004**, *45*, 165–177.
- (3) Liu, H.; Beauvoit, B.; Kimura, M.; Chance, B. *J. Biomed. Opt.* **1996**, *1*, 200–211.
- (4) Borovoi, A. G.; Naats, E. I.; Oppel, U. G. *J. Biomed. Opt.* **1998**, *3*, 364–372.
- (5) Kohl, M.; Cope, M.; Essenpreis, M.; Bocker, D. *Opt. Lett.* **1994**, *19*, 2170–2172.
- (6) Maier, J. M.; Walker, S. A.; Fantini, S.; Franceschini, M. A.; Gratton, E. *Opt. Lett.* **1994**, *19*, 2062–2064.
- (7) Bruselma, J. T.; Hayard, J. E.; Farrell, T. J.; Patterson, M. S.; Heinemann, L.; Berger, M.; Koschinsky, T.; Sandahl-Christiansen, J.; Orskov, H.; Essenpreis, M.; Schmelzeisen-Redeker, G.; Bocker, D. *Opt. Lett.* **1997**, *2*, 190–192.
- (8) Fine, I.; Fikhte, B.; Shvartaman, L. D. *SPIE Proc.* **2000**, *4162*, 130–139.
- (9) Cohen, O.; Fine, I.; Monashkin, E.; Karasik, A. *Diabetes Technol. Ther.* **2003**, *5*, 11–17.

* To whom correspondence should be addressed. Fax: 319-353-1115. Phone: 319-335-1368. E-mail: mark-arnold@uiowa.edu.

[†] Present address: Sanofi-Aventis Pharmaceuticals, 25 Great Valley Parkway, Malvern, PA 19355.

indirect approaches. While glucose might strongly impact the scattering properties of tissue, such properties are not exclusively modulated by glucose.

The principal objective of the work reported here is to characterize and understand the glucose-dependent transmission properties of whole blood in order to develop better spectroscopic methods for measuring glucose nondestructively and without reagents. Given the concerns noted above regarding the nonspecific nature of light scattering, we have focused this investigation on wavelength regions known to contain glucose-specific absorption features.^{10–12} Specifically, measurements are obtained over the overtone and combination regions of the near-infrared spectrum. The overtone region is 10 000–5400 cm^{-1} and corresponds to the first-overtone and high-order overtones of fundamental stretching vibrations of primarily C–H bonds. The combination region is 5000–4000 cm^{-1} and corresponds to the combination of bending and stretching vibrations associated with C–H, N–H, and O–H bonds. In this work, spectra are collected independently from 9000 to 5400 cm^{-1} and from 5000 to 4000 cm^{-1} . The first region is termed the first-overtone and short-wavelength spectral region and the second is the combination region.

EXPERIMENTAL SECTION

Apparatus and Materials. Near-infrared spectra were obtained by using either a Nicolet Nexus 670 Fourier transform (FT) spectrometer or a Nicolet Magna 560 FT spectrometer (Thermo Nicolet Corp., Madison, WI). The Nexus spectrometer was configured for measurements in the short-wavelength and first-overtone spectral region. This spectrometer was equipped with a 20-W tungsten–halogen lamp, calcium fluoride beam splitter, and cryogenically cooled indium antimonide (InSb) detector. The effective spectral range was 9000–5400 cm^{-1} , where the natural wavelength responsivity of the InSb detector limited measurements at the high wavenumber (9000 cm^{-1}) and absorption by water-limited measurements at the lower wavenumber (5400 cm^{-1}). The Magna spectrometer was used for measurements in the combination region. This spectrometer was equipped with a 37-W tungsten–halogen lamp powered by an Agilent E3633 power supply (Agilent Technology, Inc., Palo Alto, CA), a calcium fluoride beam splitter, and a cryogenically cooled InSb detector. A K-band interference filter (Barr Associates, Westford, MA) was used to isolate the combination spectral region for these measurements.

All spectra were collected as 256 coadded, double-sided interferograms. Fourier processing with triangular apodization, one level of zero-filling, and Mertz phase correction was accomplished with resident software on the Nicolet spectrometers. The resulting overtone single-beam spectra were saved over 9000–5400 cm^{-1} with 4- cm^{-1} point spacing, and the combination single-beam spectra were saved over 5000–4000 cm^{-1} with 4- cm^{-1} point spacing. The nominal spectral resolution was 8 cm^{-1} for both data sets.

Experimental Methods and Procedures. Bovine blood was used throughout as a model for human blood. Bovine blood was

collected from a local abattoir immediately after the animal was sacrificed. Blood was collected in 1-L containers with 1.5 g/L EDTA added as an anticoagulant. The disodium salt of EDTA was used as received from Aldrich Chemical Co. (Milwaukee, WI). The resulting blood was stored on ice or in a refrigerator at 4 °C until needed.

Standard blood solutions used for calibration and prediction purposes were prepared in the following manner. A 50-mL aliquot of blood was warmed for 30 min to room temperature. A small sample of this warmed blood was analyzed for glucose and lactate concentrations with a YSI model 2300 STAT plus (Yellow Springs Instrument Co., Yellow Springs, OH) glucose/lactate analyzer. Base glucose values were in the range of 3–4 mM. A syringe pump (model 44, Harvard Apparatus, Holliston, MA) was used to flow the remaining blood through the sample cell while spectra were collected. The flowing system was required to maintain a consistent suspension of blood cells in the optical path. The sample cell was a demountable liquid transmission cell with a 20-mm-diameter circular aperture (model 118-3, Wilmad Glass, Buena, NJ), sapphire windows (Meller Optics, Providence, RI), and a fixed path length of 0.7 mm for the combination region and 1.0 mm for the overtone region. No attempts were made to control the sample temperature during the data collection. The room temperature remained constant at 22 °C throughout the data acquisition experiments.

After spectra were collected for this unmodified aliquot of blood, a predetermined quantity of dried solid glucose was mixed with the blood to produce another sample. Prior to use, glucose powder (Sigma Chemical Co., St. Louis, MO) was dried at 60 °C for several hours in a Yamato DX 300 oven (Yamato Scientific America, Inc. San Francisco, CA).

Spectra were acquired as this modified blood flowed through the sample cell. This procedure was repeated until a total of five glucose additions were made and the corresponding spectra collected. For each measurement type (overtone and combination), this process was repeated for 16 aliquots of blood, thereby providing spectra from 80 samples (16 × 5). The experimental design was targeted to achieve sampling of the 3–30 mM range of glucose concentration over these 80 samples. Individual concentrations were assigned to the 16 sample blocks in such a way that targeted concentrations could be achieved by successive addition of solid glucose to the base blood. Three replicate spectra were scanned consecutively while the blood was continuously pumped through the sample cell. These replicate spectra represent the same sample from the viewpoint of biochemical content but not necessarily the same structure of red blood cell aggregates passing through the optical path. In this regard, the replicate spectra are more independent than a similar set of spectra collected from a static homogeneous solution. The blood flow rate was maintained at 3 mL/min during the data collection.

The reference glucose concentration for each sample was calculated as the sum of the initial glucose concentration measured with the glucose/lactate analyzer plus the weighed amount of solid glucose diluted to the assumed final volume of 50 mL. To evaluate the computed concentrations, the glucose/lactate analyzer was used periodically to determine the glucose concentration in the final blood sample produced after the five additions of solid glucose. For example, this procedure was applied to 10 of the 16

(10) Arnold, M. A.; Small, G. W. *Anal. Chem.* **1990**, *62*, 1457–1464.

(11) Hazen, K. H.; Arnold, M. A.; Small, G. W. *Appl. Spectrosc.* **1998**, *52*, 1597–1605.

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

sample blocks produced during the acquisition of the combination region spectra. In each case, the measured reference concentration was slightly lower than the computed concentration, indicating the presence of some systematic error. The concentration differences for these 10 cases ranged from 0.22 to 1.39 mM, with a standard error of 0.37 mM. This standard error provides an estimate of the precision of the reference concentrations used in the development of the glucose calibration models.

Calibration models for the near-infrared spectral data were constructed by using the partial least-squares (PLS) algorithm. All calculations were performed in Matlab (Version 7.0, The MathWorks, Inc., Natick, MA). Some procedures used routines from the PLS Toolbox (Version 3.0, Eigenvector Research, Inc., Manson, WA). Each full data set of 80 samples and 240 spectra was randomly split into calibration and prediction data sets. The calibration set comprised 60 samples (12 blocks) while the remaining 20 samples (4 blocks) formed the prediction set. The assignment of blocks to the prediction samples was performed such that two blocks lay outside the time span of the calibration data and the calibration blocks encompassed two prediction blocks. For PLS calibration models, the calibration spectra were further divided by blocks into separate training and monitoring data sets. The monitoring data set consisted of all spectra associated with 3 blocks of 15 selected samples from the calibration data set. The division of the training and monitoring data sets was performed three times to produce three calibration/monitoring subsets. The results from these three rearrangements of the calibration data were pooled and used to find the optimal calibration model.

A grid search was used to establish the optimal spectral range for submission to the PLS algorithm. Similar procedures were used for the combination and first-overtone ranges. For example, in the combination region, contiguous ranges of 1000, 900, 800, 700, 650, 600, 550, 500, 450, 400, 350, 300, 250, 200, 150, and 100 cm^{-1} were moved across the 5000–4000- cm^{-1} range in 25- cm^{-1} steps. For each range, PLS model sizes of 1–15 factors were investigated. This produced a total of 5100 combinations of spectral range and model size. The optimal spectral range and model size was then selected on the basis of the minimum prediction residual error sum of squares (PRESS) computed across the three monitoring sets and calculated as

$$\text{PRESS} = \sum_i (\hat{c}_i - c_i)^2 \quad (1)$$

where c_i and \hat{c}_i are reference and predicted concentrations, respectively, and i denotes the spectrum index. Once obtained, the optimal parameters were applied to the full calibration data set to obtain the final calibration model. Model performance was evaluated by comparing the standard error of calibration (SEC) and standard error of prediction (SEP) calculated as

$$\text{SEC} = \sqrt{\frac{\sum_{i=1}^n (\hat{c}_i - c_i)^2}{n - h - 1}} \quad (2)$$

$$\text{SEP} = \sqrt{\frac{\sum_{i=1}^m (\hat{c}_i - c_i)^2}{m}} \quad (3)$$

where n , m , and h denote number of spectra in the calibration set, number of spectra in the prediction set and number of PLS factors used to build the model.

In studies involving glucose uptake by red blood cells, solutions were prepared by suspending pellets of red blood cells in plasma. Plasma and pellets of red blood cells were generated by centrifuging the whole blood for 15 min at 2600 rpm. The radius from the centrifuge drive shaft to the outermost part of the sample tube when it was in the spinning position was 16 cm, which resulted in an effective g force of 1276. Centrifugations were performed with a refrigerated Marathon 21K/R centrifuge (Hermle Labor-technik GmbH, Wehingen, Germany), and samples were maintained at 4 °C during the centrifugation process.

RESULTS AND DISCUSSION

Basic Response to Glucose. Figure 1 presents a series of single-beam spectra collected with different concentrations of glucose in a sample of whole blood. Step increases in the concentration of glucose result in lower transmission of light through the blood. This effect was observed for the combination (Figure 1A), as well as the first-overtone and short-wavelength (Figure 1B) spectral regions. The spectra in Figure 1B are divided into two parts by the water absorption band in the vicinity of 6900 cm^{-1} . The magnitude of the observed intensity changes is much greater than can be explained by glucose absorptivities over these spectral regions. In addition, the decrease in measured intensity is uniform across wavenumbers, which demonstrates a lack of spectral specificity. Furthermore, a regression analysis of these spectra¹³ indicates that the effective aqueous path length ranges from 1.1 to 1.2 mm compared to the 1.0-mm geometric path length of the sample cells. These facts strongly imply that differences in the scattering properties of the blood, as opposed to absorption properties of glucose, are responsible for the measured decreases in transmission with the step changes in glucose concentration.

The strong dependency between glucose concentration and scattering shown in Figure 1 might involve variations in the refractive index mismatch between the red blood cells and plasma as well as variations in the size and shape of the red blood cells. The observed increase in scattering with higher glucose concentrations would be consistent with either an increase in the refractive index mismatch between the plasma and red blood cells or a reduction in the effective size of the red blood cells (leading to an increased packing density).

If an increase in the refractive index mismatch is responsible for the observed decrease in transmission, then one must consider the effect of glucose on the refractive index of each component (plasma and red blood cells). The refractive index of red blood cells is known to be higher than the surrounding plasma.^{14,15}

(13) Burmeister, J. J.; Chung, H.; Arnold, M. A. *Photochem. Photobiol.* **1998**, *67*, 50–57.

(14) Mazaverica, G.; Freivalds, T.; Jurka, A. J. *Biomed. Opt.* **2002**, *7*, 244–247.

(15) Shvartsman, L. D.; Fine, I. *SPIE Proc.* **2000**, *4162*, 120–129.

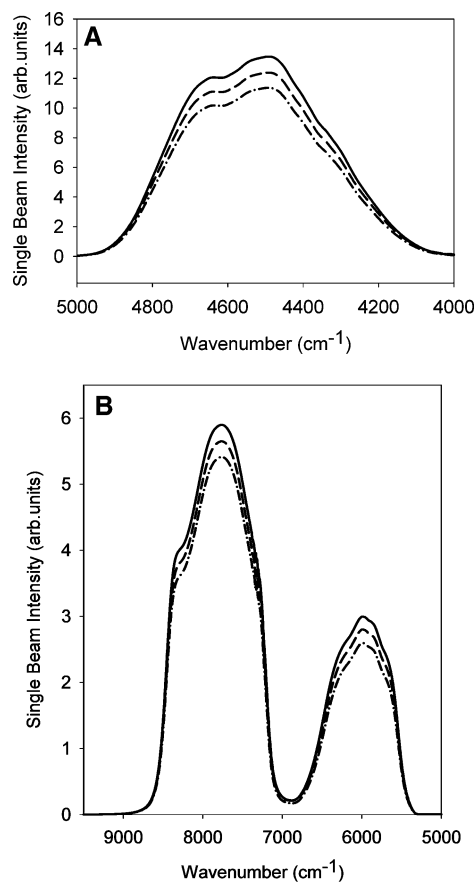


Figure 1. Single-beam near-infrared spectra of whole blood with different concentrations of glucose: 9.6 (solid), 19.2 (dash), and 27.4 mM (dash-dot) over the combination spectral region (A) and the first-overtone and short-wavelength spectral regions (B).

Furthermore, it is also known that glucose increases the refractive index of plasma. The impact of glucose on the refractive index (n) of plasma can be estimated by the following equation, which has been established for aqueous solutions: $n = 1.325 + 2.73 \times 10^{-5} \times [C_g]$, where $[C_g]$ is the millimolar concentration of glucose.^{5,6}

The observed increase in scattering with increasing glucose concentration requires a larger difference in the refractive index mismatch. If the addition of glucose increases the refractive index of the plasma, then the refractive index of the red blood cells or other cellular components of the blood must increase even more. An increase in the refractive index of a cell requires a chemical interaction between glucose and the contents of the cell. For red blood cells, one possibility is the nonenzymatic glycation of the free amino groups of hemoglobin.¹⁸ Hemoglobin glycation involves an initial rapid formation of a Schiff's base followed by a slower Amadori rearrangement. The final product is an amino-linked 1-deoxyfructose derivative of hemoglobin, which should have a higher refractive index in comparison with untreated hemoglobin. Studies to explore other possible interactions between glucose and cells are needed to provide further insight into possible mechanisms for changes in refractive index.

(16) Amerov, A. K.; Chen, J.; Small, G. W.; Arnold, M. A. *SPIE Proc.* **2004**, *5330*, 101–111.

(17) Amerov, A. K.; Arnold, M. A. *SPIE Proc.* **2003**, *4965*, 7–16.

(18) Goldstein, D. E.; Little, R. R.; Wiedmeyer, H. M.; England, J. D.; McKenzie, E. M. *Clin. Chem.* **1986**, *32*, 64–67.

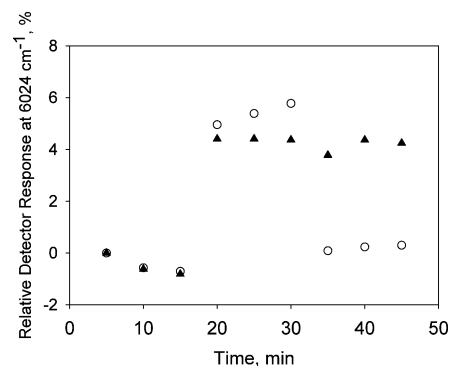


Figure 2. Effect of glucose on transmission of near-infrared light at 6024 cm^{-1} for whole blood in the presence of DMSO alone (open circles) and DMSO with cytochalasin B (triangles).

The involvement of glucose penetration into the red blood cells was examined by monitoring the degree of light scattering in the presence and absence of cytochalasin B, a well-known inhibitor of glucose transport into red blood cells.¹⁹ In this experiment,¹⁶ a stock solution of cytochalasin B was prepared by dissolving 2 mg of cytochalasin B into $250 \mu\text{L}$ of dimethyl sulfoxide (DMSO). While repeatedly collecting first-overtone spectra, $50 \mu\text{L}$ of this stock solution was added to a sample of whole blood. Blank measurements were also performed with only DMSO in order to separate the effects of DMSO and cytochalasin B.

The results of this experiment are presented in Figure 2, where the relative single-beam spectral intensity at 6024 cm^{-1} is plotted as a function of time. As shown in Figure 1B, 6024 cm^{-1} corresponds to the maximum single-beam intensity for blood in the first-overtone range and is thus an effective location for illustrating the effects of scattering. Figure 2 shows results for whole blood with and without cytochalasin B. Initially, no change in the monitored intensity is observed before the addition of either blank DMSO or DMSO with cytochalasin B. In both cases, the detector response increases $\sim 6\%$ upon adding the DMSO or DMSO with cytochalasin B. This is consistent with literature reports that DMSO reduces the refractive index mismatch between red blood cells and the surrounding plasma.²⁰

In the absence of cytochalasin B (blank DMSO), a sharp drop in response is observed after the addition of glucose (10 mM concentration). In contrast, no change in spectral intensity is observed in the presence of cytochalasin B following the same addition of glucose.

Similar results were obtained when this experiment was repeated with a 16% solution of red blood cells as opposed to whole blood. As shown in the Supporting Information (Figure S-1), the relative detector response increased by $\sim 8\%$ upon the addition of DMSO with cytochalasin B and then decreased slightly after 5 mM additions of glucose. In this experiment, a higher total glucose concentration was used and it was found that the detector response was slightly below the initial point at 40 mM glucose. This is in contrast to measured intensities under the same conditions in the absence of cytochalasin B, where the intensity dropped by 7% after each 10 mM glucose loading. The slight decrease in detector response with glucose loadings shown in

(19) Peterson, J. R.; Mitchinson, T. J. *Chem. Biol.* **2002**, *9*, 1275–1285.

(20) Galanzha, E. I.; Tuchin, V. V.; Solovieva, A. V.; Stepanova, T. V.; Luo, Q.; Cheng, H. J. *Phys. D: Appl. Phys.* **2003**, *36*, 1739–1746.

Figure S-1 is consistent with competitive inhibition of glucose transport by cytochalasin B. It is known that additional external glucose reduces cytochalasin B binding to red cells.²¹ Moreover, an osmotic shrinkage mechanism, which was experimentally demonstrated earlier,¹⁷ reduces the diameter of the red blood cells and, as a result, increases the light attenuation of the blood sample even when glucose transport is blocked.

Finally, the time profile for the glucose-dependent scattering of whole blood was measured by following the intensity at 6024 cm^{-1} as a function of time for different glucose concentrations. Results were obtained for 10, 20, and 30 mM glucose, and the time trace is included in the Supporting Information as Figure S-2. In this study, time zero corresponded to the intensity before adding glucose. In all cases, the intensity dropped immediately after adding glucose to the sample. Several minutes were required before the intensity values stabilized somewhat. Starting at ~ 60 min, the intensity dropped slightly over the last 30 min of the experiment.

The results presented in Figures 2, S-1, and S-2 provide strong evidence that the reduction in light intensity upon addition of glucose to blood is a phenomenon of light scattering that is directly tied to the uptake of glucose by the red blood cells. When the time and intensity profiles are considered together (Figures 2 and S-2), the effect of the glucose uptake inhibition on altering the scattering phenomenon is evident.

Glucose Measurements in Blood. A PLS analysis was used to establish calibration models for glucose with three different sets of data. These data sets included the following: (1) single-beam spectra collected over the combination spectral range, (2) single-beam spectra corresponding to the first-overtone range, and (3) single-beam spectra in which only the short-wavelength spectral region was extracted.

Molecular absorption features in the short-wavelength spectral range correspond to high-order overtones and combinations of C–H, N–H, and O–H vibrations. These high-order transitions result in weak, broad, and poorly distinguishable absorption bands that are much weaker in comparison to those found in the first-overtone and combination spectral regions. Despite low absorptivities, measured responses to glucose over the short-wavelength spectral region are larger compared to those in either the first-overtone or combination regions (as illustrated in Figure 1B). These large spectral changes correspond to variations in the scattering properties of the samples as opposed to molecular absorptions. For this reason, spectra collected over the short-wavelength spectral region are used here to model calibrations based on light-scattering information.

Calibration models were generated individually with spectra in each spectral range. Models were computed with single-beam spectra after computing the negative base-10 logarithm of the intensity ($-\log(I)$). This logarithm function puts the spectra in absorbance units relative to a constant background. Each spectrum was mean centered before submission to the PLS algorithm. The principal findings are summarized below for each spectral range.

Combination Spectral Range. The quality of the combination region spectra can be assessed by an analysis of the root-mean-

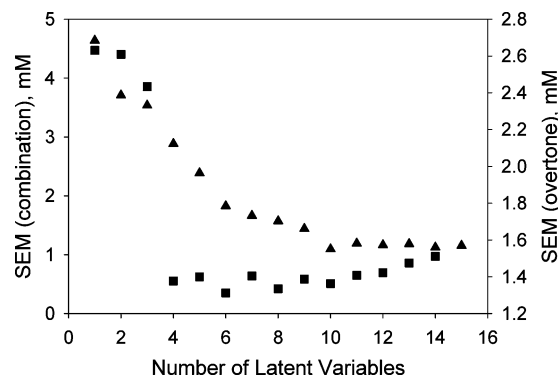


Figure 3. Effect of number of latent variables on SEM for PLS calibration models generated from combination spectra (triangles) and first-overtone spectra (squares).

square (rms) noise of 100% lines. In this analysis, 100% lines were computed for each blood sample by dividing replicate spectra for a given sample by each other. The resulting spectra were converted to absorbance units and fitted to a second-order polynomial. The rms noise was then computed about the polynomial fit over a restricted spectral range. For the spectra used in this study, the average rms noise determined from all samples was $15\ \mu\text{AU}$ over the $4500\text{--}4300\text{-cm}^{-1}$ spectral range.

The best spectral range for the PLS models was $4850\text{--}4200\text{ cm}^{-1}$. Calibration models from spectra restricted to this range gave the lowest pooled SEP for the monitoring sets (termed the standard error of monitoring, SEM). This spectral range incorporates known glucose absorption features centered at 4710, 4400, and 4300 cm^{-1} .¹⁰ All subsequent calibration models used this optimized spectral range.

The optimal number of latent variables was determined by comparing SEM values for models prepared with 1–15 latent variables or factors. The ideal number of factors corresponded to the fewest number that produced an SEM not statistically different from the minimum SEM. A plot of SEM as a function of the number of latent variables is presented in Figure 3. The minimum SEM is observed with 10 factors. An *F*-test was used to compare SEM values obtained with fewer factors to determine whether these smaller models produced significantly different prediction errors. The SEM value obtained with 10 latent variables is significantly lower than that with nine factors at the 95% confidence level. For this reason, 10 latent variables were taken as optimal for this data set.

After the spectral range and number of latent variables were optimized, the calibration and monitoring data sets were recombined into a single calibration data set that was used to build the final PLS calibration model. This model was used to predict the concentrations of glucose within the prediction data set in order to judge model performance.

The concentration correlation plot provided in Figure 4 corresponds to the final PLS calibration model for glucose measurements in whole blood from combination region spectra. Both the calibration and prediction points are plotted and are observed to follow the unity line. The values of the coefficient of determination (R^2) describing the correlation plot for the calibration and prediction data were 0.988 and 0.960, respectively. The SEC and SEP values associated with the final model were 0.56 and 0.96 mM, respectively.

(21) Cloherty, E. K.; Heard, K. S.; Carruthers, A. *Biochem.* **1996**, *35*, 10411–1–421.

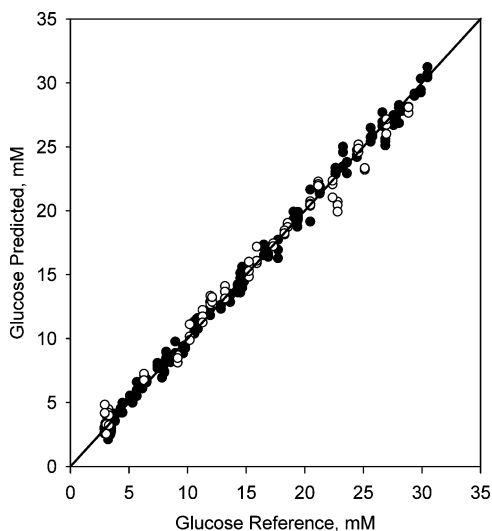


Figure 4. Concentration correlation plot for the PLS calibration model based on combination region spectra, showing calibration (closed circles) and prediction (open circles) data points.

First-Overtone and Short-Wavelength Spectral Range. The same methods described above were used for the analysis of the first-overtone and short-wavelength spectral data. Optimization of the spectral range started with spectra collected over the full range from 9000 to 5400 cm^{-1} . This full range encompasses the entire first-overtone range and the available short-wavelength spectral region. The best spectral range found through the grid search was 6450–5700 cm^{-1} , which corresponds to a selected region within the first-overtone spectrum. This spectral region corresponds to several well-known glucose absorption features centered at 6200, 5920, and 5775 cm^{-1} .¹¹ The average rms noise for this optimal spectral range was 46 μAU , which is higher than the noise level for the combination spectral region. Higher rms noise levels for the first-overtone data are consistent with a lower intensity source in the Nexus spectrometer used to acquire these data and the larger geometrical path length of the cell used to compensate for the lower glucose absorptivities found in the first-overtone range.

The optimum number of latent variables was determined by the same procedure noted above. As shown in Figure 3, a large decrease in SEM is noted between three and four factors and a minimum SEM occurs at six factors. An *F*-test analysis revealed that, at the 95% confidence level, the four-factor model produced a value of SEM that was not statistically distinguishable from the minimum at six factors. Hence, four latent variables were taken as optimal for this data set.

The concentration correlation plot for the resulting optimized first-overtone glucose calibration model was similar in character to that shown in Figure 4 for the combination region but with slightly more scatter. This plot is included in the Supporting Information as Figure S-3. The SEC and SEP values obtained with the final model were 1.14 and 1.20 mM, respectively. No systematic prediction bias was evident from the distribution of points about the unity line. The values of R^2 describing the correlation plot for the calibration and prediction data were 0.984 and 0.980, respectively.

A noteworthy point is the exclusion of short-wavelength spectral information within the optimized spectral range. The

short-wavelength spectral range is dominated by scattering and is essentially void of glucose-specific absorption information. Although the scattering properties of whole blood depend on glucose concentration (see Figure 1B), the PLS algorithm excludes this information in preference to the first-overtone wavelengths. This finding suggests that scattering alone is inferior to absorption information for building multivariate glucose calibration models. This finding is consistent with the need to include glucose-specific information into the calibration model and the limited utility of using nonspecific scattering information for such purposes.

For comparison, a PLS calibration model was also developed with spectral information restricted to the short-wavelength spectral region. Specifically, the spectral range used was 8500–7300 cm^{-1} , where the absolute variation of the single-beam intensity was the greatest, thereby signifying the largest signal arising from the scattering phenomena discussed previously. Four latent variables were used to be consistent with the model generated from first-overtone spectra (three factors were found optimal for this range according to the factor optimization strategy described above). The resulting concentration correlation plot is included in the Supporting Information as Figure S-4. Much greater scatter was observed in this plot compared to either of the corresponding plots for the combination (Figure 4) or first-overtone regions (Figure S-3). The values of SEC (1.83 mM) and SEP (2.53 mM) for this model similarly reflect its poorer overall performance.

Selectivity of Spectral Measurements. The PLS algorithm is a statistical method that maximizes the overlap between variances in the spectral and concentration data matrices. A more detailed analysis of the spectral data is justified in order to assess the chemical basis of model selectivity. As stated previously, scattering is not spectrally selective for glucose, which limits the generation of useful calibration models. Models based on intrinsic glucose-specific information should be more robust, and therefore, such models are preferred if reliable glucose-specific information can be identified.

A net analyte signal (NAS) approach^{22,23} was used to investigate the glucose-specific content within our blood spectra. This approach was used previously in our laboratory to identify glucose spectral features in *in vivo* tissue spectra.²⁴ To determine non-glucose variations in the blood spectra, principal component analysis (PCA) was applied to the 16 blood samples (48 spectra) that contained only endogenous glucose. This calculation assumes the absence of significant glucose concentration variations during the time of the experiment. This data set incorporated 48 log-transformed single-beam spectra, which were not included in the PLS analysis, and corresponds to the initial blood aliquots before any glucose additions were made. Principal components of this data set describe systematic variations of the optical properties of blood passing through the sample cell during the experiment, variations caused by uncontrollable changes in the position of the sample cell, effects of solution temperature variations, and instrumental drift. Spectra of these blood samples with endogenous glucose can be considered as baseline blood spectra. A

(22) Lorber, A. *Anal. Chem.* **1986**, *58*, 1167–1172.

(23) Lorber, A.; Faber, K.; Kowalskii, B. R. *Anal. Chem.* **1997**, *69*, 1620–1626.

(24) Olesberg, J. T.; Liu, L.; Van Zee, V.; Arnold, M. A. *SPIE Proc.* **2004**, *5325*, 11–19.

set of significant principal components that describes the systematic non-noise variations in these spectra can be identified by plotting the residual rms absorbance of factors calculated over the spectral region of interest as a function of the number of these factors:

$$\text{residual rms} = \sum_j \sqrt{\sum_i A_{\perp}^{ij}} \quad (4)$$

$$\mathbf{A}_{\perp} = (\mathbf{I} - \mathbf{U}\mathbf{U}^+) \mathbf{A} \quad (5)$$

where \mathbf{A}_{\perp} (residual absorbance) denotes the portion of the baseline spectra, \mathbf{A} , that is orthogonal to the subspace of significant principal components \mathbf{U} determined from the baseline spectra, \mathbf{I} is the identity matrix, \mathbf{U}^+ is the Moore–Penrose pseudoinverse of \mathbf{U} , i is the sample index, and j denotes the spectral (wavenumber) index. Example plots of log (residual rms absorbance) versus number of factors obtained by applying this calculation to data from the combination and first-overtone spectral ranges are included in the Supporting Information as Figure S-5. An F -test analysis of these data indicated that seven and five principal components were sufficient to describe the nonrandom spectral variance within the combination and first-overtone baseline blood spectra, respectively.

Unique spectral changes caused by adding glucose to the blood can be determined by analyzing these spectra in the following manner. Spectra collected from samples with high glucose concentration are projected onto the subspace spanned by the principal components computed from the baseline blood spectra. The corresponding residual spectrum is computed according to the following expression:

$$\mathbf{a}_{g\perp} = (\mathbf{I} - \mathbf{U}\mathbf{U}^+) \mathbf{a}_g \quad (6)$$

In eq 6, \mathbf{a}_g is a spectrum corresponding to the baseline blood plus added glucose, $\mathbf{a}_{g\perp}$ is the portion of \mathbf{a}_g that is orthogonal to the subspace spanned by the baseline principal components, and \mathbf{I} , \mathbf{U} , and \mathbf{U}^+ are the same as previously defined in eq 5.

The NAS for glucose can be obtained in the same manner by establishing the residual spectrum for a standard absorptivity spectrum for glucose relative to the baseline variance, as defined by Lorber.^{22,23} The glucose absorptivity spectra used in this analysis have been reported previously for both the first-overtone and combination spectral regions.²⁵

Figure 5 provides a set of spectra for comparison. This figure superimposes (1) a representative residual spectrum computed from blood samples with high glucose concentrations, (2) the NAS for glucose relative to the blood matrix, and (3) the glucose absorptivity spectrum for glucose. Panels A and B in Figure 5 present these spectra for the combination and first-overtone spectral regions, respectively. For both spectral regions, shapes are similar for the blood residual spectra and the corresponding NAS for glucose relative to the blood matrix. This similarity is important because it indicates that glucose-specific information is available within these blood spectra. Glucose features are more

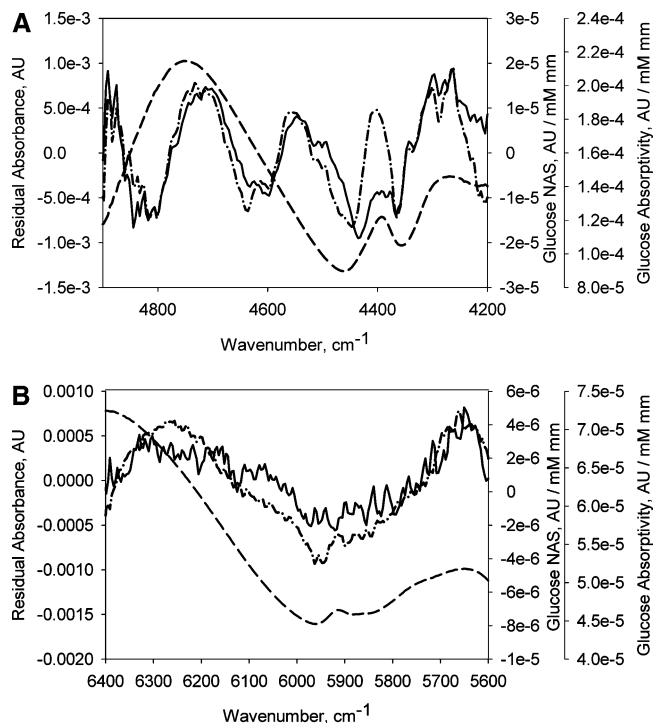


Figure 5. Blood residual spectrum for a 30 mM glucose blood sample (solid), NAS for glucose relative to blood (dash–dot), and glucose absorptivity spectrum (dash) for measurements in the combination (A) and overtone (B) spectral regions.

pronounced in the combination spectral range compared to the first-overtone region. All three carbohydrate absorption bands centered at 4710, 4400, and 4300 cm^{-1} are represented in both the blood residual spectrum and the NAS spectrum. By contrast, analysis of the first-overtone spectral range fails to reveal obvious glucose-specific information. Indeed, the small glucose peak centered at 5920 cm^{-1} is not evident in either the blood residual spectrum or the NAS spectrum. The more intense scattering in the first-overtone region and the presence of three times higher rms noise relative to the combination region can mask glucose absorption features in the blood spectra and thereby degrade performance.

Further investigation of the selectivity of the spectral measurements was performed by comparing the NAS for glucose with the regression vector associated with the PLS calibration model. In the ideal case when only glucose-specific variations are used to build the PLS calibration model, the resulting regression vector will have the same spectral profile as the NAS. In practice, the PLS algorithm uses all variations in the calibration spectral data, whether or not this information derives from the spectral response of the glucose molecule. Any data artifacts that happen to correlate with glucose concentration can thus be incorporated into the calibration model. A direct comparison between the NAS in Figure 5 and the PLS regression vector is complicated by the differences in procedure used to obtain these two spectral profiles. The NAS calculation requires the estimation of the background blood response through the application of PCA to the baseline blood spectra. The PLS calculation employs the full calibration set of log-transformed single-beam spectra. The PLS regression vector is derived from the full range of glucose loadings and thus encodes a greater range of possible data variance. For example, the PLS

(25) Amerov, A. K.; Chen, J.; Arnold, M. A. *Appl. Spectrosc.* **2004**, *58*, 1195–1204.

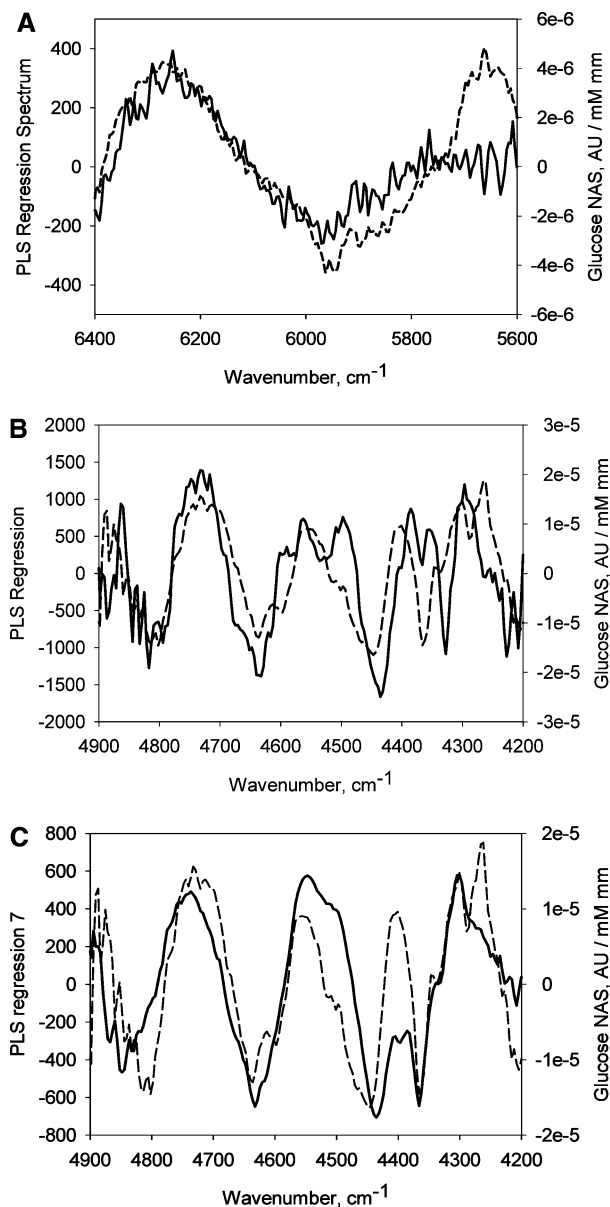


Figure 6. PLS regression vectors (solid) and glucose NAS spectra (dashed) based on 4 PLS factors and 5 PCA factors, respectively, with overtone spectra (A), 10 PLS factors and 7 PCA factors, respectively, for combination spectra (B), and 7 PLS factors and 7 PCA factors, respectively, for combination spectra (C).

model will attempt to describe any nonlinearities or interactions present in the full calibration data set. This information would not be taken into account in the NAS calculation. However, a comparison of the NAS and PLS vectors is valid as an initial assessment of the presence of glucose absorption information in the computed calibration model.

The resulting PLS regression vectors are presented in Figure 6 along with the NAS vectors. Despite slight differences, the major features are reproduced in both spectra for the first-overtone (Figure 6A) and combination (Figure 6B) regions. There are additional spectral features in the PLS calibration spectra that are not present in the NAS spectra. Differences are noted in the vicinity of 4500 and 4600 cm^{-1} , and also, a shift is observed of the

4300- cm^{-1} band in the combination range. A change of the spectral shape in the vicinity of 5700 cm^{-1} is noted in the overtone region. One reason for these differences in the combination region is the effect of an additional three factors in the PLS model compared to the seven factors associated with the PCA approach. These last three factors in the PLS model accumulate some additional spectral variation. The PLS regression vector and the NAS spectrum are more similar when both are computed with seven factors. These spectra are presented in Figure 6C. This finding confirms that additional information is incorporated into the optimal PLS model to explain the full range of glucose concentrations. More important, however, is the fact that glucose spectral measurements in whole blood are based mostly on available glucose-specific information.

CONCLUSIONS

Glucose is shown to greatly impact the optical properties of whole blood. The principal impact of glucose is associated with the scattering properties of the whole blood matrix. The scattering of light depends on the concentration of glucose in the sample and is likely caused by changes in the refractive index mismatch between red blood cells and plasma as well as by changes in the dimensions of the red blood cells. Glucose predominantly affects the scattering properties of the blood matrix.

Multivariate calibrations for glucose were generated for three different wavenumber regions (combination, first-overtone, and short-wavelength near-infrared). The best calibration model was obtained from the combination spectra with an SEP of 0.96 mM for a 10-factor model. The first-overtone calibration model provided an SEP of 1.2 mM with only four factors. Inferior calibration models resulted when spectra were restricted to the short-wavelength region. This implies that scattering information is inferior to absorption information for the selective measurement of glucose. Finally, analysis of the PLS regression vector and computation of the NAS for glucose relative to the blood matrix revealed direct evidence of glucose-specific spectral features in the calibration models based on the combination and first-overtone regions.

ACKNOWLEDGMENT

The authors acknowledge Dr. J. T. Olesberg of the Optical Science and Technology Center at the University of Iowa for fruitful discussions and help with the introduction of the net analyte signal analysis approach used in this study. The authors thank D. Havel of Bud's Custom Meats, Inc. for assistance in obtaining the fresh bovine blood used in these experiments. Professor John Wienczek provided access to the refrigerated Marathon 21K/R centrifuge, and his generosity is greatly appreciated.

SUPPORTING INFORMATION AVAILABLE

Figures S-1, S-2, S-3, S-4, and S-5. This material is available free of charge via the Internet at <http://pubs.acs.org>.

Received for review March 9, 2005. Accepted May 9, 2005.

AC0504161