

In Vivo Near-Infrared Spectroscopy of Rat Skin Tissue with Varying Blood Glucose Levels

Jonathon T. Olesberg,* Lingzhi Liu, Valerie Van Zee, and Mark A. Arnold*

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

We have performed in vivo measurements of near-infrared rat skin absorption in the 4000–5000-cm⁻¹ spectral range (2.0–2.5- μ m wavelength) during a glucose clamp experiment in order to identify the presence of glucose-specific spectral information. Spectra were collected during an initial 3-h period where the animal's blood glucose concentration was held at its normal value. The blood glucose level was then increased above 30 mM by venous infusion of glucose and held for 2 h, after which it was allowed to return to normal. Spectra were recorded continuously during the procedure and are analyzed to identify spectral changes associated with changes in glucose concentration. Because the change in absorbance due to an increase in glucose concentration is small compared to changes due to other variations (e.g., the thickness of the skin sample), a simple subtraction of absorbance spectra from the hyperglycemic and euglycemic phases is not instructive. Instead, a set of principal components is established from the euglycemic period where the glucose concentration is constant. We then examine the change in absorbance during the hyperglycemic period that is orthogonal to these principal components. We find that there are significant similarities between these orthogonal variations and the net analyte signal of glucose, which suggests that glucose spectral information is present. The analysis described here provides a procedure by which the analytical significance of a multivariate calibration can be evaluated.

Diabetes is one of the fastest growing diseases around the world. The World Health Organization estimates that 171 million people worldwide currently have diabetes, and this number is projected to increase to more than 366 million people by the year 2030.¹ The costs associated with diabetes—which include premature death, pain, and financial burden—are tremendous. These costs are directly related to the medical complications associated with chronic hyperglycemia. Early detection and tight glycemic control are paramount to controlling the costs of the diabetes epidemic.^{2,3}

The cornerstone to tight glycemic control is frequent glucose monitoring, where blood glucose concentrations are measured to help administer proper levels of insulin and maintain euglycemic conditions. Glucose sensing technology has advanced considerably in recent years, thereby providing excellent tools for home glucose monitoring and establishing opportunities for tight glycemic control. Unfortunately, the cost and pain associated with current glucose test-strip technology generally restrict the number of daily measurements performed by the average person with diabetes. Recent instrumental advances are focused on reducing these technological barriers. Examples of next-generation technologies include implantable glucose biosensors,^{4,5} which are capable of continuous glucose monitoring over multiple days, and the GlucoWatch biographer, which provides a new glucose reading every 20 min.⁶

Noninvasive optical sensing of glucose has been proposed by many research groups for the frequent and painless measurement of glucose in people with diabetes.^{7,8} The concept is to pass a selected band of infrared radiation through a vascular region of the body and extract the glucose concentration from the resulting spectral information. Near-infrared spectroscopy is one promising approach for noninvasive glucose sensing because of the unique near-infrared absorption spectrum of glucose and the significant penetration of near-infrared light into tissue.

To date, all reported attempts to measure glucose noninvasively with near-infrared spectroscopy have involved collecting spectra from human subjects and using multivariate calibration techniques to correlate variations in the spectral information to blood glucose concentrations.^{9–18} Partial least-squares regression (PLS) is the

* To whom correspondence should be addressed. (e-mail) mark-arnold@uiowa.edu, (voice) 319-335-1368, (fax) 319-353-1115; (e-mail) jonathon-olesberg@uiowa.edu, (voice) 319-335-3514, (fax) 319-353-1115.

(1) <http://www.who.int/diabetes>.

(2) Diabetes Control and Complications Trial Research Group. *N. Engl. J. Med.* **1993**, *329*, 977–986.

(3) *Diabetes in America*, 2nd ed.; Publication No. 95-1468, National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases: Washington, DC, 1995.

(4) Feld, B.; Brazg, R.; Schwartz, S.; Weinstein, R. *Diabetes Technol. Ther.* **2003**, *5*, 769–777.

(5) Wilson, G. S. *Nat. Biotechnol.* **1997**, *15*, 322–323.

(6) Tamada, J. A.; Garg, S.; Jovanovic, L.; Pitzer, K. R.; Fermi, S.; Potts, R. O. *JAMA, J. Am. Med. Assoc.* **1999**, *282*, 1839–1844.

(7) Khalil, O. S. *Clin. Chem.* **1999**, *45*, 165–177.

(8) Khalil, O. S. *Diabetes Technol. Ther.* **2004**, *6*, 660–697.

(9) 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*, 1618–1621.

(10) Marbach, R.; Koschinsky, T. H.; Gries, F. A.; Heise, H. M. *Appl. Spectrosc.* **1993**, *7*, 875–881.

(11) Heise, H. M.; Marbach, R.; Koschinsky, T. H.; Gries, F. A. *Artif. Organs* **1994**, *18*, 439–447.

(12) Heise, H. M. *Horm. Metab. Res.* **1996**, *28*, 527–534.

(13) Jagemann, K. U.; Fischbacher, C.; Danzer, K.; Müller, U. A.; Mertes, B. Z. *Phys. Chem.* **1995**, *191*, 179–190.

(14) Fischbacher, C. H.; Jagemann, K. U.; Danzer, K.; Müller, U. A.; Papenkordt, L.; Schüller, J. *Fresenius J. Anal. Chem.* **1997**, *359*, 78–82.

most popular multivariate calibration algorithm for this purpose. PLS relies on a regression step in which spectral variance is statistically correlated with assigned glucose concentrations. For the most part, published accounts fail to convincingly prove that concentration predictions from these multivariate calibration models are based on glucose-specific spectral information, as opposed to spurious correlations within the data set.^{7,8,19} This is a problem because the PLS algorithm is designed to accentuate all signals that even weakly correlate with analyte concentration. The PLS algorithm cannot distinguish or enhance glucose-specific spectral variations relative to all other sources of spectral variation that happen to correlate with concentration. Unfortunately, it is common to generate apparently functional PLS calibration models on the basis of nonanalyte-dependent variance, such as noise and instrument drift.¹⁹ This problem is particularly troublesome when the glucose concentration varies systematically as a function of time.

The primary goal of this work is to identify glucose-specific absorbance features in *in vivo* spectra in order to establish a solid analytical grounding for evaluating glucose calibration models generated with multivariate techniques. We also hope to provide a means by which other researchers can evaluate the analytical significance of their own multivariate calibrations. An auxiliary goal is to gather evidence toward answering the question of the distribution volume of glucose in tissue. At question is whether intracellular glucose concentrations track with extracellular and capillary glucose concentrations, especially during large concentration variations. The answer has significant implications for a transcutaneous spectroscopic measurement where the ratio of intracellular to extracellular volume is 2:1.²⁰

EXPERIMENTAL SECTION

Adult male Sprague–Dawley rats were used as an animal model in this work. Previous studies in our laboratory have demonstrated that the skin on the upper shoulders of these animals has absorption properties similar to that of a skin fold on the back of a human hand. Retired breeder rats weighing more than 400 g were used. The animals were anesthetized using ketamine for the surgical preparation. Anesthesia was maintained during the course of the procedure by administration of chloralose. A catheter was placed in the right femoral vein for the infusion of glucose, saline, and anesthetic during the experiment. The right femoral artery was cannulated to provide access to the arterial blood stream. All procedures were approved by the University of Iowa Animal Care and Use Committee.

Rats were fasted overnight prior to the clamp experiments. After anesthetization, rectal temperature and pulse-oximetry probes were inserted to monitor body temperature, pulse rate, and oxygenation during the cannulation procedure, which was performed on a heated surgical station. After surgery, the animal

was transferred to the spectroscopy station. Animal temperature was maintained at 38.0 °C using a closed-loop temperature controller. Pulse rate and oxygenation were monitored continuously. Supplemental oxygen was provided to the animal at a rate of ~1 L/min.

Tissue spectra were collected using a Nicolet Nexus 670 FT-IR spectrometer through a skin fold on the animal's back. A 50-W tungsten–halogen bulb with an integrated, gold-coated reflector was used as a broadband light source. A custom fiber-optic interface was used to bring light from the interferometer to the animal and to couple the transmitted light to an external detector. A 1-mm-diameter solid-core low-OH silica fiber terminated with a 4-mm-diameter sapphire ball lens was used for light delivery to the skin fold. Transmitted light was collected using a second ball lens into another solid-core fiber. The collected light was then coupled using an aspheric lens pair onto a two-stage thermoelectrically cooled extended-wavelength InGaAs detector with a 2.6- μm cutoff (Judson Technologies). The current output from the detector was converted to a voltage signal using a low-noise, variable-gain transimpedance amplifier (Femto Messtechnik GmbH), the output of which was connected to the external detector port of the spectrometer. Spectra were recorded as 60-s averages (128 coadded interferograms) with 16-cm⁻¹ (8 nm) spectral resolution.

Spectral noise values were determined from back-to-back measurements of a 1-mm water sample. The spectral noise in the 4400–4500-cm⁻¹ range for a 60-s scan was 4 μAU . Noise values for a 1-mm water sample and a 1.8 AU neutral density filter were ~10–15 μAU .

Body hair was removed with an electric razor before positioning the interface across a fold of skin. After the fiber-optic interface was attached to the animal, spectra were collected for ~3 h while the blood glucose levels were held constant. During this time, isotonic saline was infused into the venous line at 2 mL/h, which is the average fluid intake rate of this size rat. After a sufficient set of baseline spectra were collected, the saline infusion was replaced with an infusion of a 50% glucose solution at a rate of 2 mL/h for 2 h. After that, glucose infusion was stopped and saline infusion was resumed. The overall infusion rate was constant over the course of the experiment.

The fiber-optic interface was located across a skin fold on the animal's back near the shoulders. After initial placement, the interface was not moved during the experiment, except for a period between 13:00 and 13:45, where it was repositioned before the collection of each spectrum. The average thickness of the skin fold during the course of the experiment was ~1.5 mm.

Blood samples were collected from the arterial cannula at 5–15-min intervals. Arterial glucose readings were measured using a HemoCue Glucose 201 Analyzer. When blood glucose values exceeded the functional range of the device (24 mM), the blood samples were diluted with saline. Calibrations were performed with a set of diluted and nondiluted blood samples within the functional range of the device to correct for a proportional error due to the reduction of hematocrit in the diluted samples.

RESULTS AND DISCUSSION

Tissue Composition and Aqueous Path Length Estimation. Skin can vary significantly in thickness from animal to animal and for different sampling sites on a single animal. In addition,

(15) Burmeister, J. J.; Arnold, M. A.; Small, G. W. *Diabetes Technol. Ther.* **2000**, *2*, 5–16.

(16) Samann, A.; Fischbacher, C. H.; Jagemann, K. U.; Danzer, K.; Schüller, J.; Papenkorrdt, L.; Müller, U. A. *Exp. Clin. Endocrin. Diabetes* **2000**, *108*, 406–413.

(17) Malin, S. F.; Ruchti, T. L.; Blank, T. B.; Thennadil, S. N.; Monfre, S. L. *Clin. Chem.* **1999**, *45*, 1651–1658.

(18) Maruo, K.; Tsurugi, M.; Tamura, M.; Ozaki, Y. *Appl. Spectrosc.* **2003**, *57*, 1236–1244.

(19) Arnold, M. A.; Burmeister, J. J.; Small, G. W. *Anal. Chem.* **1998**, *70*, 1773–1781.

(20) Albert, S. N. *Blood volume and extracellular fluid volume*, 2nd ed.; Thomas: Springfield, IL, 1971.

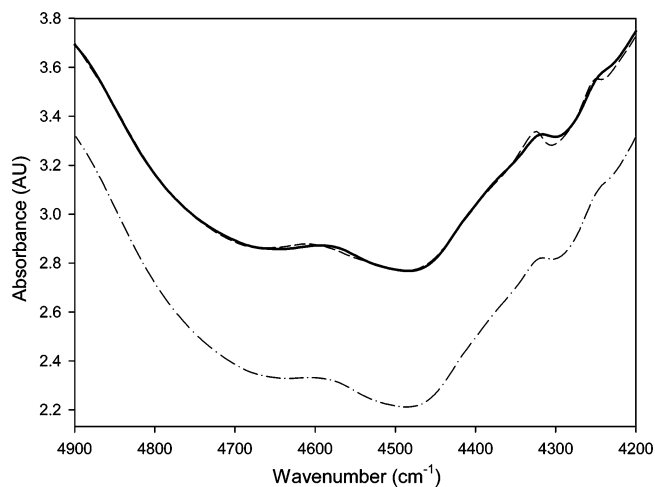


Figure 1. Measured (solid) and regressed rat skin spectra (dash). The purpose of the regression is solely to estimate the effective aqueous path length of the tissue. A transmission spectrum of human skin on the back of the hand is shown for reference (dash-dot). The overall shape of the rat and human spectra are very similar.

the thickness of a skin fold can be modified by varying the pressure applied to the tissue. To account for these variations, we use an estimation of the aqueous path length based on a simple linear regression of the tissue absorbance spectra in terms of the known absorptivities of water, type-I collagen, keratin, and fat. To account for variations in the total skin scattering, a constant spectrum is also included in the regression basis. In addition, to account for variations due to the strong temperature dependence of water, a spectrum obtained from a principal component analysis of water spectra collected at different temperatures is also included. An example *in vivo* absorbance spectrum (solid line) with the regressed fit (dashed line) is shown in Figure 1. The primary structure of the measured spectra is reproduced by the fit, although there is some deviation around the absorption features at 4250 and 4350 cm^{-1} , which are due to fat.

Also shown in Figure 1 is a transmission spectrum of human skin from a site on the back of the hand (dash-dot line). The shapes of the human and rat skin spectra are very similar, which supports the use of rats as an animal model for this study. Regressed skin component thicknesses for human tissue are similar to those from rat tissue, with the exception of the scattering component. The scattering component is typically smaller in human tissue than rat tissue, as indicated by the lower absorbance values across the spectral range in Figure 1.

The regressed skin-component thicknesses are shown as a function of time in Figure 2. The magnitude of each coefficient is steady over the majority of the experiment, except for the period between 13:00 and 13:45, where the magnitudes of all of the coefficients vary significantly. This is the same period of time where the fiber-optic interface was repositioned between scans, as described previously.

The magnitudes of the regressed coefficients are consistent with the expected tissue composition. Over most of the experiment, the sum of the thicknesses of the nonscattering components are ~ 1.5 mm, which matches the approximate physical thickness of the skin fold. The dominant source of optical attenuation is the spectrally featureless contribution of scattering, which accounts for half of the total attenuation. The average aqueous path length

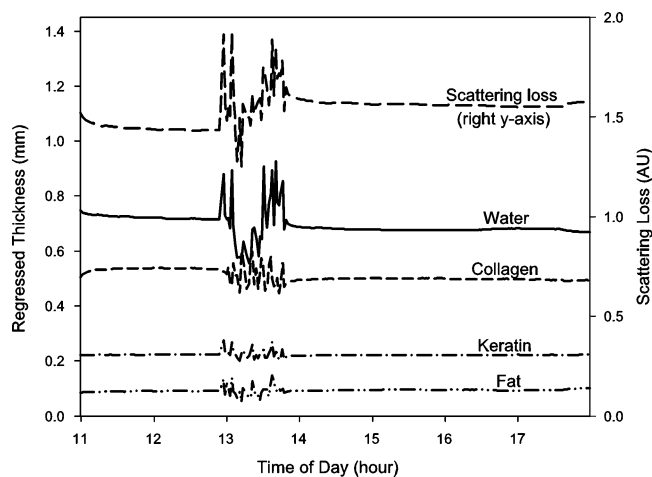


Figure 2. Regressed tissue components as a function of time during the experiment. Thicknesses of water, collagen, keratin, and fat are shown on the left y-axis. Scattering loss is shown on the right axis.

for all of the spectra in this study was 0.70 ± 0.04 mm. Aqueous path lengths can be extracted reliably using this approach because water is the dominant absorptive component in tissue. The aqueous thickness extracted using this procedure will be used as a normalizing factor to account for variations in tissue thickness. The next most significant absorber in skin is protein. Collagen, which is the primary dermal protein, is present in larger quantities than keratin, which is primarily found in the epidermis. The effective thickness of keratin is ~ 200 μm , which is consistent with two passes through a 100 - μm -thick epidermal layer. The lack of significant fat content on a skin fold on the animal's upper shoulders has been verified by postmortem dissections.

Because the water content of tissue even at one site is not constant with time, the regressed aqueous thickness is used as an internal standard. This is accomplished by scaling the concentrations by the regressed aqueous thickness. For example, in the PLS analysis described in the following section, the modeled quantity is the concentration times the aqueous path length rather than just the concentration. The product is proportional to the number of glucose molecules encountered by the light as it propagates through the tissue. After the analysis is complete, the concentration-length product is converted back to a concentration by dividing by the path length.²¹

Partial Least-Squares Analysis. The *in vivo* spectra have been analyzed with PLS using normal precautions against over-modeling. For all chemometric analyses, a fixed spectral range of 4200 – 4900 cm^{-1} was used. A set of 54 randomly selected spectra were removed from the data set and put aside for an independent measure of the standard error of prediction after the calibration is constructed. The remainder of the spectra (306) were used to build the calibration. The number of factors was chosen by performing 50 repetitions of a randomly selected leave-one-third-out cross-validation of the calibration spectra after mean centering.²² The average standard error of cross-validation (SECV) from the 50 repetitions is shown in Figure 3. Nine factors were determined to be optimal on the basis of a standard *F*-test

(21) Chen, J. Spectral processing tools to improve and characterize near-infrared spectroscopic measurements of glucose in complex biological matrices. Thesis, University of Iowa, 2004.

(22) Baumann, K.; von Korff, M.; Albert, H. J. *Chemom.* **2002**, *16*, 351–360.

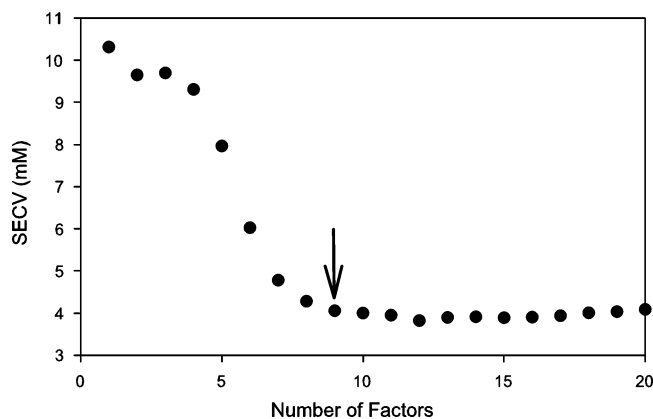


Figure 3. Standard error of cross-validation vs the number of factors used to build the PLS calibration model. A nine-factor model was determined to be optimal using an *F*-test with a 90% confidence interval.

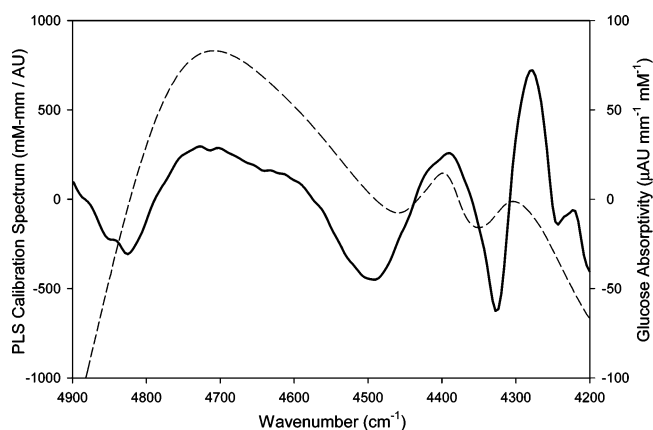


Figure 4. Calibration spectrum generated by the PLS algorithm using nine factors (solid line). Absorptivity spectrum of glucose (dashed line).

statistical analysis with a 90% confidence level, which results in 34 independent spectra per latent variable.

The resulting PLS calibration spectrum is shown in Figure 4 along with the glucose absorptivity spectrum. There is little high-frequency noiselike structure in the PLS calibration spectrum. Such noiselike structure is often an indication of overmodeling. However, there is only vague similarity between the PLS calibration spectrum and the glucose absorptivity spectrum. Both spectra have a broad feature at 4700 cm^{-1} and a pair of peaks at 4300 and 4400 cm^{-1} . It is difficult to say with confidence that the spectral information encoded in the PLS calibration spectrum is keyed to glucose absorbance features or the combination of some other variables.

The result of applying the PLS calibration model to the *in vivo* spectra are shown in Figure 5. Modeled concentrations for the spectra used to build the calibration are shown as open circles while those for the independent prediction set are shown as solid circles. The solid line shows the glucose concentrations of blood samples taken from the arterial line.

Overall, the correlation between the modeled concentrations and those determined from the arterial blood samples shows good agreement. The main exception to this occurs in the period between 13:00 and 13:45, where the tissue interface was repositioned between samples. The added spectral variation introduced by this motion causes difficulty for the PLS model.

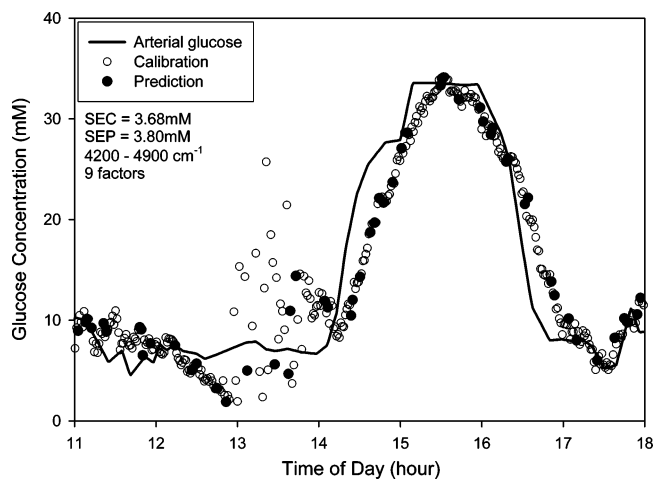


Figure 5. Predicted glucose concentrations using PLS. The open circles represent predictions for spectra from the calibration set. The solid circles are for the independent prediction spectra, which were not used in the model-building process.

In the period between 14:00 and 17:00, where the glucose concentrations are changing rapidly, the modeled glucose concentrations appear to lag those of the arterial samples by ~ 15 min. The delay can be understood as the time required for glucose changes in the arterial system to propagate to the intracellular and intercellular spaces.

The standard error of calibration is 3.7 mM, and the standard error of prediction is 3.8 mM. These errors are too large for a clinically useful device. It should be noted, however, that a significant portion of the standard error is caused by the time lag. If the data are reanalyzed with a 15-min shift in the arterial concentrations, the standard errors drop to 2.20 and 1.98 mM, respectively, which are significantly better.

Even though there is a significant correlation between the arterial glucose concentrations and those predicted using PLS, it is not clear that this correlation is due to glucose spectral information rather than secondary effects or chance correlations. It cannot be overemphasized that surprisingly good correlations can be obtained even with systems where no glucose information is present.^{7,8,19} From an analytical perspective, it is essential to know that the calibration is based on glucose spectral features.

PLS Analysis of Synthetic Spectra. One way to identify the significance of features in the PLS calibration spectrum is to perform a calibration using a set of numerically synthesized spectra. With synthesized spectra, parameters such as instrumental drift and scattering can be controlled or eliminated. To provide spectra analogous to the *in vivo* tissue spectra, we construct our synthesized spectra based on the regressions of the tissue spectra in terms of the major tissue components described previously under Tissue Composition and Aqueous Path Length Estimation. That is, each synthesized spectrum is a linear combination of the spectra of glucose, water, collagen, keratin, fat, a constant (flat) spectrum, and a spectrum representing the temperature dependence of water. The amounts of each component used in the synthetic spectra are determined by the regression coefficients from the *in vivo* spectra. The synthesized spectra are modeled after the *in vivo* spectra except that they do not contain spectral information beyond the components included in the regression analysis.

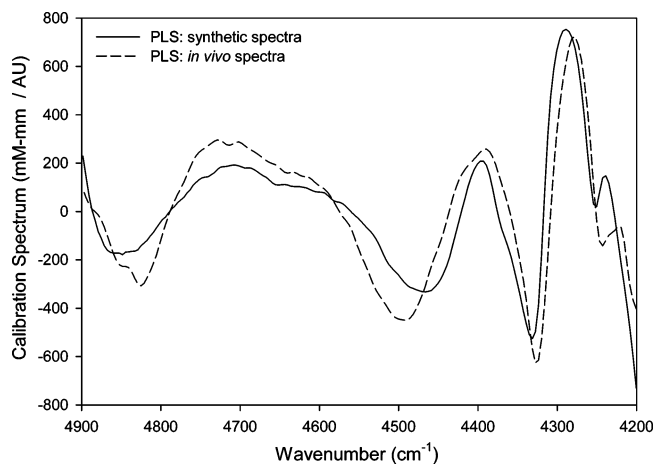


Figure 6. PLS calibration spectra. The solid curve is the calibration spectrum generated using the synthetic spectra. The dashed curve is the PLS calibration spectrum generated from the in vivo spectra. The two are very similar in their dominant features, indicating that the calibration for the in vivo spectra is appropriate for extracting glucose information.

A PLS calibration model was generated in a manner similar to that for the in vivo spectra except that the rank of the model was limited to seven factors in order to not exceed the number of spectral components. As required mathematically, the standard errors are zero since the spectra contain only seven degrees of freedom. The calibration spectrum generated for the synthetic spectra is shown along with the calibration spectrum for the in vivo spectra in Figure 6. There is significant similarity between the two calibration spectra. This finding indicates that the features seen in the calibration spectrum generated for the in vivo spectra are due primarily to the set of spectra included in our regression basis.

There is still the possibility that the calibrations shown in Figure 6 are due to correlations between the concentration of glucose and the quantity of one of the regression basis components. For the case of the synthetic spectra, the correlation coefficients between glucose and all spectral variables are known exactly. The correlations between glucose and all of the other components is small. The greatest correlations are with the quantities of protein and water, which have a correlation coefficient (R^2) of 0.12 and 0.11, respectively. All other components have a correlation less than 0.09.

An additional way to confirm that model performance is not based on correlations is to repeat the PLS analysis on a set of synthetic spectra constructed without the glucose spectral contribution. If the original model is based on glucose spectral information rather than correlations with the other tissue components, the model of the spectra without glucose spectral information should perform poorly compared to the calibration for spectra containing glucose information. This is in fact the case, as is shown in Figure 7, which shows the predicted glucose concentrations as a function of time. The calibration is not able to model the nominal glucose concentrations. The standard errors of calibration and prediction are 7.5 and 8.1 mM, which are much worse than those of the model based on the in vivo spectra. A model with no predictive ability would be expected to have a standard error equal to the standard deviation of the data set, which in this case is 10.5 mM.

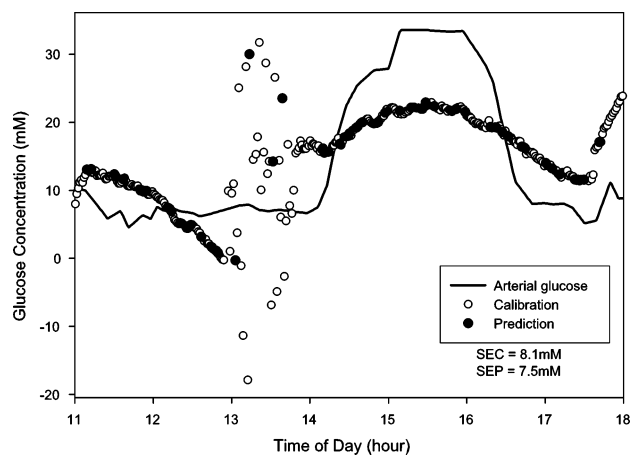


Figure 7. PLS calibration results for synthetic spectra containing no glucose. The calibration is not able to model the nominal glucose concentrations, as is expected. This suggests that the PLS calibration process is not able to build a calibration model for glucose based solely on the correlation of glucose with the primary tissue components.

Spectral Variations due to Elevated Glucose Concentrations

The analysis described in the previous section suggests that the PLS model built from the in vivo spectra is likely based on glucose absorbance information in the presence of water, collagen, keratin, and fat. However, PLS is inherently statistical in nature and relies on a blind correlation analysis between the nominal glucose concentrations and spectral variations. The clearest way to establish the analytical specificity of the measurement would be to directly identify the glucose absorption spectrum in in vivo spectra. In an in vitro experiment, this would be done by subtracting the absorbance of a sample with high glucose concentration from that with a low glucose concentration. The remainder would be the pure-component absorption spectrum of glucose. Unfortunately, this procedure does not work with in vivo spectra because the change in absorbance due to glucose concentration changes (on the order of $14 \mu\text{AU}$ for a 1 mM change in glucose concentration) is much smaller than the natural spectral variability of tissue. In our experience, it is extremely difficult to avoid changes in tissue absorbance due to factors such as protein and water content even when attempting to maintain a constant fiber-optic sampling position, pressure, and sample thickness. Compounding these difficulties is the fact that it takes tens of minutes to induce a significant change in the animal's arterial glucose level, in addition to the time required for glucose transport from the core blood stream into the peripheral skin volume.

Instead of attempting to avoid spectral variations experimentally, it is possible to quantify those variations and look for changes that appear in the high-glucose spectra that are not described by the variations observed in a set of constant-glucose spectra. This is accomplished by collecting a set of baseline spectra while the blood glucose concentration is held constant. During this time, the normal tissue variations are observed and summarized using a set of principal components. This set of principal components describes the normal changes due to animal temperature, skin pressure, protein and fat content, and spectrometer drift. We can then ask if there is categorically different spectral information that arises due to the infusion of glucose. This procedure is a

generalization of the simple subtraction described above. Instead of subtracting a single spectrum, however, a subspace of spectra is used.

An optimal subspace that incorporates the primary systematic variations in a set of spectra can be obtained from a principal component analysis. The number of significant (i.e., non-noise) factors depends on the sample under investigation and the noise characteristics of the spectroscopy system. Experimentally, the spectral variations described by the principal components are due to variations in tissue composition induced by the pressure of the clamp (e.g., water and protein content), variations in animal physiology, and instrumental drift.

Spectra from the baseline period of the experiment are concatenated as columns of a matrix:

$$B = [b_1 \dots b_n] \quad (1)$$

where B is the matrix of baseline spectra b_1 through b_n . The spectra are mean centered using the average baseline spectrum, \bar{b} :

$$B' = B - \bar{b} \quad (2)$$

A singular value decomposition is used to obtain a set of orthogonal spectra that describe the systematic variations in the baseline. The most significant factors are collected into a matrix, U .

After identifying a set of factors that describe the systematic variations in the baseline spectra, additional variations not described by the principal components can be identified in a subsequent spectrum by removing the projection of that spectrum onto the baseline factors from the spectrum. If a matrix A consists of a columnwise concatenation of spectra from the high-glucose portion of the experiment, the portion of A that is orthogonal to the subspace defined by the columns of U can be written as

$$A_{\perp} = (I - UU^+)A \quad (3)$$

where I is the identity matrix and U^+ is the Moore–Penrose pseudoinverse of U . In this study, A_{\perp} is any change in the tissue absorbance observed for high-glucose concentrations that has not already been observed in the baseline spectra.

Figure 8 shows the time profile of arterial glucose concentration in the animal over the course of the experiment. The lower gray regions on the figure indicate the times over which spectra were used to form the baseline (11:00–13:50 and 17:30–18:00). The latter set of spectra are included in the calculation of the baseline factors in order to correct for long-term drift. The baseline set includes the period between 13:00 and 13:45 where the fiberoptic clamp was repositioned on the animal between each scan. This was done in order to include variations due to interface placement and pressure in the baseline spectra in order to reduce the impact of unintentional variations during the high-glucose portion of the experiment.

A principal component analysis was performed on the baseline spectra. It was determined that a set of nine factors described the majority of the non-glucose systematic variations in the spectra.

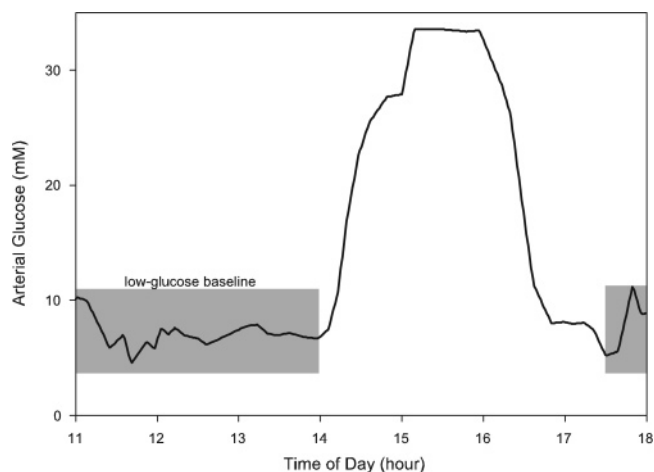


Figure 8. Arterial glucose concentrations (solid line) over the course of the experiment. The gray rectangles indicate the spectra used to calculate the low-glucose baseline.

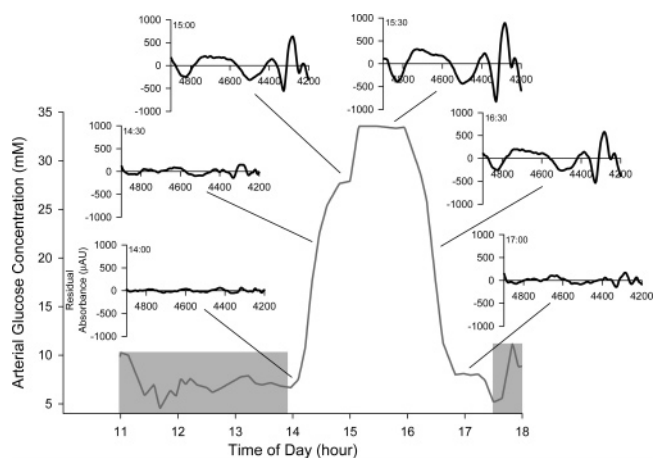


Figure 9. Spectral residuals after the removal of nine baseline factors. The lines indicate the time from which the spectra were taken.

The number of factors was chosen based on a visual inspection of the residuals after removing an increasing number of factors.

These nine factors define a subspace that contains the vast majority of spectral variance observed in the euglycemic portion of the experiment. We can then ask what additional spectral changes exist in the other parts of the experiment. The residual tissue spectra after the removal of the nine factors are shown in Figure 9 at a selection of times during the procedure. Spectra from the period just after 14:00 (but not included in the baseline set) show very little residual absorbance. As the glucose concentration increases, residual structure begins to appear and increases in amplitude, reaching a maximum around 15:30. The residual structure then decreases in amplitude as the glucose concentration returns to normal levels. By 17:00, the residual structure is almost gone.

The average tissue residual from the period around 15:30 is shown as the solid line in Figure 10. The dashed line in Figure 10 is the absorptivity spectrum of glucose. There is some similarity between them, with peaks around 4700, 4400, and 4300 cm^{-1} , but it is not obvious that the spectral residual is due to glucose. That the tissue residuals and the glucose absorptivity spectrum do not match is not surprising, however. What should be compared to the tissue residual spectrum is the absorptivity of glucose after

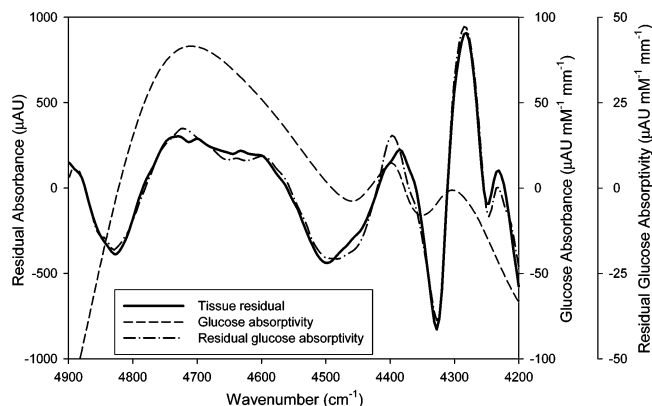


Figure 10. Spectral residuals after the removal of nine baseline factors from spectra collected while the glucose concentration was elevated (solid). Absorptivity spectrum of pure glucose (dash). Glucose absorptivity after removing the nine baseline factors (dash-dot).

the removal of the nine baseline factors:

$$g_{\perp} = (I - UU^+)g \quad (4)$$

where g is the absorptivity spectrum of glucose. This glucose residual spectrum, g_{\perp} is shown as the dash-dot line in Figure 10. A much stronger similarity is seen between the tissue and glucose residual spectra. This similarity indicates that the residual tissue spectra are due to the presence of glucose. In other words, the measured changes in in vivo tissue absorbance have the shape that would be expected based on the addition of glucose into the animal's tissue space.

The residual spectrum after the removal of the baseline factors from the glucose absorbance spectrum is simply the net analyte signal, as defined by Lorber.^{23,24} The net analyte signal is the portion of the analyte absorption that is orthogonal to all other variations in the sample matrix (including absorbance due to other chemical components, effects such as those due to temperature variations and instrumental variations). The similarity of the residuals of the high-glucose spectra to the net analyte signal of glucose is evidence of glucose-specific spectral information in the in vivo spectra.

Net Analyte Signal Calibration. The net analyte signal can be used to directly quantify the analyte concentration. In the case of an ideal multivariate calibration, the calibration spectrum from a method such as PLS should be identical to the net analyte signal, though in practice, they will differ slightly. Both the net analyte signal and the PLS calibration spectrum are estimations of the ideal calibration spectrum. The PLS calibration spectrum is calculated using a statistical regression of the analyte concentrations with the PLS factors, while the net analyte signal is calculated using the analyte absorbance spectrum and a set of baseline factors. In general, the PLS calibration will be more optimal because it is constructed using a regression, but the net analyte signal approach provides a more robust calibration because it is based directly on analyte absorption spectrum.

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

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

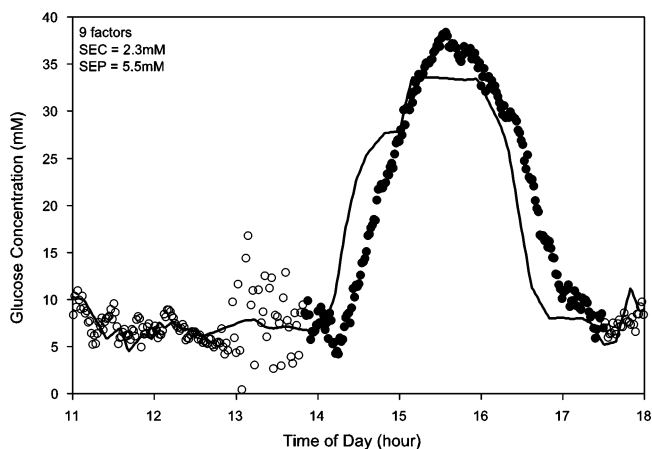


Figure 11. Predicted glucose concentrations using the net analyte signal. The open circles are predictions for the baseline spectra. The solid circles are for the nonbaseline spectra. The solid line is the arterial blood glucose concentration measured with a glucose analyzer.

A properly scaled calibration spectrum can be generated from the net analyte signal:

$$\beta = \frac{g_{\perp}}{|g_{\perp}|^2} \quad (5)$$

where β is the scaled calibration spectrum and g_{\perp} is the net analyte signal. The predicted glucose concentration can then be calculated as

$$c = \frac{\beta \cdot (a - \bar{b})}{l} + \bar{c}_B \quad (6)$$

where a is the absorbance spectrum of the sample being quantified, l is the aqueous path length of the sample, and \bar{c}_B is the average concentration of glucose in the baseline spectra.

The results of this procedure are shown in Figure 11, where the open circles are for the spectra used in the calculation of the baseline factors and the solid circles are predictions for spectra not in the baseline set. The significant scatter in the period between 13:00 and 13:45 is due to added spectral variability caused by moving the fiber-optic interface between scans. The predicted glucose concentrations follow the arterial glucose trend with a 15-min delay. The rms prediction error for the baseline spectra is 2.3 mM, including the time period between 13:00 and 14:45. The prediction error for the nonbaseline region is 5.5 mM. However, a significant portion of the nonbaseline prediction error is due to the time lag between the arterial and optically predicted values. If the predicted values are compared to arterial values from 15 min earlier, the rms error drops to 2.2 mM, which is essentially equivalent to the result for the PLS calibration. The magnitude of the prediction errors is obviously too large for clinical use at this point; however, our emphasis is on obtaining analytical glucose-specific information. We believe that accuracy can be improved as the design of our experiment and equipment improves.

It is important to emphasize that the scaling of the predicted glucose values is derived entirely from the amplitude of the absorptivity spectrum of glucose and not from a regression. In

other words, not only does the shape of the tissue residual match the expected shape for glucose but the amplitude matches as well. No fitting parameters or weighting coefficients are necessary to provide accurate predictions of glucose concentration from noninvasive tissue spectra. Only sample path length information is required in addition to the noninvasive tissue spectrum. The result is an absolute calibration of glucose concentration in the tissue matrix.

Considering the absolute nature of the calibration noted above and the fact that such spectroscopic measurements cannot distinguish between different tissue compartments, the results presented in Figure 11 imply that the glucose concentration is uniform throughout the sampled tissue matrix. To understand this, consider a model of the tissue matrix in terms of the glucose-containing aqueous pools. The simplest model of the aqueous compartments of skin is a combination of intracellular fluid, interstitial fluid, and capillary blood. Approximately 66% of the total aqueous fluid volume of skin is intracellular fluid, 33% is interstitial fluid, and only 1% is capillary blood.²⁰ As noted above, the NAS calibration is sensitive to the total molar concentration of glucose throughout the entire tissue matrix, which is the sum of glucose in these three compartments. The results presented in Figure 11 indicate that the highest measured tissue glucose concentration is approximately the same as the arterial blood glucose level. This finding implies that all three compartments have approximately equal concentrations of glucose and these concentrations are approximately equal to that in the arterial blood. For example, if the concentration of glucose in the blood and interstitial volumes were the same but the concentration in the intracellular fluid was zero, the expected maximum glucose concentration in Figure 11 would be 1/3 that found in the arterial blood.

Equivalent glucose concentration inside and outside cells within the skin matrix is interesting because it suggests that glucose metabolism is slow within these skin cells. Clearly, this finding is preliminary and must be explored further. In addition, this finding could be an artifact of the abnormally high levels of glucose used in this experiment. The impact of the anesthesia must also be determined.

A final step is the direct comparison of the net analyte calibration spectrum to the PLS calibrations using the *in vivo* and synthetic spectra, which are shown in Figure 12. The dominant features of all of these calibration spectra agree in shape and amplitude. This is significant since the three calibrations were generated using significantly different approaches. The PLS calibrations are based on a blind regression of factors in order to establish a correlation with glucose. In contrast, the net analyte signal calibration is based on a set of non-glucose variations and the known absorptivity spectrum of glucose without recourse to a regression step. The convergence of the approaches gives us confidence that all three models are in fact based on glucose-specific spectral information present in the *in vivo* spectra.

CONCLUSIONS

Strategies to explore the chemical basis of selectivity of PLS and other multivariate calibration methods are necessary to advance the development of useful noninvasive sensing technology.²⁵ One interesting approach is to compare the calibration regression vector to analyte spectra²⁶ or to spectral signatures derived from

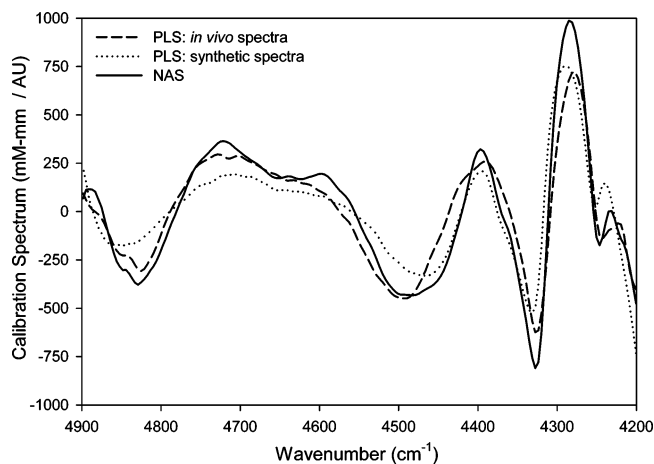


Figure 12. Comparison of the PLS calibrations and the calibration generated using the net analyte signal of glucose.

analyte spectra. Here, we have performed *in vivo* measurements of rat skin absorbance during a glucose clamp experiment in order to identify a spectral signature for glucose. By removing the projection of spectra with high-glucose concentration onto a set of nine baseline factors, we observe significant spectral differences between high- and low-glucose spectra relative to that set of nine factors. By comparing these differences to the net analyte signal of glucose with respect to the same set of factors, we have identified glucose-specific features in these *in vivo* spectra. This structure provides a picture of what a valid PLS calibration spectrum should look like under the conditions for this experiment.

A calibration spectrum was determined by scaling the net analyte signal. This calibration was based solely on the set of baseline spectra with near-constant glucose concentration and the absorbance of glucose. When applied to the measured spectra, the rms error for the nonbaseline spectra was 5.5 mM. If allowance is made for a 15-min delay between the blood and tissue concentrations, the rms error is 2.2 mM. The scaling of this calibration was based entirely on the absorptivity of glucose with no slope correction or regression of the predicted versus known concentrations. The amplitude of the predicted values is consistent with glucose being distributed throughout the entire aqueous volume (intracellular and extracellular fluids).

Calibration spectra determined using PLS are very consistent with the net analyte signal of glucose, which suggests that the PLS model generated for this experiment is based on glucose-specific information and not on secondary effects, such as changes in scattering or chance correlations.

The current experimental procedure was designed to minimize the variance in the *in vivo* spectra. The experiments are very short term in nature (~7 h), involve only a single animal, require a large set of baseline spectra, and involve primarily one tissue site. Ongoing work is addressed at relaxing these constraints in a systematic manner and examining how each factor affects the net analyte signal.

The findings presented here demonstrate a glucose-derived spectral signature in noninvasive *in vivo* spectra. It is important to realize, however, that our findings do not necessarily justify or

(25) Arnold, M. A.; Small, G. W. *Anal. Chem.* **2005**, *77*, 5429–5439.

(26) Enejder, A. M.; Seccina, T. G.; Oh, J.; Hunter, M.; Shih, W.-C.; Sasic, S.; Horowitz, G. L.; Feld, M. S. *J. Biomed. Opt.* **2005**, *10*, 031114-1–031114-9.

validate previously published accounts of noninvasive glucose measurements.⁸ None of these previous experiments were performed under the same conditions of instrumental signal-to-noise ratio, sample thickness, or tissue matrix used in this work. Moreover, most of this previous work was not performed with the same spectral wavelengths. Finally, our results from an animal model under extreme hyperglycemic conditions may not accurately extrapolate to typical conditions for people with diabetes.

ACKNOWLEDGMENT

The authors acknowledge Judy Herlein of the VA Medical Center of Iowa City and Donald Morgan of the University of Iowa

Department of Medicine for their gracious assistance with the design and implementation of the rat glucose clamp procedure. This research was supported by grants from the National Institute of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health (DK-60657 and DK-02925).

Received for review June 11, 2005. Accepted October 28, 2005.

AC051036I