

# Near-Infrared Spectrometric Determination of Hydrogen Ion, Glucose, and Human Serum Albumin in a Simulated Biological Matrix

James K. Drennen, Brian D. Gebhart, Elizabeth G. Kraemer, and Robert A. Lodder University of Kentucky

Near-infrared (near-IR) spectroscopy is used for the determination of pH and glucose and human serum albumin (HSA) concentrations in a simulated biological matrix. Aqueous samples were equilibrated thermally with a heat exchanger prior to analysis in the spectrometer. The liquid samples were analyzed using mathematical techniques in which spectra were correlated with known analyte concentrations. The prediction equations were evaluated using two criteria: the strength of these correlations and the ability of the prediction equations to calculate concentrations of "unknown" solutions. Good correlations between near-IR spectra and analytes were observed at concentrations an order of magnitude below those ordinarily considered amenable to near-IR analysis.

Near-infrared (near-IR) spectroscopy is a method of analysis that uses the signal from a sample at several wavelengths to determine the sample's composition (1). Some advantages of the technique are:

- Analysis times of <1 min are common.
- Either liquid or solid samples can be analyzed; little or no sample preparation is required.
- It is noninvasive and nondestructive.
- It determines a wide variety of sample properties; not only sample component identities and concentrations, but also molecular weight, taste, hardness, and even thermodynamic parameters such as heat of formation (2).

- The equipment is easy to operate and inexpensive.

Simultaneous determinations of the properties described above are made by collecting near-IR signals at a variety of wavelengths during a single sample scan and analyzing the signals using a number of calibration equations. These calibration equations are obtained through a modeling process that employs a training set of samples (analyzed by a chemical procedure) to "teach" a computer to recognize relationships between diminutive spectral features and sample composition. The resulting calibration equations are stored in a computer. A single set of spectra often has several sets of associated reference values; thus, a number of sample properties can be simultaneously determined from the spectra.

The near-IR detection limit for most substances is ~0.1% (3), but in most near-IR applications the analyte is present at a concentration >1%. This article's goal is to determine the feasibility of using near-IR spectroscopy for the determination of pH and glucose and human serum albumin (HSA) concentrations in phosphate-buffered saline (PBS) solutions. A near-IR spectrum was collected for each liquid sample. The resulting spectra were analyzed using computerized regression techniques in which the spectra were correlated with known analyte concentration values to produce prediction equations for each analyte. The dependability of each prediction equation was measured by the strength of the correlation between the near-IR spectra and the known analyte concentrations and by the ability of the prediction equations to calculate the concentrations of "unknown" PBS solutions (cross-validation).

## EXPERIMENTAL

**Samples.** Nineteen different PBS solutions (pH 7.0) with HSA (Sigma Chemical Co., St. Louis, Missouri) concentrations ranging from 10 to 20,000 ppm were prepared and examined spectrometrically. Nineteen different

PBS solutions (pH 7.0) with glucose (Sigma) concentrations from 1.1 mM to 0.44 M were also examined, as were 19 different sample-matrix solutions with pH values from 4.5 to 9.25. Standard laboratory acid and base (HCl and NaOH) were used to adjust the pH values of these variable-pH matrix solutions prior to recording near-IR spectra. A blank sample-matrix solution was also scanned with each group of 19 samples.

**Equipment.** The spectral data were collected using an InfraAlyzer 500 monochromator-based scanning near-IR spectrometer (Bran + Luebbe, Elmsford, New York) with a temperature-regulated (34.5 °C) disposable sample cell. Sample cells and solutions were equilibrated on a 34.5 °C heat exchanger (constructed from a recirculating water bath and a steel reservoir inside a styrofoam-insulated housing) for 1 h before collecting spectra. Each spectral scan covered the 1100–2500-nm wavelength range in 4-nm increments. The data analysis was performed using a MicroVAX II computer (Digital Equipment Corp., Maynard, Massachusetts), a 3090/300E parallel vector supercomputer (IBM, Armonk, New York), and programs written in Speakeasy IV Epsilon (Speakeasy Computing Corp., Chicago, Illinois).

## RESULTS AND DISCUSSION

**Reproducibility of samples in the liquid cell.** A major problem in near-IR analysis of minor components of aqueous solutions is the difficulty in reproducing the appearance of the liquid sample to the instrument. Small shifts in the most intense spectral bands can easily obscure the signals from concentration changes of minor sample components. Temperature changes (often a major source of near-IR spectral variation in pure samples) seem to particularly affect the water peaks in aqueous solutions because water peaks are by far the strongest peaks in the spectrum of an aqueous solution.

Received for review 1/15/90.

© 1990 by American Chemical Society

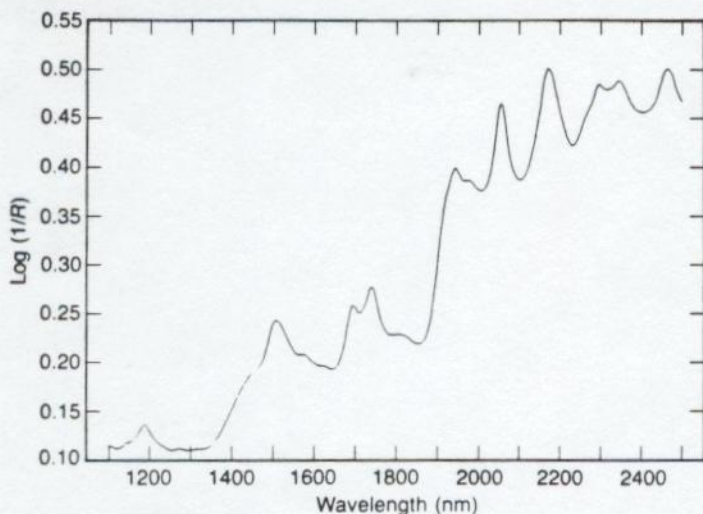
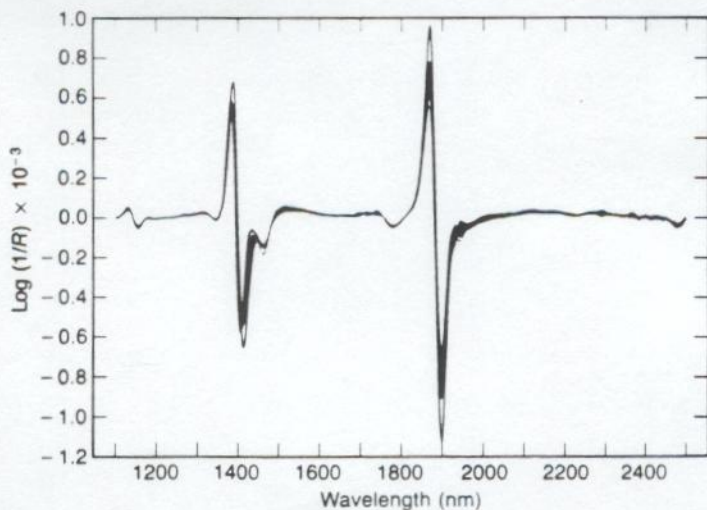


Figure 1. Spectrum of freeze-dried solid HSA.

Table I. Analysis of full spectra (1100–2500 nm) of HSA solutions.

Spectra	Moving average width (nm)	SEE (ppm)	RSD (%)
Log (1/R) spectra	12	1600	8
	20	1100	5
	28	1400	7
Smoothed first derivative	20	2600	13
Smoothed second derivative	20	700	4



The temperature-controlled disposable liquid cell and heat-exchanger system was tested with water samples to determine the instrumental noise level (spectral reproducibility). Twenty water samples were loaded into the cell using a precision pipette, and 20 scans were run of each sample. The variance at each wavelength was calculated, and the mean noise level across the water spectrum was determined to be 63  $\mu$ AU (8  $\mu$ AU referenced to 0 AU). The 70- $\mu$ L liquid cell and heat-exchanger system was stable (maintained this noise level) over a period of  $\sim$ 7 h.

**Analysis of HSA in PBS solution.** Figure 1 shows the spectrum of freeze-dried solid HSA. HSA is not apparent in the solution spectra, which are determined primarily by the water solvent.

The sample solution spectra (composed of HSA, glucose, and test solutions of varying pH) are actually made up of 21 scans of the corresponding sample solution. The 21 spectral scans were obtained without moving the sample between scans. Medians (by wavelength) of three spectra were taken to produce seven median spectra of each sample solution: these seven spectra were then averaged to produce the final sample spectrum that was used in the regression and prediction process. Use of the median spectra removes spikes that appear in the spectra as a result of a high detector-gain setting that causes overflow of the analog-to-digital (A/D) converter (the high gain setting improves overall spectral reproducibility at the expense of an occasional A/D "glitch"). The averaging process was performed simply to increase the signal-to-noise ratio (S/N).

Once the solution spectra were obtained, further data treatments could be employed. Moving-average filters and calculation of first and second derivatives were used as pre-processing treatments for principal-component regression. The best window for moving-average smoothing was found to be 20 nm.

The results in Table I were obtained from calibrations constructed using the full spectra (1100–2500 nm) from HSA in test solutions. Frequently, however, much of the information in a spectrum will be unrelated to the analyte of interest. In these cases, it is generally best to isolate the spectral regions containing analyte information and use only these regions in calibration and prediction. Two bands were used in this way to predict HSA: 1680–1780 nm and 2020–2320 nm. Smoothed (20-nm window) second derivatives (that is, smooth, take first derivative, smooth, take second derivative, then smooth twice) of these regions produced the best linear prediction results for HSA in test solutions: standard error of calibration (SEE) = 700 ppm, standard error of prediction (SEP) = 800 ppm, and relative standard deviation (RSD) = 4% (obtained from cross validation). It is interesting to note that when the smoothed second derivative of a partial spec-

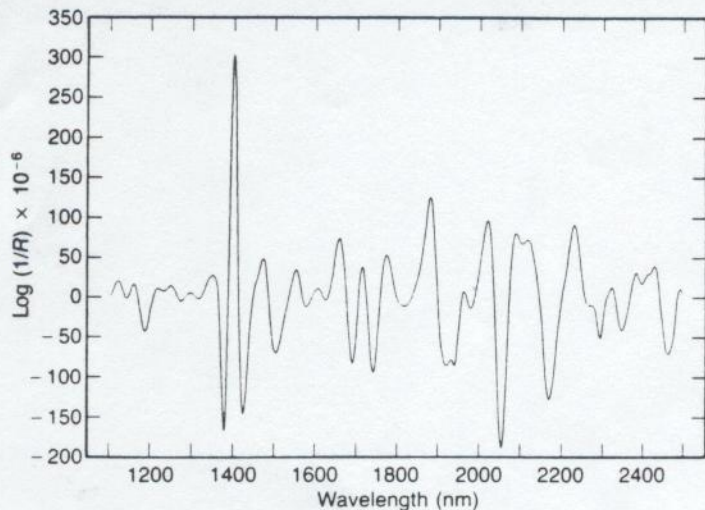


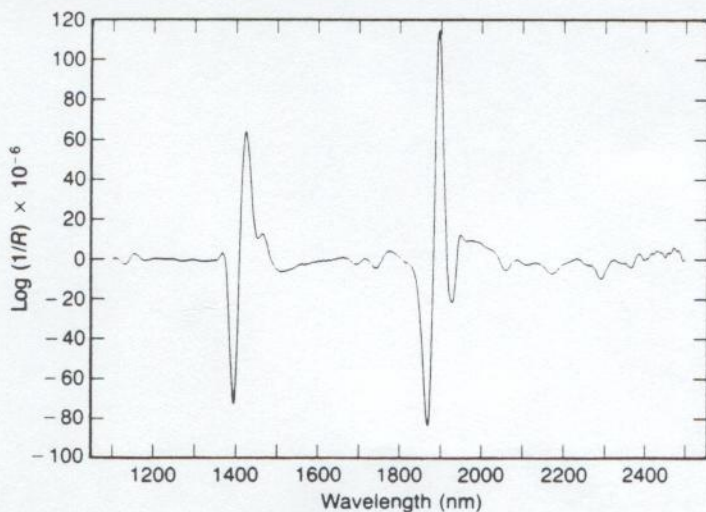
Figure 3. Second-derivative spectrum of pure HSA.

trum is used with principal-axis transformation, the first axis correlates strongly to HSA concentration (along with the fourth and fifth axes). The first axis contains 70% of the total spectral variation while the fourth and fifth principal components account for 3%. In ordinary  $\log(1/R)$  spectra, it is not unusual for the first principal axis to describe only base-line variations and to contain 95% or more of the total spectral variation.

The cross-validation samples superimposed on the calibration line for HSA suggested that a polynomial fit was more appropriate for HSA determinations than a linear one. HSA changes its conformation with increasing concentration. A nonlinear response

was therefore expected. A second-degree polynomial fit to the HSA spectral data gives an SEE of 470 ppm (equivalent to a 2% RSD).

Identification of the spectral regions that are relevant to HSA determination is easier if one compares the second-derivative spectrum of pure HSA with the spectrum of HSA reconstructed from the solution spectra. The second-derivative spectra of HSA in test solution appear in Figure 2. The second-derivative spectrum of pure HSA appears in Figure 3. The spectral reconstruction (through cross correlation) of HSA in test solutions appears in Figure 4. Comparing Figures 3 and 4 shows that HSA signals appear in aqueous solution in the regions examined above (1680–



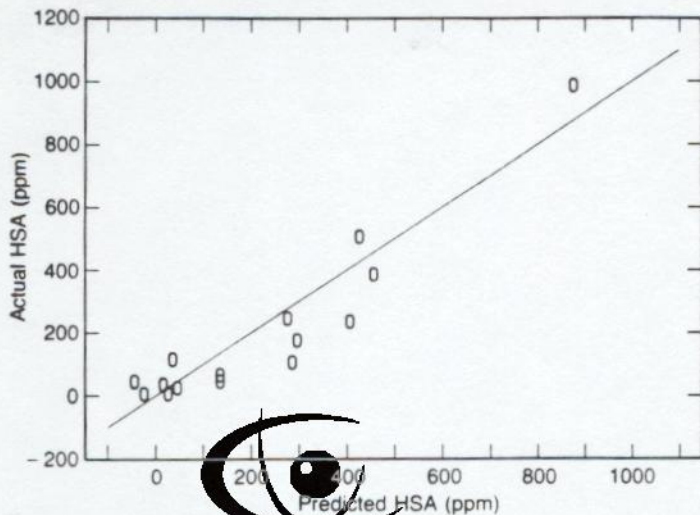


Figure 5. HSA calibration using only the HSA concentrations from 0 to 1000 ppm, calculated from smoothed second-derivative spectra.

ing an approximate detection limit of 300 ppm.

The use of second derivatives as a data pre-processing treatment for principal-component regression seems to compensate somewhat for the use of different disposable sample cells (which often have slightly different path lengths). Following second-derivative treatment, base-line variation is no longer the major source of spectral variation, and the multiplicative effect of path length changes on spectra is largely eliminated. Nevertheless, two of the cells used (no. 4 and no. 6) did show a regular tendency to err in HSA concentration (no. 4 by 400 ppm and no. 6 by 2400 ppm) on the first principal axis calculated from full spectra. This problem was eliminated by not using the first principal axis in the full-spectral calibrations. However, the use of matched cells and a dual-beam background correction should permit the use of the first principal axis in most calibrations (if needed).

**Analysis of glucose in PBS solution.** The analysis of glucose solutions was conducted in much the same manner as the HSA solution analysis. Glucose solutions ranging in concentration from 0 to 0.44 M were analyzed using the smoothed second derivative of their full near-IR spectra (1100–2500 nm) (see Figure 6). Validation samples (samples not used to develop the calibration) are shown superimposed on the calibration line in Figure 6. The SEP calculated by cross vali-

1780 nm and 2020–2320 nm).

An HSA calibration was also constructed using only the HSA concentrations from 0 to 1000 ppm. This calibration appears in Figure 5 and was calculated from smoothed second-

derivative spectra (that is, smooth, smooth, first derivative, smooth, second derivative, smooth, smooth). The parabolic behavior of HSA signals is still detectable in the fit. The SEE for this calibration is 100 ppm, indicat-

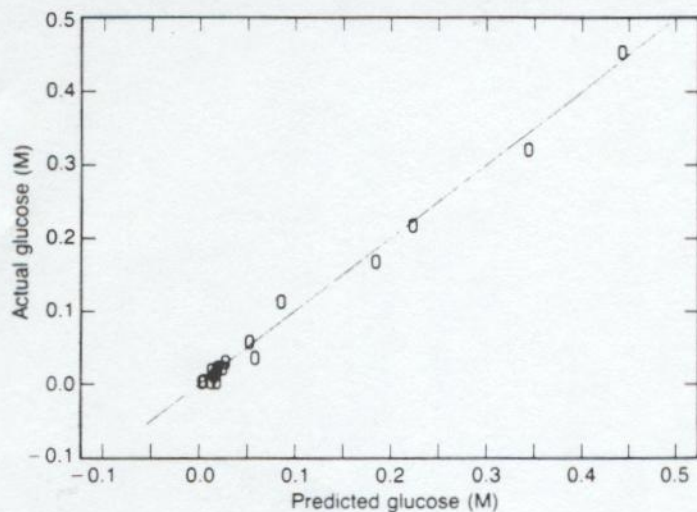


Figure 6. Calibration line of the smoothed second derivative of glucose.

ation is 0.01 M over this range (an RSD of <3%).

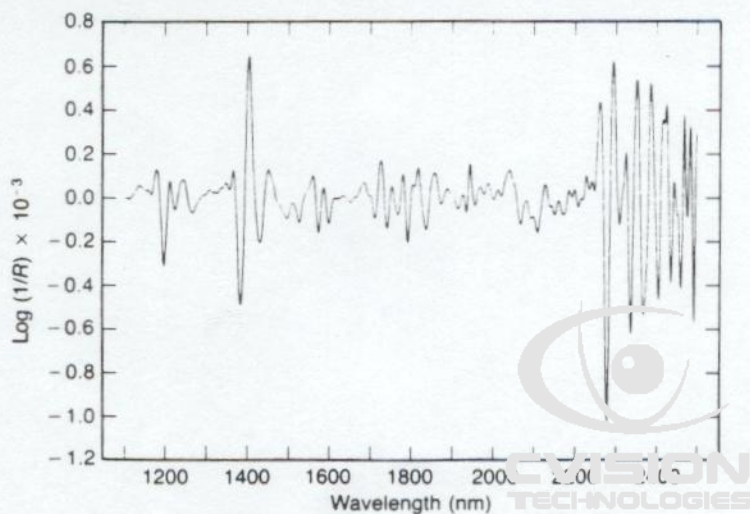
No spectral regions were found to correlate to glucose concentration in both the solution and the solid phase. Figure 7 shows the second-derivative spectrum of solid glucose reagent, while Figure 8 depicts the second-derivative spectrum of glucose, reconstructed by cross correlation from the sample solutions. A comparison of the two figures demonstrates that they share no common peaks, and therefore analysis cannot easily be limited to a small and relatively noise-free region.

A calibration equation was constructed for glucose using the samples with concentrations from 0 to 0.03 M. The SEE for this cali-

bration is 0.0044 M, corresponding to an approximate detection limit of 0.013 M.

**Analysis of pH of PBS solution.** The determination of pH by near-IR spectroscopy was also conducted in a manner similar to that used in the HSA experiment reported above. Smoothed second-derivative spectra were employed in principal-component regression. The pH of the PBS solutions ranged from 4.5 to 9.3, and the results from full-spectral analysis produced a 0.3-pH SEP (from cross validation). The calibration line is shown in Figure 9 with the validation samples superimposed upon it.

Four spectral regions were found to contain the information necessary to extract pH from near-IR spectra of PBS solutions: 1130-



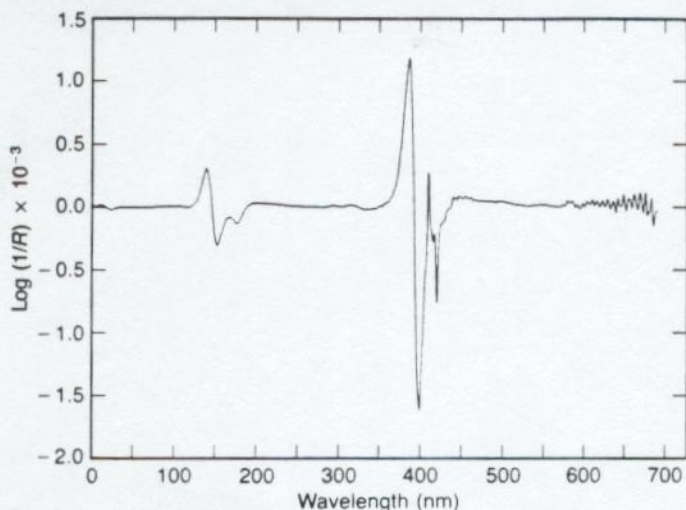


Figure 8. Second-derivative spectrum of glucose, reconstructed by cross correlation from sample solutions.



1220 nm, 1310–1370 nm, 1450–1500 nm, and 1950–1980 nm. The analysis to these regions produced the same SEE as the full spectral analysis.

**Glucose determination using a fiber-optic cell.** Recent work (4) with a fiber-optic probe has led to an improvement in S/N of the glucose determination from 25,000 (using the drawer on solid samples) to 150,000 (using the probe on solid samples). An attempt was made to analyze aqueous glucose solutions in a disposable cell (5) built around the fiber probe for the following reasons:

- A PbS cell that is physically smaller than the one in the sample drawer is used as the

detector in the fiber probe. Smaller detectors are generally less noisy in the near-IR because thermal gradients in the semiconductor material are reduced.

- The temperature of the liquid microcell is easier to control with a heat exchanger and heat sink situated away from the heated spectrometer compartments and electronics. The optical-fiber configuration allows the sample to be analyzed up to five feet away from the spectrometer.

The liquid microcell in the sample drawer gave an average S/N of 14,000. The calculated S/N of the fiber-optic liquid microcell is 27,000.

Ten glucose solutions, with concentrations ranging from 0 to 0.005 M, were prepared for analysis in the fiber-optic microcell. The calibration process produced an SEE of 0.009 M glucose.

A closer inspection of the data revealed an interesting trend in the fitting process residuals. In the analysis of the HSA data, two sample cells (no. 4 and no. 6) showed large systematic prediction errors, and as a result these cells were not used in the new glucose calibration. However, as the reproducibility of the sampling apparatus increased, it became apparent that other sample cells showed systematic errors as well. The regularity in the residuals appears to arise from the fact that the six disposable sample cells used in this study did not match perfectly and, in fact, showed systematic errors: no. 1 on average predicted 0.0003 M high, no. 2 predicted 0.0001 M high, no. 3 predicted 0.0006 M low, and no. 5 predicted 0.0003 M low.

A calibration plot corrected for errors introduced by the sample cells appears in Figure 10. This calibration produced an SEE of 0.0005 M ( $r^2 = 0.96$ ) and an SEP of 0.0007 M, which corresponds to an approximate detection limit of 0.002 M.

Finally, it should be noted that an examination of the elements of the transformation matrix corresponding to the second principal component (the principal component that correlated most strongly to glucose concentration) showed that the strongest peak was in the 1400–1460-nm region, a region that was not accessible (in other words, showed no glucose signal) when the liquid microcells were used in the sample drawer.

## CONCLUSIONS

The weaker correlation noted between pH and near-IR spectra may be due to the use of a low-volume liquid cell with a large surface area, the time the solutions spent equilibrating in the cell (1 h), heat transfer between the monochromator compartment and the sample-cell drawer during scanning, and not prewashing the sample cell with the sample solution before recording spectra.

A true double-beam experiment, using a PBS reference solution, would probably help to compensate for changes in spectra caused by temperature variations in water. Focusing the near-IR light down on a small transmission cell should also permit the use of longer optical path lengths in analyses.

The calculated noise level for the sample cell-heat exchanger system used in this study is about one-half of that reported for commercial cells. The preliminary data from the new fiber-optic cell indicate that the cell has ~25% the amount of noise reported for aqueous solutions in most commercial cells. Good correlations between near-IR spectra and analytes have been observed at concentration levels as low as 10<sup>-5</sup> M, which is low that ordinarily considered amenable to near-IR analysis. It seems likely that the trends observed in this study (for example, that certain wavelength

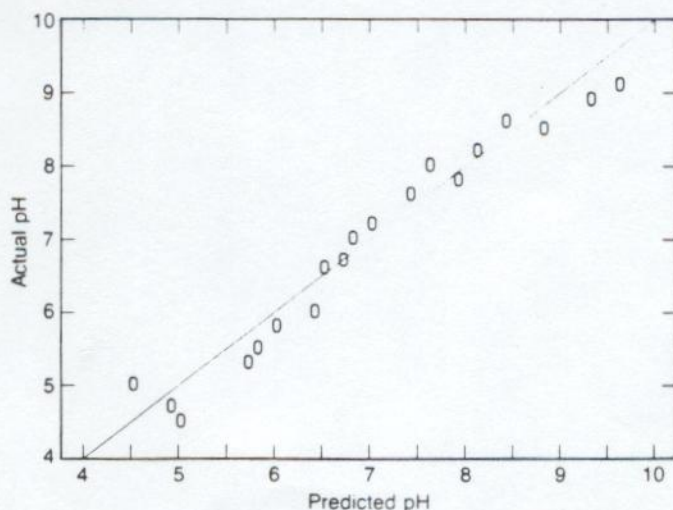


Figure 9. Calibration line using the smoothed second-derivative spectra for the pH of

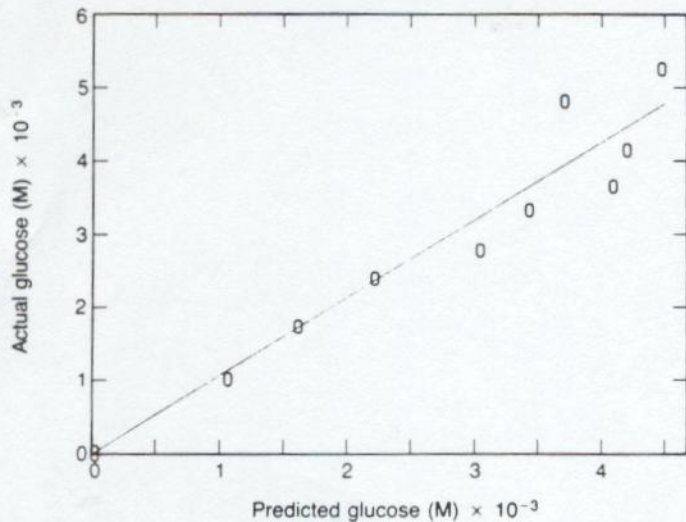


Figure 10. A calibration plot corrected for errors introduced by the sample cells.

centration of HSA in test solutions) will also be observed in future experiments using improved equipment.

#### ACKNOWLEDGMENTS

This research was supported in part by the National Science Foundation Grant No. RII-

8610671, the Commonwealth of Kentucky through the University of Kentucky's Center for Computational Sciences, and BRSG S07 RR05857-09 awarded by the Biomedical Research Support Grant Program, Division of Research Resources, National Institutes of Health.

#### REFERENCES

- (1) R.A. Lodder, M. Selby, and G.M. Hieftje, *Anal. Chem.* **59**, 1921 (1987).
- (2) D.E. Honigs, T.B. Hirschfeld, and G.M. Hieftje, *Anal. Chem.* **57**, 443 (1985).
- (3) D.L. Wetzel, *Anal. Chem.* **55**, 1165A (1983).
- (4) R.A. Lodder, Paper no. 1096, presented at the Pittsburgh Conference, Atlanta, Georgia, March 1989.
- (5) R.A. Lodder and G.M. Hieftje, *Appl. Spectrosc.* **42**, 518 (1988).

**James K. Drennen** is an assistant professor of pharmaceutical chemistry and pharmaceuticals at the Duquesne University School of Pharmacy, where he is exploring the theoretical limits of near-infrared spectroscopy.

**Brian D. Gebhart** received his MS in analytical chemistry from Indiana University in 1989. He is currently engaged in Fourier transform discriminant analysis studies at the Midwest Research Institute in Kansas City, Missouri.

**Elizabeth J. Kraemer** received a BS in psychology from Susquehanna University. She is an experimental psychologist and is currently working on biological applications of near-IR spectrometry.

**Robert A. Udelson** is an assistant professor in the College of Pharmacy at the University of Kentucky Chandler Medical Center, Rose Street, Lexington, Kentucky 40536-0082. He received the Thomas Hirschfeld Award in near-IR spectroscopy in 1988. Please address correspondence to Dr. Lodder. ■■