

Comparison of Combination and First Overtone Spectral Regions for Near-Infrared Calibration Models for Glucose and Other Biomolecules in Aqueous Solutions

Jun Chen,[†] Mark A. Arnold,^{*,†} and Gary W. Small[‡]

Department of Chemistry and Optical Science and Technology Center, University of Iowa, Iowa City, Iowa 52242, and Center for Intelligent Chemical Instrumentation, Department of Chemistry, Ohio University, Athens, Ohio 45701

Partial least squares calibration models are compared for the measurement of glucose, lactate, urea, ascorbate, triacetin, and alanine in aqueous solutions from single-beam spectra collected over the first overtone (6500–5500 cm⁻¹) and the combination (5000–4000 cm⁻¹) regions of the near-infrared spectrum. Spectra are collected under two sets of conditions with one designed for combination spectra and the other designed for first overtone spectra. As part of the optimization of conditions, an exponential function is presented that accurately characterizes the strong dependency between spectral quality and sample thickness. Sample thickness set for the first overtone and combination spectra are 7.5 and 1.5 mm, respectively. Independent calibration models are established for each solute from both combination and first overtone spectra. Direct comparison reveals superior performance by models generated from combination spectra, particularly for glucose and urea. Standard error of prediction (SEP) values are 1.12 and 0.45 mM for glucose models generated from first overtone and combination spectra, respectively. SEP values for urea are 7.33 and 0.10 mM for first overtone and combination spectra, respectively. Such high SEP values for urea with first overtone spectra correspond to an inability to quantify urea from these spectra because of a lack urea-specific molecular absorption features in this spectral region. Net analyte signal (NAS) is used to quantify the degree of selectivity provided within the first overtone and combination spectral regions. The superior selectivity of combination spectra is confirmed by comparing the length of the NAS vectors for each matrix component.

Near-infrared spectroscopy is widely used for analytical measurements in complex matrixes.^{1–6} A key feature of this methodology is the ability to extract quantitative information for numerous

substances from a single near-infrared spectrum. In addition, near-infrared spectroscopic analyses are nondestructive, simple, fast, and require no sample pretreatment, which makes this technology ideally suited for *on line* process monitoring and quality control.

Biological and clinical applications of near-infrared spectroscopy frequently involve measurements in samples composed primarily of water. An important special case is the measurement of glucose in complex biological matrixes.⁷ Even in matrixes as complex as whole blood, water is the single most important component that defines the fundamental analytical performance. Measurements in aqueous samples are complicated by the strong near-infrared absorption properties of water,^{8,9} which accentuate the impact of sample thickness,¹⁰ instrumental signal-to-noise ratio, and spectral range.

Figure 1 presents the near-infrared transmission spectrum for a 1.5-mm thick sample of water maintained at 37 °C. As illustrated in this figure, three windows of transmission are available between the dominating water absorption bands centered at approximately 6900, 5200, and 3800 cm⁻¹.^{11,12} These windows correspond to the combination (4800–4200 cm⁻¹), first overtone (6700–5400 cm⁻¹), and the short wavelength near-infrared (12 000–7100 cm⁻¹) spectral regions. Absorption features in the combination region correspond to the combination of bending and stretching vibrations associated with C–H, O–H, and N–H bonds. Absorptions in the first overtone region correspond to the first overtone of C–H bending vibrations and higher order combinations and overtones associated with C–H, N–H, and O–H bonds are found in the short wavelength region.⁷

- (3) Haaland, D. M.; Robinson, M. R.; Eaton, R. P.; Koepp, G. W. *Appl. Spectrosc.* **1992**, *46*, 1575–1578.
- (4) Riley, M. R.; Crider, H. M.; Nite, M. E.; Garcia, R. A.; Woo, J.; Wegge, R. M. *Biotechnol. Prog.* **2001**, *17*, 376–378.
- (5) Heise, H. M.; Marbach, R.; Bittner, A.; Koschinsky, T. *J. Near Infrared Spectrosc.* **1998**, *6*, 361–374.
- (6) LaFrance, D.; Lands, L. C.; Hornby, L.; Burns, D. H. *Appl. Spectrosc.* **2000**, *54*, 300–304.
- (7) Heise, H. M. In *Handbook of Vibrational Spectroscopy*; Chalmers, J. M., Griffiths, P. R., Eds.; John Wiley & Sons, Inc.: New York, 2002; Vol. 5, pp 3280–3294.
- (8) Dull, G. G.; Giangiacomo, R. *J. Food Sci.* **1984**, *49*, 1601–1603.
- (9) Arnold, M. A.; Small, G. W. *Anal. Chem.* **1990**, *62*, 1457–1464.
- (10) Jensen, P. S.; Bak, J. *Appl. Spectrosc.* **2002**, *56*, 1600–1606.
- (11) Bayly, J. G.; Kartha, V. B.; Sevens, W. H. *Infrared Phys.* **1963**, *3*, 211–223.
- (12) Bertie, J. E.; Lan, Z. *Appl. Spectrosc.* **1996**, *50*, 1039–1046.

* Corresponding author. E-mail: mark-arnold@uiowa.edu. Fax: 319-353-1115.

[†] University of Iowa.

[‡] Ohio University.

- (1) Arnold, M. A. In *Handbook of Clinical Laboratory Automation, Robotics, and Optimization*; Kost, G. J., Ed.; John Wiley & Sons: New York, 1996; pp 631–647.
- (2) Marbach, R.; Koschinsky, T. H.; Gries, F. A. *Appl. Spectrosc.* **1994**, *47*, 875–881.

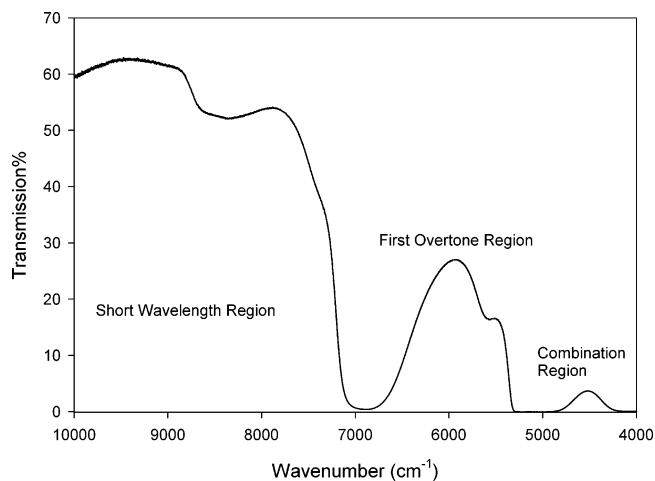


Figure 1. Near-infrared transmission spectrum of water relative to air collected with an optical path length of 1.5 mm.

Relative transmittance differs greatly for these three transmission windows. As illustrated in Figure 1, the highest transmittances are in the short wavelength region with lower values in the first overtone region and still lower transmittance levels in the combination region. Low molecular absorptivities in the short wavelength region result in high optical throughput. This high throughput is critical for the success of pulse-oximetry and related oxygen measurements, which are based on the electronic spectra of oxy- and deoxyhemoglobin.^{13–15} Despite some reported success,¹⁶ analytical measurements are severely limited in the short wavelength region because of low absorptivities for molecular vibrations. For this reason, the first overtone and combination regions are principally available for measurements in complex biological matrices.

Figure 1 indicates that the peak transmittance is approximately 5-fold higher for the first overtone region compared to the combination region. This 5-fold increase in optical throughput allows for thicker samples, which increases the optical path length and, according to the Beer–Lambert law, enhances sensitivity of the measured absorbance signal. On the other hand, molecular absorptivities are lower in the first overtone spectral region compared to the combination region, thereby offsetting this enhancement.

From a practical standpoint, the difference in transmittance between the first overtone and combination spectral regions makes it difficult to collect optimum spectra in both regions simultaneously. Longer sample thickness is necessary to achieve optimum signal-to-noise for first overtone spectra compared to combination spectra. Typically, sample thickness is 5 to 10 mm for the first overtone region, which is too thick for quality combination spectra. Conversely, sample thicknesses of 0.5 to 2 mm are commonly used for the collection of combination spectra, which is too thin for high-quality first overtone spectra. Indeed, Hazen et al. demonstrated that a sample thickness of 2 mm is insufficient for measuring glucose in simple buffer solutions.¹⁷

Given this mismatch in sample thickness, the use of a single sample thickness is not possible for collecting high signal-to-noise ratio spectra for each spectral region simultaneously. A fundamental question is which of these two spectral regions provides superior analytical results for the measurement of these analytes.

This report focuses on the fundamental selectivity properties of near-infrared spectroscopy over the first overtone and combination spectral regions. Calibration models are generated from separately optimized first overtone and combination spectra. Partial least squares (PLS) calibration models are generated from a set of sample solutions composed of different concentrations of glucose, lactate, alanine, urea, ascorbate, and triacetin. Individual calibration models are generated for each solute over each spectral region. Results indicate that combination spectra provide greater chemical distinction, thereby resulting in superior calibration models.

EXPERIMENTAL SECTION

Instrumentation and Apparatus. Spectra were collected with a Nicolet 670 Nexus Fourier transform infrared spectrometer. The spectrometer was equipped with a 20-watt tungsten–halogen light source, a calcium fluoride beam-splitter, and a cryogenically cooled InSb detector. When collecting combination spectra, a K-band interference filter (5000–4000 cm^{-1} , Barr and Associates, Inc.) was positioned before the sample. Similarly, an H-band interference filter (6500–5500 cm^{-1} , Barr and Associates, Inc.) was used to restrict light to the first overtone region. Samples were placed in a water-jacketed thermostated sample cell equipped with sapphire windows. Sample temperature was maintained at 37.0 ± 0.1 °C using a VWR circulating water bath. A copper–constantan thermocouple probe (Omega, Inc., Stamford, CT) was used to monitor the sample temperature while collecting spectra. The optical path length was 1.5 mm and 7.5 mm in the combination and first overtone spectral regions, respectively. These values match those used in previously reported work and represent a reasonable compromise between magnitude of the absorbance signal and spectral noise.^{17–21} Path lengths were set with Teflon spacers between sapphire windows.

Reagents. Glucose, sodium lactate, sodium ascorbate, and alanine were obtained from Aldrich Chemical Co., Inc. Urea and triacetin were purchased from Sigma Chemical Co. Phosphate buffer solution was prepared by dissolving 3.4023 g of potassium dihydrogen phosphate and 3.5495 g of sodium monohydrogen phosphate into 1 L of distilled water. In addition, 0.2 g/L 5-fluorouracil was added as an antimicrobial agent. The pH of the final solution was 6.86 at 25 °C.

Procedures. Eighty unique sample solutions were prepared by carefully weighing masses of standard material and dissolving in known volumes of buffer. Error propagation indicates a relative concentration uncertainty of 0.2% for these solutions, which corresponds to a maximum uncertainty of 0.07 mM. This level of uncertainty ultimately limits the analytical performance of all PLS calibration models. Solutions were prepared shortly before use

(13) Zijlstra, W. G.; Buursma, A.; Meeuwse-Van der Roest, W. P. *Clin. Chem.* **1991**, *37*, 1633–1638.

(14) Jeon, K. J.; Kim, S.; Park, K. K. *J. Biomed. Opt.* **2002**, *7*, 45–50.

(15) Zhang, S.; Soller, B. R.; Kaur, S.; Perras, K.; Vander Salm, T. J. *Appl. Spectrosc.* **2000**, *54*, 294–299.

(16) Bittner, A.; Marbach, R.; Heise, H. M. *J. Mol. Struct.* **1995**, *349*, 341–344.

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

(18) Burmeister, J. J.; Arnold, M. A.; Small, G. W. *LEOS Newsl.* **1998**, 6–9.

(19) Heise, H. M.; Bittner, A. *Fresenius' J. Anal. Chem.* **1998**, *362*, 141–147.

(20) Marquardt, L. A.; Arnold, M. A.; Small, G. W. *Anal. Chem.* **1993**, *65*, 3271–3278.

(21) Mattu, M. J.; Small, G. W. *Anal. Chem.* **1997**, *69*, 4695–4702.

Table 1. Correlation Coefficients between Component Concentrations

	lactate	urea	ascorbate	alanine	triacetin
glucose	0.0065	0.0047	0.0002	0.0167	0.0085
lactate		0.0177	0.0119	0.0008	0.0002
urea			0.0024	0.0067	0.0001
ascorbate				0.0031	0.0326
alanine					0.0001

by dissolving the desired amount of each component directly into the phosphate buffer. Solution pH was monitored during preparation and the maximum change in pH was less than 0.05. No adjustments were made to correct for these minor pH changes. Prior to collecting spectra, the sample of interest was placed in the thermostated cell located within the instrument. A 5-min incubation period was imposed to allow the solution to thermally equilibrate at the targeted temperature.

Triplicate spectra were collected consecutively without removing the sample from the spectrometer. These repeated spectra were used to compute 100% lines from which RMS noise values were obtained to characterize spectral quality. Spectra were collected as 256 coadded double-sided interferograms that were subsequently converted to single-beam spectra with a 1.9 cm^{-1} point spacing. Interferograms were triangularly apodized with Mertz phase correction. Zero-filling was not used, which resulted in a spectral resolution of 4 cm^{-1} . In total, 240 single-beam spectra were collected in each spectral region. Neither the spectral resolution nor the apodization function was optimized for this experiment. In previous work, measurement performance was relatively insensitive to spectral resolution.²² Little is reported on the impact of the apodization function on analytical performance of multivariate calibration models originating from near-infrared spectra. Although the exact nature of the apodization function is known to affect band shapes, particularly for mid-infrared and Raman spectra,²³ the effect of the apodization function is minimal on the performance of multivariate calibration functions, such as partial least squares, when applied to mid-infrared spectra.²⁴

PLS calibration models were built on the basis of single-beam spectra. A previously described procedure was used to optimize calibration models in terms of spectral range and number of latent variables.²⁵ Briefly, the 80 samples were randomly divided into a calibration data set with 195 spectra (65 samples) and an independent prediction data set with 45 spectra (15 samples). The calibration data set was then randomly divided into a training data set with 150 spectra (50 samples) and a monitoring data set with 45 spectra (15 samples). The division of the training and monitoring data sets was done three times. For each rearrangement of training and monitoring data sets, one optimized calibration model was generated based on standard error of training (SET) and standard error of monitoring (SEM). The optimum calibration model was determined from the results of the three rearrangements of the calibration data set. A modified grid search was used

to identify the best spectral range, and a minimum SEM with the fewest number of significant factors was used to identify model rank. The resulting calibration model was validated by predicting concentrations from spectra in the independent prediction data set. Model performance was evaluated by the standard error of calibration (SEC), standard error of prediction (SEP), and mean percent error (MPE), according to the following equations:

$$\text{SEC(SET)} = \sqrt{\frac{\sum(C_a - C_p)^2}{n_c - f - 1}} \quad (1)$$

$$\text{SEP(SEM)} = \sqrt{\frac{\sum(C_a - C_p)^2}{n_p}} \quad (2)$$

$$\text{MPE} = \frac{\sum(|C_a - C_p|/C_a)}{n_p} \times 100\% \quad (3)$$

where n_c and n_p correspond to the number of spectra in the calibration (training) and prediction (monitoring) data sets, respectively, f is the number of latent variables or PLS factors, and C_a and C_p represent the reference and predicted concentrations of analyte, respectively. All statistical testing of significance used the standard t -test at the 95% confidence level.

RESULTS AND DISCUSSION

Analyte Concentration Covariance. Concentration of each component was randomly assigned in order to minimize concentration covariance between components. Covariance between components must be minimized to avoid the use of nonanalyte spectral information in the PLS calibration model.²⁶ Correlation coefficients are listed in Table 1 for the resulting concentration correlations between every two components in the six-component mixtures. These correlation coefficients indicate no significant covariance at the 95% confidence level between any two components in this experiment.

Comparison of Spectral Features. Individual absorbance spectra are presented in Figure 2 for each solute. Each spectrum was collected from a standard solution of the indicated solute dissolved in the aqueous buffer solution. The resulting spectral features correspond to a combination of solute absorptivity and solvent (water) displacement.²⁷ Overtone spectra are presented in Figure 2A, and combination spectra are plotted in Figure 2B. Several critical features are highlighted by these spectra. First, the absorbance magnitude is small for all absorption bands. Small absorptivities are characteristic of these vibrational transitions, which demands high signal-to-noise ratios for millimolar limits of detection. Second, the spectral features are highly overlapping in nature. This high degree of overlap is characteristic of both combination and overtone vibrational spectra. Selective measurements from such complex overlapping structures demand multivariate calibration techniques. The third feature of interest highlighted within this figure is the uniqueness of the individual spectra. Although highly overlapping, each spectrum is different,

(22) Eddy, C. V.; Flanigan, M.; Arnold, M. A. *Appl. Spectrosc.* **2003**, *57*, 1230–1235.

(23) Parker, S. F.; Tooke, P. B. *Spectrochim. Acta* **1997**, *53*, 2245–2252.

(24) Reeves, J. B.; Reeves, V. B. *Spectrosc. Lett.* **2002**, *35*, 663–680.

(25) Riley, M. R.; Rhiel, M.; Zhou, X.; Arnold, M. A.; Murhamimer, D. W. *Biotechnol. Bioeng.* **1997**, *55*, 11–15.

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

(27) Amerov, A. K.; Chen, J.; Arnold, M. A. *Appl. Spectrosc.*, in press.

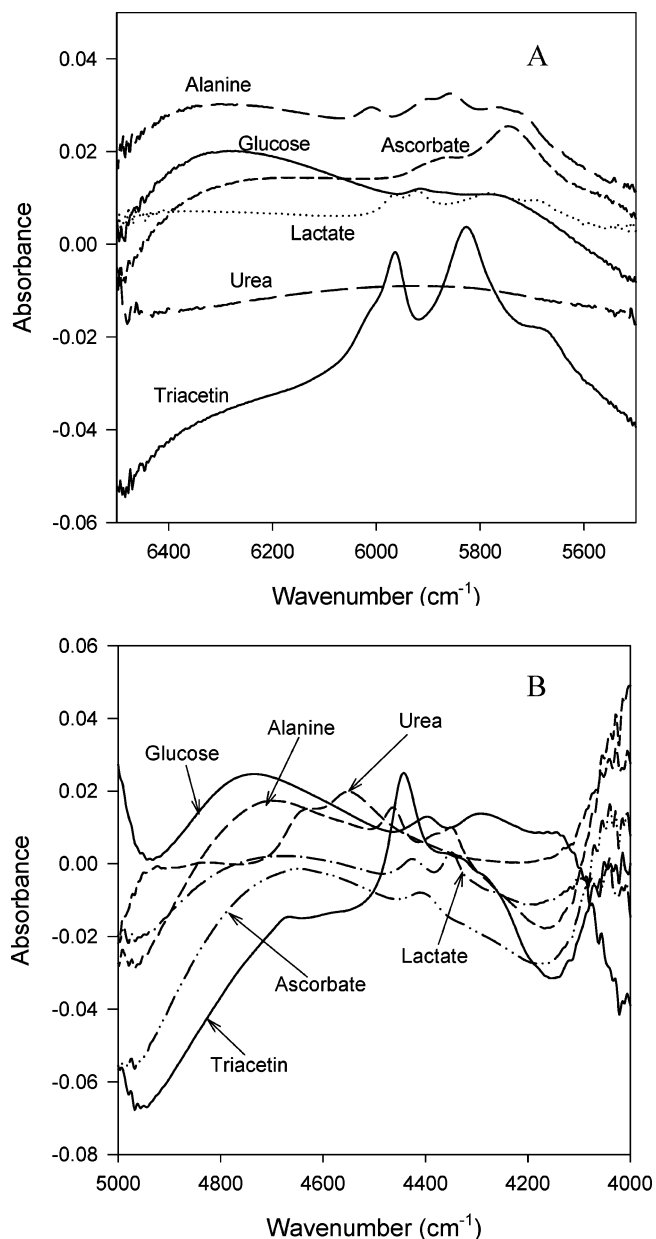


Figure 2. Absorbance spectra of glucose, lactate, alanine, ascorbate, triacetin, and urea over the first overtone (A) and combination (B) spectral regions. Spectra were collected using 100 mM solutions of each components and blank buffer as the reference material.

which is the fundamental molecular basis of selectivity for near-infrared measurements.

Differences in the first overtone and combination spectra are apparent by comparing the spectra in parts A and B of Figure 2. As noted above, the absorption bands are highly overlapping in both cases. The first overtone absorbance spectra are somewhat less complicated, however. Absorption features corresponding to N–H and O–H bonds are not present within the first overtone spectral region. Both N–H and O–H overtones are located under the nearby water absorption band and are not observable within the available first overtone wavelengths. The reduction of complexity is exemplified by the spectrum of urea, where the absence of C–H bonds within the urea molecule results in no significant absorption features in the first overtone spectral region. In fact, the first overtone absorption bands for the urea N–H stretches

are located around 6800 cm^{-1} .²⁸ On the other hand, characteristic absorption bands are clearly evident for urea in the combination spectrum in Figure 2B. Two features are apparent for urea, which correspond to the combination of the symmetric and asymmetric stretching vibrations coupled with the bending vibration mode.

The first overtone and combination spectra also differ in terms of the magnitude and shape of the absorption bands. As noted above, absorptivities are larger for combination features. In fact, absorptivities are generally 4- to 5-fold higher for combination spectra.²⁷ The impact of higher absorptivities is diminished because sample thickness is restricted for aqueous samples. Differences in band shape are also apparent in a comparison of parts A and B of Figure 2. Absorption bands in the combination spectral region are narrower and sharper, particularly at the lower wavenumbers ($4200\text{--}4500\text{ cm}^{-1}$), which correspond to different C–H bonding environments.

Critical differences in these spectral regions can be quantified on the basis of the net analyte signal (NAS) for each solute. NAS, as defined by Lorber,²⁹ is the portion of the solute spectrum that is orthogonal to all other sources of spectral variance in the data matrix. The NAS is easily calculated by the following equation:

$$a_g^* = (\mathbf{I} - \mathbf{A}_g \mathbf{A}_g^+) a_g \quad (4)$$

where a_g represents the pure component spectrum of glucose, a_g^* represents the NAS for glucose in the six-component mixture, and \mathbf{I} represents the identity matrix that has the same dimension as $\mathbf{A}_g \mathbf{A}_g^+$. \mathbf{A}_g represents the matrix of spectra for all the components in the mixture except glucose. The spectral variation caused by fluctuation of instrumental and environmental conditions is also included in the \mathbf{A}_g matrix. \mathbf{A}_g^+ represents the Moore–Penrose pseudo inverse of \mathbf{A}_g .³⁰ The matrix $(\mathbf{I} - \mathbf{A}_g \mathbf{A}_g^+)$ is a projection matrix that projects a_g onto the null space of the rows of \mathbf{A}_g , which is the orthogonal complement of the column space of \mathbf{A}_g . The reasoning for this definition stems from the properties of solving a set of linear equations because the part of the spectrum that is not orthogonal to the other components can be represented as a linear combination of the spectral features of the others.

From the standpoint of a multivariate spectral analysis, such as principal components regression or partial least squares (PLS) regression, selectivity corresponds to differences in spectral shape, or differences in the direction of the corresponding spectral vectors. Likewise, NAS indicates the degree of uniqueness of the analyte spectrum in a highly overlapping system. The length of the NAS vector per unit concentration and unit path length represents the degree of difference between the analyte and all other sources of spectral variance. NAS vectors can be computed over both the first overtone and combination spectral regions and the lengths compared. Larger normalized vectors represent greater distinction between the analyte and the chemical matrix.

NAS vector lengths are listed in Table 2 for each solute and for each spectral range. Each value is normalized for both

(28) Shenk, J. S.; Workman, J. J., Jr.; Westerhaus, M. O. In *Handbook of Near-Infrared Analysis*, 2nd ed.; Burns, D. A., Ciurczak, E. W., Eds.; Marcel Dekker: New York, 2001; Vol. 27, pp 419–474.

(29) Lorber, A. *Anal. Chem.* **1997**, *69*, 1620–1626.

(30) Lawson, C. L.; Hanson, R. J. *Solving Least Squares Problems*; Prentice Hall: Englewood Cliffs, NJ, 1974.

Table 2. Length of Net Analyte Signal Vectors^a

	first overtone region ^b	combination region ^c	ratio ^d
glucose	9.17×10^{-6}	3.52×10^{-5}	3.84
lactate	4.38×10^{-5}	1.46×10^{-4}	3.33
urea	3.84×10^{-6}	2.44×10^{-4}	63.5
ascorbate	2.62×10^{-5}	4.34×10^{-5}	1.66
alanine	2.24×10^{-4}	2.69×10^{-4}	1.20
triacetin	1.36×10^{-4}	9.13×10^{-4}	6.17

^a Normalized for concentration and path length. Units: AU·mM⁻¹·mm⁻¹. ^b 6500–5800 cm⁻¹. ^c 4800–4200 cm⁻¹. ^d Combination/first overtone.

concentration and path length. In addition, the ratio is provided for the combination NAS vector length relative to the first overtone NAS vector length. For each solute, the NAS vector is larger for the combination spectra, which indicates that for this set of solutes the combination spectral region provides greater distinction relative to the first overtone spectral region. The ratio indicates a 3.8-fold improvement in selectivity for glucose by using combination spectra. The ratio for urea is even higher, which is consistent with the lack of urea absorption bands in the first overtone spectral region. Overall, this selectivity analysis predicts superior performance for multivariate calibration models consisting of combination spectra compared to first overtone spectra.

Spectral Quality. Spectral quality is critical for both selectivity and sensitivity.^{17,18,22,31} In this work, spectral quality was characterized as the root-mean-square (RMS) noise on 100% lines computed for replicate spectra for each sample.¹⁷ RMS values were computed in absorbance units relative to a second-order polynomial fit of the 100% line data across sequential 100 cm⁻¹ segments. Figure 3 shows the distribution of RMS noise levels across the first overtone (Figure 3A) and combination (Figure 3B) spectra collected in this experiment. Values plotted in Figure 3 for the first overtone and combination spectra correspond to the 5900–6000 and 4400–4500 cm⁻¹ spectral regions, respectively.

The data plotted in Figure 3A,B indicate consistent spectral quality across all samples for both data sets. Mean values (\pm standard deviation) are $7.8 \pm 1.1 \mu\text{AU}$ and $2.0 \pm 0.4 \mu\text{AU}$ for the first overtone and combination data sets, respectively. The corresponding relative standard deviations are 14.1% and 20.0%, respectively. Such consistency requires precise control of sample temperature for each replicate spectrum. In addition, these RMS noise levels are inversely related to the measurement signal-to-noise ratio (SNR), where the $\text{SNR} = (2.303 \times \text{RMS})^{-1}$.^{22,32} Correspondingly, average SNRs are 56 000 and 217 000 over the 5900–6000 and 4500–4400 cm⁻¹ spectral regions, respectively. As discussed previously,^{17,22} higher RMS noise levels (lower SNRs) are observed outside these spectral regions because of the reduced optical throughput caused by the absorption of light by water.

For aqueous solutions, spectral quality is strongly dependent on sample thickness. RMS noise increases as the sample thickness increases, owing to water absorption. On the other hand, shorter path lengths limit sensitivity of the absorbance measurement, thereby reducing analytical performance. In practice, the optical

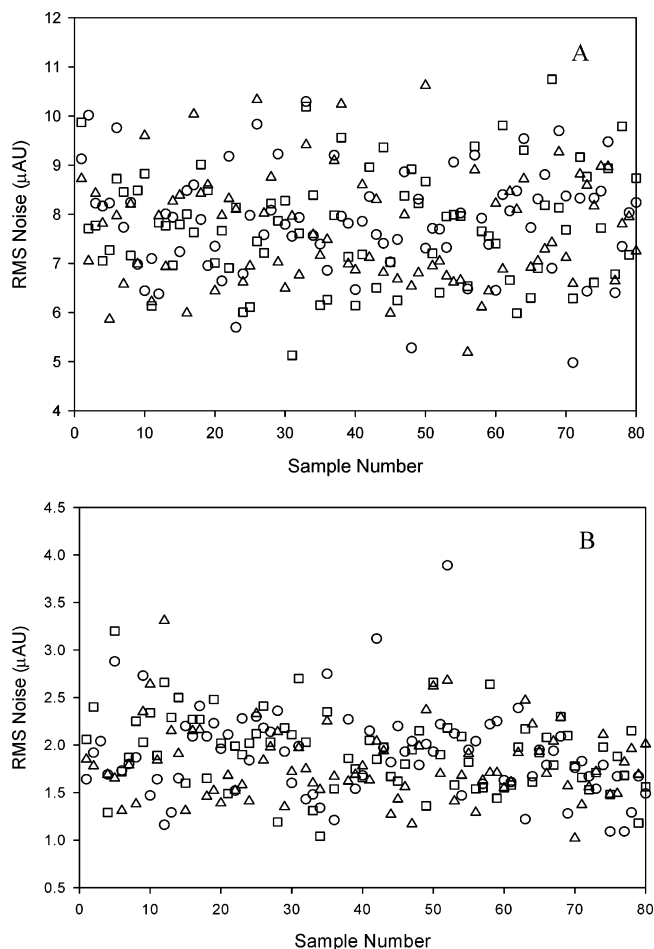


Figure 3. Distribution of RMS noise levels for spectra collected over the first overtone (A) and combination (B) spectral regions. Values are presented for noise using the first and second (circles), second and third (triangles), and first and third (squares) spectra of the triplicate measurements.

path length or sample thickness must be a compromise between low RMS noise and high measurement sensitivity.

Sample thickness and RMS noise are related by the following expression:

$$\text{RMS} = \frac{N}{2.303I_0} e^{2.303\epsilon bc} \quad (5)$$

where N is the spectral noise underlying the measurement, I_0 is the incident light intensity, and ϵ , b , and c , correspond to the molar absorptivity, thickness, and concentration of water, respectively. Accordingly, the RMS noise should increase exponentially as a function of sample thickness. Figure 4 presents the measured RMS noise as a function of sample thickness for both the first overtone and combination spectral regions. Individual values are plotted as the mean (\pm standard deviation) for 24 measurements at each sample thickness. RMS noise is computed from 100% lines in absorbance units over the 5900–6000 and 4400–4500 cm⁻¹ spectral region for plots A and B, respectively. In addition, these data were fit to eq 5. The fitted regression coefficients are provided in the figure, and the magnitude of the resulting function is plotted as a solid line in each plot.

(31) Wetzel, D. L.; Sweat, J. A. *Mikrochim. Acta, Suppl.* **1997**, *14*, 325–327.

(32) Griffiths, P. R.; de Haseth, J. A. *Fourier Transform Infrared Spectrometry*; John Wiley & Sons: New York, 1984.

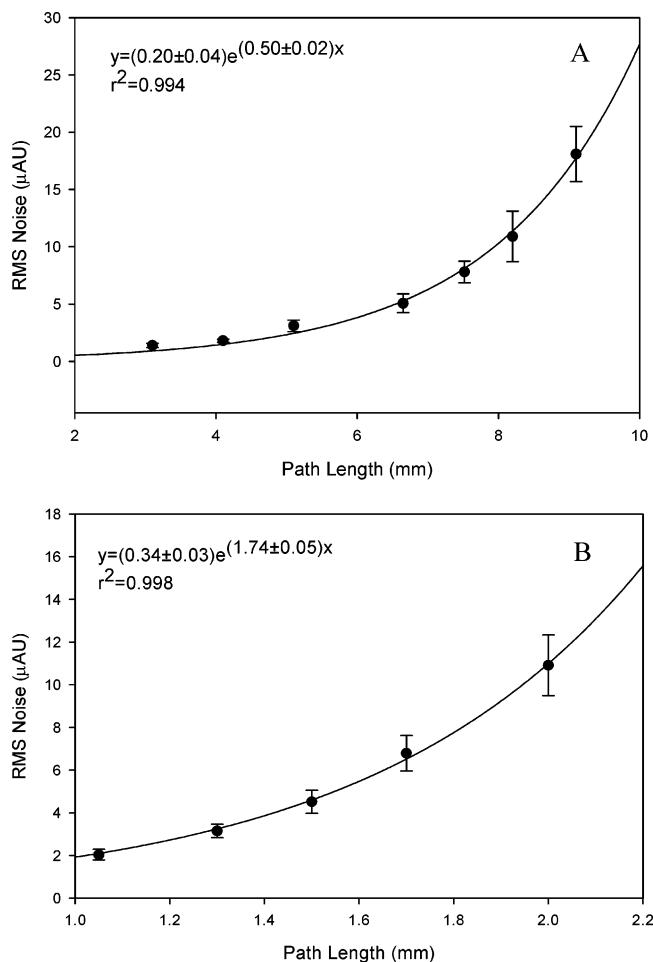


Figure 4. RMS noise as a function of sample thickness for the first overtone (A) and combination (B) spectral regions. Plotted points correspond to the mean (\pm standard deviation) of 24 measurements at each sample thickness. RMS noise is computed from 100% lines in absorbance units over the 5900–6000 and 4400–4500 cm^{-1} spectral regions for plots A and B, respectively. Solid lines correspond to exponential fits according to eq 5, and the fitted equation is provided together with the regression coefficient (R -squared).

In both cases, the data are well modeled by the exponential function in eq 5. Regression coefficients for the exponential terms ($2.303 \epsilon c$) are 0.50 ± 0.02 and $1.74 \pm 0.05 \text{ mm}^{-1}$, respectively, for the first overtone and combination spectral regions. These values compare favorably to the values of 0.54 mm^{-1} and 1.90 mm^{-1} , which were computed by using the mean absorptivity of water over these spectral regions¹¹ and the density of water.³³

Optical path lengths of 7.5 and 1.5 mm were used for the first overtone and combination spectra, respectively. These sample thicknesses are consistent with previous work in these spectral ranges^{17–21} and are known to provide spectra with sufficiently low noise and high sensitivities for the measurement of millimolar concentrations.

Calibration Models. Separate PLS calibration models were generated for each solute over the two spectral regions. Table 3 summarizes the optimized spectral range, number of latent variables, SEC, and SEP for each solute over each spectral region. To varying degrees, each solute can be quantified by this method, with the notable exception of urea from first overtone spectra. With the exclusion of the first overtone urea model, prediction

errors range from 0.33 to 1.12 mM for first overtone models and from 0.10 to 0.53 mM for combination models. Generally, combination models outperform first overtone models, which is consistent with greater distinction of combination spectra as discussed above (see Table 2).

Glucose results are characterized by lower measurement errors for the combination model compared to the first overtone model. The corresponding concentration correlation plots are presented in Figure 5A,B. Consistent with lower SEC and SEP values, scatter about the unity line is noticeably reduced for the combination model. The combination model provides a MPE of 7.3% for glucose. The best spectral range for glucose measurements is 4560–4100 cm^{-1} , which corresponds to the C–H bonding region of the combination spectral range.

Measurement bias as a function of cosolute concentration is demonstrated by the residual plots presented in Figure 6. This figure shows residuals for the prediction data set as a function of the concentration of the other solution components. Overall, prediction residuals are independent of companion solute concentration. No systematic bias is evident from these plots, and there is no significant correlation between the prediction residual and the concentration of any other component at the 95% confidence level.

Urea results are also excellent for the combination model, where the MPE is only 1.3%. From combination spectra, urea can be measured very well with the lowest SEC and SEP values (0.16 and 0.10 mM, respectively) of all solutes. The optimized spectral range is 4650–4380 cm^{-1} , which includes a unique absorption peak for urea at 4550 cm^{-1} . It is noteworthy that no calibration model is possible for urea from first overtone spectra. This finding is consistent with the lack of significant urea absorption features in the available spectral data. The corresponding SEP and SEC values are 7.33 and 7.32 mM, respectively, which are similar to the standard deviation of the urea distribution in the prepared samples (7.98 mM). The inability to model urea from first overtone spectra demonstrates the effectiveness of the experimental design to minimize concentration correlations between solutes.

Calibration models for the other components (lactate, ascorbate, triacetin, and alanine) are similar to those for glucose, in that functional calibration models are obtained from both first overtone and combination spectra and no systematic correlations are evident in the corresponding residual plots. Although not as drastic as for glucose, calibration models from combination spectra generally outperform models from first overtone spectra. This finding is evident from the listing of SEC and SEP values in Table 3.

Comparison to Previous Reports. Other researchers have compared the first overtone and combination spectral ranges for analyte measurements in aqueous solutions. Examples include reports by Kasemsumran et al.³⁴ and Heise and co-workers.⁵ In the Kasemsumran paper, a moving window algorithm is used to find the ideal spectral range for measuring glucose in a phosphate buffer matrix containing albumin and globulin protein. In the Heise work, multiple components, including protein, glucose, cholesterol, and triglycerides, are measured in samples of human plasma.

(33) In *Handbook of Chemistry and Physics*, 59 ed.; CRC Press Inc.: Boca Raton, FL, 1979; pp F-11.

(34) Kasemsumran, S.; Du, Y.; Murayama, K.; Huehne, M.; Ozaki, Y. *Analyst* **2003**, *128*, 1471–1477.

Table 3. Summary of Calibration Models with Combination and First Overtone Spectra

	optimized range (cm ⁻¹)		number of factors		SEC (mM)		SEP (mM)	
	combination	first overtone	combination	first overtone	combination	first overtone	combination	first overtone
glucose	4560–4100	6030–5650	11	9	0.31	1.24	0.45	1.12
urea	4650–4380	6150–5950	11	7	0.16	7.32	0.10	7.33
lactate	4550–4250	6220–5780	10	9	0.20	0.27	0.26	0.33
triacetin	4510–4400	5920–5620	7	9	0.13	0.27	0.18	0.22
alanine	4540–4350	6080–5800	7	8	0.18	0.39	0.20	0.25
ascorbate	4560–4000	5980–5580	12	8	0.33	0.55	0.53	0.45

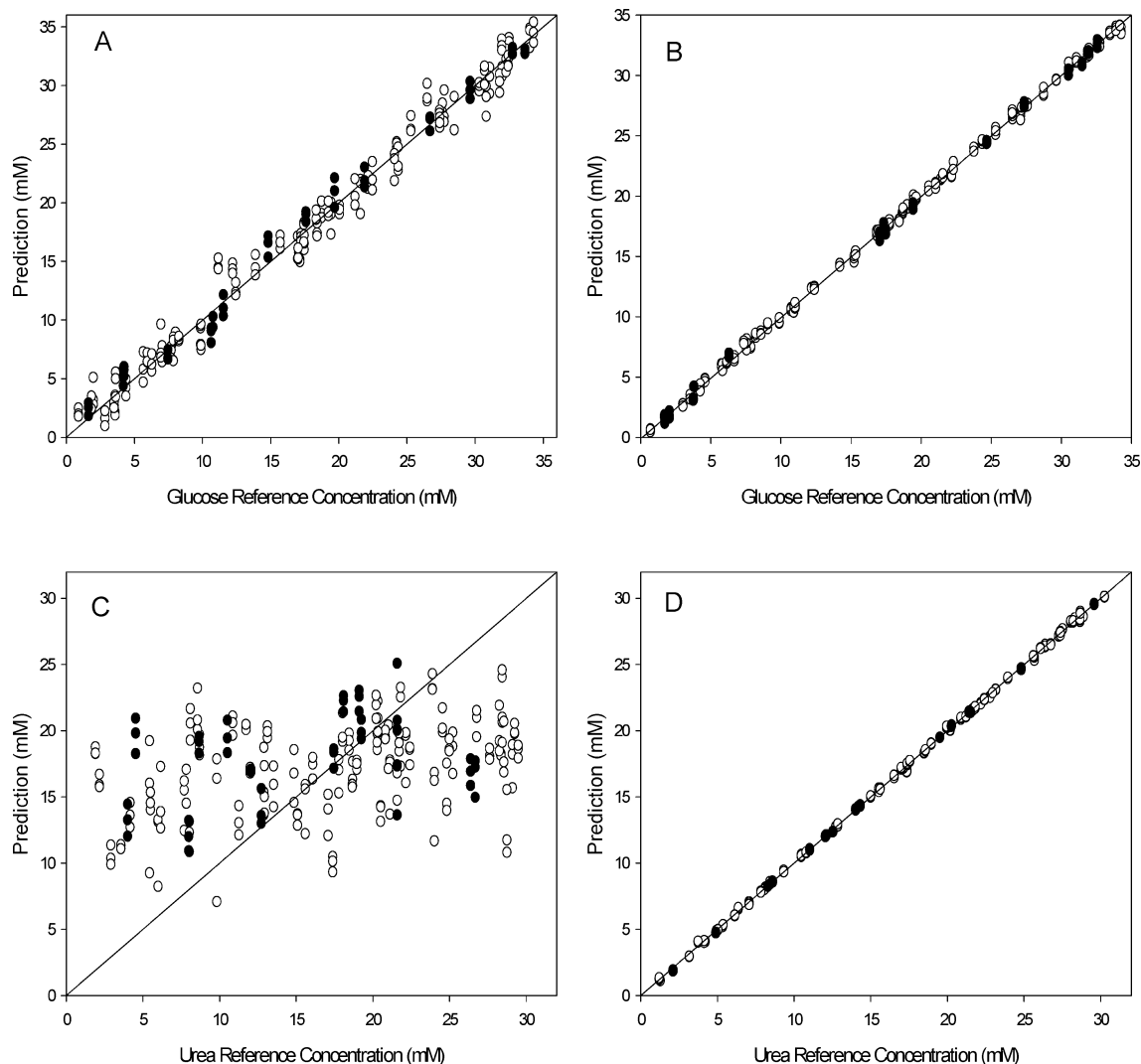


Figure 5. Concentration correlation plots for PLS calibration models for glucose from first overtone (A) and combination (B) spectra and for urea from first overtone (C) and combination (D) spectra. Open and closed circles correspond to calibration and prediction points, respectively, and solid lines correspond to the ideal, unity correlation between prediction and reference concentrations.

In both cases, near-infrared spectra were collected over a wide spectral range with a single sample thickness. In the Kasemsumran paper, spectra were collected over a spectral range of 12 000–4000 cm⁻¹ with a sample thickness, and corresponding optical path length, of 1 mm. Similarly, the Heise work involved collecting spectra over the 6800–4200 cm⁻¹ spectral range with a 1-mm optical path length. In both cases, calibration models from different spectral regions (first overtone and combination) were compared and performance assessed on the basis of SEP values. Kasemsumran found that the combination spectral range (4912–4303

cm⁻¹) provides the best results for glucose, while Heise reported that a region consisting of two ranges (6788–5461 and 4736–4212 cm⁻¹) provides the best results for glucose.

A major limitation of the above-mentioned work is the fact that a single optical path length was used for both spectral regions. It is not possible to get the best analytical performance from both regions by using a single path length. A sample thickness of only 1 mm is simply too thin for acceptable glucose measurements from first overtone spectra collected from aqueous samples. The basis of this conclusion is the report by Hazen et al.,¹⁷ where PLS

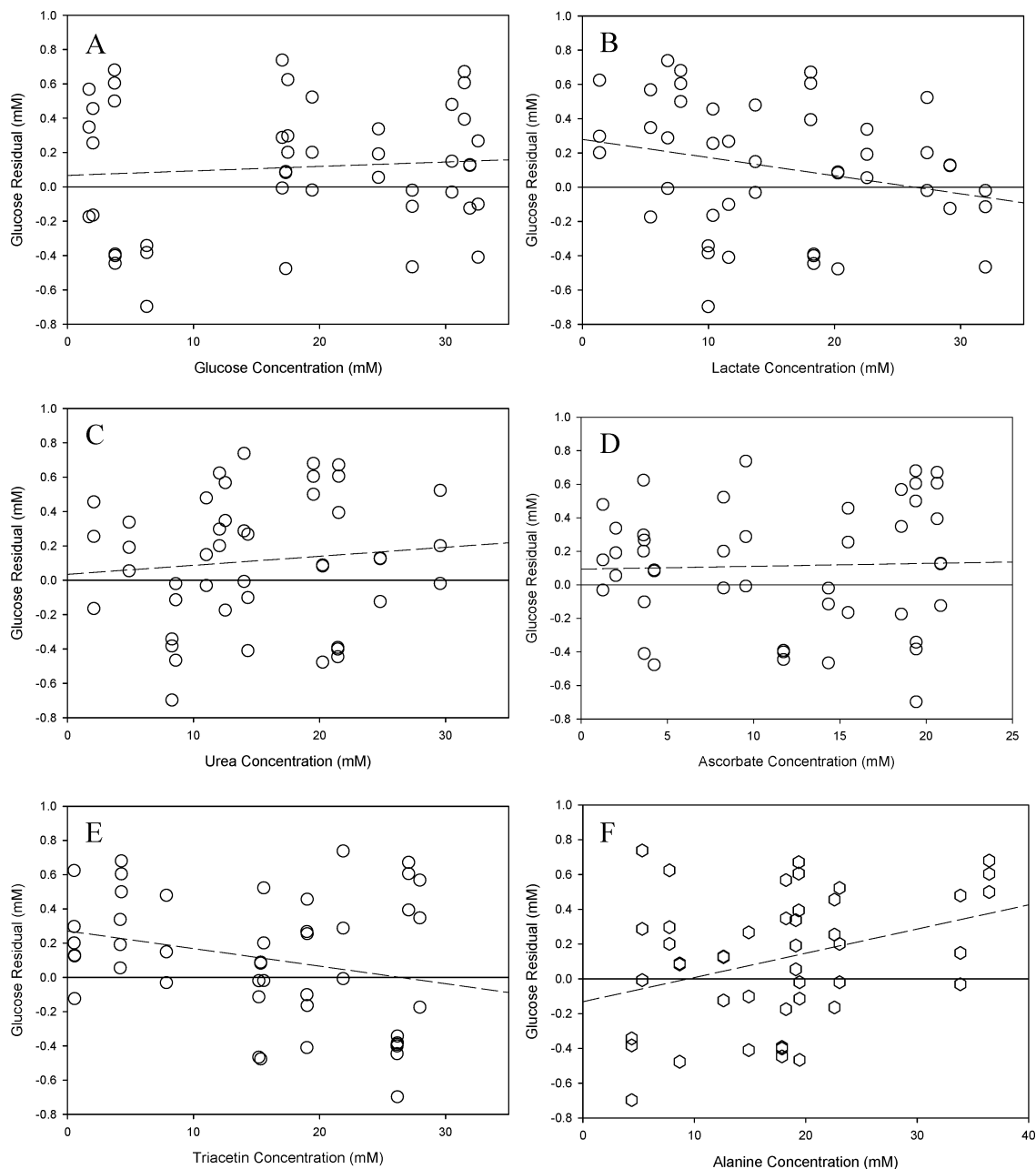


Figure 6. Residual plots for the glucose combination model, showing glucose prediction residuals as a function of the concentrations of glucose (A), lactate (B), urea (C), ascorbate (D), triacetin (E), and alanine (F). Solid line highlights the ideal zero-response and dashed line shows the linear regression model.

calibration models for glucose provided SEP values of 1.02, 0.55, and 0.42 mM under optimal conditions with optical path lengths of 2, 5.2, and 10 mm, respectively. Conversely, essentially no light would be transmitted through 7-mm thick aqueous samples over the combination spectral region.

The Kasemsumran results noted above³⁴ are consistent with the fact that a 1-mm sample thickness is insufficient for quality glucose measurements from first overtone spectra. Results in Figure 2c of the Kasemsumran paper³⁴ clearly demonstrate the lack of glucose-specific information in the first overtone spectra with this path length. For this reason, their moving window search algorithm isolated the combination spectral region for glucose measurements.

As noted above, Heise et al. reported a dual spectral range (6788–5461 and 4736–4212 cm^{-1}) as optimal for glucose with a 1-mm

path length.⁵ As part of their investigation, separate calibration models were constructed with each range individually. In this case, the first overtone range (6788–5461 cm^{-1}) outperformed the combination range (4736–4212 cm^{-1}), and the combination of these spectral ranges outperformed either of the individual ranges. These findings are inconsistent with those reported here and suggest that the first overtone range is superior. It is noteworthy that, in the same paper,⁵ urea calibration models were improved by including first overtone spectral data with combination spectral data (6788–5461 cm^{-1} and 4736–4212 cm^{-1}). As noted above, urea absorptivities in the first overtone spectral range are very weak, as the principal urea absorption features are located under the nearby large water absorption band.²⁷ Improvement in urea measurements with the inclusion of first overtone spectral information is inconsistent with the known absorptivities of urea.

Overall, improvements realized from the use of spectra collected from too thin of samples and spectra with no analyte information suggest overmodeling of the calibration data.

CONCLUSIONS

The results presented demonstrate the superior nature of combination spectra relative to first overtone near-infrared spectra for the quantification of solutes in aqueous solutions. PLS calibration models for glucose are particularly better when generated from combination spectra compared to first overtone spectra. SEP values are approximately 3-fold lower for the glucose model generated from combination versus first overtone spectra. Calibrations for urea are not possible from first overtone spectra owing to the absence of analyte-specific absorption features. PLS calibration models for urea from combination spectra display excellent measurement accuracy, with SEP values of 0.1 mM. Each of the six cosolutes (glucose, urea, lactate, ascorbate, triacetin, and alanine) can be quantified from a PLS analysis of combination spectra.

Visual comparison of first overtone and combination spectra indicate greater molecular distinction for combination spectra compared to first overtone spectra. The NAS vector for each analyte allows quantification of spectral overlap and distinction. Such an analysis clearly reveals the superior selectivity of combination spectra over first overtone spectra. An example is glucose where the length of the NAS vector is 3.8 times larger for combination spectra relative to first overtone spectra.

Our general conclusion is that combination spectra are preferred over first overtone for direct measurements of organic-based solutes in aqueous solutions. The basis of this conclusion is the enhanced selectivity afforded by combination spectra. NAS vector analysis confirms the difference in chemical selectivity of these spectral regions.

Received for review February 4, 2004. Accepted June 17, 2004.

AC0498056