

Temperature-Independent Near-Infrared Analysis of Lysozyme Aqueous Solutions

Shih-Yao B. Hu and Mark A. Arnold*

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

John M. Wiencek

Department of Chemical and Biochemical Engineering and Optical Science and Technology Center, University of Iowa, Iowa City, Iowa 52242

Digital Fourier filtering is used to produce a temperature-insensitive univariate calibration model for measuring lysozyme in aqueous solutions. Absorbance spectra over the 5000–4000 cm⁻¹ spectral range are collected for lysozyme standards maintained at 14 °C. These spectra are used to compute the calibration model while a set of spectra collected at temperatures ranging from 4 to 24 °C are used to validate the accuracy of this model. The root-mean-square error of prediction (RMSEP) is 0.279 mg/mL over a tested lysozyme concentration range of 0.036–51.6 mg/mL. The detection limit is 0.68 mg/mL. In addition, multivariate calibration models based on partial least-squares regression (PLS) are evaluated and compared to the results from the univariate model. PLS outperforms the univariate model by providing a RMSEP of 0.090 mg/mL. Analysis of variance showed that both calibration methods effectively eliminate the adverse affects created by variations in solution temperature.

Optimized protein crystal growth rates are necessary to provide large, high-quality crystals for X-ray diffraction analysis. Temperature can be used to control crystal growth rates by regulating protein saturation levels and, thereby, controlling the extent of supersaturation. The effectiveness of using temperature to control crystal growth rates has been established for selected proteins for which an appropriate temperature–time profile can be developed.^{1,2} Such a profile requires detailed knowledge of the temperature-dependent solubility properties of the protein as well as an understanding of the kinetic parameters that govern nucleation and crystal growth for this protein. Such details are not generally known a priori, thereby limiting the utility of this approach.

Alternatively, in situ crystal growth rates can be used to drive a temperature-controlled protein crystallization process. In this method, in situ crystal growth rates are obtained by measuring changes in the soluble protein concentration in real time during the crystallization process. Consumption of soluble protein during crystal formation results in lower soluble protein concentrations. Accurate measurement of the soluble protein concentration

permits computation of the crystal growth rate from simple mass balance, and this information can be used as a feedback-control parameter for adjusting system temperature. This approach demands a suitable analytical method to provide accurate in situ protein concentrations in a nondestructive, continuous, and temperature-insensitive manner.

Near-infrared (NIR) spectroscopy is proposed as a method for monitoring soluble protein levels in situ during crystallization. Conceptually, the measurement is made by passing a selected beam of NIR radiation through the sample of interest and extracting the desired analytical information from the resulting spectral information. Unlike other analytical methods for protein measurements in aqueous solutions,^{3,4} NIR spectroscopic measurements are reagentless and nondestructive. Both of these properties are critical for in situ control of crystallization where structural alterations at the molecular level cannot be tolerated. The feasibility of using NIR spectroscopy for measuring protein levels in aqueous solutions is well-established, particularly for applications in agricultural science⁵ and clinical chemistry.^{6–11} Recently, we measured the limit of detection for lysozyme measurements by NIR spectroscopy under fixed temperature conditions.¹²

The well-known temperature sensitivity of NIR spectroscopy is a major issue that bears directly upon the proposed application. This sensitivity originates mostly from the temperature sensitivity

- (3) Hohnadel D. C.; Koller, A. In *Methods in Clinical Chemistry*; Pesce, A. J., Kaplan, L. A., Eds.; C. V. Mosby Co.: St. Louis, MO, 1987; Chapter 6.
- (4) Sophianopoulos, A. J.; Rhodes, C. K.; Holcomb, D. W.; VanHolde, K. E. *J. Biol. Chem.* **1962**, *237*, 1107–12.
- (5) Burns, D. A.; Ciurczak, E. W. *Handbook of Near-Infrared Analysis*; Marcel Dekker: New York, 1992.
- (6) Flewelling, R. *The Biomedical Engineering Handbook*; Bronzino, J. D., Ed.; CRC Press: Boca Raton, FL, 199; pp 1346–56.
- (7) Harthum, S.; Matischak, K.; Friedl, P. *Anal. Biochem.* **1997**, *251*, 73–8.
- (8) Marquardt, L. A.; Arnold, M. A.; Small, G. W. *Anal. Chem.* **1993**, *65*, 3279–89.
- (9) Pan, S.; Chung, H.; Arnold, M. A.; Small, G. W. *Anal. Chem.* **1996**, *68*, 1124–35.
- (10) Hall, J. W.; Pollard, A. *Clin. Biochem.* **1993**, *26*, 483–90.
- (11) Hazen, K. H. Analysis of Albumin and Globulin Protein, Triglyceride, Cholesterol, Urea, Glucose and Lactate in Human Serum Using Near Infrared Spectroscopy. Ph.D. Thesis, University of Iowa, 1995; Chapter 5.
- (12) Hu, S. B.; Lillquist, A.; Arnold, M. A.; Wiencek, J. M. *Appl. Biochem. Biotechnol.*, in press.

(1) Schall, C. A.; Riley, J. S.; Li, E.; Edward, A.; Wiencek, J. M. *J. Cryst. Growth* **1996**, *165*, 299–307.

(2) Rosenberger, F.; Meehan, E. J. *J. Cryst. Growth* **1988**, *90*, 74–8.

of the underlying water absorption bands.¹³ The water absorption bands correspond to combinations and overtones associated with fundamental vibrational transitions. As such, these absorption bands are sensitive to hydrogen bonding, which is affected by sample temperature. These bands shift to higher frequencies with an increase in temperature. The effect is most prominent with aqueous samples, as in clinical and biomedical samples, thereby requiring precise temperature control to minimize measurement errors.^{14–16} Changes in solution temperature are inevitable in the present application. To optimize protein crystallization, solution temperatures must systematically vary between 4 and 24 °C. Successful implementation of NIR spectroscopy as an on-line monitor for controlling protein crystallization demands that the predictive results from the NIR calibration model are insensitive to solution temperature over this range.

A spectral processing method has been described to eliminate the adverse effects of temperature variation on the NIR spectroscopic measurement of glucose in aqueous solutions.^{17,18} Spectral variations created by differences in solution temperature can be diminished considerably by subjecting the raw absorbance spectra to a digital Fourier filtering step prior to the calibration model. In the digital Fourier filtering process, a fast Fourier transform is performed on an absorbance spectrum to reveal the underlying frequencies of signals that make up this spectrum. The transformed spectral data are then multiplied by a digital filter. Generally, this filter is a Gaussian function centered on the analytically relevant information. The width of the Gaussian function determines the bandwidth of the filter. Appropriate selection of the position and width of the Gaussian function is critical for successful implementation. Typically, the Gaussian function is selected to discriminate, or filter out, both high- and low-frequency components of the transformed spectral data. High frequencies typically correspond to choppy spectral noise while low frequencies correspond to broad baseline variations, such as those created by differences in solution temperature. Eventually, the filtered spectral data are subjected to an inverse Fourier transform and the data are restored to the original domain of optical frequencies.

A likely alternative to digital Fourier filtering is a factor-based calibration method such as partial least-squares (PLS) regression. In PLS, the spectral and concentration information is represented as vectors and correlations are established between these information vectors. Non-concentration-specific information, such as temperature-induced spectral variations, is removed. An effective PLS calibration model requires a comprehensive set of calibration standards that encompass all possible variations in concentration and temperature. For a simple system without interfering substances such as ours, PLS may not be the most economical way for analysis unless a significant improvement in performance is achieved.

The goal of the work presented here is to develop a temperature-independent algorithm for measuring soluble levels of

lysozyme dissolved in aqueous solutions with NIR spectroscopy. The simplicity of the sample matrix (only lysozyme in buffer) permits the use of a univariate calibration model where the protein concentration is related to the magnitude of an integrated peak area. In addition, digital Fourier filtering is used to eliminate temperature effects and produce a robust calibration method. The proposed NIR method is validated by generating a calibration model from spectra collected with standards set at 14 °C and using this model to predict lysozyme levels from spectra collected from solutions with temperatures ranging from 4 to 24 °C. Performance of this univariate model is compared to a multivariate PLS calibration model designed to account for sample temperature variations.

EXPERIMENTAL SECTION

Spectra were collected with a Nicolet 550 Fourier transform spectrometer (Nicolet Instrument, Madison, WI). The light source was a 75 W tungsten–halogen lamp. A calcium fluoride beam splitter and a cryogenically cooled indium antimonide (InSb) detector were used. The incident light was restricted to the 5000–4000 cm^{-1} (2.00–2.50 μm) range by using a multilayer optical interference filter (Barr Associates, Westford, MA). The aperture was set to 55 and the gain was set to 8 for all measurements. Samples were placed in an aluminum-jacketed sample cell (Wilma Glass Co., Inc., Buena, NJ) with sapphire windows (Meller Optics, Inc., Providence, RI). The path length of the cell was 1.5 mm. Temperatures were maintained at the desired value (± 0.1 °C) by circulating thermostated fluid through the aluminum jacket. A programmable temperature-controlled circulator bath model 9110 (Fisher Scientific, Pittsburgh, PA) was used. An external temperature probe was placed in the jacket, and this temperature was taken as the solution temperature.

Hen egg white lysozyme solutions (Sigma Chemicals, St. Louis, MO) were prepared by dialysis in a molecular filtration membrane (MWCO 3500, Spectrum Medical Industries, Inc., Houston, TX) against deionized water and then against 0.1 M sodium acetate buffer, pH 4.6, with two buffer changes. The final concentrations of the lysozyme stock solutions range between 40 and 60 mg/mL. All solutions were passed through disposable 0.2 μm sterile syringe filters before use. Lysozyme stock solution was then mixed with the appropriate quantities of acetate buffer to obtain the desired final concentration. Reference lysozyme concentrations were determined by using a published value for the molar absorptivity ($2.635 \text{ (mg/mL)}^{-1} \cdot \text{cm}^{-1}$ at 280 nm) and a Shimadzu UV–visible scanning spectrophotometer M model UV-2101PC (Shimadzu Corp., Kyoto, Japan).⁴ The acetate buffers were prepared from acetic acid (glacial, ACS Plus grade) and sodium acetate trihydrate (HPLC grade), both purchased from Fisher Scientific.

Spectra were collected as double-sided, 16Kbyte interferograms with 128 coadded scans. Interferograms were triangularly apodized and Fourier transformed to produce single-beam spectra with 1.9 cm^{-1} point spacing. A total of 17 independent samples were arranged into the calibration data set and the prediction data set according to Table 1. In Table 1, all concentrations are in units of milligrams per milliliter. A temperature of 1 indicates the sample was analyzed at a single temperature of 14 °C. A temperature of 7 means the sample was analyzed at seven different temperatures between 4 and 24 °C (4, 7.3, 10.6, 14, 17.3, 21.6, and 24 °C).

(13) Libnau, F. O.; Kvalheim, O. M.; Christy, A. A.; Toft, J. *Vib. Spectrosc.* **1994**, *7*, 243–54.

(14) Arnold, M. A.; Small, G. W. *Anal. Chem.* **1990**, *62*, 1457–64.

(15) Bauer, B.; Floyd, T. A. *Anal. Chim. Acta* **1987**, *197*, 295–301.

(16) Hall, J.; Pollard, A. *Clin. Chem.* **1992**, *38*, 1623–31.

(17) Horlick, G. *Anal. Chem.* **1972**, *44*, 943–7.

(18) Hazen, K. H.; Arnold, M. A.; Small, G. W. *Appl. Spectrosc.* **1994**, *48*, 477–83.

Table 1. Sample Arrangements in the Univariate Study and the PLS Study

sample		univariate		PLS	
conc	temp	calib	pred	calib	pred
51.61	7		✓	✓	
32.25	1	✓		✓	
20.67	7		✓		✓
16.15	7		✓	✓	
10.21	1	✓		✓	
8.39	7		✓		✓
5.52	7		✓	✓	
3.19	1	✓		✓	
1.08	7		✓		✓
0.541	7		✓	✓	
0.311	1	✓		✓	
0.131	7		✓		✓
0.121	7		✓	✓	
0.087	1	✓		✓	
0.077	7		✓		✓
0.049	7		✓	✓	
0.036	7		✓	✓	
no. of samples		5	84	47	42

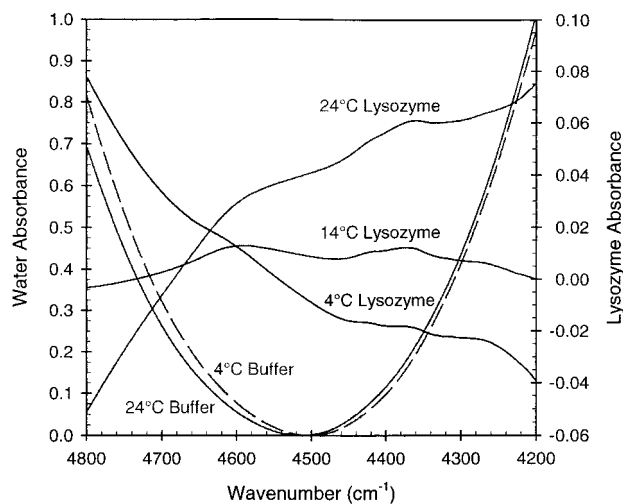


Figure 1. Effect of temperature on NIR absorbance spectra of lysozyme aqueous solutions and acetate buffer.

Triplicate spectra (three consecutive, 128 coadded scans) were collected for a given sample at each temperature during which the sample was continuously maintained in the spectrometer. Triplicate spectra were treated as multiple observations for each sample during analysis. All spectra for a given sample were moved together into either the calibration data set or the prediction data set, but not both. The order of temperature and sample concentration analyzed was randomized to eliminate possible effects from instrumental drift. Background spectra were collected with buffer solutions at 14 °C before each sample. Single-beam spectra were converted into absorbance before processing.

RESULTS AND DISCUSSION

Effect of Temperature on Lysozyme Absorbance Spectra.

Temperature is a critical parameter for NIR spectroscopic analysis of aqueous samples. As can be seen from Figure 1, the absorption band of acetate buffer shifts toward higher frequency as the sample temperature changes from 4 to 24 °C. This spectral change contributes to the baseline variation of lysozyme absorbance

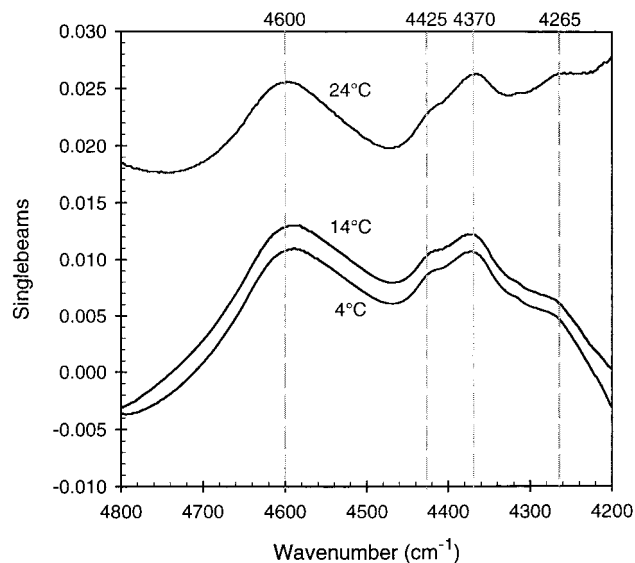


Figure 2. Lysozyme absorbance spectra at three different temperatures with both sample and background spectra collected at the same temperature.

spectra, also shown in Figure 1. Using the 14 °C buffer spectrum as a reference, the three curves represent absorbance of a 10 mg/mL lysozyme aqueous solution at 4, 14, and 24 °C. The absorbance of water above 4500 cm^{-1} decreases as temperature increases while it increases with temperature below 4500 cm^{-1} . As a result, the 24 °C spectrum displays a positive slope in the baseline while a negative slope is observed in the 4 °C spectrum.

Baseline variations are significantly reduced when the sample and reference spectra are collected at the same temperature, as shown in Figure 2. Four distinct absorption bands are evident in the lysozyme spectra collected at 4, 14, and 24 °C. These bands are centered at 4260, 4365, 4420, and 4600 cm^{-1} and are consistent with our earlier findings.¹² Although baseline shifts are evident in Figure 2, the sloping nature of the baseline variation is no longer significant. More importantly, there is no apparent shift in the position of the lysozyme absorption bands. When temperature changes, NIR spectra can change due to chemical effects (variation in association and H-bonding) as well as physical effects (density variation of the solution). While all the absorption bands of all the components are affected, some are more sensitive than others.¹⁹ The above findings suggest that the variations observed in Figure 1 are mainly caused by the temperature-dependent absorption properties of water. The absorption characteristics of lysozyme appear only slightly affected by solution temperature.

Temperature-dependent baseline variations must be removed to achieve robust calibration models that are capable of accurate analyte predictions across solution temperatures. Digital Fourier filtering is a demonstrated spectral preprocessing tool that is capable of eliminating solution temperature effects from NIR spectra.¹⁴ In this earlier work, raw NIR absorbance spectra were collected for a series of glucose-containing samples at temperatures between 32 and 41 °C while the reference spectrum was collected from a blank solution held at 37 °C. By preprocessing these raw absorbance spectra with an optimized digital Fourier filter, accurate glucose measurements were achieved from calibra-

(19) Wülfert, F.; Kok, W. T.; Smilde, A. K. *Anal. Chem.* **1998**, *70*, 1761–7.

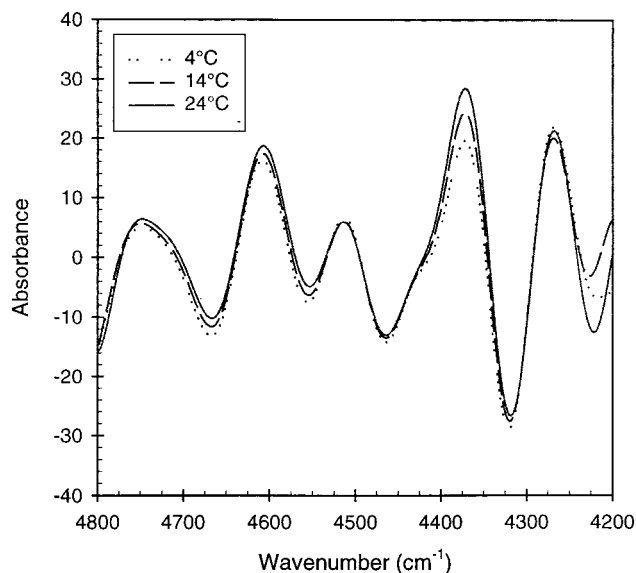


Figure 3. The same lysozyme absorbance spectra as in Figure 1 after digital Fourier filtering.

tion models constructed with only 37 °C spectra. Similar calibration robustness is required for a practical temperature-controlled protein crystallization system. The question is, can digital Fourier filtering effectively eliminate temperature variations in lysozyme absorbance spectra over the 20-deg temperature range from 4 to 24 °C?

Digital Fourier Filtering. Digital Fourier filtering is a widely used data-processing technique. The key to a successful Fourier filter is the selection of the filter function. An inferior filtering function removes useful information from the spectrum while retaining unnecessary data and noise. In general, a smooth function is preferred because it introduces fewer artifacts into the resulting spectra.¹⁷ In particular, the effectiveness of Gaussian-shaped filters has been shown in various studies.^{14,17} A Gaussian-shaped filter function is defined by two values: a mean position value and a standard deviation. Figure 3 illustrates the result of subjecting the lysozyme absorbance spectra in Figure 1 to a digital Fourier filter defined by a mean position of $0.02f$ and a standard deviation of $0.005f$. These filter parameters represent common values used before and are given in digital frequency units.²⁰ After Fourier filtering, the effects of temperature are mostly eliminated while the lysozyme information is emphasized. Five spectral features are evident for lysozyme with maximums around 4750, 4610, 4510, 4370, and 4270 cm^{-1} .

Area Calculation. Fourier-filtered NIR absorbance spectra possess features that resemble derivative spectra with maximums and minimums corresponding to the underlying sinusoidal functions passed by the Gaussian-shaped filter response function. A univariate calibration model can be envisioned where the integrated area under one or more of these spectral features is related to analyte concentration. Output from the filtering process varies dramatically depending on the filter parameters. Figure 4 shows three Fourier filtering results of a single spectrum. With a typical parameter combination ($0.02f$ mean, $0.005f$ standard deviation), the absorbing features in the original spectrum are emphasized

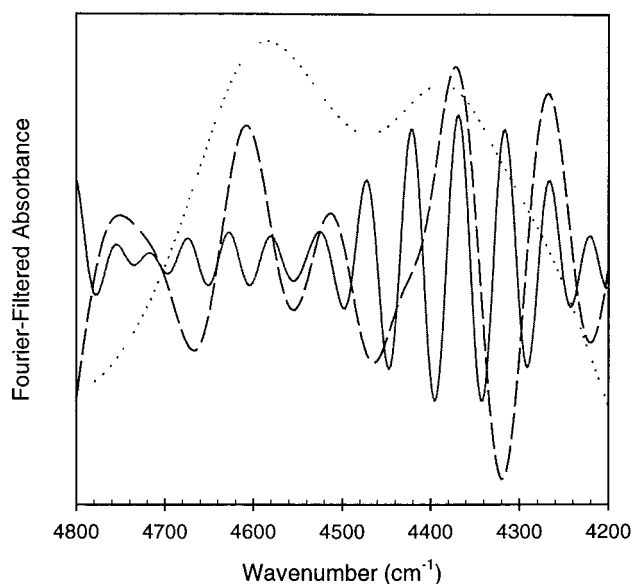


Figure 4. Effect of Gaussian filter parameters. The solid line results from a filter combination of $0.04f$ mean position and $0.005f$ standard deviation; the dashed line, $0.02f$ and $0.005f$; and the dotted line, $0.00f$ and $0.01f$. The original spectrum is from a 32.249 mg/mL sample collected at 14 °C. The magnitude of the curves have been adjusted for better visualization; the scales are -0.6 to 0.6 , -0.1 to 0.1 , and -0.03 to 0.03 for the dotted, dashed, and solid lines, respectively.

with few artifacts. When a broad filter centered at zero is used ($0.00f$ mean, $0.01f$ standard deviation), a wide range of digital frequencies pass and the filtered spectrum closely resembles the original spectrum. No enhancement is realized under such conditions. Filters centered at high digital frequencies (i.e., mean $0.04f$ and standard deviation $0.005f$) pass only choppy noise while the analytical information is lost. Nevertheless, such a situation must be accounted for when the filter parameters are optimized (to be discussed in the next section). Scales used for the three filtered spectra in Figure 4 are adjusted to aid visualization. The actual scale of the $0.04/0.005$ line is much less than the scales for the other two spectra.

Because the features of filtered spectra vary dramatically, precise area measurements require a dynamic band-locating step, followed by a baseline correction step, and finally the area integration.¹⁴ On the basis of a series of experiments, it was found that the absorption feature at 4600 cm^{-1} remains consistent over the widest range of filter parameters. A possible reason for such observation is that the profile of this particular band is least sensitive to temperature change; Fourier filtering does not introduce additional artifacts during the removal of baseline variation caused by changes in water absorbance. Hence, this particular band was selected for our univariate calibration model. The actual integration process starts by locating the maximum absorbance closest to 4600 cm^{-1} in a filtered spectrum. Next, two inflection points at each side of the maximum are located and used to define a linear baseline. This linear baseline is subtracted across the feature, thereby further reducing baseline variations. These maximums and minimums do not reside at exact frequencies and must be located dynamically for each filtered spectrum. A search algorithm was implemented in our data-processing scheme to locate these extremes. After the minimums are located

(20) Small, G. W.; Marquardt, L. A.; Arnold, M. A. *Anal. Chem.* **1993**, *65*, 3279-89.

and baseline corrected, the integrated area under the resulting spectral band is computed with Simpson's 1/3 and 3/8 rules.⁸

Calibration Model. Since lysozyme is the only analyte in the solution matrix, an univariate calibration method should be sufficient for the analysis. Based on the Beer–Lambert law, a univariate calibration model has a general form

$$\mathbf{A} = b_0 + b_1\mathbf{C} \quad (1)$$

where \mathbf{A} is a vector of integrated absorbance area, \mathbf{C} is a vector of the corresponding lysozyme concentrations, and b_0 and b_1 are regression coefficients. After Fourier filtering, the area of one or more of the absorption bands can be integrated and related to the sample concentration. Across the range of concentrations, the use of optimal filter parameters will lead to a calibration model with maximum linearity. Common techniques for finding optimal filter parameters include numerical optimization and grid search.^{14,18} In numerical optimization, one first defines an objective function that indicates the goodness of fit. The filter parameters are then optimized numerically to maximize the objective function. With the grid-search method, one first defines the range for each filter parameter to be searched. An objective function such as mean square prediction error is then evaluated at all possible combinations of the parameters in a stepwise fashion point by point. It has been found that, when Gaussian-shaped filters are used, the optimal filter has a mean position around $0.02f$ and standard deviation around $0.005f$ in general. The exact optimal values of these parameters, however, are likely to be sample-dependent.¹⁸ We thus choose to use these two values ($0.02f$ mean and $0.005f$ standard deviation) directly as a more universal approach, and no effort was made to further optimize the parameters.

The sample solutions were grouped into two sets, one for calibration and one for prediction. For the calibration samples, triplicate spectra were collected for each sample at 14 °C. For prediction samples, triplicate spectra were collected at seven different temperatures between 4 and 24 °C for each sample. Background spectra were collected from buffer solutions at 14 °C before each sample was analyzed. The above-stated filter parameters were used to process the spectra in the calibration sample set. The band area of the filtered spectra was then integrated and b_0 and b_1 were calculated through a simple linear regression. A relation of $\mathbf{A} = (0.034 \pm 9.5 \times 10^{-3}) + (0.2887 \pm 1.6 \times 10^{-4}) \times \mathbf{C}$ is obtained by this regression and the R^2 of the regression is 0.999 94. Note that only the calibration samples were used in generating this calibration line.

Spectra of the prediction samples were processed with the same filter parameters, and the results are plotted in Figure 5 (open circles) along with results from the calibration samples (solid circles). The line represents the regressed correlation between area and calibration sample concentrations. As can be seen, both data sets follow the regression line closely. The standard error of calibration (SEC) is 0.102 mg/mL while the root-mean-square error of prediction (RMSEP) is 0.279 mg/mL. Although the highest prediction sample concentration (51.61 mg/mL) is far higher than the highest calibration sample concentration (32.25 mg/mL), the prediction holds up satisfactorily. This finding suggests that the method is able to extrapolate to higher values. The ability to extrapolate is a function of the univariate nature of

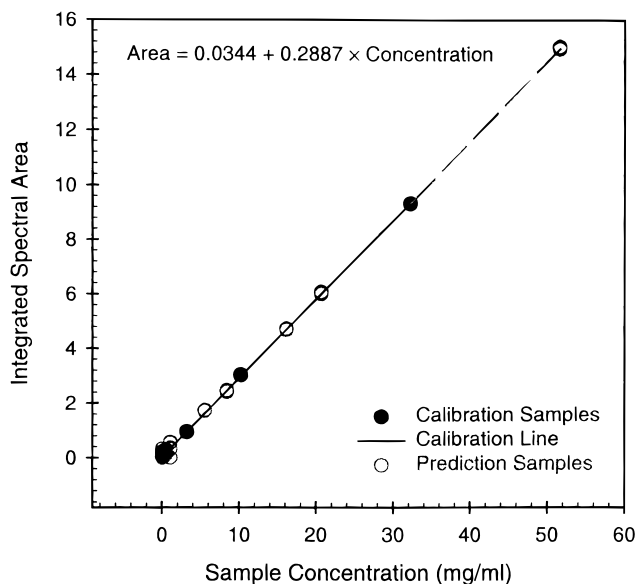


Figure 5. Correlation between area and sample concentrations. The dashed line represents an extrapolation of the calibration line.

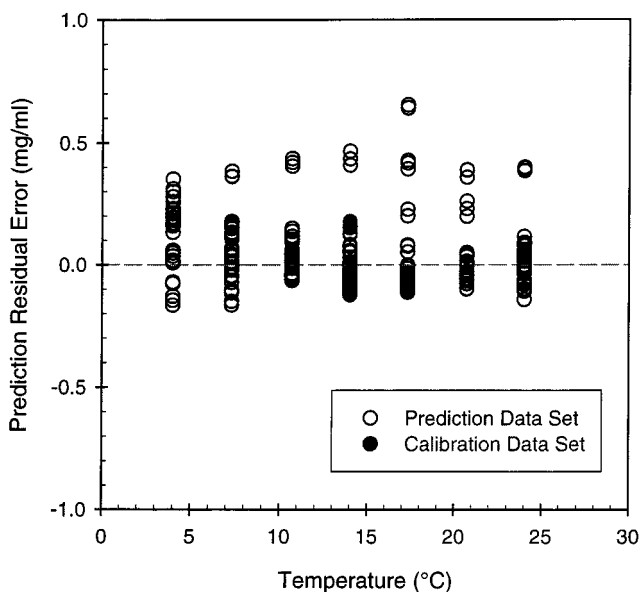


Figure 6. Effect of temperature on the calibration model. Residuals from the calibration model are plotted against sample temperatures. The solid circles indicate calibration samples while the open circles represent prediction samples.

the sample and the effectiveness of the Fourier filtering step to remove nonlinear baseline effects. These results also prove that the calibration generated at a single temperature can be used successfully to analyze samples collected over a wide range of temperatures.

Effect of Temperature. To ensure the data-processing procedures indeed eliminate temperature effects, measurement residuals are plotted as a function of temperature in Figure 6. No correlation between temperature and the residual error is evident. Data close to 14 °C (temperature of the reference spectrum) show just as much deviation as data collected at more extreme temperatures. At the 95% confidence interval ($\alpha = 0.05$) and the given sample size, an F statistic of 2.13 is required in an analysis of variance (ANOVA) to prove that the effect of temperature is

significant. The F statistic of our model is 0.77, far less than the critical value. This suggests that spectral variations caused by temperature changes are efficiently eliminated by digital Fourier filtering coupled with baseline correction.

Detection Limit. One of the objectives of the study is to determine the detection limit of the analysis, which is usually discussed in terms of orders of magnitude. Thus, the samples were prepared such that the concentrations spaced more evenly on a log scale rather than a linear scale. Quantitatively, the most common estimation of detection limits involves the analysis of several blank samples. The minimum distinguishable analytical signal A_m (area in this case) is taken as the mean blank signal \bar{b}_0 plus a multiple k of the standard deviation of the blank s_{b_0} :

$$A_m = \bar{b}_0 + ks_{b_0} \quad (2)$$

Combining this equation with eq 1, one can derive the detection limit C_m :

$$C_m = (A_m - \bar{b}_0)/b_1 \quad (3)$$

A value of 3 is commonly accepted for k .²¹ Thirty blank samples, none of them used in generating the calibration, are analyzed by this approach. A \bar{b}_0 of 0.0768 and a ns_{b_0} of 0.0652 were obtained, which led to a computed detection limit of 0.68 mg/mL. The blank sample spectra used in this calculation were collected during an extended four-month period following the collection of the calibration spectra. The detection limit might be improved if calibration is carried out just prior to analysis.

Partial Least-Squares Regression. PLS is commonly used for calibration models with NIR spectra. Being a multivariate method, PLS is predominately used for calibrations with multiple analytes or complex sample matrices. However, a factor-based calibration method such as PLS should also be able to handle spectral baseline variations resulting from temperature changes. The key to a successful PLS analysis is to provide comprehensive calibration data that encompasses all possible spectral variations. The original calibration data set consists only of spectra collected at 14 °C. For PLS analysis, additional calibration samples are necessary to account for temperature and concentration variations across samples. It is important to have at least five or six samples for each factor so the PLS model is valid statistically. Sample arrangements are listed in Table 1. Calibration models based on 1–10 PLS factors were examined, and RMSEP is plotted against PLS model rank in Figure 7 (diamond symbols). The curve follows a typical rank–RMSEP response. Initially, RMSEP decreases as the number of factors increases and more analyte-specific spectral variation is incorporated into the calibration model. With six factors, all meaningful spectral variations have been modeled and additional factors do not significantly improve the quality of calibration. The spectral range for each model represented in Figure 7 was optimized by a previously described method.¹² In the case of the six-factor model, the ideal range is 4705–4475 cm^{-1} and the RMSEP is 0.090 mg/mL.

Comparison of Univariate and PLS Models. The above results showed that PLS performs better than the univariate

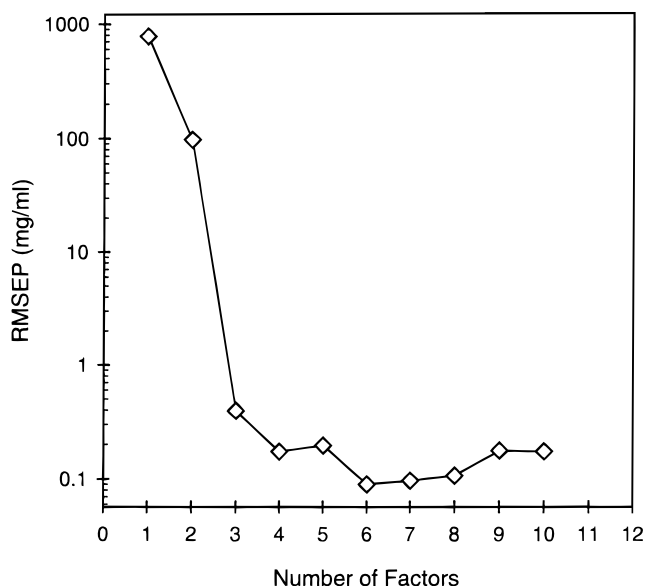


Figure 7. Effect of model rank on prediction errors for PLS calibration model.

method. In the univariate model, an RMSEP of 0.279 mg/mL was obtained, versus the 0.090 mg/mL obtained with PLS. To assess the capability of a model in removing the effect of temperature, ANOVA analyses using temperature as the factor were conducted. At the 95% confidence interval ($\alpha = 0.05$), an F statistic greater than 2.13 is required to prove that temperature does have an effect on a calibration model. The univariate model results in an F of 0.77 while the PLS model results in an F of 1.31. This means that the filtered univariate method and the PLS method are both capable of eliminating temperature effects. One should choose a calibration method based on the demands of a particular analysis. If accuracy is the ultimate goal, PLS should be used. On the other hand, a filtered univariate model can be generated with far fewer samples and may excel in extrapolation in terms of both temperature and concentration.

CONCLUSION

This work demonstrates the effectiveness of FT-NIR spectroscopy for measuring lysozyme levels in aqueous solutions with varying temperatures. Water absorption in NIR spectra is highly sensitive to temperature. Such effects translate into baseline variations, particularly when the sample and the reference spectra are collected at different temperatures. One must remove such baseline variations for quantitative analysis to succeed. Digital Fourier filtering is a technique that is able to remove low-frequency spectral variations such as those introduced by temperature variations. After filtering, the exact position of the absorption band of interest must be located before a simple baseline correction can then be implemented to further characterize the area of the absorption bands. This data-processing scheme employs four adjustable parameters: the mean position and standard deviation of the Gaussian-shaped digital filter, and the slope and intercept of a linear correlation between band area and lysozyme concentration. A good combination of the parameters must lead to a large slope (high sensitivity) with small uncertainty (low noise).

A univariate calibration model was generated by using a Gaussian-shaped digital filter of 0.02 f mean position and 0.005 f

(21) Skoog, D. A.; Leary, J. J. *Principles of Instrumental Analysis*, 4th ed.; Saunders College Publishers: Philadelphia, PA, 1992; Chapter 1.

standard deviation, in conjunction with a set of calibration samples collected at 14 °C. This calibration model produced a RMSEP of 0.279 mg/mL for 11 prediction samples collected at various temperatures between 4 and 24 °C. This low RMSEP suggests that, with digital Fourier filtering, one can generate a calibration model at one temperature and use it to analyze samples at other temperatures. Residual errors for prediction do not correlate with sample temperature. The detection limit is about 0.68 mg/mL based on the analysis of spectra from blank samples. PLS analysis outperforms the univariate model by providing a RMSEP of 0.90

mg/mL. Both calibration methods effectively eliminate the adverse affects created by variations in solution temperature.

ACKNOWLEDGMENT

This work was supported by grants from NASA (NASA 8-1352) and the Iowa Space Grant Consortium.

Received for review June 28, 1999. Accepted November 2, 1999.

AC9907080