

Temperature-Insensitive Near-Infrared Method for Determination of Protein Concentration during Protein Crystal Growth

J. T. Olesberg, M. A. Arnold,* and Shih-Yao B. Hu

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

John M. Wiencek

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

A temperature-insensitive method for measuring protein concentration in aqueous solutions using near-infrared spectroscopy is described. The method, which is based on identification of the net analyte signal of single-beam spectra, can be calibrated using a single protein absorbance measurement and is thus well suited for crystallization monitoring where the quantity of protein is limited. The method is applied to measurements of glucose-isomerase concentration in a sodium phosphate buffer that is actively varied over the temperature range of 4–24 °C. The standard error of prediction using the optimized spectral range of 4670–4595 cm⁻¹ is 0.12 mg/mL with no systematic trend in the residuals with solution temperature. The method is also applied to previously collected spectra of hen egg-white lysozyme and yields a standard error of prediction of 0.14 mg/mL. Spectra sampled at discrete wavelengths can also be used for calibration and prediction with performance comparable to that obtained with spectral bands. A set of four wavelengths are identified that can be used to predict concentrations of both proteins with a standard error less than 0.14 mg/mL.

Near-infrared spectroscopy and chemometric analysis have seen increasing use for a range of applications including clinical measurements and chemical process monitoring in agricultural, pharmaceutical, textile, and food industries. Two of the primary advantages of near-infrared spectroscopy are that the measurement is both noninvasive and nondestructive (reagentless). These properties allow on-line measurements to be made of systems that would be disrupted by more invasive techniques. One particular application for which near-infrared spectroscopy is attractive is the monitoring of dissolved protein concentration during the growth of protein crystals.

One avenue for the systematic growth of protein crystals involves the use of temperature to regulate the protein crystal-

lization process.^{1–10} It has been shown that active control of the crystal growth rate can produce larger, higher-quality crystals than passive growth.^{8–11} If the concentration of protein in solution can be continuously monitored in situ, the crystal growth rate can be inferred from a mass balance and used in an active control algorithm. The primary complication to the measurement is the active variation of solution temperature that is used to drive the crystallization process. Changes in the water absorption spectrum due to temperature are significantly larger than the magnitude of protein absorption at concentrations relevant to crystallization (tens of milligrams per milliliter). It is essential that the method used for extracting protein concentration from the transmission spectra be highly temperature-insensitive.

A number of methods have been used for temperature-insensitive near-infrared measurements. The most straightforward is a global calibration model where temperature is included as an active variable in the multivariate calibration set.¹² For example, in a partial least squares (PLS) analysis, spectra are collected from a large calibration set containing samples whose temperature and protein concentration are varied independently and span their expected range of variation. The multivariate calibration procedure identifies spectral variations that are due to concentration differences rather than temperature changes and uses these to predict protein concentration. A drawback to this method is that the calibration data set will be relatively large and must be designed to avoid unintentional correlations between temperature and

- (1) Rosenberger, F. *J. Cryst. Growth* **1986**, *76*, 618–636.
- (2) Rosenberger, F.; Meehan, E. J. *J. Cryst. Growth* **1988**, *90*, 74–78.
- (3) DeMattei, R. C.; Feigelson, R. S. *J. Cryst. Growth* **1992**, *122*, 21–30.
- (4) Lorber, B.; Giegé, R. *J. Cryst. Growth* **1992**, *122*, 168–175.
- (5) Ward, K. B.; Zuk, W. M.; Perozzo, M. A.; Walker, M. A.; Birnbaum, G. L.; Kung, W.; Cavaliere, A.; Uffen, D. R.; Scholaert, H. *J. Cryst. Growth* **1992**, *122*, 235–241.
- (6) DeMattei, R. C.; Feigelson, R. S. *J. Cryst. Growth* **1993**, *128*, 1225–1231.
- (7) Rosenberger, F.; Howard, S. B.; Sowers, J. W.; Nyce, T. A. *J. Cryst. Growth* **1993**, *129*, 1–12.
- (8) Schall, C. A.; Riley, J. S.; Li, E.; Arnold, E.; Wiencek, J. M. *J. Cryst. Growth* **1996**, *165*, 299–307.
- (9) Bray, T. L.; Kim, L. J.; Askew, R. P.; Harrington, M. D.; Rosenblum, W. M.; Wilson, W. W.; DeLucas, L. J. *J. Appl. Crystallogr.* **1998**, *31*, 515–522.
- (10) Jones, W. F.; Wiencek, J. M.; Darcy, P. A. *J. Cryst. Growth*, in press.
- (11) Wiencek, J. M. *Annu. Rev. Biomed. Eng.* **1999**, *1*, 505–534.
- (12) Wülfert, F.; Kok, W. T.; Smilde, A. K. *Anal. Chem.* **1998**, *70*, 1761–1767.

* Corresponding author. E-mail: mark-arnold@uiowa.edu.

concentration.¹³ In addition, protein that is diluted for use as calibration samples is normally not useful for the actual crystallization process.

Another strategy for temperature-invariant calibration involves the selection of a subset of spectral wavelengths for use in building the calibration model. The set of wavelengths are chosen so that variations in different portions of the spectra with temperature cancel out. This technique has been shown by Swierenga et al. to give comparable results to those of a global calibration model.¹⁴ However, it requires a relatively large calibration set and substantial computation to select an optimal set of wavelengths. Other techniques for temperature-insensitive calibration fall into the general category of filters or transformations that are applied to the measured spectra before the multivariate regression. As such, they overlap with techniques used for calibration transfer. Examples of prefiltering strategies include continuous piecewise direct standardization,¹⁵ derivative analysis,¹⁶ orthogonal signal correction,^{17–19} and Fourier filtering.^{20–22} Fourier filtering, in particular, has been applied successfully to the problem of protein crystallization monitoring.²²

Another approach to temperature-insensitive measurements can be devised on the basis of the identification of the net analyte signal.^{23,24} Net analyte signal methods are mathematically similar to the orthogonal signal correction methods, except that they are used to directly calculate a calibration vector rather than serve as a prefiltering step. Net analyte signal calibration for a multivariate problem was originally described by Lorber et al.;²⁴ variants of the method include hybrid linear analysis²⁵ and the optimum-factor-selection-free method of Xu and Schechter.^{26,27} In all of the net analyte signal methods, it is necessary to identify the variations in the measured spectra due to all factors other than the analyte. In hybrid linear analysis, this is done by carefully removing the known analyte absorbance from the calibration set. In the other methods, the removal is performed mathematically by identifying the components of the spectra orthogonal to the analyte concentration.

In the present paper, we describe an experimental approach to identifying the nonanalyte components of the measured spectra. It is applicable whenever one has access to calibration samples that contain all of the nonanalyte variation but do not contain the

analyte. This experimental approach is particularly well suited to the application of protein crystallization monitoring, because calibration can be performed reliably using only one sample containing protein. This means that it is possible to perform the calibration with the protein solution that will be used in the actual crystal growth, with no protein lost in the calibration process. The performance of this method is evaluated using measurements of glucose isomerase and hen egg-white lysozyme. Performance for each protein is determined using optimized spectral bands. Model performance using spectra consisting of 2–6 discrete wavelengths is also determined in order to evaluate the possibility of using a spectrometer system based on a laser diode array. The ability to predict concentration of both proteins using a single wavelength set is also evaluated as a first step in developing a monitoring system that will work with many different proteins.

METHOD DESCRIPTION

The method is based on the construction of a calibration vector that has significant overlap with the analyte absorbance spectrum, but which is orthogonal to the background (e.g., the crystal growth buffer solution) absorption spectra at all temperatures. Briefly, the principal components of the background absorption spectra over the temperature range of interest are determined. An absorption spectrum for a single sample containing the target protein is made orthogonal to the principal components of the background absorption spectra by removing its projection onto the principal component basis. Protein concentration can then be quantified by taking the inner product of the resulting calibration vector with the spectra from a sample with unknown protein concentration.

The measured spectra are a product of the light-source spectrum, buffer transmission, protein transmission, and detector response. If we assume that the light source and detector response are constant throughout the measurement, we can write the logarithmized single-beam spectra, $\mathbf{s}(T, C)$, as a function of sample temperature, T , and protein concentration, C , as

$$\mathbf{s}(T, C) = \mathbf{s}_{\text{source}} - \mathbf{a}_{\text{buffer}}(T) - \mathbf{a}_{\text{protein}}(C) + \mathbf{r}_{\text{detector}} \quad (1)$$

where $\mathbf{s}_{\text{source}}$ is the logarithm of the source spectral distribution, $\mathbf{a}_{\text{buffer}}(T)$ is the absorbance spectrum of the buffer, $\mathbf{a}_{\text{protein}}(C)$ is the absorbance of the protein, and $\mathbf{r}_{\text{detector}}$ is the logarithm of the response of the detector. Other factors affecting the single-beam spectra, such as additional losses or variation in instrument response with wavelength, can be incorporated into the source, detector, or buffer absorbance terms. The goal is to extract a measure of the absorbance of the protein, which is assumed to be proportional to the protein's concentration, in a way that is insensitive to the large, temperature-dependent absorbance of the buffer. Formally, we seek a calibration vector, \mathbf{f} , that has the properties that it is orthogonal to the nonprotein terms of eq 1 but has a large overlap with the protein absorbance spectrum. The protein concentration will then be proportional to the inner product of the calibration vector with the logarithmized single-beam spectrum of a sample containing an unknown amount of protein

$$C = \mathbf{f}^T \mathbf{s}_{\text{unknown}} \quad (2)$$

- (13) Arnold, M. A.; Burmeister, J. J.; Small, G. W. *Anal. Chem.* **1998**, *70*, 1773–1781.
- (14) Swierenga, H.; Wulfert, F.; de Noord, O. E.; de Weijer, A. P.; Smilde, A. K.; Buydens, L. M. C. *Anal. Chim. Acta* **2000**, *411*, 121–135.
- (15) Wulfert, F.; Kok, W. T.; de Noord, O. E.; Smilde, A. K. *Anal. Chem.* **2000**, *72*, 1639–1644.
- (16) Swierenga, H.; Haanstra, W. G.; de Weijer, A. P.; Buydens, L. M. *Appl. Spectrosc.* **1998**, *52*, 7–16.
- (17) Wold, S.; Antti, H.; Lindgren, F.; Öhman, J. *Chemom. Intell. Lab. Syst.* **1998**, *44*, 175–185.
- (18) Sjöblom, J.; Svensson, O.; Josefson, M.; Kullberg, H. Wold, S. *Chemom. Intell. Lab. Syst.* **1998**, *44*, 229–244.
- (19) Fearn, T. *Chemom. Intell. Lab. Syst.* **2000**, *50*, 47–52.
- (20) Horlick, G. *Anal. Chem.* **1972**, *44*, 943–947.
- (21) Hazen, K. H.; Arnold, M. A.; Small, G. W. *Appl. Spectrosc.* **1994**, *48*, 477–483.
- (22) Hu, S.-Y. B.; Arnold, M. A.; Wienczek, J. M. *Anal. Chem.* **2000**, *72*, 696–702.
- (23) Lorber, A. *Anal. Chem.* **1986**, *58*, 1167–1172.
- (24) Lorber, A.; Faber, K.; Kowalski, B. R. *Anal. Chem.* **1997**, *69*, 1620–1626.
- (25) Berger, A. J.; Koo, T.-W.; Itzkan, I.; Feld, M. S. *Anal. Chem.* **1998**, *70*, 623–627.
- (26) Xu, L.; Schechter, I. *Anal. Chem.* **1997**, *69*, 3722–3730.
- (27) Goicoechea, H. C.; Olivieri, A. C. *Anal. Chem.* **1999**, *71*, 4361–4368.

The method as described here assumes that the absorbance spectrum of the protein is independent of temperature. The experiments described later indicate that this is a satisfactory assumption. In situations where this is not true, it is possible to add correction terms to account for variation of the protein's absorption spectrum with temperature.

Construction of the calibration vector is performed using two sets of measurements. The first is a series of single-beam measurements on a buffer sample that does not contain protein over the range of temperatures expected during crystallization. The principal components of these spectra are determined from an eigenvector decomposition of the spectra. If \mathbf{B} denotes the matrix of buffer single-beam spectra (each spectrum is a column in \mathbf{B}), the matrix containing the principal components of the buffer spectra, \mathbf{B}' , is given by

$$\mathbf{B}' = \mathbf{B}\mathbf{E} \quad (3)$$

where \mathbf{E} is the matrix of the first few normalized eigenvectors of the product $\mathbf{B}'\mathbf{B}$. The principal components can be used as a basis for describing the buffer absorbance spectra at any temperature. The first few of these components, ordered by descending eigenvalue, contain significant information about the variation of the buffer spectra with temperature, whereas the remainder describe only the noise in the buffer spectra.

The second step in constructing the calibration vector involves an absorbance measurement of a sample containing a known concentration of protein. It is assumed that the protein concentration of this sample will have been measured using a standard method, such as ultraviolet spectroscopy or an enzymatic assay. A net analyte signal vector, \mathbf{g} , is constructed by removing the projection of the spectrum of a sample containing protein, $\mathbf{s}_{\text{calibration}}$, onto the buffer principal components from the spectrum of the sample with protein

$$\mathbf{g} = \left(\mathbf{I} - \sum_i \frac{\mathbf{B}'_i \mathbf{B}'_i^T}{\mathbf{B}'_i^T \mathbf{B}'_i} \right) \mathbf{s}_{\text{calibration}} \quad (4)$$

where \mathbf{I} is the identity matrix and \mathbf{B}'_i is the i th principal component of the buffer (the i th column of \mathbf{B}'). The calibration vector, \mathbf{f} , is then determined by scaling the net analyte signal vector so that its inner product with $\mathbf{s}_{\text{calibration}}$ gives the known concentration, $C_{\text{calibration}}$

$$\mathbf{f} = \left(\frac{C_{\text{calibration}}}{\mathbf{g}^T \mathbf{s}_{\text{calibration}}} \right) \mathbf{g} \quad (5)$$

This calibration vector is, by construction, orthogonal to the buffer spectra at all temperatures (within the approximation that the buffer spectra are represented by the finite number of principal components). The inner product of the calibration vector with a measured spectra will be proportional to the analyte concentration, but independent of the buffer temperature. Although the previous description is given in terms of logarithmized single-beam spectra, the procedure is also valid for raw single-beam spectra over limited concentration ranges. The validity of the method for nonlogarith-

mized single-beam spectra is important if the spectral resolution of the optical measurement system is not narrow with respect to the sample's spectral features.

EXPERIMENTAL TRIALS WITH GLUCOSE ISOMERASE

To evaluate the effectiveness of this method for monitoring protein concentration, measurements were performed using the protein glucose isomerase in a simple buffer solution. Samples were prepared from concentrated glucose isomerase obtained from Genecore, Int. (Rolling Meadows, IL). The protein was purified by dialysis with molecular porous membrane tubing with a molecular weight cutoff of 6000–8000. The tubing containing 37 mL of glucose isomerase was put in 800 mL of sodium phosphate buffer with a pH of 6.0 and dialyzed overnight in a refrigerator. The buffer was then exchanged with fresh buffer and dialysis again continued overnight. The buffer solution consisted of 0.1 M sodium phosphate, 10 mM magnesium chloride, and 0.05 mM cobalt chloride. After dialysis, the concentration of the glucose isomerase was 69 mg/mL as determined by ultraviolet spectroscopy at 280 nm. Samples were prepared by gravimetrically diluting the purified stock using the phosphate buffer. Samples were refrigerated until the optical measurements were complete. The total time from protein purification to the completion of the data collection was 4 days.

Single-beam spectra were collected using a Nicolet Magna 550 FTIR spectrometer (Nicolet, Madison, WI) with a 20 W tungsten-halogen lamp, a calcium fluoride beam splitter, and a liquid nitrogen cooled InSb detector. The near-infrared beam was passed through a multilayer interference filter (Barr Associates, Westford, MA) before being incident on the sample to isolate the 5000–4000 cm^{-1} range and avoid unnecessary sample heating. Sample temperature was controlled using a water-jacketed sample cell (Wilma, Buena, NJ) connected to a thermostated water bath (Cole-Parmer, Vernon Hills, IL). The sample cell had sapphire windows and a path length of 1.5 mm. The water bath was regulated to within ± 0.1 °C. All reported temperatures are that of the bath.

Initial measurements were performed on the phosphate buffer as a function of temperature. Single-beam spectra were collected as the sample temperature was decreased from 24 to 4 °C in 4 °C steps. Spectra collection began immediately after the water bath reached the desired set point. A single spectrum originating from 128 co-added interferograms was recorded at each temperature.

Measurements on samples containing protein were then made at 4, 8, 12, 16, 20, and 24 °C for 10 protein concentrations ranging from 0.25 to 50.2 mg/mL. Spectra from all concentrations at a given sample temperature were measured before moving to the next temperature. The order of sample temperatures and the order of sample concentrations within each temperature were randomized to avoid unintentional correlations with time.¹³ After placement in the spectrometer, each sample was allowed to equilibrate for four minutes before data collection began. Triplicate spectra of 128 co-added interferograms were recorded with a resolution of 4 cm^{-1} (the resulting point spacing was 2 cm^{-1}). The total collection time for each triplicate set was 3.5 min. The total time for the spectral measurements was 4 days. The spectrometer was not realigned by the operator during this period. One spectrum ($C = 50.2$ mg/mL, $T = 20$ °C) was discarded due to significant

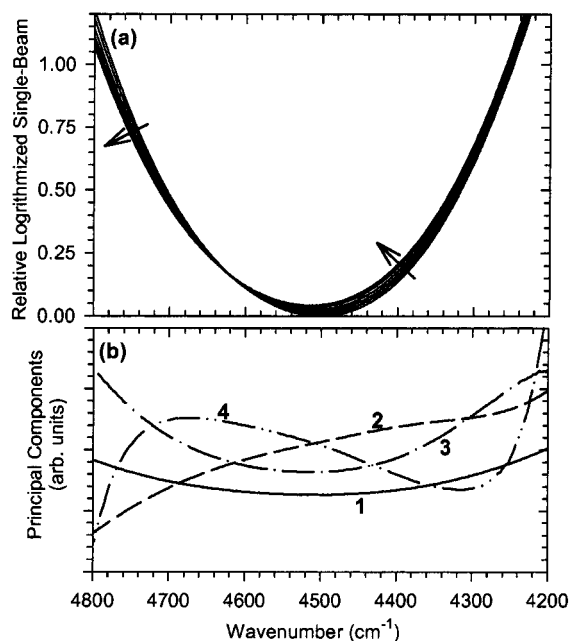


Figure 1. (a) Logarithmized single-beam spectra for 1.5 mm of sodium phosphate buffer at temperatures ranging from 4 to 24 °C. The arrows indicate the direction of increasing temperature. Note that there is a region around 4620 cm^{-1} where there is very little change with temperature. (b) The first four principal components of the spectra from (a), ordered by decreasing eigenvalue. These functions describe almost the entirety of the spectra in (a).

differences between itself and the other two spectra in its triplet; the other two spectra in the triplet were retained.

The logarithmized single-beam spectra for the buffer-only samples at the six temperatures are shown in Figure 1a. Because the transmission of the buffer between 4200 and 4000 cm^{-1} and 5000–4800 cm^{-1} is very small, only the range from 4800 to 4200 cm^{-1} is included in the analysis. The first four principal components of these spectra are shown in Figure 1b, numbered by decreasing eigenvalue. The remaining two principal components describe spectral noise. The first principal component is roughly the mean of the buffer spectra. The second corresponds to the shift to higher energy of the water absorbance minimum with increasing temperature. The third represents the broadening of the minimum with increasing temperature. The fourth does not appear to be a function of sample temperature and possibly represents baseline variations in the FT-IR spectrometer.

A calibration vector, shown in Figure 2a, was constructed from a single measurement of the sample with the highest protein concentration (50.2 mg/mL) at the first sample temperature (12 °C) using eqs 4 and 5. Four principal components were used to represent the buffer absorption spectra. This calibration vector contains all of the information required to predict protein concentration at any temperature within the calibration range. For reference, Figure 2b shows the absorbance spectrum of glucose isomerase from a representative sample solution. The basic structure of the absorbance spectrum is preserved in the calibration vector (e.g., peaks at 4610, 4370, and 4260 cm^{-1}).

The concentrations of the remaining 59 samples were calculated using this calibration vector by calculating its inner product with each of the logarithmized single-beam spectra. Predicted protein concentrations are shown as circles in Figure 3. The

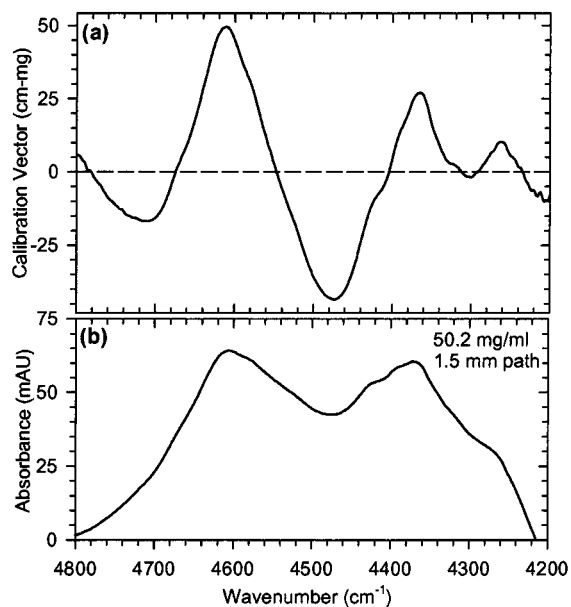


Figure 2. (a) Calibration vector for glucose isomerase. (b) Absorbance spectrum for a 50.2 mg/mL solution of glucose isomerase at 24 °C.

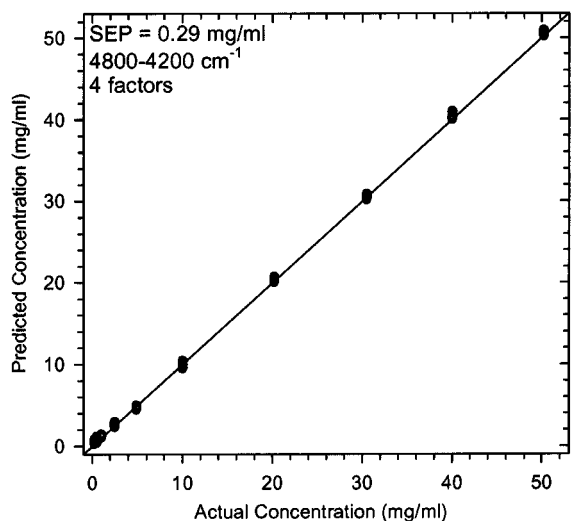


Figure 3. Prediction results for glucose isomerase in a sodium phosphate buffer over a 4 to 24 °C temperature range.

standard error of prediction (SEP) using the 4800–4200 cm^{-1} spectral range is 0.29 mg/mL. Residual errors as a function of actual protein concentration and sample temperature are shown in Figure 4. In Figure 4a, the diamond symbols indicate the average of the residuals across all temperatures for each concentration; the diamond symbols in Figure 4b indicate the average of the residuals across all concentrations for each temperature. There is little trend in the residuals with either parameter.

Model performance can be improved by optimizing the spectral range used in the calibration. By focusing on a narrow spectral range containing a distinct analyte feature, measurement noise, baseline variations, and temperature variations of the buffer absorbance are effectively reduced. Calculations were performed for all possible spectral ranges between 4900 and 4100 cm^{-1} . A minimum SEP of 0.12 mg/mL was obtained when the spectral range was restricted to 4670–4595 cm^{-1} . This region includes both the large peak at 4610 cm^{-1} for glucose isomerase (shown in

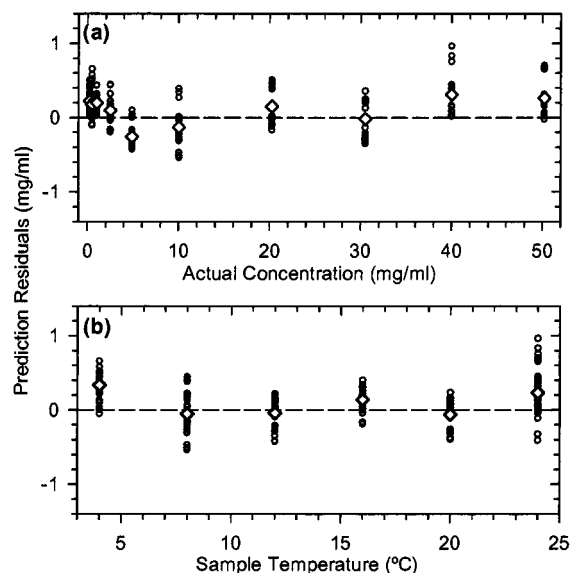


Figure 4. Prediction residuals for glucose isomerase as a function of (a) actual concentration and (b) sample temperature. The circles indicate residuals for each spectrum, while the diamonds are the average residual across all spectra with the same concentration or temperature.

Table 1. Minimum SEP Values for Glucose Isomerase for Spectra Consisting of 2–6 Wavelengths^a

| wavelengths in calibration | no. of principal components | min SEP (mg/mL) |
|----------------------------|-----------------------------|-----------------|
| 2 | 1 | 13.1 |
| 3 | 1 | 0.27 |
| 3 | 2 | 0.18 |
| 4 | 2, 3 | 0.13 |
| 5 | 2, 3 | 0.13 |
| 5 | 4 | 0.12 |
| 6 | 2–5 | 0.12 |

^a The second column indicates the number of principal components used to describe the temperature-dependent buffer absorption spectra.

Figure 2b) and the smallest temperature variation in the buffer absorption (Figure 1a). The large 4610 cm^{-1} peak is due to combinations of bending and stretching modes of the N–H bond in the peptide chain and will be present in all proteins.

It is also possible to predict protein concentrations using spectra sampled at irregularly spaced wavelengths. This could be useful for compact instruments based on laser diode arrays. To establish the range of model performance, SEP values were calculated using sets of three wavelengths drawn at random from the FT-IR spectra. The minimum SEP values involve wavelengths near the 4610 cm^{-1} peak of the glucose isomerase absorption spectra, shown in Figure 2b. A representative wavelength set producing a SEP of 0.18 mg/mL includes 4596, 4609, and 4638 cm^{-1} . Similar results are obtained when four wavelengths are used. Minimum SEP values obtained using 2–6 wavelengths are shown in Table 1. These results indicate that accurate measurements of glucose isomerase concentration are possible with a small number of wavelengths despite the large, temperature-sensitive interference of the buffer. Except for the extreme case of spectra comprised of two wavelengths, SEP values are superior to those obtained using the 4800–4200 cm^{-1} range.

Table 2. Minimum SEP Values for Measurements of Lysozyme with Varying Numbers of Wavelengths

| wavelengths in calibration | no. of principal components | minimum SEP (mg/mL) |
|----------------------------|-----------------------------|---------------------|
| 2 | 1 | 9.1 |
| 3 | 1 | 0.45 |
| 3 | 2 | 0.26 |
| 4 | 2, 3 | 0.13 |
| 5 | 2 | 0.14 |
| 5 | 3, 4 | 0.13 |
| 6 | 2 | 0.14 |
| 6 | 3–5 | 0.13 |

EXPERIMENTAL TRIALS WITH LYSOZYME

Model performance was evaluated on a second protein in order to determine the general applicability of the method. Hen egg-white lysozyme is a well studied protein whose thermodynamics, growth kinetics, and crystallization properties are well-known. Previous experiments have demonstrated that superior lysozyme crystals can be grown using predetermined temperature profiles based on this information.^{8,10}

The lysozyme spectra used here were originally collected for evaluation of Fourier filtering as a temperature-insensitive measurement of protein concentration.²² Samples of lysozyme in a sodium acetate buffer of pH 4.6 were prepared with concentrations ranging from 0 to 52 mg/mL. Spectra were collected using the same FT-IR spectrometer used in the glucose isomerase study. Measurements were made at seven temperatures between 4 and 24 °C with a 1.5-mm sample path length. Further details of the experiment are given in ref 22. Fourier filtering followed by linear regression against the area of the peak near 4600 cm^{-1} was shown to be effective for removing the temperature interference in the measurement of lysozyme, producing an SEP of 0.28 mg/mL. Multivariate PLS models were also constructed using an optimized spectral range that had a SEP of 0.09 mg/mL. However, it should be noted that the PLS calibration process requires a much more extensive calibration set than does the Fourier filtering/linear regression model (more than half of the spectra were used for calibration of the PLS model).

In the present work, the buffer absorption basis was determined using spectra from the lowest concentration samples at each of the seven temperatures between 4 and 24 °C (the actual concentration of these samples was 0.036 mg/mL, which is significantly smaller than typical SEP values). The buffer absorption basis functions are nearly identical to those obtained in the glucose isomerase analysis. The calibration vector was then constructed from measured spectra for the highest concentration sample at 24 °C. The SEP using the 4800–4200 cm^{-1} spectral range is 0.20 mg/mL, while that for an optimized wavelength range (4750–4570 cm^{-1} , determined as for glucose isomerase) is 0.14 mg/mL. The results for measurements with 2–6 wavelengths are shown in Table 2.

To determine whether the same discrete wavelengths can be used for both glucose isomerase and lysozyme, sets of wavelengths were again drawn at random and used for analysis of both proteins. After calibration and prediction, the larger SEP value from either protein was retained. The minimum SEP using a set of three wavelengths is 0.26 mg/mL, which is comparable to that for analysis of lysozyme by itself. Optimal calibrations can be

constructed using wavelengths near the 4610 cm^{-1} absorption feature. A representative wavelength set producing a SEP of 0.26 mg/mL contains 4592 , 4611 , and 4630 cm^{-1} . When four wavelengths are used, the minimum SEP value drops to 0.15 mg/mL . There is a significant range of wavelengths that can produce SEP values of approximately 0.2 mg/mL . There are two optimal spectral bands for joint analysis of both proteins. One is from 4695 to 4435 cm^{-1} and the other is from 4758 to 4504 cm^{-1} . Both yield SEP values of 0.14 mg/mL and include the 4600 cm^{-1} absorption feature.

PROTEIN CRYSTALLIZATION CALIBRATION PROCEDURE

The primary advantage of this technique over a typical net analyte signal method or the more established PLS approach is the relatively simple calibration procedure. This is particularly relevant to the problem of concentration monitoring during protein crystallization. A possible calibration procedure for a protein crystallization experiment can be envisioned where no protein is lost during the calibration procedure. The sample chamber can first be filled with the crystallization buffer, but no protein. Spectral measurements made on the buffer over the expected temperature range can be used to determine the principal components of the buffer absorption. The buffer can then be replaced with the growth solution that contains protein. A single spectral measurement can be made while the sample is at its highest temperature (or lowest temperature, if the protein in the growth buffer is exothermic), which, when combined with the principal components of the buffer absorption, yields the net analyte signal vector. If the protein

concentration of the initial growth solution is known, the calibration vector can be scaled and the calibration process is complete. The temperature of the sample can then be changed to initiate nucleation and growth with no loss of protein.

CONCLUSIONS

The calibration method described here is a simple means to obtain temperature-insensitive measurements of protein concentration from aqueous samples using noninvasive, near-infrared spectroscopy. It delivers performance comparable to multivariate PLS calibration methods, with the benefit of a simplified calibration process. For the two proteins included in this analysis, it is possible to choose a single set of four wavelengths that give very satisfactory monitoring capability. It is likely, on the basis of the physical origin of the 4600 cm^{-1} absorption band, that these wavelengths will also work for a number of other proteins. The simplified calibration process makes it ideal for monitoring the crystallization process, especially when the quantity of the protein for crystallization is limited.

ACKNOWLEDGMENT

This research was supported in part by the National Aeronautics and Space Administration, Office of Life and Microgravity Sciences (contracts NAG8-1352 and NAG8-1386).

Received for review April 7, 2000. Accepted August 2, 2000.

AC000406U