Determination of Glucose in Human Aqueous Humor Using Raman Spectroscopy and Designed-Solution Calibration

CHRISTINE C. PELLETIER, JAMES L. LAMBERT,* and MARK BORCHERT

Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, California 91109 (C.C.P., J.L.L.); and Division of Ophthalmology, Childrens Hospital Los Angeles and Department of Ophthalmology, Keck School of Medicine, University of Southern California, Los Angeles, California 90027 (M.B.)

Glucose concentrations of *in vitro* human aqueous humor (HAH) samples from cataract patients were determined using 785 nm Raman spectra and partial least squares (PLS) calibration. PLS models were created from spectra of prepared calibration solutions rather than aqueous humor samples. Spectra were obtained with an excitation energy (100 mW for 150 s), which was higher than can be applied *in vivo*, to decrease the models' contribution to prediction uncertainty. The solutions contained experimentally designed levels of glucose, bicarbonate, lactate, urea, and ascorbate. Multiplicative signal correction of spectra helped compensate for the ±20% drift in laser power observed at the sample over six noncontiguous days of data collection. Seventeen HAH samples containing 38–775 mg/dL of glucose exhibited a root-mean-square error (RMSEP) of 22 mg/dL, coefficient of determination (r^2) of 0.989, and bias of 6 mg/dL when predicted from lower energy (30 s) spectra collected contemporaneously with fifty calibration spectra. Similar results were obtained even when spectral data were gathered separately for human aqueous humor samples and calibration samples: 10 HAH samples, calibrated on 25 solutions measured 3.6 weeks earlier, exhibited an RMSEP of 23 mg/dL, r^2 of 0.992, and bias of 9 mg/dL. The results demonstrate progress toward the determination of glucose levels in patient-derived aqueous humor using laboratory-derived "artificial aqueous humor" calibration solutions.

Index Headings: **Raman; Partial least squares; PLS; Multivariate; Calibration; Multiplicative signal correction; MSC; Quantitative; Glucose; Aqueous humor.**

INTRODUCTION

The aqueous humor in the anterior chamber of the eye is an attractive target for noninvasive spectroscopic glucose monitoring. Glucose monitoring is important for the management of diabetes mellitus[1] and may provide a paradigm for broader multiple-metabolite monitoring. Aqueous humor in the eye is optically more accessible than blood in tissues, since the anterior chamber is close to the exterior of the body and bounded by the transparent cornea. Glucose levels in aqueous humor appear to correlate with blood glucose levels,[2–5] and the time lag for reflecting a change in blood glucose is probably short.† In addition, aqueous humor is a less complex fluid than is blood, since the blood–aqueous barrier filters out most large molecules.

Infrared absorption spectroscopy has received the most attention for noninvasive glucose determination in blood. However, polarimetry, stimulated Raman spectroscopy, and normal Raman spectroscopy are the techniques that have been primarily investigated for glucose determination in aqueous humor.[10,11] We, among others, have chosen Raman spectroscopy with multivariate analysis—here, partial least squares (PLS)[12]—for determining glucose in aqueous humor because this combination can quantitatively extract the signal due to relatively low levels of glucose despite the presence of interfering components.[4,13–16]

Ideally, a multivariate calibration model is created from a set of spectra (or other data) that contain all of the variation that will be encountered in future spectra. The future variation includes changes in instrumental and environmental parameters as well as the range of each chemical component's concentration. To deal with the potential overload of variables to consider, several approaches are often used together. Some potential variables are at least approximately controlled, such as the use of a buffer to maintain a constant pH. Some variables are corrected for, as in the calibration of an instrument before using it or the preprocessing of spectra to correct for a (known) non-constant detector response. Some variables, such as instrument-specific characteristics, are dealt with using model transfer techniques after development of the multivariate model. The remaining variables are hopefully represented in the multivariate calibration model itself. For example, a calibration model for determining glucose in the aqueous humor of patients could conceivably be built on spectra from a very large number of volunteers, with the expectation that the spectra so obtained would represent the spectra of future patients' aqueous humor.

Providing a calibration set based directly on *in vivo* human aqueous humor has at least two disadvantages. The first includes the expense, difficulty, and potential hazards of obtaining spectra from a large number of volunteers, especially while the technique is still being developed, and the problems associated with repeating the procedure at a later time if recalibration were necessary. The second is that the variation in glucose level would not necessarily be independent from the variation in the levels of other spectrally active components in the aqueous humor. For example, the model's prediction of glucose concentration could actually be based on bicarbonate and/or lactate Raman peaks. This collinear-variation problem can lead to an initial model that appears to perform well, but which does not maintain its performance later when the collinearity fails to hold.

Received 6 October 2004; accepted 25 May 2005.

* Author to whom correspondence should be sent. E-mail: James.L.Lambert@jpl.nasa.gov.

† Aqueous humor glucose response to a change in blood glucose has been reported as being almost immediate above 200 mg/dL blood glucose based on nine rabbits,[4] nearly instantaneous in one human,[6] −2.2 to 11.2 min (average <5 min) in five rabbits,[7] an immediate change in three rabbits,[8] and on the order of 30 min in one rabbit.[9]

0003-7028/05/5908-1024$2.00/0
© 2005 Society for Applied Spectroscopy

Since the number and molecular size of the major components of human aqueous humor are relatively small, PLS calibration for *in vivo* glucose might be possible based on the spectra of laboratory-derived solutions rather than patient-derived spectra. Such a calibration could be repeated as needed to reflect changes in the instrument or procedure and is much more accessible during the development stage. Potentially, higher quality spectra for constructing the PLS model can be obtained when laboratory samples are used, since they can be probed with a much higher laser power than is safe in an eye.[17,18] In addition, component concentrations in laboratory samples can be experimentally designed to avoid the problem of concentration collinearity.

Since we did not yet have access to *in vivo* human aqueous humor for this study, we examine here *in vitro* human aqueous humor glucose determination. Although Raman spectra of human and rabbit aqueous humor have previously been examined qualitatively,[19] and rabbit aqueous humor glucose has been quantified with a Raman-PLS method using the same samples for calibration and prediction,[4] we report the first (to our knowledge) successful quantitative determination of glucose from Raman spectra of natural (*in vitro*) aqueous humor samples calibrated with laboratory-derived ''artificial aqueous humor'' (AAH) solutions. The results not only show that aqueous humor glucose concentration can be accurately predicted in samples that are completely separate from the PLS calibration solutions, but that a predictive PLS model can be stable over a time period of several weeks.

Future extension of this approach to the measurement of *in vivo* aqueous humor may require several additional steps. Spectral interference may be amenable to correction by augmenting the PLS model with appropriate spectra (e.g., ascorbate, cornea, lens spectra). *In vivo* pH and temperature differences relative to *in vitro* may be compensated by adjusting the conditions under which calibration samples are measured. Instrument design will have to consider eye-movement issues. A review of obstacles related to clinical implementation is included elsewhere.[18] Nevertheless, the results reported here indicate that overcoming the hurdles of accessible model recalibration, PLS model stability, and eye-compatible laser energy requirements is probably possible.

Multivariate calibration allows information from the entire Raman spectrum to contribute to the determination of an analyte. The net result of PLS calibration can be expressed as a calibration vector, **b** (also called a regression vector or calibration ''spectrum''). The dot product of a sample spectrum **x** with **b** predicts the glucose concentration, c, of that sample:

$$\mathbf{x}\cdot\mathbf{b} = c \tag{1}$$

In practice, the spectrum is mean-centered before applying Eq. 1 by subtracting the average of the spectra used in the calibration. The average glucose concentration of the calibration samples is then added to the value of c found in Eq. 1 to obtain the final concentration of the sample.

Preprocessing the spectra before mean-centering with multiplicative signal correction (MSC),[12] as suggested by Goetz et al.,[16] can correct for multiplicative interference and/or different spectral offsets. MSC linearly rescales each spectrum to make the corrected spectrum similar to a reference spectrum (e.g., the average of the spectra used in the calibration). Additive and multiplicative scaling factors for each measured spectrum are obtained by regressing the spectrum's intensities on the reference spectrum's intensities. The dominance of Raman scattering of water in the spectra of the low-concentration aqueous solutions considered here justifies this approach, since corrections based on the whole spectrum are unlikely to be distorted by large peaks in some of the individual spectra.

EXPERIMENTAL

Raman Instrument. A titanium:sapphire laser, pumped by an argon-ion laser, produced ~400 mW of 785 nm excitation, which was coupled via a single-mode optical fiber to a Kaiser Optical Systems (KOSI) confocal Raman microscope.[18] A 20× objective, producing a depth of field of 130 μm full width at half-maximum (FWHM) and a beam waist radius of 5 μm in air, focused ~100 mW (±20% due to drift) of laser power into a sample. The same objective collected the Raman-scattered photons (180° geometry), which were transferred through an optical fiber to the f/1.8 KOSI HoloSpec spectrograph and imaged onto a liquid-nitrogen-cooled charge-coupled device (CCD) camera (Princeton Instruments). The camera electronics were interfaced to a shutter, which blocked the laser beam except during spectral data collection, as well as to a computer. The microscope interface included a video camera for viewing a real-time image of the laser impinging on the sample.

An in-house Matlab (MathWorks, Inc.) computer program coordinated laser-power monitoring with collection of spectral data, which were processed and stored using HoloGRAMS 4.0 (KOSI) software. Fifty spectra having an acquisition time of 3 s (~300 mJ) per spectrum were collected in series for each sample. These could be added together later to create equivalent exposure times of 3–150 s (0.3–15 J). Laser power was continually monitored with a power meter having a detector head mounted on top of the microscope to receive a fraction of the laser light, which was picked off by a custom-installed quartz beam-splitter. Periodic laser power measurements at the sample position were used to calibrate the picked-off fraction, indicating that the laser power impinging on samples varied between 79 and 124 mW (~100 mW ± 20%) over the course of data gathering.

Spectra of neon emission and white light, provided by a HoloLab (KOSI) calibration accessory, were employed to calibrate spectral wavelength and intensity axes. The measured laser wavelength for converting wavelengths to wavenumbers was determined from the cyclohexane peak at 801.8 cm^{-1}. Spectral calibrations were refreshed at the beginning of each day of experimentation, and whenever the instrument alignment had been disturbed, by collecting new white-light and neon spectra. The HoloGRAMS software automatically dark-count subtracted and spectrally calibrated the sample spectra as they were collected. Intensity calibration (flat-fielding) was carried out by dividing the raw sample spectrum by the white-light spectrum corrected to the black-body response, thus reducing fixed pattern noise and yielding spectra that were more instrument independent.

Partial Least Squares Model Calibration Samples. Two sets of 25 artificial aqueous humor (AAH) samples, Set I and Set II, were planned using a 25-sample, 5-level partial factorial experimental design[20] to have independently varying concentrations of five components. Four of the components had the same nominal concentrations in corresponding samples of the two sets. The glucose concentration levels were different in the two sets, but the mean of the glucose levels in each set was kept the same, at 400 mg/dL. This arrangement made the two sets combinable into a single set of 50 samples, theoretically having zero covariance between the components. The two sets of AAH samples were prepared independently, from separate stock solutions, on separate days, 3.6 weeks apart.

As prepared, glucose, bicarbonate, lactate, urea, and ascorbate concentration levels spanned the ranges from zero to 770, 250, 200, 180, and 90 mg/dL, respectively (~13×, 2×, 5×, 7×, and 5× relative to concentrations in normal human aqueous humor[3,19,21]). The high ends of these ranges were chosen to include extreme concentrations that might be encountered in aqueous humor based on the concentrations found in blood in various disease states.[22] Samples were prepared by diluting volumes of stock solutions of the components with calculated volumes of saline solution (aqueous sodium chloride, 0.77 mg/dL[21]). Stock solutions were made by dissolving known masses of powdered reagents in saline solution, adjusting the pH of the solutions to approximate *in vitro* human aqueous humor pH, and diluting to volume in volumetric flasks. To get an idea of what the *in vitro* HAH pH would be, the pH of five human aqueous humor samples was measured, yielding 8.8 ± 0.3, consistent with a previous report.[23] The post-dilution pH values for the AAH samples were, for Set I, 8.2 ± 1.4, and for a 10-sample subset of Set II, 7.9 ± 1.3.

Human Aqueous Humor Samples. Specimens of human aqueous humor (HAH) were collected from patients undergoing cataract surgery at Los Angeles County/University of Southern California Medical Center, with Institutional Review Board (IRB) approval. The specimens were stored at −30 °C until the day of analysis, when they were thawed and vortexed. Samples were kept on ice until an hour before measurement, when they were allowed to come to room temperature.

Set A contained seven natural, *in vitro* HAH samples that were interspersed with Set I AAH samples for spectral and reference-concentration measurement. Set B contained ten *in vitro* HAH samples, six of which were augmented with glucose to represent the hyperphysiological concentrations that may be encountered in uncontrolled diabetes. Glucose augmentation was carried out by (1) evaporating 10–70 μL of an 800 mg/dL aqueous glucose solution in micro-centrifuge tubes, (2) adding about 90 μL of HAH, and (3) vortexing to redissolve the glucose. Set B samples were interspersed with Set II AAH samples for measurement.

The reference concentrations of D-glucose in the HAH samples were measured using a 2700 Select Biochemistry Analyzer (YSI Life Sciences). Internal recalibration of the instrument was completed before every sample measurement to avoid correlated errors. Periodically, the YSI-supplied calibration solution used for internal recalibration was measured as if it were a sample, and all AAH samples were measured with the Analyzer as well. Reference concentrations for the 17 HAH samples spanned the range from 38 to 775 mg/dL of glucose.

Sample Handling. Spectra were obtained from an aliquot of a sample while it was contained in a physical model of the anterior chamber of the eye.[14] The physical model consisted of a poly(methyl methacrylate) contact lens overturned on a quartz microscope slide. The slide was affixed to a 9-mm-thick metal base so that the assembly (''sample carrier'') could be interchanged with the 1-cm cuvette of cyclohexane used as a spectroscopic reference without disturbing the microscope stage height. Before each sample measurement, the sample carrier was aligned to the laser beam by translating the microscope stage in the horizontal plane. The height of the stage was carefully set at the beginning of each day of gathering sample spectra and whenever the stage had been moved vertically, which rarely occurred. First the height was found where the laser reflection off the top surface of the contact lens was in focus in the video image. Then the stage was raised 1100 μm so that the focus of the laser beam was within the sample.

Small air bubbles were observed to collect on the inner surface of the contact lens over time after receiving the sample, whether or not it was exposed to the laser beam. To reduce the number and size of these bubbles, sample solutions were brought to ambient room temperature and centrifuged 5 min before being loaded into the contact lens. Reference glucose concentration values were determined with the Biochemistry Analyzer using a 10 μL aliquot from the micro-centrifuge tube originally holding the sample as well as directly from the contact lens immediately after spectral data collection. The average of the two Analyzer results was used as the glucose reference concentration.

Data Processing. Individual 3-s spectra (50 per sample) were imported into Matlab for cosmic-ray correction and summation to an equivalent exposure time, t. In general, the first $t/3$ spectra for each sample were summed, while the remainder of the fifty 3-s spectra was ignored (e.g., for a 12-s equivalent exposure time, the first four 3-s spectra were added together). Resulting spectra were preprocessed and then multiplied according to Eq. 1 to yield concentrations. Statistics to assess performance, such as the root-mean-square error of prediction (RMSEP), coefficient of determination (r^2), and bias (mean difference between predicted and reference concentrations), were generally calculated from a set of concentration predictions containing one prediction per sample.

The 17-HAH concentration standard deviations in Fig. 6 below (denoted by ♦) and the statistics describing glucose predictions from 30-s and 12-s spectra in Table II, however, were calculated from a larger number of predictions, N, than the number of samples in a set. This was done to reduce the uncertainty inherent in statistics calculated from sets containing small numbers of results. The extra predictions came from sample spectra of equivalent exposure time t, which were obtained by summing separate, previously ignored 3-s spectra. For example, fifty 3-s spectra were collected for each of the 17 HAH samples. The sum of the first ten 3-s spectra for each

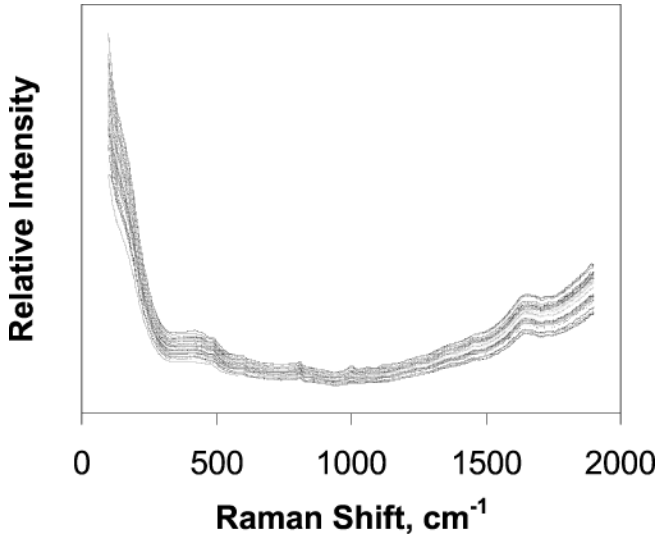

Fig. 1. Spectra of the 50 designed-solution (AAH) samples in Sets I and II. Each spectrum is the result of summing (after correction for cosmic rays) fifty 3-s spectra that had been collected in series, yielding an equivalent exposure time of 150 s. The 300–1500 cm^{-1} region was selected before calculating the 50-AAH PLS model.

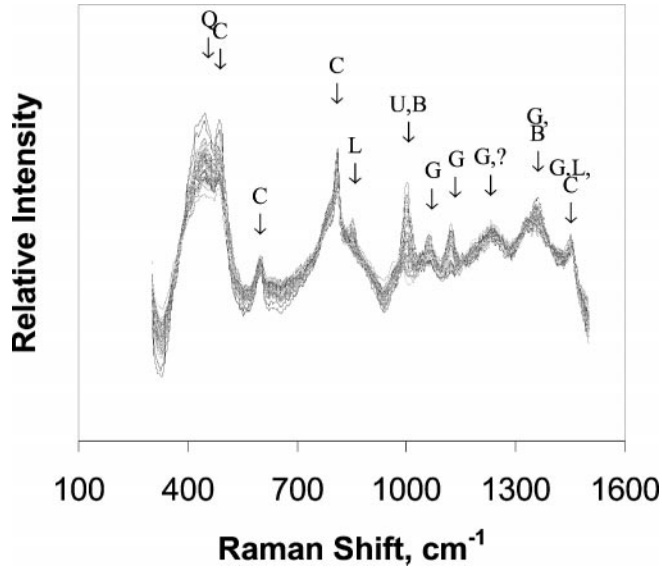

Fig. 2. The spectra of Fig. 1 after subtraction of a smooth background. Peaks are marked to identify the major contributor(s): (**G**) glucose, (**Q**) quartz, (**C**) contact lens, (**L**) lactate, (**U**) urea, (**B**) bicarbonate, and (**?**) unidentified contributor. See Fig. 4 for glucose peak locations.

sample yielded 30-s spectra from which 17 concentrations were predicted with the 50-AAH model. This process was then repeated using the second through fifth sets of ten spectra for each sample to obtain a total of $N = 5 \times 17 = 85$ predictions. The root-mean-square error of prediction calculated from the 85 predictions yielded 22.0 mg/dL for the 30-s RMSEP, as recorded in the fourth row of Table II. The 17-HAH standard deviations in Fig. 6 were similarly calculated, setting $n = N$ in Eq. 2.

PLS_Toolbox 3.0 (Eigenvector Research, Inc.) routines were employed for preprocessing the spectra (i.e., selection of the 300 to 1500 cm^{-1} region and multiplicative signal correction (MSC), followed by mean-centering) and for constructing partial least squares (PLS) models. For display purposes only, a smooth background was subtracted from the spectra with the Pearson's algorithm in KOSI's HoloReact program.

RESULTS AND DISCUSSION

Artificial Aqueous Humor Spectra. Spectra of the 50 experimentally designed ''artificial aqueous humor'' (AAH) samples of Sets I and II (150-s equivalent exposure time) are displayed in Fig. 1. Their background-subtracted counterparts are shown in Fig. 2. While the latter look better, models and predictions calculated from spectra with background subtraction generally performed no better than those calculated from spectra without subtraction.

A slight over-correction at higher wavenumbers during intensity calibration may explain the increasing baselines of the spectra of Fig. 1. An uncalibrated white-light spectrum is roughly a parabola over the 300–1500 cm^{-1} region, with a maximum near 825 cm^{-1}. Detector (CCD) sensitivity is lower at wavenumbers farther from the maximum. Since the intensity calibration essentially magnifies spectral intensities at the ends of spectra to have equal weight with those in the middle, intensity accuracy and precision degrade toward wavenumber extremes. To see if the intensity correction was adversely affecting glucose predictions, a number of PLS models were constructed using spectra that were not intensity calibrated. For the models tested, using spectra without intensity calibration did not improve glucose prediction. Therefore, only intensity-calibrated spectra were used to obtain the results in this report.

Spectral Normalization. Raman spectral intensity is proportional to laser power, and PLS model predictions of glucose concentration depend on spectral intensities, so laser power variation can result in prediction uncertainty if not compensated. Attempts were made, using several normalization methods, to correct the Raman spectra for the rather large laser power variations (±20%) that occurred during data collection. The area of the baseline-subtracted 1640 cm^{-1} water band and the continuously monitored laser power picked off by the custom-installed beam-splitter were each tried as normalization factors. However, preprocessing the spectra using MSC resulted in smaller prediction errors than employing either of these normalizations. The success of MSC may be related to its accommodation of spectral offsets as well as multiplicative effects such as laser power drift. The correction by MSC of some other error source, such as subtle fluorescence or the possibly varying elastic scattering that might occur in aqueous humor samples, may also have contributed to its success.

Designed-Solution (AAH) Components. Since glucose determination was the main objective during experimentation, extra care was taken to record the gravimetric concentration of glucose in the AAH samples. The concentrations of all components (including glucose) of the designed solutions were, however, the result of volumetric transfers of stocks solutions. Gravimetric glucose concentrations in AAH samples were calculated from the mass of the glucose stock solution added, the mass of the final sample solution, and tabulated[25] densities.

The YSI Analyzer glucose values averaged 97.7% of

TABLE I. AAH component prediction statistics using PLS models[a] based on 150-s, non-contemporaneous spectra of 25 AAH samples and volumetric reference concentrations.

Calibration set, prediction set	Statistic[b]	Components				
		Urea	Lactate	Ascorbate	Bicarbonate	Glucose
Set I, Set II	RMSEP[c]	7	17	41	21	18
Set I, Set II	r^2	0.990	0.956	0.007	0.962	0.996
Set I, Set II	bias[c]	−2.4	2.1	−9.9	−12.1	2.1
Set II, Set I	RMSEP[c]	10	16	30	14	28
Set II, Set I	r^2	0.977	0.938	0.168	0.972	0.990
Set II, Set I	bias[c]	6	4	10	6	−6

[a] All models employ five factors (''latent variables'') each.
[b] Statistics were calculated from 25 predictions in every case.
[c] Components' RMSEP and bias are in mg/dL.

the gravimetric AAH concentrations and of the YSI calibration solution concentration (manufacturer's specification = 250 mg/dL). Consequently, final YSI Analyzer reference glucose concentrations in this paper are the raw Analyzer glucose determinations divided by 0.977. Omitting this correction would not have changed the values of statistics (RMSEP, bias, r^2) reported in this paper, however, because the same reference scale was used during model calibration and for calculating prediction errors.

Reference glucose concentration uncertainty is likely to be consistent with YSI's Analyzer linearity specification of ±5%. Out of 155 Analyzer glucose determinations, 94% of the raw glucose values lay within this range (97% of the values after division by 0.977). The 155 determinations included non-blank measurements of AAH, which were compared to gravimetric concentrations, and measurements of the YSI-supplied calibration solution, which were compared to the specified concentration (250 mg/dL).

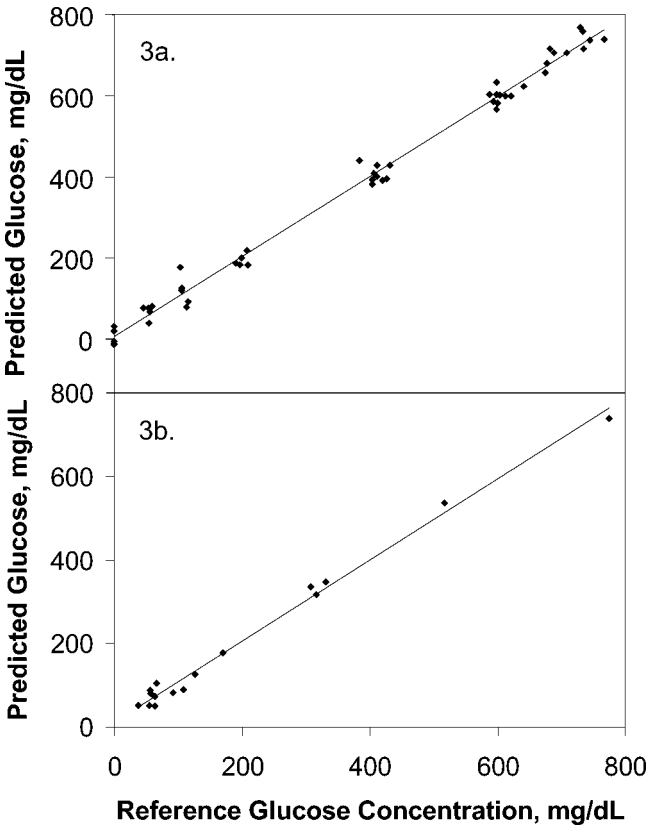

Fig. 3. Determination of glucose concentrations from Raman spectra. (**a**) The line fit to the leave-one-out cross-validation results for the 50-AAH PLS model (5 factors) based on the spectra of Fig. 1 has a slope of 0.986 and intercept of 5.8 mg/dL. The r^2 is 0.992; RMSECV is 24 mg/dL; bias is zero. (**b**) Glucose concentrations of 17 *in vitro* HAH samples (Sets A and B) predicted from their 150-s spectra using the 50-AAH PLS model are plotted versus reference concentration. The slope is 0.972; intercept is 11 mg/dL; r^2 is 0.991; RMSEP is 20.4 mg/dL; bias is 5.9 mg/dL.

A comparison of the predictive ability of non-contemporaneous PLS models for AAH components is shown in Table I. Models are based on 150-s spectra, use five factors each, and were calculated one component at a time (PLS-1). All reference concentrations used in generating Table I are the volumetric concentrations, since YSI and gravimetric measurements were not available for most of the analytes. The component prediction results in Table I are generally similar to the glucose results based on Analyzer reference concentrations in Table II, except for ascorbate. There was almost no correlation between predicted ascorbate and prepared ascorbate concentrations. This is probably due to a combination of the following observations: (1) the strongest (dominant) Raman peak in aqueous ascorbate is at 1650 cm^{-1},[26] which is outside of the spectral region selected for optimal glucose prediction, (2) the prediction vectors from Sets I and II had a couple of peaks in common, but were very noisy, indicating calibration on extraneous spectral information, and (3) ascorbate is unstable in solution at neutral pH.[27]

Example Aqueous Humor Glucose Predictions. A representative glucose calibration is shown in Fig. 3a. The 50-AAH PLS model (using five factors) was calculated from the spectra in Fig. 1 after selection of the 300–1500 wavenumber region, MSC, and mean-centering. Leave-one-out cross-validation resulted in a coefficient of determination (r^2) of 0.992, a root-mean-square error of cross-validation (RMSECV) of 24.0 mg/dL, and bias of zero. The weight vector of the PLS factor accounting for the largest amount of concentration variance is compared to an authentic glucose spectrum in Fig. 4. Reassurance that the PLS model is responding to glucose, rather than to some other serendipitous correlation, is provided by their similarity.[24] Glucose self-prediction errors (concentration residuals) of the AAH samples making up the 50-AAH model do not seem to have a noticeable trend with concentration, as illustrated in Fig. 5. Constant error variance relative to concentration, as appears to be the case here, validates the use of PLS.[12]

The glucose concentrations of the 17 Set A and Set B

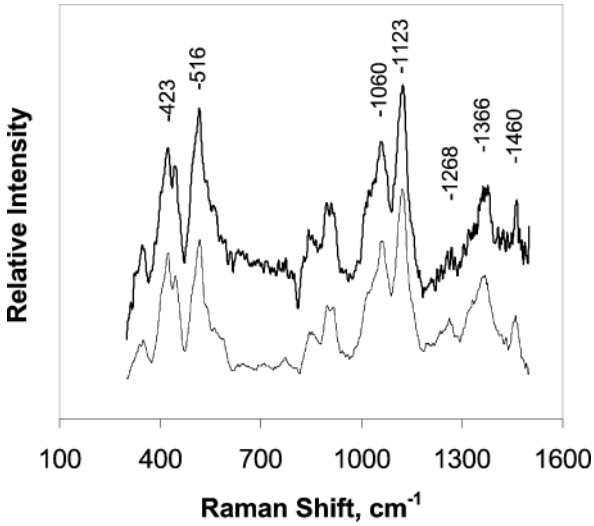

Fig. 4. Comparison of glucose spectrum with PLS-model second weight vector. Top trace is a measured glucose spectrum (772 mg/dL, 150-s equivalent exposure, offset after subtraction of a saline solution ''blank'' spectrum); bottom trace is the weight vector (scaled) associated with the second PLS factor in the 50-AAH model. The second factor accounted for 19% of the X (spectral) variance and 67% of the Y (concentration) variance in this model.

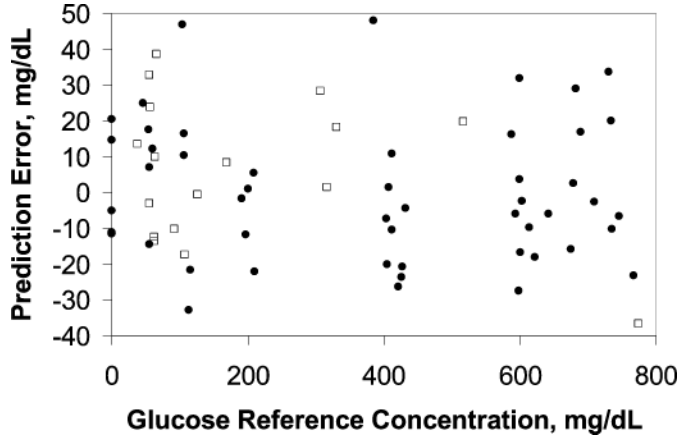

Fig. 5. Circles represent errors of predicting the samples in the 50-AAH model (concentration residuals from self-prediction, not cross-validation). Squares represent errors (predicted glucose minus reference glucose) for the 17 HAH samples found using 150-s spectra.

HAH samples, as predicted from 150-s spectra by the 50-AAH PLS model represented in Fig. 3a, are plotted in Fig. 3b and summarized in the first row of Table II. Six glucose-augmented and 11 natural *in vitro* human aqueous humor samples are represented in the total of 17 samples. The r^2 of 0.991, RMSEP of 20.4 mg/dL, and bias of 5.9 mg/dL are similar to model values. HAH concentration errors appear consistent with AAH errors, as shown in Fig. 5. These results indicate that the components of *in vitro* HAH that are not represented in the calibration solutions are interfering very little with glucose determination. The idea that glucose in human aqueous humor can be determined from Raman spectra using a relatively straightforward PLS model created from laboratory-derived artificial aqueous humor samples is therefore supported.

Model Stability Over Time. Practical use of a PLS model for aqueous humor glucose determination will require that spectra of samples (e.g., spectrum of a patient's aqueous humor) be collected separately from the spectra used to create the PLS model. To mimic this situation, a model (5 factors) was prepared from the spectra (150-s, 300–1500 cm^{-1}, MSC and mean-center preprocessed) of only the 25 AAH samples in Set I. The resulting 25-AAH PLS model was then used to predict the glucose in Set B HAH samples measured 3.6 weeks later, as shown in the third row of Table II. For the ten 150-s HAH spectra, RMSEP = 21.6 mg/dL, r^2 = 0.994, and bias = 9.4 mg/dL. These encouraging results are nearly the same as those found using the 50-AAH model to predict the 150-s Set B HAH spectra: RMSEP = 21.5, r^2 = 0.993, and bias = 7.7 mg/dL. The similarity of these results validates (1) the use of a smaller (25-sample) calibration set and (2) the separation of the calibration step from the gathering of HAH spectra to be predicted.

If there is no systematic error (random, normally distributed errors only), then the magnitude of the bias should be statistically indistinguishable from zero. Concentration variance (s^2) can be calculated from RMSEP and bias according to

$$s^2 = (n/n - 1)\cdot(\mathrm{RMSEP}^2 - \mathrm{bias}^2) \qquad (2)$$

where n is the number of samples in the set.[12] The uncertainty of the bias is then

$$u = \pm\ \mathrm{t} \times (s)/n^{1/2} \qquad (3)$$

where t is Student's t value for 95% confidence and

TABLE II. Statistics of glucose predictions of *in vitro* human aqueous humor samples and, for comparison, of 25 AAH samples that were not used to create the 25-AAH PLS model.

Samples	Exposure time[a] (s)	Model	RMSEP (mg/dL)	r^2	Bias (mg/dL)	N for statistics[c]
17 HAH	150	50-AAH	20.4	0.991	5.9	17
25 Set II AAH	150	25 Set I AAH[b]	23.0	0.995	6.3	25
10 Set B HAH	150	25 Set I AAH[b]	21.6	0.994	9.4	10
17 HAH	30	50-AAH	22.0	0.989	5.9	85
25 Set II AAH	30	25 Set I AAH[b]	25.2	0.995	6.3	125
10 Set B HAH	30	25 Set I AAH[b]	22.9	0.992	9.4	50
17 HAH	12	50-AAH	26.0	0.984	6.0	204
25 Set II AAH	12	25 Set I AAH[b]	28.5	0.994	6.7	300
10 Set B HAH	12	25 Set I AAH[b]	26.4	0.988	9.6	120

[a] Equivalent exposure time of the predicted spectra. Both models were created from 150-s calibration spectra.
[b] This model was calculated from the 25 designed-solution (AAH) samples of Set I, the spectra of which were collected 3.6 weeks earlier than the spectra of the 25 Set II AAH and the 10 Set B HAH samples.
[c] N is the number of predictions used to calculate RMSEP, r^2, and bias, as described in the Data Processing subsection.

$n - 1$ degrees of freedom. For the 10 HAH samples predicted by the Set I 25-AAH model, $u = \pm 15$ mg/dL. So the bias of 9.4 mg/dL cannot be statistically distinguished from a zero bias. This result demonstrates that a model can remain fairly stable over a period of weeks.

As further evidence of model stability, the same calculations were performed on predictions of glucose in the 25 Set II AAH samples. In this case, the larger number of samples narrows the uncertainty limits and leads to $u = \pm 9.3$ mg/dL. The observed bias of 6.3 mg/dL is still not distinguishable from zero bias, supporting the idea that bias is negligible.

Noise. Long (150-s) equivalent exposure times and relatively high laser power were used for the AAH and HAH Raman spectra discussed above. The resultant high excitation energy was chosen in an effort to reduce spectral noise as much as possible. If all other noise sources are successfully eliminated, Raman spectra become photon-counting-noise (shot-noise) limited. While spectral intensity is directly proportional to exposure time, t, spectral shot noise (standard deviation of intensity) is proportional to $t^{1/2}$. Normalization of spectra (including by MSC preprocessing) rescales the data by dividing each spectrum by a number roughly proportional to t. The shot noise on the scaled spectra then varies as $1/t^{1/2}$, which means that the variance of intensity is proportional to $1/t$. PLS predicted concentrations are the sum of the scaled spectral intensities multiplied by the calibration vector elements (see Eq. 1), so concentration variance (s^2) is also expected to be roughly proportional to ($1/t$). One can arrive at the same expected time dependence by viewing concentrations predicted based on longer-exposure-time spectra as the average of several concentration predictions based on shorter-exposure-time spectra. Then the concentration variance is the square of the standard error of the mean and thus is proportional to $1/n$.

Standard deviations of glucose concentrations determined (predicted) from the 50-AAH PLS model are plotted in Fig. 6 versus the equivalent exposure times of their spectra (one spectrum per sample for all points except those denoted by ♦). Concentration standard deviations (s) were quantified from replicate measurements of saline solution ''blank'' samples, from replicate measurements of a separate AAH sample (a 173 mg/dL glucose sample was analyzed repeatedly among Set I samples; a 177 mg/dL glucose sample was analyzed repeatedly among Set II samples), or were calculated for same-set HAH samples using Eq. 2. A smoother indication of the dependence of HAH precision on exposure time is shown by the solid diamonds (♦) in Fig. 6 (plotted at 6-s intervals for clarity). These points represent the concentration standard deviations for the set of 17 HAH samples (A and B combined), calculated from as much of the spectral data as possible, resulting in a larger number of predictions from which to calculate the standard deviation (see the Data Processing subsection above). The dashed line in Fig. 6 represents the $1/t^{1/2}$ dependence of the largest standard deviation plotted at 3 s. As can be seen, none of the sample sets' standard deviations decrease at the theoretical rate, although the two sets of AAH replicates appear closer to doing so than the other sets. Even in the theoretical best case, increasing exposure time as a way to decrease noise has diminishing effectiveness at longer times, as shown by the dashed line in Fig. 6. However, the noise of some of the data sets in Fig. 6 seems to become independent of exposure time past a certain point. This can happen when noise that varies randomly on the time scale of the exposure is no longer the major source of uncertainty. Further improvement to the precision of a measurement may still be possible by averaging replicate results obtained over a longer period of time if a more slowly varying but still apparently random noise is significantly influencing results.

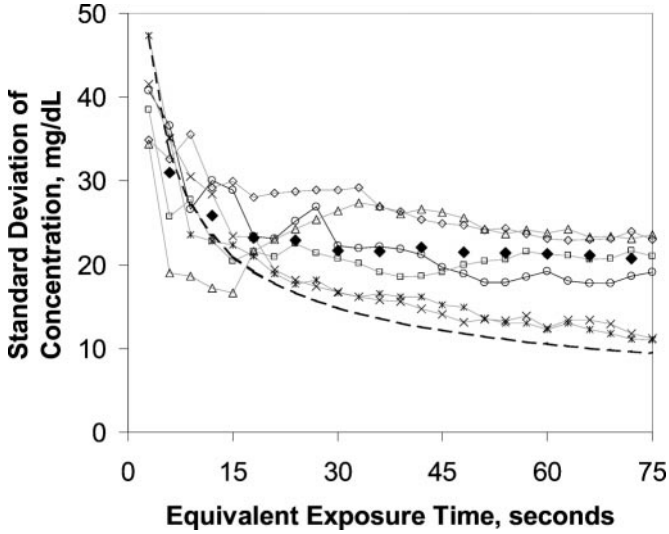

Fig. 6. Standard deviation of glucose concentration predicted using the 50-AAH PLS model (150-s calibration spectra). The standard deviation was calculated from independent replicate measurements of saline solution ''blank'' samples, from replicate measurements of separate AAH samples (eleven aliquots of a 173 mg/dL solution were analyzed among Set I samples; nine aliquots of a 177 mg/dL solution were analyzed among Set II samples), or from Eq. 2 for HAH Sets A and B. Sample standard deviation is plotted versus the equivalent exposure time of the spectra used for glucose prediction: (◇) 7 Set I blanks; (□) 9 Set II blanks; (*) 11 replicate AAH measurements with Set I; (×) 9 replicate AAH measurements with Set II; (○) 7 Set A HAH samples; and (Δ) 10 Set B HAH samples. (♦) Standard deviations calculated using Eq. 2 and N glucose predictions of the 17 HAH samples (see the Data Processing subsection), plotted at 6-s intervals for clarity. (- - -) Standard deviation proportional to ($1/t^{1/2}$).

Further work is needed to identify and reduce the sources of uncertainty in this study, or at least to discover what time scale is needed to collect independent replicates. Sources of uncertainty could include instability in laser-beam intensity, wavelength, pointing, or polarization; vertical and/or horizontal sample positioning issues; formation of small bubbles in the sample; or possibly temperature effects. Reference concentration uncertainty may also contribute but is not likely to be the dominant source of the observed prediction uncertainty since the former should be proportional to concentration, while the latter apparently is not, as can be seen in Fig. 5 for long exposure times.

Reduced Aqueous Humor Laser Exposure. From the leveling off of the curves in Fig. 6, one might conclude that 30-s HAH spectra should give results similar to 150-s HAH spectra. To test this, 30-s spectra of the 17 HAH samples were predicted by the 50-AAH model, yielding the results listed in the fourth row of Table II. In addition, 30-s spectra of the 10 Set B HAH samples

were predicted by the Set I 25-AAH model. These results appear in the sixth row of Table II. To test the effect of further reducing exposure time, 12-s spectra were similarly predicted and appear in the last rows of Table II. The number of predictions, N, from which RMSEP, r^2, and bias were calculated is listed in the last column of each row.

A comparison between the 150-s and 30-s results reveals that they are essentially the same, confirming the hypothesis (based on Fig. 6) that 30 s of exposure time would do as well as 150 s of exposure time for HAH glucose determination in this study. Further reduction of exposure time, to 12 s, results in somewhat larger determination errors. This may be due to the dwindling signal-to-noise ratio of the 12-s sample spectra beginning to overshadow the uncertainty contributed by the calibration.[17]

Whereas the prediction errors of the 12-s glucose determinations may be slightly larger than desirable for clinical work,[18] they could be reduced to acceptable levels by averaging two such measurements. Because of divergence within the human eye of the laser beam from our Raman instrument, the irradiance of a 12-s exposure (1200 mJ) on the retina would be below the International Commission on Non-Ionizing Radiation Protection (ICNIRP) and the American National Standards Institute (ANSI) thresholds for retinal toxicity if delivered over more than 40 s. So, for example, two doses of laser light, 25 mW for 48 s each, should give *in vivo* aqueous humor spectra capable of producing a clinically useful glucose determination. In practice, multiple shorter duration exposures would probably work better, since obviously unusable spectra, such as those collected during blinking, could be discarded before processing.

Conclusions. Using 785 nm, ~15 J Raman spectra, PLS models created from AAH calibration solutions containing experimentally designed levels of glucose, bicarbonate, lactate, urea, and ascorbate were shown to predict glucose levels of *in vitro* human aqueous humor (HAH) with $r^2 > 0.99$ relative to reference concentrations measured with a commercial glucose-oxidase analyzer. Such a high laser energy level was employed to obtain calibration spectra with good signal-to-noise ratios. We explored how reduction in the laser energy used for collecting HAH spectra affected glucose prediction. Table II summarizes the statistics found for HAH glucose predictions, along with AAH predictions for comparison. The last row of each exposure group represents the prediction of glucose in ten *in vitro* HAH samples using a PLS model created from 25 ''artificial aqueous humor'' calibration solutions, the spectra of which had been collected 3.6 weeks earlier. The results indicate that the PLS calibration was stable for several weeks and that the components of *in vitro* HAH not represented in the calibration solutions did not appear to cause interference with the determination of glucose levels above ~50 mg/dL. Though there remain a number of challenges to solve before this approach can be applied *in vivo*, the results of this study support the idea that a calibration model based on high signal-to-noise spectra may successfully predict appropriately gathered low-laser-power spectra to make Raman spectroscopic determination of glucose in human eyes feasible.

ACKNOWLEDGMENTS

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, and at Childrens Hospital Los Angeles, under a contract with the National Aeronautics and Space Administration (NASA).

1. National Institute of Diabetes and Digestive and Kidney Disorders, *The Diabetes Control and Complications Trial, NIH Publication No. 97–3874* (National Institutes of Health, Bethesda, MD, 1994).
2. W. F. Schrader, R. S. Grajewski, P. Meuer, and W. Kiefer, Investigative Ophthalmology & Visual Science Supplement **41,** S77 (2000).
3. S. Pohjola, Acta Opthalmol. Suppl. **88,** 11 (1966).
4. M. S. Borchert, M. C. Storrie-Lombardi, and J. L. Lambert, Diabetes Technol. Therapeutics **1,** 145 (1999).
5. B. Rabinovitch, W. F. March, and R. L. Adams, Diabetes Care **5,** 254 (1982).
6. P. G. Steffes, Diabetes Technol. Therapeutics **1,** 129 (1999).
7. B. D. Cameron, J. S. Baba, and G. L. Coté, Diabetes Technol. Therapeutics **3,** 201 (2001).
8. W. F. March, B. Rabinovitch, and R. L. Adams, Diabetes Care **5,** 259 (1982).
9. C. Chou, C.-Y. Han, W.-C. Kuo, Y.-C. Huang, C.-M. Feng, and J.-C. Shyu, Appl. Opt. **37,** 3553 (1998).
10. R. J. McNichols and G. L. Coté, J. Biomed. Opt. **5,** 5 (2000).
11. G. L. Coté, R. M. Lec, and M. V. Pishko, IEEE-Sens. J. **3,** 251 (2003).
12. H. Martens and T. Naes, *Multivariate Calibration* (John Wiley and Sons, New York, 1989).
13. J. Lambert, M. Storrie-Lombardi, and M. Borchert, IEEE Lasers & Electro-Optics Society Newsletter **4,** 19 (1998).
14. J. L. Lambert, J. M. Morookian, S. J. Sirk, and M. S. Borchert, J. Raman Spectrosc. **33,** 524 (2002).
15. A. J. Berger, Y. Wang, and M. S. Feld, Appl. Opt. **35,** 209 (1996).
16. M. J. Goetz, Jr., G. L. Coté, R. Erckens, W. March, and M. Motamedi, IEEE Trans. Biomed. Eng. **42,** 728 (1995).
17. A. J. Berger and M. S. Feld, Appl. Spectrosc. **51,** 725 (1997).
18. J. L. Lambert, C. C. Pelletier, and M. Borchert, ''Glucose Determination in Human Aqueous Humor with Raman Spectroscopy'', J. Biomed. Opt. **10,** May/June (2005), paper in press.
19. R. J. Erckens, M. Motamedi, W. F. March, and J. P. Wicksted, J. Raman Spectrosc. **28,** 293 (1997).
20. R. G. Brereton, *Chemometrics, Data Analysis for the Laboratory and Chemical Plant* (John Wiley and Sons, Chichester, 2003), Sect. 2.3.4.
21. M. L. Sears, ''Dynamics of Ocular Fluids and Control of Intraocular Pressure: Formation of Aqueous Humor,'' in *Principles and Practice of Ophthalmology*, D. M. Albert, F. A. Jakobiec, and N. L. Robinson, Eds. (W.B. Saunders, Philadelphia, 1994).
22. *Harrison's Principles of Internal Medicine*, E. Braunwald, A. S. Fauci, D. L. Casper, S. L. Hauser, D. L. Longo, and J. L. Jameson, Eds. (McGraw Hill, New York, 2002), 15th ed.
23. J. Veselovsky, Z. Olah, A. Vesela, and S. Gressnerova, Bratisl. Lek. Listy **104,** 14 (2003).
24. D. M. Haaland, M. R. Robinson, G. W. Koepp, E. V. Thomas, and R. P. Eaton, Appl. Spectrosc. **46,** 1575 (1992).
25. CRC Handbook of Chemistry and Physics, R. C. Weast, Ed. (The Chemical Rubber Co., Cleveland, 1971), 52nd ed., pp. D-191, D-213, and F-5.
26. S. Y. Wang, C. E. Hasty, P. A. Watson, J. P. Wicksted, R. D. Stith, and W. F. March, Appl. Opt. **32,** 925 (1993).
27. I. Koshiishi and T. Imanari, Anal. Chem. **69,** 216 (1997).