Blood analysis by Raman spectroscopy

Annika M. K. Enejder, Tae-Woong Koo, Jeankun Oh, Martin Hunter, Slobodan Sasic, and Michael S. Feld

George R. Harrison Spectroscopy Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Gary L. Horowitz

Department of Pathology, Beth Israel Deaconess Medical Center, Boston, Massachusetts 02215

Received June 12, 2002

Concentrations of multiple analytes were simultaneously measured in whole blood with clinical accuracy, without sample processing, using near-infrared Raman spectroscopy. Spectra were acquired with an instrument employing nonimaging optics, designed using Monte Carlo simulations of the influence of light-scattering–absorbing blood cells on the excitation and emission of Raman light in turbid medium. Raman spectra were collected from whole blood drawn from 31 individuals. Quantitative predictions of glucose, urea, total protein, albumin, triglycerides, hematocrit, and hemoglobin were made by means of partial least-squares (PLS) analysis with clinically relevant precision (r^2 values >0.93). The similarity of the features of the PLS calibration spectra to those of the respective analyte spectra illustrates that the predictions are based on molecular information carried by the Raman light. This demonstrates the feasibility of using Raman spectroscopy for quantitative measurements of biomolecular contents in highly light-scattering and absorbing media.

OCIS codes: 170.5660, 170.1470, 170.4580, 170.1610, 170.7050, 170.1580.

Blood analysis is frequently performed for medical diagnosis. Approximately 600 million cholesterol tests are performed annually worldwide to assess the risk and severity of cardiovascular disease. More than 100 million diabetics are advised to monitor their glucose levels several times each day. There is a great interest in optical measurement that would permit simultaneous analysis of multiple components (analytes) in whole blood without the need for conventional sample processing, such as centrifuging and adding reagents. Several techniques for measuring analytes in biological media are being studied.[1,2] The major challenge in analysis of whole blood samples lies in the presence of numerous low-concentration components, all with weak signals that are further distorted by the strong light absorption and scattering caused by the red blood cells. Raman spectroscopy, by generating a distinct spectrum for each analyte, can resolve the individual components of this complex mixture. Its potential in studies of biological media was previously shown, primarily with the objective of the characterizing tissue rather than obtaining quantitative information about the constituents.[3–6] In previous research, Berger *et al.* accurately predicted concentrations of analytes in blood serum. However, correspondingly good results were not achieved in whole blood.[7] Here we report what we believe to be the first quantitative observation of multiple analytes in whole blood at clinically relevant precision, made using a new highly sensitive Raman instrument optimized for turbid samples. The analytes quantified are glucose, urea, cholesterol, albumin, total protein, triglycerides, hematocrit (hct), hemoglobin, and bilirubin, all of which are frequently ordered diagnostic tests used in connection with common medical conditions.[8]

Our experiment employed a specially designed instrument optimized for collecting near-infrared Raman spectra from whole blood (Fig. 1). A Monte Carlo model for the propagation of excitation and Raman-scattered light was used for system design. It employed three steps: (i) propagation of a photon in a medium specified by its absorption and scattering coefficients (μ_a, μ_s, and g) at the excitation wavelength,[9] (ii) isotropic launch of a Raman photon with a probability depending on the Raman cross section of the analyte being studied, followed by (iii) propagation of the Raman photon in the sample with optical properties at the Raman wavelength(s). This information enabled us to estimate the size and brightness of the effective Raman source created in the sample by the

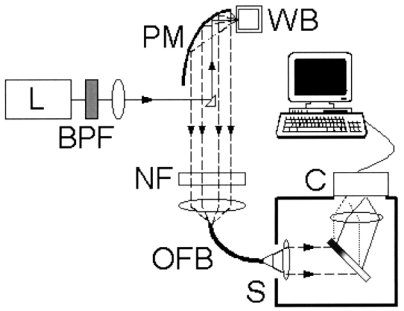

Fig. 1. Schematic diagram of the Raman instrument. A beam of 830-nm light from a diode laser (L; PI-ECL-830-500, Process Instruments, Salt Lake City, Utah) is passed through a bandpass filter (BPF; Kaiser Optical Systems, Ann Arbor, Mich.), directed toward a paraboloidal mirror (Perkin Elmer, Fremont, Calif.) by means of a small prism, and focused onto a quartz cuvette containing a whole blood sample (WB). Raman-scattered light emitted from the whole blood surface (1-mm^2 area) is collected by the mirror, passed through a notch filter (NF; Kaiser Optical Systems) to reject backreflected 830-nm light, and coupled into an optical fiber bundle (OFB; Romack Fiber Optics, Williamsburg, Va.), which converts the circular shape of the collected light to rectangular to match the entrance slit of the spectrograph (S, Holospec $f/1.8i$; Kaiser Optical Systems). The spectra are collected by a cooled CCD array detector (C, 1152 × 770 pixels, Roper Scientific, Trenton, N.J.) and binned along the vertical direction, resulting in a 1152-pixel spectrum.

0146-9592/02/222004-03$15.00/0

excitation light under the influence of turbidity. The etendue of the instrument (area–solid angle product) was determined by the spectrograph–CCD detector. We then optimized the trade-off between the area and the solid angle of the collection optics to maximize the collected Raman signal. Raman signals calculated for various collection areas at three solid angles are shown in Fig. 2. The combination of a collection half-angle of ~35° and a radius of ~0.5 mm maximizes the signal. This collection geometry was achieved by means of a nonimaging optical element, a gold-coated, half-paraboloidal mirror, with specifications of which (f = 15.9 mm, collection half-angles 46°/30° resulting in an effective N.A. of 0.5) were determined by an optical design code (Zemax, Focus Software, Tucson, Ariz.). The mirror collected approximately 25% of the total Raman-scattered light from the whole blood surface, approximately four times more efficiently than conventional lens-type optics used in a previous system.[7]

Whole blood samples for routine clinical diagnosis were collected from 31 patients. For each sample, 30 consecutive 10-s spectra were collected over a 5-min period. Conventional clinical laboratory methods, including absorbance spectrophotometry and automated cell counting, were used to assess the nine analyte concentrations. These reference concentrations were correlated with the recorded Raman spectra and used for multivariate calibration and validation.

The background subtracted (low-order polynomial fit) Raman spectra (Fig. 3) were analyzed over the range 650–1650 cm^{-1} by use of partial least squares (PLS),[11] with leave-one-sample-out cross validation.[12] Depending on the analyte, 6–21 loading vectors were used. We chose cross validation to utilize fully the small data set for both calibration and validation. A larger sample size would allow external validation, providing a more robust calibration model. The accuracy of PLS was determined from the difference of and the correlation between the predicted and reference concentrations, evaluated in terms of the root mean square error of prediction (RMSEP) and the squared correlation coefficient (r^2). Table 1 lists the analysis results for PLS cross validation of the whole blood data set collected for 5 min. Accurate predictions were made for seven diagnostic tests parameters. Although similar PLS analysis was attempted, no prediction was possible for bilirubin. A typical concentration of bilirubin is at least an order of magnitude lower than that of the other analytes studied, and its weak signal appears difficult to resolve. All test parameters show strong correlation between the predicted and the reference concentrations (Fig. 4), with r^2 values of 0.93 or higher except for total cholesterol (r^2 = 0.66). Generally, r^2 values higher than 0.9 indicate that the method under investigation is clinically accurate.[13] Based on this criterion, our method measured seven diagnostic test parameters with clinical accuracy. It appears that the reduced correlation of total cholesterol is due to the fact that the reference measurements in plasma do not accurately reflect the total cholesterol concentration in whole blood, since approximately 40% of the cholesterol in whole blood resides in the cell membranes.[14] Improved predictions (RMSEP, 20 mg/dL, r^2 = 0.74) were obtained with an approximate correction based on the measured hct value, which indicates that even better correlation is expected with accurate reference values of whole blood cholesterol. For 1-min collection time, prediction errors for total protein, albumin, total cholesterol, triglyceride, hemoglobin, and hct were only 30% higher than those for 5 min. However, the improvement in RMSEP with 5-min data integration was more significant for glucose and urea (by factors of 4 and 2, respectively), whose concentrations are lower than those of the other analytes measured.

In addition to demonstrating the clinical accuracy of our predictions, we can use physical information to interpret the calibration spectra determined by PLS. The calibration spectra are equivalent to the

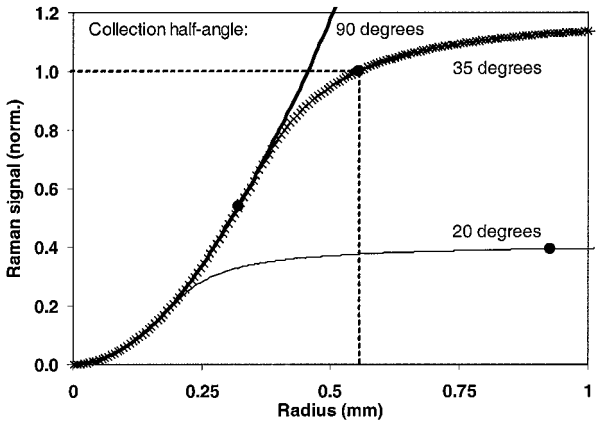

Fig. 2. Raman signals from whole blood at collection half-angles 20°, 35°, and 90°, obtained from Monte Carlo simulations: For hct, 45%; μ_a = 1.25 mm^{-1}, μ_s = 253 mm^{-1}, and g = 0.991 at the 830-nm excitation wavelength and μ_a = 1.46 mm^{-1}, μ_s = 222 mm^{-1}, and g = 0.989 (Ref. 10) at 920 nm, a representative Raman emission wavelength. The resulting Raman intensities at the collection radii matching the etendue of the spectrograph–CCD detector are indicated. The largest signal is obtained for 35° half-angle and a 0.55-mm radius.

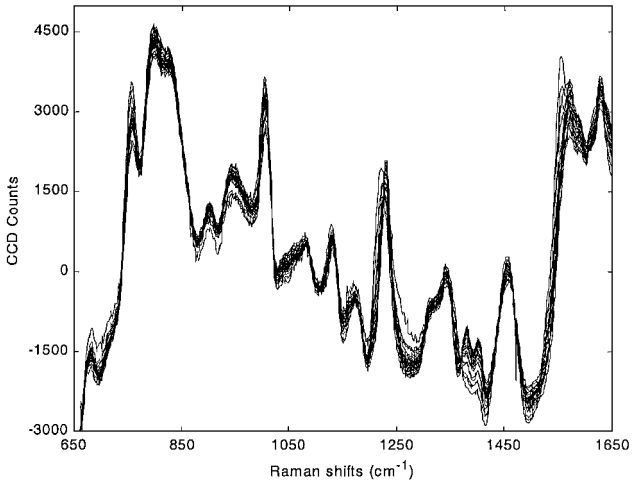

Fig. 3. Raman spectra of 31 whole blood samples after polynomial background subtraction.

Table 1. PLS Predictions

Analyte	Error: Prediction	Error: Ref.	Physiological Range	r^2
Glucose (mg/dL)	21	3	45–180	0.97
Urea (mg/dL)	4.9	0.9	17–50	0.94
Cholesterol (mg/dL)	30	4	150–250	0.66
Triglycerides (mg/dL)	38	3	10–190	0.92
Total Protein (g/dL)	0.31	0.1	6–8.3	0.94
Albumin (g/dL)	0.11	0.09	3.2–4.5	0.98
Hemoglobin (g/dL)	0.66	0.17	14–17.5	0.94
Hematocrit (%)	1.7	0.4	35.9–50.4	0.94
Bilirubin	N/A	0.02	0.1–1.2	N/A

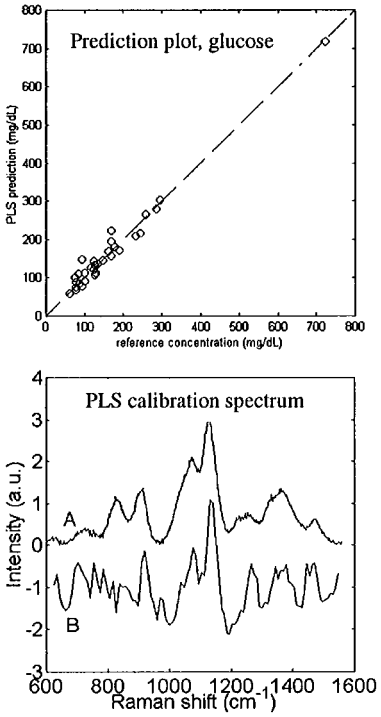

Fig. 4. Prediction plot (top) for glucose and (bottom) the corresponding PLS calibration spectrum. Good agreement is obtained between the latter (curve B, offset) and the Raman spectrum of pure glucose in water (curve A).

regression coefficients that are generated. Each calibration spectrum yields the concentration of its corresponding analyte when it is projected onto a background-subtracted Raman spectrum. Figure 4 compares the Raman spectrum of glucose with the corresponding PLS calibration spectrum. As can be seen, the promient Raman features near 1070 and 1130 cm^{-1} that are characteristic of glucose are well represented. The similarity of spectral features indicates that our predictions are indeed based on molecular information provided by the Raman spectra.

The successful measurement of multiple analytes in whole blood demonstrates that Raman spectroscopy can be used to extract quantitative information about biomolecular contents with good accuracy in an optically complex medium with strong and varying light-scattering and absorption properties by means of the linear PLS algorithm, despite the generally nonlinear influence of optical properties, primarily due to variations in hct. Monte Carlo simulations indicate that in the hct range 30–50% the increased absorption and scattering accompanying an increase in hct give rise to an approximately linear decrease in the backscattered Raman light. Further improvement in prediction accuracy may be obtained by correction for variations in scattering and absorption. A method of extracting the intrinsic Raman signal is currently under development in our laboratory.

The method and instrumentation described here demonstrate the capability of Raman spectroscopy for quantitative measurements in highly light-scattering and absorbing media as well as its potential clinical utility.

This work was performed at the Massachusetts Institute of Technology, Laser Biomedical Research Center and supported by National Institutes of Health, National Center for Research Resources (grant P41-RR02594) and the Bayer Corporation. A. Enejder (e-mail: enejder@mit.edu) acknowledges the support of the Swedish Research Council.

References

1. O. S. Khalil, Clin. Chem. **45,** 165 (1999).
2. R. J. McNichols and G. L. Cote, J. Biomed. Opt. **5,** 5 (2000).
3. G. J. Puppels, F. F. M. de Mul, C. Otto, J. Greve, M. Robert-Nicoud, D. J. Arndt-Jovin, and T. M. Jovin, Nature **347,** 301 (1990).
4. M. G. Shim, L. M. W. M. Song, N. E. Marcon, and B. C. Wilson, Photochem. Photobiol. **72,** 146 (2000).
5. H. P. Buschman, G. Deinum, J. T. Motz, M. Fitzmaurice, J. R. Kramer, A. van der Laarse, A. V. Bruschke, and M. S. Feld, Cardiovasc. Pathol. **10,** 69 (2001).
6. K. E. Shafer-Peltier, A. S. Haka, M. Fitzmaurice, J. Crowe, J. Myles, R. R. Dasari, and M. S. Feld, J. Raman. Spectrosc. **33,** 552 (2002).
7. A. J. Berger, T.-W. Koo, I. Itzkan, G. L. Horowitz, and M. S. Feld, Appl. Opt. **38,** 2916 (1999).
8. Although hemoglobin concentration and hct are two different analytes, they are highly correlated, as observed in our study ($r^2 = 0.97$).
9. L.-H. Wang, S. L. Jacques, and L.-Q. Zheng, Comput. Methods Programs Biomed. **47,** 131 (1995).
10. A. Roggan, M. Friebel, K. Dörschel, A. Hahn, and G. Müller, J. Biomed. Opt. **4,** 36 (1999).
11. D. M. Haaland and E. V. Thomas, Anal. Chem. **60,** 1193 (1988).
12. H. Martens and T. Naes, *Multivariate Calibration* (Wiley, New York, 1989).
13. J. B. Henry, *Clinical Diagnosis and Management by Laboratory Methods* (Saunders, Philadelphia, Pa., 1996).
14. B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell* (Garland, New York, 1989).