Multivariate Reference Technique for Quantitative Analysis of Fiber-Optic Tissue Raman Spectroscopy

ARTICLE *in* ANALYTICAL CHEMISTRY · OCTOBER 2013

Impact Factor: 5.64 · DOI: 10.1021/ac402059v · Source: PubMed

4 AUTHORS:

Mads Bergholt
Imperial College London
37 PUBLICATIONS **436** CITATIONS
SEE PROFILE

Shiyamala Duraipandian
National University of Singapore
11 PUBLICATIONS **131** CITATIONS
SEE PROFILE

Wei Zheng
National University of Singapore
160 PUBLICATIONS **1,826** CITATIONS
SEE PROFILE

Zhiwei Huang
National University of Singapore
174 PUBLICATIONS **2,448** CITATIONS
SEE PROFILE

analytical chemistry

Article

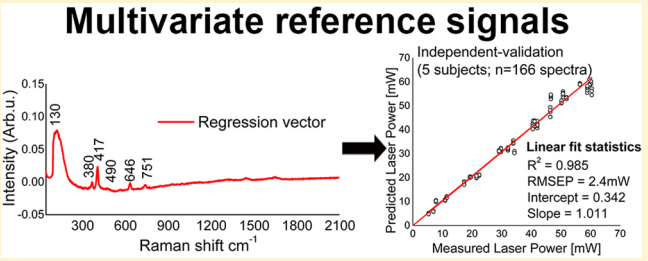
pubs.acs.org/ac

Multivariate Reference Technique for Quantitative Analysis of Fiber-Optic Tissue Raman Spectroscopy

Mads Sylvest Bergholt, Shiyamala Duraipandian, Wei Zheng, and Zhiwei Huang*

Optical Bioimaging Laboratory, Department of Biomedical Engineering, Faculty of Engineering, National University of Singapore, Singapore 117576

ABSTRACT: We report a novel method making use of multivariate reference signals of fused silica and sapphire Raman signals generated from a ball-lens fiber-optic Raman probe for quantitative analysis of in vivo tissue Raman measurements in real time. Partial least-squares (PLS) regression modeling is applied to extract the characteristic internal reference Raman signals (e.g., shoulder of the prominent fused silica boson peak (~130 cm^{-1}); distinct sapphire ball-lens peaks (380, 417, 646, and 751 cm^{-1})) from the ball-lens fiber-optic Raman probe for quantitative analysis of fiber-optic Raman spectroscopy. To evaluate the analytical value of this novel multivariate reference technique, a rapid Raman spectroscopy system coupled with a ball-lens fiber-optic Raman probe is used for in vivo oral tissue Raman measurements (n = 25 subjects) under 785 nm laser excitation powers ranging from 5 to 65 mW. An accurate linear relationship (R^2 = 0.981) with a root-mean-square error of cross validation (RMSECV) of 2.5 mW can be obtained for predicting the laser excitation power changes based on a leave-one-subject-out cross-validation, which is superior to the normal univariate reference method (RMSE = 6.2 mW). A root-mean-square error of prediction (RMSEP) of 2.4 mW (R^2 = 0.985) can also be achieved for laser power prediction in real time when we applied the multivariate method independently on the five new subjects (n = 166 spectra). We further apply the multivariate reference technique for quantitative analysis of gelatin tissue phantoms that gives rise to an RMSEP of ~2.0% (R^2 = 0.998) independent of laser excitation power variations. This work demonstrates that multivariate reference technique can be advantageously used to monitor and correct the variations of laser excitation power and fiber coupling efficiency in situ for standardizing the tissue Raman intensity to realize quantitative analysis of tissue Raman measurements in vivo, which is particularly appealing in challenging Raman endoscopic applications.



Near-infrared (NIR) Raman spectroscopy is a molecular specific vibrational analytical technique that can provide quantitative and qualitative information about compositions of samples.[1,2] This technique has excelled in analytical sciences including pharmaceutics,[3] forensics,[4] and biomedical applications.[1,2,5−12] With recent technical advances in fiber-optic Raman probe designs, NIR Raman spectroscopy has been applied for in vivo tissue diagnosis and characterization in a variety of organs (e.g., bladder, nasopharynx and larynx, cervix, lung, esophagus, stomach, etc.).[5−12] While much effort has been devoted to the multivariate analysis of tissue Raman spectra in biomedicine, only very limited work has been reported on the quantitative analysis of tissue compositions by Raman spectroscopy.[13−15] Since tissue pathogenesis is associated with prominent changes in biomolecular and biochemical compositions, the development of quantitative Raman spectroscopy has great potential to significantly improve tissue characterization and diagnosis. However, quantitative fiber-optic tissue Raman spectroscopy is challenging and requires careful control and monitoring of experimental parameters, including laser excitation power, fiber coupling efficiency, measurement reproducibility, etc. The Raman intensity of the sample is linearly proportional to the laser excitation power used and the fiber coupling efficiency of the system, which also remain the most common sources of variances in fiber-optic quantitative analysis.[16,17] An introduction of reference signals is therefore an important measure in standardizing the absolute Raman peak intensities in Raman spectroscopy. The internal reference Raman signals based on univariate references (i.e., single Raman peaks) have been reported in fiber-optic Raman applications.[13,14,16,17] For example, Morris and Okagbare[13,14] reported the use of fluorocarbon fibers or a fluorinated ethylene propylene copolymer cap on the excitation fiber of the fiber-optic Raman probes to generate reference Raman signals (e.g., C−F stretch at 692 cm^{-1} or 732 cm^{-1}) proportional to delivered laser power for quantitative analysis of tissue Raman spectroscopy. Self-referencing fiber-optic Raman probes with embedded diamond in the fiber-tip or distal lens have been studied for quantitative analysis and monitoring of laser excitation power delivered to transparent chemical samples.[16,17] Although inserting additional reference materials or impurities into the fiber-optic Raman probe produces well-defined reference Raman peaks, these methods could lead to undesirable

Received: July 5, 2013
Accepted: October 25, 2013
Published: October 25, 2013

interference with characteristic tissue Raman peaks in the spectral region of interest.[1,2,5−12] A more intuitive approach utilizing the elastic scattered laser light (i.e., Rayleigh light) as an internal reference has been proposed for quantitative analysis of simple transparent chemical solutions.[18] But for fiber-optic biomedical applications, the intuitive method may have difficulty in keeping the Rayleigh light detected within the dynamic range of the charge-coupled device (CCD) camera. The situation in fiber-optic tissue Raman spectroscopy for quantitative analysis of biomedical tissue in vitro or in vivo is even more complicated due to several confounding factors: (i) Tissue spectra contain overlapped Raman and autofluorescence signals with prominent variance that could obscure the internal reference signals. (ii) The induced fiber probe background (i.e., fluorescence and Raman scattering) could be related to the varying tissue optical properties (e.g., refractive index, absorption, and scattering) which influence the intensity of the elastically scattered laser light back into the collection fibers from the tissue sample, thereby resulting in varying reference signals. (iii) Quantitative analysis for in vivo biomedical endoscopic applications is further complicated because the bending of fibers can induce the change of laser delivery power and collection efficiency that is difficult to monitor in situ. This is particularly of concern in some endoscopic screening situations (e.g., endoscopic retroflex, which requires a sharp bending of fiber probes for the inspection of the cardia anatomical region in the stomach). (iv) Fiber coupling efficiency might also vary in clinical settings. To tackle the above problems, in this work, we report a novel multivariate reference technique making use of the multivariate nature of fused silica and sapphire Raman signals generated from a ball-lens fiber-optic Raman probe as an internal reference for quantitative fiber-optic tissue Raman measurements. To evaluate the utility of the multivariate reference technique, a rapid Raman spectroscopy system coupled with a ball-lens fiber-optic Raman probe[19] is used for in vivo oral tissue Raman measurements under different 785 nm laser excitation powers. A partial least-squares (PLS) regression is applied to uncover the characteristic Raman signals of fused silica and sapphire ball lens from complex tissue Raman spectra. We prove that the multivariate reference technique developed is much more accurate and robust than the univariate reference method for quantitative tissue Raman measurements. The accurate and quantitative analysis of gelatin tissue phantoms independent of laser excitation power variations is also demonstrated using the multivariate reference technique.

MATERIALS AND METHODS

Raman Instrumentation. A rapid fiber-optic Raman spectroscopy technique was used for in vivo tissue Raman measurements.[20,21] Briefly, the Raman spectroscopy system consists of a spectrum stabilized 785 nm diode laser (maximum output: 300 mW, B&W TEK Inc., Newark, Delaware) and a transmissive imaging spectrograph (Holospec f/1.8, Kaiser Optical Systems Inc., Ann Arbor, Michigan) equipped with a liquid nitrogen-cooled, back-illuminated and deep depletion CCD camera (1340 × 400 pixels at 20 × 20 μm per pixel; Spec-10:400BR/LN, Princeton Instruments, Trenton, New Jersey). The system also consists of a specially designed fused silica fiber-optic Raman endoscopic probe (1.8 mm in outer diameter and 1.30 m in length) that comprises 9 × 200 μm collection fibers (N.A. = 0.22) surrounding the central light delivery fiber (200 μm in diameter, N.A. = 0.22).[19] A 1.0 mm sapphire ball lens (refractive index 1.76) is coupled to the fiber tip of the Raman probe for enhancing epithelial tissue Raman measurements.[19] The system acquires Raman spectra over the range of 800−1800 cm^{-1} with spectral resolution of 9 cm^{-1}. Each Raman spectrum in this study was measured with an integration time of 0.5 s under 785 nm laser excitation. The rapid Raman spectroscopy technique was wavelength calibrated using an argon/mercury spectral lamp (AR-1 and HG-1, Ocean Optics Inc., Dunedin, Florida). All wavelength-calibrated spectra were corrected for the intensity response of the system using a tungsten-halogen calibration lamp (RS-10, EG&G Gamma Scientific, San Diego, CA).

Data Preprocessing. Customized software was developed to control the Raman spectroscopy system for real time data acquisition and analysis.[22] The raw Raman spectra measured from in vivo tissue represent a combination of weak tissue Raman signal, intense autofluorescence background, and noise. The raw spectra are preprocessed by a first-order Savitzky−Golay smoothing filter (a window width of three pixels selected to match the spectral resolution) to reduce the spectral noise. In the fingerprint region (800−1800 cm^{-1}), a fifth-order polynomial was found to be optimal for fitting the autofluorescence background in the noise-smoothed spectrum, and this polynomial is then subtracted from the raw spectrum to yield the tissue Raman spectrum alone. All the aforementioned preprocessing is completed within 100 ms, and the processed results can be displayed on the computer screen in real time.[22]

Partial Least-Squares Regression. We employ the partial least-squares (PLS) regression as a multivariate method to extract characteristic internal reference background signals from the fiber-optic Raman probe. Since the laser excitation power and the fiber coupling efficiency have a similar effect on the change of the Raman intensity of the sample in fiber-optic Raman spectroscopy, we fix the fiber coupling but use the varying laser power as a dependent variable for PLS modeling. Briefly, PLS utilizes the fundamental principle of principal component analysis (PCA) but further rotates the components (latent variables (LVs)) by maximizing the covariance between the spectral variation and the dependent variable (e.g., laser excitation power), so that the LV loadings explain the relevant variations rather than the most prominent variations in the spectral data set.[23−25] This ensures that the important spectral reference variations related to the laser excitation power are retained in the first few LVs. In this study, mean-centering was performed before modeling to reduce the complexity of the PLS regression model. The optimal complexity of the PLS regression model was determined through leave-one-subject-out cross-validation, and the performance of the PLS regression model was examined by calculating the coefficient of determination (R^2), root-mean-square error of calibration (RMSEC), root-mean-square error of cross validation (RMSECV), and root-mean-square error of prediction (RMSEP). Note that an optimal PLS model has a high R^2 but with a low RMSEC, RMSECV, and RMSEP. The PLS regression model developed for resolving the internal reference signals in this study was also implemented as an online laser excitation power predictor in our real-time clinical Raman software and tested prospectively in an unbiased manner.[22] Multivariate statistical analysis was conducted in the Matlab (Mathworks Inc., Natick, MA) programming environment.

Subjects. A total of 30 normal healthy subjects (16 female and 14 males) were recruited for in vivo tissue Raman

measurements in the oral cavity. Informed consent forms were obtained from all participating subjects. Prior to in vivo tissue Raman spectroscopy measurements, all subjects underwent extensive mouthwash to reduce confounding factors (e.g., food debris, microbial coatings etc.). To develop an internal reference, we measured in vivo tissue Raman spectra (n = 783) of the inner lip of 25 subjects. For the 25 subjects, in vivo oral tissue Raman spectra (n = ~5) were acquired at six power levels in the range of 5−65 mW (intervals of ~10 mW). Before each tissue Raman measurement, the laser excitation power level was measured at the distal tip of the fiber-optic probe using a power meter with a linearity of ±0.5% and accuracy of ±3% (range of 0.1 to 100 mW). Other confounding factors (e.g., probe pressure rendering on the tissue surface, photobleaching, tissue optical properties, and bending of the fiber optic probe) were not monitored purposely but modeled by the PLS regression. After deployment of the developed PLS model in the online Raman acquisition framework, the prospective and independent validation of the multivariate internal reference signals for laser excitation power monitoring was performed on the five new subjects (n = 166 spectra) in real time.

Tissue Phantoms. To further validate the quantitative value of the multivariate internal reference method developed in this work, we also conducted tissue phantom experiments. Tissue phantoms of various gelatin concentrations were prepared using bovine skin, Type-B gelatin (G9391, Sigma, USA). The gelatin was dissolved in predefined concentrations (20, 25, 30, 35, 40, 45, and 50% by weight) in distilled H_2O. The dissolved gelatin was heated to 50 °C for 1 h in a water bath with continuous stirring. Subsequently, the molten gelatin was poured into a prechilled mold (4 °C) and stored for 2−3 h to produce solid gelatin phantoms. Quantitative fiber-optic Raman spectroscopic analysis of the tissue phantoms was then conducted. A total of 133 Raman spectra were measured from the various tissue phantoms using the fiber-optic Raman probe under different laser powers. The laser excitation powers were changed in the range 10−60 mW, and the measured spectra were normalized to the internal PLS reference.

RESULTS

Figure 1 shows the background spectrum of a ball-lens fiber-optic Raman probe used when excited by a 785 nm diode laser. The prominent sapphire (Al_2O_3) Raman peaks originating from the distal ball lens can be found at 417 and 646 cm^{-1} (phonon mode with A_{1g} symmetry) and 380 and 751 cm^{-1} (E_g phonon mode).[25] There are two dominant Raman components from the fused silica fiber as well as fiber autofluorescence background likely caused by metal ion impurities in the fused silica fibers. The sharp "defect peaks" of fused silica denoted as D_1 and D_2 at 490 and 606 cm^{-1} have been assigned to breathing vibrations of oxygen atoms in four- and three-membered rings, respectively.[26] The shoulder (~130 cm^{-1}) of an intense boson Raman band related to the general feature of amorphous silica substances[26,27] is also observed from the background spectrum of the fiber-optic Raman probe. The silica boson band is peaking near ~60 cm^{-1}, but only the shoulder is apparent due to the optical filterings of our Raman probe design.[19,21] These characteristic background Raman peaks (shorter than fingerprint region (800−1800 cm^{-1})) from the fiber-optic Raman probe itself could serve as an internal reference for in vivo tissue Raman measurements as demonstrated in the following.

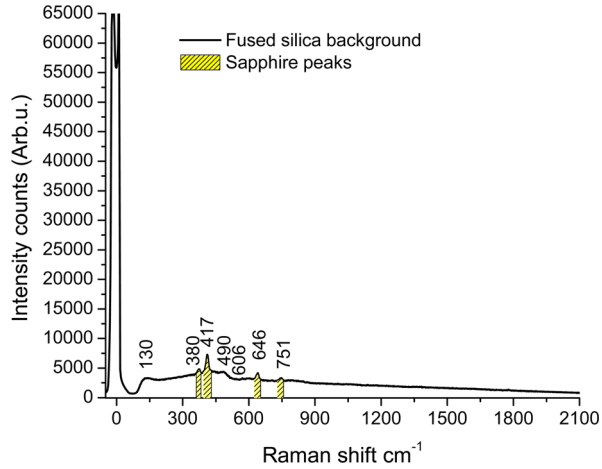


Figure 1. Background Raman spectrum of a ball-lens fiber-optic Raman probe containing the characteristic Raman spectra and broadband fiber autofluorescence of fused silica as well as distinct sapphire Raman signals originating from the distal ball lens when excited by a 785 nm laser with an excitation power of 10 mW.

We measured in vivo Raman spectra of 25 subjects in the oral cavity with the laser excitation power as a dependent parameter. For each subject, in vivo tissue Raman spectra (n = ~5) were acquired with different power levels in the range of 5−65 mW (intervals of ~10 mW). Figure 2a shows an example of the mean in vivo raw Raman spectra ± 1 standard deviation (SD) measured from the inner lip using different laser excitation powers (e.g., 10, 30, and 60 mW). The weak tissue Raman signals superimposed on the varying broad autofluorescence background can be observed. Figure 2b shows the calibrated background-subtracted mean Raman spectra ± 1 SD. Similar to the literature report,[28] the in vivo Raman spectrum of the inner lip shows Raman peaks at around 853 cm^{-1} (ν(C−C)), 1004 cm^{-1} (ν_s(C−C)), 1245 cm^{-1} (amide III ν(C−N) and δ(N−H) of proteins), 1302 cm^{-1} (CH_3CH_2 twisting and wagging), 1443 cm^{-1} ($\delta(CH_2)$ deformation), and 1655 cm^{-1} (amide I ν(C═O) of proteins). On the other hand, the raw in vivo tissue Raman spectra (Figure 2a) also contained the prominent fused silica and sapphire Raman peaks from the fiber-optic Raman probe, that is, 380, 417, 490, 606, 646, and 751 cm^{-1}. We evaluated the usefulness of the univariate internal reference based on the single strong sapphire Raman peak intensity at 417 cm^{-1} (after tissue autofluorescence background subtraction). Figure 3 shows the regression results (i.e., measured laser power vs predicted laser power). This Raman peak was well correlated with the laser excitation power (R^2 = 0.972) but gave a relatively poor power reference (RMSE of 6.2 mW). This indicates that simply based on a single peak, the univariate reference method is accompanied by a large error for quantitative analysis.

To tackle the above problems, we utilized a PLS regression model to extract multivariate reference signals from the oral tissue Raman spectra. The Rayleigh scattered light was excluded from PLS analysis because of difficulty in keeping the Rayleigh light detected within the dynamic range of the CCD camera. The measured in vivo raw tissue Raman spectra were arranged in a matrix with row-wise spectra and column-wise wavenumbers. The reference laser power levels were arranged in a column vector representing the dependent variables. After mean-centering, a PLS regression model was developed using the leave-one-subject-out cross-validation in order to establish

dx.doi.org/10.1021/ac402059v | Anal. Chem. 2013, 85, 11297−11303

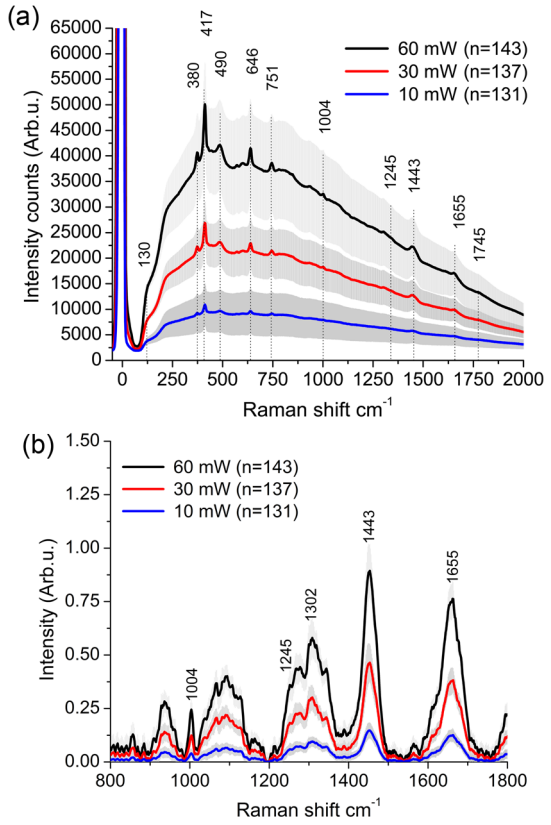


Figure 2. (a) The mean in vivo raw tissue Raman spectra ± 1 standard deviation (SD) of the inner lip by using different 785 nm laser excitation powers (i.e., 10, 30, and 60 mW). Characteristic Raman peaks from the fused silica fiber and sapphire ball-lens are also be observed. (b) The autofluorescence-subtracted and intensity-calibrated mean in vivo tissue Raman spectra ± 1 SD of inner lip by using different 785 nm laser excitation powers (i.e., 10, 30, and 60 mW).

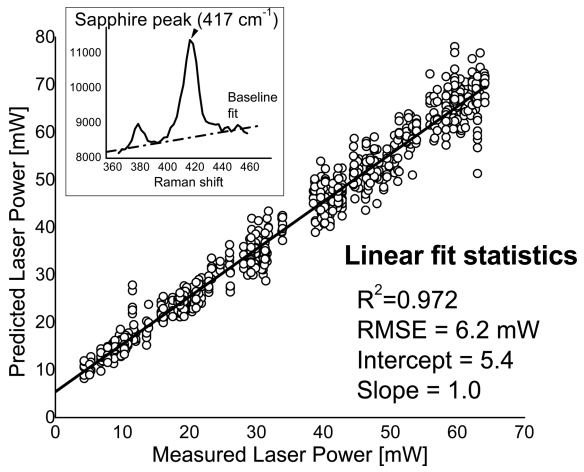


Figure 3. The relationship between the laser excitation power and the intensity of the prominent sapphire Raman peak at 417 cm^{-1} after background subtraction ($y = 5.4 + 1.0x$). A correlation coefficient of $R^2 = 0.972$ is obtained with RMSE of 6.2 mW.

the optimum algorithm for rendering a robust internal reference. Figure 4 shows the RMSEC and RMSECV as a function of retained LVs. The PLS regression analysis showed that an optimal model (RMSECV = 2.5 mW) could be obtained using four LVs. Obviously, including more than four LVs did not improve the cross-validated model significantly.

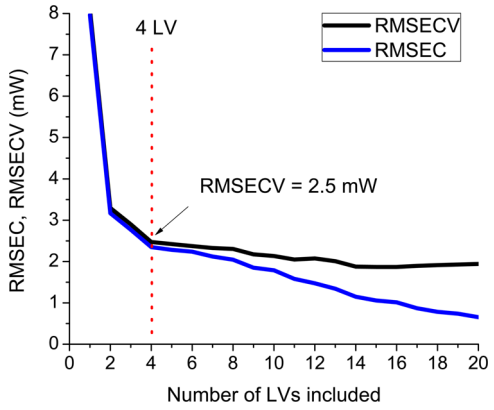


Figure 4. The relationship among a number of PLS factors: latent variables (LVs), the root-mean-square error of calibration (RMSEC), and the root-mean-square error of cross validation (RMSECV) for the prediction of laser excitation powers (mW) used.

Figure 5 displays the first four LV loadings accounting for the largest Raman spectral variance (i.e., LV1, 94.8%; LV2, 3.0%;

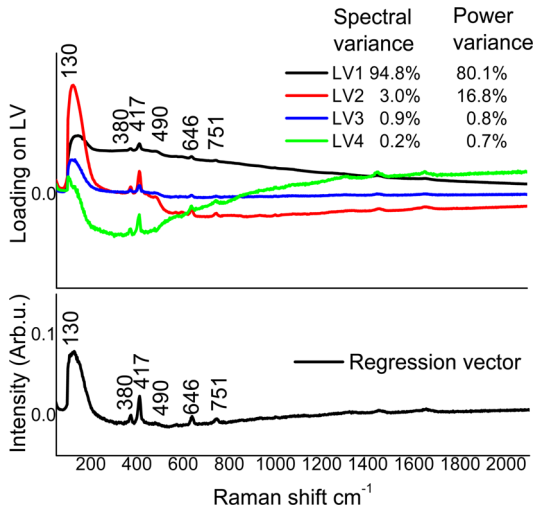


Figure 5. LV loadings for the component PLS regression model with the spectral variance captured (LV1, 94.8%; LV2, 3.0%; LV3, 0.9%; and LV4, 0.2%) and the corresponding laser excitation power variance (LV1, 80.1%; LV2, 16.8%; LV3, 0.8%; LV4, 0.7%). Also shown is the calculated PLS regression vector.

LV3, 0.9%; and LV4, 0.2%) and laser excitation power variance (LV1, 80.1%; LV2, 16.8%; LV3, 0.8%; LV4, 0.7%). Also shown is the calculated PLS regression vector. Figure 6a shows how the multivariate reference technique can be used for in vivo laser power monitoring (i.e., measured laser power vs predicted laser power) using leave-one-subject-out cross validation. The data can be best fit by using a linear function ($y = 0.551 + 0.984x$; $R^2 = 0.981$). The PLS model complexity of four LVs offered an accurate internal reference for laser excitation power monitoring with a RMSECV of 2.5 mW. The same PLS regression model was subsequently implemented online in our Raman software for independent validation of the five new subjects ($n = 166$ spectra) in real time. Figure 6b shows the relationship between the actual laser excitation power measured and the predicted laser excitation power using the developed PLS regression model. The RMSEP of 2.4 mW and a linear relationship ($y = 0.342 + 1.011x$; $R^2 = 0.985$) can be obtained.

dx.doi.org/10.1021/ac402059v | Anal. Chem. 2013, 85, 11297−11303

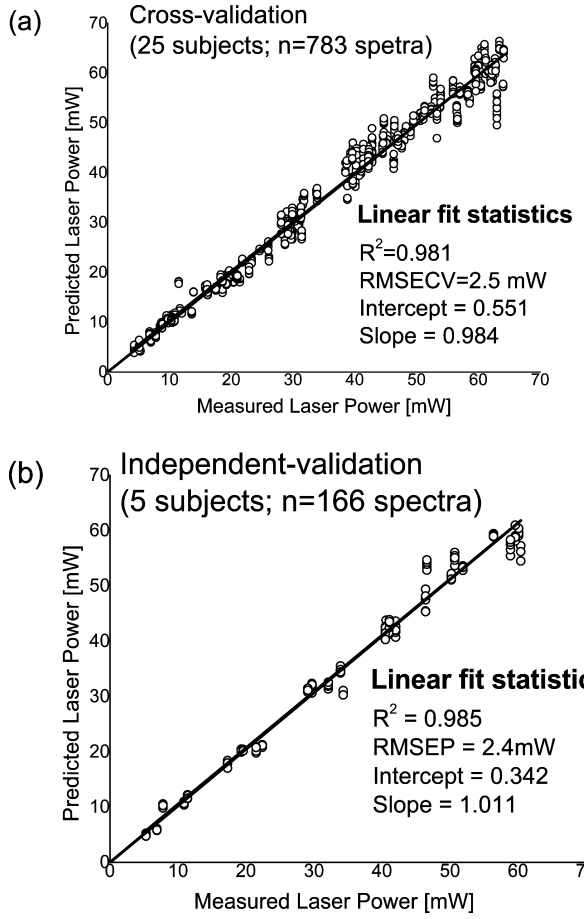


Figure 6. (a) The relationship between the actual and the predicted laser excitation powers using a PLS regression model based on the leave-one-subject-out cross-validation. Also shown is the linear fit to the data. A substantially linear relationship (y = 0.551 + 0.984x; R^2 = 0.981; RMSECV = 2.5 mW) can be obtained. (b) The relationship (y = 0.342 + 1.011x; R^2 = 0.985; RMSEP = 2.4 mW) between the actual and the predicted laser excitation powers using PLS regression together with independent validation.

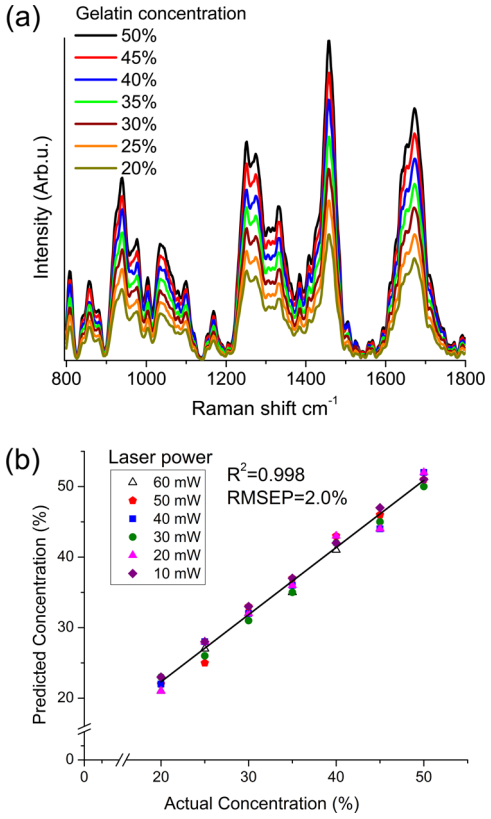


Figure 7. (a) Raman spectra of gelatin tissue phantoms prepared with different concentrations (i.e., 20, 25, 30, 35, 40, 45, and 50% by weight) measured at 60 mW laser excitaiton power. (b) The correlation (R^2 = 0.998; RMSEP = 2.0%) between the actual and predicted gelatin concentrations in tissue phantoms after the correction with the predicted laser power.

Hence, the multivariate internal reference technique is much more accurate than that obtained through the univariate internal reference technique.

We further validate the multivariate reference technique developed for quantitative spectral analysis of tissue phantoms in a laboratory setting. Seven tissue phantoms composed of gelatin with different concentrations (i.e., 20, 25, 30, 35, 40, 45, and 50% by weight) were constructed. The above methodology of generating an internal reference was repeated for the tissue phantom samples. Raman spectra (n = 133 spectra) from gelatin phantoms were subsequently measured and normalized to the internal reference predicted in real time. Figure 7a shows the Raman spectra measured from gelatin tissue phantoms with different concentrations at 60 mW excitation laser power. As expected, these Raman spectra show a linear relationship (R^2 = 0.992) between the Raman peak intensities and gelatin concentrations. Figure 7b shows the correlation between the actual gelatin concentrations and the predicted concentrations (defined as the 1445 cm^{-1} peak intensity) with varying excitation laser powers (from 10 to 60 mW). It is evident that by correcting the laser power variation using the multivariate internal reference technique, accurate quantitative analysis of gelatin tissue phantoms can be realized (RMSEP = 2.0% and R^2 = 0.998). The above results demonstrate that the multivariate internal reference technique developed can realize robust quantitative compositional analysis in fiber-optic tissue Raman spectroscopy.

DISCUSSION

Raman spectroscopy is a compelling technique for quantitative analysis in analytical chemistry.[16,17] In biomedical applications of fiber-optic Raman spectroscopy, the implementation of reference signals accounting for instrumental variations is essential to realizing accurate quantitative analysis. We have presented a novel multivariate self-referencing technique that takes advantage of the multivariate characteristics of the fiber-optic probe background. We validated the technique using in vivo oral tissue as an example that represents the challenging experimental scenario in fiber-optic biomedical Raman spectroscopy. By assigning the laser excitation power as a dependent parameter, PLS regression was able to resolve the internal reference peaks originating from the fiber-optic probe (e.g., fused silica fiber and sapphire ball-lens), even though these were embedded in the in vivo tissue Raman spectra (Figure 2a). Not surprisingly, the loading of LV1 represented the largest spectral variation (94.8% spectral variance accounting for ~80% of the laser excitation power variance), which contained the highly specific probe background peaks (i.e., <130, 380, 417, 490, 606, 646, 751 cm^{-1}) as well as the fused silica fiber autofluorescence background (Figure 1). In contrast, loadings on LV2−4 (4.1% spectral variance

dx.doi.org/10.1021/ac402059v | Anal. Chem. 2013, 85, 11297−11303

accounting for 18% of the laser excitation power variance) were associated with complex relationships between the fused silica background and ball lens sapphire peaks possibly related to fiber bendings and optical/spectral properties of the in vivo tissue. By examining the regression vector more closely (Figure 5), it was found that the distal sapphire ball-lens and fused silica boson peaks were the most significant reference features. The prominent weight of the boson peak was probably caused by the absence of intense tissue Raman and autofluorescence in the low-wavenumber region (~60−130 cm^{-1}). The PLS regression also resolved subtle tissue Raman peaks in the fingerprint region (i.e., 800−1800), that is 1302 cm^{-1} (CH_3CH_2 twisting and wagging), 1443 cm^{-1} ($\delta(CH_2)$ deformation), and 1655 cm^{-1} (amide I ν(C═O) of proteins). The magnitudes of the tissue Raman peaks were relatively small (<5%) compared to the reference Raman peaks from the fiber-optic probe. Nevertheless, it is interesting that tissue Raman spectral variations can serve as an internal reference, suggesting that some tissue biochemical components are quite consistent among different subjects. Our further analysis (Figure 6a) showed that by using the multivariate internal reference method, the good linear relationship (R^2 = 0.981; RMSECV = 2.5 mW) could be achieved for laser excitation power monitoring based on leave-one-subject-out cross validation, which is superior to the univariate reference method in Figure 3 (RMSE = 6.2 mW). These results were subsequently validated independently on the five new subjects (Figure 6b (RMSEP = 2.4 mW; R^2 = 0.985)), verifying that we did not overfit the tissue Raman data acquired. Therefore, besides providing important insights into the spectral characteristics of fiber-optic Raman probes, the use of PLS regression can also offer an accurate and robust multivariate reference for in vivo tissue measurements. One should notice that some of the errors observed can be explained by the errors in the reference method (i.e., laser power meter) and probe handling variations. We further proved the analytical value of this technique for standardizing the absolute intensity for quantitative spectral analysis of gelatin tissue phantoms in the presence of highly varying laser excitation powers (Figure 7(a,b)). Without normalizing the intensity of the gelatin Raman spectra using the multivariate reference technique, the linearity of the quantitative analysis at different laser excitation powers was very poor (R^2 = 0.385, data not shown). On the other hand, after standardizing the intensity of the gelatin Raman spectra using the multivariate reference technique, the quantitative analysis can be accurately realized with a linear relationship (R^2 = 0.998) independent of laser excitation power (Figure 7b).

The use of PLS regression to extract multivariate reference signals has several advantages: (i) The multivariate reference technique is less prone to spectral noise compared to the univariate reference method, but containing a broad range of spectral information for improved accuracy. This is especially evident for noisy spectra using low laser power levels (Figure 6). (ii) PLS regression is able to model complex instrument−tissue relationships whereby the sample or instrument interferes with the background references. (iii) No external reference materials are required to be incorporated into the fiber-optic probe design, and hence, the method presented herein substantially eliminates interfering reference signals in the fingerprint region (i.e., 800−1800 cm^{-1}) that contains a wealth of biochemical and biomolecular information for better tissue diagnosis and characterization.[1−15]

In summary, we present a novel multivariate reference technique for quantitative analysis of fiber-optic Raman spectroscopy. We prove that the multivariate characteristics of background Raman signals from a ball-lens fiber-optic Raman probe can serve as an internal reference for quantitative Raman analysis, which is superior to the univariate reference method. By purposely varying the laser-excitation power, partial least-squares (PLS) regression is able to uncover fused-silica fiber and ball-lens sapphire Raman reference signals embedded in tissue Raman spectra. We demonstrate that multivariate internal reference signals of fiber-optic Raman probes can be advantageously used for standardizing the tissue Raman intensity to realize quantitative analysis of tissue Raman measurements in vivo, which is particularly appealing in the challenging Raman endoscopic applications.

AUTHOR INFORMATION

Corresponding Author

*Tel.: +65- 6516-8856. Fax: +65- 6872-3069. E-mail: biehzw@nus.edu.sg.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This research was supported by the National Medical Research Council and the Biomedical Research Council, Singapore.

REFERENCES

(1) Stone, N.; Kendall, C.; Smith, J.; Crow, P.; Barr, H. *Faraday Discuss.* **2004**, *126*, 141−157.
(2) Mahadevan-Jansen, A.; Richards-Kortum, R. *J. Biomed. Opt.* **1996**, *1*, 31−70.
(3) Buckley, K.; Matousek, P. *J. Pharm. Biomed. Anal.* **2011**, *55*, 645−652.
(4) Sikirzhytski, V.; Virkler, K.; Lednev, I. K. *Sensors.* **2010**, *10*, 2869−2884.
(5) Draga, R. O. P.; Grimbergen, M. C. M.; Vijver- berg, P. L. M.; van Swol, C. F. P.; Jonges, T. G. N.; Kummer, J. A.; Ruud Bosch, J. L. H. *Anal. Chem.* **2010**, *82*, 5993−5999.
(6) Bergholt, M. S.; Lin, K.; Zheng, W.; Lau, D.; Huang, Z. *J. Biomed. Opt.* **2012**, *17*, 077002.
(7) Kanter, E. M.; Majumbder, S.; Vargis, E.; Robichaux-Viehoever, A.; Kanter, G. J.; Shappell, H.; Jones, H. W., III; Mahadevan-Jansen, A. *J. Raman Spectrosc.* **2009**, *40*, 205−211.
(8) Short, M. A.; Lam, S.; McWilliams, A. M.; Ionescu, D. N.; Zeng, H. *J. Thorac. Oncol.* **2011**, *6*, 1206−1214.
(9) Shim, M. G.; Wong Kee Song, L. M.; Marcon, N. E.; Wilson, B. C. *Photochem. Photobiol.* **2000**, *72*, 146−150.
(10) Bergholt, M. S.; Zheng, W.; Lin, K.; Ho, K. Y.; Teh, M.; Yeoh, K. G.; So, J. B. Y.; Huang, Z. *Technol. Cancer Res. Treat.* **2011**, *10*, 103−112.
(11) Bergholt, M. S.; Zheng, W.; Lin, K.; Ho, K. Y.; Teh, M.; Yeoh, K. G.; So, J. B. Y.; Huang, Z. *Analyst* **2010**, *135*, 3162−3168.
(12) Bergholt, M. S.; Zheng, W.; Lin, K.; Ho, K. Y.; Teh, M.; Yeoh, K. G.; So, J. B. Y.; Shabbir, A.; Huang, Z. *J. Biophoton.* **2013**, *6*, 49−59.
(13) Okagbare, P. I.; Morris, M. D. *Analyst* **2012**, *137*, 77−81.
(14) Okagbare, P. I.; Morris, M. D. *Appl. Spectrosc.* **2012**, *66*, 728−730.
(15) Bergholt, M. S.; Zheng, W.; Lin, K.; Ho, K. Y.; Teh, M.; Yeoh, K. G.; So, J. B. Y.; Huang, Z. *Biosens. Bioelectron.* **2011**, *26*, 4104−4110.
(16) Zheng, X.; Fu, W.; Albin, S.; Wise, K. L.; Javey, A.; Cooper, J. B. *Appl. Spectrosc.* **2001**, *55*, 382−388.
(17) Xiao, H.; Dai, S.; Young, J. P.; Feigerle, C. S.; Edwards, A. G. *Appl. Spectrosc.* **1998**, *52*, 626−628.
(18) Dai, S.; Mamantov, G.; Begun, G. M.; Coffield, J. E.; Young, J. P. *Appl. Spectrosc.* **1993**, *47*, 1286−1288.

(19) Wang, J.; Bergholt, M. S.; Zheng, W.; Huang, Z. *Opt. Lett.* **2013**, *38*, 2321−2323.
(20) Huang, Z.; Zeng, H.; Hamzavi, I.; McLean, D. I.; Lui, H. *Opt. Lett.* **2001**, *26*, 1782−1784.
(21) Huang, Z.; Teh, S. K.; Zheng, W.; Mo, J.; Lin, K.; Shao, X.; Ho, K. Y.; Teh, M.; Yeoh, K. G. *Opt. Lett.* **2009**, *34*, 758−760.
(22) Duraipandian, S.; Bergholt, M. S.; Zheng, W.; Ho, K. Y.; Teh, M.; Yeoh, K. G.; So, J. B. Y.; Shabbir, A.; Huang, Z. *J. Biomed. Opt.* **2012**, *17*, 081418.
(23) Geladi, P.; Kowalski, B. R. *Anal. Chim. Acta* **1986**, *185*, 1−17.
(24) Lorber, A.; Wangen, L. E.; Kowalski, B. R. *J. Chemom.* **1987**, *1*, 19−31.
(25) Porto, S. P. S.; Krishnan, R. S. *J. Chem. Phys.* **1967**, *47*, 1009−1013.
(26) Papatheodorou, G. N.; Kalampounias, A. G. *J. Phys.: Condens. Matter* **2009**, *21*, 205101.
(27) Gouadec, G.; Colomban, P. *Prog. Cryst. Growth. Charact.* **2007**, *53*, 1−56.
(28) Bergholt, M. S.; Zheng, W.; Huang, Z. *J. Raman Spectrosc.* **2012**, *43*, 255−262.