Automated Method for Subtraction of Fluorescence from Biological Raman Spectra

CHAD A. LIEBER and ANITA MAHADEVAN-JANSEN*

Department of Biomedical Engineering, Vanderbilt University, Station B, Box 351631, Nashville, Tennessee 37235

One of the challenges of using Raman spectroscopy for biological applications is the inherent fluorescence generated by many biological molecules that underlies the measured spectra. This fluorescence can sometimes be several orders of magnitude more intense than the weak Raman scatter, and its presence must be minimized in order to resolve and analyze the Raman spectrum. Several techniques involving hardware and software have been devised for this purpose; these include the use of wavelength shifting, time gating, frequency-domain filtering, first- and second-order derivatives, and simple curve fitting of the broadband variation with a high-order polynomial. Of these, polynomial fitting has been found to be a simple but effective method. However, this technique typically requires user intervention and thus is time consuming and prone to variability. An automated method for fluorescence subtraction, based on a modification to least-squares polynomial curve fitting, is described. Results indicate that the presented automated method is proficient in fluorescence subtraction, repeatability, and in retention of Raman spectral lineshapes.

Index Headings: **Raman spectroscopy; Fluorescence rejection; Background removal; Tissue diagnosis.**

INTRODUCTION

Recent years have seen an explosion in the use of Raman spectroscopy for biological purposes such as tissue diagnosis, blood analyte detection, and cellular examination. The greatest benefit of this technique is its high sensitivity to subtle molecular (biochemical) changes, as well as its capability for nonintrusive application. However, biological applications of Raman spectroscopy involve turbid, chemically complex, and widely varying target sites. Thus, the challenge in using Raman spectroscopy for biological purposes is not only the acquisition of viable Raman signatures but also the suppression of inherent noise sources present in the target media. Perhaps the greatest contributor of noise to biological Raman spectra is the intrinsic fluorescence of many organic molecules in biological materials. This fluorescence is often several orders of magnitude more intense than the weak chemical transitions probed by Raman spectroscopy and, if left untreated, can dominate the Raman spectra and make analysis of sample biochemistry impractical. Therefore, in order to extract Raman signal from the raw spectrum acquired, it is necessary to process the spectrum to remove this fluorescence.

A number of methods, both instrumentational and mathematical, have been proposed for fluorescence subtraction from raw Raman signals. Methods implemented in hardware such as wavelength shifting and time gating have been shown to effectively minimize fluorescence interference in Raman spectra, but these methods require modifications to the spectroscopic system.[1–3] Mathematical methods implemented in software require no such system modifications and thus have become the norm for fluorescence reduction. These methods include first- and second-order differentiation,[4,5] frequency-domain filtering,[2] wavelet transformation,[6,7] and polynomial fitting.[8–10] Though each of these methods has been shown to be useful in certain situations, they are not without limitations. Differentiation is an unbiased and efficient method for fluorescence subtraction, yet this method severely distorts Raman line shapes and relies on complex mathematical fitting algorithms to reproduce a traditional spectral form.[2] Frequency-based techniques can under- or over-filter, or cause artifacts to be generated in the processed spectra if the frequency elements of the Raman and noise features are not well separated.[2] Wavelet transformation is highly dependent on the decomposition method used and the shape of the fluorescence background.[6] Polynomial fitting is often used for its preservation of traditional Raman line shapes, yet most published records rely on sample-dependent user intervention for assignment of ''non-Raman'' locations on which to fit the curve.

POLYNOMIAL CURVE-FITTING

Polynomial curve-fitting has a distinct advantage over other fluorescence reduction techniques in its ability to retain the spectral contours and intensities of the input Raman spectra. However, simply fitting a polynomial curve to the raw Raman spectrum in a least-squares manner does not efficiently reproduce the fluorescence background, as the fit will be based on minimizing the differences between the fit and the measured spectrum, which includes both the fluorescence background and the Raman peaks. Thus, subsequent subtraction of this fit polynomial results in a spectrum that varies about the zero baseline. It is therefore paramount that the polynomial fits the spectral regions containing only background fluorescence while ignoring spectral regions containing Raman bands. The polynomial fitting technique traditionally relies on user-selected spectral locations on which to base the fit. Unfortunately, this subjective intervention has several drawbacks. It is time consuming, as the user must process each spectrum individually to identify non-Raman-active spectral regions to be used in the fit. In addition, identification of non-Raman-active frequencies is not always trivial, as biological Raman spectra sometimes contain several adjacent peaks or peaks that are not immediately obvious. The end effect is a method that is prone to variability.

Received 29 January 2003; accepted 17 June 2003.

* Author to whom correspondence should be sent. E-mail: anita.mahadevan-jansen@vanderbilt.edu.

0003-7028/03/5711-1363$2.00/0

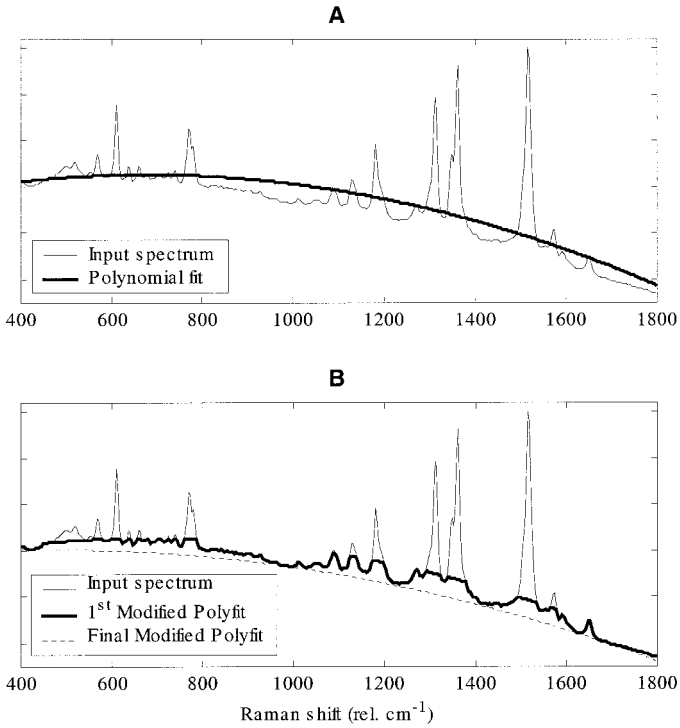

FIG. 1. Polynomial curve fitting as it is applied to rhodamine 6G. (**A**) Raw measured Raman spectrum and a least-squared polynomial curve fit of the baseline fluorescence. (**B**) Raw measured Raman spectrum and two iterations of the modified curve fit applied until a smoothed polynomial curve along the base of the measured spectrum is obtained.

MODIFIED POLYFIT METHOD

To address these limitations, the modified polyfit method for fluorescence subtraction was developed. This method smoothes the spectrum in such a way that Raman peaks are automatically eliminated, leaving only the baseline fluorescence intact, to be subtracted from the raw spectrum.

The basis for this method is a least-squares-based polynomial curve-fitting function. However, to eliminate the Raman bands from the fit, this function is modified such that all data points in the generated curve that have an intensity value higher than their respective pixel value in the input spectrum are automatically reassigned to the original intensity (Fig. 1). This process (curve fitting and subsequent reassignment) is repeated typically between 25 and 200 iterations (depending on factors such as the relative amount of fluorescence to Raman), gradually eliminating the higher-frequency Raman peaks from the underlying baseline fluorescence. In the ideal case, the filtering process would cease when there are no longer any data points in the fit curve that require reassignment (all values equal to or less than respective smoothed spectrum intensities). However, due to noise factors and other intrinsic artifacts, it is sometimes impossible to affect all pixel values with the filter. Thus, the filtering process can also be halted if there is convergence in the number of data points affected by each iteration, as determined by a standard root test for convergence. The processed baseline spectrum is then subtracted from the raw spectrum to yield the Raman bands on a near-null baseline.

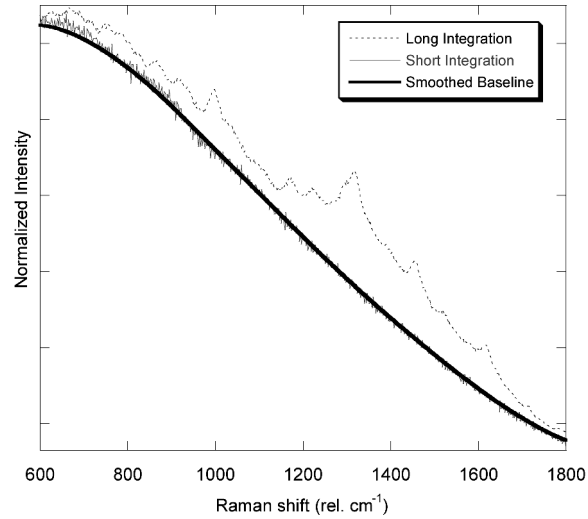

FIG. 2. Spectra of the blue sponge used as the fluorescence phantom measured using a long integration time (1 min) and moderate power (80 mW), using a short integration time (8 ms) and low excitation power (5 mW), and using a sliding window mean filter on the short integration spectrum. Measured spectra were normalized to their mean intensities for visualization.

TEST AND VALIDATION

Measurement System. All test spectra were recorded using a fiber-probe-based Raman spectroscopic system. In this system, a 300-mW, 785 nm GaAlAs diode laser (Process Instruments, UT) is used to excite samples using a 400-μm-core-diameter fused-silica optical fiber. The delivery fiber is contained in an optical probe (Visionex, GA) consisting of seven 300-μm fused-silica beam-steered collection fibers around the central delivery fiber, with in-line filters in the delivery and collection fibers for rejection of signals generated in the fibers themselves.

The collection fibers are aligned linearly and imaged on to a 100 μm entrance slit of the detection system, consisting of an imaging spectrograph (Kaiser Optical Systems Inc, MI) and a deep-depletion, back-illuminated,

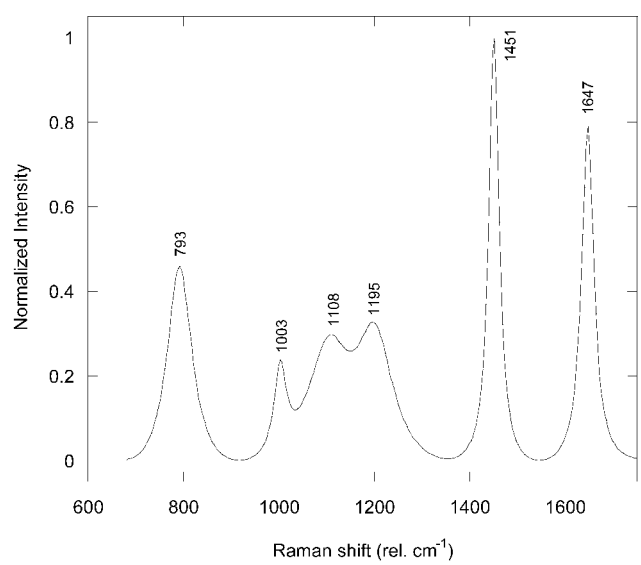

FIG. 3. Mathematically generated pure Raman spectrum created using a series of Lorentzian peaks on a null baseline in a distribution similar to that seen in tissue spectra; peak locations are labeled.

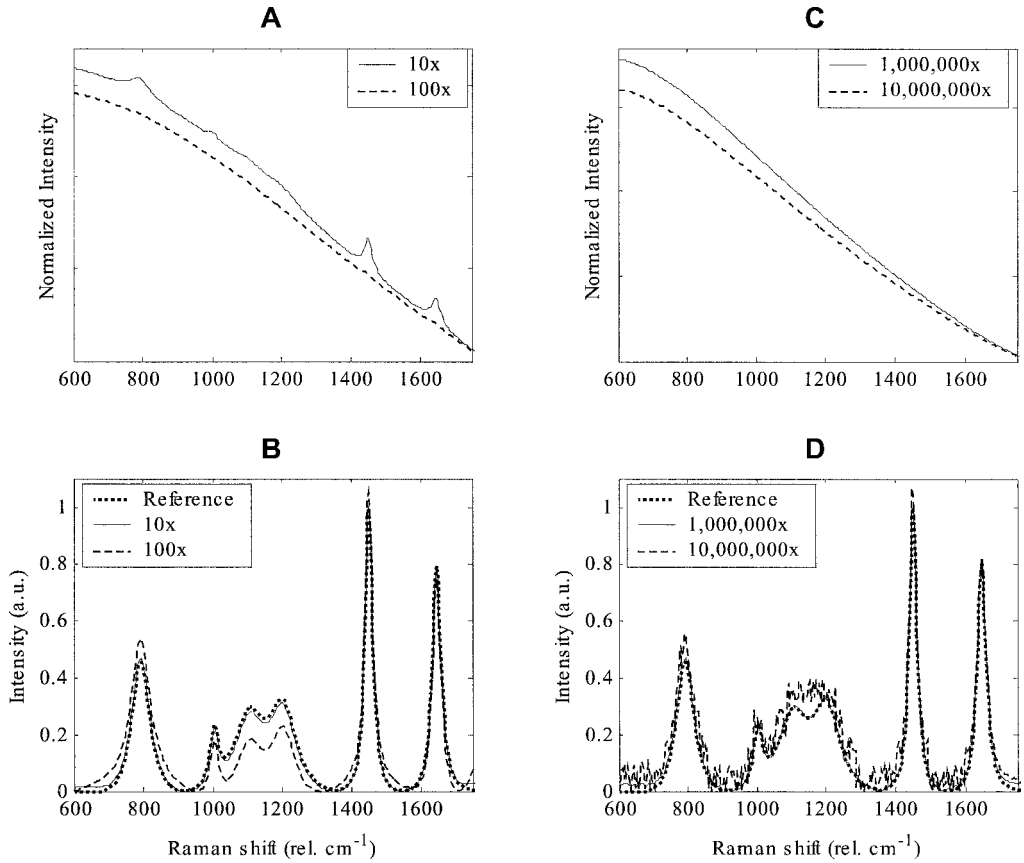

FIG. 4. Phantom spectra of varying fluorescence/Raman ratios processed using the modified polyfit technique. (**A**) Smoothed sponge fluorescence spectrum with pure Raman phantom in a ratio of 10:1 and 100:1, and (**B**) the corresponding extracted Raman spectra along with the pure Raman phantom, which have $R^2 = 0.994$ and 0.726 compared to the pure phantom, respectively. (**C**) Using a fifth-order polynomial fit of the smoothed sponge spectrum as the fluorescence with the pure Raman phantom in a ratio of 1 000 000:1 and 10 000 000:1, and (**D**) the corresponding extracted Raman spectra along with the pure Raman phantom, which have $R^2 = 0.997$ and 0.961 compared to the pure phantom, respectively.

liquid-nitrogen-cooled charge-coupled device (CCD) camera (Roper Scientific Instruments, NJ). The detection system yields a spectral resolution of 7 cm^{-1}.

Spectral calibration of each measured spectrum was performed using a neon–argon lamp. Raman shift was calibrated using 4-acetamidophenol and naphthalene. All spectra were binned to half the spectral resolution of the detection system for direct comparison. Since the modified polyfit technique presented in this paper is designed only to reduce the low-frequency background fluorescence, a first-order Savitzky–Golay filter was used to minimize high-frequency noise (such as shot noise, readout noise, dark current, etc.) present in all Raman spectra. After noise smoothing and subsequent binning, the spectra were treated with the modified polyfit method to remove background fluorescence.

All mathematical processing was performed using in-house code written in Matlab (The MathWorks Inc., MA).

Phantom Design. In order to test the efficacy of the modified polyfit fluorescence removal technique, a phantom system with known Raman and fluorescence spectral contributions was needed. This posed a unique challenge in that nearly all known fluorescent substances are also Raman active. Literature searches and personal communications yielded no likely candidates that were fluorescent but Raman inactive. In studying various materials that were available in a typical biomedical photonics lab, a potential candidate was found in the blue-colored sponge sometimes used in optics packaging. This substance yields moderately high intrinsic fluorescence in very short integration times, with no immediately resolvable Raman peaks. Furthermore, the spectral shape of the fluorescence emission closely resembles that of human tissue. To further guarantee a Raman-free background phantom, the sponge spectrum was measured at a low excitation power (5 mW) with an integration time of 8 ms (the minimum value for the detector and shutter used). This creates a pseudo-time-gated spectrum in that any Raman features present are obscured by thermal and readout noise of the CCD array. Several iterations of a simple sliding-window mean filter (with window width = 2× spectral resolution of the system) served to both smooth the readout noise in the spectrum and to further guarantee elimination of Raman contributions. Figure 2 shows the smoothed sponge spectrum used as the fluorescence phantom, along with the raw pseudo-time-gated spectrum and a raw spectrum obtained at 80 mW and 1 min integration (both measured raw spectra normalized to their mean intensity), which clearly shows both the fluorescence and Raman features.

The Raman spectral components in the phantom were mathematically generated as a series of Lorentzian peaks (Fig. 3), with a distribution similar to that seen in Raman

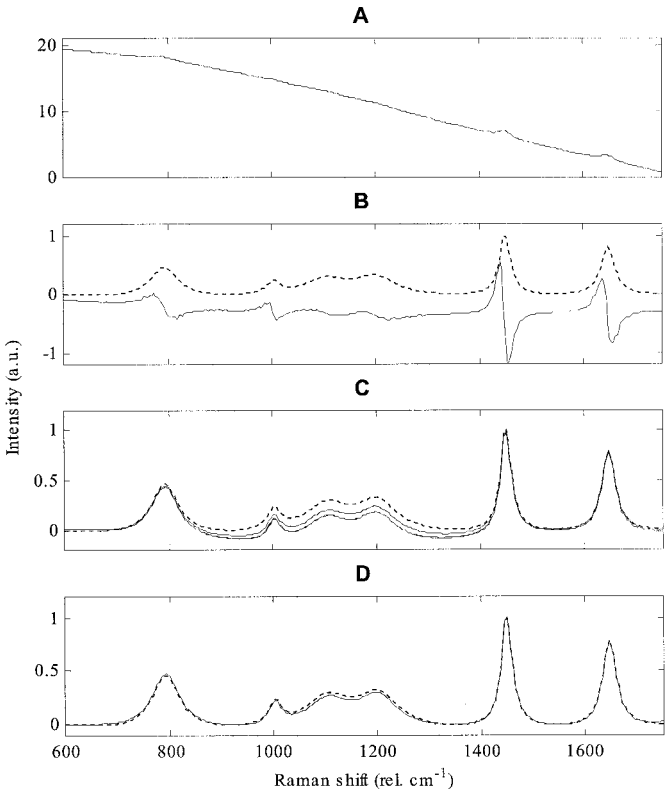

FIG. 5. Comparison of conventional fluorescence subtraction techniques with the modified polyfit techniques using the sponge phantom with a 20:1 fluorescence-to-Raman ratio. (**A**) Raw phantom spectrum; (**B**) processed spectrum using first derivative (solid line) along with the reference Raman component; (**C**) processed spectrum using the manual polynomial fit method by three independent users (solid lines) along with the generated pure Raman component; and (**D**) processed spectrum using the modified polyfit method (solid line) along with the generated pure Raman component.

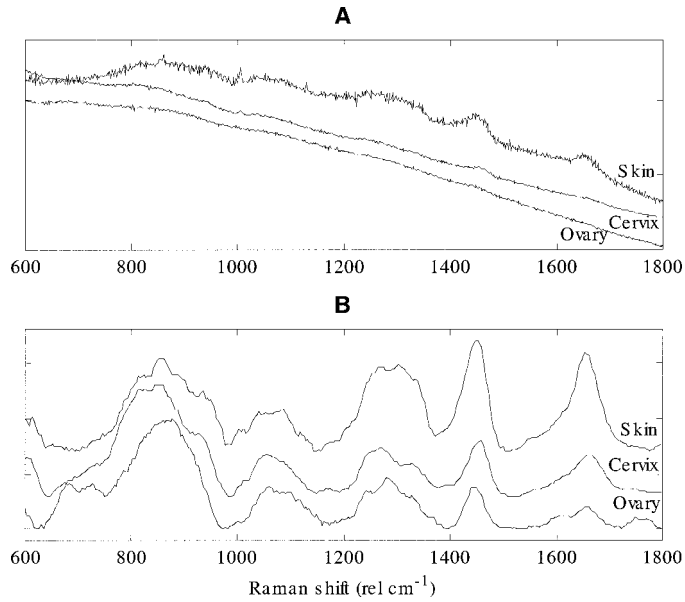

FIG. 6. Tissue Raman spectra of the human skin, cervix, and ovary acquired *in vitro* (**A**) before and (**B**) after modified polyfit processing. Each spectrum is normalized to its mean intensity and all spectra are offset for visualization.

spectra of human tissue.† A linear fit of the baseline offset produced by overlapping peaks was subtracted to ensure a null baseline. A series of phantom spectra were then created by adding the artificial Raman spectrum to the smoothed sponge fluorescence spectrum in varying ratios.

RESULTS AND DISCUSSION

The phantom spectra with various fluorescence-to-Raman ratios were processed using the developed modified polyfit technique. Since the phantom spectra contained known quantities of both fluorescence background and Raman scatter, the performance of the processing technique could be evaluated by its ability to return only the added Raman scatter. Figure 4A shows the raw spectra of two different phantoms prior to fluorescence subtraction where maximum fluorescence intensity was 10 and 100 times more than the maximum Raman intensity (at 1451 rel. cm^{-1}), respectively. For illustration purposes, each phantom spectrum was normalized to its mean intensity and the spectra were offset for clarity. After modified polyfit processing (Fig. 4B), the filtered spectra yielded coefficients of determination (R^2) of 0.994 and 0.726, respectively, with respect to the generated pure Raman component added to the phantom spectra; this indicates that the Raman bands can be extracted from signals in which fluorescence is at least one hundred times more intense than the Raman features.

The modified polyfit technique relies on the observations by many groups that biological fluorescence is best approximated by a high-order polynomial. Thus, there are two separate factors that affect the processing of a measured Raman spectrum using the modified polyfit method: (1) the ability to extract faint Raman spectral features from a much higher background intensity, and (2) the dependence upon the background fluorescence's strict adherence to the contours of a high-order polynomial. In order to test only the former factor, phantom spectra were created by adding the generated pure Raman spectral component to a fifth-order polynomial fit of the smoothed sponge spectrum and were processed using the modified polyfit technique. Figure 4C shows the phantom spectra created for this experiment, each spectrum normalized to its mean intensity and offset, for illustration purposes only. The resultant modified polyfit-processed spectra (Fig. 4D) show that Raman bands can be extracted even at 10 000 000:1 ratios of fluorescence to Raman intensity ($R^2 = 0.961$). This data also illustrates that approximation of the fluorescence background by a high-order polynomial is a limiting factor of this method. While biological fluorescence has been reported to be best approximated by a high-order polynomial on the order of 4–5,[8,10] this may not always be the case (in non-biological spectra this is even more plausible). However, the results presented here are directed towards the application of this method for biological materials and thus show its validity using a polynomial fit. For other types of materials, the basic premise of the modified polyfit method can also be implemented using non-polynomial background line shapes,

† Centers (and respective full widths at half-maximum) of peaks used were 568.75 (70), 792.75 (59.5), 1002.75 (35), 1107.75 (105), 1195.25 (94.5), 1450.75 (28), and 1646.75 (31.5) rel. cm^{-1}.

such as Gaussian, Weibull, logarithmic, exponential, or other distributions.

Figure 5 compares the modified polyfit method with two other conventional fluorescence subtraction techniques typically used to process biological spectra, first-order differentiation and manual polynomial curve fitting. The smoothed sponge spectrum was used as the phantom fluorescence background, at a 20:1 intensity ratio with the generated pure Raman spectrum. This ratio was selected for its approximation to that encountered in biological measurements. Figure 5 clearly illustrates the advantages of the modified polyfit method as compared to the other methods in terms of consistency, peak retention, and baseline drift. The output of the modified polyfit method is seen to accurately retain the Raman spectral components of the phantom. In comparison, the derivative method makes the lower intensity Raman bands (1050–1300 rel. cm^{-1}) difficult to identify in the processed spectrum, and the conversion of Raman peaks to sinusoids makes traditional peak identification difficult, especially between peaks of close proximity. The manual polynomial fit was performed by three separate individuals so that variability due to user intervention could be realized. The individuals were given the raw phantom spectrum with no *a priori* knowledge of the Raman peak locations. Even in this small number of cases, the variability in the fit is evident.

While demonstration of the modified polyfit technique on a phantom demonstrates its capability, the technique was designed for application to biological samples, specifically human tissues. Several human tissue specimens (ovary, skin, and cervix) were obtained fresh-frozen for this study, thawed in room temperature phosphate buffered saline (PBS) solution, and measured using the spectroscopic system described (P = 80 mW, 30 s integration). Although the modified polyfit method can be used to automate fluorescence subtraction from measured Raman signals of different types of materials, it was designed specifically for use with tissue spectra. Tissue spectra typically display low signal-to-noise ratios (as compared to chemical standards) and a fluorescence background that has been described as best approximated by a fourth- or fifth-order polynomial.[8,10] Figure 6 shows the Raman spectra of these *in vitro* human tissues before (Fig. 6A) and after processing with the modified polyfit method (Fig. 6B) and demonstrates the effectiveness of the presented method in subtracting tissue fluorescence and retaining Raman signatures. Although the intensity of the broad peak in the 800–900 rel. cm^{-1} region varies considerably due to deviations in silica contamination from within the collection optics, the dominant tissue peaks at ~1200–1400, ~1450, and ~1660 rel. cm^{-1} are clearly retained.

CONCLUSION

The goal of any fluorescence subtraction technique is the unbiased removal of all fluorescence contribution, with no effect on the Raman bands. The derivative method is quite adept at the former requirement; however, it fails miserably at the latter. Polynomial curve-fitting, in principle, provides a solution to both, but its implementation thus far has relied on user intervention, which makes this technique unreliable and time consuming. The modified polyfit method described here automates the manual curve-fitting process, thus fulfilling the requirements of an effective fluorescence subtraction method. This method can be completely automated and has been shown to be proficient in producing a null baseline and retaining Raman spectral contours, even at exceedingly small signal-to-noise ratios.

1. R. Van-Duyne, D. Jeanmaire, and D. Shriver, Anal. Chem. **46,** 213 (1974).
2. P. Mosier-Boss, S. Lieberman, and R. Newberry, Appl. Spectrosc. **49,** 630 (1995).
3. J. Zhao, M. M. Carrabba, and F. S. Allen, Appl. Spectrosc. **56,** 834 (2002).
4. D. Zhang and D. Ben-Amotz, Appl. Spectrosc. **54,** 1379 (2000).
5. A. O'Grady, A. Dennis, D. Denvir, J. McGarvey, and S. Bell, Anal. Chem. **73,** 2058 (2001).
6. V. Barclay and R. Bonner, Anal. Chem. **69,** 78 (1997).
7. T. Cai, D. Zhang, and D. Ben-Amotz, Appl. Spectrosc. **55,** 1124 (2001).
8. A. Mahadevan-Jansen, M. F. Mitchell, N. Ramanujam, A. Malpica, S. Thomsen, U. Utzinger, and R. Richards-Kortum, Photochem. Photobiol. **68,** 123 (1998).
9. T. Vickers, J. R. E. Wambles, and C. Mann, Appl. Spectrosc. **55,** 389 (2001).
10. J. F. Brennan, Y. Wang, R. R. Dasari, and M. S. Feld, Appl. Spectrosc. **51,** 201 (1997).