Journal of Cellular Biochemistry Supplement 39:125–137 (2002)

Model-Based Biological Raman Spectral Imaging

Karen E. Shafer-Peltier,[1] Abigail S. Haka,[2] Jason T. Motz,[2] Maryann Fitzmaurice,[3] Ramachandra R. Dasari,[2] and Michael S. Feld[2]*

[1]Department of Biomedical Engineering, Northwestern University, Evanston, Illinois 60208
[2]G.R. Harrison Spectroscopy Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139
[3]Department of Pathology, University Hospitals of Cleveland and Case Western Reserve University, Cleveland, Ohio 44106

Abstract Raman spectral imaging is a powerful tool for determining chemical information in a biological specimen. The challenge is to condense the large amount of spectral information into an easily visualized form with high information content. Researchers have applied a range of techniques, from peak-height ratios to sophisticated models, to produce interpretable Raman images. The purpose of this article is to review some of the more common imaging approaches, in particular principal components analysis, multivariate curve resolution, and Euclidean distance, as well as to present a new technique, morphological modeling. How to best extract meaningful chemical information using each imaging approach will be discussed and examples of images produced with each will be shown. J. Cell. Biochem. Suppl. 39: 125–137, 2002. © 2002 Wiley-Liss, Inc.

Key words: Raman spectroscopy; confocal microscopy; imaging; cells; breast; artery; tissue; principal component analysis; Euclidean distance; multivariate curve resolution; least-squares modeling

Raman spectroscopy can provide detailed qualitative and quantitative information about a sample being studied. It is an inelastic scattering process in which photons incident on a sample transfer energy to or from the sample's vibrational or rotational modes. The difference in energy between the incident and exiting photons corresponds to the transition of a molecule from one state to another. Since the energy levels are unique for each molecule, Raman spectra are chemical specific [McCreery, 2000]. The wealth of information obtainable from a Raman spectral image has led to its use in many fields, including environmental [Nelson et al., 2001], industrial [Andrew et al., 1998], polymers [Appel et al., 2000], semiconductors [Schaeberle et al., 2001], food science [Archibald et al., 1998], and pharmaceutical [Clarke et al., 2001] applications. Raman spectral imaging has only begun to be applied to the chemical analysis of complex biological samples. In the past few years, however, a number of papers have been published using Raman spectral imaging to monitor chemical contributions at the cellular and sub-cellular level.

Several approaches have been employed to acquire Raman imaging data sets. The three standard approaches are point scanning, line scanning, and direct imaging [Delhaye and Dhamelincourt, 1975; Turrell and Corset, 1996]. Direct imaging involves the collection of a full image with a single spectral component. Wavelength selectivity is achieved by using either an acousto-optic or a liquid crystal tunable filter that sweeps through specified wavelength intervals capturing a frame at each. Line scanning and point scanning collect a full spectrum (usually covering Raman shifts between 400 and 1,800 cm^{-1} for biological media), either while imaging a line or a single point. The resultant data set from each of these approaches can be thought of as a hypercube of Raman intensity as a function of Raman shift and two spatial axes.

Grant sponsor: NIH; Grant number: P41-RR 02594; Grant sponsor: Pathology Associates of University Hospitals.

*Correspondence to: Michael S. Feld, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139, USA. E-mail: msfeld@mit.edu

Received 16 October 2002; Accepted 16 October 2002
DOI 10.1002/jcb.10418
Published online in Wiley InterScience (www.interscience.wiley.com).

One of the first tissues to be explored using Raman imaging was human breast tissue. In 1996, Treado et al. published Raman images of foreign polymer inclusions (from silicone breast implants) anchored in the fibrous breast tissue surrounding the implant [Schaeberle et al., 1996]. Although the Raman spectrum obtained from the normal biopsy tissue had very weak, uninterpretable features, the 1,615 cm^{-1} band of Dacron polyester was easily identified, enabling imaging of the polymer material within the breast tissue matrix. A subsequent study of breast tissue demonstrates the use of Raman imaging to develop a model to understand the chemical and morphological origins of Raman spectral features observed in normal and diseased tissue [Shafer-Peltier et al., 2002]. This morphological model, consisting of spectra acquired from the cell cytoplasm, cell nucleus, fat, β-carotene, collagen, calcium hydroxyapatite, calcium oxalate dihydrate, cholesterol-like lipid deposits of normal and diseased samples can be used in a linear combination to fit macroscopic breast tissue spectra. The model can in turn be used to generate Raman maps with highly specific information content.

In addition to mapping tissue architecture, Raman imaging can be used for in situ chemical investigation of disease processes. One such example is atherosclerosis where the end product of the disease, ceroid, is defined as an autofluorescent lipid product whose chemical composition is unknown. Recently, van de Poll et al. [2002] have explored the chemistry of ceroid using Raman imaging. This study has provided new insight into the chemical nature of ceroid and the disease processes that produce it. In another experiment, designed to study fatigue-related microdamage in bone, spectroscopically distinct microstructures were correlated with tissue damage [Timlin et al., 2000]. Raman images have even been acquired to study sub-cellular chemistry. Arikan et al. [2002] have used Raman imaging as a tool to monitor beta-carotene in live corpus luteum cells, while Freeman et al. [1998] have investigated the sub-cellular localization of zinc phthalocyanines, photosensitizing agents used in photodynamic therapy. Surface-enhanced Raman spectroscopy in conjunction with imaging has been used to study the chemical composition of live cells [Kneipp et al., 2002]. In particular, the DNA and phenylalanine contents of the cells were monitored.

A time-honored technique for creating spectral images is by examination of a specific peak height. In this approach, the intensity of a particular Raman band at each spatial location is plotted to produce an image [McCreery, 2000]. This method has been widely used and provides information about the spatial location of every molecule in the sample that contributes intensity to the vibrational frequency chosen. However, this approach only takes advantage of a small portion of the data available. In complex biological samples, where several distinct moieties may contribute intensity to a particular Raman band, it is necessary to incorporate all of the spectral information in order to differentiate them. This is achieved by the application of a model that utilizes the full spectrum, as is done with point and line scanning, when creating an image. The key is to compress the information into a manageable, yet still informative form. Some common data compression techniques, which will be presented here, are principal component analysis (PCA), multivariate curve resolution (MCR), and Euclidean distance. Morphological modeling is a new approach that will also be presented.

Each one of these techniques relies on the basic assumption that the Raman spectrum of a mixture of chemicals can be represented as a linear combination of the mixture's component spectra [Buschman et al., 2001a,b; Shafer-Peltier et al., 2002]. Raman images are generated by fitting basis spectra contained within the model to the Raman spectrum obtained at each position in the image. Generally, the more a basis spectrum contributes to a data spectrum, the larger the fit coefficient and the brighter that spot appears in the image of the component being examined. In the cases of PCA and MCR, basis spectra are mathematically derived, whereas for Euclidean distance and morphological modeling, basis spectra are experimentally determined.

In PCA, singular-value decomposition is used to calculate basis spectra [Wold et al., 1987]. The first basis spectrum, or principal component, accounts for the maximum variance in the data if the data is mean-centered prior to analysis. The second basis spectrum accounts for the next most variance, and so on, until the basis spectra account only for the noise in the data. These basis spectra are created such that they are orthogonal to each other, and therefore contain no overlapping spectral information. The fit

coefficients obtained when these principal components are fit to the imaging data set can be used to create a two-dimensional image. This image will provide a map of how the spectral features represented by the principal components are distributed in the sample. In turn, this map can be correlated with morphological features observed through another optical technique, such as phase contrast microscopy or light microscopy with histological staining. The lineshapes of the principal components might also be correlated with the Raman spectra of known chemicals, however, this is difficult as the principal components contain both negative and positive spectral features.

MCR is designed to extract basis spectra that are similar to the real Raman spectra of the chemicals present in the sample [Tauler et al., 1994; Andrew and Hancewicz, 1998]. An initial estimate of the concentrations or basis spectra present in the sample is used in a constrained, alternating least-squares optimization. New estimates for the concentrations and basis spectra are generated by iterating between least-squares solutions for basis spectra and concentrations. These equations can be solved subject to non-negativity constraints to ensure that both the basis spectra and concentrations are all positive and thus physically relevant. Optimization continues until the changes in the concentrations and basis spectra from one iteration to the next are minimal. The more complex the system, the better the initial estimates need to be to obtain meaningful solutions to these equations. Due to the high-degree of overlap in the spectral features of different components and the noise inherent in the data, MCR cannot always converge on the correct solution. However, when a solution is found, the basis spectra produced resemble the Raman spectra of the individual chemicals present in the sample. Once again, the fit coefficients of the basis spectra can be used to produce an image.

Both PCA and MCR are useful techniques when little is known about the sample a priori. They enable one to extract spectral information without knowing its chemical origin. Euclidean distance measurements and morphological modeling both use information about the known chemistry of a sample to create an image. Euclidean distance only requires the knowledge of a few chemicals present whereas morphological modeling requires knowledge of all of the major contributors to the sample's Raman spectrum. Despite requiring the most prior knowledge of the sample, morphological modeling produces the most easily interpretable results.

Euclidean distance classifies spectral variance in the image data from a basis spectrum, usually a pure chemical spectrum, according to the data's geometric distance [Potter et al., 2001]. The distance is calculated using the equation: $\sqrt{\sum_{\lambda} (S(\lambda) - P(\lambda))^2}$ where d is the Euclidean distance, S is the sample data, P is the pure chemical spectrum, and λ represents the wavelengths over which the spectra are acquired. The more a spectrum in the image differs from the basis spectrum, the larger the distance.

Morphological modeling is a new technique for analyzing Raman images, which uses ordinary least-squares to fit a set of basis spectra to the data [Buschman et al., 2001a; Shafer-Peltier et al., 2002]. The origin of the basis spectra is what makes this approach so useful. The basis spectra are acquired from the major morphological features found in a set of representative samples using a Raman confocal microscope. By using a spectrum of a morphological feature acquired in situ, one obtains a spectrum that represents that morphological component in its chemical microenvironment. The basis spectra should account for all of the major chemicals present in the sample, but both the signal to noise of the data as well as the degree of overlap of the basis spectra must be considered to determine whether they can be accurately be resolved. Although basis spectra can be acquired from pure chemical compounds [Brennan et al., 1997], morphologically derived components [Buschman et al., 2001b; Shafer-Peltier et al., 2002] are preferable as they are derived from actual samples, and are thus closer than pure chemical spectra to what is observed in situ. Sometimes, a combination of pure chemical components and morphologically derived components will produce the best result if the chemicals of interest do not occur independently within a sample. If a model is well chosen, the images produced can reveal detailed morphological and chemical structure in the sample.

In this article, the application of morphological modeling to Raman images of human colonic carcinoma cells as well as human breast and artery samples will be demonstrated. This new

technique of morphological modeling will then be compared with other commonly used techniques, primarily: peak height analysis, PCA, MCR, and Euclidean distance. The advantages and disadvantages of each technique, as well as when to use them, will be discussed.

MATERIALS AND METHODS

Tissue Handling

Breast tissue samples were obtained from excisional biopsy specimens while artery samples were obtained from explanted hearts at the time of transplant. Once removed, the tissue was snap frozen in liquid nitrogen and stored at -80°C. The tissue samples were then mounted on a cryostat chuck using Histoprep (Fisher Diagnostics, Orangeburg, NY) and cut into 6 μm thick sections using a cryomicrotome (International Equipment Company, Needham Heights, MA). Several consecutive sections were cut, one mounted on a MgF_2 slide (Moose Hill Enterprises, Inc., Sperryville, VA) for Raman data acquisition and at least two others on glass slides for histological staining. The stained slides were used for pathological confirmation of features observed in the Raman maps. During measurements, the tissue was kept moist with phosphate buffered saline (PBS), pH = 7.4. In addition to the Raman micro-images, phase contrast images of the unstained tissue were recorded via a CCD camera.

Cell studies were performed using the human colonic carcinoma cell line HT29 (gift of the Harvard Digestive Diseases Center, Boston, MA). They were grown using high-glucose Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal calf serum, 100 U/ml penicillin, and 100 μg/ml streptomycin (all Gibco BRL products, Life Technologies, Grand Island, NY). Cells were grown to confluency at 37°C in a humidified atmosphere of 5% CO_2 in air and dispersed into suspension using trypsin. Cell suspensions were placed on MgF_2 flats, rinsed with PBS, buffered at pH = 7.4, and allowed to air dry. Drying of the sample was necessary in order to immobilize the cells for the entire mapping experiment. The dried samples were then rewet with PBS and Raman maps were subsequently acquired. Raman imaging microscope data collected from the dried cells were compared to data collected from viable cells still in suspension using a bulk Raman system. The spectra acquired from the dried cells were used to model the spectra obtained from the viable cells. No residual from the model fit was observed.

Instrumentation

The Raman micro-imaging set-up used to collect the data for the images presented here was a point scan system. Raman excitation was provided by an argon ion laser-pumped Ti:sapphire laser (Coherent Innova 90/Spectra Physics 3900S, Coherent, Inc., Santa Clara, CA). Typically 50–150 mW of 830 nm excitation light was focused through a microscope objective (63 × Zeiss Achroplan, infinity corrected, water immersion, numerical aperture 0.9) to a spot on the sample with a diameter of < 2 μm. The experimental system has been described previously [Shafer-Peltier et al., 2002]. The spectral resolution was ~ 8 cm^{-1}. Spectral maps of the tissue were created by raster scanning the translation stage (Prior Scientific Instruments Ltd., Cambridge, MA) under the microscope objective. Maps were normally acquired with a step size of 2 μm, consistent with the spatial resolution of the confocal microscope. Although data collection time depended on several user defined parameters, such as the image step size, number of steps, and spectral acquisition time, an entire Raman image was typically generated in 2–5 h. A CCD camera atop the microscope allowed for registration of the focused laser spot with a white light trans-illuminated or phase contrast image.

Data Processing

All spectral data processing was preformed using MATLAB (MathWorks, Inc., Natick, MA). The data were corrected for the spectral response of the system using a tungsten light source and then frequency calibrated using the known Raman lines of toluene. Cosmic rays were removed with a derivative filter and the small background from the MgF_2 flat was subtracted. Data were then fit with a fourth or fifth order polynomial, which was subtracted from the spectrum in order to remove any fluorescence background. All data was peak-height normalized to one. Finally, MATLAB was used to implement the various data compression techniques: PCA, MCR, Euclidean distance, and morphological modeling. Algorithms for PCA and ordinary least-squares (used as the fitting algorithm for morphological modeling) already exist in MATLAB, while the algorithm

for MCR was a part of PLS_Toolbox (Eigenvector Research, Inc., Manson, WA). The pure chemicals used for spectroscopic modeling of the HT29 cells: triolein, phosphatidyl choline, cholesterol, and DNA (calf thymus), were purchased from Sigma (St. Louis, MO) [Buschman et al., 2001a]. The morphologically derived spectra used to analyze the artery, breast, and HT29 data have been presented previously [Buschman et al., 2001b; Shafer-Peltier et al., 2002].

In order to obtain improved image contrast a smoothing algorithm based on spatial filtering was applied to all data presented here [Baxes, 1984; Bock and Krischer, 1998]. Spatial filtering relies on the assumption that adjacent pixels in a digital image contain related information. A group of pixels surrounding and including the central pixel is called a kernel. We base our smoothing algorithm on a kernel size of 3×3. Our algorithm uses a mask that weights the contributing pixels according to the reciprocal of their geometric distance from the center of the kernel. The resultant mask is

2/28	3/28	2/28
3/28	8/28	3/28
2/28	3/28	2/28

where each fraction represents the weight of a pixel in the kernel.

RESULTS

Morphological modeling is a powerful tool for collecting architectural and chemical information on a small scale. In Figure 1, features such as the cell membrane, nucleus, and cytoplasm are easily identified when spectra of human colonic carcinoma cells (HT29) are fit with the pure chemical spectra of phosphatidyl choline (A), DNA (B), cholesterol linoleate (C), triolein (D), and "cell cytoplasm" (E), a morphologically derived spectrum developed for the breast tissue model, mostly actin) [Shafer-Peltier et al., 2002]. The spectrum (·) corresponding to the voxel indicated in Figure 1E can be seen in G, along with the corresponding fit (–) and residual (below). The fit contributions of the individual model elements are also shown. The spectral images agree with the phase contrast image, demonstrating that using a simple model of five basis spectra, it is possible to obtain structural and chemical information about a sample at the sub-cellular level. As the cell shown in the image is evenly bisected by the plane of focus of the confocal microscope, the cell membrane (mostly phosphatidyl choline and cholesterol) is observed as a ring structure with the cell cytoplasm and DNA contributions observed clearly as distinct features within. The average nuclear size for HT29 cells is 10 μm, consistent with the dimensions provided by the Raman image of the cell DNA content [Wax et al., 2002].

Morphological modeling can be applied to human tissue samples as well. Figure 2 shows phase contrast images (2A and G) of a mildly atherosclerotic artery along with Raman images depicting the distribution of some of the morphological structures (2B–F). The images clearly show that the cholesterol (2B), foam cells, and necrotic core (2C) are solely confined to the intima while the smooth muscle cells (2E) are more prominently found in the media. This finding is consistent with the known architecture of atherosclerotic vessels. There is only a slight demarcation between one smooth muscle cell and the next because they are so closely spaced and even overlapping in the media. The images demonstrate the high spatial resolution of this technique and show evidence of fenestration of the elastic lamina, a process known to occur with the development of atherosclerosis [Braunwald et al., 2001]. The fenestration can be observed in the Raman image of the internal elastic lamina (IEL), Figure 2D. The smooth muscle cells, shown in Figure 2E, can be seen migrating through the break in the IEL into the intima. Smooth muscle cell migration is a characteristic of atherosclerotic disease progression. In addition, one can identify a prominent collagen fiber (2F) in the media atop a diffuse connective tissue background, a feature that is difficult to fully appreciate from the phase contrast image.

Figure 3 shows Raman images of a normal human breast duct obtained using a morphological model [Shafer-Peltier et al., 2002] created specifically to analyze breast tissue (Fig. 3A–D). These images can be compared with those created by plotting the intensities of two Raman bands (Fig. 3E and F) characteristic of the DNA phosphate stretch (1,094 cm^{-1}) and the amide I band (1,664 cm^{-1}). The morphologically based Raman images represent the regions where a particular component (cell cytoplasm (3A), cell nucleus (3B), fat (3C), or collagen (3D)) contribute strongly to the spectrum (bright

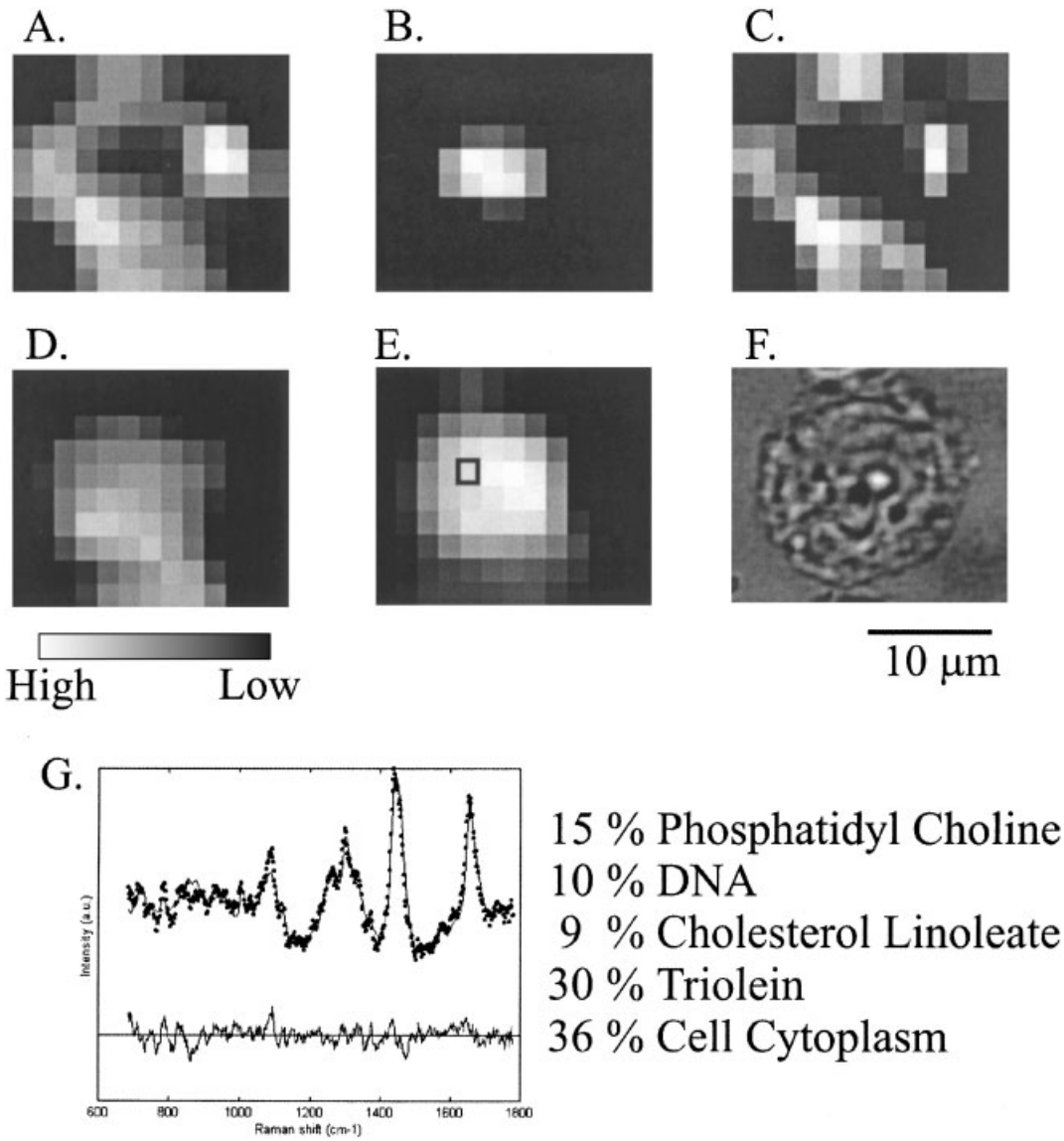

Fig. 1. Raman images (**A**–**E**) of HT29 cells with corresponding phase contrast image (**F**). Raman spectra are fit with phosphatidyl choline (A), DNA (B), cholesterol linoleate (C), triolein (D), and morphologically derived cell cytoplasm (E) spectra to produce chemical maps of the cells. **G**: shows the spectrum (·) acquired from within the box indicated in image E along with the corresponding fit (–) and residual (below, with zero line drawn). The fit contributions of each model element are listed to the side.

regions). Histological analysis of the tissue sample showed a normal breast duct with a diameter of approximately 25 μm. A typical breast duct of this size consists of a ring of epithelial cells surrounded by a basement membrane (primarily collagen). Within and surrounding the duct is some fat. The morphological model images clearly show the architecture of the duct, whereas the peak height images produced using the Raman bands found at 1,094 and 1,664 cm^{-1} are much less informative. Although the DNA phosphate stretch (1,094 cm^{-1}, Fig. 3E) should be found primarily in cellular regions, while the amide I band (1,664 cm^{-1}, Fig. 3F), indicative of protein, should be found mainly in collagenous regions, the images produced show neither the cellular component nor the collagen as clearly as the morphological model images do. This is because the amide I band can be found in many proteins, including those that form the cell cytoskeleton, whereas the phosphate stretch overlaps with bands present in the collagen spectrum. The inability of peak height analysis to accurately

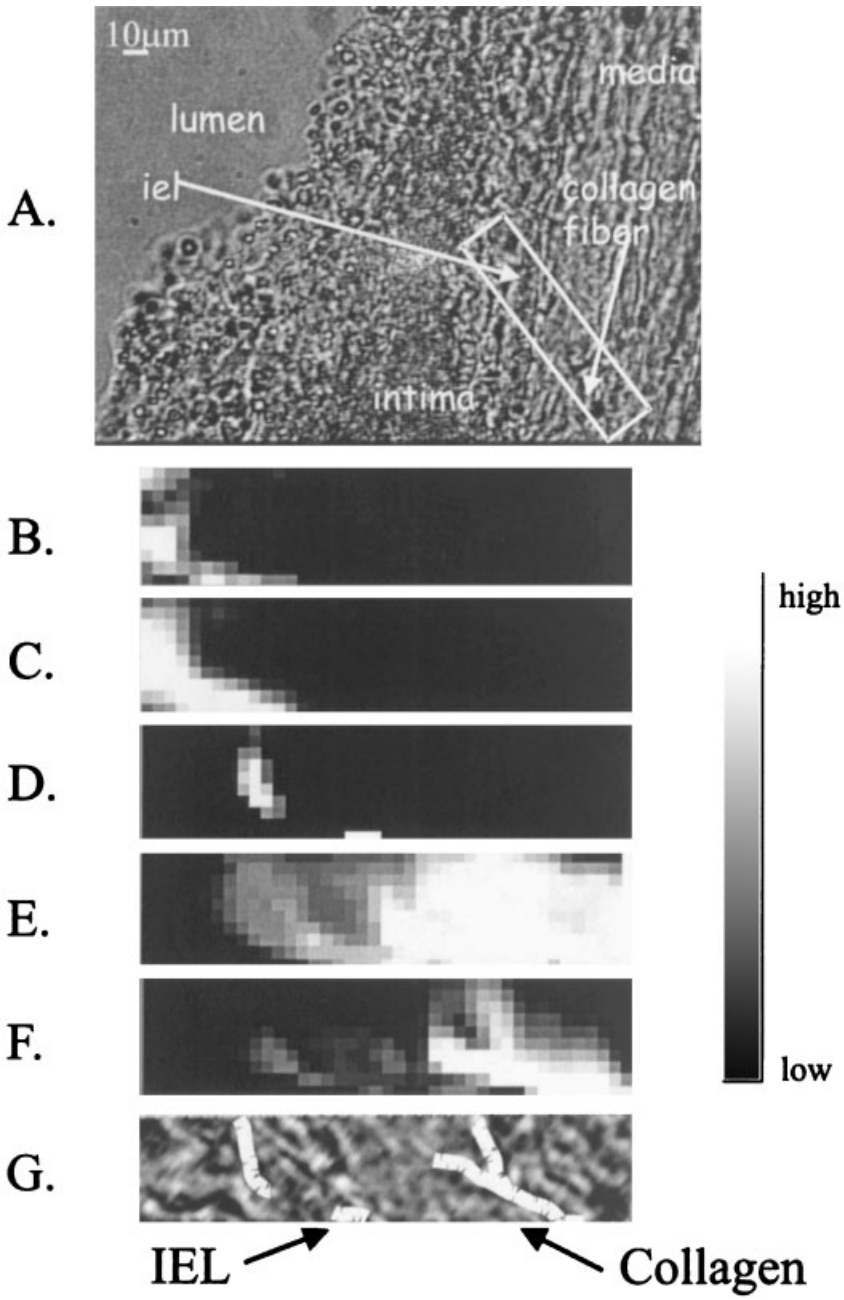

Fig. 2. Phase contrast images (**A** and **G**) of a mildly atherosclerotic artery, with the internal elastic lamina (IEL) and collagen fibers highlighted in G. Also shown are the Raman images of cholesterol (**B**), foam cells and necrotic core (**C**), IEL (**D**), smooth muscle cells (**E**), and collagen (**F**). Key morphological features, such as the fenestration of the IEL can be observed.

distinguish morphological features due to spectral overlap results in a much less informative image.

The Raman spectrum in Figure 3G represents a single point in the Raman image. The spectrum is a mixture of many chemical components, all of which contribute to the Raman spectrum. By fitting the spectrum with a morphological model it is possible to account for the major spectral features in the data. The residual of the fit, also shown in Figure 3G, is predominately noise, indicating that all of the information in the Raman imaging data hypercube can be represented by model-based images.

Although morphological modeling is an effective means of representing Raman images, it requires much advanced knowledge of the sample being studied. As discussed earlier, PCA, MCR, and Euclidean distance can also be used to compress the data into a manageable form and are much more effective when little is known about the system. Figure 4 shows a side-by-side comparison of PCA, MCR, Euclidean distance, and morphological modeling. The images, generated from the same data set, are of a sample of normal breast tissue containing three ductal units (mostly cells) surrounded by a collagen matrix. As can be seen, the images created by all four techniques are similar. The Euclidean distance images are shown as inverses (as they represent differences from input spectra rather than similarities as the other methods do) for easy comparison with the other techniques. On the left, the contributions attributable to collagen are shown, while on the right, the more subtle contributions of the cell nucleus (mostly DNA) are displayed. Both PCA (Fig. 4A) and MCR (Fig. 4B) were able to find seven independently varying basis spectra. Our complete morphological model for breast tissue has nine basis spectra, however, this includes several elements, such as microcalcifications, that are pathologically very important but that are observed only rarely in human breast tissue and not at all in this specimen [Haka et al., 2002].

The first two principal components, two of the spectra derived using MCR, and the collagen and cell nucleus basis spectra are shown in Figure 4E. The first principal component and the MCR spectra are similar to the collagen spectrum, the largest contributor to the image. The second principal component and the spectrum produced by MCR both contain some features of the cell nucleus spectrum (as negative peaks), but as can be seen from the image produced (Fig. 4A,B, right), they are much less effective at extracting the nuclear content within the ductal units than the morphological model (Fig. 4D, right). The filled-in rounded shape of the ductal units observed in

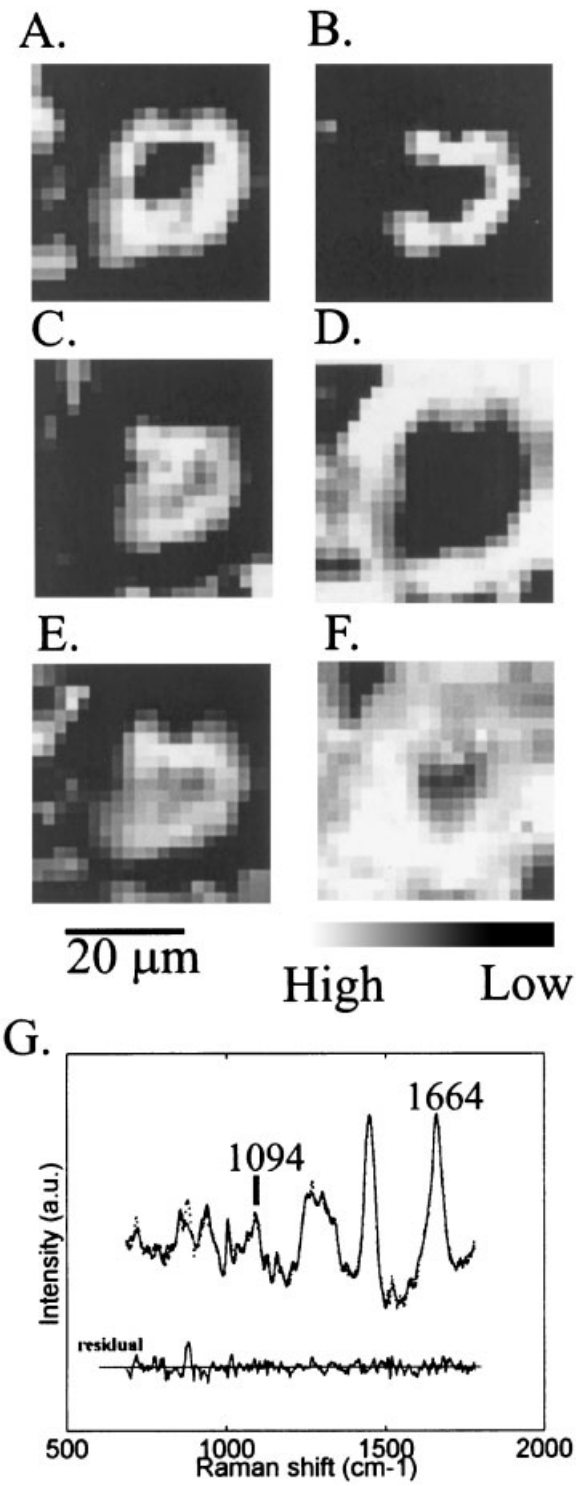

Fig. 3. Raman images of normal breast duct based on ordinary least-squares fitting of morphologically derived components: cell cytoplasm (**A**), cell nucleus (**B**), fat (**C**), and collagen (**D**). Images **E** and **F** plot the intensity of single bands: the DNA phosphate (1,094 cm^{-1}) and the protein-based amide I (1,664 cm^{-1}) peaks respectively. Demonstration of the fitting of a morphologically based model (·) to the spectrum of an individual pixel (located in a region with cellular content) in a Raman image (—) is shown in **G**. The residual of the fit is plotted below the spectrum (with the zero line drawn).

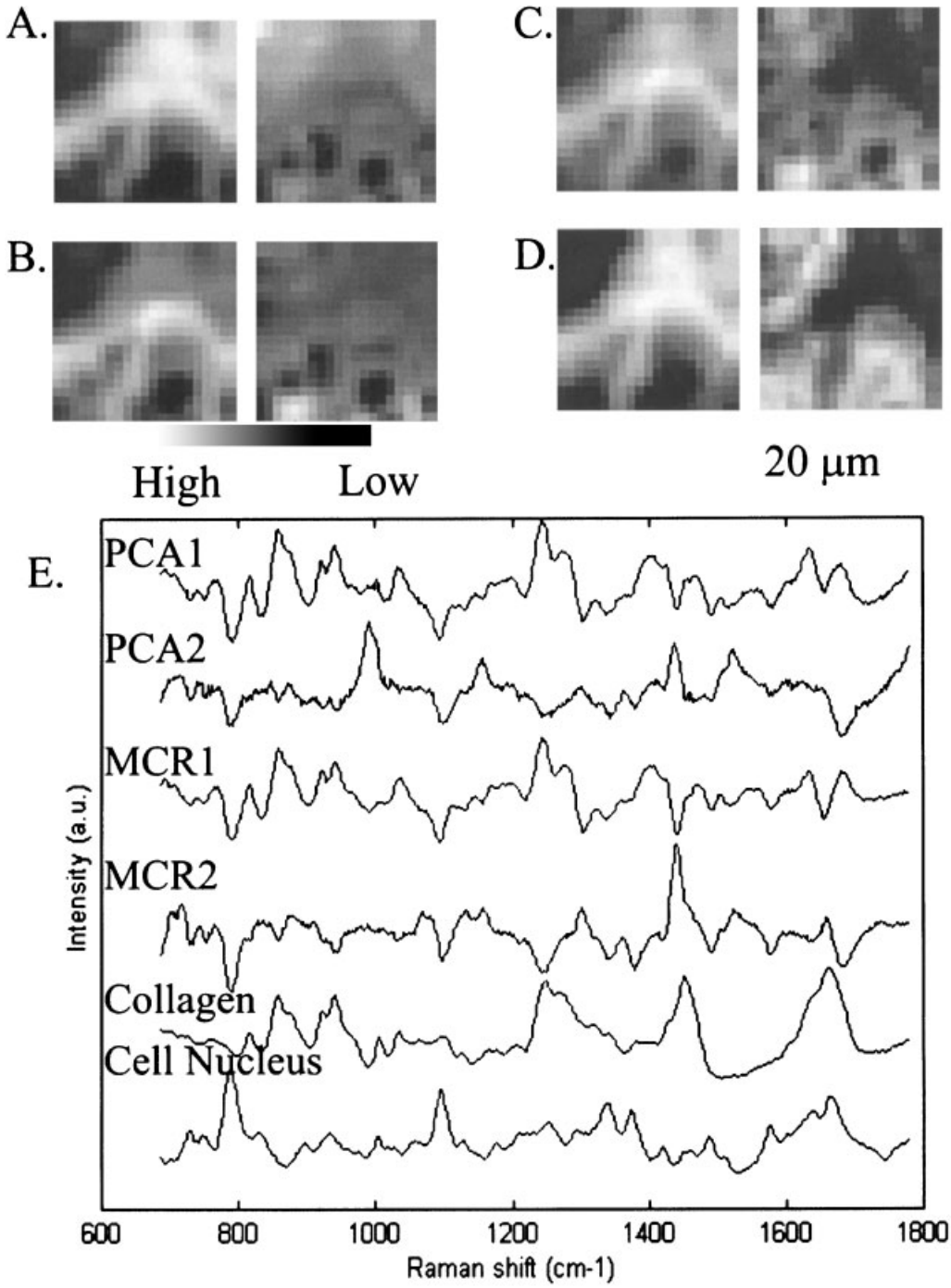

Fig. 4. Comparison of four different methods for analyzing Raman images of a region with multiple ductal units, separated by collagen. The images produced by the fit coefficients of the first two principal components are shown in **A**. **B**: This shows the two corresponding images produced by multivariate curve resolution (MCR). **C**: This shows images based on Euclidean distance, using the collagen (left) and cell nucleus (right) spectra from the morphological model. The images in **D** are produced using the fit coefficients produced by ordinary least-squares fitting with the morphological model, only collagen (left) and cell nucleus (right) are shown, but the complete model was used. **E**: shows the basis vectors used to create the images, from **top** to **bottom**: the first two principal components, the corresponding spectra produced by MCR, the morphologically derived spectrum of collagen and the morphologically derived spectrum of the cell nucleus. The last two spectra were used in both the Euclidean distance measurements and morphological modeling.

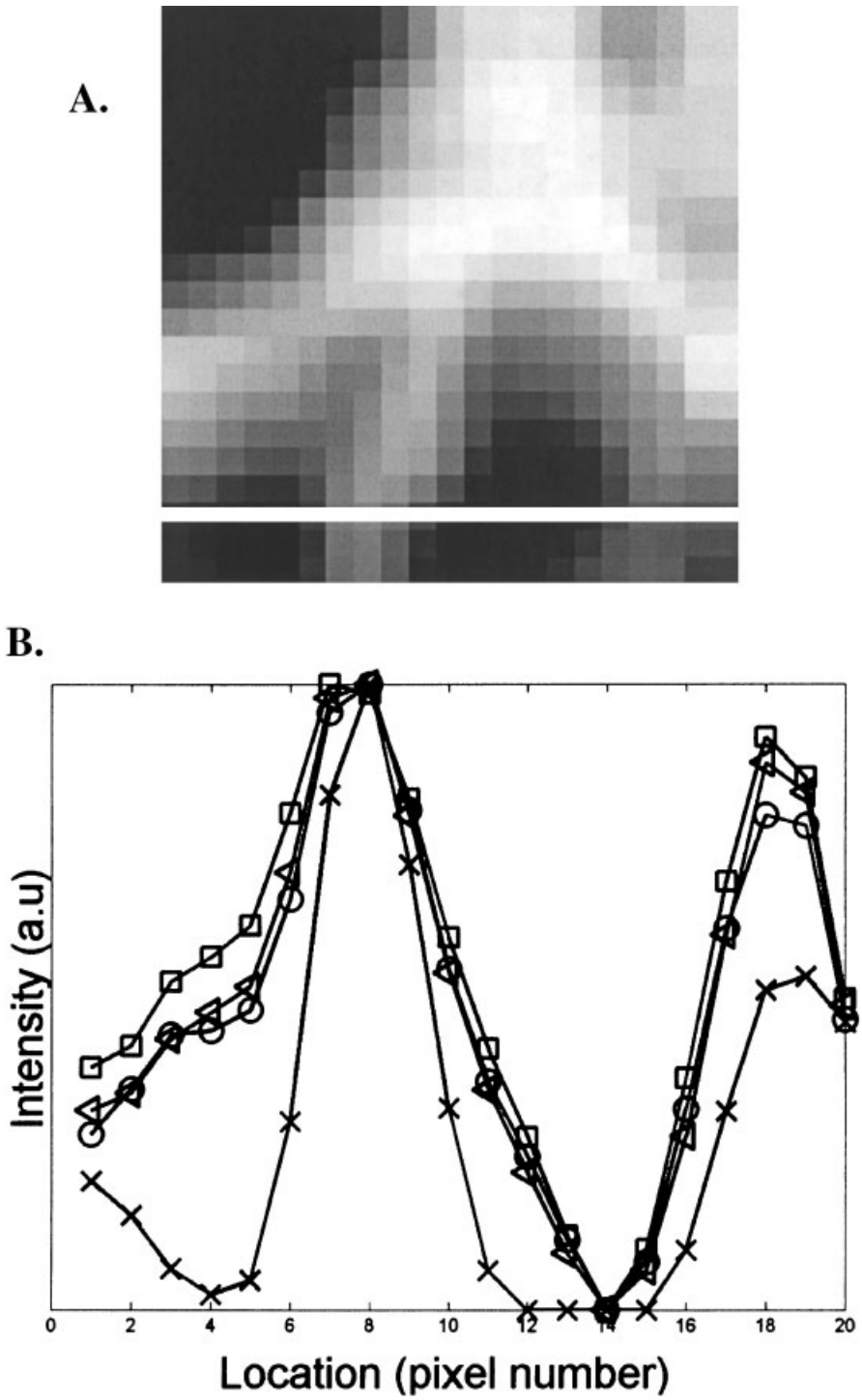

Fig. 5. **A**: Raman image (same as Fig. 4D, **left**) with third row indicated by white line and (**B**) heights for corresponding fit coefficients for the indicated row obtained using the four different models: PCA (△), MCR (□), Euclidean distance (○), and morphological model (X).

Figure 4D (right) is consistent with the pathology of this tissue slice.

Figure 5 shows the normalized fit coefficients of a particular row of the Raman images used in Figure 4A–D, left (row indicated in Fig. 5A). Although PCA (△), MCR (□), Euclidean distance (○), and morphological model (X) all display some form of transition from the

collagenous to cellular regions of the tissue, indicated by a change in the intensity of the fit coefficient, the transition is sharpest when using the morphological model. Therefore, not only does the morphological model provide information about more of the constituents of the sample (e.g., cell nucleus), but it also produces images with a higher resolution.

DISCUSSION

The simplest method for displaying a Raman image is to plot the intensity of a particular Raman band, or alternatively the ratio of two Raman bands. This method of analysis only takes advantage of a small portion of the data and because most biological samples contain many compounds with similar spectral features, is not applicable to biological systems. Spectral overlap makes it difficult to obtain structural or chemical information about a sample from a Raman image based solely on peak height.

In this article, four techniques which utilize the full spectrum for creating Raman images are presented: PCA, MCR, Euclidean distance, and morphological modeling. These imaging techniques are applicable not only to Raman, but also to many other spectroscopic imaging techniques, such as fluorescence. Each technique has its advantages and disadvantages. Some require no (PCA) or little (MCR) prior knowledge of the sample being studied, while others require some (Euclidean distance) or complete (morphological modeling) knowledge. The quality of the images produced is usually related to how much information is known.

PCA requires the least input from the user and consequently is the best tool for studying new types of samples. PCA is used to map out regions based on their spectral variance. Due to the mathematical process by which they are created, the principal components will always explain all of the spectral features present in the data. However, as the principal components themselves are mathematical constructs, they can be difficult or impossible to correlate with known chemicals. Despite this drawback, information gained from PCA can be used to build more sophisticated models, such as the morphological models developed for breast and artery tissues.

While MCR is also mathematically driven, non-negativity constraints can be applied to ensure that the basis spectra developed have more identifiable features than those produced by PCA. In fact, spectra determined using MCR can be very similar to the true chemical spectra. The disadvantage of MCR is that the more complex the system being studied, especially if there is much overlap in spectral features, the more difficult it is to perform the analysis. A skilled user can recognize when MCR has failed and adjust the parameters accordingly if the system is simple enough, but this too becomes more challenging as more component spectra are added to the sample mixture. In addition, as more curves are resolved in a complex system, noise plays a larger and larger role. Nonetheless, MCR is extremely useful for obtaining spectral lineshapes that can be used to direct further analysis of a sample.

When some, but not all, of the components of a sample are known, Euclidean distance is very effective. For example, it is not uncommon to have a sample in which the spectrum of the specific chemical being studied is known, but where the background chemicals are unknown. In this case, Euclidean distance can map the distribution of that particular chemical within the sample, unencumbered by the lack of knowledge of the background.

For detailed analysis of a system, especially for producing images with similar information content to pathology slides, morphological modeling is the best technique. However, development of a good morphological model can take time and requires much data acquisition in its own right. If the model is incomplete, the images will give less accurate information. Therefore, morphological modeling is best used when extensive studies are being performed and model development is a part of the experiment.

CONCLUSIONS

Raman spectral imaging is a powerful tool for determining chemical information in a biological specimen. The challenge is to capitalize on all of the spectral information, condensing it into an image with maximal information content. In this article, we introduced a new technique, morphological modeling, and reviewed three of the more common imaging approaches: PCA, MCR, and Euclidean distance. Each technique has its time and place. PCA and MCR are excellent for studying samples about which little is known a priori, whereas Euclidean distance can

produce improved images when some information about the sample is known. Using a morphological model, it is possible to obtain structural and chemical information about subcellular features, although it is ineffective if the system being studied is not well understood.

The ability to combine Raman confocal microscopy with imaging modalities to produce images of tissue or cells will be important for future biological studies. Although, Raman has only recently begun to be used as a tool for studying biological processes, because Raman is nondamaging at low laser powers, it will be used more and more to study biological samples, including live cells. Soon researchers will be monitoring sub-cellular processes using Raman imaging.

ACKNOWLEDGMENTS

The work was carried out at the MIT Laser Biomedical Research Center supported by NIH P41-RR 02594 grant. We thank Pathology Associates of University Hospitals for funding part of this study. Tissue was provided by the Cleveland Clinic Foundation and the Cooperative Human Tissue Network of the National Cancer Institute. HT29 cells were provided by Dr. Kamran Badizadegan (Department of Pathology, Children's Hospital and Harvard Medical School).

REFERENCES

Andrew JJ, Hancewicz TM. 1998. Rapid analysis of Raman image data using two-way multivariate curve resolution. Appl Spec 52:797–807.

Andrew JJ, Browne MA, Clark IE, Hancewicz TM, Millichope AJ. 1998. Raman imaging of emulsion systems. Appl Spec 52:790–796.

Appel R, Zerda TW, Waddell WH. 2000. Raman microimaging of polymer blends. Appl Spec 54:1559–1566.

Archibald DD, Kays SE, Himmelsbach DS, Barton FE. 1998. Raman and NIR spectroscopic methods for determination of total dietary fiber in cereal foods: A comparative study. Appl Spec 52:22–31.

Arikan S, Sands HS, Rodway RG, Batchelder DN. 2002. Raman spectroscopy and imaging of beta-carotene in live corpus luteum cells. Anim Reprod Sci 71:249–266.

Baxes GA. 1984. Digital image processing: A practical primer. Englewood Cliffs, NJ: Prentice-Hall.

Bock RK, Krischer W. 1998. The data analysis breifbook. Berlin: Springer-Verlag.

Braunwald E, Zipes DP, Libby P. 2001. Heart disease: A textbook of cardiovascular medicine. New York: W.B. Saunders Company.

Brennan JF, Romer TJ, Lees RS, Tercyak AM, Kramer JR, Feld MS. 1997. Determination of human coronary artery composition by Raman spectroscopy. Circulation 96:99–105.

Buschman HP, Deinum G, Motz JT, Fitzmaurice M, Kramer JR, van der Laarse A, Bruschke AV, Feld MS. 2001a. Raman microspectroscopy of human coronary atherosclerosis: Biochemical assessment of cellular and extracellular morphologic structures in situ. Cardio Pathol 10:69–82.

Buschman HP, Motz JT, Deinum G, Romer TJ, Fitzmaurice M, Kramer JR, van der Laarse A, Bruschke AV, Feld MS. 2001b. Diagnosis of human coronary atherosclerosis by morphology-based Raman spectroscopy. Cardio Path 10:59–68.

Clarke FC, Jamieson MJ, Clark DA, Hammond SV, Jee RD, Moffat AC. 2001. Chemical imaging fusion. The synergy of FT-NIR and Raman mapping microsocpy to enable a more complete visualization of pharamaceutical formulations. Anal Chem 73:2213–2220.

Delhaye M, Dhamelincourt P. 1975. Raman microprobe and microscope with laser excitation. J Raman Spectrosc 3: 33–43.

Freeman TL, Cope SE, Stringer MR, Cruse-Sawyer JE, Brown SB, Batchelder DN, Birbeck K. 1998. Investigation of the subcellular localization of zinc phthalocyanines by Raman mapping. Appl Spec 52:1257–1263.

Haka AS, Shafer-Peltier KE, Fitzmaurice M, Crowe J, Dasari RR, Feld MS. 2002. Identifying microcalcifications in benign and malignant breast lesions by probing differences in their chemical composition using Raman spectroscopy. Cancer Res 62:5375–5380.

Kneipp K, Haka AS, Kneipp H, Badizadegan K, Yoshizawa N, Boone C, Shafer-Peltier KE, Motz JT, Dasari RR, Feld MS. 2002. Surface-enhanced Raman spectroscopy in single living cells using gold nanoparticles. Appl Spec 56:150–154.

McCreery RL. 2000. Raman spectroscopy for chemical analysis. New York: John Wiley and Sons.

Nelson MP, Zugates CT, Treado PJ, Casuccio GS, Exline DL, Schlaegle SF. 2001. Combining Raman chemical imaging and scanning electron microscopy to characterize ambient fine particulate matter. Aerosol Sci Tech 34: 108–117.

Potter K, Kidder LH, Levin IW, Lewis EN, Spencer RGS. 2001. Imaging of collagen and proteoglycan in cartilage sections using Fourier transform infrared spectral imaging. Arthritis Rheum 44:846–855.

Schaeberle MD, Kalasinsky JL, Luke JL, Lewis EN, Levin IW, Treado PJ. 1996. Raman chemical imaging: Histopathology of inclusions in human breast tissue. Anal Chem 68:1829–1833.

Schaeberle MD, Tuschel DD, Treado PJ. 2001. Raman chemical imaging of microcrystallinity in silicon semiconductor devices. Appl Spec 55:257–266.

Shafer-Peltier KE, Haka AS, Fitzmaurice M, Crowe J, Myles J, Dasari RR, Feld MS. 2002. Raman microspectroscopic model of human breast tissue: Implications for breast cancer diagnosis in vivo. J Raman Spec 33:552–563.

Tauler R, Smilde AK, Henshaw JM, VBurgess LW, Kowalski BR. 1994. Multicomponent determination of chlorinated hydrocarbons using a reaction-based chemical sensor. 2. Chemical speciation using multivariate curve resolution. Anal Chem 66:3337–3344.

Timlin JA, Garden A, Morris MD, Rajachar RM, Kohn DH. 2000. Raman spectroscopic imaging markers for fatigue-related microdamage in bovine bone. Anal Chem 72: 2229–2236.

Turrell G, Corset J. 1996. Raman spectroscopy developments and applications. London: Academic Press.

van de Poll SWE, Bakker Schut TC, van der Laarse A, Puppels GJ. 2002. In situ investigation of the chemical composition of ceroid in human athersclerosis by Raman spectroscopy. J Raman Spec 33:544–551.

Wax A, Yang C, Backman V, Badizadegan K, Boone CW, Dasari RR, Feld MS. 2002. Cellular organization and substructure measured using angle-resolved low-coherence interferometry. Biophys J 82:2256–2264.

Wold S, Esbensen K, Geladi P. 1987. Principal component analysis. Chemometrics Intell Lab Sys 2:37–52.