Classification of atherosclerotic rabbit aorta samples by mid-infrared spectroscopy using multivariate data analysis

Liqun Wang
Georgia Institute of Technology
School of Chemistry and Biochemistry
Atlanta, Georgia 30332

Jessica Chapman
Richard A. Palmer
Duke University
Department of Chemistry
Durham, North Carolina 27708

Olaf van Ramm
Duke University
School of Biomedical Engineering
Durham, North Carolina 27708

Boris Mizaikoff
Georgia Institute of Technology
School of Chemistry and Biochemistry
Atlanta, Georgia 30332
E-mail: boris.mizaikoff@chemistry.gatech.edu

Abstract. Atherosclerotic and normal rabbit aorta samples show a marked difference in chemical composition governed by the water, lipid, and protein content. The strongly overlapping infrared absorption features of the different constituents, and the complexity of the tissue matrix, render tissue classification by direct evaluation of molecular spectroscopic characteristics obtained from IR reflectance or attenuated total reflectance (ATR) measurements virtually impossible. We apply multivariate analysis and classification techniques based on partial least squares regression (PLS) and linear discriminant analysis to IR spectroscopic data obtained by IR-ATR measurements and reflectance IR microscopy with high predictive accuracy during blind testing. Training data are collected from atherosclerotic and normal rabbit aorta samples. These results demonstrate the potential of IR spectroscopy combined with multivariate classification strategies for the *in-vitro* identification of normal and atherosclerotic aorta tissue. The prospect for future *in-vivo* measurement concepts is also discussed.
[DOI: 10.1117/1.2714030]

Keywords: atherosclerosis; partial least squares data analysis; principal components regression; Mahalanobis distance; infrared reflectance microscopy; attenuated total reflectance infrared spectroscopy.

Paper 05362RR received Dec. 3, 2005; revised manuscript received Nov. 19, 2006; accepted for publication Dec. 18, 2006; published online Mar. 26, 2007.

1 Introduction

Histochemical analysis is the classical method for studying atherosclerotic lesions and their pathophysiological progression. However, this method usually requires trained personnel for the sample preparation, which includes slicing artery wall tissue and staining for optical microscopy, rendering this procedure complex, time consuming, and limited to *in-vitro* conditions.

Optical spectroscopy is a powerful characterization tool sensitive to the variation of molecular components in the sample, and has been applied for rapid classification of cell and tissue samples.[1–13] Recent studies have shown that the vulnerability of atherosclerotic plaque largely depends on its chemical composition and ultrastructure. Different spectroscopic techniques, including fluorescence spectroscopy, Raman techniques, and near-infrared (NIR) spectroscopy, have been used to characterize normal tissues and plaques in human artery samples. Fluorescence spectroscopy has been used to study normal and atherosclerotic tissues based on endogenous or exogenous tissue chromophores,[14–27] successfully classifying normal and plaque artery tissues *in vitro*. In a more recent study, Marcu et al. demonstrated a catheter-based time-resolved fluorescence spectroscopic technique for *in-vivo* differentiating and demarking macrophage content versus collagen content in a rabbit atherosclerotic model.[28] Christov et al. have shown a catheter-based fluorescence emission analysis technique applied to the detection of Russell's viper venom-induced atherosclerotic plaque disruption in rabbit models during *in-vitro* and *in-vivo* studies.[29] The same fluorescence technique was also utilized for *in-vivo* analyzing of quantitative changes in collagen and elastin during arterial remodeling in rabbit models.[30] However, fluorescence techniques provide limited discriminatory information at a molecular level due to broad and frequently overlapping absorption and emission spectra obtained from tissue chromophores. Fourier-transform (FT) Raman with near-infrared (NIR) excitation has extensively been applied for qualitative and quantitative studies on the chemical composition of atherosclerotic plaques, and appears to be among the most promising techniques at present for the identification of vulnerable plaques.[31–42] Recently, van de Poll et al. applied Raman spectroscopy to studying the effects of diet and lipid-lowering therapy on plaque development in apoloprotein (APO) E^*3-Leiden transgenic mice.[43] Furthermore, *in-vivo* Raman spectroscopy techniques have gained importance for intravascular detection. The group of Boschman et al.[44] has utilized an *in-vivo* fiber optic probe for obtaining high-quality Raman spectra characterizing the artery wall in lambs and sheep. Fur-

Address all correspondence to: Boris Mizaikoff, School of Chemistry and Biochemistry, Georgia Institute of Technology, 770 State Street, Atlanta, GA 30332-0400; Tel: 404-894-4030; E-mail address: borismiz@aol.com

1083-3668/2007/12(2)/024006/11/$25.00 © 2007 SPIE

ther progress on *in-vivo* detection was achieved by Motz et al. demonstrating a fiber optic probe-based Raman system applied to real-time *in-vivo* collection of Raman spectra in the human atherosclerosis system.[45–47]

In addition, a variety of IR spectroscopic techniques including diffuse reflectance NIR spectroscopy,[48–51] conventional transmission Fourier transform infrared (FT-IR) spectroscopy,[52] attenuated total reflectance (ATR) spectroscopy,[53] and FT-IR microscopy[54] have been used for characterizing and identifying atherosclerotic plaques. A variety of spectroscopic mapping/imaging techniques, such as fluorescence,[55] Raman,[56,57] reflectance NIR,[58] transmission FT-IR microscopy,[59] and ATR FT-IR techniques[60] have also been used to characterize atherosclerotic plaques. Among these imaging techniques, micro-ATR FT-IR imaging,[60] as recently demonstrated by Colley et al., provides the inherent advantage of superior sensitivity and significantly faster data acquisition compared to Raman imaging techniques, and simultaneously higher resolution than other FT-IR-based imaging techniques. In this study, the cross section of atherosclerotic rabbit arteries is analyzed at a spatial resolution of 3 to 4 μm, thereby revealing the distribution heterogeneity of cholesterol esters in plaque. Consequently, among the optical techniques for studying atherosclerotic plaque, IR-ATR techniques are of particular interest due to their surface sensitivity and rapid data acquisition, which renders them ideal for thick and strongly absorbing materials such as tissue. In addition, ATR techniques are suitable for miniaturization, providing the potential to obtain spectroscopic signals and diagnostic information *in vivo*, if coupled with fiber optic signal delivery systems.

In our study, reflectance IR microscopy and IR-ATR spectroscopy have been applied for the investigation of normal and atherosclerotic rabbit aorta samples, in preparation for the development of an IR-ATR-based catheter system[61–63] for future *in-vivo* applications. All data presented in this study were obtained from intact aorta samples, and all spectra were generated from the inner surface of intima. Atherosclerotic and normal rabbit aorta samples show a significant difference in chemical composition governed by the water, lipid, and protein content. However, initial reflectance IR studies on hydrated rabbit aorta samples revealed that the difference between plaque and normal aorta tissue is very subtle due to averaging of the spectra within the measured area, as determined by the ATR element. Therefore, tissue classification by direct evaluation of the spectroscopic differences is virtually impossible for such IR-ATR catheter technology. Hence, instead of evaluating a few individual spectroscopic features for identification of rabbit aorta samples, multivariate data analysis strategies were adopted and applied to the spectral range of the data (900 to 4000 cm^1). Principle components analysis (PCA) was combined with Raman spectroscopy in a study by Deinum et al. to identify three classes of human coronary artery.[36] Discriminant analysis using Mahalanobis distance was applied to PCA scores extracted from Raman spectra of human artery tissue, enabling classification into three categories.[37] Cacheux et al. and Weinmann et al. coupled partial least square (PLS) regression with Raman spectroscopy for quantifying the cholesterol and cholesterol ester concentration in human and rabbit aorta tissue,[38,39] suitable for identifying lipid-rich plaques prone to disruption.

In the present study, we have successfully applied PLS discriminate analysis (PLS-DA) and linear discriminant analysis, along with Mahalanobis distance calculations, to data obtained by reflectance IR microscopy for the classification of lesion and nonlesion rabbit aorta tissue, demonstrating 100% predictive accuracy of the developed multivariate classification models during blind testing. Training data were collected from atherosclerotic and normal rabbit aorta samples. The spectra collected using the presently developed ATR FT-IR catheters[63] in our research group inherently present an average across a certain tissue area, defined by the contact area of the ATR element with the lesion or the aorta wall surface. However, the results in this study demonstrate that IR-ATR spectroscopy combined with multivariate classification techniques has the potential to identify normal and atherosclerotic aorta, which provides a sound basis for the development of *in-vivo* IR-ATR diagnostic devices.

2 Materials, Methods, and Multivariate Data Analysis

2.1 *Tissue Samples*

Five New Zealand White male rabbits were used to obtain the training sample set for building the classification models in this study: four were approximately 13 weeks old; one was approximately 6 months old. The six months old and one of the 13-weeks-old rabbits were fed a normal diet of rabbit chow. The remaining three rabbits were fed rabbit chow supplemented with 1% (w/w) cholesterol (Harlan Teklad, Indianapolis, Indiana) daily for eight weeks to induce atherosclerotic lesions.[64] One additional normal-fed and one additional cholesterol-fed rabbit (approximately 13 weeks old) were used to obtain the first set of test samples (12 in total) for validation of the established classification models. Two more normal-diet and two more cholesterol-fed rabbits (approximately 13 weeks old) were used to obtain the second set of test samples (56 in total) to further validate the classification models. Their weight and blood cholesterol levels were monitored every other week. For harvesting the aorta tissue, the rabbit was anesthetized and given an overdose of sodium pentobarbital. After euthanasia, the aorta tissue was excised and stored in 0.9% sodium chloride (NaCl) solution. Normal and atherosclerotic aortas (or aorta areas) were identified by visual inspection. Aortas from the rabbits on a normal diet appeared inconspicuous without evident lesions. One cholesterol-diet rabbit revealed lesion streak scattering along the inner wall of the aorta; two cholesterol-diet rabbits were characterized by atherosclerotic aortas, where the aorta inner wall was entirely covered by lesions. Tissue samples were cut into segments with a diameter of 4 mm using a biopsy device (Bio-punch, Health Link, Jacksonville, Florida) for spectroscopic measurement.

2.2 *Reflectance Infrared Microscopy*

Reflectance spectra (single beam), which were collected with an FT-IR spectrometer (Thermo Nicolet, Nexus 470, Thermo Electron Corporation, Somerset, New Jersey) coupled to an IR microscope (Spectra-Tech IR Plan, Vermont Optechs Incorporated, Charlotte, Vermont) were used as training data to build multivariate models for classifying lesion and nonlesion

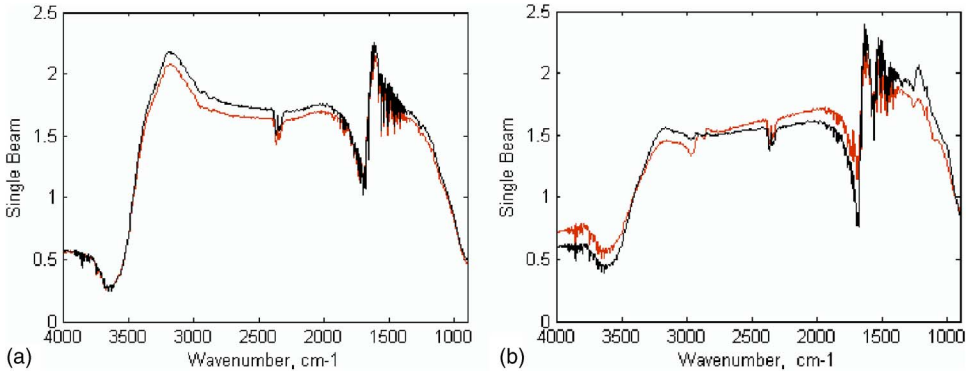

Fig. 1 (a) Red—average **a** spectrum of lesion samples; black—average **a** spectrum of nonlesion samples. (b) Red—average **e** spectrum of lesion samples; black—average **e** spectrum of nonlesion samples (Color online only).

aorta tissue. The biopsy sample (diam 4 mm) was placed on a glass slide, and the slide was positioned on the microscope stage. Spectra were collected at 4 cm^{-1} resolution from 650 to 4000 cm^{-1}, averaging 32 interferometer scans per measurement from a 100×100-μm spot. All lesion aorta samples were obtained from one of three cholesterol-diet rabbits; nonlesion aorta samples were prepared from the 13 months old normal-diet rabbit. A total of 14 biopsies from each kind of sample (lesion and nonlesion) were taken, and five IR reflectance spectra were recorded for each biopsy. The five measurements of each biopsy are denominated **a** through **e** in the remainder of this study. The first spectrum of each **a** through **e** set was measured three minutes after removal of the sample from the saline. The remaining spectra (**b** through **e**) were measured at 2-min intervals thereafter. By standardizing the data collection in this way, the effects of loss of water to evaporation were presumed to be reproducible from sample to sample for each spectrum **a** through **e**. Since the maximum penetration depth for MIR radiation into tissue is approximately 10 μm (or less in the presence of water), it can also be presumed that the reflectance signals obtained were generated entirely or at least predominantly from the intima.[6]

Two sets of test samples were independently investigated following the same procedure described before. The obtained data were then classified utilizing the multivariate classification models developed in the first phase of this study.

2.3 *Infrared Attenuated Total Reflectance Spectroscopy*

IR-ATR spectra were collected with a 45-deg single reflection diamond ATR accessory (Golden Gate, Specac Limited, Orrington, United Kingdom) in the same FT-IR spectrometer. In total, 29 dehydrated biopsy samples with a diameter of 4 mm were investigated, comprising ten lesion samples from the atherosclerotic aorta of the second cholesterol-diet rabbit, and ten nonlesion samples from the aorta of the 6 months old rabbit. The remaining nine samples were taken from nonlesion regions from the normal aorta regions of the third cholesterol-diet rabbit. Prior to the measurement, each biopsy sample was prepared by rinsing with deionized (DI) water, drying with lens paper, and then exposure to air for approximately 10 min. The dehydrated tissue samples were centered on the top of the circular diamond ATR element. To ensure sufficient contact between the tissue sample and the diamond, a constant pressure was applied via a built-in adjustable plunger. Spectra were collected at 4-cm^{-1} resolution from 400 to 4000 cm^{-1}, averaging 16 spectra per measurement.

2.4 *Multivariate Data Analysis*

PLS_Toolbox_3.5 (Eigenvector Incorporated, Wenatchee, Washington) was used to generate the classification models. Principal components regression (PCR), partial least squares (PLS), partial least squares linear discriminant analysis (PLS-DA), and Mahalanobis distance were applied on the first (**a**) and last (**e**) spectra of each dataset obtained with IR reflectance microscopy, and on hydrated and dehydrated tissue data obtained with the IR-ATR method. The obtained spectra for each particular set of experiments were always mean centered prior to multivariate analysis. Cross-validation (leaving one sample out) was performed to determine the optimal number of principal components (PC) or latent variables (LV).

3 Results and Discussion

3.1 *Reflectance Infrared Microscopy*

3.1.1 *Average spectra of classification data*

Average spectra of the first (**a**) and last (**e**) measurements of the training set aorta samples are shown in Fig. 1. From these plots, it is clearly evident that the spectral differences between lesion and nonlesion tissue samples are very subtle. The experimental results obtained in this study convincingly demonstrate that sophisticated multivariate data analysis and classification techniques are essential to robust and reliable sample classification for diagnostic purposes.

3.1.2 *Multivariate classification results using* **a** *data*

In the following multivariate classification, lesion samples were assigned class 1, and nonlesion samples class 2. IR reflectance spectra were preprocessed by meancentering prior to further analysis.[65]

PLS-DA is a discrimination method developed from PLS regression models.[66] Based on the root mean square error for cross validation (RMSECV) results for PLS-DA shown in Fig. 2(a), four latent variables (LVs) are selected as optimal numbers to minimize error during classification and prediction. In general, four or six LVs were tested to build the sta-

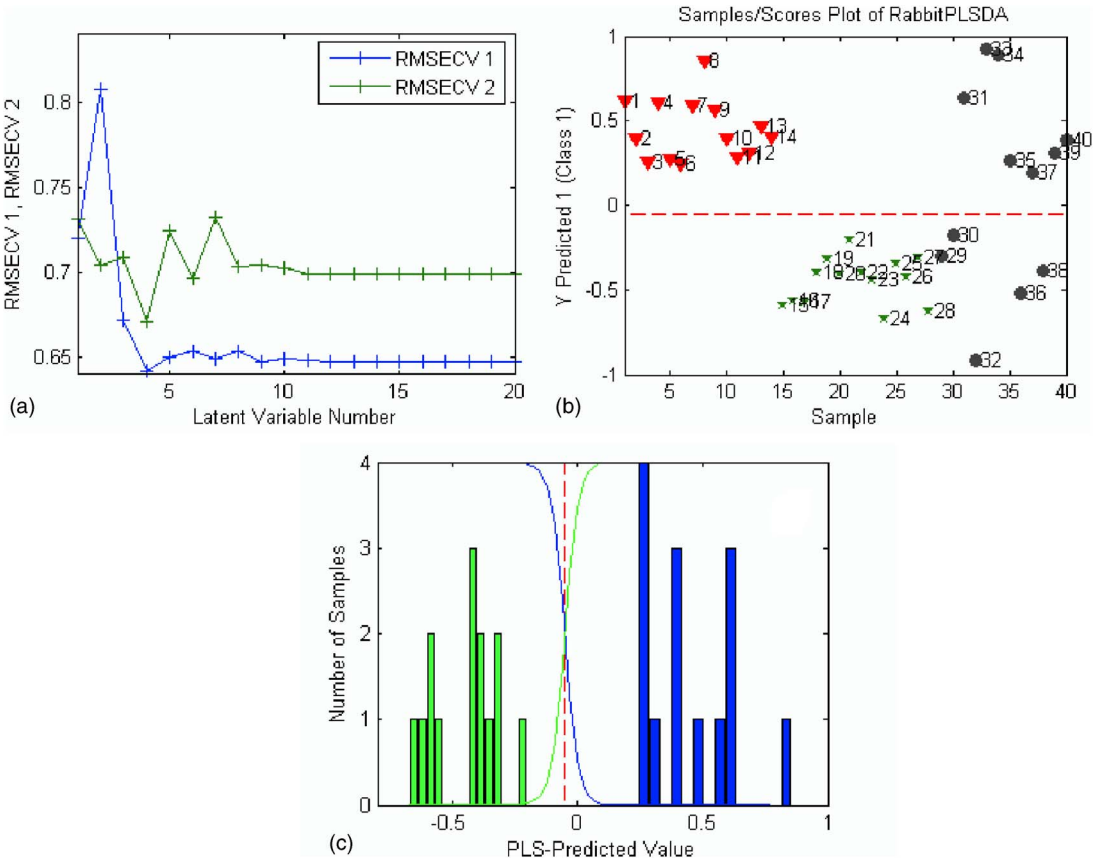

Fig. 2 (a) RMSECV versus LV number using **a** data of training set samples. RMSECV—root mean square error for cross-validation. Class 1 (blue)—lesion samples; class 2 (green)—nonlesion samples. The minimum theoretically indicates the optimum number of LVs to build the model. (b) Classification and prediction results for PLS-DA model 6 LVs using **a** data. Red triangles—lesion training samples; green stars—nonlesion training samples; black dots—blind samples; red line—threshold (−0.0507). (c) Histograms for PLS-DA 6 LVs model using **a** data of training set samples. Threshold is −0.0507 (Color online only).

tistical models. The corresponding classification and prediction results are shown in Figs. 2(b) and 2(c). Ideally, lesion samples have a value of 0.5, and nonlesion samples have a value of −0.5. However, the predicted values frequently deviate from the ideal hit values due to the variations of the samples within the same class.

In all plots shown next, points 1 to 14 represent lesion training samples (class 1); 15 to 28 nonlesion training samples (class 2); and 29 to 40 samples from the first test set. The establishment of the model using the training samples (1 to 28) by Wang preceded the measurement of the unknown samples (by Chapman) by six months owing to tissue availability schedules. For the 12 samples from the first test set, only the raw single beam IR spectra were provided for evaluation without any indication of the number of lesion versus nonlesion cases among the 12 samples. The identity of the test samples was shared only after the classification had been made.

Threshold values were calculated using the observed distribution of the predicted values and the Bayesian theorem for discriminating the two different classes. As shown in Fig. 2(c), blue bars are a histogram of the predicted values for class 1 samples; green bars are a histogram of the predicted values for class 2 samples. The threshold is the cross point of two normally fitted histograms.

The Bayesian statistics also provide the probability that a sample is a member of a certain class given the predicted value. The prediction probability results for both four LV and six LV PLS-DA models based on all investigated samples are shown in Table 1. Given a sample, its probability belonging to class 1 is calculated using Eq. (1).

$$\text{probability(class 1)} = \frac{P(y,1)}{[P(y,1) + P(y,2)}, \qquad (1)$$

where y is the predicted value from the PLS-DA model for the sample in question, $P(y,1)$ is the probability of this sample being a member of class 1 given the value of y, and $P(y,2)$ is the probability of this sample being a member of class 2 given the value of y. Consequently, a sample with a predicted value at the threshold has a 50% probability belonging to either class.

In the model using four LVs, sample 10 cannot be unambiguously classified, but its probability of belonging to class 1 is >50% (see Table 1). Using this model, only test sample 30 was incorrectly classified. If six LVs were applied to establish the model, all samples could be correctly classified or predicted.

Linear discrimination analysis (LDA) used before is a method to maximize the among-class difference relative to the

Table 1 Prediction probability results for PLS-DA models using **a** data. 1 to 28: training sample set; 1 to 14: lesion sample set; 15 to 28: nonlesion sample set; and 29 to 40: first set of test samples.

Sample	Revealed class	Prediction probability			
		Four LVs		Six LVs	
		Class 1	Class 2	Class 1	Class 2
1	1	0.9848	0.0152	1	0
2	1	0.9921	0.0079	1	0
3	1	0.9274	0.0726	1	0
4	1	0.9995	0.0005	1	0
5	1	0.9489	0.0511	1	0
6	1	0.8146	0.1854	1	0
7	1	0.9959	0.0041	1	0
8	1	1	0	1	0
9	1	0.9989	0.0011	1	0
10	1	0.5937	0.4063	1	0
11	1	0.9682	0.0318	1	0
12	1	0.9267	0.0733	1	0
13	1	0.9967	0.0033	1	0
14	1	0.9778	0.0222	1	0
15	2	0.2024	0.7976	0	1
16	2	0.0017	0.9983	0	1
17	2	0.0001	0.9999	0	1
18	2	0.0413	0.9587	0	1
19	2	0.1192	0.8808	0.0001	0.9999
20	2	0.0146	0.9854	0	1
21	2	0.0001	0.9999	0.005	0.995
22	2	0.0185	0.9815	0	1
23	2	0.3082	0.6918	0	1
24	2	0.0011	0.9989	0	1
25	2	0.0218	0.9782	0.0001	0.9999
26	2	0.0069	0.9931	0	1
27	2	0.324	0.676	0.0002	0.9998
28	2	0.0241	0.9759	0	1
29	2	0.0991	0.9009	0.0002	0.9998
30	2	1	0	0.0126	0.9874
31	1	0.9999	0.0001	1	0
32	2	0.0001	0.9999	0	1
33	1	1	0	1	0
34	1	1	0	1	0
35	1	0.9767	0.0233	1	0
36	2	0.0044	0.9956	0	1
37	1	0.7582	0.2418	0.9999	0.0001
38	2	0.0006	0.9994	0	1
39	1	0.9957	0.0043	1	0
40	1	0.9992	0.0008	1	0

within-class difference. The Mahalanobis distance[67,68] is a specific linear discriminant analysis method particularly suitable for classification, which was performed here by first compressing the spectral data to six latent variables and corresponding scores of a 6-D vector using PLS. Following this, the mean score vector $\mathbf{S}_{mn}$ and the mean-centered scores $\mathbf{S}_{mc}$ for each class (lesion or nonlesion) were calculated, and the covariance matrix (6×6) $\mathbf{M}$ of $\mathbf{S}_{mc}$ for each class was computed. For the prediction of a blind sample, its score would be calculated from the measured spectrum and latent variables, and mean centered by $\mathbf{S}_{mn}$ of one class. The distance $\mathbf{D}_j^2$ of the mean-centered unknown score $\mathbf{t}_j$ from $\mathbf{S}_{mn}$ of this class was computed and normalized by $\mathbf{M}$ following Eq. (2).

$$D_j^2 = (t_j)M^{-1}(t_j)', \tag{2}$$

where $M = S'_{mn}S_{mn}/m - 1$, with m indicating the number of training samples in one class.

The distance of an unknown sample to the classes determines which class the unknown belongs to. The class that has less distance to the unknown will incorporate the unknown sample. From Fig. 3 it is evident that the Mahalanobis distance method has provided 100% successful classification and prediction of all samples in the first test set, similar to PLS-DA.

3.1.3 *Multivariate classification results using* **e** *data*

Based on the RMSECV results (not shown), six LVs have been determined as the optimal number for the PLS-DA classification model. The corresponding classification results are shown in Fig. 4(a). All training samples could be clearly classified with this method, and only test sample 40 could not be classified with sufficient certainty. Most probably, it would be incorrectly classified as a nonlesion sample.

The corresponding histograms and the prediction probability results for the PLS-DA model using six LVs and **e** data are shown in Fig. 4(b) and Table 2.

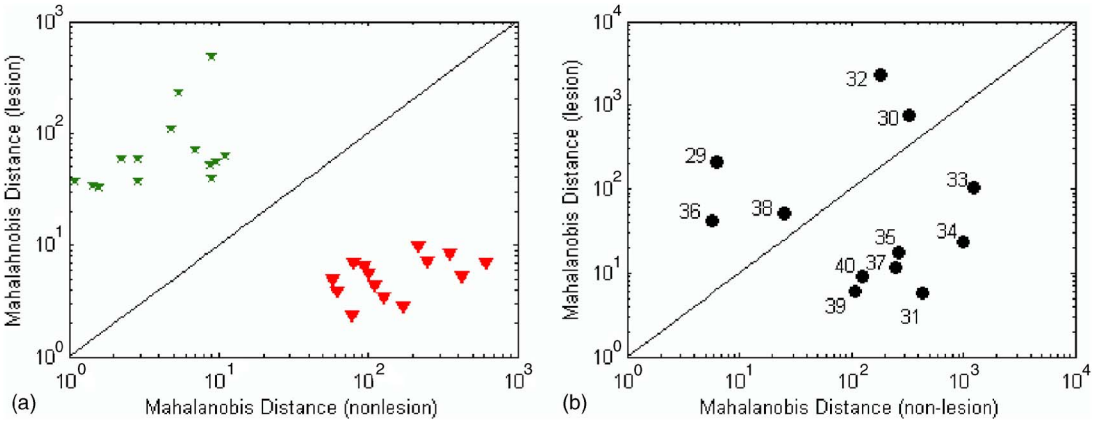

Fig. 3 (a) Classification results of 28 training samples using the Mahalanobis distance method and **a** data of training set samples. Green stars—nonlesion training samples; red triangles—lesion training samples; diagonal line—discriminant line. (b) Prediction results of 12 blind samples using the Mahalanobis distance method and **a** data of blind samples. Diagonal line—discriminant line (Color online only).

The Mahalanobis distance method was also applied to classify **e** data. The classification results are shown in Fig. 5. Again, test sample 40 could not be correctly classified. Sample 30 could be classified more clearly using the Mahalanobis distance in contrast to using PLS-DA.

Using the PLS-DA models developed before, classification of the second set of test samples was attempted, however, with reduced hit quality using both **a** and **e** data. Yet 74% of the samples were classified correctly using **a** data, and 60% of the samples were classified correctly using **e** data. The sensitivity and specificity of the PLS-DA model for the test samples were calculated using the method introduced by Balchum et al.,[69] and are summarized in Table 3. The lower classification rate is attributed to the limited diversity of the training sample set for developing the predictive models. Consequently, it is essential for introducing significantly more spectra for covering the variation among animals by using spectral imaging techniques, enabling the collection of large sets of model data in a reasonable period of time.

The possible reason that using **a** data provides (marginally) more accurate predictive results in contrast to using **e** data may result from the fact that the sample had significantly changed during ambient exposure and the experimental procedure. It has to be considered that the **e** dataset has been recorded as the fifth consecutive measurement starting after 11 min of an entire measurement series. Hence, due to water evaporation the sample was significantly drier compared to the beginning of the measurement series. In turn, this indicates that classification during hydrated conditions, which more closely resemble the *in-situ* environment, is more accurate.

Alternatively, the application of principal components regression (PCR) techniques was investigated for the **a** and **e** data series of the first test set to discriminate between lesion and nonlesion classes. 1 was the preset value for all lesion samples, and 0 was for all nonlesion samples.[70] All spectra were again mean centered prior to PCR. The predicted lesion value ideally centers at 0.5, and the nonlesion at −0.5. However, PCR-based classification failed in accurately classifying **a** data. Figure 6 shows the PCR results using **e** data. A total of nine PCs were selected for the model, and all training samples could be accurately classified. Test sample 40 was incorrectly classified as nonlesion, similar to PLS-DA and the Mahalanobis distance method. In addition, test sample 35 could not be

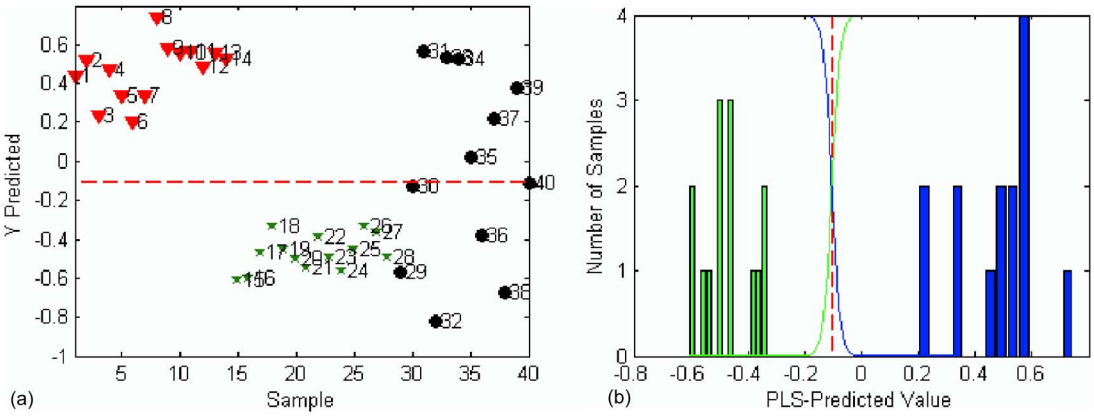

Fig. 4 (a) Classification and prediction results for PLS-DA 6 LVs model using **e** data. red triangles—lesion training samples (class 1); green stars—nonlesion training samples (class 2); black dots—blind samples; red line—threshold (−0.1042). (b) Histogram for PLS—DA 6 LVs model using **e** data. Threshold is −0.1042 (Color online only).

Table 2 Prediction probability results for PLS-DA models using **e** data. 1 to 28: training sample set; 1 to 14: lesion sample set; 15 to 28: nonlesion sample set; and 29 to 40: first set of test samples.

		Prediction probability	
	Revealed	Six LVs	
Sample	class	Class 1	Class 2
1	1	1	0
2	1	1	0
3	1	1	0
4	1	1	0
5	1	1	0
6	1	1	0
7	1	1	0
8	1	1	0
9	1	1	0
10	1	1	0
11	1	1	0
12	1	1	0
13	1	1	0
14	1	1	0
15	2	0	1
16	2	0	1
17	2	0	1
18	2	0	1
19	2	0	1
20	2	0	1
21	2	0	1
22	2	0	1
23	2	0	1
24	2	0	1
25	2	0	1
26	2	0	1
27	2	0	1
28	2	0	1
29	2	0	1
30	2	0.1634	0.8366
31	1	1	0
32	2	0	1
33	1	1	0
34	1	1	0
35	1	0.9999	0.0001
36	2	0	1
37	1	1	0
38	2	0	1
39	1	1	0
40	1	0.3682	0.6318

clearly predicted with the horizontal zero line as the discriminator, as the prediction value was only slightly above zero.

In contrast to PCR, the PLS-DA method not only considers the changes in the spectra, but instantaneously also considers the changes in concentration of the various constituents (or class difference in our case). Due to uncertainties introduced by the sample preparation process and ambient effects during the measurements, the among-group difference is not always larger than the within-group difference. Hence, PCR appeared to be the least able to provide satisfactory classification results.

The sensitivity and specificity of the investigated multivariate methods for test samples of the first test set without any *a priori* knowledge are summarized and compared in Table 3.

3.2 *Infrared Attenuated Total Reflectance Spectroscopy*

The average spectra for lesion samples and for nonlesion samples using single reflection ATR spectroscopy are shown in Fig. 7. Spectral differences are most evident in the region 2700 to 3000 cm^{-1}. However, spectra collected from individual dehydrated nonlesion samples also show relatively strong absorptions in the spectral region of 2700 to 3000 cm^{-1} and at approximately 1650 cm^{-1}. These characteristics appear smoothed out in the average spectra, and the classification of these samples might render difficult if these spectral features are used as only identifiers. Therefore, chemometric analysis is essential for obtaining reliable tissue classification models.

PLS-DA was applied on IR-ATR data after preprocessing of the spectra by mean centering. Lesion samples were assigned class 1, and nonlesion samples class 2. Five LVs were selected for building of PLS-DA classification model. The corresponding classification results are shown in Fig. 8(a). The prediction probability calculated using the Bayesian theorem is 1 for all tissue samples. Alternatively, the Mahalanobis

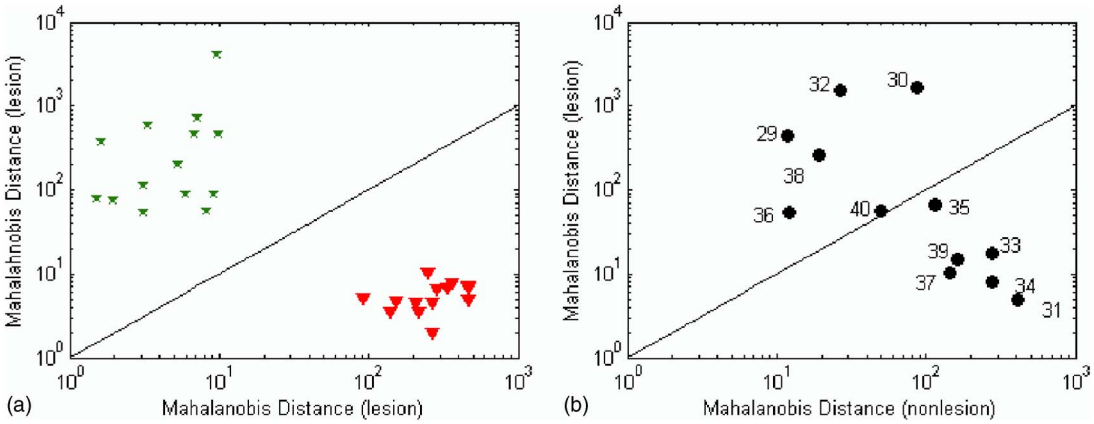

Fig. 5 (a) Classification results of 28 training samples using the Mahalanobis Distance method and **e** data. Green stars—nonlesion training samples; red triangles—lesion training samples; diagonal line—discriminant line. (b) Prediction results of 12 blind samples based on the PLS-DA 6 LVs model for **e** data (Color online only).

distance was applied on dehydrated tissue data collected using IR-ATR. The classification results based on the five latent variables derived from the PLS-DA are shown in Fig. 8(b).

Alternatively to building a model using all samples in the dataset as training samples, the data were separated into a training set and a validation set. The validation set was used to test the robustness of model established with the training set. This operation was performed five times, each time with a different set of five or six samples selected as validation data (two lesion, and three or four nonlesion samples). The remaining 23 or 24 samples were used as training data. Eventually, each sample was selected into the validation dataset once, and tested once. Five LVs were applied for all five calibration models, similar to the model using all data. All five models turned out sufficiently robust and predicted the corresponding validation samples with 100% hit quality. Alternatively, PCR was tested also on the IR-ATR samples; however, it failed to accurately classify the samples, as previously discussed.

4 Conclusion

PLS-DA (or PLS) and Mahalanobis distance linear discriminant analysis methods are applied to mid-infrared microspecular reflectance data and mid-infrared ATR data of lesion and nonlesion biopsy samples of rabbit aorta. Both methods achieve 100% hit quality with outstanding sensitivity and specificity during tests on small sets of samples. More diverse test sets reveal that larger training datasets, such as those provided by IR imaging techniques, are required for accurate classification, although up to 89% correct classification results are obtained. Consequently, the overall results reveal a promising prospect for successful classification of lesion versus nonlesion tissue samples. The fundamentals of the approach presented in this study are currently being expanded and tested with an IR-ATR catheter system for future *in-vivo* diagnostics during plaque ablation.[63]

Acknowledgments

The authors would like to thank Aya Eguchi (Duke University) for support during the data collection and discussion, and Ellen Dixon-Tulloch (Duke University) for rabbit care, euthanasia, and tissue harvesting. This study was in part supported by NIH grant R01 HL067111 and R01 EB000508.

References

1. K. Stehfest, J. Toepel, and C. Wilhelm, "The application of micro-FTIR spectroscopy to analyze nutrient stress-related changes in biomass composition of phytoplankton algae," *Plant Physiol. Bioch.* **43**, 717–726 (2005).
2. S. Kim, B. Reuhs, and L. Mauer, "Use of fourier transform infrared

Table 3 Sensitivity and specificity of the investigated multivariate data analysis methods for training and test samples.

	Sensitivity,%				Specificity, %			
	PLS-DA(6LV) and M distance		PCR		PLS-DA(6LV) and M disance		PCR	
	a	**e**	**a**	**e**	**a**	**e**	**a**	**e**
Training samples	100	100	NA	100	100	100	100	100
First set of testing samples	100	85.7		71.4	100	100	NA	100
Second set of testing samples	60.7	12.4		NA	78.6	89.3		NA

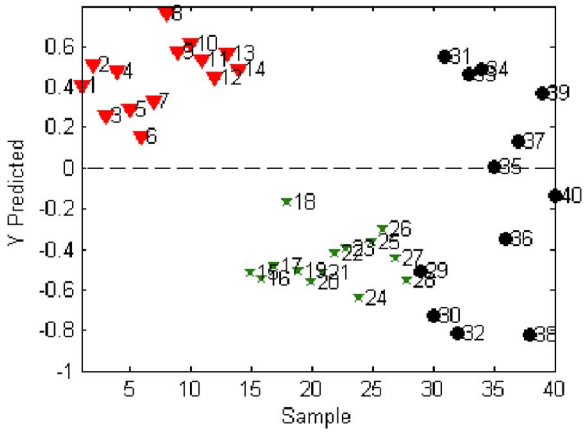

Fig. 6 Classification and prediction results for PCR model with 9 PCs using **e** data. Red triangles—lesion training samples (class 1); green stars—nonlesion training samples (class 2); black dots—blind samples (Color online only).

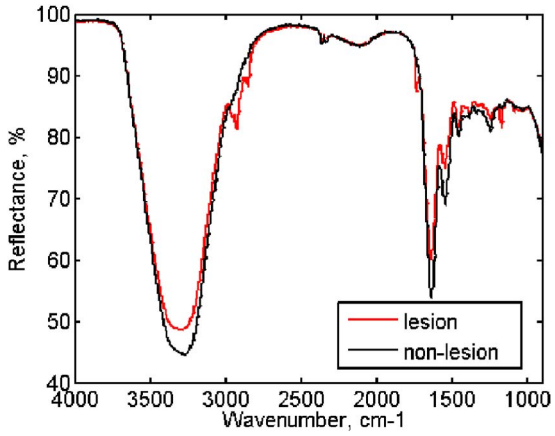

Fig. 7 Red—average of dehydrated lesion sample spectra using IR-ATR; black—average spectrum of dehydrated nonlesion samples using IR-ATR (Color online only).

spectra of crude bacterial lipopolysaccharides and chemometrics for differentiation of Salmonella enterica serotypes," *J. Appl. Microbiol.* **99**, 411–417 (2005).
3. J. Moutant, K. Short, S. Carpenter, N. Kunapareddy, L. Coburn, M. Tamara, and J. Freyer, "Biochemical differences in tumorigenic and nontumorigenic cells measured by Raman and infrared spectroscopy," *J. Biomed. Opt.* **10**, 031106/1–031106/15 (2005).
4. M. Alam, J. Timlin, L. Martin, D. Willians, C. Lyons, K. Garrison, and B. Hjelle, "Spectroscopic evaluation of living murine macrophage cells before and after activation using attenuated total reflectance infrared spectroscopy," *Vib. Spectrosc.* **34**, 3–11 (2004).
5. C. Krafft, S. Sobottka, G. Schackert, and R. Salzer, "Analysis of human brain tissue, brain tumors and tumor cells by infrared spectroscopic mapping," *Analyst (Cambridge, U.K.)* **129**, 921–925 (2004).
6. E. Gazi, J. Dwyer, P. Gardner, A. Ghanbari-Siahkali, A. Wade, J. Miyan, N. Lockerman, N. Clarke, J. Shanks, L. Scott, C. Hart, and M. Brown, "Applications of Fourier transform infrared microspectroscopy in studies of benign prostate and prostate cancer. A pilot study," *J. Pathol.* **201**, 99–108 (2003).
7. G. Steiner, A. Shaw, L. Choo-smith, M. Abuid, G. Schackert, S. Sobottka, W. Steller, R. Sazlzer, and H. Mantsch, "Distinguishing and grading human gliomas by IR spectroscopy," *Biopolymers* **72**, 464–471 (2003).
8. A. Dalman, V. Erukhimovitch, M. Talyshinsky, M. Huleihil, and M. Huleihel, "FTIR spectroscopic method for detection of cells infected with Herpes viruses," *Biopolymers* **67**, 406–412 (2002).
9. M. Mossoba, F. Khambaty, and F. Fry, "Novel application of a disposable optical film to the analysis of bacterial strains: a chemometric classification of mid-infrared spectra," *Appl. Spectrosc.* **56**, 732–736 (2002).
10. P. Lasch, W. Haensch, E. Lewis, L. Kidder, and D. Naumann, "Characterization of colorectal adenocarcinoma sections by spatially resolved FT-IR microspectroscopy," *Appl. Spectrosc.* **56**, 1–9 (2002).
11. L. Mclntosh, M. Jackson, H. Mantsch, M. Stranc, D. Pilavdzic, and A. Crowson, "Infrared spectra of basal cell carcinomas are distinct from non-tumor-bearing skin components," *J. Invest. Dermatol.* **112**, 951–956 (1999).
12. C. Schultz and H. Mantsch, "Biochemical imaging and 2D classification of keratin pearl structures in oral squamous cell carcinoma," *Cell Mol. Biol. (Paris)* **44**, 203–210 (1998).
13. M. Nilson, D. Heinrich, J. Olajos, and S. Andersson-Engels, "Near infrared diffuse reflection and laser-induced fluorescence spectroscopy for myocardial tissue characterization," *Spectrochim. Acta, Part A* **51**, 1901–1912 (1997).
14. L. I. Deckelbaum, J. K. Lam, H. S. Cabin, K. S. Clubb, and M. B. Long, "Discrimination of normal and atherosclerotic aorta by laser-induced fluorescence," *Lasers Surg. Med.* **7**, 330–335 (1987).
15. L. I. Deckelbaum, M. L. Stetz, K. M. O'Brien, F. W. Cutruzzola, A. F. Gmitro, L. I. Laifer, and G. R. Gindi, "Fluorescence spectroscopy guidance of laser ablation of atherosclerotic plaque," *Lasers Surg.*

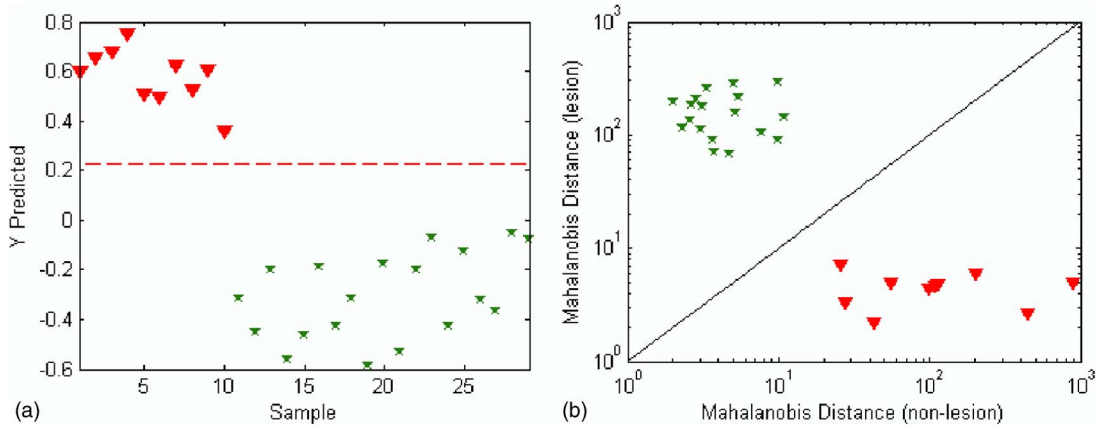

Fig. 8 (a) Classification results for PLS-DA 5 LVs model on dehydrated data using IR-ATR. Red triangles—lesion samples (class 1); green stars—nonlesion samples (class 2); red line—threshold (0.2234). (b) Classification results of 29 dehydrated training samples measured using IR-ATR based on the Mahalanobis distance method. Green stars—nonlesion samples; red triangles—lesion samples; diagonal line—discriminant line (Color online only).

Med. **9**, 205–214 (1989).
16. M. Leon, D. Lu, L. Prevosti, W. Macy, P. Smith, M. Granovsky, R. Bonner, and R. Balaban, "Human arterial surface fluoresce: atherosclerotic plaque identification and effects of laser atheroma ablation," *J. Am. Coll. Cardiol.* **12**, 94–102 (1988).
17. L. Laifer, K. O'Brien, M. Stetz, G. Gindi, T. Garrand, and L. Deckelbaum, "Biochemical basis for the difference between normal and atherosclerotic arterial fluorescence," *Circulation* **80**, 1893–1901 (1989).
18. J. J. Baraga, R. P. Rava, P. Taroni, C. Kittrell, M. Fizmaurice, and M. S. Feld, "Laser induced fluorescence spectroscopy of normal and atherosclerotic aorta using 306–310 nm excitation," *Lasers Surg. Med.* **10**, 245–261 (1990).
19. D. Murhy-Chutorian, J. Kosek, W. Mok, S. Quay, W. Huestis, J. Mehigan, D. Profitt, and R. Ginburg, "Selective absorption of ultraviolet laser energy by human atherotic plaque treated with tetracycline," *Am. J. Cardiol.* **55**, 1293–1297 (1985).
20. R. Clark, J. Isner, T. Gauthier, K. Nakagawa, F. Cerio, E. Hanion, E. Gaffney, E. Rouse, and S. Dejesus, "Spectroscopic characterization of cardiovascular tissue," *Lasers Surg. Med.* **8**, 45–59 (1988).
21. S. Anderson-Engels, J. Johannson, U. Stenram, S. Svanberg, and K. Svanberg, "Time-resolved laser-induced fluorescence spectroscopy for enhanced demarcation of human atherosclerotic plaques," *J. Photochem. Photobiol., B* **4**, 363–369 (1990).
22. S. Anderson-Engels, J. Johannson, and S. Svanberg, "The use of time-resolved fluorescence for diagnosis of atherosclerotic plaque and malignant tumors," *Spectrochim. Acta, Part A* **46A**, 1203–1210 (1990).
23. L. Marcu, M. Fishbein, J. Maarek, M. Jean-Michel, and W. Grundfest, "Discrimination of human coronary artery atherosclerotic lipid-rich lesions by time-resolved laser-induced fluorescence spectroscopy," *Arterioscler., Thromb., Vasc. Biol.* **21**, 1244–1250 (2001).
24. Q. Fang, T. Papaioannou, J. A. Jo, R. Vaitha, and K. Shastry, "Time-domain laser-induced fluorescence spectroscopy apparatus for clinical diagnostics," *Rev. Sci. Instrum.* **75**, 151–162 (2004).
25. N. Anastassopoulou, B. Arapoglou, P. Demakakos, M. Makropoulou, A. Paphiti, and A. Serafetinides, "Spectroscopic characterisation of carotid atherosclerotic plaque by laser induced fluorescence," *Lasers Surg. Med.* **28**, 67–73 (2001).
26. T. G. Papazolou, W. Q. Liu, A. Katsamouris, and C. Fotakis, "Laser-induced fluorescence detection of cardiovascular atherosclerotic deposits via their natural emission and hypocrellin (HA) probing," *J. Photochem. Photobiol., B* **22**, 139–144 (1994).
27. B. Zhu, F. A. Jaffer, V. Ntziachristos, and R. Weissleder, "Development of a near infrared fluorescence catheter: operating characteristics and feasibility for atherosclerotic plaque detection," *J. Phys. D* **38**, 2701–2707 (2005).
28. L. Marcu, Q. Fang, J. A. Jo, T. Papaioannou, A. Dorafshar, T. Reil, J. H. Qiao, J. D. Baker, J. A. Freischlag, and M. C. Fishbein, "In vivo detection of macrophages in a rabbit atherosclerotic model by time-resolved laser-induced fluorescence spectroscopy," *Atherosclerosis* **181**, 295–303 (2005).
29. A. Christov, E. Dai, M. Drangova, L. Liu, G. S. Abela, P. Nash, G. McFadden, and A. R. Lucas, "Optical detection of triggered atherosclerotic plaque disruption by fluorescence emission analysis," *Photochem. Photobiol.* **72**, 242–252 (2000).
30. A. Christov, R. Korol, E. Dai, L. Liu, H. Guan, M. A. Bernards, P. B. Cavers, D. Susko, and A. R. Lucas, "In vivo optical analysis of quantitative change in collagen and elastin during arterial remodeling," *Photochem. Photobiol.* **81**, 457–466 (2005).
31. R. P. Rava, J. J. Baraga, and M. S. Feld, "Near infrared Fourier transform Raman spectroscopy of human artery," *Spectrochim. Acta, Part A* **47A**, 509–512 (1991).
32. J. J. Baraga, M. S. Feld, and R. P. Rava, "In situ optical histochemistry for human artery using near infrared Fourier transform Raman spectroscopy," *Proc. Natl. Acad. Sci. U.S.A.* **89**, 3473 (1992).
33. J. J. Baraga, M. S. Feld, and R. P. Rava, "Rapid near-infrared Raman spectroscopy of human tissue with a spectrograph and CCD detector," *Appl. Spectrosc.* **46**, 187–190 (1992).
34. J. F. Brennan, T. J. Römer, R. S. Lees, A. M. Tercyak, J. R. Kramer, and M. S. Feld, "Determination of human coronary artery composition by Raman spectroscopy," *Circulation* **96**, 99–105 (1997).
35. R. Manoharan, J. J. Baraga, M. S. Feld, and R. P. Rava, "Quantitative histochemical analysis of human artery using Raman spectroscopy," *J. Photochem. Photobiol., B* **16**, 211–233 (1992).
36. G. Deinum, D. Rodriguez, T. J. Römer, M. Fizmaurice, J. R. Kramer, and M. S. Feld, "Histological classification of Raman spectra of human coronary artery atherosclerosis using principal component analysis," *Appl. Spectrosc.* **53**, 938–942 (1998).
37. H. P. Buchman, J. T. Motz, G. Deinum, T. J. Römer, M. Fizmaurice, J. R. Kramer, A. van der Laarse, A. V. Bruschke, and M. S. Feld, "Diagnosis of human coronary atherosclerosis by morphology-based Raman spectroscopy," *Cardiovasc. Pathol.* **10**, 59–68 (2001).
38. P. Weinmann, M. Joouan, Q. D. Nguyen, B. Lacroix, C. Groiselle, J. P. Bonie, and L. Gérald, "Quantitative analysis of cholesterol and cholesterol esters in human atherosclerotic plaques using near-infrared Raman spectroscopy," *Atherosclerosis* **140**, 81–88 (1988).
39. P. L. Cacheux, G. Ménard, H. N. Quaang, N. Q. Dao, and A. G. Roach, "Quantitative determination for free and esterified cholesterol concentrations in cholesterol-fed rabbit aorta using near-infrared Fourier transform-Raman spectroscopy," *Spectrochim. Acta, Part A* **52A**, 1619–1627 (1996).
40. J. F. Brennan, III, Y. Wang, R. R. Dasari, and M. S. Feld, "Near-infrared Raman spectroscopy systems for human tissue studies," *Appl. Spectrosc.* **51**, 201–207 (1996).
41. H. P. Buchman, G. Deinum, J. T. Motz, M. Fitzmaurice, J. R. Kramer, A. van der Laarse, A. V. Bruschke, and M. S. Feld, "Raman microspectroscopy of human coronary atherosclerosis: Biochemical assessment of cellular and extracellular morphologic structures in situ," *Cardiovasc. Pathol.* **10**, 69–82 (2001).
42. S. W. E. van de Poll, K. Kastelijn, T. C. Bakker Schut, C. Strijder, G. Pasterkamp, G. J. Puppel, and A. van der Laarse, "On-line detection of cholesterol and calcification by catheter based Raman spectroscopy in human atherosclerotic plaque ex vivo," *Heart* **89**, 1078–1082 (2003).
43. S. W. E. van de Poll, T. J. Römer, O. L. Volger, D. J. M. Delsing, T. C. Bakker Schut, H. M. G. Princen, L. M. Havekes, J. Wouter Jukema, A. van der Laarse, and G. J. Puppels, "Raman spectroscopic evaluation of the effects of diet and lipid-lowering therapy on atherosclerotic plaque development in mice," *Arterioscler., Thromb., Vasc. Biol.* **21**, 1630–1635 (2001).
44. H. P. Buschman, E. T. Marple, M. L. Wach, B. Bennett, T. C. Bakker Schut, H. A. Bruining, A. V. Bruschke, A. Van der Laarse, and G. J. Puppels, "In vivo determination of the molecular composition of artery wall by intravascular Raman spectroscopy," *Anal. Chem.* **72**, 3771–3775 (2000).
45. J. T. Motz, M. Hunter, L. H. Galindo, J. A. Gardecki, J. R. Kramer, R. R. Dasari, and M. S. Feld, "Optical fiber probe for biomedical Raman spectroscopy," *Appl. Opt.* **43**, 542–554 (2004).
46. J. T. Motz, S. J. Gandhi, O. R. Scepanovic, A. S. Haka, J. R. Kramer, R. R. Dasari, and M. S. Feld, "Real-time Raman system for in vivo disease diagnosis," *J. Biomed. Opt.* **10**, 031113 (2005).
47. J. T. Motz, M. Fizmaurice, A. Miller, S. J. Gandhi, A. S. Haka, L. H. Garlindo, R. R. Dasari, J. R. Kramer, and M. S. Feld, "In vivo Raman spectral pathology of human atherosclerosis and vulnerable plaque," *J. Biomed. Opt.* **11**, 021003 (2006).
48. R. A. Lodder, L. Cassis, and E. W. Ciurczak, "Artery analysis with a novel near-IR fiber-optic probe," *Spectroscopy (Amsterdam)* **5**, 12–17 (1990).
49. W. Jaross, V. Neumeister, P. Lattke, and D. Schuh, "Determination of cholesterol in atherosclerotic plaques using near infrared diffuse reflection spectroscopy," *Atherosclerosis* **147**, 327–337 (1999).
50. J. Wang, Y. Geng, B. Guo, T. Klim, B. N. Lal, J. T. Willerson, and W. Casscells, "Near-infrared spectroscopic characterization of human advanced atherosclerotic plaques," *J. Am. Coll. Cardiol.* **39**, 1305–1313 (2002).
51. V. Neumeister, M. Scherbe, P. Lattke, and W. Jaross, "Determination of the cholesterol-collagen ratio of arterial atherosclerotic plaques using near infrared spectroscopy as a possible measure of plaque stability," *Atherosclerosis* **165**, 251–257 (2002).
52. T. Arai, K. Mlzuno, A. Fujikawa, M. Nakagawa, and M. Kikuchl, "Infrared absorption spectra ranging from 2.5 to 10 μm at various layers of human normal abdominal aorta and Fibrofatty atheroma in vitro," *Lasers Surg. Med.* **10**, 357–362 (1990).
53. J. J. Baraga, M. S. Feld, and R. P. Rava, "Detection of atherosclerosis in human artery by mid-infrared attenuated total reflectance," *Appl. Spectrosc.* **45**, 709–711 (1991).
54. R. Manoharan, J. J. Baraga, R. P. Rava, R. R. Dasari, M. Fitzmaurice, and M. S. Feld, "Biochemical analysis and mapping of atherosclerotic human artery using FT-IR microspectroscopy," *Atherosclerosis*

103, 181–193 (1993).
55. M. Sartori, D. Weilbaecher, G. L. Valderrama, S. Kubodera, R. C. Chin, M. J. Berry, and F. K. Tittel, "Laser-induced autofluorescence of human arteries," *Circ. Res.* **63**, 1053–1059 (1988).
56. S. W. E. van de Poll, T. C. B. Schut, A. Van den Laarse, and G. J. Puppels, "In situ investigation of the chemical composition of ceroid in human atherosclerosis by Raman spectroscopy," *J. Raman Spectrosc.* **33**, 544–551 (2002).
57. K. E. Shafer-Peltier, A. S. Haka, J. T. Motz, M. Fitzmaurice, R. R. Dasari, and M. S. Feld, "Model-based biological Raman spectral imaging," *J. Cell Biochem. Suppl.* **39**, 125–137 (2002).
58. L. A. Cassis and R. A. Lodder, "Near-IR imaging of atheromas in living artery tissue," *Anal. Chem.* **65**, 1247–1256 (1993).
59. F. Alo, P. Bruni, A. Cavalleri, C. Conti, E. Dasari, C. Rubini, and G. Tosi, "Infrared microscopy characterization of carotid plaques and thyroid tissue biopsies," *J. Mol. Struct.* **651**, 419–426 (2003).
60. C. S. Colley, S. G. Kazarian, P. D. Weinberg, and M. J. Lever, "Spectroscopic imaging of arteries and atherosclerotic plaques," *Biopolymers* **74**, 328–335 (2004).
61. B. A. Hooper, A. Maheshwari, A. C. Curry, and T. M. Alter, "Catheter for diagnosis and therapy with infrared evanescent waves," *Appl. Opt.* **42**, 3205–3214 (2003).
62. B. A. Hooper, G. C. LaVerde, and O. T. von Ramm, "Design and construction of an evanescent optical wave device for the recanalization of vessels," *Nucl. Instrum. Methods Phys. Res. A* **475**, 645–649 (2001).
63. L. Wang, J. Chapman, R. A. Palmer, T. M. Alter, B. A. Hooper, O. V. Ramm, and B. Mizaikoff, "Classification of atherosclerotic rabbit samples with an infrared attenuated total reflectance catheter and multivariate data analysis," *Appl. Spectrosc.* **60**, 1121–1126 (2006).
64. I. Dabanoglu, "A quantitative study of the aorta of the New Zealand Rabbit (Oryctolagus cuniculus L.)," *Anat. Histol. Embryol.* **29**, 145–147 (2000).
65. R. Kramer, *Chemometric Techniques for Quantitative Analysis*, pp. **173**, Marcel Deker, New York (1998).
66. M. Barker and W. Rayens, "Partial least squares for discrimination," *J. Chemom.* **17**, 166–173 (1993).
67. H. L. Mark and D. Tunnell, "Qualitative near-infrared reflectance analysis using mahalanobis distance," *Anal. Chem.* **57**, 1449–1456 (1985).
68. H. Mark, "Normalized distance for qualitative near-infrared reflectance analysis," *Anal. Chem.* **58**, 379–384 (1986).
69. O. J. Balchum, D. R. Doiron, A. E. Profil, and G. C. Huth, "Fluorescence bronchoscopy for localizing early bronchial cancer and carcinoma in situ," *Recent Results Cancer Res.* **82**, 97–120 (1982).
70. D. M. Haaland, H. D. T. Jones, and E. V. Thomas, "Multivariate classification of the infrared spectra of cell and tissue samples," *Appl. Spectrosc.* **51**, 340–345 (1997).