Raman Microscopy for the Chemometric Analysis of Tumor Cells

Ahcene Taleb,*,† James Diamond,† John J. McGarvey,‡ J. Renwick Beattie,‡ Carla Toland,† and Peter W. Hamilton†

Biomedical Imaging and Informatics Research Group, Department of Pathology, Queen's University Belfast, Grosvenor Road, Belfast BT12 6BL, Northern Ireland, United Kingdom, and Clinical Raman Microscopy Group, School of Chemistry and Chemical Engineering, Queen's University Belfast, Belfast BT9 5AG, Northern Ireland, United Kingdom

Received: March 30, 2006; In Final Form: June 22, 2006

Raman spectroscopy is recognized as a tool for chemometric analysis of biological materials due to the high information content relating to specific physical and chemical qualities of the sample. Thirty cells belonging to two different prostatic cell lines, PNT1A (immortalized normal prostate cell line) and LNCaP (malignant cell line derived from prostate metastases), were mapped using Raman microscopy. A range of spectral preprocessing methods (partial least-squares discriminant analyses (PLSDAs), principal component analyses (PCAs), and adjacent band ratios (ABRs)) were compared for input into linear discriminant analysis to model and classify the two cell lines. PLSDA and ABR were able to correctly classify 100% of cells into benign and malignant groups, while PLSDA correctly classified a greater proportion of individual spectra. PCA was used to image the distribution of various biochemicals inside each cell and confirm differences in composition/distribution between benign and malignant cell lines. This study has demonstrated that PLSDAs and ABRs of Raman data can identify subtle differences between benign and malignant prostatic cells in vitro.

1. Introduction

Raman spectroscopy is a valuable tool for probing biochemical changes in histological and cytological samples.[1−3] This is largely due to the highly specific information content that Raman spectroscopy provides, with the ability to provide unique fingerprints of the biochemical constituents of a sample and predict the quantity of a wide range of these biochemical constituents. Since most diseases display complex perturbations in biochemical processes at the cellular level, often driven by underlying genetic changes,[4] Raman spectroscopy appears to be a promising tool for probing these often subtle alterations and thereby distinguishing abnormal tissues. This has been illustrated in numerous studies where Raman spectroscopy can be used to detect tissue abnormalities.[4−7] The underlying cellular basis for these observations is not fully known, and further studies need to be carried out to explore Raman signatures associated with well-defined and characterized cell populations. This study assesses the role that Raman spectroscopy might play in the investigation of well-defined benign and malignant prostate cell lines.

Raman spectroscopy is rapid, nondestructive, and noninvasive under appropriate choices of laser wavelength and power,[8] with the ability to detect target compounds within a complex matrix. It enables samples to be investigated with little or no prior preparation. Moreover, water as a solvent is a weak Raman scatterer. Consequently, Raman spectroscopy can be applied in vivo[9,10] as well as in vitro.[5,7,11] Combining Raman spectroscopy with optical confocal microscopy (confocal Raman microscopy) adds spatial to spectral resolution and opens up the possibility of mapping, for instance, biological samples for chemical and physical composition at a cellular level. This spatial resolution is important as biological samples are often inherently heterogeneous with biochemical distributions that may differ from sample to sample and also from point to point within the same sample.

The primary advantage that Raman microscopy holds over optical microscopy rests in the amount of information that can be simultaneously recorded and analyzed in a purely objective manner. Developments in Raman technology have made modern instruments much more ergonomic and accessible to the biomedical community for cellular research.

Previous studies using Raman techniques have shown that specific biochemical changes can be detected in cancerous tissue.[4−6,9−12] These have been carried out in either an in vivo setting where access with a Raman probe can be achieved or in an ex vivo setting where access would be more restricted in vivo. However, analysis of gross tissue samples in this way is problematic in that it corresponds to a complex combination of cell types with different biochemical profiles. In malignancy, abnormalities usually arise in a single cell type, often presenting as focal change in a background of visually normal cells. Analyzing average spectra from this complex biochemical arena without selectivity could result in a dilution of information and deterioration in the sensitivity for identifying abnormalities. Sensitivity may be further compromised by the variation in biochemical constituents within a particular cell type, the impact that localized impurities in the sample may have on signal distortion, and the noise resulting from the sample container in which tissues are presented to the Raman detector and possibly by the methods of tissue/cell preparation. The challenge is to

* Author to whom correspondence should be addressed. Fax: +442890632763. E-mail: a.taleb@qub.ac.uk.

† Biomedical Imaging and Informatics Research Group, Department of Pathology, Queen's University Belfast.

‡ Clinical Raman Microscopy Group, School of Chemistry and Chemical Engineering, Queen's University Belfast.

10.1021/jp061981q CCC: $33.50
Published on Web 09/12/2006

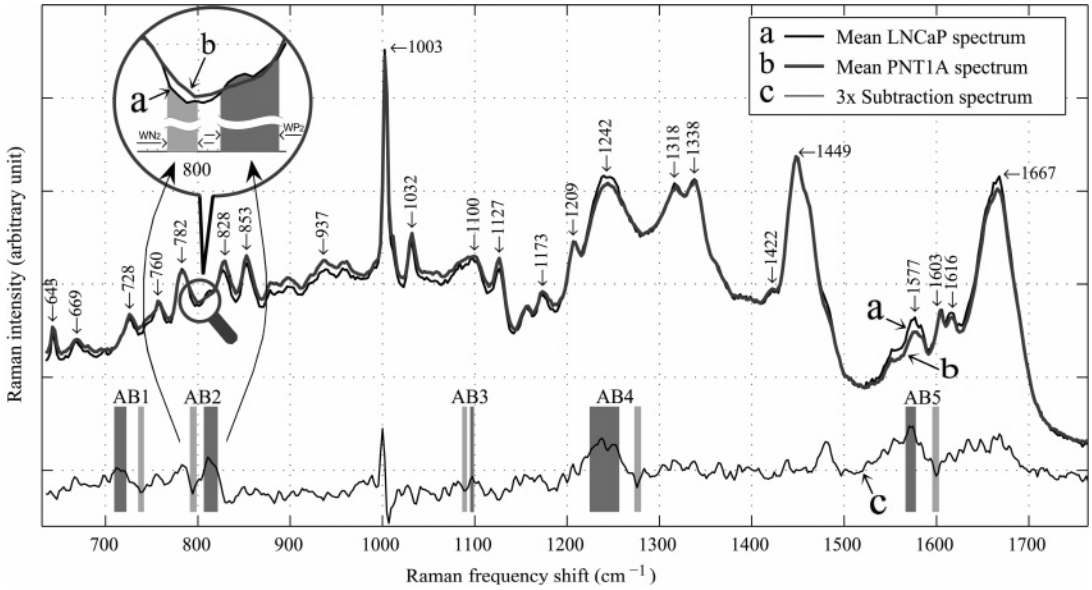

Figure 1. Normalized average spectra and their subtraction spectrum where a represents the average spectrum of LNCaP cells, b the average spectrum of PNT1A cells and c the amplified (3×) subtraction spectrum (a − b). AB*i* shows the locations of adjacent positive and negative bands. WP2 and WN2 in the enlarged region represent the widths of the positive (WP2) and negative (WN2) adjacent bands 2 (AB2). The areas delimited by a and AB2 are highlighted. ANOVA and *t*-test of means show that the differences in the highlighted bands are statistically significant between the two cell types ($p < 0.001$).

accurately characterize Raman spectra associated with individual cells and understand their contribution to the spectra obtained from tissue samples.

Probing biochemicals and their conformational structures inside single cells, using the family of techniques based on Raman scattering, is a developing research domain.[3,13−20] Many Raman cancer detection studies on tissue samples have shown a degree of success.[4−7,9,11,12] In this study we wished to investigate whether Raman scattering is capable of providing adequate discrimination in vitro on a system where there is a limited number of cell types and where no ambiguity exists, thereby allowing us to investigate the consequences of a range of commonly used spectral preprocessing steps. Probing biochemical changes in cultured cancerous cells may provide a better understanding of the relationship between a cancer cell and diagnosis of the disease using Raman spectroscopy. In this study a visible excitation laser wavelength of 633 nm was used to probe for biochemical differences between two prostatic cell lines LNCaP (malignant) and PNT1A (normal) and to assess and compare the linear discrimination and prediction rate using multivariate analysis techniques.

2. Materials and Methods

2.1. Cell Preparation. Two cell lines, PNT1A (immortalized with SV40 virus, normal prostate) and LNCaP (derived from prostate metastases), were examined. These cells were supplied in 1 mL Nunc cryotubes frozen in liquid nitrogen. The cells were seeded and incubated in the culture medium RPMI 1640 (Sigma Aldrich, Dorset, England) supplemented with fetal calf serum (FCS) (Labtech International, Sussex, England) 4 mM L-glutamine (Sigma-Aldrich), 0.06 mg/mL benzylpenicillin, and 0.1 mg/mL streptomycin (Sigma-Aldrich) in T75 flasks at 37 °C in 5% CO_2.

To minimize possible artifacts due to variation in cell cycle we took a number of steps to ensure no bias was created due to investigations being conducted at different stages in cell growth. The Raman data were recorded from six separate cultures, all of which were grown to a similar level of confluency (approximately 70%).

Drops of trypsin and EDTA were added to detach cells from the flask and enhance the activities of the enzymes, for about 1 min, then the trypsin reaction was neutralized by adding 5 mL of culture medium before the flask was returned to the incubator for 3 min at 37 °C. The cells were subsequently fixed in 95% ethanol, washed twice in double-distilled water, and cytospun onto quartz slides.

2.2. Confocal Raman Microscopy. Raman measurements were performed on 30 cells and recorded using a LabRam HR800 confocal Raman microscope and LabSpec software (Jobin-Yvon, Villeneuve d'Ascq, France). The spectra were generated using the 633 nm line of a HeNe laser (20 mW at sample) focused on the sample with a 100× objective (N.A. = 0.9) in a confocal arrangement (confocal hole diameter = 100 μm). The microscope was equipped with a (*xyz*) motorized stage for automatic positioning of the sample in *xyz*-space. The excitation laser was focused to a ca. 1.25 μm diameter spot with the scattered light collected in a 180° geometry.

A 633 nm notch filter was used to reject elastic scattering, and a 600 groove mm^{-1} diffraction grating was used to provide spectral resolution of 6 cm^{-1} over the spectral range from 635 to 1765 cm^{-1} on a Peltier-cooled Andor CCD detector (model no. DU420). Under the foregoing conditions, no significant sample degradation was observed. All optical images and spectral maps were recorded using the LabSpec software, with automatic cosmic ray removal.

The unstained cytospun cells were identified on quartz slides under the Raman microscope equipped with a 40× phase contrast objective. Fifteen apparently intact cells were selected for each cell line, giving a total of 30 cells. Subsequently, 42 measurements on average per cell were acquired using point-by-point scanning, with a spatial interval of 1.6 μm along the *x*- and *y*-axes over the full area of the cell. The quartz substrate contribution to the observed Raman signal was corrected by subtracting a standard quartz spectrum, acquired under identical conditions using LabSpec software.

2.3. Multivariate Discriminant Analysis. Adjacent positive and negative bands in the subtraction spectrum of PNT1A (normal) from LNCaP (malignant) (Figure 1) are significantly

different between the cell lines and therefore expected to be appropriate for discrimination purposes. The areas of these bands were calculated both for the individual spectra and for the average spectra of each cell, using a trapezoidal integration algorithm, and the ratios then were taken to give adjacent band ratios (ABRs). ABRs were developed by this group to extract features from Raman spectra for the purposes of cell classification. The method is described in more detail in the Results section, and the algorithm was programmed specifically for this study in Matlab (Mathworks, Inc., Natick, MA). Alternatively multivariate methods were used to reduce the data, using principal component analyses (PCAs) and partial least-squares discriminant analyses (PLSDAs) to generate "scores" that summarize the main variation within the spectral data.

For the individual spectra a randomly selected training set of 232 spectra from the total of 1304 acquired (after thresholding) was input to the linear discriminant analysis (LDA) classifier to classify and predict the group membership of the remaining 1072 spectra. Each individual spectrum was classified in one of the two groups. A cell is classified in the group in which the majority of its spectra are classified. To investigate the use of the mean spectrum for each cell for classification the spectra of each cell were averaged after the 3% threshold was applied and before any further processing steps were carried out. Three LNCaP cells and three PNT1A cells were used as a training set to classify and predict the group membership of the remaining 24 cells.

A scores map of the extracted principal components (PCs) was constructed, representing the scores by shades of gray, varying from black for the lowest score spectra to white for the highest score. Matlab was used for baselining, normalizing, smoothing, and PCA-preprocessing of the Raman data. Simca P8.0 (Umetrics, Umeå, Sweden) was used for PLSDA preprocessing of the Raman data. Spectra were mean-centered prior to the PCA and PLSDA procedures. SPSS V13.0 (SPSS, Inc. Chicago, IL) was employed for LDA of the reduced Raman data.

2.4. Data Preprocessing. Commonly used spectral data preprocessing methods were applied to the data prior to data reduction to assess their influence on the classification and prediction ability.

2.4.1. Filtering. A Savitzky−Golay smoothing filter [21] with a window width of 5 points was used to preserve most of the spectral features while removing noise.

2.4.2. Baselining. Three-point linear baselining was used for background removal in the Raman data.

2.4.3. Thresholding. Spectra with intensity <3% of the average CH_2 scissor intensity (1449 cm^{-1}) were rejected from the dataset.

2.4.4. Normalization. The absolute signal intensity is not very reproducible in Raman spectroscopy as it is affected by a wide range of experimental factors. Generation of ABRs inherently involves normalization of one band to another. For multivariate preprocessing, the area of the commonly employed 1449 cm^{-1} band,[8,16,22,23] corresponding to C−H deformation vibrations in all the biochemical constituents of the cells,[16] was used (Table 3).

3. Results

The average Raman spectra for both cell lines are compared in Figure 1, along with the subtraction spectrum for PNT1A from LNCaP. The non-Raman background observed in the raw data is relatively low, thus enabling acquisition of spectra with good quality signal-to-noise ratio at short accumulation times.

TABLE 1: Raman Peak Assignment[a]

frequency shift (cm^{-1})	DNA/RNA	proteins	lipids
643		C−C twist Tyr[6]	
669	T, G[3]	Cys[6]	
726	A[22]		
760		Trp[16]	
782	C, T, U[16]		
828	ν(O−P−O) bk[6,16]	Tyr[6]	
853		Tyr[6,22], Pro[6]	
937		ν(C−C) α-helix[16]	
1003		Phe[22]	
1032		Phe[19]	
1100	PO_2^- [24]		
1127		ν(C−N) bk[22]	ν(C−C)[7]
1177	T[22]	Tyr[22]	
1209		Phe, Tyr[6]	
1242		amide III β-sheet[16]	
1318	G[24]		
1338	A[24]		
1422	A, G[22]		
1449	δCH_2 in-plane r. ν[17]	δCH_2[3,17]	δCH_2[16,17]
1484	A, G[24]		
1577	A, G[3,24]		
1603		Phe, Tyr[6]	
1616		Tyr, Trp[6]	
1667	T[24]	amide I β-sheet[10]	fatty acids[10]

[a] C is cytosine, T is thymine, G is guanine, A is adenine, U is uracil, Phe is phenylalanine, Tyr is tyrosine, Trp is tryptophan, Cys is cysteine, Pro is proline, r is ring, bk is backbone, ν is stretch, and δ is deformation.

The spectra predominantly show the typical pattern for proteins, with contributions from DNA/RNA and many bands related to individual amino acids such as Phe, Trp, and Tyr. A detailed list of assignments for the bands is given in Table 1.

The positive bands in the subtraction spectrum relate to the LNCaP cell line and may arise from amino acids (proteins) and A-DNA. For example, the bands at 711, 813, 1100, 1243, and 1572 cm^{-1} are all typical of A-DNA. In contrast, the PNT1A cell line shows more intensity (corresponding to the minima in the subtraction spectrum) in bands that can be related to B-DNA, as shown by the negative bands at 733, 798, and 1091 cm^{-1}. The differences between the Raman spectra of the two cell lines are small compared with the overall signal levels, but the differences were indeed found to be statistically significant (analysis of variance (ANOVA) and *t*-test of means gave $p < 0.001$). While enhancement methods exist such as surface-enhanced Raman scattering (SERS) and resonance Raman scattering, these involve addition of invasive materials and use of specific excitation wavelengths, respectively. Since we wished to minimize sample processing we chose not to investigate the use of SERS, while resonance Raman spectroscopy of DNA and proteins requires use of UV wavelengths, which would increase the risk of sample photodegradation.

These results suggest that the most significant spectral change between the cell lines is an increase in DNA content for the malignant cells, with a concomitant change in DNA conformation from the common B-DNA form to the A-DNA form.

3.1. Adjacent Band Ratios. Figure 1 also shows the locations and widths of the five pairs of adjacent bands (ABs) that were determined to be most significant in discriminating between the cell lines. The ratios of the positive to negative bands are computed for each individual spectrum in the full dataset as well as for the average spectrum of each cell. The resulting ABRs are input to the LDA for classification.

Table 2 and Figure 2a show the classification results for the cell lines, using the ABRs calculated after various preprocessing

TABLE 2: LDA Classification/Prediction Based on the ABR Relative to Different Data Preprocessing[a]

training spectra classification	test spectra prediction	cross-validation	cells correctly predicted	preprocessing
77.2%	74.3%	76.3%	30/30	
77.6%	74.5%	76.7%	30/30	SG
75.4%	71.5%	74.1%	28/30	LBasl 3
75.4%	70.1%	75.0%	28/30	SG, LBasl 3
72.0%	69.0%	71.1%	28/30	LBasl 6
76.3%	70.4%	75.4%	27/30	SG, LBasl 6

[a] SG is the Savitsky−Golay five-point smoothing filter, and LBasl 3 and LBasl 6 are three- and six-point linear baselining, respectively.

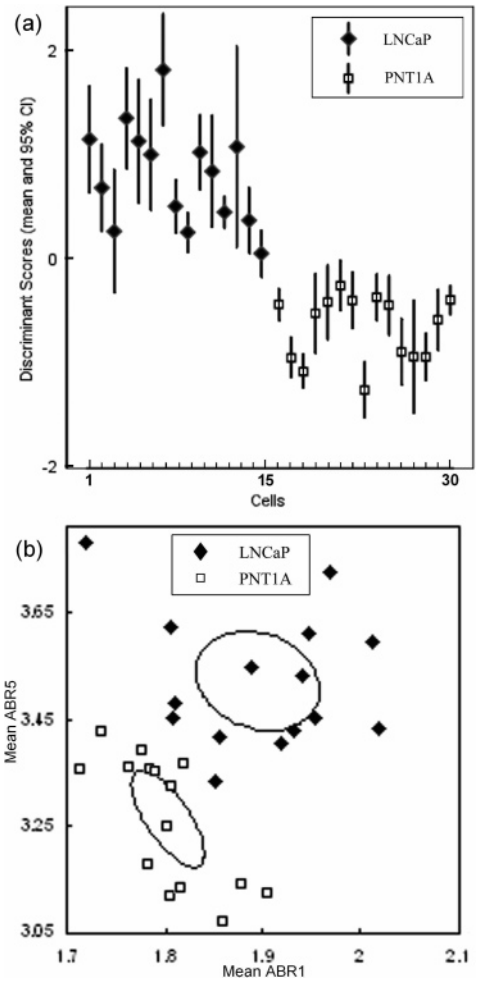

Figure 2. (a) Mean and 95% confidence interval of the test set discriminant scores derived from the ABRs plotted for each cell. (b) Scatter plot of the mean values of the most significant ABRs (ABR1 and ABR5, see Figure 1). The 95% confidence ellipse for each cell type is plotted, showing a clear separation between cell types.

methods were employed. As the table shows, it was found that in our particular study there is no need for any preprocessing procedure to achieve a high classification rate using the ABR technique. Indeed overprocessing of the spectra reduced the classification rate. Through the use of the ABR derived from individual spectra as input to the LDA classifier, 100% correct classification/prediction of the group membership of cells was achieved in both unprocessed data and smoothed data. However, further preprocessing reduces the classification/prediction performance (Table 2). The 95% confidence ellipses for the individual ABRs (in the scatter plot of the two selected band ratios shown in Figure 2b) are well-separated, confirming the ability of the ABRs selected above (Figure 1) to discriminate between the two cell lines.

In addition, the ABRs were computed for the averaged spectrum of each cell. To avoid overfitting (overtraining) the LDA model, the number of training spectra is chosen to be at least 2 times greater than the number of features.[9] Three ABRs were used on a training set of 6 spectra and a prediction set of 24 spectra. In this scenario the predictive capability degraded significantly (4 cells on average are misclassified) except in the case where the data were linearly baselined using 6 points, with or without smoothing.

3.2. Multivariate Analysis of spectra. The same training set of 232 spectra employed above was reduced using PCA. The first 12 PC scores are input into the LDA to classify and predict the group membership of the test set. The results in Table 3 show that in most cases only one cell is misclassified. The best prediction rating (90% of individual scores correctly predicted) was obtained for the case of baselined and normalized spectra. The ABR method showed the opposite behavior, with these processing steps reducing performance.

Employing PLSDA, in which the reduction of the spectral data is directed by PLS regression, enabled 100% correct classification of the cells and 94% correct classification of the individual spectra. This high prediction rate for the individual spectra is an indication of the power of PLSDA for elucidating complex patterns in spectral data. Spectral differences are not only observed as arising from the two cell types but also as a result of the heterogeneous composition within each cell line. PLSDA is able to deal with the two sets of differences between cell lines for both the cytoplasm and the nucleus without separating the spectra for the two regions.

Multivariate scores relate to information concerning the biochemical composition of the volume probed by the laser. Visual mapping of Raman scores will reveal the distribution of these biochemicals within the cell. As an example, two cells, from the total number investigated, one malignant (LNCaP type) and the other normal (PNT1A) in the test set, were analyzed separately to spatially visualize the combination of chemical constituents represented by the first five PCs. The plots of the first five extracted PCs (Figure 3) illustrate the major Raman bands used in the data reduction. For instance, PC4 is a typical reproduction of a DNA spectrum. The corresponding score maps are plotted in Figure 4, showing the spatial distribution of linear combinations of biochemicals represented by the extracted PCs.

4. Discussion

The differences observed between the spectra of the cell lines (Figure 1) are in accordance with histopathologic assessment where the nuclei are larger in malignant cells than for normal cells, consistent with the increased spectral intensity of DNA-related bands. It is noteworthy that many of the bands characteristic of A-DNA/RNA conformation, in particular at 706 $\pm$ 5, 807 $\pm$ 3, and 1099 $\pm$ 1 cm^{-1} are found in the LNCaP (malignant) cells, while the more common B-DNA/RNA

TABLE 3: Prediction Performance Based on PC Scores[a]

training scores classification	test scores prediction	cross-validation	cells correctly predicted	preprocessing
92.2%	87.1%	91.4%	29/30	
94.4%	88.4%	93.5%	29/30	SG, PCA
89.7%	84.5%	88.4%	29/30	B, PCA
90.5%	84.3%	89.7%	29/30	N, PCA
95.7%	88.1%	93.5%	29/30	SG, N, PCA
92.2%	87.2%	91.4%	29/30	SG, N, B, PCA
90.1%	89.7%	90.1%	29/30	B, N, PCA
90.5%	86.1%	90.5%	29/30	SG, B, N, PCA

[a] SG is the Savitsky−Golay five-point smoothing filter, B is three-point linear baselining, and N is normalization about the scissor band at 1449 cm^{-1}.

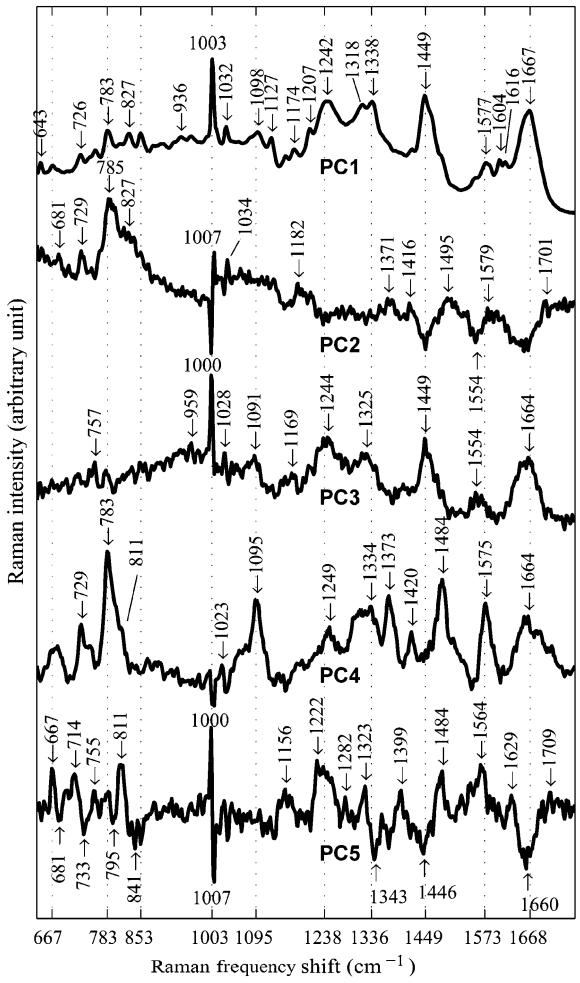

Figure 3. First five principal components. PC1 can be identified with the average spectrum, PC2 represents variation in DNA and protein, PC3 is mainly due to the cytosol with a small contribution from nucleic acids, PC4 corresponds to typical spectra of nucleic acids, and PC5 is a combination of nucleic acid, lipid, and protein spectra.

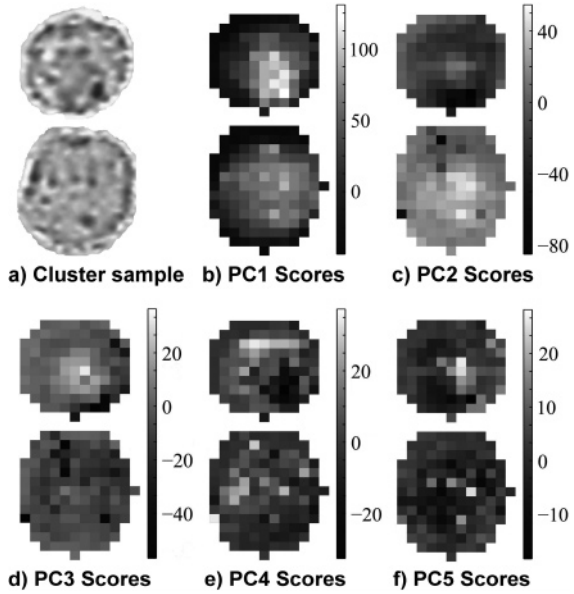

Figure 4. (a) Score maps generated by the first five PCs on a cluster sample image of two cells, LNCaP (top) and PNT1A (bottom). (b) PC1 scores could be identified with the overall average cellular spectrum intensity. (c) PC2 scores represent possibly a variation in DNA and protein. (d) PC3 scores can be linked mainly to the cytosol. (e) PC4 scores show the concentration distribution of nucleic acids. (f) PC5 scores represent a variety of biochemicals linked to the differences between the two cell types.

conformation bands, in particular at 790 ± 5, 835, and 1092 ± 1 cm^{-1}, are mostly found in the PNT1A (normal) cells. More complete assignments of DNA conformation bands can be found elsewhere. [24,25]

4.1 Classification. The LDA classifier is easily overfitted, giving overly optimistic prediction rates that prove unreliable when tried on independent samples. In practice, it is essential to provide the LDA model with at least twice as many samples as there are variables, and as the full Raman spectrum contains hundreds (in our case 512) of data points this number needs to be reduced to carry out training with a realistic number of samples. The model that is generated must be tested on an independent test set to determine the reliability of the model accurately.

The results indicate that both PLSDA and ABR data reduction methods are capable of accurately predicting the cell line to which a cell belongs. In terms of predicting the class membership of individual spectra within the cell, PLSDA was found to have enhanced sensitivity compared with the computationally simpler ABR method. This is attributed to the fact that the PLSDA data reduction retains a much greater proportion of the original information than the ABR method.

The contrasting influence of data preprocessing on the ABR and PLSDA techniques is not surprising as the method of calculating the ABR inherently compensates for many of the variations that these preprocessing methods are designed to reduce. In this study the non-Raman spectral background was relatively weak and only showed some subtle variations (our dataset consisted of only two pure cell types), which meant that because the bands are adjacent they have effectively the same baseline, which will be canceled out in a ratio. As long as the proportion of the accumulated data attributable to background interference does not vary relative to the true Raman signal, ABRs are relatively insensitive to changes in background shape.

However, in many sample types there can be considerable variation in the ratio of the background to the true Raman signal, and this can quickly compromise the use of unbaselined band ratios, even if they are adjacent to one another. Similarly, the calculation of ratios is self-normalizing, and the calculation of band areas is inherently smoothing. In the case of multivariate data reduction algorithms, preprocessing methods are required to reduce irrelevant variation that can potentially confound any model constructed from the training set. For example, when the background is non-Raman and so is independent of the Raman signal, the absolute intensity of the Raman signal itself is hard to reproduce accurately as it is dependent on a wide range of factors. These variations can easily mask the variations within the Raman part of the acquired data, and so it is generally more helpful to remove this variation, since it is not related to the Raman spectrum.

The decrease in prediction performance when using average spectra rather than individual spectra can be explained by the fact that minor impurities with high Raman/fluorescent intensity could distort significantly the average spectra, while by screening each individual spectrum the contamination is restricted to a very low proportion of the spectra and only those spectra arising from impurities would be misclassified. In addition, the average spectra could also vary according to the cell cycle. Considering each individual spectrum separately could also provide useful information about the biochemical distribution in a cellular sample. It is true that the biochemistry of cells changes as they move through the cell cycle and this may contribute to the Raman signals. When benign and malignant cell types are compared, it is difficult to selectively sample cells in different phases of the cell cycle without extensive developments of techniques such as flow cytometric cell sorting. Certainly mitotic cells (which can be morphologically identified) were not included in the study. As for the remaining cells that were analyzed, no attempt was made to identify and analyze them by their position in the cell cycle. While there are numerous underlying factors that may be in part responsible for spectral differences, it nevertheless remains true that Raman microscopy, on account of the high information content of the spectra, is capable of distinguishing spectral changes associated with distinct biochemical processes, and here it is effective in being able to distinguish between benign and malignant cells. The ultimate aim is to extend this work into the analysis of human tissue samples, which would make cycle phase specific analysis virtually impossible, and so for this study cells were analyzed regardless of their cell cycle phase. However, understanding the influence of cell cycle phase on Raman metrics remains an interesting area for future research.

A high proportion of the spectra (90%) were correctly classified when using PCA scores, but one cell was persistently misclassified (Figure 5). The increased prediction rate for the individual spectra relative to that of the ABR method reflects the ability of PCA to reduce the influence of noise considerably, allowing the lower-intensity spectra to be more reliably predicted. However, the reduction in the overall prediction rate for full cells suggests that some of the variation observed in the PCA reduction is not related to the difference between the two cell lines. PLSDA directs the data reduction to obtain maximum group separation and is more able to determine the variations important in discriminating the cell lines. As a consequence the prediction rates for the individual spectra and overall cells are enhanced to 94% and 100% respectively.

4.2. Raman Mapping of Scores. PCA generates an efficient and compact representation of the data, particularly useful in Raman maps containing hundreds of collection points, with 512 variables comprising the spectra at each of these points. This multivariate technique reduces the data to a small number of new nonredundant and manageable variables, which account for much of the variance of the original dataset, and enables elucidation of some hidden patterns in the original data.

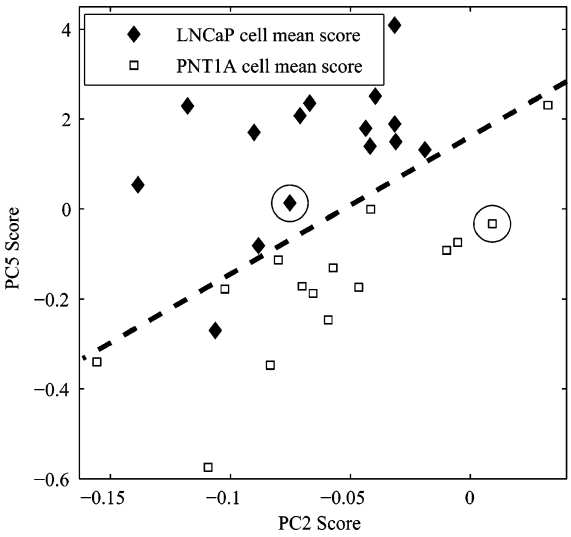

Figure 5. Mean scores for each cell generated by PC2 and PC5, with the two imaged cell scores encircled.

The two cell lines exhibit mainly Raman bands related to amino acids (protein) and DNA. In the PCs sorted in descending order, the first 4 PCs correlate well with bands of proteins and DNA/RNA.

The protein populations appear different between the two cell lines, as demonstrated by the bands known to be sensitive to the secondary structure of proteins. In the malignant cells the intensities of the amide I and III features are stronger in the regions near 1668 and 1242 cm^{-1}, respectively, than those in the case of the benign cells (Figure 1). These positions are characteristic of proteins in β-sheet conformations as compared to positions corresponding to α-helices (the most stable conformation of protein), which occur at 1650−1658 cm^{-1} and at 1265−1274 cm^{-1} for amide I and amide III, respectively, and are more intense in the benign cells.

The 4 PCs show band patterns typical of DNA/RNA, and the scores suggest that the malignant cells have a greater amount of DNA than the benign cells, corroborating the evidence from the subtraction spectrum discussed above (Figure 1).

For the purpose of imaging the distribution of biochemical constituents in two cells of different types, 5 PCs shown in Figure 3 were extracted from unprocessed, mean-centered data. PC1 could be identified with the overall average cellular spectrum. This first component effectively normalizes the spectra, and this PC is eliminated by normalizing the data prior to PCA. PC2 possibly represents variation in DNA and protein with a smooth background contribution. PC3 can be linked mainly to the cytosol, with smaller contributions from nucleic acids. PC4 corresponds to spectra typical of nucleic acids. PC5 is a busy spectrum including a combination of nucleic acids and lipids and protein positive and negative bands. This PC includes primarily the bands used to calculate AB1 and AB2. (See the subtraction spectrum in Figure 1.)

The score maps generated from these five PCs are shown in Figure 4. These maps show the distribution of the various biochemicals discussed above. For instance, the score map for PC4 reveals the distribution of DNA, with gray levels indicating differences in concentration (i.e., white highest, black lowest).

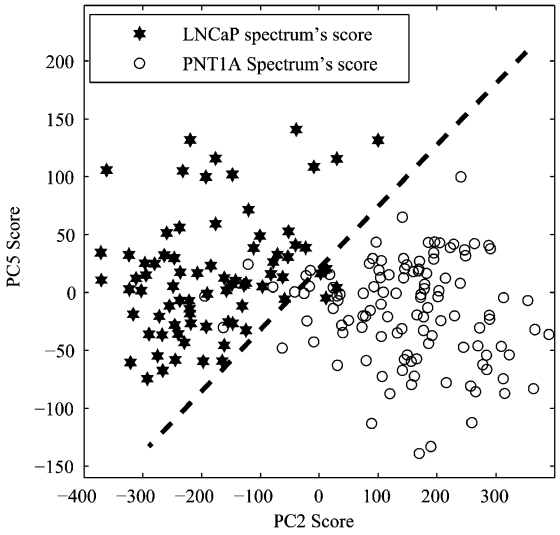

Figure 6. Scores generated by PC2 and PC5 in the imaged cells.

Figure 5 shows a scatter plot of the mean scores of the 30 cells generated by the projection of the normalized and smoothed data onto PC2 and PC5 extracted from the spectra of the two illustrative imaged cells indicated by circles in Figure 5. The separation of the two cell lines is clear, as indicated by the straight line drawn in Figure 5, except for one malignant cell that lies on the wrong side. This demonstrates that the differences, though small, are reproducible. PC5 and PC2 scores provided the highest separation between the two cell lines, accounting for the majority of the discriminant functions used for Table 3. The equivalent scores from the individual spectra of the two cells illustrated in Figure 4 are plotted in Figure 6, demonstrating the consistent level of discrimination across the full area of the cells. This was also confirmed by the LDA discriminant function where the scores of these two components also had the highest discrimination coefficients. This is not surprising as PC5 could be linked to the subtraction spectrum for the two cell types and PC2 to the variation in DNA and protein between the two cell lines.

5. Conclusion

This study demonstrates the potential of Raman microscopy to accurately discriminate and predict the group membership of benign and malignant prostatic cell lines. It was found that by reducing the number of variables by simply taking the ratios of selected adjacent bands in the Raman signal or by employing PLSDA it was possible to accurately classify individual cells of the cell lines with 100% accuracy. However, it was found that PLSDA data reduction was also able to accurately classify a higher proportion of individual spectra within each cell (94% vs 76% of spectra correctly classified) as it is more capable of ignoring irrelevant variations. Raman microscopy can be used to probe the chemometric characteristics of individual cells and demonstrate subtle differences in underlying pathophysiology. This proved valuable in the present instance in distinguishing cells with benign and malignant phenotypes. More significantly the technique could ultimately have potential applications in the identification of cell markers in diagnostic pathology that are not visible to the naked eye and also to probe for changes in cell biochemistry without the use of chromogenic, fluorescent, or bioluminescent markers.

Acknowledgment. This research was supported by funds from the Northern Ireland Health and Personal Social Services Research and Development Office (Grant No. SPI/2384/03) and the Prostate Cancer Charitable Trust. Purchase of the Raman microscope was assisted by funding from the Biotechnology and Biological Sciences Research Council (Joint Research Equipment Initiative Grant No. 18471).

References and Notes

(1) Buschman, H. P.; Deinum, G.; Motz, J. T.; Fitzmaurice, M.; Kramer, J. R.; van der Laarse, A.; Bruschke, A. V.; Feld, M. S. *Cardiovasc. Pathol.* **2001**, *10*, 69.

(2) Choo-Smith, L. P.; Edwards, H. G.; Endtz, H. P.; Kros, J. M.; Heule, F.; Barr, H.; Robinson, J. S., Jr.; Bruining, H. A.; Puppels, G. J. *Biopolymers* **2002**, *67*, 1.

(3) Krafft, C.; Knetschke, T.; Siegner, A.; Funk, R. H. W.; Salzer, R. *Vib. Spectrosc.* **2003**, *32*, 75.

(4) Min, Y.-K.; Yamamoto, T.; Kohda, E.; Ito, T.; Hamaguchi, H.-o. *J. Raman Spectrosc.* **2005**, *36*, 73.

(5) Crow, P.; Stone, N.; Kendall, C. A.; Uff, J. S.; Farmer, J. A.; Barr, H.; Wright, M. P. *Br. J. Cancer* **2003**, *89*, 106.

(6) Stone, N.; Kendall, C.; Shepherd, N.; Crow, P.; Barr, H. *J. Raman Spectrosc.* **2002**, *33*, 564.

(7) Nijssen, A.; Bakker Schut, T. C.; Heule, F.; Caspers, P. J.; Hayes, D. P.; Neumann, M. H.; Puppels, G. J. *J. Invest. Dermatol.* **2002**, *119*, 64.

(8) Notingher, I.; Selvakumaran, J.; Hench, L. L. *Biosens. Bioelectron.* **2004**, *20*, 780.

(9) Bakker Schut, T. C.; Witjes, M. J.; Sterenborg, H. J.; Speelman, O. C.; Roodenburg, J. L.; Marple, E. T.; Bruining, H. A.; Puppels, G. J. *Anal. Chem.* **2000**, *72*, 6010.

(10) Utzinger, U.; Heintzelmam, D.; Mahadevan-Jansen, A.; Malpica, A.; Follen, M.; Richards-Kortum, R. *Appl. Spectrosc.* **2001**, *55*, 955.

(11) Mahadevan-Jansen, A.; Mitchell, M. F.; Ramanujam, N.; Malpica, A.; Thomsen, S.; Utzinger, U.; Richards-Kortum, R. *Photochem. Photobiol.* **1998**, *68*, 123.

(12) Nunes, L. D. O.; Martin, A. A.; Silveira, L., Jr.; Zampieri, M. *Spectroscopy* **2003**, *17*, 597.

(13) Eliasson, C.; Loren, A.; Engelbrektsson, J.; Josefson, M.; Abrahamsson, J.; Abrahamsson, K. *Spectrochim. Acta, Part A* **2005**, *61*, 755.

(14) Feofanov, A. V.; Grichine, A. I.; Shitova, L. A.; Karmakova, T. A.; Yakubovskaya, R. I.; Egret-Charlier, M.; Vigny, P. *Biophys. J.* **2000**, *78*, 499.

(15) Kneipp, K.; Haka, A. S.; Kneipp, H.; Badizadegan, K.; Yoshizawa, N.; Boone, C.; Shafer-Peltier, K. E.; Motz, J. T.; Dasari, R. R.; Feld, M. S. *Appl. Spectrosc.* **2002**, *56*, 150.

(16) Notingher, I.; Verrier, S.; Haque, S.; Polak, J. M.; Hench, L. L. *Biopolymers* **2003**, *72*, 230.

(17) Omberg, K. M.; Osborn, J. C.; Zhang, S. L.; Freyer, J. P.; Mourant, J. R.; Schoonover, J. R. *Appl. Spectrosc.* **2002**, *56*, 813.

(18) Uzunbajakava, N.; Lenferink, A.; Kraan, Y.; Volokhina, E.; Vrensen, G.; Greve, J.; Otto, C. *Biophys. J.* **2003**, *84*, 3968.

(19) Uzunbajakava, N.; Lenferink, A.; Kraan, Y.; Willekens, B.; Vrensen, G.; Greve, J.; Otto, C. *Biopolymers* **2003**, *72*, 1.

(20) Yazdi, Y.; Ramanujam, R.; Lotan, R.; Mitchell, M. F.; Hittelman, W.; Richards-Kortum, R. *Appl. Spectrosc.* **1999**, *53*, 82.

(21) Savitzky, A.; Golay, M. J. E. *Anal. Chem.* **1964**, *36*, 1627.

(22) Hayashi, H.; Nishimura, Y.; Katahira, M.; Tsuboi, M. *Nucleic Acids Res.* **1986**, *14*, 2583.

(23) Le Bihan, T.; Blochet, J. E.; Desormeaux, A.; Marion, D.; Pezolet, M. *Biochemistry* **1996**, *35*, 12712.

(24) Benevides, J. M.; Overman, S. A.; Thomas, G. J., Jr. *J. Raman Spectrosc.* **2005**, *36*, 279.

(25) Krafft, C.; Benevides, J. M.; Thomas, G. J., Jr. *Nucleic Acids Res.* **2002**, *30*, 3981.