Variability in Raman Spectra of Single Human Tumor Cells Cultured in Vitro: Correlation with Cell Cycle and Culture Confluency

QUINN MATTHEWS, ANDREW JIRASEK,* JULIAN LUM, XIAOBO DUAN, and ALEXANDRE G. BROLO

Dept. of Physics & Astronomy, University of Victoria, Victoria, B.C., Canada (Q.M., A.J.); Deeley Research Centre, BC Cancer Agency, Victoria, B.C., Canada (J.L., X.D.); and Dept. of Chemistry, University of Victoria, Victoria, B.C., Canada (A.G.B.)

In this work we investigate the capability of Raman microscopy (RM) to detect inherent sources of biochemically based spectral variability between single cells of a human tumor cell line (DU145) cultured in vitro. Principal component analysis (PCA) is used to identify differences in single-cell Raman spectra. These spectral differences correlate with (1) cell cycle progression and (2) changing confluency of a cell culture during the first 3 to 4 days after sub-culturing. Cell cycle regulatory drugs are used to synchronize the cell cycle progression of cell cultures, and flow cytometry is used to determine the cell cycle distribution of cell cultures at the time of Raman analysis. Spectral variability arising from cell cycle progression is (1) expressed as varying intensities of protein and nucleic acid features relative to lipid features, (2) well correlated with known biochemical changes in cells as they progress through the cell cycle, and (3) shown to be the most significant source of inherent spectral variability between cells. Furthermore, the specific biomolecules responsible for the observed spectral variability due to both cell cycle progression and changes in cell culture confluency can be identified in the first and second components of principal component analysis (PCA). Our characterization of the inherent sources of variability in Raman spectra of single human cells will be useful for understanding subtle spectral differences in RM studies of single cells.

Index Headings: Raman microscopy; Spectral variability; Human cells; Cell biochemistry; Cell cycle; Confluency; Principal component analysis; PCA.

INTRODUCTION

In the last decade, Raman microscopy (RM) has emerged as a prominent tool for the biochemical analysis of human cells and tissues. A recent study found that Raman spectroscopy provides biochemical information at comparable levels of accuracy and sensitivity as established techniques such as NMR spectroscopy and flow cytometry.1 RM is an attractive modality both for its ability to attain biochemical information from proteins, nucleic acids, lipids, and carbohydrates in a single acquisition and for its noninvasiveness and nondestructiveness when used with an appropriate choice of laser wavelength and power.2–4 Furthermore, the use of high-power focusing objectives, confocal optics, and sub-micrometer resolution stepping stages enables the analysis of single cells in vitro, whereas fiber-optic-based Raman systems may, in the future, allow the technology to be applied to clinical patients in vivo.

There has been considerable interest in applying RM to live cells and tissues for cancer detection and diagnosis. Due to the complexity of a single biological Raman spectrum, RM is commonly used in conjunction with multivariate statistical methods such as principal component analysis (PCA), linear discriminant analysis (LDA), or cluster analysis. RM with the use of multivariate methods has been successfully applied to discriminate between healthy and cancerous skin,5–7 bladder,8 and gastric9,10 tissues. Applying RM and multivariate methods to the analysis of single cells has shown the ability to distinguish between healthy and tumorigenic rat fibroblast cells,11 human bone cells,12 and human epithelial cells from a variety of organs.13 Similar techniques have been employed to discriminate between different types of human tumor cell lines from a mixed sample set.14–16

The single-cell studies mentioned above have focused on the differences in Raman spectra between different cell types (i.e., healthy vs. tumor, prostate vs. bladder). In addition, several studies have used RM for analyzing biochemical differences arising within a population of a single type of cell. One pair of studies detected spectral changes from cell death via apoptosis17 and spectral differences between live and dead cells,18 and another pair of studies detected spectral changes due to cell death via necrosis19 and spectral differences between exponentially growing (proliferative) and plateau phase (non-proliferative) cells.20 The former pair of studies compared the averaged Raman spectra of many single cells from each sample, and the latter pair of studies compared spatially averaged spectra, obtained from many cells at once, from a pellet of cells. One recent study applied PCA and LDA to Raman spectra of single live cells to discriminate between cells synchronized in the G0/G1, S, and G2/M phases of the cell cycle and demonstrated accurate discrimination of G0/G1 cells from S and G2/M cells.21

Within a given population of cells, one can expect biochemical differences between individual cells due to a number of reasons, such as cells existing at different points in their cell cycle (in the case of an asynchronously growing cell culture) or cells growing in cultures of different confluence (in the case of multiple cell cultures). In all the studies mentioned above, natural biochemical differences exist between individual cells in a sample. In all these previous studies (with the exception of the recent cell cycle discrimination study21), spectral variability arising from inherent biochemical differences was either averaged during spectral acquisition or post-processing or was not a relevant or necessary consideration for the purpose of the study. However, the spectral differences between cell populations may be very subtle, and it is important to thoroughly investigate the existing sources of spectral variability within a given population. To our knowledge, there have been no systematic studies on the

Received 22 December 2009; accepted 12 May 2010.

* Author to whom correspondence should be sent. E-mail: jirasek@uvic.ca.
inherent sources of spectral variability that may arise due to biochemical differences between single cells.

In this work, we have undertaken an investigation of the capability of RM to detect inherent sources of spectral variability within a human tumor cell line (DU145) cultured in vitro. PCA is used to observe differences in Raman spectra that correlate with cells existing in different phases of the cell cycle, as well as differences correlating with the confluence of the cell culture at the time of Raman analysis. Furthermore, the biochemical changes detected by RM that correlate with cell cycle progression are consistent with known biochemical changes within cells. We also show that the variability from cell cycle and culture confluency comprises almost all of the inherent variability in a multi-culture data set and is primarily explained by the first two PCA components.

The results of this work are presented in two studies. In the first study, the inherent variability between cultures of varying confluence is investigated by acquiring Raman spectra of cells collected from asynchronously growing cell cultures harvested 1–8 days after sub-culturing. In the second study, the variability due to cell cycle in sub-confluent cultures is directly examined by acquiring spectra from cells collected from cultures synchronized at specific points in the cell cycle. Flow cytometry is used to monitor the cell cycle distribution and viability of all cultures at the time of Raman analysis. We use adherent cells that have been resuspended and centrifuged into a pellet, from which individual cells are selected using a high-power focusing objective and 785 nm laser excitation. This technique provides a very high quality Raman spectrum of a single cell, while eliminating any spectral variability (caused by inconsistent cell structure and varying local cell density) arising when cells are grown and analyzed directly in a monolayer. 25 Both low-wavenumber (LWN), 600–1800 cm−1, and high-wavenumber (HWN), 2700–3100 cm−1, spectral windows are used in order to determine whether information can be attained equivalently from either window, as some authors have found to be the case for certain applications.25

METHODS

Cell Preparation. Standard Cell Culture. Human prostate tumor cells (cell line DU145 (ATCC, Manassas, VA)) were cultured in a sterile environment and grown in T-75 flasks with 15 mL of Dulbecco Modified Eagle Medium (HyClone, Logan, UT) supplemented with 10% fetal bovine serum (FBS) (HyClone). Cultures were kept in an incubator at 5% CO2 and 37 °C to promote growth. Cell stocks were sub-cultured every 3 to 4 days by rinsing the cells in phosphate buffered saline (PBS), adding trypsin (HyClone) to detach the cells from the flask, and transferring 10–20% of the harvested cells to a new flask containing fresh media. Before re-incubation, the new cell suspensions were pipetted several times to ensure an evenly distributed monolayer of cells throughout the flask.

Asynchronous Cell Cultures. A single T-75 flask was grown to ∼90% confluency and sub-cultured equally into eight identical T-75 flasks. The first flask was harvested for RM analysis 24 hours after sub-culturing, and each remaining flask was harvested every 24 hours thereafter. Each time a flask was harvested, the confluency, cell cycle distribution, and viability of the culture was measured, as described below.

Synchronized Cell Cultures. Using established protocols,24,25 2 mM of thymidine (Sigma-Aldrich, Oakville, ON, Canada) was used to inhibit DNA replication and arrest cells at the end of G1 phase, before the onset of S phase, and 100 ng/mL of nocodazole (Sigma-Aldrich) was used to prevent formation of the mitotic spindle and arrest cells at the end of G2 phase, prior to M phase. In the second study presented in this work, cell cultures were treated with thymidine and nocodazole to obtain cultures synchronized at four points during the cell cycle: (1) at the G1/S boundary (with double-thymidine treatment), (2) during S phase (with double-thymidine treatment and re-incubation with drug-free media for 3 hours), (3) at the G2/M boundary (with thymidine treatment for 24 hours, re-incubation with drug-free media for 3 hours, nocodazole treatment for 12 hours, and mitotic shake-off) and (4) during early G1 phase (with mitotic shake-off from a G2/M arrested culture and re-incubation of detached cells with fresh media for 5 hours).

Cell Cycle, Viability, and Confluence Analysis. The cell cycle distribution and viability of each culture was measured using flow cytometry,26 following established protocols.27,28 For cell cycle analysis, over 200,000 cells were extracted from a culture and fixed in 70% ethanol to permeabilize the cell membrane. RNase A (Qiagen Inc., Mississauga, ON, Canada) was added to a concentration of 1 mg/mL in order to degrade cellular RNA. Propidium iodide (PI) (Sigma) was subsequently added to a final concentration of 50 μg/mL. After 30 minutes the suspension was centrifuged, re-suspended in a buffer (PBS + 1% FBS), and kept on ice until analysis. For flow cytometry collection, the relative DNA content of 100,000 cells was measured using a BD FACSCalibur Flow Cytometer (BD Biosciences, Mississauga, ON, Canada). Cell counts were recorded by the flow cytometer during sample acquisition. Relative fractions of cells in each phase of the cell cycle (G1, S, or G2) were determined by performing a nonlinear least-squares fit to the measured data using functions representative of the expected distributions of DNA content for each cell cycle phase (Matlab, The Mathworks, Natick, MA).

For culture viability assessment, over 100,000 cells were extracted from the culture and split into two equal parts. One part was stained with PI at a concentration of 5 μg/mL, and the other part was left untreated to serve as a control. 20,000 cells from each sample underwent flow cytometry analysis within 15 minutes of staining, and the fractions of live (no PI signal) and dead (positive PI signal) cells in the PI stained sample were determined.

Cell culture confluency is defined as the percentage of the surface area of the culture flask covered by cells. Confluency estimates of each of the asynchronous cell cultures were obtained prior to harvesting by acquiring low magnification images of five different regions of the cell culture. Each image was imported into Matlab and the fraction of the image covered by cells was calculated. Because the confluency is never consistent throughout the entire surface area of the flask, the averaged confluency from the five regions was used as an estimate of the overall confluency of the culture.

Raman Microscopy. To prepare cells for Raman analysis, cultures were harvested by rinsing with PBS to remove dead cells and debris, adding trypsin to detach the remaining live cells, and centrifuging to discard the trypsin supernatant. Cells were re-suspended in growth media and centrifuged into a pellet in a 200 μL plastic vial. Vials were kept on ice until RM analysis (1–6 hours), upon which the chosen pellet was transferred to a low-fluorescence quartz disc (Technical Glass
Products, Painesville, OH) in order to minimize spectral contributions from the sample substrate. All Raman spectra were acquired within 2 hours of transferring the pellet to the quartz disc. We have observed no spectral variations that correlate with the time of sample removal from the ice bath, suggesting that any effect of removing cells from the ice bath to an exposed environment has negligible impact on our RM analysis within this two-hour time interval.

Raman analysis was performed on an InVia Raman Microscope (Renishaw Inc., Hoffman Estates, IL) with a 100× dry objective (N.A. = 0.9) (Leica Microsystems, Wetzlar, Germany) and a 1200 lines/mm diffraction grating. A 785 nm continuous wave diode laser (Renishaw) was used for sample excitation, providing a laser power density at the sample of \( \sim 0.5 \) mW/\( \mu \)m\(^3\). The size of the sampling volume was measured to be \( \sim 2 \times 5 \times 10 \) \( \mu \)m; these dimensions allow a single acquisition to represent the spectrum of a single cell. Raman spectra were acquired from 20 individual cells from each sample, chosen from the top layer of the cell pellet (Fig. 1). Spectra were collected at 30-second acquisitions per cell; the LWN window (600–1800 cm\(^{-1}\)) and the HWN window (2700–3100 cm\(^{-1}\)) spectra were acquired in succession for each cell. The wavenumber range for each spectral window was covered in a single acquisition using the Renishaw’s SynchroScan operation mode.

**Spectral Processing.** Prior to PCA analysis, each spectrum was processed to remove cosmic rays, increase the signal-to-noise ratio via spectral smoothing, subtract a baseline arising from the quartz substrate and biological fluorescence, and normalize to the amount of biological material within the sampling volume. All data processing was performed with Matlab.

Spectral smoothing was performed with an in-house version of the previously described two-point maximum entropy method,\(^{29,30}\) which has been particularly successful when applied to Raman spectra.\(^{31,32}\) In this work, a very modest amount of smoothing was applied in order to maintain fidelity of the sharpest Raman peaks in the spectra.

The large number of spectra collected in this study necessitated the use of automated baseline removal methods. An effective and robust baseline removal method is critical for this work in order to remove sources of variability arising from varying levels of fluorescence or quartz substrate contamination (addressed further in the Discussion section). For the LWN spectral window, we used a modified signal removal method.\(^{33}\) This method was chosen due to the mixture of sharp and broad features throughout the spectral window and the need for a highly conformal baseline around the regions of quartz contamination (\( \sim 800 \) cm\(^{-1}\) and \( \sim 1050 \) cm\(^{-1}\)) and a broader baseline from 1100 to 1800 cm\(^{-1}\), where many overlapping peaks give rise to broad Raman features. For the HNW spectral window, a three-point linear interpolated baseline was found to be sufficient for baseline removal.

After baseline removal, the principal remaining source of variability between spectra is the overall intensity of the Raman features, arising from the variable amount of biological material within the sampling volume. In the present work, this variability arises from slightly different physical shapes and orientations of each cell in the cell pellet. To remove this source of variability, each spectrum was normalized to the total area under the baseline-corrected spectrum. Other authors have addressed this issue of intensity variability by normalizing to the area under the CH deformation peak at 1450 cm\(^{-1}\), thought to be proportional to the total amount of biological material within the sampling volume.\(^{17,18}\) In our work, we have found that the CH deformation peak can vary independently of other Raman peaks and therefore may not be suitable as a normalization peak in all cases. For example, in this study we report that one of the most significantly varying Raman peaks between cells arises from CH\(_2\) deformation in lipids at 1438 cm\(^{-1}\), which affects the area of the CH deformation peak as well due to its close spectral proximity. Furthermore, we have found that the method of normalizing to the total area under the baseline-corrected spectrum is suitable for both the LWN and HWN spectral windows.

Principal component analysis was performed using standard algorithms (Matlab). PCA calculations were performed separately on the LWN and HNW window data sets to facilitate an independent comparison and corroboration of results obtained from each window. In this work, spectral variability arising from the quartz substrate was easily identifiable in a single PCA component from the LWN window; the quartz component was therefore removed and the PCA calculation was repeated on the filtered set of spectra. It is important to note that this action does not affect the other LWN window components, but only redistributes the variance explained by the excluded component among the remaining components.

**RESULTS**

**Single DU145 Cell Spectrum.** In the LWN spectral window, the Raman spectrum of a single DU145 cell (Fig. 2a) contains multiple contributions from proteins, lipids, and nucleic acids. Spectral features of proteins arise from aromatic amino acids (phenylalanine, tryptophan, and tyrosine), amide groups of secondary protein structures (\( \alpha\)-helices, \( \beta\)-sheets, and random coils), and the stretching or deformation of carbon atoms bonded with nitrogen, hydrogen, or other carbon atoms. Nucleic acid features include contributions from individual RNA and DNA bases (adenine, thymine, guanine, cytosine, and uracil), as well as from the sugar-phosphate backbone of DNA. A number of different lipid features are also detectable throughout the spectral window. In the HNW window, the
spectrum (Fig. 2b) is a superposition of broad features dominated by the stretching of various lipid and protein CH$_2$ and CH$_3$ groups. There are also weak contributions from =CH stretching in lipids and from aromatic groups in both nucleic and amino acids. A detailed listing of the molecular assignments$^{11,18,34–38}$ for all spectral features observed in this work is provided.$^{11,18,34–38}$

Study #1: Asynchronous Cell Cultures. Cell Cycle, Confluency, and Viability. The cell cycle distributions and culture confluencies for the eight samples in this study (Fig. 3) are typical for asynchronous cells growing to confluency in culture. From 24 to 72 hours after sub-culturing the distribution among the three phases is fairly constant at ~50% G1, ~20% G2, and ~30% S. Between 72 and 96 hours we see an increase in the G1 phase and a decrease of both the G2 and S phases. After 96 hours the G2 content remains relatively constant, whereas the S content decreases, and the G1 content increases, until about 168 hours. One element that is not measurable with this method of cell cycle analysis is the fraction of cells in “G0” phase, a state of cellular quiescence. Cellular quiescence is only achievable during G1 phase, usually soon after cell division; therefore, G0 cells are indistinguishable from G1 cells by the flow cytometry methods used in this study.

The viability of the harvested cells was determined with flow cytometry prior to Raman analysis. Dead cells will usually detach from the growth substrate and subsequently be rinsed off and discarded during the harvesting procedure. However, a small percentage of dead cells will always remain in a harvested culture. For this study, viability tests proved that all of the first seven samples (24–168 hours after sub-culturing) had a viability of >98% (i.e., less than 2% dead cells), and the eighth and final sample (192 hours after sub-culturing) had a viability of >95%.

First Principal Component. For the 160 cell spectra collected in this study, the first PCA components (Fig. 4) represent the most significant source of spectral variability in each data set (52.6% of the total variance for the LWN window, 88.6% for the HWN window). By comparison with the known Raman shifts (Fig. 2, Table I), the features in the PCA components for both the LWN and HWN window are identifiable as arising from variability in the Raman intensity of peaks in the original data set; therefore, one can assign a molecular origin to the features in the components. The PCA components consist of both positive and negative features; any spectrum that is assigned a higher (i.e., more positive) PCA score for a given component will have a proportionately higher amount of the positive features and a lower amount of the negative features from that component. It should be noted that the positive or negative nature of the features is purely arbitrary and only holds meaning with respect to the sign of the corresponding PCA scores. Any component can be reflected about zero, with a corresponding change of sign for all scores for that component, without altering the results of the PCA transformation.

The negative features in the first PCA component for the LWN window (Fig. 4a) are dominated by lipid contributions from cholesterol, CH$_2$ twisting, CH$_2$ and CH deformation, and C–C, C=C, and C=O stretching. The C–C features at 1065, 1080, and 1127 cm$^{-1}$, the CH feature at 1460 cm$^{-1}$, and the C=C feature at 1656 cm$^{-1}$ have the same Raman shifts as the protein contributions from C–N stretching, CH deformation, and $\alpha$-helix amide groups, respectively. As such, it is impossible to determine whether or not this variability is solely lipid or protein in nature, or some combination of both; however, as the rest of the negative features in the component are uniquely lipid in origin, it follows that the negative nature of these features arises in part from lipids as well. The negative feature at 844 cm$^{-1}$ was not identifiable. The positive features in the LWN component are almost exclusively nucleic acid and protein in origin, with the exception of a weak positive contribution from choline at 719 cm$^{-1}$. Nucleic acid features arise from DNA and RNA bases and from the DNA backbone. Protein features arise from aromatic amino acids (phenylalanine, tryptophan, and tyrosine) and from $\beta$-sheet amide groups.

FIG. 2. Raman spectra of a single DU145 cell, for the (a) LWN and (b) HWN spectral windows; the Raman shift and molecular origin of identifiable features are provided.$^{11,18,34–38}$ Abbreviations: (p) protein, (l) lipid, (d) DNA/RNA, (A) adenine, (T) thymine, (G) guanine, (C) cytosine, (U) uracil, (Phe) phenylalanine, (Tyr) tyrosine, (Trp) tryptophan, (bk) backbone, (def) deformation, (tw) twist, (sym) symmetric, (asym) asymmetric, and (str) stretch.
Interestingly, it is known that the aromatic amino acids are most likely to be found in a β-sheet conformation and less likely to be found in an α-helix or coiled structure.\(^{39}\) As such, an increase in signal from the aromatic amino acids should be correlated with an increase in signal from β-sheet amide groups, which we see to be the case here. For the HWN window (Fig. 4b), the positive features can be assigned to the symmetric and asymmetric stretching of CH3 groups in both proteins and lipids. The negative features, however, arise from the symmetric and asymmetric stretching of CH2 groups in lipids alone. To summarize, the negative features in both the LWN and the HWN window are primarily due to lipids, whereas the positive features in both the LWN and the HWN window are primarily due to nucleic acids and proteins (in particular from amino acids and β-sheet amides for the LWN window).

The PCA scores (Fig. 5) determine how much of the variability explained by the first components (Fig. 4) is expressed in each of the 160 cell spectra in each data set. Note that positive scores are correlated with increased nucleic acid and protein content, and negative scores are correlated with increased lipid content. The scores for both windows

### TABLE I. Molecular assignments for spectra of DU145 cells. Superscript numbers indicate references used for particular assignment.\(^a\)

<table>
<thead>
<tr>
<th>Raman shift (cm(^{-1}))</th>
<th>DNA/RNA</th>
<th>Proteins</th>
<th>Lipids</th>
</tr>
</thead>
<tbody>
<tr>
<td>622</td>
<td></td>
<td>C-C tw Phe(^ {34,35})</td>
<td></td>
</tr>
<tr>
<td>644</td>
<td></td>
<td>C-C tw Tyr(^ {34,35})</td>
<td></td>
</tr>
<tr>
<td>669</td>
<td></td>
<td>Trp ring br(^ {34,35})</td>
<td></td>
</tr>
<tr>
<td>700</td>
<td></td>
<td>Tyr ring br(^ {34–37})</td>
<td></td>
</tr>
<tr>
<td>728</td>
<td></td>
<td>C-C sym str bk, α-helix(^ {11,34–37})</td>
<td></td>
</tr>
<tr>
<td>759</td>
<td></td>
<td>CH def(^ {18,34})</td>
<td></td>
</tr>
<tr>
<td>784</td>
<td></td>
<td>CH def(^ {18,34})</td>
<td></td>
</tr>
<tr>
<td>811</td>
<td></td>
<td>Phe sym ring br(^ {11,34–37})</td>
<td></td>
</tr>
<tr>
<td>828</td>
<td></td>
<td>C–N str(^ {11,34–36})</td>
<td></td>
</tr>
<tr>
<td>853</td>
<td></td>
<td>C–N str(^ {11,34–36})</td>
<td></td>
</tr>
<tr>
<td>877</td>
<td></td>
<td>C–N str(^ {11,34–36})</td>
<td></td>
</tr>
<tr>
<td>936</td>
<td></td>
<td>C–N str(^ {11,34–36})</td>
<td></td>
</tr>
<tr>
<td>975</td>
<td></td>
<td>C–N str(^ {11,34–36})</td>
<td></td>
</tr>
<tr>
<td>1003</td>
<td></td>
<td>C–N str(^ {11,34–36})</td>
<td></td>
</tr>
<tr>
<td>1032</td>
<td></td>
<td>C–N str(^ {11,34–36})</td>
<td></td>
</tr>
<tr>
<td>1063</td>
<td></td>
<td>C–N str(^ {11,34–36})</td>
<td></td>
</tr>
<tr>
<td>1080</td>
<td></td>
<td>C–N str(^ {11,34–36})</td>
<td></td>
</tr>
<tr>
<td>1094</td>
<td></td>
<td>C–N str(^ {11,34–36})</td>
<td></td>
</tr>
<tr>
<td>1100</td>
<td></td>
<td>C–N str(^ {11,34–36})</td>
<td></td>
</tr>
<tr>
<td>1127</td>
<td></td>
<td>C–N str(^ {11,34–36})</td>
<td></td>
</tr>
<tr>
<td>1158</td>
<td></td>
<td>C–N str(^ {11,34–36})</td>
<td></td>
</tr>
<tr>
<td>1175</td>
<td></td>
<td>C–N str(^ {11,34–36})</td>
<td></td>
</tr>
<tr>
<td>1208</td>
<td></td>
<td>C–N str(^ {11,34–36})</td>
<td></td>
</tr>
<tr>
<td>1230</td>
<td></td>
<td>C–N str(^ {11,34–36})</td>
<td></td>
</tr>
<tr>
<td>1246</td>
<td></td>
<td>C–N str(^ {11,34–36})</td>
<td></td>
</tr>
<tr>
<td>1255</td>
<td></td>
<td>C–N str(^ {11,34–36})</td>
<td></td>
</tr>
<tr>
<td>1267</td>
<td></td>
<td>C–N str(^ {11,34–36})</td>
<td></td>
</tr>
<tr>
<td>1300</td>
<td></td>
<td>C–N str(^ {11,34–36})</td>
<td></td>
</tr>
<tr>
<td>1320</td>
<td></td>
<td>C–N str(^ {11,34–36})</td>
<td></td>
</tr>
<tr>
<td>1340</td>
<td></td>
<td>C–N str(^ {11,34–36})</td>
<td></td>
</tr>
<tr>
<td>1374</td>
<td></td>
<td>C–N str(^ {11,34–36})</td>
<td></td>
</tr>
<tr>
<td>1421</td>
<td></td>
<td>C–N str(^ {11,34–36})</td>
<td></td>
</tr>
<tr>
<td>1438</td>
<td></td>
<td>C–N str(^ {11,34–36})</td>
<td></td>
</tr>
<tr>
<td>1450</td>
<td></td>
<td>C–N str(^ {11,34–36})</td>
<td></td>
</tr>
<tr>
<td>1460</td>
<td></td>
<td>C–N str(^ {11,34–36})</td>
<td></td>
</tr>
<tr>
<td>1486</td>
<td></td>
<td>C–N str(^ {11,34–36})</td>
<td></td>
</tr>
<tr>
<td>1577</td>
<td></td>
<td>C–N str(^ {11,34–36})</td>
<td></td>
</tr>
<tr>
<td>1607</td>
<td></td>
<td>C–N str(^ {11,34–36})</td>
<td></td>
</tr>
<tr>
<td>1618</td>
<td></td>
<td>C–N str(^ {11,34–36})</td>
<td></td>
</tr>
<tr>
<td>1656</td>
<td></td>
<td>C–N str(^ {11,34–36})</td>
<td></td>
</tr>
<tr>
<td>1660</td>
<td></td>
<td>C–N str(^ {11,34–36})</td>
<td></td>
</tr>
<tr>
<td>1669</td>
<td></td>
<td>C–N str(^ {11,34–36})</td>
<td></td>
</tr>
<tr>
<td>1685</td>
<td></td>
<td>C–N str(^ {11,34–36})</td>
<td></td>
</tr>
<tr>
<td>1743</td>
<td></td>
<td>C–N str(^ {11,34–36})</td>
<td></td>
</tr>
<tr>
<td>253</td>
<td></td>
<td>CH def(^ {18,34})</td>
<td></td>
</tr>
<tr>
<td>2858</td>
<td></td>
<td>CH def(^ {18,34})</td>
<td></td>
</tr>
<tr>
<td>2935</td>
<td></td>
<td>CH def(^ {18,34})</td>
<td></td>
</tr>
<tr>
<td>2960–2980</td>
<td></td>
<td>CH def(^ {18,34})</td>
<td></td>
</tr>
<tr>
<td>3008</td>
<td></td>
<td>CH def(^ {18,34})</td>
<td></td>
</tr>
<tr>
<td>3060</td>
<td></td>
<td>CH def(^ {18,34})</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) Abbreviations: (A) adenine, (U) uracil, (C) cytosine, (T) thymine, (G) guanine, (Trp) tryptophan, (Tyr) tyrosine, (Phe) phenylalanine, (br) breathing, (bk) backbone, (def) deformation, (tw) twist, (sym) symmetric, (asym) asymmetric, and (str) stretch.
show the same overall trend; between 24 and 48 hours after sub-culturing there is an increase in the average nucleic acid and protein content relative to the average lipid content, followed by a steady decrease in the average nucleic acid and protein content relative to the average lipid content from 48 to 192 hours after sub-culturing. Furthermore, the relative positions of the individual cell scores are consistent between the LWN and HWN windows. For example, cells #30 and #39
Fig. 5. PCA scores for the first components from the asynchronous cell cultures study, for the (a) LWN and (b) HWN window. Scores for all 160 cells are grouped by time of harvest after sub-culturing. The average score and standard deviation are shown for each sample for visualization of the trends in the data. The Raman spectra of cell #30 and cell #148 are shown in Fig. 6.

(Fig. 5) have, respectively, the highest and lowest scores from the 48-hour LWN window sample set, and the same two cells have, respectively, the highest and lowest scores from the corresponding HWN window sample set. It is worth emphasizing that the intra-sample variability in the PCA scores for a given PCA component arises from the same source of spectral variability as the inter-sample variability and simply reflects the intrinsic biochemical heterogeneity of each cell culture.

To show that the variability predicted by the PCA analysis is directly observable in the original data, the Raman and difference spectra for two cells (cells #30 and #148) having a large separation in their PCA scores (Fig. 5) are shown in Fig. 6, along with the PCA components for comparison. All of the major features in the components are directly observable in the corresponding difference spectrum for each spectral window, without any rescaling of the difference spectra.

**Second Principal Component.** The second PCA component for the LWN window (Fig. 7a) explains 10.1% of the total variance, and the corresponding component for the HWN window (Fig. 7b) explains only 1.7% of the total variance. Assigning a molecular origin to the features in the second components is more difficult than for the first components,

![Fig. 5](image-url)  
![Fig. 6](image-url)

**Fig. 6.** Raman and difference spectra for two cells (#30 and #148) having a large difference in PCA score (Fig. 5) for the first PCA component. The first PCA components have been offset and rescaled for comparison with the unscaled difference spectra. Wavenumbers are provided for any known features in the components (Fig. 4) that are also observable in the difference spectra.
especially for the HWN window, where the percent variance explained is very low and there is a small number of known molecules contributing to the HWN spectra (Fig. 2b, Table I). The only feature in the HWN window that corresponds with a known wavenumber is the symmetric stretching of CH$_3$ groups at 2935 cm$^{-1}$ (Fig. 7b), although the accuracy of this assignment is uncertain. However, for the LWN window (Fig. 7a) almost all of the major features can be assigned with confidence. The positive features arise from amino acids, amide groups in $\beta$-sheet and random coil conformation, and a combined contribution from the nucleic acid bases A and G and CH deformation in proteins. The origin of the positive feature at 1120 cm$^{-1}$ is unknown. The negative features include a strong contribution from choline, as well as contributions from O–P–O stretching in lipids and RNA, the nucleic acid bases A and G, and a combined contribution from lipid $\equiv$CH deformation and $\alpha$-helix amide groups. The sharp negative feature at 1660 cm$^{-1}$ arises from amide groups as well, but whether it arises from a certain protein conformation, or from amide groups in general, is unknown. It is also unclear as to why contributions from the nucleic acids A and G appear in both the positive and negative features of the component. Despite the uncertainty of the molecular origins of the features in the second PCA components (especially for the

Fig. 7. Second PCA components from the asynchronous cell cultures study: (a) LWN window (10.1% of total variance), (b) HWN window (1.7% of total variance). The Raman shift and molecular origin of identifiable features are provided. Abbreviations: (p) protein, (l) lipid, (d) DNA/RNA, (A) adenine, (G) guanine, (Phe) phenylalanine, (Tyr) tyrosine, (def) deformation, (sym) symmetric, and (str) stretch.

Fig. 8. PCA scores for the second components from the asynchronous cell cultures study, for the (a) LWN and (b) HWN window. Scores for all 160 cells are grouped by time of harvest after sub-culturing. The average score and standard deviation is shown for each sample for visualization of the trends in the data. The Raman spectra of cells indicated by arrows are shown in Fig. 9.

878 Volume 64, Number 8, 2010
HWN window), the scores for both windows still show the same general trend: from 24 to ~120 hours after sub-culturing there is an overall increase in the average scores, and after ~120 hours the average scores appear to remain relatively constant until decreasing slightly between 168 and 192 hours. However, the relative positions of the individual cell scores between the LWN and HWN windows are not consistent. Therefore, the similar trends between the two windows may not be the result of the same biomolecular changes occurring within the cells.

To determine whether the variability predicted by the second PCA components is directly observable in the original data (as it was with the first PCA components (Fig. 6)), the Raman and difference spectra for two cells (cells #137 and #19 for the LWN window, and cells #114 and #31 for the HWN window) having a large separation in their PCA scores (Fig. 8) are shown in Fig. 9, along with the PCA components for comparison. For the LWN window, all of the major features in the component are observable in the LWN difference spectrum. However, the features in the HWN component are not observable in the HWN difference spectrum.

Other Principal Components. The third PCA component for the LWN window (not shown) explains 7.6% of the total variance and is dominated by a sharp derivative-like feature centered at the wavenumber of the sharp phenylalanine ring breathing peak at 1003 cm\(^{-1}\). This feature in the PCA component indicates variability arising from a shift in the calibration of the Raman system over time. The trend displayed by the score plots for this component (also not shown) correlates well with a known drift in the Raman calibration over the eight-day sample collection period, which was monitored by measuring peak shifts of the 520 cm\(^{-1}\) feature of an instrument-based silicon sample before and after the daily Raman collection. During daily collections, it was verified that the initial calibration of the system was within 0.5 cm\(^{-1}\) of the calibration performed on the first day of collection. By inspection of various pairs of spectra with large differences in their scores for the third PCA component, it was found that the maximum shift in the position of the phenylalanine peak at 1003 cm\(^{-1}\) was less than one pixel (1 pixel \(\approx 0.9\) cm\(^{-1}\)).

Fig. 9. Raman and difference spectra for two cells (#137 and #19 for the LWN window, and #114 and #31 for the HWN window) having a large difference in PCA score (Fig. 8) for the second PCA component. The second PCA components have been offset and rescaled for comparison with the unscaled difference spectra. Wavenumbers are provided for any known features in the components (Fig. 7) that are also observable in the difference spectra.
the combined fraction of cells in G1 or S phase was estimated to be \( \sim 26\% \), with at least 74% of the cells successfully arrested at the G2/M boundary (Fig. 10c, “G2/M”). Five hours after harvesting and re-incubating cells from an identical G2/M arrest, \( \sim 21\% \) of the fourth culture was determined to be left in G2 phase, while \( \sim 75\% \) of the culture was now found in G1 phase (Fig. 10d, “G2/M +5 hrs”). Since the fourth culture was seeded with cells that were primarily at the G2/M boundary, the G1 cells in the fourth culture must be less than 4 to 5 hours into G1 phase.

**First Principal Component.** The first PCA component for the LWN window (Fig. 11a) explains 51.6% of the total variance and is very similar to the corresponding component from the asynchronous cell cultures study (Fig. 4a), which explained 52.6% of the total variance. As in the previous study, the negative features in the component are dominated by lipid contributions from cholesterol, \( \text{CH}_2 \) twisting, \( \text{CH}_2 \) and CH deformation, and \( \text{C–C}, \text{C=} \), and \( \text{C}–\text{O} \) stretching, with an additional negative contribution from choline, which previously contributed as a weak positive feature in the asynchronous study. There is also a new negative feature at 1267 cm\(^{-1}\), which is a combined contribution from lipid \( \text{CH} \) deformation and \( \alpha \)-helix amide groups; this feature correlates with the existing negative combined contribution from lipid \( \text{C}–\text{C} \) stretching and \( \alpha \)-helix amides at 1656 cm\(^{-1}\). The previously observed negative features at 844 and 1127 cm\(^{-1}\) are not observed here. The positive features in the LWN component, as in the previous study, are exclusively nucleic acid and protein in origin, with contributions from DNA and RNA bases, the DNA backbone, aromatic amino acids, and \( \beta \)-sheet amide groups. In this study there are additional positive contributions from tyrosine at 853 cm\(^{-1}\), thymine at 1374 cm\(^{-1}\), and random coil amide groups at 1230 cm\(^{-1}\). The previously observed positive feature at 811 cm\(^{-1}\) is not observed here. The first

---

**Fig. 10.** Flow cytometry analysis of cell cycle distributions for the synchronized cell cultures. Synchronization was performed using thymidine and nocodazole as described in the Materials and Methods section.

**Fig. 11.** First PCA components from the synchronized cell cultures study: (a) LWN window (51.6% of total variance), (b) HWN window (86.6% of total variance). The Raman shift and molecular origin of identifiable features are provided.11,18–34-38 Abbreviations: (p) protein, (l) lipid, (d) DNA/RNA, (A) adenine, (T) thymine, (G) guanine, (C) cytosine, (U) uracil, (Phe) phenylalanine, (Tyr) tyrosine, (Trp) tryptophan, (bk) backbone, (def) deformation, (tw) twist, (sym) symmetric, (asym) asymmetric, and (str) stretch.
PCA component for the HWN window (Fig. 11b) explains 86.6% of the total variance and is nearly identical to the corresponding component from the asynchronous cell cultures study (Fig. 4b), which explained 88.6% of the total variance. As before, the positive features arise from the symmetric and asymmetric stretching of CH₃ groups in both proteins and lipids, whereas the negative features arise from the symmetric and asymmetric stretching of CH₂ groups in lipids alone.

The PCA scores for the first components (Fig. 12) show the same trend for both the LWN and the HWN window. Between the G1/S culture and the S-phase culture, there is a slight increase in the average nucleic acid and protein content relative to the average lipid content. There is no observable difference in the average scores between the S-phase culture and the G2/M culture. However, between the G2/M culture and the early G1-phase culture, there is a decrease in the average nucleic acid and protein content relative to the average lipid content. As was the case for the PCA scores for the first components from the asynchronous study (Fig. 5), the relative positions of the individual cell scores are consistent between the LWN and HWN windows. For example, cells #63 and #75 have, respectively, the highest and lowest scores from the “G2/M +5 hrs” LWN window sample set, and the same two cells have,
respectively, the highest and lowest scores from the corresponding HWN window sample set (Fig. 12).

**Second Principal Component.** The second component for the LWN window (Fig. 13a) explains 7.7\% of the total variance, and the corresponding component for the HWN window (Fig. 13b) explains only 2.1\% of the total variance. Neither the LWN nor HWN window components have any similarity to the second components from the asynchronous cell cultures study (Fig. 7). For the LWN window all features are easily identifiable, except for the feature at 1402 cm⁻¹. The negative features include multiple contributions from the aromatic amino acids, with additional contributions from choline and O–P–O stretching in nucleic acids. The positive features are made up of contributions from nucleic acid bases and the DNA backbone, α-helix and β-sheet amide groups in proteins, and CH₂ twisting, C=C stretching, and both CH₂ and =CH deformation in lipids. For the HWN window, two broad negative features are observed, which possibly arise from the asymmetric stretching of CH₂ groups in lipids and the symmetric stretching of CH₃ groups in proteins and lipids.

The PCA scores for the LWN window (Fig. 14a) show a distinct increase in the average score for the G2/M culture. This increase is correlated with increased amounts of nucleic acid bases, DNA, conformational proteins, and CH₂ and C=C groups in lipids, and decreased amounts of aromatic amino acids, choline, and O–P–O groups in nucleic acids. The scores for the HWN window do not have any relationship to the LWN window scores and do not appear to provide much meaningful biochemical information, except that the highest scores are mostly observed in the early G1-phase culture.

**Other Principal Components.** The third PCA component for the LWN window (not shown) explains 5.1\% of the total variance. Some features in this component are similar to those in the second PCA component from the asynchronous cell cultures study (Fig. 7a), including a strong negative contribution from choline at 719 cm⁻¹ and a positive contribution from phenylalanine at 1003 cm⁻¹. However, the PCA scores for this component (also not shown) do not show any significant trend or discrimination between samples. The fourth and fifth components show features representative of slight x-axis calibration shifts, but as all the spectra in this study were collected in a single day, the system calibration was very consistent for all samples; as such, each component explains only ~3\% of the total variance. Each of the remaining components for the LWN window explain less than 2\% of the total variance and likely have little to no biological significance and account for any residual variance arising from random sources of variability. The same can be said for all the remaining components for the HWN window, each of which explains less than 1\% of the total variance.

**DISCUSSION**

**Study #1: Asynchronous Cell Cultures.** The results of this 8-day study show that when Raman spectra are acquired from single DU145 cells taken from multiple cell cultures over multiple days, with different times between sub-culturing and Raman acquisition for each culture, there are primarily two independent sources of inherent variability observed in the Raman spectra. These two sources of variability are represented in this study by the first and second PCA components (Figs. 4 and 7).

**First Principal Component.** For the entire 8-day data set in this study, the first PCA component explains 52.6\% of the total variance for the LWN window data set. When searching for a biological origin for this component, an important consideration is that no matter which subset of the total 8-day data set is input into PCA, this same component is always observed as the primary source of variability and typically explains 35 to 60\% of the total variance. For example, if the data for only the first four days is input into PCA, the variance explained is 37.3\%; however, if only the data for the last four days is used, the variance explained is 51.3\%. No matter how many days worth of data are input into PCA, or which days are chosen, the primary features of the component do not change; namely, the positive features arise from the same nucleic acid and protein molecules and the negative features arise from the same lipid...
molecules as those assigned to the first component in this study (Fig. 4a). These properties of the LWN component are also true for the HWN component (Fig. 4b), except that the percent variance explained is typically 75 to 90% of the total variance. It is also important to note that over the course of previous experiments (not shown) we have collected Raman spectra from thousands of single cells. No matter which subset of previously collected data is input into PCA, the first PCA components presented in this study (Fig. 4) are reproduced as the primary source of variability.

The most likely biological origin for the source of variability expressed by the first PCA components is the biochemical variability due to cell cycle (examined further below in the discussion of study #2). In this study, the PCA scores for the first PCA component, for both the LWN and the HWN window, show the trend of a steady decrease in the average cellular nucleic acid and protein content relative to the average lipid content from 48 to 192 hours after sub-culturing (Fig. 5). Furthermore, there is a definite correlation between the steadily increasing fraction of cells in G1 phase as measured by flow cytometry (Fig. 3) and the steady decrease in the nucleic acid and protein content of individual cells relative to the lipid content as measured by RM and calculated by PCA (Fig. 5). Interestingly, the flow cytometry results show that the fraction of cells in G1 phase begins to stop steadily increasing around 120 to 144 hours after sub-culturing, whereas the RM and PCA analyses show that the relative nucleic acid and protein content continues to decrease steadily from 120 to 192 hours (Fig. 5). This discrepancy is likely in part due to a continual increase in the fraction of G1 cells that have entered into a non-proliferating quiescent “G0” phase from 120 to 192 hours, which is not detectable by the methods used in this study. It is known that quiescent cells have a much lower RNA content than actively cycling G1 cells, as well as a decreased amount of certain proteins required for cell cycle progression.40,41 The discrepancy is also likely in part due to the flow cytometry observation that the fraction of cells in S phase continues to decrease from 120 to 192 hours and reaches a minimum of ~8% at 192 hours after sub-culturing. However, an increased fraction of quiescent cells and a decreased fraction of S-phase cells are both indicators of a less proliferative cell culture, which is an expected trend as cells are left for longer periods of time after sub-culturing.

It is important to note that the results of this study are in agreement with two previous Raman studies,20,42 both of which compared the average bulk Raman spectra of exponentially growing cells (G1 fraction ≤ 50%) to plateau-phase cells (G1 fraction ≥ 80%). One of these studies20 found that the protein/lipid, RNA/lipid, and DNA/lipid ratios were all statistically higher for exponentially growing cells, as determined by fitting biochemical component spectra to the measured LWN and HWN window averaged Raman spectra. This same study also identified the spectral regions (and the corresponding molecules assigned to those regions) that yielded significant averaged spectral differences between samples. With similar methods, the other study42 demonstrated that increased fractions of both protein and nucleic acid content in exponentially proliferating cells were correlated with decreased fractions of lipid and glycogen content, as compared to plateau-phase cells. The results presented here on RM of single cells corroborate and extend these previous Raman results for bulk samples. Our study identifies which individual molecular sub-groups are most responsible for the observed changes in Raman spectra, such as the strong contribution from CH2 deformation in lipids in the first PCA component for the LWN window (Fig. 4a). Our PCA analysis also demonstrates that changes in the relative lipid content in a cell are mathematically anti-correlated with changes in both the protein and nucleic acid content in a cell; this result is consistent both with previous Raman results42 and with the prior knowledge that the RNA-to-protein ratio is relatively constant within a cell throughout the cell cycle.40 Our results extend previous Raman studies by showing that the changes in biochemical composition due to cell cycle can be directly observed in single cell spectra (i.e., Fig. 6) and that the changes can be readily observed as a continuous process as a cell culture moves from an exponentially growing culture (24 to 96 hours after sub-culturing) to a confluent non-exponential culture (120 to 192 hours after sub-culturing). Finally, as discussed below, our study shows that there is another significant source of variability (arising from cell culture confluency) that is detectable when performing RM on single cultured cells, in addition to the variability in the nucleic acid and protein content relative to the lipid content.

Second Principal Component. For the entire 8-day data set in this study, the second PCA component explains 10.1% of the total variance for the LWN window data set. However, unlike the first PCA component, the amount of variance explained by this component is highly dependent on which subsets of the total data set are input into PCA. For example, the variance explained is maximized at 16.7% when only the data for the first five days is input into PCA. However, if the data for the first two days are excluded, the variance explained drops from 10.1% to 4.8%, and if the first three days are excluded, the variance explained drops further to 2.4%. When the first four days, or more, are excluded, the variance explained becomes less than 2% and the component is no longer recognizable. These properties of the LWN component are also true for the HWN component (Fig. 7b); in the HWN case the percent variance explained is maximized at 3.3% when only the data for the first five days is input into PCA, yet the component is not observed when the first four or more days are excluded, as was the case for the LWN component. The dependency of the second PCA component on the choice of sample subset is consistent with the corresponding PCA scores (Fig. 8), which steadily increase up to five days after sub-culturing and remain fairly constant from five to eight days after sub-culturing.

A definitive biological origin for the second PCA component is unclear, especially for the HWN window where the molecular origin of the features is unknown (Fig. 7b). However, there is a strong correlation between the trend of the PCA scores (Fig. 8) and the measured confluency of the cell cultures (Fig. 3), which is in turn related to the amount of time the culture was left to incubate after sub-culturing. In this study, the cell cultures steadily increase their confluency until approximately five days after sub-culturing, after which there is very little room left to grow and the confluency remains relatively constant at ~90%. The confluency trend matches the trend of the PCA scores, which steadily increase up to five days after sub-culturing and remain fairly constant afterwards. Furthermore, as discussed above, if only the data from days 5 to 8 is input into PCA (i.e., only the data collected once the culture had reached ~90% confluency), then the second PCA
component is not observed at all. This implies that the component is directly caused by biochemical changes related to a sub-confluent culture growing during the first three to four days after sub-culturing. The primary biomolecules responsible for this source of variability in the Raman spectra can be identified in the second PCA component for the LWN window (Fig. 7a). To the author’s knowledge this study is the first spectroscopic measurement of inherent biochemical variability in a cell culture that is correlated with the changing confluency of a cell culture during the first three to four days after sub-culturing, which is shown here to be independent of the existing variability arising from cell cycle progression. We have observed this source of variability in several previous experiments with DU145 cells (not shown) in which cell cultures were harvested for Raman analysis one to two days after sub-culturing. It should be noted that although this source of variability is shown here to be significant for DU145 cells, it may not be a characteristic of all in vitro cultured cell lines.

Study #2: Synchronized Cell Cultures. In theory, the effects of the cell cycle regulatory drugs thymidine and nocodazole are reversible, such that when the drug is removed and replaced by fresh media the cells in the culture will progress synchronously through the rest of their mitotic cycle. In practice, whole culture synchronization is unfeasible and cell cultures become desynchronized very quickly. Furthermore, a certain fraction of the culture will not be immediately released (or released at all) from the drug-induced arrest. However, large fractions of cells (i.e., \(~75\)%) can indeed be synchronized for short periods of time (typically less than 24 hours), and the cell cycle distribution can indeed be significantly altered from that of an untreated asynchronous population. Despite the known shortcomings of using drugs for cell cycle control, drug treatment is still the easiest and simplest technique for significantly altering the cell cycle distribution and achieving a high yield of mostly synchronized cells. It is important to note that the level of synchronization reported in this study (Fig. 10) is comparable to the level of synchronization reported by both the recent study using RM for cell cycle discrimination, where synchronization was performed by serum starvation and cell cycle regulatory drugs, and another previous study that investigated the infrared spectroscopic differences between cells in different stages of the cell cycle, where synchronization was performed by centrifugal elutriation.

A main goal of this second study is to directly examine whether the variability expressed by the first PCA components (addressed above in the discussion for study #1) is indeed due to biochemical differences between cells at different points in the cell cycle. The first PCA components for this study have primarily the same features as the first PCA components for the asynchronous cell cultures study, and both the LWN and HWN window components explain approximately the same amount of the total variance as the corresponding components in the asynchronous cell cultures study. However, in this study all four cultures were harvested on the same day (after each were synchronized) and the confluency of each culture was \(<60\%\); as such, all four cultures should have very few quiescent cells, and we would not expect to measure any variability due to the differences in confluency between the cultures.

The variability in the nucleic acid and protein content relative to the lipid content in single cells, between the four synchronized cell cultures, is expressed by the PCA scores for the first components (Fig. 12). The slight shift to a higher average relative nucleic acid and protein content between the “G1/S” and the “G1/S +3 hrs” cultures correlates with the flow cytometry measured shift from \(~83\%\) of the first culture existing at the G1/S boundary to \(~64\%\) of the second culture progressing through S phase (Fig. 10). This shift is consistent with expected changes in the biochemical content for S-phase cells, which contain increased levels of RNA and protein as compared to G1 cells, and an increased amount of DNA due to the active DNA replication that occurs during S phase. There is no observable shift in the scores between the “G1/S +3 hrs” and the “G2/M” cultures (Fig. 12), even though the “G2/M” culture has over \(74\%\) of its cells at the G2/M boundary compared to only \(~16\%\) of the cells in G2 phase for the “G1/S +3 hrs” culture. This lack of separation in the scores for the first components may seem at odds with a known increase in the overall RNA and protein content of G2/M cells as compared to late S-phase cells; however, the PCA scores for the first components only represent changes in nucleic acid and protein content relative to the total lipid content, which is also increasing throughout G2 phase in preparation for mitotic division.

The most significant change in the scores for the first PCA components occurs as a decrease in the average nucleic acid and protein content between the “G2/M” and the “G2/M +5 hrs” cultures, which undergo a transition from a culture with \(>74\%\) of its cells in a G2/M phase to a culture with \(~75\%\) of its cells existing within the first five hours of G1 phase. The observed change in the relative nucleic acid and protein content is consistent with previous biochemical experiments that have shown that the lowest levels of RNA and protein are found within the first few hours of G1 phase, immediately following cell division. Our observations are also consistent with the results of the recent RM study for cell cycle discrimination, in which the successful discrimination between S or G2/M cells and G0/G1 cells was due to increased nucleic acid and protein content relative to lipid content in both S and G2/M cells, as measured in the LWN spectral window. This previous study also reported poor discrimination between S and G2/M cells based on nucleic acid and protein content relative to lipid content, which we also observe here for both spectral windows (Fig. 12).

In our study, it is interesting to note that the PCA scores for the LWN window (Fig. 12a) for the “G2/M +5 hrs” culture are well split into two subgroups: \(60\%\) of the cells have PCA scores \(<-1\) (low relative amount of nucleic acid and protein), and \(40\%\) of the cells have PCA scores \(>0\) (high relative amount of nucleic acid and protein). This split is matched by the cell cycle distribution for this culture (Fig. 10), which is distinctly separated into two groups: \(~75\%\) of the cells in early G1 phase and \(~21\%\) of the cells in G2 phase, with only \(~4\%\) of the cells in S phase. The relative positions of the scores for this culture are similar for the HWN window, but the separation between the two subgroups is less distinct (Fig. 12b). In summary, these results confirm that the most significant source of Raman spectral variability between cells in a culture, which is expressed in this work by the first PCA components, can be confidently attributed to biochemical changes arising from the progression of individual cells through their mitotic cycle.

The features in the second PCA components for this study (Fig. 13) are different from the features in the second PCA
components for the asynchronous cell cultures study (Fig. 7). The scores for the LWN and HWN windows do not show similar trends, which suggests they each arise from different sources of biochemical variability between cultures. The scores for the HWN component (Fig. 14b) do not show any clear trend or separation between cultures, and the features in the HWN component (Fig. 13b) are not clearly attributed to a unique set or class of biomolecules. As such, it is difficult to assign a biological meaning to the HWN window results. However, the scores for the LWN component (Fig. 14a) distinctly separate the “G2/M” cells from the other cultures with an increase in the average PCA score. According to the corresponding PCA component (Fig. 13), the increase in scores for the G2/M cells corresponds primarily with a decrease in aromatic amino acids, choline, and O–P–O groups in RNA and/or lipids, correlating with an increase in nucleic acid bases, DNA, α-helix and β-sheet amide groups, and CH2, =CH, and C=C lipid groups. The biological reason for these changes is unclear, but may be related to changes in the cell biochemistry in preparation for mitotic division. Alternatively, the variability could arise as a temporary cellular response to the nocodazole treatment for synchronization of the G2/M culture. If these measured changes are indeed due to a natural source of variability inherent to G2/M cells, this component was likely not observed in the asynchronous cell culture study due to the low fraction of G2 cells in each of the asynchronous cultures and the presence of other larger sources of variability such as the changing confluency and shifts in the x-axis calibration. Further work using different synchronization techniques would be required to determine whether this source of variability is indeed an inherent characteristic of G2/M cells.

Spectral Variability and Principal Component Analysis. Many of the results in this work depend on the accurate assignment of a molecular origin to features in the PCA components. However, achieving confidence in the validity of such assignments is only possible if all external sources of spectral variability that are not inherent to the biochemical composition of the cells have been removed prior to PCA implementation. Sources of variability that arise include variability in the intensity and shape of the fluorescent baseline (originating from a sample substrate or from the cellular material itself) or variability induced in a data set by an improper spectral normalization technique. If an external source of variability happens to contribute preferentially to a certain sample (or samples) in the data set, the PCA algorithm will faithfully correlate the variability from the external source with any variability that is inherent to the sample in question, which is possibly the variability of interest in the experiment.

An example of this issue, which arose during the course of this work, is the variability in intensity of spectral contributions from the quartz substrate. For the “G2/M +5 hrs” culture from the synchronized cell culture study, the pellet of cells used for Raman acquisition was only a few cell layers thick, as opposed to tens of cell layers thick for the other cultures in the study. As such, there was a slightly greater quartz contribution observed in all the spectra collected from the “G2/M +5 hrs” cells. If a very conformal baseline was not applied for baseline correction of the lower half of the LWN window (~600–1200 cm⁻¹), it was found that the variability due to quartz became significant to the degree that the quartz variability became correlated with other sources of variability inherent to the “G2/M +5 hrs” culture. As the conformity of the baseline was reduced, the resultant increase in quartz variability became observable in the first PCA component as recognizable quartz features, and many biological features from the first PCA component began to appear in the component that was originally dominated by quartz features alone.

These considerations require extreme care when developing and implementing automated spectral processing methods, such as spectral smoothing or baseline correction algorithms, when large multi-sample data sets are prepared for PCA analysis. However, the PCA components themselves can aid in the identification of external sources of variability during the development and implementation of spectral processing techniques, as long as the spectral features of the external sources are known. It should be noted that performing Raman analysis with the HWN window is significantly simpler, since baseline removal is easier due to the absence of substrate contributions and fluorescence in this spectral region. However, the simplicity advantage comes at the cost of a significant decrease in the amount of biochemical information available, as compared to the LWN window. Furthermore, the strong spectral contributions from water in the HWN window may become significant if the methods presented here are applied to the Raman analysis of cells in an aqueous environment.

Low-Wavenumber versus High-Wavenumber Spectral Windows. Both the LWN and HWN windows were analyzed independently throughout this work to determine whether information can be obtained equivalently from either window. We have found that biochemical variability due to cell cycle is clearly observable in either window, and the spectral differences are directly observable in the original data for both windows (Figs. 6a and 6b). However, the LWN window provides information from many more biomolecules, including multiple contributions from nucleic acids, which are not observed as sources of cell cycle variability in the HWN window. The variability due to changes in cell culture confluency after sub-culturing is more apparent in the LWN window, due to the strong contributions from the features identified in the second PCA component from the asynchronous cells study (Fig. 4a). Although the trends of the PCA scores for the second components are similar for both the LWN and HWN windows (Fig. 5), the molecular origins of the corresponding features in the HWN window component are uncertain (Fig. 4b). Furthermore, the spectral differences arising from this source of variability are directly observable in the original data only for the LWN window (Fig. 9). Therefore, in this case the LWN window provides spectroscopic information that is not available in the HWN window. We have also shown that the LWN window is sensitive to biochemical changes unique to the G2/M sample from the synchronized cell cultures study, whereas in the HWN window no spectroscopic differences were observed for the same sample.

Spectral Variability and Cell Size. All of our single-cell RM measurements are acquired with a fixed sampling volume (~2 × 5 × 10 μm, in x-y-z) that is aligned with the center of the selected cell (see Fig. 1). Therefore, there is the possibility of observing spectral differences that correlate simply with size differences in the cell population. For example, previous authors have noted that a smaller cell will have a higher surface area to volume ratio than a larger cell and may therefore yield more biochemical signals from cell membrane lipids and proteins relative to cytoplasmic and nuclear biomolecules.
However, these authors used direct measurements of the size of each selected cell, obtained in suspension during optical tweezers RM acquisition, to show that cell size had no correlation with the ability of RM and PCA to biochemically discriminate between two cell lines of different average size.\textsuperscript{16}

In our work, monitoring absolute cell size via direct optical measurements of the cells selected would be inaccurate due to the lack of three-dimensional cell geometry information in the pellet. However, the relative cell size distribution for a given culture is well described by the flow cytometry measurement of forward scatter intensity, which is acquired from cells in suspension simultaneously with PI fluorescent intensity acquired for cell cycle analysis.

To verify that the dominant sources of inherent spectral variability observed in this work (as described by the first and second PCA components in study \#1) are not simply due to changing cell size, we have analyzed the forward scatter intensity distributions of all eight cell cultures used in study \#1. From 24 to 72 hours after sub-culturing, there is no detectable change in the cell size distribution between cultures. From 72 to 96 hours, there is a detectable shift in the measured distribution towards lower forward scatter intensities, which is indicative of a higher proportion of smaller cells in the culture.\textsuperscript{1}

From 96 to 192 hours after sub-culturing, there is no further detectable change in the cell size distribution between cultures. However, our PCA analysis shows that the largest source of spectral variability observed in this study (first PCA component) displays a steady trend of continuing spectral differences occurring from 48 to 192 hours after sub-culturing (Fig. 5).

Furthermore, the second largest source of spectral variability (second PCA component) displays a trend of continuing spectral differences occurring from 24 to 120 hours after sub-culturing (Fig. 8). Neither of these sources of variability correlate with the observed changes in the relative cell size distribution. Therefore, any spectral variability arising from differences in cell size must be explained by one of the many lower variance PCA components, each of which explains less than 3\% of the total variance for the LWN window and less than 1\% of the total variance for the HWN window. Differences in cell size may introduce significant spectral variability when comparing cell lines with large differences in average size, but within a single cell line our results show that cell size is not a significant source of spectral variability.

\section*{CONCLUSION}

We have shown that the inherent variability in Raman spectra of single human tumor cells cultured \textit{in vitro} is correlated with biochemical changes arising from (1) cell cycle progression and (2) the confluency of a cell culture during the first three to four days after sub-culturing.

The variability between single-cell Raman spectra arising from cell cycle progression is expressed as varying intensities of protein and nucleic acid features relative to lipid features. Raman spectra acquired from synchronized cell cultures show a continual increase in the average nucleic acid and protein content relative to lipid content as cells progress from early G1 phase to the G1/S boundary and into S phase. The molecular origins of the Raman features affected by cell cycle progression have been identified for both the LWN and the HWN spectral windows, by the features of the first PCA components (Figs. 4 and 11). Our PCA analysis has shown that in the absence of additional external sources of variability, cell cycle variability typically accounts for 40–60\% of the total variance if the LWN window is used and 75–90\% if the HWN window is used. Because there will always be some level of variability in biochemical composition between cells due to the cell cycle, the characterization of cell cycle variability presented in this work may be useful for future Raman studies in order to distinguish the inherent cell cycle variability between cells from other independent sources of variability.

The molecular origins of the Raman features that produce variability correlated with the changing confluency of a cell culture have been identified for the LWN spectral window by the features of the second PCA component from the study of asynchronous cell cultures (Fig. 7a). In our work with DU145 cells, this source of variability can explain up to 17\% of the total variance if the LWN window is used. The characterization of this variability, as presented in this work, may be an important consideration for future Raman studies involving comparisons between cell cultures harvested at different time intervals after sub-culturing. For example, if cell cultures are allowed to incubate after sub-culturing for three to four days before Raman analysis, the variance explained by this source of variability is greatly reduced and may facilitate the observation of other more subtle spectral differences between cell cultures.

\section*{ACKNOWLEDGMENTS}

The authors gratefully acknowledge funding from the National Science and Engineering Research Council, the Canadian Foundation for Innovation, and the Western Economic Diversification program. We would also like to thank the staff of the Decleey Research Centre at the BC Cancer Agency’s Vancouver Island Centre for providing the initial DU145 cell stocks and technical assistance with cell culture and flow cytometry.

\begin{thebibliography}{100}
\bibitem{13}N. Stone, C. Kendall, N. Shepherd, P. Crow, and H. Barr, J. Raman Spectrosc. 33, 564 (2002).
\bibitem{14}C. Krishna, G. Stockelign, G. Kegelaer, S. Rubin, V. Kartha, and M. Manfait, Vib. Spectrosc. 38, 95 (2005).
\end{thebibliography}

\footnotetext[1]{This shift is consistent with our cell cycle analysis (Fig. 3) where between 72 and 96 hours we observe a sharp increase in the fraction of G1 phase cells, which are typically smaller than S-phase and G2-phase cells.}