

Comparison of principal component analysis and biochemical component analysis in Raman spectroscopy for the discrimination of apoptosis and necrosis in K562 leukemia cells

Yi Hong Ong, Mayasari Lim, and Quan Liu*

School of Chemical and Biomedical Engineering, Nanyang Technological University, 70 Nanyang Drive, 637457 Singapore

**quanliu@ntu.edu.sg*

Abstract: Raman spectroscopy has been explored as a promising label-free technique in discriminating apoptosis and necrosis induced cell death in leukemia cells. In addition to Principal component analysis (PCA) as commonly employed in Raman data analysis, another less commonly used but powerful method is Biochemical Component Analysis (BCA). In BCA, a Raman spectrum is decomposed into the contributions from several known basic biochemical components, such as proteins, lipid, nucleic acids and glycogen groups etc. The differences in terms of classification accuracy and interpretability of resulting data between these two methods in Raman spectroscopy have not been systematically investigated to our knowledge. In this study, we utilized both methods to analyze the Raman spectra measured from live cells, apoptotic and necrotic leukemia cells. The comparison indicates that two methods yield comparable accuracy in sample classification when the numbers of basic components are equal. The changes in the contributions of biochemical components in BCA can be interpreted by cell biology principles in apoptosis and necrosis. In contrast, the contributions of most principle components in PCA are difficult to interpret except the first one. The capability of BCA to unveil fine biochemical changes in cell spectra and excellent accuracy in classification can impel the broad application of Raman spectroscopy in biological research.

©2012 Optical Society of America

OCIS codes: (300.6450) Spectroscopy, Raman; (180.5655) Raman microscopy; (070.4790) Spectrum analysis.

References and links

1. P. Chen, A. G. Shen, X. D. Zhou, and J. M. Hu, "Bio-Raman spectroscopy: a potential clinical analytical method assisting in disease diagnosis," *Anal. Methods* **3**(6), 1257–1269 (2011).
2. C. L. Zavaleta, B. R. Smith, I. Walton, W. Doering, G. Davis, B. Shojaei, M. J. Natan, and S. S. Gambhir, "Multiplexed imaging of surface enhanced Raman scattering nanotags in living mice using noninvasive Raman spectroscopy," *Proc. Natl. Acad. Sci. U.S.A.* **106**(32), 13511–13516 (2009).
3. N. Kunapareddy, J. P. Freyer, and J. R. Mourant, "Raman spectroscopic characterization of necrotic cell death," *J. Biomed. Opt.* **13**(5), 054002 (2008).
4. B. R. Lutz, C. E. Dentinger, L. N. Nguyen, L. Sun, J. W. Zhang, A. N. Allen, S. Chan, and B. S. Knudsen, "Spectral Analysis of Multiplex Raman Probe Signatures," *ACS Nano* **2**(11), 2306–2314 (2008).
5. J. R. Mourant, J. Dominguez, S. Carpenter, K. W. Short, T. M. Powers, R. Michalczyk, N. Kunapareddy, A. Guerra, and J. P. Freyer, "Comparison of vibrational spectroscopy to biochemical and flow cytometry methods for analysis of the basic biochemical composition of mammalian cells," *J. Biomed. Opt.* **11**(6), 064024 (2006).
6. J. R. Mourant, K. W. Short, S. Carpenter, N. Kunapareddy, L. Coburn, T. M. Powers, and J. P. Freyer, "Biochemical differences in tumorigenic and nontumorigenic cells measured by Raman and infrared spectroscopy," *J. Biomed. Opt.* **10**(3), 031106 (2005).

7. A. G. Ryder, "Classification of narcotics in solid mixtures using Principal Component Analysis and Raman spectroscopy," *J. Forensic Sci.* **47**(2), 275–284 (2002).
8. B. G. M. Vandeginste, D. L. Massart, L. M. C. Buydens, S. De Jong, P. J. Lewi, and J. Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics* (Elsevier Science, 1998).
9. H. Shinzawa, K. Awa, W. Kanematsu, and Y. Ozaki, "Multivariate data analysis for Raman spectroscopic imaging," *J. Raman Spectrosc.* **40**(12), 1720–1725 (2009).
10. K. E. Shafer-Peltier, A. S. Haka, M. Fitzmaurice, J. Crowe, J. Myles, R. R. Dasari, and M. S. Feld, "Raman microspectroscopic model of human breast tissue: implications for breast cancer diagnosis in vivo," *J. Raman Spectrosc.* **33**(7), 552–563 (2002).
11. D. Van de Sompel, E. Garai, C. Zavaleta, and S. S. Gambhir, "A Hybrid Least Squares and Principal Component Analysis Algorithm for Raman Spectroscopy," *PLoS ONE* **7**(6), e38850 (2012).
12. I. A. Boere, T. C. B. Schut, J. van den Boogert, R. W. F. de Bruin, and G. J. Puppels, "Use of fibre optic probes for detection of Barrett's epithelium in the rat oesophagus by Raman spectroscopy," *Vib. Spectrosc.* **32**(1), 47–55 (2003).
13. A. Molckovsky, L. M. Song, M. G. Shim, N. E. Marcon, and B. C. Wilson, "Diagnostic potential of near-infrared Raman spectroscopy in the colon: differentiating adenomatous from hyperplastic polyps," *Gastrointest. Endosc.* **57**(3), 396–402 (2003).
14. J. W. Chan, D. S. Taylor, S. M. Lane, T. Zwerdling, J. Tuscano, and T. Huser, "Nondestructive identification of individual leukemia cells by laser trapping Raman spectroscopy," *Anal. Chem.* **80**(6), 2180–2187 (2008).
15. H. Yao, Z. Tao, M. Ai, L. Peng, G. Wang, B. He, and Y.-Li, "Raman spectroscopic analysis of apoptosis of single human gastric cancer cells," *Vib. Spectrosc.* **50**(2), 193–197 (2009).
16. J. L. Pichardo-Molina, C. Frausto-Reyes, O. Barbosa-García, R. Huerta-Franco, J. L. González-Trujillo, C. A. Ramírez-Alvarado, G. Gutiérrez-Juárez, and C. Medina-Gutiérrez, "Raman spectroscopy and multivariate analysis of serum samples from breast cancer patients," *Lasers Med. Sci.* **22**(4), 229–236 (2007).
17. M. Moreno, L. Raniero, E. A. L. Arisawa, A. M. D. Santo, E. A. P. dos Santos, R. A. Bitar, and A. A. Martin, "Raman spectroscopy study of breast disease," *Theor. Chem. Acc.* **125**(3–6), 329–334 (2010).
18. G. Pyrgiotakis, O. E. Kundakcioglu, K. Finton, P. M. Pardalos, K. Powers, and B. M. Moudgil, "Cell death discrimination with Raman spectroscopy and support vector machines," *Ann. Biomed. Eng.* **37**(7), 1464–1473 (2009).
19. I. Nottingher, S. Verrier, H. Romanska, A. E. Bishop, J. M. Polak, and L. L. Hench, "In situ characterisation of living cells by Raman spectroscopy," *Spectrosc. Int. J.* **16**(2), 43–51 (2002).
20. J. S. U. Hjorth, "Cross validation," in *Computer Intensive Statistical Methods: Validation, Model Selection, and Bootstrap* (Chapman and Hall/CRC, 1993), 27–28.
21. N. Stone, C. Kendall, J. Smith, P. Crow, and H. Barr, "Raman spectroscopy for identification of epithelial cancers," *Faraday Discuss.* **126**, 141–157, discussion 169–183 (2004).
22. A. Zoladek, F. C. Pascut, P. Patel, and I. Nottingher, "Non-invasive time-course imaging of apoptotic cells by confocal Raman micro-spectroscopy," *J. Raman Spectrosc.* **42**(3), 251–258 (2011).
23. S. Malladi, M. Challa-Malladi, and S. B. Bratton, "Apoptosis," in *Comprehensive Toxicology (Second Edition)*, A. M. Editor-in-Chief: Charlene, ed. (Elsevier, 2010), 543–578.
24. T. M. Sauerwald, A. Lewis, H. Dorai, and M. J. Betenbaugh, "Apoptosis: The Signaling Pathways and Their Control," in *Comprehensive Biotechnology (Second Edition)*, M.-Y. Editor-in-Chief: Murray, ed. (Academic Press, 2011), 483–494.
25. A. Vaculova and B. Zhivotovsky, "Caspases: Determination of their activities in apoptotic cells," in *Programmed Cell Death, General Principles for Studying Cell Death, Pt A*, R. KhosraviFar, Z. Zakeri, R. A. Lockshin, and M. Piacentini, eds. (Elsevier Academic Press Inc, 2008), 157–181.
26. K. S. Saini and N. I. Walker, "Biochemical and molecular mechanisms regulating apoptosis," *Mol. Cell. Biochem.* **178**(1/2), 9–25 (1998).
27. M. Leist and P. Nicotera, "Cell Death: Apoptosis versus Necrosis," in *Primer on Cerebrovascular Diseases* (Academic Press, 1997), 101–104.
28. S. Elmore, "Apoptosis: a review of programmed cell death," *Toxicol. Pathol.* **35**(4), 495–516 (2007).
29. N. Kunapareddy, S. Carpenter, J. P. Freyer, and J. R. Mourant, "Biochemical characterization of cell-death via Raman spectroscopy - art. no. 60930V," in *Biomedical Vibrational Spectroscopy III: Advances in Research and Industry*, A. MahadevanJansen, and W. H. Petrich, eds. (SPIE, 2006).
30. S. Verrier, I. Nottingher, J. M. Polak, and L. L. Hench, "In situ monitoring of cell death using Raman microspectroscopy," *Biopolymers* **74**(1–2), 157–162 (2004).

1. Introduction

Raman spectroscopy is a laser-based spectroscopic technique that is capable of measuring the inelastic scattering of photons induced by the intrinsic molecular bonds present in a sample. It has the ability to obtain rich biochemical information from single cells without the need of staining or labeling thus this technique has been explored in a variety of cell studies including the investigation of cell death. Cell death induced by either apoptosis or necrosis has characteristic signatures that distinguish one from the other. The unique biochemical

fingerprint of cell death types in leukemia cells using Raman spectra can help us readily identify the pathway of drug-induced cell death. A Raman spectrum contains rich biochemical information that can provide a molecular structural signature to a substance [1]. However, biological macromolecules such as proteins, nucleic acids, lipids and polysaccharides are complex and relatively large, generating numerous peaks which are overlapping in a broad band. Qualitative studies of cell spectra by assigning one or a few Raman peaks to specific basic biochemical components often introduce inaccuracies to the analysis. Visual inspection of specific peaks usually involves guessing, to determine which biochemical components that the intensity changes in the peaks correspond to, because several components may contribute to the same peaks at the same excitation wavelength. Changes in separate Raman bands thus provide little quantitative understanding of how the relative amounts of basic biochemical components change across each cell type. Furthermore, selective studies on specific Raman bands often result in loss of important spectra information carried in the unselected regions.

Due to the drawbacks of classical visual inspection method, chemometrics methods emerged to be a better approach for implementing spectral diagnosis by utilizing the entire Raman spectral. By using statistical or mathematical techniques, the essence of information present in the spectral data can be highlighted and represented in a matrix with reduced dimension for easier quantitative interpretation. Commonly used methods include Biochemical Component Analysis (BCA) [2–6], and principal component analysis (PCA) [1, 7–9]. BCA method uses least square regression, by assuming the sample's spectrum is the linear summation of all basic components' spectra, to estimate the contribution of each component. This method is named differently in the literature, which includes spectral deconvolution, basis spectral analysis method [3], direct classical least squares method (DCLS) [2], and spectral fitting method [5, 6]. For the purpose of convenience, this method will be regarded as Biochemical Component Analysis (BCA) in this paper. Kunapareddy et al. has demonstrated this method by fitting the basis spectra of protein, lipid, RNA, DNA and glycogen to full cell spectra to estimate the biochemical changes in necrotic human malignant melanoma cell (MEL-28). They reported a decrease in the relative amount of lipid and RNA in necrotic cells, and observed an increase in the relative amount of protein [3]. This technique has also been reported earlier at the tissue level in a study of breast cancer diagnosis [10]. The quantification of biochemical changes can provide useful information of the structural and pathological states of cells and help to classify different cell and disease types. While this is a fast and quantitative approach, BCA requires prior knowledge of the pure constituents of the sample to supply an explicit background model to the algorithm. Moreover, spectral fitting algorithm in BCA is also sensitive to changes in the background spectrum as it cannot adapt to peak shifts and alterations in the relative intensities of peaks [11].

The principal component analysis (PCA) method is a non-parametric method that does not require an explicit background model. PCA performed mathematical decomposition of the spectral data that reduce the data dimensions of a highly complex chemical system to a smaller number of scores and principal components (PCs) or loadings that effectively carries all the important information of the spectra [8]. Classification of spectral data can be easily done by choosing different combinations of PCs to build a new coordinate system. PCA is widely used in Raman spectroscopy studies for pathological classification, such as to discriminate between Barrett's and normal epithelium [12], to differentiate adenomatous from hyperplastic polyps of the colon [13] and also to classify T and B lymphocytes of normal and leukemic patients [14]. In a cell death study, Yao et al. has demonstrated the use of PCA method to distinguish between live and apoptotic human gastric cancer cells [15]. However, while useful as a classification method, PCA does not reveal physically or chemically interpretable information of the sample. It provides only abstract information representing

entire features broadly distributed in the data and is unable to explicitly quantify the biochemical changes in a biological sample [1, 9].

Although both BCA and PCA have been widely used in Raman spectroscopic studies, there have been no studies that systematically compared the classification accuracy based on the features extracted from both methods in distinguishing between two distinctive cell death modes, i.e. apoptosis and necrosis. In this work, we will first compare the efficacies of classical visual inspection and BCA methods in extracting features from Raman spectra of live, apoptotic and necrotic human chronic myelogenous leukemia cells (K562 cell line). In BCA study, the Raman spectra of cells will be decomposed to estimate the contribution from basic cellular constituents, such as proteins, nucleic acids, lipids and polysaccharides, to quantify the biochemical differences among live, apoptotic and necrotic leukemic cells. These results will be validated against the literature. Then, PCA will be performed on the spectral data to yield principal components and their scores. Finally, a linear discriminant analysis (LDA) method [16, 17] will be employed to distinguish cell death modes based on the features extracted from both BCA and PCA to compare their classification accuracies. Furthermore, the principal components from PCA will be decomposed by using BCA method in an attempt to interpret the biochemical information behind them.

2. Materials and methods

2.1 Sample preparation and drug treatment

Human chronic myelogenous leukemia cells (K562 cell line) were purchased from American Type Culture Collection (Manassas, VA, US). K562 cells were cultured in Iscove's Modified Dulbecco's Medium supplemented with 10% fetal bovine serum, and were incubated in the incubator at 37°C and with 5% CO₂. Cultures were maintained by the addition or replacement of fresh medium every 2-3 days to maintain the cell density between 10⁵ and 10⁶ cells per ml. Cells were then transferred to a 6-well culture plate at cell density of 10⁵ per ml and were incubated at 37°C and with 5% CO₂ for 24 hours. Cytosine arabinoside (Sigma Aldrich, Singapore) was then added to two wells to reach a final concentration of 300 µM to induce apoptosis. Triton X-100 was added to another two wells at a concentration of 100 µM to induce necrosis. This concentration will induce damage to cell membrane without completely lysing it in the first 24 hours upon induction [18]. The other two wells were used as the control group without drug treatment. All groups were incubated for another 72 hours at 37°C and with 5% CO₂. Cells from the control group and the treated groups, were washed twice, rinsed and immersed in phosphate-buffered saline (PBS). A small cell sample from each treated group was taken to validate the occurrence of apoptosis and necrosis while the remaining cells were left unstained for Raman measurements. The apoptosis of the cells treated with cytosine arabinoside were tested and validated by applying a dye, Hoechst 33258. Necrotic cells treated with Triton X-100 were validated by using Trypan blue dye, which is a membrane permeation assay.

2.2 Evaluation of apoptosis and necrosis

Apoptotic cell death was evaluated by Hoechst 33258, a blue fluorescent dye that stains nucleic acids. A cell sample taken from the well plate where Cytosine arabinoside was applied was first centrifuged and resuspended in PBS solution to wash away culture medium. Hoechst dye was added to the sample to reach a final concentration of 2 µg/ml. The mixture was incubated for 15 minutes at 37°C before being washed twice and immersed in PBS solution. The sample was observed under a fluorescence microscope with 350-nm excitation light. Apoptotic cells were confirmed by nuclear condensation and fragmentation while the nucleolus and chromatin of control cells remained intact.

To evaluate cell death due to necrosis, 10 µl of 0.4% (w/v) Trypan blue solution was added to 10µl of the cell sample. The mixture was left for 5 minutes in room temperature

before the observation under optical microscope. Necrotic cells were stained blue while live and apoptotic cells remained unstained as their plasma membrane remained intact.

2.3 Raman spectroscopy

Raman spectra were measured using a micro-Raman spectrometer system (inVia, Renishaw, UK) coupled to a microscope (Alpha 300, WITec, Germany) in a backscattering geometry. A Czerny-Turner type spectrograph ($f = 250$ mm) equipped with a holographic grating (1800 groove/mm) and a RemCam CCD detector (inVia, Renishaw, UK), which yields a spectral resolution of 2 cm^{-1} , were selected for this study. A 785-nm diode laser (about 50 mW on the sample) was used for excitation and the illumination time was 60 seconds. It was reported in another study [19] that no visible effect on cell viability was observed when cells were illuminated by a laser at the same wavelength with the identical power for 120 seconds. Similarly, no visible effect on cell viability was observed in this study either during or after each measurement. The laser spot of around $1.6\text{ }\mu\text{m}$ in diameter was formed on individual cells by a microscope objective lens (50x, NA = 0.6, Leica). The cell sample was prepared on an aluminum substrate to achieve minimal Raman and fluorescence background and covered by a thin quartz cover slip as shown in Fig. 1 to reduce the distortion on focusing due to evaporation. A small well was created using a cellophane tape underneath aluminum foil that wrapped around a glass slide to limit the movement of cells during focusing adjustment, while keeping the sample in suspension. This well design also helped reduce the pressure on the cell exerted by the cover slip thus keeping cell morphology unchanged.

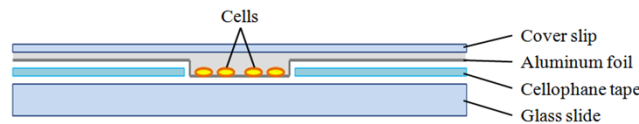


Fig. 1. Cross sectional view of the sample prepared in a small well created on an aluminum substrate.

2.4 Data analysis

Twenty cell spectra from each group were measured over a range of 600 cm^{-1} to 1800 cm^{-1} . Ten spectra were measured from the central region of the cells (nucleus) and another ten spectra were measured from the peripheral region of cells (cytoplasm and membrane) to obtain average spectra representing each cell type. The background spectrum was measured from PBS on the aluminum substrate as in Fig. 1 and subsequently subtracted from cell spectra. Data processing was performed using MATLAB (Version 7.6, MathWorks, Natick, MA, US). Firstly, the narrow spikes caused by cosmic rays were removed. Then, the broad and slowly varying fluorescence background was estimated by using the fifth order polynomial fitting and subtracted. Afterwards, each spectrum was smoothed using a Savitzky-Golay smoothing algorithm before the subsequent analysis was performed.

PCA was first performed on measured Raman spectra, using princomp function in MATLAB. PCA is a statistical analysis method which can reduce the dimension of the data while accounting for most of the variance in the original data. Kruskal-Wallis one-way analysis of variance was performed on the scores of the first ten principal components to determine which PC has significant differences in the mean of scores among three groups of cells. Two-dimensional plots were constructed with different combination of scores for the first three principal components and a three-dimensional plot was also constructed with the three sets of scores.

For BCA, the Raman spectra of pure basic biochemical components, including actin, albumin, triolein, phosphatidylcholine, DNA, RNA, and glycogen, were measured with the same configuration as in cell measurements. These spectra were normalized by dividing the intensity at each wavenumber by the maximum intensity of each spectrum and the normalized

spectra were used in BCA. All components were purchased from Sigma Aldrich, Singapore and used without further purification. These components were chosen to represent the major biochemical groups in cellular constituents, which include proteins, lipids, polysaccharides, and nucleic acids. The intensities of the component spectra were assumed to be linearly proportional to their concentrations and the cell spectra were the linear combination of these basic component spectra at various concentrations. To find out the contribution of each component to the cell spectra, a least square regression method was employed by using a nonlinear curve fitting function, *lsqcurvefit*, in MATLAB with the option of the trust-region-reflective algorithm. This function finds the coefficients of all basic component spectra that best fit the cell spectra. These coefficients reflect the relative amount of biochemical components in cells. Wilcoxon signed-rank test was performed to evaluate the statistical significance of difference in the amount of biochemical components between different groups of cells.

Then, we compared the performance of cell death classification using the features extracted from BCA and the PCA. Linear discriminant analysis (LDA) classification models were built using the fitting coefficients of biochemical components from BCA and principal components from PCA. A leave-one-out cross validation method [20] was employed to train the classifier. The performance of these models was compared in term of accuracy in classifying different cell death modes. Furthermore, the first ten principal components from PCA were decomposed using BCA in an attempt to interpret those principal components showing significant differences among different groups.

3. Results

Figure 2 shows the Raman spectra of seven selected basic biochemical components. Note that the spectra of each individual biochemical component were vertically segregated for clarity. These basic components are used to represent the four major organic bio-macromolecules in cells, i.e. proteins, lipids, nucleic acids, and polysaccharides. Among the seven components, actin and albumin represent proteins, triolein and phosphatidylcholine represent lipids, DNA and RNA are nucleic acids and glycogen represents polysaccharides. It is noted that the Raman spectra of actin, triolein and DNA are quite similar to those of albumin, phosphatidylcholine and RNA, respectively. The selection of these components will be discussed in detail in the Discussion section.

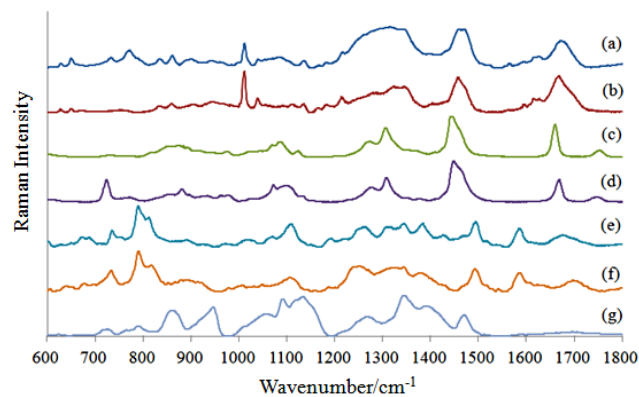


Fig. 2. Basic biochemical components' spectra used in the fitting of Raman spectra measured from K562 cells. (a) Actin; (b) Albumin; (c) Triolein; (d) Phosphatidylcholine; (e) DNA; (f) RNA; (g) Glycogen

The Raman spectra of live, apoptotic, and necrotic K562 cells were shown in Fig. 3. The spectrum for each group was obtained by averaging over 20 different cells and all spectra were offset accordingly on the y dimension for the clarity purpose. The statistical power for

the differences between cell groups is larger than 0.9 for representative Raman peaks at 1011 cm^{-1} and 1672 cm^{-1} with an α value of 0.05. The standard deviations of the spectra were calculated and were superimposed on the averaged spectra as shown by the grey dotted line around curve (a), (b) and (c). The standard deviations of the spectra are small for most wavenumbers in the spectral range of interest except in the region around 1650 cm^{-1} , which is assigned to amide I bonding of proteins. This observation suggests that the inter-cell variation was low in cells of the same type due to the well regulation of contents in the cells. The large standard deviation at the protein band 1650 cm^{-1} may be due to the possibility that K562 cells were not synchronized in the cell cycle thus demonstrated various levels of protein expression in the different stages of the cycle.

By visual inspection, it can be noted in Fig. 3 that the Raman peak at 734 cm^{-1} increased in apoptotic cells and decreased in necrotic cells compared to normal cells. This band was assigned to choline groups of phospholipids and it indicated the increase of membranous lipids in apoptotic cells and the decrease of membranous lipids in necrotic cells.

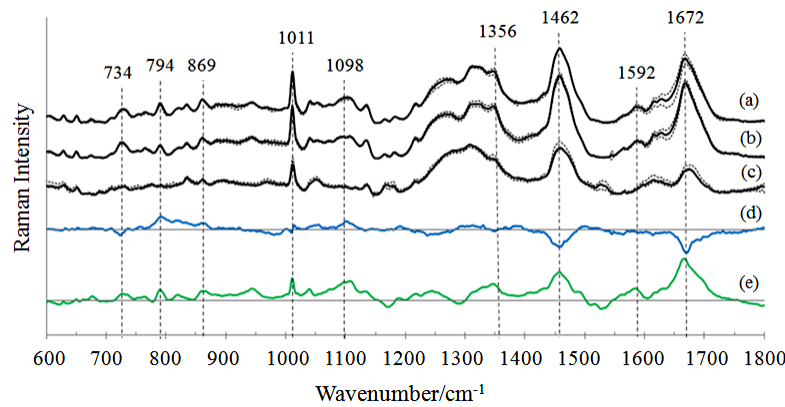


Fig. 3. Averaged Raman spectra of twenty (a) live K562 cells, (b) apoptotic cells and (c) necrotic cells. The standard deviations of the spectra were superimposed on the averaged spectra as shown by the grey dotted lines, which are too small to observe at most wavenumbers. Curve (d) was the difference between spectra (a) and (b) while curve (e) was the difference between spectra (a) and (c).

Raman peaks at 794 cm^{-1} , 1098 cm^{-1} and 1356 cm^{-1} were assigned to O-P-O, PO_4^{2-} nucleic acids backbone vibration and DNA-purine bases bonds of polynucleotide chain. From the basic components spectra, the peak around 1592 cm^{-1} is solely associated with nucleic acids. These bands were consistently lower in necrotic cells compared to normal cells indicating a significant reduction in both DNA and RNA concentration in necrotic cells, whereas the DNA/RNA related peak reduction in apoptotic cells was observed only at peak 794 cm^{-1} and the region around 1098 cm^{-1} .

Proteins have a prominent peak at 1011 cm^{-1} that was assigned to the symmetric ring breathing mode of phenylalanine [21] and does not overlap with Raman peaks of other components. This peak intensity notably decreased in necrotic cells but no significant change in intensity was observed in apoptotic cells. Raman peak at 1462 cm^{-1} can be assigned to CH_2 bending mode found primarily in proteins and lipids. The intensity at this wavenumber decreased in the necrotic cell spectra whereas an increase in this peak was observed in the apoptotic cell spectra compared to live cells. At the region of 1672 cm^{-1} , there was a significant increase in intensity for apoptotic cells while on the other hand there was a relatively large drop in intensity for necrotic cells. This region is assigned to C = O stretching mode of proteins and also C = C lipids stretch [15]. The Raman peaks of proteins and lipids overlap significantly at 1462 cm^{-1} and 1672 cm^{-1} . It is impossible to determine if changes in

the peak intensity at these wavenumbers are due to the alteration in protein level alone or lipid level alone, or both of them.

Figure 4 shows the average Raman spectra of live, apoptotic and necrotic K562 cells and the fittings to the combination of basic biochemical component spectra as in Fig. 2. The Raman spectrum of empty sample cell was also included in the fitting to improve the goodness of fitting. Each fitting coefficient was divided by the sum of all coefficients (excluding the coefficient for the spectrum of empty sample cell) and converted to a percentage, to represent the relative contributions of basic biochemical component spectra to the bulk cell spectra. For example, 61% of the live cell spectrum was contributed by protein, 7.4% by triolein, 16.7% by phosphatidylcholine, 5.1% by DNA, 7.2% by RNA and 2.7% by glycogen according to Fig. 4(a). The contributions of protein, DNA and glycogen were lower, whereas lipids and RNA contents were higher in apoptotic cells as compared to live cells. In contrast, the protein content in necrotic cells was higher than live and apoptotic cells while lipids, nucleic acids and glycogen content were lower as compared to others.

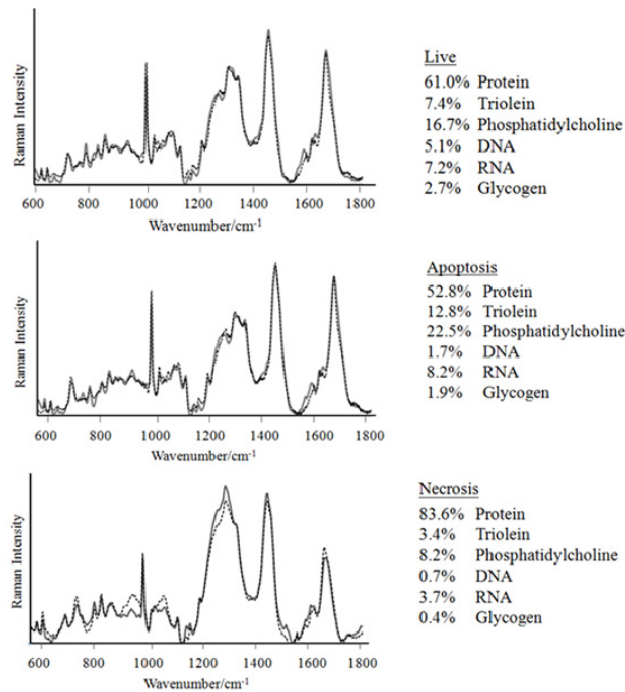


Fig. 4. Mean Raman spectra of (a) live, (b) apoptotic and (c) necrotic cells and the corresponding fittings. The thick solid lines represent the measured spectra of cells while the thin dotted lines represent the fitting. The percentages to the right of each plot represent the fitting coefficients of the basic components that have been divided by the sum of all coefficients to represent the relative contributions of basic biochemical component spectra to the bulk cell spectra.

Figure 5 shows the fitting coefficients for each basic biochemical component in live, apoptotic and necrotic cells. The error bars indicate the standard deviations of the fitting coefficients measured in twenty different cells. The star symbol (*) above the error bars indicates that the relative amount of the specific biochemical component is statistically different between two groups at a significance level of $p < 0.05$. The amounts of triolein, phosphatidylcholine and RNA were significantly higher while DNA content was lower in apoptotic cells than in live cells. However, no significant differences were observed in the levels of protein and glycogen in apoptotic cells as compared to the control group (live cells). Moreover, all of the biochemical components in necrotic cells were significantly lower than

live cells. In this study, we performed a direct comparison in the amounts of each biochemical component between apoptotic and necrotic cells. The analysis in Fig. 5 indicates that the amounts of all the biochemical components in necrotic cells were lower than that in apoptotic cells. A Wilcoxon signed-rank test result shows that the differences in all components between these two death modes are significant to a level of $p < 0.05$.

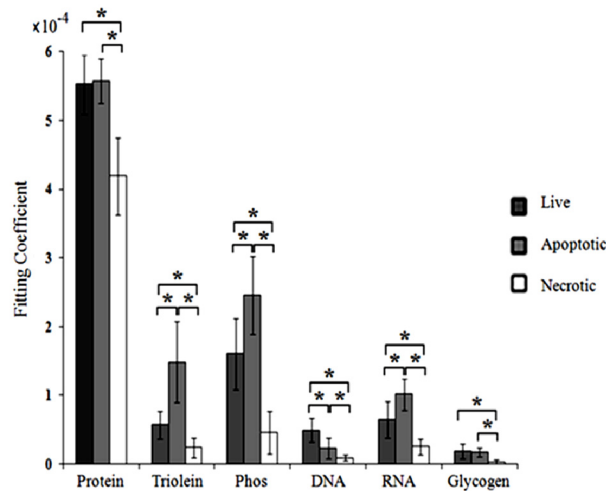


Fig. 5. Fitting coefficients of each basic biochemical component in live, apoptotic and necrotic cells. (*) indicates a significance level of $p < 0.05$ obtained by Wilcoxon signed-rank test. "Phos" represents "Phosphatidylcholine".

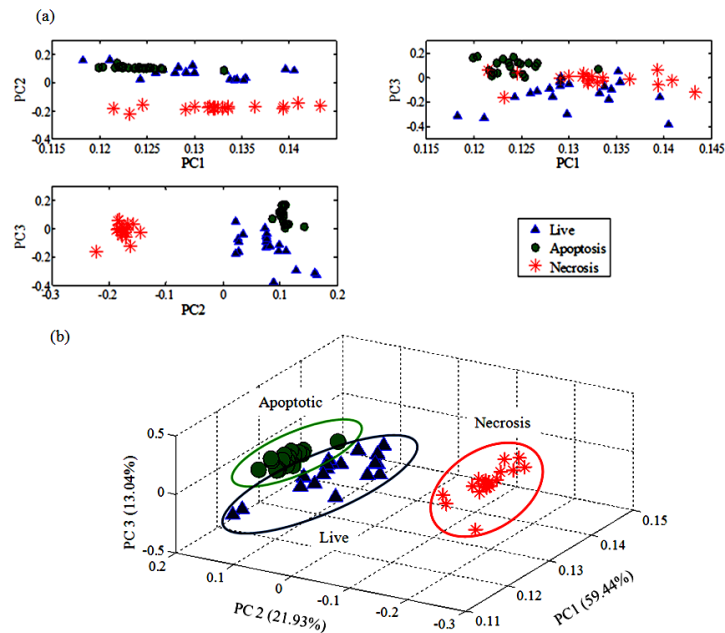


Fig. 6. (a) 2-D and (b) 3-D PCA plots show the separation of data based on different modes of cell death. The percent variance captured by each PC is shown in parenthesis along each axis in (b).

PCA was performed on the raw cell spectra and the scores for the first ten principal components were analyzed with Kruskal-Wallis one-way analysis of variance. It was found

that the means of PC 2 and PC 3 scores show significant differences among the three groups of cells. Although PC 1 accounts for the most variance in the data sets, it does not show significant difference across the groups in Kruskal-Wallis analysis. Two-dimensional plots were constructed with different combination of scores for the first three principal components (PC1, PC2 and PC3) and a three-dimensional plot was also constructed with the three sets of scores as shown in Fig. 6. From Fig. 6(a), it was found that necrotic cells can be easily distinguished from live and apoptotic cells by the first pair of principal components. The main discriminant was the score of PC 2, where necrotic cells have negative scores while live and apoptotic cells have positive scores. The PC 1 score distribution of apoptotic cells was narrower than that of live and necrotic cells where all cells score positively in the first principal component. Live and apoptotic K562 cells were not separable with PC 1 and PC 2 but they could be separated by the scores of PC 3, which were negative for live cells and positive for apoptotic cells. However, necrotic cells have a scattered distribution over the range of PC 3 scores for both live and apoptotic cells and thus there is not any pair in the first three principal components that could be used to separate three groups of cells effectively. A three dimensional PCA plots employing all the first three PCs was then constructed as in Fig. 6(b) and showed excellent separation of apoptotic and necrotic cells with live cells in between the two groups.

4. Discussion

4.1 Selection of basic biochemical components

Seven basic biochemical components were selected because of the following reasons. Actin and albumin were chosen as representative of proteins as the two types of protein account for the varieties and complexities of cellular proteins; albumin is the most common plasma protein while actin is found mostly in the cytoskeleton. Although the addition of albumin spectra has significantly reduced fitting residuals, the Raman spectra of albumin and actin look similar. Therefore the relative contribution of actin and albumin are summed and regarded as protein in this paper. Phosphatidylcholine and triolein were chosen to represent membranous lipids [22] and non-membranous lipids in cells, respectively. Deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) are two main types of nucleic acids found in all living organisms. In spite of the close chemical similarity, they were highly specific in carrying out their unique functions. The amount of DNA gives indication of nuclear content and amount of RNA gives indication of protein synthesizing activity in cells. Glycogen is the main form of polysaccharides in cells, which is the secondary long-term cellular energy storage form after lipids. The amount of glycogen in cells is relatively low compared to lipids because it is less compact and has lower energy storage efficiency.

4.2 Biochemical Component Analysis revealing biochemical changes underlying apoptosis and necrosis

A. Biochemical changes in apoptotic cells

It can be seen in Fig. 5 that the level of protein in apoptotic cells was not significantly different from live cells. It is known that the initiation of caspase cascade reaction during apoptosis activates effector caspase proteins that cleave other protein substrates within the cell, for example, breaking down cytoskeletons [23, 24]. Consequently, the increase in caspase protein is likely counteracted by the breakdown of other cellular proteins, which would result in a minute change in the total protein amount in cell [25, 26].

The levels of triolein and phosphatidylcholine in apoptotic cells showed significant increases relative to the control group in Fig. 5, which can be explained by their roles in apoptosis. Triolein, a non-membranous lipid, increases in apoptotic cells due to the accumulation of unsaturated lipids in cytoplasm forming lipids bodies. This finding agrees with a previous study [22], where Zoladek et al. took Raman images in live and apoptotic

human breast cancer cells using 1659 cm^{-1} band to visualize lipids distribution over 6 hours. In contrast, the increase in phosphatidylcholine indicates the accumulation of membranous lipids in cells mainly attributed to the formation of apoptotic bodies. The packaging of organelles and cellular contents into apoptotic bodies and the event of membrane blebbing in apoptosis require the synthesis of new membrane lipids [27].

Decreased DNA content in apoptotic cells relative to live cells in Fig. 5 suggests that most cells in this study were in the late apoptosis stage. In the early apoptosis stage, nuclear condensation occurs where chromatin is compacted against perinuclear envelope in the hallmark process of apoptosis called pyknosis and DNA content is increased as confirmed in a previous Raman study [22]. However, nuclear condensation in apoptosis is a short process. In the fragmentation process after condensation, DNA is cleaved into short fragments to be packed in apoptotic bodies, which causes reduction in the intensity of DNA peaks mainly at O-P-O phosphodiester backbones band at 794 cm^{-1} [15]. The breakdown of phosphodiester bonds and DNA bases during nuclear fragmentation is the hallmark event of late apoptosis. In addition, cytosine arabinoside used in this study to induce apoptosis in K562 cells will induce erythroid differentiation in K562 cells where they progressively lose their DNA content even before apoptosis cascade is activated. This expected drop in DNA content is confirmed by the observed decrease in DNA content based on the fitting result in Fig. 5. RNA shows a slight increase in apoptotic cells compared to live K562 cells. This can be attributed to the increase in RNA that directs the synthesis of caspase proteins [28].

Even though apoptosis is an active process that requires energy, no significant reduction in the relative amount of glycogen in apoptotic cells is observed. It may suggest that glucose supply in the culture medium used for culturing these apoptotic cells was sufficient in this process. Since glucose is favored as the primary and immediate source of energy in apoptosis, there is no need to initiate glycogenolysis to break down glycogen for energy if the glucose supply is adequate.

B. Biochemical changes in necrotic cells

In necrotic cells, the relative amounts of all the basic biochemical components showed significant reduction compared to the control group. The reduction in DNA content is due to DNA degradation by the breakdown of both phosphodiester bonds and DNA bases. Different from apoptosis where DNA is first broken down into nuclear fragments, DNA degradation in necrosis is a random event and nuclear fragments are not packed into vesicles to facilitate the uptake and incorporation into neighboring cells. RNA level in the cells reduces in necrotic cells because necrosis is a passive mechanism of cell death that requires no new protein needs to be synthesized. The decrease in the amount of phosphatidylcholine that represents membranous lipids suggests the loss of membrane integrity in necrotic cells. The rupture of cell plasma membrane causes the leakage of cell content into the surroundings as the ruptured membrane allows the random diffusion of cell content across the membrane. This explains the decrease in cellular lipids, proteins content and glycogen level in necrotic cells even though necrosis does not require energy.

4.3 Comparison of PCA and BCA in cell death classification

We have explored the use of biochemical component analysis (BCA) and principal component analysis (PCA) in analyzing Raman spectra of different groups of live and dead cells. Previous studies have demonstrated that either methods is capable in highlighting the variance in Raman spectra across different groups of cells [3, 15, 29]; however, little effort has been made to compare the classification performance and interpretability of the two methods. Hence, we have fed the features extracted from both methods, i.e. the (fitting coefficients of each biochemical coefficient in BCA and the scores of principal components in PCA, into a linear discriminant classifier and a leave-one-out cross validation method was employed to train the classifier. First, we compared the resulting classification accuracies of

PCA and BCA by using a single feature extracted from each method. When the score of a single principal component was used, the classification accuracy is 68.3% for PC 1, 91.7% for PC 2, 63.3% for PC 3, 60.0% for PC 4 and 63.3% for PC5. The classification accuracy using the fitting coefficient of a single basic biochemical component is 70.0% for protein, 83.3% for both triolein and phosphatidylcholine, 80.0% for DNA, 73.3% for RNA and 66.7% for glycogen.

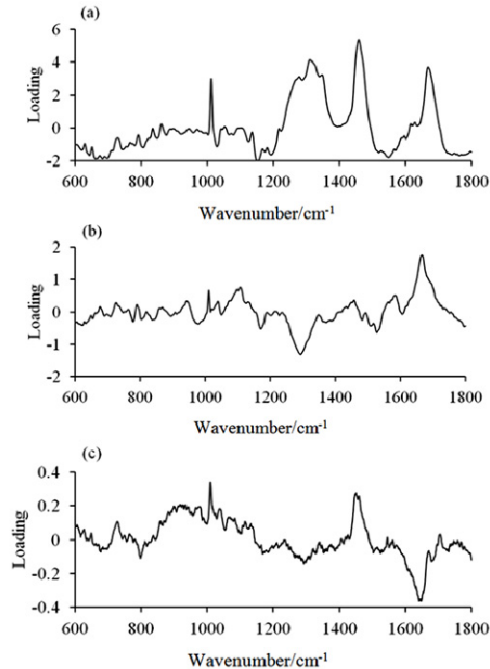


Fig. 7. The spectra of first three principal components in PCA, where (a) is PC 1, (b) PC 2 and (c) PC 3.

Figure 7(a) shows that PC 1 resembles the average cell spectrum. It is noted that the classification with PC 1 score is 68.3% which is lower than the classification accuracy of PC 2 even though PC 1 accounts for most of the variance in cell spectra. By decomposing PC 1 with regards to the basic biochemical components, i.e. performing BCA on PC 1, we find that the majority of PC 1 is contributed by protein (58%) and phosphatidylcholine (20%) while other biochemical components contribute less than 10% each to PC 1. Since there is no significant change in the relative amount of protein in live and apoptotic cells as shown in Fig. 5 and PC 1 is largely contributed by protein spectrum, the low classification accuracy using PC 1 could be attributed to its high protein content. Decomposing PC 2 spectrum with regards to the basic biochemical components returns a large fitting residual (result not shown). This is likely due to the fact that PC 2 is orthogonal to PC 1 in the vector space thus does not show strong correlation to the basic biochemical components spectra to which PC 1 can be fit well. However, it is noted that PC 2 as in Fig. 7(b) captures two prominent Raman peaks at 1098 cm^{-1} which corresponds to O-P-O DNA backbone vibration and 1672 cm^{-1} which corresponds to the C = O stretching mode of proteins and C = C lipids stretch. This observation explains why PC 2 can be used to classify live and dead cells with good accuracies, considering that both lipids and DNA demonstrate significance differences between live and dead cells as shown in Fig. 5. Similar to PC 2, the fitting of PC 3 using BCA also returns a large fitting residual. PC 3 as in Fig. 7(c) captures three peaks, respectively, at 734 cm^{-1} which corresponds to choline group of phospholipids, 1011 cm^{-1} which can be

assigned to the symmetric ring breathing mode of phenylalanine and 1462 cm^{-1} which corresponds to the CH_2 bending mode of proteins and lipids. This result suggests that the combination of proteins' and lipids' peaks is not a good parameter in distinguishing cell death modes as the classification accuracy using PC 3 is only 63.3%.

When using the fitting coefficient of BCA to classify live, apoptotic and necrotic cells, both lipids (triolein and phosphatidylcholine) score the highest accuracies among all basic biochemical components with 83.3%, followed by DNA which scores 80%. Proteins and glycogen have lowest classification accuracies, which are 70% and 66.7% respectively. This can be due to the fact that the fitting coefficients for both proteins and glycogen show no significant differences among cell groups in Wilcoxon signed-rank test as shown in Fig. 5.

Table 1. Classification accuracies using two principal component scores obtained from PCA

	<i>PC 1</i>	<i>PC 2</i>	<i>PC 3</i>	<i>PC 4</i>	<i>PC 5</i>
PC 1	68.3%	95.0%	75%	85.0%	78.3%
PC 2		91.7%	96.7%	95.0%	95.0%
PC 3			63.3%	80.0%	80.0%
PC 4				60.0%	78.3%
PC 5					63.3%

Table 2. Classification accuracies using two fitting coefficients obtained from BCA

	Protein	Triolein	Phos	DNA	RNA	Glycogen
Protein	70.0%	86.7%	93.3%	90.0%	90.0%	86.7%
Triolein		83.3%	93.3%	96.7%	90.0%	86.7%
Phos			83.3%	86.7%	83.3%	76.7%
DNA				76.7%	90.0%	90.0%
RNA					73.3%	80.0%
Glycogen						66.7%

To explore the minimum number of components for high classification accuracy, the use of two PC scores or fitting coefficients in conducting LDA analysis has been tried and the classification results are presented in Tables 1 and 2. Generally, when two components from PCA or BCA are used, the classification accuracies will be improved. For example, when the score of PC 1 alone is included in the classifier, the accuracy is 68.3%. In contrast, when PC 2 score is also incorporated into the classification, the accuracy will be improved to 95%. Similarly, the accuracy of using only PC 3 score will be improved from 63.3% to 75% when the score of PC 1 is included in the classifier. The highest accuracy achieved is 96.7% when scores of PC 2 and PC 3 are used.

The classification accuracies for BCA also improved when two fitting coefficients are used. For example, the classification accuracy by using the fitting coefficient of proteins is improved from 70% to 86.7% when the fitting coefficient of triolein is added, while the addition of fitting coefficient of phosphatidylcholine can improve the classification accuracy to 93.3%. The highest classification accuracy that can be achieved is 96.7% when the fitting coefficients of DNA and triolein are used. This result again indicates that the combination of DNA and membranous lipids can be used to discriminate between live and dead cells with excellent accuracies. This combination has been explored recently by Zoladek et. al [22] for Raman imaging of apoptosis in human breast cancer cells. Besides DNA and lipids, several studies [3, 15, 30] has also highlighted the changes in protein amount between live and dead cells. For PCA method, a combination of scores from the first three principal components results in a classification accuracy of 98.3%. The addition of another score from the fourth PC can further improve the classification accuracy to 100%. In BCA, the combination of three fitting coefficients for proteins, DNA and lipids yields a classification accuracy of 100% in this study.

It is worth noting that the highest classification accuracy achieved using the fitting coefficients from BCA method is slightly better than that achieved with the scores of

principal components from PCA method. However, our results have also showed that the classification of live, apoptotic and necrotic cells based on both PCA and BCA had excellent accuracies, and neither method has showed a clear superiority over the other method in this respect. However, principal component analysis has a disadvantage in that it does not reveal the biological basis of the changes observed in the Raman spectra. Most principal components (except PC 1) extracted in PCA do not carry physically meaningful information about biochemical changes in cells. The BCA method, on the other hand, is capable of extracting the relative concentrations of biochemical components contributing to the spectra that could provide useful information about the physiological and structural changes in different cell death modes. Proper utility of the BCA method however requires good knowledge of the underlying biochemical constituents in a cell in order to achieve good fitting result with minimal residual. Thus, PCA analysis method can be favorable when the spectra of basic constituents of samples are unavailable.

5. Conclusions

In this study, we demonstrate a simple Raman spectroscopic technique to measure and analyze the different death modes of individual leukemic cells in suspension. The comparison of cell death classification accuracy between PCA and BCA has been performed by using LDA based on features extracted in these two methods. Both methods showed excellent performance in classifying live, apoptotic and necrotic cells. In particular, 100% accuracy is achieved by using a combination of the first four principal components in PCA or a combination of fitting coefficients of proteins, DNA and lipids in BCA. Neither PCA nor BCA shows a clear superiority over the other in achieving higher classification accuracy. BCA is shown to be a robust and reproducible method that could provide insight into the biological changes in different cell groups. The detection of fine biochemical changes in cell spectra has been demonstrated and this will help develop Raman spectroscopy techniques to monitor cell death in fields such as pharmacology and tissue engineering. PCA can be favorable when the spectra of basic constituents of samples are unavailable. This study clearly demonstrates the pros and cons of the two methods of data analysis and points out their applicability in various scenarios in Raman spectroscopy for cell studies.

Acknowledgment

We gratefully acknowledge the financial support from the New Investigator Grant (Project No. NMRC/NIG/1044/2011) funded by the National Medical Research Council (NMRC) in Singapore.