Raman spectroscopy: a potential platform for the rapid measurement of carbon nanotube-induced cytotoxicity

Peter Knief  
*Technological University Dublin*

Colin Clarke  
*Technological University Dublin*, Colin.Clarke@tudublin.ie

Eva Herzog  
*Technological University Dublin*, eva.herzog@tudublin.ie

Maria Davoren  
*Technological University Dublin*, Maria.davoren@tudublin.ie

Fiona Lyng  
*Technological University Dublin*, Fiona.lyng@tudublin.ie

Follow this and additional works at: [https://arrow.tudublin.ie/radart](https://arrow.tudublin.ie/radart)

**Recommended Citation**

Authors
Peter Knief, Colin Clarke, Eva Herzog, Maria Davoren, Fiona Lyng, Aidan Meade, and Hugh Byrne

This article is available at ARROW@TU Dublin: https://arrow.tudublin.ie/radart/11
Raman spectroscopy – a potential platform for the rapid measurement of carbon nanotube-induced cytotoxicity

Peter Knief*, Colin Clarke, Eva Herzog, Maria Davoren, Fiona M. Lyng, Aidan D. Meade and Hugh J. Byrne

Received (in XXX, XXX) [Xth XXXXXXXXX 200X], Accepted [Xth XXXXXXXXX 200X]
First published on the web [Xth XXXXXXXXX 200X]
DOI: 10.1039/b000000x

In this study the suitability of Raman spectroscopy for the determination of carbon nanotube mediated toxicity on human alveolar carcinoma epithelial cells (A549) is explored. The exposure of this cell line represents the primary pathway of exposure in humans, that of inhalation. Peak ratio analysis demonstrates a dose dependent response which correlates to previous toxicological studies. Principal component analysis is employed to further classify cellular response as a function of dose and to examine differences between spectra as a function of exposed concentration. To further illustrate the potential of Raman spectroscopy in this field, Partial Least Squares (PLS) regression and genetic algorithm feature selection have been utilised to demonstrate that clonogenic end points, and therefore toxic response, can be potentially predicted from spectra of cells exposed to un-determined doses, removing the need for costly and time consuming biochemical assays. This preliminary study demonstrates the potential of Raman spectroscopy as a probe of cytotoxicity to nanoparticle exposure.

* Author to whom correspondence should be addressed.
1 School of Physics/Focas Research Institute, Dublin Institute of Technology, Kevin Street, Dublin
2 Radiation and Environmental Science Centre, Focas Research Institute, Dublin Institute of Technology, Kevin Street, Dublin 8
3 Focas Research Institute, Dublin Institute of Technology, Kevin Street, Dublin 8
**Introduction**

Nanomaterials are considered to be a new class of materials with unusual characteristics, not only due to the chemistry of the materials themselves, but because their dimensions result in new physical characteristics that have a significant impact on their chemical properties. In addition, chemical modifications can change the optical, magnetic or electric properties of these materials. Thus nanomaterials are a new class of materials that can be employed in wide ranging applications in nanoscience, medicine and engineering.

Carbon nanotubes, in both their multi- and single-walled forms, have attracted significant attention since their emergence in 1991. They are one dimensional macro molecules of rolled graphene sheets, either single or multi walled, with diameter of the order of nanometers and a length up to several centimetres. It is anticipated that their huge potential will see them in mass production in the near future. Given the likely widespread applications of these materials, concerns exist regarding potential toxic responses and an evaluation of their biological effects is required. Various studies have already been published reporting toxicological or inflammatory responses in animal models. On a cellular level, some in vitro studies have shown less dramatic effects in terms of viability and proliferative capacity although chemical modification or functionalisation of the nanotubes can increase the toxic response. The diversity of the reports of the extent and mechanism of the toxic response however clearly demonstrates the need for more systematic investigations.

Recent studies have shown that common colorimetric assays interact with single walled carbon nanotubes (SWCNT) themselves, compromising their overall suitability in toxicological assays. The cell culture medium itself is altered by the interaction with SWCNTs, challenging conventional cytology in general. Clonogenic studies have been demonstrated to be a potentially more appropriate toxicological assay and it has been demonstrated that colony-size rather than colony number is a more powerful measurement of cellular toxicity post exposure to SWCNTs. The implications are that the toxic response is possibly one of reduced proliferative capacity of the cells due to medium depletion rather than a reduced viability due to direct interaction with the cells. Such an indirect toxicological response has more recently been demonstrated. While the clonogenic assay yields a reliable assessment of the toxic response to SWCNT exposure in vitro, the endpoints of colony size and number are time consuming, phenomenological and limited in their potential to elucidate underlying biochemical mechanisms. Therefore alternative assay in designing are desirable for the rapid evaluation of cellular toxicity post exposure to carbon nanotubes.

Raman spectroscopy is a very versatile analytical tool, known for its strengths in the physical and chemical characterisation of materials and systems. This technique has previously been employed for the analysis of interactions of SWCNT’s with biomolecules as well as the effect of external toxins as agents for induced cell damage. The modality potentially offers analytical and diagnostic information at a high sub-cellular spatial resolution. It derives additional benefit from the minimal need for processing of biological materials. It has already been shown to be a viable tool for disease diagnosis as well as for the detection of alterations on a cellular level caused by external influences.

The Raman spectrum of a cell also contains chemical information regarding its constituents, providing a complete biochemical fingerprint of the cell, and ultimately exhibiting signatures that are indicative of cell state, e.g. proliferating, apoptotic, necrotic, etc. While changes to individual or combinations of spectral features may give clear indications of cellular response, the complex biochemical changes are often manifest as multivariate changes to the overall spectral response. Multivariate analysis of spectroscopic data delivers an appropriate means to analyse multiple dependent and independent singular features occurring in vibrational spectra of biological materials, delivering a detailed view of the overall response, allowing classification of, for example, tissue pathologies and quantification of response to external stimuli with the additional option of modelling the spectral features for prediction of a biological result.

In this study, Raman spectroscopy is employed as a probe of the toxic response of cellular systems to SWCNT exposure in vitro. A human lung cell line was chosen for experimental purposes, as it represents the potential scenario of inhalation, one of the first steps in the route of exposure. In order to allow a realistic relationship between the experimental data presented here and that in vivo, the SWCNT samples were minimally processed to mimic inhalation of airborne SWCNT dust particles. SWCNT dispersion was carried out in an identical fashion to that employed in previous cytotoxicity studies and cell growth and exposure conditions were identical to those employed in previous clonogenic studies. Dose dependent responses are examined in terms of peak ratios utilised in previous toxicity studies as well as principal component analysis. Finally, the suitability of genetic algorithm optimised Partial Least Squares (PLS) regression as a quantitative model to predict clonogenic endpoints is assessed. Coupled with such a predictive model, spectroscopic analysis is demonstrated to be a potentially powerful analytical technique avoiding time consuming and expensive biochemical assays.

**Experimental Procedures**

### Cell culture

Quartz slides (24.5mm x 24.5mm, UQG Optics Ltd.) were coated for 24 h at 4°C with a sterile solution of 2% gelatine (Type-B from bovine skin) in deionised water (dH2O) solution. Such substrates have previously been shown to be optimal for cell growth and subsequent spectral analysis. Cells of the human alveolar carcinoma epithelial line A549 (ATCC, CCL-185), were cultivated in Dulbecco’s modified minimum essential medium (DMEM, Cambrex). All media were supplemented with 10% foetal bovine serum (FBS) and 45 IU ml⁻¹ penicillin and 45 μg ml⁻¹ streptomycin and cells were maintained at 37°C in a 5% CO₂ humidified incubator. The cells were allowed to attach to the quartz substrates at a concentration of approximately 2 x 10⁶ cells per slide for 24h. After the 24h incubation period, the unattached cells were rinsed off with PBS. An ultrasonic tip...
processor VCX-750 watt) operating at 40% was employed to disperse the SWCNTs in four exposure suspensions of single wall carbon nanotubes (0 mg/l (Control), 1.56mg/l, 6.25mg/l, 25.0

Nanotubes (Carbon Nanotubes Inc.) were employed for the study for consistency with previous studies. The tip was operated at a medium level of output for a total time of 30s carried out in 10s sequential steps to minimise sample heating. The cells were then exposed to 3 ml of each of the different SWCNT suspensions for 96 hours. After the exposure period the slides were rinsed with PBS and fixed in 4% formalin in PBS solution for 10 minutes, rinsed once again in dH$_2$O, and finally stored in dH$_2$O at 4°C prior to conducting the measurements. Three independent sets of cells were exposed at each concentration and all measurements were repeated for each exposure batch.

Spectroscopy

Raman Spectroscopy was carried out with an Instruments S.A. (Horiba Jobin-Yvon) Labram 1B Raman confocal microscope using 514.5 nm laser excitation with a grating of 1800 l/mm, providing a spectral dispersion of about 1 cm$^{-1}$ per pixel. Spectra were recorded using a water immersion lens (Olympus Lum-Plan FL 100x) from substrates immersed in water in a climatic chamber to prevent desiccation of the samples. The immersion reservoir was constructed by inserting a quartz window into the bottom of a Petri dish filled with dH$_2$O. The x100 water immersion objective gave a spot size approximately 1µm at the sample.

All recordings were performed as an average of three individual measurements of one point to reduce the influence of spectral noise. The system was previously calibrated to the spectral line of crystalline silicon, at 520.7 cm$^{-1}$ at a constant room temperature of 21°C. The measurement range was set to an interval of ~250-1750 cm$^{-1}$ in order to detect spectra within the fingerprint region of the cell samples and the characteristic SWCNT features. Before spectral acquisition, the dark current of the system and the system intensity response (using the NIST fluorescent intensity standard SRM 2243), were recorded in triplicate. After a series of spectral measurements on a particular slide, the spectral background of the substrate was acquired. The laser power was set to 23 mW at the sample and the acquisition time was set to 90s which delivered reasonable spectra.

In total, approximately 75 spectra (25 per sample in triplicate) were recorded from the nuclear portion of multiple cells at each concentration. Principal component analysis (PCA- see following section for details) was employed to identify outlier spectra. Cells across the whole area of the sample slide were chosen for measurement in an attempt to ensure a true representation of the sample. This technique was designed to limit variability that might occur due to the spatial position of the laser focal spot within the nuclear portion of cells, and biological variability that could occur between samples of the cell line. It was noted however, that even after repeated washing with PBS, some single wall carbon nanotube aggregates could be visibly observed attached to the cells, although no SWCNTs were observed inside the cells themselves. All measurements reported here were taken away from regions where large aggregates were visible.

Data analysis

In total 321 valid spectra were acquired for the 5 distinct concentrations (Table 1) with a spectral range from 248 to 1751 cm$^{-1}$. The raw spectra were imported into Matlab 7.3 (Mathworks CA, USA) for pre-processing and analysis. Every spectrum was corrected for system intensity response, according to the guidelines of NIST (SRM 2243). Prior to the subtraction of the underlying quartz signature, each spectrum was normalised to the characteristic underlying quartz peak at 486 cm$^{-1}$, without the application of any filtering. Finally the spectra were cropped to a spectral window of 599-1700 cm$^{-1}$ to isolate the fingerprint region. In order to minimise the electronic noise associated with the CCD detector, the spectra were smoothed using the Savitzky Golay algorithm with a 15 point window and a polynomial order of 3 for further analysis.

<table>
<thead>
<tr>
<th>Concentration</th>
<th>Recorded Replicates</th>
<th>Validated Replicates</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0 mg/l</td>
<td>I II III</td>
<td>I II III</td>
</tr>
<tr>
<td>1.56 mg/l</td>
<td>25 25 25</td>
<td>22 21 20</td>
</tr>
<tr>
<td>6.25 mg/l</td>
<td>25 25 25</td>
<td>22 20 21</td>
</tr>
<tr>
<td>25.0 mg/l</td>
<td>25 25 25</td>
<td>22 24 21</td>
</tr>
<tr>
<td>100.0 mg/l</td>
<td>25 25 25</td>
<td>21 23 21</td>
</tr>
</tbody>
</table>

Principal component analysis

Principal component analysis (PCA) is a method of multivariate statistical analysis broadly used with datasets of multiple dimensions. It allows the reduction of the number of variables in a multidimensional dataset while retaining most of the variation within the dataset. The principal components (PCs) obtained are not correlated to each other and are also called eigenvectors or loadings. The lack of correlation means that the PCs represent valuable different ‘dimensions’ of the data. The order of the PCs describes their relative importance for the dataset. PC1 describes the highest amount of variation, PC2 the second greatest and so on. The variance of the PC is sometimes called the eigenvalue or scores of the PC. When PCA is successfully applied, the eigenvalues of the PCs are progressively lower and ideally the variation in the data set can be described sufficiently by a few PCs whose eigenvalues are most significant. The actual number of PCs that feeds into further analysis is dependent on the variance covered by the eigenvalue of a PC, visualized by e.g. a Scree plot, and the threshold one applies, in order to give a satisfactory representation of the original dataset, explaining most of the variance within. As the scores of a sample in PCs are orthogonal to each other and uncorrelated they represent coordinates along the dimensions of the PCs in e.g. a three dimensional space for 3PCs, used to access possible separation of certain groups within samples. Analysing the loadings of a PC can give information about the variable based source for the variance in a PC. Variables of the loadings within the PC with positive value indicate a positive contribution to the nature and dimension of the PC whereas negative values show an inverse relation to the PC, not giving a positive contribution to the variance covered by the PC and therefore not contributing to the dimension of the PC.
In this study, the datasets for all exposure concentrations were pre-processed as described previously, arranged conforming to the required SAISIR structure (SAISIR (2008)(c). Package of function for chemometrics in the MATLAB (Registered) environment. Dominique Bertrand coordinator. Unité de Sensométrie et de Chimiométrie 27) and then fed into PCA analysis using Matlab 7.3 (Mathworks CA, USA) with the accompanying statistics toolbox. In a second PCA analysis the spectra were doubly derivatized, a common technique in dealing with spectral data to further reduce the baselines and backgrounds28, and also subjected to PCA using Matlab.

Partial least squares modelling

First described by Wold in 1960, partial least squares (PLS) is a popular and well known tool in the field of chemometrics29-31. The aim of PLS is the construction of a model to describe the response variables (i.e. analyte concentration) in terms of the observed variables (spectra) from a set of training data. The least squares model is given by:

\[ Y = XB + E \]  

Equation 1

where \( Y = n \times m \) are the dependent variables (i.e. concentration), \( X = n \times p \) are the independent variables (i.e. Raman spectra), \( B = p \times m \) is the matrix of regression parameters for each component in \( Y \) and \( E \) are the matrix of residuals (differences between measured and predicted variables). PLS decomposition is similar to that of principal component analysis (PCA). PCA produces scores and loading matrix, a series of weight vectors are the targets. The initial weight vector is calculated as follows:

\[ \text{Initial scores vector is calculated as: } \]

\[ w_i = X^T y / \left[ \| X^T y \| \right] \]  

Equation 2

The initial scores vector is calculated as:

\[ t_i = Xw_i \]  

Equation 3

and the loadings:

\[ p_i = X^T t_i / \left[ \| t_i \| \right] \]  

Equation 4

The regression parameters are calculated as follows:

\[ \hat{b} = y^T t_i / \left( t_i t_i^T \right) \]  

Equation 5

The residual matrix is calculated as:

\[ E_i = X - t_i p_i^T \]  

Equation 6

The algorithm continues for each factor used, taking \( E_i \), instead of the weight matrix to calculate the second set of weights. When presented with an unknown spectrum, \( y \) is determined using \( W \) and \( P \) to compute scores for the unknown spectrum along with the regression parameters allowing the concentration of \( y \) to be determined from equation 1. PLS calibration models were constructed in this work using the SIMPLS algorithm with root mean squared error of cross validation (RMSECV) as the fitness value. Leave one out cross validation was used to select the number of latent variables (LVs) to retain. To construct the PLS-models the Eigenvector toolbox 3.5 for MATLAB was used.

Feature selection using genetic algorithms

Calibration models are known to be greatly improved through the application of efficient feature selection methods, increasing the predictive ability and reducing model complexity. One such method is the adaptive search technique known as the genetic algorithm (GA). Here a GA based variable selection procedure is used to reduce the original spectra to a subset of wavenumbers to correlate Raman spectra to response. The first generation for evaluation is a random population consisting of a number of individuals or “chromosomes”, each containing a subset of the original variables. Each chromosome is composed of a vector of 1s and 0s, corresponding to the wavenumbers in the X matrix, (1 if selected and 0 if not) where each wave number is termed a “gene”. The performance of models resulting from each chromosome is determined by means of a fitness function (here the root mean square error of cross validation is used). Once each generation is evaluated a new set of chromosomes is produced by retaining and “crossing” the fittest individuals from the previous generation. “Mutations” are also produced which force the evaluation of new combinations avoiding saturation with similar sets of events and can further lower the number of variables and fitness values. The process continues until the difference in mean fitness level between successive generations is below a certain threshold the GA is terminated to avoid over-training and avoid over fitting risk in the PLS model32-34. Feature selection in this work was achieved using GA optimisation (with the genpols MATLAB toolbox by Ledardi) over 100 runs requiring approximately 60mins (see table 1 for GA settings).

Each calibration model was evaluated using root mean squared error of cross validation (RMSECV) and root mean squared error of calibration (RMSEC) performed on the calibration set. The root mean squared error of prediction (RMSEP) of the independent testing is also calculated. The root mean squared error is calculated for each dataset as follows:

\[ RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_{act} - y_{pred})^2} \]  

Equation 7

In this study PLS regression enhanced through GA feature selection is used to construct a calibration model to predict the end points of clonogenic studies directly from dose dependant Raman spectra. In order to ensure over fitting is avoided the data is split into two sets, a calibration and testing set. Firstly the optimum number of latent variables is chosen with cross validation and the model constructed using the calibration set. The test set is unseen during the training phases and is used as an independent test set validation of the constructed model.

Results and Discussion
Preliminary Spectral Evaluation

The Raman spectrum of a SWCNT sample (suspended in water at concentrations similar to those used during this study) exhibits characteristic radial breathing modes (RBM) in the region of 200 cm\(^{-1}\) (Figure 1). These features describe the synchronous oscillation of the atoms of the nanotube in the radial direction and can be used to define structural characteristics of SWCNTs such as their diameter, metallicity, and helicity \(^{36}\). The so-called “disorder-included” D band appears at 1330-1390 cm\(^{-1}\) and is reputedly an indicator for disorder in the graphene sheet. The tangential mode, or G-Band, appears at 1583-1605 cm\(^{-1}\) originating from tangential oscillations of the carbon atoms in the nanotubes \(^{36-38}\).

Spectra of A549 cells Figure 2 (A) show classic features within the amide I band area of 1656-1690 cm\(^{-1}\), consisting of ~80% of CO stretching, ~10% CN stretching, and ~10% NH bending vibration modes, indicating protein based \(\alpha\)-helix, random coil and \(\beta\)-sheet structures. In the amide III area at about 1238 cm\(^{-1}\) \(\beta\)-sheet and random coil structures are indicated by ~30% CN stretching and ~30% NH bending vibrations, as well as ~10% CO stretching and ~10% O=C-N bending vibrations. Vibrational features of amino acids and amino acid hydro halides appear in the area of 1485 - 1660 cm\(^{-1}\) (NH deformation vibrations and \(\alpha\)-form C=O stretching of polypeptides).

Characteristic signals of lipid related groups appear at 965 cm\(^{-1}\) (CN asymmetric stretching vibrations), 1170 cm\(^{-1}\) (weak CO-O-C symmetric stretching) and 1451 cm\(^{-1}\) (CH\(_2\) scissoring and CH\(_3\) bending vibrations) \(^{38, 39, 41}\). As the samples were rinsed before fixing as described in the cell culture section above, it is assumed that all features are cellular in origin. In Figure 2 (B), an average spectrum of a high concentration exposure cellular sample (25mg/l) is shown. Strong contributions of the G-Line & D-Line features of SWCNTs as well as common cellular spectral features are clearly visible, although the SWCNTs were washed off thoroughly and were not visible microscopically. In a previous study, in samples prepared under identical conditions, no SWCNTs could be observed internalised in the cells and so it is assumed that small bundles or ropes adhere to the cell surface \(^9\).

The strongest peak of the typical SWCNT spectrum, the G-line at about 1585 cm\(^{-1}\) \(^{36}\), overlaps strongly with the amide I region of the cellular spectra (1637, 1656-1690 cm\(^{-1}\) \(^{39, 42}\). This makes it difficult to utilise this band for analysis of cellular response to the SWCNT exposure without deconvolution.

After background removal, the region of 1502-1700 cm\(^{-1}\), was extracted and fitted with a mixture of Gaussian/Lorenzian band functions to extract the relative contributions of the SWCNT G-line and the cellular amide I band. Figure 3 shows the intensity of the SWCNT G-line as a function of exposure dose in terms of concentration (mg/l). Although the Raman intensity is approximately linear as a function of dose up to ~30mg/l, the maximum dose of 100mg/l shows significant deviation from this. This apparent saturation of the response may be a result of over dosage, the nanotubes not being effectively dispersed throughout the sample, and/or an effect of the increased optical density of the residual carbon nanotubes which are resonant at the Raman wavelength, limiting penetration of the light into the sample and absorption of the light scattered by the sample. Figure 4 shows the intensity of the amide I Raman band as a function of SWCNT dosage. The intensity is seen to be only weakly dependent on dosage, indicating that the reduced intensity of the SWCNT G-line has origin primarily in saturation of dosage rather than optical effects, although the slight reduction at large doses points towards some optical effects. Such a saturation of dosage resulting in large aggregates of nanotubes adhered to the cells has indeed been seen in other studies \(^9\).
A direct visual comparison of the cellular Raman spectra demonstrates clearly that several individual peaks are altered as a result of exposure (Figure 5). Examination of the spectra reveals changes to the 1030-1060 cm$^{-1}$ lipid related bands caused by C-O-P stretching$^{18,40,41}$ and CO-O-C sym. stretching$^{18,40,41}$ vibrations, an observation which correlates well with the recent work of Davoren et al.$^{9}$ which, using Transmission Electron Microscopy (TEM), demonstrated an increase of surfactant storing lamellar bodies in A549 cells after exposure to SWCNTs, which supports the assumption of a change in the overall lipid content in the exposed cells. It should be noted that although the nucleus is primarily targeted, the spot of ~2-3mm diameter will pick up some of the neighbouring and overlaying cytoplasm. Changes to the RNA ribose C-O vibration at 930-960cm$^{-1}$ and 1295-1304 cm$^{-1}$, as well as changes to the adenine and guanine activity at ~1345 cm$^{-1}$ are also observed. It is possible, however, that the D-line of the SWCNTs may be masking the response in this latter region, but changes to the cellular spectra are clearly visible after deconvolution of the amide III band area, far from the spectral features of the SWCNTs, with its different conformations at approximately 1238, 1258, 1271 cm$^{-1}$ by a combined Gaussian and Lorenzian fit, known as a pseudo-Voigt function$^{44,45}$ to approximate the Voigt profile, with a total of nine individual centred peaks, identified from the second derivative of the unfitted region.$^{15}$

Although difficult to precisely assign the many overlapping bands, the use of the ratio of Raman peak heights for CH$_2$ deformation modes at ~1302cm$^{-1}$, DNA bases guanine, adenine and thymine at 1287and 1338cm$^{-1}$ versus the amide III band at 1238cm$^{-1}$ have previously been used for estimation of cellular toxicity.$^{15}$ The three ratios exhibit an general trend of an overall increase with an increase of dose with the exception of the largest exposure dose of 100mg/ml which has been shown to have a saturated spectral/exposure response (Figure 3). All spectral features, with the exception of the 1338cm$^{-1}$ band are far from any SWCNT bands, and the fact that the 1338cm$^{-1}$ band exhibits the same trends indicates that there is minimal interference from the underlying tail of the SWCNT D-line.

Figure 6 shows an approximately linear relationship between the ratio of 1338 cm$^{-1}$/amide III as a function of G-line intensity which should more accurately represent the actual SWCNT dose. The ratios of bands previously identified as cytotoxic markers clearly show a dose dependent response. This dose dependence correlates well with that previously observed for colony size in clonogenic assays on the same samples.$^{12}$ The dose dependent response of the colony size endpoint of the clonogenic study is plotted in (Figure 10). A monotonic decrease in colony size with increasing dose up to ~30mg/l is observed with a saturated response at higher doses. This toxic response has been attributed to a reduced proliferative capacity as a result of medium depletion caused by adsorption of components of the cell growth medium to the SWCNTs.$^{12}$ Figure 11 demonstrates a clear correlation of the dose dependent 1287cm$^{-1}$/amide III peak ratio with toxic response as determined by the colony size endpoint of reference$^{12}$. 

![Figure 4. Intensity of amide I at 1656cm$^{-1}$ versus concentration](image4.png)

![Figure 5. Detailed magnification of Raman Spectra of the A549 Cells at different concentrations smoothed with Savitzky Golay Filter order 3, 15 points.](image5.png)

![Figure 6. Peak ratio of 1287 cm$^{-1}$/amide III versus concentration](image6.png)
The results clearly indicate that dose dependent spectral markers can be identified in the Raman spectra of cellular samples exposed to SWCNTs. However, the intrinsic influences of inhomogeneity of the spatial dispersion of SWCNTs in e.g. cell culture medium and the SWCNT residues adhering to the cells, as well as the complex changes to the spectral response of the cells, demand more elaborate data analysis methods, moving from the univariate approaches described above to the analysis of the spectral data by multivariate analysis. Principle component analysis will thus be employed as a more powerful classification tool, potentially elucidating a more detailed signature of the cellular response.

**Multivariate Analysis**

The loadings from the PCA of the un-derivatized data (Figure 12) are used to monitor the spectral features according to their contribution to the variance in the dataset. The highest variance, describing 68.2% of the overall variance, is represented by PC1, which is dominated by the strong features of the control variable, SWCNTs, as expected. The largest variances related to biological response due to exposure with SWCNTs, are expressed by PC2 and PC3, although they represent only a further 25% variance. Within the first five components, compared to the control cell spectrum (Figure 2), component three shows the most similar features, indicating a defined response at ~1030, ~1300,~1450 cm$^{-1}$ implying a change of spectral variance due to activity in
lipid related bands\textsuperscript{39} corroborating the peak ratio analysis of Figure 6-8. In detail, changes in the region 1230\textsuperscript{cm\textsuperscript{-1}} to 1350\textsuperscript{cm\textsuperscript{-1}} associated with the Amide III band at 1238\textsuperscript{cm\textsuperscript{-1}}, DNA bases guanine, adenine and thymine at 1287 \textsuperscript{cm\textsuperscript{-1}} and 1338\textsuperscript{cm\textsuperscript{-1}}, and lipid deformation modes at 1302\textsuperscript{cm\textsuperscript{-1}}, feature strongly positive in PC\textsubscript{3} and less strong in PC\textsubscript{5}, although multiple features of loadings of a variable in different PCs are complex to interpret.

The scores plot of the PCA shows a degree of separation into two classes between exposed and unexposed populations (Figure 13) in the 2D vector space spun by PC\textsubscript{2} and PC\textsubscript{3}. However, many of the exposed populations group with the unexposed populations, indicative of a non-uniform exposure. It is clear from visual observation at high concentrations\textsuperscript{9} and the variability of the contribution of the G-line as shown in figure 9, that the spatial distribution and thus local concentration of the SWCNTs varies considerably from point to point in the sample at each dose. The separation or distinction between the five different exposure doses is not therefore very clear and a continuous variation of dosages as measured using the high spatial resolution of the laser is inferred. By doubly derivatizing the data, the scores plot of the PCA shows distinct separations down to the level of exposure concentration of the samples, giving a defined cellular response and spatially denser co-localisation of each group (Figure 14). The plot demonstrates a well defined dose dependent response but again highlights the difficulties of establishing completely homogeneous exposure doses.

### Partial Least Squares Analysis

This study uses the same cell populations described in Herzog et al\textsuperscript{12} and the cells under investigation here were identical in exposure to those used in the previous work. Therefore these cells should reach identical clonogenic endpoints. The results of the clonogenic study are given in (Figure 10). These targets have therefore been used in a PLS model with the aim of demonstrating that the technique can potentially be extended to the determination of toxicity through the presentation of spectra of samples of unknown dosage. It is hoped that this work will have advantages in the analysis of the toxic response of SWCNTs, and indeed other nanoparticles.

In total 130 Raman spectra, \~60\% (192 spectra) of the full set formed the training data for construction of the PLS model described in this work. A GA was applied to reduce the number of wavenumbers required for prediction. Table 1 gives details of the Genetic algorithm parameters.
Table 2: Genetic algorithm parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chromosome size</td>
<td>30</td>
</tr>
<tr>
<td>Max. genes per chromosome</td>
<td>30</td>
</tr>
<tr>
<td>Mutation probability</td>
<td>0.01</td>
</tr>
<tr>
<td>Crossover probability</td>
<td>0.5</td>
</tr>
<tr>
<td>Pre-processing</td>
<td>None</td>
</tr>
<tr>
<td>Max LV</td>
<td>15</td>
</tr>
<tr>
<td>#runs</td>
<td>100</td>
</tr>
</tbody>
</table>

The aim of the genetic algorithm was to minimize the RMSECV for the calibration model in predicting the clonogenic endpoints of CNT induced toxicity. The GA was performed over 100 runs; the fittest individuals used 178 variables, reducing the original dataset by 559 wavenumbers. In order to choose the optimum number of LVs to be retained, rigorous cross validation was carried out on the calibration set. (Figure 15) shows the results of the cross validation. Ten LVs were retained for model construction as the RMSECV did not decrease significantly after this point.

The RMSECV was calculated to be 2.10. Table 2 shows the performance of the GA PLS.

Table 3: Performance of PLS and GA optimized PLS. The RMSE for the test set (129 spectra) and training set is shown. 10 latent variables were retained for each model.

<table>
<thead>
<tr>
<th>#wavenumbers</th>
<th>LVs retained</th>
<th>RMSECV</th>
<th>RMSEC</th>
<th>RMSEP</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLS</td>
<td>178</td>
<td>10</td>
<td>4.31</td>
<td>3.17</td>
</tr>
<tr>
<td>GA-PLS</td>
<td>149</td>
<td>10</td>
<td>2.53</td>
<td>2.10</td>
</tr>
</tbody>
</table>

Using 10 latent variables of the GA, PLS clearly outperforms multivariate calibration using the entire wavelength range, showing a decrease in all RMSE values. The independent testing set held back from training was used to determine whether over fitting had occurred. Upon presentation of the testing set, the RMSEP was calculated to be 2.78 indicating an accurate model, and furthermore no over fitting was observed. Figure 16 shows a plot of the predicted toxicity based on the spectral variations of the test set versus the cellular toxicity as measured in the clonogenic study. Therefore, an accurate GA optimised PLS model has been created correlating Raman spectra to clonogenic endpoints thereby potentially reducing toxicity analysis time and the cost of analysis by negating the need for post exposure cell culture. While the RMSE values observed in this study are encouraging, a further reduction in these values would be beneficial. Work is also ongoing toward the inclusion of various exposure time points, greater range of concentrations and the evaluation of Raman spectroscopy for the determination of cellular toxicity mediated by other types of nanomaterials.

Raman spectroscopy can thus be considered as a potential technique for monitoring SWCNT induced biochemical changes at the cellular level. It is hoped that this work will have advantages in the analysis of the toxic response of SWCNTs, and indeed other nanoparticles.

Conclusions

Although the data shown are preliminary, the potential of Raman spectroscopy as a viable tool to assess toxicology is demonstrated. A good correlation is seen between previously identified spectral markers of toxicity and the exposure dose. The study uses the same cell populations described in Herzog et al and the cells under investigation here were identical in exposure to those used in the previous work. A similarly good correlation between the spectral markers and the clonogenic endpoint of proliferative capacity is observed, indicating that the technique can potentially overcome the previously identified problems with colorometric assays in determining the cytotoxicity of carbon nanotubes. The changes in the spectra are visibly observable, dose dependent and associated with cytological data throughout, emphasizing that Raman spectroscopy is a precise analytical method for the examination of chemical and biological properties of cells.

Although the multivariate statistical approaches may appear complex and user unfriendly, they are becoming standard tools of biospectroscopy and increasingly user friendly packages are becoming available. Principal Component Analysis as such shows good dependent separation of spectra. Raman spectroscopy and multivariate calibration via GA optimised PLS for prognosis of cellular toxicity and proliferation shows much promise. The use of GA for variable selection increases the accuracy of the PLS model and reduces the number of factors required. Accurate quantitative prediction of the endpoints of clonogenic assays was possible using Raman spectroscopy and
GA PLS. It is hoped that this work will lead to rapid Raman based methods for the determination of SWCNT toxicity. Further experiments are planned to corroborate these analyses and to conquer the limitations of the model as a result of dispersion inhomogeneity within the SWCNT suspensions.

References