

Raman spectroscopy and support vector machines for quick toxicological evaluation of titania nanoparticles

Georgios Pyrgiotakis,^a O. Erhun Kundakcioglu,^b Panos M. Pardalos^c and Brij M. Moudgil^{a,d*}

With the rapid development of nanotechnology products, there is a significant concern on the adverse effects that might be associated with them. Traditional biological assays are typically used to assess the toxicity *in vitro*. There are, however, questions regarding the suitability of these assays for this purpose, mainly due to the potential interaction of the particles with the utilized dyes. In addition, this process can be costly and time consuming, as a large number of different assays have to be used. To address some of these issues, Raman spectroscopy is used in this study to investigate the particle-cell interactions. The spectrum of a living cell is a very complex and rich collection of data directly related to its chemical composition. To enhance the data resolution and make the detection of toxicity more robust, data-mining techniques have been deployed. Furthermore, data-mining techniques enable full automation of the entire process, minimizing user input. The Raman spectroscopy successfully evaluated the toxicity of TiO₂ nanoparticles by both the peak-by-peak analysis and with the implementation of support vector machines. The particles were found to display cytotoxicity after 36 h of exposure. The results were confirmed by MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-Diphenyltetrazolium Bromide) assay and are in agreement with the existing literature on the subject. Overall, Raman spectroscopy appears to be among the very few techniques that exhibit low levels of interferences (obscurance, fluorescence, emission, etc.) from the particle addition. Since it does not rely on biomarkers, it can be used *in situ* for an extended period with minimal effects on the cellular biochemistry. Copyright © 2011 John Wiley & Sons, Ltd.

Keywords: Raman spectroscopy; support vector machines; titania nanoparticles; cells; toxicity

Introduction

The rapid growth of nanotechnology in the last few years promises great technological advances and further applications of nanomaterials in very diverse fields, including diagnosis and therapy.^[1] The potential for biomedical and environmental applications has raised concern about the fate and toxicological implications of nanomaterials in biological systems. This has driven numerous *in vitro* and *in vivo* toxicity studies over the last few years.^[2] However, interactions between biological systems and nanomaterials are complex events that depend on various factors not yet fully understood. Such factors include the nature of the nanoparticles (e.g. size, shape, composition, surface chemistry, etc.), the cell type, and the environment.^[3,4] Recognition of cellular death pathways (e.g. inflammation, apoptosis, and necrosis) is directly linked to toxicity studies, as death is the ultimate response to a toxic event unless mitigated by other bioprocesses. Being able to monitor cell death and identify its different mechanisms is an important step in understanding the pathological processes induced by the interactions of cells with nanoparticles. In recent literature, cell death has been classified into three distinct forms: apoptotic (or type I programmed), autophagic (or type II programmed), and necrotic (or oncotic) cell death.^[5] Apoptosis is marked, among other indicators, by caspases activation, chromatin condensation, formation of apoptotic bodies, cell shrinkage, and nucleus fragmentation.^[6,7] Autophagic death is marked by the engulfment of organelles and particles inside separate compartments within the dying cell. Although it can be considered as a subdivision of the apoptotic death, it is argued that autophagic

death can be a non-apoptotic pathway.^[5] Necrosis is indicated by organelle swelling with the eventual loss of plasma, reduction of membrane integrity, and subsequent inflammation.^[5,8]

Except for cell morphology, there is no single *in vivo/in vitro* assay that can unambiguously identify the cell death type (if such an absolute delineation can be drawn), which makes characterizing cell death very difficult.^[9] Although a wide variety of biochemical methods (e.g. MTT, neutral red uptake, lactate dehydrogenase (LDH) assays, etc.) are available for measuring cells' biological responses, these conventional techniques all have their respective shortcomings and are not fully suitable for nanoparticle toxicology studies. Available assay kits with the ability to detect and differentiate cell death processes are destructive, time consuming, and expensive, and, most importantly, have very poor time resolution. Furthermore, many factors need to be considered when interpreting the results. For instance, since cytotoxicity assays rely on chemicals and biomarkers, problems may arise due to

* Correspondence to: Brij M. Moudgil, Particle Engineering Research Center and Materials Science & Engineering, University of Florida, Gainesville, FL, USA. E-mail: bmoudgil@perc.ufl.edu

a Particle Engineering Research Center, University of Florida, Gainesville, FL, USA

b Industrial Engineering, University of Houston, Houston, TX, USA

c Industrial & Systems Engineering, University of Florida, Gainesville, FL, USA

d Materials Science & Engineering, University of Florida, Gainesville, FL, USA

undesired interferences and interactions between the biomarkers and particles. In addition, for assays dependent upon enzymatic reactions (e.g. MTT, LDH), results may be skewed by promiscuous enzymatic inhibitors.^[10] Especially in the case of nanoparticles, there are additional limitations, such as the adsorption of reagents on the particle's surface, preferential catalysis, or hindrance of the reaction on the particle surface. Kanduc *et al.* compared many of the conventional cytotoxicity assays and found that the reported viability of treated cells differed depending on the assay used.^[11] Specific protocols that can increase the reliability of these assays and make possible the evaluation of particle toxicity through the conventional biological assays have been empirically developed. However, the two major limitations that still remain are sensitivity and time resolution. The required length of time for most of these assays is in the order of hours (typically anywhere between 0.5 and 4 h). Interactions in the same or shorter time frame cannot be monitored. Moreover, events occurring in seconds or minutes will not usually be detected by these assays, masking important information at the cellular level that would help explain the interaction mechanisms between the cells and nanoparticles.

Raman spectroscopy has emerged as an alternative method to investigate the particle–cell interaction in real time, without the aid of any chemicals or biomarkers.^[12,13] The spectrum of a cell is a representation of its chemical composition and provides insight into the biomolecular changes resulting from cellular processes. There are four main advantages of using Raman spectroscopy: (1) The technique is noninvasive and does not require additional reagents or markers; (2) Measurements can be made rapidly, in real time, *in vivo* or *in vitro* and noninvasively; (3) It can be focused on a single cell or averaged over the entire population; (4) The asymmetric nature of water gives a weak Raman signal that, unlike in infrared spectroscopy, does not interfere with the spectrum of interest. Raman spectroscopy has already been used for identifying bacteria^[14] and eukaryotic cells^[15] and has been investigated as a toxicity evaluation tool.^[16] In the past, we have demonstrated the ability to resolve potential particle–cell interactions^[17] by assessing toxicity and shown that the technique is indifferent to the cell line used.^[17,18] It is among the very few techniques that result in only a small interference (if any at all) from the particle addition. Since it does not rely on biomarkers, it can be used *in situ* for a very long period without affecting the cellular biochemistry. The Raman spectrum of a living cell is a very complex and rich collection of data directly related to its chemical composition. Data-mining techniques can enable the utilization of all spectral features and fully automate the entire process by minimizing user input. This results in a tool that can accurately and reliably detect and classify cellular processes and death directly linked to the toxicity of nanomaterials.

One of the limitations of the Raman spectroscopy technique is the data interpretation, which can be very tedious and time consuming. Although many researchers have employed new statistical techniques such as principal components analysis to cluster and analyze the data, the information is still not fully extracted and human input is required. To overcome some of these limitations, an alternative method has been employed that utilizes data-mining techniques such as support vector machines (SVMs). For scientific experiments that generate an immense amount of complicated and convoluted data, *machine learning techniques* have recently emerged as a prominent tool that can be used to analyze the data and derive conclusions. *Supervised learning* refers to the capability of a system to learn from a group of examples, which is a set of input (i.e. *vector of features*)/output (i.e.

label) pairs. A set of objects with an input and the corresponding output is called a *training set*. The labels for the objects are provided in the training set. These labels have to be determined by an outside source, e.g. the user, usually referred to as the *supervisor*, hence the term *supervised learning*. These sets can be used to derive classification or regression functions and, in that way, conclude the cell death type. In a previous publication, we showed the enhancement of Raman spectroscopy via deployment and development of data-mining techniques to assess the cellular death type.^[19] Since cell death type is linked to the toxicological impact, it can make the detection of toxicity more robust. The methodology utilized in this study is the use of SVMs. Considering that each spectrum is a vector in the hyperspace, SVMs will divide the hyperspace into two regions with a hyperplane (linear case) or hypersurface (nonlinear case). A new unknown spectrum will now be classified based on its relative position to the plane. Although similar approaches have been used in radically different tissue samples by Widjaja *et al.*,^[20] in our approach we distinguish among states of the same cells. The results are also confirmed by an appropriately modified MTT assay.

Materials and Methods

In this section, the main techniques, instrumentation, and materials used in the present research endeavor are described.

Anatase nanoparticles

The nanoparticles were purchased from Sigma Aldrich (Lot #: 637254) and, according to the manufacturer, the primary particle size is 25 nm. They were measured with the NanoTrac (Microtrac, Montgomeryville, PA), and transmission electron microscopy (TEM) images were used to investigate the primary particle size.

Cell line

Since one of most common pathway points for particles is through the lungs, for this set of experiments A549 lung epithelia cells were used (from ATCC; cell line number CCL-185). This is a hypo-triploid human cell line with the modal chromosome number 12, which occurs in 24% of cells. The line was initiated in 1972 by D. J. Giard *et al.* through explants culture of lung carcinomatous tissue from a 58-year-old Caucasian male.^[21,22]

Cell culture protocols and plating

The growth medium was made by 89% RPMI-1640 with L-glutamine (from Cellgro; Cat #: 25-053-CL), 10% fetal bovine serum (four times filtered through 0.1 μ m filter, from Hyclone; Cat. #: SH30070.03), and 1% antibiotic-antimycotic solution (from Cellgro; Cat. #: 30-004-CL). The ingredients were mixed and passed through a 0.4- μ m filter. The cells were grown with complete growth medium in a 25 cm² cell culture flask at 37 °C and 5% CO₂. In order to harvest the cells for various experimental applications, the culture medium was removed and the cells were rinsed with 1 \times Hank's Balanced Salt Solution (HBSS), without Ca⁺² or Mg⁺² (from ATCC; Cat: 30-22/3), to remove all traces of serum containing the trypsin inhibitor. Next, 1 ml of 0.25% trypsin/2.2 mM EDTA in HBSS, without sodium bicarbonate, calcium or magnesium (from Cellgro; Cat. #: 25-053-CL), was added to the flask in order to detach the cells from the bottom of the flask. The process was observed under an inverted microscope until at least 80% of the cell layer

was detached (usually within 2–5 min). Once the majority of the cells (80% and above, visually estimated) have detached, 5 ml of complete growth medium was added to deactivate the trypsin, and the cell suspension was then centrifuged for 3 min at 125 g. The supernatant was removed and 1 ml of fresh complete medium added. Then the cells were gently aspirated, resulting in a homogeneous solution of $\sim 2\text{--}3 \times 10^6$ cells/ml and viability ratio above 95%, as measured by the ViCell from Beckmann-Coulter (Fullerton, CA). This solution was used for seeding with approximately $\sim 5 \times 10^3$ cells on a 5×5 mm MgF_2 substrate (custom-made by Red-Optronics) used in the Raman experiments. The MgF_2 crystal was used to reduce the background radiation from the Petri dish. The seeded MgF_2 substrates were placed on a six-well plate (9.6 cm² per well) in an incubator for 45 to 60 min, which was sufficient for the cells to attach onto the MgF_2 . Three milliliters of the growth medium was added and the seeded cells were incubated at 37 °C and 5% CO₂. For 24 h, the particles or toxins were administered to the cells according to the protocol described in later sections.

The process is similar for the 96-well plates used in executing the assay. In this case, the final cell suspension, harvested from three 75-cm² flasks, was adjusted to 10⁵ cells/ml. The solution was used to seed the cells in the 96-well plates with 100 μ l of media and approximately 10 000 cells in each well. The cells were allowed to adhere for 12–18 h before the particles were administered to the cells.

Toxic agent standards

The triggers for the two different cellular deaths were etoposide (apoptosis) and Triton X-100 (necrosis). Etoposide is a strong chemotherapeutic drug used as a form of treatment for malignancies such as Ewing's sarcoma, lung cancer, testicular cancer, lymphoma, non-lymphocytic leukemia, and glioblastoma multiforme. Etoposide is known to form a ternary complex with DNA and topoisomerase II causing double stranded DNA breaks, which is one of the defined mechanisms of apoptosis.^[23] It has also been shown to upregulate p53, an initiator of apoptosis.^[24,25] Triton X-100 is used as a benchmark in various assays because it can rupture the cellular membrane, resulting in the necrotic death of the cell. Triton X-100 exposure has been reported to increase the expression of apoptosis inhibitors and is known to solubilize and destabilize the cell membrane.^[26]

The toxin concentrations were selected based on the literature values, which suggest that these values will impact the cells without leading to complete cell lysis in the first 24 h of the experiment. For these experiments, the agents' concentrations were 100 μ M for Triton-X^[27] and 80 μ M^[7,23,28] for etoposide. The solution was prepared immediately prior to dosing by mixing a stock solution of 100 mM. Etoposide is insoluble in water, so a stock solution was prepared with 100 mM etoposide in dimethylsulfoxide.

Particle and toxin dosing protocols

The particles were used as received from the manufacturer. One hundred milligrams of the particles was added in 10 ml of nanopure water (18.2 M Ω) and was sonicated for 15 min in an ultra sonicator (15 s sonication, 10 s break). The suspension was then autoclaved at 121 °C for 30 min and the particles were again sonicated briefly for 3 min in a water bath sonicator before use. The particles at this stage were used for dosing and the various characterizations.

Since different experiments require different cell vessels, the dosing remains constant with respect to the amount of particles

(mass) per surface area of the well ($m_{\text{TiO}_2}/S_{\text{well}}$) and the amount of the medium per well surface area ($m_{\text{media}}/S_{\text{well}}$). Uptake of the particles by cells is shown to occur via endocytosis.^[29]

Particle dosing protocols for the MTT assay

The same stock solution as before was used to prepare a final solution of 1.28 mg ml⁻¹ in the medium. The initial stock solution was consecutively diluted in centrifuge tubes to reach final concentrations of 640, 320, 160, 80, 40, 20, 10, and 5 μ g ml⁻¹. After the elapse of the required time for cell attachment, 50 μ l of the medium was removed and 50 μ l of the appropriate particle solution was added. The assay was developed with the final concentrations of particles of 640, 320, 160, 80, 40, 20, 10, 5 and 2.5 μ g ml⁻¹.

Biological assay – MTT

There are no viability assays available to specifically address and investigate particle-related toxicity. Thus, in the present research, the widely used MTT assay was utilized. The MTT was designed to measure metabolic dysfunction and mitochondria activity; therefore, in this experiment, it was developed and read in the same well with the cells. The assay protocols could be used as such, but the particles can act as obscurants, increasing the adsorption and altering the results. For the present study, the protocol was modified to accommodate the particle presence. More specifically, after the formazan crystal dissolution, the cell plates were centrifuged, and 100 μ l of the supernatant was obtained and used for the assay reading. The assays were executed in 96-well plates with eight replicates of each sample.

MTT

The MTT assay used was the #TOX-1 (7H258) kit from Sigma-Aldrich. The assay ingredients were prepared according to the manufacturer's instructions. The cells were plated in the 96-well plate, with approximately 10 000 cells/well (32 000 for 100% confluency) in 100 μ l of the full medium, and were left in the incubator for 24 h to attach to the bottom of the well. Fifty microliters of the medium was removed and another 50 μ l of fresh medium containing the particles (or the Triton-X 100) was added. The cells were left in the incubator for a designated amount of time (12, 24, and 36 h). Before the development of the assay, the medium was removed from the wells and 100 μ l of phosphate buffered saline (PBS) was added, as phenol red can interfere with the assay. Following that, 10 μ l of the M-5655 (MTT in serum) was added. The cultures were then returned to the incubator for 3 h for the formazan crystals to form. Following the incubation, 100 μ l of M-8910 solution (MTT solubilization solution: 10% Triton X-100 with 0.1 N HCl in anhydrous isopropanol) was added and the plates were then aspirated and placed on a shaker for 30 min to dissolve the formazan crystals. Once they were fully dissolved, the plates were centrifuged at 250 rpm for 10 min to ensure that the titania particles were not in the solution. Finally, 100 μ l of the suspension was moved to a new 96-well plate to be read at 570 nm with reference at 690 nm.

Particle dosing protocols for the Raman experiment

For Raman spectroscopy, a particle concentration of 20 μ g ml⁻¹ was selected because it was determined to have a significant effect (by the MTT assay) on the cells without interfering with the acquired spectrum. The appropriate amount of the stock solution

was added to the full medium to generate final solutions of 60 g ml^{-1} . After reaching $\sim 80\%$ confluency on the MgF_2 plate, the growth medium was removed and the cells were rinsed twice with HBS to remove any traces of proteins. Next, the cells were rinsed with the full medium twice to ensure that any remaining traces of HBS would not be significant enough to alter the final concentration of growth medium. Finally, 2 ml of the medium was added and the cells were transferred in the incubator where they were allowed to recover for 4 h. After the 4 h of incubation, 1 ml of the full growth medium containing the particles at concentrations 60 g ml^{-1} was added to reach the final particle concentration of 20 g ml^{-1} , with a final growth medium volume of 3 ml.

The dosing of the toxins utilized to simulate the various cell death mechanisms that was used for the machine learning techniques database was similar to the particle dosing. Prior to each experiment, the medium was removed, the cells were rinsed with HBS twice, and 2 ml of fresh RPMI 1640 was added so that the proteins in the serum did not interfere with the obtained spectrum. The absence of proteins was determined not to have any measurable effect on the cells over the experimental period ($\sim 1 \text{ h}$).^[30] The RPMI 1640 provides all the necessary growth hormones and sugars essential for the cell viability.

Raman spectroscopy protocols and procedures

The Raman microscope used is the InVia system by Renishaw, consisting of a Leica microscope attached to a Renishaw 2000 spectrometer. When used with the $63\times$ lens (48 mW), the high-power diode laser (250 mW) produces laser light of 785 nm and does not cause any damage to the cells, even after 40 min of exposure time. The MgF_2 plate, after being rinsed with the PBS, is moved onto a Delta T culture dish (from Biotech; Cat #: 04200415C), and 2 ml of RPMI 1640 is added. The dish is placed on a heating stage (Delta T4 Culture Dish Controller, Biotech, Butler, PA, USA) in order to maintain 37°C throughout the entire measurement. Through the Leica microscope, the laser is focused at the center of the cell with the crosshair. The spot size is $20 \times 40 \text{ }\mu\text{m}$ when focused on a dry Si wafer and $20 \times 30 \text{ }\mu\text{m}$ when in a water-based solvent. Given the fact that the intensity profile of the beam is a Gaussian, it can be assumed that the laser spot covers one cell at a time ($\sim 20 \times 20 \text{ }\mu\text{m}$ when 80% confluent, $\sim 40 \times 40 \text{ }\mu\text{m}$ when isolated). However, for the isolated cells, the relative position of the laser can potentially affect the spectrum. Thus, for all the isolated cells used, the position of the spot was kept the same in relation to the cell. The 785-nm laser beam passes through the $63\times$ water immersion Leica objective and the final output power is 48 mW.

Raman spectroscopy peak-by-peak analysis

The obtained spectra had to be baseline-corrected and normalized in order to make them comparable. The data processing was performed using a custom-made mathematical code written in Mathematica (Wolfram Research, USA), which has been previously described by Pyrgiotakis *et al.*^[17] and Bhomwick *et al.*^[18] First, the background from the cell medium was subtracted using nonlinear subtraction according to the method developed by Maquelin *et al.*^[31] For this, a spectrum was obtained from a cell-free region on the MgF_2 plate while the focus was still on the cell plane. This spectrum was used as the medium background. Following that, the spectra were treated with the algorithm developed by Lieber and Mahadevan-Jansen to remove the baseline.^[32]

Next, the spectra were normalized with respect to the 1450 cm^{-1} peak originating from the CH_2 bond-stretching from all the cell components, which is assumed to remain constant over the measurement time. It represents a bond that is not susceptible to the titania nanoparticles or the various toxins. Thus, it can be assumed constant during the time of the experiment. A second normalization process follows using the standard normal variate method to further normalize the spectra and eliminate any intensity differences that might have been induced by the previous process. The processed spectra were then averaged and the standard deviation was calculated. The average is considered a representative spectrum of the cells at a certain point in time, and the standard deviation is the error of the measurement. The steps of the aforementioned algorithm are optimized to achieve minimum deviation between the fitted curves and the experimental data. This process was monitored via the χ^2 and R^2 factors. In addition, the fitting algorithm should converge to a minimum or a local minimum.

Assignment of the peaks was done according to Nottingher *et al.*^[27,33] (Table 1). The deconvolution of the spectra was done through curve fitting, and the area under the individual peaks is computed using the GRAMS software (ThermoGalactic, USA). The peak position, ω_0 , was set within a span of 4 wavenumbers ($\omega_0 \pm 2 \text{ cm}^{-1}$) and the full-width at half-maximum (FWHM) of each peak was constrained at 30 wavenumbers. The relative intensities were unconstrained. The profile of the peak was Gaussian and the baseline was assumed linear (second and third degree polynomials were attempted with no significant difference in the result). The curve-fitting was carried out until convergence or a local minimum was reached. The area under the curve is used as the intensity of the peak. The same process was repeated for the standard deviation spectrum and the resultant areas are used as the error of the peak intensity measurement.

Support vector machines

SVMs, developed by Vapnik,^[34] are considered as one of state-of-the-art supervised machine learning methods. SVMs have a wide range of applications ranging from pattern recognition,^[35] text categorization,^[36] biomedicine,^[37–41] and cell death discrimination^[19] to brain–computer interface^[42,43] and finance.^[44,45] The theoretical strength of SVM classifiers comes from the fact that they are *universally consistent*.¹ The basic principle of SVM classifiers with two classes is to maximize the distance between the separating hyperplane and the closest instances of both sets. In order to avoid infeasibility due to outliers, a *goal programming* approach is employed by penalizing points that are on the wrong side of their *supporting hyperplane*. Note that SVMs can also be extended to nonlinear classification by using dual formulation and replacing the dot products for input vectors with a suitable *kernel function*. This is equivalent to implicitly embedding the original data into a nonlinear space using a kernel trick.^[46]

In our experiments, only primal SVM formulations (i.e. linear classifiers) are used. A leave-one-out cross-validation technique^[47] is employed to estimate the generalization performance of our classifiers. To visually illustrate the employed cross-validation technique, the functional distance of each instance from the

¹ A classifier is consistent if the probability of misclassification converges in expectation to a Bayes optimal rule as the sample size is increased. A classifier is universally consistent if it is consistent for all distributions of data.

Table 1. The main Raman peaks representing the cells

Spectrum region	Peak location	Assigned peaks for the molecules			
		DNA/RNA	Protein	Lipid	Carbohydrates
770–803	782	U, C, T ring br. ^[52–54]			
	788	O–P–O ^[52–55]			
995–1020	1005		Sym. Ring br. Phe ^[53,56–58]		
	1013	C–O deoxyribose ^[53]			C–O str. ^[53]
1020–1150	1033		C–H in plane Phe ^[54,59]		
	1049				C–O str. ^[53]
	1066		C–N str. ^[60]	Chain C–C str. ^[60–63]	
	1080		C–N str. ^[54]	Chain C–C str. ^[59,60]	C–O ^[60]
	1095	PO ₂ [–] str. ^[56,59]		Chain C–C str. ^[61,62]	C–C str. ^[60]
	1128		C–N str. ^[59]	Chain C–C str. ^[54,60–62]	
1190–1385	1209		C–C ₆ H ₅ , Phe, Trp ^[54,59]		
	1231		Amide III, random coil ^[64]		
	1242		Amide III, β -sheet ^[64]		
	1258		Amide III, β -sheet ^[64]	=CH def. ^[60–62]	
	1284		Amide III, α -helix ^[64]	=CH def. ^[60–62]	
	1301			CH ₂ Twist ^[59,60,62–64]	
	1320	G ^[56]	C–H def. ^[52]		
	1342	A, G ^[52,55,56,59,63]	C–H def. ^[52,56]		
	1367			Sym. Str. CH ₃ ^[62]	

The spectrum is divided into four major areas. The recognized peaks are those that have minimum overlap with other peaks and provide a very well defined contribution to the cell spectrum.

O–P–O, phospho-di-ester; PO₂[–], phospho-di-oxy; Phe, phenylalanine; Trp, tryptophan; A, adenine; G, guanine; C, cytosine; T, thymine; U, uracil; Sym, symmetric; str, stretching; def, deformation.

separating hyperplane, which is obtained without that instance in the training set, is presented.

For this application, the data is used in its raw form without any further processing. The baseline, fluorescence, and intensity range of the spectra, although vary from case to case, are characteristics of the cellular state. Therefore removing them will reduce the convoluted spectral information.

Results and Discussion

Titania nanoparticle characterization

When suspended in water, the average particle size was 92.80 ± 26.93 nm (Fig. 1(a)). The particles were also characterized in PBS and in the full medium. In PBS, although there are no ions present since the pH is close to the isoelectric point of TiO₂ (pH at 5.8), the particles were expected to be somewhat aggregated. In the full medium, in addition to the pH, the proteins further increase the aggregation. The Nanotracer analysis also revealed particle sizes of 262.10 ± 40.80 nm (in PBS, Fig. 1(b)) and 301.00 ± 62.30 nm (in medium Fig. 1(c)). The TEM results showed that the particle size is 50 nm and of irregular shape (Fig. 1(d)).

MTT assay

The MTT assay shows the cytotoxicity of the particles, which is in agreement with other reports claiming that titania nanoparticles are potentially toxic.^[10,48–50] Although the assay cannot reveal the exact cell death pathway, it is observed that the toxicity is dose and time dependent (Fig. 2). The concentration used for the Raman experiment is 20 g ml^{-1} and is highlighted in the graph. It

should be mentioned here that LDH assay was also used, but it was inconclusive. The results were not only inconsistent with regard to the dose and exposure time, but it also seems that there was a significant interference with the titania nanoparticles on the LDH reaction (reaction catalysis). Although the observation was not further investigated to assess the reason for the incompatibility, it highlights the inadequacy of conventional techniques in understanding nanoparticle cytotoxicity.

Raman spectra analysis

For both methods of analysis, the same spectra were used. The spectra, in the case of the peak-by-peak analysis, were preprocessed as described below. In the case of the SVMs, the data were used in their raw form. In total, 15 spectra were selected from each time point and case. Although the peak-by-peak analysis revealed certain trends in the DNA/RNA, the protein concentration was inconclusive due to the measurement error (standard deviation). To address this issue, machine learning techniques were employed. For this attempt, SVMs were used to compare the obtained results to a database created with Triton-X 100 and etoposide.

Peak-by-peak analysis

Figure 3 shows the spectra after they have been processed with the appropriate algorithms described in the Methods section. This step is required to ensure the same base of comparison. It is obvious that the spectra are drastically different from each other, especially when these results are compared with those of previous reported work.^[17,18] This holds true not only for the different time

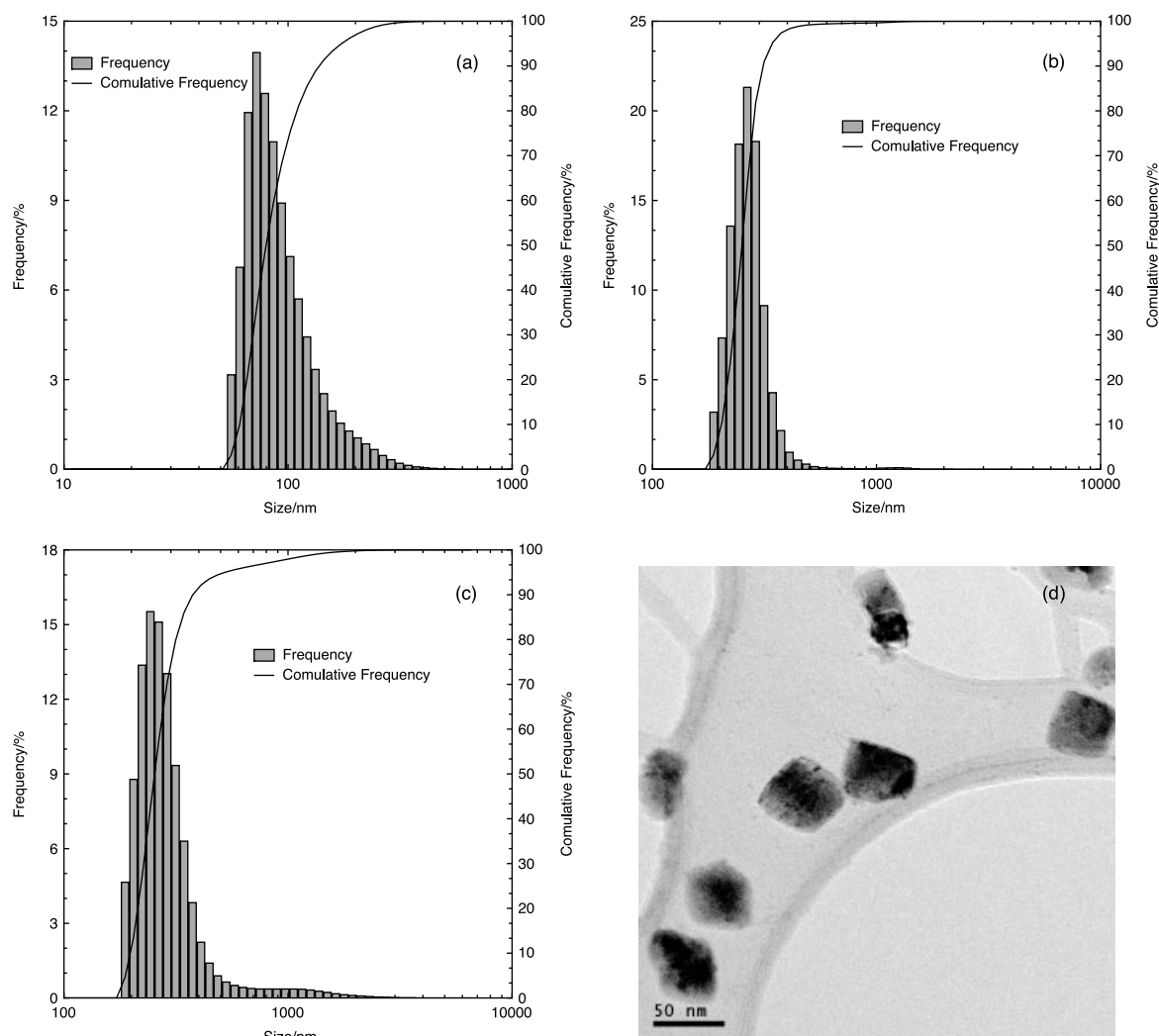


Figure 1. Particle sizing of the anatase nanoparticles as measured by NanoTrac™. (a) Particle size distribution of the anatase TiO₂ in water (average particle size is 92.80 ± 26.99 nm). (b) Particle size distribution of the anatase TiO₂ in phosphate buffered solution (PBS) (average particle size is 262.10 ± 0.80 nm). (c) Particle size distribution of the anatase TiO₂ in full medium (average particle size is 301.00 ± 62.30 nm). (d) TEM image of the titania particles.

points but also for the anatase-treated cells *versus* control cells. This indicates that the potential toxicity of the particles should be quite low at this given concentration, which is in agreement with the MTT assay that suggests results along these lines (Fig. 2).

Figure 4(a) shows the changes in the intensities of the most important cell components for cells exposed to anatase nanoparticles. Figure 4(b) shows the same components for the control cells. The analysis is executed according to the protocols described in the Methods section. The comparison is done by assessing the intensities of the control cells against the anatase-treated cells.

The 782 cm⁻¹ peak: This peak represents the ring breathing mode of the DNA bases cytosine and thymine and the RNA base uracil. The damping of this vibration over time denotes the intensity reduction that originates from the destruction of the ring structure, indicating degradation of the DNA. The peak for the control cells does not show any significant alteration over time, but for the anatase-treated cells there is a continuously decreasing trend. This is an indication of DNA/RNA degradation.

The 788 cm⁻¹ peak: This peak represents the phospho-di-ester bond ···O–P–O··· the backbone of the DNA/RNA. As shown in

Fig. 4(a), this peak follows the same trend as the previous peak, i.e. decreasing over time. Although it is notable that the changes occur within the error of the measurements, there is a distinct trend which indicates the breaking of the DNA/RNA. This is a very important indication since the breaking of this bond is associated with DNA laddering, which is a very specific characteristic of apoptosis. An identical peak for the control cells remains constant over the same time period.

The 1005 cm⁻¹ peak: This peak represents the phenylalanine ring, which is a very important building block for proteins. For both the control and the anatase-treated cells, it appears unaffected. In the literature, it was shown that this peak is very sensitive to the necrotic death of cells.^[17,18,27,28,33,51] This seems occur irrespective of the source of the toxicity, since it was confirmed both in the case of toxins^[17,27,28,33] and particles.^[17,18]

The 1095 cm⁻¹ peak: This peak represents the PO₂⁻ bond that is used to attach the various bases through the deoxyribose sugar onto the DNA/RNA backbone and some protein components. However, the protein signal has been shown to be extremely weak,^[13] and so most of the contribution comes from the DNA. This component, as before, appears to be reduced over time

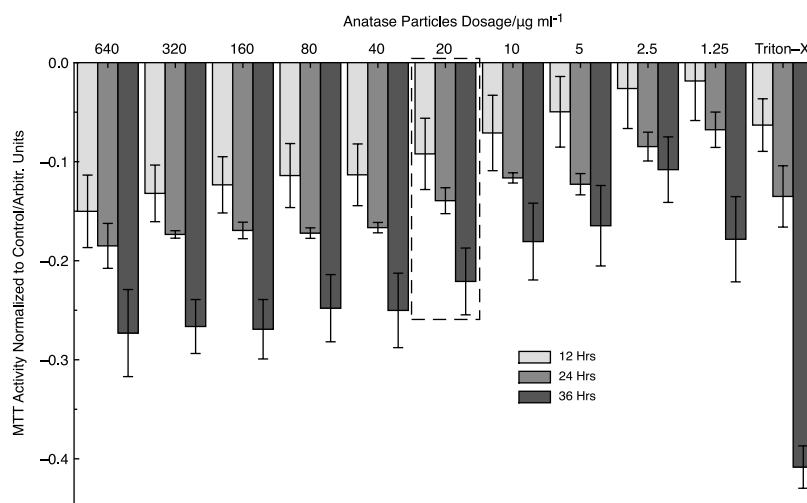


Figure 2. Results from the MTT assay normalized with the control. The cytotoxicity appears to be both dose and time dependent. Highlighted is the dose concentration that is used for the Raman spectroscopy.

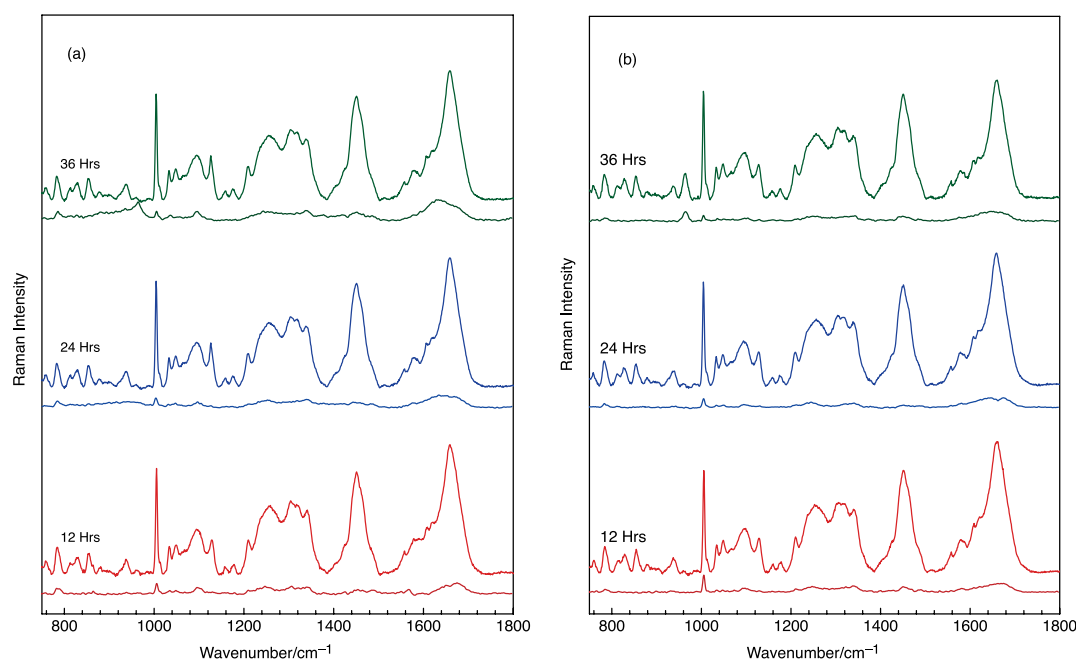


Figure 3. Raman spectra averaged from the cells with their error. (a) Anatase-exposed cells. (b) Control cells.

and, in this case, the trend is not obscured by the error of the measurement. The band for the control cells remains unaffected.

The 1231, 1242, 1258, and 1284 cm^{-1} peaks: These peaks represent amide III, which is a basic component of the protein structure and is extremely sensitive to the structure of the protein. The 1231 cm^{-1} represents the random coil, the 1242 cm^{-1} and the 1258 cm^{-1} the β -sheets, and the 1284 cm^{-1} the α -helix. As can be seen from the figures, control cells have a small but steady trend. However, the cells that have been exposed to the anatase nanoparticles show a more drastic conformational change. Although this can denote denaturing, this information is inadequate for a conclusive assessment, since this information is likely to relate to other biological processes as well.

When the information from the peak analysis is collectively examined, it is observed that the DNA/RNA components are being reduced, especially their backbone, while the proteins do not

change significantly other than structurally. This is a significant indication that the cells exposed to the anatase particles undergo apoptotic death, characterized by the DNA/RNA fragmentation but with no significant degradation of the protein. This is in agreement with the literature, which reports DNA laddering and breakage when A549 cells are exposed to nanosized anatase particles.^[48,49]

Support vector machines

The successful implementation of SVMs for the interpretation of Raman spectra has been demonstrated before^[19] for three-class classification by cross-examining the three cases. Similarly, in this approach the idea is to cross-examine the cell spectra with the two most common cell death pathways: apoptotic and necrotic. There are three datasets of 20 cells: (1) based on Triton-X100 to

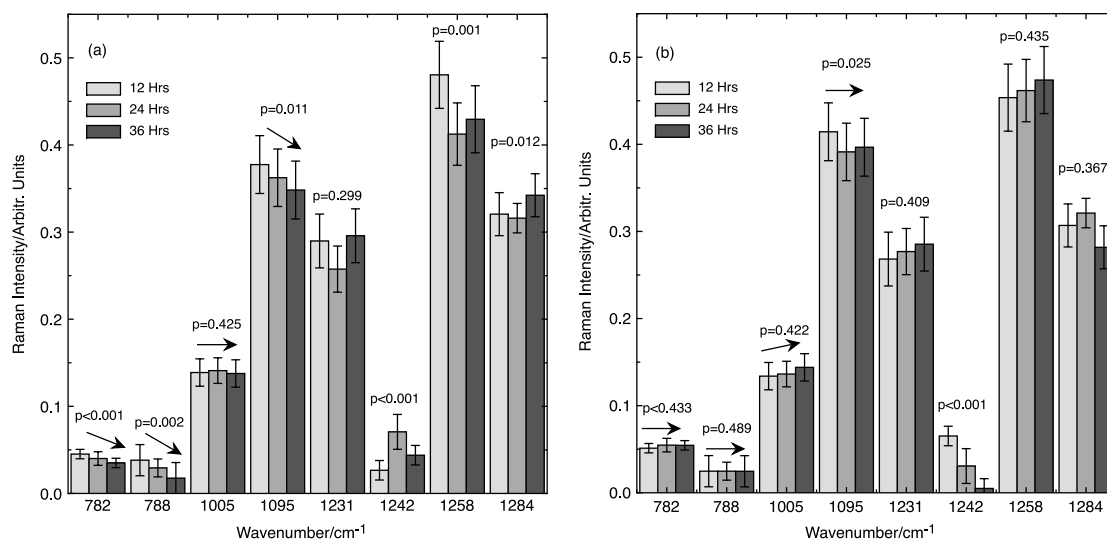


Figure 4. Evolution over time for the most important cellular components. (a) Anatase-exposed cells. It shows the degradation of the DNA/RNA components while the protein components remain invariant. (b) Control cells. The components show only minute changes remaining virtually invariant.

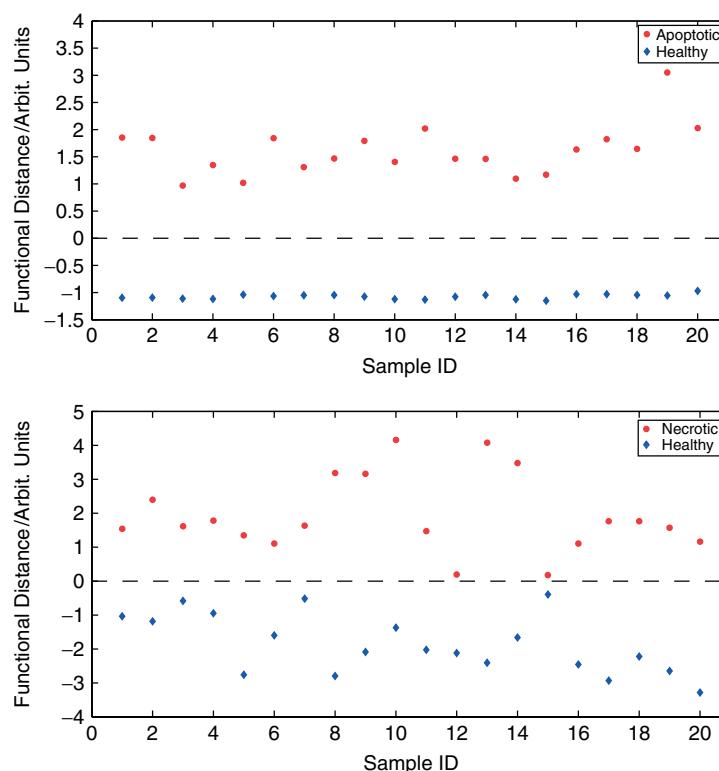


Figure 5. Cross-validation results with leave-one-out method. Although both classifiers have 100% accuracy, apoptotic *versus* healthy classifier seems more promising in terms of generalization ability.

simulate the necrotic death, (2) based on etoposide to simulate the apoptotic death, and (3) healthy control cells. All data instances consist of 1140 features, which represent different frequencies of the spectra.

There are two sets of data that were used for *training*. One training set consists of apoptotic and healthy cells, whereas the other consists of necrotic and healthy cells. The leave-one-out cross-validation method was employed, where 1 instance out of 40 is left out for validation and the remaining 39 for training. This process was repeated for 40 points in the set, and the functional

distances were computed (Fig. 5). Cross-validation reveals 100% accuracy in both training sets classifying all 40 instances correctly. This also justifies that a linear SVM classifier can classify both training sets with promising generalization ability. However, it should be noted that the apoptotic *versus* healthy classifier seems more promising in terms of generalization ability since the absolute distances are larger on the desired side.

Next, SVMs identify two hyperplanes separating the two complete *training sets* of 40 instances each. Based on these classifiers, the goal is to observe how instances in the *test set*

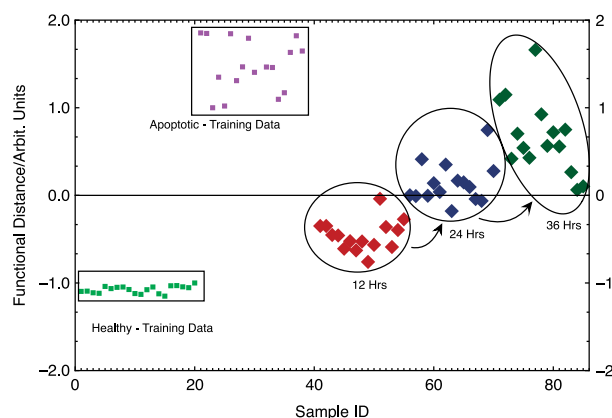


Figure 6. Clustering of the data points as they are grouped among the apoptotic and healthy cells. Again, as in the previous case, there is a progression from healthy to apoptotic along the course of 36 h. At 12 h, most of the data points are clustered in the healthy region (average distance = -0.458 ± 0.175). At 24 h, the points are scattered in both regions (average distance = 0.139 ± 0.232). Finally, at the 36-h mark, most of the data points are already at the apoptotic region (average distance = 0.663 ± 0.421).

(i.e. anatase-treated cells) behave within the context of necrotic *versus* healthy and apoptotic *versus* healthy. Labeling is done for each classifier by computing the functional distance between test instances and the separating hyperplanes. The larger the distance, the more definitive the classification.

Apoptotic–Healthy: The classification between apoptotic and healthy cells is in agreement with the previous observations (Fig. 6). At 12 h, the cells appear to be classified mostly as healthy, with one sample very close to the separating hyperplane. At 24 h, the cells are slowly moving from the healthy side to the apoptotic side. There are, however, still some cells that are classified as healthy. At the 36 h mark, the cells have completely crossed over to the apoptotic side. The average functional distance from the separating hyperplane follows a distinct trend; at 12 h, most of the data points are clustered at the healthy region (average functional distance = -0.458 ± 0.175 a.u., arbitrary units); at 24 h, the points are scattered among both regions (average functional distance = 0.139 ± 0.232 a.u.); and at the 36 h mark, most of the data points are already at the apoptotic region (average functional distance = 0.663 ± 0.421 a.u.).

Necrotic–Healthy: The classification according to necrotic *versus* healthy cells did not reveal any particular trend in the anatase-treated cells (Fig. 7). Furthermore, the average functional distance from the separating hyperplane displays no particular tendency. The average functional distances are 12 h: 2.867 ± 5.0863 ; 24 h: -0.022 ± 2.743 ; and 36 h: 1.787 ± 4.738 . Since the functional distances from the separating hyperplane are scattered along both regions, it is safe to conclude that the linear SVM classifier does not shed light on the necrotic death pathway. Alternatively, it can be hypothesized that the cells are indeed dying through the necrotic pathway and that there are always a certain number of cells undergoing necrotic death. However, such an assumption will contradict the MTT observation, which showed toxicity is a function of the time after dosing.

Conclusions

A novel (Raman spectroscopy-based) technique to assess nanoparticle toxicity is presented and evaluated. Among the most distinct

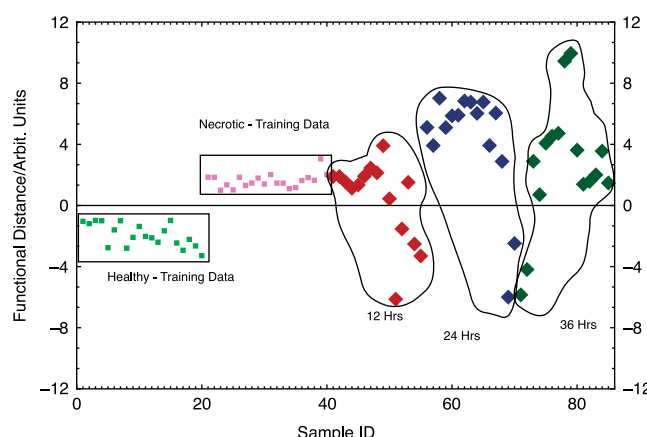


Figure 7. Clustering of the data points as they are grouped among necrotic and healthy cells. This result is inconclusive since the distances from the separating hyperplane are scattered along both regions. The average distances are 12 h: 2.867 ± 5.0863 ; 24 h: -0.022 ± 2.743 ; and 36 h: 1.787 ± 4.738 . This result however is not consistent with the MTT assay.

features of this method is the fact that it does not require any special chemicals or markers, making it completely noninvasive and allowing for the monitoring of cell populations for an extended period without altering the cells or their environment.

In the present study, this technique is used to investigate the effect of anatase nanoparticles (primary particle size < 25 nm). The initial data analysis is done with traditional peak-by-peak analysis, which reveals changes in the various cellular components individually. The peak-by-peak analysis shows that the particles can potentially induce apoptosis to the cells over the course of 36 h. However, among the most important drawbacks of the peak-by-peak analysis is the intensive amount of human input required to operate and supervise the mathematical analysis, as well as the lengthy processing time (4–5 h post data acquisition). In addition, since the focus is on the peaks that present low interference or overlapping, much of the information is lost.

To address these issues and increase the spectral resolution, a novel method utilizing the entire spectrum and determining the state of the cells based upon the similarities/differences of the examined spectra *versus* an established database is proposed. In the current study, the SVM technique is utilized. SVMs, generally referred to as machine learning, use a database to categorize the obtained data. In this particular case, the classification intends to identify the cell mortality type, by distinguishing among apoptotic *versus* healthy cells and necrotic *versus* healthy cells. The utilized database includes living healthy cells, necrotic death induced from Triton-X100, and apoptotic death induced by etoposide. The results again confirm the previous conclusions and indicate the evolution of the cell's health, slowly progressing from fully healthy to apoptotic. Although an extra step for the creation of the database is required, the resolution and reliability of detection increase, since in this case the cell-spectra are examined collectively and not just focused on specific peaks. The combination of data-mining techniques with the noninvasive nature of the Raman spectroscopy results in a very promising tool that can be used to evaluate and prescreen the toxicity of particles. Such a technique will be a valuable tool to help design better nanoparticles by reducing their toxicity and their impact on the environment.

Data-mining techniques, such as SVMs, can significantly enhance the capabilities of Raman spectroscopy in recognizing the cellular changes and evaluating toxicology.

Acknowledgements

The authors would like to acknowledge Dr. Krishna for the titania nanoparticles and K. Seibien for the TEM operation. This work was supported by the Particle Engineering Research Center (PERC), University of Florida, the Center for Surfactants and Particulate Systems (CPaSS), University of Florida and Columbia University, the State of Florida, the National Science Foundation (NSF Grant EEC-94-02989, NSF-NIRT Grant EEC-0506560, National High Field Magnet Laboratory), the National Institutes of Health (Grants 1-P20-RR020654-01, RO1HL75258, RO1HL78670) and the Industrial Partners of the PERC. PMP was partially supported by the NSF and Air Force grants.

References

- [1] G. Whitesides, *Nat. Biotechnol.* **2003**, *21*, 1161.
- [2] K. Donaldson, V. Stone, C. Tran, W. Kreyling, P. Borm, *Br. Med. J.* **2004**, *61*, 727.
- [3] K. Powers, M. Palazuelos, B. Moudgil, S. Roberts, *Nanotoxicology* **2007**, *1*, 42.
- [4] K. Powers, S. Brown, V. Krishna, S. Wasdo, B. Moudgil, S. Roberts, *Toxicol. Sci.* **2006**, *90*, 296.
- [5] A. Edinger, C. Thompson, *Curr. Opin. Cell Biol.* **2004**, *16*, 663.
- [6] D. Kanduc, A. Mittelman, R. Serpico, E. Sinigaglia, A. A. Sinha, C. Natale, R. Santacroce, M. G. Di Corcia, A. Lucchese, L. Dini, P. Pani, S. Santacroce, S. Simone, R. Bucci, E. Farber, *Int. J. Oncol.* **2002**, *21*, 165.
- [7] G. Yogalingam, A. M. Pendergast, *Apoptosis* **1997**, *2*, 199.
- [8] P. Golstein, G. Kroemer, *Trends Biochem. Sci.* **2007**, *32*, 37.
- [9] H. Jaeschke, J. S. Gujral, M. L. Bajt, *Liver Int.* **2004**, *24*, 85.
- [10] S. Lanone, F. Rogerieux, J. Geys, A. Dupont, E. Maillot-Marechal, J. Boczkowski, G. Lacroix, P. Hoet, *Part. Fibre. Toxicol.* **2009**, *6*, 14.
- [11] D. Kanduc, F. Capuano, S. A. Capurso, J. Geliebter, D. Guercia, A. Lucchese, A. Mittelman, S. M. Simone, A. A. Sinha, R. Tiwari, E. Farber, *J. Exp. Ther. Oncol.* **2003**, *3*, 108.
- [12] I. Nottingher, J. Selvakumaran, L. L. Hench, *Biosens. Bioelectron.* **2004**, *20*, 780.
- [13] I. Nottingher, *Sensors* **2007**, *7*, 1343.
- [14] P. Rösch, M. Harz, M. Schmitt, K.-D. Peschke, O. Ronneberger, H. Burkhardt, H.-W. Motzkus, M. Lankers, S. Hofer, H. Thiele, J. Popp, *Appl. Environ. Microbiol.* **2005**, *71*, 1626.
- [15] P. Rösch, M. Harz, K.-D. Peschke, O. Ronneberger, H. Burkhardt, J. Popp, *Biopolymers* **2006**, *82*, 312.
- [16] C. Owen, I. Nottingher, R. Hill, M. Stevens, L. Hench, *J. Mater. Sci.: Mater. Med.* **2006**, *17*, 1019.
- [17] G. Pyrgiotakis, T. K. Bhowmick, K. Finton, P. M. Pardalos, K. Powers, B. M., *Biopolymers* **2008**, *89*, 555.
- [18] T. K. Bhowmick, G. Pyrgiotakis, K. Finton, A. K. Suresh, S. G. Kane, J. R. Bellare, B. M. Moudgil, *J. Raman Spectrosc.* **2008**, *39*, 1859.
- [19] G. Pyrgiotakis, O. E. Kundakcioglu, K. Finton, P. M. Pardalos, K. Powers, B. M. Moudgil, *Ann. Biomed. Eng.* **2009**, *37*, 1464.
- [20] E. Widjaja, W. Zheng, Z. Huang, *Int. J. Oncol.* **2008**, *32*, 653.
- [21] D. J. Giard, S. A. Aaronson, G. J. Todaro, P. Arnstein, J. H. Kersey, H. Dosik, W. P. Parks, *J. Natl. Cancer Inst.* **1973**, *51*, 1417.
- [22] ATCC: The Global Bioresource Center. <http://www.atcc.org/>. **2009**.
- [23] N. O. Karpnich, M. Tafani, R. J. Rothman, M. A. Russo, J. L. Farber, *J. Biol. Chem.* **2002**, *277*, 16547.
- [24] P. Huang, W. Plunkett, *Anal. Biochem.* **1992**, *207*, 163.
- [25] V. Solovyan, Z. Bezvenyuk, V. Huotari, T. Tapiola, T. Suuronen, A. Salminen, *Mol. Brain Res.* **1998**, *62*, 43.
- [26] D. Boesewetter, J. Collier, A. Kim, M. Riley, *Cell. Biol. Toxicol.* **2006**, *22*, 101.
- [27] I. Nottingher, S. Verrier, S. Haque, J. M. Polak, L. L. Hench, *Biopolymers* **2003**, *72*, 230.
- [28] C. A. Owen, J. Selvakumaran, I. Nottingher, G. Jell, L. L. Hench, M. M. Stevens, *J. Cell. Biochem.* **2006**, *99*, 178.
- [29] R. Stearns, J. Paulauskis, J. Godleski, *Am. J. Respir. Cell Mol. Biol.* **2001**, *24*, 108.
- [30] G. Yogalingam, A. M. Pendergast, *J. Biol. Chem.* **2008**, *283*, 35941.
- [31] K. Maquelin, L. P. Choo-Smith, T. van Vreeswijk, H. P. Endtz, B. Smith, R. Bennett, H. A. Bruining, G. J. Puppels, *Anal. Chem.* **2000**, *72*, 12.
- [32] C. Lieber, A. Mahadevan-Jansen, *Appl. Spectrosc.* **2003**, *57*, 1363.
- [33] I. Nottingher, S. Verrier, H. Romanska, A. E. Bishop, J. M. Polak, L. L. Hench, *Spectrosc. Int. J.* **2002**, *16*, 43.
- [34] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag: Red Bank, NJ, **1995**.
- [35] H. Byun, S.-W. Lee, in *Pattern Recognition with Support Vector Machines* (Eds: S. Lee, A. Verri), Springer: Niagara Falls, **2002**, pp. 186.
- [36] T. Joachims, in *Proceedings of the European Conference on Machine Learning* (Eds: C. Nédellec, C. Rouveirol), Springer: Berlin, **1998**, pp. 137.
- [37] M. Brown, W. Grundy, D. Lin, N. Cristianini, C. Sugne, T. Furey, M. Ares, D. Haussler, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 262.
- [38] C. Cifarelli, G. Patrizi, *Optim. Methods Softw.* **2007**, *22*, 25.
- [39] W. S. Noble, in *Kernel Methods in Computational Biology* (Eds: B. Schölkopf, K. Tsuda, J.-P. Vert), MIT Press: Paris, **2004**, pp. 71.
- [40] M. R. Guarracino, S. Cuciniello, D. Feminiano, G. Toraldo, P. M. Pardalos, in *Data Mining and Mathematical Programming* (Eds: P. M. Pardalos, P. Hansen), American Mathematical Society: Gainesville, FL, **2008**, pp. 109.
- [41] M. Bewernitz, G. Ghacibeh, O. Seref, P. M. Pardalos, C.-C. Liu, B. Uthman, in *Data Mining, Systems Analysis and Optimization in Biomedicine* (Eds: O. Seref, O. E. Kundakcioglu, P. M. Pardalos), American Institute of Physics: Gainesville, FL, **2008**, pp. 206.
- [42] T. N. Lal, M. Schroeder, T. Hinterberger, J. Weston, M. Bogdan, N. Birbaumer, B. Schölkopf, *IEEE Trans. Biomed. Eng.* **2004**, *51*, 1003.
- [43] G. Garcia, T. Ebrahimi, J. Vesin, *EURASIP J. Adv. Signal Process.* **2003**, *7*, 713.
- [44] Z. Huang, H. Chen, C. J. Hsu, W. H. Chen, S. Wuc, *Decis. Support Syst.* **2004**, *37*, 543.
- [45] T. B. Trafalis, H. Ince, in *IJCNN 2000: Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks*, vol. 6 (Ed.: S.-I. Amari), IEEE Computer Society, **2000**, pp. 348.
- [46] J. Shawe-Taylor, N. Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge University Press: Cambridge, **2004**.
- [47] A. Lunts, V. Brailovskiy, *Eng. Cybern.* **1967**, *3*, 98.
- [48] A. Simon-Deckers, B. Gouget, M. Mayne-L'Hermite, N. Herlin-Boime, C. Reynaud, M. Carrière, *Toxicology* **2008**, *253*, 137.
- [49] S. Li, R. Zhu, H. Zhu, M. Xue, X. Sun, S. Yao, S. Wang, *Food Chem. Toxicol.* **2008**, *46*, 3626.
- [50] S. Barillet, A. Simon-Deckers, N. Herlin-Boime, M. Mayne-L'Hermite, C. Reynaud, D. Cassio, B. Gouget, M. Carrière, *J. Nanopart. Res.* **2010**, *12*, 1.
- [51] I. Nottingher, C. Green, C. Dyer, E. Perkins, N. Hopkins, C. Lindsay, L. L. Hench, *J. R. Soc. Interface* **2004**, *1*, 79.
- [52] G. Puppels, H. Garritsen, G. Segers-Nolten, F. de Mul, J. Greve, *Biophys. J.* **1991**, *60*, 1046.
- [53] C. Rajani, J. R. Kincaid, D. H. Petering, *Biophys. Chem.* **2001**, *94*, 219.
- [54] N. Stone, C. Kendall, J. Smith, P. Crow, H. Barr, *Faraday Discuss.* **2004**, *126*, 141.
- [55] X. Yiming, Z. Zhixiang, Y. Hongying, X. Yan, Z. Zhiyi, *J. Photochem. Photobiol., B: Biol.* **1999**, *52*, 30.
- [56] S. Overman, K. Aubrey, K. Reilly, O. Osman, S. Hayes, P. Serwer, G. Thomas, *Biospectroscopy* **1998**, *4*, 47.
- [57] G. Puppels, J. Greve, in *Biomedical Applications of Spectroscopy* (Eds: R. J. H. Clark, R. E. Hester), Wiley: Chichester, **1996**, pp. 1.
- [58] R. Rasmussen, R. Painter, *Nature* **1964**, *203*, 1360.
- [59] K. Osberg, J. Osborn, S. Zhang, J. Freyer, J. Mourant, J. Schoonover, *Appl. Spectrosc.* **2002**, *56*, 813.
- [60] D. Naumann, in *Infrared and Raman Spectroscopy of Biological Materials, Practical Spectroscopy Series*, vol. 24 (Eds: H. U. Gremlich, B. Yan), CRC: New York, **2000**, pp. 323.
- [61] D. Borchman, D. Tang, M. Yappert, *Biospectroscopy* **1999**, *5*, 151.
- [62] R. Lakshmi, V. Kartha, C. Krishna, J. Solomon, G. Ullas, P. Devi, *Radiat. Res.* **2002**, *157*, 175.
- [63] A. Mahadevan-Jansen, R. Richards-Kortum, *J. Biomed. Opt.* **1996**, *1*, 31.
- [64] H. G. M. Edwards, E. A. Carter, in *Infrared and Raman Spectroscopy of Biological Materials* (Eds: H. U. Gremlich, B. Yan), CRC: New York, **2000**, pp. 421.