

FULL ARTICLE

# RMieS-EMSC correction for infrared spectra of biological cells: Extension using full Mie theory and GPU computing

Paul Bassan<sup>1</sup>, Achim Kohler<sup>2</sup>, Harald Martens<sup>2,3</sup>, Joe Lee<sup>4</sup>, Edward Jackson<sup>1</sup>, Nicholas Lockyer<sup>1</sup>, Paul Dumas<sup>5</sup>, Michael Brown<sup>6</sup>, Noel Clarke<sup>6,7,8</sup>, and Peter Gardner\*,<sup>1</sup>

<sup>1</sup> School of Chemical Engineering and Analytical Science, Manchester Interdisciplinary Biocentre, University of Manchester, 131 Princess Street, Manchester, M1 7DN, UK

<sup>2</sup> Nofima Mat, Centre for Biospectroscopy and Data Modelling, Osloveien 1, 1430 Ås, Norway

<sup>3</sup> CIGENE, Department of Mathematical Sciences and Technology, Norwegian University of Life Sciences, 1430 Ås, Norway

<sup>4</sup> School of Chemical Engineering and Analytical Science, The Mill, The University of Manchester, Oxford Road, Manchester, M13 9PL, UK

<sup>5</sup> Synchrotron SOLEIL, L'Orme des Merisiers, BP48 – Saint Aubin, 91192 Gif-sur-Yvette Cedex, France

<sup>6</sup> Genito Urinary Cancer Research Group, School of Cancer and Enabling Sciences, Paterson Institute for Cancer Research, The University of Manchester, Manchester Academic Health Science Centre, The Christie NHS Foundation Trust, Manchester, UK, M20 4BX, UK

<sup>7</sup> Department of Urology, The Christie NHS Foundation Trust, Manchester, UK, M20 4BX, UK

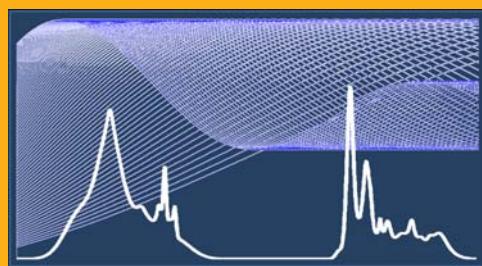
<sup>8</sup> Department of Urology, Salford Royal NHS Foundation Trust, Salford, UK, M6 8HD, UK

Received 26 February 2010, revised 4 April 2010, accepted 5 April 2010

Published online 26 April 2010

**Key words:** FTIR, single cells, cell nucleus, infrared microspectroscopy, anomalous dispersion artefact, signal correction, EMSC, Mie scattering, resonant Mie scattering

In the field of biomedical infrared spectroscopy it is often desirable to obtain spectra at the cellular level. Samples consisting of isolated single biological cells are particularly unsuited to such analysis since cells are strong scatterers of infrared radiation. Thus measured spectra consist of an absorption component often highly distorted by scattering effects. It is now known that the predominant contribution to the scattering is Resonant Mie Scattering (RMieS) and recently we have shown that this can be corrected for, using an iterative algorithm based on Extended Multiplicative Signal Correction (EMSC) and a Mie approximation formula. Here we present an iterative algorithm that applies full Mie scattering theory. In order to avoid noise accumulation in the iterative algorithm a curve-fitting step is implemented on the new reference spectrum. The new algorithm increases the computational time when run on an equivalent processor. Therefore parallel processing by a Graphics Processing Unit (GPU) was employed to re-



duce computation time. The optimised RMieS-EMSC algorithm is applied to an IR spectroscopy data set of cultured single isolated prostate cancer (PC-3) cells, where it is shown that spectral distortions from RMieS are removed.

\* Corresponding author: e-mail: Peter.Gardner@manchester.ac.uk, Tel.: 0161 306 4463, Fax: 0161 306 5201

## 1. Introduction

In the field of biomedical infrared spectroscopy there is a drive towards obtaining spectra at high spatial resolution such that biologically relevant information can be obtained at the single cell level [1–12]. This desire for spatially resolved chemical information applies equally to cells in tissue or isolated cells deposited onto an infrared compatible substrate. The goal is to compare the biochemistry associated with one cell with that of neighbouring cells in the tissue or culture, usually in an attempt to obtain specific signatures or markers for disease. Unfortunately this is rarely possible since a measured spectrum often contains contributions from absorption, reflection and scattering [13, 14]. Biological cells are typically a few micrometres to a few tens of micrometres in diameter. This is comparable in size to the wavelength of mid-infrared radiation used in a conventional infrared microspectroscopy system. Consequently the cells in the sample scatter the infrared radiation being used to probe them very efficiently. Thus the resulting spectrum contains contributions from both effects. Given that cell size and shape (morphology) is variable within any given sample, obtaining subtle differences in biochemistry alone, from a measured spectrum, is very difficult to achieve. It is therefore imperative that methods are developed to separate the scattering contributions and the pure absorption contribution from the measured spectrum [15–16]. In order to do this one must first understand the predominant cause of the scattering phenomenon. In 2005, Mohlenhoff et al. postulated that Mie scattering from the cell nucleus was the cause of the strong broad oscillations in the spectrum baseline [17]. In 2008, Kohler et al. published the first reliable correction for these Mie type scattering oscillations in the baseline [18], but the sharp reduction in intensity on the high wavenumber side of the Amide I band, often mistakenly called the “dispersion artefact” frequently remained uncorrected [18]. Recently, Bassan et al. demonstrated that both the broad oscillation in the baseline and the “dispersion artefact” could be described by the phenomenon of Resonant Mie Scattering (RMieS) [16, 19]. It should be noted that the term “dispersion artefact” is a misnomer since the pronounced dip in intensity is only indirectly related to anomalous dispersion and it is not an artefact. It is also not solely a reflection component, as is often mistakenly believed, and although there may well be a pure reflection contribution, this is adequately incorporated in the resonant Mie scattering model (as reflection is a special case of scattering which is explained by Mie theory [20]).

Previous papers on the subject of correcting the effects of Mie scattering in IR spectra of scattering samples have utilised the van de Hulst approxima-

tion equation [20], for the calculation of the Mie scattering efficiency [18, 19]. This equation is an approximation used instead of full Mie theory [21], which is significantly more complicated and computationally expensive to implement. There are a number of conditions under which the van de Hulst equation should be used: for homogenous non-absorbing spherical particles. Highly scattering samples measured with IR (i.e. single cells and tissues) are inhomogeneous non-spherical absorbing particles rendering the equation of limited applicability.

Previous work by Bassan et al. [19] used an adapted form of the van de Hulst equation where the real refractive,  $n$ , used was not a constant, but variable. This was a “spectrum” obtained from the Kramers–Kronig transform of a reference spectrum. A comparison of measured and theoretical data using PMMA microspheres shows that this adaptation of the van de Hulst equation was a reasonable substitute for the full Mie theory [16].

In this paper we introduce full Mie theory and a method of optimising the reference spectrum iteratively. Full Mie theory is considerably more computationally expensive to implement, and so we have harnessed the power of graphics processing unit (GPU) computing to reduce this time.

The full Mie theory (RMieS-EMSC) algorithm is applied to a typical data set consisting of spectra from single isolated prostate cancer (PC-3) cells which have been a focus of our research recently [6–8, 22, 23].

## 2. Experimental

### 2.1 Cell culture

Cells from the PC-3 human prostate adenocarcinoma cell line were seeded in 25 ml sterile plastic culture flasks (Fisher Scientific, UK). After culturing (37 °C, 5% CO<sub>2</sub>, in air) in PC-3 media (Ham's F12, 7% FCS and 2 mM L-glutamine, Sigma Aldrich, UK) to approximately 80% confluence, the media was removed. The cells were then washed in Dulbecco's Phosphate Buffered Saline (DPBS) (10 ml, Sigma Aldrich, UK), after which 1 ml trypsin solution (0.5 mg/ml trypsin; 0.2 mg/ml EDTA in PBS, Sigma Aldrich, UK) was added and the cells were incubated (3 minutes, 37 °C, 5% CO<sub>2</sub>, in air). The cells were then re-suspended in the growth media (4 ml) and removed for centrifugation. Once removed, the cell suspension was centrifuged (800 g, 5 minutes) and the supernatant discarded. The cells were then re-suspended in DPBS (5 ml) and centrifuged as before. Finally, the DPBS was removed and the cells were suspended – and fixed – in ethanol (4 ml, 70%

in HPLC grade water, Sigma Aldrich, UK). Suspensions of the PC-3 cells (150 µl) were cytopspun (800 RPM, 5 minutes, low acceleration) onto calcium fluoride ( $\text{CaF}_2$ ) slides for FTIR analysis.

## 2.2 Synchrotron infrared micro-spectroscopy

The synchrotron FTIR micro-spectroscopy data were recorded at the synchrotron SOLEIL on the SMIS beamline, details of which can be found elsewhere [24]. The spectra were obtained using a Nicolet Nic-Plan microscope equipped with an MCT detector. Spectra were recorded by co-adding 512 interferograms at  $8 \text{ cm}^{-1}$  resolution. The size of the aperture was  $20 \times 20 \mu\text{m}^2$ , which roughly matched the diameter of the cells such that each cell was fully illuminated.

## 3. RMieS-EMSC

### 3.1. Mie theory

In 1908 Gustav Mie wrote his theory for the scattering of light by homogenous spherical absorbing particles [21]. Until the advent of high speed computing, calculations were considered impractical due to the heavy computation required, particularly when a range of wavelength or sphere diameter is involved. For the purposes of correcting IR spectra from highly scattering samples a number of assumptions and simplifications need to be made, the most important of which is that the non-spherical particles can be approximated as several spheres of different radii. Mie theory requires knowledge of the scattering particle radius,  $r$ , and the complex refractive index ratio,  $\eta$ , of scattering particle to air. The real and imaginary parts of the refractive index of air can be assumed to be the same as vacuum,  $n_{\text{air}} = 1$ ,  $k_{\text{air}} = 0$ . This simplifies the complex refractive index ratio to the index of the scattering particle and hence this convention will be used throughout this paper.

Mie theory describes the total loss of light caused by a scattering particle due to both the absorption characteristics of the particle and scattering; the sum of the absorption and scattering efficiencies ( $Q_{\text{abs}} + Q_{\text{sca}}$ ) is termed the extinction efficiency,  $Q_{\text{ext}}$ . Of interest to us are  $Q_{\text{sca}}$  values of the particle over a range of wavelengths, the full mathematical definition of which is given in the Appendix.

### 3.2. Implementation of full Mie theory in the RMieS-EMSC algorithm

Bassan et al. [19] published the first algorithm to correct for resonant Mie scattering of highly scattering samples which was based upon a non-resonant Mie scattering algorithm by Kohler et al. [18]. Both of these algorithms build upon the extended multiplicative signal correction (EMSC) by Martens et al. [25]. Here we show an optimisation of the RMieS-EMSC which uses full Mie theory instead of the van de Hulst approximation for the computation of scattering efficiencies. The RMieS-EMSC model is

$$\vec{Z}_{\text{Raw}} = c + m\vec{v} + h\vec{Z}_{\text{Ref}} + \sum_{i=1}^N g_i \vec{p}_i + \vec{E} \quad (1)$$

where  $Z_{\text{Raw}}$  = the raw spectrum to be corrected,  $c$  = constant offset baseline,  $m$  = gradient of sloping baseline,  $Z_{\text{Ref}}$  = reference spectrum,  $h$  = magnitude of  $Z_{\text{Ref}}$ ,  $N$  = number of principal components used,  $g$  = weight applied to loadings,  $p$ , of the Mie scattering matrix and  $E$  = un-modelled residual variance. Symbols with arrows above denote vectors, i.e. a column or row of numbers.

The algorithm assumes that the measured spectrum is a linear combination of a constant and sloping baseline, a reference spectrum, a scattering curve and a spectrum of features that could not be modelled. The sigma term in the equation is where the variation due to RMieS is accounted for in the form of weights applied to loadings from the principal component analysis (PCA) decomposition of a matrix containing many scattering curves of different particle radius and refractive index.

In this paper we change the algorithm such that the matrix of RMieS curves is calculated using full Mie theory instead of the van de Hulst approximation. The use of full Mie theory requires certain manipulations to deal with the use of a complex refractive instead of just the real part of the refractive as previously implemented. The complex refractive index,  $\eta$ , is defined as

$$\eta = n - ik \quad (2)$$

where  $n$  = real part of the refractive index,  $k$  = imaginary part of the refractive index and  $i$  = the imaginary unit,  $\sqrt{-1}$ .

To calculate the  $Q_{\text{sca}}$  curves, both real and imaginary refractive indices are needed, the starting point of which is the reference spectrum. The reference spectrum is taken and normalised such that the Amide I band peak is at an absorbance of 1. At this stage it is useful to introduce two terms which construct the real refractive index,  $a$  and  $b$  which are

the average refractive index and “resonance parameter” respectively, such that:

$$n = a + b n_{KK} \quad (3)$$

where  $n_{KK}$  is the Kramers–Kronig transform of the reference spectrum minus an average refractive index such that it is centred around zero allowing us to add our own desired average refractive index.

To create a database of  $Q_{sca}$  curves covering many possibilities, 10 equidistant values of particle radius,  $r$ , are used between 2 and 8  $\mu\text{m}$ . A range of complex refractive indices also needs to be covered, and so 10 equidistant coefficients are multiplied with the Amide I normalised reference spectrum ranging from 0 to 0.4. The value of 0 represents a non-absorbing particle, and the upper limit of 0.4 is somewhat arbitrary but an educated guess suggests that the value of the Amide I peak in the  $k$  spectrum is not likely to exceed a value of 0.4. For each of these 10,  $k$ , possibilities, a Kramers–Kronig,  $n_{KK}$  has to be calculated and then an average real refractive index,  $a$ , has to be added. The range of values that  $a$  can take is limited by the  $n_{KK}$  spectrum, e.g., if the lowest value in the  $n_{KK}$  spectrum is  $-0.1$ , then the lowest value  $a$  can take is 1.1 ensuring that the real refractive index never goes below 1. The upper limit for  $a$  has been chosen to be 1.5 which is an appropriate upper limit for biological cells. Using 10 equidistant values of  $r$ ,  $a$  and  $b$  between the stated ranges gives 1000 permutations and hence 1000  $Q_{sca}$  curves which cover many possible particle radii and complex refractive indices within a realistic range.

As the Appendix shows, the calculation of  $Q_{sca}$  requires the computation of spherical Bessel and Hankel functions which are considered computationally intensive and slow. Therefore, an algorithm based on a stepwise adaptation of Bessel and Hankel functions to spectra of highly scattering samples by changing scattering particle radius,  $r$ , and the complex refractive index ratio,  $\eta$ , of scattering particle to air is not suitable in practice. The computation involved would be too time-consuming for practical purposes. In order to circumvent tedious calculations for the adaptation of each spectrum we approximate the database of 1000 Mie curves  $Q_{sca}$  by a subspace model using a few principal components according to Kohler et al. [18] and build this subspace model in the EMSC model of Eq. (1). Since only principal components are used in the model of Eq. (1), the estimation of parameters by linear regression model used for the estimation is stable. The subspace modelling according to Kohler et al. [18] has been employed by us for the correction of RMieS scattering using the approximate formula by van de Hulst [20], but shows its full power here, where calculation of RMieS scatter curves using the full Mie theory is computationally very intensive.

In order to decrease computation time for building the database, a method is discussed to optimise the calculation of these spherical functions to save time using graphics processing unit (GPU) computing in the next section.

### 3.3 Graphics processing unit (GPU) computing

Modern GPUs contain many processing cores each of which can be used in a similar manner to conventional central processing units (CPUs). The computer system used here is an Intel 2.4 GHz Quad Core Xeon with 12 GB of RAM and an NVIDIA Tesla 1060c GPU. This GPU has 240 processing cores which operate at 1.3 GHz each and hence is ideally suited to parallel processing, such as the calculation of Bessel functions which can be done independently from one another. To calculate a 1000 curve  $Q_{sca}$  database using Matlab 2009a (Mathworks, Natick, MA, USA) on our computer takes 89.91 seconds for a spectrum of 1601 intensity values. The only way this calculation can be speeded up on the CPU is by having a higher frequency CPU, i.e., greater than 2.4 GHz. The computation time of the  $Q_{sca}$  database using the van de Hulst approximation for the same example is 0.324 seconds. However due to heat production issues, CPU speeds are not being increased above typically 3.33 GHz, and instead multiple cores (such as Quad core processors) are being produced which does not necessarily help certain computation. In the case of the spherical Bessel and Hankel functions within Matlab 2009a, these functions are not able to access multiple cores meaning that multiple core machines produce no further speed gains.

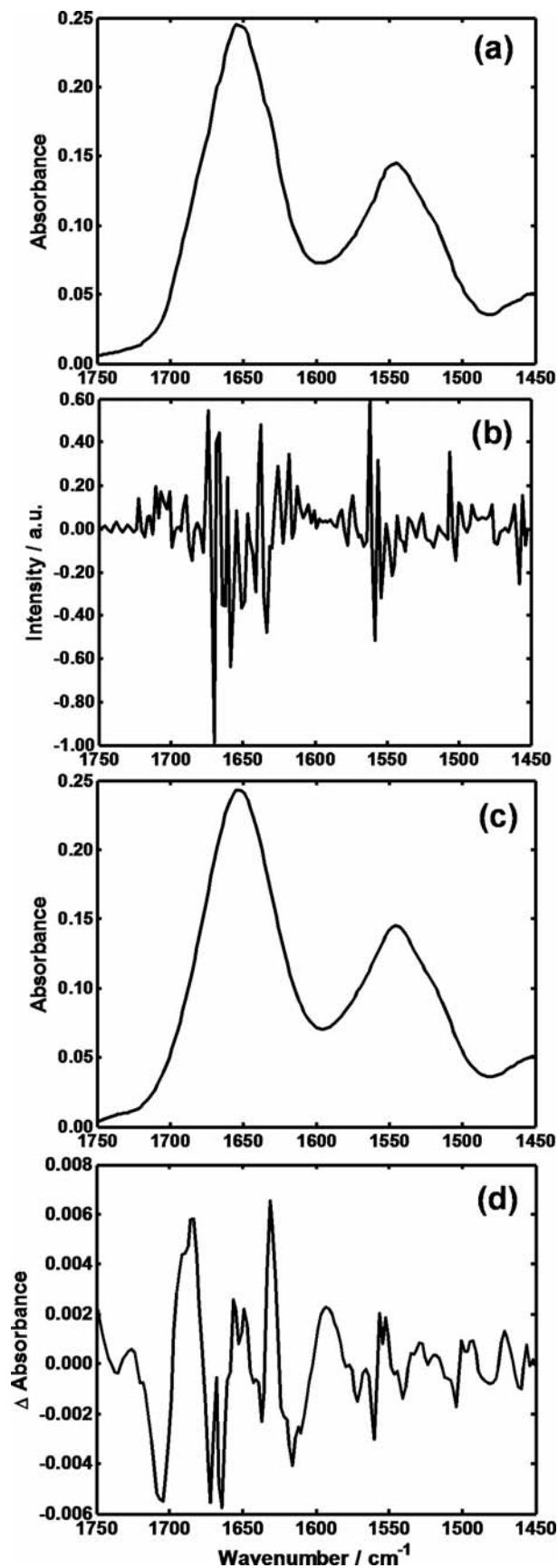
A toolbox “The Jacket Toolbox” (Acclereyes, Atlanta, GA, USA) available for Matlab contains a programme that acts as an intermediate layer between Matlab and the GPU. Of particular interest here are the spherical Bessel and Hankel functions. The equivalent computation of the  $Q_{sca}$  database mentioned above can be completed in 35.94 seconds, a speed gain of 2.48. Although this speed gain may seem trivial for a single spectrum, when using the RMieS-EMSC algorithm for correcting IR images with many pixels, this speed gain will save many hours. The current version of the Jacket Toolbox (v1.2.2) is going to be improved in the near future such that the speed gains of these functions will increase by at least an order of magnitude which is very significant. In the future more sections of the algorithm will be transferred to the GPU to take advantage of its computational potential.

### 3.4. Reference spectrum optimisation

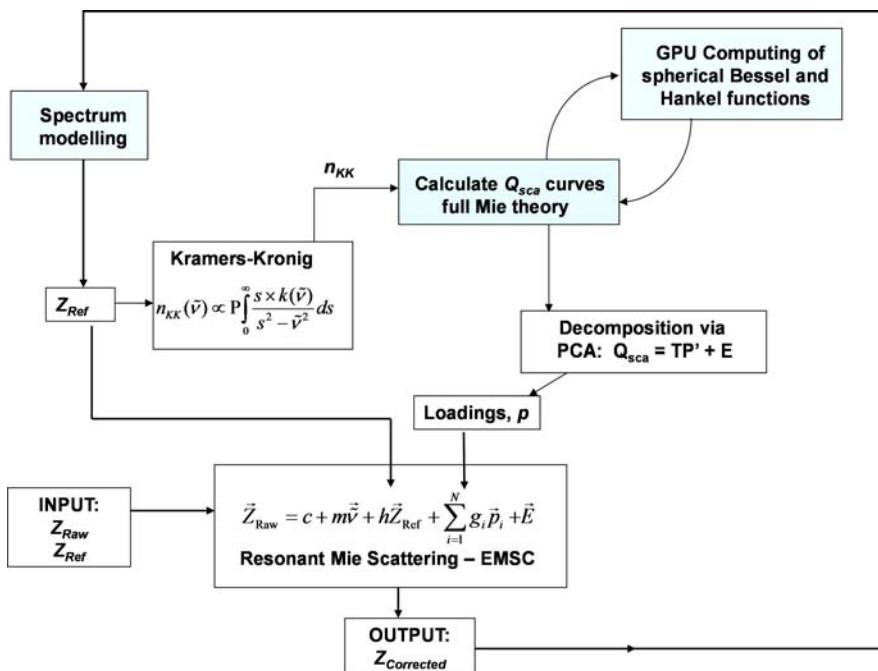
The weakest part of the RMieS-EMSC is the use of a reference spectrum to provide stability during the linear regression stage to estimate the scatter contributions. The important point is that the perfect reference spectrum to use for a correction is the answer that is being calculated, i.e., the pure absorbance spectrum of the sample. Since this is never available, a reference spectrum that will be close to the answer being sought is required. The result of using a reference spectrum that is “approximately” correct is that the corrected spectrum contains elements of the reference spectrum that may not be present in the true absorption spectrum of the sample [19]. A solution to this problem is to iteratively improve the reference spectrum by using the corrected spectrum of one iteration as the new reference spectrum for another; the rationale being that eventually the corrected spectrum would converge on the pure absorbance spectrum of the sample [19].

The RMieS-EMSC was previously demonstrated on a simulated data set constructed by adding many Gaussian curves together which were subsequently distorted by adding theoretical RMieS curves [19]. These spectra were noise free and showed proof of principle, but there are several potential issues that can arise when attempting to iteratively correct a real measured spectrum.

The algorithm requires the real refractive index to be calculated by applying the Kramers–Kronig transform to the newly corrected spectrum. Measured data often contain noise, and for spectra acquired using a Synchrotron source, the noise is often increased at the low-wavenumber end of the spectrum due to the diffraction limit and/or substrate transmittance. The Kramers–Kronig transform of a poor S/N spectrum will yield an  $n$  spectrum of ever poorer S/N, (in a similar way that first derivative spectra are more noisy than the original spectrum). If a noisy  $n$  spectrum is inserted into the Mie formulae, the resultant Mie scattering curves would exhibit similar noise properties which is highly undesirable. In this situation, corrected spectra can often come out of the algorithm with features of the scattering remaining, and/or absorbance values below zero in certain regions of the spectrum.



**Figure 1** (a) Measured spectrum of Matrigel film. (b) Second derivative spectrum of the Matrigel film. (c) Resulting fitted spectrum. (d) The difference between the measured and fitted spectra.



**Figure 2** (online color at: [www.biophotonics-journal.org](http://www.biophotonics-journal.org)) Schematic of the optimized RMieS-EMSC algorithm with modifications highlighted.

To reduce the issue of noise and remaining scattering effects within a corrected spectrum a fitting step was introduced. The newly corrected spectrum is taken and interpolated such that the data spacing correspond to  $0.5 \text{ cm}^{-1}$ , after which a Savitzky-Golay second derivative with 9 smoothing is implemented (to locate peak positions to an accuracy of  $\pm 0.25 \text{ cm}^{-1}$ ). The region between  $1800$  and  $2500 \text{ cm}^{-1}$  contains no biochemical information but still contains remnants of the scattering effects, and hence is replaced with zeros. Using a peak minima finding algorithm, the second derivative spectrum yields the wavenumber position of peaks including secondary structure peak information of, e.g., the Amide I band. Using these peak positions, a database is constructed comprising of Gaussian curves of various widths centred at each peak position. A non-negative linear regression – as peaks can only be positive in an absorbance spectrum – is then performed on the corrected spectrum using the Gaussian curves as descriptive vectors. Figure 1 shows the results of this fitting process applied to a measured spectrum of a thin layer of Matrigel.

Figure 1 shows the quality of the fitting procedure, where Figure 1(d) shows the difference between the measured and fitted spectrum. The difference between the spectra is 3 orders of magnitude lower than the measured spectrum and hence it can be considered a good representation. This fitted spectrum now contains zero noise and since each Gaussian curve was centred on a peak position from the second derivative spectrum, it contains no negative values or any scatter contributions.

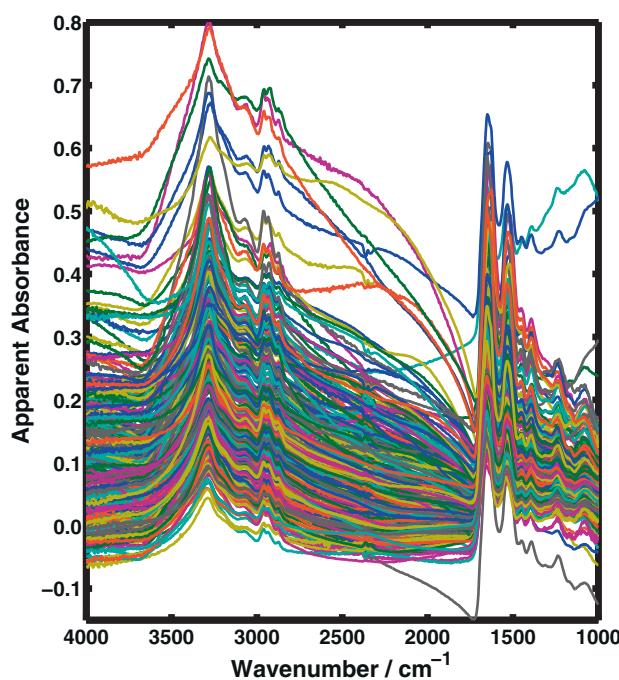
### 3.5 Optimisation of the RMieS-EMSC algorithm

In this paper the RMieS-EMSC algorithm has been optimised to incorporate full Mie theory which is computed using GPU computing. The iterative improvement of the reference spectrum has been optimised by modelling the corrected spectrum using a Gaussian database to reduce noise. The remainder of the algorithm has not been stated because it can be found in our previous publication [19], however a concise summary is presented in flow chart form below. Figure 2 shows a schematic of the RMieS-EMSC algorithm with the new optimisations presented in this paper highlighted. The callout and return from the calculation of the  $Q_{\text{sca}}$  database to the GPU represents the calculation of the spherical Bessel and Hankel functions on the GPU so that they can be parallel processed for speed optimisation.

## 4. Results

### 4.1 Correction of single prostate cancer (PC-3) cells

Figure 3 shows 280 raw FTIR spectra of single isolated PC-3 cells on  $\text{CaF}_2$ . Despite all the spectra belonging to cells from a single cell line, they are ob-

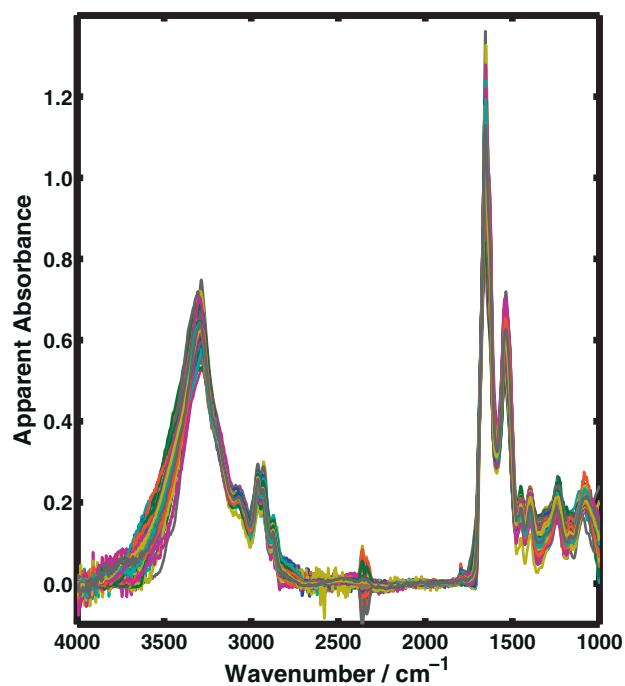


**Figure 3** (online color at: [www.biophotonics-journal.org](http://www.biophotonics-journal.org)) Raw transmission FTIR microscopy spectra of 280 isolated single PC-3 cells.

viously not all the same, showing significant variation over the whole wavenumber range. The spectra are clearly distorted by resonant Mie scattering which results in broad oscillating baselines, and the decrease in intensity around  $1700\text{ cm}^{-1}$ . The biochemistry of these cells is not expected to be identical, due to natural variation and different stages of the cell cycle but the spectra were expected to be very similar in appearance.

As we have previously shown [16, 19], the differences observed in Figure 3 can be fully explained and are caused by resonant Mie scattering. The morphological characteristics of each cell, namely the size and shape of the cell, the size of the nucleus and other organelles varies significantly amongst the 280 cells measured. RMieS is dependent on the scattering particle radius and refractive index properties [16, 19] so it is not surprising that the spectra look very different. These scattering effects distort the pure absorbance spectrum of each cell which is of interest to us and yields the spectrum of limited use, and the biochemical information cannot be trusted.

Using the RMieS-EMSC algorithm, the presence of the scattering effects can be significantly reduced yielding corrected spectra which more closely resemble the pure absorbance spectrum of each cell [19]. Figure 4 shows the 280 spectra shown in Figure 3, corrected using the RMieS-EMSC algorithm incorporating the full Mie theory. Five iterations of the algorithm were used after which there was no signifi-



**Figure 4** (online color at: [www.biophotonics-journal.org](http://www.biophotonics-journal.org)) RMieS-EMSC corrected spectra of the data set shown in Figure 3.

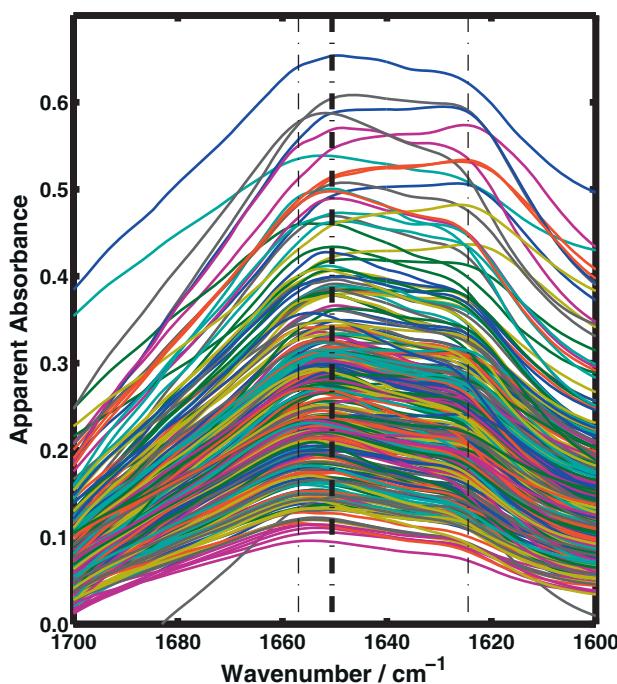
cant change in the resulting spectra. It is immediately obvious that the removal of scattering contributions produces a much more internally consistent looking set of spectra.

Although some variation in the spectra is still present, this is to be expected since there will be natural variation within a cell line.

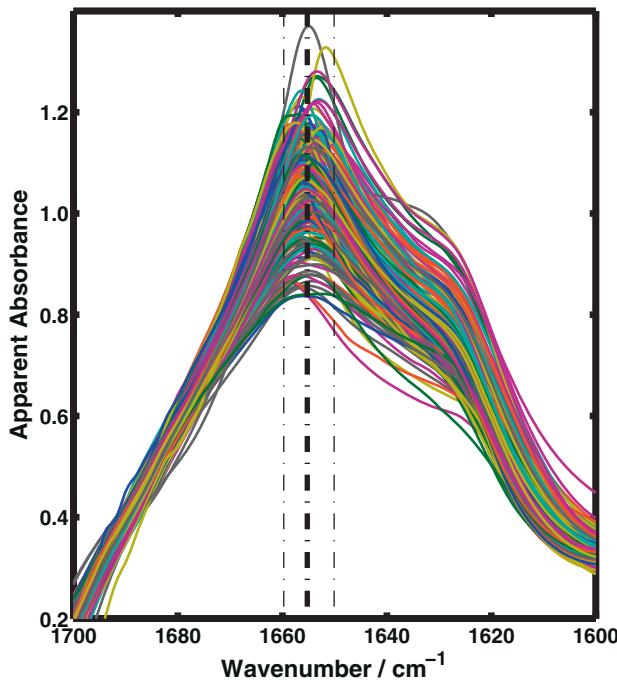
An important feature of biological spectra is the band shape and peak maximum position of the Amide I band. This is important since under non-scattering conditions the position and shape of the Amide I band can be indicative of protein secondary structure. Figure 5 shows an expanded plot of the Amide I band of the raw spectra in Figure 3, the mean peak position is  $1650.5\text{ cm}^{-1}$  (central dashed trace). The lowest and highest peak position of the Amide I bands is  $1624.4$  and  $1656.9\text{ cm}^{-1}$ , a range of  $32.5\text{ cm}^{-1}$ .

This range of peak positions is far greater than would be expected, for a sample of cells from the same cell line as is the case here. The true Amide I peak position has been distorted due to resonant Mie scattering and if taken at face value, incorrect conclusions regarding protein secondary structure may be drawn inappropriately. To produce reliable peak positions the effects of resonant Mie scattering have to be removed, Figure 6 shows an expanded plot of the Amide I band for the corrected spectra from Figure 4.

The Amide I peak positions of the corrected spectra are significantly different from the raw spectra, with



**Figure 5** (online color at: [www.biophotonics-journal.org](http://www.biophotonics-journal.org)) Amide I bands of the raw spectra from Figure 3. Dashed traces from left to right show, maximum, mean and minimum peak positions respectively.



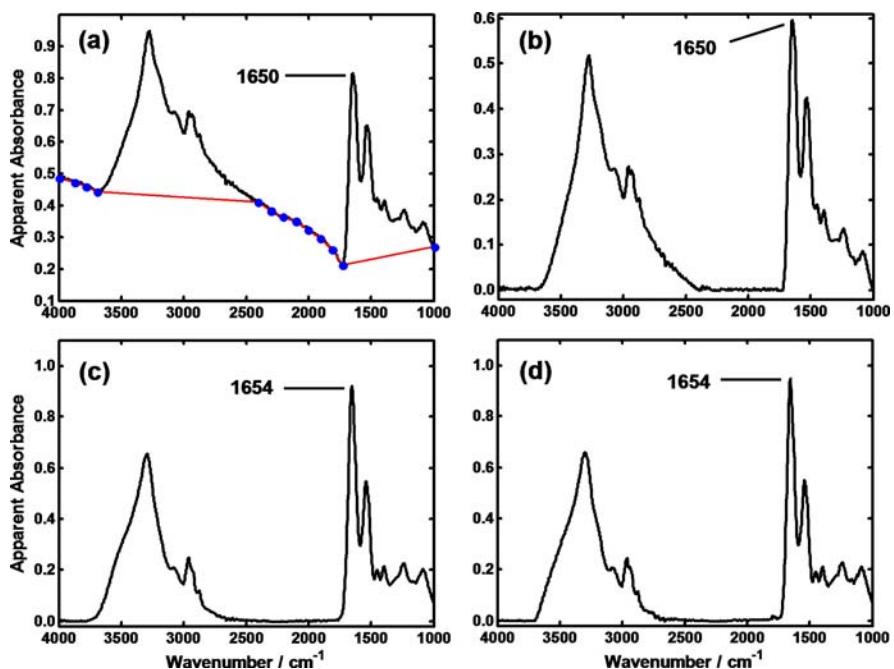
**Figure 6** (online color at: [www.biophotonics-journal.org](http://www.biophotonics-journal.org)) Amide I peak positions of the corrected spectra from Figure 4. Dashed traces from left to right show: maximum, mean and minimum peak positions.

the mean peak position now at  $1655.2\text{ cm}^{-1}$ . The minimum and maximum peak positions are now  $1650.1$  and  $1659.7\text{ cm}^{-1}$  respectively, a range of  $9.6\text{ cm}^{-1}$ .

#### 4.2 Comparison with “Rubber Band” and previous Mie scattering correction

Despite the well known problem of scattering artefacts in spectra of biological cells, in the absence of a scattering correction algorithm a simple “baseline correction” has often been employed using the so-called “rubber band” method [26, 27].

Figure 7(a) shows one of the raw spectra from the data set in Figure 3 and is very typical of a single cell spectrum exhibiting the classic sloping baseline between the C–H stretching region and the Amide I band. The blue dots indicate a typical baseline correction using the rubber band method. The resulting corrected spectrum is shown in Figure 7(b). As can be seen the spectrum *looks* very good and is similar to the RMieS-EMSC corrected spectra in Figure 7(c and d). What is important however, is that the position of the Amide I band is not corrected properly by the rubber band method and is still significantly low-wavenumber shifted due to the RMieS. The rubber band method corrected peak position is at  $1650\text{ cm}^{-1}$ , compared with  $1654\text{ cm}^{-1}$  using the RMieS-EMSC algorithms. Although one could argue that we do not know what the “correct” position of this band should be, extensive testing of the RMieS algorithm on simulated data strongly indicates that the RMieS-EMSC correction results in an Amide I band position within  $1\text{ cm}^{-1}$  of the correct value [19]. Figure 7(c) is the spectrum corrected with the RMieS-EMSC algorithm using the van de Hulst approximation. Figure 7(d) shows the same spectrum corrected using the algorithm presented in this paper, incorporating the full Mie scattering theory. A comparison of the two spectra shows that they are essentially identical with the Amide I band position at  $1654\text{ cm}^{-1}$  in both cases and the same intensity distribution in the diagnostically important region from  $1500$ – $1000\text{ cm}^{-1}$ . Thus, it is clear that as a spectral correction method the previous RMieS-EMSC algorithm is as good as the new algorithm using the full Mie theory, and importantly it is significantly quicker. This begs the question, why develop the new algorithm? The answer to this is two-fold. Firstly, the fact that the algorithms based on the approximation and the full Mie theory agree almost exactly demonstrate the validity of using the van de Hulst approximation. Secondly, although the van de Hulst approximation works, it transpires that the scattering parameters, e.g., radii of scatterers and refractive index combinations that lead to the cor-



**Figure 7** (online color at: [www.biophotonics-journal.org](http://www.biophotonics-journal.org)) (a) Spectrum of a single PC-3 cell exhibiting RMieS. Blue markers and connecting dashed lines show the points used for performing a rubber band correction (b) Spectrum corrected using rubber band correction. (c) Spectrum corrected using the van de Hulst based RMieS-EMSC algorithm (5 iterations). (d) Spectrum corrected using the full Mie theory based RMieS-EMSC algorithm (5 iterations).

rected spectra are slightly different in each case. Pijanka et al. [28] have shown that this scattering information can also be of diagnostic value. It is important therefore to obtain the best possible estimate of the scattering parameters if these are to be subsequently used [28]. This will be the subject of a future paper.

## 5. Conclusion and discussion

In this paper we have introduced full Mie theory and the manipulations required to deal with the complex refractive index in the calculation of  $Q_{\text{sca}}$ . These functions are slow to compute due to spherical Bessel and Hankel functions, however we have shown that using the Accelereyes Jacket toolbox for GPU computing within Matlab can speed calculations up significantly. This speed gain will improve in the near future as the Jacket toolbox is optimised further.

The use of the reference spectrum in the RMieS-EMSC which is the weak point of the algorithm has been optimised such that before the newly corrected spectrum is used as the new reference, it is modelled and fitted using Gaussians placed at specific positions located from the second derivative spectrum. This produces a noise free spectrum with no negative values or any scattering remnants that were previously a potential problem.

We have subsequently shown that the optimised RMieS-EMSC algorithm incorporating the full Mie theory can be used to correct a typical data set recorded from 280 single prostate cancer cells. The

correction converts unusable raw data into a series of spectra that reflect the natural variations within a typical cell line population. Comparison with baseline correction methods that are freely available within standard commercial FTIR software show that if biochemical information relating to proteins is required some form of RMieS correction should be used. Importantly we demonstrate that the previous RMieS-EMSC correction using the van de Hulst approximation corrects the spectra as well as the RMieS-EMSC based on the full Mie theory but the latter (although not shown here) will give better more accurate scattering parameters.

## APPENDIX

### Mie theory

The Mie scattering efficiency,  $Q_{\text{sca}}$ , is defined as:

$$Q_{\text{sca}} = \frac{2}{x^2} \sum_{n=1}^{\infty} (2n+1) \left\{ |a_n|^2 + |b_n|^2 \right\} \quad (\text{A1})$$

$$a_n = \frac{\psi'_n(y) \psi_n(x) - \eta \psi_n(y) \psi'_n(x)}{\psi'_n(y) \zeta_n(x) - \eta \psi_n(y) \zeta'_n(x)} \quad (\text{A2})$$

$$b_n = \frac{\eta \psi'_n(y) \psi_n(x) - \psi_n(y) \psi'_n(x)}{\eta \psi'_n(y) \zeta_n(x) - \psi_n(y) \zeta'_n(x)} \quad (\text{A3})$$

$$\psi_n(z) = \sqrt{\frac{\pi z}{2}} J_{n+1/2}(z) \quad (\text{A4})$$

$$\xi_n(z) = \sqrt{\frac{\pi z}{2}} H_{n+1/2}^{(2)}(z) \quad (\text{A5})$$

$$x = 2\pi r \tilde{v} \quad (\text{A6})$$

$$y = 2\pi r \tilde{v} \eta \quad (\text{A7})$$

$$\eta = \eta_1 / \eta_2 \quad (\text{A8})$$

Where,  $\eta_1$  = complex refractive index of scattering particle,  $\eta_2$  = complex refractive index of surrounding medium,  $r$  = particle radius;  $J$  and  $H^{(2)}$  denote the 1st kind Bessel and 2nd kind Hankel functions respectively.  $z = x$  or  $y$ .

**Acknowledgements** We would like to thank (i) the EPSRC/RSC analytical fund for studentships for PB and EJ; (ii) James Malcolm from Accelereyes for his help with the Bessel and Hankel functions in the Accelereyes Jacket toolbox for Matlab; (iii) all the staff on the SMIS beamline for their assistance during our visit and (iv) the Agricultural Food Research Foundation.



**Paul Bassan** received his B.Sc. in Physics at the University of Warwick U.K. in 2006 and his an M.Sc. in Medical Physics at the University of Manchester U.K. in 2007. In October 2007 he joined the Gardner group at

the University of Manchester and is currently researching the fundamental principles of mid-IR light scattering (namely Mie scattering) in single cell and tissue samples as used in FTIR spectroscopy. In the future he plans to apply the RMieS-EMSC to FTIR imaging data with the use of parallel computing using graphics processing unit (GPU) computing.



**Achim Kohler** received his master degree in physics at the University of Freiburg in Germany in 1997 and obtained his Ph.D. in the field of Quantum Chaos in theoretical physics in 1999. After a short period as a programmer in software industry developing graphical information systems, he started as a research scientist at Nofima Mat. At Nofima Mat his research focuses on biospectroscopy and data modelling for biospectroscopy and biology in general. He is the leader of the research program on data modelling and the founder of the virtual Centre for Biospectroscopy and Data modelling at Campus Ås.

trscopy and data modelling for biospectroscopy and biology in general. He is the leader of the research program on data modelling and the founder of the virtual Centre for Biospectroscopy and Data modelling at Campus Ås.



**Harald Martens** has a background in biochemical engineering and has worked in the field of chemometrics since the early 70-ies, focusing on the conversion of high-dimensional biological and chemical measurements into meaningful chemical information. He took part in developing the PLS regression method in the early 80-ies and the Multiplicative Signal Correction (MSC). He defended his dr.techn. thesis at the Norwegian University of Science and Technology, Trondheim, Norway in 1985. In the 90-ies he worked on model-based compression of video signals. He is currently Professor II, CIGENE/Inst.Math.Sci. & Technol., Norwegian U. of Life Sciences, Ås and Senior research scientist, Norwegian Food Research Institute, Ås, Norway, and Adjunct professor, University of Copenhagen, Faculty of Life Sciences, Denmark.



**Joe Lee** received his B.Sc. in chemistry at the University of Leeds in 1951 and his Ph.D. in physical chemistry at the same university in 1955. In 1956 he was appointed as an ICI Research Fellow in the Inorganic and Physical Chemistry Department of the University of Liverpool. In 1959, he moved to the then Manchester College of Science and Technology as Lecturer in Chemistry. Before retiring in 1997 as a Senior Lecturer at the University of Manchester Institute of Science and Technology (UMIST), he had lectured in all the standard aspects of physical chemistry and in the proper application of mathematics and statistics to chemistry. His research interests included high resolution NMR, UV-visible emission spectroscopy of flow discharge systems and theoretical aspects of temperature-programmed gas chromatography. Since retirement, as a Visiting Research Fellow in UMIST and later in the University of Manchester, he has been involved with theoretical aspects of infrared spectroscopy of thin-films and more recently of cancer cells, particularly in the understanding of scattering distortions.



**Edward Jackson** obtained a B.Sc. in Zoology from Leeds University in 2005 and a master of Science in Analytical Biosciences and Drug Design from the University of Salford in 2007. Currently, Edward is studying

for a Ph.D. in analytical science at the University of Manchester, based at the Manchester Interdisciplinary Biocentre. Edward's research uses a combination of time-of-flight secondary ion mass spectrometry (ToF-SIMS) and synchrotron source based infrared microscopy to study cancer at a cellular level.



**Nicholas Lockyer** received a Ph.D. in 1996 from the UMIST on laser ionisation mass spectrometry of biomolecules. After a period of post-doctoral work developing imaging time-of-flight secondary ion mass spectrometry (ToF-SIMS) instrumentation he was awarded a 2 year Special Research Fellowship from the Leverhulme Trust. In 2002 he was appointed Lecturer in the Department of Chemistry at UMIST. In 2004 he joined the School of Chemical Engineering and Analytical Science at the University of Manchester where he currently holds a Senior Lectureship. His research interests include the development of imaging mass spectrometry, particularly applications in biology and medicine and associated fundamental aspects of the technique.



**Paul Dumas** is Research Director at the CNRS, and beamline manager of the Infrared Station at the French National Synchrotron facility, SOLEIL. He promoted this analytical technique in France, at LURE, where he was responsible for the design and construction of an infrared microscope beamline at LURE, the French National synchrotron center,

until its closure in December 2003. His research program, apart for surface science, includes studies of individual cells in biology, geological inclusions, astrophysics, soft matter and archeology.



**Mick Brown** obtained a Ph.D. in molecular virology in 1995 from the Institute for Animal Health, Compton/University of Reading, UK. He then spent five years within the Department of Molecular Biology at the Paterson Institute for Cancer Research, Manchester, UK using avipox viral vec-

tors as potential immunotherapeutic agents against cervical cancer. He then moved to the ICRF Clinical Centre, St James Hospital in Leeds, UK where he continued his interests in cancer immunotherapy utilising embryonic stem cells and avipox viral vectors. Since 2002 he has been an associate scientist within Noel Clarke's Genito Urinary Cancer Research Group where he leads the laboratory research team investigating urological malignancies with a particular interest in the mechanism of metastasis and the development of novel diagnostic/prognostic biomarkers.



**Noel Clarke** is Professor of Urological Oncology at Manchester University. He qualified in London in 1981, completing his Fellowship of the Royal College of Surgeons in 1985 and his research thesis in prostate cancer biology and metastasis in 1990. He was appointed as Consultant Urological Surgeon to the Christie and Salford Royal Hospitals in Manchester in 1993 and has specialized in the management of Urological Cancer and complex pelvic and retroperitoneal tumours. He is currently the Director of the Manchester University GU Cancer Research Group and the pan-Manchester MCRC biobank. His research encompasses basic science, translational cancer studies and Phase 2/3 trials in Urological Cancer.



**Peter Gardner** obtained a Ph.D. in surface vibrational spectroscopy in 1988 from the University of East Anglia. After four years at the Fritz Haber Institute (MPG) in Berlin he joined the Chemistry Department at the University of Cambridge in 1992 studying catalytic systems and then in 1994 moved to the Chemistry Department at UMIST. In 2004, he joined the School of Chemical Engineering and Analytical Science and is part of the centre for instrumentation and analytical science (CIAS). He has a keen interest in understanding the underlying basis of spectral discrimination in biological systems. His group have also been influential in solving the long standing problem of spectral distortions in infrared spectra of single eukaryotic cells and in 2009 were the first to attribute the problem to the phenomenon of resonant Mie scattering (RMieS).

## References

- [1] N. Jamin, P. Dumas, J. Moncuit, W. H. Fridman, J. L. Teillaud, L. G. Carr, and G. P. Williams, *P. Nat. Acad. Sci. USA* **95**, 4837–4840 (1998).
- [2] P. Lasch, A. Pacifico, and M. Diem, *Biopolymer* **67**, 335–338 (2002).
- [3] P. Lasch, M. Boese, A. Pacifico, and M. Diem, *Vib. Specrosc.* **28**, 147–157, (2002).
- [4] P. Dumas and L. Miller, *Vib. Spectroc.* **32**, 3–21 (2003).
- [5] P. Dumas, N. Jamin, J. L. Teillaud, L. M. Miller, and B. Beccard, *Faraday Discuss.* **126**, 289–302 (2004).
- [6] E. Gazi, J. Dwyer, N. P. Lockyer, P. Gardner, J. Miyan, C. A. Hart, M. D. Brown, J. H. Shanks, and N. W. Clarke, *Biopolymers* **77**, 18–30 (2005).
- [7] E. Gazi, J. Dwyer, N. P. Lockyer, J. Miyan, P. Gardner, C. A. Hart, M. D. Brown, and N. W. Clarke, *Vib. Spectrosc.* **38**, 193–201 (2005).
- [8] E. Gazi, P. Gardner, N. P. Lockyer, C. A. Hart, N. W. Clarke, and M. D. Brown, *J. Lipid Res.* **48**, 1846–1856 (2007).
- [9] D. A. Moss, M. Keese, and R. Pepperkok, *Vib. Spectrosc.* **38**, 185–191 (2005).
- [10] M. J. German, A. Hammiche, N. Ragavan, M. J. Tobin, L. J. Cooper, N. J. Fullwood, S. S. Matenhelia, A. C. Hindley, C. M. Nicholson, N. J. Fullwood, H. M. Pollock, and F. L. Martin, *Biophys. J.* **90**, 3783–3795 (2006).
- [11] J. Sule-Suso, D. Skingsley, G. D. Sockalingum, A. Kohler, G. Kegelaer, M. Manfait, and A. El Haj, *Vib. Spectrosc.* **38**, 179–184 (2005).
- [12] M. J. Baker, C. Clarke, D. Démoulin, J. M. Nicholson, F. M. Lyng, H. J. Byrne, C. A. Hart, M. D. Brown, N. W. Clarke, and P. Gardner, *Analyst*, DOI: 10.1039/b920385k (2010).
- [13] M. Romeo, B. Mohlenhoff, and M. Diem, *Vib. Spectrosc.* **42**, 9–14 (2006).
- [14] P. Bassan, H. J. Byrne, J. Lee, F. Bonnier, C. Clarke, P. Dumas, E. Gazi, M. D. Brown, N. W. Clarke, and P. Gardner, *Analyst* **134**, 1171–1175, (2009).
- [15] J. Lee, E. Gazi, J. Dwyer, M. D. Brown, N. W. Clarke, and P. Gardner, *Analyst* **132**, 750–755 (2007).
- [16] P. Bassan, H. J. Byrne, F. Bonnier, J. Lee, P. Dumas, and P. Gardner, *Analyst* **134**, 1586–1593 (2009).
- [17] B. Mohlenhoff, M. Romeo, M. Diem, and B. R. Wood, *Biophys. J.* **88**, 3635–3640 (2005).
- [18] A. Kohler, J. Sule-Suso, G. D. Sockalingum, M. Tobin, F. Bahrami, Y. Yang, J. Pijanka, P. Dumas, M. Cotte, D. G. van Pettius, G. Parkes, and H. Martens, *Appl. Spectrosc.* **62**, 259–266 (2008).
- [19] P. Bassan, A. Kohler, H. Martens, J. Lee, H. J. Byrne, P. Dumas, E. Gazi, M. Brown, N. Clarke, and P. Gardner, *Analyst* **135**, 268–277 (2010).
- [20] H. C. van de Hulst, *Light Scattering by Small Particles* (John Wiley and Sons, New York; Chapman and Hall, London, 1957) 179.
- [21] G. Mie, *Ann. Phys.* **330**, 377–445 (1908).
- [22] T. J. Harvey, A. Henderson, E. Gazi, N. W. Clarke, M. Brown, E. Correia Faria, R. D. Snook, and P. Gardner, *Analyst* **132**, 292–295 (2007).
- [23] E. Gazi, T. J. Harvey, M. D. Brown, N. W. Clarke, N. P. Lockyer, and P. Gardner, *Vib. Spectrosc.* **50**, 99–105 (2009).
- [24] P. Dumas, F. Polack, B. Lagarde, O. Chubar, J. L. Giorgetta, and S. Lefrancois, *Infrared Phys. Technol.* **49**, 152–160 (2006).
- [25] H. Martens, J. P. Nielsen, and S. B. Engelsen, *Anal. Chem.* **75**, 394 (2003).
- [26] V. Erukhimovich, M. Talyshinsky, Y. Souprun, and M. Huleihel, *J. Mol. Struct.* **792–793**, 99–103 (2006).
- [27] W. Tanthanuch, K. Thumanu, C. Lorthongpanich, R. Parnpai, and P. Heraud, *J. Mol. Struct.* **967**, 189–195 (2010).
- [28] J. K. Pijanka, A. Kohler, Y. Yang, P. Dumas, S. Chiosrichan, M. Manfait, G. D. Sockalingum, and J. Sule-Suso, *Analyst* **134**, 1176–1181 (2009).