View Article Online / Journal Homepage / Table of Contents for this issue

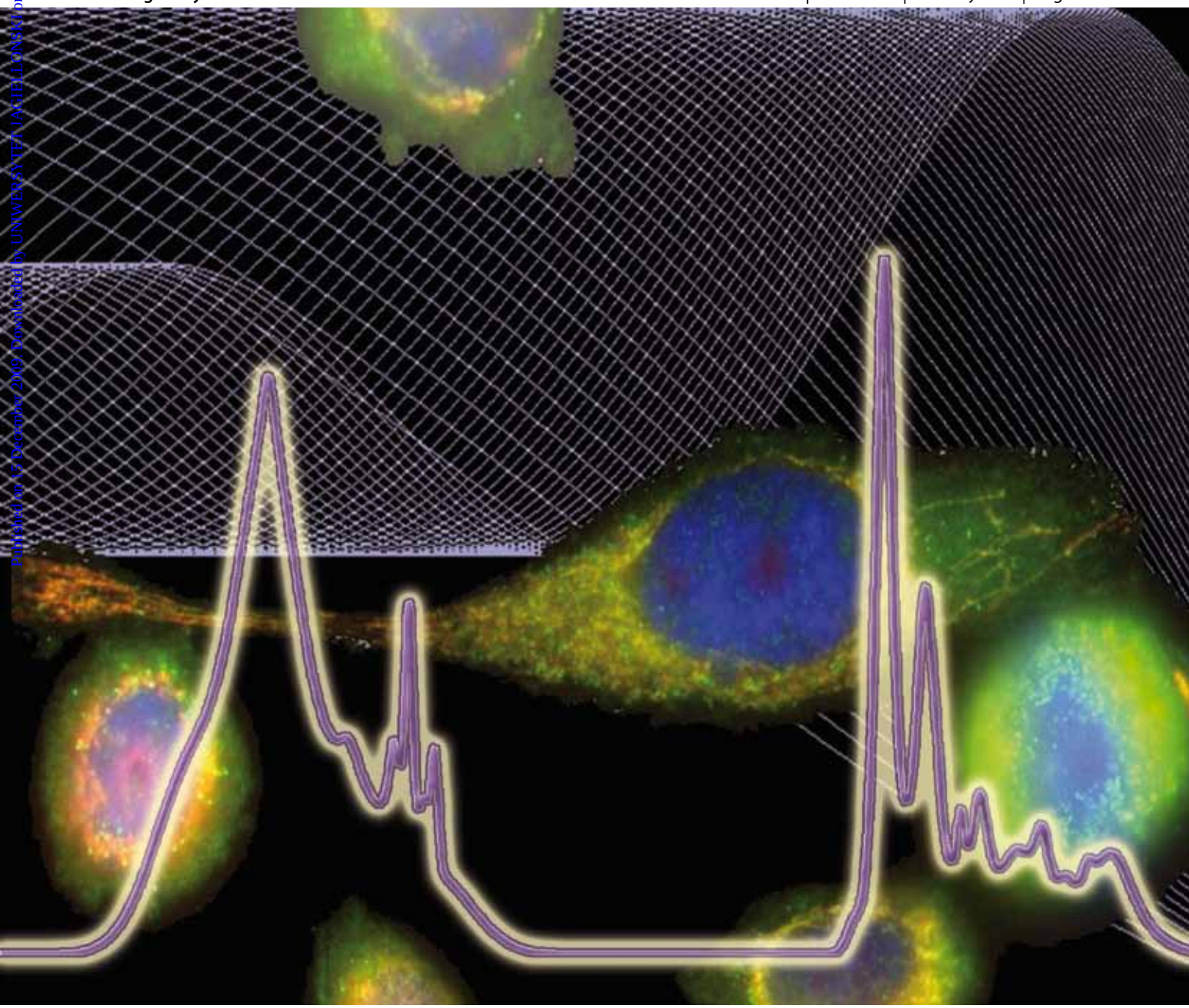
Published on 15 December 2009. Downloaded by UNIVERSANTEA IACIB LIONS on 11/4/2022 1:53:39 PM.

Analyst

Interdisciplinary detection science

www.rsc.org/analyst

Volume 135 | Number 2 | February 2010 | Pages 197–424



ISSN 0003-2654

RSC Publishing

EDITORIAL
Lisa Hall
Reflections: quality and globalisation in detection science

HOT ARTICLE
Peter Gardner *et al.*
Resonant Mie Scattering (RMieS) correction of infrared spectra from highly scattering biological samples

0003-2654(2010)135:2;1-N

View Article Online

Resonant Mie Scattering (RMieS) correction of infrared spectra from highly scattering biological samples

Paul Bassan,[a] Achim Kohler,[bc] Harald Martens,[bc] Joe Lee,[a] Hugh J. Byrne,[d] Paul Dumas,[e] Ehsan Gazi,[f] Michael Brown,[f] Noel Clarke[fgh] and Peter Gardner*[a]

Received 7th October 2009, Accepted 30th November 2009
First published as an Advance Article on the web 15th December 2009
DOI: 10.1039/b921056c

Published on 15 December 2009. Downloaded by UNIWERSYTET JAGIELLONSKI on 11/4/2022 1:53:39 PM.

Infrared spectra of single biological cells often exhibit the 'dispersion artefact' observed as a sharp decrease in intensity on the high wavenumber side of absorption bands, in particular the Amide I band at ~1655 cm^{-1}, causing a downward shift of the true peak position. The presence of this effect makes any biochemical interpretation of the spectra unreliable. Recent theory has shed light on the origins of the 'dispersion artefact' which has been attributed to resonant Mie scattering (RMieS). In this paper a preliminary algorithm for correcting RMieS is presented and evaluated using simulated data. Results show that the 'dispersion artefact' appears to be removed; however, the correction is not perfect. An iterative approach was subsequently implemented whereby the reference spectrum is improved after each iteration, resulting in a more accurate correction. Consequently the corrected spectra become increasingly more representative of the pure absorbance spectra. Using this correction method reliable peak positions can be obtained.

1. Introduction

Cytological examination for disease diagnosis is being used more extensively since the analysis of a few cells to diagnose a disease state can obviate the need for more invasive intervention, for example, taking a biopsy. It is still the case, however, that the analysis of cells by eye under an optical microscope, even by a highly trained cytopathologist is both a time-consuming and subjective process. Specific disease recognition tools such as immunohistochemical stains offer some assistance in the diagnosis procedure but these too are not without problems,[1] thus there is a real need to develop more objective methods of analysis. In recent years, there has been increasing interest in using infrared micro-spectroscopy to study single biological cells.[2–13] Infrared spectroscopy could, in principle, improve the sensitivity and specificity of such analysis since it is completely objective, based upon biochemical changes rather than cellular architecture. Several studies have already shown that FTIR can be used to detect spectral changes in malignant and pre-malignant cells[14,15] and have shown results that compare very favourably with visual cytology.[16] Measuring an infrared absorbance spectrum from a single biological cell, however, is an inherently flawed process. Given that the diameter of typical human cells are in the region of ~8–30 μm and the size of nuclei and other organelles ranges from 1–10 μm, it is clear that mid-infrared radiation with wavelengths of 3–10 μm will scatter strongly from such samples. This scattering, dominated by Mie scattering, will significantly distort the measured spectrum such that it appears significantly different from the pure absorbance spectrum.[13,17–24] Under such circumstances, in the absence of some form of correction, band shape, spectral positioning and intensity of signature features are unreliable and cannot be used with any certainty to evaluate cellular biochemistry.[13,22] This problem is compounded in comparative studies of drug–cell interaction where the action of a drug induces a change in cell morphology that alters the scattering profile of the cell compared with a control.[12] An increase in scattering is often observed as cells become more rounded, for example, due to the action of a cytotoxic agent. This causes the so-called Amide I band, usually the strongest band in the spectrum, to exhibit an apparent loss of intensity and a downward shift of the band centre. This can be misinterpreted as a change in protein structure induced by the drug. It is essential therefore that such distortions in IR spectra, arising from scattering effects, should be removed to facilitate recovery of the pure absorbance spectrum. Only then can spectra be reliably compared from one cell to another and the influence of anticancer agents and other cytotoxins be evaluated. To date, the most successful method for removing spectral distortions, including the effects of Mie scattering, has been the application of the Extended Multiplicative Signal Correction (EMSC)

[a]*School of Chemical Engineering and Analytical Science, Manchester Interdisciplinary Biocentre, University of Manchester, 131 Princess Street, Manchester, UK M1 7DN. E-mail: Peter.gardner@manchester.ac.uk; Fax: +44 (0) 161 306 5201; Tel: +44 (0) 161 306 4463*
[b]*Nofima Mat, Centre for Biospectroscopy and Data Modelling, Ås, Osloveien 1, 1430 Ås, Norway*
[c]*CIGENE, Department of Mathematical Sciences and Technology, Norwegian University of Life Sciences, 1430 Ås, Norway*
[d]*Focas Research Institute, Dublin Institute of Technology, Kevin Street, Dublin 8, Ireland*
[e]*Synchrotron SOLEIL, L'Orme des Merisiers, BP48 - Saint Aubin, 91192 Gif-sur-Yvette Cedex, France*
[f]*Genito Urinary Cancer Research Group, School of Cancer, Enabling Sciences and Technology, Paterson Institute for Cancer Research, The University of Manchester, Manchester Academic Health Science Centre, The Christie NHS Foundation Trust, Manchester, UK M20 4BX*
[g]*Department of Urology, The Christie NHS Foundation Trust, Manchester, UK M20 4BX*
[h]*Department of Urology, Salford Royal NHS Foundation Trust, Salford, UK M6 8HD*

 This journal is © The Royal Society of Chemistry 2010

View Article Online

Published on 15 December 2009. Downloaded by UNIWERSYTET JAGIELLONSKI on 11/4/2022 1:53:39 PM.

algorithm,[20] which works well in most cases, particularly where the Mie scattering is weak and where the spectra do not show strong distortion (dispersion artefact) of the Amide I band.[12] Where the dispersion artefact is strong, however, conventional EMSC struggles to correct this important region of the spectrum.

Bassan *et al.* have recently shown that the principal origin of the dispersion artefact is a process termed *resonant* Mie scattering (RMieS)[22] in both FTIR transmission and transflection mode measurements. Briefly, this relates to the fact that the Mie scattering efficiency is dependent upon the refractive index of the sample and this changes on passing through an absorption resonance.[22] Measurements in transflection mode are subject to further distortion when the cellular reflection is comparable to the (doubly) transmitted intensity.[23] In Part A of this paper we illustrate the problem of spectral distortions with infrared data collected from prostate cancer cells, known to give rise to a strong Mie scattering/dispersion artefact, and show that the spectral distortions can be corrected for using a modified version of EMSC (RMieS-EMSC). However, as with all correction algorithms it is important to know 'how well has the distorted spectrum been corrected'? To this end, in Part B of the paper, simulated data are artificially distorted and then corrected using the new correction algorithm. Since the simulated pure absorption spectrum is known, the effectiveness of the correction procedure can be evaluated.

Part A

2. Experimental

2.1 Cell culture

Cultures of PC-3 cells, a human prostate cancer cell line, were grown on 70% v/v ethanol-sterilised, CaF_2 plates (Crystran Pool, UK) using standard protocols.[7,26] The cells were cultured in Ham's F12 with 7% FCS and 2 mM L-glutamine at 37 °C in a humidified atmosphere of 5% CO_2. Reagents were purchased from Sigma-Aldrich (Poole, UK) and tissue culture media were obtained from Invitrogen (Paisley, UK). Once the cells were 70% confluent the CaF_2 plates were removed from the growth medium and fixed in 4% formalin in phosphate-buffered saline (PBS) for 20 min at room temperature, washed in distilled water to remove residual PBS from their surface, dried under ambient conditions and stored in a desiccator prior to analysis.

2.2 Infrared microscopy

The synchrotron FTIR micro-spectroscopy data were recorded at the synchrotron SOLEIL on the SMIS beamline, details of which can be found elsewhere.[26] The spectra were obtained using a Nicolet Continuum XL microscope equipped with a 50 μm MCT detector. Spectra were recorded at 4 cm^{-1} resolution. The size of the aperture was adjusted to match the diameter of the cell such that it was fully illuminated.

3. Infrared results and data analysis

3.1 Raw infrared spectrum

Fig. 1a(i), shows an optical image of two PC-3 cells. The cell on the right is typical of this cell line, being slightly flattened and having a diameter of ~26 μm. The cell on the left is smaller, ~13 μm, more rounded with very little cytoplasm and is optically denser. Although very much in the minority, these morphologically smaller cells are often observed in PC-3 cultures and are believed to be cells that have divided just prior to fixation. PC-3 cells are unusual in that they often divide perpendicularly to the substrate and the unattached daughter cell is often washed away during the fixation process. The infrared spectrum of the small rounded cell is shown in Fig. 1a(ii). The spectrum looks highly distorted compared to a pure absorbance† spectrum, showing a broad hump in the baseline between 2000 and 4000 cm^{-1}, part of the broad oscillation of the baseline, associated with classic (non-resonant) Mie scattering, as well as the dispersion artefact;

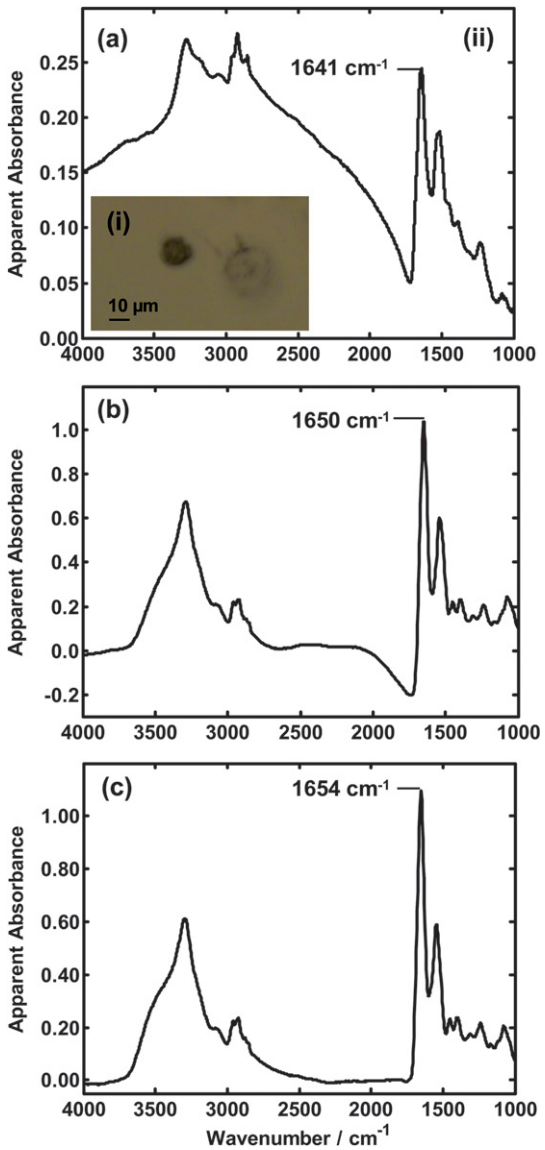


Fig. 1 (a) Optical image of a PC-3 cells (i) and an IR spectrum of the smaller cell (ii). (b) Corrected IR spectrum using an existing Mie scattering-EMSC algorithm. (c) Corrected spectrum using new RMieS-EMSC algorithm.

† All absorbances quoted in this paper are decadic, *i.e.* $-\log_{10}(I/I_0)$.

View Article Online

Published on 15 December 2009. Downloaded by UNIWERSYTET JAGIELLONSKI on 11/4/2022 1:53:39 PM.

the pronounced dip in absorbance at 1750 cm^{-1} is predominantly due to RMieS. This profile is typical of many small single-cell infrared spectra and supports the assumption that scattering in cells is primarily from the cell nucleus.[22,23] In Fig. 1a(ii) we observe that the position of the Amide I band is at 1641 cm^{-1}, which if interpreted as a signal with no scattering artefact would be indicative of a predominance of β-sheet and random coil secondary protein structure within the cell.[27]

3.2 EMSC correction

An explanation of the EMSC will be given fully in Part B of this paper. However, the Mie scattering-EMSC[20,25] has been applied to the raw spectrum in Fig. 1a, to produce the corrected spectrum shown in Fig. 1b. As can be seen, the large baseline oscillation has been removed but the dispersion artefact indicated by the reduced intensity on the higher wavenumber side of the Amide I band is still present. The peak of the Amide I band has up-shifted by 9 cm^{-1} to 1650 cm^{-1}. This is significant since, again taken at face value, it would imply predominance of random coil plus turns and bends and rather less β-sheet protein secondary structure. However, from our previous work on PMMA microspheres the presence of the dispersion artefact is likely to significantly influence both the shape and position of the Amide I band.[22]

3.3 Resonant Mie scattering (RMieS)/EMSC correction

In order to correct for the dispersion artefact caused by RMieS we have developed a new modified version of EMSC which is also discussed in further detail in Part B. This correction algorithm has been applied to the data in Fig. 1a to produce the spectrum in Fig. 1c. As can be seen, the spectrum now resembles a conventional FTIR absorbance spectrum with a relatively flat baseline and no apparent dispersion artefact. In addition, the Amide I band has shifted further to 1654 cm^{-1}, which suggests that the secondary structure present is largely α-helix.

At this point it is reasonable to assume that the RMieS corrected spectrum closely represents the pure absorbance spectrum that would have been measured in the absence of any scattering. Thus for the first time it would appear that the band positions can be interpreted in terms of the real biochemistry present in the cell. However, since we can never know for sure the 'correct' absorbance spectrum of an inherently strong scattering sample, we cannot test the above assumption using real biological cells. We have, therefore attempted to evaluate the accuracy of the correction algorithm using simulated data.

Part B

4. Evaluating the algorithm

It is important for the understanding of this paper to briefly outline the various processes performed by the correction algorithms. Since the RMieS-EMSC algorithm builds on the standard EMSC algorithm it is useful to start our discussion at this point.

4.1 Extended Multiplicative Signal Correction (EMSC)

The EMSC is a model-based multivariate data pre-processing method and based on linear statistical regression modelling.[25] However, it can also be extended to handle non-linear effects, *e.g.* physical effects such as Mie scattering, that otherwise require non-linear mathematical modelling. This is done by representing the non-linear mathematical modelling by a low-rank multivariate bi-linear model.[20] Offsets and baseline slopes are removed effectively, whilst the multiplicative part of the algorithm compensates for optical path length differences, essentially normalising the spectra. This is done by taking a reference spectrum which can be the mean spectrum of the sample data set or a spectrum with similar spectral features, hereafter referred to as Z_{Ref}. The algorithm takes the reference spectrum and attempts to recreate the raw spectrum to be corrected (Z_{Raw}) by adding an offset, a slope and amplifying the reference by multiplication. This can be summarised algebraically, where symbols with arrows above represent vectors, *i.e.* a column of spectral intensity values

$$\vec{Z}_{Raw} = c + m\vec{\tilde{\nu}} + h\vec{Z}_{Ref} + \vec{E} \quad (1)$$

where Z_{Raw} = the raw spectrum, Z_{Ref} = reference spectrum, c = constant value for spectrum offset, m = gradient of the sloping baseline, $\tilde{\nu}$ = wavenumber (reciprocal wavelength, *i.e.* λ^{-1}), h = multiplicative scaling factor, and $\vec{E}$ = un-modelled residual information.

By finding the values of c, m and h using a least squares linear regression method, the spectrum can be corrected. An extension to this was published by Kohler *et al.*[20] where oscillating baseline variations due to Mie scattering, obtained by a linear sub-space model of Mie scattering effects as described in the following section, were corrected for by adding an additional term to eqn (1).

4.2 Non-resonant Mie scattering EMSC

In 1957, van de Hulst[28] published an approximation equation for the Mie scattering efficiency, Q, which is a simpler and less computationally intensive version of the original theory published in 1908 by Mie.[29] The Mie approximation, eqn (2), explained further in eqns (3)–(5) is given as

$$Q = 2 - \frac{4}{\rho}\sin\rho + \frac{4}{\rho^2}(1 - \cos\rho) \quad (2)$$

where

$$\rho = \frac{2\pi d(n-1)}{\lambda} \quad (3)$$

and where, n and d denote the ratio of the real refractive indices of particle and surrounding medium, and the diameter of the scattering particle respectively. In this case, the medium is air for which the real refractive index is essentially 1, hence n simplifies to the real refractive index of the scattering particle. For convenience, the ρ term was simplified to

$$\rho = \frac{\alpha}{\lambda} \quad (4)$$

so

$$\alpha = 2\pi d(n-1) \quad (5)$$

 This journal is © The Royal Society of Chemistry 2010

View Article Online

Published on 15 December 2009. Downloaded by UNIWERSYTET JAGIELLONSKI on 11/4/2022 1:53:39 PM.

The introduction of this α parameter enabled a single variable to describe the product of the refractive index and particle diameter. The model of the expected Mie contribution in eqn (2) is a non-linear function of ρ which in turn is a non-linear function of $\tilde{\nu}$ and α in eqn (4). If these non-linear functions were to be implemented directly into the pre-processing, a non-linear and complicated parameter estimation problem would have to be solved. Instead, the non-linear function can be approximated by a multivariate bi-linear model and incorporated into the linear EMSC model;[20] thereby, the parameter estimation is done by a simple one-step multivariate linear regression. To this purpose, 200 different α values were chosen to cover the range of parameter values for d and n considered relevant for the present application:

d: 2 to 20 μm

n: 1.1 to 1.5

This resulted in 200 output spectra as a function of $\tilde{\nu}$. The 200 'Q curves' are decomposed using a non-mean-centred principal component analysis (PCA) to find the principal components of the data matrix. The first six loadings, $p_1 \cdots p_6$, summarised 99.99% of the sum-of-squares in the 200 simulated spectra. They were thus considered to be an adequate bi-linear approximation of the non-linear Mie model under the present conditions. These loadings are used to produce a new correction algorithm, summarised algebraically below:

$$\vec{Z}_{\mathrm{Raw}} = c + m\vec{\tilde{\nu}} + h\vec{Z}_{\mathrm{Ref}} + \sum_{i=1}^{6} \vec{g}_i p_i + \vec{E} \tag{6}$$

The fourth term in the equation (containing the summation) is where the variations from Mie scattering are covered. A precise weighting for each of the six loadings is added, controlled by the parameters 'g_i' which are calculated during the least squares linear regression. Since the loading vectors are orthogonal, the parameter estimation by least squares regression is very stable. The vector $\vec{E}$ represents the un-modelled residual variance which could not be described by the EMSC model. Ideally, the residual should be zero, indicating that the model described all features; however, this is rarely the case in practice; in fact, all the interesting chemical variations remain in the residual spectra, unless they have been modelled explicitly by including, for example, analyte spectra in the EMSC model (not used in this paper). The described algorithm is successful at removing the smooth oscillations due to Mie scattering; however, the dispersion artefact often remains.

4.3 Resonant Mie scattering (RMieS)/EMSC correction algorithm

Recently, the origins of the so-called dispersion artefact have been understood and linked to resonant Mie scattering (RMieS), connecting the broad oscillations and the sharp decrease in apparent absorbance on the higher wavenumber side of absorption bands.[22] Knowledge of the origin of the phenomenon enables a new correction algorithm to be constructed, presented in this paper that removes both broad oscillations in the baseline and the dispersion artefact, both of which derive from resonant Mie scattering (RMieS).

Bassan *et al.* documented that the dispersion artefact is predominantly due to resonant Mie scattering (RMieS) caused by a changing real refractive index near an absorption band.[22] This causes some degree of index matching meaning that the efficiency with which the photons are scattered at this wavenumber is reduced to almost zero, visually interpreted as a sharp decrease in absorbance.

The algorithm by Kohler *et al.*[20] therefore needs modification to correct for resonant Mie scattering. The spectrum of the real refractive index of a material can be calculated from its absorbance spectrum, which is proportional to the imaginary refractive index (k):

$$n(\tilde{\nu}) = \langle n \rangle + \frac{2}{\pi}\mathbf{P}\int_0^\infty \frac{s \times k(\tilde{\nu})}{s^2 - \tilde{\nu}^2}\mathrm{d}s \tag{7}$$

where $\langle n \rangle$ is the average real refractive index, and $\mathbf{P}$ denotes the Cauchy principal of improper integrals, needed in this case when $s = \tilde{\nu}$ as a division by zero occurs creating a singularity. The solution to this is to perform two integrations either side of the singularity.

The output from the Kramers–Kronig transform[30,31] is the refractive index spectrum minus the average real refractive index, hereafter referred to as n_{KK}.

$$n_{\mathrm{KK}} = n(\tilde{\nu}) - \langle n \rangle = \frac{2}{\pi}\mathbf{P}\int_0^\infty \frac{s \times k(\tilde{\nu})}{s^2 - \tilde{\nu}^2}\mathrm{d}s \tag{8}$$

The k term can be replaced by the reference spectrum as they are nearly proportional:

$$Z_{\mathrm{Ref}}(\tilde{\nu}) \propto k_{Z_{\mathrm{Ref}}}(\tilde{\nu}) \tag{9}$$

The $2/\pi$ factor can also been omitted as we are only interested in the proportional relationship of n_{KK} and the Kramers–Kronig transform of Z_{Ref}:

$$n_{\mathrm{KK}}(\tilde{\nu}) \propto \mathbf{P}\int_0^\infty \frac{s \times k(\tilde{\nu})}{s^2 - \tilde{\nu}^2}\mathrm{d}s \tag{10}$$

The average refractive index of each sample is again unknown, as is the imaginary refractive index. The n_{KK} spectrum is arbitrarily normalised so that its minimum value is -1, the reason for this is explained later. To construct a refractive index for insertion into eqn (11), two terms a and b need to be defined as the average refractive index and an amplification factor for n_{KK} respectively:

$$n = a + bn_{\mathrm{KK}} \tag{11}$$

The parameter b is required as the Z_{Ref} used is not the correct input for the Kramers–Kronig transform. It is, however, directly proportional and so a scaling parameter can be used to compensate. The refractive index cannot go below a value of 1, and this is ensured by carefully controlling the value of b. If an average refractive index $a = 1.3$ is used (typical for a biological sample), then b can range from 0 to 0.3, resulting in the minimum value being 1. The particle diameter d is the last parameter which needs to be varied to cover many scattering possibilities giving a total of three parameters: a, b and d.

For each parameter, 10 equidistant values were used between the ranges:

View Article Online

Published on 15 December 2009. Downloaded by UNIWERSYTET JAGIELLONSKI on 11/4/2022 1:53:39 PM.

a: 1.1 to 1.5
b: 0 to (*a* − 1)
d: 4 to 40 μm

This results in 1000 permutations. This equivalence data set is compressed by PCA and approximated by a small number of loadings (seven in this case, explaining 99.9% of the total sum-of-squares in the 10^3 simulated spectra). The total number of 'descriptive vectors' in the linear EMSC model is now ten, consisting of seven loadings, the reference spectrum, plus the constant and sloping baselines. All the model parameters are estimated simultaneously by multiple linear regression solved by least squares estimation. As in the Mie scattering-EMSC,[22] the parameter estimation is stable due to the orthogonality of the loadings.

The remainder of the algorithm is exactly the same as the previously published Mie scattering-EMSC,[20] except that seven loadings are now used from the data matrix of 10^3 RMieS Q curves:

$$\vec{Z}_{\mathrm{Raw}} = c + m\vec{\tilde{\nu}} + h\vec{Z}_{\mathrm{Ref}} + \sum_{i=1}^{7} \vec{g}_i p_i + \vec{E} \qquad (12)$$

5. Evaluating the RMieS algorithm

5.1 Evaluation methodology

The new algorithm subtracts a curve which is the sum of a constant value offset, c, a sloping baseline, $m\tilde{\nu}$ and the RMieS curve, Q which is described by the summation term in eqn (12). If the pure absorbance spectra of some of the major and most strongly varying chemical constituents in the samples correlate (overlaps) with some of the spectra in the EMSC model, the EMSC algorithm potentially can remove chemical information from the pure absorbance spectrum of the sample which is undesirable as results may be distorted and unreliable. This problem can be alleviated by including important constituent spectra in the EMSC model. However, for single-cell spectra these are presently considered unknown. Hence, to validate the algorithm, a simulated data set where all constituent effects are known was created. The set was simulated to form two groups (clusters) of data, each with 25 spectra. These spectra were created by adding together a number of Gaussian curves with various peak positions, heights and widths. The spectra were created so that they visually appeared to be similar to the fingerprint region of typical biomedical IR spectra; however, no two spectra were identical. A spectrum of a thin layer of Matrigel (an artificial basement membrane consisting mainly of protein[32]) was used as a 'template' to acquire peak parameters. A random number generator was used to vary the positions (±1 cm^{-1}), heights (±20%) and widths (±2.5%) of peaks within each spectrum. The second data set was subject to the same random variation as the first but was intentionally given a higher absorbance by 0.1 at the 1300 cm^{-1} and 1740 cm^{-1} peaks so that the two groups of data would appear different when analysed using PCA. These simulated data and the corresponding score plot for the first two PCA components are shown in Fig. 2. As expected, two distinct clusters of spectra can be observed. This two-component model accounted for 81.9% of the variance in these 'ideal' data.

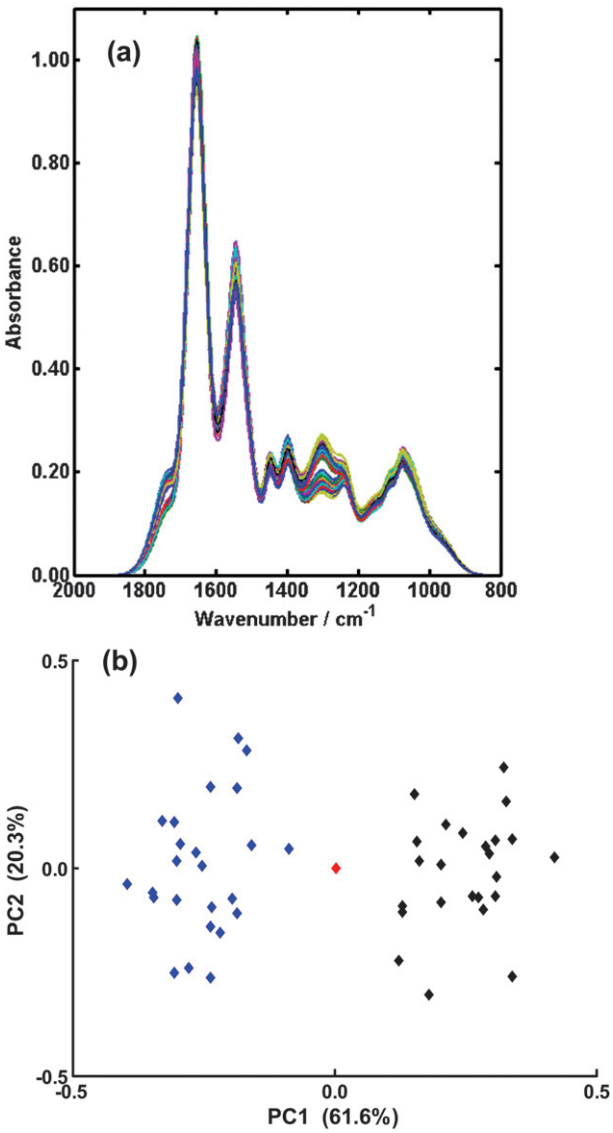


Fig. 2 (a) All 50 simulated 'pure absorbance' spectra shown on one plot. (b) The corresponding PCA scores plot of the data.

To create a data set affected by RMieS from multiple scattering particles, each spectrum from the pure absorbance spectra data set was taken and 10 unique scattering curves, Q, were added to each one. The refractive index used was the correct one corresponding to each spectrum, and a random number generator was used to create different scattering particle diameters ranging from 4 to 10 μm and an average refractive index varying between 1.3 and 1.4.

The resultant spectra are shown in Fig. 3a and closely resemble the type of data typically observed for single cells. Fig. 3b shows that, despite the fact that the original pure spectra consist of two distinct groups, the spectra no longer separate using PCA even when the 2nd derivative of the data is used. This is because the spectral distortions induced by scattering are generally significantly larger than real spectral differences associated with subtle differences in biochemistry. This data set of scattered spectra and the corresponding pure absorbance are subsequently used to test the algorithm.

 This journal is © The Royal Society of Chemistry 2010

View Article Online

Published on 15 December 2009. Downloaded by UNIWERSYTET JAGIELLONSKI on 11/4/2022 1:53:39 PM.

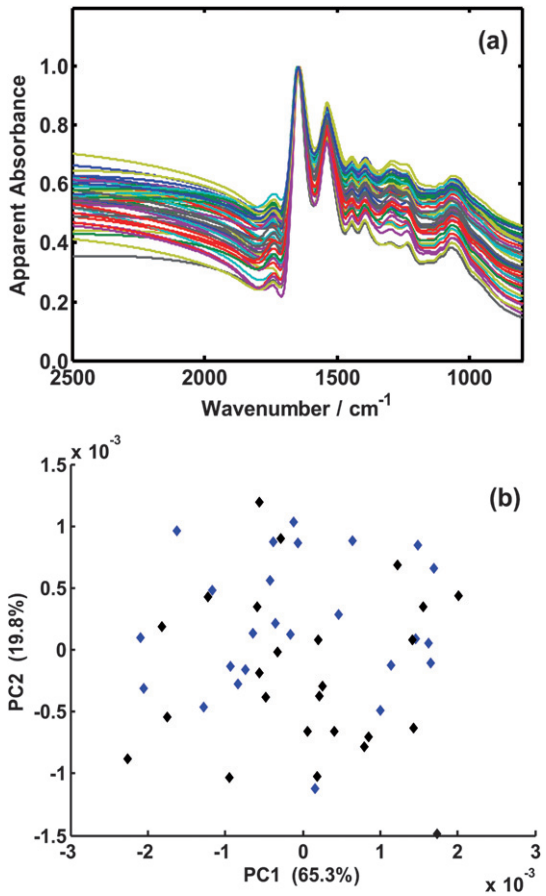


Fig. 3 (a) The 50 simulated 'pure absorbance' spectra from Fig. 2a, after the superposition of 10 unique artificial Mie scattering curves. (b) PCA scores plot for the total data set of the 2nd derivative of raw spectra and EMSC normalisation.

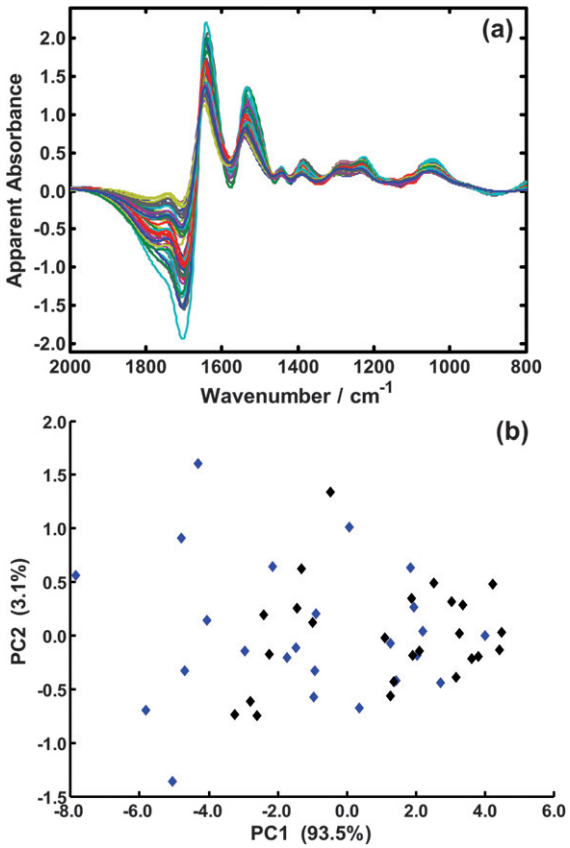


Fig. 4 (a) The artificial spectra corrected using original Mie scattering-EMSC algorithm. (b) The corresponding PCA scores plot of the corrected spectra.

5.2 Correction results

5.2.1 Correction results using existing Mie scattering-EMSC. Fig. 4a shows the resulting spectra after the Mie scattering-EMSC correction algorithm is applied. As can be seen the baseline oscillations have been removed but the dispersion artefact is still very prominent in many of the spectra. The two data sets still cannot be distinguished using PCA, Fig. 4b, because the variation associated with the dispersion artefact that still remains is still larger than the known true spectral differences.

5.2.2 Correction using RMieS model

5.2.2.1 Ideal correction. The reference spectrum used to correct each scattered spectrum was the pure absorbance spectrum of each sample, giving the algorithm the most ideal conditions. The resulting corrected spectra look almost identical to those in Fig. 2a, as expected, and hence they are not shown. The PCA scores plot for these corrected spectra is also practically identical to that of the pure absorbance spectra in Fig. 2b. By subtracting the corrected spectra from the pure absorbance spectra it was observed that the differences were four orders of magnitude smaller than the pure absorbance spectra, hence negligibly different.

This result is non-trivial as it demonstrates that a scattered spectrum can be corrected using our method if the reference spectrum used is a perfect match, even though 10 scattering curves of unknown scattering particle diameter were added. The mathematics of the algorithm have been verified and the concept of using PCA to describe the majority of variance within 1000 Q curves into 7 loading spectra has proved a success.

5.2.2.2 Using an imperfect reference spectrum. Although the previous section shows that correcting a spectrum with an ideal reference spectrum will give the correct result, it is clear that the perfect reference spectrum will almost never be available in practice. Thus a compromise has to be made. One option is to use the mean spectrum of the whole data set under investigation. The mean spectrum was used for the correction of each spectrum meaning that none of the spectra has the perfect conditions for correction.

The resultant PCA scores plot shown in Fig. 5a is independent of the pure absorbance scores plot in Fig. 2b, making a direct comparison difficult; however, projecting the corrected spectra onto the loadings from the pure absorbance spectra produces a directly comparable scores plot.

Such a plot is shown in Fig. 5b where the scores of the corrected spectra are shown in the same 'sub-space' as those of the pure absorbance spectra, achieved by projecting the corrected spectra on the loadings from the pure absorbance spectra. This

View Article Online

Published on 15 December 2009. Downloaded by UNIWERSYTET JAGIELLONSKI on 11/4/2022 1:53:39 PM.

essentially means that they are being viewed from the same point of view, mathematically speaking, making a direct comparison of positions possible. Fig. 5c shows the scores plot on the same scale with the arrows indicating the movement of each point from its ideally corrected position. The main observation here is that the corrected spectra separate into two separate groups, so from a pragmatic classification point of view, the pre-processing has been successful. However, since we also want to interpret the spectral details of the pre-processed spectra, it is a problem that the pre-processed spectra are not in their expected positions. Firstly, they have moved somewhat towards the reference spectrum, which is the mean spectrum located at the origin.

Secondly, this result shows that using a non-ideal reference spectrum to some extent does affect the quality of the corrected spectra; they exhibit greater similarity to the reference spectrum. Although the spectra have not been corrected perfectly the results do still show that there are two clear groups of data which may be sufficient for certain applications.

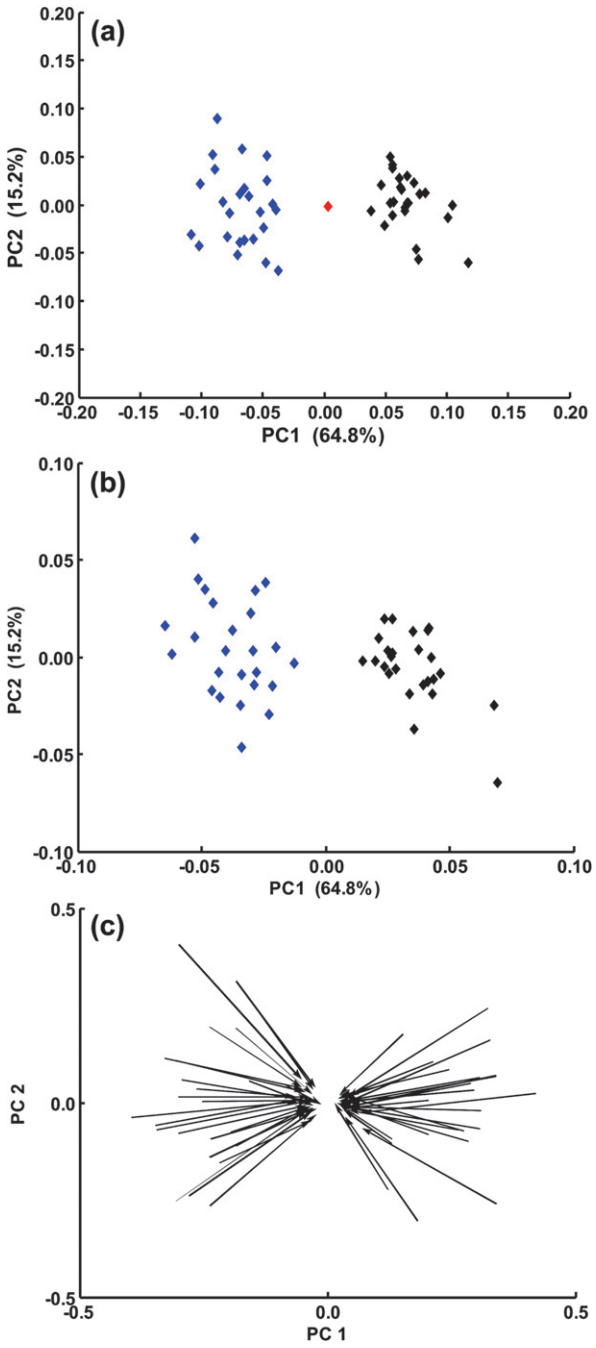


Fig. 5 (a) PCA scores plot of the corrected spectra. (b) Scores plot of the non-ideal corrected data projected onto the loadings from the pure absorbance spectra PCA. (c) A plot showing the shift of the non-ideal reference corrected spectra from their correct pure absorbance positions.

5.2.2.3 Iterative correction method. In order to improve the spectral correction process even further, the model spectra going into the EMSC model may be optimized for the given purpose. This is achieved here by iteratively improving the reference spectrum, by letting the original, non-ideal reference spectrum (the mean of the input spectra) be replaced by the corrected spectrum from the previous iteration for each corresponding spectrum. The algorithm is run once more correcting the raw spectrum again using the new reference. This iterative approach is depicted schematically in Fig. 6.

Fig. 7 shows the effect of increasing iterations on the accuracy of the correction, calculated by projecting the corrected spectra onto the loadings from the pure absorbance spectra. Using this iterative approach, the corrected spectra move towards their pure absorbance spectra in score space with increasing number of iterations, indicating an improvement in the quality of the correction. For this particular data set, convergence of the algorithm is reached after 8 iterations before the corrected spectra have moved to their pure absorbance positions; however, there is a significant improvement upon the first iteration. This is illustrated further in Fig. 8 which shows the sum of Pythagorean distances of the PCA scores plot of the corrected spectra, projected onto the original pure absorbance spectra sub-space, as a function of iteration. For comparison, the data point for the previous Mie Scattering-EMSC is also shown. As can be seen, using the RMieS-EMSC algorithm produced a significant improvement which continues with each iteration, in this case up to 8 iterations. The number of iterations required will of course

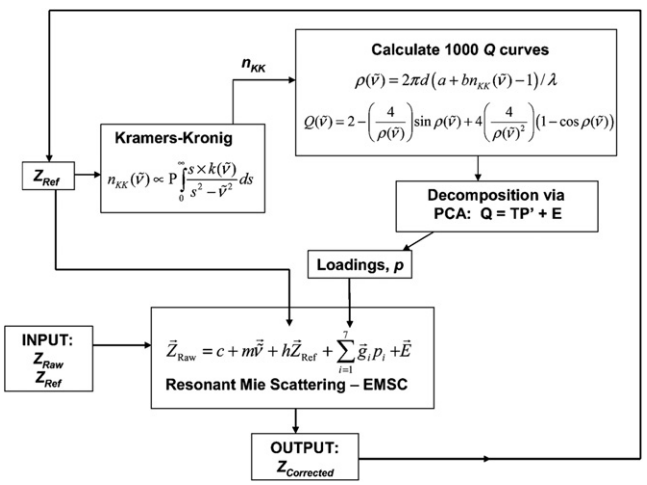


Fig. 6 Flow chart illustrating the iterative procedure implemented to use the corrected spectrum as the new reference spectrum and running the algorithm once more.

 This journal is © The Royal Society of Chemistry 2010

View Article Online

Published on 15 December 2009. Downloaded by UNIWERSYTET JAGIELLONSKI on 11/4/2022 1:53:39 PM.

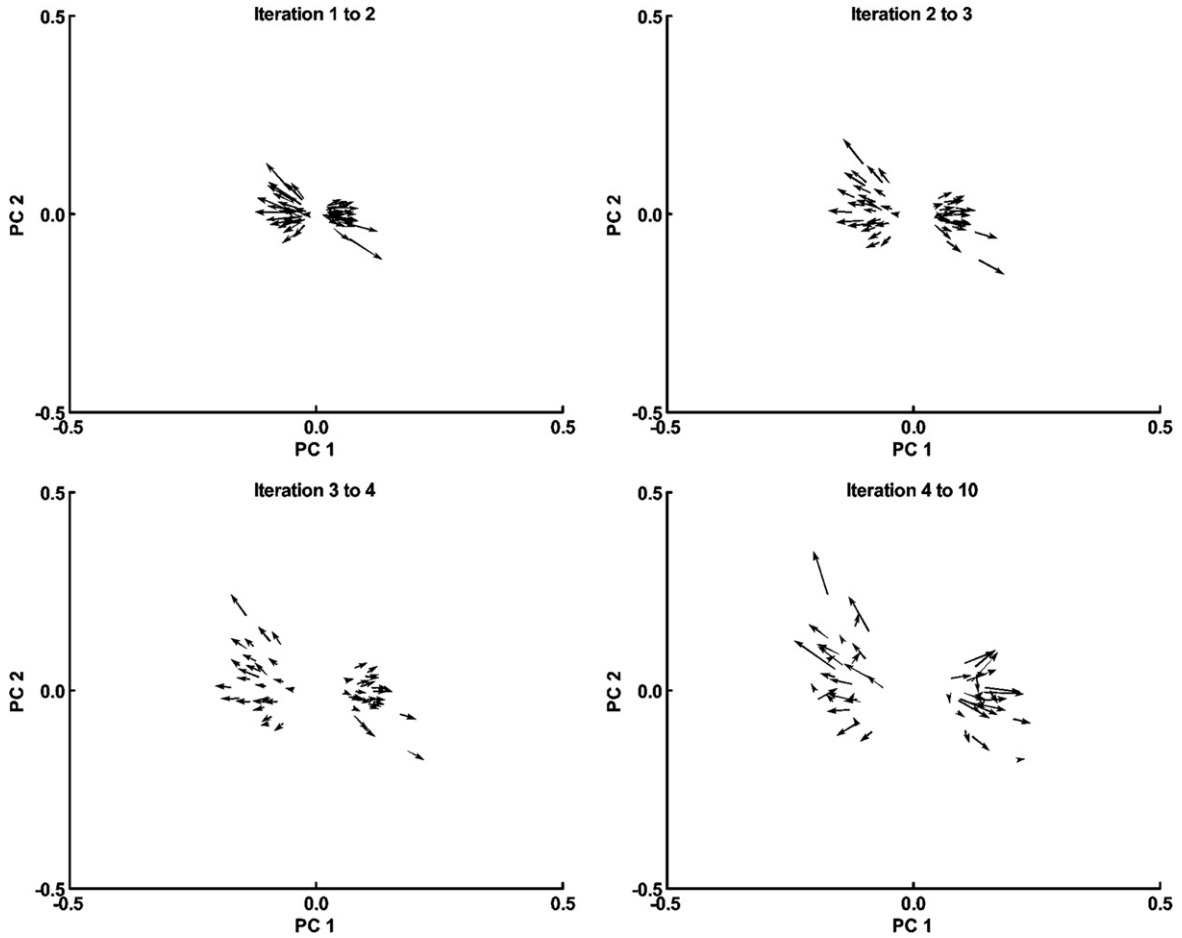


Fig. 7 Scores plot showing the scores shift of the iteratively corrected spectra from iteration 1 to 2, 2 to 3, 3 to 4 and 4 to 10. Arrows show that each spectrum is moving towards its true absorbance spectrum position. All spectra were projected onto the loadings from the pure absorbance spectra.

depend on the data set but further work, to be published elsewhere, suggests that 10 iterations should be sufficient in most cases.[33]

Although the sum of the Pythagorean distances give a measure of how well the correction algorithm works, it is useful to consider other qualitative measures. Fig. 9a shows the Amide I band for the original uncorrected simulated data. The original position of the peak was set to 1655 ± 1 cm^{-1} indicated by the leftmost vertical line. The actual peak positions of the simulated scattering data range from 1635.9 to 1647.0 cm^{-1} with a mean of 1642.3 cm^{-1}. Thus it is clear that the significant shift in peak wavenumber is induced by the RMieS. The hitherto existing Mie scattering-EMSC correction significantly improves the overall data and brings down the sum of the Pythagorean distances (from the pure absorbance spectra) from a value of 95 to 22 but has little impact on the Amide I peak position Fig. 9b. The peak position of the EMSC corrected spectra range from 1636.8 to 1647.1 cm^{-1} with a mean of 1642.7 cm^{-1}. It is only when the RMieS-EMSC correction is performed that a band position close to the correct wavenumber value is obtained. Fig. 9c shows that the Amide I bands are now closely aligned, ranging from 1653.0 to 1656.4 cm^{-1} with a mean of 1654.5 cm^{-1}. This is a potent illustration of the fact that although the Mie Scattering-EMSC may successfully enable data to be separated into groups (which is often the aim of the experiment), any biological interpretation of the data, particularly with respect to the Amide I band and associated protein structure, cannot be made unless the RMieS-EMSC correction is applied.

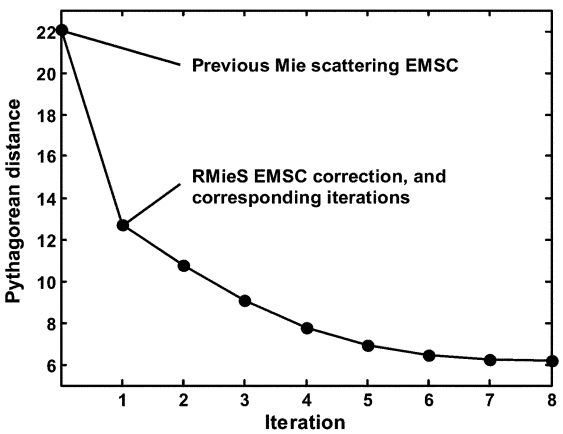


Fig. 8 Plot of sum of the Pythagorean distance of PCA scores away from the score positions for the pure absorbance spectra (measured on a common sub-space) *vs.* iteration. The first point on the plot is for the Mie Scattering-EMSC algorithm.

View Article Online

Published on 15 December 2009. Downloaded by UNIWERSYTET JAGIELLONSKI on 11/4/2022 1:53:39 PM.

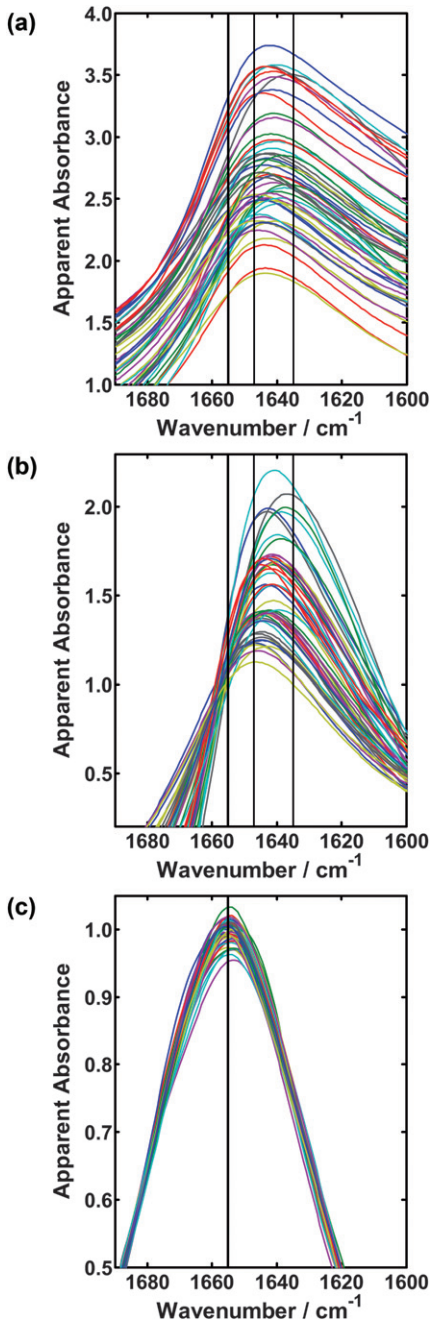


Fig. 9 The Amide I band shown for (a) the uncorrected, (b) Mie Scattering-EMSC corrected and (c) RMieS-EMSC corrected spectra.

6. Conclusion

In this work we have presented a model for the dynamics of Mie scattering in single-cell spectra, and a correction method based on those dynamics. We have shown that using the true absorbance spectrum as the reference spectrum for each correction, every scattered spectrum can be corrected essentially perfectly. This result is non-trivial as it illustrates the concept of compressing 1000 scattering Mie curves into a small number of principal component loading spectra and using these in a least squares fitting algorithm to estimate the scattering contributions.

The second and more interesting test was the correction of the spectra using a non-ideal reference spectrum which was non-ideally suited to any spectrum as would be the case in real life as the true spectrum is unknown. This method yielded corrected spectra that still separated the test data set into two groups as they should when analysed with PCA.

Using an iterative correction process whereby the corrected spectrum becomes the new reference spectrum it has been shown that the new corrected spectrum resembles its true pure absorbance spectrum even further. The limitation of this method is that convergence is reached before each spectrum is corrected perfectly; however, each corrected spectrum is a significantly better representation of its true spectrum compared with that before the correction.

Most importantly, we have shown that the true position of the biologically significant Amide I band can only be obtained with the RMieS-EMSC algorithm. It follows therefore that interpretation of previously uncorrected single-cell infrared spectra, in terms of protein secondary structure, must be viewed with extreme caution.

Further work is underway to improve on the incorporation of the theory of the scattering process into the algorithm, and new ways to provoke a stronger convergence during the iterative correction procedure. This will be the subject of a paper in the near future.

Acknowledgements

We acknowledge the EPSRC-RSC Analytical Science Studentship scheme for support for P. B., and the EU-Special support action (DASIM project) for facilitation of meetings. We also acknowledge the EU for funding travel to SOLEIL and we thank all the staff at SOLEIL particularly those associated with the SMIS beamline. A. K. and H. M. acknowledge support from the Norwegian Agricultural Food Research Foundation.

References

1 L. J. Fowler and W. A. Lachar, *Arch. Pathol. Lab. Med.*, 2008, **132**(3), 373–383.
2 N. Jamin, P. Dumas, J. Moncuit, W. H. Fridman, J. L. Teillaud, L. G. Carr and G. P. Williams, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 4837–4840.
3 P. Lasch, A. Pacifico and M. Diem, *Biopolymers*, 2002, **67**, 335–338.
4 P. Lasch, M. Boese, A. Pacifico and M. Diem, *Vib. Spectrosc.*, 2002, **28**, 147–157.
5 P. Dumas and L. Miller, *Vib. Spectrosc.*, 2003, **32**, 3–21.
6 P. Dumas, N. Jamin, J. L. Teillaud, L. M. Miller and B. Beccard, *Faraday Discuss.*, 2004, **126**, 289–302.
7 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*, 2005, **77**, 18–30.
8 E. Gazi, J. Dwyer, N. P. Lockyer, J. Miyan, P. Gardner, C. A. Hart, M. D. Brown and N. W. Clarke, *Vib. Spectrosc.*, 2005, **38**, 193–201.
9 E. Gazi, P. Gardner, N. P. Lockyer, C. A. Hart, N. W. Clarke and M. D. Brown, *J. Lipid Res.*, 2007, **48**, 1846–1856.
10 D. A. Moss, M. Keese and R. Pepperkok, *Vib. Spectrosc.*, 2005, **38**, 185–191.
11 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.*, 2006, **90**, 3783–3795.

 This journal is © The Royal Society of Chemistry 2010

View Article Online

12 J. Sulé-Suso, D. Skingsley, G. D. Sockalingum, A. Kohler, G. Kegelaer, M. Manfait and A. J. El Haj, *Vib. Spectrosc.*, 2005, **38**, 179–184.
13 M. Romeo, B. Mohlenhoff and M. Diem, *Vib. Spectrosc.*, 2006, **42**, 9–14.
14 B. R. Wood, L. Chiriboga, H. Yee, M. A. Quinn, D. McNaughton and M. Diem, *Gynecol. Oncol.*, 2004, **93**, 59.
15 M. Romeo, C. Matthaus, M. Miljkovic and M. Diem, *Biopolymers*, 2004, **74**, 168.
16 B. Bird, M. J. Romeo, M. Diem, K. Bedrossian, N. Laver and S. Naber, *Vib. Spectrosc.*, 2008, **48**, 101–106.
17 B. Mohlenhoff, M. Romeo, M. Diem and B. R. Wood, *Biophys. J.*, 2005, **88**, 3635–3640.
18 S. Boydston-White, T. Gopen, T. Houser, J. Bargonetti and M. Diem, *Biospectroscopy*, 1999, **5**, 219–227.
19 M. Romeo and M. Diem, *Vib. Spectrosc.*, 2005, **38**, 129–132.
20 A. Kohler, J. Sulé-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.*, 2008, **62**, 259–266.
21 J. Lee, E. Gazi, J. Dwyer, M. D. Brown, N. W. Clarke and P. Gardner, *Analyst*, 2007, **132**, 750–755.
22 P. Bassan, H. J. Byrne, F. Bonnier, J. Lee, P. Dumas and P. Gardner, *Analyst*, 2009, **134**, 1586–1593.
23 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*, 2009, **134**, 1171–1175.
24 J. K. Pijanka, A. Kohler, Y. Yang, P. Dumas, S. Chio-Srichan, M. Manfait, G. D. Sockalingum and J. Sulé-Suso, *Analyst*, 2009, **134**, 1176–1181.
25 H. Martens, J. P. Nielsen and S. B. Engelsen, *Anal. Chem.*, 2003, **75**, 394–404.
26 P. Dumas, F. Polack, B. Lagarde, O. Chubar, J. L. Giorgetta and S. Lefrancois, *Infrared Phys. Technol.*, 2006, **49**, 152–160.
27 B. Stuart, *Biological Applications of Infrared Spectroscopy*, John Wiley and Sons Ltd, Chichester, UK, 1997.
28 H. C. van de Hulst, *Light scattering by small particles*, Dover Publications, Mineola, NY, 1981.
29 G. Mie, Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen, *Ann. Phys.*, 1908, **330**, 377–445.
30 R. de L. Kronig, *J. Opt. Soc. Am.*, 1926, **12**, 547–557.
31 H. A. Kramer, *Atti Congr. Intern. Fisica Como*, 1927, **2**, 545–557.
32 H. K. Kleinman, M. L. McGarvey, J. R. Hassell, V. L. Star, F. B. Cannon, G. W. Laurie and G. R. Martin, *Biochemistry*, 1986, **25**, 312–318.
33 P. Bassan, A. Kohler, H. Martens, J. Lee and P. Gardner, to be published.

Published on 15 December 2009. Downloaded by UNIWERSYTET JAGIELLONSKI on 11/4/2022 1:53:39 PM.