



Fluorescence EEM Spectroscopy for Rapid Identification and Quality Evaluation of Cell Culture Media Components.

Title	Fluorescence EEM Spectroscopy for Rapid Identification and Quality Evaluation of Cell Culture Media Components.
Author(s)	Li, Boyan;Ryan, Paul W.;Shanahan, Michael;Ryder, Alan G.
Publication Date	2011-11
Publisher	Ingenta
Repository DOI	http://dx.doi.org/10.1366/11-06383

FLUORESCENCE EEM SPECTROSCOPY FOR RAPID IDENTIFICATION AND QUALITY EVALUATION OF CELL CULTURE MEDIA COMPONENTS.

Boyan Li,¹ Paul W. Ryan,¹ Michael Shanahan,¹ Kirk J. Leister,² and Alan G. Ryder.^{1*}

¹ Nanoscale Biophotonics Laboratory, School of Chemistry, National University of Ireland, Galway, Galway, Ireland.

² Process Analytical Sciences. Bristol-Myers Squibb Company, East Syracuse, New York, USA.

* To whom all correspondence should be addressed.

Tel: +353 91 49 2943 **Fax:** +353 91 49 4596 **Email:** alan.ryder@nuigalway.ie (A.G.R)

Running Title: Fluorescence spectroscopic analysis of cell culture media

Published Citation: Fluorescence EEM Spectroscopy for Rapid Identification and Quality Evaluation of Cell Culture Media Components. B. Li, P.W. Ryan, M. Shanahan, K.J. Leister, and A.G. Ryder. *Applied Spectroscopy*, 65(11), 1240-1249, (2011).

DOI: <http://dx.doi.org/10.1366/11-06383>

Note: This is the final accepted version that includes corrections but may not include all the proofing corrections. The final definitive version of the manuscript is available on the Applied Spectroscopy website.

ABSTRACT:

The application of fluorescence excitation-emission matrix (EEM) spectroscopy to the quantitative analysis of complex, aqueous solutions of cell culture media components was investigated. These components, yeastolate, phytone, recombinant human insulin, eRDF basal medium, and four different chemically defined (CD) media, are used for the formulation of basal and feed media employed in the production of recombinant proteins using a Chinese Hamster Ovary (CHO) cell based process. The comprehensive analysis (either identification or quality assessment) of these materials using chromatographic methods is time consuming and expensive, and not suitable for high throughput quality control. The use of EEM in conjunction with multiway chemometric methods provides for a rapid, non-destructive analytical method suitable for the screening of large numbers of samples. Here we use multiway robust principal component analysis (MROBPCA) in conjunction with N-way partial least-squares-discriminant analysis (NPLS-DA) to develop a robust routine for both the identification and quality evaluation of these important cell culture materials: These methods are applicable to a wide range of complex mixtures since they do not rely on any predetermined compositional or property information, thus making them potentially very useful for sample handling, tracking, and quality assessment in biopharmaceutical industries.

INDEX HEADINGS:

Fluorescence, Chemometrics, ROBPCA, Cell culture, chemically defined, media, quality.

INTRODUCTION:

In industrial therapeutic protein production using mammalian cell culture, the cell culture media are vital for generating the correct physiological environment to allow effective large scale, reproducible culturing. For effective cell growth, the formulation of appropriate media requires the careful selection and blending of a wide variety of components which include amino acids, carbohydrates, cofactors, and other materials. These aqueous based media must provide all of the cell nutrients required for growth including an energy source, while also maintaining pH and osmolarity.¹ The formulation and use of various media are strictly controlled to optimize growth rate, product yield, and quality, while preventing the formation of unwanted metabolites. Media are, in general, complex aqueous solutions containing inorganic salts, sugars, amino acids, vitamins, metals, and other cell culture nutrients. Media composition varies according to cell type, product type, and are, often propriety formulations for individual manufacturers.^{2,3,4}

Media may also be formulated using components which themselves are complex, chemically defined (CD) media mixtures. For example, a 1:1 mixture of RPMI 1640 and DMEM/F12 formulations is termed RDF medium⁵ and an enriched basal RDF (eRDF) can be formulated by increasing the levels of amino acids and glucose two-fold over the original RDF formulation to sustain high density growth. In mammalian cell culture, eRDF is often used as a

basal medium to optimize cell culture growth and achieve rapid cell proliferation.⁶ CD media are also available for Chinese hamster ovary (CHO) cells used in the production of recombinant proteins in suspension culture.⁷ Other media constituents include complex natural products such as yeast (yeastolate) and soya (phytone) extracts which are used as nutritional supplements. Yeastolate is a filtered, aqueous extract of baker's or brewer's yeast from autolysis while phytone is an ultra-filtered enzymatic digest of soybean meal. Both are largely undefined mixtures of amino acids, peptides, vitamins, polysaccharides and minerals. Another constituent of many media is recombinant human insulin (H.R. Insulin) which stimulates cell proliferation, aiding carbohydrate metabolism. The efficacy of these complex materials are very dependent on the specific production process and measuring lot-to-lot variability (particularly for subtle changes) can be an issue because of the difficulty in characterization of these complex mixtures.

Analysis of these complex media is challenging because comprehensive analysis generally requires a combination of chromatographic separations and mass spectrometry. This can make routine analysis expensive and time-consuming, particularly for simple requirements such as materials identification, determination of lot-to-lot consistency, and the monitoring of batch-to-batch preparations of media. Thus there is a need for rapid, holistic methods that can be used for the early stage identification and material quality assessment. Spectroscopic methods offer an ideal solution because of the advantages ranging from speed and sensitivity, to facile automation, and inexpensive unit test costs. Fluorescence in particular, offers high sensitivity, low signal-to-noise ratios, and relatively large linear ranges for quantitative analysis.⁸ For dilute aqueous solutions, the combination of multidimensional fluorescence spectroscopy with multiway chemometric methods is one option. For example, an excitation-emission matrix (EEM) spectrum which is a tri-parametric assay,⁹ providing excitation wavelength, emission wavelength, and fluorescence intensity can be mathematically analyzed to provide information about individual components in mixtures.^{10,11,12} EEM measurements typically takes a couple of minutes, which allows an almost continuous measurement (on-line, *in situ* and *in vivo*) of samples. For complex, turbid mixtures containing biogenic fluorophores (amino acids, proteins, vitamins, etc.) EEM spectroscopy can enable the simultaneous detection, qualitative and quantitative analysis of multiple analytes. For complex cell culture media where there are multiple fluorophores and other photophysically active molecules present in relatively high concentrations, Energy Transfer (radiative and non-radiative) and quenching (both static and dynamic) play a large part in determining the shape and intensity of EEM plots, thus providing a unique fingerprint suitable for qualitative characterization. There is an extensive recent published literature on the utilization of multidimensional fluorescence measurements for complex mixture analysis.^{13, 14, 15, 16, 17, 18} The use of multiway chemometric methods for the qualitative and quantitative analysis of EEM data is well established in water analysis,^{19, 20, 21} food science,²² and for quantitative determination of analytes ($\mu\text{g/mL}$) in various fluids such as urine.²³

Previous work has demonstrated the efficacy of Raman spectroscopy for the identification / quality assessment of cell culture media components,^{24, 25} and that EEM data when combined

with chemometric methods can be used to predict cell culture media performance.²⁶ This study complements the previous works because some cell culture media components such as yeastolate and phytone, are too fluorescent for Raman analysis, produce highly detailed EEM. We suggest that by using a combination of these fluorescence and Raman-based methods, a comprehensive and integrated analysis system can be easily implemented for the majority (and many of the most significant analytes) of the components used for the manufacture of complex cell culture media. This EEM-Chemometric method is therefore an ideal Process Analytical Technology candidate for the assessment of critical quality and performance attributes of the complex in-process materials used to prepare cell culture media.^{27, 28}

MATERIALS AND METHODS

Media samples: Materials used in the preparation of cell culture media for an industrial biopharmaceutical process (CHO cell fermentation) were supplied by Bristol-Myers Squibb (BMS). Most of the materials contained a wide variety of components and were usually supplied as aqueous solutions. Dissolved solids (in many cases) were estimated to be in the 0.5–2.0% range. The precise compositional information of these materials is proprietary to BMS and their third-party suppliers. These media components are blended together (in varying proportions) to produce the feed and basal media for the manufacturing process. All samples were stored (including during transport from US to Ireland) at low temperatures (~4 °C for CD media and -70 °C for the others) as they are potentially labile and composition can change with time. The chemically defined media were (number of samples in brackets): CD-A1, mostly amino acids (31), CD-A2, mostly inorganic salts (31), CD-S1, inorganic and organic salts (28), CD-S2, mostly organic materials (30), and eRDF, a diverse mix of organic & inorganic materials (21). The other materials were yeastolate (a hydrolyzed yeast extract, 21 samples), phytone (a hydrolyzed soya extract, 3 samples), and H. R. Insulin (19). In addition we were supplied with a test set of 17 samples for validation purposes. The eRDF, yeastolate, and phytone were supplied as solid materials and were made up in aqueous solutions under aseptic conditions at concentrations of 0.84, 5.0, and 5.0 g/L with Millipore water. The HR Insulin concentration was 2 mg/mL.

Because all the cell culture media component samples are potentially labile and the composition may change with time, they have to be stored at low temperatures. For these complex aqueous-based solutions, the solubility of their individual components is temperature dependant and their fluorescence measurements are sensitive to temperature fluctuations. This can be observed as precipitates and/or solid suspensions in the samples on removal from cold storage (particularly evident in the eRDF samples). The presence of these particles causes significant scatter artifacts in the EEM spectra (and also Raman spectra).²⁴ These changes can be mistaken for real compositional changes in the sample and therefore it is very important to

thermally equilibrate cell culture media samples at room temperature for several minutes after removal from cold storage to ensure complete dissolution prior to EEM measurement.

Instrumentation and data collection: Fluorescence measurements were at 25°C made using a Cary Eclipse (Varian, now Agilent) fluorescence spectrometer fitted with a thermostatted 4-position multi-cell holder. EEMs were obtained by combining a series of emission scans made from 270 to 600 nm while exciting at wavelengths ranging between 230 and 520 nm with an interval of 5 nm. The excitation and emission bandwidths were 5 nm. Semi micro quartz cuvettes with a long axis of 10 mm (excitation axis) and a short axis of 4 mm (emission axis) were employed for all of the fluorescence measurements.²⁶ A background spectrum of deionized water was measured at the beginning and end of each sample collection run.

Data pre-treatment: EEM data was pre-processed using established methods,^{29, 30, 31} some of which have been described previously.²⁶ Since Rayleigh and Raman scatter are generally unrelated to the compositional properties of the normal samples and the scatter peaks do not behave linearly (or trilinearly), they may complicate and bias fluorescence data modeling unless dealt with appropriately.²⁹ For example the Rayleigh scattering bands (including the second order peaks) were replaced with a curve fit, connecting points either side of the peak using imputation.^{30, 31} It was not possible to remove the Raman bands as they were convoluted with the much stronger fluorescence signal. For the majority of the analyses undertaken, PLS_Toolbox 4.0 (Eigenvector Research Inc., Wenatchee, WA, USA), supplemented by in-house-written codes was used as a source of standard procedures. All calculations were performed using the MATLAB, (Mathworks Inc., Cambridge, MA, USA) platform version 7.4 on a standard desktop computer.

Chemometric Methods: For the robust qualitative (identification) and quantitative (quality assessment) analysis of these cell culture media and raw materials a combination of chemometric methods were employed. N-way partial least-squares-discriminant analysis (NPLS-DA) was used initially to discriminate the samples, after which the relative quality of these discriminated samples were determined using dedicated multiway robust principal component analysis (MROBPCA). PLS-DA is an extension of the PLS methodology for classification, where one defines a set of dummy variables to characterize different groups within the spectral variables and then seek a set of latent variables to maximize between-group variability relative to a pooled measure of within-group variability for discriminant analysis.^{32, 33, 34, 35, 36, 37} N-way PLS-DA was used to analyze multidimensional EEM data to rapidly identify different types of media and raw materials. The second step in the integrated chemometric analysis method was to determine the quality of the various samples using chemometric methods based on PCA. Robust PCA (ROBPCA) is a robust statistical method which is more resistant to outliers than classical PCA.^{38, 39, 40, 41} MROBPCA is an extension of ROBPCA for handling multiway data. In essence the method unfolds the multidimensional EEM data into a two-dimensional matrix before

undertaking a standard ROBPCA. Classical PCA tries to explain the maximum variance in the data using a few significant principal components (PCs), which are often biased towards outlying points (samples or data). ROBPCA on the other hand, tries to determine a small number of PCs which are not sensitive to anomalous observations in the data by obtaining a robust estimate of the centre of the data, a set of robust loadings, eigenvalues and robust scores. The robust PCs are calculated using projection-pursuit techniques,^{42, 43, 44} and the minimum covariance determinant method.^{45, 46, 47} An outlier diagnostic map is the ROBPCA product which highlights and distinguishes outliers in the data. Once the robust PCs were determined, two significant statistics, Hotelling's T^2 and Q residual, were calculated. T^2 and Q values of each sample were then compared with predefined thresholds, and thus the magnitude of the measured/threshold ratio for T^2 and Q gives a very useful and quick quantitative indicator of sample quality. For ratios of ≤ 1.0 , this indicates good quality samples, and conversely as the ratio becomes larger the more the sample deviates from the normal group of samples. Assessing whether a sample is a true outlier due to compositional change is the key objective of these studies, thus setting the magnitude of the predefined thresholds is a critical parameter. In the context of these complex cell culture media, these thresholds would be defined via an analysis of the end use performance (typically via cell culture studies) of the normal or good media.

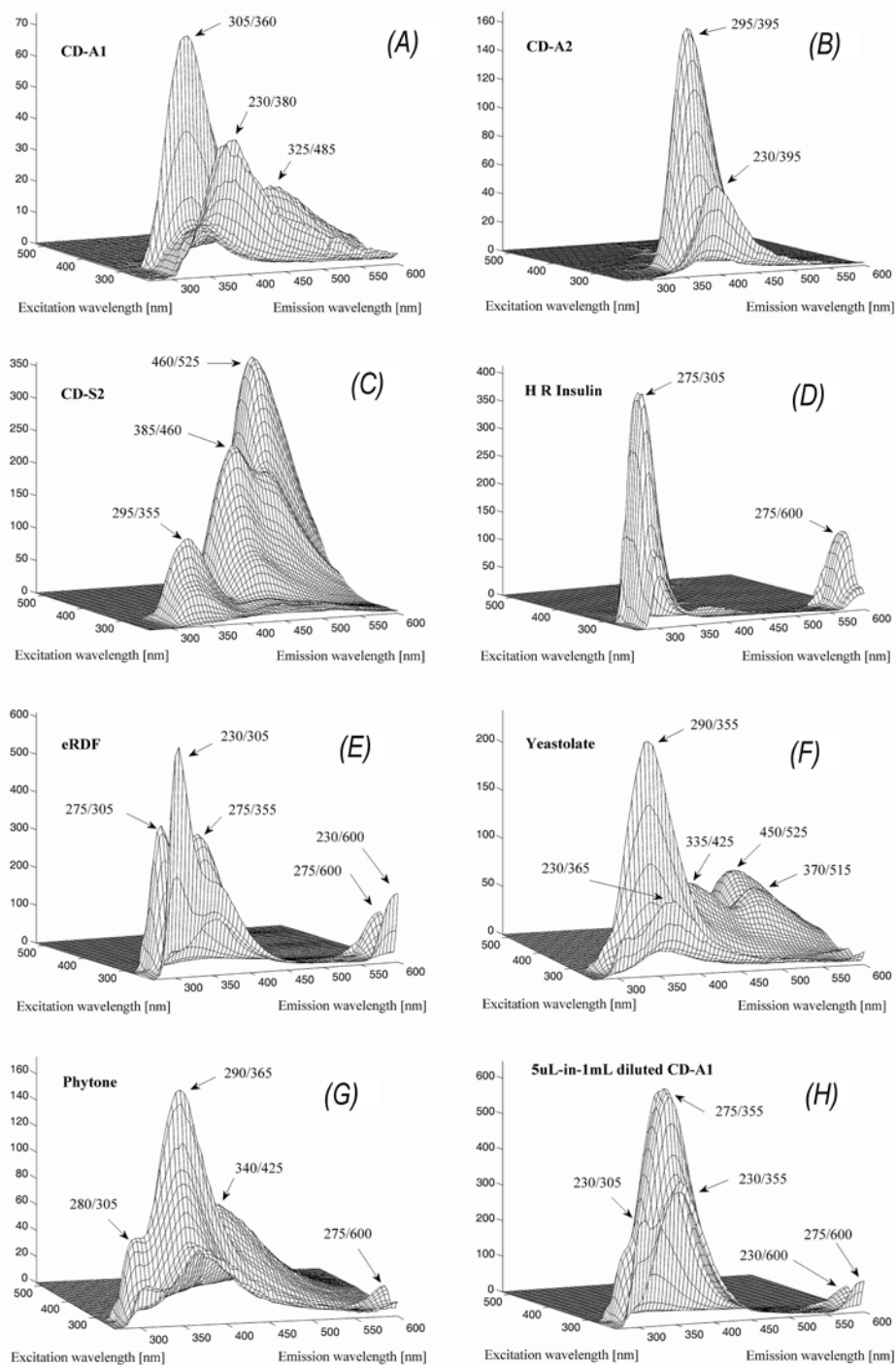


Figure 1: EEM landscapes (after Rayleigh scattering correction) of the various media samples studied. The local maxima are marked by arrows and the excitation/emission values.

RESULTS AND DISCUSSION:

All samples except for CD-S1 (which contain primarily inorganic and organic salts) are reasonably fluorescent (Fig. 1) and produce elaborate EEM spectra indicating the presence of multiple fluorophores. There is significant overlap of the different emission bands from individual constituent fluorophores for all samples apart from insulin. The relatively simple shape of the insulin EEM is due to the fact that it is a low concentration ($\sim 3 \times 10^{-4}$ M), single component sample. Furthermore, the insulin molecule contains only a single phenylalanine (very weakly fluorescent) and four tyrosine residues, therefore the emission is mainly from tyrosine ($\lambda_{\text{ex}}/\lambda_{\text{em}} = 275/305$ nm). The other relatively simple EEM are the CD-A2 samples, where the primary fluorophore is pyridoxine hydrochloride (vitamin B₆, $\lambda_{\text{ex}}/\lambda_{\text{em}} = \sim 295/395$ nm).⁴⁸ For CD-A1, CD-S2, and eRDF, accurate band assignment is less certain, because of their complex composition (7 to 40+ components), and low concentration (0.5–2.0% dissolved solids). We are reasonably confident (based on *a priori* knowledge of the media composition) that the significant fluorescence of CD-A1 mostly originates from tryptophan, tyrosine, and phenylalanine, which are present in appreciable quantities: ~ 190 , ~ 200 , and ~ 200 mg/L, respectively. In contrast, CD-S2 and eRDF fluorescence derives from a wider variety of fluorophores which include tryptophan, tyrosine, riboflavin, folic acid, and vitamin B₁₂. For yeastolate and phytone, where there is no accurate compositional data available, we expect that emission originates largely from tryptophan and tyrosine.

This type of sampling geometry employed for these measurements does not prevent or minimize inner-filter (IF) effects, but it provides a very reproducible optical system for the collection of EEM data from widely differing complex samples (see supplemental information). This is important, because we are only concerned with the identification and analysis of variance from specified formulations at their *working concentrations* for operational reasons such as minimizing sample handling. These working concentrations are very concentrated from a traditional fluorescence spectroscopy standpoint, and thus IF effects are intrinsic to the data being collected. Any observed changes in the EEM spectra caused by IF variations, therefore, must arise from some form of compositional change. Even small changes caused by media preparation in a slightly more dilute or concentrated form, may be significant to know/observe from a manufacturing quality standpoint. In practice, all these media and the raw materials are supposed to have very narrowly defined compositions, and as such measureable deviations from the norm are potentially very important in process management. Therefore, for these biopharmaceutical materials there is no need to implement any IF correction methods to the EEM data.⁴⁹ Our preference would be to utilize an external standard for both instrument bias correction and calibration transfer in this situation because these issues of instrument bias coupled and IF are very significant.⁵⁰

The interplay of IF, energy transfer (ET) and quenching between all of the various media components thus produces a unique EEM profile which can be used to characterize these complex mixtures. However, since IF, ET and quenching are concentration-dependant, any significant sample manipulation such as pre-concentration or dilution will cause large changes in

the EEM. This is illustrated by the very significant differences (profile and intensity) between the EEM of the undiluted (Fig. 1A) and diluted (5 μ L-in-1mL) CD-A1 solutions (Fig. 1H). This sensitivity to both concentration and composition makes the EEM method ideal for analyzing complex mixtures where one requires a rapid method of characterizing the similarity between different batches of cell culture media components or of the media itself. Another advantage of the EEM method is that for these materials, the fluorescence emission intensity is sufficient to allow sample analysis in their formulated, undiluted state, thus enabling online or at-line measurements with minimal sample handling. Reproducibility of the EEM measurements is of substantial importance in establishing the efficacy of fluorescence-based methods for robust cell culture media analysis. In this study, data was collected in triplicate over a six-month period and reproducibility was examined by an MPCA study on each of the three separate EEM datasets for each sample type. In all cases reproducibility was excellent, and this is illustrated in the eRDF case where MPCA of the three separate fluorescence EEM datasets (after Rayleigh scattering correction) generated three significant PCs (Fig. S-3, Supplemental Information). The correlation coefficients were then calculated and compared for each PC (Supplemental information) and we found that PC1 and PC2 from the three datasets are nearly identical. The only significant difference between the different data collections is in terms of PC3. However, this difference is not very significant because PC3 from each of these three runs explains a mere $\sim 0.03\%$ of the total variance in the data. We would suggest that this PC is most likely to include the minor fluctuations inducted by instrumental effects such as excitation light intensity variations. Similar results were obtained for the other samples (Supplemental Information). These measurements (and previous studies²⁶) indicate that EEM measurements are sufficiently reproducible and can be utilized for detailed chemometric testing for quality assessment and sample identification purposes.

One of the key steps in media component analysis is the evaluation of spectral variance for each sample class and the use of this information for the development of a sample quality analysis tool. The CD-A1 samples are used as an example to demonstrate this application because they have detailed and relatively strong EEM spectra. The first step was to undertake MROBPCA using the triplicate data collections from the 31 samples (93 EEM measurements, after Rayleigh scattering removal). The EEM data were first robust mean-centred and then the first three robust PCs were selected to create a MROBPCA model (Fig. 2). A 95% confidence level was set to determine the two critical discriminant threshold parameters: $T_{0.05}^2 = 8.39$ and $Q_{0.05} = 4.75 \times 10^3$. Hotelling's T^2 and Q residual were then calculated for each EEM measurement and visualized (Fig. 2A) in an outlier diagnostic map. The T^2 values are related to the scores whereas the Q residuals reflect the amount of data which is not explained by the MROBPCA model. Taken together this gives a very quick method for evaluating spectral variance of each measurement and thus sample quality. This plot clearly identifies 14 anomalous measurements, some of which may not always be evident when one is evaluating the individual scores plots (Fig. 2B & 2C). These outlying measurements either have larger T^2 values (e.g. #30), or have

larger Q residuals (e.g. #24), or have both $T^2 > T_{0.05}^2$ and $Q > Q_{0.05}$ (e.g. #21). These 14 outlier measurements originated from nine different samples of which four (#13, #21, #24 and #30) gave two or three outlying measurements. This behavior suggests that these four samples have significantly different compositions, which results in their discrimination. The five other outlying measurements were singular events (i.e. only one of the three replicate spectra was an outlier) which are most likely due to experimental/instrumental error and the stringent 95% confidence level. Decreasing the sensitivity by changing the confidence level to 99% decreases the number of outliers to eight of which five relate to samples #24 and #30, with singular outliers for samples #2, #20, and #21. This strongly indicates that these two samples are definitely different compared to the normal samples which should all have the exact same composition. The correct confidence level to employ for the analysis of these types of materials in the context of their industrial use is not known explicitly and would have to be established using some type of performance data, e.g. bioreactor yield. To establish an appropriate level one would have to have access to performance data to determine which media are good and which are poorly performing when used in actual cell culture, this data would then be used to establish the normal sample population. In the case of these particular materials, they were blended across various lots and thus it is not feasible to attribute important process data to individual samples and establish the normal sample population.

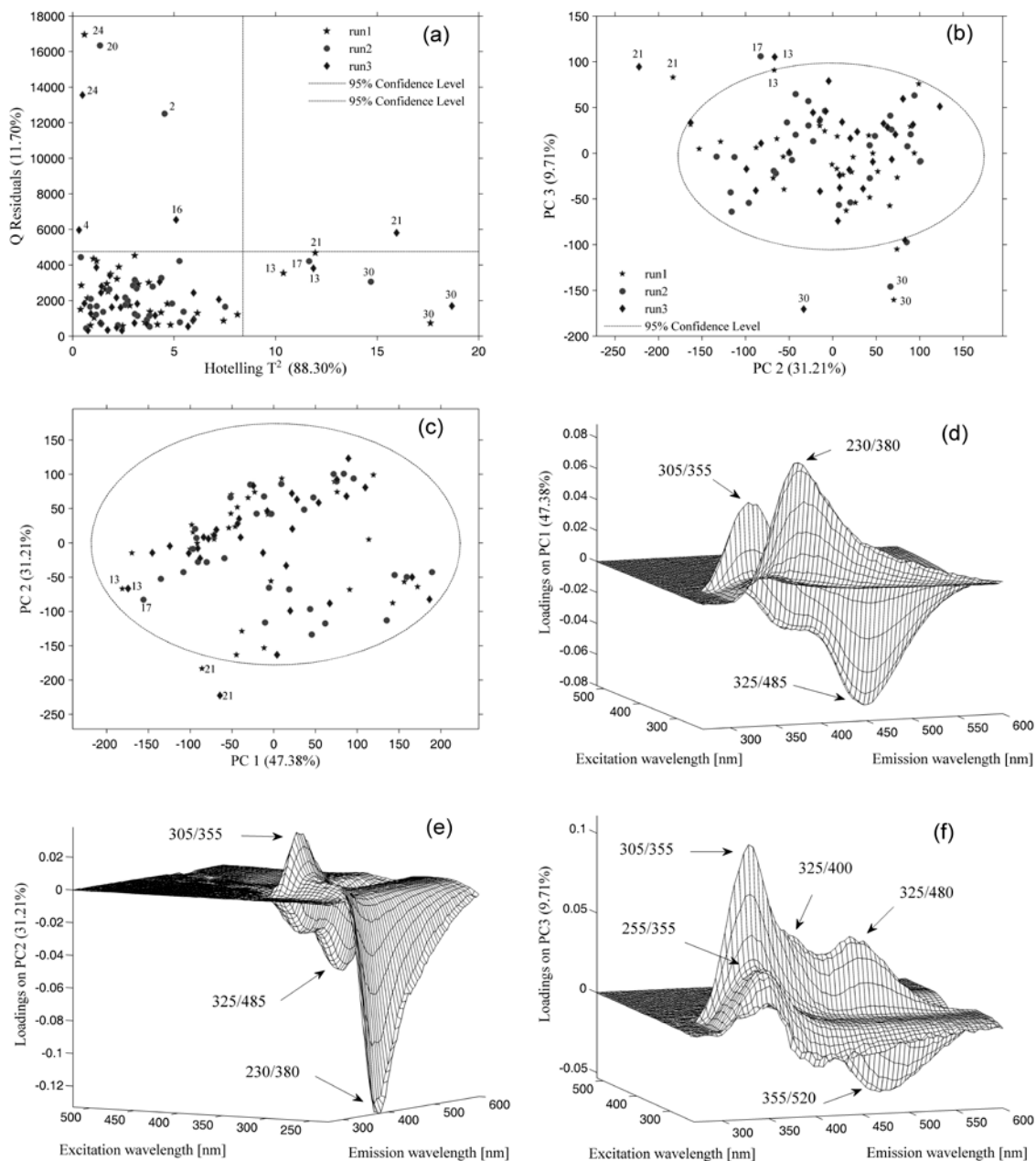


Figure 2: Overview of MROBPCA model, 3 PCs, from 93 EEM measurements of 31 CD-A1 samples from triplicate data collections after being Rayleigh scattering corrected: (a) outlier diagnostic map (Q versus T^2 statistics), (b-c) PC scores plots (ellipses show the 95% confidence level.), and (d-f) loadings plots. In (a) the straight lines define two threshold values of $T^2_{0.05}=8.39$ and $Q_{0.05}=4.75 \times 10^3$ at a confidence level

The PC1 loadings plot (Fig. 2D) indicates that separation of the samples along PC1 is due to a combination of increased fluorescence from two bands ($\lambda_{ex}/\lambda_{em} = 230/380$ nm and 305/355

nm,) and a decrease in another ($\lambda_{\text{ex}}/\lambda_{\text{em}} = 325/485$ nm). For PC2 the largest effect is a sharp negative band at ($\lambda_{\text{ex}}/\lambda_{\text{em}} = 230/380$ nm) whereas the PC3 loadings shows an increase in four different bands and a decrease in a broad band ($\lambda_{\text{ex}}/\lambda_{\text{em}} = 355/520$ nm). As yet we have not identified with complete certainty the identities of the species responsible for each of these bands, because the loadings plots are not representative of pure components and also IF and ET effects cause significant changes to emission band profiles. Despite this, we can use this to assess the changes in samples, so for instance when we compare the significant we find that #21 is different from #13 along PC2, while #13 is different from #30 mostly along PC3.

In order to determine the reason why these samples were abnormal, a closer examination was made in terms of the spectral variance explained by each robust PC. First, the 79 normal CD-A1 EEM measurements identified by the MROBPCA model were averaged into one *mean* EEM (this is equivalent to the robust centre used for mean-centering in the MROBPCA modeling). At the same time, another four average EEM were generated for each of the outliers from the three spectra collected for each sample. The difference spectra between the outliers and the *mean* normal spectrum show that each of the samples is an outlier because of different spectral reasons (supplemental information). However, correlating the bands observed in the PCs with specific species is not always feasible because in many cases the PCs within the MROBPCA model are not representative of pure constituents. We can ascribe the differences as being caused by changes in a combination of IF, ET, and quenching processes, induced by compositional changes. This leads to changes in the fluorescence emission from the tryptophan, tyrosine, and phenylalanine present in the CD-A1 sample, which in turn changes the overall spectral profile.

Table 1: Variance explained by the individual robust PCs for the four CD-A1 outlier samples (i.e. samples #13, #21, #24, and #30) compared to the normal sample #10. Calculation undertaken using the averaged Rayleigh scattering corrected EEM spectrum for each sample which had the mean spectrum (calculated from the 79 normal EEM measurements) subtracted.

variance explained (%)	by PC1	by PC2	by PC3	cumulative
sample #13	64.36	9.07	19.59	93.02
sample #21	9.48	69.10	13.23	91.81
sample #24	0.32	11.45	0.15	11.92
sample #30	2.97	4.12	87.78	94.87
sample #10	84.71	0.08	8.99	93.78

The correlation between the difference spectra and the PC loadings from the MROBPCA model provide some additional clues to the source of sample variability. To assist in the comparison, one typical, normal CD-A1 sample, #10, was selected and its average spectrum generated. This particular sample was selected because all 3 measurements were: close to each other (good reproducibility and small variability), all had small T^2 and Q values in the

MROBPCA model, and their average spectrum approached (but was not identical to) the mean spectrum of the normal samples (i.e., the robust centre used for mean-centering in the MROBPCA model). Table 1 shows how much of the sample variance for each of the difference spectra (Fig. S-4, supplemental info) is explained by each of the individual robust PCs generated by MROBPCA (Fig. 2). From this table, one can observe that compared to the normal sample #10, the difference spectra for the four outlier samples show very different variance patterns where #13 has most explained variance in PC1 (64.36%), #21 is very different with PC2 (69.10%) being the dominant PC, whereas #30 has most explained variance in PC3 (87.78%). The cumulative explained variance for the difference spectrum generated from #24 is only 11.92%, which means that it is very different from the other samples. This large difference is due to compositional variation and is the reason why it has a very large Q value (Fig. 2A) and is easily identified by MROBPCA as an outlier.

For some CD-A1 samples, quantitative analysis was undertaken at BMS, Syracuse to quantify some of the amino acid species that possibly contribute most to sample variability. For the majority of the amino acid analyses (n=22) on the samples tested (n=7), the amino acid concentrations were all within 5% of the mean value. However, for the outlier #30, it was found that there was a very significantly lower aspartic acid concentration (~22%) and a significantly larger asparagine monohydrate (~12%). For #24, the aspartic acid concentration was low by ~23% and the asparagine monohydrate high by ~12%. In terms of the more strongly fluorescent amino acids, these two outliers, #30 and #24, had only marginally lower concentrations of tyrosine (3.3% and 4% respectively), while for phenylalanine the decrease was greater (4% and ~5% respectively). In the case of tryptophan, #30 has a higher concentration by ~1.5%, while for #24 it is lower by ~3.5%.

Table 2: Overview of the MROBPCA models created for CD-A1, CD-A2, CD-S2, insulin, eRDF, yeastolate, and phytone sample solutions.

PCA model	total measurements	normal measurements	samples	outlier samples	PCs	explained variance (%)
CD-A1	93	79	30	4	3	88.30
CD-A2	93	80	28	4	2	94.72
CD-S2	90	78	27	3	3	96.79
H. R. insulin	57	46	16	3	2	88.89
eRDF	63	52	18	4	2	92.90
yeastolate	63	56	19	2	3	87.55
phytone	9	9	3	n/a	2	92.69

For all sample classes, similar MROBPCA analyses were undertaken, the real compositional outlier samples were identified, and their quality assessed. The limited sample set sizes dictated that we use all of the triplicate spectral data for MROBPCA modeling (Table 2). For each model, the $T_{0.05}^2$ and $Q_{0.05}$ thresholds were compared to the Hotelling's T^2 and Q residual statistics of the individual samples for sample quality evaluation. Based on the individual MROBPCA identifications, a total of 68 outlier measurements were excluded, and the 400 *normal* measurements (Table 2) were then combined to develop a single NPLS-DA model. The best NPLS-DA model used six latent variables (LVs) to explain a total of 99.39% of the spectral variance and 95.13% of the variance associated with dummy variables for each sample class. This enabled clear and distinct separation between all seven sample types (Fig. 3A), thus the method can be used as the basis for rapid identification of these materials. While this may seem rather trivial, it should be noted that all of these solutions appear virtually identical to the human eye and can be easily confused. In a biopharmaceutical manufacturing context where there may be compliance requirements for specific identity tests before media blending, this relatively simple method may be of value.

Multivariate analysis of variance (MANOVA)^{51,52} methods were implemented to calculate the within-group and between-group variances for these seven classes of aqueous materials and to detail numerically the reason why these seven classes of media components can be discriminated (Table 3). We find that the within-group and between-group variances vary greatly in terms of magnitude with the former being much smaller than the latter for each class. This indicates that one can use the NPLS-DA procedure as a reliable method for the identification of these complex mixtures using fluorescence EEM data. Table 3 also gives an indication about sample quality as the within-group variance will be larger for the samples which have the largest compositional variation. CD-S2 gives the largest degree of variation with a within-group variance, nearly 3.5 times that of the next most variable sample type, eRDF. Spectral overlap is also an issue which can be observed here as we see that the between-group variances of the CD-A1, CD-A2, and phytone samples are relatively small because all three sample types have a common intense band spanning the 300~320 nm excitation and 350~450 nm emission range. Yeastolate and phytone solutions also show small between-group variance, this is because their major emission bands overlap in similar regions (300~320 nm excitation and 350~400 nm emission range, Supplemental Information).

Table 3: Evaluation of the NPLS-DA model performance in terms of variances within and between the sample classes. YE = yeastolate. The bold values correspond to within-group variance.

variance ($\times 10^5$)	CD-A1	CD-A2	CD-S2	insulin	eRDF	YE	phytone
CD-A1	0.16	2.96	58.42	14.69	38.48	6.23	1.93

CD-A2	2.96	0.11	63.18	17.42	38.54	7.48	2.16
CD-S2	58.42	63.18	2.78	71.40	89.20	37.03	19.20
insulin	14.69	17.42	71.40	0.14	26.22	22.06	9.82
eRDF	38.48	38.54	89.20	26.22	0.80	35.84	13.78
YE	6.23	7.48	37.03	22.06	35.84	0.43	1.27
phytone	1.93	2.16	19.20	9.82	13.78	1.27	0.13

In order to test the efficacy of these fluorescence-based methods from a standpoint of robustness, performance and applicability, a validation sample set was generated by BMS in Syracuse and was analyzed blind in Galway (Table S-9, supplemental information). These 17 samples mostly comprised of CD media of various classes that had been spiked with varying amounts of different materials, some of which will have a direct impact on the fluorescence emission while others will not. For each validation sample, EEM data was collected from two separate aliquots, these measurements were then averaged and the Rayleigh scattering correction was then performed. Finally, they were analyzed using the NPLS-DA calibration model (Fig. 3A) to identify these unknown samples. Eleven samples were assigned to various groups, but six (A, H, K, L, O, and P) were not identifiable as belonging to any class. Samples A, K, L, and P were not fluorescent, and thus are unclassifiable using EEM. Samples H and O were fluorescent, but could not be identified as belonging to any of the classes. It should be noted that the NPLS-DA calibration models used were built from the pure samples and as such the method should only allow for the ID and quality assessment of pure materials and not mixtures of these materials.

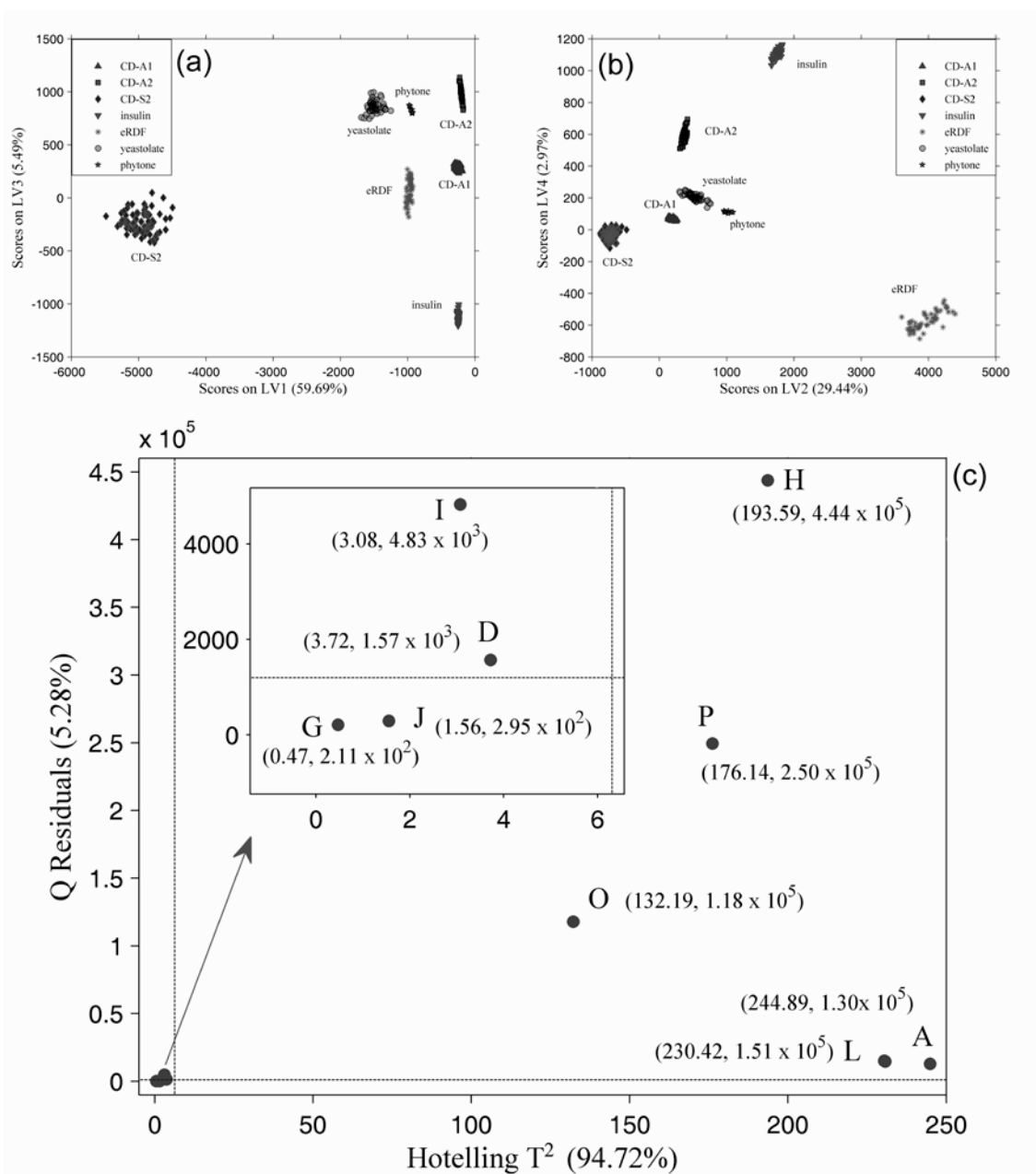


Figure 3: (a) NPLS-DA discrimination (scores plots for LV3 vs. LV1 (left), and (b) LV2 vs. LV4 (right)) of the 400 measurements using the Rayleigh scattering corrected fluorescence EEM spectra of the seven sample types, (c) Quality evaluation of samples A, D, G, H, I, J, L, O, and P by the CD-A2 MROBPCA calibration model. The straight lines define two threshold values of $T_{0.05}^2 = 6.31$ and $Q_{0.05} = 1.20 \times 10^3$ at a confidence level of 95% for the MROBPCA analysis with two robust PCs.

Now we must also consider the quality of the validation samples, to see if we can identify the adulterated samples. For this we use a detailed MROBPCA analysis of the samples and consider the T^2 and Q values generated for the validation samples (Table S-9, supplemental). For example, D, G, I, and J were identified as belonging to CD-A2, and indeed in every case CD-A2 is the main component. However, only sample G is 100% pure, and this is reflected by the fact that the corresponding T^2 and Q values are much smaller than the critical thresholds ($T_{0.05}^2 = 6.31$ and $Q_{0.05} = 1.20 \times 10^3$). Sample J was spiked with 1 mg/mL KH_2PO_4 , but this obviously has no effect on the fluorescence emission as the T^2 and Q values are much smaller than the critical thresholds. This illustrates one of the issues with an EEM based quality assessment method, in that in order to observe adulteration or deviations from specifications, the species must have some impact on the fluorescence emission. In many cases this will be so, however, in this case, we cannot observe the adulteration. In the case of samples D and I, they have also been spiked with non-fluorescent materials, but these materials have a relatively large impact on the fluorescence emission and this is evident in the T^2 and Q values (Fig. 3c and Table S-9, supplemental information). When we consider the diagnostic map of these and the unidentified A, H, L, O, and P samples generated from the 2-PC MROPCA calibration model of the pure CD-A2 materials, it is easy to see which samples are good, and which samples are clearly bad. Using this map, one can clearly see that A, H, L, O, and P were not CD-A2 type materials because all of them had far larger T^2 and Q values than the CD-A2 class critical thresholds ($T_{A,H,L,O,P}^2 \gg T_{0.05}^2 = 6.31$ and $Q_{A,H,L,O,P} \gg Q_{0.05} = 1.20 \times 10^3$).

Repeating this analysis for all MROBPCA models shows similar results with only sample Q (100% yeastolate) passing the T^2 and Q criteria, but sample E (pure CD-A1) comes very close. For the samples identified as belonging to the CD-S2 grouping (B, C, F, N), sample C (the only other 100% pure material) had high T^2 and Q values, indicating that it is not of the same quality as the samples used to make the model. However, considering that CD-S2 samples have the largest within-class variance of all the sample types (Table 3 and Fig. 3a), then this may be due to sample stability. This is not surprising for CD-S2 because it contains folic acid and riboflavin, both of which are fluorescent, but which are also chemically unstable and this stability is linked to environmental factors (e.g. temperature, light, etc.). Sample B (poloxamer spiked) also failed the quality assessment, which may be due to the reasons given above, since we would not expect that poloxamer would have any detectable effect on the fluorescence at the concentrations used. Sample F (NaCl spiked) likewise failed the quality assessment, but again not by much. Sample N (identified as a CD-S2 type) failed the quality assessment by a large margin which is not surprising since it was spiked with a significant amount of CD-A1 (20% by vol.) and this has a large impact on fluorescence emission which is reflected in large T^2/Q values. When the mixtures are changed to 1:1 (sample O) or 1:4 (sample H) the NPLS-DA models fails to identify the mixtures because the impact on the fluorescence emission is much greater.

The validation study shows that the EEM method is suitable for identifying a variety of potential contaminants in the various media components. For operational use in media characterization one would need to first risk assess the potential for contamination by specific

materials, or incorrect addition of known components (this is an open ended problem). Then one could rapidly generate specific models to determine the levels at which they cause significant differences in the EEM data. This will not be suitable for every possible eventuality, but it may suffice for detecting many of the most probable errors in the preparation of complex cell culture media.

CONCLUSIONS:

This study has demonstrated that using a combination of fluorescence EEM spectroscopy and robust chemometric methods one can generate an effective, rapid (<5 minutes) and inexpensive analytical method for the identification and quality assessment of complex cell culture media components. This is achieved with minimal sample preparation and the use of standard off-the-shelf spectrometers. Since the industrial application need is focused on identification and quality assessment, we do not have to undertake any corrections for IF effects of the EEM data. However, in the application of these methods to different sampling geometries (such as in-line measurements), then care will need to be taken to account for the excitation/emission pathlength effects.

The use of a combination of supervised and unsupervised learning methods provides for robust data analysis. Unsupervised methods like MROBPCA are first used to build up a picture of sample variances and determine which samples should be used for identification and quality assessment purposes. Second, once the correct samples have been selected, definitive models can be constructed for use in a supervised learning method like NPLS-DA for the routine identification of incoming materials. Finally, a more detailed MROBPCA model incorporating batch record data (product yield and performance) can be utilized to assess product quality before usage and thus predict media/raw material performance. In practice, these methods can be improved and refined if there are larger numbers of samples, process information, and analytical data, available for each sample class from which to inform and build the identification and quality calibration models.

5. Acknowledgements: This work was funded by the Irish Industrial Development Authority and Bristol-Myers Squibb under the Centre for BioAnalytical Sciences. Bryan Ray, Lindy Smith, and John deVincentis are thanked for their work in sample management.

6. References:

- 1 T. Cartwright and G.P. Shah, Culture Media. In: Davis, J.M. Editor, *Basic Cell Culture*, 2nd ed. (Oxford University Press Inc., New York, 2002, pp 69-106).
- 2 G. Sitton and F. Srienc, *J. Biotechnol.* **135**, 174 (2008).
- 3 E. Jain and A. Kumar, *Biotechnol. Adv.* **26**, 46 (2008).
- 4 K.F. Wlaschin and W.S. Hu, *Fedbatch culture and dynamic nutrient feeding*, in *Cell Culture Engineering*, (Springer: New York, 2006, pp 43-74).
- 5 H. Murakami, T. Shimomura, T. Nakamura,; H. Ohashi,; K. Sinohara,; H. Omura, *J. Agricult. Chem. Soc. Japan*, **58**, 575 (1984).
- 6 D. Jayme, T. Watanabe, and T. Shimada, *Cytotechnology*, **23**, 95 (1997).
- 7 M. Schroder, K. Matischak, and P. Friedl, *J. Biotechnol.* **108**, 279 (2004).
- 8 J.R. Lakowicz, *Principles of Fluorescence Spectroscopy*, 3rd ed. (Springer Science and Business Media LLC, New York, 2006).
- 9 G. Weber, *Nature* **190**, 27 (1961).
- 10 I. M. Warner, G. D. Christian, E. R. Davidson, and J. B. Callis, *Anal. Chem.* **49** 564 (1977).
- 11 I.M. Warner, E. R. Davidson, and G. D. Christian, *Anal. Chem.* **49** 2155 (1977).
- 12 C.-N. Ho, G. D. Christian, and E. R. Davidson, *Anal. Chem.* **50**, 1108 (1978).
- 13 A.G. Ryder, *J. Fluoresc.* **14**, 99 (2004).
- 14 S. Elliott, J.R. Lead, and A. Baker, *Water Res.* **40**, 2075 (2006).
- 15 C. Lindemann, S. Marose, H.O. Nielsen, and T. Scheper, *Sens. Actuators, B*, **51**, 273 (1998).
- 16 E. Sikorska, T. Gorecki, I.V. Khmelinskii, M. Sikorski, and D. de Keukeleire, *Food Chem.* **96**, 632 (2006).
- 17 A.F. Zuluaga, U. Utzinger, A. Durkin, H. Fuchs, A. Gillenwater, R. Jacob, B. Kemp, J. Fan, and R. Richards-Kortum, *Appl. Spectrosc.* **53**, 302 (1999).
- 18 D. Patra and A.K.Mishra, *TrAC-Trends Anal. Chem.* **21**, 787 (2002).
- 19 P.G. Coble, *Mar. Chem.* **51**, 325 (2006).
- 20 G.J. Hall, K.E. Clow, and J.E. Kenny, *Environ. Sci. Technol.* **39**, 7560 (2005).
- 21 G.J. Hall and J.E. Kenny, *Anal. Chim. Acta*, **581**, 118 (2007).
- 22 J. Sadecka and J. Tothova, *Czech J. Food Sci.* **25**, 159 (2007).
- 23 A.M. de la Pena, A.E. Mansilla, N.M. Diez, D.B. Gil, A.C. Olivieri, and G.M. Escandar, *Appl. Spectrosc.* **60**, 330 (2006).
- 24 B. Li, P.W. Ryan, B.H. Ray, K.J. Leister, N.M.S. Sirimuthu, and A.G. Ryder, *Biotechnol. Bioeng.* **107**, 290 (2010).
- 25 A.G. Ryder, J. de Vincentis, B. Li, P.W. Ryan, N.M.S. Sirimuthu, and K.J. Leister, *J. Raman Spectrosc.*, **41**, 1266 (2010).
- 26 P.W. Ryan, B. Li, M. Shanahan, K.J. Leister, and A.G. Ryder, *Anal. Chem.* **82**, 1311 (2010).
- 27 E.K. Read, J.T. Park, R.B. Shah, B.S. Riley, K.A. Brorson, and A.S. Rathore, *Biotechnol. Bioeng.* **105**, 276 (2010).

- 28 E.K. Read, R.B. Shah, B.S. Riley, J.T. Park, K.A. Brorson, and A.S. Rathore, *Biotechnol. Bioeng.*, **105**, 285 (2010).
- 29 A. Rinnan and C.M. Andersen, *Chemometr. Intell. Lab. Syst.* **76**, 91 (2005).
- 30 R.G. Zepp, W.M. Sheldon, and M.A. Moran, *Mar. Chem.*, **89**, 15 (2004).
- 31 M. Bahram, R. Bro, C. Stedmon, and A. Afkhami, *J. Chemometr.* **20**, 99 (2006).
- 32 L. Ståhle and S. Wold, *J. Chemometr.* **1**, 185 (1987).
- 33 M. Sjöström, S. Wold, and B. Söderström, In *Pattern Recognition in Practice II*; Gelsema, E. S., Kanal, L. N.; Eds.; Elsevier: Amsterdam, 1986, p 486.
- 34 E.K. Kemsley, *Chemom. Intell. Lab. Syst.* **33**, 47, (1996).
- 35 S. Wold, J. Trygg, A. Berglund, and H. Antti, *Chemom. Intell. Lab. Syst.* **58**, 131-150, (2001).
- 36 M. Barker and W. Rayens, *J. Chemometr.* **17**, 166 (2003).
- 37 H. Nocairi, E. Mostafa Qannari, E. Vigneau, and D. Bertrand, *Comput. Stat. Data Anal.* **48**, 139 (2005).
- 38 J.E. Jackson and G.S. Mudholkar, *Technometrics* **21**, 341, (1979).
- 39 I.T. Joliffe, *Principal Component Analysis*, 2nd ed. (Springer-Verlag: New York, 2002).
- 40 M. Hubert and S. Engelen, *Bioinformatics*, **20**, 1728 (2004).
- 41 Hubert, M.; Rousseeuw, P.J.; Vanden Branden K. *Technometrics*, **47**, 64 (2005).
- 42 G.Y. Li and Z.L. J. Chen, *Am. Stat. Assoc.* **80**, 759 (1985).
- 43 C. Croux and A. Ruiz-Gazen, In *COMPSTAT 1996: Proceedings in Computational Statistics*, Prat, A., Ed. (Physica-Verlag: Heidelberg, 1996, pp 211-217).
- 44 M. Hubert, P.J. Rousseeuw, and S. Verboven, *Chemom. Intell. Lab. Syst.* **60**, 101 (2002).
- 45 P.J. Rousseeuw and K. van Driessen, *Technometrics*, **41**, 212 (1999).
- 46 R.Y. Liu, J.M. Parelius, and K. Singh, *Ann. Stat.* **27**, 783 (1999).
- 47 Y. Zuo, H. Cui, and D. Young, *Ann. Stat.* **32**, 189 (2004).
- 48 S.O. Algar, N.R. Martos, and A.M. Díaz, *Spectrosc. Lett.* **36**, 133 (2003).
- 49 Larsson, T.; Wedborga, M.; Turner, D. *Anal. Chim. Acta.* **2007**, 583, 357-363.
- 50 R. D. Holbrook, P.C. DeRose, S.D. Leigh, A.L. Rukhin, and N.A. Heckert, *Appl. Spectrosc.* **60**, 791 (2006).
- 51 H.R. Barker and B.M. Barker, *Multivariate Analysis of Variance (MANOVA): A Practical Guide to Its Use in Scientific Decision-Making*, (University of Alabama Press, Birmingham, 1984).
- 52 C.J. Huberty and S. Olejnik, *Applied MANOVA and Discriminant Analysis*, 2nd ed. (John Wiley & Sons: Hoboken, New Jersey, 2006).