<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Rapid quantification of tryptophan and tyrosine in chemically defined cell culture media using fluorescence spectroscopy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Ryder, Alan G.; Calvet, Amandine; Li, Boyan</td>
</tr>
<tr>
<td><strong>Publication Date</strong></td>
<td>2012</td>
</tr>
<tr>
<td><strong>Publication Information</strong></td>
<td>Calvet, A, Li, BY, Ryder, AG (2012) 'Rapid quantification of tryptophan and tyrosine in chemically defined cell culture media using fluorescence spectroscopy'. Journal Of Pharmaceutical And Biomedical Analysis, 71 :89-98.</td>
</tr>
<tr>
<td><strong>Link to publisher's version</strong></td>
<td><a href="http://dx.doi.org/10.1016/j.jpba.2012.08.002">http://dx.doi.org/10.1016/j.jpba.2012.08.002</a></td>
</tr>
<tr>
<td><strong>Item record</strong></td>
<td><a href="http://hdl.handle.net/10379/3937">http://hdl.handle.net/10379/3937</a></td>
</tr>
<tr>
<td><strong>DOI</strong></td>
<td><a href="http://dx.doi.org/10.1016/j.jpba.2012.08.002">http://dx.doi.org/10.1016/j.jpba.2012.08.002</a></td>
</tr>
</tbody>
</table>
Rapid Quantification of Tryptophan and Tyrosine in Chemically Defined Cell Culture Media using Fluorescence Spectroscopy. A. Calvet, B. Li, and A. G. Ryder. * 
Nanoscale BioPhotonics Laboratory, School of Chemistry, National University of Ireland, Galway, Galway, Ireland
* Corresponding author: Email: alan.ryder@nuigalway.ie Phone: +353-91-492943.

ABSTRACT:
The rapid and inexpensive analysis of the complex cell culture media used in industrial mammalian cell culture is required for quality and variance monitoring. Excitation-emission matrix (EEM) spectroscopy combined with multi-way chemometrics is a robust methodology applicable for the analysis of raw materials, media, and bioprocess broths. We have shown that the methodology can identify compositional changes and predict the efficacy of media in terms of downstream titre [1]. Here we describe how to extend the measurement methodology for the quantification of specific tryptophan (Trp), tyrosine (Tyr) in complex chemically defined media. The sample type is an enriched basal RDF medium in which five significant fluorophores were identified: Trp, Tyr, pyridoxine, folic acid, and riboflavin. The relatively high chromophore concentrations and compositional complexity lead to very significant matrix effects which were assessed using PARAllel FACtor analysis2 (PARAFAC2). Taking these effects into account, N-way Partial Least Squares (NPLS) combined with a modified standard addition method was used to build calibration models capable of quantifying Trp and Tyr with errors of ~4.5 and 5.5% respectively. This demonstrates the feasibility of using the EEM method for the rapid, quantitative analysis of Trp and Tyr in complex cell culture media with minimal sample handling as an alternative to chromatographic based methods.

Keywords: Fluorescence, cell culture media, chemometrics, amino acids, inner filter effect.

1. Introduction
The use of chemically defined cell culture media (CD-media) for industrial mammalian cell culture is of significant importance because of the need to eliminate the use of complex, ill-defined biogenic materials like bovine serum. These CD-media are often highly complex mixtures, containing a variety of amino acids, carbohydrates, cofactors, and other materials [2]. Media composition varies according to cell type, product type, and are, often propriety formulations for individual manufacturers [3-5]. Media are often formulated using...
components which themselves are complex, chemically defined (CD) media mixtures e.g. RDF [6]. RDF is a 2:1:1 mixture of RPMI(Roswell Park Memorial Institute) 1640, DMEM (Dulbecco's Modified Eagles Medium), and F12 (a Ham's Nutrient Mixture medium) media, which is often used as a basal medium for mammalian cell fermentation. An enriched basal RDF (eRDF) is formulated by increasing the levels of amino acids and glucose to sustain high density growth. In mammalian cell culture, eRDF is widely used as a basal medium [7]. eRDF can vary in composition but typically comprises over 30 compounds including inorganic salts, amino acids, vitamins, HEPES buffer, glucose, and various others [8, 9].

The analysis of these complex CD-media is challenging since comprehensive analysis requires a variety of chromatographic separations coupled to methods like mass spectrometry. This can make routine analysis expensive and time-consuming particularly for simple requirements such as raw materials or media constituent identification, determination of lot-to-lot consistency, and the monitoring of batch-to-batch preparations of media. Thus there is a need for a rapid, holistic analytical method that can be used for the early stage screening and analysis of these materials. Spectroscopic methods offer the ideal solution because of advantages ranging from speed, sensitivity, facile automation, and inexpensive unit test costs. Raman spectroscopy can be used for the identification and quality monitoring of cell culture media components including eRDF [10, 11]. However, quantification of specific components is difficult because the analyte signals are typically weak and there may be fluorescent interferences. For aqueous solutions, the combination of Excitation-emission matrix (EEM) spectroscopy with multi-way chemometrics methods is one very attractive option [12].

EEM spectroscopy when applied to complex, turbid mixtures containing biogenic fluorophores such as amino acids, proteins, coenzymes, and vitamins can enable simultaneous detection, qualitative and quantitative analysis [13-15]. EEM based analyses, which take only a few minutes to perform also benefit from the intrinsic advantages of fluorescence spectroscopy which include: high sensitivity, high signal-to-noise ratios, and relatively large linear ranges for quantitative analysis [16]. For complex CD-media mixtures where there are multiple fluorophores and other photophysically active molecules, Energy Transfer and quenching (both static and dynamic) play a large part in determining the shape and intensity of the EEM profile, thus providing a unique fingerprint suitable for qualitative characterization of these materials [17-22]. The use of multi-way chemometrics methods for the qualitative and quantitative chemometrics analysis of EEM data is now well established in environmental analysis of water [23-28], food science [29], and for quantitative determination of analytes (µg/mL) in various fluids such as urine [30] or plasma [31]. It was demonstrated that EEM combined with chemometric methods can be used to identify and characterize cell culture media and to predict the performance of cell culture media in terms of product yield [1, 12] as well as retrieve key parameters for on-line monitoring of yeast cultures [32]. An alternative methodology Surface Enhanced Raman Scattering (SERS) can also provide valuable information about complex media variance to complement EEM measurements [33]. However SERS is not a sufficiently mature or reproducible technique at
the present time. Thus EEM-chemometric method is an ideal process analytical technology
candidate for the assessment of critical quality and performance attributes of the complex
materials used to prepare cell culture media [34-37].
This study extends previous work [1, 12] and demonstrates the next stage in the realization of
EEM based analytical methods for biopharmaceutical manufacturing, namely the
quantification of specific media constituents. In CD-media such as eRDF the fluorescent
amino acids, tryptophan (Trp) and tyrosine (Tyr), are often important components [9]. Tyr in
particular is a key amino-acid used for protein synthesis and Tyr concentration in the media
has been shown to be correlated with titre in CHO based antibody production [38]. Another
recent study describing the optimization of media for CHO cell culture by metabolic flux
analysis shows clearly that Trp is also an important, limiting component [39]. Thus the
ability to rapidly quantify variances in the concentrations of these amino acids in the complex
cell culture media is of interest from a quality control point of view.
Our objective in this study was to develop an accurate, fast, inexpensive, with minimal
sample handling, generic method for the quantification of Trp and Tyr in a complex CD-
media (eRDF) in its prepared, concentrated state. The EEM technique described in this
paper, is fast in order to provide real time analysis (~ 5 min per sample), non-destructive, and
does not require any complex sample handling, apart from pipetting the sample into a cuvette.
It should be noted that a single EEM measurement can be analyzed using a variety of
different chemometric models for different purposes, like identification [12], then quality
control/variance analysis [12, 33], and even process performance prediction [1]. While some
effort is required in the initial phase of developing the chemometric models, for routine day-
to-day application, the models can be integrated into the spectrometer software to enable
semi-autonomous use. Finally, because all the data is digital and the measurements are
reproducible, one can collect data throughout the life of the bioprocess and continually update
the chemometric models during the process lifecycle. This should result in increased
prediction accuracy as the bioprocess matures.
However, the multitude and relatively high concentrations of chromophores and fluorophores
present in CD-media at the working concentrations leads to a variety of complications in the
EEM method in terms of matrix effects which have to be dealt with. Here we use multi-way
chemometric analysis to deal with the matrix effects, enabling media analysis without
dilution, so minimizing sample handling. The goal of the study is to accurately quantify
Trp/Tyr concentrations where there is little a priori knowledge of the analyte concentrations.

2. Materials and Methods
2.1 Materials:
eRDF was obtained from Kyokuto Pharmaceuticals Industrial (Japan). NaOH (97+ %),
NaHCO₃ (99.7+ %), L-Tyrosine (≥ 98%), L-Tryptophan (≥ 98%), Pyridoxine, (-)-Riboflavin
and Folic acid dehydrate (97 %) were obtained from Sigma-Aldrich and used without further
purification. An aliquot of sterilized high purity water was used to dissolve 0.885 g of eRDF powder, to which was added 0.0565 g of sodium bicarbonate before making the solution up to a final volume of 50 mL (17.7 g/L working concentration, eRDF Stock). The solution was immediately sterilized by membrane (0.22 µm) filtration and then dispensed into sterile containers before storage at −70ºC. The pH of the eRDF Stock solution was 7.0 with Trp and Tyr concentrations of 90.1 µM and 480 µM respectively (manufacturer reported formulation). Stock solutions of Trp (18.0 mM) in water and Tyr (96.0 mM) in an alkaline solution were prepared so that the addition of 5 µL of stock solution would double the concentration of these amino acids in a 1 mL aliquot of the eRDF stock. After spiking 1mL of eRDF stock solution with 10 µL of Tyr stock solution the sample pH was 7.3. A variation from pH 7.0 to 7.3 is not expected to affect significantly the fluorescence properties of Tyr and Trp [40].

2.2 Instrumentation and data collection:
EEM were measured over spectral ranges of 220-400 nm (excitation) and 250-600 nm (emission) with a data interval of 5 nm at 25ºC using a Cary Eclipse (Varian) fluorescence spectrometer [1]. Semi-micro quartz cuvettes (Lightpath Optical Ltd., UK) with a 4 mm axis (excitation) and a long 10 mm axis (emission) were used. All EEM collected constituted 37x71 matrices where each data point was the fluorescence intensity of the sample at a specific excitation/emission pair. Fig. S-1 (Supplementary data) shows the landscape and contour plots of a typical eRDF EEM. All calculations were performed using PLS_Toolbox 4.0®, supplemented by in-house-written codes for MATLAB® (ver. 7.4). All figures were generated using MATLAB® (ver. 7.4) or Origin ver. 7.0, OriginLab Corporation, Northampton, MA.

2.3 Samples
Two EEM datasets were collected: dataset1 was used to assess matrix effects and validate the band assignments and comprised of EEM collected from a series of ten eRDF solutions with concentrations between 3.5 and 35 g/L. Each sample was measured in triplicate (30 spectra) to form a 37x71x30 tri-dimensional array. For the quantification of Tyr and Trp in eRDF, a sample set (dataset2a) spanning the nominal to three times the nominal specified concentration (for a 17.7 g/L eRDF soln.) was prepared by means of a standard addition protocol: 1 mL portions of the eRDF stock solution were successively spiked with 1 µL aliquots of either Trp (up to a maximum of 8 µL) or Tyr (up to a maximum of 10 µL) stock solutions. Each sample was prepared in triplicate (Table 1). When more than 8 µL of Trp was added the signal became saturated under the measurement conditions used and this therefore defines the measurement/concentration range for the method. We note that if a

---

1 NaOH was added to improve solubility. 6.25 mL of ~1M NaOH was added to 434.8 mg of Tyr and the solution was made up to a final volume of 25 mL with sterile millipore water. The final pH was 11.5.
classical standard addition method were implemented with NPLS, then the absolute concentrations are used for calibration, and thus the models produced will have calibration ranges which have a lower limit equal to the initial analyte concentration (i.e. the specified concentration in sample 0). Consequently, it would not be possible to accurately quantify samples where the analyte concentrations are lower than the specified concentration.

Table 1: Description of the standard addition samples. Sample 0 is the test sample.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Tyrosine</th>
<th>Tryptophan</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Added concentration (µM)</td>
<td>Total concentration (µM)</td>
</tr>
<tr>
<td>Sample 0</td>
<td>0</td>
<td>480.2</td>
</tr>
<tr>
<td>Sample 1</td>
<td>95.4</td>
<td>575.6</td>
</tr>
<tr>
<td>Sample 2</td>
<td>190.6</td>
<td>670.8</td>
</tr>
<tr>
<td>Sample 3</td>
<td>285.6</td>
<td>765.8</td>
</tr>
<tr>
<td>Sample 4</td>
<td>380.4</td>
<td>860.6</td>
</tr>
<tr>
<td>Sample 5</td>
<td>475.1</td>
<td>955.3</td>
</tr>
<tr>
<td>Sample 6</td>
<td>569.5</td>
<td>1049.7</td>
</tr>
<tr>
<td>Sample 7</td>
<td>663.8</td>
<td>1144.0</td>
</tr>
<tr>
<td>Sample 8</td>
<td>757.9</td>
<td>1238.1</td>
</tr>
<tr>
<td>Sample 9</td>
<td>851.7</td>
<td>1331.9</td>
</tr>
<tr>
<td>Sample 10</td>
<td>945.4</td>
<td>1425.6</td>
</tr>
</tbody>
</table>

For each series of samples, the non-spiked samples (stock eRDF solutions, e.g. sample 0 in Table 1) were designated as Test samples, the spiked ones as Standard addition (SA) samples (samples 1 to 8/10 in Table 1). The calibration dataset comprised of 54 spiked samples while the test dataset comprised of 6 stock eRDF samples. An 11 sample prediction dataset (dataset2b) was also generated comprising 1 stock eRDF (sample 0), 3 samples spiked only with Trp at concentrations equal to samples 1 to 3 (Table 1), 5 spiked with only Tyr at concentrations equal to sample 1 to 5 and an additional sample spiked with both Trp and Tyr at the concentrations specified for sample 3.
2.4 Chemometric methodology:

Prior to analysis the first and second order Rayleigh scattering bands were removed from the EEM data by replacing with missing values. This was done because the Rayleigh scatter is largely unrelated to the chemical properties of the sample and the scatter peaks do not behave linearly (or tri-linearly) which can bias the EEM data modeling. The advantages of using missing data in order to deal with Rayleigh scattering in the case of the application of a PARAFAC algorithm were described by Andersen [41]. The eRDF solutions were significantly fluorescent and there were signals over the majority of the EEM space and there was no evidence of any Raman bands in the EEM, so Raman scatter correction was not implemented.

As the goal of this study is to determine the concentrations of specific amino acids in a complex cell culture media, we used a variety of methods. Initially, PARAFAC2 [42, 43] was implemented on dataset1 to assess the various spectral factors. This method was used in preference to PARAFAC because PARAFAC is highly dependent on data tri-linearity [44]. This means that the fluorescence response measured for one analyte in two different samples should have the same profile (excitation and emission profiles) but scaled differently according to the analyte concentration. Here we observed significant concentration induced spectral profile changes with the presence of a varying saddle in the Trp excitation profile (fig. S-5, Supplementary data); consequently the simple PARAFAC algorithm was not suitable. Instead we used PARAFAC2 because the method accommodates changes in the response profiles of one of the dimensions. PARAFAC2 often used to deal with shift issues which are of great importance in NMR, or chromatographic data handling [42, 45-50]. The use of PARAFAC2 combined with EEM is less popular but it was previously used to model pH induced fluorescence emission shifts [51]. Here we implemented PARAFAC2 with the excitation mode set as the ‘changing’ mode.

Finally an N-way partial least squares (NPLS) calibration model [52] was developed from the data in dataset2, using a modified standard addition method (MSAM), [53, 54] to show that it was feasible to predict the concentrations of Trp/Tyr varying about the specified formulation concentrations without relying on precise a priori knowledge of analyte concentration. In this method one subtracts the EEM matrix of the Test sample from all the ‘raw’ EEMs (Fig. 1a/b) of the S4 samples to build the calibration model. We designated these test matrix subtracted EEM data as the TMS-EEM (Fig. 1c/d). Quantitation was then implemented by an external calibration procedure carried out on the ‘raw’ (un-subtracted) EEMs. Using this method the nominal concentration of analytes in the Test sample was not required as the calibration was performed on the residual data (i.e. the TMS-EEM) which originated primarily from the change in concentration and relates to the added concentrations. Sample quality was assessed using multi-way (unfolded) classic principal component analysis (U-PCA) to ensure that no abnormal samples were included in the calibration sets.

The key factors in developing accurate calibration models were to: 1). determine the best sample set, 2). choose the best data pre-processing (PreP) methods [55] and 3). select an
optimal spectral range. For the sample sets, we had to decide on whether to build models using only samples spiked with a single analyte or use a selection of samples spiked alternatively with both. A second consideration in the use of NPLS [56] was the inclusion of just Trp or Tyr concentrations or both in the Y variables element of the model. Eventually we selected five different cases shown in Table 2. Calibration model performance was assessed using a combination of the root mean square error of calibration (RMSEC), RMSE of validation (RMSEV) and RMSE of prediction (RMSEP).

RMSEC represents the ability of the model to predict the added concentration of Tyr and Trp after subtraction of the reference Test sample. The RMSEV represents the ability that the model has to predict the total concentration of Trp and Tyr in the SA samples within the calibration range (as determined by the concentration range of the added analyte). For validation the ‘raw’ EEMs from dataset2b were used to calculate RMSEV. The RMSEP for the Test samples (RMSEPT) was calculated as a final model assessment step. To determine the model complexity in terms of appropriate number of latent variables (LVs), we calculated and compared the RMSEC, RMSEV and RMSEP values up to 15 LVs. The appropriate number of LVs for the best models was then chosen where both RMSEV and RMSEP reached a common minimum (e.g. LV = 4 for Trp on the full EEM, Table 4).

3. Results and Discussion
3.1 Spectral & Variance analysis:
A typical eRDF EEM landscape plot (Fig. 1 and Fig. S-1, Supplementary data) shows a relatively complex topography indicating the presence of multiple fluorophores. Comparing this data with fluorescence spectra from individual fluorophores and literature data we identified Tyr and Trp as the strongest fluorophores, with weaker contributions from pyridoxine (Py), riboflavin (RF), and folic acid (FA) (Table 2, Figs. S-2/3/4 Supplementary data). All of these are specified components of eRDF with concentrations varying from 0.53 to 480 µM.
**Table 2:** Summary of the fluorophores identified from the EEM data of eRDF solutions and relevant photo physical parameters [1-6].

<table>
<thead>
<tr>
<th>Fluorophore</th>
<th>Conc. (a) (µM)</th>
<th>λ&lt;sub&gt;max&lt;/sub&gt; excitation (nm)</th>
<th>λ&lt;sub&gt;max&lt;/sub&gt; emission (nm)</th>
<th>I&lt;sub&gt;max&lt;/sub&gt; (a. u.)</th>
<th>Quantum yield (c)</th>
<th>Extinction coefficient (d) (cm&lt;sup&gt;-1&lt;/sup&gt; mM&lt;sup&gt;-1&lt;/sup&gt;)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tyr</td>
<td>480</td>
<td>275</td>
<td>310</td>
<td>305</td>
<td>448.9</td>
<td>0.14 (i) 1.45 (275nm)</td>
</tr>
<tr>
<td>Trp</td>
<td>90.1</td>
<td>260 – 285 (b)</td>
<td>355</td>
<td>350</td>
<td>513.1</td>
<td>0.13 (ii) 6.24 (279 nm)</td>
</tr>
<tr>
<td>Pyridoxine</td>
<td>4.86</td>
<td>320</td>
<td>390</td>
<td>390</td>
<td>36.79</td>
<td>0.11 (iii) 8.78 (325 nm)</td>
</tr>
<tr>
<td>Riboflavin</td>
<td>0.53</td>
<td>270, 370 and 445</td>
<td>520</td>
<td>520</td>
<td>10.96</td>
<td>0.267 (iv) 35.1 (265 nm)</td>
</tr>
<tr>
<td>Folic acid</td>
<td>19.9</td>
<td>355</td>
<td>~355 (acid) / 375 (base)</td>
<td>445</td>
<td>445 (acid) / 455 (base)</td>
<td>10.26</td>
</tr>
</tbody>
</table>

(a) In a 17.7 g/L eRDF solution.
(b) Band centred at 275 nm.
(c) i) ± 0.01 in H<sub>2</sub>O at 275 nm and 23°C, ii) ± 0.01 in H<sub>2</sub>O at 280 nm and 23°C, iii) ± 0.01 in 0.05 M PPB at 25°C, iv) ± 0.01 in H<sub>2</sub>O at pH=7, v) in aqueous solution at pH =5.0 and pH=10.
(d) Calculated at the maximum absorption wavelength in brackets from the absorption spectrum (Fig. 2d).

Typical Trp and Tyr concentration variation observed in CD-media used in industrial applications are expected to be lower than that used here because the media are supposed to be manufactured to a specific recipe with very high tolerances. There has been little data published work in the literature on the lot-to-lot variation in CD-media, apart from our own studies [1, 11, 12] and other recent studies [57]. An earlier EEM study [12] on another eRDF type media (21 lots) used in an industrial process was undertaken on samples at a much lower concentration (~0.84 g/L). The measured lot-to-lot fluorescence variation can be compared by the inter- and intra-sample variability. This can be done by comparing the standard deviation of specific band intensity measurements across all samples and for the triplicate measurements of the same sample. Here we took the points of maximum intensity in the Trp/Tyr bands from normalized EEM spectra and this yielded intra-/inter-sample variability ratios of 1:2.8 and 1:3.7 respectively for these dilute eRDF samples [12]. In other words, the observed lot-to-lot EEM differences were ~3 times the measurement error, based on this...
simple ratio measurement. In contrast, for dataset2a (much more concentrated), the
1 equivalent values were 1:35 (Trp) and 1:47 (Tyr).
2 However, we have to recognize that these two sample systems have different photophysical
3 behavior (and EEM spectra) and thus it is not proper to compare only the changes in
4 fluorescence intensity at a few discrete points. We have to consider the shape changes in the
5 EEM contours which are much more sensitive to inner filter effects (IFE) caused by the
6 medium complexity. We combined the eRDF dataset from this work with the dilute sample
7 datasets from the previous study [12], and then normalized the spectra. We then undertook
8 an Unfolded Principal Component Analysis (U-PCA) study to gauge the degree of change
9 that can be expected in reality. The U-PCA results (Supplementary data) shows that the
10 magnitude of the lot-to-lot variation was comparable to the spectral variation generated by
11 the spiked samples.
12
13 We must also address the issue of fluorescent contaminants and what factors may adversely
14 affect the implementation of these quantitative methods. Some media components can have a
15 very large effect on EEM signal via quenching or energy transfer, while others will have
16 comparatively minor or no effect at all [12]. In the context of CD-media prepared under
17 Good Manufacturing Practice (GMP) for large scale industrial processes one expects a high
18 degree of reproducibility in formulation and thus reproducible EEM spectra. Any significant
19 deviations in the formulation process should be easily identified using U-PCA prior to the
20 implementation of the quantitative analysis for Trp/Tyr. An example of how one can use U-
21 PCA to check sample quality is the photodegradation of CD-media where it was easy to
22 identify (Supplementary data, Fig. S-13) those samples which have photodegraded. 
23 Photodegradation of eRDF solution leads to the formation of a new fluorescence band at 355
24 / 460 nm (excitation/emission pair) from an unknown component. In practice, these
25 identified abnormal samples can then be rejected or sent for a more-comprehensive and time-
26 consuming detailed chromatographic analysis. Alternatively, if one has knowledge of
27 intrinsic variability of a known fluorophore in the media, then this can be accounted for in the
28 calibration dataset by careful experimental design, as was implemented here for dataset2a
29 (Table 3). Since every industrial CHO based process uses different, bespoke CD-media, we
30 did not develop this method for any specific bioprocess and thus we used a large calibration
31 range to demonstrate the feasibility of the concept. The method can easily be modified to
32 lower variation ranges by careful experimental design and the preparation of new calibration
33 samples.
34
35 Table 3: Description of the different calibration cases (CalC) built based on the nature of the
36 X (EEM) and Y (concentration) datasets used for calibration.

<table>
<thead>
<tr>
<th>Calibration cases</th>
<th>X includes samples spiked with …</th>
<th>Number of samples in X</th>
<th>Y includes added concentrations of …</th>
<th>Number of column in Y</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Page 9 of 24
3.2 PARAFAC2 modelling.

To assess the matrix effects and validate peak assignments (Table 2), dataset1 was modeled using PARAFAC2 [41, 58]. For PARAFAC2 to generate individual factors/components corresponding to different compounds those compounds must be present with varying relative concentrations in the different samples, i.e. the measured responses must not be collinear. In dataset 1 which is simply composed of samples made by diluting a known concentration of a complex mixture, the relative concentrations of the analytes do not change and thus you have a collinear concentration system. However, since we are measuring the fluorescence response (the EEM), the situation is very different. The fluorescence response of the individual fluorophore is not collinear, in other words each fluorophore’s emission behaves differently as the samples are diluted because of IFE. Thus the IFE mediated non-collinearity of dataset1 enabled PARAFAC2 to generate a unique component (loadings and scores plots are shown in Fig. 2) for the major fluorophores.

PARAFAC2 required four components to adequately model eRDF (dataset1) and the emission loadings plots (Fig. 2b) clearly ascribe the first three components to Tyr, Trp, and Py emission respectively while the fourth component is a composite. This fourth component represents fluorescence from RF and FA with emission maxima at 420 nm and 515 nm. The corresponding excitation loading features maxima at 250, 355, and 370 nm which also correspond to a composite of the excitation spectra of FA and RF (Table 2, Fig. S-2 and 4 supplementary data). FA and RF have much weaker fluorescence, with poor signal-to-noise ratios, leading to imprecision in the discrimination of emission maxima. They may also display a similar fluorescence behavior in response to concentration variations, and this co-linearity in the EEM data from dataset1 prevented PARAFAC2 from resolving the individual components. The use of MCR-ALS (supplementary data) did not improve the resolution of these two fluorophores.

Fig. 2a shows the effect of varying eRDF concentration on the excitation spectra and as expected, major changes were observed in the Trp and Tyr loadings with a pronounced saddle visible in the Trp excitation loading. The saddle deepens with the growing IFE caused by the increasing eRDF concentration. In the electronic absorption spectrum of the stock eRDF solution (Fig. 2d), the maximum absorption (275 nm) corresponds to the excitation wavelength of the Trp saddle. This relatively high absorption reduces both excitation and emission intensity, and the effect varies according to wavelength, affecting each fluorophore differently. The PARAFAC2 scores plot (Fig. 2c), shows the quantitative impact of IFE on the fluorescence intensities of the four components. It is obvious that there are very
significant non-linearity, particularly for Tyr emission intensity (Fig. 2c) and Trp excitation profile (Fig. 2a). There is no obvious evidence in this data of any significant non-radiative energy transfer; in particular there is no Trp emission blue shift with Tyr or Trp addition. IFE is particularly evident in Tyr and Trp fluorescence because of the high intrinsic concentrations and the very significant spectral overlap of the media absorbance maximum with their fluorescence excitation and emission spectra.

3.3 Quantitative analysis:

There are several second-order calibration algorithms available for use with EEM of which PARAFAC is one of the most common [41, 59]. Unfortunately, the extensive IFE present in these types of samples precludes its use, and the use of many other algorithms [60]. However, unfolded partial least squares (U-PLS) [61] and NPLS, have the benefit of being better able to handle the tri-linearity deviations caused by IFE [62]. The use of U-PLS with Residual Bi-linearization to deal with IFE has been demonstrated for simple relatively mixtures [59]. In the NPLS modeling we used dataset2a, for calibration and validation, and dataset2b, for prediction. Fig. 1a/b shows typical raw EEM and Fig. 1c/d shows the TMS-EEM used for building the calibration model. It is obvious that the topographical changes in the EEM spectra are very large and varied when spiked with double the nominal concentration.

Plotting the impact of spiking the eRDF stock solution with Trp/Tyr (Fig. 3) clarifies the matrix effects solely introduced by the presence of the spiked analytes. Fig. 1 and 3 shows the effect on the excitation spectra of Trp (c) and Tyr (d) when eRDF is spiked with increasing amounts of each amino acid. Plotting the changes in fluorescence intensity of the various fluorophore with Trp/Tyr addition (Fig. 3a/b) showed that there was very little impact from the changing matrix on the fluorescence behavior of the vitamins. A more significant impact can be noted from the amino acid on each other; however it will be shown that, when necessary, this problem can be handled by considering the two analytes simultaneously when building a calibration model. Overall the deviations to linearity are less dramatic than that observed when the eRDF concentration itself was varied (Fig. 2c). We conclude that NPLS could deal with these effects via the use of additional LV’s in the modeling process [59].

Table 4: Summary of results from NPLS modeling of the full EEM data for prediction of Trp and Tyr concentrations.

<table>
<thead>
<tr>
<th>Analyte</th>
<th>Model type</th>
<th>LVs</th>
<th>PreP(^{(a)})</th>
<th>PreP(^{(a)})</th>
<th>RMSEC(^{(b)})</th>
<th>RMSEP(^{(b)})</th>
<th>REC(^{(b)})</th>
<th>REV(^{(b)})</th>
<th>REP(^{(b)})</th>
<th>REPT(^{(b)})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trp</td>
<td>CalC1</td>
<td>3</td>
<td>C1</td>
<td>C1</td>
<td>0.97 (39.44)</td>
<td>35.23 (38.63)</td>
<td>1.20</td>
<td>29.2</td>
<td>33.1</td>
<td>42.9</td>
</tr>
<tr>
<td></td>
<td>CalC3</td>
<td>5</td>
<td>C1</td>
<td>C1</td>
<td>2.32 (7.19)</td>
<td>8.68 (6.50)</td>
<td>6.47</td>
<td>5.33</td>
<td>8.16</td>
<td>7.21</td>
</tr>
<tr>
<td></td>
<td>CalC4</td>
<td>4</td>
<td>C1</td>
<td>C1</td>
<td>1.64 (2.99)</td>
<td>5.36 (2.99)</td>
<td>4.57</td>
<td>2.22</td>
<td>5.04</td>
<td>3.32</td>
</tr>
</tbody>
</table>

Page 11 of 24
In the first phase of the quantitative analysis we used the full EEM spectra after verifying the quality of the calibration sample datasets using a 2-component U-PCA model (supplemental info, fig. S-6). Table 4 shows a selection of NPLS results obtained using the full EEM data and details the combination of pre-processing, LV’s and datasets used for calibration used to obtain the best predictions. It is clear that we can generate a reasonably accurate model for Trp with relatively low errors (Relative Error of Prediction, REP ~ 5%) using the full EEM. However, using this data for Tyr prediction generates poor predictability, irrespective of the conditions used. This is a consequence of the extensive non-linearity in the data arising from the fact that Tyr emits at shorter wavelengths and is thus more sensitive to IFE in both the excitation and emission modes. For Trp, IFE is more significant for the excitation relative to the emission mode because Trp emits at longer wavelengths (> 305 nm) where the absorbance of the medium is relatively low. In most cases it was necessary only to centre the EEM data, which is a common pre-processing step implemented when performing NPLS on EEM data. It is interesting to note how the addition of the Tyr spiked samples in the calibration dataset (CalC3 and CalC4) improved the Trp prediction. As Tyr and Trp signals are partially overlapping and because we do not use a model presenting the second order advantage, the presence and variation in concentration of the secondary analyte could lead to poor model robustness. The secondary analyte (Trp or Tyr) can be considered here as a known and expected interferant and by including this extra interferant information in the calibration sets the model can be trained to take into account its presence during the prediction step. Including the samples spiked with both Trp and Tyr (CalC3, CalC4 and CalC5) can give information to the system that the variations in a specific region of the EEM are unrelated to the observed concentration and thus the NPLS algorithm will ascribe a lower weighting to the variables from these regions.

To improve prediction accuracy particularly for Tyr determination, we sub-sampled the EEM data and then re-ran an extensive series of NPLS models under different conditions. Ultimately, two regions of interest (Fig. S-9, Supplementary data) were selected. These regions (R1/R2) were chosen in order to try and minimize IFE, avoid the Rayleigh scatter, while still retaining sufficient spectral information (See details Table 5). R1’s wide excitation range includes 6.0% missing values (compared to 6.7 % for the full EEM). R2...
used a much more restricted, longer-wavelength excitation range (only 4 excitation wavelengths), and included 3.7% missing values. Thus the proportion of missing values in the datasets are relatively low and do not seem to have much of an effect on prediction accuracy.
Table 5: Summary of results for the models giving best prediction of Trp and Tyr using selected EEM regions.

<table>
<thead>
<tr>
<th>An.</th>
<th>Model type</th>
<th>LV</th>
<th>(R_{\text{ex}}^{(a)}) (nm)</th>
<th>(R_{\text{em}}^{(a)}) (nm)</th>
<th>Pre-P^{(b)}</th>
<th>REC^{(c)} (%)</th>
<th>REV^{(c)} (%)</th>
<th>REP^{(c)} (%)</th>
<th>REPT^{(c)} (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trp (R1)</td>
<td>CalC3</td>
<td>10</td>
<td>220-310</td>
<td>285-495</td>
<td>C1S2,AS</td>
<td>2.03</td>
<td>2.99</td>
<td>4.23</td>
<td>3.29</td>
</tr>
<tr>
<td>Tyr (R2)</td>
<td>CalC5</td>
<td>6</td>
<td>285-300</td>
<td>295-495</td>
<td>C1S2,C</td>
<td>4.88</td>
<td>5.04</td>
<td>3.76</td>
<td>4.60</td>
</tr>
</tbody>
</table>

(a) \(R_{\text{ex}}\) and \(R_{\text{em}}\) are the excitation and the emission range defining the sub-sample of the original EEM used.
(b) PreP, Pre-processing on \(X, Y\). Centering (C or C1), Scaling on mode 2 and/or 3 (S2 and/or 3), Autoscaling (AS).
(c) REC, REV, REP are the relative error of calibration, validation and prediction respectively; those errors are calculated from the RMSEC, RMSEV and RMSEP relatively to the mean of the expected values in each case.

To avoid IFE, the absorbance should always be less than 0.05; however this is not possible with media, as we do not wish to implement a dilution step in the method. The eRDF stock absorption is greater than 0.05 at all wavelengths shorter than 495 nm for a 4 mm pathlength (See Fig. 2d). Thus we cannot select an IFE free region without discarding all the spectral information pertaining to Trp/Tyr. However, low wavelength cut-off at 285 or 295nm, where the solution absorbance starts to increase exponentially (Fig. 2d), was successfully applied on the emission mode for R1/R2 and on the R2 excitation mode limiting the IFE effect on model quality.

NPLS models for predicting Trp (using R1) and Tyr (using R2) with errors better than 5.5 % and 4.5% respectively (Table 5) were generated (Fig. 4). For these models more complicated pre-processing was required with scaling of the emission and excitation modes in addition to the classical centering. Scaling (here to the variance unit) standardizes variables so they all display the same variability [55]. Scaling within the excitation mode was necessary in many cases and this is consistent with the hypothesis that scaling reduces some of the non-linearity arising from IFE induced intensity changes which are more evident in the excitation mode due to the spectral ranges affected. Indeed, the strongest IFE occurred between 220 and 310 nm which overlapped the Trp and Tyr excitation bands. Scaling on the emission mode was also proved to be of use in some cases for Tyr prediction (Table S-1, Supplementary data).

For both Tyr and Trp prediction, it was noticeable that the addition of more variance information in the calibration dataset (CalC3-5) also improved the quality of the predictions as observed on the full EEM data (Table S-1, Supplementary data). This extra information given to the calibration dataset (X) can also be supplemented by the introduction of the secondary analyte concentration in \(Y\) (CalC3 case).
4. Conclusions

This study shows that fluorescence EEM coupled with multi-way chemometrics methods is a rapid, effective and inexpensive method requiring minimum sample handling, for the quantitative analysis of tyrosine and tryptophan in a complex, concentrated, chemically defined cell culture media. PARAFAC2 was used to identify the various fluorophores and show the very significant IFE present in these samples. Despite the extensive IFE present, it was possible to quantify both Trp and Tyr using NPLS coupled with a modified standard addition method, and by careful selection of the EEM range used. Both Trp and Tyr could be quantified with an accuracy of 4.5 and 5.5% using different chemometric models applied to different spectral regions of the same EEM measurement. Thus a single rapid (<7 minutes) measurement can be used for the quantification of these two important amino acids, in contrast traditional chromatographic analysis may require a much longer analysis time. This when coupled to previous work on qualitative analysis of media components [12], and correlating media variance with process performance [1] allows for the use of fluorescence EEM as a comprehensive analytical tool for biotechnology manufacturing. These results are significant in an industrial QC environment, because the method does not require any significant sample handling and can be implemented on a wide variety of complex aqueous solutions. Further improvement in the prediction accuracy of the method is limited by the experimental design used here because the calibration samples spanned a large concentration range (up to twice nominal concentration). With such a large compositional change we induce significantly large IFE and non-linearity in the data which will cause difficulties in obtaining higher accuracy. If greater accuracy is desired then, the calibration range can be reduced to cover a smaller (more realistic) concentration range and a larger number of calibration samples used.

5. ACKNOWLEDGEMENTS

This work was supported by funding from IRCSET (the Irish Research Council for Science, Engineering & Technology) to AC. AC would also like to thank Dr. Denisio Togashi for assistance to AC.

SUPPLEMENTAL INFORMATION AVAILABLE

Supporting information is available including further details on spectral analysis of the EEM data, identification of the fluorophores, PCA studies of eRDF variability, and additional figures and data.

REFERENCES

DOI: 10.1016/j.jpba.2012.08.002


DOI: [10.1016/j.jpba.2012.08.002](https://doi.org/10.1016/j.jpba.2012.08.002)


1 Figures

Figure 1: Landscape plots showing the changes in the EEM spectra of the eRDF media (17.7 g/L) spiked with (a) Trp (to twice nominal concentration, 179.4 µM) and (b) Tyr (to twice nominal concentration, 955.3 µM) and (c and d) the corresponding test matrix subtracted EEM (TMS-EEM) spectra used in the calibration modeling.
Figure 2: Summary of the PARAFAC2 model generated using dataset1 (varying eRDF concentration). Loadings plots for the excitation mode (a), emission mode (b), and sample mode (c). The absorption spectrum (d) of the 17.7 g/L eRDF stock solution (4 mm pathlength). Legend: Trp (□), Tyr (○), pyridoxine (△), and the mixture of folic acid and riboflavin (▽). The average of the loadings and scores of the 3 replicates for each sample are shown.
Figure 3: Plots of changes in fluorescence intensity for specific excitation/emission wavelength pairs, Tyr (■, 275/310 nm), Trp (●, 275/355 nm and ▲, 285/355 nm), pyridoxine (□, 320/390 nm), riboflavin (○, 365/520 nm) and folic acid (Δ, 355/445 nm), as the Trp (a) or Tyr (b) concentration is changed, and plots of the eRDF excitation spectra as Trp and Tyr concentration are increased at the (c) Trp maximum emission wavelength (360 nm) and at the (d) Tyr maximum emission wavelength (310 nm).
Figure 4: Plots of predicted versus expected Trp (top) and Tyr (bottom) concentrations for the model calculated using the restricted range EEMs. The calibration (○) and validation (△) plots (a and c), and prediction plots (b and d) are shown, and the line represents the diagonal where the predicted concentration is equal to the expected concentration.