



**HAL**  
open science

## Qualitative and quantitative analysis of therapeutic solutions using Raman and infrared spectroscopy

Alaa Makki, Franck Bonnier, Renaud Respaud, Fatma Chtara, Ali Tfayli, Clovis Tauber, Dominique D. Bertrand, Hugh Byrne, Elhadi Mohammed, Igor Chourpa

### ► To cite this version:

Alaa Makki, Franck Bonnier, Renaud Respaud, Fatma Chtara, Ali Tfayli, et al.. Qualitative and quantitative analysis of therapeutic solutions using Raman and infrared spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* [1994-..], 2019, 218, pp.97-108. 10.1016/j.saa.2019.03.056 . hal-02866199

**HAL Id: hal-02866199**

**<https://univ-tours.hal.science/hal-02866199>**

Submitted on 22 Oct 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

## **Qualitative and quantitative analysis of therapeutic solutions using Raman and Infrared spectroscopy**

Alaa A. Makki<sup>1</sup>, Franck Bonnier<sup>\*1</sup>, Renaud Respaud<sup>2</sup>, Fatma Chtara<sup>1</sup>, Ali Tfayli<sup>3</sup>, Clovis Tauber<sup>4</sup>, Dominique Bertrand<sup>5</sup>, Hugh J. Byrne<sup>6</sup>, Elhadi Mohammed<sup>7</sup>, Igor Chourpa<sup>1</sup>

<sup>1</sup>Université François-Rabelais de Tours, EA 6295 Nanomédicaments et Nanosondes, 31 avenue Monge, 37200 Tours, France

<sup>2</sup>Université François-Rabelais de Tours, UMR 1100, CHRU de Tours, Service de Pharmacie, F-37032 Tours, France

<sup>3</sup>U-Psud, University of Paris-Saclay, Lip (Sys)<sup>2</sup>, EA7357, UFR-Pharmacy, Châtenay-Malabry, France

<sup>4</sup>Université François-Rabelais de Tours, INSERM UMR 1253 iBrain, 37000 Tours, France.

<sup>5</sup>Data\_Frame, 25 rue Stendhal, 44300 Nantes, France

<sup>6</sup>FOCAS Research Institute, Technological University Dublin, Kevin Street, Dublin 8, Ireland

<sup>7</sup>Department of Pharmacognosy, Faculty of Pharmacy, University of Gezira, Sudan

### **\*Corresponding author**

Franck Bonnier

EA 6295 NMNS "Nanomédicaments et Nanosondes"

UFR Sciences Pharmaceutiques

31, avenue Monge, 37200 Tours

E-Mail: [franck.bonner@univ-tours.fr](mailto:franck.bonner@univ-tours.fr)

Tel: +33 2 47 3 67307

Key words: Infrared spectroscopy, Raman spectroscopy, quantitative analysis, discriminative analysis, anthracyclines.

## Abstract:

Anticancer drugs are prescribed and administered to an increasing number of patients on a daily basis. As a consequence, a number of concerns have been raised about the patient health and safety in the case that the drugs administered are not at the required concentration or even worse not the correct ones. Quality control of therapeutic solutions has therefore been extensively implemented in hospital environments, in order to avoid any failure in the intense workflow faced by administering pharmacists. In the present study, infrared (IR) and Raman spectroscopy have been employed for the analysis of 3 commercially available therapeutic solutions TEVA<sup>®</sup>, MYLAN<sup>®</sup>, CERUBIDINE<sup>®</sup>, respectively containing doxorubicin, epirubicin and daunorubicin. They perfectly illustrate the analytical difficulties encountered, as these 3 chemotherapeutic drugs are isomers, hardly distinguishable with conventional approaches such as UV/VIS spectrometry. Any analytical failure to identify these molecules can lead to delays in patient treatment. While Partial Least Squares Regression analysis demonstrates that both Raman and IR can deliver satisfactory quantitative analysis in the clinical range, with respective Root Mean Square Error of Cross Validation (RMSECV) between 0.0127 – 0.0220 g.L<sup>-1</sup> and 0.0573 – 0.0759 g.L<sup>-1</sup>, the identification rate between the 2 techniques differs substantially. Indeed, Principal Component Analysis – Factorial Discriminant Analysis (PCA-FDA) highlights that, depending on the data preprocessing applied to Raman spectra, the discrimination between the 3 drugs is decreased, with in some cases specificity and sensitivity below 50%. However, IR analysis displays encouraging results with an overall specificity and sensitivity between 99-100%, suggesting that reliable validation of the therapeutic solution for administration to patients can be achieved. IR and Raman spectroscopy could assist and support quality control of chemotherapeutic solutions prepared in personalised concentrations for each patient. The effective and reliable characterisation of therapeutic solutions could have a lot to offer to improve current practices in a near future.

## **I. Introduction:**

Evolutions of clinical procedures and protocols in recent years are motivated by improvement of patient safety. Anticancer drugs are extensively used on a daily basis, the number of patients treated still increasing annually. Errors in medications are the main risk leading to therapeutic failure, which, considering the hazardous nature of most anti-cancers drugs, can have disastrous repercussion for a patient's health. As a result, numerous hospitals have implemented internal regulations to standardise their own procedures and minimise the risk of erroneous treatment of a patient. High performance liquid chromatography (HPLC) remains the gold standard for quantitative analysis of organic compounds as part of daily quality control and is therefore one of most commonly found methods in clinical analytical laboratories attached to preparation facilities. Mass spectrometry is also a powerful technique for drug screening and analysis, extensively employed in the development of new pharmaceutical compounds or patient monitoring through detection and quantification in body fluids such as serum, urine or saliva [1, 2]. However, it is not currently an option for therapeutic screening applications, due to time constraints of the clinical workflow, requirements for sample preparation, and simply the cost involved. However, some limitations such as high operating costs, relatively low output and the requirement of specialised skills make it less than ideal for routine analysis of chemotherapy preparations [3, 4]. Flow-injection analysis (FIA) spectroscopic based methods are the most commonly implemented, with Ultraviolet-coupled to Fourier-transform infrared spectrometry (UV/vis-FTIR) probably the most widespread [5, 6]. Recent evolutions have seen the infrared detection replaced by Raman technologies to overcome some of the limitations, but implementation in clinics still requires an aliquot to be taken for injection into the analytical device [7, 8]. Moreover, investigations are ongoing about the possibility to use Raman spectroscopy as a non-invasive tool to overcome some current limitation such as large volume required for FIA, but mostly to avoid any risk in relation to exposure of staff to hazardous solutions [9].

Vibrational spectroscopic techniques, both infrared (IR) absorption and Raman scattering, deliver specific molecular fingerprinting signatures of samples. They have been extensively studied for biomedical applications [10, 11] can be used and exploited as novel and accurate methods for histopathology [12], cytology [13], biopsy characterisation [14], surgical guidance [15], treatment

monitoring [16] and drug studies [17]. The rapid, cost effective, label free and non-destructive advantages are indeed often pointed out to highlight the benefits of those techniques, but it is the analytical capabilities in terms of both qualitative and quantitative performances are also most relevant for a wide range of clinically oriented applications [18, 19]. IR and Raman spectroscopy are powerful analytical tools widely used for characterisation of organic molecules such as, for example, therapeutic drugs [20]. Although primarily used in industry and research laboratories as screening tools to monitor whether a synthesis process has been successful [21], coupling the molecular specificity of the spectra collected with optimised quantitative approaches strongly promotes the techniques for therapeutic solution screening in the clinical environment [22]. In previous studies, although encouraging results are documented for monitoring a number of molecules such as ganciclovir [23] or 5FU [24], directly in the infusion bag, the material used for the packaging, often polystyrene like, also has strong contributions to the spectra that can limit the usable analytical range and lead to difficulties in quantifying and/or identifying the chemotherapeutic drugs accurately [25]. Therefore, such developments remain only at the research level and no translation into the clinic has been achieved as yet, suggesting that the present workflow, entailing analysis of aliquots from the therapeutic solutions is still the most realistic approach.

In the present study, three commercial formulations, TEVA<sup>®</sup> (Doxorubicin), MYLAN<sup>®</sup> (epirubicin) and CERUBIDINE<sup>®</sup> (daunorubicin) have been used to evaluate the potential of Raman and infrared spectroscopy for the analysis of therapeutic solutions in such a workflow. The chemotherapeutic agents are chemically similar, and thus present difficulties in terms of their identification and/or quantification with UV absorbance spectroscopy. Consequently, a comparative study has been conducted as a demonstration of the potential of vibrational spectroscopy as quality control tools for clinical use. A two-step approach to this demonstration has been implemented, firstly evaluating the discriminative capabilities of both techniques, irrespective of concentrations, by means of Principal Component Analysis (PCA) coupled to Factorial Discriminant Analysis (PCA-FDA); and secondly determination of the precision of the quantitative analysis of the approaches using the Partial Least Squares Analysis (PLSR).

## 2. Materials and Methods:

### 2.1 Anticancer drugs selected and sample preparation:

The study includes three commercial anthracycline forms, TEVA<sup>®</sup> (doxorubicin, injectable solution, 2 g.L<sup>-1</sup>), MYLAN<sup>®</sup> (epirubicin injectable solution, 2 g.L<sup>-1</sup>) and CERUBIDINE<sup>®</sup> (daunorubicin, lyophilised powder, 20 mg) which were provided through a collaboration established with the University Hospital of Tours. While TEVA<sup>®</sup> and MYLAN<sup>®</sup> contain small amounts of HCl and NaOH (<1%), CERUBIDINE<sup>®</sup> has mannitol added to the formulation, in concentrations as high as 100 mg per 20 mg daunorubicin. Mannitol is a commonly found excipient in commercialized daunorubicin-based chemotherapeutics approved for clinical use. The UCBO unit (Unité de Biopharmacie Clinique Oncologique, Tours, France) deals with the preparation of anticancer drug formulations directly on site for rapid administration to patients. Based on recent clinical protocols, a set of 8 aqueous solutions, with drug concentrations of 2 g.L<sup>-1</sup>, 1.5 g.L<sup>-1</sup>, 1 g.L<sup>-1</sup>, 0.5 g.L<sup>-1</sup>, 0.25 g.L<sup>-1</sup>, 0.125 g.L<sup>-1</sup>, 0.0625 g.L<sup>-1</sup> to 0.03125 g.L<sup>-1</sup>, corresponding to clinically relevant concentrations, have been prepared for each chemotherapeutic drug. The samples were prepared from the stock solution using serial dilutions, which were analysed directly with Raman and infrared spectroscopy. The procedure has been repeated 5 times in order to obtain 5 independent sets of spectra. All quoted concentrations refer to the amount of drug present in the solution without taking into account excipients. All samples were prepared freshly on the day of analysis and stored at 4°C in a dark room to avoid photo-damage prior to spectral analysis. [ATR-IR spectra have also been recorded for mannitol 10 g.L<sup>-1</sup> and 0.15625 g.L<sup>-1</sup> which correspond to the concentrations of mannitol found in CERUBIDINE<sup>®</sup> 2 g.L<sup>-1</sup> and 0.03125 g.L<sup>-1</sup>, respectively.](#)

### 1.2 Data collection

The current work flow in place at the UCBO unit (Unité de Biopharmacie Clinique Oncologique, Tours, France) is based on a Multipsec<sup>®</sup> analyser (MICRODOM, France). The pharmacists receive between 100 and 150 anticancer drug solutions on a daily basis which require analysis and clearance before they are administered to patients. For each individual sample, 1.5 mL is extracted to an aliquot and injected in the analyser with a syringe. The protocols employed in this study for

Raman and Infrared analysis have been derived from the current procedures and designed to improve on the analytical performance.

### 1.2.1 Raman spectroscopy analysis

Raman spectra were collected using a Labram spectrometer (Horiba Jobin-Yvon, France) equipped with a 690 nm laser source delivering ~10 mW at the sample. 500  $\mu$ L of the solution were placed in a quartz cuvette and the data was collected by illuminating the sample through the wall of the cuvette. In order to ensure maximum reproducibility between measurements, a macro sampling holder, consisting of a cuvette holder attached to the turret of the microscope was employed. Basically, the laser coming out of the turret is reflected by a 45° mirror and directed through the quartz cuvette. Although the set up requires larger volumes to perform the analysis, the laser is not tightly focused and thus the risk of any photothermal damage is minimised. The spectral range was set between 150- 3700  $\text{cm}^{-1}$ , resulting in a spectral resolution of about 2.5  $\text{cm}^{-1}$ , achieved using a 300 lines/mm grating. Two accumulations of 20 seconds were taken for each spectrum. 5 independent sets of dilutions have been prepared for each formulation and analyzed on different days to capture possible instrumental variability in measurements. Additionally, 5 spectra have been measured for each concentration tested. In total, 200 spectra were recorded over the range of concentrations for each anticancer drug included in this study.

### 1.2.2 Infrared spectroscopy analysis

Infrared spectra were recorded with a Frontier spectrometer (Perkin Elmer, France) equipped with a multi-reflection ( $n = 10$ ) Attenuated Total Reflectance (ATR) accessory (Eurolabo, France). The IR spectra collected result from 16 scans performed over the range 600 – 4000  $\text{cm}^{-1}$  with a spectral resolution of 4  $\text{cm}^{-1}$ . A background was acquired (64 scans) and ratioed with the sample spectra by the software. 2  $\mu$ l drops were deposited directly on the diamond crystal and allowed to air dry prior to analysis. 5 independent replicates have been measured for each concentration tested, 5 spectra recorded each time. In total, 200 spectra were recorded over the range of concentrations

for each anticancer drug included in this study. The time required to record spectra from one drop is about 5 mins, including drying time and data collection, which is comparable to protocols currently employed at the Hospital of Tours.

### 2.2.3 Data handling:

The Raman and IR data sets have been pre-processed and analysed using MATLAB (Mathworks, USA). For the purpose of the study, although the preprocessing and data analysis steps have been broken down as individual procedures subsequently applied to data, one should keep in mind that the optimised data processing and analysis protocols are easily incorporated in instrumental software and implemented as automated functions. Once the protocol has been established, total computational time to apply quantitative and discriminative analysis is less than 1 minute

a) Discriminative analysis: Raman spectra have been cut in order to focus on the **fingerprint** region ( $300\text{-}1800\text{ cm}^{-1}$ ) in which the most relevant spectral features are observed. Pre-processing of Raman spectra included baseline correction, using Lieber or Extended Multiplicative Scattering Correction (EMSC), followed by vector normalisation. While the Lieber algorithm estimates the background by a polynomial function and by an iterative method based on least squares [26], the EMSC is more advanced approach allowing to correct for additive baseline effects, multiplicative scaling effects, and interference effects [27]. For instance, spectra of deionised water were provided to neutralise its contribution to the Raman spectra collected from the different solutions. For consistency, infrared spectra have also been reduced to the fingerprint region ( $600\text{-}1800\text{ cm}^{-1}$ ) prior to being subjected to baseline correction (rubber band) and a vector normalisation. **The lower limit of the spectral range for ATR-IR is defined by the crystal material to be at  $600\text{ cm}^{-1}$ . In contrast, Raman spectroscopy allows to analyse the finger print region down to tens of  $\text{cm}^{-1}$ , 200 in this study.** Once pre-processing was completed, both Raman and IR spectra were analysed by PCA [28] and PCA-FDA (PCA-Factorial-Discrimination-Analysis) [29]. Although PCA is a well-established method for rapid evaluation of the data, highlighting the spectral variability present in data sets, more advanced approaches such as PCA-FDA are required to deliver information about the discriminative performance of the techniques. PCA-FDA includes a 100 fold cross validation through a leave K-out cross validation routine. This approach ensures no replicate of a given sample can be found in both the calibration and validation sets and avoids over optimistic outcome

of the quantitative analysis [30]. 1/3 of the data is selected as a test set, while the 2/3 of the remaining samples were used as calibration and validation sets. The test set is kept constant and at each iteration, different combinations of randomly selected calibration/validations sets are tested. Ultimately, the results are presented in the form of confusion matrices, allowing calculation of the specificity and sensitivity of the discrimination analysis performed.

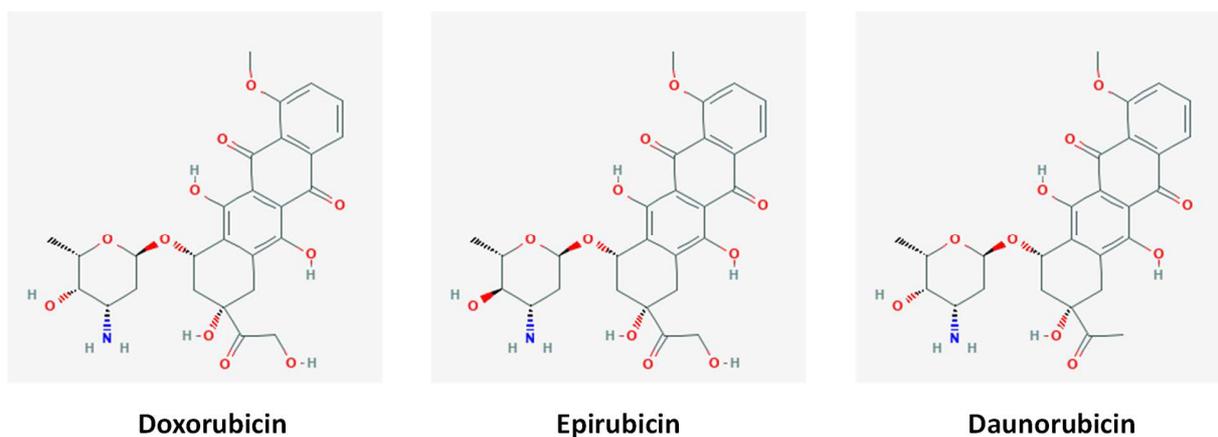
b) Quantitative analysis: Normalisation procedures should be avoided to preserve the intensity to concentration relationship in the data. Therefore, only an offset correction has been applied for the spectra pre-processing. PLSR remains the most relevant analysis to evaluate the possibility of performing quantitative analysis of systematically varying vibrational spectra [31, 32]. It is a supervised multivariate technique that is highly specific for quantitative methods, and is usually performed via two steps: calibration and validation. Therefore, similarly to the PCA-FDA a 100 fold leave K-out cross validation was implemented. The output of PLSR provides performance criteria defined under the linearity ( $R^2$ ) and precision, also called Root Mean Square Error for Cross Validation (RMSECV). Additionally, the lower Limit Of Detection – lower LOD has been estimated based on the work published by Allegrini et al. [33].

### 3. Results and Discussion

#### 3.1 UV characterisation of anthracycline formulation

Anthracyclines are a group of cytotoxic compounds used as treatment of a wide variety of cancers, from leukemia, thyroid, lung and other neoplasia, ovarian, breast, lung and gastric cancer [34] leading to a large number of solutions being prepared on a daily basis for patients under chemotherapeutic treatment. Doxorubicin, epirubicin and daunorubicin are composed of tetracycline aromatic moiety attached to a sugar moiety of daunsamine via a glycosidic bond. As displayed in Figure 1, these 3 anthracyclines have the same aromatic chromophore and the difference between them concerns the side chains. Notably, epirubicin is a stereoisomer of doxorubicin, and differs from it only in the orientation of the hydroxyl group of the hexapyranosyl sugar [34]. Daunorubicin differs from doxorubicin only in the shorter side chain terminated with a methyl group instead of a primary alcohol group. As a direct consequence, those molecules cannot be discriminated using UV spectrophotometry. The absorbance spectra were collected for the 3 molecules in their commercial forms (doxorubicin from TEVA®, epirubicin from MYLAN®

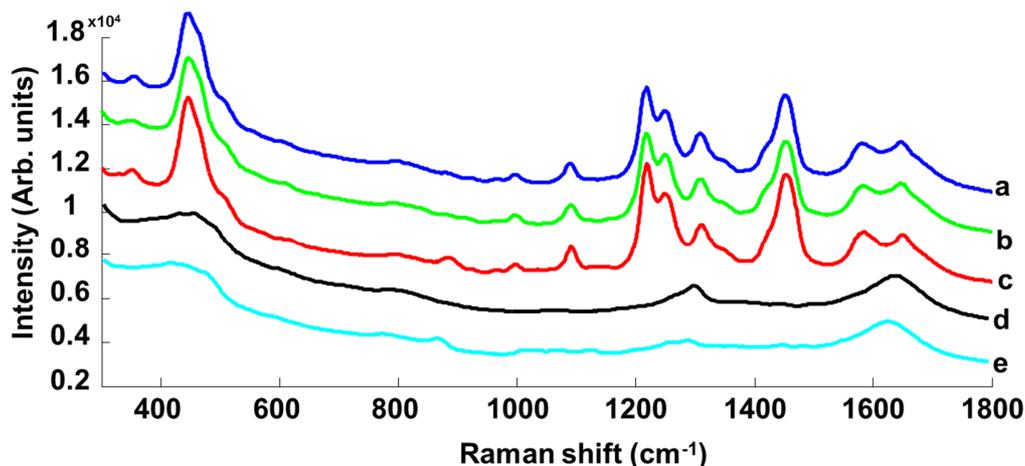
and daunorubicin from CERUBIDINE®. These spectra have the same profile, with the two main maxima located at ca. 480 and 500 nm. NaOH, HCl and mannitol (a sugar) present in the commercial forms do not have direct contribution to the UV/Vis spectra in this range, and thus the absorbance is solely derived from the drugs. While quantitative analysis is expected to be achievable with UV absorbance spectroscopy, difficulties to perform specific identification of the chemotherapeutic drugs are understandable. As they contain information of all the moieties of the molecule, rather than just the chromophoric  $\pi$ -conjugated structures, vibrational spectra are richer in specific molecular structural information than UV-vis absorption spectra. This study aims to demonstrate this and optimise protocols to implement the analysis in a clinical workflow.



*Figure 1: Chemical structures of the 3 anticancer drugs of interest.*

### 3.2. Quantitative and discriminant analysis using Raman spectroscopy

#### 3.2.1. Raman characterisation of anthracycline formulations



*Figure 2: Mean macro-Raman spectra recorded from TEVA® (a), MYLAN® (b) and CERUBIDINE® (c) commercial forms diluted to drug concentrations of 2 g.L<sup>-1</sup> in deionised water; deionised water (d) and manitol solution at 10 g.L<sup>-1</sup> in deionised water (e). Spectra are offset for clarity.*

Figure 2 presents mean Raman spectra collected from the commercial forms of the three chemotherapeutic drugs, TEVA® (spectrum a), MYLAN® (spectrum b) and CERUBIDINE® (spectrum c) diluted to concentrations of 2 g.L<sup>-1</sup> in deionised water. As expected, due to their similar chromophores, spectral signatures display strong similarities, all major peaks being found in all 3 formulations. To further represent the vibrational modes observed, [Table 1](#) provides a direct comparison of their positions with corresponding assignments. The differences observed are relatively minor, especially for TEVA® (spectrum a) vs MYLAN® (spectrum b), which show only a 2 cm<sup>-1</sup> shift between 1089 cm<sup>-1</sup> vs 1091 cm<sup>-1</sup>. With the exception of the band at 886 cm<sup>-1</sup> observed in the case of CERUBIDINE®, all other features appear consistent for the 3 molecules, as seen in [Figure 2](#). Moreover, the recording of Raman spectra from liquid samples can lead to some underlying contribution from water. Although the Raman bands observed have strong intensities, the profile of the signal and the band around 1600-1650 cm<sup>-1</sup> ([Figure 2](#)) suggest that H<sub>2</sub>O (spectrum d) also contributes to the spectra of the drug solutions. The features observed around 450 cm<sup>-1</sup> and 1300 cm<sup>-1</sup> in the water spectrum are respectively assigned to the quartz cuvette and the glass of the focusing lens of the macro adapter used. In comparison to the short side chains, the intensity of the peaks associated to the conjugated chromophores is dominant, due to their high

Raman activity. Consequently, no major spectral differences can be observed between the 3 formulations. Moreover, in contrast, molecules such as NaOH and HCl do not have any significant contribution to the spectra. In the solution of 2 g.L<sup>-1</sup> CERUBIDINE®, mannitol is found at 10 g.L<sup>-1</sup>. Despite the relatively high concentration, the Raman spectrum collected from the mannitol aqueous solution does not exhibit strong features, and the only difference with the water spectrum is at 886 cm<sup>-1</sup>, where a weak peak arises from the broad background (spectrum 2e).

*Table 1: Bands assignments of main features observed in the finger print region of TEVA®, MYLAN® and CERUBIDINE® Raman spectra collected from 2 g.L<sup>-1</sup> solutions in deionized water [34-38].*

TEVA® (cm <sup>-1</sup> )	MYLAN® (cm <sup>-1</sup> )	CERUBIDI NE® (cm <sup>-1</sup> )	Assignments
355	350	351	-
446	446	447	C-C-O vibrations , C=O in plane deformation
-	-	886	-
1000	1000	998	C-C stretching of alicyclic ring
1089	1091	1091	Ring breathing and aromatic C-H bending vibrations, methoxy C-O stretching and C-C aliphatic chain vibrations
1219	1219	1219	C-O-H vibrations and in plane bending of O-H
1250	1250	1251	C-O vibrations
1309	1309	1311	C-O-H vibrations
1445	1445	1453	Aromatic ring stretching modes
1582	1582	1586	Aromatic ring stretching
1647	1647	1650	Hydrogen bonded stretching vibrations of C=O

### 3.2.2 Discriminant analysis using PCA-FDA

Discriminant and quantitative analysis are two aspects of the data mining process not necessarily achieved through similar approaches, although they remain strongly connected. One could consider that any measurements of concentrations performed cannot be validated without identifying the molecule in solution being analysed. Therefore, it appears to be more relevant to first identify the drug and then aim for the quantification. While the quantitative analysis can establish a correlation between the intensity of the signal and the concentrations measured, when it comes to discriminating data sets based on the molecular information of the spectra, the overall intensity should be disregarded in favour of band positions, number and intensity ratios. Consequently, in the present study, it has been considered more relevant to subject the spectra to both baseline correction and vector normalisation prior to **PCA-FDA**. Thus, the concentration parameter has been removed from the analysis, focusing solely on the variations described above. Moreover, two methods for baseline correction have been evaluated, the Lieber correction [39] and EMSC [40].

PCA is a well-established unsupervised approach for rapid evaluation of spectral variability in complex data sets. Intensively used for analysis of IR and Raman spectra collected from biological samples, it readily allows visualisation of possible discrimination of samples using scatter plots [41, 42]. Figure 3A displays the scatter plot obtained for Raman spectra collected from 2 g.L<sup>-1</sup> TEVA®, MYLAN® and CERUBIDINE® subjected to Lieber baseline correction followed by vector normalisation. Although spectral differences are not conspicuous after preprocessing, the scatter plot presents three well separated groups, corresponding to each formulation. It can be seen that, according to Principal Component 1 (PC1), spectra collected from CERUBIDINE® are well separated from the others (Figure 3A red cluster), while TEVA® and MYLAN® data are only discriminated according to Principal Component 2 (PC2) (Figure 3A blue and green clusters).

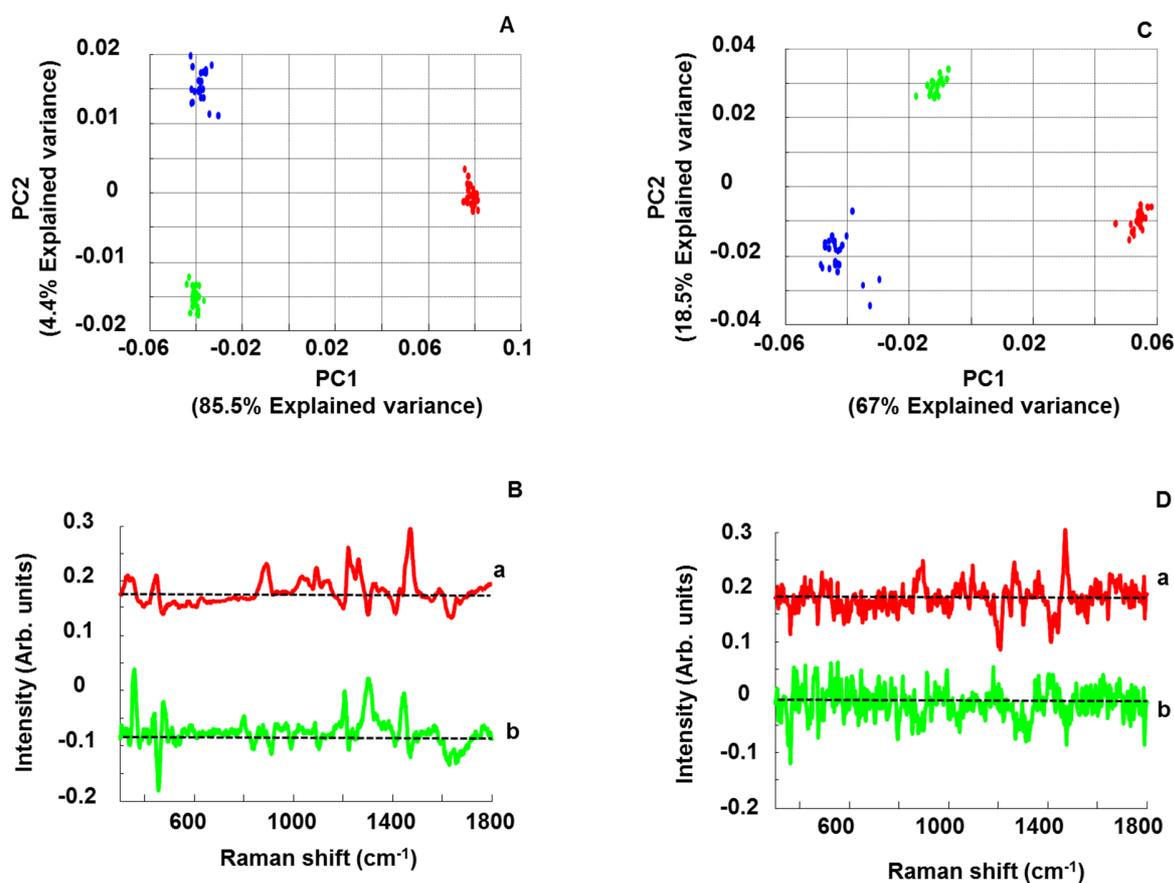


Figure 3: PCA of Raman data collected from TEVA® (blue), MYLAN® (green) and CERUBIDINE® (red) 2 g.L<sup>-1</sup> solutions in deionised water. A and B: scatter plot of PC1 and PC2 with corresponding loadings 1 and 2 obtained from data subjected to Lieber baseline correction and vector normalization; C and D: scatter plot of PC1 and PC2 with corresponding loadings 1 and 2 obtained from data subjected to EMSC correction and vector normalization. Loading are offset for clarity.

The loadings corresponding to each PC are displayed in Figure 3B, highlighting the spectral features leading to the distribution of the data points in the scatter plot. Dominant features in the loading of PC1 observed at 447 cm<sup>-1</sup>, 891 cm<sup>-1</sup>, 1096 cm<sup>-1</sup>, 1221 cm<sup>-1</sup>, 1263 cm<sup>-1</sup> and 1301 cm<sup>-1</sup>, 1411 cm<sup>-1</sup>, 1472 cm<sup>-1</sup> and 1641 cm<sup>-1</sup> correspond to dominant peaks previously identified in the

mean spectra (Table 1). Loading 2, only accounting for 4.4% of the explained variance discriminating TEVA® (*blue*) and MYLAN® (*green*), displays a lower number of strong features ( $357\text{ cm}^{-1}$ ,  $475\text{ cm}^{-1}$ ,  $1210\text{ cm}^{-1}$ ,  $1306\text{ cm}^{-1}$ , and  $1448\text{ cm}^{-1}$ ). These 2 forms are really similar with only NaOH and HCl as excipients which have no contribution to the spectral signatures recorded. The features of the loading of PC2 are common to the Raman spectra of both compounds, as shown in Figure 2, but the differentiating features suggest that small shifts can result from the different stereo-chemistry of the molecules.

Although PCA performed using a single concentration (i.e.  $2\text{ g.L}^{-1}$ ) exhibits clear separation of the three clusters and thus possible identification of the different chemotherapeutic forms, it is also important to address the need to perform discriminant analysis as an overall classifier, including all concentrations. The spectra in Figure 2 illustrate how similar the Raman spectra of the three formulations are, and the PCA in Figure 3 indicates that any variations between them are minor differences in peaks positions and intensities. Applying PCA to the range of possible combinations of concentrations would require considerable effort, and thus other multivariate approaches such as PCA-FDA need to be implemented. All the different concentrations and drug types have been considered as distinct groups to classify ( $n = 24$ ). The approach is supervised and implementation of a cross validation procedure is necessary to ensure the robustness of discrimination observed. Specifically, 1/3 of the data was used as a validation set and for each iteration of the 100-fold leave K-out cross validation, calibration and validation sets were constituted from the 2/3 of the remaining spectra. To better appreciate the specificity and sensitivity of the discrimination, at each iteration of the cross validation, results obtained with a model based on 4 principal components are reported in a confusion matrix (Table 3). This step of the analysis aims purely to discriminate the different anti-cancer drugs without any significance given to their concentrations. Therefore, successful classification was judged solely on the ability to identify the correct drug corresponding to the test spectra, irrespective of the concentration assignment. Although the overall specificity and sensitivity obtained are above 90%, the results are concentration and molecule dependent, and in some cases exhibit rather low values, as for example the TEVA® S7 (specificity = 85%), MYLAN® S1 (specificity = 75%) or CERUBIDINE® S1 (sensitivity 50%) (Table 3).

Although the results are encouraging, observation made in Figure 2 of minimal spectral differences between TEVA® and MYLAN® formulations, the **specificity** of the classification achieved is

somehow surprising. Raman spectroscopy is a technique particularly sensitive to changes of the light scattering when performing analysis on liquid samples. While the dynamic range of concentrations tested could lead to inconsistencies in the measurements, solutions of commercial formulations also contain excipients which possibly contribute to the spectra collected, not necessarily in terms of spectral features considering the concentrations, but rather affecting the baseline (or background) through scattering [43]. For this reason, more advanced baseline correction, taking into account the water spectrum has been performed by means of EMSC coupled to vector normalisation prior to PCA (Figure 3C and D). While the scatter plot from the PCA still displays clear separation between data groups according to formulation (Figure 3C), the corresponding loadings of PC1 and PC2 are strongly affected, and retain only residual features at 895 cm<sup>-1</sup>, 918 cm<sup>-1</sup>, 993 cm<sup>-1</sup>, 1096 cm<sup>-1</sup>, 1180 cm<sup>-1</sup>, 1265 cm<sup>-1</sup>, 1471 cm<sup>-1</sup>. Clearly, a standard baseline correction does not take into account the changes in band ratios induced by underlying background in the spectra possibly leading to misinterpretation of results. Following EMSC, both the residual water and background are removed and relative band intensities are more accurately corrected. Consequently, when performing the PCA-FDA, the outcome is also affected and the overall specificity and sensitivity are now found to respectively be between 64.4% - 72.5% and 53.3% - 59.2% with the classification model constructed using 3 principal components (Table 4). A number of concentrations exhibit values below 50%, which indicates that a non-negligible number of misclassifications have occurred, thus raising some serious concerns about the choice of data preprocessing having high impact on the analysis outcome and consequently the robustness of Raman spectroscopy measurements for drug identification can be questioned. In some cases, such as TEVA® S3 and CERUBIDINE® S4, there is obviously a reduced reliability in the drug identification, which for high concentrations can generate higher risks of dose intoxication of patients. Moreover, at lower concentrations, the accuracy is also not suitable for accurate classification, suggesting the specificity of Raman is not sufficient. EMSC is an advanced correction method able to remove instrumental and/or experimental interferences. However, despite applying such correction, the outcome of the PCA-FDA remains quite disappointing, suggesting that the sensitivity and specificity observed are directly linked to the ability of the

technique to discriminate the 3 therapeutic solutions rather than the underlying source of variability.

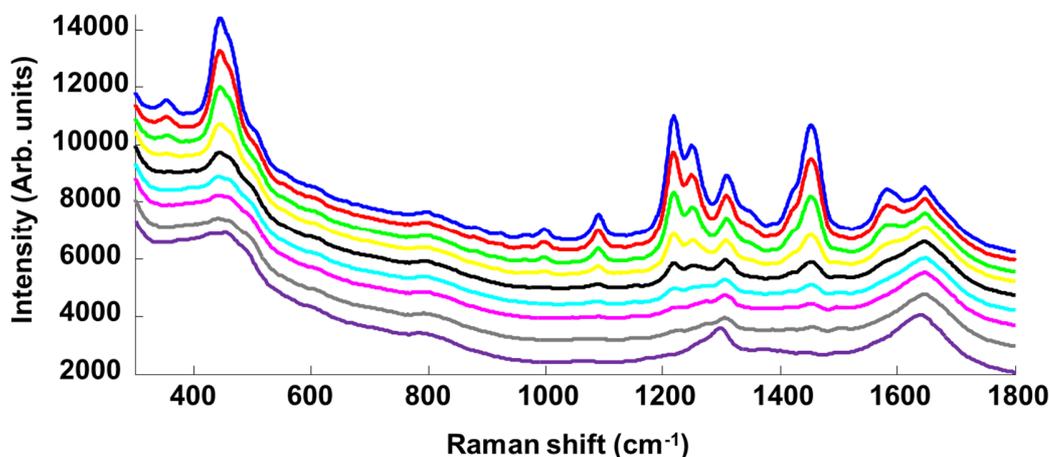
*Table 3: Specificity and sensitivity % of PCA\_FDA for Raman data processed without EMSC.*

Sample concentration (g.L <sup>-1</sup> )	TEVA®		MYLAN®		CERUBIDINE®	
	Specificity %	Sensitivity %	Specificity %	Sensitivity %	Specificity %	Sensitivity %
S1 (0.03125)	93.3	100	75	90	100	50
S2 (0.0625)	100	100	80	100	100	60
S3 (0.125)	88.8	100	90	80	100	80
S4 (0.25)	100	100	100	100	100	100
S5 (0.5)	100	100	100	100	100	100
S6 (1)	100	100	100	100	100	100
S7 (1.5)	85	100	100	70	100	100
S8 (2)	100	100	100	100	100	100
Overall	95.9	100	93.2	92.5	100	86.6

*Table 4: Specificity and sensitivity % of PCA\_FDA for Raman data with EMSC.*

Sample concentration (g.L <sup>-1</sup> )	TEVA®		MYLAN®		CERUBIDINE®	
	Specificity %	Sensitivity %	Specificity %	Sensitivity %	Specificity %	Sensitivity %
S1 (0.03125)	50	26.7	33.3	53.3	57.1	13.3
S2 (0.0625)	40.1	46.7	40	53.3	78.9	6.7
S3 (0.125)	93.8	0	31.8	53.3	36.4	46.7
S4 (0.25)	36.8	40	80	6.7	29.1	0.4
S5 (0.5)	58.8	13.3	58.8	13.3	17.4	53.3
S6 (1)	100	100	100	100	100	100
S7 (1.5)	100	100	100	93.3	96.7	100
S8 (2)	100	100	100	100	100	100
Overall	72.5	53.3	72.3	59.2	64.4	57.5

### 3.2.3 Quantitative analysis using PLSR



*Figure 4: Mean Raman spectra of the finger print regions collected from TEVA® following serial dilutions. Corresponding drug concentrations are 2g.L<sup>-1</sup> (blue), 1.5g.L<sup>-1</sup> (red), 1g.L<sup>-1</sup> (green), 0.5 g.L<sup>-1</sup> (yellow), 0.25g.L<sup>-1</sup> (black), 0.125 g.L<sup>-1</sup> (cyan), 0.0625 g.L<sup>-1</sup> (magenta), 0.03125 g.L<sup>-1</sup> (grey), water (purple). Spectra are offset for clarity.*

While examination of spectra allows identification of the anti-cancer drug with a reasonably high degree of accuracy, the linear relationship existing between the intensity of the peaks and the concentrations prepared can be examined and used as a quantitative prediction model using PLSR. Due to the experimental macro set up used in this study, the reproducibility achieved during the measurements enables unambiguous monitoring of the decrease of the peak intensities according to the dilutions performed, as illustrated for the example of TEVA® in Figure 4. PLSR is a reliable method to evaluate the quantitative information contained in the data sets collected. However, in order to ensure maximum statistical relevancy, cross validation procedures are generally implemented [30], in order to demonstrate that the analysis is not biased, but also to test the robustness of the predictive model used. Considering the size of the number of spectra available for each concentration, a 100-fold leave K out cross validation with 2/3 of spectra used as calibration and the remaining 1/3 as validation has been preferred for this study. The calibration and validation are therefore completely independent and selected randomly at each iteration of the cross validation, enabling testing of numerous data combinations. The first graph generated during

the analysis represents the Root Mean Square Error of Calibration (RMSEC) according to the number of dimensions used to regress the data (Figure 5A). As a supervised method, and aiming to deliver the lowest RMSEC and thus the best precision, a gradual decrease towards 0 is expected as the number of dimensions employed in the model increases. Once the model has been established using the calibration set, it is then evaluated with the validation spectra. The Figure 5B exhibits first a steep decrease in the **RMSECV** (validation) before slightly increasing a stabilizing around 0.015 g.L<sup>-1</sup>. All dimensions calculated from the calibration set are not relevant to the quantification, and according to Figure 5B, above 3 dimensions over fitting of the data could actually affect the precision of the predictive model. Consequently, the regression plot has been implemented using three dimensions, and is presented in Figure 5C. The correlation between observed concentration (true concentrations) and predicted concentrations (experimental concentrations) delivers a linear trend characterised by  $R^2$  value of 0.9997 (Table 5). The error bars illustrate the standard deviation calculated from the 100 fold cross validation and further support the high reproducibility of the measurements. The overall **RMSECV** reached was found to be 0.0127 g.L<sup>-1</sup> for TEVA® (Table 5). The numbers of dimensions used to construct the PLSR predictive models, respectively 3, 5 and 5 for TEVA®, MYLAN® and DNR, are also indicated in **Table 5**. Figure 5D shows the weighting vector corresponding to the PLSR analysis performed on TEVA® solutions. It represents the wavenumbers used to construct the predictive models, confirming the molecular specificity of the analysis has been preserved. The features with major contribution in the PLSR model can be found at 354 cm<sup>-1</sup>, 445 cm<sup>-1</sup>, 1000 cm<sup>-1</sup>, 1090 cm<sup>-1</sup>, 1219 cm<sup>-1</sup>, 1249 cm<sup>-1</sup>, 1309 cm<sup>-1</sup>, 1453 cm<sup>-1</sup>, 1581 cm<sup>-1</sup>, 1647 cm<sup>-1</sup> which are the major spectral features previously observed in **Figure 3** and **Table 1**. Finally, a similar approach has been employed for MYLAN® and CERUBIDINE® formulations, yielding respective **RMSECV** of 0.0173 g.L<sup>-1</sup> and 0.0220g.L<sup>-1</sup>, as summarised in **Table 5**. Although inconsistent between patients and hospital, the lowest administered concentration for doxorubicin is roughly 0.15 g.L<sup>-1</sup>, and therefore the corresponding **RMSECV** found of 0.0127 g.L<sup>-1</sup> suggest overall imprecision of about 8.5%. Similarly for MYLAN® and CERUBIDINE®, the **RMSECV** respectively account for 5.8% and 2.35% compared to the lowest doses typically prescribed (0.3 g.L<sup>-1</sup> and 1.2 g.L<sup>-1</sup>). **The lower limit of detection (LOD) calculated from the PLSR analysis indicates that Raman spectroscopy can pick up concentrations as low as 0.0290 g.L<sup>-1</sup>, 0.0383 g.L<sup>-1</sup> and 0.0778 g.L<sup>-1</sup> respectively for TEVA®,**

MYLAN® and CERUBIDINE®, values which are all more than 10 times below the lowest dose administered to patients.

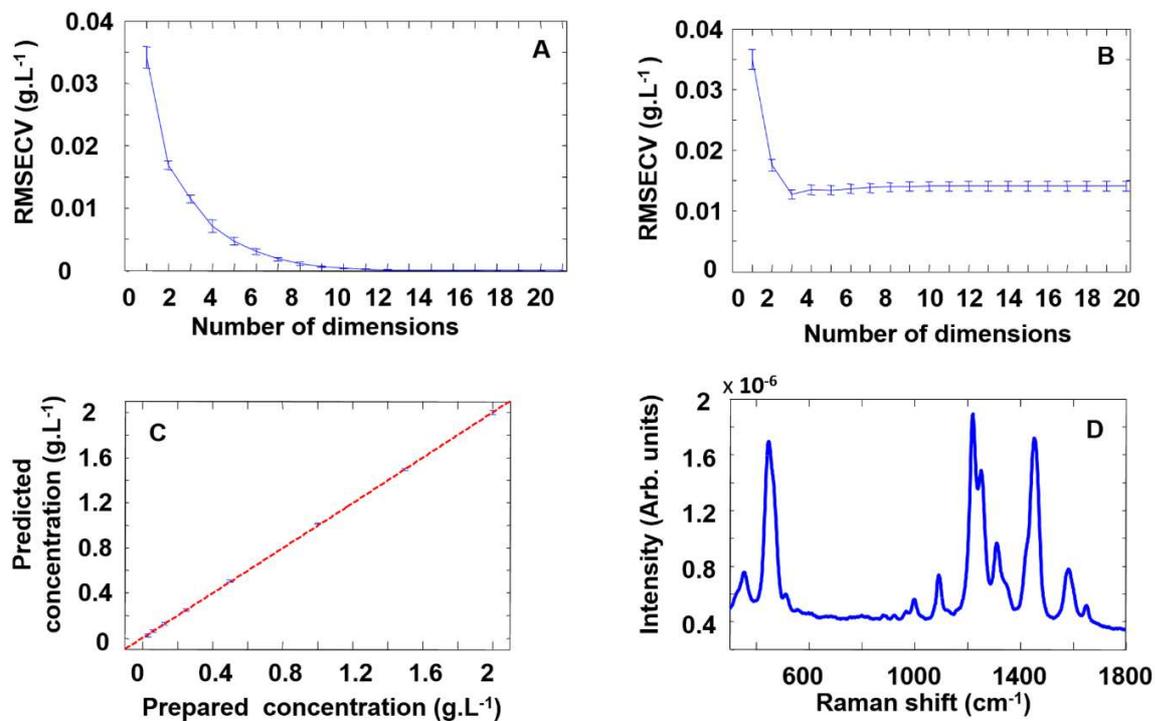


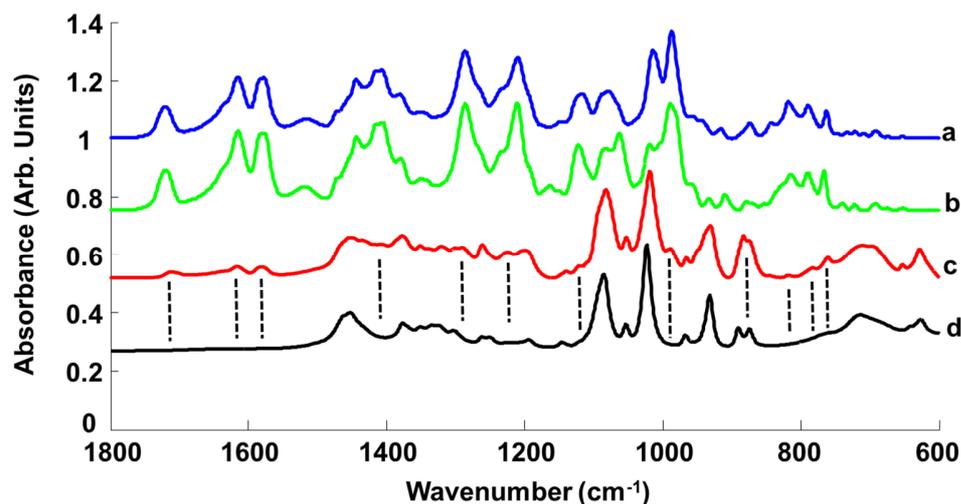
Figure 5: PLSR analysis performed on TEVA®. A: Standard deviation error of calibration; B: Standard deviation error of validation; C: Regression fitting and D: Weighting vector. Error bars represent the standard deviation

Table 5: PLSR results obtained from the 3 chemotherapeutic drug solutions tested

Commercialized formulation	Active drug	$R^2$	RMSECV (g.L <sup>-1</sup> )	Dimensions
TEVA®	DOX	0.9996	0.0127	3
MYLAN®	EPI	0.9993	0.0173	5
CERUBIDINE®	DAU	0.9990	0.0220	5

### 3.3. Discriminant and quantitative analysis using ATR-IR spectroscopy

#### 3.3.1. ATR-IR characterisation of anthracycline formulations



*Figure 6: Mean ATR-IR spectra recorded from TEVA® (A), MYLAN® (B), CERUBIDINE® (C) formulations and Mannitol (D). All spectra were recorded from solutions with final concentration of anticancer drugs at 2g.L<sup>-1</sup>. Spectra are offset for clarity.*

Although recent work has supported the feasibility to perform IR quantitative analysis with ATR-IR from liquid samples [32, 44], in the present study the anti-cancer drug concentrations involved are not sufficient to overcome the strong contribution of the water bands, which completely swamp the features of interest (data not shown). Consequently, air drying of samples following deposition onto the ATR crystal is required to collect exploitable data. Figure 6 displays the fingerprint region of mean spectra collected from the 3 different formulations. TEVA® and MYLAN® (Figure 6 A and B) deliver spectral signatures with high degree of similarity, but comparison of the peaks positions would suggest some variations in the range 700-1200 cm<sup>-1</sup> can be observed (Table 6). Interestingly, the major difference compared to Raman spectroscopy is the weaker contribution of conjugated chromophores in the signatures collected. Consequently, other contributions from side

chains and excipients can be somehow enhanced and more easily detected. For instance, the CERUBIDINE® mean spectrum differs from the other two without any possible ambiguity. This can be explained by the presence of a high amount of mannitol in the formulation, obviously delivering strong features in the IR signature (Figure 6d). Summarised in Table 6, the most intense mannitol features at 1454 cm<sup>-1</sup>, 1085 cm<sup>-1</sup>, 1023 cm<sup>-1</sup>, 875 cm<sup>-1</sup>, 932 cm<sup>-1</sup> and 890 cm<sup>-1</sup> clearly dominate in the spectrum. However, a comparison between the IR spectrum collected from CERUBIDINE® (Figure 6c) and the pure spectrum of mannitol highlights that, despite the lower band intensities, specific contributions from daunorubicin can be identified at 1712 cm<sup>-1</sup>, 1617 cm<sup>-1</sup>, 1582 cm<sup>-1</sup>, 1411 cm<sup>-1</sup>, 1202 cm<sup>-1</sup>, 1110 cm<sup>-1</sup>, 996 cm<sup>-1</sup> and 984 cm<sup>-1</sup>.

Table 6: Bands positions and assignments in ATR\_IR spectra collected from TEVA®, MYLAN® and CERUBIDINE®.

TEVA® (cm <sup>-1</sup> )	MYLAN® (cm <sup>-1</sup> )	CERUBIDINE® (cm <sup>-1</sup> )	Mannitol (cm <sup>-1</sup> )	Assignments
-	-	-	712	-
762	767	-	-	=C-H (aromatic) bending vibrations
789	791	783	-	
817	814	-	875	-N-H out of plane bending vibrations, =C-H (aromatic) bending vibrations
884	882	883	890	=C-H (aromatic) bending vibrations
-	-	931	932	-
987	989	984	-	C=O, C-O-H bending vibrations
-	-	996	-	
1014	1019	1019	1023	
1079	1063	1082	1085	C-O (alcoholic) bending vibrations
1117	1122	1110	-	Ring breathing coupled with C-C=O bending vibrations, C-O (alcoholic) bending vibrations
1210	1211	1202	-	C-O (alcoholic) bending vibrations
1286	1286	-	-	H-bonding vibrations of O-H...O coupled with ring vibrations
1407	1407	1411	-	
-	-	-	1454	-CH <sub>3</sub> bending vibrations
1578	1578	1582	-	-N-H bending vibrations, Ring vibrations
1616	1616	1617	-	
				-N-H bending vibrations
1709	1709	1712	-	C=O stretching vibrations
1722	1722	-		

### 3.3.2 Discriminant analysis using PCA-FDA

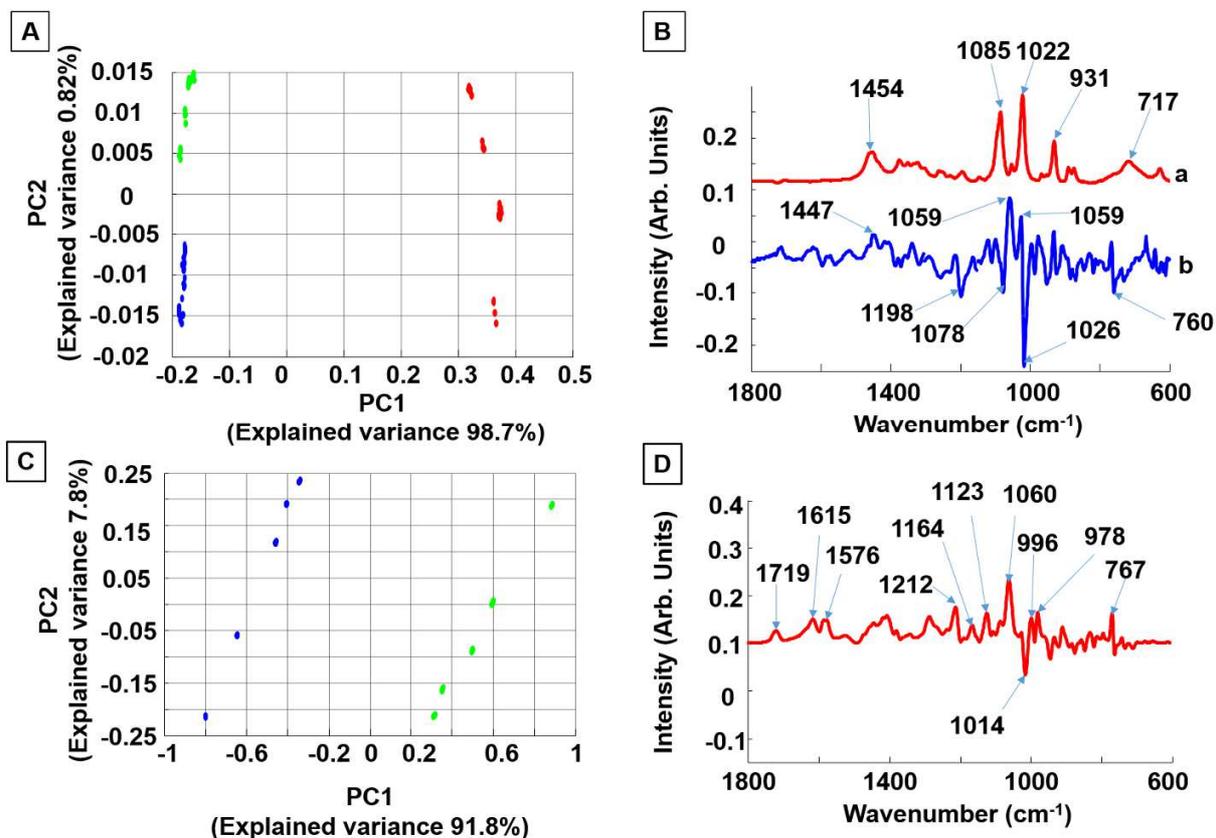
Similar to Raman spectra, observations based on PCA enable a rapid visualisation of the spectral variability contained in the data set collected by ATR-IR. While Raman spectra from liquid samples can to some extent be affected by underlying background, the collection of infrared spectra using the ATR set up coupled to air dried drops drastically reduces any distortion of the

baseline. Although recording of IR spectra from biological material such as cell or tissues in transmission can lead to Mie or resonant Mie scattering effects [45], direct deposition of the solution on the ATR crystal minimises such phenomena. Consequently, only some offset in the data collected is observed according to the concentrations analysed, but such effects are easily compensated by implementation of a simple straight baseline subtraction such as the rubber band employed here. A rubber band correction with only 2 nodes at  $1800\text{ cm}^{-1}$  and  $600\text{ cm}^{-1}$  has the effect of bringing the spectral baseline down to 0 across the spectral range. Prior to performing the discriminant analysis, the concentration parameters have been removed using a vector normalisation to rescale all spectra.

The PCA scatter plot displays strong separation of the three formulations according to PC1 and PC2, accounting respectively for 98.7% and 0.82% of the explained variance (Figure 7A). Similarly to Raman spectra, PC1 unambiguously differentiates TEVA® (Figure 7A blue) and MYLAN® (Figure 7A green) formulations from CERUBIDINE® (Figure 7A red). The corresponding loading suggests that PC1 is basically the spectral signature of mannitol, which is expected due to the strong features observed in the spectra of CERUBIDINE®. Notably, the bands at  $1454\text{ cm}^{-1}$ ,  $1085\text{ cm}^{-1}$ ,  $1022\text{ cm}^{-1}$ ,  $931\text{ cm}^{-1}$  and  $717\text{ cm}^{-1}$  support this statement. PC2 discriminates TEVA® (Figure 7A blue) from MYLAN® (Figure 7A green). The PCA tends to maximise the explained variance expressed by the first principal components. As a consequence of the strong contribution from the features assigned to mannitol, PC2 only accounts for 0.82%. This observation supports the presence of small variations in band positions, as previously witnessed and reported in Table 6. The features of loading 2 confirm that the discrimination is based on the differing spectroscopic profiles of the compounds but, due to the scatter of CERUBIDINE® data points over the 2 other clusters in that dimension, the exact determination of the origin of the features of the loading of PC2 is not trivial. Pairwise PCA is the most direct and unambiguous analysis to highlight the presence of discriminative wavenumbers between 2 datasets [28]. Figure 7C and 7D display respectively the scatter plot and loading 1 for the PCA applied to TEVA® and MYLAN®. In the absence of the CERUBIDINE® and associated strong mannitol bands, the scatter plot displays a clear separation of TEVA® and MYLAN® along PC1 with an explained variance of 91.8%. The loading 1 corresponding to PC1 exhibits a quite large number of features but interestingly some of the most intense, located at  $767\text{ cm}^{-1}$ ,  $1014\text{ cm}^{-1}$ ,  $1060$

$\text{cm}^{-1}$ ,  $1123 \text{ cm}^{-1}$ ,  $1212 \text{ cm}^{-1}$ ,  $1576 \text{ cm}^{-1}$ ,  $1615 \text{ cm}^{-1}$  and  $1719 \text{ cm}^{-1}$  either match or are closely related to features found in TEVA® and MYLAN® as listed in Table 6. Although pairwise PCA is a relevant approach to investigate inter and intra data sets variability, it becomes relatively difficult to evaluate all combinations of data and estimate the discrimination rate when the number of samples is increasing. Therefore, more advanced methods such as PCA-FDA can be used.

The outcome of the PCA-FDA obtained using 3 principal components is presented in Table 7, whereby, once again, the specificity and sensitivity has been calculated for each concentration analysed and for each drug. Using the ATR-IR spectra for discriminative analysis leads to high rates of correct classification with overall values equal or above 99.4%. A more detailed breakdown of the results shows that only the lowest concentrations tested, corresponding to  $0.03125 \text{ g.L}^{-1}$ , do not deliver 100% specificity and sensitivity for all 3 formulations. However, with 90% as the lowest outcome, the results remain quite satisfactory for accurate and reliable identification of formulations tested in respective formulations.



*Figure 7: PCA of ATR\_FTIR data collected from chemotherapeutic solutions at 2 g.L<sup>-1</sup>. A: Scatter plot of PC1 and PC 2 with TEVA® (blue), MYLAN® (green) and CERUBIDINE® (red). B: Loading corresponding to PC1 (a) and PC2 (b). C: Pairwise PCA of ATR\_FTIR from TEVA® (blue) and MYLAN® (green) and D: Loading corresponding to loading PC1*

Table 7: Specificity and sensitivity % of PCA\_FDA for ATR\_FTIR.

Sample concentration (g.L <sup>-1</sup> )	TEVA®		MYLAN®		CERUBIDINE®	
	Specificity %	Sensitivity %	Specificity %	Sensitivity %	Specificity %	Sensitivity %
S1 (0.03125)	95	100	100	90	100	100
S2 (0.0625)	100	100	100	100	100	100
S3 (0.125)	100	100	100	100	100	100
S4 (0.25)	100	100	100	100	100	100
S5 (0.5)	100	100	100	100	100	100
S6 (1)	100	100	100	100	100	100
S7 (1.5)	100	100	100	100	100	100
S8 (2)	100	100	100	100	100	100
Overall	99.4	100	100	98.75	100	100

The main concern of the study is to investigate the discrimination of clinical formulations of the anticancer drugs, rather than their simple solutions. Considering the strong sugar features in the IR spectra collected from CERUBIDINE®, it is obvious that mannitol plays a major role in the discrimination. One could question the specificity of differentiating the therapeutic solution containing CERUBIDINE® from pure solution of mannitol. This can be demonstrated thanks to a pair wise PCA performed from CERUBIDINE® solution at 0.03125 g.L<sup>-1</sup>, the lowest concentration presently tested, and mannitol solution at 0.15625g.L<sup>-1</sup>. The concentration of mannitol has been adjusted accordingly to the dilution factor of the stock solution of CERUBIDINE®. The scatter plot displayed in figure 8A exhibits a clear separation of the data along PC1 with 85.4% of the explained variance. Naturally, in a pair wise PCA, PC2, with 6.5% represents the intragroup variability. Loading 1 compared to the pure spectrum of mannitol highlights that numerous spectral features are contributed by other molecules such as daunorubicin in the solutions. Therefore, the discrimination of CERUBIDINE® observed in the PCA-FDA is a combined contribution of both mannitol and other ingredients including the anticancer drug. Similar observations are made with all concentrations included in this study (data not shown) which is perfectly understandable as the mannitol/daunorubicin ratio remains identical for all samples due to the same dilution factor applied during the preparation.

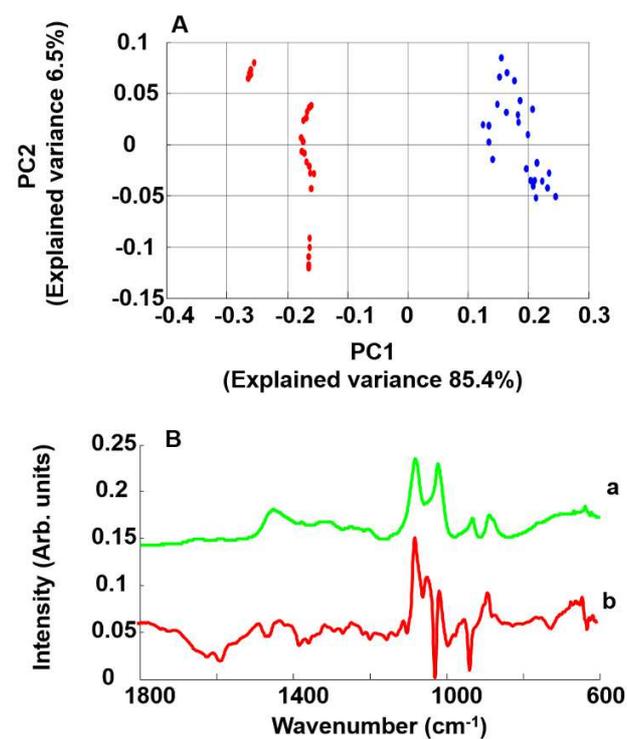
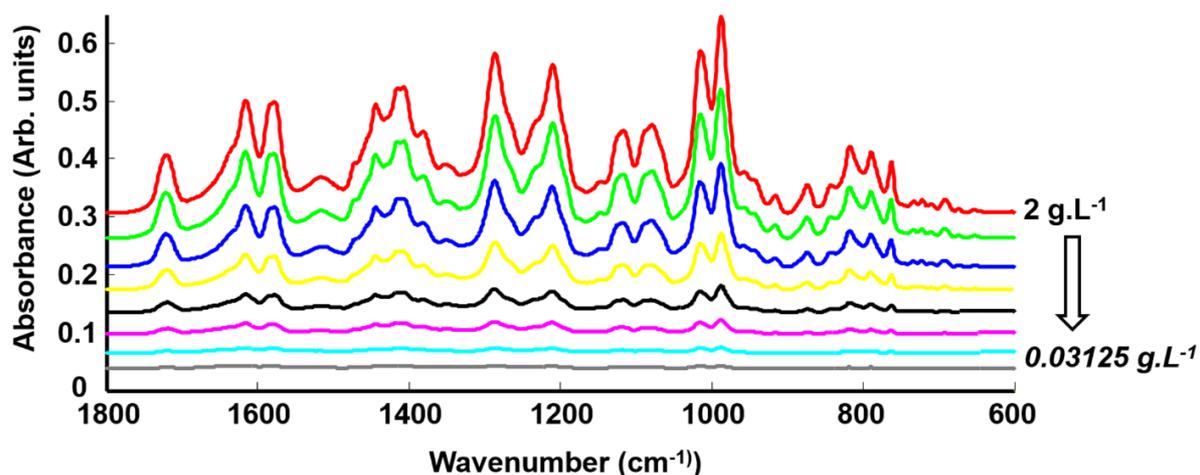


Figure 8: PCA of ATR\_FTIR data collected from CERUBIDINE<sup>®</sup> 0.03125 g.L<sup>-1</sup> and Mannitol 0.15625g.L<sup>-1</sup>. A: Scatter plot of PC1 and PC 2 with CERUBIDINE<sup>®</sup> (red) and Mannitol (blue). B: Reference spectrum of mannitol (a) and loading corresponding to PC1 (b). Spectra offset for clarity.

### 3.3.3 Quantitative analysis using PLSR



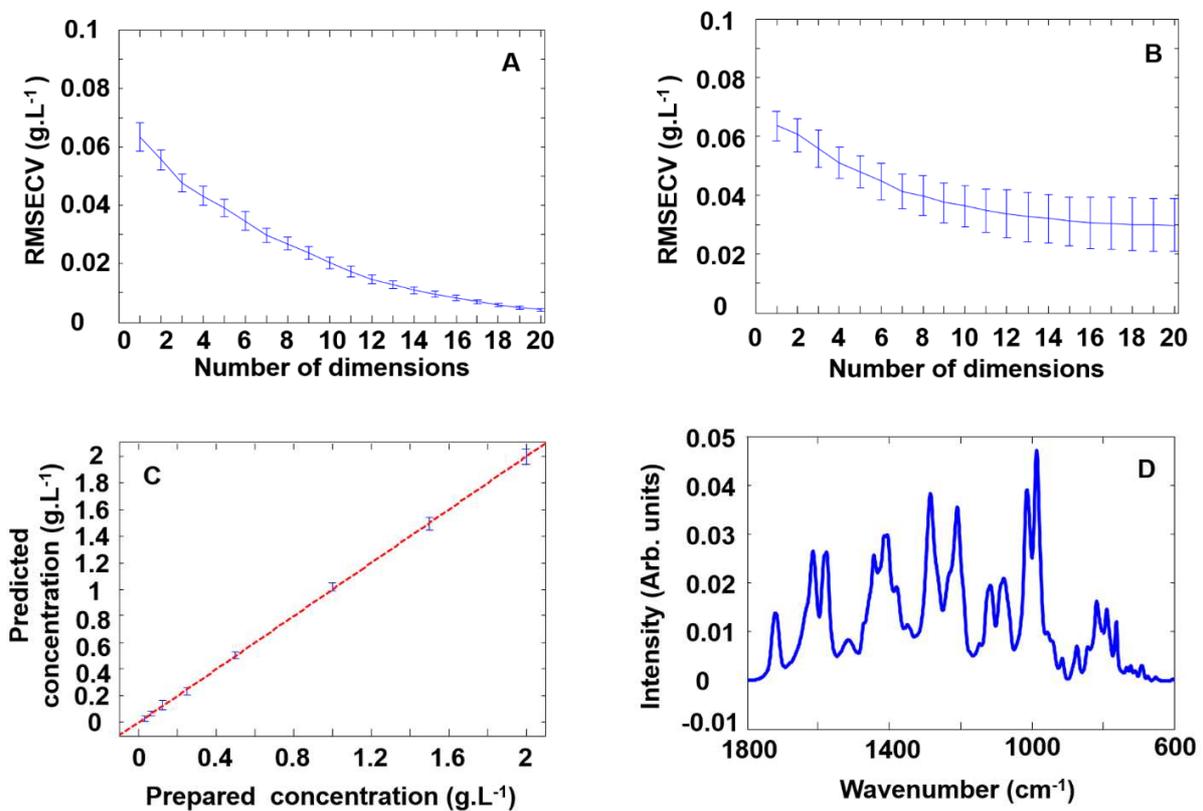
*Figure 9: Mean Raman spectra of fingerprint regions of TEVA® following serial dilutions. Spectra are organised as decreasing concentrations from the top to bottom, respectively corresponding to doxorubicin concentrations of  $2\text{g.L}^{-1}$  (red),  $1.5\text{g.L}^{-1}$  (green),  $1\text{g.L}^{-1}$  (blue),  $0.5\text{g.L}^{-1}$  (yellow),  $0.25\text{g.L}^{-1}$  (black),  $0.125\text{g.L}^{-1}$  (magenta),  $0.0625\text{g.L}^{-1}$  (cyan), and  $0.03125\text{g.L}^{-1}$  (grey). Spectra are offset for clarity*

Observation of the ATR-IR spectra collected from a set of TEVA® solutions highlights that the intensity of the absorbance tends to decrease according to concentration (Figure 9). It has been demonstrated that, while ATR-IR measurements can be used for quantitative analysis, the approach has a limited range of concentrations over which the Beer- Lambert law can be directly applied, whereby high concentration samples lead to a plateau effect with a loss of the linear relationship [44]. As the concentration increases, the thickness of the deposited residue film increases, to the extent that is comparable to the spatial extent of the evanescent field of the ATR. This can be accounted for by additional dilution of the higher concentrations. However, in the present study, it appears the dynamic range of concentrations is within the range of linearity of absorbance response. Similar to the case for Raman, the ATR-IR spectra have been subjected to PLSR analysis in order to evaluate how well correlated the spectral variations are with the concentration of the anticancer formulations. Figure 10 presents the different steps performed on the TEVA® form, as described previously in section 3.2.3. As observed with the Raman, the

weighting vector extracted from the PLSR analysis is almost an identical copy of the mean spectra presented in **Figure 10D**, once more illustrating the molecular specificity of the analysis performed. However, the behaviour of the RMSECV as function of the dimensions number strongly differs in IR from that in Raman analysis. It tends to gradually decrease, with no clearly defined optimal number of dimensions which would ensure no overfitting of the model. Examination of the latent variables suggests that only the first 3 ones exhibit strong relevant features, while subsequent ones become increasingly dominated by noise and random bands. Therefore, considering the limited volume of data in this investigation, it was deemed preferable to limit the number of dimensions to 3, in order to not deliver over optimistic results for the ATR-IR analysis. Ultimately, the outcome of the PLSR analysis delivers RMSECV higher than those observed with Raman spectroscopy, respectively 0.0558 g.L<sup>-1</sup>, 0.0573g.L<sup>-1</sup> and 0.0759 g.L<sup>-1</sup> for TEVA®, MYLAN® and CERUBIDINE® formulations (**Table 8**). Nevertheless, with  $R^2$  values all above 0.99, the high degree of linearity in the PLSR model constructed with the IR data is demonstrated.

However, although a multi-reflection crystal has been used, compared to the macro set up for the Raman experiments, the repeatability in the IR analysis remains lower, leading to loss of precision. In order to avoid over fitting of the results, the number of dimensions has been limited to 3 for the 3 drugs. **Ultimately the results demonstrate the limited quantitative capabilities of the approach when applied to anticancer formulations. For instance, the LOD calculated from the PLSR analysis are considerably higher than for Raman analysis. LOD of 0.5802g.L<sup>-1</sup> (TEVA®), 0.4043 g.L<sup>-1</sup> (MYLAN®) and 0.3514 g.L<sup>-1</sup> (CERUBIDINE®) are calculated which are in same order of the lowest administered doses (respectively 0.15 g.L<sup>-1</sup>, 0.3 g.L<sup>-1</sup>, 1.2 g.L<sup>-1</sup> . It appears uncertainties in determination of low concentrations raises questions regarding the reliability of ATR-IR for accurate quantification.** The Beer Lambert law is commonly employed to describe the linear relationship between concentration of an analyte in solution and its absorbance, facilitating quantitative analysis. When performing analysis of liquid samples, it is easy to comply with the basic requirements of uniform optical path length of the samples and sample homogeneity, to produce reliable results. As most of other protocols of sample preparation for IR analysis, air drying the samples onto the ATR crystal also raises questions about the uniformity and repeatability of the sample layer thickness and area [45]. In recent years, concerns over the so-called “coffee ring effect” have been raised, in relation to ATR FTIR analysis of dried droplets.

The effect refers to the tendency of molecular species to accumulate at the edge of the drop during the drying process [43-44]. Depending on the sample concentration, the coffee ring effect can also lead to complete loss of the signal simply because the sample accumulates outside the area of measurement of the ATR crystal [41]. Different types of crystals are commercially available, defined by the material of the crystal (for example diamond or germanium) but also the number of internal reflections. Since a single reflection accessory delivered no signal from the dried droplets, due to the coffee ring effect (data not shown), we used in the present study the multi-reflection diamond crystal enabling the 2 $\mu$ L to be completely comprised within the recording area. Nevertheless, the heterogeneity of the dried hand-made deposits remained significant.



*Figure 10: PLSR analysis performed on TEVA®. A: Standard deviation error of calibration; B: Standard deviation error of validation; C: Regression fitting and D: Weighting vector. Error bars represent the standard deviation*

Table 8: PLSR results obtained from the 3 chemotherapeutic solutions tested

Commercialized formulation	Active drug	$R^2$	RMSECV (g.L <sup>-1</sup> )	Dimensions
TEVA®	DOX	0.9935	0.0580	3
MYLAN®	EPI	0.9936	0.0573	3
CERUBIDINE®	DAU	0.9882	0.0759	3

## Overall discussion

Direct comparison between infrared and Raman spectroscopy has been conducted for the identification and quantification of 3 anthracyclines (TEVA®, MYLAN® and CERUBIDINE®) in solution, at clinically relevant concentrations (i.e. in the range of concentrations administered to patients). Data mining of spectral data sets with the help of multivariate analysis demonstrates that discrimination of drugs and predicting concentrations can be achieved with both techniques, although with different reliabilities. On the one hand, Raman spectroscopy displays the best precision for quantitative analysis, with **RMSECV** between 0.127-0.220 g.L<sup>-1</sup> obtained from PLSR analysis, while on the other hand IR spectroscopy delivers higher sensitivity and specificity (above 95%) for drug identification with PCA-FDA. Both techniques behave differently due to differences in the intrinsic nature of the signal collected. Raman spectroscopy is strongly dependent on polarisability of the molecules which, in the case of TEVA®, MYLAN® and CERUBIDINE®, results in spectra clearly dominated by contributions from the conjugated ring structures. In contrast, IR is sensitive to small variations in side-chains but also to additives. [Commercial chemotherapeutic solutions are formulae or mixtures with different degrees of complexity, depending on the number of ingredients present. The drug is usually the most concentrated chemical constituent of the solution, by many orders of magnitude, therefore with little or no contribution from the other constituents. However, in some cases the presence of a specific](#)

molecule at higher concentrations can change the balance and generate additional features in the spectral signatures. In CERUBIDINE®, for example, the presence of mannitol contributes significantly to the profile of the infrared spectra, which are therefore representative of mixed contributions of the drug and the sugar used as additive. This does not represent a limitation of the technique, but rather a strength of ATR-IR compared to Raman, as manifested in the significantly improved discriminative power of the technique and its higher sensitivity in classification. Some ingredient or additive signatures, once combined with the spectral bands of the drug, can play the role of spectral markers for more reliable identification of the stock solution used to prepare the diluted solution. In the present study, TEVA®, MYLAN® and CERUBIDINE® are the ones and only brands used for respectively doxorubicin, epirubicin and daunorubicin. Thus the spectral bands identified and used for discrimination will be reproducible and consistent over time. At present, quality control laboratories in hospitals typically employ flow injection analysis which is able to deliver accurate quantification for the majority of the 100+ solutions tested daily. The technique lacks the required specificity, however, when challenged with the discrimination of isomers such as TEVA®, MYLAN® and CERUBIDINE®, which display similar absorbance in the visible range, which can result in significant delays in administration of the solutions on a daily basis. As an augmentation to the current workflow, ATR-IR provides a potential solution for the rapid differentiation of the solutions. Performing the ATR analysis requires small volumes (few  $\mu\text{L}$ ) compared to other current systems with injection of aliquots of at least 1.5 mL necessary. Considering the whole process of data analysis can be automated and completed in less than 1 minute, the overall time to get the result does not exceed 5 mins, which is comparable to systems already in use. Ultimately, ATR-IR spectroscopy is a cost effective technique with compact bench top apparatus that does not represent a huge capital investment and are increasingly user friendly, such that the data preprocessing and analysis routines can be easily incorporated into the instrument software and implemented in an automated protocol.

#### **4. Conclusion**

Both Raman and Infrared spectroscopy are valuable analytical tools with demonstrated potential for clinical applications. Although often proposed as complementary, instrumental set up would

have a strong influence of their performance to a given application. The quantitative analysis of chemotherapeutic drugs can be achieved in solution with Raman spectroscopy or following air drying with ATR-IR. In contrast, the discrimination of anthracyclines such as TEVA® (doxorubicin), MYLAN® (epirubicin) and CERUBIDINE® (daunorubicin) as part of quality control of therapeutic solutions suggests specificity of Raman analysis is more limited than that of IR. Although air drying is required for IR, analysis results are promising and enable both identification and quantification of all 3 molecules tested. With specificity and sensitivity above 95%, ATR-IR spectrophotometry appears to be the most valuable approach to address current needs for discrimination of **therapeutic solutions containing isomeric drugs** and should be positioned as a highly relevant technique. Having an ATR just beside on the bench as a complementary tool to current techniques could provide an unambiguous identification in 5 mins. To summarise, ATR-IR spectrophotometry remains a cost effective, rapid and user friendly method which with optimization of methods would lead to fast implantation in the hospital routine.

## Bibliography

- [1] N.D. Popowicz, S.J. O'Halloran, D. Fitzgerald, Y.C.G. Lee, D.A. Joyce, A rapid, LC-MS/MS assay for quantification of piperacillin and tazobactam in human plasma and pleural fluid; application to a clinical pharmacokinetic study, *Journal of Chromatography B*, 1081-1082 (2018) 58-66.
- [2] A.J. McShane, D.R. Bunch, S. Wang, Therapeutic drug monitoring of immunosuppressants by liquid chromatography–mass spectrometry, *Clinica Chimica Acta*, 454 (2016) 1-5.
- [3] A. Delmas, J.B. Gordien, J.M. Bernadou, M. Roudaut, A. Gresser, L. Malki, M.C. Saux, D. Breilh, Quantitative and qualitative control of cytotoxic preparations by HPLC-UV in a centralized parenteral preparations unit, *J Pharm Biomed Anal*, 49 (2009) 1213-1220.
- [4] J.D. Hofer, B.A. Olsen, E.C. Rickard, Is HPLC assay for drug substance a useful quality control attribute?, *J Pharm Biomed Anal*, 44 (2007) 906-913.
- [5] C. Bazin, B. Cassard, E. Caudron, P. Prognon, L. Havard, Comparative analysis of methods for real-time analytical control of chemotherapies preparations, *International Journal of Pharmaceutics*, 494 (2015) 329-336.
- [6] C. Bazin, V. Vieillard, A. Astier, M. Paul, Implementation of real-time identification analysis and quantification of chemotherapies preparations with a Multispec® analyser, *Annales Pharmaceutiques Françaises*, 72 (2014) 33-40.
- [7] F. Nardella, M. Beck, P. Collart-Dutilleul, G. Becker, C. Boulanger, L. Perello, A. Gairard-Dory, B. Gourieux, G. Ubeaud-Sequier, A UV-Raman spectrometry method for quality control of anticancer preparations: Results after 18 months of implementation in hospital pharmacy, *Int J Pharm*, 499 (2016) 343-350.

- [8] P. Bourget, A. Amin, F. Vidal, C. Merlette, F. Lagarce, Comparison of Raman spectroscopy vs. high performance liquid chromatography for quality control of complex therapeutic objects: Model of elastomeric portable pumps filled with a fluorouracil solution, *Journal of Pharmaceutical and Biomedical Analysis*, 91 (2014) 176-184.
- [9] P. Bourget, A. Amin, F. Vidal, C. Merlette, P. Troude, A. Baillet-Guffroy, The contribution of Raman spectroscopy to the analytical quality control of cytotoxic drugs in a hospital environment: Eliminating the exposure risks for staff members and their work environment, *International Journal of Pharmaceutics*, 470 (2014) 70-76.
- [10] K. Kong, C. Kendall, N. Stone, I. Notingher, Raman spectroscopy for medical diagnostics--From in-vitro biofluid assays to in-vivo cancer detection, *Adv Drug Deliv Rev*, 89 (2015) 121-134.
- [11] M.J. Baker, J. Trevisan, P. Bassan, R. Bhargava, H.J. Butler, K.M. Dorling, P.R. Fielden, S.W. Fogarty, N.J. Fullwood, K.A. Heys, C. Hughes, P. Lasch, P.L. Martin-Hirsch, B. Obinaju, G.D. Sockalingum, J. Sulé-Suso, R.J. Strong, M.J. Walsh, B.R. Wood, P. Gardner, F.L. Martin, Using Fourier transform IR spectroscopy to analyze biological materials, *Nature protocols*, 9 (2014) 1771-1791.
- [12] B. Bird, M.S. Miljkovic, S. Remiszewski, A. Akalin, M. Kon, M. Diem, Infrared spectral histopathology (SHP): a novel diagnostic tool for the accurate classification of lung cancer, *Lab Invest*, 92 (2012) 1358-1373.
- [13] I.R. Ramos, A.D. Meade, O. Ibrahim, H.J. Byrne, M. McMennamin, M. McKenna, A. Malkin, F.M. Lyng, Raman spectroscopy for cytopathology of exfoliated cervical cells, *Faraday Discuss*, 187 (2016) 187-198.
- [14] V. Vuiblet, M. Fere, E. Bankole, A. Wynckel, C. Gobinet, P. Birembaut, O. Piot, P. Rieu, Raman-based detection of hydroxyethyl starch in kidney allograft biopsies as a potential marker of allograft quality in kidney transplant recipients, *Sci Rep*, 6 (2016) 33045.
- [15] I.P. Santos, E.M. Barroso, T.C. Bakker Schut, P.J. Caspers, C.G.F. van Lanschot, D.H. Choi, M.F. van der Kamp, R.W.H. Smits, R. van Doorn, R.M. Verdijk, V. Noordhoek Hegt, J.H. von der Thusen, C.H.M. van Deurzen, L.B. Koppert, G. van Leenders, P.C. Ewing-Graham, H.C. van Doorn, C.M.F. Dirven, M.B. Busstra, J. Hardillo, A. Sewnaik, I. Ten Hove, H. Mast, D.A. Monserez, C. Meeuwis, T. Nijsten, E.B. Wolvius, R.J. Baatenburg de Jong, G.J. Puppels, S. Koljenovic, Raman spectroscopy for cancer detection and cancer surgery guidance: translation to the clinics, *Analyst*, 142 (2017) 3025-3047.
- [16] A. Jaworska, S. Fornasaro, V. Sergio, A. Bonifacio, Potential of Surface Enhanced Raman Spectroscopy (SERS) in Therapeutic Drug Monitoring (TDM). A Critical Review, *Biosensors (Basel)*, 6 (2016).
- [17] S. Wartewig, R.H. Neubert, Pharmaceutical applications of Mid-IR and Raman spectroscopy, *Adv Drug Deliv Rev*, 57 (2005) 1144-1170.
- [18] F. Bonnier, H. Blasco, C. Wasselet, G. Brachet, R. Respaud, L.F. Carvalho, D. Bertrand, M.J. Baker, H.J. Byrne, I. Chourpa, Ultra-filtration of human serum for improved quantitative analysis of low molecular weight biomarkers using ATR-IR spectroscopy, *Analyst*, 142 (2017) 1285-1298.
- [19] D. Rohleder, W. Kiefer, W. Petrich, Quantitative analysis of serum and serum ultrafiltrate by means of Raman spectroscopy, *Analyst*, 129 (2004) 906-911.
- [20] Z. Farhane, F. Bonnier, H.J. Byrne, Monitoring doxorubicin cellular uptake and trafficking using in vitro Raman microspectroscopy: short and long time exposure effects on lung cancer cell lines, *Anal Bioanal Chem*, 409 (2017) 1333-1346.
- [21] D. Šahnić, E. Meštrović, T. Jednačak, I. Habinovec, J. Parlov Vuković, P. Novak, Monitoring and Quantification of Omeprazole Synthesis Reaction by In-Line Raman Spectroscopy and Characterization of the Reaction Components, *Organic Process Research & Development*, 20 (2016) 2092-2099.
- [22] L.M.M. Lê, A. Tfayli, J. Zhou, P. Prognon, A. Baillet-Guffroy, E. Caudron, Discrimination and quantification of two isomeric antineoplastic drugs by rapid and non-invasive analytical control using a handheld Raman spectrometer, *Talanta*, 161 (2016) 320-324.

- [23] A. Amin, P. Bourget, F. Vidal, F. Ader, Routine application of Raman spectroscopy in the quality control of hospital compounded ganciclovir, *International Journal of Pharmaceutics*, 474 (2014) 193-201.
- [24] M. Kiehntopf, B. Monch, R. Salzer, M. Kupfer, M. Hartmann, Quality control of cytotoxic drug preparations by means of Raman spectroscopy, *Pharmazie*, 67 (2012) 95-96.
- [25] P. Bourget, A. Amin, A. Moriceau, B. Cassard, F. Vidal, R. Clement, La Spectroscopie Raman (SR) : un nouvel outil adapté au contrôle de qualité analytique des préparations injectables en milieu de soins. Comparaison de la SR aux techniques CLHP et UV/visible-IRTF appliquée à la classe des anthracyclines en oncologie, *Pathologie Biologie*, 60 (2012) 369-379.
- [26] T. Lan, Y. Fang, W. Xiong, C. Kong, Automatic baseline correction of infrared spectra, *Chin. Opt. Lett.*, 5 (2007) 613-616.
- [27] N.K. Afseth, A. Kohler, Extended multiplicative signal correction in vibrational spectroscopy, a tutorial, *Chemometrics and Intelligent Laboratory Systems*, 117 (2012) 92-99.
- [28] F. Bonnier, H.J. Byrne, Understanding the molecular information contained in principal component analysis of vibrational spectra of biological systems, *Analyst*, 137 (2012) 322-332.
- [29] D. Bertrand, P. Courcoux, J.C. Autran, R. Meritan, P. Robert, Stepwise canonical discriminant analysis of continuous digitalized signals: Application to chromatograms of wheat proteins, *Journal of Chemometrics*, 4 (1990) 413-427.
- [30] S. Guo, T. Bocklitz, U. Neugebauer, J. Popp, Common mistakes in cross-validating classification models, *Analytical Methods*, 9 (2017) 4410-4417.
- [31] M.E. Keating, H. Nawaz, F. Bonnier, H.J. Byrne, Multivariate statistical methodologies applied in biomedical Raman spectroscopy: assessing the validity of partial least squares regression using simulated model datasets, *Analyst*, 140 (2015) 2482-2492.
- [32] L. Miloudi, F. Bonnier, D. Bertrand, H.J. Byrne, X. Perse, I. Chourpa, E. Munier, Quantitative analysis of curcumin-loaded alginate nanocarriers in hydrogels using Raman and attenuated total reflection infrared spectroscopy, *Anal Bioanal Chem*, 409 (2017) 4593-4605.
- [33] F. Allegrini, A.C. Olivieri, IUPAC-Consistent Approach to the Limit of Detection in Partial Least-Squares Calibration, *Analytical Chemistry*, 86 (2014) 7858-7866.
- [34] S. Charak, D.K. Jangir, G. Tyagi, R. Mehrotra, Interaction studies of Epirubicin with DNA using spectroscopic techniques, *Journal of Molecular Structure*, 1000 (2011) 150-154.
- [35] Z. Farhane, F. Bonnier, H.J. Byrne, Monitoring doxorubicin cellular uptake and trafficking using in vitro Raman microspectroscopy: short and long time exposure effects on lung cancer cell lines, *Analytical and Bioanalytical Chemistry*, 409 (2017) 1333-1346.
- [36] L.M.M. Lê, M. Berge, A. Tfayli, J. Zhou, P. Prognon, A. Baillet-Guffroy, E. Caudron, Rapid discrimination and quantification analysis of five antineoplastic drugs in aqueous solutions using Raman spectroscopy, *European Journal of Pharmaceutical Sciences*, 111 (2018) 158-166.
- [37] D. Nieciecka, A. Królikowska, I. Setiawan, P. Kryszewski, G.J. Blanchard, Interactions of Doxorubicin with Organized Interfacial Assemblies. 2. Spectroscopic Characterization, *Langmuir*, 29 (2013) 14570-14579.
- [38] E. Szafraniec, K. Majzner, Z. Farhane, H.J. Byrne, M. Lukawska, I. Oszczapowicz, S. Chlopicki, M. Baranska, Spectroscopic studies of anthracyclines: Structural characterization and in vitro tracking, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 169 (2016) 152-160.
- [39] N.L. Marc, G.R. Alan, Comparison of Derivative Preprocessing and Automated Polynomial Baseline Correction Method for Classification and Quantification of Narcotics in Solid Mixtures, *Applied Spectroscopy*, 60 (2006) 182-193.
- [40] K.H. Liland, A. Kohler, N.K. Afseth, Model-based pre-processing in Raman spectroscopy of biological samples, *Journal of Raman Spectroscopy*, 47 (2016) 643-650.

- [41] Z. Farhane, F. Bonnier, A. Casey, A. Maguire, L. O'Neill, H.J. Byrne, Cellular discrimination using in vitro Raman micro spectroscopy: the role of the nucleolus, *Analyst*, 140 (2015) 5908-5919.
- [42] G.L. Owens, K. Gajjar, J. Trevisan, S.W. Fogarty, S.E. Taylor, B.D. Gama-Rose, P.L. Martin-Hirsch, F.L. Martin, Vibrational biospectroscopy coupled with multivariate analysis extracts potentially diagnostic features in blood plasma/serum of ovarian cancer patients, *Journal of Biophotonics*, 7 (2014) 200-209.
- [43] F. Bonnier, A. Mehmood, P. Knief, A.D. Meade, W. Hornebeck, H. Lambkin, K. Flynn, V. McDonagh, C. Healy, T.C. Lee, F.M. Lyng, H.J. Byrne, In vitro analysis of immersed human tissues by Raman microspectroscopy, *Journal of Raman Spectroscopy*, 42 (2011) 888-896.
- [44] F. Bonnier, G. Brachet, R. Duong, T. Sojinrin, R. Respaud, N. Aubrey, M.J. Baker, H.J. Byrne, I. Chourpa, Screening the low molecular weight fraction of human serum using ATR-IR spectroscopy, *J Biophotonics*, 9 (2016) 1085-1097.
- [45] P. Bassan, H.J. Byrne, F. Bonnier, J. Lee, P. Dumas, P. Gardner, Resonant Mie scattering in infrared spectroscopy of biological materials - understanding the 'dispersion artefact', *Analyst*, 134 (2009) 1586-1593.

