2018

Enabling Quantification of Protein Concentration in Human Serum Biopsies Using Attenuated Total Reflectance – Fourier Transform Infrared (ATR-FTIR) Spectroscopy

Kate Spalding
University of Strathclyde, WestCHEM, Glasgow, United Kingdom

Franck Bonnier
Technological University Dublin, Franck.Bonnier@tudublin.ie

Clément Bruno
Universite Francois-Rabelais Tours, Tours, France

See next page for additional authors

Recommended Citation

Spalding, K., Bonnier, F. & Bruno, C. (2018). Enabling quantification of protein concentration in human serum biopsies using attenuated total reflectance _ Fourier transform infrared (ATR-FTIR) spectroscopy. *Vibrational Spectroscopy, vol. 99. pg. 50-58. doi:10.1016/j.vibspec.2018.08.019*

Enabling Quantification of Protein Concentration in Human Serum Biopsies using Attenuated Total Reflectance – Fourier Transform Infrared (ATR-FTIR) Spectroscopy

Katie Spalding [1], Franck Bonnier [2], Clément Bruno [2], Hélène Blasco [3], Ruth Board [4], Isabelle Benz-de Bretagne [3], Hugh J. Byrne [5], Holly J. Butler [1], Igor Chourpa [2], Pretheepan Radhakrishnan [1,6], and Matthew J. Baker [1*]

[1]*WestCHEM, Department of Pure & Applied Chemistry, Technology and Innovation Centre, University of Strathclyde, Glasgow, G1 1RD, UK*

[2] *Université François-Rabelais de Tours, Faculté de Pharmacie, EA 6295 Nanomédicaments et Nanosondes, 31 avenue Monge, 37200 Tours, France*

[3]*Laboratoire de biochimie et biologie moléculaire, CHRU Tours, France, Inserm U930, Université François Rabelais, Tours, France*

[4]*Rosemere Cancer Centre, Lancashire Teaching Hospitals NHS Trust, Royal Preston Hospital, Sharoe Green Lane, Preston, PR2 9HT, UK*

[5] *FOCAS Research Institute, Dublin Institute of Technology (DIT), Kevin Street, Dublin 8, Ireland*

[6] *CDT in Medical Devices & Health Technologies, Department of Biomedical Engineering, Technology and Innovation Centre, University of Strathclyde, Glasgow, G1 1RD, UK*

**e-mail: matthew.baker@strath.ac.uk @ChemistryBaker*

ABSTRACT

Changes in protein concentrations within human blood are used as an indicator for nutritional state, hydration and underlying illnesses. They are often measured at regular clinical appointments and the current analytical process can result in long waiting times for results and the need for return patient visits. Attenuated total reflectance – Fourier transform infrared (ATR-FTIR) spectroscopy has the ability to detect minor molecular differences, qualitiatively and quantitatively, in biofluid samples, without extensive sample preparation. ATR-FTIR can return an analytical measurement almost instantaneously and therefore could be deemed as an ideal technique for monitoring molecular alterations in blood within the clinic.

To determine the suitability of using ATR-FTIR spectroscopy to enable protein quantification in a clinical setting, pooled human serum samples spiked with varying concentrations of human serum albumin (HSA) and immunoglobulin G (IgG) were analysed, before analysing patient clinical samples. Using a validated partial least squares method, the spiked samples (IgG) produced a linearity as high as 0.998 and a RMSEV of 0.49 ± 0.05 mg mL^{-1}, with the patient samples producing R^2 values of 0.992 and a corresponding RMSEV of 0.66 ± 0.05 mg mL^{-1}. This claim was validated using two blind testing models, leave one patient out cross validation and k-fold cross validation, achieving optimum linearity and RMSEV values of 0.934 and 1.99 ± 0.79 mg mL^{-1}, respectively.

This demonstrates that ATR-FTIR is able to quantify protein within clinically relevant complex matirces and concentrations, such as serum samples, rapidly and with simple sample preparation. The ability to provide a quantification step, along with rapid disease classification, from a spectroscopic signature will aid clinical translation of vibrational spectroscopy to assist with problems currently faced with patient diagnostic pathways.

Keywords *– Protein, Infrared, Attenuated Total Reflection, Serum, Clinical, Quantification*

1. Introduction

The analysis of biofluids such as serum using vibrational spectroscopy is considered a potential solution to current problems with early and accurate diagnosis of many diseases [1] and promises improved patient mortality, morbidity and quality of life [2]. Biofluids are routinely obtained following a minimally invasive procedure, providing a large sample volume that contains biomolecular

components such as proteins, amino-acids, lipids and carbohydrates in relative concentrations which are highly dependent on demographical characteristics and physiological or pathological status [3]. Clinicians establish a diagnosis from several criteria, including; medical history, clinical symptoms, imaging data and biological exploration. Numerous diseases are characterised by a qualitative or quantitative modification of a specific biological parameter, while others are associated with a biological signature,changes in multiple biological paramters [4],[5].

Proteomics, peptidomics and metabolomics are often studied through nuclear magnetic resonance [6], mass spectrometry [7] or capillary electrophoresis [8]. A large number of proof-of-principle studies have identified diagnostic markers for cancers [9],[10],[11],[12]. However, there is extensive sample preparation associated with these techniques. ATR-FTIR can provide a spectral profile of all the macromolecular classes contained within serum and a signature, as opposed to single markers, could be advantageous when analysing a heterogenous disease such as cancer. Vibrational spectroscopic investigations have resulted in a large numer of proof of principle studies that show promising results [13].

The diagnosis of gliomas (high-grade and low-grade) from non-cancer through a combination of ATR-FTIR and multivariate support vector machine analysis (SVM), was achieved with average sensitivities and specificities of 93.75 and 96.53 % respectively for human serum samples [14]. In 2016, a large serum study using FTIR spectroscopy was completed, reporting the discrimination of cancer vs non-cancer patients with a sensitivity of 91.5 % and specificity of 83.0 %, as well as deciphering cancer severity and the primary site of metastasis [15]. These classification values were then improved to 92.8 and 91.5 %, sensitivity and specificity, by executing Random forest and 2D correlation analysis in combination [16]. The application of vibrational spectroscopy to analyse tissue sections, as well as single cells [17], [18] has also been hugely successful. The advatanges of vibrational spectroscopy, such as ATR-FITR, and high classification values demonstrates a potential use as the gold standard for patient disease screening using serum [19], [20], [21].

To facilitate the translation of an infrared spectroscopy based diagnostic test, the incorporation of a quantification step could be regarded as beneficial and complementary to current clinical practice as the majority of clinical tests are currently based upon quantitative values as opposed to signatures or fingerprints. Protein vibrations are often the most prominant in a biological infrared spectrum [22]. Furthermore, protein concentrations are systematically measured in routine practice; they are useful to interpret biological parameters, discuss nutritional status, extracellular hydration status or to help in the diagnosis of some diseases. Specific proteins such as human serum albumin (HSA) and immunoglobulinG (IgG), (as well as the ratio of the two), may be altered in the case of inflammation, infection, unexplained weight loss, fatigue or act as symptoms of kidney or liver disease [23],[24]. HSA constitutes between 57 – 71 % of the serum composition, and globulins 8 – 26 % [25]. HSA and IgG could be regarded as ideal to produce models in order to demonstrate an ATR-FTIR spectroscopic test capable of quantifying proteins.

Infrared spectroscopy enables the production of a unique spectrum representative of the fundamental molecular vibrations that occur within the sample, that provides a ‘fingerprint’ of the sample [26], [27]. The combination of the rapid collection method obtained through the FTIR systems and spectroscopic method development has accelerated biomedical research using infrared spectroscopy. In particular, ATR-FTIR spectroscopy has been shown to be suitable for biological materials, due to the minimal sample preparation and the ability to analyse a variety of samples types, including serum [1], [28], [29], [30] . An advantageous property of IR based techniques, is that they obey the principles of the Beer Lambert law [31], allowing quantification of a given molecule relative to the absorbance of light in the sample it is travelling through. This enables ATR-FTIR spectroscopy to

quantify specific biomolecule concentrations, as the proportion of light absorbed by the sample will correlate with the concentration of molecules within a sample.

This is evident from the wide variety of research carried out, quantifying particular biomarkers from biofluid samples [32], [33], [34]. For example, the analysis of dried serum deposits using transmission spectroscopy highlighted the ability to quantify eight serum analytes [35] and the simultaneous quantification of glucose and urea analytes in addition to malaria parasitemia from a single drop of blood dried on a glass slide [36]. The latter highlights the capability of using ATR-FTIR spectroscopy to determine disease and metabolic state, through the identification and quantification of chemical parameters associated with the disease diagnosis. Furthermore, the concentration of *in situ* DNA within cells [37], as well as the metabolite concentrations in urine [38] and saliva [39], could be determined using ATR-FTIR and bovine IgG was quantified using transmission and ATR-FTIR spectroscopy [40]. The quantification of glycine, a low molecular weight fraction (LMWF), provided evidence that ATR-FTIR spectroscopy can monitor systemic spectral modifications created by spiking human serum with lyophilised glycine [41]. Additionally, the removal of high molecular weight fractions (HMWF), through centrifugal filtration, led to an increased precision and accuracy of the quantitative models based on the partial least squares algorithm [42]. Research carried out by Perez-Guaita in 2012 [43], showed the possibility of determining total albumin, total globulin and immunoglobulin concentrations through the analysis of 50 µl liquid serum samples deposited on an ATR crystal cell. This work highlighted the potential for ATR-FTIR to act as a green alternative to current methods used within hospitals, through the removal of reagents and implementation of relatively cheap and simple instrumentation. However no sample preparation study was performed in order to establish the optimum sample preparation with minute volumes of serum.

Infrared spectral datasets are information rich, highlighting underlying biological and structural differences. Coupled with powerful multivariate analysis approaches, they have the ability to differentiate between disease classes by extracting relevant information. Multiple data mining approaches have been used in spectral data analysis, such as Principal Component Analysis (PCA), Random Forest (RF) and Support Vector Machine (SVM), all demonstrating the ability to discriminate diseased from non-diseased biofluid samples [44]. Currently, Partial Least Squares Regression analysis (PLSR) is one of the most frequently used techniques for the production of quantitative models, due to its ability to identify systematic variations of contributing factors and generate quantitative predictive models. This allows the prediction of unknowns, using the latent variables extracted from the regression model [40], [32], [45], [46].

ATR-FTIR spectroscopy has the ability to detect minor differences in biofluid samples, with minimal sample preparation, and multiple proof-of-principle studies have highlighted the potential clinical use for such a technique. However, translation of ATR-FTIR spectroscopy has not occurred due to multiple factors, including the lack of acceptance from clinical environments.

We show, for the first time, an optimised methodology to enable protein quantification in single and complex mixtures using a PLSR approach, detailing the in-depth progression of determining protein concentration from spiked samples, to patient samples, before blind testing methods. The incorporation of this new quantification step within biofluid diagnostic methodologies would enable a direct comparison to gold standard diagnostic methods and highlight the clinical excellence of vibrational spectroscopic analysis of biofluids and facilitate translation.

2. Materials and Methods

2.1. Sample Preparation Methodology

For the first time, an in depth methodological investigation was performed in order to establish the optimum sample preparation protocol for quantification from serum based ATR-FTIR spectroscopy. This study was performed using 2 models samples sets 1) Whole Serum Dilution Study and 2) Spiked Human Serum Models, before moving onto patient samples. Table 1 and subsections, 2.1.1 - 2.1.3, below provide further information on experimental details.

2.1.1. *Whole Serum Dilution Study*

To determine the ability of ATR-FTIR spectroscopy to detect variable protein concentrations, 1000 µl of commercially available, whole, sterile, filtered, mixed pool human serum (TCS Biosciences, UK) were used to create a set of seven 2- fold dilutions using deionised water (Milli-Q water (Millipore Elix S)

2.1.2.*Spiked Human Serum Models*

HSA and IgG (Sigma-Aldrich, UK) were used to create two spiked pooled serum models. Due to their abundant nature within human blood, current clinical use and detection over a range of concentrations from patient samples, these two proteins were deemed ideal for comparative measurements. Preparation of the two separate spiked models were created by the addition of each protein independently, while maintaining the concentration of the other. The initial concentrations of HSA and IgG of 46.3 and 13.53 mg/ml, respectively, were also taken into consideration. Further details of this can be found in Table 1 and Supporting Information (SI Materials and Methods, *Spiked Human Serum Models and Table S*$^{-1}$).

2.2. *Patient Sample Protein Levels*

Serum samples collected at the Biochemical laboratory at the University Hospital CHU Bretonneau de Tours, for the measurement of total protein, HSA and IgG were used for research, obeying the ethical procedures implemented by the hospital. The concentrations of total protein, HSA and IgG were measured using a Cobas 6000 analyser series (Roche Diagnostics) with a measurement precision of 1g/L – (Table S-2). The samples were obtained following an in-house standard operating procedure, developed by the hospital for the routine analysis of serum samples. Whole blood was collected using a dry tube with separate gel and coagulation activator. After at least one hour of clotting, blood was centrifuged for 10 minutes at 3000 g to isolate the serum from the other blood components. Sera were then analysed by immunoturbidimetry (IgG and HSA) and a colorimetric assay based on copper reaction for total proteins. The remainder of the blood serum was stored at -20°C until ATR-FTIR experiments were carried out.

2.3. *Data Collection Using ATR – FTIR Spectrometer*

ATR-FTIR spectra were recorded using a diamond crystal and a single reflection golden gate accessory (Specac, UK) attached to a Bruker Vector 22 (Bruker, Germany). 32 co-added scans, covering a wavenumber range of 4000 – 400 cm^{-1}, were combined to produce the spectrum, using a spectral resolution of 4 cm^{-1}. A background spectrum (32 co-added scans), using the same spectral range, of the ambient conditions was automatically subtracted by the OPUS package (Bruker,Germany) to create the sample spectrum.

The sample preparation approaches used are liquid, air dried and liquid samples which have been diluted by 10 %using deionised water, and then air dried (10 % air dried). Spectra from liquid serum samples were obtained in triplicate, immediately after the drop was deposited on to the crystal. Dry serum samples (air dried and 10% air dried) were also obtained in triplicate, following a drying time of between 5 – 8 minutes. The triplicate analysis of each serum drop accounted for any instrumental variance This process was repeated five times to encompass any biological variance between the

samples. The serum was dropped at a perpendicular angle using a micropipette to ensure a high level of reproducibility. This led to the production of a spiked data set containing 105 spectra and a patient data set containing 300 spectra, for analysis. Information on how the samples were confirmed as dry and the triplicate analysis is included in Supporting Information (Materials and Methods, *Data Collection Using ATR – FTIR Spectrometer*).

2.4. *Data Pre-Processing and Analysis*

Matlab (Mathworks, USA) was used to carry out all pre-processing and data analysis. A rubber-band baseline correction and vector normalisation using University of Strathclyde, in house written software was applied to the fingerprint region (1800 - 900 cm^{-1}) (Figure 1). Pre-processing allowed the systematic increase of the two protein concentrations to be observed and assessed, by removing any non-biochemical components of the spectra and enabling clearer analysis of spectral variations in the amide region.

PLSR was used to quantify the prepared protein concentrations from the spiked samples as well as estimate the serum protein levels in patient samples. The algorithm is a supervised method, whereby the concentrations are provided to the model prior to running the analysis. The PLSR models discussed have been built from the pre-processed data sets. The analysis gives an estimated value for the model accuracy [47] and is termed the Root Mean Square Error (RMSE), as well as an R^2 value indicating the linearity between the experimental and the predicted concentrations.

2.5. Whole Serum Dilution Study PLSR Optimisation

During the PLSR optimisation step, mean spectra were identified following pre-processing to calculate the area under the curve (AUC) and determine whether ATR-FTIR can detect and quantify protein concentrations. This was carried out on pure air dried and pure liquid, 2-fold dilution, samples to determine if a dilution factor was required.

2.6. *Spiked and Patient Model Validation*

To validate the robustness of the PLSR predictive models, the optimum number of cross validation loops was determined, by re-sampling 512 cross validation iterations, 1000 times. This produced three convergence plots, for the outputs of the RMSE calibration (RMSEC), RMSE validation (RMSEV) and R^2 values (Figure 2). From this plot, the optimum number of cross validation loops could be determined to produce minimal variation in the output from the PLSR analysis. This was done by curve fitting a one term power series model to the data and calculating when the gradient was <0.0001. The highest number of iterations was selected from the three convergence plots, with the result that the number of cross validation loops was not the same for each protein, but was consistently selected as the largest. For each iteration, the calibration set was compiled from 50 % of the data, selected randomly, leaving the remaining 50 % to be used for the validation set for the quantitative predictions. The mean and standard deviation of the RMSE and R^2 were calculated from each iteration. This methodology was carried out prior to all PLSR analysis in order to validate any results obtained, as PLSR is a supervised method and may be prone to overfitting the data. Once the ability of ATR-FTIR spectroscopy to determine the protein concentration of spiked as well as patient samples was identified, the patient set was used to blindly test the models.

2.7. Blind Testing Model Validation

The patient sample dataset was then used to create both the calibration and validation sets. To begin with, a leave one patient out cross validation (LOPOCV) method was employed, whereby 19 patient samples were used as the training set and the remaining 1 was used to test the model. A similar

methodology was then repeated, whereby 15 patients were selected to act as the training set, leaving the remaining 5 to act as the test set and be blindly predicted, in a process termed K-fold cross validation. Both approaches were optimised ensuring the maximum number of combinations were carried out as cross validation iterations. As such, this led to the former approach being repeated to cover all 20 possible combinations of selecting one patient out of 20, and the latter approach carried out over all the 15,504 possible combinations of selecting five from 20 patients. As the IgG concentration from patient 18 was not available (Table S-2), the model validation for the IgG concentrations was based on 19 patients as opposed to 20. Thus, the training sets contained 18 and 14 patients, for the LOPOCV and the K-fold methodologies respectively, reducing the number of possible combinations to 19 and 11,628, respectively.

3. **Results and discussion**

3.1. Quantification of protein concentrations in spiked human serum

3.1.1. Determining dilution factor

The analysis of biofluids, such as serum, using ATR-FTIR, produces high quality spectra with clearly defined spectral features [48]. In an ATR-FTIR spectrum, spectral peaks can be assigned to particular biomolecules, in order to allow the function, structure and biochemical signature of the sample to be identified [28]. Due to the strong water absorbance of IR light, air dried samples are generally preferred over liquid samples, although the biomolecular composition of the serum is unchanged [49]. The spectra of air dried pooled human serum exhibits the expected spectral features and assignments associated with human serum (Figure 3). These can be briefly described as; 3280 cm^{-1} (H-O-H stretching), 2957 cm^{-1} (asymmetric CH_3 stretching), 2920 cm^{-1} (asymmetric CH_2 stretching), 2872 cm^{-1} (symmetric CH_3 stretching), 1650 cm^{-1} (amide I of proteins), 1536 cm^{-1} (amide II of proteins), 1453 cm^{-1} (CH_2 scissoring), 1394 cm^{-1} (C=O stretch of COO-), 1242 cm^{-1}- (asymmetric PO_2 stretch), 1171 cm^{-1} (ester C-O symmetric stretch) and 1080 cm^{-1} (C-O stretch) [50]. The spectra are strongly dominated by the abundant proteins contained in the serum, which are present in high concentration compared to the other low molecular weight (LMW) components. In fact, the amide I peak at 1650 cm^{-1} has the highest intensity within the entire spectrum.

Figure 3 also shows the spectra of serum solutions which have been serially diluted before drying. The spectral peak centroids identified and assigned above remain unchanged by the dilution process, and the impact of the dilution procedure can be monitored by plotting the integrated area under the curve of the fingerprint region, as shown in Figure 3 (right). The curve shows an approximate linear dependence of integrated absorbance as a function of concentration in the low concentration region, but the behaviour rapidly deviates from linearity above 30% dilution. Notably, after 50 % dilution of the stock solution, the integrated absorbance decreases by only less than 5 %. The nonlinearity and saturation of the absorbance of dried deposits measured by ATR-FTIR, as a function of solution concentration has previously been discussed by Bonnier et al. [41]. Importantly, for the methodology employed in the current study, in order to produce the models spiked with protein, further protein will need to be added to the pooled serum to incorporate a wide concentration range. The minimal change in absorption above a 50 % dilution factor, shown in Figure 3 suggests the identification of an upper detection limit for the volume deposited. Therefore, for the analysis of the air dried serum, a dilution factor of 10 %, is required to ensure different protein concentrations are observable. This could also could have been combatted experimentally by depositing smaller volumes. However, to satisfy the requirement to cover the entire internal reflection element (IRE) and reproducible pipetting, diluting the larger volume was determined to be the optimum experimental approach.

In contrast, following the analysis of the liquid samples, biomolecular spectral assignments are difficult, due to the dominant water contribution from the broad O-H stretching band around

3300 cm^{-1} and bending vibration around 1680 cm^{-1}, (Figure 4). The AUC plot, shown in Figure 4 (right), indicates that serum dilution has a minimal effect on the integrated absorbance over the concentration range, although a linear decrease is observed below the 6.25 % dilution, with an R^2 value of 0.9979, which may be associated with a disaggregation phenomenon [51], therefore no dilution is necessary.

3.1.2.Construction of the quantitative model: partial least square regression (PLSR

Prior to the analysis of patient samples, a quantitative model using the PLSR algorithm was produced to evaluate the ability of ATR-FTIR to quantify protein concentration within a complex medium, such as human serum. This was applied to the protein spiked human serum models that reflect the clinically relevant protein concentrations, which tend to lie outside the normal ranges of 34 – 54 mg/ml for HSA and 8.1 – 23 mg/ml for IgG, within human blood in order to optimise the protocol.

Figure 5 shows the mean (n = 9) ATR-FTIR spectral fingerprint region for spectra obtained from IgG spiked, 10% diluted air dried serum samples. The data shows an increasing absorbance trend moving from the stock solution (red) to the highest concentration of protein (black), highlighting the systematic increase in the protein amide bands at 1640 and 1560 cm^{-1} associated with the increased concentration of IgG.

To determine any relationship between variations in the spectra and the protein concentrations (HSA and IgG), PLSR analysis was conducted. The optimum number of dimensions was selected by plotting the RMSE from the validation set *vs*. the number of dimensions, and an example of such a plot, for the 10% diluted air dried albumin analysis, is shown in Figure 6. For consistency, the minimum point on the curve was selected, in this case the 8^{th} dimension, determined at 2.347 ± 0.2788 mg mL^{-1}. This information is fed into a predictive model to compare the estimated concentrations from the spectral data set to the known concentrations from the produced solutions.

Table 2 summaries the predictive values from the protein spiked models, also provided in Supporting Information (SI Results and Discussion, *Figure S-2*). By comparing the R^2 values, as well as the RMSE of the validation set (RMSEV), the best overall result came from the 10 % diluted air dried samples. These results show that concentrations can be estimated unambiguously and that dilution ensures that the protein absorbances are within the range of validity of the Beer-Lambert Law. Results highlight that, after air drying, consistent and reproducible spectra are obtained. The best individual result came from the pure air dried IgG spiked samples, the linearity being, $R^2 = 0.998$ and the RMSEV being 0.49 ± 0.05 mg mL^{-1}.

Summarising, the standard deviations across all the predictive values represent good repeatability between the cross-validation iterations. As the 10 % diluted air dried samples produce the best overall predictive values, this sample preparation protocol was adopted for the analysis of the patient samples. Interestingly, the PLSR models from the liquid samples produce comparable predictive results to the air dried samples. Due to the speed and ease of acquiring liquid ATR-FTIR spectra, given the removal of the rate determining step (5-8 minute drying time), the liquid sample state was also considered in the patient sample analysis steps.

It was necessary to carry out these methodological analysis steps before progressing to patient samples and model blind testing, to establish the optimum sample preparation protocol, leading to the best possible predictive values. While the work carried out by Perez *et al*. showed excellent results, this particular type of methodology development procedure was not considered. This led to protein quantification of 50 µl liquid serum samples using an ATR crystal cell, potentially missing the demonstrated potential of diluted serum sample analysis [43].

3.2. Protein level quantification in patient samples

Due to the natural biological variation between individuals, the analysis of patient samples can be considered more challenging than spiking commercially available pooled human serum. Spiking a sample with a known amount of a specific biological component can model one physiological change, whilst everything else remains consistent. Between patients, the composition of blood can vary for multiple reasons, including diet, time of sample collection, as well as their disease state. During routine blood analysis, multiple biomolecular concentrations are measured, including the total protein concentration made up mainly of HSA and immunoglobulins. It is therefore important to analyse patient samples in order to understand the potential variance in the spectral response in order for these spectra to be used for clinical purposes. As the 10 % diluted air dried samples produced the best predictive models, the patient samples were analysed in this sample state. In addition, the patient samples were also analysed in the liquid form, due to the promising performance of these samples during calibration and the shorter collection time which is an important parameter in clinical situations.

3.2.1.10 % Diluted Air Dried Patient Samples

A new PLSR model was calculated based on the 20 patient samples using the 10 % diluted air dried sample preparation, these are listed in Table 3 and provided in Supporting Information (SI Results and Discussion, *Figure S-3*). For the quantification of total protein concentration, a RMSEV of 0.662 ± 0.046 mg mL^{-1} and an R^2 value of 0.992 was achieved. This result suggests that, despite moving to a more complex serum sample, a high level of predictive power is maintained and the relationship between the spectral variations and the total protein concentration is linear, within standard deviation. Quantification of the individual protein concentrations resulted in the HSA performing better than the spiked model and the IgG performing poorer than the spiked model, when examining the RMSEV and R^2 values detailed in Table 2 (spiked) and 3 (patient).

3.2.2.Liquid Patient Samples

Similarly, the results from the PLSR analysis of the 20 patient liquid samples are detailed in Table 3 and provided in Supporting Information (SI Results and Discussion, *Figure S-3*). It is evident that the RMSEV values are relatively consistent with those achieved for the spiked samples. The HSA patient model produced a result of 2.56 ± 0.35 mg mL^{-1}, compared to the HSA spiked models result of 3.065 ± 0.290 mg mL^{-1}. However, when comparing the RMSEV values of the patient liquid to the patient 10% diluted air dried samles, the results are dramatically higher. This suggests that the analysis of the liquid patient samples produce models with a reduced predictive power.

When considering the R^2 values, results show that the values for the liquid patient samples are considerably lower than both the spiked models and the 10 % diluted air dried patient samples. The drop-in linearity implies the spectral variations and the protein concentrations show less correlation; the best result achieved was 0.831 for the total protein concentration, dramatically less than the 0.962 achieved for the analysis of the sample patient set in the 10 % diluted air dried state. The error bars displayed on the plot (Figure 7) are increased in size and in some areas, show overlap. From this analysis, it is evident that the patient sample concentrations cannot be quantified unambiguously.

3.3. Model Validation

Following the determination of the optimal sample states from predicting the spiked samples and then clarifying using the patient sample concentrations, the next stage was to determine the ability of ATR-FTIR spectroscopy to predict unknown protein concentrations from serum. This was done by blind testing, removing knowledge of the concentrations from the model.

3.3.1.Leave One Patient Out Cross Validation (LOPOCV) of Patient Based Model

The first method of validating the use of ATR-FTIR spectroscopy to predict serum protein concentrations involved the use of a LOPOCV process. From the results in Table 4, it is evident that the prediction of the total protein concentration produced the best results, with an RMSEV of 1.534 ± 1.14 mg mL^{-1} and an R^2 value of 0.926. The prediction of individual HSA and IgG concentrations were not as effective as the total protein concentration, represented by higher RMSECV values for both proteins. The high standard deviation of the IgG data (± 2.14), results in the model not being able to identify individual patient concentrations with precision. The R^2 values for both proteins are also lower than those of the total protein concentration, showing correlation between spectral variations and concentrations decreases.

3.3.2.K-fold Cross Validation of Patient Based Model

A similar trend to the LOPOCV is apparent in Table 4. The total protein content allows for the best predictive values, followed by the HSA and then the IgG. The RMSECV for the total protein concentration prediction is higher (1.99 ± 0.78 mg mL^{-1}) than the LOPOCV model, but the lower standard deviation suggests that this is a more precise method. For this reason, the R^2 value of 0.934 is also higher than that of the previous method, showing more linearity between the predicted and observed concentrations. The prediction of the individual proteins shows the same trend. For both HSA and IgG, the k-fold blind testing produces less accurate results, in comparison to the true result but with more precision and reduced statistical variability. The linearity of the models decreases and again highlights that a linear relationship between the spectral variations and the concentrations.

Both the LOPOCV and k-fold blind testing methods produced promising results with similar trends, specifically for the prediction of total protein concentration. The reason for the poor IgG results could be due to the inability to differentiate between the variable contributions of the five major types of immunoglobulin present within human blood (IgA, IgG, IgM, IgE and IgD) [52].

4. Conclusions

Vibrational spectroscopy is widely used for the analysis of biofluids and many proof-of-principle studies have shown the capability of techniques like ATR-FTIR spectroscopy to enable disease detection, as well as quantification of biomolecules However, the translation of vibrational spectroscopy into a clinical environment is dramatically impacted by the inability to perform a direct comparison to the current quantitative diagnostic measurements. Current clinical practice uses blood protein concentration as a non-specific disease indicator, possibly leading to further investigation and potentially a diagnosis, highlighting the advantageous nature of protein quantification.

The work presented showcases for the first time the development of the optimal methodology for the quantification of protein biomarkers in a complex background (namely 10% diluted air dried serum analysis), the inclusion of this methodology into vibrational spectroscopic diagnostics of biofluids could bridge the gap between vibrational spectroscopy and clinical practice. In addition, the drying process could be accelerated through the implementation of heating mantle, the use of a smaller sample volume or batch drying, before analysis.

This study shows how ATR-FTIR spectroscopy can be used to quantify proteins in spiked and patient samples, rapidly, economically and with simple sample preparation. Linearity values as high as 0.992, in addition to high accuracy and precision demonstrated by RMSEV values such as 0.662 ± 0.046 mg mL^{-1}, indicate that quantification of clinically relevant molecules can be conducted using this approach.

The blind testing of patient clinical samples, while maintaining desirable linearity (R^2 = 0.934), precision and accuracy (RSMEV = 1.986 ± 0.778 mg mL^{-1}), illustrates the potential use of this technique within a clinical setting and its incorporation could bridge the gap between vibrational spectroscopy and current clinical analyses. The development of a quantification step in addition to disease

differentiation shows great promise to enable a dynamic clinical diagnostic platform that can improve the current patient diagnostic pathway.

Acknowledgements
The authors would like to acknowledge funding from the Postgraduate and Early Career Researcher Exchanges (PECRE) and the Pool Engagement in European Research (PEER) schemes, as well as support from Rosemere Cancer Foundation and the Spectral Analytics Laboratory, University of Strathclyde.

Competing interests statement
The authors declare that no competing interests exist.

References

[1] M.J. Baker, S.R. Hussain, L. Lovergne, V. Untereiner, C. Hughes, R.A. Lukaszewski, G. Thiéfin, G.D. Sockalingum, Developing and understanding biofluid vibrational spectroscopy: a critical review, Chem. Soc. Rev. 45 (2015) 1803–1818. doi:10.1039/C5CS00585J.

[2] J.R. Hands, P. Abel, K. Ashton, T. Dawson, C. Davis, R.W. Lea, A.J.S. McIntosh, M.J. Baker, Investigating the rapid diagnosis of gliomas from serum samples using infrared spectroscopy and cytokine and angiogenesis factors, Anal. Bioanal. Chem. 405 (2013) 7347–7355. doi:10.1007/s00216-013-7163-z.

[3] D.W. Greening, R.J. Simpson, Serum/Plasma Proteomics, 728 (2011) 259–265. doi:10.1007/978-1-61779-068-3.

[4] J. Liu, Y. Duan, Saliva: A potential media for disease diagnostics and monitoring, Oral Oncol. 48 (2012) 569–577. doi:10.1016/j.oraloncology.2012.01.021.

[5] R.J. Perry, V.T. Samuel, K.F. Petersen, G.I. Shulman, N. Haven, N. Haven, HHS Public Access, 510 (2015) 84–91. doi:10.1038/nature13478.The.

[6] S. Beauclercq, L. Nadal-Desbarats, C. Hennequet-Antier, A. Collin, S. Tesseraud, M. Bourin, E. Le Bihan-Duval, C. Berri, Serum and Muscle Metabolomics for the Prediction of Ultimate pH, a Key Factor for Chicken-Meat Quality, J. Proteome Res. 15 (2016) 1168–1178. doi:10.1021/acs.jproteome.5b01050.

[7] C. Bruno, D. Dufour-Rainfray, F. Patin, P. Vourch, D. Guilloteau, F. Maillot, F. Labarthe, M. Tardieu, C.R. Andres, P. Emond, H. Blasco, Validation of amino-acids measurement in dried blood spot by FIA-MS/MS for PKU management, Clin. Biochem. 49 (2016) 1047–1050. doi:10.1016/j.clinbiochem.2016.07.008.

[8] E. Sato, T. Mori, E. Mishima, A. Suzuki, S. Sugawara, N. Kurasawa, D. Saigusa, D. Miura, T. Morikawa-Ichinose, R. Saito, I. Oba-Yabana, Y. Oe, K. Kisu, E. Naganuma, K. Koizumi, T. Mokudai, Y. Niwano, T. Kudo, C. Suzuki, N. Takahashi, H. Sato, T. Abe, T. Niwa, S. Ito, Metabolic alterations by indoxyl sulfate in skeletal muscle induce uremic sarcopenia in chronic kidney disease, Sci. Rep. 6 (2016) 36618. doi:10.1038/srep36618.

[9] G.L. Wright Jr, L. Cazares, S.-M. Leung, S. Nasim, B.-L. Adam, T. Yip, P. Schellhammer, L. Gong, A. Vlahou, ProteinChip® surface enhanced laser desorption/ionization (SELDI) mass spectrometry: A novel protein biochip technology for detection of prostate cancer biomarkers in complex protein mixtures, 2000. doi:10.1038/sj.pcan.4500384.

[10] A. Vlahou, P.F. Schellhamrner, S. Mendrinos, K. Patel, F.I. Kondylis, L. Gong, S. Nasim, G.L. Wright, Development of a novel proteomic approach for the detection of transitional cell carcinoma of the bladder in urine, Am. J. Pathol. 158 (2001) 1491–1502. doi:10.1016/S0002-9440(10)64100-4.

[11] E.F. Petricoin, A.M. Ardekani, B.A. Hitt, P.J. Levine, V.A. Fusaro, S.M. Steinberg, G.B. Mills, C. Simone, D.A. Fishman, E.C. Kohn, L.A. Liotta, Use of proteomic patterns in serum to identify ovarian cancer, Lancet. 359 (2002) 572–577. doi:10.1016/S0140-6736(02)07746-2.

[12] H. Zhang, G. Wu, H. Tu, F. Huang, Discovery of serum biomarkers in astrocytoma by SELDI-TOF MS and proteinchip technology, J. Neurooncol. 84 (2007) 315–323. doi:10.1007/s11060-007-9376-5.

[13] K. Spalding, R. Board, T. Dawson, M.D. Jenkinson, M.J. Baker, A review of novel analytical diagnostics for liquid biopsies: spectroscopic and spectrometric serum profiling of primary and secondary brain tumors, Brain Behav. 6 (2016) 1–8. doi:10.1002/brb3.502.

[14] J.R. Hands, K.M. Dorling, P. Abel, K.M. Ashton, A. Brodbelt, C. Davis, T. Dawson, M.D. Jenkinson, R.W. Lea, C. Walker, M.J. Baker, Attenuated Total Reflection Fourier Transform Infrared (ATR-FTIR) spectral discrimination of brain tumour severity from serum samples, J. Biophotonics. 7 (2014) 189–199. doi:10.1002/jbio.201300149.

[15] J.R. Hands, G. Clemens, R. Stables, K. Ashton, A. Brodbelt, C. Davis, T.P. Dawson, M.D. Jenkinson, R.W. Lea, C. Walker, M.J. Baker, Brain tumour differentiation: rapid stratified serum diagnostics via attenuated total reflection Fourier-transform infrared spectroscopy, J. Neurooncol. 127 (2016) 463–472. doi:10.1007/s11060-016-2060-x.

[16] B.R. Smith, K.M. Ashton, A. Brodbelt, T. Dawson, M.D. Jenkinson, N.T. Hunt, D.S. Palmer, M.J. Baker, Combining random forest and 2D correlation analysis to identify serum spectral signatures for neuro-oncology, Analyst. 141 (2016) 3668–3678. doi:10.1039/C5AN02452H.

[17] G. Clemens, J.R. Hands, K.M. Dorling, M.J. Baker, Vibrational spectroscopic methods for cytology and cellular research., Analyst. 139 (2014) 4411–44. doi:10.1039/c4an00636d.

[18] 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. doi:10.1039/C5AN01157D.

[19] J. Backhaus, R. Mueller, N. Formanski, N. Szlama, H.G. Meerpohl, M. Eidt, P. Bugert, Diagnosis of breast cancer with infrared spectroscopy from serum samples, Vib. Spectrosc. 52 (2010) 173–177. doi:10.1016/j.vibspec.2010.01.013.

[20] G.L. Owens, K. Gajjar, J. Trevisan, S.W. Fogarty, S.E. Taylor, B. Da 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, J. Biophotonics. 7 (2014) 200–209. doi:10.1002/jbio.201300157.

[21] M.J. Baker, C.S. Hughes, K.A. Hollywood, Biophotonics: Vibrational Spectroscopic Diagnostics, 2016. doi:10.1088/978-1-6817-4071-3.

[22] A. Barth, Infrared spectroscopy of proteins, Biochim. Biophys. Acta - Bioenerg. 1767 (2007) 1073–1101. doi:10.1016/j.bbabio.2007.06.004.

[23] R.A.M.& M.R. Pincus, Henry's Clinical Diagnosis and Management by Laboratory Methods, 23rd Edition, 2011.

[24] J.T. Busher, Serum Albumin and Globulin, Clin. Methods Hist. Phys. Lab. Exam. (1990) 497–499.

[25] J. Barth, J.K. Rae, Harmonisation of Reference Intervals President, Association for Clinical Biochemistry, (2011).

[26] P. Larkin, Introduction, Infrared Raman Spectrosc. (2011) 1–5. doi:10.1016/B978-0-12-386984-5.10001-1.

[27] R.D. Schoenwald, Basic Principles, Ther. Drug Monit. 1 (2002) 4–33. doi:10.1016/B978-0-12-386984-5.10002-3.

[28] 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, Nat. Protoc. 9 (2014) 1771–1791. http://dx.doi.org/10.1038/nprot.2014.110.

[29] A.L. Mitchell, K.B. Gajjar, G. Theophilou, F.L. Martin, P.L. Martin-Hirsch, Vibrational spectroscopy of biofluids for disease screening or diagnosis: Translation from the laboratory to a clinical setting, J.

Biophotonics. 7 (2014) 153–165. doi:10.1002/jbio.201400018.

[30] C. Hughes, M. Brown, G. Clemens, A. Henderson, G. Monjardez, N.W. Clarke, P. Gardner, Assessing the challenges of Fourier transform infrared spectroscopic analysis of blood serum, J. Biophotonics. 7 (2014) 180–188. doi:10.1002/jbio.201300167.

[31] P.R. Griffiths, J. a. de Haseth, Chapter-2 Theoretical background, 2007. doi:10.1002/047010631X.

[32] R.A. Shaw, H.H. Mantsch, Infrared Spectroscopy in Clinical and Diagnostic Analysis, Encycl. Anal. Chem. (2006) 1–20. doi:10.1002/9780470027318.a0106.

[33] A. Oleszko, J. Hartwich, A. Wójtowicz, M. Gąsior-Głogowska, H. Huras, M. Komorowska, Comparison of FTIR-ATR and Raman spectroscopy in determination of VLDL triglycerides in blood serum with PLS regression, Spectrochim. Acta - Part A Mol. Biomol. Spectrosc. 183 (2017) 239–246. doi:10.1016/j.saa.2017.04.020.

[34] G. Sankari, E. Krishnamoorthy, S. Jayakumaran, S. Gunasekaran, V. Vishnu Priya, S. Subramaniam, S. Subramaniam, S.K. Mohan, Analysis of serum immunoglobulins using Fourier transform infrared spectral measurements, Biol. Med. 2 (2010) 42–48.

[35] R.A. Shaw, S. Kotowich, M. Leroux, H.H. Mantsch, Multianalyte Serum Analysis Using Mid-Infrared Spectroscopy, Ann. Clin. Biochem. An Int. J. Biochem. Lab. Med. 35 (1998) 624–632. doi:10.1177/000456329803500505.

[36] S. Roy, D. Perez-Guaita, D.W. Andrew, J.S. Richards, D. McNaughton, P. Heraud, B.R. Wood, Simultaneous ATR-FTIR Based Determination of Malaria Parasitemia, Glucose and Urea in Whole Blood Dried onto a Glass Slide, Anal. Chem. 89 (2017) 5238–5245. doi:10.1021/acs.analchem.6b04578.

[37] D.R. Whelan, K.R. Bambery, L. Puskar, D. Mcnaughton, B.R. Wood, Quantification of DNA in simple eukaryotic cells using Fourier transform infrared spectroscopy, J. Biophotonics. 6 (2013) 775–784. doi:10.1002/jbio.201200112.

[38] H.M. Heise, G. Voigt, P. Lampen, L. Küpper, S. Rudloff, G. Werner, Multivariate calibration for the determination of analytes in urine using mid-infrared attenuated total reflection spectroscopy, Appl. Spectrosc. 55 (2001) 434–443. doi:10.1366/0003702011951948.

[39] S. Khaustova, M. Shkurnikov, E. Tonevitsky, V. Artyushenko, A. Tonevitsky, Noninvasive biochemical monitoring of physiological stress by Fourier transform infrared saliva spectroscopy, Analyst. 135 (2010) 3183–3192. doi:10.1039/C0AN00529K.

[40] I. Elsohaby, J.T. McClure, C.B. Riley, R.A. Shaw, G.P. Keefe, Quantification of bovine immunoglobulin G using transmission and attenuated total reflectance infrared spectroscopy, J. Vet. Diagnostic Investig. 28 (2016) 30–37. doi:10.1177/1040638715613101.

[41] 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. doi:10.1002/jbio.201600015.

[42] F. Bonnier, H. Blasco, C. Wasselet, G. Brachet, R. Respaud, L.F.C.S. 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. (2017). doi:10.1039/C6AN01888B.

[43] D. Perez-Guaita, J. Ventura-Gayete, C. Pérez-Rambla, M. Sancho-Andreu, S. Garrigues, M. De La Guardia, Protein determination in serum and whole blood by attenuated total reflectance infrared spectroscopy, Anal. Bioanal. Chem. 404 (2012) 649–656. doi:10.1007/s00216-012-6030-7.

[44] H.J. Byrne, M. Baranska, G.J. Puppels, N. Stone, B. Wood, K.M. Gough, P. Lasch, P. Heraud, J. Sulé-Suso, G.D. Sockalingum, Spectropathology for the next generation: quo vadis?, Analyst. 140 (2015) 2066–73. doi:10.1039/c4an02036g.

[45] Y. Xu, H. Muhamadali, A. Sayqal, N. Dixon, R. Goodacre, Partial least squares with structured output for modelling the metabolomics data obtained from complex experimental designs: A study into the ??-

block coding, Metabolites. 6 (2016). doi:10.3390/metabo6040038.

[46] A. Khoshmanesh, M.W.A. Dixon, S. Kenny, L. Tilley, D. McNaughton, B.R. Wood, Detection and Quantification of Early-Stage Malaria Parasites in Laboratory Infected Erythrocytes by Attenuated Total Reflectance Infrared Spectroscopy and Multivariate Analysis, Anal. Chem. 86 (2014) 4379–4386. doi:10.1021/ac500199x.

[47] B. a Walther, J.L. Moore, The concept of bias, precison and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimators, Ecography (Cop.). 28 (2005) 815–829. doi:10.1111/j.2005.0906-7590.04112.x.

[48] J. Coates, Interpretation of Infrared Spectra, A Practical Approach, Encycl. Anal. Chem. (2000) 10815–10837. doi:10.1002/9780470027318.

[49] F. Bonnier, F. Petitjean, M.J. Baker, H.J. Byrne, Improved protocols for vibrational spectroscopic analysis of body fluids, J. Biophotonics. 7 (2014) 167–179. doi:10.1002/jbio.201300130.

[50] Z. Movasaghi, S. Rehman, D.I. ur Rehman, Fourier Transform Infrared (FTIR) Spectroscopy of Biological Tissues, Appl. Spectrosc. Rev. 43 (2008) 134–179. doi:10.1080/05704920701829043.

[51] B.R. Priya, H.J. Byrne, Investigation of Sodium Dodecyl Benzene Sulfonate Assisted Dispersion and Debundling of Single-Wall Carbon Nanotubes, J. Phys. Chem. C. 112 (2008) 332–337. doi:10.1021/jp0743830.

[52] A.A. Marghoob, K. Koenig, F. V. Bittencourt, A.W. Kopf, R.S. Bart, Breslow thickness and Clark level in melanoma: Support for including level in Pathology Reports and in American Joint Committee on Cancer Staging, Cancer. 88 (2000) 589–595. doi:10.1002/(SICI)1097-0142(20000201)88:3<589::AID-CNCR15>3.0.CO;2-I.

Figure 1 – Highlighting spectral pre-processing steps. From left to right, raw data, baseline corrected data and finally baseline corrected and vector normalised.

Figure 2 –A representative convergence plot of the R^2 value +/- SE vs. the no. of iterations from the 10 % diluted air dried globulin analysis. This particular plot led to the selection of 26 iterations which was compared to the number selected from the RMSEC and the RMSEV plots, before the highest value was selected and taken forward to the PLSR analysis

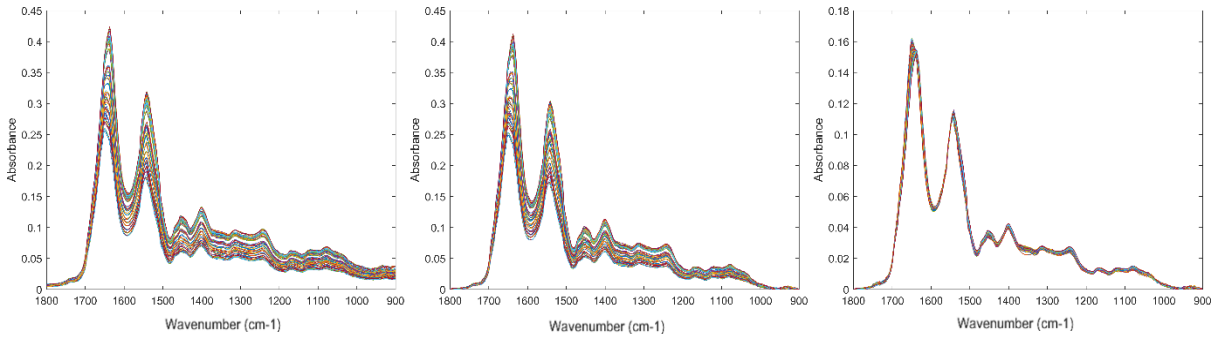
Figure 3 – Left: Mean ATR-FTIR spectra collected from the analysis of the air dried 2-fold dilution set of pooled serum. Red: 0 % serum, Pink: 3.125 %, Yellow: 6.25 %, Orange: 12.5 %, Green: 25 %, Blue: 50 % and Black: 100 % Stars highlight peaks of interest. Right: AUC plot from the fingerprint region.

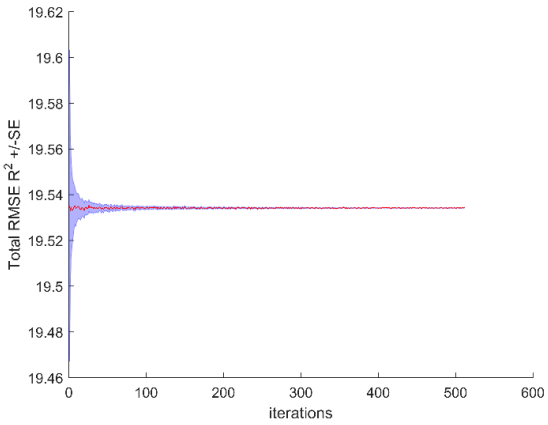
Figure 4 - Left: Mean ATR-FTIR spectra collected from the analysis of the liquid 2-fold dilution set of pooled serum Red: 0 % serum, Pink: 3.125 %, Yellow: 6.25 %, Orange: 12.5 %, Green: 25 %, Blue: 50 % and Black: 100 % Stars highlight peaks of interest. Right: AUC plot from the fingerprint region.

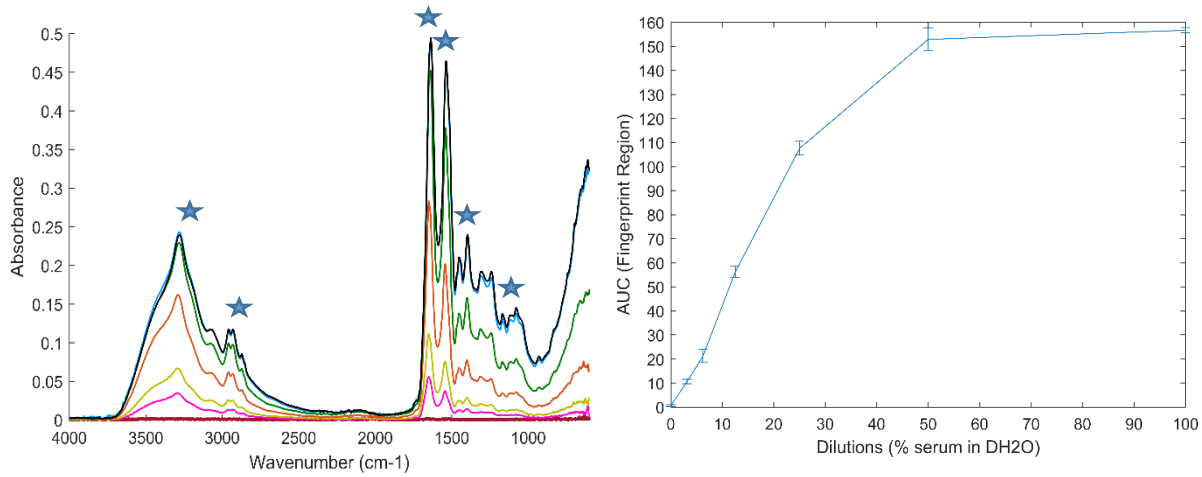
Figure 5 – Mean ATR-FTIR fingerprint spectra following the analysis of the 10 % diluted air dried, IgG spiked samples. Red: 13.53 mg/ml, Pink: 18.48 mg/ml, Yellow: 23.58 mg/ml, Orange: 28.53 mg/ml, Green: 33.48 mg/ml, Blue: 38.58 mg/ml and Black: 43.53 mg/ml.

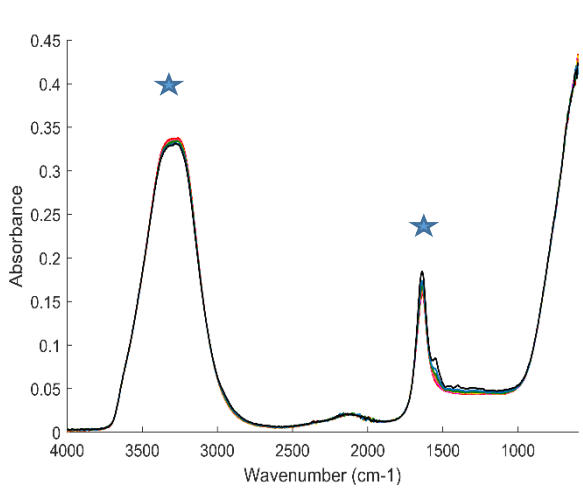
Figure 6 - Evolution of the root mean square error on the validation set (RMSEV). In this case, values are averaged from the 234 cross validations.

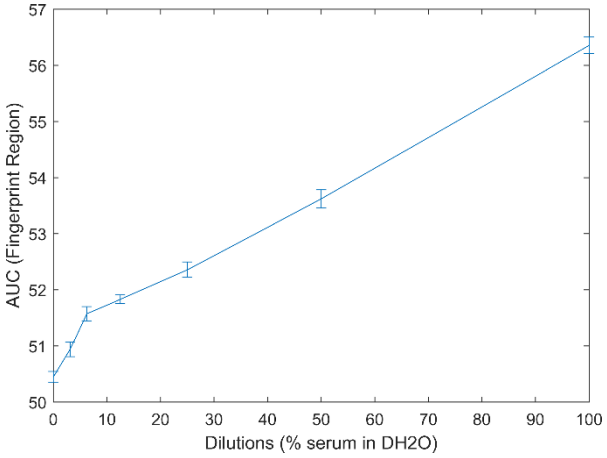
Figure 7 - Predictive model built from the PLS analysis of the liquid IgG patient samples. For each concentration the values displayed are an average of the concentration predicted from the iterations of the cross validation. Shown on the plot is the RMSEV, R^2 and the standard deviation corresponding to each of the values.

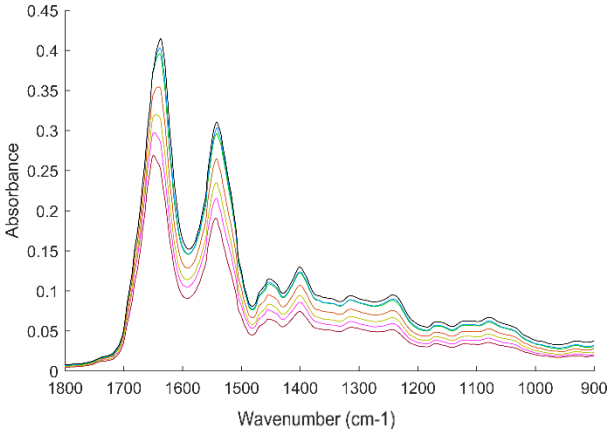

Absorbance
Wavenumber (cm-1)
Absorbance
Wavenumber (cm-1)
Absorbance
Wavenumber (cm-1)

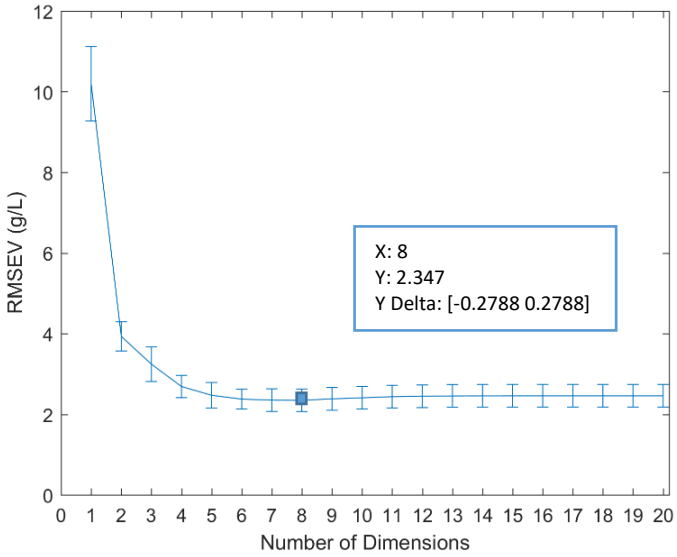

19.62
19.6
19.58
19.56
19.54
19.52
19.5
19.48
19.46
Total RMSE R2 +/-SE
0
100
200
300
400
500
600
iterations


Absorbance
Wavenumber (cm-1)
AUC (Fingerprint Region)
Dilutions (% serum in DH2O)


Absorbance
Wavenumber (cm-1)
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
4000
3500
3000
2500
2000
1500
1000


AUC (Fingerprint Region)
Dilutions (% serum in DH2O)
57
56
55
54
53
52
51
50
0
10
20
30
40
50
60
70
80
90
100


0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Absorbance
1800
1700
1600
1500
1400
1300
1200
1100
1000
900
Wavenumber (cm-1)


RMSEV (g/L)
Number of Dimensions
X: 8
Y: 2.347
Y Delta: [-0.2788 0.2788]
0
2
4
6
8
10
12
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

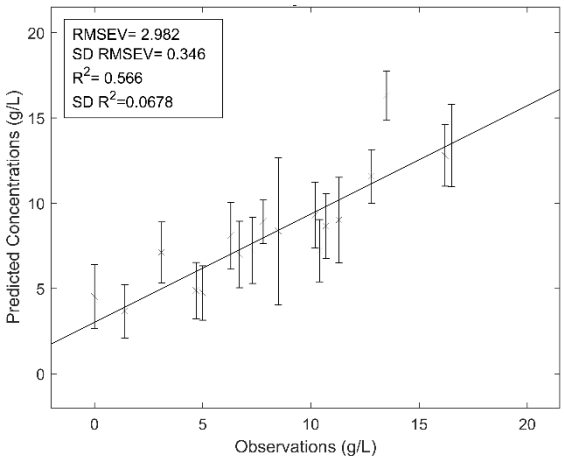

RMSEV= 2.982
SD RMSEV= 0.346
R2= 0.566
SD R2=0.0678
Predicted Concentrations (g/L)
Observations (g/L)
0
5
10
15
20

Table 1 - Experimental details

Table 2 – Summary of the RMSEV ± STD and R2 values from the predictive models, for the two protein spikes

Table 3 - Summary of the RMSEV ± STD and R^2 values from the three predictive models, for the two patient sample states

Table 4 - Summary of the RMSEV ± STD and R^2 values from the two blind predictive models, for the three different protein concentrations of the 10 % diluted air dried samples

		Spiked Human Serum Models	
	Whole Serum Dilution Study	**HSA**	**IgG**
Sample Preparation	2 - fold dilutions from pure human pooled serum	0.14 g HSA/2000 µl human pooled serum	0.06 g IgG/2000 µl human pooled serum
Sample Concentrations	100, 50, 25, 12.5, 6.25, 3.125, 0 %	116.3, 106.29, 96.28, 86.27, 76.33, 66.32, 46.3 mg mL^{-1}	43.53, 38.58, 33.48, 28.53, 23.58, 18.48, 13.53 mg mL^{-1}
Sample States Analysed	10 µl liquid, 1 µl pure air dried	10 µl liquid, 1 µl pure air dried, 2 µl 10% diluted air dried	10 µl liquid, 1 µl pure air dried, 2 µl 10% diluted air dried

Protein	Air Dried		Liquid		10 % Diluted Air Dried	
	RMSECV ± STD (mg mL^{-1})	R^2	RMSECV ± STD (mg mL^{-1})	R^2	RMSECV ± STD (mg mL^{-1})	R^2
HSA	4.585 ± 0.568	0.959	3.065 ± 0.290	0.982	2.347 ± 0.287	0.989
IgG	0.487 ± 0.053	0.998	2.365 ± 0.194	0.947	0.861 ± 0.104	0.993

Protein	10 % Diluted Air Dried		Liquid	
	RMSECV ± STD (mg mL^{-1})	R^2	RMSECV ± STD (mg mL^{-1})	R^2
Total	0.662 ± 0.046	0.992	3.080 ± 0.483	0.831
HSA	0.848 ± 0.064	0.976	2.556 ± 0.351	0.780
IgG	1.945 ± 0.134	0.812	2.982 ± 0.346	0.566

Protein	LOPOCV		K-Fold CV	
	RMSEV ± STD (mg mL^{-1})	R^2	RMSEV ± STD (mg mL^{-1})	R^2
Total	1.534 ± 1.140	0.926	1.986 ± 0.778	0.934
HSA	2.029 ± 1.260	0.890	2.491 ± 0.849	0.805
IgG	3.582 ± 2.140	0.827	4.464 ± 1.460	0.454