

# Detection of Glycine as a Model Protein in Blood Serum using 2D-IR Spectroscopy

Samantha H. Rutherford,<sup>‡,†,\*</sup> Gregory M. Greetham,<sup>§</sup> Paul M. Donaldson,<sup>§</sup> Michael Towrie,<sup>§</sup> Anthony W. Parker,<sup>§</sup> Matthew J. Baker,<sup>#</sup> Neil T. Hunt,<sup>||</sup>

<sup>‡</sup> Department of Physics, University of Strathclyde, SUPA, 107 Rottenrow East, Glasgow, G4 0NG, UK

<sup>§</sup> STFC Central Laser Facility, Research Complex at Harwell, Rutherford Appleton Laboratory, Harwell Campus, Didcot, OX11 0QX, UK

<sup>#</sup> WestCHEM, Department of Pure and Applied Chemistry, University of Strathclyde, Technology and Innovation Centre, 99 George Street, Glasgow, G1 1RD, UK

<sup>||</sup> Department of Chemistry and York Biomedical Research Institute, University of York, Heslington, York, YO10 5DD, UK

\*Corresponding author email: Samantha.hume-rutherford@strath.ac.uk

---

**ABSTRACT:** Glycine (Gly) is used as a model system to evaluate the ability of ultrafast 2D-IR spectroscopy to detect and quantify the low-molecular weight proteinaceous components of blood serum. Combining data acquisition schemes to suppress absorption bands of H<sub>2</sub>O that overlap with the protein amide I band with analysis of peak patterns appearing in the off-diagonal region of the 2D-IR spectrum allows separation of the Gly spectral signature from that of the dominant protein fraction of serum in a transmission-mode 2D-IR measurement without any sample manipulation, e.g. filtration or drying. 2D-IR spectra of blood serum samples supplemented with varying concentrations of Gly were obtained and a range of data analysis methods compared, leading to a detection limit of ~ 3 mg/mL for Gly. The reported methodology provides a platform for a critical assessment of the sensitivity of 2D-IR for measuring the concentrations of amino acids, peptides and low molecular weight proteins present in serum samples. We conclude that, in the case of several clinically relevant diagnostic molecules and their combinations, the potential exists for 2D-IR to complement IR absorption methods as the benefits of the second frequency dimension offered by 2D-IR spectroscopy outweigh the added technical complexity of the measurement.

---

Blood serum can be easily obtained and provides a means to acquire important biomedical information relating to health. By virtue of coming into contact with most of the major organs of the body, serum provides a molecular picture of metabolic processes. Analysis of the chemical composition of serum therefore gives access to a number of potentially-useful diagnostic markers, originating from the circulatory proteome and the low molecular weight peptidome as well as lipids, sugars and nucleic acids.<sup>1-3</sup>

Extracting the information content of serum is technically challenging. The total protein content of serum lies in the range of 60-80 mg/mL and comprises more than fifty individual proteins.<sup>4</sup> The molecular weights and concentrations of individual proteins span a large range while the latter can also show considerable patient to patient variation.<sup>5</sup> The large molecule fraction of serum includes dominant species such as serum albumin (66.5 kDa; 35-50 mg/mL), immunoglobulins such as IgG (160 kDa; 8-18 mg/mL) and transferrin (76 kDa; 1.5-4 mg/mL) as well as low concentration proteins like IgE (190 kDa; 6 x 10<sup>-4</sup> mg/mL).<sup>6</sup> The low molecular weight portion of serum is typically defined as containing molecules with masses < 25 kDa and includes cytokines (< 6-30 kDa), peptide hormones (< 30 kDa) and amino acids (75-204 Da).<sup>7</sup>

Current analytical technologies employ antibody assays to measure protein levels in serum and rely on the availability of specific antibodies for proteins of interest. These methods

require time for sample preparation and measurement that can hinder treatment. Furthermore, diseases can be heterogeneous, meaning that knowledge of changes in concentration of a broad range of molecules can be more useful diagnostically than that of a single marker protein or a small antibody panel. In this respect, infrared (IR) spectroscopy is potentially powerful in that a spectrum, in principle, provides a detailed molecular fingerprint of a sample. However, the water content of serum means that the informative amide I band of proteins is obscured using IR transmission absorption spectroscopy. Alternative measurement methods, such as attenuated total reflection (ATR), transfection or the use of Quantum Cascade lasers have been applied to overcome this problem.<sup>1,8-10</sup> Similarly, measurements on dried serum droplets have been widely used.<sup>11,12</sup> Although effective, for example in the identification of samples from patients diagnosed with cancer, clinical monitoring of the cancer severity or identifying the presence of metastatic disease,<sup>13-16</sup> using quantitative IR approaches to measure protein levels has remained elusive.

In addition to information relating to overall protein content, access to the low molecular weight (LMW) fraction of serum is desirable from a diagnostic perspective. As disease states progress, changes in LMW composition can occur. For example, significant changes in amino acid levels (75 – 204 Da) have been documented for patients with neurodegenerative diseases.<sup>17,18</sup> The LMW peptide neuropeptide Y (4 kDa), is a

sympathetic neurotransmitter that regulates processes relevant to tumour biology in neuroblastomas and has been identified as a serum biomarker, where high levels of the peptide (> 0.5 ng/mL) are associated with poor prognosis and survival rates.<sup>19,20</sup> Analysis of the LMW proteinaceous fraction with IR spectroscopy is however hampered further by dominant signals from albumin and the globulins. Filtering samples to remove the larger species enables an increased sensitivity to LMW molecules using ATR-IR spectroscopy,<sup>21,22</sup> but at the expense of time and the possible unintentional removal of some LMW components. In particular, the role of albumin as a transport protein means that its removal during filtration could also result in the coincidental removal of bound LMW species and subsequently the much needed diagnostic information.<sup>7,23</sup>

It was demonstrated recently that ultrafast 2D-IR spectroscopy offers a route to circumventing the problem of water absorption in the amide I region of the IR spectrum.<sup>24</sup> The non-linear nature of the 2D-IR signal amplifies the high molar extinction coefficient protein amide I band relative to the weaker (per molecule) water bending vibrational mode, while the 50 fs time resolution of 2D-IR spectroscopy enables separation of the protein and water signals based on their different vibrational relaxation timescales. Using this approach, 2D-IR was able to measure the albumin to globulin ratio of as-received serum samples and showed sensitivity to changes in concentration of individual large globulin proteins IgG, IgA and IgM.<sup>24</sup>

Recent advances in laser and pulse control technology have significantly decreased both the experimental complexity and the time required to measure a 2D-IR spectrum to only a few seconds per spectrum.<sup>25-27</sup> These innovations in instrumentation provide a route to high throughput screening applications, using low volume samples (20  $\mu$ L), as recently demonstrated,<sup>28,29</sup> making this technique worthy of investigating as a potential clinical tool. Despite the innovations described, 2D-IR is more expensive and technically more complex to implement than IR absorption spectroscopy and therefore the initial steps in the pathway to impact is a thorough understanding of the range of molecular targets in serum that 2D-IR might be used to measure. This requires an evaluation of the sensitivity limits of 2D-IR to a range of molecular species. Using the simplest amino acid, Gly, as a model of a low molecular weight protein, we have measured serum spiked with a range of Gly concentrations, showing that, using current 2D-IR technology, the detection limit is  $\sim$ 3 mg/mL in as received serum. Key to this measurement is the characteristic network of off-diagonal peaks in the Gly 2D-IR spectrum, which allows spectral separation of its amide I contribution. Physiologically, Gly is present in serum at much lower levels than are detected here ( $\sim$ 20  $\mu$ g/mL), with current detection technologies (antibody assays and liquid chromatography tandem mass spectroscopy) allowing detection within the physiological range (100-390  $\mu$ mol/dm<sup>3</sup>).<sup>30,31</sup> However this low molecular weight study shows how the off-diagonal region of the serum 2D-IR spectrum could be diagnostic of smaller protein species and provides a means to evaluate which protein molecules and peptide fractions in serum are accessible.<sup>32</sup> Ultimately, we report here that 2D-IR does have the potential to complement IR absorption methods for specific biomedical problems in the clinical environment.

## EXPERIMENTAL SECTION

*Sample preparation:* Pooled equine serum, Gly, bovine serum albumin (BSA) and  $\gamma$ -globulins were obtained from Sigma

Aldrich and used without further purification. Individual samples of Gly, albumin and  $\gamma$ -globulin proteins were prepared in a Tris buffer, in H<sub>2</sub>O, at pH=7.5 to mimic the pH of the pooled serum samples. Serum samples containing Gly were prepared by adding Gly to pooled equine serum at a range of concentrations (Table 1).

*IR Spectroscopy:* For all IR and 2D-IR spectroscopy measurements, samples were placed between two CaF<sub>2</sub> windows in a standard transmission cell. To avoid saturation of the H-O-H bending ( $\delta_{\text{H-O-H}}$ ) mode of water near 1650 cm<sup>-1</sup> the sample thickness was carefully controlled. No spacer was used between the windows, but the tightness of the sample holder was adjusted to obtain approximately consistent absorbance values of  $\sim$ 0.1 OD for the combination band of the  $\delta_{\text{H-O-H}}$  and librational modes of water, located at 2130 cm<sup>-1</sup>.<sup>24</sup> Based upon the measured molar extinction coefficient of water, this corresponded to a sample thickness of  $\sim$ 3  $\mu$ m. IR absorption spectra were measured using a Thermo Scientific Nicolet iS10 Fourier Transform spectrometer. Spectra were the result of 20 co-added scans at a resolution of 1 cm<sup>-1</sup> in the spectral region 400 - 4000 cm<sup>-1</sup>.

Sample	[Gly] mg/mL	[Gly] mol/dm <sup>3</sup>
<sup>39</sup> [Gly]	0.78	0.01
<sup>78</sup> [Gly]	1.56	0.02
<sup>157</sup> [Gly]	3.13	0.04
<sup>313</sup> [Gly]	6.25	0.08
<sup>625</sup> [Gly]	12.5	0.17
<sup>1250</sup> [Gly]	25	0.33
<sup>2500</sup> [Gly]	50	0.67
<sup>5000</sup> [Gly]	100	1.33

**Table 1.** Summary of Gly-spiked serum samples studied, showing [Gly] in both mg/mL and mol/dm<sup>-3</sup>. Notation <sup>x</sup>[Gly] uses x to denote the relative ratio of Gly in spiked sample to that of physiological Gly serum levels irrespective of concentration units.

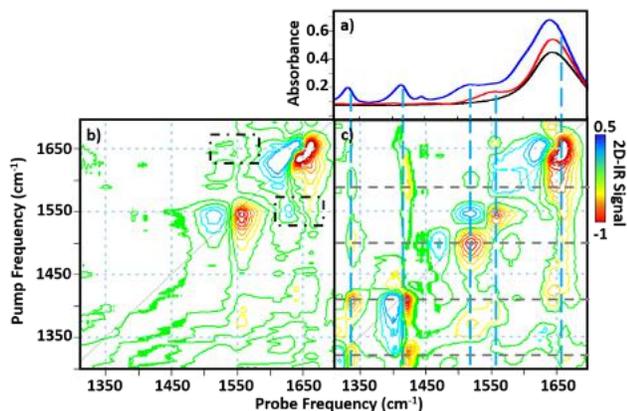
2D-IR spectra were recorded using the Ultra B laser spectrometer at the STFC Central Laser Facility using the Fourier Transform 2D-IR technique in which a sequence of three mid-IR laser pulses are arranged in the pseudo pump-probe geometry, which has been described previously.<sup>26,33-35</sup> Mid-IR pulses with a temporal duration of <50 fs with a central frequency of 1650 cm<sup>-1</sup> and a bandwidth of  $\sim$  400 cm<sup>-1</sup> were obtained. A waiting time of 250 fs between pump and probe pulses using a parallel pump-probe polarisation was employed throughout, which has been shown to minimise the water signal relative to that from the protein content of the serum.<sup>24</sup> All 2D-IR spectra were pre-processed and analysed using a custom made script written using the statistical analysis software programme, R.<sup>36</sup> Each measurement was made in triplicate using identical sample conditions to those used for IR absorption measurements and a baseline subtraction was performed using a 2<sup>nd</sup> order polynomial function and all spectra were normalised to the albumin  $\nu=0-1$  diagonal peak at 1656 cm<sup>-1</sup>.

## RESULTS & DISCUSSION

For ease of reference, each serum sample (Table 1) will be identified using the notation <sup>x</sup>[Gly], where x denotes the factor by which the Gly concentration exceeds that of typical serum

(20  $\mu\text{g/mL}$ ),<sup>32</sup> irrespective of units. In this manner, serum samples without additional Gly added are labelled  $^1[\text{Gly}]$ .

The IR absorption spectra of  $^1[\text{Gly}]$ ,  $^{5000}[\text{Gly}]$  and  $\text{H}_2\text{O}$  in the wavenumber range  $1300 - 1700 \text{ cm}^{-1}$  are shown in Fig.1(a). The  $\text{H}_2\text{O}$  spectrum (black) exhibits a broad band near  $1650 \text{ cm}^{-1}$  due to the  $\delta_{\text{H-O-H}}$  mode. This band is also dominant in the spectrum of  $^1[\text{Gly}]$  (red), where it overlaps, but does not fully obscure, the amide I band of the protein. The protein amide II band ( $1551 \text{ cm}^{-1}$ ) is also visible in the spectrum of  $^1[\text{Gly}]$ . The addition of Gly to serum ( $^{5000}[\text{Gly}]$ , blue) leads to the presence of additional bands due to the amino acid. Bands at  $1330$ ,  $1415$  and  $1445 \text{ cm}^{-1}$  are assigned to  $\text{CH}_2$  wagging, symmetric carboxylate stretching and  $\text{CH}_2$  scissor modes respectively.<sup>37</sup> A band at  $1515 \text{ cm}^{-1}$  is assigned to the in-plane N-H bending mode while the shoulder visible on the low frequency side of the  $1650 \text{ cm}^{-1}$  peak is due to the carboxylate C=O stretching mode.



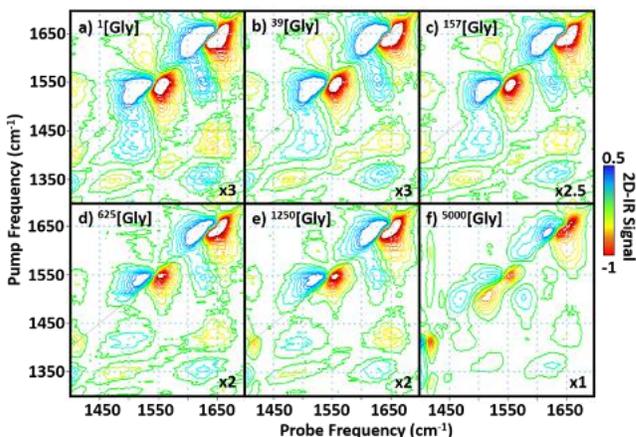
**Figure 1.** a) Infrared absorption spectra of  $\text{H}_2\text{O}$  (black), serum ( $^1[\text{Gly}]$ , red) and serum spiked with 100 mg/mL of glycine ( $^{5000}[\text{Gly}]$ , blue). b) 2D-IR spectrum of serum ( $^1[\text{Gly}]$ ), black dashed boxes indicate peaks due to coupling/energy transfer between amide I and II modes of the protein component (the peak above the diagonal is not clearly visible at these contour levels). c) 2D-IR spectrum of serum spiked with 100 mg/mL of glycine ( $^{5000}[\text{Gly}]$ ). Blue dashed vertical lines identify peak positions relative to IR absorption measurements. Intersections of blue lines with horizontal grey dashed lines indicate positions of off-diagonal peaks linking diagonal peaks due to Gly (see text).

The 2D-IR spectra of  $^1[\text{Gly}]$  and  $^{5000}[\text{Gly}]$  are shown in Figs.1(b) and (c) respectively. In the  $^1[\text{Gly}]$  spectrum, negative peaks (red) are visible on the 2D-IR spectrum diagonal at wavenumber coordinates (pump, probe) ( $1650, 1650 \text{ cm}^{-1}$ ) and ( $1550, 1550 \text{ cm}^{-1}$ ). These are assigned to the  $\nu=0-1$  transitions of the amide I and II bands of the protein content of the serum respectively. Positive (blue) peaks assigned to  $\nu=1-2$  transitions of the same bands are shifted to lower probe frequencies by the respective vibrational anharmonicities of the two modes. Weak off-diagonal peaks are observed linking these two diagonal peaks (black boxes), consistent with vibrational coupling and energy transfer processes between the amide I and II modes.<sup>38</sup> No contribution from water is observed in this spectrum, as reported previously.<sup>24,39</sup>

The 2D-IR spectrum of the  $^{5000}[\text{Gly}]$  sample (Fig.1(c)) contains a number of diagonal and off-diagonal peaks in addition to those already assigned to the serum proteins. Diagonal peaks corresponding to the vibrational modes of Gly observed in the IR absorption spectrum are located at (pump = probe)  $1330$ ,  $1445$  and  $1515 \text{ cm}^{-1}$ . For convenience these are identified in

Fig.1(c) by dashed vertical lines (blue) extending from the IR absorption spectrum to the 2D-IR diagonal. In addition, a network of off-diagonal peaks is observed linking these diagonal peaks, forming a characteristic 2D peak pattern for Gly (see grey horizontal guidelines in Fig.1(c)). Of particular note are several off-diagonal peaks linking the low frequency modes of Gly to the large diagonal peak at  $1650 \text{ cm}^{-1}$ , which indicate a contribution from Gly to this peak and provide a route to quantifying the contribution of Gly to the ‘amide I band’ of the serum sample.

**Figure 2.** 2D-IR spectra of serum spiked with Gly after normalisa-

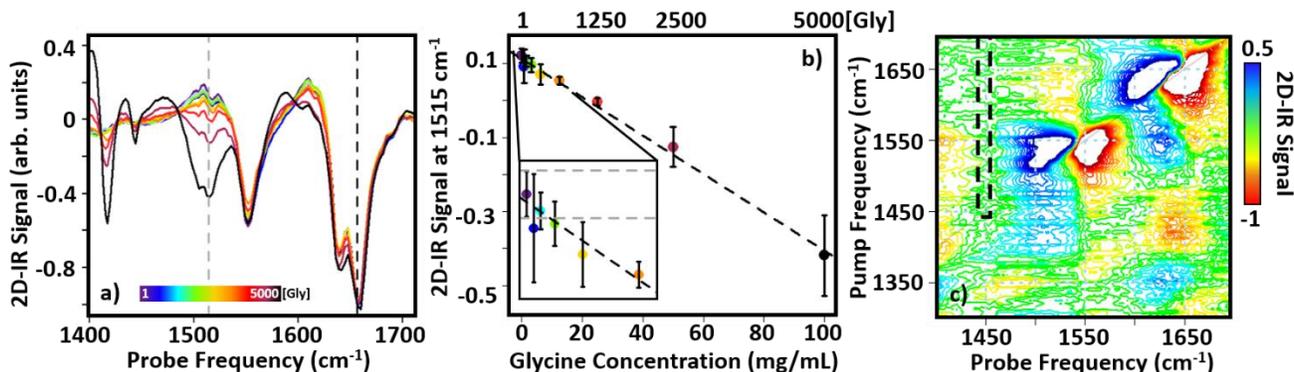


tion to the albumin signal at  $1656 \text{ cm}^{-1}$ . The figures show the average of 3 spectra for each Gly sample. All spectra are magnified by a factor indicated in bottom right hand corner, to enhance on and off diagonal peaks due to Gly.

In order to evaluate the detection limit for Gly using 2D-IR, spectra of a range of  $^x[\text{Gly}]$  samples were measured (Fig.2). Qualitatively, it can be seen that the peaks identified with Gly (Fig.1(c)) reduce in magnitude as the  $[\text{Gly}]$  diminishes, as would be expected. To develop a quantitative analytical method, we assess three different approaches to correlate the Gly contribution to the spectrum with concentration. It is anticipated that, in each case, the result will show a linear correlation with  $[\text{Gly}]$  until the limit of detection is reached.

i) The first method correlated the magnitude of the 2D-IR diagonal peak assigned to Gly at (pump = probe)  $1515 \text{ cm}^{-1}$  (Fig.3(a), grey dashed line) with  $[\text{Gly}]$ . The variation in amplitude of the  $1515 \text{ cm}^{-1}$  peak with  $[\text{Gly}]$  is shown in Fig.3(b). Although peaks associated with  $\nu=0-1$  transitions are negative in the 2D-IR plots shown, at low  $[\text{Gly}]$  a positive contribution at  $1515 \text{ cm}^{-1}$  is observed due to the  $\nu=1-2$  transition of the amide II mode of serum proteins (Fig.3(a)). As  $[\text{Gly}]$  increased, the  $\nu=0-1$  transition of the in-plane N-H bending mode of Gly increased progressively in amplitude, eventually dominating the spectrum in this region. Plotting the amplitude of the  $1515 \text{ cm}^{-1}$  diagonal peak produced a linear trend (Fig.3(b);  $R^2 = 0.994$ ) as expected.

Other diagonal peaks arising from the Gly content of the samples also yielded similar linear correlations, however as the diagonal peak at  $1515 \text{ cm}^{-1}$  produced the largest change in amplitude, it provided the strongest correlation and was found to offer the best position along the 2D-IR diagonal for Gly detection.

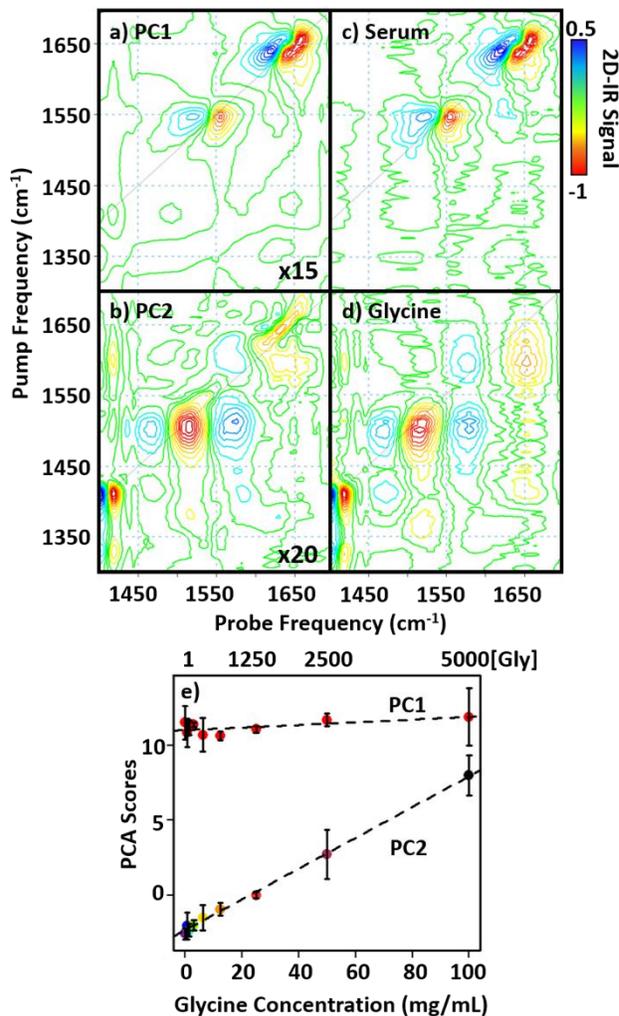


**Figure 3.** a) 2D-IR diagonal slices scaled to the amide I albumin peak at  $1656\text{ cm}^{-1}$  (black dashed vertical line). b) The average amplitude at  $1515\text{ cm}^{-1}$  (grey dashed vertical line in a)) plotted as a function of Gly concentration, error bars show  $1\sigma$  variation from measurement in triplicate. The optimised linear regression fit is shown as black dashed line with  $R^2=0.994$ . The inset shows the data for the range  $1\text{--}313\text{ [Gly]}$  expanded for clarity and dashed horizontal grey lines show noise levels for the spectra (see text). c) 2D-IR spectrum of serum ( $^{1}\text{[Gly]}$ ) plotted on same scale as in Figure 2, magnified by a factor of 5. The area identified by the black dashed box was used to calculate the RMS noise level shown in the inset of b).

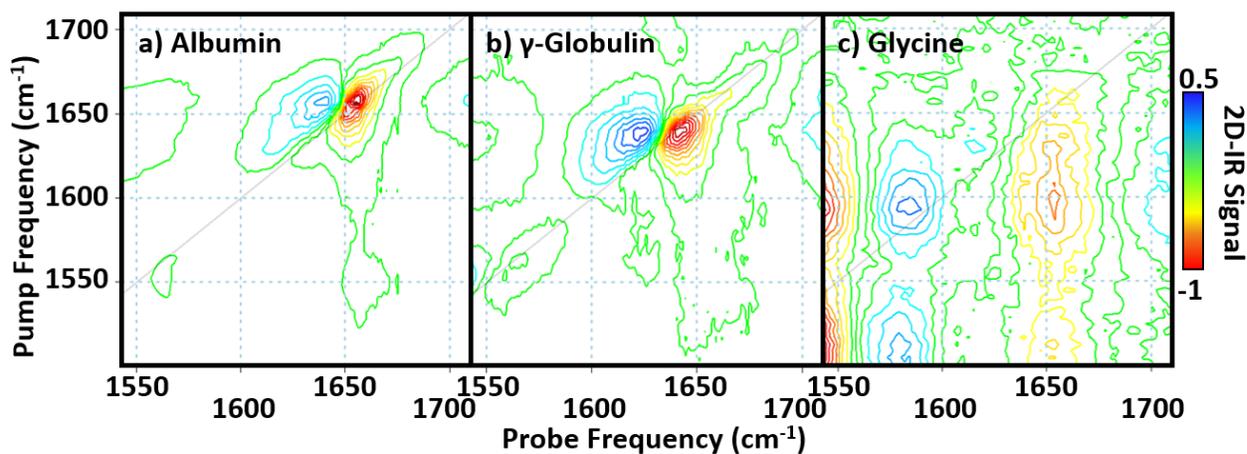
From the linear correlation in Fig 3b, it is possible to estimate the [Gly] value at which its 2D-IR signature becomes detectable against the background of experimental noise. By taking an area of the 2D-IR spectrum of the  $^{1}\text{[Gly]}$  sample free from any peaks (dashed box, Fig.3(c)), the root mean square (RMS) noise level was found to be  $\pm 9.6 \times 10^{-3}$  (the 2D-IR signal has both positive and negative components). Making the conservative assumption that only peaks exceeding twice this noise level will be visible consistently, leads to a detection limit for Gly of  $^{150}\text{[Gly]}$  ( $3\text{ mg/mL}$ ) (Fig.3(c)).

ii) The diagonal component of the 2D-IR spectrum does not utilise the additional spectral information from the off-diagonal peaks that were identified in Fig.1(c). Including this network of Gly-specific peaks may provide an improvement in the detection limit and this was tested by applying principal component analysis (PCA) to the set of 2D-IR spectra obtained as a function of [Gly], including the whole 2D spectral region studied. The results (Fig.4) showed that the variance in the data could be attributed mainly to the first two principal components, PC1 and PC2 (Fig.4), accounting for 87.25% and 7.5% of the variance, respectively. In the case of PC1, the relative loading was found to be largely invariant with [Gly] (Fig.4(e)), while the spectral information associated with this PC (Fig.4(a)) showed excellent agreement with the spectrum of neat serum ( $^{1}\text{[Gly]}$ ), Fig.4(c)). In contrast, the average PC2 scores showed a slightly superior linear correlation with [Gly] ( $R^2=0.997$ , Fig.4(e)) than method i) while the spectrum associated with PC2 (Fig.4(b)) clearly identified the network of diagonal and off-diagonal peaks assigned to Gly (Fig.4(d)).

As PCA is a dimension reduction technique it is often used as a noise removal method by actively removing principal components beyond the first one or two. This means that PCA can make small features more visible in the spectra of the major PCs. An example of this is present in PC1 where a very small peak appears on the diagonal at  $1415\text{ cm}^{-1}$  (Fig.4(a)).



**Figure 4.** Comparison of separate serum ( $^{1}\text{[Gly]}$ ) and Gly 2D-IR spectra with results of PCA of Gly-spiked serum samples. a) PC1 loading plot. b) PC2 loading plot. c) Serum ( $^{1}\text{[Gly]}$ ) spectrum. d) Gly spectrum ( $100\text{ mg/mL}$  in  $\text{H}_2\text{O}$ ). e) Averaged score results for PC1 (red) and PC2 (multi-coloured). Error bars show  $2\sigma$  variation. Best linear fit is shown as black dashed line, PC2  $R^2=0.997$ .



**Figure 5.** a-c) 2D-IR spectra used in the linear combination analysis method (iii): a) BSA in H<sub>2</sub>O, b)  $\gamma$ -globulins in H<sub>2</sub>O and c) Gly in H<sub>2</sub>O. Spectra a) and b) were scaled to their amide I peak and c) was scaled to the largest peak at 1543 cm<sup>-1</sup>. All spectra are plotted on the same scale, see colour bar. d) Result of linear combination analysis, showing variation in the amount of each spectra required to recreate the 2D-IR spectrum of each <sup>x</sup>[Gly] sample. Both BSA (light blue) and  $\gamma$ -globulins (blue) show almost constant contributions while Gly (red) increases with increasing Gly concentration. e) Expanded region of d) showing Gly coefficient only. Black circles are results of leave one out cross-validation tests (see text). Grey dashed line shows linear fit with R<sup>2</sup>=0.942. Error bars show 1 $\sigma$  variation from triplicate measurement.

Referring to the IR absorption spectrum (Fig.1(a)), the presence of a small peak near this frequency in the serum spectrum is present. Another possible explanation for this is a small contribution to PC1 from the nearby Gly band at 1415 cm<sup>-1</sup>. However, the absence of any other Gly contributions to the PC1 spectrum alongside the fact that the 1415 cm<sup>-1</sup> diagonal peak (and corresponding cross peaks) is also present in the serum 2D-IR spectrum at very low amplitudes (ESI Fig.S1) suggests that it has been enhanced via the removal of noise by the PCA process.

A series of leave-one-out cross validation tests were performed on the data. In this process each sample (of which there are 27, 3 for each concentration) was removed from the dataset and the remaining data used to re-calculate the PCA. Using the linear correlation of PC2 with [Gly], the concentration of the omitted sample was then predicted. This technique predicted all

concentrations of Gly larger than <sup>157</sup>[Gly] (inclusive) accurately to within an error of 1 $\sigma$  from the original triplicate measurement, however lower concentrations of Gly were not predicted accurately. This reinforces the result from i) regarding the current detection 3 mg/mL limit for Gly using 2D-IR in this manner.

It is instructive to compare these PCA results with data obtained using IR absorption spectroscopy. A similar approach, applying PCA results to IR absorption spectra of the same samples (Fig.S2) showed that, although the PC2 yields a strong linear correlation (R<sup>2</sup>=0.996), the spectral components associated with PC1 and PC2 do not show the same agreement with the spectra of serum and Gly as was observed for the 2D-IR dataset. Thus, IR absorption spectroscopy is unable to differentiate PC1 and PC2 cleanly into serum and Gly contributions. The reason for this is likely to be related to the fact that IR absorption measurements still contain strong contributions from the water spectrum and lack the extra information arising from off-diagonal peaks.

While the off-diagonal 2D-IR peak amplitudes are less intense than the diagonal peaks, their amplitudes displayed linear correlations with Gly concentration. In the case of some of the smaller peaks they become obscured by the noise at lower concentrations. However the results of PCA clearly identify a concentration dependence of the of the cross peaks (Fig 4(b,e) associated with the Gly contribution.

iii) The third analysis method applied a linear model to fit each measured <sup>x</sup>[Gly] spectrum to the sum of component 2D-IR spectra in an attempt to separate and quantify contributions to

the spectra. In the first iteration, a linear combination of 2D-IR spectra of neat serum ( $^1$ [Gly]) and the spectrum of Gly in buffer was used (Fig.S3), using the full spectral region. The results were similar to those from PCA, though a slightly improved  $R^2$  value of 0.998 was obtained. To determine the importance of the off-diagonal peaks in the analysis of serum spectra, the linear combination approach was repeated in a modified fashion to focus on the amide I region of the spectrum. The second iteration of the method employed fitting the  $^x$ [Gly] spectra to the sum of 2D-IR spectra of BSA,  $\gamma$ -globulins and Gly, all measured in buffer solution in the spectral range 1550-1700  $\text{cm}^{-1}$ .

The component spectra used in the linear combination are shown in Fig.5(a-c). The 2D-IR diagonal peak due to the amide I band of BSA appears at (pump = probe) 1656  $\text{cm}^{-1}$ , consistent with a secondary structure that is dominated by  $\alpha$ -helices. The amide I diagonal contribution from the  $\gamma$ -globulins appears at a slightly lower wavenumber (1639  $\text{cm}^{-1}$ ) due to the predominance of  $\beta$ -sheet structures. The spectrum of Gly in the same region is distinct from those of the large proteins in that it features a diagonal peak at 1650  $\text{cm}^{-1}$  and a prominent off diagonal peak to a weak diagonal peak near 1600  $\text{cm}^{-1}$ , the latter is visible as a shoulder in the IR absorption spectrum of the  $^1$ [Gly] sample (Fig.1(a)).

Applying this linear combination to the  $^x$ [Gly] spectra shows (Figs.5(d,e)) that the contributions of BSA (Fig.5(d), light blue) and the  $\gamma$ -globulins (Fig.5(d), dark blue) remain approximately constant across the range of  $^x$ [Gly] samples studied, while that of Gly increased in a linear fashion ( $R^2=0.942$ , Fig.5(d), red). It is clear that, although this approach was successful, based on the  $R^2$  value removing the 2D network of Gly peaks from the analysis lowers the quality of the quantitative analysis of the serum spectra. However, this does show that 2D-IR has the potential not only to identify a low molecular weight contributor to the serum spectrum, but also to deliver a more nuanced analysis of the protein content, based on the amide I region.

The outcome of analysis methods i) and ii) above imply a detection limit for Gly of  $\sim 3$  mg/mL using 2D-IR spectroscopy. This value is many orders of magnitude larger than the physiological levels of Gly itself. However, these initial results serve to provide a useful indicator of the capability 2D-IR has for biomedical applications. It is instructive to compare results from 2D-IR with those obtained using IR absorption methods. For example, the evaluation of detection limits using ATR-FTIR spectroscopy have been reported.<sup>22</sup> Studies of Gly detection using wet and dry serum samples were carried out and compared with serum samples following depletion of the high molecular weight fraction by centrifugal filtration. Analysis of wet serum allowed linear regression models to be built, permitting prediction of a 0.5 mg/mL glycine sample in serum to within an accuracy of 10% ( $0.45 \pm 0.16$  mg/mL). Removal of the high molecular weight fraction allowed quantification of Gly levels higher than 0.01 mg/mL, 50 times lower than that of whole serum. This suggests that the 2D-IR detection limit is slightly higher than ATR methods though, in the absence of any sample filtration, the two methods are comparable in terms of the order of magnitude of the detection limits and neither is capable of detecting Gly at physiological levels.<sup>22,40</sup> ATR-FTIR was unable to discriminate between different components of the amide I band as 2D-IR has been shown to be able to do.

Overall, our results provide a useful starting point from which the 2D-IR technique can improve further. Detection of single amide units in proteins and peptides have been demonstrated at

concentrations slightly lower than we report for glycine in serum.<sup>41</sup> These measurements have however employed deuterated solvents, isotopic labels or have probed other vibrational modes of the molecules in question, in contrast to the current measurements in a complex aqueous serum sample. This does indicate that, with further development there is potential for detection limits in serum to be improved. The 2D-IR instrument used for this work has seen significant improvements since the time of this data collection and it is possible that the introduction of a pulse shaper to generate the pump pulse pair as well as improved referencing will increase the 2D-IR spectrometer's sensitivity. Specifically, the study demonstrates that 2D-IR spectroscopy of serum can be employed to identify the presence of low molecular weight components in addition to the major protein constituents of the serum matrix. Furthermore, it is plausible that 2D-IR methods can be used to break down the total protein content into smaller contributions, taking the first steps along the road to producing a broad biomolecular fingerprint from a single optical measurement. Typically in disease states, associated changes in serum protein levels can involve more than one peptide/protein. With further investigation 2D-IR coupled with PCA has the potential to produce, in an ideal scenario, different PCs each corresponding to the peptide/protein causing changes from that of a normal/healthy serum sample. With PCA being an unsupervised method, the obtained principal component loading spectra would have to be cross referenced against a library of 2D-IR of peptide/protein spectra for identification, however for multi-peptide analysis the use of more sophisticated and supervised techniques may prove superior. The fact that 2D-IR spectra are acquired using as-received serum samples and that individual spectra can be obtained in similar amounts of time to an FTIR spectrum mean that there is potential for 2D-IR to act as a complementary spectral analysis tool.

The intention behind using Gly as a model low molecular weight peptide or protein system was that it provides a means of estimating concentrations of larger species which could be detected. Gly has a MW of 75 Da and 3 mg/mL corresponds to a concentration of  $\sim 40$  mM. By inference, assuming a similar 2D-IR signal size per residue, a 200-residue protein would, based on the reported Gly results, be detectable at concentrations of  $\sim 200$   $\mu\text{M}$ . It should be noted however, that the magnitude of the amide I band of a protein in a 2D-IR spectrum is increased on a per residue basis by vibrational coupling within the protein secondary structure.<sup>42,43</sup> This effect does not apply to IR absorption spectroscopy, meaning that the 200  $\mu\text{M}$  estimate is an upper sensitivity limit, based on a single amino acid with no secondary structure amplification of the 2D-IR signal, meaning that the actual detection level may well be lower. This can be substantiated further by comparison with the 2D-IR spectrum of BSA (Fig.5(a)), which indicates a detection limit, based on the spectral noise of  $\sim 1$  mg/mL, indicating that intramolecular coupling can be expected to play a significant role in determining the detection limits for proteins using 2D-IR methods.

Assuming a detection range of 1-3 mg/mL in this way enables a survey of likely problems to which 2D-IR may be applied by considering proteinaceous components of serum that either are found at levels greater than 1-3 mg/mL or which show fluctuations at this level in response to a medical condition. Examples include the presence of nephrosis (albumin,  $\alpha$ -macroglobulin, transferrin,  $\beta$ -lipoprotein); neoplasms ( $\alpha$ 1-acid glycoprotein, transferrin); liver disease (haptoglobin,  $\alpha$ -macroglobulin, IgG,

$\alpha$ -lipoprotein) and the presence of inflammatory conditions ( $\alpha$ 1-acid glycoprotein,  $\alpha$ 1-antitrypsin, haptoglobin).<sup>4</sup>

A key finding of this work is that the off-diagonal region of the 2D-IR spectrum has a significant role to play in differentiating protein species. Although the Gly content could be observed using just the amide I region (as in iii, above) the sensitivity and specificity was improved significantly when the off-diagonal spectral signature of the molecule was taken into the analysis. It is interesting to note that the 2D-IR spectrum of neat serum (Fig. 3(c)) shows considerable off-diagonal structure that has so far not been fully characterised. This offers the scope for a library of 2D spectra of significant proteins, such as those identified above, alongside linear combination fitting, to be used to create a 2D-IR-based protein analysis panel for a serum sample. This will require further development of the 2D-IR methodology, including sample handling and measurement protocols<sup>39</sup> in order to produce methods capable of determining absolute, rather than relative concentrations, but on the basis of this model protein study, the potential for 2D-IR to contribute in a meaningful way seems realistic.

## CONCLUSION

Using Gly as a model system, we have evaluated the ability of ultrafast 2D-IR spectroscopy to detect and quantify the low-molecular weight proteinaceous components of blood serum. The characteristic network of peaks appearing in the off-diagonal region of the 2D-IR spectrum of Gly allowed spectroscopic separation from that of the dominant protein fraction in a transmission-mode 2D-IR measurement without prior filtration or sample drying. 2D-IR spectra of blood serum samples supplemented with a range of concentrations of Gly were obtained and analysed using three different approaches including the magnitude of diagonal peaks, PCA and spectral fitting to a linear combination of component spectra. The results indicated a detection limit of  $\sim 3$  mg/mL for Gly. Using this result as the basis for a critical assessment of the sensitivity of 2D-IR for measuring the concentrations of amino acids, peptides and proteins present in serum samples showed that, in the case of several clinically relevant diagnostic molecules and their combinations, the potential exists for 2D-IR to complement IR absorption methods because the benefits of the second frequency dimension offered by 2D-IR spectroscopy outweigh the added technical complexity of the measurement.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.

FTIR analysis and further LCA analysis. (doc)

## AUTHOR INFORMATION

### Corresponding Author

\* Samantha H. Rutherford<sup>†</sup>

Email: samantha.hume-rutherford@strath.ac.uk

### Present Addresses

<sup>†</sup>WestCHEM, Department of Pure and Applied Chemistry, University of Strathclyde, Technology and Innovation Centre, 99 George Street, Glasgow, G1 1RD, UK

## Author Contributions

The manuscript was written through contributions of all authors

## ACKNOWLEDGMENT

Funding is gratefully acknowledged from EPSRC (EP/T014318/1 and EP/T014245/1) and STFC for programme access to Central Laser Facility systems.

## REFERENCES

- (1) Baker, M. J.; Sockalingum, G. D.; Hughes, C.; Lukaszewski, R. A. Developing and Understanding Biofluid Vibrational Spectroscopy: A Critical Review. *Chem. Soc. Rev.* **2016**, *45*, 1803–1818.
- (2) Petricoin, E. F.; Belluco, C.; Araujo, R. P.; Liotta, L. A. The Blood Peptidome: A Higher Dimension of Information Content for Cancer Biomarker Discovery. *Nat. Rev. Cancer* **2006**, *6*, 961–967.
- (3) Hu, S.; Loo, J.; Wong, D. Human Body Fluid Proteome Analysis. *Proteomics*. 2006, pp 6326–6353.
- (4) Putnam, F. W. *The Plasma Proteins - Structure, Function and Genetic Control*, 2nd Editio.; Academic Press: London, 1975.
- (5) Katie Crosley, L.; Duthie, S. J.; Polley, A. C.; Bouwman, F. G.; Heim, C.; Mulholland, F.; Horgan, G.; Johnson, I. T.; Mariman, E. C.; Elliott, R. M.; Daniel, H.; De Roos, B. Variation in Protein Levels Obtained from Human Blood Cells and Biofluids for Platelet, Peripheral Blood Mononuclear Cell, Plasma, Urine and Saliva Proteomics. *Genes Nutr.* **2009**, *4*, 95–102.
- (6) Busher, J. T. Serum Albumin and Globulins. In *Clinical Methods: The History, Physical, and Laboratory Examinations.*; Hurst, J. ., Walker, H. ., Hall, W. ., Eds.; Butterworths: Boston, 1990; pp 497–500.
- (7) Tirumalai, R. S.; Chan, K. C.; Prieto, D. A.; Issaq, H. J.; Conrads, T. P.; Veenstra, T. D. Characterization of the Low Molecular Weight Human Serum Proteome. *Mol. Cell. Proteomics* **2003**, *2*, 1096–1103.
- (8) Naseer, K.; Ali, S.; Qazi, J. ATR-FTIR Spectroscopy as the Future of Diagnostics: A Systematic Review of the Approach Using Bio-Fluids. *Appl. Spectrosc. Rev.* **2020**, *55*, 1–13.
- (9) Isensee, K.; Kröger-Lui, N.; Petrich, W. Biomedical Applications of Mid-Infrared Quantum Cascade Lasers—a Review. *Analyst* **2018**, *143*, 5888–5911.
- (10) Perez-Guaita, D.; Heraud, P.; Marzec, K. M.; De La Guardia, M.; Kiupel, M.; Wood, B. R. Comparison of Transfection and Transmission FTIR Imaging Measurements Performed on Differentially Fixed Tissue Sections. *Analyst* **2015**, *140*, 2376–2382.
- (11) Bonnier, F.; Baker, M. J.; Byrne, H. J. Improved Protocols for Vibrational Spectroscopic Analysis of Body Fluids. *J. Biophotonics* **2014**, *7*, 167–179.
- (12) Butler, H. J.; Brennan, P. M.; Cameron, J. M.; Finlayson, D.; Hegarty, M. G.; Jenkinson, M. D.; Palmer, D. S.; Smith, B. R.; Baker, M. J. Development of High-Throughput ATR-FTIR Technology for Rapid Triage of Brain Cancer. *Nat. Commun.* **2019**, *10*, 4501.
- (13) Sheng, D.; Wu, Y.; Wang, X.; Huang, D.; Chen, X.; Liu, X. Comparison of Serum from Gastric Cancer Patients and from Healthy Persons Using FTIR Spectroscopy. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2013**, *116*, 365–369.
- (14) Backhaus, J.; Mueller, R.; Formanski, N.; Szlama, N.; Meerpohl, H.-G.; Eidt, M.; Bugert, P. Vibrational Spectroscopy Diagnosis of Breast Cancer with Infrared Spectroscopy from Serum Samples. *Vib. Spectrosc.* **2010**, *52*, 173–177.
- (15) Cameron, J. M.; Butler, H. J.; Smith, B. R.; Hegarty, M. G.; Jenkinson, M. D.; Syed, K.; Brennan, P. M.; Ashton, K.; Dawson, T.; Palmer, D. S.; Baker, M. J. Developing Infrared Spectroscopic Detection for Stratifying Brain Tumour Patients: Glioblastoma Multiforme vs . Lymphoma. *Analyst* **2019**, *144*,
- (16) Hands, J. R.; Clemens, G.; Stables, R.; Ashton, K.; Brodbelt, A.; Davis, C.; Dawson, T. P.; Jenkinson, M. D.; Lea, R. W.; Walker, C.; Baker, M. J. Brain Tumour Differentiation: Rapid Stratified Serum Diagnostics via Attenuated Total Reflection Fourier-Transform Infrared Spectroscopy. *J. Neurooncol.* **2016**, *127*,

- 463–472.
- (17) Socha, E.; Koba, M.; Kośliński, P. Amino Acid Profiling as a Method of Discovering Biomarkers for Diagnosis of Neurodegenerative Diseases. *Amino Acids* **2019**, *51*, 367–371.
- (18) Li, N.; Liu, W.; Li, W.; Li, S.; Chen, X.; Bi, K.; He, P. Plasma Metabolic Profiling of Alzheimer's Disease by Liquid Chromatography/Mass Spectrometry. *Clin. Biochem.* **2010**, *43*, 992–997.
- (19) Galli, S.; Naranjo, A.; Van Ryn, C.; Tilan, J. U.; Trinh, E.; Yang, C.; Tsuei, J.; Hong, S. H.; Wang, H.; Izycka-Swieszewska, E.; Lee, Y. C.; Rodriguez, O. C.; Albanese, C.; Kitlinska, J. Neuropeptide Y as a Biomarker and Therapeutic Target for Neuroblastoma. *Am. J. Pathol.* **2016**, *186*, 3040–3053.
- (20) Rascher, W.; Kremens, B.; Wagner, S.; Feth, F.; Hunneman, D. H.; Lang, R. E. Serial Measurements of Neuropeptide Y in Plasma for Monitoring Neuroblastoma in Children. *J. Pediatr.* **1993**, *122*, 914–916.
- (21) Bonnier, F.; Blasco, H.; Wasselet, C.; Brachet, G.; Respaud, R.; Carvalho, L. F. C. S.; Bertrand, D.; Baker, M. J.; Byrne, H. J.; Chourpa, I. Ultra-Filtration of Human Serum for Improved Quantitative Analysis of Low Molecular Weight Biomarkers Using ATR-IR Spectroscopy. *Analyst* **2017**, *142*, 1285–1298.
- (22) Bonnier, F.; Brachet, G.; Duong, R.; Sojinrin, T.; Respaud, R.; Aubrey, N.; Baker, M. J.; Byrne, H. J.; Chourpa, I. Screening the Low Molecular Weight Fraction of Human Serum Using ATR-IR Spectroscopy. *J. Biophotonics* **2016**, *9*, 1085–1097.
- (23) Adkins, J. N.; Varnum, S. M.; Auberry, K. J.; Moore, R. J.; Angell, N. H.; Smith, R. D.; Springer, D. L.; Pounds, J. G. Toward a Human Blood Serum Proteome. *Mol. Cell. Proteomics* **2002**, *1*, 947–955.
- (24) Hume, S.; Hithell, G.; Greetham, G. M.; Donaldson, P. M.; Towrie, M.; Parker, A. W.; Baker, J.; Hunt, N. T. Measuring Proteins in H<sub>2</sub>O with 2D-IR Spectroscopy. *Chem. Sci.* **2019**, *10*, 6448–6456.
- (25) Donaldson, P. M.; Greetham, G. M.; Shaw, D. J.; Parker, A. W.; Towrie, M. A 100 KHz Pulse Shaping 2D-IR Spectrometer Based on Dual Yb:KGW Amplifiers. *J. Phys. Chem. A* **2018**, *122*, 780–787.
- (26) Shim, S. H.; Strasfeld, D. B.; Ling, Y. L.; Zanni, M. T. Automated 2D IR Spectroscopy Using a Mid-IR Pulse Shaper and Application of This Technology to the Human Islet Amyloid Polypeptide. *Proc. Natl. Acad. Sci.* **2007**, *104*, 14197–14202.
- (27) Luther, B. M.; Tracy, K. M.; Gerrity, M.; Brown, S.; Krummel, A. T. 2D IR Spectroscopy at 100 KHz Utilizing a Mid-IR OPCPA Laser Source. *Opt. Express* **2016**, *24*, 4117.
- (28) Fritzsche, R.; Donaldson, P. M.; Greetham, G. M.; Towrie, M.; Parker, A. W.; Baker, M. J.; Hunt, N. T. Rapid Screening of DNA-Ligand Complexes via 2D-IR Spectroscopy and ANOVA-PCA. *Anal. Chem.* **2018**, *90*, 2732–2740.
- (29) Tracy, K. M.; Barich, M. V.; Carver, C. L.; Luther, B. M.; Krummel, A. T. High-Throughput Two-Dimensional Infrared (2D IR) Spectroscopy Achieved by Interfacing Microfluidic Technology with a High Repetition Rate 2D IR Spectrometer. *J. Phys. Chem. Lett.* **2016**, *7*, 4865–4870.
- (30) Li, X.; Sun, L.; Zhang, W.; Li, H.; Wang, S.; Mu, H.; Zhou, Q.; Zhang, Y.; Tang, Y.; Wang, Y.; Chen, W.; Yang, R.; Dong, J. Association of Serum Glycine Levels with Metabolic Syndrome in an Elderly Chinese Population. *Nutr. Metab.* **2018**, *15*, 1–9.
- (31) NHS Foundation Trust. Tests and Investigations - Amino Acids <https://www.gloshospitals.nhs.uk/our-services/services-we-offer/pathology/tests-and-investigations/amino-acids/> (accessed Nov 16, 2020).
- (32) de Vries, A. De; Alexander, B.; Quamo, Y. Studies on Amino Acid Metabolism. II. Blood Glycine and Total Amino Acids in Various Pathological Conditions, with Observations on the Effects of Intravenously Administered Glycine. *J. Clin. Invest.* **1948**, *27*, 655–664.
- (33) Greetham, G. M.; Burgos, P.; Cao, Q.; Clark, I. A. N. P.; Codd, P. S.; Farrow, R. C.; George, M. W.; Matousek, P.; Parker, A. W.; Pollard, M. R.; Robinson, A.; Xin, Z.; Towrie, M. ULTRA: A Unique Instrument for Time-Resolved Spectroscopy. *Appl. Spectrosc.* **2010**, *64*, 1311–1319.
- (34) Deflores, L. P.; Nicodemus, R. A.; Tokmakoff, A. Two-Dimensional Fourier Transform Spectroscopy in the Pump – Probe Geometry. *Opt. Lett.* **2007**, *32*, 2966–2968.
- (35) Hunt, N. T. 2D-IR Spectroscopy: Ultrafast Insights into Biomolecule Structure and Function. *Chem. Soc. Rev.* **2009**, *38*, 1837–1848.
- (36) Team, R. D. C. R: A Language and Environment for Statistical Computing. *R Found. Stat. Comput.* **2010**.
- (37) Lutz, O. M. D.; Messner, C. B.; Hofer, T. S.; Canaval, L. R.; Bonn, G. K.; Huck, C. W. Computational Vibrational Spectroscopy of Glycine in Aqueous Solution – Fundamental Considerations towards Feasible Methodologies. *Chem. Phys.* **2014**, *435*, 21–28.
- (38) Deflores, L. P.; Ganim, Z.; Nicodemus, R. A.; Tokmakoff, A. Amide I-II' 2D IR Spectroscopy Provides Enhanced Protein Secondary Structural Sensitivity. *J. Am. Chem. Soc.* **2009**, *131*, 3385–3391.
- (39) Hume, S.; Greetham, G. M.; Donaldson, P. M.; Towrie, M.; Parker, A. W.; Baker, M. J.; Hunt, N. T. 2D-Infrared Spectroscopy of Proteins in Water: Using the Solvent Thermal Response as an Internal Standard. *Anal. Chem.* **2020**, *92*, 3463–3469.
- (40) Yang, H.; Yang, S.; Kong, J.; Dong, A.; Yu, S. Obtaining Information about Protein Secondary Structures in Aqueous Solution Using Fourier Transform IR Spectroscopy. *Nat. Protoc.* **2015**, *10*, 382–396.
- (41) Kim, Y. S.; Hochstrasser, R. M. Applications of 2D IR Spectroscopy to Peptides, Proteins, and Hydrogen-Bond Dynamics. *J. Phys. Chem. B* **2009**, *113*, 8231–8251.
- (42) Grechko, M.; Zanni, M. T. Quantification of Transition Dipole Strengths Using 1D and 2D Spectroscopy for the Identification of Molecular Structures via Exciton Delocalization: Application to Alpha-Helices. *J. Chem. Phys.* **2012**, *137*, 184202.
- (43) Minnes, L.; Shaw, D. J.; Cossins, B. P.; Donaldson, P. M.; Greetham, G. M.; Towrie, M.; Parker, A. W.; Baker, M. J.; Henry, A. J.; Taylor, R. J.; Hunt, N. T. Quantifying Secondary Structure Changes in Calmodulin Using 2D-IR Spectroscopy. *Anal. Chem.* **2017**, *89*, 10898–10906.

---

For Table of Contents only

