Electronic Supplementary Material (ESI) for Chemical Science.
This journal is © The Royal Society of Chemistry 2019

Measuring proteins in H_2O with 2D-IR spectroscopy

Samantha Hume,[1] Gordon Hithell,[1] Gregory M. Greetham,[2] Paul M. Donaldson,[2] Michael Towrie,[2] Anthony W. Parker,[2] Matthew J. Baker,[3] Neil T. Hunt,[4,*]

[1] Department of Physics, University of Strathclyde, SUPA, 107 Rottenrow East, Glasgow, G4 0NG, UK

[2] STFC Central Laser Facility, Research Complex at Harwell, Rutherford Appleton Laboratory, Harwell Campus, Didcot, OX11 0QX, UK

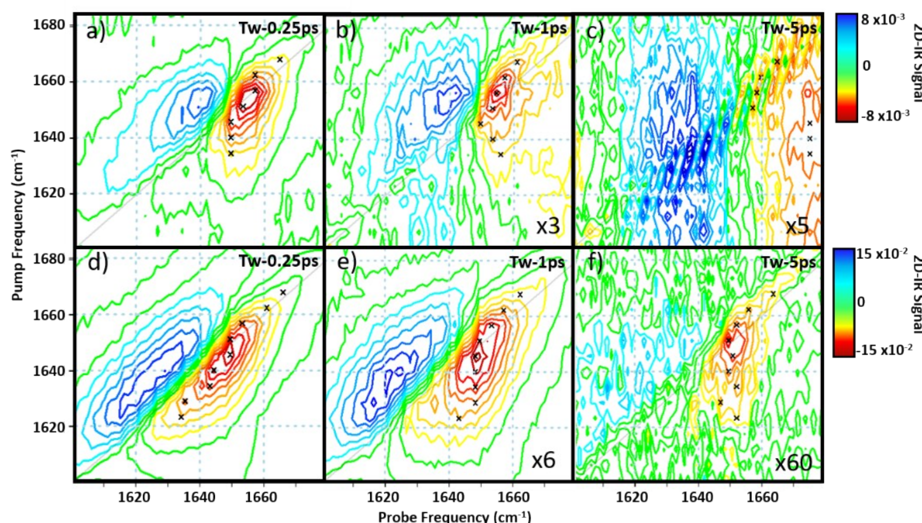
[3] WestCHEM, Department of Pure and Applied Chemistry, University of Strathclyde, Technology and Innovation Centre, 99 George Street, Glasgow, G1 1RD, UK

[4] Department of Chemistry and York Biomedical Research Institute, University of York, Heslington, York, YO10 5DD, UK

Supporting Information

Comparison of Lineshapes and Spectral Diffusion of Protein Amide I in H_2O and D_2O:

A series of waiting times studies for serum albumin (30 mg/mL) in both H_2O (~2.75 µm spacer) and D_2O (25 µm spacer) solvents were carried out to compare the effect of H/D exchange on spectral diffusion dynamics. The centre line slope method (CLS) [1] was utilised to extract these dynamics. Although the amide I lineshape in H_2O appears more circular in nature (Fig S1(a-c), suggesting different inhomogeneous broadening or altered spectral diffusion dynamics, the CLS analysis did not clearly reflect this trend. At late waiting times (Fig.S1(c)) the water



response due to heating dominates the 2D-IR signal and thus obscures the protein signal.

Figure S1. *2D-IR spectra of serum albumin in H_2O (a-c) and D_2O (d-f) at the waiting times 0.25 ps (a,d), 1 ps (b, e) and 5 ps (c,f). All H_2O spectra have been plotted on the same scale and the later waiting times have been magnified to reflect this. All D_2O spectra have also been plotted on the same scale and magnified as shown. Black crosses denote the CLS of each spectra.*

2D-IR Data Analysis and Measurement of AGR:

All 2D-IR spectral processing and analysis was carried out using a custom made script on the statistical analysis software programme, R.[2] Prior to the analysis described in the text to obtain the AGR values, a 2nd order polynomial baseline subtraction was performed

Three methods were used to obtain the AGR from the 2D spectra: i) the 2D-IR spectrum diagonal, ii) pump-frequency slices and iii) linear combination of 2D-IR spectra.

i) 2D-IR spectrum diagonal: the diagonal of each 2D spectrum was extracted, showing two distinct peaks at 1656 cm^{-1} and 1639 cm^{-1}, which were assigned to the albumin and globulin fractions respectively. The ratio of the absolute values of the peak amplitudes of these features was used to determine the AGR following scaling of the globulin amplitude by a factor of 1.8 (see Fig.S3 below).

ii) The pump-slice method utilised slices through the 2D-IR spectrum at 1656 cm^{-1} and 1639 cm^{-1}, assigned to peaks of the albumin and globulin signals respectively. The ratio of the absolute values of the amplitude of the globulin pump slice, at a probe frequency of 1639 cm^{-1}, and that of the albumin slice at 1656 cm^{-1} was used to determine the AGR following application of the scaling factor (1.8) to the globulin signal.

iii) Linear combination: All 2D-IR spectra were normalised to the albumin peak at 1656 cm^{-1}. Linear combination analysis fitted the serum 2D-IR spectrum to the linear sum of the independent 2D-IR spectra of albumin and globulins. The coefficients for the relative contribution of the two protein spectra were then used to give the AGR, following scaling of the globulin fraction by 1.8.

The AGR values were measured from the 2D-IR spectra of the spiked serum samples using these approaches while the variation over triplicate repeat measurements was used to reflect the repeatability of each method (Fig.5(a,c,e)). These are compared to the real value of the AGR (Fig.5(a,c,e), solid black line), as determined by traditional analysis of the neat serum sample along with the added γ-globulin spike. The non-spectroscopic AGR measurement of the neat serum was repeated on three individual batches of serum to ensure repeatability.

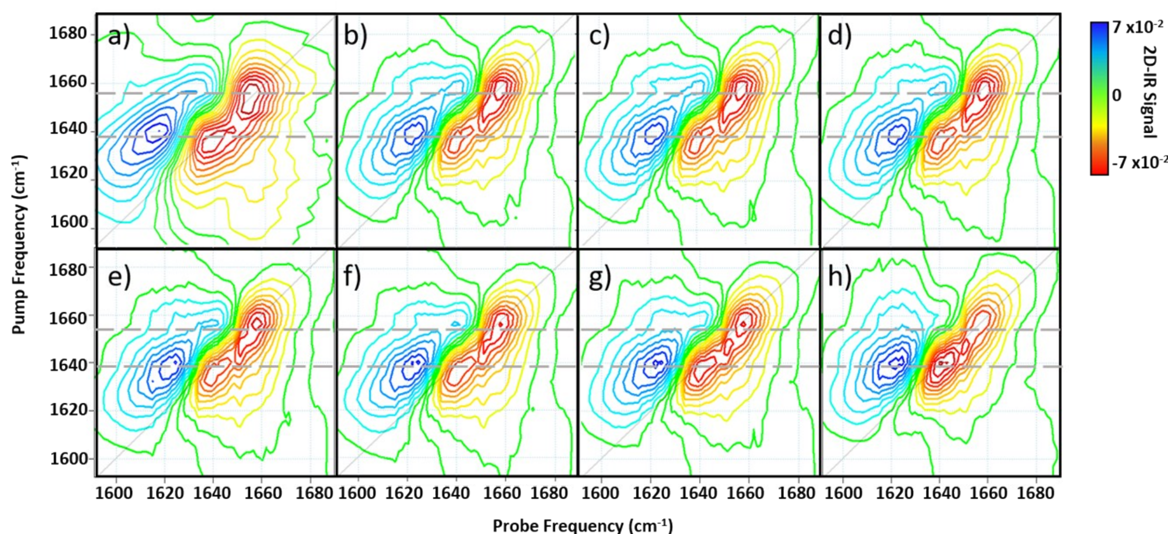


Figure S2. *2D-IR spectra of serum spiked with γ-globulins in the following concentrations (mg/mL): a) 0 b) 0.5 c) 0.9 d) 1.9 e) 3.8 f) 7.5 g) 15 h) 30. Dashed grey horizontal lines show how the peak positions of albumin and γ-globulins discussed in the text. All spectra have been plotted on the same scale, see colour bar.*

Scaling Factor:

Having assigned the two peaks on the 2D-IR spectrum diagonal to albumin and γ-globulins, the relative concentrations of albumin and γ-globulins could be obtained. To do this, an indication of the relative 2D-IR signal amplitudes of albumin and the γ-globulins for a given concentration (mg/ml) was required. This is owing to the different molecular weights (albumin ~66kDa; γ-globulin average ~150kDa) and the fact that differing secondary structural compositions give rise to changes in vibrational coupling of amide I oscillators, which in turn influences the amplitude of the 2D-IR amide I band via the transition dipole moment. [3] The quantitative ratio of the serum albumin and γ-globulin 2D-IR signals were obtained by measurements of known concentrations of the two protein components under as close as possible to identical conditions (Fig.S3). From the relative maximum amplitudes of the 2D-IR spectrum diagonals of serum albumin and γ-globulins, we can establish that, per unit concentration (mg/mL), the albumin signal is 1.8 times larger (±0.1) than the γ-globulin response.

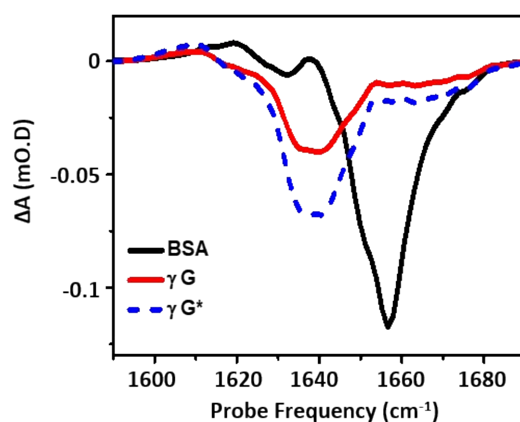


Figure S3. *Diagonal slices through 2D-IR spectra of 50 mg/mL of serum albumin (black), 30 mg/mL of γ-globulins (red). Blue dashed lines shows the slice due to the γ-globulins scaled to reflect a 50 mg/ml concentration*.

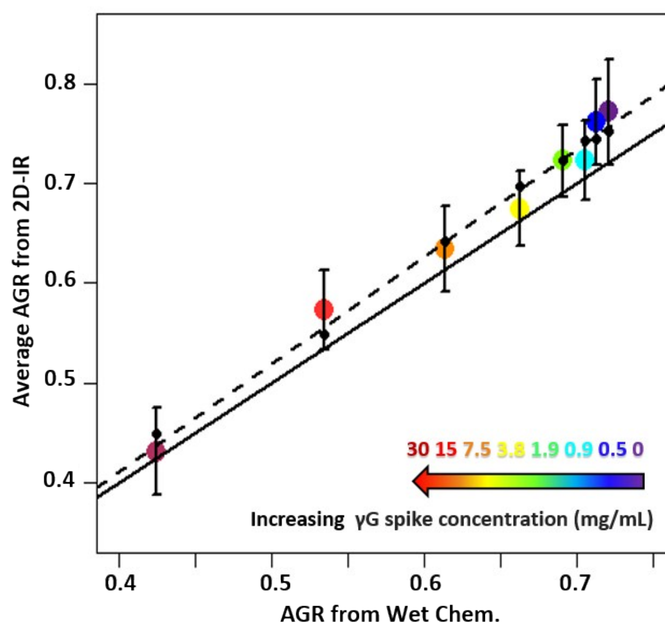


Figure S4. *Results of leave one out-type tests. Every data point (comprised of the average of the three analysis methods on the average of three measurements) is left out in turn and the linear fit is recalculated. The AGR is then predicted for each data point using the actual AGR and the linear model, as shown by the solid black dots.*

The solid black line indicates the actual AGR of the samples. Error bars show 2σ variation. The dashed line shows a linear fit to the experimental AGR values.

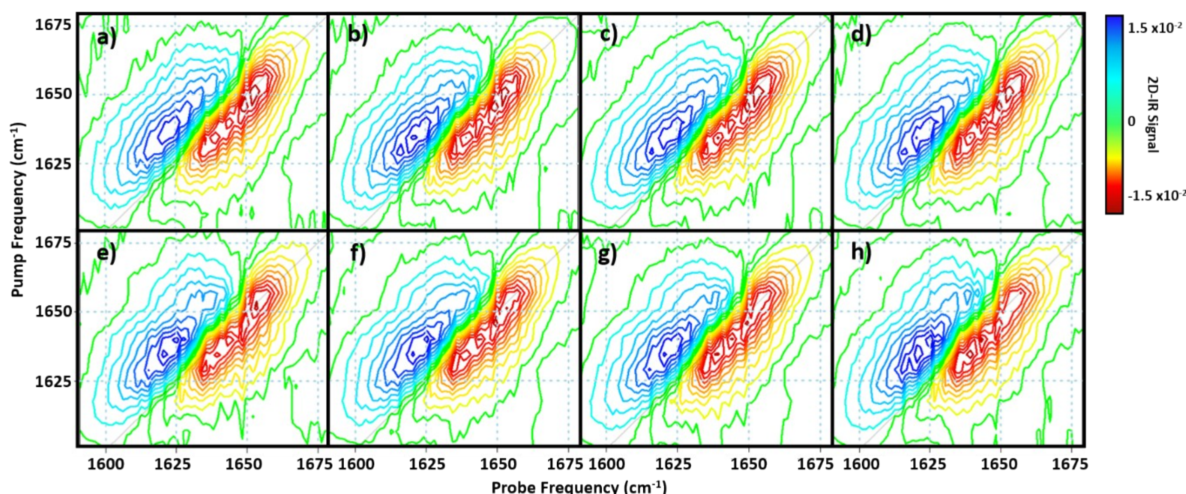


Figure S5. *2D-IR spectra of serum spiked with IgA in the following concentrations (mg/mL): a) 0 b) 0.25 c) 0.5 d) 1 e) 2.5 f) 5 g) 7.5 h) 15. All spectra have been plotted on the same scale, see colour bar.*

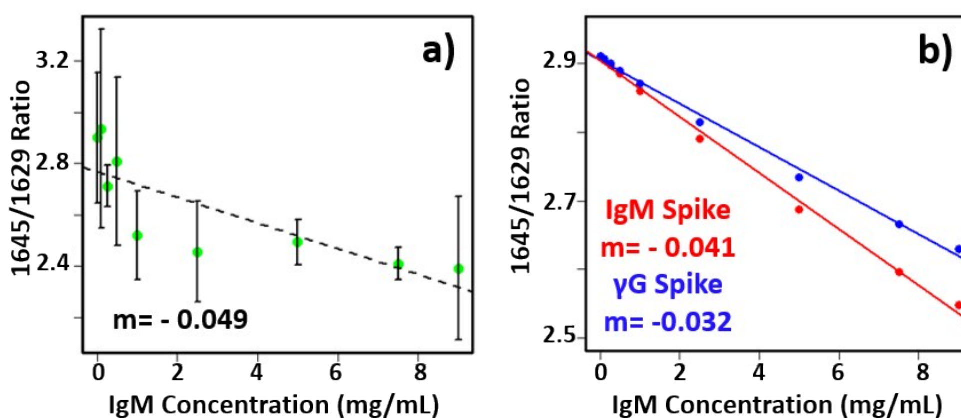


Figure S6. *a) The variation in 2D-IR signal at the peak of the difference spectral response for IgM-γ-globulin (1645:1629) as a function of added IgM. The linear increase in the negative signal shows a correlation with protein content. b) Results of the simulations with IgM or γ-globulins added to serum. Gradients of the straight line fits are shown.*

References:

1. Guo, Q.; Pagano, P.; Li, Y. L.; Kohen, A.; Cheatum, C. M., Line shape analysis of two-dimensional infrared spectra. *J Chem Phys* **2015,** *142* (21).

2. R Development Core Team *R: A language and environment for statistical computing*, R Foundation for Statistical Computing. Vienna, Austria: 2010.

3. Dunkelberger, E. B.; Grechko, M.; Zanni, M. T., Transition Dipoles from 1D and 2D Infrared Spectroscopy Help Reveal the Secondary Structures of Proteins: Application to Amyloids. *J Phys Chem B* **2015,** *119* (44), 14065-14075.