Studies of Biofluid Analysis using 2D-IR Spectroscopy

By Samantha M. Hume Rutherford

Department of Physics

University of Strathclyde

A thesis submitted in the requirement for the degree of Doctor of Philosophy

2020

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Signed: 

Date: 24-09-2020

Contents

List of Abbreviations

2D-IR	Two-Dimensional Infrared
AOM	Acousto-Optic Modulator
AGR	Albumin to Globulin Ratio
$AgGaS_2$	Silver Thiogallate
ATR	Attenuated Total Reflection
BBO	Barium Borate
BSA	Bovine Serum Albumin
CLS	Centre Line Slope
D_2O	Deuterium Oxide
DFG	Difference Frequency Generation
DFT	Density Functional Theory
DMSO	Dimethyl Sulfoxide
FID	Free Induction Decay
FTIR	Fourier Transform Infrared
Ge	Germanium
Gly	Glycine
H/D	Hydrogen/Deuterium
H_2O	Hydrogen Oxide
He:Ne	Helium:Neon
HMW	High Molecular Weight
Ibu	Ibuprofen
Ig'X'	Immunoglobulin 'X'
IR	Infrared
IRE	Internal Reflective Element
KBr	Potassium Bromide
LCA	Linear Combination Analysis
LMW	Low Molecular Weight
MCT	Mercury Cadmium Telluride
Nd:YAG	Neodynium doped Yttrium Aluminium Garnet
OPA	Optical Parametric Amplifier
PCA	Principal Component Analysis
Pmol	Paracetamol
PP	Pump Probe
PTFE	Polytetrafluoroethylene
QCL	Quantum Cascade Laser
RAL	Rutherford Appleton Laboratory
TEM	Transmission Electron Microscope
TDC	Transition Dipole Coupling
Ti:Sapph	Titanium Sapphire
Warf	Warfarin

Abstract

In this thesis, the application of two-dimensional infrared (2D-IR) spectroscopy is evaluated to provide a quantitative analysis of the protein content of blood serum. The foremost challenge is obtaining protein measurements in physiologically relevant solvents as the most informative infrared peak relating to protein studies, the amide I band, overlaps with the bending mode of H_2O, making label-free detection of the protein content in serum challenging. This project demonstrates that 2D-IR can surmount the major obstacle to serum protein analysis as the 2D-IR amide I signature of proteins is shown to dominate that of water. Furthermore the link between protein secondary structure and the 2D-IR amide I lineshape allows differentiation of protein signals in serum leading to clinically relevant measurements of the biomedically important proteins. Detection limits for 2D-IR are also established allowing projection of the sensitivity of 2D-IR for future applications.

Quantification of proteins is key for diagnosis and prognosis outcomes, however using 2D-IR, standardisation of measurement protocols needs to be addressed in order to achieve this. A new method is demonstrated for 2D-IR spectroscopy to internally normalise spectral signals, allowing normalisation of the protein response to the thermal response of water which is temporally separate from the protein signal, reducing the impact of measurement fluctuations on the data. Furthermore, normalisation of sample signals enables calibration curves of serum albumin to be produced allowing absolute protein concentrations to be obtained using 2D-IR spectroscopy.

The application of 2D-IR is explored further and attempts to detect drug-binding at clinically relevant levels to serum albumin are made. Changes in the secondary structure of albumin are detected upon drug-binding however the complex nature of serum makes assigning changes observed using 2D-IR challenging.

1. Introduction

1.1. Biofluids

A biofluid is a biological liquid that is either excreted (e.g. sweat), secreted (e.g. milk) or obtained via a needle (e.g. blood) from the body. They contain information regarding the human body that pertains to a patient's health, which can be extracted and utilised to help monitor health, aid in disease diagnosis, predict treatment outcomes and their studies have the potential to develop personalised medicine.[1–3]

For certain diseases e.g. cancer, tissue biopsies are often required in order to formulate a correct diagnosis however these procedures are invasive and can produce unappealing side effects for patients.[4] This promotes the need for a minimally invasive procedure to provide a diagnosis and to monitor disease progression.

Easily obtained, serum is a biofluid that can be routinely collected in the form of a venepuncture procedure which collects a vial of blood from patients and the removal of blood cells and clotting factors produces serum. Serum accesses virtually all parts of the body and major organs, providing data from the circulatory system which contains a range of potentially diagnostic chemical markers including the circulatory proteome and low molecular weight peptidome as well as lipids, sugars and nucleic acids.[1,5,6] Serum is primarily made up of water (~92%) and proteins (~7-8%)[7] of which it contains > 20,000.[8] The total protein concentration ranges from 60-80 mg/mL[9,10] where the serum albumin (35-50 mg/mL) and globulins (25-35 mg/mL) are the most abundant.[11]

Increases in the total protein concentration have been linked to breast cancer compared with healthy patients and decreased albumin-to-globulin ratios have been linked to increased mortality in cancer patients.[12–14] Often low levels of serum albumin in critical illness is associated with a poor outcome[15]. Studies have documented the association of low albumin concentration to cognitive decline and chronic liver disease[15–17] and chronic inflammations typically result in the production of high globulin levels.[12,18,19] Protein concentrations such as serum albumins, creatinine, immunoglobulins and prolactin are few of the many proteins routinely tested in UK hospitals.[20]

Current processes performed in clinics and hospitals throughout the UK for the analysis of serum proteins typically involve the use of biochemistry and electrophoresis techniques.[8] Immunoassays like the Biuret method is a common wet chemistry technique used for the detection of total protein in solution however like many other wet chemistries, this

technique requires extensive sample preparation, large volumes (1 mL) of the biofluid being assessed and can take up to several hours to achieve results. They also rely critically on the availability of specific antibodies for proteins of interest.[21,22]

Electrophoresis involves the separation of molecules present in a biofluid via migration through a matrix upon application of an electric field, separating the biofluid into various fractions based on their molecular weight and electric charges.[23] Electrophoresis techniques are also subject to interference with substances such as free lipids, drugs, haemoglobin and bilirubin which can affect the results.[24] Again this technique can often take several minutes to hours to obtain results, requires lengthy sample preparation and the accessibility of different support media depending on the type of test being performed.

In terms of developing a rival technique suitable for clinical environments, a robust, simple and rapid approach is desired and efforts have been made to develop infrared (IR) absorption spectroscopy techniques for the study of biofluids as IR absorption spectroscopy is especially sensitive to the secondary structure of proteins.[25,26]

1.2. Infrared Spectroscopy of Proteins

Infrared spectroscopy is a powerful tool used for structural determination of molecular compounds including proteins. The absorption of IR radiation excites vibrational modes of molecules, producing IR spectra with variations in intensity and frequency dependent on the vibration of their interatomic bonds, orientation and their environment.[27,28]

1.2.1. Molecular Vibrations

Molecular vibrations are a result of the balance between attractive and repulsive forces between two covalently bonded atoms, and can be described using the Morse oscillator.[23,29,30] This balance is characterised by a minimum in the potential energy, $V(r_0)$, of the system which occurs at an atomic spatial separation that is equal to the natural bond length, otherwise known as the equilibrium distance, r_0.

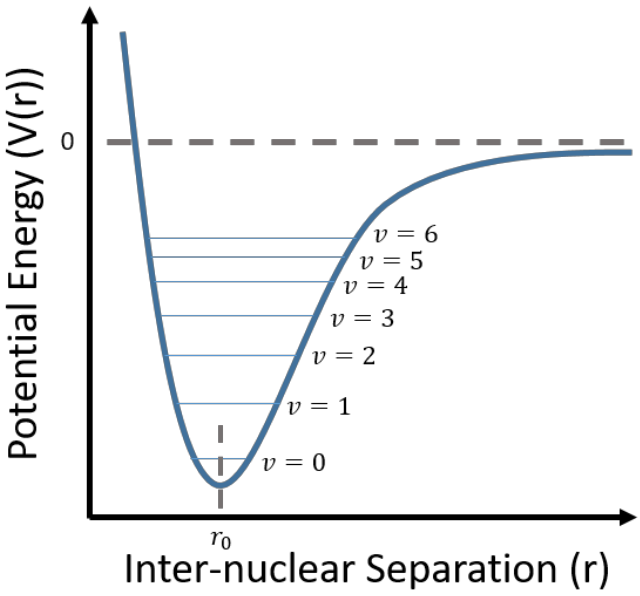

Figure 1.1: *Morse potential of two covalently bonded atomic nuclei. Allowed vibrational energy levels are shown.*

Displacements from this equilibrium position are opposed by a restoring force, similar to that of a mechanical spring, leading to oscillatory, vibrational motion. A Morse potential of a chemical bond between two atoms is shown in Figure 1.1. With increasing displacement from the equilibrium the bond can eventually be broken which is reflected as anharmonicity in the Morse potential. With increasing displacements from r_0 the potential energy between the two nuclei tends to 0, resulting in an anharmonic potential well.

Bond vibrations are described quantum-mechanically and in doing so only specific vibrational energy levels within the potential are permitted. The allowed potential energy of these discrete levels in the potential well (E_v) is described below:

$$E_v = \left(v + \frac{1}{2}\right) v_e - \left(v + \frac{1}{2}\right)^2 x_e v_e + \left(v + \frac{1}{2}\right)^3 y_e v_e + higher\ terms \qquad (1.1)$$

where v is the vibrational quantum number, v_e is the harmonic energy and x_e and y_e are the anharmonicity constants of the Morse potential. The ground vibrational state, where v = 0 has the non-zero energy of $\frac{1}{2}v_e$ as required by the Heisenberg uncertainty principle. The anharmonicity of vibrational modes is vital in understanding 2D-IR spectra. With each successive vibrational level the average bond length increases and so the energy required to

make the transition to the subsequent transition, $v = 1 - 2$, (and so on) is also lowered compared to the $v = 0 - 1$ (previous) transition.

The fundamental frequency ($v = 0 - 1$ transition) of a diatomic vibration can be described as function of the masses of the atoms that make up that bond, known as the reduced mass, μ, and the electron distribution across the bond. The fundamental vibrational frequency is:

$$\tilde{v} = \frac{1}{2\pi c}\sqrt{\frac{k}{\mu}} \tag{1.2}$$

where $\tilde{v}$, is the wavenumber (cm^{-1}) which is the frequency unit used in IR spectroscopy, c,is the speed of light and k, is the interatomic force constant. The IR vibration selection rules determine which transitions are allowed. For absorption of IR light to occur, the vibrational excitation must induce a change in the transitional dipole moment, μ, of the molecule:

$$\frac{\partial \mu}{\partial q} \neq 0 \tag{1.3}$$

where q is the displacement from the equilibrium state along the aligned axis of the molecule, these vibrations are known to be IR active.[31,32] For a harmonic potential, only transitions between adjacent energy levels, $\Delta v = \pm 1$, produce a non-zero transition dipole moment. However, using the better approximated anharmonic potential (Figure 1.1), higher order transitions ($\Delta v = \pm 2, \pm 3$, etc, known as overtones) can be accessed. In this regime, their transition dipole moment is now non-zero, although it is weak which results in low intensity peaks in comparison to the $v = 0 - 1$ transition under normal experimental conditions (e.g. room temperature).[28]

In general, bonds with a greater polarity have bigger dipole moment changes and result in a stronger absorption of IR light.[33] Stretching vibrations of completely symmetric bonds, e.g. N_2, do not result in a change of the transition dipole and therefore do not absorb IR light.

Molecular vibrations are also heavily influenced by their chemical environment including neighbouring functional groups and the surrounding solvent molecules.[28,34] This offers IR spectroscopy the ability to differentiate between different parts of one molecular structure, for example there are several vibrational modes of proteins however the most informative when utilising IR is the amide I mode (~ 1650 cm^{-1}) which is comprised mainly of the C=O and N-H bend. Proteins in a deuterated solvent absorb light at slightly lower frequencies than

those in H_2O owing to the heavier deuterium atoms creating different molecular distributions from the bound D atoms and fleeting solvent deuterium bonding, therefore red-shifting the fundamental frequencies.[35] Molecular bonds can thus be identified by the frequencies at which they absorb light allowing a molecular fingerprint of functional groups to be established.

1.2.2. Protein Structure

Proteins have four different categories of structure known as the primary, secondary, tertiary and quaternary, each contributing to the final complex three-dimensional arrangement. The building blocks of every protein are amino acids and it is the unique ordering of these that differentiate individual proteins. There are 20 naturally occurring amino acid species and each contains a carboxyl and amino group as well as an 'R_n' group (Figure 1.2) which denotes the amino acid functional group where n numbers the functional group along the chain.[36] Each amino acid is linked via a peptide (amide) bond between the carboxyl group of one amino acid and the amino group of another. This is known as a proteins primary structure and at this stage can be referred to as a polypeptide chain. The polypeptide chains can vary vastly in length depending on the nature and function of the proteins.[37] For example, human serum albumin is comprised of 585 amino acids,[38] it is a complex protein designed to take part in many functions throughout the body, such as transporting fatty acids, drugs and calcium, and regulating blood pressure and pH, in comparison to glucagon, whose sole function is to stimulate conversion of stored glycogen in the liver to glucose, has only 29 amino acids.[39]

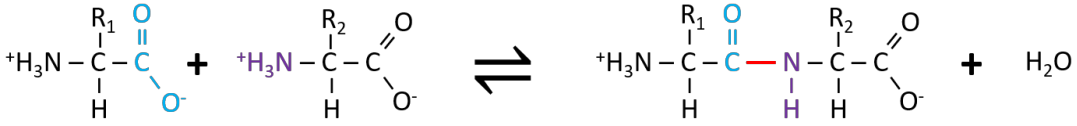

Figure 1.2: *Amide bond formation. Linking of the carboxyl group (blue) of one amino acid and the amino group (purple) of another results in an amide bond (red) and the production of one water molecule. The 'R' group denotes the side chain specific to each amino acid, of which there are 20.*

The secondary structure of a protein refers to the coiling or folding of this polypeptide chain. There are many types of secondary structure that are observed in proteins, however the two most common are known as the α-helix and the β-sheet (Figure 1.3).[37] The α-helix is typically a right handed helix, however left handed ones do occur but are rare in nature,[40] which is shaped like a coiled spring where every N-H group is hydrogen bonded to the C=O group of the amino acid 3 or 4 residues along the polypeptide chain, resulting in a helical formation (Figure 1.3(a)). It is sometimes referred to as the 3.6_{13}-helix, where 3.6 is the average number of residues per helical turn and the number of atoms in the hydrogen loop is 13. The α-helix is the most prevalent helical structure, however similar structures do occur, including the 3_{10}-helix and π-helix which contain 3 and 5 residues per turn, respectively.[41]

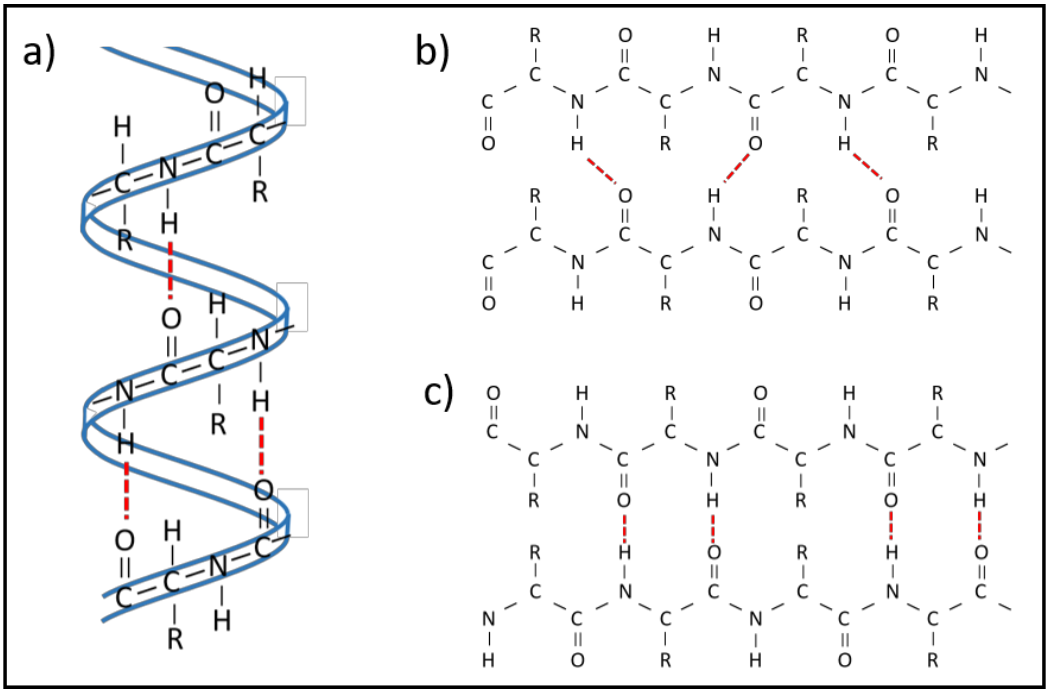

Figure 1.3: *The most common protein secondary structures of a polypeptide chain, a) α-helix b) parallel β-sheet c) anti-parallel β-sheet. Red dashed lines indicate hydrogen bonding between N-H terminus of one amino acid to the C=O of another.*

β-sheet structures involve the folding of the polypeptide chain where the adjacent sections are linked via hydrogen bonding between the N-H group of one chain and the C=O group of the adjacent section (Figure 1.3(b,c)).[37] Two main types of the β-sheet structure exist and are known as the parallel and antiparallel sheets and both structures are shown in Figure 1.3(b,c). A parallel β-sheet is one where both sections of the polypeptide chain are pointing in the

same direction, i.e. when both ends are the C or N terminus, if the sections are pointing in opposite directions then the sheet is anti-parallel. Irregularities in the β-sheet structure known as bulges have also been documented, more commonly in the anti-parallel formation and occur when two residues on one strand hydrogen bond with one residue of the opposite strand often causing a change in the directionality of the chain.[42] Other structures include the β-hairpin and β-turns, the former utilises two β-strands oriented in the anti-parallel direction which are linked by a loop containing 2 amino acids and cause a reversal in the direction of the polypeptide chain and the latter utilises 4 residue loop and also effects the directionality of the chain.[41] A polypeptide chain that lacks secondary structure is known as a random coil, where the chain samples all possible confirmations randomly with the only fixed relationship being the joining of adjacent amino acids via a peptide bond. Random coils are often the result of protein denaturing.[41,43]

The tertiary structure of the protein is the three-dimensional shape primarily due to the interactions between the functional R groups of the amino acids that make up the protein. The tertiary structure has a single polypeptide chain, known as the backbone, with one or more protein secondary structures. The final stage involves the interaction between two or more polypeptide chains, however not all proteins require the final quaternary structure as many proteins exist as a single polypeptide chain.[37,41]

1.2.2.1. Amide I vibrations

The amide I vibrations of the polypeptide backbone are one of the most commonly used in IR spectroscopic studies as IR absorption spectroscopy has been shown to be sensitive to the difference in α-helix and β-sheet protein secondary structures as their hydrogen bonds patterns are different producing characteristic frequencies for each structure.[34] The production of the proteins secondary structure results in group vibrations of the backbone which report on the protein secondary conformation. Isolated proteins typically display an IR absorption at ~1650 cm^{-1} and this peak originates largely from the backbone C=O stretch with smaller contributions from the NH bending mode and is referred to as the "amide I" vibration. Other amide group modes are also produced however the correlation between secondary structure and frequency is less straight forward.[33]

The hydrogen bonding created from the secondary structure gives rise to the interaction of vibrational modes along the polypeptide chain in a highly fixed geometry (Figure 1.3) via

through space mechanisms.[44–46] As the amide I vibrations are all at near-resonant frequencies with one another, this results in a highly delocalised vibrational component along the three-dimensional polypeptide chain, which is known as transition dipole coupling (TDC), ultimately producing large absorption features.[33] This leads to the characteristic amide I resonances that result from secondary structure elements, making amide I studies highly valuable to the study of proteins.[34]

As α-helices have on average 3.6 residues per helical turn, adjacent amino acid couplings along the polypeptide chain are not the only prominent coupling terms. As a result of the intramolecular hydrogen bonding throughout the helix which renders the structure very stable, there are two allowed amide I modes of an α-helix: one that lies parallel to the helical axis, the A mode, and the E mode which lies perpendicular.[30,47,48] The energy separation of these modes depends on the number of oscillators per turn on the helix which results in a different separation for different helix types. The parallel A mode is more intense than the E mode which is not easily observed in IR spectra and involves in the in-phase oscillation of all residues along the peptide chain; this local mode frequency occurs at around 1650 cm^{-1}. However the frequency at which an α-helix will absorb is dependent on the solvent, chain length, environment and long range disorder of the structure. Studies have shown that for a helix made of only 5-10 residues, the mode frequency is red shifted (to ~1660 cm^{-1}) compared to helices composed of more than 20 residues (1650 cm^{-1}).[30,47]

For the case of β-sheet structures, a parallel and perpendicular mode are present.[30,34,47] For anti-parallel β-sheets the perpendicular vibration is a result of the in-phase oscillation of the amide I modes on adjacent strands which produce a strong transition dipole that lies perpendicular to the β-strands. The parallel mode has a lower intensity than the perpendicular mode and this is because it involves the in-phase oscillations of the residues along each strand which are in turn out-of-phase with respect to the hydrogen bonded neighbouring reside on the adjacent strand. The perpendicular mode is the lower frequency vibration and appears as a narrow band between 1620-1640 cm^{-1}, the parallel mode is centred near 1670 cm^{-1}, however the frequencies at which the β-sheet will absorb are again dependent on the solvent environment, chain length, and long range disorder of the structure. Studies have observed that with an increasing polypeptide chain length, the higher frequency parallel peak does not change frequency position but the intensity of the peak does increase. The lower frequency perpendicular mode intensity also increases with

increasing size and experiences a red shift of up to ~20 cm^{-1}. The lower frequency mode has been established as a good indicator of β-sheet structure, with larger and well-ordered β-sheets having a typical absorption frequency situated at 1640 cm^{-1}.[49] Table 1.1 lists the common secondary structures and their amide I IR absorption peaks.[30,31,34,43,47,49,50]

Table 1.1: *Common protein secondary structures found in the amide I vibrational modes of molecular vibration.*

Secondary Structure	**IR absorption Wavenumbers (cm^{-1})**	**Secondary Structure**	**IR absorption Wavenumbers (cm^{-1})**
β-sheet (parallel)	*1620-1640* *1660-1670*	β-sheet (anti-parallel)	*1620-1640* *1670-1695*
α-helix	*1648-1660*	β-hairpin	*1635-1645* *1665-1680*
3_{10}-helix	*1660-1666*	β-turns	*~1680*
Random coil	*1645-1655*		

As a result of this, amide I modes are effectively probes of the secondary structure of proteins as the hydrogen bonding defines the unique amide I oscillations allowing the unique absorption features to be obtained and used to assign structures within a protein, which can lead to protein identification and changes that occur during different processes.[33,36,37]

1.3. Applications of Vibrational Spectroscopies to Serum Analysis

Biomedical applications of vibrational spectroscopy investigations have already been performed to study disease diagnostics[51–55], treatment monitoring,[56,57] molecular pathologies[58] and pharmaceutical applications.[59] IR spectroscopy provides a broad chemical coverage, typically measuring a few thousand wavenumbers at any time. Owing to the heterogeneous nature of disease,[1,5] the broad bio-molecular fingerprint obtained from vibrational spectroscopies of biofluids can be preferential over single biomarker detection by producing spectral signatures which reflects the total biochemical composition of the sample.

1.3.1. Overcoming Water Absorptions: Current Spectroscopic Methods

As biofluids are composed primarily of water and the IR absorption of the O-H bend of water, ~1645 cm^{-1}, overlaps the amide I band of proteins, ~1650 cm^{-1}, making their study challenging. In a bid to overcome this problem a variety of IR absorption spectroscopic methods and techniques have been explored.

Attenuated total reflectance (ATR) Fourier transform infrared (FTIR) spectroscopy is a common technique used for the analysis of biofluids. ATR allows spectroscopic interrogation of solid samples as the beam is no longer dependent on being transmitted through the sample but instead the attenuated reflection of the IR beam is recorded. The fraction of light penetrating the sample is known as an evanescent wave and only enters a few µm into the sample, as the sample absorbs energy the wave is attenuated. After one or multiple internal reflections the IR beam leaves the reflective element (Figure 1.4) and is directed towards the detector.

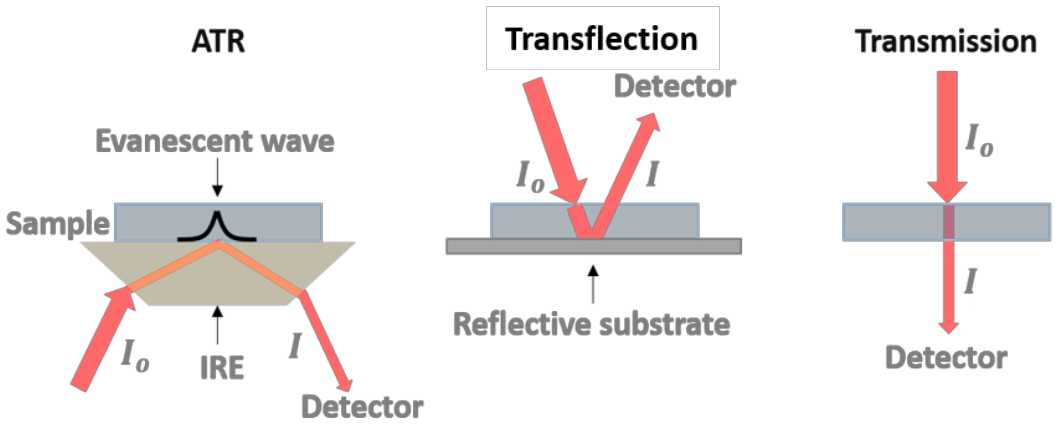

Figure 1.4: *Schematics of attenuated total reflectance (ATR), transflection and transmission modalities of FTIR spectroscopy. ATR uses an internal reflective element (IRE) shown in yellow, to reflect the beam back to the detector.* I_O *and* I *indicate the incident and reflected or transmitted, depending on modality, IR beams, respectively.*

The intensity of the evanescent wave decays exponentially with distance from the surface of the internal reflective element (IRE), as only a few µm of the sample is tested, this technique is not affected by the sample thickness provided that the sample thickness is greater than the depth of penetration of the wave. This has resulted in the use of ATR towards dried

biofluids in an effort to overcome the competing absorptions from both the serum constituents and water, as drying the sample prior to investigation reduces the signal from the serum water.[60] However inhomogeneous drying patterns are observed from dried samples which can introduce experimental difficulties. The application of serum droplets to an artificial surface (the IRE) results in the Vroman effect, this is the competitive adsorption of serum proteins to the surface which results in the chemical and physical inhomogeneity in the deposited dried layer.[61–63] Many studies have demonstrated the effects of sample drying, including the Vroman and coffee-ring effects, cracking and gelation patterns produced as well as the development of optimum sample preparation steps, including sample filtration and drying conditions to help overcome these problems.[53,64–67] Despite the analysis of dried biofluids being negatively affected by inconsistencies in droplet drying, ATR-FTIR has shown clear discrimination between healthy and non-healthy patients among increased sensitivities to detection limits when compared to other IR absorption modalities.[4,55,66,68,69]

Steps towards high-throughput analysis have also been made. The development of disposable internal reflectance elements (IRE) for ATR-FTIR modalities combined with an automated slide indexing unit allows accurate movement of the IRE across the spectrometer for high-throughput analysis.[70] The use of an automated FTIR combined with robotically spotting of serum samples onto a 384 multi-well titer plates have also been established allowing high-throughput spectroscopic analysis of serum.[71] The robotic spotting was also found to produce more homogenous droplets once dried.

FTIR experiments in the transflection geometry have also been used to analyse serum as samples can be dried prior to use thus reducing the water content of the sample. This mode detects the absorbed IR light after it has passed through the sample, reflected off a substrate and passed back through the sample, essentially doubling the thickness (Figure 1.4), allowing analysis of thin (1-4 μm) samples which is beneficial for biofluids.[72–74] However, spectral shifts and distortions are often noted with this modality due to the production of electric-field standing waves at the sample-substrate interface, and when using dried samples the effects of droplet drying need also be considered.[75–80]

Transmission mode FTIR experiments of wet serum are typically obtained alongside a background water spectrum in order to subtract the strong water contribution however this process is subjective.[81] Difficulties in reproducing optical pathlengths less than 10 μm have also been documented, making the analysis of serum using transmission mode challenging

as larger path lengths leads to the absorption of all of the IR beam.[82] Transmission mode FTIR measurements of proteins have also been made in deuterated solvents which involves exchanging solvent hydrogens for deuteriums. This H/D exchange provides a clearer observation of the protein amide I band by shifting the overlapping water peak to lower wavenumbers. This has allowed a more detailed study of the structural and conformational changes in proteins to be monitored as a function of disease states or drug binding.[33,83–86] However the use of D_2O also leads to the practical issue surrounding incomplete proton exchange due to the complex nature of serum and elevated sample preparation and costs.[33,81] Deuteration has also been shown to alter or disrupt physiological processes again making it unsuitable for use of biofluid diagnostics.[87–89]

High brightness quantum cascade lasers (QCLs) are tuneable semiconductor lasers that can emit discrete frequencies across the mid-IR spectrum, with tuning ranges around 1000 wavenumbers having been reported[90,91] with the high spectral power of QCLs allowing thicker samples to be analysed.[90–92] QCLs are often combined with microscopy techniques allowing full hyperspectral datasets covering large areas, over 1000 cm^{-1}, to be obtained in a matter of seconds.[91] Synchrotron based light sources have also been utilised for biomedical analysis and are often combined with imaging modalities. Synchrotron sources produce light up to 1000 times greater than benchtop IR spectrometers and thus have the ability to produce a superior signal-to-noise ratio and allows greater spatial resolutions to be obtained when compared with traditional globar techniques.[74] Although transmission measurements of aqueous protein solutions are possible, the difficulties of analysing these samples in water based solvents remain with this methodology.

Protocols have been established within the biomedical applications of IR spectroscopy community in order to standardise sample pre-treatment, measurement methods and procedures, as well as data pre-processing techniques.[53,64,74,93]

Applications of Raman spectroscopy have also been utilised to interrogate biofluids. Raman spectroscopy is an inelastic scattering technique based on the Raman effect and is specifically well suited to working in water based solvents due to the relatively low scattering that is produced from water molecules, allowing signals from the analytes in solution to be obtained.[94,95] However the Raman effect is generally quite weak as only a small proportion (around 1 in 10 million) of incident photons are Raman scattered with a corresponding change in frequency and approaches are often combined with imaging, labelling or signal

enhancement strategies.[64,94,96,97] Due to the different fundamental theory of both IR and Raman they are generally considered complementary techniques. [98,99]

Sample protocols in the form of filtration have also been investigated to help improve sensitivity to low molecular weight (LMW) disease markers in serum which involves the removal of high molecular weight (HMW) species by centrifugal filtration.[60,66,67,100] Although, with the filtration methodologies the unintentional removal of LMW components may occur as serum albumin (MW = 66.5 kDa) have been well documented as a transport protein of low molecular weight species and thus removal of albumins during the filtration process could also result in the concomitant removal of the LMW peptidome.[101,102]

1.3.2. Molecular and Disease Applications

The analysis of serum using IR spectroscopies have been successful in detection of specific molecular targets. The detection and quantification of 8 major constituents of serum (total protein, glucose, cholesterol, high and low density lipoproteins, triglycerides, urea and uric acid) has been made from a study containing 247 patients using Raman and IR spectroscopies.[103,104] Concentrations of the analyte concentrations are compared to concentrations determined via typical laboratory methods. Although accuracies are not yet better than laboratory based techniques, strong advantages of IR techniques lie in the small volume of sample required, comparatively less sample preparation and thus associated costs as well as the ability to gain information on many parameters from a single spectroscopic measurement compared to eight different laboratory methods.

Studies by Bonnier *et al.* have shown the use of filtration methodologies to increase the sensitivity of low molecular weight (LMW) species by the removal of the high molecular weight (HMW) fraction of serum, namely the serum albumins and globulins.[105] Using ATR-FTIR spectroscopy, glycine has been detected in the LMW fraction up to 50 times lower than whole serum.[66] The detection of glucose in the absence of the HMW using multivariate analysis tools produces a linear relationship between glucose concentrations in patient samples and the intensity of spectral features; an improvement of the root mean square error of the validation set by a factor of 5 was noted when compared to whole unfiltered serum.[67]

Mahmood *et al.* have shown the ability of Raman spectroscopy combined with multivariate analyses to identify patients infected by the dengue virus and distinct features in the Raman

spectra assigned to different immunoglobulin proteins allowed monitoring of the disease progression. Results attained were comparable to the specificities and sensitivities achieved using traditional enzyme linked immunosorbent assay and polymerase chain reaction techniques, however on a much quicker timescale and with lower associated costs.[106] Using similar analytical techniques, Sahu *et al.* have predicted the recurrence of oral cancers from serum collected before and after surgery.[107] Colorectal cancers have also been classified from control patients in serum using Raman spectroscopy coupled with principal component analysis, where decreases in spectrum peaks associated with tryptophan, phenylalanine and beta carotene were noted.[108]

Applications of vibrational spectroscopies as a potential tool for disease diagnostics have been demonstrated due to their high specificity and ability to cover a broad range of molecular species.[6,95,109,110] Advanced analytical techniques allow pattern recognition and changes in the broad spectral fingerprint associated with disease to be detected. An FTIR study of 389 patients with acute chest pain was shown to diagnose acute myocardial infarction from patients with minor chest pains of other origins allowing therapeutic attention to be given to those in need of it most. This transflection mode study led by Petrich *et al.* conceives a diagnostic pattern recognition method that has the potential to aid the diagnosis within the first 6 hours after the onset of chest pain and suggest spectral changes due to an increased role of carbohydrates. The produced spectral results were found to be comparable to routine cardiac laboratory markers on the same dataset.[111]

In the field of oncology, proof-of-principle methods have been established for the detection of brain tumours using ATR-FTIR spectroscopy and further identification of tumour types using a study of 433 patients, achieving sensitivities and specificities greater than 80%.[4,112] ATR-FTIR has also been shown to quantify protein concentrations in serum using diluted and dried serum samples.[69] Backhaus *et al.* demonstrated the use of transflection and transmission FTIR combined with cluster analysis and artificial neural network analysis to distinguish between breast cancer patients and controls. Spectral comparisons with 11 other diseases, covering a total of 3119 patients, assured spectral changes were in fact from breast cancer.[113] Sheng *et al.* have shown the ability of FTIR spectroscopy to diagnose gastric cancer from healthy controls from a ratio of amplitudes provided by the stretching vibrations of RNA and DNA found in serum.[114]

However the studies discussed utilise a broad range of the IR absorption spectrum as little variation is often seen between the overlapping amide I peak of proteins in serum.[115,116] Despite the advances made with IR absorption spectroscopy for serum analysis, the overlapping water absorptions across the IR spectra is still a notable issue, either directly through absorptions or indirectly through inhomogenous drying patterns of protein solutions.

Ultrafast 2D-IR spectroscopy offers increased structural information of the biomolecules being studied by utilising short pulses of IR light to excite molecular vibrations of a system. By spreading the information obtained over a second frequency axis a correlation map of excitation versus detection frequency is produced. This gives rise to unique spectral patterns via intramolecular couplings of proteins, which allows the unravelling of convoluted protein structures,[117,118] resulting in a 2D-IR signal that is sensitive to protein secondary structure to a greater degree than IR absorption spectroscopy.[30] The ultrafast duration of IR pulses used in 2D-IR provides the ability to observe molecular processes occurring on the femtosecond timescale; examples include hydrogen bond fluctuations, protein folding and energy transfer between vibrational modes within a sample.[30,119–124] The theory and methodology of 2D-IR spectroscopy will be discussed in the following chapter. Until now, the absorptions associated with the overlapping water contributions of the protein amide I have also been met with sample deuteration. However in this thesis, 2D-IR is shown to bypass the water absorption signals owing to its signal dependence on the transition dipole moment and time resolution offered by the ultrafast pulses.

1.4. Application of Two-Dimensional IR Spectroscopy to Protein Studies

Since the development of 2D-IR there have been many studies of biological molecules, providing structural information that is currently not possible using IR absorption methods.[50,125–131] The ultrafast timescale of 2D-IR provides observation of dynamics on a femtosecond timescale which combined with the dispersal of both the pump (excitation) and probe (detection) frequencies reveals coupling and the transfer of energy between vibrational modes presenting as off-diagonal peaks on the 2D-IR spectrum. The theory and instrumentation of 2D-IR will be discussed in the following chapter.

Several 2D-IR studies have been made to understand the unique 2D-IR signatures of protein secondary structure. 2D-IR signals are dependent on the 4th power of the transition dipole

moment (compared to the 2nd for IR absorption spectroscopy) allowing enhancement of the strong amide I mode of biological molecules and assignment of proteins into varying degrees of α-helix and β-sheet content.[117] 2D-IR spectral signatures of β-sheets have a characteristic 'Z' shaped pattern, compared with the double peak seen in IR absorption, resulting from the formation of cross-peaks between the two IR active modes.[50,132] α-helical structures have been shown to produce a flattened 'figure of 8' signature , while random coil features tend to be unstructured diagonally elongated peaks.[50] These typical α-helical and β-sheet structures are shown in Figure 1.5.

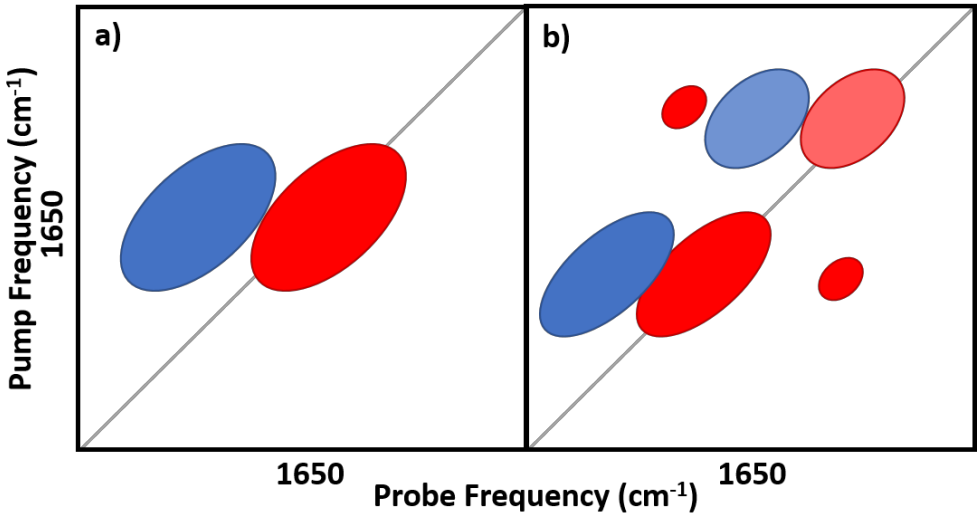

Figure 1.5: *Typical 2D-IR lineshapes observed for a) α-helix and b) β-sheet secondary protein structures.* α-helical structures typically show a flattened 'figure of 8' signature and the *β-sheet shows a characteristic 'Z' pattern from the low frequency perpendicular and higher frequency parallel modes.*

Isotopic labelling of the protein backbone has been shown to alter the amide I vibration frequency, shifting the new labelled amide I vibration to a spectral region where no other absorptions occur, allowing the characterisation of the dynamics of proteins, enzymes and protein-water interactions.[133–136] Works led by Zanni have shown insights into amyloid aggregates, in particular the observation of fibril formation which have been linked to many diseases including type 2 diabetes and Alzheimer's disease.[49,136–139] Protein dynamic studies have reported lifetimes of vibrational modes that differ depending on protein secondary structure. By modifying the 2D-IR pulses to increase the time between excitation and detection pulses, suppression of signals from the fast decaying isotopically labelled random coil modes of amyloid fibrils has been reported, leaving only β-sheet vibrational features.[140]

The ultrafast time resolution of 2D-IR has also been utilised to observe the misfolding of proteins into β-sheet rich amyloid fibrils, the production of which has been associated with multiple diseases including those related to cognitive decline, and capturing the secondary structural changes during this process.[138,139] This isotopic labelling approach, while powerful, is not desired for biofluids as the label used has the potential to interact with the sample which could in theory alter protein structure thus altering the spectrum produced.

2D-IR has also been shown to be sensitive to small structural changes, by using the ratio of the 2D-IR amide I band to that of IR absorption spectroscopy the transition dipole moment can be measured which demonstrates the extent of coupling within a molecular structure.[129,141] Studies of the messenger protein calmodulin have shown the ability to quantify protein secondary structural changes affecting only 7 of the 150 residue protein upon calcium binding, and provide comparable results to circular dichroism techniques which currently sets the standard for protein secondary structural determination.[122]

Studies of protein-ligand binding have shown how 2D-IR can map the migration of ligands between sites on a protein, as well as the ability to couple 2D-IR with powerful analytical techniques. [142,143] Label-free 2D-IR has been used to study the effects of drug binding on protein structure and dynamics in InhA, an enzyme which is in part responsible for cell wall integrity in the Myobacterium tuberculosis bacterium. Changes upon drug-binding induced enzyme inhibition were noted in the off-diagonal region of the amide I 2D-IR band.[144] Quantification of ligand binding has been shown via the labelling of peptides with IR active amino acid side chains which act as probes of local structure. Observation of the labelled peptide binding to a domain within the protein tyrosine phosphatase 1E formed calculation of dissociation constants comparable to isothermal titration calorimetry techniques.[145] Instrumental techniques have also been employed to implement fast scanning times and the reduction of scattering and phasing issues to produce a lower detection limit of ~100 μM of the labelled peptide.

Using both the off- and on-diagonal 2D-IR features, studies of ex-vivo porcine cataracts and human lens tissue was shown to reveal amyloid β-sheet-like secondary structures where the current standard technique, transmission electron microscopy (TEM), failed to resolve them when the fibres were too short.[137,146] Comparisons made between juvenile and mature human lenses with no diagnosed cataracts suggests that the amyloid structures begin formation prior to onset of cataracts, establishing a link between formation and structure

with disease pathology and progression. This highlights the possible use of 2D-IR to understand how diseases are manifest, potentially allowing more accurate and effective treatments to be explored.

Experimental and theoretical investigations into the structure and dynamics of the stretching and bending modes of H_2O, HOD and D_2O have also been studied allowing population relaxation of the modes examined to be calculated.[147–151] Apart from the results published from this thesis, 2D-IR studies of the amide I mode of proteins have all been characterised in D_2O.[118] However deuteration has been shown to alter or disrupt physiological processes again making it unsuitable for its use in biofluid diagnostics.[87] The 2D-IR studies presented here show the potential of 2D-IR to gain an insight into the fundamental mechanisms that dictate the behaviour of proteins, among other biological molecules. Despite the advances made with 2D-IR spectroscopy, the amount of structural information obtained is largely qualitative and challenges regarding the complexity of biofluids are still prevalent.

1.5. Aims & Objectives

This thesis aims to expand the use of 2D-IR to date by exploring its ability to detect, identify and quantify proteins in physiologically relevant H_2O solutions, e.g. biofluids. In Chapter 3, the spectroscopic differences between water and deuterated water is explored, and the first 2D-IR spectra of the amide I band of proteins in water are shown. Analysis of individual proteins and water alone allow establishes the ability of 2D-IR to measure proteins in water without the background water signal, allowing the detection of clinically relevant proteins in serum. The aim of this chapter is to determine whether the two most dominant proteins present in serum, albumins and globulins which are have different secondary structures (α-helix and β-sheet, respectively) can be distinguished. Attempts are also made to identify and ratio-metrically quantify structurally similar proteins (Immunoglobulin A, G and M which are all primarily β-sheet proteins) within serum. Chapter 4 aims to study the low molecular weight fraction of serum to determine whether smaller and less abundant proteins can still be detected in the presence of the heavy and dominant protein fraction. By using glycine as a model low molecular weight protein, the detection limits of 2D-IR are evaluated. The aim of this chapter is to assess which physiologically relevant proteins could be detected within serum using their typical concentration ranges for future 2D-IR measurements. In Chapters 3 and 4, the use of ratios within a dataset are implemented due to the minimal control of the

sample thickness which does not allow direct comparability between spectra. The aim of Chapter 5 is to overcome this problem and sees the introduction of a new method to normalise 2D-IR spectra which has been devised and is explored. This takes the analysis of 2D-IR to the next level, allowing cross dataset evaluation to be performed without the need for ratio metrics. In the final chapter, Chapter 6, the major objective is to employ this normalisation method in order to detect allosteric changes of serum albumin during drug binding. Several common drugs are utilised to determine whether 2D-IR is suitable to detect drug binding with albumin at clinically relevant concentrations.

1.6. References

1. Hu, S., Loo, J. & Wong, D. Human body fluid proteome analysis. *Proteomics* **6,** 6326–6353 (2006).

2. Veenstra, T. D., Conrads, T.P., Hood, B.L., Avellino, A.M. *et al.* Biomarkers: mining the biofluid proteome. *Mol. Cell. Proteomics* **4.4,** 409–418 (2005).

3. Li, Y. & Bahassi, E. M. Biofluid-based circulating tumor molecules as diagnostic tools for use in personalized medicine. *J. Mol. Biomarkers Diagnosis* **5,** 157 (2013).

4. Hands, J. R., Clemens, G., Stables, R., Ashton, K. *et al.* Brain tumour differentiation: rapid stratified serum diagnostics via attenuated total reflection Fourier-transform infrared spectroscopy. *J. Neurooncol.* **127,** 463–472 (2016).

5. 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* **6,** 961–967 (2006).

6. Baker, M. J., Sockalingum, G. D., Hughes, C. & Lukaszewski, R. A. Developing and understanding biofluid vibrational spectroscopy: A critical review. *Chem. Soc. Rev.* **45,** 1803–1818 (2016).

7. Tortura, G. J. & Nielsen, M. T. *Principles of Human Anatomy*. (John Wiley & Sons, 2012).

8. De, M., Rana, S., Akpinar, H., Miranda, O.R. *et al.* Sensing of proteins in human serum conjugates of nanoparticles and green fluorescent protein. *Nat. Chem.* **1,** 461–465 (2009).

9. Gutman, A. B., Moore, D. H., Gutman, E. B., McClellan, V. *et al*. Fractionation of serum proteins in hyperproteinemia with special reference to multiple myeloma. *J. Clin. Invest.* **20,** 765–783 (1941).

10. Marrack, J. R. & Hoch, H. Serum Proteins: A Review. *J. Clin. Pathol.* **2,** 161–192 (1949).

11. Busher, J. T. Serum albumin and globulins. in *Clinical Methods: The History, Physical, and*

Laboratory Examinations. (eds. Hurst, J., Walker, H. & Hall, W.) 497–500 (Butterworths, 1990).

12. Suh, B., Park, S., Shin, D.W., Yun, J.M. *et al.* Low albumin-to-globulin ratio associated with cancer incidence and mortality in generally healthy adults. *Ann. Oncol.* **25,** 2260–2266 (2014).

13. Liu, J. Prognostic value of pretreatment albumin – globulin ratio in predicting long-term mortality in gastric cancer patients who underwent D2 resection. *Onco. Targets. Ther.* **10,** 2155–2162 (2017).

14. Al-Muhtaseb, S. I. Serum and saliva protein levels in females with breast cancer. *Oncol. Lett.* **8,** 2752–2756 (2014).

15. Llewellyn, D., Langa, K., Friedland, R. & Lang, I. Serum Albumin Concentration and Cognitive Impairment. *Curr. Alzheimer Res.* **7,** 91–96 (2010).

16. Kim, T. S., Pae, C., Yoon, S., Jang, Y.W. *et al.* Decreased plasma antioxidants in pateints with Alzheimer's disease. *Int. J. Geriatr. Psychiatry* **21,** 344–348 (2006).

17. Chojkier, M. Inhibition of albumin synthesis in chronic diseases: Molecular mechanisms. *J. Clin. Gastroenterol.* **39,** 143–146 (2005).

18. Lin, S., Sun, Q., Mao, W. & Chen, Y. Serum immunoglobulin A (IgA) level is a potential biomarker indicating cirrhosis during chronic hepatitis B infection. *Gastroenterol. Res. Pract.* **2016,** (2016).

19. Göke, M., Hoffmann, J. C., Evers, J., Krüger, H. & Manns, M. P. Elevated serum concentrations of soluble selectin and immunoglobulin type adhesion molecules in patients with inflammatory bowel disease. *J. Gastroenterol.* **32,** 480–486 (1997).

20. NHS. Pathology Tests. Available at: https://www.southtees.nhs.uk/services/pathology/tests/. (Accessed: 14th May 2020)

21. Lubran, M. M. The measurement of total serum proteins by the Biuret method. *Ann. Clin. Lab. Sci.* **8,** 106–110 (1978).

22. Zheng, K., Wu, L., He, Z., Yang, B. & Yang, Y. Measurement of the total protein in serum by biuret method with uncertainty evaluation. *Measurement* **112,** 16–21 (2017).

23. Atkins, P. & Paula, J. de. *Atkins' Physical Chemistry*. (Oxford University Press, 2006).

24. Gupta, A. & Stockham, S. L. Negative interference of icteric serum on a bichromatic biuret total protein assay. *Vet. Clin. Pathol.* **43,** 422–427 (2014).

25. Byler, D. M. & Susi, H. Examination of the secondary structure of proteins by deconvolved FTIR spectra. *J. Biopolym.* **25,** 469–487 (1986).

26. Surewicz, W. K. & Mantsch, H. H. New insight into protein secondary structure from resolution-enhanced infrared spectra. *Biochim. Biophys. Acta* **952,** 115–130 (1988).

27. Kumar, V., Coluccelli, N. & Polli, D. Coherent optical spectroscopy microscopy and applications. in *Molecular and Laser Spectroscopy* (Elsevier, 2018).

28. Stuart, B. *Infrared Spectroscopy: Fundamentals and Applications*. (J.Wiley and Sons, 2009).

29. Telle, H. H., Urena, A. G. & Donovan, R. J. *Laser Chemistry*. (John Wiley & Sons, 2007).

30. Hamm, P. & Zanni, M. T. *Concepts and Methods of 2D Infrared Spectroscopy*. (Cambridge University Press, 2011).

31. Diem, M. *Modern Vibrational Spectroscopy and Micro-Spectroscopy*. (Wiley, 2015).

32. Thompson, J. M. *Infrared Spectroscopy*. (Pan Stanford, 2018).

33. Barth, A. & Zscherp, C. What vibrations tell us about proteins. *Q. Rev. Biophys.* **35,** 369–430 (2002).

34. Barth, A. Infrared spectroscopy of proteins. *Biochim. Biophys. Acta - Bioenerg.* **1767,** 1073–1101 (2007).

35. Susi, H. & Timasheff, S. N. Infrared spectra and protein conformations in aqueous solutions 1. The amide I band in H2O and D2O solutions. *J. Biol. Chem.* **242,** 5460–5467 (1967).

36. Schulz, G. E. & Schrimer, R. H. *Principles of Protein Structure*. (Springer).

37. Petsko, G. A. & Ringe, D. *Protein Structure and Function*. *Annual Review of Physical Chemistry* (New Science Press Ltd, 2004).

38. Meloun, B., Moravek, L. & Kostka, V. Complete Amino Acid Sequence of Human Serum Albumin. *FEBS Lett.* **58,** 134–137 (1975).

39. Bromer, W. W., Boucher, M. E., Patterson, J. M., Pekar, A. H. *et al*. Glucagon structure and function. *J. Biol. Chem.* **247,** 2581–2585 (1972).

40. Shepherd, N. E., Hoang, H. N., Abbenante, G. & Fairlie, D. P. Left- and right-handed alpha-helical turns in homo- and hetero-chiral helical scaffolds. *J. Am. Chem. Soc.* **131,** 15877–15886 (2009).

41. Richardson, J. S. The Anatomy & Taxonomy of Protein Structure. *Anat. Taxon. Protein Struct.* **34,** 167–339 (1981).

42. Craveur, P., Joseph, A. P., Rebehmed, J. & De Brevern, A. G. β-Bulges: Extensive structural analyses of β-sheets irregularities. *Protein Sci.* **22,** 1366–1378 (2013).

43. Smith, L. J., Fiebig, K. M., Schwalbe, H. & Dobson, C. M. The concept of a random coil : Residual structure in peptides and denatured proteins. *Fold. Des.* **1,** 95–106 (1996).

44. Kubelka, J., Kim, J., Bour, P. & Keiderling, T. A. Contribution of transition dipole coupling to amide coupling in IR spectra of peptide secondary structures. *Vib. Spectrosc.* **42,** 63–73 (2006).

45. Krimm, S. & Abe, Y. Intermolecular Interaction Effects in the Amide I Vibrations of Beta Polypeptides. *Proc. Natl. Acad. Sci.* **69,** 2788–2792 (1972).

46. Moran, A. & Mukamel, S. The origin of vibrational mode couplings in various secondary structural motifs of polypeptides. *Proc. Natl. Acad. Sci.* **101,** 506–510 (2004).

47. Baiz, C. R., Reppert, M. & Tokmakoff, A. Introduction to protein 2D-IR spectroscopy. in *Ultrafast Infrared Vibrational Spectroscopy* (ed. Fayer, M. D.) (CRC Press, 2012).

48. Nguyen, K. T., Le Clair, S. V., Ye, S. & Chen, Z. Orientation determination of protein helical secondary structures using linear and nonlinear vibrational spectroscopy. *J. Phys. Chem. B* **113,** 12169–12180 (2009).

49. Strasfeld, D. B., Ling, Y. L., Shim, S. & Zanni, M. T. Tracking fiber formation in human islet amyloid polypeptide with automated 2D-IR spectroscopy. *J. Am. Chem. Soc.* **130,** 6698–6699 (2008).

50. Ganim, Z. *et al.* Amide I two-dimensional infrared spectroscopy of proteins. *Acc. Chem. Res.* **41,** 432–441 (2008).

51. Peters, A. S. *et al.* Serum-infrared spectroscopy is suitable for diagnosis of atherosclerosis and its clinical manifestations. *Vib. Spectrosc.* **92,** 20–26 (2017).

52. Kendall, C. *et al.* Vibrational spectroscopy: a clinical tool for cancer diagnostics. *Analyst* **134,** 1029–1045 (2009).

53. Cameron, J. M., Butler, H. J., Palmer, D. S. & Baker, M. J. Biofluid spectroscopic disease diagnostics : A review on the processes and spectral impact of drying. *J. Biophotonics* (2018). doi:10.1002/jbio.201700299

54. Socha, E., Koba, M. & Kośliński, P. Amino acid profiling as a method of discovering biomarkers for diagnosis of neurodegenerative diseases. *Amino Acids* **51,** 367–371 (2019).

55. Krafft, C. & Sergo, V. Biomedical applications of Raman and infrared spectroscopy to diagnose tissues. *Spectroscopy* **20,** 195–218 (2006).

56. Caixeta, D. C., Aguiar, E.M.G., Cardoso-Sousa, L., Coelho, L.M.D. *et al.* Salivary molecular spectroscopy : A sustainable , rapid and non-invasive monitoring tool for diabetes mellitus during insulin treatment. *PLoS One* **15,** e0223461 (2020).

57. Sule-Suso, J., Skingsley, D., Sockalingum. G.D., Kohler, A. *et al.* FT-IR microspectroscopy as a tool to assess lung cancer cells response to chemotherapy. *Vib. Spectrosc.* **38,** 179–184 (2005).

58. Diem, M., Mazur, A., Lenau, K., Schubert, J. *et al.* Molecular pathology via IR and Raman spectral imaging. *J. Biophotonics* **6,** 855–886 (2013).

59. Wartewig, S. & Neubert, R. H. H. Pharmaceutical applications of Mid-IR and Raman spectroscopy. *Adv. Drug Deliv. Rev.* **57,** 1144–1170 (2005).

60. Bonnier, F., Baker, M. J. & Byrne, H. J. Improved protocols for vibrational spectroscopic analysis of body fluids. *J. Biophotonics* **7,** 167–179 (2014).

61. Noh, H. & Vogler, E. A. Volumetric interpretation of protein adsorption: Competition from mixtures and the Vroman effect. *Biomaterials* **28,** 405–422 (2007).

62. Vroman, L., Adams, A. L., Fischer, G. C. & Munoz, P. C. Interaction of high molecular weight Kininogen, Factor XII, and Fibrinogen in plasma at interfaces. *Blood* **55,** 156–159 (1980).

63. Vroman, L. & Adams, A. L. Findings with the recording ellipsometer suggesting rapid exchange of specific plasma proteins at liquid/solid interfaces. *Surf. Sci.* **16,** 438–446 (1969).

64. Baker, M. J. *et al.* Developing and understanding biofluid vibrational spectroscopy: a critical review. *Chem. Soc. Rev.* **45,** 1803–1818 (2016).

65. Lovergne, L., Clemens, G., Untereiner, V., Lukaszweski, R.A. *et al.* Investigating optimum sample preparation for infrared spectroscopic serum diagnostics. *Anal. Methods* **7,** 7140–7149 (2015).

66. Bonnier, F., Brachet, G., Duong, R., Sojinrin, T. *et al.* Screening the low molecular weight fraction of human serum using ATR-IR spectroscopy. *J. Biophotonics* **9,** 1085–1097 (2016).

67. Bonnier, F., Blasco, H., Wasselet, C., Brachet, G. *et al.* Ultra-filtration of human serum for improved quantitative analysis of low molecular weight biomarkers using ATR-IR spectroscopy. *Analyst* **142,** 1285–1298 (2017).

68. 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.* **0,** 1–13 (2020).

69. Spalding, K., Bonnier, F., Bruno, C., Blasco, H. *et al.* Enabling quantification of protein concentration in human serum biopsies using attenuated total reflectance – Fourier transform infrared (ATR-FTIR) spectroscopy. *Vib. Spectrosc.* **99,** 50–58 (2018).

70. Butler, H.J., Brennan, B.M., Cameron, J.M., Finlayson, D. *et al.* Development of high-throughput ATR-FTIR technology for rapid triage of brain cancer. *Nat. Commun.* **10,** 4501 (2019).

71. Ollesch, J., Heinze, M., Heise, H.M., Behrens, T. *et al.* It's in your blood: spectral biomarker candidates for urinary bladder cancer from automated FTIR spectroscopy. *J. Biophotonics* **7,** 210–221 (2014).

72. Baker, M. J., Hughes, C. S. & Hollywood, K. A. *Biophotonics: Vibrational Spectroscopic Diagnostics*. (Morgan & Claypool Publishers, 2016).

73. Smith, B. C. *Fundamentals of Fourier Transform Infrared Spectroscopy*. (CRC Press, 2011).

74. Baker, M.J., Trevisan, J., Bassan, P., Bhargava, R. *et al.* Using Fourier transform IR spectroscopy to analyze biological materials. *Nat. Protoc.* **9,** 1771–1791 (2014).

75. Bassan, P., Lee, J., Sachdeva, A., Pissardini, J. *et al.* The inherent problem of transflection-mode infrared spectroscopic microscopy and the ramifications for biomedical single point and imaging applications. *Analyst* **138,** 144–157 (2013).

76. Filik, J., Frogley, M. D., Pijanka, J. K., Wehbe, K. & Cinque, G. Electric field standing wave artefacts in FTIR micro-spectroscopy of biological materials. *Analyst* **137,** 853–861 (2012).

77. Davis, B.J., Carney, P.S. & Bhargava, R. Theory of midinfrared absorption microspectroscopy: I. Homogeneous samples. *Anal. Chem.* **82,** 3474–3486 (2010).

78. Davis, B.J., Carney, P.S. & Bhargava, R. Theory of mid-infrared absorption microspectroscopy: II. Heterogeneous samples. *Anal. Chem.* **82,** 3487–3499 (2010).

79. Cui, L., Butler, H. J., Martin-Hirsch, P. L. & Martin, F. L. Aluminium foil as a potential substrate for ATR-FTIR, transflection FTIR or Raman spectrochemical analysis of biological specimens. *Anal. Methods* **8,** 481–487 (2016).

80. Staniszewska-Slezak, E., Rygula, A., Malek, K. & Baranska, M. Transmission versus transflection mode in FTIR analysis of blood plasma: Is the electric field standing wave effect the only reason for observed spectral distortions? *Analyst* **140,** 2412–2421 (2015).

81. Kong, J. & Yu, S. Fourier transform infrared spectroscopic analysis of protein secondary structures. *Acta Biochim. Biophys. Sin. (Shanghai).* **39,** 549–559 (2007).

82. Lasch, P. & Kneipp, J. *Biomedical Vibrational Spectroscopy*. (Wiley-Interscience, 2010).

83. Fasano, M., Curry, S., Terreno, E., Galliana, M. *et al.* The extraordinary ligand binding properties of human serum albumin. *IUBMB Life* **57,** 787–796 (2005).

84. Huang, Y. T., Liao, H. F., Wang, S. L. & Lin, S. Y. Glycation and secondary conformational changes of human serum albumin: Study of the FTIR spectroscopic curve-fitting technique. *AIMS Biophys.* **3,** 247–260 (2016).

85. Malferrari, M., Venturoli, G., Francia, F. & Mezzetti, A. A new method for D2O/H2O exchange in infrared spectroscopy of proteins. *Spectrosc. An Int. J.* **27,** 337–342 (2012).

86. Baenziger, J. E. & Methot, N. Fourier transform infrared and hydrogen/deuterium exchange reveal an exchange-resistant core of α-helical peptide hydrogens in the nicotinic acetylcholine receptor. *J. Biol. Chem.* **270,** 29129–29137 (1995).

87. Kselíková, V., Vítová, M. & Bišová, K. Deuterium and its impact on living organisms. *Folia Microbiol. (Praha).* **64,** 673–681 (2019).

88. Kong, J. & Yu, S. Fourier Transform Infrared Spectroscopic Analysis of Protein Secondary Structures. *Acta Biochim. Biophys. Sin. (Shanghai).* **39,** 549–559 (2007).

89. 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.* **10,** 382–396 (2015).

90. Pecharromán-Gallego, R. An overview of quantum cascade lasers: Origins and developments. in *Quantum Cascade Lasers* (ed. Stavrou, V. N.) (IntechOpen, 2017).

91. Childs, D.T.D., Hogg, R.A., Revin, D.G., Rehman, I.U. *et al.* Sensitivity advantage of QCL tunable-laser mid-infrared spectroscopy over FTIR spectroscopy. *Appl. Spectrosc. Rev.* **50,** 822–839 (2015).

92. Isensee, K., Kröger-Lui, N. & Petrich, W. Biomedical applications of mid-infrared quantum cascade lasers-a review. *Analyst* **143,** 5888–5911 (2018).

93. Lovergne, L., Bouzy, P., Untereiner, V., Garnotel, R. *et al.* Biofluid infrared spectro-diagnostics: pre-analytical considerations for clinical applications. *Faraday Discuss.* **187,** 521–537 (2016).

94. Smith, E. & Dent, G. *Modern Raman Spectroscopy. A Practical Approach*. *Journal of Raman Spectroscopy* **36,** (J.Wiley and Sons, 2005).

95. Parachalil, D. R., McIntyre, J. & Byrne, H. J. Potential of Raman spectroscopy for the analysis of plasma/serum in the liquid state: recent advances. *Anal. Bioanal. Chem.* **412,** 1993–2007 (2020).

96. Bonifacio, A., Cervo, S. & Sergo, V. Label-free surface-enhanced Raman spectroscopy of biofluids : fundamental aspects and diagnostic applications. *Anal. Bioanal. Chem.* **407,** 8265–8277 (2015).

97. Kong, K., Kendall, C., Stone, N. & Notingher, I. Raman spectroscopy for medical diagnostics - From in-vitro biofluid assays to in-vivo cancer detection. *Adv. Drug Deliv. Rev.* **89,** 121–134

(2015).

98. Leal, L. B., Nogueira, M. S., Canevari, R. A. & Carvalho, L. F. C. S. Vibration spectroscopy and body biofluids: Literature review for clinical applications. *Photodiagnosis Photodyn. Ther.* **24,** 237–244 (2018).

99. Lyng, F. M., Ramos, I. R. M., Ibrahim, O. & Byrne, H. J. Vibrational microspectroscopy for cancer screening. *Appl. Sci.* **5,** 23–35 (2015).

100. Parachalil, D.R., Bruno, C., Bonnier, F., Blasco, H. *et al.* Raman spectroscopic screening of high and low molecular weight fractions of human serum. *Analyst* **144,** 4295–4311 (2019).

101. Adkins, J. N., Varnum, S.M., Auberry, K.J., Moore, R.J. *et al.* Toward a human blood serum proteome. *Mol. Cell. Proteomics* **1,** 947–955 (2002).

102. Tirumalai, R.S., Chan, K.C., Prieto, D.A., Issaq, H.J. *et al.* Characterization of the low molecular weight human serum proteome. *Mol. Cell. Proteomics* **2,** 1096–1103 (2003).

103. Rohleder, D., Kocherscheidt, G., gerber, K., Kiefer, W. *et al.* Comparison of mid-infrared and Raman spectroscopy in the quantitative analysis of serum. *J. Biomed. Opt.* **10,** 31108 (2005).

104. Rohleder, D., Kiefer, W. & Petrich, W. Quantitative analysis of serum and serum ultrafiltrate by means of Raman spectroscopy. *Analyst* **129,** 906–911 (2004).

105. Bonnier, F., Baker, M. J. & Byrne, H. J. Vibrational spectroscopic analysis of body fluids: Avoiding molecular contamination using centrifugal filtration. *Anal. Methods* **6,** 5155–5160 (2014).

106. Mahmood, T., Nawaz, H., Ditta, A., Majeed, M.I. *et al.* Raman spectral analysis for rapid screening of dengue infection. *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.* **200,** 136–142 (2018).

107. Sahu, A., Nandakumar, N. & Krishna, C. M. Recurrence prediction in oral cancers: a serum Raman spectroscopy study. *Analyst* **140,** 2294–2301 (2015).

108. Li, X., Yang, T. & Li, S. Discrimination of serum Raman spectroscopy between normal and colorectal cancer using selected parameters and regression-discriminant analysis. *Appl. Opt.* **51,** 5038–5043 (2012).

109. Ellis, D. I., Cowcher, D. P., Ashton, L., O'Hagan, S. & Goodacre, R. Illuminating disease and enlightening biomedicine: Raman spectroscopy as a diagnostic tool. *Analyst* **138,** 3871–3884 (2013).

110. *Ewing's Analytical Instrumentation Handbook*. (CRC Press, 2019).

111. Petrich, W., Lewandrowski, K.B., Muhlestein, J.B., Hammond, M.E.H. *et al.* Potential of mid-infrared spectroscopy to aid the triage of patients with acute chest pain. *Analyst* **134,** 1092–1098 (2009).

112. Hands, J.R., Dorling, K.M., Abel, P., Ashton, K.M. *et al.* Attenuated Total Reflection Fourier Transform Infrared (ATR-FTIR) spectral discrimination of brain tumour severity from serum samples. *J. Biophotonics* **7,** 189–199 (2014).

113. Backhaus, J., Mueller, R., Formanski, N., Szlama, N. *et al.* Vibrational Spectroscopy Diagnosis of breast cancer with infrared spectroscopy from serum samples. *Vib. Spectrosc. J.* **52,** 173–177 (2010).

114. Sheng, D., Wu, Y., Wang, X., Huang, D. *et al.* Comparison of serum from gastric cancer patients and from healthy persons using FTIR spectroscopy. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **116,** 365–369 (2013).

115. Hume, S., Hithell, G., Greetham, G.M., Donaldson, P.M. *et al.* Measuring Proteins in H2O with 2D-IR Spectroscopy. *Chem. Sci.* **10,** 6448–6456 (2019).

116. Baiz, C. R., Reppert, M. & Tokmakoff, A. Introduction to Protein 2D IR Spectroscopy. *Ultrafast Infrared Vib. Spectrosc.* 361–404 (2012).

117. Baiz, C. R., Peng, C. S., Reppert, M. E., Jones, K. C. & Tokmakoff, A. Coherent two-dimensional infrared spectroscopy: Quantitative analysis of protein secondary structure in solution. *Analyst* **137,** 1793–1799 (2012).

118. Fritzsch, R., Hume, S., Minnes, L., Baker, M.J. *et al.* Two-dimensional infrared spectroscopy: An emerging analytical tool? *Analyst* **145,** 2014–2024 (2020).

119. Hithell, G., Shaw, D.J., Donaldson, P.M., Greetham, G.M. *et al.* Long-range vibrational dynamics are directed by Watson-Crick base pairing in duplex DNA. *J. Phys. Chem. B* **120,** 4009–4018 (2016).

120. Bredenbeck, J., Helbing, J., Kolano, C. & Hamm, P. Ultrafast 2D-IR spectroscopy of transient species. *ChemPhysChem* **8,** 1747–1756 (2007).

121. Hunt, N. T. Transient 2D-IR spectroscopy of inorganic excited states. *Dalt. Trans.* **43,** 17578–17589 (2014).

122. Minnes, L., Shaw, D.J., Cossins, B.P., Donaldson, P.M. *et al.* Quantifying secondary structure changes in calmodulin using 2D-IR spectroscopy. *Anal. Chem.* **89,** 10898–10906 (2017).

123. Adamczyk, K., Simpson, N., Greetham, G.M., Gumiero, A. *et al.* Ultrafast infrared spectroscopy

reveals water-mediated coherent dynamics in an enzyme active site. *Chem. Sci.* **6,** 505–516 (2015).

124. Kropman, M. F., Nienhuys, H. K., Woutersen, S. & Bakker, H. J. Vibrational relaxation and hydrogen-bond dynamics of HDO:H2O. *J. Phys. Chem. A* **105,** 4622–4626 (2001).

125. Simpson, N. & Hunt, N. T. Ultrafast 2D-IR spectroscopy of haemoproteins. *Int. Rev. Phys. Chem.* **34,** 361–383 (2015).

126. Hunt, N. T. 2D-IR spectroscopy: Ultrafast insights into biomolecule structure and function. *Chem. Soc. Rev.* **38,** 1837–1848 (2009).

127. Kim, Y. S. & Hochstrasser, R. M. Applications of 2D IR spectroscopy to peptides, proteins, and hydrogen-bond dynamics. *J. Phys. Chem. B* **113,** 8231–8251 (2009).

128. Baiz, C. R., Reppert, M. & Tokmakoff, A. Amide I two-dimensional infrared spectroscopy: Methods for visualizing the vibrational structure of large proteins. *J. Phys. Chem. A* **117,** 5955–5961 (2013).

129. 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.* **137,** 184202 (2012).

130. Hamm, P., Lim, M. & Hochstrasser, R. M. Structure of the Amide I band of peptides measured by femtosecond nonlinear-infrared spectroscopy. *J. Phys. Chem. B* **102,** 6123–6138 (1998).

131. Hamm, P., Lim, M. & Hochstrasser, R. M. Structure of the amide I band of peptides measured by femtosecond nonlinear-infrared spectroscopy. *J. Phys. Chem. B* **102,** 6123–6138 (1998).

132. Demirdöven, N., Cheatum, C., Chung, H.S., Khalil, M. *et al.* Two-dimensional infrared spectroscopy of antiparallel beta-sheet secondary structure. *J. Am. Chem. Soc.* **126,** 7981–7990 (2004).

133. Strasfeld, D. B., Ling, Y. L., Gupta, R., Raleigh, D. P. & Zanni, M. T. Strategies for extracting structural information from 2D IR spectroscopy of amyloid: Application to islet amyloid polypeptide. *J. Phys.* **113,** 15679–15691 (2009).

134. Sam Kim, Y., Liu, L., Axelsen, P. H. & Hochstrasser, R. M. 2D IR provides evidence for mobile water molecules in Beta-amyloid fibrils. *Proc. Natl. Acad. Sci.* **106,** 17751–17756 (2009).

135. Shim, S.H., Gupta, R., Ling, Y.L., Strasfeld, D.B. *et al.* Two-dimensional IR spectroscopy and isotope labeling defines the pathway of amyloid formation with residue-specific resolution. *Proc. Natl. Acad. Sci.* **106,** 6614–6619 (2009).

136. Middleton, C. T., Woys, A. M., Mukherjee, S. S. & Zanni, M. T. Residue-specific structural kinetics of proteins through the union of isotope labeling, mid-IR pulse shaping, and coherent 2D IR spectroscopy. *Methods* **52,** 12–22 (2010).

137. Zhang, T. O., Alperstein, A. M. & Zanni, M. T. Amyloid β-sheet secondary structure identified in UV-induced cataracts of porcine lenses using 2D IR spectroscopy. *J. Mol. Biol.* **429,** 1705–1721 (2017).

138. Wang, L., Middleton, C.T., Singh, S., Reddy, A.S. *et al.* 2DIR spectroscopy of human amylin fibrils reflects stable β-sheet structure. *J. Am. Chem. Soc.* **133,** 16062–16071 (2011).

139. Moran, S. D. & Zanni, M. T. How to get insight into amyloid structure and formation from infrared spectroscopy. *J. Phys. Chem. Lett.* **5,** 1984–1993 (2014).

140. Middleton, C. T., Buchanan, L. E., Dunkelberger, E. B. & Zanni, M. T. Utilizing lifetimes to suppress random coil features in 2D IR spectra. *J. Phys. Chem. Lett.* **2,** 2357–2361 (2011).

141. 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* **119,** 14065–14075 (2015).

142. Bredenbeck, J., Helbing, J., Nienhaus, K., Nienhaus, G. U. & Hamm, P. Protein ligand migration mapped by nonequilibrium 2D-IR exchange spectroscopy. *Proc. Natl. Acad. Sci.* **104,** 14243–14248 (2007).

143. Fritzsch, R., Donaldson, P.M., Greetham, G.M., Towrie, M. *et al.* Rapid screening of DNA-ligand complexes via 2D-IR spectroscopy and ANOVA-PCA. *Anal. Chem.* **90,** 2732–2740 (2018).

144. Shaw, D. J., Robb, K., Vetter, B.V., Tong, M. *et al.* Disruption of key NADH-binding pocket residues of the Mycobacterium tuberculosis InhA affects DD-CoA binding ability. *Sci. Rep.* **7,** 1–7 (2017).

145. Johnson, P. J., Kozoil, K. L. & Hamm, P. Quantifying bomolecular recognition with site-specific 2D infrared probes. *J. Phys. Chem. Lett.* **8,** 2280–2284 (2017).

146. Alperstein, A. M., Ostrander, J. S., Zhang, T. O. & Zanni, M. T. Amyloid found in human cataracts with two-dimensional infrared spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **116,** 6602–6607 (2019).

147. Petersen, J., M O Ller, K. B., Rey, R. & Hynes, J. T. Ultrafast librational relaxation of H2O in liquid water. *J. Phys. Chem. B* **117,** 4541–4552 (2013).

148. Huse, N., Ashihara, S., Nibbering, E. T. J. & Elsaesser, T. Ultrafast vibrational relaxation of O – H

bending and librational excitations in liquid H2O. *Chem. Phys. Lett.* **404,** 389–393 (2005).

149. Rey, R., Møller, K. B. & Hynes, J. T. Ultrafast vibrational population dynamics of water and related systems : A Theoretical Perspective. *Chem. Rev.* **104,** 1915–1928 (2004).

150. Nibbering, E. T. J. & Elsaesser, T. Ultrafast vibrational dynamics of hydrogen bonds in the condensed phase. *Chem. Rev.* **104,** 1887–1914 (2004).

151. Møller, K. B., Rey, R. & Hynes, J. T. Hydrogen bond dynamics in water and ultrafast infrared spectroscopy : A theoretical study. *J. Phys. Chem. A* **1018,** 1275–1289 (2004).

2. Theory & Methods

2.1. IR Spectroscopy

IR spectroscopy is often used to identify the chemical composition of molecules and utilises the fact that the frequency and intensity of IR light absorption is indicative of, or influenced by, molecular structure.[1] Interatomic bond vibrations and the surrounding solvent environment all contribute to the frequencies of IR light absorbed and are thus important in understanding the behaviour and function of a given molecular species. Each experimental chapter in this thesis utilises infrared spectroscopy techniques and how they can be used to investigate protein structure and their dynamics. This chapter will focus on describing the methods used, their theoretical basis and experimental setups.

2.1.1. Infrared Absorption Spectroscopy Theory

The absorption of light by a molecule depends on the frequency of the electromagnetic radiation and fundamental vibrational frequencies of the molecules that make up the sample being studied.[2] The frequency of the excitation field must be resonant with the difference in energy between two vibrational energy levels in order to induce a polarisation of charges within a dielectric medium. This induced macroscopic polarisation creates an oscillating dipole moment in the sample being studied at a frequency matching that of the excitation radiation. This oscillation of the polarisation results in the emission of a signal from the molecule which recombines with the excitation field.[3]

The resulting signal will experience a $\pi/2$ shift in phase relative to the polarisation of the excitation field, producing destructive interference between the two fields corresponding to the bond frequencies.[4] The polarisation, $P(t)$, induced by an incident electric field, E(t), in a dielectric medium is described as:

$$P(t) = \varepsilon_0 \chi E(t) = \varepsilon_0 \left(\chi^{(1)} E(t) + \chi^{(2)} E(t)^2 + \chi^{(3)} E(t)^3 + \cdots \chi^{(n)} E(t)^n \right) \qquad (2.1)$$

where ε_0 is the permittivity of free space and χ is the optical susceptibility of the dielectric medium. The first susceptibility term, $\varepsilon_0 \chi^{(1)} E(t)$, is only considered for the case of linear optics. The higher order terms require higher intensity electric fields to make these terms substantial, and so as they are significantly smaller than the first term they are ignored for the linear case.

As χ is directly related to the polarisability of the medium, in turn it is also associated with the transition dipole moment corresponding to the fundamental transition $\nu = 0 - 1$ (μ_{01}).[4] Thus following interaction with the excitation field, the macroscopic polarisation is dependent on the excitation of molecules from the ground to first excited vibrational states.

The resulting emission of the signal field can be described by both the classical and quantum mechanical descriptions of light-matter interactions. Classically, it is the absorption of light resonant to the vibrational bonds resulting in the vibration of oscillating charges that produces the emission of the phase shifted signal field, $E_{sig}(t)$. This in turn causes destructive interference with the excitation field, $E(t)$, at the frequencies specific to the bond vibrations allowing an absorption spectrum to be produced.[3]

However in the quantum description when the excitation field interacts with the sample ensemble of oscillators, the ground and first excited states become coupled by excitation of the *bra* or *ket* which creates a coherent superposition between these two states. This produces the macroscopic polarisation in the medium which oscillates at the frequency corresponding to the fundamental transition frequency resulting in the emission signal at that frequency. Similar to the classical description, the emitted signal is phase shifted 90° relative to the polarisation field which causes the signal to destructively interfere with the excitation field at the fundamental transition frequencies of the molecule.[5] The macroscopic polarisation, $P(t)$, is time dependent and evolves according to a molecular response function, $R(t)$. This response function contains the unique information on the molecular structure being studied and the Fourier transform of $R(t)$ can extract the complex signal in the frequency domain, $S(\omega)$, containing both real and imaginary components. The real component contains information on the absorptive part of the lineshape, $A(\omega)$, where the imaginary component provides information of the dispersive vibrational lineshape $D(\omega)$.[4]

The phase of $D(\omega)$ and $A(w)$ is such that only $A(\omega)$ interferes with $E(t)$. The dispersive and absorptive vibrational lineshape components are given by the following equations:

$$D(\omega) \propto -\mu_{01}^2 \frac{\omega - \omega_{01}}{(\omega - \omega_{01})^2 + (0.5\gamma)^2} \tag{2.2}$$

$$A(\omega) \propto \mu_{01}^2 \frac{0.5\gamma}{(\omega - \omega_{01})^2 + (0.5\gamma)^2} \tag{2.3}$$

where ω is the frequency of the emitted light, ω_{01} is the frequency of the fundamental transition and γ is the natural linewidth of the vibrational transition.

2.1.1.1. Spectral Lineshapes

The natural linewidth of a vibration is intrinsic to that vibration and cannot be narrower than dictated by its vibrational lifetime; this takes shape of a Lorentzian frequency distribution (Figure 2.1) and is known as the homogeneous linewidth.[4,6]

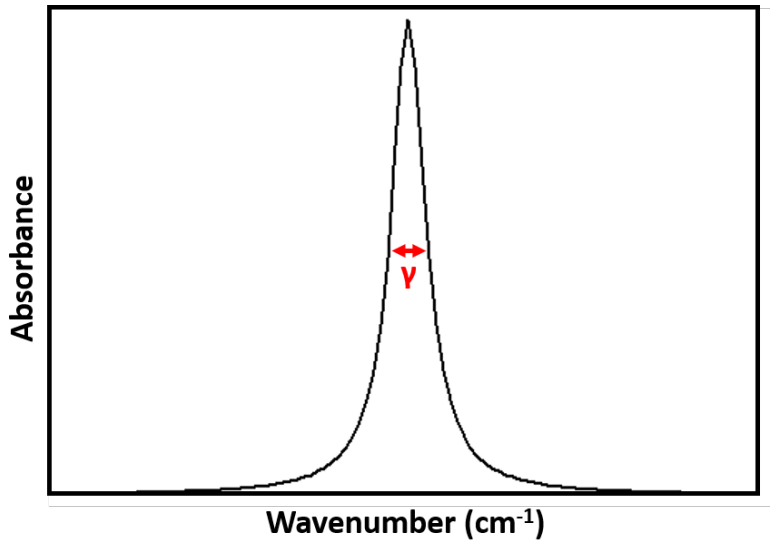

Figure 2.1: *A Lorentzian frequency distribution showing the natural homogeneous linewidth γ.*

The population of the first excited state is evolving with a finite lifetime so as the Heisenberg uncertainty principle dictates, there will be a linked uncertainty in its energy:

$$\partial E \cdot \partial t \geq \frac{h}{4\pi} \tag{2.4}$$

where ∂E is the uncertainty in the energy of the excited vibrational energy level, ∂t is the time spent in the excited state (vibrational lifetime) and h is Planck's constant. The frequency of the oscillator is directly related to the difference in energy between the initial and excited states and so an uncertainty in the energy will have a direct impact on the oscillation frequency uncertainty:

$$\partial E = hc \cdot \partial v \tag{2.5}$$

Substitution of this definition of ∂E into equation (2.4) yields:

$$\partial v \geq \frac{1}{4\pi c \partial t} \tag{2.6}$$

Therefore the error in the frequency of vibration is directly associated with the vibrational lifetime[7] and as such the vibrational frequency of a single oscillation may occupy a narrow range of frequencies. The induced macroscopic polarisation evolves with a characteristic lifetime and is subject to dephasing. The process of dephasing is a loss of coherence in the oscillations of the molecules which were excited at the same time, and this causes a loss in the macroscopic polarisation.[8] This process is known as 'pure-dephasing' and occurs at a timescale of T_2^*.[4,9] This combined with the lifetime of the vibration, T_1, dictates the minimum homogeneous linewidth of a vibrational transition:

$$\gamma \sim \frac{1}{2T_1} + \frac{1}{T_2^*} = \frac{1}{T_2} \tag{2.7}$$

where, T_2 is the total homogeneous dephasing time. If the linewidth of a vibrational transition is determined solely by T_2 then the distribution is said to be 'homogeneously broadened'.[10,11]

However this is not typically the case and other contributions to the linewidth are often present meaning that a much broader linewidth is observed in real spectra.[4,10,11]

Samples in solution phase, which are used throughout this thesis, are subject to the effects of their local environment which plays a significant role in the molecular dephasing.

Interactions such as hydrogen bonding from the solvent molecules offers a wealth of potential conformations available which would alter the oscillator's potential well and in turn adjust the frequency of the fundamental $v = 0 - 1$ transition.[12,13]

For each individual oscillator, its instantaneous frequency will therefore fall within a probability distribution with a Gaussian profile computed from the many structural conformations available from solvent interactions (Figure 2.2). When considering a molecular ensemble of oscillators it is likely that each of these many conformations are occupied producing a range of vibrational frequencies. The resulting effect is absorption across this range of frequencies that is significantly broader than the natural linewidth of the oscillators. In solution phase samples, the resulting Gaussian spectral profile is principally made up of many overlapping Lorentzian distributions from each individual oscillator in the molecular ensemble. This Gaussian profile linewidth is known as being 'inhomogeneously broadened'.[4,10,14]

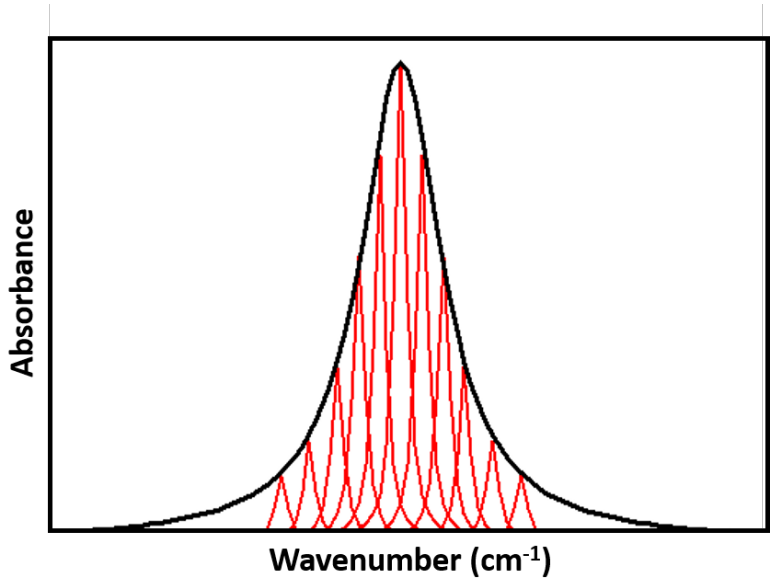

Figure 2.2: *A Gaussian frequency distribution of a molecular ensemble showing a series of individual oscillator Lorentzian profile contributions.*

The environment surrounding single oscillators however is constantly changing with the rapid creation and breaking of hydrogen bonds from water molecules meaning that as oscillators experience a change in their surroundings they experience fluctuations in their fundamental transition frequency with time.[13] Thus, the molecular ensemble is subject to intrinsic variations in frequency which affects all molecules in the same way (homogeneous

broadening) and a range of oscillation frequencies caused by the environment of which each molecule may experience differently (inhomogeneous broadening). This evolution of vibrational frequencies with time is what is known as 'spectral diffusion' and cannot be measured using continuous wave IR absorption techniques.[4] This process of spectral diffusion will be discussed in more detail when discussing 2D-IR lineshapes later in this chapter.

2.1.2. Fourier Transform Infrared Spectroscopy

An Infrared spectrum depicts how IR light interacts with matter. Fourier transform Infrared (FTIR) spectroscopy employs a blackbody radiation source, a Michelson Interferometer and a He:Ne laser to record absorbance as a function of IR frequency. Production of the mid-IR radiation is done by passing a current through a silicon carbide element resulting in heating and continuous emission of IR radiation in the frequency range 16.7-1.4 µm (600-7000 cm^{-1}). In order to create an IR beam, as light from the source is emitted in all directions, it is passed through an aperture and collimated via a parabolic mirror before being directed through a Michelson interferometer,[15] as shown in Figure 2.3.

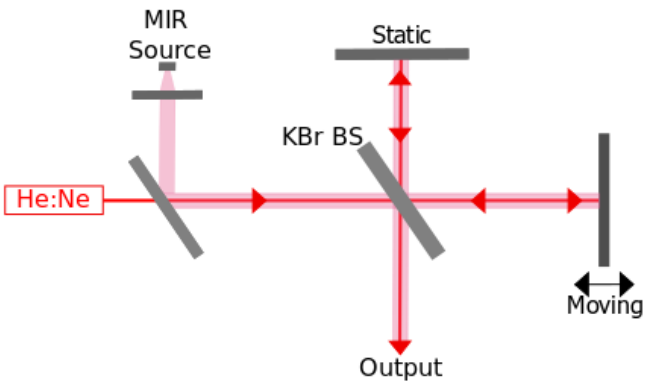

Figure 2.3: *Schematic of the He:Ne (red) and MIR (pink) progressing through the Michelson interferometer used in FTIR spectroscopy.*

The collimated light is incident on a potassium bromide (KBr) beamsplitter which is used to split the IR beam into two paths set 90° to one another. One path of the interferometer has a stationary mirror while the second path utilises a motorised delay stage to change the path length of the second beam. The light beams from the two arms are then recombined at the beam splitter and the intensity profile of this recombined beam is directly dependent on the

exact path difference between the two interferometer arms at any given time. Different frequencies will interfere either constructively or destructively thus creating a time dependent amplitude profile of the IR light - an interference pattern (or interferogram) which is measured at the detector.[2,15,16] To correctly transform this signal into the frequency domain the exact difference in path length between the two beams must be known and is calculated via the use of a He:Ne laser. The He:Ne is also propagated through the Michelson interferometer, experiencing the same optical path difference as the IR light and upon recombination generates a series of 'fringes'. These fringes can be counted by the spectrometer providing a reference for the path difference thus producing the frequency axis.

Samples studied in this thesis were housed between two calcium fluoride CaF_2 windows prior to IR investigation, as shown in Figure 2.4.

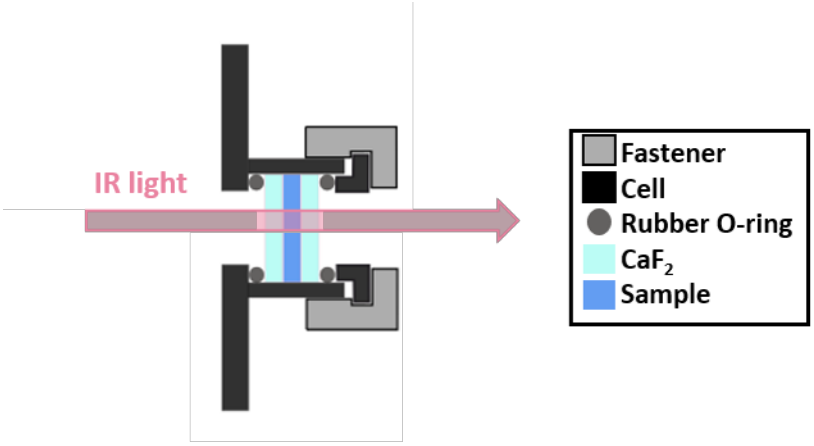

Figure 2.4: *Cross section of a typical cell set-up. The process of controlling the sample thickness is done by slowly tightening the fastening ring onto the cell body while taking FTIR measurements.*

The relative intensity of the IR light before and after passing through the sample is used to calculate the absorbance, (A), and is calculated as follows:

$$A = log_{10} \frac{I_o}{I} \tag{2.9}$$

where I_o and I are the intensities of light before and after interaction with the sample, respectively. As the absorbance is a ratio of the amount of light before and after it interacts with the sample it is unitless. In practice, the light intensity before sample interaction is taken as a 'background' or reference spectrum when no sample is in place.

Absorption from the sample is further calculated using the Beer-Lambert law:

$$A = \varepsilon c l \tag{2.10}$$

where ε is the molar extinction coefficient (L mol^{-1} cm^{-1}), c is the concentration of molecules (mol L^{-1}) and l is the path length or thickness of samples used (cm).[6,16,17] An example FTIR spectrum of water, is shown in Figure 2.5.

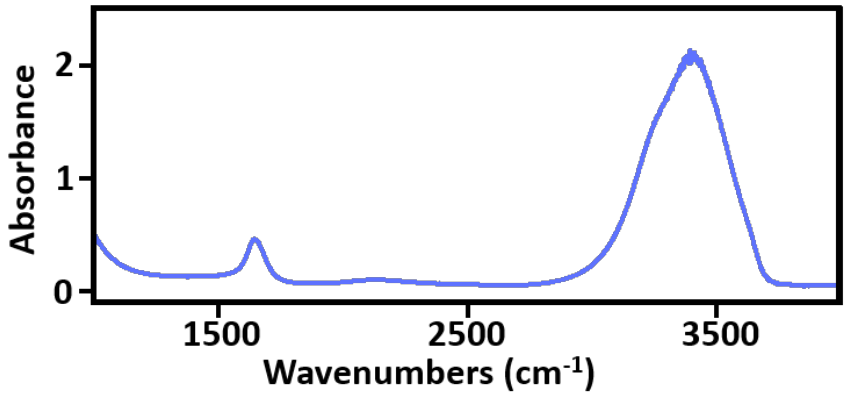

Figure 2.5: *An example FTIR spectrum obtained for H_2O.*

2.2. Non-linear Spectroscopy

So far only linear spectroscopy has been discussed, where the polarisation and response functions produced have linear dependence with the excitation field and thus only requires consideration of the first order optical susceptibility term, $\chi^{(1)}$(Equation 1.3). However, the vibrational dynamics of proteins occur rapidly, on a femtosecond (fs) to picosecond (ps) timescale,[18] meaning that laser pulses with durations faster than these dynamics are required. As a consequence of using such fast laser pulses, strong laser fields are produced which results in non-linear optical responses.[1,4] Thus the macroscopic polarisation can no longer be considered linear and now contributions from the higher order terms must be considered, resulting in production of a non-linear polarisation. A discussion of the theory and experimental techniques for pump-probe and two-dimensional infrared spectroscopies will follow an introduction to non-linear spectroscopy.

The measured response function, $R(t)$, will now primarily contain contributions from the first ($\chi^{(1)}$) and third ($\chi^{(3)}$) terms, other higher order odd-power terms are insufficient enough to not be considered here. The even power terms do not contribute to this response, as the sample does not contain a point of inversion symmetry (the samples are not centrosymmetric crystals) and so the even terms of $R(t)$ are equal to zero (e.g. $\chi^{(2n)} = 0$).

Typically, the 3rd order signals are extracted from a medium via a 'four-wave mixing' interaction.[4] This involves three excitation fields (laser pulses, E_1, E_2 and E_3) incident on a material (the molecular ensemble) which are separated in time by the intervals t_1, t_2 and t_3. As a result of using three different pulses, the response function produced from these interactions must account for the multiple orderings in which the pulses may interact with the ensemble. Thus, a more complicated response function (Equation 2.11) is observed when compared to linear spectroscopy and becomes convoluted with each field to produce the following macroscopic polarisation.

$$P^{(3)}(t) \propto \iiint_0^\infty E_1(t - t_1 - t_2 - t_3)E_2(t - t_1 - t_2)E_3(t - t_3)R^{(3)}(t_1, t_2, t_3)\, dt_1 dt_2 dt_3 \quad (2.11)$$

The three laser excitation fields and the emitted field are often referred to by their wavevectors, $\vec{k}$, which describes their rate of change of phase along the direction of propagation $\left(\frac{2\pi}{\lambda}\right)$. The fields are therefore referred to as $\vec{k}_1$, $\vec{k}_2$, $\vec{k}_3$, and $\vec{k}_4$ denotes the emitted signal, where the numbered subscript indicates the order in time of each field. Each of the excitation fields interacts with the ensemble independently and imprints a polarisation with their wavevector, this dictates the direction of the emitted field, $\vec{k}_4$ through the following equation.[1,4,19,20]

$$\vec{k}_1 + \vec{k}_2 + \vec{k}_3 + \vec{k}_4 = 0 \quad (2.12)$$

Deconvolution of the 3rd order response from the excitation fields can be done using phase matching of the pulse sequence.[4,20] This technique allows specific pathways of the molecule to be measured depending on the geometry of the excitation fields. However, phase-matching of the pulse sequence requires both statistical and quantum mechanics to describe the molecular ensemble of oscillators and so the ensemble is represented by a density matrix describing its statistical average. When one of the excitation pulses interacts with the

ensemble, the transition dipole operator, μ, is applied to the density matrix. Feynman diagrams can be used to represent the temporal evolution of the density matrix with the three excitation pulses (Figure 2.6).[4,21] The excitation and de-excitation of the *bra* and *ket* states of the ensemble for a series of three excitation pulses creates many different possible pathways through which the density matrix may evolve. Positive and negative wavevectors ($\pm \vec{k}$) excite the *ket* and *bra* states of the ensemble, respectively. The wavevectors can also de-excite the opposing *bra* or *ket* terms, meaning that positive and negative wave-vectors de-excite the *bra* and *ket* states, respectively.[4]

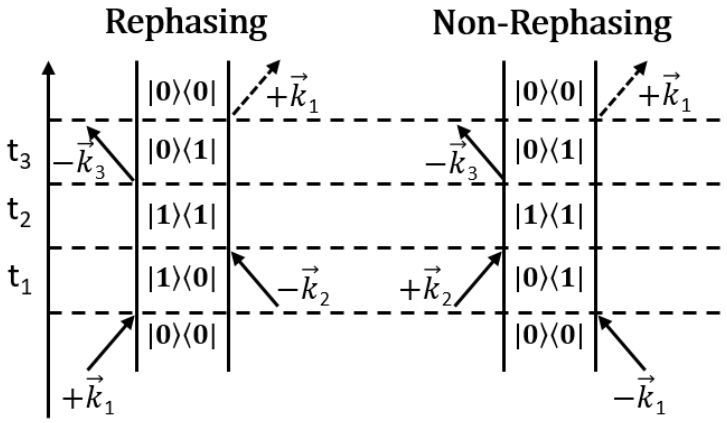

Figure 2.6: *Feynman diagrams of both the rephasing and non-rephasing pathways for a four wave mixing interaction.*

There are two types of pathways that can be produced and these are known as 'rephasing' and 'non-rephasing' pathways (Figure 2.6).[4] A 'rephasing' pathway requires excitation of either the *bra* or the *ket* state of the ensemble and the signal field, $\vec{k}_4$, must be emitted by de-excitation of the opposite state. So if $\vec{k}_1$ causes excitation of the *ket (bra)* state of the ensemble and the signal field $\vec{k}_4$ is emitted by de-excitation of the *bra (ket)* state of the ensemble, resulting in a rephasing pathway. A 3rd order response function that contains a rephasing pathway is able to restore the macroscopic polarisation induced by the initial excitation field, $\vec{k}_1$, and so emits a 'photon-echo' signal, which occurs whenever inhomogeneous broadening occurs. A 'non-rephasing' pathway is one that does not fit the criteria for a 'rephasing' pathway and after excitation of either the *bra* or the *ket* state of the

ensemble, the signal field, $\vec{k}_4$, is emitted by de-excitation of the same state. These Feynman diagrams will be used to help describe pump-probe and 2D-IR measurements below.

2.2.2. Pump-Probe Spectroscopy

Pump-probe (PP) spectroscopy is an absorption-difference technique and is carried out with only two interaction pulses.[4] However it can fully be described as a 'four-wave mixing' interaction where both $\vec{k}_1$ and $\vec{k}_2$ pulses are overlapped temporally and are degenerate. This combined pulse is referred to as the 'pump' pulse and $\vec{k}_3$ is known as the 'probe' pulse. As the $\vec{k}_1$ and $\vec{k}_2$ pulses now overlap in time during this four wave mixing interaction, $t_1 = 0$ and separation of these excitation pathways is not possible and so the interaction of the pump pulse with a molecular ensemble automatically creates a population state. Figure 2.7 shows the pulse sequence utilised.

The probe pulse, $\vec{k}_3$, can therefore interact with either a $|0\rangle\langle 0|$ or a $|1\rangle\langle 1|$ population state following the time between the pump and probe pulses, t_2, known as the pump-probe delay time. As t_2 increases the probability of relaxation of the molecules back into the initial ground state, $|0\rangle\langle 0|$, increases. During a PP experiment the pump and probe pulse separation, t_2, can be controlled and altered throughout the experiment allowing measurement of vibrational lifetimes. When $t_2 = 0$ the pump and probe pulses are temporally overlapped, any molecules that were excited from the ground to the first excited state by the pump pulse will still be there as relaxation will not have yet occurred. When t_2 is lengthened the probability of spontaneous relaxation from the first to the ground excited state is increased and so the absorption difference will become smaller with increasing t_2.[19] The final coherent state in the ensemble created by interaction with the probe pulse is therefore dependent on the amount of relaxation that has occurred during t_2 thus allowing for direct measurements of the vibrational lifetime of a given vibrational transition.

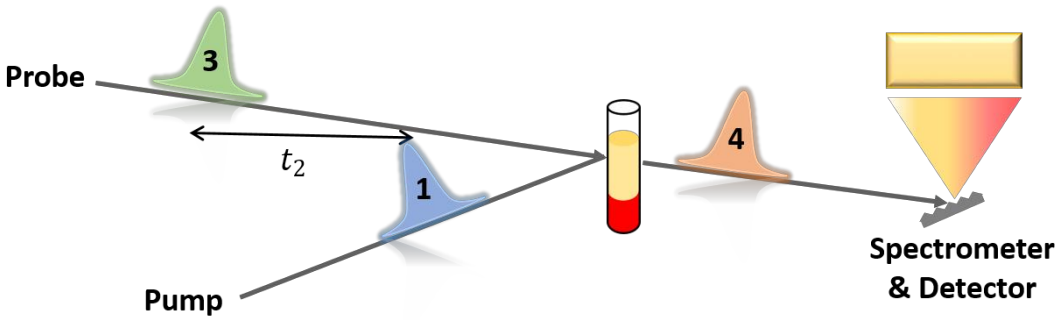

Figure 2.7: *The pulse sequence used for IR pump-probe spectroscopy, the pump pulse (1, blue) is temporally separated from the probe pulse (3, green) by the delay time t_2. The emitted signal (4, orange) is heterodyned with the residual probe light and dispersed by a grating prior to detection.*

The wavevectors of the three interaction pulses dictates the direction of the emitted field, $\vec{k}_4$. As only one pump pulse is used such that $\vec{k}_1 = \vec{k}_2$, both the rephasing and non-rephasing pathways are emitted in the same direction, simplifying the experimental setup in order to detect the complete signal.[4,11] A further benefit of this geometry is that the signal field, $\vec{k}_4$, is emitted collinearly with $\vec{k}_3$, which allows $\vec{k}_3$ to act as a local oscillator and self-heterodyne the signal. Self-heterodyning is the convolution of the emitted signal $\vec{k}_4$ with $\vec{k}_3$, the field responsible for the signal emission.[22] This allows the signal to be detected much more readily as the isolated signal itself is weak.

PP spectroscopy is an absorption-difference technique, meaning that for every pump-probe measurement taken another measurement is acquired when the pump pulse is blocked, essentially taking a background reading for each measurement. The difference between the probe in the presence and absence of the initial pump pulse results in the change of absorption due to the pump. The 'pump-off' result is subtracted from the 'pump-on' measurement yielding the final PP spectrum resulting in the change of absorption due to the pump (Figure 2.8).[4,23]

As a result of the initial excitation created by the pump, the probe pulse experiences a reduced absorption for the $v = 0 - 1$ transition owing to the depleted population of molecules occupying the ground state. The probe pulse may also cause population of the second vibrational level through excitation of the first excited state. Where the potential is anharmonic, the higher $v = 1 - 2$ transition will occur at a slightly lower frequency to the

$v = 0 - 1$ vibration potentially allowing for both transitions to be resolved separately (Figure 2.8 (blue)). Once t_2 is sufficiently large enough to allow complete relaxation then no absorption differences are measured (Figure 2.8 (red)).

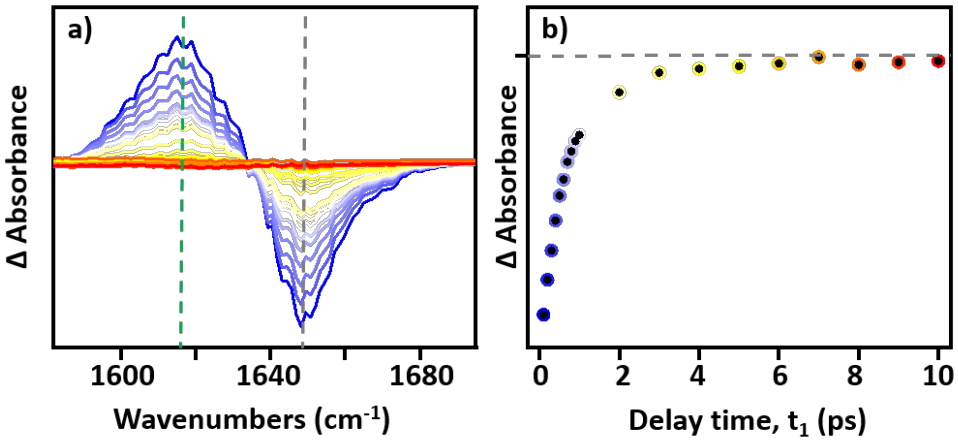

Figure 2.8: *An example dataset showing a) pump-probe spectra of bovine serum albumin. Each pixel collects data as function of* t_2*. Spectra shown from 0-1 ps in 0.1 ps increments (blue) to 2-10 ps in 1 ps increments (yellow-red). The separation of the* $v = 0 - 1$ *(grey dashed line) and* $v = 1 - 2$ *(green dashed line) transitions of a molecule are clearly visible due to the anharmonicity.* b) Shows the peak height at 1648 cm^{-1} (grey dashed line) as a function of the pump-probe delay time, t_1.

2.2.3. Two-Dimensional Infrared Spectroscopy

Two-Dimensional Infrared (2D-IR) spectroscopy builds on PP spectroscopy by utilising the three interaction pulses and controlling their sequence, allowing frequency resolution of the pump axis to be achieved. Spreading this information obtained from the molecular response over a second frequency axis provides a much more detailed picture of the molecule being examined when compared to both PP and IR absorption spectroscopies. Creation of the pump frequency axis in 2D-IR can be achieved using either the 'frequency-domain' or the 'time-domain' methodologies. The simpler experimental approach is the frequency-domain technique which utilises narrow bandwidth pulses where the centre frequency is scanned to generate the second frequency axis.[24] However as a result of the narrow bandwidth, the duration of the pulses is increased which in turn reduces the temporal resolution. With

biological systems having short-lived responses, ~1 ps,[24] the ideal approach must have short durations in order to extract more information; this is where acquisition in the time domain becomes the better approach for biological systems and has been used in all experiments throughout this thesis.

The time-domain method measures the third order response function of the ensemble using the four wave mixing approach and utilises broadband pump pulses.[4,25]

Interaction of the pulses with the molecular ensemble is as follows and the pulse sequence is shown in Figure 2.9:

- Initially the ensemble is in the ground vibrational state.
- The first pulse, $\vec{k}_1$, interacts with the ensemble, exciting either the *bra* or *ket*, this creates a coherent superposition of the ground and first excited states, which oscillates with a frequency ω_{01}.
- During t_1 between the first and second pulses, the macroscopic polarisation created by $\vec{k}_1$, decays as a result of dephasing between the molecules.
- The second pulse, $\vec{k}_2$, converts the coherent superposition into a population state by either de-excitation of the previously excited *bra* or *ket* or excitation of the opposite state to that initially excited.
- During t_2 between the second and third pulses, the population state evolves in time. This can be due to vibrational relaxation, spectral diffusion or energy transfer between coupled modes.
- The third pulse, $\vec{k}_3$, induces a second coherent superposition of states, restoring the macroscopic polarisation of the system. For rephasing pathways, the coherence between the molecule returns after a time t_3, this time is approximately equal to the time t_1, as the time taken to restore coherence of the molecules is equivalent to reversing the initial dephasing.
- Following t_3 the signal field $\vec{k}_4$, is emitted from the ensemble.

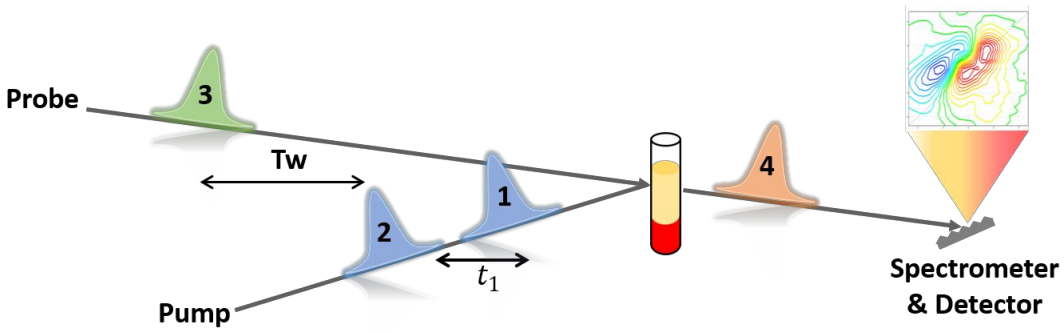

Figure 2.9: *The pulse sequence used for 2D-IR spectroscopy, the pump pulses (1,2, blue) are separated by time* delay, t_1, and are *temporally separated from the probe pulse (3, green) by* T_w*. The emitted signal (4, orange) is heterodyned and dispersed by a grating prior to detection.*

The second frequency axis is constructed through control of the temporal delay between the first two excitation fields, $\overrightarrow{k_1}$ and $\overrightarrow{k_2}$. This can be done through the use of a Michelson Interferometer, which uses a variable optical delay line between the pump pulses or alternatively the more recently developed 'pulse-shaper'. The pulse-shaper utilises acousto-optical modulation to replicate the effects of an interferometer on a much shorter timescale. Both methods are used throughout this thesis and will be described in more detail under the 2D-IR instrumentation section later in this chapter. It is by recording the emitted signal field as a function of the time delay, t_1, and performing a Fourier transform that the frequency domain data can be recovered.

The wavevectors of the three interaction pulses dictate the direction of the emitted field, $\vec{k}_4$. As in PP spectroscopy, the resulting signal, $\vec{k}_4$, is emitted collinearly with $\overrightarrow{k_3}$, allowing $\vec{k}_3$ to act as a local oscillator and self-heterodyne the signal. The PP geometry is used such that $\vec{k}_1 = \vec{k}_2$, which results in both the rephasing and non-rephasing pathways to be emitted in the same direction. Collection of both the rephasing and non-rephasing pathways is necessary for separation of linewidths and for preventing distortion of lineshapes. They produce a 'purely absorptive' 2D-IR spectrum which preserves the sign of the measured response function. The phase and time dependence of the signal is required to perform the

Fourier transform needed to obtain 2D-IR spectra and so by self-heterodyning this phase information is retained when the signal is collected, ensuring that spectra produced can be correctly phased. Phasing of 2D-IR spectra will be discussed in detail later in this chapter.

2.2.3.1. 2D-IR Spectrum Peaks

The linear absorption spectrum (Figure 2.10(a)) of an example molecule with two coupled vibrational modes is shown and is accompanied by a 2D-IR spectrum (Figure 2.10(b)), of which it is clear that significantly more peaks appear in the 2D-IR spectrum. In order to allow possible excitation of multiple modes the bandwidth of the laser pulse must be large enough to excite both modes. Assignment of the peaks is done with reference to the energy level diagram (Figure 2.10(c)) where the ground state is denoted as $|00\rangle$, the first excited state of each of the two modes as $|01\rangle$ and $|10\rangle$, the two second excited states as $|02\rangle$ and $|20\rangle$ and the combination band between the two modes as $|11\rangle$.

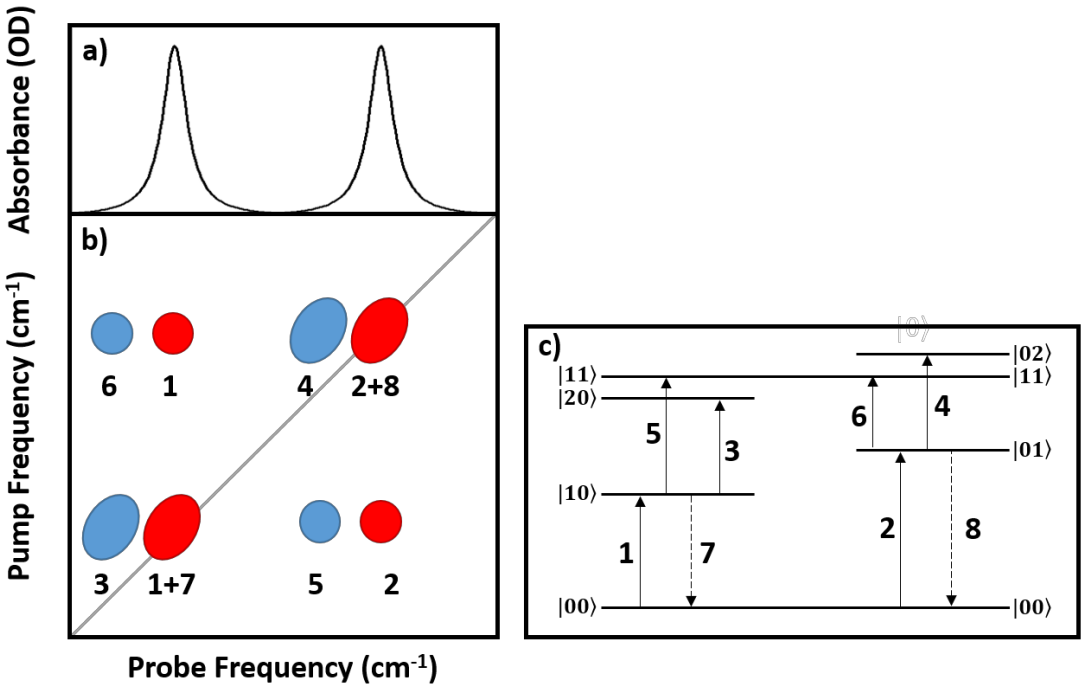

Figure 2.10: *a) A schematic IR absorption spectrum of a molecule with coupled vibrational modes b) A schematic 2D-IR spectrum of the same molecule collected at an arbitrary value of t_2. The red on-diagonal features denote the $v = 0 - 1$ transition and the $v = 1 - 2$ transition is shown in blue. c) Energy level diagram indicating excitation pathways of the molecule to create the features in the example spectrum.*

In the example spectrum shown, there are two regions that must be considered – the on-diagonal region and the off-diagonal region. The spectrum diagonal is where the probe frequency equals the pump frequency (Figure 2.10 (b, grey line) and the off-diagonal is anywhere else in the spectrum.

Firstly the on-diagonal red peaks labelled (1+7) and (2+8) will be considered. In terms of the Feynman pathways discussed, these features are the result of a pathway which contains the same coherence between the ground and first excited vibrational states (i.e. $|00\rangle\langle 01|$ or $|01\rangle\langle 00|$) during both t_1 and t_3. As a result of recording the absorption difference the probe pulse records a reduced absorption from the $v = 0 - 1$ transition which results in a negative peak on the 2D-IR spectrum diagonal.[4,21] These diagonal peaks belonging to the $v = 0 - 1$ transitions are also seen in the linear IR absorption spectrum (Figure 2.10(a)).

Accompanying each on-diagonal feature is a second feature with opposite sign to the on-diagonal feature and is shown by blue peaks labelled (3) and (4) in Figure 2.10(b). These features are the result of pathways that create a population state of the $v = 1$ vibrational energy level during t_2 (i.e. $|10\rangle\langle 10|$ or $|01\rangle\langle 01|$) allows the probe interaction to generate a coherence between $v = 1 - 2$ during t_3 (i.e. $|20\rangle\langle 10|$ or $|01\rangle\langle 02|$). For an anharmonic coupled system this feature has a lower vibrational frequency, as previously discussed in Section 1.2.1. If the system investigated were to be completely harmonic, the $v = 0 - 1$ and $v = 1 - 2$ transitions would overlap and cancel out, resulting in a featureless 2D-IR spectrum. As mentioned, the sign of this peak is the opposite to that of the on-diagonal $v = 0 - 1$ red peak. This is owing to the fact that the Feynman diagram carries a $(-1)^n$ term, where n is the number of interactions acting upon the *bra* (right hand side of the diagram).[4] Using the phase-matching geometry described for 2D-IR, this results in a positive response for the $v = 0 - 1$ transition as this transition results from an even number of interactions with the *bra* while features from the $v = 1 - 2$ transitions have an odd number of interactions with the *bra* and thus yields a negative response. Thus the spectrum diagonal hosts a pair of peaks, one negative and one positive, separated by the anharmonicity of the transitions.[26]

The off-diagonal features originate from coupling between vibrational modes, this occurs when two oscillating chemical bonds are close enough either mechanically or through space, so much so that the vibrational potentials of one molecule influences another.[4,19] If the two

modes are coupled they can be thought of as sharing a common ground state, $|00\rangle$, as shown in the energy level diagram (Figure 2.10(c)). As a result, excitation of the $|00\rangle \rightarrow |01\rangle$ depletes the now shared ground state and so a negative off-diagonal feature is observed at the $|00\rangle \rightarrow |10\rangle$ transition frequency labelled as peak (1). Peak (2) arises for the reversed case where excitation of the $|00\rangle \rightarrow |10\rangle$ produces a negative off-diagonal feature observed at the $|00\rangle \rightarrow |01\rangle$ transition frequency.[4,27] Again, accompanying the negative peak is a positive feature which is due to the transition from $|01\rangle$ or $|10\rangle$ to the an-harmonically shifted combination band, $|11\rangle$, peaks (5) and (6). The anharmonicity here is not equivalent to that of the $v = 0 - 1$ transition and instead the off-diagonal features produce the off-diagonal anharmonicity.

By recording a series of spectra at a range of different waiting times between the pump and probe pulses, dynamic information from the system can be obtained. On-diagonal features will decrease in amplitude with increasing waiting time as the probability of the molecules decaying back to the ground state increases.[4]

When modes are coupled, the same decay is seen in the off-diagonal features however other processes can occur. One of these is energy transfer and occurs when energy from one vibrational mode is transferred to another resulting in new off-diagonal peaks appearing. These new peaks corresponds to transitions between the excited mode and modes that were not accessible from the IR pulses. After excitation of $|10\rangle$, population transfer results in the $|10\rangle \rightarrow |01\rangle$ transition made possible. Subsequently, intramolecular vibrational population transfer between $|01\rangle \rightarrow |02\rangle$ which leads to a peak, that is situated between peaks (6) and (1), to grow into the spectrum. Studies to investigate the energy transfer between modes can employ laser pulses centred in two different spectral regions, where the pump is used to excite vibrational motion in one part of the molecule and the probe to detect the energy transfer within a different part of the molecule.[28]

At short waiting times, on-diagonal modes will appear elongated along the diagonal (Figure 2.11(a)) and have a narrow anti-diagonal linewidth. If the anti-diagonal linewidth is much narrower than the diagonal linewidth, as shown, then the transition is said to be inhomogenously broadened. If homogenous broadening were the only contribution however, the two linewidths would be equivalent. At short waiting times the oscillator experiences a 'static' environment as the measurement time is not long enough for the

system to have evolved appreciably between the pump and probe pulses and as such the solvent molecules are not yet fluctuating and changing their interacting with the oscillator.

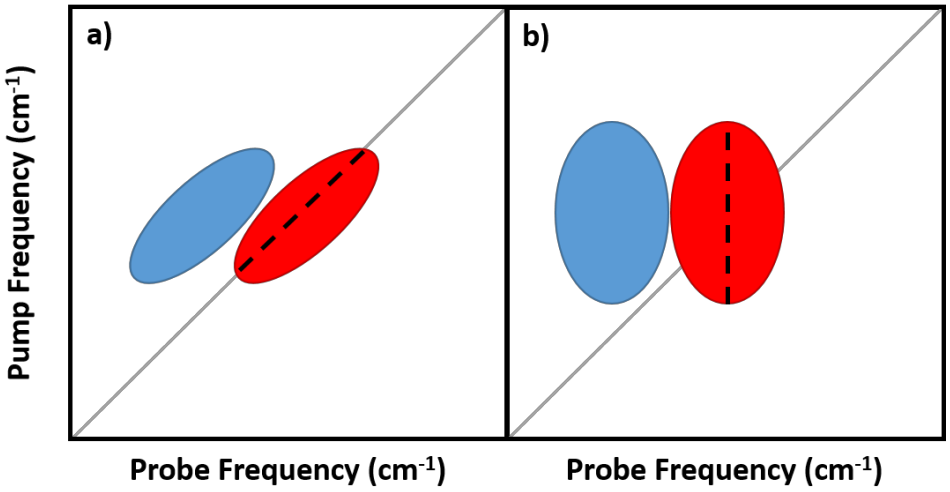

Figure 2.11: *Example 2D-IR spectra of a single vibrational mode at a) $T_w = \sim 0$ ps and b) $T_w > 0$ ps. Increasing T_w results in broadening of the vibrational modes ($v = 0 - 1$ shown in red and $v = 1 - 2$ shown in blue). Black dashed line denotes the CLS used to calculate spectral diffusion, see text for explanation.*

By extending the waiting time, the instantaneous frequencies of each oscillator becomes increasingly likely to fluctuate because of changes in its environment. As a result of this, the anti-diagonal linewidth becomes larger in the later waiting time spectrum (Figure 2.11(b)) and the spectral features becomes broadened and circular. At the longer waiting times the oscillator will experience a range of frequencies owing to the constant re-orientation of surrounding solvent molecules. This gives rise to a greater probability of the oscillator having a different detected vibrational frequency from that of which it was initially excited. This is known as spectral diffusion.[4,21,29]

The rate at which spectral diffusion occurs can be quantified by measuring the change in the lineshapes from multiple waiting time spectra. Chapter 3 sees the comparison of spectral diffusion of serum albumin in both H_2O and D_2O solvents and the central line slope (CLS) method was used to attempt quantification of the spectral diffusion.[30,31] The CLS method extracts the probe frequency where a maximum negative amplitude is found for a given range of pump frequencies and a straight line can be drawn through the feature (Figure 2.11, black dashed line). At early waiting times where little to no spectral diffusion has occurred,

the CLS will lie parallel to the 2D-IR spectrum diagonal and have a gradient of 1, relative to the pump and probe frequencies. As the waiting time is increased and the molecule undergoes spectral diffusion, the CLS will become gradually vertical and result in a gradient approaching infinity. By calculating the gradient of the CLS for each waiting time, the correlation between the initial and final vibrations can be mapped and the rate of spectral diffusion can be calculated.

2.3. 2D-IR Instrumentation

The data acquired throughout this thesis was generated on two different instruments housed at the Rutherford Appleton Laboratory (RAL) in Oxfordshire, England. Both instruments result in the production of a pump pulse pair and a probe pulse and are discussed in detail. The main differences between the two instruments are the repetition rates of the laser pulses which influences the length of time required to obtain a 2D-IR spectrum, and the bandwidths of the IR laser pulses used which limits the range of frequencies accessed. Comparing the results between the two instruments is vital as 2D-IR spectrometers are custom built and resulting signals may vary significantly. It is important to know if results from one instrument are repeatable on a different instrument with different properties. Chapter 5 focusses on comparing results between the two instruments. The data collected in Chapters 3 and 4 were carried out using the ULTRA set up. Experiments in Chapter 6 were recorded on LIFEtime.

2.3.1. ULTRA Instrument

Initially a titanium sapphire (Ti:Sapph) oscillator produces pulses with a repetition rate of 65 MHz, centred at 800 nm, each with a duration of ~20 fs.[32,33]

The pulse train is then sent through a regenerative amplifier to increase the average pulse energy (Figure 2.12). Amplification of short pulses leads to high peak intensities, and so the pulses are stretched prior to amplification which helps prevent damage to the Ti:Sapph gain medium and optics. The Ti:Sapph is pumped by two Nd:YAG lasers, each producing 55 W of power. To help prevent heat damage to the crystal, it is cryogenically cooled to -10°C.[34] Once the pulses have gained sufficient energy they are compressed before leaving the amplifier. This stage produces pulses with an average energy of 0.8 mJ at a repetition rate of 10 kHz, centred at 800 nm, each with a duration of ~50 fs.

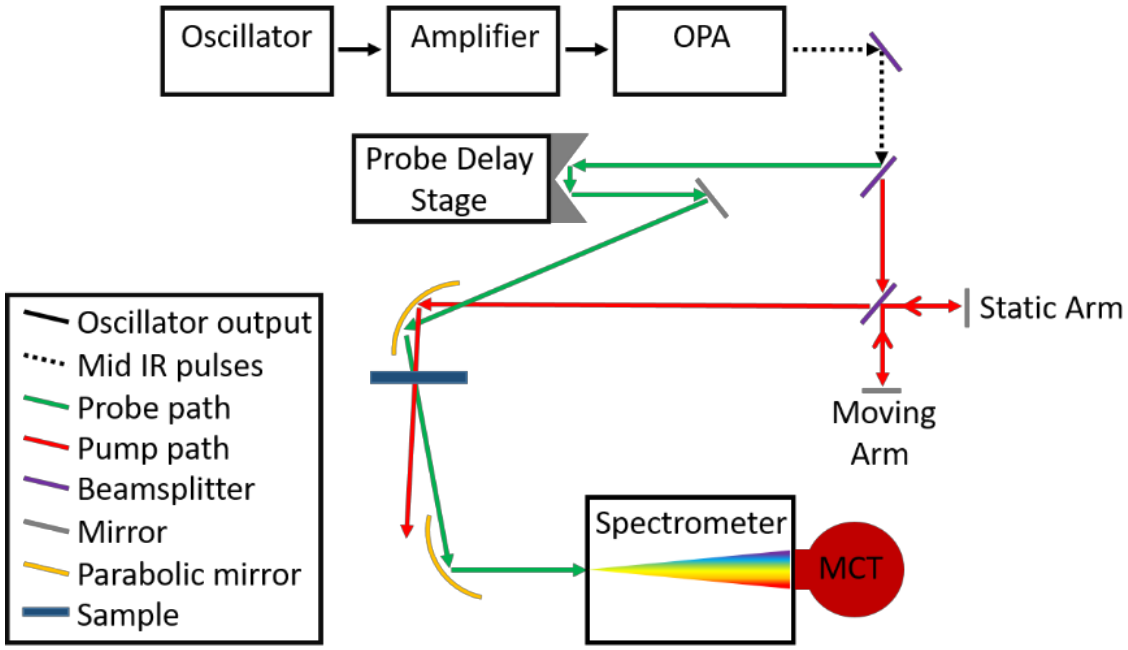

Figure 2.12: *Schematic diagram of the ULTRA system employing the Michelson interferometer for pulse pair generation.*

The next stage is to convert the wavelength of the pulses to produce light resonant with the molecular vibrations being studied. This is achieved through the use of an optical parametric amplifier (OPA). [35,36] The OPA splits the incoming pulses using a 50:50 beamsplitter, producing two identical beams known as the pump beam and the seed beam. The seed is then passed through an optic limiting its intensity to 10% and then focussed onto a sapphire crystal to generate a white light continuum. At the point of focus on the sapphire plate, the now high intensity light is sufficient enough to modify the refractive index of the crystal which causes self-focussing, this creates multiphoton ionisation which produces free electrons in the crystal, decreasing its refractive index and thus opposing the self-focussing effect.

As the refractive index is intensity-dependent, throughout the duration of the Gaussian shaped pulse the seed beam will encounter a change in refractive index. This produces an intensity-dependent phase shift as the light passes through the crystal which in turn changes the frequency of the light. As the temporal profile of the pulses pass through the crystal the light is broadened to both higher and lower frequencies. The leading half of the Gaussian pulse increases in intensity causing an increase in the refractive index, which in turn induces

broadening to lower frequencies. The latter half of the Gaussian decreases in intensity causing a decrease in the refractive index which in turn induces broadening of the light to higher frequencies. Balancing these two effects allows the production of a high intensity white light spanning ~1000 nm (400-1400 nm).

The white light pulses are then temporally stretched before focussed onto a barium borate (BBO) crystal where it meets the 800 nm pump beam and optical parametric amplification occurs, converting 800 nm light to lower frequency 'signal' and 'idler' beams:

$$\omega_p = \omega_s + \omega_i \tag{2.17}$$

where the idler beam, ω_i, is a by-product of the conversion of energy from the pump, ω_p, to the signal, ω_s, conserving photon momentum.[20] Selection of a specific signal wavelength can be reached by altering the time delay between the pump beam and the temporally dispersed white light pulse. In order to achieve efficient energy conversion, the group velocities of both signal and idler beams need to be well-matched and so the angle of the BBO crystal may need to be adjusted to achieve this.[35] Following this, the pump and idler beams are dumped onto a non-reflective surface and the signal beam is collimated before entering a second optical parametric amplification stage to increase the signal power. The only difference between the second and first OPA stages is that in the second stage the pump and signal beams propagate collinearly, producing a collinear idler beam.

Mid-IR light is produced by difference frequency generation (DFG) of the signal and idler beams resulting from this second optical parametric stage.[36,37] This process involves separating the signal and idler beams in order to control phase matching during their recombination. A silver thiogallate ($AgGaS_2$) crystal produces light at a frequency which is equal to the difference between the signal and idler beams; a range of 2.4 – 12 µm is achievable using this method.[36]

In this thesis, all experiments employ pulses centred at 6 µm (1650 cm^{-1}). The ULTRA system results in an output with a temporal duration of ~50 fs and a bandwidth of ~400 cm^{-1}.

Once the pulses are at the desired wavelength they progress through a 90:10 beamsplitter to produce a high intensity pump pulse and a low intensity probe pulse, as seen in Figure 2.12, red and green paths respectively.

The pump beam is then passed through a Michelson interferometer (as shown in Figure 2.12) to produce the pump-pulse pair, $\vec{k}_1$ and $\vec{k}_2$. The moving arm of the interferometer is controlled by a variable delay stage which moves during the data acquisition and defines the time period, t_1, between the two re-combined collinear pump pulses. Displacement of the interferometers moving arm by a distance, Δx, varies the inter-pulse timing as follows:

$$t_1 = \frac{2\Delta x}{c} \tag{2.14}$$

where c is the speed of light. As the light travels from the beamsplitter to the mirror and back again, the total path difference is $2\Delta x$ for every increment of Δx the delay stage moves. The path difference between the two pump pulses gives rise to interference when $t_1 \neq 0$ and as the stage is scanned the frequency components of the recombined light from the two arms of the interferometer shift in and out of phase depending on the path difference caused by the moving arm. Scanning this time delay during data acquisition allows the excitation frequency to be controlled in a way parallel to the scanning of the narrowband pump pulse in the frequency domain.[4,38]

The step sizes taken by the variable delay stage, as well as the maximum recorded value of t_1, help define the pump frequency axis of the 2D-IR spectrum after Fourier transform. The spectral resolution, $\Delta\omega$, of the pump frequency axis is related to the maximum value recorded for t_1, thus for a high resolution 2D-IR spectra many points are typically collected for t_1.[4]

$$\Delta\omega = \frac{1}{t_{1\,max}} \tag{2.15}$$

The step size, Δt, dictates the maximum frequency that can be detected from the Fourier transform of the time domain data, known as the Nyquist frequency, ω_N.

$$\frac{\omega_N}{2\pi} = \frac{1}{2\Delta t} \tag{2.16}$$

A smaller step size allows detection of higher frequency components and larger values of $t_{1\,max}$ provides a higher pump frequency resolution after the Fourier transform. For the data acquired in this thesis using the interferometer set up, step sizes of the interferometer arm of 3 fs (~0.45μm) with a maximum recorded t_1 value of 4ps were used.

The probe beam, $\vec{k}_3$, encounters the probe delay stage prior to reaching the sample. This delay stage controls the waiting time, T_w, between the pump and probe pulses. This stage is set prior to the experiment and does not change during data acquisition. Following the interferometer, both the pump and probe pulses are focussed and directed towards the sample by a gold parabolic mirror. After interaction with the sample the pump pulse is dumped onto a non-reflective surface. The emitted signal, $\vec{k}_4$, and the probe pulse, $\vec{k}_3$, are re-collimated by a second gold parabolic mirror before being spectrally dispersed via a diffraction grating and then onto liquid nitrogen cooled mercury cadmium telluride (MCT) pixel detector array which yields the probe frequency axis of the 2D-IR spectra. With this set up a spectral resolution of ~1.5 cm^{-1} for the probe frequency axis was achieved.

As the signals are recorded by referencing against no interaction of the pump pulse with the sample, otherwise known as 'pump-off', the pump pulses are modulated by a beam chopper which is synchronised to the regenerative amplifier output (10 kHz) at half of its repetition rate, 5 kHz. This blocks alternative pump pulses allowing the probe spectrum to be recorded both with and without the pump.

Each data point collected is averaged for 0.4 seconds to help increase the signal to noise ratio. The final signal is also averaged over three scans of t_1, again to help increase signal to noise and to help compensate for any long term drift effects in the alignment of the laser.

2.3.2. LIFEtime Instrument

The LIFEtime laser instrument is comparable to that described for the ULTRA set up, however there are a number of differences which have a significant impact on the performance of LIFEtime set-up when measuring 2D-IR spectra.[39]

Initially, one oscillator feeds two ytterbium potassium gadolinium tungstate (Yb:KGW) regenerative amplifiers which are then used to seed two OPAs: one for the pump beam and another for the probe beam (Figure 2.13). A third OPA can also be used to produce a second probe beam at different frequencies (two-colour experiment)[40,41] however for the work in this thesis a second probe pulse was not used. The pump amplifier produces pulses centred at 1032 nm with a pulse duration of 260 fs at a repetition rate of 100 kHz.[40] The probe amplifier also generates pulses centred at 1032 nm at a reputation rate of 100 kHz but have a shorter duration at 180 fs. The total power output of the pump and probe amplifiers is 15 and 6 W, respectively. The pump amplifier drives OPA1 (Light Conversion Orpheus-HP) which

is a similar OPA and DFG process as described for ULTRA and produces pump pulses tuneable within the range of 2100 - 13000 nm. The probe amplifier drives OPA2 (Light Conversion Orpheus-One) and the additional OPA3 if required. This provides one (or two) separate probe pulse output that is tuneable from 4000 - 13000 nm. The two separate probe OPAs (OPA2 & OPA3) allow probing of two different spectral regions simultaneously.

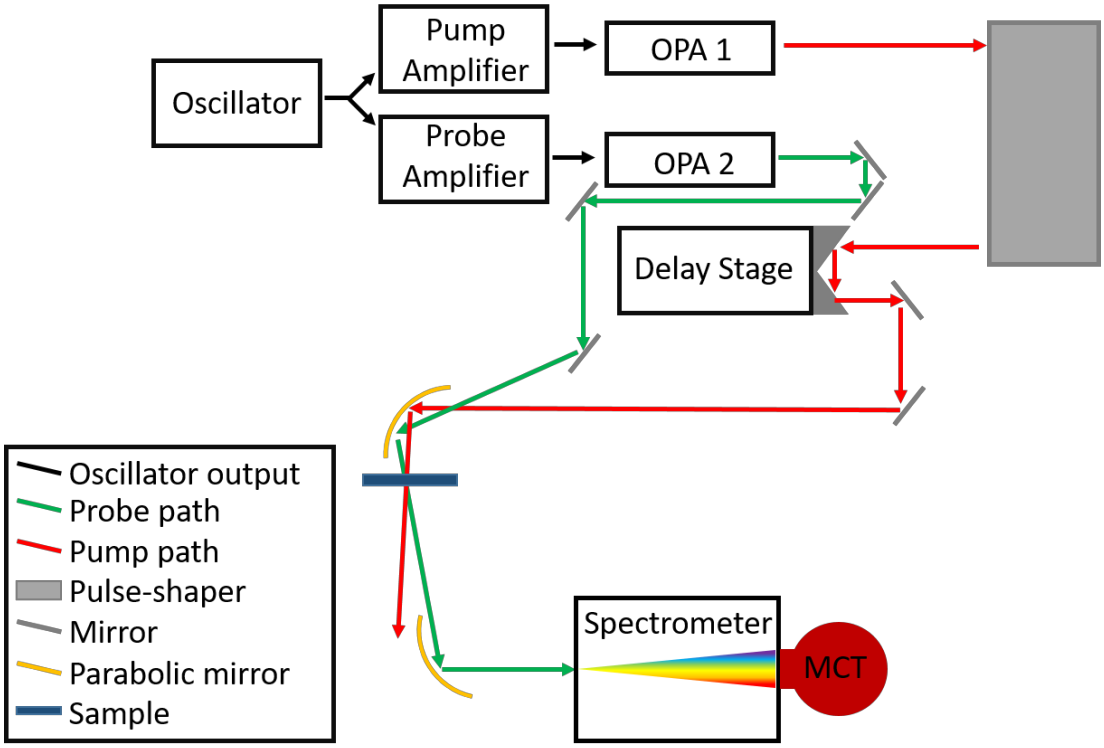

Figure 2.13: *Schematic diagram of the LIFEtime system employing a pulse shaper for pulse pair generation (grey box). As OPA 3 was not used for the experiments it has not been included here.*

One of the major differences between the ULTRA and the LIFEtime set up is the use of a MIR pulse shaper instead of a Michelson interferometer to produce the pump pulse pair (Figure 2.14).[42,43] The use of a pulse-shaper results in a programmable pump beam frequency meaning that the phase relation of the two pumps is always known.[42]

The output of OPA1, a single pump pulse, is aligned to enter the pulse shaper which results in the production of a pump pulse pair (Figure 2.14). As the pump beam enters the pulse shaper it is dispersed by a grating before being collimated by a cylindrical mirror. This collimated light is then allowed to pass through a germanium acousto-optic modulator (AOM). This modulator is coupled to a waveform generator which is programmed to pass an

acoustic wave along the crystal in order to re-shape the incoming pump pulse to the desired output.[44]

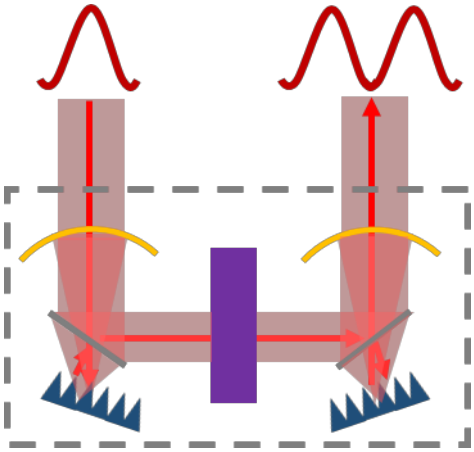

Figure 2.14: *Pulse-shaper schematic. One pulse enters the pulse shaper and is dispersed by a grating, the parabolic mirror creates a cylindrical beam which passes through the Ge crystal. The beam is then incident on another parabolic mirror before being dispersed again and leaving the pulse shaper as two pulses separated by* t_1.

The use of the pulse shaper allows exact modulation of the pump pulse to produce the pump pulse pair with known phase and intensity. With each programmed acoustic wave producing a pulse pair, the AOM is varied to alter the timing between the pulse pair. Resulting in a train of pump pulse pairs with different time delays, t_1, between the two pump pulses. Each resultant pump pulse pair is identical to those created using the interferometer except now time zero and the phase of the two pump pulses are always known and no longer need to be calculated, allowing more accurately phased 2D-IR spectra to be obtained.[4,43,45]

The new pump pulse pair, $\vec{k}_1$and $\vec{k}_2$, is then sent through a computer-controlled delay stage to set the time separation between the pump and probe, $\vec{k}_3$, pulses, the waiting time. Both the pump and probe pulses are focussed and directed towards the sample via a gold parabolic mirror (Figure 2.13). Again, after interaction with the sample the pump pulse is dumped onto a non-reflective surface. The emitted signal, $\vec{k}_4$, and the probe pulse, $\vec{k}_3$, are re-collimated by a second gold parabolic mirror before being spectrally dispersed via a diffraction grating and then onto liquid nitrogen cooled MCT pixel detector array which yields the probe frequency axis of the 2D-IR spectra. With this set up a spectral resolution of

~1.5 cm^{-1} for the probe frequency axis was achieved. If OPA3 is in use a second MCT pixel detector array can be set up to detect a different wavelength region.

The signal frequencies detected at each pixel on LIFEtime are collected as a function of t_1 and Fourier transformed with respect to t_1 to obtain the pump frequency axis of the 2D-IR spectra.

The LIFEtime system utilises pulses with a central frequency of 1650 cm^{-1} with a pulse duration of ~200 fs with a bandwidth of ~100 cm^{-1}.[40]

Due to the fast repetition rate of the laser, a high rate of production of data points is acquired which may result in memory capacity issues within the computers used and the pulseshaper cannot carry enough waveforms to scan a large pump frequency range in small steps. One way to help minimise the number of data points collected without compromising the spectral resolution and range is to under-sample the interferogram which leads to aliasing.[4] Under-sampling allows much larger step sizes to be taken between the pump pulses whilst covering the same frequency range. As fewer points are recorded to determine the frequency of the signal, the frequency components higher than the Nyquist frequency are aliased.[4] An example of aliasing is shown in Figure 2.15 below.

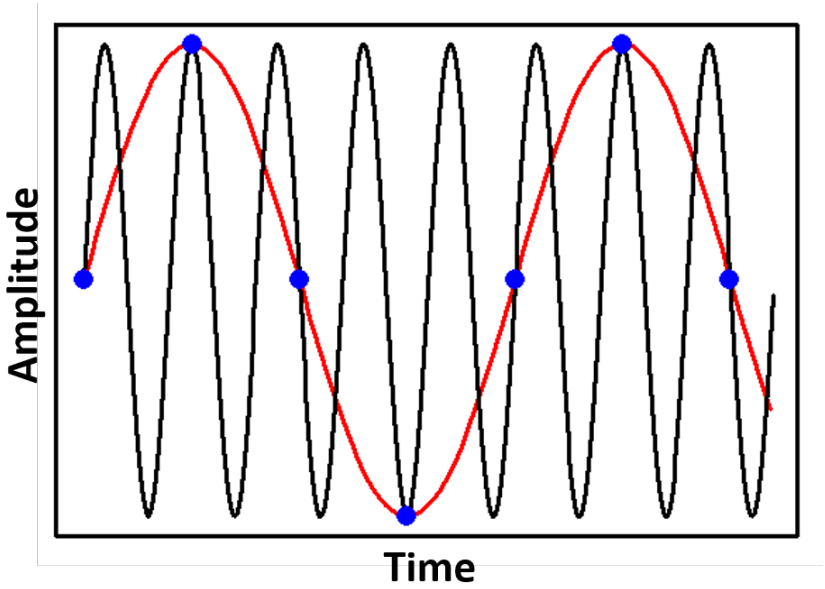

Figure 2.15: *An example of aliasing. Black trace shows 'correct' sampling of the signal and the red trace shows the under-sampled signal. Blue points denotes where the amplitude of the two traces are identical.*

The correct frequencies for features found can be retrieved given that the measured frequencies are dictated by the frequency laser spectrum. Data presented in this thesis using the LIFEtime and pulseshaper upgraded ULTRA instruments utilised 12 fs step sizes compared to the 3 fs step sizes used for the ULTRA Interferometer.

Furthermore, the pulseshaper set-up makes use of phase cycling, which is a technique employed to aid in reducing the volume of background noise and scattering effects in 2D-IR spectra by measuring additional data points at each value of t_1 with different phases ($\pm \pi$) of the pump pulse pair. This additional data is subtracted from the original in phase data allowing reduction of background noise and scatter.

2.4. 2D-IR Data Processing

The beam geometry was chosen in order to emit both the rephasing and non-rephasing pathways in the same direction and so with an accurate derivation of time zero the rephasing and non-rephasing pathways are summed correctly producing absorptive 2D-IR spectra.

Where $\vec{k}_1$ and $\vec{k}_2$ overlap exactly is important to establish the pump frequency axis and phasing of the 2D-IR spectra, this position is known as 'time-zero'. As the pulse shaper provides a pump pulse pair where the phase is always known between the two pulses, time-zero is automatically known and does not need to be calculated. Once time-zero has been established, measurements prior to time-zero are not included in the 2D-IR spectra as they can cause distortion in the observed 2D-IR line shapes and so the measured signals in t_1 are truncated to remove the negative delay times. This can only be achieved once time-zero is known and hence why accurate measurement of time-zero is necessary. [4]

Zero-padding is also implemented in order to increase the apparent spectral resolution of the obtained pump axis after Fourier transform. Prior to the Fourier transform, the number of data points in time is increased by adding data points with an intensity equal to zero. This process involves extending the maximum value of $t_{1\,max}$ in the Fourier transform which in turn produces a higher frequency resolution of the obtained pump frequency (equation 2.5).

An example of the signal recorded by each pixel in the MCT array is shown in Figure 2.16(a). Fourier transform of the interference pattern of the heterodyned signal with respect to t_1 produces a spectrum of all of the frequency contributions to the signal for each pump

frequency (Figure 2.16(b)). Fourier transform of this on each probe pixel allows compiling of the 2D-IR spectrum.

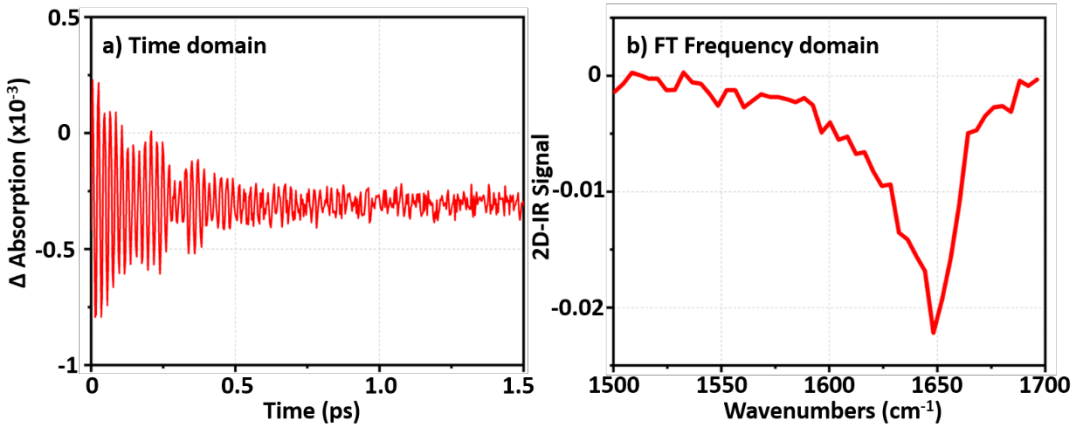

Figure 2.16: *a) Time domain signal from one of the pixels in the MCT array. The time (ps) is truncated at 1.5 ps however t_1 max continues until 4 ps. b) The frequency domain signal of a) after Fourier transform.*

An example 2D-IR spectrum using the pulse-shaper LIFEtime set-up, with automatic time zero calculation and with zero padding applied is shown in Figure 2.17. A resolution of ~4.2 cm^{-1} is obtained for the pump frequency axis.

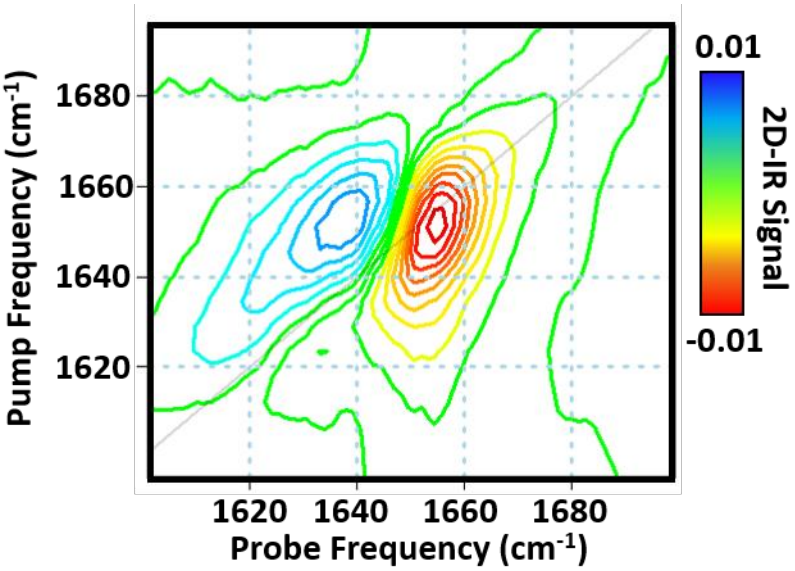

Figure 2.17: *An example 2D-IR spectrum of bovine serum albumin in H_2O using the LIFEtime instrument.*

However for the case of the ULTRA interferometer setup the phase is not known and must be calculated in order to produce correctly phased data. This process will now be discussed.

After propagation through the interferometer, a weak reflection of the pump beam is isolated and sent to a single element detector where the intensity of the recombined pump pulses is measured as a function of t_1, this is known as a pump auto-correlation and is crucial for proper phasing of 2D-IR spectra. An example of an auto-correlation is shown in Figure 2.18.

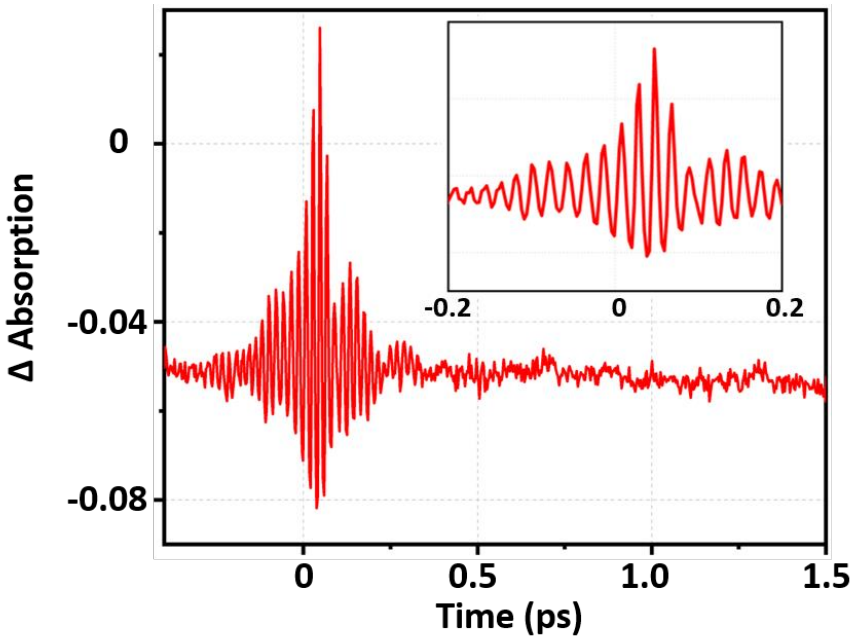

Figure 2.18: *Pump auto-correlation, acquired during 2D-IR measurements on ULTRA. Insert shows 'centre-burst' of the auto-correlation between -0.2 and 0.2 ps.* $\Delta t = 3\ fs$.

The maximum pump-pump intensity is recorded at t_1=0, meaning that the pump pulses are completely in phase and interfere constructively. In order to record the full interference pattern, and to ensure time-zero is collected, negative time delays are also recorded.

A small window of the auto-correlation between -0.2 and 0.2 ps is known as the 'centre-burst' (Figure 2.18) and is used to accurately determine time-zero and phasing of the 2D-IR spectra. The centre-burst is the temporal region in which the measured intensity is defined by the relative phase of the two pump pulses.

Accurate derivation of time-zero is achieved through analysis of the phase slope ($\partial\phi/\partial t_1$) of the analysed pump auto-correlation.[4] At time-zero, where all the frequency components are in phase, the phase slope, $\partial\phi/\partial t_1$, gradient is zero. This point is then adjusted by interpolating the auto-correlation signal to get a more accurate result, allowing fine-tuning of the spectral phase in each of the probe signals. Finally, the projection slice theorem is used to establish the final value for time-zero.[4,38] A projection of all 2D-IR pump slices onto the probe axis is overlaid with a pump-probe spectrum, when the spectrum is phased correctly the two should overlap perfectly (Figure 2.19(a)).

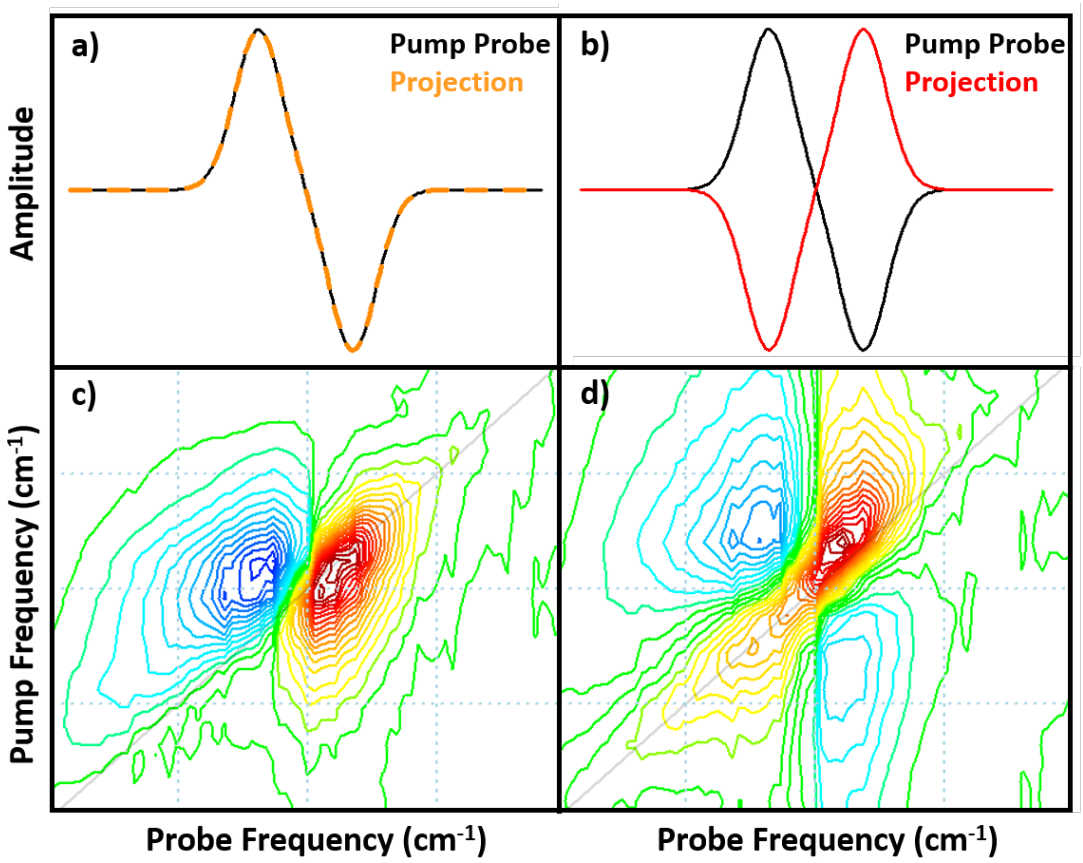

Figure 2.19: *Phasing of the 2D-IR spectra using the Projection Slice Theorem. a) the overlap of the projected slice with the pump-probe spectrum. The correct time-zero has been selected and the two spectra overlap closely, as a result the 2D-IR spectra (c) is properly phased. b) the overlap of the projected slice with the pump-probe spectrum, an incorrect time-zero has been selected and the two spectrums do not overlap, as a result the 2D-IR spectra (d) is not properly phased. 2D-IR spectra that have not been phased correctly have dispersive components that distort the 2D-IR lineshapes, splitting the spectral features into quadrants.*

Accurate derivation of time zero allows both the rephasing and non-rephasing pathways to be summed correctly which results in absorptive 2D-IR spectra. Incorrectly phased spectra contain contributions from the dispersive component derived from the negative time delays, causing a split of the 2D-IR into quadrants (Figure 2.19(d)).

Developments in ULTRA were performed throughout the measurement access time. This included the replacement of the Michelson Interferometer with a pulse shaper. As the pulseshaper data is automatically correctly phased, the techniques described for the ULTRA interferometer set up are not required.[4,46]

2.5. References.

1. Kumar, V., Coluccelli, N. & Polli, D. Coherent optical spectroscopy microscopy and applications. in *Molecular and Laser Spectroscopy* (Elsevier, 2018).
2. Stuart, B. *Infrared Spectroscopy: Fundamentals and Applications*. (J.Wiley and Sons, 2009).
3. Graham-Smith, F., King, T. A. & Wilkins, D. *Optics and Photonics: An Introduction*. (J.Wiley and Sons, 2007).
4. Hamm, P. & Zanni, M. T. *Concepts and Methods of 2D Infrared Spectroscopy*. (Cambridge University Press, 2011).
5. Jackson, J. D. & Wiley, J. *Classical Electrodynamics*. (J.Wiley and Sons, 1999).
6. Atkins, P. & Paula, J. de. *Atkins' Physical Chemistry*. (Oxford University Press, 2006).
7. Sathyanarayana, D. . *Vibrational Spectroscopy: Theory and Applications*. (New Age International, 2007).
8. Skinner, J. L. & Hsu, D. Pure dephasing of a two-level system. *J. Phys. Chem.* **90,** 4931–4938 (1986).
9. Park, S., Kwak, K. & Fayer, M. . Ultrfast 2D-IR vibrational echo spectroscopy: A probe of molecular dynamics. *Laser Phys. Lett.* **4,** 704–718 (2007).
10. Zewail, A. H. Optical molecular dephasing: principles of and probings by coherent laser spectroscopy. *Acc. Chem. Res.* **13,** 360–368 (1980).
11. Khalil, M., Demirdöven, N. & Tokmakoff, A. Vibrational coherence transfer characterized with Fourier-transform 2D IR spectroscopy. *J. Chem. Phys.* **121,** 362–373 (2004).
12. Rector, K. . & Fayer, M. . Vibrational dephasing mechanisms in liquids and glasses: Vibrational

echo experiments. *J. Chem. Phys.* **108,** 1794–1803 (1998).

13. Tokmakoff, A., Urdahl, R.S., Zimdars, D., Francis, R.S. *et al.* Vibrational spectral diffusion and population dynamics in a glass-forming liquid: Variable bandwidth picosecond infrared spectroscopy. *J. Chem. Phys.* **102,** 3919–3931 (1998).

14. Hunt, N. T. Transient 2D-IR spectroscopy of inorganic excited states. *Dalt. Trans.* **43,** 17578–17589 (2014).

15. Thompson, J. M. *Infrared Spectroscopy*. (Pan Stanford, 2018).

16. Smith, B. C. *Fundamentals of Fourier Transform Infrared Spectroscopy*. (CRC Press, 2011).

17. Telle, H. H., Urena, A. G. & Donovan, R. J. *Laser Chemistry*. (John Wiley & Sons, 2007).

18. Bloembergen, N. *Nonlinear Optics*. (World Scientific, 2002).

19. Hamm, P. Principles of nonlinear optical spectroscopy: A practical approach or mukamel for dummies. *University of Zurich* (2005). Available at: http://www.mitr.p.lodz.pl/evu/lectures/Hamm.pdf. (Accessed: 27th February 2020)

20. Boyd, R. W. *Nonlinear Optics*. (Academic Press, 2008).

21. Jansen, T. la C. Multi-dimensional spectroscopy. *University of Groningen* (2009). Available at: acmm.nl/molsim/han/2009/LaCourJansen-multidim-spectrosc-Han11jan2010.pdf. (Accessed: 27th February 2020)

22. Fleming, G. . *Chemical Applications of Ultrfast Spectroscopy*. (Oxford University Press, 1986).

23. Mukamel, S. Principles of Nonlinear Optical Spectroscopy. (1999).

24. Hamm, P., Lim, M. & Hochstrasser, R. M. Structure of the amide I band of peptides measured by femtosecond nonlinear-infrared spectroscopy. *J. Phys. Chem. B* **102,** 6123–6138 (1998).

25. Cervetto, V., Helbing, J., Bredenbeck, J. & Hamm, P. Double-resonance versus pulsed Fourier transform two-dimensional infrared spectroscopy : An experimental and theoretical comparison. *J. Chem. Phys.* **121,** 5935–5942 (2007).

26. Kim, Y. S. & Hochstrasser, R. M. Applications of 2D IR spectroscopy to peptides, proteins, and hydrogen-bond dynamics. *J. Phys. Chem. B* **113,** 8231–8251 (2009).

27. Hunt, N. T. 2D-IR spectroscopy: ultrafast insights into biomolecule structure and function. *Chem. Soc. Rev.* **38,** 1837–1848 (2009).

28. Hithell, G., Gonzalez-Jimenez, M., Greetham, G.M., Donaldson, P.M. *et al.* Ultrafast 2D-IR and optical Kerr effect spectroscopy reveal the impact of duplex melting on the structural dynamics

of DNA. *Phys. Chem. Chem. Phys.* **19,** 10333–10342 (2017).

29. Roberts, S. T., Loparo, J. J. & Tokmakoff, A. Characterization of spectral diffusion from two-dimensional line shapes. *J. Chem. Phys.* **125,** 84502 (2006).

30. Kwak, K., Park, S., Finkelstein, I. J. & Fayer, M. D. Frequency-frequency correlation functions and apodization in two-dimensional infrared vibrational echo spectroscopy: A new approach. *J. Chem. Phys.* **127,** 124503 (2007).

31. Guo, Q., Pagano, P., Li, Y.-L., Kohen, A. & Cheatum, C. M. Line shape analysis of two-dimensional infrared spectra. *J. Chem. Phys.* **142,** 212427 (2015).

32. Greetham, G.M. Burgos, P., Cao, Q., Clark, I.P. *et al.* ULTRA: A unique instrument for time-resolved spectroscopy. *Appl. Spectrosc.* **64,** 1311–1319 (2010).

33. Greetham, G.M., Matousek, P., Robinson, D.A., Parker, A.W. *et al.* ULTRA laser system: A new dual-output 10 kHz Ti:Sapphire amplifier with UV–IR generation for time-resolved spectroscopy. *Cent. Laser Facil. Annu. Rep.* 249–250 (2008).

34. Backus, S., Iii, C. G. D., Mourou, G., Kapteyn, H. C. *et al*. 0.2-TW laser system at 1 kHz. *Opt. Lett.* **22,** 1256–1258 (1997).

35. Baumgartner, R. A. & Byer, R. Optical parametric amplification. *IEEE J. Quantum Electron.* **15,** 432–444 (1979).

36. Cerullo, G. & DeSilvestri, S. Ultrafast optical parametric amplifiers. *Rev. Sci. Instrum.* **74,** (2003).

37. Brida, D. Manzoni, C., Cirmi, G., Marangoni, M. *et al.* Generation of broadband mid-infrared pulses from an optical parametric amplifier. *Opt. Express* **15,** 15035–15040 (2007).

38. Deflores, L.P., Nicodemus, R.A. & Tokmakoff, A. Two-dimensional Fourier transform spectroscopy in the pump – probe geometry. *Opt. Lett.* **32,** 2966–2968 (2007).

39. Greetham, G.M., Donaldson, P.M., Nation, C., Sazanovich, I.V. *et al.* A 100 kHz Time-resolved multiple-probe femtosecond to second infrared absorption spectrometer. *Appl. Spectrosc.* **70,** 645–653 (2016).

40. Donaldson, P.M., Greetham, G.M., Shaw, D.J., Parker, A.W. *et al*. A 100 kHz pulse shaping 2D-IR spectrometer based on dual Yb:KGW amplifiers. *J. Phys. Chem. A* **122,** 780–787 (2018).

41. Hithell, G., Shaw, D.J., Donaldson, P.M., Greetham, G.M. *et al.* Long-range vibrational dynamics are directed by Watson-Crick base pairing in duplex DNA. *J. Phys. Chem. B* **120,** 4009–4018 (2016).

42. Shim, S., Strasfeld, D.B., Fulmer, E.C. & Zanni, M.T. Femtosecond pulse shaping directly in the mid-IR using acousto-optic modulation. *Opt. Lett.* **31,** 838–840 (2006).

43. 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.* **104,** 14197–14202 (2007).

44. Hillegas, C.W., Tull, J.X., Goswami, D., Strickland, D. *et al*. Femtosecond laser pulse shaping by use of microsecond radio-frequency pulses. *Opt. Lett.* **19,** 737–739 (1994).

45. Shim, S.H. & Zanni, M.T. How to turn your pump–probe instrument into a multidimensional spectrometer: 2D IR and Vis spectroscopiesvia pulse shaping. *Phys. Chem. Chem. Phys.* **11,** 748–761 (2009).

46. Rock, W., Li, Y.L., Pagano, P. & Cheatum, C. M. 2D IR Spectroscopy using four-wave mixing, pulse shaping, and IR upconversion: A quantitative comparison. *J. Phys. Chem. A* **117,** 6073–6083 (2013).

3. Measuring Proteins in H_2O Using 2D-IR Spectroscopy

This chapter contains results published in the following publications:

Hume S, Hithell G, Greetham G.M, Donaldson P.M, Towrie M, Parker A.W, Baker M.J, Hunt N.T., *Measuring proteins in H_2O with 2D-IR spectroscopy*. Chemical Science. 10(26):6448-6456 (2019)

Fritzsch R, Hume S, Minnes L, Baker M.J, Burley G.A, Hunt N.T., *Two-dimensional infrared spectroscopy: an emerging analytical tool?* The Analyst. 145:2014-2024 (2020)

Butler H.J, Cameron J.M, Jenkins C.A, Hithell G, Hume S, Hunt N.T, Baker M.J, *Shining a light on clinical spectroscopy: translation of diagnostic IR, 2D-IR and Raman spectroscopy into the clinic.* Clinical Spectroscopy. 1:100003 (2019)

All 2D-IR serum data acquisition and analysis was performed by myself.

3.1. Abstract

The amide I infrared band of proteins is highly sensitive to secondary structure, but studies under physiological conditions are prevented by strong, overlapping absorption bands of water which motivates the widespread use of deuterated solvents. Exchange of solvent hydrogen for deuterium (H/D) raises fundamental questions regarding the impact of increased mass on protein dynamics and is impractical for biomedical or commercial applications of protein IR spectroscopy. In this chapter, it is shown that 2D-IR spectroscopy can avoid this problem as the non-linear optical nature of the 2D-IR signal suppresses the water response allowing, the amide I signature of proteins to dominate that of water even at sub-millimolar protein concentrations.

Infrared spectroscopy is sensitive to molecular structure, but its application to proteins is hindered by the broad nature of the amide I band, essentially the C=O stretching mode of the peptide link. 2D-IR spectroscopy overcomes this problem by measuring vibrational couplings between peptide units in macromolecular structures and the 2D-IR amide I signal is diagnostic of secondary structure content and ligand binding.

Using equine blood serum as a test system, the spectroscopy and dynamics of proteins in water is investigated. Measurements of vibrational relaxation dynamics of serum proteins reveals that deuteration slows down the rate of amide I vibrational relaxation indicating a dynamic impact of isotopic exchange in the proteins studied.

The unique link between protein secondary structure and 2D-IR amide I lineshape allows differentiation of signals due to albumin and globulin protein fractions in serum leading to measurements of the bio-medically important albumin to globulin ratio (AGR) over a clinically-relevant range.

Additionally, this study demonstrates that 2D-IR spectroscopy can enable differentiation of the structurally similar globulin proteins IgG, IgA and IgM, which opens up a straightforward spectroscopic approach to measuring levels of serum proteins that are currently only accessible via biomedical laboratory testing, the potential exists for more detailed characterisation of the protein content of serum.

3.2. Introduction

Successful treatment of disease is often critically dependent upon obtaining an early and accurate diagnosis and the production of new enabling technologies for healthcare is an ongoing scientific challenge.[1,2] Label-free testing strategies based on minimally-invasive sampling methods are desirable, being preferred by patients while delivering results quickly and economically is important. For these reasons spectroscopic interrogation of biofluids, such as blood serum, are particularly attractive targets for new analytical methods. Blood serum is easily obtained with minimal patient discomfort and provides data from the circulatory system which contains a range of potentially diagnostic chemical markers by virtue of contact with most of the major organs.[3–5] Current technologies use antibody assays to enhance the signal associated with a target biomolecule, relying critically on the availability of specific antibodies for proteins of interest and requiring significant sample preparation.[6]

Moreover, the heterogeneous nature of disease means that single-metabolite detection may be inferior to a broad bio-molecular fingerprint of metabolic function as an early warning of deteriorating patient health[3,4] or, for example, to indicate the presence of cancers.[7] The protein content of blood serum represents an ideal substrate for holistic analysis. Human serum contains ~70 mg/mL of proteins composed of albumin (~35–50 mg/mL) and the globulins (~25–35 mg/mL).[8–10] Diagnostically, measurement of the albumin to globulin ratio (AGR) is valuable.[8,11–13] The AGR of blood serum is a widely-used clinical marker and knowledge of this ratio is sufficient to provide important insights into general health. Changes in the AGR are associated with liver and kidney disease as well as with the presence of infection, inflammation, autoimmune disease and some forms of blood cancer.[11–14] The presence of an inflammatory response is indicated by increased globulin levels thus a low AGR. An indication of the ability of the patient to survive cancer therapy can also be indicated by higher AGR values but threshold limits vary depending on the cancer type. Despite the importance of the AGR, the globulin fraction is currently obtained indirectly using wet chemistry techniques and taking the difference between the concentration of albumin and the total protein content of serum rather than by a direct measurement.[15]

Beyond the AGR, a more detailed breakdown of the globulin composition enables deeper insights into specific health problems, however the globulins encompass a huge number of proteins. The γ-globulins constitute the bulk of the serum globulin fraction. Of these,

immunoglobulin-G (IgG) is the most abundant, accounting for ~80% of the γ-globulins, while IgA (~13%) and IgM (~6%) are the next most abundant.[16,17] As well as bulk changes in globulin concentration, changes in serum levels of each of these individual globulin components are associated with health-related issues.[17] IgG levels are found to increase in cases of liver disease or chronic infection,[18] IgA is linked to cirrhosis[19] and diabetes mellitus[20] while changes in IgM levels can warn of antibody deficiency syndrome.[21] There is therefore considerable benefit in a straightforward spectroscopic measurement that can not only deliver the AGR directly but also differentiate between the major globulin components.

Water is an obstacle common to both IR absorption and 2D-IR spectroscopy. The biological solvent absorbs widely across the mid-IR, but the H–O–H bending mode (δ_{H-O-H}, 1644 cm^{-1}) directly obscures the protein amide I transition, motivating the widespread use of deuterated solvents (D_2O) for protein IR spectroscopy (Figure 3.1). As deuterated water is heavier than H_2O, the bending mode is shifted to lower wavenumbers allowing observation of the protein amide I band in solution without the overlapping water contribution.

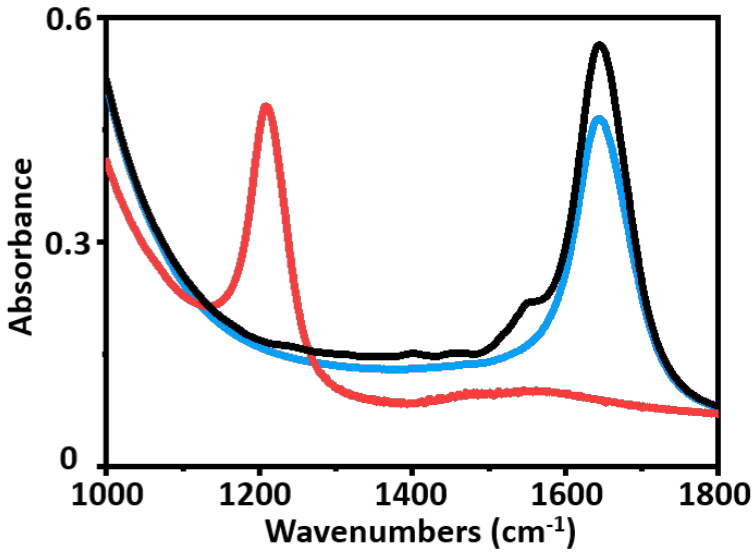

Figure 3.1: *The IR absorption spectra of D_2O (red), H_2O (blue) and serum (black).*

However, deuteration is an imperfect solution as D_2O is non-physiological, which raises fundamental questions regarding the impact of solvent isotope exchange on the ability to measure natural protein structural dynamics. The kinetic isotope effect is well known while changes in vibrational coupling between solvent and protein in deuterated systems arising

from the removal of spectral overlap between the amide I band and the δ_{H-O-H} mode may alter energy transfer dynamics and impact upon function in ways currently unknown.[22] The use of D_2O also leads to practical issues surrounding incomplete proton exchange and elevated cost. These severely limit applications of protein IR spectroscopy in the biomedical or commercial sector where a hydrogen/deuterium (H/D) exchange is either impractical or viewed as economically unfeasible.

IR absorption methods are fast, label-free and provide broad chemical coverage however current IR spectroscopic studies of blood serum employ dried samples to avoid the problem of water absorption, which can introduce artefacts from the drying process.[23] Drying or filtering samples takes time and has been linked to measurement uncertainties from sample processing.[24] Water spectra can also be obtained alongside serum spectra and subtracted, but this can be inaccurate and is a subjective process.[25] Attenuated total reflection (ATR) methods enable studies in aqueous liquids have been used to detect ovarian cancer and the severity of brain cancers however such linear spectroscopic methods cannot separate albumin and globulin signals or identify contributions from individual globulin proteins.[26,27] Thus, a practical, fast spectroscopic method capable of measuring AGR without the need for serum pre-processing alongside the ability to clearly differentiate key proteins would offer a major step forwards in healthcare technology.

In this chapter, measurements of the first 2D-IR measurements of proteins in water are explored. Using blood serum as an exemplar system, it is evident that H/D exchange is not necessary for label-free protein studies in H_2O using 2D-IR spectroscopy. Typical serum protein concentrations in humans are in the sub mM range (35–50 mg/mL albumin corresponds to 0.5–0.7 mM; 25–30 mg/mL γ-globulins ~0.15–0.25 mM), which correspond closely to those used for 2D-IR spectroscopic studies of proteins in D_2O.[28] Measurements of proteins in both H_2O and D_2O solvents show that deuteration slows down the vibrational relaxation dynamics of serum proteins. The enhanced spectral resolution of 2D-IR relative to IR absorption and the unique 2D-IR lineshapes enable accurate differentiation of protein signals of proteins with different secondary structures in the complex aqueous serum environment, allowing quantification of the AGR, and even when their secondary structure composition is similar for the cases of the γ-globulins sub groups IgG, IgA and IgM.

3.3. Experimental Techniques

3.3.1. Sample Preparation

Pooled equine serum, serum albumin (bovine), γ-globulins (bovine), IgG, IgA and IgM (human) were obtained from Sigma Aldrich and used without further purification. Measurements of individual proteins were performed using aqueous Tris buffer (pH ~7.5) to mimic the pH of the serum samples. To study the spectroscopy of serum samples at a range of AGR values, γ-globulin was spiked into pooled horse serum at concentrations of 30, 15, 7.5, 3.8, 1.9, 0.9 and 0.5 mg/mL. Yielding a total of 8 samples (7 spikes and pure serum). These concentrations cover a clinically relevant range and encompass large variations (>10 mg/mL) which are typically seen in the globulin fraction.[8–10]

To measure IR spectra in water using transmission mode, the sample thickness was carefully controlled in the cell set-up to avoid saturation of the δ_{H-O-H} mode of water at 1650 cm^{-1}. The same cells (Figure 2.4) are used in the 2D-IR experiments after FTIR data collection and so in order for the sample absorbance not to saturate both the FTIR and 2D-IR array detectors, an FTIR absorbance less than 0.6 is required. This is typically achieved by using a polytetrafluoroethylene (PTFE) spacer of known size, however the smallest available spacer, 6 µm, produced an absorption amplitude that was greater than 0.6. As a result of this, no spacer was used and instead the FTIR instrument was used to regulate the sample thickness. As the combination band of the bending mode and librational modes of water, situated at 2130 cm^{-1}, is also present undisturbed in the serum samples studied in this thesis, it was used to closely monitor the sample thickness. This was achieved by slowly tightening the fastening ring on to the cell body to gradually reduce the sample thickness and measuring the sample absorbance (Figure 3.2) at each stage.

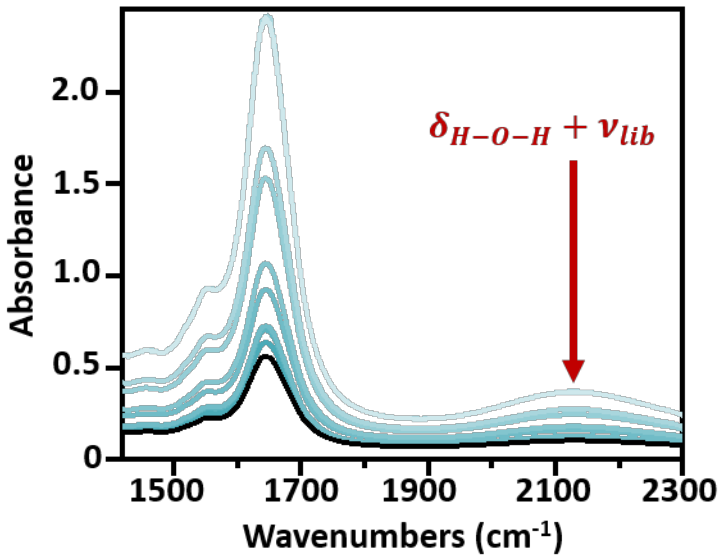

Figure 3.2: *FTIR spectra of a serum sample in the process of tightening the cell fastening ring and gradually reducing the sample thickness until the desired sample thickness (black) is achieved.*

An absorption of 0.1 at the combination band results in a serum amide I peak less than 0.6 and so this value for the combination band ($\pm$ 0.01) was achieved for each sample. In order to calculate the thickness this water combination band absorption corresponds to, the linear relationship between absorbance and path length is exploited here using a reference measurement of water taken at a sample thickness of 6 µm:

$$\frac{A_{ref}}{l_{ref}} = \frac{A_s}{l_s} \tag{3.1}$$

where the subscripts *ref* and *s* denote the absorbance and path length of the 6 µm reference and serum samples being measured. Based on the molar extinction coefficient of water, the absorbance of 0.1 ± 0.01 at 2130 cm^{-1} corresponds to a sample thickness of 2.75 ± 0.25 µm.

3.3.2 IR Absorption Spectroscopy

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^{-1} in the spectral region 400–4000 cm^{-1}. A background spectrum was measured before

each sample and subtracted following scaling to the amplitude of the δ_{H-O-H}+ υ_{libr} mode. Each measurement was made in triplicate.

3.3.3. PP and 2D-IR Spectroscopy

PP and 2D-IR spectra were recorded using the ULTRA laser spectrometer[29,30] at the Central Laser Facility using the Fourier transform 2D-IR method employing a sequence of three mid-IR laser pulses arranged in a pseudo pump–probe beam geometry, as described in Chapter 2. Comparable results were obtained irrespective of the use of a scanning interferometer or pulse shaper to deliver the sequence of two pump pulses. Mid-IR pulses with a temporal duration of <50 fs; a central frequency of 1650 cm^{-1} and a bandwidth of ~400 cm^{-1} were obtained.

All 2D-IR spectra were recorded at a waiting time (T_w) of 250 fs between pump and probe pulses. Each measurement was made in triplicate using identical sample conditions (described above) to those used for IR absorption measurements.

3.4. Results and Discussion

All 2D-IR spectral processing and analysis was carried out using a custom made script on the statistical analysis software programme, R.[31] Prior to the analysis described in the text, a 2nd order polynomial baseline subtraction was performed. Normalisation to the albumin peak was done for illustrative purposes however ratios within datasets are acquired and remain unchanged before/after normalisation. For the data in section 3.4.3.3 vector normalisation was carried out.

3.4.1. Blood Serum Spectroscopy

Infrared absorption spectra of equine serum samples spiked with 0–30 mg/mL of γ-globulins are shown in Figure 3.3(a). The dominant feature of serum is near 1650 cm^{-1} and is assignable to overlapping contributions arising from the δ_{H-O-H} mode of water (Figure 3.3(a, black)) and the amide I mode of the serum protein component. The water contribution is accounted for by spectral subtraction (Figure 3.3(b)). The resulting amide I band of the serum proteins is largely featureless, but gains in amplitude as the added quantity of γ-globulins is increased.

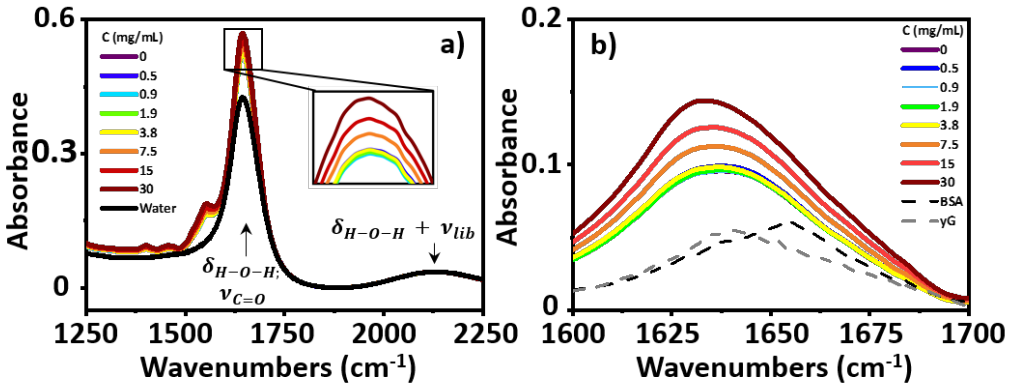

Figure 3.3: *(a) IR absorption spectra of serum samples spiked with γ-globulins, pure H_2O is shown in black. The inset shows an expansion of the tip of the peak due to the H–O–H bending mode of water and the protein amide I band at 1650 cm^{-1}. (b) IR absorption spectra of serum spiked with γ-globulins from (a) following subtraction of the H_2O spectrum. The strong overlap of their respective amide I bands prevented quantitative determination of the AGR from IR absorption data.*

The IR absorption spectra of the individual serum albumin and γ-globulin protein components are shown as black and grey dashed lines in Figure 3.3(b), respectively. These individual protein spectra indicate that the amide I absorption lineshape below 1650 cm^{-1} has a greater contribution from the γ-globulin proteins while the lineshape above 1650 cm^{-1} has a larger contribution from the serum albumin. The strong overlap of the amide I absorptions with water prevents quantitative determination of the AGR using IR absorption spectroscopy. However works demonstrating the effectiveness of second order derivatives have shown promise in protein identification and could be utilised to help resolve overlapping peaks and isolate signals.[32] For fair comparison to 2D-IR spectra, second derivatives were not utilised here.

By contrast to the IR absorption spectra (Figure 3.3 and 3.4(a)), the 2D-IR spectrum of pure serum shows considerable structure (Figure 3.4(b)). The negative feature (red) located on the 2D-IR diagonal near 1650 cm^{-1} is assigned to the $v = 0 - 1$ transitions of modes observed in the IR absorption spectrum and contains two distinct contributions with pump frequencies of 1639 and 1656 cm^{-1} (arrows). Positive (blue) peaks due to the accompanying $v = 1 - 2$ transitions are shifted to lower probe frequencies by the vibrational anharmonicity.

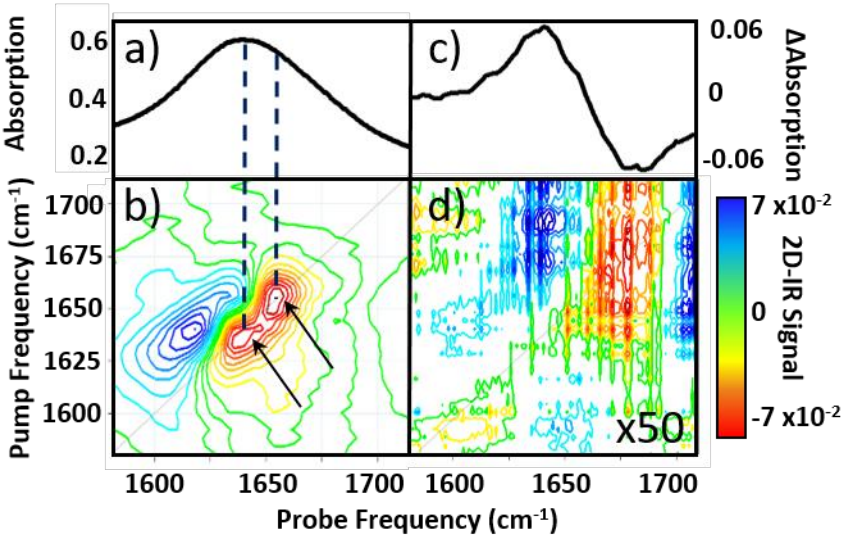

Figure 3.4: *(a) IR absorption and (b) 2D-IR spectrum of pure serum in the amide I region. (c) IR pump–probe spectrum of H_2O at a pump–probe time delay of 300 fs, black arrows show 2 distinct peaks at 1639 and 1656 cm^{-1}. (d) 2D-IR spectrum of H_2O, magnified 50x. The spectra in (b) and (d) were both obtained with a waiting time (T_w) of 250 fs and are plotted on the same scale, see colour bar.*[33]

However unlike the IR absorption data, the 2D-IR signal of water is significantly weaker than that of the serum under the same sample conditions. Comparison of 2D-IR spectra of serum and water is shown in Figure 3.4(b,d)). Although weak, the measured 2D-IR response of water was found to be in good agreement with previous observations.[34–36] The corresponding IR pump–probe spectrum at a pump-probe delay time of 300 fs is also shown for comparison (Figure 3.4(c)) and is in good agreement with previous work.[37]

As 2D-IR signals are dependent upon the 4th power of the vibrational transition dipole moment, this leads to enhancement of the strong amide I mode of the biological macromolecules relative to the more plentiful, but weakly absorbing, water molecules.[38–42] The extinction coefficient for the serum proteins is approximately 15,000 times larger than that of the water[43] at 1650 cm^{-1} and so despite the large absorbance of the δ_{H-O-H} mode of water, the 2D-IR response of proteins is the dominant feature in the serum spectrum. This results in a spectrum showing only the protein signature without the overlapping water contribution.

Comparison of the IR absorption and 2D-IR spectra of pure serum (Figure 3.5(a,d)) to the spectra of serum albumin (Figure 3.5(b,e)) and γ-globulins (Figure 3.5(c,f)) when obtained individually in water under the same conditions, allows assignment of the two peaks seen at 1656 and 1639 cm^{-1} in the serum spectrum. These two features are assigned to the albumin and the globulin components respectively.

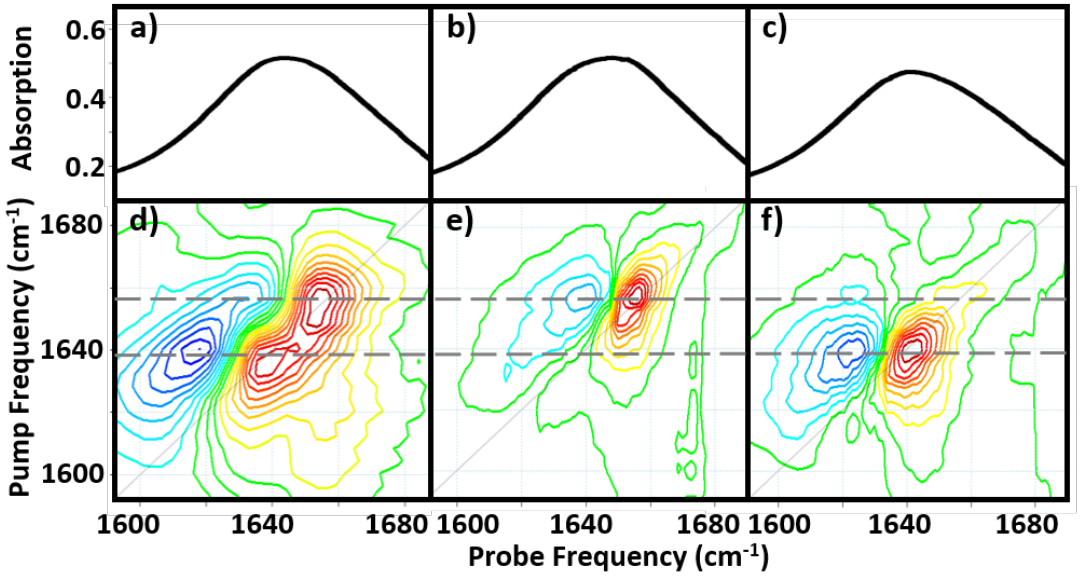

Figure 3.5: *IR absorption spectra of (a) serum (b) serum albumin (c) γ-globulins. 2D-IR spectra of (d) serum (e) serum albumin (f) γ-globulins. Dashed grey horizontal lines show peak positions of albumin and γ-globulins. The colour scale is as shown in Figure 3.8.*[33]

The difference in frequency of the two protein signals arises from the fact that serum albumin has a largely α-helical secondary structure while the globulins have a higher proportion of β-sheet, which shifts the centre of mass of the amide I band to a lower frequency.[40]

The fact that these peaks are separable in the 2D-IR spectrum, whereas they were poorly-resolved in the IR absorption spectrum is due to the non-linear nature of the 2D-IR signal. While each mode appearing in the IR absorption spectrum is visible on the diagonal of the 2D-IR plot, the higher-order dependence of the 2D-IR signal upon the transition dipole moment means that the lineshapes appearing on the diagonal of a 2D-IR spectrum are narrower than those found in the IR absorption spectrum.[39] This leads to the appearance of two well-resolved peaks along the diagonal of the 2D-IR plot (Figure 3.5(d)), where only one broad signal was observed in the IR absorption spectrum (Figure 3.5(a)).

3.4.2. Protein Dynamics

As all 2D-IR protein studies to date have been investigated in D_2O, it is important to establish the impact of solvent isotopic exchange on the vibrational dynamics of the serum proteins. To do this, IR pump–probe data were obtained for bovine serum albumin, the γ-globulins and the γ-globulin protein IgG in both H_2O and D_2O. The case of serum albumin will be discussed first followed by the γ-globulins and IgG and a table (Table 3.1) summarising the results can be found at the end of this section. Single measurements are shown however lifetimes calculated are the average of the measurement in triplicate.

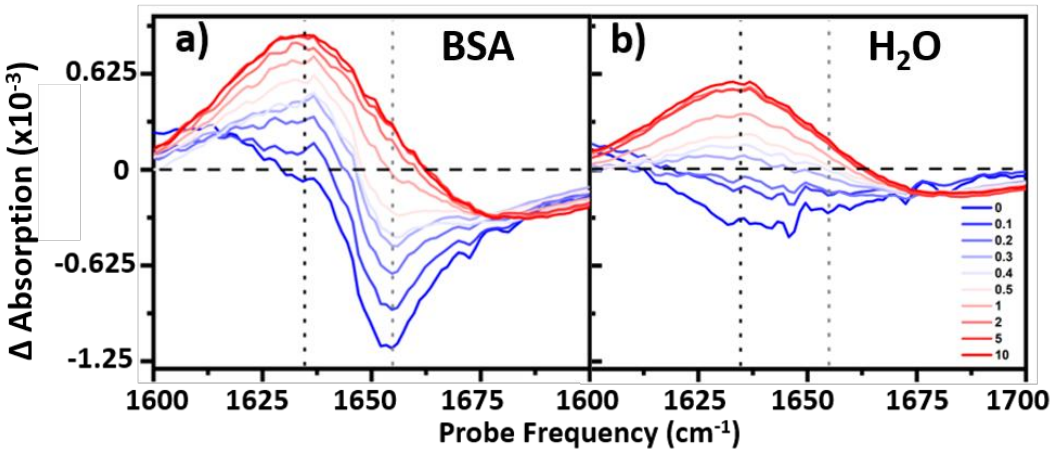

Figure 3.6: *IR pump–probe spectra of (a) serum albumin and (b) H_2O. Pump-probe delay times range from 0 ps (blue) to 10 ps (red). Horizontal dashed line indicates 0 change in absorption. Vertical dashed lines indicate peak positions at 1635 cm^{-1} (dark grey) and 1656 cm^{-1}(light grey) respectively.*

The IR pump–probe spectrum is equivalent to the 2D-IR response projected onto the probe frequency axis[39] and Figure 3.6 shows the pump-probe response of BSA (a) and H_2O (b) for a range of pump-probe delay times. In the case of BSA in H_2O, the amide I $v = 0-1$ peak is overlapped by the bleach of the $v = 0-1$ transition of the δ_{H-O-H} mode of water and so must be accounted for in the protein dynamics. The H_2O signal is significantly smaller in amplitude than the protein response (<20% of the total amplitude at T_w= 0 ps) and is well-represented by a bi-exponential function featuring a 0.22 $\pm$ 0.04 ps decay and a 1.21 $\pm$ 0.19 ps rise-time due to the effects of residual sample heating, which persists to T_w values up to 10 ps (Figure 3.7(a)). This behaviour of the water band is in agreement with previous work,[44]

and this fast decay of the water O-H bending mode is due to rapid depopulation of the $v = 0 - 1$ vibration allowing the transfer of energy into intermolecular modes. This in turn produces a heated ground state, where the temperature rise of the sample causes a red-shift in the peak frequency (Figure 3.6).

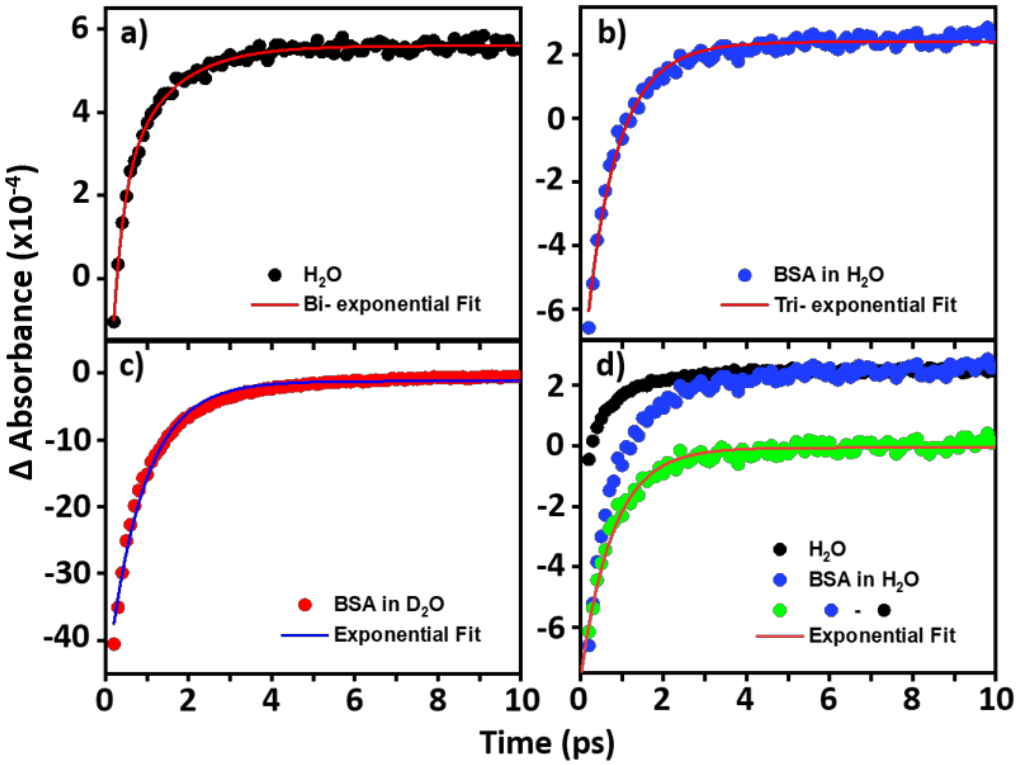

Figure 3.7: *Temporal variation of the amide I $v = 0 - 1$ bleach signal observed in the IR pump–probe spectrum of a) H_2O b) BSA in H_2O c) BSA in D_2O . d) The water response (black) has been subtracted after scaling the signals of water to the BSA (blue) at a frequency of 1650 cm^{-1}, resulting in only the protein dynamics (green). Solid lines show fits to the data using exponential functions.*

The peak dynamics of the amide I $v = 0 - 1$ transition of BSA in H_2O (Figure 3.6(a) light grey dotted line) are investigated using two different methods. The first is well-represented by a tri-exponential decay. This encompasses the bi-exponential result from the water dynamics with t_1 and t_2 fixed within error, and the third exponential gives a vibrational relaxation time of 0.78 ps (Figure 3.7(b)). In the second approach, shown in Figure 3.7(d), the water response has been subtracted after scaling the signals of water and the protein to the T_w signal at 5 ps where no protein contribution is observed. Using this method, the water signal is subtracted

and so a single exponential vibrational relaxation time of 0.81 ps is observed for the amide I $v = 0 - 1$ transition for the serum albumin in water. In the case of the serum albumin in D_2O, the dynamics of the peak of the amide I $v = 0 - 1$ transition were well-represented by single exponential decay with a lifetime of 0.89 ps (Figure 3.7(c)). Both methods result in a reduced vibrational relaxation time, approximately 10%, when the protein is surrounded by H_2O compared to D_2O.

The dynamics of the peak of the amide I $v = 0 - 1$ transition of γ-globulins and IgG are also investigated using the two methods mentioned above. The first approach using the tri-exponential fit yields a third exponential with a vibrational relaxation time of 0.75 and 0.76 ps for the γ-globulins and IgG (Figure 3.8(a,d)), respectively.

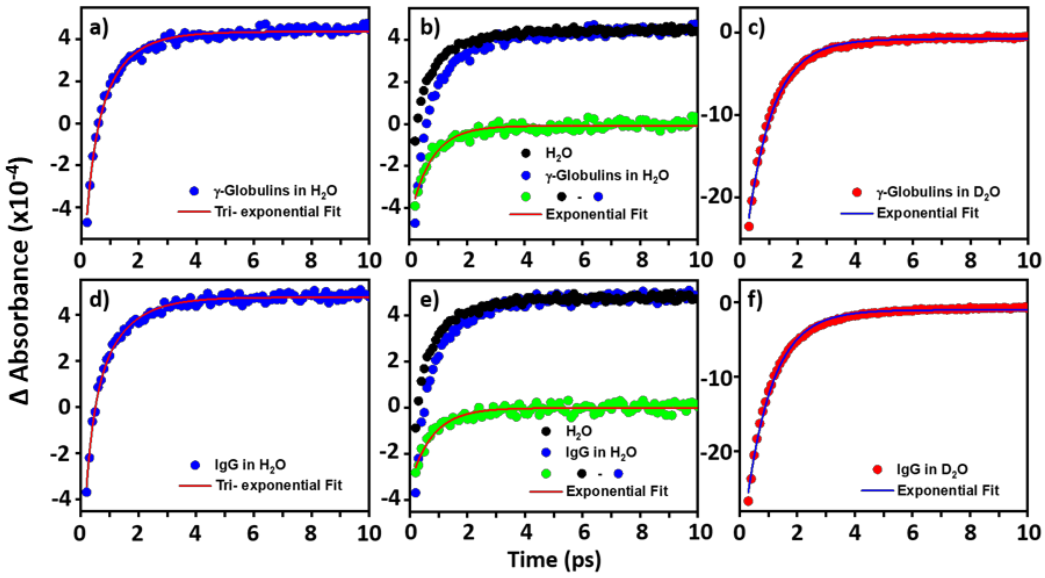

Figure 3.8: *Temporal variation of the amide I $v = 0 - 1$ bleach signal observed in the IR pump–probe spectrum for the γ-globulins (a-c) and IgG (d-f) a) γ-globulins in H_2O. b) H_2O (black) and γ-globulins in H_2O (blue). The water response has been subtracted after scaling the signals of water to the γ-globulins at the frequency of 1650 cm^{-1}, resulting in only the protein dynamics (green). c) γ-globulins in D_2O. d) IgG in H_2O. e) H_2O (black) and IgG in H_2O (blue). The water response has been subtracted after scaling the signals of water to the γ-globulins at the frequency of 1650 cm^{-1}, resulting in only the protein dynamics (green). f) IgG in D_2O. Solid lines show fits to the data using exponential functions.*

In the second approach, shown in Figures 3.8(b,e), where the water response has been subtracted, a single exponential vibrational relaxation time of 0.74 and 0.78 ps is observed for the amide I $v = 0 - 1$ transition of the γ-globulins and IgG in water, respectively. In the case of γ-globulins and IgG in D_2O, the dynamics of the peak of the amide I $v = 0 - 1$ transition was well-represented by single exponential decay with lifetimes of 0.9 and 0.93 (Figure 3.8(c,f)) ps, respectively. A summary of these results is shown in Table 3.1.

Table 3.1: *Summary of the vibrational relaxation times of the three proteins studied in both D_2O and H_2O.*

Vibrational Relaxation Times	BSA	γ-Globulins	IgG
D_2O	*0.89 ps*	*0.9 ps*	*0.93 ps*
H_2O tri-exponential	*0.78 ps*	*0.75 ps*	*0.76 ps*
H_2O subtraction	*0.81 ps*	*0.74 ps*	*0.78 ps*

It is important to note that the vibrational relaxation time of the amide I band of a large protein is, by definition, a weighted average over a large number of coupled amide I oscillators.[45] In the case of the γ-globulins, this is a mixture of proteins. It has been shown previously using fibrillar aggregates of short chain peptides that the lifetime of the amide I band is sensitive to secondary structure and to the level of solvation of a given residue.[46] However, the consistent observation of a reduction in vibrational lifetime by around 10% using both methods upon moving from D_2O to H_2O for each of the three proteins studied indicates that the average lifetime of the amide I mode of these proteins is being reduced. These findings are consistent with previous studies of solvent isotope-dependent vibrational dynamics, suggesting that isotopic exchange of the solvent may be responsible for altering the observed protein dynamics.[47]

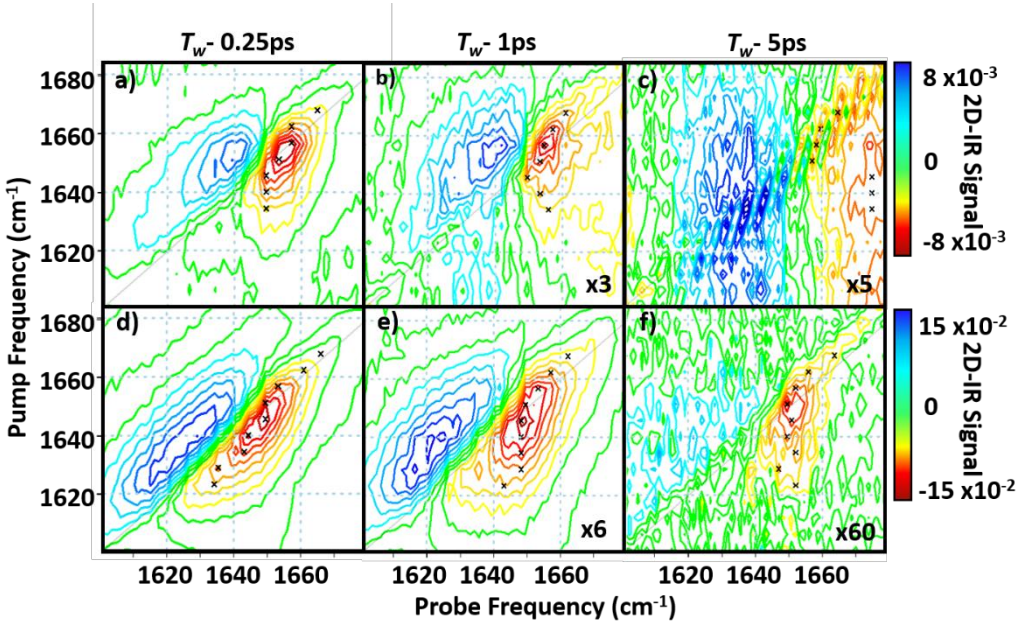

Figure 3.9: *2D-IR spectra of BSA in H_2O (a,b,c) and D_2O (d,e,f) at the waiting times 0.25 ps (a,d), 1 ps (b, e) and 5 ps (c,f). All H_2O spectra have been plotted on the same scale and the later waiting times have been magnified to reflect this. All D_2O spectra have also been plotted on the same scale and magnified as shown. Black crosses denote the centre line slope of each spectra.*

In addition to pump–probe data, spectral diffusion of the 2D-IR amide I lineshape of the serum albumin in water was compared to that in D_2O (Figure 3.9). A series of waiting times studies for BSA (30 mg/mL) in both H_2O (~2.75 µm spacer) and D_2O (25 µm spacer) solvents were studied to compare the spectral diffusion dynamics of BSA in the two different solvents. The centre line slope (CLS)[48] was utilised to attempt to extract these dynamics. Indications of changes in linewidths and spectral diffusion processes in H_2O were present in the data. The amide I lineshape in H_2O appears more circular in nature (Figure 3.9(a)) compared to the more elongated peak in D_2O (Figure 3.9(d)), suggesting different inhomogeneous broadening or changes in the spectral diffusion dynamics. However the CLS analysis did not clearly reflect this trend (Figure 3.9, black crosses). At late waiting times the water response due to heating dominates the 2D-IR signal and thus obscures any measurement of the protein CLS. Indications of this can be seen in the PP dynamics of each of the three proteins in H_2O (Figures 3.7 and 3.8) where the dynamics do not tend to zero but instead exceed zero and level off with a positive signal. This response is due to the redistribution of energy within the sample

eventually leading to energy transfer to the solvent which induces a red-shift of the absorption.[37] In relation to analytical studies of complex protein mixtures such as serum, the relative dynamics of the signals due to water and proteins can be utilised to optimise the difference between the 2D-IR protein response and that of water. Figure 3.10 shows the faster relaxation time of the δ_{H-O-H} mode (blue) than the protein amide I response of serum (black), which has a vibrational lifetime of 0.83 $\pm$ 0.09 ps at 1650 cm^{-1}. This means that at a waiting time of 250 fs, the water signal is at a minimum prior to the onset of the small rising signal due to water heating.

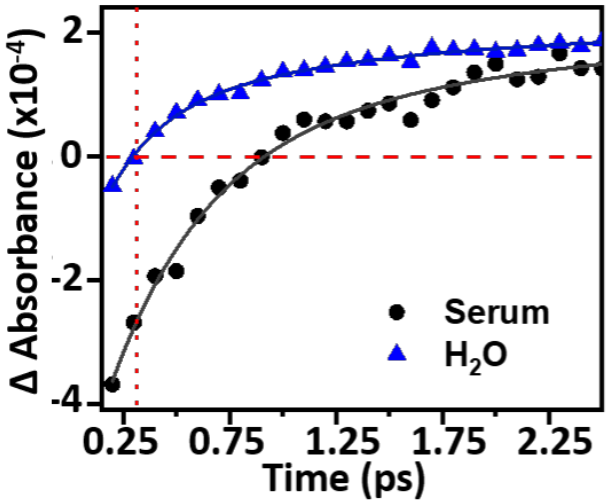

Figure 3.10: *Temporal variation of the bleach signal observed in the IR pump–probe spectrum of H_2O (blue) and serum (black) at a frequency of 1650 cm^{-1}. Solid lines show fits to the data using a single exponential function with a decay time constant of 0.22 ps (H_2O) and 0.83 ps (serum) respectively.*

This results in a spectrum showing only the protein signature at this waiting time and on this basis, the following spectral analysis of protein samples is carried out using a T_w of 0.25 ps.

3.4.3 2D-IR Biofluid Analysis

Having now measured proteins in water, the focus is turned to quantify the relative concentrations of the two major components of serum, the serum albumin and globulins.

3.4.3.1 AGR Scaling Factor

The two peaks observed in the serum 2D-IR spectra (Figure 3.5(d)) have been assigned to the albumin (1656 cm^{-1}) and γ-globulins (1639 cm^{-1}) fractions. However, the non-linearity of the 2D-IR signal intensities means that additional factors to the simple Beer-Lambert terms must be taken into account. The relative molecular weights of these two proteins are very different (albumin of ~66 kDa and γ-globulins of ~150 kDa) and their differing secondary structural elements gives 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.[41] The quantitative ratio of the serum albumin and γ-globulin 2D-IR signals were obtained by measurements of known concentrations of the two proteins under as close as possible to identical conditions (Figure 3.11). Measurements were taken at a concentration of 30 mg/mL for the γ-globulins and 50 mg/mL for the serum albumin. These concentrations were chosen to maximise the 2D-IR signal within the solubility limit of the protein. From the relative maximum amplitudes of the 2D-IR spectrum diagonals of serum albumin and γ-globulins it can be established that, per unit concentration (mg/mL), the albumin signal is 1.8 ± 0.1 times larger than the γ-globulin response.

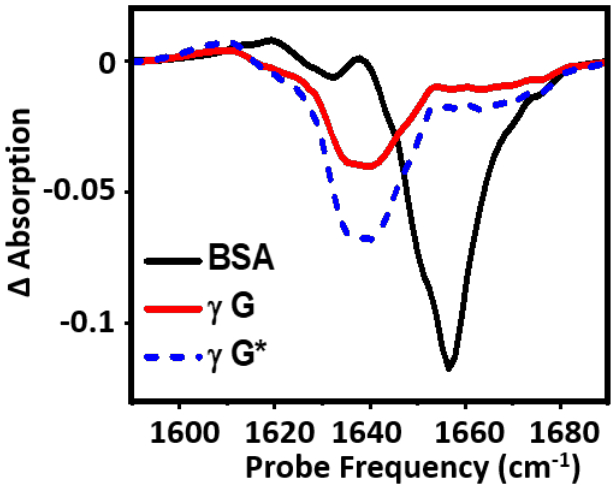

Figure 3.11: *Diagonal slices through 2D-IR spectra of 50 mg/mL of serum albumin (black), 30 mg/mL of γ-globulins (red). Blue dashed lines shows the slice due to the γ-globulins scaled to reflect a 50 mg/mL concentration.*[33]

3.4.3.2 Using 2D-IR to Measure the AGR

Using 2D-IR spectra of a range of serum samples spiked with differing quantities of γ-globulins (Figure 3.12), an attempt was made to utilise the spectral differentiation of the peaks assigned to the albumin and globulin protein components (Figure 3.5) to quantify the AGR directly from 2D-IR spectra.

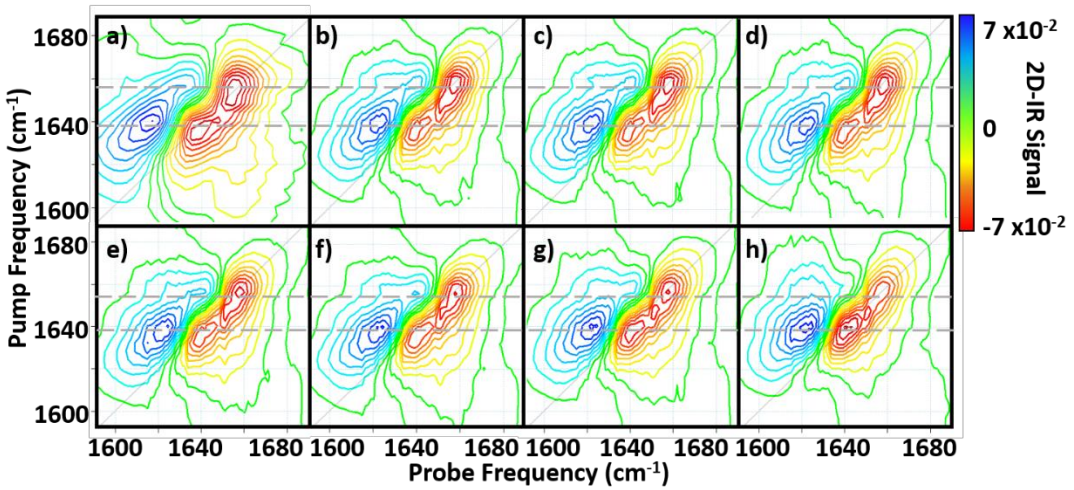

Figure 3.12: *2D-IR spectra of serum spiked with γ-globulins in the following concentrations mg/mL): a) 0 b) 0.5 c) 0.9 d) 1.9 e) 3.8 f) 7.5 g) 15 h) 30. Dashed grey horizontal lines show the peak positions of albumin and γ-globulins discussed in the preceding text. All spectra have been plotted on the same scale, see colour bar.*[33]

Three separate approaches were employed to determine the AGR values of the serum samples: (i) the 2D-IR spectrum diagonal, (ii) pump-frequency slices and (iii) linear combination analysis of 2D-IR spectra.

(i) The 2D-IR spectrum diagonal approach extracts the diagonal of each 2D-IR spectrum, resulting in two distinct peaks at 1656 cm^{-1} and 1639 cm^{-1}, which were assigned to the albumin and globulin fractions respectively (Figure 3.13 (a, b)). In this approach, the ratio of the absolute values of the amplitudes assigned to the albumin and globulin fractions were used to determine the AGR (Figure 3.13(a)). Scaling of the globulin amplitude by the scaling factor of 1.8 was implemented to account for the measured differences in signal amplitude between albumin and the γ-globulins per unit concentration.

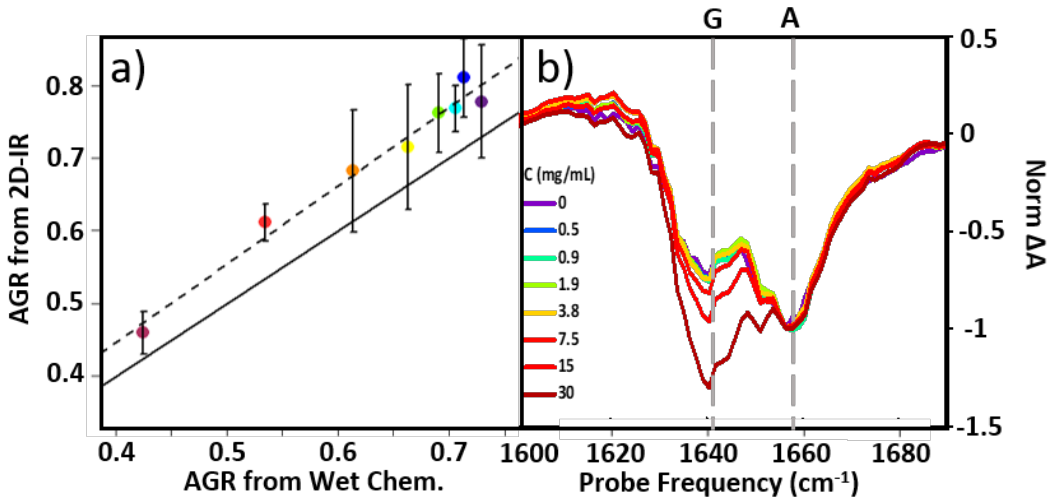

Figure 3.13: *AGR of serum samples from 2D-IR spectroscopy obtained using the 2D-IR diagonal method. (a) the solid black line indicates the actual AGR of the samples. Error bars show 2σ variation from triplicate measurement. Dashed lines show linear fits to the experimental AGR values. (b) The spectral basis of the method where the albumin and globulin signals are indicated with vertical dashed grey lines.*

(ii) The pump-frequency slice method utilises horizontal slices through the 2D-IR spectrum at 1656 cm^{-1} and 1639 cm^{-1}, assigned to peaks of the albumin and globulin signals respectively (Figure 3.14(a, b)). Using the amplitudes of the $v = 0-1$ peaks due to albumin and globulins, the ratio of the absolute values of the maximum amplitudes 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 (Figure 3.14(a)) following application of the scaling factor (1.8) to the globulin signal.

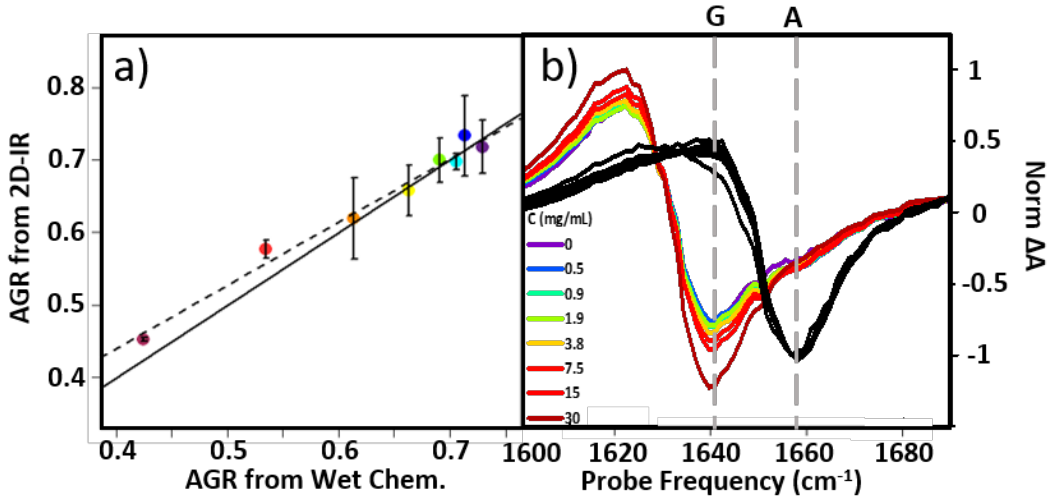

Figure 3.14: *AGR of serum samples from 2D-IR spectroscopy obtained using the 2D-IR pump-frequency slice method. (a) the solid black line indicates the actual AGR of the samples. Error bars show 2σ variation from triplicate measurement. Dashed lines show linear fits to the experimental AGR values. (b) The spectral basis of the method where the albumin and globulin signals are indicated with vertical dashed grey lines.*

(iii) Linear combination analysis fits the serum 2D-IR spectrum to the linear sum of the independent 2D-IR spectra obtained for the serum albumin and globulins (Figure 3.15(a, b)). All 2D-IR spectra were normalised to the albumin peak at 1656 cm^{-1} before the analysis and the coefficients of the relative contributions of the two protein spectra were then used to evaluate the AGR (Figure 3.15(a)), following scaling of the globulin fraction by 1.8.

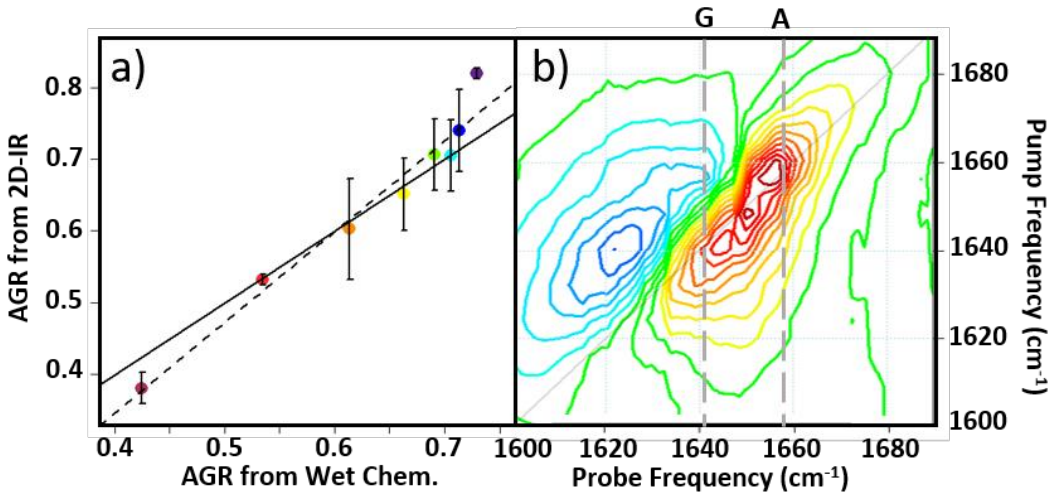

Figure 3.15: *AGR of serum samples from 2D-IR spectroscopy obtained using the 2D-IR linear combination analysis method. (a) the solid black line indicates the actual AGR of the samples. Error bars show 2σ variation from triplicate measurement. Dashed lines show linear fits to the experimental AGR values. (b) The spectral basis of the method where the albumin and globulin signals are indicated with vertical dashed grey lines.*

The AGR values were measured from the 2D-IR spectra of the spiked serum samples using all three approaches while the variation over triplicate measurements was used to reflect the repeatability of each method. The values obtained are compared to the real value of the AGR (Figures 3.13(a), 3.14(a) and 3.15(a), solid black line). The 'actual' AGR value was determined via sending a sample of the as-received equine blood serum for standard laboratory testing at the University of Glasgow School of Veterinary Medicine (Glasgow, UK) and adding the quantity of the known γ-globulin spike to the globulin component. The total protein concentration was assayed using the Biuret method[8,15] and the albumin concentration was assayed using the Bromo cresol green (BCG) method.[49] However this technique assumes the difference between the total proteins and serum albumins is globulins. This non-

spectroscopic AGR measurement of the neat serum was repeated on three individual batches of serum to ensure repeatability.

All three methods for determining the AGR spectroscopically produced a linear relationship when the measured value (Figures 3.13(a), 3.14(a) and 3.15(a), points and dashed lines) was plotted against the actual AGR, over a clinically relevant concentrations. Ideal agreement between the actual and 2D-IR-measured AGR values is represented by the solid black line in Figures 3.13(a), 3.14(a) and 3.15(a), which denotes where the predicted and actual AGR values are identical, with the aim for the 2D-IR AGR values to lie along this line.

Of the three 2D-IR methods used, the pump slice approach (Figure 3.14) was the most accurate at the higher values of the AGR, which correspond most closely to the expected human clinical range of 1-2.[8,10] The equine serum test samples showed a somewhat higher globulin concentration (40 mg/mL) than is typical for humans (between 25-35 mg/mL). At lower AGR values, the agreement obtained with the pump-slice method was less effective, which could be caused by to the very large γ-globulin spike distorting the albumin response. The results obtained from the 2D-IR diagonals were good across the full range of the samples studied (Figure 3.13), with most 2D-IR-derived values being within the measurement error of the actual AGR value, though a constant offset from the actual AGR value was noted. As the spectrum diagonal selects the centre frequency of the albumin peak, it can be seen in Figure 3.12 that the globulin signal lies slightly to the right of the diagonal and this is accredited to differences in the anharmonicities of the proteins. Finally, the linear combination (Figure 3.15) yielded excellent agreement over the mid-range of the spiked samples (AGR = 0.5 – 0.7), but was less effective at the extremities.

Importantly, all three analysis methods are superior to direct IR absorption spectroscopy, which did not produce a reliable value of the AGR. Taking an average of the three analysis approaches (Figure 3.16) produced agreement with actual AGR values across the full range of samples, within the experimental uncertainty and is the best approach.

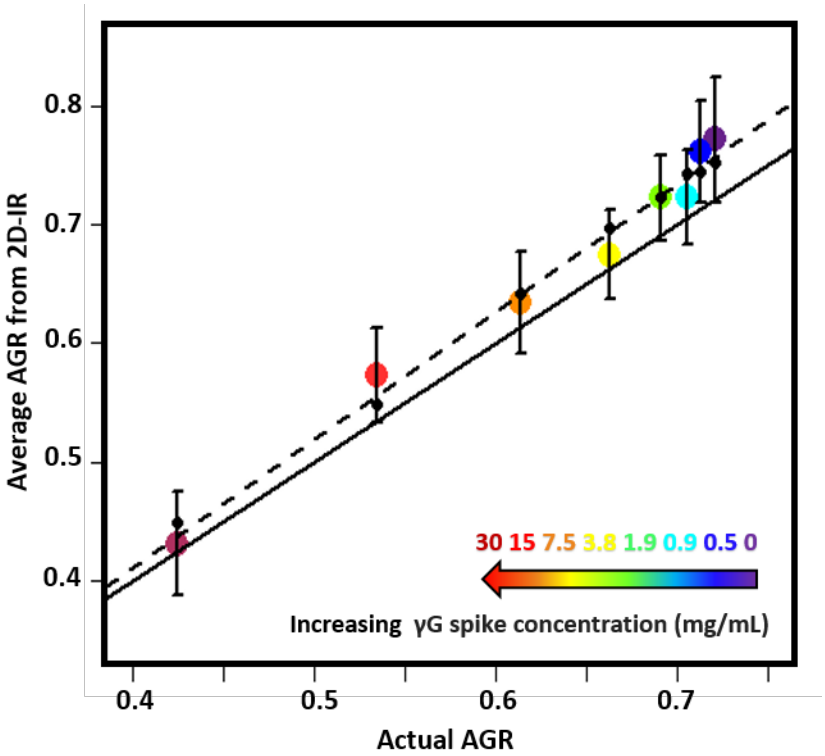

Figure 3.16: *AGR obtained from averaging the results in Figure 3.14(a, c, e). The AGR is predicted for each data point using leave one out tests, as shown by the solid black dots. The solid black line indicates the actual AGR of the samples. Error bars show 2σ variation. The dashed line shows a linear fit to the experimental AGR values.*

The accuracy of the method is tested using leave one out cross validation tests.[50] This involves removing an AGR result and repeating the linear regression and using it to predict the AGR of the removed data point. This is repeated for every AGR value and in all cases the results gave an accuracy to within the expected error of the measurement (Figure 3.16, black dots).

Overall, the 2D-IR measurements tested here show accuracy over a clinically-relevant range. Based on sample-to-sample variation, the accuracy of the 2D-IR derived AGR measurement was $\pm$ 0.03 ($\pm$ 4%). Direct comparisons with the current wet assay technique are not possible because these tests derive the AGR value from the difference in total protein and albumin concentrations and so do not directly measure globulin content, however typical quoted

accuracies are ~1%.[51] Although the spectroscopic approach is less accurate than current technologies, it is important to note that these are the first measurements of proteins in water using 2D-IR spectroscopy and this first demonstration holds promise for considerable scope for improvement of the accuracy through engineering approaches to sample path length repeatability and improved data collection protocols.

3.4.3.3. Using 2D-IR to measure Immunoglobulins in Serum

In principle, 2D-IR offers the possibility of going beyond the AGR measurement by virtue of the information-rich 2D lineshape of the proteins, which gives the opportunity to resolve more than albumin and the γ-globulins. The major protein constituents of the γ-globulins are IgG (80%), IgA (13%) and IgM (6%) and so represent realistic and clinically relevant test targets.[16,17]

2D-IR spectroscopy is used here to attempt to differentiate changes in the signal from the immunoglobulins from that of the albumin and γ-globulin contributions to a spectrum of a serum sample. The fact that IgG, IgA and IgM all have similar β-sheet rich structures to the γ-globulins is challenging and so it is necessary to identify regions of the 2D-IR spectrum where the response due to IgG, IgA and IgM proteins can be differentiated from the generic γ-globulin response. Figure 3.17(a, b, c) shows the 2D-IR amide I signal of IgG, IgA and IgM proteins, respectively.

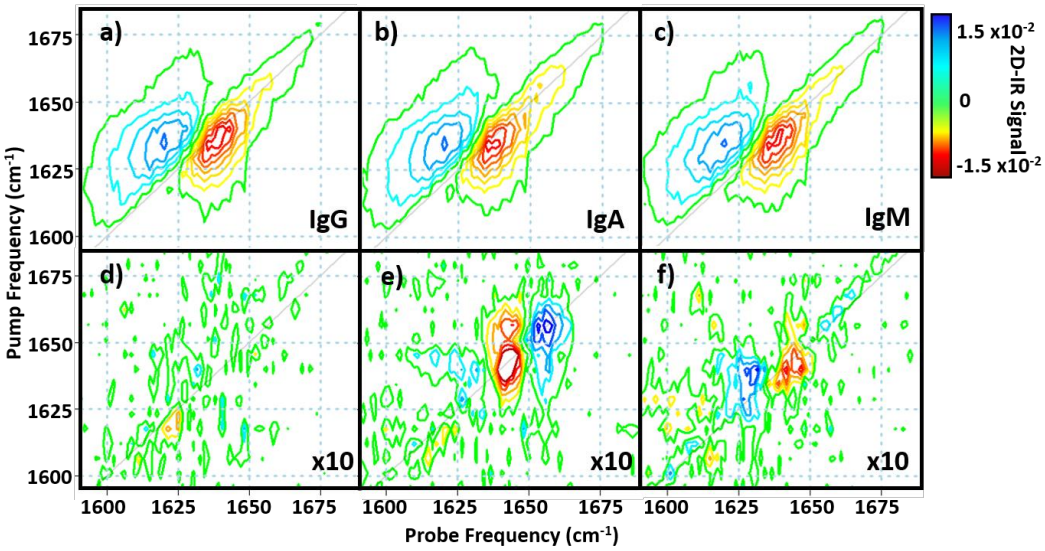

Figure 3.17: *2D-IR spectra of a) IgG, b) IgA and c) IgM in H_2O. (d, e, f) Difference spectra obtained by subtracting the 2D-IR response of the spectra in (a, b, c) from that of the γ-globulins. The spectral features have been magnified by a factor of 10 for clarity.*

As expected, the spectral features all appear very similar due to their comparable secondary protein structures and compositions. In order to highlight the subtle discrepancies between the signals, difference 2D-IR spectra (Figure 3.17(d, e, f)) were constructed by subtracting the average of triplicate measurements of IgG, IgA and IgM at a concentration of 30 mg/mL from the spectrum of the 30 mg/mL of γ-globulins.

The results of the difference spectra show that there is little variance between the spectrum of IgG and that of the γ-globulins (Figure 3.17(d)). Although somewhat trivial, this result is as expected due to IgG accounting for ~80% of the γ-globulins and acts as an effective control for the process. Owing to this, a concentration study of IgG with serum was not carried out.

By contrast, the difference spectra obtained for IgA and IgM do show regions of spectral differences with the γ-globulins. IgA in particular (Figure 3.17(e)) shows clear regions of decreased negative (red) spectral density in the diagonal region near ~1640 cm^{-1} and increases in the diagonal part of the amide I band (blue) near 1657 cm^{-1}. IgM also shows spectral differences to the γ-globulins in the diagonal region of the spectrum (Figure 3.17(f)), though the effect is less than for IgA and the amplitude of the difference signal is reduced. In order to determine whether the signals observed in the difference spectra are sufficient to quantify changes in serum levels of IgA and IgM, measurements were carried out on a range of serum samples spiked with additional quantities of IgA and IgM. The 2D-IR spectra of serum samples spiked with concentrations of IgA from 0–15 mg/mL are shown in Figure 3.18.

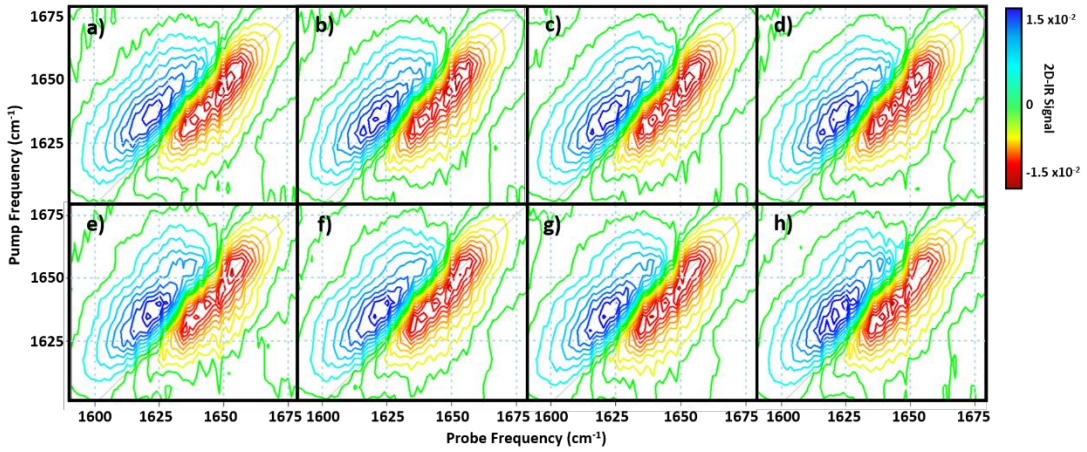

Figure 3.18: *2D-IR spectra of serum spiked with IgA in the following concentrations (mg/mL): a) 0 b) 0.25 c) 0.5 d) 1 e) 2.5 f) 5 g) 7.5 h) 15. All spectra have been plotted on the same scale, see colour bar.*[33]

It is important to note the positions of the peaks in the γ-globulins–IgA difference spectrum (Figure 3.17(e), 1640 and 1657 cm^{-1}) are not the same as those used to measure the γ-globulin fraction for the above AGR analysis (Figures 3.13, 3.14 and 3.15). Furthermore, the opposing sign of the two components gives two points of reference that can be used to separate the contributions from IgA and γ-globulins. To determine the ability of the 2D-IR spectrum to determine the IgA content, the ratio of the amplitudes on the spectrum diagonal at 1657 : 1640 cm^{-1} (the positive and negative peaks in the (γ-globulin–IgA) difference spectrum) is plotted as a function of IgA concentration (Figure 3.19(a)). A linear increase in the negative signal shows a correlation with increasing IgA content.

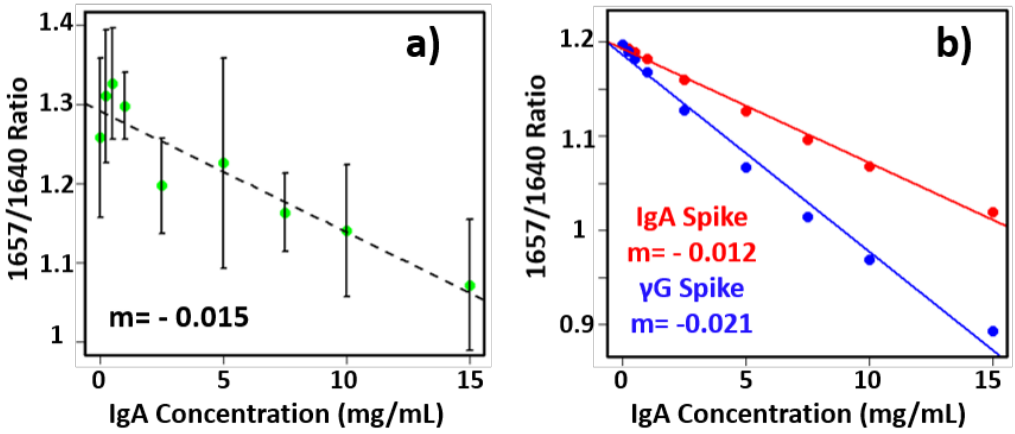

Figure 3.19: *a) The variation in 2D-IR signal at the peak of the difference spectral response for IgA as a function of added protein for the experimental data. The linear increase in the negative signal shows a correlation with protein content. b) Results of the ratio of calculations of serum with added IgA or γ-globulins.*

However, the analysis is difficult because the IgA and γ-globulin signals overlap strongly and there is no portion of the spectrum that is unique to either the γ-globulins or to IgA. In order to anticipate what to expect from the addition of IgA to serum, calculation of the spectra expected from this experiment using reconstructions from the individually-measured serum and IgA spectra (Figure 3.19(b)) are performed. The results show that if IgA is influencing the signal, the 1657 : 1640 cm^{-1} amplitude ratio should decrease as the IgA concentration is increased. Importantly, the gradient of the decrease in this ratio would be significantly shallower than that observed if just the generic γ-globulin response increased (Figure 3.19(b)).

It can be seen from a comparison of the measured (Figure 3.19(a)) and calculated (Figure 3.19(b)) data that the 2D-IR response recovered matches well with that expected for an IgA-specific signal increase. The gradient of -0.015 from the experimental data is close to the value of -0.012 derived from calculated data. Furthermore, the gradient of the measured amplitude ratio is much closer to that predicted for a change in IgA levels than for a change in γ-globulin fraction. The fact that the correlation with IgA levels persists down to ~1 mg/mL compares well with expected serum levels of IgA, which are in the range of 13% of 30 mg/mL (4 mg/mL).

Repeating this analysis to determine the ability of the 2D-IR spectrum to determine the content of IgM in serum and also showed a linear relationship between the measured 2D-IR signal at the peaks of the (IgM–γ-globulin) difference spectrum and the IgM concentration (Figure 3.20(a)). The IgM–γ-globulin difference spectra peaks shown in Figure 3.17(f) and the ratio of the amplitudes along the spectrum diagonal at the positions 1645 : 1629 cm^{-1} are used. It is again important to note the positions of the peaks in the γ-globulins–IgM difference spectrum are not the same as those used to measure the γ-globulin fraction for the above AGR analysis (Figures 3.13, 3.14, 3.15 and 3.19).

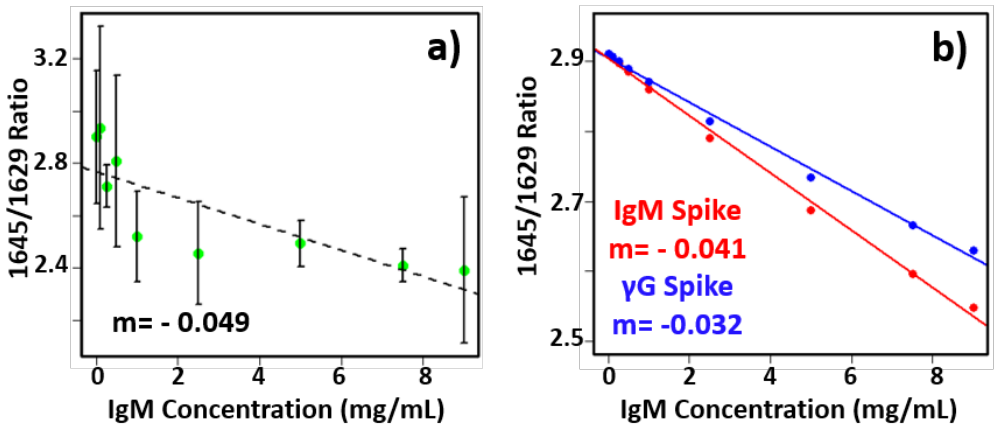

Figure 3.20: *a) The variation in 2D-IR signal at the peak of the difference spectral response for IgM as a function of added protein for the experimental data. The increase in the negative signal shows a correlation with protein content. b) Results of the simulations of serum with added IgM or γ-globulins.*[33]

However the smaller magnitude of the spectral differences between IgM and γ-globulins (Figure 3.17(f)) leads to a significantly noisier correlation when compared to the IgA result.

The calculated differences shown in Figure 3.20(b) show that a less clear separation of the IgM and γ-globulin responses would be anticipated. The gradient of -0.049 from the experimental data matches well with the value of -0.041 derived from calculated data compared to the value of -0.032 for the γ-globulin increase. Furthermore, the gradient of the measured amplitude ratio is closer to that predicted for a change in IgM levels than for a change in γ-globulin fraction. It is noted that a linear fit may not be the most appropriate given the result shown in Figure 3.20 however the use of spectral noise reduction techniques may help reveal this linear change more clearly.

3.5. Conclusions

2D-IR spectroscopy has demonstrated the capability to measure the molecular dynamics of proteins in water, without the need for H/D exchange of the solvent, thus increasing the physiological relevance of 2D-IR studies. 2D-IR instrumentation can perform accurate relative quantitative measurements of proteins in H_2O by probing the molecular system at a waiting time of 250 fs. The 2D-IR response obtained from the water background is negligible allowing isolation of the protein signal and clear separation of albumin and globulin signals in serum. These experiments have revealed that 2D-IR can measure the amide I band of protein samples at sub millimolar concentrations in water and confer a 'simple', robust approach to determining the AGR of serum using a single spectroscopic measurement, without the need for time consuming sample preparation or complex data analysis. In addition, 2D-IR offers the scope to go beyond the AGR measurement by virtue of the information-rich 2D lineshape of the proteins. The reported proof of concept experiments illustrate the promising potential for development of 2D-IR as a tool for differentiating IgA and IgM protein contributions from albumin and γ-globulins based on their 2D-IR responses.

Improvement in the accuracy could be achieved via the development of more advanced data analysis tools, spectral derivatisation, principal component analysis and partial least square regression techniques have all seen promising applications to serum analysis and can aid in maintaining spectral features while minimising noise.[2,32] Further work to standardise data collection protocols including the recent advances in pulse shaper technologies could also help improve the accuracy of data collection.

Whilst acknowledging that applying this approach to samples of unknown protein levels would be challenging, achieving absolute protein concentrations from 2D-IR spectra, as

opposed to relative ratios, is not possible using these methods. One of the main issues arises from repeatability of the thickness of the sample, meaning that, although relative peak heights within a spectrum are obtainable with high accuracy, as demonstrated, sample to sample variations in absolute concentrations are harder to achieve. In Chapter 5, the issue of sample to sample measurement repeatability is explored and a method has been devised to create an internal normalisation standard, allowing quantitative protein concentrations to be obtained.

3.6. References

1. Byrne, H.J., Baranska, N., Puppels, G.J, Stone, N. *et al.* Spectropathology for the next generation: Quo vadis? *Analyst* **140,** 2066–2073 (2015).

2. Baker, M.J., Hussain, S.R., Lovergne, L., Untereiner, V. *et al.* Developing and understanding biofluid vibrational spectroscopy: a critical review. *Chem. Soc. Rev.* **45,** 1803–1818 (2016).

3. Adkins, J.N., Varnum, S.M., Auberry, K.L, Moore, R.J. *et al.* Toward a Human Blood Serum Proteome. *Mol. Cell. Proteomics* **1,** 947–955 (2002).

4. Hu, S., Loo, J. & Wong, D. Human body fluid proteome analysis. *Proteomics* **6,** 6326–6353 (2006).

5. 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* **6,** 961–967 (2006).

6. Mire-sluis, A.R., Barret, Y.C., Devanarayan, V., Koren, E. *et al.* Recommendations for the design and optimization of immunoassays used in the detection of host antibodies against biotechnology products. *J. Immunol. Methods* **289,** 1–16 (2004).

7. Großerueschkamp, F., Bracht, T., Diehl, H.C., Kuepper, C. *et al.* Spatial and molecular resolution of diffuse malignant mesothelioma heterogeneity by integrating label-free FTIR imaging, laser capture microdissection and proteomics. *Sci. Rep.* **7,** 1–12 (2017).

8. Busher, J. T. Serum albumin and globulins. in *Clinical Methods: The History, Physical, and Laboratory Examinations.* (eds. Hurst, J. ., Walker, H. . & Hall, W. .) 497–500 (Butterworths, 1990).

9. Pathology Harmony Group & Clinical Biochemistry. Harmonisation of Reference Intervals. (2011). Available at: http://www.acb.org.uk/docs/Pathology Harmony for web.pdf. (Accessed: 25th February 2020)

10. Eaton, P. & Fuggle, S. *Clinical Biochemistry Reference Ranges Handbook*. **1.8,** (2018).

11. Liu, J. Prognostic value of pretreatment albumin – globulin ratio in predicting long-term mortality in gastric cancer patients who underwent D2 resection. *Onco. Targets. Ther.* **10,** 2155–2162 (2017).

12. Koyama, T., Kuriyama, N., Ozaki, E., Matsui, D. *et al.* Serum albumin to globulin ratio is related to cognitive decline via reflection of homeostasis: A nested case-control study. *BMC Neurol.* **16,** 253–263 (2016).

13. Beamer, N., Bruce, M.S., Sexton, G., de Garmo, P. *et al.* Fibrinogen and the Albumin-Globulin Ratio in Recurrent Stroke. *Stroke* **24,** 1133–1139 (1993).

14. Peters, A.S., Backhaus, J,. Pfutzner, A., Raster, M. *et al.* Serum-infrared spectroscopy is suitable for diagnosis of atherosclerosis and its clinical manifestations. *Vib. Spectrosc.* **92,** 20–26 (2017).

15. NHS. Total Protein Monograph of the Association for Clinical Biochemistry and Laboratory Medicine. Available at: https://www.gloshospitals.nhs.uk/our-services/services-we-offer/pathology/tests-and-investigations/protein-total/.

16. Mayo-Clinic-Laboratories. Mayo Clinic Laboratories Immunoglobulins Reference Intervals. 1–5 (2020). Available at: https://www.mayocliniclabs.com/test-catalog/Clinical+and+Interpretive/8156. (Accessed: 25th February 2020)

17. ExeterUniversity & NHS. Exeter Laboratory Blood Science Test Immunoglobulins. Available at: https://www.exeterlaboratory.com/test/immunoglobulins-iga-igg-igm/. (Accessed: 25th February 2020)

18. Fallatah, H. I. & Akbar, H. O. Elevated serum immunoglobulin G levels in patients with chronic liver disease in comparison to patients with autoimmune hepatitis. *Libyan J. Med.* **5,** 4857 (2010).

19. Lin, S., Sun, Q., Mao, W. & Chen, Y. Serum immunoglobulin A (IgA) level is a potential biomarker indicating cirrhosis during chronic hepatitis B infection. *Gastroenterol. Res. Pract.* **2016,** (2016).

20. Rodriguezsegade, S., Camifa, F. C. & Maria, S. L. High serum IgA concentrations in patients with diabetes mellitus: agewise distribution and relation to chronic complications. *Clin. Chem.* **42,** 1064–1067 (1996).

21. Burg, M. van der, Weemaes, C. M. R. & Cunnigham-Rundles, C. Isotype Defects. in *Stiehm's Immune Deficiencies* (eds. Stiehm, E. R. & Sullivan, K.) (Academic Press).

22. Czurlok, D., Gleim, J., Lindner, J. & Vo, P. Vibrational Energy Relaxation of Thiocyanate Ions in

Liquid-to-Supercritical Light and Heavy Water. A Fermi's Golden Rule Analysis. *J. Phys. Chem. Lett.* **5,** 3373–3379 (2014).

23. Cameron, J. M., Butler, H. J., Palmer, D. S. & Baker, M. J. Biofluid spectroscopic disease diagnostics: A review on the processes and spectral impact of drying. *J. Biophotonics* **11,** e201700299 (2018).

24. Bonnier, F., Baker, M. J. & Byrne, H. J. Improved protocols for vibrational spectroscopic analysis of body fluids. *J. Biophotonics* **7,** 167–179 (2014).

25. 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.* **10,** 382–396 (2015).

26. Gajjar, K. *et al.* Fourier-transform infrared spectroscopy coupled with a classi fi cation machine for the analysis of blood plasma or serum: a novel diagnostic approach for ovarian cancer. *Analyst* **138,** 3917–3926 (2013).

27. Hands, J.R., Dorling, K.M., Abwl, P., Ahston, K.M. *et al.* Attenuated total reflection Fourier transform infrared (ATR-FTIR) spectral discrimination of brain tumour severity from serum samples. *J. Biophotonics* **7,** 189–199 (2014).

28. Baiz, C. R., Reppert, M. & Tokmakoff, A. Introduction to Protein 2D IR Spectroscopy. *Ultrafast Infrared Vib. Spectrosc.* 361–404 (2012). doi:10.1201/b13972

29. Greetham, G.M. Burgos, P., Cao, Q., Clark, I.P. *et al.* ULTRA: A unique instrument for time-Resolved spectroscopy. *Appl. Spectrosc.* **64,** 1311–1319 (2010).

30. Greetham, G. M., Matousek, P., Robinson, D. A., Parker, A. W. *et al*. ULTRA laser system: a new dual-output 10 kHz Ti:Sapphire amplifier with UV–IR generation for time-resolved spectroscopy. *Cent. Laser Facil. Annu. Rep.* 249–250 (2008).

31. Team, R. D. C. R: A Language and Environment for Statistical Computing. *R Found. Stat. Comput.* (2010).

32. Baker, M.J., Trevisan, J., Bassan, P., Bhargava, R. et al. Using Fourier transform IR spectroscopy to analyse biological materials. *Nature Protocols* **9**, 1771-1791 (2014)

33. Hume, S., Hithell, G., Greetha, G.M., Donaldson, P.M. *et al.* Measuring proteins in H_2O with 2D-IR spectroscopy. *Chem. Sci.* **10,** 6448–6456 (2019).

34. Chuntonov, L., Kumar, R. & Kuroda, D. G. Non-linear infrared spectroscopy of the water bending mode: direct experimental evidence of hydration shell reorganization. *Phys. Chem.*

Chem. Phys. 13172–13181 (2014).

35. Kraemer, D., Cowan, M.L., Paarman, A., Huse, E. *et al.* Temperature dependence of the two-dimensional infrared spectrum of liquid liquid H_2O. *Proc. Natl. Acad. Sci.* **105,** 437–442 (2008).

36. Carpenter, W. B., Fournier, J. A., Biswas, R. & Voth, G. A. Delocalization and stretch-bend mixing of the HOH bend in liquid water. *J. Chem. Phys.* **147,** 84503 (2017).

37. Huse, N., Ashihara, S., Nibbering, E. T. J. & Elsaesser, T. Ultrafast vibrational relaxation of O – H bending and librational excitations in liquid H_2O. *Chem. Phys. Lett.* **404,** 389–393 (2005).

38. Hunt, N. T. 2D-IR spectroscopy: ultrafast insights into biomolecule structure and function. *Chem. Soc. Rev.* **38,** 1837–1848 (2009).

39. Hamm, P. & Zanni, M. T. *Concepts and Methods of 2D Infrared Spectroscopy*. (Cambridge University Press, 2011).

40. Baiz, C. R., Peng, C. S., Reppert, M. E., Jones, K. C. *et al*. Coherent two-dimensional infrared spectroscopy: Quantitative analysis of protein secondary structure in solution. *Analyst* **137,** 1793–1799 (2012).

41. 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* **119,** 14065–14075 (2015).

42. Minnes, L., Shaw, D.J., Cossins, B.P., Donaldson, P.M. *et al.* Quantifying secondary structure changes in calmodulin using 2D-IR spectroscopy. *Anal. Chem.* **89,** 10898–10906 (2017).

43. Venyaminov, S. Y. & Prendergast, F. G. Water (H_2O and D_2O) molar absorptivity in the 1000 – 4000 cm-1 range and quantitative infrared spectroscopy of aqueous solutions 1. *Anal. Biochem.* **248,** 234–245 (1997).

44. Ashihara, S., Huse, N., Espagne, A., Nibbering, E. T. J. *et al*. Vibrational couplings and ultrafast relaxation of the O-H bending mode in liquid H2O. *Chem. Phys. Lett.* **424,** 66–70 (2006).

45. Barth, A. & Zscherp, C. What vibrations tell us about proteins. *Q. Rev. Biophys.* **35,** 369–430 (2002).

46. Middleton, C. T., Buchanan, L. E., Dunkelberger, E. B. & Zanni, M. T. Utilizing lifetimes to suppress random coil features in 2D IR spectra. *J. Phys. Chem. Lett.* **2,** 2357–2361 (2011).

47. Adamczyk, K., Simpson, N., Greetham, G.M., Gumiero, A. *et al.* Ultrafast infrared spectroscopy reveals water-mediated coherent dynamics in an enzyme active site. *Chem. Sci.* **6,** 505–516 (2015).

48. Guo, Q., Pagano, P., Li, Y.L., Kohen, A. & Cheatum, C. M. Line shape analysis of two-dimensional infrared spectra. *J. Chem. Phys.* **142,** 212427 (2015).

49. Harding J.R & J.W, K. Bromocresol green as a reagent for serum albumin J. R. *Proc. Assoc. Clin. Biochem.* **5,** 51–53 (1968).

50. Sammut, C. & Webb, G. . Leave-One-Out-Cross-Validation. in *Encyclopedia of Machine Learning* (Springer, 2011).

51. Zheng, K., Wu, L., He, Z., Yang, B. & Yang, Y. Measurement of the total protein in serum by biuret method with uncertainty evaluation. *Measurement* **112,** 16–21 (2017).

4. Using 2D-IR Spectroscopy for the Detection of Glycine in Serum

4.1. Abstract

Vibrational spectroscopic analysis of the protein content of blood serum is a powerful biomedical diagnostic tool, providing a broad molecular fingerprint of metabolic activity encompassing both the high molecular weight circulatory proteome and the low molecular weight (LMW, < 25 kDa) peptidome. Current applications of infrared spectroscopy to blood serum analysis are however complicated by the strong overlapping absorption of H_2O in the amide I band region, while bands due to low molecular weight serum components or species present at low concentrations are obscured by those from larger or more numerically dominant constituents.

Building on Chapter 3, where ultrafast 2D-IR spectroscopy is used to analyse the protein amide I band in liquid serum, the evaluation of the ability of 2D-IR to detect the lower molecular weight content of blood serum, such as amino acids and peptides is now studied. Studying samples of blood serum supplemented with the amino acid glycine shows that the information-rich off-diagonal region of the 2D-IR spectrum can be used to facilitate separation of the glycine component from the spectrally dominant high molecular weight protein fraction in liquid serum samples without sample filtration or drying. This approach is used to establish a 2D-IR detection limit for glycine of ~3 mg/mL and provides a platform for a critical assessment of the sensitivity and utility of 2D-IR for future molecular analysis of serum samples. The results show that, in the case of some clinically-relevant diagnostic molecules, the potential exists for 2D-IR to complement IR absorption methods because the benefits of the second frequency dimension offered by 2D-IR spectroscopy outweighs the added technical complexity of the measurement.

4.2. Introduction

Early detection of disease is essential for effective healthcare whether it is to improve the patient's response to treatment and chances of survival or because some diseases are asymptomatic until it is too late for medical intervention.[1] Blood serum is an important source of biomedical information, it is easily obtained and provides a molecular snapshot of metabolic processes by virtue of coming into contact with most of the major organs.[2–4] The chemical composition of serum therefore includes a number of diagnostic markers ranging from the circulatory proteome and low molecular weight peptidome to lipids, sugars and nucleic acids.[5–7]

Extracting the information content of serum is technically challenging due to the high dynamic concentration range of the biochemical constituents.[6,8] Typical blood serum protein concentrations in humans are 60-80 mg/mL with the serum albumins and globulins dominating with levels of 35-50 mg/mL and 20-30 mg/mL, respectively,[9–12] which makes it challenging to detect low molecular weight and less abundant molecules, some of which may be of use in disease detection and diagnosis.[13–15] The low molecular weight fraction of serum is typically defined as containing molecules with masses < 25 kDa. This includes cytokines (<6-30 kDa), peptide hormones (<30 kDa), lipids (<1 kDa) and protein fragments smaller than 25 kDa and access to the LMW fraction of serum is desirable from a diagnostic perspective.[16,17]

As disease states progress, changes in the LMW composition can occur. For example, the use of principal component analysis has shown to discriminate between patients with Alzheimer's disease and healthy controls caused by significant changes in amino acid levels (78 – 204 Da)[18] and amino acid changes have also been documented for patients with different neurodegenerative diseases.[19] Studies have shown that a decrease in the lipid desmosterol (385 Da) concentration shows a high correlation with patients diagnosed with Alzheimer's when compared with healthy controls and these changes were found to be more prominent in females.[20] 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.[21,22]

As discussed in Chapter 3, current analytical technologies employ antibody assays to measure protein levels in serum, however these techniques rely critically on the availability of specific antibodies for proteins of interest and require time for sample preparation as well as a significant sample volume, ~1 mL.[7,23,24] Furthermore, as diseases can be heterogeneous the broad chemical coverage acquired via vibrational spectroscopy is preferred to information derived from a single, or small panel of marker proteins.[25–27]

Using IR spectroscopy, analysis of the LMW protein fraction of serum is challenging due to the dominant signals from the albumins and globulins.[15,28,29] Serum filtration has been utilised to remove the larger species and enables an increase in sensitivity to LMW molecules using ATR-IR spectroscopy,[29,30] but at the expense of time and the possible unintentional removal of some LMW components. In particular, the role of albumin as a carrier protein means that its removal during the filtration process could also result in the concomitant removal of

bound LMW species.[8,31] Due to the high water content of serum, the use of IR spectroscopy to examine the amide I band of proteins is often done using dried serum droplets in order to minimise the overlapping spectral peak cause by the O-H bend of water, however this introduces physical and chemical inhomogeneity within the droplet resulting in repeatability difficulties.[32]

In Chapter 3, the 2D-IR pulse sequence was exploited and shown to supress the water signal allowing direct measurement of the protein content of serum without the overlapping water absorptions.[33] Recent advances in laser and pulse control technology have significantly decreased the time required to measure a 2D-IR spectrum to a few seconds per spectrum[34–36] and high throughput screening applications, using low volume samples (20 μL), have been demonstrated,[37] making this technique potentially viable for future clinical use. However, as 2D-IR is technically more complex than IR absorption spectroscopy, it is essential to obtain a thorough understanding of the range of molecular targets that 2D-IR might be used to measure within a complex serum sample. This requires an evaluation of the sensitivity limits of 2D-IR to a range of molecular species.

Using the simplest amino acid, glycine, as a model protein, measurements of serum spiked with a range of glycine concentrations are made and using current 2D-IR technology, the detection limit is ∼3 mg/mL in as received serum. Key to this measurement is the characteristic network of off-diagonal peaks in the glycine 2D-IR spectrum, which allows spectral separation of its amide I contribution. Physiologically, glycine is present in serum at much lower levels (∼20 μg/mL) than detected here and has a molecular weight of 75 Da, making it one of the smallest LMW models to supplement serum,[38] however this 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.[38] Ultimately, it is envisaged that 2D-IR could be used to complement IR absorption methods for specific biomedical problems.

4.3. Experimental Techniques

4.3.1 Sample Preparation

Pooled equine serum, glycine, and bovine serum albumin and γ-globulins were obtained from Sigma Aldrich and used without further purification. Individual glycine, albumin and γ-

globulin proteins were prepared in a Tris buffer at pH=7.5 to mimic the pH of the pooled serum samples. Serum samples were prepared by adding glycine to pooled equine serum at concentrations shown in Table 4.1, yielding a total of 9 samples (8 spikes and neat serum). For ease of reference, each serum sample will be referred to using the notation ^{x}[Gly], where x denotes the factor, irrespective of units, by which the glycine concentration exceeds that of typical serum (20 µg/mL). Serum samples without additional Gly added will thus be labelled ^{1}[Gly]. The samples studied are shown in Table 4.1.

Table 4.1: *Summary of Gly-spiked serum samples studied, showing [Gly] in both mg/mL and mol/dm^{-3}. Notation ^{x}[Gly] uses x to denote the relative ratio of Gly in spiked sample to that of physiological Gly serum levels (~20 µg/mL) irrespective of concentration units.*

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

4.3.2. IR Spectroscopy

For all IR and 2D-IR spectroscopy measurements, samples were housed between two CaF_2 windows in a standard sample cell. To avoid saturation of the O-H bending ($\delta_{H\text{-}O\text{-}H}$) mode of water at 1650 cm^{-1} the sample thickness was carefully controlled. The same process described in Chapter 3 was used to obtain a sample thickness of ~2.75 µm based upon the measured molar extinction coefficient of water.[33]

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^{-1} in the spectral region 400 - 4000 cm^{-1}.

2D-IR spectra were recorded using the ULTRA spectrometer using the Fourier transform 2D-IR technique,[39] as described in Chapter 2. A waiting time of 250 fs was employed throughout, which has been shown to optimise protein amide I signals at the expense of the water signal, as described in Chapter 3.[33]

4.4. Results and Discussions

All 2D-IR spectra were pre-processed and analysed using a custom made script on the statistical analysis software programme, R.[40] Prior to analysis a baseline subtraction was performed using a 2nd order polynomial function and all spectra were normalised to the albumin $v = 0 - 1$ diagonal peak at 1656 cm^{-1}.

4.4.1. Glycine Peak Assignment

The IR absorption spectrum of neat equine serum (^{1}[Gly]) and that of serum spiked with 100 mg/mL glycine (^{5000}[Gly]) in the wavenumber range 1300 – 1700 cm^{-1} are shown in Figure 4.1(a) alongside that of H_2O.

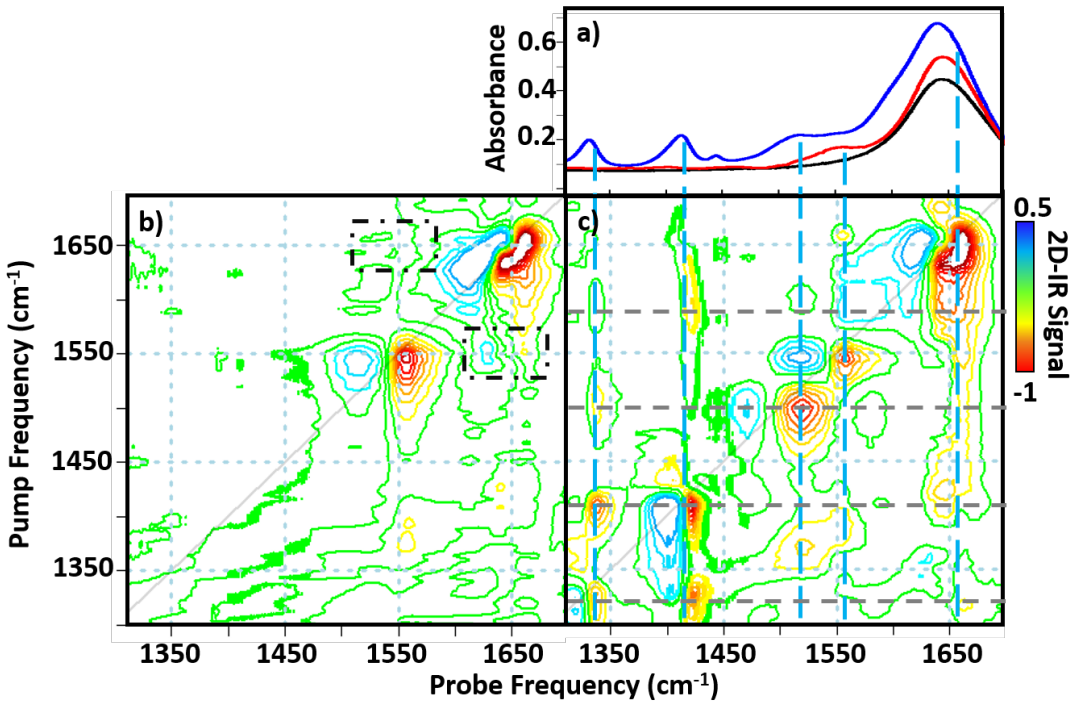

Figure 4.1: *a) Infrared absorption spectra of H_2O (black), serum (^{1}[Gly], red) and serum spiked with 100 mg/mL of glycine (^{5000}[Gly]). 2D-IR spectra of b) serum (^{1}[Gly]) and c) serum spiked with 100 mg/mL of glycine (^{5000}[Gly]). Blue dashed vertical lines in (a,c) compare peaks using both spectroscopy types. Horizontal grey dashed lines indicate coupling between on-diagonal glycine modes. Spectra are magnified by a factor of 2 to enhance off-diagonal couplings.*

In Figure 4.1(a) the H_2O spectrum (black) shows a band near 1650 cm^{-1} due to the $\delta_{H\text{-}O\text{-}H}$ mode of water. This band is also dominant in the spectrum of ^{1}[Gly], overlapping but not fully obscuring the amide I band of the protein (red). The protein amide II band is also present in the IR absorption spectrum of ^{1}[Gly] at 1550 cm^{-1}. The addition of glycine to serum (^{5000}[Gly], blue) leads to the presence of additional bands due to the amino acid. Bands at 1330, 1415 and 1445 cm^{-1} are assigned to CH_2 wagging, symmetric carboxylate stretching and CH_2 scissor modes respectively.[41] The band located at 1515 cm^{-1} is assigned to the in-plane N-H bending mode while there is a contribution at 1650 cm^{-1} from the carboxylate C=O stretching mode.

The 2D-IR spectra of ^{1}[Gly] and ^{5000}[Gly] are shown in Figure 4.1(b, c). In the as-received serum sample, (^{1}[Gly]) negative peaks (red) are visible on the spectrum diagonal at wavenumber coordinates (pump, probe) of (1650, 1650 cm^{-1}) and (1550, 1550 cm^{-1}). These are assigned to the $v = 0 - 1$ transitions of the amide I and II bands of the protein content of the serum respectively. Positive (blue) peaks are assigned to $v = 1 - 2$ transitions of the same bands

and are shifted to lower probe frequencies by the respective vibrational anharmonicities of the two modes. Weak off-diagonal peaks are observed (Figure 4.1(b), black dashed boxes) linking these two diagonal peaks, consistent with vibrational coupling and energy transfer processes between the amide I and II modes.[42] No contribution from water is observed in this spectrum, as reported previously.[33]

In the case of the 2D-IR spectrum of the ^{5000}[Gly] sample (Figure 4.1(c)), a number of diagonal and off-diagonal peaks are observed, in addition to those already assigned to the protein content of the serum. Diagonal peaks are present in the spectrum corresponding to the vibrational modes of glycine described in the IR absorption spectrum at (pump = probe) 1330, 1415 and 1515 cm^{-1}. These are identified in the figure by dashed vertical lines 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 mode pattern for glycine (see horizontal dashed lines in Figure 4.1(c)). The clear separation of the glycine spectral pattern from the amide I/II bands of the serum also allows identification of glycine in the complex serum and its quantification using 2D-IR spectroscopy.

4.4.2. Quantifying the Glycine Concentration in Serum

The impact on the 2D-IR spectrum of increasing the glycine concentration in serum is shown in Figure 4.2. It can be seen from the data that the size of the peaks identified with glycine from Figure 4.1(c) increase in magnitude as the glycine content is increased. This allows the detection threshold for glycine using this 2D-IR spectrometer to be determined. Three different analysis methods were applied to quantify the glycine levels in a given spectrum, examination of the 2D-IR spectrum diagonals, principal component analysis and linear combination analysis.

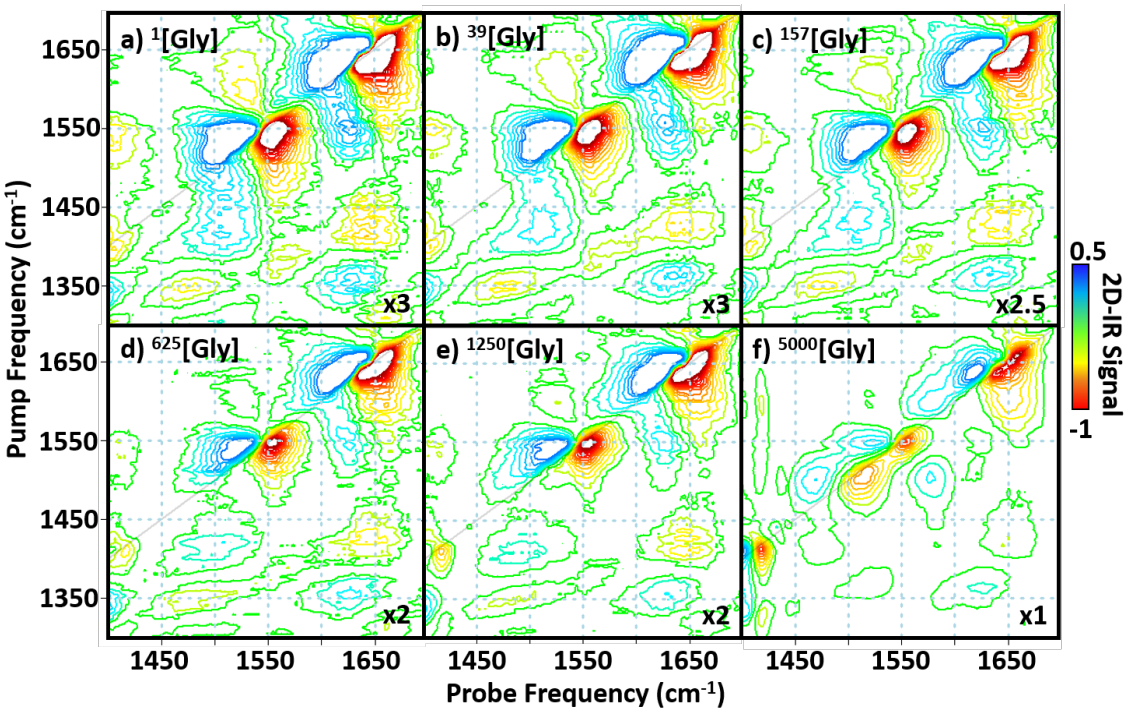

Figure 4.2: *2D-IR spectra of serum spiked with glycine after normalisation to the albumin signal at 1656 cm^{-1}. The figures show the average of 3 spectra for each glycine spikes at the following concentrations: a) ^{1}[Gly], b) ^{39}[Gly], c) ^{157}[Gly], d) ^{625}[Gly], e) ^{1250}[Gly], f) ^{5000}[Gly]. All spectra are magnified by a factor indicated in bottom right hand corner, to enhance glycine on and off diagonal peaks.*

4.4.2.1. The 2D-IR Spectrum Diagonal

The magnitude of the 2D-IR diagonal peak assigned to glycine at (pump = probe) 1515 cm^{-1} (Figure 4.3(a), grey dashed line) was plotted as a function of glycine concentration. In the ^{1}[Gly] sample, where no additional glycine was added to the serum, the positive contribution at 1515 cm^{-1} is due to the $v = 1 - 2$ transition of the amide II mode of serum proteins. As the glycine concentration increases, the $v = 0 - 1$ transition of the in plane N-H bending mode of glycine increased progressively in amplitude, eventually dominating the spectrum. This change in amplitude, plotted as a function of [Gly] produced a linear trend (Figure 4.3(b), R^2 = 0.994). Other diagonal peaks arising from the glycine content of the samples also yielded similar linear correlations however as the peak at 1515 cm^{-1} produced the largest change in amplitude from the ^{1}[Gly] sample, it provides the strongest correlation and was found to offer the best position along the 2D-IR diagonal for glycine detection.

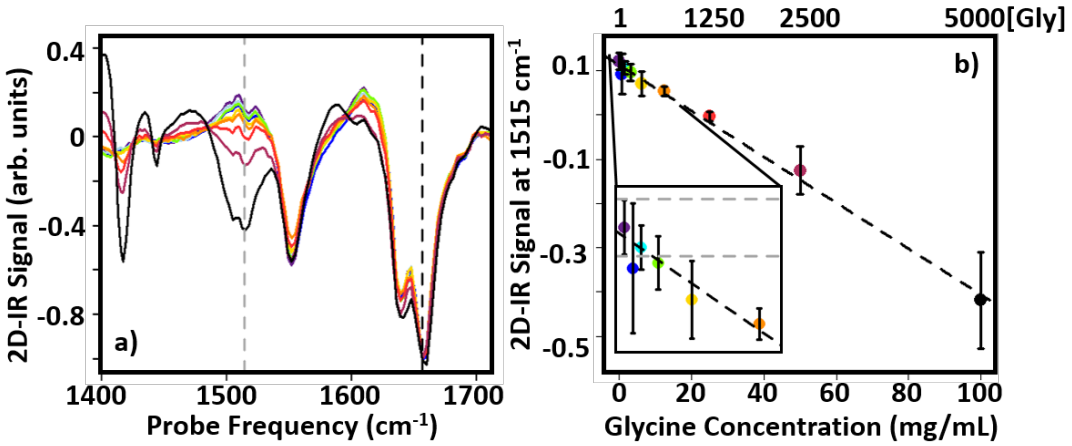

Figure 4.3: *a) 2D-IR Diagonal slices scaled to the amide I albumin signal (black dashed vertical line, 1656 cm^{-1}). b) The average amplitude at 1515 cm^{-1} (grey dashed vertical line in a)) as a function of glycine concentration. Error bars show 1σ variation from triplicate measurement. Best linear fit is shown as black dashed line with R^2=0.994.*

From this linear correlation, it is possible to estimate the glycine value at which the 2D-IR signature becomes detectable against the background of the instrumental noise levels. An area of the 2D-IR spectrum of the ^{1}[Gly] sample free from any peaks, as shown by the dashed black box in Figure 4.4(a, b) is used to calculate the noise level.

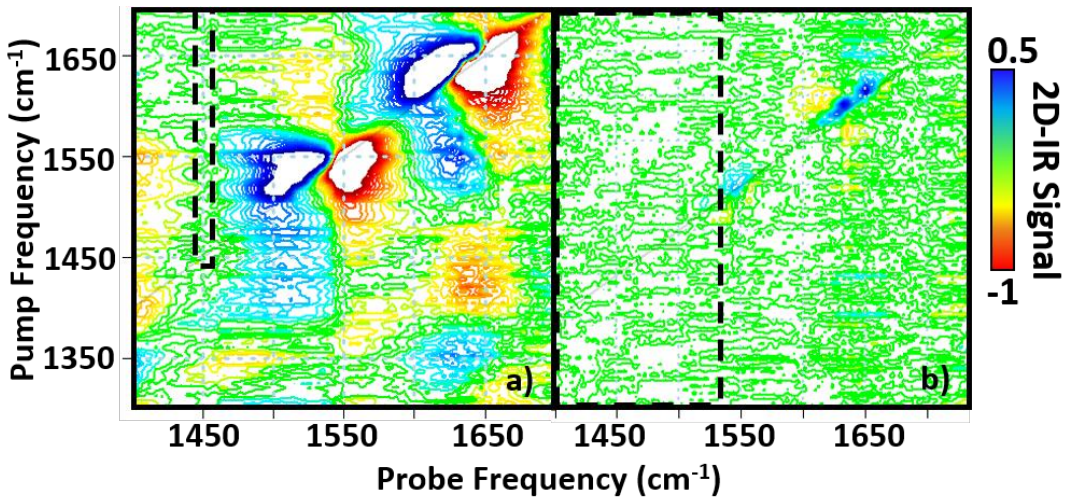

Figure 4.4: *a) 2D-IR spectrum of serum (^{1}[Gly]). b) The average 2D-IR serum difference spectra of the ^{1}[Gly] spectra. Both (a, b) are magnified by a factor of 5 for clarity. Area denoted by black dashed box used to calculate the RMS noise level.*

The average root mean square (RMS) noise level of three 2D-IR spectra of serum only was calculated to be 9.7 x 10^{-3}. However only 1.6% of data points from the whole spectrum were utilised to achieve this result (Figure 4.4(a), black dashed box). In order to compare a noise evaluation using greater number of data points, the average difference spectrum of the serum only spectra were obtained, of which the average is plotted in Figure 4.4(b). Using the area defined by the black dashed box in Figure 4.4(b), which utilises 40% of data points on the spectrum, produces an average root mean square (RMS) noise level of 9.6 x 10^{-3}, similar to that obtained for the ^{1}[Gly] sample. Making the conservative assumption that only peaks exceeding twice this noise floor will be visible, this leads to a detection limit for glycine of ^{157}[Gly] (~3 mg/mL).

4.4.2.2. Principal Component Analysis

Principal component analysis (PCA) is an unsupervised technique which extracts elements of variance within a dataset by reducing the dimensionality of the data. By geometrically projecting the dataset onto fewer dimensions, known as principal components (PCs), a new co-ordinate system is established where the largest variance is described by the first principal component (PC1) and weightings of each of these new PCs are calculated and used to reconstruct the original dataset.[43,44] Loadings and scores are the terms given to the spectral representation and weightings of each of the PCs. Mean centring is commonly used to pre-process the data prior to PCA however for the results in this thesis this technique is not used here.

PCA was applied to the data in Figure 4.2 and the results showed that the variance could be attributed mainly to the first two components, PC1 and PC2 (Figure 4.5), which account for 87.6% and 7.4% of the variance, respectively. In the case of PC1, the relative loading was found to be invariant with glycine concentration as shown in Figure 4.6.

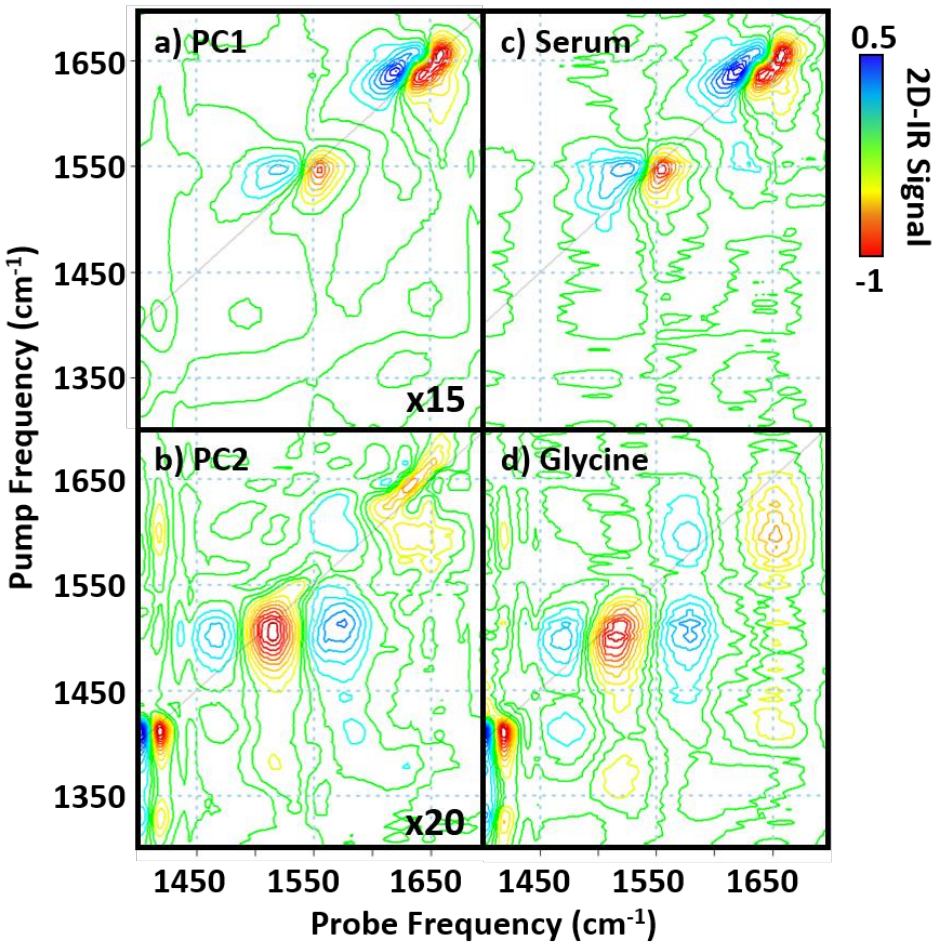

Figure 4.5: *Comparison of separate serum and glycine spectra with PCA results. a) PC1 loading plot, b) PC2 loading plot, c) Serum spectra, d) Glycine (100 mg/mL) spectra in H_2O. All spectra have been plotted on the colour scale shown.*

The spectral information associated with this principal component (Figure 4.5(a)) was in excellent agreement with the spectrum of serum proteins ([1][Gly]), Figure 4.5(c)), while the spectrum of PC2 clearly identified the glycine diagonal and off-diagonal peak network discussed earlier (Figure 4.5(b,d)).

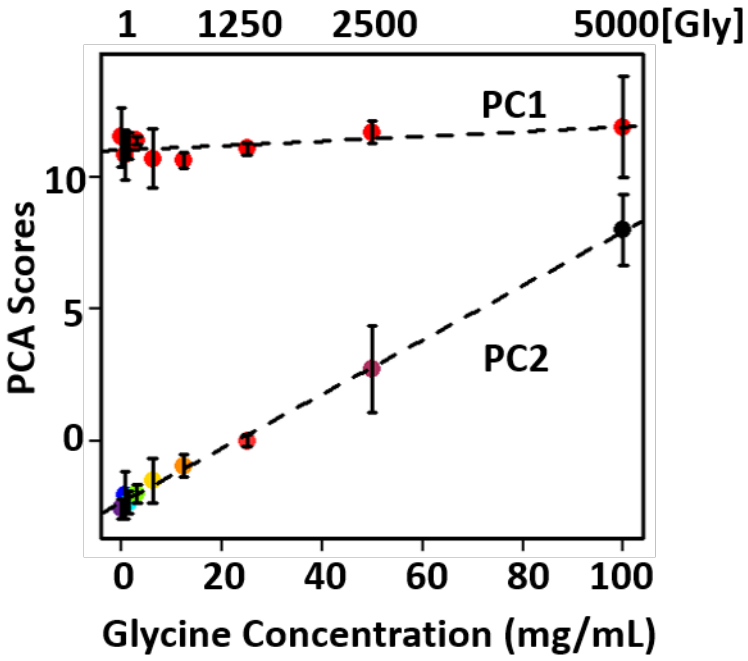

Figure 4.6 *Average PC1 (red) and PC2 (multi-coloured) scores for the glycine spiked serum. Error bars show 2σ variation from measurement in triplicate. Best linear fit is shown as black dashed line, PC2 R^2=0.997.*

For the second principal component, the average PC2 score values showed a linear correlation with glycine concentration (Figure 4.6). The straight line fitted for PC2 is done so by using the least squares linear regression technique. This calculates the best fitting line for the observed data points by minimising the sum of squares of vertical deviations between each point to the fitted line.[45] This produces an R^2=0.997 for PC2 as a function of glycine concentration.

Cross validation is performed in order to evaluate the linear regression for PC2 and is done by removing each individual data point and the linear regression is re-run on the remaining part of the data. The data that was removed can then be used to test the performance of the new model. The leave one out cross validation technique is utilised here, where each individual data point is left out in turn and the remaining data is used to create the model. The new model can then be used to predict the output value of the removed data point.

Here, the number of data points used to create the linear regression in Figure 4.6 for PC2 is 9 and so the cross validation is completed 9 separate times. All concentrations of glycine larger than ^{157}[Gly] (3.125 mg/mL) were predicted within an accuracy of the standard

deviation due to triplicate measurement however the smaller concentrations were not predicted within this tolerance. Using this method, the ^{157}[Gly] (3.125 mg/mL) sample is predicted to be 3.68 $\pm$ 1.82 mg/mL. From the noise calculation obtained above (Figure 4.4), concentrations less than ^{157}[Gly] (3.125 mg/mL) were found to be smaller than the noise level and thus not confidently detected and so the PC results are comparable with the detection limit previously established.

Comparisons with the PCA results of the IR absorption spectra (Figure 4.7) show that while the PC2 yields a strong linear correlation, (R^2=0.996), the loading plots (Figure 4.7(b)) of PC1 (99.25% variance) and PC2 (0.75% variance) do not differentiate into serum and glycine spectra as in the 2D-IR result.

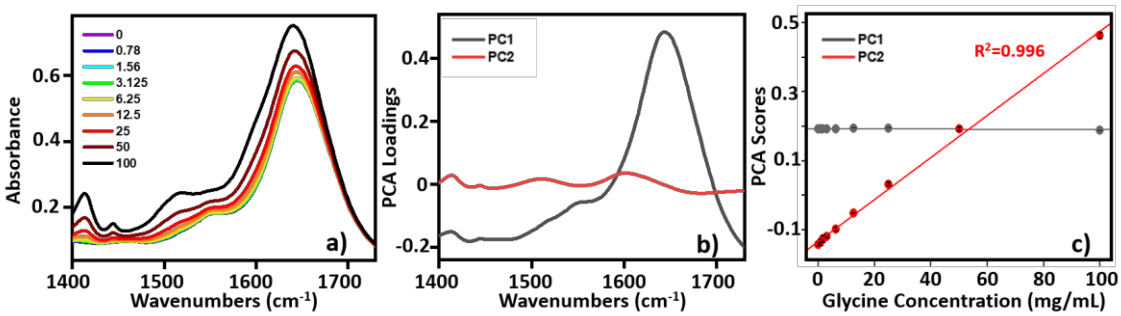

Figure 4.7: *Results of PCA of serum and glycine dataset. a) Infrared absorption spectra of serum and glycine, concentrations shown are in mg/mL. PCA loadings and scores plots for the first two principal components are shown in b) and c). PC1 is shown in black and PC2 in red. Error bars show 2σ variation from measurement in triplicate. Best linear fit is shown as solid line with and PC2 R^2=0.996.*

It is imperative to compare the results obtained from 2D-IR with those obtained using linear IR absorption methods. For example, the evaluation of detection limits using ATR-FTIR spectroscopy have been reported.[30] Studies of glycine detection using both wet and dry serum samples were carried out and compared with serum samples following depletion of the HMW fraction using ultra-filtration techniques. This ATR-FTIR study has shown analysis of wet serum allowing linear regression models to be built, permitting prediction of a 0.5 mg/mL to within an accuracy of 10% (0.45 $\pm$ 0.16 mg/mL). Removal of the HMW fraction through ultrafiltration allowed quantification of glycine levels down to 0.01 mg/mL in dried serum samples, a limit 50 times lower than that achieved prior to filtration. While this

detection limit is lower than in 2D-IR, the increased sample preparation steps need also be considered and discrimination of different protein components to the amide I band is not possible using linear methodologies.

4.4.2.3. Linear Combination Analysis

Linear combination analysis (LCA) is the third analysis method used which employs a multiple linear regression model combining the 2D-IR spectra of serum and glycine, where the linear model produced coefficients for each of these component spectra needed to recreate the original ^{x}[Gly] dataset. The model was computed three times using different input component spectra and different spectral regions to rebuild the original dataset.

i) The first iteration of LCA used two component spectra, one of serum and one of glycine to rebuild the original ^{x}[Gly] dataset. The spectral regions used were the same as those utilised for the 2D-IR diagonal and PC analysis and allowed inclusion of the lower frequency peak at 1415 cm^{-1}, as shown in Figure 4.8(a, b).

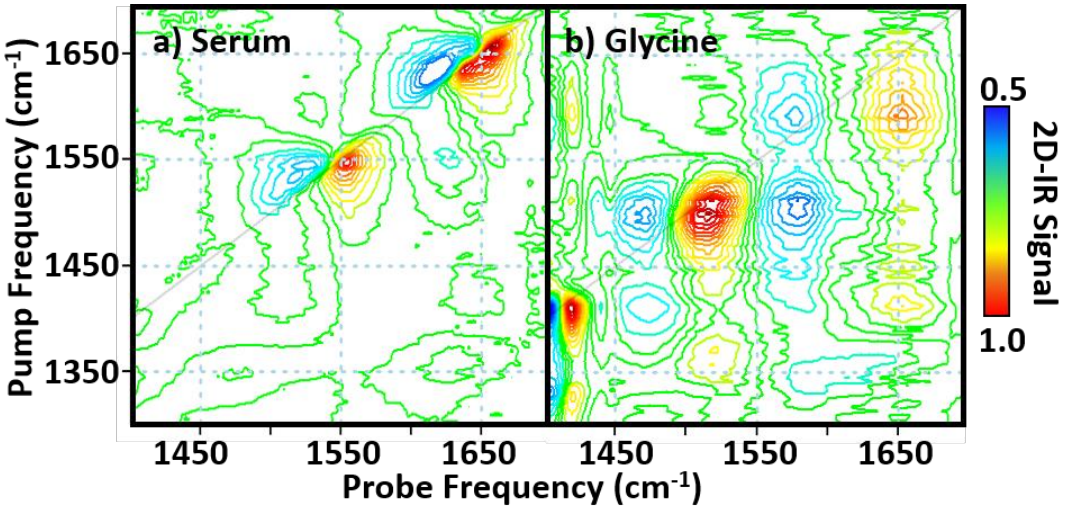

Figure 4.8: *2D-IR spectra of the LCA inputs, (a) serum (b) glycine in H_2O buffer. The spectral range uses the full spectrum acquired on one detector, with a probe frequency range of 1400 - 1700 cm^{-1}.*

The results of the LCA are shown in Figure 4.9, and the best linear fits were again constructed using the least squares linear regression method. The serum coefficient is found to be relatively constant throughout the concentrations studied (Figure 4.9(a), blue) however a slight decline is noted with increasing concentrations of the amino acid, this could be the

result of some spectral overlap of the glycine and serum peaks. The glycine coefficient shows a linear increase with increasing concentration (Figure 4.9(a- red, b) with an R^2 value of 0.998.

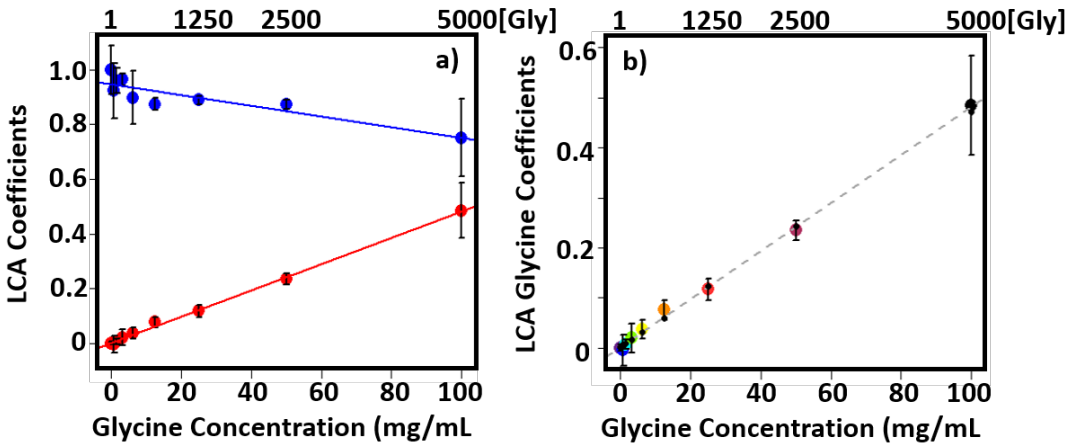

Figure 4.9: *a) Coefficients produced by the LCA for serum (blue) and glycine (red). (b) The glycine trend only is shown and grey dashed line indicates the best linear fit with R^2=0.998. Black circles are results of cross validation tests. Error bars show 2σ variation due to triplicate measurement.*

Leave one out cross validation was performed to assess the predictability, the results of which are shown as black dots in Figure 4.9(b) for each sample. This result obtained a predicted value for the ^{157}[Gly] (3.125 mg/mL) sample of 3.91 $\pm$ 3.18 mg/mL. The associated error due to the triplicate measurement is large and significantly greater than the PC result, so while this method does yield an acceptable prediction the uncertainty is too high to produce a meaningful result.

ii)The second iteration of LCA again used only two component spectra, one of serum and one of glycine to rebuild the original ^{x}[Gly] dataset as before however the spectral area used has been reduced in order to focus on the amide I region of the spectrum as shown in Figure 4.10(a, b).

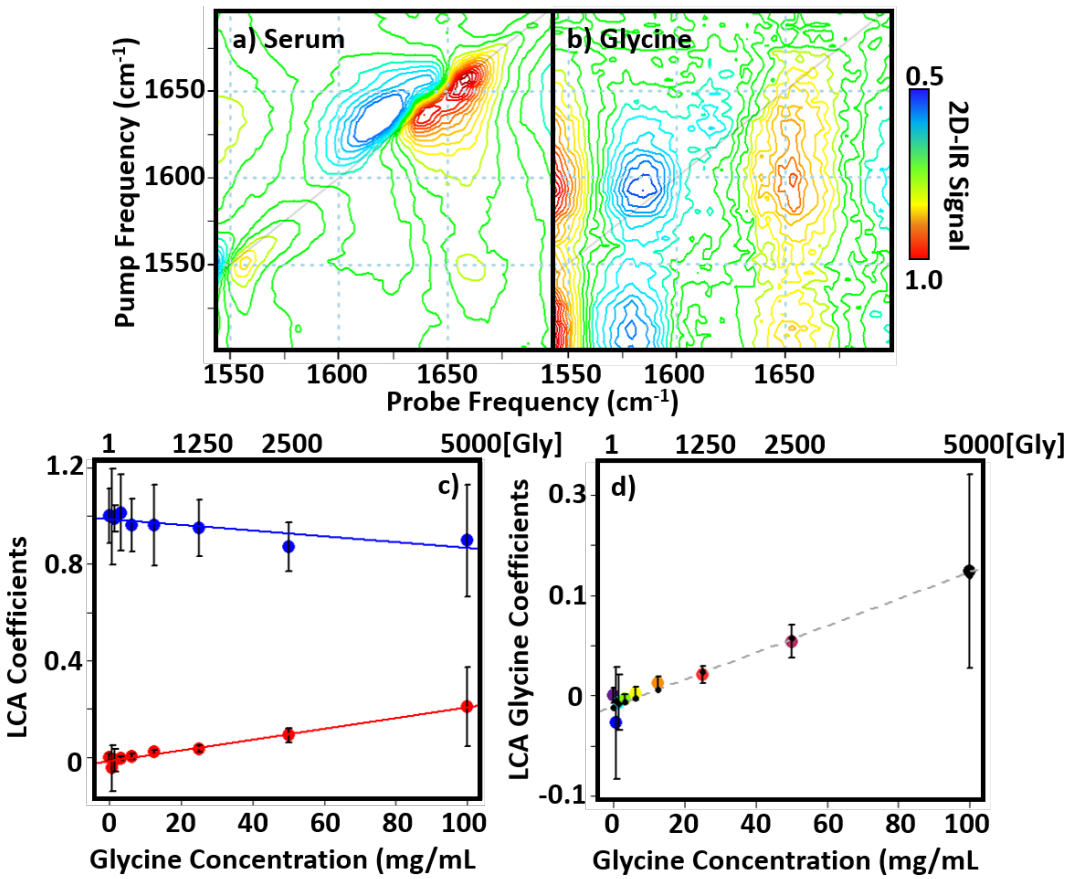

Figure 4.10: *2D-IR spectra of the LCA inputs, (a) serum (b) glycine in H_2O buffer. The spectral range of (a, b) has been shortened to focus on the amide I region, with a probe frequency range between 1535 - 1730 cm^{-1}. c) Coefficients produced by the LCA for serum (blue) and glycine (red). (d) The glycine trend only is shown and grey dashed line indicates the best linear fit with R^2=0.964. Black circles are results of cross validation tests. Error bars show 2σ variation due to triplicate measurement.*

Again, the LCA produced a serum coefficient that is relatively constant throughout the concentrations studied and the glycine coefficient shows a linear increase with increasing concentration with an R^2 value of 0.964 (Figure 4.10(d)). Leave one out cross validation tests (Figure 4.10(d), black dots) predicted a concentration of 4.99 $\pm$ 1.94 mg/mL for the ^{157}[Gly] (3.125 mg/mL) sample.

It is noted that by using the larger spectral region in Figures 4.8 that includes the 1415 cm^{-1} peak, the correlation of the glycine coefficient is strengthened to give an R^2 value of 0.998 (Figure 4.9(b)), compared to R^2=0.964 for the amide I region (Figure 4.10(d)). A more accurate

prediction of ^{157}[Gly] is obtained using the larger spectral region, suggesting that the additional glycine peaks and off-diagonal peak network is required in order to obtain a more accurate result.

iii) Focussing on the ability of 2D-IR spectroscopy to differentiate contributions of glycine within the amide I region of the spectrum in the presence of contributions from the larger, more dominant proteins, and in the third case, a multiple linear regression model combining the 2D-IR spectra of serum albumins, γ-globulins and glycine were applied to recreate the measured ^{x}[Gly] spectra. By utilising serum albumin and globulin spectra as opposed to the serum spectrum as before the method is tested in its ability to recreate the dataset using a small library of spectra. The input component 2D-IR spectra were measured independently in H_2O buffer solution and are shown in Figure 4.11.

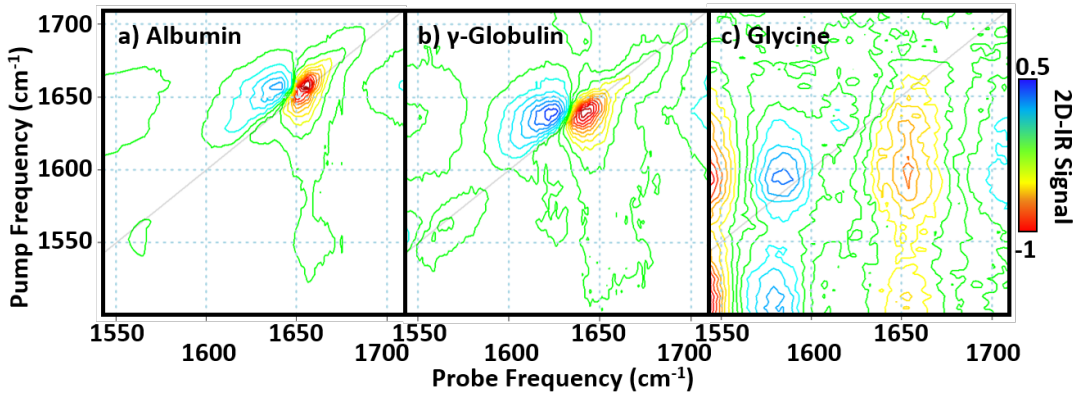

Figure 4.11: *LCA was performed using three inputs: a) BSA in H_2O. b) γ-Globulin in H_2O and c) Glycine in H_2O. Samples a) and b) were scaled to their amide I peak and c) was scaled to the signal at 1600, 1650 cm^{-1} along the pump and probe respectively. All spectra are plotted on the same scale, see colour bar.*

The result of the LCA show that the coefficient of the γ-globulins (Figure 4.12(a), blue) required to recreate each spectrum in the dataset is relatively steady, the BSA coefficient also shows a relatively constant coefficient however a slight decrease is noted with increasing concentrations of glycine. Again, this could be a result of some spectral overlap as the spectrum of the BSA amide I band (Figure 4.11(a)) and that of glycine (Figure 4.11(c)) overlay one another. As the concentration of both the albumins and globulins in serum are unchanged throughout this experiment, the coefficients of the LCA for these two components is anticipated to remain constant throughout the range of glycine

concentrations studied. The glycine component is shown to increase with increasing concentration of the amino acid (R^2=0.942, Figure 4.12(a), red).

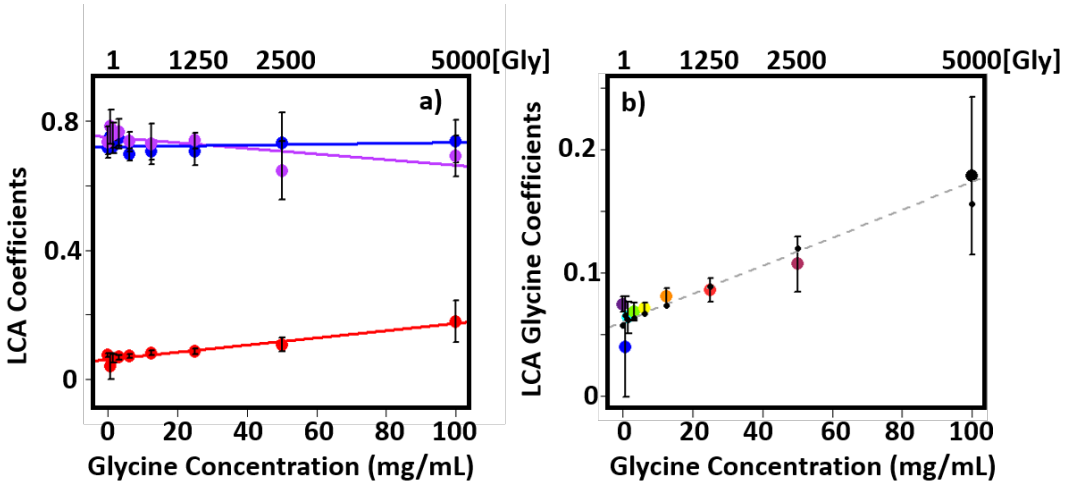

Figure 4.12: *Result of LCA, a) both BSA (purple) and y-Globulin (blue) show almost constant contributions and glycine (red) increases with increasing glycine concentration. b) Glycine coefficient only as a result of LCA, showing increasing glycine contribution with increasing concentration. Black circles are results of cross validation tests. Grey dashed line shows linear fit with R^2=0.942. Error bars show 1σ variation.*

As before, leave one out cross validation was performed, leaving each data point out of the multiple linear regression model to allow prediction of each concentrations of glycine (Figure 4.12(b, black dots)). The ^{157}[Gly] (3.125 mg/mL) sample was predicted to be 7.54 ± 6.81 mg/mL and is the poorest result obtained.

The residuals produced from the LCA yield variations between 5-35% of the normalised scale (residual peaks varying from 0.05 - 0.35) which indicates that there is indeed an error with the LCA fit. While the results obtained do in fact show a linear correlation with glycine concentration using each of the variations of LCA used, the spread of results is too large to be used conclusively.

The accuracy of this result would hope to be improved from utilising a larger spectral region of the three input component spectra used, including the network of glycine off-diagonal peaks as discussed for the previous two LCA utilising serum and glycine over different spectral regions. The use of pre-processing techniques, including noise reduction and smoothing methods, may also see an improvement in the linear regressions observed.

4.4.3. Assessment of Future 2D-IR Detection Studies in Serum

Analysis of the noise in the 2D-IR spectra allows glycine concentrations as low as 3.125 mg/mL (41 mmol/dm^3) to be detected. It is noted however that glycine, while being an ideal candidate for a model system, is a very small LMW species within the LMWF (<25 kDa) with a MW of 75 Da. If glycine is considered as a model peptide consisting of a single residue it can be inferred that a peptide consisting of 200 residues will show changes on a 2D-IR spectrum down to 0.2 mmol/dm^3 (1/200th of 41 mmol/dm^3). As glycine exists as a single molecule, amide I bonding between amino acids and transition dipole coupling throughout the secondary structure of the protein also needs to be considered as this will add to the strength of the signals produced for larger proteins as described in Chapter 1 and so 0.2 mmol/dm^3 is defined as an upper limit of detection.

In order to determine the extent of amide bonding and TDC on the signals produced from proteins rather than single amino acids, evaluation of the signal amplitude of 0.45 mmol/dm^3 BSA with the RMS of the noise would allow the detection down to 0.015 mmol/dm^3 (1 mg/mL). From this, it is suggested that proteins occurring at a minimum of ~1-3 mg/mL, e.g. α1-Antitrypison, an inflammatory response protein which occurs around 1.5-3.5 mg/mL[46,47] could be successfully detected using 2D-IR spectroscopy.

Although samples are normalised to help minimise variations from the sample path length, the issue of sample repeatability when using 2D-IR spectroscopy is clear from the large errors calculated from measurements in triplicate in particular from the final LCA method. However this method proved successful in Chapter 3 when evaluating the albumin to globulin ratio (AGR) within a serum sample, and so the AGR was calculated for this glycine dataset in order to assess whether the coefficients produced for the glycine dataset can also determine the AGR. Across the range of concentrations studied the average AGR was calculated to be 0.56 $\pm$ 0.05. This contrasts with the results obtained in Chapter 3 where using the LCA method, the serum only spectrum produced an AGR of around 0.8.

However this is a different dataset from the AGR study and the two datasets were not acquired at the same time. In order to determine whether there are problems within the dataset or the analysis method, the AGR of the samples was also calculated using the 2D-IR diagonal method (as described in Chapter 3) and an average AGR value of 0.75 was obtained across the concentrations studied; for a serum only sample a value of 0.73 is anticipated. It

is anticipated that if the two experimental datasets were of the same quality that the LCA should be able to determine the AGR as in Chapter 3. As the ratio is accurate when applying the 2D-IR diagonal method which does not account for changes in lineshapes and off-diagonal couplings it can be concluded that the two experiments do not produce spectra with a similar quality.

It is also important to note that by normalising the spectra to the albumin peak, where there is overlap with the glycine signal there is potential to cover the changes associated with the addition of glycine to the serum and so an ideal experiment would not involve normalising using any peaks of interest. The issue of sample-to-sample repeatability is addressed in Chapter 5.

4.5. Conclusions

The need for a technique that allows identification and quantification of multiple components in biofluids, without altering the sample, is essential for a fast and accurate disease diagnosis. The potential of 2D-IR spectroscopy to identify low molecular weight biomarkers in blood serum without the need for time consuming sample preparation such as filtration of the LMWF is shown. The on- and off-diagonal structure observed in 2D-IR spectroscopy allows identification of low molecular weight species, such as glycine, present at low concentrations in serum using 3 different analysis approaches. Establishing a detection limit of the amino acid glycine of ~3 mg/mL allows assessment of the sensitivity of 2D-IR for larger proteins for future analysis.

While the LCA produces the largest error it has been shown to provide additional information allowing the dataset to be rebuilt using a small library of spectra which cannot be done using traditional IR spectroscopies. This provides potential scope for a library of 2D-IR spectra of a large number of serum constituents to be used to identify serum components. However the errors associated with this analysis method were significant and results pertaining to the AGR ratio showed a lack of consistency between the dataset discussed and the previous chapter. This highlights the error in experiment-to-experiment repeatability of the 2D-IR technique. The data sets thus far have been normalised to the dominant albumin peak allowing variations across the dataset to be examined however in doing so the glycine signal in the amide I region is in effect overlooked due to their overlap and experimental fluctuations are not accounted for when using this normalisation step. Ideally, an internal normalisation

technique that reduces the experimental variations should be employed, a method for this has been developed and will be discussed in the following chapter.

This study highlights that 2D-IR spectroscopy offers a new approach to identifying low molecular weight proteins and potential biomarkers and the calculated detection limit for larger proteins means that important biomarkers of disease could potentially be detected. Establishing a detection limit of the amino acid glycine of allows assessment of the sensitivity of 2D-IR for larger proteins for future analysis.

4.6. References

1. Butler, H. J., Brennan, P. M., Cameron, J. M., Finlayson, D. *et al.* Development of high-throughput ATR-FTIR technology for rapid triage of brain cancer. *Nat. Commun.* **10,** 4501 (2019).

2. Sohn, E. Frontiers in blood testing. *Nature* **549,** S16–S18 (2017).

3. Quig, K., Wheatley, E. G. & Hara, M. O. Perspectives On Blood-Based Point-Of-Care Diagnostics. *Open Access Emerg. Med.* **11,** 291–296 (2019).

4. Song, Y., Huang, Y., Liu, X., Zhang, X. *et al.* Point-of-care technologies for molecular diagnostics using a drop of blood. *Trends Biotechnol.* **32,** 132–139 (2014).

5. Baker, M. J., Sockalingum, G. D., Hughes, C. & Lukaszewski, R. A. Developing and understanding biofluid vibrational spectroscopy: A critical review. *Chem. Soc. Rev.* **45,** 1803–1818 (2016).

6. 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* **6,** 961–967 (2006).

7. Hu, S., Loo, J. & Wong, D. Human body fluid proteome analysis. *Proteomics* **6,** 6326–6353 (2006).

8. Adkins, J. N., Varnum, S. M., Auberry, K. J., Moore, R. J. *et al.* Toward a human blood serum proteome. *Mol. Cell. Proteomics* **1,** 947–955 (2002).

9. Gutman, A. B., Moore, D. H., Gutman, E. B., McClellan, V. *et al*. Fractionation of serum proteins in hyperproteinemia with special reference to multiple myeloma. *J. Clin. Invest.* **20,** 765–783 (1941).

10. Marrack, J. R. & Hoch, H. Serum proteins: A review. *J. Clin. Pathol.* **2,** 161–192 (1949).

11. Busher, J. T. Serum albumin and globulins. in *Clinical Methods: The History, Physical, and Laboratory Examinations.* (eds. Hurst, J. ., Walker, H. . & Hall, W. .) 497–500 (Butterworths, 1990).

12. Eaton, P. & Fuggle, S. *Clinical Biochemistry Reference Ranges Handbook*. **1.8,** (2018).

13. Harper, R. G., Workman, S. R., Schuetzner, S., Timperman, A. T. *et al*. Low-molecular-weight human serum proteome using ultrafiltration, isoelectric focusing, and mass spectrometry. *Electrophoresis* **25,** 1299–1306 (2004).

14. Xu, X. & Veenstra, T. D. Analysis of biofluids for biomarker research. *Proteomics Clin. Appl.* **2,** 1403–1412 (2008).

15. Parachalil, D. R., Bruno, C., Bonnier, F., Blasco, H. *et al.* Analysis of bodily fluids using vibrational spectroscopy: A direct comparison of Raman scattering and infrared absorption techniques for the case of glucose in blood serum. *Analyst* **144,** 3334–3346 (2019).

16. Tirumalai, R. S., Chan, K. C., Prieto, D. A., Isaaq, H. J. *et al.* Characterization of the low molecular weight human serum proteome. *Mol. Cell. Proteomics* **2,** 1096–1103 (2003).

17. Chertov, O., Simpson, J.T., Biragyn, A., Conrads, T.P. *et al.* Enrichment of low-molecular- weight proteins from biofluids for biomarker discovery. *Expert Rev. Proteomics* **2,** 139–146 (2005).

18. Li, N., Liu, W., Li, W., Li, S. *et al.* Plasma metabolic profiling of Alzheimer's disease by liquid chromatography/mass spectrometry. *Clin. Biochem.* **43,** 992–997 (2010).

19. Socha, E., Koba, M. & Kośliński, P. Amino acid profiling as a method of discovering biomarkers for diagnosis of neurodegenerative diseases. *Amino Acids* **51,** 367–371 (2019).

20. Sato, Y., Suzuki, I., Nakamura, T., Bernier, F. *et al.* Identification of a new plasma biomarker of Alzheimer's disease using metabolomics technology. *J. Lipid Res.* **53,** 567–576 (2012).

21. Rascher, W., Kremens, B., Wagner, S., Feth, F. *et al.* Serial measurements of neuropeptide Y in plasma for monitoring neuroblastoma in children. *J. Pediatr.* **122,** 914–916 (1993).

22. Galli, S., Naranjo, A., Ryn, C. V., Tilan, J. U. *et al.* Neuropeptide Y as a biomarker and therapeutic target for neuroblastoma. *Am. J. Pathol.* **186,** 3040–3053 (2016).

23. Harding J.R & J.W, K. Bromocresol green as a reagent for serum albumin J. R. *Proc. Assoc. Clin. Biochem.* **5,** 51–53 (1968).

24. NHS. Total Protein Monograph of the Association for Clinical Biochemistry and Laboratory Medicine. Available at: https://www.gloshospitals.nhs.uk/our-services/services-we-offer/pathology/tests-and-investigations/protein-total/.

25. Zanin, M., Tuñas, J. M. & Menasalvas, E. Understanding diseases as increased heterogeneity: A complex network computational framework. *J. R. Soc. Interface* **15,** (2018).

26. Spalding, K., Board, R., Dawson, T., Jenkinson, M. D. *et al*. A review of novel analytical diagnostics for liquid biopsies: spectroscopic and spectrometric serum profiling of primary and secondary brain tumors. *Brain Behav.* **6,** 1–8 (2016).

27. Popescu, I. D., Codrici, E., Albulesco, L., Mihai, S. *et al.* Potential serum biomarkers for glioblastoma diagnostic assessed by proteomic approaches. *Proteome Sci.* **12,** 1–14 (2014).

28. Bonnier, F., Baker, M. J. & Byrne, H. J. Vibrational spectroscopic analysis of body fluids: Avoiding molecular contamination using centrifugal filtration. *Anal. Methods* **6,** 5155–5160 (2014).

29. Bonnier, F., Blasco, H., Wasselet, C., Brachet, G. *et al.* Ultra-filtration of human serum for improved quantitative analysis of low molecular weight biomarkers using ATR-IR spectroscopy. *Analyst* **142,** 1285–1298 (2017).

30. Bonnier, F., Brachet, G., Duong, R., Sojinrin, T. *et al.* Screening the low molecular weight fraction of human serum using ATR-IR spectroscopy. *J. Biophotonics* **9,** 1085–1097 (2016).

31. Tirumalai, R. S., Chan, K. C., Prieto, D. A., Issaq, H. J. *et al.* Characterization of the low molecular weight human serum proteome. *Mol. Cell. Proteomics* **2,** 1096–1103 (2003).

32. Cameron, J. M., Butler, H. J., Palmer, D. S. & Baker, M. J. Biofluid spectroscopic disease diagnostics : A review on the processes and spectral impact of drying. *J. Biophotonics* **11,** 1–12 (2018).

33. Hume, S. Hithell, G., Greetham, G. M., Donaldson, P. M. *et al.* Measuring proteins in H_2O with 2D-IR spectroscopy. *Chem. Sci.* **10,** 6448–6456 (2019).

34. Donaldson, P. M., Greetham, G. M., Shaw, D. J., Parker, A. W. *et al*. A 100 kHz pulse shaping 2D-IR spectrometer based on dual Yb:KGW amplifiers. *J. Phys. Chem. A* **122,** 780–787 (2018).

35. 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.* **104,** 14197–14202 (2007).

36. 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* **24,** 4117 (2016).

37. Fritzsch, R., Donaldson, P. M., Greetham, G. M., Towrie, M. *et al.* Rapid screening of DNA-ligand complexes via 2D-IR spectroscopy and ANOVA-PCA. *Anal. Chem.* **90,** 2732–2740 (2018).

38. 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.* **27,** 655–664 (1948).

39. Greetham, G. M., Burgos, P., Cao, Q., Clark, I. P. *et al.* ULTRA: A Unique Instrument for Time-Resolved Spectroscopy. *Appl. Spectrosc.* **64,** 1311–1319 (2010).

40. Team, R. D. C. R: A Language and Environment for Statistical Computing. *R Found. Stat. Comput.* (2010).

41. Lutz, O. M. D., Messner, C. B., Hofer, T. S., Canaval, L. R. *et al.* Computational vibrational spectroscopy of glycine in aqueous solution – Fundamental considerations towards feasible methodologies. *Chem. Phys.* **435,** 21–28 (2014).

42. 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.* **131,** 3385–3391 (2009).

43. Bro, R. & Smilde, A. K. Principal component analysis. *Anal. Methods* **6,** 2812–2831 (2014).

44. Lever, J., Krzywinski, M. & Altman, N. Points of significance: Principal component analysis. *Nat. Methods* **14,** 641–642 (2017).

45. Rencher, A. C. & Schaalje, G. B. *Linear Models in Statistics*. (John Wiley & Sons, 2008).

46. Fregonese, L. & Stolk, J. Hereditary alpha-1-antitrypsin deficiency and its clinical consequences. *Orphanet J. Rare Dis.* **3,** 1–9 (2008).

47. Pérez-Holanda, S., Blanco, I., Menéndez, M. & Rodrigo, L. Serum concentration of alpha-1 antitrypsin is significantly higher in colorectal cancer patients than in healthy controls. *BMC Cancer* **14,** 1–9 (2014).

5. Using the Thermal Response of Water as an Internal Standard for 2D-IR Spectroscopy of Proteins

This chapter contains results published in the following publications:

Hume S, G, Greetham G.M, Donaldson P.M, Towrie M, Parker A.W, Baker M.J, Hunt N.T., *2D-IR Spectroscopy of Proteins in Water: Using the Solvent Thermal Response as an Internal Standard*. Analytical Chemistry. 10(26):6448-6456 (2020)

All 2D-IR serum data acquisition and analysis was performed by myself.

5.1. Abstract

In Chapter 3, it was demonstrated that 2D-IR spectroscopy of the protein amide I band can be performed in aqueous (H_2O) rather than deuterated (D_2O) solvents. Ultrafast 2D-IR spectra can now be obtained in a matter of seconds which opens up the possibility of high throughput screening applications which are relevant to the biomedical and pharmaceutical sectors. Determining quantitative information from 2D-IR spectra recorded on different samples and different instruments is however made difficult by variations in beam alignment, laser intensity and sample conditions. This chapter reports a method that uses the magnitude of the associated thermal response of H_2O as an internal normalisation standard for 2D-IR spectra. Using the water response, which is temporally separated from the protein signal, to normalise the spectra allows significant reduction of the impact of measurement-to-measurement fluctuations on the data. This normalisation method enables creation of calibration curves for measurement of absolute protein concentrations and facilitates reproducible difference spectroscopy methodologies. These advances make significant progress towards the robust data handling strategies that will be essential for the realisation of automated spectral analysis tools for large scale 2D-IR screening studies of protein-containing solutions and biofluids.

5.2. Introduction

In Chapters 3 and 4 it has been demonstrated that 2D-IR spectroscopy can be used to measure the amide I vibrational band of proteins in aqueous (H_2O) solutions at sub-millimolar concentrations[1] and the LMW amino acid glycine in serum. This ability arises because the non-linear nature of the 2D-IR measurement preferentially amplifies the protein response relative to that of the overlapping water bending vibration that dominates IR absorption measurements of the amide I band.[2,3] Combining the ability to work in water with the sensitivity of the 2D-IR amide I lineshape to protein secondary structure[4,5] allowed the clinically-relevant albumin to globulin ratio of blood serum to be measured from a single transmission-mode 2D-IR spectrum without sample pre-processing. The 2D-IR lineshape sensitivity also allowed spectral separation of the serum albumin, globulins and glycine in a single measurement, resulting in a detection limit of glycine allowing assessment of 2D-IR sensitivity for future protein applications.

The advances reported in Chapters 3 and 4 establish a proof of concept for utilising 2D-IR spectroscopy for biofluid analysis which circumvents the current need for time consuming sample pre-processing or laborious wet chemistry analysis techniques. Moreover, the ability to acquire 2D-IR spectra of proteins in H_2O, rather than D_2O, removes a significant economic barrier to large-scale protein-drug screening studies making 2D-IR more accessible to clinical laboratories and the pharmaceutical sector.

Ultrafast 2D-IR spectroscopy is an established powerful tool for interrogating the structure and dynamics of molecules in the solution phase and has played a significant role in developing our understanding of biomolecular systems, such as proteins and nucleic acids.[6–14] Recent developments in laser technology and mid-infrared pulse shaping[15–18] have provided scope for 2D-IR to be applied in a more analytical manner, for example in high throughput measurements for rapid screening of multiple biomolecule-drug combinations.[19] This combined with the advances outlined in Chapters 3 and 4 indicate the potential for 2D-IR spectroscopy to undergo a transition from the high-end research laboratory to a more mainstream place in the suite of analytical techniques, in a manner that mirrors changes undergone by NMR spectroscopy. Taking such a step however, presents new challenges that must be overcome in terms of the experimental and data handling methods employed currently.

A major challenge is associated with the technical complexity of the 2D-IR measurement. In contrast to absorption spectroscopy, where a measurement of absorbance allows reliable cross-comparison of spectra obtained on different spectrometers and under varying sample conditions, each 2D-IR spectrum is subject to fluctuations in laser pulse energy, laser beam quality, focusing and spatial overlap as well as the usual variables introduced by changes in sample concentration and path length. The impact of issues such as laser energy fluctuations can be eased by the use of referencing or signal averaging, but variations in spectrometer alignment and path length make comparisons between different samples and different measurements problematic, while absolute measurements of concentrations are currently impossible.

The source of variation due to path length is owing to the absence of the spacer between the two CaF_2 windows within the sample cell. As discussed in Chapter 3, in order to control the sample thickness, the tightness of the cell was manually adjusted to obtain an approximate absorbance of 0.1 ± 0.01 OD for the combination band of the δ_{HOH} and librational modes of

water at 2130 cm^{-1} (Figure 5.1(a)). However, the absorbance of the combination band at 2130 cm^{-1} was found to differ before and after the Pump-Probe and 2D-IR measurements (Figure 5.1(a,b)).

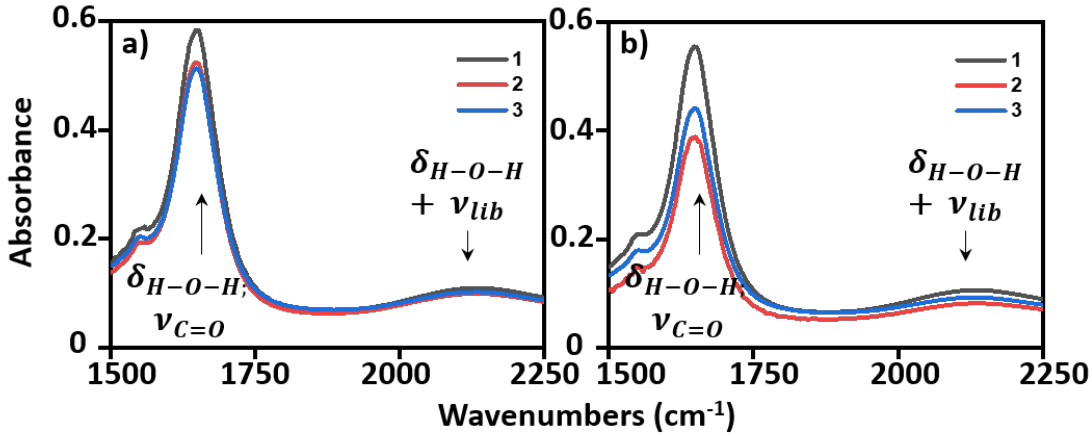

Figure 5.1: *The FTIR spectra showing the variation in the combination band of the δ_{HOH} and librational modes of water before a) and after b) PP and 2D-IR measurement, where 1,2,3 denotes measurement number. Sample thickness was adjusted manually to achieve an absorbance of ~0.1 OD at the δ_{HOH} + ν_{lib} band of water at 2130 cm^{-1}prior to IR investigation.*

It is thought that this occurs during the transport of the cell between the FTIR and 2D-IR instruments which are situated in different rooms at RAL. The 2D-IR laboratories were not spacious enough to house the FTIR spectrometer and so this change in sample thickness was unavoidable in these experiments. This coupled with the need to acquire quantitative measurements of protein concentrations prompted the following study.

Such issues due to the path length variation do not affect applications of 2D-IR spectroscopy where relative changes in lineshapes or peak heights within a single measurement are studied, but they do represent a significant barrier to analytical applications of 2D-IR spectroscopy where quantitative sample-to-sample comparisons are essential. In order to make absolute measurements of concentrations possible a technique needs to be devised to normalise spectra, allowing calibration curves to be created. Ideally the resolution to this should not involve the addition of any labels or further additives to the samples being investigated as for work of biofluids this could be detrimental to the sample or influence changes within the sample. It has been demonstrated that 2D-IR can be used in combination with IR absorption and an independent calibrant molecule to provide measurements of

transition dipole moments of peptide and protein samples.[2,3] Although this approach is powerful, avoiding the need for additional calibration steps or the addition of molecules to the sample is desirable if 2D-IR is to be used in a high throughput fashion.

In this study, the 2D-IR signals of proteins in water is shown to provide a route to normalisation of spectra via the magnitude of the thermal response of the water that follows protein amide I excitation. Such internal spectral normalisation addresses problems of measurement-to-measurement repeatability and confers the ability to determine absolute protein concentrations accurately.

5.3. Experimental Techniques

5.3.1. Sample Preparation

Bovine serum albumin (BSA) was obtained from Sigma Aldrich and used without further purification. BSA was dissolved in an H_2O buffer (pH = 7.5) to mimic the pH of the serum samples. Two stock solutions of BSA were used at concentrations of 30 mg/mL and 50 mg/mL. Serum albumin (BSA) was selected for use in this set of experiments owing to its abundance in serum. Blood serum is a complex mixture including the major protein constituent serum albumin which occurs in humans at a range of 35-50 mg/mL (0.5-0.7 mM) and so the two BSA stock solutions contained clinically-relevant concentrations of BSA. It is important to note that these concentrations correspond closely to the sub-mM levels typical of current 2D-IR studies of protein in deuterated solvents.[20,21] The same process described in Chapter 3 was used to obtain a sample thickness of ~2.75 μm.

5.3.2 IR Absorption Spectroscopy

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^{-1} in the spectral region 400–4000 cm^{-1}.

5.3.3. PP and 2D-IR Spectroscopy

For this study, IR PP and 2D-IR spectra were recorded using both the ULTRA[22,23] and the LIFEtime[16] laser spectrometers at the Central Laser Facility using the Fourier transform 2D-IR method, employing a sequence of three mid-IR laser pulses arranged in a pump–probe beam geometry, as described in Chapter 2. This approach allowed comparison of the signal from

the same sample over a range of cell and spectrometer conditions. Both systems used the pulse shaper technique to deliver sequences of the two pump pulses.[24] Mid-IR pulses with a temporal duration of <50 fs; a central frequency of 1650 cm^{-1} and a bandwidth of ~400 cm^{-1} were obtained for ULTRA. For LIFEtime, pulses with a central frequency of 1650 cm^{-1} were also used however this time with a pulse duration of ~200 fs with a bandwidth of ~100 cm^{-1}.[16]

5.4. Results and Discussions

To demonstrate how this internal spectral normalisation method can be applied to determine absolute protein concentrations, IR pump-probe and 2D-IR spectra of a series of measurements taken from each of two stock solutions of BSA in H_2O-based buffer (pH = 7.5) were taken over a period of two weeks using LIFEtime. Using ULTRA, one stock solution of BSA in H_2O-based buffer (pH = 7.5) was used and measurements were also taken over a two week period.

5.4.1. Pump-Probe Spectroscopy

The IR absorption spectrum of BSA in H_2O is shown in Figure 5.2(a), alongside IR pump-probe spectra (Figure 5.2(b)) of the same sample under the same conditions. In Chapters 3 and 4 it was shown that the IR absorption spectrum BSA in H_2O has an intense peak near 1650 cm^{-1} (Figure 5.2(a)) which is assignable to overlapping contributions from the amide I band of the protein and the H-O-H bending mode of water (δ_{HOH}).

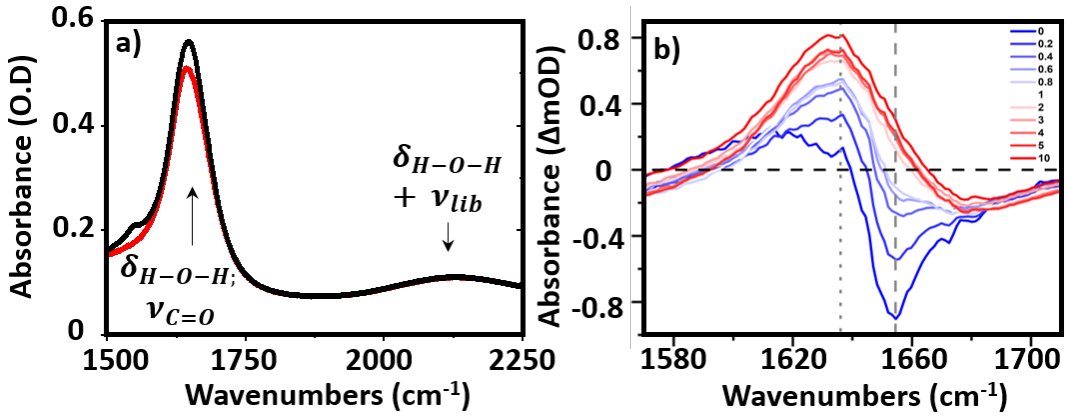

Figure 5.2: *a) IR absorption spectra of 30 mg/mL BSA in H_2O buffer (black) and the H_2O buffer only (red). b) IR pump-probe spectra of 30 mg/mL BSA in H_2O buffer at short (0 ps, blue) and long (10 ps, red) pump-probe time delays. Grey dashed and dotted vertical lines indicate peak positions of protein (1656 cm^{-1}) and water (1634 cm^{-1}) respectively.*

As discussed in section 3.4.2, the response of the protein is shown at short pump-probe delay times (Figure 5.2(b) blue) and a negative feature corresponding to the $v = 0 - 1$ bleach of the amide I band of BSA is visible at 1656 cm^{-1}. Also present is a positive peak shifted to lower wavenumbers which is assigned to the accompanying $v = 1 - 2$ transient absorption. Both features decay rapidly with increasing pump-probe time delay, corresponding to the previously reported vibrational relaxation time of the amide I band of BSA in water (0.78 ps).[15] This relaxation leads to loss of the amide I features at longer time delays and a growth of the signal due to the water response, which is present from 2 ps and is persistent to at least 10 ps (Figure 5.2(b)), which is in good agreement with previous ultrafast studies of the water response.[25]

The basis of the new normalisation method is that both the amide I and thermal H_2O responses in PP (and 2D-IR) spectroscopy originate from similar laser-sample interaction processes, namely those that give rise to the non-linear spectroscopic IR pump-probe (or 2D-IR) signals. Although separated temporally, the magnitudes of the amide I and thermal H_2O signals are influenced in an identical manner by both sample-related (concentration, path length) and instrumental factors (laser intensity, beam quality, beam alignment).[1,16,17]

Thus, the ratio of the magnitudes of the thermal response of water and the resonant amide I response of BSA in each measurement should be constant for a given BSA concentration.

When instrumental factors cause changes in the magnitudes of the signals, the two should be linearly correlated. By extension, for a set of samples in which the BSA concentration varies, the ratio of the thermal water response to the BSA amide I signal will depend linearly upon the BSA content of the sample, because both the 2D-IR and pump-probe signals scale linearly with concentration.[16] The latter is justified because the concentration of water, being the solvent, can be assumed to be constant allowing the water response to be used to normalise the data for direct comparison of the BSA amide I response.

This self-normalisation method applied here also assumes that no significant change in secondary structure of BSA occurs that could influence the 2D-IR signal intensity via changes in amide I coupling within the protein.[16] It is noted, however, that the normalisation method could in principle be used to aid comparisons of samples featuring changes in protein structure with time, for example as a result of disease progression.

With these factors satisfied, normalisation of all spectra to the thermal water response enables the absolute BSA content to be determined taking into account the common instrumental variables. It has been shown in Chapter 3 that although there is a small degree of spectral overlap of the water and protein pump-probe responses near 1650 cm^{-1}, the instantaneous response of the amide I band of BSA is a factor of five greater in magnitude than that of water under these conditions and so the latter is neglected for the purposes of this study.[15]

The vertical dashed and dotted lines in Figure 5.2(b) indicate the two wavenumber positions of a pump-probe spectra monitored as a function of pump-probe delay time. The protein signal at 1656 cm^{-1} is shown as a function of early delay times and the water signal at 1634 cm^{-1} is shown at delay times between 5-10 ps in Figures 5.3(a,b).

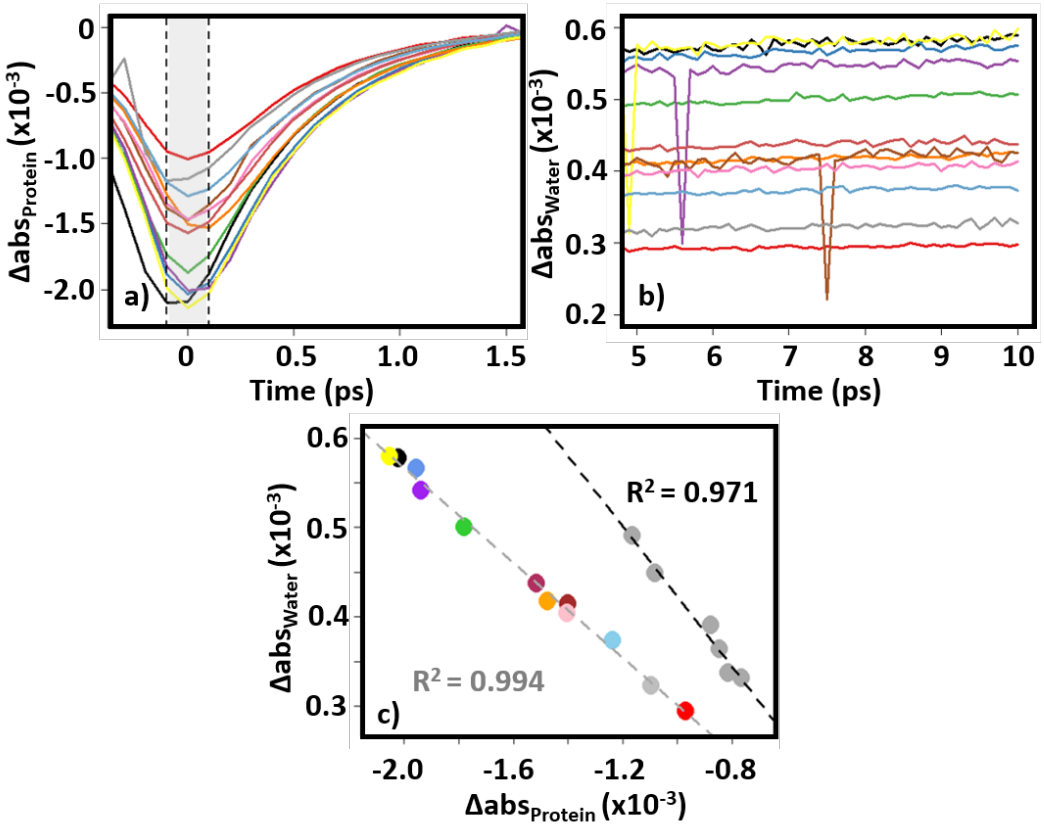

Figure 5.3: *Vibrational relaxation dynamics obtained from IR pump-probe spectroscopy (LIFEtime instrument) of 50 mg/mL BSA in H_2O buffer. a) The temporal dynamics of the amide I response at 1653 cm^{-1} at pump-probe delay times < 1.5 ps. Shaded area indicates pump-probe delay times of 0 +/- 0.1 ps. b) The water response at 1634 cm^{-1} at pump-probe delay times from 5-10 ps. Each data point indicates a different measurement of the same stock solution. c) The linear correlation arising from plotting the average protein absorbance from pump-probe delay times of +/- 0.1 ps (shaded area in (a)) against the water signal averaged between pump-probe delay times of 5-10 ps. Results for a BSA concentration of 50 mg/mL are shown as coloured circles (R^2=0.994) and data for 30 mg/mL BSA samples are shown using grey circles (R^2=0.971). Dashed lines indicate linear fits to the data.*

In order to correlate these two signals, the average response over a range of waiting times was obtained. The protein signal amplitude near the peak of the amide I $v = 0 - 1$ response was obtained from an average of the signal at pump-probe time delays of 0 ± 0.1 ps (Figure 5.3(a), shaded area). The water signal amplitude at 1634 cm^{-1} was obtained from an average between the time delays of 5-10 ps (Figure 5.3(b)). The linear correlation of the protein amide I and water IR pump-probe responses is shown in Figure 5.3(c) for data obtained using the

LIFEtime spectrometer and reveals the expected strong, linear, correlation ($R^2 > 0.971$) for both of the BSA concentrations used (Figure 5.3(c), 50 mg/mL – coloured circles, 30 mg/mL – grey circles).

Despite careful attempts to use the tightness of the sample cell to set the absorbance of the combination band at 2130 cm^{-1} to 0.1 in all cases, a large spread of experimental values were obtained for the protein (-2.0 to - 0.9 $x10^{-3}$ at [BSA] = 50 mg/mL) and thermal water signals (0.3 to 0.6 $x10^{-3}$ at [BSA] = 50 mg/mL) (Figure 5.3(c)). These variations of up to 50% in the measured amplitudes reflect not only changes in the path length but also day to day fluctuations in spectrometer alignment and laser intensity.

The same measurements and analysis was repeated on the ULTRA spectrometer for the 30 mg/mL BSA concentration (Figure 5.4). Figure 5.4(c) reveals the expected strong, linear, correlation ($R^2 = 0.951$). Again a large spread of up to 50% of the experimental values for both the protein (-1.3 to -0.6 $x10^{-3}$) and the thermal water signal (0.4 to 0.9 $x10^{-3}$) is seen for the ULTRA spectrometer. What is clear however is that, despite the large fluctuation in signals acquired for each individual measurement, the protein and water thermal response are linearly correlated independent of the instrument used (Figures 5.3(c) and 5.4(c)).

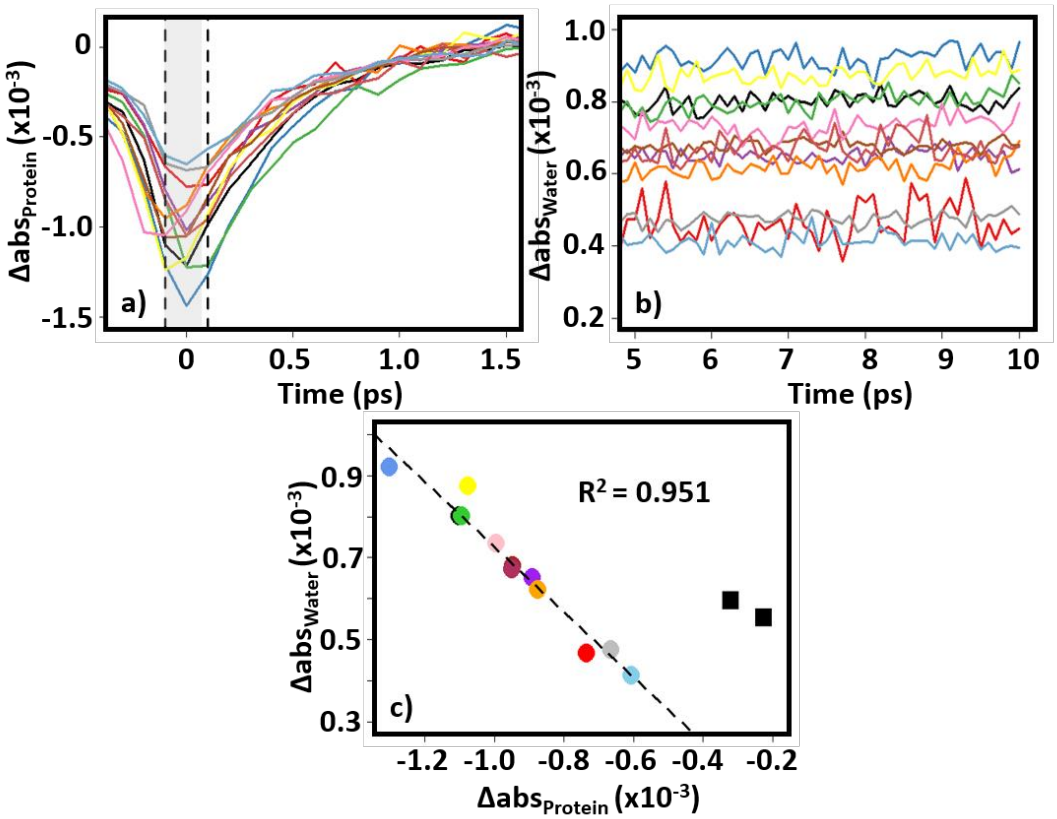

Figure 5.4: (a,b) *Vibrational relaxation dynamics obtained from IR pump-probe spectroscopy (ULTRA instrument) of 30 mg/mL BSA in H_2O buffer. a) The temporal dynamics of the amide I response at 1653 cm^{-1} at pump-probe delay times < 1.5 ps. b) The water response at 1634 cm^{-1} at pump-probe delay times from 5-10 ps. c) The linear correlation arising from plotting the water signal averaged between pump-probe delay times of 5-10 ps against the average protein absorbance from pump-probe delay times of $\pm$ 0.1 ps (shaded area in (a)). Each data point indicates a different measurement of the same stock solution. Dashed lines indicate linear fits to the data. The black squares in c) indicate results obtained from pure water samples for comparison.*

For both concentrations used, LIFEtime yields correlations greater than that obtained for ULTRA. Independent of the instrument used, the signals are averaged for 1 second and the resulting spectra is averaged over three repeated measurements in order to achieve a high signal to noise ratio. ULTRA has a repetition rate of 10 kHz which results in one PP measurement taking just under 10 minutes. However with LIFEtime's higher repetition rate (100 kHz) the time taken to acquire a PP measurement is reduced to only 1 minute. Achieving

the spectra on a much quicker timescale results in stronger correlated data as shown in Figures 5.3(c) and 5.4(c).

To ascertain that the linear relationship is due to protein content, two samples of H_2O were also treated in a similar manner (Figure 5.4(c), squares) and these data points lie significantly off the linear region describing the protein samples. This is as expected owing to the fact that at early waiting times the signal amplitude of the H_2O is <20 % of that of the proteins (Chapter 3) but still contains the same number of water molecules and so produces a late delay time thermal response comparable to the BSA samples which is dependent on the sample thickness.

5.4.2. 2D-IR Spectroscopy

As discussed previously, at early waiting times the water signal is negligible allowing measurement of the protein amide I band without the water contribution. At later waiting times where the protein signal has decayed, the water response is present and this is due to solvent heating via energy transfer, as discussed in Chapter 3. This results in the separation of the protein amide I signal at early waiting times and the water thermal response at late waiting times. To extend the normalisation method to 2D-IR, the spectra presented here were recorded at a waiting time (T_w) of 250 fs and 5 ps between pump and probe pulses. Pairs of spectra at these waiting times were obtained from each of the BSA samples, three of which are shown from the LIFEtime spectrometer in Figure 5.5.

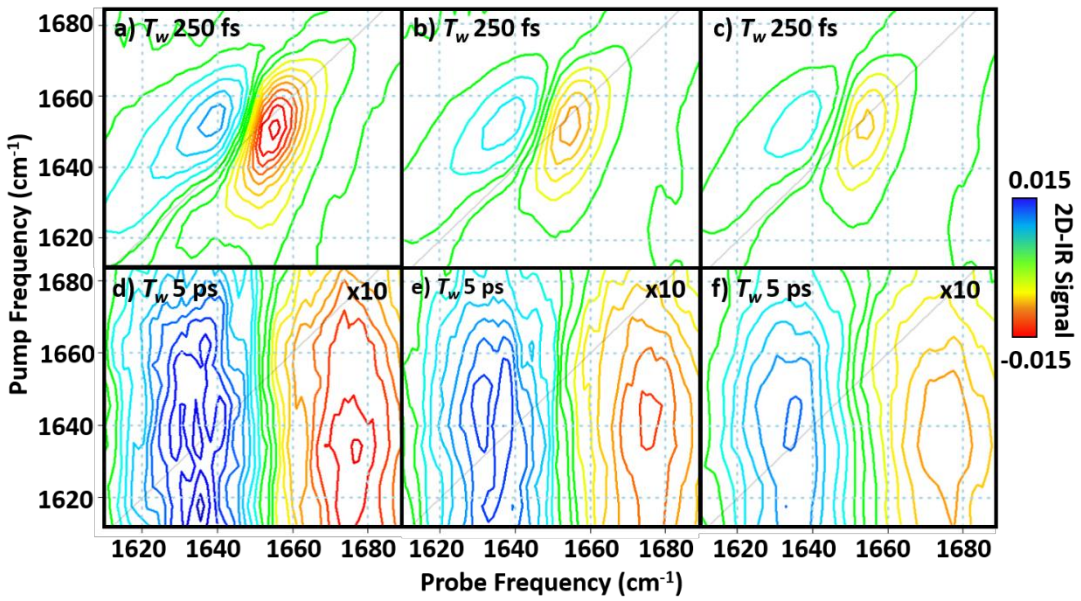

Figure 5.5: *2D-IR spectra taken for 3 measurements of the 50 mg/mL BSA stock solution on the LIFEtime spectrometer. (a-c) 2D-IR spectra taken at $T_w = 250$ fs. (d-f) 2D-IR spectra taken at $T_w = 5$ ps. All 2D-IR spectra are plotted on the same scale, colour bar shown, (d-f) are magnified by a factor of ten.*

The 2D-IR spectrum of the 50 mg/mL BSA solution at waiting times (T_w) of 250 fs (Figure 5.5(a-c)) and 5 ps (Figure 5.5(d-f)) show a similar result to the pump-probe data. In the $T_w =$ 250 fs spectrum, peaks due to the $v = 0 - 1$ (red) and $v = 1 - 2$ (blue) transitions of the amide I band of BSA are clearly visible. This is replaced at $T_w = 5$ ps by the thermal response of water, which is magnified by a factor of 10 in the figure.[26,27]

Several different methods were used to extract the protein peak intensity at a T_w value of 250 fs and the thermal water signal at $T_w = 5$ ps. An example is demonstrated for 2D-IR spectra obtained with the LIFEtime (Figure 5.6(a-c)) and ULTRA instruments (Figure 5.6(d-f)). Projection of the LIFEtime 2D-IR spectra onto the probe frequency axis of the 2D-IR spectrum (Figure 5.6(a, b)) allowed quantification of the BSA response.

As mentioned previously, the bandwidth between the two instruments is significantly different. ULTRA has a bandwidth of ~400 cm^{-1} while LIFEtime is ~100 cm^{-1} which has a significant impact on the range of molecular vibrations that can be accessed by each spectrometer. With LIFEtime having a smaller bandwidth of ~100 cm^{-1}, the pump frequencies between 1580 - 1720 cm^{-1} were used to project onto the probe axis. By increasing

LIFEtime's pump range we introduce noise to the projections and a decrease in the linear correlations produced was found. An attempt was made to utilise the same pump frequency range for ULTRA, however by not including the pump frequencies outside this range important off diagonal information that is needed to help correlate between the water and protein signals is removed. As a result of this the pump frequency range, 1400 - 1700 cm^{-1}, was used to project onto the probe axis for ULTRA data.

In order to quantify the responses, amplitudes of these projections are averaged over probe frequency ranges (Figure 5.6). This is done by averaging the signal in the $T_w = 250$ fs spectrum over a small frequency range near the peak of the BSA signal (Figure 5.6(a,d) shaded area). The value of the thermal water response was measured using the $T_w = 5$ ps spectrum near 1675 cm^{-1} (Figure 5.6(b,e), shaded area). The correlation of the signals obtained for 30 and 50 mg/mL samples (grey and coloured circles, respectively) is shown in Figure 5.6(c) for LIFEtime. Figure 5.6(f) shows the correlation produced for the 30 mg/mL sample on ULTRA.

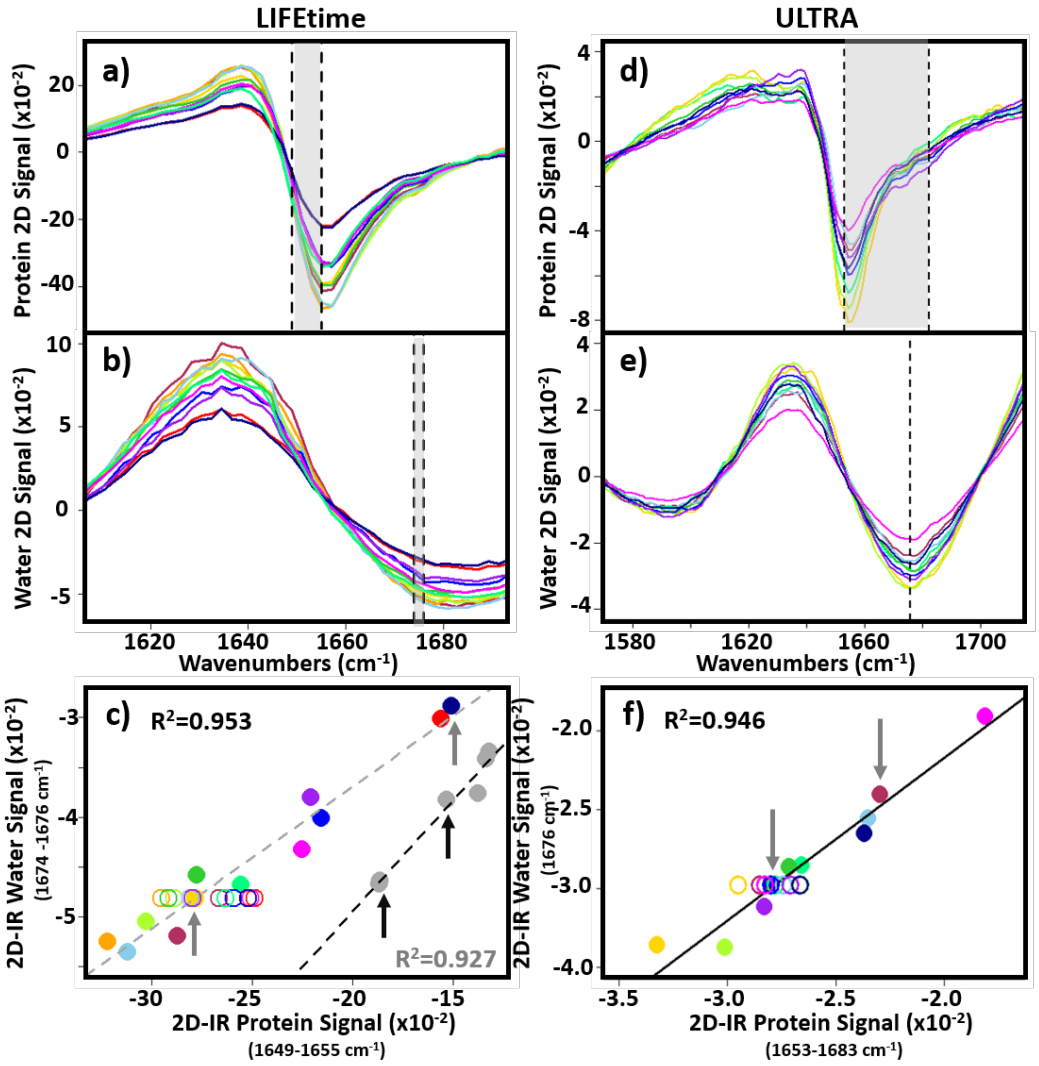

Figure 5.6: *(a-c) Projection of 2D-IR spectra of 50 mg/mL BSA in H_2O buffer over pump frequencies between 1580-1720 cm^{-1} onto the probe frequency axis at waiting times of a) 250 fs showing the protein response and b) 5 ps showing the water response. c) Correlation of the water signal between 1674-1676 cm^{-1} with the protein signal between 1649-1655 cm^{-1} (shaded areas in (b) and (a), respectively). Each solid coloured circle indicates an individual measurement. Open circles show the spread of the spectra after scaling to the water response of the gold spectrum (see text). Grey circles show the results obtained using a 30 mg/mL BSA solution. (d-f) Using the ULTRA instrument, the 2D-IR spectra of 30 mg/mL BSA in H_2O buffer are projected onto the probe axis at waiting times of d) 250 fs and e) 5 ps using the pump frequency range between 1400-1700 cm^{-1}. f) Correlation of the water signal at 1676 cm^{-1} with the protein signal between 1653-1683 cm^{-1} (shaded area in (d)). Open circles show the spread of the spectra after scaling to the water response of the blue spectrum. Arrows denote the pair of spectra used for the calculation of difference spectra in Figure 5.8.*

Although a large spread in absolute values is observed, as for the IR pump-probe data, a clear linear relationship between protein and water signals is present in the data (Figure 5.6(c,f)). Changing the spectral region within the $v = 0 - 1$ transition of the amide I band used to obtain the protein signal led to small variations in the linear correlation produced but in the majority of cases the R^2 value obtained was > 0.9. Comparable data obtained with the ULTRA spectrometer are shown in Figure 5.6(d,e,f).

The results from both IR pump-probe and 2D-IR experiments show that the intensity of the thermal water response can be used as an internal normalisation standard to account for fluctuations in spectrometer performance in both experiments. The validation of an internal normalisation method leads to two powerful applications. Firstly, normalisation of spectra from individual measurements provides a basis to extract BSA concentrations directly from 2D-IR spectra. The plots in Figure 5.6(c, f) act as calibration curves for 30 and 50 mg/mL BSA samples. This is demonstrated in Figure 5.7(a,b) where the averaged protein signals obtained for each BSA concentration using the LIFEtime instrument were used to create a linear calibration plot of 2D-IR protein signal versus BSA concentration. Leave one out analysis was then performed by individually omitting each measurement from the creation of the calibration plot and using the result to estimate the concentration of the left out sample. The results prior to the normalisation are shown in Figures 5.7(c,d). This approach is based on only two concentration points however it can be seen that applying the normalisation method leads to a significant reduction in spread of the concentrations (Figure 5.7(b) compared to Figure 5.7(d)). Using the normalisation approach leads to a measurement of the 50 mg/mL BSA concentration accurate to $\pm$4.5 mg/mL (< 10%).

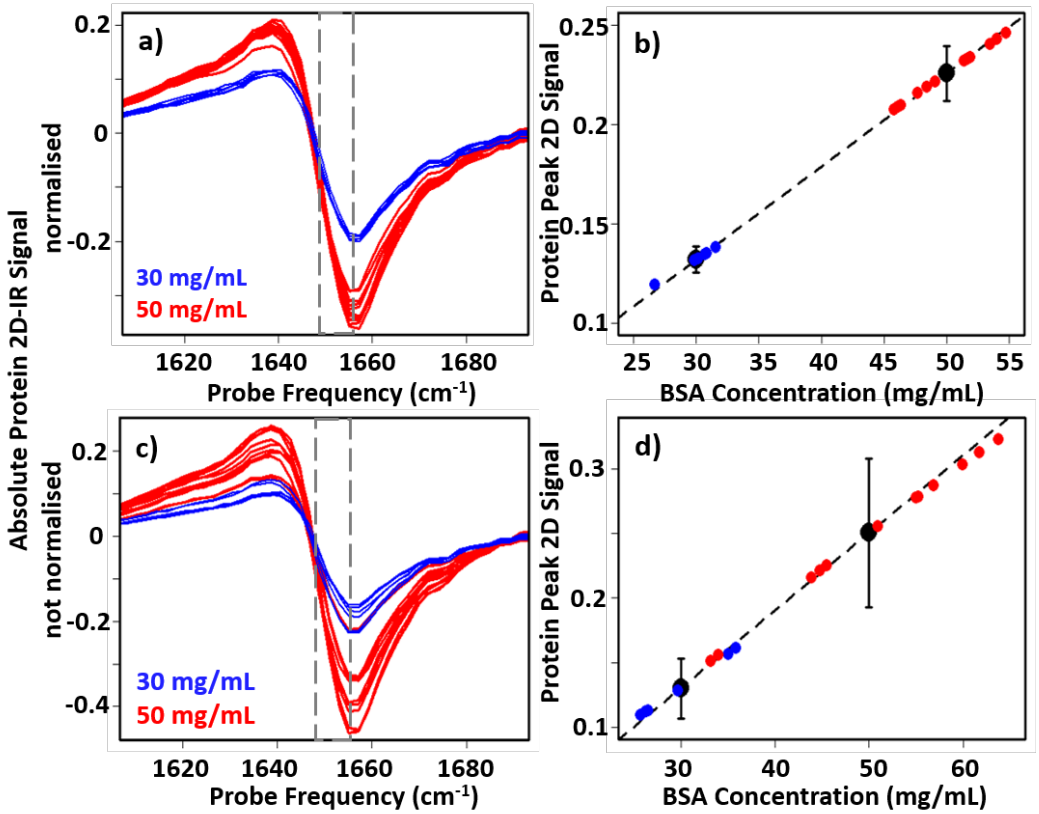

Figure 5.7: *Use of 2D-IR signal of BSA to estimate protein concentration. The 2D-IR pump projections are shown (a) normalised (b) not normalised. The average protein 2D-IR signal obtained for two known BSA concentrations (black dots) were used to create a linear calibration plot (dashed line). Each sample in turn was left out of the creation of the calibration plot and the result used to estimate the concentration of the left out sample based upon the protein signal size. The results are plotted as points for 50 mg/mL (red) and 30 mg/mL (blue) BSA concentrations respectively. Comparing the normalised data (a.b) with the non-normalised data (c,d) shows the significant increase in accuracy of the results obtained. Post normalisation shows that the BSA concentrations were obtained accurate to* $\pm$ *9%. The error bars show the range of protein signals measured with (b) and without (d) normalisation.*

If similar relationships were produced for a range of BSA concentrations then, for an unknown sample, a given ratio of the water response to the protein signal would yield the BSA concentration. Combining a similar approach using human serum albumin with the previously demonstrated method for measuring the albumin to globulin ratio of blood serum with 2D-IR[15] makes it possible to obtain the clinically relevant albumin and globulin

concentrations and so the total protein content of the serum from a single 2D-IR measurement within a few seconds.

Secondly, the internal normalisation approach can be extended to allow rapid cross-comparison of protein spectra in a high-throughput screening context. For example, an experiment is visualised in which a series of 2D-IR spectra are measured from a range of samples featuring the same protein at the same concentration in aqueous buffer in complex with a range of alternative ligands (or of the same ligand at a range of concentrations).[28] In this case, relatively small differences in the spectrum of the protein would be expected as a result of ligand binding and careful production of difference 2D-IR spectra would be needed to extract relevant information.[29,30] Normalisation of the spectra to the thermal water response would not significantly increase the measurement time, but would provide an experimentally determined route to difference spectral analysis by reducing the impact of instrumental fluctuations.

The efficiency of this approach is demonstrated using two BSA 2D-IR spectra from different points in the range of signal amplitudes measured (Figure 5.6(c,f), arrows). Creating difference spectra by simply subtracting one spectrum from the other for pairs of 50 mg/mL (Figure 5.8(a)) and 30 mg/mL (Figure 5.8(d)) LIFEtime samples, Figure 5.8(g) shows a pair of 30 mg/mL ULTRA samples which all result in a clearly visible residual BSA signal (Figure 5.8(a,d,g), arrows). This reflects the range of measured values from a common sample arising from instrumental effects. However, normalising the 2D-IR spectra to the thermal water signal prior to calculating the difference spectrum reduces the residual signal dramatically (Figure 5.8(b,e,h)). This is as would be expected for difference spectra comparing two identical samples. Indeed, magnification by a factor of 20 for data obtained using LIFEtime and 10 for data obtained using ULTRA shows how effective the normalised difference spectral measurement approach is (Figure 5.8(c, f, i)).

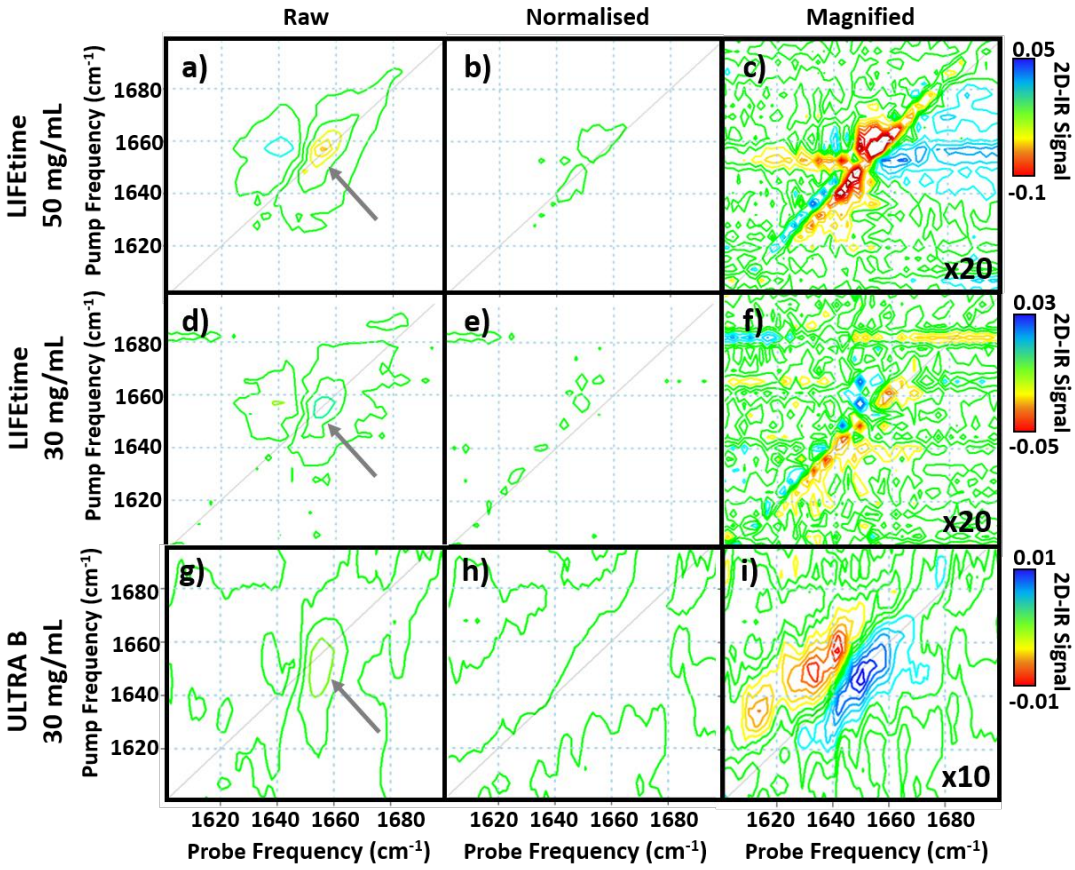

Figure 5.8: *a) 2D-IR difference spectra of two 50mg/mL BSA samples obtained using LIFEtime before any scaling. b) 2D-IR difference spectra obtained from the two spectra after scaling to the water response. c) As b) magnified by 20 times. Colour scale is shown. d) 2D-IR difference spectra of two 30mg/mL BSA samples obtained using LIFEtime before any scaling. e) 2D-IR difference spectra of the same two spectra after scaling to the water response. f) as e) magnified by 20 times. Colour scale is shown. g) 2D-IR difference spectra of two 30mg/mL BSA samples obtained using ULTRA before any scaling. h) 2D-IR difference spectra of the same two spectra after scaling to the water response. i) as h) magnified by 10 times. Colour scale is shown. Arrows shown in a,d,g) indicate residual response after subtraction.*

For measurements using the ULTRA spectrometer, the lower pulse repetition rate (10 kHz) leads to a slower data acquisition time of ~25 minutes per 2D-IR spectrum, as opposed to < 1 minute using LIFEtime. This has resulted in a less effective subtraction process, which is ascribed to slow changes in the intensity distribution of the broad bandwidth laser pulses during each spectral acquisition. This will not be accounted for by the normalisation method

and so we conclude that the normalisation approach is best suited to measurement protocols that employ rapid acquisition of 2D-IR spectra.

To show that this approach works for all of the samples studied, all 2D-IR spectra of 50 mg/mL BSA samples obtained using LIFEtime were normalised to the water response of one spectrum and the protein signals re-plotted versus the normalised water signals in Figure 5.6(c) (open circles). The same process was repeated for all 2D-IR spectra of 30 mg/mL BSA samples obtained using ULTRA (Figure 5.6(f), open circles). Now all of the samples show the same water signal (vertical axis) as expected following normalisation and the spread of protein signal sizes along the horizontal axis is significantly reduced. For LIFEtime this results in a reduction of the spread by 73% and for ULTRA a reduction of 64%.

5.5. Conclusions

In Chapters 3 and 4, experiments provide a simple technique of working around the issues faced by measuring 2D-IR in physiological solvents. However the issue of sample to sample repeatability and fluctuations in the laser system prevented absolute quantitative analysis of protein concentrations. In this chapter, these challenges are explored and a method has been devised to create an internal normalisation standard, allowing quantitative protein concentrations to be obtained.

Normalisation of each early waiting time spectra to its corresponding thermal response of water occurring at late waiting times provides a route to producing difference spectra with reduced impact from instrumental variations, without the addition of labels to the system. Furthermore, this goes significantly beyond recent work shown in Chapters 3 and 4 showing that 2D-IR can be applied to study aqueous (H_2O) biofluids in transmission without prior sample preparation steps by adding the ability to determine absolute protein concentrations via the amide I band, which cannot be achieved using absorption spectroscopy methods.

It is noted that improvements in sample cell design could be made to limit the path length variability such as when using standard transmission cells. Etching an area on a CaF_2 window would offer control of the sample thickness. These can be manufactured in such a way so that the path length between two windows is fixed at all times however precise engineering is costly. The issues of spectrometer alignment and laser fluctuation would remain and therefore this simple analytical approach provides a means of avoiding complex and expensive engineering solutions.

Instrument comparison has shown that faster acquisition time for measurements results in a more effective subtraction process, so it can be concluded that the normalisation approach is best suited to measurement protocols that employ rapid acquisition of 2D-IR spectra. Developments in laser technology and mid-infrared pulse shaping[15–18] have provided scope for 2D-IR to be applied in high throughput measurements for rapid screening of biomolecule-drug combinations.[19] This combined with the advances made in this thesis thus far removes a significant economic barrier to large-scale physiological protein-drug screening studies making 2D-IR more accessible to the pharmaceutical sector.

5.6. References

1. Hume, S., Hithell, G., Greetham, G. M. Donaldson, P. M. *et al.* Measuring proteins in H_2O with 2D-IR spectroscopy. *Chem. Sci.* **10,** 6448–6456 (2019).

2. 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.* **137,** 184202 (2012).

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* **119,** 14065–14075 (2015).

4. Baiz, C. R., Peng, C. S., Reppert, M. E., Jones, K. C. *et al*. Coherent two-dimensional infrared spectroscopy: Quantitative analysis of protein secondary structure in solution. *Analyst* **137,** 1793–1799 (2012).

5. Minnes, L., Shaw, D. J., Cossins, B. P. Donaldson, P. M. *et al.* Quantifying secondary structure changes in calmodulin using 2D-IR spectroscopy. *Anal. Chem.* **89,** 10898–10906 (2017).

6. Hamm, P. & Zanni, M. T. *Concepts and Methods of 2D Infrared Spectroscopy*. (Cambridge University Press, 2011).

7. Simpson, N. & Hunt, N. T. Ultrafast 2D-IR spectroscopy of haemoproteins. *Int. Rev. Phys. Chem.* **34,** 361–383 (2015).

8. Hunt, N. T. 2D-IR spectroscopy: ultrafast insights into biomolecule structure and function. *Chem. Soc. Rev.* **38,** 1837–1848 (2009).

9. Hunt, N. T. Transient 2D-IR spectroscopy of inorganic excited states. *Dalt. Trans.* **43,** 17578–17589 (2014).

10. Adamczyk, K., Candelaresi, M., Robb, K., Gumiero, A. *et al.* Measuring protein dynamics with ultrafast two-dimensional infrared spectroscopy. *Meas. Sci. Technol.* **23,** 62001 (2012).

11. Ghosh, A., Ostrander, J. S. & Zanni, M. T. Watching proteins wiggle: Mapping structures with two-dimensional infrared spectroscopy. *Chem. Rev.* **117,** 10726–10759 (2017).

12. Kim, H. & Cho, M. Infrared probes for studying the structure and dynamics of biomolecules. *Chem. Soc. Rev.* **113,** 817–5847 (2013).

13. Hithell, G., Ramakers, L. A. I., Burley, G. A. & Hunt, N. T. *Applications of 2D-IR Spectroscopy to Probe the Structural Dynamics of DNA* in *Frontiers and Advances in Molecular Spectroscopy* (Elsevier, 2018).

14. Rubtsova, N. I. & Rubtsov, I. V. Vibrational energy transport in molecules studied by two-dimensional infrared spectroscopy. *Annu. Rev. Phys. Chem.* **66,** 717–738 (2015).

15. Shim, S. H. & Zanni, M. T. How to turn your pump–probe instrument into a multidimensional spectrometer: 2D IR and vis spectroscopies via pulse shaping. *Phys. Chem. Chem. Phys.* **11,** 748–761 (2009).

16. 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* **122,** 780–787 (2018).

17. Strasfeld, D. B., Ling, Y. L., Shim, S. & Zanni, M. T. Tracking fiber formation in human islet amyloid polypeptide with automated 2D-IR spectroscopy. *J. Am. Chem. Soc.* **130,** 6698–6699 (2008).

18. Tracy, K. M., Barich, M. V, Carver, C. L., Luther, B. M. *et al*. 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.* **7,** 4865–4870 (2016).

19. Fritzsch, R., Donaldson, P. M., Greetham, G. M., Towrie, M. *et al.* Rapid screening of DNA-ligand complexes via 2D-IR spectroscopy and ANOVA-PCA. *Anal. Chem.* **90,** 2732–2740 (2018).

20. Baiz, C. R., Reppert, M. & Tokmakoff, A. Introduction to protein 2D IR spectroscopy. *Ultrafast Infrared Vib. Spectrosc.* 361–404 (2012).

21. Hamm, P., Lim, M. & Hochstrasser, R. M. Structure of the amide I band of peptides measured by femtosecond nonlinear-infrared spectroscopy. *J. Phys. Chem. B* **102,** 6123–6138 (1998).

22. Greetham, G. M., Burgos, P., Cao, Q., Clark, I. P. *et al.* ULTRA: A unique instrument for time-resolved spectroscopy. *Appl. Spectrosc.* **64,** 1311–1319 (2010).

23. Greetham, G. M., Matousek, P., Robinson, D. A., Parker, A. W. *et al*. ULTRA laser system: a new dual-output 10 kHz Ti:Sapphire amplifier with UV–IR generation for time-resolved spectroscopy. *Cent. Laser Facil. Annu. Rep.* 249–250 (2008).

24. 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.* **104,** 14197–14202 (2007).

25. Ashihara, S., Huse, N., Espagne, A., Nibbering, E. T. J. *et al*. Vibrational couplings and ultrafast relaxation of the O-H bending mode in liquid H2O. *Chem. Phys. Lett.* **424,** 66–70 (2006).

26. Kraemer, D., Cowan, M. L., Paarmann, A., Huse, N. *et al.* Temperature dependence of the two-dimensional infrared spectrum of liquid H_2O. *Proc. Natl. Acad. Sci.* **105,** 437–442 (2008).

27. Chuntonov, L., Kumar, R. & Kuroda, D. G. Non-linear infrared spectroscopy of the water bending mode: direct experimental evidence of hydration shell reorganization? *Phys. Chem. Chem. Phys.* 13172–13181 (2014).

28. Johnson, P. J., Kozoil, K. L. & Hamm, P. Quantifying biomolecular recognition with site-specific 2D infrared probes. *J. Phys. Chem. Lett.* **8,** 2280–2284 (2017).

29. Shaw, D. J. Hill, R. E., Simpson, N., Husseini, F. S. *et al.* Examining the role of protein structural dynamics in drug resistance in Mycobacterium tuberculosis. *Chem. Sci.* **8,** 8384–8399 (2017).

30. Hithell, G., Shaw, D. J., Donaldson, P. M., Greetham, G. M. *et al.* Long-range vibrational dynamics are directed by Watson-Crick base pairing in duplex DNA. *J. Phys. Chem. B* **120,** 4009–4018 (2016).

6. Detection of Drug Binding to Serum Albumin using 2D-IR

6.1. Abstract

In the previous chapters, the potential of 2D-IR as an analytical tool for biomedical diagnostics have been discussed. So far, dynamics of proteins in H_2O and D_2O have been compared, comparative measurements of concentrations of key proteins in serum have been made, detection of the presence of the LMW molecule glycine in serum has allowed evaluation of obtainable detection limits using 2D-IR and, by utilising the thermal response of water at late waiting times, a normalisation technique has been developed to allow quantitative measurements of proteins to be made in H_2O using 2D-IR spectroscopy.

In this chapter, the application of 2D-IR spectroscopy to blood serum measurements is explored further and applied to evaluating the ability of 2D-IR to detect changes in the spectroscopy of serum albumin upon ligand binding. As a carrier protein, serum albumin is important for transporting a wide variety of molecules around the body and upon binding structural changes in the protein can be induced. Understanding these changes could offer a means of monitoring drug behaviour in the circulatory system and holds a significant importance in pharmaceutical problems such as drug metabolism and delivery. Application of the normalisation method described in Chapter 5 allows drug-protein complexes of serum albumin with clinically relevant levels of paracetamol, ibuprofen and warfarin to be studied in aqueous serum. Typical concentrations of these drugs found in the bloodstream after a standard typical dose are in the µmol/L region which may pose challenging for 2D-IR detection. However as 2D-IR detects the coupled structure of the protein molecule rather than the drug itself, it is important to understand whether 2D-IR can detect drug binding in serum. Drug-albumin complexes are detected in serum and changes in the protein secondary structure caused by drug binding are explored using transition dipole moment theory.

6.2. Introduction

Drug-binding to serum proteins plays a major role in drug therapy as binding provides a reservoir for many compounds and affects the pharmacokinetics of drugs. As only the free (unbound) portion of the drug is responsible for the pharmacological action it is important to understand how drugs and proteins bind with one another in biological fluids.[1–4] 2D-IR has been shown to be highly sensitive to small changes in molecular structures of proteins upon ligand-binding by producing patterns in the 2D-IR spectra from coupled vibrational modes within a molecule allowing presence of ligands to be observed.[5,6] As this thesis focuses on

the use of 2D-IR spectroscopy for biofluid analysis, particularly in the amide I region of proteins, changes in the spectroscopy of the amide I band of serum albumin upon drug-binding are explored in this chapter.

2D-IR has been shown to be sensitive to small changes in secondary structure as well as the detection of drug binding. Further to previous works where 2D-IR spectroscopy has been shown to be sensitive to protein secondary structure,[7,8] its ability to differentiate small changes in protein secondary structure has also been investigated which may be powerful in determining structural changes of proteins during disease states or allosteric affects due to binding.[9–12] Grechko and Zanni have shown that the ratio of the magnitude of the amide I band using 2D-IR with that of IR absorption spectroscopy allows measurement of the amide I transition dipole moment. This is closely dependent upon molecular coupling which can indicate protein secondary structure.[9,10] A recent study of the calcium-binding messenger protein calmodulin (CaM) has shown the capability of 2D-IR to detect small secondary structure changes upon binding of Ca^{2+} ions.[11] The 2D-IR measurements observed a thermally induced reduction in α-helical content (15 %) upon binding, an induced change affecting only seven (5 %) residues. These studies were validated using circular dichroism spectroscopy which detected a comparable 13 % change.

Applications of label-free 2D-IR has also been used to study the sensitivity of drug-protein binding in InhA, an enzyme partly responsible for the maintenance of cell walls in the bacterium Mycobacterium tuberculosis. Correlation of drug-protein complexes which induces inhibition of bacterium activity with changes in the off-diagonal region of the amide I lineshapes were found.[12] Using peptides labelled with the non-natural amino acid azidohomoalanine and the protein tyrosine phosphatase 1E, quantification of ligand binding has also been observed using 2D-IR, producing dissociation constants comparable with isothermal titration calorimetry (ITC) results.[6] These studies highlight the potential of 2D-IR to detect and understand drug-protein complexes, allowing more accurate and effective drug treatments to be developed.

As the major protein constituent of serum, albumin plays an important role in the pharmacokinetic behaviour of many drugs as binding can affect their rate of distribution within the body as well as their rates of metabolism and excretion.[13,14] Serum albumin primarily functions as a carrier (or transport) protein, known to bind to a wide variety of substrates, such as steroids, hormones, fatty acids, glucose, metal ions and pharmaceuticals

including warfarin and ibuprofen.[13,15–22] X-ray crystallography studies have shown that serum albumin is composed of three major domains, I, II and III, which form to create a heart shaped molecule (Figure 6.1). Within each domain are two α-helix subdomains, A and B, which are connected by a random coil and inter-domain helices link subdomains IB to IIA and also IIB to IIIA.[13,17,23–26]

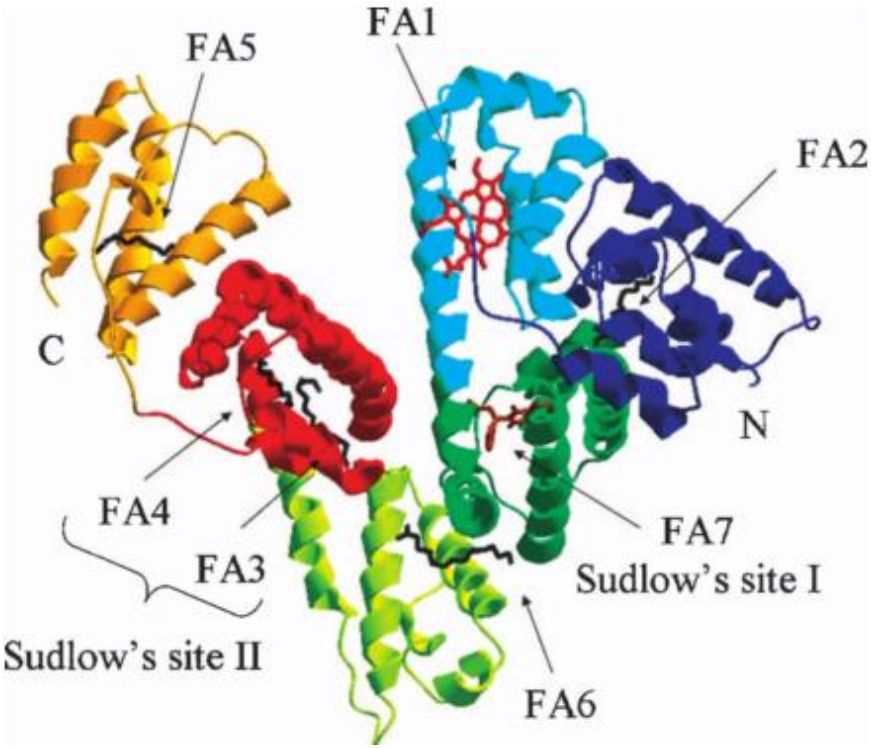

Figure 6.1: *Serum albumin structure, reprinted with permissions from reference.*[13] *Three domains and their A and B subdomains are denoted as follows: domain I - dark blue (IA) and light blue (IB), domain II - dark green (IIA) and light green and domain III - red (IIIA) and orange (IIIB). The two drug binding sites are shown as Sudlow I and II and the seven fatty acid (FA) binding sites are also labelled. Except for FA1 and FA7, all other FA sites are occupied by myristate. Sudlow's (drug) site I overlaps with FA7 and is occupied in this diagram by warfarin. Sudlow's (drug) site II overlaps with both FA3 and FA4. Myristrate is shown to occupy sites FA2-6. A heme is shown in FA1.*

There are two major binding sites for ligands located in subdomains IIA and IIIA, known as drug sites I and II, respectively (labelled as Sudlow's sites I and II in Figure 6.1).[13,14,27–29]. There are also 7 known fatty acid (FA) binding sites distributed throughout the molecule.[15,17,30–32]

Studies have documented the conformational change in serum albumin structure upon drug and fatty acid binding. In particular the system tends to undergo a change known as the neutral-basic (NB) transition. The neutral (N) form is characterised by a heart shaped structure. Between the pH values 4.3 - 8.0 and in the presence of drugs or FAs, the serum albumin transitions to the basic (B) structure which involves the breaking of salt bridges between domains I and III. This increases the proteins structural flexibility, typically resulting from a loss of α-helix and the structural change from a heart to an open ellipsoid (doughnut) shape, regarded as changes of domain I and III relative to domain II.[13,18,29,33–36]

This transition from the neutral to basic conformation of albumin has been shown by Petitpas *et al* to also be linked to a change in pH which can be induced from fatty acid binding.[36] Upon binding of warfarin to serum albumin this NB transition has been documented, as well as a 3-fold increase in the binding affinity for warfarin on the albumin molecule when the protein is already bound to FAs, at a 1:3 molar ratio of albumin:FA.[35] However an influx of fatty acids, greater than 1:3 ratio has seen the release of warfarin from the albumin binding site, allowing delivery of the drug.[13,35] Daneshgar *et al* have demonstrated a decrease in the α-helix content of human serum albumin upon paracetamol binding to drug site II using circular dichroism spectroscopy.[37] Further studies have shown that glucose bound to albumin (glycated albumin) decreases the binding affinity for paracetamol by half when compared to glucose free albumin. This has significant impact upon diabetic patients who have significantly increased levels of glycated albumin compared to non-diabetics.[21]

The free fraction of drugs in serum, which is responsible for the therapeutic action taken by the drug, is also dependent on the presence of other drugs that have been administered to the patient. The competition of multiple drugs binding to serum albumin depends on the relative binding strengths of each drug and this rivalry will modify the unbound fraction of each drug circulating the bloodstream.[3] Using fluorescence and UV absorption techniques, Sułkowska *et al* suggests that when administered together paracetamol and cytarabine competitively bind to the albumin drug site I, by increasing the concentration of one drug the binding properties of the other are altered.[38] Although many investigations have shown the effects of drug and fatty acid binding to serum albumin, the complex nature and patient to patient variation of serum makes validating true events challenging.[39–45]

In this chapter, attempts are made to further test the capabilities of 2D-IR spectroscopy in its ability to detect small changes in serum albumin's secondary structure due to protein-drug

binding. Using three common drugs, paracetamol, ibuprofen and warfarin, measurements of drug binding to serum albumin are explored. Typical concentrations of these drugs found in the bloodstream after a standard typical dose are μmol/L,[46–51] which would pose challenging for 2D-IR detection which has been shown to detect mmol/L changes in previous chapters, however in this chapter 2D-IR spectroscopy is used to detect changes in the albumin structure upon binding rather than the free drug itself. The detection of drug-binding of the three drugs studied was not conclusive at clinically-relevant levels, yet the addition of paracetamol to serum albumin is shown to induce a change in the α-helix content of albumin upon binding. Investigation into the variation of results from serum albumins within serum which is a natural environment containing other factors including fatty acids, and serum albumins dissolved in a water buffer are explored.

6.3. Experimental Techniques

6.3.1. Sample Preparation

Pooled equine serum, serum albumin (bovine), acetaminophen (paracetamol), ibuprofen and warfarin were obtained from Sigma Aldrich and used without further purification. 2D-IR measurements of individual drugs were performed in dimethyl sulfoxide (DMSO) as they are not all soluble in water, allowing clear spectra to be obtained using high concentrations (20 mg/mL) with a sample path length of 25 μm. Conversions between mg/mL and mmol/dm^3 are shown in Table 1.1.

Table 6.1: *Conversions of 20 mg/mL of paracetamol, ibuprofen and warfarin into their respective molar concentrations.*

Drug	Concentration (mg/mL)	Molar Concentration (mmol/dm^3)
Paracetamol	20	132
Ibuprofen	20	97
Warfarin	20	65

To study the spectroscopy of serum samples at a range of drug concentrations, paracetamol was spiked into pooled horse serum and the following drug : albumin molar ratios were created, as shown in Table 6.2, yielding a total of 10 samples (9 spikes and pure serum). The same molar ratios were created for the ibuprofen and serum datasets. These ratios include clinically relevant levels of each drug after a typical oral dose.[49–51] Warfarin doses are much smaller and so different concentrations were used, as shown in Table 6.2, producing a total of 9 samples (8 spikes and pure serum).

Table 6.2: *Summary of Paracetamol/Ibuprofen/Warfarin: Albumin molar ratios. Serum albumin has a concentration of 0.45 mmol/dm^3 in the body. Samples will be referred to as ^{n}[X] where n indicates an n:1 molar ratio drug, X, to albumin, as shown.*

Paracetamol			**Ibuprofen**			**Warfarin**		
Sample	**µg/mL**	**mmol/dm³**	**Sample**	**µg/mL**	**mmol/dm³**	**Sample**	**µg/mL**	**µmol/dm³**
^{4}[Pmol]	*275*	*1.8*	^{4}[Ibu]	*375*	*1.8*	^{70m}[Wfn]	*10*	*30*
^{2}[Pmol]	*137*	*0.9*	^{2}[Ibu]	*187*	*0.9*	^{54m}[Wfn]	*7.5*	*25*
^{1}[Pmol]	*69*	*0.45*	^{1}[Ibu]	*94*	*0.45*	^{35m}[Wfn]	*5*	*16*
$^{0.5}$[Pmol]	*34*	*0.22*	$^{0.5}$[Ibu]	*47*	*0.22*	^{18m}[Wfn]	*2.5*	*8*
$^{0.25}$[Pmol]	*17*	*0.11*	$^{0.25}$[Ibu]	*23*	*0.11*	^{7m}[Wfn]	*1*	*3*
$^{0.12}$[Pmol]	*9*	*0.06*	$^{0.12}$[Ibu]	*12*	*0.06*	$^{3.5m}$[Wfn]	*0.5*	*1.5*
$^{0.06}$[Pmol]	*1*	*0.03*	$^{0.06}$[Ibu]	*6*	*0.03*	$^{1.8m}$[Wfn]	*0.25*	*0.8*
$^{0.03}$[Pmol]	*2*	*0.015*	$^{0.03}$[Ibu]	*3*	*0.015*	$^{0.7m}$ [Wfn]	*0.1*	*0.3*
$^{0.015}$[Pmol]	*1*	*0.007*	$^{0.015}$[Ibu]	*1.5*	*0.007*	^{0}[Wfn]	*0*	*0*
^{0}[Pmol]	*0*	*0*	^{0}[Ibu]	*0*	*0*			

6.3.2 IR Absorption Spectroscopy

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^{-1} in the spectral region 400–4000 cm^{-1}.

6.3.3. 2D-IR Spectroscopy

2D-IR spectra were recorded using the LIFEtime laser spectrometer[52,53] at the Central Laser Facility using the Fourier transform 2D-IR method employing pulse shaper to deliver the pump pulse pair, as described in Chapter 2. Mid-IR pulses with a temporal duration of ~200 fs; a central frequency of 1650 cm^{-1} and a bandwidth of ~100 cm^{-1} were obtained.

2D-IR spectra for each sample were recorded at a waiting time (T_w) of 250 fs and 5 ps. The method used for setting up the sample path length as described in Chapter 3 is used in this chapter. However as small variations in the thickness can incur large variations in the 2D-IR signal, the method described in Chapter 5 to normalise the dataset has been applied prior to analysis. Each sample was measured in triplicate.

6.4. Results and Discussion

All 2D-IR spectral processing and analysis was carried out using a custom made script on the statistical analysis software programme, R.[54] As in previous chapters, prior to the analysis described in the text a 2nd order polynomial baseline subtraction was performed.

6.4.1. IR and 2D-IR Spectra of Paracetamol, Ibuprofen and Warfarin

Previous studies have shown that the binding of drugs to serum albumin affects the molecular conformation of serum albumin which results from a change in the α-helix content of the protein.[55] As the amide I band is used to determine the proteins secondary structure it is important to initially assess if the drugs have vibrational frequencies that overlap with the amide I region (1620-1690 cm^{-1}). In order to compare the IR and 2D-IR spectra of paracetamol, ibuprofen and warfarin, each drug was dissolved independently in DMSO at a concentration of 20 mg/mL (see Table 6.1 for molar concentrations) as shown in Figure 6.2 and Figure 6.3 (d) which shows the FTIR spectrum of each drug between the region of 1100-1800 cm^{-1}.

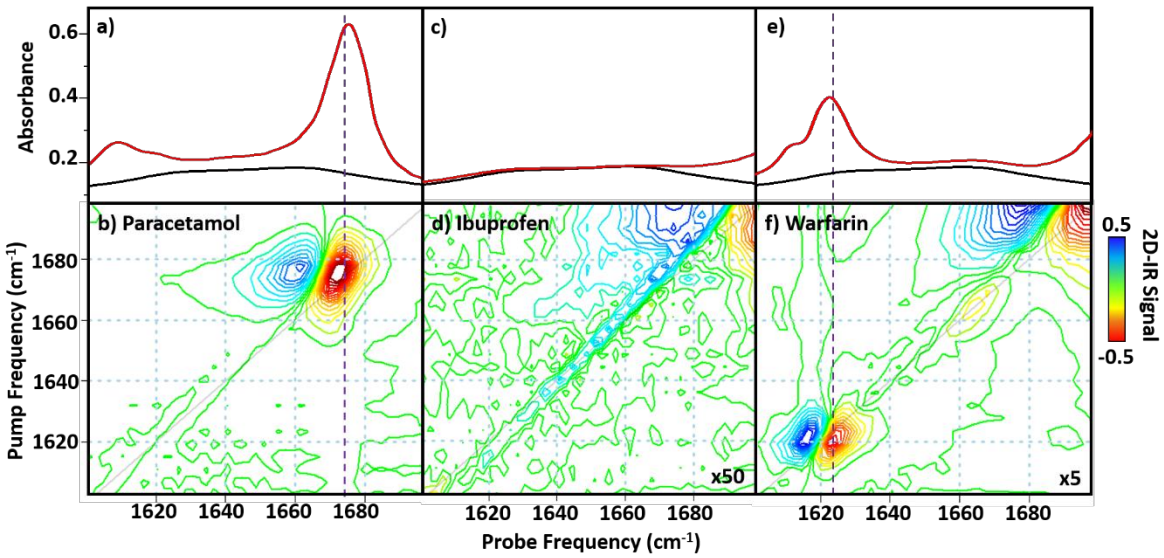

Figure 6.2: *The IR absorption and 2D-IR spectra of 20 mg/mL of paracetamol (a, b), ibuprofen (c, d) and warfarin (e,f). In the IR absorption spectrum, the drug is shown in red and the DMSO solvent only spectrum shown in black. Dashed purple vertical line in (a,b) at 1675 cm^{-1} indicates paracetamol peak in both figures. Dashed purple vertical line in (e,f) at 1623 cm^{-1} indicates warfarin peak position in both figures. All 2D-IR spectra are plotted on the same scale (see colour bar), d) and f) are magnified by factors of 50 and 5 respectively.*

This concentration was used in order to ensure signals from all 3 drugs are visible in the amide I region shown in both IR absorption and 2D-IR spectra. The spectra are not shown in water as of the three drugs chosen in this study only paracetamol has a sufficient enough solubility in water to be detectable, at around 15 mg/mL (99 mmol/dm^3) at room temperature.[56–58] Ibuprofen and warfarin each have a low solubility in water at around 21 and 17 µg/mL (0.01 and 0.005 mmol/dm^3) at room temperature, respectively.[59,60]

The IR absorption spectrum of paracetamol around the amide I region is shown along with their corresponding 2D-IR spectrum (Figure 6.2(a, b)). The IR absorption of paracetamol shows two distinct peaks at 1608 and 1675 cm^{-1}. The 2D-IR spectrum of paracetamol shows only the higher frequency peak situated at 1675 cm^{-1}. Theoretical calculations combined with FT-IR studies of solid paracetamol assigns peaks occurring at 1609 and 1625 cm^{-1} to the C=C stretching modes, and 1652 cm^{-1} to the C=O stretching mode of paracetamol, of which the molecular structures is shown in Figure 6.2(a).[61] It is likely that the peaks observed in Figure 6.2 at 1610 and 1675 cm^{-1} are due to the C=C and C=O stretching vibration. Due to hydrogen

bonding in water solvents these peaks may be shifted to lower wavenumbers when compared to that of the DMSO solvent due to the different cooperative interactions displayed by the two solvents.[62]

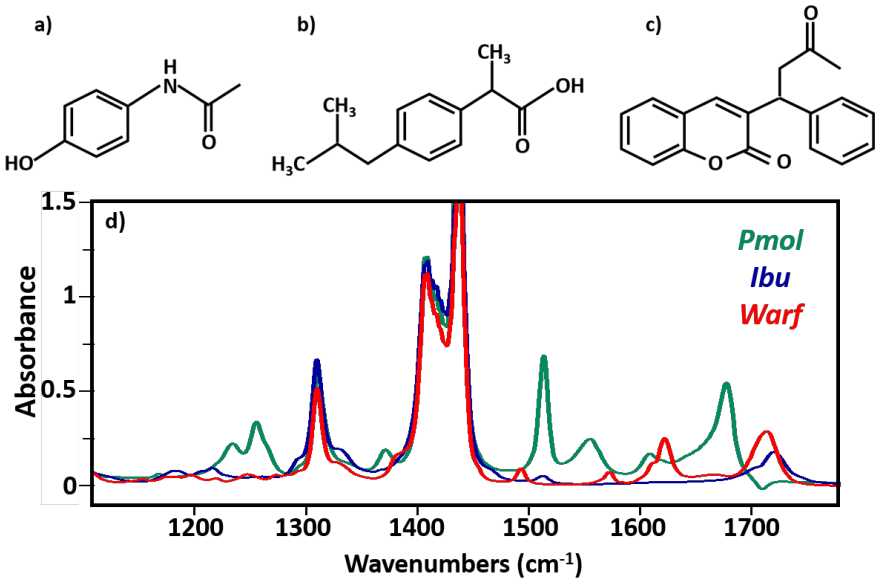

Figure 6.3: *Molecular structures of a) paracetamol, b) ibuprofen and c) warfarin and their respective FTIR spectra d) in the region of 1100-1800 cm^{-1}, where paracetamol is shown in green, ibuprofen in blue and warfarin in red.*

It is possible to infer the potential signal that would arise in the serum study using the concentration and sample thickness of the results obtained here. As the 2D-IR signal is linear with both concentration and sample thickness, a decrease in the path length by a factor of 10 will result in the decrease of the 2D-IR signal by the same factor. Furthermore the concentration used is ~70 times greater than the highest paracetamol concentration (275 µg/mL) used in the molar ratio study and so a total reduction of the signal by ~700 times would be anticipated. However, as the drug is anticipated to bind to the serum albumin the proportion of unbound drug in the system will be less than that used to infer the signal produced, and so it is unlikely that this peak will be distinguishable in the 2D-IR spectra dataset at the concentrations studied.

Infrared absorption and 2D-IR spectra of ibuprofen (Figure 6.2(c,d) and warfarin (Figure 6.2(e,f) again at concentrations of 20 mg/mL in DMSO are also shown. Within the amide I region of interest, there are no peaks that arise from the ibuprofen molecule, however the

shoulder of a higher frequency peak (1717 cm^{-1}) can be seen, DFT calculations allow assignment of this peak to the C=O stretch of the ibuprofen molecule.[63] The FTIR and 2D-IR spectra of warfarin display a peak at 1622 cm^{-1} which is tentatively assigned to the C-C stretching of the phenyl ring modes and the shoulder of a peak at 1712 cm^{-1} belonging to the C=O stretching vibrations can also be seen.[64]

In H_2O, these peaks may be shifted and overlap with the amide I region however, similar to paracetamol, the concentrations used for the serum albumin binding study are considerably smaller than that shown in DMSO. The highest concentration of warfarin and ibuprofen used in the serum binding study are 10 µg/mL (0.03 mmol/dm^3) and 375 mg/mL (1.8 mmol/dm^3), considerably smaller to that studied in DMSO (65 and 97 mmol/dm^3); this combined with the reduction of sample thickness to 2.5 µm will unlikely result in a signal strong enough to be detected for any free drug in the system.

6.4.2. Detection of Drug Binding using 2D-IR Spectroscopy

2D-IR spectra of serum samples spiked with different molar ratios of paracetamol are shown in Figure 6.4. As described in Chapter 3, the two dominant negative features (red) in the serum spectrum are assigned to the $v = 0 - 1$ transitions of the albumin (1656 cm^{-1}) and globulin (1639 cm^{-1}) fractions, as shown by the grey arrows in Figure 6.4(a). The $v = 1 - 2$ modes are shifted to lower wavenumbers due to the anharmonicity of the vibrations and is shown as a positive feature (blue). The addition of paracetamol leads to an increase in the signal intensity of the albumin (1656 cm^{-1}) amide I contribution, which is shown by the more intense red colour of the $v = 0 - 1$ feature, no additional peaks are noted.

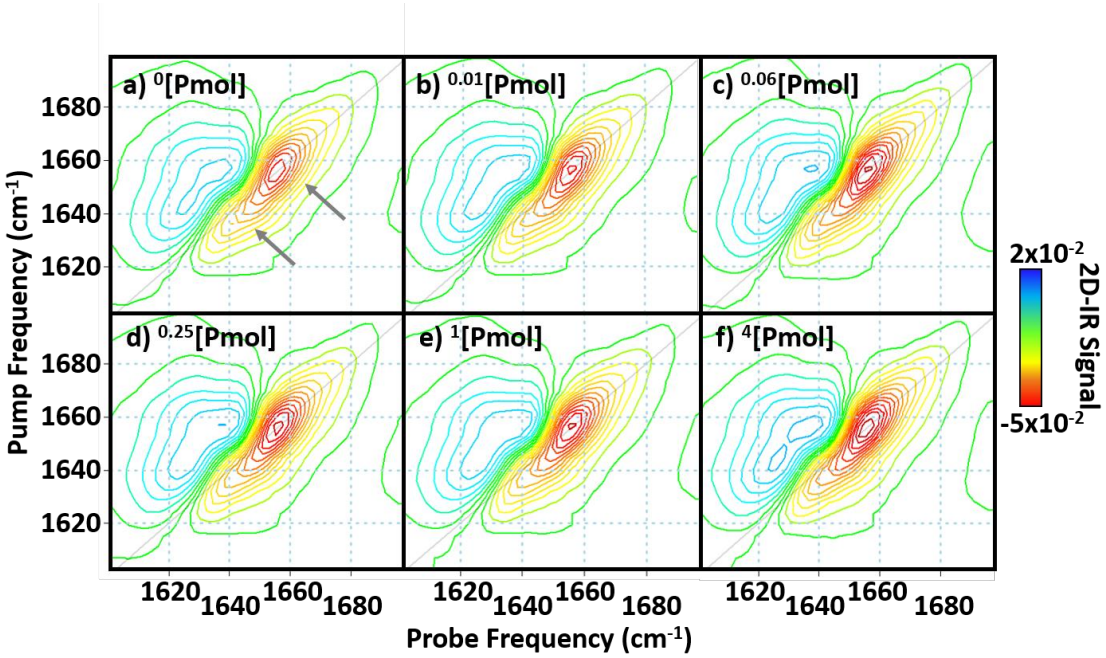

Figure 6.4: *2D-IR spectra of serum spiked with paracetamol in the following molar ratios of paracetamol and albumin: a) ^{0}[Pmol], b) $^{0.01}$[Pmol], c) $^{0.06}$[Pmol], d) $^{0.25}$[Pmol], e) ^{1}[Pmol], f) ^{4}[Pmol]. All spectra have been normalised using method described and plotted on the same scale, see colour bar. Grey arrows show the albumin and globulin responses, at 1656 and 1639 cm^{-1} respectively (see chapter 3).*

Difference spectral analysis was performed by subtracting the serum only ([^{0}Pmol]) spectrum from the spectrum of each paracetamol containing sample in order to show clearly the differences observed, and are shown in Figure 6.5.

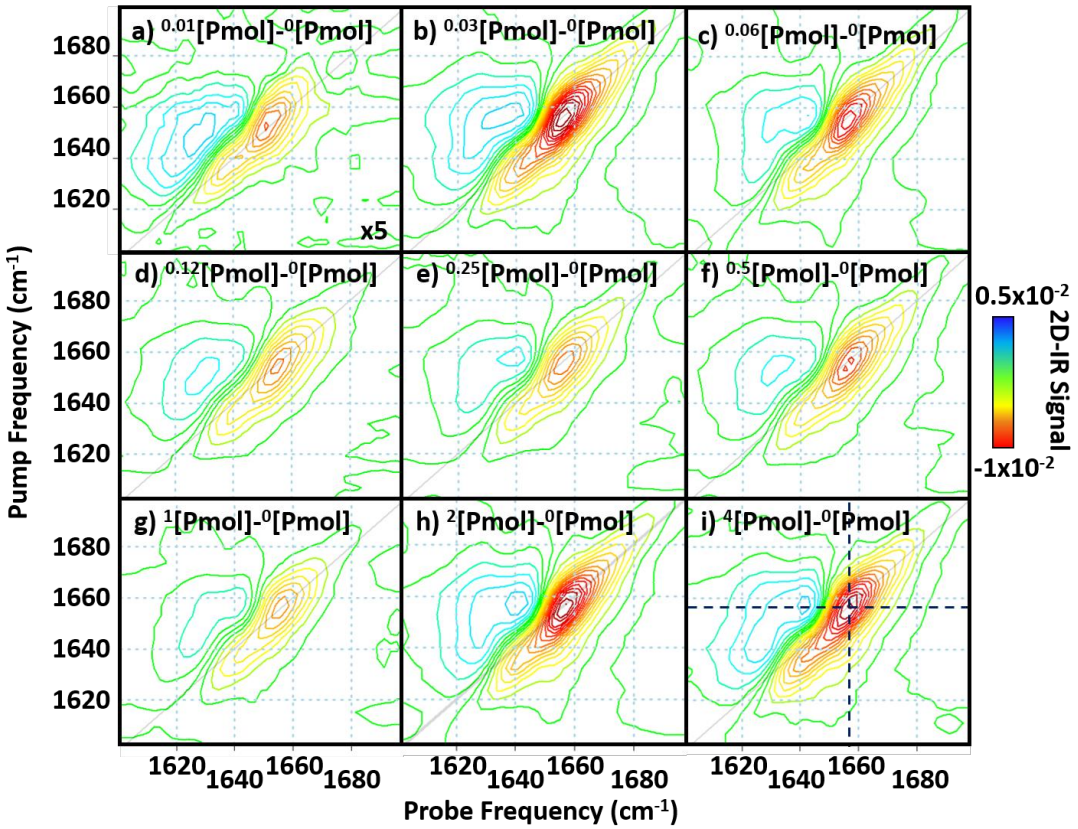

Figure 6.5: *2D-IR spectra following subtraction of serum only ^{0}[Pmol] spectra from those containing paracetamol: a) $^{0.01}$[Pmol], b) $^{0.03}$[Pmol], c) $^{0.06}$[Pmol], d) $^{0.12}$[Pmol], e) $^{0.25}$[Pmol], f) $^{0.5}$[Pmol], g) ^{1}[Pmol], h) ^{2}[Pmol], i) ^{4}[Pmol]. All spectra are plotted on the same scale, see colour bar, and have been magnified by a factor of 5 in the case of a). Crosshairs in i) show position with the largest difference.*

A negative peak at 1657 cm^{-1} is present at every quantity of paracetamol. The addition of the smallest concentration of paracetamol, as shown in Figure 6.5(a), has the smallest change in signal, as denoted by the faint signal after 5x magnification. The difference spectra of $^{0.03}$[Pmol], ^{2}[Pmol] and ^{4}[Pmol] (Figure 6.5(b, h, i) display the largest changes. The amplitude of the diagonal peak at 1657 cm^{-1} (Figure 6.5(i) purple dashed lines) as a function of paracetamol concentration is shown in Figure 6.6, and it is noted that with the addition of paracetamol to the system the magnitude of the signal at 1657 cm^{-1} increases. In order to determine whether these changes are large enough to be detected, the RMS of the noise level of the 2D-IR spectrum is calculated using the same method as described in Chapter 4, where an area free from peaks is utilised. This results in the RMS noise value of 4 x 10^{-4} to be established. The changes observed here all large enough to be detected as they are all greater

than twice this noise level, where making the conservative assumption that only peaks exceeding twice this value can be observed is made.

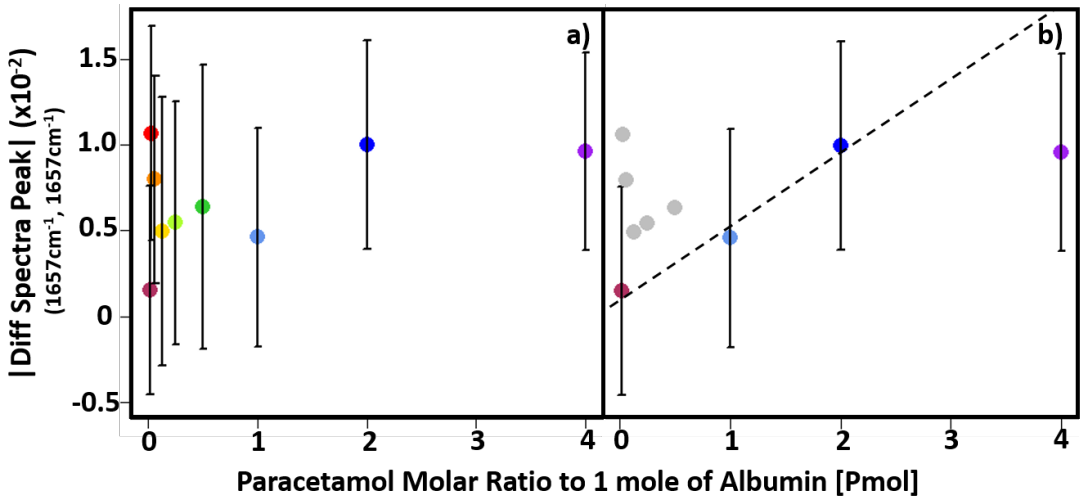

Figure 6.6: *a) The magnitude of the difference spectra peak at a pump and probe frequency of 1657 cm^{-1}, plotted as a function of paracetamol concentration for the full range of concentrations examined b) with $^{0.03}$[Pmol] – $^{0.5}$[Pmol] greyed out to observe trend from $^{0.01}$[Pmol], ^{1}[Pmol], ^{2}[Pmol] and ^{4}[Pmol]. Error bars show 1σ variation from triplicate measurement.*

However the large variation within each concentration due to triplicate measurements, of which 1 standard deviation is plotted as the error bars as shown in Figure 6.6, overlap with one another significantly. The $^{0.01}$[Pmol] (maroon) sample has the smallest change upon paracetamol addition, this is anticipated as only 1% of albumin molecules have a paracetamol bound at this stage. By removing the smaller ratios between $^{0.03}$[Pmol] – $^{0.5}$[Pmol], as seen in Figure 6.6(b), a gradual linear increase in the peak amplitude is observed from $^{0.01}$[Pmol], ^{1}[Pmol] and ^{2}[Pmol] (maroon – light blue – blue). Little change is seen between the ^{2}[Pmol] and ^{4}[Pmol].

The addition of 2 or more paracetamol molecules per albumin to the system does not change the signal observed. Fluorescence spectroscopy studies of the binding constant (K_a) of paracetamol with serum albumin produce a K_a = 1.3 x 10^4 M^{-1} and predicts 2 binding sites for paracetamol within the serum albumin molecule[37] and so as the concentration of paracetamol molecules increases above ^{2}[Pmol] no more binding should occur.[21,38] The results of the 2D-IR difference spectra also displays this however the large variation at each

concentration means that this analysis technique does not produce conclusive results. Different dependencies, including a square root dependence were also considered to help improve the trend however no notable differences in the trend was observed.

Difference spectral analysis was also carried out for both ibuprofen and warfarin datasets, the difference spectra are shown in Figures 6.7 and 6.8, respectively.

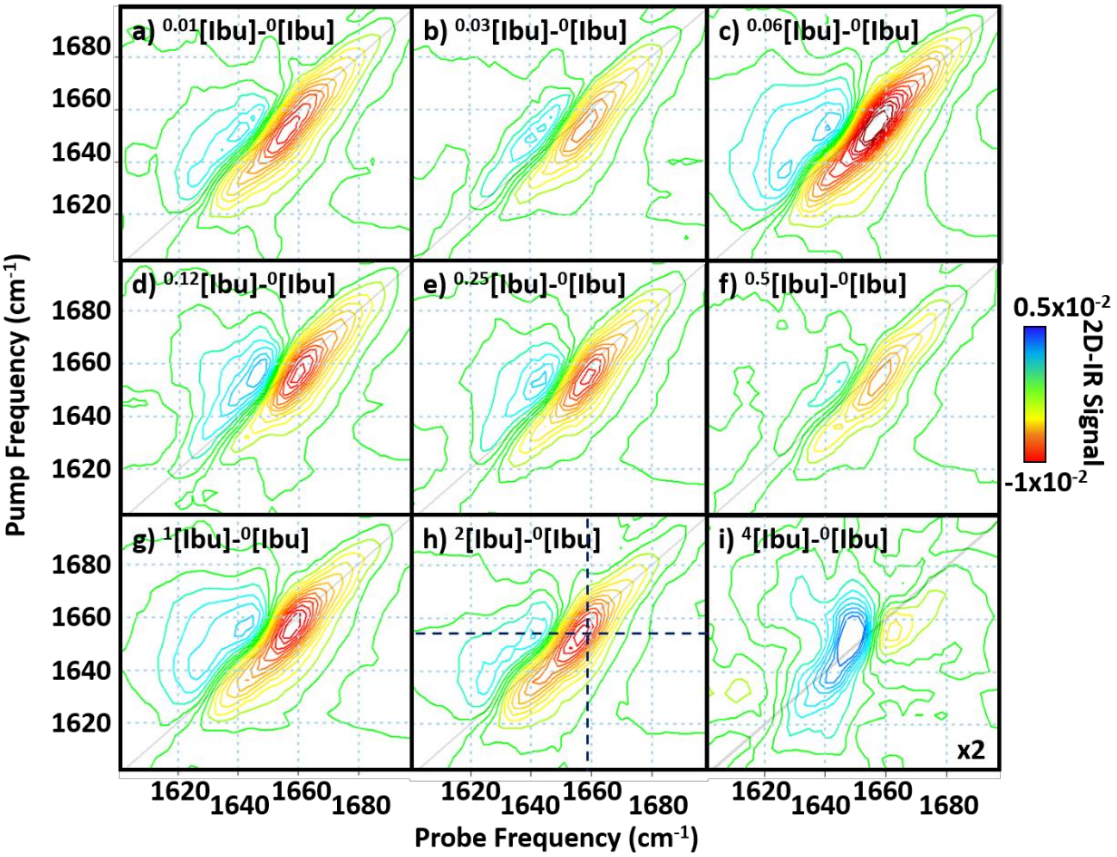

Figure 6.7: *2D-IR spectra of a) $^{0.01}$[Ibu], b) $^{0.03}$[Ibu], c) $^{0.06}$[Ibu], d) $^{0.12}$[Ibu], e) $^{0.25}$[Ibu] f) $^{0.5}$[Ibu], g) ^{1}[Ibu], h) ^{2}[Ibu], i) ^{4}[Ibu], after subtraction of serum only ^{0}[Ibu] spectra. All spectra are plotted on the same scale, see colour bar, and have been magnified by a factor of 2 in the case of i). Crosshairs in h) show position with the largest difference.*

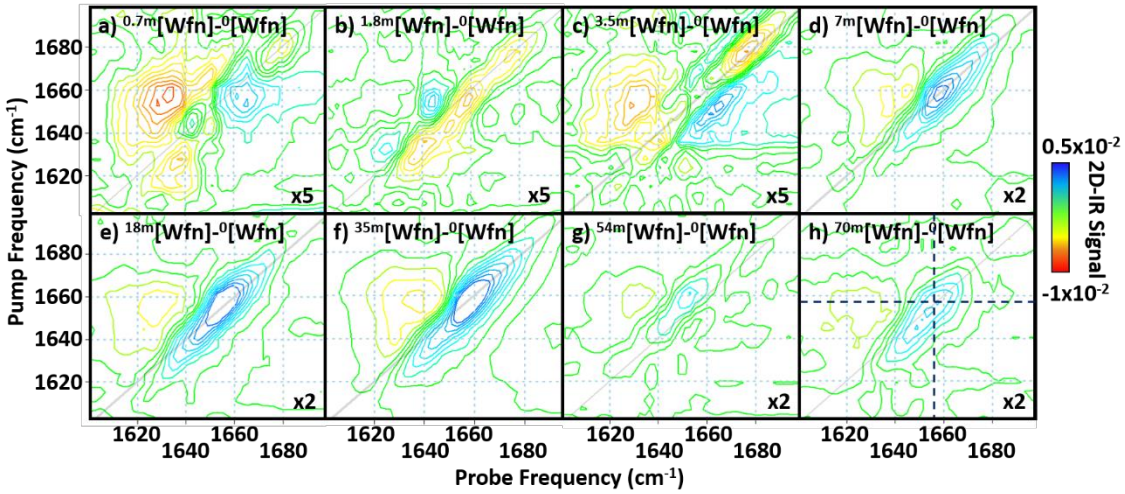

Figure 6.8: *2D-IR spectra of a) $^{0.7m}$[Wfn], b) $^{1.8m}$[Wfn], c) $^{3.5m}$[Wfn], d) ^{7m}[Wfn], e) ^{18m}[Wfn], f) ^{35m}[Wfn], g) ^{54m}[Wfn], h) ^{70m}[Wfn], after subtraction of serum only ^{0}[Wfn] spectra. All spectra are plotted on the same scale, see colour bar, and have been magnified by a factor of 2 or 5. Crosshairs in h) show position with the largest difference.*

The difference spectral analysis does not show a clear correlation with increasing drug concentration for either ibuprofen or warfarin. The ibuprofen difference spectral amplitude of the negative peak at position 1656 and 1654 cm^{-1} on the pump and probe frequency axes (Figure 6.7(h) purple dashed lines) monitored as a function of ibuprofen concentration is shown in Figure 6.9(a).

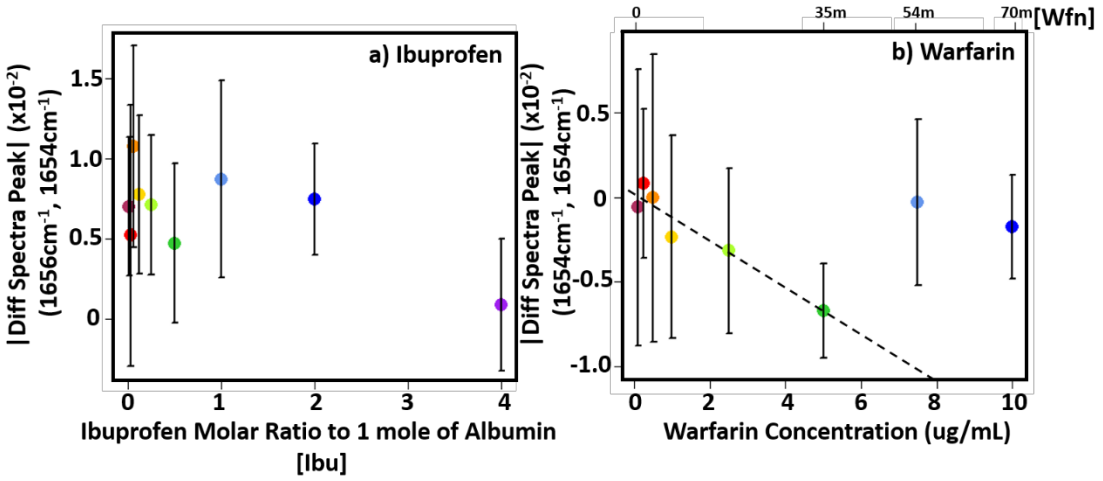

Figure 6.9: *Magnitude of the difference spectra peak plotted as a function of a) ibuprofen at the pump frequency 1656 cm^{-1} and probe frequency of 1654 cm^{-1}and b) warfarin concentration pump and probe frequency of 1654 cm^{-1}. Error bars show 1σ from triplicate measurement.*

The smallest change is suggested by the addition of the highest ibuprofen concentration, ^{4}[Ibu] however this seems unlikely. The warfarin difference spectra results are shown in Figure 6.9(b) at the peak position as defined in Figure 6.8(h). A linear decrease is seen from $^{0.7m}$[Wfn] (maroon) to ^{35m}[Wfn] (dark green) however the two greatest contributions (^{54m}[Wfn] and ^{70m}[Wfn]) indicate a close to zero difference with warfarin addition, again seeming unlikely. While the linear decrease observed between $^{0.7m}$[Wfn] to ^{35m}[Wfn] would be seemingly likely as studies have documented a decrease in α-helix content of albumin upon binding of warfarin, as the molecule undergoes the NB transition,[55,65] the two highest contributions,^{54m}[Wfn] and ^{70m}[Wfn], do not fit with this theory. Perhaps it is not possible to see the changes in warfarin as at most only 7% of the albumin molecules are bound to the drug (^{70m}[Wfn] sample). While the trend may be physically likely, the variation in signals from each concentration produces large overlapping error bars, and the detection of the changes observed from the ratios studied should be made with caution.

However, comparisons of the difference spectra produced from each of the three datasets shows that each drug is consistent in their change to the 2D-IR spectral lineshapes. Warfarin shows a decrease in signal amplitude (Figure 6.8, blue diagonal peak) upon binding and both paracetamol and ibuprofen show an increase (Figure 6.5 and 6.7, red diagonal peak). Furthermore the ibuprofen difference spectra peak is more elongated along the diagonal compared to paracetamol's more circular peak shape. This suggests that each drug upon binding to the albumin results in a slightly different structural change however it is difficult to determine if the changes observed are real.

As mentioned previously, studies have shown that upon binding the albumin molecule undergoes a conformational change from the N to the B isomer. This transition has been documented as a result of the change in the secondary structure of the protein. However the potential overlap of the drug vibrational modes with the protein amide I must be addressed. The amplitude of the changes observed in the difference spectra are considerably larger than those suggested for paracetamol at the correct concentration and thickness Section 6.4.1. The magnitude of the 2D-IR signal amplitude for paracetamol at the highest concentration used was predicted to be around 1/700th of the paracetamol in DMSO signal, and assuming binding has occurred this amplitude will be lower. This allows the assumption to made that if the drugs were free in solution that the signals due to the drug would be too small to detect (the RMS of the noise floor was calculated to be 4 x10^{-4}). As the difference spectra show

changes larger than twice the RMS noise level and in the amide I, particularly in the region belonging to α-helix signals, it is likely that the observations made are due to a secondary structure change due to ligand binding with the albumin molecule.

6.4.3. Quantification of Secondary Structure Change

The intensities of 2D-IR signals are sensitive to the transition dipole moment of the molecule which is dependent on the proteins secondary structure due to coupling throughout the entire helix/sheet. Calculation of the transition dipole moment due to small changes in a protein molecules secondary structure have been documented using 2D-IR and this method is used here to detect changes in the protein structure for the serum and paracetamol dataset.

This method combines both the linear absorption and 2D-IR diagonal spectra to extract a 'spectrum' of the transition dipole strength of the sample. The dependences of the signal intensity on the transition dipole moment of the two techniques differs, where linear absorption methods scales to μ^2 and 2D-IR $\propto \mu^4$, which allows changes in the relative amplitudes of the two signals to be seen if the degree of vibrational coupling changes within the molecule. The integrated area of an IR absorption spectrum is conserved when the coupling changes, and so for a highly coupled system with strong transition dipole strengths, like an α-helix, produces the same integrated intensity as an uncoupled random coil. However as the 2D-IR spectrum is more sensitive to the transition dipole strength, integration of the spectrum area is no longer conserved as the coupling changes.[9–11] The absolute transition dipole moment can be calculated however this requires an accurate calibration.[9]

Here, a modified approach has been used to extract the changes observed in the proteins secondary structure relative to the serum only sample in the form of a frequency-dependent ratio of the 2D- IR diagonal and IR absorption signals to calculate a spectrum of the relative transition dipole strength, $d(\omega)_{rel}$;

$$d(\omega)_{rel} = \frac{2DIR(\omega)_P}{FTIR(\omega)_P} \frac{FTIR\ OD_s}{2DIR\ \Delta OD_s} \tag{6.1}$$

where, $2DIR(\omega)_{PC}$ and $FTIR(\omega)_{PC}$ are the absolute spectra of the 2D-IR diagonal and IR absorption at paracetamol concentration, P (Figure 6.12). $FTIR\ OD_s$ and $2DIR\ \Delta OD_s$ are the absolute maxima for both the serum only IR absorption and 2D-IR spectra.[11]

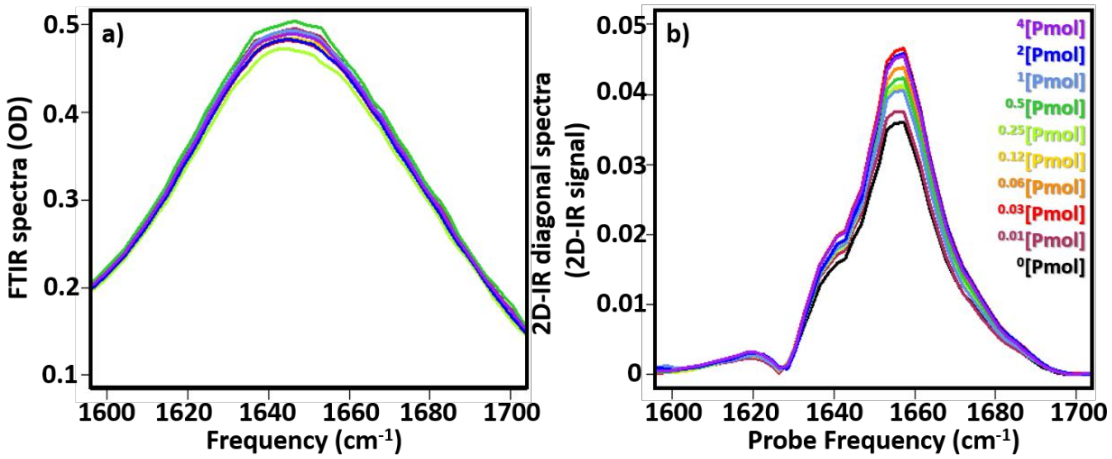

Figure 6.10: a) FTIR and b) 2D-IR diagonal spectra are shown for the ^{n}[Pmol] dataset. Each colour is indicative of paracetamol concentration, as shown in Figures 6.8 and 6.11. FTIR spectra have been normalised to the OH bending mode at 2130 cm^{-1}.

In the case of a change in vibrational coupling, a change in $d(\omega)_{rel}$ will be seen. Calculation of the spectra of the relative transition dipole strength, $d(\omega)_{rel}$, in Figure 6.11(a)) shows an increase in the transition dipole strength for every concentration of paracetamol (colour) compared to the serum only spectrum (black). Calculation of the area under the curve (AUC) between 1628-1700 cm^{-1} results in the production of similar trends to the difference spectra (Figure 6.11 (b,c)).

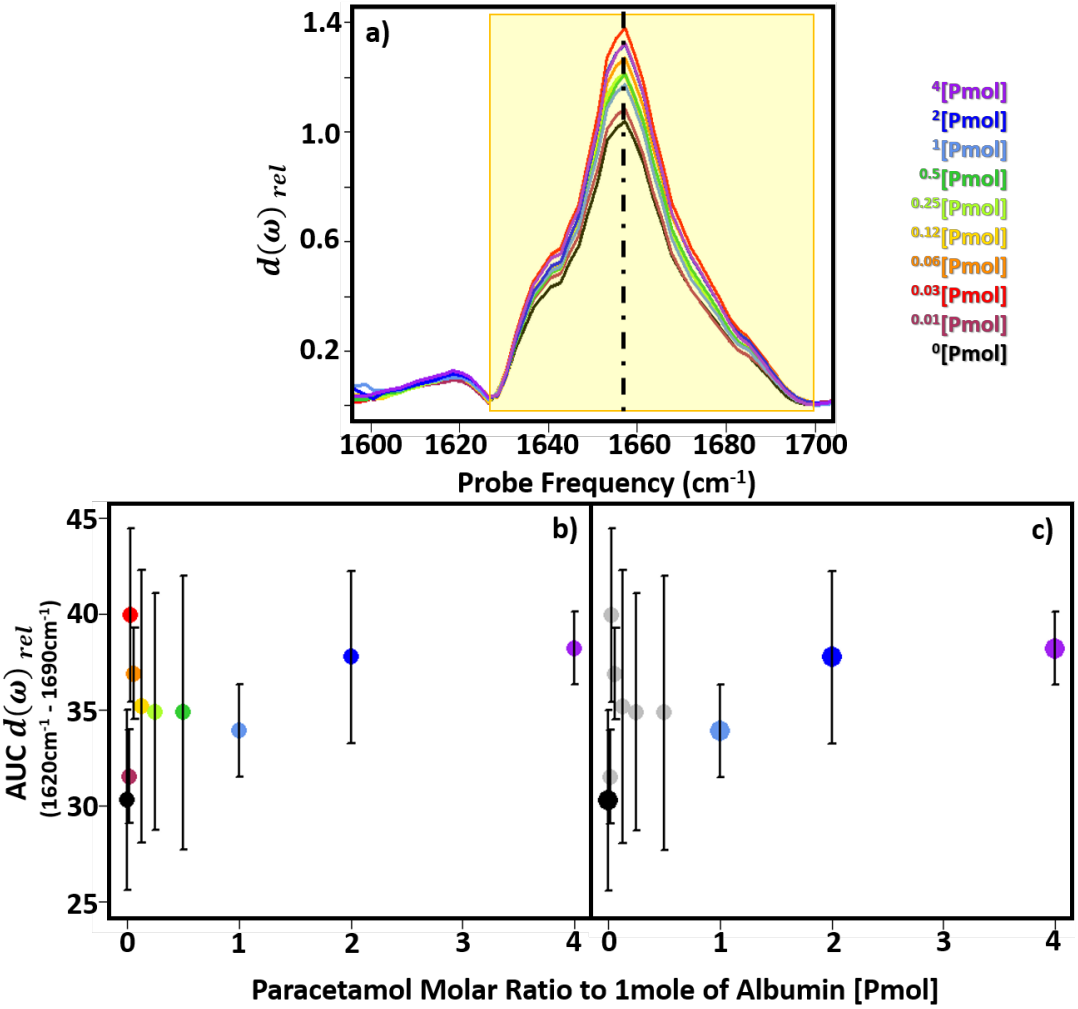

Figure 6.11: *The average transition dipole moment calculated for each paracetamol concentration. Dashed vertical line shows position of greatest change, 1657 cm^{-1}. Yellow area indicates area used to integrate for results in b). b) The area under the curve (AUC) of the transition dipole moment in (a) between 1628 and 1700 cm^{-1} for each paracetamol concentration c) with smaller concentrations greyed out for clarity showing ^{0}[Pmol], ^{1}[Pmol], ^{2}[Pmol] and ^{4}[Pmol] molar ratios. Error bars show 1σ from triplicate measurement.*

As an α-helix typically appears on infrared spectra at around 1655 cm^{-1}, this peak at 1657 cm^{-1} (Figure 6.11(a)) is an indication of a gain of the vibrational coupling associated with an α-helix. However the variation within each concentration due to triplicate measurements, of which 1 standard deviation is plotted as the error bars as shown in Figure 6.11, again overlap with one another. At this stage of the development of 2D-IR for analysis of biofluids in water it would be unlikely that the sensitivity of 2D-IR is low enough to detect a change of drug

binding to only 1% (0.01:1) of molecules. It is possible that the changes observed here at low concentrations are within the noise and not assignable to real changes within the protein.

The transition dipole analysis was also applied to ibuprofen and warfarin datasets and can be seen in Figure 6.12. Again, similar patterns were found for both ibuprofen and warfarin when compared to the difference spectra analysis.

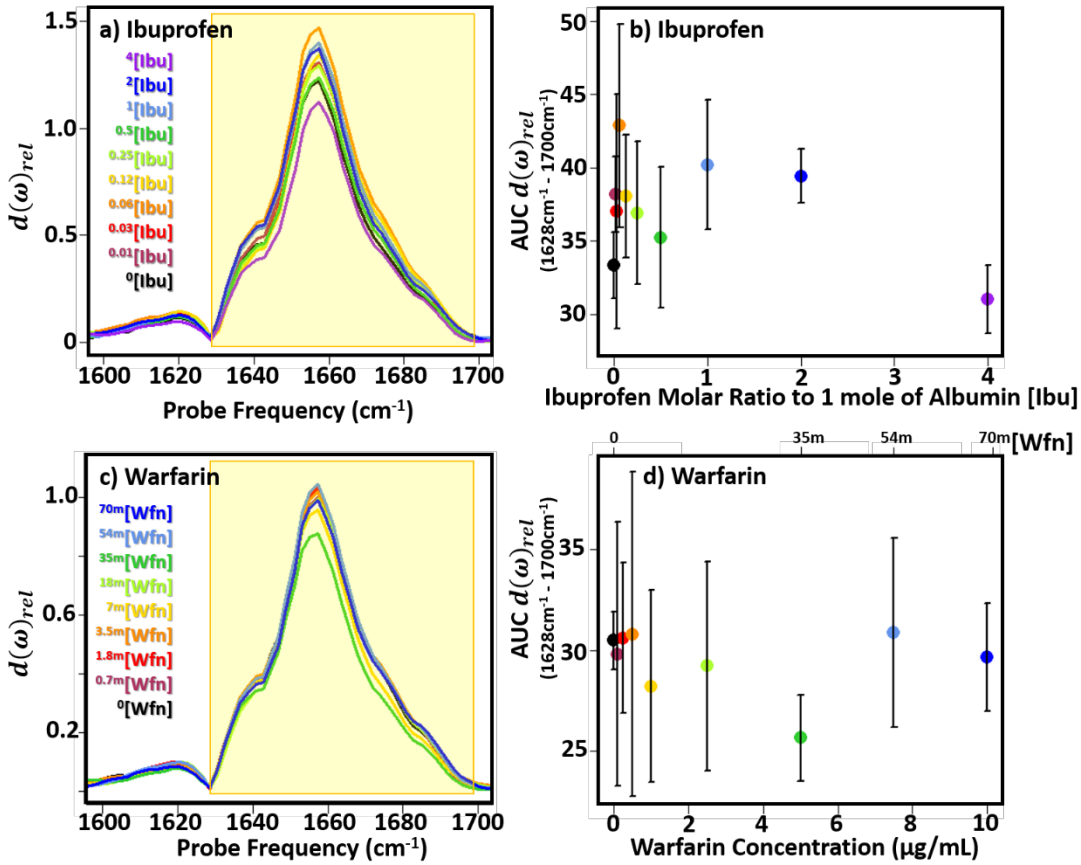

Figure 6.12: *The average transition dipole moment calculated for each ibuprofen (a,b) and warfarin (c,d) concentration. (b,d) The area under the curve (AUC) of the transition dipole moment in (yellow box in a and c) between 1628 and 1700 cm^{-1} for each ibuprofen and warfarin concentration. Error bars show 1σ from triplicate measurement.*

Evidence can be seen of a change in the secondary structure of the protein due to drug binding using 2D-IR. When paracetamol is fully bound (occupying both binding sites) to albumin there is a strong increase in the both the signal amplitude from the difference spectra and in the relative transition dipole moment. Both of these methods are consistent in producing results with an increase at 1657 cm^{-1} and so this change could be indicative of

an increase in the α-helix content of the albumin upon paracetamol binding. Both methods also show that upon the addition of paracetamol above a 2:1 molar ratio, that variations in the protein are no longer changing – consistent with other works documenting 2 binding sites for paracetamol in the albumin molecule.[21,37,38,66]

Previous studies have shown conflicting results as to whether the albumin gains or loses α-helix upon paracetamol binding however the number of albumin binding sites for paracetamol is found to be consistently 2.[21,37,38,66] Investigations of the effects of specific variables on the binding of drugs to serum albumins have been made, however it is important to note that initial parameters like solvent buffers and pH values in which proteins and drugs are dissolved in are not consistent across different studies, and different buffers and pH values have been shown to have an effect on drug binding to serum albumin.[13,21] Many studies have been carried out investigating the effects of other variables on the binding of drugs to serum albumins, including solvent buffer, pH, presence of glucose, competing drugs and of fatty acids however the quantification of albumin secondary structure change in serum *in vivo*, taking into account the overall effect of many competing factors, have not been made.[13,21,31,33,34,37,38,66–70]

6.4.4. Effects of Serum Constituents on Serum Albumin Binding

Many studies have shown that by binding a drug to the albumin, in either drug binding site I or II, causes the protein molecule to undergo the neutral to basic transition, resulting in the open ellipsoid structure and this change in structure is a result of a reduction in α-helix content of the serum albumin.[13,34,65,69,71,72]

The studies shown in this chapter may however have a different starting conformation to those documenting this change. The form in which the serum albumin presents prior to the addition of any of the three drugs has not been investigated and as albumin is a carrier protein and present in serum among fatty acids, metals, electrolytes, amino acids, and many other proteins it is probable to assume that some of these components may already be bound to the albumin affecting the albumin structure. As mentioned previously, the addition of fatty acids to serum albumin induces the NB transition of the protein, meaning that it is likely that the albumin is not in the N (heart) conformation prior to albumin binding. Competitive binding must also be considered, as it is possible that the drugs are competing against the other serum constituents to bind to the serum albumin which in turn may also

leave the number of 'free' (unbound) serum albumins in solution to be few.[33,35] Nevertheless what can be shown in particular for the paracetamol dataset is the change observed due to the addition of the drug, and the two analysis methods used suggest that an increase in coupling due to an increase in the α-helix content is seen with increasing paracetamol concentration, until no more binding occurs.

In order to attempt to understand the changes associated with the serum albumin upon drug binding, a sample dataset of BSA and paracetamol using the same serum albumin concentration and the same molar ratios was investigated and the difference spectra are shown in Figure 6.13. The albumin concentration in serum was determined via the Bromocresol Green test[73] (conducted by University of Glasgow School of Veterinary Medicine) which revealed the albumin content to be 30 mg/mL.

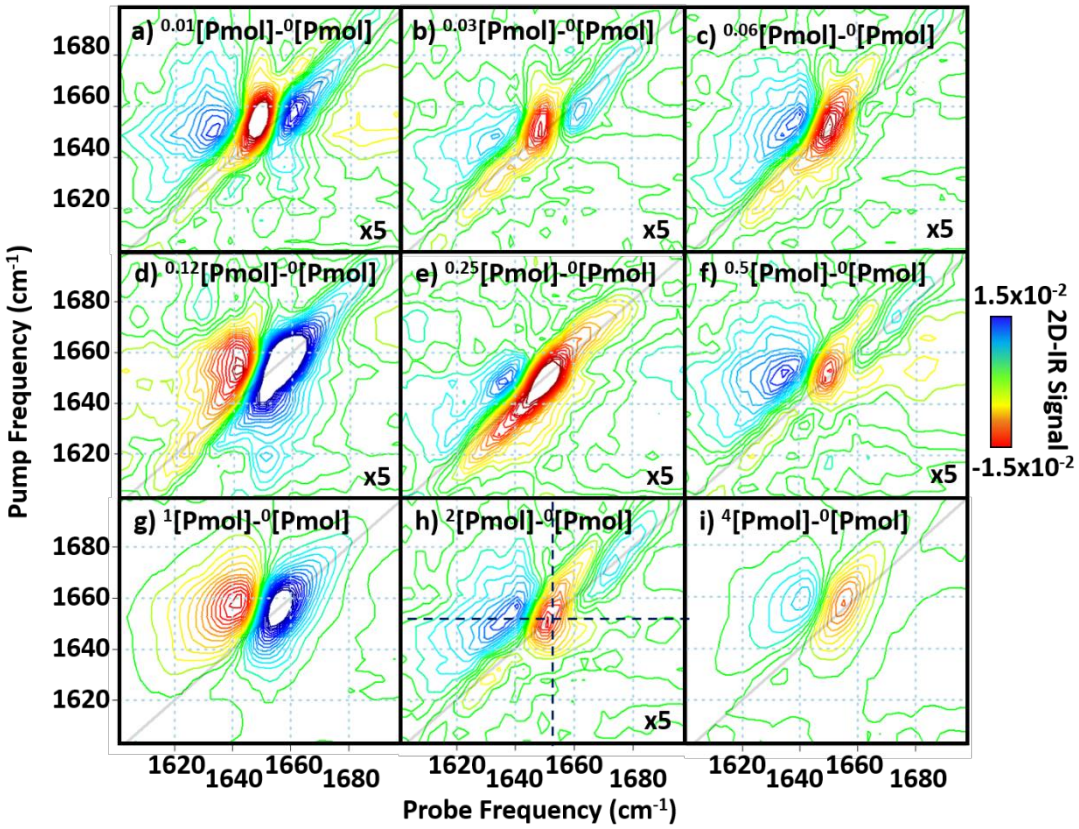

Figure 6.13: *2D-IR difference spectra of the BSA and paracetamol datset following subtraction of BSA only ^{0}[Pmol] spectra from those containing paracetamol: a) $^{0.01}$[Pmol], b) $^{0.03}$[Pmol], c) $^{0.06}$[Pmol], d) $^{0.12}$[Pmol], e) $^{0.25}$[Pmol], f) $^{0.5}$[Pmol], g) ^{1}[Pmol], h) ^{2}[Pmol], i) ^{4}[Pmol]. All spectra are plotted on the same scale, see colour bar, and some have been magnified by a factor of 5, as shown. Crosshairs in h) show position with a consistent difference throughout the dataset.*

The difference spectra peak position and sign varies, however in 6 out of 9 difference spectra a peak at 1650 cm^{-1} persists, as highlighted by the crosshairs in Figure 6.13(h). Results for both the difference spectra and transition dipole analysis for the BSA dataset are compared with that of serum and are shown in Figure 6.14.

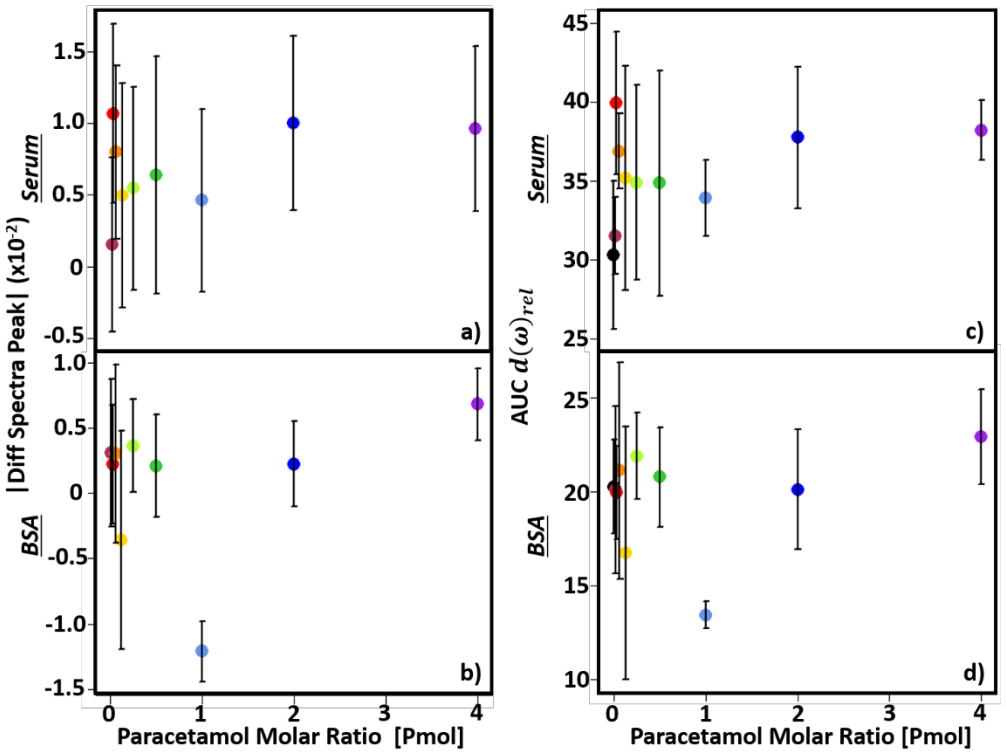

Figure 6.14: *Comparison of the difference spectra peak (a,b) and transition dipole moment (c,d) results from paracetamol in both serum (a,c) and BSA (b,c). Error bars show 1σ variation from triplicate measurement.*

The BSA obtained from Sigma has undergone charcoal treatment and organic solvent precipitation in order to remove the fatty acids, thus providing a close to 'pure' BSA.[74] However as mentioned, studies have shown that the presence of bound fatty acids to serum albumin can change its affinity for specific drugs.[29,36,40] The two analysis methods used, difference spectra and transition dipole strength as before, produce similar trends for the BSA and paracetamol dataset. The BSA difference spectra peak amplitude (Figure 6.14(b)) as a function of paracetamol concentration fluctuates around 0, where the $^{0.12}$[Pmol] - ^{0}[Pmol] and particularly the ^{1}[Pmol] - ^{0}[Pmol] spectra show large decreases in signal. The ^{4}[Pmol] - ^{0}[Pmol] spectra yields the largest increase in peak height, whereas the serum dataset sees a

consistent increase for all additions of paracetamol (Figure 6.14(a)). The differences between the two datasets could potentially be attributed to the wealth of additional molecules bound to albumin in the serum dataset, e.g. fatty acids.

As mentioned earlier, the addition of fatty acids to serum albumin results in a conformational change of the molecule from the N to the B isomer.[13,29,75] In order to determine whether there are any 2D-IR spectral differences between the known fatty acid free BSA and serum albumin in serum (both 30 mg/mL), their spectra and the difference of the two (Serum-BSA) is shown in Figure 6.15.

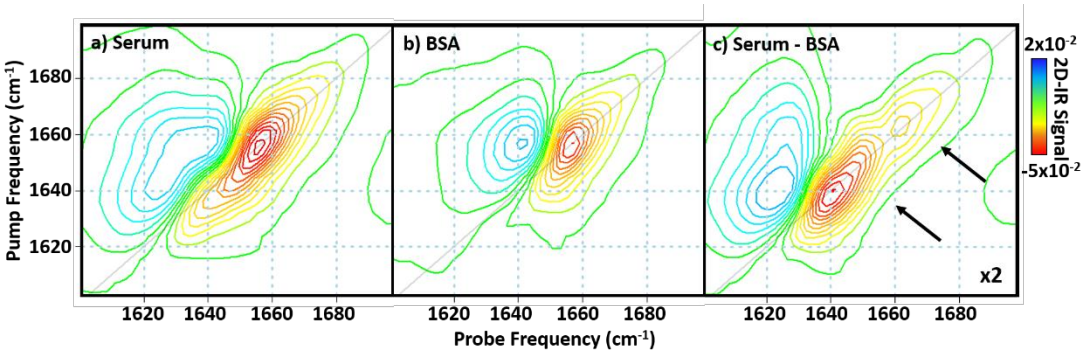

Figure 6.15: *The average 2D-IR spectra of a) serum, b) 30 mg/mL of BSA and c) subtraction of BSA from serum resulting in a 2D-IR difference spectra. All spectra are plotted on the same scale, see colour bar. Black arrows in (c) show residual peaks after BSA subtraction of serum.*

Both the serum and BSA spectra shown in Figure 6.15 are averages of a total of nine spectra taken throughout the experiment run (three for each drug dataset) in order to obtain the most reliable difference spectra possible. The serum – BSA difference spectra reveals two peaks on the 2D-IR diagonal. The lower frequency peak at 1640 cm^{-1} which has been assigned to the globulin fraction (Chapter 3) however there is a significant residual signal at around 1660 cm^{-1} (Figure 6.15(c)).

The remaining higher frequency signal left in the serum-BSA difference spectra suggests that there are contributions existing in the serum at 1660 cm^{-1} that are not present in the BSA sample for which there could be many reasons. The accumulation of many proteins/peptides containing α-helical structures could in fact be the reason for this residual signal as there are many other proteins present in the serum sample and is the most likely answer. However it could also suggest a change in isomer due to the addition of fatty acids and other serum molecules in the serum sample prior to the addition of drugs, meaning that the initial albumin

structure prior to drug binding may differ from the N conformation as previously stated. While this suggestion is likely, it is not clear if the two isomers would produce such a change in the 2D-IR spectrum and further investigations into the ability of 2D-IR to determine different albumin isomers would be needed to clarify this.

6.5. Conclusions

As demonstrated in previous chapters, 2D-IR spectroscopy can perform quantitative measurements of proteins in H_2O due to its ultrafast nature. In this chapter, the boundaries of 2D-IR spectroscopy have been explored to measure drug-protein binding in aqueous serum, increasing the physiological relevance of both drug-binding and 2D-IR studies.

The change in 2D-IR spectral lineshapes of serum albumin was found to be consistent for each drug as shown by the difference spectral analysis. Using FTIR spectra and 2D-IR spectrum diagonals, transition dipole moment analysis was performed producing trends similar to those obtained for the difference spectra for each drug analysed. Results indicate that the albumin molecule houses two binding sites for paracetamol, which is comparable to previous drug-binding studies, and in doing so the protein molecule increases its α-helix content which suggests a conformational change in the protein molecule. Studies of paracetamol-albumin binding in its natural serum environment have not been studied, presumably due to the complexity of serum, and so comparisons of the structural change due to paracetamol binding in serum cannot be made. Assessment of the differences between serum albumin and BSA in buffer via spectral analysis indicates there are contributions present in the serum that could affect the result of the serum albumin binding when compared to the BSA binding alone. However what isn't clear is whether this change is due to conformational changes in the protein due to binding of serum molecules (eg. fatty acids), and thus hinders the ability to compare the serum albumin binding results with those not taking into consideration other serum components. Developments of more advanced data analytics may improve the results obtained and although careful consideration has been taken in the set-up of the sample thickness and normalisation methods have been devised, other works to help standardise data collection protocols could also lead to improvement in data quality.

6.6. References

1. Kragh-Hansen, U. Molecular aspects of ligand binding to serum albumin. *Pharmacol. Rev.* **33,** 17–53 (1981).

2. Bos, O. J. M., Remijn, J. P. M., Fischer, M. J. E., Wilting, J. *et al.* Location and characterization of the suramin binding sites of human serum albumin. *Biochem. Pharmacol.* **40,** 1595–1599 (1990).

3. Sułkowska, A., Bojko, B., Równicka, J. & Sułkowski, W. Competition of drugs to serum albumin in combination therapy. *Biopolymers* **74,** 256–262 (2004).

4. Chasman, D. I. *Protein Structure, Determination, Analysis and Applications for Drug Discovery*. (Marcel Dekker, Inc., 2003).

5. Baumann, T., Hauf, M., Schildhauer, F., Eberl, K.B. *et al.* Site-resolved observation of vibrational energy transfer using a genetically encoded ultrafast heater. *Angew. Chemie - Int. Ed.* **58,** 2899–2903 (2019).

6. Johnson, P. J., Kozoil, K. L. & Hamm, P. Quantifying biomolecular recognition with site-specific 2D infrared probes. *J. Phys. Chem. Lett.* **8,** 2280–2284 (2017).

7. Baiz, C. R., Reppert, M. & Tokmakoff, A. Introduction to protein 2D IR spectroscopy. *Ultrafast Infrared Vib. Spectrosc.* 361–404 (2012).

8. Ganim, Z., Chung, H. S., Smith, A. W., Deflores, L. P. *et al.* Amide I two-dimensional infrared spectroscopy of proteins. *Acc. Chem. Res.* **41,** 432–441 (2008).

9. 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.* **137,** 184202 (2012).

10. 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* **119,** 14065–14075 (2015).

11. Minnes, L., Shaw, D. J., Cossins, B. P., Donaldson, P. M. *et al.* Quantifying secondary structure changes in calmodulin using 2D-IR spectroscopy. *Anal. Chem.* **89,** 10898–10906 (2017).

12. Shaw, D. J., Robb, K., Vetter, B. V., Tong, M. *et al.* Disruption of key NADH-binding pocket residues of the mycobacterium tuberculosis InhA affects DD-CoA binding ability. *Sci. Rep.* **7,** 1–7 (2017).

13. Fasano, M. *et al.* The extraordinary ligand binding properties of human serum albumin. *IUBMB Life* **57,** 787–796 (2005).

14. Sudlow, G., Birkett, D. J. & Wade, D. N. The characterization of two specific drug binding sites on human serum albumin. *Mol. Pharmacol.* **11,** 824–832 (1975).

15. Vusse, G. J. van der. Albumin as fatty acid transporter. *Drug Metab. Pharmacokinet.* **24,** 300–307 (2009).

16. Tortura, G. J. & Nielsen, M. T. *Principles of Human Anatomy*. (John Wiley & Sons, 2012).
17. Curry, S. Beyond expansion: structural studies on the transport roles of human serum albumin.

Vox Sang. **83 Suppl 1,** 315–319 (2002).

18. Peters, T. *All about Albumin: Biochemistry, Genetics and Medical Applications*. (Academic Press, 1995).

19. Marrack, J. R. & Hoch, H. Serum Proteins: A Review. *J. Clin. Pathol.* **2,** 161–192 (1949).

20. Petsko, G. A. & Ringe, D. *Protein Structure and Function*. *Annual Review of Physical Chemistry* (New Science Press Ltd, 2004). doi:10.1146/annurev.pc.23.100172.001121

21. Daneshegar, P., Moosavi-Movahedi, A. A., Norouzi, P., Ganjali, M.R. *et al.* Characterization of paracetamol binding with normal and glycated human serum albumin assayed by a new electrochemical method. *J. Braz. Chem. Soc.* **23,** 315–321 (2012).

22. Maciążek-Jurczyk, M., Szkudlarek, A., Chudzik, M., Pożycka, J. & Sułkowska, A. Alteration of human serum albumin binding properties induced by modifications: A review. *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.* **188,** 675–683 (2018).

23. Dockal, M., Carter, D. C. & Rüker, F. The three recombinant domains of human serum albumin. *J. Biol. Chem.* **274,** 29303–29310 (1999).

24. Dockal, M., Carter, D. C. & Rüker, F. Conformational transitions of the three recombinant domains of human serum albumin depending on pH. *J. Biol. Chem.* **275,** 3042–3050 (2000).

25. Sugio, S., Kashima, A., Mochizuki, S., Noda, M. *et al*. Crystal structure of human serum albumin at 2.5 Å resolution. *Protein Eng.* **12,** 439–446 (1999).

26. Thompson, T. J. U., Islam, M. & Bonniere, M. A new statistical approach for determining the crystallinity of heat-altered bone mineral from FTIR spectra. *J. Archaeol. Sci.* **40,** 416–422 (2013).

27. Brodersen, R., Sjodin, T. & Sjoholm, I. Independent binding of ligands to human serum albumin. *J. Biol. Chem.* **252,** 5067–5072 (1977).

28. Ghuman, J., Zunszain, P. A., Petitpas, I., Bhattacharya, A. A. *et al.* Structural basis of the drug-binding specificity of human serum albumin. *J. Mol. Biol.* **353,** 38–52 (2005).

29. Birkett, D. J., Myers, S. P. & Sudlow, G. Effects of fatty acids on two specific drug binding sites on human serum albumin. *Mol. Pharmacol.* **13,** 987–992 (1977).

30. Bhattacharya, A. A., Grüne, T. & Curry, S. Crystallographic analysis reveals common modes of binding of medium and long-chain fatty acids to human serum albumin. *J. Mol. Biol.* **303,** 721–732 (2000).

31. Goodman, D. W. S. The Interaction of human serum albumin with long-chain fatty acid anions. *J. Am. Chem. Soc.* **80,** 3892–3898 (1958).

32. Spector, A. A. Fatty acid binding to plasma albumin. *J. Lipid Res.* **16,** 165–179 (1975).

33. Yamasaki, K., Maruyama, T., Yoshimoto, K., Tsutsumi, Y. *et al.* Interactive binding to the two principal ligand binding sites of human serum albumin: Effect of the neutral-to-base transition. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* **1432,** 313–323 (1999).

34. Baroni, S., Mattu, M., Vannini, A., Cipollone, R. *et al.* Effect of ibuprofen and warfarin on the allosteric properties of haem-human serum albumin: A spectrospic study. *Eur. J. Biochem.* **268,** 6214–6220 (2001).

35. Wilting, J., van der Giesen, W. F. & Janssen, L. H. M. The effect of human albumin conformation on the binding of warfarin. *Acta Pharm. Suec.* **17,** 74–75 (1980).

36. Petitpas, I., Bhattacharya, A. A., Twine, S., East, M. *et al*. Crystal structure analysis of warfarin binding to human serum albumin. Anatomy of drug site I. *J. Biol. Chem.* **276,** 22804–22809 (2001).

37. Daneshgar, P., Moosavi-Movahedi, A. A., Norouzi, P., Ganjali, M. R. *et al.* Molecular interaction of human serum albumin with paracetamol: Spectroscopic and molecular modeling studies. *Int. J. Biol. Macromol.* **45,** 129–134 (2009).

38. Sułkowska, A., Bojko, B., Równicka, J. & Sułkowski, W. W. Paracetamol and cytarabine binding competition in high affinity binding sites of transporting protein. *J. Mol. Struct.* **792–793,** 249–256 (2006).

39. Kang, J. S. & Lee, M. H. Overview of therapeutic drug monitoring. *Korean J. Intern. Med.* **24,** 1–10 (2009).

40. Vorum, H. & Honoré, B. Influence of fatty acids on the binding of warfarin and phenprocoumon to human serum albumin with relation to anticoagulant therapy. *J. Pharm. Pharmacol.* **48,** 870–875 (1996).

41. Olsen, G. D. Methadone binding to human plasma albumin. *Science (80-.).* **176,** 525–526 (1972).

42. Reidenberg, M. M. & Affrime, M. Influence of Disease on Binding of Drugs To Plasma Proteins. *Ann. N. Y. Acad. Sci.* **226,** 115–126 (1973).

43. Fredholm, B. B., Rane, A. & Persson, B. Diphenylhydantoin binding to proteins in plasma and its dependence on free fatty acid and bilirubin concentration in dogs and newborn infants. *Pediatr. Res.* **9,** 26–30 (1975).

44. Campagnoli, M., Kragh-Hansen, U., Pedersen, A. O., Amoresano, A. *et al.* Structural analysis, fatty acid and thyroxine binding properties of Vancouver and Naskapi variants of human serum albumin. *Clin. Biochem.* **36,** 597–605 (2003).

45. Minchiotti, L. *et al.* Structural characterization, stability and fatty acid-binding properties of two French genetic variants of human serum albumin. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* **1431,** 223–231 (1999).

46. Kwon, M. J., Kim, H. J., Kim, J. W., Lee, K. H. *et al.* Determination of plasma warfarin concentrations in Korean patients and its potential for clinical application. *Korean J. Lab. Med.* **29,** 515–523 (2009).

47. Marx, J., Hockberger, R. & Walls, R. *Rosen's Emergency Medicine - Concepts and Clinical Practice*. (Elsevier, 2013).

48. Janssen, G. M. & Venema, J. F. Ibuprofen: Plasma Concentrations in man. *J. Int. Med. Res.* **13,** 68–73 (1985).

49. RICHTER, A. & SMITH, S. E. Bioavailability of different preparations of paracetamol. *Br. J. Clin. Pharmacol.* **1,** 495–498 (1974).

50. Dewland, P. M., Reader, S. & Berry, P. Bioavailability of ibuprofen following oral administration of standard ibuprofen, sodium ibuprofen or ibuprofen acid incorporating poloxamer in healthy volunteers. *BMC Clin. Pharmacol.* **9,** 1–10 (2009).

51. Wittkowsky, A. K. Generic warfarin: Implications for patient care. *Pharmacotherapy* **17,** 640–643 (1997).

52. Donaldson, P. M., Greetham, G. M., Shaw, D. J., Parker, A. W. *et al*. A 100 kHz pulse shaping 2D-IR spectrometer based on dual Yb:KGW amplifiers. *J. Phys. Chem. A* **122,** 780–787 (2018).

53. Greetham, G. M., Donaldson, P. M., Nation, C., Sazanovich, I. V. *et al.* A 100 kHz time-resolved multiple-probe femtosecond to second infrared absorption spectrometer. *Appl. Spectrosc.* **70,** 645–653 (2016).

54. Team, R. D. C. R: A Language and Environment for Statistical Computing. *R Found. Stat. Comput.* (2010).

55. Bos, O. J. M., Labro, J. F. A., Fischer, M. J. E., Wilting, J. & Janssen, L. H. M. The molecular mechanism of the neutral-to-base transition of human serum albumin. *J. Biol. Chem.* **264,** 953–959 (1989).

56. Granberg, R. A. & Rasmuson, Å. C. Solubility of paracetamol in pure solvents. *J. Chem. Eng. Data* **44,** 1391–1395 (1999).

57. Yalkowsky, S. H., He, Y. & Jain, P. *Handbook of Aqueous Solubility Data*. (CRC Press, 2010).

58. Paracetamol. Available at: https://www.drugbank.ca/drugs/DB00316. (Accessed: 21st April 2020)

59. Warfarin. Available at: https://www.drugbank.ca/drugs/DB00682. (Accessed: 21st April 2020)

60. Ibuprofen. Available at: https://www.drugbank.ca/drugs/DB01050. (Accessed: 21st April 2020)

61. Amado, A. M., Azevedo, C. & Ribeiro-Claro, P. J. A. Conformational and vibrational reassessment of solid paracetamol. *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.* **183,** 431–438 (2017).

62. Varfolomeev, M. A., Abaidullina, D. I., Klimovitskii, A. E. & Solomonov, B. N. Solvent effect on stretching vibration frequencies of the N-H and O-H groups of diphenylamine and phenol in complexes with various proton acceptors: Cooperative effect. *Russ. J. Gen. Chem.* **77,** 1742–1748 (2007).

63. Vueba, M. L., Pina, M. E. & Batista de Carvalho, L. A. E. Conformational stability of ibuprofen: Assessed by DFT calculations and optical vibrational spectroscopy. *J. Pharm. Sci.* **79,** 845–859 (2008).

64. Kostova, I., Amalanathan, M. & Hubert, J. I. Molecular first order hyperpolarizability and vibrational spectral investigationof Warfarin sodium. *The Journal of Chemical Physics* **378,** 88/102 (2010).

65. Bos, O. J. M., Fischer, M. J. E., Wilting, J. & Janssen, L. H. M. Mechanism by which warfarin binds to human serum albumin. Stopped-flow kinetic experiments with two large fragments of albumin. *Biochem. Pharmacol.* **38,** 1979–1984 (1989).

66. Rzaei-tavirani, M., Hassan, S. & Ranjbar, B. The effects of acetaminophen on human serum albumin (HSA). *Iran. J. Pharm. Res.* **4,** 239–244 (2005).

67. Bojko, B., Sułkowska, A., Maciazek-Jurczyk, M., Równicka, J. *et al.* Investigations of acetaminophen binding to bovine serum albumin in the presence of fatty acid: Fluorescence

and 1H NMR studies. *J. Mol. Struct.* **924–926,** 332–337 (2009).

68. Wilding, G., Feldhoff, R. C. & Vesell, E. S. Concentration-dependent effects of fatty acids on warfarin binding to albumin. *Biochem. Pharmacol.* **26,** 1143–1146 (1977).

69. Petersen, C. E., Ha, C. E., Curry, S. & Bhagavan, N. V. Probing the structure of the warfarin-binding site on human serum albumin using site-directed mutagenesis. *Proteins Struct. Funct. Genet.* **47,** 116–125 (2002).

70. Wilting, J., Weideman, M. M., Roomer, A. C. J. & Perrin, J. H. Conformational changes in human serum albumin around the neutral pH from circular dichroic measurements. *BBA - Protein Struct.* **579,** 469–473 (1979).

71. Wilting, J., Hart, B. J. & De Gier, J. J. The role of albumin conformation in the binding of diazepam to human serum albumin. *BBA - Protein Struct.* **626,** 291–298 (1980).

72. He, S., Huang, M., Ye, W., Chen, D. *et al.* Conformational change of bovine serum albumin molecules at neutral pH in ultra-diluted aqueous solutions. *J. Phys. Chem. B* **118,** 12207–12214 (2014).

73. Harding J.R & J.W, K. Bromocresol green as a reagent for serum albumin J. R. *Proc. Assoc. Clin. Biochem.* **5,** 51–53 (1968).

74. Aldrich, S. Bovine Serum Albumin. Available at: https://www.sigmaaldrich.com/content/dam/sigma-aldrich/docs/Sigma/Product_Information_Sheet/a2153pis.pdf. (Accessed: 27th April 2020)

75. Hamilton, J. A. Fatty acid interactions with proteins: What X-ray crystal and NMR solution structures tell us. *Prog. Lipid Res.* **43,** 177–199 (2004).

7. Conclusions and Future Work

In Chapter 1, 2D-IR spectroscopy is presented as a valuable technique for the examination of the amide I band of proteins. The sensitivity of the 2D-IR signal to the transition dipole moment allows the examination of protein secondary structures which is not possible using linear IR absorption methods. However, physiologically relevant studies of proteins are highlighted as an area absent in current 2D-IR studies as the strong water absorptions which overlap peaks pertaining to the important amide I band of proteins promotes measurements to be performed in deuterated solvents.

The theory and methodology relevant to 2D-IR spectroscopy was described in detail in Chapter 2. The two laser systems used in this thesis which are housed at the Rutherford Appleton Laboratories were described and details of the data acquisition for each instrument were also discussed.

In Chapter 3, measurements of the molecular dynamics of proteins in water were made which demonstrate the ability of 2D-IR spectroscopy to measure proteins in water without the need for H/D exchange of the solvent, thus increasing the physiological relevance of 2D-IR studies. The 2D-IR response of H_2O is shown to be negligible at a waiting time of 250 fs which allows isolation of the protein signal and accurate relative quantitative measurements of proteins in H_2O are made. Signals from proteins with different secondary structures, albumin (α-helix) and globulins (β-sheet) are shown to be clearly identifiable in serum and accurate relative quantitative measurements of these proteins are made. Based on their 2D-IR responses, identification of structurally similar proteins, IgA and IgM, were made in serum, this demonstrates the promising development of 2D-IR as a tool for further protein identification and their relative quantifications.

The ability of 2D-IR to detect low molecular weight species in the presence of high molecular weight proteins is explored in Chapter 4. The on- and off-diagonal structure observed in 2D-IR spectroscopy allows identification of the low molecular weight species glycine in serum using 3 different analysis approaches. As many important biomarkers for disease are typically in the low molecular weight fraction this study highlights the potential of 2D-IR to be used to identify low molecular weight proteins and potential biomarkers. A limit of detection is made using the amino acid which allows critical assessment of the sensitivity of 2D-IR for larger proteins in future analysis. However where analysis utilises the entire 2D-IR spectrum as in the linear combination analysis approach, results were not comparable to the study conducted in Chapter 3 and so the issue of repeatability between datasets is highlighted.

The advanced capability of 2D-IR to analyse biofluids has potential for further analytical applications able to exploit the additional information content of 2D-IR spectroscopy in the healthcare arena. However for the determination of many species further experimentation in defining lineshapes associated with specific proteins and biomarkers is imperative. IR absorption methods have seen success using multivariate analysis techniques to classify and quantify healthy and non-healthy serum samples and so applying these techniques to a similar 2D-IR dataset would be a logical next step to push 2D-IR spectroscopy forward.

Thus far, 2D-IR experiments of proteins in H_2O have overcome the issue of overlapping water absorptions by utilising the timing between pump and probe pulses in order to minimise the associated water signal. However the issue of sample to sample repeatability and fluctuations in the laser system prevented absolute quantitative analysis of protein concentrations. In Chapter 5, an internal normalisation method has been created, by further utilising the water response at late waiting times, allowing quantitative protein concentrations to be obtained. While current 2D-IR spectrometers are not designed for routine high throughput work, this type of application demonstrated in Chapter 5 serves to motivate a trajectory towards faster, more hands-free 2D-IR data acquisition, stimulating more commercially-available instruments. Additional investigation into other pre-processing techniques would also serve useful for future applications.

Tighter controls of the sample path length could be employed in order to aid the normalisation method. However as spacers becomes increasingly difficult to use with their smaller size, experimentation into the sample cell would need to be employed in order to create a fixed path between the CaF_2 windows for the sample to flow to and protocols establishing cleaning techniques in order to prevent contamination would also need to be established.

Finally, in Chapter 6, the limits of 2D-IR spectroscopy are explored in measuring protein-drug binding in aqueous serum in an attempt to increase the physiological relevance of both drug-binding and 2D-IR studies. Using the normalisation method in Chapter 5, changes in the 2D-IR spectral lineshapes of serum albumin are associated with drug binding, and different lineshapes are noted for the three different drugs used. Changes in the signal amplitude of serum albumin are also noted, in particular for the paracetamol dataset which indicates that the albumin molecule houses two binding sites for paracetamol, which is comparable to previous drug-binding studies, and in doing so the protein molecule increases its α-helix

content suggesting a conformational change in the protein molecule. Conformational changes of serum albumin when binding to drugs is well documented in solvent buffers however studies documenting these changes in the natural serum environment have not yet been made and so comparisons of the structural change due to paracetamol binding in serum cannot be made. Further experimentation into the 2D-IR signatures of different albumin isomers would be beneficial for albumin-drug binding studies. Calculations of the vibrational decays of the amide I mode due to binding could help in the identification of bound and unbound albumin. Understanding the mechanism of drug-binding could be aided by isotopic labelling of the drug and the use of two colour 2D-IR experiments which have often been employed to demonstrate binding where by an additional probe is set up at a different wavelength region in order to detect molecular couplings, which could be an interesting experiment in the detection of bound and unbound species.

2D-IR is a label-free technique that generates 2D frequency correlation maps as a function of time and produces vast amounts of information quickly and throughout this thesis, it has shown to provide an alternative approach to existing methods for biofluid analysis. The benefits of using IR spectroscopy over traditional wet chemistries allows the information from more than one biological molecule to be obtained from one measurement, as opposed to several measurements and sample volumes required for traditional methods. However in order to advance 2D-IR spectroscopy for use in a clinical environment, technological developments allowing more compact instrumentation would be beneficial as well as improvements with laser stability. This may help provide further insights into the analysis of biofluids using 2D-IR.

Acknowledgements

Throughout my PhD I have received the support and guidance from many people. First and foremost I would like to thank my supervisor Neil, for always answering my somewhat silly questions and for all of those interesting data discussions, even if I did end up with more questions and work afterwards. I'm very grateful to have had you as a supervisor, so thank-you!

To my second supervisor, Matt thank you for all of your help and acronyms along the way! I have learned a great deal about the analytics from you and the Baker team! Speaking of, Sayali, Holly-Jean and Hamez, thank you for keeping the laughter going.

To the life-long friends I've made along the way in the form of the Strathclyde UCP Group before we were disbanded; Gordon, Lucy & Robby. Your enthusiasm for all things punny and delivery of high quality cakes has been very encouraging and jam-packed with fun, laughter and of course 2D-IR... Thank you Danny and the rest of the team at York for all your help and discussions. I'm not sure I would have managed to get through the long days and nights at RAL without you all.

The team at RAL; Tony, Mike, Greg, Ian, Emma, Paul & Igor, I owe a massive thank-you to all, for all of your help in data acquisition, troubleshooting and spectrometer discussions. I appreciate all of the nights each of you stayed behind to help out with matters big or small.

A special thanks to my parents and sister, for putting up with my stress head over the last however many years in education. Thank you for giving me all your love and support and I hope I've made you proud! I know you will be very happy to hear that at the ripe-old age of 28, my studies are finally coming to an end (for now). My wonderful friends deserve a shout out and a big thank-you for helping me de-stress; Gin and Wine… no but seriously, Catriona, Jenna, Brodie, and Becky, you four especially have helped me through some tough times and I can't thank you enough. Cheers to you!

To my precious puppy dog, Albert, thank you for being so cute and cuddly and the number one good doggo for stress relief. Last but most definitely not least to my darling husband Alasdair, your constant love, support and belief that I can achieve anything and everything I chose to do is beyond encouraging and most appreciated in times of need. Thank-you.