Unravelling the Complexities of Protein Conformational Stability Using Raman Spectroscopy and Two-Dimensional Correlation Analysis

Thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Lancaster University

by

Ilokugbe Ettah

November 2020

For El-Elyon; Your priceless gift of life made this endeavour possible.

All science is necessarily prophetic, so truly so that the power of prophecy is the test, the infallible criterion by which any presumed science is ascertained to be actually and verily science……..

Samuel Taylor Coleridge

Declaration

This thesis has not been submitted in support of an application for another degree at this or any other university. It is the result of my work and includes nothing that is the outcome of work done in collaboration except where specifically indicated.

Excerpts of this thesis have been published as follows:

Determination of Phosphorylation and Deprotonation Induced Higher Order Structural Transitions in α_s-Caseins.

Ilokugbe Ettah and Lorna Ashton

Abstract

Unravelling the Complexities of Protein Conformational Stability Using Raman Spectroscopy and Two-Dimensional Correlation Analysis.

Submitted by Ilokugbe Ettah for the degree of Doctor of Philosophy,

November 2020.

Protein drug development is a risky, lengthy and expensive process. Accordingly, there are significant time and cost implications for stakeholders regardless of the stage at which a prospective candidate is discontinued. Mitigating this risk depends critically on making informed decisions as early as possible during the process. Some of the most indispensable decision-making aids are analytical techniques used in the characterisation of protein stability. Protein stability is a multifaceted and crucial determinant for progressing a candidate through development to commercialisation. The inherent complexity of proteins, the evolving pipeline of engineered protein constructs, as well as limitations of current techniques place a demand for the expansion and development of more analytical probes for protein stability.

This thesis focuses on the development of Raman spectroscopy and 2D-correlation analysis (2DCA) for monitoring conformational stability during early protein drug development. Although Raman spectroscopy has been used for many decades to study protein conformational changes in proteins; it is highly underutilized during protein drug development, owing partly to the complexity of Raman protein spectra. 2DCA is a mathematical technique that can significantly improve the

visualization of and clarify the interpretation of complicated spectral features arising from an applied perturbation. The specificity of 2DCA in extracting such spectral variations directly linked to a perturbation is apt for protein stability testing, where the identification of a protein's response to stress is pivotal.

In this work, we demonstrate this combined technique as a sensitive means of distinguishing the conformational stability profiles of proteins under an applied perturbation on the basis of post-translation modification, metal-ligand and surface charge mutation. In all cases, we show that the technique is capable of informing the choice of the most stable candidate and or conditions, even where subtle differences exist.

Acknowledgements

Contents

List of Figures

List of Tables

List of Equations

Word count: 47210

List of Abbreviations and Acronyms

2DCA	2D-Correlational Analysis
2DMW	2D-Moving Windows
2DPCMW	2D-Perturbation Correlation Moving Windows
Ac	Sodium Acetate Buffer
AP	Asynchronous Plot
apohTf	apo-Transferrin
CD	Circular Dichroism
CIMEMT	Centre for Innovative Manufacturing in Emergent Macromolecular Therapies
Cit	Sodium Citrate Buffer
dα_s-casein	Dephosphorylated α_s-Casein
Fab	Antigen Binding Fragment
Fc	Fragment Crystallisable Portion
holohTf	holo-Transferrin
HOS	Higher-Order Structure
IgG	Immunoglobulin G
IR	Infrared
NaP	Sodium Phosphate Buffer
NHS	National Health Service
NMR	Nuclear Magnetic Resonance
PCA	Principal Component Analysis
PTM	Post-Translational Modification
R&D	Research and Development
RNase A	Ribonuclease A
SAS	Solvent Accessible Area
SDM	Site-Directed Mutagenesis
SP	Synchronous plot
UCL	University College London
UV	Ultraviolet
UVRR	Ultraviolet Resonance Raman

1 Introduction

1.1 Background

Protein-based therapeutics continue to account for the highest number of approved biopharmaceuticals (Walsh, 2014, Agyei et al., 2017, Walsh, 2018). The high specificity and efficacy of these products (including antibodies, enzymes, fusion proteins and interferons) often provide life-saving options for patients with complex and chronic diseases, such as cancer, autoimmune disorders, diabetes, inflammatory conditions, and neurological diseases. Alongside health benefits, their sales have been attended with notable success (Walsh, 2014, Walsh, 2018). Despite these positive outcomes, the time and cost of research and development (R&D) remain a considerable challenge (Kannt and Wieland, 2016). On average, it takes between 10-15 years for a drug to reach the market and costs $2.6 billion for R&D (PhRMA, 2016). A major reason for this high cost is that the structural complexity of protein molecules is delicately linked to their stability and activity, thus making their development a more rigorous and expensive process compared to smaller chemical entities (Lagassé et al., 2017).

In the United Kingdom, high costs of biopharmaceuticals place pressure on the National Health Service (NHS) budget, limiting affordability and wider access to patients (NHS, 2017, Russell et al., 2018). The recent introduction of biosimilars (a drug with a similar biological structure to an already licensed drug) is promising with respect to reducing individual drug cost and widening patient access in the

NHS (Aladul et al., 2017). However, a significant cost burden still resides in the initial drug development stage (Samanen, 2013).

Characterising a protein's conformational stability during drug development is a critical element of developability assessment. Developability assessment is aimed at early de-risking of the drug development process and focuses on minimizing the chance that a drug candidate will fail to reach the market (Saxena et al., 2009, Yang et al., 2013). There is considerable potential for cost savings if such characterisation can be reliably carried out at an early stage before huge investments are made (DiMasi, 2002, Zurdo, 2013) because development costs influence the overall setting of a company's drug prices, whether or not a particular drug eventually reaches the market (DiMasi et al., 2003, Samanen, 2013). Therefore, developing analytical techniques that assist in making early decisions about the feasibility of developing and formulating a new protein molecule can influence the reduction of R&D costs and wastage during production.

Raman spectroscopy can identify and monitor protein conformational changes in a non-destructive manner. Its utility over a wide concentration range can obviate the need for empirical extrapolation in the context of high dose protein formulations, which generally have to be diluted to lower concentrations using most of the current routine techniques including Ultraviolet (UV) spectroscopy, Circular dichroism (CD), and Fluorescence spectroscopy (Lewis et al., 2014, Zhou et al., 2015). Raman spectroscopy is also particularly suited for the study of aqueous biological samples compared to Infrared (IR) spectroscopy due to reduced water interference in the spectra (Lewis et al., 2014, Manning, 2005). The flexibility of applying Raman spectroscopy to examine different sample states and

concentrations can potentially be used to provide a conformational stability profile using a single technique from early drug development through to market. Although the use of different techniques to probe a distinct aspect of stability is advantageous in obtaining complementary information, there is a possibility of obtaining a fragmented stability profile if a technique cannot be used when the sample is in a different form or state (Filipe et al., 2013).

In view of these advantages, Raman spectroscopy is attractive as a routine technique for monitoring protein conformational stability during protein drug development. However, the complexity of data obtained from studies of proteins can be challenging to analyse, requiring an appropriate choice of data preprocessing methods and chemometric tools (Buckley and Ryder, 2017). Standard chemometric techniques for analysing protein Raman data include principal component analysis (PCA) (Thiagarajan et al., 2015, Hernández-Vidales et al., 2019, Schack et al., 2019), partial least squares regression (Brewster et al., 2011, Lewis et al., 2014) and multivariate curve resolution-alternating least squares (Ota et al., 2016, Noothalapati et al., 2017). The preference of a data analysis method depends on the experimental design and consequently, the nature of acquired data. In the context of stability testing, data analysis methods that specifically relate the spectral changes to the applied stress are desirable. One versatile technique that has been applied to condense and interpret perturbation-induced spectral data is 2D-correlation analysis (2DCA).

2DCA is a versatile mathematical technique that can be applied to enhance the visualization and aid interpretation of data generated under the influence of a designated perturbation (Noda, 1993). Its specific advantages include delineating

overlapped peaks, clarifying peak assignments, and improving spectral resolution due to spreading the peaks over a second dimension (Noda, 1990). Several studies have demonstrated the potential of combining 2DCA and Raman spectroscopy to study conformational changes in proteins (section 1.7). However, this joint technique is yet to be developed for the analysis of conformational stability during early protein drug development.

1.2 Aim and Objectives

The main aim of this project is to develop the application of Raman spectroscopy combined with two-dimensional correlation analysis to characterise protein conformational stability under varying pH, temperature, storage and agitation conditions during early drug development. Samples of model proteins shall initially be investigated using Raman spectroscopy to determine and compare their conformational stability under specific environmental conditions. The model proteins are examined in pairs, each sample in a pair slightly differing from the other due to a specific chemical moiety. 2DCA shall be subsequently applied to the Raman data to uniquely analyse perturbation induced changes in the spectra of each pair of proteins. The same procedure will then be applied to a therapeutically relevant protein and its variants.

1.3 Thesis Overview

The subsequent sections of this introductory chapter will discuss protein conformational stability and Raman spectroscopy as well as its application to the study of protein structure and stability. In chapter 2, the materials and methods

are discussed. In chapter 3, the techniques are applied to understand the influence of post-translational modification on the stability of an intrinsically unfolded model protein. Chapter 4 covers the application of the techniques to the study of the influence of a ligand on the stability of a model protein with a well-defined folded structure. Chapter 5 will discuss the study of a therapeutically relevant protein molecule and four of its mutants in different buffers and sample states. Chapter 6 shall cover the introduction of a variant form of Raman spectroscopy, Ultraviolet Resonance Raman spectroscopy (UVRRS) and 2DCA to examine the samples studied in chapter 5. Finally, chapter 7 will cover general discussion and future work.

1.4 Protein Conformational Stability

The conformation of a protein denotes the specific three-dimensional shape which a protein assumes when it is folded (Price, 2000, Creighton, 2010). This conformation is stabilized by a number of weak interactions including hydrogen bonding, hydrophobic interactions, electrostatic interactions and Van der Waals' forces (Fersht, 1999, Price and Nairn, 2009). The disruption of these forces under environmental conditions such as exposure to solvent, temperature and pH, can result in a change in the conformation of a protein and subsequent denaturation (Murphy, 2001, Franks, 1993).

Consequently, the preservation of a protein's active conformation is well understood to be crucial in maintaining its normal biological function. In living systems, the conditions that make such preservation possible are in-built and managed by well-designed mechanisms. In contrast, proteins which are expressed

in other hosts or synthesized chemically do not have such homeostatic controls and therefore require a different approach to maintain the stability of their active conformation (Franks, 1993, Price, 2000). This approach requires a thorough understanding of the protein structure as well as the range of environmental conditions in which its stability is maintained.

Protein conformational stability can be defined as the free energy change (ΔG) that occurs in the reaction between the folded and unfolded forms of a protein (Figure 1.1).
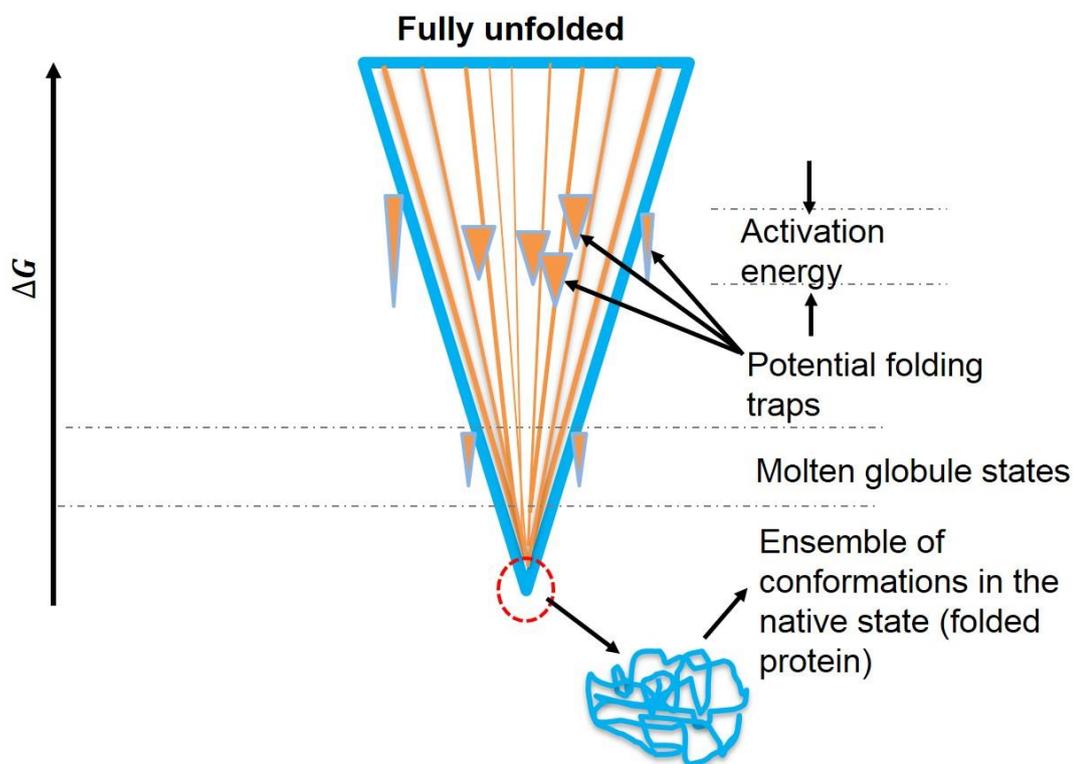
Figure 1.1 Schematic depicting funnel-shaped protein folding energy landscape and free energy change (ΔG). Figure adapted from Houde and Berkowitz (2014).

The typical value for this is approximately 20-60 kJ mol^{-1}; thus, there is only a narrow stability margin between folded and unfolded forms (Pace et al., 1996,

Huyghues-Despointes et al., 2001). In both the folded and unfolded forms, the concept of a single stable conformation is far removed from the complex and dynamic nature of the three-dimensional protein structure. A group of conformations exists, and in fact, at any point, a stable protein in solution will contain a mixture of folded and even some unfolded conformations (Price, 2000, Houde and Berkowitz, 2014, Kazlauskas, 2018). Therefore, the challenge is to stabilize this group of conformations such that the protein continues to maintain its structure and function.

Surmounting this challenge is crucial during drug development and manufacturing, where a protein may undergo conformational changes during processes such as purification, freeze-drying, formulation, storage and transportation (Ejima et al., 2007, Wang, 1999, Berkowitz et al., 2012). Such conformational changes may not only result in product instability and loss of activity but in immunogenic reactions, which could be life-threatening (Wang, 1999, Jiskoot et al., 1999). Furthermore, the protein itself could undergo covalent (chemical) structural changes involving fragmentation and more commonly, post-translation modification (PTM). The most common PTM encountered in the manufacture of protein drugs is glycosylation. However, the structure and stability of the protein drug could also be affected by less frequently encountered PTMs such as carboxylation, amidation, and oxidation (Walsh, 2010). Whilst much progress has been made in identifying PTMs presence using mass spectrometry and related techniques, the influence of a PTM on a protein's conformation is yet to be fully understood (Berkowitz et al., 2012).

Identifying changes to protein conformation during drug development typically involves stress testing to identify changes in the folding/unfolding status (Tamizi and Jouyban, 2016, Chang and Hershenson, 2002). Stress testing establishes a baseline for the stable form of a functioning protein and then assesses any deviations from that fingerprint. Whereas ΔG provides insight into the intactness of the folded state, it does not provide information about structural changes that are occurring in the protein (Huyghues-Despointes et al., 2001). Assessing conformational stability, as used in this thesis, refers to variations in protein higher-order structure (HOS) which can indirectly provide information about changes in a protein's three-dimensional shape (conformation).

The current methods of monitoring changes in HOS during the development of protein drugs are well described by Houde and Berkowitz (2014). A tool kit comprising different techniques with varying sensitivity limits and operating modes is used for this process. Many of these techniques fall into their 'standard' category because they can be used rapidly and routinely during development. The authors emphasize that the use of multiple techniques is advantageous not only to increase the chances of detecting a change in conformation but also in identifying the unique contribution of each technique. Furthermore, where the information from techniques overlap, the consistency of results can serve as a useful confirmation of a specific quality.

Due to the innate complexity of protein structure, limitations of current techniques as well as the development of diverse engineered protein constructs, there is an ongoing need to expand and develop this toolbox to meet the challenge of characterising protein drug conformational stability (Jorgensen et al., 2013, Zurdo,

2013, Houde and Berkowitz, 2014). A reliable assessment of protein conformational stability has implications for vital procedures such as candidate selection, optimization, compatibility and comparability studies.

1.5 Raman Spectroscopy

1.5.1 The Raman Effect

About 9 decades ago, Raman and Krishan (1928) demonstrated the Raman effect in very unrefined conditions. They focused sunlight using a telescope on to several liquid and vapour samples and used complementary light-filters (blue-violet and yellow-green) to extinguish the light passing through the sample. By placing a yellow filter between the sample and the detector (human eye), they observed that the track of light reappeared thus, demonstrating a light scattering effect.
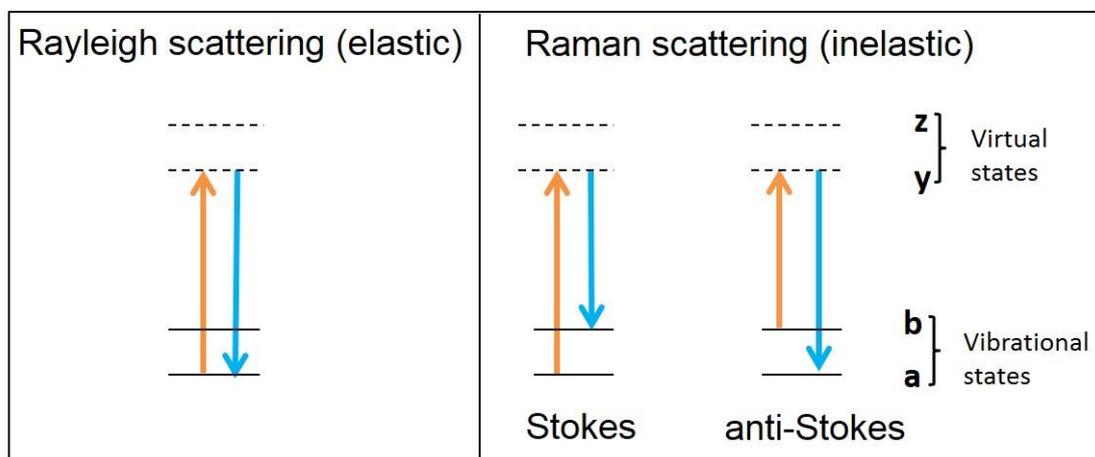
Figure 1.2 Schematic showing elastic (Rayleigh) and inelastic (Raman) scattering. Image adapted from Smith and Dent (2005).

The Raman effect refers to the inelastic scattering of light when light (photons) interacts with a molecule, channelling energy to it (Long, 2002, Ferraro et al.,

2003). This energy is not directly absorbed but causes a distortion of the electron cloud of the molecule and its transition to a virtual state. The virtual state is a highly unstable, transient union between the light and the molecule; and ceases to exist when the energy is released as scattered light in all directions. Only about 1 in 10^6 photons are scattered inelastically, making Raman scattering a weak process. In addition to receiving energy, molecules can also behave as an independent source of radiation and influence the scattering process by transferring energy to the scattered photons (Smith and Dent, 2005, Long, 2002).

Inelastic scattering occurs when the vibrational energy of the molecule is altered during the scattering process (Figure 1.2). Where the photons are scattered with a lesser energy than the incident radiation, it is termed Stokes scattering. On the other hand, when scattered photons have a higher energy, the process is referred to as anti-Stokes scattering. Stokes scattering is routinely used for Raman spectroscopic analysis because it occurs more frequently than anti-Stokes scattering. The higher intensity is due to a higher proportion of molecules existing in the ground state and not an excited vibrational state at room temperature (Smith and Dent, 2005, Vandenabeele, 2013). Anti-Stokes scattering may be used for studies conducted at higher temperatures where a higher proportion of molecules will possess a raised vibrational energy or in other related Raman techniques such as coherent anti-Stokes scattering Raman spectroscopy. If the energy of the molecules remains unchanged, the scattering process is elastic (Rayleigh scattering). This elasticity arises from a distortion of the electron cloud in the absence of any nuclear displacement. Compared to Stokes and anti-Stokes

scattering, Rayleigh scattering is a stronger and more dominant scattering process (Smith and Dent, 2005).

1.5.2 Types and Modes of Vibrations

There are two main types of vibrations, stretching and bending vibrations (Chalmers and Dent, 1997, Günzler and Gremlich, 2002, Stuart, 2004). Stretching involves movement along the bond axis and may lengthen or shorten the distance between atoms. Bending refers to a movement that results in a change of bond angle.
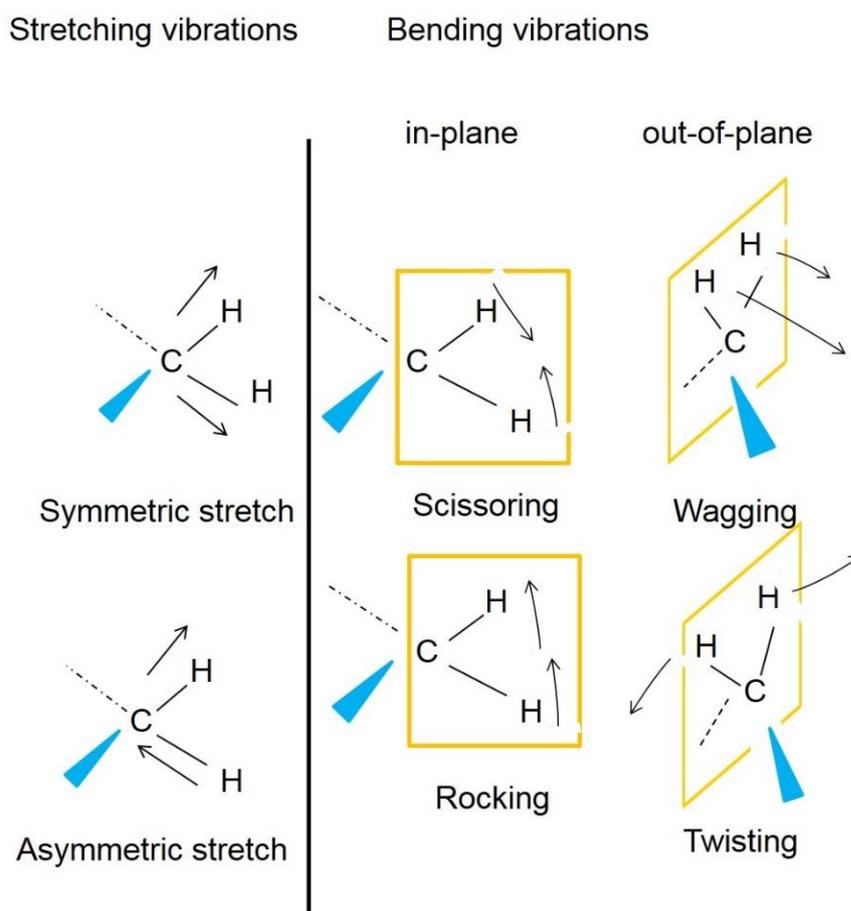
Figure 1.3 Vibrational modes for methylene group. Image adapted from Pavia et al. (2009).

These vibrations can further be divided into in-plane and out-of-plane for bending vibrations and symmetric and asymmetric stretch for stretching vibrations, as shown in Figure 1.3. The number of vibrations a molecule can undergo depends on the number of atoms and the number of degrees of freedom (Ferraro et al., 2003, Smith and Dent, 2005).

1.5.3 Raman Spectroscopy of Proteins

A protein Raman spectrum (Figure 1.4) links the vibrations that occur in a protein to the structural components of the protein (Table 1.1).
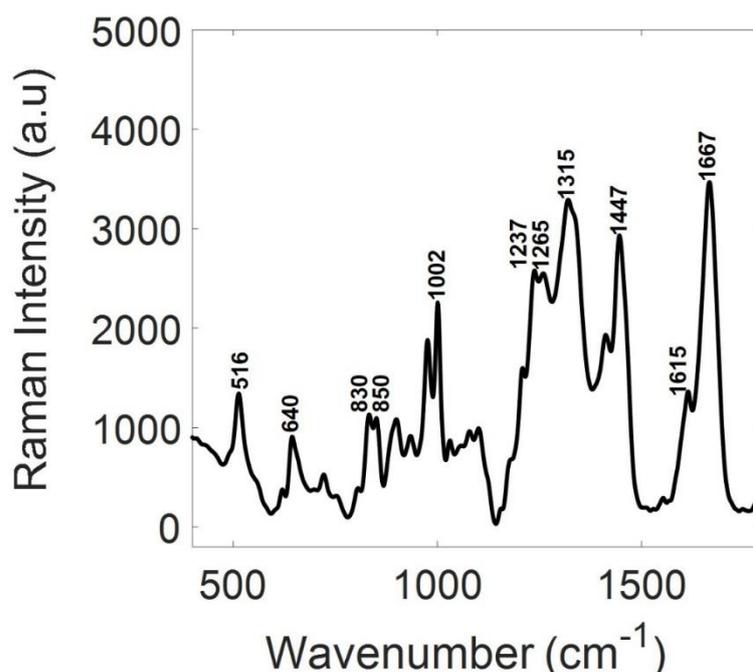
Figure 1.4 Averaged (n=6) Raman spectrum of 10 mg/ml Ribonuclease A (RNase A) dried film. Preprocessing was carried out by baseline correction, normalization and smoothing. Raman spectrum was collected in 10s using an excitation wavelength of 785 nm. Assignments are listed in Table 1.1.

These correlations between the vibrations and structural elements are referred to as assignments. These assignments are based on theoretical vibrational mode calculations, validated with experimental results from peptide related model compounds such as N-methylacetamide; amino acids and peptides (Miyazawa et al., 1958, Jansen et al., 2006, Hernández et al., 2016, Schweitzer-Stenner, 2001). Proteins comprise one or more polypeptide chains, and their structure is defined in levels of hierarchy and complexity. The simplest or blueprint structure is that of the amino acids linked in a linear sequence by peptide bonds. The next level refers to the secondary structure, which is the local spatial conformation of the polypeptide backbone, excluding the side chain (Fersht, 1999). The most common secondary structures are the α-helices and β-sheets. The third level is the tertiary structure, which is a three-dimensional shape of all peptide atoms in relation to one another. This shape enables bonding between groups that would not otherwise come in contact (Walsh, 2002). Quaternary structures refer to the final specific arrangement of the subunits (polypeptide chains) or monomers that constitute a protein. Two other structures in this hierarchy system are domains and super-secondary structures (Lesk, 2001). Domains are tightly folded sub-regions of a single polypeptide. They are structurally distinct and are made up of structural motifs, also referred to as super secondary structures. Super secondary structures involve an interaction between regions of secondary structures which are in proximity in the sequence (Lesk, 2001, Littlechild, 2013). Protein Raman vibrations can broadly be grouped into polypeptide backbone and sidechain vibrations.

The vibrational modes that are commonly used for structural analysis are the amide I (1600-1700 cm^{-1}) and amide III (1200-1340 cm^{-1}) bands in the polypeptide backbone region as well as the modes from aromatic amino acids and disulfide bonds (Spiro and Gaber, 1977, Tuma, 2005). The amide bands relate to the peptide bonds. Amide I region is assigned to C=O stretching of carbonyl groups while the amide III region is assigned to NH bending and C-N stretching modes (Lord and Yu, 1970, Wen, 2007). Side-chain peaks mainly include aromatic residues: Trp, Phe and Tyr. Spectral changes including peak intensities, peak shifts or peak shape can be linked with specific vibrational modes identifying variations in secondary, and side-chain structure, and consequently the conformational stability of a protein.

Table 1.1 Common protein structure assignments from Raman spectra

Peak (cm^{-1})	Assignment	References
505/510/516	S-S stretch	(Chen and Lord, 1976, Yu and Jo, 1973)
640	Tyr	(Chen and Lord, 1976, Maiti et al., 2004)
760	Trp, indole ring vibrations	(Chen et al., 1973, Liang et al., 2006)
830	Tyr, in-plane ring breathing motion	(Hernández et al., 2016)
850	Tyr, C-H out of plane bending	(Hernández et al., 2016)
830/850	Tyr Fermi doublet	(Siamwiza et al., 1975)
880	Sidechain/indole ring vibration	(Miura et al., 1988, Kitagawa et al., 1979)
930	α-helix, N-C$_\alpha$-C stretch	(Monti et al., 1998, Beattie et al., 2008)
940	α-helix, C-C-N stretching	(Frushour and Koenig, 1974, Nonaka et al., 1993)
980	Phosphate stretch	(Zhang et al., 2005, Ashton et al., 2011).
1003	Phe	(Lord and Yu, 1970, Hernández et al., 2013)
1030	Phe	(Chen and Lord, 1976, Hernández et al., 2013)
1065	C-C stretch, Asp, Glu, Lys	(Vohník et al., 1998, Durig et al., 1987)
1124	Trp, C-N stretch	(Takeuchi, 2003)
1173	Tyr	(Lord and Yu, 1970, Hernández et al., 2016)
1208	Tyr	(Hernández et al., 2016)
1237	β-sheet (Amide III- NH deformation with C-N stretch	(Li and Li, 2009, Lippert et al., 1976, Yu et al., 1972)

1250	Disordered structure	(Ashton et al., 2007, Jarvis et al., 2007)
1262	α-helix, Amide III/C-N stretching and NH in-plane bending	(Lee and Krimm, 1998, Chen and Lord, 1974)
1315	α-helix	(Tsuboi et al., 2000, Rygula et al., 2013)
1337	Trp, α-helix	(Tsuboi et al., 2000, Rygula et al., 2013)
1415	Ionization of carboxyl side chains	(Lord and Yu, 1970)
1447	CH_2 deformation	(Chen et al., 1973, Mangialardo et al., 2012)
1460	Proline side chain	(Jordan et al., 1996)
1550	Trp	(Miura et al., 1989, Takeuchi, 2003)
1615	Tyr	(Maiti et al., 2004, Hernández et al., 2016)
1600	Tyr/Phe	(Mangialardo et al., 2012, Hernández et al., 2016)
1640	H-O-H bending	(Lord and Yu, 1970, Byler et al., 1988)
1660	Disordered structure (amide I)	(Pelton and McLean, 2000, Rygula et al., 2013)
1667	β-sheet (amide I)	(Shao et al., 1999, Maiti et al., 2004)

1.6 2D-Correlation Analysis (2DCA)

Two-dimensional correlation analysis was first proposed by Isao Noda (1986) as being applicable to IR spectroscopy; however, its application was later extended to other forms of spectroscopy and beyond spectroscopy (Noda, 1986, Noda, 1993). This approach was known as generalised two-dimensional spectroscopy at the time but is now just referred to as 2D-correlation spectroscopy or 2D-correlation analysis. In the last three decades, there have been significant advances in the types of 2DCA available including, projection 2D-correlation; hetero-correlation analysis; concatenated 2D-correlation; sample-sample correlation, moving windows, perturbation correlation moving windows, scaling moving–windows, 2D-codistribution, quadrature 2D correlation and two trace 2D correlation analysis (Noda, 2014a, Park et al., 2016b, Park et al., 2018).

2DCA is an adaptable technique applied to enhance the visualization and aid interpretation of spectral data generated under the influence of a defined perturbation (Noda, 1993, Noda et al., 2000). A variety of perturbations can be used provided that the perturbation can induce an observable spectral variation (Noda, 1993). Applied perturbations may be static such as temperature, concentration and pressure or dynamic, including chemical or biological reactions and exposure to stimuli (Noda and Ozaki, 2005, Park et al., 2016b). Although the source of electromagnetic radiation is not limited to any spectroscopic method or the field of spectroscopy, spectroscopic methods still dominate the applications of 2DCA. A survey of the analytical probes used in 461 2DCA studies revealed that 61% used IR spectroscopy; 10% used Raman spectroscopy; 9% used Near-IR

spectroscopy; 9% used Nuclear magnetic resonance spectroscopy (NMR) and miscellaneous techniques accounted for the rest (Noda, 2014a). This trend continues to be reflected in more recent surveys (Park et al., 2016b, Park et al., 2018). There is also a considerable versatility in the fields of application of 2DCA comprising but not limited to the study of food products, chemicals, biological materials, polymers and plant metabolites (Noda and Marcott, 2002, Molenda et al., 2005, Geitner et al., 2015, Baranska et al., 2006, Ashton et al., 2008).

1.6.1 2DCA Theory

The general operation of 2DCA is displayed in Figure 1.5. Electromagnetic radiation via an analytical probe is applied to a sample or system.
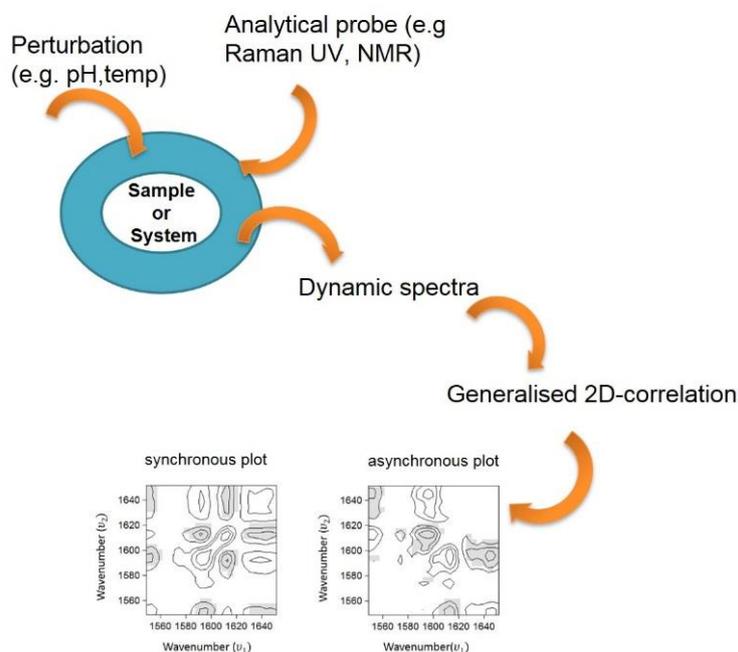
Figure 1.5 Schematic showing a general operation of 2D-correlation analysis. Image adapted from Noda (1993) and Noda (2012).

A designated perturbation is applied simultaneously to the same system to generate a response which is captured in the *dynamic spectra*. A *dynamic spectrum* characterises the fluctuations in the spectral signals selectively induced by the applied perturbation to the system. 2DCA is then applied to the *dynamic spectra* to generate the synchronous and asynchronous plots (Noda, 1993, Noda and Ozaki, 2005). The synchronous plot captures similar spectral responses as a result of the perturbation, whereas the asynchronous plot represents dissimilar responses as a result of the perturbation. These responses are plotted as a function of two independent wavenumbers. The dynamic spectrum $\tilde{y}(v, t)$ can be obtained, as shown in equation 1.1. where $\bar{y}(v)$ is the reference spectrum.

$$\tilde{y}(v, t) = \begin{cases} y(v, t) - \bar{y}(v) & \text{for } T_{min} \leq t \leq T_{max} \\ 0 & \text{otherwise} \end{cases} \qquad (1.1)$$

The reference spectrum is usually an average or the ground state spectrum. It may also be set at zero depending on the type of 2DCA. Setting it at zero is useful for mathematical consistency. However, an average spectrum has been reported to be a more accurate representation, especially when the specific physical origin of the dynamic spectrum is unknown (Noda, 1993, Noda and Ozaki, 2005). The formula for the average spectrum is shown in equation 1.2.

$$\bar{y}(v) = \frac{1}{T_{max-}T_{min}} \int_{T_{min}}^{T_{max}} y(v,t) \, dt \qquad (1.2)$$

where $T_{max-}T_{min}$ is the interval between the highest and lowest variable (e.g. time, pressure, temperature) and $y(v,t)$= spectral intensity variation

The 2D correlation function which applies to several systems is given as

$$\Phi(v_1 \, v_2) + i\psi(v_1, v_2) = \frac{1}{\pi(T_{max}-T_{min})} \int_0^\infty \tilde{Y}_{1}(\omega).\tilde{Y}_2^*(\omega)d\omega \qquad (1.3)$$

$\Phi(v_1 \, v_2)$ and $i\psi_\omega(v_1, v_2)$ represent the synchronous and asynchronous correlation intensities while $\tilde{Y}_{1(\omega)}$ and $\tilde{Y}_{2(\omega)}^*$ are the forward Fourier transform and conjugate of the Fourier transform of $\tilde{y}(v_1, t)$ and $\tilde{y}(v_2,t)$ respectively. $\tilde{y}(v_1, t)$ and $\tilde{y}(v_2, t)$ correspond to the spectral intensity variations observed at given spectral variables v_1 and v_2 (Noda, 1993).

1.6.2 Synchronous Plot

The synchronous plot is the real component of the 2D-correlation function (Noda, 1993, Noda and Ozaki, 2005). The synchronous plot shows the degree of similarity between two distinct spectral intensity variations occurring at two independent

wavenumbers (v_1, v_2). It has a diagonal line where $v_1=v_2$. The peaks that occur on

the diagonal are referred to as autopeaks, and those occurring on either side of the

diagonal are known as cross peaks. As the plot is symmetric along the diagonal,

the cross peaks on one side of the diagonal have the reverse co-ordinates of those

on the opposite side. A schematic is shown in Figure 1.6.
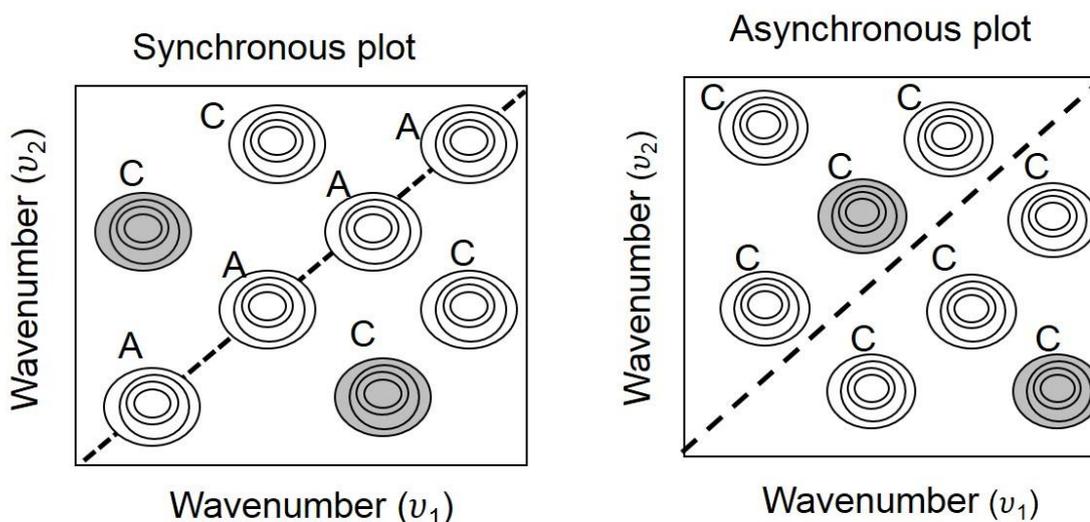
Figure 1.6 Schematic showing synchronous and asynchronous plots. Shaded contours indicate negative cross peaks and unshaded contours indicate positive cross peaks. Positive cross peaks indicate two wavenumbers with perturbation induced intensity changes in the same direction (increasing or decreasing together). Negative cross peaks indicate two wavenumbers with perturbation induced intensity changes in opposite directions (one increasing and the other decreasing Contours labelled A are autopeaks while contours labelled C, are cross peaks.

An autopeak always has a positive magnitude and reveals the degree of spectral

intensity variation at a specific spectral wavenumber. A plot of the values of the

autopeaks is referred to as an autocorrelation plot. A high intensity autopeak

implies a significant amount of change in that spectral region and *vice versa*. An

autocorrelation plot is a useful tool for extracting an overview of perturbation

induced changes across the full data set. Cross peaks indicate co-ordinated

spectral changes observed at two different spectral variables, v_1 and v_2. The sign of a cross peak can either be positive or negative. It is positive if the spectral intensities at the two spectral wavenumbers corresponding to the coordinates of the cross peak are both increasing or decreasing as a function of the external perturbation. It is negative if one of the spectral intensities is decreasing, and the other is increasing (Noda, 1993, Noda et al., 2000).

1.6.3 Asynchronous Plot

The asynchronous plot is the imaginary component of the 2D-correlation function (Noda, 1993). It shows the degree of dissimilarity between two distinct spectral intensity variations occurring at two independent wavenumbers (v_1, v_2). It has no autopeaks, only cross peaks. A schematic is also shown in Figure 1.6. These cross peaks only appear if the spectral changes occurring at distinct spectral variables are different. This feature is useful for differentiating overlapping bands (Noda, 1993, Noda and Ozaki, 2005). The interpretation of the asynchronous spectrum requires a knowledge of the sign of the corresponding cross peak in the synchronous spectrum. If a spectral intensity change occurs predominantly at v_1 before v_2, the sign of the associated cross peaks in both the synchronous and asynchronous plots are the same. If the reverse occurs, the cross peak signs are different (Noda, 1986, Noda, 2012). If the spectral intensity change occurs simultaneously at v_1 and v_2, only synchronous cross peaks will be observed. The order of spectral intensity change is indeterminate in cases where only an asynchronous relative intensity is observed with no corresponding synchronous cross peaks. These rules that govern the interpretation of sequential order intensity changes are known as Noda's rules. The usefulness of these rules

depends on the extent to which the cross peaks in both the synchronous and asynchronous plots can be identified and interpreted. The occurrence of multiple trends or the complexity of the data such as the formation of intermediates species, overlapped responses may result in 2DCA plots with weak and congested cross peaks making it challenging to interpret the sequential order of intensity changes unambiguously (Ashton and Blanch, 2010, Noda, 2014b).

Furthermore, Noda's rules are limited to determining the sequential order of intensity changes. They do not account for the sequential order of the population of contributing species which can provide more insight into the mechanisms associated with the formation, or disappearances of species or structures (Noda, 2014b, Noda, 2016a). Two recent 2DCA techniques seek to address these problems. The first is the merged correlation spectrum combines the synchronous and asynchronous plot with the advantage of an improved resolution as well as a straightforward interpretation of the sequential order intensity variations (Noda, 2016). The second is the two-dimensional co-distribution spectroscopy which helps to identify the order in which distinct populations of species appear (Noda, 2014b). The benefits of these two techniques are yet to be widely explored for general applications as well as in the context of protein research.

1.6.4 Two-Dimensional Perturbation-Correlation Moving-Windows (2DPCMW)

A more extensively used 2DCA method that attempts to overcome the complexity of interpreting congested 2DCA plots is the two-dimensional perturbation-correlation moving-windows (2DPCMW). 2DPCMW is a 2D-correlation technique used to monitor spectral variations along the perturbation axis. It directly relates

the spectral intensity change to a variation in perturbation (Morita, 2006). 2DPCMW builds on an earlier 2DCA technique, two-dimensional moving windows analysis (2DMW) which involves dividing the data set into a number of small sections referred to as windows (Thomas and Richardson, 2000). For example, in a data set of 15 spectra using a window size of 4, spectra will be grouped into 1st-4th; 2nd-5th; 3rd-6th; 4th-7th; 5th–8th and 6th-9th until all spectra fit. This segmentation is particularly useful in the study of complex protein data where determination of the direction of sequential intensity changes are unclear or challenging to follow (Noda, 2017). 2DMW analysis creates a moving window to distinguish spectral intensity changes across the perturbation axis. However, the resulting plot does not indicate the direction in which the spectral intensity is changing.
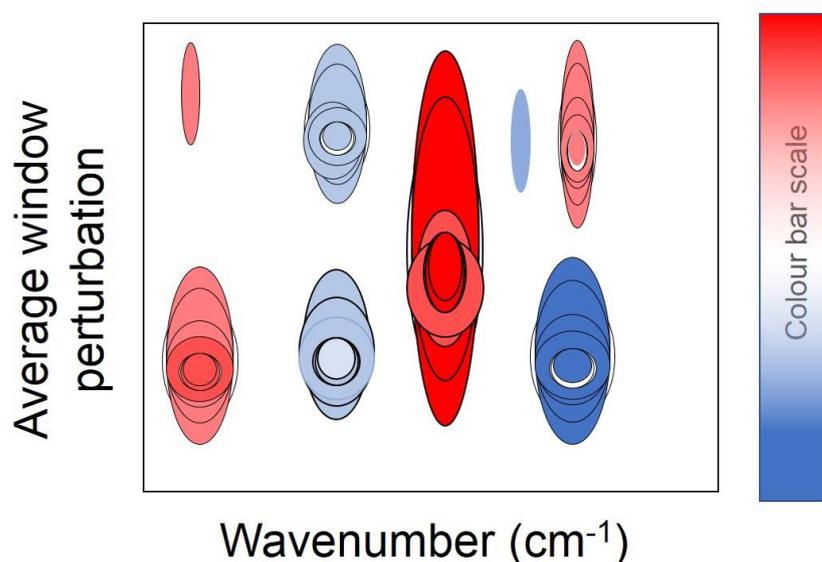
Figure 1.7 Schematic showing a 2DPCMW plot. Positive contours are shaded in red while negative contours are shaded in blue. Positive contours indicate increasing peak intensities in the direction of the perturbation while negative contours indicate decreasing peak intensities in relation to the perturbation. A darker shade and an increased number of contours indicate a relatively greater change in intensity. The scale on the colour bar is arbitrary.

Furthermore, the 2DMW requires careful mathematical adjustments to separately describe synchronous and synchronous correlation intensities (Morita, 2006). 2DPCMW creates a similar moving window but uniquely incorporates the synchronous and asynchronous aspects of the 2D-correlation function, thereby maintaining the central concept of the 2D-correlation analysis in addition to relating the direction of spectral intensity changes to the perturbation (Morita, 2006, Lee, 2017). A 2DPCMW synchronous plot provides insight into how spectral changes relate to the applied perturbation whilst the 2DPCMW asynchronous plot gives a more accurate indication of the transition points in a data set (Noda, 2017). The smaller identified regions can then be further analysed using the synchronous and asynchronous plots. A schematic of a 2DPCMW contour map plotted as the average window perturbation against the wavenumber is shown in Figure 1.7.

From the outlined 2DCA techniques and in general, 2DCA may be viewed as comprising a range of tools that can be selected and applied alone or in combination based on a working knowledge of any intrinsic limitations and according to user requirements.

1.7 Protein Studies

Proteins are the most commonly investigated biomolecule using 2DCA (Park et al., 2018). The enhanced visualisation of spectral changes that can be achieved using 2DCA is particularly attractive for the elucidation of the complex and intricate structural changes that may occur as part of a proteins' response to an applied perturbation (Noda, 2017, Wu et al., 2018). One of the earliest protein 2DCA and Raman spectroscopic study examined changes in the secondary structure of β-

lactoglobulin as a result of increasing concentration. Using 2DCA, they identified and proposed assignments for overlapped peaks in the amide III region (1200-1340 cm^{-1}). Furthermore, from the asynchronous plots, it was possible to identify the sequential order of intensity changes corresponding to the secondary structural elements. These results were consistent with the 2DCA and IR spectroscopy findings in the same study (Jung et al., 2000). Another concentration-related study investigated the concentration-dependent entanglement of the glycoprotein, mucin. They identified an essential residue involved in the entanglement transition using spectral variations analysed with 2DPCMW (Ashton et al., 2013a). This observation was in agreement with previous studies using the considerably higher resolution technique, NMR (Coltart et al., 2002). Although it would be advantageous to use a high-resolution technique, NMR is not routinely used in the monitoring of protein structure during drug development due to constraints including long processing times and low size limits of ~25kDa (Arzhantsev et al., 2012, Houde and Berkowitz, 2014).

There have also been a few studies investigating transition states and intermediates in proteins and polypeptides. A study involving silk fibroin identified the optimum quantity of calcium and pH necessary to produce conformations of silk I, II and their intermediates (Zhou et al., 2004). Ashton and Blanch (2010) identified three pH-induced transition states in α-lactalbumin utilizing Raman spectroscopy and 2DMW. However, a later study by Park et al. (2016a) identified only two transition states for α-lactalbumin; pH (1-3) corresponding to the acid state and pH (4-7) corresponding to the native state.

In the former study, 2DMW which is plotted as the average window pH versus wavenumber was used to identify three transition states while in the latter study, a PCA scores plot as a function of pH was used to identify the transition states. The point of transition from the acid to native-state as deduced from the PCA scores plot was pH 3.5 which corresponds to the beginning of the middle transition point (~pH 3.6-4.6) identified by Ashton and Blanch (2010). The identification of this intermediate transition state by the moving windows plot shows the potential of 2DCA to identify subtle spectral variations. The presence of three transition states has also been confirmed by vibrational circular dichroism (VCD) 2D gradient mapping (Ryu et al., 2012).

In contrast to their fluorescence spectroscopic identification of one transition state for RNase A, Brewster et al. (2013) confirmed the presence of two transitions states in the amide I region of RNase A using 2DMW and Raman spectroscopy. The 2DMW results also provided more clarity about the number of transition states in RNase A compared to the folding curves obtained from Raman spectra alone.

Although most of the documented studies have involved model proteins, there have been a few involving more complex proteins such as antibodies and heme proteins which have more physiologic and or therapeutic relevance. Lu et al. (2014) monitored the concentration of methaemoglobin in blood in varying concentrations. Their Raman spectra showed differences between high, medium and low concentrations. However, by using 2DCA, peaks not evident from the Raman spectra were identified and assigned to the change from haemoglobin to methaemoglobin. Gömez De La Cuesta et al. (2014) studied temperature-induced aggregation in antibodies with varying aggregation tendencies. Despite the low

signal to noise ratio of the acquired Raman spectra, their use of 2DPCMW enabled the identification of sample variations linked to conformational transitions and aggregation.

The outlined studies demonstrated the capacity of 2DCA and Raman spectroscopy to probe and interpret dynamic structural changes of proteins primarily in a homogenous environment. Beyond the complexity that proteins possess, biopharmaceuticals require formulation into a heterogeneous system. Therefore, it is expected that some of the Raman spectral variations may arise from non-protein constituents in the sample environment. This challenge has been highlighted in the monitoring of a recombinant antibody production using Ultraviolet Resonance Raman spectroscopy (UVRR) and 2DMW. The spectra showed peaks of various components in the changing cell medium, making it difficult to follow changes in the main protein component (Ashton et al., 2013b). However, by the application of 2DMW, it was possible to relate the spectral variations to time as well as identify changes in protein and nucleic acid bands thus demonstrating the role of 2DCA in the visualization of complex spectra.

1.8 Conclusion

The complexity of protein structure and its marginal stability remains a challenge to protein drug development. There is a demand to continue developing analytical techniques that address these constraints. As discussed, protein vibrations detected by Raman spectroscopy can provide insight concerning structural variations arising from an applied perturbation. The application of 2DCA is uniquely designed to identify and extract such changes even when they are not

apparent from visual observation of the spectra. While 2DCA has been used for over two decades to study proteins, the novelty of this research is the investigation of its specific application in combination with Raman spectroscopy as a conformational stability screening tool during early protein drug development.

This project seeks to develop this technique to meet the analytical demands of monitoring protein conformational stability in the biopharmaceutical industry.

A thorough understanding of the theory of the 2DCA technique, as well as the analytical probe (Raman spectroscopy), is required to arrive at the correct interpretation of the data and subsequent conclusions. Furthermore, their application in any new context might present challenges that need to be resolved as part of developing the technique.

2 Materials and Methods

2.1 Samples

Two groups of protein samples were investigated in this study. The first group (chapter 3 and 4) was made of four model proteins purchased as lyophilised powders from Sigma Aldrich and used without further purification. The samples were dissolved in deionized or Milli-Q water and examined as dried films or solutions. The second group (chapter 5 and 6) comprised of five therapeutically relevant samples (liquid and freeze-dried) and three buffers of varying pH obtained from CIMEMT, UCL London. The freeze-dried samples required no additional preparation while the liquid samples were prepared by mixing in buffer or water to obtain the desired concentration.

2.2 Raman Spectra

2.2.1 Dried Films

Spectra of the proteins were collected using a Renishaw inVia Raman microscope (Renishaw plc, Wotton-Under Edge, UK) coupled with excitation wavelengths of 532 or 785 nm and a x50 objective. Diffraction gratings were 1200 and 2400 l/mm for the 532 and 785 nm lasers respectively. Wavelength calibration was performed using a 1s static spectrum of a silicon wafer centred at $\sim 520 \pm 0.5$ cm^{-1}.

5 µl of aqueous protein solutions were spotted onto a calcium fluoride window and left to dry at room temperature for 2 hours. Spotting was repeated once over the previous area to increase the amount of sample on the window. Spectra were

collected from varying points across the dried film. The spectral parameters for each sample are described in the respective chapters.

2.2.2 Solutions

All Raman spectra were collected using the previously stated Renishaw inVia confocal Raman microscope, coupled to an excitation wavelength of 785 nm. Instrument calibration was performed as described in section 2.2.1. For the pH, studies, spectra of aqueous protein solutions (400 μl) were obtained using a 96-well quartz plate and a long-distance x15 objective. Exposure time was 5s with 360 accumulations (total collection time of 30 min) and ~60 mW laser power at the source. The parameters for the temperature-dependent Raman spectra and UVRR spectra are described in the respective chapters.

2.3 Data Preprocessing

It was necessary to perform various preprocessing procedures on the raw data to reduce the influence of interfering effects making the data more suitable for analysis. Furthermore, preprocessing methods have been shown to improve the quality of 2DCA plots significantly, making them clearer to interpret (Ashton et al., 2008). Most of the procedures were carried out using MATLAB R2016a. Water subtraction, calculation of peak height and scaling value were performed using Microsoft Excel 2010. Cosmic rays were first removed from the spectra using the WiRE 4.2 software (Renishaw Plc) before any further preprocessing. The order used for the application of the remaining preprocessing methods was decided by trialling a combination of methods in different sequences and observing their effect on the data set.
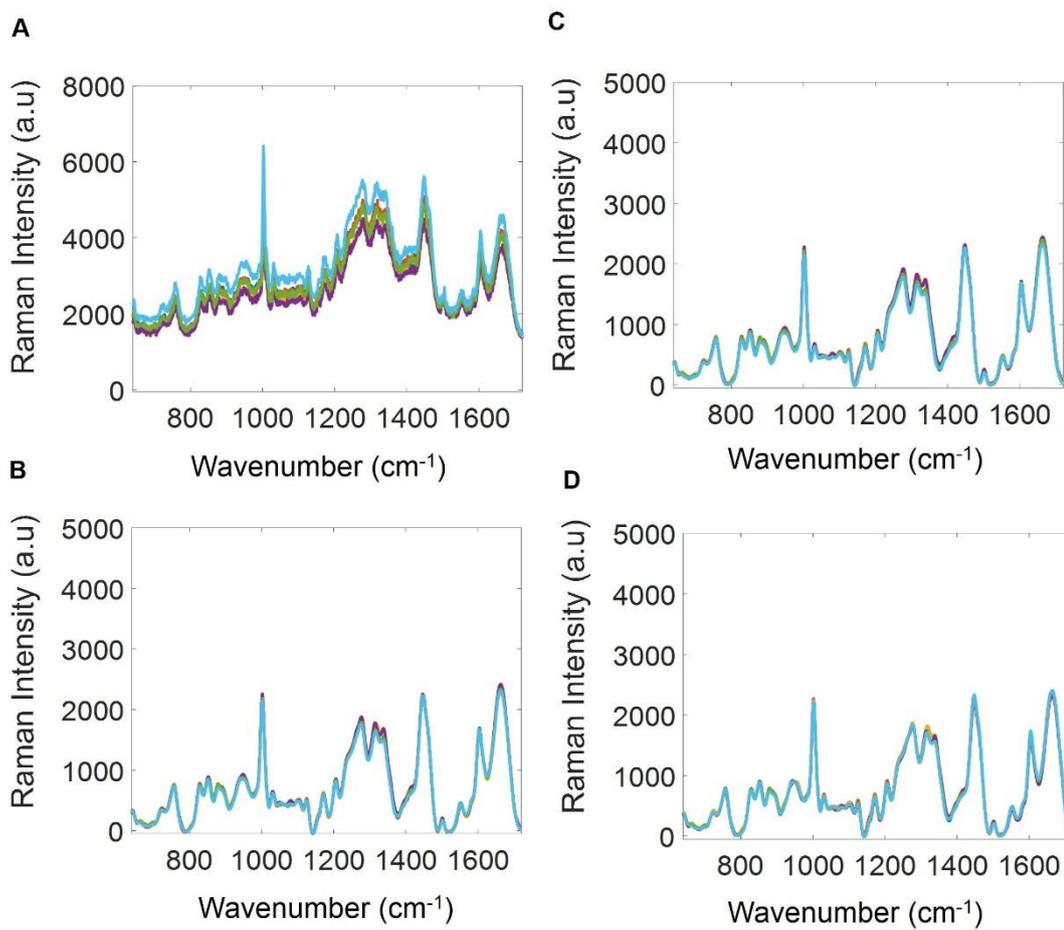
Figure 2.1 Preprocessing sequences trialled for the Raman spectra of dried film protein samples (A) Raw (B) baseline correction, smoothing and normalisation (C) smoothing, normalisation and baseline correction (D) baseline correction, normalisation and smoothing.

Although there is no fixed method for deciding on the order to preprocess a set of data, the order used may influence the results (Lasch, 2012, Engel et al., 2013). It is therefore essential to consider how a preprocessing method or the order of application may affect the final interpretation of the data.

Despite being a subjective approach, visual inspection is the most commonly used method due to ease and accessibility (Lasch, 2012, Engel et al., 2013). However, the volume of the data could make it impracticable to observe subtle variations in

the data. Visual inspection was performed after each preprocessing step to ascertain the degree of spectral overlap and identify any artefacts or differences in spectra.

For the dried film sample spectra, at least three different sequences of preprocessing methods (normalization, baseline correction and smoothing) produced very similar effects (Figure 2.1). The final order chosen was, baseline correction, normalization, smoothing and averaging. Similar trials were also performed for the solution spectra. However, the effect on the data proved to be more sequence-dependent than in the case of the dried film samples. Therefore, unless otherwise stated elsewhere in this thesis, the final preprocessing order for the solution spectra was baseline correction, smoothing, followed by normalization, water subtraction, further baseline correction and averaging.

2.3.1 Baseline Correction

Baseline correction was performed with asymmetric least squares (Eilers and Boelens, 2005) to reduce the effects of background interference in the spectra. In the case of solutions, it was observed that baseline correction was the preferred first step because the uniform baseline resulted in minimal adjustments of the water spectral intensities before water subtraction (Figure 2.2A and 2.2B). Further baseline subtraction was performed after water subtraction to reduce baseline variations following water subtraction. Additional baseline subtraction has also been shown to enhance the clarity of the 2D-correlation plots (Ashton et al., 2008).

2.3.2 Smoothing

The smoothing procedure was carried out using a fast-smoothing function (Haver, 2006) to reduce the noise present in spectra. For the solution spectra, smoothing followed by normalization also proved a suitable option because it was easier to determine the peak intensity range after smoothing of the spectra rather than before smoothing is applied (Figure 2.2C and D).
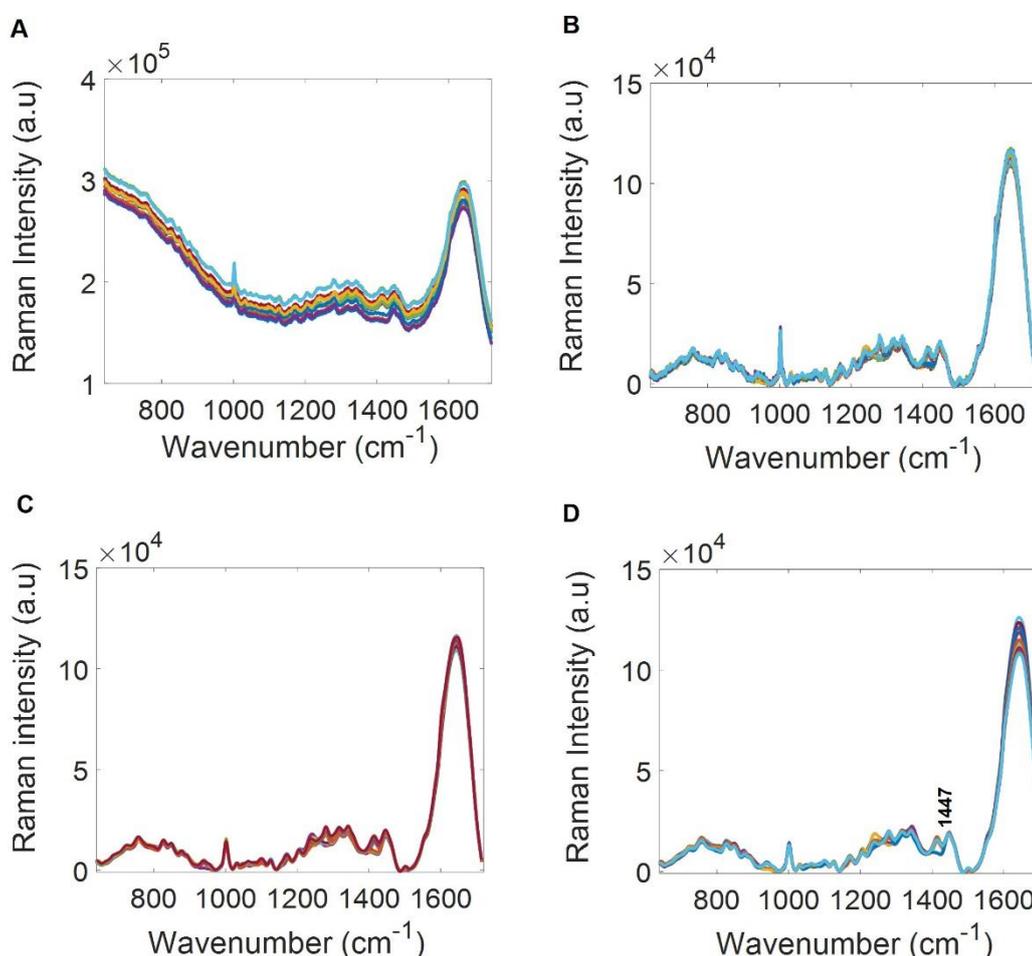
Figure 2.2 Preprocessing sequence for solution spectra before water subtraction (A) Raw (B) baseline corrected (C) baseline corrected and smoothed (D) baseline corrected, smoothed and normalised to ~1447cm^{-1}.

2.3.3 Normalisation

Normalization was carried by scaling the peak height at ~1447 cm^{-1} assigned to methylene deformation (Table1.1). Peak heights were calculated from the difference in maximum and minimum peak intensities. Scaling of spectral intensities allows spectra possessing varying intensities due to instrumental fluctuations to be compared objectively. The peak at ~1447 cm^{-1} is a widely used choice for normalisation because it is reported to be minimally influenced by peptide backbone conformation due to its association with amino acid sidechain structure, external to the peptide backbone (Church et al., 1998, Schlücker et al., 2006, de Vasconcelos Nasser Caetano et al., 2017). The standard normal variate (SNV) method of normalisation discussed in section 6.2.3 was also used in this study. For the water samples, only baseline correction and smoothing were applied. The water spectra were excluded from the normalisation procedure because they do not have the ~1447 cm^{-1} peak.

2.3.4 Water Subtraction

Water subtraction was carried out in order to remove the solvent contribution to the spectra. Although water has a weak Raman effect, it gives a broad peak at ~1640 cm^{-1} in biological samples (Lord and Yu, 1970, Maeda and Kitano, 1995, Smith and Dent, 2005). The high proportion of water compared to proteins in low concentrated solutions makes it difficult to observe other spectral features arising from proteins peaks in the same region (Figure 2.3A and 2.3C). For water subtraction to be effectively carried out, the appropriate intensity of the water spectrum to be subtracted must be determined. Due to experimental limitations

or the intrinsic nature of the protein solution, the peak intensity at ~1640 cm^{-1} could vary between the protein and the water spectra.
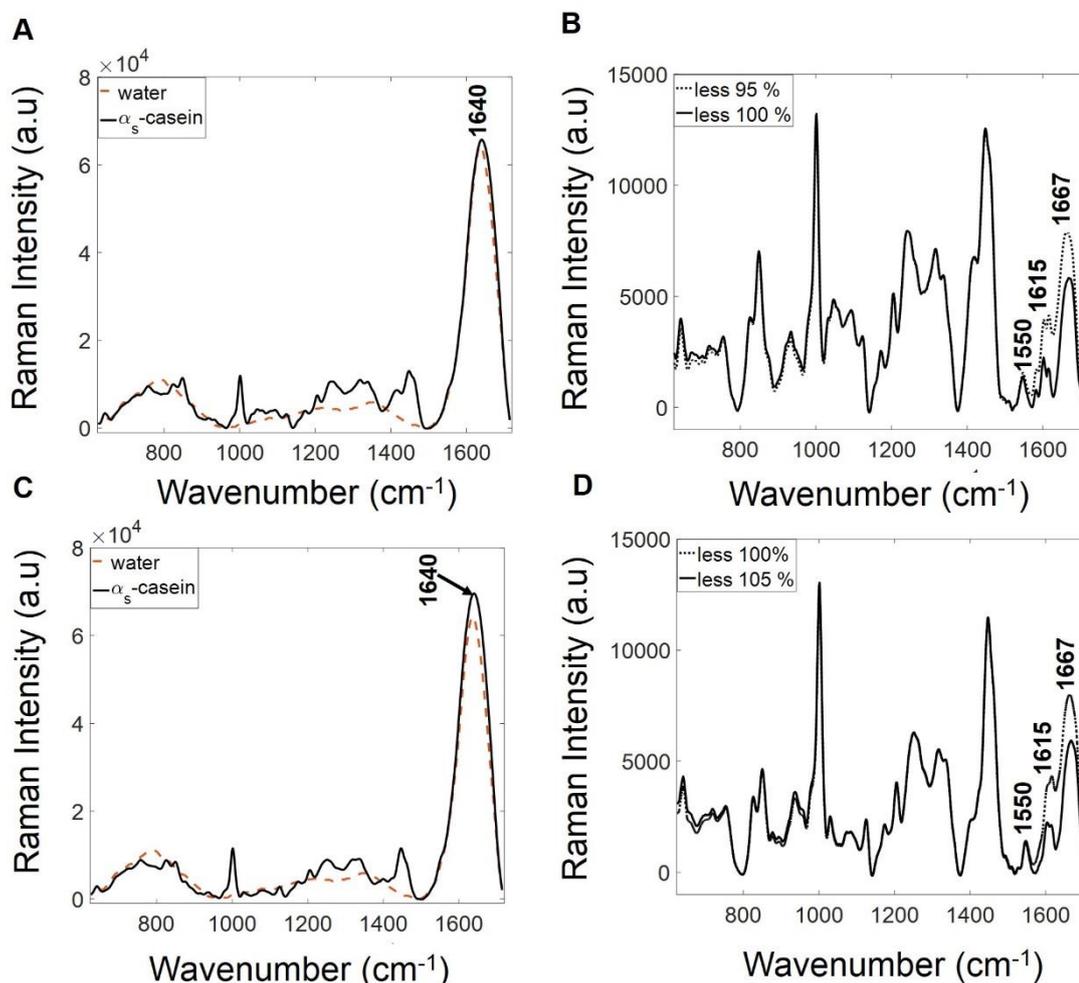
Figure 2.3 Water subtraction from the Raman spectra of α_s-casein (10 mg/ml) in water. (A) α_s-casein and pure water with similar peak intensities at ~1640 cm^{-1} (B) α_s-casein in water less 95 and 100% of water spectrum (C) α_s-casein and pure water with varying peak intensities at ~1640 cm^{-1} (D) α_s-casein in water less 100 and 105% of water spectrum.

If the peak intensity at ~1640 cm^{-1} varies slightly between the protein and water spectrum, the water spectrum may be subtracted without any adjustments. In Figure 2.3A, the α_s-casein and water spectrum have similar peak intensities at ~1640 cm^{-1}. Subtracting the entire water spectrum without adjustments produced

a spectrum that revealed the formerly obscure peaks at ~1550, 1615 and 1667 cm^{-1} (Figure 2.3B). Subtracting a reduced amount of the water spectrum (95%) produces a spectrum with an increased baseline at ~1615 and 1667 cm^{-1} (Figure 2.3B). If the peak intensity at ~1640 cm^{-1} varies considerably between the protein and water spectra, then the maximum intensities of the water spectrum are adjusted depending on whether the water spectrum has the lower or higher intensity. In Figure 2.3C, the water spectrum has a lower peak intensity at ~1640 cm^{-1} compared to the same peak in the protein spectrum. In this case, a higher amount of the water spectrum (105%) was subtracted (Figure 2.3D). By subtracting an increased proportion of the water spectrum (105%), it can be observed that the resulting spectrum has peaks at ~1615 and 1667 cm^{-1} with lower intensities and a baseline that is uniform to that of other peaks in the spectrum. Despite further baseline correction after water subtraction, intensity variations were commonly observed above ~1600 cm^{-1}. These fluctuations were easily minimized by repeating the water subtraction and using the intensity of the α_s-casein~1615 cm^{-1} peak (or for other samples any of the peaks between ~1600-1615 cm^{-1}) as a marker to ensure uniform subtractions not only for a set of repeat spectra but across the entire data set.

An alternative to the adjustments made to the water spectrum would be to normalise the protein and water spectra to the ~1640 cm^{-1} peak. However, this is not an attractive option because as seen in Figure 2.3B and 2.3D, there are hidden peaks in this region of the protein spectra. Consequently, such normalization could compromise the observation of any spectral variations arising from these peaks.

2.4 Temperature

2.4.1 Linkam Temperature Stage

For the temperature investigations, the samples were heated from ~28 to 65 °C, and a spectrum was simultaneously acquired at defined temperature points. Heating and temperature control were achieved using a Linkam stage (THMS600), Linkam scientific instruments Ltd, Waterfield, Epsom, UK). It was necessary to adapt the original sample holder for the Linkam stage due to evaporation and condensation observed during heating. The occurrence of these processes in the limited volume of liquid in the sample holder (200 μl), led to the sample not being retained long enough for spectral acquisition throughout the experiment.
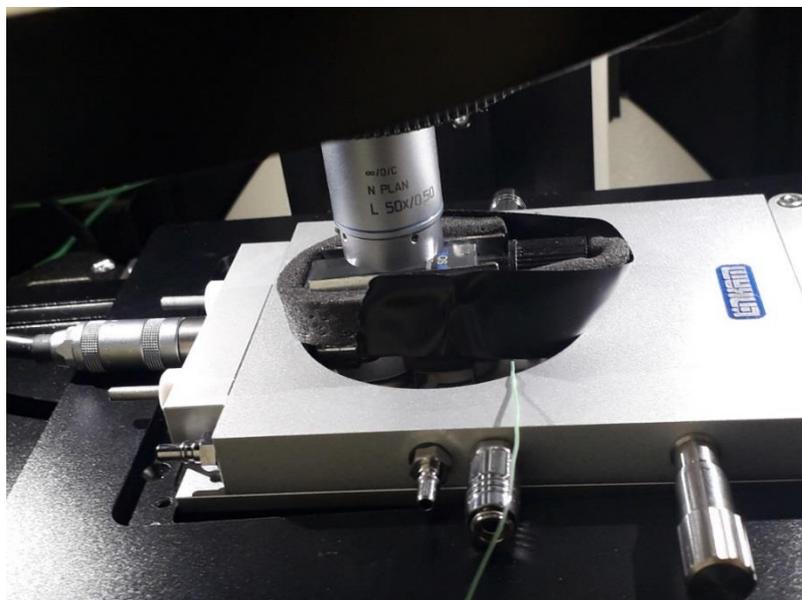
Figure 2.4 Photograph of adapted Linkam temperature stage (THMS600). Insulated 3ml sealed cuvette is placed on the Linkam stage on the Renishaw invia Raman microscope stage under a long-distance objective. Thermocouple (green cable) is inserted into the cuvette.

In order to overcome these challenges, quartz cuvettes with volumes ranging from 500-1000 µl were sealed using insulation tapes and trialled as sample holders. The heating surface area was maximised by placing the cuvette in a horizontal position.

Although this approach solved the issues of evaporation and condensation, it introduced a longer heating time due to the increased volume and the exposure of the metal plate to ambient temperature. Furthermore, the sealing of cuvettes using insulation tapes did not prevent leakage of the sample. Keeping the cuvette in an upright position would have overcome the leakage problem, but this position would limit the heated area of the cuvette. The final option was to use a 3 ml cuvette with a screw lid. The cuvette was placed in a stainless-steel holder covered with insulated foam secured by tape, as shown in Figure 2.4. The main drawback of this arrangement was a further prolongation of the heating time, making it challenging to obtain temperatures above 65 °C.

2.4.2 Calibration of Temperature Stage

For calibration, the cuvette containing Milli-Q water was positioned on the temperature stage and placed under the microscope in the Raman spectrometer (Figure 2.4). A spectrum of the sample was collected at ~20 °C before the temperature stage setting was ramped from ~28 to 104 °C at a rate of ~0.5 °C per min. A thermocouple (Oko Lab) was inserted into the liquid to measure its temperature for the entire duration of the experiment. Initial holding time was 30 mins at ~28 °C and 15 mins at subsequent temperatures. After the holding period, a single spectrum was collected at the same temperature for 30 mins.

The calibration curve is shown in Figure 2.5. It is linear but shows increasing standard deviations of up to ± 2.4 °C at higher temperatures. The increased variation is likely due to a more significant influence of the ambient temperature at a high temperatures in the adapted set-up.

Although this observation could have negative implications for reproducibility, this study was deemed feasible because the investigation involved the comparison of sample responses across the temperature range and not necessarily a focus on specific structural changes observed at a single temperature.
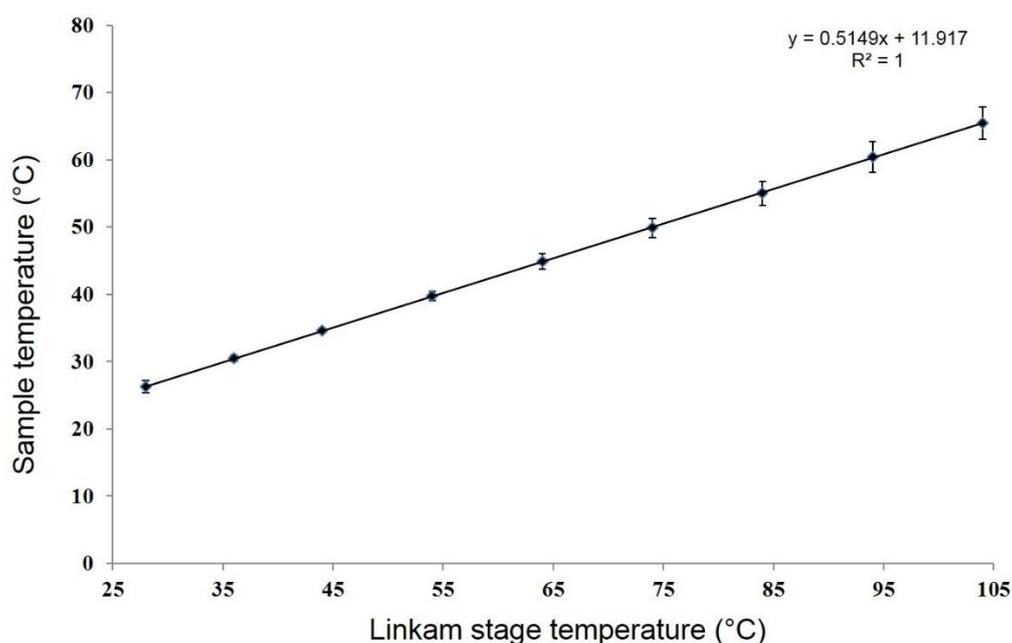
Figure 2.5 Calibration curve for the Linkam temperature stage (THMS600). Graph values are the mean of four repeats. Error bars represent mean standard deviation values. Sample temperatures ranged from 26-65 °C.

2.5 2D-Correlation Analysis (2DCA)

2.5.1 Interpolation

The computation method of the 2DCA technique requires the perturbation range to vary in steps with fixed increments (Noda & Ozaki, 2005). The perturbation range for each sample used in this study is unevenly spaced and so, does not meet this criterion. Therefore, an interpolation procedure (Savitzky-Golay spline estimation) was performed on the preprocessed data using MATLAB R2016a to modify the data to an evenly spaced one. This procedure assumes that the perturbed system undergoes minimal changes across the intervals between the original data points (Ashton et al., 2008). If these intervals are large, this assumption may become somewhat unreliable. In this experiment, significant variations in spectral intensity were not observed between sample intervals; hence, it was appropriate to use the interpolation procedure. The interpolated data steps for each sample are detailed in the respective chapters.

2.5.2 Synchronous, Asynchronous and 2DPCMW plots.

2D shige software developed by Prof. Shigeaki Morita (freely available from https://sites.google.com/site/Shigemorita/home/2dshige) was used to create synchronous and asynchronous plots. The autocorrelation and 2DPCMW were also obtained using this software before the subsequent generation of the respective plots using MATLAB R2016a. The 2DPCMW technique (section 1.6.4) was applied in order to relate the spectral intensity variation to pH or temperature. In this method, the interpolated data was divided into small subsets

(windows). The specific window size applied as well as the maximum level of contours is stated for each sample in the respective chapters.

3 α_s-Casein Model

3.1 Introduction

For several years, casein proteins have attracted significant research interest for reasons including their abundance and availability from milk, utility in the food industry, drug delivery applications and their unique natively unfolded structure (Swaisgood, 2003, Holland and Boland, 2014, Elzoghby et al., 2012, Horne, 2002). Caseins are a family of phosphorylated proteins broadly designated as α_{s1}, α_{s2}, β and κ based on their amino acid sequences (Farrell et al., 2004, Holt et al., 2013). Phosphorylation of casein amino residues is a post-translational modification (PTM) that is vital for casein's biological functions such as the stabilization of calcium phosphate complexes in micelles, formation of micelle internal structure, molecular chaperone activities and the maintenance of a flexible and open structure (Holt et al., 2013, Holland and Boland, 2014). α_s-casein is comprised of α_{s1} and α_{s2} forms of casein. α_{s1} usually has phosphate residues on 8 or 9 serine amino acids, while α_{s2}-casein can have up to 13 phosphate groups (Farrell et al., 2004, Holland and Boland, 2014). They exist in bovine milk in a ratio of 4:1.

As with other casein proteins, α_s-casein is yet to be crystallised thus insight into their secondary structure has mainly been obtained from spectroscopic studies (Kumosinki et al., 1991, Alaimo et al., 1999). However, there have been conflicting result regarding quantities of specific structure types even by the same technique (Horne, 2002). A Raman study of α_{s1}-casein reported small amounts of β sheet (20%) and α-helix (13%) and higher amounts of turns (33%) and irregular structures (33%)(Byler et al., 1988). CD and FTIR studies of α_{s2} casein reported

about (24-32%) α-helix, (30%) β-sheet, 24% turns and 22% irregular structure (Hoagland et al., 2001). Smyth et al. (2001) remarked that their characterisation of α_s-casein was inadequate because the sample had unknown amounts of both isoforms. Nevertheless, in addition to α-helix and β-structure they identified the polyproline structure (PPII) which is a frequently occurring conformation in unfolded proteins (Thorn et al., 2015). Although a variety of amino acids can form the PPII structure, Pro is more likely to form this conformation and they are strongly associated with the open unfolded structure of caseins due to their ability to disrupt regular structures (Horne, 2017).

The assessment of the influence on a PTM on protein stability is a crucial aspect of drug development due to their effect on protein conformation and function (Walsh, 2010, Walsh and Jefferis, 2006). Raman spectroscopic studies comparing dephosphorylated and phosphorylated forms of α_s-casein have suggested that the absence of phosphate residues is associated with conformational changes (Jarvis et al., 2007, Ashton et al., 2011). Chakraborty and Basak, (2007) studied pH-induced structural transitions in α_s-casein using Fluorescence and CD spectroscopies. They demonstrated that at high alkaline pH there was extensive loss of secondary structure whereas at acidic pH 4-5, changes in secondary structure were minimal. They attributed this observation to increased inter residue repulsion as a result of increasing high negative charge at alkaline pH and reduced repulsion due to an initial neutralisation of the high negative charge at moderately acidic pH, resulting in a compact folded structure ~pH 4-5. Their findings are consistent with Liu & Gou (2008) who observed a more compact casein micelle structure at ~pH 5.5 compared to at ~pH 12. However, both studies

did not examine the structural changes in the absence of phosphate groups. In this study, α_s-casein has been used as a model to investigate the influence of a PTM on the pH-dependent conformational stability of a protein. The aim is to investigate the application of Raman spectroscopy and 2DCA to the study of protein conformational stability in the presence and absence of a specific PTM. Raman spectral variations in phosphorylated α_s-casein and dephosphorylated α_s-casein (dα_s-casein) were initially compared in solid and liquid sample states. pH-induced Raman spectral variations were then analysed using 2DPCMW as well as synchronous and asynchronous plots.

3.2 Experimental

3.2.1 Sample Preparation

3.2.1.1 Dried Film Samples

Samples of α_s-casein and dα_s-casein (bovine milk) were purchased from Sigma Aldrich and used without further purification. 1 mg/ml solutions were prepared by dissolving the lyophilised powder in deionised water. Dried films were prepared on a calcium fluoride window as described in section 2.2.1.

3.2.1.2 Liquid Samples

α_s-casein and dα_s-casein (bovine milk) samples of different pH were prepared by dissolving the lyophilised powder in deionised water and then adjusting the pH using dilute NaOH or HCl. The final concentration of each sample was ~10 mg/ml. The pH of each sample (Table 3.1) was measured using a pH meter (HI2210, Hanna Instruments). No pH adjustment was made for native samples of α_s-casein (pH 6.5) and dα_s-casein (pH 7).

Table 3.1 pH values (± 0.2) for α_s-casein and dα_s-casein solutions.

α_s-casein	5.7	6.5	7.7	8.7	9.2	10.2	11.1	12.1
dα_s-casein	6.1	7.0	7.6	7.7	8.9	10.1	10.8	12.3

3.2.2 Spectra Collection

All liquid sample Raman spectra were collected as described in section 2.2.2. A total of 6 spectra were acquired for the dried film samples using ~15 mW laser power at source. The exposure time was 5s with 6 accumulations (collection time of 30s).

3.2.3 Data Preprocessing

Spectral preprocessing was carried out as discussed in section 2.3. For each pH value, the spectra were averaged (n=3), except for the pH 10.2 spectra where n=1 due to distortion at ~1065 cm^{-1} in two of the spectra. For the generation of 2DPCMW plots, spectra were interpolated using the Savitzky-Golay spline estimation to generate evenly spaced interpolated data in steps of ~0.4.

3.3 Results & Discussion

3.3.1 Dried Film Samples

Figure 3.2 shows the dried film spectra of α_s-casein and dα_s-casein. Table 1.1 lists the proposed assignments for the observed peaks. A comparison of their spectra reveals that most of the peak intensities vary minimally between the phosphorylated and dephosphorylated spectra.
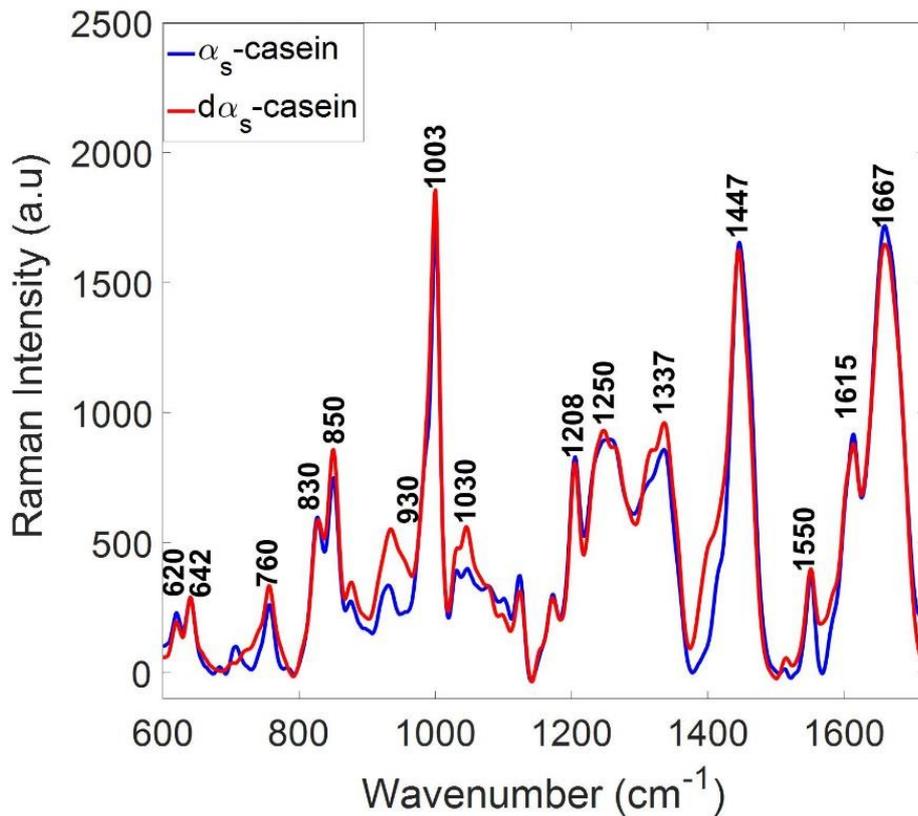
Figure 3.1 Averaged (n=6) Raman spectra of 10 mg/ml α_s-casein and dα_s-casein dried films acquired on a calcium fluoride window. Preprocessing was carried out by baseline correction, normalization and smoothing. Each Raman spectrum was collected in 30s using an excitation wavelength of 532 nm. Assignments are listed in Table 1.1.

The decreased peak intensity at ~930 cm^{-1} suggests a lower helical content in the phosphorylated sample (Monti et al., 1998, Ashton et al., 2007, Beattie et al., 2008) while the peak intensity variations at 1030 cm^{-1} assigned to Phe residues suggest a difference between the Phe residues in the α_s-casein and dα_s-casein samples. A higher number of Phe residues in one sample could lead to a higher intensity of the Raman spectra. In each sample, there may be varying amounts of Phe residues based on the ratio of α_{s1}-casein (n=8) to α_{s2}-casein (n=6) (Morris, 2002, Horne, 2017). However, the exact ratio of α_{s1} to α_{s2}-casein in each sample is unknown and could not be confirmed by the supplier.

The peak intensity variations at 1337 and 1667 cm^{-1} suggest changes in Trp/α-helix and β-sheet respectively. However, these changes reveal some apparent contribution from baseline fluctuations and cannot be unambiguously attributed to structural differences between α$_s$-casein and dα$_s$-casein. In their study of α$_s$-casein and dα$_s$-casein, Jarvis et al. (2007) also observed differences in the spectral intensity of phosphorylated casein compared to dephosphorylated casein (Figure 3.2). They observed that compared to the α$_s$-casein, dα$_s$-casein had increased intensities at ~989, 1003, and 1400 cm^{-1} which were assigned to Tyr, Phe and His sidechain conformation respectively. Furthermore, decreased peak intensities were observed at 1250, 1265 and 1655-1670 cm^{-1} in the dα$_s$-casein spectrum compared to the α$_s$-casein spectrum. Apart from the decreased spectral intensity at ~1655-1670 for dα$_s$-casein, the spectral intensity variations in this study are generally not consistent with the Jarvis et al., 2007 study. Another notable difference between both studies is that a peak at ~989 cm^{-1} was hardly detectable in Figure 3.1. In contrast, it was observed to increase in the dephosphorylated casein sample in the Jarvis et al., 2007 study (Figure 3.2). Their assignment of the peak at ~989 cm^{-1} to Tyr does not correspond to any established Raman markers of Tyr (Hildebrandt et al., 1988, Takeuchi, 2011, Hernández et al., 2016). A peak at ~990 cm^{-1} observed in the Raman spectra of a protein buffered in sodium phosphate (section 5.3.1) has been assigned to a phosphate stretch vibration. However, the presence of a phosphate group in the dα$_s$-casein samples is contrary to what would be expected for a dephosphorylated protein. Therefore the ~989 cm^{-1} peak cannot be reassigned to a phosphate vibration.
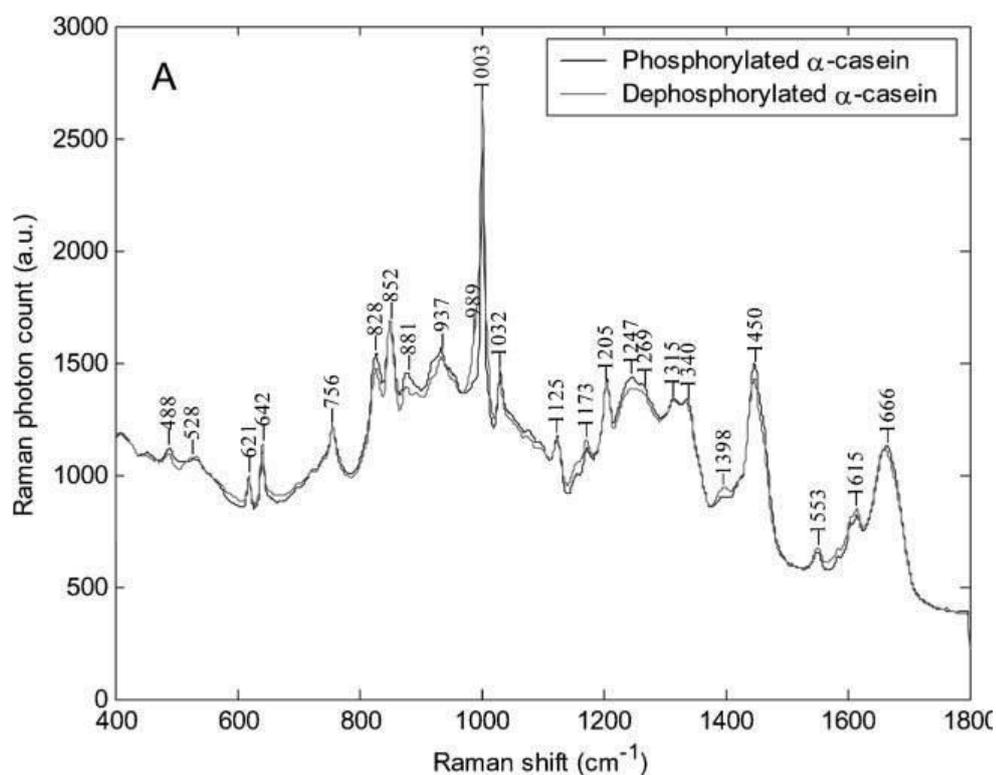
Figure 3.2 Averaged spectrum (n=12) of α_s-casein and dα_s-casein acquired on a hydrophobic slide. Republished with permission of the Royal Society of Chemistry from Quantification of Casein phosphorylation with conformational interpretation using Raman spectroscopy by Jarvis et al., 2007. Permission conveyed through Copyright Clearance centre Inc.

One reason for the dissimilarity of the results from both studies could be the use of a non-hydrophobic slide (calcium fluoride window) in this study. Hydrophobic slides usually cause a preconcentration effect on the samples resulting in increased detection limits (Zhang et al., 2003, Jarvis et al., 2007). Another reason may be variations in purity or components of the examined samples. To reduce the influence of variations arising from drying, both samples (α_s-casein and dα_s-casein) were also examined as solutions.

3.3.2 Solution Samples
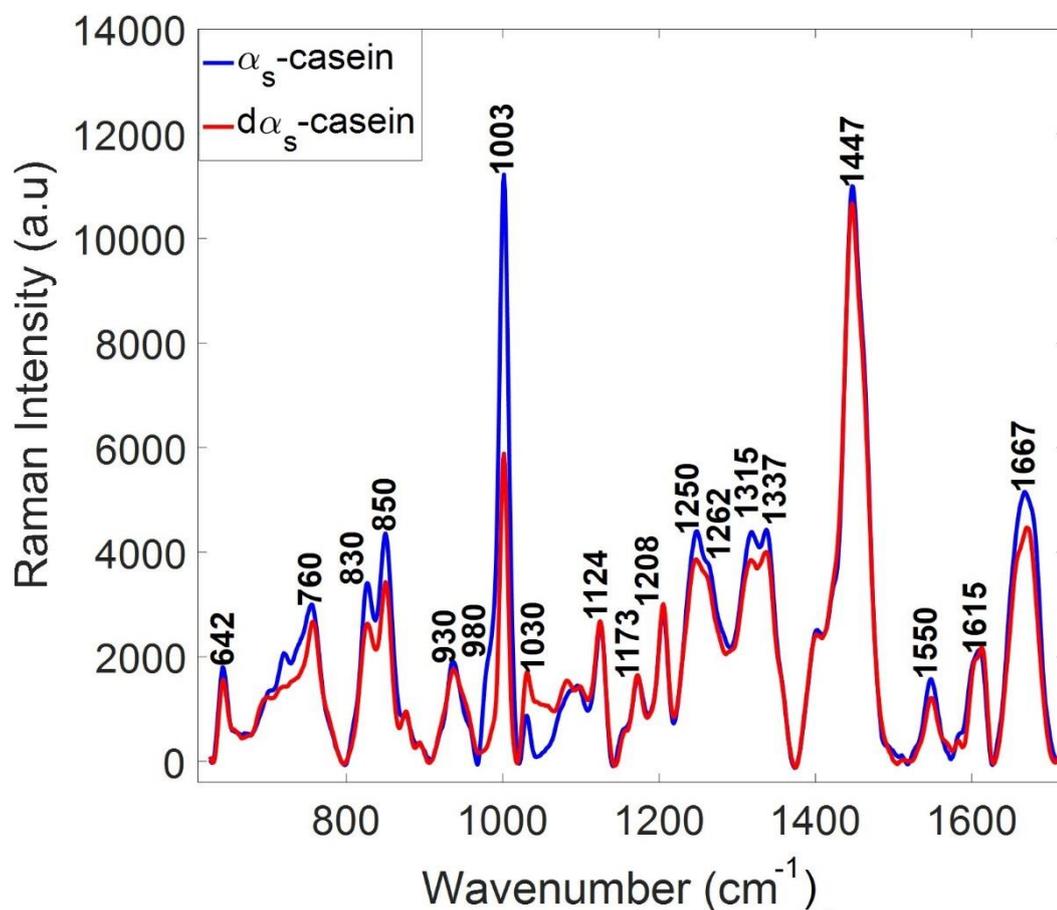
Figure 3.3 Averaged (n=3) Raman spectra of 10 mg/ml α_s-casein (pH 6.5) and dα_s-casein (pH 7.0) solutions. Preprocessing was carried out by baseline correction, smoothing, normalization, water subtraction and further baseline correction. Each Raman spectrum was collected in 30 mins using an excitation wavelength of 785 nm. Assignments are listed in Table 1.1.

Figure 3.3 displays the Raman spectra of α_s-casein and dα_s-casein solutions. In contrast to the dried film spectra (Figure 3.1), the peak at ~980 cm^{-1} assigned to dibasic phosphate stretch (Zhang et al., 2005, Ashton et al., 2011) can be observed as a small shoulder in the α_s-casein solution spectrum only. The challenge of observing phosphate peaks in proteins has been documented and was earlier

attributed to the presence of overlapping peptide bands in the same region as well as the lower solubility of phosphorylated proteins in aqueous solutions compared to phosphorylated amino acids (Xie et al., 2005). However, it has been successfully demonstrated that by increasing or decreasing the pH, phosphate peaks at ~980 and 1080 cm^{-1} can be easily detected in phosphorylated peptides and amino acids using Raman spectroscopy (Xie et al., 2005, Ashton et al., 2011) and phosphorylated proteins using Raman optical activity (Ashton et al., 2011). In comparison to α_s-casein, the dα_s-casein spectrum has a slightly lower intensity at some of the Raman peaks assigned to aromatic side chains (Table 1.1) including 830/850 cm^{-1} (Tyr Fermi doublet); 760, 1337 and 1550 cm^{-1} (Trp); and 1003 cm^{-1} (Phe). This observation differs from the dried film sample spectra in Figure 3.1. No peak intensity variations can be observed at ~830, and 1550 cm^{-1} whereas slightly increased peak intensities can be observed at ~760 and 1337 cm^{-1} in the dα_s-casein spectrum compared to the α_s-casein spectrum (Figure 3.1). The physical state of the sample can affect the intensity and band shape of the spectra (Smith and Dent, 2005). Therefore the differences between the dried film and solution spectra are most likely arising from variations in hydration states (Torii et al., 1998, Oliver et al., 2016).

The peaks at ~830 and 850 cm^{-1} have been historically assigned to the Tyr Fermi doublet, and the intensity ratio (I_{850}/I_{830}) has been used as a marker of hydrogen bonding and solvent exposure (Siamwiza et al., 1975). From the spectra, the calculated values of I_{850}/I_{830} (dα_s-casein=1.4; α_s-casein=1.4) are equal suggesting that there is no difference in hydrogen bonding or solvent exposure of Tyr residues in both samples despite the loss in intensity at ~830 and 850 cm^{-1}. As mentioned

previously, the samples may contain varying amounts of α_{s1}-casein and α_{s2}-casein. Consequently, the loss of intensity may be due to the number of Tyr residues present in each sample. α_{s1}-casein has 10 Tyr residues, and α_{s2}-casein has 12 Tyr residues (Horne, 2017, Morris, 2002). However other peaks ~1173, 1208 and 1615 cm^{-1} assigned to Tyr reveal no variations in peak intensity. Therefore, the loss of intensity at ~830 and 850 cm^{-1} is not fully explained by the difference in the number of residues. Hernandez et al. (2016) recently suggested a separate reassignment of these peaks to the in-plane ring breathing motion (830 cm^{-1}) and C-H out of plane bending (850 cm^{-1}) stating that the previous assignments are limited because the calculations were made using a simple model compound rather than Tyr. These separate assignments suggest the possibility of an alternative explanation for the intensity loss despite similar ratios. Nevertheless, the actual interpretation of the observed peak intensity loss remains unassigned in the literature.

Decreased peak intensity changes in the Trp peak at ~760 and 1550 cm^{-1} have been associated with changes in solvent exposure (Jacob et al., 1998, Gömez De La Cuesta et al., 2014). Therefore, the peak intensity changes may be as a result of increased exposure of the Trp residues in the dα_s-casein sample compared to the α_s-casein, which has the same number of Trp residues (n=2) (Morris, 2002, Horne, 2017). Variations in the peak intensity at ~1337 cm^{-1} are usually interpreted using the hydrophobicity marker ratio, I_{1360}/I_{1340}. Nevertheless, one of the two peaks may be absent in a Raman protein spectrum as is the case in Figures 3.1 and 3.3, thereby limiting the use of this ratio to interpret peak intensity variations for the observed peak. However, this peak may also be assigned to α-helix as discussed

further below with secondary structure changes. In contrast to the dried film spectra (Figure 3.1), the Phe peak at ~1003 cm^{-1} in the α_s-casein spectra has a significantly increased intensity compared to dα_s-casein. Similar intensity changes have been linked with increased short-range molecular interactions of Phe residues (Ota et al., 2016). However, as discussed previously, peak intensity variations may also be associated with a variation in the number of Phe residues present in each sample. Another possible explanation may be the influence of nearby phosphate vibrations (Jarvis et al., 2007). The peak intensity at ~1030 cm^{-1} (in the dα_s-casein spectrum) also assigned to Phe has an increased intensity compared to α_s-casein. This observation at ~1030 cm^{-1} is similar in the dried film spectra suggesting that this change is independent of the physical state of the sample (Figure 3.1).

In the dα_s-casein spectrum, there are other decreased peak intensities compared to the α_s-casein spectrum, associated with protein secondary structure (Table 1.1), including ~1250 cm^{-1} (disordered structure),~1315 cm^{-1} (α-helix), ~1337 cm^{-1} (α-helix) and ~1667 cm^{-1} (β-sheet). There is a limited number of studies comparing the secondary structure of phosphorylated and dephosphorylated α_s-casein in solutions. One such study by Koudelka et al. (2009) suggested an increase in both α-helix and β-sheet upon dephosphorylation. Variations in the ~1320-1345 cm^{-1} region have been linked with changes in α-helix symmetry (Tsuboi et al., 2000). Thus, the decreased peak intensities at ~1315 and 1337 cm^{-1} may suggest that there is a difference in the α-helix symmetry in the absence of phosphate residues.

Additionally, the decreased peak intensity at these α-helical peaks may also be linked with changes in the Phe peak at ~1003 cm^{-1}. The position of Phe in the

amino acid sequence of proteins in relation to itself other aromatic residues (Tyr, Trp and His) has been shown to determine specific conformations of α-helix when either of these residues is 1, 3 or 4 residues apart (Thomas et al., 2002). Of the eight Phe residues in α_{s1}-casein, 7 residues fall within this criterion of proximity to Tyr or Phe while two of the 6 Phe residues in α_{s2}-casein meet this criterion (Morris, 2002, Horne, 2017). Therefore the decreased intensities at these peaks may suggest that the dα_s-casein has a different conformation of α-helix possibly due to the arrangement of Phe in its sequence.

Unlike the Raman peaks at ~1315 and 1337 cm^{-1} which can indicate a change in helical content as well as orientation, the Raman peak at ~930 cm^{-1} is mainly reported to change as a result of helical content (Monti et al., 1998, Beattie et al., 2008). In Figure 3.3, no peak intensity variation can be observed at ~930 cm^{-1} between the α_s-casein and dα_s-casein spectra, suggesting that the overall helical content is similar in both samples. In contrast, the dried film spectrum of dα_s-casein has a decreased peak intensity at ~930 cm^{-1} (α-helix) compared to the α_s-casein spectrum (Figure 3.1). This observation may be due to different states or batch variations. The decreased intensity at ~1667 cm^{-1} in the α_s-casein spectra is associated with a reduced amount of β-sheet in the absence of phosphate residues. This observation is consistent between both dried film and solution spectra. The increased peak intensity at ~1250 cm^{-1} in the phosphorylated samples suggests an increase in disordered structure compared to dephosphorylated casein (Ashton et al., 2011). An explanation for the increase in disordered structure may be due to the role of phosphorylated residues in maintaining the open structure of casein (Holt et al., 2013).

Overall the results suggest differences between the Raman spectra of α_s-casein and dα_s-casein samples associated with aromatic side-chain residues and secondary structure. As discussed, some of these differences are possibly influenced by variations in particular components of the sample, sample states or detection limits. Notwithstanding, the results suggest conformational distinctions in the absence and presence of the PTM (phosphorylation).

3.3.3 2D-Correlation Analysis (2DCA)

3.3.3.1 α_s-Casein Autocorrelation

The pH-stability of α_s-casein was investigated by analysing the Raman spectra of its solutions ranging from ~pH 6-12 using 2DCA. As discussed previously in section 1.6.2, one of the most useful 2D-correlation plots is the autocorrelation which reveals the regions of the spectra or data that have changed significantly as a result of the perturbation. Figure 3.4A shows the autocorrelation plot for the α_s-casein pH-dependent spectra. The strongest autopeaks can be observed at ~1415, and 1600 cm^{-1} while a much weaker group of autopeaks can be observed at ~850, 980, 1003, 1065, 1234 and 1460 cm^{-1}. The assignments for the corresponding Raman peaks at these wavenumbers are listed in Table.1.1. A Raman peak at ~1415 cm^{-1} is assigned to the ionization of carboxyl groups. The dependence of ionization on pH makes this peak a visual marker of pH-induced spectral variation. At high pH, its increased intensity suggests increasing ionization. The Raman peak at ~1600 cm^{-1} is associated with the asymmetric ring vibrations of Tyr and Phe but can also be assigned to the deprotonation of Tyr residues.
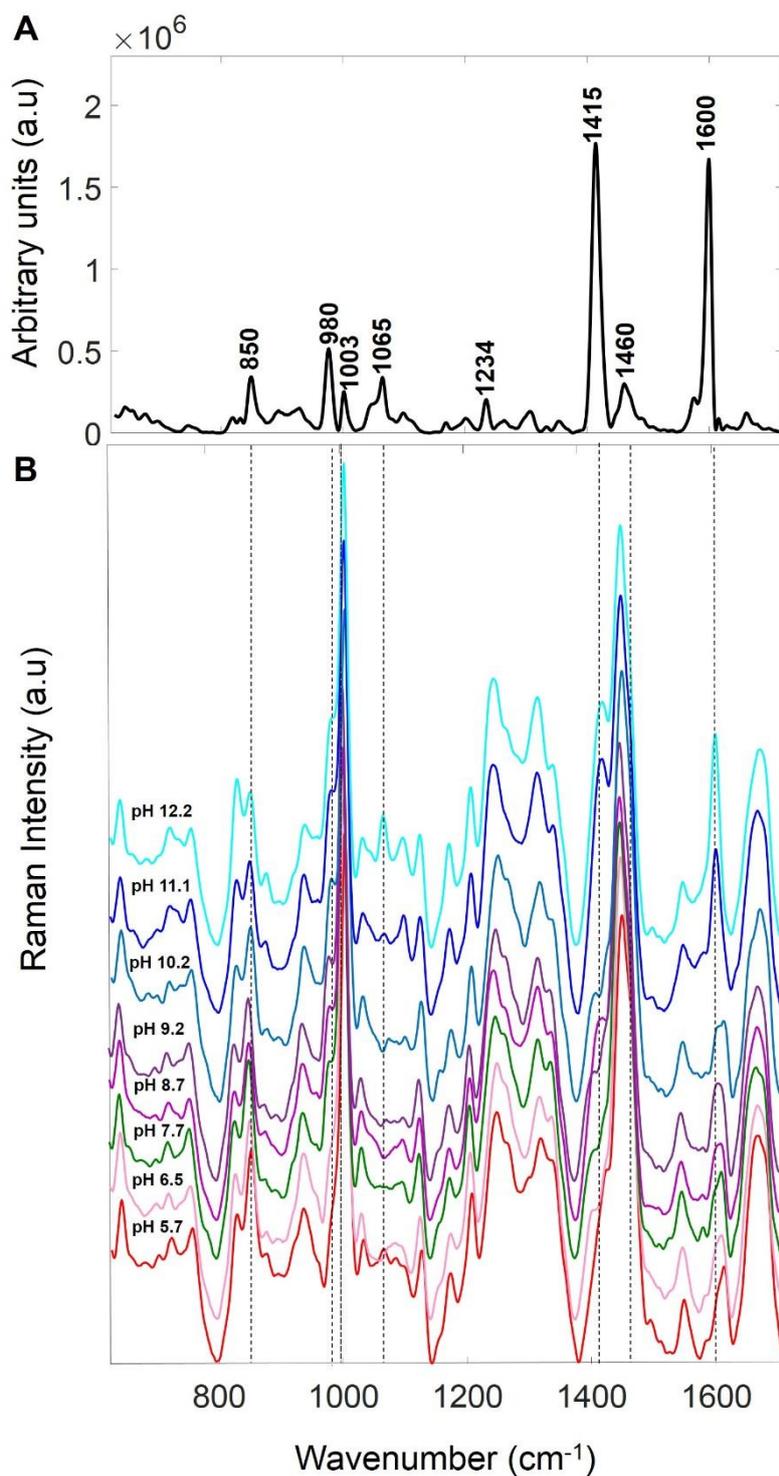
Figure 3.4 pH-dependent Raman spectral variations of 10 mg/ml α_s-casein from ~pH 5.7-12.2 (A) Autocorrelation plot (B) Averaged (n=3) Raman spectra except at pH 10.2 where n=1. Spectra were collected at pH intervals ranging from ~0.5-1.2. Preprocessing was carried out by baseline correction, smoothing, normalization, water subtraction and further baseline correction. Each Raman spectrum was collected in 30 min using an excitation length of 785 nm. Assignments are listed in Table 1.1.

In Figure 3.4B, the peak at ~1600 cm^{-1} is hardly detectable below ~pH 10.2. Instead a peak at ~1615 cm^{-1} is visible. From ~pH 10.2-12, there appears to be a peak shift from ~1615 cm^{-1} to 1600 cm^{-1}. This observation is associated with the deprotonation of Tyr residues (Ludwig and Asher, 1988, Hernández et al., 2016). An increase in the peak intensity at ~1065 cm^{-1} can also be observed to occur at high pH, suggesting changes in charged amino acids (Table 1.1). As previously mentioned, the improved visualization of the phosphate peak at ~980 cm^{-1} is pH-dependent. It can be observed that the autopeak at ~980 cm^{-1} indicates its increasing peak intensity between ~pH 6.5-9 (Figure 3.4B). The other spectral changes indicated by the remaining autopeaks at ~850,1003,1234 and 1460 cm^{-1} are harder to observe in the spectra by visual inspection alone.

3.3.3.2 α_s-Casein 2DPCMW

A key advantage of the 2DPCMW (section 1.6.4) is its utility in relating the spectral changes to the applied perturbation, which in this case is pH. In combination with the autocorrelation plot which identifies the main spectral variations, the way these spectral changes vary with pH can be carefully examined and interpreted. Figure 3.5 shows the synchronous 2DPCMW for α$_s$-casein. The negative contours shaded in blue represent decreasing peak intensity with increasing pH while the positive contours shaded in red indicate peaks that increase in intensity with increasing pH. The number of contours and the darker the shade of the contours indicates a greater magnitude of change. In Figure 3.5, the contours can be broadly grouped into two groups, ~pH 6-10 and 10-12.
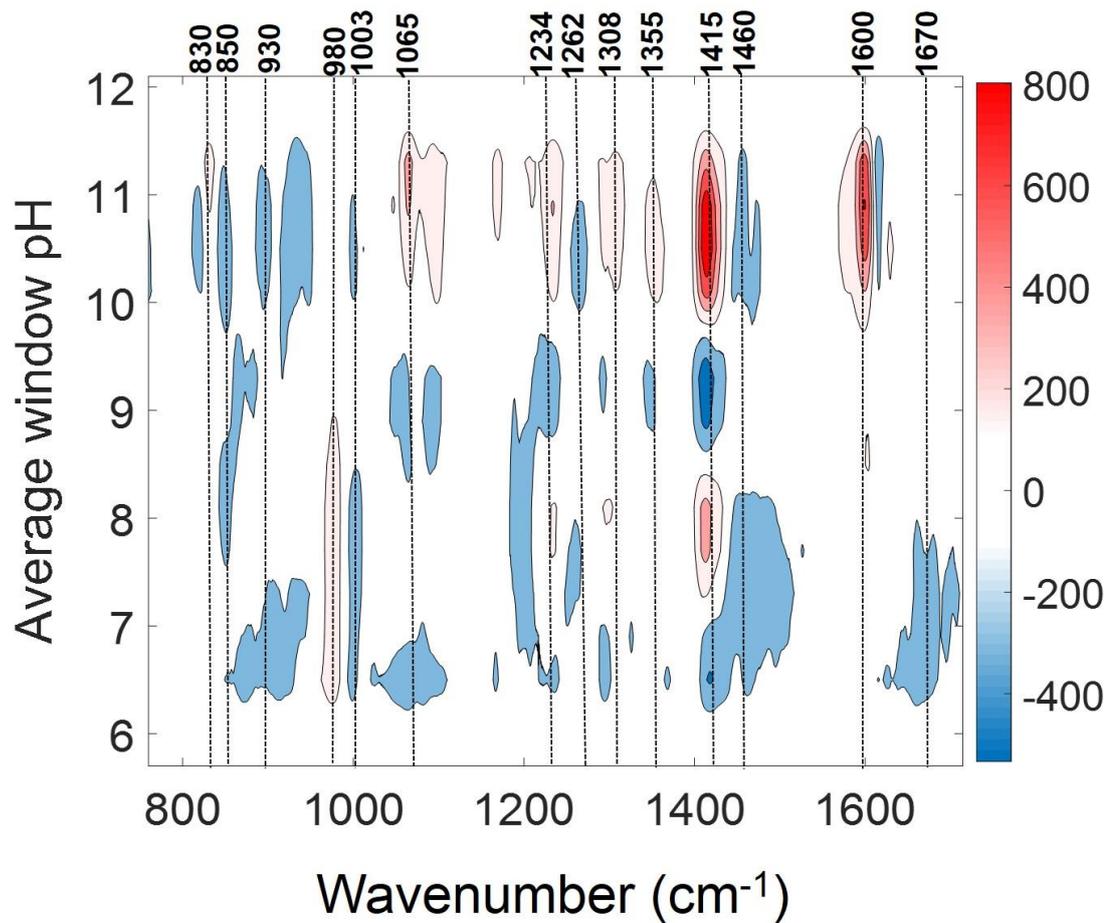
Figure 3.5 Synchronous 2DPCMW plot generated from pH-dependent Raman spectra of α_s-casein. Red shaded contours indicate increasing Raman peak intensity with pH, and blue shaded contours indicate decreasing Raman peak intensity with increasing pH. A darker shade of colour and a higher number of rings in the contour indicates a more significant change in intensity as indicated by the colour bar. The scale on this bar is of arbitrary units. A moving window size of 5 and a maximum of 6 contours was applied.

The positive contour at ~980 cm^{-1} indicates that the phosphate peak (~980 cm^{-1}) increases in intensity from ~pH 6.5-9 signifying increasing amounts of dibasic phosphate ions. Negative contours can be observed at ~1003 cm^{-1} which is assigned to the Phe Raman peak suggesting an increased solvent exposure of Phe side chains (Xu et al., 2008). The decreasing peak intensity at ~1460 cm^{-1} indicated by the presence of a negative contour suggests variations related to X-Pro peptide bond where X-refers to any amino residue bonded to Pro. Jordan et al. (1996)

studied proteins with Pro residues in their native and unfolded states. They demonstrated that a peak at ∼1460 cm^{-1} is associated with an unfolded protein state where X-Pro carbonyl moieties are strongly hydrogen-bonded with solvent. α_s-casein is natively unfolded and is known to have high proline content. There are 17 and 10 Pro residues in α_{s1} and α_{s2}-casein respectively (Morris, 2002). Furthermore, Pro residues in a PPII conformation, adopt a backbone conformation that is mainly exposed to water and the Pro-carbonyl moiety is known to be a good hydrogen bond acceptor (Thorn et al., 2015). Thus, the decreasing intensity at ∼1460 cm^{-1} might signal a weakening of such X-Pro carbonyl interactions with the surrounding solvent.

Overall, contours positioned at wavenumbers assigned to protein secondary structure are negative, which indicates a loss of intensity and consequently, loss of structure. The negative contours observed at ∼930 cm^{-1} (∼pH 9) and 1262 cm^{-1} (∼pH 8) suggest a loss of α-helix structure. Negative contours can also be found at ∼1234 and 1669 cm^{-1} implying loss of β-structure. However, from the spectra (Figure 3.4B), no peak can be observed at ∼1234 cm^{-1} rather a peak can be seen at ∼1250 cm^{-1} which is assigned to disordered structure.

In the 2DCA, autopeaks indicating spectral changes may occur at wavenumbers which are slightly different, but close to the original peak wavenumbers in the spectra. This variation occurs because autopeaks represent regions in the spectra, where the highest degree of change has occurred. Consequently, in addition to or without changes in peak intensity, the position of the autopeak could also reflect variations in peak shape such as broadening or narrowing (Noda and Ozaki, 2005,

Noda, 2017). Therefore, the negative contour at ~1234 cm^{-1} likely suggests a loss of disordered structure alongside the loss of α-helix and β-sheet.

One of the unique features in the ~pH 10-12 range is the absence of a contour at ~980 cm^{-1} indicating no further increase in dibasic phosphate groups. The darker shade and number of positive contours at ~1065, 1415 and 1600 cm^{-1} indicate a significant increase in the intensity of the Raman peaks assigned to charged amino acids, ionization of carboxyl groups and deprotonation of Tyr side chains respectively (Table 1.1). In contrast to the ~pH 6-10 range, positive contours can be observed at ~1308 and 1355 cm^{-1}. The closest original Raman peaks in the spectra occur at ~1315 and 1337 cm^{-1} and have been assigned to changes α-helix symmetry (Figure 3.4). As discussed, the position of the contours could vary from the peak position in the actual spectra. Consequently, the contours at the 1308 and 1355 cm^{-1} suggest that geometrical changes in α-helix are occurring in addition to the loss of the α-helix structure indicated by negative contours at ~930 and 1262 cm^{-1} also assigned to α-helix. Similarly, the positive contour at ~1234 cm^{-1} earlier linked with variations in the Raman peak at ~1250 cm^{-1} implies an increase in disordered structure, which is consistent with a loss of secondary structure.

The absence of changes in the phosphate peak and the corresponding increased loss of secondary structure from ~pH 10-12 compared to ~pH 6-9 is possibly accounted for by the presence of phosphorylated serine residues providing a 'shielding' effect during the deprotonation of the phosphate groups below pH 10. As shown in Figure 3.6, the hydrogen atom on the hydroxyl group is substituted by the dibasic phosphate group (OPO_3^{2-}) in the serine side chain.
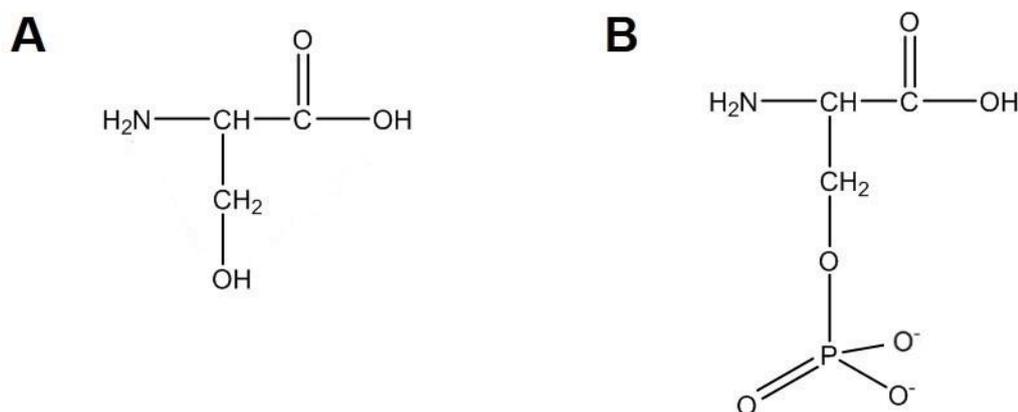
Figure 3.6 (A) Structure of serine amino acid (B) Phosphoserine structure showing attached deprotonated phosphate group.

3.3.3.3 dα_s-Casein Autocorrelation

Figure 3.7A and 3.7B shows the autocorrelation and pH-dependent Raman spectra of the dα_s-casein, respectively. As expected, there is no autopeak indicating any changes linked with phosphate vibrations. Another notable observation is that unlike in the α_s-casein autocorrelation, there is an autopeak at 830 cm^{-1} and not 850 cm^{-1}. As with α_s-casein sample, the autopeaks at ~1065, 1415 and 1600 cm^{-1} indicate Raman peak intensity changes at ~1065, 1415 and 1600 cm^{-1} assigned to charged amino acids, ionization of carboxyl groups and deprotonation of Tyr residues respectively. These spectral changes are evident at ~pH 10-12 (Figure 3.7B). As previously mentioned in section 3.3.3.1, the autopeak at ~1600 cm^{-1} is likely due to a peak shift from ~1615 to 1600 cm^{-1}. Synchronous and asynchronous plots are particularly useful in confirming the presence of peak shifts, so it was interesting to use this tool to clarify this observation in both α_s-casein and dα_s-casein spectra.
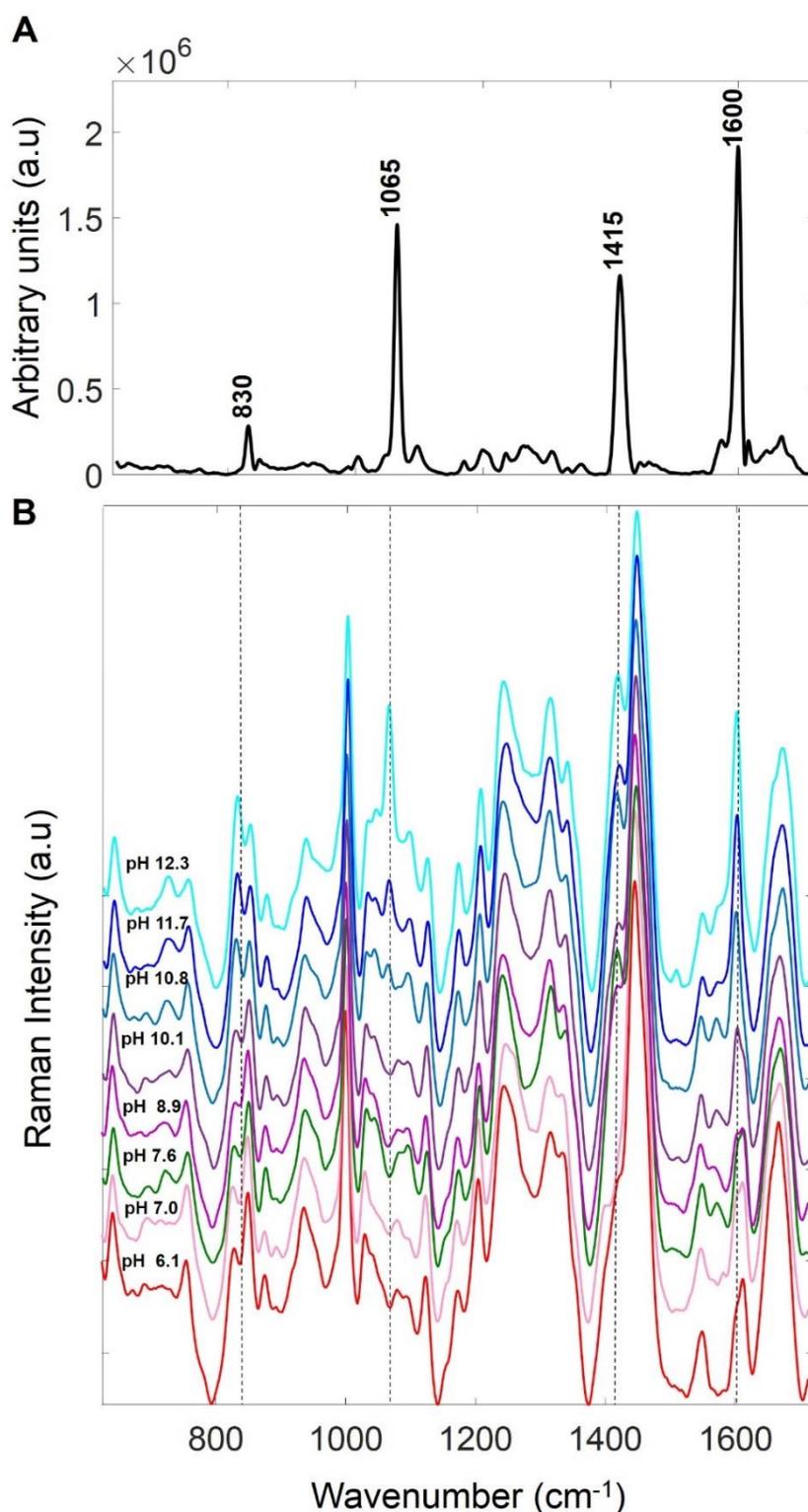
Figure 3.7 pH-dependent Raman spectral variations of 10 mg/ml dα_s-casein solution from ~pH 6.1-12.3 (A) Autocorrelation plot (B) Averaged (n=3) Raman spectra. Spectra were collected at pH intervals ranging from ~0.6-1.3. Preprocessing was carried out by baseline correction, smoothing, normalization, water subtraction and further baseline correction. Each Raman spectrum was collected in 30 mins using an excitation length of 785 nm. Assignments are listed in Table 1.1.
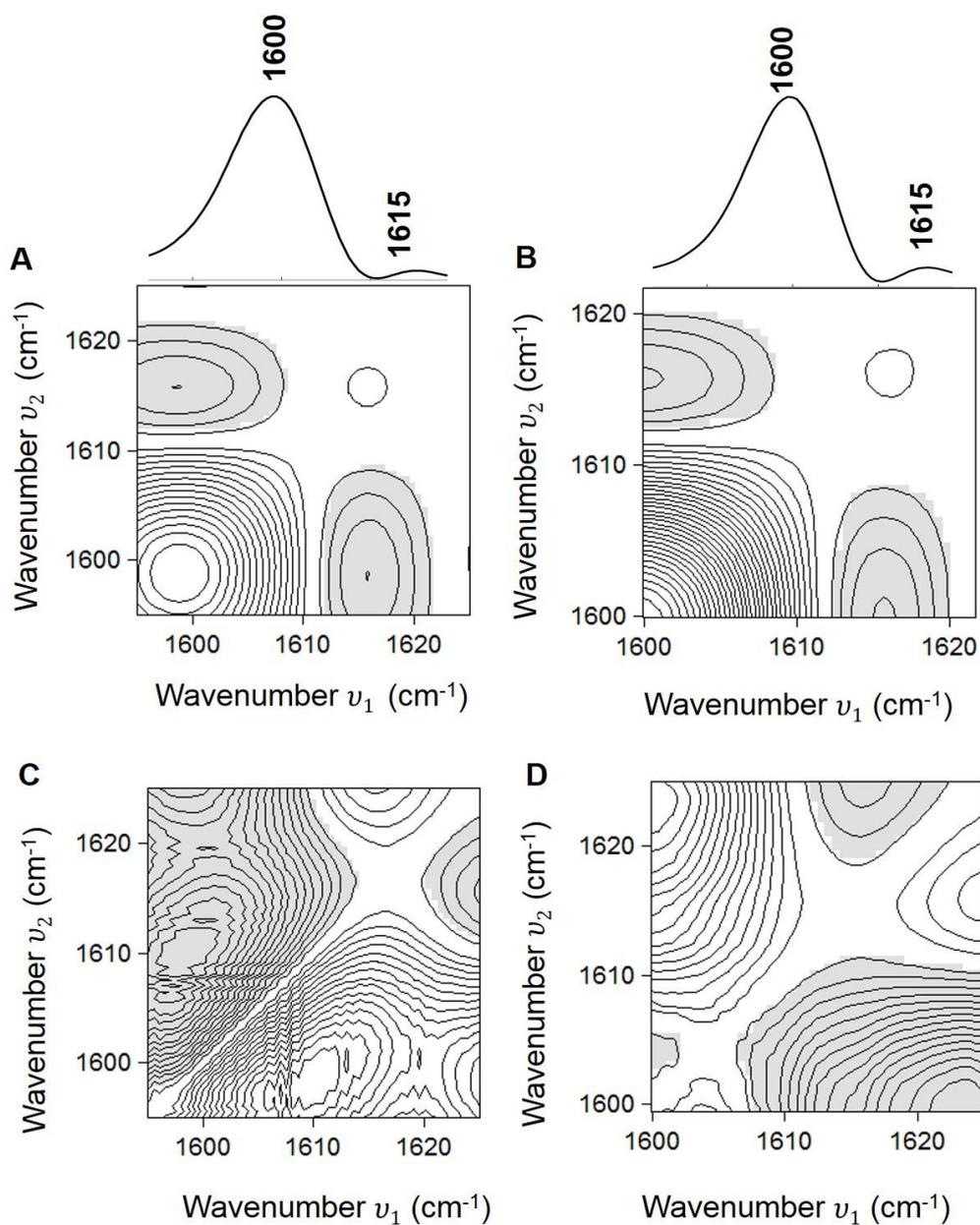
Figure 3.8 Synchronous plot (SP) and Asynchronous plot (AP) generated from pH-dependent spectra of 10 mg/ml solutions of αs-casein (~pH 10-12) and dαs-casein (~pH 9-12). (A) SP αs-casein (B) SP dαs-casein (C) AP αs-casein (D) AP dαs-casein. Shaded contours indicate negative cross peaks and unshaded contours indicate positive cross peaks. Positive cross peaks indicate two wavenumbers with perturbation induced intensity changes in the same direction (increasing or decreasing together). Negative cross peaks indicate two wavenumbers with perturbation induced intensity changes in opposite directions (one increasing and the other decreasing). Maximum contour level=16.

Figure 3.8 shows the synchronous and asynchronous plots for both samples. The synchronous plots depict an *angel* pattern which has two negative cross peaks of similar strength (wings) and two autopeaks with one significantly weaker (head) than the other (body). The *angel* pattern can be either indicative of two overlapped peaks with markedly different intensities or a peak shift (Noda & Ozaki, 2005).

The exact interpretation is dependent on the pattern of the cross peaks in the corresponding asynchronous plot (Figure 3.8C and 3.8D). The appearance of a butterfly pattern in the asynchronous plot is indicative of a peak shift. Since both asynchronous plots lack a butterfly pattern, no peak shift is indicated. Consequently, the *angel* pattern here indicates the two overlapped Tyr peaks (1600 and 1615 cm^{-1}) with markedly different intensities, changing in opposite directions as indicated by the two well-resolved cross peaks in the asynchronous plots. As discussed above, the appearance of the peak at ~1600 cm^{-1} and the disappearance of the ~1615 cm^{-1} peak at high alkaline pH spectral change is generally reported as a being due to a peak shift. In fact, by examining the spectra alone (Figure 3.4B and 3.7B), one may easily arrive at the same conclusion.

However, these plots clearly show that the peak at ~1615 cm^{-1} does not actually shift, but it becomes very weak at high alkaline pH, whereas the ~1600 cm^{-1} peak intensity increases significantly. Based on Noda's rules for sequential order intensity changes (section 1.6.3), the cross peak signs in Figure 3.8 indicate that the peak intensity change at ~1600 cm^{-1} comes predominantly before 1615 cm^{-1} for α_s-casein whereas, for dα_s-casein, the change at ~1615 cm^{-1} occurs predominantly before 1600 cm^{-1}. This difference may be linked to their phosphorylation status.

3.3.3.4 dα_s-Casein 2DPCMW

Figure 3.9 displays the synchronous 2DPCMW plot for dα_s-casein. It appears very similar to the α_s-casein 2DPCMW plot except that the highest number of contours appear at a lower pH rather than at a higher pH as observed with α_s-casein.
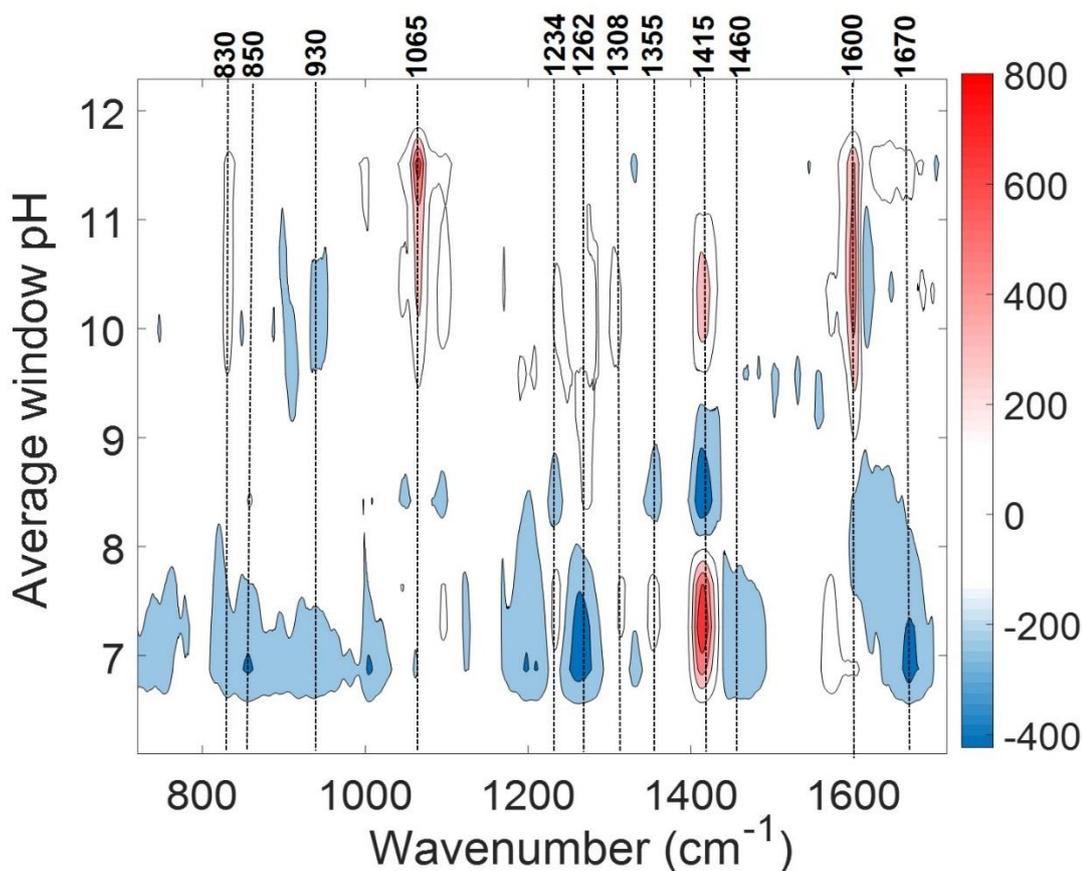
Figure 3.9 Synchronous 2DPCMW plot generated from pH-dependent Raman spectra of dα_s-casein. Red shaded contours indicate increasing Raman peak intensity with pH, and blue shaded contours indicate decreasing Raman peak intensity with increasing pH. A darker shade of colour and a higher number of rings in the contour indicates a significant change in intensity as indicated by the colour bar. The scale on this bar is of arbitrary units. A moving window size of 5 and a maximum of 6 contours was applied.

From ~pH 6-9, the pattern of contours from ~1234 to 1355 cm^{-1} in the dα_s-casein plot is analogous to the same region in the α_s-casein 2DPCMW plot at ~pH 10-12 (Figure 3.5).The dα_s-casein 2DPCMW plot can also be grouped into two ranges, ~pH 6-9 and 9-12. In the lower range, there are negative contours at ~830/850, 1003 and 1460 cm^{-1} indicating decreasing peak intensity at the Raman peaks assigned to Tyr, Phe and X-Pro carbonyl groups (Table 1.1). The presence of the negative contours at ~830 and 850 cm^{-1} between ~pH 6-9 contrasts with the α_s-casein 2DPCMW plot where no such contours can be observed. This difference is possibly linked to the absence of phosphorylated serine residues. Furthermore, the negative contours at ~930, 1262 and 1670 cm^{-1} also suggest a loss of secondary structure at Raman peaks assigned to α-helix and β-sheet respectively. As discussed previously, the positive contours at ~1234 cm^{-1} suggest an increase in disordered structure, whereas those at 1308 and 1355 cm^{-1} suggest a change in the symmetry of the α-helix structure.

While there are some striking similarities between the pattern of contours in the ~pH 6-9 range of the dα_s-casein 2DPCMW plot (Figure 3.9) and α_s-casein 2DPCMW plot (Figure 3.5); it is worth noting that the size and shade of the negative contour at ~1262 cm^{-1} suggests that in the dα_s-casein sample, the loss of α-helix is occurring to a greater extent at low pH compared to α_s-casein. This observation suggests that some of the pH-dependent secondary structural changes are more significant at a lower pH in the absence of dibasic phosphate residues.

In dα_s-casein, none of the serine residues are phosphorylated, and so their sidechains do not have the protection afforded by the deprotonation of the

phosphate residues. Their exposure may account for the earlier onset of the changes in secondary structure in dα$_s$-casein compared to α$_s$-casein.

As discussed earlier, there are no crystallised structures for caseins. Kumosinski et al., (1991) proposed a working energy minimised molecular model for α-$_{s1}$ casein providing insight into its folding mechanism in relation to the phosphate groups as well as Pro residues. They showed that 7 of the phosphoserine residues are located on β-turns and the polypeptide folds to bring these residues into proximity between residues 43 and 84. In addition, the phospho serine residues are further bound by Pro residues at 29 and 87.
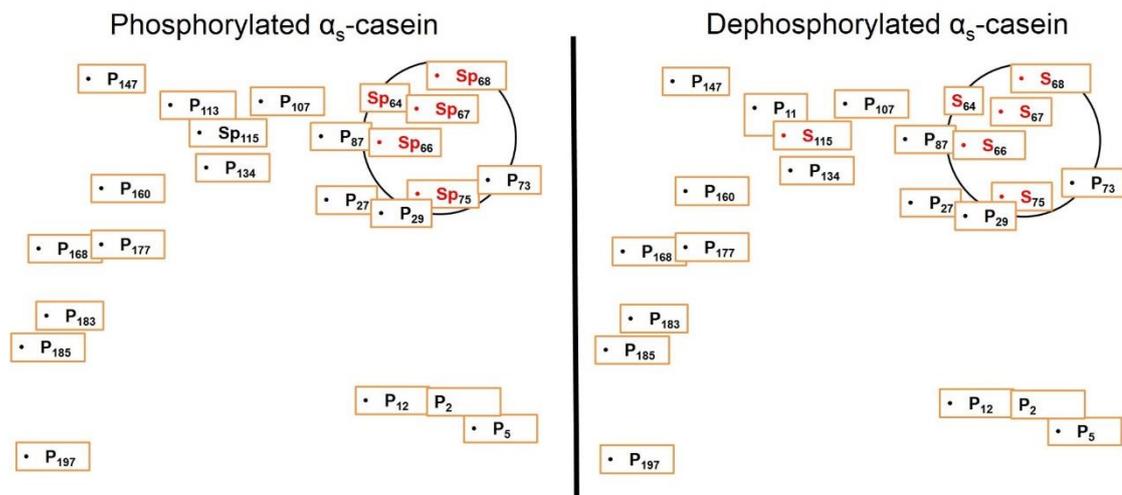
Figure 3.10: Schematic depicting energy minimised model of α$_{s1}$-casein adapted from Kumosinski et al., (1991) with labelled representative positions Pro (P) and PhosphoSer residues (Sp). Black ring indicates cluster of phosphoserines. Only 5 Ser residues are shown in the cluster instead of 7 as the position of the 2 others could not be deduced from the original model.

Farrell et al., (2009) generated a homologous molecular model for α$_{s2}$ -casein. However, it has not been used in this discussion because the position of the phosphoserines on the model are not easy to visualise. In addition, the model does not give any insight to how these residues fold. Figure 3.10 shows a schematic of

the energy model with clustered phosphoserine residues. Using this model, it can be proposed that the folding of α_{s1}-casein will be altered in the absence of phosphate groups. Phosphate groups on amino acids can significantly alter the local and global and conformational properties of a protein (Hunter, 2012, Kumar et al., 2020). As a hydrogen bond acceptor, phosphoserine has been reported to form stronger salt bridges with Arg compared to Asp or Glu (Rapp et al., 2013). Furthermore, conformational states of intrinsically disordered proteins have been shown to be influenced by the net charge per residue (Mao et al., 2010). With increasing deprotonation of its phosphate groups, it is likely that the net charge per residue of α_s-casein is altered, modifying its conformational stability compared to dα_s-casein. Therefore, observed difference in stability between α_s-casein and dα_s-casein at lower pH has been attributed to the presence of deprotonating phosphate groups on the serine residues.

The pattern and position of contours in the ~pH 9-12 range suggest that the spectral changes mainly relate to sidechain structure. As discussed previously, the positive contours at ~1065, 1415 and 1600 cm^{-1} suggest increased changes in the Raman peaks assigned to charged amino acids (deprotonation of Lys, Glu or Asp), ionization of carboxyl groups and deprotonation of Tyr sidechains respectively. These observations are consistent at high pH regardless of phosphorylation status.

Two interesting observations relating to side chains that appear to be dependent on phosphorylation are the absence of a contour at ~850 and 1460 cm^{-1}. As observed in Figure 3.7A, there is an autopeak at ~830 cm^{-1} suggesting peak intensity changes at ~830 cm^{-1} which is assigned to Tyr. This observation contrasts with the autocorrelation plot of α_s-casein (Figure 3.4A) which reveals an

autopeak at ~850 cm^{-1} and is possibly explained by the separate assignments for the peaks at ~830 and 850 cm^{-1} (Hernández et al., 2016) rather than the traditional assignment to Fermi resonance (Siamwiza et al., 1975); since it reflects the intensity changes of separate peaks rather than the intensity ratio of the doublet. The exact interpretation of the decreased intensities for protein structure and the influence phosphorylation will require further investigation. Nonetheless, for both α_s-casein and dα_s-casein, the ratio of peak intensities at ~830/850 cm^{-1} (I_{850}/I_{830}) in the Raman spectra (Table 3.2) decreased from ~1.5 at slightly acidic/moderately basic pH to ~0.9 at highly alkaline pH.

This ratio has been linked to changes in hydrogen bonding (Siamwaza et al., 1975) and most recently, hydrophobic/hydrophilic balance of interactions involving the Tyr Phenol ring (Hernandez et al.,2016). The latter study argues that hydrogen bonding variations does not sufficiently explain the limits of the ratio which correspond to when Tyr is a strong H-bond acceptor (~2.5) or donor (~0.3). The range of the ratios suggests that regardless of phosphorylation, most of the Tyr residues in α_s-casein are exposed on the surface of the protein and that hydrogen bonding is moderate to weak (Siamwaza et al., 1975). The ratio values are also consistent with the observations by Hernandez et al., (2016) who observed (I_{850}/I_{830}) values above 1 in a Tripeptide where the Tyr phenol ring is thought to be mostly exposed to solvent. This surface location is expected due to the unfolded nature of α_s-casein. The lower (I_{850}/I_{830}) values at alkaline pH is likely accounted for by the increasing ionization of Tyr from ~pH 10.2-12 (Siamwiza et al., 1975).

Table 3.2 I_{850}/I_{830} values calculated from the pH-dependent Raman spectra of α_s-casein and dα_s-casein.

α_s-casein		dα_s-casein	
pH	**I_{850}/I_{830}**	**pH**	**I_{850}/I_{830}**
5.7	1.5	6.1	1.4
6.5	1.4	7.0	1.4
7.7	1.5	7.6	1.4
8.7	1.4	8.9	1.4
9.2	1.3	10.1	1.23
10.2	1.3	10.8	0.98
11.1	1.1	11.7	0.92
12.2	0.9	12.3	0.85

The absence of the autopeak at $\sim$1460 cm^{-1} suggests no changes in the hydrogen bonding interactions of X-Pro carbonyl groups implying a limited extent of unfolding in this pH range (Jordan et al., 1996). As mentioned previously, there are a total of 27 proline groups between α_{s1} and α_{s2-} casein. Based on the involvement of the phosphoserine cluster in the energy minimised model folding (Figure 3.10), the Pro groups associated with this change at $\sim$1460 cm^{-1} may be those near this cluster (29 and 87) in α_{s1}-casein. This limited unfolding is possibly linked to the

limited changes in secondary structure observed in this pH range. Apart from the negative contour at ~930 cm^{-1} signalling a decreased peak intensity in the Raman peak assigned to α-helix, there appears to be minimal changes in secondary structure in this upper range, demonstrating increased stability from ~pH 9-12 in the absence of phosphate residues.

On the other hand, for a dα$_s$-casein sample whose pH is gradually increased from ~pH 6-12, this observation could indicate that since extensive changes have already occurred at lower pH, no further changes would be expected at higher pH.

3.4 Conclusion

The application of 2DCA to the pH-dependent Raman spectra of α$_s$-casein and dα$_s$-casein identified differences in their conformational stability attributable to their phosphorylation status. An initial comparison of the Raman spectra of α$_s$-casein and dα$_s$-casein in their native state suggests that in the absence of phosphate residues, α$_s$-casein has increased exposure of Trp residues to solvent, variation in α-helix conformation as well as a decrease in β-sheet and disordered structure.

Although the observed pH-induced variations in both samples' spectra compared closely by visual inspection, 2DPCMW analysis of the spectra highlights differences in the behaviour of dα$_s$-casein and α$_s$-casein across the examined pH range. The results suggest that dα$_s$-casein is less stable at lower pH (6-9) compared to α$_s$-casein. At the lower pH range, extensive loss of secondary structure including α-helix and β-sheet, were observed for dα$_s$-casein whereas comparably limited changes in secondary structure were observed at higher pH. α$_s$-casein, on the other hand, showed extensive secondary structure loss at pH 10-12 compared to at the

lower pH range. This difference in stability between α_s-casein and dα_s-casein at lower pH has been linked to the presence of phosphate groups on the serine residues playing a protective role through their deprotonation in α_s-casein. Furthermore, the relative stability of dα_s-casein at ~pH 10-12 has been linked to limited unfolding of the X-Pro carbonyl groups. Synchronous and asynchronous plots generated from the Raman spectra of both samples indicate that the generally reported peak shift from ~1615 to 1600 cm^{-1} signalling deprotonation is due to changes in the intensity of overlapping peaks at ~1600 and 1615 cm^{-1}.

4 Human Serum Transferrin Model

4.1 Introduction

Human serum Transferrin (hTf) and other serum transferrins are well known for their role in iron transport through receptor-mediated endocytosis (Thorstensen and Romslo, 1990, Baker, 1994). The efficiency of the iron-binding and transportation of hTf has been exploited for its potential applications as iron-binding therapy, metallodrug carriers, targeting ligands and drug delivery machinery (von Bonsdorff et al., 2001, Wagner et al., 1994, Li and Qian, 2002).

Figure 4.1 Cartoon image of 3QYT (Yang, N, Zhang, H, Wang, M Hao, Q, Sun, H (2012) Iron and bismuth bound human serum transferrin reveals a partially opened conformation in the N-lobe. *Sci Rep* 2:999) created with NGL (A.S. Rose, A.R. Bradley, Y. Valasatava, J.D. Duarte, A. Prlić, P.W. Rose (2018) NGL viewer: web-based molecular graphics for large complexes. *Bioinformatics* 34: 3755–3758).

In particular, the transport of hTf conjugated to anticancer agents into malignant cells has resulted in desirable pharmacological properties including decreased drug toxicity, enhanced cytotoxic activity and modified drug release profiles (Szwed et al., 2014, Zhang et al., 2015). Additionally, hTf coupled liposomes have exhibited improved drug release and stability behaviour compared to the liposomes alone (Singh, 1999, Leto et al., 2016).

HTf is a glycoprotein with a molecular weight of ~80 kDa. It is a single polypeptide chain with two similar lobes (C and N-lobes). Each lobe has a Fe^{3+} binding site situated in a cleft formed by two subdomains (Figure 4.1). The lobes are linked by a flexible short stretch of a peptide (Baker and Baker, 2004, MacGillivray et al., 1998, Wally and Buchanan, 2007). The Fe^{3+} binding sites include two tyrosine residues, one histidine residue, one asparagine residue as well as a bicarbonate ion (MacGillivray et al., 1998, Baker and Baker, 2004, Eckenroth et al., 2011).

In the endosome, the presence and release of bound iron are associated with conformational changes that involve the closing and opening of the domains around the iron-binding site respectively. The closure is facilitated by the tight reversible binding of iron via ligands situated in two of the domains and the backbone strands that link the domains (Baker, 1994). During the opening of the site, the domains on each lobe rotate away from each other, weakening the interactions of the bound iron and consequently resulting in its release (Andersen et al., 1990, Jeffrey et al., 1998). The iron free form (apohTf) is thought to have an open conformation due to the vacant iron-binding site whereas holohTf has a closed conformation because the domains close to surround the bound iron (Baker et al., 1990, Baker, 1994, Navati et al., 2003). The release of iron from hTf is also

pH-dependent, occurring below ~pH 6 in the endosome; and also observed in-vitro (MacGillivray et al., 1998, Day et al., 1992, Baker and Baker, 2004).

Raman spectroscopy has been used to distinguish apohTf from holohTf because the holohTf spectra show specific peaks that can be linked to iron vibrations. Earlier studies of hTf using resonance Raman spectroscopy, have revealed its utility in the identification of iron peaks providing insight into the Fe^{3+} binding site (Tomimatsu et al., 1976, Gaber et al., 1974). More recently, conventional Raman spectroscopy (RS) has been used in the identification of structural modifications in hTf as a result of iron-binding and glycosylation (Ashton et al., 2017). The differing conformations of hTf and its potential role in drug delivery, combined with the sensitivity of Raman spectroscopy to detect structural differences in the presence and absence of iron, makes hTf a suitable model to examine the ability of 2DCA and Raman spectroscopy to investigate differences in protein conformational stability during early drug development.

The aim of this study is to examine and compare the conformational stability of apohTf and holohTf under varying pH and temperature conditions using this technique. As with the α_s-casein model discussed in chapter 3, this study mimics a drug development screening scenario where a protein exists in variant forms, and there is a need to compare the stability of the variants to aid the selection and optimization of the most stable candidate.

4.2 Experimental

4.2.1 Dried Film Samples

Samples of apohTf and holohTf were purchased from Sigma Aldrich and used without further purification. Dried films of aqueous solutions were prepared as discussed in section 2.2 1 and 3.2.1.1.

4.2.2 Liquid Samples (pH-study)

Samples preparation and pH measurements (Table 4.1) are as described in section 3.2.1.2. At pH 2.6 and 3.6, a complete loss of colour was observed in the holo-hTf samples after preparation. The loss of colour is likely due to complete iron dissociation at low acidic pH (Figure 4.2).

Table 4.1 pH values (± 0.2) for hTf samples

apohTf	2.6	3.6	4.6	5.6	6.6	7.6	8.6	10.4	11.4
holohTf	2.6	3.6	4.6	5.7	7.0	8.1	9.4	10.5	11.5
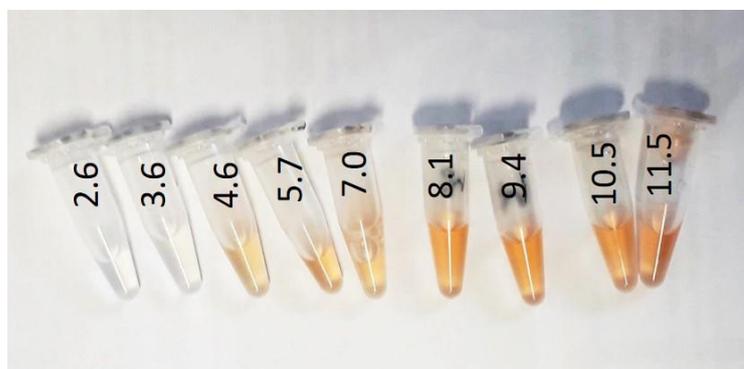
Figure 4.2 Photograph showing a complete colour loss in 10 mg/ml holohTf solution samples at ~pH 2.6 and 3.6.

Liquid Samples (Temperature study)

3 ml of each sample was prepared by dissolving the lyophilised powder in Milli-Q water. The final concentration of each sample was ~10 mg/ml.

4.2.4 Spectra Collection

pH-dependent Raman spectra were collected as described in section 2.2.2.

Three spectra were collected for each sample at pH values listed in Table 4.1. The exposure time for dried film samples was 10s with 1 accumulation (collection time of 10s) and 30 mW laser power at source. A total of 6 spectra were collected from different positions on the sample spot. For temperature studies, a single spectrum was collected through a 3 ml quartz cuvette at ~20 °C and ten other temperature points (section 2.4.2) using a long-distance x50 objective. A spectrum was also obtained from the heated samples after cooling down to ~20 °C. Spectral acquisition time was also 30 min as in the pH study. Two repeat experiments were carried out for each sample.

4.2.5 Data Preprocessing

Dried film and liquid spectra (pH study) were preprocessed as discussed in section 2.3. The same preprocessing approach was initially attempted for the temperature-related spectra, despite significant interfering signals from the quartz cuvette, especially at ~800 and 1200-1400 cm^{-1} (Figure 4.3). It was reasonable to expect that water subtraction from the protein spectra would reduce the interference of the quartz signals since the water spectra also had some quartz interference (Figure 4.3). However, compared to apohTf spectra, water subtraction alone did not significantly reduce the quartz interference in the holohTf spectra (Figure 4.4C and 4.4D).
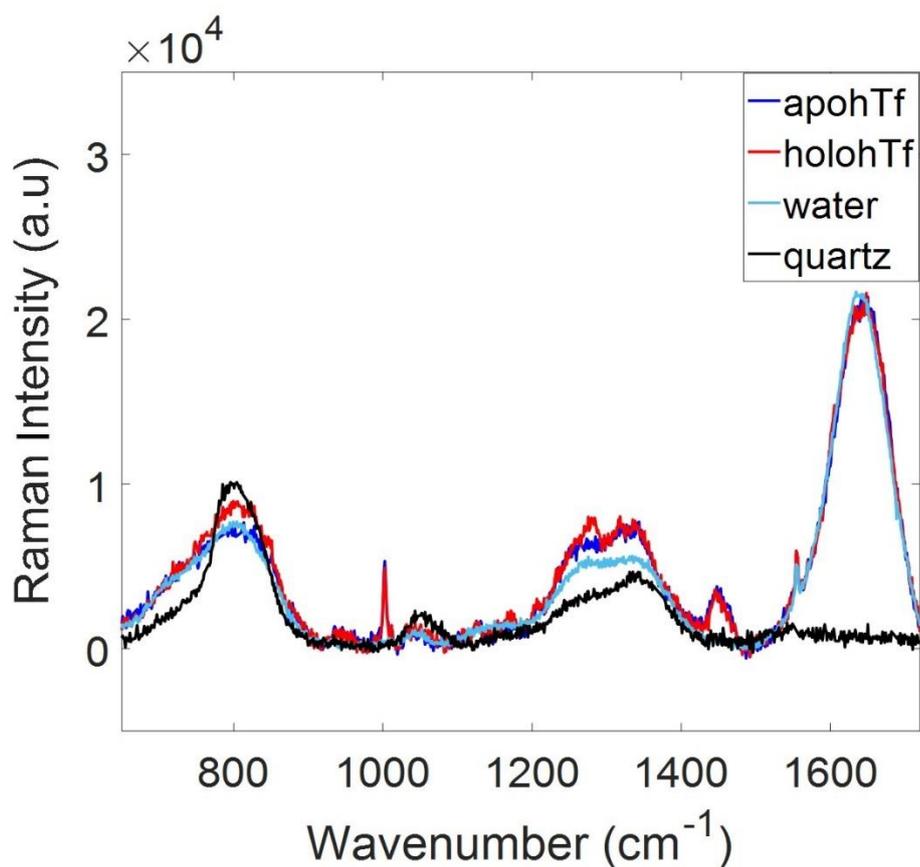
Figure 4.3 Raw Raman spectra of water and 10 mg/ml solutions of apohTf and holohTf showing interference of quartz signals at ~800 and 1200-1400 cm^{-1}. Each spectrum was collected in 30 mins using an excitation length of 785 nm.

This discrepancy is possibly due to the position of the objective focus on the cuvette or the nature of the sample. For consistency, a quartz spectrum subtraction was performed for apohTf and holohTf samples before water subtraction (Figure 4.4C). Using this approach, it was also necessary to subtract quartz from the water spectrum before water subtraction to prevent the occurrence of negative peaks in the resulting protein spectrum. Therefore, the final order of preprocessing steps for the temperature spectra was baseline correction followed by subtraction of the quartz spectrum, smoothing, normalisation to the peak at ~1447 cm^{-1}, water subtraction and further baseline correction.

The ~30 °C holohTf spectrum was excluded from preprocessing due to distortion by a cosmic ray. Due to the cosmic ray location, it was not possible to remove it without interfering with the spectra. The ~30 °C apohTf spectrum was also removed for consistency. For the generation of synchronous 2DPCMW plots, spectra were interpolated using the Savitzky-Golay spline estimation to generate evenly spaced interpolated data in steps of ~0.5 and 2.4 for pH and temperature studies respectively.
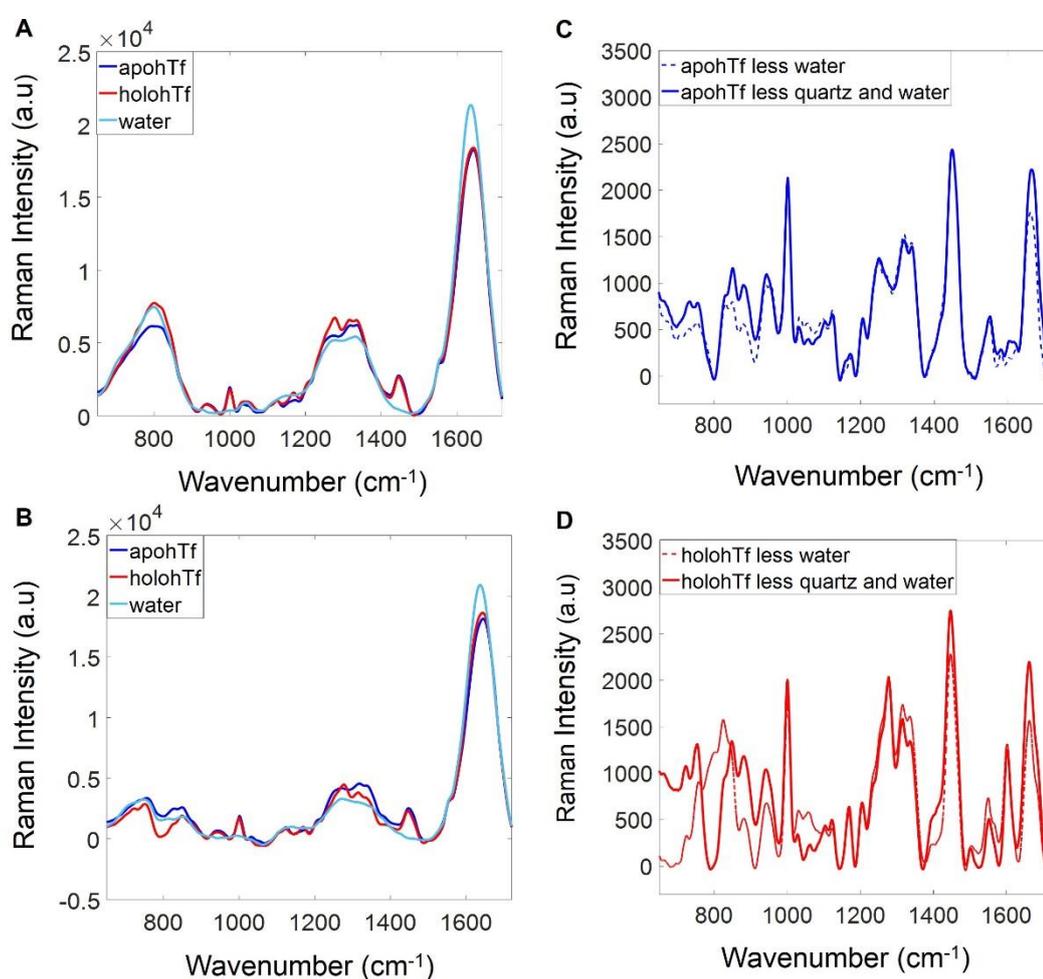
Figure 4.4 Raman spectra collected at 28° C after baseline subtraction, smoothing and normalisation. (A) apohTf, holohTf and water (B) Quartz subtracted apohTf, holohTf and water (C) apohTf after water subtraction with and without quartz subtraction (D) holohTf after water subtraction with and without quartz subtraction. Quartz subtraction was performed after initial baseline subtraction for both water and protein spectra. A further baseline correction was applied after water subtraction in all cases. Concentration of apohTf and holohTf was 10 mg/ml.

4.3 Results & Discussion

4.3.1 Dried Sample Spectra.

Figure 4.5 displays the preprocessed Raman spectra of apohTf and holohTf. Typical peak assignments for Raman spectra are listed in Table 1.1 (section 1.5.3). The main differences between the two spectra arise from the peaks at ~1172, 1280, 1505 and 1605 cm^{-1} observed in the holohTf spectra but very weak in the spectra of apohTf.
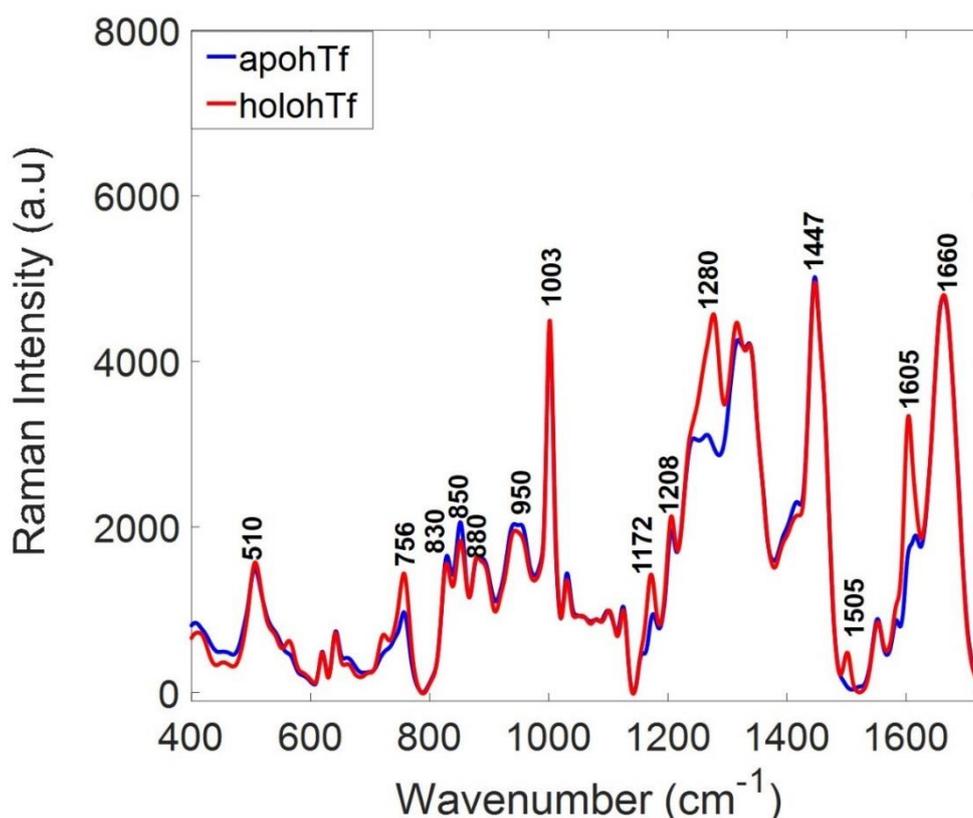
Figure 4.5 Averaged (n=6) Raman spectra of 10 mg/ml apohTf and holohTf dried film samples. Preprocessing was carried out by baseline correction, normalization and smoothing. Each Raman spectrum was collected in 10s using an excitation length of 785 nm. Assignments are listed in Table 1.1.

These holohTf peaks have been assigned using resonance Raman spectroscopy to CH or CO bending; C-O; C-C and C-C ring stretching modes of iron-bound Tyrosinate respectively (Gaber et al., 1974, Tomimatsu et al., 1976). Miura et al. (2001) also observed these iron-bound peaks in the resonance Raman spectrum of an Fe(III) bound amyloid peptide, distinguishing it in the absence of Fe(III) with spectral variations at ~1176, 1278, 1505 and 1605 cm^{-1}. The maximum absorption of Fe(III) occurs at ~465 nm; thus resonant Raman spectra collected at 488 and 514 nm are dominated by the iron-bound tyrosinate peaks (Gaber et al., 1974, Tomimatsu et al., 1976, Miura et al., 2001). In a more recent study of hTf using non-resonance Raman spectroscopy (785 nm), the holohTf spectra also had a higher intensity at ~1170, 1270, 1500 and 1603 cm^{-1} compared to the apohTf spectra, consistent with the observations in this study (Ashton et al., 2017).

4.3.2 2D-Correlation Analysis

4.3.2.1 apohTf Autocorrelation

Figure 4.6 shows the autocorrelation and pH-dependent spectra of apohTf from ~pH 2.6-11.4. Autopeaks can be observed at ~1237, 1415 and 1655 cm^{-1} indicating the regions of the spectra where the most significant change has occurred in relation to increasing pH. At ~pH 5.6 (native pH) and from ~pH 7.6-11.4, the 1237 cm^{-1} peak assigned to β-sheet (Table 1.1) has a weak intensity compared to spectra at other pH values. At pH 2.6 (Figure 4.6B), the high peak intensity at ~1237 cm^{-1}, indicates the highest amount of β-sheet is present at low acidic pH.
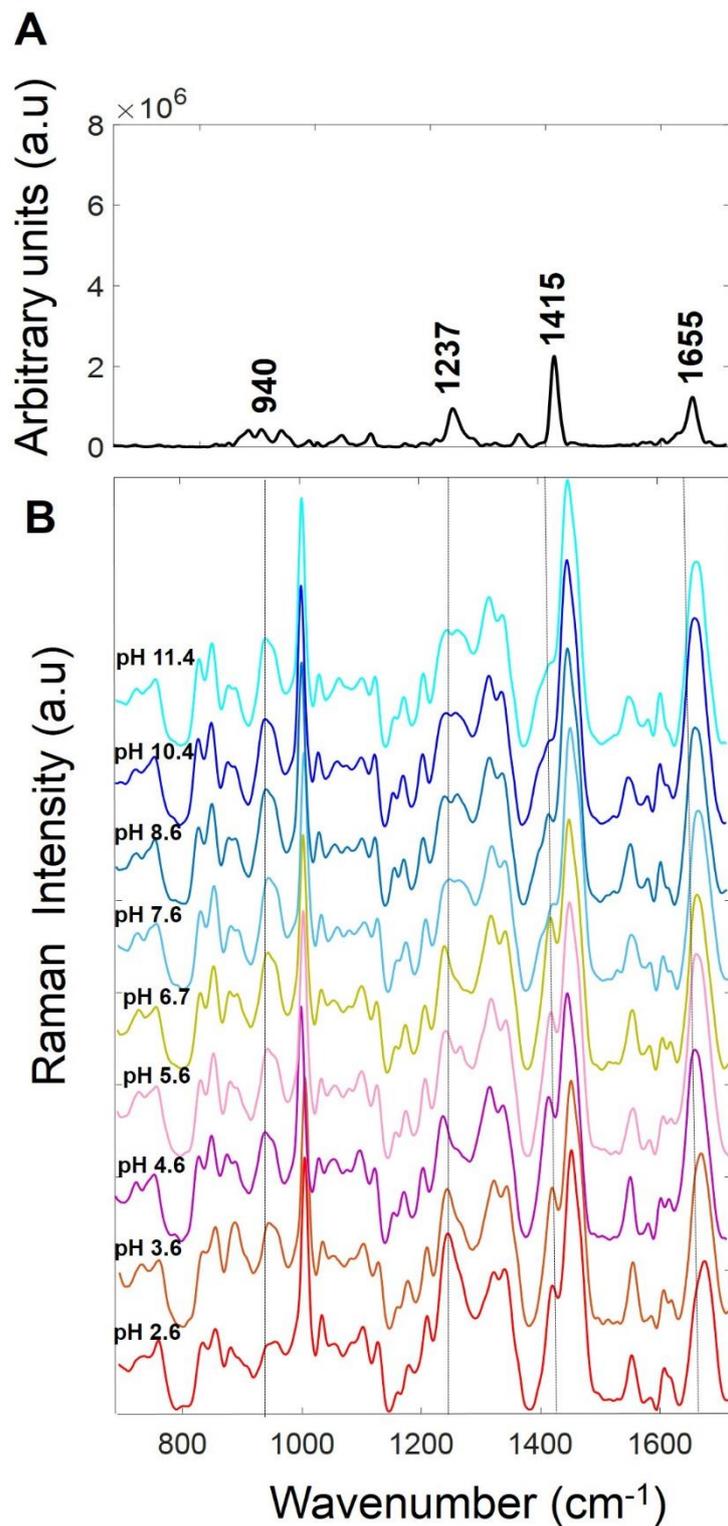
Figure 4.6 pH-dependent Raman spectral variations of 10 mg/ml apohTf solution from ~pH 2.6-11.4. (A) Autocorrelation plot (B) Averaged (n=3) Raman spectra. Spectra were collected at pH intervals ranging from ~1.0-1.8. Preprocessing was carried out by baseline correction, smoothing, normalization, water subtraction and further baseline correction. Each Raman spectrum was collected in 30 mins using an excitation length of 785 nm. Assignments are listed in Table 1.1.

As discussed in section 3.3.3.1, an autopeak at ~1415 cm^{-1} indicates Raman peak intensity variations associated with the ionization of carboxyl groups (Table 1.1). In the apohTf spectra (Figure 4.6B), the peak intensity at ~1415 cm^{-1} has a lower intensity above pH 7.6, suggesting reduced ionization of the carboxyl groups from pH 7.6. There is also an autopeak at ~1655 cm^{-1} indicating that some spectral changes have occurred near the ~1660 cm^{-1} Raman peak. As discussed previously in section 3.3.3.2, the actual position of an auto peak may vary from the position of a Raman peak because the autocorrelation identifies the region of the most significant change which may involve changes in peak width or intensity. However, it is not possible to follow the direction of spectral intensity variation from the stacked spectra (Figure 4.6B). This difficulty in visualizing and interpreting some spectral variations across multiple spectra is a key motivation for the use of 2DCA.

4.3.2.2 holohTf autocorrelation

The autocorrelation of the holohTf spectra (Figure 4.7A) differs from that of apohTf primarily due to the presence of additional autopeaks at ~1280 and 1605 cm^{-1} which indicate spectral variations in the Raman peaks associated with bound iron. In Figure 4.7B, the peak at ~1280 cm^{-1} assigned to C-O stretch of iron bound Tyrosinate is barely observable from pH 2.6 to 3.6 but gradually increases in intensity between ~pH 4.6-5.7. Above pH 5.7, no visible changes in peak intensity are observed. In addition, the peak intensity at 1605 cm^{-1} assigned to the Tyr-Fe-C-C stretch can be observed to increase significantly at pH 4.6 compared to at pH 2.6 and 3.6. These peak intensity changes with increasing pH indicate a dissociation of iron between pH 4.6 and 5.7.
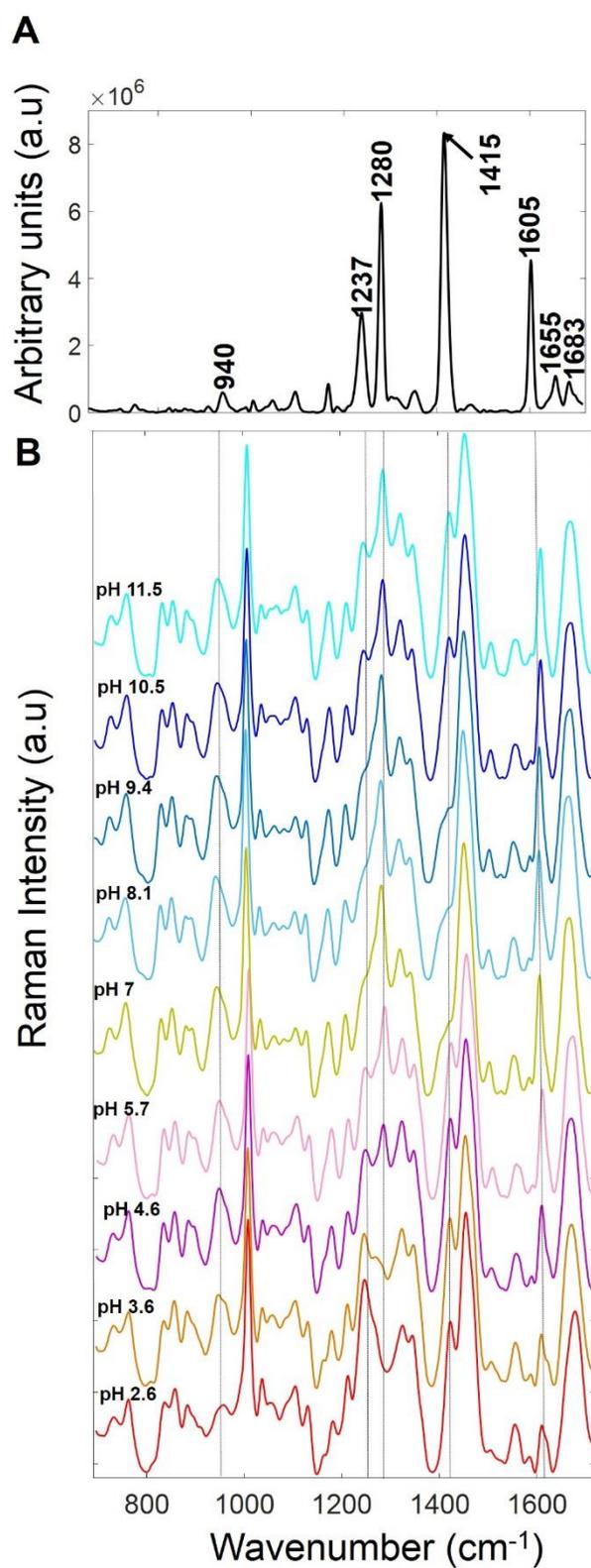
Figure 4.7 pH-dependent Raman spectral variations of 10 mg/ml holohTf from ~pH 2.6-11.5 (A) Autocorrelation plot (B) Averaged (n=3) Raman spectra. Spectra were collected at pH intervals ranging from ~1-1.3. Preprocessing was carried out by baseline correction, smoothing, normalization, water subtraction and further baseline correction. Each Raman spectrum was collected in 30 mins using an excitation length of 785 nm. Assignments are listed in Table 1.1.

The complete dissociation of iron by ~pH 2.6 and 3.6 are evidenced by a drastic colour change in the samples shown in Figure 4.2. As with apohTf, the autopeak at ~1237 cm^{-1} is likely indicative of the highest amount of β-sheet at ~pH 2.6 (Figure 4.7B). This similarity is due to the release of iron from holohTf below ~pH 5.7, thus rendering the sample iron-free (apohTf). Nevertheless, the increased ~1237 cm^{-1} peak intensity changes observed at ~pH 10.5 and 11.5 are also represented by this singular autopeak (Figure 4.7A). This observation demonstrates the composite nature of autopeaks in relation to spectral variations in a data set. Autopeaks represent the change across the entire data set and do not provide insight on how the variations relate to the perturbation. This drawback is overcome using 2DPCMW analysis discussed in section 1.6.4.

It is noteworthy that at native pH (~pH 7), holohTf spectra has no visible peak at ~1237 cm^{-1}. However, there is a 1237 cm^{-1} peak at ~pH 2.6 whose intensity decreases with increasing pH up to 5.7. The termination of this spectral change at ~pH 5.7 corresponds to the end of the iron dissociation, as indicated by no further change in peak intensity at ~1280 cm^{-1} above pH 5.7 (Figure 4.7B). This observation, coupled with the presence of a 1237 cm^{-1} peak at neutral pH in the apohTf spectra (Figure 4.6B) and its decreasing intensity with increasing pH between ~pH 2.6-5.6 suggests a possible link between the presence of iron and a significant decrease in β-sheet. The structural changes accompanying pH-dependent iron release have been reported as mainly involving the tertiary structure (Hadden et al., 1994, Mecklenburg et al., 1997). The observed spectral variations suggest a change featuring secondary structure. Two antiparallel β-strands (backbone β-strands) lie behind the iron-binding cleft, and they facilitate

the opening and closing of the cleft during iron release and binding via a hinge-like structure (Baker et al., 2003). A possible explanation for the observed decrease in β-sheet is that while iron remains bound, the exposure of these β-strands is limited. However, in acidic conditions, when the iron is released, the strands become more exposed to solvent and become denatured. The autopeaks at ~1655 and 1683 cm^{-1} in Figure 4.7A indicate spectral changes near the ~1660 cm^{-1} peak, and are also difficult to follow from the stacked spectra (Figure 4.7B).

4.3.2.3 2DPCMW of apohTf and holohTf.

Figure 4.8 displays the synchronous 2DPCMW plots generated from the pH-dependent Raman spectra of apohTf and holohTf. Apart from the positive contours at ~1280 and 1605 cm^{-1} (Figure 4.8B) representing spectral variations relating to bound iron, the plots have a similar pattern of contours from ~pH 4-8. In Figure 4.8B, the positive contours at 1280 and 1605 cm^{-1} imply an increase in peak intensity with increasing pH in the presence of iron. At ~pH 4, both plots have a positive contour at ~940 cm^{-1}, indicating an increased intensity with increasing pH of the Raman peak at ~940 cm^{-1}. Although the autopeak for ~940 cm^{-1} is very weak (Figure 4.6A and 4.7A), an increased peak intensity at ~940 cm^{-1} can be seen above ~pH 2.6 in the apohTf and holohTf spectra (Figure 4.6B and 4.7B). This observation suggests a reduced amount of α-helix in hTf at ~pH 2.6 compared to higher pH values (Frushour and Koenig, 1974, Ikeda and Li-Chan, 2004). A positive contour at ~1655 cm^{-1} centred at ~pH 4 can be observed in the holohTf 2DPCMW plot (Figure 4.8B). In the apohTf plot, there is also a positive contour at ~1655 cm^{-1} from ~pH 3-5, indicating an increased peak intensity in the nearby Raman peak at ~1660 cm^{-1} with increasing pH (Figure 4.8A). As mentioned previously, this

spectral change is difficult to observe in the stacked spectra (Figure 4.6 and 4.7).

An alternative method of displaying the spectra is to overlay them, as shown in

Figure 4.9.
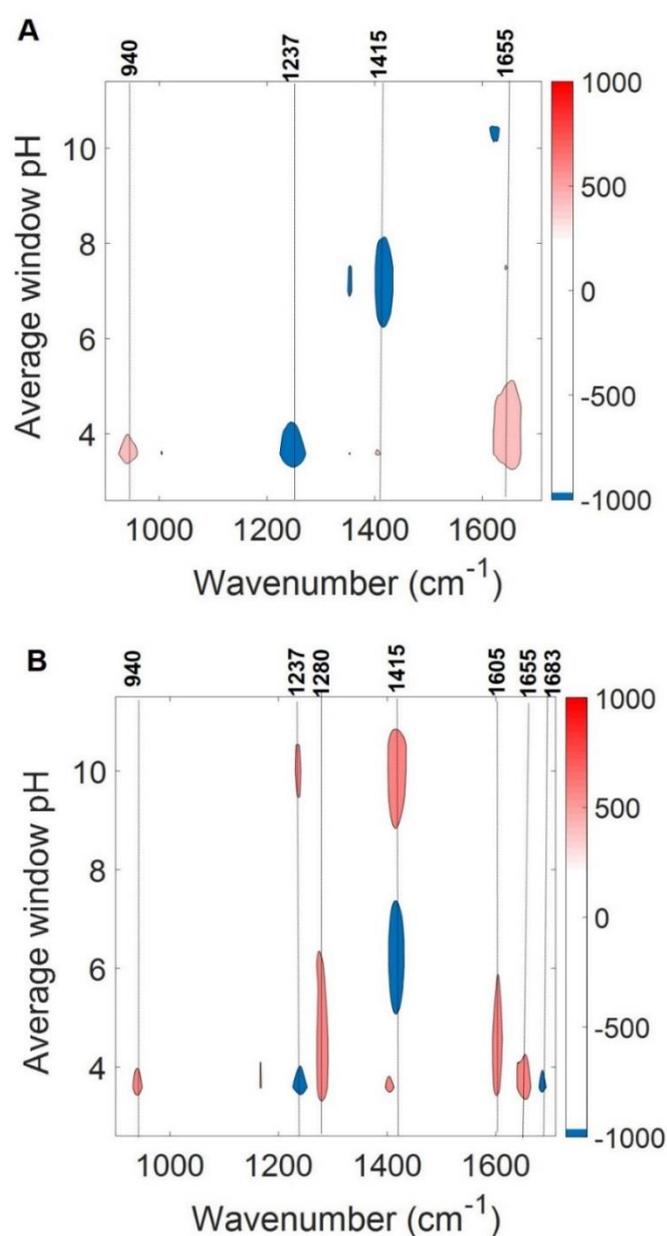
Figure 4.8 Synchronous 2DPCMW plot of hTf generated from pH-dependent Raman spectra (A) apohTf (B) holohTf. Blue shaded contours indicate peaks that are decreasing in intensity with increasing pH. Red shaded contours show peaks that are increasing in intensity with increasing pH. The scale on the colour bar has arbitrary units. A moving window size of 5 and a maximum of 2 contours was applied.
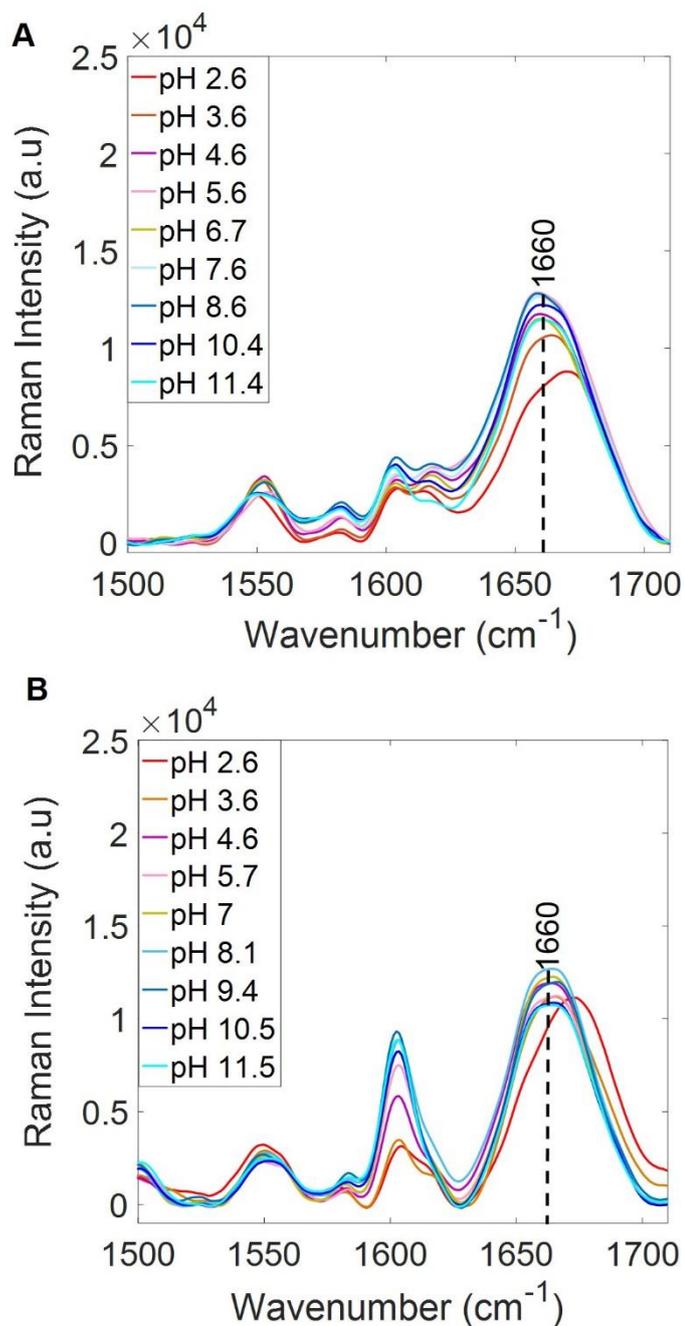
Figure 4.9 Averaged (n=3) pH-dependent Raman spectra of 10 mg/ml hTf solutions from ~pH 2.6-11.5 in the ~1500-1720 cm^{-1} region (A) apohTf (B) holohTf. Preprocessing was carried out by baseline correction, smoothing, normalization, water subtraction and further baseline correction. Each Raman spectrum was collected in 30 mins using an excitation length of 785 nm.

The difference in the pH range occupied by the ~1655 cm^{-1} contours in the apohTf

and holohTf PCMW plots is likely due to the increased baseline observed at ~1630

cm^{-1} in the apohTf spectra collected at ~pH 3.6 and 4.6 (Figure 4.9A). Nevertheless,

a decrease in intensity and a possible shift at ~1660 cm^{-1} can be observed in the ~pH 2.6 spectra for apohTf and holohTf (Figure 4.9). The peak at ~1660 cm^{-1} has been assigned to disordered structure (Pelton and McLean, 2000, Rygula et al., 2013) but also to α-helix structure in globular proteins (Ikeda and Li-Chan, 2004). Therefore, a lower peak intensity at ~pH 2.6 compared to higher pH values suggests a loss of disordered or α-helical structure.

It is interesting to note that the negative contour at ~1683 cm^{-1} is only present in the holohTf plot (Figure 4.8B). This difference is likely due to the pronounced right shift of the peak at ~1660 cm^{-1} in the holohTf spectra compared to apohTf spectra at ~pH 2.6 (Figure 4.9). Although the sample is equivalent to apohTf at ~pH 2.6, the increased perturbation associated with complete loss of iron from the holohTf sample (Figure 4.2) is a possible explanation for this observed peak shift. Thus, the peak shift may indicate the accompanying conformational change due to the vacant iron site involving α-helix or disordered structure.

As in section 3.3.3.3, the occurrence of a peak shift was clarified by examining synchronous and asynchronous plots (Figure 4.10). The synchronous plot of apohTf has a single autopeak at ~1655 cm^{-1} with no cross peaks, and that of holohTf has two autopeaks (~1655 and 1683 cm^{-1}) and two cross peaks referred to as a *four-leaf* clover pattern (Noda and Ozaki, 2005). Their corresponding asynchronous plots do not indicate a peak shift but rather overlapping peaks with changing intensities. The positions of at least three overlapped peaks in the broad peak centred at ~1660 cm^{-1} (Figure 4.9) can be deduced as ~1655, 1672 and 1683 cm^{1}. However, in the case of holohTf, the cross peaks at ~1655 and 1672, and 1683 cm^{-1} are elongated and distorted compared to apohTf possibly due to increased

perturbation at the Raman peak at ~1665 cm^{-1} assigned to α-helix (Ikeda and Li-Chan, 2004) or disordered structures (Pelton and McLean, 2000, Rygula et al., 2013) as well as at ~1672 and 1683 cm^{-1} assigned to intermolecular β-strand or irregular structures (Li and Li, 2009).
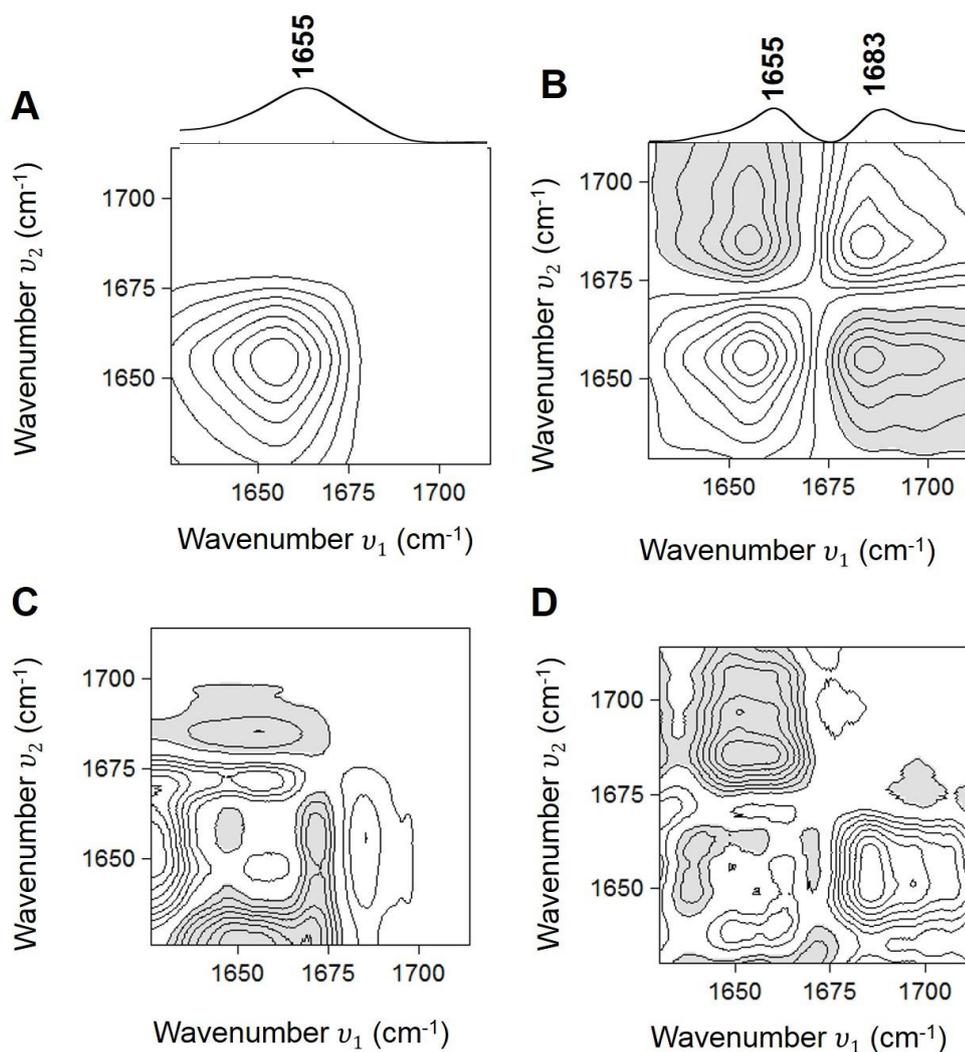
Figure 4.10 Synchronous plot (SP) and Asynchronous plot (AP) generated from pH-dependent spectra of 10 mg/ml hTf solutions (~pH 2.6-11.5). (A) SP apohTf (B) SP holohTf (C) AP apohTf (D) AP holohTf. Shaded contours indicate negative cross peaks and unshaded contours indicate positive cross peaks. Positive cross peaks indicate two wavenumbers with perturbation induced intensity changes in the same direction (increasing or decreasing together). Negative cross peaks indicate two wavenumbers with perturbation induced intensity changes in opposite directions (one increasing and the other decreasing). Maximum contour level=8.

Consequently, the negative contour at ~1683 cm^{-1} in Figure 4.8B suggests that a degree of denaturation due to loss of intermolecular β-strand or irregular structures is occurring only in the holohTf sample as illustrated in Figure 4.11C. The main difference in the 2DPCMW plots can be observed at ~pH 9-11 where a positive contour can be observed at ~1237 and 1415 cm^{-1} in the holohTf 2DPCMW plot only. In the apohTf spectra, the peak at ~1237 cm^{-1} varies minimally in intensity from ~pH 7.6 to 11.4. Conversely, in the holohTf spectra, there is no peak at ~1237 cm^{-1} from ~pH 7-9.4 but appears from ~pH 10.5-11.4. The appearance of this peak at high alkaline pH in the holohTf spectra may also be signalling an increased exposure of β-sheet structure as earlier discussed for acidic pH. In the apohTf and holohTf spectra, a pattern can be observed in spectral variations at ~1237 and 1415 cm^{-1} where they both change in a similar direction (Figure 4.6B and 4.7B). This occurrence suggests that the variation in the β-sheet peak is closely influenced by the ionization of carboxyl groups in hTf. At acidic pH, carboxyl groups of amino acids such as Asp and Glu become protonated whereas, at alkaline pH, carboxyl groups of amino acids such as Lys, Arg and His become deprotonated (Messori et al., 1997). Consequently, the main structural change in hTf (Figure 4.11A) when the carboxyl groups are protonated at low acidic pH is a high amount of β-sheet accompanied by a low amount of α-helix suggesting a transition from α-helix to β-sheet (Tanaka et al., 2009, Litvinov et al., 2012, Amani and Naeem, 2014) as depicted in Figure 4.11B. The positive contour at ~1415 cm^{-1} above ~pH 9 in the holohTf plot is indicative of increasing Raman peak intensity at ~1415 cm^{-1} due to increasing deprotonation (Figure 4.8B).

A

Iron binding sites

B

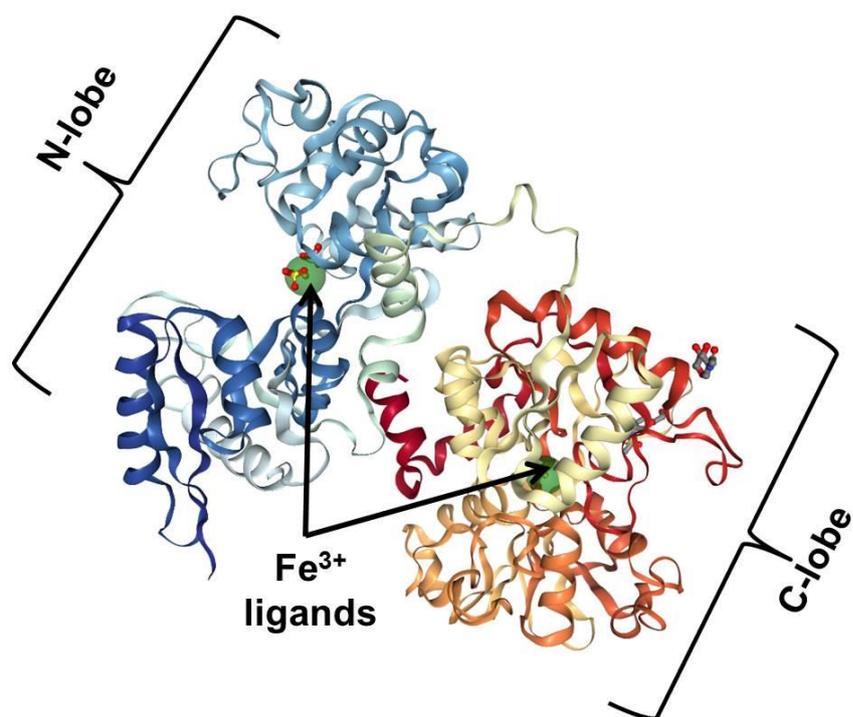(apohTf and holohTf at pH 2.6)

C

(holohTf at pH 2.6)
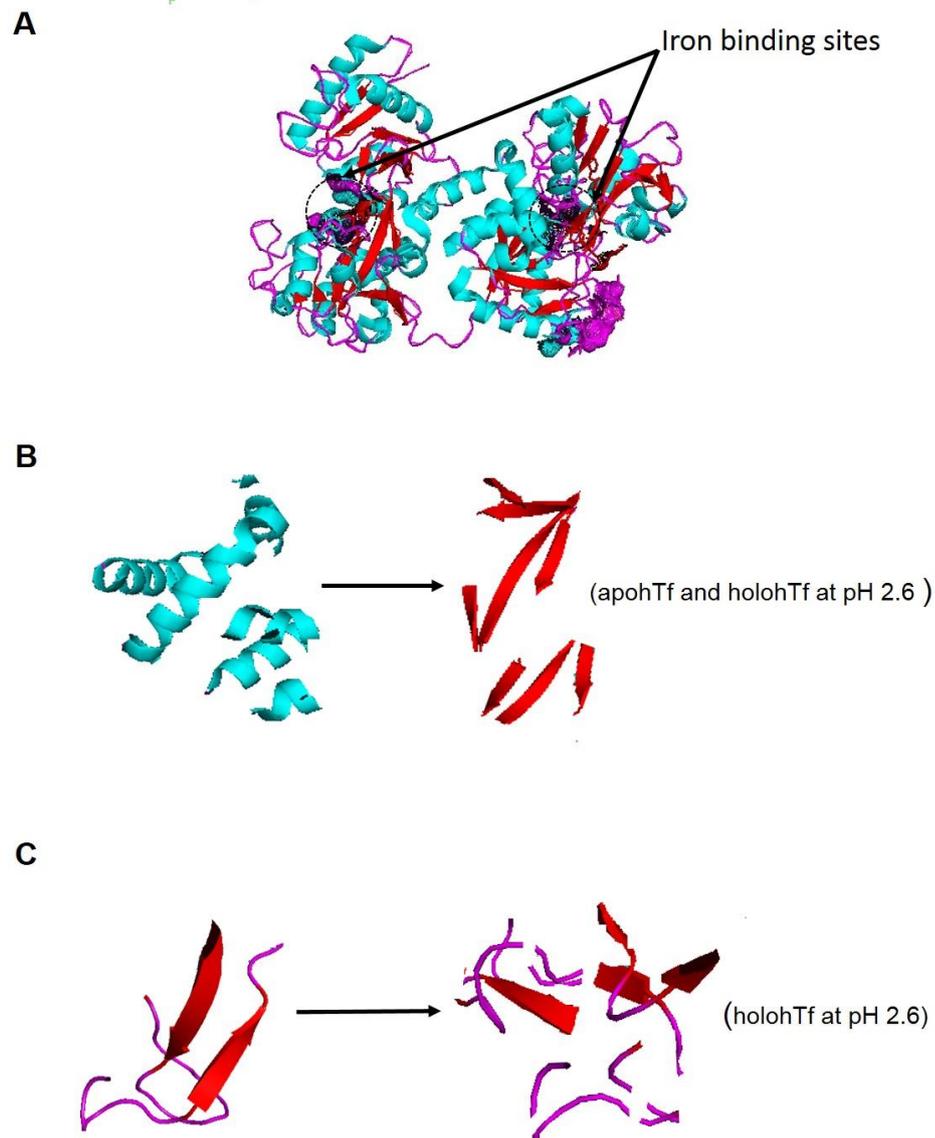
Figure 4.11 (A) Cartoon image of 3QYT (Yang, N, Zhang, H, Wang, M Hao, Q, Sun, H (2012) Iron and bismuth bound human serum transferrin reveals a partially opened conformation in the N-lobe. *Sci Rep* 2:999) created with the PyMol Molecular Graphics System version 2.0. Schrödinger, LLC. Adapted images of 3QYT showing (B) α-helix to β-sheet transition and (C) β-strand or irregular structure loss possibly occurring around Fe binding site of holohTf.

Overall, the results suggest that apohTf and holohTf are stable from ~pH 5-9 and 6-9, respectively. However, at ~pH 10-11, holohTf appears less stable because its spectra reveal a slight increase in β-sheet compared to its native structure and apohTf. Messori et al.(1997) analysed holohTf and observed minimal spectral

changes using CD analysis from ~pH 6-11. However, these changes were not linked with any structural components or compared with apohTf samples. Their study also suggested the loss of iron above ~pH 12; thus providing a possible explanation that the increase in β-sheet from ~pH 10-11.5 in the holohTf sample is occurring prior to the dissociation of iron where the highest amount of β-sheet is expected (Figure 4.8B). However, the dissociation of iron at high alkaline pH was not observed in this study because the holohTf samples were not examined beyond ~pH 12.

4.3.3 Temperature Stability

The stability of apohTf and holohTf samples at native pH were also examined under varying temperatures between ~26-65 °C. Figure 4.12 shows the autocorrelation and temperature-dependent Raman spectra of apohTf. The peak assignments are listed in Table 1.1 (section 1.5.3). As previously discussed, the autocorrelation plot identifies the most significant spectral variations in the data set regardless of the point they occur on the perturbation axis.

The apohTf autocorrelation plot (Figure 4.12A) closely resembles the apohTf spectra in Figure 4.12B because the spectral intensity variations represented by the autopeaks is mainly arising from the spectra collected at ~65 °C. At 65 °C, the peaks at ~722 assigned to methionine C-S stretch (Chen and Lord, 1976, Lord and Yu, 1970), 760, 1337 and 1552 cm^{-1} assigned to Trp have significantly increased intensities compared to lower temperatures suggesting a decreased exposure of Trp residues to solvent upon heating (Gömez De La Cuesta et al., 2014). In addition, increased peak intensities with increasing temperature can be observed at 1270 and 1315 cm^{-1}, assigned to α-helix structure.
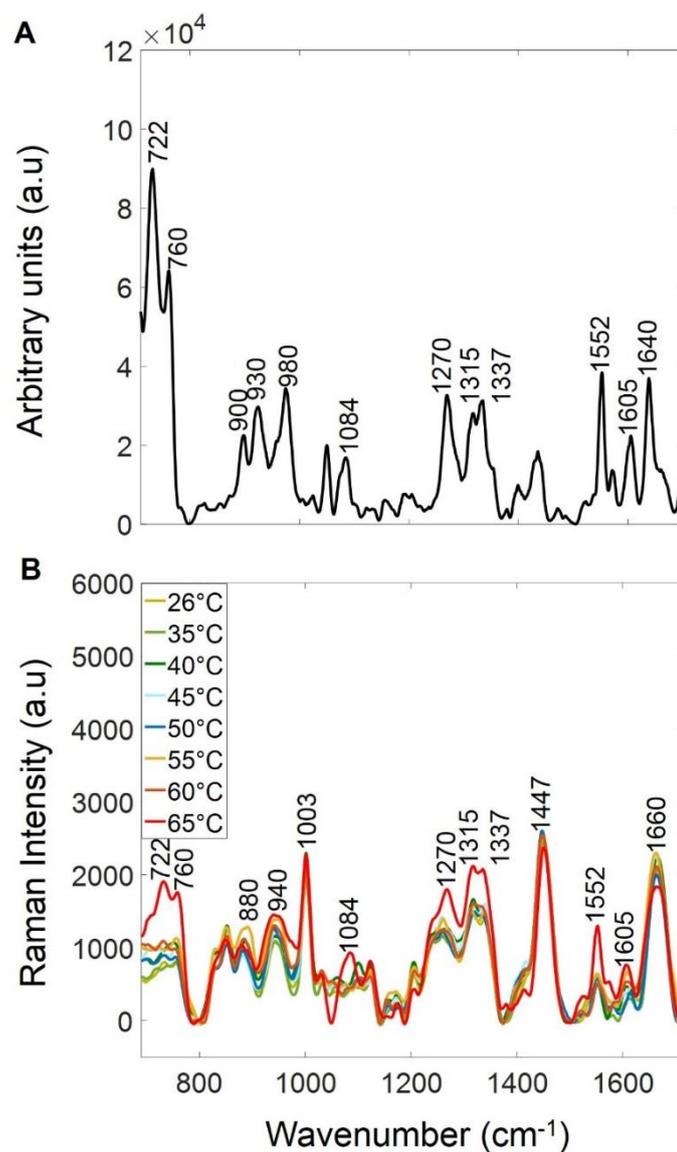
Figure 4.12 Temperature-dependent Raman spectral variations of 10 mg/ml apohTf solutions between ~28-65 °C. (A) Autocorrelation plot (B) Raman spectra. Preprocessing was carried out by baseline correction followed by subtraction of the quartz spectrum, smoothing, normalisation to the peak at ~1447 cm^{-1}, water subtraction and further baseline correction. Each Raman spectrum was collected in 30 mins using an excitation length of 785 nm. Assignments are listed in Table 1.1.
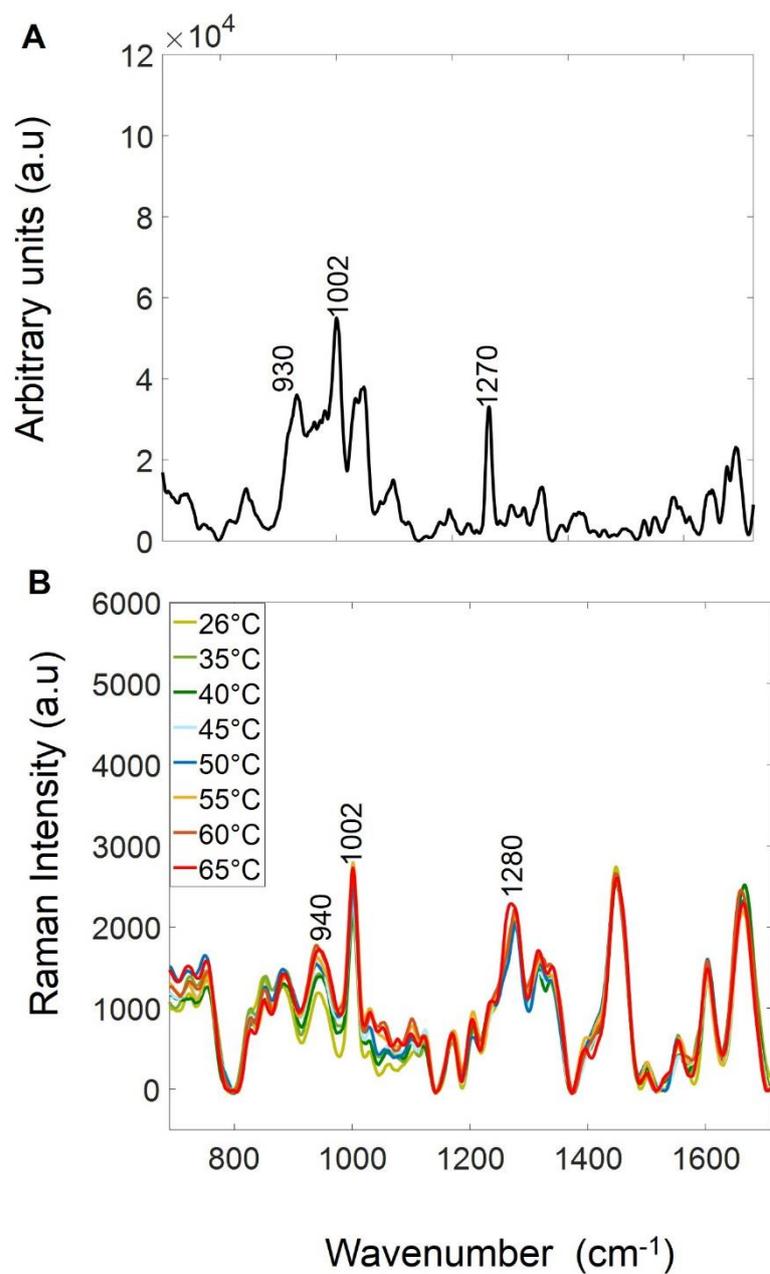
Figure 4.13 Temperature dependent Raman variations of 10 mg/ml holohTf between ~28-65 °C (A) Autocorrelation (B) Raman spectra. Preprocessing was carried out by baseline correction followed by subtraction of the quartz spectrum, smoothing, normalisation to the peak at ~1447 cm^{-1}, water subtraction and further baseline correction. Assignments are listed in Table 1.1.

As discussed in Chapter 3, peaks at ~1315-1337 cm^{-1} have been associated with changes in the geometry of α-helix (Tsuboi et al., 2000). Therefore, these peak intensity changes suggest not only an unfolding of the α-helix structure (~1270 cm^{-1}) but also a change in the α-helical geometry (1315-1340 cm^{-1}) at ~65 °C. Figure 4.13 shows the autocorrelation and temperature-dependent Raman spectra for holohTf collected at the same temperature range. The autocorrelation (Figure 4.13A) has fewer autopeaks compared to the apohTf autocorrelation plot in Figure 4.12A. There is also an autopeak at ~1270 cm^{-1} possibly arising from the broadening at ~1280 cm^{-1} (Tyr-C-O stretch) in the holohTf spectra at ~65 °C compared to lower temperatures. Overall, apoTf and holohTf show minimal differences in their response to temperature except at ~65 °C suggesting that apohTf is less stable than holohTf (Figure 4.12B and 4.13B).

Due to the limitations of the temperature calibration, spectra were not obtained above 65 °C; nevertheless, holohTf shows minimal spectral changes at this temperature compared to apohTf and this observation was consistent with two repeat experiments (65 ± 2.4 °C). Using FTIR, Hadden et al. (1994) also demonstrated that apohTf is thermally unstable above ~60 °C showing a loss in α-helix and β-sheet structure; while little change in α-helix or β-sheet content was observed for holohTf between 50-80 °C. Their Differential scanning calorimetry results showed that holohTf has higher thermal stability than apohTf by up to ~30 °C. ApohTf had two transition points at ~54.5/66.8 °C (onset) and 61/73.2°C (peak) respectively whereas holohTf had a single transition point with onset and peak temperatures at ~83.4 and 90.9 °C respectively.

Similar results have also been obtained for apohTf at ~65°C using optical density measurements (Azari and Feeney, 1958). The higher thermal stability of holohTf is attributed to the additional hydrogen bonding interactions and dynamics that result from the presence of iron (Kumar and Mauk, 2012) The spectral broadening observed at ~1280 cm^{-1} could be signalling the weakening or loosening of these bonds before other structural changes become evident at higher temperatures.
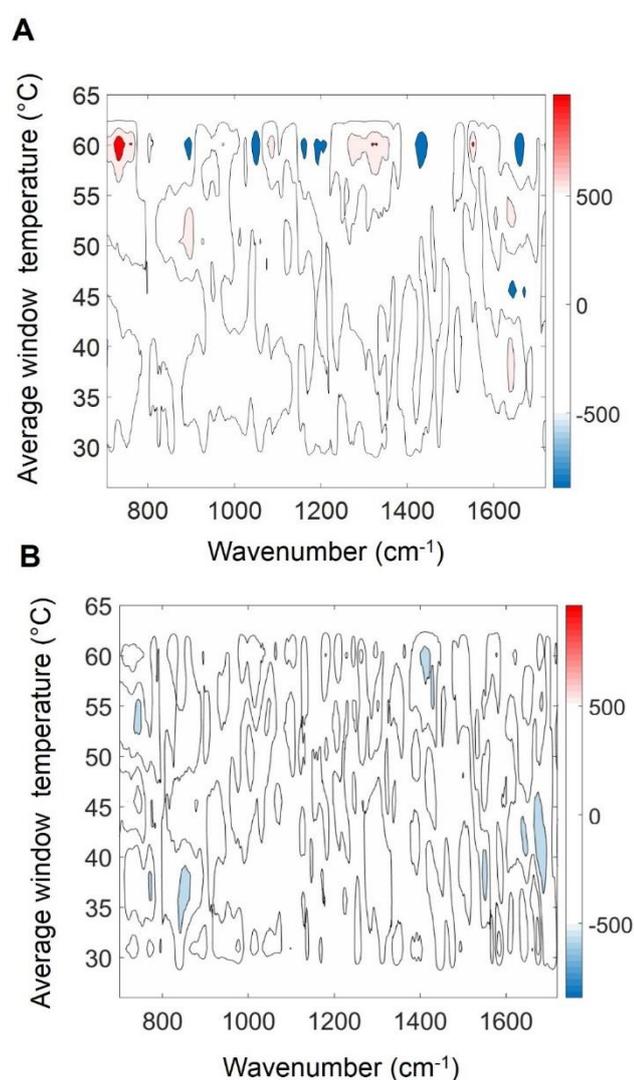
Figure 4.14 Synchronous 2DPCMW of hTf temperature-dependent Raman spectra (A) apohTf (B) holohTf. Blue shaded contours indicate peaks that are decreasing in intensity with increasing pH. Red shaded contours show peaks that are increasing in intensity with increasing pH. The scale on the colour bar has arbitrary units. A moving window size of 5 and a maximum of 4 contours was applied.

Due to the primary spectral change occurring at ~65 °C, the apohTf 2DPCMW plot (Figure 4.14A) shows the most intense contours at ~60-65 °C. In contrast, the holohTf 2DPCMW plot (Figure 4.14B) shows much weaker and random positioning of the contours across the entire temperature range, indicating that no significant change has occurred with respect to increasing temperature. Compared to the pH study, the use of 2DPCMW plots did not necessarily present any advantage over the spectra alone because the main temperature-induced change is occurring at one specific point and can be unambiguously observed in the spectra.

A comparison of the spectra of an unheated holohTf sample with the spectra of the heated holohTf sample cooled down to 20 °C reveals relatively less variations compared to apohTf (Figure 4.15A), providing further evidence for its higher stability (Figure 4.15B). On the contrary, there was a significant difference between apohTf samples examined under similar conditions, indicating that the observed denaturation at ~65 °C is not fully reversible (Figure 4.15A). In Figure 4.15, there is a difference between the peak intensities at 930 and 1315/1337 cm^{-1} (assigned to α-helix structure) in the two samples. With apohTf, there is minimal intensity difference between the unheated and cooled after heating spectra at ~930 cm^{-1} but there is a significantly large intensity variation at ~1315/1337 cm^{-1}. In contrast, the holohTf peak at 1315/1337 cm^{-1} varies only slightly between unheated and cooled after heating spectra whereas the peak intensity at ~930 cm^{-1} is notably lower. A lower peak intensity at ~930 cm^{-1} suggests a reduction in α-helical content in holohTf upon heating whereas the change at ~1315/1337 cm^{-1} suggests a marked change in the α-helical symmetry associated with the

denaturation of apohTf. There are also intensity variations of some peaks assigned to side chains. The two apohTf spectra (Figure 4.15A) reveal large peak intensity variations at ~722-760 cm^{-1} but no peak intensity variations between 830-880 cm^{-1} while the holohTf spectrum has slightly lower peak intensity in both regions. The differences in the peak variations between apo and holohTf in the 722-760 cm^{-1} after heating and cooling suggest a decreased exposure to solvent of Met and Trp residues in apohTf and an increased exposure of these residues in holohTf. In the 830-880 cm^{-1} region, the (I_{850}/I_{830}) ratio of apohTf (1.3) was unchanged between the unheated and heated and cooled spectra. While with holohTf, the ratio decreased slightly from 1.3 to 1.1. These values suggest that for both apohTf and holohTf, the implicated Tyr residues are solvent exposed with a moderate to weak hydrogen bonding that is unaffected by temperatures up to 65 °C. However, the spectral variations at 880 cm^{-1} suggest that Trp side chains have increased exposure to solvent after heating and cooling of holohTf but not apohTf. These opposing spectral variations implicating, Met, and Trp residues and α-helix structure could possibly reflect the differing nature of their conformations.

The reason for observing some spectral changes with holohTf Raman spectra despite not reaching the transition point of holohTf as determined by Hadden et al.(1994) is likely due to difference in techniques and the spectral regions analysed. DSC does not provide information about secondary structure changes may occur before reaching a transition point. Their FTIR study also examined the amide I band which is mainly used for secondary structural analysis with minimal input from overlapping Asn and Gln side chains (Manning, 2005). In addition, they did not examine the cooled spectra after heating. Furthermore, FTIR and Raman

often provide complementary information due to their selection rules (Pelton and McLean, 2000, Smith and Dent, 2005). In contrast to the FTIR study by Hadden et al.(1994), the spectral changes in this Raman study revealed a loss of α-helix but not β-sheet upon denaturation of apohTf and a reduction in α-helix content for holohTf after heating and cooling. It also provided information about the sidechain exposure and α-helical structure upon heating and cooling that possibly reflect the nature of their opposing conformations.
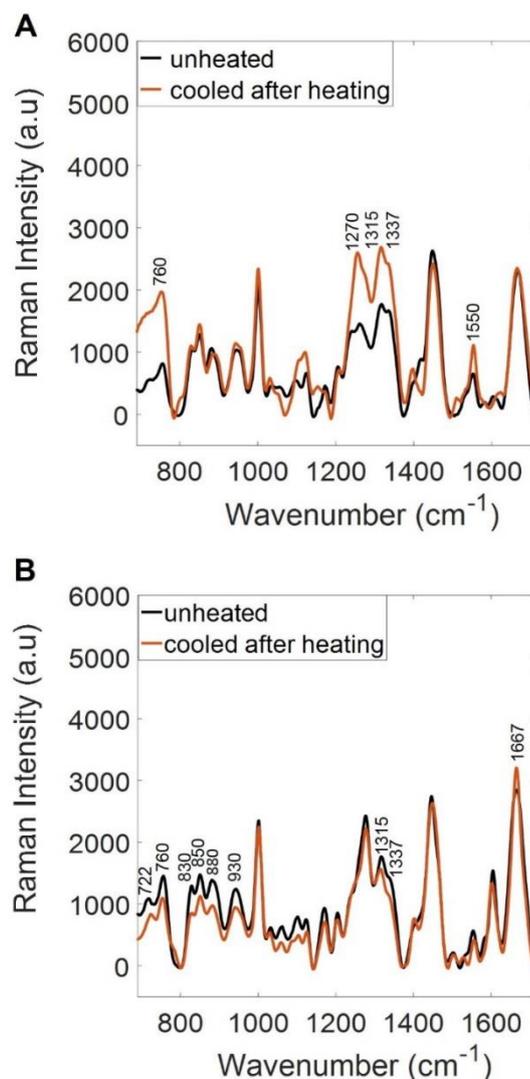
Figure 4.15 Raman spectra of 10 mg/ml hTf solutions before heating and after cooling (A) apohTf (B) holohTf. Unheated sample temperature was ~20 °C. Heated samples (~65 °C) were cooled down to ~20 °C.

4.4 Conclusion

The conformational stability of apohTf and holohTf were examined under varying pH and temperature conditions using Raman spectroscopy and 2DCA. At low acidic pH, an α-helix to β-sheet transition was observed for apohTf, but no comparison could be made with holohTf due to the dissociation of iron below ∼pH 5.7. Despite the similarity between the iron dissociated holohTf sample and apohTf, their ∼pH 2.6 spectra could be distinguished by a more pronounced spectral change which has been assigned to the loss of β-strand sheet or irregular structure in the holohTf sample using synchronous and asynchronous plots.

This spectral change has been proposed to be indicative of a conformational change due to a complete iron removal. An evaluation of the Raman spectra of both samples suggests a link between the increase in β-sheet at acidic pH and the absence of iron, possibly due to increased exposure of the backbone β-strands in the iron cleft upon the removal of iron.

The synchronous 2DPCMW plots were a useful and convenient method to visually compare the stability behaviour of both samples as a function of their response to pH. From the plots, apohTf and holohTf exhibited no significant signs of instability at ∼pH 5-9 and 6-9, respectively. However, above ∼pH 9 holohTf showed a slight increase in β-sheet compared to its native pH and apohTf at the same pH. Overall apohTf was observed to be more stable over a wider pH range compared to holohTf.

The inability to examine temperatures beyond ∼65 °C, limited the possibility of observing more temperature-induced spectral changes and consequently a more

extensive comparison of apohTf and holohTf spectra using the synchronous 2DPCMW plots. Nevertheless, the results suggest that apohTf is less stable than holohTf beyond ~60-65 °C. Using the combination of Raman spectroscopy and 2DPCMW analysis, the stability behaviour of apohTf and holohTf was observed to compare closely over similar regions of the applied perturbation range. Notwithstanding, the differences highlighted by the 2DPCMW plots support apohTf as a more stable candidate overall because compared to holohTf, it appears stable over a wider pH range (~pH 5-11.5) as well as at temperatures which are likely to be encountered in protein drug in use.

5 Antibody Fragment Model

5.1 Introduction

The results from previous studies have demonstrated the capacity of 2DCA and Raman spectroscopy to screen and distinguish the conformational stability profiles of well-studied model proteins. This chapter seeks to extend this application to the investigation of therapeutically relevant proteins amenable to drug development.
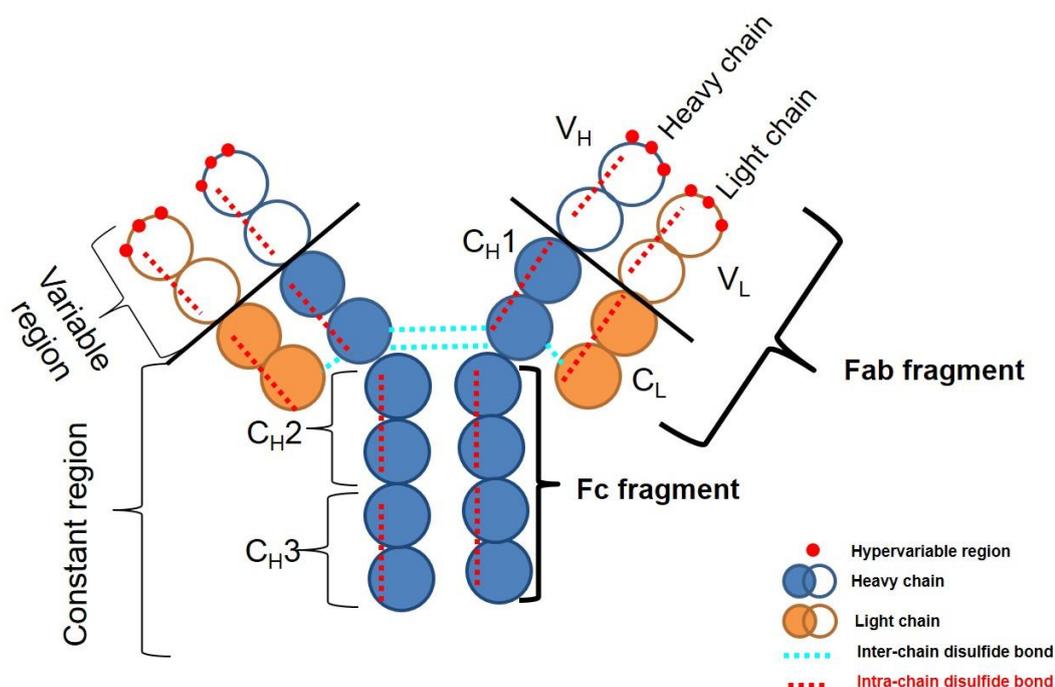
Figure 5.1 Schematic depicting the IgG type antibody structure. VH-variable heavy domain, VL-variable light domain, CH-constant heavy domain and CL-constant light chain domain. Image adapted from Dhar and Das (2018).

As a protein drug class, monoclonal antibodies have steadily dominated the landscape of approved protein drugs (Ecker et al., 2015, Irani et al., 2015); accounting for more than 50% of newly approved biopharmaceuticals between 2015 and 2018 (Walsh, 2018). Virtually all the constituent antibodies of these drugs belong to the IgG class (Beck et al., 2010, Nelson and Reichert, 2009, Buss et al., 2012). A typical IgG antibody structure consists of two identical heavy chains and two identical light chains (Edelman et al., 1969, Williams and Barclay, 1988, Schroeder and Cavacini, 2010). Both heavy and light chains span across a variable region and a constant region, as shown in Figure 5.1. Through proteolytic cleavage, an IgG antibody molecule can be split into three fragments: two antigen fragment binding (Fab) molecules and one Fc (fragment crystallisable) portion. Fab is a monovalent antibody fragment (Figure 5.1) with a molecular weight of ~50 kDa. It consists of four domains; two variable domains (V_H and V_L) and two constant domains, (C_H1 and C_L) linked by an inter-chain disulfide bond (Edelman et al., 1969, Fleischman et al., 1963).

A significant incentive for the development of therapeutic antibody fragments stems from their documented advantages over full-sized antibodies including their smaller size, ease of genetic manipulation, binding affinity, improved tissue penetration as well as cheaper manufacturability (Nelson and Reichert, 2009, Nelson, 2010, Xenaki et al., 2017). Viable measures have been developed to mitigate limitations which are attributable to their lack of Fc domain such as short half-life and reduced affinity for Protein A (commonly used in the purification of whole antibodies). For instance, in the case of short half-life; conjugation to other moieties (Holt et al., 2008) or their use in applications where a long half-life is not

desirable have been explored (Holliger and Hudson, 2005, Weisser and Hall, 2009). Regarding purification, resins such as protein L or non-affinity-based primary capture have been promoted as alternatives for protein A (Rodrigo et al., 2015). In general, these driving advantages have not always translated into improved clinical efficacy or superior commercial success compared to intact antibodies (Nelson, 2010, Dhar and Das, 2018, Xenaki et al., 2017). Nevertheless, the diversity that is readily accessible through the myriad of antibody fragments designs has created an invaluable set of tools that serve to extend the knowledge frontiers in critical aspects such as structure, immunogenicity, manufacture and purification of the entire antibody molecule and its fragments (Nelson, 2010, Dhar and Das, 2018).

Fab molecules were the first group of antibody fragments to be clinically developed and approved for therapeutic use (Nelson and Reichert, 2009, Nelson, 2010). Other antibody fragment classes such as single-chain variable fragments and its variants (diabodies, bi-specific single-chain variable fragments (bis-scFvs) and tandem scFv (tasFv) and third-generation antibody fragment molecules such as nanobodies and domain antibodies have been less successful in reaching the market (Strohl and Strohl, 2012, Bates and Power, 2019). Thus, Fab is the most studied class of antibody fragments and therefore, a representative candidate for testing the utility of 2DCA and Raman spectroscopy in the stability screening of therapeutic proteins and engineered variants.

Protein engineering has steadily advanced over the years to become a dynamic field driving the rational design and optimization of proteins for various applications, including food, drug therapy, and medical diagnostics. This drive is

based on the ever present demand for proteins with the best possible combination of physical and functional properties, suited for their application (Lagassé et al., 2017, Kazlauskas, 2018). Rational design and optimization approaches, generally embrace computational and experimental measures to improve protein expression and stability, reduce aggregation or modulate protein activity (Kamionka, 2011).

One of the most utilized protein engineering machinery is the application of site-directed mutagenesis (SDM) to introduce point or multiple mutations in the nucleotide sequence resulting in the alteration of the amino acid sequence due to substitution of amino acids, deletion or insertion of single or multiple amino acids in the mutant protein (Botstein and Shortle, 1985, Xia and Xun, 2017). These proteins are then screened to understand the effect of the mutation on their activity or other desirable physical properties. Due to the specificity of the SDM technique, specific properties can be targeted during the mutation design based on prior knowledge of the protein as well as the documented effect of a mutation. One such property which can be targeted for mutation is the alteration of the charges on the proteins by substituting, removing or introducing charged or neutral amino acids residues into an existing sequence (Schweiker and Makhatadze, 2009, Raghunathan et al., 2013). Charged amino acid residues can be found on protein surfaces, active sites and the hydrophobic interior of the protein; and may exist alone in pairs or groups (Gitlin et al., 2006, Miller et al., 1987). They are mainly concentrated on the protein surface where they interface with the surrounding water and are readily ionizable (Gitlin et al., 2006, Price and Nairn, 2009, Isom et al., 2010). Their location can affect the nature of their interactions

with other charged groups and their contribution to the electrostatic free energy of the protein (Gitlin et al., 2006, Sanchez-Ruiz and Makhatadze, 2001).

Electrostatic interactions in proteins have been associated with critical processes including, folding, stability, protein-protein interactions and protein-solvent interactions (Simonson, 2003, Nakamura, 2009). Notably, improved thermal stability has been demonstrated by effecting charge mutations (Loladze et al., 1999, Loladze and Makhatadze, 2002, Chan et al., 2012). However, the role of charge-charge interactions in folding and stability has been challenged by some studies which conclude that despite the removal of all charges from the protein, folding and stability are not affected (Loladze and Makhatadze, 2002, Kurnik et al., 2012, Højgaard et al., 2016). These studies argue that the roles of charges lie mainly in solubility and molecular interaction functions. It is well established that even these latter roles have implications for therapeutic proteins. A protein must be soluble for it to be expressed, formulated, and an unfavourable change in molecular interactions could promote aggregation, which is undesirable.

Notwithstanding, the possibility of a conformational change cannot be excluded due to the complexity, long-range nature and non-specificity of charge-charge interactions (Makhatadze, 2017). Therefore, the impact of any charge mutations requires examination on its merit as it may influence protein stability mechanisms including folding, denaturation, protein-protein and protein-solvent interactions.

A wild type Fab (A33Fab-C226S) and four surface charge mutants: HC-K65M, HC-K133M, HC-S75K and L50K were obtained from the Centre for Innovative Manufacturing in Emergent Macromolecular Therapies (CIMEMT), University College London (UCL). For ease of identification, the samples shall be subsequently

referred to as written in brackets: A33Fab-C226S (WTFab), HC-K65M (MT1-1); HC-K133M (MT2-1); L50K (MT3+1) and HC-S75K (MT4+1).

WTFab is a humanised antibody fragment that is reported to be stable at ∼pH 5-7; demonstrating increased apparent conformational stability, retention of secondary structure and minimal monomer loss (Chakroun et al., 2016). MT1-1 and MT2-1 have been mutated to possess a reduced net positive charge while MT3+1 and MT4+1 have an increased net positive charge compared to WTFab. A previous study on the mutant samples reported MT1-1 and MT2-1 as having a decreased stability to aggregation compared to WTFab, while MT3+1 and MT4+1 have increased stability to aggregation compared to WTFab (Ahmad, 2011). A subsequent study by Zhang (2016) on the freeze-dried mutant samples showed that overall, pH 5 had the most impact of monomer loss with MT1-1 having the highest monomer loss, and MT4+1, the least monomer loss. None of these Fab samples have been previously examined using Raman spectroscopy. The study of WTFab and its mutants using Raman spectroscopy is in pursuit of the goal of CIMEMT to develop multiple techniques to assess the cost and challenge of manufacturing a protein molecule, thereby improving access to these products.

The aim of this study was to extend the application of Raman spectroscopy and 2DCA to protein conformational stability screening of therapeutically relevant samples. To this effect, pH-induced spectral variations in the Raman spectra of WTFab and its mutants were investigated and compared using different buffers to identify the most stable sample and or conditions. Furthermore, the impact of freeze-drying on the structure and pH-induced spectral variations were also

assessed. Temperature-dependent spectra were also compared for WTFab and MT1-1.

5.2 Experimental

5.2.1 Sample Preparation:

12.5 mg/ml solutions of Fab samples (WTFab, MT-1, MT2-1, MT3+1 and MT4+1) were supplied by CIMEMT along with 100 mM buffers of sodium acetate (pH 4.5 and 5.5); sodium citrate (pH 4.5 and 5.5) and sodium phosphate (pH 7). To obtain the desired pH with a 20 mM ionic strength, a final concentration of 10 mg/ml of each sample was prepared by mixing the sample and appropriate buffer in a ratio of 4:1. The pH of the protein in buffer solution was measured using a pH meter (section 3.2.1.2) to an accuracy of ±0.1 units.

Freeze-dried forms of all the Fab samples in acetate, citrate and phosphate buffers were also supplied by CIMEMT. 1 mg of each Fab sample was examined without further preparation in its solid form. Reconstitution was performed by dissolving, 5 mg of freeze-dried sample in 0.5 ml of solvent to obtain a concentration of ~10 mg/ml.

5.2.2 Spectra Collection

All Raman spectra for the liquid samples were collected as described for the pH and temperature study of the model proteins in section 2.2.2 and 4.2.4. In addition, a single spectrum of each buffer was collected. For the freeze-dried samples, 8 spectra were collected, and the exposure time was 10s with 12 accumulations (total collection time of 2 mins) and 30 mW laser power at the source.

5.2.3 Data Preprocessing

Except for buffer subtraction which is discussed in section 5.3.2. Data preprocessing was carried out as established for the model samples (section 2.3) The spectrum collected at ~45 °C was excluded from both WTFab and MT1-1 samples due to distortion by a cosmic ray in the former that could not be removed without interfering with the spectrum. For 2DCA, spectra were interpolated in steps of ~0.3 and 2.4 for pH and temperature studies, respectively.

5.3 Results & Discussion

5.3.1 Buffer Influence

Figure 5.2A displays the raw spectra of WTFab in water and water only. In the WTFab spectra, peaks can be observed at ~760, 1003, 1209, 1237, 1447, 1550 and 1640 cm^{-1}. The assignments are listed in Table 1.1 (section 1.5.3). The water spectrum has a broad peak at ~1640 cm^{-1}. In sodium phosphate buffer (Figure 5.2B), there are additional peaks labelled in red in the WTFab spectra at ~875, 990 and 1077 cm^{-1} that can be directly linked to buffer contribution. Figure 5.3 displays the raw spectra of WTFab in sodium acetate and sodium citrate buffer as well as pure buffer samples at pH 4.5 and 5.5. In the WTFab spectra, most of the peaks can be assigned to protein structure. However, there are also additional peaks labelled in red at ~838, 890, 927, 946, 1347 and 1415 cm^{-1} which can directly be linked to buffer contribution. The assignments for the buffer peaks are listed in Table 5.1. As discussed previously in section 1.5.3, a typical Raman protein spectrum also has peaks at similar wavenumbers to these buffer peaks. Protein peaks at ~830/850, 880, 927/946, 1347 and 1415 cm^{-1} can be assigned to Tyr Fermi doublet

(Siamwiza et al., 1975); Trp (Miura et al., 1988);α-helix (Xie et al., 2003) or Trp (Rygula et al., 2013) and ionisation of carboxyl groups (Lord and Yu, 1970) respectively.
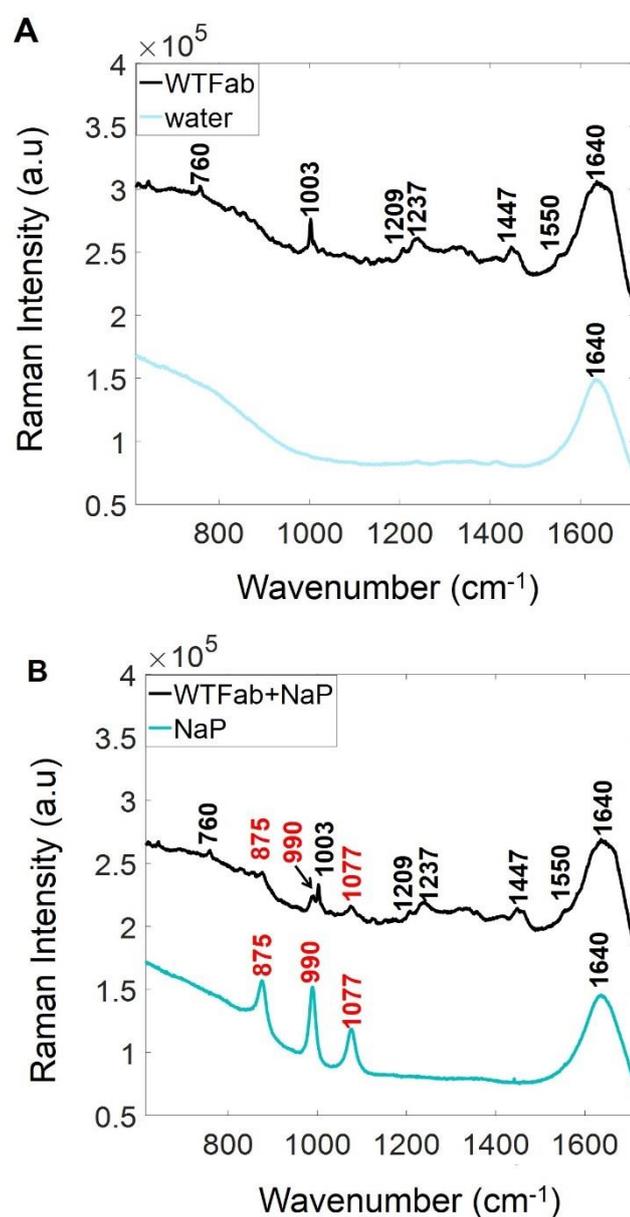
Figure 5.2 Averaged (n=3) raw Raman spectra of 10 mg/ml WTFab, water and sodium phosphate buffer (NaP). (A) WTFab in water and water (B) WTFab in NaP and NaP.Raman spectrum was collected in 30 mins using an excitation length of 785 nm. Protein assignments are listed in Table 1.1 and buffer assignments are listed in Table 5.1.
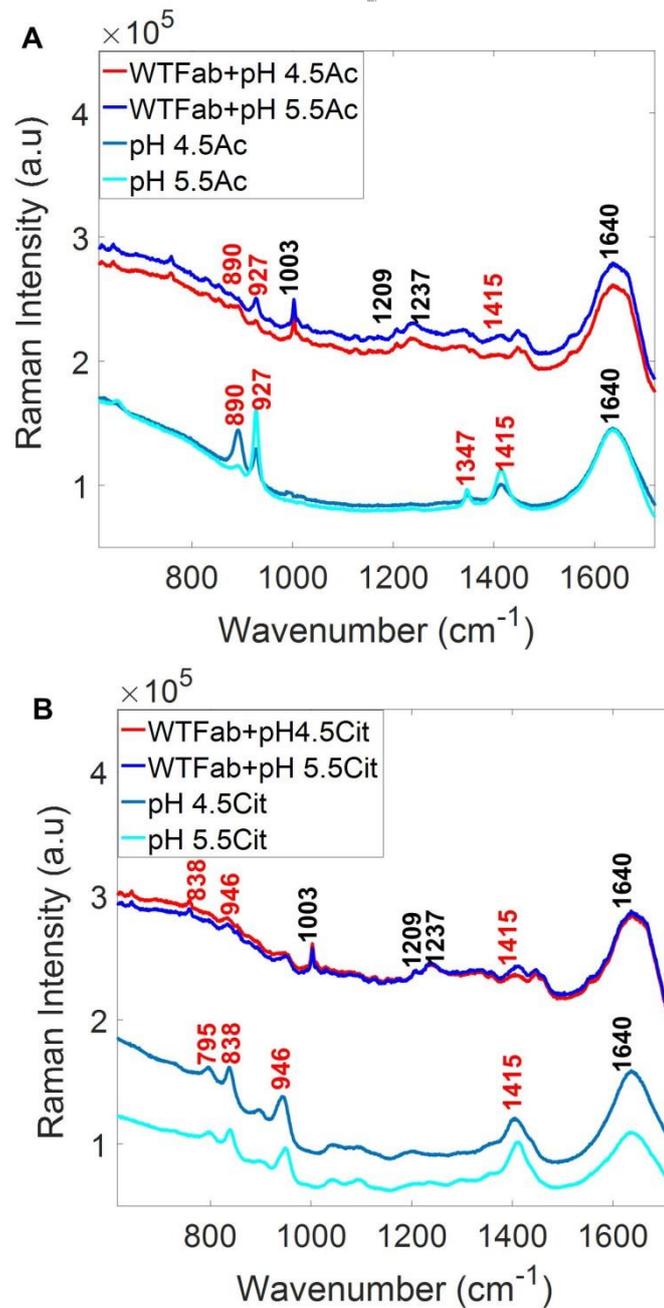
Figure 5.3 Averaged (n=3) raw Raman spectra of 10 mg/ml WTFab in sodium acetate buffer (Ac) and sodium citrate buffer (Cit). (A) WTFab in Ac and Ac at pH 4.5 and 5.5 (B) WTFab in Cit and Cit at pH 4.5 and 5.5. Raman spectrum was collected in 30 mins using an excitation length of 785 nm. Protein assignments are listed in Table 1.1 and buffer assignments are listed in Table 5.1.

This overlap of buffer and protein peaks necessitates the clarification of the degree of buffer influence on any observed spectral variation. Moreover, there is a potential for any change in the structure to be masked as a result of interfering

buffer peaks. One way to overcome this interference is to perform a buffer

subtraction.

Table 5.1 Buffer Raman peaks and proposed assignments.

Buffer	Peak (cm^{-1})	Group	Assignment	References
Sodium acetate	890	CH_3COOH	C-C stretch	(Ng and Shurvell, 1987, Semmler and Irish, 1988, Frost and Kloprogge, 2000)
	927	CH_3COO^-	C-C stretch	(Fournier et al., 1998, Frost and Kloprogge, 2000, Wang et al., 2005).
	1347	CH_3	deformation	(Frost and Kloprogge, 2000, Wang et al., 2005)
	1415	COO^-	C-O stretching	(Ito and Bernstein, 1956, Wang et al., 2005)
Sodium Citrate	797	$H_2C_6H_5O_7^-$	C-C stretch	(Elbagerma et al., 2015)
	838	$H_2C_6H_5O_7^-$	C-C stretch	
	947	$HC_6H_5O_7^-$	C-C stretch	
	1415	COO^-	Symmetric stretching	
Sodium Phosphate	875	$P(OH)_2$	Symmetric stretching	(Marshall and Begun, 1989, Shaver et al., 1998, Xie et al., 2005)
	990	HPO_4^{-2}	Symmetric stretching	
	1077	PO_2	Symmetric stretching	

5.3.2 Solvent Subtraction

In the previously discussed pH experiments, solvent subtraction involved the subtraction of a water spectrum containing a single peak at ~1640 cm^{-1} (section 2.3.4). In this study, three buffers with differing sets of Raman peaks were used to control pH (Figure 5.2B and 5.3). Furthermore, due to the difference in the quantities of basic and acidic components required to obtain a specific pH of the buffer solution; sodium acetate and sodium citrate buffer spectra showed pH-dependent changes at pH 4.5 and 5.5 (Figure 5.3). Therefore, the solvent subtraction process was relatively more complex and was addressed using a range of approaches, including buffer subtraction, water subtraction, as well as a combination of buffer and water subtraction in sequence.

Buffer subtraction was performed by subtracting the corresponding buffer spectrum from the spectrum of WTFab in buffer solution. A water spectrum was also subtracted from WTFab in water as a control sample. The resulting spectra are shown in Figure 5.4. The occurrence of negative buffer peaks is due to using the buffer spectra from undiluted solutions. In the case of water subtraction, an appropriate adjustment of spectral intensity is adequate to reverse the appearance of a negative water peak. This approach for adjusting spectral intensities was applied to reverse the appearance of negative buffer peaks. As an example, Figure 5.5A shows the subtracted spectra obtained after adjusting the spectral intensities of sodium acetate buffer (pH 4.5). After subtracting 50,100 and 150% of the pure buffer spectrum intensity; negative peaks can still be observed in the protein spectra. Although subtracting 150% of pure buffer spectrum may initially appear to be an excessive amount, it should be noted that due to the

varying peak intensities in a spectrum, each peak does not decrease by the same

magnitude during spectral subtraction.
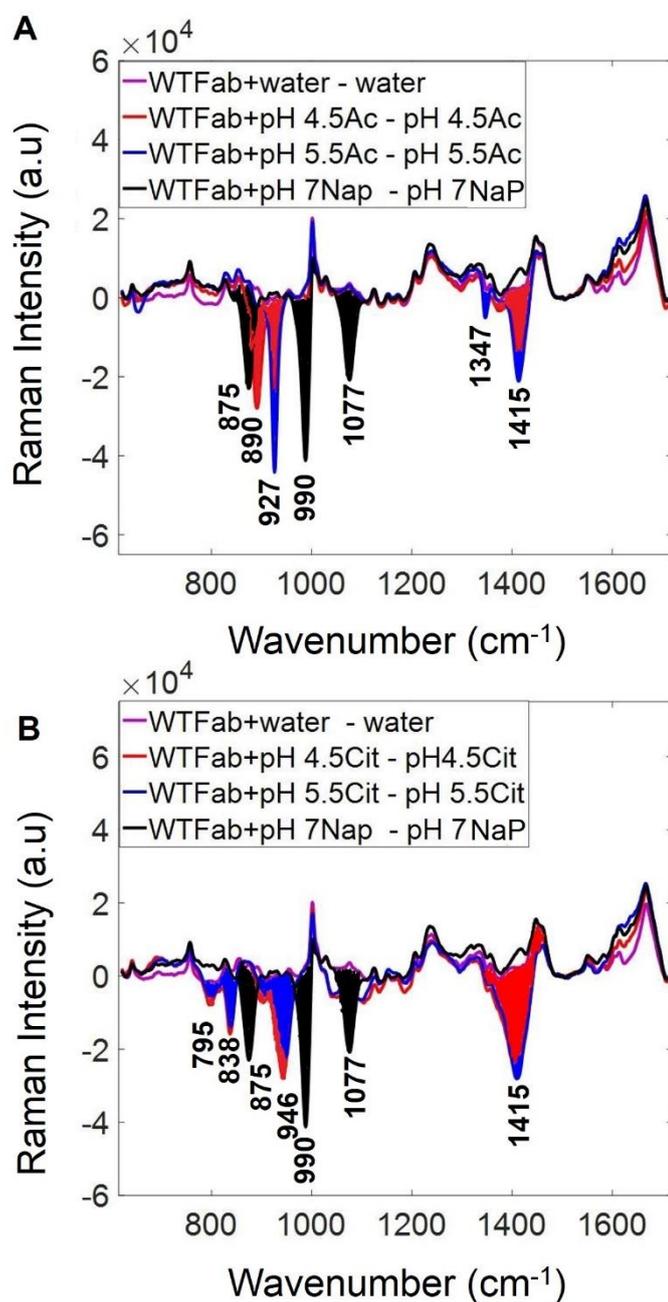
Figure 5.4 Buffer subtracted 10 mg/ml WTFab spectra showing retained negative buffer peaks of sodium phosphate buffer (NaP), sodium citrate buffer (Cit) and sodium acetate buffer (Ac). Preprocessing was carried out by baseline correction, smoothing, normalization and buffer subtraction. Raman spectrum was collected in 30 mins using an excitation length of 785 nm. Buffer peaks are shaded and assignments are listed in Table 5.1.

The spectrum obtained by subtracting 50% of the buffer spectrum intensity had negative buffer peaks with the least intensity; nevertheless, it also reveals a dominant water peak. There appears to be a compromise between the disappearance of buffer peaks and the appearance of the broad water peak.
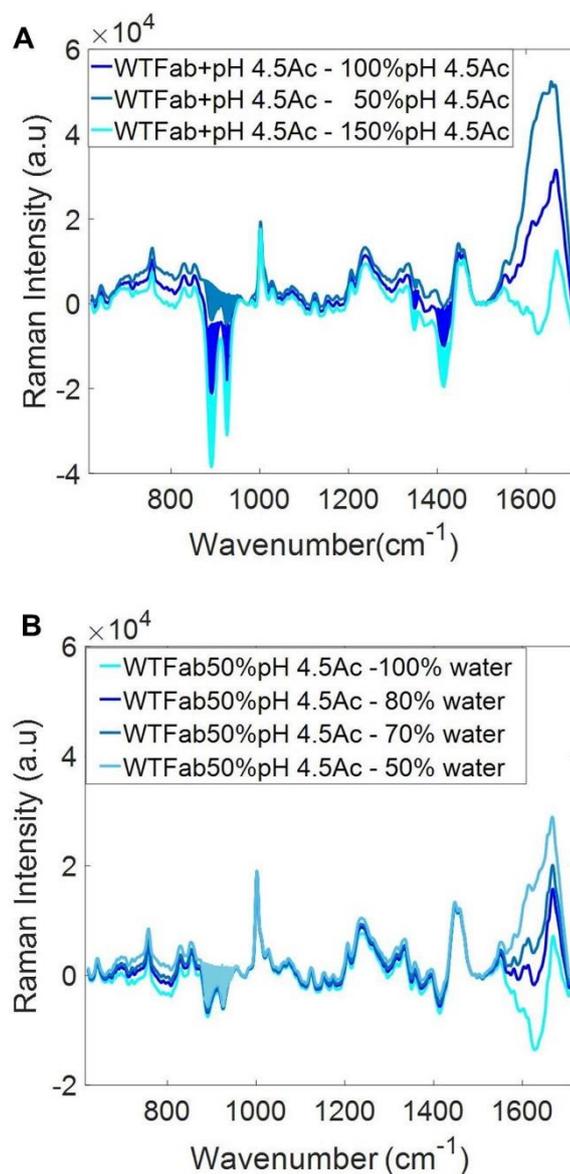
Figure 5.5 Adjustment of spectral intensities for solvent subtraction. (A) 10 mg/ml WTFab in pH 4.5Ac after subtraction of varying proportions of buffer spectrum intensities (B) 10 mg/ml WTFab in pH 4.5Ac after subtraction of 50% of pH 4.5 Ac spectrum and varying proportions of a water spectrum. Buffer peaks are colour shaded. Preprocessing was carried out by baseline correction, smoothing, normalization, and solvent subtraction.

Therefore, the next approach was to try a subsequent water subtraction, as shown in Figure 5.5B. This approach still resulted in the appearance of negative buffer peaks

5.3.2.1 Subtraction of Diluted Buffer Spectrum
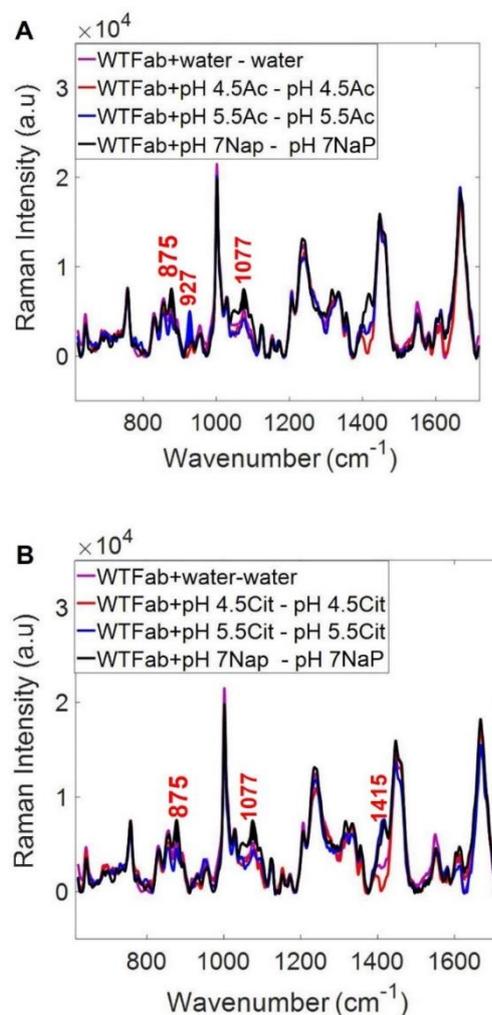
Figure 5.6 Averaged (n=3) 10 mg/ml WTFab spectra after buffer (dilute) subtraction. (A) WTFab in water, Ac and NaP (B) WTFab in water, Cit and NaP. Buffer peaks are colour shaded and labelled in red. WTFab in water is the control sample. Preprocessing was carried out by baseline correction, smoothing, normalization, buffer subtraction and further baseline correction. Raman spectrum was collected in 30 mins using an excitation length of 785 nm.

A further approach was to subtract a buffer spectrum collected at the same concentration as the buffer in the examined protein-buffer solution (1 in 4). Using a diluted buffer spectrum resulted in a partial subtraction of the buffer peaks from the protein spectra (Figure 5.6). Buffer peaks at ~795, 838, 890 and 990 cm^{-1} are no longer evident in the spectra. However positive peaks at ~875 and 1077 cm^{-1} (from sodium phosphate buffer); ~927 cm^{-1} (from acetate buffer) and 1415 cm^{-1} (from sodium citrate buffer) are still retained in their respective spectra. Partial subtraction indicates a difference in the response of individual peaks to the dilution approach. The partial and random retention of some of the buffer peaks in the subtracted spectra was considered to increase the tendency for a biased interpretation of the observed spectral variations. Therefore, water subtraction was considered as an alternative approach.

5.3.2.2 Water Subtraction

Water subtraction was chosen as the preferred method of solvent subtraction for the WTFab spectra due to the incomplete buffer subtraction. Buffer and protein spectra both contain a water peak at ~1640 cm^{-1}. Therefore, it was possible to carry out a water subtraction across all spectra. Figure 5.7 shows the averaged spectra of WTFab after water subtraction. All the buffer peaks are still retained in the spectra and have been shaded for clarity.
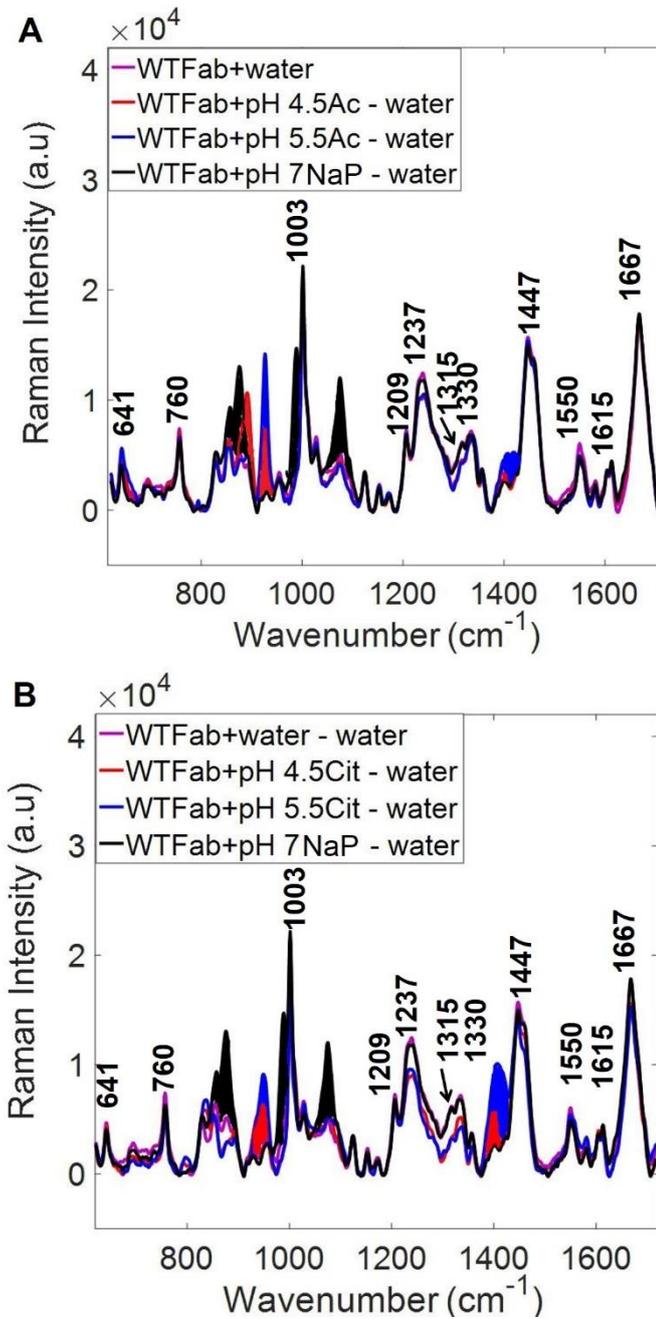
Figure 5.7 Averaged (n=3) 10 mg/ml WTFab spectra after water subtraction only (A) WTFab in water , Ac and NaP (B) WTFab,in water, Cit and NaP. Preprocessing was carried out by baseline correction, smoothing, normalization, water subtraction and further baseline subtraction. Raman spectrum was collected in 30 mins using an excitation length of 785 nm. Buffer peaks are shaded, and protein peaks assignments are listed in Table 1.1.

5.3.3 pH-induced Spectral Variations

The analysis of the pH-induced spectral variations were restricted to the wavenumber regions at ~640-760; 1230-1330; and 1447-1667 cm^{-1} because they contain only protein peaks (Figure 5.7). The most distinct spectral change is the loss of peak intensity at ~1237 cm^{-1} assigned to the β-sheet structure (Lippert et al., 1976, Li and Li, 2009). The loss of β-sheet structure in WTFab at pH 4.5 and 5.5 compared to at ~pH 7 and in water suggests a degree of denaturation is occurring at lower pH. At pH 4.5 and 5.5, there is also a loss of peak intensity at ~1315 cm^{-1} and 1315-1337 cm^{-1} in acetate and citrate buffer respectively suggesting a change in the α-helical symmetry (Tsuboi et al., 2000). Previous studies by Chakroun et al., 2016 using CD showed that there was no significant change in the secondary structure of WTFab at 20°C between pH 3.5-9, with ionic strength up to 100 mM. Changes in the secondary structure were only observed on heating WTFab, with it being most stable between ~pH 5-7. However, it was difficult to resolve the observed CD spectral changes into actual component structures. A potential reason for the difference in results is that CD is known to be less sensitive in estimating the secondary structure content particularly in proteins which have a mixture of α-helix and β structures (Micsonai et al., 2015, Khrapunov, 2009, López-Peña et al., 2015).

Two recent studies on WTFab using other techniques have confirmed conformational changes at low pH that are consistent with these results. Using small-angle x-ray scattering, Codina et al. (2019a) reported the occurrence of a more expanded conformation at pH 3.5, 4.5 and 5.5 compared to at ~pH 7. This conformational change was demonstrated to be limited to the C_L domain using

molecular dynamic modelling and single-molecule Förster resonance energy transfer. The loss of β-sheet following domain unfolding and loss of domain interfacial contacts at ~pH 3.5 has also been demonstrated using molecular dynamic simulation (Codina et al., 2019b).

5.3.3.1 Effect of Buffer Type on pH-induced Variations

WTFab spectra collected at pH 4.5 and 5.5 in two buffers (sodium acetate and sodium citrate) were compared to understand the effect of buffer type on the observed pH-induced spectral variations. In both buffers, WTFab shows a decreased peak intensity at ~1237 and 1315 cm-1 at pH 4.5 and 5.5 compared to at pH 7 and in water (Figure 5.8). However, the spectra of WTFab in citrate buffer has a decreased peak intensity at ~1337 cm-1 compared to in acetate buffer. This difference in the case of WTFab in citrate buffer suggests a distinction between the interaction of sodium citrate and sodium acetate ions with the protein.
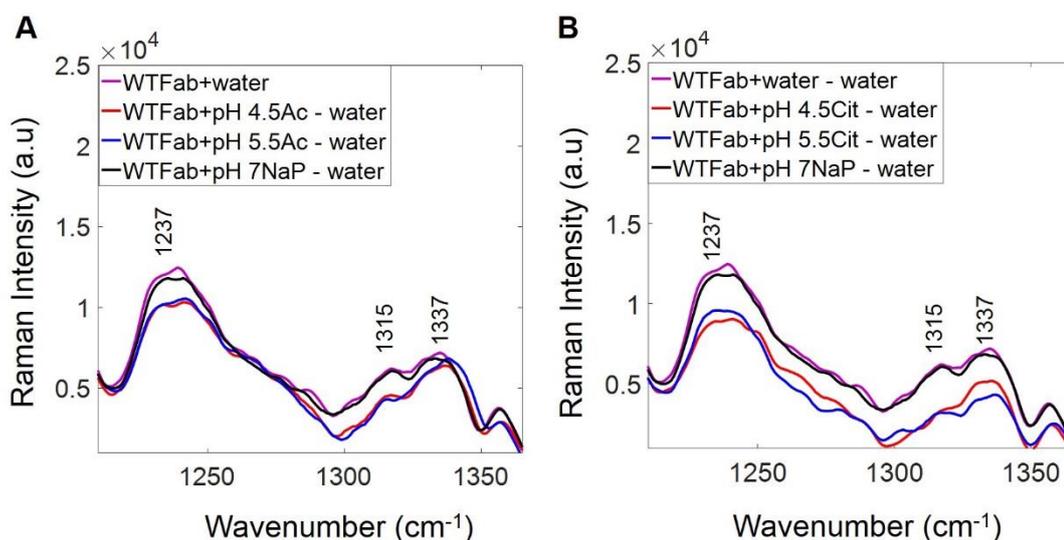
Figure 5.8 Averaged (n=3) Raman spectra (~1210-1370 cm^{-1}) of 10 mg/ ml WTFab (A) WTFab in water, Ac (pH 4.5 and 5.5) and NaP (pH 7) (B) WTFab in water, Cit (pH 4.5 and 5.5) and NaP (pH 7). Preprocessing as stated for Figure 5.7.

Citrate ions have been shown to preferentially accumulate on the protein surface compared to acetate ions (Barnett et al., 2015). Their surface location could facilitate enhanced interaction with α-helix side chains.

5.3.4 2D Correlation Analysis

As with the protein models in chapter 3 and 4, 2DCA was applied to the Raman spectra to further analyse the observed pH-induced variations. The overlapping buffer peaks, which are also changing in intensity, present a challenge compared to previously discussed samples with only protein peaks in the Raman spectra.

The autocorrelation and synchronous 2DPCMW plots generated from WTFab Raman spectra (pH 4.5-7) are shown in Figure 5.9. As expected, the dominance of the buffer peaks in the spectra is reflected in both plots. The strongest auto peaks (875, 946, 990, 1077 and 1415 cm^{-1}) and contours (2DPCMW plot) can be linked to the buffer whereas the weaker autopeaks and contours at ~1237 and 1337 cm^{-1} arise from protein structural changes. Their positive sign, however, indicates that both peaks increase with increasing pH. The Raman spectra (pH 4.5 and 5.5) used to generate the plots were collected in citrate buffer. However, based on the spectra, the same dominance of buffer peaks would be expected even with acetate buffer (Figure 5.7). An examination of the spectral changes in both buffers using synchronous and asynchronous plots confirms that peaks at ~1237 and 1337 cm^{-1} are changing in the same direction as indicated by the positive cross peaks (synchronous plots). The cross peaks in the asynchronous plots indicate that the spectral change at ~1315-1337 cm^{-1} occurs predominantly before that at 1237 cm^{-1}, in both acetate and citrate buffers.
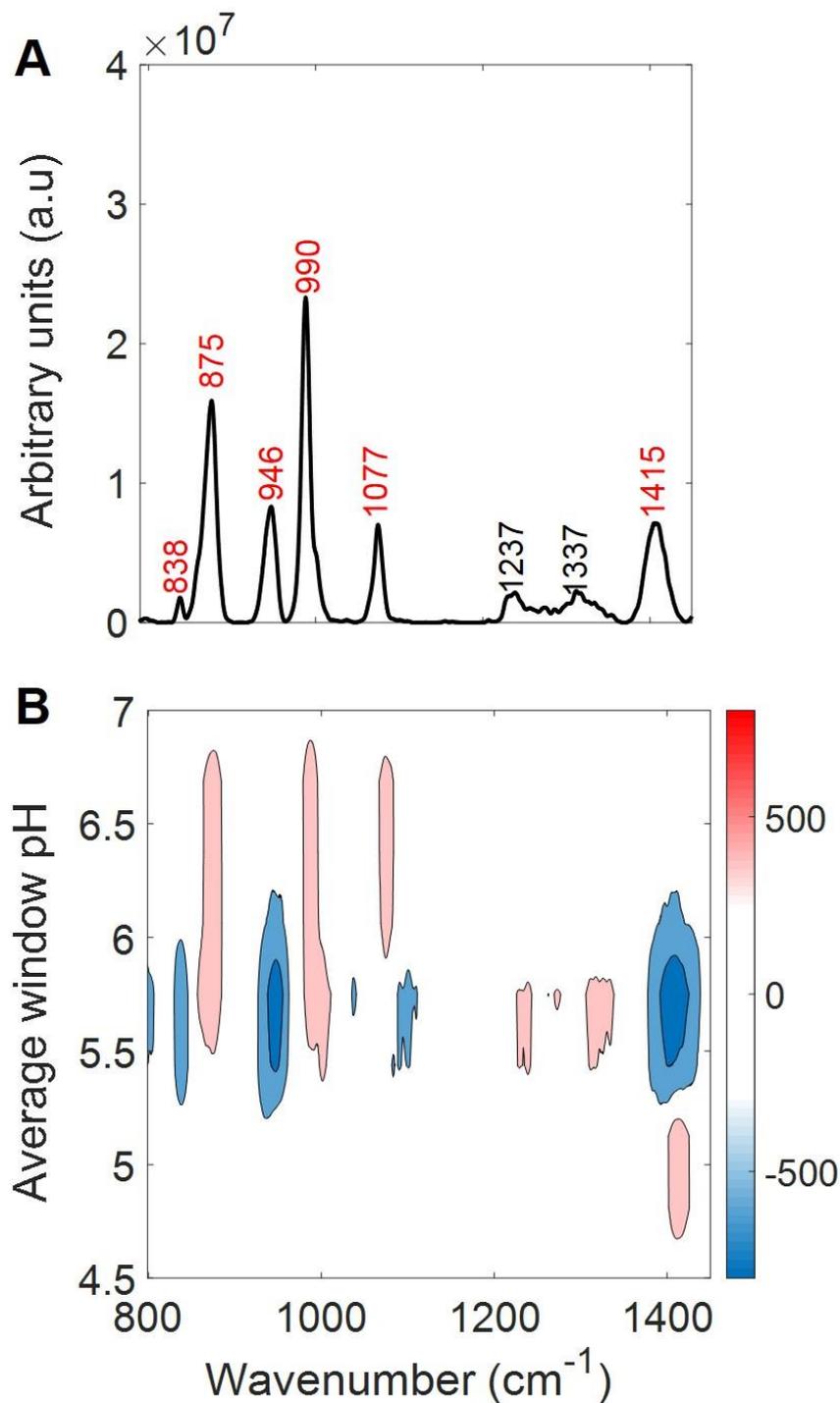
Figure 5.9 (A) Autocorrelation and (B) synchronous 2DPCMW plot generated from the pH-dependent Raman spectra of WTFab in water, Cit (pH 4.5 and 5.5), NaP (pH 7) and water. Buffers autopeaks are labelled in red. Protein autopeaks are labelled in black. Positive contours are shaded in red while negative contours are shaded in blue. Positive contours indicate increasing peak intensities in the direction of the perturbation while negative contours indicate decreasing peak intensities in relation to the perturbation. A darker shade and an increased number of contours indicate a relatively greater change in intensity. The scale on the colour bar is arbitrary. A moving window size of 3 and a maximum number of 4 contours was applied.
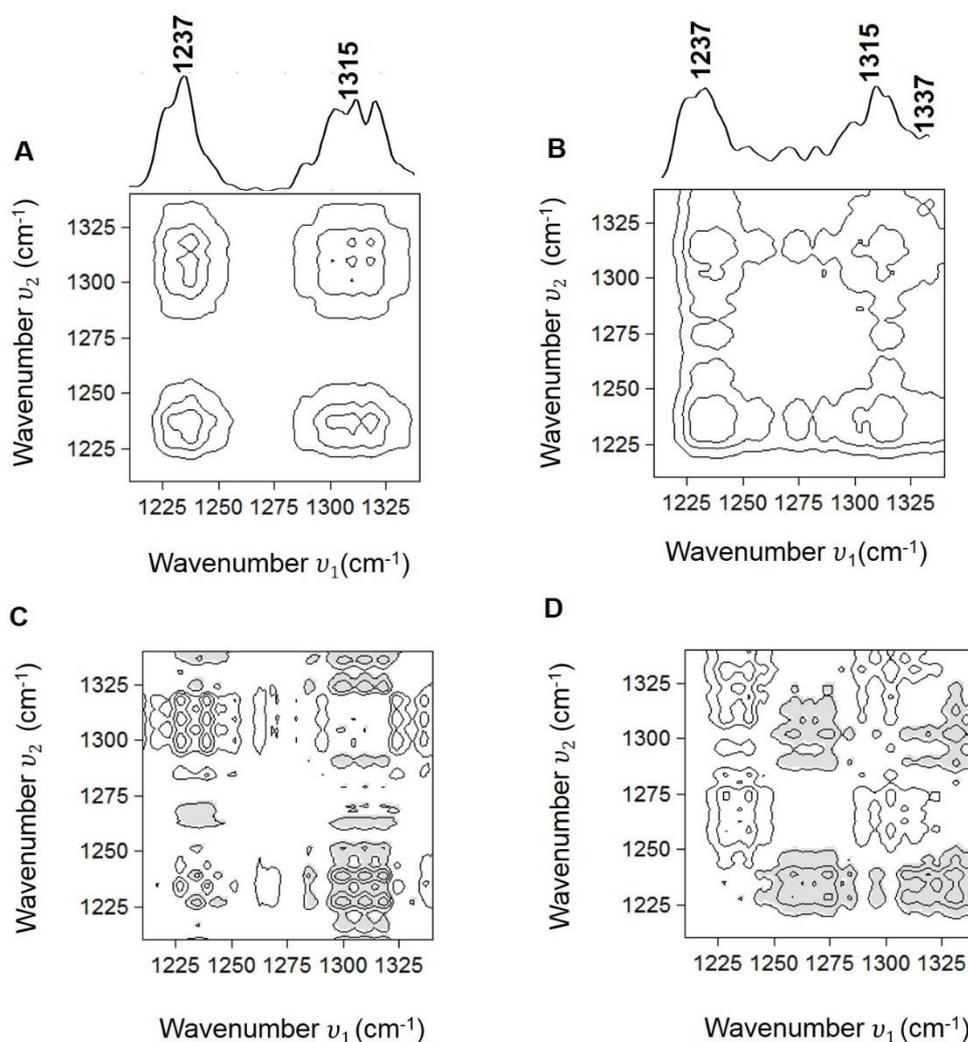
Figure 5.10 Synchronous plot (SP) and Asynchronous plot (AP) generated from pH-dependent spectra of 10 mg/ml WTFab (pH 4.5-7) (A) S P WTFab in water, Ac and NaP (B) SP WTFab in water, Cit and NaP (C) AP WTFab in water, Ac and NaP (D) AP WTFab in water, Cit and NaP. Shaded contours indicate negative cross peaks and unshaded contours indicate positive cross peaks. Positive cross peaks indicate two wavenumbers with perturbation induced intensity changes in the same direction (increasing or decreasing together). Negative cross peaks indicate two wavenumbers with perturbation induced intensity changes in opposite directions (one increasing and the other decreasing). Maximum number of contours=4.

The limited number and simplicity of the observed spectral changes arising solely from protein structure compared to the previously discussed samples did not provide an occasion to fully explore the benefits of 2DCA, which performs to capacity in the context of multiple or complex responses.

5.3.5 Raman Spectra of Mutant Fab Samples and WTFab

The four charge mutants were prepared, as stated in section 5.2.1. Figure 5.11A compares the Raman spectra of the aqueous solutions of the Fab charge mutants and WTFab, while Figure 5.11B compares the difference spectra of the four mutant samples after subtraction of WTFab. It is difficult to identify any significant difference between the mutant samples or in comparison to WTFab from Figure 5.11A alone. However, in the difference spectra, there are some spectral variations at $\sim$1003 and 1415 cm^{-1} (Figure 5.11B). MT1-1 and MT2-1 have a positive peak intensity at $\sim$1003 cm^{-1}, whereas MT3+1 and MT4+1 have a negative peak intensity. A Raman peak at $\sim$1003 cm^{-1} is assigned to Phe, and its increased peak intensity has been associated with increased hydrophobicity (Xu et al., 2005, Xu et al., 2008) as well as pre-aggregation molecular interactions (Ota et al., 2016).

The varying difference peaks suggest that compared to WTFab, the reduced positive charged mutants (MT1-1 and MT2-1) may have a more hydrophobic environment around the implicated Phe residues and an increased tendency to aggregate whereas the increased positive charge mutants (MT3+1, MT4+1) have a more solvent-exposed environment and a less tendency to aggregate. This deduction is consistent with the previous observations by Ahmad (2011). Nonetheless, it has been obtained from the simplicity of aqueous solutions without buffers and in non-aggregating conditions highlighting the predictive capability of Raman spectroscopy. Furthermore, these results suggest that the environment of Phe residues is indicative of each mutant's aggregation tendency. The negative peak intensity at $\sim$1415 cm^{-1} assigned to ionized carboxyl groups can be observed in every mutant's difference spectrum except MT4+1 (Figure 5.11B).
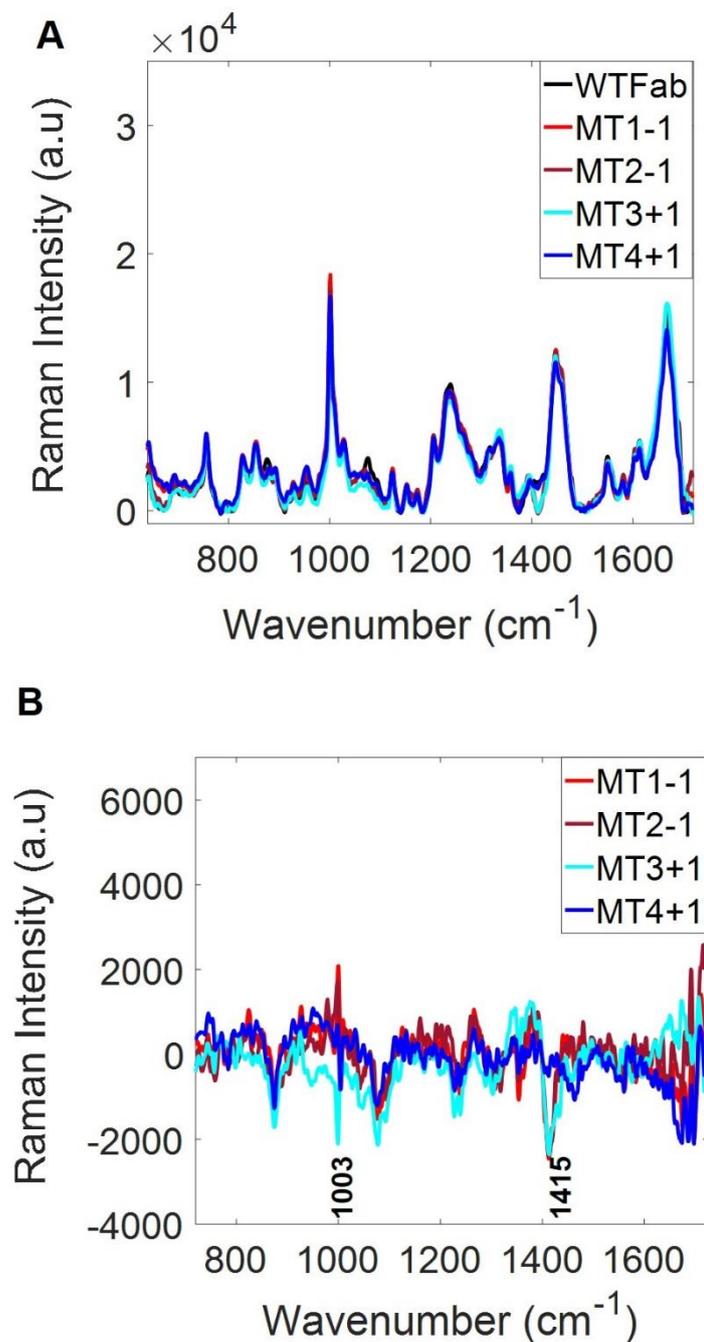
Figure 5.11 (A) Averaged (n=3) Raman spectra of 10 mg/ml WTFab, MT1-1, MT2-1, MT3+1 and MT4+1 solutions (B) Difference spectra of surface charged mutant Fab samples. Preprocessing was carried out by baseline correction, smoothing, normalization, water subtraction and further baseline subtraction. Raman spectrum was collected in 30 mins using an excitation length of 785 nm. Difference spectra were obtained by subtracting the WTFab spectrum from each mutant spectrum.

This difference can be confirmed by an examination of their spectra (Figure 5.11A) which shows that whereas MT4+1 and WTFab have a shoulder at ~1415 cm^{-1}, the other samples have a defined peak instead. These observations suggest that MT4+1 has a similar profile of ionized carboxyl groups to WTFab, and this could have implications for their pH-dependent behaviour since this peak varies significantly with pH.

5.3.6 pH-induced Variations

As with the pH-dependent WTFab spectra discussed previously, only limited regions of the spectra were analysed due to buffer interference (Appendix A).

Figure 5.12 compares the ~1210-1370 cm^{-1} region for all four mutants in water, sodium acetate and sodium phosphate buffer solutions. For MT1-1, slightly increased peak intensities can be observed at ~1237, 1315 and 1337 cm^{-1} at pH 7 (sodium phosphate buffer) compared to at pH 4.5 and 5.5 (acetate buffer) and in water. This spectral variation suggests that MT1-1 has a slightly increased β-sheet and a different α-helix symmetry at ~pH 7 (sodium phosphate buffer) than at the other examined pH (Figure 5.12A). In contrast, a loss of peak intensity at ~1237 cm^{-1} (loss of β-sheet) and 1315 cm^{-1} (α-helix) at pH 4.5 can be observed for MT2-1 (Figure 5.12B).

No significant peak intensity variations at either ~1237 and 1315-1337 cm^{-1} can be observed in the MT3+1 spectrum (Figure 5.12 C), suggesting that MT3+1 does not undergo a similar loss of β-sheet as WTFab in acetate buffer (pH 4.5 and 5.5).
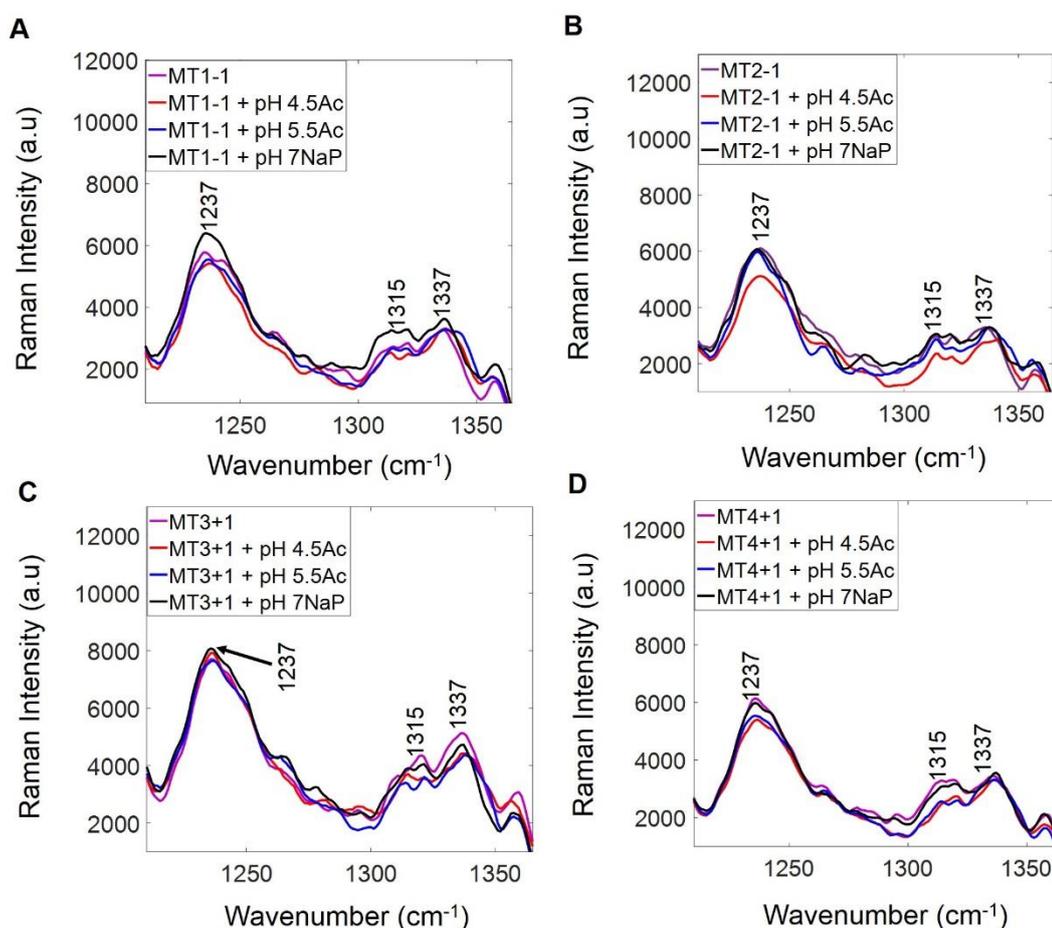
Figure 5.12 Averaged (n=3) Raman spectra of 10 mg/ml mutant Fab samples in sodium acetate (Ac), sodium phosphate (NaP) and water (A) MT1-1 (B) MT2-1 (C) MT3+1 (D) MT4+1 Preprocessing was carried out by baseline correction, smoothing, normalization, water subtraction and further baseline subtraction. Raman spectrum was collected in 30 mins using an excitation length of 785 nm. Protein assignments are listed in Table 1.1.

On the contrary, a loss of peak intensity at ~1237 and 1315 cm^{-1} can be observed at pH 4.5 and 5.5 for MT4+1 (Figure 5.12D). The results suggest that apart from MT4+1, each mutant has a slightly different pH-dependent behaviour in comparison to WTFab. MT1-1 and MT3+1 exhibited no loss of β-sheet at pH 4.5 and 5.5 in acetate buffer, and MT2-1 only shows a loss of β-sheet at ~pH 4.5.
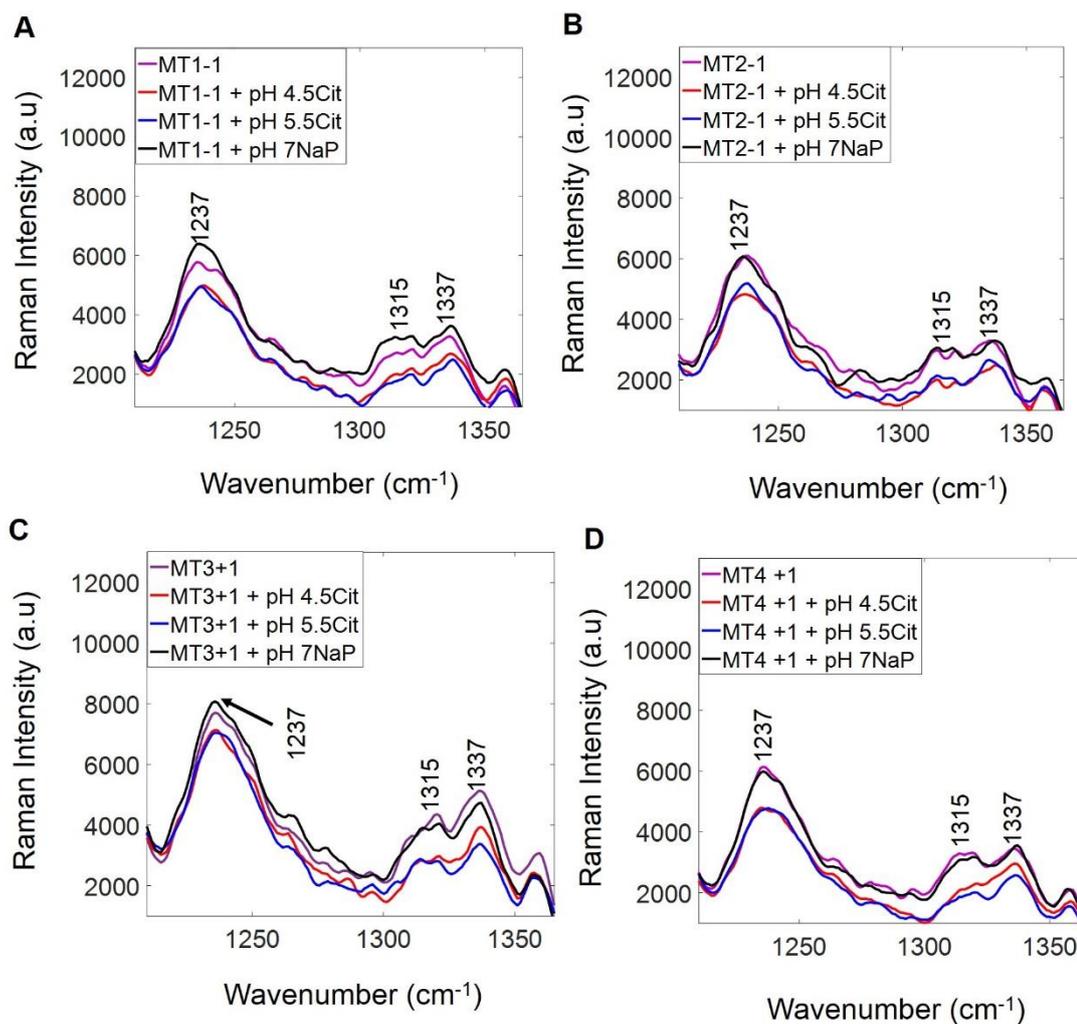
Figure 5.13 Averaged (n=3) Raman spectra of 10 mg/ml mutant Fab samples in sodium citrate (Cit), sodium phosphate (NaP) and water (A) MT1-1 (B) MT2-1 (C) MT3+1 (D) MT4+1. Preprocessing was carried out by baseline correction, smoothing, normalization, water subtraction and further baseline subtraction. Raman spectrum was collected in 30 mins using an excitation length of 785 nm. Protein assignments are listed in Table 1.1.

Only MT1-1 showed an increased β-sheet at pH 7, possibly linked to an increased tendency to aggregate compared to the other mutants (Ahmad, 2011). MT3+1 is distinguished by having minimal structural changes in acetate and phosphate buffers. This observation may be linked to the solvent accessibility and location of its mutation compared to other mutants. The MT3+1 mutation is situated in a protein region with low solvent exposure on the variable light chain. All other

mutants have their mutations on the heavy chain; MT1-1 and MT4+1 on the variable heavy chain and MT2-1 on the constant heavy chain (Ahmad, 2011).

For all mutants in citrate buffer (pH 4.5 and 5.5), a loss of peak intensity can be observed at ~1237, 1315 and 1337 cm^{-1} suggesting that they all exhibit similar structural changes to WTFab in citrate buffer at pH 4.5 and 5.5 (Figure 5.13 and Figure 5.8B). This observation may be linked to differences in acetate and citrate ions, namely their polyprotic vs monoprotic nature (Krukowski et al., 2017), as well as their preferential accumulation on the protein surface compared to acetate ions (Barnett et al., 2015).

The results demonstrate that the alteration of surface charges whether by reduction or addition of positive charges had variable effects on the pH-dependent structural changes previously observed with WTFab in sodium acetate and phosphate buffers. Of all the mutants, MT4+1 shows the most similar spectral changes to WTFab in both sodium acetate and citrate buffers; possibly due to their similar ionization profiles mentioned earlier (section 5.3.5). Furthermore, WTFab and MT4+1 have been reported to have the closest predicted and highest solvent accessible area (SAS) values. The predicted SAS for WTFab and MT4+1 are 21,275.1 and 21,359 respectively while MT1-1, MT2-1 and MT3+1 have predicted SAS values of 21,208.3, 21,241 and 21,260.4 respectively (Ahmad, 2011). Consequently, WTFab and MT4+1 are likely to engage in more solvent interaction than other mutants. Compared to WTFab, MT3+1 and MT1-1 demonstrate the most improved stability in acetate buffer (pH 4.5 and 5.5) followed by MT2-1. MT1-1 however, shows the least improved stability in sodium phosphate buffer compared to all other mutants. An important issue raised from this study is the

need to distinguish whether the observed structural changes originate from buffer or from pH changes. A combination of buffer and pH related structural changes have been identified. Modifying pH using acid and base as carried out with the model proteins in chapters 3 and 4 can be a convenient method to remove the buffer spectral signature; allowing only pH related spectral changes to be observed.

Owing to the similar profile of the pH-induced spectral changes in the mutant Raman spectra with WTFab especially in citrate buffer, 2DCA was not included in this section as it did not provide any additional information (Appendix B).

5.3.7 Freeze-dried Samples.

Freeze-drying is a widely accepted and frequently utilized method to retain the stability of protein therapeutics, providing a longer shelf life than its liquid counterpart. However, the impact of freeze-drying, as well as the optimum conditions required, must be assessed during drug development to ensure that it is not detrimental to the protein's stability and that the reconstituted product retains the stable structure of the protein (Franks and Auffret, 2007, Emami et al., 2018). Accordingly, freeze-dried samples of WTFab and its mutants were examined. Figure 5.14 shows the spectra of the freeze-dried WTFab and its mutant samples. The main spectral variations can be observed at ~760 and $\sim1315\text{-}1337$ cm^{-1}. The peak at ~760 cm^{-1} is assigned to Trp, and its increased intensities in the spectra of WTFab and MT3+1 compared to the other mutants suggest some of their Trp residues are in a more hydrophobic environment. This finding does not build on any previous observation from the solution spectra. No information about Trp residues in any of the samples was obtained. It is interesting to note that in

contrast to the liquid Fab sample spectra (Figure 5.11A), the peaks at ~1315-1337 cm^{-1} generally have a higher intensity than the 1237 cm^{-1} peaks. This observation is possibly due to the different sample states. At ~1337 cm^{-1}, MT3+1 has the highest intensity followed by MT2-1 and MT4+1 then MT1-1 and WTFab. The association of this peak with changes in the α-helix geometry has been discussed previously. Consequently, these spectral variations may therefore be monitoring such geometrical changes induced by the freeze-drying process or differences in the sample due to different charge interactions.
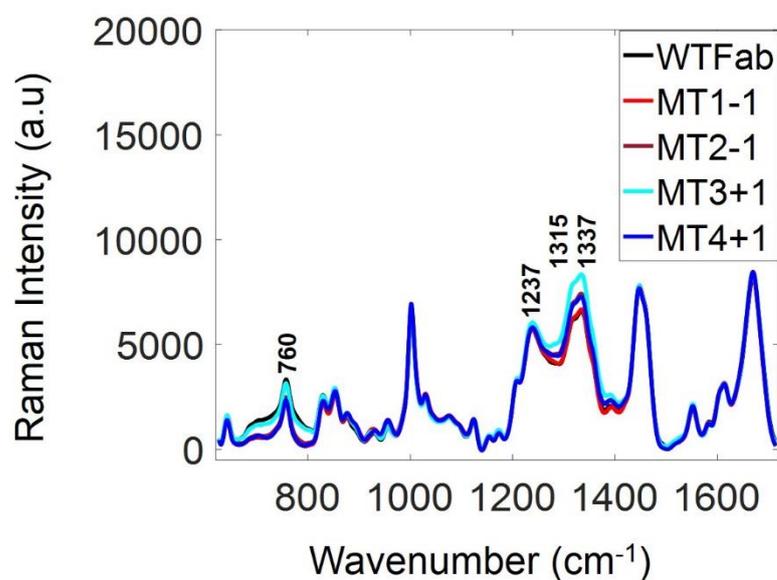
Figure 5.14 Averaged (n=8) Raman spectra of 10 mg/ml freeze-dried WTFab and four surface charged mutant Fab samples. 1mg samples were examined in solid-state. Preprocessing was carried out by baseline correction, normalization and smoothing. Raman spectrum was collected in 120s using an excitation wavelength of 785 nm.

5.3.7.1 pH-induced Spectral Variations

WTFab and the mutants were examined as freeze-dried forms in the different buffers: acetate (pH 4.5 and 5.5); sodium citrate (pH 4.5 and 5.5) and sodium

phosphate (pH 7). However, only the MT2-1 and MT3+1 samples are discussed because of the significant intensity variations in the other sample data sets arising from a change of the laser.
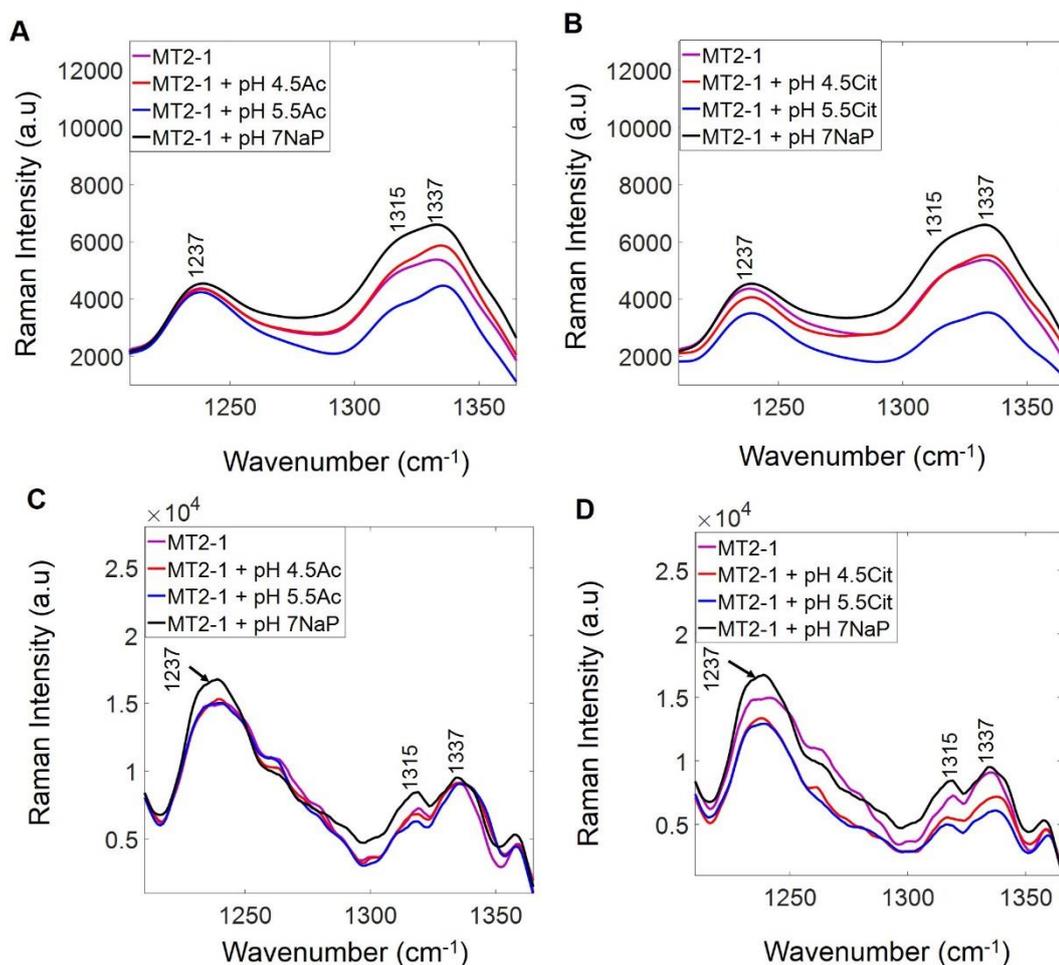
Figure 5.15 Averaged Raman spectra (n=8) of 10 mg/ml freeze-dried Fab samples and averaged Raman spectra (n=3) of reconstituted freeze-dried samples (A) freeze-dried MT2-1 in water, Ac and NaP (B) freeze-dried MT2-1 in water, Cit and NaP (C) reconstituted freeze-dried MT2-1 in water, Ac and NaP (D) reconstituted freeze-dried MT2-1 in water, Cit and NaP. Preprocessing of freeze-dried samples was carried out by baseline correction, normalization and smoothing. Raman spectrum was collected in 120s using an excitation wavelength of 785 nm. Preprocessing of reconstituted samples was as stated for Figure 5.13.

All the spectra for the buffered freeze-dried samples of MT2-1 and MT3+1 were collected using the new laser. However, the spectral data set of the buffered freeze-

dried samples for WTFab, MT1-1 and MT4+1 were acquired using both old and new lasers. The spectra from the new laser had a much higher intensity than the ones collected with the old laser despite normalisation of both data sets. For this reason, it was not possible to compare these mutants (MT2-1 and MT3+1) with the freeze-dried WTFab. However, MT2-1 and MT3+1 have been compared in different buffers as freeze-dried and reconstituted forms.

Figure 5.15 displays the freeze-dried and reconstituted sample spectra of MT2-1. With the freeze-dried sample spectra, the peak intensity at ~1315-1337 cm^{-1} varies significantly in intensity with the sample in phosphate buffer having the highest intensity compared to in water and the other buffers (Figure 5.15A and 5.15B). Additionally, this peak intensity is increased at pH 4.5 compared to pH 5.5 in both acetate and citrate buffers, suggesting a similar variation in α-helix orientation between the respective samples. Furthermore, the minimal intensity variations between the MT2-1 spectra at pH 4.5 (acetate and citrate buffer) and in water suggests that the α-helix symmetry is unchanged for these sample formulations but is different at pH 5.5 (acetate and citrate buffer) where the ~1337 cm^{-1} peak has a lower intensity. These observations suggest that in the freeze-dried samples of MT2-1, the α-helix symmetry is different in phosphate buffer (pH 7) as well as in acetate and citrate buffer (pH 5.5).

The freeze-dried MT2-1 sample spectra contrasts with the MT2-1 solution spectra (Figure 5.12B), where compared to the control, there was no change in the ~1315-1337 cm^{-1} peak intensity in sodium phosphate buffer. With respect to changes in the β-sheet peak (~1237 cm^{-1}), a loss of intensity, indicating a loss of β-sheet can be observed mainly at pH 5.5 in citrate buffer (Figure 5.15B). This spectral change

also differs from observations in the MT2-1 solution spectra (Figure 5.12B and 5.13B) where a loss of β-sheet was indicated at pH 4.5 (acetate and citrate buffer) in addition to at pH 5.5 (citrate buffer). Upon reconstitution, the main distinction in the spectra (Figure 5.15C) compared to the freeze-dried sample spectra (Figure 5.15A) is an increase in the peak intensity at 1237 cm^{-1} (MT2-1 in phosphate buffer) compared to MT2-1 in water and other buffers. This observation is possibly due to a degree of aggregation occurring in sodium phosphate buffer upon freeze-drying. The reconstituted spectra (Figure 5.15C) also differs from the solution spectra (Figure 5.12B) as there is no spectral evidence for a loss of β sheet observed at ~pH 4.5 (acetate buffer). The lack of consistency highlighted by the absence of a spectral change suggesting a loss of β-sheet in the freeze-dried and reconstituted MT2-1 spectra but not in solution spectra suggests a difference between the samples either initially or as a result of freeze-drying. However, the loss of peak intensity at ~1237 cm^{-1} in citrate buffer (pH 4.5 and 5.5), is consistent with the results from the solution study (Figure 5.13B), suggesting that the freeze-drying process did not affect the loss of β-sheet in MT2-1 at pH 4.5 and 5.5 (citrate buffer).

In accounting for the reasons for differences between the freeze-dried, reconstituted solution and solution sample spectra; the extent to which factors including the sample state, pH and freeze-drying affect the observed variations are difficult to delineate completely. This issue raises the question of how to objectively compare the peaks in the solution and freeze-dried spectra supposedly at the same pH. As observed from the reconstituted sample spectra, the difference in the shape and height of the peaks at ~1315-1337 cm^{-1} between the solution and

freeze-dried sample spectra is likely due to differences in sample state because in general, the reconstituted sample spectra (Figures 5.15C and 5.15D) regain the same peak shape as the solutions.
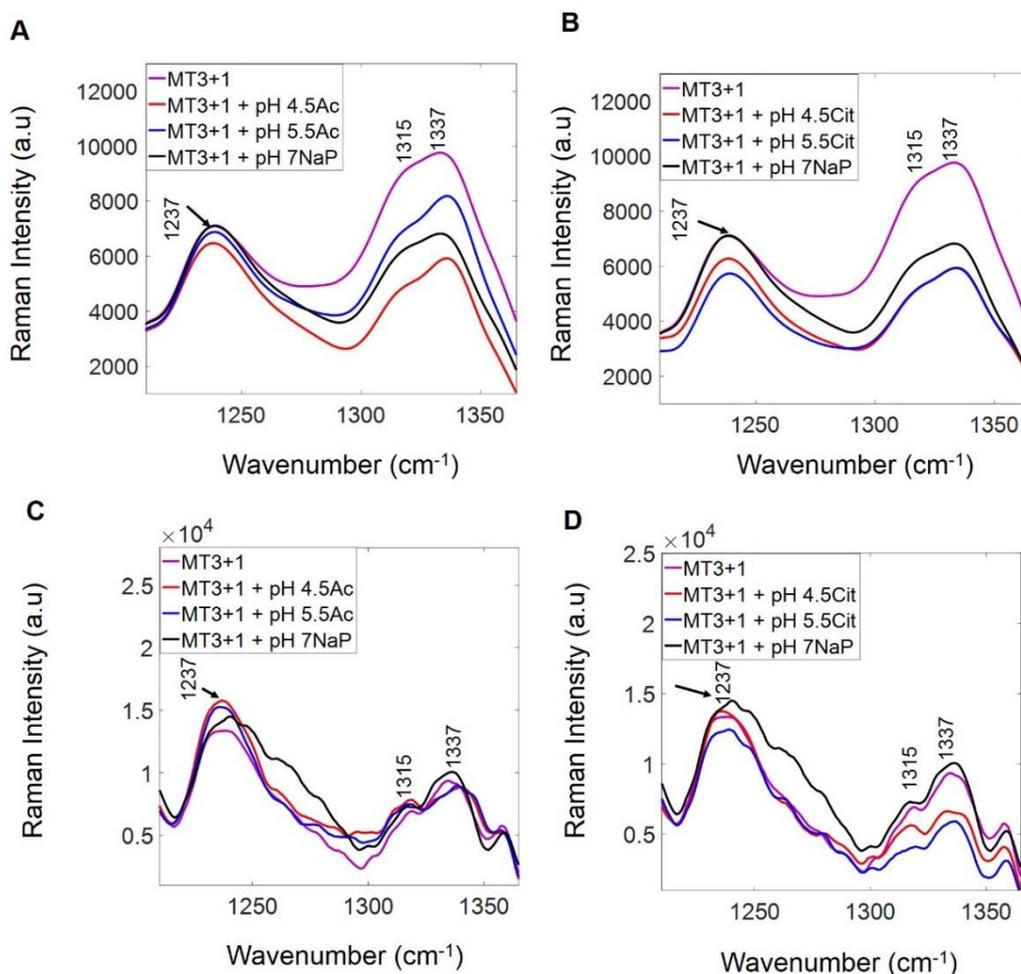
Figure 5.16 Averaged Raman spectra (n=8) of 10 mg/ml freeze-dried Fab samples and averaged Raman spectra (n=3) of reconstituted 10 mg/ml freeze-dried samples (A) freeze-dried MT3+1 in water, Ac and NaP (B) freeze-dried MT3+1 in water, Cit and NaP (C) reconstituted freeze-dried MT3+1 in water, Ac and NaP (D) reconstituted freeze-dried MT3+1 in water, Cit and NaP. Preprocessing of freeze-dried samples was carried out by baseline correction, normalization and smoothing. Raman spectrum was collected in 120s using an excitation wavelength of 785 nm. Preprocessing of reconstituted samples was as stated for Figure 5.13.

A comparison of the WTFab and MT2-1 in these sample states will be useful to gain more insight into the observed spectral differences. Overall results suggest that

MT2-1 is least stable in phosphate buffer following freeze-drying compared to in acetate or citrate buffers.

Figure 5.16 shows the freeze-dried and reconstituted sample spectra of MT3+1. Again, the peaks at ~1315-1337 cm^{-1} have the most significant spectral variations. with MT3+1 in water (pH 7) having the highest intensity.

The results suggest that the α-helical geometry of freeze-dried MT3+1 in the buffers differs from that in water (Figure 5.16A and 5.16B). At pH 4.5 and 5.5 (citrate buffer), there is a loss of peak intensity at ~1237 cm^{-1} indicating a loss of β-sheet (Figure 5.16B) consistent with the observations in the solution spectra (Figure 5.13C). This spectral change appears weaker in the case of acetate buffer (Figure 5.16A and Figure 5.12C). Upon reconstitution, a broadening of the peak at ~1237 cm^{-1} can be observed in the MT3+1 spectrum at ~pH 7, sodium phosphate buffer (Figure 5.16C). The broadening may signal β-sheet aggregation (Li and Li, 2009).

In MT2-1, with a reduced positive surface charge, no such broadening upon reconstitution was observed in phosphate buffer. However, a slight increase in peak intensity, also suggestive of aggregation was observed (Figure 5.15C). Zhang (2016) showed that the ~pH 7 phosphate buffer was not optimal for the freeze-drying of WTFab or its mutants, due to considerable monomer loss. Consequently, these spectral changes at pH 7, which can only be seen in the reconstituted and not the liquid sample spectra, suggest that the freeze-drying of MT2-1 and MT3+1 with sodium phosphate buffer induces additional structural changes compared to in solution. Contrary to spectral variations in the MT3+1 solution spectrum (Figure 5.12C), an increased peak intensity at 1237 cm^{-1} suggestive of aggregation was

also observed at pH 4.5 and 5.5 (acetate buffer) compared to in water (Figure 5.16C).

Acetate buffer has also been reported as favouring monomer loss upon freeze-drying compared to citrate buffer at the same pH (Zhang, 2016). It should also be noted that previously, MT3+1 demonstrated improved stability at pH 4.5 and 5.5 in acetate solutions compared to citrate solutions (Figure 5.12C and 5.13C). Therefore, this aggregation tendency has likely been introduced through freeze-drying in acetate buffer. The results show that MT2-1 has a similar stability profile in solution and upon reconstitution in citrate buffer (pH 4.5 and 5.5) and acetate buffer (pH 5.5). MT3+1 however, shows a different behaviour between solution and reconstituted states in both buffers suggesting decreased stability in acetate buffer (pH 4.5 and 5.5) and phosphate buffer.

5.3.7.2 2D-Correlation Analysis

In addition to peak shifts discussed in sections 3.3.3.3 and 4.3.2.3, spectral features such as broadening can be identified using generalised 2D-correlation plots (Noda and Ozaki, 2005). Therefore, synchronous and asynchronous plots (Figure 5.17) were generated to examine the broad feature in the reconstituted sample spectra of MT3+1 at pH 7. There is no autopeak at 1237 cm^{-1}, but at ~1268 cm^{-1} in the synchronous plot and the contour pattern does not a suggest broadening of the ~1237 cm^{-1} peak is occurring. A four-way symmetric pattern (including a dominant central peak as well as smaller auto peaks and negative cross peaks) in the synchronous plot and a *cross* pattern in the asynchronous plot is associated with broadening of a single peak (Noda and Ozaki, 2005).
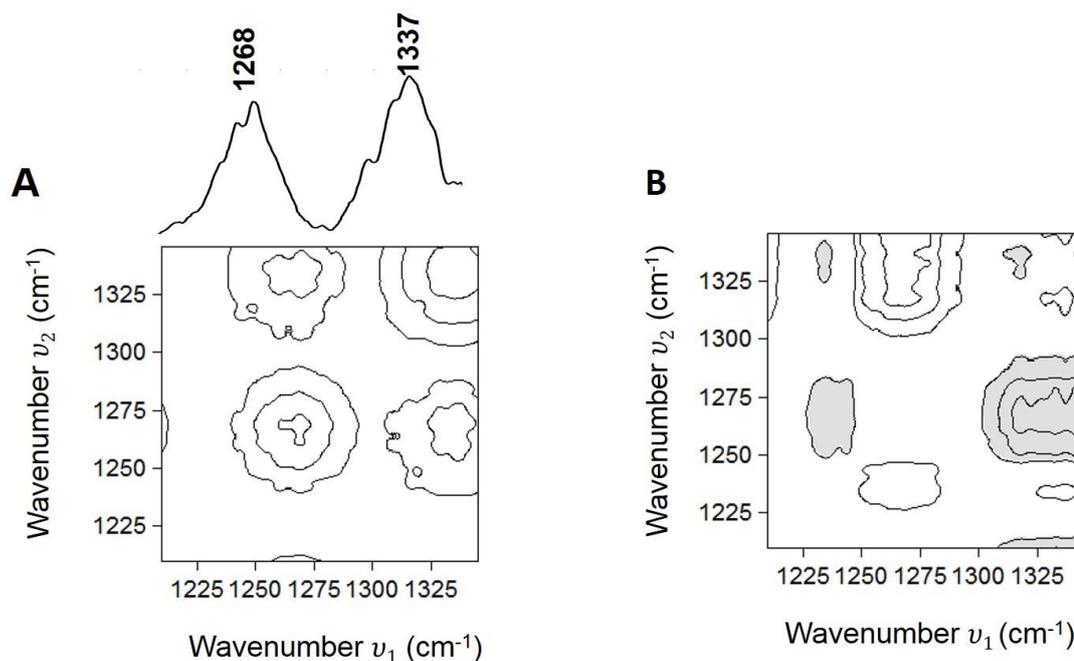
Figure 5.17 2D-correlation analysis of the pH-dependent Raman spectra of reconstituted MT3+1 from pH 4.5-7. pH 4.5 and 5.5 (sodium citrate buffer) and pH 7 (sodium phosphate buffer). (A) Synchronous plot (B) Asynchronous plot. Shaded contours indicate negative cross peaks and unshaded contours indicate positive cross peaks. Positive cross peaks indicate two wavenumbers with perturbation induced intensity changes in the same direction (increasing or decreasing together). Negative cross peaks indicate two wavenumbers with perturbation induced intensity changes in opposite directions (one increasing and the other decreasing). Maximum number of contours=4.

The asynchronous plot has a negative cross peak (1237,1268), suggesting the presence of an overlapped peak at ~1268 cm^{-1}. A Raman peak at ~1268 cm^{-1} is associated with α-helix (Yu et al., 1972). Consequently, the broad feature in the pH 7 (phosphate buffer) MT3+1 spectra, which has an increased intensity compared to the other spectra is likely arising from α-helix unfolding and not aggregation. This observation is possibly linked to the different α-helix orientation of MT3+1 as suggested from a comparison of the freeze-dried Raman spectra of WTFab and all the mutants (Figure 5.14).

5.3.8 Comparison of WTFab and Surface Charge Mutants

Table 5.2 shows a summary and comparison of the pH-dependent behaviour of WTFab and the four mutants in the different sample states. The black triangles indicate a variation in the peak intensity at ~1237 and or 1315-1337 cm^{-1} compared to the control and the black dash implies that no change was observed at any of these peaks with respect to the control. A combination of both symbols separated by a slant indicates the response at pH 4.5 and 5.5, respectively. The control is the specific sample in water in the respective sample states.

A stability ranking for the samples in buffer solutions has been proposed based on the control as a reference. The comparison assumes the control (pH 7) is the reference for good stability based on the observations from this study and previous work that WTFab is stable at pH 7. This ranking does not apply to the freeze-dried or reconstituted freeze-dried samples because the WTFab data has not been included in the table for reasons mentioned earlier. A ranking of 4 means a significant variation in the peak intensities at ~1237 and 1315/1315-1337 cm^{-1} compared to the control indicating less stability than the control while a ranking of 1 means no change in the peak intensities with respect to the control implying comparable stability with the control. A ranking of 2 represents a situation where there is a variation in these peak intensities at either ~ pH 4.5 or 5.5.

In solutions, MT1-1 and MT3+1 rank the most stable in acetate buffer followed by MT2-1 whereas WTFab and MT4+1 are the least stable in acetate buffer. MT1-1 is the least stable in phosphate buffer, and all the samples have similar stability in citrate buffer.

Table 5.2 Comparison of WTFab and surface charge mutants. The black triangles and black dash indicate a variation and no change with respect to the control, respectively. Ac-(sodium acetate buffer, pH 4.5 and 5.5); Cit-(sodium citrate buffer, pH 4.5 and 5.5); NaP-(sodium phosphate buffer, pH 7). A ranking of 4 means the sample is less stable compared to control. A ranking of 1 means as stable as control. A ranking of 2 means less stable or as stable as control at pH 4.5 or 5.5.

Sample	Solution			Freeze-dried			Reconstituted			Ranking (solutions)		
Buffer	Cit	Ac	NaP	Cit	Ac	NaP	Cit	Ac	NaP	Cit	Ac	NaP
WTFab	▼	▼	—	*	*	*	*	*	*	4	4	1
MT1-1	▼	—	▼	*	*	*	*	*	*	4	1	4
MT2-1	▼	▼/—	—	—/▼	—/▼	▼	▼	—	▼	4	2	1
MT3+1	▼	—	—	▼	▼	▼	—/▼	▼	▼	4	1	1
MT4+1	▼	▼	—	*	*	*	*	*	*	4	4	1

*data not included due to significant intensity variations.

This ranking indicates that in the solution samples, the pH-dependent behaviour of mutants compared to WTFab is not divided along the lines of the nature of the mutation, i.e. reduced positive charge or increased positive charge. The different rankings for the mutant samples (MT1-1, MT2-1 and MT3+1) compared to the similar rankings for WTFab and MT4+1 in acetate and citrate buffer (pH 4.5 and 5.5) suggest that the variations in their pH-dependent behaviour are influenced by the type of buffer as well by the ionization profile of the carboxyl groups.

5.4 Temperature Stability

WTFab and only one mutant, MT1-1 were subjected to temperature studies due to the limited time available. 10 mg/ml of WTFab and MT1-1 solutions in water (pH 7) were heated as described in section 2.4.2. Their autocorrelation plots and Raman spectra are displayed in Figure 5.18 and 5.19 respectively. As indicated from the spectra, there are some slight spectral intensity variations across the WTFab and MT1-1 spectra (Figure 5.18B and 5.19B). However, the indistinct nature of the spectral variations, as shown by the autocorrelation plots, makes it difficult to interpret these changes in terms of protein structure. At pH values above 5, WTFab has been shown to have apparent conformational stability temperatures ($T_{m\,app}$) above 75 °C (Chakroun et al., 2016). No similar studies have been reported for MT1-1.

Figure 5.20 displays the synchronous 2DPCMW plots for both samples. The pattern of contours in the synchronous 2DPCMW plot is random supporting the observation that no significant temperature-dependent change has occurred and that both samples are relatively stable up to 65 °C.
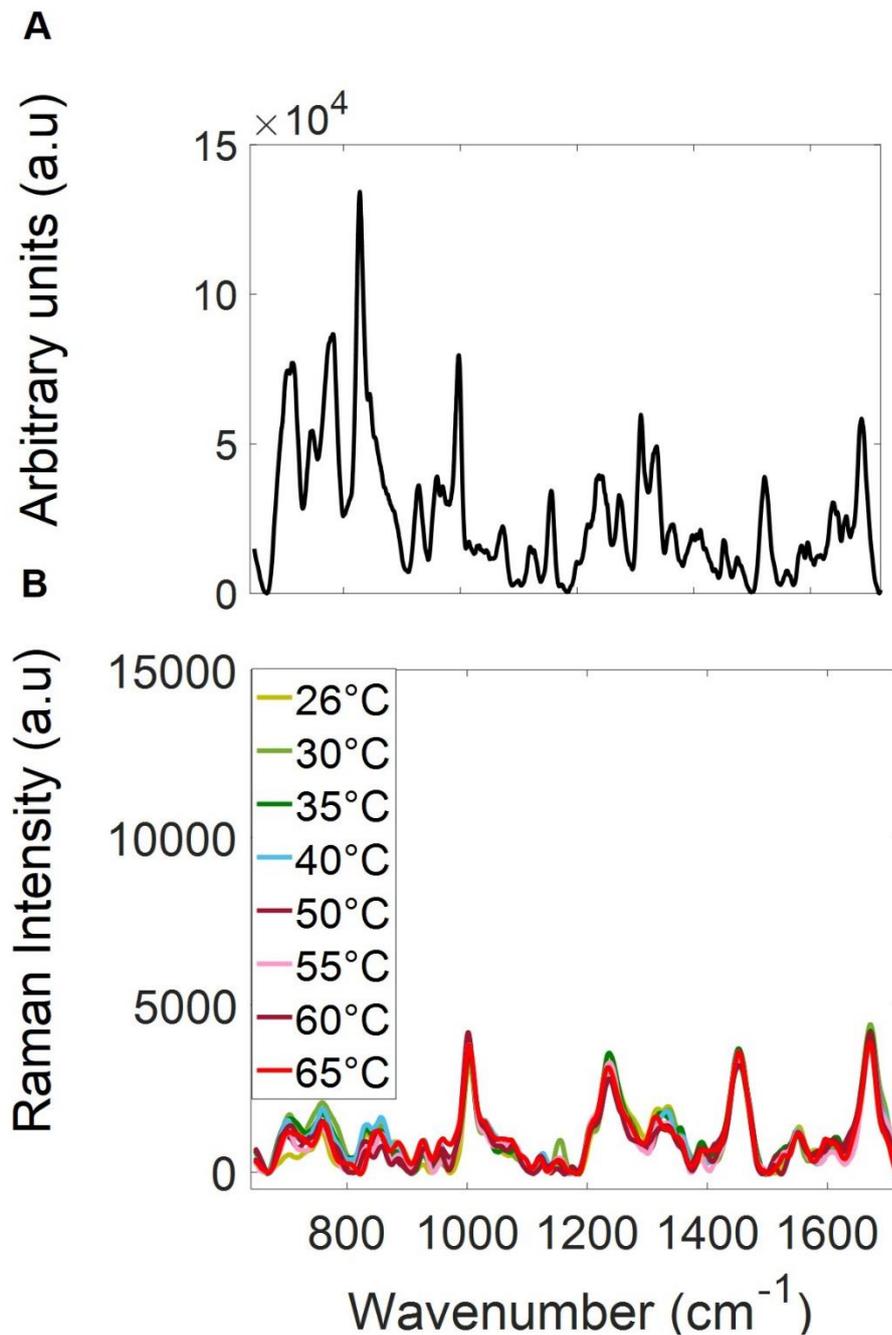
Figure 5.18 Temperature-dependent Raman spectral variations of 10 mg/ml WTFab between 28-65 °C (A) Autocorrelation (B) Raman spectra. Preprocessing was carried out by baseline correction followed by subtraction of the quartz spectrum, smoothing, normalisation to the peak at ~1447 cm⁻¹, water subtraction and further baseline correction. Each Raman spectrum was collected in 30 mins using an excitation length of 785 nm.
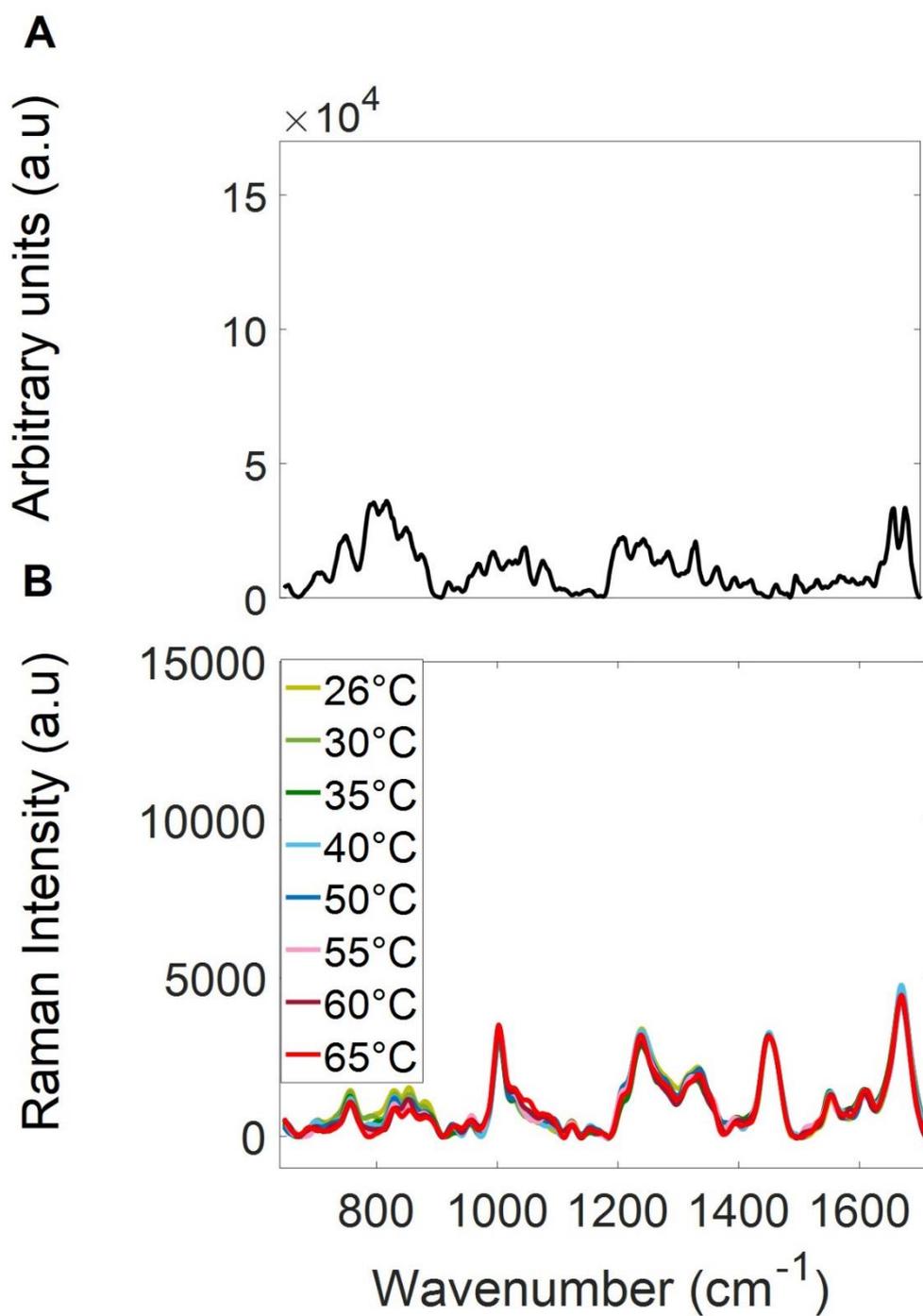
Figure 5.19 Temperature-dependent Raman spectral variations of 10 mg/ml MT1-1 between 28-65 °C (A) Autocorrelation (B) Raman spectra. Preprocessing was carried out by baseline correction followed by subtraction of the quartz spectrum, smoothing, normalisation to the peak at ~1447 cm^{-1}, water subtraction and further baseline correction. Each Raman spectrum was collected in 30 mins using an excitation length of 785 nm.
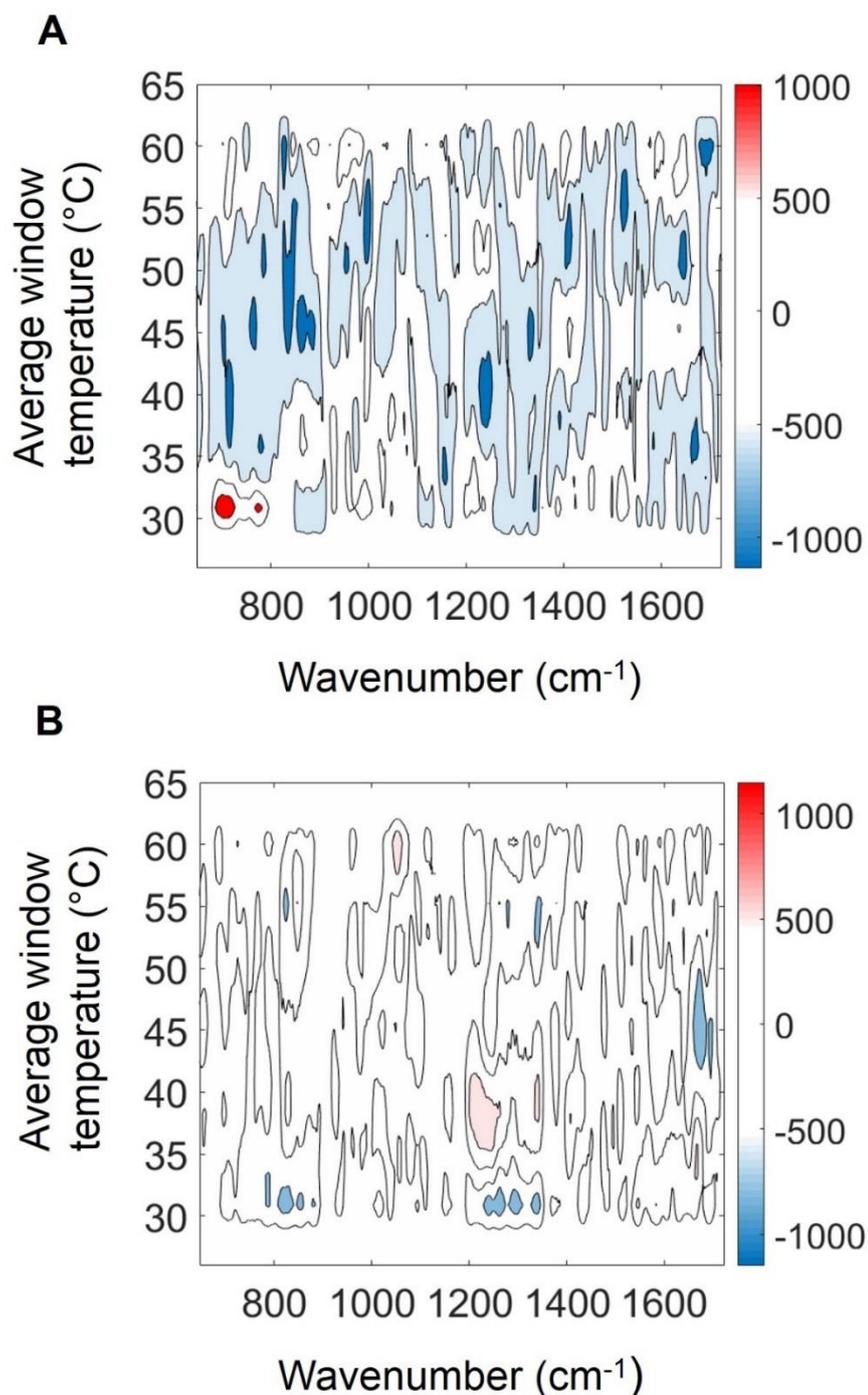
Figure 5.20 Synchronous 2DPCMW plots generated from Temperature-induced Raman spectra (A) WTFab (B) MT1-1. Positive contours are shaded in red while negative contours are shaded in blue. Positive contours indicate increasing peak intensities in the direction of the perturbation while negative contours indicate decreasing peak intensities in relation to the perturbation. A darker shade and an increased number of contours indicate a relatively greater change in intensity. The scale on the colour bar is arbitrary. A moving window size of 5 and a maximum number of 4 contours was applied.

5.5 Conclusion

This study investigated and compared the conformational stability profiles of five therapeutically relevant models, WTFab and four surface charge mutants (MT1-1, MT2-1, MT3+1 and MT4+1). An initial comparison of their difference Raman spectra in aqueous solutions suggested variations in the solvent exposure of Phe residues that could be linked to the type of mutation and aggregation tendency. The use of different buffers to control the protein sample pH was accompanied by additional complexity in the Raman protein spectra due to the presence of strong peaks arising from each buffer as well as pH-induced changes of some buffer peaks. Despite the use of buffers, water spectral subtraction was used as an alternative to buffer spectral subtraction due to the partial subtraction of some buffer peaks. Consequently, the application of 2DCA analysis was limited by the few wholly protein assigned spectral intensity variations. Due consideration should therefore be given to how the use of multiple buffers or a single buffer at different pH will affect the final spectral interpretation or subsequent 2DCA analysis.

The WTFab results suggest a pH-dependent loss of β-sheet and variation in α-helical symmetry at pH 4.5 and 5.5 compared to at pH 7 (sodium phosphate buffer and water). The spectral intensity variations in peaks assigned to α-helix symmetry were also observed to be buffer dependent at pH 4.5 and 5.5.

The four mutant samples exhibited similar behaviour to the WTFab at ~pH 4.5 and 5.5 in citrate buffer. In acetate buffer, only MT4+1 compared closely with WTFab revealing the same pH-dependent changes possibly due to having a similar profile

of ionized carboxyl groups. MT3+1 was the most stable sample overall in acetate buffer solution and MT1-1 was the least stable sample in phosphate buffer compared to WTFab and the other mutants.

Since only MT2-1 and MT3+1 were analysed in the case of the freeze-dried and reconstituted samples, the proposal of the most stable candidate is limited to solution samples. Notwithstanding, the results indicate that the choice sample needs to be examined in the context of the solvent not just in solutions but for any subsequent stage such as freeze-drying. Citrate buffer appears to be less aggregation-prone compared to acetate and phosphate buffer.

With respect to thermal stability, WTFab and MT1-1 showed no significant difference in their temperature-dependent behaviour up to 65 °C; nonetheless, the results suggest that MT1-1 was less perturbed than WTFab under the examined conditions.

6 Antibody Fragment Model using Ultraviolet Raman Resonance (UVRR) Spectroscopy and 2DCA

6.1 Introduction

As demonstrated by the initial investigation into the pH-dependent structural changes of WTFab and its mutants (chapter 5), a significant challenge to overcome is the overlapping of buffer peaks with protein peaks in the Fab Raman spectrum. The interference of the buffer peaks limits the analysis of spectral variations arising from protein structure. UVRR spectroscopy was utilised to study WTFab, and four of its surface charge mutants because it has been shown to have weaker vibrational modes for aqueous buffers compared to proteins due to differences in their maximum absorption wavelengths (Knee and Mukerji, 2009, López-Peña et al., 2015).

UVRR is a form of Raman spectroscopy which uses laser excitation wavelengths in the UV region typically between (~180-415 nm) to obtain resonantly enhanced Raman signals up to 10^3-10^6 due to the occurrence of a molecular electronic absorption band near the applied laser wavelength (Asher, 1993, Oladepo et al., 2012, Chadwick et al., 2015). As shown in Figure 6.1, resonant Raman scattering arises within an excited electronic transition state, whereas normal Raman scattering does not involve excited electronic transition states.
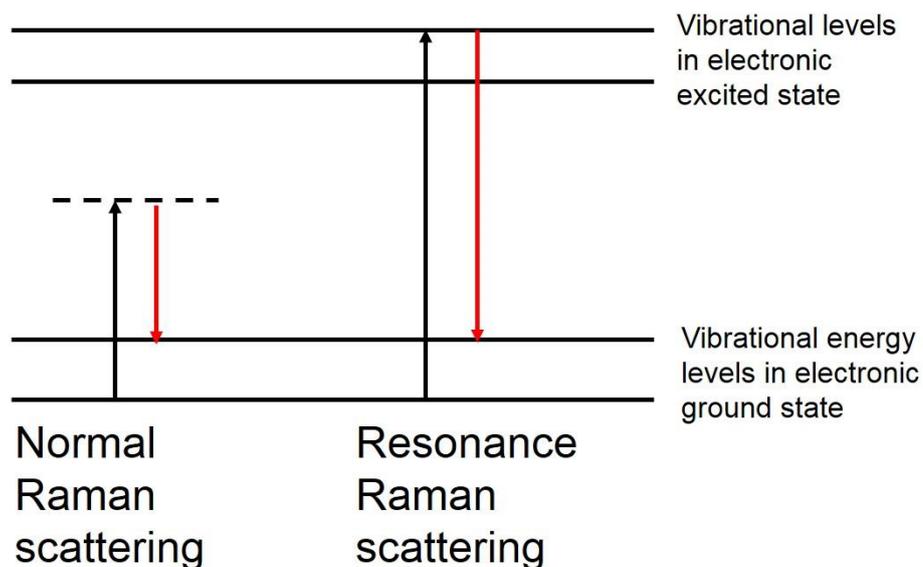
Figure 6.1 Schematic showing the difference between normal resonance and resonance Raman scattering. Image adapted from Asher (1993) and Smith and Dent (2005).

Due to this wavelength dependence, vibrations arising from the affected electronic absorption bands will dominate the spectrum (Johnson et al., 1984, Wen and Thomas Jr., 1998, Takeuchi, 2011). In general, peptide bond vibrations dominate the spectrum from ~193-206 nm while aromatic amino acids vibrations predominate from ~220-250 nm (Asher et al., 1986, Asher, 2006, López-Peña et al., 2015). However, the level of peak intensity enhancement for a specific amino acid often varies significantly depending on the specific wavelength or surrounding environment in a protein (Ludwig and Asher, 1988, Wen and Thomas Jr., 1998, Chi and Asher, 1998). Apart from resonance enhancement of spectral intensities, UVRR spectroscopy also exhibits minimal fluorescence interference, particularly in the deep UV region, ~180-260 nm (Asher and Johnson, 1984, Smith and Dent, 2005). There is, however, an increased risk of sample photodecomposition due to the high energy of UV lasers. The detrimental impact of the laser on the sample can be minimised through sample rotation, replenishing

of sample volume and adjustment of spectral parameters (Smith and Dent, 2005, Ashton et al., 2013b, Jarvis and Goodacre, 2004).

The results from the previous studies in chapter 5 suggested that all the mutant samples and WTFab share a similar stability profile in citrate buffer. They all showed a slight loss of β-sheet at pH 4.5 and 5.5. However, the autocorrelation and synchronous 2DPCMW plots were dominated by the buffer peaks. This study was aimed at investigating the presence of any additional stability differences between the mutant samples and WTFab that may have been masked by the buffer interference, using a combination of UVRR spectroscopy and 2DCA.

6.2 Experimental

6.2.1 Spectra Collection

For the UVRR spectra collection, the Renishaw inVia confocal Raman microscope was adapted by exchanging the three visible holographic filters and grating (1200 l/mm) with UV compatible filters and grating (3600 l/mm). The instrument was also coupled to an excitation wavelength of 244 nm (Lexer SHG laser). Calibration was performed using a diamond spectrum centred at ~1332 cm^{-1} (acquisition time of 1s). All UVRR spectra were acquired using a x40 UV objective lens. The spectra of the Fab solutions (~50 μL) were collected using a polystyrene 96-well plastic lid. The 96-well plastic lid has been previously used despite the strong UVRR signal of polystyrene since with appropriate positioning of the objective focus; it is possible to rule out any interfering polystyrene signals in the sample spectrum (Ashton et al., 2013b). For spectra collection, the correct objective focal position was determined by comparing a spectrum of the 96-well plastic lid with a

spectrum of WTFab placed in the same plastic lid. The absence of the polystyrene peaks (~1004, 1196 and 1604 cm^{-1}) in the protein spectra confirms the correct focal position (Figure 6.2).
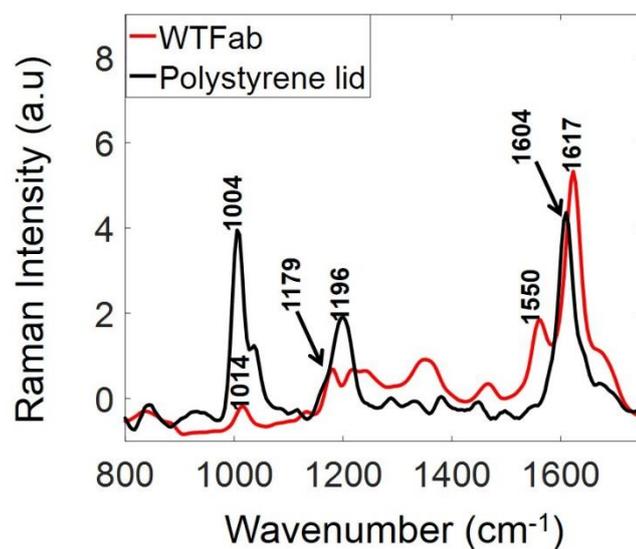
Figure 6.2 UVRR spectra collection comparing Raman spectra of 10 mg/ml WTFab in a polystyrene 96-well plate lid and polystyrene 96-well plate lid. Preprocessing carried out by normalisation followed by smoothing. UVRR spectra were collected in 2 min using an excitation wavelength of 244 nm.

The sample was rotated at 15 rpm on the well axis throughout the time of spectral acquisition to prevent photodecomposition of the sample (Jarvis and Goodacre, 2004, Ashton et al., 2013b). Photodecomposition was monitored by examining white light images (Figure 6.3) as well as by examining spectra after collection. When the samples are not rotated, there is a significant loss of spectral features over a range of laser power settings as well as photodegradation in the white light image (Figure 6.3A). However, with rotation, the Fab sample retains its spectral features, and no significant signs of photodecomposition were observed with the white light image (Figure 6.3B).
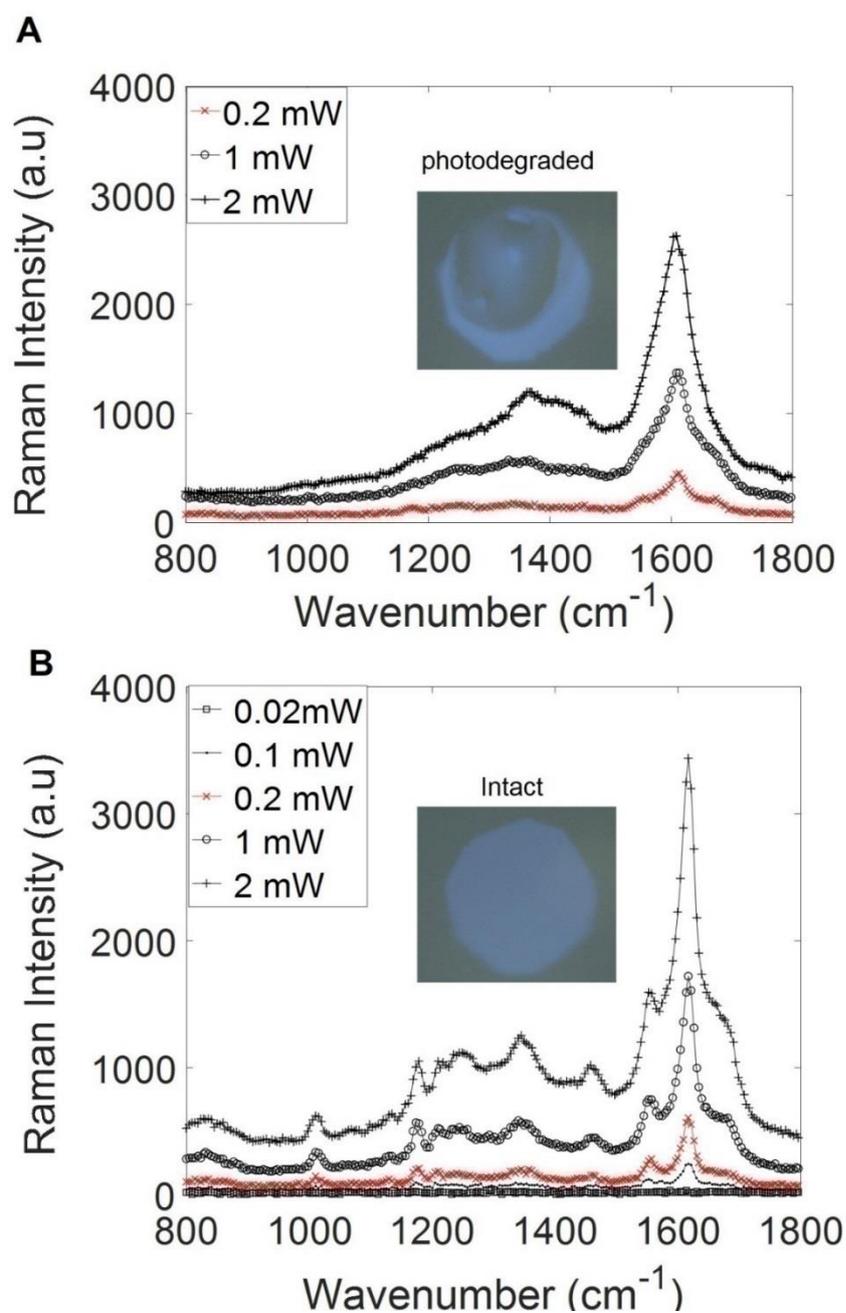
Figure 6.3 Laser power adjustment for raw UVRR spectra with insert of white light images monitoring photodecomposition. (A) static MT2-1 samples (B) rotated MT2-1 samples. UVRR spectra were collected in 2 min using an excitation wavelength of 244 nm.

Consequently, no significant signs of photodecomposition were observed when samples were rotated during spectral collection. For each spectrum, the exposure time was 10s with 12 accumulations (total exposure time of 2 minutes) and ~0.2

mW laser power at the sample. This low power was chosen despite the accompanying lower intensity spectrum because of the higher risk of photodecomposition with increasing power. A total of nine spectra were collected for each sample. In addition, three to four spectra of each buffer were also collected using the same instrument parameters.

6.2.2 Sample Preparation

12.5 mg/ml solution of Fab samples (WTFab, MT-1, MT2-1, MT3+1 and MT4+1) were supplied by CIMEMT along with 100 mM buffers of sodium acetate (pH 4-5.5), sodium citrate (pH 4-6.5) and sodium phosphate (pH 7-7.5). Samples were prepared, and pH was measured as described in section 5.2.1.

6.2.3 Data Preprocessing

Compared to the previously discussed Raman spectra (section 2.2), the UVRR spectra required fewer preprocessing steps. Baseline correction and water subtraction were not required for the UVRR spectra of the Fab solutions because of the reduced background and absence of the water peak at ~1640 cm^{-1} in the protein spectra. However, normalisation and smoothing procedures were still carried out to minimise the intensity variations and reduce noise in the spectra, respectively. UVRR spectra were normalised using the standard normal variate method (SNV), followed by smoothing using a fast-smooth algorithm (Haver, 2006).

The SNV method (Barnes et al., 1989) involves the calculation of the mean spectral intensity of the spectrum. The mean intensity is subtracted from the original intensity values and divided by the standard deviation. It does not require a

reference intensity like the peak normalisation because its operation on each spectrum is independent of the remaining data set (Rinnan et al., 2009). This approach is more suitable for the UVRR spectra than the peak normalisation because of the selective resonant enhancement of different peaks. Data was also interpolated (section 2.5.1) to obtain evenly spaced data of ~0.4 before 2DCA. The number of spectra used for averaging varied between 5-9 across Fab samples due to extremely weak signals in some spectra detected during preprocessing (Table 6.1).

Table 6.1 Number (n) for UVRR Fab spectra averaging. pH 4-6.5 (citrate buffer), pH 7-7.5 (Sodium phosphate buffer). The pH of all Fab samples in water was ~7.

Spectra	WTFab	MT1-1	MT2-1	MT3+1	MT4+1
Water	7	7	8	9	8
pH 4	7	9	7	8	8
pH 4.5	5	8	8	8	7
pH 5	5	8	9	9	9
pH 5.5	6	9	9	9	9
pH 6	7	9	9	8	9
pH 6.5	5	9	8	8	9
pH 7	6	9	7	9	9
pH 7.5	5	8	7	8	8

6.3 Results & Discussion

6.3.1 Buffer Influence

Figure 6.4 compares the raw UVRR spectra of WTFab in water and selected buffers.

In all cases, buffer peaks are not evident in the protein spectra (Figure 6.4B-D).
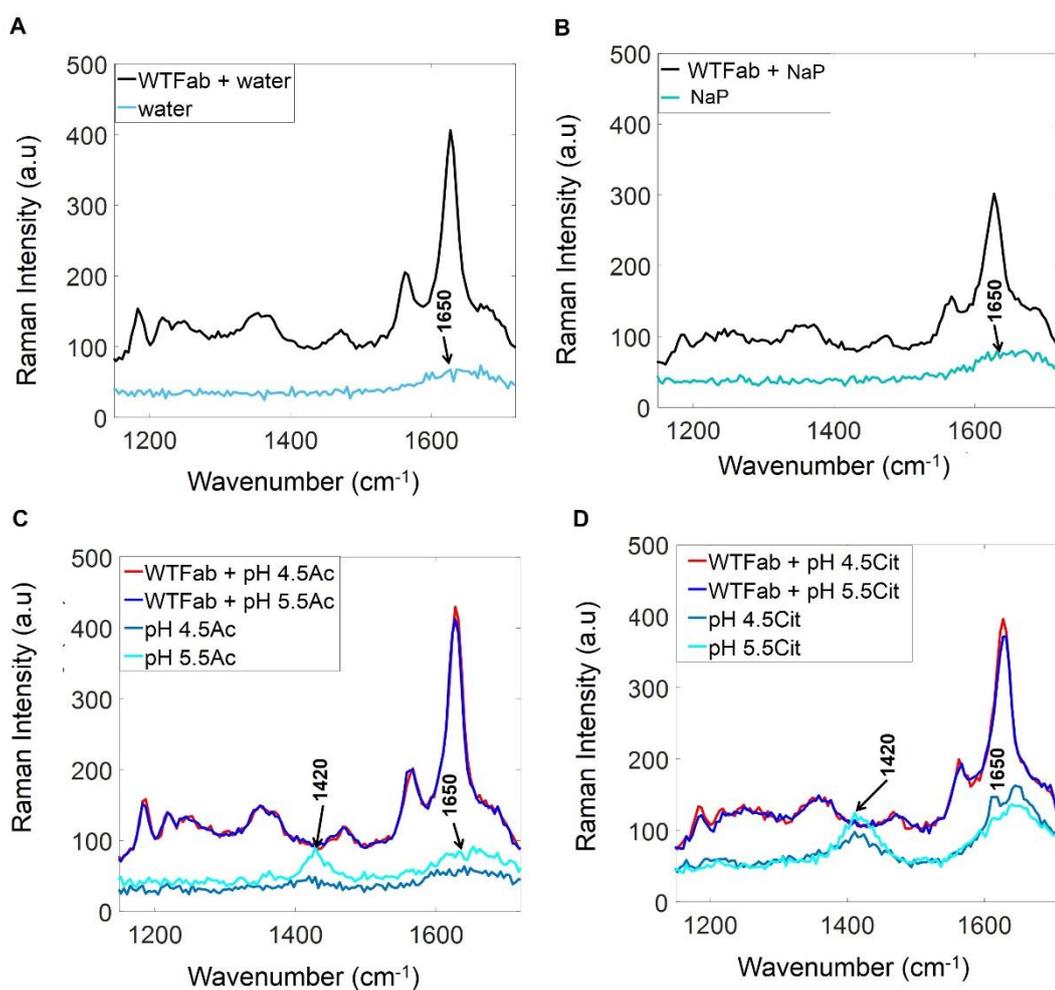
Figure 6.4 Averaged raw UVRR spectra of 10 mg/ml WTFab and buffer (A) WTFab in water and water only (B) WTFab in sodium phosphate buffer (NaP) and NaP at pH 7 (C) WTFab in sodium acetate buffer (Ac) and Ac at pH 4.5 & 5.5 (D) WTFab in sodium citrate buffer (Cit) and Cit at ~pH 4.5 & 5.5. n=3 for buffer; 8 and 9 for WTFab in pH 4.5 Ac and 5.5 Ac respectively. Other n values are listed in Table 6.1. UVRR spectra were collected in 2 min using an excitation wavelength of 244 nm. Preprocessing carried out by normalisation and smoothing.

Sodium citrate and sodium acetate buffer spectra have peaks at ~1420 cm^{-1} assigned to the ionization of carboxyl groups (Wang et al., 2005, Elbagerma et al., 2015). The peak at 1650 cm^{-1} in the water and buffer spectra is assigned to bending vibrations of water (Ludwig and Asher, 1988). Regardless of buffer and across the pH-range (4-7.5), the buffer spectra are of considerably lower intensity compared to the protein spectra. Moreover, there are no visible buffer peaks in the protein spectra, indicating minimal interference from the buffers in protein spectra.

6.3.2 pH-induced Spectral Variations (WTFab)
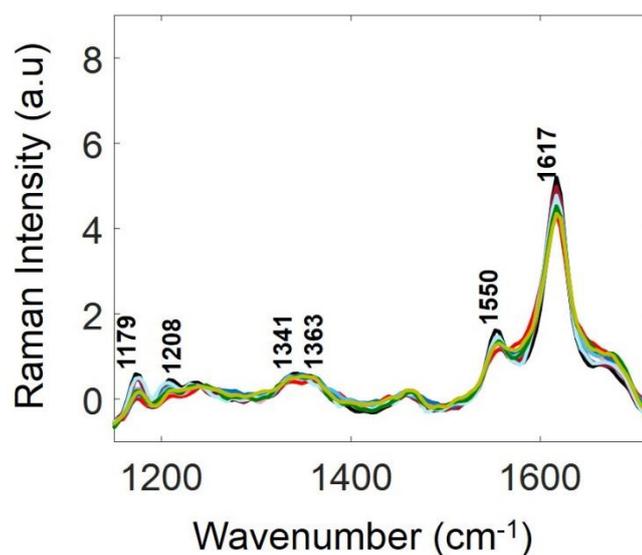
Figure 6.5. Averaged UVRR spectra of WTFab in water (~pH 7.1), Cit (pH 4-6.5) and NaP (pH 7-7.5). Preprocessing carried out by normalisation and smoothing. UVRR spectra were collected in 2 min using an excitation wavelength of 244 nm. Assignments are listed in Table 6.2. Values for n are listed in Table 6.1.

— pH 4 — pH 4.5 — pH 5 — pH 5.5 — pH 6 — pH 6.5 — pH 7 — pH 7.1 — pH 7.5

With the buffer inference overcome, the next step was to examine WTFab and all four mutant samples using UVRR spectroscopy combined with 2DCA.

The Raman spectra of WTFab was collected in sodium citrate (pH 4-6.5) and sodium phosphate buffers (pH 7-7.5) and water. Spectral intensity variations can be observed mainly at ~1179, 1550 and 1617 cm^{-1} (Figure 6.5). The peaks at ~1179, 1550 and 1617 cm^{-1} are assigned to Tyr, Trp and Tyr/Trp respectively (Table 6.2).

Table 6.2 Proposed peak assignments for WTFab UVRR spectra.

Peak (cm^{-1})	Residue	Assignment	References
1179	Tyr	In-plane CH bend and ring stretching	(Ludwig and Asher, 1988, Hildebrandt et al., 1988, Rodriguez-Mendieta et al., 2005)
1208	Tyr	Symmetric stretch	(Ludwig and Asher, 1988, Wen and Thomas Jr., 1998)
1341/1363	Trp	Fermi-doublet	(Takeuchi, 2003, Takeuchi, 2011, Hashimoto et al., 1997)
1550	Trp	Indole ring vibration C-C stretch	(Johnson et al., 1984, Hashimoto et al., 1997, Takeuchi, 2003)
1617	Trp/Tyr	In-plane ring stretching	(Hildebrandt et al., 1988, Chi et al., 1998, Asher et al., 1986, Takeuchi, 2003, Ludwig and Asher, 1988).

The nature of the peak intensity at ~1179 cm^{-1} has been reported as a marker for hydrogen bonding (Takeuchi, 2011). Low intensities suggest that Tyr is acting as a proton donor whereas high intensities suggest that Tyr is acting as a proton acceptor. Medium intensities are indicative of donor-acceptor or non-hydrogen bonding states. Overall, the peak intensities at ~1179 cm^{-1} (Figure 6.5) are low, so

one may assume that Tyr is generally acting as a proton acceptor considering that Tyr is expected to be protonated in the examined pH range.

Increasing and decreasing peak intensities of peaks assigned to aromatic residues are usually interpreted as markers signalling changes in their exposure to solvent. Increased intensities generally indicate a more hydrophobic environment while decreased intensities indicate an increased exposure to solvent (Kim et al., 2003, El-Mashtoly et al., 2005). The pH-induced spectral variations in Figure 6.5 suggest a change in the exposure of some Tyr and Trp residues. However, it is difficult to identify distinct pH trends visually. To gain further insight into the pH-induced changes, 2DCA was applied to the data.

6.3.3 2D-Correlation Analysis of WTFab

6.3.3.1 Autocorrelation and Synchronous 2DPCMW

Autocorrelation and synchronous 2DPCMW plots were generated from the UVRR spectra of WTFab (Figure 6.6). The autopeaks at 1175, 1208, 1550 and 1617 cm^{-1} appear at similar positions to the Raman peaks in Figure 6.5 (Figure 6.6A).

There are also additional autopeaks at ~1588 and 1653 cm^{-1} which do not have corresponding peaks in the Raman spectra (Figure 6.5). A careful examination of the spectra (Figure 6.5) reveals that these autopeaks have possibly arisen from a broadening of the peak at ~1617 cm^{-1}. Broadening may suggest the presence of other species or different conformations (Pieridou and Hayes, 2009, Copeland and Spiro, 1985); it may also arise due to fluctuations of laser power and optic parameters (Lednev et al., 2005, Alden et al., 1990).

In the 2DPCMW plots, the contour sign changes at approximately every pH unit interval, indicating that there is no continuous change across any pH range for any

single peak (Figure 6.6B). These intensity variations may be due to the level of protonation in citrate and phosphate buffer as this has been shown to shift the maximum UV-absorption wavelength of aromatic molecules. (Alata et al., 2012, Grante et al., 2014).
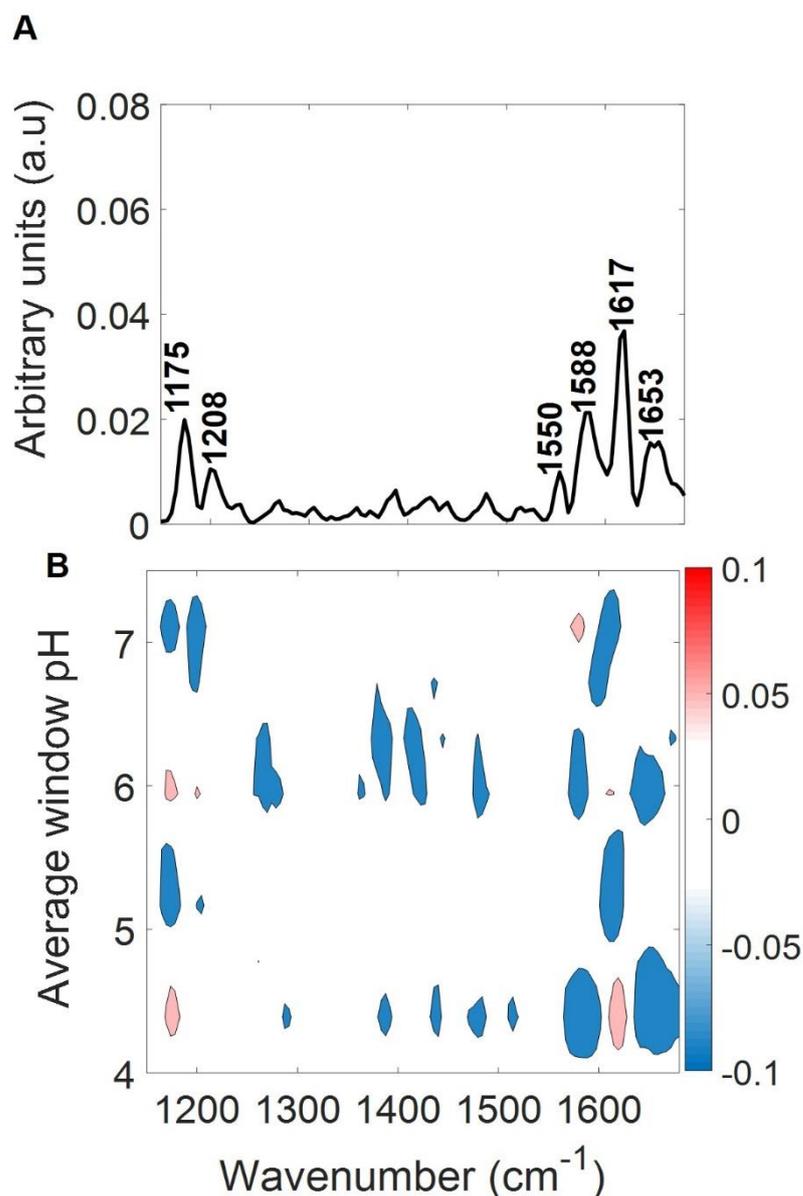
Figure 6.6 (A) Autocorrelation and (B) Synchronous 2DPCMW plots generated from pH-dependent UVRR WTFab spectra. Blue coloured contours indicate peak intensity decrease with increasing pH. Red shaded contours show peak intensity increase with increasing pH. The scale on the colour has arbitrary units. A moving window size of 3 and a maximum number of 2 contours was applied.

In aqueous solutions, a mixture of mono, di and tribasic citrate ion forms may be present depending on the environmental factors such as pH and temperature. In contrast, monoprotic buffers such as acetate buffer form only a mono ion (Krukowski et al., 2017). Consequently, the maximum absorption wavelength of polyprotic acids such as phosphate and citric acid tends to increase with pH, whereas that of acetate buffer remains relatively constant with pH (McConnell et al., 1993). As mentioned previously (section 5.3.3.1), citrate ions have been observed to accumulate on the protein surface preferentially. The pH response, as well as preferential accumulation of citrate ions on the protein surface of WTFab, might account for a more significant resonance effect and consequently a high sensitivity to changes in intensity.

It is interesting to note that there is a similarity between the contour patterns for the autopeaks at ~1175 and 1617 cm^{-1} (Figure 6.6B). Inferring from Noda's rules about the appearance of synchronous cross peaks (section 1.6.2), this similarity indicates that both autopeaks have a similar origin or may be assigned to the same type of residue. This ability to clarify peak assignments is undoubtedly one of the significant advantages of applying 2DCA.

In the UVRR spectroscopy of aromatic residues, one of the most resonantly enhanced electronic transitions is the L_a π-π* transition, affecting Tyr side chains at ~1618 (Y_{8a}) and 1175 cm^{-1} (Y_{9a}) (Ludwig and Asher, 1988). Trp residues can also be excited at this wavelength with overlapping peaks at ~1614-1620 cm^{-1} (Table 6.2). Nevertheless, based on the similar behaviour with the autopeak at 1175 cm^{-1} which is wholly associated with a Tyr Raman peak (Table 6.2), the peak

at ~1617 cm^{-1} in the Fab spectra is likely associated with Tyr residues rather than Trp residue.

The positive contours at ~1175 and 1617 cm^{-1} suggest an increasing peak intensity from pH 4.5 and 6 compared at neutral pH (native pH) where negative contours are observed. This lower peak intensity at lower pH suggests that these Tyr residues are in a less hydrophobic environment with increased solvent exposure. Whereas at neutral pH, the Tyr residues are less solvent-exposed.

This observation interestingly complements the results from chapter 5, where information about the aromatic residues was not obtained due to overlapping buffer peaks. Thus, in addition to the loss of β-sheet at pH 4.5 and 5.5, there is possibly an increased solvent exposure of aromatic residues. These results are consistent with the previous observation of increased solvent exposure of an aggregation-prone region on the C$_L$ domain at low pH (Codina et al., 2019a, Codina et al., 2019b).

6.3.3.2 Asynchronous 2DPCMW

As discussed, the synchronous 2DPCMW plots reveal how the spectral intensity changes with the perturbation, whereas the asynchronous 2DPCMW plot aids the identification of specific transition regions in the data (section 1.6.4).

Due to the change at every pH unit in the synchronous 2DPCMW, it was necessary to identify the transition points in the data using the asynchronous 2DPCMW plot so that the data could be grouped appropriately for further analysis. Specifically, clarification of nature spectral changes occurring from ~1570-1710 cm^{-1} using 2D-correlation plots. As discussed in section 5.3.7.2, spectral changes such as broadening can be clearly identified from patterns in synchronous and

asynchronous plots. Figure 6.7 shows the asynchronous 2DPCMW plot. The broken red line indicates the inflection point in the data at ~pH 5.2. This pH value links with the previously discussed stability of WTFab above ~pH 5.5 and divides the data into just two regions, making it easier to analyse using generalised 2DCA.
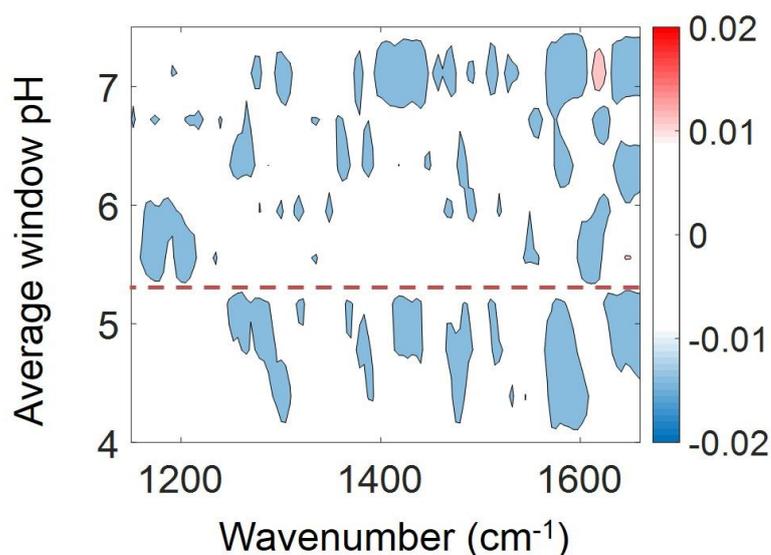
Figure 6.7 Asynchronous 2DPCMW plot generated from the pH-dependent UVRR spectra of WTFab. The red broken line indicates the transition point along the perturbation axis. The scale on the colour bar has arbitrary units. A moving window size of 3 and a maximum of 2 contours was applied.

6.3.3.3 Synchronous and Asynchronous plots

Synchronous and asynchronous plots were generated for the spectral region, ~1570-1710 cm^{-1}. In Figure 6.8A, the synchronous plot (pH 4-5.2) depicts a pattern of contours that is typical of single peak broadening at ~1617 cm^{-1} (Noda and Ozaki, 2005). This broadening can also be observed at a higher pH range (Figure 6.8B). However, the corresponding asynchronous plots are not a cross pattern as expected in the case of a single peak broadening (Noda and Ozaki,

2005). This occurrence suggests that broadening is only part of the spectral change occurring. The asynchronous plot instead reveals a *butterfly* pattern with elongated peaks close to the diagonal, which is a strong indication of the presence of a peak shift with intensity changes (Noda and Ozaki, 2005).

Figure 6.8 Synchronous (SP) and Asynchronous plots (AP) generated from pH-dependent UVRR spectra of WTFab. (A) SP, pH 4-5.2 (B) SP, pH 5.5-7.5 (C) AP, pH 4-5.2 (D) AP, pH 5.5-7.5. The red arrows indicate the proposed peak shift direction. Shaded contours indicate negative cross peaks and unshaded contours indicate positive cross peaks. Positive cross peaks indicate two wavenumbers with perturbation induced intensity changes in the same direction (increasing or decreasing together). Negative cross peaks indicate two wavenumbers with perturbation induced intensity changes in opposite directions (one increasing and the other decreasing). Maximum contour level=8.

In addition to the elongated cross peaks, one would also expect to observe smaller cross-peaks of opposite signs away from the diagonal (next to the elongated cross peaks). In Figure 6.8C & 6.8D, the cross peaks away from the diagonal are of opposing signs but are of similar size or larger than the elongated cross peaks near the diagonal. This difference could be as a result of the combination of broadening with a peak shift. The proposed peak shift direction is interpreted using a different rule from the standard set of Noda's rules (section 1.6.3). Consequently, the sequential order interpretation does not apply in the case of peak shifts. Where the elongated cross peak above the diagonal in the asynchronous plot is negative and that below is positive then the peak shift occurs from right to left on the horizontal axis and from top to bottom on the vertical axis (Noda and Ozaki, 2005). Therefore, in Figure 6.8, the proposed peak shift occurs from higher wavenumber to lower wavenumber in the ~1588 to 1653 cm^{-1} region at low pH (pH 4-5.2) and in the opposite direction at higher of pH 5.5-7. The opposing directions may account for differences in the stability of WTFab in the two pH ranges.

The peak at ~1617 cm^{-1} appears central to the peak shift. However, it is difficult to unambiguously interpret whether or how other wavenumbers are involved in the observed shift; because the synchronous plot is complicated by broadening, and there are multiple cross peaks in the asynchronous spectrum. Nevertheless, the appearance of such cross peaks has been highlighted as not being a valid reason to dismiss a single peak shift, since the *butterfly* pattern is a strong indicator of a peak shift (Noda and Ozaki, 2005).

Although only the 1617 cm^{-1} peak is visible in the spectra, it is essential to note that Tyr peaks may occur between ~1554-1620 cm^{-1} (Table 6.3). UVRR Tyr peaks

at ~1617 cm^{-1} (Y$_{8a}$), and ~1600 (Y$_{8b}$) are assigned to asymmetric and anti-asymmetric vibrational modes arising from the substituted benzene ring in the Tyr side chain (Copeland and Spiro, 1985). They have been studied widely for their peak shifts signalling the Tyr protonation, deprotonation and hydrogen bonding. Compared to (Y$_{8a}$), (Y$_{8b}$) tends to be generally less resolved in spectra, often appearing as a weak shoulder (Ludwig and Asher, 1988, Pieridou and Hayes, 2009). During deprotonation as in high alkaline pH, downshifts have been reported to occur to ~1600 cm^{-1} for (Y$_{8a}$) and ~1554 cm^{-1} for (Y$_{8b}$) (Hildebrandt et al., 1988, Pieridou and Hayes, 2009, Takeuchi, 2011). On the other hand, protonation has been associated with slight upshifts of ~1600 cm^{-1} (Y$_{8b}$) with a corresponding decrease in the peak intensity at ~1617 cm^{-1} (Y$_{8a}$) (Rodgers et al., 1992). Downshifts of (Y$_{8a}$) have also been associated with increased hydrogen bond donation, and upshifts are linked with decreased hydrogen bond donation (El-Mashtoly et al., 2005). It is unlikely that Tyr is deprotonated within this examined pH range pH 4-5-7 due to the high pK$_a$ (~10.2) of aqueous Tyr (Pieridou and Hayes, 2009). Also, the ~1617 cm^{-1} peak does not disappear at any point in the spectra as would be expected in the case of deprotonation. Furthermore, it is difficult to link the peak shift to protonation since the associated spectral changes were not observed.

Therefore, the proposed direction of the peak shifts possibly suggests a change in hydrogen bonding donation. At higher pH (5.5-7.5), there is decreased hydrogen bond donation due to an upshift of Y$_{8a}$ (1617 cm^{-1}) whereas an increased hydrogen bond donation is occurring at lower pH. This decreased hydrogen bond donation may explain the increased stability of WTFab above ~pH 5.5. The loss of domain interfacial contacts has been reported to precede domain unfolding and loss of

secondary structure in WTFab (Codina et al., 2019b). Thus, the decreased hydrogen bond donation featuring Tyr at high pH may indicate the retention of such interfacial contacts particularly in the V_H-V_L region where Tyr residues (n=6) outnumber Trp or Phe residues (n=2) (Codina et al., 2019b).

Table 6.3 Reported UVRR Tyr peak positions from ~1550-1620 cm^{-1}.

Peak (cm^{-1})	Vibration	References
1554,1587,1593,1597,1600	(Y$_{8b}$)	(Ludwig and Asher, 1988, Hildebrandt et al., 1988, Pieridou and Hayes, 2009, El-Mashtoly et al., 2005, Takeuchi, 2011, Rodriguez-Mendieta et al., 2005, Jones et al., 2014)
1600,1606,1615 1617,1621	(Y$_{8a}$)	

Whereas it has been repeatedly mentioned in previous chapters that autopeaks may have a position close to an observed Raman peak in the spectra; it is difficult to arrive at that conclusion for the autopeaks at ~1588 cm^{-1} and 1653 cm^{-1}.

The closest Raman peaks in the spectra are ~1550 and 1617 cm^{-1}. However, there are autopeaks at ~1550 and 1617 cm^{-1}, respectively (Figure 6.6A). There is a possibility that the autopeaks represent spectral intensity variations in overlapped Raman peaks at ~1588 and 1653 cm^{-1} assigned to Tyr/Phe or Trp (Hildebrandt et al., 1988, Chi et al., 1998) and β-turn/sheet (Chi et al., 1998, Huang et al., 2006) respectively. A Tyr Raman peak at ~1587 cm^{-1} (Y$_{8b}$) has been attributed to a strong hydrogen bond with another residue (Hildebrandt et al., 1988). In the case of Tyr residue being linked to the autopeak at ~1588 cm^{-1}, its shared origin with ~1653 cm^{-1} as indicated from the contour pattern in the synchronous 2DPCMW plot, as well as their joint involvement in the broadening,

may represent the formation of hydrogen bonds between the Tyr residue(s) and the carbonyl groups in the folded β-turns (Figure 6.6B).

6.3.4 pH-induced Variations (Fab Mutants)
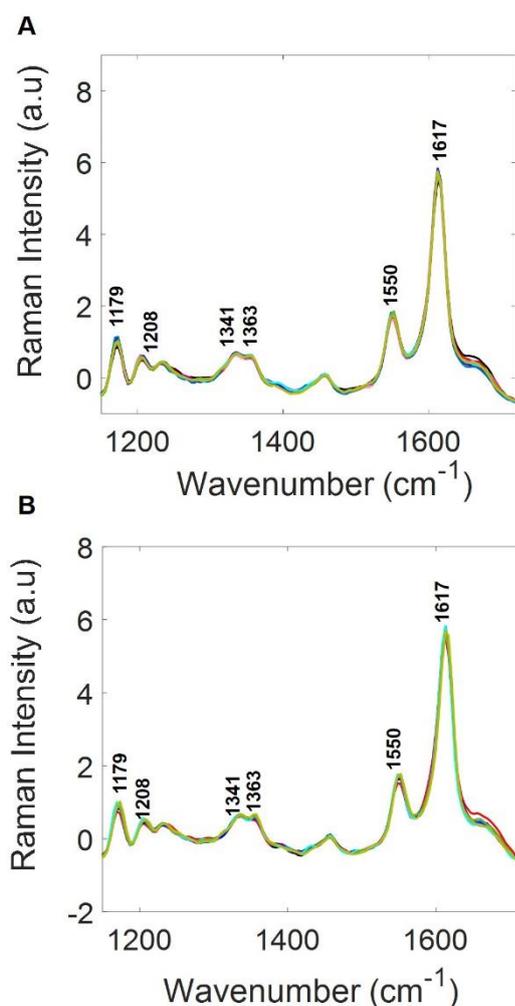
Figure 6.9 Averaged UVRR spectra of 10 mg/ml mutant Fab in water (pH 7.1), Cit, (pH 4-6.5) and NaP (pH 7 and 7.5). (A) MT1-1 (B) MT2-1. Values for n are listed in Table 6.1. Preprocessing carried out by normalisation and smoothing. UVRR spectra were collected in 2 min using an excitation wavelength of 244 nm. Assignments are listed in Table 6.2

pH 4   pH 4.5   pH 5   pH 5.5   pH 6   pH 6.5   pH 7   pH 7.1   pH 7.5

The four mutant samples, MT1-1, MT2-1, MT3+1 and MT4+1, were also examined using UVRR spectroscopy and 2DCA. As mentioned in section 5.1, the surface charges of these samples have been modified by either reducing or increasing positive charges. MT1-1 and MT2-1 both have one less positive charge in the V_H and C_H domains, respectively, whereas MT3+1 and MT4+1 have one more positive charge in the V_L and V_H domains, respectively.
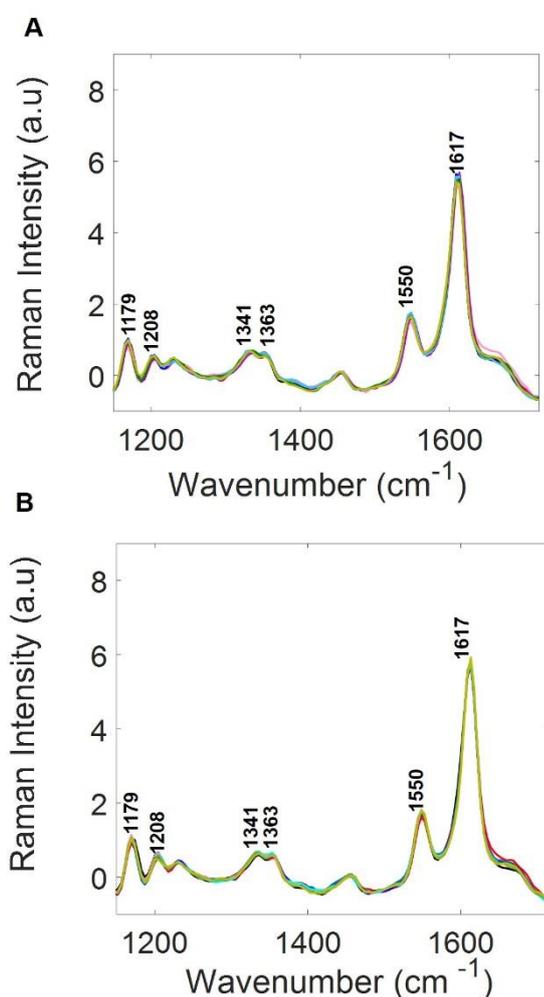
Figure 6.10 Averaged UVRR spectra of 10 mg/ml mutant Fab in water (pH 7.1), Cit, (pH 4-6.5) and NaP (pH 7 and 7.5). (A) MT3+1 (B) MT4+1. Values for n are listed in Table 6.1. Preprocessing carried out by normalisation and smoothing. UVRR spectra were collected in 2 min using an excitation wavelength of 244 nm. Assignments are listed in Table 6.2

— pH 4 — pH 4.5 — pH 5 — pH 5.5 — pH 6 — pH 6.5 — pH 7 — pH 7.1 — pH 7.5

Figure 6.9 and 6.10 shows the averaged preprocessed spectra for the mutant samples. The main peak assignments are listed in Table 6.2. In contrast to WTFab, the spectral sets corresponding to each mutant sample reveal minimal differences upon visual inspection.

6.3.5 2D-Correlational Analysis (MT1-1 and MT2-1)

6.3.5.1 Autocorrelation and Synchronous 2DPCMW
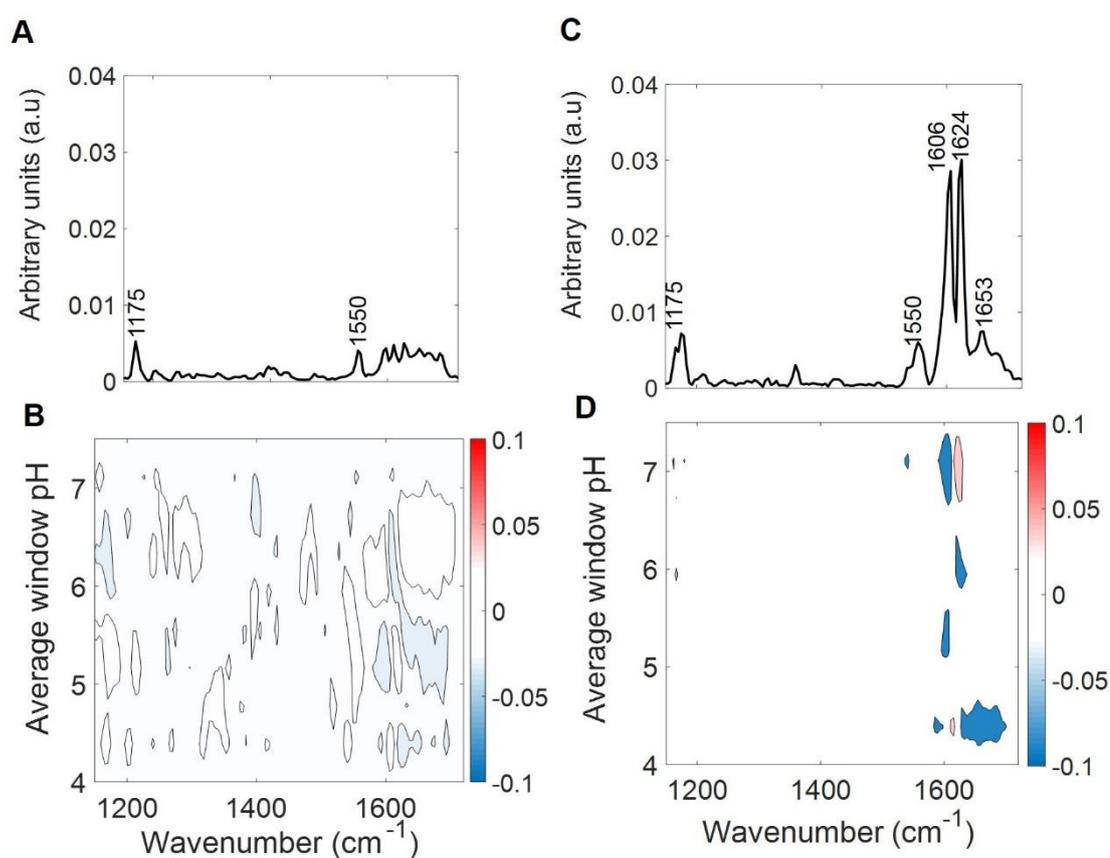
Figure 6.11 Autocorrelation and Synchronous 2DPCMW plots generated from the pH-dependent UVRR spectra of mutant Fab samples (A & B) MT1-1 (B & D) MT2-1. Blue coloured contours indicate a peak intensity decrease with increasing pH. Red shaded contours show a peak intensity increase with increasing pH. The scale on the colour bar has arbitrary units. A moving window size of 3 and a maximum of 2 contours was applied.

Figure 6.11 displays the autocorrelation plots and the corresponding synchronous 2DPCMW plots generated from the UVRR spectra of MT1-1 and MT2-1. The autopeaks from the mutant samples reflect changes not evident from the spectra (Figure 6.9). As discussed previously, the autopeaks at 1175 and $\sim$1550 cm^{-1} in the MT1-1 and MT2-1 autocorrelation plots indicate variations in the intensities of Raman peaks assigned to Tyr and Trp respectively suggesting changes in hydrogen bond donation and solvent exposure with pH. MT2-1 has additional autopeaks at $\sim$1606 and 1624 cm^{-1} compared to MT1-1 suggesting spectral variations near the Tyr peak at $\sim$1617 cm^{-1}.

6.3.5.2 Asynchronous 2DPCMW (MT1-1 and MT2-1)

As with WTFab, asynchronous 2DPCMW and 2D-correlation plots were generated further to examine the spectral changes from $\sim$1570-1710 cm^{-1}. Transition points at $\sim$pH 6.3 and 5.1 were identified for MT1-1 and MT2-1, respectively (Figure 6.12). Although there appears to be a middle transition point for MT2-1 at $\sim$pH 6 (Figure 6.12B), it was not considered separately because it did not contain the minimum number of three spectra recommended for generalised 2D-correlation plots (Noda, 2016b, Noda, 2018). The contour pattern in the 2D-correlation synchronous and asynchronous plots was the same regardless of the inclusion or exclusion of the corresponding spectra from the upper range.
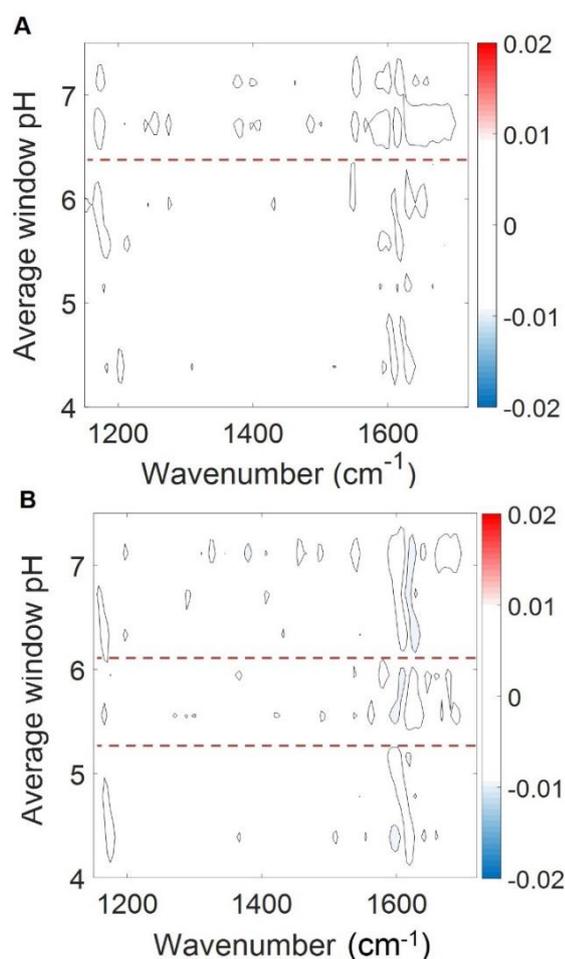
Figure 6.12 Asynchronous 2DPCMW plots generated from the pH-dependent UVRR Raman spectra of mutant Fab samples (pH 4.5-7.5). (A) MT1-1 (B) MT2-1. Red broken line indicates transition points along the perturbation axis. The scale on the colour bar has arbitrary units. A moving window size of 3 and a maximum of 2 contours was applied.

6.3.5.3 Synchronous and Asynchronous plots (MT1-1 and MT2-1)

Figure 6.13 displays the 2D-correlation plots (~1570-1710 cm^{-1}) for MT1-1. The similarity between the plots in the two pH ranges indicate that the spectral change is similar across the entire pH range. An autopeak at ~1593 cm^{-1} is now evident, indicating possible changes with Tyr residues (Table 6.3). However, the remainder of the plot is difficult to interpret likely due to the weakness of the spectral changes as shown in its autocorrelation and synchronous 2DPCMW plots (Figure 6.11A and 6.11B). Figure 6.14A displays the MT2-1 synchronous plots. There is a cross pattern which indicates the broadening of the ~1617 cm^{-1} peak (pH 4-5.2). The

corresponding asynchronous plot pattern indicates a peak downshift occurring in each pH range.
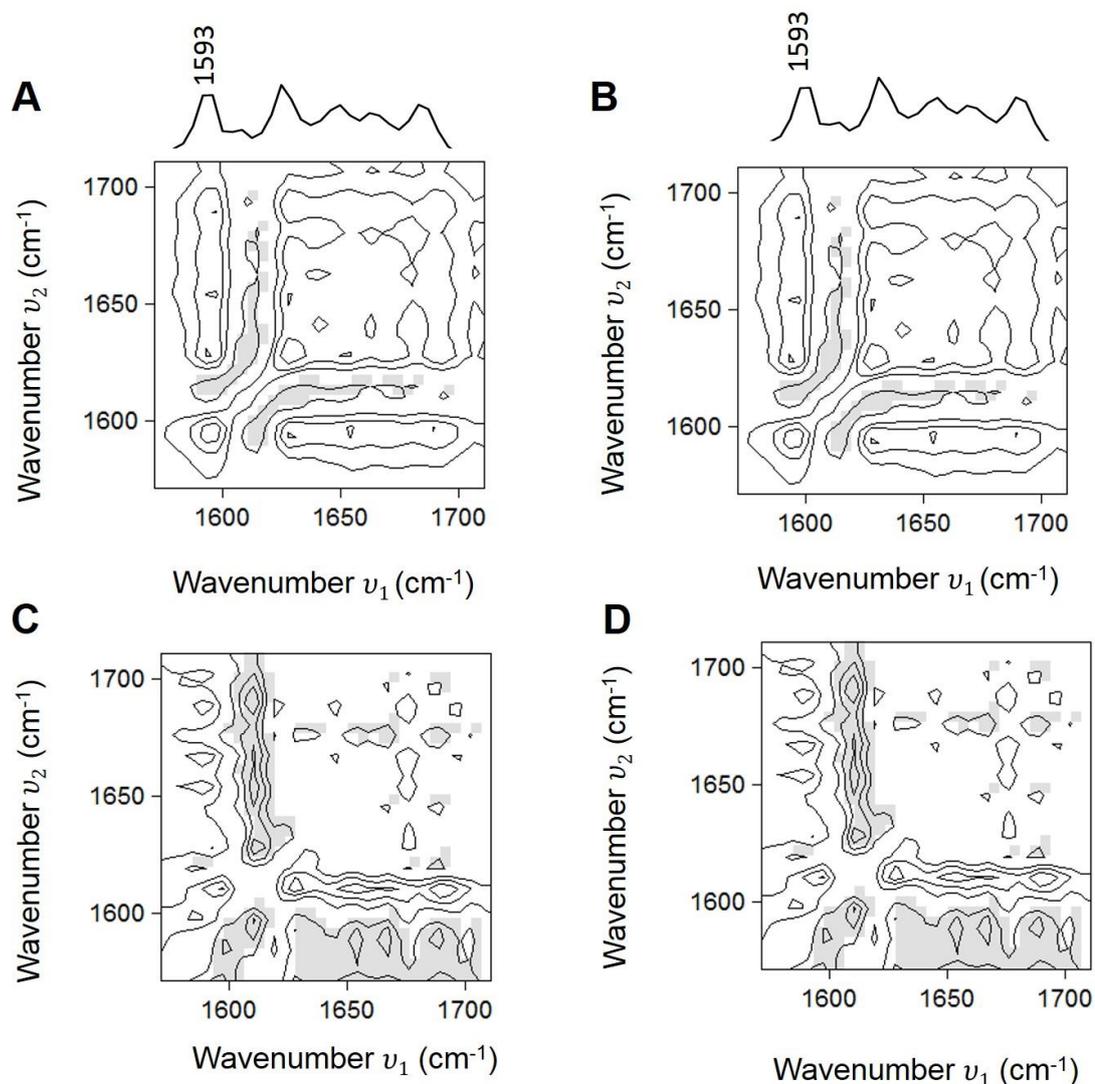
Figure 6.13 Synchronous plot (SP) and Asynchronous plot (AP) generated from the pH-dependent UVRR spectra of MT1-1 (A) SP, pH 4-6.3 (B) SP, pH 6.7-7.5 (C) AP, pH 4-6.3 (D) AP, pH 6.7-7.5. Shaded contours indicate negative cross peaks and unshaded contours indicate positive cross peaks. Positive cross peaks indicate two wavenumbers with perturbation induced intensity changes in the same direction (increasing or decreasing together). Negative cross peaks indicate two wavenumbers with perturbation induced intensity changes in opposite directions (one increasing and the other decreasing). Maximum contour level=4.

Unlike with WTFab where broadening was also observed, a comparison of the autopeaks in the two pH ranges, 4-5.2 and 5.5-7.5 reveals different sets of wavenumbers. The different values of the autopeaks in the pH 4-5.2 (1593, 1615) and pH 5.5-7.5 (1606, 1624) ranges suggest a shift at two different wavenumbers.

As previously discussed in section 6.3.3.3, a *butterfly* pattern in the corresponding asynchronous plot clearly indicates a peak shift. The MT2-1 synchronous spectrum (Figure 6.14B) is uncomplicated by broadening. It features what is referred to as a *four-leaf-clover* cluster pattern indicative of two overlapped bands with changing intensities or a peak shift ((Noda and Ozaki, 2005). The corresponding asynchronous plot (Figure 6.14D) has a *butterfly pattern* confirming the presence of a peak shift.

As discussed previously, Tyr peaks may appear between $\sim$1600-1620 cm^{-1} (Table 6.3). Thus, the two sets of autopeaks possibly represent Tyr peak upshifts from 1593 to $\sim$1606 cm^{-1} (Y_{8b}) and $\sim$ 1615-1624cm^{-1} (Y_{8a}) as pH is increased from 4-7.5. Using UVRR spectroscopy, Jones et al. (2014) investigated the closed-open transition effected by calcium-induced changes in calmodulin protein. They observed similar overlapping Tyr peak shifts at $\sim$1600-1620 cm^{-1} which they associated with the donation of a hydrogen bond by Tyr138 to Glu82 following the increased solvent exposure of Tyr138. The original UVRR Tyr peak was at $\sim$1616 cm^{-1} (closed transition state) however upon calcium-binding (open transition state) UVRR difference spectra and singular value decomposition (SVD) analysis revealed peaks at $\sim$1593, 1620 and 1589, 1600, 1615 and 1621 respectively.
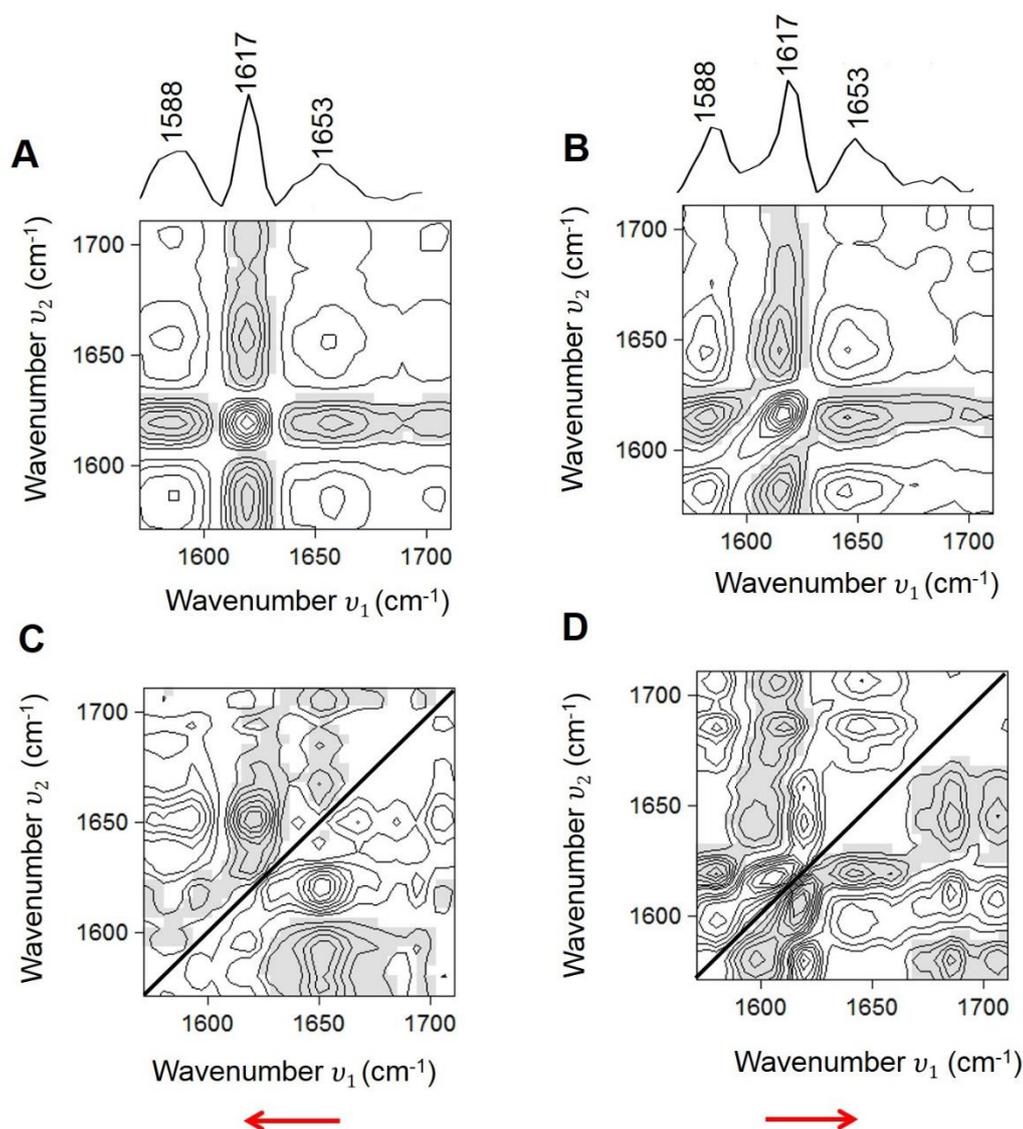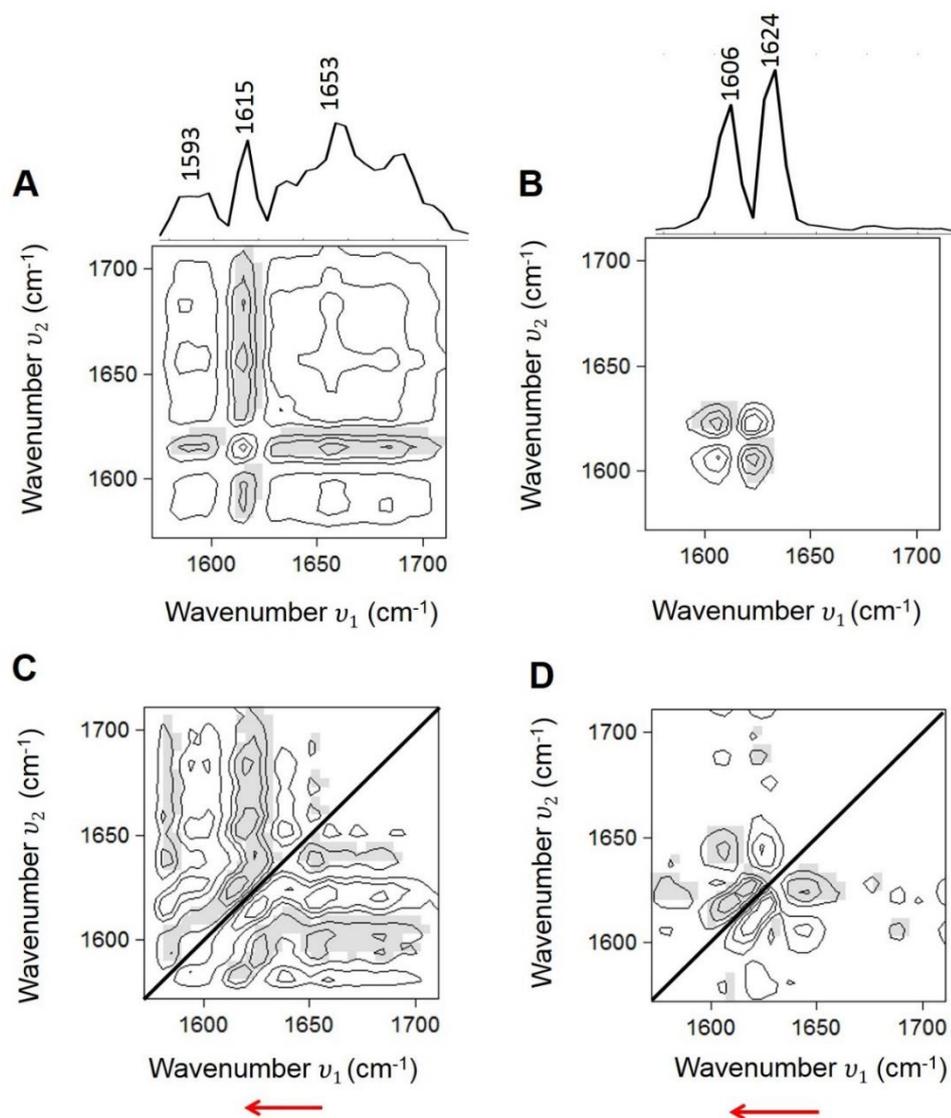
Figure 6.14 Synchronous plot (SP) and Asynchronous plot (AP) generated from the pH-dependent UVRR spectra of MT2-1. (A) SP, pH 4-5.2 (B) SP, pH 5.5-7 (C) AP, pH 4-5.2 (D) AP, pH 5.5-7.5. The red arrow shows the proposed peak shift directions. Shaded contours indicate negative cross peaks and unshaded contours indicate positive cross peaks. Positive cross peaks indicate two wavenumbers with perturbation induced intensity changes in the same direction (increasing or decreasing together). Negative cross peaks indicate two wavenumbers with perturbation induced intensity changes in opposite directions (one increasing and the other decreasing). Maximum contour level=4.

As previously discussed in section 6.3.3.3, the upshift of the Tyr peaks (Y_{8a}) is associated with decreased hydrogen donation (El-Mashtoly et al., 2005). Upshifts of (Y_{8b}) have also been partly associated with protonation (Rodgers et al., 1992). It is expected that protonation will be linked to a decreased hydrogen bond

donation. Consequently, these overlapping peak shifts may represent a decreased hydrogen bond donation from ~pH 5.5-7 compared to at ~pH 4-5.2. As suggested for WTFab, the decreased hydrogen bond donation at higher pH may also be linked to its higher stability above pH 5.5 observed in citrate buffer (section 5.3.5).

The broad autopeak at ~1653 cm^{-1} may also signal hydrogen bond interactions of Tyr with carbonyl groups as discussed for WTFab.

6.3.6 2D-Correlation Analysis (MT3+1 and MT4+1)

6.3.6.1 Autocorrelation and Synchronous 2DPCMW
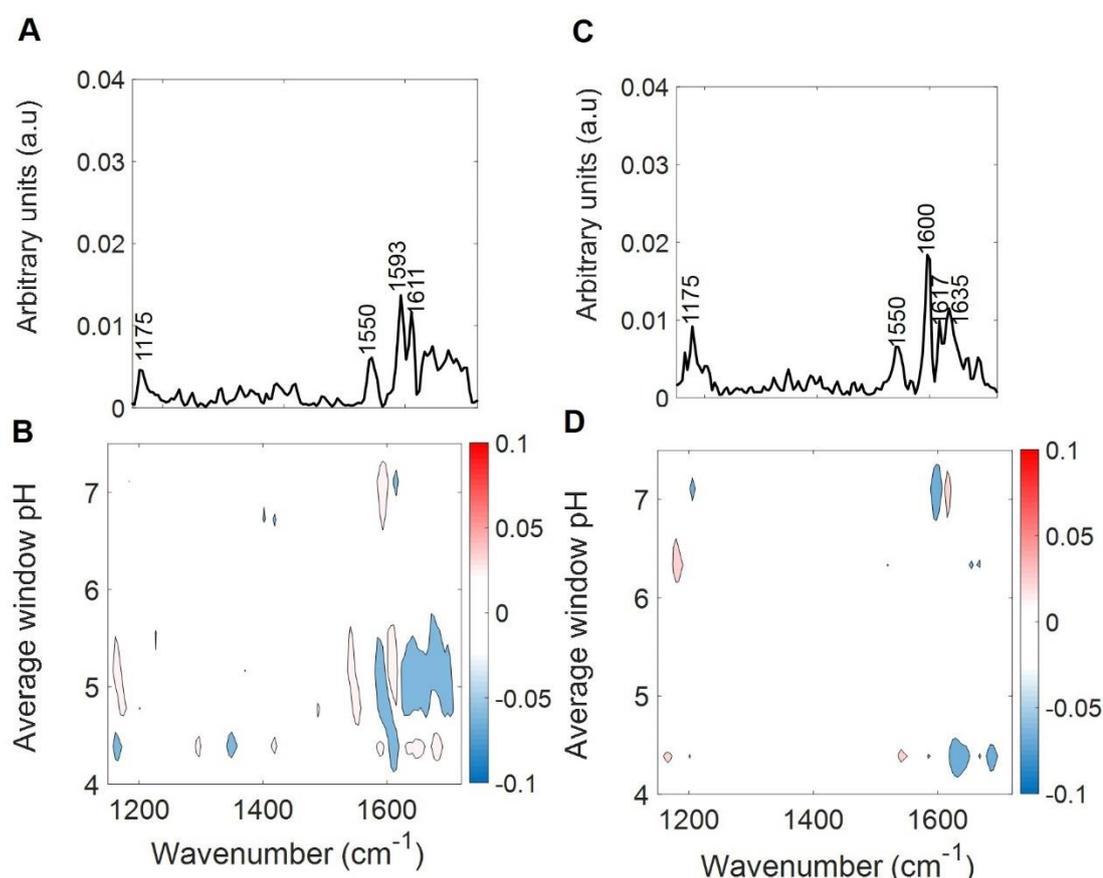
Figure 6.15 Autocorrelation and Synchronous 2DPCMW plots generated from the pH-dependent UVRR spectra of mutant Fab samples (A & B) MT3+1 (B & D) MT4+1. Blue coloured contours indicate a peak intensity decrease with increasing pH. Red shaded contours show a peak intensity increase with increasing pH. The scale on the colour bar has arbitrary units. A moving window of size of 3 and a maximum number of 2 was applied.

Autocorrelation and synchronous 2DPCMW plots were also generated for MT3+1 and MT4+1 (Figure 6.15). There are autopeaks at 1611 and 1635 cm^{-1}. The other autopeaks are at similar positions to those obtained for MT1-1, MT2-1 and WTFab.
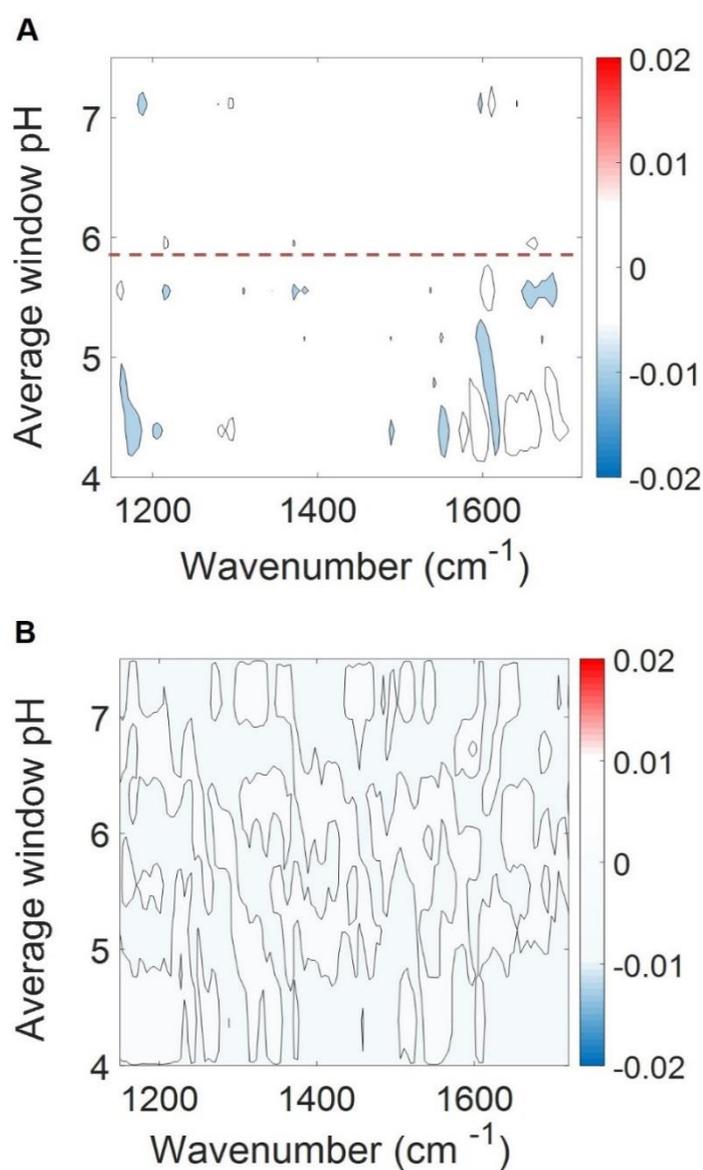
Figure 6.16 Asynchronous 2DPCMW plot generated from pH-dependent UVRR Raman spectra. (A) MT3+1 (B) MT4+1. Red broken line indicates a transition point along the perturbation axis. The scale on the colour bar has arbitrary units. A moving window of size of 3 and a maximum number of 2 was applied.

The pattern of contours in their synchronous 2DPCMW plots is different from MT2-1 and WTFab, suggesting some differences in their pH-dependent behaviour. Transition points were identified at ~pH 5.9 for MT3+1, and none was identified for MT4+1 (Figure 6.16B).

6.3.6.2 Synchronous and Asynchronous plots

Synchronous and asynchronous plots were also generated from the pH-dependent UVRR spectra of MT3+1 and MT4+1. The MT3+1 synchronous plot (Figure 6.17A) depicts a pattern indicative of broadening at ~1611 cm^{-1} as observed with WTFab and MT2-1. In contrast to WTFab, the corresponding asynchronous plot indicates the presence of overlapped peaks (Figure 6.17C). The two resolved cross peaks in the asynchronous plot near ~1600 cm^{-1} suggests the presence of overlapped UVRR peaks (1593 and 1611 cm^{-1}) with changing intensities (Noda and Ozaki, 2005). These peak intensities (from pH 4-5.9) can be observed to decrease and increase respectively in the synchronous 2DPCMW plot (Figure 6.15B) and have been linked to Tyr vibrations (Y_{8a}) and (Y_{8b}) respectively (Table 6.3).

As earlier discussed in section 6.3.2, increased peak intensities indicate a more hydrophobic environment, whereas decreased peak intensities indicate a more solvent-exposed environment. Thus, these opposing spectral intensity changes may signal two different environments of some Tyr residues. Conversely, the spectral changes may reflect opposing Tyr vibrations; symmetric (Y_{8a}) and antisymmetric (Y_{8b}), of Tyr residues in the same environment.

In the ~pH 6.3-7.5 range, the synchronous plot depicts an *angel* pattern (Figure 6.17B). As previously discussed in section 3.3.3.3, the *angel* pattern can be either

indicative of two overlapped peaks with markedly different intensities or a peak shift (Noda & Ozaki, 2005). The exact interpretation is dependent on the pattern of the cross peaks in the corresponding asynchronous plot (Figure 6.17D).

Since the asynchronous plot lacks a butterfly pattern, no peak shift is indicated. Therefore, the *angel* pattern here indicates two overlapped peaks with markedly different intensities. Again, this is evident from the synchronous 2DPCMW plot in Figure 6.15B with a large positive contour ~1593 cm^{-1} and a weak negative contour at ~1611 cm^{-1}. There is a striking difference between the autocorrelation plots in Figure 6.17A and 6.17B. It can be observed that whereas the autopeaks (1593 and 1611 cm^{-1}) are of similar strengths between pH 4-5.9, the autopeak at ~1611 cm^{-1} is much weaker than ~1593 cm^{-1} above ~pH 6. It is also interesting to note in the 2D synchronous PCMW plot, the switch in contours signs between the lower and higher pH range in addition to the significant weakening of the autopeak at ~1611 cm^{-1} (Figure 6.15B).

Both observations suggest that there is a marked difference between the two pH ranges (4-5.9) and (6.3-7.5) possibly due to changes in solvent exposure of the Tyr residue. However, it is difficult to suggest whether a hydrophobic environment dominates at the higher pH range or *vice versa* because of the opposing spectral intensity changes which may be arising from the same set of Tyr residues.

What is obvious, however, is that MT3+1 is behaving differently from WTFab and MT2-1 where peak shifts linked to changes in hydrogen bond donation were observed. Consequently, it is reasonable to suggest that MT3+1's higher stability above ~pH 5.5 might be linked to a significant change in solvent exposure.

With MT4+1, there is only one set of synchronous and asynchronous plots due to the lack of identification of a transition point. The synchronous plot has a *four-leaf clove*r cluster, albeit with a broad autopeak at ~1635 cm^{-1} (Figure 6.18A).
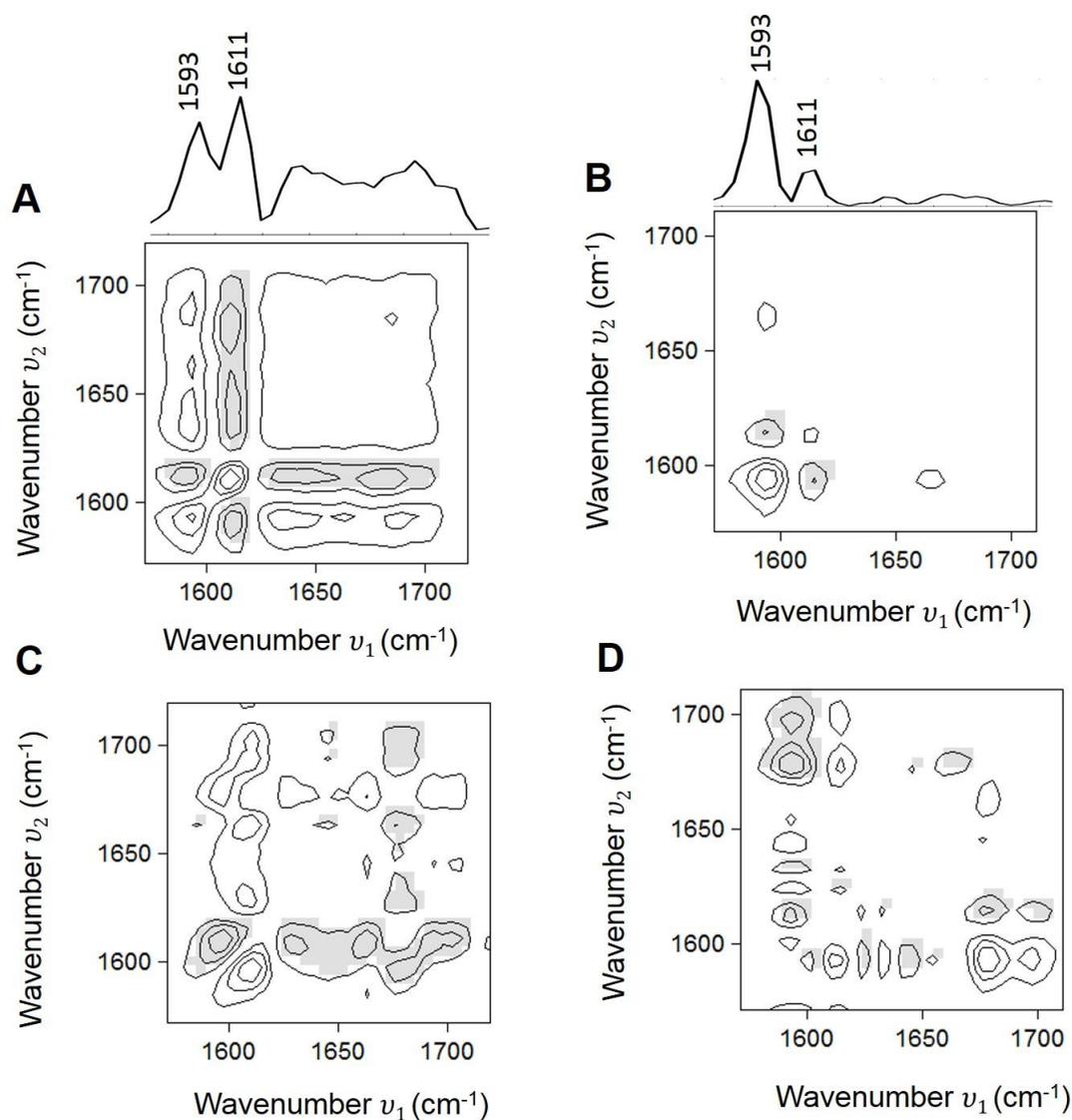
Figure 6.17 Synchronous plot (SP) and Asynchronous plot (AP) generated from the pH-dependent UVRR spectra of MT3+1. (A) SP, pH 4-5.9 (B) SP, pH 6.3-7.5 (C) AP, pH 4-5.9 (D) AP, pH 6.3-7.5. Shaded contours indicate negative cross peaks and unshaded contours indicate positive cross peaks. Positive cross peaks indicate two wavenumbers with perturbation induced intensity changes in the same direction (increasing or decreasing together). Negative cross peaks indicate two wavenumbers with perturbation induced intensity changes in opposite directions (one increasing and the other decreasing). Maximum contour level=4.

The asynchronous plot indicates a peak upshift and the maxima of the peak involved in the shift corresponds to the position between the cross peaks in the synchronous plot while the cross peaks indicate the boundaries of the shifted peak (Noda and Ozaki, 2005). Accordingly, the results indicate an upward peak shift from $\sim$1617 cm^{-1} suggesting decreased hydrogen bond donation. From the synchronous 2DPCMW plot (Figure 6.15D), this peak shift occurs at $\sim$pH 7. As previously mentioned, the decreased hydrogen bond donation might account for increased stability.
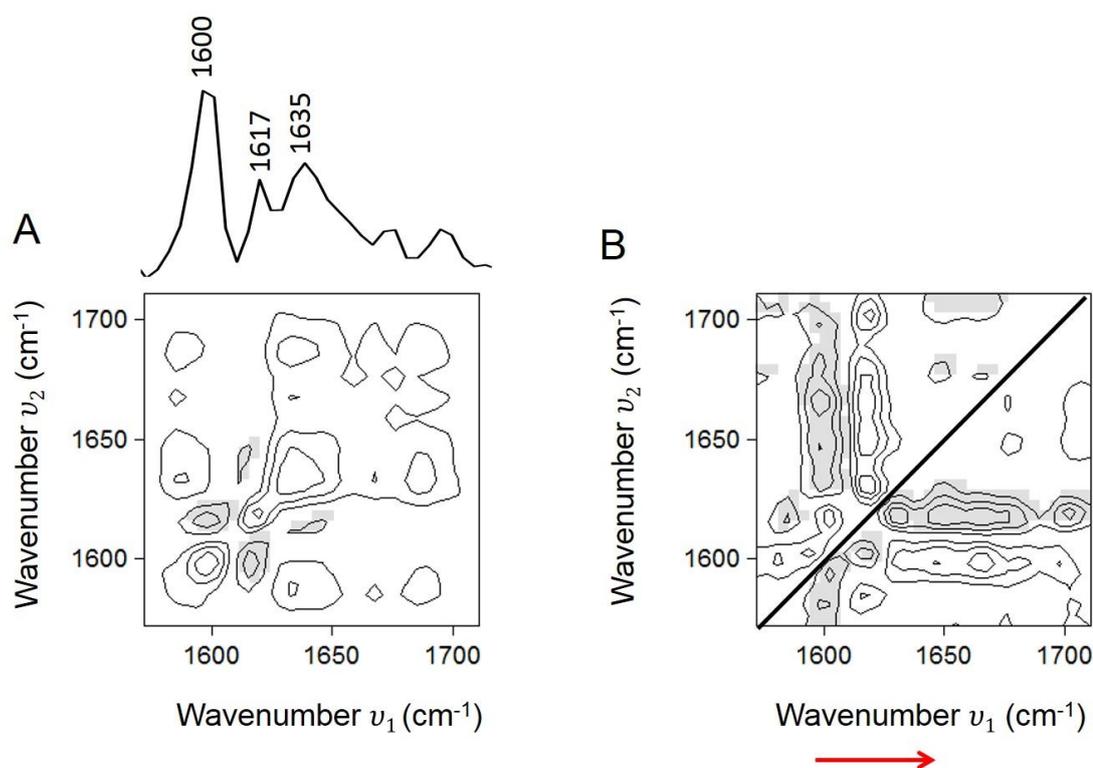
Figure 6.18 Synchronous plot (SP) and Asynchronous plot (AP) generated from the pH-dependent UVRR spectra of MT4+1 (A) SP, pH 4-7.5 (B) AP, pH 4-7.5. The red arrow shows the proposed peak shift direction. Shaded contours indicate negative cross peaks and unshaded contours indicate positive cross peaks. Positive cross peaks indicate two wavenumbers with perturbation induced intensity changes in the same direction (increasing or decreasing together). Negative cross peaks indicate two wavenumbers with perturbation induced intensity changes in opposite directions (one increasing and the other decreasing). Maximum contour level=4.

The lack of a transition point in the MT4+1 asynchronous 2DPCMW plot does not allow for a suggestion as to the reason for its lower stability below pH 5.5. The inability to detect a transition point might be due to its high solvent accessible area (SAS) compared to other mutants and WTFab (Ahmad, 2011). Its already high exposure to solvent might result in less marked spectral changes compared to a mutant with a lesser SAS. The autopeak at ~1635 cm^{-1} may also suggest changes in β-sheet hydrogen bonding as a Raman peak at ~1635 cm^{-1} has also been associated with β-sheet (Huang et al., 2006).

6.4 Comparison of WTFab and Mutant Samples

The combination of the 2DPCMW and 2D-correlation plots revealed the complexity of pH-dependent variations in the UVRR spectra of WTFab and its mutants. The results mainly suggest changes in hydrogen bonding and solvent exposure primarily due to the selective enhancement of Tyr residues in the spectra. Compared to all the mutants, WTFab appears particularly sensitive to variations in spectral intensities possibly due to pH-induced shifts in the maximum UV absorption wavelength of citrate buffer. The reduced effect in the mutants may be due to differences in the distribution of the citrate ions on the protein surface as a result of altered charge interactions. This sensitivity may also account for the enhanced clarity of the mutant 2DCA plots in general compared to that of WTFab.

A decrease in hydrogen bond donation by Tyr appears to account for the increased stability of WTFab and MT2-1 above their pH transition point (Table 6.4). However, with MT2-1, the peak upshifts suggestive of decreased hydrogen bond donation is evidenced by autopeaks at different wavenumbers, demonstrating the

increased clarity of spectral changes provided by 2DCA. Decreased hydrogen

donation was also attributed to the peak shift observed with MT4+1 at ~pH 7.

Hydrogen bonding via Tyrosine OH groups have been shown to contribute to

protein stability (Pace et al., 2001). Tyrosine can usually form two hydrogen bonds

as hydrogen bond acceptor and donor via its sidechain hydroxyl group (Figure

6.19).
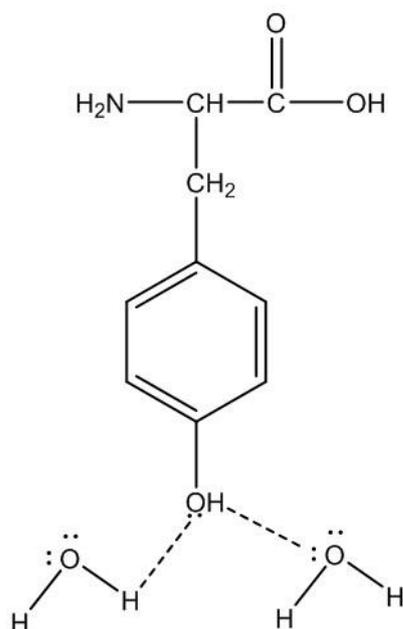
Figure 6.19 Structure of Tyrosine showing hydrogen bonding with water molecules as donor (using H atom) and acceptor using lone pair of electrons on its oxygen atom.

However, it is twice as likely to act as a hydrogen bond donor than an acceptor;

and the acceptor bonds are weaker possibly due to the delocalisation of one of the

lone electron pairs in the aromatic ring (McDonald and Thornton, 1994). Baker &

Hubbard 1994 showed that Tyr OH-groups are more likely to form bonds with

water molecules than with back bone or side chain groups. They also showed that

number of H-bonds associated with a Tyr side chain could vary from 0 to > 3

Therefore, the decreased hydrogen bond donation between pH 5.5-7 likely refers

to the reduced number of H-bonds donated by the implicated Tyr residues to solvent. As discussed, the Fab proteins are more stable between pH 5.5-7. Thus, the reduced number of Tyr H-bonds being donated possibly signal a folded state of the protein and consequently, increased conformational stability.

MT3+1 is distinguished from WTFab and the other mutants as its pH transition point and consequently, stability between pH 5-7 corresponds to a marked change in solvent exposure. This difference may be due to the location of its mutation on the variable light chain (V_L) compared to other mutants with mutations on the heavy chain (Table 6.4). The MT1-1 results provided minimal insight due to very weak spectral variations, and this has been attributed to it having a lower SAS. However, there was evidence of Tyr peak changes as observed with the other mutants and WTFab.

Table 6.4 pH transition points and proposed stability mechanisms from 2DCA analysis

	WTFab	MT1-1	MT2-1	MT3+1	MT4+1
Mutation site/residue		V_H/Lys	C_H/Lys	V_L/Leu	V_H/Ser
pH-transition point	5.2	6.3	5.5	5.9	ND
Proposed mechanism for increased stability (>pH 5.5)	Decreased hydrogen bond donation	Not detected	Decreased hydrogen bond donation	Change in solvent exposure	Decreased hydrogen bond donation

6.5 Conclusion

WTFab and its four surface charge mutants were examined using UVRR spectroscopy and 2DCA. At 244 nm, the buffer peaks did not interfere with the protein peaks in the spectra, thus allowing for the unrestricted analysis of the entire UVRR Fab spectra using 2DCA. Unlike the previous chapters, the spectral changes were generally less obvious, making the autocorrelation and synchronous 2DPCMW plots more complex to interpret in conjunction with the spectra. The fragmented pattern of contours along the perturbation axis was addressed by using the asynchronous 2DPCMW plots to identify transition points in the data. This data segmentation improved the understanding of the synchronous 2DPCMW plots as well as the conventional 2D-correlation plots (synchronous and asynchronous). The identification of transition points between pH 5-6 is consistent with the previously observed stability points of WTFab and the four mutant samples above ~pH 5.5. However, 2DCA analysis of the UVRR spectra in these transition regions suggest a link between the stability of the samples and interactions involving Tyr residues. Specifically, the higher stability of WTFab, MT2-1 and MT4+1 above pH 5.5 might be linked to a decrease in hydrogen bond donation, whereas that of MT3+1 is possibly due to marked changes in solvent exposure. It was not possible to propose a mechanism for MT1-1 due to the weak nature of the observed spectral changes. Furthermore, the results suggest that the mutation of the charges may affect the sensitivity of UVRR signals as indicated by the spectral intensity variations of WTFab compared to the mutants.

7 General Discussion and Future Work

The purpose of this research was to develop the application of Raman spectroscopy and 2D-correlation analysis to aid candidate selection and optimization during protein conformational stability screening, applicable in early drug development. By using different 2DCA tools (autocorrelation, synchronous and asynchronous plots as well as 2DPCMW synchronous and asynchronous plots) in combination with the acquired Raman spectra; it has been possible to obtain extensive information about the observed spectral changes to support the choice of the most stable protein(s) and/or conditions. The first two experimental chapters focused on applying the technique to model proteins. In subsequent chapters, the technique was applied to therapeutically relevant samples, and UVRR spectroscopy was used instead of conventional Raman spectroscopy in combination with 2DCA.

7.1 Model Protein Samples

Four model proteins were examined in pairs: α_s-casein; dα_s-casein and apohTf; holohTf. The rationale for the pairing was to investigate and compare the conformational stability of samples which have the same basic structure but with a known structural difference. By subjecting them to similar conditions, any observed differences in their stability behaviour may then be attributed to their distinct structural differences. This experimental design created a platform to investigate the capacity of the technique to aid the selection of a stable protein

candidate from amongst variants. Stability testing was carried out by perturbing the samples using pH and/or temperature. Due to time constraints, other perturbations such as storage and agitation were not applied as initially intended. Sample pH was conveniently adjusted over an extensive range with either NaOH or HCl. Significant spectral changes were visually observed at extremely acidic values for the hTf pair and alkaline pH values for the α_s-casein pair. In contrast to pH studies, the temperature range was limited because it was not possible to attain temperatures beyond 65 °C. Nevertheless, some distinct spectral changes were also observed in the hTf temperature study.

The application of 2DCA to the Raman spectra uniquely improved the elucidation of both subtle and visually evident spectral changes; allowing for the identification of transition points and the examination of limited regions or spectral features. Specifically, spectral changes indicative of peaks shifts and broadening have been clarified, thus enhancing the robustness of the overall spectral interpretation.

7.1.1 Transition Points

The identification of transition points along the perturbation axis of the synchronous 2DPCMW plots was invaluable in analysing the data. With the α_s-casein pair of samples, the transition points provided markers for differences in stability between the phosphorylated and dephosphorylated forms (Figure 3.5 and 3.9). Secondary structure changes were observed to be less prominent in the presence of the phosphate residues from ~pH 6-9 than at 10-12. The limited unfolding of α_s-casein at an earlier pH compared to dα_s-casein has been attributed to the 'shielding' by the deprotonating dibasic phosphate groups on the serine side chain. Although the transition points in the case of the hTf samples could be better

identified from the spectra compared to the α_s-casein samples, the 2D synchronous 2DPCMW plots offered a condensed and improved visualisation of these transition points. HolohTf and apohTf showed no significant difference in stability from pH 6-9. However, only holohTf showed an increase in β-sheet at ~pH 9-11 possibly due to increased exposure of the backbone β-strands in the iron-binding site prior to iron dissociation.

7.1.2 Elucidation of Peak Shifts.

With the α_s-casein model (chapter 3), the initially proposed peak shift from ~1615 to 1600 cm^{-1} from pH 10-12 was shown using synchronous and asynchronous plots to be due to two overlapping peaks with varying intensities. This understanding is notable as it proposes that the marker for deprotonation is not a peak shift as previously thought, but a drastic loss of intensity of the ~1615 cm^{-1} peak at the expense of the ~1600 cm^{-1} peak.

The use of these plots also led to a similar observation with holohTf (chapter 4), where the supposed peak shift at ~1660 cm^{-1} was shown to arise from variations in the intensities of overlapping peaks at ~1655 and 1683 cm^{-1} suggesting an additional conformational change at ~pH 2.6 linked with the removal of iron. These findings demonstrate the benefit of 2DCA, even where the interpretation of the spectral change appears evident from visual inspection.

7.1.3 Preprocessing Specifications

By far the most time-consuming aspect of this research was deciding on the specific parameters for each preprocessing method (baseline correction, smoothing, normalisation and solvent subtraction) as well as the sequence of application to the Raman spectra. The choices were decided by trialling different

sequences and parameters across the sample data sets. Kevin and Ryder (2017) previously pointed out the cumbersome nature of preprocessing the Raman spectra of biologic samples and emphasised the need to exercise restraint in attempting to try out several algorithms, but to instead focus on a few tested and tried processes.

Preprocessing was not only necessary for reducing unwanted effects in the Raman spectra but also in the 2DCA plots. The importance of data treatment for obtaining reliable 2DCA plots has been extensively reviewed (Noda, 2004, Noda, 2014). For generalised 2D-correlation and 2DMW plots generated from the Raman spectra of model protein samples, preprocessing steps in the order of baseline correction, normalisation, smoothing and further baseline correction has been demonstrated as being adequate (Ashton et al., 2007, Ashton et al., 2006, Ashton et al., 2008). The sequence used in this study retained the order of the initial and further baseline correction. However, smoothing was carried out before normalisation and the performance of solvent subtraction before the final baseline correction (section 2.3). Once the parameters and sequence were chosen, the methods were applied across the spectral sets, with necessary adjustments as in the case of performing a quartz spectrum subtraction after baseline correction for temperature-dependent spectra, in addition to the other preprocessing methods.

7.2 Therapeutically Relevant Samples

Five antibody fragment (Fab) samples including a wildtype (WTFab) and four surface charge mutants (MT1-1, MT2-1, MT3+1 and MT4+1) were obtained from CIMEMT. Again, the samples had the same basic structure except for the different

charge mutations in the samples. In these studies, the pH-perturbation range was much narrower and near the neutral region and so compared to the model proteins, the spectral changes were less evident visually. The choice of the perturbation range was informed by suggestions from CIMEMT based on previous work using other techniques.

7.2.1 Buffer Effects

In addition to a narrow pH range, another difference between the study of the therapeutically relevant samples and that of the model proteins was the method of pH control. Three different buffers (sodium acetate, sodium citrate and sodium phosphate) were used to control pH. The reason for the number was partly due to the buffering capacity of the specific buffers, as well as the need to investigate and compare the buffer dependent effects at pH 4.5 and 5.5 on the Fab stability.

The use of these buffers introduced additional spectral complexity compared to the pH control using acid and base solutions. Each buffer had at least three separate Raman peaks in the same region ($\sim$800-1500 cm^{-1}) as the protein peaks. Furthermore, at pH 4.5 and 5.5, the spectra of sodium acetate and sodium citrate showed pH-dependent peak intensity variations.

The buffer subtraction process proved challenging and was at its best partial. Sodium phosphate retained two (875 and 1077 cm^{-1}) out of the three peaks while acetate and citrate buffer retained one peak (927 and 1415 cm^{-1} respectively) out of four peaks at pH 5.5 (section 5.3.2). Thus, the subtraction of the different buffer spectra from their associated Fab spectra did not produce a uniform outcome, which is imperative for an objective comparison of the spectra. While it may have

been possible to try out buffer subtraction using other levels of buffer dilution, the different response at this dilution implies that the issue is not necessarily with the dilution but with the intrinsic nature of the buffer peaks. The variations in the buffer subtraction process were addressed by performing water subtraction instead. The retention of buffer peaks in the water subtracted Fab spectra potentially provided a view of the presence of the buffer ions with the protein in solution and not just the effect of the buffer protein interaction; as would be applicable with buffer subtracted Fab spectra where buffer peaks are removed. Differences between the interaction of citrate and acetate ions were observed in chapter 5, where WTFab showed a decreased peak intensity at ~1337 cm^{-1} (pH 4.5 and 5.5) in citrate buffer but not in acetate buffer (Figure 5.8). This observation has been attributed to a more extensive interaction of citrate ions with α-helix-side chains due to their preferential accumulation on the WTFab surface.

Furthermore, all the mutant samples exhibited similar stability behaviour to WTFab in citrate buffer (pH 4.5 and 5.5) but not in acetate buffer, thus demonstrating buffer dependent interactions. The compromise in retaining the buffer peaks was that buffer overlapped protein peaks could not be unambiguously analysed for pH-induced changes in the Fab. Therefore, the use of 2DCA in chapter 5 was limited due to the identification of only two wholly protein peak intensity variations. Understandably, the issue of the different overlapping buffer peaks interfering with 2DCA analysis will be negligible where the applied perturbation is not pH. For example, where the perturbation is temperature, concentration, agitation or storage, a single buffer at a specific pH can be used. If buffer subtraction is not feasible or partial so that the buffer peaks still overlap the

protein peaks, they will not be expected to change significantly across the spectra and therefore will have a minimal impact on the 2DCA analysis. Nevertheless, the presence of overlapped buffer peaks with protein peaks will still be challenging for spectral interpretation.

7.2.2 Stability Deductions

Notwithstanding the buffer influence, the pH-induced Raman spectral variations revealed that WTFab and the four mutant samples exhibited a loss of β-sheet and a change in α-helix orientation at pH 4.5 and 5.5 (citrate buffer) compared to neutral pH (water and phosphate buffer). Only WTFab and MT4+1 showed similar pH-dependent behaviour in acetate buffer (pH 4.5 and 5.5). This observation has been attributed to the similarity in the profiles of their ionized carboxyl groups as well as their solvent accessible area. Despite having different charge mutations, MT3+1 and MT1-1 were both stable in acetate buffer at pH 4.5 and 5.5, whereas MT2-1 was only stable at ~pH 5.5. Since all three samples displayed no difference in their ionized carboxyl profile, MT2-1 may also be stable at ~pH 4.5, as shown from the reconstituted freeze-dried sample spectra. Therefore, the peak at ~1415 cm^{-1} may be used as a predictor of buffer-dependent behaviour.

Apart from MT1-1, all the other Fab samples showed no increase in β-sheet in phosphate buffer, and this has been linked with its increasing tendency to aggregate compared to the other mutants (Ahmad, 2011). Using the Phe peak (~1003 cm^{-1}) in the difference spectra, MT1-1 and MT2-1 were predicted to have an increased tendency to aggregate compared to MT3+1 and MT4+1 (Figure 5.14). The distinct behaviour of MT1-1 in phosphate buffer may also be signalling this tendency. A direct comparison of the freeze-dried sample spectra with the solution

spectra was challenging due to the differences in the peak shapes. However, the peak shapes in the spectra of the reconstituted samples were analogous with the solution spectra. The stability of freeze-dried MT2-1 when reconstituted was not significantly different from the MT2-1 solution, except in phosphate buffer where an increase in β-sheet was observed. MT3+1 also showed a marked difference in stability between solution and freeze-dried reconstituted forms in phosphate buffer and acetate buffer linked to α-helix unfolding and β-sheet aggregation respectively.

7.3 UVRR spectroscopy

UVRR spectroscopy, a form of Raman spectroscopy was used in chapter 6 to address the issue of buffer interference since weak UVRR vibrational modes are typically observed for buffers. In addition to overcoming the buffer interference, there was also a considerable reduction of the solution spectra collection time from 30 min per spectra in the Raman spectroscopic studies to ~2 mins in the UVRR study. The time savings offered by UVRR spectroscopy is accompanied with limited but specific information based on the wavelength applied. Furthermore, due to the reduced background interference, the preprocessing methods applied to the UVRR spectra were minimal and thus determining the sequence of their application was less time consuming compared to the Raman spectra.

7.3.1 Transition Points.

In contrast to the Raman spectroscopic studies where it was adequate to use the synchronous 2DPCMW plots to identify transitions points, the fragmented pattern of the contours necessitated the use of asynchronous 2DPCMW plots to identify

the transition points in the data. The corresponding pH values were ~5.2, 6.3, 5.5 and 5.9 for WTFab, MT1-1, MT2-1 and MT3+1, respectively. Interestingly, none was identified for MT4+1 and this is possibly explained by it having the highest solvent accessible area (SAS) compared to the other samples such that its spectral changes are less marked compared to one with a lower SAS. It was also notable that the transition pH values ranged between ~5-6 coinciding with the previous observations of increased stability above pH 5.5 using Raman spectroscopy (chapter 5).

This consistency of the UVRR spectroscopy results with the Raman spectroscopy results despite the dominance of the UVRR spectra by aromatic residues demonstrates the unique contribution of 2DCA. It is interesting to note that not only did the transition points link with increased stability, but they were also markers for changes in peak shift direction (WTFab), peak intensity direction (MT3+1) and autopeak positions (MT2-1).

7.3.2 Elucidation of Peak Shifts

Unlike the model protein studies were dramatic spectral changes led to an initial proposal of peak shifts, the spectral changes in the UVRR spectra could not be detected visually and so were only considered following the application of 2DCA. The presence of peak shifts was detected using synchronous and asynchronous plots generated from the groups of spectra as defined by the respective transition points. The difficulty of using the generalised 2D-correlation asynchronous plots to distinguish between a single shifting peak and a peak shift due to overlapping peak intensities in real-life spectra has been extensively discussed (Noda, 2014a,

Noda, 2012). Indeed, there has been the development of PCA based methods which can verify the exact nature of the peak shifts (Ryu et al., 2010, Ryu and Jung, 2011).

In many cases where such methods have been applied to temperature and concentration induced changes in IR experimental data, the observed peak shifts have been reported as being due to changes in overlapping peak intensities (Noda, 2012, Ryu et al., 2011). However, when such methods are applied to simulated spectra with a single peak shift, the results are consistent with the characteristic asynchronous contour patterns for peak shifts (Ryu et al., 2010, Ryu et al., 2011). Hence, it has been recommended that these PCA based methods be used in addition to synchronous and asynchronous plots to clarify if a peak shift change is arising from overlapped peaks with intensity changes (Ryu and Jung, 2011, Noda, 2014a). As discussed for the Fab samples (Table 6.3), there are overlapping Tyr peaks in the examined region ($\sim$1550-1620 cm^{-1}). No peak shifts were indicated for MT1-1 and MT3+1. The contours for MT1-1 were weak while the contour patterns for MT3+1 clearly indicated overlapping peak intensity changes.

In contrast, the asynchronous *butterfly* contour pattern for MT2-1 signalled a peak shift. In addition, the two sets of autopeaks with different positions in each of the spectra groups (defined by the transition point) was attributed to the presence of overlapping peak shifts. The asynchronous *butterfly* contour pattern in the WTFab plot appeared to be influenced by spectral broadening while MT4+1 had a characteristic butterfly pattern attributable to a peak shift.

While it was not possible to delineate the nature of the peak shift using the verification methods, it is worth noting the presence of paired contours with opposite signs in the synchronous 2DPCMW plots of WTFab, MT2-1, MT3+1 and

MT4+1 (Figures 6.6B, 6.11D, 6.15B and 6.15D). These contours correspond to the wavenumbers of Tyr peaks and are indicative of changing peak intensities in opposite directions. Nevertheless, only the asynchronous plots for MT3+1 clearly specifies the presence of overlapping peak intensity changes. In contrast, the other sample plots denote peak shifts thus highlighting a degree of discrimination between the ability of the synchronous 2DPCMW and 2D-synchronous and asynchronous plots to identify peak shifts (Ryu et al., 2011). Furthermore, the clarification of the apparent peak shifts in α_s-casein and holohTf spectra as being due to the presence of overlapping peaks with changing intensities based on the asynchronous contour patterns also demonstrates that peak shifts due to the presence of overlapping peaks with changing intensities can be clearly distinguished (Figures 3.8 and 4.10). Therefore, it is likely that the asynchronous *butterfly* patterns observed in the UVRR Fab spectra are signalling single shifting peaks and thus can be interpreted using the rule described in section 6.3.3.3. Accordingly, peak shift directions in the UVRR Fab spectra were determined and linked with changes in hydrogen bonding. MT2-1, MT4+1 and WTFab demonstrated a decrease in hydrogen bond donation above their transition pH. This observation has been proposed as accounting for their higher stability above pH 5.5. The stability of MT3+1 above pH 5.5, on the other hand, has been linked to a marked change in solvent exposure.

7.4 Future Work

The crucial role of pH in monitoring protein stability and the central place of buffers in the pH control of protein drugs places a demand on the utility of Raman spectroscopy and 2DCA in protein drug development. The complication of protein

spectra by the presence of other components including buffers is a frequently encountered problem even with more routinely used techniques like CD, UV and IR spectroscopies (Houde and Berkowitz, 2014, Pelton and McLean, 2000). Such interferences are usually addressed by measures including solvent subtraction, using low concentrations and restricting the wavelength range (Dong et al., 1995, Pelton and McLean, 2000, Greenfield, 1996). In addition, through prior testing and screening the choice of the sample components can be substituted for others with less interference where possible (Houde and Berkowitz, 2014, Greenfield, 1996, Pelton and McLean, 2000).

Consequently, evaluation and development of similar approaches relevant to Raman spectroscopy will prove useful in embracing its application for protein stability screening. The results have shown that in addition to having Raman peaks in the protein peak regions buffers also influence structural changes. Therefore, it will be worthwhile to establish a baseline in aqueous solutions before the addition of any buffers. In this case, pH control can be initially achieved using acid and base as was shown for the model proteins without the spectral signature of buffers interfering with the protein Raman spectra or 2DCA. UVRR spectroscopy is promising as it offers a way to overcome the buffer influence on the spectra. The choice of wavelength can be informed by the observed structural changes using Raman spectroscopy.

Furthermore, other perturbations relevant to protein stability, including concentration, agitation and storage, can be explored. Although the concentration of the solution samples (~10 mg/ml) was higher than is applicable for most of the currently used techniques; there is scope to exploit the sensitivity of Raman

spectroscopy at even higher protein concentrations to achieve shorter acquisition times. The possibility of acquiring spectra simultaneously at multiple perturbation points will also be less time consuming and attractive for routine use. Finally, it would be advantageous to incorporate some statistical testing with 2DCA so that the statistical significance of any identified perturbation-induced changes can be evaluated.

7.5 Concluding Remarks

This research demonstrates that the combined technique of Raman spectroscopy and 2DCA can facilitate the identification of the most stable protein candidate(s) or conditions. In addition to standard Raman spectral interpretation, the use of 2DCA to identify transitions points and examine spectral regions of interest provided detailed information about complex spectral changes. This utility offers a compelling incentive for proposing their addition to the current toolbox for monitoring protein conformational stability during early drug development. The research emphasizes the central role of preprocessing Raman spectra in the adoption of these techniques. The level of clarity obtainable via preprocessing influences Raman spectra interpretation and ultimately 2DCA. As perturbation is central to the operation of 2DCA, the ideal preprocessing outcome is that the data should only include perturbation-induced spectral changes that are associated with the protein. Where the spectral changes indicate that a non-protein component of the samples varies significantly with the perturbation, and dominates the 2DCA plots, the benefit of combining 2DCA with Raman spectroscopy may not be realised. The same limitation is applicable if the spectral changes are restricted to a certain perturbation point or if they can be adequately

explained by visual inspection alone. In the case of overlapping buffer peaks, UVRR spectroscopy is an alternative technique that can be applied. In summary, the successful application of Raman spectroscopy and 2DCA to conformational stability screening during early drug development will be substantially contingent on the nature of the perturbation-induced spectral variations whether they are competing, complex, or confined.

Appendix A

pH-dependent Raman spectra (~640-1700 cm^{-1}) of Fab mutant samples (MT1-1, MT2-1, MT3+1 and MT4+1) in water and buffers (sodium phosphate (NaP), sodium citrate (Cit) and sodium acetate (Ac).
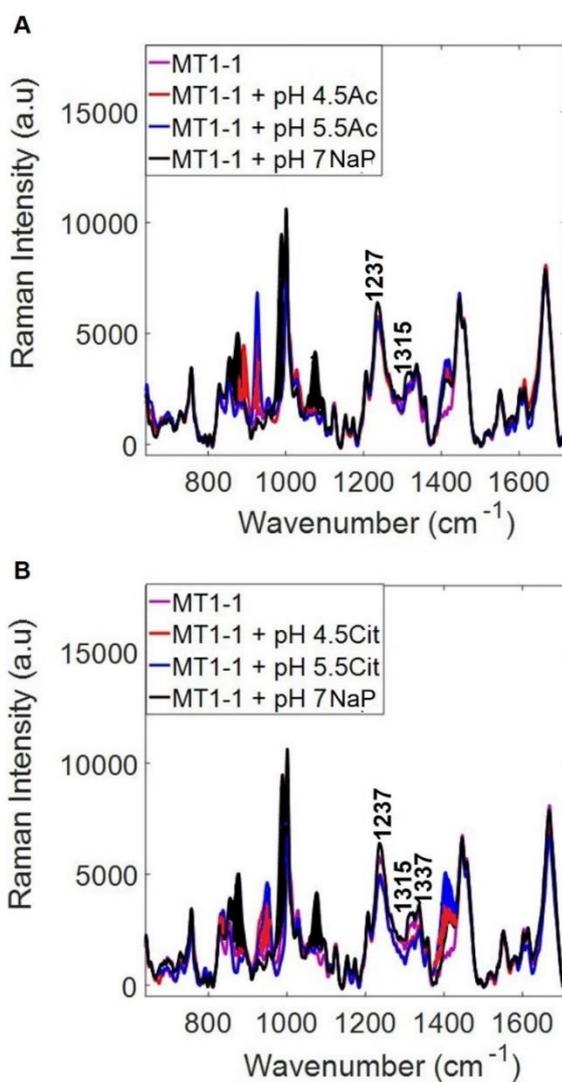
Figure A-1 Averaged (n=3) Raman spectra of 10 mg/ml MT1-1 (A) MT1-1 in water (pH 6.9), Ac (pH 4.5 and 5.5) and NaP (pH 7) (B) MT1-1 in water (pH 6.9), Cit (pH 4.5 and 5.5) and NaP (pH 7). Preprocessing was carried out by baseline correction, smoothing, normalization, water subtraction and further baseline subtraction. Each Raman spectrum was collected in 30 mins using an excitation length of 785 nm. Buffer peaks are shaded and selected protein peaks are labelled. Protein peak assignments are listed in Table 1.1.
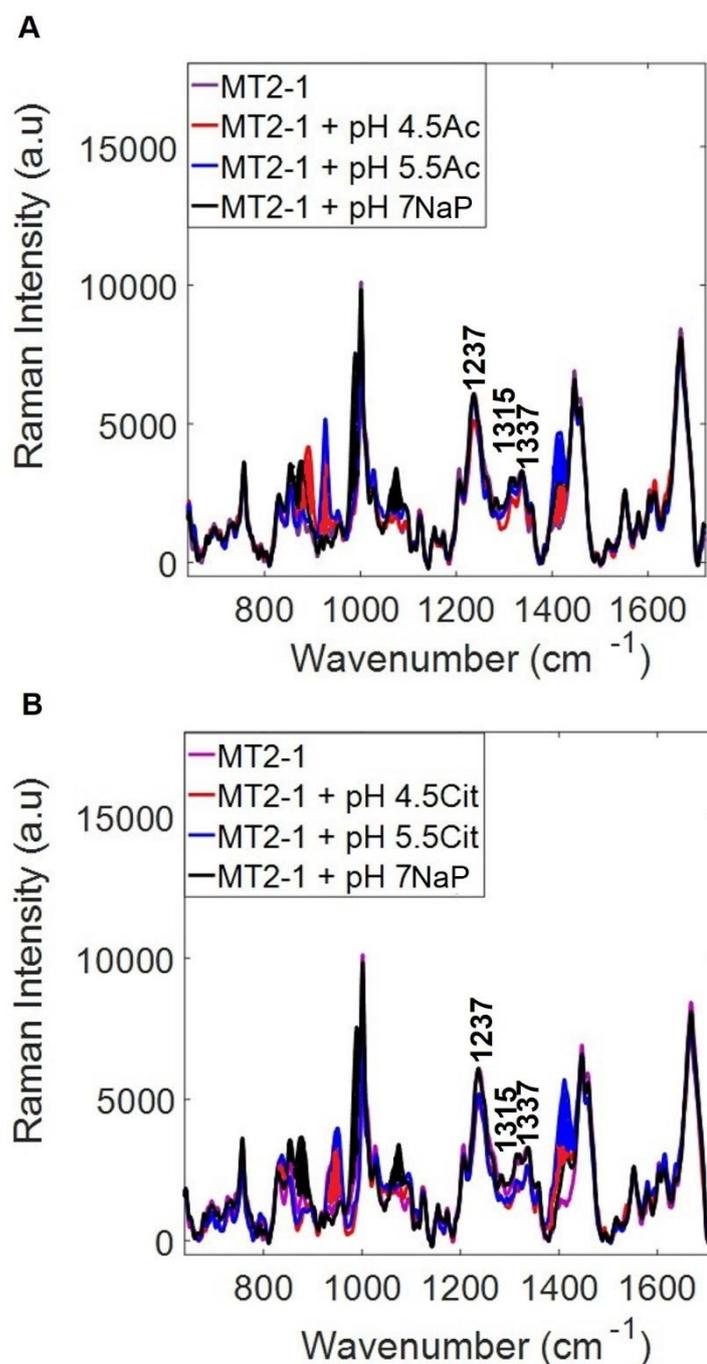
Figure A-2 Averaged (n=3) Raman spectra of 10 mg/ml MT2-1 (A) MT2-1 in water (pH 7.1), Ac (pH 4.5 and 5.5) and NaP (pH 7) (B) MT2-1 in water (pH 7.1), Cit (pH 4.5 and 5.5) and NaP (pH 7). Preprocessing was carried out by baseline correction, smoothing, normalization, water subtraction and further baseline subtraction. Each Raman spectrum was collected in 30 mins using an excitation length of 785 nm. Buffer peaks are shaded and selected protein peaks are labelled. Protein peak assignments are listed in Table 1.1.
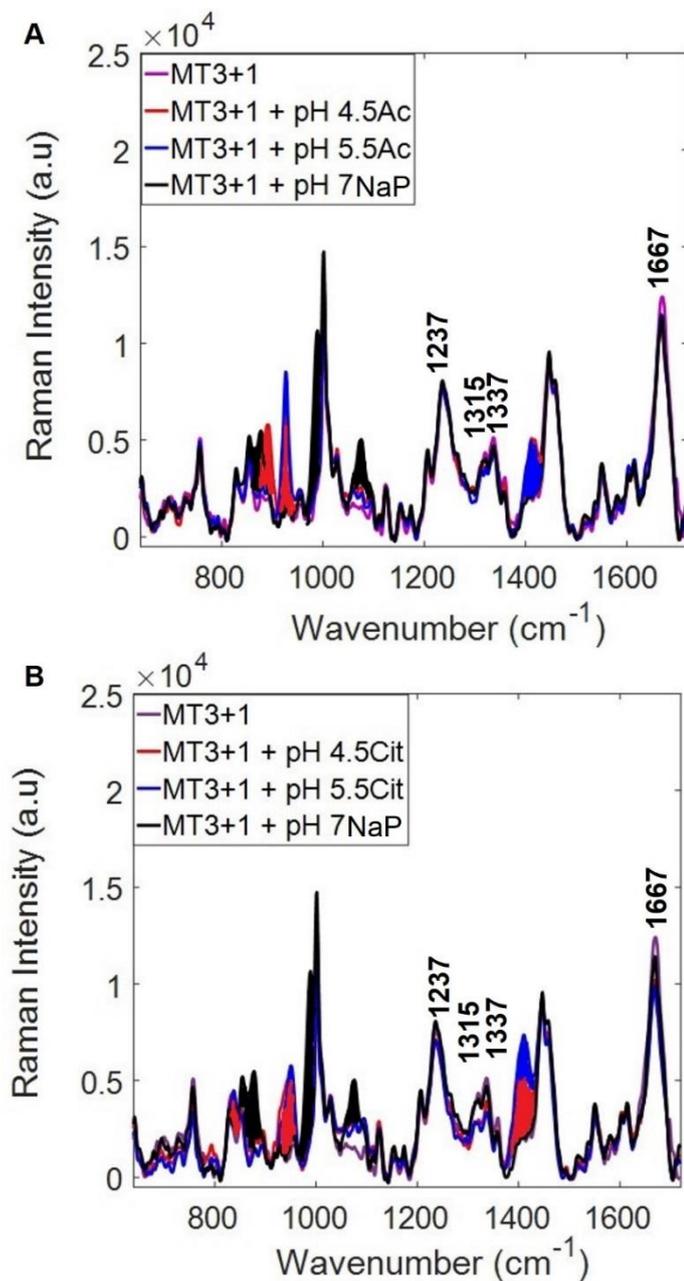
Figure A-3 Averaged (n=3) Raman spectra of 10 mg/ml MT3+1 (A) MT3+1 in water (pH 6.9), Ac (pH 4.5 and 5.5) and NaP (pH 7) (B) MT3+1 in water (pH 6.9), Cit (pH 4.5 and 5.5) and NaP (pH 7). Preprocessing was carried out by baseline correction, smoothing, normalization, water subtraction and further baseline subtraction. Each Raman spectrum was collected in 30 mins using an excitation length of 785 nm. Buffer peaks are shaded and selected protein peaks are labelled. Protein peak assignments are listed in Table 1.1.
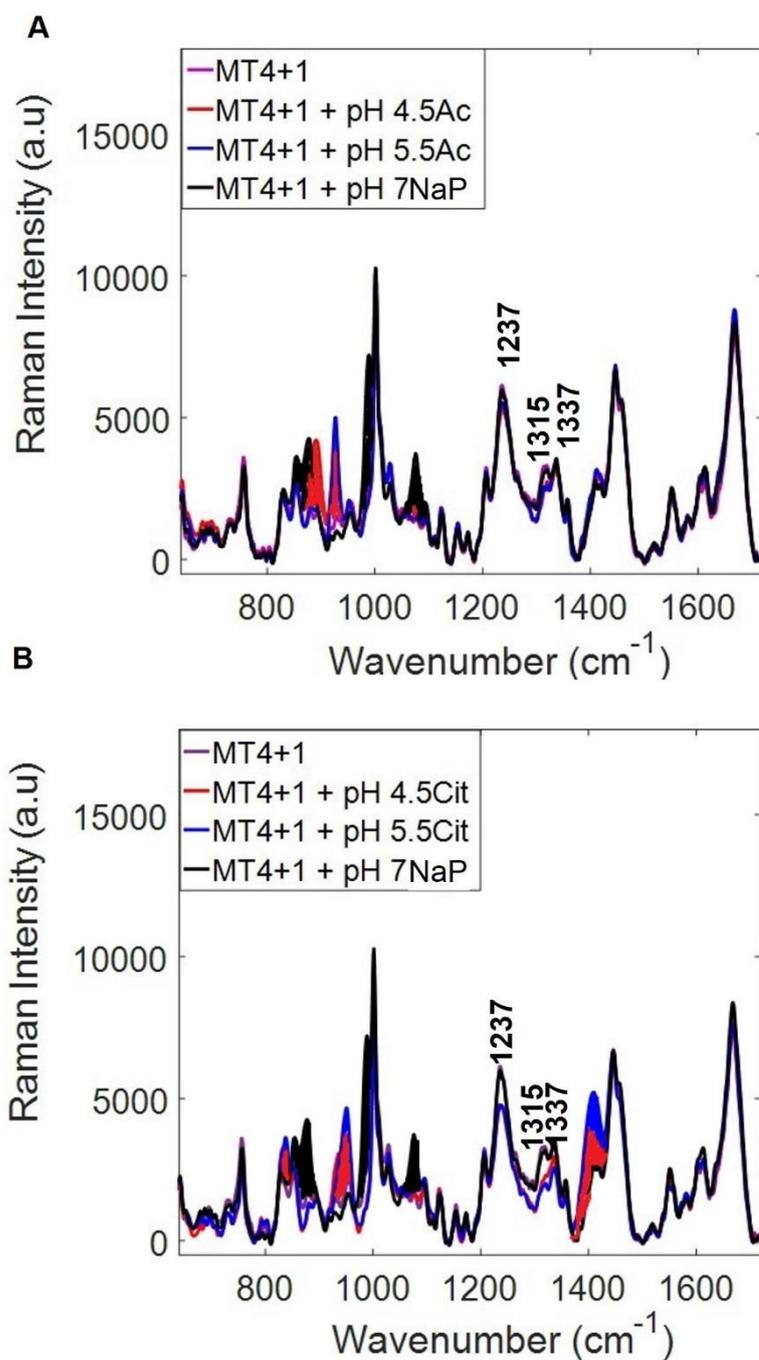
Figure A-4 Averaged (n=3) Raman spectra of 10 mg/ml of MT4+1 (A) MT4+1 in water (pH 6.8), Ac (pH 4.5 and 5.5) and NaP (pH 7) (B) MT4+1 in water (pH 6.8), Cit (pH 4.5 and 5.5) and NaP (pH 7). Preprocessing was carried out by baseline correction, smoothing, normalization, water subtraction and further baseline subtraction. Each Raman spectrum was collected in 30 mins using an excitation length of 785 nm. Buffer peaks are shaded, and selected protein peaks are labelled. Protein peak assignments are listed in Table 1.1.

Appendix B

2DCA of the pH-dependent Raman spectra of Fab mutant samples (MT1-1, MT2-1, MT3+1 and MT4+1) in water, sodium citrate (Cit) and sodium phosphate (NaP) buffers.
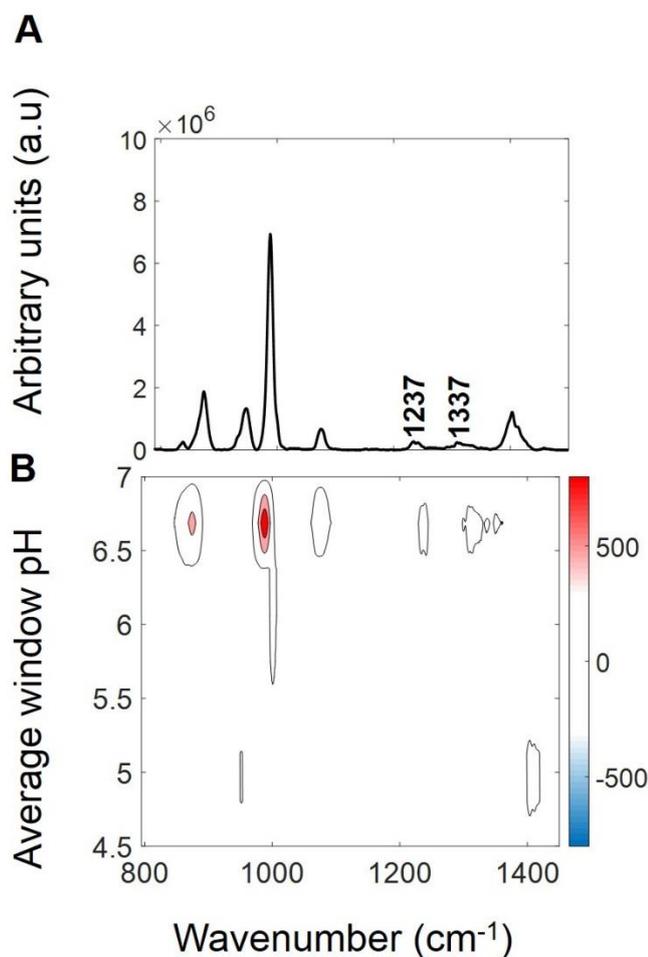
Figure B-1 2DCA of the pH-dependent Raman spectra of MT1-1 in water (pH 6.9), Cit (pH 4.5 and 5.5) and NaP (pH 7). (A) Autocorrelation (B) Synchronous 2DPCMW plot. Protein autopeaks are labelled in black. Positive contours are shaded in red while negative contours are shaded in blue. Positive contours indicate increasing peak intensities in the direction of the perturbation while negative contours indicate decreasing peak intensities in relation to the perturbation. A darker shade and an increased number of contours indicate a relatively greater change in intensity. The scale on the colour bar is arbitrary. A moving window size of 3 and a maximum number of 3 contours was applied.
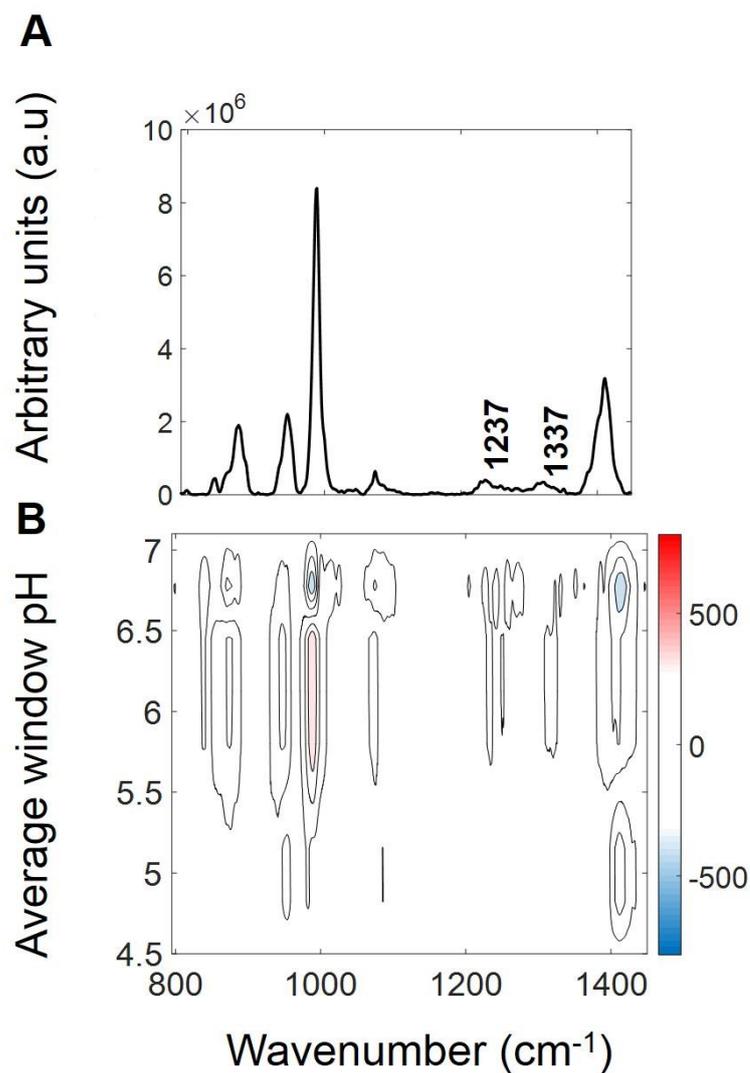
Figure B-2 2DCA of the pH-dependent Raman spectra of MT2-1 in water (pH 7.1), Cit (pH 4.5 and 5.5) and NaP (pH 7). (A) Autocorrelation (B) Synchronous 2DPCMW plot. Protein autopeaks are labelled in black. Positive contours are shaded in red while negative contours are shaded in blue. Positive contours indicate increasing peak intensities in the direction of the perturbation while negative contours indicate decreasing peak intensities in relation to the perturbation. A darker shade and an increased number of contours indicate a relatively greater change in intensity. The scale on the colour bar is arbitrary. A moving window size of 3 and a maximum number of 6 contours was applied.
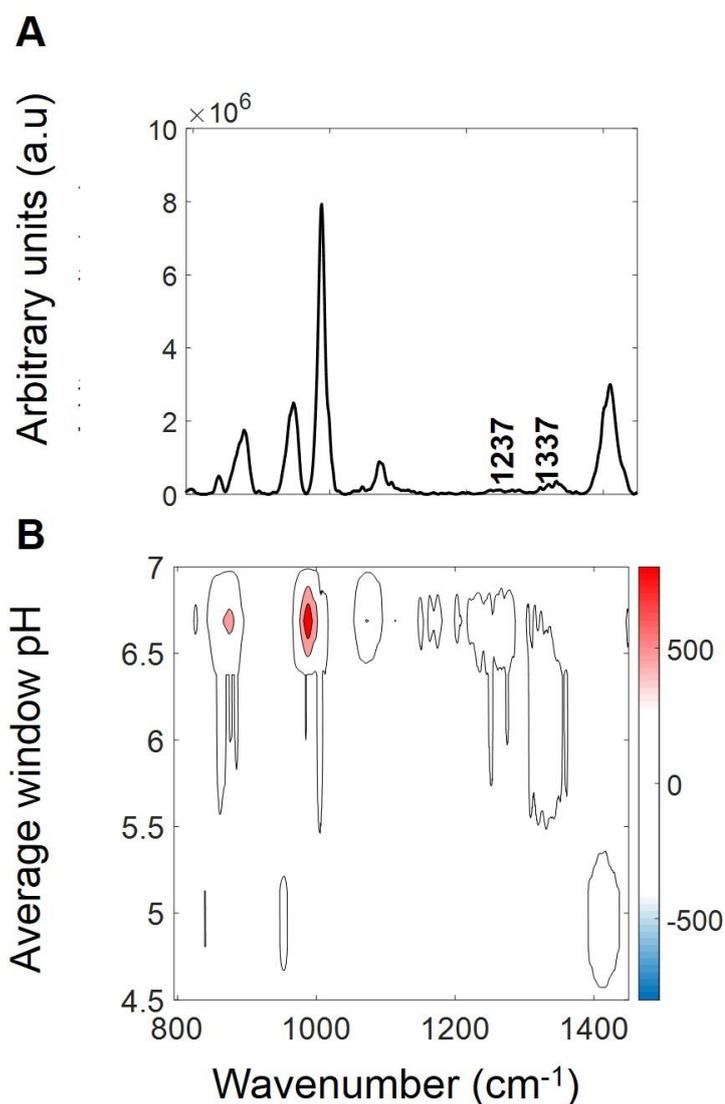
Figure B-3 2DCA of the pH-dependent Raman spectra of in MT3+1 in water (pH 6.9), Cit (pH 4.5 and 5.5) and NaP (pH 7). (A) Autocorrelation (B) synchronous 2DPCMW plot. Protein autopeaks are labelled in black. Positive contours are shaded in red while negative contours are shaded in blue. Positive contours indicate increasing peak intensities in the direction of the perturbation while negative contours indicate decreasing peak intensities in relation to the perturbation. A darker shade and an increased number of contours indicate a relatively greater change in intensity. The scale on the colour bar is arbitrary. A moving window size of 3 and a maximum number of 3 contours was applied.
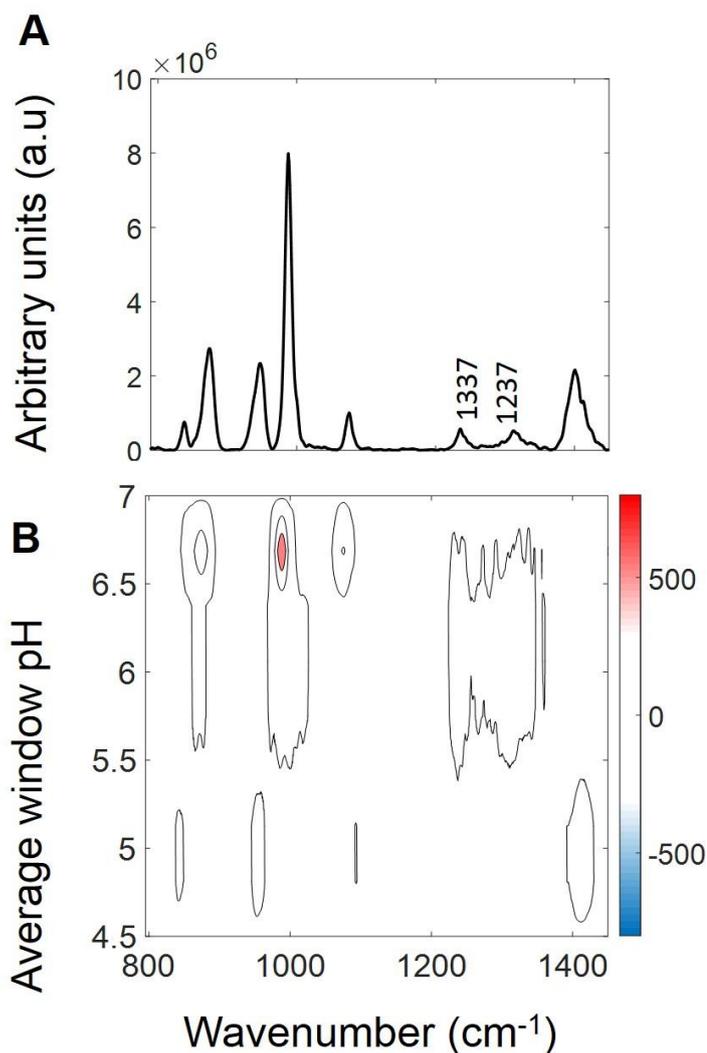
Figure B-4 2DCA of the pH-dependent Raman spectra of MT4+1 in water (pH 6.8), Cit (pH 4.5 and 5.5) and NaP (pH 7). (A) Autocorrelation (B) synchronous 2DPCMW plot. Protein autopeaks are labelled in black. Positive contours are shaded in red while negative contours are shaded in blue. Positive contours indicate increasing peak intensities in the direction of the perturbation while negative contours indicate decreasing peak intensities in relation to the perturbation. A darker shade and an increased number of contours indicate a relatively greater change in intensity. The scale on the colour bar is arbitrary. A moving window size of 3 and a maximum number of 4 contours was applied.

8 References

AGYEI, D., AHMED, I., AKRAM, Z., IQBAL, H. M. N. & DANQUAH, M. K. 2017. Protein and Peptide Biopharmaceuticals: An Overview. *Protein and Peptide Letters,* 24**,** 94-101.

AHMAD, S. 2011. *Analysis of the Precipitation and Aggregation of Engineered Proteins.* Doctor of Philosophy, University College London.

ALADUL, M. I., FITZPATRICK, R. W. & CHAPMAN, S. R. 2017. Impact of Infliximab and Etanercept Biosimilars on Biological Disease-Modifying Antirheumatic Drugs Utilisation and NHS Budget in the UK. *BioDrugs,* 31**,** 533-544.

ALAIMO, M. H., FARRELL, H. M. & GERMANN, M. W. 1999. Conformational Analysis of the Hydrophobic Peptide αs_1-Casein (136–196). *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology,* 1431**,** 410-420.

ALATA, I., BROQUIER, M., DEDONDER, C., JOUVET, C. & MARCECA, E. 2012. Electronic Excited States of Protonated Aromatic Molecules: Protonated Fluorene. *Chemical Physics,* 393**,** 25-31.

ALDEN, R. G., ONDRIAS, M. R., COURTNEY, S., FINDSEN, E. W. & FRIEDMAN, J. M. 1990. Power-induced Broadening of Transient Resonance Raman Spectra of Deoxyhemoglobin. *The Journal of Physical Chemistry,* 94**,** 85-90.

AMANI, S. & NAEEM, A. 2014. Transition of Transferrin from Native to Fibrillar State: An Implication for Amyloid-linked Diseases. *Biochemical Engineering Journal,* 91**,** 120-128.

ANDERSEN, B. F., BAKER, H. M., MORRIS, G. E., RUMBALL, S. V. & BAKER, E. N. 1990. Apolactoferrin Structure Demonstrates Ligand-induced Conformational Change in Transferrins. *Nature,* 344**,** 784-787.

ARZHANTSEV, S., VILKER, V. & KAUFFMAN, J. 2012. Deep-Ultraviolet (UV) Resonance Raman Spectroscopy as a Tool for Quality Control of Formulated Therapeutic Proteins. *Applied Spectroscopy,* 66**,** 1262-1268.

ASHER, S. A. 1993. UV Resonance Raman Spectroscopy for Analytical, Physical, and Biophysical Chemistry. Part 1. *Analytical Chemistry,* 65**,** 59A-66A.

ASHER, S. A. 2006. Ultraviolet Raman Spectrometry. *In:* CHALMERS, M., GRIFFITHS, P.R. (ed.) *Handbook of Vibrational Spectroscopy.* Chichester: John Wiley & Sons.

ASHER, S. A. & JOHNSON, C. R. 1984. Raman Spectroscopy of a Coal Liquid shows that Fluorescence Interference is Minimized with Ultraviolet Excitation. *Science,* 225**,** 311-313.

ASHER, S. A., LUDWIG, M. & JOHNSON, C. R. 1986. UV Resonance Raman Excitation Profiles of the Aromatic Amino acids. *Journal of the American Chemical Society,* 108**,** 3186-3197.

ASHTON, L., BARRON, L. D., CZARNIK-MATUSEWICZ, B., HECHT, L., HYDE, J. & BLANCH, E. W. 2006. Two-Dimensional Correlation Analysis of Raman Optical Activity Data on the α-helix-to-β-sheet Transition in Poly(L-lysine). *Molecular Physics,* 104**,** 1429-1445.

ASHTON, L., BARRON, L. D., HECHT, L., HYDE, J. & BLANCH, E. W. 2007. Two-Dimensional Raman and Raman Optical Activity Correlation Analysis of the α-helix-to-disordered Transition in Poly(L-glutamic acid). *Analyst,* 132**,** 468-479.

ASHTON, L. & BLANCH, E. W. 2010. pH-induced Conformational Transitions in α-lactalbumin Investigated with Two-Dimensional Raman Correlation Variance Plots and Moving Windows. *Journal of Molecular Structure,* 974**,** 132-138.

ASHTON, L., BREWSTER, V. L., CORREA, E. & GOODACRE, R. 2017. Detection of Glycosylation and Iron-binding Protein Modifications using Raman Spectroscopy. *Analyst,* 142**,** 808-814.

ASHTON, L., HOBRO, A., CONN, G. L., ROUHI, M. & BLANCH, E. W. 2008. Investigations of Conformational Transitions in Proteins and RNA Using 2DCOS Raman and 2DCOS Raman Optical Activity Spectroscopies. *Journal of Molecular Structure,* 883–884**,** 187-194.

ASHTON, L., JOHANNESSEN, C. & GOODACRE, R. 2011. The Importance of Protonation in the Investigation of Protein Phosphorylation using Raman Spectroscopy and Raman Optical Activity. *Analytical Chemistry,* 83**,** 7978-7983.

ASHTON, L., PUDNEY, P. D. A., BLANCH, E. W. & YAKUBOV, G. E. 2013a. Understanding Glycoprotein Behaviours Using Raman and Raman Optical Activity Spectroscopies: Characterising the Entanglement Induced Conformational Changes in Oligosaccharide Chains of Mucin. *Advances in Colloid and Interface Science,* 199-200**,** 66-77.

ASHTON, L., XU, Y., BREWSTER, V. L., COWCHER, D. P., SELLICK, C. A., DICKSON, A. J., STEPHENS, G. M. & GOODACRE, R. 2013b. The Challenge of Applying Raman Spectroscopy to Monitor Recombinant Antibody Production. *Analyst,* 138**,** 6977-6985.

AZARI, P. R. & FEENEY, R. E. 1958. Resistance of Metal Complexes of Conalbumin and Transferrin to Proteolysis and to Thermal Denaturation. *Journal of Biological Chemistry,* 232**,** 293-302.

BAKER, E. N. 1994. Structure and Reactivity of Transferrins. *In:* SYKES, A. G. (ed.) *Advances in Inorganic Chemistry.* San Diego: Academic Press.

BAKER, E. N., ANDERSON, B. F., BAKER, H. M., HARIDAS, M., NORRIS, G. E., RUMBALL, S. V. & SMITH, C. A. 1990. Metal and Anion Binding Sites in Lactoferrin and related Proteins. *Pure and Applied Chemistry,* 62**,** 1067-1070.

BAKER, E. N. & HUBBARD, R. E. 1984. Hydrogen Bonding in Globular Proteins. *Progress in Biophysics and Molecular Biology,* 44**,** 97-179.

BAKER, H. M., ANDERSON, B. F. & BAKER, E. N. 2003. Dealing with Iron: Common Structural Principles in Proteins that Transport Iron and Heme. *Proceedings of the National Academy of Sciences,* 100**,** 3579-3583.

BAKER, H. M. & BAKER, E. N. 2004. Lactoferrin and Iron: Structural and Dynamic Aspects of Binding and Release. *Biometals,* 17**,** 209-216.

BARANSKA, M., SCHULZ, H., JOUBERT, E. & MANLEY, M. 2006. In situ Flavonoid Analysis by FT-Raman Spectroscopy: Identification, Distribution, and Quantification of Aspalathin in Green Rooibos (Aspalathus linearis). *Analytical Chemistry,* 78**,** 7716-7721.

BARNES, R. J., DHANOA, M. S. & LISTER, S. J. 1989. Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra. *Applied Spectroscopy,* 43**,** 772-777.

BARNETT, G. V., RAZINKOV, V. I., KERWIN, B. A., LAUE, T. M., WOODKA, A. H., BUTLER, P. D., PEREVOZCHIKOVA, T. & ROBERTS, C. J. 2015. Specific-Ion Effects on the Aggregation Mechanisms and Protein–Protein Interactions for Anti-streptavidin Immunoglobulin Gamma-1. *The Journal of Physical Chemistry B,* 119**,** 5793-5804.

BATES, A. & POWER, C. A. 2019. David vs. Goliath: The Structure, Function, and Clinical Prospects of Antibody Fragments. *Antibodies (Basel, Switzerland),* 8**,** 28.

BEATTIE, J. R., BELL, S. E. J., BORGGAARD, C. & MOSS, B. W. 2008. Preliminary Investigations on the Effects of Ageing and Cooking on the Raman Spectra of Porcine Longissimus Dorsi. *Meat Science,* 80**,** 1205-1211.

BECK, A., WURCH, T., BAILLY, C. & CORVAIA, N. 2010. Strategies and Challenges for the Next Generation of Therapeutic Antibodies. *Nature Reviews Immunology,* 10**,** 345-352.

BERKOWITZ, S. A., ENGEN, J. R., MAZZEO, J. R. & JONES, G. B. 2012. Analytical Tools for Characterizing Biopharmaceuticals and the Implications for Biosimilars. *Nature Reviews Drug Discovery,* 11**,** 527-540.

BOTSTEIN, D. & SHORTLE, D. 1985. Strategies and Applications of In Vitro Mutagenesis. *Science,* 229**,** 1193.

BREWSTER, V. L., ASHTON, L. & GOODACRE, R. 2011. Monitoring the Glycosylation Status of Proteins Using Raman Spectroscopy. *Analytical Chemistry,* 83**,** 6074-6081.

BREWSTER, V. L., ASHTON, L. & GOODACRE, R. 2013. Monitoring Guanidinium-Induced Structural Changes in Ribonuclease Proteins Using Raman Spectroscopy and 2D Correlation Analysis. *Analytical Chemistry,* 85**,** 3570-3575.

BUCKLEY, K. & RYDER, A. G. 2017. Applications of Raman Spectroscopy in Biopharmaceutical Manufacturing: A Short Review. *Applied Spectroscopy,* 71**,** 1085-1116.

BUSS, N. A. P. S., HENDERSON, S. J., MCFARLANE, M., SHENTON, J. M. & DE HAAN, L. 2012. Monoclonal Antibody Therapeutics: History and Future. *Current Opinion in Pharmacology,* 12**,** 615-622.

BYLER, D. M., FARRELL, H. M. & SUSI, H. 1988. Raman Spectroscopic Study of Casein Structure. *Journal of Dairy Science,* 71**,** 2622-2629.

CHADWICK, C. T., WILLITSFORD, A. H., PHILBRICK, C. R. & HALLEN, H. D. 2015. Deep Ultraviolet Raman Spectroscopy: A Resonance-absorption Trade-off Illustrated by Diluted Liquid Benzene. *Journal of Applied Physics,* 118**,** 243101.

CHAKRABORTY, A. & BASAK, S. 2007. pH-Induced Structural Transitions of Caseins. *Journal of Photochemistry and Photobiology B: Biology,* 87**,** 191-199.

CHAKROUN, N., HILTON, D., AHMAD, S. S., PLATT, G. W. & DALBY, P. A. 2016. Mapping the Aggregation Kinetics of a Therapeutic Antibody Fragment. *Molecular Pharmaceutics,* 13**,** 307-319.

CHALMERS, J. M. & DENT, G. 1997. Introduction, Basic Theory, and Principles. *In:* BARNETT, N. W. (ed.) *Industrial Analysis with Vibrational Spectroscopy [Online].* Cambridge: The Royal Society of Chemistry. Available at: https://pubs-rsc-org.ezproxy.lancs.ac.uk/en/content/ebook/978-0-85404-565-5.

CHAN, C.-H., WILBANKS, C. C., MAKHATADZE, G. I. & WONG, K.-B. 2012. Electrostatic Contribution of Surface Charge Residues to the Stability of a Thermophilic Protein: Benchmarking Experimental and Predicted pKa Values. *PLOS ONE,* 7**,** e30296.

CHANG, B. S. & HERSHENSON, S. 2002. Practical Approaches to Protein Formulation Development. *In:* CARPENTER, J. F. & MANNING, M. C. (eds.) *Rational Design of Stable Protein Formulations: Theory and Practice.* Boston, MA: Springer US.

CHEN, M., LORD, R. & MENDELSOHN, R. 1973. Laser-excited Raman Spectroscopy of Biomolecules: IV. Thermal Denaturation of Aqueous Lysozyme. *Biochimica et Biophysica Acta (BBA)-Protein Structure,* 328**,** 252-260.

CHEN, M. C. & LORD, R. C. 1974. Laser-Excited Raman Spectroscopy of Biomolecules. VI. Polypeptides as Conformational Models. *Journal of the American Chemical Society,* 96**,** 4750-4752.

CHEN, M. C. & LORD, R. C. 1976. Laser Raman Spectroscopic Studies of the Thermal Unfolding of Ribonuclease A. *Biochemistry,* 15**,** 1889-1897.

CHI, Z. & ASHER, S. A. 1998. UV Raman Determination of the Environment and Solvent Exposure of Tyr and Trp Residues. *The Journal of Physical Chemistry B,* 102**,** 9595-9602.

CHI, Z., CHEN, X. G., HOLTZ, J. S. W. & ASHER, S. A. 1998. UV Resonance Raman-Selective Amide Vibrational Enhancement: Quantitative Methodology for Determining Protein Secondary Structure. *Biochemistry,* 37**,** 2854-2864.

CHURCH, J. S., CORINO, G. L. & WOODHEAD, A. L. 1998. The Effects of Stretching on Wool Fibres as Monitored by FT-Raman Spectroscopy. *Journal of Molecular Structure,* 440**,** 15-23.

CODINA, N., HILTON, D., ZHANG, C., CHAKROUN, N., AHMAD, S. S., PERKINS, S. J. & DALBY, P. A. 2019a. An Expanded Conformation of an Antibody Fab Region by X-Ray Scattering, Molecular Dynamics, and smFRET Identifies an Aggregation Mechanism. *Journal of Molecular Biology,* 431**,** 1409-1425.

CODINA, N., ZHANG, C., CHAKROUN, N. & DALBY, P. A. 2019b. Insights into the Stability of a Therapeutic Antibody Fab Fragment by Molecular Dynamics and its Stabilization by Computational Design. *BioRxiv***,** 644369.

COLTART, D. M., ROYYURU, A. K., WILLIAMS, L. J., GLUNZ, P. W., SAMES, D., KUDUK, S. D., SCHWARZ, J. B., CHEN, X.-T., DANISHEFSKY, S. J. & LIVE, D. H. 2002. Principles of Mucin Architecture: Structural Studies on Synthetic Glycopeptides Bearing Clustered Mono-, Di-, Tri-, and Hexasaccharide Glycodomains. *Journal of the American Chemical Society,* 124**,** 9833-9844.

COPELAND, R. A. & SPIRO, T. G. 1985. Ultraviolet Resonance Raman Spectra of Cytochrome C Conformational States. *Biochemistry,* 24**,** 4960-4968.

CREIGHTON, T. E. 2010. Biophysical Chemistry of Nucleic Acids and Proteins. [Online]. Knovel. Helvetian Press. Available at: https://app.knovel.com/hotlink/pdf/id:kt00TXWTQ2/biophysical-chemistry/protein-structure.

DAY, C. L., STOWELL, K. M., BAKER, E. N. & TWEEDIE, J. W. 1992. Studies of the N-terminal Half of Human Lactoferrin Produced from the Cloned cDNA Demonstrate that Interlobe Interactions Modulate Iron Release. *Journal of Biological Chemistry,* 267**,** 13857-13862.

DE VASCONCELOS NASSER CAETANO, L., DE OLIVEIRA MENDES, T., BAGATIN, E., AMANTE MIOT, H., MARQUES SOARES, J. L., SIMOES E SILVA ENOKIHARA, M. M. & ABRAHAO MARTIN, A. 2017. In Vivo Confocal Raman Spectroscopy for Intrinsic Aging and Photoaging Assessment. *Journal of Dermatological Science,* 88**,** 199-206.

DHAR, S. K. & DAS, M. 2018. Engineering Antibodies. *Journal of the Indian Institute of Science,* 98**,** 5-20.

DIMASI, J. A. 2002. The Value of Improving the Productivity of the Drug Development Process. *PharmacoEconomics,* 20**,** 1-10.

DIMASI, J. A., HANSEN, R. W. & GRABOWSKI, H. G. 2003. The Price of Innovation: New Estimates of Drug Development Costs. *Journal of Health Economics,* 22**,** 151-185.

DONG, A., PRESTRELSKI, S. J., ALLISON, S. D. & CARPENTER, J. F. 1995. Infrared Spectroscopic Studies of Lyophilization and Temperature-induced Protein Aggregation. *Journal of Pharmaceutical Sciences,* 84**,** 415-424.

DURIG, J. R., GEYER, T. J., LITTLE, T. S. & KALASINSKY, V. F. 1987. Spectra and Structure of Small Ring Compounds. XLVIII. Conformational Stability of Methylcyclobutane from Low Frequency Raman Data of the Gas. *The Journal of Chemical Physics,* 86**,** 545-551.

ECKENROTH, B. E., STEERE, A. N., CHASTEEN, N. D., EVERSE, S. J. & MASON, A. B. 2011. How the Binding of Human Transferrin Primes the Transferrin Receptor Potentiating Iron Release at Endosomal pH. *Proceedings of the National Academy of Sciences,* 108**,** 13089-13094.

ECKER, D. M., JONES, S. D. & LEVINE, H. L. 2015. The Therapeutic Monoclonal Antibody Market. *mAbs,* 7**,** 9-14.

EDELMAN, G. M., CUNNINGHAM, B. A., GALL, W. E., GOTTLIEB, P. D., RUTISHAUSER, U. & WAXDAL, M. J. 1969. The Covalent Structure of an Entire GammaG Immunoglobulin Molecule. *Proceedings of the National Academy of Sciences of the United States of America,* 63**,** 78-85.

EILERS, P. H. C. & BOELENS, H. F. M. 2005. Baseline Correction with Asymmetric Least Squares Smoothing. The Netherlands: Biosystems Data Analysis Group, Swammerdam Institute for Life Sciences (SILS), University of Amsterdam.

EJIMA, D., TSUMOTO, K., FUKADA, H., YUMIOKA, R., NAGASE, K., ARAKAWA, T. & PHILO, J. S. 2007. Effects of Acid Exposure on the Conformation, Stability, and Aggregation of Monoclonal Antibodies. *Proteins: Structure, Function, and Bioinformatics,* 66**,** 954-962.

EL-MASHTOLY, S. F., YAMAUCHI, S., KUMAUCHI, M., HAMADA, N., TOKUNAGA, F. & UNNO, M. 2005. Structural Changes during the Photocycle of Photoactive Yellow Protein Monitored by Ultraviolet Resonance Raman Spectra of Tyrosine and Tryptophan. *The Journal of Physical Chemistry B,* 109**,** 23666-23673.

ELBAGERMA, M. A., ALAJTAL, A. I., EDWARDS, H. G. M., AZIMI, G. H., VERMA, K. D. & SCOWEN, I. J. 2015. Raman Spectroscopic and Potentiometric Studies of Acidity level and Dissociation of Citric acid in Aqueous Solution. *Journal of Applied Chemical Science International,* 2**,** 1-11.

ELZOGHBY, A. O., SAMY, W. M. & ELGINDY, N. A. 2012. Protein-based Nanocarriers as Promising Drug and Gene Delivery Systems. *Journal of Controlled Release,* 161**,** 38-49.

EMAMI, F., VATANARA, A., PARK, E. J. & NA, D. H. 2018. Drying Technologies for the Stability and Bioavailability of Biopharmaceuticals. *Pharmaceutics,* 10**,** 131.

ENGEL, J., GERRETZEN, J., SZYMAŃSKA, E., JANSEN, J. J., DOWNEY, G., BLANCHET, L. & BUYDENS, L. M. 2013. Breaking with Trends in Pre-Processing? *Trends in Analytical Chemistry,* 50**,** 96-106.

FARRELL, H. M., JR., JIMENEZ-FLORES, R., BLECK, G. T., BROWN, E. M., BUTLER, J. E., CREAMER, L. K., HICKS, C. L., HOLLAR, C. M., NG-KWAI-HANG, K. F. & SWAISGOOD, H. E. 2004. Nomenclature of the Proteins of Cows' Milk -Sixth Revision. *Journal of Dairy Science,* 87**,** 1641-1674.

FARRELL, H. M., MALIN, E. L., BROWN, E. M. & MORA-GUTIERREZ, A. 2009. Review of the Chemistry of α_{S2}-Casein and the Generation of a Homologous Molecular Model to Explain its Properties. *Journal of Dairy Science,* 92**,** 1338-1353.

FERRARO, J. R., NAKAMOTO, K. & BROWN, C. W. 2003. *Introductory Raman Spectroscopy,* Amsterdam, Elsevier.

FERSHT, A. 1999. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding,* New York, W. H. Freeman.

FILIPE, V., HAWE, A., CARPENTER, J. F. & JISKOOT, W. 2013. Analytical Approaches to Assess the Degradation of Therapeutic Proteins. *Trends in Analytical Chemistry,* 49**,** 118-125.

FLEISCHMAN, J. B., PORTER, R. R. & PRESS, E. M. 1963. The Arrangement of the Peptide Chains in Gamma-Globulin. *The Biochemical Journal,* 88**,** 220-228.

FOURNIER, P., OELKERS, E. H., GOUT, R. & POKROVSKI, G. 1998. Experimental Determination of Aqueous Sodium-Acetate Dissociation Constants at Temperatures from 20 to 240°C. *Chemical Geology,* 151**,** 69-84.

FRANKS, F. 1993. Conformational Stability of Proteins *In:* FRANKS, F. (ed.) *Protein Biotechnology: Isolation, Characterization, and Stabilization. [Online]. Springer.* Available at: https://link-springer-com.ezproxy.lancs.ac.uk/book/10.1007%2F978-1-59259-438-2.

FRANKS, F. & AUFFRET, T. 2007. Freeze-Drying of Pharmaceuticals and Biopharmaceuticals: Principles and Practice. [Online]. Knovel. Available at: https://app.knovel.com/web/toc.v/cid:kpFPB00001/viewerType:toc//root_slug: viewerType%3Atoc/url_slug:root_slug%3Afreeze-drying-pharmaceuticals?kpromoter=federation.

FROST, R. L. & KLOPROGGE, J. T. 2000. Raman Spectroscopy of the Acetates of Sodium, Potassium and Magnesium at Liquid Nitrogen Temperature. *Journal of Molecular Structure,* 526**,** 131-141.

FRUSHOUR, B. G. & KOENIG, J. L. 1974. Raman Spectroscopic Study of Tropomyosin Denaturation. *Biopolymers,* 13**,** 1809-1819.

GABER, B. P., MISKOWSKI, V. & SPIRO, T. G. 1974. Resonance Raman Scattering from Iron (III)-and Copper (II)-Transferrin and an Iron (III) Model Compound. Spectroscopic Interpretation of the Transferrin Binding Site. *Journal of the American Chemical Society,* 96**,** 6868-6873.

GEITNER, R., KÖTTERITZSCH, J., SIEGMANN, M., BOCKLITZ, T., HAGER, M., SCHUBERT, U., GRÄFE, S., DIETZEK, B., SCHMITT, M. & POPP, J. 2015. Two-Dimensional Raman Correlation Spectroscopy Reveals Molecular Structural Changes During Temperature-induced Self-healing in Polymers Based on the Diels–Alder Reaction. *Physical Chemistry Chemical Physics,* 17**,** 22587-22595.

GITLIN, I., CARBECK, J. D. & WHITESIDES, G. M. 2006. Why Are Proteins Charged? Networks of Charge–Charge Interactions in Proteins Measured by Charge Ladders and Capillary Electrophoresis. *Angewandte Chemie International Edition,* 45**,** 3022-3060.

GÖMEZ DE LA CUESTA, R., GOODACRE, R. & ASHTON, L. 2014. Monitoring Antibody Aggregation in Early Drug Development Using Raman

Spectroscopy and Perturbation-Correlation Moving Windows. *Analytical Chemistry,* 86**,** 11133-11140.

GRANTE, I., ACTINS, A. & OROLA, L. 2014. Protonation Effects on the UV/Vis Absorption Spectra of Imatinib: A Theoretical and Experimental Study. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy,* 129**,** 326-332.

GREENFIELD, N. J. 1996. Methods to Estimate the Conformation of Proteins and Polypeptides from Circular Dichroism Data. *Analytical Biochemistry,* 235**,** 1-10.

GÜNZLER, H. & GREMLICH, H.-U. 2002. *IR spectroscopy. An Introduction,* Chichester UK, John Wiley & Sons

HADDEN, J. M., BLOEMENDAL, M., HARIS, P. I., SRAI, S. K. S. & CHAPMAN, D. 1994. Fourier Transform Infrared Spectroscopy and Differential Scanning Calorimetry of Transferrins: Human Serum Transferrin, Rabbit Serum Transferrin and Human Lactoferrin. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology,* 1205**,** 59-67.

HASHIMOTO, S., OBATA, K., TAKEUCHI, H., NEEDLEMAN, R. & LANYI, J. K. 1997. Ultraviolet Resonance Raman Spectra of Trp-182 and Trp-189 in Bacteriorhodopsin: Novel Information on the Structure of Trp-182 and Its Steric Interaction with Retinal. *Biochemistry,* 36**,** 11583-11590.

HAVER, T. O. 2006. Fast Smoothing Function. MATLAB.

HERNÁNDEZ-VIDALES, K., GUEVARA, E., OLIVARES-ILLANA, V. & GONZÁLEZ, F. J. 2019. Characterization of Wild-type and Mutant p53 Protein by Raman Spectroscopy and Multivariate Methods. *Journal of Raman Spectroscopy,* 50**,** 1388-1394.

HERNÁNDEZ, B., COÏC, Y. M., PFLÜGER, F., KRUGLIK, S. G. & GHOMI, M. 2016. All Characteristic Raman markers of Tyrosine and Tyrosinate Originate from Phenol Ring Fundamental Vibrations. *Journal of Raman Spectroscopy,* 47**,** 210-220.

HERNÁNDEZ, B., PFLÜGER, F., KRUGLIK, S. G. & GHOMI, M. 2013. Characteristic Raman lines of Phenylalanine Analyzed by a Multiconformational Approach. *Journal of Raman Spectroscopy,* 44**,** 827-833.

HILDEBRANDT, P. G., COPELAND, R. A., SPIRO, T. G., OTLEWSKI, J., LASKOWSKI JR, M. & PRENDERGAST, F. G. 1988. Tyrosine Hydrogen-bonding and Environmental Effects in Proteins Probed by Ultraviolet Resonance Raman Spectroscopy. *Biochemistry,* 27**,** 5426-5433.

HOAGLAND, P., UNRUH, J., WICKHAM, E. & FARRELL, H. 2001. Secondary Structure of Bovine α_{s2}-Casein: Theoretical and Experimental Approaches. *Journal of Dairy Science,* 84**,** 1944-1949.

HØJGAARD, C., KOFOED, C., ESPERSEN, R., JOHANSSON, K. E., VILLA, M., WILLEMOËS, M., LINDORFF-LARSEN, K., TEILUM, K. & WINTHER, J. R. 2016. A Soluble, Folded Protein without Charged Amino Acid Residues. *Biochemistry,* 55**,** 3949-3956.

HOLLAND, J. & BOLAND, M. 2014. Post-Translational Modifications of Caseins. *In:* SINGH, H., BOLAND, M. & THOMPSON, A. (eds.) *Milk proteins: From Expression to Food. [Online]. Knovel.* 2nd ed.: Available at: https://app.knovel.com/web/view/khtml/show.v/rcid:kpMPFEFE09/cid:kt00U BB0U3/viewerType:khtml//root_slug:5-post-translational-modifications-of-caseins/url_slug:post-translational-modifications?kpromoter=federation&b-toc-cid=kpMPFEFE09&b-toc-root-slug=&b-toc-url-slug=post-translational-modifications&b-toc-title=Milk%20Proteins%20-%20From%20Expression%20to%20Food%20(2nd%20Edition)&page=1&view=collapsed&zoom=1.

HOLLIGER, P. & HUDSON, P. J. 2005. Engineered Antibody Fragments and the Rise of Single Domains. *Nature Biotechnology,* 23**,** 1126-1136.

HOLT, C., CARVER, J. A., ECROYD, H. & THORN, D. C. 2013. Invited Review: Caseins and the Casein micelle: Their Biological Functions, Structures, and Behavior in Foods. *Journal of Dairy Science,* 96**,** 6127-6146.

HOLT, L. J., BASRAN, A., JONES, K., CHORLTON, J., JESPERS, L. S., BREWIS, N. D. & TOMLINSON, I. M. 2008. Anti-serum Albumin Domain Antibodies for Extending the Half-lives of Short Lived Drugs. *Protein Engineering, Design and Selection,* 21**,** 283-288.

HORNE, D. S. 2002. Casein Structure, Self-assembly and Gelation. *Current Opinion in Colloid & Interface science,* 7**,** 456-461.

HORNE, D. S. 2017. A Balanced View of Casein Interactions. *Current Opinion in Colloid & Interface Science,* 28**,** 74-86.

HOUDE, D. J. & BERKOWITZ, S. A. 2014. *Biophysical Characterization of Proteins in Developing Biopharmaceuticals,* Amsterdam, Elsevier Science.

HUANG, C.-Y., BALAKRISHNAN, G. & SPIRO, T. G. 2006. Protein Secondary Structure from Deep-UV Resonance Raman Spectroscopy. *Journal of Raman Spectroscopy,* 37**,** 277-282.

HUNTER, T. 2012. Why Nature Chose Phosphate to Modify Proteins. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences,* 367**,** 2513-2516.

HUYGHUES-DESPOINTES, B. M., PACE, C. N., ENGLANDER, S. W. & SCHOLTZ, J. M. 2001. Measuring the Conformational Stability of a Protein by Hydrogen Exchange. *In:* MURPHY, K. (ed.) *Protein Structure, Stability, and Folding. Methods in Molecular Biology™. [online]. Springer.* Available

at: https://link-springer-com.ezproxy.lancs.ac.uk/protocol/10.1385/1-59259-193-0:069.

IKEDA, S. & LI-CHAN, E. C. Y. 2004. Raman Spectroscopy of Heat-induced Fine-stranded and Particulate β-lactoglobulin Gels. *Food Hydrocolloids,* 18**,** 489-498.

IRANI, V., GUY, A. J., ANDREW, D., BEESON, J. G., RAMSLAND, P. A. & RICHARDS, J. S. 2015. Molecular Properties of Human IgG Subclasses and Their Implications for Designing Therapeutic Monoclonal Antibodies Against Infectious Diseases. *Molecular Immunology,* 67**,** 171-182.

ISOM, D. G., CASTAÑEDA, C. A., CANNON, B. R., VELU, P. D. & GARCÍA-MORENO E, B. 2010. Charges in the Hydrophobic Interior of Proteins. *Proceedings of the National Academy of Sciences,* 107**,** 16096.

ITO, K. & BERNSTEIN, H. J. 1956. The Vibrational Spectra of the Formate, Acetate, and Oxalate ions *Canadian Journal of Chemistry,* 34**,** 170-178.

JACOB, J., KRAFFT, C., WELFLE, K., WELFLE, H. & SAENGER, W. 1998. Melting Points of Lysozyme and Ribonuclease A Crystals Correlated With Protein Unfolding: A Raman Spectroscopic Study. *Acta Crystallographica. Section D, Biological Crystallography,* 54**,** 74-80.

JANSEN, T. L. C., DIJKSTRA, A. G., WATSON, T. M., HIRST, J. D. & KNOESTER, J. 2006. Modeling the Amide I Bands of Small Peptides. *The Journal of Chemical Physics,* 125**,** 044312.

JARVIS, R. M., BLANCH, E. W., GOLOVANOV, A. P., SCREEN, J. & GOODACRE, R. 2007. Quantification of Casein Phosphorylation With Conformational Interpretation Using Raman Spectroscopy. *Analyst,* 132**,** 1053-1060.

JARVIS, R. M. & GOODACRE, R. 2004. Ultra-Violet Resonance Raman Spectroscopy for the Rapid Discrimination of Urinary Tract Infection Bacteria. *FEMS Microbiology Letters,* 232**,** 127-132.

JEFFREY, P. D., BEWLEY, M. C., MACGILLIVRAY, R. T. A., MASON, A. B., WOODWORTH, R. C. & BAKER, E. N. 1998. Ligand-Induced Conformational Change in Transferrins: Crystal Structure of the Open Form of the N-Terminal Half-Molecule of Human Transferrin. *Biochemistry,* 37**,** 13978-13986.

JISKOOT, W., RANDOLPH, T. W., VOLKIN, D. B., RUSSELL MIDDAUGH, C., SCHÖNEICH, C., WINTER, G., FRIESS, W., CROMMELIN, D. J. A. & CARPENTER, J. F. 1999. Protein Instability and Immunogenicity: Roadblocks to Clinical Application of Injectable Protein Delivery Systems for Sustained Release. *Journal of Pharmaceutical Sciences,* 101**,** 946-954.

JOHNSON, C. R., LUDWIG, M., O'DONNELL, S. & ASHER, S. A. 1984. UV Resonance Raman Spectroscopy of the Aromatic Amino acids and Myoglobin. *Journal of the American Chemical Society,* 106**,** 5008-5010.

JONES, E. M., BALAKRISHNAN, G., SQUIER, T. C. & SPIRO, T. G. 2014. Distinguishing Unfolding and Functional Conformational Transitions of Calmodulin using Ultraviolet Resonance Raman Spectroscopy. *Protein Science,* 23**,** 1094-1101.

JORDAN, T., MUKERJI, I., WANG, Y. & SPIRO, T. G. 1996. UV Resonance Raman Spectroscopy and Hydrogen Bonding of the Proline Peptide Bond. *Journal of Molecular Structure,* 379**,** 51-64.

JORGENSEN, L., GROENNING, M., ROEST, N., PINHOLT, C. & VAN DE WEERT, M. 2013. The Challenges in and Importance of Analysing Protein Structure and Physical Stability in Complex Formulations. *Journal of Drug Delivery Science and Technology,* 23**,** 347-353.

JUNG, Y. M., CZARNIK-MATUSEWICZ, B. & OZAKI, Y. 2000. Two-Dimensional Infrared, Two-Dimensional Raman, and Two-Dimensional Infrared and Raman Heterospectral Correlation Studies of Secondary Structure of β-lactoglobulin in Buffer Solutions. *The Journal of Physical Chemistry B,* 104**,** 7812-7817.

KAMIONKA, M. 2011. Engineering of Therapeutic Proteins Production in Escherichia coli. *Current Pharmaceutical Biotechnology,* 12**,** 268-274.

KANNT, A. & WIELAND, T. 2016. Managing Risks in Drug Discovery: Reproducibility of Published Findings. *Naunyn-Schmiedeberg's Archives of Pharmacology,* 389**,** 353-360.

KAZLAUSKAS, R. 2018. Engineering More Stable Proteins. *Chemical Society Reviews,* 47**,** 9026-9045.

KHRAPUNOV, S. 2009. Circular Dichroism Spectroscopy has Intrinsic limitations for Protein Secondary Structure Analysis. *Analytical Biochemistry,* 389**,** 174-176.

KIM, J. E., PAN, D. & MATHIES, R. A. 2003. Picosecond Dynamics of G-Protein Coupled Receptor Activation in Rhodopsin from Time-Resolved UV Resonance Raman Spectroscopy. *Biochemistry,* 42**,** 5169-5175.

KITAGAWA, T., AZUMA, T. & HAMAGUCHI, K. 1979. The Raman Spectra of Bence-Jones Proteins. Disulfide Stretching Frequencies and Dependence of Raman Intensity of Tryptophan Residues on Their Environments. *Biopolymers,* 18**,** 451-465.

KNEE, K. M. & MUKERJI, I. 2009. Real Time Monitoring of Sickle Cell Hemoglobin Fiber Formation by UV Resonance Raman Spectroscopy. *Biochemistry,* 48**,** 9903-9911.

KOUDELKA, T., HOFFMANN, P. & CARVER, J. A. 2009. Dephosphorylation of α_s- and β-Caseins and Its Effect on Chaperone Activity: A Structural and Functional Investigation. *Journal of Agricultural and Food Chemistry,* 57**,** 5956-5964.

KRUKOWSKI, S., KARASIEWICZ, M. & KOLODZIEJSKI, W. 2017. Convenient UV-Spectrophotometric Determination of Citrates in Aqueous Solutions with Applications in the Pharmaceutical Analysis of Oral Electrolyte Formulations. *Journal of Food and Drug Analysis,* 25**,** 717-722.

KUMAR, A., NARAYANAN, V. & SEKHAR, A. 2020. Characterizing Post-Translational Modifications and Their Effects on Protein Conformation Using NMR Spectroscopy. *Biochemistry,* 59**,** 57-73.

KUMAR, R. & MAUK, A. G. 2012. Protonation and Anion Binding Control the Kinetics of Iron Release from Human Transferrin. *The Journal of Physical Chemistry B,* 116**,** 3795-3807.

KUMOSINKI, T., BROWN, E. M. & FARRELL, H. M. J. 1991. Three-Dimensional Molecular Modeling of Bovine Caseins: α_{s1}-Casein. *Journal of Dairy Science,* 74**,** 2889-2895.

KURNIK, M., HEDBERG, L., DANIELSSON, J. & OLIVEBERG, M. 2012. Folding Without Charges. *Proceedings of the National Academy of Sciences,* 109**,** 5705-5710.

LAGASSÉ, H. A. D., ALEXAKI, A., SIMHADRI, V. L., KATAGIRI, N. H., JANKOWSKI, W., SAUNA, Z. E. & KIMCHI-SARFATY, C. 2017. Recent Advances in (Therapeutic Protein) Drug Development. *F1000Research,* 6**,** 113.

LASCH, P. 2012. Spectral Pre-Processing for Biomedical Vibrational Spectroscopy and Microspectroscopic Imaging. *Chemometrics and Intelligent Laboratory Systems,* 117**,** 100-114.

LEDNEV, I. K., ERMOLENKOV, V. V., HE, W. & XU, M. 2005. Deep-UV Raman Spectrometer Tunable between 193 and 205 nm for Structural Characterization of Proteins. *Analytical and Bioanalytical Chemistry,* 381**,** 431-437.

LEE, S.-H. & KRIMM, S. 1998. Polarized Spectra of Oriented Films of Raman α-Helical and its Poly (L-alanine) N-Deuterated Analogue. *Journal of Raman Spectroscopy,* 29**,** 73-80.

LEE, Y. J. 2017. Analytical and Numerical Characterization of Autocorrelation and Perturbation-Correlation Moving-Window Methods. *Applied Spectroscopy,* 71**,** 1321-1333.

LESK, A. M. 2001. *Introduction to Protein Architecture: The Structural Biology of Proteins,* Oxford, Oxford University Press.

LETO, I., CORONNELLO, M., RIGHESCHI, C., BERGONZI, M. C., MINI, E. & BILIA, A. R. 2016. Enhanced Efficacy of Artemisinin Loaded in Transferrin-Conjugated Liposomes versus Stealth Liposomes against HCT-8 Colon Cancer Cells. *ChemMedChem,* 11**,** 1745-1751.

LEWIS, E. N., QI, W., KIDDER, L. H., AMIN, S., KENYON, S. M. & BLAKE, S. 2014. Combined Dynamic Light Scattering and Raman Spectroscopy Approach for Characterizing the Aggregation of Therapeutic Proteins. *Molecules,* 19**,** 20888-905.

LI, C. H. & LI, T. 2009. Application of Vibrational Spectroscopy to the Structural Characterization of Monoclonal Antibody and its Aggregate. *Current Pharmaceutical Biotechnology,* 10**,** 391-399.

LI, H. & QIAN, Z. M. 2002. Transferrin/Transferrin Receptor-mediated Drug Delivery. *Medicinal Research Reviews,* 22**,** 225-250.

LIANG, M., CHEN, V. Y., CHEN, H.-L. & CHEN, W. 2006. A Simple and Direct Isolation of Whey Components from Raw Milk by Gel Filtration Chromatography and Structural Characterization by Fourier Transform Raman Spectroscopy. *Talanta,* 69**,** 1269-1277.

LIPPERT, J. L., TYMINSKI, D. & DESMEULES, P. J. 1976. Determination of the Secondary Structure of Proteins by Laser Raman Spectroscopy. *Journal of the American Chemical Society,* 98**,** 7075-7080.

LITTLECHILD, J. A. 2013. Protein Structure and Function. *In:* GANELLIN, R., ROBERTS, S. & JEFFERIS, R. (eds.) *Introduction to Biological and Small Molecule Drug Research and Development [Online]. Elsevier* Available from:https://ebookcentral.proquest.com/lib/lancaster/reader.action?docID=1187669&ppg=78.

LITVINOV, R. I., FAIZULLIN, D. A., ZUEV, Y. F. & WEISEL, J. W. 2012. The α-helix to β-sheet Transition in Stretched and Compressed Hydrated Fibrin Clots. *Biophysical Journal,* 103**,** 1020-1027.

LIU, Y. & GUO, R. 2008. pH-Dependent Structures and Properties of Casein Micelles. *Biophysical Chemistry,* 136**,** 67-73.

LOLADZE, V. V., IBARRA-MOLERO, B., SANCHEZ-RUIZ, J. M. & MAKHATADZE, G. I. 1999. Engineering a Thermostable Protein via Optimization of Charge−Charge Interactions on the Protein Surface. *Biochemistry,* 38**,** 16419-16423.

LOLADZE, V. V. & MAKHATADZE, G. I. 2002. Removal of Surface Charge-Charge Interactions from Ubiquitin Leaves the Protein Folded and Very Stable. *Protein Science,* 11**,** 174-177.

LONG, D. A. 2002. *The Raman Effect: A Unified Treatment of the Theory of Raman Scattering by Molecules,* Chichester, England, John Wiley & Sons Ltd.

LÓPEZ-PEÑA, I., LEIGH, B. S., SCHLAMADINGER, D. E. & KIM, J. E. 2015. Insights into Protein Structure and Dynamics by Ultraviolet and Visible Resonance Raman Spectroscopy. *Biochemistry,* 54**,** 4770-4783.

LORD, R. & YU, N.-T. 1970. Laser-Excited Raman Spectroscopy of Biomolecules: I. Native Lysozyme and Its Constituent Amino Acids. *Journal of Molecular Biology,* 50**,** 509-524.

LU, M., ZHAO, L., WANG, Y., YOU, G., KAN, X., ZHANG, Y., ZHANG, N., WANG, B., GUO, Y.-J. & ZHOU, H. 2014. Measurement of the Methemoglobin Concentration Using Raman Spectroscopy. *Artificial cells, Nanomedicine, and Biotechnology,* 42**,** 63-69.

LUDWIG, M. & ASHER, S. A. 1988. Ultraviolet Resonance Raman Excitation Profiles of Tyrosine: Dependence of Raman Cross Sections on Excited-state Intermediates. *Journal of the American Chemical Society,* 110**,** 1005-1011.

MACGILLIVRAY, R. T. A., MOORE, S. A., CHEN, J., ANDERSON, B. F., BAKER, H., LUO, Y., BEWLEY, M., SMITH, C. A., MURPHY, M. E. P., WANG, Y., MASON, A. B., WOODWORTH, R. C., BRAYER, G. D. & BAKER, E. N. 1998. Two High-Resolution Crystal Structures of the Recombinant N-Lobe of Human Transferrin Reveal a Structural Change Implicated in Iron Release. *Biochemistry,* 37**,** 7919-7928.

MAEDA, Y. & KITANO, H. 1995. The Structure of Water in Polymer Systems as Revealed by Raman Spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy,* 51**,** 2433-2446.

MAITI, N. C., APETRI, M. M., ZAGORSKI, M. G., CAREY, P. R. & ANDERSON, V. E. 2004. Raman Spectroscopic Characterization of Secondary Structure in Natively Unfolded Proteins: α-Synuclein. *Journal of the American Chemical Society,* 126**,** 2399-2408.

MAKHATADZE, G. I. 2017. Linking Computation and Experiments to Study the Role of Charge–Charge Interactions in Protein Folding and Stability. *Physical Biology,* 14**,** 013002.

MANGIALARDO, S., PICCIRILLI, F., PERUCCHI, A., DORE, P. & POSTORINO, P. 2012. Raman Analysis of Insulin Denaturation Induced by High-Pressure and Thermal Treatments. *Journal of Raman Spectroscopy,* 43**,** 692-700.

MANNING, M. C. 2005. Use of Infrared Spectroscopy to Monitor Protein Structure and Stability. *Expert Review of Proteomics,* 2**,** 731-743.

MAO, A. H., CRICK, S. L., VITALIS, A., CHICOINE, C. L. & PAPPU, R. V. 2010. Net Charge Per Residue Modulates Conformational Ensembles of Intrinsically Disordered Proteins. *Proceedings of the National Academy of Sciences,* 107**,** 8183.

MARSHALL, W. L. & BEGUN, G. M. 1989. Raman Spectroscopy of Aqueous Phosphate Solutions at Temperatures up to 450 °C. Two Liquid Phases, Supercritical Fluids, and Pyro- to ortho-Phosphate Conversions. *Journal of the Chemical Society, Faraday Transactions 2: Molecular and Chemical Physics,* 85**,** 1963-1978.

MCCONNELL, J. S., MCCONNELL, L. M. & HOSSNER, L. R. 1993. Ultraviolet Spectra of Acetic Acid, Glycine, and Glyphosate. *Journal of the Arkansas Academy of Science,* 47**,** 73-76.

MCDONALD, I. K. & THORNTON, J. M. 1994. Satisfying Hydrogen Bonding Potential in Proteins. *Journal of Molecular Biology,* 238**,** 777-793.

MECKLENBURG, S. L., DONOHOE, R. J. & OLAH, G. A. 1997. Tertiary Structural Changes and Iron Release from Human Serum Transferrin. *Journal of Molecular Biology,* 270**,** 739-750.

MESSORI, L., DAL POGGETTO, G., MONNANNI, R. & HIROSE, J. 1997. The pH Dependent Properties of Metallotransferrins: A Comparative Study. *Biometals,* 10**,** 303-313.

MICSONAI, A., WIEN, F., KERNYA, L., LEE, Y.-H., GOTO, Y., RÉFRÉGIERS, M. & KARDOS, J. 2015. Accurate Secondary Structure Prediction and Fold Recognition for Circular Dichroism Spectroscopy. *Proceedings of the National Academy of Sciences,* 112**,** E3095-E3103.

MILLER, S., JANIN, J., LESK, A. M. & CHOTHIA, C. 1987. Interior and Surface of Monomeric Proteins. *Journal of Molecular Biology,* 196**,** 641-656.

MIURA, T., SUZUKI, K. & TAKEUCHI, H. 2001. Binding of Iron(III) to the Single Tyrosine Residue of Amyloid β-Peptide Probed by Raman Spectroscopy. *Journal of Molecular Structure,* 598**,** 79-84.

MIURA, T., TAKEUCHI, H. & HARADA, I. 1988. Characterization of Individual Tryptophan Side Chains in Proteins Using Raman Spectroscopy and Hydrogen-Deuterium Exchange Kinetics. *Biochemistry,* 27**,** 88-94.

MIURA, T., TAKEUCHI, H. & HARADA, I. 1989. Tryptophan Raman Bands Sensitive to Hydrogen Bonding and Side-Chain Conformation. *Journal of Raman Spectroscopy,* 20**,** 667-671.

MIYAZAWA, T., SHIMANOUCHI, T. & MIZUSHIMA, S. I. 1958. Normal Vibrations of N-Methylacetamide. *The Journal of Chemical Physics,* 29**,** 611-616.

MOLENDA, M., DZIEMBAJ, R., PODSTAWKA, E. & PRONIEWICZ, L. M. 2005. Changes in Local structure of Lithium Manganese Spinels (Li:Mn=1:2) Characterised by XRD, DSC, TGA, IR, and Raman Spectroscopy. *Journal of Physics and Chemistry of Solids,* 66**,** 1761-1768.

MONTI, P., FREDDI, G., BERTOLUZZA, A., KASAI, N. & TSUKADA, M. 1998. Raman Spectroscopic Studies of Silk Fibroin from Bombyx mori. *Journal of Raman Spectroscopy,* 29**,** 297-304.

MORITA, S. 2006. Perturbation-Correlation Moving-Window Two-Dimensional Correlation Spectroscopy. *Applied Spectroscopy,* 60**,** 398-406.

MORRIS, G. A. 2002. The Self-Assembly and Structure of Caseins in Solution. *Biotechnology and Genetic Engineering Reviews,* 19**,** 357-376.

MURPHY, K. P. 2001. Stabilization of Protein Structure. *In:* MURPHY, K. P. (ed.) *Protein Structure, Stability, and Folding. Methods in Molecular Biology™ [Online]. Springer.* Available at: https://link-springer-com.ezproxy.lancs.ac.uk/protocol/10.1385/1-59259-193-0:001.

NAKAMURA, H. 2009. Roles of Electrostatic Interaction in Proteins. *Quarterly Reviews of Biophysics,* 29**,** 1-90.

NAVATI, M. S., SAMUNI, U., AISEN, P. & FRIEDMAN, J. M. 2003. Binding and Release of Iron by Gel-Encapsulated Human Transferrin: Evidence for A Conformational Search. *Proceedings of the National Academy of Sciences,* 100**,** 3832-3837.

NELSON, A. L. 2010. Antibody Fragments: Hope and Hype. *mAbs,* 2**,** 77-83.

NELSON, A. L. & REICHERT, J. M. 2009. Development Trends for Therapeutic Antibody Fragments. *Nature Biotechnology,* 27**,** 331-337.

NG, J. B. & SHURVELL, H. F. 1987. Application of Factor Analysis and Band Contour Resolution Techniques to the Raman Spectra of Acetic Acid in Aqueous Solution. *The Journal of Physical Chemistry,* 91**,** 496-500.

NHS 2017. Commissioning Framework for Biological Medicines (Including Biosimilar Medicines). *In:* MEDICINES, D. A. P. M. P. T., NATIONAL MEDICAL DIRECTORATE (ed.). United Kingdom: NHS England.

NODA, I. 1990. Two-Dimensional Infrared (2D IR) Spectroscopy: Theory and Applications. *Applied Spectroscopy.,* 44**,** 550-561.

NODA, I. 1993. Generalized Two-Dimensional Correlation Method Applicable to Infrared, Raman, and Other Types of Spectroscopy. *Applied Spectroscopy,* 47**,** 1329-1336.

NODA, I. 2012. Close-Up View on the Inner Workings of Two-Dimensional Correlation Spectroscopy. *Vibrational Spectroscopy,* 60**,** 146-153.

NODA, I. 2014a. Frontiers of Two-Dimensional Correlation Spectroscopy. Part 1. New Concepts and Noteworthy Developments. *Journal of Molecular Structure,* 1069**,** 3-22.

NODA, I. 2014b. Two-Dimensional Codistribution Spectroscopy to Determine the Sequential Order of Distributed Presence of Species. *Journal of Molecular Structure,* 1069**,** 50-59.

NODA, I. 2016a. Techniques Useful in Two-Dimensional Correlation and Codistribution Spectroscopy (2DCOS and 2DCDS) Analyses. *Journal of Molecular Structure,* 1124**,** 29-41.

NODA, I. 2016b. Two-Dimensional Correlation Spectroscopy (2DCOS) Analysis of Polynomials. *Journal of Molecular Structure,* 1124**,** 53-60.

NODA, I. 2017. Vibrational Two-Dimensional Correlation Spectroscopy (2DCOS) Study of Proteins. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy,* 187**,** 119-129.

NODA, I. 2018. Two-Trace Two-Dimensional (2T2D) Correlation Spectroscopy – A Method for Extracting Useful Information from a Pair of Spectra. *Journal of Molecular Structure,* 1160**,** 471-478.

NODA, I., DOWREY A.E, MARCOTT , C. & STORY, G. M. 2000. Generalized Two-Dimensional Correlation Spectroscopy. *Applied Spectroscopy,* 54**,** 236A-248A.

NODA, I. & MARCOTT, C. 2002. Two-Dimensional Raman (2D Raman) Correlation Spectroscopy Study of Non-Oxidative Photodegradation of β-Carotene. *The Journal of Physical Chemistry A,* 106**,** 3371-3376.

NODA, I. & OZAKI, Y. 2005. *Two-Dimensional Correlation Spectroscopy: Applications in Vibrational and Optical Spectroscopy,* Chichester, England, John Wiley & Sons.

NODA, I. I. 1986. Two-Dimensional Infrared (2D IR) Spectroscopy of Synthetic and Biopolymers. *Bull. Am. Phy,* 31**,** 520.

NONAKA, M., LI-CHAN, E. & NAKAI, S. 1993. Raman Spectroscopic Study of Thermally Induced Gelation of Whey Proteins. *Journal of Agricultural and Food Chemistry,* 41**,** 1176-1181.

NOOTHALAPATI, H., IWASAKI, K. & YAMAMOTO, T. 2017. Biological and Medical Applications of Multivariate Curve Resolution Assisted Raman Spectroscopy. *Analytical Sciences: The International Journal of the Japan Society for Analytical Chemistry,* 33**,** 15-22.

OLADEPO, S. A., XIONG, K., HONG, Z., ASHER, S. A., HANDEN, J. & LEDNEV, I. K. 2012. UV Resonance Raman Investigations of Peptide and Protein Structure and Dynamics. *Chemical Reviews,* 112**,** 2604-2628.

OLIVER, K. V., MARÉCHAL, A. & RICH, P. R. 2016. Effects of the Hydration State on the Mid-Infrared Spectra of Urea and Creatinine in Relation to Urine Analyses. *Applied spectroscopy,* 70**,** 983-994.

OTA, C., NOGUCHI, S., NAGATOISHI, S. & TSUMOTO, K. 2016. Assessment of the Protein–Protein Interactions in a Highly Concentrated Antibody Solution by Using Raman Spectroscopy. *Pharmaceutical Research,* 33**,** 956-969.

PACE, C. N., HORN, G., HEBERT, E. J., BECHERT, J., SHAW, K., URBANIKOVA, L., SCHOLTZ, J. M. & SEVCIK, J. 2001. Tyrosine Hydrogen Bonds make a Large Contribution to Protein Stability. *Journal of Molecular Biology,* 312**,** 393-404.

PACE, C. N., SHIRLEY, B. A., MCNUTT, M. & GAJIWALA, K. 1996. Forces Contributing to the Conformational Stability of Proteins. *The FASEB Journal,* 10**,** 75-83.

PARK, Y., JIN, S., NODA, I. & JUNG, Y. M. 2018. Recent Progresses inTwo-Dimensional Correlation Spectroscopy (2D-COS). *Journal of Molecular Structure,* 1168**,** 1-21.

PARK, Y., KIM, Y., VIKRAM, K., JUNG, Y. M. & CZARNIK-MATUSEWICZ, B. 2016a. Two-Dimensional Correlation Analysis of pH-induced Raman Spectral Changes of α-Lactalbumin. *Bulletin of the Korean Chemical Society,* 37**,** 783-785.

PARK, Y., NODA, I. & JUNG, Y. M. 2016b. Novel Developments and Applications of Two-Dimensional Correlation Spectroscopy. *Journal of Molecular Structure,* 1124**,** 11-28.

PAVIA, D. L., LAMPMAN, G. M., KRIZ, G. S. & VYVYAN, J. A. 2009. *Introduction to Spectroscopy,* USA, Brooks/Cole, Cengage Learning.

PELTON, J. T. & MCLEAN, L. R. 2000. Spectroscopic Methods for Analysis of Protein Secondary Structure. *Anal Biochem,* 277**,** 167-76.

PHRMA 2016. Pharmaceutical Research and Manufacturers of America Biopharmaceutical Research Industry Profile. Washington DC.

PIERIDOU, G. K. & HAYES, S. C. 2009. UV Resonance Raman Spectroscopy of TTR(105–115): Determination of the pKa of Tyrosine. *Physical Chemistry Chemical Physics,* 11**,** 5302-5309.

PRICE, N. C. 2000. Conformational Issues in the Characterization of Proteins. *Biotechnology and Applied Biochemistry,* 31**,** 29-40.

PRICE, N. C. & NAIRN, J. 2009. *Exploring Proteins: A Student's Guide to Experimental Skills and Methods,* New York, Oxford University Press

RAGHUNATHAN, G., SOKALINGAM, S., SOUNDRARAJAN, N., MADAN, B., MUNUSSAMI, G. & LEE, S.-G. 2013. Modulation of Protein Stability and Aggregation Properties by Surface Charge Engineering. *Molecular BioSystems,* 9**,** 2379-2389.

RAMAN, C. V. & KRISHNAN, K. S. 1928. A New Type of Secondary Radiation. *Nature,* 121**,** 501-502.

RAPP, C., KLERMAN, H., LEVINE, E. & MCCLENDON, C. L. 2013. Hydrogen Bond Strengths in Phosphorylated and Sulfated Amino Acid Residues. *PLOS ONE,* 8**,** e57804.

RINNAN, Å., BERG, F. V. D. & ENGELSEN, S. B. 2009. Review of the Most Common Pre-Processing Techniques for Near-Infrared Spectra. *Trends in Analytical Chemistry,* 28**,** 1201-1222.

RODGERS, K. R., SU, C., SUBRAMANIAM, S. & SPIRO, T. G. 1992. Hemoglobin R—T Structural Dynamics from Simultaneous Monitoring of Tyrosine and Tryptophan Time-Resolved UV Resonance Raman Signals. *Journal of the American Chemical Society,* 114**,** 3697-3709.

RODRIGO, G., GRUVEGÅRD, M. & VAN ALSTINE, J. M. 2015. Antibody Fragments and Their Purification by Protein L Affinity Chromatography. *Antibodies,* 4**,** 259-277.

RODRIGUEZ-MENDIETA, I. R., SPENCE, G. R., GELL, C., RADFORD, S. E. & SMITH, D. A. 2005. Ultraviolet Resonance Raman Studies Reveal the Environment of Tryptophan and Tyrosine Residues in the Native and Partially Folded States of the E colicin-Binding Immunity Protein Im7. *Biochemistry,* 44**,** 3306-3315.

RUSSELL, M. D., BUKHARI, M. & GALLOWAY, J. 2018. The Price of Good Health Care. *Rheumatology,* 58**,** 931-932.

RYGULA, A., MAJZNER, K., MARZEC, K., KACZOR, A., PILARCZYK, M. & BARANSKA, M. 2013. Raman Spectroscopy of Proteins: A Review. *Journal of Raman Spectroscopy,* 44**,** 1061-1076.

RYU, S. R., CZARNIK-MATUSEWICZ, B., DUKOR, R. K., NAFIE, L. A. & JUNG, Y. M. 2012. Analysis of the Molten Globule State of Bovine α-Lactalbumin by Using Vibrational Circular Dichroism. *Vibrational Spectroscopy,* 60**,** 68-72.

RYU, S. R. & JUNG, Y. M. 2011. Moving Window Principal Component Analysis for Detecting Positional Fluctuation of Spectral Changes. *Bulletin of the Korean Chemical Society,* 32**,** 2332-2338.

RYU, S. R., NODA, I. & JUNG, Y. M. 2010. What is the Origin of Positional Fluctuation of Spectral Features: True Frequency Shift or Relative Intensity Changes of Two Overlapped Bands? *Applied Spectroscopy,* 64**,** 1017-1021.

RYU, S. R., NODA, I., LEE, C.-H., LEE, P. H., HWANG, H. & JUNG, Y. M. 2011. Two-Dimensional Correlation Analysis and Waterfall Plots for Detecting Positional Fluctuations of Spectral Changes. *Applied Spectroscopy,* 65**,** 359-368.

SAMANEN, J. 2013. The Structure and Business of Biopharmaceutical Companies Including the Management of Risks and Resources. *In:* GANELLIN, R., ROBERTS, S. & JEFFERIS, R. (eds.) *Introduction to Biological and Small Molecule Drug Research and Development [Online] ProQuest.* Available at: https://ebookcentral.proquest.com/lib/lancaster/reader.action?docID=1187669&ppg=246.

SANCHEZ-RUIZ, J. M. & MAKHATADZE, G. I. 2001. To Charge or Not to Charge? *Trends in Biotechnology,* 19**,** 132-135.

SAXENA, V., PANICUCCI, R., JOSHI, Y. & GARAD, S. 2009. Developability Assessment in Pharmaceutical Industry: An Integrated Group Approach for Selecting Developable Candidates. *Journal of Pharmaceutical Sciences,* 98**,** 1962-1979.

SCHACK, M. M., DAHL, K., RADES, T., GROENNING, M. & CARPENTER, J. F. 2019. Spectroscopic Evidence of Tertiary Structural Differences Between Insulin Molecules in Fibrils. *Journal of Pharmaceutical Sciences,* 108**,** 2871-2879.

SCHLÜCKER, S., LIANG, C., STREHLE, K. R., DIGIOVANNA, J. J., KRAEMER, K. H. & LEVIN, I. W. 2006. Conformational Differences in Protein Disulfide Linkages Between Normal Hair and Hair from Subjects with Trichothiodystrophy: A Quantitative Analysis by Raman Microspectroscopy. *Biopolymers,* 82**,** 615-622.

SCHROEDER, H. W. & CAVACINI, L. 2010. Structure and Function of Immunoglobulins. *Journal of Allergy and Clinical Immunology,* 125**,** S41-S52.

SCHWEIKER, K. L. & MAKHATADZE, G. I. 2009. Protein Stabilization by the Rational Design of Surface Charge-Charge Interactions. *Methods in Molecular biology (Clifton, N.J.),* 490**,** 261-283.

SCHWEITZER-STENNER, R. 2001. Visible and UV-Resonance Raman spectroscopy of Model Peptides. *Journal of Raman Spectroscopy,* 32**,** 711-732.

SEMMLER, J. & IRISH, D. E. 1988. Vibrational Spectral Studies of Solutions at Elevated Temperatures and Pressures. IX. Acetic Acid. *Journal of Solution Chemistry,* 17**,** 805-824.

SHAO, Z., VOLLRATH, F., SIRICHAISIT, J. & YOUNG, R. J. 1999. Analysis of Spider Silk in Native and Supercontracted States Using Raman Spectroscopy. *Polymer,* 40**,** 2493-2500.

SHAVER, J. M., CHRISTENSEN, K. A., PEZZUTI, J. A. & MORRIS, M. D. 1998. Structure of Dihydrogen Phosphate Ion Aggregates by Raman-Monitored Serial Dilution. *Applied Spectroscopy,* 52**,** 259-264.

SIAMWIZA, M. N., LORD, R. C., CHEN, M. C., TAKAMATSU, T., HARADA, I., MATSUURA, H. & SHIMANOUCHI, T. 1975. Interpretation of the Doublet

at 850 and 830 cm^{-1} in the Raman Spectra of Tyrosyl Residues in Proteins and Certain Model Compounds. *Biochemistry,* 14**,** 4870-4876.

SIMONSON, T. 2003. Electrostatics and Dynamics of Proteins. *Reports on Progress in Physics,* 66**,** 737-787.

SINGH, M. 1999. Transferrin As A Targeting Ligand for Liposomes and Anticancer Drugs. *Current Pharmaceutical Design,* 5**,** 443-451.

SMITH, E. & DENT, G. 2005. *Modern Raman Spectroscopy: A Practical Approach,* Chichester, England, John Wiley & Sons Ltd.

SPIRO, T. G. & GABER, B. P. 1977. Laser Raman Scattering as a Probe of Protein Structure. *Annual Review of Biochemistry,* 46**,** 553-570.

STROHL, W. R. & STROHL, L. M. 2012. Antibody Fragments as Therapeutics. *In:* STROHL, W. R. & STROHL, L. M. (eds.) *Therapeutic Antibody Engineering [Online]. Elsevier.* Available at: https://www.sciencedirect.com/science/article/pii/B9781907568374500129.

STUART, B. 2004. *Infrared Spectroscopy: Fundamentals and Applications. [Online]. Proquest*, Available at: https://ebookcentral.proquest.com/lib/lancaster/detail.action?docID=194354.

SWAISGOOD, H. E. 2003. Chemistry of the Caseins. *In:* FOX, P. F. & MCSWEENEY, P. L. H. (eds.) *Advanced Dairy Chemistry—1 Proteins: Part A / Part B.* Boston, MA: Springer US.

SZWED, M., MATUSIAK, A., LAROCHE-CLARY, A., ROBERT, J., MARSZALEK, I. & JOZWIAK, Z. 2014. Transferrin as a Drug Carrier: Cytotoxicity, Cellular Uptake and Transport Kinetics of Doxorubicin Transferrin Conjugate in the Human Leukemia Cells. *Toxicology In Vitro,* 28**,** 187-197.

TAKEUCHI, H. 2003. Raman Structural Markers of Tryptophan and Histidine Side Chains in Proteins. *Biopolymers,* 72**,** 305-317.

TAKEUCHI, H. 2011. UV Raman Markers for Structural Analysis of Aromatic Side Chains in Proteins. *Analytical Sciences,* 27**,** 1077-1077.

TAMIZI, E. & JOUYBAN, A. 2016. Forced Degradation Studies of Biopharmaceuticals: Selection of Stress Conditions. *European Journal of Pharmaceutics and Biopharmaceutics,* 98**,** 26-46.

TANAKA, M., ABIKO, S., KOGA, T., KOSHIKAWA, N. & KINOSHITA, T. 2009. Aggregation Induced α-Helix/β-Sheet Transition of the Poly(ethylene glycol)-attached Peptide. *Polymer journal,* 41**,** 153-156.

THIAGARAJAN, G., WIDJAJA, E., HEO, J. H., CHEUNG, J. K., WABUYELE, B., MOU, X. & SHAMEEM, M. 2015. Use of Raman and Raman Optical Activity

for the Structural Characterization of a Therapeutic Monoclonal Antibody Formulation Subjected to Heat Stress. *Journal of Raman Spectroscopy,* 46**,** 531-536.

THOMAS, A., MEURISSE, R. & BRASSEUR, R. 2002. Aromatic Side-Chain Interactions in Proteins. II. Near- and Far-sequence Phe-X pairs. *Proteins: Structure, Function, and Bioinformatics,* 48**,** 635-644.

THOMAS, M. & RICHARDSON, H. H. 2000. Two-Dimensional FT-IR Correlation Analysis of the Phase Transitions in a Liquid Crystal, 4′-n-Octyl-4-Cyanobiphenyl (8CB). *Vibrational Spectroscopy,* 24**,** 137-146.

THORN, D. C., ECROYD, H., CARVER, J. A. & HOLT, C. 2015. Casein Structures in the Context of Unfolded Proteins. *International Dairy Journal,* 46**,** 2-11.

THORSTENSEN, K. & ROMSLO, I. 1990. The Role of Transferrin in the Mechanism of Cellular Iron Uptake. *Biochemical Journal,* 271**,** 1-9.

TOMIMATSU, Y., KINT, S. & SCHERER, J. R. 1976. Resonance Raman Spectra of Iron (III)-, Copper (II)-, Cobalt (III)-, and Manganese (III)-Transferrins and of Bis (2, 4, 6-Trichlorophenolato) Diimidazolecopper (II) Monohydrate, a Possible Model for Copper (II) Binding to Transferrins. *Biochemistry,* 15**,** 4918-4924.

TORII, H., TATSUMI, T. & TASUMI, M. 1998. Effects of Hydration on the Structure, Vibrational Wavenumbers, Vibrational Force field and Resonance Raman Intensities of N-Methylacetamide. *Journal of Raman Spectroscopy,* 29**,** 537-546.

TSUBOI, M., SUZUKI, M., OVERMAN, S. A. & THOMAS, G. J. 2000. Intensity of the Polarized Raman Band at 1340-1345 cm^{-1} As an Indicator of Protein α-Helix Orientation: Application to Pf1 Filamentous Virus. *Biochemistry,* 39**,** 2677-2684.

TUMA, R. 2005. Raman Spectroscopy of Proteins: From Peptides to Large Assemblies. *Journal of Raman Spectroscopy,* 36**,** 307-319.

VANDENABEELE, P. 2013. *Practical Raman Spectroscopy : An Introduction,* Chichester, England, John Wiley & Sons, Ltd.

VOHNÍK, S., TUMA, R., THOMAS, G. J., HANSON, C., FUCHS, J. A. & WOODWARD, C. 1998. Conformation, Stability, and Active-Site Cysteine Titrations of Escherichia coli D26A Thioredoxin Probed by Raman Spectroscopy. *Protein science,* 7**,** 193-200.

VON BONSDORFF, L., TÖLÖ, H., LINDEBERG, E., NYMAN, T., HARJU, A. & PARKKINEN, J. 2001. Development of a Pharmaceutical Apotransferrin Product for Iron Binding Therapy. *Biologicals,* 29**,** 27-37.

WAGNER, E., CURIEL, D. & COTTEN, M. 1994. Delivery of Drugs, Proteins and Genes into Cells using Transferrin as a Ligand for Receptor-Mediated Endocytosis. *Advanced Drug Delivery Reviews,* 14**,** 113-135.

WALLY, J. & BUCHANAN, S. K. 2007. A Structural Comparison of Human Serum Transferrin and Human Lactoferrin. *BioMetals,* 20**,** 249.

WALSH, G. 2002. *Proteins: Biochemistry and Biotechnology,* Chichester,UK, Wiley.

WALSH, G. 2010. Post-Translational Modifications of Protein Biopharmaceuticals. *Drug Discov Today,* 15**,** 773-80.

WALSH, G. 2014. Biopharmaceutical Benchmarks 2014. *Nature Biotechnology,* 32**,** 992-1000.

WALSH, G. 2018. Biopharmaceutical Benchmarks 2018. *Nature Biotechnology,* 36**,** 1136-1145.

WALSH, G. & JEFFERIS, R. 2006. Post-Translational Modifications in the Context of Therapeutic Proteins. *Nature Biotechnology,* 24**,** 1241-1252.

WANG, L.-Y., ZHANG, Y.-H. & ZHAO, L.-J. 2005. Raman Spectroscopic Studies on Single Supersaturated Droplets of Sodium and Magnesium Acetate. *The Journal of Physical Chemistry A,* 109**,** 609-614.

WANG, W. 1999. Instability, Stabilization, and Formulation of Liquid Protein Pharmaceuticals. *International Journal of Pharmaceutics,* 185**,** 129-188.

WEISSER, N. E. & HALL, J. C. 2009. Applications of Single-Chain Variable Fragment Antibodies in Therapeutics and Diagnostics. *Biotechnology Advances,* 27**,** 502-520.

WEN, Z. Q. 2007. Raman Spectroscopy of Protein Pharmaceuticals. *J Pharm Sci,* 96**,** 2861-78.

WEN, Z. Q. & THOMAS JR., G. J. 1998. UV Resonance Raman Spectroscopy of DNA and Protein Constituents of Viruses: Assignments and Cross Sections for Excitations at 257, 244, 238, and 229 nm. *Biopolymers,* 45**,** 247-256.

WILLIAMS, A. F. & BARCLAY, A. N. 1988. The Immunoglobulin Superfamily— Domains for Cell Surface Recognition. *Annual Review of Immunology,* 6**,** 381-405.

WU, Y., ZHANG, L., JUNG, Y. M. & OZAKI, Y. 2018. Two-Dimensional Correlation Spectroscopy in Protein Science, A Summary for Past 20 Years. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy,* 189**,** 291-299.

XENAKI, K. T., OLIVEIRA, S. & VAN BERGEN EN HENEGOUWEN, P. M. P. 2017. Antibody or Antibody Fragments: Implications for Molecular Imaging and Targeted Therapy of Solid Tumors. *Frontiers in Immunology,* 8**,** 1-6.

XIA, Y. & XUN, L. 2017. Revised Mechanism and Improved Efficiency of the QuikChange Site-Directed Mutagenesis Method [Online]. Springer. *In:* REEVES, A. (ed.) *In Vitro Mutagenesis: Methods and Protocols.* Available at: https://link-springer-com.ezproxy.lancs.ac.uk/protocol/10.1007/978-1-4939-6472-7_25.

XIE, C., LI, Y.-Q., TANG, W. & NEWTON, R. J. 2003. Study of Dynamical Process of Heat Denaturation in Optically Trapped Single Microorganisms by Near-Infrared Raman Spectroscopy. *Journal of Applied Physics,* 94**,** 6138-6142.

XIE, Y., JIANG, Y. & BEN-AMOTZ, D. 2005. Detection of Amino acid and Peptide Phosphate Protonation Using Raman Spectroscopy. *Analytical Biochemistry,* 343**,** 223-230.

XU, M., ERMOLENKOV, V. V., HE, W., UVERSKY, V. N., FREDRIKSEN, L. & LEDNEV, I. K. 2005. Lysozyme Fibrillation: Deep UV Raman Spectroscopic Characterization of Protein Structural Transformation. *Biopolymers,* 79**,** 58-61.

XU, M., ERMOLENKOV, V. V., UVERSKY, V. N. & LEDNEV, I. K. 2008. Hen Egg white Lysozyme Fibrillation: A Deep-UV Resonance Raman Spectroscopic Study. *Journal of Biophotonics,* 1**,** 215-229.

YANG, X., XU, W., DUKLESKA, S., BENCHAAR, S., MENGISEN, S., ANTOCHSHUK, V., CHEUNG, J., MANN, L., BABADJANOVA, Z., ROWAND, J., GUNAWAN, R., MCCAMPBELL, A., BEAUMONT, M., MEININGER, D., RICHARDSON, D. & AMBROGELLY, A. 2013. Developability Studies before Initiation of Process Development. *mAbs,* 5**,** 787-794.

YU, N.-T. & JO, B. 1973. Comparison of Protein Structure in Crystals and in Solution by Laser Raman Scattering. II. Ribonuclease A and Carboxypeptidase A. *Journal of the American Chemical Society,* 95**,** 5033-5037.

YU, N.-T., JO, B. & LIU, C. 1972. Laser Raman Spectroscopic Study of the Effect of Solvation on the Conformation of Ribonuclease A. *Journal of the American Chemical Society,* 94**,** 7572-7575.

ZHANG, C. 2016. *Freeze-drying of Engineered Proteins Using Protein Modelling Tools and Experimental Validation.* Doctor of Philosophy, University College London.

ZHANG, D., ORTIZ, C., XIE, Y., DAVISSON, V. J. & BEN-AMOTZ, D. 2005. Detection of the Site of Phosphorylation in a Peptide Using Raman Spectroscopy and Partial Least Squares Discriminant Analysis. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy,* 61**,** 471-475.

ZHANG, D., XIE, Y., MROZEK, M. F., ORTIZ, C., DAVISSON, V. J. & BEN-AMOTZ, D. 2003. Raman Detection of Proteomic Analytes. *Analytical Chemistry,* 75**,** 5703-5709.

ZHANG, H., JI, Y., CHEN, Q., JIAO, X., HOU, L., ZHU, X. & ZHANG, Z. 2015. Enhancement of Cytotoxicity of Artemisinin Toward Cancer Cells by Transferrin-Mediated Carbon Nanotubes Nanoparticles. *Journal of Drug Targeting,* 23**,** 552-567.

ZHOU, C., QI, W., NEIL LEWIS, E. & CARPENTER, J. F. 2015. Concomitant Raman Spectroscopy and Dynamic Light Scattering for Characterization of Therapeutic Proteins at High Concentrations. *Analytical Biochemistry,* 472**,** 7-20.

ZHOU, P., XIE, X., KNIGHT, D. P., ZONG, X.-H., DENG, F. & YAO, W.-H. 2004. Effects of pH and Calcium Ions on the Conformational Transitions in Silk Fibroin Using 2D Raman Correlation Spectroscopy and 13C Solid-State NMR. *Biochemistry,* 43**,** 11302-11311.

ZURDO, J. 2013. Developability Assessment As an Early De-Risking Tool for Biopharmaceutical Development. *Pharmaceutical BioProcessing,* 1**,** 29–50.