Creating an Automated Analysis Toolbox for Ultrafast 2D-IR Spectroscopy

By Lucy Minnes

Department of Physics

University of Strathclyde

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

2019

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

Signed:

Date:

1 Table of Contents

Parts of this thesis are based on the following publications:

(1) Minnes, L.; Shaw, D. J.; Cossins, B. P.; Donaldson, P. M.; Greetham, G. M.; Towrie, M.; Parker, A. W.; Baker, M. J.; Henry, A. J.; Taylor, R. J.; et al. Quantifying Secondary Structure Changes in Calmodulin Using 2D-IR Spectroscopy. *Anal Chem*. **2017**, *89*, 10898–10906.

(2) Minnes, L.; Greetham, G.; Shaw, D. J.; Clark, I.; Fritzsch, R.; Towrie, M.; Parker, A.; Henry, A. J.; Taylor, R. J.; Hunt, N. T. Uncovering the Early Stages of Domain Melting in Calmodulin with Ultrafast Temperature-Jump Infrared Spectroscopy. *J Phys Chem B*. **2019**, *123,* 8733-8739.

2 Acknowledgements

Firstly, I would like to thank my supervisor Prof. Neil Hunt for all the support and guidance he has provided me with over the past four years. I cannot thank him enough for all the discussions and teachings about 2D-IR spectroscopy and all the helpful feedback he has given with regards to this thesis. I would also like to extend my thanks to my other supervisors Prof. Tony Parker and Dr. Matthew Baker for their insightful discussions throughout my PhD.

Next I would like to thank the fellow members of the UCP group I had the pleasure of doing my PhD with Robby, Gordon, Sam, and Niall, who made the long days and nights at RAL tolerable, and occasionally enjoyable. Thank you for all the chats about science, and, some would say more importantly, thank you for sharing my love of food. I could not have asked to spend my time during my PhD with a better bunch of people.

I cannot mention RAL without thanking Greg, Mike, Paul, Ian, Igor and Emma for all their help during my many weeks there. Without your tireless work and help in acquiring and interpreting the data this thesis would not be possible and I appreciate each and everyone one of you. Thank you for not only the lasers but also for making me feel welcome whenever I was at RAL.

I would like to extend my thanks to my collaborators at UCB. Thank you, Danny, Alistair, Richard, and Ben. Without you this project would not be what it is today. I want to especially thank Danny for all his time and effort on this calmodulin project from preparing protein to acquiring additional data.

I would like to thank my parents for all the belief and love they have shown me not just during my PhD but throughout my life. Without your unwavering support I would not be who I am today, and I cannot thank you enough. I would especially like to thank my mum for putting up with me for 25 years even though I am the crankiest person ever in the mornings!

Finally, I would like to thank Sean for being with me on this journey and always encouraging me and believing in me, even when I do not believe in myself. I could not have done this without you. Thank you for being there for me during all the long days and nights of writing and keeping me sane throughout. I cannot wait to see what the future holds for us.

3 List of Abbreviations

2D-IR	Two-dimensional infrared
2D-NMR	Two-dimensional nuclear magnetic resonance
ANOVA-PCA	Analysis of variance - principal component analysis
AOM	Acousto-optic modulator
Apo-CaM	Calcium free calmodulin
BBO	Barium borate
CaM	Calmodulin
CD	Circular dichroism
DNA	Deoxyribonucleic acid
DSC	Differential scanning calorimeter
ESA	Excited state absorption
FSD	Fourier self-deconvolution
FT	Fourier transform
FTIR	Fourier transform infrared
H-bonding	Hydrogen-bonding
He:Ne	Helium:Neon
Holo-CaM	Calcium bound calmodulin
HSA	Human serum albumin
IR	Infrared
MCT	Mercury cadmium telluride
MD	Molecular Dynamics
MgO:PPLN	Magnesium oxide periodically poled lithium niobite
MVA	Multivariate analysis
NMR	Nuclear magnetic resonance
OD	Optical density
OPA	Optical parametric amplifier
OPO	Optical parametric oscillator
PCA	Principal component analysis
PDB	Protein data bank
PTFE	Polytetrafluoroethylene

RAL	Rutherford Appleton Laboratory
RMS	Root mean squared
SVD	Singular value decomposition
TDC	transition dipole coupling
TFA	Trifluoroacetic acid
TFP	Trifluoperazine
Ti:Sapph	Titanium:Sapphaire
T-jump	Temperature-jump
TRIR	Time-resolved infrared
TRMPS	time-resolved multiple probe spectroscopy
T_w	Waiting time

4 Abstract

Understanding protein structure is intrinsic to understanding their biological mechanisms. The technique of infrared (IR) spectroscopy can be used to analyse secondary structure of proteins using the amide I vibrational mode. However, the broad peak associated with this mode can be convoluted and extracting structural information can be difficult. The technique of two-dimensional infrared (2D-IR) spectroscopy can assist by spreading the spectral content across two frequency axes. It is a time-resolved method based on a series of sub-picosecond laser pulses that allows for ultrafast dynamics to be probed providing an advantage over more established structural analysis techniques, such as circular dichroism (CD) spectroscopy and X-ray crystallography.

The aims of this thesis are to assess the ability of 2D-IR spectroscopy to extract information pertaining to changes in structure of proteins in tandem with the multivariate analysis method of principal component analysis (PCA). This has been performed as a means to understand the spectral features of structure change in proteins upon perturbation and provides the basis for analysing protein dynamics using time-resolved infrared (TRIR) spectroscopy. As a testbed for this analysis, the previously well-characterised protein calmodulin (CaM) has been studied. CaM is a calcium binding messenger protein that is predominantly α-helical in structure. Ca^{2+} free CaM (*apo*) can bind up to four Ca^{2+} ions (*holo*) which produces a change in structure and α-helical content.

Thermal studies on both *apo*- and *holo*-CaM were performed and the quantitative results from 2D-IR spectroscopy produced good agreement with CD spectroscopy with a reduction in α-helix structure by 13% (CD) and 15% (2D-IR) observed for *apo*-CaM. Accurate differentiation between melting transitions and generic heating effects was achieved using the thermally stable *holo*-CaM as a reference.

A temperature-jump (T-jump) TRIR experiment was then established. The system was calibrated and determined to induce a 9 °C temperature rise in the sample. Comparison of non-equilibrium relaxations of *apo*-CaM and *holo*-CaM showed domain melting of *apo*-CaM begins on microsecond timescales with α-helix destabilisation. These observations enable the assignment of previously reported dynamics of CaM on hundreds of microsecond timescales to thermally activated melting, producing a complete mechanism for thermal unfolding of CaM.

CaM-drug binding studies were performed using IR and 2D-IR spectroscopy and small changes in structure and electrostatic properties were extracted utilising PCA. Separation into groups of ligand binding was achieved for IR spectroscopy with 2D-IR spectroscopy providing spectral evidence for the changes occurring.

The limitations of 2D-IR data processing for elucidating the small spectral changes and the developments made to ensure accurate extraction of spectral features are discussed.

1 Introduction

1.1 Proteins

Proteins are large biomolecules that are essential to life with many different types performing several functions within all organisms[1]. The structures of proteins are intrinsically linked to their function where changing the structure of the protein can provide a different function. Therefore, understanding the structure of proteins has been an area of interest for a number of years[2,3]. This thesis will endeavour to understand changes in structure of proteins upon perturbation using two-dimensional infrared (2D-IR) spectroscopy.

Proteins are composed of a series of amino acid residues, with below thirty amino acids in the sequence considered to be a peptide while a larger biomolecule of thirty amino acids or more is a protein. An amino acid is composed of an amine group ($-NH_2$), a carboxyl functional group (-COOH), and a side-chain group (-R) which is unique to each amino acid, of which there are twenty naturally occurring (Figure 1-1).

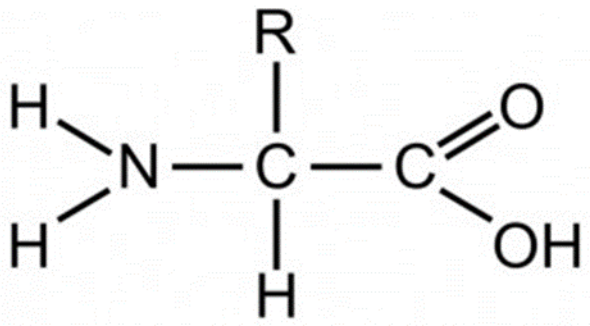


Figure 1-1. Generic structure of a single amino acid, where the "R" group is the side chain of amino acids.

The amine group of one amino acid can bind to the carboxyl group of another in a dehydration reaction forming an amide bond (Figure 1-2, orange CN bond), and multiple amide bonds can form resulting in a long sequence of amino acids. This is the basis of the protein backbone, with the unbonded amine group being referred to as the N-terminus and the unbonded carboxyl group being referred to as the C-terminus.

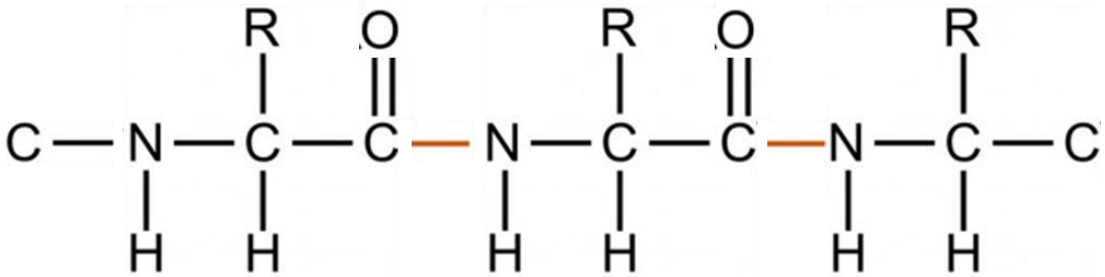


Figure 1-2. A series of amino acids joined by amino bonds (orange C-N bonds) forming a polypeptide chain.

The sequence of the amino acids in the protein is called the primary structure, and the secondary structure is the local repeating units within the protein made up of these amino acids. Examples of secondary structures are α-helix and β-sheet (Figure 1-3). It can be seen from the figure that α-helix is a spiral conformation where, on average, there are 3.7 residues per turn and every N-H group on the backbone undergoes hydrogen bonding (H-bonding) with a C=O group on the backbone located four residues earlier in the sequence. H-bonding is also seen in β-sheets which consists of β-strands connected by backbone H-bonding. β-strands are chains of residues usually three to ten residues in length.

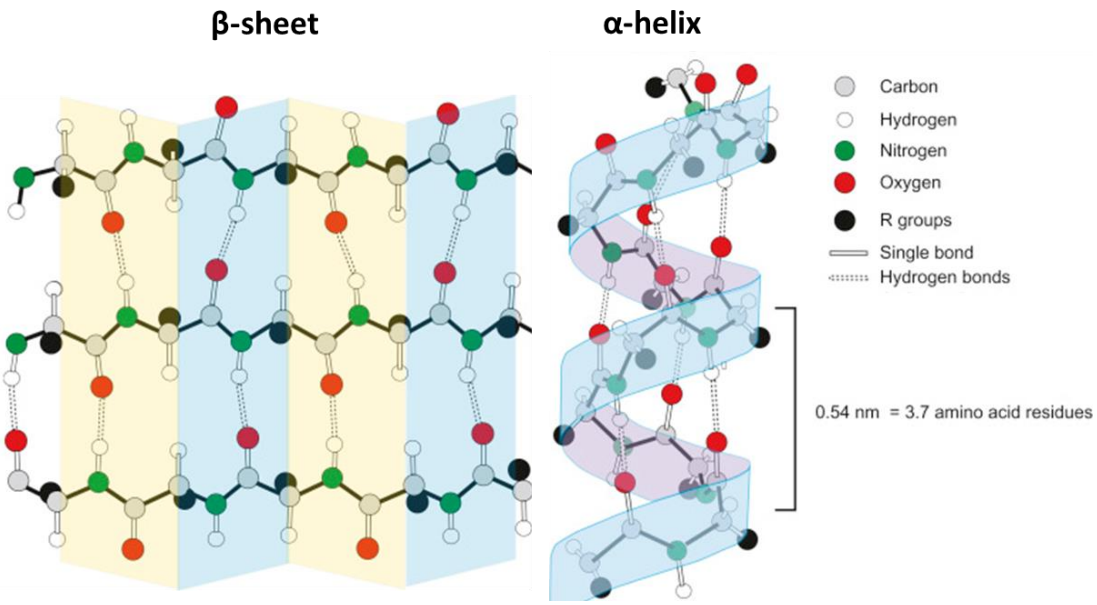


Figure 1-3. Protein secondary structures of α-helix and β-sheet. Adapted from [4].

The tertiary structure of a protein is the overall way the protein folds and, finally, the quaternary structure is the way in which multiple protein units arrange in a complex system. This multi-layered complexity to the structure highlights how small changes can result in

significant changes in the action of proteins and the way in which protein-ligand and protein-drug binding occurs.

Investigating the structure of proteins has been a prominent area of research in many fields including biotechnology[5,6] and drug discovery[7,8] due to its link with function and understanding how proteins will interact with other structures. Within drug development, the structure of proteins is important in understanding its druggability[9,10] and the pharmacokinetics and pharmacodynamics in protein-drug binding[11]. This allows for more efficient development of drug design. Examples of this include using the structure of aspartic protease to model a target for antihypertensives, renin[12], and using the structure of neuraminidase to design the flu drug, Relenza[13]. Some proteins are involved in many biological functions, for example human serum albumin (HSA)[14] and α_1-acid glycoprotein[15] are both responsible for numerous activities.

Proteins can also undergo structural changes after metabolic processes and the resulting conformational change can affect the binding affinity of ligands and how the protein functions in the body. Examples of changes in binding affinity from structural changes include the modifications of glycosylation to HSA[16], and carbamylation of uraemia along with addition of endogenous compounds to the protein leads to changes in ligand binding[17].

1.1.1 Calmodulin

One of the aims of this thesis is to investigate the changes in structure of proteins, specifically the secondary structure, using linear infrared (IR) spectroscopy and two-dimensional infrared (2D-IR) spectroscopy. This will be used as a basis to investigate dynamics of proteins. Utilised throughout this work will be the well-studied calcium binding messenger protein calmodulin (CaM). This protein is used as a test for the experiments and analysis conducted due to the wealth of knowledge available already of its structure and behaviour.

CaM was first discovered in 1970[18,19] and is found in many eukaryotic cells. It is comprised of 148 amino acid residues with an approximate molecular weight of 16.7 kDa[20]. It is involved in many regulatory effects[21,22] of Ca^{2+} including smooth muscle contraction[23], intracellular signalling[24], and cell cycle regulation[25].

The structure of the protein is predominantly α-helical and is largely conserved across a range of organisms[26] including, but not limited to, human, sea urchin, bovine, and rat[27–30]. The structure of CaM comprises of two globular domains, a C-terminal domain and an N-terminal

domain, which are separated by a flexible central linker[31]. Each domain consists of helix-loop-helix E-F hand motifs[32–34] which can accommodate two Ca^{2+} ion binding sites with a maximum of four Ca^{2+} ion binding sites per protein molecule. This E-F hand motif is common to other Ca^{2+} binding proteins including troponin C, a Ca^{2+} binding protein involved in the function of muscle contraction in the skeletal and cardiac muscles[35,36].

In Ca^{2+} free (*apo*) CaM the central linker contains two short α-helices joined by a random coil resulting in structural flexibility. The 3D structure of *apo*-CaM (Figure 1-4. (a)) has been previously determined by NMR studies[37–39]. Uptake of four Ca^{2+} ions to CaM (*holo*-CaM) occurs in a two-step positive cooperative process with binding of the first two ions occurring in the C-domain first, due to a six times greater affinity to this domain compared to the N-domain[40], which the latter two Ca^{2+} ions bind. This binding results in the CaM structure changing from a closed, compact structure to an open "dumbbell" structure where the central linker becomes fully helical. Hydrophobic binding sites for other proteins or small molecules are thus exposed by the open structure of *holo*-CaM[41]. The 3D structure of *holo*-CaM has been established through X-ray crystallography studies[42–44] (Figure 1-4 (b)).

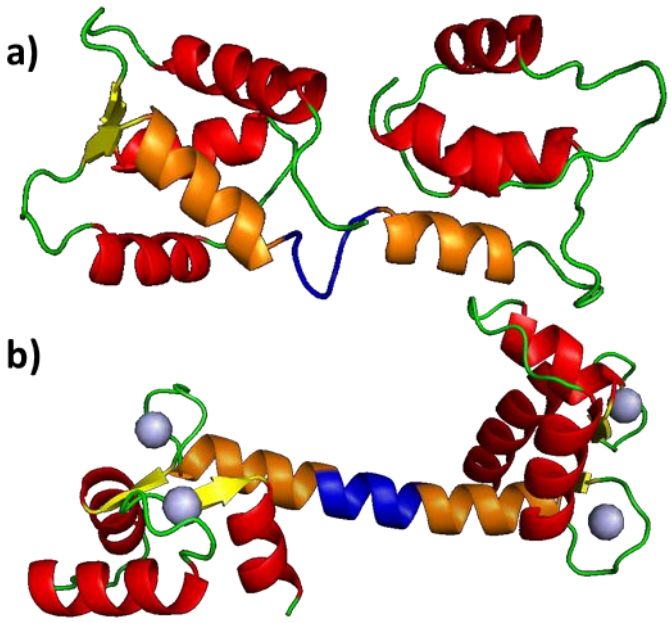


Figure 1-4. 3D molecular structures of a) apo-CaM (PDB 1CFD[39]) and b) holo-CaM (PDB 1CLL[44]) showing α-helices (red), β-strands (yellow), flexible central linker (orange and blue), and Ca^{2+} ions (grey).

Thermal studies on the *apo*- form have been performed which show reversible conformation changes occurring up to 80 °C[45]. The *holo*-CaM form has been reported to be more thermally

stable with Fourier transform infrared (FTIR) spectroscopy and circular dichroism (CD) spectroscopy experiments reporting no conformational changes below 80 °C[46,47].

Many studies have been performed on CaM to understand the structure of the protein and its interaction with ligands therefore making it an ideal candidate for further developing the technique of 2D-IR spectroscopy coupled to multivariate analysis (MVA).

1.1.2 Current techniques of analysis of protein structure

A variety of methods are currently available for analysing the structure of proteins, each with their own advantages and disadvantages. A few such methods shall be discussed here but the following is not an exhaustive list of all possible techniques.

The most widely used technique is X-ray crystallography, with approximately 129,000 entries on the protein data bank (PDB)[48] out of a total of approximately 143,000 entries arising from this technique. It has advanced the understanding of structural biology and has resulted in Nobel prizes for determining the structures, for example, of vitamin B12[49] and insulin[50].

The method of X-ray crystallography involves the use of high-powered X-rays which are aimed at small crystals to investigate the structure of said crystals which cause scattering of the X-rays. The scattered rays are diffracted onto an electronic detector and the pattern and intensity of the rays allows for mapping of each atom in the crystallised molecule. This method gives precise and high-quality detail on secondary and tertiary structure through the use of 3D models with little error observed. It is able to detect atomic bonds due to the short wavelengths of 10^{-10} m that X-rays possess, allowing spatial resolution down to <1.5 Å[51].

However, the technique requires for the samples to be crystalline and therefore gives limited insight into the flexibility of biological molecules as dynamics information is difficult to obtain in this form[52], as opposed to a liquid sample. Crystallising proteins can also prove to be a challenging task which can result in a long time before a measurement can be obtained[53].

A method not limited by the need to have a crystalline structure is circular dichroism (CD) spectroscopy. This method is the measured difference between the absorbance of left and right circularly polarised light[54,55]. Far-UV CD spectroscopy is a “gold standard” technique used to investigate and quantify the secondary structure of proteins[56]. Chiral molecules can absorb left or right-handed circularly polarised light at different ratios and CD spectroscopy is able to measure this. The alignment of the chromophores in the amides of the backbone

of the protein give rise to specific spectra – therefore the different alignments of secondary structure have different spectral features.

Spectral features for α-helices include negative peaks at 208 nm and 222 nm and a positive peak at 193 nm. Anti-parallel β-sheets give rise to a negative peak at 218 nm and positive peak at 195 nm, and disordered structures have low signals around 210 nm and negative at 195 nm[57] (Figure 1-5).

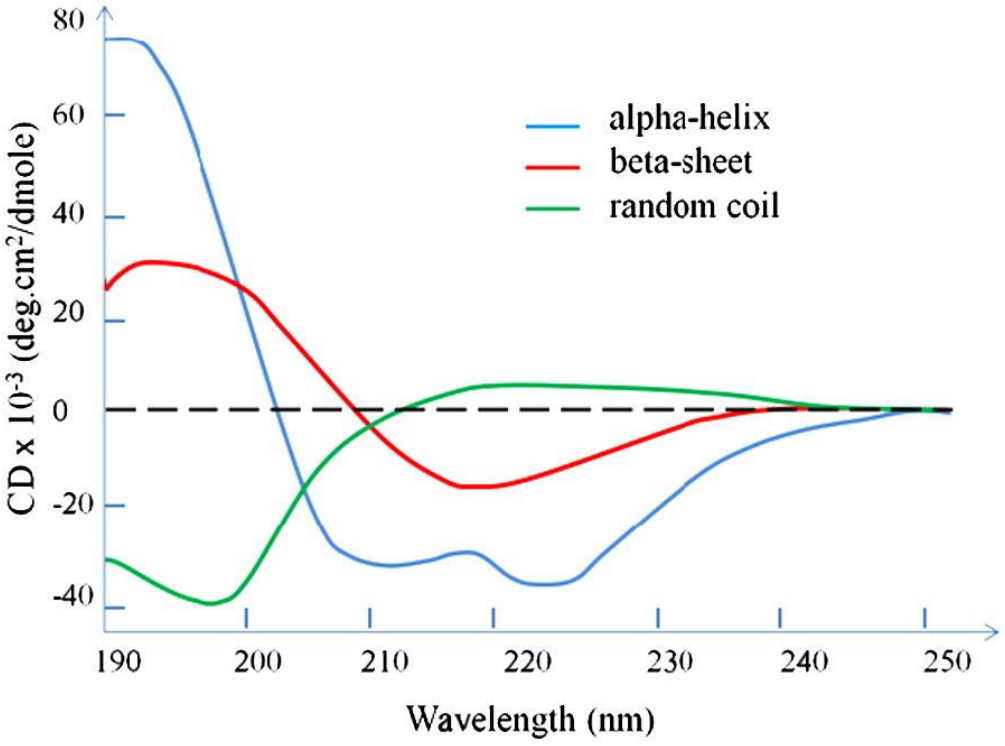


Figure 1-5. Typical CD spectra of secondary structures of α-helix (yellow), β-sheet (blue) and random coil (red)[58].

Although CD spectroscopy is a gold-standard technique for determining and quantifying the secondary structure of proteins, it is limited in its ability to monitor the dynamics of proteins.

The method of Molecular Dynamics (MD) simulations allows for the theoretical evaluation of proteins and their dynamics (Figure 1-6). MD simulations can probe both structure and fast dynamics simultaneously[59]. The method works well with experimental data as they complement each other. However, disadvantages of MD simulations include being prone to inaccuracies in force fields[60] and it has high computational costs for even relatively short simulated time scales.

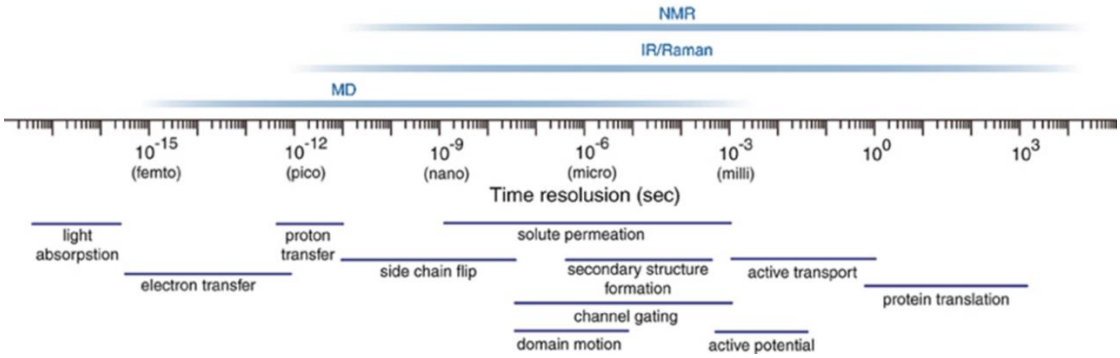


Figure 1-6. Typical protein dynamics and the associated timescales, and the experimental techniques used to probe them. Adapted from[61].

Nuclear magnetic resonance (NMR) spectroscopy is an experimental technique that can probe the structure of proteins and also follow dynamics on the nanosecond to minute timescale. These timescales encompass protein unfolding[62], ligand binding[63] and redox reactions[64]. NMR can analyse proteins to atomic resolution and can be performed in solution[65,66], in contrast to the crystalline requirements of X-ray crystallography.

NMR uses radio wavelengths to detect magnetic resonance of certain atomic nuclei, only those with a nuclear spin greater than zero, and therefore have an odd number of protons or neutrons, are able to be detected with NMR; common nuclei that are investigated include ^{1}H, ^{13}C, ^{15}N and ^{31}P. The radio frequency perturbs the magnetic field of a nucleus and the response is a frequency in resonance with the atomic nucleus that was affected[67].

Dynamics of proteins can be probed by expanding the NMR experimental method to include a series of radio pulses which allows selective excitation of nuclei. This results in a second axis being introduced resulting in a two-dimensional NMR (2D-NMR) spectrum[68]. Disadvantages of NMR spectroscopy include difficulties in analysing very large proteins as it results in poorly resolved peaks. This can be overcome by isotopic labelling, but it is an expensive method[69]. Another disadvantage is that NMR is limited by its temporal resolution, and therefore rapid protein dynamics on the sub-nanosecond timescale, such as energy transfer, cannot be probed.

The focus of this thesis therefore will be investigating a technique that can probe secondary structure changes, with the added benefit of following ultrafast dynamics. 2D-IR spectroscopy is a relatively new technique compared to those previously mentioned that has emerged in the last twenty years and builds upon 2D-NMR by using a series of pulses to obtain a spectrum in two dimensions. However, it utilises infrared (IR) light which has shorter

wavelengths than the NMR counterpart of radio waves, and this gives rise to an increased temporal resolution. Therefore, the timescales provide a complementary experimental technique to the fast time scales of MD simulations.

1.2 IR Spectroscopy

IR spectroscopy is a type of vibrational spectroscopy involving the interaction of both light and matter. Mid-IR radiation covers the range 4000 cm^{-1} to 400 cm^{-1} in the electromagnetic spectrum. The technique is based upon vibrations in molecules whereby the light causes chemical bonds to vibrate as a result of absorption of photons with the same energy as the vibration[70]. Understanding molecular vibrations is paramount to understanding IR spectroscopy.

By first considering a diatomic molecule, the frequency with which the bond oscillates, ν, can be determined using Hooke's Law:

$$\nu = \frac{1}{2\pi}\sqrt{\frac{k}{\mu}} \qquad \text{(Equation 1.1)}$$

Where k is the spring constant for the bond and μ is the reduced mass.

$$\mu = \frac{m_A m_B}{m_A + m_B} \qquad \text{(Equation 1.2)}$$

Where m_A is the mass of one atom in the molecule, and m_B is the mass of the other atom in the molecule. Hooke's Law relates the strength of the bond and the masses of the atoms to the frequency with which they oscillate, and therefore the wavenumber that the molecule absorbs IR radiation.

The energy of the photon absorbed by the molecule vibration is equal to the energy required to transition from one energy level to the next. This transition arises from quantum mechanics and can be calculated from the energy of harmonic motions equation:

$$E = \left(n + \frac{1}{2}\right) h\nu \qquad \text{(Equation 1.3)}$$

Where n is the quantum number and h is Planck's constant.

A harmonic oscillator has evenly spaced energy levels (Figure 1-7, black trace), however no molecule follows this oscillator as increasing the internuclear distance, r, leads to an

inevitable bond dissociation. Molecules are anharmonic oscillators, and the diatomic molecule follows a Morse potential, whereby the energy between levels decreases with increasing levels and at higher energies the bond can dissociate (Figure 1-7, red trace). Anharmonicity plays an important part in time resolved infrared spectroscopy and will be discussed later.

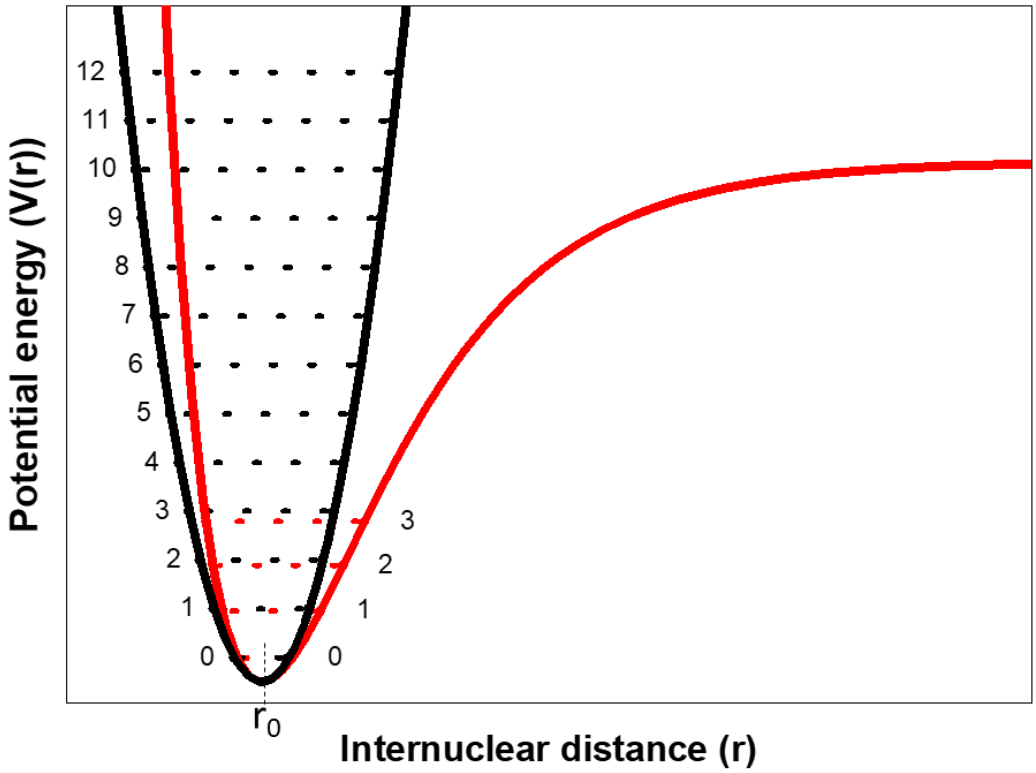


Figure 1-7. Harmonic oscillator (black curve) with evenly distributed energy levels. Morse potential (red curve) demonstrating an anharmonic oscillator where increasing the internuclear distance, r, can result in dissociation of the bond and unevenly spaced levels.

The simplest example of a molecular vibration is small displacements from the bond length of a diatomic molecule. However, larger molecules exhibit a greater degree of vibrational modes with the exact number determined by 3N-5 for linear molecules and 3N-6 for non-linear molecules, where N is the number of atoms present in the molecule. This results in the average protein molecule having more than 20,000 vibrational modes[71]. For a vibrational mode of a molecule to be IR active there must be a change in the dipole moment of the molecule.

The common molecular vibrations can be described by the vibrations seen in a methylene group (-CH_2). These include both stretching and bending vibrations and examples can be seen in Figure 1-8.

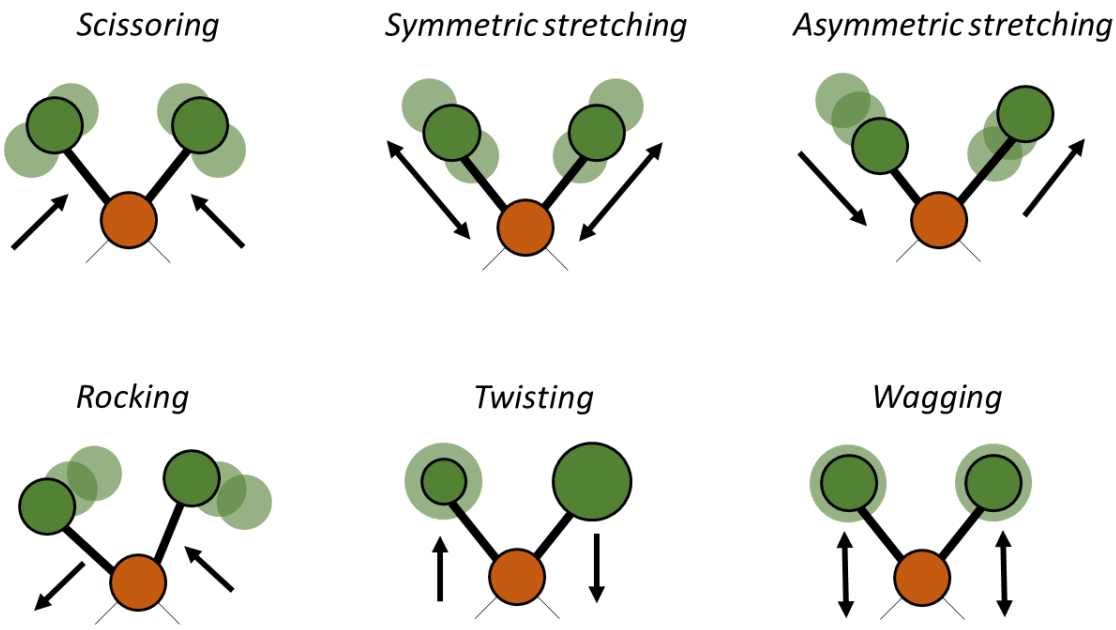


Figure 1-8. Common vibrations, both stretching (symmetric and asymmetric) and bending (scissoring, rocking, twisting, and wagging).

1.2.1 IR spectroscopy of proteins

There are a number of vibrational modes associated with the amide group of proteins and these give rise to distinct spectral peaks. These are amide I, amide II, amide III, amide A, and amide B. The most intense of these modes is the amide I which absorbs at 1600 to 1700 cm^{-1} and is the result of the C=O stretching vibration (Figure 1-9), with smaller contributions from the out-of-phase CN stretching vibration, the CCN deformation, and the NH in-plane bend[71].

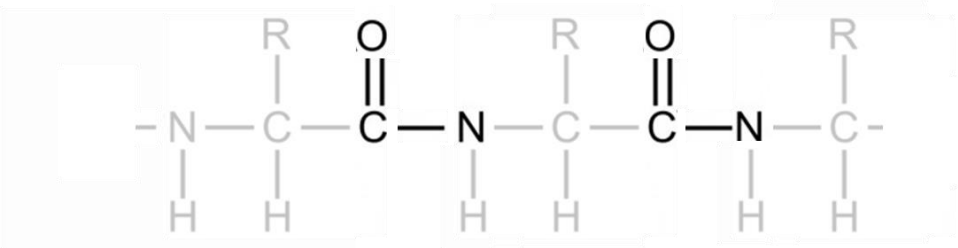


Figure 1-9. Polypeptide chain with the carbonyl bond highlighted as it is the major vibrational contribution to the amide I mode. Smaller contributions are observed from the C-N stretching vibration.

The amide I mode is sensitive to secondary structure and is not strongly influenced by side chains therefore making it the most commonly used mode for secondary structure analysis. The β-sheet has a strong absorption at 1630 cm^{-1} and a weaker absorption at higher wavenumbers (> 1680 cm^{-1}), whereas the α-helix and random coil have overlapping absorptions at 1640 - 1660 cm^{-1} and 1640 – 1650 cm^{-1}, respectively[72].

Many effects on the protein can have an impact of the results observed in the amide I region including through-bond coupling, H-bonding, and transition dipole coupling (TDC). Through-bond coupling only has a small impact on the amide I mode due to only a small displacement occurring on the C_α atom, however TDC is vital to secondary structure analysis and quantification[73–75].

The TDC effect has been shown to be an important factor in analysis and quantification of secondary structure. It is a resonance interaction of adjacent amide vibrations along the polypeptide chain and the extent of coupling depends on the orientation and spatial distribution of the interacting dipoles. This interaction leads to delocalisation of vibrational modes across multiple residues. The TDC effect was first described in 1972[74] where it was found that the amide I splitting observed for the pleated sheets in polyglycine I can be accounted for by the introduction of a direct interaction force constant between adjacent C=O groups in neighbouring chains which arises from TDC.

This effect is also present in α-helical proteins with a previous study[75] indicating that interactions with the three neighbouring groups along the chain are most essential for the amide I band. The length of the helix also appears to affect the frequency of the maximum intensity of the amide I with increased length resulting in a lower maximum wavenumber. The study also stated that H-bonding occurred in α-helices with the third group along in the chain.

H-bonding to the C=O group can be shown to affect the amide I vibration. This has been exhibited through polypeptides with a shift to lower frequency observed for poly-β-L-Glu compared to poly-β-L-Ala that was explained by the former having stronger hydrogen bonds[76]. H-bonding has an impact on the maximum absorption frequency of a secondary structure, with a great degree of H-bonding exhibited correlating to a lower IR frequency. For example, the C=O bond of a protein can be stabilised through the donation of a H-bond from water to the oxygen in the amide group. This results in a lower IR frequency and in turn causes the amide I peak to down-shift in wavenumber. Temperature has been shown to

affect the H-bonds, and results have shown that the amide I band tends to shift to higher frequency with increasing temperature due to weakening of hydrogen bonds[77].

As mentioned, there are other protein vibrational modes. The amide II mode returns a peak at approximately 1550 cm^{-1} and has major contributions from the NH in-plane bend and the CN stretching vibration. Smaller contributions come from the CO in-plane bend and the CC and NC stretches[71] (Figure 1-10).

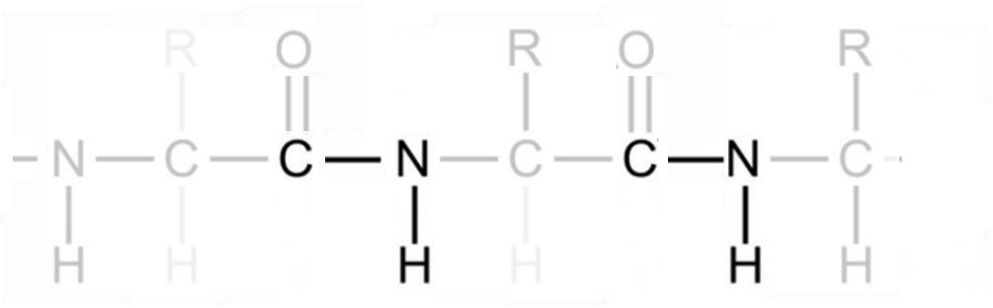


Figure 1-10. Polypeptide chain with the NH and CN bonds highlighted which represent the major vibrational contribution to the amide II mode.

As with the amide I bond there is little contribution from amino acid side chains to the amide II vibration, and in contrast to the amide I bond, the relationship between secondary structure and amide II bond is complex due to added vibrational contributions that are associated with it. In deuterated conditions (amide II') the N^2H bend contribution is down shifted in frequency which results in no coupling between the vibration and the CN stretch. This results in the peak position down shifting to the frequency of the CN stretch (approximately 1470 cm^{-1}). Many IR studies of proteins are performed in deuterated conditions as the H-O-H bending mode at 1650 cm^{-1} interferes with the amide I region. The D-O-D bending mode is down-shifted to 1210 cm^{-1}, eliminating the interference with the protein vibrational mode (Figure 1-11).

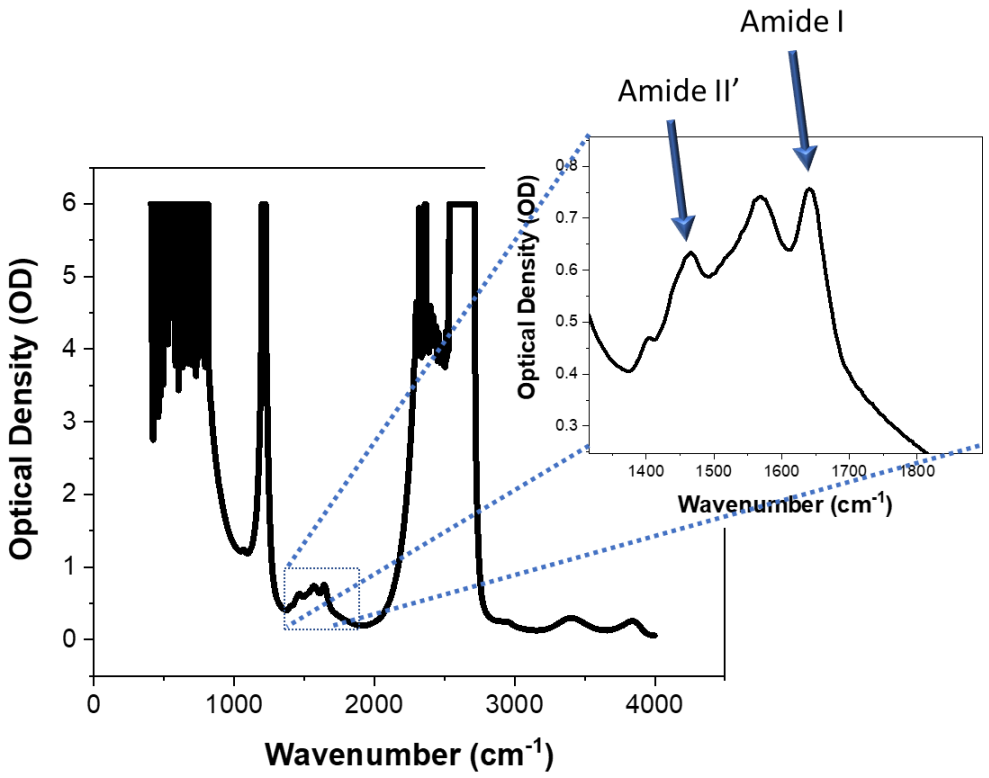


Figure 1-11. IR spectrum of a protein measured in a D_2O solvent with large absorption peaks at approximately 600 cm^{-1}, 1200 cm^{-1} and 2500 cm^{-1} associated with vibrational modes of D_2O. Zoomed portion of the spectrum highlights the amide I and amide II' peaks with the amide II' bond downshifted to approximately 1470 cm^{-1} compared to the amide II frequency of approximately 1550 cm^{-1}.

Many studies have been performed using IR spectroscopy to investigate proteins and their structure. This includes the use of examining H/D exchange[78], for example a study was performed monitoring the H/D exchange for tertiary structure unfolding by changing the pD of a BSA sample following changes to the amide II[79]. Also included is probing protein stability through chemical and thermal denaturation[80,81].

A key part in protein studies by IR spectroscopy is elucidating information about structure using data processing. This is of particular importance for analysing the amide I region due to overlapping contributions from different secondary structure content, which creates a broad indistinguishable peak. Multiple different methods have been employed to extract secondary structure content from the amide I and amide II modes including curve fitting, Fourier self-deconvolution (FSD), difference spectroscopy, second derivatives, and MVA.

Curve fitting of the amide I region involves fitting multiple Gaussian curves within the peak to simulate different secondary structure components. This has been used to quantify the

secondary structure composition of proteins[82,83], although inaccuracies in this method have been apparent[84], and relies on subjective decisions in the fitting process which can affect the results[85,86].

FSD is a signal-processing tool which reduces the bandwidths of overlapping peaks in broad indistinct bands and deconvolutes them to allow for resolution between peaks[87]. This has been used to calculate the fractions of secondary structure components in proteins[88,89].

Difference spectroscopy[90,91] can be used to monitor changes in different protein states by measuring a spectrum in one state, A, and then measuring a spectrum in another state, B, and then subtracting one from the other, i.e. B minus A. This produces a difference spectrum which highlights the changes between the two states.

Second derivative analysis[92] is a process in which the baseline fluctuations are reduced and overlapping bands are deconvoluted. It has been used successfully in the identification of biomarkers for bladder cancer detection in serum[93].

MVA is the statistical analysis of multiple variables at one time. IR spectroscopy can produce many measurements that require analysis across a range of wavelengths leading to software packages that can interpret the complex data[94]. Principal component analysis (PCA) is one such method of MVA that reduces the dimensionality of data and extracts trends in data[95]. It has been used in previous IR studies, for example in the determination of proteins in milk samples[96]. PCA will be utilised in this thesis for extracting trends in datasets and the method will be discussed in further detail in chapter (2).

Although linear IR spectroscopy has been proven to be a useful tool in analysing structure of proteins, it is not without its limitations. One key disadvantage of IR spectroscopy is its inability to probe dynamics, however, by expanding on the technique, 3rd order non-linear IR spectroscopy, a method based on the response from three laser pulses, can delve deeper and gain insights into dynamic information of proteins.

Another disadvantage is that, although the amide I bond is information rich in terms of secondary structure, the broad absorption lineshapes masks the coupling that is present which means extracting structural information is difficult. 2D-IR has the ability to also measure the coupling present in the amide I bond and spreads the information across two axes, making extraction of structural features more accessible.

1.3 Non-linear Spectroscopy

2D-IR spectroscopy is a 3rd order non-linear spectroscopic method that is relatively new with the first measurement performed in 1998[97]. It was first used to examine the amide I vibration and anharmonicity of a peptide linkage molecule, N-methylacetamide, using a frequency domain technique, which shall be described later. The development of the technique is in part due to progression in ultrafast lasers with sub-picosecond pulses as in order to observe fast dynamics, a laser pulse of duration shorter than those fast dynamics is required. Prior to discussing the theory of 2D-IR spectroscopy, it is best to consider the simpler approach, pump-probe spectroscopy, first.

1.3.1 Pump-probe Spectroscopy

IR pump-probe spectroscopy is also a 3rd order non-linear spectroscopic method and is based upon two pulses; a pump (or excitation) pulse which undergoes two interactions between light and matter, and a probe (detection) pulse. A typical pulse sequence schematic is seen in Figure 1-12.

A sub-picosecond pump pulse first interacts with the sample. This will excite molecules within the same frequency region of the bandwidth from the ground vibrational state (v = 0), which is assumed to be the only state significantly populated at room temperature, to the first vibrational state (v = 1). After a short waiting time (T_w) a probe pulse then interacts with the sample.

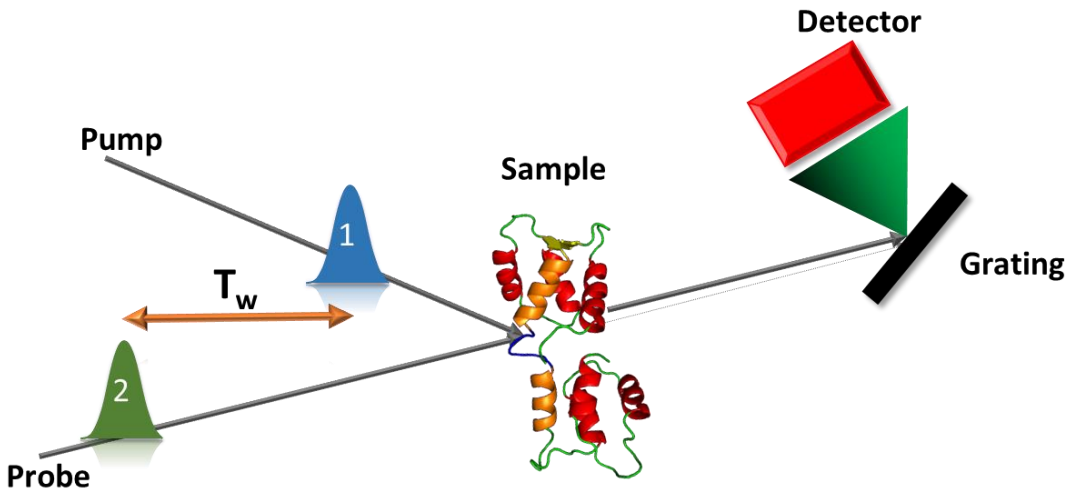


Figure 1-12. Pulse sequence for IR pump-probe spectroscopy. A pump pulse will excite photons in the sample and after a short period of time, T_w, a probe pulse will interact with the sample. An emitted signal along with the probe pulse will be dispersed by a grating prior to detection.

The probe pulse is measured with the pump pulse both on and off – often this is accomplished by chopping the pump pulse to half the repetition rate of the probe pulse. The pump-probe spectrum is the difference spectrum of the probe spectrum measured with pump on minus pump off. The spectrum comprises of pairs of peaks resulting from the difference in absorbance spectra – a positive peak at a lower frequency which corresponds to the excited state absorption (ESA) from the first excitation state to the second as a result of the probe pulse. The second peak is a negative peak (bleach) at higher frequency with corresponds to the stimulated emission from the first excited state back to the ground state and also caused by the ground state bleach (Figure 1-13). The positive ESA is red shifted from the bleach depending on the anharmonicity of the molecule.

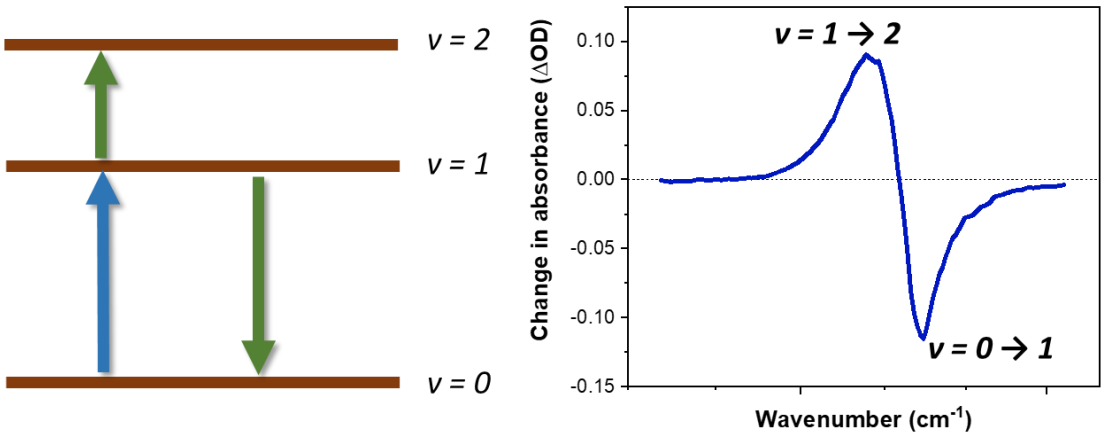


Figure 1-13. Example vibrational excited states for IR pump-probe spectroscopy (left) with blue arrow indicating the transition from the ground excitation state to first excitation state from the pump pulse. The green arrows indicate the two responses from the probe pulse: ESA from the first excitation state to the second, and the stimulated emission from the first state back down to the ground state. An example spectrum (right) shows a pair of peaks: a positive peak at lower frequency corresponding to the v = 1 → 2 transition, and a negative peak for the v = 0 → 1 transition. The difference in frequency between the two peaks corresponds to the anharmonicity of the molecule.

The time between the two pulses, T_w, can be altered in pump-probe spectroscopy. If T_w is zero then no relaxation in the sample will have occurred, resulting in no decaying of the molecules from the first excited state back to the ground state between the pulses. Increasing the time increases the probability that spontaneous decaying of the molecules in the first excitation state back to the ground state will occur. A result of this happening is that the signal size of the pump-probe spectrum becomes smaller as the difference between the pump on and pump off signal is decreased. Scanning the T_w across a range of times allows for the vibrational lifetime of the vibrational mode to be determined. This is only true if magic angle polarisation is used. If another polarisation is used, for example "parallel" where all beams are linearly polarised in the same axis, then the signal will decrease with T_w due to both relaxation from the first excited and from re-orientational motion of the dipoles in the sample. Due to the large sizes of proteins the motion is very slow relative to relaxation and is often neglected[98].

The spectrum in Figure 1-13 is an example of exciting and detecting one vibrational mode in a sample. However, a more complex system could involve multiple modes and coupling between the modes would affect the vibrational lifetimes. Pump-probe spectroscopy is

broadband excitation and therefore all vibrational modes in the selected window are excited simultaneously and information on coupling becomes convoluted with the relaxations. For a multimode system information on coupling and energy transfer are useful and therefore more insight can be gained through 2D-IR spectroscopy.

1.3.2 2D-IR Spectroscopy

2D-IR spectroscopy expands on pump-probe spectroscopy and is able to separate the information from coupling and relaxations into two axes allowing for extraction of different contributions to occur.

There are two ways of collecting 2D-IR spectra, in the frequency domain and time domain. The frequency domain method involves the use of a narrowband system and was utilised in early 2D-IR experiments; however, the time resolution is limited due to the narrowband width[99]. The more common method currently used is the broadband time-domain approach, which has been utilised in this thesis, however, the frequency domain method shall be discuss first as it allows for more comparisons between pump-probe spectroscopy and 2D-IR.

The frequency domain method is based upon the pump-probe method, however, differs in that the normal broadband pump width used in pump-probe spectroscopy is reduced to approximately 10 cm^{-1} prior to reaching the sample, mainly through the use of a Fabry-Perot filter[100]. The pump frequency is scanned and narrowband pump, broadband probe is measured for a given waiting time. As with pump-probe spectroscopy the pump pulse is reduced to half the repetition rate of the probe pulse through the use of a chopper and the probe pulse is dispersed on a grating before being detected by an array detector. The frequency of the pump pulse is scanned across the desired region which creates a 2D-IR spectrum comprising of a series of pump-probe slices at each detected pump frequency.

As with pump-probe spectroscopy, the spectrum comprises of a pair of peaks corresponding to the energy level transitions resulting from the laser pulses and arises from the difference in absorbance between pump on and pump off pulses (Figure 1-14). The negative bleach peak, arising from ground state bleaching lies on the diagonal of the spectrum where the probe frequency (ω_{probe}) is equal to the pump frequency (ω_{pump}) such that,

$$\omega_{probe} = \omega_{pump} = \omega_{01} \qquad \text{(Equation 1.6)}$$

Where ω_{01} is the frequency of the ground state to first excited state energy transition. The positive ESA peak has a lower probe frequency corresponding to the excitation from the first excited state to the second (Equation 1.7, where ω_{12} is the frequency of the transition between the first excited state to the second excited state). The pump frequency of the ESA peak remains the same as the bleach peak (Equation 1.8).

$$\omega_{probe} = \omega_{12} \quad \text{(Equation 1.7)}$$

$$\omega_{pump} = \omega_{01} \quad \text{(Equation 1.8)}$$

A representative linear IR spectrum is shown in Figure 1-14 (above 2D-IR spectrum) that would correspond to the same sample as shown in the 2D-IR spectrum. For every peak in the linear IR spectrum, there are a pair of peaks in the 2D-IR spectrum, assuming anharmonicity in the vibrational mode. The frequency difference between the ESA and bleach peak is the anharmonicity, with an increase in frequency difference indicating a greater degree of anharmonicity. If the system was harmonic the two peaks would overlap and as they have opposite signs, they would destructively interfere and cancel out.

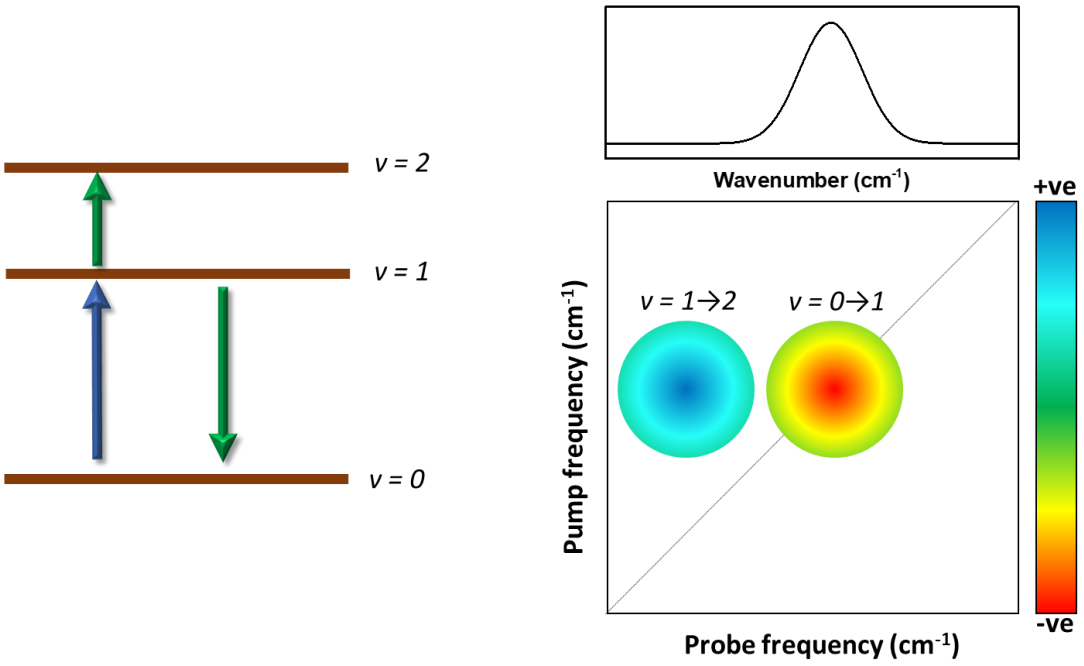


Figure 1-14. Example energy levels for a simple vibrational system (left) with blue arrow indicating the transition from the ground excitation state to first excitation state from the pump pulse. The green arrows indicate the two responses from the probe pulse: ESA from the first excitation state to the second, and the stimulated emission

from the first state back down to the ground state. An example 2D-IR spectrum (bottom right). A positive peak at lower frequency corresponding to the v = 1 → 2 transition, and a negative diagonal peak for the v = 0 → 1 transition, and an equivalent IR spectrum (top right). The two peaks are the result of the changes in absorbance between the pump on and pump off pulses.

Above is assuming a simple system with excitation of one vibrational mode, however 2D-IR spectroscopy has the ability to detect coupling between two vibrational modes which results in off-diagonal contributions. Coupling occurs when two bonds are close enough to each other that the oscillations from one causes vibrations in the second to occur. Figure 1-15 shows a 2D-IR spectrum with coupling between two modes and the associated energy levels.

Both modes share a common ground state, $|00\rangle$, and there are two first excited states, one for each mode, $|01\rangle$ and $|10\rangle$. As previously, the diagonal peaks arise from the ground state to first excited state transitions (negative peak) and the first to second excited state transitions (positive peak). Additional off-diagonal peaks come from exciting from the first excited state to the combination band ($|11\rangle$). It should be noted that $T_w = 0$ for this 2D-IR spectrum example. The peak intensities of the cross-peaks give rise to information about the distance and angle between the two modes which provides details about the molecular structure.

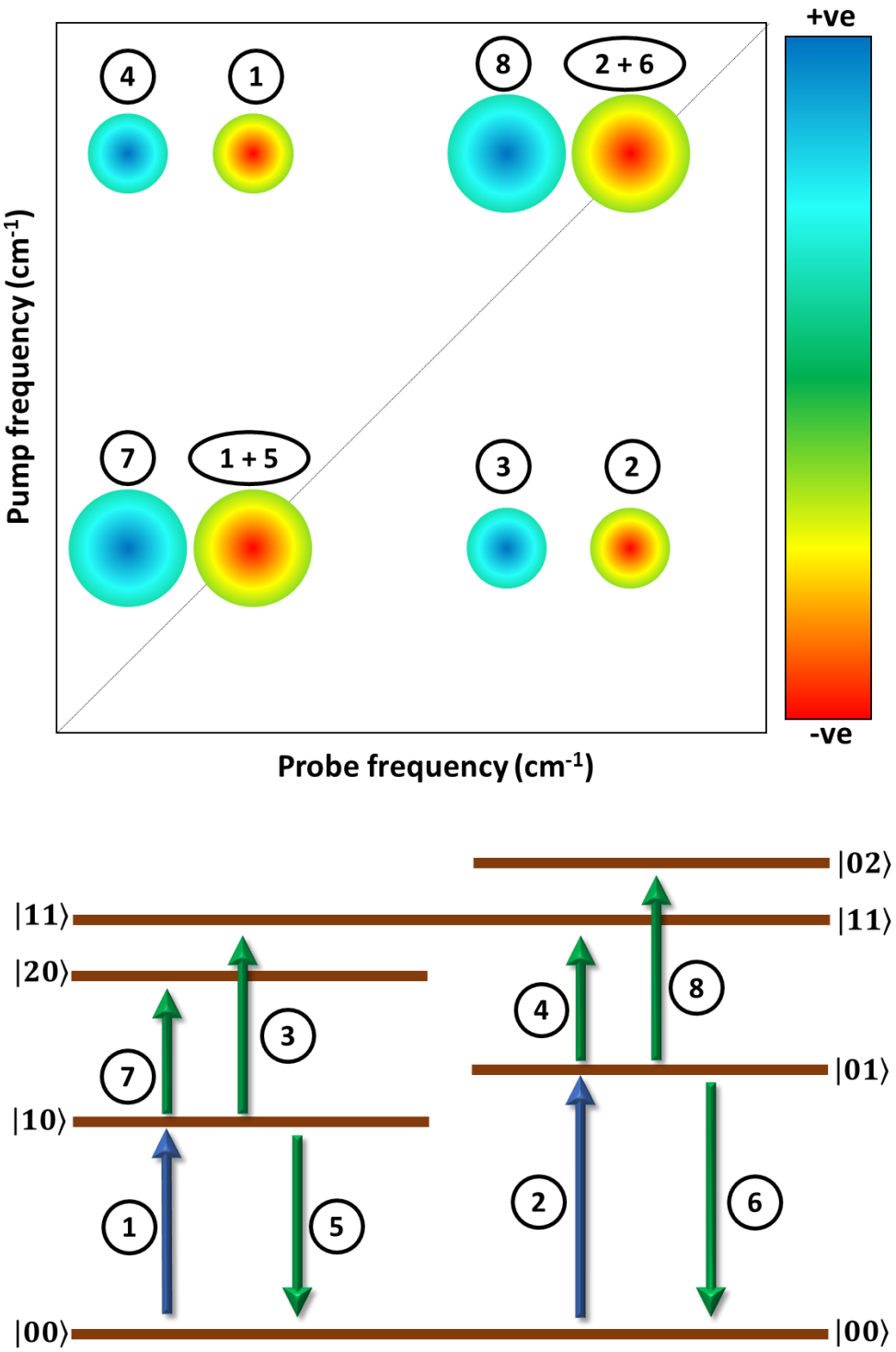


Figure 1-15. 2D-IR spectrum of coupling between two modes showing to resulting cross peaks. The energy levels associated with each peak in the 2D-IR spectrum is illustrated below.

Due to the narrow band width associated with the frequency domain 2D-IR method, the temporal resolution is in the picosecond range. Therefore, any dynamics present in sub-picosecond timescales are not detected. This can be overcome with the use of time-domain 2D-IR. It is based on three laser pulses (Figure 1-16) with the first two pulses being pump pulses which are separated by a time, τ, and the second and third, which is a probe pulse, being separated by a time, T_w. The time between the final probe pulse and the emission of a signal is separated by a time, t.

The process with which the signal is emitted is described by the following. The interaction of the first pump pulse with the sample creates a coherent superposition of the ground and first excited states. After this first pulse the initial phase coherence is lost and the macroscopic polarisation which was created with the first pulse decays. This occurs during the time delay between the first and second pump pulses τ. The method in which the timings between the two pulses are varied can be done in multiple approaches including with the use of an interferometer and the use of a pulse shaper. Both methods will be explained in further detail in chapter (2). The time delay between the first and second pump pulses is scanned across a range of times during the 2D-IR measurement. For example, the τ range used throughout this thesis was -0.3 to 4.0 ps.

A second pump pulse then interacts with the sample and converts the superposition state into a population state, either of the ground or first excited state. During the second time delay T_w the population state evolves allowing processes such as vibrational relaxation, spectral diffusion and energy transfer to occur. This second time delay, T_w, remains constant throughout the 2D-IR measurement. A series of measurements with different T_w times are required to evaluate the processes listed above.

After T_w a third pulse then interacts with the sample which creates a coherent superposition state again, resulting in the emission of a signal, commonly referred to as a photon echo pulse, after a further time t. Different 2D-IR geometries, such as boxcar and pump-probe, produce the emitted signal in different ways. As a pseudo-pump probe geometry was used in this thesis, only this geometry will be considered. The emitted signal after time t is collinear and indistinguishable with the probe pulse and is therefore self-heterodyned, meaning it is correctly phased across the detection axis. This signal is dispersed by a spectrometer onto an array detector. This dispersed signal results in the detection, or probe, axis in the 2D-IR

spectrum. Signals are extracted by recording probe spectra across the detector both with and without preceding pump pulse interactions, allowing for the signal from the probe pulse to be subtracted from the detected signal. The excitation, or pump, frequency axis is a result of Fourier transforming the signal as a function of the time delay between the two pump pulses, τ.

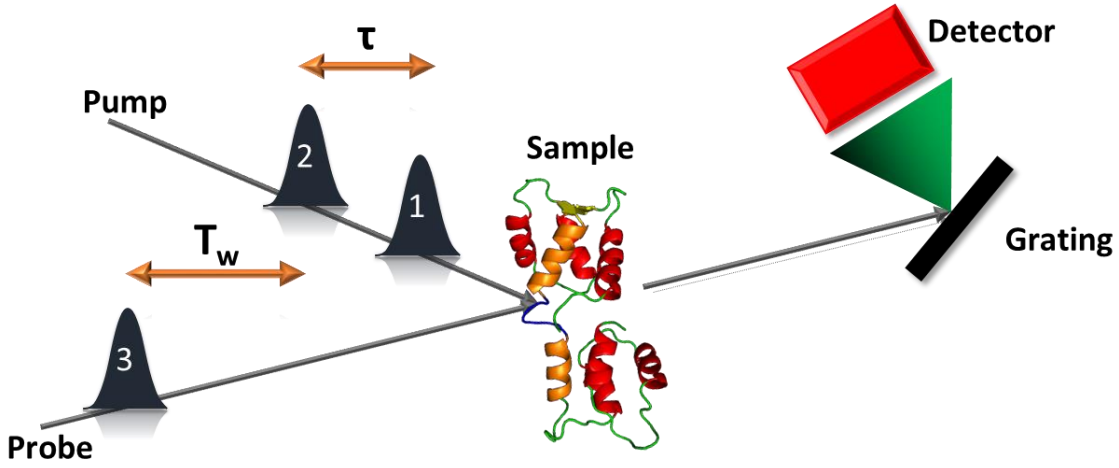


Figure 1-16. Schematic of pulses used in generation of a 2D-IR spectrum. Two pump pulses interact with the sample first and these are separated by a time, τ, then a probe interacts with the sample a set time after the second pump pulse, T_w. A signal pulse is then emitted from the sample (dashed line) before travelling to a grating with separates the pulse into an array before being detected.

1.3.2.1 2D-IR spectroscopy of proteins

Since the first 2D-IR experiment, analysis of proteins and biological samples have been performed extensively using this technique[99,101]. Analysis of protein structure usually focuses on the amide I vibrational mode. Examples include the ability to distinguish between the helical dimer and dimeric helix conformers of the peptide gramicidin D which have only subtle differences in structure[102]. Both conformers appeared to be similar to β-sheets using conventional FTIR spectroscopy, however, coupling observed within 2D-IR was able to determine the dimeric structures.

2D-IR spectroscopy has also been used to investigate β-type structures, including hairpins, parallel and antiparallel sheets[103–105] where the TDC effect was shown to play an important role in the spectral features. One study reported 2D-IR to be a more sensitive method for determining structural differences between β-sheet proteins than traditional linear IR

spectroscopy[105]. As the spectrum is spread across two axes coupling can be determined from off-diagonal features resulting from the two amide I spectral contributions of β-sheets at ~1630 cm^{-1} and ~1680 cm^{-1} leading to a distinct "Z" pattern.

The amide I bond of an α-helix has been investigated through 2D-IR spectroscopy previously[106]. Measuring of isotope labels allowed for evaluation of coupling between amide I groups of the α-helix and through the use of off-diagonal features the TDC effect was shown to be the cause of coupling between groups.

Further work has been performed using a combination of linear IR and 2D-IR spectroscopy for identification of structure and quantification of transition dipole strengths[107,108]. 2D-IR spectroscopy is more sensitive to secondary structures that have delocalised vibrational excitons than linear IR as the area of a linear IR bond is independent of coupling, whereas 2D-IR is dependent. As a result, 2D-IR is able to distinguish between structure that is coupled, such as an α-helix in this study, and structure that is uncoupled, such as random coil. By ratioing the areas for linear IR and 2D-IR, transition dipole strengths were calculated and the values confirmed structure identification[107].

The prior reported studies were performed in deuterated solvents, however recently a study was performed in un-deuterated serum to determine the ratio of albumin to globulin proteins present in the samples, and this was possible due to the amide I mode dominating the 2D-IR signal at approximately 1650 cm^{-1} compared to the H-O-H bend which is more weakly absorbing[109].

Measuring 2D-IR spectra as a function of waiting time is an important technique and can offer structural information on the sub-picosecond timescale; however, this method only probes systems at equilibrium. Combining 2D-IR with non-equilibrium methods, transient 2D-IR, allows for 2D-IR time resolution of systems perturbed from equilibrium. This is performed by addition of another pulse to the experiment to perturb the system prior to the 2D-IR pulse sequence, which then acts as a probe to the change in equilibrium[110]. Different types of transient experiments are possible including UV excitation for photoinduced reactions[111], and pH-jump methods to induce pH-dependent conformational changes[112].

One such transient experiment that shall be the focus of chapter (4), albeit in a time-resolved infrared (TRIR) spectroscopy experiment and not a transient 2D-IR spectroscopy experiment, is temperature jump (T-jump) spectroscopy. This experiment studies thermal denaturation

and protein folding and unfolding dynamics by using a laser pulse to excite the solvent, typically the deuterated overtone band at 5000 cm^{-1}, which in turn heats the sample. 2D-IR spectroscopy can then be used to probe the response of the heating and measure folding and unfolding dynamics. T-jump 2D-IR spectroscopy has been utilised to analyse the unfolding dynamics of ubiquitin using a nanosecond pulse laser to excite the D_2O overtone band resulting in a 9 K T-jump. The results from this experiment returned two unfolding timescales, a shorter one of a few microns and a longer one of a few milliseconds[113]. Another example of T-jump 2D-IR includes insulin dimer dissociation which resulted in four dynamic timescales being observed. These were fast water thermalisation and H-bond weakening (< 5 ns), β-sheet dissociation (250 – 1000 μs), monomer disordering (5 – 150 μs), and finally re-association (> 5 ms)[114]. This highlights the complex nature of protein folding and unfolding. Further details on T-jump experiments will be provided in chapters (2) and (4).

2D-IR spectroscopy experiments result in complex data where a large amount of information is held in one spectrum. This results in a large amount of time being invested by the analyst to interpret the results. In recent years advances in the processing of 2D-IR spectra have been made with the following studies examples in different processing methods.

A method using MVA for quantifying secondary structure from 2D-IR spectra has been developed using singular value decomposition (SVD), which is a dimension reducing technique that separates data in components, similar to PCA. By creating a library of protein amide I spectra from sixteen different proteins of known secondary structure content, SVD was used to quantify the secondary structure of an unknown protein[115]. This analysis was performed on unlabelled proteins and assumed that individual secondary structure had a unique spectrum, and the total protein spectrum for each sample is the sum of each secondary structure spectrum linearly weighted by its content. The study was also performed using FTIR which returned results with greater error than 2D-IR, highlighting the greater sensitivity the technique has for secondary structures. This is a method that works well by comparing samples measured in the same conditions but would deviate if different environmental conditions were used such as buffer or pH.

MVA of 2D-IR spectra has also been performed on a large dataset of DNA spectra by a technique called analysis of variance combined with PCA (ANOVA-PCA)[116]. The spectra consist of different DNA sequences, both bound and unbound to a small molecule, Hoechst 33258, and all measured at different waiting times. The study showed that the DNA

sequences could be determined along with if the biomolecule was bound or unbound. Finally, the vibrational lifetime for each system was determined from the ANOVA-PCA method. This highlights how sophisticated analysis methods can be utilised for interpreting large datasets, which is required as the development of 2D-IR laser systems continue. Development of 2D-IR laser systems that are able to measure spectra quickly, such as the LIFEtime[117] laser system at the STFC facility at the Rutherford Appleton Laboratory (RAL), which can measure a 2D-IR spectrum in less than ten seconds, means it is possible to obtain large quantities of data very quickly. Therefore, a key requirement is developing processing methods to facilitate the extraction of information from 2D-IR spectra.

This thesis will further the knowledge of time resolved IR spectroscopy of proteins using the amide I region by using PCA to help quantify changes in secondary structure upon heating. Subsequently, a T-jump TRIR setup will be developed and used to probe the early stages of domain melting and then small changes in protein structure will be probed upon ligand binding using PCA.

1.4 References

(1) Littlechild, J. A. Protein Structure and Function. In *Introduction to Biological and Small Molecule Drug Research and Development: Theory and Case Studies*; 2013.

(2) Wright, P. E. 50 Years of Protein Structure Determination. *J. Mol. Biol.* **2009**, *392* (1), 1.

(3) Richards, F. M. Areas, Volumes, Packing and Protein Structure. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151–176.

(4) Feher, J. *Quantitative Human Physiology: An Introduction*; 2016.

(5) Walker, J. Protein Biotechnology: Isolation, Characterisation and Stabilization. *FEBS Lett.* **1994**, *337* (3), 308–308.

(6) Marks, D. S.; Hopf, T. A.; Sander, C. Protein Structure Prediction from Sequence Variation. *Nat. Biotechnol.* **2012**, *30*, 1072–1080.

(7) Congreve, M.; Murray, C. W.; Blundell, T. L. Keynote Review: Structural Biology and Drug Discovery. *Drug Discov. Today* **2005**, *10* (13), 895–907.

(8) Śledź, P.; Caflisch, A. Protein Structure-Based Drug Design: From Docking to Molecular Dynamics. *Curr. Opin. Struct. Biol.* **2018**, *48*, 93–102.

(9) Keller, T. H.; Pichota, A.; Yin, Z. A Practical View of "Druggability." *Curr. Opin. Chem. Biol.* **2006**, *10* (4), 357–361.

(10) Hajduk, P. J.; Huth, J. R.; Tse, C. Predicting Protein Druggability. *Drug Discov. Today* **2005**, *10* (23–24), 1675–1682.

(11) Mahmood, I.; Green, M. D. Pharmacokinetic and Pharmacodynamic Considerations in the Development of Therapeutic Proteins. *Clin. Pharmacokinet.* **2005**, *42*, 553–571.

(12) Blundell, T.; Sibanda, B. L.; Pearl, L. Three-Dimensional Structure, Specificity and Catalytic Mechanism of Renin. *Nature* **1983**, *304*, 273–275.

(13) Varghese, J. N. Development of Neuraminidase Inhibitors as Anti-Influenza Virus Drugs. *Drug Dev. Res.* **1999**, *46*, 176–196.

(14) Quinlan, G. J.; Martin, G. S.; Evans, T. W. Albumin: Biochemical Properties and Therapeutic Potential. *Hepatology* **2005**, *41*, 1211–1219.

(15) Widmer, N.; Decosterd, L. A.; Csajka, C.; Leyvraz, S.; Duchosal, M. A.; Rosselet, A.; Rochat, B.; Eap, C. B.; Henry, H.; Biollaz, J.; et al. Population Pharmacokinetics of Imatinib and the Role of A1-Acid Glycoprotein. *Br. J. Clin. Pharmacol.* **2006**, *62*, 97–112.

(16) Nishikawa, M.; Ohtsubo, Y.; Ohno, J.; Fujita, T.; Koyama, Y.; Yamashita, F.; Hashida, M.; Sezaki, H. Pharmacokinetics of Receptor-Mediated Hepatic Uptake of Glycosylated Albumin in Mice. *Int. J. Pharm.* **1992**, *85*, 75–85.

(17) Dengler, T. J.; Rohertz-Vaupel, G. M.; Dengler, H. J. Albumin Binding in Uraemia: Quantitative Assessment of Inhibition by Endogenous Ligands and Carbamylation of Albumin. *Eur. J. Clin. Pharmacol.* **1992**, *43*, 491–499.

(18) Cheung, W. Y. Cyclic 3',5'-Nucleotide Phosphodiesterase. Demonstration of an Activator. *Biochem. Biophys. Res. Commun.* **1970**, *38*, 533–538.

(19) Kakiuchi, S.; Yamazaki, R. Calcium Dependent Phosphodiesterase Activity and Its Activating Factor (PAF) from Brain. Studies on Cyclic 3',5'-Nucleotide Phosphodiesterase (III). *Biochem. Biophys. Res. Commun.* **1970**, *41*, 1104–1110.

(20) Wallace, R. W.; Cheung, W. Y. Calmodulin. Production of an Antibody in Rabbit and Development of a Radioimmunoassay. *J. Biol. Chem.* **1979**, *254*, 6564–6571.

(21) Means, A. R.; VanBerkum, M. F. A.; Bagchi, I.; Kun Ping Lu; Rasmussen, C. D. Regulatory Functions of Calmodulin. *Pharmacol. Ther.* **1991**, *50*, 255–270.

(22) Brostrom, C. O.; Wolff, D. J. Properties and Functions of Calmodulin. *Biochem. Pharmacol.* **1981**, *30*, 1395–1405.

(23) Itoh, T.; Ikebe, M.; Kargacin, G. J.; Hartshorne, D. J.; Kemp, B. E.; Fay, F. S. Effects of Modulators of Myosin Light-Chain Kinase Activity in Single Smooth Muscle Cells. *Nature* **1989**, *338*, 164–167.

(24) Tebar, F.; Villalonga, P.; Sorkina, T.; Agell, N.; Sorkin, A.; Enrich, C. Calmodulin Regulates Intracellular Trafficking of Epidermal Growth Factor Receptor and the MAPK Signaling Pathway. *Mol. Biol. Cell* **2002**, *13*, 2057–2068.

(25) Means, A. R. Calcium, Calmodulin and Cell Cycle Regulation. *FEBS Lett.* **1994**, *347*, 1–4.

(26) Halling, D. B.; Liebeskind, B. J.; Hall, A. W.; Aldrich, R. W. Conserved Properties of Individual Ca2+-Binding Sites in Calmodulin. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (9), 1216–1225.

(27) Park, H. Y.; Kim, S. A.; Korlach, J.; Rhoades, E.; Kwok, L. W.; Zipfel, W. R.; Waxham, M. N.; Webb, W. W.; Pollack, L. Conformational Changes of Calmodulin upon Ca(2+) Binding Studied with a Microfluidic Mixer. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (2), 542–547.

(28) Trewhella, J.; Liddle, W. K.; Heidorn, D. B.; Strynadkaf, N. Calmodulin and Troponin C Structures Studied by Fourier Transform Infrared Spectroscopy : Effects of Ca 2 + and Mg2 + Binding. *Biochemistry* **1989**, *28* (1982), 1294–1301.

(29) Hayashi, N.; Matsubara, M.; Takasaki, A.; Titani, K.; Taniguchi, H. An Expression System of Rat Calmodulin Using T7 Phage Promoter in Escherichia Coli. *Protein Expr. Purif.* **1998**, *12*, 25–28.

(30) Li, X. J.; Wu, J. G.; Si, J. L.; Guo, D. W.; Xu, J. P. High-Level Expression of Human Calmodulin in E. Coli and Its Effects on Cell Proliferation. *World J Gastroenterol* **2001**, *6*, 588–592.

(31) Spoel, D. Van Der; Groot, B. L. De; Hayward, S.; Berendsen, H. J. C.; Vogel, H. J. Bending of the Calmodulin Central Helix: A Theoretical Study. *Protein Sci.* **1996**, *5*, 2044–2053.

(32) Kretsinger, R. H.; Nockolds, C. E. Carp Muscle Calcium-Binding Protein. II. Structure Determination and General Description. *J. Biol. Chem.* **1973**, *248*, 3313–3326.

(33) Nakayama, S.; Kretsinger, R. H. Evolution of the EF-Hand Family of Proteins. *Annu. Rev. Biophys. Biomol. Struct.* **1994**, *23*, 473–507.

(34) Lewit-Bentley, A.; Réty, S. EF-Hand Calcium-Binding Proteins. *Curr. Opin. Struct. Biol.* **2000**, *10*, 637–643.

(35) Takeda, S.; Yamashita, A.; Maeda, K.; Maéda, Y. Structure of the Core Domain of Human Cardiac Troponin in the Ca2+-Saturated Form. *Nature* **2003**, *424*, 35–41.

(36) Heidorn, D. B.; Trewhella, J. Comparison of the Crystal and Solution Structures of Calmodulin and Troponin C. *Biochemistry* **1988**, *27* (3), 909–915.

(37) Barbato, G.; Ikura, M.; Kay, L. E.; Pastor, R. W.; Bax, A. Backbone Dynamics of Calmodulin Studied by 15N Relaxation Using Inverse Detected Two-Dimensional NMR Spectroscopy: The Central Helix Is Flexible. *Biochemistry* **1992**, *31*, 5269–5278.

(38) Ikura, M.; Spera, S.; Barbato, G.; Kay, L. E.; Krinks, M.; Bax, A. Secondary Structure and Side-Chain H-1 and C-13 Resonance Assignments of Calmodulin in Solution By Heteronuclear Multidimensional NMR-Spectroscopy. *Biochemistry* **1991**, *30*, 9216–9228.

(39) Kuboniwa, H.; Tjandra, N.; Grzesiek, S.; Ren, H.; Klee, C. B.; Bax, A. Solution Structure of Calcium-Free Calmodulin. *Nat. Struct. Mol. Biol.* **1995**, *2*, 768–776.

(40) Linse, S.; Helmersson, A.; Forsen, S. Calcium Binding to Calmodulin and Its Globular Domains. *J. Biol. Chem.* **1991**, *266*, 8050–8054.

(41) LaPorte, D. C.; Wierman, B. M.; Storm, D. R. Calcium-Induced Exposure of a Hydrophobic Surface on Calmodulin. *Biochemistry* **1980**, *19*, 3814–3819.

(42) Zhang, M.; Tanaka, T.; Ikura, M. Calcium-Induced Conformational Transition Revealed by the Solution Structure of Apo Calmodulin. *Nat. Struct. Biol.* **1995**, *2* (9),

758–767.

(43) Finn, B. E.; Forsén, S. The Evolving Model of Calmodulin Structure,Function and Activation. *Structure* **1995**, *3*, 7–11.

(44) Chattopadhyaya, R.; Meador, W. E.; Means, A. R.; Quiocho, F. A. Calmodulin Structure Refined at 1.7 Å Resolution. *J. Mol. Biol.* **1992**, *228*, 1177–1192.

(45) Browne, J. P.; Strom, M.; Martin, S. R.; Bayley, P. M. The Role of Beta-Sheet Interactions in Domain Stability, Folding, and Target Recognition Reactions of Calmodulin. *Biochemistry* **1997**, *36* (31), 9550–9561.

(46) Jackson, M.; Haris, P. I.; Chapman, D. Fourier Transform Infrared Spectroscopic Studies of Ca(2+) -Binding Proteins. *Biochemistry* **1991**, *30*, 9681–9686.

(47) Brzeska, H.; Venyaminov, S. V.; Grabarek, Z.; Drabikowski, W. Comparative Studies on Thermostability of Calmodulin, Skeletal Muscle Troponin C and Their Tryptic Fragments. *FEBS Lett.* **1983**, *153* (1), 169–173.

(48) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The Protein Data Bank: A Computer-Based Archival File for Macromolecular Structures. *Arch. Biochem. Biophys.* **1978**, *112*, 535–542.

(49) Brink, C.; Hodgkin, D. C.; Lindsey, J.; Pickworth, J.; Robertson, J. H.; White, J. G. Structure of Vitamin B 12 : X-Ray Crystallographic Evidence on the Structure of Vitamin B 12. *Nature* **1954**, *174*, 1169–1171.

(50) Blundell, T. L.; Dodson, G. G.; Dodson, E.; Hodgkin, D. C.; Vijayan, M. X-Ray Analysis and the Structure of Insulin. *Recent Progress in Hormone Research*. 1971.

(51) Drenth, J.; Mesters, J. *Principles of Protein X-Ray Crystallography: Third Edition*; 2007.

(52) Gorfman, S. Sub-Microsecond X-Ray Crystallography: Techniques, Challenges, and Applications for Materials Science. *Crystallogr. Rev.* **2014**, *20*, 210–232.

(53) Pusey, M. L.; Liu, Z. J.; Tempel, W.; Praissman, J.; Lin, D.; Wang, B. C.; Gavira, J. A.; Ng, J. D. Life in the Fast Lane for Protein Crystallization and X-Ray Crystallography. *Prog. Biophys. Mol. Biol.* **2005**, *88*, 359–386.

(54) Johnson, W. C. Secondary Structure of Proteins through Circular Dichroism Spectroscopy. *Annu. Rev. Biophys. Biophys. Chem.* **1988**, *17*, 145–166.

(55) Johnson, W. C. Protein Secondary Structure and Circular Dichroism: A Practical Guide. *Proteins Struct. Funct. Bioinforma.* **1990**, *7*, 205–214.

(56) Whitmore, L.; Wallace, B. A. Protein Secondary Structure Analyses from Circular Dichroism Spectroscopy: Methods and Reference Databases. *Biopolymers* **2008**, *89* (5), 392–400.

(57) Greenfield, N. J. Using Circular Dichroism Spectra to Estimate Protein Secondary Structure. *Nat. Protoc.* **2007**, *1*, 2876–2890.

(58) Wei, Y.; Thyparambil, A. A.; Latour, R. A. Protein Helical Structure Determination Using CD Spectroscopy for Solutions with Strong Background Absorbance from 190

to 230 Nm. *Biochim. Biophys. Acta - Proteins Proteomics* **2014**, *1844*, 2331–2337.

(59) Dror, R. O.; Dirks, R. M.; Grossman, J. P.; Xu, H.; Shaw, D. E. Biomolecular Simulation: A Computational Microscope for Molecular Biology. *Annu. Rev. Biophys.* **2012**, *41*, 429–452.

(60) Chen, J.; Brooks, C. L. Can Molecular Dynamics Simulations Provide High-Resolution Refinement of Protein Structure? *Proteins Struct. Funct. Genet.* **2007**, *67*, 922–930.

(61) Ode, H.; Nakashima, M.; Kitamura, S.; Sugiura, W.; Sato, H. Molecular Dynamics Simulation in Virus Research. *Front. Microbiol.* **2012**, *3*, 258(1)-(9).

(62) Kay, L. E. NMR Studies of Protein Structure and Dynamics. *J. Magn. Reson.* **2005**, *173*, 193–207.

(63) Boehr, D. D.; Dyson, H. J.; Wright, P. E. An NMR Perspective on Enzyme Dynamics. *Chem. Rev.* **2006**, *106*, 3055–3079.

(64) Crowley, P. B.; Ubbink, M. Close Encounters of the Transient Kind: Protein Interactions in the Photosynthetic Redox Chain Investigated by NMR Spectroscopy. *Acc. Chem. Res.* **2003**, *36*, 723–730.

(65) Wüthrich, K. Protein Structure Determination in Solution by NMR Spectroscopy. *J. Biol. Chem.* **1990**, *265*, 22059–22062.

(66) Clore, G. M.; Gronenborn, A. M. Structures of Larger Proteins in Solution: Three- and Four-Dimensional Heteronuclear NMR Spectroscopy. *Science.* **1991**, *252*, 1390–1399.

(67) Cavanagh, J.; Fairbrother, W.; III, A. P.; Skelton, N. *Protein NMR Spectroscopy, 2nd Edition Principles and Practice*; 2007.

(68) Jeener, J.; Meier, B. H.; Bachmann, P.; Ernst, R. R. Investigation of Exchange Processes by Two-Dimensional NMR Spectroscopy. *J. Chem. Phys.* **1979**, *71*, 4546–4553.

(69) Jiang, Y.; Kalodimos, C. G. NMR Studies of Large Proteins. *J. Mol. Biol.* **2017**, *429*, 2667–2676.

(70) Sheppard, N. The Historical Development of Experimental Techniques in Vibrational Spectroscopy. In *Handbook of Vibrational Spectroscopy*; 2006.

(71) Barth, A.; Zscherp, C. What Vibrations Tell about Proteins. *Q. Rev. Biophys.* **2002**, *35*, 369–430.

(72) Barth, A. Infrared Spectroscopy of Proteins. *Biochim. Biophys. Acta - Bioenerg.* **2007**, *1767*, 1073–1101.

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

(74) Abe, Y.; Krimm, S. Normal Vibrations of Crystalline Polyglycine I. *Biopolymers* **1972**, *9*, 1817–1839.

(75) Nevskaya, N. A.; Chirgadze, Y. N. Infrared Spectra and Resonance Interactions of Amide-I and II Vibrations of A-helix. *Biopolymers* **1976**, *15*, 637–648.

(76) Krimm, S.; Bandekar, J. Vibrational Spectroscopy and Conformation of Peptides, Polypeptides, and Proteins. *Adv. Protein Chem.* **1986**, *38*, 181–364.

(77) Manas, E. S.; Getahun, Z.; Wright, W. W.; Degrado, W. F.; Vanderkooi, J. M. Infrared Spectra of Amide Groups in α-Helical Proteins: Evidence for Hydrogen Bonding between Helices and Water. *J. Am. Chem. Soc.* **2000**, *122*, 9883–9890.

(78) Raschke, T. M.; Marqusee, S. Hydrogen Exchange Studies of Protein Structure. *Curr. Opin. Biotechnol.* **1998**, *9*, 80–86.

(79) Militello, V.; Casarino, C.; Emanuele, A.; Giostra, A.; Pullara, F.; Leone, M. Aggregation Kinetics of Bovine Serum Albumin Studied by FTIR Spectroscopy and Light Scattering. *Biophys. Chem.* **2004**, *107*, 175–187.

(80) Haris, P. I.; Severcan, F. FTIR Spectroscopic Characterization of Protein Structure in Aqueous and Non-Aqueous Media. *J. Mol. Catal. - B Enzym.* **1999**, *7*, 207–221.

(81) Sagle, L. B.; Zhang, Y.; Litosh, V. A.; Chen, X.; Cho, Y.; Cremer, P. S. Investigating the Hydrogen-Bonding Model of Urea Denaturation. *J. Am. Chem. Soc.* **2009**, *131*, 9304–9310.

(82) Byler, D. M.; Susi, H. Examination of the Secondary Structure of Proteins by Deconvolved FTIR Spectra. *Biopolymers* **1986**, *25*, 469–487.

(83) Taddei, P.; Monti, P. Vibrational Infrared Conformational Studies of Model Peptides Representing the Semicrystalline Domains of Bombyx Mori Silk Fibroin. *Biopolymers* **2005**, *78*, 249–258.

(84) Haris, P. I.; Chapman, D.; Benga, G. A Fourier-Transform Infrared Spectroscopic Investigation of the Hydrogen-Deuterium Exchange and Secondary Structure of the 28-kDa Channel-Forming Integral Membrane Protein (CHIP28). *Eur. J. Biochem.* **1995**, *233*, 659–664.

(85) Jackson, M.; Mantsch, H. H. The Use and Misuse of FTIR Spectroscopy in the Determination of Protein Structure. *Crit. Rev. Biochem. Mol. Biol.* **1995**, *30*, 95–120.

(86) Goormaghtigh, E.; Ruysschaert, J. M.; Raussens, V. Evaluation of the Information Content in Infrared Spectra for Protein Secondary Structure Determination. *Biophys. J.* **2006**, *90*, 2946–2957.

(87) Tooke, P. B. Fourier Self-Deconvolution in IR Spectroscopy. *Trends Anal. Chem.* **1988**, *7*, 130–136.

(88) Hu, X.; Kaplan, D.; Cebe, P. Determining Beta-Sheet Crystallinity in Fibrous Proteins by Thermal Analysis and Infrared Spectroscopy. *Macromolecules* **2006**, *39*, 6161–6170.

(89) Mouro, C.; Jung, C.; Bondon, A.; Simonneaux, G. Comparative Fourier Transform Infrared Studies of the Secondary Structure and the CO Heme Ligand Environment in Cytochrome P-450cam and Cytochrome P-420cam. *Biochemistry* **1997**, *36*, 8125–8134.

(90) Zscherp, C.; Barth, A. Reaction-Induced Infrared Difference Spectroscopy for the Study of Protein Reaction Mechanisms. *Biochemistry* **2001**, *40*, 1875–1883.

(91) Mäntele, W. Reaction-Induced Infrared Difference Spectroscopy for the Study of

Protein Function and Reaction Mechanisms. *Trends Biochem. Sci.* **1993**, *18*, 197–202.

(92) Rieppo, L.; Saarakkala, S.; Närhi, T.; Helminen, H. J.; Jurvelin, J. S.; Rieppo, J. Application of Second Derivative Spectroscopy for Increasing Molecular Specificity of Fourier Transform Infrared Spectroscopic Imaging of Articular Cartilage. *Osteoarthr. Cartil.* **2012**, *20*, 451–459.

(93) Ollesch, J.; Drees, S. L.; Heise, H. M.; Behrens, T.; Brüning, T.; Gerwert, K. FTIR Spectroscopy of Biofluids Revisited: An Automated Approach to Spectral Biomarker Identification. *Analyst* **2013**, *138*, 4092–4102.

(94) Rajalahti, T.; Kvalheim, O. M. Multivariate Data Analysis in Pharmaceutics: A Tutorial Review. *Int. J. Pharm.* **2011**, *417*, 280–290.

(95) Karamizadeh, S.; Abdullah, S. M.; Manaf, A. A.; Zamani, M.; Hooman, A. An Overview of Principal Component Analysis. *J. Signal Inf. Process.* **2013**, *4*, 173–175.

(96) Etzion, Y.; Linker, R.; Cogan, U.; Shmulevich, I. Determination of Protein Concentration in Raw Milk by Mid-Infrared Fourier Transform Infrared/Attenuated Total Reflectance Spectroscopy. *J. Dairy Sci.* **2004**, *87*, 2779–2788.

(97) Hamm, P.; Lim, M.; Hochstrasser, R. M. Structure of the Amide I Band of Peptides Measured by Femtosecond Nonlinear-Infrared Spectroscopy. *J. Phys. Chem. B* **1998**, *102* (98), 6123–6138.

(98) Hamm, P.; Zanni, M. *Concepts and Methods of 2D Infrared Spectroscopy*; 2011.

(99) Hunt, N. T. 2D-IR Spectroscopy: Ultrafast Insights into Biomolecule Structure and Function. *Chem. Soc. Rev.* **2009**, *38*, 1837–1848.

(100) Cervetto, V.; Helbing, J.; Bredenbeck, J.; Hamm, P. Double-Resonance versus Pulsed Fourier Transform Two-Dimensional Infrared Spectroscopy: An Experimental and Theoretical Comparison. *J. Chem. Phys.* **2004**, *121*, 5935–5942.

(101) Le Sueur, A. L.; Horness, R. E.; Thielges, M. C. Applications of Two-Dimensional Infrared Spectroscopy. *Analyst* **2015**, *140*, 4336–4349.

(102) Stevenson, P.; Tokmakoff, A. Distinguishing Gramicidin D Conformers through Two-Dimensional Infrared Spectroscopy of Vibrational Excitons. *J. Chem. Phys.* **2015**, *142*, 212424.

(103) Cheatum, C. M.; Tokmakoff, A.; Knoester, J. Signatures of β-Sheet Secondary Structures in Linear and Two-Dimensional Infrared Spectroscopy. *J. Chem. Phys.* **2004**, *120*, 8201–8215.

(104) Smith, A. W.; Tokmakoff, A. Amide I Two-Dimensional Infrared Spectroscopy of ß-Hairpin Peptides. *J. Chem. Phys.* **2007**, *126*, 045109/1-045109/11.

(105) Demirdöven, N.; Cheatum, C. M.; Chung, H. S.; Khalil, M.; Knoester, J.; Tokmakoff, A. Two-Dimensional Infrared Spectroscopy of Antiparallel Beta-Sheet Secondary Structure. *J. Am. Chem. Soc.* **2004**, *126* (25), 7981–7990.

(106) Fang, C.; Wang, J.; Charnley, A. K.; Barber-Armstrong, W.; Smith, A. B.; Decatur, S. M.; Hochstrasser, R. M. Two-Dimensional Infrared Measurements of the Coupling between Amide Modes of an α-Helix. *Chem. Phys. Lett.* **2003**, *382*, 586–592.

(107) Grechko, M.; Zanni, M. T. Quantification of Transition Dipole Strengths Using 1D and 2D Spectroscopy for the Identification of Molecular Structures via Exciton Delocalization: Application to Alpha-Helices. *J. Chem. Phys.* **2012**, *137* (18), 184202/1-184202/9.

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

(109) Hume, S.; Hithell, G.; Greetham, G. M.; Donaldson, P. M.; Towrie, M.; Parker, A. W.; Baker, M. J.; Hunt, N. T. Measuring Proteins in H2O with 2D-IR Spectroscopy. *Chem. Sci.* **2019**, *10*, 6448–6456.

(110) Bredenbeck, J.; Helbing, J.; Kolano, C.; Hamm, P. Ultrafast 2D-IR Spectroscopy of Transient Species. *ChemPhysChem* **2007**, *8*, 1747–1756.

(111) Kolano, C.; Helbing, J.; Kozinski, M.; Sander, W.; Hamm, P. Watching Hydrogen-Bond Dynamics in a β-Turn by Transient Two-Dimensional Infrared Spectroscopy. *Nature* **2006**, *444*, 469–472.

(112) Donten, M. L.; Hamm, P. PH-Jump Induced α-Helix Folding of Poly-L-Glutamic Acid. *Chem. Phys.* **2013**, *422*, 124–130.

(113) Chung, H. S.; Khalil, M.; Smith, A. W.; Tokmakoff, A. Transient Two-Dimensional IR Spectrometer for Probing Nanosecond Temperature-Jump Kinetics. *Rev. Sci. Instrum.* **2007**, *78* (6), 063101/1-063101/10.

(114) Zhang, X.; Jones, K. C.; Fitzpatrick, A.; Peng, C. S.; Feng, C.; Baiz, C. R.; Tokmako, A. Studying Protein − Protein Binding through T - Jump Induced Dissociation : Transient 2D IR Spectroscopy of Insulin Dimer. *J. Phys. Chem. B* **2016**, *120*, 5134–5145.

(115) Baiz, C. R.; Peng, C. S.; Reppert, M. E.; Jones, K. C.; Tokmakoff, A. Coherent Two-Dimensional Infrared Spectroscopy: Quantitative Analysis of Protein Secondary Structure in Solution. *Analyst* **2012**, *137* (8), 1793–1799.

(116) Fritzsch, R.; Donaldson, P. M.; Greetham, G. M.; Towrie, M.; Parker, A. W.; Baker, M. J.; Hunt, N. T. Rapid Screening of DNA-Ligand Complexes via 2D-IR Spectroscopy and ANOVA-PCA. *Anal. Chem.* **2018**, *90*, 2732–2740.

(117) Greetham, G. M.; Donaldson, P. M.; Nation, C.; Sazanovich, I. V.; Clark, I. P.; Shaw, D. J.; Parker, A. W.; Towrie, M. A 100 KHz Time-Resolved Multiple-Probe Femtosecond to Second Infrared Absorption Spectrometer. *Appl. Spectrosc.* **2016**, *70*, 645–653.

2 Methods

2.1 Calmodulin expression and purification

The protein calmodulin (CaM) was used in chapters (3) – (5). It was expressed and purified in collaboration with UCB, Slough.

Escherichia coli DH5α (Invitrogen) competent cells were used in the transformation of human CaM (*h.calmodulin*) plasmid DNA and *Escherichia coli* BL21 (DE3) Star (Novagen) competent cells were used in the expression of the protein.

H.calmodulin plasmid DNA was transformed using DH5α competent cells in super optimal broth with catabolite repression (SOC) media and using BL21 (DE3) competent cells, incubated overnight in 2YT broth with Carb 100 antibiotic at 37 °C, 220 RPM. The resultant culture was over-expressed (2YT broth with additions of Carb 100) and monitored by A_{600} until a value of 0.4 OD was observed at which it was induced with Isopropyl-thio-β-D-galactopyranoside (IPTG) to make a final concentration of 0.5 mM and left to cultivate over 5-6 hours at 37 °C until an A_{600} OD of 2.5 was observed. The inoculated media was then centrifuged (Beckman centrifuge, JA 90) at 5,000 RPM for 15 minutes at 4 °C. The supernatant was discarded, and the pellets were re-dissolved in 50 mL of suspension buffer (50 mM Tris-HCl pH 7.5, 2 mM EDTA). Lysis was performed using a French high-pressure cell at pressure of ~1000 p.s.i. The collected lysis was centrifuged at 10,000 RPM for 30 minutes at 4 °C. The supernatant was collected and subjected to AKTA purification using a HiTrap Phenyl column (Phenyl FF, 5mL). Equilibration and elution buffers containing 50 mM Tris-HCl, pH 7.5,100 mM NaCl, 5 mM CaCl2 and 50 mM Tris-HCl, pH 7.5, 10 mM EDTA were used and the protein was penultimately purity checked via size-exclusion chromatography before final analysis using mass spectrometry to verify the monomeric molecular weight.

2.2 FTIR Spectroscopy

A Fourier transform infrared (FTIR) spectrometer uses a blackbody radiation source to generate IR light and a Helium:Neon (He:Ne) laser with a Michelson interferometer to measure IR intensity. The interferometer splits the IR radiation into two beams with the use of a beam splitter. One beam of the light is reflected off a moving mirror and the other beam is reflected off a static mirror. The two beams then recombine at the beam splitter and travel together to the sample (Figure 2-1). The beams will either combine constructively or destructively, depending on the path difference which is dependent on the moving mirror,

and the resulting oscillations will produce an interferogram that correlates each frequency component to a point in time.

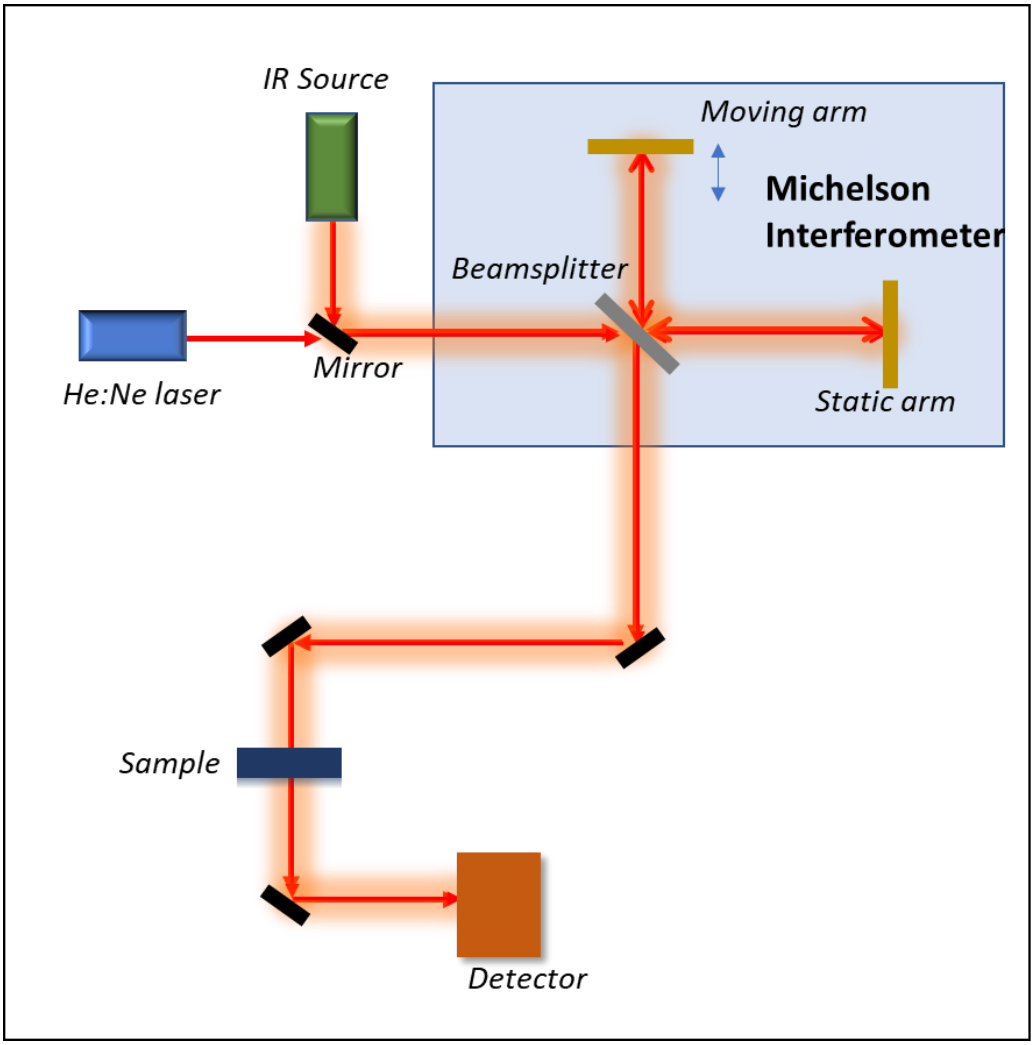


Figure 2-1. Typical schematic of FT-IR spectrometer setup.

The He:Ne laser is used as a reference beam alongside the IR beam to determine what the path difference is between the two beams. This information is required to calculate the frequency axis. The He:Ne beam follows the same paths as the IR light and the interference pattern generated by the 632.8 nm wavelength beam is detected and counted, providing a reference for the path difference.

An FTIR spectrometer gives a measure of the absorbance of a sample and it is related to the intensity of the signal before and after passing through the sample (Equation 2.1).

$$A = log_{10} \frac{I_0}{I} \qquad \text{(Equation 2.1)}$$

Where A is the absorbance measured in Optical Density (OD), I_0 is the intensity of the light before the sample, and I is the intensity of the light after the sample. Absorbance can also be calculated from the Beer-Lambert Law (Equation 2.2).

$$A = \varepsilon c l \qquad \text{(Equation 2.2)}$$

Where ε is the molar extinction coefficient (M^{-1} cm^{-1}), c is the concentration of the sample (mol l^{-1}), and l is the sample path length (cm).

The concentrations and path lengths of the samples used throughout this thesis were chosen to give good intensities of the peaks of interest and specific values will be given in each experimental chapter.

2.3 IR sample setup

Every IR sample in this thesis was prepared in the same manner within a Harrick cell (Figure 2-2). Approximately 30 µL of sample was placed between two CaF_2 windows. A polytetrafluoroethylene (PTFE) spacer of known thickness was used to separate the windows and provides the path length of the sample. The cells were either static, for room temperature measurements, or temperature dependent cells controlled by an external power unit. The same sample setup was used for the time-resolved measurements as well as the FTIR measurements.

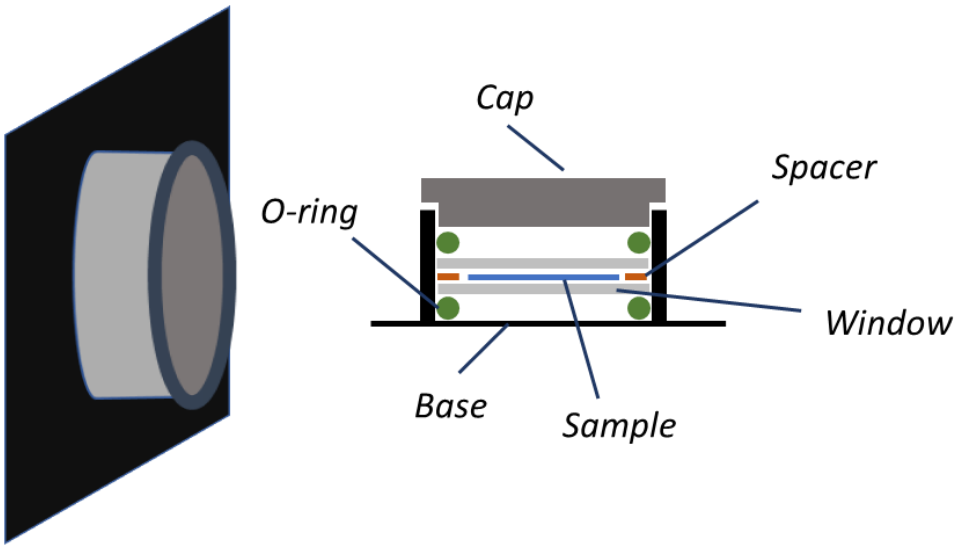


Figure 2-2. Schematic of prepared IR sample housed in a Harrick cell (left) with an example cross section of the cell setup (right).

2.4 2D-IR Spectroscopy

The 2D-IR spectroscopy data used throughout this thesis were collected on the ULTRA femtosecond laser system based at the STFC Central Laser Facility at the Rutherford Appleton Laboratory (RAL) (Harwell)[1,2].

A custom built titanium sapphire (Ti:Sapph) oscillator was used to generate 800 nm pulsed light, with a 20 fs duration, a 50 nm bandwidth and a repetition rate of 65 MHz. The short pulses were then amplified using a custom-built regenerative amplifier system. First the short pulses were stretched temporally using a grating. This was performed as amplification of short pulses can damage the Ti:Sapph crystal with high pulse peak intensities. The stretched pulse then enters the amplification cavity, with input and output controlled by electro-optic modulators (Pockel cells). The Ti:Sapph crystal was pumped by two Nd:YAG lasers each with approximately 55 W of power, and the crystal underwent cryogenic cooling to reduce thermal lensing.

Amplification in the cavity was achieved by numerous round trips through the Ti:Sapph and, once achieved, the Pockel cell allows exit from the cavity. After the pulse has exited the cavity it is compressed temporally with a second grating to obtain short pulse durations of 40 – 80 fs at 800 nm with a 10 kHz repetition rate.

After amplification the 800 nm pulses are converted into mid-IR light with the use of an optical parametric amplifier (OPA) and difference-frequency generation (DFG). In the case of this thesis, all the laser pulses produced from the DFG were centred at approximately 1650 cm^{-1} (~6 µm) for the amide I region with pulse duration of ~60 fs and pulse energy of 1-3 µJ. The output from the OPA to the DFG was not measured as the beam passed immediately into the DFG from the OPA.

The first step in optical parametric amplification is splitting the 800 nm light into two pump beams with the use of a beam splitter. One pump beam is used in the second stage of the amplification; the other is further split into two beams with another beam splitter in the ratio 90:10. The lower intensity beam is used to generate a white light continuum (WLC) through a sapphire plate. This WLC beam is then focused onto a 2^{nd} order nonlinear crystal (barium borate, BBO) along with the other portion of this split pump light. Optical parametric amplification occurs within this crystal which converts the pump 800 nm light into two beams

of lower energy, the signal and idler beams, where the idler beam is produced as a by-product of the conversion of energy from the pump to the signal:

$$\omega_3 = \omega_1 + \omega_2 \qquad \text{(Equation 2.3)}$$

Where ω_3 is the frequency of the input pump beam, ω_1 is the frequency of the signal beam, and ω_2 is the frequency of the idler beam. The frequency of the signal light generated can be controlled by altering the temporal delay between the pump beam and WLC beam and altering the angle of the crystal. The idler beam and residual pump beam are blocked after the crystal with a beam blocker while the signal beam is collimated and continues to the second step of amplification to generate more signal power.

As with the first step, the beam is passed through a BBO crystal, along with the other pump beam that was split by the initial beam splitter. The interaction of the signal and pump beams in the crystal amplifies the signal beam and produces a collinear idler beam.

The signal and idler beams then undergo mixing to produce the desired mid-IR wavelength through DFG with the use of a silver thiogallate ($AgGaS_2$) crystal. Tuning of the difference frequency pulse wavelength is achieved by rotating the crystal in the vertical plane and changing the angle between incoming beams.

After the OPA-DFG the generated mid-IR light is passed through a 90:10 beam splitter creating two beams, a pump beam and a probe beam, respectively. After splitting of the mid-IR light into two beams, two different laser setups were used on ULTRA with regards to the pump beam: an interferometer approach and a pulse shaper approach.

2.4.1 Interferometer Approach

All 2D-IR spectra collected in chapter (3) used an interferometer approach to collect the data. Spectra in the technical note of chapter (5) also used this approach. Both approaches used a "parallel" beam polarisation.

The set-up is a pseudo pump-probe geometry[3] whereby the pump beam is passed through an interferometer where a 50:50 beam splitter divides the beam for the fixed arm and the moving arm (Figure 2-3). The moving arm is controlled by a delay stage (Delay stage 2), and the static arm passes through a chopper, which chops the beam to half the repetition rate (5 kHz). The time delay between the two pump arms, t_1, is variable due to the delay stage.

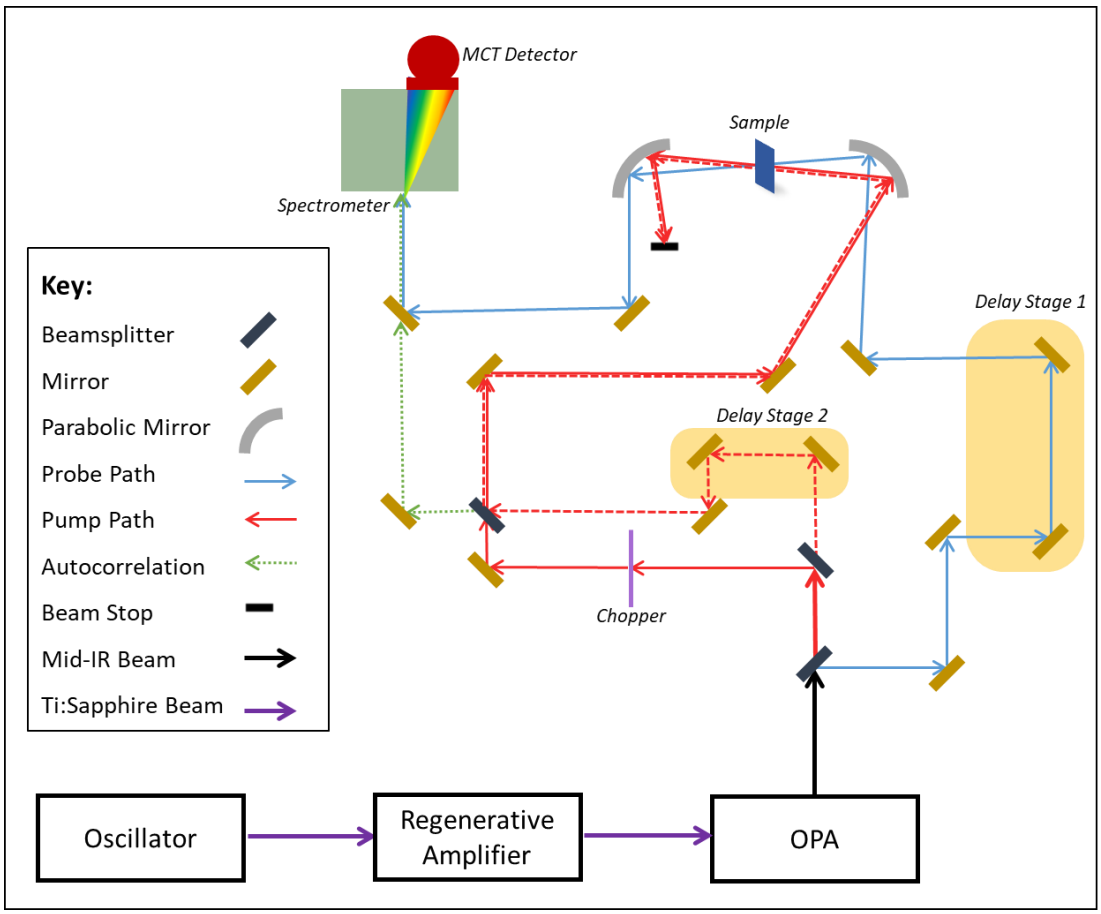


Figure 2-3. Schematic of the ULTRA laser optical setup with interferometer in the pump path.

The movements of the delay stage have to be precise as the step sizes it takes, along with the maximum time between the two pump pulses recorded, dictates the resolution of the pump frequency axis generated through Fourier-transformation of the interference pattern generated by the two pump pulses. This is otherwise known as the pump field autocorrelation (Figure 2-3, green dashed line) and was collected in parallel with the main spectrum. This interference pattern is also required to properly "phase" the 2D-IR spectrum. This will be discussed in further detail below. The value of t_1 where the two pump pulses are completely in phase and interfere constructively, results in an intensity maximum, a "centre-burst", in the field autocorrelation (Figure 2-4). This time is referred to as t_0. The pump light for the autocorrelation is directed onto a single pixel on the MCT detector.

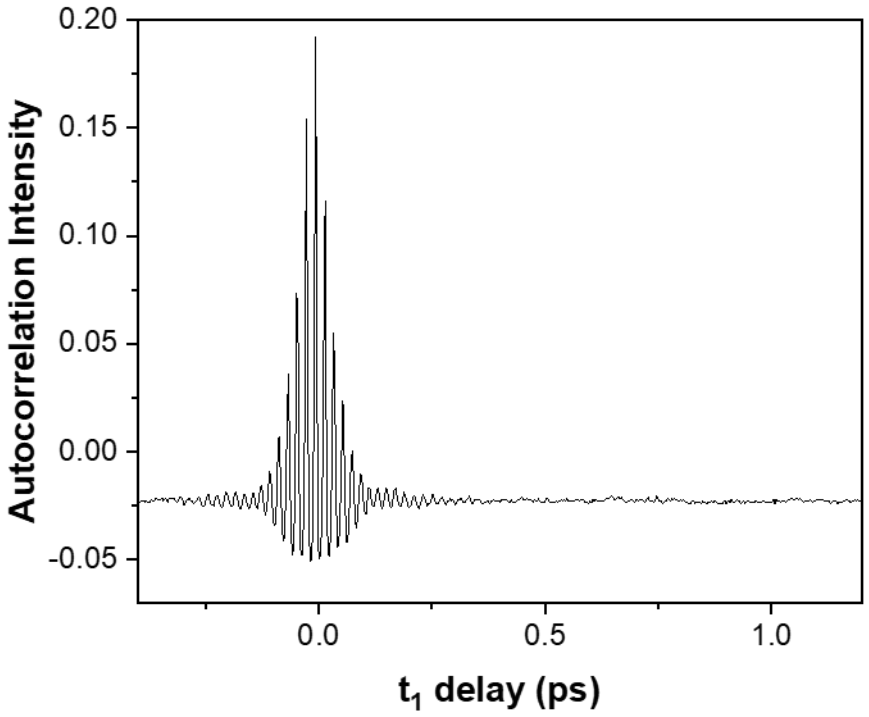


Figure 2-4. Field autocorrelation versus t_1 delay plot highlighting the "centre-burst" at t_0.

The step size is also important in defining the maximum frequency detected, also known as the Nyquist frequency (ω_N). This can be defined by the following equation:

$$\frac{\omega_N}{2\pi} = \frac{1}{2\Delta t} \qquad \text{(Equation 2.4)}$$

Where Δt is the step size. This means that for a vibrational mode with a frequency of 1650 cm^{-1} the vibrational timing will be approximately 19 fs. Therefore, steps sizes of less than 10 fs will be required to capture the frequency components of the emitted signal. The step sizes used in the interferometer approach 2D-IR measurements in this thesis were 3 fs, which is equivalent to movements of the delay stage 2 of approximately 0.45 µm. The maximum t_1 values recorded were 4 ps.

The probe beam is also controlled by a delay stage and altering this allows the time delay between the second pump pulse and the probe pulse to be controlled. This is defined as the waiting time (T_w).

After the interferometer stage, both the pump and probe beam are reflected off a parabolic mirror and directed towards the sample. Upon interaction with the sample the remaining

pump light is stopped by a beam block. The probe light and the emitted signal from the sample is collimated by another parabolic mirror and then directed towards the spectrometer which contains a grating to diffract the light onto a liquid nitrogen cooled 128-element MCT detector. The is the basis of the probe frequency axis for the 2D-IR spectrum. Resolution of the probe axis for all 2D-IR measurements is approximately 1.4 cm^{-1}. The spectrometer grating angle was adjusted to ensure the centre of the probe pulse was at the centre of the MCT pixels. An example of the signal on one pixel is shown in Figure 2-5. The system was purged with nitrogen gas to eliminate interferences and absorption of water vapour which could have a negative impact of the signal intensity.

The probe light is detected both with and without pump pulses interacting with the sample, which is achieved using the chopper. For each probe spectrum detected on the array detector the absorbance difference of pump on minus pump off is calculated. Fourier transformation of the free induction decay at each probe pixel is then calculated with respect to t_1 results in a spectrum for each pump frequency, and therefore a series of probe slices at each pump frequency are calculated and combined to generate a 2D-IR spectrum. This FT and generation of 2D-IR spectra is performed using custom data processing software in Labview created by Dr. Gregory M. Greetham. This was adapted from processes created by Prof Neil Hunt, Dr. Rafal Kania, and Dr. Katrin Adamczyk.

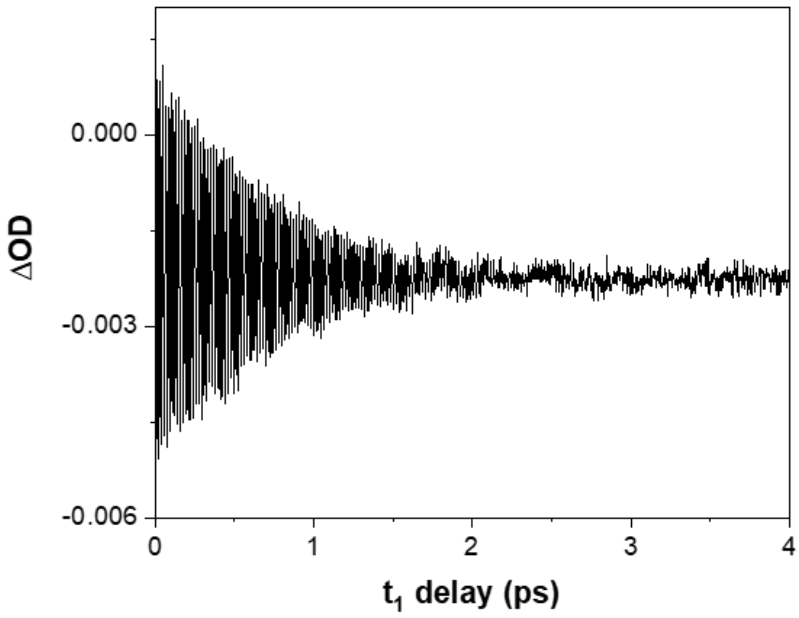


Figure 2-5. Signal collect from a single pixel during a 2D-IR acquisition.

2.4.1.1 2D-IR data processing for interferometer approach

This custom programme is used to generate the 2D-IR spectra by generating the pump frequency axis, finding t_0, and phasing the data.

The 2D-IR spectra presented in this thesis are the result of averaging of three scans with each taking approximately 15 minutes to obtain. Over the duration of the three scans beam drifting can occur that has to be compensated for. This is done by averaging the three field autocorrelation signals that are recorded from the three scans and is the first stage in the data processing.

The second stage is accurately determining t_0 using the averaged field autocorrelation signal. This is an important stage as signals recorded prior to t_0 can cause distortion of the 2D-IR lineshapes. An estimate of t_0 is first calculated by Fourier transforming the averaged field autocorrelation signal. At t_0 all the frequency components are in phase, therefore using the phase slope ($\delta\varphi/\delta t_1$) generated from the FT, the moment where the slope is zero is the estimated t_0 value. An accurate t_0 is then determined using the projection slice theorem[4,5].

The field autocorrelation data are interpolated to allow for accurate tuning of t_0. A pump-probe signal is collected during the measurement, from an interaction of the second pump pulse and the probe pulse. This pump probe signal is removed from the 2D-IR signal during the FT of the signals on the probe pixel, and therefore has no influence on the resulting 2D-IR spectrum. However, the pump probe signal is extracted during the data processing and compared to a projected 2D-IR spectrum. Phasing of the spectrum through adjustments of the interpolated data points is performed until the projected spectrum overlaps with the pump-probe signal (Figure 2-6). Incorrect determination of t_0 produces out of phase pump-probe spectra and therefore results in distortion to the 2D-IR lineshapes (Figure 2-7, same data used as shown in Figure 2-6).

The pump axis frequency resolution was also increased in the software through the use of zero-padding. This increases the t_{max} which is subsequently Fourier transformed.

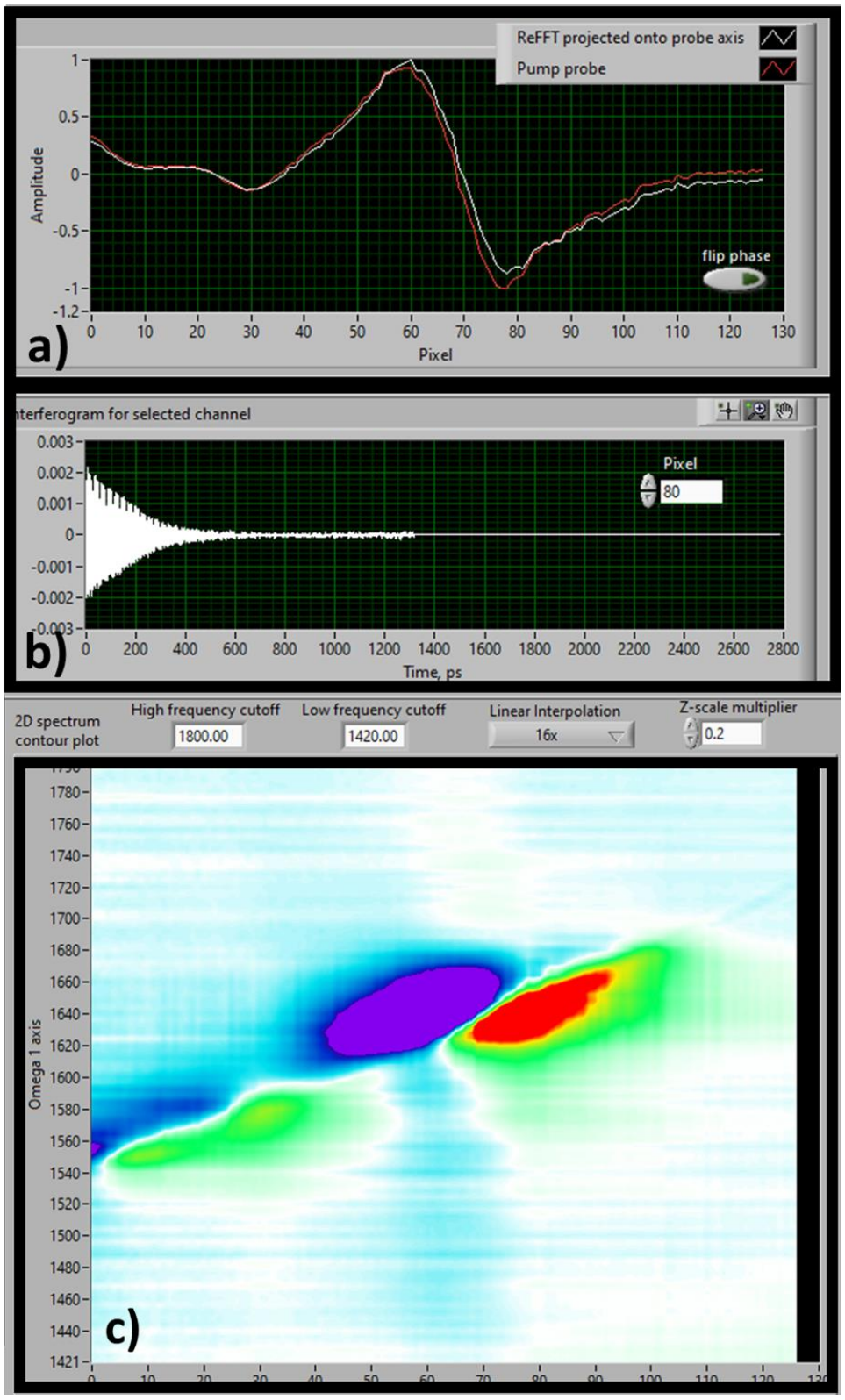


Figure 2-6. Example of Labview processing software phasing a 2D-IR spectrum by a) comparing the expected pump-probe spectrum to the experimental. b) interferogram for each pixel is determined. c) the resulting spectrum from the phasing of the data.

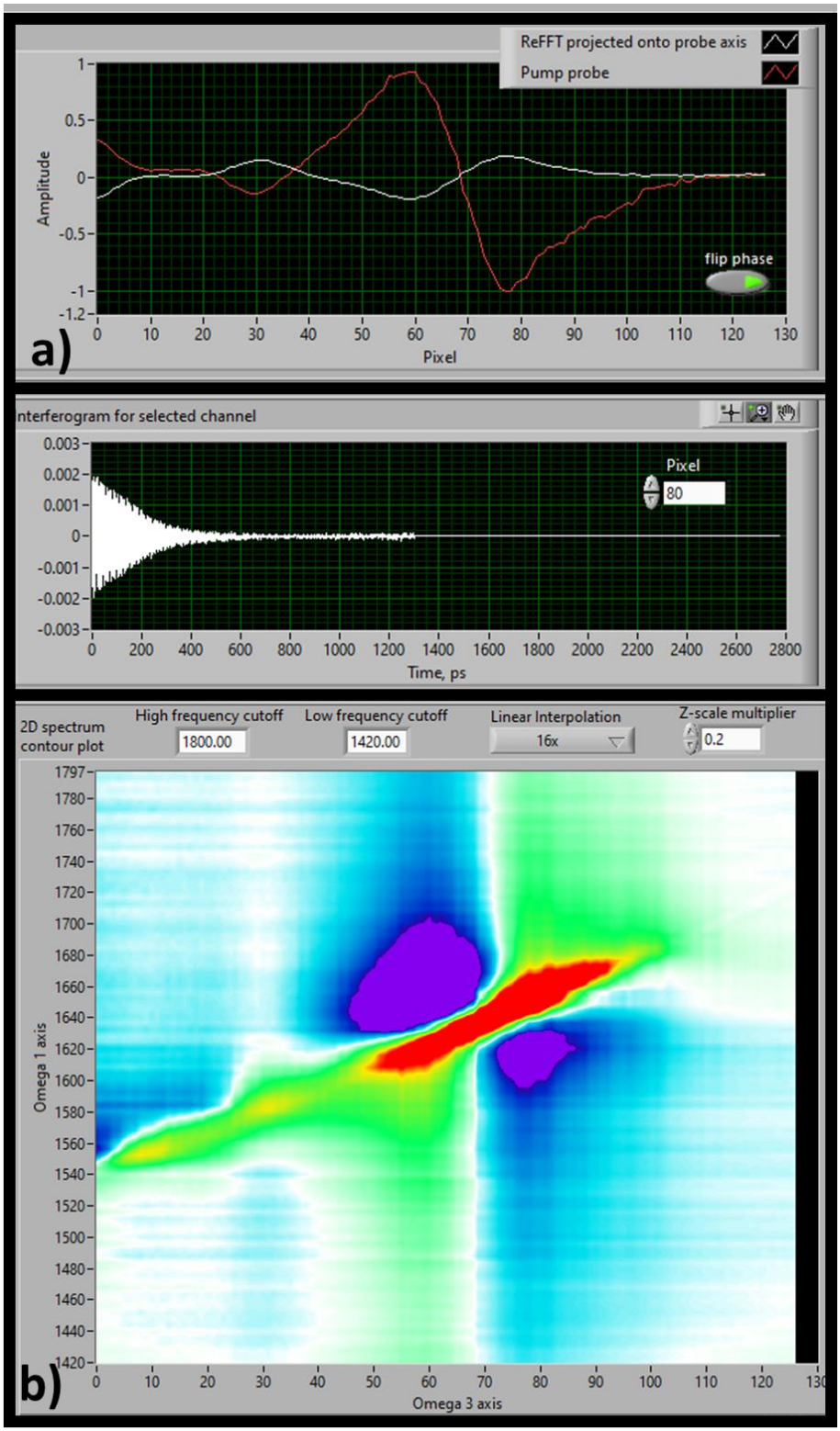


Figure 2-7. The same data as Figure 2-6 has been used. Selecting the wrong t_0 results in a) incorrectly phased data where the expected pump-probe spectrum is out of phase with the experimental pump-probe spectrum. b) the resulting 2D-IR spectrum has distorted lineshapes.

2.4.2 Pulse shaper

2D-IR spectra in chapter (5) were acquired using a pulse shaper approach[6]. In this setup the pump beam is directed into a pulse shaper after splitting from the probe beam (Figure 2-8).

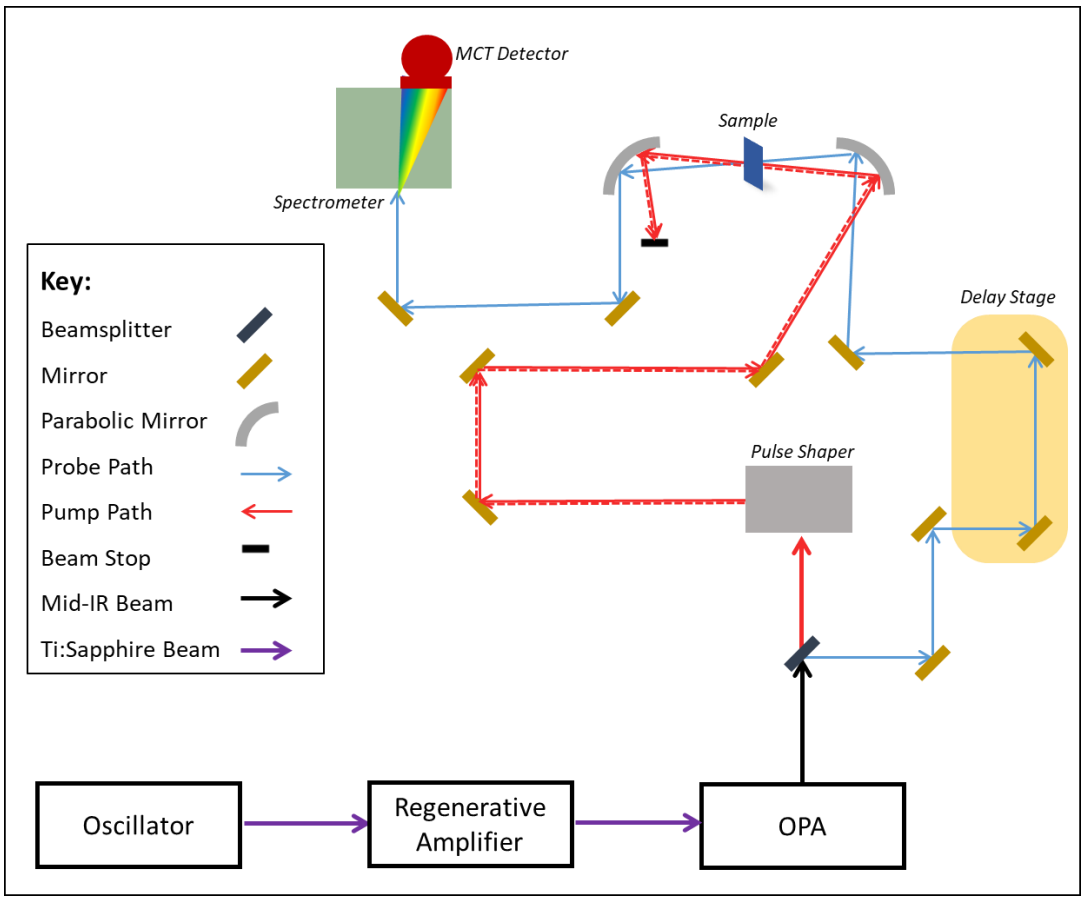


Figure 2-8. Schematic of ULTRA laser optical setup with pulse shaper.

The beam is dispersed on to a grating to allow separation of different frequencies within the pulse, then onto a cylindrical mirror where the beam is collimated. The collimated beam is then passed onto an acousto-optic modulator (AOM) which is coupled with a programmable acoustic wave. This wave is controlled by a waveform generator in an attached computer, and acts as a grating and diffracts the desired frequencies with the correct intensity and phase. Once the beam has passed through the AOM, a second cylindrical mirror and grating pair transform the light back into a temporal pulse. A pulse pair is the output from the pulse shaper and altering the AOM allows for different timings between the two pump pulses (Figure 2-9). The pulse shaper is responsible for pump on/pump off.

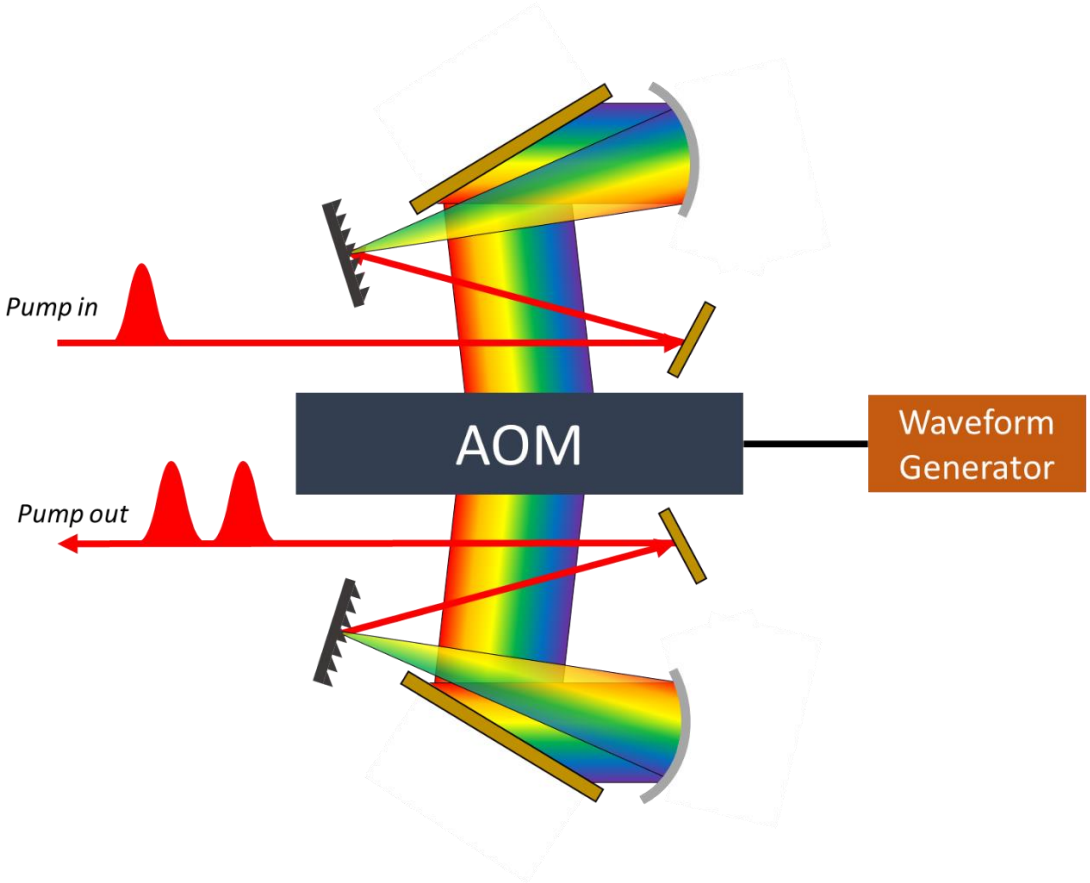


Figure 2-9. Optical setup of the pulse shaper whereby the single pump pulse in is transformed into two pump pulses.

Once passed through the pulse shaper the laser setup is similar to the interferometer approach with the pump and probe pulses overlapped on the sample. The probe beam and signal are then directed onto the grating in the spectrometer then dispersed onto the liquid nitrogen cooled 128-pixel MCT detector. As with the interferometer approach the signal is Fourier transformed to determine the pump frequency axis.

In the interferometer approach it was stated that sampling the signal below the Nyquist limit must occur to detect the frequency component of vibrational modes. In the case of the 1650 cm^{-1} region this is approximately 10 fs, however it was not possible to sample at this small step size for the pulse shaper approach due to issues with memory capacity. The measurements were therefore under-sampled with 24 fs step sizes which results in aliasing and phantom peaks are removed.

Processing of the pulse shaper data is simpler than the interferometer counterpart as the phase is dictated by the pulse shaper and is already known, therefore eliminating the need

for the complex phasing of data using the projection slice theorem. The only data processing required to generate the 2D-IR spectra is to perform a FT of signal.

2.5 T-Jump

This temperature-jump (T-jump) setup was utilised in chapter (4), and it is a time-resolved infrared (TRIR) spectrometer setup with a T-jump pump, mid IR probe[7]. The system used was the ULTRA system (RAL) and the generation of the 1650 cm^{-1} (~6 µm) probe pulse was the same as described for the 2D-IR measurements (section 2.3).

The nanosecond T-jump pump pulse was generated using a custom-built Nd:YAG pumped optical parametric oscillator (OPO). The Nd:YAG laser produced a pulse with a duration of approximately 1.5 ns, an energy of 4 mJ and a repetition rate of 1 kHz. This pulse was then directed into a singly resonant OPO cavity and focussed onto a magnesium oxide periodically poled lithium niobite (MgO:PPLN) crystal. The MgO:PPLN crystal is a quasi-phase matched material where the orientation within is periodically inverted (poled). As the pump beam is passed through the crystal, photons are generated periodically 180° out of phase with the generated photon that would have been created at that point in the crystal if it had not been poled, therefore leading to phase matching. This results in photons interfering constructively and therefore high conversion efficiency within the crystal. A signal beam and an idler beam were generated from the crystal and the idler beam was allowed to exit the OPO cavity while the residual pump beam and signal beam were reflected back into the cavity.

The wavelength of the pump beam was tuned by adjusting the temperature of the crystal and the poling period. In chapter (4) a wavelength of approximately 3.75 µm was used as this allows for pumping the O-D stretch of the D_2O solvent. The pump energy at 3.75 µm was 70 µJ with a pulse duration of ~1.5 ns.

The pump pulse was mechanically chopped to 500 Hz from 1 kHz with a chopper, then the pump pulse was directed towards the sample and overlapped spatially with the probe pulse in the sample. As the pump pulse was at a wavelength of 3.75 µm there would be little interaction with atmospheric water vapour allowing the beam to travel through a different portion of the laser setup to the probe beam, which required purging with nitrogen gas. This allowed the pump pulse to pass through the sample in the opposite direction compared to the probe pulse, i.e. the pump pulse passed through the back side of the sample and the probe passed through the front side.

The data were acquired using a time-resolved multiple probe spectroscopy (TRMPS)[8] method which allowed for acquisition from the nanosecond timescale to the millisecond timescale. The output pulse from the OPO was synchronised with the 10 kHz probe pulse using a computer-controlled delay generator. The 10 kHz probe clock was divided down to 1 kHz and this was used to trigger the Nd:YAG laser. A further 500 Hz output was used to trigger the pump chopper. This 500 Hz signal was used in the data acquisition system to correctly identify the sequence of probe pulses.

Due to the differences in repetition rates between the pump and probe pulses (0.5 kHz and 10 kHz, respectively), 20 probe spectra can be recorded for each pump pulse (Figure 2-10). Assuming the delay between the pump pulse and the first probe pulse is x ns, then the subsequent probe pulses are recorded with delays of x + 100 µs, x + 200 µs, x + 300 µs, with the final probe pulse before the next pump pulse recorded with a delay of x + 1900 µs. The last probe pulse is used as the pump-off measurement for the other 19 probe spectra.

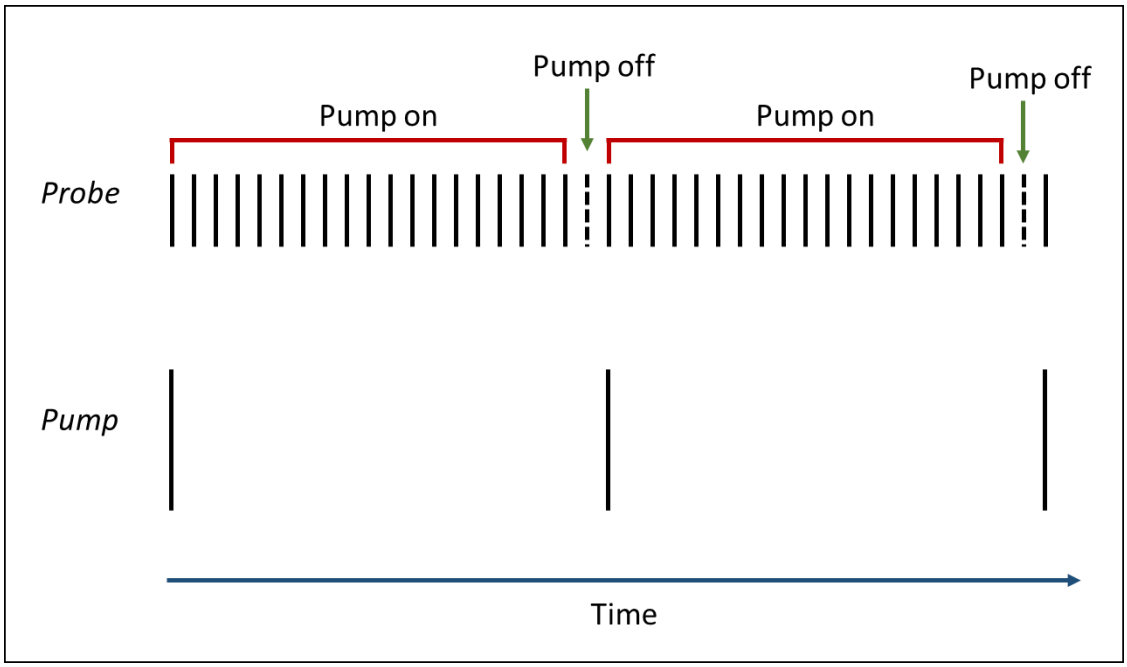


Figure 2-10. Schematic of TRMPS pulse sequence. For every pump pulse there will be 19 'pump on' probes and one 'pump off' pulse.

2.6 Principal Component Analysis

Principal Component Analysis (PCA) is a multivariate analysis method that is used extensively in spectral processing[9–11] to reduce the dimensionality of the data and extract trends within the dataset. It involves orthogonal transformation of the original dataset to get linearly

uncorrelated variables, called Principal Components (PCs), that best describes the variance within the data – referred to as covariance. The first PC (PC1) is calculated to show as much variability in the data as possible. The next PC (PC2) is then calculated to show as much variability in the remaining data not described in PC1. This is continued for each PC with each one representing less data than the previous. Often, the latter PCs represent just noise in the data.

The original data matrix (X) of all the measurements to be analysed, comprises of the samples (columns) and the variables (rows) which in the case of this thesis will be wavenumbers (Figure 2-11). The matrix is then transformed into two smaller matrices[9,12]:

$$X = T.P + E \qquad \text{(Equation 2.5)}$$

Where *T* is the resulting scores of the PCs which describes the relationship between the measurements in the form of PC spectra, *P* is the resulting loadings which describes the relationship between the samples in the form of trends, and *E* is the residual contribution from random noise.

The results of the PCA allows for easier visualisation of patterns within the dataset. The original spectra can be recalculated from the PC scores and loadings in the following manner:

$$Original\ Spectrum\ A = (PC1\ scores * PC1\ loadings\ for\ A) + (PC2\ scores * PC2\ loadings\ for\ A) + \cdots \qquad \text{(Equation 2.6)}$$

All PCA will be performed in the software package Origin Pro (versions 2015 and 2018) using the in-built method.

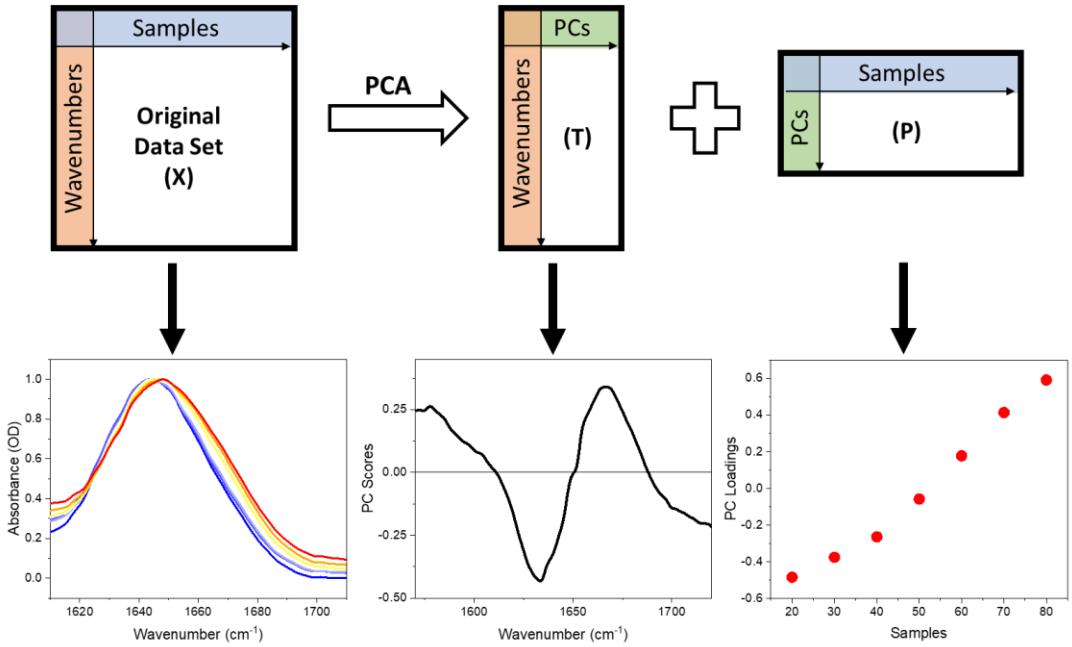


Figure 2-11. PCA data transformation of original data matrix (top), and the corresponding visual outputs (bottom).

2.7 References

(1) Greetham, G. M.; Burgos, P.; Qian, C.; Clark, I. P.; Codd, P. S.; Farrow, R. C.; George, M. W.; Kogimtzis, M.; Matousek, P.; Parker, A. W.; et al. ULTRA: A Unique Instrument for Time-Resolved Spectroscopy. *Appl. Spectrosc.* **2010**, *64* (12), 1311–1319.

(2) Shaw, D. J.; Adamczyk, K.; Frederix, P. W. J. M.; Simpson, N.; Robb, K.; Greetham, G. M.; Towrie, M.; Parker, A. W.; Hoskisson, P. A.; Hunt, N. T. Multidimensional Infrared Spectroscopy Reveals the Vibrational and Solvation Dynamics of Isoniazid. *J. Chem. Phys.* **2015**, *142* (21), 212401/1-212401/9.

(3) DeFlores, L. P.; Nicodemus, R. A.; Tokmakoff, A. Two-Dimensional Fourier Transform Spectroscopy in the Pump-Probe Geometry. *Opt. Lett.* **2007**, *32* (20), 2966.

(4) Hunt, N. T. 2D-IR Spectroscopy: Ultrafast Insights into Biomolecule Structure and Function. *Chem. Soc. Rev.* **2009**, *38*, 1837–1848.

(5) Tokmakoff, A. Coherent 2D IR Spectroscopy : Molecular Structure and Dynamics in Solution. **2003**, *107*, 5258–5279.

(6) Shim, S.-H.; Strasfeld, D. B.; Ling, Y. L.; Zanni, M. T. Automated 2D IR Spectroscopy Using a Mid-IR Pulse Shaper and Application of This Technology to the Human Islet Amyloid Polypeptide. *Proc Natl Acad Sci U S A* **2007**, *104* (36), 14197–14202.

(7) Greetham, G. M.; Clark, I. P.; Young, B.; Fritsch, R.; Minnes, L.; Hunt, N. T.; Towrie, M. Time-Resolved Temperature-Jump Infrared Spectroscopy at a High Repetition Rate. *Appl. Spectrosc.* **2020**, *0*, 1–8.

(8) Greetham, G. M.; Sole, D.; Clark, I. P.; Parker, A. W.; Pollard, M. R.; Towrie, M. Time-Resolved Multiple Probe Spectroscopy. *Rev. Sci. Instrum.* **2012**, *83* (10), 103107/1-

103107/5.

(9) Rajalahti, T.; Kvalheim, O. M. Multivariate Data Analysis in Pharmaceutics: A Tutorial Review. *Int. J. Pharm.* **2011**, *417*, 280–290.

(10) Ellis, D. I.; Goodacre, R. Metabolic Fingerprinting in Disease Diagnosis: Biomedical Applications of Infrared and Raman Spectroscopy. *Analyst* **2006**, *131* (8), 875–885.

(11) Wood, B. R.; Quinn, M. a.; Burden, F. R.; McNaughton, D. An Investigation into FTIR Spectroscopy as a Biodiagnostic Tool for Cervical Cancer. *Biospectroscopy* **1998**, *2* (3), 143–153.

(12) Karamizadeh, S.; Abdullah, S. M.; Manaf, A. A.; Zamani, M.; Hooman, A. An Overview of Principal Component Analysis. *J. Signal Inf. Process.* **2013**, *4*, 173–175.

3 Quantifying Thermally Induced Secondary Structure Changes in Calmodulin using 2D-IR Spectroscopy

This chapter contains results published in the following publication:

> Minnes, L.; Shaw, D. J.; Cossins, B. P.; Donaldson, P. M.; Greetham, G. M.; Towrie, M.; Parker, A. W.; Baker, M. J.; Henry, A. J.; Taylor, R. J.; et al. Quantifying Secondary Structure Changes in Calmodulin Using 2D-IR Spectroscopy. **2017**, *89*, 10898–10906.

This chapter was a collaborative project with scientists from the pharmaceutical company, UCB, who provided additional biophysical measurements, sample preparation and MD simulations. I performed all IR spectroscopy work and data analysis.

3.1 Abstract

Molecular structure is closely linked to function in biological molecules. Understanding the details of mechanisms in the solution phase requires tools to monitor structural changes to the molecule upon perturbation. The ability of 2D-IR spectroscopy, coupled with multivariate analysis, to monitor and quantify secondary structural changes of the calcium binding protein, calmodulin (CaM), as a function of temperature is demonstrated. 2D-IR spectroscopy is able to detect the α-helix to random coil transition of *apo*-CaM upon heating and changes in helical coupling due to structural rigidity, which have been shown to be only accessible through 2D-IR and not FTIR spectroscopy. Quantification of the change in structure is also investigated with good agreement observed between the results of circular dichroism (CD) spectroscopy and 2D-IR spectroscopy, although CD spectroscopy analysis is not optimised for quantification at elevated temperatures. This is partially overcome by using the Ca^{2+} bound state (*holo*-CaM) as a reference as it does not undergo a structural change, only heating effects.

3.2 Introduction

It is known that the structure of a protein is intrinsically linked to its biological function but further information about the molecular mechanisms of changes in structure occurring is desired. This can be achieved through quantifying conformational changes upon perturbation of the protein. The complex and interconnected nature of proteins means that, in certain cases, large-scale structural fluctuations involving secondary structure elements or domains of a protein may be pivotal to function. In this chapter the thermal effect on the structure of a protein will be examined by changing it from room temperature to elevated temperatures.

Circular dichroism (CD) spectroscopy, along with X-ray crystallography, is a gold standard technique for determining the secondary structure of biological molecules at room temperature. Understanding protein mechanisms has been established theoretically with Molecular Dynamics (MD) simulations by predicting the behaviour of proteins in aqueous conditions. The confidence in the simulations can be improved with corroboration from experimental techniques that are able to look at the ultrafast dynamics of femtosecond to picosecond such as 2D-IR spectroscopy.

2D-IR spectroscopy has been previously used to extract site-specific information on structure, dynamics and function of biological molecules using isotopic substitution and non-

natural labelling[1–5]. However, there is scope for 2D-IR spectroscopy of the whole molecule of unlabelled proteins to be utilised in exploring these large-scale changes in proteins and probe conformational changes. This can be helped by the key benefit that 2D-IR has unique coupling patterns occurring between the amide I vibrational modes of peptide units that are visible in the off-diagonal region of the 2D-IR spectrum and it is possible to quantify secondary structure content under equilibrium conditions[6–10].

A challenge is to correlate changes in the spectra to biological functions. 2D-IR is also technically demanding with complex data analysis required. To overcome these problems the use of multivariate data analysis, specifically Principal Component Analysis (PCA), will be developed as a tool in tandem with 2D-IR spectroscopy to extract quantitative spectral information as a function of temperature perturbation. Quantification of protein secondary structure from 2D-IR spectroscopy has already been achieved using SVD as the analysis method[10], however the use of PCA with the technique is novel. The protein used to interrogate these changes will be the human calmodulin (CaM) protein – both in the presence and absence of Ca^{2+} ions.

CaM is a calcium binding messenger protein found in many eukaryotic cells. The Ca^{2+} free structure (*apo*-CaM) is predominantly α-helical and comprises two (C- and N-) globular domains that each contain two Ca^{2+} binding sites[11–14]. The two domains are separated by a central linker, comprising of α-helices and random coil structure (Figure 3-1 (a)). Once bound to Ca^{2+} (*holo*-CaM), the central linker becomes helical and transitions from a closed structure to an open one[15] (Figure 3-1 (b)).

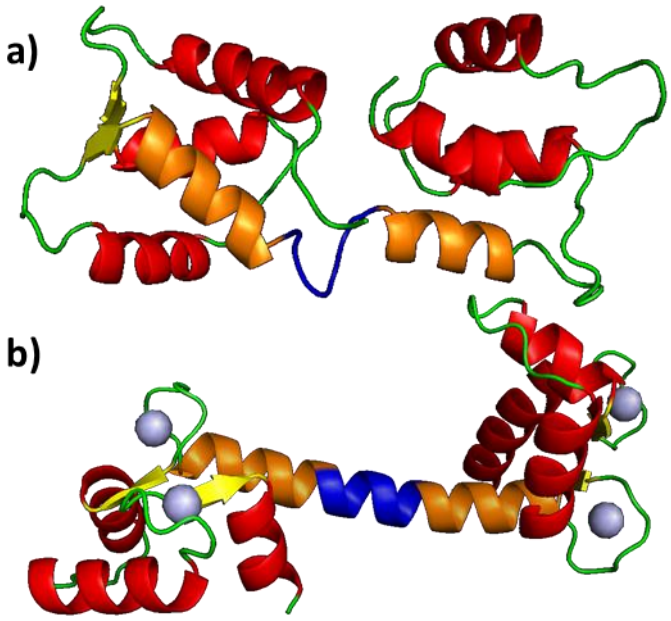


Figure 3-1. Molecular structures of a) apo-CaM (PDB 1CFD) and b) holo-CaM (PDB 1CLL) showing α-helices (red), β-strands (yellow), flexible central linker (orange and blue), and Ca^{2+} ions (grey).

Previous CD studies on CaM have been published[16–19] with thermal experiments reporting unfolding of *apo*-CaM in a two-step process where the C-domain transitions at approximately 43 °C with the more stable N-domain unfolding at a higher temperature of approximately 57 °C[18,20]. A sigmoidal trend in the data was observed for the transitions of the two domains for *apo*-CaM. This is comparable with previous DSC thermal studies on *apo*-CaM where melting of the C- and N-domains was reported to be at 48 °C and 60 °C, respectively[21].

CaM has been previously studied by FT-IR spectroscopy[22–26] with a thermal stability study of *apo*-CaM reporting distorted helical structure and intermolecular β-sheet formation at 70 °C, but no rearrangement observed for *holo*-CaM. 2D-IR has also been used to study CaM which focused on the effect to the carboxylate groups with ion binding and mutations[27,28] and not on the secondary structure changes.

The challenge of this chapter will be to use 2D-IR spectroscopy and PCA to examine and identify large structural changes in the protein and assess its ability and accuracy in quantifying changes in structure compared to the other techniques, namely CD spectroscopy, MD simulations, and FTIR spectroscopy. This will form the basis for probing smaller structural changes in later chapters of this thesis.

3.3 Method

3.3.1 IR Spectroscopy

D_2O, Trizma base, DCl, and $CaCl_2$ were all purchased from Sigma-Aldrich and used without further purification. The human CaM was expressed and purified in house and in collaboration with UCB. *Apo*-CaM was prepared at a concentration of 0.8 mM in deuterated 50mM Tris-DCl pD 7.4 and *holo*-CaM was prepared with the addition of 8 mM $CaCl_2$. Samples were housed in a temperature-controlled Harrick cell. All samples were prepared between two CaF_2 windows separated by a PTFE spacer with a thickness of 50 µm.

All linear infrared spectra were collected on a Bruker Vertex spectrometer at a resolution of 1 cm^{-1}. All 2D-IR spectra were collected using the ULTRA setup at RAL utilising an interferometer approach.

Samples were measured from 20 °C to 80 °C in 10 °C steps and resulting linear IR and 2D-IR spectra were buffer subtracted, baselined, and normalised.

3.3.2 CD Spectroscopy

CD measurements were performed by Daniel Shaw (UCB).

All spectra were recorded on a Chirascan qCD spectrometer (Applied Photophysics) equipped with Peltier heating-cooling control to measure temperature-dependent behaviour with CaM prepared at a concentration of 12 µM in Tris-HCl buffer, pH 7.4. Samples were measured in triplicate over a 185-260 nm range with a resolution of 1 nm. As with FTIR and 2D-IR, measurements were acquired over 20 – 80 °C in 10 °C steps.

Data analysis was performed using the algorithms SELCON3[29,30] and CONTINLL[31,32] from Dichroweb[33–35] and CDNN[36] to predict secondary structure percentages.

CONTINLL is a ridge regression method which fits by comparing the experimental spectra to a linear combination of the spectra from a large database of proteins with known conformations. It has shown to be a good estimate of α-helices.

SELCON3 is a self-consistent method, which means it uses an initial guess of first approximation then performs SVD. The process is repeated until “self-consistency” is achieved. This method has been shown to give a good estimate of globular proteins.

CDNN is a method that uses neural networks. This involves using an artificial intelligence program to find correlations in the experimental data.

3.3.3 Differential Scanning Calorimetry (DSC)

Differential scanning calorimetry (DSC) measurements were performed and analysed by Daniel Shaw (UCB).

All measurements were performed via previously described methods[21]. Briefly, the measurements were performed using a VP-Capillary DSC microcalorimeter (MicroCal, Malvern Instruments) with CaM prepared at a concentration of 180 µM in 20 mM HEPES buffer, 100 mM KCl, pH 7.5, with the addition of 2.5 mM EDTA and 2.5 mM EGTA for the *apo*-CaM samples, and the addition of 2 mM $CaCl_2$ for the *holo*-CaM samples. The measurements were performed in continuous scanning mode with a 15-minute pre-scan at 10 °C then a 10 to 125 °C scan at a rate of 90 °C/hr. Each buffer was also measured and the appropriate buffer for each sample was subtracted from the measurements.

3.3.4 MD simulations

MD simulations were performed by Daniel Shaw (UCB).

The structures used as the starting point for MD simulations of *apo*- (1CFD[37]) and *holo*-CaM (1CLL[38]) were obtained from the Protein Data Bank (PDB). The AMBER ff14SB[39] forcefield, TIP3P[40], and Merz[41] parameters were used for the protein, water and Ca^{2+} ions, respectively.

Simulations at 20 °C were performed using GROMACS 5.04 with 2 fs time-steps with appropriate potential truncation and smoothing, and with simulation snapshots taken every 10 ps. An energy minimisation step was used to relax the structure avoiding any steric clashes or incorrect geometries and a position restraining force was applied to the heavy atoms of the protein to stabilise the protein whilst equilibrating. The production MD was run over the course of 1 µs. The processed trajectories were analysed using the Define Secondary Structure of Proteins (DSSP) hydrogen bonding algorithm[42] to identify backbone hydrogen bonds and classify secondary structure. The root-mean squared (RMS) structural deviation was calculated with reference to the PDB structure.

For 80 °C temperature simulations, the multiple-replica exchange model based on a Hamiltonian exchange method combined with solute scaling was applied called REST2[43], instead of GROMACS which unable to predict structural changes at elevated temperatures.

Further details on the methods used for the MD simulations can be found elsewhere[44].

3.4 Results

3.4.1 CD Spectroscopy

CD spectroscopy was performed and spectra of *apo*-CaM over the temperature range 20 °C to 80 °C (Figure 3-2 (a)) agree well with previously reported measurements[18,19] and feature two negative peaks at 208 nm and 220 nm which are consistent with α-helix structure[45], which CaM is predominately comprised of[11]. The peak sizes diminish with increasing temperature (brown to green) with the 80 °C spectrum having very little peak structure to the trace.

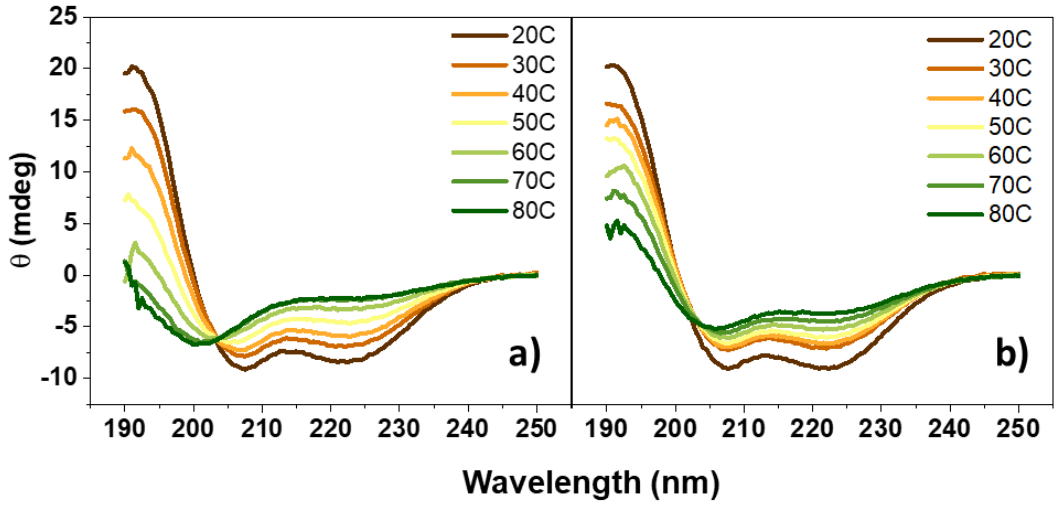


Figure 3-2. CD spectra of a) apo-CaM and b) holo-CaM as a function of temperature from 20 °C to 80 °C (brown to green traces).

The peak at 220 nm was plotted against temperature (Figure 3-3 (black points)) and the resulting profile was approximately sigmoidal in form with the transition occurring near 50 – 60 °C and flattening occurring above 70 °C.

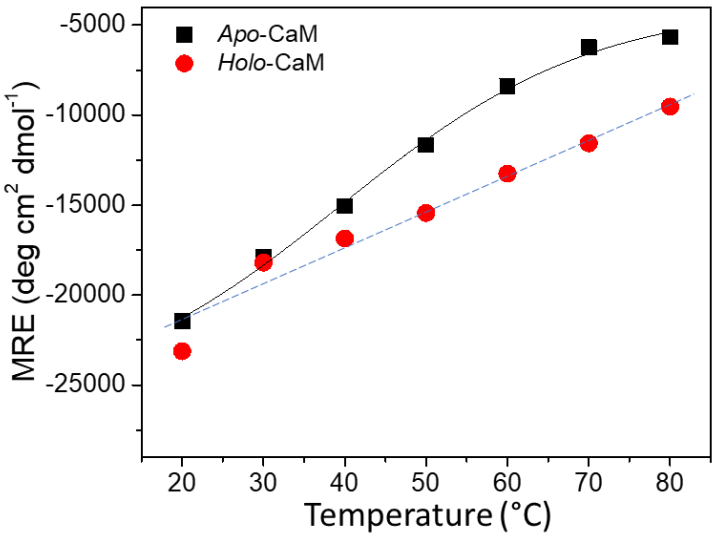


Figure 3-3. Peak at 220 nm from CD spectra of apo-CaM (black) and holo-CaM (red) with sigmoidal fit to apo-CaM. Dashed line on holo-CaM to guide the eye.

The *holo*-CaM spectra also showed the two negative peaks consistent with α-helical structure at 208 and 220 nm (Figure 3-2 (b)). With increasing temperature, the two peaks diminished in size, though to a lesser extent compared to the *apo*-CaM peaks. The 220 nm peak for *holo*-CaM was also plotted and appeared to have a linear decrease with increasing temperature (Figure 3-3 (red points)).

As CD spectroscopy is an established method of evaluating secondary structure at room temperature, there are freely available data analysis algorithms which can be used to predict secondary structure content of a protein from spectra. For each temperature for both *apo*- and *holo*-CaM the secondary structure content was estimated using three algorithms: SELCON3, CONTINLL and CDNN. It must be noted that these algorithms are optimised for use for room temperature spectra, hence are not directly applicable for studying temperature dependent results, but instead are used as a first estimate of protein structure.

The predicted spectra from the algorithms were compared to experimental spectra and the values accepted if the spectra were in good agreement. The values from each method were averaged and summarised results for each temperature noted in Table 3-1, with the standard deviation between the results from each algorithm shown in brackets. For *apo*-CaM there was a predicted decrease in α-helix content (67% to 27%) and an increase in random coil (17% to 38%) with increasing temperature. Likewise, for *holo*-CaM there was a decrease in

α-helix (71% to 39%) and an increase in random coil (16% to 32%). However, it is suggested from the large error values between the methods that the predictions are not wholly accurate, which can be expected as they are derived from algorithms based on room temperature spectra and not from elevated temperature spectra.

Table 3-1. Predicted secondary structure percentages of α-helix and random coil from CD data for all temperatures for both apo-CaM and holo-CaM. Standard deviation errors derived from the variation of results for the different analysis algorithms are in brackets.

Temperature (°C)	*Apo*-CaM		*Holo*-CaM	
	α-helix	Random coil	α-helix	Random coil
20	**67 (±2)**	**17 (±3)**	**71 (±12)**	**16 (±7)**
30	57 (±2)	23 (±2)	60 (±10)	21 (±7)
40	50 (±2)	27 (±3)	58 (±9)	23 (±7)
50	43 (±6)	35 (±10)	54 (±8)	25 (±6)
60	36 (±8)	39 (±7)	52 (±5)	24 (±3)
70	25 (±4)	50 (±16)	44 (±4)	31 (±6)
80	**27**	**38**	**39 (±**1)	**32 (±5)**

3.4.2 DSC Spectroscopy

DSC is a thermal technique which can show transitions in structure in proteins. The results of *apo*-CaM over a temperature range of 20 °C to 125 °C (Figure 3-4 (red trace)) show a peak near 60 °C. Fitting of two Gaussian functions to the peak returned thermal transition midpoints (T_m) of 46 °C and 58 °C. The fitting of two Gaussians to this peak can be justified as attempts to fit a single Gaussian function were performed but deviations from the experimental trace above 60 °C were observed as the peak is asymmetrical. The fitting of the second Gaussian is a subtle change but results in a closer fit to the data.

For *holo*-CaM two peaks were observed at significantly higher temperatures compared to *apo*-CaM. Fitting of Gaussian functions were performed on the *holo*-CaM peaks with resulting T_m values of 96 °C and 111 °C (Figure 3-4 (blue trace)).

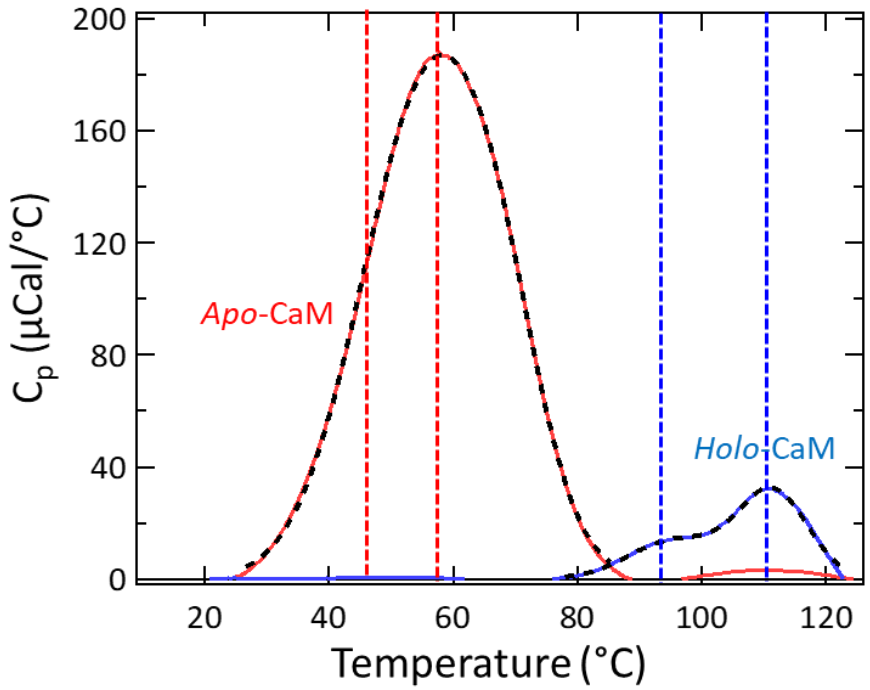


Figure 3-4. DSC data for apo-CaM (red) and holo-CaM (blue). Vertical dashed lines represent the thermal transition midpoints (T_m) determined via Gaussian fitting.

Comparison of *apo*-CaM DSC data with previously published results[18,19] allows assignment of the features observed to thermal unfolding of the protein in a two-step process, with the C-terminal domain unfolding at 46 °C and the more stable N-terminal domain unfolding at 58 °C. This agrees with the CD results which shows a sigmoidal transition at approximately 50-60 °C which is also comparable with previous results[20].

For *holo*-CaM, the results confirm that the addition of Ca^{2+} leads to a significantly more thermally stable protein, with no transition observed below 80 °C from DSC measurements and lack of a sigmoidal profile in the CD data. However, the overall results between the reported CD and DSC data are contradictory with the CD spectra predicting a helical change of 71% to 39% from 20 °C to 80 °C and DSC observing no change below 80 °C for *holo*-CaM. As stated above, the algorithms used for predicting the secondary structure from the CD spectra are not optimal for temperature-dependent results and the effect of temperature on the CD spectra will be discussed further in this chapter.

3.4.3 FTIR

Linear IR spectra of *apo*-CaM between 20 °C and 80 °C show a broad peak near 1645 cm^{-1} which is consistent with the amide I mode of a protein that is primarily α-helical in structure

(Figure 3-5 (a), normalised spectra). The samples were prepared by buffer exchanging H_2O to D_2O which resulted in a peak being observed at 1455 cm^{-1} corresponding the amide II' band as opposed to a peak at 1550 cm^{-1} which would correspond to an un-deuterated amide II band. This suggests essentially complete H/D exchange had occurred prior to analysis (see Appendix 1 for example *apo*-CaM spectrum from 1300 cm^{-1} to 1800 cm^{-1}). Cyclic spectra of measuring at 20 °C then at 80 °C then cooling back to 20 °C were also obtained and the two 20 °C measurements were essentially identical confirming that no further H/D exchange occurred during the thermal cycle and that the unfolding is fully reversible.

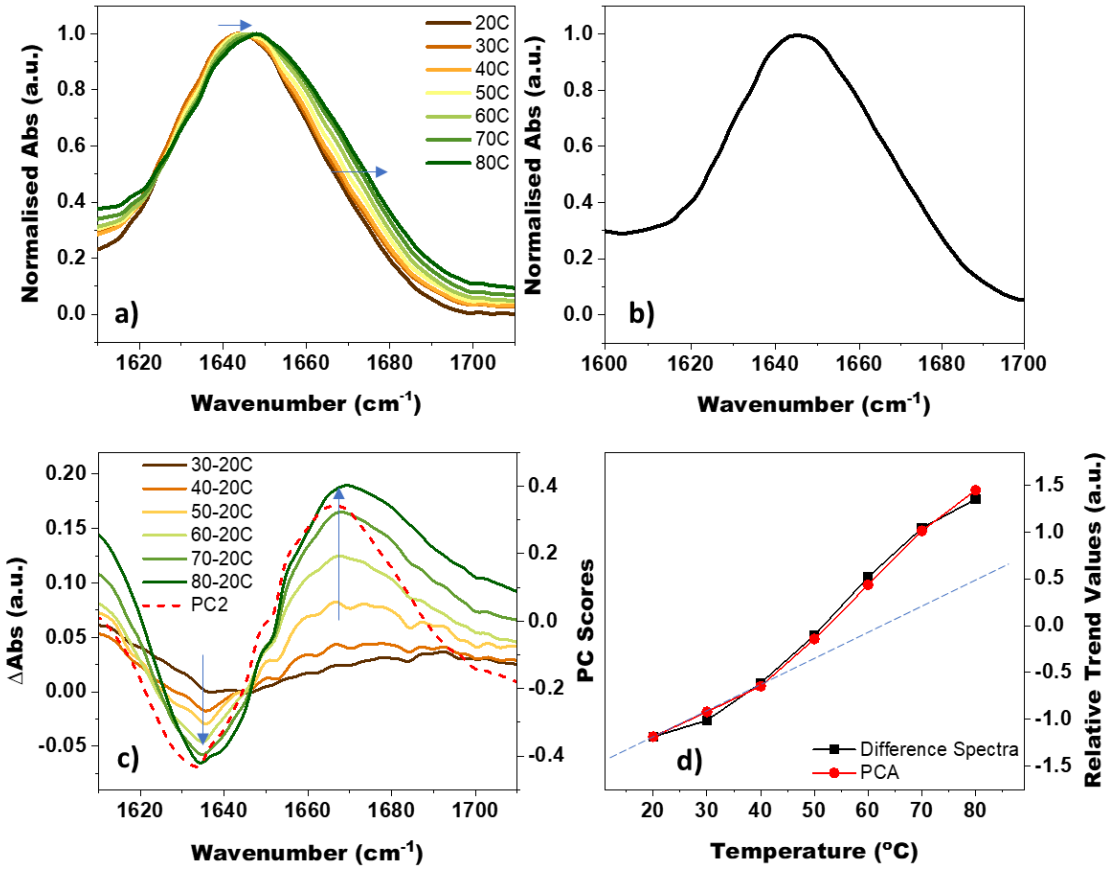


Figure 3-5. Linear IR results of apo-CaM with a) normalised spectra of amide I region over temperature range 20 °C to 80 °C (brown to green), b) average spectrum from all temperatures, c) difference spectra of 20 °C measurement subtracted from all other temperatures (brown to green traces) and PC2 trace (red dashed line), and d) PC2 trend compared to (red) difference spectra magnitudes (peak-to-peak) (black) with blue dashed line to guide the eye.

Qualitatively, the band was observed to broaden, and the peak shifted slightly to higher wavenumbers as the temperature was increased (Figure 3-5 (a), arrows indicating shift). A

clear isosbestic point is seen at 1650 cm^{-1}. This shift is visualised more clearly using difference IR spectra (Figure 3-5 (c)) which have been calculated by subtracting the lowest temperature spectrum (20 °C) from the higher temperature spectra. Fitting the resulting difference IR spectra to two Gaussian functions reveals a negative peak, corresponding to a loss, at approximately 1636 cm^{-1}, and a positive peak, corresponding to a gain, at approximately 1671 cm^{-1}. (Figure 3-6 (a), Table 3-2). The variation in the peak-to-peak magnitude of these two peaks across the different temperature spectra can be seen to have a sigmoidal profile (Figure 3-5 (d), black points).

Table 3-2. Results of fitting linear IR and 2D-IR spectra of apo- and holo-CaM to Gaussian lineshape functions. Values of amplitude (A_i), central wavenumber (ω_i), width (σ_i), and anharmonic shift ($\Delta\omega$) derived from fitting two Gaussian functions are reported for: the linear IR PCA PC2 spectrum ($FTIR_{PCA}$), the average difference IR spectrum ($FTIR_{Diff}$), and slices through 2D-IR spectra at pump frequencies of 1640 cm^{-1} and 1660 cm^{-1} ($2D\text{-}IR_{,pump,}$).

	Apo-CaM				***Holo-CaM***			
Fitting	**$FTIR_{PCA}$**	**$FTIR_{Diff}$**	**$2DIR_{1640cm-1}$**	**$2DIR_{1660cm-1}$**	**$FTIR_{PCA}$**	**$FTIR_{Diff}$**	**$2DIR_{1640cm-1}$**	**$2DIR_{1660cm-1}$**
A_1	-0.3	-0.1	0.2	0.09	-0.3	-0.08	0.2	0.09
ω_1 (cm^{-1})	1634	1636	1627	1635	1633	1633	1626	1634
σ_1 (cm^{-1})	8	10	14	18	8	9	15	19
A_2	0.5	0.04	-0.2	-0.1	0.2	-	-0.2	-0.1
ω_2 (cm^{-1})	1667	1671	1644	1658	1658	-	1643	1657
σ_2 (cm^{-1})	12	11	8	9	15	-	8	10
$\Delta\omega$ (cm^{-1})			17	23			17	23

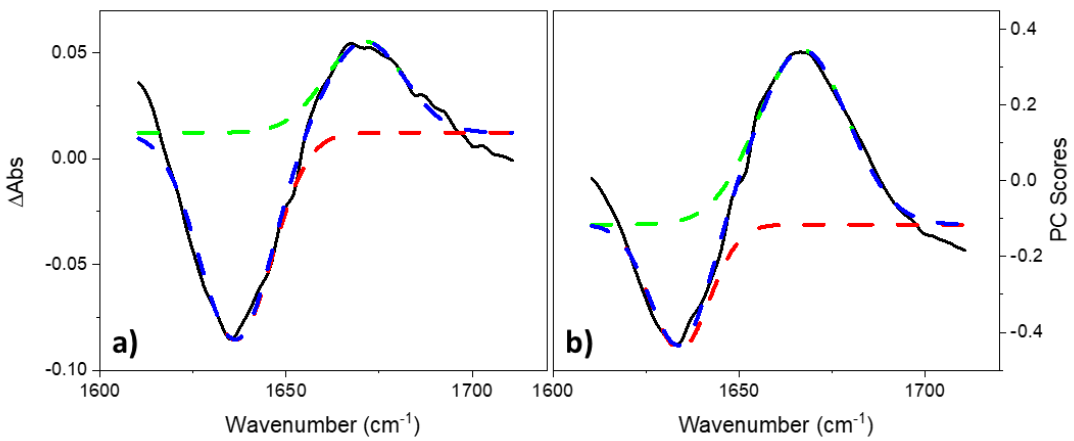


Figure 3-6. Fitting of two Gaussian lineshape functions to apo-CaM linear IR results a) the average difference spectrum, and b) the PC2 spectrum.

Although the temperature-dependent spectral response of the buffer solution has been subtracted from the spectra, the effect of a temperature-dependent baseline, or other temperature-dependent spectral features is clearly visible below 1620 cm^{-1} in the difference spectra.

As an aim of this chapter is to benchmark the use of PCA and compare it to a standard analysis method of difference spectroscopy, PCA was applied to the normalised amide I region of the *apo*-CaM dataset. The results show that the majority (99.97%) of the variance in the data could be described by two principal components (PC) (Figure 3-7 (a)). The major PC spectrum (PC1, 99.29% of the variance) shows a broad peak at 1645 cm^{-1} (Figure 3-7 (c)) with the structure similar to the average of the FTIR spectra (Figure 3-5 (b)). This PC1 is temperature independent as indicated by the unchanging trend over temperature (Figure 3-7 (b), black points). The minor of the two PCs, (PC2, 0.68%) represents the changes in the spectrum as a function of temperature (Figure 3-7 (b), red points).

		PCs		
		PC1	PC2	Cumulative
% of variance		99.29%	0.68%	99.97%
Loadings	20	0.3771	-0.48376	
	30	0.37788	-0.37523	
	40	0.37868	-0.26383	
	50	0.37926	-0.05771	
	60	0.37897	0.17781	
	70	0.37775	0.41385	
	80	0.3761	0.59106	

a)

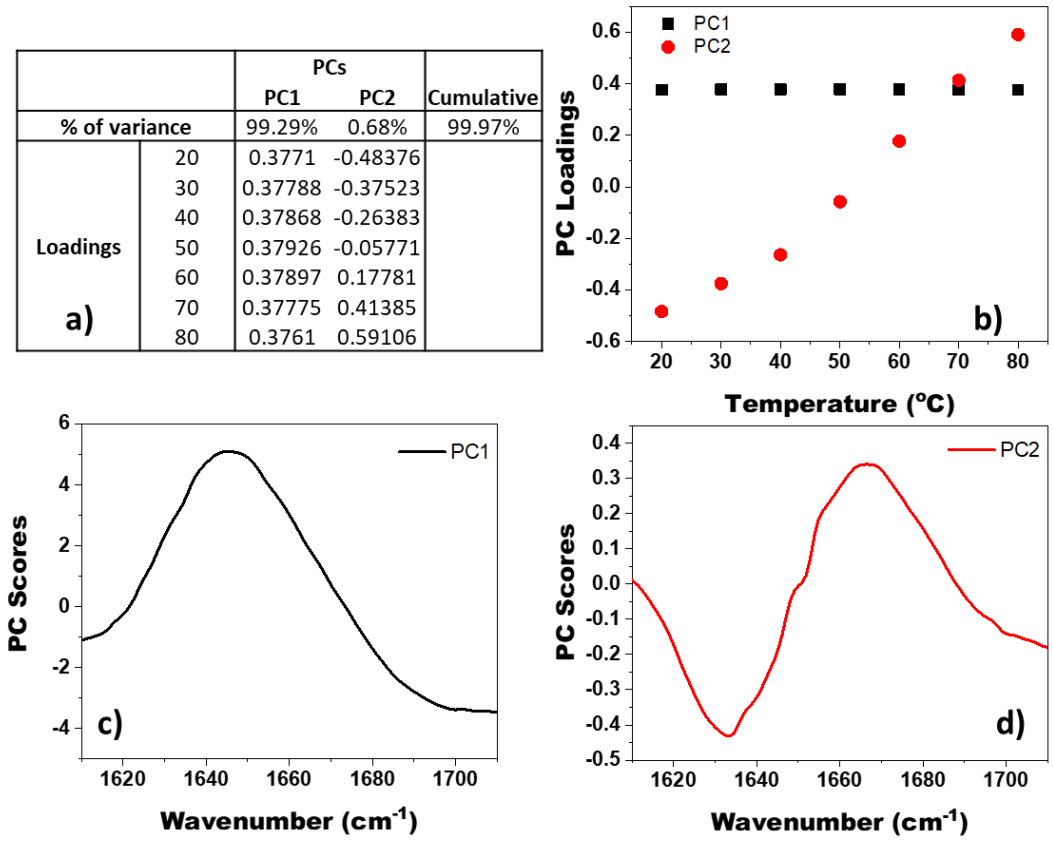


Figure 3-7. Results of applying PCA to FTIR spectra of apo-CaM a) table showing the percentage variance of PC1 and PC2 and the cumulative percentage, and the loading values used in graph b). c) plot showing the spectral contributions to PC1, which is an average spectrum across the temperature range, and d) plot showing the spectral contributions to PC2 with a decrease in intensity evident at 1634 cm^{-1} and an increase at 1667 cm^{-1}.

The PC2 spectrum closely resembled that of the average of the temperature dependent difference spectra in that it could be well-represented by the sum of two Gaussian functions (Figure 3-6 (b)) centred at 1634 cm^{-1} (negative peak) and 1667 cm^{-1} (positive peak). The normalised quantitative contribution of PC2 (loadings) were compared to the peak-to-peak magnitudes of the difference spectra and showed that the two methods are in excellent agreement (Figure 3-5 (c)), confirming that PCA offers a valid approach to extracting temperature-dependent effects in the data. As with the difference spectra, the PC2 trend showed a sigmoidal profile centred at approximately 55 °C, similar to that seen in the CD experiments.

Linear IR experiments on *holo*-CaM showed similar results to *apo*-CaM. The spectra featured a broad peak centred at 1643 cm^{-1} with shifts to higher wavenumber with increasing temperature (Figure 3-8 (a)). PCA was performed on the spectra and again the variance in

data was largely (99.98%) described in two PCs. The major PC (PC1, 99.72%) was again temperature-independent and represented the average spectrum of the spectra in the dataset (Figure 3-9). The minor PC (PC2, 0.26%) showed a strong temperature-dependence with a loss near 1633 cm^{-1} and a gain near 1658 cm^{-1}.

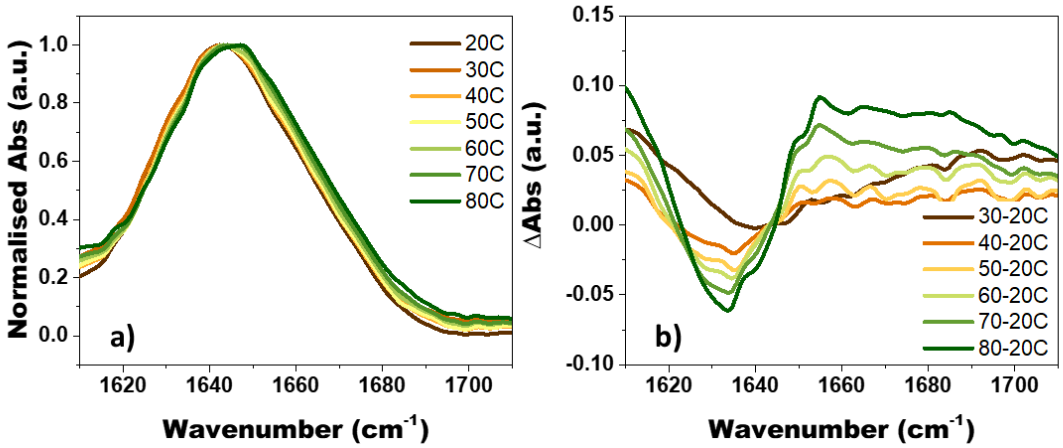


Figure 3-8. Results from linear IR of holo-CaM a) amide I region spectra from 20 °C to 80 °C, b) difference spectra.

The higher frequency feature was significantly smaller than that seen in *apo*-CaM, with the 1633 cm^{-1} feature dominating. A result of this is that the difference spectroscopy approach to IR spectra of *holo*-CaM, calculated in the same manner as *apo*-CaM difference spectra, could not be fit well to two Gaussian functions due to the presence of additional spectral contributions near 1680 cm^{-1}(Figure 3-8 (b)). This additional contribution is not reflected in the PC2 spectrum.

The trend for *holo*-CaM was also different to *apo*-CaM trend (Figure 3-9 (d)), with *holo*-CaM following close to a linear trend compared to the sigmoidal trend seen with *apo*-CaM. The variation with increasing temperature was also decreased with *holo*-CaM. These compare well with the results of the CD spectroscopy, suggesting a more thermally stable protein after Ca^{2+} binding.

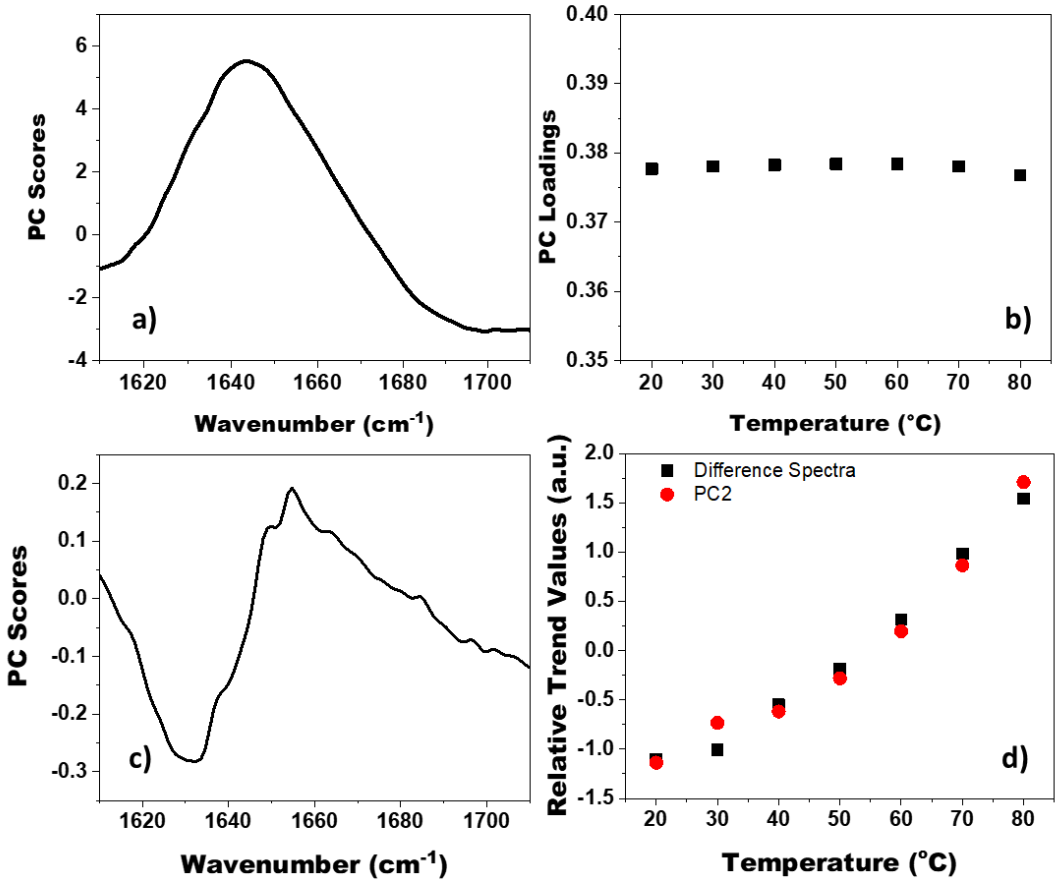


Figure 3-9. a) PC1 spectrum and b) PC1 trend for FTIR results of holo-CaM. c) PC2 spectrum and d) peak-to-peak magnitudes of difference spectra compared to PC2 trends for holo-CaM.

3.4.4 2D-IR Spectroscopy

The 2D-IR spectrum of *apo*-CaM obtained at 20 °C comprises of a negative diagonal peak (red) at a pump and probe frequency of approximately 1645 cm^{-1} (Figure 3-10 (a)). This corresponds well to the peak observed in the 20 °C linear IR spectrum and is assigned to the v=0-1 transition of the amide I mode. This peak is accompanied by an off-diagonal positive peak (blue) which is due to the v=1-2 transition. Fitting slices through the 2D-IR spectrum at pump frequencies of 1640 cm^{-1} and 1660 cm^{-1} (highlighted slice positions shown on Figure 3.10 by dashed lines) to the sum of two Gaussian functions returned values for the anharmonic shift of 17 cm^{-1} and 23 cm^{-1}, respectively (Table 3-2, Figure 3-11). The separation along the probe frequency axis is caused by the anharmonicity of the mode, which is not constant across the amide I band.

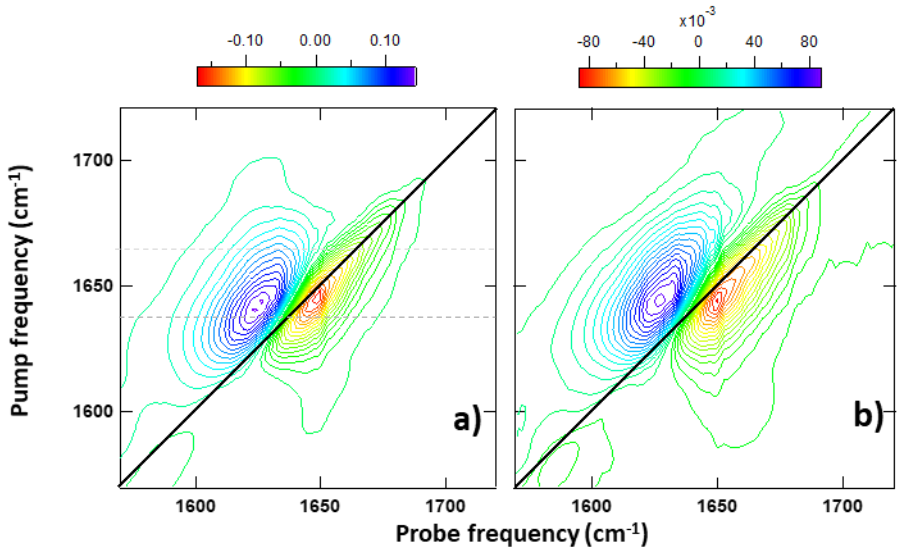


Figure 3-10. 2D-IR spectra for apo-CaM at a) 20 °C and b) 80 °C. Negative peaks (red) present on the diagonal assigned to the v=0-1 transition of the amide I mode, and the positive off-diagonal peaks (blue) assigned to the v=1-2 transition. Dashed orange line highlighting where 1640 cm^{-1} pump slice from Figure 3.11 arises. Dashed blue line highlighting where 1660 cm^{-1} pump slice from Figure 3.11 arises.

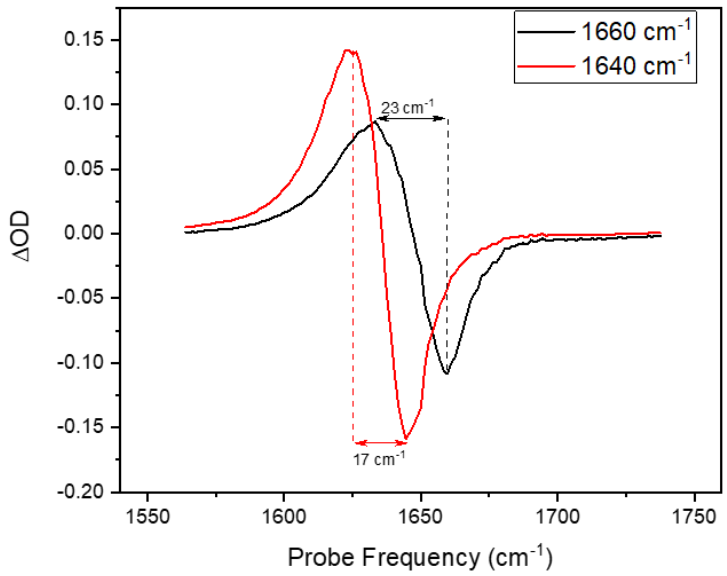


Figure 3-11. Pump slices at 1640 cm^{-1} and 1660 cm^{-1} used to determine the anharmonic shift variance across the 2D-IR diagonal peak of apo-CaM at 20 °C.

The 80 °C *apo*-CaM spectrum is qualitatively similar to the 20 °C spectrum, with a negative diagonal peak at 1645 cm^{-1} (red) and the corresponding off-diagonal positive peak (blue)

(Figure 3-10 (b)). Difference spectroscopy, performed in the same manner as the linear IR method by subtracting the 20 °C spectrum from all other temperature spectra, was employed to visualise differences in the spectra that are otherwise difficult to detect, and temperature-dependent changes were revealed by the increasing signal size of the difference spectra as the temperature increased (Figure 3-12).

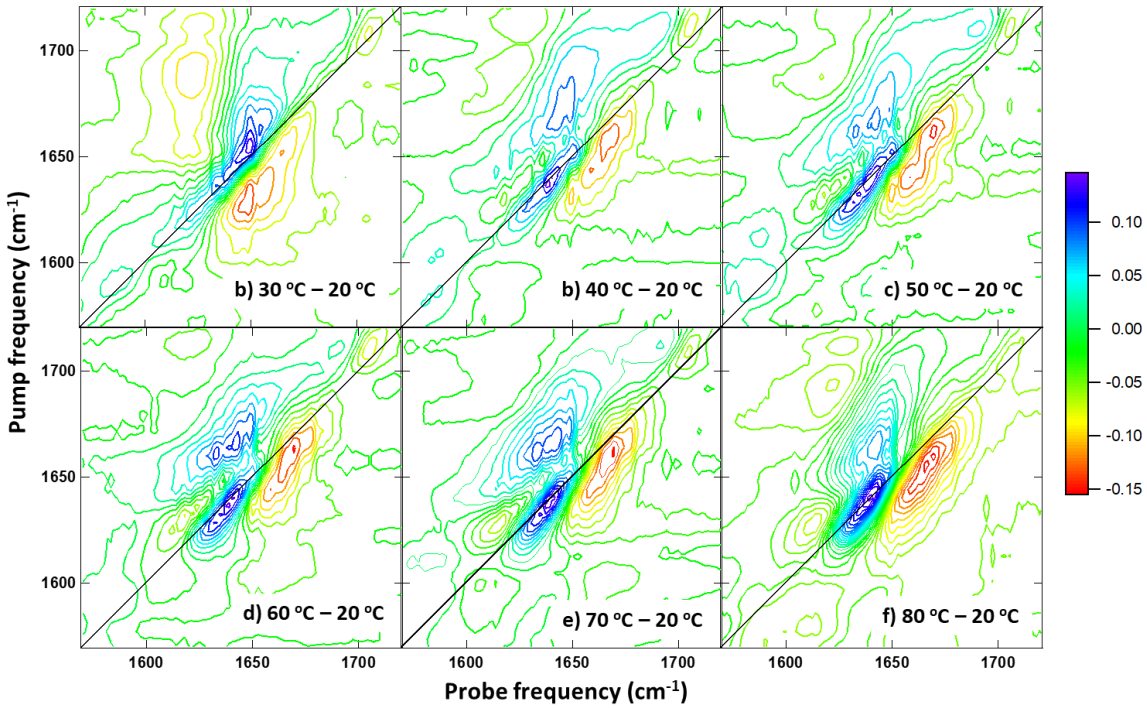


Figure 3-12. Full set of 2D-IR difference spectra for apo-CaM from 20 °C to 80 °C by subtracting the lowest temperature spectrum (20 °C) from all elevated temperatures.

The average of the difference spectra (Figure 3-13 (a)) contains a positive feature located along the diagonal near 1635 cm^{-1} and a negative feature located just below the diagonal near 1660 cm^{-1}. Both features occur in the region of the 2D-IR spectrum occupied by the amide I v=0-1 band (black and red arrows). These indicate a loss of spectral feature at 1635 cm^{-1} and a gain at 1660 cm^{-1}, corresponding to the changes observed in the linear IR difference spectra. Accompanying these features are peaks in the position of the v=1-2 transition with the opposite sign to the diagonal peaks as would be expected (blue and yellow arrows). The values of the anharmonic shift between the v=0-1 and v=1-2 components of the difference spectral features at 1635 and 1660 cm^{-1} were found by Gaussian fitting to be 16 and 28 cm^{-1} respectively. The former is very similar to that obtained from the 2D-IR spectrum,

the latter is somewhat larger, reflecting the slight shift of the 1660 cm^{-1} feature off the spectrum diagonal.

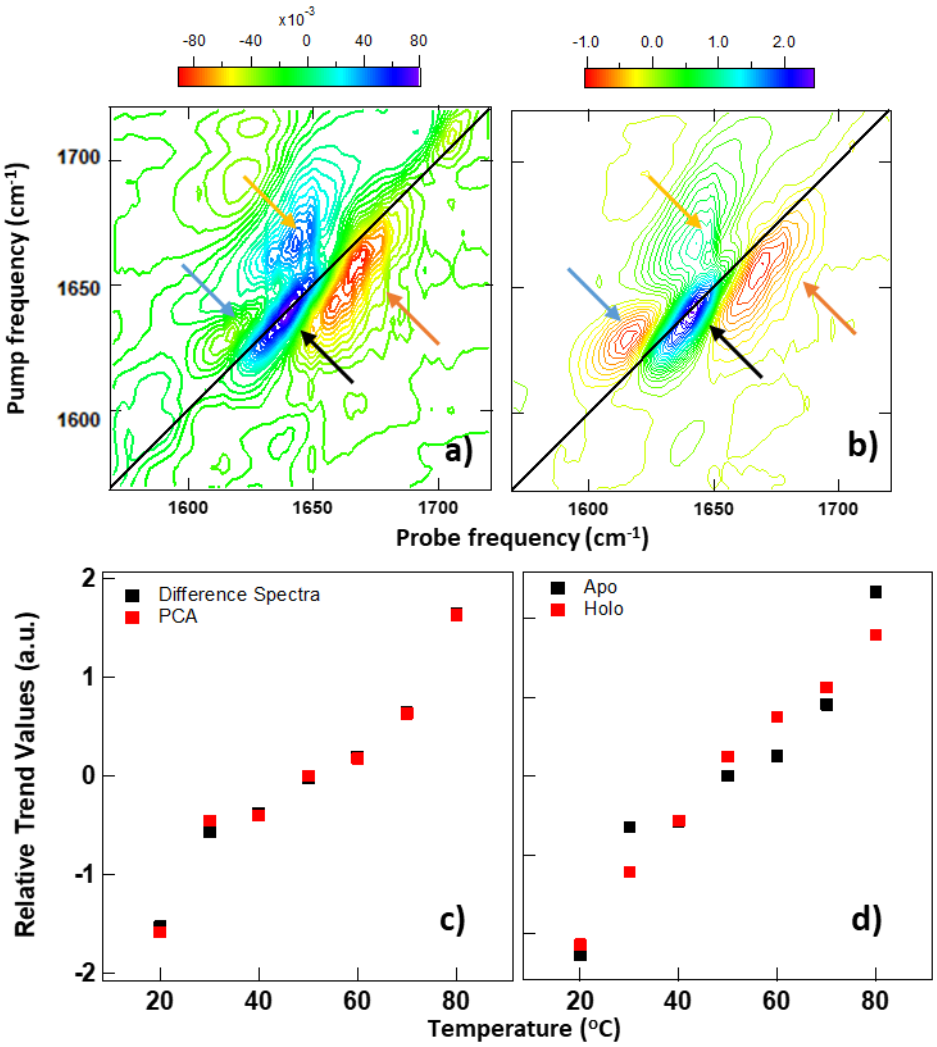


Figure 3-13. Apo-CaM 2D-IR results showing a) the average difference spectrum and b) the PC2 spectrum. Arrows used to indicate main spectral features, c) trend comparison between the difference spectra magnitudes and the PC2 loadings, and d) comparison of trends between apo-CaM and holo-CaM.

The dataset was then analysed using PCA and revealed two PCs (99.39%) (Figure 3-14 (a)). The major PC (PC1, 98.62%) was temperature independent as evident by the lack of change in loadings with increasing temperature (Figure 3-14 (b), black points). The spectrum was the average of the spectra in the dataset (Figure 3-14 (c)). The minor PC (PC2, 0.77%) had a strong temperature-dependence and the spectrum was in good agreement with the average difference spectrum (Figure 3-13 (b)).

		PCs		
		PC1	PC2	Cumulative
% of variance		98.62%	0.77%	99.39%
Loadings	20	0.37542	-0.64876	
	30	0.37687	-0.18589	
	40	0.37975	-0.1643	
	50	0.3799	0.000231	
	60	0.3794	0.07163	
a)	70	0.37902	0.25938	
	80	0.37536	0.66719	

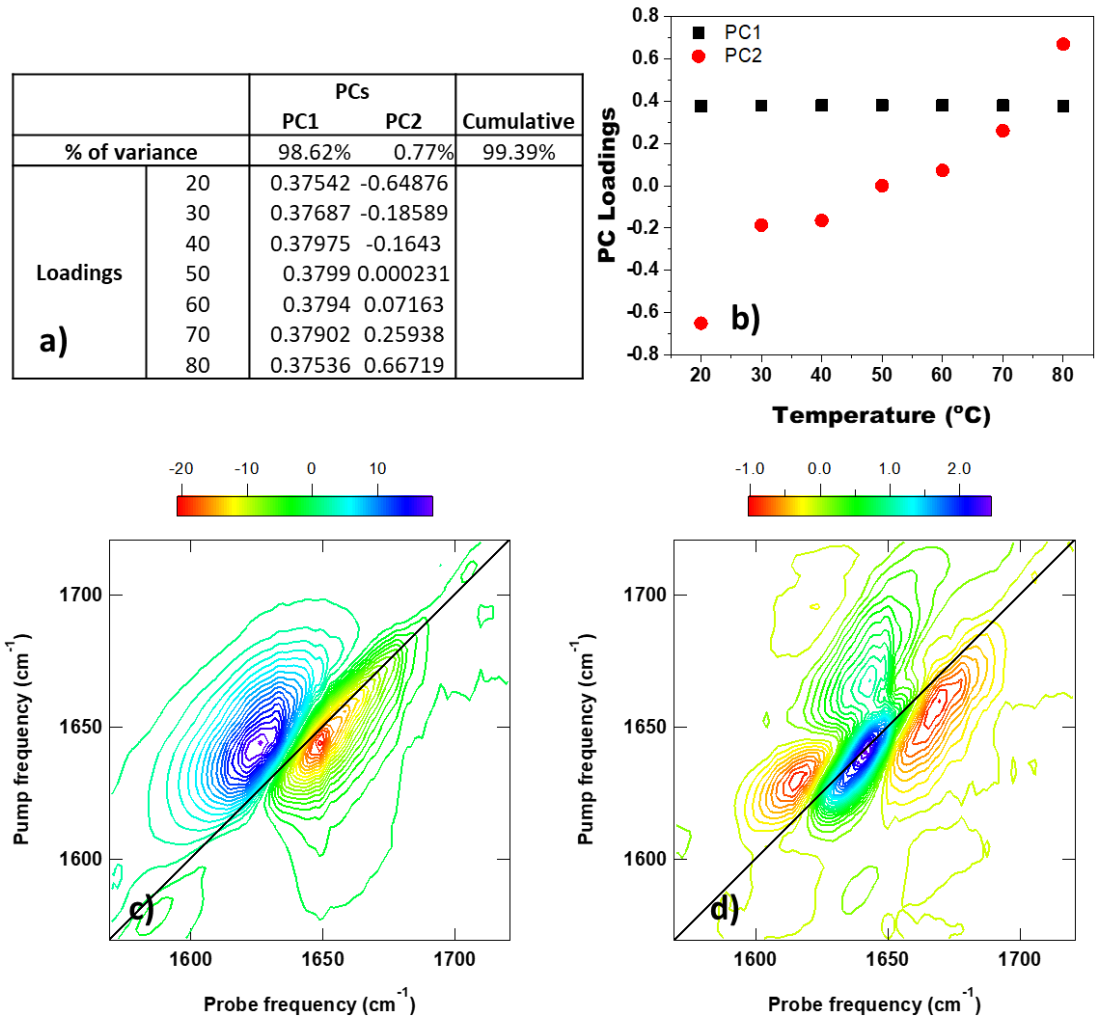


Figure 3-14. Summary of the PCA results of 2D-IR apo-CaM a) table showing the percentage variance of PC1 and PC2 and the cumulative percentage, and the values of the loadings used in graph b). c) plot showing the spectral contributions to PC1, which is an average spectrum across the temperature range, and d) plot showing the spectral contributions to PC2 with a diagonal peak evident at 1640 cm^{-1} and a peak just below the diagonal at 1660 cm^{-1}.

The relative contribution of PC2 with increasing temperature is displayed in (Figure 3-13 (c), red points), which closely matched the temperature dependence of the features obtained from the difference spectra from the peak-to-peak magnitudes (Figure 3-13 (c), black points). The spectral features used to calculate the peak-to-peak magnitude trends from the difference spectra were the diagonal peaks at 1640 cm^{-1} and 1660 cm^{-1}. The comparison of PC2 with the difference spectra for *apo*-CaM 2D-IR supports the validity of using PCA to extract the important temperature dependent features in the 2D-IR spectra. Another benefit of using PCA is its ability to improve signal to noise as seen in the comparison between the average difference spectrum and PC2 spectrum. This occurs as variations from random

instrumental or sample-related noise is not temperature-dependent and therefore is not present in PC2 but will be present in another PC. This is a known quality of PCA and has resulted in being widely used as a noise reduction method[46,47].

A difference from the linear IR results is the lack of a sigmoidal profile for the temperature dependent trends. The reasons for this difference will be discussed later in this chapter.

2D-IR analysis was also performed on *holo*-CaM over the temperature range 20 °C to 80 °C with the 20 °C spectrum qualitatively similar to the 80 °C. Both spectra possessed a negative diagonal peak (red) at 1645 cm^{-1} corresponding to the v=0-1 transition, and a positive off-diagonal peak (blue) corresponding to the v=1-2 transition (Figures 3-15 (a) and (b)). Difference spectroscopy and PCA were performed on the *holo*-CaM dataset with the resulting average difference spectrum (Figure 3-15 (c)) and the PC2 spectrum (Figure 3-15 (d)) resembling each other. The spectra were similar to the *apo*-CaM counterparts, though the change in the v=0-1 feature at 1635 cm^{-1} was dominant resulting in a reduced contribution near 1660 cm^{-1}, consistent with the linear IR results. This suggests a different thermal response of *holo*-CaM as to *apo*-CaM.

The temperature dependences of the signals extracted from linear IR and 2D-IR using difference spectroscopy and PCA agreed well, showing an increasing trend with temperature (Figure 3-15 (e)). As in the case of the FTIR data, *holo*-CaM displayed a weaker overall temperature dependence than *apo*-CaM (Figure 3-13 (d)), though not to the same extent as reported earlier. The reasons for this will be discussed later in the chapter.

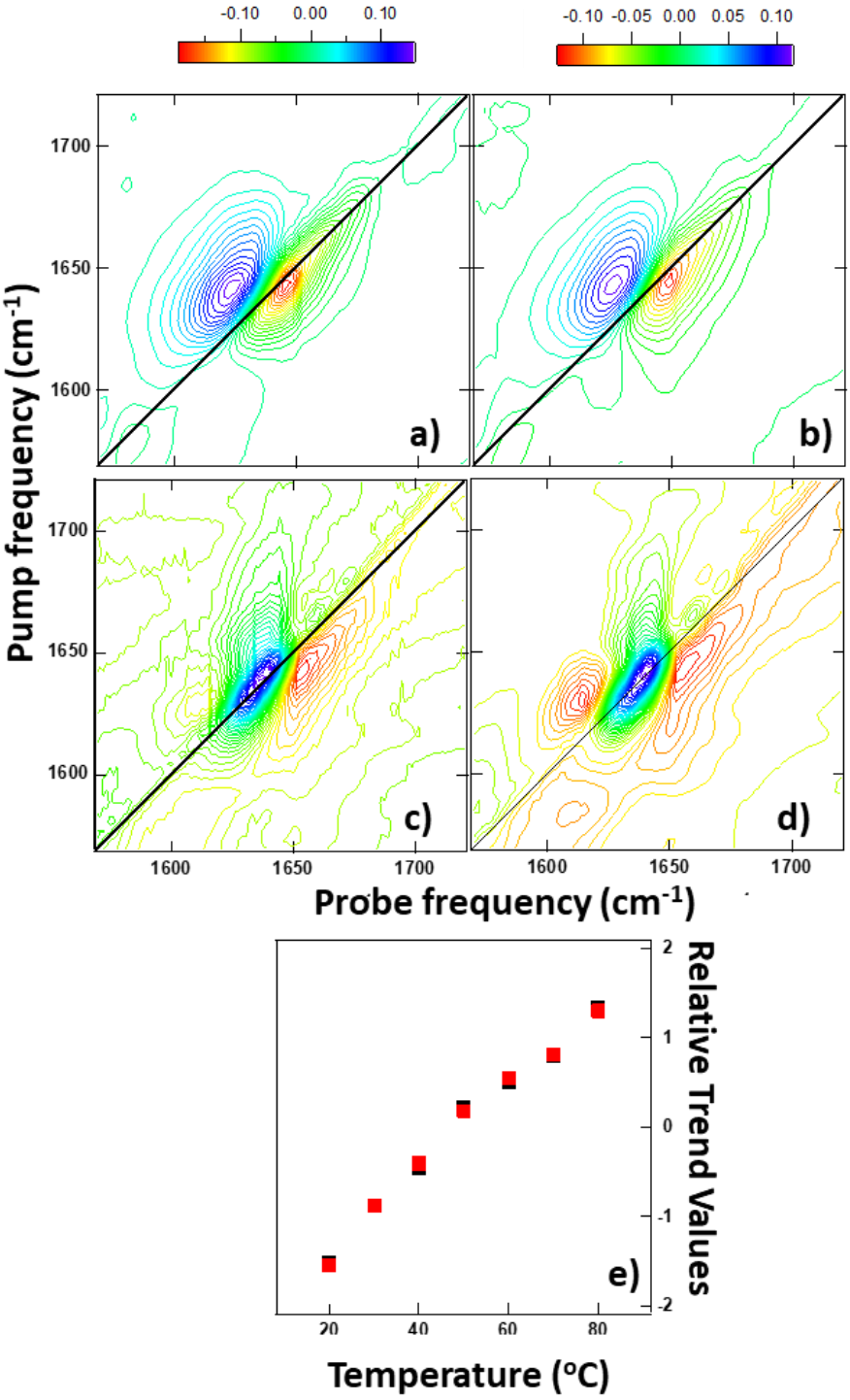


Figure 3-15. 2D-IR results of holo-CaM a) the 20 °C spectrum, b) the 80 °C spectrum, c) the average difference spectrum, d) the PC2 spectrum, and e) the normalised trends comparing the magnitudes of the difference spectra over the temperature range (black) with the PC2 loadings over the given temperature range (red).

3.5 Discussion

3.5.1 Spectral assignment

The IR data (both linear and 2D-IR) for *apo*-CaM indicate that, as the temperature increases, the amide I band of the protein shifts from lower to higher frequency. For both types of experiments a loss of intensity near 1640 cm^{-1} and a gain near 1660 cm^{-1} were observed with increasing temperature. These changes were seen in *holo*-CaM also, though with slight differences in the thermal response.

There are two possible assignments for the effects seen in the IR data; the first is the effect of general heating on the spectrum. Increasing the temperature would be expected to induce a shift to higher frequency of the amide I band through a weakening of H-bonds both within the protein and between the protein and the solvent. The other possible effect is a change in the secondary structure of the protein from α-helix to random coil which may cause a spectral shift. It has been seen in previous CD studies that a decrease in α-helix signal is due to a combination of these effects[48].

The DSC data suggest that heating *apo*-CaM leads to denaturing of the C- and N- domains at 46 °C and 57 °C, respectively. However, it also suggests that the *holo*-CaM protein is thermally stable until 96 °C, where C-domain melting occurs. These results are seemingly contradictory to the results of IR spectroscopy making it vital that accurate assignment of the changes seen in the IR spectra is determined.

Examination of the IR data shows an isosbestic point in the linear IR data that correlates to a transition between two states having occurred but is not conclusive without other corroborating data. The component of the difference and PC2 spectra decreasing in amplitude with increasing temperatures located near 1640 cm^{-1} in both the linear IR and 2D-IR data is close to the frequency expected for the amide I mode of an α-helix. The feature gaining in amplitude at 1665 cm^{-1} is likely to contain two contributions, firstly from the generic effects of increased temperature weakening the H-bonding in the system, but also from a gain in random coil secondary structure as a result of the loss of the α-helix.

Although random coil frequency is expected near 1650 cm^{-1} at room temperature, the elevated temperatures would produce a shift to higher wavenumber. In addition, this positive component will overlap with the negative component of a loss of α-helix resulting in a bias towards higher wavenumber – resulting in the random coil feature appearing at

approximately 1665 cm^{-1} instead of 1650 cm^{-1}. This feature appears just below the diagonal (as opposed to the α-helix feature that appears on the diagonal) in the 2D-IR spectrum which suggest it arises from a more flexible secondary structure, as would be expected from random coil at elevated temperatures, because fluctuations of inhomogeneously-broadened systems lead to anti-diagonal broadening of the 2D-IR lineshapes. Therefore, examination of the IR data shows multiple factors that are indicative of the changes being a result of a secondary structure transition.

Insight into the spectral features for *apo*-CaM can also be obtained through the differences in the anharmonicity (Figure 3-11). The feature at 1635 cm^{-1} has an anharmonicity of approximately 17 cm^{-1}, which is small in comparison to the feature at 1665cm^{-1} (approximately 23 cm^{-1}). As it is known that the coupling of amide I oscillators in the α-helix is known to decrease the anharmonicity with increasing size of the helix, it further suggests that there is a transition from a high quantity of well-structured α-helix to the less structured random coil.

The IR data imply that an α-helix to random coil transition is occurring for *apo*-CaM, however, the *holo*-CaM has produced similar results which does not undergo a transition. Therefore, to provide confidence for the assignment of α-helix to random coil in *apo*-CaM with elevated temperatures and to differentiate between the two states, a transition dipole moment method has been applied. This is a method which combines linear IR and 2D-IR measurements to extract a spectrum of the relative transition dipole strength of the sample for each temperature, d(ω)[49]. It relies on the fact that the dependences of the signal intensity on the transition dipole moment of the linear absorption ($\propto|\mu|^2$) and 2D-IR techniques ($\propto|\mu|^4$) are different. This leads to changes in the relative magnitudes of the FTIR and 2D-IR signals if a change in secondary structure induces variations in the vibrational coupling within the molecule. A key result of this is that the integrated area of a linear IR spectrum is conserved when the coupling changes, whereas the integrated area of a 2D-IR spectrum is not conserved. This difference can be reflected in the frequency-dependent ratio of the IR signal and the 2D-IR diagonal, where the former will decrease relative to the latter in the event of increased vibrational coupling.

To perform the published method of extracting the absolute transition dipole moment, FTIR and 2D-IR measurements of a molecule with a known or measurable transition dipole is required as a calibration at the time of the protein measurements. A slightly different

approach is applied here, as the absolute transition dipole moment is not required. Changes in the transition dipole strength relative to a reference temperature (20 °C) have been used in the relationship stated in equation (3-1). This allows qualification of the transition dipole moment and not accurate quantification.

$$d(\omega)^{rel} = \frac{2DIR^{T}(\omega)}{FTIR^{T}(\omega)} \frac{OD^{20}}{\Delta OD^{20}} \qquad \text{(Equation 3-1)}$$

Where $d(\omega)^{rel}$ is the spectrum of the relative transition dipole strength as compared to that at a reference temperature (20 °C). $2DIR(\omega)^{T}$ is the diagonal of the 2D-IR spectrum at temperature T, $FTIR^{T}(\omega)$ is the IR spectrum at temperature T. OD^{20} is the maximum value obtained from the IR spectrum at 20 °C, and ΔOD^{20} is the maximum ΔOD value from the 2D-IR diagonal at 20 °C.

Examining the amide I region of the linear IR spectra without normalisation (Figure 3-16 (a)) it can be seen, that although the shape of the spectrum changes with temperature, the integral is constant over the same range, as would be expected (Figure 3-16 (c), black points).

The 2D-IR diagonal was extracted for each 2D-IR spectrum without normalisation and showed a positive peak centred at approximately 1645 cm^{-1} (Figure 3-16 (b)). The integral of the 2D-IR diagonal is not conserved with increasing temperature (Figure 3-16 (c), red points). The integral decreases in intensity. This is due to decreasing signal size on the low frequency edge of the amide I lineshape but no change to the signal size of the higher frequency edge.

The calculated $d(\omega)^{rel}$ spectra for *apo*-CaM shows a peak near 1635 cm^{-1} which decreases with increasing temperature (Figure 3-16 (d)). This is a strong indication of a loss of vibrational coupling associated with an α-helix. The region near 1660 cm^{-1}, assignable to random coil contributions to the spectrum, is unchanged with increasing temperature as, although the quantity of random coil increases as the temperature is elevated, the impact upon the transition dipole from changes in coupling is negligible.

The results of the transition dipole method therefore strongly support the other indicators that the temperature-induced changes in the IR and 2D-IR spectra of *apo*-CaM are assignable to a helix-coil transition accompanying melting of the C and N-terminal domains. These results are only accessible via 2D-IR measurements and not through FTIR spectroscopy directly due to the differences in transition dipole moment signal dependencies.

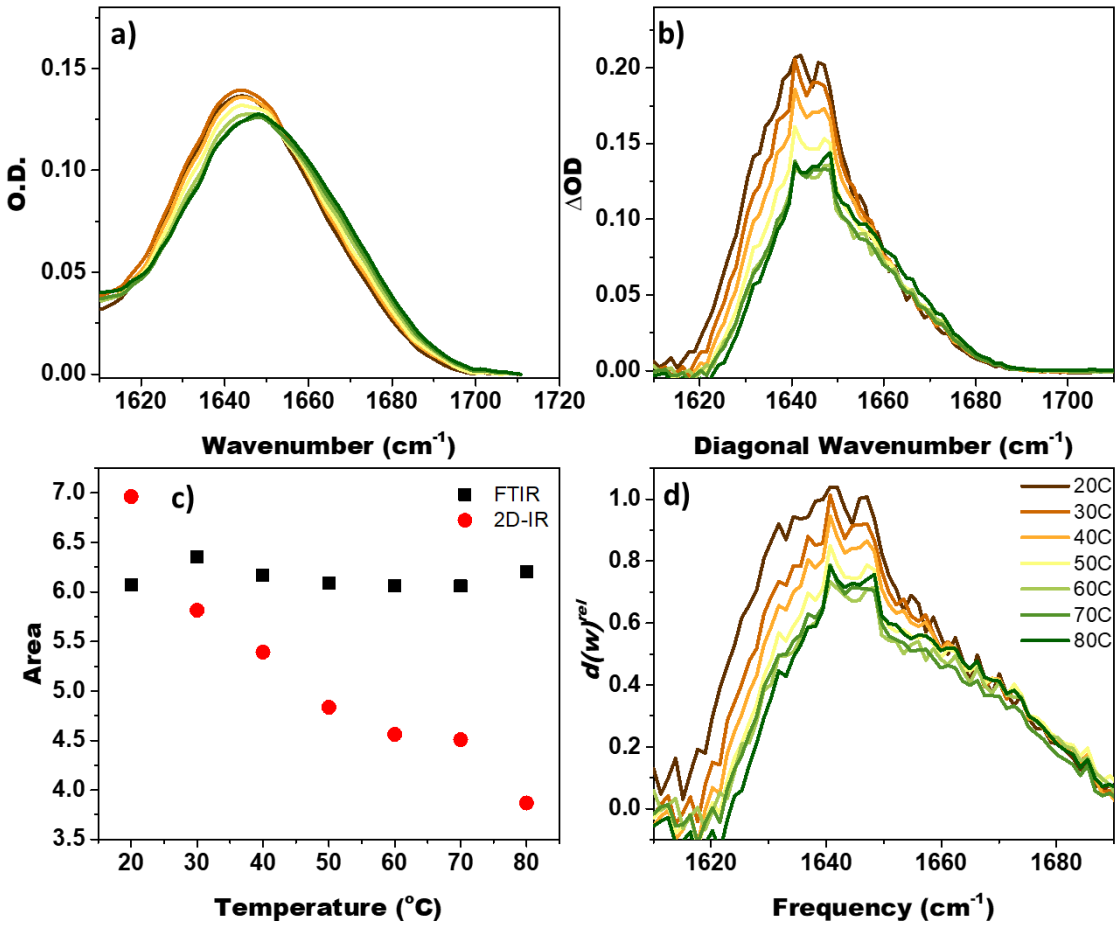


Figure 3-16. Transition dipole analysis for apo-CaM. a) Buffer-subtracted unnormalised linear IR spectra, and b) 2D-IR diagonal spectra of apo-CaM. Spectra have been inverted from negative to positive to compare with the positive FTIR spectra. c) integrated areas for each temperature for FTIR (black) and 2D-IR diagonal (red), and d) calculated $d(\omega)^{rel}$ spectra for apo-CaM.

The transition dipole analysis was also performed for the *holo*-CaM data (Figure 3-17) and again the integral of the IR spectra was conserved with increasing temperature, and the integral of the 2D-IR diagonal decreases in intensity (3.2 to 2.2, Figure 3-17 (c)), though to less of an extent than *apo*-CaM (7.0 to 3.75, Figure 3 -16 (c)). The $d(\omega)^{rel}$ spectra were calculated and a change in helical coupling due to changes in structural rigidity implied.

The smaller signal size of *holo*-CaM and the smaller gain of random coil as suggested from the PCA analysis hints that although the protein states have similarities, they behave in different manners.

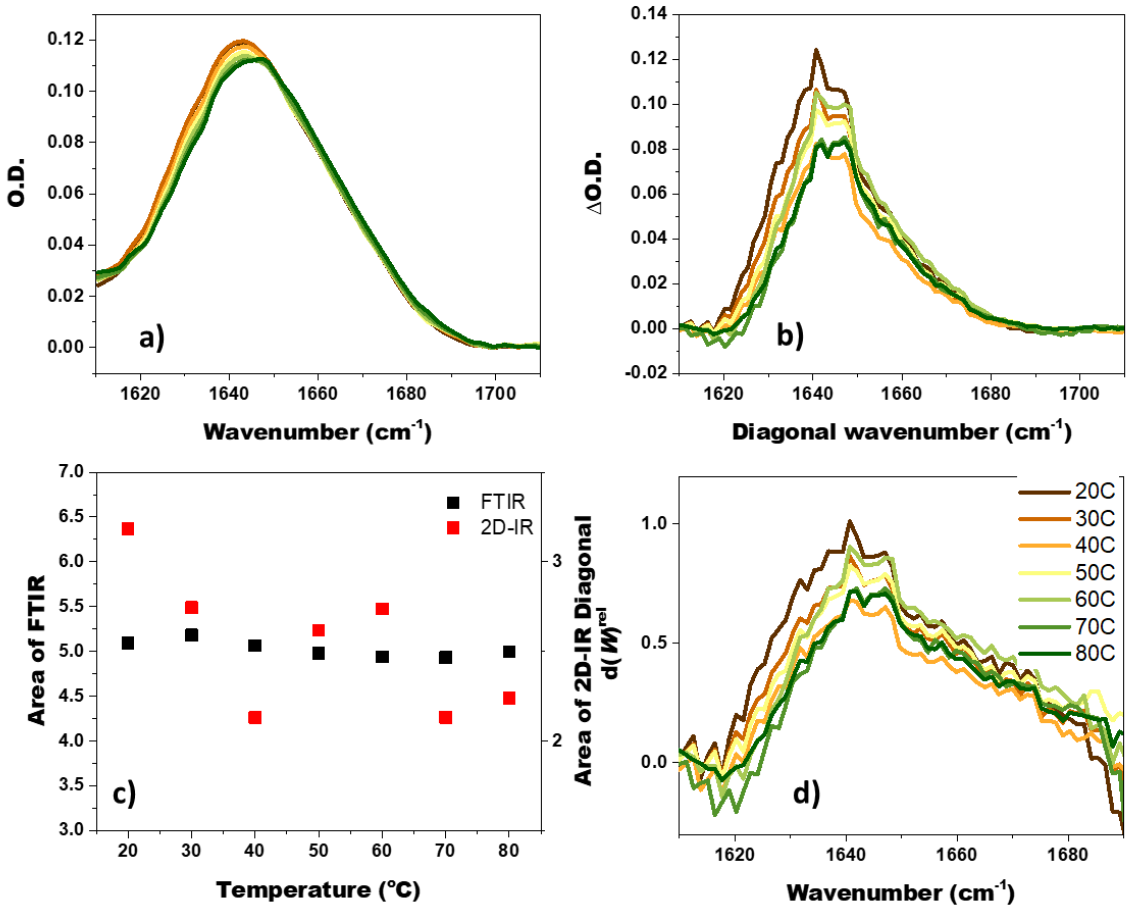


Figure 3-17. Transition dipole analysis for holo-CaM. a) Buffer-subtracted unnormalised linear IR spectra, and b) 2D-IR diagonal spectra of holo-CaM. Spectra have been inverted from negative to positive to compare with the positive FTIR spectra. c) integrated areas for each temperature for FTIR (black) and 2D-IR diagonal (red), and d) calculated $d(\omega)^{rel}$ spectra for holo-CaM.

3.5.2 MD simulations

The theoretical changes in secondary structure predicted by MD simulations for *apo*-CaM and *holo*-CaM agree well with previous experiments[37,38] in terms of the overall structure (Figures 3-18 (a) and (b)). The 20 °C simulation found the predicted α-helical content to be 52% (vs 56% from experimental) for *apo*-CaM and 58% (vs 62%) for *holo*-CaM (Table 3-3).

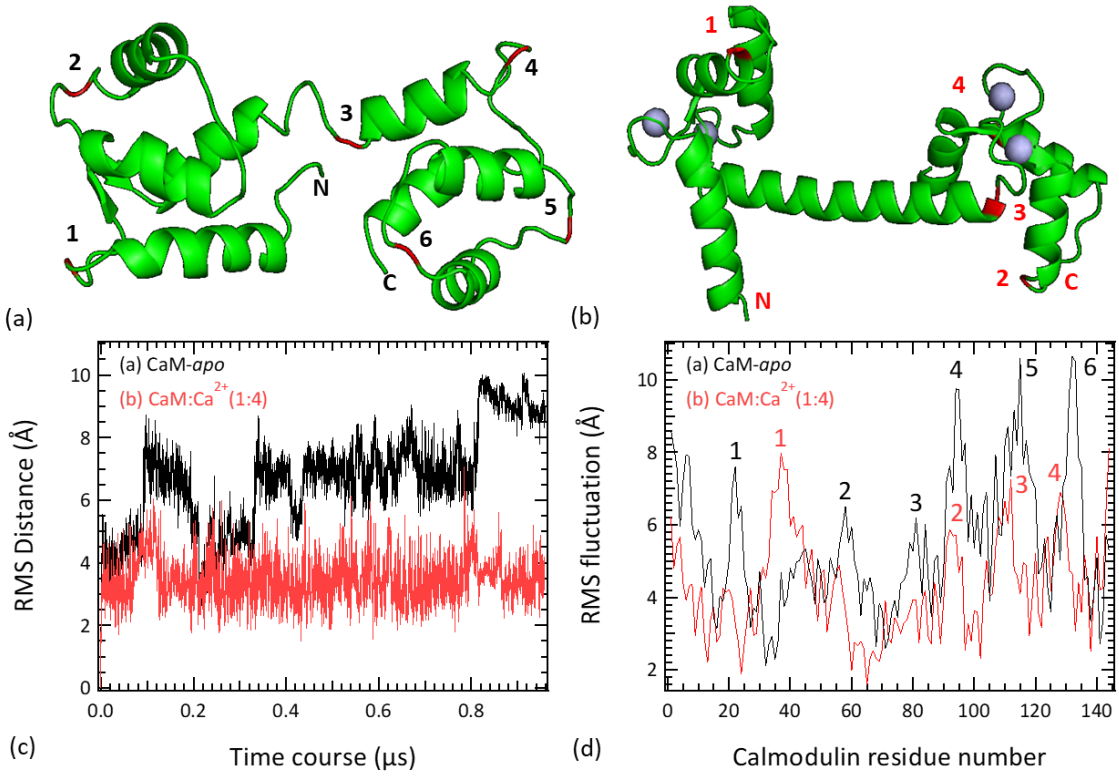


Figure 3-18. MD simulations at 20 °C from GROMACS results a) predicted apo-CaM structure, b) predicted holo-CaM structure, c) RMS distance time course plot for apo-CaM (black) and holo-CaM (red), and d) RMS fluctuations per residue for apo- (black) and holo-CaM (red). Positions for the amino acids displaying the greatest fluctuations are marked in the structures using numbers in (a) and (b).

The root mean square (RMS) distance parameter, effectively a direct measure of the relative stability of each CaM conformation, is the RMS of differences in distance between atom-pairs in the reference structure (PDB entry structure, 1CFD for *apo*-CaM and 1CLL for *holo*-CaM) and the calculated structure at time t. The high RMS-distance average of *apo*-CaM (7 Å) demonstrates a high degree of variance from the starting structure, indicative of a flexible structure. The average RMS-distance of *holo*-CaM (3.5 Å) is lower and is indicative of a more rigid structure compared to *apo*-CaM (Figure 3-18 (c)).

The RMS-fluctuation parameter, which identifies the residues displaying the greatest fluctuation, provides further insight into the differences between *apo*-CaM and *holo*-CaM. For *apo*-CaM (black), residues 90-148 have the largest deviation representing flexibility in the C-domain. *Holo*-CaM shows a marked decrease in fluctuations for residues 90-148, representing a more rigid C-domain (Figure 3-18 (d)).

The simulations were repeated at 80 °C with the GROMACS method but returned negligible structural changes, mainly due to the simulation being unable to sufficiently sample the timescales required at higher temperatures over the 1 μs simulated timescale. An enhanced form of sampling, which increases the sampling rate using several replicates at different temperatures known as REST2 sampling was applied to 20 °C and 80 °C simulations. Comparing the two approaches for *apo*-CaM at 20 °C showed good agreement with 54% and 52% α-helix content from REST2 and GROMACS, respectively. The REST2 simulation predicted a decrease of 9% in α-helix content at 80 °C for *apo*-CaM, and a smaller decrease of 4% in α-helix content at 80 °C for *holo*-CaM (Table 3-3).

Table 3-3. Predicted secondary structure content from MD simulations for both GROMACS and REST2 methods for apo- and holo-CaM at room temperature and elevated temperature. Errors are standard deviations of mean across the 1 μs simulation.

Structure	**Apo-CaM**	**Holo-CaM**	**Apo-CaM**	**Apo-CaM**	**Holo-CaM**
Temp (°C)	20	20	20	80	80
Model	GROMACS	GROMACS	REST2	REST2	REST2
Random coil (%)	17 ± 2	13 ± 1	16 ± 3	21 ± 4	15 ± 2
β-sheet (%)	7 ± 1	5 ± 1	7 ± 2	2 ± 1	4 ± 1
α-helix (%)	52 ± 3	58 ± 2	54 ± 4	45 ± 7	53 ± 3

3.5.3 Quantification of secondary structure change

It is clear from PCA and transition dipole analysis of the 2D-IR data that a transition from α-helix to random coil occurs in *apo*-CaM and this is in keeping with the results seen from CD measurements and MD simulations. Now the aim is to determine the quantitative accuracy of the 2D-IR methods. To evaluate this the assignment of the 1635 cm^{-1} and 1665 cm^{-1} features in the IR data (α-helix and random coil, respectively) have been used to allow direct comparison with the quantitative results from the fitting tools used for CD spectroscopy and from MD simulations.

Although previous publications have shown that it is possible to obtain absolute values of secondary structure content from IR data[10], it is not a trivial process. Instead, this chapter will

focus on quantifying changes in secondary structure. It must be noted that the methods used here are a first attempt at quantifying changes in structure of α-helix, with a critical analysis of the results discussed later. Comparisons are made to the quantified CD spectra results where it has already been discussed that the algorithms used for assessing secondary structure content were not optimal for elevated temperature measurements, and spectral changes can be the result of both a structural change and weakening of H-bonding without a change in secondary structure content[48].

The starting point for the FTIR and 2D-IR data for the α-helix and random coil values were taken from the predicted results of the REST2 method from the MD simulations at 20 °C, as these values are consistent with structural studies. The peak positions of the α-helix and random coil peaks in the PC2 spectrum were used to evaluate the original data for both FTIR and 2D-IR spectroscopy. The percentage difference from the starting measurement, 20 °C, then was calculated for each of the temperatures using the relative change in the magnitude of the signals. The percentage differences were utilised in the calculation of the predicted secondary structure content at each temperature. The predicted secondary structure content was calculated in the same manner for both *apo*-CaM and *holo*-CaM.

The 2D-IR data for *apo*-CaM reproduces the general trend observed from CD data of decreasing α-helix and increasing random coil with increasing temperature, apparently somewhat more effectively than FTIR, which shows a shallower gradient with temperature for both the α-helix and random coil percentages than 2D-IR. Closer inspection however shows that the detailed structure of the CD data, in terms of the sigmoidal decrease in α-helix content of apo-CaM is apparently recovered more effectively by linear IR than by 2D-IR spectroscopy (Figure 3-19).

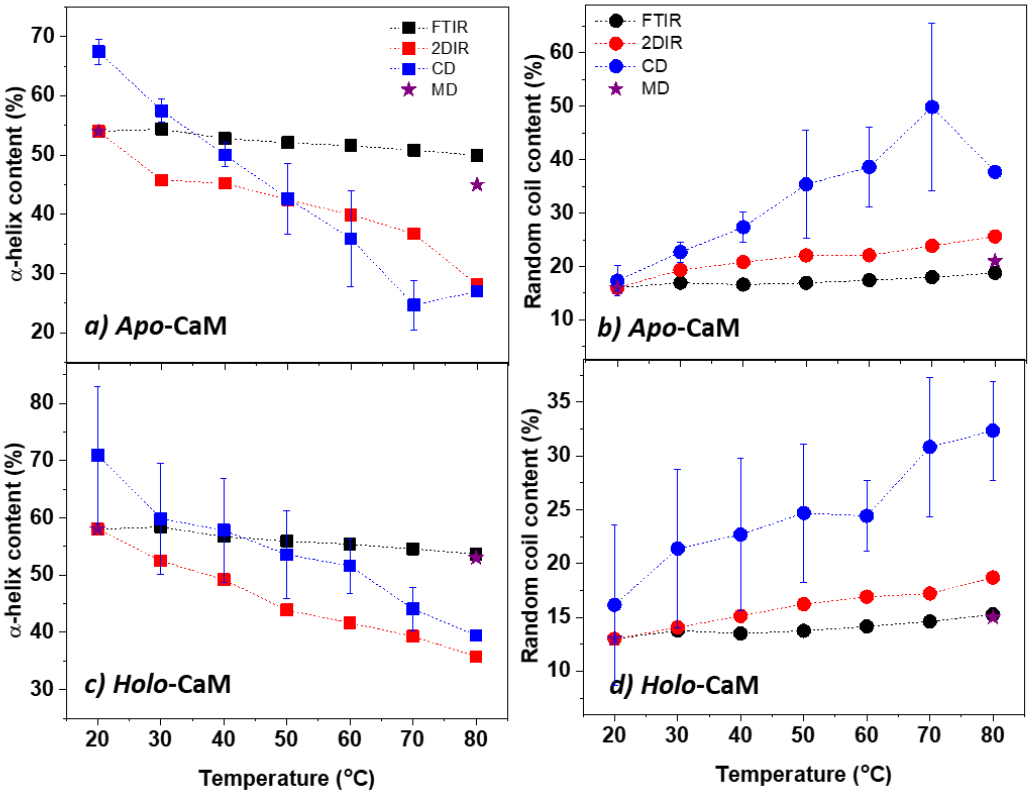


Figure 3-19. Quantitative comparison of secondary structure contributions obtained as a function of temperature for a) α-helix predictions for apo-CaM, b) random coil predictions for apo-CaM, c) α-helix predictions for holo-CaM, and d) random coil predictions for holo-CaM.

A large change in α-helix content is observed for both the CD data and 2D-IR data in *apo*-CaM, loss of 40% and 26%, respectively. A slightly smaller loss is apparently observed for *holo*-CaM but still a significant amount, 32% and 22% for CD and 2D-IR, respectively (Table 3-4).

An important note when considering the results of the CD spectroscopy is that the different data analysis algorithms return a spread of results, leading to an uncertainty in the values recovered. This is reflected in the large error bars on the CD data.

Table 3-4. Summary of percentage changes of α-helix and random coil of apo-CaM and holo-CaM from 20 °C to 80 °C.

	Method	α-helix (%)	random coil (%)
apo-CaM	CD	67-27%	17-38%
	2D-IR	54-28%;	16-26%
	FTIR	54-50%	16-19%
	Rest2 MD	54-45%	16-21%
holo-CaM	CD	71-39%	16-32%
	2D-IR	58-36%	13-19%
	FTIR	58-54%	16-19%
	Rest2 MD	58-53%	13-15%

The predicted secondary structure content shows similar values for *apo*-CaM and *holo*-CaM between 20 °C and 80 °C with *holo*-CaM displaying a slightly smaller change, whereas the DSC data suggests a marked difference between the two states with elevated temperatures, with *holo*-CaM not showing a transition until 96 °C. This indicates that changing the temperature has two effects: 1) heating the α-helix without triggering a secondary structure transition; 2) the occurrence of a transition. It can be seen from the comparison of data that the former is present in both *apo*-CaM and *holo*-CaM while the unfolding transition is only present in *apo*-CaM.

In analysing the data to extract quantitative information on secondary structure content, 2D-IR and CD data are quantitatively comparable in that they reflect similar aspects of the molecular structure, i.e. the interactions of residues within the α-helix. An important consideration is that heating of the protein will increase the average inter-residue separations, thereby reducing the interactions and signal amplitudes without a reduction in α-helix content. This is especially true for the 2D-IR approach where coupling-induced spectral changes are targeted, as discussed earlier. Therefore, it is an oversimplification to interpret the reduction of α-helix signal as a quantitative loss of structural content, as opposed to thermal modification of the signal.

The challenge is to differentiate between the two effects in the data analysis. DSC results suggest that *holo*-CaM does not undergo a transition in the 20 to 80 °C range measured by IR and CD spectroscopy, which both show a temperature dependent effect. The use of *holo*-

CaM as reference to differentiate the two thermal effects will be investigated. Comparing the α-helix content from CD and 2D-IR methods for *apo*- and *holo*-CaM shows that the magnitude of differences between the two protein states peaks at 60 °C, where domain melting occurs in *apo*-CaM, observed from the DSC data (Figure 3-20).

Taking the difference in predicted α-helix content at 60 °C indicates a reduction of 13% (CD) and 15% (2D-IR) which can be attributed to the unfolding transition. This brings the predicted structural change of CD and 2D-IR closer to the values of linear IR (5%) and MD simulations (9%) which are conservative in their predictions.

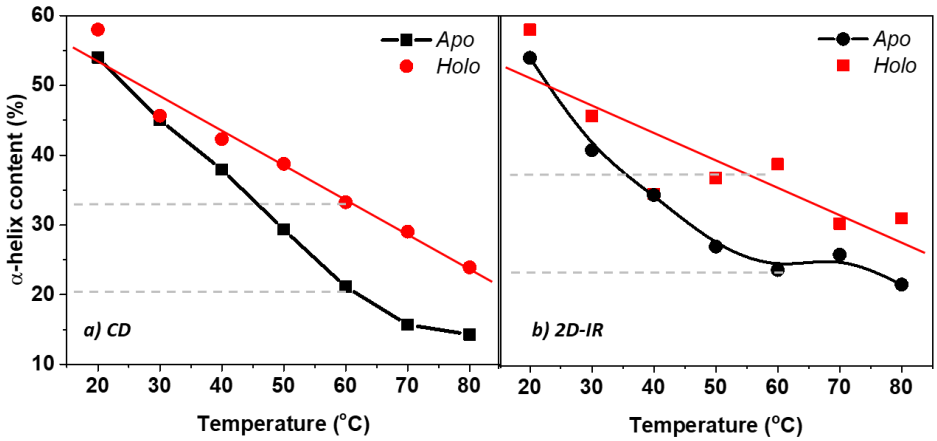


Figure 3-20. Predicted α-helix percentages as a function of temperature for apo-CaM (black) and holo-CaM (red) for a) CD spectroscopy and b) transition dipole analysis of 2D-IR data. Lines are to guide the eye.

3.5.4 Detecting the *apo*-CaM to *holo*-CaM Transition

The testing of small changes in secondary structure content will be investigated. This will be performed as the large change in temperature (20 °C to 80 °C) would not likely be viable in a temperature-jump (T-jump) type experiment, therefore it is imperative to determine if 2D-IR is sensitive enough to small changes in secondary structure before a full-scale T-jump experiment is conducted, which will be the focus of chapter (4). The *apo*-CaM to *holo*-CaM transition will be investigated as the binding of Ca^{2+} results in just a 5% change in the α-helix, equivalent to a change in just 7-8 residues.

The 20 °C measurements were used in the analysis which eliminates the contribution from thermal effects that affected the prediction of secondary structure changes above. For both FTIR and 2D-IR the difference spectrum between *apo*-CaM and *holo*-CaM was calculated. For the FTIR spectrum (Figure 3-21 (a)) the signal to noise ratio is poor but a loss of spectral

density near 1650 cm^{-1} is accompanied by a gain at 1635 cm^{-1}, as would be expected from the central linker of the protein changes from random coil to α-helix.

The effect is much clearer in the 2D-IR difference spectrum where it appears to be the reverse of that obtained for increasing the temperature of the *apo*-protein. A gain of signal is observed near 1640 cm^{-1} (black arrow) with a loss of signal at higher wavenumber (orange arrow) (Figure 3-21 (b)). It is noted that the random coil portion of the signal is significantly smaller in this case and that the α-helix signal dominates, as would be expected arising from the larger transition dipole moment of the α-helix transition.

The 1635 cm^{-1} peak was used to calculate the estimated % change of α-helix between *apo* and *holo*-CaM. This was calculated by dividing the difference in peak amplitudes by the *apo*-CaM 1635 cm^{-1} amplitude. The estimated change in α-helix was then determined using the *apo*-CaM MD simulation value of 54% α-helix as a starting point and was found to be 12% in the 2D-IR data, though this value could be potentially distorted due to different coupling. This somewhat overestimates the data from structural studies and MD predictions, as well as that published using FTIR spectroscopy (~6%)[24].

Although the 2D-IR data does not provide an accurate quantified change in secondary, the signal to noise ratio is much better than for the FTIR which is an important consideration for kinetic measurements, reinforcing the benefit of 2D-IR spectroscopy.

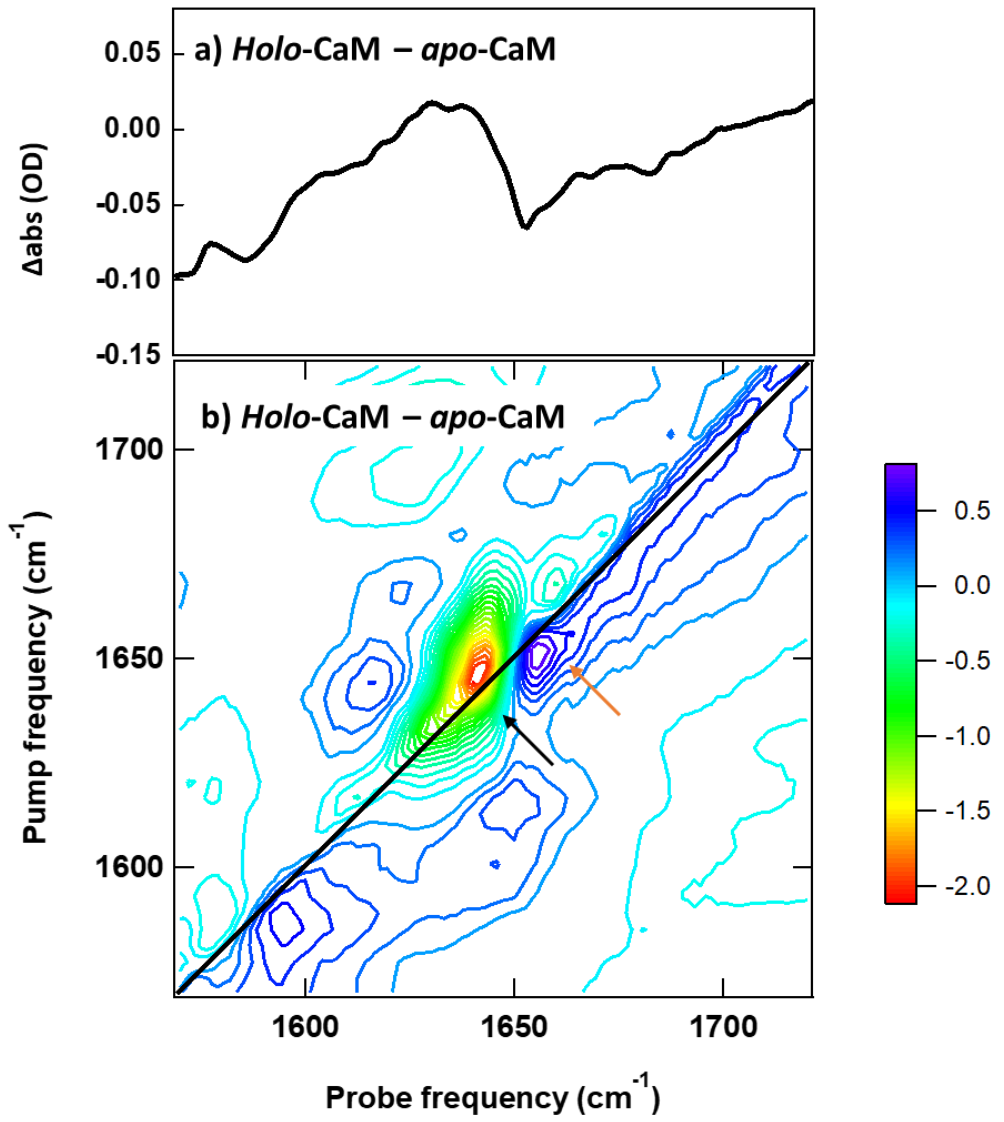


Figure 3-21. Apo-CaM to holo-CaM transition detected by a) FTIR difference spectroscopy (holo- minus apo-) and b) 2D-IR difference spectroscopy.

3.6 Conclusion and further work

Linear IR spectroscopy and 2D-IR spectroscopy have been used, coupled with difference spectroscopy and PCA, to detect changes in secondary structure in *apo*- and *holo*-CaM upon thermal perturbations between 20 °C to 80 °C. Both difference spectroscopy and PCA gave comparable results, with PCA producing clearer spectra due to the noise cancelling ability of the method. The results inferred a loss of α-helix structure and a gain of random coil structure with increasing temperature.

Due to overlapping contributions of α-helix and random coil in the amide I region, transition dipole analysis was employed to calculate the relative transition dipole spectra, $d(\omega)^{rel}$, by ratioing the linear IR and 2D-IR diagonal spectra and introducing a relative calibrant. The resulting spectra confirmed the assignment of α-helix and of random coil due to changing in coupling observed for the α-helix and not the random coil peak with increasing temperature.

MD simulations were performed at both low and high temperature and the predicted helical change was more conservative at 54% to 45% for *apo*-CaM. Alongside, DSC data provided evidence of unfolding transitions of *apo*-CaM at 46 °C and 58 °C, whereas the *holo*-CaM protein was much more stable with transitions not occurring until 96 °C and 111 °C.

It was determined that the changes in α-helix from CD and 2D-IR spectroscopy were a combination of both a structural transition and thermally induced alterations to the coupling that does not result in secondary structure content. *Holo*-CaM was deemed to possess only the latter, and through comparison of *apo*- and *holo*-CaM it was calculated that the change in α-helix was approximately 13-15% for *apo*-CaM.

Foundations have now been set for the use of 2D-IR spectroscopy with PCA to detect small secondary structure changes. One of the main benefits of the technique over CD spectroscopy is the ability to observe dynamics of proteins. Therefore, the information learnt from equilibrium measurements of CaM can be applied to time-resolved vibrational spectroscopy.

In addition, relatively large changes in structure were observed over a temperature range of 20 °C to 80 °C, and although investigating smaller changes was applied with respect to the *apo*- to *holo*-CaM transition, further analysis into the ability of 2D-IR detecting smaller changes in structure can be explored.

3.7 References

(1) Manor, J.; Mukherjee, P.; Lin, Y.-S.; Leonov, H.; Skinner, J. L.; Zanni, M. T.; Arkin, I. T. Gating Mechanism of the Influenza A M2 Channel Revealed by 1 and 2D-IR Spectroscopies. *Struc* **2009**, *17*, 247–254. https://doi.org/10.3816/CLM.2009.n.003.Novel.

(2) Strasfeld, D. B.; Ling, Y. L.; Shim, S.-H.; Zanni, M. T. Tracking Fiber Formation in Human Islet Amyloid Polypeptide with Automated 2D-IR Spectroscopy. *J. Am. Chem. Soc.* **2008**, *130*, 6698–6699. https://doi.org/10.1038/jid.2014.371.

(3) Ghosh, A.; Qiu, J.; DeGrado, W. F.; Hochstrasser, R. M. Tidal Surge in the M2 Proton Channel, Sensed by 2D IR Spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (15), 6115–6120. https://doi.org/10.1073/pnas.1103027108.

(4) Moran, S. D.; Woys, A. M.; Buchanan, L. E.; Bixby, E.; Decatur, S. M.; Zanni, M. T. Two-Dimensional IR Spectroscopy and Segmental 13C Labeling Reveals the Domain Structure of Human ГD-Crystallin Amyloid Fibrils. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (9), 3329–3334. https://doi.org/10.1073/pnas.1117704109.

(5) Cheng, M.; Brookes, J. F.; Montfort, W. R.; Khalil, M. PH-Dependent Picosecond Structural Dynamics in the Distal Pocket of Nitrophorin 4 Investigated by 2D IR Spectroscopy. *J. Phys. Chem. B* **2013**, *117*, 15804–15811.

(6) Demirdöven, N.; Cheatum, C. M.; Chung, H. S.; Khalil, M.; Knoester, J.; Tokmakoff, A. Two-Dimensional Infrared Spectroscopy of Antiparallel Beta-Sheet Secondary Structure. *J. Am. Chem. Soc.* **2004**, *126* (25), 7981–7990.

(7) Ganim, Z.; Chung, H. S.; Smith, A. W.; Deflores, L. P.; Jones, K. C.; Tokmakoff, A. Amide I Two-Dimensional Infrared Spectroscopy of Proteins. *Acc. Chem. Res.* **2008**, *41* (3), 432–441. https://doi.org/10.1021/ar700188n.

(8) Smith, A. W.; Cheatum, C. M.; Chung, H. S.; Demirdoven, N.; Khalil, M.; Knoester, J.; Tokmakoff, A. Two-Dimensional Infrared Spectroscopy of Beta-Sheets and Hairpins. *Biophys. J.* **2004**, *86* (1), 619A-619A.

(9) Woys, A. M.; Lin, Y. S.; Reddy, A. S.; Xiong, W.; De Pablo, J. J.; Skinner, J. L.; Zanni, M. T. 2D IR Line Shapes Probe Ovispirin Peptide Conformation and Depth in Lipid Bilayers. *J. Am. Chem. Soc.* **2010**, *132* (8), 2832–2838. https://doi.org/10.1021/ja9101776.

(10) Baiz, C. R.; Peng, C. S.; Reppert, M. E.; Jones, K. C.; Tokmakoff, A. Coherent Two-Dimensional Infrared Spectroscopy: Quantitative Analysis of Protein Secondary Structure in Solution. *Analyst* **2012**, *137* (8), 1793–1799.

(11) Heidorn, D. B.; Trewhella, J. Comparison of the Crystal and Solution Structures of Calmodulin and Troponin C. *Biochemistry* **1988**, *27* (3), 909–915.

(12) Wu, G.; Gao, Z.; Dong, A.; Yu, S. Calcium-Induced Changes in Calmodulin Structural Dynamics and Thermodynamics. *Int. J. Biol. Macromol.* **2012**, *50* (4), 1011–1017.

https://doi.org/10.1016/j.ijbiomac.2012.02.017.

(13) Tripathi, S.; Portman, J. J. Inherent Flexibility Determines the Transition Mechanisms of the EF-Hands of Calmodulin. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (7), 2104–2109. https://doi.org/10.1073/pnas.0806872106.

(14) Linse, S.; Helmersson, A.; Forsen, S. Calcium Binding to Calmodulin and Its Globular Domains. *J. Biol. Chem.* **1991**, *266*, 8050–8054.

(15) Zhang, M.; Tanaka, T.; Ikura, M. Calcium-Induced Conformational Transition Revealed by the Solution Structure of Apo Calmodulin. *Nat. Struct. Biol.* **1995**, *2* (9), 758–767.

(16) Wang, Q.; Liang, K. C.; Czader, A.; Waxham, M. N.; Cheung, M. S. The Effect of Macromolecular Crowding, Ionic Strength and Calcium Binding on Calmodulin Dynamics. *PLoS Comput. Biol.* **2011**, *7* (7), e1002114/1-e1002114-16.

(17) Protasevich, I.; Ranjbar, B.; Lobachov, V.; Makarov, A.; Gilli, R.; Briand, C.; Lafitte, D.; Haiech, J. Conformation and Thermal Denaturation of Apocalmodulin: Role of Electrostatic Mutations. *Biochemistry* **1997**, *36* (8), 2017–2024. https://doi.org/10.1021/bi962538g.

(18) Kleinjung, J.; Fraternali, F.; Martin, S. R.; Bayley, P. M. Thermal Unfolding Simulations of Apo-Calmodulin Using Leap-Dynamics. *Proteins Struct. Funct. Genet.* **2003**, *50* (4), 648–656. https://doi.org/10.1002/prot.10331.

(19) Masino, L.; Martin, S. R.; Bayley, P. M. Ligand Binding and Thermodynamic Stability of a Multidomain Protein, Calmodulin. *Protein Sci.* **2000**, *9* (8), 1519–1529. https://doi.org/10.1110/ps.9.8.1519.

(20) Browne, J. P.; Strom, M.; Martin, S. R.; Bayley, P. M. The Role of Beta-Sheet Interactions in Domain Stability, Folding, and Target Recognition Reactions of Calmodulin. *Biochemistry* **1997**, *36* (31), 9550–9561.

(21) Yamniuk, A. P.; Ishida, H.; Lippert, D.; Vogel, H. J. Thermodynamic Effects of Noncoded and Coded Methionine Substitutions in Calmodulin. *Biophys. J.* **2009**, *96* (4), 1495–1507. https://doi.org/10.1016/j.bpj.2008.10.060.

(22) Trewhella, J.; Liddle, W. K.; Heidorn, D. B.; Strynadkaf, N. Calmodulin and Troponin C

Structures Studied by Fourier Transform Infrared Spectroscopy : Effects of Ca(2+) and Mg(2+) Binding. *Biochemistry* **1989**, *28*, 1294–1301.

(23) Yu, T.; Wu, G.; Yang, H.; Wang, J.; Yu, S. International Journal of Biological Macromolecules Calcium-Dependent Conformational Transition of Calmodulin Determined by Fourier Transform Infrared Spectroscopy. *Int. J. Biol. Macromol.* **2013**, *56*, 57–61. https://doi.org/10.1016/j.ijbiomac.2013.02.004.

(24) Jackson, M.; Haris, P. I.; Chapman, D. Fourier Transform Infrared Spectroscopic Studies of Ca(2+) -Binding Proteins. *Biochemistry* **1991**, *30*, 9681–9686.

(25) Rainteau, D.; Wolf, C.; Lavialle, F. Effects of Calcium and Calcium Analogs on Calmodulin: A Fourier Transform Infrared and Electron Spin Resonance Investigation. *Biochim Biophys Acta* **1989**, *1011*, 81–87.

(26) Creon, A.; Josts, I.; Niebling, S.; Huse, N.; Tidow, H. Conformation-Specific Detection of Calmodulin Binding Using the Unnatural Amino Acid p-Azido-Phenylalanine (AzF) as an IR-Sensor. *Struct. Dyn.* **2018**, *5* (6), 064701/1-064701/12.

(27) Edington, S. C.; Gonzalez, A.; Middendorf, T. R.; Halling, D. B.; Aldrich, R. W.; Baiz, C. R. Coordination to Lanthanide Ions Distorts Binding Site Conformation in Calmodulin. *Proc. Natl. Acad. Sci.* **2018**, *115* (14), E3126–E3134. https://doi.org/10.1073/pnas.1722042115.

(28) Edington, S. C.; Halling, D. B.; Bennett, S. M.; Middendorf, T. R.; Aldrich, R. W.; Baiz, C. R. Non-Additive Effects of Binding Site Mutations in Calmodulin. *Biochemistry* **2019**, *58* (24), 2730–2739. https://doi.org/10.1021/acs.biochem.9b00096.

(29) Sreerama, N.; Woody, R. W. A Self-Consistent Method for the Analysis of Protein Secondary Structure from Circular Dichroism. *Anal. Biochem.* **1993**, *209*, 32–44. https://doi.org/10.1006/abio.1993.1079.

(30) Sreerama, N.; Venyaminov, S. Y.; Woody, R. W. Estimation of the Number of Alpha-Helical and Beta-Strand Segments in Proteins Using Circular Dichroism Spectroscopy. *Protein Sci.* **1999**, *8* (2), 370–380. https://doi.org/10.1110/ps.8.2.370.

(31) van Stokkum, I. H.; Spoelder, H. J.; Bloemendal, M.; van Grondelle, R.; Groen, F. C. Estimation of Protein Secondary Structure and Error Analysis from Circular Dichroism Spectra. *Anal. Biochem.* **1990**, *191* (1), 110–118.

https://doi.org/http://dx.doi.org/10.1016/0003-2697(90)90396-Q.

(32) Provencher, S. W.; Glöckner, J. Estimation of Globular Protein Secondary Structure from Circular Dichroism. *Biochemistry* **1981**, *20* (1), 33–37. https://doi.org/10.1021/bi00504a006.

(33) Whitmore, L.; Wallace, B. A. Protein Secondary Structure Analyses from Circular Dichroism Spectroscopy: Methods and Reference Databases. *Biopolymers* **2008**, *89* (5), 392–400.

(34) Whitmore, L.; Wallace, B. A. DICHROWEB, an Online Server for Protein Secondary Structure Analyses from Circular Dichroism Spectroscopic Data. *Nucleic Acids Res.* **2004**, *32*, 668–673. https://doi.org/10.1093/nar/gkh371.

(35) Lees, J. G.; Miles, A. J.; Wien, F.; Wallace, B. A. A Reference Database for Circular Dichroism Spectroscopy Covering Fold and Secondary Structure Space. *Bioinformatics* **2006**, *22* (16), 1955–1962. https://doi.org/10.1093/bioinformatics/btl327.

(36) Böhm, G.; Muhr, R.; Jaenicke, R.; Bohm, G.; Muhr, R.; Jaenicke, R. Quantitative Analysis of Protein Far UV Circular Dichroism Spectra by Neural Networks. *Protein Eng.* **1992**, *5* (3), 191–195. https://doi.org/10.1093/protein/5.3.191.

(37) Kuboniwa, H.; Tjandra, N.; Grzesiek, S.; Ren, H.; Klee, C. B.; Bax, A. Solution Structure of Calcium-Free Calmodulin. *Nat. Struct. Mol. Biol.* **1995**, *2*, 768–776.

(38) Chattopadhyaya, R.; Meador, W. E.; Means, A. R.; Quiocho, F. A. Calmodulin Structure Refined at 1.7 Å Resolution. *J. Mol. Biol.* **1992**, *228*, 1177–1192.

(39) Kokubo, H.; Tanaka, T.; Okamoto, Y. Two-Dimensional Replica-Exchange Method for Predicting Protein-Ligand Binding Structures. *J. Comput. Chem.* **2013**, *34* (30), 2601–2614. https://doi.org/10.1002/jcc.23427.

(40) Jorgensen, W. L. Revised TIPS for Simulations of Liquid Water and Aqueous Solutions. *J. Chem. Phys.* **1982**, *77* (8), 4156. https://doi.org/10.1063/1.444325.

(41) Li, P.; Roberts, B. P.; Chakravorty, D. K.; Merz, K. M. Rational Design of Particle Mesh Ewald Compatible Lennard-Jones Parameters for +2 Metal Cations in Explicit Solvent. *J. Chem. Theory Comput.* **2013**, *9* (6), 2733–2748.

https://doi.org/10.1021/ct400146w.

(42) Kabsch, W.; Sander, C. Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* **1983**, *22* (12), 2577–2637. https://doi.org/10.1002/bip.360221211.

(43) Wang, L.; Friesner, R. A.; Berne, B. J. Replica Exchange with Solute Scaling: A More Efficient Version of Replica Exchange with Solute Tempering (REST2). *J. Phys. Chem. B* **2011**, *115* (30), 9431–9438. https://doi.org/10.1021/jp204407d.

(44) Minnes, L.; Shaw, D. J.; Cossins, B. P.; Donaldson, P. M.; Greetham, G. M.; Towrie, M.; Parker, A. W.; Baker, M. J.; Henry, A. J.; Taylor, R. J.; et al. Quantifying Secondary Structure Changes in Calmodulin Using 2D-IR Spectroscopy. *Anal. Chem.* **2017**, *89*, 10898–10906.

(45) Greenfield, N. J. Using Circular Dichroism Spectra to Estimate Protein Secondary Structure. *Nat. Protoc.* **2007**, *1*, 2876–2890.

(46) Rajalahti, T.; Kvalheim, O. M. Multivariate Data Analysis in Pharmaceutics: A Tutorial Review. *Int. J. Pharm.* **2011**, *417*, 280–290.

(47) Karamizadeh, S.; Abdullah, S. M.; Manaf, A. A.; Zamani, M.; Hooman, A. An Overview of Principal Component Analysis. *J. Signal Inf. Process.* **2013**, *4*, 173–175.

(48) Luo, P.; Baldwin, R. L. Mechanism of Helix Induction by Trifluoroethanol: A Framework for Extrapolating the Helix-Forming Properties of Peptides from Trifluoroethanol/Water Mixtures Back to Water. *Biochemistry* **1997**, *36*, 8413–8421.

(49) Grechko, M.; Zanni, M. T. Quantification of Transition Dipole Strengths Using 1D and 2D Spectroscopy for the Identification of Molecular Structures via Exciton Delocalization: Application to Alpha-Helices. *J. Chem. Phys.* **2012**, *137* (18), 184202/1-184202/9.

4 Investigating unfolding rates of calmodulin using temperature-jump IR spectroscopy

This chapter contains results published in the following publication:

Minnes, L.; Greetham, G.; Shaw, D. J.; Clark, I.; Fritzsch, R.; Towrie, M.; Parker, A.; Henry, A. J.; Taylor, R. J.; Hunt, N. T.
Uncovering the Early Stages of Domain Melting in Calmodulin with Ultrafast Temperature-Jump Infrared Spectroscopy. *J Phys Chem B*. **2019**, *123,* 8733-8739.

4.1 Abstract

Understanding folding and unfolding dynamics of proteins is a key aspect of understanding their biological functions. This has resulted in developments being made into experiments which look at the folding of peptides and proteins upon perturbation by a Temperature-jump (T-jump). Commonly, the T-jump setup uses an IR laser pulse to heat the water in the sample and time-resolved spectroscopy is used to probe the folding / unfolding dynamics. The experiments undertaken within this chapter utilise a newly developed laser setup with the excitation of the O-D stretch within the samples at 3.7 μm by a nanosecond laser creating the T-jump. The protein calmodulin (CaM) was used to assess the suitability of the system. The dynamics underlying the static structure rearrangements in the previous chapter with a probe of the amide I region measuring the unfolding and subsequent refolding of the predominately α-helix protein were investigated. Comparison of the *apo*-CaM (Ca^{2+} free) and *holo*-CaM (Ca^{2+} bound) protein complements the previously revealed data on the C- and N-domain transitions by DSC and 2D-IR spectroscopy (Chapter 3). New dynamics information on *apo*-CaM uncovered a fast unfolding decay constant of 5 μs representative of α-helix destabilisation and a refolding which follows the cooling of the sample. These observations enable the assignment of previously reported dynamics of CaM on hundreds of microsecond timescales to thermally activated melting, producing a complete mechanism for thermal unfolding of CaM.

4.2 Introduction

In chapter (3) the temperature induced changes in structure of the protein calmodulin (CaM) were measured using 2D-IR. These changes have been measured at equilibrium heating and established the temperature-dependence of the CaM unfolding to lay the groundwork for this chapter which measures the dynamics. One of the large benefits of time-resolved spectroscopy is the ability to gain insight into protein dynamics.

Whereas time-resolved spectroscopy such as pump-probe and 2D-IR spectroscopy are able to observe ultrafast equilibrium dynamics on the femtosecond to picosecond timescales of proteins at a particular temperature, it is limited in the information that can be acquired at this timescale by the picosecond vibrational lifetime of the amide I. Consequently, valuable information on the non-equilibrium dynamics of folding or unfolding kinetics of proteins

cannot be obtained, where the folding of proteins can provide insight into their biological functions. These non-equilibrium dynamics will be probed in this chapter.

Rapid mixing methods such as stopped-flow are capable of measuring kinetics down to several milliseconds[1] and have seen improvements which allow processes as fast as tens of microseconds to be observed[2] but miss out on folding/ unfolding information which have timescales of nanosecond to microsecond[3–5] due to restrictions on the mechanical processes. Photo or photochemical experiments can induce folding/unfolding of proteins[6,7] with the ability to probe picosecond to nanosecond dynamics[8,9], however this requires chemical modification, such as adding a photoswitch to the peptide, to have occurred which is undesirable. Photoswitches give extremely accurate control of the location of the change and the type of perturbation that the peptide is subjected to. The downside is that they typically only work on small peptides and are likely to perturb the native structure and to change it in a way that is non-natural. A method of looking at the folding and unfolding of proteins without modification being required which is not restricted by the mechanical processes that are found in rapid mixing is a temperature jump (T-jump) experiment[10]. This combines the heating aspect of the previous chapter with the ability to look at dynamics on the timescales on the order of nanoseconds out to a few microseconds.

The stability of proteins is highly dependent on the temperature of its environment and, as observed in chapter (3), a change in temperature can induce unfolding to occur in the protein. A rise in temperature that is fast relative to the natural unfolding dynamics of the protein allows us to measure the T-jump induced folding/ unfolding kinetics. This method does not require any chemical additives or labelling to track folding, but the use of labelling has proven useful in the tracking of specific sites and regions within the protein[11]. Adding a dye to the protein solution and heating the dye has also given positive results[12], where heating a crystal violet dye in the sample with a T-jump of 3-6 °C induced β-sheet disruption of RNase A. However, this can lead to interference as high concentrations of dye have been added to the sample. The best method for biological samples, therefore is to heat the water in the sample directly[13].

The T-jump experiment has been performed within many types of spectroscopy including, but not limited to, fluorescence[14], CD[15], and vibrational[16]. A usual T-jump setup employs the use of a continuous wave (CW) laser in conjunction with a fast detector, or a nanosecond laser as the pump to induce the temperature rise. This nanosecond laser usually has a lower

repetition rate than the probe laser which means measurements can be recorded out until the millisecond timescales with the use of electronic delays. Frequently, the pump laser will excite the O-D overtone band at 2 μm, with a repetition rate of 1 - 20 Hz[17–19]. This repetition rate results in increased acquisition time and reduced signal to noise. However, achieving a usable T-jump, for example 10 °C, requires a large amount of energy, 9 mJ at the sample has been previously reported for a T-jump of this magnitude[20], and the wavelengths are low in intensity. The use of high powered lasers, such as a 1064 nm Nd:YAG, are capable of producing the energy required for a T-jump using the O-D overtone, and these have low repetition rates in order to generate the pulse energies. Also to note, the near IR spectra of H_2O and D_2O are temperature dependent[21], i.e. the absorption increases with temperature which results in larger temperature rises at higher initial temperatures.

The first IR laser T-jump setup was constructed in 1996[16] and involved the use of a 10 ns pulse duration Nd:YAG laser to generate the 18 °C temperature rise by exciting the D_2O overtone band. This type of laser setup has been utilised in subsequent studies[11,18] that produced a 5 °C rise and determined that the C-domain of α-helix peptides folds more rapidly than the N-domain[11]. Also created a 10 °C rise in a β-hairpin peptide, Trpzip2, which resulted in a decay constant of a few microseconds[18]. Further developments have been made with a transient 2D-IR setup for T-jump[19,20,22,23]. These studies have involved the unfolding of ubiquitin, insulin dimer formation and Trpzip2 peptide folding.

This chapter focuses on measurements using a newly developed T-jump laser setup at RAL which varies from previous applications by utilising a 1 kHz pump laser exciting the 3.7 μm region of the O-D stretch. By exciting this region, changing the initial set temperature should have less of an impact on the magnitude of the temperature jump as opposed to exciting the O-D overtone. In addition, the use of a higher repetition rate laser will reduce the acquisition time than if a conventional few Hz nanosecond laser was used. This will in turn shorten the timescale the probe can acquire, but the relevant unfolding transitions will be captured on the early timescales in the microsecond region.

Unlike chapter (3), which involved 2D-IR spectroscopy, the work presented here will be time-resolved infrared spectroscopy (TRIR), as the new system required testing and calibrating on the simpler TRIR system for unfolding of proteins prior to being developed into a T-jump 2D-IR setup.

Many previous T-Jump studies have focused on the folding and unfolding of small peptides containing a single α-helix[11,24,25] with most results producing sub-microsecond timescales. This includes T-jump results from IR, fluorescence[26], and ultraviolet resonance Raman spectroscopy[27]. Not only monomeric α-helices, but also monomeric β-hairpins[18,28] have undergone studies where the fast folding rates of these structures give insights on how proteins fold and behave. However, the work in this chapter will represent heating of whole proteins, not purely small monomeric structures.

A significant point of interest, which has garnered numerous studies, has been the formation of α-helix from random coil[24,29,30]. This formation of helix is by a nucleation event followed by a series of propagation steps[31]. Multiple factors affect the timescale of folding, including the peptide length[32], the type of residues making up the peptide or protein[33], what domain the structure is in[11,34], and the environmental conditions of the protein or peptide with which the measurements occurred in[35,36]. This last factor means comparing the exact results between numerous groups is challenging unless the conditions, such as pD and salt concentration, used are the same.

Although most studies have shown helices to have short timescales on the sub-microsecond range, a previous experiment has shown a transition from picosecond to microsecond for a 16-residue peptide with measurements between 8 °C and 49 °C[29], which highlights the complexity of kinetics. This is further emphasised in a study which found that the helix from a ribosomal protein, L9, produced kinetics of approximately 2 μs which is considerably slower than alanine-rich peptide helices[33].

The formation and unfolding of β-hairpins have also been the subject of many studies[37–39], for example a range of 1 μs to 6 μs for the unfolding/ folding has been observed. Expanding on TRIR experiments, a 2D-IR T-jump setup has been used to investigate the unfolding of β-hairpins which was found to occur in a stepwise manner from folded to misfolded and then disordered[23]. The unfolding of ubiquitin has been reported to occur on two timescales. Hundreds of μs dynamics were assigned to the crossing of energetic barriers in response to the elevated temperature, but rapid unfolding processes taking just a few μs were also observed and attributed to downhill unfolding, constituting the first steps of the structural rearrangement[40].

The comparison between T-jump studies of α-helix structures and β-hairpin structures shows that folding and unfolding occurs quicker in less stable peptides and proteins.

Unlike the many helix formation studies, this work will investigate the unfolding of domains at elevated temperatures. Initial calibration of the setup will occur with the simple molecule trifluoracetic acid (TFA) which has carboxylate stretch at 1673 cm^{-1} which is within the region of interest, the amide I (Figure 4-1).

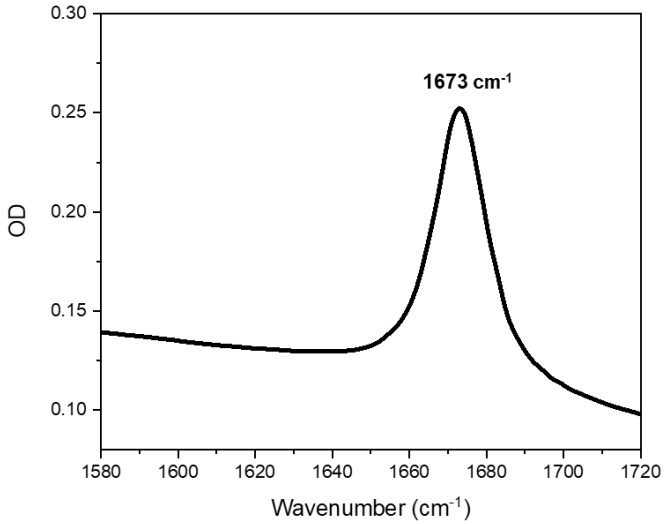


Figure 4-1. Linear IR spectrum of TFA from 1580 cm^{-1} to 1720 cm^{-1} focussing on the carboxylate stretch.

The protein to test the system will be calmodulin (CaM). It has previously undergone a T-jump experiment[41] using changes in the absorbance of tyrosine residues in CaM to probe melting where the timescale reported for C-domain unfolding was a relaxation rate of ~ 100 µs. This experiment had a measurement range of 1 µs to 16 ms with a signal rise time of 5.6 µs indicating slow heating of the sample which resulted in dynamics below this rise time being undetected.

This chapter will involve measurements of both the Ca^{2+} bound (*holo*-CaM) and the *apo*-CaM forms at a series of different temperatures that have been chosen to capture the C- and N-domain transitions which were determined from the DSC data in chapter (3). Comparisons will be made to the results in the previous chapter and previously published results.

4.3 Methods

4.3.1 Sample preparation

Trifluoroacetic acid (TFA), D_2O, Trizma base, DCl, and $CaCl_2$ were all purchased from Sigma-Aldrich and used without further purification. The CaM was expressed and purified by Daniel Shaw at UCB. The T-jump calibration sample consisted of 130 mM TFA in D_2O. The protein

samples consisted of 1.2 mM CaM in 50 mM deuterated Tris-DCl pD 7.4 for *apo*-CaM, and the addition of 4.8 mM $CaCl_2$ for a 1:4 CaM:Ca^{2+} ratio for *holo*-CaM. All samples were housed in a temperature-controlled Harrick cell and prepared between two CaF_2 windows separated by a PTFE spacer with a thickness of 6 µm.

4.3.2 Infrared Spectroscopy

All linear infrared spectra were collected on a Thermo Scientific Nicolet iS10 spectrometer at a resolution of 1 cm^{-1}. TFA sample was measured from 20 °C to 36 °C in 1 °C steps to produce a calibration curve.

4.3.3 T-jump setup

The T-jump pump laser[42] was a 1kHz nanosecond Wedge Short Pulse Q- Switched DPSS Laser from Bright Solutions with a maximum power of 4 W or 4 mJ per pulse. 80 % of the 1064 nm output is then chopped (using a TTi chopper) to 500 Hz before pumping a PPLN-OPO crystal tuned to the O-D stretch at 3.76 µm. The pump laser was controlled by a Stanford delay generator. Pump power reaching the sample was 35 mW with a 75% absorbance for the 6 µm samples with pulse duration of approximately 1.5 ns. The large absorbance of the pump pulse is attributed to pumping the fundamental region of 3.76 µm. Typical T-jump experiments pump at the O-D overtone at 2 µm with only 10% absorbance of the pulse resulting in a more homogeneous heating of the sample[19]. The use of a small path length (6 µm) for the samples within this chapter was used to elevate some of the inhomogeneous heat distribution that is inherent with an absorbance of 75% of the pump pulse, though will still be present as the sample at the beginning of the cell will absorb most of the pulse energy. A publication on the T-jump laser used in this chapter has simulated the heat distributed present in the short path length samples[42] and summarises the decay timescales observed are similar to that of a lower absorbance and thicker cell. The 3.76 µm region was chosen as the pump pulse wavelength to minimise the risk of self-focusing in the sample that occurs with the higher energy 2 µm region. Pumping at the 2 µm region has also resulted in inhomogeneous heating of the sample[19], therefore heat homogeneity is an important consideration for any T-jump study.

The 10 kHz probe laser at RAL from a custom built Ti:Sapph oscillator and OPA, details of which can be found in Chapter 2.5, is tuned to 6 µm region, and time-resolved multiple probe spectroscopy (TRMPS) was utilised to measure the whole range from 1 ns to 2 ms. The set temperatures (T_0) are 20, 40, 44, 54, 60 °C with the use of a water bath to heat the cell.

4.4 Results

4.4.1 Calibration with TFA

To calibrate the temperature rise gained from the pump laser the change in absorbance of the TFA carbonyl stretch in D_2O as a function of temperature was used. First IR spectroscopy measurements were performed before progressing to T-jump experiments. TFA has a carbonyl group present which results in a carbonyl stretch vibration at a wavenumber of 1673 cm^{-1} (Figure 4-1). FTIR spectra were obtained by heating the sample 20 °C to 36 °C in 1 °C steps. The 1673 cm^{-1} peak decreased in intensity with a slight shift to higher wavenumbers with increasing temperature. Difference spectra were calculated by subtracting the 20 °C spectrum from all the other spectra (Figure 4-2 (a)). This was performed without further processing of the data including buffer subtraction or normalisation, to replicate the results that will be observed from the T-jump experiment. The difference spectra show a large negative peak at 1668 cm^{-1} and a small feature with a negative value at 1686 cm^{-1} with the intensity of each peak increasing with increasing temperature difference. A calibration plot was created by taking the difference in intensity between the negative 1668 cm^{-1} peak and the feature at 1686 cm^{-1} (Figure 4-2 (a), black arrow) and plotting against temperature. This resulted in a linear line with increasing temperature and a straight line was then fitted (Figure 4-2 (b)). The equation of the straight line can be seen in equation 4.1.

$$y = 0.020 - 0.00092x \qquad \text{(Equation 4.1)}$$

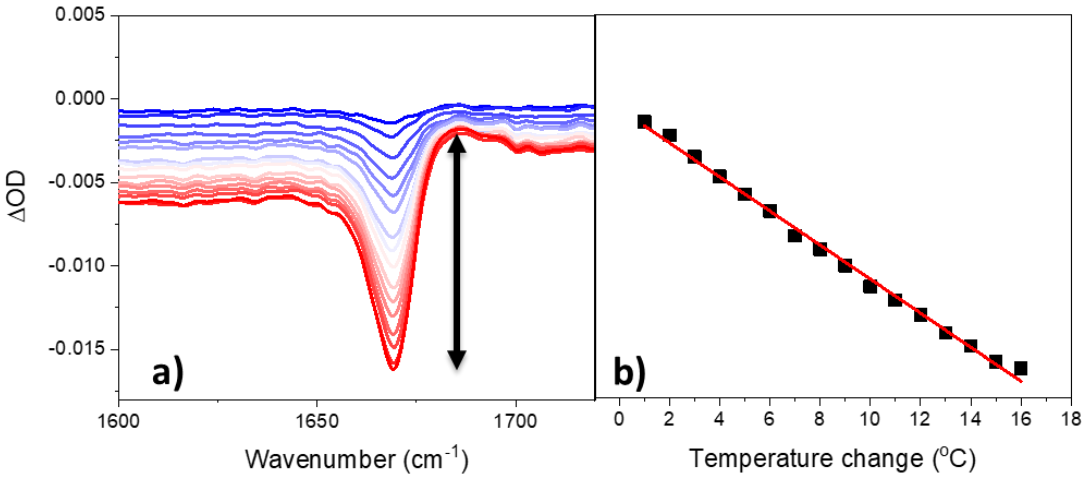


Figure 4-2. FTIR calibration of TFA, a) difference spectra of the carbonyl stretch from 20 °C (blue) to 36 °C (red) with the 20 °C spectrum subtracted. A negative peak present at 1668 cm^{-1} and a small feature present at 1686 cm^{-1} with the difference in size between these peaks plotted on b) an a straight line fitted (red).

TFA at T_0 = 20 °C was measured on the T-jump setup and resulted in a series of spectra from 1 ns to 1.9 ms (Figure 4-3). The spectra were very similar to the difference spectra produced by FTIR spectroscopy; there is a sharp negative peak at 1668 cm^{-1} and a slight positive peak at 1686 cm^{-1} (Figure 4-3(a)). These arise from the carbonyl peak of TFA decreasing in intensity and shifting to higher wavenumber with increased temperature. Plotting the 1668 cm^{-1} peak as a function of time shows that this starts large and negative before decaying to zero after 1ms as the sample cools (Figure 4-3 (b)). From this and the calibration plot, the maximum T-jump is concluded to occur just after the pump pulse. The timepoint with the maximum signal was used to calculate the T-jump. The magnitude between the negative peak and positive peak was 8.54 mOD. Using the straight line fitted from the FTIR calibration, the temperature rise from the T-jump experiment is calculated to be 9 °C (ΔT = 9 °C).

The decay of the 1668 cm^{-1} peak relaxed in a manner well-described by a stretched exponential function:

$$\Delta A = \alpha \cdot exp\left(-\frac{\tau_{pp}}{\tau_1}\right)^{\beta} \qquad \text{(Equation 4.2)}$$

with a lifetime of τ_1 ~47 µs and β parameter of 0.54. In the equation, ΔA indicates the measured change in absorbance; τ_{pp} the T-jump-probe delay time and α the amplitude of the signal. The stretched exponential character originates largely from a loss of heat from the sample due to conductivity of buffer and CaF_2 cells[43,44].

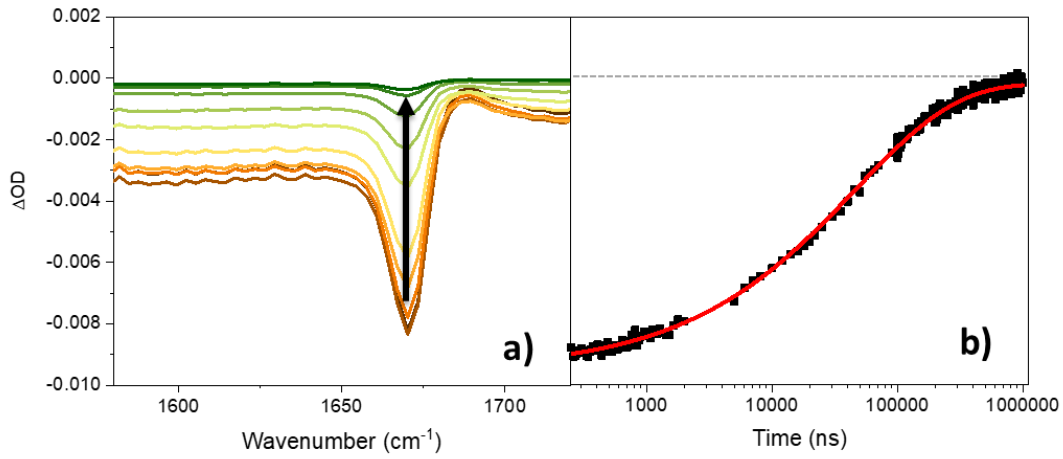


Figure 4-3. a) T-jump spectra of TFA at T_0 = 20 °C at selected waiting times from early times (250 ns, brown) to late times (900 µs, green), with arrow indicating increasing time, show a large negative peak at 1668 cm^{-1} and a small positive peak at 1680 cm^{-1} with both peaks decreasing in size as time increases, and b) showing the dynamics of the 1668 cm^{-1} absorbance change. Dashed line represents signal intensity of zero OD.

4.4.2 Spectral differences between *apo*-CaM and *holo*-CaM

Apo-CaM and *holo*-CaM were both measured at T_0 values of 20 °C, 40 °C, 44 °C, 54 °C, and 60 °C. These temperatures were chosen after analysis of the DSC curve for *apo*-CaM from chapter (3). T_0 = 20 °C, T_0 = 40 °C, and T_0 = 60 °C measurements for *apo*-CaM have been plotted in Figure 4-4 and are representative of all the data. Selected time points are shown only as to not overwhelm the figure. The figure shows the T-jump data in the lower panel and the corresponding FTIR-derived difference spectrum in the upper panel.

The FTIR difference spectrum for *apo*-CaM T_0 = 20 °C, which mimics a 9 °C step (Figure 4-4 (a)), shows a negative peak at approximately 1640 cm^{-1} and very little other spectral features. The T_0 = 40 °C FTIR difference spectrum (Figure 4-4 (b)) and the T_0 = 60 °C FTIR difference spectrum (Figure 4-4 (c)) both have a negative peak at approximately 1640 cm^{-1} which is more prominent than the peak from the T_0 = 20 °C spectrum. Both also have a positive peak at 1670 cm^{-1} (blue arrow) which is smaller in magnitude compared to the 1640 cm^{-1} peak. The T_0 = 40 °C 1670 cm^{-1} peak is slightly larger than the T_0 = 60 °C counterpart.

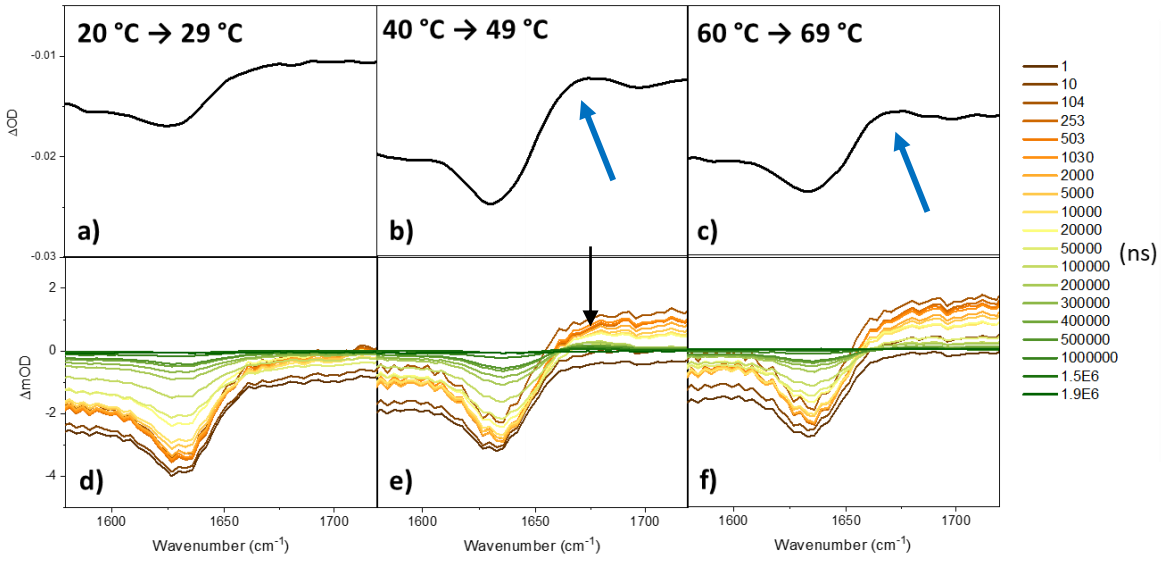


Figure 4-4. Selected apo-CaM temperature transitions for FTIR difference spectra (a-c) and selected T-jump spectra (d-f). The 20 to 29 °C temperature change is shown in a) and d) with both showing a negative peak at 1635 cm^{-1}. The 40 to 49 °C transition is shown in b) and e) with the negative peak shown in both and a distinct positive peak at 1660 cm^{-1} for the FTIR spectra – arrow in e) highlights positive peak is present in T-jump as well. The 60 to 69 °C transition in c) and f) shows similar trends as with the 40 to 49 °C transition.

The CaM signal is small and therefore the T-jump results will initially focus on establishing if a CaM signal is present and then the dynamics will be discussed later. T-jump results for *apo*-CaM are shown in Figure 4-4 (d-f) for the rises from T_0 = 20 °C, 40 °C, and 60 °C, respectively

where ΔT = 9 °C. As with the FTIR difference spectra, there is a negative peak at 1640 cm^{-1} present in all T-jump measurements and no other peak in the 20 °C measurement. However, for the 40 °C and 60 °C measurements there is a background shift at higher wavenumbers. This large peak is most likely a result of an experimental artefact and not a response of the protein itself. Taking into account the thermal effects of the solvent, although Figure 4-4 suggests that the major features are similar for the T-jump measurements for *apo*-CaM the T_0 = 40 °C data shows an additional peak identified by the black arrow at later times.

The results from the *holo*-CaM T_0 = 20 °C, T_0 = 40 °C and T_0 = 60 °C measurements are shown in Figure 4-5. The FTIR difference spectra (Figures 4-5 (a-c)) are all shown to be very similar. The spectra show a negative peak at 1640 cm^{-1} and no other spectral features, which is comparable to the results seen in the 20 to 29 °C FTIR difference spectrum for *apo*-CaM (Figure 4-5 (a)). From the FTIR difference spectra for 40 to 49 °C and 60 to 69 °C it can be seen that *holo*-CaM is different from *apo*-CaM as there is no positive peak at 1670 cm^{-1} detected when Ca^{2+} is present.

As with *apo*-CaM, the T-jump results for *holo*-CaM 20 to 29 °C, 40 to 49 °C and 60 to 69 °C show a negative peak at 1640 cm^{-1}. At the higher temperatures a solvent response is present at 1670 cm^{-1}.

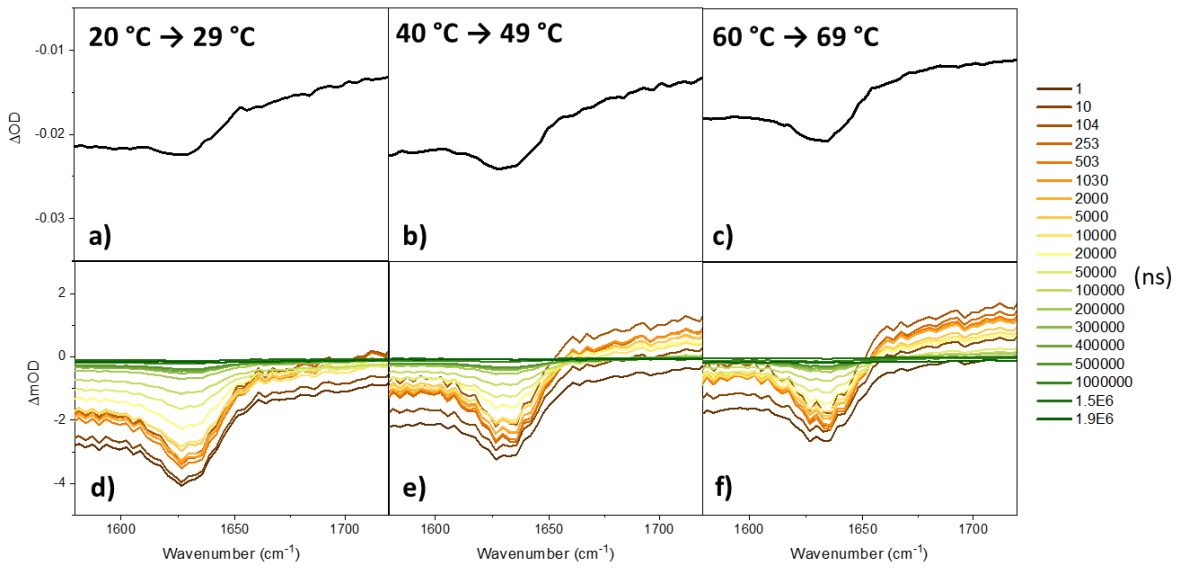


Figure 4-5. Selected holo-CaM temperature transitions for FTIR difference spectra (a-c) and selected T-jump spectra (d-f). All spectra show a strong negative peak at 1635 cm^{-1}, however, unlike apo-CaM there is no positive peak in higher temperature spectra. Each temperature spectrum has similar structure, for both FTIR and time spectra.

Differences between *apo*-CaM and *holo*-CaM are difficult to determine from Figures 4-4 and 4-5 as the large response from the solvent suppresses the effect the protein has on the results. However, by looking at the single time point of 30 µs, differences between the samples can be clearly seen (Figure 4-6). This time point has been selected for visual purposes as the broad baseline shift observed at high wavenumbers arising from experimental artefact somewhat masks the effect, but similar differences are still observed in the spectra at other times.

For *apo*-CaM the positive peak at 1670 cm^{-1} that is present in the FTIR difference spectra for T_0 = 40 °C and T_0 = 60 °C is now present in the 40 °C measurement (black arrow) and is in the 60 °C measurement but is small in intensity. At the same time, it is clear there is no positive peak in the T_0 = 20 °C T-jump measurement. The *holo*-CaM 30 µs time points highlight that there is very little difference between each of the measurements. There is no positive peak uncovered at 1670 cm^{-1} as is seen for *apo*-CaM.

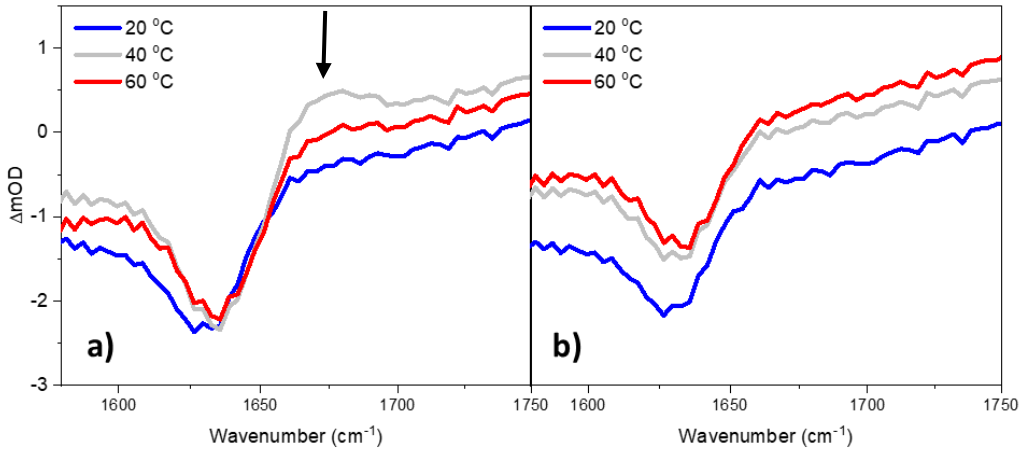


Figure 4-6. T-jump spectra from the 30 µs time point for both a) apo-CaM and b) holo-CaM. Only selected temperatures are shown (20 °C, 40 °C, and 60 °C). Arrow in a) highlights the additional positive peak at 1660 cm^{-1} that is prominent in the 40 °C spectrum.

To further extract dynamic differences between *apo*-CaM and *holo*-CaM, difference spectra are calculated. To normalise the data, each measurement is scaled to the 1600 cm^{-1} wavenumber position, a region with only a response from the solvent, for the 1000 ns time point. Each *holo*-CaM measurement is subtracted from the corresponding spectrum for *apo*-CaM. The resulting spectra are a double difference calculation as the T-jump spectra are the difference caused by the T-jump. The results of *holo*-CaM minus *apo*-CaM for the selected

measurements of T_0 = 20 °C, T_0 = 40 °C, and T_0 = 60 °C are seen in Figure 4-7. For the difference between the 20 °C measurements (Figure 4-7(a)) no distinct spectral features are present at any time point – highlighting that, for the T_0 = 20 °C temperature, there is very little difference between *apo*-CaM and *holo*-CaM. For 40 °C and 60 °C difference spectra (Figure 4-7 (b-c)) there is a negative peak at 1640 cm^{-1} (red arrow) and a positive peak at 1670 cm^{-1} (black arrow). The negative peak with the frequency of 1640 cm^{-1} suggests that there is a reduction in amide I structure for *apo*-CaM compared to *holo*-CaM that corresponds to this frequency, and subsequently, the positive peak suggests there is an increase in amide I structure for *apo*-CaM compared to *holo*-CaM that corresponds to this 1670 cm^{-1} frequency. The next section will endeavour to assign what these structural changes are.

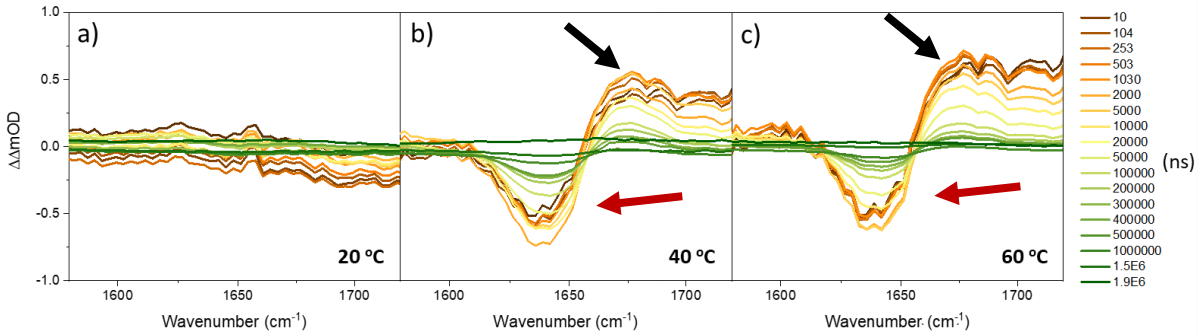


Figure 4-7. Apo-CaM minus holo-CaM difference T-jump spectra from early times (orange) to late times (green) for selected initial temperatures. a) is the 20 °C starting temperature which shows very little difference between apo-CaM and holo-CaM, b) 40° C which has two distinct peaks, a negative peak at 1635 cm^{-1} and a positive peak at 1660 cm^{-1}, and c) with starting temperature of 60 °C has a similar two peaks as 40 °C spectra.

4.5 Discussion

4.5.1 Spectral assignment

From chapter (3) it was observed that heating of CaM resulted in a loss of spectral features at 1640 cm^{-1} and a gain of spectral features at 1660 cm^{-1}. These were attributed to α-helix and random coil features, respectively. CaM is stable around 20 °C, and transitions in the structure are not observed until around 46 °C. This is when the C-domain for *apo*-CaM "melts"; the N-domain "melt" transition is at the higher temperature of 58 °C. Therefore, at temperatures below 46 °C both *apo*-CaM and *holo*-CaM should have similar behaviour upon heating. *Holo*-CaM, in comparison, does not undergo a transition until a higher temperature of around 96 °C. *Holo*-CaM was used in chapter (3) as a reference for thermal heating without a transition occurring, and subtracting from *apo*-CaM resulted in accurate changes in α-helix structure for the heating of *apo*-CaM.

Figure 4-7(a) shows there is no difference between *apo*-CaM and *holo*-CaM for the 20 °C T-jump measurements. Following on from this, differences between *apo*-CaM and *holo*-CaM become evident with the 40 °C measurement and the behaviour upon heating is different between the samples, with the appearance of the loss of α-helix (1640 cm^{-1}) and gain of random coil (1660 cm^{-1}) suggesting a transition in secondary structure that occurs in *apo*-CaM and not *holo*-CaM. As the set temperature is 40 °C, the temperature after the T-jump pulse is 49 °C; this is now above the temperature for the C-domain transition for *apo*-CaM. As *holo*-CaM is significantly thermally stable and does not have a transition below 90 °C, it is used as a control in the difference spectra and eliminates responses from the solvent and the thermal effects on hydrogen bonding which do not result in a structural change. Therefore, the response from the 40 °C to 49 °C difference spectra can be assigned to be a result of the C-domain "melt" of the unfolding of the α-helix to random coil.

The T_0 = 44 °C difference spectrum again shows the same spectral features as the T_0 = 40 °C difference spectrum. As this is still within the C-domain transition window it is expected to have similar results. The T_0 = 54 °C measurements difference spectra show similar positive and negative peaks indicating an unfolding transition from α-helix to random coil. This temperature range is now within the N-domain transition for *apo*-CaM, so expect similar results as for the C-domain transition temperatures.

The T_0 = 60 °C measurement is now past both the C- and N-domain transitions, and less of a change is observed in the difference spectrum compared to that of the T_0 = 40 °C, T_0 = 44 °C, and T_0 = 54 °C measurements. However, as this is a comparison between post-transition point *apo*-CaM and pre-transition point *holo*-CaM, there are expected spectral differences.

4.5.2 Peak-to-peak magnitudes

To give a direct comparison of the difference spectra for all temperatures, the 100 μs time point for all has been plotted in Figure 4-8 (a). This time point has been selected as it is far enough into the heating of the sample that no artefacts of heating will interfere with the sample but still shows a protein unfolding response, even though the temperature increase has dropped to ~35% of the initial value. The spectra were then normalised where the largest peak-to-peak magnitude corresponds to an intensity value of one. Therefore, the signal size in Figure 4-8 is not comparable to Figure 4-7. A clear difference can be seen between the 20 °C measurement, where the signal is close to the baseline, and all other measurements, where distinct negative, α-helix peaks and positive, random coil peaks are present.

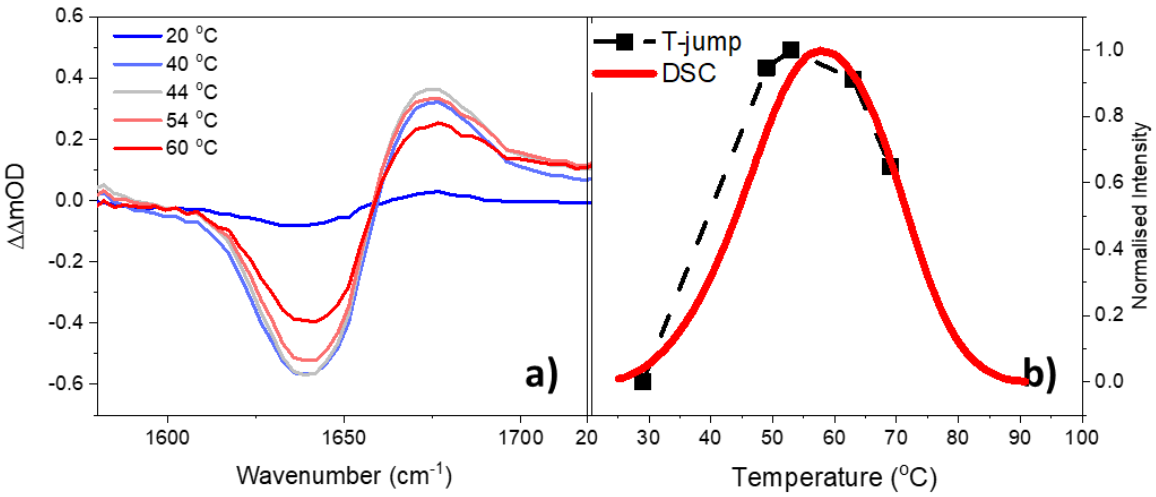


Figure 4-8. a) Difference spectra, apo-CaM minus holo-CaM, for the 100 µs time point for all temperatures measured with a negative peak at 1635 cm^{-1} and a positive peak at 1660 cm^{-1}, and b) the peak-to-peak magnitude trend (black) from 20 °C to 60 °C for the peaks shown in a) and the DSC data for apo-CaM (red) which been selected to only show the corresponding temperatures as the T-jump data.

The peak-to-peak magnitudes (Figure 4-8 (b)) for all temperatures for the 100 µs second time point have been calculated by simply subtracting the intensity of the negative 1640 cm^{-1} peak from the intensity of the positive 1670 cm^{-1} peak. Plotting the magnitude for each temperature produces a curve like the DSC curve for *apo*-CaM that was described in chapter (3) and reproduced here in Figure 4-8 (b) for clarity, along with the T-jump peak-to-peak magnitudes. The temperatures for the T-jump trend have been adjusted to reflect the maximum temperature reached for each measurement, resulting in good overlap of T-jump and DSC.

The maximum signal appears in the T_0 = 44 °C measurement. The intensity has started to fall with the T_0 = 60 °C measurement which follows the trend of DSC. Therefore, the DSC results have confirmed the spectral analysis allowing for the dynamics of the C-domain melting transition of *apo*-CaM to be determined by T-jump spectroscopy.

4.5.3 Dynamics of melting

The laser system allowed for measurements to 1.9 ms and it is shown that the temperature has returned to the initial starting temperature within this time scale as the signal has returned to zero at ~1.6 ms (Figure 4-9).

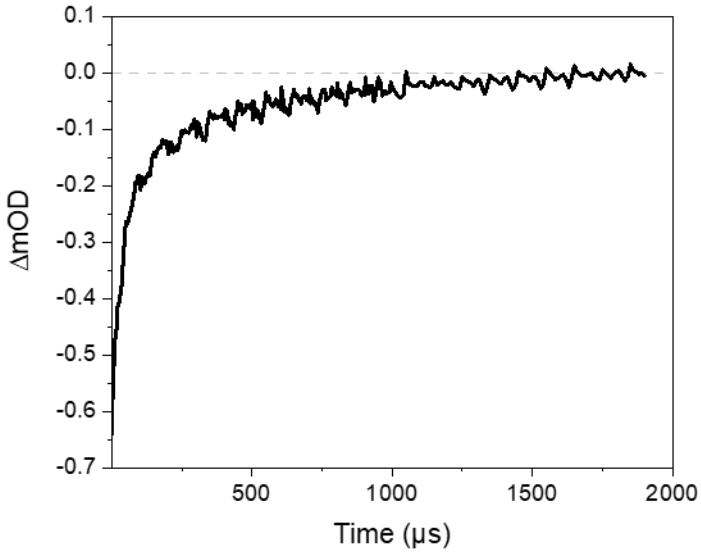


Figure 4-9. T-jump spectrum for apo-CaM with T0 = 20 °C. Time axis displayed with a linear scale showing the signal has returned to zero at ~1.6 ms, before the next pump pulse arrives at 2 ms.

Plotting of dynamics begins at 250 ns as prior to 250 ns there are oscillations in the data that are the result of electronic noise from the pump laser at early times. It is known to be from the pump laser as a blank measurement was collected with the probe laser blocked and no sample included, and the oscillations were still present in the measurement.

The dynamics of the baseline (black line) have been plotted from the *apo*-CaM measurements at 1600 cm^{-1} where no contribution from the protein is present. Also plotted is the 1641 cm^{-1} α-helix peak from *apo*-CaM (green), and the same 1641 cm^{-1} peak from *holo*-CaM (red). All times are plotted using a logarithmic scale (Figure 4-10).

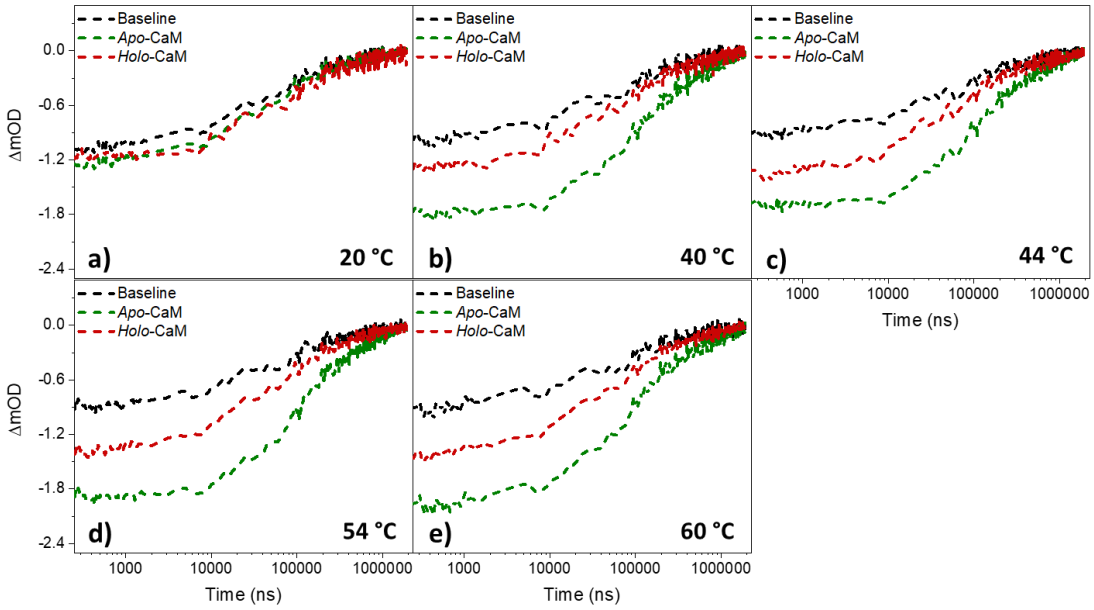


Figure 4-10. Dynamic plots of time dependent absorbance change for baseline (black) from the apo-CaM measurements at 1600 cm^{-1} where no contribution from protein is, apo-CaM α-helix peak (green), and holo-CaM α-helix peak (red) for all temperatures 20 °C to 60 °C (a-e) from 250 ns to 1.9 ms.

The 20 °C measurement (Figure 4-10 (a)) shows that the amide I peak for *apo*-CaM and *holo*-CaM decay at the same rate as the buffer, with a maximum signal change of approximately 1.2 mOD happening immediately for all dynamics. These plots decay back to zero and fitting stretched exponentials to all signals gives a time of around 50 µs. This lack of difference between buffer, *apo*-CaM and *holo*-CaM is not unexpected as discussions above state that CaM is stable over this temperature range for both *apo*-CaM and *holo*-CaM.

The 40 °C measurement (Figure 4-10 (b)) shows deviations for the *apo*-CaM and *holo*-CaM plot from the baseline. Although *holo*-CaM is a stable protein with no changes in the secondary structure below 80 °C and therefore no changes from the buffer in terms of dynamics would be expected, the deviation from the baseline can be explained from what was learnt from chapter (3). It was concluded that as *holo*-CaM was thermally perturbed it resulted in changes in the H-bonding that did not result in a structural change. Therefore, the deviation from the buffer is a result of this thermal effect and this is confirmed as the fitting of stretched exponentials to the *holo*-CaM give the same time scales as the buffer of around 50 µs. The difference between the buffer and *holo*-CaM results is the amplitude of the signal at the beginning of the plot. As a result of the noise, fitting would be inaccurate prior to 250 ns. H-bonding changes from thermal effects occur on very fast time scales of nanoseconds and therefore the initial change due to H-bonding cannot be seen as it occurs quicker than

can reliably be measured. By scaling the plots (Figure 4-11) it can visually be seen that *holo*-CaM is behaving in the same manner as the buffer.

The *apo*-CaM plot for the 40 °C measurement has an even larger deviation from the baseline compared to *holo*-CaM and the fitting results in different timescales - around 200 µs. This trend of deviations from the buffer continue for *apo*-CaM and *holo*-CaM for the subsequent higher temperature measurements which can be seen in Figure 4-10 (b-e).

The differences between *apo*-CaM and *holo*-CaM can further be accentuated by scaling (Figure 4-11). While *holo*-CaM and the baseline provide a good overlap, *apo*-CaM deviates from approximately 10 µs with a slower temperature relaxation rate. The maximum difference occurs in the 44 °C measurement (Figure 4-11 (c)), highlighted by the black arrow. The 60 °C measurement sees the difference decreasing again as both the C-domain and N-domain transitions have occurred for *apo*-CaM.

The implication of these fitting results is that for the *apo*-CaM where measurements are around the DSC transitions of 46 °C and 58 °C there is an additional dynamic component that is not observed in the *holo*-CaM data. This additional component appears to lengthen the decay timescale.

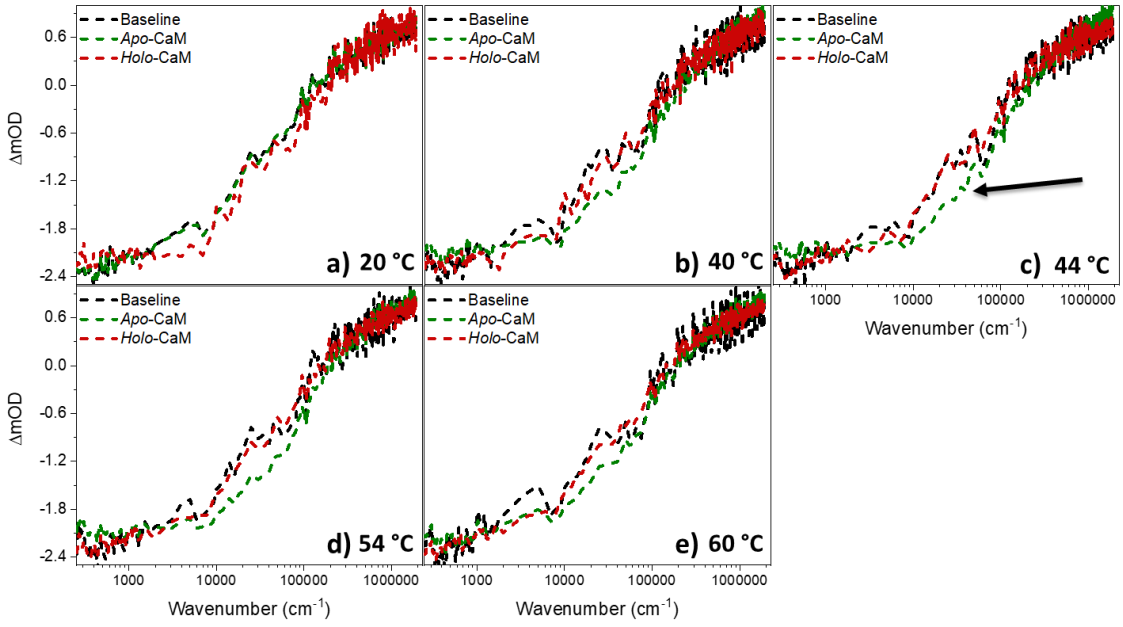


Figure 4-11. Normalised dynamic plots for baseline (black), apo-CaM α-helix peak (green), and holo-CaM α-helix peak (red) for all temperatures 20 °C to 60 °C (a-e) from 250 ns to 1.9 ms. Arrow in c) highlights that apo-CaM dynamics are different than holo-CaM and baseline, which are the same upon normalising.

As well as looking at the 1640 cm^{-1} α-helix peak dynamics, the peak to peak magnitude dynamics between the α-helix peak (1640 cm^{-1}) and the random coil peak (1660 cm^{-1}) have been investigated. This has been determined by simply subtracting the random coil dynamics at 1660 cm^{-1} from the α-helix dynamics at 1640 cm^{-1} for all measurements. Fitting stretched exponentials to the peak to peak magnitudes gives the same cooling rates of approximately 50 µs for *holo*-CaM. The resulting data is less noisy as random noise within the measurement has been cancelled out by the subtraction.

The comparison of the dynamics of *apo*-CaM and *holo*-CaM across all the temperatures measured can be seen in Figure 4-12. For the *apo*-CaM peak to peak dynamics (Figure 4-12 (a)), the 20 °C measurement (blue) is easily seen to be different from the rest with a much smaller overall signal size (from around -0.6 mOD compared to around -2 mOD for the other temperatures). The decay constants (Figure 4-12 (c)) show the 20 °C has a shorter timescale (70 µs) than the other temperatures which have decay constants of 130 µs to 200 µs.

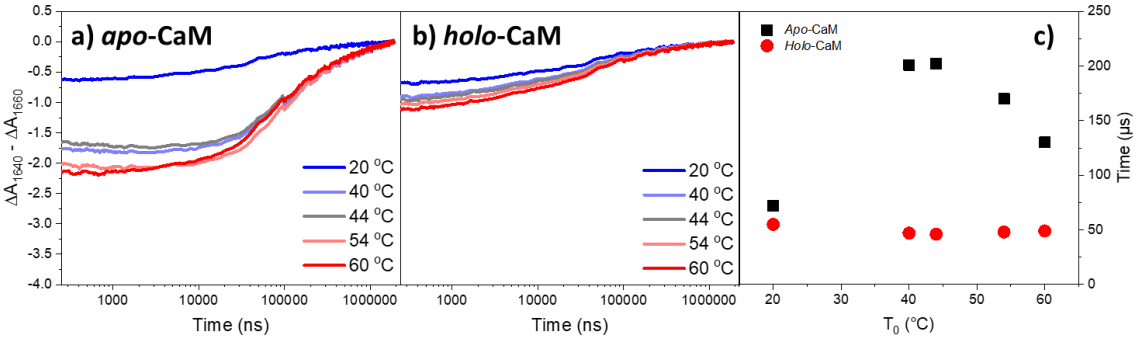


Figure 4-12. Plots of time dependent absorbance change for signals at 1640 cm^{-1} minus signals at 1660 cm^{-1} for a) apo-CaM and b) holo-CaM for 20 °C (blue) to 60 °C (red) T_0. c) decay constants for apo-CaM (black) and holo-CaM (red) from fitting stretched exponentials to the plots in a) and b). Decay constants for holo-CaM are shorter times compared to apo-CaM.

Table 4-1. Stretched exponential β terms for each of the apo- and holo-CaM dynamics fittings in Figure 4-12.

Temperature (°C)	β *apo*-CaM	β *holo*-CaM
20	0.49	0.45
40	0.72	0.42
44	0.74	0.42
54	0.75	0.42
60	0.65	0.43

For *holo*-CaM peak to peak dynamics (Figure 4-12 (b)) there is less distinction between the temperatures – they only have increasing signal size from 20 °C (blue at -0.6 mOD) to 60 °C (red at -1.2 mOD) which is the result of the fluctuation in H-bonding due to thermal effects.

This thermal effect on the H-bonding is also present in the *apo*-CaM data and in chapter (3) it was discussed that to eliminate it from *apo*-CaM the *holo*-CaM spectra would be subtracted from it. As with chapter (3) the difference between *apo*-CaM and *holo*-CaM reveals the secondary structure changes that are obscured by the heating of the H-bonding. The dynamics of this difference (Figure 4-13 (a)) shows that for the 20 °C measurement (blue) there is no difference between *apo*-CaM and *holo*-CaM across all times. From 40 °C and above there is a significantly revealing profile for the difference dynamics. There is a growth (which appears negative due to the dominating α-helix peak in the peak to peak magnitudes) which peaks around 20 µs at -1.2 mOD before decaying back towards zero before the end of the 1.9 ms measurement. No fitting was performed on the 20 °C plot but stretched bi-exponential fitting was performed on all other temperatures which resulted in a short decay constant for the growth portion, and a longer decay constant for the decay portion of the plots (Figures 4-13 (b) and 4-14, Table 4-2).

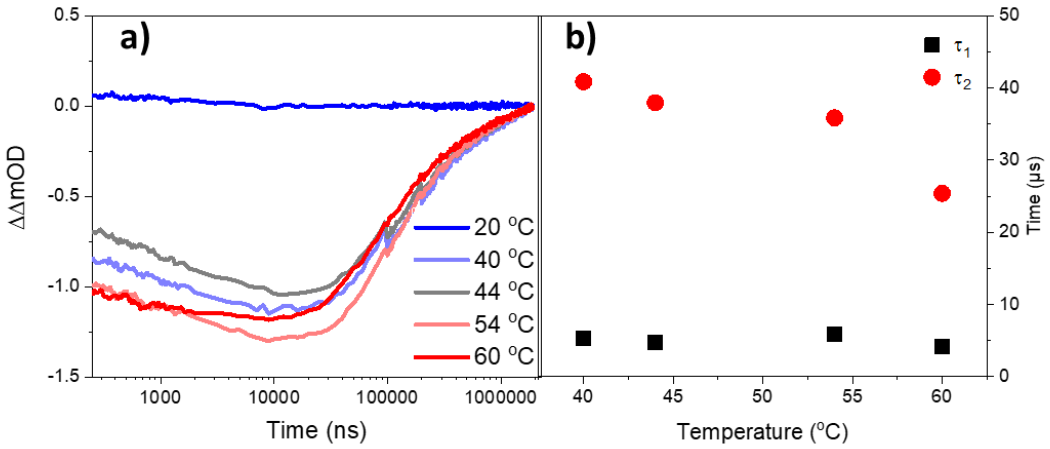


Figure 4-13. a) Time dependent absorbance change for difference spectra (apo-CaM minus holo-CaM) from the 1640 cm^{-1} peak for 20 °C (blue) to 60 °C (red), and b) the resulting decay constants, t_1 (black) and t_2 (red), from stretched bi-exponential fitting of the dynamics from a). The 20 °C time dependent absorbance change was approximately zero across the complete time range, therefore no fitting occurred for this temperature plot.

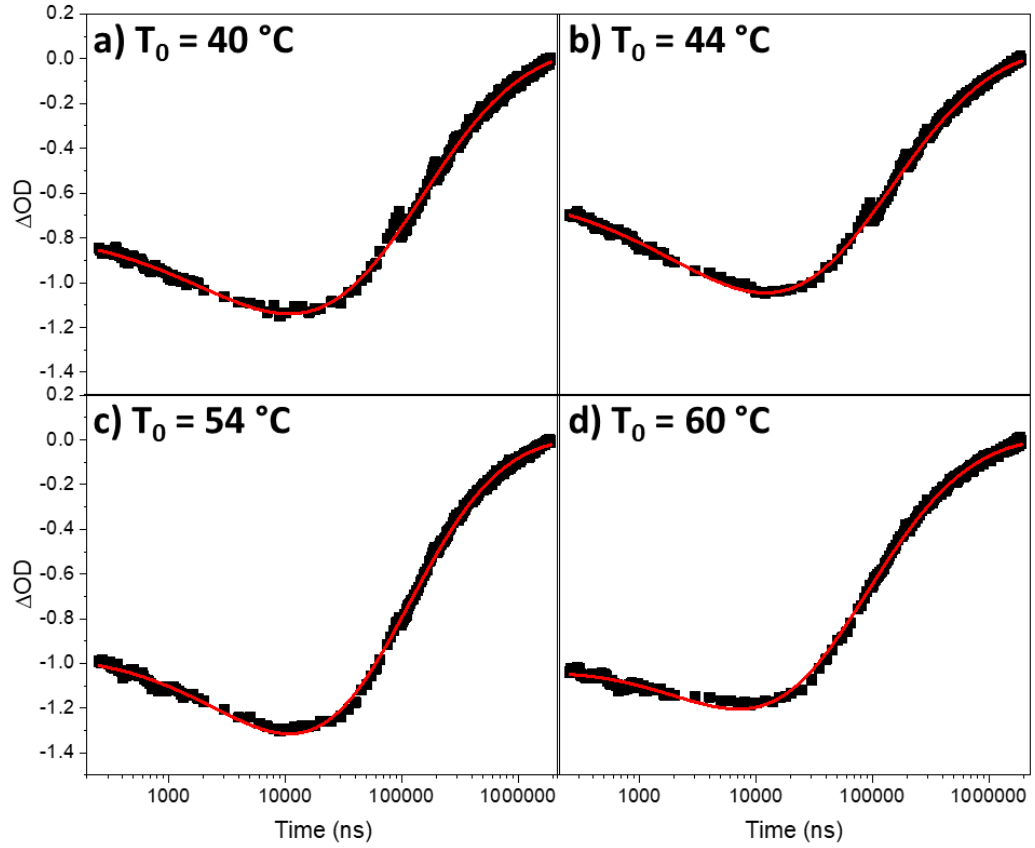


Figure 4-14. Stretched bi-exponential fittings to the apo- minus holo-CaM dynamics for a) T_0 = 40 °C, b) T_0 = 44 °C, c) T_0 = 54 °C, and d) T_0 = 60 °C.

Table 4-2. Stretched bi-exponential β terms for each of the apo- minus holo-CaM dynamics fittings in Figure 4-14.

Temperature (°C)	β1	β2
40	0.46	0.36
44	0.46	0.35
54	0.53	0.39
60	0.53	0.35

The growth was on the timescale of ~5 μs which fits with previous literature of helical peptide melting and hairpin unfolding. There is also no clear distinction in the growth timescales of the C-domain unfolding and the N-domain unfolding of CaM. Reasons for this could be the unclear distinction between the two domain transitions in the DSC which leads to an overlap

of the unfolding which results in contributions of C- and N-domain unfolding for all measurements with a growth profile. Another reason could be that unfolding kinetics are usually activated, and as the two unfolding events occur at different temperatures, N-domain at 58 °C and C-domain at 46 °C, they could have the same rate but the N-domain being intrinsically slower.

Another factor is that this is believed to be only a partial unfolding of the α-helix. The maximum signal size for the difference dynamics is roughly -1.2 mOD and is seen at around 20 μs, however, from the buffer dynamics in Figure 4-10 it is shown that the signal has already begun to decrease back towards zero. This suggests that the temperature has started decreasing back to the original set temperature before the maximum unfolding transition has occurred, resulting in only a partial unfolding of the α-helix. The buffer signal has decreased by about 45% at 20 μs and it can be assumed the temperature has also dropped by about 45%. This suggests that there is now only a 5 °C rise from the set temperature and this is still falling.

A compromise had to be made to the experimental setup which affected the heat dissipation out of the sample. The path length could have been increased from 6 μm to ensure the temperature remains in the sample for longer giving more opportunity for a larger degree of unfolding to occur. This is because the heat conductivity of CaF_2 is more than ten times higher than that of D_2O[44], therefore the heat dissipates quicker the closer it is to the CaF_2 windows. However, if there was a longer pathlength then the sample would not have had enough time to cool back to the original temperature before the next pump pulse. This would result in heat build-up in the sample. An alternative strategy for this would be to chop the pump pulse, reducing the repetition rate further, which allows for measurements past 2 ms, however, the signal to noise would decrease and to compensate the acquisition time would have to increase to average for longer. This would be detrimental for developing the setup into a 2D-IR T-jump method as the acquisition time would drastically increase.

Even if the above setup changes were to occur, full unfolding of CaM would not be observed. This is confirmed in chapter (3) which has shown that heating the sample from 20 °C to 80 °C results in only a 13 to 15% α-helix secondary structure decrease out of over 50% α-helix content, therefore not full unfolding even over the large temperature range. Consequently, the best compromise in this case is the current setup of the system.

The ~5 μs timescales observed in the T-jump melting experiments are comparable with fast steps reported in the melting of ubiquitin[40], which were assigned to initial 'burst phase' or downhill melting of the protein following T-jump perturbation of the potential energy surface for some of the proteins. In those experiments, slower timescale dynamics on hundreds of microsecond timescales were assigned to activated barrier crossing caused by proteins re-equilibrating to the new temperature. In this current study, although the early processes are visible, sample cooling competes with any activated barrier crossing such that the results are dominated by the fast initial phase of the unfolding. This is supported by the lack of a T_0 dependence of the 5 μs unfolding timescale.

The longer decay timescale is around 40 μs which fits with the decay timescale which is seen for *holo*-CaM and the buffer. The large decay constant suggests this is the re-association of the α-helix as the protein cools down to the starting temperature.

Comparing with the one previous T-jump study of *apo*-CaM[41], where melting processes with timescales of ~100 μs were observed, would suggest a conflict in the results. However, based on the information above, these are most likely to be assignable to the activated barrier crossing processes such as those reported for ubiquitin[40]. Thus, by probing the fastest steps in CaM domain melting, the new results from this chapter complement this prior study to produce an overall picture of CaM domain melting. Taken jointly these two studies also reinforce the need to observe protein dynamics over as wide a range of timescales as possible to gain a complete molecular picture.

4.6 Conclusion and further work

Development of a new T-jump TRIR laser setup at RAL was performed with testing and calibration of the temperature rise by the sample TFA. Through this calibration the T-jump was determined to be 9 °C. After calibration of the system, CaM was measured in the first TRIR T-jump experiment of CaM. Both *apo*-CaM and *holo*-CaM systems were measured at a series of initial set temperatures, and, as with chapter (3), there was a large benefit of having a reference sample (*holo*-CaM in both cases) as a control to measure the thermally-induced changes in H-bonding without a secondary structure transition being present. This allowed the extraction of unfolding timescales of α-helix to be separated out from non-transition changes for *apo*-CaM, which gave a decay constant of 5 μs for thermal destabilisation of the α-helix. These results are complementary to previous CaM T-jump results with long timescales of hundreds of microseconds associated with activated barrier crossing processes.

These fast timescales were only observed for T_0 values of 40 °C to 60 °C which correlated to the C- and N- terminal domain transitions which were determined in chapter (3) from DSC as being 46 °C and 57 °C, respectively. No short, unfolding timescale was observed for the room temperature measurement of *apo*-CaM as expected due to no secondary structure transition being present at this temperature. This T-jump experiment is complementary to DSC as both can detect thermally induced transitions in secondary structure, with the T-jump experiment having the added benefit of detecting unfolding dynamics.

The measurements of the well characterised protein have provided positive evidence that accurate measurements of fast initial protein unfolding timescales from proteins with unknown characteristics is possible using this current T-jump TRIR laser.

This setup is different from previous infrared spectroscopy T-jump lasers as it employs a pump laser that is chopped at 500 Hz. This allows for typically quicker acquisition of data and leads onto further development of the setup to become a 2D-IR T-jump laser. The ULTRA laser where this T-jump setup operates has previously measured non-transient 2D-IR spectra at the probe region of interest – the amide I 6 μm region. However, these measurements were collected using an interferometer setup which has limitations. Over long periods of time data can undergo phase drift which would distort the spectrum. This would be a problem as T-jump 2D-IR measurements involve long acquisition times – hence why it was imperative that the acquisition time of the measurements in this chapter were kept at a minimum. The implementation of a pulse shaper instead of an interferometer would increase the stability of the laser over time and is a requirement for the development of a 2D-IR T-jump setup at RAL. Chapter (5) begins this implementation and has a comparison of 2D-IR measurements using an interferometer setup and a pulse shaper setup.

4.7 References

(1) Johnson, K. A. Advances in Transient-State Kinetics. *Curr. Opin. Biotechnol.* **1998**, *9*, 87–89.

(2) Roder, H.; Maki, K.; Cheng, H. Early Events in Protein Folding Explored by Rapid Mixing Methods. *Chem. Rev.* **2006**, *106* (5), 1836–1861.

(3) Jennings, P. A.; Wright, P. E. Formation of a Molten Globule Intermediate Early in the Kinetic Folding Pathway of Apomyoglobin. **1993**, *262*, 892–897.

(4) Roder, H.; Colnt, W. Kinetic Role of Early Intermediates in Protein Folding. *Curr.*

Opin. Struct. Biol. **1997**, *7*, 15–28.

(5) Matthews, C. R. Pathways of Protein Folding. *Annu. Rev. Biochem.* **1993**, *62*, 653–683.

(6) Telford, J. R.; Wittung-Stafshede, P.; Winkler, J. R.; Gray, H. B. Protein Folding Triggered by Photoinduced Electron Transfer. *Abstr. Pap. Am. Chem. Soc.* **1998**, *216*, 3–5.

(7) Chen, J.; Rempel, D. L.; Gross, M. L. T-Jump and Fast Photochemical Oxidation Probe Sub Millisec Protein Folding. *J Am Chem Soc .* **2011**, *132* (44), 15502–15504.

(8) Hansen, K. C.; Rock, R. S.; Larsen, R. W.; Chan, S. I. A Method for Photoinitating Protein Folding in a Nondenaturing Environment. *J. Am. Chem. Soc.* **2000**, *122* (46), 11567–11568.

(9) Gooding, E. A.; Kholodenko, Y.; DeGrado, W. F.; Volk, M.; Lu, H. S. M.; Hochstrasser, R. M. Peptide Conformational Dynamics and Vibrational Stark Effects Following Photoinitiated Disulfide Cleavage. *J. Phys. Chem. B* **2002**, *101* (42), 8607–8616.

(10) Kubelka, J. Time-Resolved Methods in Biophysics . 9 . Laser Temperature-Jump Methods for Investigating Biomolecular Dynamics. *Photochem. Photobiol. Sci.* **2009**, *8*, 499–512.

(11) Volk, M.; Ramajo, A. P. Time-Resolved Isotope-Edited IR Spectroscopy – Detailed Insight into the Fast Folding Dynamics of α -Helical Peptides. *Cent. Laser Facil. Annu. Rep.* **2005**, *1*, 173–175.

(12) Phillips, C. M.; Mizutani, Y.; Hochstrasser, R. M. Ultrafast Thermally Induced Unfolding of RNase A. **1995**, *92*, 7292–7296.

(13) Beitz, J. V.; Flynn, G. W.; Turner, D. H.; Sutin, N. The Stimulated Raman Effect. A New Source of Laser Temperature-Jump Heating. *J. Am. Chem. Soc.* **1970**, *92* (13), 4130–4132.

(14) Rigler, R.; Rabi, C.-R.; Jovin, T. M. A Temperature-Jump Apparatus for Fluorescence Measurements. *Rev. Sci. Instrum.* **1974**, *45*, 580–588.

(15) Volk, M.; Mendonça, L.; Solinas, X.; Sharma, S.; Hache, F.; Khuc, M.-T. Measurement of Circular Dichroism Dynamics in a Nanosecond Temperature-Jump Experiment.

Rev. Sci. Instrum. **2011**, *82* (5), 054302.

(16) Williams, S.; Causgrove, T. P.; Gilmanshin, R.; Fang, K. S.; Callender, R. H.; Woodruff, W. H.; Dyer, R. B. Fast Events in Protein Folding : Helix Melting and Formation in a Small Peptide. *Biochemistry* **1996**, *2*, 691–697.

(17) Petty, S. A.; Volk, M. Fast Folding Dynamics of an a -Helical Peptide with Bulky Side Chains. *Phys. Chem. Chem. Phys.* **2004**, *6*, 1022–1030.

(18) Krejtschi, C.; Huang, R.; Keiderling, T. A.; Hauser, K. Time-Resolved Temperature-Jump Infrared Spectroscopy of Peptides with Well-Defined Secondary Structure: A Trpzip Beta-Hairpin Variant as an Example. *Vib. Spectrosc.* **2008**, *48*, 1–7.

(19) Chung, H. S.; Khalil, M.; Smith, A. W.; Tokmakoff, A. Transient Two-Dimensional IR Spectrometer for Probing Nanosecond Temperature-Jump Kinetics. *Rev. Sci. Instrum.* **2007**, *78* (6), 063101/1-063101/10.

(20) Jones, K. C.; Ganim, Z.; Peng, C. S.; Tokmakoff, A. Transient Two-Dimensional Spectroscopy with Linear Absorption Corrections Applied to Temperature-Jump Two-Dimensional Infrared. **2012**, *29* (1), 118–129.

(21) Murphy, K. P.; Hofrichter, J. Laser Temperature-Jump Methods for Studying Folding Dynamics. *Protein Struct. Stability, Fold.* **2003**, *168*, 159–191.

(22) Zhang, X. X.; Jones, K. C.; Fitzpatrick, A.; Peng, C. S.; Feng, C. J.; Baiz, C. R.; Tokmakoff, A. Studying Protein-Protein Binding through T-Jump Induced Dissociation: Transient 2D IR Spectroscopy of Insulin Dimer. *J. Phys. Chem. B* **2016**, *120* (23), 5134–5145.

(23) Jones, K. C.; Peng, C. S.; Tokmakoff, A. Folding of a Heterogeneous β -Hairpin Peptide from Temperature-Jump 2D IR Spectroscopy. *Proc Natl Acad Sci U S A* **2013**, *110*, 2828–2833.

(24) Williams, S.; Causgrove, T. P.; Gilmanshin, R.; Fang, K. S.; Callender, R. H.; Woodruff, W. H.; Dyer, R. B. Fast Events in Protein Folding: Helix Melting and Formation in a Small Peptide. *Biochemistry* **1996**, *35* (3), 691–697.

(25) Thompson, P. A.; Eaton, W. A.; Hofrichter, J. Laser Temperature Jump Study of the Helix⇆coil Kinetics of an Alanine Peptide Interpreted with a “kinetic Zipper” Model.

Biochemistry **1997**, *36* (30), 9200–9210.

(26) Gruebele, M.; Sabelko, J.; Ballew, R.; Ervin, J. Laser Temperature Jump Induced Protein Refolding. **1998**, *31* (11), 699–707.

(27) Lednev, I. K.; Karnoup, A. S.; Sparrow, M. C.; Asher, S. A. Alpha-Helix Peptide Folding and Unfolding Activation Barriers: A Nanosecond UV Resonance Raman Study. *J Am Chem Soc* . **1999**, *121*, 8074–8086.

(28) Popp, A.; Scheerer, D.; Chi, H.; Keiderling, T. A.; Hauser, K. Site-Specific Dynamics of Beta-Sheet Peptides with D Pro – Gly Turns Probed by Laser-Excited Temperature-Jump Infrared Spectroscopy. *ChemPhysChem* **2016**, *17*, 1273–1280.

(29) Hamm, P.; Helbing, J.; Bredenbeck, J.; Kumita, J. R.; Woolley, G. A. Helix Formation in a Photoswitchable Peptide Tracked from Picoseconds to Microseconds by Time-Resolved IR Spectroscopy. *Proc. Natl. Acad. Sci.* **2005**, *102* (7), 2379–2384.

(30) Werner, J. H.; Dyer, R. B.; Fesinmeyer, R. M.; Andersen, N. H. Dynamics of the Primary Processes of Protein Folding: Helix Nucleation. *J. Phys. Chem. B* **2002**, *106* (2), 487–494.

(31) Bai, Y.; Nussinov, R. *Protein Folding Protocols*; 2007.

(32) Wang, T.; Zhu, Y.; Getahun, Z.; Du, D.; Haung, C.-Y.; DeGrado, W. F.; Gai, F. Length Dependent Helix-Coil Transition Kinetics of Nine Alanine-Based Peptides. *J. Phys. Chem. B* **2004**, *108* (39), 3831–3840.

(33) Mukherjee, S.; Chowdhury, P.; Bunagan, M. R.; Gai, F. Folding Kinetics of a Naturally Occurring Helical Peptide: Implication of the Folding Speed Limit of Helical Proteins. *J. Phys. Chem. B* **2008**, *112* (30), 9146–9150.

(34) Ramajo, A. P.; Petty, S. A.; Starzyk, A.; Decatur, S. M.; Volk, M. The Alpha-Helix Folds More Rapidly at the C-Terminus Than at the N-Terminus. *J Am Chem Soc* . **2005**, *127*, 13784–13785.

(35) Gooding, E. A.; Sharma, S.; Petty, S. A.; Fouts, E. A.; Palmer, C. J.; Nolan, B. E.; Volk, M. PH-Dependent Helix Folding Dynamics of Poly-Glutamic Acid. *Chem. Phys.* **2013**, *422*, 115–123.

(36) Ramajo, A. P.; Petty, S. A.; Volk, M. Fast Folding Dynamics of Alpha-Helical Peptides –

Effect of Solvent Additives and PH. *Chem. Phys.* **2006**, *323*, 11–20.

(37) Santiveri, C. M.; Santoro, J.; Rico, M.; Jiménez, M. A. Thermodynamic Analysis of β-Hairpin-Forming Peptides from the Thermal Dependence of 1H NMR Chemical Shifts. *J. Am. Chem. Soc.* **2002**, *124* (50), 14903–14909.

(38) Causgrove, T. P.; Maness, S. J.; Dyer, R. B.; Gibbs, A. C.; Franzen, S. Nanosecond Temperature Jump Relaxation Dynamics of Cyclic β-Hairpin Peptides. *Biophys. J.* **2009**, *84* (6), 3874–3882.

(39) Muñoz, V.; Eaton, W.; Thompson, P. a; Hofrichter, J. Folding Dynamics and Mechanism of Bold Beta-Hairpin Formation. *Nature* **1997**, *390* (6656), 196–199.

(40) Chung, H. S.; Tokmakoff, A. Temperature-Dependent Downhill Unfolding of Ubiquitin. I. Nanosecond-to-Millisecond Resolved Nonlinear Infrared Spectroscopy. *Proteins Struct. Funct. Bioinforma.* **2008**, *72*, 474–487.

(41) Rabl, C. R.; Martin, S. R.; Neumann, E.; Bayley, P. M. Temperature Jump Kinetic Study of the Stability of Apo-Calmodulin. *Biophys. Chem.* **2002**, *101–102*, 553–564.

(42) Greetham, G. M.; Clark, I. P.; Young, B.; Fritsch, R.; Minnes, L.; Hunt, N. T.; Towrie, M. Time-Resolved Temperature-Jump Infrared Spectroscopy at a High Repetition Rate. *Appl. Spectrosc.* **2020**, *0*, 1–8.

(43) Fritzsch, R.; Greetham, G. M.; Clark, I. P.; Minnes, L.; Towrie, M.; Parker, A. W.; Hunt, N. T. Monitoring Base-Specific Dynamics during Melting of DNA–Ligand Complexes Using Temperature-Jump Time-Resolved Infrared Spectroscopy. *J. Phys. Chem. B* **2019**, *123*, 6188–6199.

(44) Wray, W. O.; Aida, T.; Dyer, R. B. Photoacoustic Cavitation and Heat Transfer Effects in the Laser-Induced Temperature Jump in Water. *Appl. Phys. B Lasers Opt.* **2002**, *74* (1), 57–66.

5 Optimisation of detecting changes in protein structure by IR spectroscopy in protein-drug binding

5.1 Abstract

Understanding changes in protein structure upon protein-drug binding is an important consideration in determining if a biological response will be elicited. Small molecule binding is expected to generate relatively small changes in IR signal compared to the previous unfolding work in chapter (3). The development of data analysis for detecting changes due to protein-drug interactions will be assessed using CaM as previous knowledge of binding to a small molecule, trifluoperazine (TFP), is available from previous publications. TFP binds to *holo*-CaM, resulting in a tightly compact structure, and is expected to bind to *apo*-CaM, albeit via a different mechanism. In this chapter, the interaction of CaM with Ca^{2+} and TFP in different ratios will be analysed with FTIR and 2D-IR spectroscopy and processed using PCA as a large dataset. The results show that PCA of FTIR measurements can separate the binding of calcium and TFP to CaM into distinguishable groups. Analysis of spectral features from 2D-IR confirms the changes observed to be a loss of α-helix with addition of TFP and small changes in the electrostatic properties. *Apo*-CaM binds TFP also but requires a larger amount of TFP to induce the same changes in electrostatic properties due to intra-domain binding sites being inaccessible. Evidence from 2D-IR spectral features are able to corroborate the separation by PCA that is observed by FTIR spectroscopy showing the requirement for both techniques in determining changes due to binding. These results are probing small changes in structure and the interferometer 2D-IR approach used in chapter (3) was deemed unsuitable for detecting such small changes due to insufficiencies in phasing the spectra. This problem was eliminated with the use of a pulse shaper approach.

5.2 Introduction

The importance of understanding protein-drug binding has garnered a large degree of attention in recent years[1–3]. Both Infrared (IR) spectroscopy and Two-dimensional Infrared (2D-IR) spectroscopy have been previously utilised in examining protein-drug binding[4–10], with FTIR spectroscopy showing that the secondary structure of proteins is altered upon binding of a drug molecule, with particular focus on the amide I region. The 2D-IR experiments focused on probing the structural changes of the drug molecule as opposed to the protein.

This chapter will concentrate on analysing large sample sizes with measurements focusing on structural changes of the protein in protein-drug binding, and again the protein of interest will be calmodulin (CaM). CaM is involved in many processes and is targeted by several drug

molecules[11,12]. An example of a process is the activation of phosphodiesterases (PDEs), which are enzymes that break bonds in second messenger molecules and play an important part in signal transduction[13]. This activity can be inhibited by certain drug molecules. One such drug is trifluoperazine[14–22] (TFP) (Figure 5-1). TFP is a well-known drug, with many different uses, including the inhibition of the activation of PDE[23], as above, and its use in treating schizophrenia[20]. Since its discovery in the 1980's, it has been shown to inhibit many other functions of CaM. This is done by restricting the access to hydrophobic binding sites within CaM that are exposed upon binding of Ca^{2+} to CaM.

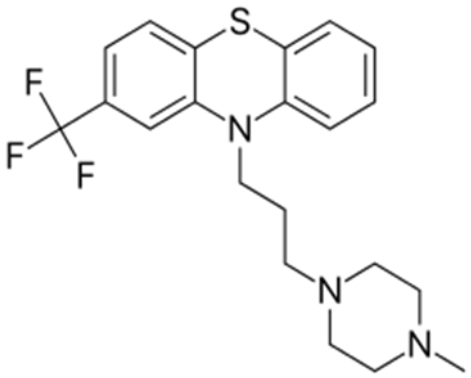


Figure 5-1. Trifluoperazine (TFP) molecular structure.

The binding of TFP to CaM has been previously well studied and there is a wealth of information proposed on the mechanics of the binding. As the structural changes of binding TFP to *holo*-CaM have been well established, it provides a suitable starting point for testing the capabilities of 2D-IR spectroscopy coupled to PCA analysis in detecting small effects observed in the protein from protein-drug binding interactions. This chapter will also see further development of the 2D-IR ULTRA instrument located at the Rutherford Appleton Laboratory by progressing the optical setup from an interferometer approach used in chapter (3) that involves manually "phasing data" in the data processing to the use of a more automatic approach, a pulse shaper[24]. The benefits and limitations of both methods will be investigated and compared to the capabilities of linear IR spectroscopy for both Ca^{2+} and TFP binding analysis.

The sequence of structure changes for CaM binding to both Ca^{2+} and TFP have been determined from previous experimental studies. The sequence can be seen in Figure 5-2 and is from *apo*-CaM (closed) to *holo*-CaM (open) to TFP bound to *holo*-CaM (closed), with each additional equivalent TFP bound producing a more compact structure (Figure 5-2). These

results from structural experiments provide insight into the binding of TFP to CaM[18,25,26]. The binding ratio of 1:1 *holo*-CaM to TFP shows TFP binding to the exposed hydrophobic pocket of the C-domain. This binding of one equivalent TFP induces a structural change with disruption to the central linker resulting in the open dumbbell-like structure of *holo*-CaM becoming closed, similar to that of the *apo*-CaM structure. The binding ratio 1:2 (*holo*-CaM to TFP) results in a more enclosed structure with TFP occupying sites in both domains. The final structure available, the 1:4 ratio indicates a tightly compact state[18,25–28].

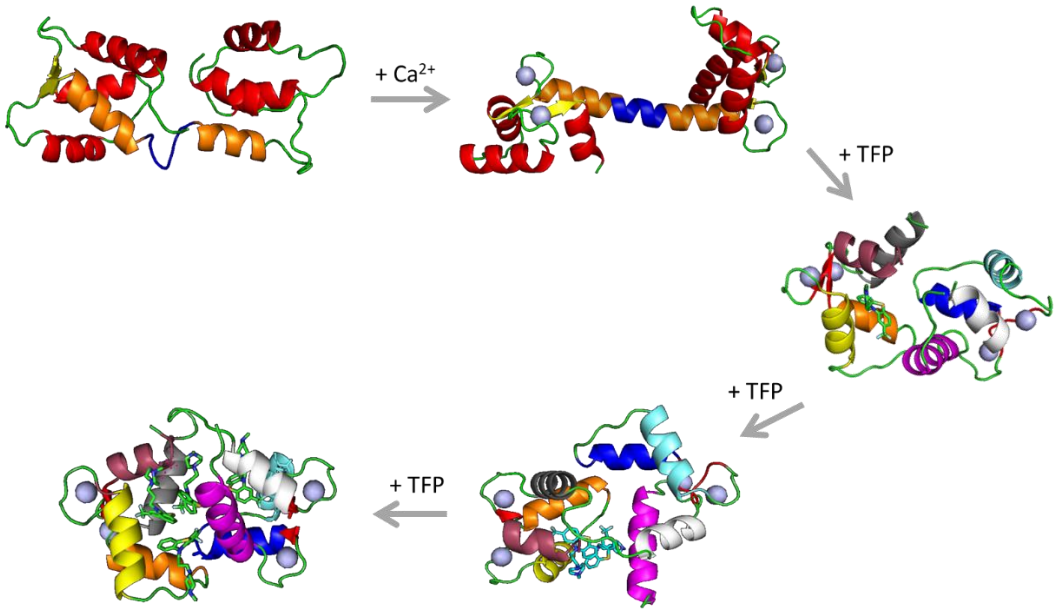


Figure 5-2. Sequential change in crystallography structure of CaM upon addition of Ca^{2+} and TFP[18,25–28]. PDB references: 1CFD, 1CLL, 1CTR, 1A29, 1LIN.

There are two types of sites which TFP binds to, intra-domain sites for both the C- and N-domains, and inter-domain sites within the central linker region[14] and binding occurs both electrostatically[29] and by hydrophobic interactions between the CaM and TFP. Contrary to earlier publications[21,30], TFP can bind to *apo*-CaM, as demonstrated by small-angle X-ray scattering and NMR experiments[15,19]. However, as the binding sites are not exposed from the extension of the central linker, the binding works in a different manner to that of TFP bound to *holo*-CaM. The C-domain is semi-open in its *apo*-CaM form – it switches between semi-open and closed which is suggested by MD simulations[31] which show *apo*-CaM switching to open briefly over the millisecond simulation. This semi-open state allows TFP to bind to the intra-domain binding site of *apo*-CaM C-domain[19]. However, the N-domain for *apo*-CaM is closed and therefore TFP cannot access this intra-domain site. Binding can also occur in the interdomain sites for *apo*-CaM situated in the central linker region.

There is contradictory information about the number of equivalent TFP molecules that can bind to a single CaM[21,32], with up to seven equivalent TFP to one CaM suggested[33]. A value that recurs often in the literature is four equivalent TFP binding to one CaM in the *holo*-form[14,19,21]. With Ca^{2+} bound to CaM, four binding sites are exposed, two in the C-domain and two in the N-domain, with binding occurring preferentially to the C-domain first[16]. Alongside 2D-IR method development, this chapter will endeavour to confirm how many TFP molecules can bind to one CaM molecule.

In addition to binding TFP to *holo*-CaM, the effect on binding TFP to *apo*-CaM will be examined to evaluate for different effects of binding and if IR spectroscopy has the capabilities of detecting said changes or if there will be no distinction made.

5.3 Method

5.3.1 Sample Preparation

TFP, D_2O, Trizma base, DCl, and $CaCl_2$ were all purchased from Sigma-Aldrich and used without further purification. The human CaM was expressed and purified by Daniel Shaw at UCB. CaM was prepared at a concentration of 0.8 mM in deuterated 50mM Tris-DCl pD 7.4 and Ca^{2+} added at ratios of 2, 4 and 10 per CaM. TFP was added at ratios of 0.25, 0.5, 0.75, 1, 2, 3, 4, and 10 molecules per CaM. Temperature dependent measurements were performed by heating the samples from 20 °C to 80 °C in 10 °C steps. All concentrations were doubled for the 2D-IR pulse shaper measurements. Room temperature samples were housed in a Harrick cell, and temperature dependent samples were housed in a temperature-controlled Harrick cell. All samples were prepared between two CaF_2 windows separated by a PTFE spacer with a thickness of 50 μm.

5.3.2 FTIR Spectroscopy

All linear infrared spectra were collected on a Thermo Scientific Nicolet iS10 spectrometer at a resolution of 1 cm^{-1}.

5.3.3 2D-IR Spectroscopy

2D-IR measurements were performed on the ULTRA instrument[34] at RAL using both an interferometer and a pulse shaper method, as described in chapter (2). Both interferometer and pulse shaper data were measured with a 2nd order diffraction grating for the detection of the spectra. This gave comparable probe resolutions of 1.4 cm^{-1}. All 2D-IR measurements were performed at a T_w of 250 fs.

5.4 Results and Discussion

5.4.1 Ca^{2+} ratios comparison

Prior to analysis of different ratios of TFP bound to CaM it is imperative to understand what occurs when different numbers of Ca^{2+} are bound to CaM as the binding of Ca^{2+} ions to CaM is important in determining how TFP binds to CaM. To summarise the results from chapter (3), *apo*-CaM is a closed globular structure with a flexible central linker. Upon heating, *apo*-CaM undergoes structural change with a loss of approximately 13-15% α-helix to random coil with two melting transitions occurring at 46 °C and 58 °C for the C- and N- domains, respectively. Upon excess binding of Ca^{2+} (*holo*-CaM) the protein opens, and the central linker becomes fully helical with the overall structure becoming more rigid, which is confirmed by the lack of thermal unfolding below 80 °C for *holo*-CaM.

CaM can bind up to four Ca^{2+} and chapter (3) aimed to highlight the gross changes from *apo*-CaM to *holo*-CaM. However, this chapter focuses on analysing subtle changes between different ratios of Ca^{2+} ions to CaM. As binding occurs cooperatively with two ions binding to the C-domain first and the following two to the N-domain, analysis on CaM bound to Ca^{2+} in the following ratios has been performed by interferometric 2D-IR spectroscopy: *apo*-CaM, 1:2, 1:4, and 1:10 (*holo*-CaM, excess Ca^{2+}) CaM:Ca^{2+}. Both 1:4 and 1:10 were analysed to understand if excess Ca^{2+} would influence the structure differently than a 1:4 ratio.

The 2D-IR spectra for each ratio at 20 °C can be seen in Figure 5-3. The spectra are visually similar with each comprising of a negative diagonal peak (red) attributable to the V=0-1 amide I transition and an off-diagonal peak (blue) from the v=1-2 transition.

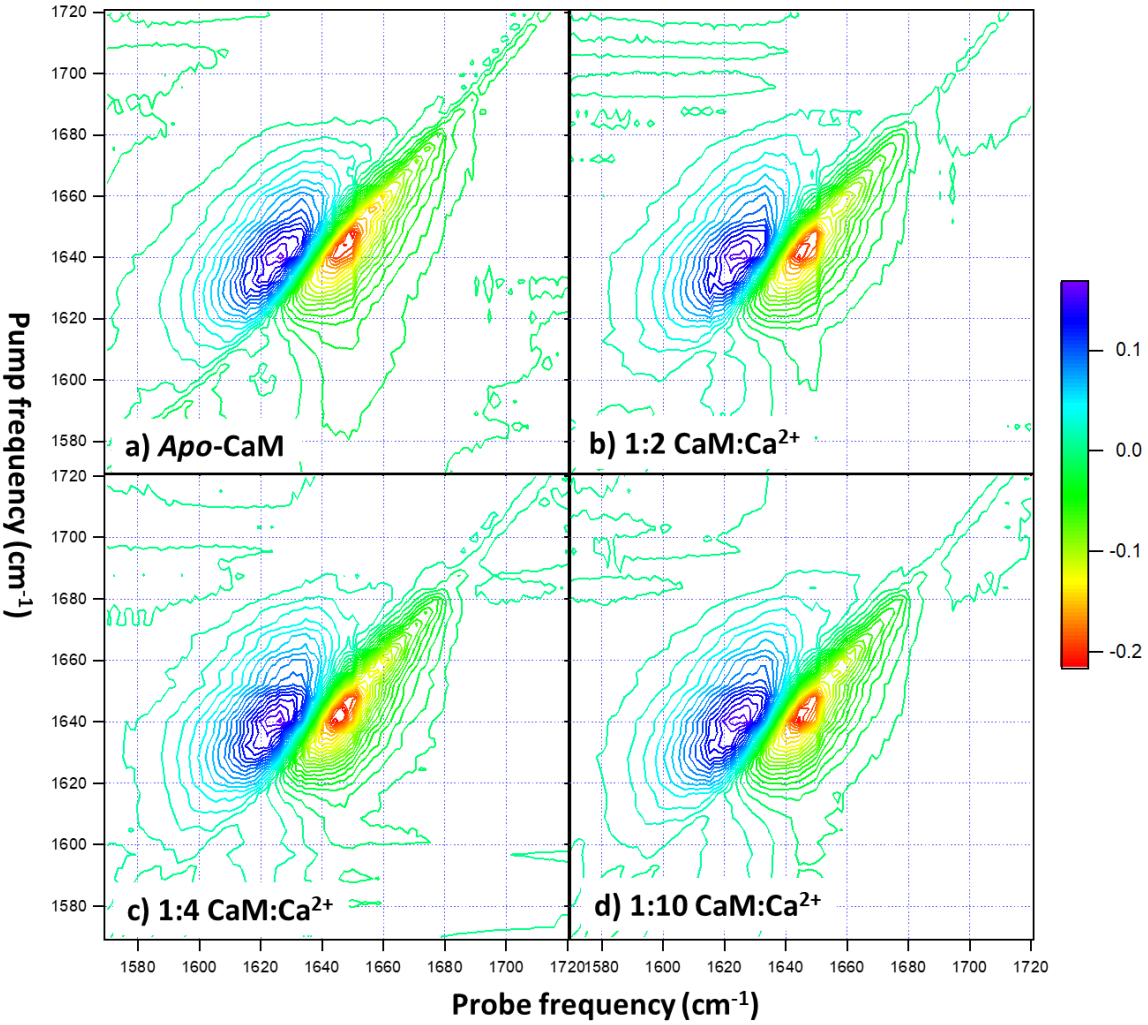


Figure 5-3 - Raw 2D-IR spectra for a) apo-CaM, b) 1:2 CaM:Ca^{2+}, c) 1:4 CaM:Ca^{2+}, and d) 1:10 CaM:Ca^{2+}, all measured at 20 °C.

To evaluate the differences in structure, difference spectra have been calculated by subtracting the *apo*-CaM spectrum from all other Ca^{2+} ratio spectra (Figure 5-4). The spectra were normalised to the maximum point of the amide I bleach at around 1640 cm^{-1} first to account for slight variations in laser intensity and concentration of the protein.

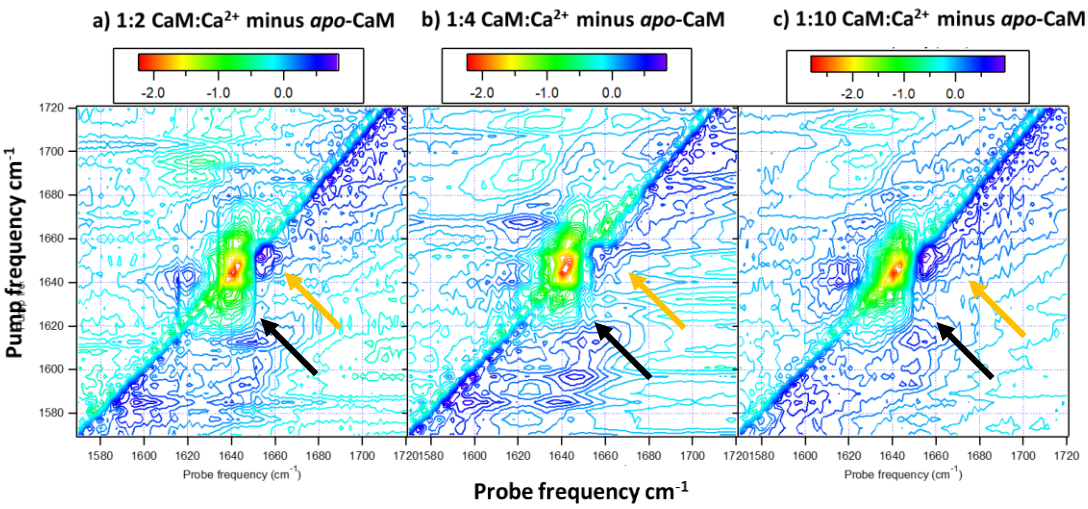


Figure 5-4. 2D-IR difference spectra performed using an interferometer approach at room temperature (20 °C). Apo-CaM subtracted from a) 1:2 CaM to Ca^{2+}, b) 1:4 CaM to Ca^{2+}, and c) 1:10 CaM to Ca^{2+}. All raw spectra were normalised prior to subtraction.

Apo-CaM minus 1:10 (Figure 5-3, (c)) was previously described in chapter (3). The red peak at 1640 cm^{-1} (as indicated by the black arrow) shows a gain of α-helix structure, and the blue peak at 1655 cm^{-1} (orange arrow) shows a loss of random coil for the 1:10 CaM:Ca^{2+} sample compared to *apo*-CaM. This was explained by the change in the central linker from a mix of random coil and α-helix to fully α-helical upon complete Ca^{2+}binding. The 1:2 difference spectrum (Figure 5-3, (a)) and the 1:4 difference spectrum (Figure 5-3, (b)) show similar features of a decrease in random coil (orange arrow) and an increase in α-helix (black arrow) for the Ca^{2+} bound proteins. The large difference in signal of the two secondary structure features is explained by α-helix having a larger transition dipole moment than random coil, and therefore dominating the signal, as described in chapter (3). Differences between the 1:10 and 1:4 results are minimal which is to be expected as a maximum of four Ca^{2+} can bind to one equivalent CaM. It is therefore suggested that the use of excess Ca^{2+} within *holo*-CaM has no adverse impact on the results, and it is deduced that the results from 1:10 and 1:4 are qualitatively the same.

The 1:2 result shows very little difference between itself and that of the 1:4 and 1:10 difference spectra, indicating that two Ca^{2+} bound to CaM has a more similar structure to that of *holo*-CaM compared to *apo*-CaM. This finding is not unexpected as previous studies have indicated that Ca^{2+} bind to the C-domain first[35], and this results in the protein structure changing from a closed state to an open state with the central linker becoming fully helical[36] – as is observed in the structure of *holo*-CaM.

An observation to note in all three difference spectra is the presence of a negative signal across the diagonal. This is a result of light scattering reaching the detector after passing through the sample. This is not observed in the raw spectra (Figure 5-3) as the signal size of the amide I dominates this lower scatter signal. As the differences in secondary structure are small (5-6%, from chapter (3)) this has enhanced the scatter signal.

Thermal measurements of CaM in the presence of different ratios of Ca^{2+} were also performed by heating the protein samples from 20 °C to 80 °C, as in chapter (3). PCA, a multivariate analysis method which extracts trends in datasets, was carried out on these data with analysis performed individually for each ratio dataset. Two PCs were used to describe the data for each analysis and PC spectra and trends for the 1:2 analysis are shown as an example in Figure 5-5.

The PC1 spectra for each analysis consisted of two peaks, a negative peak on the diagonal (Figure 5-5 (a), red peak) and a positive peak off the diagonal (blue peak). This correlated to the average spectra for all spectra in the datasets, which is confirmed by the temperature independent trend from PCA (Figure 5-5 (c), black) and comparison between the average spectrum for the 1:2 dataset and the PC1 spectrum (Figure 5-6).

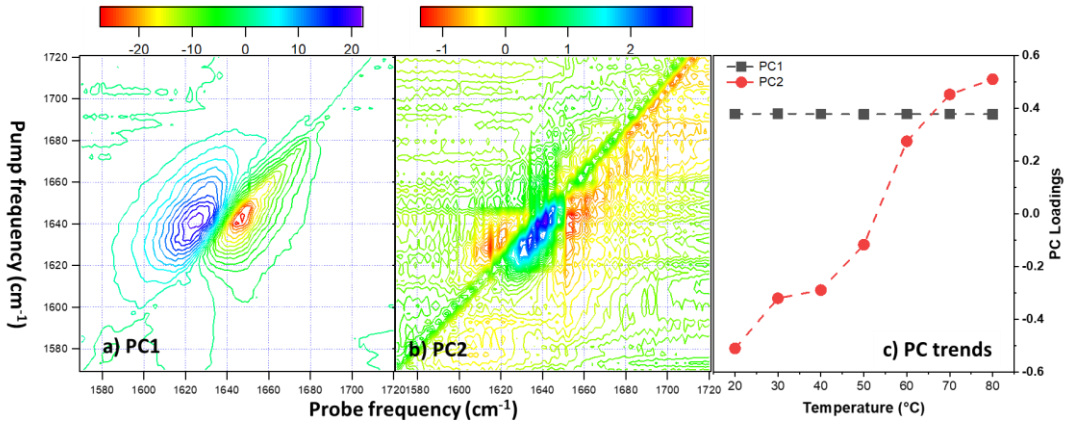


Figure 5-5 PCA results for 1:2 CaM:Ca^{2+} showing a) PC1 spectrum, b) PC2 spectrum, and c) trends for PC1 and PC2.

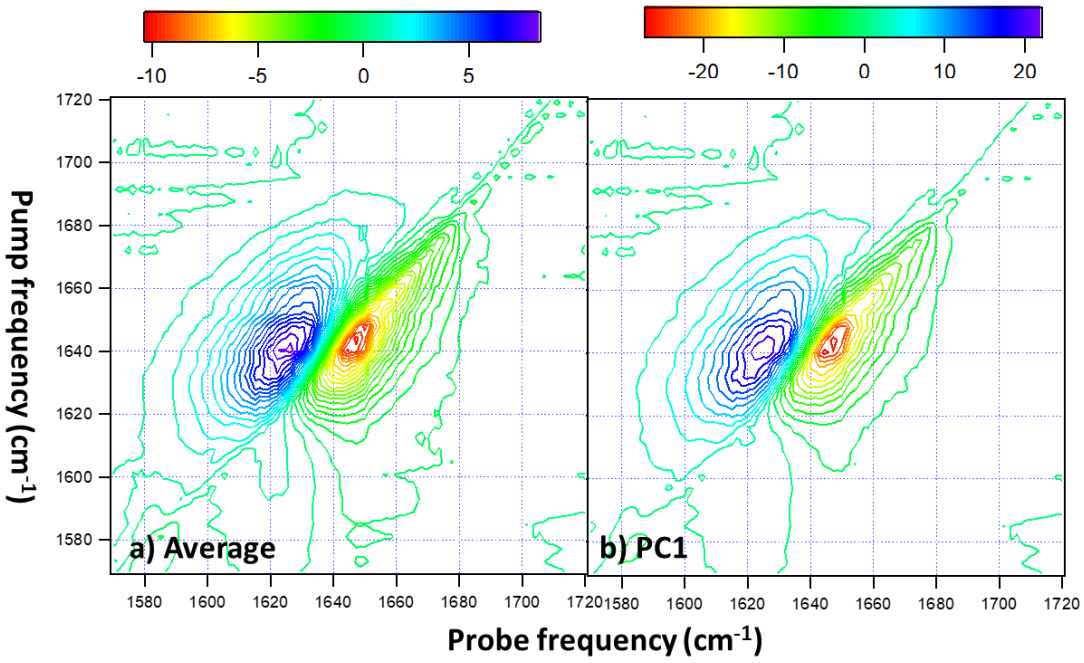


Figure 5-6. Comparison between a) average spectrum, and b) PC1 spectrum for 1:2 CaM:Ca^{2+}.

PC2 was temperature dependent (Figure 5-5 (c), red) with the spectra for all ratios showing a peak pair at 1640 cm^{-1} indicating a loss of α-helix, and a peak pair at 1655 cm^{-1} indicating a gain of random coil with increasing temperature (Figure 5-7). The results from these thermal measurements will not be discussed in detail here in this chapter as they are comparable to those achieved in chapter (3). Briefly, the 1:2 PC2 spectrum (Figure 5-7 (b)) shows more similarities with the PC2 spectra of 1:4 (Figure. 5-7 (c)) and 1:10 (Figure 5-7 (d)) CaM:Ca^{2+} than the PC2 spectrum of *apo*-CaM (Figure 5-7 (a)). This is determined by the position of the random coil peak where is it positioned at higher wavenumber (~1660 cm^{-1}) for *apo*-CaM compared to the Ca^{2+} bound spectra (~1650 cm^{-1}). This is highlighted by the black arrows in Figure 5-7.

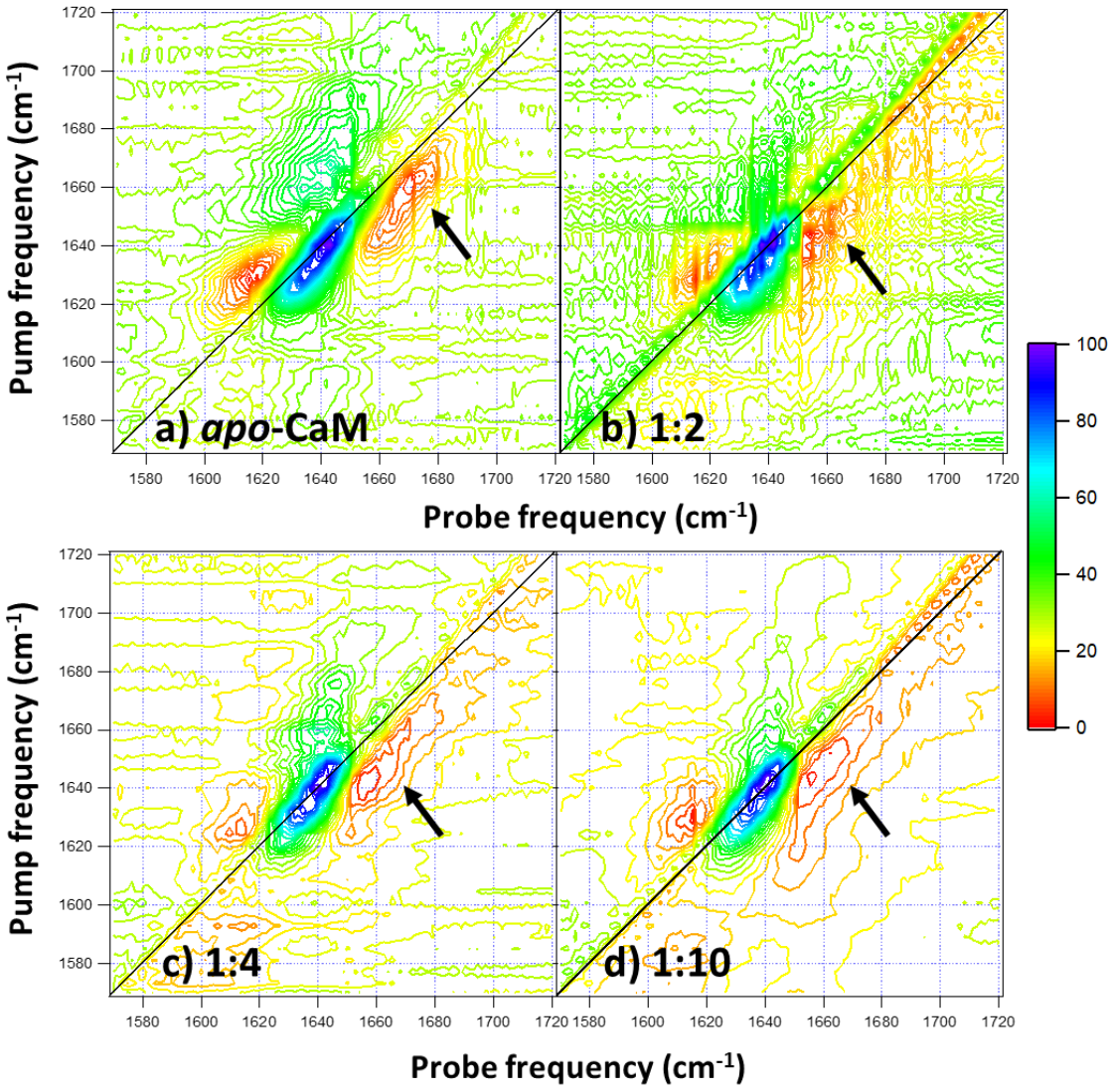


Figure 5-7. Individual PC2 spectra from 2D-IR thermal measurements (20 °C to 80 °C) of a) apo-CaM, b) 1:2 CaM to Ca^{2}, c) 1:4 CaM to Ca^{2+}, and 1:10 CaM to Ca^{2+}.

Analysis of the percent change in the 1640 cm^{-1} signal amplitude over the temperature range measured for the Ca^{2+} bound CaM at different ratios was calculated using the transition dipole moment, $d(\omega)^{rel}$ discussed in chapter (3). Briefly, a ratio of the linear IR spectrum and the diagonal of the 2D-IR spectrum is calculated for each measurement using equation 5-1, with each measurement multiplied by the maximum signal of the FTIR plot at 20 °C divided by the maximum signal of the 2D-IR diagonal at 20 °C.

$$d(\omega)^{rel} = \frac{2DIR^{T}(\omega)}{FTIR^{T}(\omega)} \frac{OD^{20}}{\Delta OD^{20}} \qquad \text{(Equation 5-1)}$$

The resulting spectrum was fitted to two Gaussian functions and the amplitude of the 1640 cm^{-1} peak was used to calculate the percentage difference in signal size from the lowest temperature measurement (20 °C). The results of the percentage change in signal size provided good agreement across all three ratios with a decrease in signal size observed of approximately 65% (Figure 5-8). This is not an indication of predicted decrease in α-helix content or helical population, but it does show that the changes in signal intensity from the FTIR and 2D-IR diagonal data are comparable across the CaM bound to Ca^{2+} datasets.

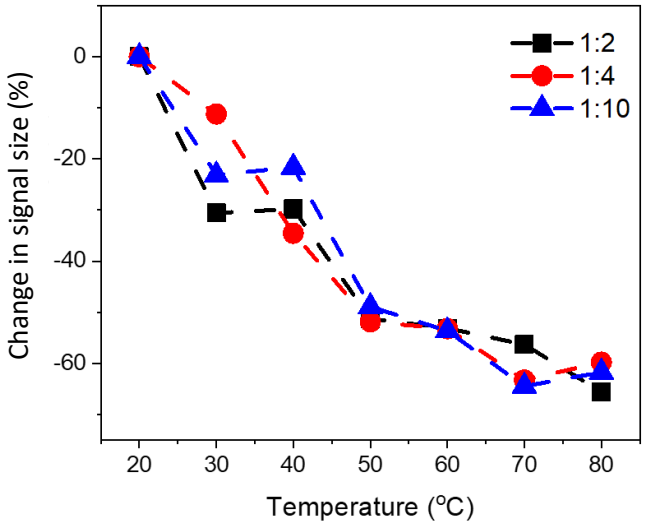


Figure 5-8. Transition dipole moment, $d(\omega)^{rel}$, calculations of percent change in signal amplitude over the 20 °C to 80 °C temperature range for the three different ratios of CaM to Ca^{2+} measured: 1:2 (black), 1:4 (red) and 1:10 (blue).

These CaM to Ca^{2+} ratio results suggests that binding of just two Ca^{2+} to CaM produces a similarly rigid structure to that of *holo*-CaM. This can be explained from not only the change in the central linker but also from the findings from MD simulations of *apo*-CaM (reproduced here in Figure 5-9 specifically the RMS-fluctuation (RMS-f) results which show that the C-domain (residues 90-148) is the more flexible domain (Figure 5-9 (d)). The *apo*-CaM (black) C-domain has higher RMS-f values than the *holo*-CaM C-domain (red). The N-domain, by contrast, does not undergo a drastic change in flexibility upon Ca^{2+} binding. Binding of Ca^{2+} to the C-domain rigidifies it, bringing the flexibility of the domain to a similar degree as that of the *apo*-CaM N-domain. Therefore, when the final two Ca^{2+} are bound to the N-domain, the degree of additional rigidification is minimal. Hence why little difference between the 1:2 CaM to Ca^{2+} and the *holo*-CaM results are detected.

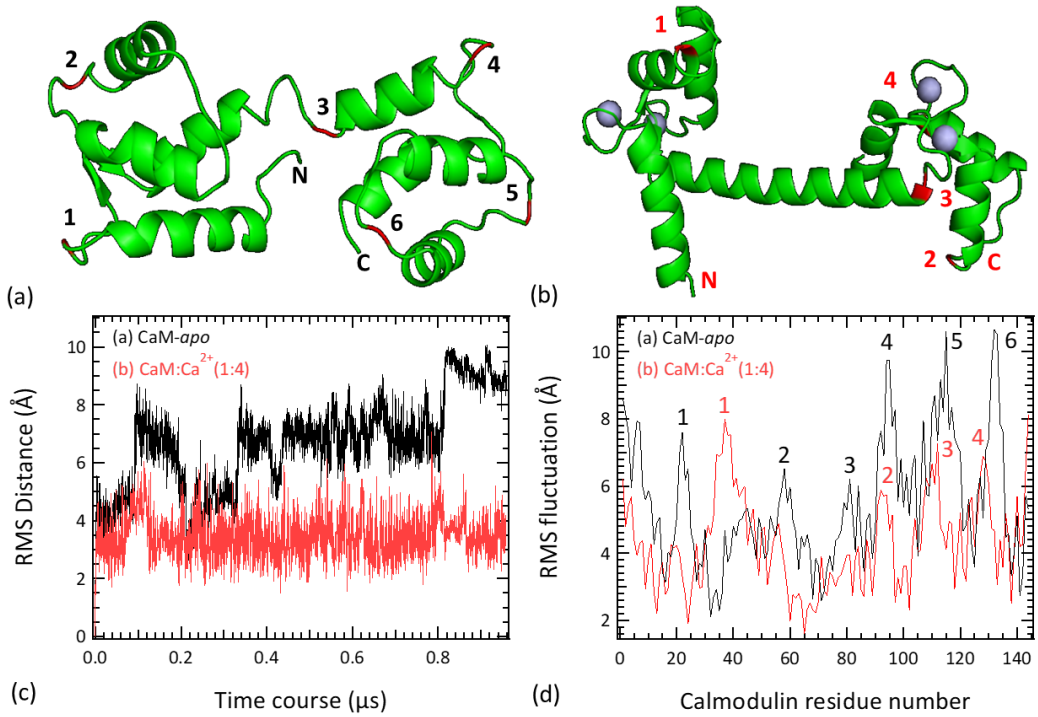


Figure 5-9. MD simulation results of apo-CaM and holo-CaM over 1 millisecond with a) the average structure of apo-CaM over the simulation, b) the average structure of holo-CaM over the simulation, c) the RMS-distance results for both states indicating the degree of flexibility over the simulation, and d) the RMS-fluctuation indicating the flexibility in individual residues in apo-CaM (black) and holo-CaM (red).

5.4.2 TFP binding to CaM

A series of FTIR spectroscopy measurements were performed with different ratios of TFP and Ca^{2+} bound to CaM (Table 5-1), resulting in 36 separate measurements at room temperature. Analysis of the spectra was performed on the amide I region only (1600 cm^{-1} to 1700 cm^{-1}), with each spectrum normalised to the maximum of the amide I.

The spectra show a peak centred at approximately 1645 cm^{-1} (Figures 5-10 (a) and (b) showing *apo*-CaM and 1:10 spectra, respectively) with the peaks broadly similar across all measurements, although when zoomed in, the amide I peak shifts to higher wavenumber with increased binding of TFP (Figures 5-10 (c) and (d), brown to green). The shift is subtle with only a 1.6 cm^{-1} shift occurring for *apo*-CaM from zero TFP molecules present (peak position 1644.6 cm^{-1}) to ten TFP molecules present per CaM (peak position of 1646.0 cm^{-1}).

Table 5-1. List of different ratios of Ca^{2+} and TFP bound to CaM which constitutes the measurements by FTIR and 2D-IR, all performed at room temperature.

	CaM:Ca^{2+}[1:0]	CaM:Ca^{2+}[1:2]	CaM:Ca^{2+}[1:4]	CaM:Ca^{2+}[1: 10]
TFP:CaM	0	0	0	0
	0.25	0.25	0.25	0.25
	0.5	0.5	0.5	0.5
	0.75	0.75	0.75	0.75
	1	1	1	1
	2	2	2	2
	3	3	3	3
	4	4	4	4
	10	10	10	10

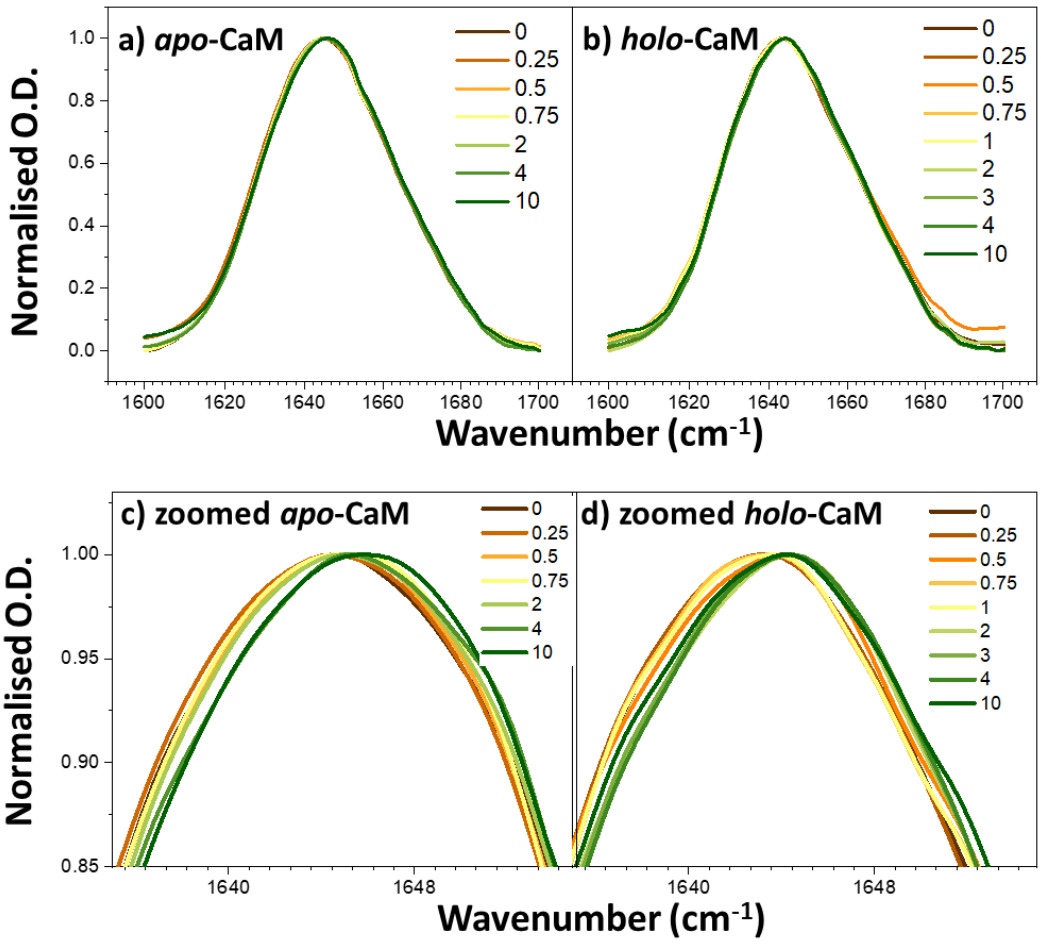


Figure 5-10. Normalised amide I region FTIR spectra for a) apo-CaM (zoomed (c)) and b) holo-CaM (1:10 CaM:Ca^{2+}) (zoomed (d)). Legend numbers are TFP:CaM ratios.

The 36 measurements (Table 5-1) were also performed with 2D-IR spectroscopy using a pulse shaper approach. A pulse shaper does not require manual phasing – it uses predetermined

waveforms; therefore, no autocorrelation is needed to calculate time zero. All the measurements were taken using the same waveforms.

Visually, the spectra are similar (Figure 5-11 showing representative spectra) with each spectrum comprising of two peaks, a negative on diagonal peak (red) and a positive off diagonal peak (blue). The diagonals were extracted (Figure 5-12) and it can be seen that, like the FTIR spectra, the peaks are broadly similar and centred at approximately 1645 cm^{-1} but enhancing the peak highlights a shift to higher wavenumber with increasing TFP ratio (Figures 5-12 (c) and (d)).

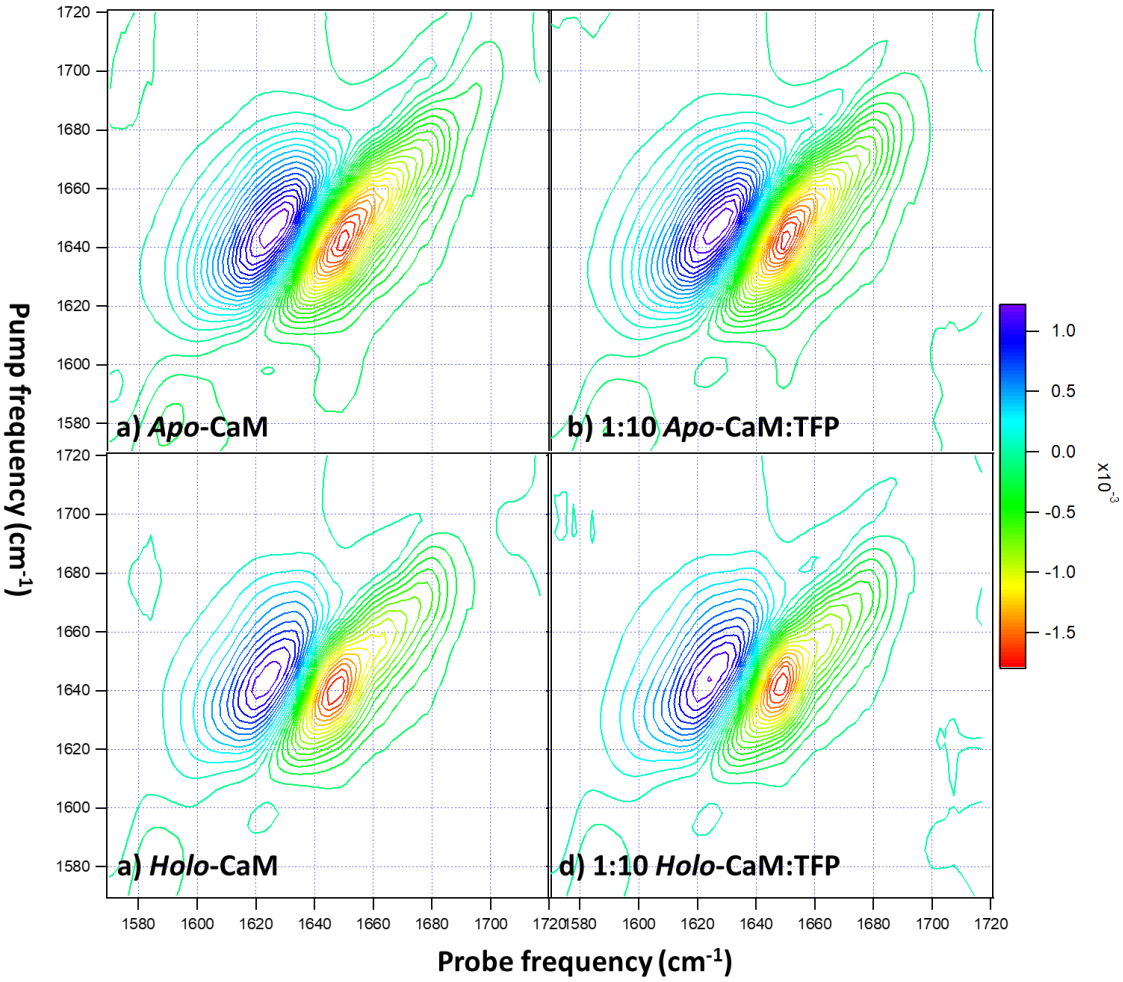


Figure 5-11. 2D-IR spectra from pulse shaper system for a) apo-CaM without TFP, b) 1:10 apo-CaM:TFP, c) holo-CaM (1:10 CaM:Ca^{2+}) without TFP, and d) 1:10 holo-CaM:TFP.

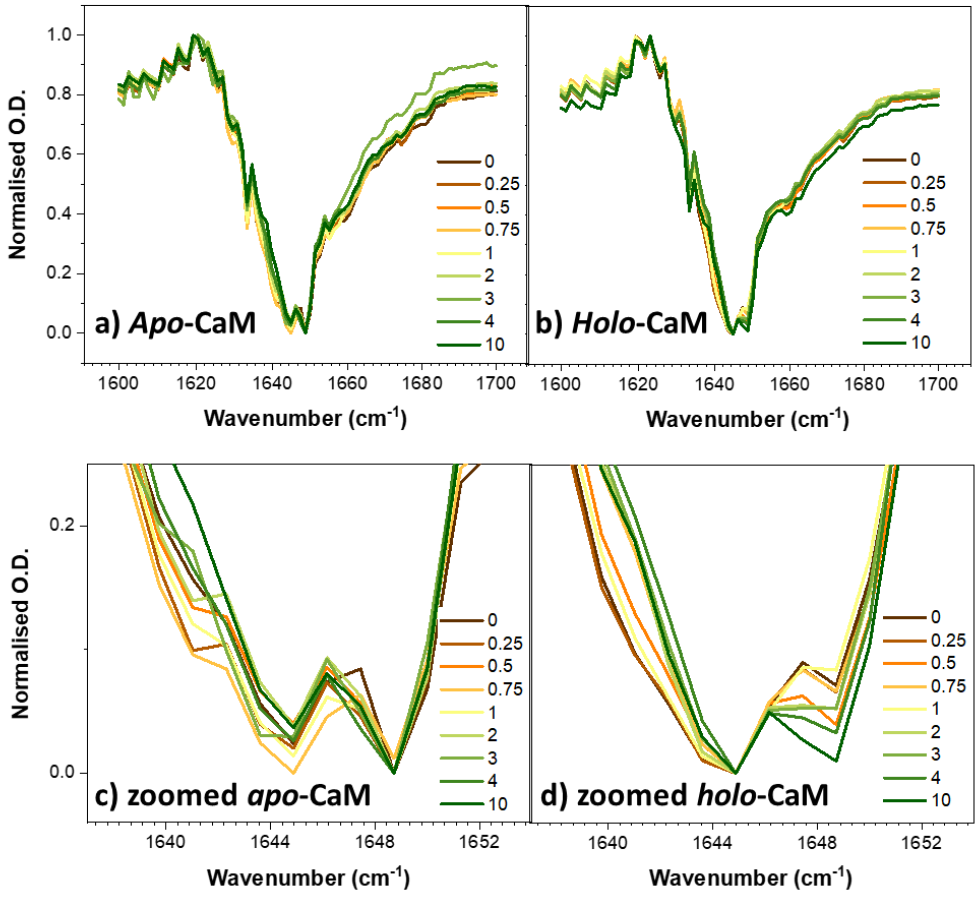


Figure 5-12. 2D-IR diagonals for a) apo-CaM (zoomed (c)) and b) holo-CaM (1:10 CaM:Ca^{2+}) (zoomed (d)). Legend numbers are TFP:CaM ratios.

5.4.3 PCA

PCA was performed on each of the FTIR, 2D-IR and 2D-IR diagonal CaM:Ca^{2+}:TFP datasets. The results of PCA were three distinct PCs which were responsible for describing the majority of the variance within each dataset (99.99%, 99.82%, and 99.88% of the variance for the FTIR, 2D-IR and 2D-IR diagonals datasets, respectively). The results of each individual PC will be discussed in detail later in the chapter.

The PC2 and PC3 components for each dataset have been used to separate the spectra into groups based upon their binding of Ca^{2+} or TFP to CaM to give an overview of the PCA results (Figure 5-13). The five groups present are: *apo*-CaM (black), *apo*-CaM with TFP bound (*apo*-CaM; TFP, red), Ca^{2+} bound CaM (CaM:Ca^{2+} in the ratios 1:2, 1:4 and 1:10, blue), Ca^{2+} and TFP bound CaM (Ca^{2+}-bound CaM;TFP , pink), Ca^{2+} bound CaM with TFP present with a ratio of less than 1:1 CaM:TFP (Ca^{2+}-bound CaM/TFP < 1, green).

The PCA groupings for the FTIR spectra provide good distinction between groups except for Ca^{2+}-bound CaM and Ca^{2+}-bound CaM/TFP < 1 which were unable to be separated (Figure 5-13 (a)). This is not surprising as a ratio of less than 1:1 CaM:TFP would suggest minimal binding has occurred. Both PC2 and PC3 appear to contribute to the separation of the groups with the biggest separation occurring between *apo*-CaM spectra and Ca^{2+} bound spectra. PC2 appears to contribute more to the separation between *apo*-CaM and TFP binding, while PC3 appears to contribute more to the separation between Ca^{2+}-bound CaM and TFP binding.

The PCA groupings for the whole 2D-IR spectra can only provide good separation between the Ca^{2+} free and Ca^{2+} bound spectra. This requires both PC2 and PC3 to completely separate the groups. There is overlap between Ca^{2+}-bound CaM spectra and Ca^{2+}-bound CaM bound to TFP spectra which are not able to be separated.

The 2D-IR diagonal analysis can provide some separation of data, a clear separation of Ca^{2+} binding, with TFP binding of *holo*-CaM being unable to be separated completely. Separation of *apo*-CaM and *apo*-CaM/TFP is minimal.

It initially appears that FTIR is better suited for protein-ligand analysis compared to 2D-IR spectroscopy. However, the spectral features associated with PC2 and PC3 will now be interrogated for each of the three analyses to comprehend how each contributes to the understanding of protein-ligand binding.

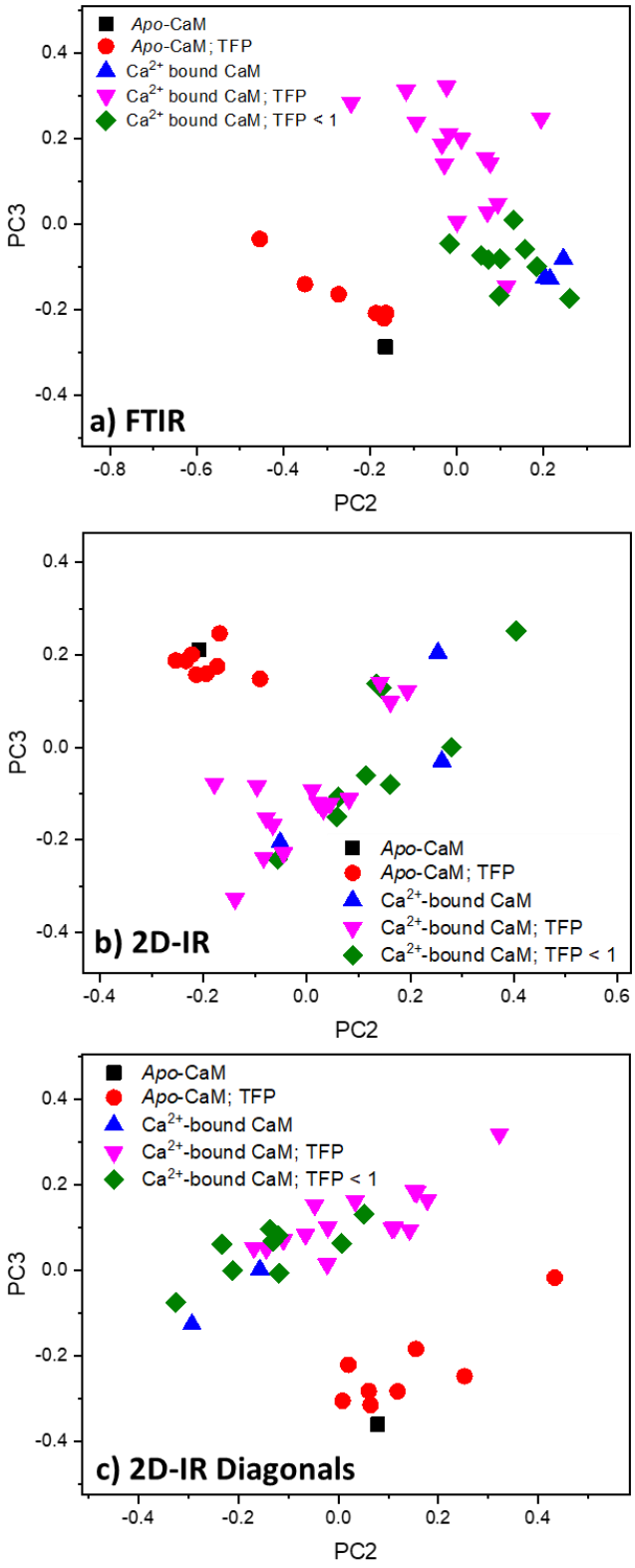


Figure 5-13. PCA grouping between PC2 and PC3 for a) FTIR results, b) 2D-IR results, and c) the extracted 2D-IR diagonals. Ca^{2+}-bound CaM refers to CaM:Ca^{2+} ratios of 1:2, 1:4 and 1:10.

5.4.4 Principal Component one

The major component of the FTIR PCA (PC1, 99.26% of variance) was independent of changes in concentration of TFP and Ca^{2+} (Figure 5-14 (d)), with the corresponding spectrum showing the average of the amide I peak in all measurements (Figure 5-14 (a)). Deviations in the trend were seen for the 1:2 CaM:Ca^{2+} set (red) at TFP ratios of three and ten due to bubbles in the samples affecting the spectral features and therefore removed from analysis.

PC1 of the 2D-IR data accounted for 98.58% of the variance. This component is the average spectrum of the 36 measurements, as with the results from the FTIR PCA and is independent of changes in concentrations between the samples (Figures 5-14 (b) and (e)).

PC1 of 2D-IR diagonals accounts for a variance of 99.52%. As with the FTIR and 2D-IR results it is independent of concentration of TFP and Ca^{2+} (Figure 5-14 (f)) and the spectrum represents the average diagonal spectrum (Figure 5-14 (c), inverted to show consistency with the FTIR spectrum).

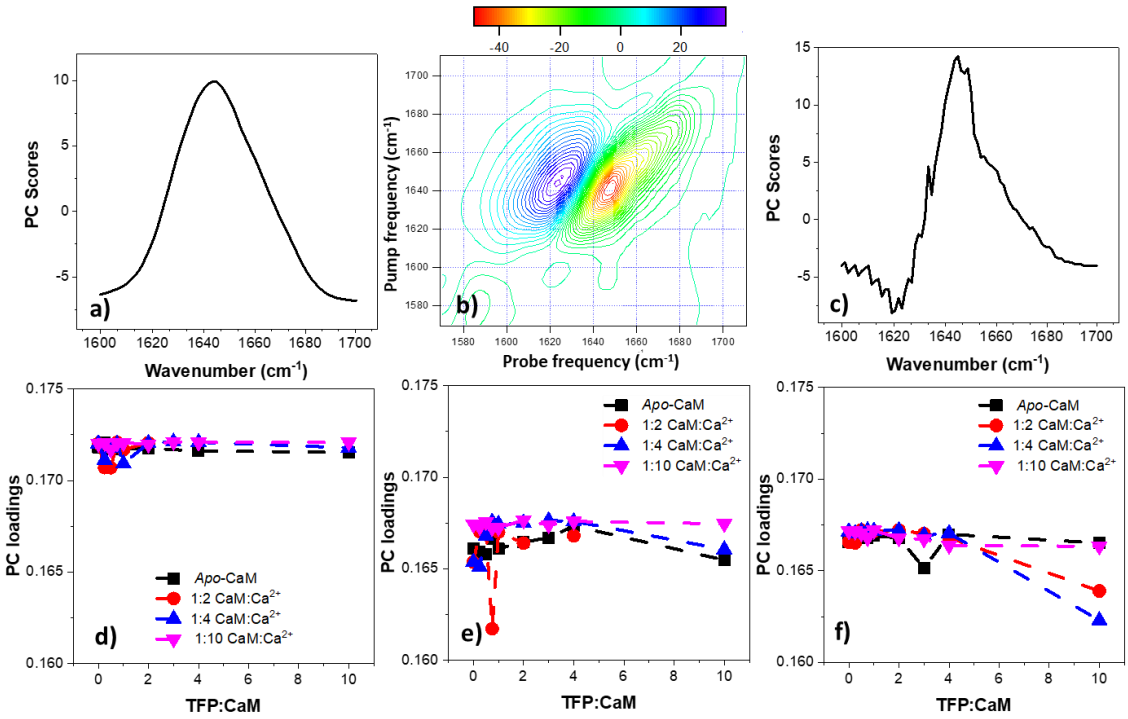


Figure 5-14. PC1 spectra for a) FTIR, b) 2D-IR and c) 2D-IR diagonals (inverted), and their associated trends, d) FTIR, e) 2D-IR and f) 2D-IR diagonals.

5.4.5 Principal Component two

The FTIR PC2 results describes 0.67% of the variance in the data. The spectrum shows two distinct peaks, a positive peak at 1635 cm^{-1} and a negative peak at 1655 cm^{-1} (Figure 5-15 (a)). The component is dependent on the concentration of both TFP and Ca^{2+}, as indicated by the

trends (Figure 5-15 (d)). The *apo*-CaM trend (black line) is significantly different from the Ca^{2+} bound trends, with an amplitude range of approximately -0.2 to -0.45, compared to all the Ca^{2+} bound trends which range from approximately 0.2 to -0.1. All Ca^{2+} bound trends are similar which is unsurprising after the results of the Ca^{2+} ratio measurement described earlier; the central linker is fully helical in all three Ca^{2+} states with the protein in an open conformation, therefore exposing the TFP binding sites.

After the CaM:TFP 1:1 ratios the trends tend towards negative values with an apparent levelling off at 1:3 CaM:TFP for the Ca^{2+} bound CaM (except for 1:2 which is still observing the same bubble affects present in the PC1 data).

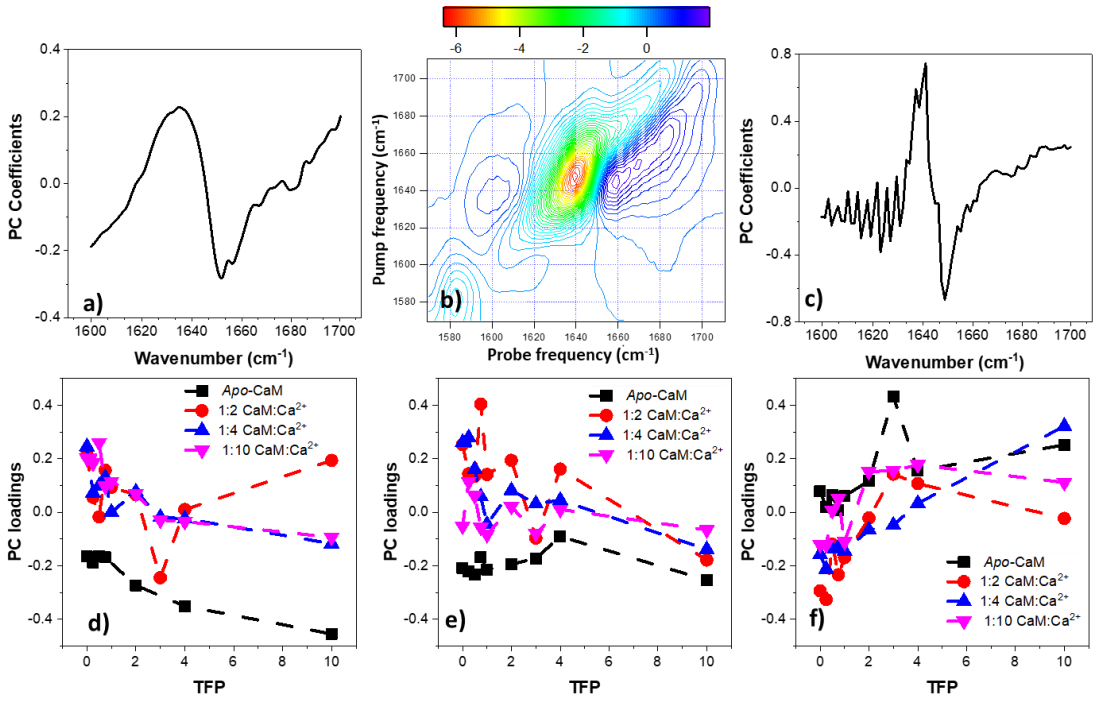


Figure 5-15. PC2 spectra for a) FTIR, b) 2D-IR and c) 2D-IR diagonals, and their associated trends, d) FTIR, e) 2D-IR and f) 2D-IR diagonals.

PC2 of the 2D-IR dataset accounts for 1.12% of variance. The spectrum comprises of three horizontal peaks, a positive peak at 1640 cm^{-1}, 1600 cm^{-1} (pump, probe), the second peak is negative at 1640 cm^{-1}, 1640 cm^{-1} (pump, probe), and the third peak is 1640 cm^{-1}, 1660 cm^{-1} (pump, probe), (Figure 5-15 (b)). This series of positive-negative-positive peaks off the diagonal suggests a shift to higher wavenumbers along the pump axis with binding of both Ca^{2+} and TFP to CaM. This is confirmed by simulating the difference spectrum between two 2D-IR spectra with peaks at different pump frequencies (Figure 5-16). Simulating the difference spectrum was done by creating 2D-IR spectra by using a 2D-Guassian function

(Equation 5-2), where A is amplitude of peak, x_0, x_1, y_o, and y_1 are the centres of the ESA and bleach peaks, and σ_x and σ_y are the spectral spread.

$$f(x,y) = Ae^{\left(-\frac{(x-x_0)^2}{2\sigma_x{}^2}\right)+\left(\frac{(y-y_0)^2}{2\sigma_y{}^2}\right)} - Ae^{\left(-\frac{(x-x_1)^2}{2\sigma_x{}^2}\right)+\left(\frac{(y-y_1)^2}{2\sigma_y{}^2}\right)} \quad \text{(Equation 5-2)}$$

The spectra were simulated with different pump frequencies but the same amplitudes. The spectrum of higher pump frequency was subtracted from the spectrum of lower pump frequency, resulting in the simulated difference spectrum. This shift to higher wavenumber is similar to what is observed for the PC2 spectrum of the FTIR data.

The trend for PC2 is nosier than the FTIR trend, however it does still separate out *apo*-CaM from the Ca^{2+} bound trends with a decrease in PC loadings for the Ca^{2+} bound trends (Figure 5-15 (e)). The *apo*-CaM loadings, however, appears to have very little difference from 1:0 to 1:10 CaM:TFP ratios, remaining at approximately -0.2.

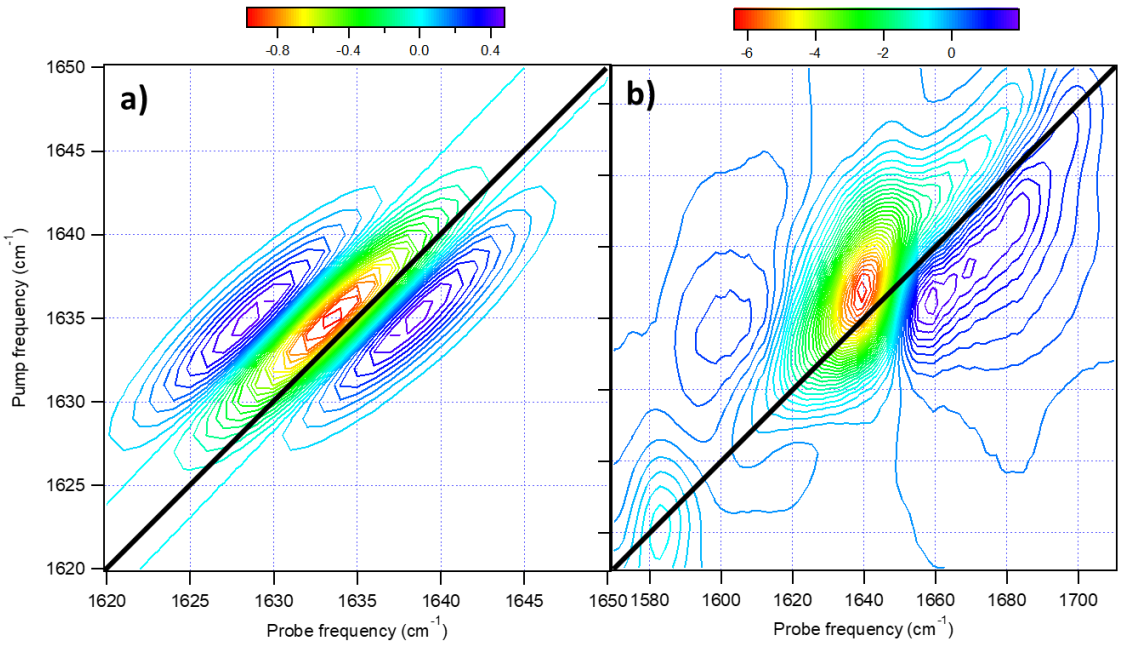


Figure 5-16. a) Simulated difference spectrum between two 2D-IR spectra with peaks at different pump frequencies, and b) the PC2 2D-IR spectrum.

PC2 of the 2D-IR diagonals accounts for 0.19% of the variance in the data. The spectrum comprises of a positive peak at 1640 cm^{-1} and a negative peak at 1660 cm^{-1} (Figure 5-15 (c)). The trend for the PC2 of the diagonals does not suggest a difference between *apo*-CaM and Ca^{2+} bound, unlike the FTIR and 2D-IR counterparts, however there does appear to be a shift with increasing TFP (Figure 5-15 (f)).

The positive peak in all three PC2 spectra, along with the associated PC2 loading plot, corresponds to a loss of α-helix structure, and the negative peak corresponds to a gain in random coil as the ratio of TFP increases. As the PC2 loadings for all of the Ca^{2+} bound 1:10 CaM:TFP is approximately the same as the *apo*-CaM 1:0 CaM:TFP PC loading, and it is known that *apo*-CaM has less α-helix structure than *holo*-CaM, it shows that with TFP binding to CaM, a loss of α-helix occurs. *Holo*-CaM bound fully to TFP has a similar amount of α-helix as *apo*-CaM does. This loss of α-helix can be seen in Figure 5-2 where binding of TFP causes CaM to close and become more compact with each TFP bound.

5.4.6 Principal Component three

A significant difference between this work and that of chapter (3) is there is a further PC from this dataset highlighting the complexity compared to just the two PCs which are present in chapter (3).

PC3 of the FTIR data accounts for 0.06% of the variance. This equates to a PC spectrum 90 times smaller in intensity than PC1. The spectrum is a series of positive (1630 cm^{-1}, 1645 cm^{-1}, and 1664 cm^{-1}) and negative peaks (1621 cm^{-1}, 1637 cm^{-1}, 1652 cm^{-1}, and 1682 cm^{-1}) (Figure 5-17 (a)). This suggests either red shifting or blue shifting of components underlying the amide I band; however, it is not possible to determine which direction shifting is occurring as there is approximately even spacing of seven or eight wavenumbers between the peaks.

The trends for the PC3 of the FTIR data show a steep increase in the PC loadings from 1:1 to 1:4 CaM:TFP for Ca^{2+} bound CaM, then a levelling off to 1:10 CaM:TFP (Figure 5-17 (d)). In contrast, the *apo*-CaM trend is a constant small gradient up until 1:10 CaM:TFP. These differences in trends give a compelling argument that binding of TFP to *apo*-CaM and *holo*-CaM is different.

PC3 of the 2D-IR dataset accounts for 0.11% of the variance. The spectrum comprises of a series of negative and positive peaks along the diagonal (Figure 5-17 (b)) with the 1640 cm^{-1} (positive) and the 1660 cm^{-1} (negative) peaks notably more intense than the other peaks present along the diagonal. Although PC3 for both FTIR and 2D-IR accounts for only a small portion of the data, confidence in the results can be taken as both comprise of the same spectral features.

The trends for PC3 of the 2D-IR data are again noisier than the FTIR PC3 trend and the clear trend from 1:1 to 1:4 CaM:TFP is not present for the Ca^{2+} bound spectra (Figure 5-17 (e)).

Also, there is no gradual change for *apo*-CaM, it appears there is very little change with the PC loadings remaining at approximately 0.2 across the different TFP ratios. The separation between *apo*-CaM and Ca^{2+} bound CaM is not as clear to see as in the FTIR PC3 trends.

PC3 of the 2D-IR diagonals accounts for 0.17% of variance and, again, the spectrum comprises of a series of positive and negative peaks (Figure 5-17 (c)). As with the whole 2D-IR PC3 spectrum, the 1640 cm^{-1} and 1660 cm^{-1} peaks are more intense than the rest. The clear trend from 1:1 to 1:4 CaM:TFP in the FTIR PC3 trends is not present again in the 2D-IR diagonal PC3, however, there is a clear separation between *apo*-CaM and Ca^{2+} bound CaM which was not present in whole 2D-IR analysis (Figure 5-17 (f)).

Between the FTIR and 2D-IR diagonal trends changes in the Ca^{2+} bound CaM from a 1:1 to 1:4 CaM:TFP ratio are observed, with the levelling off from 1:4 to 1:10 indicating fully bound CaM with four TFP molecules as previously suggested in the literature[14,19,21]. The continued rise of the *apo*-CaM from 1:1 to 1:10, albeit at a less steep gradient, suggests that binding occurs to *apo*-CaM across this TFP concentration range, as the PC3 spectral changes are present, but a larger quantity of TFP is required to result in the same binding effect.

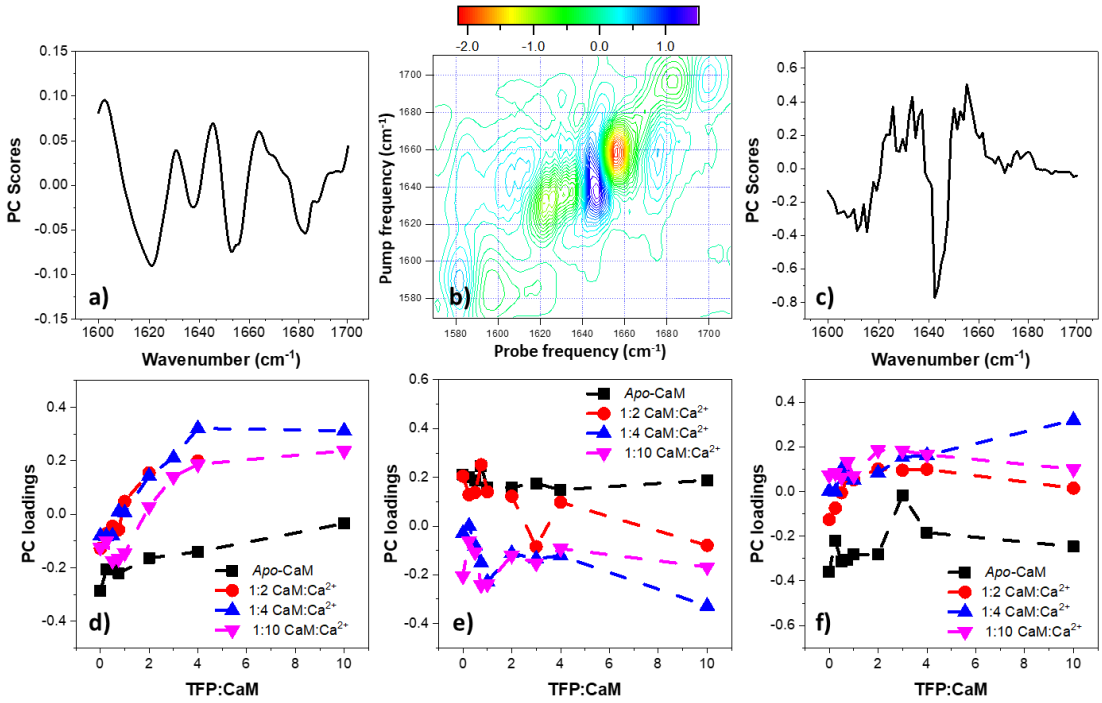


Figure 5-17. PC3 spectra for a) FTIR, b) 2D-IR and c) 2D-IR diagonals, and their associated trends, d) FTIR, e) 2D-IR and f) 2D-IR diagonals.

Although the 2D-IR trends do not provide distinction between binding as seen for the FTIR trends, 2D-IR is able to confirm that the spectral features observed are the result of binding through comparable PC3 spectral features and the presence of cross peaks.

A difference spectrum was calculated by subtracting the *apo*-CaM spectrum from the 1:4:4 CaM:Ca^{2+}:TFP 2D-IR spectrum (Figure 5-18 (a)). This difference spectrum was chosen as the PC3 trends suggest that there is a distinct difference between these two spectra; *apo*-CaM has a PC3 loading value of -0.2 and 1:4:4 has a value of 0.3. This resulted in a spectrum which comprised of the series of positive and negative peaks which are observed in PC3.

Two slices from the pump axis correlating to the frequencies of two PC3 peaks, 1621 cm^{-1} and 1664 cm^{-1} were examined (Figures 5-18 (b) and (c), respectively). Series of peaks were present across both slices associated with the spectral features observed in PC3 which confirm the presence of off-diagonal cross peaks in the 2D-IR difference spectrum (peaks indicated on the 2D-IR spectrum by *). Seven peaks were observed from the PC3 and the same seven frequencies were observed in both of the pump slices (including on-diagonal peaks). Cross peaks confirm that coupling has occurred between difference vibrational modes. Coupling arises when two chemical bonds are close enough that the vibrations from one mode influence the other and as such cause the second mode to vibrate. This suggests that the features observed in PC3 are the result of changes due to binding and not an artefact of random noise, which is further evident by the series of peaks being present in other difference spectra between multiple raw spectra (Appendix 2). Thus, the next step shall be assigning the peaks.

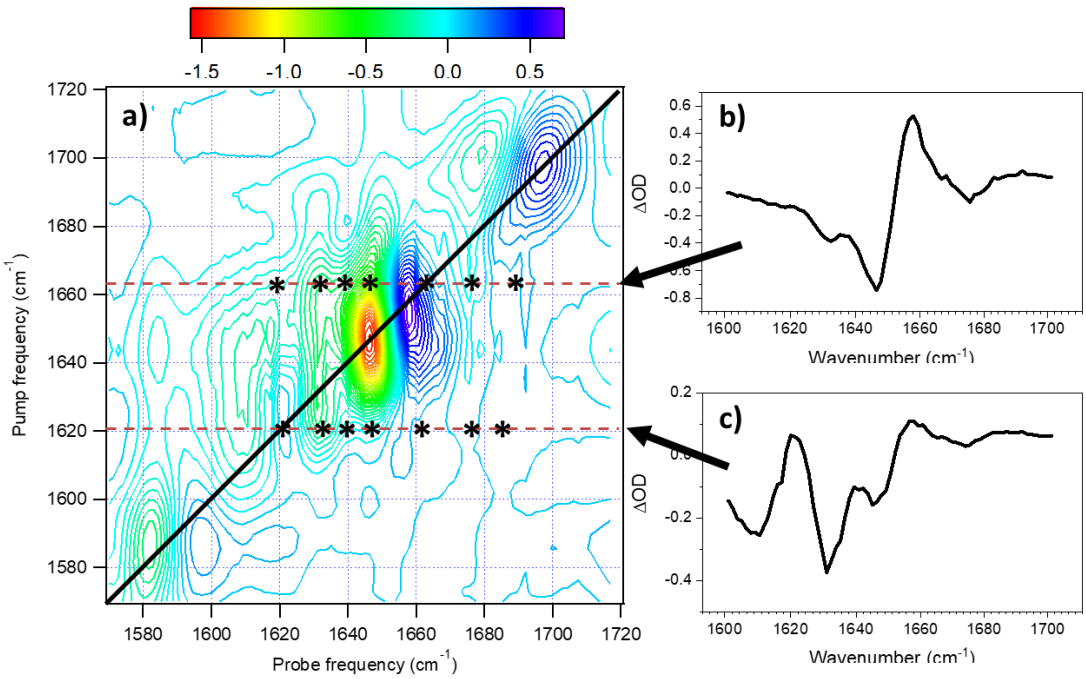
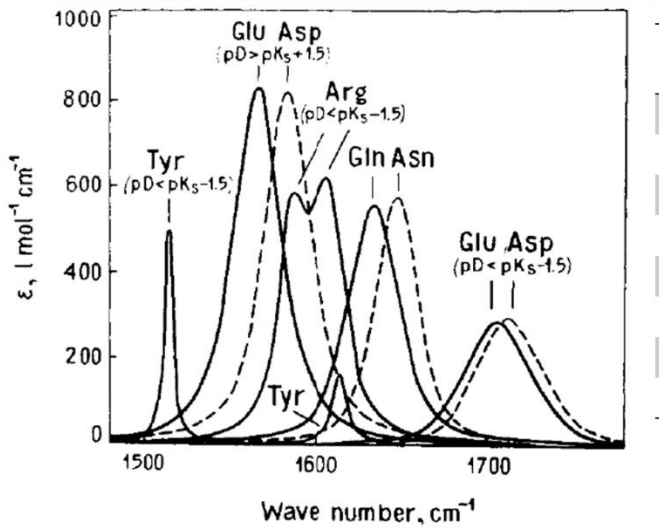


Figure. 5-18. 2D-IR difference spectrum of 1:4:4 (CaM:Ca^{2+}:TFP) minus apo-CaM. Black dots indicating small cross peaks present at frequency of electrostatic peaks.

These series of peaks in the PC3 spectra have been ruled out as individual amino acids changes due to binding of TFP. This has been determined as only two amino acids have side-chain absorption in the 1600 to 1700 cm^{-1} region in deuterated solvents which correspond to the wavenumbers of interest. These are glutamine (Gln) and Asparagine (Asn)[37,38] (Figure 5-19). There is a total of twelve of these amino acids in CaM which comprise 8% of the sequence (Figure 5-19, 12 out of 148 residues). However, only one Gln residue is involved in TFP binding (Figure 5-19) and therefore this series of seven peaks from PC3 are unlikely to be the result of changes in one residue which only has side-chain absorption between 1635 cm^{-1} and 1654 cm^{-1}.

ADQLTEEQIAEFKEAFSLFDKDGDGTITTKELGTVMRSLGQNPTEAELQDMINEVDADGNGTIDFPEFL
TMMARKMKDTDSEEEIREAFRVFDKDGNGYISAAELRHVMTNLGEKLTDEEVDEMIREADIDGDGQ
VNYEEFVQMMTAK



Amino acid involved in TFP binding	No. of residues
Glu (E)	6
Met (M)	6
Leu (L)	2
Ile (I)	2
Val (V)	2
Phe (F)	2
Ala (A)	1
Gln (Q)	1

Figure 5-19. Amino acid sequence of CaM (top). Residues in red have side-chain absorption in the frequencies of the PC3 peaks, with the bold residue involved in binding of TFP (only the second Gln in the sequence). Spectrum (left) from[37] showing amino acids with side chain absorption in the amide I region. Table (right) listing the amino acids in CaM involved in TFP binding.

These types of peaks have been determined to be a result of changes in the protein structure from the reorientation of the secondary structures present in CaM to accommodate further TFP molecules. As it is a result of binding, and not just structure change, there is a bigger difference between the FTIR trends of *apo*-CaM and Ca^{2+} bound CaM than from the difference in trends observed in PC2.

For both the 2D-IR and 2D-IR diagonals PC3 intense peaks at approximately 1640 cm^{-1} and 1660 cm^{-1} are observed (Figure 5-17 (b) and (c)). These can be assigned to α-helix and random coil peaks, respectively, which show a change in structure with binding of both Ca^{2+} and TFP to CaM.

It is important, therefore, to have access to the whole 2D-IR spectrum for confirming details via cross peak analysis and analyse both the whole 2D-IR spectrum and its diagonal to confirm that structural changes are occurring. To get the complete understanding of the binding effect of the Ca^{2+} and TFP to CaM, analysis of both 2D-IR and FTIR is required, this follows from one of the conclusions from chapter (3) that utilises both linear IR and 2D-IR in calculating the relative transition dipole spectra for determining the structural change occurring.

5.5 Conclusions and further work

The effects of binding different ratios of Ca^{2+} and TFP to CaM on the structure of CaM has been investigated by both linear IR spectroscopy and 2D-IR spectroscopy, utilising PCA as a method to elucidate the structural changes from the large dataset. 2D-IR was first used to analyse the impact of binding different ratios of Ca^{2+} to CaM would have on the structure of CaM as it is known to occur in a two-step cooperative process. The results built upon those from Chapter (3) where distinct differences between *apo*-CaM and *holo*-CaM (Ca^{2+} saturated CaM) were detected in the structure and stability. There was little difference between the ratios 1:2, 1:4 and 1:10 CaM:Ca^{2+} due to the fact that the more flexible domain, C, binds two Ca^{2+} first allowing this domain to become rigid, and the central linker becomes fully helical which changes the tertiary structure from closed to open – exposing potential binding sites for other molecules. Binding of two further Ca^{2+} to the N-domain does not induce as large of a change as the N-domain was more rigid in the *apo*-CaM form, as indicated from MD simulations. As a maximum of four Ca^{2+} can bind to an equivalent one CaM, there is little difference from binding four Ca^{2+} as opposed to saturating with ten Ca^{2+} ions.

Different ratios of Ca^{2+} and TFP bound to CaM were investigated with FTIR and 2D-IR spectroscopy. The large datasets were analysed using PCA and two binding dependent PCs were extracted. FTIR measurements were able to be separated into distinct binding groups for both calcium and TFP binding. 2D-IR and their diagonals produced less separations, with only distinguishing between calcium bound and unbound spectra, however, were able to provide evidence for the changes observed.

The PCA spectral features and trends were examined for each PC and comparisons were made between FTIR, 2D-IR and 2D-IR diagonals. Two distinct changes with binding of TFP were established in PC2 and PC3. The largest of the two changes was a loss of α-helix and a gain of random coil with the addition of each equivalent TFP up to four equivalent TFP for the Ca^{2+} bound CaM. There was little change with less than one TFP bound. The second, smaller change detected with TFP binding were changes in the overall structure of CaM possibly due to reorientation to accommodate further TFP molecules, again up to four equivalent TFP for Ca^{2+} bound CaM. This confirms the maximum number of Ca^{2+} that can bind to *holo*-CaM which had been an area of uncertainty in previous publications. Binding to *apo*-CaM was also confirmed, though as the trend differs from that of the Ca^{2+} bound CaM data

it confirms that the binding is performed by a different mechanism and suggests that the binding affinity of TFP to *apo*-CaM is lower compared to Ca^{2+}.

PC3 which describes the changes in overall structure only contributed to a small amount of variation in the data, however 2D-IR cross-peak examination confirmed that the spectral features were real structural changes and not the result of noise in the spectra. Therefore, the best approach of extracting maximum information of the protein-drug binding has been shown to be with a combination of FTIR and 2D-IR spectroscopy.

5.6 Technical Note

The 2D-IR spectra presented in this chapter were originally performed using an interferometer approach which was utilised in chapter (3). However, issues arose in the data quality when trying to analyse the small spectral differences that were present with TFP binding to CaM. These issues and the reason why the implementation of a pulse shaper was required will be discussed in this technical note.

The same set of 36 measurements were performed using a 2D-IR interferometer approach, prior to the pulse shaper setup. The data was processed via the PCA method and three distinct PCs were determined (99.54% of total data). As with previous results, PC1 was the average spectrum of all measurements and accounted for 98.90% of variance and was independent of concentration changes (Figure 5-20(a) and (d)). PC2 consisted of a series of positive and negative peaks on the anti-diagonal. This PC accounted for 0.49% of the variance (Figure 5-20 (b) and (e)). In comparison to the pulse shaper data, the trends for PC2 were varied and showed no pattern from the measurements. PC3 consisted of two peaks on the diagonal, a positive peak at approximately 1640 cm^{-1}, attributable to a change in α-helix structure, and a negative peak at approximately 1660 cm^{-1}, attributable to a change in random coil structure (Figure 5-20 (c) and (f)). This PC3 accounted for 0.16% of the variance. The trend for PC3 in comparison to PC2 follows more of a pattern to the measurements, with the same levelling off trend for the Ca^{2+} bound CaM datasets at a CaM:TFP ratio of 1:4 that was seen previously in the FTIR data and the pulse shaper data.

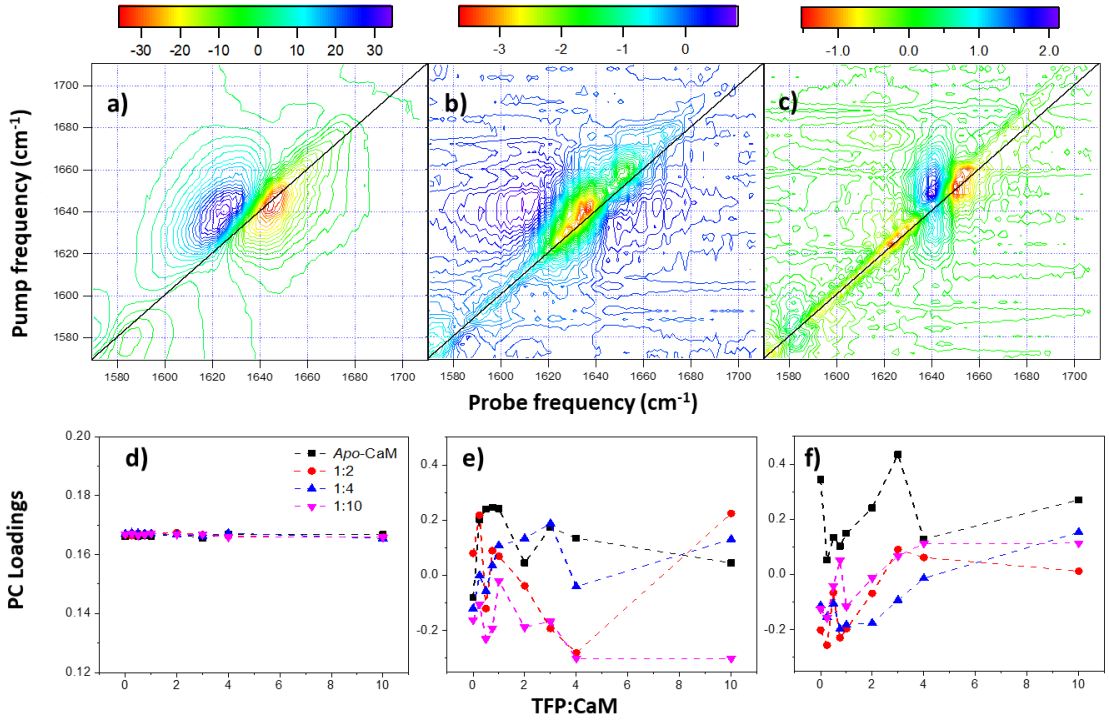


Figure 5-20. PCA results of 2D-IR data collected using an interferometric approach with PC spectra (top) and trends (bottom) for PC1 (left), PC2 (middle) and PC3 (right).

The spectra produced from PCA were noisy and the trends produced do not follow the same pattern as FTIR, especially the PC2 trend. The PC2 accounted for more variance than PC3 but appears to be a random trend that does not fit with the measurements. One of the benefits of PCA is its ability to pick out trends from random noise, therefore it is assumed that the PC2 spectrum and trend are not the result of random noise. To further develop the accuracy of the data, it is important to understand where this information is produced from.

Very small changes in the protein are now being detected with this dataset – electrostatic properties are small changes therefore any inaccuracies in the processing of the data will impact the results that are acquired through PCA.

The interferometric method used involves a processing step of "phasing" all the spectra manually using the projection slice theorem[40,41]. It is stated that the phase should be accurate to 0.2 to 0.4 fs[42], i.e. time zero must be determined to a high accuracy in order to get the correct resulting 2D-IR spectrum. The method used to phase the interferometric data is performed in a way that time zero is accurate to below 0.4 fs, therefore within the recommended window. However, the differences that are currently being extracted are so small that an accuracy of 0.2 to 0.4 fs for time zero is not enough.

A phasing test was performed on one spectrum where it was purposely phased incorrectly by a small amount three times. These three spectra plus the predicted correctly phased spectrum were analysed with PCA (Figure 5-21). The resulting PC1 spectrum was the average spectrum of the four spectra as expected, however the PC2 spectrum was a series of three peaks on the anti-diagonal, similar to the PC2 spectrum from the interferometer TFP dataset. This suggests that there are errors in the phasing of this TFP dataset.

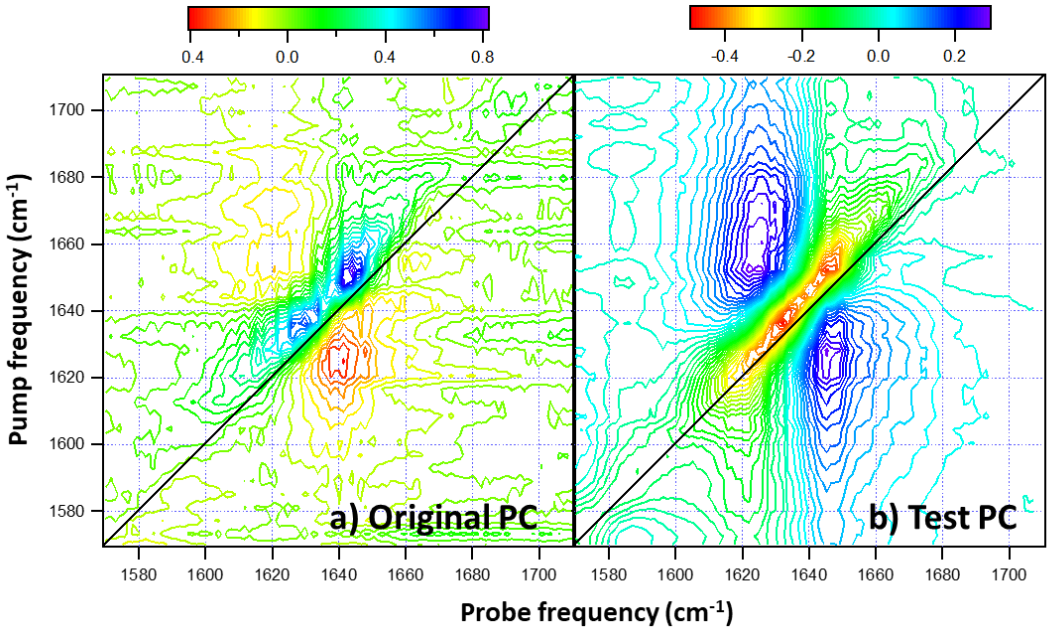


Figure 5-21. Phasing test comparison of "real" dataset (PC3 spectrum from CaM:Ca^{2+} 1:4, left) and the generated spectrum from the test data (right).

There is difficulty is producing a more accurate time zero manually as the projection slice theorem does not always have a “clean” spectrum to compare against a predicted spectrum (Figure 5-22). This results in inaccuracies between the predicted spectrum and the experimental spectrum. Therefore, a secondary check is made by examining the resulting 2D-IR spectrum by eye. However, it is difficult to determine a more accurate time zero below 0.4 fs by eye as all the spectra at this level appear to be phased correctly.

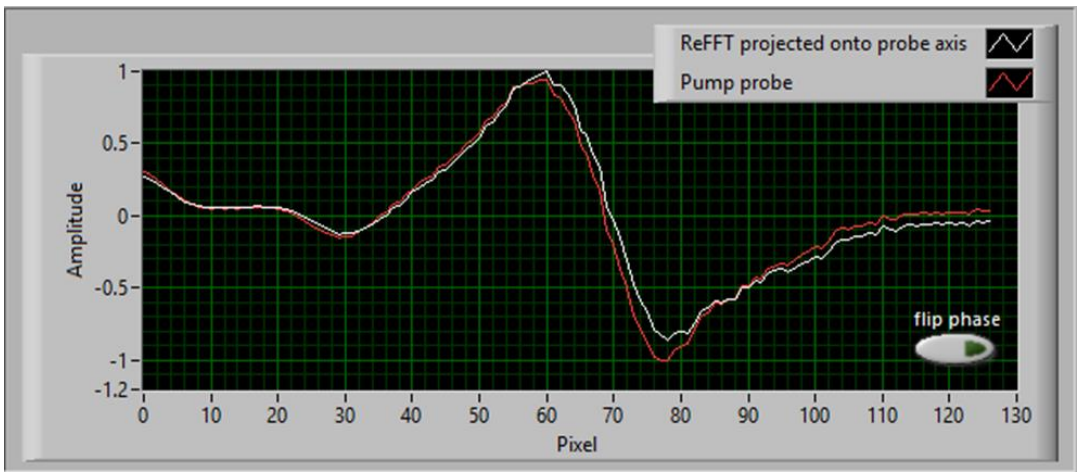


Figure 5-22. Display from data analysis showing projected slice (white) against experimental data (red).

A method has been developed using PCA to reduce these discrepancies in phasing. The absolute correct phase cannot be determined for the correct time zero, but a common phase can be determined to ensure all spectra are phased to the same degree. This has been done by "phasing" the data ± 0.05 rad and ± 0.1 rad from the predicted phase calculated by the projection slice theorem. This change is still within the range of 0.2 to 0.4 fs and all spectra appear to be correctly phased by visual inspection with no obvious distortion to the spectra.

This results in five plots of differently phased data for each of the 36 measurements, i.e. 180 spectra in total. PCA analysis was performed on these 180 spectra and the resulting PC2 spectrum showed a distinct phasing pattern (Figure 5-23).

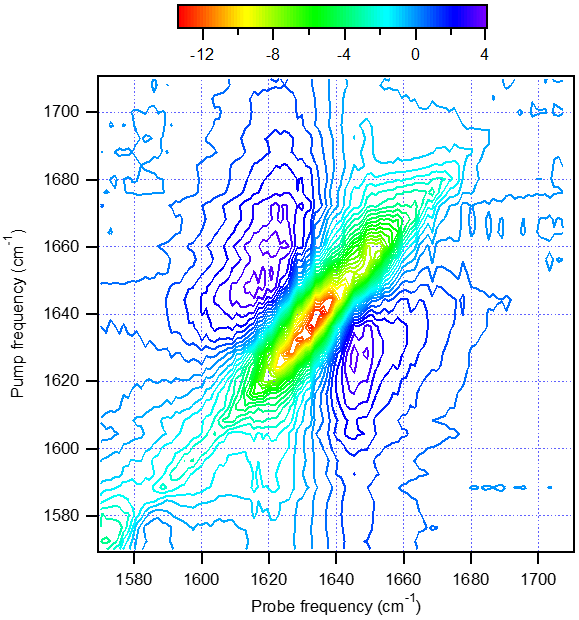


Figure 5-23. Resulting PC2 spectrum of PCA phasing elimination data.

Where the loading value is closest to zero for each of the 36 measurements out of the five "phasing" plots, i.e. the points closest to the horizontal line for each TFP ratio measurement (Figure 5-24), this is deemed the common phase for all spectra. These spectra selected from the zero loadings and are commonly phased for each measurement were processed via PCA.

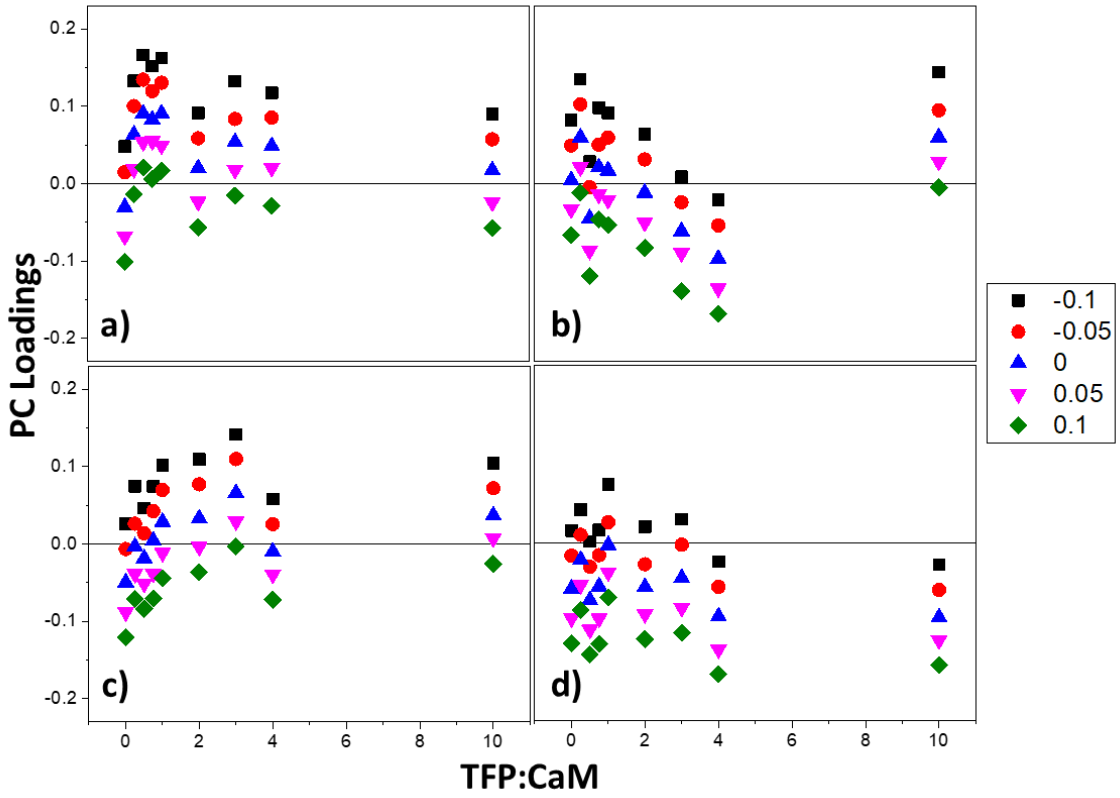


Figure 5-24. PC loadings of phasing data for each of the four datasets, a) apo-CaM, b) 1:2, c) 1:4, and d) 1:10. Horizontal line demonstrates point of common phase.

The resulting PC2 spectrum from this commonly phased dataset (Figure 5-25 (b)) shows a reduced phasing error compared to the previous PC2 of the interferometer 2D-IR data (Figure 5-25 (a)). The electrostatic peaks are now highlighted in this newly acquired PC2, however, the trends as still poorly resolved suggesting phasing issues may still be affecting the data (Figure 5-25 (c) showing old PC2 trend and (d) showing new PC2 trend). This method of finding a common phase via PCA is also time consuming and does not eliminate the error of not having the true time zero. It was for this reason that the decision was made to develop the pulse shaper 2D-IR method, as this eliminated the ambiguity of phasing. The pulse shaper

method also allowed for faster data acquisition (45 minutes versus 6 minutes per 2D-IR spectrum) which would be beneficial if a T-jump 2D-IR experiment was to be conducted.

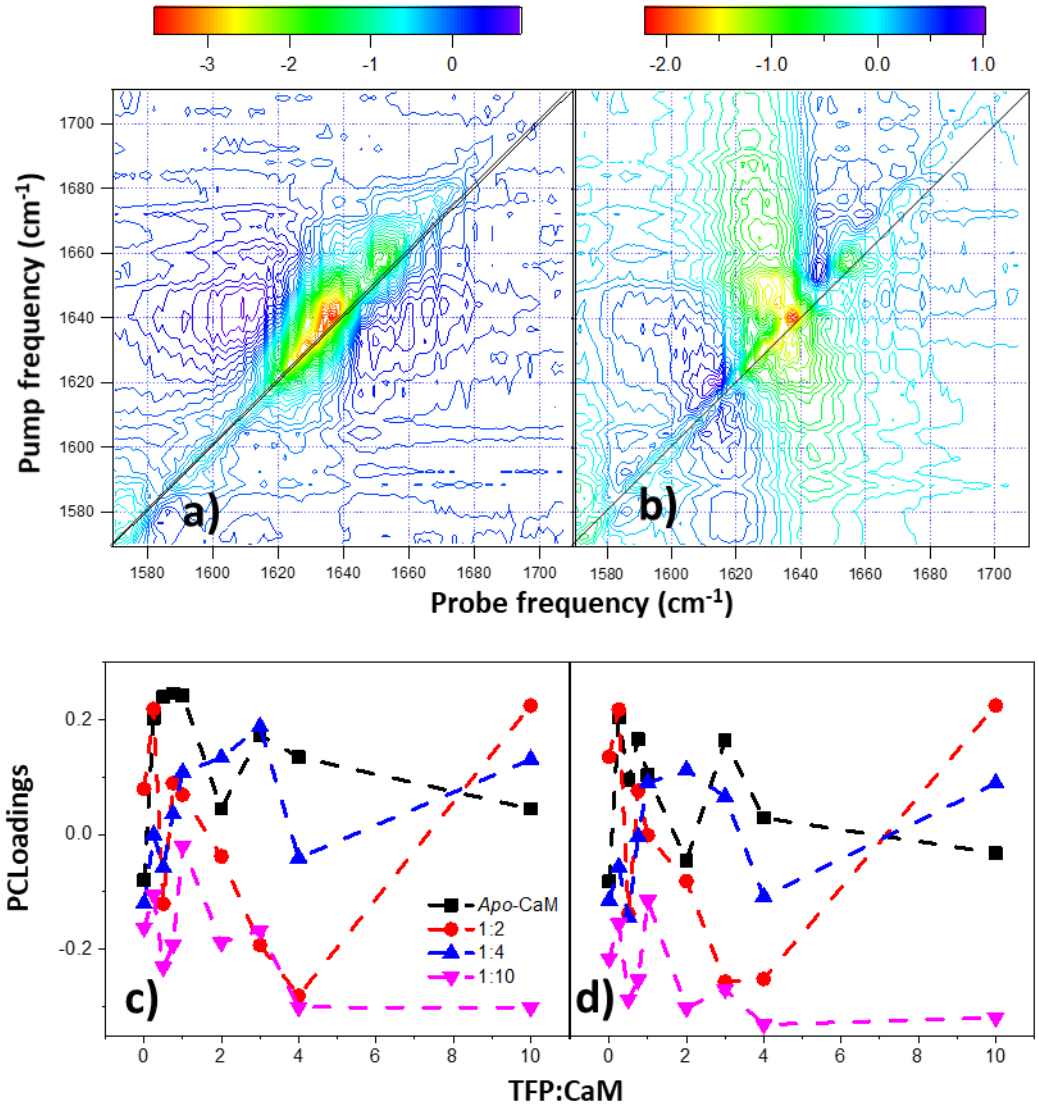


Figure 5-25. Comparison of PC2 results for interferometer data before (left) and after (right) phasing elimination.

5.7 References

(1) Vuignier, K.; Schappler, J.; Veuthey, J.; Carrupt, P.; Martel, S. Drug – Protein Binding : A Critical Review of Analytical Tools. *Anal. Bioanal. Chem.* **2010**, *398*, 53–66.

(2) Lampe, J. N. Advances in the Understanding of Protein-Protein Interactions in Drug Metabolizing Enzymes through the Use of Biophysical Techniques. *Front. Pharmacol.* **2017**, *8*, 1–13.

(3) Eyers, C. E.; Vonderach, M.; Ferries, S.; Jeacock, K.; Eyers, P. A. Understanding Protein – Drug Interactions Using Ion Mobility – Mass Spectrometry. *Curr. Opin. Chem. Biol.* **2018**, *42*, 167–176.

(4) Neault, J. F. Interaction of Cisplatin with Human Serum Albumin . Drug Binding Mode and Protein Secondary Structure. *Biochim Biophys Acta* **1998**, *1384*, 153–159.

(5) Xiao, J.; Shi, J.; Cao, H.; Wu, S. Analysis of Binding Interaction between Puerarin and Bovine Serum Albumin by Multi-Spectroscopic Method. *J. Pharm. Biomed. Anal.* **2007**, *45*, 609–615.

(6) Li, Y.; He, W.; Dong, Y.; Hu, Z. Human Serum Albumin Interaction with Formononetin Studied Using Fluorescence Anisotropy , FT-IR Spectroscopy , and Molecular Modeling Methods. *Bioorganic Med. Chem.* **2006**, *14*, 1431–1436.

(7) Liu, Y.; Xie, M.; Kang, J.; Zheng, D. Studies on the Interaction of Total Saponins of Panax Notoginseng and Human Serum Albumin by Fourier Transform Infrared Spectroscopy. **2003**, *59*, 2747–2758.

(8) Sowley, H.; Liu, Z.; Davies, J.; Peach, R.; Guo, R.; Sim, S.; Long, F.; Willison, K.; Zhuang, W.; Klug, D. R. Detection of Drug Binding to a Target Protein Using EVV 2DIR Spectroscopy. *J. Phys. Chem. B* **2019**, *123*, 3598–3606.

(9) Kuroda, D. G.; Bauman, J. D.; Challa, J. R.; Patel, D.; Troxler, T.; Das, K.; Arnold, E.; Hochstrasser, R. M. Snapshot of the Equilibrium Dynamics of a Drug Bound to HIV-1 Reverse Transcriptase. *Nat. Chem.* **2013**, *5* (3), 174–181.

(10) Shaw, D. J.; Hill, R. E.; Simpson, N.; Husseini, F. S.; Robb, K.; Greetham, G. M.; Towrie, M.; Parker, A. W.; Robinson, D.; Hirst, J. D.; et al. Examining the Role of Protein Structural Dynamics in Drug Resistance in: Mycobacterium Tuberculosis. *Chem. Sci.* **2017**, *8*, 8384–8399.

(11) Harmat, V. The Structure of the Complex of Calmodulin with KAR-2. *J. Biol. Chem.* **2005**, *280* (9), 8266–8274.

(12) Prozialeck, W. C.; Wallace, T. L. Commentary Interaction of Drugs with Calmodulin. *Biochem. Pharmacol.* **1982**, *31* (13), 2217–2226.

(13) Klee, C. B.; Crouch, T. H.; Richman, P. G. Calmodulin. *Annu. Rev. Biochem.* **1980**, *49*,

489–515.

(14) Matsushima, N.; Hayashi, N.; Jinbo, Y.; Izumi, Y. Ca2+-Bound Calmodulin Forms a Compact Globular Structure on Binding Four Trifluoperazine Molecules in Solution. *Biochem. J.* **2000**, *347*, 211–215.

(15) Matsushima, N.; Hayashi, N.; Watanabe, N. Binding of Trifluoperazine to Apocalmodulin Revealed by a Combination of Small-Angle X-Ray Scattering and Nuclear Magnetic Resonance. *J. Appl. Crystallogr.* **2007**, *40*, 179–183.

(16) Yamaotsu, N.; Suga, M.; Hirono, S. Molecular Dynamics Simulation of the Calmodulin–Trifluoperazine Complex in Aqueous Solution. *Biopolymers* **2001**, *58*, 410–421.

(17) Tanokura, M.; Yamada, K. Effects of Trifluoperazine on Calcium Binding by Calmodulin. *J. Biol. Chem.* **1986**, *261* (23), 10749–10752.

(18) Cook, W. J.; Walter, L. J.; Walter, M. R. Drug Binding by Calmodulin: Crystal Structure of a Calmodulin-Trifluoperazine Complex. *Biochem. Pharmacol.* **1994**, *33*, 15259–15265.

(19) Feldkamp, M. D.; Donnell, S. E. O.; Yu, L.; Shea, M. A. Allosteric Effects of the Antipsychotic Drug Trifluoperazine on the Energetics of Calcium Binding by Calmodulin. *Proteins Struct. Funct. Bioinforma.* **2010**, *78*, 2265–2282.

(20) Kang, S.; Hong, J.; Lee, J. M.; Moon, H. E.; Jeon, B.; Yoon, N. A.; Paek, S. H.; Roh, E. J.; Lee, C. J.; Kang, S. S. Trifluoperazine, a Well-Known Antipsychotic, Inhibits Glioblastoma Invasion by Binding to Calmodulin and Disinhibiting Calcium Release Channel IP 3 R. *Mol. Cancer Ther.* **2017**, *16*, 217–227.

(21) Massom, L.; Lee, H.; Jarrett, H. W. Trifluoperazine Binding to Porcine Brain Calmodulin and Skeletal Muscle Troponin C. *Biochemistry* **1990**, *29*, 671–681.

(22) Gresh, N. Theoretical Studies of the Binding of Trifluoperazine Derivatives to Site (82-93) of Calmodulin : Effect of Lengthenings of the Methylene Linker Chain on the Binding Affinity. *Mol. Pharmacol.* **1987**, *31*, 617–622.

(23) Levin, M.; Weiss, B. Binding of Trifluoperazine to the Calcium-Dependent Activator of Cyclic Nucleotide Phosphodiesterase. *Mol. Pharmacol.* **1977**, *13*, 690–697.

(24) Shim, S.-H.; Strasfeld, D. B.; Ling, Y. L.; Zanni, M. T. Automated 2D IR Spectroscopy Using a Mid-IR Pulse Shaper and Application of This Technology to the Human Islet Amyloid Polypeptide. *Proc Natl Acad Sci U S A* **2007**, *104* (36), 14197–14202.

(25) Vandonselaar, M.; Hickie, R. A.; Quail, W.; Delbaere, L. T. J. Trifluoperazine-Induced Conformational Change in Ca(2+)-Calmodulin. *Nat. Struct. Mol. Biol.* **1994**, *11*, 795–801.

(26) Vertessy, B. G.; Harmat, V.; Bo, Z. Simultaneous Binding of Drugs with Different Chemical Structures to Ca 2 + -Calmodulin : Crystallographic and Spectroscopic Studies. *Biochemistry* **1998**, *37*, 15300–15310.

(27) Kuboniwa, H.; Tjandra, N.; Grzesiek, S.; Ren, H.; Klee, C. B.; Bax, A. Solution Structure of Calcium-Free Calmodulin. *Nat. Struct. Mol. Biol.* **1995**, *2*, 768–776.

(28) Chattopadhyaya, R.; Meador, W. E.; Means, A. R.; Quiocho, F. A. Calmodulin Structure Refined at 1.7 Å Resolution. *J. Mol. Biol.* **1992**, *228*, 1177–1192.

(29) Sengupta, P.; Tebar, F.; Golebiewska, U.; Zaitseva, I.; Enrich, C.; Mclaughlin, S.; Villalobo, A. Bind to Membranes, Changing the Electrostatic Surface Potential. *J. Biol. Chem.* **2007**, *282* (11), 8474–8486.

(30) Cox, J. A. Interactive Properties of Calmodulin. *Biochem. J.* **1988**, *249* (3), 621–629.

(31) Minnes, L.; Shaw, D. J.; Cossins, B. P.; Donaldson, P. M.; Greetham, G. M.; Towrie, M.; Parker, A. W.; Baker, M. J.; Henry, A. J.; Taylor, R. J.; et al. Quantifying Secondary Structure Changes in Calmodulin Using 2D-IR Spectroscopy. *Anal. Chem.* **2017**, *89*, 10898–10906.

(32) Klevit, R E, Levine, B A, Williams, R. J. P. A Study of Calmodulin and Its Interaction with Trifluoperazine by High Resolution 1H NMR Spectroscopy. *FEBS Lett.* **1981**, *123*, 25–29.

(33) Jackson, A. E.; Puett, D. Binding of Trifluoperazine and Fluorene-Containing Compounds to Calmodulin and Adducts. *Biochem. Pharmacol.* **1986**, *35*, 4395–4400.

(34) Greetham, G. M.; Burgos, P.; Qian, C.; Clark, I. P.; Codd, P. S.; Farrow, R. C.; George, M. W.; Kogimtzis, M.; Matousek, P.; Parker, A. W.; et al. ULTRA: A Unique Instrument for Time-Resolved Spectroscopy. *Appl. Spectrosc.* **2010**, *64* (12), 1311–

1319.

(35) Linse, S.; Helmersson, A.; Forsen, S. Calcium Binding to Calmodulin and Its Globular Domains. *J. Biol. Chem.* **1991**, *266*, 8050–8054.

(36) Yu, T.; Wu, G.; Yang, H.; Wang, J.; Yu, S. Calcium-Dependent Conformational Transition of Calmodulin Determined by Fourier Transform Infrared Spectroscopy. *Int. J. Biol. Macromol.* **2013**, *56*, 57–61.

(37) Chirgadze, Y. N.; Fedorov, O. V.; Trushina, N. P. Estimation of Amino Acid Residue Side-chain Absorption in the Infrared Spectra of Protein Solutions in Heavy Water. *Biopolymers* **1975**, *14*, 679–694.

(38) Barth, A.; Zscherp, C. What Vibrations Tell about Proteins. *Q. Rev. Biophys.* **2002**, *35*, 369–430.

(39) Dupeyrat, F.; Vidaud, C.; Lorphelin, A.; Berthomieu, C. Long Distance Charge Redistribution upon Cu,Zn-Superoxide Dismutase Reduction: Significance for Dismutase Function. *J. Biol. Chem.* **2004**, *279* (46), 48091–48101.

(40) Tokmakoff, A. Coherent 2D IR Spectroscopy : Molecular Structure and Dynamics in Solution. **2003**, *107*, 5258–5279.

(41) Asbury, J. B.; Steinel, T.; Stromberg, C.; Gaffney, K. J.; Piletic, I. R.; Fayer, M. D. Hydrogen Bond Breaking Probed with Multidimensional Stimulated Vibrational Echo Correlation Spectroscopy. *J. Chem. Phys.* **2003**, *119*, 12981–12997.

(42) Hamm, P.; Zanni, M. *Concepts and Methods of 2D Infrared Spectroscopy*; Cambridge University Press: Cambridge, U.K., 2011.

6 Conclusions and Further Work

In chapter (1) 2D-IR spectroscopy coupled to the multivariate analysis method PCA to extract changes in secondary structure of proteins upon perturbations, such as heating and ligand binding, was introduced. The importance of understanding the structure of proteins was presented along with the current methods of analysing secondary structure. Analysis using the amide I vibrational mode from IR spectroscopy and the subsequent data processing was reviewed with spectral features of the band encompassing information about a variety of secondary structures. Extracting the assignments from the convoluted band was deemed to be difficult and the technique of 2D-IR spectroscopy was introduced as a means of elucidating information by spreading the spectral contents onto two frequency axes, a detection axis and an excitation axis. The addition of 2D-IR spectroscopy's ability to measure ultrafast dynamics on the femtosecond to picosecond timescale was discussed as an advantage of the method over conventional secondary structure techniques such as CD spectroscopy and X-ray crystallography.

2D-IR spectra are complex plots which contain large amounts of information relating to the sample and subsequently data processing to extract information on structural features is desired. The multivariate analysis method of PCA has been proven to be an effective method of extracting trends in datasets from linear IR spectroscopy.

In chapter (3) thermal studies on the previously well-characterised protein, calmodulin, were investigated by 2D-IR spectroscopy with the use of PCA and results corroborated by other techniques. CaM is a calcium binding messenger protein that is predominately α-helical in structure with two globular domains separated by a flexible central linker. Upon binding up to four Ca^{2+} ions the structure becomes elongated and the flexible central linker rigidifies and adopts a fully helical structure. Previous studies have reported the inherent thermal stability of CaM, especially the *holo*-CaM state.

2D-IR in combination with PCA reproduces the temperature-dependent trends recovered from CD spectroscopy. The results revealed a situation in that changes in the 2D-IR signal caused by large changes in temperature included thermally induced alterations in coupling as well as secondary structure transitions. *Holo*-CaM was used as a reference to extract only information relating to secondary structure from the *apo*-CaM state. The 2D-IR results go beyond conventional linear IR spectroscopy whereby the signal is not as sensitive to

thermally induced relaxation of coupling within the α-helices as 2D-IR and CD spectroscopy. Combining the linear and 2D-IR measurements of *apo*- and *holo*-CaM via the transition dipole approach and cross-referenced with DSC data suggested that the derived lower values for the secondary structure change are likely to be closer to the actual values than numbers derived independently from CD or 2D-IR obtained without comparing to a reference sample.

The structural change between *apo*- and *holo*-CaM was also probed by 2D-IR spectroscopy. Without the thermal effects needing to be accounted for, an accurate change of a few percent in α-helices was determined, proving 2D-IR spectroscopy to be a sensitive technique for structural change determination.

In chapter (4), following on from the equilibrium measurements from chapter (3), the non-equilibrium dynamics of CaM were probed with the development of a T-jump TRIR experiment which was capable of measuring the initial processes of protein unfolding. The system was calibrated with TFA and the temperature rise was determined to be 9 °C. As with chapter (3), the use of a reference sample (*holo*-CaM) as a control to measure the thermally induced changes in H-bonding without a secondary structure transition being present was found to be of benefit. This allowed extraction of domain melting timescales beginning on microsecond timescales with α-helix destabilisation. These observations enabled the assignment of previously reported dynamics of CaM on hundreds of microsecond timescales to thermally activated melting. Therefore, a complete mechanism for thermal unfolding of CaM was produced.

This setup is different from previous T-jump TRIR experiments as it employs a pump laser chopped at 500 Hz allowing for quicker acquisition of data. This would provide the basis for developing a 2D-IR T-jump setup in the future. Although, future setups would benefit with a pulse shaper approach as opposed to the interferometer approach utilised in chapter (3) due to the long acquisition times required for 2D-IR T-jump.

In chapter (5) FTIR and 2D-IR measurements using a pulse shaper were obtained. CaM was bound to a series of different Ca^{2+} and TFP ratios creating large datasets. PCA was able to effectively separate the FTIR results into distinguishable group of different types of binding to CaM. The 2D-IR whole spectra analysis and the diagonals did not provide separation to the same degree, with only the calcium bound and unbound spectra split. However, the spectral features of 2D-IR were able to provide evidence for the structural changes that were observed with binding. With TFP binding and calcium binding a loss of α-helix structure was

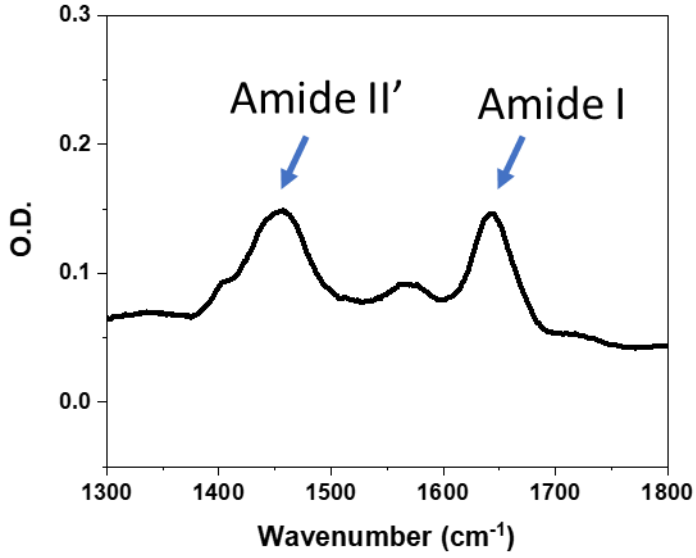
seen and a change in the electrostatic properties of the protein. Therefore, the complete picture on protein binding was achieved by utilising both FTIR and 2D-IR spectroscopy.

The changes observed in chapter (5) were smaller than the large thermal transitions seen in chapter (3). The 2D-IR laser system was developed further during this chapter from an interferometric approach to the pulse shaper. This was a necessary development as the manual phasing of the interferometer approach limits the accurate detection of small changes in structure upon protein binding. This manual phasing issue was eliminated with the pulse shaper approach.

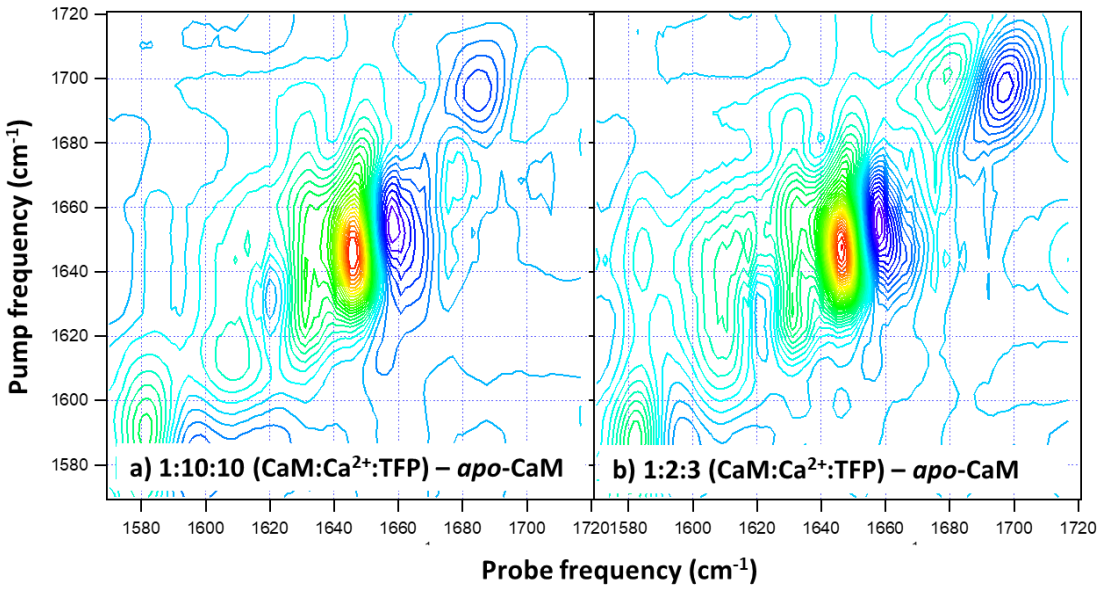
This updated pulse shaper system is still a developing laser system and further work in optimising the parameters to extract the best data is required. Factors of concern that can be optimised to influence the data include the acquisition times, the bandwidth, smoothing of data by different methods including apodization, and limiting the amount of atmospheric water in the system.

Overall this project has shown PCA coupled to 2D-IR spectroscopy to be an effective method for extracting and quantifying trends from protein datasets. Continued improvements on laser systems will increase the effectiveness of PCA on detecting small changes in protein structure. New laser setups have been developed through this project including a T-jump TRIR system that has been shown to uncover fast initial unfolding dynamics on the microsecond timescales. Future projects could include the development of a T-jump 2D-IR experiment for the unfolding of proteins by utilising the findings from this project.

Appendix



Appendix 126. Example spectrum of apo-CaM at 20 °C showing full range of amide II' to amide I.



Appendix 2. Additional difference spectra of CaM:Ca^{2+}:TFP ratios highlighting the series of peaks observed in PC3 are present in multiple difference spectra (Chapter 5.4.6). Difference spectra are a) 1:10:10 (CaM:Ca^{2+}:TFP) minus apo-CaM, and b) 1:2:3 (CaM:Ca^{2+}:TFP) minus apo-CaM.