Accessing Ultrafast Protein Dynamics through 2DIR Spectroscopy of Intrinsic Ligand Vibrations

By Niall Simpson

Department of Physics
University of Strathclyde

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

2015

‘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:

Table of Contents

Parts of this thesis are based on the following publications:

(i) Simpson, N.; Adamczyk, K.; Hithell, G.; Shaw, D. J.; Greetham, G. M.; Towrie, M.; Parker, A.W.; Hunt, N. T., *The effect on structural and solvent water molecules of substrate binding to ferric horseradish peroxidase*, Faraday Discussions, (2015)

(ii) Adamczyk, K.; Simpson, N.; Greetham, G. M.; Gumiero, A.; Walsh, M. A.; Towrie, M.; Parker, A. W.; Hunt, N. T., *Ultrafast infrared spectroscopy reveals water-mediated coherent dynamics in an enzyme active site*. Chemical Science, **6**, 505-516, (2015)

(iii) Simpson, N.; Shaw, D. J.; Frederix, P. W. J. M.; Gillies, A. H.; Adamczyk, K.; Greetham, G. M.; Towrie, M.; Parker, A. W.; Hoskisson, P. A.; Hunt, N. T., *Infrared spectroscopy of Nicotinamide Adenine Dinucleotides in One and Two Dimensions*, The Journal of Physical Chemistry B, 117, 16488-16478, (2013)

List of Abbreviations

2DIR – two-dimensional infrared spectroscopy

BHA – benzohydroxamic acid

CARS – coherent anti-Stokes Raman scattering

Cc – cytochrome c

CLS – central line slope

CT – charge transfer

CW – continuous wave

DFG – difference frequency generation

DFT – density functional theory

ESA – excited state absorption

FFCF – frequency-frequency correlation function

FID – free induction decay

FTIR – fourier transform infrared spectroscopy

FWHM – full width at half maximum

HPLC – high performance liquid chromatography

HRP – horseradish peroxidase

ITC – isothermal titration calorimetry

Mb – myoglobin

MCT – mercury cadmium telluride

MD – molecular dynamics

MIR – mid infrared

MS – mass spectrometry

NAD – nicotinamide adenine dinucleotide

NLS – nodal line slope

NMN - nicotinamide mononucleotide

NMR – nuclear magnetic resonance

OD – optical density

OPA – optical parametric amplifier

PDB – protein data bank

PTFE – polytetrafluoroethylene

T3SS – type three secretion system

TDC – transition dipole coupling

TPR – tetratricopeotide repeat

TRIR – time-resolved infrared

UV - ultraviolet

Vis – visible

Abstract

Proteins are complex molecular machines that facilitate the chemical reactions fundamental to life. Their functions are encoded in a linear sequence of amino acids, of which only 20 species are found in nature. Yet the functional and structural diversity accessible through these building blocks is vast.

Molecular and atomic-level protein studies have been crucial to our understanding of health and treatment of disease, with increasingly sophisticated experimental and computational methods continuing to provide new information with which to advance medicine. However, the requirement for more detailed understanding of proteins has risen through the emergence of multi-antibiotic-resistant bacteria and also through the potential to design synthetic proteins of novel function.

Paradigms of protein function have evolved significantly since early studies, though few all-encompassing descriptions have been proposed, owing to the complex, dynamic structures of these large biomolecules. Presently, the relationship between protein structural motions at different timescales appears to hold vital significance to the elusive aspects of biological mechanisms. No single measurement technique is capable of accessing the multitude of timescales over which protein motions occur, and thus concerted investigation is necessary.

Observation of dynamics at the femtosecond-picosecond timescale has only recently become possible through the development of new experimental techniques, allowing a new class of protein motions to be investigated.

In this thesis, the advanced technique of two-dimensional infrared spectroscopy (2DIR) is employed to study three biomolecular systems with implications to ubiquitous protein interactions. The aims of these investigations are, firstly, to demonstrate the suitability of 2DIR spectroscopy in gathering novel dynamic

information from biological systems that is not accessible *via* other methods, and secondly, to derive the potential physical significance of these dynamics as they relate to biological function.

A description of the underlying theory of 2DIR is presented in this Chapter, along with the considerations that must be made in the application of such a technique to complex biological case-studies.

In Chapter (2), descriptions are given for the experimental setups used to acquire infrared spectra, specifically, Fourier transform infrared (FTIR), pump-probe and 2DIR spectroscopies.

In Chapter (3) the catalytic-site dynamics of two closely-related haem proteins are each studied by monitoring the vibrational evolution of a nitric oxide (NO) probe molecule bound to the haem centre. A comparison of the active site dynamics is performed in order to correlate the observed differences with discrepancies between the protein reaction mechanisms.

Chapter (4) explores the potential of a coenzyme with high protein-binding promiscuity to serve as an intrinsic reporter of the dynamics that occur at substrate binding sites. Infrared analysis and categorisation of the free coenzyme molecule is performed in order to establish its effectiveness as a probe.

In Chapter (5), method-development strategies are proposed for the extraction of 2DIR data from large, complex protein-protein systems, with the objective of expanding the range of interactions on which 2DIR can effectively report. Both well-established and novel strategies are employed, and the potential and limitations of the technique are discussed in the context of these demanding case-studies.

Chapter (6) draws together conclusions and an overview of progress made and discusses future directions.

1. Introduction

1.1 Investigating biological structure

The structure and function of biological molecules have long been seen as correlated, resulting in the hypothesis that protein active-site structure are well-adapted to accommodate their respective substrates with high specificity[1]. The three-dimensional geometries of biomolecules have consequently been of vast interest to researchers for a number of decades[2].

X-ray crystallography[2–4] is the archetypal technique in the study of biological structures. In this method, high energy X-rays are diffracted by a crystallised protein, mapping electron density in the crystal. This complex diffraction pattern can subsequently be converted into an intuitive three-dimensional model that resolves protein secondary and tertiary structure, side chain species and orientation, ligands and structural solvent molecules to a spatial resolution of <1.5 Å (1.5×10^{-10} m); at which little error is observed in the positions of atoms[5]. Moreover, proteins can be crystallised in complex with biological and synthetic ligands, revealing key sites on the protein structure at which reactions are either performed or inhibited. Since the earliest successful application of X-ray crystallography, in 1958, this method has brought about major advances in medicine and biological structural theory[6]. For example, the determination of single protein structure such as insulin[7] has brought about tangible improvements to the prognoses of countless diabetes patients, and with a present total of ~102,400 protein structures catalogued in the Protein Data Bank, X-ray diffraction is arguably the most powerful tool in structural biology to date. Despite its success, the technique is not without limitations[8]. Crystallisation requires that proteins are removed from aqueous solution. When dissolved in the ubiquitous biological solvent, water, the three-dimensional structures of proteins may exhibit many types of motion over a broad range of timescales (Figure (1)). Consequently, the structural conformation measured by X-ray diffraction, following protein crystallisation is likely to reflect the average of many structural states that exist in aqueous solution[9].

The underlying flexibility or rigidity of protein scaffolds is also not immediately apparent from X-ray experiments[10]; a property of protein structure that is believed to hold importance to substrate binding. Early descriptions of binding were likened to a “key-lock” model where a substrate is geometrically congruent with a protein site, allowing it to bind[11]. Re-evaluation of the key-lock mechanism was soon prompted by the inability of this model to fully address enzyme specificity or justify observed enzyme kinetics[12]. The alternate, “induced-fit” model was subsequently proposed[13], suggesting that both protein and substrate are altered in the process of binding, allowing the ligand structure to more closely resemble its transition state, lowering the thermodynamic barrier for catalysis[14]. Soon after, an alternate explanation was offered in the form of the "conformation selection" model[15] that postulated that only specific protein sub-conformations are capable of ligand-binding[16]. Again, this latter proposal provides a more accurate thermodynamic model in many specific instances, though with the benefit of multiple studies, it appears that the applicability of “induced-fit” and “conformation selection” varies on a case-by-case basis, with some systems showing intermediate characteristics[17]. Fundamentally, each of these proposed mechanisms incorporates some degree of structural dynamics[18]. As X-ray crystallography collects only a single protein conformation, corresponding to the ensemble average, multiple X-ray experiments using multiple substrates are required to identify these more adaptive responses of protein architecture. The process of crystallisation is time-consuming, and can also require an approach of systematic trial and error[19,20]. As a result, the success of crystallisation has become a rate-limiting factor in X-ray crystallography, though a number of technological advances show potential in overcoming this constraint[21–26].

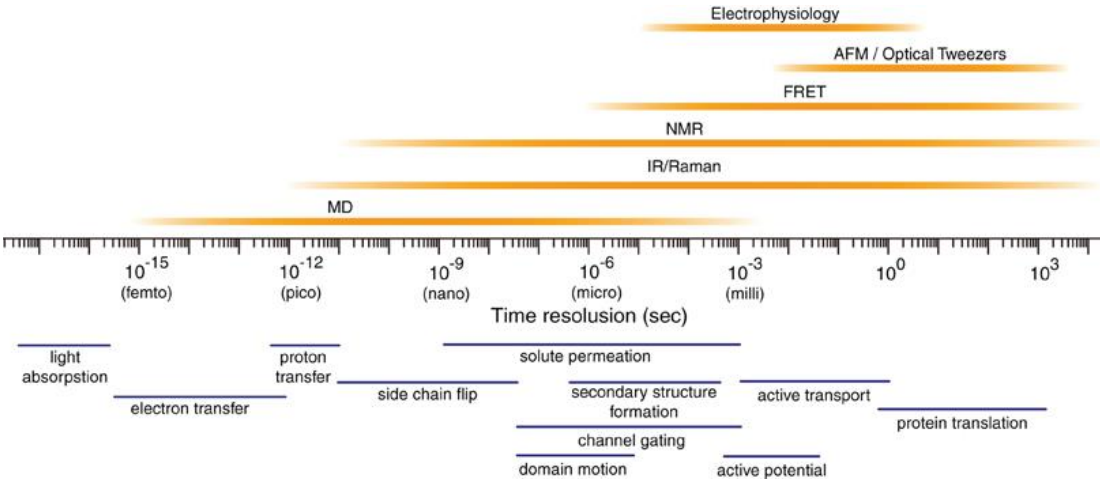


Figure (1): Various species of protein conformational motion and their corresponding dynamic timescales[36]*. These motions are best understood through the concerted application of both experimental techniques such as NMR, Förster resonance energy transfer (FRET) and computational techniques such as molecular dynamics (MD)*

Nuclear magnetic resonance (NMR) spectroscopy can provide additional detail on protein structure that is inaccessible to X-ray crystallography[27–29]. Using this method, proteins can be analysed in the solution-phase, dispensing with the difficult process of crystallisation. NMR scans across radio-wavelengths to detect the magnetic resonances of single nuclei, and can be "tuned" to the resonant frequencies of specific elements such as ^{1}H, ^{13}C, ^{15}N and ^{31}P which, in combination with variable pulse-sequences, can provide many types of information from protein structures. Atoms can be individually assigned, giving inherently site-specific information. Magnetic coupling information further yields the three-dimensional distribution of nuclei [30].

Time-resolved NMR studies complement the static structures produced by X-ray crystallography, as many biologically-important mechanisms are fundamentally dynamic[31,32], such as multi-stage enzymatic reactions, substrate recognition and side chain rotation (Figure (1)). NMR is capable of capturing dynamics in the range of ns-min[33], during which many biological processes such as protein unfolding[27], ligand binding[34] or redox reactions occur[35].

The detail available from nuclear resonance experiments was further expanded through the implementation of pulse sequences that allow selective excitation of nuclei at specific frequencies. With respect to conventional NMR spectroscopy, this selective excitation effectively labels a particular resonance, and with a second radio-frequency pulse, the interactions originating from the labelled resonance may be isolated[30]. Thus, by introducing an excitation pulse of controllable frequency, it is possible to transform a conventional, linear NMR spectrum into a two-dimensional map of the correlated nuclear resonances within a molecule. This technique, 2D-NMR, allowed for vastly more sophisticated analysis of biological systems.

Regardless, very large proteins ≥40 KDa may be challenging to analyse *via* NMR spectroscopies, particularly ^{1}H NMR methods, as spectra contain a vast quantity of data, such that peaks from individual atoms are no longer well-resolved[27]. Isotopic-labelling may be necessary to extract information from local protein regions, an approach that is often expensive to implement with full protein sequences. Further, though NMR has access to dynamics on a broad range of timescales, a subset of rapid protein motions lies beyond the temporal resolution of the method.

The ensemble of conformational states available to a protein includes non-native structures, and explanations of how proteins exploit thermally-driven motion to promote access to the functional, “native”, sub-states are elusive[37]. Significant changes in protein structure, such as folding or unfolding, involve relatively large thermodynamic barriers and resultantly occur over long timescales (seconds to hours[38]). Such motions are crucial to the adoption of the protein’s native state, defining structural domains. Within the sub-ensemble of folded states, proteins may explore a range of further conformations with different three-dimensional domain orientations. Domain motion is vital to substrate binding and protein-protein interactions[39,40], as it allows the sampling of various conformations; a particular ligand may selectively stabilise one of these conformations, while a different ligand stabilises an alternate configuration[41]. This interaction allows proteins to accommodate a range of substrates. Domain motions occur at timescales ~μs, though they have been seen to link to more rapid process, such as the collective motions of side chains[42]. Indeed, the binding of a substrate is in many cases observed to restrict the flexibility of side chains far from the substrate binding site[40]. Reciprocally, small, local changes in protein structure have been shown to contribute to global changes in flexibility[43–45], demonstrating the possible concerted action of side chain motion to produce large, global structural changes. Side chains may also operate in concerted fashion within local environments such as catalytic sites, where the concerted interactions of residues with a bound ligand give stabilises reaction intermediates throughout a multi-stage enzymatic process[46,47]. Side chain reorientations occur on the ps-ns timescale[48,49], and thus extend beyond the highest temporal resolution currently available to NMR spectroscopy. Additionally, many other dynamic processes accompany the functional roles of side chains in protein reactions, though the real-time observation of these events has paralleled the development of ultrafast laser systems. Vibrational motion, hydrogen bond switching and rapid side chain rearrangement are presently not well enough understood to incorporate into the paradigm of protein motion[50], though these dynamics may be linked to important events such as such as those that give access to a transition state or which underpin or initiate slower changes in molecular structure[51]. Further, the ubiquitous biological solvent, water, undergoes a diverse range of dynamics in the ultrafast (fs-ps) domain[52–54], and thus, the ability to understand the potential biological roles of so-called non-structural water has not been well categorised. Although computational techniques have allowed modelling of rapid molecular motions[36], the experimental observation of these dynamics has been dependent on the availability of femtosecond laser systems[55].

Two-dimensional infrared spectroscopy (2DIR) is a relatively recent technique that follows on from the highly successful method of 2D-NMR. In 2DIR, the principles of a 2D-NMR pulse sequence are applied using higher frequency infrared light[56,57]. The short wavelengths of mid-infrared (MIR) (~2.5 – 20 μm) light give access to far greater temporal resolutions than are obtainable with the radio-wavelengths (0.3 – 5 m) of NMR. While each technique exploits the resonance of electromagnetic radiation with chemical compounds, infrared experiments can simultaneously excite the resonances of many molecular functional groups within the same experiments, whereas multiple NMR measurements are required to observe different species of nuclei. Using infrared radiation, chemical bonds can be selectively excited into vibration and the susceptibility of the vibration to nearby structural reconfiguration manifests as deviations in the bond's oscillation frequency with time[58]. This allows for the detection of structural dynamics on the timescale of hydrogen bond reconfiguration[59], complementing the model of protein structure and dynamics developed my X-ray and NMR.

2DIR has shown significant potential in the field of structural biology from its first application, to simple peptide systems[60]. This study demonstrated the potential of 2DIR to improve upon the linear infrared technique of Fourier transform infrared spectroscopy (FTIR) in the same way that 2D-NMR improved upon its precursor technique. Additionally, it provided new information on the inherent delocalisation of vibrations in protein structure, and how these may be used to probe the conformations adopted by polypeptide backbones. Further investigations of coupled amino acid vibrations in peptide chains[61–63] have provided significant detail with which to model and interpret the IR spectra of complex proteins, and also developed strategies for the retrieval of site-specific information. Peptide systems were also employed to illustrate sensitivity of 2DIR to both local and global structural dynamics on sub-picosecond timescales[64]. This type of analysis was further exploited by the Fayer group in the first application of 2DIR to a haem protein, where the dephasing dynamics of a carbon monoxide ligand bound to the haem centre of myoglobin was used to report on dynamics of structural substates within the haem cavity[65]. Further investigations of the myoglobin system resolved transitions between the observed substates, allowing inference of the timescale for structural reorientation of the functionally-important distal histidine residue[66,67].

The ability to observe vibrational couplings, structural fluctuations and chemical exchange processes with picosecond time-resolution has established 2DIR as a powerful method in biological studies. To date, the technique has provided a wealth of new information on protein behaviours such as drug binding[68], substrate binding[69,70] and recognition[71], side chain motion[65,72,73], the formation of amyloid fibrils[74–76], triggered ligand dissociation[77], solvent interaction[78,79] and polypeptide conformation changes[61,80–82]. The coincidence of 2DIR timescales with those of molecular dynamics (MD) simulations can also provide further support to these experimental conclusions, or potentially assist the refinement of theoretical models[72,83].

As with X-ray crystallography and NMR spectroscopy, 2DIR faces challenges from the fundamental complexity of protein systems. Though for this relatively new technique, there remains significant scope for method development and refinement, with the ultimate objective of extracting site-specific information from native-state biological systems.

1.2 Infrared Spectroscopy

1.2.1 Molecular Vibrations

The balance of attractive and repulsive forces between two covalently-bonded atoms creates an equilibrium spatial separation between the atoms; the bond length. This equilibrium distance, r_0, is characterised by a minimum in the potential energy, V(r), of the system (Figure (2)). Small displacements of the atoms from r_0 are opposed by a restoring force, akin to the restoring force of a mechanical spring, and lead to oscillatory, vibrational motion[84]. With much larger displacements, the chemical bond can be broken. This is reflected in the potential energy well by the less sharply increasing value of V(r) at higher inter-atom separations than r_0. The potential well is therefore anharmonic.

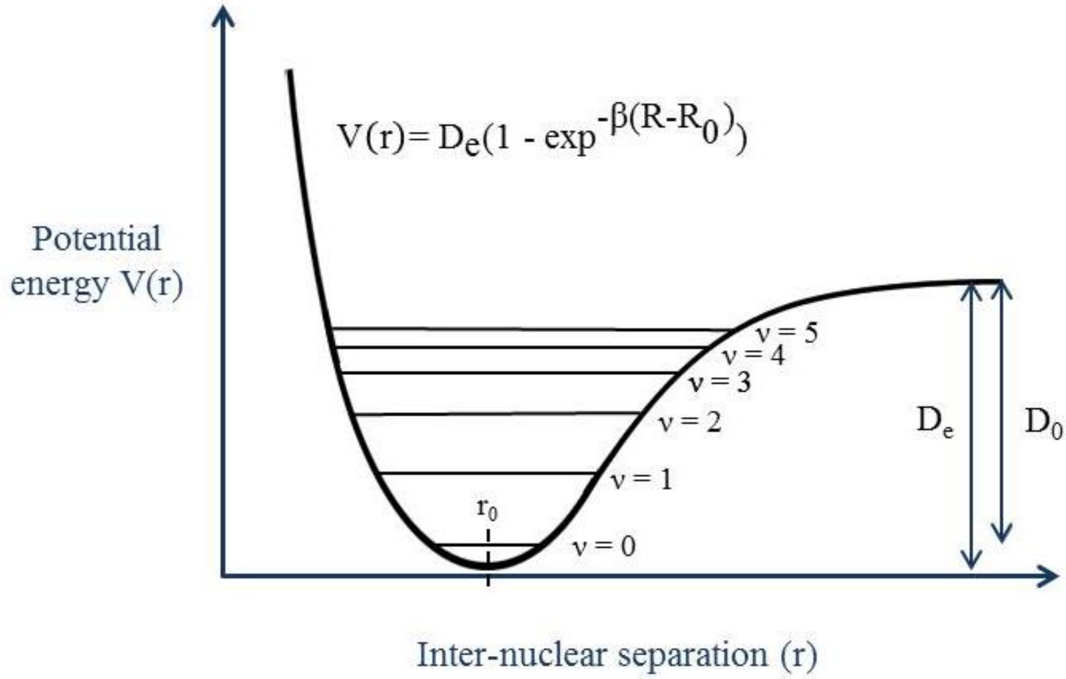


Figure (2): The Morse Potential for a diatomic molecule. Allowed potential energy states are calculated using the Born-Oppenheimer Approximation; the exclusion of nuclear kinetic energy. D_0 *indicates the dissociation energy of the bond and* D_e *indicates the well depth.*

In the case of atomic interactions, bond vibrations must be described quantum-mechanically, from which characteristically discretised potential energy levels are predicted. Excitation of a molecule from the $\nu = 0$ to $\nu = 1$ state therefore allows the inter-nuclear separation to deviate from r_0, inducing a vibration in the bond. Expressing the energy separation of two states in terms of frequency, it is apparent that vibrational transitions may be induced by electromagnetic radiation of corresponding frequency to the energy separation of $\nu = n$ and $\nu = n+1$;

$$E = hf = \frac{hc}{\lambda} = hc\cdot\tilde{\upsilon} \qquad (1)$$

where is the energy gap between vibrational levels, h is Planck’s Constant, f is the radiation frequency (in Hz), λ is the wavelength of the radiation and $\tilde{\upsilon}$ is the light frequency expressed in the manner conventional to IR spectroscopy, i.e. wavenumber

(cm^{-1}). The allowed potential energy levels of the well, E_v, with increasing values of ν, follows the series[84]:

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

where $\tilde{\upsilon}_e$ is the harmonic vibrational frequency and the terms x_e and y_e are inversely related to the dissociation energy, D_e. This gives rise to the concept of vibrational anharmonicity, in which bonds that are more favourably broken are subject to greater misshaping of their potential wells from a harmonic regime. This anharmonicity leads to smaller separations between successive energy levels with increasing ν.

The "fundamental" vibrational frequency of a bond ($\tilde{\upsilon}$ for the $\nu = 0\text{-}1$ transition) is influenced by the masses of its constituent atoms and by the electron distribution across the bond i.e. the molecular functional group[84]. For a diatomic molecule:

$$\tilde{\upsilon} = \frac{1}{2\pi c}\sqrt{\frac{f(m_1+m_2)}{(m_1 m_2)}} \qquad (3)$$

where c is the speed of light, m_1 and m_2 are the masses of atoms 1 and 2 and f is the force constant, reflecting the bond order. Applied to nitric oxide (NO), for which 1600 N/m[85], Equation (2) yields a vibrational frequency of 1905.8 cm^{-1}. Light at this frequency may therefore raise the molecule from the vibrational ground state ($\nu = 0$) to the first excited state ($\nu = 1$), causing oscillation of the bond at the fundamental frequency. Transitions to higher energy states may occur from $\nu = 0$,[86] though the probability of such a transition is negligible in comparison with the fundamental transition under "normal" conditions (i.e. at room temperature)[87]. This is the case as, in order for radiation to excite a transition, the resultant change in the molecular dipole (the transition dipole moment) must fundamentally be non-zero[84,88]. Relative to transitions where $\Delta\nu = \pm 1$, such as the fundamental transition, the change in dipole moment becomes increasingly small, and the light-matter interaction less probable.

When Equation (2) is applied to another diatomic molecule, carbon monoxide (CO), for which f = 1855 N/m, a vibrational frequency of 2141.3 cm^{-1} is predicted[85]. This serves to establish the fundamental principle of in IR spectroscopy; that different molecular functional groups can be readily distinguished by the frequencies at which they absorb light.

Diatomic molecules are the most basic molecular structures, and are resultantly the simplest case studies of IR spectroscopy. As the number of atoms within a molecule increases, the information retrieved by IR spectroscopy becomes rapidly more complex. The vibrations available to a molecule are often classified in terms of changes in the internal coordinates of the system. For a linear molecule consisting of N atoms, there are a total of 3N-5 available vibrations, and 3N-6 for a nonlinear molecule[88]. Figure (3) illustrates several typical vibrational modes.

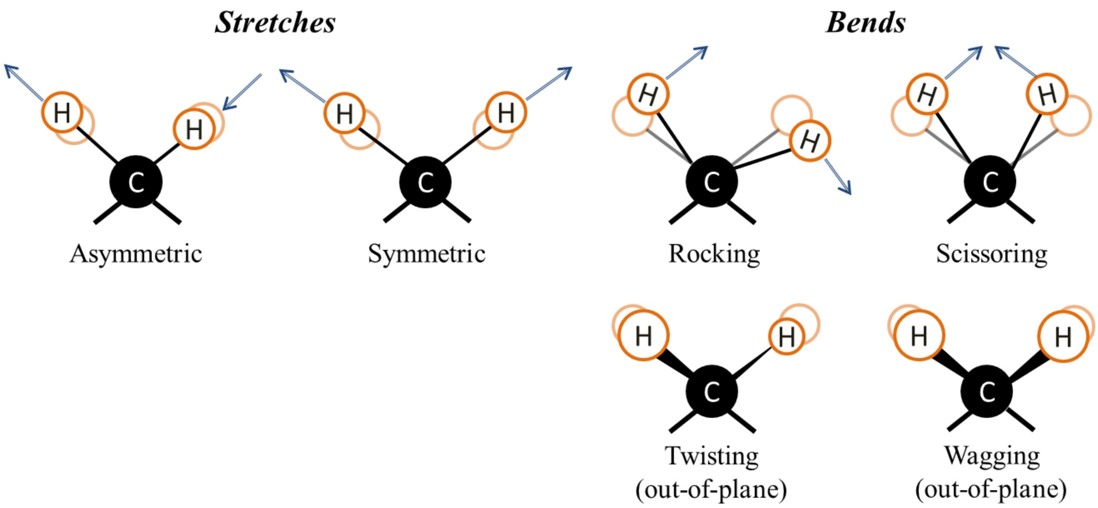


Figure (3): Common modes of molecular vibration, shown for a methylene functional group. The orientations of different bending and stretching species, with respect to the internal coordinates of the system, are differentiated by the given nomenclature.

The symmetric or asymmetric nature of a vibration has a significant effect on vibrational frequency, as symmetric stretches involve the motion of charges in opposite directions, and are thus of typically higher energy[89].

Molecular vibrational frequencies are influenced by their neighbouring functional groups[88,90], and thus two molecules with highly similar structures may be readily distinguished from their IR absorption spectra. Even between identical molecules, the instantaneous frequency of a particular species of vibration may vary as a result of inhomogeneous intermolecular contacts, creating a distribution of possible oscillation frequencies. Resultantly, infrared spectroscopy can extract information on local chemical environment from the frequency and width of a vibrational lineshape.

1.2.2 Linear IR spectroscopy

The true “absorption” of light in infrared spectroscopy does not directly occur as a result of energy transferred to the bond. Rather, the excitation field, E(t), induces a macroscopic polarisation, P(t), in the ensemble of molecules, which in turns causes emission of a signal from the molecular ensemble, $E_{sig}(t)$. A $\pi/2$ phase shift in $E_{sig}(t)$ relative to P(t) causes destructive interference between the emitted field and E(t), reducing the intensity of E(t) at frequencies corresponding to bond frequencies[91].

The absorption process can be divided into several events for more detailed description: (i) the induction of P(t), (ii) the emission of $E_{sig}(t)$ and (iii) the resulting spectrum.

(i) The polarisation induced in a dielectric material by an incident electric field E(t) is described by[92]:

$$P = \varepsilon_0 \chi E = \varepsilon_0(\chi^{(1)}E + \chi^{(2)}E^2 + \chi^{(3)}E^3 \ldots + \chi^{(n)}E^n) \quad (4)$$

where ε_0 is the electric permittivity of free space and χ is the dielectric susceptibility of the medium. χ describes the extent to which dipoles may be induced in the system by the excitation field. When considering an ensemble of molecular oscillators, χ is directly related to the polarizability of the molecule, which in turn is related to the transition dipole moment associated with the fundamental transition[93], μ_{01}.

The transition dipole is a quantum mechanical operator that yields the non-equilibrium charge distribution of the system, P(t), when applied to the molecular wavefunction[91]. The macroscopic polarisation is therefore dependent on the excitation of molecules from the $\nu = 0$ to $\nu = 1$ level. Following interaction with E(t), either vibrational state can be occupied, where the population of each state is related to the transition dipole moment.

(ii) Classically, the oscillating charges associated with the vibrating nuclei are sufficient to generate an electromagnetic field at the frequency of vibration, and at 90° to the phase of the polarisation, P(t)[94]. This emitted field, $E_{sig}(t)$, interferes destructively with frequency components of E(t) equal to bond vibrational frequencies, producing an absorption spectrum[95].

In a quantum mechanical description however, the excitation light couples the $\nu = 0$ and $\nu = 1$ states, permitting a superposition state. Non-superposition states that exist after excitation are seen to be related only to the static dipoles of the system, and so do not contribute to P(t) [91]. Coupling of the two vibrational levels *via* E(t) defines a strong phase relation between the states so that the superposition state is coherent. Light is radiated at the beat frequency of the $\nu = 0$ and $\nu = 1$ levels, i.e. at the energy of the fundamental transition. This radiation interferes with E(t) in the same manner as in the classical description.

The time-dependent decay of P(t) is propagated by a molecular response function, R(t), that encapsulates the intrinsic evolution of the molecule as it returns to equilibrium[96]. R(t) contains the absorption information unique to the molecule under measurement, and the Fourier transformation of this function yields the complex frequency-domain signal, S(ω).

(iii) The imaginary component of S(ω) yields the dispersive component of the vibrational lineshape, D(ω), while the real component yields a component describing the frequency-dependent attenuation of the system, i.e. the absorption spectrum, A(ω)[97]. The phase of D(ω) and A(ω) is such that only A(ω) interferes with E(t) [91].

$$D(\omega) \propto -\mu_{01}^2 \frac{\omega-\omega_{01}}{(\omega-\omega_{01})^2+(0.5\gamma)^2} \tag{5}$$

$$A(\omega) \propto \mu_{01}^2 \frac{0.5\gamma}{(\omega-\omega_{01})^2+(0.5\gamma)^2} \tag{6}$$

Where ω is the frequency of the light, ω_{01} is the resonant frequency of the vibration, and γ is the natural linewidth of the vibrational transition. The measured IR absorption spectrum, $A(\omega)$ follows a Lorentzian distribution, as shown in Figure (4).

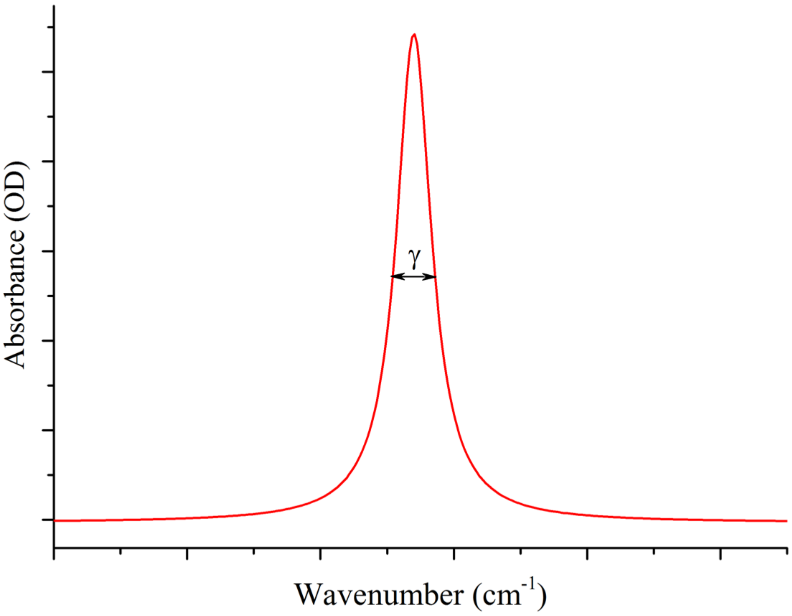


Figure (4): Illustration of a homogeneously broadened lineshape; a Lorentzian frequency distribution also referred to as the natural lineshape

1.2.2.1 Homogenous lineshapes

The width of this distribution in Figure (4), γ, originates from intrinsic quantum properties of the system. The $\nu = 1$ vibrational level is a transient quantum state, with a finite lifetime, T_1, and therefore carries an associated uncertainty in its energy through the Heisenberg Uncertainty principle[98].

$$\partial E \cdot \partial t \geq \frac{h}{4\pi} \tag{7}$$

where ∂t is excited state lifetime, ∂E is the uncertainty in the energy of the excited vibrational energy level and $h = 6.63\times10^{-34}$ J·s. As the vibrational frequency of the oscillator depends on the energy separation between the $\nu = 0$ and $\nu = 1$ levels, uncertainty in E has a direct effect on the oscillation frequency through:

$$\partial E = hc \cdot \partial \tilde{\upsilon} \tag{8}$$

From this, it can be seen that:

$$\partial \tilde{\upsilon} = \frac{1}{4\pi\delta t} \approx \frac{5.3\ \mathrm{cm}^{-1}}{\partial t\ (\mathrm{ps})} \tag{9}$$

$\delta\tilde{\upsilon}$ is therefore the contribution of the vibrational lifetime to the full width at half maximum (FWHM) of the Lorentzian frequency distribution[98]. Decay of the macroscopic polarisation is facilitated by the lifetime of the vibration, T_1; the spontaneous relaxation of the system, but over more rapid timescales, each molecule is also subject to dephasing processes that further affect the polarisation state[91]. The vibrational frequency of single molecule may occupy a narrow range of frequencies, as determined from the lifetime, and thus the instantaneous oscillation frequency may fluctuate rapidly. As this occurs, the phase-relation of the ground and excited state wavefunctions changes, resulting in loss of coherence and the decay of the polarisation[99,100]. This process, known as pure dephasing, occurs with a timescale T_2^*, and further contributes to the width of the Lorentzian spectral line. The combined effects of relaxation and pure dephasing dictate the minimum observable width of the vibrational band[91]; the natural (or homogeneous) linewidth, γ;

$$\gamma \sim \frac{1}{2T_1} + \frac{1}{T_2^*} = \frac{1}{T_2} \tag{10}$$

where T_2 is the homogeneous dephasing time, encompassing both of these intrinsic lineshape contributions. When the width of a vibrational line is determined only by T_2, the distribution is said to be “homogeneously-broadened”.

1.2.2.2 Inhomogeneous lineshapes

In a solution-phase system, the local environment of an oscillator plays a significant role in dephasing of the system[101], as many potential conformational or orientation states are available to the molecule. Effects such as hydrogen bonding may deform the potential well of an oscillator, altering the energy of the $\nu = 1$ state, and thus the frequency associated with the $\nu = 0 - 1$ transition[102]. The instantaneous frequency of a single oscillator therefore falls within a probability distribution of Gaussian profile, originating from the various structural sub-states that can be occupied by the molecule.

For a molecular ensemble, each accessible energy state in the Gaussian distribution is occupied, and the instantaneous frequencies of the oscillators are distributed inhomogeneously, resulting in an absorption band with FWHM necessarily greater than that of the natural linewidth. Indeed, the Gaussian distribution is composed of multiple overlapped Lorentzian distributions, as illustrated by Figure (5). The resulting vibrational band is referred to as “inhomogeneously-broadened”.

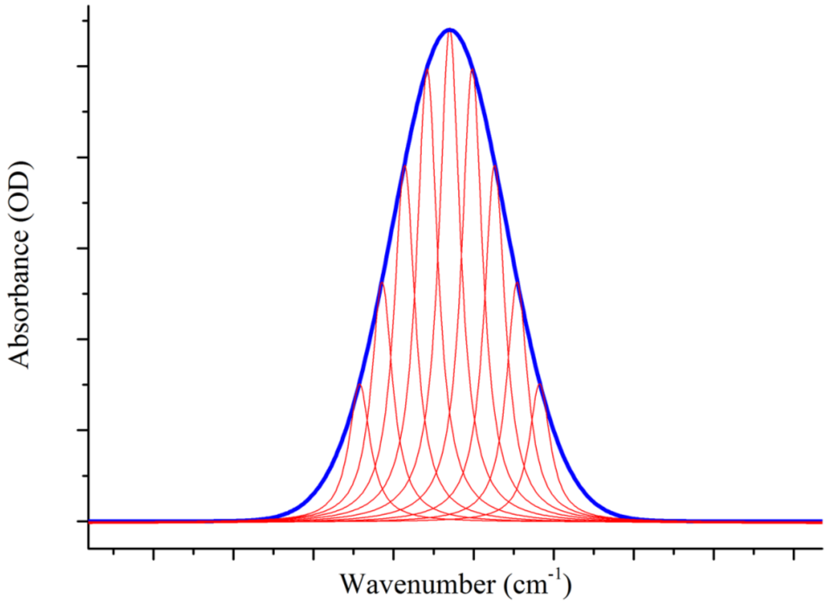


Figure (5): An inhomogeneously broadened vibrational band, originating from an ensemble of molecules, each of which contributes a Lorentzian lineshape to the ensemble absorption

Inhomogeneous broadening widens the distribution of vibrational frequencies for the molecular ensemble. Since the macroscopic polarisation of the system contains contributions from each molecule, a broader range of accessible frequencies leads to increasingly out-of-phase oscillations between molecules[91,103,104]. Thus the static inhomogeneity of a vibrational band contributes to dephasing of the polarisation.

The processes that give rise to inhomogeneous IR bands are, however, dynamic[58,105]. Single oscillators are in constant inter-exchange between energy subs-states of the distribution. This evolution of vibrational frequencies with time is the process of “spectral diffusion”. The molecular ensemble is therefore subject to rapid, intrinsic fluctuations in frequency (homogeneous dephasing), comparatively slow[106] fluctuations from exchange between structural sub-states (spectral diffusion) and instantaneous discrepancies between oscillation frequencies (inhomogeneous dephasing). In a solution-phase system, where molecules are in contact with a rapidly fluctuating solvent network, all three of these processes are likely to manifest in the IR lineshape[91].

Each effect here requires only that each molecular bond in the ensemble has access to a single potential well. However, other chemical events such as protonation[77], isomerisation[107], ligand binding[108], or H/D exchange can again widen the number of energy states available to oscillators. In these cases, the molecules may occupy more than one potential well, each of which is subject to the three processes above, though the inhomogeneity of these states and spectral diffusion timescales may differ within each well[83]. Additionally, the ability of the molecular ensemble to occupy two or more potential wells generates a corresponding number of vibrational bands[109]. Again, the molecules may be in constant inter-exchange between the available states, where the rate of interconversion is linked to the energetic favourability of one state relative to another[110].

Solution-phase vibrational bands therefore contain a significant amount of information regarding the local chemical environment of an oscillator. Though conventional IR spectroscopy theoretically extracts all of this information from the lineshape, the homogeneous and inhomogeneous components are irretrievably convoluted[91]. Separation of these contributions is a recent development that has only been made possible by advanced spectroscopic techniques utilising femtosecond-duration laser pulses. One such method, 2DIR spectroscopy, vastly increases the detail that can be extracted from molecular vibrations, and allows molecules to act as probes of local environment. 2DIR therefore holds a significant amount of potential in biological research, where the conformation states available to a molecular functional group may hold significance to a protein’s reaction mechanism.

1.3 The infrared absorptions of proteins

Proteins utilise simple building blocks to produce complex three-dimensional structures[111]. The polypeptide chains of even the simplest proteins may still contain ~100 of the peptide groups shown in Figure (6).

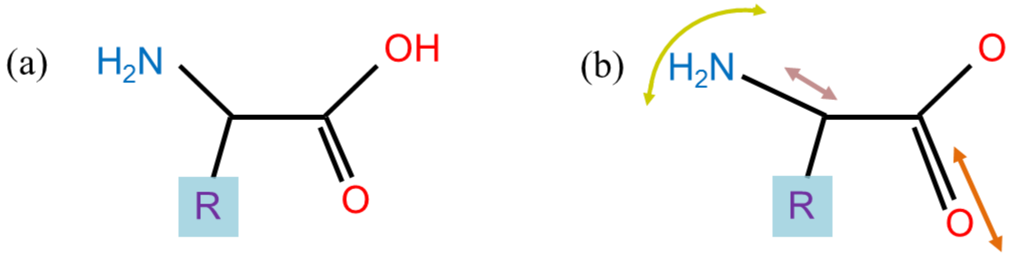


Figure (6): (a) The generic structure of a single peptide. The "R" group denotes the amino acid side chain of which there are twenty naturally-occurring species (see Appendix (A.1.1)) (b) Illustration of the different motions that may contribute to the so-called Amide I band of peptides; C=O stretching, C-N stretching and NH_2 *bending.*

Isolated peptides display a strong infrared absorption band ~1650 cm^{-1}. This peak originates from stretches of the backbone C=O bond, with small contributions from CN stretching, NH bending and CCN deformation, and is referred to as the "Amide I"[90].

All 20 amino acid species have an identical backbone structure, and thus Amide I vibrations are only made differentiable, in terms of frequency, by the local peptide environment. Resultantly, Amide I modes are effective probes of protein secondary structure, as the regular H-bonding geometry that occurs across, for example, an α-helix, modulates all Amide I oscillations across the helix in a regular fashion. Characteristic Amide I frequencies are therefore observed for peptides in an α-helical arrangement. This is equally true for other ordered structures such as β-sheets[89].

The spatial distribution and orientation of Amide I modes within a secondary structure domain allows interaction of adjacent vibrations along the polypeptide chain, by both through-bond (covalent) and through-space (electrostatic) mechanisms[112]. This effect, transition dipole coupling (TDC), is readily facilitated by the near-resonant frequencies of successive Amide I vibrations in secondary structure, and thus extends to all Amide I vibrations within the structural region, resulting in a highly delocalised vibrational component that is sensitive to conformational differences along the secondary structure domain[89].

Ultimately, many peptide residues with similar H-bonding geometries produce large absorption features that provide readily-interpretable data on sub-regions protein structure. Sensitivity of the Amide I band as a result of TDC has been historically important in the analysis and quantification of secondary structure[89,113]. Further modulation of the Amide I by effects such as solvent exposure may also allow the extraction of tertiary structure characteristics, giving the Amide I slightly enhanced scope for reflecting the unique environments between proteins.

Amide I studies are highly valuable to the study of proteins, and to other peptide systems. However, individual Amide I vibrations become indistinguishable in protein spectra, and molecular specificity is lost[114,115]. Distinction of the Amide I contribution from one residue over the contribution of any other is not of arbitrary importance in protein studies, as protein function takes place at well-defined local sites, giving certain residues a far greater functional significance than others.

Though other vibrational modes of the polypeptide backbone can provide more detail information on protein structure, such as the Amide II and Amide III modes, these bands have similarly complex absorption signals, if not more so, posing the same restrictions on access to site-specific or local information using infrared spectroscopy[89,90]. Backbone vibrations, such as the Amide I and II, have broad energy distributions, causing their lineshape to overlap with many of the higher-frequency side chain modes. Further, since there are 20 possible peptide side chains, the number of oscillators per mole of protein will be significantly higher for Amide modes than for any particular side chain, yielding higher intensity ratios that vastly favour the Amide vibrations. Side chain bands can therefore be difficult to isolate from backbone absorptions though, even in the absence of Amide features, a mixture of side chains would also give rise to a dense collection of spectral features[116,117].

Molecular vibrations can relay information on many chemical processes, which is beneficial to the understanding of the functionally-relevant interactions that take place at protein active sites. However, few intrinsically occurring protein modes are well-equipped to report on local environments, either due to delocalisation of vibrations, or due to spectral convolution. In order to access site-specific information, it is therefore necessary to employ an IR “probe” group whose spectral characteristics are predominantly determined by the interactions specific to the protein sub-site of interest.

1.4 IR probes of protein structure

The sensitivity of probe molecules is fundamental. There are two considerations in this regard:

(1) Proximity of the reporter group within the molecular system: the probe vibration may only provide information on molecular structures with which it interacts.

(2) Intrinsic sensitivity of the vibration. Increasingly delocalised vibrations, with contributions from greater numbers of atoms, become less sensitive to small fluctuations around a single atom. For this reason, the C=O stretch of a furan ring is unlikely to be as sensitive as the C=O stretch of acetone. The number of atoms involved in a particular vibrational band is not always known however, and for complex molecules, DFT[118] or MD[119] calculations are likely to be beneficial in this regard.

Within a protein study, probe groups are likely to occur in either of two forms: as modified versions of vibrations that occur within a protein molecule, or as external ligands that bind to a desirable site on a protein. Either approach should result in an IR signature, potentially composed of only a single peak, readily distinguishable from the spectral features of the protein. This can be achieved by employing probes with molecular functional groups that are not found in typical aqueous, protein-buffer systems, as this places the reporter mode in an uncluttered spectral region, and thereby eliminates the necessity for the probe molecule to exhibit particular intense absorption.

Considering firstly the strategy of protein structural modification to produce reporter groups; the Amide I modes of individual residues in the polypeptide sequence can be successfully isolated through the use of isotope-labelling. The relation between vibrational frequency and atomic mass provides an elegant solution to band convolution that does not interfere with the natural protein chemistry. Isotope-labelling of the backbone C=O is an effective means of altering the Amide I frequency, due to the large contribution of the carbonyl stretch in this absorption. Strategies such as $^{13}C{=}^{16}O$[61,120] and $^{13}C{=}^{18}O$[74,121–123] isotope-labelling have proved useful for smaller polypeptide systems by down-shifting the Amide I frequency of the labelled residue by 30-40 cm^{-1} in the case of a $^{13}C{=}^{16}O$ group, or by 55-65 cm^{-1} in the case of a $^{13}C{=}^{18}O$ label. Ideally, these shifts are sufficient to move the labelled vibration into a spectral region where no other absorptions occur, allowing information to be extracted directly from the labelled residue.

To implement this approach within a full protein is impractical however, and in the case of $^{13}C{=}^{16}O$ labels the natural abundance of 13C atoms (~1%) will also generate weak down-shifted bands from each residue. Moreover, in a full protein, where multiple secondary structure types are present, it is not a certainty that a $^{13}C{=}^{16}O$ will be adequately isolated. The expense of this approach is further limiting, particularly in the case of $^{13}C{=}^{18}O$ labels. Labelling is therefore best employed in simpler peptide systems, though these are often used as models for more complex polypeptides. In any case, isotope-labelling of main-chain carbonyl groups is an effective means of extracting local information on protein secondary structure.

Side chain atoms can be similarly isotope-labelled[124], though this approach is more temperamental than Amide I labelling, as not all side chain species possess localised, intense vibrational modes. Further, local protein environment can play a significant role in modulating side chain vibrational frequencies, particularly when that vibration is sensitive enough to be considered as a reporter[116]. It is therefore possible that the unlabelled frequency of a particular side chain is not accurately known, making the approach of isotope replacement somewhat dubious without a detailed band assignment.

Functional group substitution provides a more effective probe of side chain environment[125,126]. In this approach, an amino acid residue in the protein sequence is replaced with an unnatural variant of alanine such as cyano-homoalanine or azido-homoalanine. Cyano (C≡N) and azido ($N=N^+=N^-$) functional groups each possess simple stretching vibrations that occur around 2200 cm^{-1} and 2120 cm^{-1}, respectively; readily distinguishable from protein absorption bands. The native protein structure is compromised slightly in exchange for an IR tag group, and so care must be taken to ensure that side chain modification does not alter the set of interactions that it is intended to report on.

Protein ligands are potentially ideal probe molecules, as they directly engage with the protein's functional sites. Biological ligands come in a broad range of sizes however, from single protons, hydride ions or diatomic molecules, to large molecules such as dinucleotides or even other proteins, in the case of chaperone systems. The same considerations apply to binding ligands as for other local IR probes, but with several additions. Primarily, ligands may form intrinsically unstable complexes with their target proteins[127]. This is unsurprising, as the dissociation or chemical conversion of the ligand is often of crucial functional importance. Alternately, the binding molecule may be IR inactive, as in the case O_2 ligands. Under these circumstances, it is necessary to employ a structural analogue of the ligand to probe the active site. Analogues that are selected for their IR properties alone may have undesirable biological effects, modifying local side chain conformation or pK_a. Information gathered from a reporter under these circumstances may therefore not reflect the native biological state[128]. Analogues may also exhibit low binding affinity with the target protein, resulting in IR probe signatures whose primary contributions originate from an unbound form of the ligand analogue, in which case site-specificity is compromised. These considerations present something of a trade-of between the ideal IR characteristics of a system and its degree of comparability to its intrinsic biological state.

We now consider the information that is relayed from a local protein region *via* the lineshape of the IR signal of the probe. Thermally-activated protein motions span a broad range of timescales[108], and correspondingly involve both small and large structural fluctuations. As a result of this, inhomogeneous broadening of a probe's lineshape may not originate from a single dynamic process. If the probe is bound at a catalytic site that performs a multi-stage reaction, then several local contacts may indeed be expected, creating some ambiguity as to the chemical origins of lineshape broadening.

In a "steady-state" technique such as FTIR, the spectral acquisition time is many orders of magnitude longer than the overwhelming majority of dynamic processes that contribute to inhomogeneous broadening. The result is the time-averaged picture that has been described thus far; where conformational states essentially appear as static probability distributions. However, time-resolved infrared methods allow significantly more detailed to be extracted, enhancing the information that can be drawn from local probe molecules.

1.5 3rd-Order Spectroscopy

1.5.1 Nonlinear optical responses

Thus far, the polarisation and response functions have been considered to have linear dependence with the excitation field, as the time-dependent relaxation and dephasing of oscillators can be described sufficiently without introducing more complex terms. However, as previously stated, vibrational dynamics occur on rapid timescales that can only be accessed with short-duration laser pulses. Such strong excitation fields readily permit nonlinear optical responses[92]. With reference to Equation (4), when the magnitude of E is large, the macroscopic polarisation of the system contains non-negligible from higher-order terms, producing a nonlinear polarisation[91]. When we measure the response function in a non-linear regime such as this, R(t) may therefore contain contributions from both the linear and 3rd-order responses, as well as higher-order, odd terms that have not been detailed here. However, even terms for the dielectric susceptibility are zero in centrosymmetric media (e.g. $\chi^{(2n)} = 0$), making $\chi^{(3)}$ the simplest non-linear term in a typical solution-phase system[91].

3rd-order optical signals are typically extracted *via* a "four-wave mixing" interaction, where three excitation fields are incident with a material (in this case the molecular ensemble) at temporal separations of t_1, t_2 and t_3[129]. As a result, the response function contains highly complex information that relates to the multiple ways in which the pulses may interact with the ensemble. This is the basis for the additional information provided by 3rd-order spectroscopy[91,130], though the 3rd-order response function becomes convoluted with each field to produce the macroscopic polarisation;

$$P^{(3)}(t) \propto \int_0^\infty dt_3 \int_0^\infty dt_2 \int_0^\infty dt_1 E_3(t-t_3)\, E_2(t-t_2-t_1) E_1(t-t_3-t_2-t_1) R^{(3)}(t_3,t_2,t_1) \qquad (11)$$

Deconvolution of the 3rd-order response from the excitation fields can be achieved through the approach of "phase-matching" of the multi-pulse sequence[86].

The waves in four-wave mixing consist of the three laser pulses and the emitted field, which are commonly denoted by their wave-vectors, $\vec{k}$, which describes their rate of change of phase along the direction of propagation, i.e. $2\pi/\lambda$. The four waves are therefore referred to as $\vec{k}_1$, $\vec{k}_2$, $\vec{k}_3$ and $\vec{k}_4$, where the subscript denotes the temporal order of the waves, such that $\vec{k}_4$ is the emitted field. Each of the excitation fields interacts with the molecular ensemble independently, i.e. the excitation pulses do not overlap with each other in time, and each imprints the polarisation with its wave-vector, which ultimately dictates the propagation direction of the emitted field, $\vec{k}_4$, through:

$$\vec{k}_4 + \vec{k}_3 + \vec{k}_2 + \vec{k}_1 = 0 \qquad (12)$$

Significantly, the excitation states accessed by the molecular ensemble are also controlled by the wave vectors $\vec{k}_1$, $\vec{k}_2$, and $\vec{k}_3$. The origin of phase-matching lies in the statistical, quantum nature of the molecular ensemble; it is fundamentally a mixed quantum state, and so is represented by a density matrix that describes the statistical average of the system[130]. In its most basic explanation, when the expectation value for the ensemble polarisation is derived by applying the transition dipole operator to the density matrix, different excitation states may emerge from interaction with $\pm\vec{k}$. Positive wave-vectors excite *ket* states of the ensemble, negative wave-vectors excite *bra* states. A reciprocal relation is also true, where the wave-vectors de-excite the opposing *bra* or *ket* terms[91,131].

Excitation and de-excitation of either the *bra* or *ket*, between the $\nu = 0,1$ and 2 levels creates a number of possible excitation pathways for a series of three consecutive pulses. In the specific cases for which $\vec{k}_1$ causes the excitation of one state (for example, bra), but $\vec{k}_4$ originates from the de-excitation of the opposite state (ket), the pathway is said to be rephasing. A 3rd-order response function that rephases is able to restore the macroscopic polarisation induced by the pulse $\vec{k}_1$ (and at time $_{t1}$), emitting a "photon-echo" signal[107,132]. This occurs through the interaction of either the *bra* or *ket* state with its complex conjugate that allows the system coherence to evolve as a mirror image of the coherent evolution during t_1, so that the initial coherence is recovered. Rephasing pathways are essential to the separation of homogeneous and inhomogeneous broadening, as reintroduction of the macroscopic polarisation essentially occurs as a mirror image of the initial polarisation decay[91]. So, just as the rate of coherence-loss is dependent on the inhomogeneous distribution of vibrational frequencies in the ensemble, the rate of coherence-recovery is similarly dependent, and contributes to the photon-echo signal.

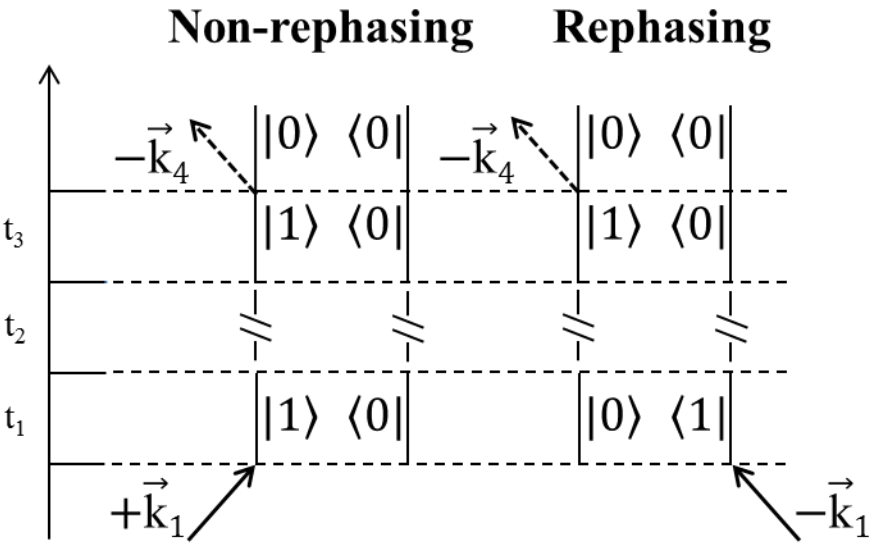


Figure (7): Depiction of the generic differences between rephasing and non-rephasing pathways using a 2-sided Feynman diagram.

Pathways that do not fulfil the rephasing criterion (and are de-excited from the same *bra* or *ket* state after t_3 as was initially excited at t_1) are said to be non-rephasing. In this case $\vec{k}_3$ does not allow recovery of the coherent state produced by $\vec{k}_1$, but instead propagates the ensemble "forwards" in time to produce the second coherent state. Non-rephasing pathways are insensitive to inhomogeneous dephasing, and therefore collection of data from both rephasing and non-rephasing pathways is essential to distinguishing between the homogeneous and inhomogeneous linewidths, and indeed to preventing distortion of the 3rd-order lineshape[86,91].

By selecting a phase-matching condition in which $\vec{k}_1 = \vec{k}_2$, signals from rephasing and non-rephasing pathways are emitted collinearly, simplifying the experimental arrangement that must be used to collect the complete 3rd-order response[133]. This condition is employed throughout the measurements performed in this thesis, and thus the lineshape will not be resolved explicitly in terms of these components in later discussion. Likewise; general non-linear spectroscopic theory often employs the density matrix formalism in conjunction with two-sided Feynman diagrams to map each of the alternate excitation pathways available in an nth-order experiment[134].

While a thorough categorisation of available pathways is certainly necessary in the modelling of spectra, it is not strictly necessary to justify each excitation pathway for an experimental discussion, as not all pathways are measured. There are several forms of 3rd-order spectroscopy[135–137], and also 5th-order spectroscopy[102,138], each of which is defined by a different phase-matching geometry. In all cases, to deconvolute the 3rd-order response from the excitation fields, the phase relationships of the pulses must be known, though each approach takes a different approach to acquiring this information. Due to this, the experimental arrangements and considerations in each method are different[131]. Detailed discussions of pulse-shaping, transient-grating, peak-shift and CARS techniques collectively provide a thorough explanation of the various ways in which data can be collected. In this investigation however, two specific 3rd-order IR spectroscopies are employed: pump-probe and 2DIR (in a pump-probe geometry), and so the theoretical concepts above will only be developed where relevant to these methods.

Figure (9) provides an illustration of how a three-pulse sequence may be used to produce a photon echo from a molecular ensemble.

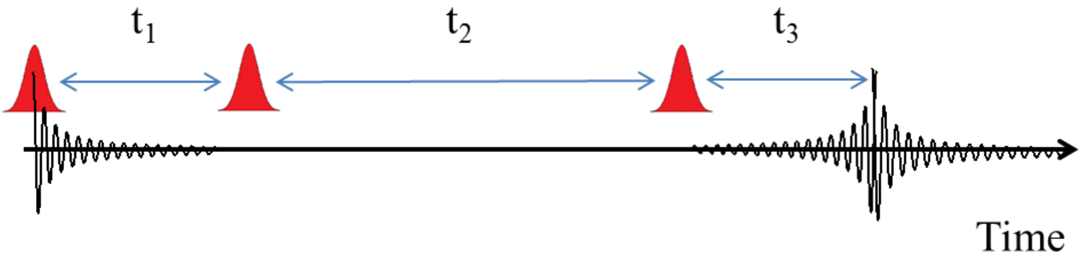


Figure (8): The production of a photon-echo signal using a three-pulse sequence; the first pulse creates a coherence, which decays during t1 (black trace). The second pulse switches the ensemble to a non-coherent population state for the period t_2, during which population dynamics occur. The final pulse triggers a rephrasing of the system, restoring coherence after a time t_3. This second coherence gives rise to a photon-echo signal (black trace) which contains information on the dynamics that occur during t_2.

Coherence fundamentally involves a superposition of states. Within the context of the molecular ensemble, a coherent superposition occurs when the *bra* and *ket* states have different excitation levels ($\nu = 0$ and $\nu = 1$). If it is assumed that the distinction between *bra* and *ket* excitation is the positive or negative phase of the excitation pulse, then for two pulses with $\vec{k}_1 = -\vec{k}_2$, the ensemble state after irradiation with both pulses is likely to have equal populations in both the *bra* and *ket* states. Under this condition there is no coherent state, but rather a population state. Therefore, the system varies as follows through interaction with three pulses:

- The ensemble is initially in the ground state

- The $\vec{k}_1$ pulse is interacts with the system to create a coherent state through the excitation of either the *bra* or *ket* state of the ensemble, which oscillates at the frequency, ω_{01}

- During t_1 the macroscopic polarisation decays as a result of dephasing between molecules

- The second pulse, $\vec{k}_{2,}$ creates a population state, either by de-exciting the initially excited state, or by exciting the opposite state to that previously excited.

- During t_2 the populations evolve in time. Physically, this can occur from a number of sources, such as: spontaneous relaxation, spectral diffusion, chemical reactions or from molecular couplings that allow a population of one mode to be transferred to another.

- The third pulse, $\vec{k}_{3,}$ induces a second superposition state, restoring a macroscopic polarisation. For a rephasing pathway, this causes emission of a photon-echo pulse a time, t_3, after the third pulse, where t_3 is approximately equal to the separation of the first and second pulse.

With respect to a two-sided Feynman diagram, the pulse-matter interactions give rise to a number of potential excitation pathways. Illustrates of these alternate pathways are shown in Figure (10)

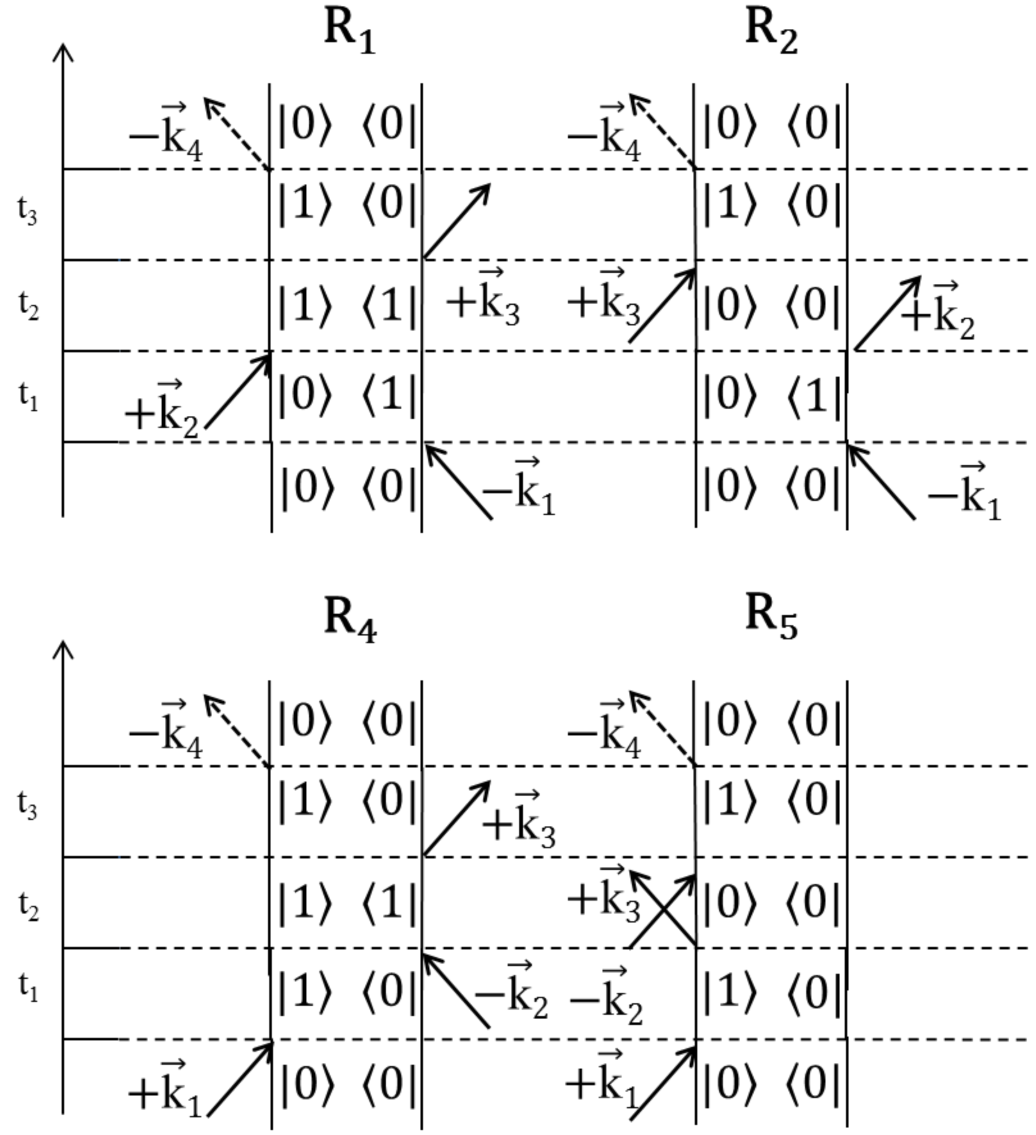


Figure (9): Four possible excitation pathways (two-sided Feynman diagrams) for a four-wave mixing interaction. Feynman diagrams in the top row illustrate rephasing pathways, while those in the bottom row illustrate non-rephasing pathway.

1.5.3 Pump-Probe Spectroscopy

Pump-probe spectroscopy is conducted with only two laser pulses, but can be fully described as a four wave mixing interaction in which the $\vec{k}_1$ and $\vec{k}_2$ pulses are degenerate and temporally overlapped[91]. The combined $\vec{k}_1$ and $\vec{k}_2$ pulse is referred to as the "pump" pulse, while $\vec{k}_3$ is the "probe". In this four-wave mixing interaction, $t_1 = 0$, and so the separation of excitation pathways during this interval is not possible. For this reason, the interaction of the pump pulse with a molecular ensemble automatically creates a population state. The coherent state created by $\vec{k}_3$ may therefore originate from the excitation of a population $|0\rangle\langle 0|$ state, or through de-excitation of a $|1\rangle\langle 1|$ population state. The population of each state is dependent on the pump-probe temporal separation, t_2, as the relaxation that occurs in this period may repopulate the ground level by depleting the $|1\rangle\langle 1|$ population state. The final coherence is thus directly affected by the vibrational relaxation that occurs within t_2. Indeed, pump-probe spectroscopy allows for the direct measurement of vibrational lifetimes.

Pump-spectroscopy is an absorption-difference technique in which the effects of initial pumping of the molecular ensemble must be established. The difference between the absorption spectrum of the probe, in the presence and absence of an initial pump beam, yields the change in absorption caused by pumping, allowing isolation of the transient absorption features whose amplitude is determined by population relaxation.

The only variable in a pump-probe experiment is therefore the inter-pulse separation, t_2. When $t_2 = 0$, no relaxation will have been allowed to occur in the system, i.e. no vibrations previously excited to the $\nu = 1$ level will have decayed to the ground state. The probe pulse may therefore excite vibrations with $\nu = 1$ to the $\nu = 2$ level, or, cause transitions from $\nu = 1$ to $\nu = 0$, stimulating emission of photons from the first excited state. As a result of this initial excitation created by the pump, the probe spectrum detects reduced absorption for the $\nu = 0\text{-}1$ transition, owing both to the stimulated emission at this frequency and to the slightly depleted population of molecules that occupy the $\nu = 0$ level. Further, the probe spectrum contains absorption contributions from the second transition, $v = 1\text{-}2$. In the event of an anharmonic potential, the vibrational frequency associated with this higher transition will lie at a lower frequency to that of the $\nu = 0\text{-}1$, potentially allowing it be resolved as a separate absorption band.

If the time delay between the pump and probe is incrementally lengthened, then the probability of spontaneous relaxation from $\nu =1$ to $\nu =0$ increases. In turn, the absorption difference becomes smaller. Ultimately, when t_2 is sufficiently large to allow complete relaxation then the absorption registered by the probe is independent of the pump pulse, and no absorption differences are measured. The transient absorption intensity as a function of t_2, yield the vibrational relaxation, T_1.

For a simple, three-level vibrational system, such as that of a diatomic molecule, pump-probe spectroscopy measures only T_1. In the case that many vibrations occur within the spectral window of the laser pulses, the response is not as simple[91], though this will be discussed explicitly below. In general, however, pump-probe spectroscopy provides information on the relaxation mechanisms of molecular functional groups, which in turn can be used to infer properties of the functional group's local environment.

The phase-matching condition of pump-probe spectroscopy ($\vec{k}_1 = \vec{k}_2$) necessarily results in the emission of $\vec{k}_4$ in the same direction as $\vec{k}_{3.}$ The signal also contains information from both rephasing and non-rephasing pathways[133] (Figure (10)).

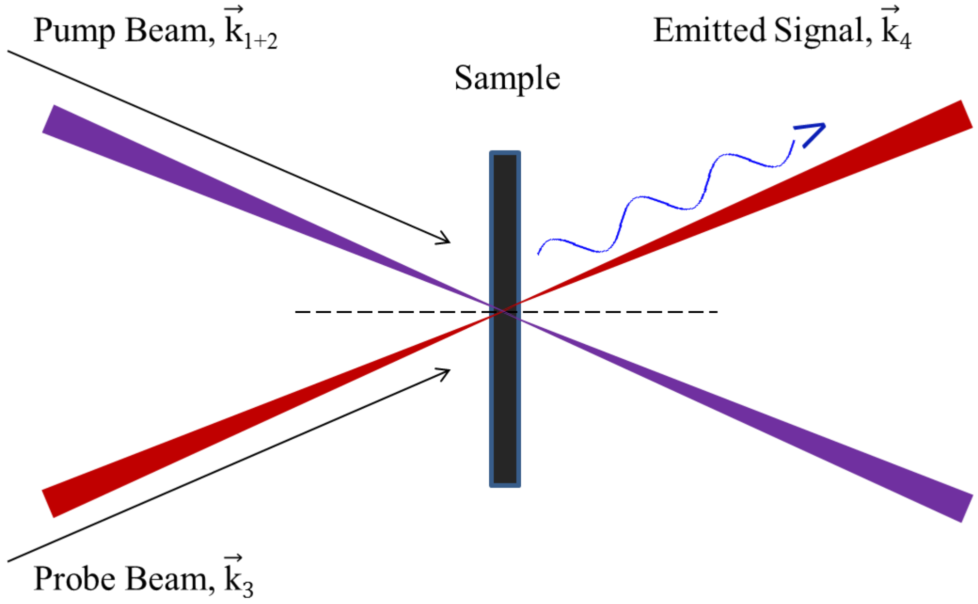


Figure (10): The beam geometry in a phase-matching approach for which $\vec{k}_1 = \vec{k}_2$*, from a view parallel with the sample plane*

Convolution of the emitted signal and the probe field is far from detrimental[56]. Since the interference term between $\vec{k}_3$ and $\vec{k}_4$ is proportional to the product of both the probe and signal field magnitudes, the residual probe light allows $\vec{k}_4$ to be detected more readily than is possible for the weak, isolated signal[91]. This process is known as heterodyne-detection. In the condition that the field responsible for signal emission also interferes with the emitted signal, as occurs between $\vec{k}_3$ and $\vec{k}_4$ in a pump-probe experiment, the signal is further referred to as “self-heterodyned”. The loss of probe intensity through interference with the signal field also only reflects the purely absorptive part of the lineshape[139,140]. When additional pulses are employed, the emitted signal is not necessarily self-heterodyned, which creates additional difficulty in measurement and phasing of the signal. A local oscillator beam ($\vec{k}_{LO}$) can be employed in these circumstances to introduce an interference term that scales up the signal intensity in a similar manner to $\vec{k}_3$ in a self-heterodyning regime. When $\vec{k}_1$ and $\vec{k}_2$ are degenerate, a self-heterodyning scenario is achieved, and residual light from $\vec{k}_3$ effectively serves as $\vec{k}_{LO}$.

Other dynamics besides relaxation occur within t_2, though pump-probe is not well-equipped to resolve this information, particularly in the event that a molecule with a complex vibrational spectrum is under investigation. As previously stated, a multi-oscillator system may give rise to a complex response during a pump-probe measurement, one that cannot solely be attributed to T_1. For coupled vibrations, the selective excitation of a mode, α, may create an oscillation in the dipole of mode β, allowing the population of α to vary as a consequence of its intrinsic relaxation, T_1, and as a consequence of population transfer to β. Provided that the frequency separation of α and β is greater than their combined FWHM, it should in principle be possible to detect coupling of the modes through the selective excitation of either α or β. Since pump-probe employs a broadband excitation source, it simultaneously excites all oscillators within the detection window and this coupling information becomes convoluted with population relaxation[91].

The ability to resolve molecular couplings from an IR probe is highly desirable, as it may indicate specific dipole orientations between modes which in turn can be used to infer molecular geometries within the system, or, coupling patterns may exemplify the exchange of an oscillator between conformational or chemical states. This additional information can be obtained by approaching the four-wave mixing interaction in a different manner.

1.5.3 Two-Dimensional Infrared (2DIR) Spectroscopy

2DIR expands on pump-probe spectroscopy by allowing frequency resolution of the pump axis *via* control of the pulse sequence. This permits the differentiation of a greater number of the 3rd-order excitation pathways than is otherwise possible[56,91,141]. Construction of a pump frequency axis can be approached using either a "time-domain" or "frequency-domain" approach[142,143]. Though a time-domain method has been used throughout the measurements discussed here, the "frequency-domain" approach will be first described, as it allows more comparisons to be drawn between 2DIR and pump-probe spectroscopies.

1.5.3.1 Frequency-Domain 2DIR

We begin with the pump-probe setup where $\vec{k}_1$ and $\vec{k}_2$ are degenerate and collinear, so that the following conditions are still valid:

- The signal is emitted in the direction of the probe pulse

- The signal is self-heterodyned, and so contains only information on the absorptive components of the lineshape

- A spectrometer is used to spectrally disperse the signal onto a detector

- Rephasing and non-rephasing pathways of the molecular response are collected together (Interchangeability of the pump beams effectively means that it is impossible to distinguish whether the *bra* and *ket* states of the ensemble are acted upon first)

As with pump-probe spectroscopy, the inter-pulse timing, t_1, cannot be controlled. Instead, the frequency of the pump pulse is regulated. One initially creates two broadband pulses as in the pump-probe method, though the pump is subsequently manipulated *via* an etalon filter or pulse-shaper to produce a spectrally narrow pulse with controllable central frequency. Since the pump originates from an identical light source to the probe, the pump frequency can be scanned over the full spectral range of the probe, in narrow increments.

A hypothetical case is now considered where frequency-domain 2DIR is used to study a molecule with the linear spectrum shown in Figure (11)(a). The spectral width of the pump pulse is chosen to be smaller than the inhomogeneous bandwidth of either mode, and the following steps are performed:

- The pump frequency is scanned in steps across the full frequency range of the probe pulse

- A probe spectrum is acquired at each pump frequency and, just as in pump-probe spectroscopy, the absorption difference is isolated through a comparison of the probe spectrum, with and without the initial pump excitation

- The compiled probe spectra generate a 2DIR spectrum

Effectively, the same excitation range has been used as in the pump-probe measurement, though the individual response originating from a specific excitation frequency can now be resolved. This allows construction of a pump frequency axis, ω_1. Figure (11)(b) shows a schematic 2DIR spectrum of the hypothetical system under investigation. Unlike a pump-probe measurement, the pump and probe pulses in 2DIR remain at a fixed temporal separation while the excitation frequency is scanned. This allows the population dynamics that occur within t_2 to be resolved by successive collection of spectra as function of t_2. In the context of time-domain 2DIR, t_2 is often referred to as the "waiting time", T_w. This term will be employed here, owing to the use of time-domain 2DIR in these studies.

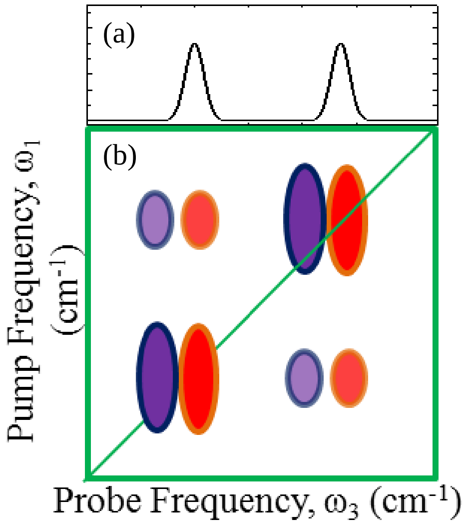
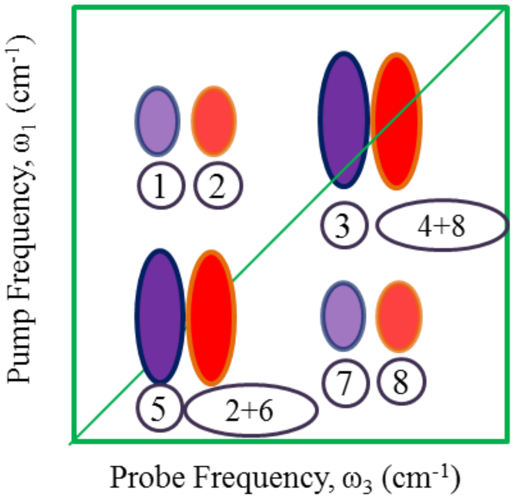


Figure (11): Schematic IR spectra of a hypothetical molecule in (a) one and (b) two dimensions. The figure in (b) reflects a frequency-domain 2DIR spectrum acquired at an *arbitrary value of t_2, or T_w. The red on-diagonal features denote the ν = 0-1 transition, while the ν = 1-2 is shown in purple.)*

2DIR spectra are typically discussed in terms of two-regions: the on-diagonal and the off-diagonal[91,141]. The spectrum diagonal ($\omega_1 = \omega_3$) reflects the linear spectrum, since by definition it only resolves vibrations that occur at the same frequency as the excitation light. Off-diagonal features originate from non-linear responses, though they may be interpreted somewhat intuitively; pumping at the ν = 0-1 frequency allows the broadband probe pulse to excite ν = 1-2 transitions, and so the ν = 1-2 peak occurs at coordinates of ($\omega_{probe} = \omega_{1\text{-}2}$, $\omega_{pump} = \omega_{0\text{-}1}$). Molecular couplings appear in a similar fashion; pumping at a specific on-diagonal frequency, ω_α, allows an absorption to be detected at the pump frequency ω_β. With respect to the vibrational levels of the oscillators,

A reciprocal relation also exists, where the pumping of mode β excites α. For this reason cross-peaks appear at coordinates of both (ωα, ωβ) and (ωβ, ωα). Additional excitation pathways originate from the coupled oscillators, giving rise to additional off-diagonal spectral features. Figure (12)(a) illustrates the available transitions for two anharmonically-coupled vibrations in a four-wave mixing experiment. Peaks in the previously seen 2DIR spectrum (Figure (11)(b)) are annotated with respect to these transitions in Figure (12)(b).

(a)



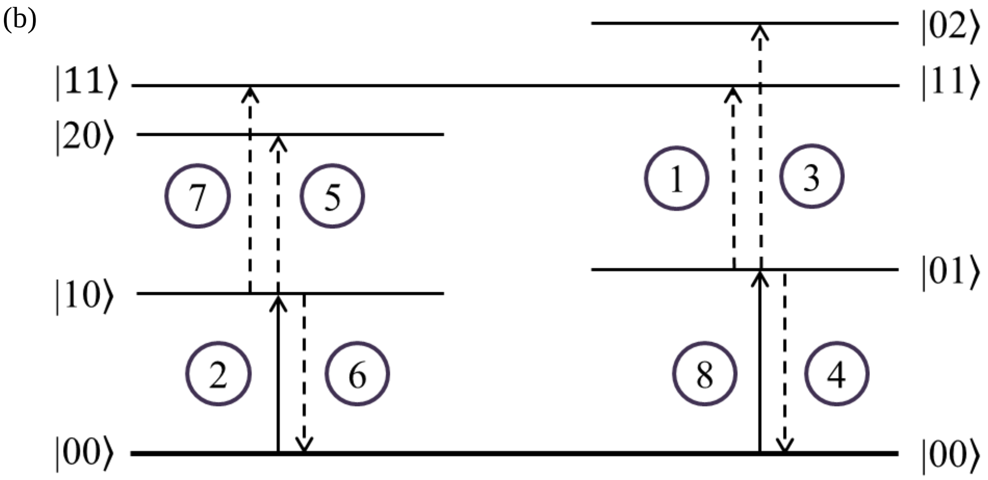


Figure (12): (a) The schematic spectrum presented in Figure (11) with peaks annotated according to their corresponding transitions, shown in (b) with respect to an anharmonically-coupled system of two oscillators.

Many processes may occur during T_w, such as those that give rise to the cross-peak patterns discussed here. Although the description of frequency domain can be used to describe significantly more features of 2DIR spectroscopy, it is used here to introduce the concept of correlation between ω_1 and ω_3, and the experimental arrangement that is required to produce a second frequency axis. For the purposes of this work, frequency-domain 2DIR is not entirely interchangeable with time-domain, as the smallest value of t_2 that can be measured in a 2DIR experiment is influenced by pulse duration; overlap between pump and probe will lead to a more complex response function. Since the frequency-domain method of 2DIR utilises a narrowband pump, the temporal profile of this pulse can be >1 ps. Many vibrations within biological systems are short-lived, with vibrational lifetimes in the order of ~1 ps[60,80], and so this variant of 2DIR spectroscopy will restrict the amount of information that can be extracted from a protein system. The time-domain approach to 2DIR spectroscopy is therefore better suited to biological investigations.

1.5.3.2 Time-Domain 2DIR

In the time-domain approach to 2DIR, a broadband pump pulse is used, permitting a higher temporal resolution than in frequency-domain 2DIR[131]. Control of the pump frequency can be implemented using a Michelson Interferometer, or alternatively using the more recently developed technique of acousto-optic “pulse-shaping”[144,145]. The former approach has been used in these studies, and so the provided theory will pertain to pump frequency control *via* a MIR Michelson Interferometer. In this configuration, the pump pulse is split into two equally intense components that propagate through a Michelson Interferometer and emerge collinearly, but at controllable temporal separation, t_1. Displacement of the interferometer’s mobile arm by a distance Δx, varies the inter-pulse timing, t_1, as:

$$t_1 = \frac{2\Delta x}{c} \qquad (13)$$

where c is the speed of light. For each increment of Δx that the stage is displaced, the pulse path varies by $2\Delta x$, where the factor of 2 originates from the round-trip to and from an optical delay stage in the mobile arm of the interferometer. Δx gives rise to a path difference between the two pump pulses, and results in interference when $t_1 \neq 0$. The remaining frequency components of the recombined pulse are therefore dependent on the phase difference induced by stepping of the interferometer. Scanning of t_1 allows the excitation frequency to be controlled in an analogous manner to the scanning of ω_{pump} in the frequency-domain.

Using a pulse-shaper, the frequency of the pump beam is programmable, and the phase relation of the two pump beams is known. However, when employing a Michelson Interferometer, the pump frequency axis must be constructed through Fourier-transformation of the inter-pump interference pattern that originates from the scanning of t_1[133]. This requires additional considerations in the experimental design, as the pump-pump interference pattern, referred to as the pump auto-field correlation, must be collected in parallel to the measurement. More detail on this procedure is given in Chapter (2).

The point of zero phase difference (t_0) between $\vec{k}_1$ and $\vec{k}_2$ can be established from the pump auto-field correlation. This permits the Fourier-transformation along t_1, which in turn yields the pump frequency range, $\omega_{1.}$ At each value of t_1, a probe spectrum is collected. As in the previously discussed techniques, the emitted 3rd-order signal is heterodyned with residual probe light and is then spectrally dispersed, allowing it to be collected in the frequency-domain. However, the probe spectra collected here cannot immediately be compiled according to their pump frequencies, as is the case in frequency-domain 2DIR. The collected probe data contains contributions from oscillatory time-domain signals from the rephasing pathways (the photon-echo) and non-rephasing pathways (the free induction decay). These signals are Fourier transformed using the phase-relation established from the pump auto-field correlation to convert the time probe intensities measured across t_1 into frequency spectra across $\omega_{1.}$ This final "phasing" step yields a 2DIR spectrum with equivalent information to the compiled probe spectra in frequency-domain 2DIR[131]. Differences in pump frequency resolution and data acquisition time between the two methods are discussed in Appendix (A.1.2).

As probe data is acquired in the frequency-domain, it is instructive to clarify the origins of the oscillatory time-domain data that is phased to produce time-domain 2DIR spectra. In reality, the collected probe spectra contain several contributions. In addition to the free induction decay and photon-echo, the probe data also contains two pump-probe signals. Since $\vec{k}_1$ and $\vec{k}_3$ are fixed in time, they generate a pump probe spectrum of constant amplitude across t_1. This can subtracted directly from the probe data through referencing with a probe spectrum that is acquired when $\vec{k}_2$ is blocked and prevented from interacting with the sample. The second pump probe contribution to signal occurs as a result of the interaction between $\vec{k}_2$ and $\vec{k}_{3.}$ The intensity of this signal is not fixed in time, and but is an exponential function with a time-constant equal to the vibrational lifetime of the vibration. Fourier-transformation of the oscillatory time-domain signal reduces this exponential to a

Lorentz lineshape with width determined by the lifetime, T_1. The second pump probe signal is therefore “removed” through conversion to the frequency-domain.

T_w (t_2) remains fixed during these 2DIR measurements, as it does in the frequency-domain variant. In the previously shown schematic 2DIR spectrum, the effect of dynamics occurring in T_w was not considered. Another hypothetical molecular system is thus considered to illustrate how dynamic processes in T_w contribute to the 2D lineshape. Figure (13) shows two schematic 2DIR spectra of this system, at two waiting times. Inset above each 2DIR spectrum is a representation of the molecule’s linear absorption.

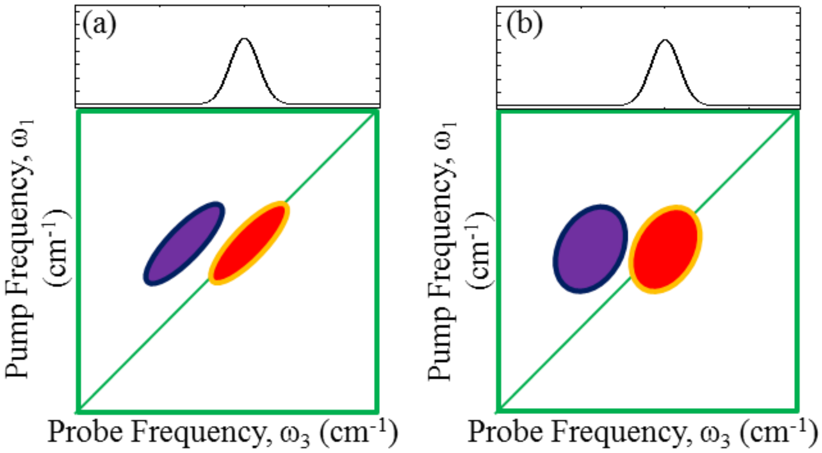


Figure (13): Two schematic time-domain 2DIR spectra of a simple molecule with one vibrational mode (inset above each spectrum), at waiting times of (a) $T_w = 0$ and (b) $T_w > 0$

The oscillator depicted is anharmonic, allowing both the ν = 0-1 and ν = 1-2 transitions to be resolved in each spectrum. The case in Figure (13)(a) is firstly considered.

At $T_w = 0$ both ν = 0-1 and ν = 1-2 bands are elongated along the spectrum diagonal, so that the peak width perpendicular to the spectrum diagonal is much smaller than the diagonal width. This effect originates from the narrow spectral width with which the vibration is excited; only a subset of oscillators in the band has an instantaneous frequency that is resonant with the excitation light. Since the minimum width of an absorption band on the probe axis is dictated by the homogeneous linewidth, this quantity can be directly observed in a 2DIR experiment where the frequency range used to excite the system is smaller than γ. The anti-diagonal width of the absorption band therefore reflects γ at waiting times approaching zero.

It can be inferred from Figure (13)(a) that this vibration is inhomogeneously broadened, as the FWHM of the band along $\omega_1 = \omega_3$ is larger than the anti-diagonal width. If homogeneous broadening were the only contribution to lineshape then the width in each axis would be equivalent. Since the probe only detects absorptions at the sub-set of frequencies that were excited by the pump, the inhomogeneous distribution of states can be thought of as static across the molecular ensemble; they do not exchange or evolve in time[91,106]. Since $T_w = 0$ in this spectrum, then no population dynamics may have occurred. By extending the waiting time, the instantaneous frequencies of each oscillator becomes increasingly likely to fluctuate as a result of dynamic inhomogeneous processes[91].

These interactions occur with the excited molecule during T_w, and by extension of the waiting time, we create the spectral differences between Figure (13)(a) and (b). The anti-diagonal width of the probed absorption is larger in the later waiting time spectrum, and is now greater than γ. As in the $T_w = 0$ spectrum, the absorption contributions can be separated from a sub-region of the spectrum with a FWHM of γ, though in this case, when the ensemble rephases, the oscillators no longer vibrate with the same frequencies as at the moment of excitation. The interactions during T_w cause this frequency modulation, and as T_w increases, the anti-diagonal width recovered from each sub-population of molecules increases, approaching the inhomogeneous limit. When T_w is sufficiently large that structural reconfiguration about the vibration are certain to occur, then the instantaneous frequency of any oscillator at t_1 is different from that at t_3. In this case the ensemble returns to a macroscopic polarisation state, but some “memory” of the initial state has been lost[107]. Structural fluctuations may occur slowly in comparison with vibrational relaxation, though this will be discussed in more detail below.

The dynamic process of lineshape broadening, spectral diffusion, provides a means of extracting highly specific information from the local environment of a probe[146–149]. Structural fluctuations around a reporter molecule often drive spectral diffusion, and thus, by determining the timescales over which broadening occurs it is possible to infer the specific molecular contacts that cause reconfiguration of the probe environment[59,73,77,150,151].

For an inhomogeneous system, elongation of the 2D band along the spectrum diagonal will give way to a more circular profile at later waiting times, and so spectral diffusion is quantified *via* a description of the lineshape "tilt", i.e. its degree of correlation between the ω_1 and ω_3 axes. Several methods have been established for the extraction of this information, such as the central line slope (CLS)[152–154] and nodal line slope (NLS)[155,156] methods, both of which provide a highly intuitive, visual representation of spectral diffusion.

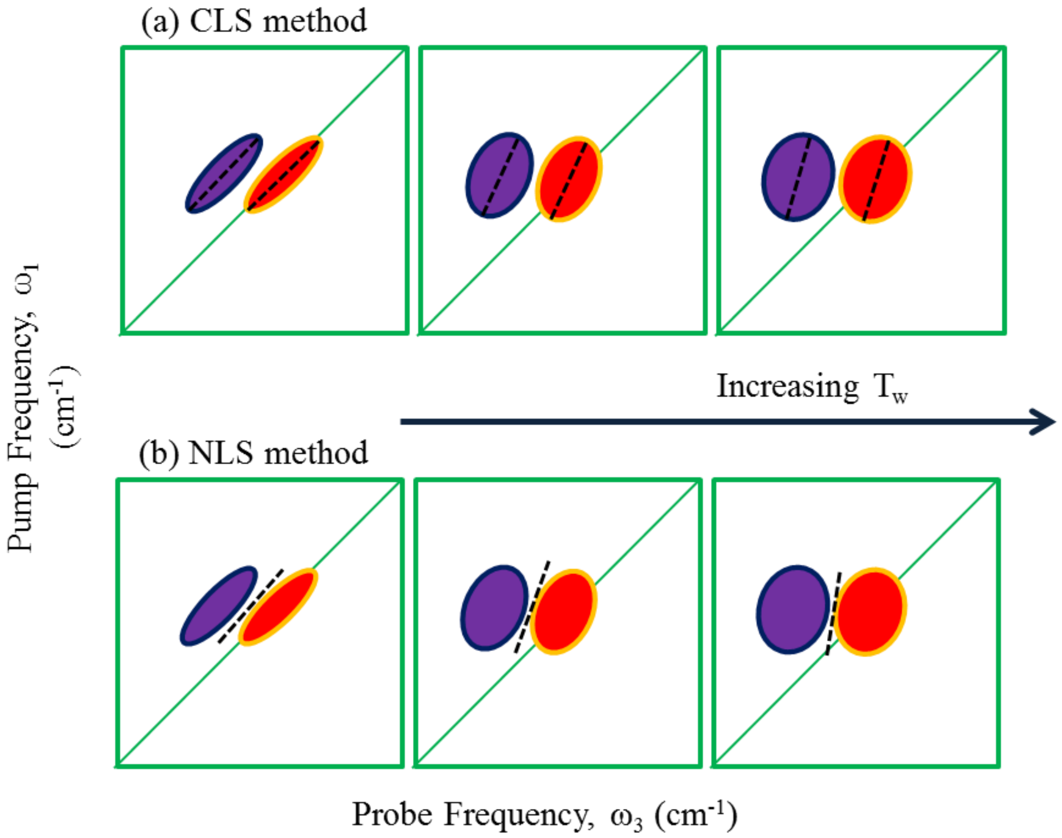


Figure (14): Illustrations of (a) central line slope (CLS) and (b) nodal line slope methods of spectral diffusion analysis. The dashed lines indicate either the central or nodal lines from which lineshape correlation is extracted.

For example, using the CLS approach, the ν = 0-1 and ν = 1-2 bands are each overlaid with a straight line that passes through their absolute peak maxima on each pump slice. When $T_w = 0$ and no spectral diffusion has occurred, these lines will theoretically have a gradient of 1 with respect to the pump and probe axes; the lines will be parallel to the $\omega_1 = \omega_3$ diagonal. As oscillators interconvert between inhomogeneous substates, the central line through peak maxima becomes increasingly orientated with the pump axis, i.e. vertical. The correlation is therefore proportional to the inverse gradient of the central line, yielding a correlation of 0 when the central line has a gradient of ∞.

The nodal line slope method applies a similar principle to CLS. Here, the point is located on each pump axis where the negative ν = 0-1 feature changes sign to the positive ν = 1-2 signal. The inverse gradient of the straight line connecting these points resultantly gives the spectral correlation, as in CLS analysis.

Spectral diffusion analysis can also be performed by calculating band ellipticity[86,157], or by fitting two-dimensional Gaussian distributions[59,158,159] to each spectral peak. In the case of Gaussian fitting, covariant Gaussian functions along the pump and probe axes are used to reproduce the 2D lineshape, and the correlation is obtained from the degree of covariance between the functions.

In principle, these methods each provide identical information, though in practice the specific vibrational spectrum under study may limit the approach that can be taken. For example, when analysing a weak vibration with low signal to noise, the peak centre at long T_w may not be well defined, and the CLS method may provide a more approximate description of the band correlation. On the other hand, in a spectrum with several closely-packed bands, fitting of 2D-Gaussian functions becomes unreliable, as the fitting function must be expanded to fit the ν = 0-1 and ν = 1-2 bands of each overlapping peak simultaneously. In any case, each of these methods yields a 2D lineshape correlation at each value of T_w.

By measuring the spectral correlation across a range of waiting times, we construct a profile of the time-dependent decay of correlation from the system (Figure (14)). This plot describes the frequency-frequency correlation function (FFCF)[91,154] of the molecular system; the time-dependent changes in vibrational frequency that occurs through spectral diffusion.

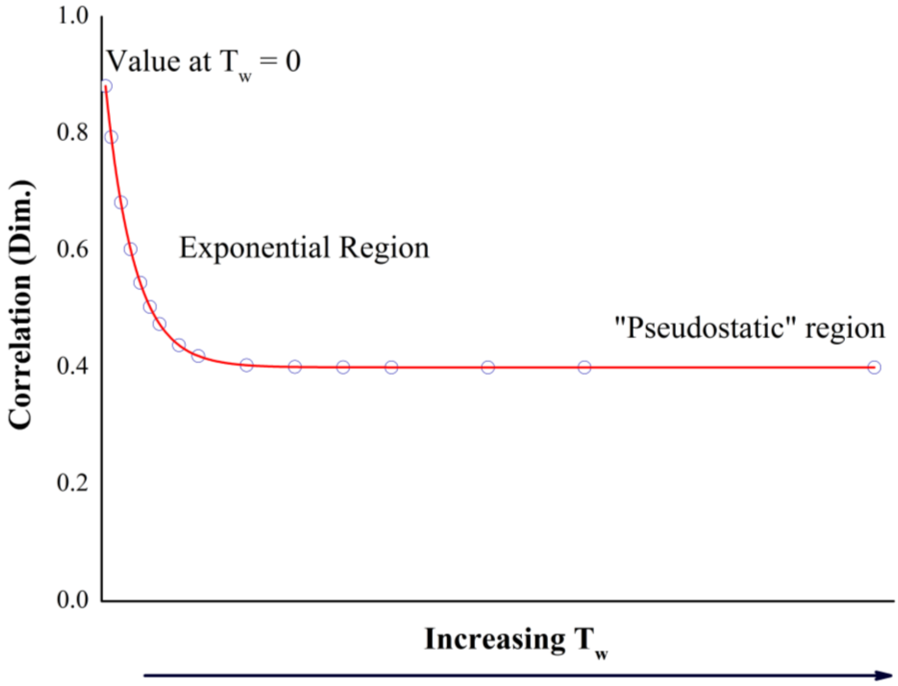


Figure (14): A schematic plot of spectral diffusion, showing the evolution of lineshape correlation within the measured range of waiting times. Information may be inferred from three regions: the exponential region, pseudostatic region and the initial correlation at the first T_w (shown as $T_w = 0$).

The FFCF is sensitive to structural reconfiguration around the excited vibration, and allows characterisation of the characteristic timescales of sub-state interconversion *via* the spectral diffusion rate. From this, the IR reporter group can provide information on hydrogen bonding dynamics, side chain motion or solvent interactions which may have functional significance in the case that the probe is deployed at a protein's substrate binding region.

Interpretation of the FFCF will be explained in greater detail in Chapter (3) where the lineshape evolution of a nitric oxide probe molecule is used to extract information from the active site of a haem protein, though broad overview will now be provided. Information can be extracted from three regions of the FFCF, foremost of which are the changes that occur across the measured range T_w. In this region the FFCF can be described using exponential functions, from which number of exponential terms and their respective time-constants can be used to isolate the contributions to spectral diffusion and the rate at which these occur. In turn, these can be assigned to dynamic processes in the vicinity of the excited vibration[77,149,160]. The second region of interest is the so-called pseudostatic region[161,159,162], represented by the static offset of the exponential function. Structural states that are not explored within T_w contribute to this offset, and so the contribution of slow dynamic processes to spectral diffusion may be inferred from this pseudostatic region. Reciprocally, dynamics may occur over more rapid timescales than can be measured. Waiting times of zero are theoretically useful, though in practice the overlap of the pump and probe pulses leads to coherent effects that add to the complexity of the nonlinear response. From this, the earliest measurable waiting time is dictated by the temporal width of the pulses; typically ~200 fs. Fast effects such as electronic fluctuation will reduce correlation of the 2DIR band at the earliest measurable waiting time. The lineshape at the first measured value of T_w may therefore reflect rapid structural fluctuations. In this type of analysis, the vibrational lifetime of the probe group dictates the temporal range over which dynamics can be observed. Reporters with short-lived vibrations will reduce more dynamic processes to static offsets in the FFCF, and thus prevent useful quantitative description of structural fluctuation timescales.

Structural information lends vital context to interpretation of the FFCF. Correlating lineshape changes to physical interactions is most readily achieved when the environment of the probe is known to some degree. For example, if the FFCF in

Figure (14) represents the spectral diffusion of a simple ion in aqueous solution, the exact interpretation of the result is entirely different than the case where the data originates from a cyano-labelled amino acid side chain in a substrate binding site. In protein studies, X-ray crystallography is therefore a powerful addition to spectral diffusion analysis, as it resolves potentially complex structural environments and allows greater detail in the assignment of dynamics. In this sense, static structural information reveals which structures are located in close proximity to the reporter group, and 2DIR reveals how these bodies move.

The concepts included in this chapter do not constitute an exhaustive review of 3rd-order spectroscopy, but provide the theoretical background for the following experiments, whose focus is to extract local dynamics from functionally-significant protein sites, using IR probe molecules in conjunction with 2DIR spectroscopy.

1.6 References

1. Weibel, E. R., Taylor, C. R. & Hoppeler, H. The concept of symmorphosis: a testable hypothesis of structure-function relationship. *Proc. Natl. Acad. Sci.* **88,** 10357–10361 (1991).
2. Strandberg, B., Dickerson, R. E. & Rossmann, M. G. 50 Years of protein structure analysis. *J. Mol. Biol.* **392,** 2–32 (2009).
3. Smyth, M. S. & Martin, J. H. J. X-Ray Crystallography. *Mol. Pathol.* **53,** 8 (2000).
4. Ilari, A. & Savino, C. in *Bioinformatics* (ed. Keith, J. M.) **452,** 63–87 (Humana Press, 2008).
5. Blow, D. *Outline of crystallography for biologists*. (Oxford University Press, 2002).
6. Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G. & Wyckoff, H. A three-dimensional model of the Myoglobin molecule obtained by X-ray analysis. *Nature* **181,** 662–666 (1958).
7. Adams, M. J. *et al.* Structure of rhombohedral 2 zinc insulin crystals. *Nature* **224,** 491–495 (1969).
8. Davis, A. M., Teague, S. J. & Kleywegt, G. J. Application and limitations of X-ray crystallographic data in structure-based ligand and drug design. *Angew. Chem. Int. Ed.* **42,** 2718–2736 (2003).

9. Fraser, J. S. *et al.* Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proc. Natl. Acad. Sci.* **108,** 16247–16252 (2011).
10. Petsko, G. A. & Ringe, D. *Protein structure and function*. (New Science Press, 2004).
11. Jorgensen, W. L. Rusting of the lock and key model for protein-ligand binding. *Science*, **254**, 954-955, 1991.
12. Koshland Jr, D. E. Application of a theory of enzyme specificity to protein synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **44,** 98 (1958).
13. Monod, J., Wyman, J. & Changeux, J.-P. On the nature of allosteric transitions : a plausible model. *J Mol Biol* **12,** 88-188, (1965).
14. Cooper, G. M. & Hausman, R. E. *The cell: a molecular approach*. (ASM Press , 2007).
15. Straub, F. B. & Szabolcsi, G. On the dynamic aspects of protein structure. *Mol. Biol. Probl. Perspect.* 182–187 (1964).
16. Vogt, A. D. & Di Cera, E. Conformational selection or induced fit? A critical appraisal of the kinetic mechanism. *Biochemistry (Mosc.)* **51,** 5894–5902 (2012).
17. Gianni, S., Dogan, J. & Jemth, P. Distinguishing induced fit from conformational selection. *Biophys. Chem.* **189,** 33–39 (2014).
18. Frauenfelder, H., Sligar, S. G. & Wolynes, P. G. The energy landscapes and motions of proteins. *Science* **254,** 1598–1603 (1991).
19. Alberts, B. *et al. Molecular biology of the cell*. (Garland Science, 2008).
20. Carpenter, E. P., Beis, K., Cameron, A. D. & Iwata, S. Overcoming the challenges of membrane protein crystallography. *Curr. Opin. Struct. Biol.* **18,** 581–586 (2008).
21. Li, D., Boland, C., Walsh, K. & Caffrey, M. Use of a robot for high-throughput crystallization of membrane proteins in lipidic mesophases. *J. Vis. Exp.* (2012).
22. Totir, M. *et al.* Macro-to-micro structural proteomics: native source proteins for high-throughput crystallization. *PLoS ONE* **7,** (2012).
23. Luft, J. R., Snell, E. H. & DeTitta, G. T. Lessons from high-throughput protein crystallization screening: 10 years of practical experience. *Expert Opin. Drug Discov.* **6,** 465–480 (2011).
24. Hui, R. & Edwards, A. High-throughput protein crystallization. *J. Struct. Biol.* **142,** 154–161 (2003).
25. Nagel, R. M., Luft, J. R. & Snell, E. H. AutoSherlock: a program for effective crystallization data analysis. *J. Appl. Crystallogr.* **41,** 1173–1176 (2008).
26. Abola, E., Kuhn, P., Earnest, T. & Stevens, R. C. Automation of X-ray crystallography. *Nat. Struct. Biol.* **7 Suppl,** 973–977 (2000).

27. Kay, L. E. NMR studies of protein structure and dynamics. *J. Magn. Reson.* **173,** 193–207 (2005).
28. Göbl, C. & Tjandra, N. Application of solution NMR spectroscopy to study protein dynamics. *Entropy* **14,** 581–598 (2012).
29. Fenwick, R. B., van den Bedem, H., Fraser, J. S. & Wright, P. E. Integrated description of protein dynamics from room-temperature X-ray crystallography and NMR. *Proc. Natl. Acad. Sci.* **111,** E445–E454 (2014).
30. Clore, G. M. & Gronenborn, A. M. Determination of three-dimensional structures of proteins and nucleic acids in solution by nuclear magnetic resonance spectroscopy. *Crit. Rev. Biochem. Mol. Biol.* **24,** 479–564 (1989).
31. Henzler-Wildman, K. A. *et al.* A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* **450,** 913–916 (2007).
32. Gutteridge, A. & Thornton, J. Conformational change in substrate binding, catalysis and product release: an open and shut case? *FEBS Lett.* **567,** 67–73 (2004).
33. Kay, L. E. Protein dynamics from NMR. *Biochem. Cell Biol.* **76,** 145–152 (1998).
34. Boehr, D. D., Dyson, H. J. & Wright, P. E. An NMR perspective on enzyme dynamics. *Chem. Rev.* **106,** 3055–3079 (2006).
35. 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.* **36,** 723–730 (2003).
36. Ode, H., Nakashima, M., Kitamura, S., Sugiura, W. & Sato, H. Molecular dynamics simulation in virus research. *Front. Microbiol.* **3,** (2012).
37. Wand, J. A., Dynamic activation of protein function: a view emerging from NMR spectroscopy. *Nat. Struct. Biol.* **8,** 926–931 (2001).
38. Hartl, F. U., Bracher, A. & Hayer-Hartl, M. Molecular chaperones in protein folding and proteostasis. *Nature* **475,** 324–332 (2011).
39. He, Y., Chen, J. Y., Knab, J. R., Zheng, W. & Markelz, A. G. Evidence of protein collective motions on the picosecond timescale. *Biophys. J.* **100,** 1058–1065 (2011).
40. Goh, C. S., Milburn, D. & Gerstein, M. Conformational changes associated with protein–protein interactions. *Curr. Opin. Struct. Biol.* **14,** 104–109 (2004).
41. Thielges, M. C., Chung, J. K. & Fayer, M. D. Protein dynamics in cytochrome P450 molecular recognition and substrate specificity using 2D IR vibrational echo spectroscopy. *J. Am. Chem. Soc.* **133,** 3995–4004 (2011).
42. Mittermaier, A. The response of internal dynamics to hydrophobic core mutations in the SH3 domain from the Fyn tyrosine kinase. *Protein Sci.* **13,** 1088–1099 (2004).

43. Xiao, Y. *et al.* Phosphorylation releases constraints to domain motion in ERK2. *Proc. Natl. Acad. Sci.* **111,** 2506–2511 (2014).
44. Lorimier, R. D., Hellinga, H. W. & Spicer, L. D. NMR studies of structure, hydrogen exchange, and main-chain dynamics in a disrupted-core mutant of thioredoxin. *Protein Sci.* **5,** 2552–2565 (1996).
45. Verma, D., Jacobs, D. J. & Livesay, D. R. Changes in lysozyme flexibility upon mutation are frequent, large and long-ranged. *PLoS Comput. Biol.* **8,** (2012).
46. Azevedo, A. M., Cabral, J. M. S. & Fonseca, L. P. Horseradish peroxidase : a valuable tool in biotechnology. *Biotech. Ann. Rev.* **9,** (2003).
47. Derat, E. & Shaik, S. The Poulos−Kraut mechanism of compound I formation in horseradish peroxidase: a QM/MM study. *J. Phys. Chem. B* **110,** 10526–10533 (2006).
48. Shaw, D. E. *et al.* atomic-level characterization of the structural dynamics of proteins. *Science* **330,** 341–346 (2010).
49. McCammon, J. A., Wolynes, P. G. & Karplus, M. Picosecond dynamics of tyrosine side chains in proteins. *Biochemistry (Mosc.)* **18,** 927–942 (1979).
50. Vos, M. H. & Martin, J. L. Femtosecond processes in proteins. *Biochim. Biophys. Acta.* **1411,** 1–20 (1999).
51. Brooks, C. L., Gruebele, M., Onuchic, J. N. & Wolynes, P. G. Chemical physics of protein folding. *Proc. Natl. Acad. Sci.* **95,** 11037–11038 (1998).
52. Pal, S. K., Peon, J., Bagchi, B. & Zewail, A. H. Biological water: femtosecond dynamics of macromolecular hydration. *J. Phys. Chem. B* **106,** 12376–12395 (2002).
53. Bakulin, A. A. *et al.* Dynamics of water confined in reversed micelles: multidimensional vibrational spectroscopy study. *J Phys Chem B* **117,** 15545–15558 (2013).
54. Ramasesha, K., De Marco, L., Mandal, A. & Tokmakoff, A. Water vibrations have strongly mixed intra- and intermolecular character. *Nat. Chem.* **5,** 935–940 (2013).
55. Keller, U. Recent developments in compact ultrafast lasers. *Nature* **424,** 831–838 (2003).
56. Hochstrasser, R. M. Two-dimensional spectroscopy at infrared and optical frequencies. *Proc. Natl. Acad. Sci.* **104,** 14190–14196 (2007).
57. Thielges, M. C. & Fayer, M. D. Protein dynamics studied with ultrafast two-dimensional infrared vibrational echo spectroscopy. *Acc. Chem. Res.* **45,** 1866–1874 (2012).
58. Koziński, M., Garrett-Roe, S. & Hamm, P. Vibrational spectral diffusion of CN− in water. *Chem. Phys.* **341,** 5–10 (2007).

59. Asbury, J. B. *et al.* Hydrogen bond breaking probed with multidimensional stimulated vibrational echo correlation spectroscopy. *J. Chem. Phys.* **119,** 12981 (2003).
60. Hamm, P., Lim, M. & Hochstrasser, R. M. Structure of the amide I band of peptides measured by femtosecond nonlinear-infrared spectroscopy. *J. Phys. Chem. B* **102,** 6123–6138 (1998).
61. Woutersen, S. & Hamm, P. Isotope-edited two-dimensional vibrational spectroscopy of trialanine in aqueous solution. *J. Chem. Phys.* **114,** 2727 (2001).
62. Hamm, P., Lim, M., DeGrado, W. F. & Hochstrasser, R. M. The two-dimensional IR nonlinear spectroscopy of a cyclic penta-peptide in relation to its three-dimensional structure. *Proc. Natl. Acad. Sci.* **96,** 2036–2041 (1999).
63. Woutersen, S., Mu, Y., Stock, G. & Hamm, P. Hydrogen-bond lifetime measured by time-resolved 2D-IR spectroscopy: N-methylacetamide in methanol. *Chem. Phys.* **266,** 137–147 (2001).
64. Woutersen, S., Mu, Y., Stock, G. & Hamm, P. Subpicosecond conformational dynamics of small peptides probed by two-dimensional vibrational spectroscopy. *Proc. Natl. Acad. Sci.* **98,** 11254–11258 (2001).
65. Merchant, K. A. *et al.* Myoglobin-CO conformational substate dynamics: 2D vibrational echoes and MD simulations. *Biophys. J.* **82,** 3277–3288 (2002).
66. Ishikawa, H., Kwak, K., Chung, J. K., Kim, S. & Fayer, M. D. Direct observation of fast protein conformational switching. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 8619–8624 (2008).
67. Merchant, K. A. *et al.* Myoglobin-CO substate structures and dynamics : multidimensional vibrational echoes and molecular dynamics simulations. *J Am Chem Soc* **125,** 13804–13818 (2003).
68. Kuroda, D. G. *et al.* Snapshot of the equilibrium dynamics of a drug bound to HIV-1 reverse transcriptase. *Nat. Chem.* **5,** 174–181 (2013).
69. Simpson, N. *et al.* The effect on structural and solvent water molecules of substrate binding to ferric horseradish peroxidase. *Faraday Discuss.* **177**, 163-179 (2014).
70. Finkelstein, I. J., Ishikawa, H., Kim, S., Massari, A. M. & Fayer, M. D. Substrate binding and protein conformational dynamics measured by 2D-IR vibrational echo spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 2637–2642 (2007).
71. Gotoh, O. Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *J. Biol. Chem.* **267,** 83–90 (1992).
72. Bagchi, S., Thorpe, D. G., Thorpe, I. F., Voth, G. A. & Fayer, M. D. Conformational switching between protein substates studied with 2D IR

vibrational echo spectroscopy and molecular dynamics simulations. *J Phys Chem B* **114,** 17187–17193 (2010).

73. Adamczyk, K. *et al.* The effect of point mutation on the equilibrium structural fluctuations of ferric Myoglobin. *Phys. Chem. Chem. Phys. PCCP* **14,** 7411–7419 (2012).
74. Strasfeld, D. B., Ling, Y. L., Gupta, R., Raleigh, D. P. & Zanni, M. T. Strategies for extracting structural information from 2D IR spectroscopy of amyloid: application to islet amyloid polypeptide. *J. Phys. Chem. B* **113,** 15679–15691 (2009).
75. Falvo, C. *et al.* Frequency distribution of the amide-I vibration sorted by residues in amyloid fibrils revealed by 2D-IR measurements and simulations. *J. Phys. Chem. B* **116,** 3322–3330 (2012).
76. Wang, L. *et al.* 2DIR Spectroscopy of human amylin fibrils reflects stable β-sheet structure. *J. Am. Chem. Soc.* **133,** 16062–16071 (2011).
77. 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* **117,** 15804–15811 (2013).
78. 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.* **108,** 6115–6120 (2011).
79. Adamczyk, K. *et al.* Ultrafast infrared spectroscopy reveals water-mediated coherent dynamics in an enzyme active site. *Chem Sci* **6,** 505–516 (2015).
80. Rubtsov, I. V., Wang, J. & Hochstrasser, R. M. Dual-frequency 2D-IR spectroscopy heterodyned photon echo of the peptide bond. *Proc. Natl. Acad. Sci.* **100,** 5601–5606 (2003).
81. Gnanakaran, S., Hochstrasser, R. M. & García, A. E. Nature of structural inhomogeneities on folding a helix and their influence on spectral measurements. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 9229–9234 (2004).
82. Kim, Y. S., Wang, J. & Hochstrasser, R. M. Two-dimensional infrared spectroscopy of the alanine dipeptide in aqueous solution. *J. Phys. Chem. B* **109,** 7511–7521 (2005).
83. Bagchi, S., Nebgen, B. T., Loring, R. F. & Fayer, M. D. Dynamics of a myoglobin mutant enzyme: 2D IR vibrational echo experiments and simulations. *J. Am. Chem. Soc.* **132,** 18367–18376 (2010).
84. Hollas, J. M. *Modern spectroscopy*. (J. Wiley, 2004).
85. Reddy, V., K. *Symmetry and spectroscopy of molecules*. (New Age International, 2009).

86. Khalil, M., Demirdöven, N. & Tokmakoff, a. Coherent 2D IR spectroscopy: molecular structure and dynamics in solution. *J. Phys. Chem. A* **107,** 5258–5279 (2003).

87. Zhang, M. *et al.* Temperature dependence of IR absorption of hydrous/hydroxyl species in minerals and synthetic materials. *Am. Mineral.* **92,** 1502–1517 (2007).

88. Stuart, B. *Infrared spectroscopy: fundamentals and application*. **1,** (Wiley, New York, 2004).

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

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

91. Hamm, P. & Zanni, M. T. *Concepts and methods of 2D infrared spectroscopy*. (Cambridge University Press, 2011).

92. Bloembergen, N. Nonlinear optics. *World Sci. Publ.* (1965).

93. Bishop, D. M. & Cheung, L. M. Vibrational contribution to molecular dipole polarisabilities. *J Phys Chem Ref Data* **11,** (1982).

94. Pethrick, R. A. & Richards, R. W. *Static and dynamic properties of the polymeric solid state*. (Proceedings of the NATO Advanced Study Institute, 1981).

95. Jackson, J. D. *Classical electrodynamics*. (Wiley, New York, 1998).

96. Fecko, C. J., Loparo, J. J., Roberts, S. T. & Tokmakoff, A. Local hydrogen bonding dynamics and collective reorganization in water: ultrafast infrared spectroscopy of HOD/D_2O. *J. Chem. Phys.* **122,** (2005).

97. Marshall, A. G. & Roe, D. C. Dispersion versus absorption: spectral line shape analysis for radiofrequency and microwave spectrometry. *Anal. Chem.* **50,** 756–763 (1978).

98. Sathyanarayana, D. N. *Vibrational spectroscopy: theory and applications*. (New Age International, 2007).

99. Zewail, A., H. Optical Molecular Dephasing: Principles of and Probings by Coherent Laser Spectroscopy. *Acc. Chem. Res.* **13,** 353–360 (1980).

100. Khalil, M., Demirdöven, N. & Tokmakoff, A. Vibrational coherence transfer characterized with fourier-transform 2D IR spectroscopy. *J. Chem. Phys.* **121,** 362–73 (2004).

101. Berg, M. A. in *Ultrafast Infrared And Raman Spectroscopy* (CRC Press, 2001).

102. Fulmer, E. C., Ding, F. & Zanni, M. T. Heterodyned fifth-order 2D-IR spectroscopy of the azide ion in an ionic glass. *J. Chem. Phys.* **122,** (2005).

103. Gershgoren, E., Wang, Z., Ruhman, S., Vala, J. & Kosloff, R. Investigating pure vibrational dephasing of I3- in solution: Temperature dependence of T2*for the fundamental and first harmonic of ν1. *J. Chem. Phys.* **118,** 3660 (2003).

104. Kubo, R. in *Advances in Chemical Physics* 101–127 (John Wiley & Sons, Inc., 2007).

105. Hamm, P., Lim, M. & Hochstrasser, R. M. Non-markovian dynamics of the vibrations of ions in water from femtosecond infrared three-pulse photon echoes. *Phys. Rev. Lett.* **81,** 5326 (1998).

106. Tokmakoff, A. *et al.* Vibrational spectral diffusion and population dynamics in a glass-forming liquid: Variable bandwidth picosecond infrared spectroscopy. *J. Chem. Phys.* **102,** 3919–3931 (1995).

107. Finkelstein, I. J. *et al.* Probing dynamics of complex molecular systems with ultrafast 2D IR vibrational echo spectroscopy. *Phys. Chem. Chem. Phys. PCCP* **9,** 1533–1549 (2007).

108. Bandaria, J. N. *et al.* Characterizing the dynamics of functionally relevant complexes of formate dehydrogenase. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 17974–17979 (2010).

109. Choi, J. H., Kwak, K. & Cho, M. Computational IR and 2D IR photon echo spectroscopy of both wild-type and double mutant myoglobin-CO proteins. *J. Phys. Chem. B* **117,** 15462–15478 (2013).

110. Fayer, M. D. Dynamics of liquids, molecules, and proteins measured with ultrafast 2D IR vibrational echo chemical exchange spectroscopy. *Annu. Rev. Phys. Chem.* **60,** 21–38 (2009).

111. Floudas, C. A., Fung, H. K., McAllister, S. R., Mönnigmann, M. & Rajgaria, R. Advances in protein structure prediction and de novo protein design: A review. *Chem. Eng. Sci.* **61,** 966–988 (2006).

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

113. Krimm, S. Vibrational analysis of conformation in peptides, polypeptides, and proteins. *Biopolymers* **22**, 217-225 (1983)

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

115. Haris, P. I. Characterization of protein structure and stability using fourier transform infrared spectroscopy. *Pharm. Pharmacol. Commun* 15–25 (1999).

116. Barth, A. The infrared absorption of amino acid side chains. *Prog. Biophys. Mol. Biol.* **74,** 141–173 (2000).

117. Wolpert, M. & Hellwig, P. Infrared spectra and molar absorption coefficients of the 20 alpha amino acids in aqueous solutions in the spectral range from 1800 to 500 cm^{-1}. *Spectrochim. Acta. A.* **64,** 987–1001 (2006).

118. Wong, M. W. Vibrational frequency prediction using density functional theory. *Chem. Phys. Lett.* **256**, 391-399 (1996).

119. Thomas, M., Brehm, M., Fligg, R., Vöhringer, P. & Kirchner, B. Computing vibrational spectra from ab initio molecular dynamics. *Phys. Chem. Chem. Phys.* **15,** 6608-6622 (2013).

120. Surewicz, W. K., Mantsch, J. H. H. & Chapman, D., Determination of protein secondary structure by fourier transform infrared spectroscopy : a critical assessment. *Biochemistry* **32,** (1993).

121. Tucker, M. J. *et al.* Nonequilibrium dynamics of helix reorganization observed by transient 2D IR spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 17314–17319 (2013).

122. Fang, C. & Hochstrasser, R. M. Two-dimensional infrared spectra of the ^{13}C=^{18}O isotopomers of alanine residues in an α-helix. *J. Phys. Chem. B* **109,** 18652–18663 (2005).

123. Maekawa, H., Ballano, G., Formaggio, F., Toniolo, C. & Ge, N. H. ^{13}C= ^{18}O/^{15}N Isotope dependence of the amide-I/II 2D IR cross peaks for the fully extended peptides. *J. Phys. Chem. C* **118,** 29448–29457 (2014).

124. Takahashi, R. *et al.* FTIR study on the hydrogen bond structure of a key tyrosine residue in the flavin-binding blue light sensor TePixD from Thermosynechococcus elongatus. *Biochemistry (Mosc.)* **46,** 6459–6467 (2007).

125. Choi, J. H., Raleigh, D. & Cho, M. Azido homoalanine is a useful infrared probe for monitoring local electrostatistics and sidechain solvation in proteins. *J. Phys. Chem. Lett.* **2,** 2158–2162 (2011).

126. Bloem, R. *et al.* Ligand binding studied by 2D IR spectroscopy using the azidohomoalanine label. *J. Phys. Chem. B* **116,** 13705–13712 (2012).

127. Spiro, T. G., Soldatova, A. V. & Balakrishnan, G. CO, NO and O2 as vibrational probes of heme protein interactions. *Coord. Chem. Rev.* **257,** 511–527 (2013).

128. Henriksen, A. *et al.* Structural interactions between horseradish peroxidase C and the substrate benzhydroxamic acid determined by X-ray crystallography. *Biochemistry (Mosc.)* **37,** 8054–8060 (1998).

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

130. Cho, M., Scherer, N. F., Fleming, G. R. & Mukamel, S. Photon echoes and related four-wave-mixing spectroscopies using phase-locked pulses. *J. Chem. Phys.* **96,** 5618–5629 (1992).

131. Cervetto, V., Helbing, J., Bredenbeck, J. & Hamm, P. Double-resonance versus pulsed fourier transform two-dimensional infrared spectroscopy: an experimental and theoretical comparison. *J. Chem. Phys.* **121,** 5935–5942 (2004).

132. Adamczyk, K. *et al.* Measuring protein dynamics with ultrafast two-dimensional infrared spectroscopy. *Meas. Sci. Technol.* **23,** 062001 (2012).

133. Deflores, L. P., Nicodemus, R. A. & Tokmakoff, A. Two-dimensional fourier transform spectroscopy in the pump-probe geometry. *Opt. Lett.* **32,** 2966–8 (2007).

134. Kuehn, W., Reimann, K., Woerner, M. & Elsaesser, T. Phase-resolved two-dimensional spectroscopy based on collinear n-wave mixing in the ultrafast time domain. *J. Chem. Phys.* **130,** 164503 (2009).

135. Agarwal, R., Prall, B. S., Rizvi, A. H., Yang, M. & Fleming, G. R. Two-color three pulse photon echo peak shift spectroscopy. *J. Chem. Phys.* **116,** 6243 (2002).

136. Cho, M. *et al.* The integrated photon echo and solvation dynamics. *J. Phys. Chem.* **100,** 11944–11953 (1996).

137. Plutzer, C. & Kleinermanns, K. Tautomers and electronic states of jet-cooled adenine investigated by double resonance spectroscopy. *Phys. Chem. Chem. Phys.* **4,** 4877–4882 (2002).

138. Zhang, Z., Wells, K. L., Seidel, M. T. & Tan, H.-S. Fifth-order three-dimensional electronic spectroscopy using a pump–probe configuration. *J. Phys. Chem. B* **117,** 15369–15385 (2013).

139. Teich, M. C. Infrared heterodyne detection. *Proc. IEEE* **56,** 37–46 (1968).

140. Xiong, W., Laaser, J. E., Mehlenbacher, R. D. & Zanni, M. T. Adding a dimension to the infrared spectra of interfaces using heterodyne detected 2D sum-frequency generation (HD 2D SFG) spectroscopy. *Proc. Natl. Acad. Sci.* **108,** 20902–20907 (2011).

141. Noda, I. Two-Dimensional infrared (2D IR) spectroscopy: theory and applications. *Appl. Spectrosc.* **44,** 550–561 (1990).

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

143. Zanni, M. T. & Hochstrasser, R. M. Two-dimensional infrared spectroscopy: a promising new method for the time resolution of structures. *Curr. Opin. Struct. Biol.* **11,** 516–522 (2001).

144. Strasfeld, D., Shim, S.-H. & Zanni, M. Controlling vibrational excitation with shaped mid-IR pulses. *Phys. Rev. Lett.* **99,** 038102 (2007).

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

146. Skinner, J. L., Auer, B. M. & Lin, Y. S. Vibrational line shapes, spectral diffusion, and hydrogen bonding in liquid water. *Adv. Chem. Phys.* **142,** 59 (2009).
147. Kuroda, D. G. & Hochstrasser, R. M. Dynamic structures of aqueous oxalate and the effects of counterions seen by 2D IR. *Phys. Chem. Chem. Phys. PCCP* **14,** 6219–24 (2012).
148. Wong, D. B. *et al.* Water dynamics in water/DMSO binary mixtures. *J. Phys. Chem. B* **116,** 5479–5490 (2012).
149. Ishikawa, H. *et al.* Neuroglobin dynamics observed with ultrafast 2D-IR vibrational echo spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 16116–16121 (2007).
150. Bagchi, S., Boxer, S. G. & Fayer, M. D. Ribonuclease S dynamics measured using a nitrile label with 2D IR vibrational echo spectroscopy. *J. Phys. Chem. B* **116,** 4034–4042 (2012).
151. Mukherjee, P., Kass, I., Arkin, I. T. & Zanni, M. T. Picosecond dynamics of a membrane protein revealed by 2D IR. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 3528–3533 (2006).
152. Chung, J. K., Thielges, M. C. & Fayer, M. D. Conformational dynamics and stability of HP35 studied with 2D IR vibrational echoes. *J. Am. Chem. Soc.* **134,** 12118–12124 (2012).
153. Kim, Y. S., Liu, L., Axelsen, P. H. & Hochstrasser, R. M. 2D IR provides evidence for mobile water molecules in beta-amyloid fibrils. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 17751–17756 (2009).
154. Kwak, K., Park, S., Finkelstein, I. J. & Fayer, M. D. Frequency-frequency correlation functions and apodization in two-dimensional infrared vibrational echo spectroscopy: A new approach. *J. Chem. Phys.* **127,** 124503 (2007).
155. Moran, S. D. *et al.* Two-dimensional IR spectroscopy and segmental ^{13}C labeling reveals the domain structure of human γD-crystallin amyloid fibrils. *Proc. Natl. Acad. Sci.* **109,** 3329–3334 (2012).
156. Lee, M. W., Carr, J. K., Göllner, M., Hamm, P. & Meuwly, M. 2D IR spectra of cyanide in water investigated by molecular dynamics simulations. *J. Chem. Phys.* **139,** (2013).
157. Okumura, K., Tokmakoff, A. & Tanimura, Y. Two-dimensional line-shape analysis of photon-echo signal. *Chem. Phys. Lett.* **314** 488–495 (1999).
158. Yu, P., Yang, F., Zhao, J. & Wang, J. Hydration dynamics of cyanoferrate anions examined by ultrafast infrared spectroscopy. *J. Phys. Chem. B* **118,** 3104–14 (2014).

159. Hunt, N. T., Greetham, G. M., Towrie, M., Parker, A. W. & Tucker, N. P. Relationship between protein structural fluctuations and rebinding dynamics in ferric haem nitrosyls. *Biochem. J.* **433,** 459–468 (2011).

160. Woutersen, S. *et al.* Peptide conformational heterogeneity revealed from nonlinear vibrational spectroscopy and molecular-dynamics simulations. *J. Chem. Phys.* **117,** 6833 (2002).

161. Kaziannis, S. *et al.* Femtosecond to Microsecond Photochemistry of a [FeFe] hydrogenase Enzyme Model Compound. *J. Chem. Phys.* **136** 15370–15379 (2010).

162. Brookes, J. F., Slenkamp, K. M., Lynch, M. S. & Khalil, M. Effect of solvent polarity on the vibrational dephasing dynamics of the nitrosyl stretch in an Fe(II) complex revealed by 2D IR spectroscopy. *J. Phys. Chem. A* **117,** 6234–6243 (2013).

2. Methods

Each experimental Chapter in this thesis utilises the core infrared techniques of FTIR, pump-probe and 2DIR spectroscopy. The presented results also rely on a number of supplemental methods that are used on a case-specific basis. Only core methods will be discussed in this section; supplemental techniques are listed within their appropriate experimental chapter, and are described in additional detail in the Appendix.

2.1 Linear IR Spectroscopy

FTIR spectrometers use a blackbody radiation source in conjunction with a Helium:Neon laser and a Michelson Interferometer[1] to record optical density as a function of Infrared frequency.

Initially, an electrical current is passed through a silicon carbide (Globar) element, causing emission of light with a continuous, blackbody distribution in the frequency range 1.4 – 16.7 μm (7,000 – 600 cm^{-1})[2]. Light from the source is emitted in all directions, and so, to produce a beam, the light is focused through a narrow aperture (4mm), after which a parabolic mirror is used to collimate the radiation. Following this, the beam is incident on a potassium bromide (KBr) beamsplitter that divides the light into two components of equal intensity, one of which propagates along the original beam path while the other is reflected at 90˚. Each of these rays travels through separate arms of a Michelson Interferometer (Figure (1)).

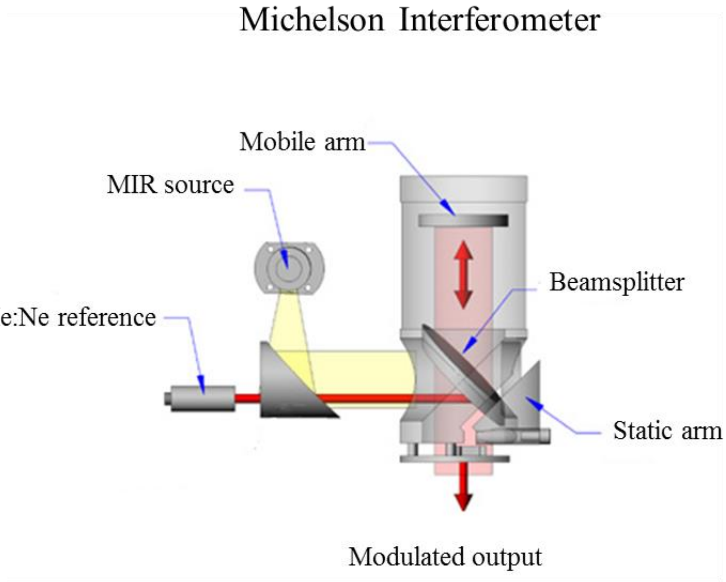


Figure (1): schematic of the Mid-Infrared Michelson Interferometer employed in FTIR spectroscopy. This design is reproduced in 2DIR spectroscopy, though the He:Ne reference is not required, as the mobile arm is moved in automated steps of known size (Image adapted from: http://www.gascell.com/htmls/primer.htm)

During measurements, periodic oscillation of the mobile arm of the interferometer is implemented via a motorised delay stage, so as to vary the path length between mobile and fixed arms at the scanning velocity. The beams are recombined after the interferometer, and travel collinearly towards the sample position. The intensity profile of recombined beam is dependent on the exact path difference between the interferometer arms, as each of its frequency components can interfere destructively or constructively. The constant scanning velocity of the mobile arm gives rise to a periodic, time-dependent intensity pattern (or interferogram) that correlates each frequency component to a point in time. In order to translate the interferogram into a frequency-domain signal, i.e. the FTIR frequency axis, it is necessary that the exact path difference of the interferometer arms be known at any specific time[2]. For this purpose, a helium neon (He:Ne) laser is used. The He:Ne beam is also propagated through the Michelson interferometer, subjecting it to the same path difference as the MIR light. The comparatively short wavelength (λ = 632.8 nm), low-bandwidth, high-intensity output of the He:Ne laser generates regular interferences fringes that are "counted" by the spectrometer, providing a reference for the path difference.

The absorbance, A (OD), of a sample is determined from:

$$A = \log_{10}\frac{I_0}{I} \tag{1}$$

where I_0 is the original intensity of the MIR light at an arbitrary frequency, collected in practice *via* a reference, or background, measurement. Absorption from a sample is further determined by properties of the sample by the Beer-Lambert Law:

$$A = \varepsilon N l \tag{2}$$

where N is the concentration of the infrared chromophore (molar), ε is the molar extinction coefficient ($M^{-1}cm^{-1}$) and l is the path length of the sample (cm).

In these investigations, all FTIR samples were housed between two CaF_2 windows separated by a polytetrafluoroethylene (PTFE) spacer that determined the path length, l, of the sample (Figure (2)). Specific path lengths and concentrations vary per experiment, and are listed in their corresponding chapters.

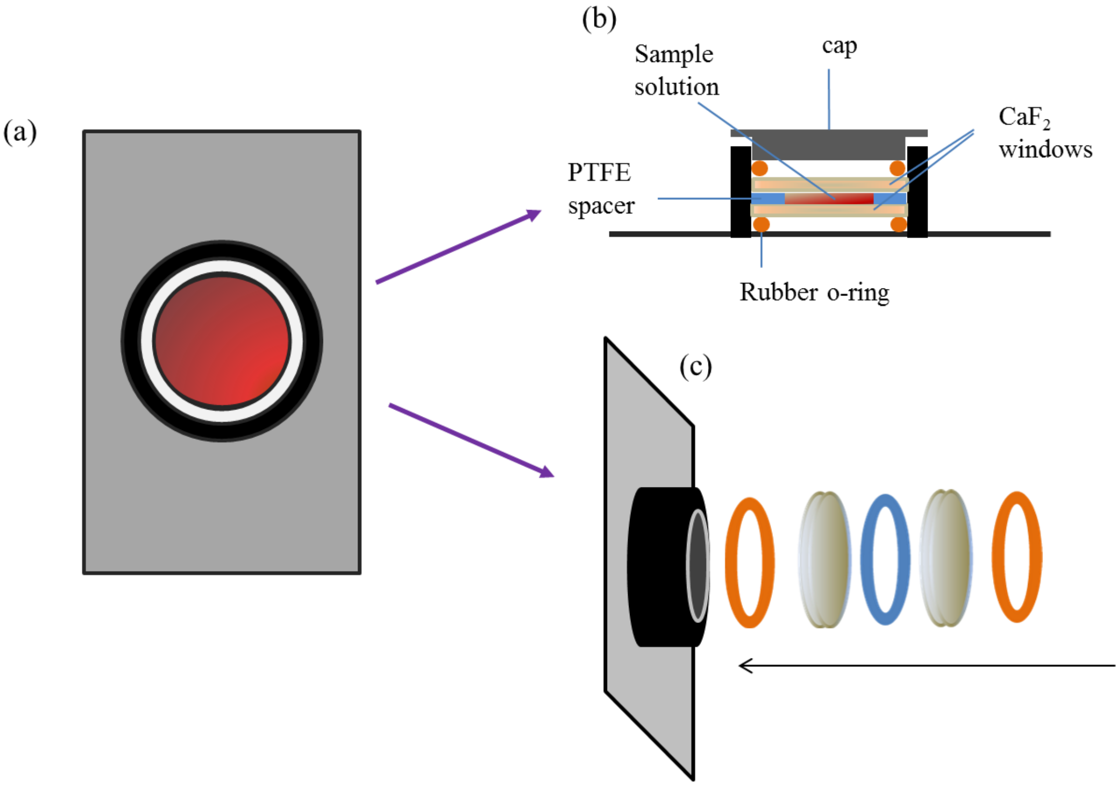


Figure (2): Illustration of the "Harrick" sample cells used for FTIR, pump-probe and 2DIR measurements. (a) Head-on view of the cell, (b) cross-section of the prepared cell and (c) an illustration of the assembly (excluding the sample, which is added after the PTFE spacer (blue)).

The linear absorption spectra in Chapters (3) and (4) of this thesis were collected using a Bruker Vertex 70 FTIR, while the spectra in Chapter (5) were collected using a Nicolet IS-10. All spectra were taken at room temperature (~23˚C), at a spectral resolution ≤ 1 cm^{-1}. The sample protocols used for FTIR spectroscopy (with respect to path length and concentration) were also observed in all time-resolved measurements.

2.2 Time-Resolved Spectroscopy

Two ultrafast lasers systems were used to collect the time-resolved data presented here; the ULTRA spectrometer[3,4] (Rutherford Appleton Laboratory) and the Ultrafast Chemical Physics (UCP) laser system (University of Strathclyde). Each of these setups is configured to carry out pump-probe and 2DIR spectroscopies in a similar fashion, though the performance of the two systems varies in a number of parameters. A single broad description will be provided to describe both setups from a qualitative standpoint, though quantitative differences will be specified.

Ultrashort pulses are produced by an oscillator with a titanium: sapphire (Ti^{3+}:Sapph) gain medium. Continuous-wave (CW) operation is achieved through pumping of the gain medium with the 2nd harmonic of a solid-state $Nd:YVO_4$ (neodymium-doped yttrium orthovanadate; $\lambda/2 = 532$ nm) pump laser. Modelocked operation, i.e. pulsed-output, is triggered through a Kerr-lensing effect in conjunction with "soft-aperture" modelocking, whereby the overlap between the pump beam and Ti:Sapph emission is enhanced at higher intensities[5]. Short duration pulses are of higher peak intensity than CW radiation, and so amplification is improved for broad-wavelength radiation. An acousto-optic modulator supplies an initial perturbation from which additional frequency side-bands are generated, biasing the amplification away from CW operation. Dispersion of the beam through the gain medium is balanced by a prism-pair in the oscillator cavity. Pulses from the UCP oscillator have a FWHM of ~70 nm, a duration of ~30 fs, and are produced at a repetition rate of ~80 MHz, while the custom-built ULTRA oscillator produces pulses with 20 fs duration and 50 nm bandwidth, at a repetition rate of 65 MHz. Pulses have a central wavelength of 800 nm in either configuration.

Pulse amplification in the UCP system is achieved with a Legend Elite regenerative amplifier (Coherent). Amplification serves to increase the average beam power from the oscillator from ~500 mW to ~2.5 W through the process of chirped pulse amplification[5]. The resultant pulse energies and repetition rates from the UCP amplifier are 2.5 mJ and 1kHz, respectively. In the ULTRA setup, pulses are produced with a 10 kHz repetition rate, with an energy of 0.8 mJ per pulse. This is achieved using a custom-built double-pass regenerative amplifier system, where the deleterious thermal effects of the high average laser power are mitigated through cryogenic cooling of the Ti:Sapph gain medium[6].

In each amplifier unit, ultrashort pulses from the oscillator are initially dispersed by a grating, though the resulting frequency components are reflected so as to travel collinearly, effectively stretching the temporal profile of the pulse. Temporal-stretching reduces the peak pulse intensity, minimising beam distortion through Kerr-lensing in the Ti:Sapph crystal, and also preventing damage to the gain medium. A single pulse is selected from the oscillator pulse train by means of polarisation, and is propagated into the linear cavity of the amplifier. The cavity termini consist of electro-optic modulators (Pockels cells) that allow switching of the pulse polarisation. As the pulse resonates between the two modulators it makes multiple passes through the gain medium.

In the UCP amplifier, the Ti:Sapph rod is pumped by a frequency-doubled Nd:YLF (Neodinium-doped yttrium lithium fluoride) laser with output power ~25 W and a central wavelength of 527 nm. Maximum amplification is achieved within 10 roundtrips of the cavity, after which the pulse polarisation is modulated by the Pockels cell, causing it to be ejected from the amplification cavity into a pulse compression unit. In this final compression stage, a second reflective grating is used to compensate for the initial pulse dispersion that occurred prior to amplification, compressing the amplified pulses to ≤40 fs. In the ULTRA system, the Ti:Sapph rod is pumped by two 55 W Nd:YAG lasers. The central wavelength and spectral width of the amplified pulses are 800 nm and ~70 nm for each laser system.

Amplified pulses are converted from 800nm to mid-infrared radiation by means of an optical parametric amplifier (OPA)[5,7,8]. Optical parametric amplification is performed to produce light resonant with the molecular vibrations under study. For the investigations conducted in this thesis, the frequency spectrum of MIR pulses was thus centred at either ~5.3 μm (~1890 cm^{-1}), or at ~6.2 μm (1560 cm^{-1}), for Chapter (3) and for Chapters (4) and (5), respectively.

In the process of optical parametric amplification, the 800 nm beam is first incident on a 50:50 beamsplitter, generating two pump beams. One of the pump beams (A) is incident on a further beamsplitter with significantly greater reflectivity, and only a small fraction (5%) of the pump intensity is transmitted through the optic. The weakly transmitted light is then focused onto a sapphire (Al_2O_3) plate. The high intensity of the focused beam results in a significant nonlinear contribution to the refractive index of the material, which further causes self-focusing of the beam. However, the pulse power at the beam centre is sufficiently great to cause multiphoton ionisation of the sapphire, creating free electrons in the crystal. In turn, this ionisation results in a negative contribution to refractive index, opposing beam self-focusing. Ultimately, the two effects balance to yield a narrow, high-intensity beam. The intensity-dependent component of refractive index introduces a corresponding intensity-dependent shift in the phase of the propagating field. In turn, this alters the rate of change of phase in the pulse, i.e. the frequency. Focusing, and self-focusing, of the pulse thereby create a significant intensity-dependent frequency modulation of the pulse (Equation (3)).

$$\omega = \omega_0 + n_2 \frac{\omega_0}{c} L \frac{\delta I(t)}{\delta t} \qquad (3)$$

The spectral profile of the field is broadened to both lower and higher frequencies as a result of the (Gaussian) temporal profile of the pulse (Figure (3)); initially, the "front edge" of the pulse will increase the refractive index of the material (resulting in a modulation to lower frequency). However, as the centre of the pulse, i.e. the most intense region, passes through a particular point in the sapphire crystal, the nonlinear component of refractive index decreases. The "trailing edge" of the pulse is subject to modulation to positive frequency through this effect. The resulting white light pulse spans ~1000nm (400-1400nm)[9].

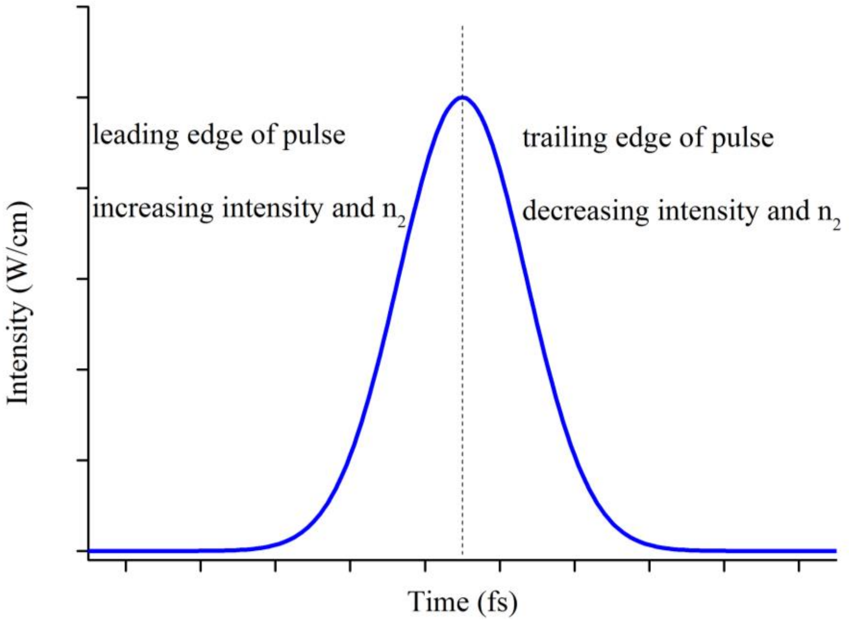


Figure (3): Temporal signal of an ultrafast pulse of Gaussian intensity profile

A high refractive index crystal is used to disperse the white light pulse temporally, after which, the beam is focused onto a 2nd order nonlinear optical crystal (barium borate; BBO). The 800nm pump beam is similarly focused onto this crystal, but travels non-collinearly with the white light pulse. Within the 2nd-order crystal, it is possible to convert single photons of 800nm radiation into two lower energy photon via the process of optical parametric amplification. In order for energy conversion to occur, photon momentum must be conserved such that:

$$\omega_1 + \omega_2 = \omega_3 \tag{3}$$

Where ω_3 is the frequency of the pump beam, and ω_1 and ω_2 are the frequencies of the so-called “signal” and “idler” beams which, respectively, are the amplified frequency components of the white light pulse and the energetic by-product of the conversion. The white light pulse is used to seed optical parametric amplification by providing a broad range of potential signal frequencies. Selection of a specific signal wavelength is achieved through adjustment of the temporal delay between the pump beam and the temporally dispersed white light pulse. Efficient energy conversion can only be achieved when the group velocities of the signal and idler beams are matched. Give the broad range of potential signal wavelengths, it is thus necessary to adjust the orientation of the birefringent BBO crystal to match these group velocities and achieve phase-matching.

Following optical parametric amplification, the signal beam is collimated using a telescope arrangement, while the pump and idler beams are dumped onto a non-reflective surface.

A second optical parametric amplification stage is then implemented to increase the power of the signal beam. Again, this is performed through a three-wave mixing process in a BBO crystal. Unlike the first OPA stage, however the pump and signal beams propagate collinearly through the crystal, thus producing a collinear idler beam.

Mid-infrared radiation is produced from the signal and idler beams through the phenomenon of difference frequency generation (DFG)[10]. This third three-wave mixing interaction involves separation of the signal and idler so that their recombination can be controlled from the standpoint of phase-matching. A silver thiogallate ($AgGaS_2$) crystal allows conversion of the signal and idler into a low frequency, broadband beam. Phase-matching of the initial amplification stage allows the signal and idler frequencies to be controlled which, by the final three-wave mixing step, allows the OPA to produce pulses with a broad range of frequencies. MIR pulses were of ~60fs duration and FWHM ~300cm^{-1} in the UCP system, while the respective parameters for ULTRA were ~40fs and ~400cm^{-1}.

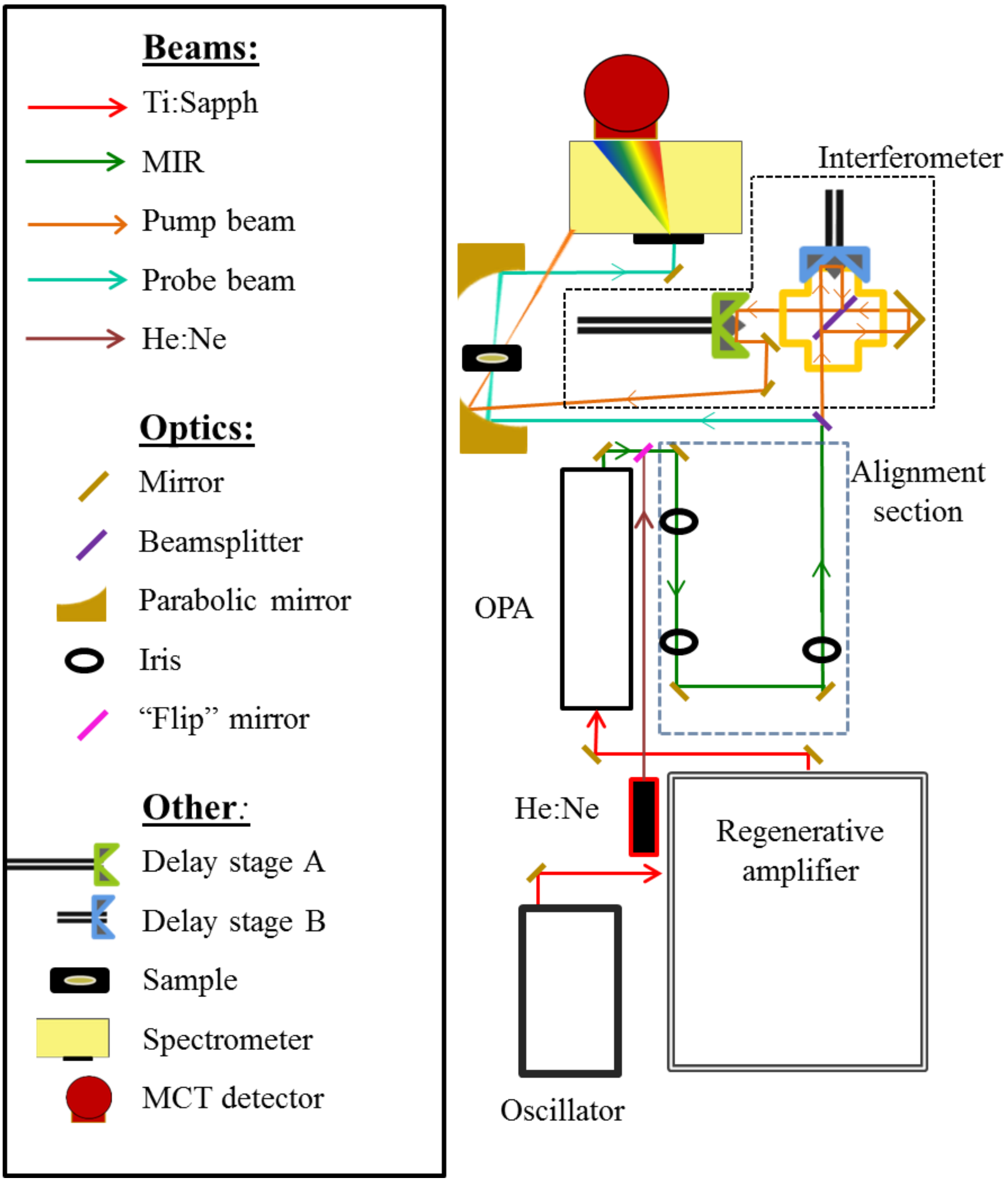


Figure (4): Schematic diagram of the Ultrafast Chemical Physics laser system. Depicted in 2DIR operation; only the fixed arm of the interferometer is required for pump-probe spectroscopy. The He:Ne laser propagates along the same path as the MIR upon activation *of the "flip" mirror, though this intersects the MIR such that the two cannot be used simultaneously.*

In the UCP setup, illustrated in Figure (2), the OPA output is directed towards a mid-infrared Interferometer *via* a set of circular apertures of adjustable diameter (irises). In combination with either a single element, liquid nitrogen-cooled mercury cadmium telluride (MCT), or a laser energy-meter, the direction of the MIR beam is adjusted so as to pass through the centre of each iris. In parallel, a He:Ne (Helium Neon) laser is also aligned with each iris centre. This process assists further beam alignment, as it creates an arrangement where the visible He:Ne beam follows the same optical path as the MIR pulse.

Following alignment, and, in both the UCP and ULTRA configurations, the MIR radiation is incident on a 90:10 beamsplitter, creating two beams: the pump and probe, the former of which is of higher intensity. The pump beam is directed into the MIR interferometer, while the probe travels towards the sample. Prior to reaching the sample position, each beam encounters a single optical delay stage; the probe beam is fed through stage A, while the pump is fed through stage B. In both 2DIR and pump-probe spectroscopies, delay stage A gives rise to the inter-pulse separation, t_2; referred to as either the waiting time, T_w, in the former method, or the pump-probe delay in the latter. Additionally, stage B is used in 2DIR spectroscopy to generate interference between the two pump pulses by modulating the inter-pulse timing, t_1. In 2DIR operation, the pump beam is split into two equal intensity components *via* a beamsplitter, and each propagates through a separate arm of the interferometer. The mobile arm, comprised of stage B, introduces temporal separation between $\vec{k}_1$ and $\vec{k}_2$.

Delay stages are programmed prior to an experiment so that the delay ranges and stepping intervals can be controlled. As previously discussed, stage A is in motion during a pump-probe study, while delay stage B is not required. For this reason, the mobile interferometer arm is effectively blocked during a pump-probe study; a mirror is employed in place of a 50:50 beamsplitter. In 2DIR, stage A is fixed at a desired delay value for the duration of the experiment while stage B is scanned. A minimum step size of ~50fs is attainable with stage A (corresponding a physical translation of ~15µm). A higher degree of precision is required for stage B, as the pump frequency axis generated through Fourier-transformation of the pump-pump interference pattern (the pump auto-field correlation) is dependent on both the step-increment of the stage (Δt_1) and the range of time delays observed (the max value of t_1)[11]. Smaller step-sizes allow the detection of higher frequency components in the Fourier-transform, while larger values of t_{1max} increase the pump frequency resolution (see Appendix (A.2.1)).

Figure (3) shows a typical pump auto-correlation, obtained with a sampling interval of 3fs (~0.45µm) steps, and maximum t_1 values of 5ps (though it is only displayed to 1.2ps). This step size (3fs) constitutes the minimum value obtainable from stage B in the UCP system, though the sensitivity of ULTRA is somewhat higher, allowing minimum step-sizes of ~1fs (0.15µm) from the equivalent delay line.

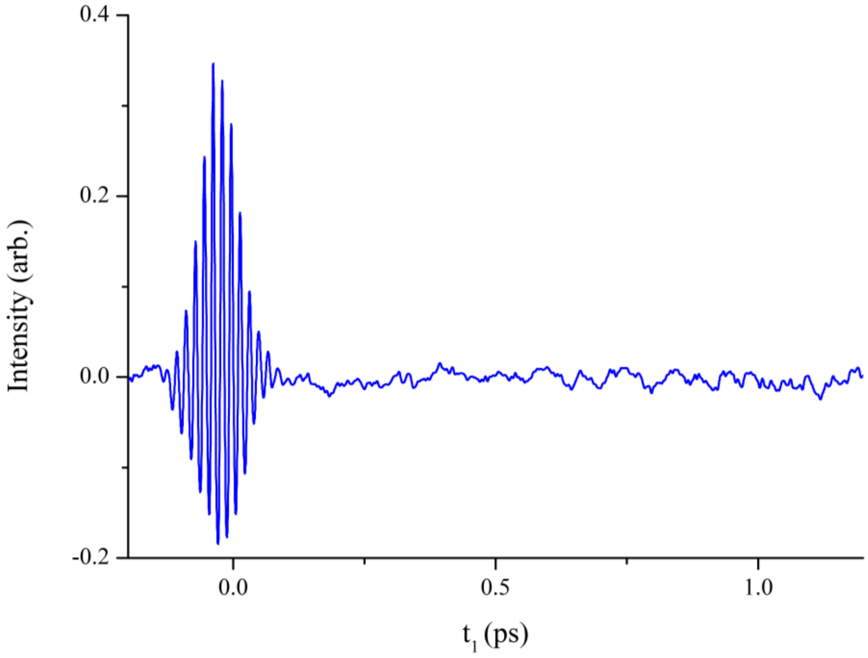


Figure (5): A pump auto-field correlation, acquired in parallel to a 2DIR measurement.

The pump auto-field correlation is measured by isolating a weak reflection of the recombined pump beams, following propagation through the MIR interferometer. At each stage position along t_1, an intensity value is obtained by directing the nondispersed pump beam onto a single detector pixel. Over the course of the 2DIR measurement, the motion of stage B across its full programmed range generates the auto-field correlation, as shown in Figure (3). It can be seen in this plot that, ~ $t_1 = 0$, a maxima in the pump-pump intensity pattern occurs, signifying that each frequency component is in-phase and interferes constructively. Establishing this so-called time-zero (t_0) point is essential to the production of the pump frequency axis, and to the "phasing" of time-domain 2DIR spectra, though this is discussed in greater detail below, following a description of the signals measured by the probe beam.

Immediately prior to interaction with the sample, the pump and probe are incident upon a parabolic mirror. This parabolic mirror reflects both pulses towards the sample, whilst simultaneously focussing the beams. After propagation through the sample, the probe beam is re-collimated using a second parabolic mirror and spectrally dispersed *via* a diffraction grating, to be measured as a frequency spectrum using a liquid nitrogen-cooled MCT detectors. In the case of the ULTRA setup, 128-element MCT detectors were employed, allowing the acquisition of probe spectra with ~2cm^{-1} resolution. Probe resolutions of ~3.2cm^{-1} were obtained from the UCP system, using 64-element MCT detectors. The spectrometer grating angle and probe alignment are adjusted to centre this pulse's frequency spectrum with respect to the centre pixel of the MCT detector. Reduction of the probe intensity from the IR absorption of water vapour is minimised by purging of the laser path with nitrogen gas. After interaction with the sample, the pump beam is "dumped" onto a non-reflective surface.

Overlap of pump and probe beams at the sample position is initially achieved through the use of a 75μm aperture at the sample position. The sample mounting block is translatable through the focal point (z-axis) of the incoming beams, and can further be adjusted in the plane perpendicular to the beams (x and y axes). The sample mounting block is thus translated so that visible radiation He:Ne beam (which follows a path collinear with the probe) passes through the aperture. Subsequently, the pump He:Ne is directed towards the aperture through the use of a steering mirror, prior to the first parabolic mirror. Optimisation of the overlap is achieved in conjunction with a test sample; for studies conducted in the nitric oxide stretching region (1900cm^{-1}; $\lambda \approx 5.3\mu\text{m}$), tungsten hexacarbonyl (WCO_6) was employed as the test sample, while n-methylacetamide (NMA) was used for this purpose in studies conducted in the 1600cm^{-1} ($\lambda \approx 6\mu\text{m}$) region. Test samples provided a means optimising the “transient absorption” signal.

This “transient absorption” is defined by the manner of data collection; the probe spectrum in all measurements (pump-probe and 2DIR) is referenced against the case when there is no initial interaction of the pump with the sample. The transient signal (the probe spectrum when the pump is “off”, subtracted from the probe spectrum when the pump is “on”) therefore contains only spectral changes that are specific to the three wave-mixing interaction. Beam choppers are used to block alternate pump pulses, allowing probe spectra to be acquired with and without the pump. Chopper oscillations were synchronised with the regenerative amplifier, though at half the frequency of the amplifier repetition rate, i.e. 0.5 kHz in the UCP laser system and 5 kHz in the ULTRA system.

In time-domain 2DIR, as there are two pump pulses, only $\vec{k}_2$ is chopped. The probe signal obtained when $\vec{k}_2$ is blocked is equivalent to a pump-probe signal with constant temporal separation; the interval between $\vec{k}_1$ and $\vec{k}_3$. This is subtracted from each probe frequency spectrum. The resulting signal at each probe pixel, across the time range, t_1, takes the form of a free induction decay, as depicted in Figure (4). A second pump-probe signal is also generated through the interaction of the mobile pump pulse, $\vec{k}_2$, with $\vec{k}_3$, though the removal of this component is discussed below.

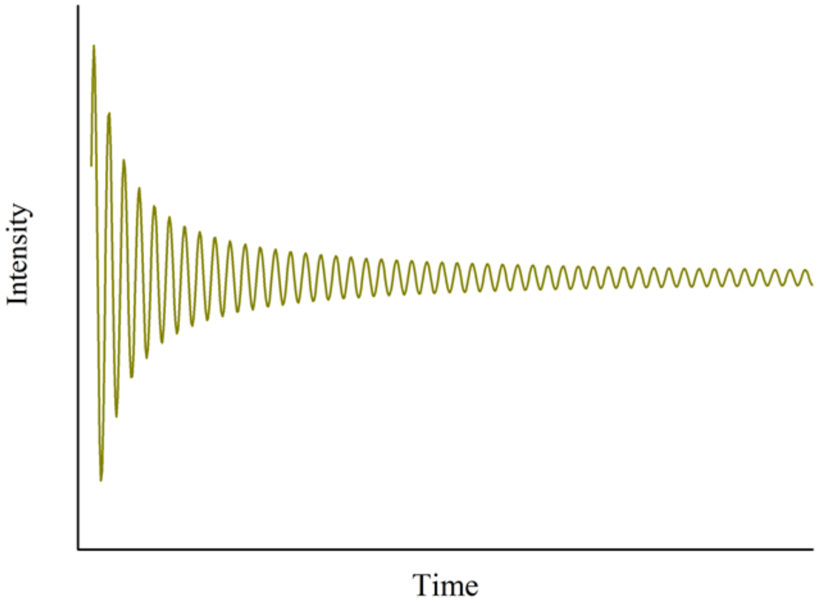


Figure (6): A schematic free-induction decay signal. Each pixel on the probe axis (ω_3) initially contains such a time-domain signal along t_1, until Fourier transformation converts this to a frequency-domain signal in ω_1

Fourier-transformation of the free induction decay across t_1 yields a frequency spectrum in $\omega_{1.}$ The second pump-probe signal is removed by this transformation, as its contribution to the free induction decay is reduced to a single exponential component with decay constant T_1, the frequency-domain counterpart of which is a Lorentz lineshape with width equal to the lifetime-determined component of the natural linewidth[12]. This final step allows probe slices to be compiled, generating a 2DIR spectrum. However, in order to preserve the desired spectral contributions and exclude unwanted information, it is necessary to truncate the probe data taken across t_1 by removing data points collected prior to t_0. This truncation time is determined from the pump-auto field correlation using a bespoke software package that will now be discussed.

All pump-probe and 2DIR spectra presented in this thesis were obtained using “parallel” polarisation, i.e. all beams are linearly polarised in the same axis. Modification of the relative beam polarisation is commonly used to supress or enhance various Feynman pathways, or to remove the contributions of rotational motion from the obtained data[11]. However, this approach was not employed here, as the molecules investigate were sufficiently large to make the effects of rotation negligible during the experiment. Verification of this assumption has previously been performed by conducting anisotropy measurements on a nitroslyated myoglobin sample.

2.3 Data processing software

2DIR spectra were processed using a LabVIEW program authored by: Dr Neil Hunt, Dr Rafal Kania, Dr Katrin Adamczyk and Dr Gregory M. Greetham. This program was used to extract the pump auto-field correlation from the data in order to construct the pump frequency axis of 2D spectra by means of Fourier-transform, and thus “phase” each spectrum.

Where multiple 2DIR measurements were performed, and an equal number of pump auto-correlations were obtained, it was first necessary to compensate for small amounts of beam drift that occur over the duration of the experiment. This was performed by transposing each interference pattern in the dataset to maximise their overlap in t_1. Following this step, an average was taken of the autocorrelation signals. The averaged signal was then interpolated by a factor of 10 before further processing, the reason for which will be made apparent during later analysis stages.

In order to define the pump frequency axis, the phase relationships of each frequency component in the pump spectrum were determined in t_1. Initially, an upper temporal cut-off was applied to the full auto-field correlation, defining a relatively small window (for example, -0.2 to 0.2ps) which contains the “centre-burst” of the pump auto-field correlation; the temporal region in which the measured intensity is defined by the correlation product of the two pump fields[13]. Fourier-transformation of the signal within this window yielded the power spectrum of the pump field as well as the phase spectrum.

In order to extract only absorptive contributions of the 3rd-order molecular response from the collected probe data, the measured signals in t_1 weree truncated to remove “negative” delay times. This can only be achieved when the point of zero phase-difference between the two pump pulses is known, and without truncation of the data in this manner, dispersive components of the lineshape may contribute to the spectrum. However, a close estimate of the t_0 value was derived through analysis the phase-slope ($\delta\varphi/\delta t_1$) of the analysed pump auto-field correlation as, at time-zero, all frequency components are in-phase, and thus the phase-slope gradient is zero.

Derived truncation points were applied to the probe data prior to Fourier-transformation in t_1 to “phase” the spectrum[12]. The previously-mentioned interpolation factor allowed fine-tuning of spectral phasing by permitting selection of additional truncation points than those measured, yielding greater precision. Determination of a “correct” phasing-time was conducted with reference to the projection slice of the 2D spectrum, overlaid with the pump-probe spectrum, and also by observing the “phase-twist” of 2DIR lineshapes. Incorrectly-phased spectra contain contributions from the dispersive component of the spectral line, causing a split of 2DIR spectra into quadrants. Figures (5)-(7) provide a visualisation of the steps involved in data processing.

As with the pump auto-field signals, probe data collected *via* repeat measurements was also averaged, to improve signal-to-noise and to compensate for beam drift during the experiment. Multiple measurements were performed for each sample, at each waiting time, and spectra presented in this thesis are the resulting average of the collected datasets.

Zero-padding was implemented within this program to increase spectral resolution along the pump axis[11]. In this procedure, the t_{max} of the Fourier transform is extended by adding a range of $y = 0$ points after the measured data in t_1. The length of the added dataset is chosen to be equivalent to the t_1 range, thereby doubling the t_{max} of the Fourier transform (see Appendix equation (A.2.1))

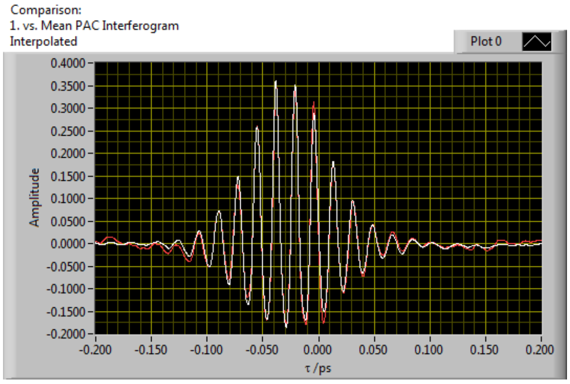


Figure (7): A pump auto-field correlation (red) obtained from a 2DIR measurement, overlaid with the average of auto-field correlation of three signals (white).

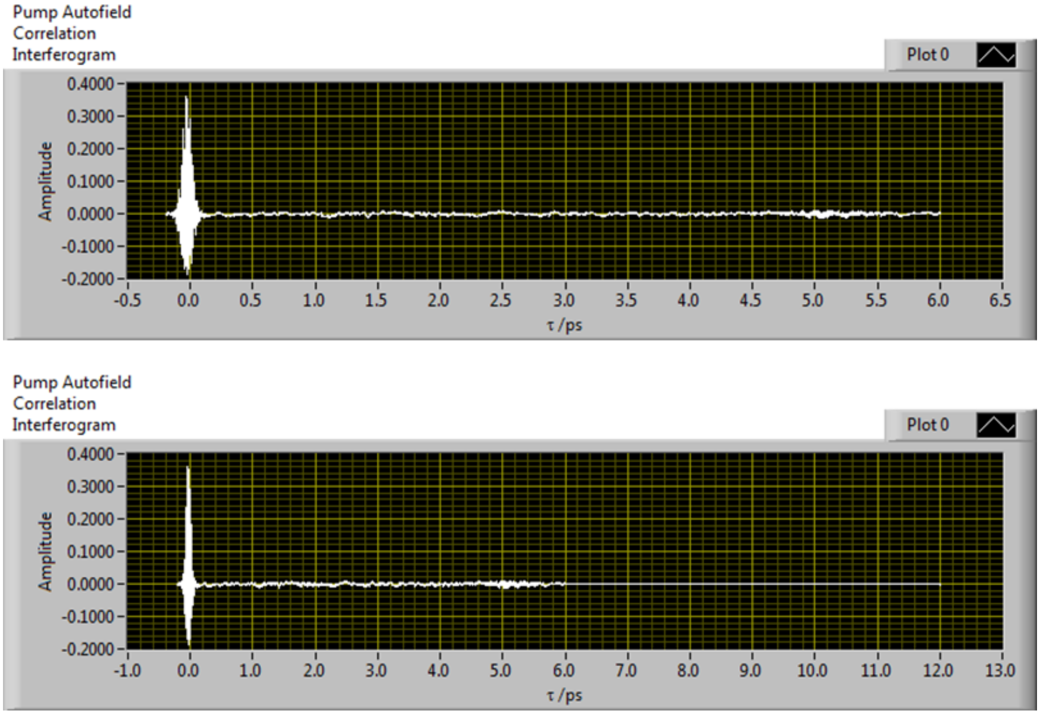
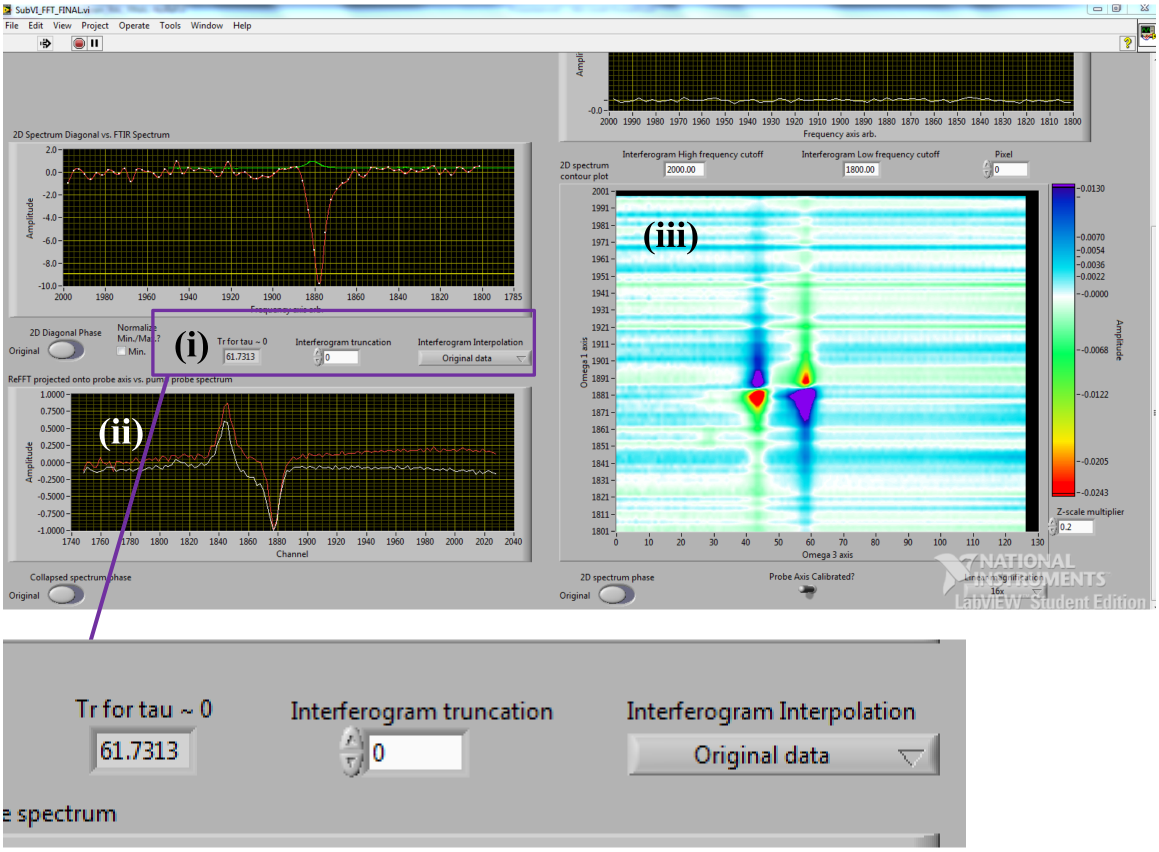


Figure (8): (a) An averaged pump auto-field correlation, shown across the full scan range of t_1 *and (b) the same signal as in (a), with zero-padding implemented*



Tr for tau ~ 0: 61.7313

Interferogram truncation: 0

Interferogram Interpolation: Original data

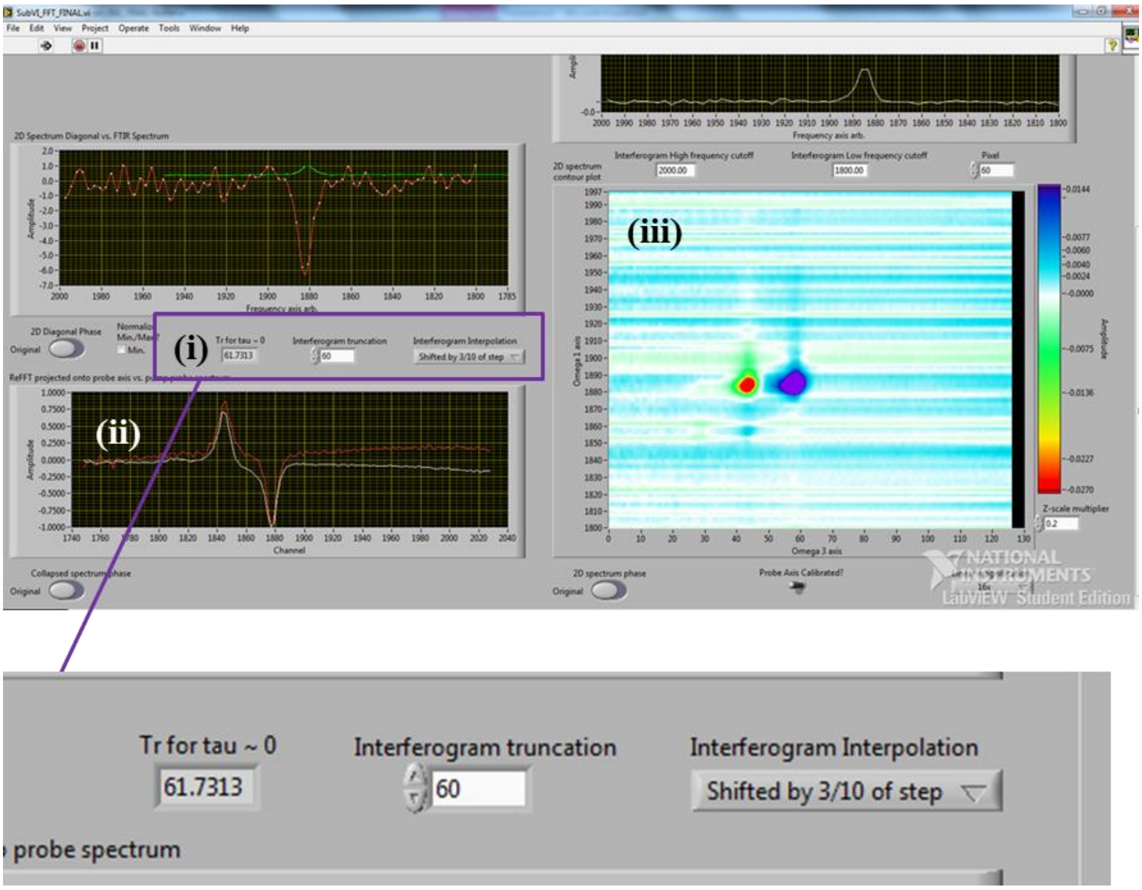
Interferogram High frequency cutoff: 2000.00

Interferogram Low frequency cutoff: 1800.00

Pixel: 0

Z-scale multiplier: 0.2

Figure (9): (a) (i) The estimated value of t0 (Tr for tau ~ 0) and the selected truncation point. In this case, a default value is selected, and so data is incorrectly phased. The projection slice of the 2D spectrum (white) is shown in (ii) alongside a pump-probe signal (red) for the same sample. In this case, the overlap between the two signals is not optimal, though the effects of poor phasing are more apparent in the 2DIR spectrum shown in (iii). (b) (Below) as in (a), but with a truncation time selected close to the predicted t_0 *(in (i)). The projection slice in (ii) bears closer resemblance to the pump-probe signal, and each of the 2DIR peaks in (iii) are of a single sign; they no longer contain a mixture of positive and negative values. Phasing of the spectra is more readily performed in this simple vibrational system (with fewer peaks) then for a complex spectrum with multiple overlapping bands.*



Interferogram High frequency cutoff 2000.00

Interferogram Low frequency cutoff 1800.00

Pixel 60

Z-scale multiplier 0.2

Tr for tau ~ 0 61.7313

Interferogram truncation 60

Interferogram Interpolation Shifted by 3/10 of step

probe spectrum

2.4 References

1. Jain, M. C. *Textbook of engineering physics*. (PHI Pvt. Ltd., 2009).
2. Stewart, J. E., Richmond, J. C., Infrared emission spectrum of silicon carbide heating elements, J. Res. Nat. Bur. Stand., **59**, 405-409, (1957).
3. Lepetit, L., Cheriaux, G. & Joffre, M. Linear techniques of phase measurement by femtosecond spectral interferometry for applications in spectroscopy. *J. Opt. Soc. Am. B* **12,** 2467–2474 (1995).

4. Greetham, G. M. *et al.* ULTRA laser system: a new dual-output 10 kHz Ti:Sapphire amplifier with UV–IR generation for time-resolved spectroscopy. *Cent. Laser Facil. Annu. Rep.* 249–250 (2008).
5. Greetham, G. M. *et al.* ULTRA: a unique instrument for time-resolved spectroscopy. *Appl. Spectrosc.* **64,** 1311–1319 (2010).
6. Svelto, O. *Principles of lasers*. (Springer US, 2010).
7. Muller, D., Backus, S., Read, K., Murnane, M. & Kapteyn, H. Very high-peak-power lasers: cryogenic cooling multiplies output of Ti:sapphire laser. *Laser Focus World* **41,** (2005).
8. Bloembergen, N. Nonlinear optics. *World Sci. Publ.* (1965).
9. Shen, Y. R. *The principles of nonlinear optics*. (Wiley, 1984).
10. Cerullo, G. & Manzoni, C. in *Coherent Vibrational Dynamics* 1–47 (CRC Press, 2007).
11. Giguère, M. Difference frequency generation makes tunable mid-IR lasers possible. *SPIE Newsrooms* (2013).
12. Hamm, P. & Zanni, M. T. *Concepts and methods of 2D infrared spectroscopy*. (Cambridge University Press, 2011).
13. Deflores, L. P., Nicodemus, R. A. & Tokmakoff, A. Two-dimensional fourier transform spectroscopy in the pump-probe geometry. *Opt. Lett.* **32,** 2966–2968 (2007).
14. Dorrer, C., Belabas, N., Likforman, J. P. & Joffre, M. Spectral resolution and sampling issues in Fourier-transform spectral interferometry. *JOSA B* **17,** 1795–1802 (2000).

3. The influence of solvent in the structural fluctuations of the horseradish peroxidase and catalase active sites

This chapter contains results published in the following papers:

(i) Simpson, N.; Adamczyk, K.; Hithell, G.; Shaw, D. J.; Greetham, G. M.; Towrie, M.; Parker, A.W.; Hunt, N. T., *The effect on structural and solvent water molecules of substrate binding to ferric horseradish peroxidase*, Faraday Discussions, (2015)

(ii) Adamczyk, K.; Simpson, N.; Greetham, G. M.; Gumiero, A.; Walsh, M. A.; Towrie, M.; Parker, A. W.; Hunt, N. T., *Ultrafast infrared spectroscopy reveals water-mediated coherent dynamics in an enzyme active site*. Chemical Science, **6**, 505-516, (2015)

3.1 Abstract

The haem group is a versatile macromolecule whose function is strongly directed by its chemical environment. Horseradish peroxidase (HRP) and catalase are closely-related haem proteins with structural and functional similarities. Both enzymes catalyse the conversion of hydrogen peroxide to water and molecular oxygen, *via* an oxy-ferryl reaction intermediate, compound I. However, the reactions of peroxide and catalase diverge after compound I formation. X-ray crystallography reveals no clear structural basis for this redirection of enzyme function, though it has recently been proposed that differences may be influenced by varying interactions with coproduct water molecules released at the haem site.

In this study, a nitric oxide probe molecule is used to extract ultrafast dynamics from the haem cavity of horseradish peroxidase, both in the presence and absence of an aromatic substrate. Parallel measurements of the catalase active site allow detailed comparison of the two haem environments. The incorporation of solvent isotope-dependence in this study provides evidence for the functional role of water in each haem protein, and further, identifies marked differences between the distal cavity solvent regimes.

3.2 Introduction

Concerted advances in the understanding of haems proteins bring into focus a powerful, core entity of biological systems[1]. The haem macromolecule (Figure (1)(a)) serves as a versatile template for reactions with diatomic molecules, many of which are physiologically vital[2].

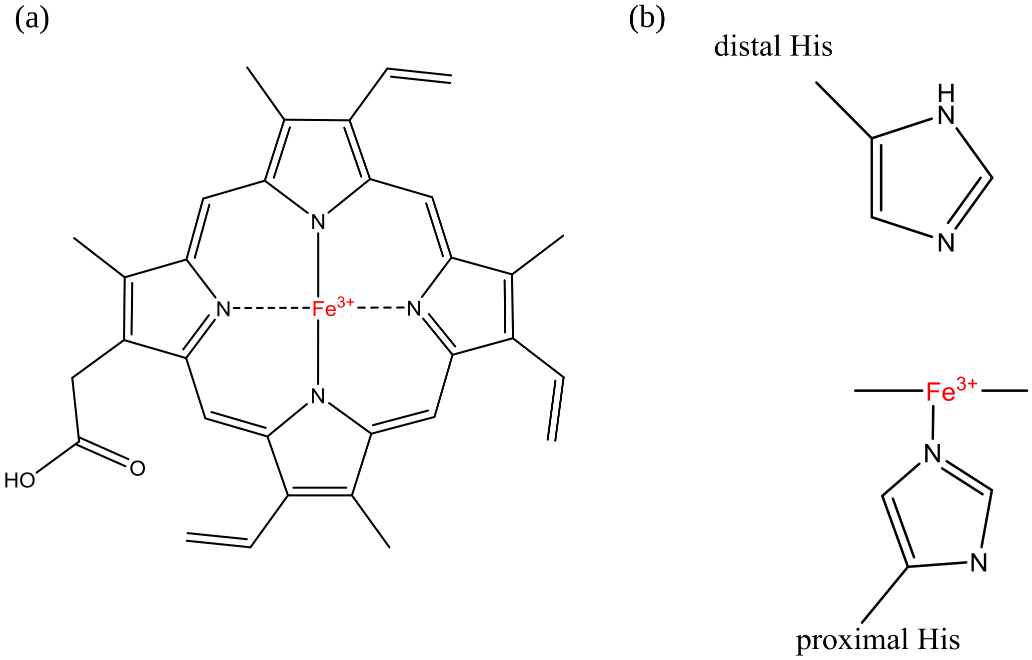


Figure (1): (a) Above view of the haem macromolecule with a ferric (3+) oxidation state. (b) The side on view of a haem with a proximal histidine ligand and a distal histidine. Proximal and distal histidine residues are highly conserved amongst haem proteins, though the exact chemical role of the distal side chain is seen to vary.

Collectively, the haem species performs roles in ligand transport[3], signalling[4,5], substrate oxidation[6] and electron transport[7] through interaction with small molecules, including oxygen (O_2), nitric oxide (NO) and hydrogen peroxide (H_2O_2). Despite this adaptability, haem proteins display a number of recurring structural attributes, or conserved features. Notably, proximal and distal histidine ligands (Figure (1)(b)) are highly conserved amongst globin and peroxidase families[8–10]. Proximity of the distal histidine residue to the sixth coordinate ligand, i.e. the active ligand, typically gives the histidine a key functional role within the active cycle, though the nature of this action is seen to vary between proteins in the same family. For example, the distal histidine in globins is believed to prevent auto-oxidation of the haem group by oxygen[11], allowing these proteins to carry out ligand transport and storage more successfully. However, the conformations accessible to the distal group in each protein, with respect to the active site ligand, seem dissimilar. This was exemplified through comparisons of neuroglobin (Ngb) and myoglobin (Mb)[12]; oxygen storage proteins found in brain and muscle tissue, respectively. In this comparison, it was further shown though point-mutation that similar histidine conformation could be achieved by modifying other residues in the distal scaffold, though the underlying explanation of how different conformational states are more conducive to function within the respective environments of the proteins remains elusive.

Haem function can be influenced through physical traits such as solvent accessibility[7], substrate recognition[13,14] and pH dependence[15]; traits that can originate from the wider protein environment such as the regions of secondary structure that surround the haem site. Ultimately, haem protein function is a complex amalgamation of highly localised, residue-specific interactions at the active site and the broader properties of potentially non-successive regions of the amino acid sequence. Point-mutation studies can readily provide information on the over-arching change to biological function that accompanies specific structural alterations, and though such studies are invaluable, they are not always able to justify functional changes with reference to the chain of events that may promote modified activity[16]. X-ray crystallography serves to report on conformation differences that occur between native proteins and their mutant forms[17], or sub-species of proteins within the same family[18,19], from which alternate functions may be inferred. However, pronounced conformational differences are not always observed in the comparisons of many highly-conserved haem proteins. To determine which effects drive conserved structures towards different functions, it is essential to incorporate dynamics into our view of proteins[3]. NMR studies have provided valuable information in this regard[20–22], though many of the processes vital to enzyme function occur on more rapid timescales than those resolvable with radio-frequency wavelengths.

With the aid of diatomic reporter molecules, significant characterisation of haem environments has been performed *via* Raman and 1st-order infrared experiments[23]. Extraction of site-specific information in this manner is well-facilitated by the intrinsic function of haem proteins. However, the sensitivity of linear IR spectroscopy to structural features is somewhat undermined by the highly conserved nature of haem proteins; histidine residues are commonly found across the haem family[8,24], often in both proximal and distal positions[25], giving rise to similar absorption bands between proteins of disparate function. While subtle features of the IR lineshape may still illustrate differences between proteins, steady-state IR is unable to access some of the most informative characteristics of vibrations, owing to a lack of time-resolution. For this reason, the contrast of haem environments is not always apparent through conventional IR spectroscopy.

Since its first application to haem-ligand complexes, 2DIR spectroscopy has demonstrated major advantages over steady-state IR techniques in terms of the detail that may be extracted from molecular vibrations[26]. The sub-picosecond time-resolution of 2DIR allows separation of the homogeneous and inhomogeneous components of the ligand lineshape[27,28]. Temporal evolution of inhomogeneous lineshape contributions can be characterised to establish the underlying structural fluctuations that occur within the distal site[12,29,30]. 2DIR is sensitive to distal side chain identity and orientation, the presence or absence of coupling between the ligand and distal architecture[29,31,32], the stabilisation of structural states through substrate binding[14], and solvent exposure of a ligand[30]. Further, the decomposition of spectral diffusion dynamics into processes occurring over different timescales allows the detection of both local and global protein motion[33,34]. 2DIR is thus highly complementary to the information that we wish to extract from iron-bound probe molecules; the ligand experiences comparable structural dynamics to the protein's natural substrate, encompassing all of the influences of local and global protein motion which may hold functional significance. And, while FTIR, Raman and UV-Vis spectroscopies have been used extensively to study haem proteins, 2DIR has only presently been applied to a small number of haem species.

The absorptions of commonly-used probe ligands, such as NO[30,31,35], CO[26,36,37], CN^-[38–40] and N_3[41–43], are spectrally well-separated from vibrations of the protein backbone and side chains[23,44]. This results in minimal convolution of IR peaks, for which analysis can be more readily performed. These "typical" probe ligands are not always interchangeable, however. The ability of a small molecule to reach the haem cavity, by successfully traversing "gating" residues or solvent channels that regulate active-site access, is not always indicative that the ligand is a perfect mimic of the natural substrate; studies have demonstrated the capacity of ligands to modify the pK_a values of distal residues[45], giving rise to configurations that may not represent the protein native state. Ideally, the natural ligand of the protein would be employed as a probe, however, the resulting complex is often transient in nature, necessitating the use of an analogue.

Additional considerations exist in infrared spectroscopy, and further, time-resolved infrared spectroscopy. These originate from varying transition dipole strengths and vibrational lifetimes between ligands. Preservation of a protein haem protein's native oxidation state, in conjunction with the ease of signal acquisition from a particular ligand, gives rise to two main approaches in the study of haem proteins *via* 2DIR. ferrous (Fe^{2+}) haems are typically ligated with CO[23,37,46], a diatomic molecule with a large transition dipole moment[47] and long vibrational lifetime. CO it is unable to ligate ferric (Fe^{3+}) haems[48], leading to the adoption of NO as a probe of ferric proteins[30,49]. NO molecules may form stable complexes with both ferrous and ferric haem groups[50], though they are a less commonly used probe of ferrous sites owing to their weaker transition dipole[47] and shorter vibrational lifetime, relative to CO[29,51], and to the coincidence of the Fe^{2+}-NO stretching vibration (~1600 cm^{-1}) with protein absorption bands[23]. It is possible to examine the haem protein at a non-native oxidation state, allowing for more versatility in employed probe ligand[36], though this is not without risk, as studies have demonstrated a link between Fe oxidation state and flexibility of the haem interior[52].

Horseradish peroxidase[53,54] is a ferric haem protein with conserved distal and proximal histidine residues (Figure (2)(a)). Despite the extensive study and application of HRP, key aspects of its function remain unknown. Its reaction is understood to consist of hydrogen peroxide degradation and small molecule oxidation, though the identity of its natural substrate and the underlying mechanism of its active cycle remain elusive. Many parallels have been drawn between horseradish peroxidase and the closely-related enzyme, catalase[55–58]. It is thought that a reaction intermediate common to each haem protein (compound I) reflects a similar action of the conserved histidine residues with the bound ligand, namely the abstraction of a proton from the H_2O_2 ligand[59,60]. This proposed route of compound I formation is described by Equation (1), below, and is illustrated with respect to HRP in Figure (3).

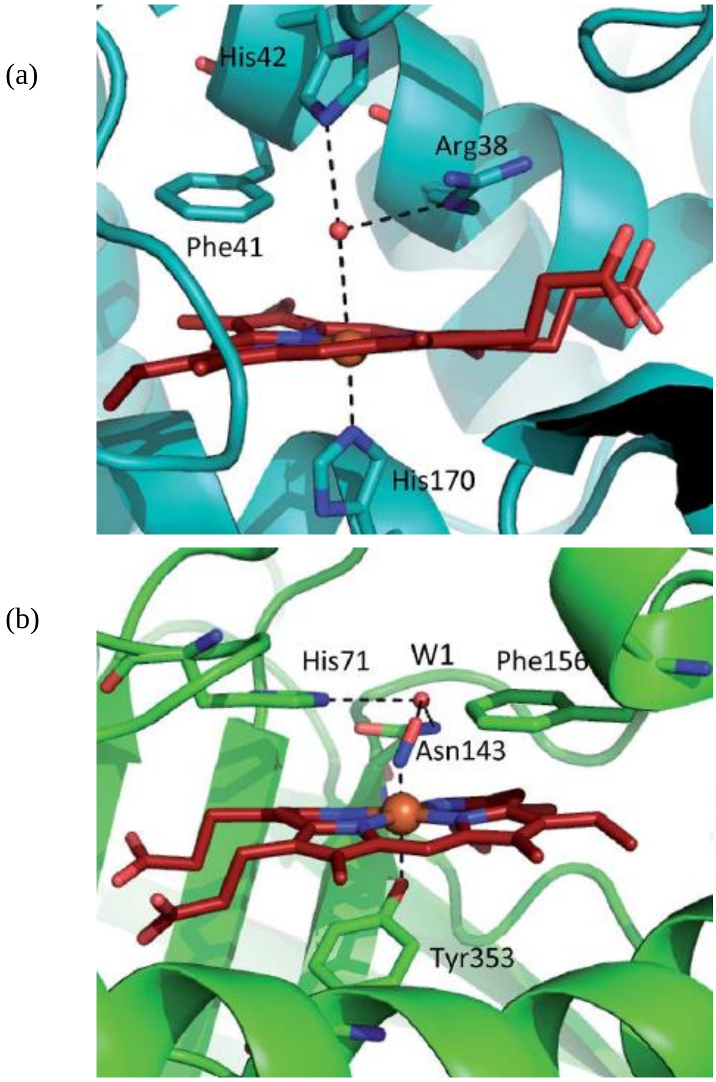


Figure (2): X-ray crystal structures of the (a) horseradish peroxidase[61] *and (b) bacterial catalase*[62] *distal cavities. (PDB codes: 1ATJ and 4B7F, respectively)*

Initial reaction of ferric HRP and catalase with a hydrogen peroxide liberates a stoichiometric quantity of coproduct water molecules;

$$\mathrm{Fe(III)} + \mathrm{H_2O_2} \Rightarrow \mathrm{O{=}Fe(IV)Por^{+\bullet}} + \mathrm{H_2O} \qquad (1)$$

where $Fe(IV)Por^{+\bullet}$ is an oxy-ferryl haem with a radical delocalised on the porphyrin ring (Por); compound I[63]. For HRP, following reaction steps follow the mechanism:

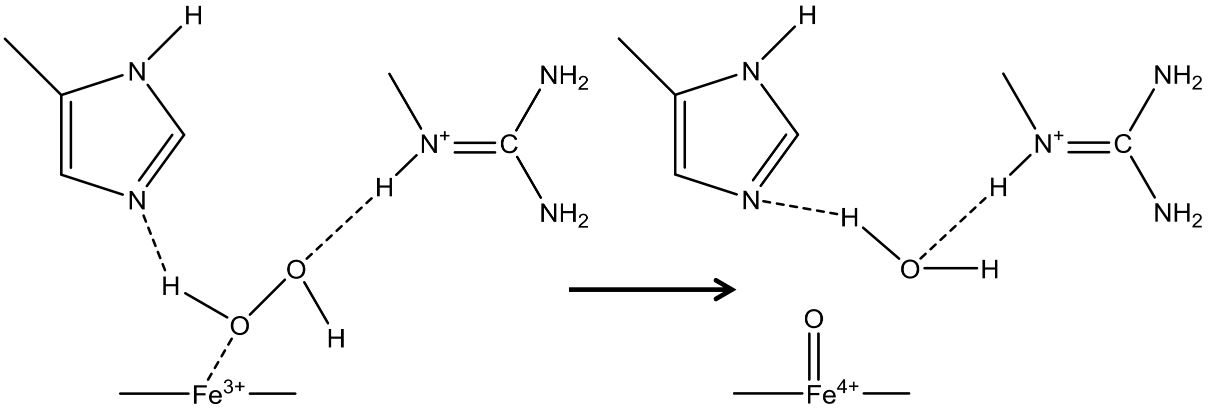
$$O{=}Fe(IV)Por^{+\bullet} + HS \Rightarrow O{=}Fe(IV)Por + S^{\bullet} \tag{2}$$

$$O{=}Fe(IV)Por + HS \Rightarrow Fe(III)Por + S^{\bullet} + H_2O \tag{3}$$

where S is an aromatic substrate. By contrast, following equation (1), catalase returns to its resting state through the following reaction:

$$O{=}Fe(IV)Por^{+\bullet} + H_2O_2 \Rightarrow Fe(III)Por + H_2O + O_2 \tag{4}$$

It can be seen that horseradish peroxidase is returned to its resting state through two one-electron reduction steps facilitated by aromatic substrates, while catalase undergoes a single two-electron reduction step facilitated by a second molecule of hydrogen peroxide. Further, the oxy-ferryl intermediate of HRP shows no capacity to react with an additional molecule of H_2O_2, and likewise, catalase exhibits only weak peroxidase activity[64].



Native HRP

HRP compound I

Figure (3): The proposed role of distal residues R38 and H42 in the formation of HRP compound I [53]. An electron from the iron atom is transferred to the compound oxygen atom, altering the oxidation state.

Comparison of the horseradish peroxidase and catalase X-ray structures show a possible route by which different functions may be achieved; the active site of HRP is heavily solvent exposed *via* an access channel[65], while access to the haem group of catalase is restricted by gating residues at the end of a long, funnel-like channel[66,67]. Substrate oxidation in HRP is directly facilitated by the solvent access channel, as this permits binding of aromatic substrates at the edge of the porphyrin ring[65]. As a secondary consequence, the channel also apparently gives rise to a contrasting degree of solvent-haem interaction between HRP and catalase.

It has been suggested that the coproduct water molecule formed with compound I reduces the favourability of a second oxidation reaction with hydrogen peroxide, should the released water molecule remain at the haem[56]. This gives rise to the classification of "wet" and "dry" reaction intermediates; the presence or absence of water molecules at the active site after compound I formation. Given the high degree of solvent exposure in HRP, this proposal appears consistent with the limited ability of peroxidase compound I to perform catalase-like function[64], leading to the designation of HRP as haem with a wet reaction intermediate. In catalase however, the high efficiency of the enzyme would require a dry intermediate, from which it has been proposed that the catalase structure is capable of rapidly expelling water molecules from the haem[56]. The existence of such an expulsion mechanism is supported by crystallographic studies that reveal water molecule displacement from the active site of the compound I form of cytochrome c peroxidase[68], relative to its resting state. Further, it has also been demonstrated through UV-Vis spectroscopy that compound I is susceptible to nucleophilic attack from water molecules[69], necessitating a means of "treating" degradative structural, disordered or coproduct water molecules. In dry intermediates (catalase), treatment of water is proposed to occur from a "priming cycle"; [56] the expulsion of water from the haem, and wet intermediates (HRP) are proposed to immobilise resident water molecules while retaining them at the haem site. More recently, molecular dynamics (MD) studies have suggested a more involved role for water in the haem cavity, on the basis that the energetic favourability of compound I formation in catalase and HRP is inconsistent with observed kinetic data[70] when water is absent from the distal cavity. This challenges the assignment of the distal histidine residue as a proton "sink" in the reaction that generates compound I.

In the case of catalase, lack of solvent accessibility implies that resident, i.e. structural, water may play a functional role. Interestingly, crystallographic studies of the catalases reveal a number of structurally-conserved water molecules in the distal cavity, most notably the solvent molecule designated W1, which is located close to the haem centre[57,62]. A similarly-placed structural water is found at the active site of HRP, in addition to "disordered" solvent molecules that access the distal cavity by virtue of the substrate access channel[71,72]. In horseradish peroxidase, additional subtleties of the reaction mechanism may also originate from structural changes accompanying complex formation[73].

The substrate binding site of HRP is known to accommodate a broad range of aromatic compounds[53]. Crystal structures of various horseradish peroxidase-substrate complexes have demonstrated flexibility of the protein structure at the haem periphery and solvent access channel, to which the versatility in substrate binding has been attributed[73]. Without knowledge of the natural substrate of horseradish peroxidase, probing of this structural region has been undertaken using numerous other aromatic molecules. From these investigations, benzohydroxamic acid (BHA) has been documented as possessing a particularly high affinity for the substrate binding site, leading to the adoption of BHA as the somewhat standard aromatic probe of HRP[74–77]. X-ray studies of the HRP:BHA complex have shown the polar head group of BHA orientated into the peroxidase distal site[73], suggesting that it has the capacity to modify the hydrogen bonding interactions of functional residues, such as R38, within the cavity (Figure (4)). Formation of the HRP:BHA complex has also been suggested to introduce conformational changes at the mouth of the solvent access channel[76,78,79] and tightening of the distal structure near the haem[36,80]. Each of these structural modifications by the substrate has the potential to carry mechanistic implications.

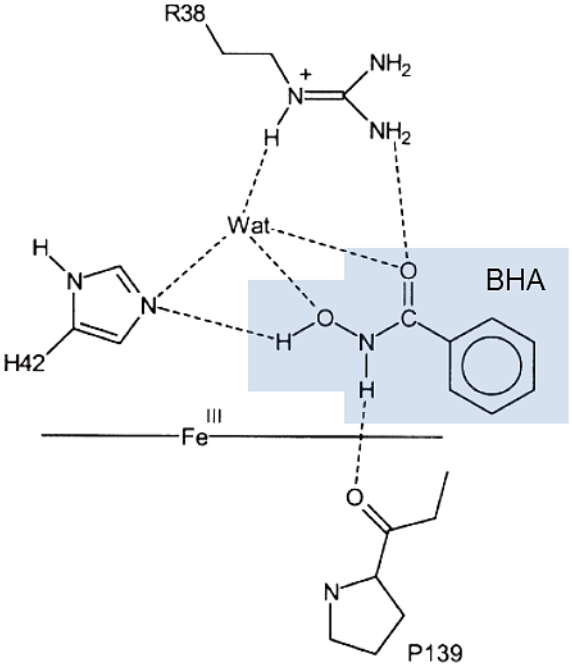


Figure (4): The intra-cavity hydrogen bond network of HRP:BHA[53]*, in the absence of an active site ligand. "Wat" denotes a resident water molecule in the HRP distal cavity.*

The influence of water at the haem sites of catalase and HRP was explored *via* isotope-dependent measurements of the structural dynamics and vibrational lifetimes. The ferric form of each protein was studied under both H_2O (aqueous) and D_2O (deuterated) environments, using an NO probe group. In the case of HRP, the effects of substrate binding were also investigated by performing comparable isotope-dependent 2DIR and pump-probe measurements on a horseradish peroxidase-benzohydroxamic acid complex (HRP:BHA).

3.3 Experimental

All chemicals were supplied by Sigma Aldrich and, with the exception of catalase from *Corynebacterium glutamicum*, all were used without further purification. Catalase solutions were centrifuged using an Amicon Ultra filtered centrifuge column with a 10 KDa molecular weight cut-off to remove glycerol and ethanol additives. This was followed by a further phase of buffer addition and centrifugation to ensure adequate washing of organic solvents from the sample. Centrifugation was also required to enhance the concentration of catalase by a factor of ~20 times the supplied concentration, giving a final value of ~3.5 mM (per tetramer).

Deuteration was conducted by diluting the protein with neutrally buffered deuterium oxide to a concentration 10 times lower than that required (HRP; 0.75 mM, catalase; 0.35 mM). The dilute protein solution was stored in the fridge for a period of 24 hours, after which re-concentration was performed using the filtered centrifuge columns specified above. This process of diluting and re-concentrating was observed at least twice for each sample prior to measurement. Before measurement, a final buffer exchange step was performed in order to flush HOD from the solution.

Nitrosylation of the ferric haem proteins was achieved through the the NO donor molecule, NOC-9, or MAHMA NONOate (6-(2-hydroxy-1-methyl-2-nitrosohydrazino)-N-methyl-1-hexanamine). The concentration of NOC-9 added to each sample resulted in a 4 times molar excess of the donor, relative to the protein. The relative concentration of NO molecules was therefore 8 times greater than that of the protein. This large molar excess of nitric oxide was intended to maximise the concentration of nitrosylated HRP. NO release was spontaneously triggered upon addition of NOC-9 to the neutrally-buffered sample solution, after which adequate time (20-30 minutes) was provided for complete release of NO gas, given a release half-life of 3 minutes at room temperature, at neutral pH[81]. Addition of the NO donor was performed under a N_2 environment for deuterated samples, in order to minimise isotope re-exchange as a result of contact between atmospheric water and the protein solution.

3.3.1 Infrared Spectroscopy

All linear spectra were obtained at room temperature, using a Bruker Vertex 70 FTIR spectrometer. Samples were housed in a typical IR cell (Harrick), between two CaF_2 windows separated by a PTFE (polytetrafluoroethylene) spacer. A path length of 100 μm was used in all IR measurements. Spectra were acquired with a resolution of $1 \mathrm{cm}^{-1}$.

2DIR spectra presented in this chapter were collected using the ULTRA spectrometer are the average of 3 separate measurements. Within measurements, the probe frequency spectra measured at each value of the interferometer stage (i.e. at each value of t_1) are the average of 4000 laser shots. Pump-probe data was obtained using the UCP laser system. 20 individual measurements were performed, and the final dataset represents an average of these experiments. For pump-probe studies, the signal at each delay time was averaged over 5000 laser shots.

“Representative” 2DIR spectra are chosen for display in this chapter, in preference of complete datasets, as the full dataset for each sample, of which there are 6, consists of >10 spectra. For this reason, two spectra are selected to represent each sample; one at an early waiting time, and one at a late waiting time. The selection of these waiting times is intended to best illustrate spectral diffusion occurring with the range of waiting times measured.

3.3.2 UV-Vis Spectroscopy

UV-Vis spectra were obtained to confirm the successful binding of both Nitric Oxide and benzohydroxamic acid with horseradish peroxidase (Appendix (A.3.1)). A custom mounting block was attached within the spectrometer cavity to permit the use of a Harrick cell in UV-Vis measurements. Hence, a path length of 100 μm was also employed here, though the concentration of HRP was reduced to 0.12 mM. Spectra were acquired with a Perkin Elmer Lamda 2 spectrometer.

3.4 Results and Discussion

3.4.1 FTIR Spectroscopy

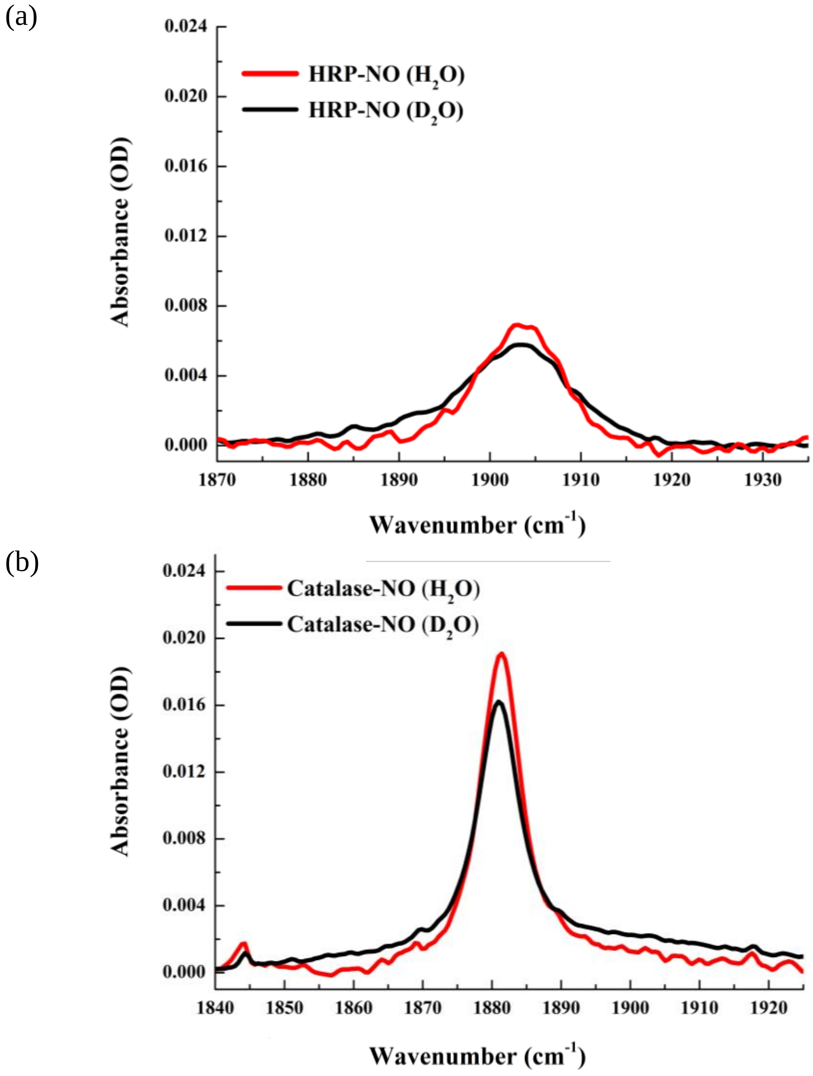


Figure (5): FTIR spectra of the nitric oxide stretching modes of (a) HRP and (b) catalase, in H_2O *(red) and* D_2O *(black).*

The linear IR spectra of HRP-NO and catalase-NO are shown in Figure (5) under aqueous (H_2O) and deuterated conditions (D_2O). Spectra in the displayed range were optimally fit with a single Gaussian function, signifying that the NO ligand in each haem protein occupies one primary conformational state with an inhomogeneous distribution of accessible sub-states. Parameters derived from peak-fitting are shown in Table (1). For HRP, the distribution of vibrational energies about this central conformation was seen to be isotope-sensitive, with FWHM values of ~13 cm^{-1} in H_2O and 15 cm^{-1} in D_2O. The central frequency of the nitric oxide stretch in HRP is also slightly affected by the solvent exchange; blue-shifted by ~1 cm^{-1} from the deuterated to aqueous systems.

Table (1): Vibrational properties of the NO ligands in HRP, HRP:BHA and catalase, established through fitting of FTIR and pump-probe data. Vibrational lifetimes were obtained from fitting of pump-probe data with exponential decay functions, and the errors in these derived values are included below

	HRP-NO		**HRP:BHA-NO**		**Catalase-NO**	
	H_2O	D_2O	H_2O	D_2O	H_2O	D_2O
Central Frequency (cm^{-1})	1904	1903	1912	1912	1881	1881
FWHM (cm^{-1})	13	15	8	9	8	7
Vibrational Lifetime, T_1 (ps)	14.8 ± 0.3	18.5 ± 0.5	16.5± 0.8	16.3 ± 0.2	9.9 ± 0.3	14.8 ± 0.3
Anharmonicity (cm^{-1})	31	31	26	26	32	33

It is noteworthy that the NO absorption band of HRP-NO is both particularly weak and broad in comparison to other Fe^{3+}NO complexes, making the peak challenging to identify. Consequently, this study offers the first characterisation of an HRP-NO complex using IR spectroscopy. The low intensity of the band is potentially explicable *via* a low dissociation constant for the HRP-NO complex, or through poor haem inclusion (% of enzymes containing porphyrin groups) in the supplied chemical sample. However, the reported dissociation constant (K_d) for nitric oxide binding to horseradish peroxidase[82] is 64 μM, which is somewhat intermediate to the reported dissociation constants of other nitric oxide complexes with ferric haems[83]; haemoglobin (Hb); 0.25 mM, sperm whale myoglobin (Mb); (0.26 mM), catalase; 0.5 μM and cytochrome c (Cc); 30 μM. The haem inclusion of the sample can also be approximated through a comparison of the measured OD of HRP-NO at a wavelength of 420 nm, relative to the OD predicted from the corresponding molar extinction coefficient[84] ($\varepsilon_{420} = 3.3\times10^5$ cm^{-1}M^{-1}). Good agreement was found between the measured value (0.403) and the predicted OD (0.396), from a protein sample of concentration 0.12 mM and path length 100 μm, suggesting a high haem inclusion. The weak transition of HRP-NO may therefore be the result of an alternate back-bonding configuration in the haem cavity, giving rise to a lower transition dipole moment relative to other nitroslyated haem proteins such as Cc[85], Mb[31] and catalase[62].

The NO band of catalase lies ~20 cm^{-1} lower than that of HRP-NO, and in the former protein, the significantly narrower peak indicates a narrower continuum of structures that lead to inhomogenous broadening in this haem cavity. Parameters obtained from Gaussian fitting of the catalase band show a slightly narrower band for the deuterated system (7 cm^{-1}) relative to the aqueous (8 cm^{-1}).

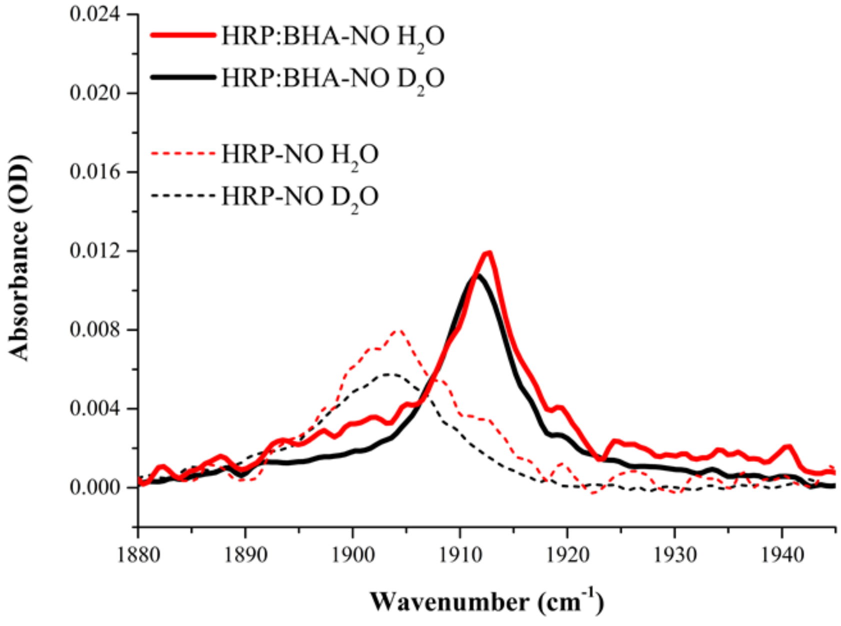


Figure (6): FTIR spectra of the NO band of HRP:BHA, in H_2O (red) and D_2O (black)

Binding of benzohydroxamic acid to HRP-NO causes sufficient changes in ligand environment to modify characteristics of the NO band (Figure (6)). The probe peak blue-shifts to ~1912 cm^{-1}, and becomes substantially narrower, to approximately 8 cm^{-1}. Changes of this magnitude are comparable with those arising from significant structural changes, such as those associated with a distal residue point-mutation[12,23,31,86]

Ligand vibrational frequency has been well-correlated with the identities of distal and proximal residues[23], and in this case, where each protein incorporates a distal histidine group, it is likely that the 20 cm^{-1} separation of ligand frequencies is at least partly attributable to the difference in proximal group, as the Fe atom of catalase is ligated by tyrosine[62], rather than the more highly conserved histidine. Additionally, it is possible that the degree of contact with the distal histidine varies substantially between catalase and HRP; interaction of the distal histidine's lone pair with a ligand has proven to be a pervasive effect in determining the IR frequency of the ligand, and when the flexibility of the distal group is such that two conformational sub-states can co-exist, with and without this lone pair interaction, the resulting frequency difference between the ligand sub-state is observed to be comparable to the 20 cm^{-1} frequency separation noted here[23]. Due to the difference in proximal ligands between catalase and HRP, it is more appropriate to compare the linear spectrum of HRP-NO to Mb-NO. The vibrational frequency of the nitric oxide group in Mb-NO is approximately 20 cm^{-1} greater than that of HRP (1921cm^{-1}), which has been attributed to the lone-pair interaction of a distal histidine with the bound NO (Figure (7)(a))[31]. When the distal residue of myoglobin is unable to offer a lone-pair interaction with the ligand, the Mb-NO frequency shifts to ~1900 cm^{-1},[23] and becomes significantly broader[31] (to ~ 10 cm^{-1}) as a consequence of the larger range of conformational states available to the ligand (Figure (7)(b)) [28]. This suggests that the lower frequency vibrational band in HRP may originate from the absence of a lone-pair interaction.

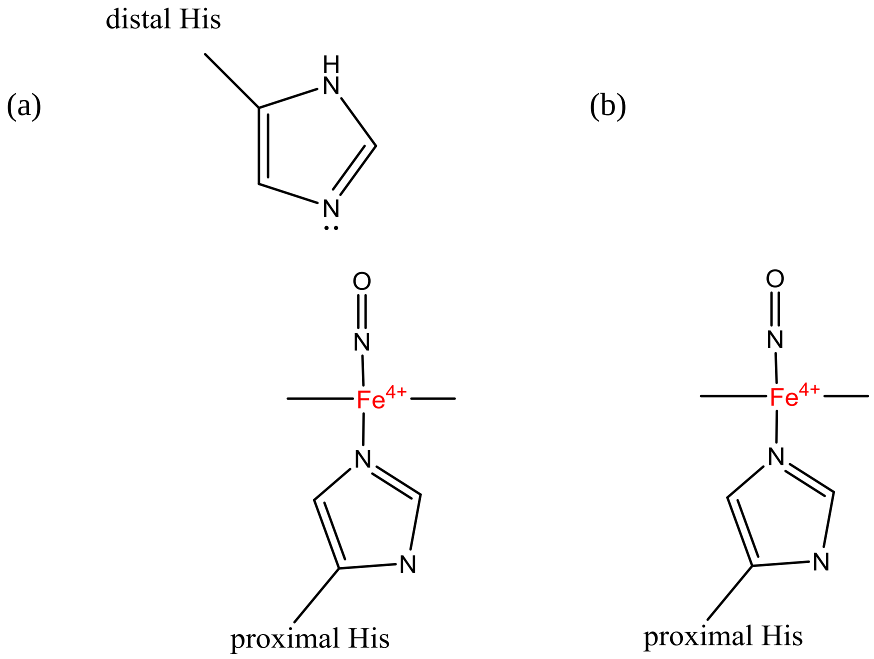


Figure (7): Possible structural interactions of a bound ligand with a distal histidine residue; (a) in close proximity, the nitrogen lone-pair of the distal group may contact the ligand, or (b) the distal residue may "swing" away from the bound ligand. The lack of an interaction in (b) allows the ligand to sample a broader range of orientations, producing a broader vibrational band.

3.4.2 Pump-Probe Spectroscopy

Pump-probe signals for the $\nu = 0\text{-}1$ (Bleach) and $\nu = 1\text{-}2$ (excited state absorption; ESA) transitions of HRP-NO and catalase-NO are displayed in Figure (8), in both H_2O and D_2O. Each of the relaxation curves was well-described by a single exponential function, and relaxation rates derived from exponential fitting are listed in Table (1). The residuals of each pump-probe signal after exponential fitting can be found in Appendix (A.3.2)

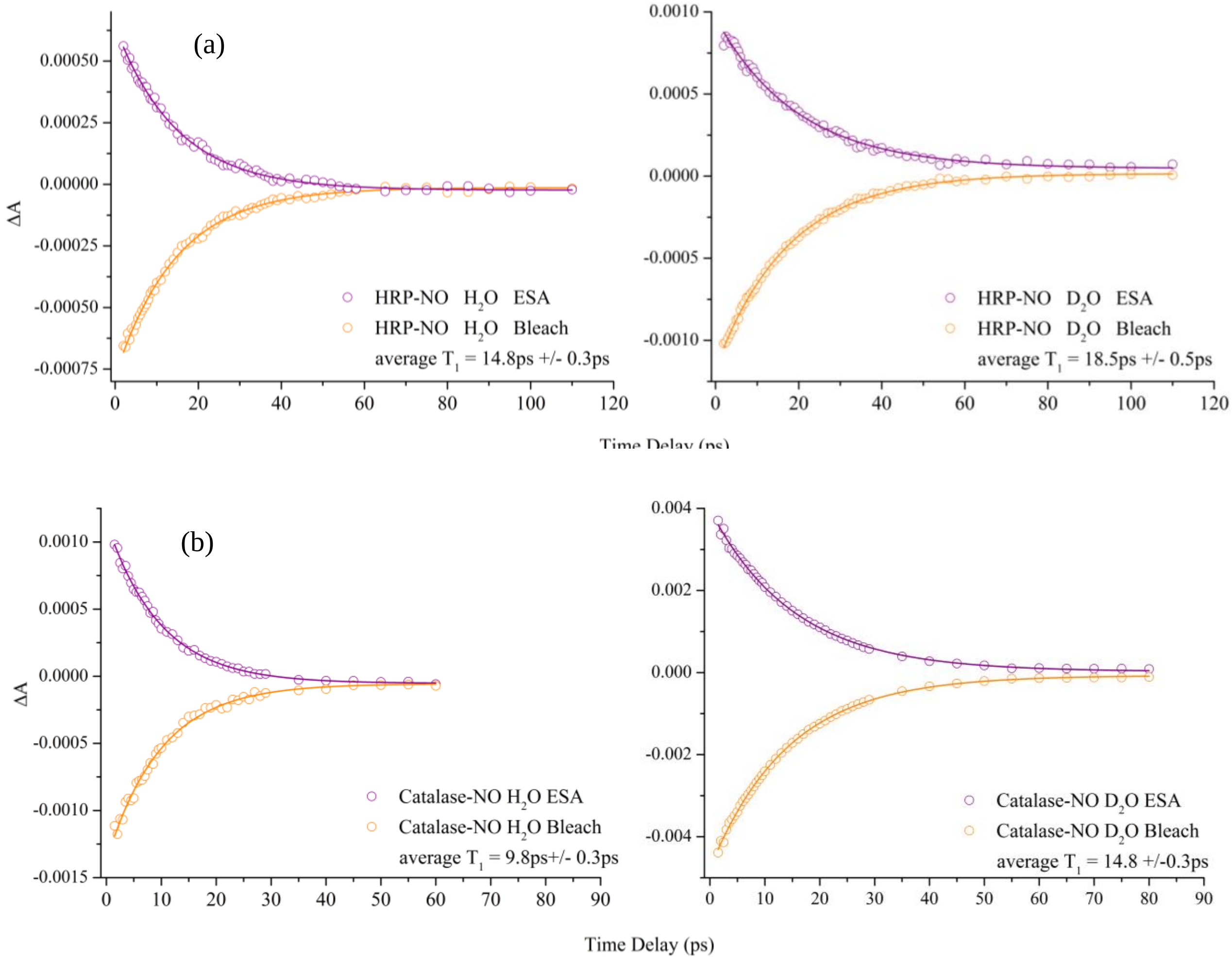


Figure (8): Relaxation dynamics of the v = 0-1 (Bleach; orange) and v = 1-2 (ESA; purple) transitions of (a) HRP-NO and (b) catalase-NO in H_2O (left) and D_2O (right).

Isotope dependence is observed in the vibrational lifetimes of each haem protein, though the effect is somewhat more pronounced in catalase, where the lifetime varies by a factor of 1.5 between the aqueous (9.9 ps) and deuterated (14.9 ps) systems. By comparison, the relaxation time-scale for HRP varies by a factor of 1.25 between 14.8 (H_2O) and 18.5 ps (D_2O); still demonstrating a significant isotope effect.

Vibrational relaxation may occur through both intra and intermolecular energy transfer processes[87–90], the latter of which is highly dependent on the chemical environment of the excited bond. It is generally believed that energy from an excited vibration is dissipated into lower frequency modes of either the oscillating molecule or other molecules with which it interacts. This is most favourably achieved when energy is conserved and the sum of the receiving mode frequencies is equivalent to that of the higher frequency mode. Where energy discrepancies occur, so-called bath modes may serve to compensate for the mismatch. The action of solvent molecules in this energy mediation has been well-documented[87,91,92], though the full complexity of the interaction is still emerging, particularly in the case of aqueous solvents, where vibrational modes are coupled with other degrees of freedom[93]. The OH stretching mode of water is particularly important in intermolecular energy transfer[89], as the hydrogen bonds made by this functional group to a solute molecule couples the solute vibrations with those of the solvent. For this reason, the relation between the solute and solvent vibrations holds importance in the intermolecular transfer of energy. Likewise, in the context of this study, coupling of the ligand and protein vibrations offers another pathway for energy dissipation.

It is apparent from the isotope sensitivity of both nitroslyated proteins that intermolecular energy transfer has a significant effect on vibrational relaxation. Deuteration will invariably alter vibrational modes of both the protein and the solvent, and thus it is not immediately apparent which of these bodies contributes to NO relaxation. In the case of the protein, H/D exchange may lead to a significant change in the frequencies of distal cavity residues or of the protein scaffold, affecting the favourability of energy transfer. Alternately, a solvent shell surrounding the active site ligand may result in a “water-like” relaxation mechanism, originating from the differences in the IR spectra of H_2O and D_2O, relative to nitric oxide.

At the frequency of the nitric oxide stretch, ~1900 cm^{-1}, H_2O is substantially more optically dense than D_2O (Figure (9)). The position of the combination band in H_2O (~2135 cm^{-1}) gives rise to this large absorption that underlies the NO peak, and as result, H_2O has greater number of vibrational states with equal energy to the ligand vibration. Resultantly, H_2O contains a higher density of accepting modes into which energy may be dissipated[94].

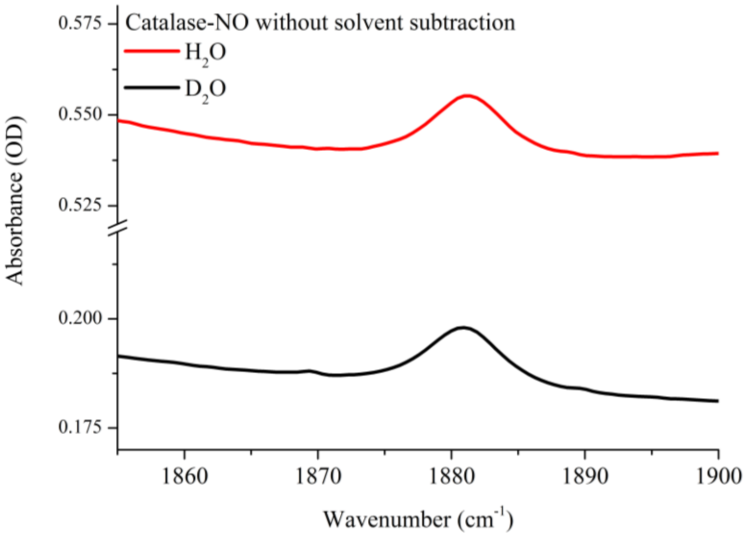


Figure (9): An illustration of the greater optical density of H_2O (red) in the nitric oxide stretching region relative to D_2O (black), shown for catalase-NO

Solvent-assisted relaxation has been well-categorised in the context of solution-phase ions, where intermolecular energy transfer plays a prevalent role in determining the rate of vibrational relaxation. In both HRP-NO and catalase-NO, the effect of solvent isotope substitution is comparable to that of a solvated ion, where vibrational lifetime have been seen to vary by factors of 1.6 to 2.9 after substitution of D_2O for H_2O[95]. This may indicate a solvent-shell-like arrangement of resident or disordered water molecules around the NO ligand in each protein. However, further information is required to assign the large isotope dependence of vibrational relaxation to protein deuteration, solvent-assisted relaxation, or a combination of the two effects.

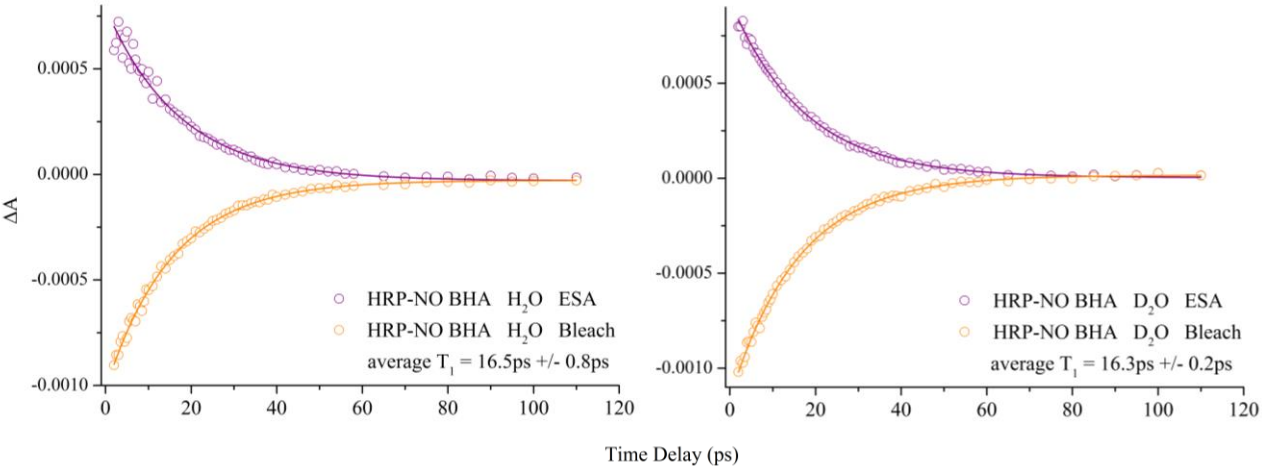


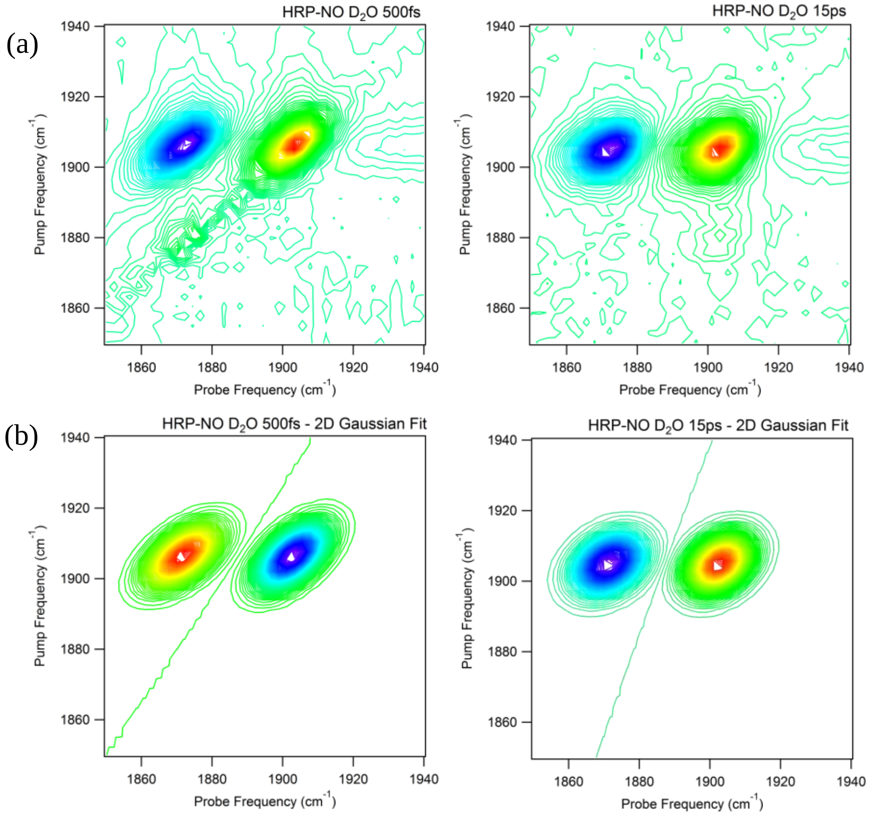
Figure (10): Relaxation dynamics of the v = 0-1 (orange) and v = 1-2 (purple) transitions of HRP:BHA-NO, in H_2O *(left) and* D_2O *(right).*

The lifetimes obtained for NO in the HRP:BHA complex appear isotope-independent, with time constants of 16.5 and 16.3 ps obtained under aqueous and deuterated conditions, respectively (Figure (10)). Since deuteration of the protein precedes complex formation it is unlikely that the previously measured isotope dependence for HRP-NO has any significant contribution from side chains with exchangeable protons, or from the protein scaffold, as these effects would equally be measured in the protein-substrate complex. It can therefore be inferred that distal water molecules assist in vibrational relaxation of the ligand in free HRP. Further support of this is given by: charge transfer (CT)[79], fluorescence[76] and X-ray crystallography[53,73] data suggesting that BHA restricts access to the haem by blocking the solvent channel. In the absence of an active site ligand, extension of the polar head group into the distal cavity results in coordination of a water molecule to the sixth ligand position to form an 6-aqua complex[75]. The results of these pump-probe studies suggest that, with the NO ligand in this sixth position, the resident solvent molecule that would otherwise form the sixth ligand, is removed from the cavity or is sufficiently displaced to no longer interact with nitric oxide, eliminating any isotope dependence from the vibrational lifetime.

It is interesting to note that NO lifetimes for HRP:BHA still fall within ±2 ps of those of the substrate free-protein as, in the former case, a significant pathway for intermolecular energy transfer appears to have been removed from the haem cavity. From this, it may be expected that the vibrational relaxation rate becomes substantially slower. The intermediate lifetimes of NO in HRP:BHA may therefore indicate an alternate, or underlying intermolecular interaction that assists relaxation of the ligand vibration in an isotope-dependent manner.

3.4.3 2DIR Spectroscopy

Representative 2DIR spectra of nitroslyated HRP and catalase are shown in Figure (11) at the waiting times 500 fs and 15 ps. All spectra in the range of T_w were fit with 2D Gaussian functions to retrieve correlation values from the vibrational lineshapes. Inset below each displayed spectrum are examples of the fit data at the corresponding waiting time. Figure (12) shows the resulting spectral diffusion profiles for HRP-NO and catalase-NO. Parameters obtained from fitting of the data are summarised in Table (2).



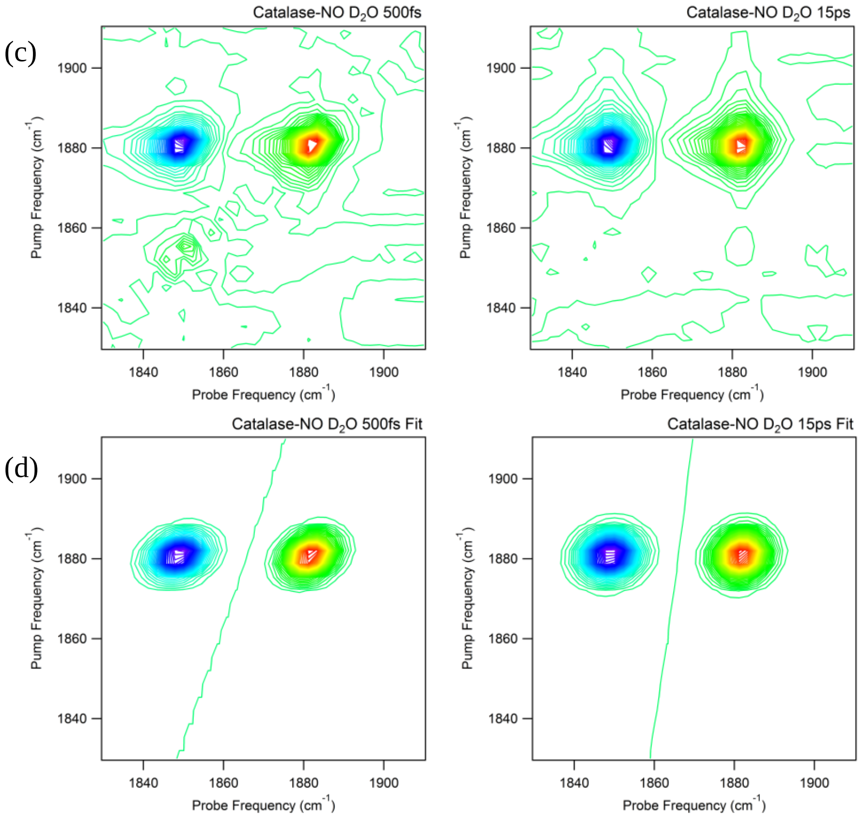
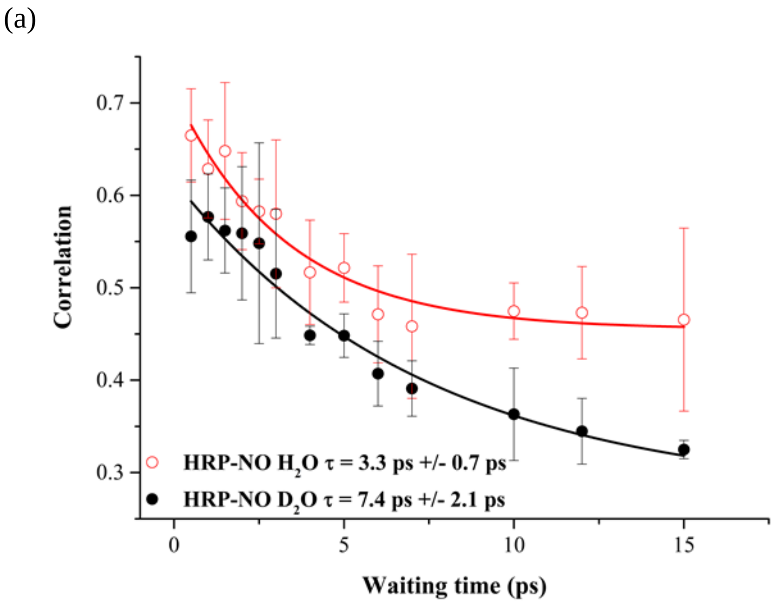


Figure (11): 2DIR spectra of (a) HRP-NO and (c) catalase-NO, at the waiting times of 500 fs and 15 ps. Corresponding 2D Gaussian fitting functions are shown in (b) and (d)



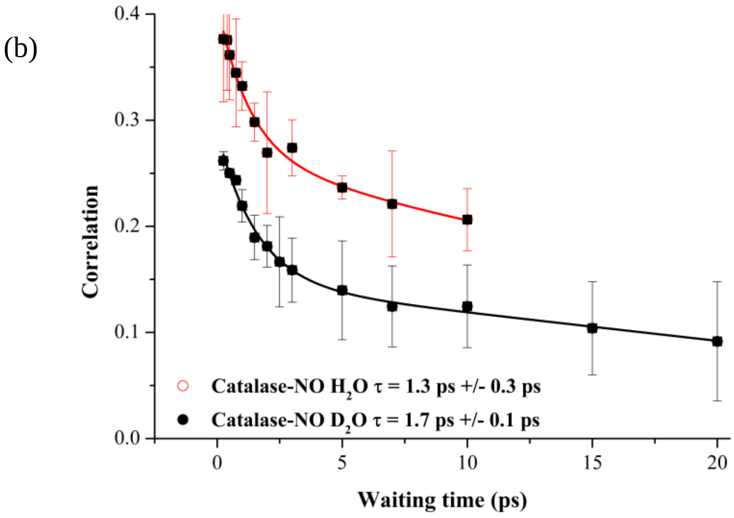


Figure (12): Spectral diffusion of the nitric oxide ligand in (a) HRP-NO and (b) catalase-NO in H_2O (red) and D_2O (black). The decrease in 2D lineshape correlation as function of T_w in each sample is described through the fitting of exponential functions (solid lines). The time-constants from exponentials are inset on each figure, along with the error in the fit. The error bars in each plot represents the maximum spread between the measured correlation values at each waiting time

Spectral diffusion of the NO band is seen to be isotope dependent in both HRP and catalase, though a comparison between the proteins identifies considerably faster spectral diffusion rates in catalase. The majority of conformational sampling of the NO group in catalase is expected to occur within ~30 ps, giving rise to a low static offset in the FFCF. Two time constants are observed in the inhomogeneous broadening of catalase-NO: a rapid process with a time constant of 1.3 ps in H_2O; 1.7 ps in D_2O, and a slower component that is not well-resolved within the temporal range of 20 ps. However, as a quantitative description of slower spectral diffusion processes could not be obtained, focus is placed on the rapid ~1ps dynamics that are well-resolved in the 20ps temporal range. The data in Figure (12) is replotted with logarithmic scale on the x-axis to justify the use of exponential fitting (see Appendix (A.3.3)).

Structural fluctuations of HRP are significantly slower, and are ascribable to a single exponential process. The variation with solvent isotope exchange is considerably larger for HRP however, with H/D exchange causing a change in spectral diffusion rate from 3.3 to 7.4 ps. At the final waiting time, a large static offset is measured, denoting the presence of slower conformational changes that are not observable within the measurement period.

Table (2): Spectral diffusion parameters for HRP-NO, HRP:BHA-NO and catalase-NO, established from exponential fitting of spectral diffusion data

	HRP-NO		**HRP:BHA-NO**		**Catalase-NO**	
	H_2O	D_2O	H_2O	D_2O	H_2O	D_2O
Spectral Diffusion Time (ps)	3.3 ± 0.6	7.4 ± 1.0	6.3 ± 4.4	3.8 ± 1.4	1.3 ± 0.2	1.7 ± 0.1
Static offset	0.45	0.27	0.29	0.36	~0.2	~0.1

3.4.3.1 Spectral diffusion of catalase-NO

The NO ligand in catalase undergoes structural fluctuations that resemble those of sodium nitroprusside (SNP), a six-coordinate Fe complex ligated by 5 cyanide groups and a single nitrosyl group (shown in Figure(13)).

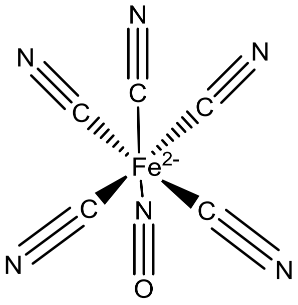


Figure (13): Structure of the nitroprusside ion.

Spectral diffusion of SNP has been categorised in a broad range of solvents in order to serve as a model for the effects of polarity in nitroslyated haem proteins[51]. Given the solvent-exposed nature of the nitrosyl group in SNP, the effects of H_2O to D_2O substitution are unsurprisingly large in relation to the spectral diffusion rate. The isotope dependence of inhomogeneous broadening in catalase is not distantly removed from the behaviour of SNP, where time constants of 0.8 ps are obtained under aqueous conditions and 1.4 ps under deuterated conditions[51]. This is suggestive of a bulk-like solvent environment in the catalase active site, indicating the presence of non-structural solvent molecules that are not resolved by X-ray crystallography. An approximation of the catalase distal cavity volume was performed using a "rolling sphere probe"[96], yielding a value of ~200 Å^3; sufficient to accommodate 5-6 free solvent molecules, in addition to structural waters. Yet, replication of bulk-like solvent environments in confined regimes, from few solvent molecules, has been documented.

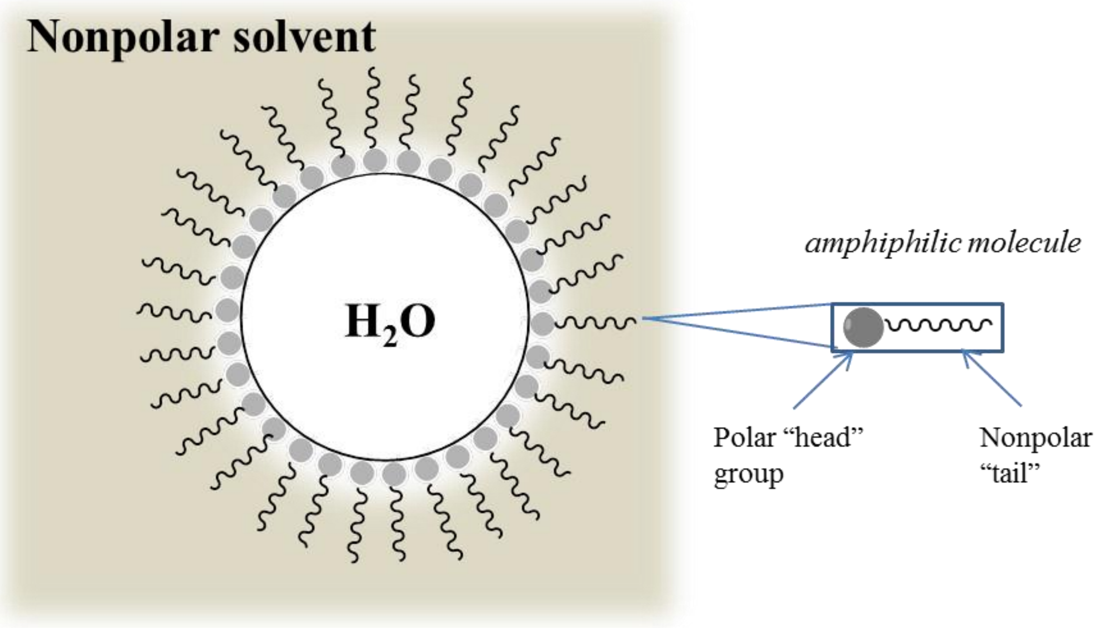


Figure (14): Illustration of a reverse micelle macromolecule. The diameter of these macromolecules can be controlled to produce highly confined water pools made up of few solvent molecules.

Amphiphilic molecules, containing both polar and nonpolar regions, have the ability to aggregate into spherical macromolecular structures known as micelles[97]. When formed in polar solvents, the charged region of each amphiphilic molecule faces outwards on the micelle, constituting the exterior wall of the structure, while the hydrophobic force drives the nonpolar regions of each molecule into the centre of the micelle, screening it from the solvent. However, when micelles form in nonpolar solvents that contain small concentrations of a polar solvent, the orientation of each amphiphilic molecule in the structure is reversed, giving rise to a “reverse micelle” (Figure (14)). Further, in these circumstances, polar solvent molecules can also be trapped in the polar interior of the macromolecule. Reverse micelle systems have been shown to display a mixture of slow (~10 ps) spectral diffusion times associated with “shell” water molecules; those that contact the internal walls of the micelle, and rapid >2 ps spectral diffusion times associable with “core” water molecules; those that directly contact only other water molecules[97]. In catalase, the small solvent pool in the distal cavity may provide an analogous scenario, with slow spectral diffusion times originating from the contact of distal side-chains with the solvent, and fast times obtained from water-water interactions that also connect to the ligand.

3.4.3.2 Spectral diffusion of substrate-free HRP-NO

Predictions of the distal cavity volumes of horseradish peroxidase suggest that it would be likely to accommodate fewer water molecules than catalase, and that this solvent pool would exist near the haem edge. However, it is clear that the isotope dependence of ligand structural fluctuations in HRP still reflect an environment with bulk-solvent characteristics. By analogy to reverse-micelle systems, the reduced haem cavity volume of HRP may give rise to a solvent regime that reflects only water molecules with “shell” behaviour. In the absence of “core” waters the ligand would be subjected to structural fluctuations that are heavily solvent dependent, but show no rapid (~1 ps) processes associated with free water molecules. In turn this suggests that frequency fluctuations of the NO ligand are determined by the distal residues, though the structural fluctuations are relayed to the ligand through a dynamically constrained water network. Solvent molecules at the shell of reverse micelles display slow structural fluctuations as a consequence of being “bound” to hydrophilic head groups; the same scenario is potentially replicable in horseradish peroxidase *via* the two polar distal residues. It is interesting to note that, despite the functional role of the distal histidine and arginine groups in HRP, contact of the active site ligand is not seen to be made directly with either, but rather through distal waters. This is consistent the large width of the HRP-NO band in comparison to the linear absorptions of other nitroslyated haem proteins with distal histidine residues, as it illustrates the far broader continuum of inhomogeneous environments available to NO in the peroxidase cavity. While it is not unreasonable that the histidine group in HRP may be able to remain in contact with the ligand over a wider range of orientations, it would suggest that the histidine does not provide as strong an influence in defining the ligand conformation than compared to other haem proteins. In such a case, it might be expected that a relatively low energy barrier exists for the transition of histidine into a structural sub-state with no ligand contact, resulting in an additional NO absorption band that is not observed here.

Previous studies have suggested that the active site structure of HRP is loosely-bound[73,98,99]. This is supported by the large pseudostatic component of spectral diffusion detected here. Structural fluctuations associated with the protein scaffold often occur over timescales too long to be fully sampled within T_w, giving rise to static offsets. In this case, the offset measured for HRP is comparable to that reported for myoglobin; a protein for which structural flexibility may have direct functional importance[85]. The loose structure of HRP may also hold functional implications, as the substrate binding site of the peroxidase is known to accommodate a vast range of substrates, for which flexibility may be advantageous[3].

It has been suggested that the distal histidine of catalases and peroxidases is insufficiently basic to readily abstract a proton from H_2O_2 during compound I formation[56]. Further, it has been noted from molecular dynamics simulations of the HRP cavity that the distance between the proximal proton of an Fe-bound H_2O_2 molecule and the distal histidine is too great (~3.8 Å) for energetically favourable deprotonation of the active site ligand[59]. In the event that the active site ligand and histidine are bridged by a single solvent molecule, MD simulations predict a significantly lower energy barrier for deprotonation[71,72] that lies close to the value derived from kinetic studies[70]. High isotope sensitivity of the spectral diffusion rates in HRP-NO and catalase-NO support the proposal of water-mediated ligand-histidine contact, providing experimental evidence for the important functional role of water in peroxidase and catalase reaction mechanisms.

3.4.3.3 Spectral diffusion of HRP:BHA-NO

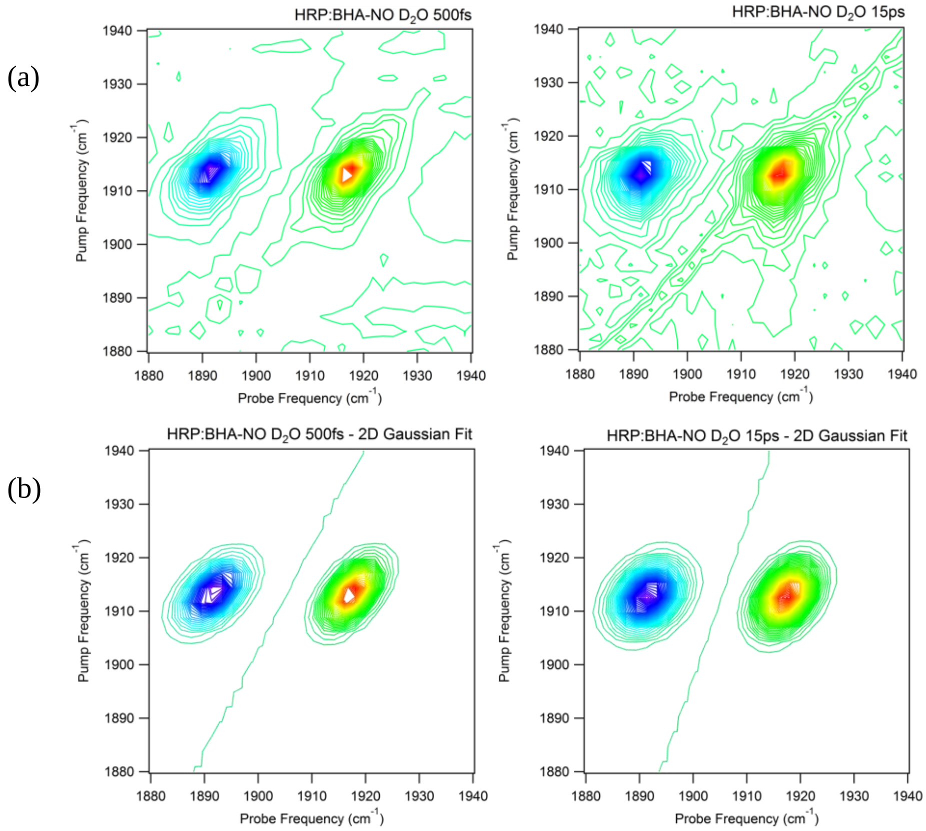


Figure (15): (a) The 2DIR spectra of HRP:BHA-NO at 500 fs and 15 ps with (b) 2D Gaussian fits of the NO lineshape at the corresponding waiting times

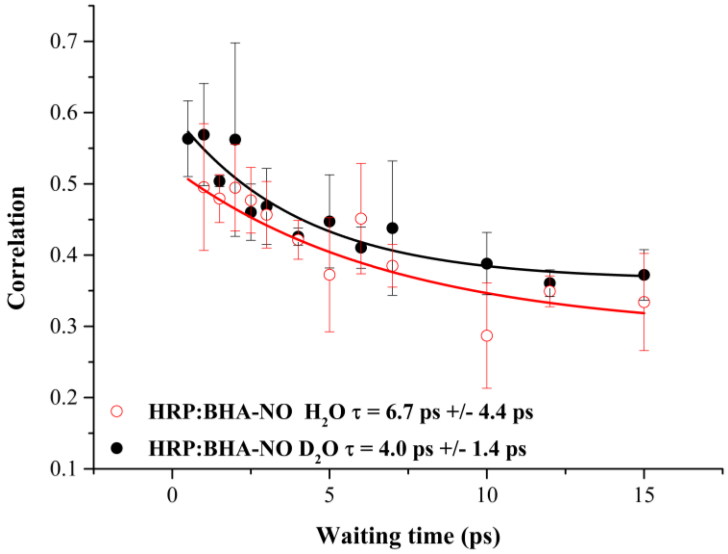


Figure (16): Spectral diffusion of nitric oxide in HRP:BHA-NO, in H_2O *(red) and* D_2O *(black). The decrease in 2D lineshape correlation as function of* T_w *in these samples are described through the fitting of exponential functions (solid lines). The time-constants from exponentials are inset on each figure, along with the error in the fit. The error bars in each plot represents the maximum spread between the measured correlation values at each waiting time*

2DIR spectra of deuterated and aqueous HRP:BHA-NO are shown in Figure (15). Structural fluctuation rates of the HRP-substrate system are characterised in Figure (16). Spectral diffusion times for each HRP:BHA complex falls between the previously measured values for the aqueous and deuterated forms of the substrate-free enzyme; ~4-6 ps. Unlike the vibrational relaxation rates of the HRP complex, rates of spectral diffusion are not interchangeable between the H_2O and D_2O systems; faster dynamics with a spectral diffusion rate of 3.8 ps are observed in the deuterated system while, by contrast, dynamics of the aqueous analogue are measured at 6.3 ps. The rate of slow conformational changes in HRP seems only weakly affected by complex formation, as indicated by static offset values of 0.3 and 0.35 ps for the aqueous and deuterated system, respectively.

It has been proposed that binding of BHA leads to a constriction of HRP's solvent access channel[78]. Ligand spectral diffusion has previously been shown to be sensitive to changes in the global protein structure[33,34], and so, in principle, tightening of the access channel could result in a lower static offset in the FFCF of HRP:BHA-NO relative to the substrate-free form. However, evidence of such a change is not supported here, as dynamics falling within the experimental window are not observed to differ greatly after substrate binding, suggesting the HRP structure remains similarly loose with and without a substrate.

3.4.3.4 Comparison of structural dynamics

A network of solvent molecules dominates spectral diffusion of the NO ligand in substrate-free horseradish peroxidase. Though the time-scale of spectral diffusion does not appear to vary drastically upon the addition of BHA, the subsequent isotope dependence is uncharacteristic of a solvent network. Taken with the vibrational lifetime data for HRP:BHA, it is unlikely that the NO ligand remains in direct contact with any distal water molecules after the binding of BHA. Again, infrared data provides support for the obstruction of the HRP solvent access channel by BHA, and verifies that solvent presence in the distal cavity is diminished. This is consistent with proton NMR studies that report a significantly reduced H/D exchange rate for distal-cavity residues in HRP following substrate binding[100]. The displacement of water from HRP may be conducive to the second, one-electron reduction stage of HRP's reaction; the reduction of compound II[53], though subsequent proton donation to the bound oxygen atom, to complete the functional cycle would likely require the reintroduction of water into the cavity.

The catalase active site also appears to constrain solvent molecules at the vicinity of the haem, though the dynamic regime in this case is not identical to HRP. Fast structural dynamics of ~1 ps suggest that water-water interactions take place in the catalase distal cavity, giving analogous dynamics to those of "core" water molecules in reverse-micelles. Isotopic dependence of the vibrational lifetime in catalase may be explicable through coupling of the NO stretching mode to protein vibrations with sensitivity to H/D exchange. However, reference experiments were performed using a third haem protein, myoglobin (Mb), in order to determine the effect of protein deuteration on the rate of Mb-NO. The active site of myoglobin contains both proximal and distal histidine residues, and the resulting vibrational spectrum and dynamics of an Mb-NO complex reflects a lone pain interaction between the latter histidine group and the ligand; comparable to catalase. Additionally, the haem group of myoglobin is solvent inaccessible, from which the effects of protein deuteration can be established without convolution with water molecule dynamics. The resulting pump-probe experiment reveals an isotope-independent relaxation rate of 27 ps (Appendix A.3.4), suggesting that effect of protein scaffold deuteration alone cannot justify the large effect of H/D observed for the vibrational relaxation of catalase-NO.

Studies by Jones[56] have tentatively proposed the immobilisation of water in peroxidases as a counter-measure of water-facilitated haem degradation. It is clear from these 2DIR studies that the resident water molecules of HRP are highly constrained, which is in agreement with this concept. The equivalent mechanism of water treatment in catalase, a priming cycle that expels water from the haem, is not as directly observable, however.

3.5 Conclusions

The ferric haem group of horseradish peroxidase was probed *via* an iron-bound NO molecule, with and without an aromatic substrate at the haem periphery. In the absence of a substrate, vibrational relaxation of nitric oxide occurred more rapidly in H_2O than D_2O. Structural fluctuations within the haem cavity similarly occur at faster rates under aqueous conditions. This was attributed to a network of solvent molecules at the distal site. These waters are display highly constrained dynamics, likely imposed through interactions with polar amino acid side chains in the cavity. Direct contact of the ligand and the functional residue H42 and R38 is not well-represented by these results.

Binding of a substrate to HRP eliminates isotope sensitivity in the dynamics previously mentioned; the vibrational lifetimes and inhomogeneous broadening rates of NO imply that a substrate is capable of removing water molecules from the vicinity of the ligand. Despite this, the measured timescales for structural fluctuation lie close to those for substrate-free HRP, suggesting that dynamic timescales may be dominated by motion of the polar distal residues in each case.

Pump-probe and 2DIR measurements demonstrate that the catalase active site similarly hosts a solvent network. A combination of fast and slow time components for structural rearrangement indicates a rigid distal structure that is occupied by a number of resident solvent molecules in addition to structural waters. The dynamics of this solvent pool show limited bulk-solvent characteristics due to interactions with the protein architecture. Crucially, in both catalase and HRP, contact between the distal scaffold and the active site ligand appears to be mediated by solvent. This provides strong support for the role of water in peroxidase and catalase functional cycles.

3.6 Future work

HRP has the potential to bind numerous substrates at its aromatic binding site. HRP:BHA complex formation is predicted to be particularly favourable due to the polar interaction of the substrate's head group with distal cavity residues. Investigations of substrates with ranging K_d values and polarities may therefore provide additional detail on the nature of the water network at the haem by exploring the variation of structural dynamics near the threshold at which substrate polarity is suitable to displace the intra-cavity solvent. Likewise, spectral diffusion studies of alternate HRP complexes may reveal the importance of specific substrates in determining structural fluctuation times, exemplifying a substrate recognition mechanism in the peroxidase.

3.7 References

1. Reedy, C. J. & Gibney, B. R. Heme protein assemblies. *Chem. Rev.* **104,** 617–649 (2004).
2. Forrester, M. T. & Foster, M. W. Protection from nitrosative stress: a central role for microbial flavohemoglobin. *Free Radic. Biol. Med.* **52,** 1620–1633 (2012).
3. Laberge, M. & Yonetani, T. Common dynamics of globin family proteins. *IUBMB Life* **59,** 528–534 (2007).
4. Mense, S. M. & Zhang, L. Heme: a versatile signalling molecule controlling the activities of diverse regulators ranging from transcription factors to MAP kinases. *Cell Res.* **16,** 681–692 (2006).
5. Wendehenne, D., Pugin, A., Klessig, D. F. & Durner, J. Nitric oxide: comparative synthesis and signalling in animal and plant cells. *Trends Plant Sci.* **6,** 177–183 (2001).
6. Suslick, K. S. Shape-selective oxidation by metalloporphyrins. *Porphyr. Handb.* **4,** 41–63 (2000).
7. Michel, H., Behr, J., Harrenga, A. & Kannt, A. Cytochrome c oxidase: structure and spectroscopy. *Annu. Rev. Biophys. Biomol. Struct.* **27,** 329–356 (1998).
8. Lesk, A. M., Chothia, C. How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins. *J. Mol. Biol.* **136,** 225–270 (1980).
9. Welinder, K. G. Plant peroxidases: their primary, secondary and tertiary structures, and relation to cytochrome c peroxidase. *Eur. J. Biochem.* **151,** 497–504 (1985).
10. Dayhoff, M. O. Atlas of protein sequence and structure. **5,** (1978).
11. Suzuki, T., Watanabe, Y. H., Nagasawa, M., Matsuoka, A. & Shikama, K. Dual nature of the distal histidine residue in the autoxidation reaction of myoglobin and hemoglobin: Comparison of the H64 mutants. *Eur. J. Biochem.* **267,** 6166–6174 (2000).
12. Ishikawa, H. *et al.* Neuroglobin dynamics observed with ultrafast 2D-IR vibrational echo spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 16116–16121 (2007).
13. Gotoh, O. Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *J. Biol. Chem.* **267,** 83–90 (1992).
14. Thielges, M. C., Chung, J. K. & Fayer, M. D. Protein Dynamics in Cytochrome P450 Molecular Recognition and Substrate Specificity Using 2D IR Vibrational Echo Spectroscopy. *J. Am. Chem. Soc.* **133,** 3995–4004 (2011).

15. Doster, W. *et al.* Control and pH dependence of ligand binding to heme proteins. *Biochemistry (Mosc.)* **21,** 4831–4839 (1982).

16. Roncone, R. *et al.* Engineering peroxidase activity in myoglobin: the haem cavity structure and peroxide activation in the T67R/S92D mutant and its derivative reconstituted with protohaemin-l-histidine. *Biochem. J.* **377,** 717–724 (2004).

17. Urakubo, Y., Ikura, T. & Ito, N. Crystal structural analysis of protein-protein interactions drastically destabilized by a single mutation. *Protein Sci.* **17,** 1055–1065 (2008).

18. Pesce, A. *et al.* Human brain neuroglobin structure reveals a distinct mode of controlling oxygen affinity. *Structure* **11,** 1087–1095 (2003).

19. Pesce, A. *et al.* Neuroglobin and cytoglobin. *EMBO Rep.* **3,** 1146–1151 (2002).

20. Ropp, J. S. D., Mandal, P., Brauer, S. L. & Mar, G. N. L. Solution ^{1}NMR study of the electronic and molecular structure of the heme cavity in high-spin, resting state horseradish peroxidase. *J Am Chem Soc* **7863,** 5400–5407 (1997).

21. Thanabal, V., Ropp, J. S. D. & La, G. N. Proton NMR characterization of the catalytically relevant proximal and distal hydrogen-bonding networks in ligated resting state horseradish peroxidase. *J. Am. Chem. Soc.* **110** 3027–3035 (1988).

22. Wang, X. Two-dimensional NMR study of the heme active site structure of chloroperoxidase. *J. Biol. Chem.* **278,** 7765–7774 (2003).

23. Spiro, T. G., Soldatova, A. V. & Balakrishnan, G. CO, NO and O2 as Vibrational Probes of Heme Protein Interactions. *Coord. Chem. Rev.* **257,** 511–527 (2013).

24. Bracken, C. S., Baer, M. T., Abdur-Rashid, A., Helms, W. & Stojiljkovic, I. Use of heme-protein complexes by the Yersinia enterocolitica HemR receptor: histidine residues are essential for receptor function. *J. Bacteriol.* **181,** 6063–6072 (1999).

25. Hou, S. *et al.* Globin-coupled sensors: a class of heme-containing sensors in Archaea and Bacteria. *Proc. Natl. Acad. Sci.* **98,** 9353–9358 (2001).

26. Merchant, K. A. *et al.* Myoglobin-CO conformational substate dynamics: 2D vibrational echoes and MD simulations. *Biophys. J.* **82,** 3277–3288 (2002).

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

28. Hamm, P. & Zanni, M. T. *Concepts and methods of 2D infrared spectroscopy*. (Cambridge University Press, 2011).

29. Bagchi, S., Nebgen, B. T., Loring, R. F. & Fayer, M. D. Dynamics of a myoglobin mutant enzyme: 2D IR vibrational echo experiments and simulations. *J. Am. Chem. Soc.* **132,** 18367–18376 (2010).

30. 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* **117,** 15804–15811 (2013).

31. Adamczyk, K. *et al.* The effect of point mutation on the equilibrium structural fluctuations of ferric myoglobin. *Phys. Chem. Chem. Phys. PCCP* **14,** 7411–7419 (2012).

32. Bagchi, S., Thorpe, D. G., Thorpe, I. F., Voth, G. A. & Fayer, M. D. Conformational switching between protein substates studied with 2D IR vibrational echo spectroscopy and molecular dynamics simulations. *J Phys Chem B* **114,** 17187–17193 (2010).

33. Ishikawa, H., Kim, S., Kwak, K., Wakasugi, K. & Fayer, M. D. Disulfide bond influence on protein structural dynamics probed with 2D-IR vibrational echo spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 19309–19314 (2007).

34. Thielges, M. C., Chung, J. K., Axup, J. Y. & Fayer, M. D. Influence of histidine tag attachment on picosecond protein dynamics. *Biochemistry (Mosc.)* **50,** 5799–5805 (2011).

35. Ye, X., Yu, A. & Champion, P. M. Dynamics of nitric oxide rebinding and escape in horseradish peroxidase. *J. Am. Chem. Soc.* **128,** 1444–1445 (2006).

36. Finkelstein, I. J., Ishikawa, H., Kim, S., Massari, A. M. & Fayer, M. D. Substrate binding and protein conformational dynamics measured by 2D-IR vibrational echo spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 2637–2642 (2007).

37. Williams, R. B., Loring, R. F. & Fayer, M. D. Vibrational dephasing of carbonmonoxy myoglobin. *J. Phys. Chem. B* **105,** 4068–4071 (2001).

38. Zeng, W., Sun, Y., Benabbas, A. & Champion, P. M. Investigations of ferric heme cyanide photodissociation in myoglobin and horseradish peroxidase. *J. Phys. Chem. B* **117,** 4042–4049 (2013).

39. Yoshikawa, S., O'Keeffe, D. H. & Caughey, W. S. Investigations of cyanide as an infrared probe of hemeprotein ligand binding sites. *J. Biol. Chem.* **260,** 3518–3528 (1985).

40. Zimmermann, J., Thielges, M. C., Seo, Y. J., Dawson, P. E. & Romesberg, F. E. Cyano groups as probes of protein microenvironments and dynamics. *Angew. Chem. Int. Ed.* **50,** 8333–8337 (2011).

41. Blumenthal, C. Azide binding to the cytochrome c ferric heme octapeptide. *J. Biol. Chem.* **254,** 9617–9620 (1979).

42. Keilin, D. & Hartree, E. F. Properties of azide-catalase. *Biochem. J.* **39,** 148–157 (1945).

43. Maurus, R., Bogumil, R., Nguyen, N. T., Mauk, A. G. & Brayer, G. Structural and spectroscopic studies of azide complexes of horse heart myoglobin and the His-64 Thr variant. **74,** 67–74 (1998).

44. McCoy, S. & Caughey, W. S. Infrared studies of azido, cyano, and other derivatives of metmyoglobin, methemoglobin, and hemins. *Biochemistry (Mosc.)* **9,** 2387–2393 (1970).

45. Thanabal, V., Ropp, J. S. D. & La, G. N. Proton NMR characterization of the catalytically relevant proximal and distal hydrogen-bonding networks in ligated resting state horseradish peroxidase. *Biochem. J.* **15** 3027–3035 (1988).

46. Kim, S. *et al.* Native and unfolded cytochrome C comparison of dynamics using 2D-IR vibrational echo spectroscopy. *J. Phys. Chem. B* **112,** 10054–10063 (2008).

47. Sathyanarayana, D. N. *Vibrational spectroscopy: theory and applications*. (New Age International, 2007).

48. Bickar, D., Bonaventura, C. & Bonaventura, J. Carbon monoxide-driven reduction of ferric heme and heme proteins. *J. Biol. Chem.* **259,** 10777–10783 (1984).

49. Bellota-Antón, C. *et al.* Spectroscopic analysis of protein Fe-NO complexes. *Biochem. Soc. Trans.* **39,** 1293–8 (2011).

50. Nutt, D. R. & Meuwly, M. Ferric and ferrous iron in nitroso-myoglobin: computer simulations of stable and metastable States and their infrared spectra. *Chemphyschem Eur. J. Chem. Phys. Phys. Chem.* **8,** 527–536 (2007).

51. Brookes, J. F., Slenkamp, K. M., Lynch, M. S. & Khalil, M. Effect of solvent polarity on the vibrational dephasing dynamics of the nitrosyl stretch in an Fe(II) complex revealed by 2D IR spectroscopy. *J. Phys. Chem. A* **117,** 6234–43 (2013).

52. Wand, J. A. Dynamic activation of protein function: a view emerging from NMR spectroscopy. *Nat. Struct. Biol.* **8,** 926–931 (2001).

53. Azevedo, A. M., Cabral, J. M. S. & Fonseca, L. P. Horseradish peroxidase : a valuable tool in biotechnology. *Biotechnol. Annu. Rev.* **9,** 199–247 (2003).

54. Veitch, N. C. Horseradish peroxidase: a modern view of a classic enzyme. *Phytochemistry* **65,** 249–259 (2004).

55. Vidossich, P., Alfonso-Prieto, M. & Rovira, C. Catalases versus peroxidases: DFT investigation of H2O2 oxidation in models systems and implications for heme protein engineering. *J. Inorg. Biochem.* **117,** 292–297 (2012).

56. Jones, P. Roles of water in heme peroxidase and catalase mechanisms. *J. Biol. Chem.* **276,** 13791–13796 (2001).

57. Jones, P. & Dunford, H. B. The mechanism of compound I formation revisited. *J. Inorg. Biochem.* **99,** 2292–2298 (2005).

58. Chuang, W. J. & Van Wart, H. E. Resonance raman spectra of horseradish peroxidase and bovine liver catalase compound I species. *J. Biol. Chem.* **267,** 13293–13301 (1992).

59. Derat, E. & Shaik, S. The poulos−kraut mechanism of compound I formation in horseradish peroxidase: a QM/MM study. *J. Phys. Chem. B* **110,** 10526–10533 (2006).

60. Poulos, T. L. & Kraut, J. The stereochemistry of peroxidase catalysis. *J. Biol. Chem.* **255,** 8199–8205 (1980).

61. Gajhede, M., Schuller, D. J., Henriksen, A., Smith, A. T. & Poulos, T. L. Crystal structure of horseradish peroxidase C at 2.15 A resolution. *Nat. Struct. Biol.* (1997). doi:10.1038/nsb1297-1032

62. Candelaresi, M. *et al.* A structural and dynamic investigation of the inhibition of catalase by nitric oxide. *Org. Biomol. Chem.* 7778–7788 (2013).

63. Gumiero, A., Metcalfe, C. L., Pearson, A. R., Raven, E. L. & Moody, P. C. E. Nature of the ferryl heme in compounds I and II. *J. Biol. Chem.* **286,** 1260–1268 (2011).

64. Hernandez-Ruiz, J., Arnao, M., Hiner, A., Garcia-Canovas, F. & Acosta, M. Catalase-like activity of horseradish peroxidase: relationship to enzyme inactivation by H2O2. *Biochem J* **354,** 107–114 (2001).

65. Berglund, G. I. *et al.* The catalytic pathway of horseradish peroxidase at high resolution. *Nature* **417,** 463–468 (2002).

66. Jakopitsch, C. *et al.* Role of the main access channel of catalase-peroxidase in catalysis. *J. Biol. Chem.* **280,** 42411–42422 (2005).

67. Kalko, S. G., Gelpi, J. L., Fita, I. & Orozco, M. Theoretical study of the mechanisms of substrate recognition by catalase. *J Am Chem Soc* **123,** 9665–9672 (2001).

68. Fülöp, V. *et al.* Laue diffraction study on the structure of cytochrome c peroxidase compound I. *Structure* **2,** 201–208 (1994).

69. Bretscher, K. R. & Jones, P. Kinetics of the decomposition of coprohaem compound I in aqueous solution. *J. Chem. Soc. Dalton Trans.* 2273–2278 (1988).

70. Rodríguez-López, J. N., Gilabert, M. A., Tudela, J., Thorneley, R. N. F. & García-Cánovas, F. Reactivity of horseradish peroxidase compound II toward substrates: kinetic evidence for a two-step mechanism. *Biochemistry (Mosc.)* **39,** 13201–13209 (2000).

71. Vidossich, P. *et al.* On the role of water in peroxidase catalysis: a theoretical investigation of HRP compound I formation. *J. Phys. Chem. B* **114,** 5161–5169 (2010).

72. Derat, E., Shaik, S., Rovira, C., Vidossich, P. & Alfonso-Prieto, M. The effect of a water molecule on the mechanism of formation of compound 0 in horseradish peroxidase. *J. Am. Chem. Soc.* **129,** 6346–6347 (2007).

73. Henriksen, a *et al.* Structural interactions between horseradish peroxidase C and the substrate benzhydroxamic acid determined by X-ray crystallography. *Biochemistry (Mosc.)* **37,** 8054–8060 (1998).

74. Chang, Y., Veitch, N. C. & Loew, G. H. A theoretical study of benzhydroxamic acid binding modes in horseradish peroxidase. *J. Am. Chem. Soc.* **120,** 5168–5178 (1998).

75. Howes, B. D. *et al.* Mutation of residues critical for benzohydroxamic acid binding to horseradish peroxidase isoenzyme C. *Biopolymers* **62,** 261–267 (2001).

76. Laberge, M., Osvath, S. & Fidy, J. Aromatic substrate specificity of horseradish peroxidase C studied by a combined fluorescence line narrowing/energy minimization approach: the effect of localized side-chain reorganization. *Biochemistry (Mosc.)* **40,** 9226–9237 (2001).

77. Szigeti, K., Smeller, L., Osváth, S., Majer, Z. & Fidy, J. The structure of horseradish peroxidase C characterized as a molten globule state after Ca(2+) depletion. *Biochim. Biophys. Acta* **1784,** 1965–1974 (2008).

78. Khajehpour, M. *et al.* Accessibility of oxygen with respect to the heme pocket in horseradish peroxidase. *Proteins* **53,** 656–666 (2003).

79. Zelent, B. *et al.* Water channel of horseradish peroxidase studied by the charge-transfer absorption band of ferric heme. *J. Phys. Chem. B* **108,** 10317–10324 (2004).

80. Kaposi, A. D. *et al.* Solvent dependent and independent motions of CO-horseradish peroxidase examined by infrared spectroscopy and molecular dynamics calculations. *Biophys. Chem.* **106,** 1–14 (2003).

81. Seccia, M., Perugini, C., Albano, E. & Bellomo, G. Inhibition of Cu 2+ induced LDL oxidation by nitric oxide: A study using donors with different half-time of NO release. *Biochem. Biophys. Res. Commun.* **220,** 306–309 (1996).

82. Acton, Q. A. *Peroxidases: Advances in Research and Application*. (Scholarly Editions, 2011).

83. Cooper, C. E. Nitric oxide and iron proteins. *Biochim. Biophys. Acta - Bioenerg.* **1411,** 290–309 (1999).

84. Kikuchi, K., Nagano, T. & Hirobe, M. Novel detection method of nitric oxide using horseradish peroxidase. *Biol. Pharm. Bull.* **19,** 649–651 (1996).

85. Hunt, N. T., Greetham, G. M., Towrie, M., Parker, A. W. & Tucker, N. P. Relationship between protein structural fluctuations and rebinding dynamics in ferric haem nitrosyls. *Biochem. J.* **433,** 459–68 (2011).

86. Rector, K. D. *et al.* Mutant and wild-type myoglobin-CO protein dynamics : vibrational echo experiments. *J. Phys. Chem*. B **101,** 1468–1475 (1997).

87. Bakker, H. J. Vibrational relaxation in the condensed phase. *J Chem Phys* **121,** 10088–10095 (2004).

88. Whitnell, R. M., Wilson, K. R. & Hynes, J. T. Fast vibrational relaxation for a dipolar molecule in a polar solvent. *J. Phys. Chem.* **94,** 8625–8628 (1990).

89. Woutersen, S. & Bakker, H. J. Resonant intermolecular transfer of vibrational energy in liquid water. *Nature* **402,** 507–509 (1999).

90. Li, M. *et al.* Vibrational and rotational relaxation times of solvated molecular ions. *J. Chem. Phys.* **98,** 5499–5507 (1993).

91. Piletic, I. R., Tan, H.-S. & Fayer, M. D. Dynamics of nanoscopic water: vibrational echo and infrared pump−probe studies of reverse micelles. *J. Phys. Chem. B* **109,** 21273–21284 (2005).

92. Tokmakoff, A., Sauter, B. & Fayer, M. D. Temperature-dependent vibrational relaxation in polyatomic liquids: picosecond infrared pump–probe experiments. *J. Chem. Phys.* **100,** 9035–9043 (1994).

93. Ramasesha, K., De Marco, L., Mandal, A. & Tokmakoff, A. Water vibrations have strongly mixed intra- and intermolecular character. *Nat. Chem.* **5,** 935–940 (2013).

94. King, J. T., Ross, M. R. & Kubarych, K. J. Water-assisted vibrational relaxation of a metal carbonyl complex studied with ultrafast 2D-IR. *J. Phys. Chem. B* **116,** 3754–3759 (2012).

95. Yu, P., Yang, F., Zhao, J. & Wang, J. Hydration dynamics of cyanoferrate anions examined by ultrafast infrared spectroscopy. *J. Phys. Chem. B* **118,** 3104–3114 (2014).

96. Kozlikova, B. *et al.* CAVER Analyst 1.0: graphic tool for interactive visualization and analysis of tunnels and channels in protein structures. *Bioinformatics* **30,** 2684–2685 (2014).

97. Bakulin, A. A. *et al.* Dynamics of water confined in reversed micelles: multidimensional vibrational spectroscopy study. *J Phys Chem B* **117,** 15545–15558 (2013).

98. Gilfoyle, D. J., Rodriguez-Lopez, J. N. & Smith, A. T. Probing the aromatic-donor-binding site of horseradish peroxidase using site-directed mutagenesis and the suicide substrate phenylhydrazine. *Eur. J. Biochem.* **236,** 714–722 (1996).

99. Henriksen, A., Smith, A. T. & Gajhede, M. The structures of the horseradish peroxidase C-ferulic acid complex and the ternary complex with cyanide suggest how peroxidases oxidize small phenolic substrates. *J. Biol. Chem.* **274,** 35005–35011 (1999).

100. Peroxidase, C. H., Ropp, J. S. D., Mandal, P. K. & Mar, G. N. L. Solution ^{1}H NMR investigation of the heme cavity and substrate binding site in cyanide-inhibited horseradish peroxidase, *Biochemistry* **38,** 1077–1086 (1999).

4. Infrared Spectroscopy of Nicotinamide Adenine Dinucleotides in One and Two Dimensions

This chapter contains results published in the following paper:

Simpson, N.; Shaw, D. J.; Frederix, P. W. J. M.; Gillies, A. H.; Adamczyk, K.; Greetham, G. M.; Towrie, M.; Parker, A. W.; Hoskisson, P. A.; Hunt, N. T., *Infrared spectroscopy of Nicotinamide Adenine Dinucleotides in One and Two Dimensions*, The Journal of Physical Chemistry B, 117, 16488-16478, (2013)

4.1 Abstract

The chemical reactions of dinucleotides underpin a variety of physiological pathways. In stark contrast to proteins, which are finely-tuned to perform a specific task, these comparatively small biomolecules are able to execute multiple functions in partnership with enzymes, without the necessity of numerous sub-species. As a consequence of this versatility in binding, a single dinucleotide may potentially be exploited as a probe to access information of the structures of its many target molecules.

In this study, we categorise the IR properties of nicotinamide adenine dinucleotide (NAD), a common biological substrate with distinctive absorption characteristics in the Amide I region. As ubiquitous ligand, NAD has the potential to be an incredibly practical IR probe-molecule that can be employed without compromising the applicability to a biological system. With this objective a full vibrational band assignments are constructed for the dinucleotide and its structural analogues, with support from parallel DFT calculations and ^{1}H NMR spectroscopy. The effectiveness of NAD as an intrinsic probe of local protein dynamics is discussed.

4.2 Introduction

The coenzyme, NAD, is a crucial component of cells[1,2,3], participating in many biological pathways by virtue of a flexible structure that is recognised by a multitude of proteins[4–8]. Within its roles in cellular respiration, electron transport and glycolysis, NAD undergoes transition between two forms: the redox partners NAD+ and NADH[2] (Figure (1)).

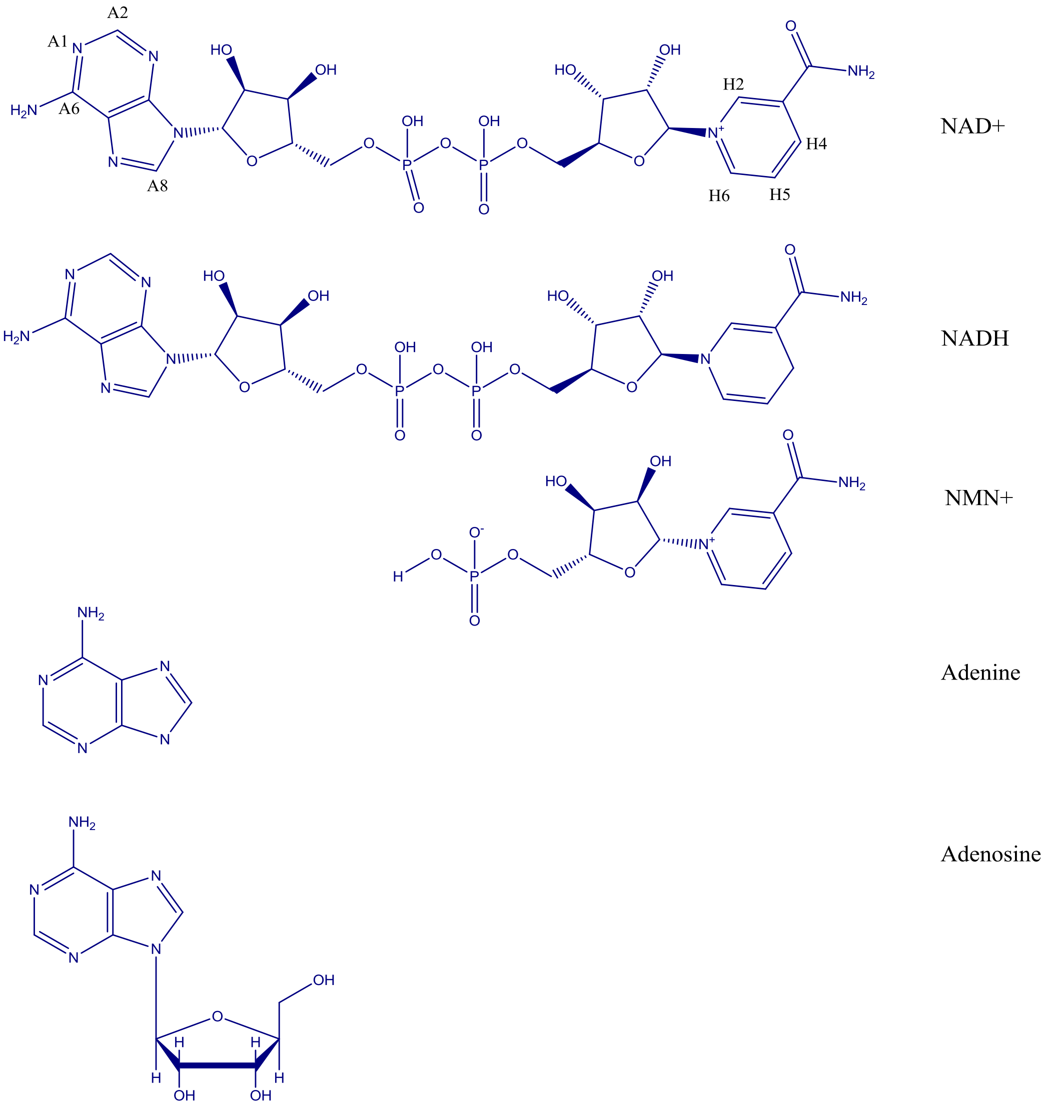


Figure (1): Chemical structures NAD+, NADH and their studied structural analogues. Numbering of the atoms in NAD+ denotes proton NMR assignments.

Interconversion between NADH and NAD+ occurs through the loss of a hydride ion from the nicotinamide group. The "active" form of the coenzyme is seen to vary, as both protein oxidation and reduction are required in the electron transport chain, facilitated by NAD+ and NADH, respectively[9,10]. Structural recognition of the coenzyme structure, and further, discrimination of its oxidation states, is resultantly vital to the transport and storage of energy in biological systems.

Proteins bind nicotinamide adenine dinucleotide with a structural motif known as a "Rossman Fold"[11], composed up to seven secondary structure domains and containing a mixture of α-helices and β-sheets, though normally favouring the latter[12,13]. Specific residues within the Rossman Fold are strictly conserved, though many other features appear variable. Despite this, the bound conformations of NAD are typically seen to fall into only a few broad sub-categories[14].

In Chapter (3), nitric oxide was used to report on the structural dynamics of enzyme active sites. This experimental approach is highly effective in the study of haem proteins, as it exploits the substrate-binding affinity of the haem group to deploy an IR probe at a functionally-relevant site. Of course, diatomic molecules cannot be used to study proteins outside of this sub-class. However, it may be possible to exploit the Rossman Fold in an analogous manner to the haem group, allowing the observation of protein dynamics at structural regions linked to energy transfer and storage. Demonstration of NAD as an effective IR probe would thus create a new category of 2DIR investigations. This requires a thorough categorisation of the underlying IR spectroscopy of NAD+ and NADH.

Many biological assays infer protein activity from the rate of conversion of NADH to NAD+[15]. Oxidation alters the pyridine ring from a di-hydro configuration, with two localised C=C bonds, to an aromatic ring system with delocalised bonds, causing significant change to the nicotinamide group's electronic arrangement. This creates pronounced differences in the UV-Vis[16] absorption of NADH and NAD+, which have served as a historically-important reference for protein activity studies. Categorisation of NAD redox states has also been documented *via* fluorescence[17], NMR[18,19, 20,21] and infrared spectroscopies[9]. In the latter case, though clear spectral changes have been associated with coenzyme oxidation and reduction, these have not been thoroughly assigned to specific functional group vibrations. Further, despite the range of techniques with which the coenzyme has been studied, its solution-phase conformation has not been demonstrated conclusively. Much attention has been paid to protein-bound forms of the coenzyme, whose crystallised complexes reveals an overwhelming preference for an extended, or unfolded, geometry[22,23,24,25,26]. In the solution-phase, however, experimental data from circular dichroism (CD)[27], fluorescence spectroscopy[28,22,24] and NMR[19,29] predict that NAD occupies a tightly-folded conformation, with the adenine and nicotinamide moieties linked *via* a parallel stacking interaction. Alternate studies[18,30,31,32] propose a dynamic equilibrium between the folded and extended forms, with a slight preference for folded conformation observed under neutral, aqueous conditions. Most recently, a description has been proposed that lies somewhere between the previous stances: NAD rests in a folded conformation, but with a less compact structure than formerly suggested[4,5,13]. These contemporary molecular dynamics (MD) simulations identify nicotinamide solvation as a force acting against parallel ring stacking and complete burial of the head-groups from the solvent. A limited degree of nucleobase solvent-exposure provides a possible explanation for the differing points of view of NAD's structure. Accordingly, MD results support the findings of several experimental studies retrospectively[31,33], though not in all cases.

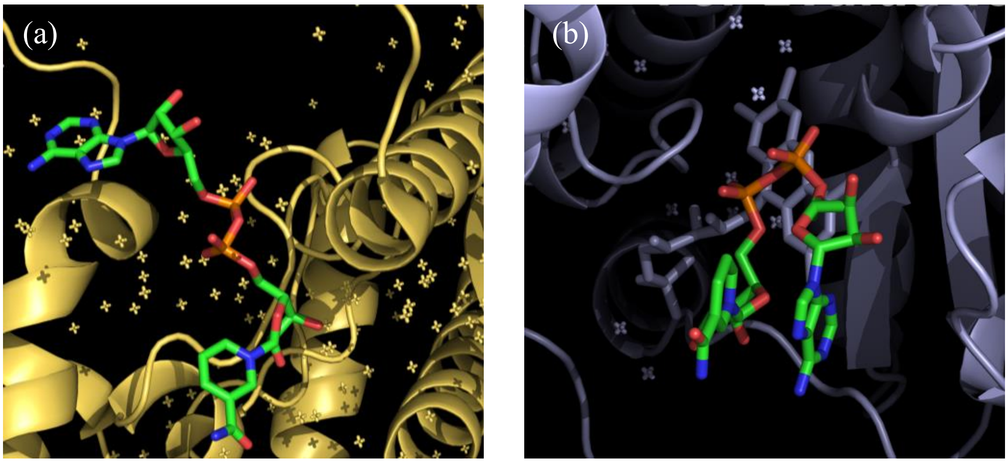


Figure (2): (a) NAD, bound in an extended conformation to the enzyme, NADS (nicotinamide adenine dinucleotide synthetase) from E. coli (PDB Code: 1WXH)[12]*, and (b) NAD+ bound to flavin reductase p in a less commonly encountered, folded conformation (PDB Code: 2BKJ)*[26]

It has been suggested that the hydrophobic environments of protein binding sites trigger a conformation change in the coenzyme. Studies of NAD in nonpolar solvents such as methanol and urea[7,18] have given support to the concept of such a conformation transition that occurs from polar to nonpolar environments. From the previously discussed MD simulations it is consistent that less polar environments could be more favourable for adenine and nicotinamide group solvation, in turn making stacking of the rings less favourable. It has also been suggested that, at low pD (<4), protonation of the adenine A1 atom in NAD+ is sufficient to repel the positively charged head groups, creating an extended conformation[18,29].

Nicotinamide adenine dinucleotide is an atypical IR probe, owing to its complex structure. Nonetheless, its ability to bind a multitude of proteins without compromising the native state of the complex makes it a highly desirable reporter. Further, NAD is formed of moieties (and resultantly functional groups) that are not found in polypeptide structures, potentially giving rise to coenzyme vibrations that are easily distinguishable from those of a protein. With the objective of using NAD as a probe molecule in the infrared, we conduct a full band assignment of NAD and its structural analogues through pD dependence experiments that mirror important ^{1}H NMR studies of the dinucleotide. Further, the coenzyme is studied in non-polar solvents in order to determine IR indicators for folded and unfolded conformations. The findings throughout are supported with DFT calculations. The results of these DFT calculations are also used to determine the viability of specific modes as probes; localisation of a vibration is highly beneficial in an IR probe, as localised vibrations are more likely to be sensitive to local chemical environment. In the absence of a computational method such as density functional theory, determining the ectent of localisation in a vibration is not possible. Computational results and spectroscopic data from FTIR, pump-probe and 2D-IR are used to discuss the suitability of NAD a biomolecular probe.

4.3 Experimental

All samples (NAD+, NADH, adenine, adenosine, nicotinamide mononucleotide and nicotinamide) were purchased from Sigma Aldrich and used without further purification. Deuterated solvents were employed throughout: either deuterium oxide or a mixture of methanol D-4 with deuterium oxide. Similarly, in pD dependence studies, solution pD was corrected using deuterium chloride and sodium deuteroxide to minimise H-D exchange and subsequent spectral changes.

4.3.1 Infrared spectroscopy

All linear infrared absorption spectra were acquired at room temperature (~23 °C) using a Bruker Vertex 70 spectrometer. Samples were housed between two CaF_2 windows separated by a PTFE spacer of 50 µm thickness. Sample concentrations were 100 mM throughout, with the sole exception of methanol dependence studies for which the existing protocols of Catterall *et al.*[18] were observed, resulting in solute concentrations of 50 mM. Spectra were collected at a resolution of 1 cm^{-1}.

All pump-probe and 2DIR measurements were similarly performed with a 50µm path length, using solute concentrations of 100mM. 2DIR spectra presented in this chapter were collected using the ULTRA laser system, and each spectrum is the average of 8 separate measurements. Within measurements, the probe frequency spectra measured at each value of the interferometer stage (i.e. at each value of t_1) are the average of 3000 laser shots. Pump probe data was obtained by averaging 3 individual measurements, with the signal at each value of the delay stage established from 20,000 laser shots.

4.3.2 ^{1}H NMR

Spectra were acquired with an Oxford Instruments NMR AS400 operating at 400 MHz. ^{1}H NMR samples were prepared from the same solutions as used for FTIR measurements to ensure consistency, and were measured within 24 hours of the equivalent IR scans. Additional information on this technique is provided in Appendix (A.4.1)

4.3.3 Density Functional Theory (DFT) Calculations

Vibrational band assignment was assisted *via* DFT calculations for NADH, NAD+ and their structural analogues, using Gaussian09 v.AO2[34]. Geometry optimisation and IR frequency calculation was carried out using a B3LYP density functional and a Pople valence triple zeta basis set (6-311g) with one diffuse and polarisation function (+G(d,p))[35]. A correction factor of 0.9679 was applied to vibrational frequencies computed under this basis set[36]. Calculations incorporated solvent (water) in the form of a polarisable continuum model (PCM) and accounted for exchangeable protons, as described below.

Preliminary geometry optimisation and frequency calculation for NAD found that the initial conformation of the coenzyme (i.e. folded or unfolded) made no significant difference to the subsequent results. Firstly, geometry optimisation did not lead to any conformation change during the course of the calculation, regardless of whether the starting conformation was folded or extended. Secondly, folding of the NAD was not seen to have any effect on the spectrum. For this reason all subsequent calculations were carried out with NAD in an open, extended form.

The potential effects of folding were deliberately exacerbated in one calculation where the coenzyme was placed in a folded conformation despite protonation of its adenine group. Only in this situation, and additionally, only when modelled in the gas-phase, did folding have an apparent effect on the spectrum. This result was dismissed when considering the importance of folding in the initial conditions of DFT calculations, as it is documented that protonation of the adenine ring of NAD+ results in a repulsive effect between the two nucleobase groups that causes unfolding of the molecule[18,29]. This circumstance is therefore physically unattainable, and was used solely as a means of identifying the vibrational modes most sensitive to folding through an exaggeration of the effect.

4.4 Results and Discussion

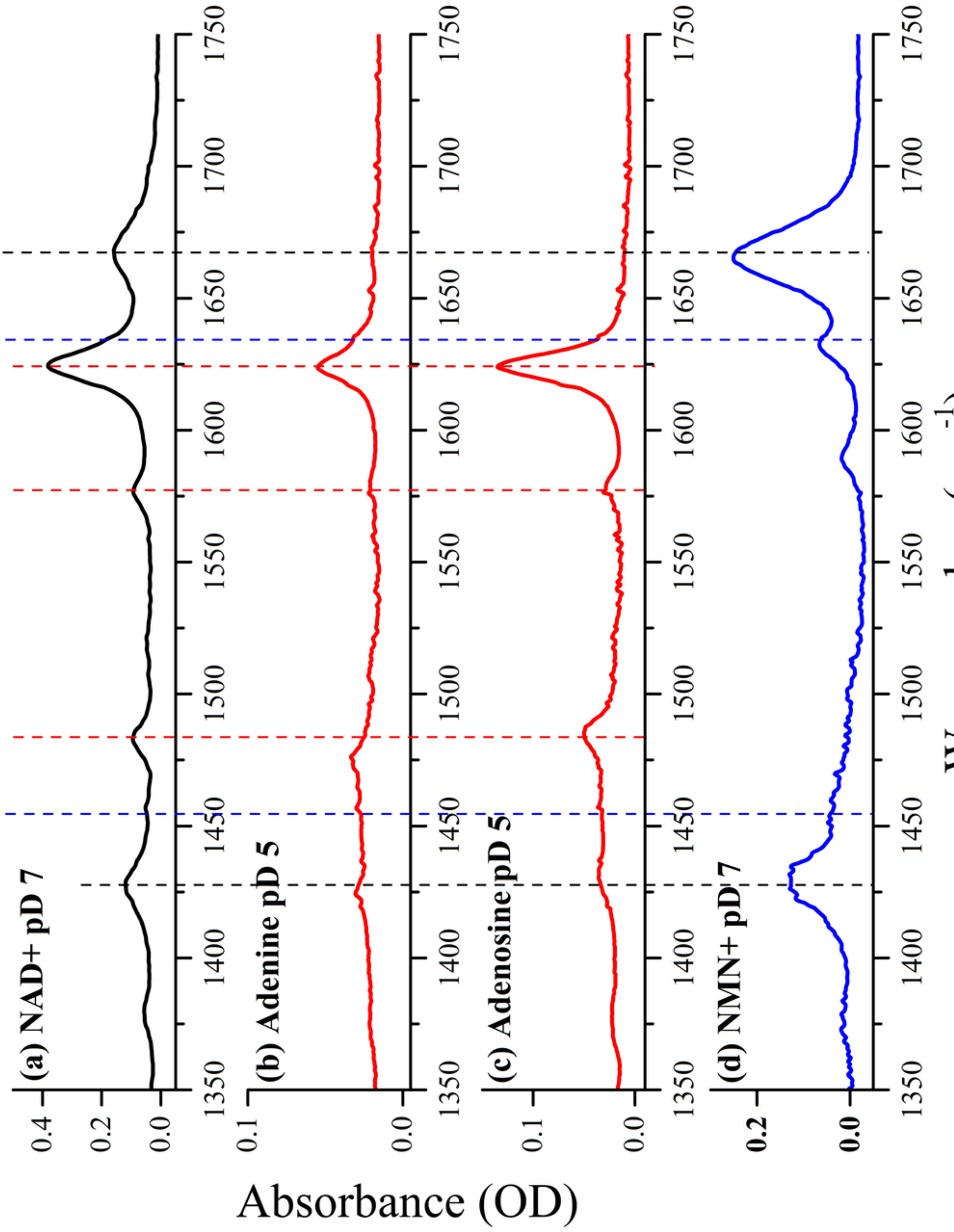


Figure (3): The FTIR spectra of NAD+ and its structural analogues: (a) NAD+ at pD 7, (b) adenine at pD 5, (c) adenosine at pD 5, and (d) NMN+ at pD 7. Dashed vertical lines show the breakdown of NAD+ modes to each to its structural analogues.

4.4.1 Vibrational band assignment of NAD+

The FTIR spectrum of NAD+ at neutral pD is displayed in Figure (3). Attention is only paid to the absorption bands between 1400 and 1750 cm^{-1}; those that would be captured within the spectral range of a typical 2DIR study in the Amide I’ region. Within this region, the NAD+ spectrum has five prominent absorption bands, occurring at: 1428, 1482, 1577, 1624, and 1666 cm^{-1}, in addition to two weaker features at 1457 and 1633 cm^{-1}.

A side-by-side comparison of NAD+ is made with its structural analogues: adenine, adenosine and nicotinamide mononucleotide (NMN+) (Figure (3)(b)-(d)). NMN+ is shown at neutral pD, though the spectra of adenine and adenosine are included at pD 5. The solubilities of adenine and adenosine were seen to decrease rapidly at pD > 2.5, with concentrations of ~100mM only obtainable at low pD. Between pD 2 and 7, reductions in the signal to noise ratio were ~5.8 and 16 for adenosine and adenine, respectively, and a pD of 5 was ultimately chosen for display as this represents the lowest pD (highest solubility) at which the moieties are still completely consistent with their neutral spectra. In Figure (4) adenine and adenosine are scaled to match the intensities of their corresponding 1428 cm^{-1} mode at pD 2. The choice of analogues shown in this figure demonstrates the sizeable contribution of nucleobase vibrations in the NAD+ spectrum, particularly in comparison to those of the ribose and pyrophosphate groups.

A rudimentary assignment of oxidised coenzyme bands is achieved by assuming that the vibrations of each constituent molecule translate directly into the NAD+ spectrum without any spectral changes associated with dinucleotide formation. Following this, it is seen that the NAD+ peaks at 1428 cm^{-1} and 1457 cm^{-1} likely contain contributions from both NMN+ and adenine. Higher-frequency vibrational modes appear more clearly-separated, allowing the assignment of peaks at 1482, 1577 and 1625 cm^{-1} to adenosine (or adenine) bands alone. The primary feature of the NMN+ spectrum, at 1666 cm^{-1} remains prominent in the NAD+ spectrum, though the mononucleotide features at 1588 and 1633 cm^{-1} are less pronounced in NAD+. Asymmetry of the adenine peak at 1625 cm^{-1} suggests that the 1633 cm^{-1} band of nicotinamide mononucleotide occupies a similar frequency within NAD+, though it is heavily obscured by the strong absorption of adenine in this region. The 1588 cm^{-1} vibration of NMN+ is not resolved within NAD+, though, from a quantitative standpoint, this might be expected; if the relative intensities of the NMN+ peaks were retained in NAD+, and the 1666 cm^{-1} intensity were used as a point of reference, one would expect only a minor absorption at 1588 cm^{-1}.

It is noted that the spectral differences between adenine and adenosine are minor; demonstrating how small an effect the addition of a ribose group has on the spectrum, both in terms of adding new features and altering the prior vibrations of adenine. This is consistent with previous data[9], and thus the spectroscopy of the free ribose group is not considered this study. The effect of the ribose upon nicotinamide vibrations was likewise not investigated here, as an oxidised form of the free nicotinamide head group is not readily attainable.

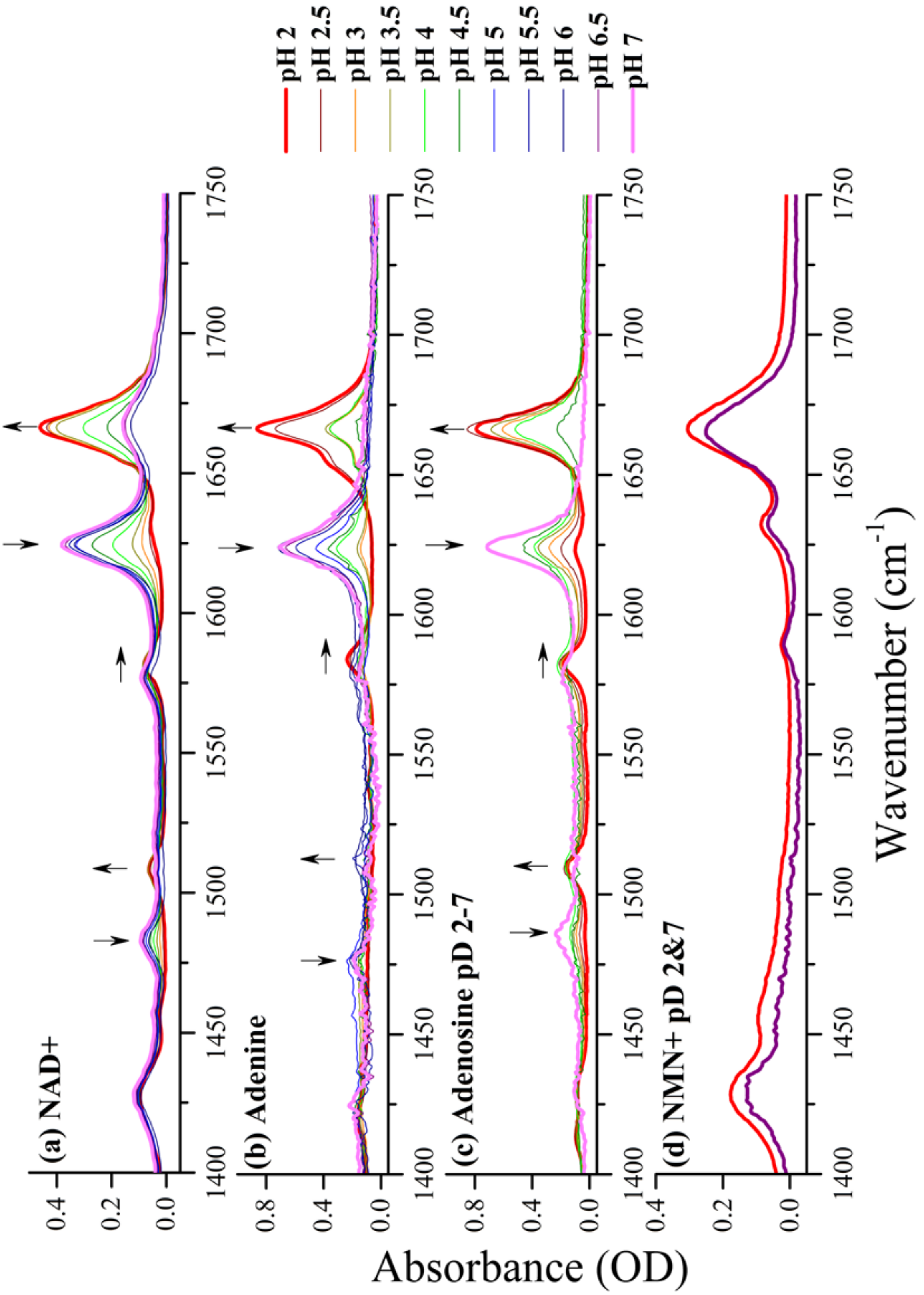

(a) NAD+
(b) Adenine
(c) Adenosine pD 2-7
(d) NMN+ pD 2&7
Absorbance (OD)
0.4
0.2
0.0
0.8
0.4
0.0
0.8
0.4
0.0
0.4
0.2
0.0
Wavenumber (cm^{-1})
1400
1450
1500
1550
1600
1650
1700
1750
pH 2
pH 2.5
pH 3
pH 3.5
pH 4
pH 4.5
pH 5
pH 5.5
pH 6
pH 6.5
pH 7

Figure (4): The pD-resolved FTIR spectra of (a) NAD+ and the structural analogues (b) Adenine, (c) Adenosine and (d) NMN+. Arrows indicate spectral changes that occur with decreasing pD.

Refinement of the initial vibrational assignment was conducted through a pD-dependence study. Figure (4) shows the effect of acidic solution pD on NAD (a) and its analogues (b-d). The chosen pD range in this experiment is advised by the ^{1}H NMR study of NAD+ that demonstrated unfolding through adenine protonation; a consequence of like-charge repulsion between the oxidised pyridine ring and protonate adenine group. A parallel set of NMR measurements was performed using the same samples as used for FTIR, allowing us to verify the consistency of this study with previous publications[9].

The spectrum of NAD+ appears highly sensitive to pD changes within the studied range; all modes are affected, with the exception of those at 1428 and 1455 cm^{-1}. The most prominent effects occur at high frequency; the amplitude of the 1666 cm^{-1} mode increases, while the intensity at 1625 cm^{-1} decreases. Likewise, the 1488 cm^{-1} band is depleted with reduced pD and a new absorption grows in at 1508 cm^{-1}. A small blue-shift is also seen to affect the 1577 cm^{-1} vibration. It is noteworthy that migration of the 1625 cm^{-1} band of NAD+ at low pD now allows the 1633 cm^{-1} absorption from the nicotinamide mononucleotide spectrum to be resolved, at the same frequency as in NMN+.

The pD sensitivity of NAD+’s structural analogues appears confined to the adenine and adenosine samples, with only minor spectral changes occurring for NMN+. At low pD, the 1666 cm^{-1} mode of the NMN+ shifts to a frequency ~2 cm^{-1} higher than seen in neutral solution, and the 1589 cm^{-1} peak is slightly narrowed and reduced in intensity. However, the latter effect is potentially attributable to the effect of D/H exchange rather than the acidity of the solution; previous studies have assigned this 1589 cm^{-1} band to an overlap of two weak modes, one of which is susceptible to D/H exchange (an NH_2 mode), and which is not (a nicotinamide ring mode). In this case the apparent narrowing of the single 1589 cm^{-1} may stem from the migration of the exchangeable NH_2 vibration, leaving only the ring mode at this position. Such a D/H exchange may also provide explanation for the shift of the 1666 cm^{-1} mode. If assignment of the 1666 cm^{-1} vibration is made to the nicotinamide C=O (see DFT section below, Table (1)), one would expect a small dependence between the frequencies of the NH_2 and C=O vibrations of the amide group. It is difficult to observe either of these behaviours in the modes of NAD+, though the relative pD insensitivity in the NMN+ spectrum allows confident assignment of the coenzyme modes to either adenine or nicotinamide.

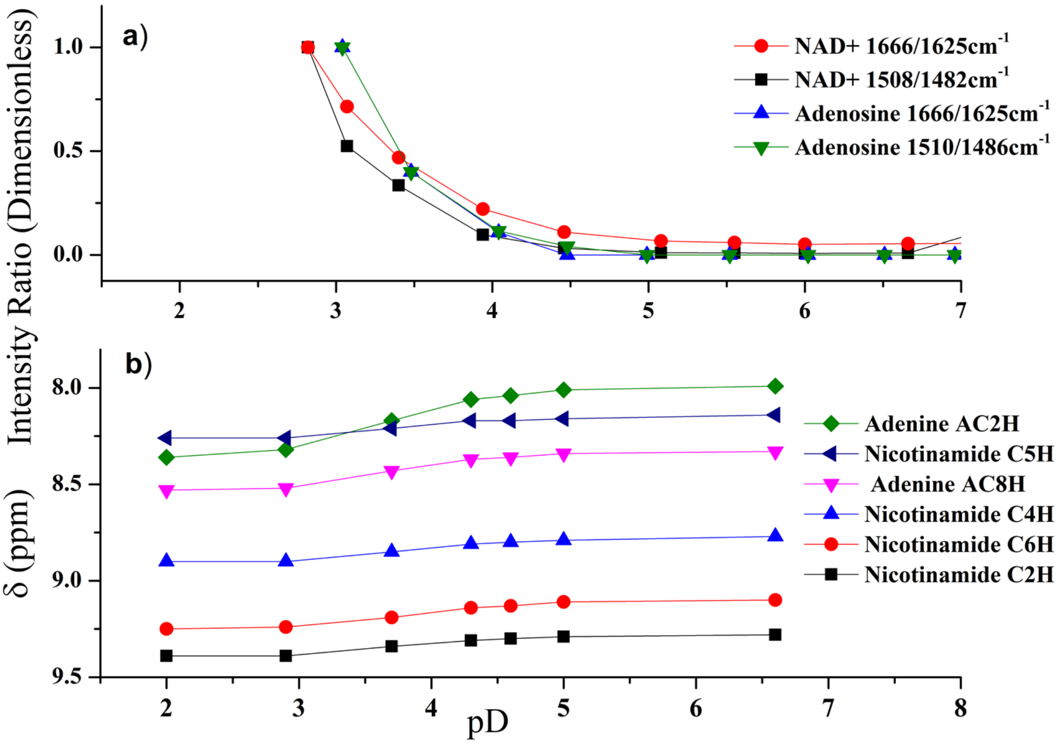


Figure (5): (a) Intensity ratios of the IR peak pairs of NAD+ and Adenosine, as a function of pD. (b) Chemical shift values for the assigned protons of NAD+, as a function of pD.

The pD dependence of adenine encompasses that observed for the full coenzyme; the 1625 and 1488 cm^{-1} peaks are reduced in intensity as solution acidity increases, intensity at 1666 cm^{-1} increases along with that of a nascent peak at 1508 cm^{-1}, and a 5 cm^{-1} blue-shift occurs for the 1577 cm^{-1} mode.

Figure (5) (a) shows the 1666/1625 cm^{-1} and 1508/1488 cm^{-1} peak intensity ratios as a function of pD, for adenosine and NAD+. From this plot, it can be seen that the peak-pairs show correlated pD dependences, and thus likely arise from the same process. The most significant changes in peak ratio occur within the pD range 3-4; reflective of the pKa of adenine’s A1 atom (Figure (6)). Spectral features seen at lower pD, such as the 1508 cm^{-1} band and stronger 1666 cm^{-1} absorption, are therefore attributed to vibrations of a protonated adenine ring, while the 1488cm^{-1} and 1625 cm^{-1} modes are assigned to the analogous modes of a deprotonated form.

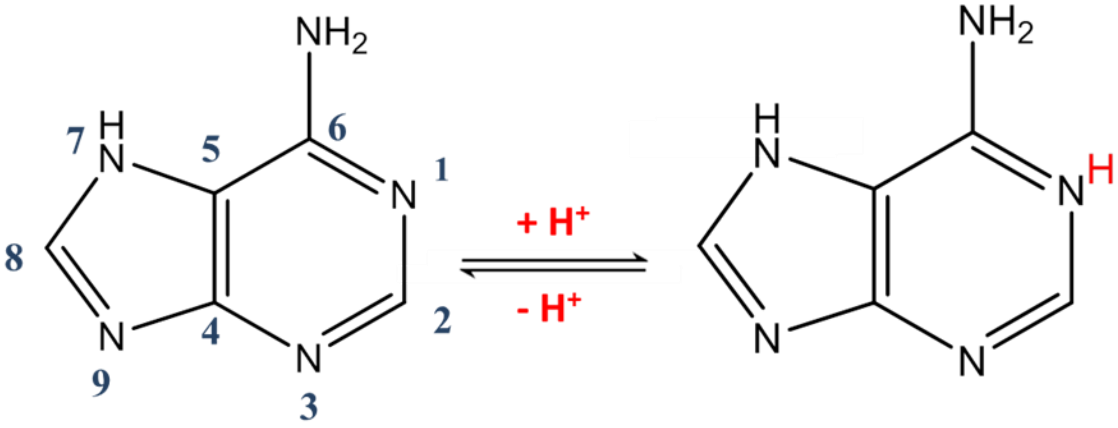


Figure (6): Reaction scheme for the protonation of adenine at the A1 atom.

Figure (5)(b) contains ^{1}H NMR chemical shift values for the nicotinamide and adenine moieties of NAD+, from pD 2-7. These data are consistent with those reported by Catterall[9]. To summarise briefly: all nicotinamide and adenine group protons exhibit lower chemical shift values with increasing pD. The most significant changes to chemical-shift occur for adenine protons, most notably A2, and all observed changes occur with a single pK_a in the pD region 3-4. Adenine’s A1 atom has the only pK_a in this range; a value of ~3.8. Thus the effects of pD on the ^{1}H NMR spectrum were assigned to protonation of the adenine ring.

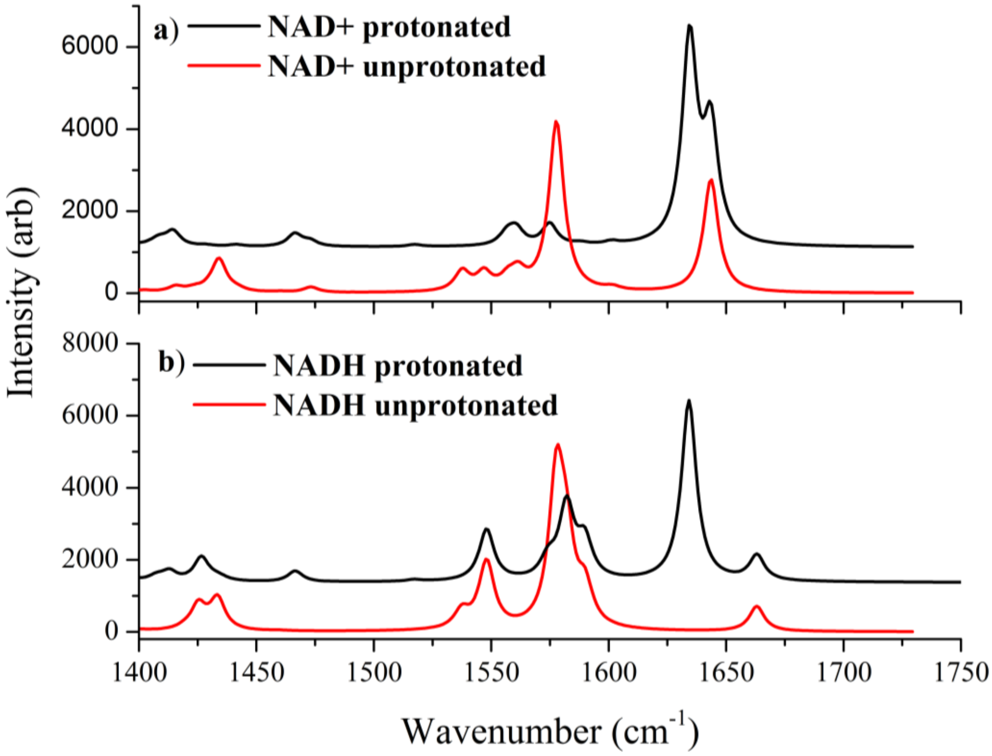


Figure (7): The DFT-calculated IR spectra of (a) NAD+ and (b) NADH, with and without protonation of the Adenine A1 atom. Arrows mark the effect of decreasing pD, i.e. protonation.

DFT calculations were used to predict the vibrational spectra of NAD+ and NADH with and without adenine A1 protonation, allowing definitive band assignment. Figure (7) shows the generated spectra, plotted under their "corrected" wavenumber values[36]. The approximations applied in DFT calculations typically lead to an overestimation of vibrational frequencies. As such, several publications have been made regarding optimum "correction factors" for various functionals, basis sets and their corresponding diffuse and polarisation functions[36,37]. Correction factors are derived from minimisation of the difference between measured and predicted frequencies for a large set of molecules. The factor is therefore an average value, and its true accuracy will vary between molecules.

When applied in this case, the correction factor leads to a slight underestimation of vibrational frequencies, though the relative profile of each absorption spectrum correlates well with their measured equivalent. For example, the protonation state of

NAD+ is well-represented in the calculation; each of the previous three reference points discussed for protonation in the FTIR is reproduced here. Firstly, the strongest absorption in the DFT spectrum, at 1577 cm^{-1}, is displaced to 1634 cm^{-1} as a result of adenine protonation; reflective of the increased 1666/1625 cm^{-1} intensity ratio of the experimental data at pD2, relative to the intensity ratio at pD7. Secondly, the DFT calculation anticipates a the loss of a band at 1434 cm^{-1}, accompanied by the formation of a new band at 1466 cm^{-1}, through protonation. This secondary band displacement follows the pD relationship of the 1488 cm^{-1} and 1508 cm^{-1} experimental peaks. Finally, a set of two vibrations around ~1540 cm^{-1} is eliminated by protonation while a new band is formed at 1574 cm^{-1}. In this third case, it could be argued that a new mode has been created, and not that the same vibration is moved to a higher frequency upon protonation. The displacement of the 1577 cm^{-1} band to 1634 cm^{-1}, however, can be directly assigned to the same vibration in each case.

The predicted 1577 cm^{-1} mode is assigned to the 1625 cm^{-1} peak in the measured pD 7 spectrum, allowing this vibration to be assigned to a localised mode of adenine. Though the structure of adenine is likely produce mostly delocalised ring vibrations, this strong, high frequency feature is seen to originate from the NH_2 group of its 6-membered ring. The band is formed through a stretch of the C-N bond connecting carbon A6 with the NH_2, which in turn generates a contribution from the NH_2 scissor motion, both motions encompassed in the same band. Adjacency of the A6 carbon to the A1 protonation site justifies the magnitude of the blue-shift caused by protonation. In its high frequency state, after protonation, the C-N/NH_2 is predicted to overlap with the band ~1650 cm^{-1}, assignable to the C=O stretch of the nicotinamide moiety. Again, this follows the measured results, though in practice the overlap is significantly closer than predicted, with the C=O band almost entirely obscured at low pD. It is interesting to note that the most intense band in NAD+ does not originate from the C=O of nicotinamide, to which it is widely assigned.

The nicotinamide C=O mode is predicted to contain contributions from the NH_2 group, providing an explanation for the previously described effect of H/D exchange on the band's frequency. This was investigated through a further calculation in which the amide protons were replaced by deuterons, resulting in a small blue-shift of the C=O band (~9 cm^{-1}), in reasonable agreement with the reported data. Conversely, the shift generated through an equivalent substitution at the NH$_2$ group of adenine led to predicted spectrum unlike that measured here, which suggests that H/D exchange of this functional group occurs over significantly longer timescales than the period between sample preparation and measurement (maximum 24 hours), or that exchange simply does not occur.

Six-membered ring vibrations in the adenine moiety give rise to a predicted absorption band at either 1444 cm^{-1} (deprotonated) or 1466 cm^{-1} (protonated), subsequently assigned to the experimental bands at 1482 cm^{-1} (pD7) and 1508 cm^{-1} (pD2), respectively. The vibrational modes before and after protonation are not as closely related as those of the C-N/NH$_2$ mode. Instead, the 1488 cm^{-1} band features a collection of vibrations around the 6-membered ring that are lost at high pD, in favour of a more localised stretch of the core C=C bond between the 5 and 6-membered rings, of frequency 1508 cm^{-1}. As might be expected from the structure, the remaining modes of NAD+ are attributable to delocalised ring modes, the origins of which are categorised in Table (1), forming the full vibrational assignment of NAD+ at neutral and acidic pD.

NAD+ (experiment, cm^{-1})		**Assignment**	**NAD+ (DFT, cm^{-1})**	
pD 7	**pD 2**		**Deprotonated**	**Protonated**
1666 (br)	1666 (br)	Nic C=O str	1655	1642
1633 (w)	1633 (w)	Nic CONH$_2$ NH$_2$ bend	1572	1560
1624 (st)	1666 (st)	Ad C-N/NH$_2$	1588	1634
1577	1582	Ad NH$_2$ + ring	1557	1584
1482	1508	Ad ring	1444	1466
1457	1457	Ad/Nic ring	1393-1484	
1428	1428	Ad/Nic ring	1393-1484	

Table (1): Vibrational band assignments for the DFT and experimental spectra of NAD+, taking into account the effects of protonation.

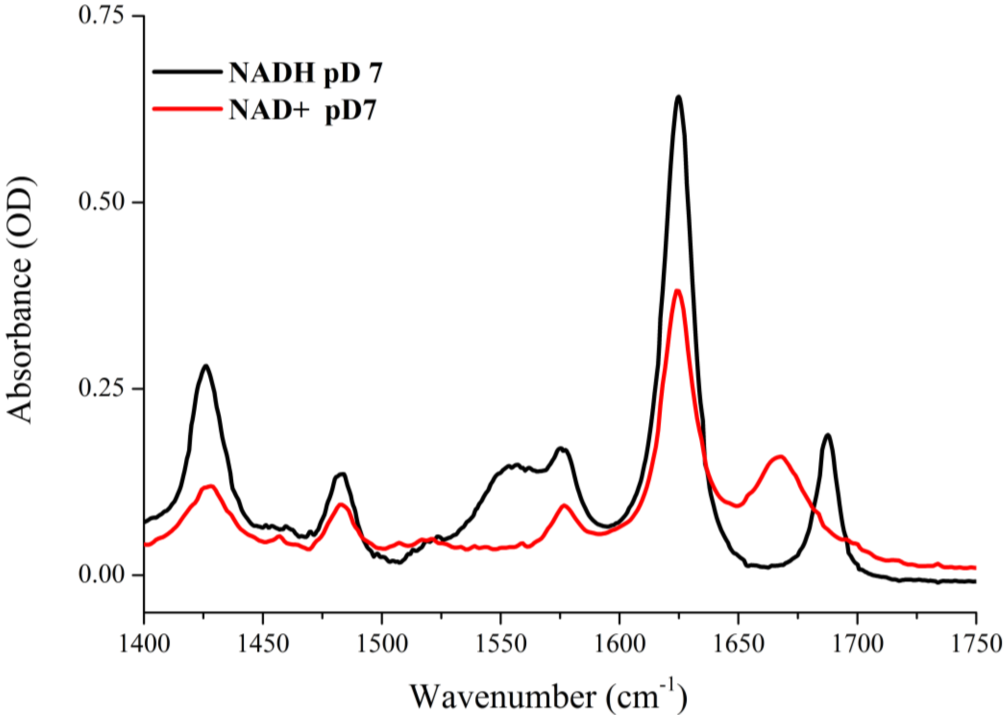


Figure (8): Comparison FTIR spectra of NADH (black) and NAD+ (red) at neutral pD.

4.4.2 Vibrational band assignment of NADH

Though a direct comparison of NAD+ to the reduced form of the coenzyme is instructive to the assignment of bands in NADH, spectroscopic differences brought about through reduction are not trivial (Figure (8)). As previously stated, all vibrations within this frequency range originate from the nicotinamide and adenine groups, therefore this relatively small change in the overall coenzyme structure, localised to the nicotinamide ring, still constitutes a significant change to structural components under measurement.

NADH shares similar absorption features with NAD+ at the frequencies: 1426, 1455, 1482, 1575 and 1624 cm^{-1}. By contrast, no 1666 cm^{-1} C=O band is observed for NADH, and two new peaks are found in the reduced coenzyme, at 1555 cm^{-1} (broad) and 1688 cm^{-1}. It is unclear whether the 1633 cm^{-1} mode remains in the spectrum of NADH. Since the adenine moiety is unchanged between NAD+ and NADH, assignment of these bands is left unchanged. The peak at 1455 cm^{-1}, with shared contributions from each of the nucleobases, appears to be unaffected by reduction of the nicotinamide ring, though this would appear to be the exception amongst nicotinamide modes.

Reduction of the nicotinamide group in NAD+ results in a loss of conjugation around the pyridine ring, creating two localised C=C bonds in NADH. This is referred to as a dihydro-nicotinamide ring, owing to the two hydrogen atoms bonded to atom C4 of the pyridine group. The new peaks, at 1555 cm^{-1} and 1688 cm^{-1} are therefore assumed to correspond with ring vibrations unique to the dihydro species of nicotinamide. Similarly, the stronger 1426 cm^{-1} absorption of NADH, relative to NAD+, is also attributed to conjugation of the pyridine ring.

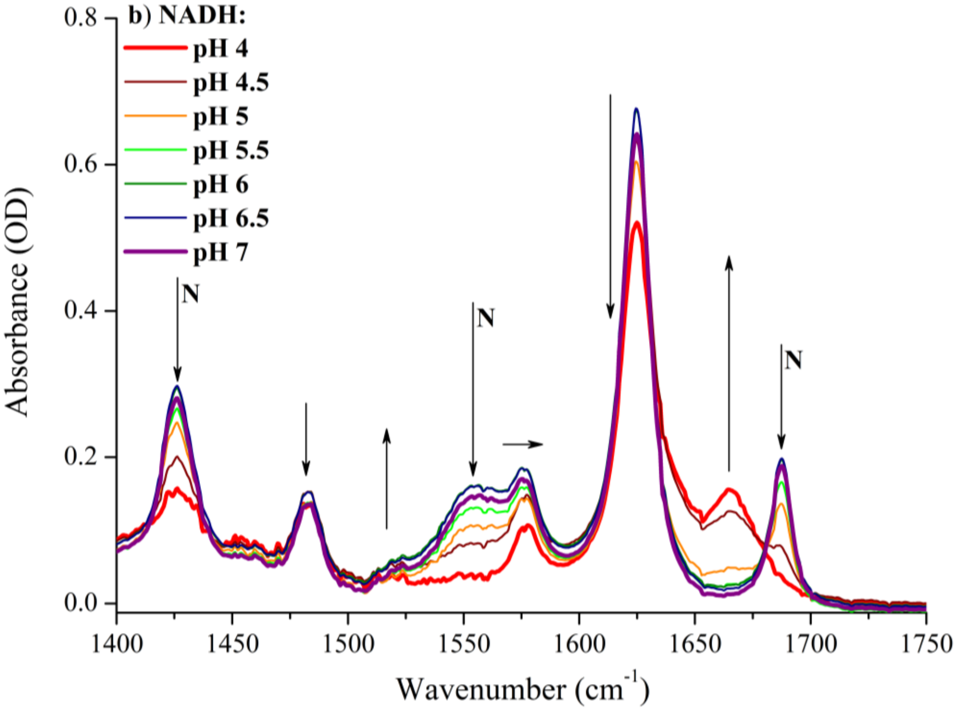
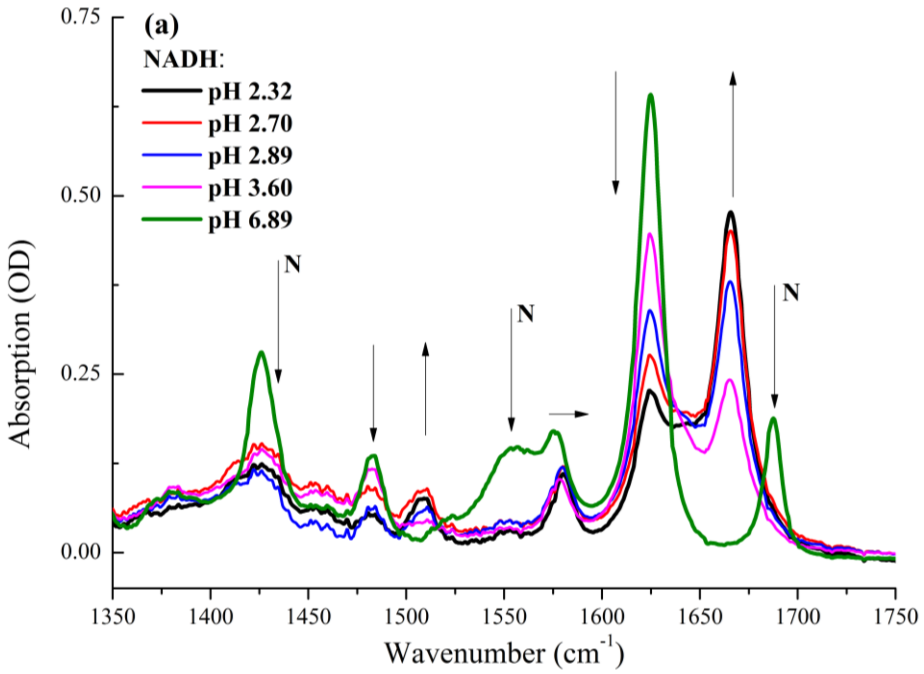


Figure (9): The pD-resolved FTIR spectra of NADH. Arrows indicate the effects of decreasing pD; unmarked arrows correspond to changes assigned to the Adenine moiety, while those marked "N" are correlated with breakdown of the reduced pyridine ring. Colour scale indicates solution acidity, from red (acidic) to purple (neutral).

Support for these assignments is provided through a duplication of the previous pD dependence study, using NADH (Figure (9)).

The reduced coenzyme undergoes substantial IR changes in the pD range 3-4 through adenine protonation: the 1624 and 1482 cm^{-1} peaks are exchanged for new bands at 1666 and 1509 cm^{-1}, respectively, while a small blue-shift affects the weak band at 1575 cm^{-1}. This pD response closely resembles that of NAD+, though this is expected, given the equivalent structure of adenine in either redox state.

A number of additional spectral changes occur through pD reduction, outwith the pk_a of adenine's A1 atom. Each of the NADH modes at 1426, 1555 and 1688 cm^{-1} is seen to decrease in amplitude at low pD; observable in Figure (9) from pD 5.5. These changes are assigned to dihydro-nicotinamide ring destruction, and more specifically, to the acid-catalysed formation of NADH's "primary acid rearrangement product" [38–42]. This degradation compound (a tetrahydropyridine species, shown in Figure (11)) has been categorised *via* NMR spectroscopy, amongst other techniques, and is concluded to form through protonation of the C5 carbon of the pyridine ring, resulting in the destruction of the unsaturated C=C bond between positions C5 and C6 of the nicotinamide ring. ^{1}H NMR measurements performed in parallel to this FTIR study of NADH identified signals at: 7.24, 2.20, 2.05, 1.33 and 1.26 ppm, consistent with formation of the acid rearrangement product. Figure (10) contains the NMR and FTIR spectra of NADH across the full pD range measured.



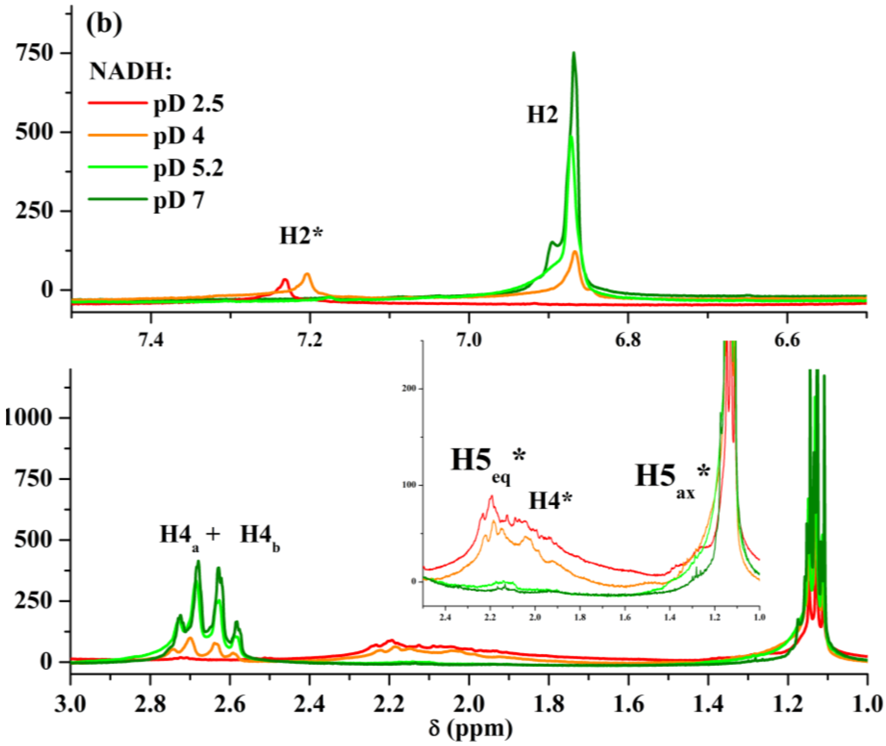


Figure (10): (a) FTIR spectra of NADH across the full pD range investigated. Arrows indicate the effect of decreased pD. Arrows marked with "N" signify changes that take place through breakdown of the reduced pyridine ring. (b) NMR spectra of NADH, complimentary to the FTIR data in (a).

Disappearance of the $H4_a$ and $H4_b$ protons is seen to parallel decreased intensity of the FTIR bands at 1555 and 1688 cm^{-1}. These vibrations are thus assigned to ring modes of nicotinamide that are fundamentally linked to the reduced, dihydro-nicotinamide ring. It is further proposed that both peaks originate from coupling of the two C=C stretching vibrations, either in an anti-symmetric (1555 cm^{-1}) or symmetric (1688 cm^{-1}) fashion. This is inferred from the fact that only a single C=C bond is broken to produce the acid rearrangement compound (Figure (11)), though the intensities at both 1555 and 1688 cm^{-1} are reduced to zero.

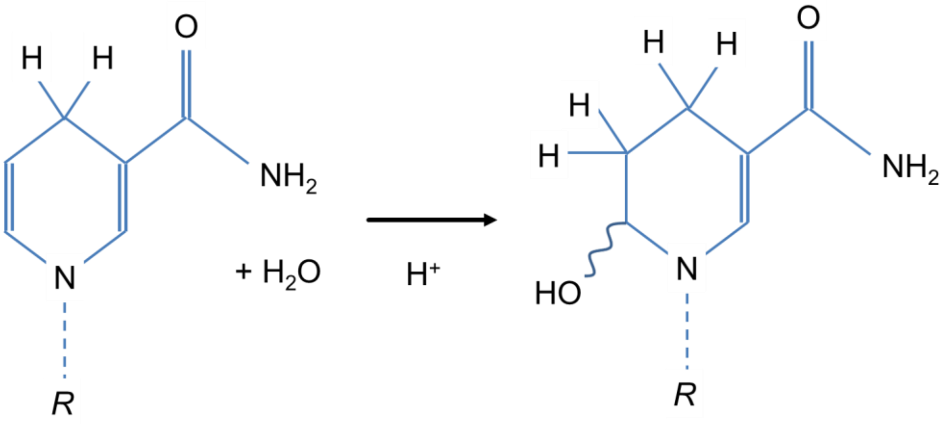


Figure (11): The acid-catalysed hydration of NADH's reduced pyridine ring to form the "primary acid rearrangement compound"[38]

It is not clear whether the band at 1426 cm^{-1} exhibits exactly the same behaviour, due to spectral overlap with an adenine peak. Although the nicotinamide 1426 cm^{-1} is diminished along with the two dihydro-pyridine bonds, it cannot be firmly stated that the band is completely lost by pD 4. However, at neutral pD, this mode is certainly linked to at least one of the nicotinamide C=C bonds.

The 1688 cm^{-1} mode of NADH has previously been assigned to the C=O stretch of nicotinamide[9]. However, direct comparison to the 1666 cm^{-1} band in NAD+, similarly assigned to the carbonyl stretch, shows a substantial variation in linewidth between the two peaks. The 1688 cm^{-1} mode of NADH also exhibits no response to H/D exchange, conflicting with the observed isotope sensitivity reported of the NAD+ and NMN+ C=O stretches. Finally, though hydration of the reduced nicotinamide ring at low pD could realistically have an impact on the vibrational frequency of the C=O, one would expect a response similar to that of adenine, with two separate peaks (1624 and 1666 cm^{-1}), representing deprotonated and protonated states of the molecule. It would seem unlikely that formation of the acid rearrangement product equates to loss of the C=O stretch entirely. The assignment of the NADH 1688 cm^{-1} to a ring vibration provides a far more complete picture: the mode's narrow width suggests an environment that is not strongly solvent-influenced. A significant degree of inhomogeneous broadening would be anticipated from interaction of a highly localised vibration (such as a C=O) with solvent, though the effect would likely be less significant in the case of a delocalised ring vibration. Further, there is little basis to expect a lower degree of C=O solvent-exposure in NADH relative to NMN+ or NAD+, and thus the FWHM of this band should be comparable for each molecule.

It is proposed that reduction of the nicotinamide ring red-shifts the C=O vibration by ~40 cm^{-1}, ultimately placing this band beneath the far stronger adenine C-N/NH_2 absorption. This assignment could be confirmed by protonation of the adenine ring, causing a shift of the C-N/NH_2 to 1666 cm^{-1}, potentially leaving the residual 1624 cm^{-1} exposed. Spectra taken at low pD appear to show identical changes in the adenine bands as seen in NAD+, though it must be acknowledged that the new bands of the NADH breakdown product prevents the same analysis of peak ratios as applied in NAD+.

Although somewhat approximate, it is clear that the amplitude of the 1624 cm^{-1} peak in NADH has greater intensity than the C-N/NH_2 of NAD+ at 1624 cm^{-1}, but has comparable amplitude to the dual band of NAD+ at 1666 cm^{-1}, after adenine protonation. This provides some support for the overlap of the nicotinamide C=O and adenine C-N/NH_2 in NADH, ~1624 cm^{-1}. The slight asymmetry of the NADH 1624 cm^{-1} suggests that the two bands may not occupy identical positions, however. Ideally, the reduced analogue of NMN+ (NMNH) could have been included in this investigation, though the dihydro form of the mononucleotide is not commercially available, and not readily produced from NMN+ in the laboratory. An IR study conducted by Jackson *et al.*[9] includes spectra for NMNH in both D_2O and H_2O. Under deuterated conditions NMNH features a peak at 1627 cm^{-1}, with an equivalent frequency of 1648cm^{-1} in H_2O. This peak is a realistic candidate for the C=O stretching mode. Though slightly weaker than the equivalent vibration in NMN+, and the frequency shift between D_2O and H_2O is significantly larger than predicted from DFT calculations, the width, position and H/D sensitivity of the mode loosely meet the criteria suggested here.

The effects of adenine A1 protonation in NADH were modelled using DFT calculations, and the resulting IR spectrum is shown in Figure (7)(b) alongside the neutral pD coenzyme spectrum. There is close agreement between the experimental and predicted spectra. Changes arising from the protonation of adenine are reproduced identically for NADH as in NAD+. At ~1575 cm^{-1} the deprotonated form of NADH is seen to have a non-concentric overlap of two bands, the lower frequency of which is assigned to the C-N/NH_2 stretch of adenine. The higher frequency band is designated as the C=O vibration of nicotinamide, consistent with the proposed assignment of 1627 cm^{-1}. An NH_2 scissor vibration from the nicotinamide $CONH_2$ is expected to occur in NADH at a comparable to that observed in NAD+ or NMN+ (predicted~ 1572 cm^{-1}, measured 1633 cm^{-1}). This weak band is not resolvable in the NADH experimental spectrum under any measured conditions however, and is therefore left unassigned.

The effects of nicotinamide ring degradation were not well-represented by DFT calculation, though formation of the primary acid rearrangement compound may involve intermediate compounds who structures are not known. Predicted 1547 and 1663 cm^{-1} modes, analogous to the 1555 and 1688 cm^{-1} experimental bands, are seen to originate from the anti-symmetric and symmetric stretches of the C=C bonds of the reduced ring, respectively, which is consistent with the observation that both bands are depleted by destruction of one C=C bond. The complete band assignment for NADH is given in Table (2).

NADH (experiment, cm^{-1})		**Assignment**	**NADH (DFT, cm^{-1})**	
pD 7	**pD 2**		**Deprotonated**	**Protonated**
1688	1688	Nic dihydr ring sym(C=C)	1674	1674
1624 (br)	1624 (br)	Nic C=O str	1600	1601
1624 (st)	1666 (st)	Ad C-N/NH_2	1589	1645
1575	1580	Ad NH_2 + ring	1556	1585
1555 (br)	1555 (br)	Nic dihydr ring antisym (C=C)	1558	1558
1482	1509	Ad ring	1443	1476
1455	1455	Ad/Nic ring	1401-1435	
1426	1426	Ad/Nic ring	1401-1435	

Table (2): Vibrational band assignments for the DFT and experimental spectra of NADH, taking into account the effects of protonation

4.4.3 IR detection of NAD conformations

The solution-phase conformation of NAD is presently not well-understood. Attempts were made in the course of this study to contribute to the existing literature regarding the molecular fold of NAD, and a complete summary of these investigations is provided in Appendix (A.4.2). In the consideration of NAD as a probe of protein structure however, the correlation of coenzyme conformation and IR absorption is primarily of interest due to the unfolded-to-folded transition that is believed to occur upon protein binding. If the free NAD molecule undergoes significant spectral changes as a result of this conformation change, it is necessary to deconvolute this from other spectral changes that may be specific to protein binding.

^{1}H NMR studies[9] have previously demonstrated the folding of both NAD+ and NADH in nonpolar solvents such as methanol and urea, with significant consequences for the chemical shifts of both nicotinamide and adenine protons. An FTIR analogue of these experiments was performed here. In summary, both the oxidised and reduced dinucleotides were solvated in water: methanol mixtures of varying ratio (up to 70%, or 17 molar, methanol), though the resulting solvatochromatic effect on the widths and positions of each assigned IR band were deemed to be of greater magnitude than spectral changes originating from a conformation change. The effect of increasing methanol concentration on IR absorption were minor, as shown in Figure (12), suggesting that folding has a negligible effect on the vibrational spectra of NAD.

This conclusion is also supported by the results of the NAD+ pD study; protonation of the adenine group in the oxidised coenzyme is believed to lead to a like-charge repulsion of the two positively charged head groups, resulting in an extended conformation. Through NMR, significant changes can be detected in the environment of the nicotinamide group through adenine protonation, whereas the equivalent FTIR study sees little change in the nicotinamide vibrations of NAD+. Similarly, the adenine group of the coenzyme is only sensitive to protonation, not unfolding; all of NAD+’s spectral changes are reflected in the free molecules, adenine and adenosine, neither of which has the capacity to fold or unfold.

One consequence of this result is that the potential limited for IR spectroscopy to determine the conformation in which NAD binds to a protein, in the case that no crystallographic information is available. Conversely, this ensures that spectral differences between free and protein-bound forms of NAD are likely to reflect local interactions made by the coenzyme.

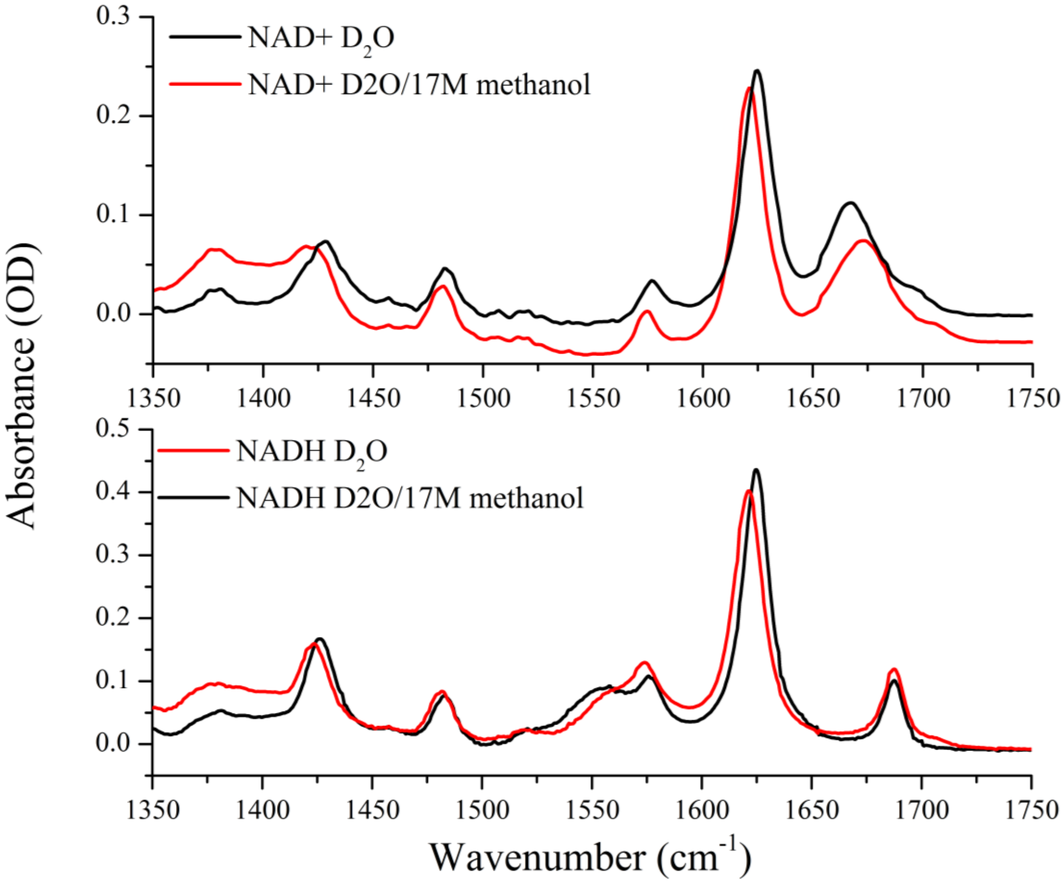


Figure (12): The spectra of NAD+ and NADH at neutral pD in water, and in a 30:70 solution of water: methanol D-4.

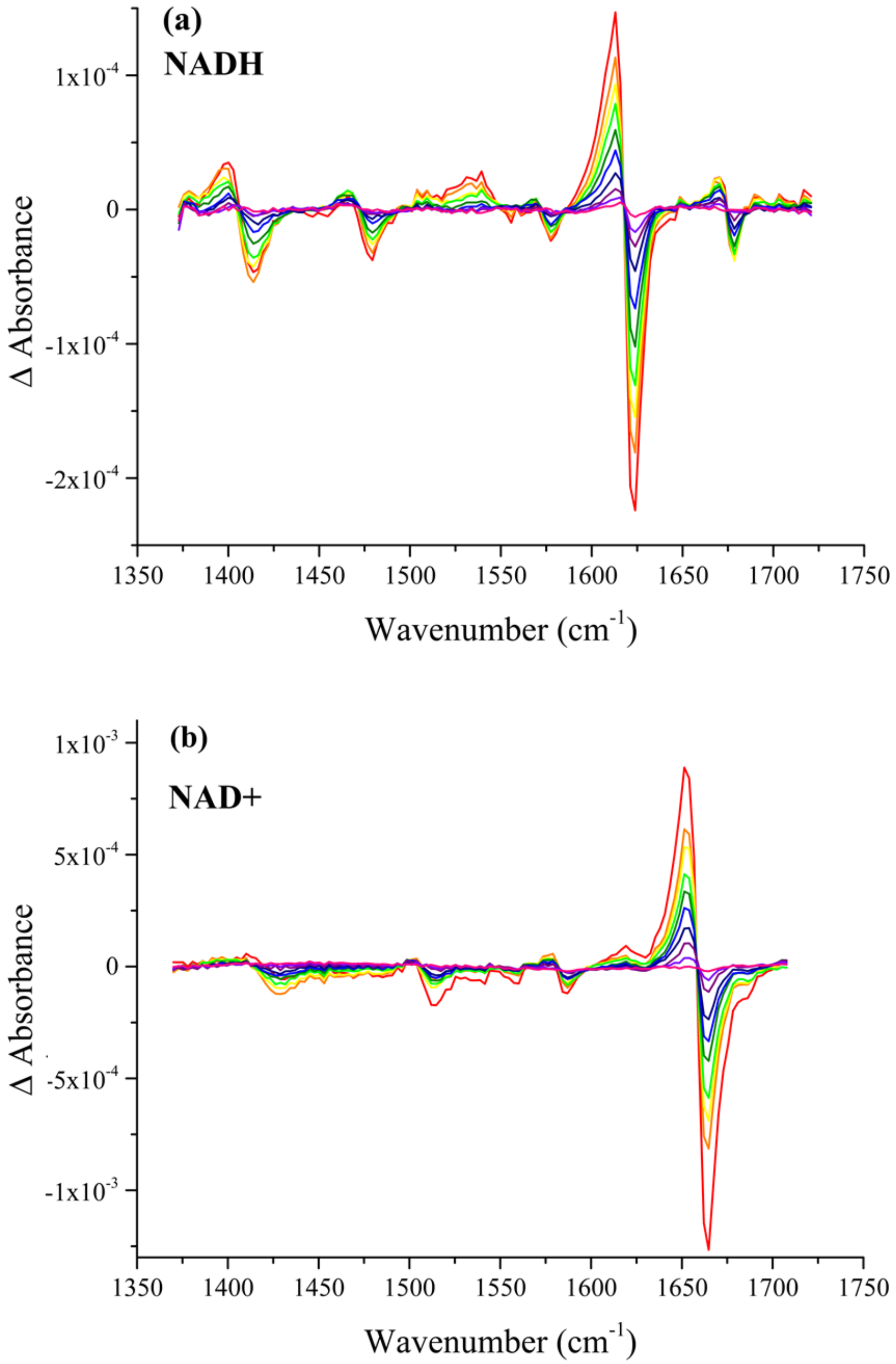


Figure (13): The time-resolved infrared (TRIR) spectra of (a) NADH (pD 7) and (b) NAD+ (pD 2.5.). The coloured traces depict the transient absorption of each molecule as a function of pump-probe delay, with red indicating short delay times (~250fs) and pink indicating long delay times (>50ps).

4.4.4 Pump-Probe Spectroscopy

Figure (13) contains the time-resolved infrared spectra of NAD+ (pD 2) and NADH (pD 7). By considering the temporal dependence of each vibration (the decay of the signal perpendicular to the plane of the graph), the vibrational lifetimes of each mode can be extracted.

The pump-probe signals for each IR band were optimally fit fit using exponential functions to obtain corresponding vibrational lifetimes. It was observed in several cases that a mono-exponential function was insufficient to describe the signal at a particular frequency, though the addition of a second exponential component vastly improved the quality of fitting. This occurred for the 1426 and 1624 cm^{-1} bands of NADH, and for the 1428 and 1666 cm^{-1} bands of NAD+, which is consistent with the proposal that each of these IR bands contains contributions from two species of vibration. Additional support for this is provided by the relative amplitudes of each exponential component in the relaxation of overlapped vibrational bands. More complex, non-exponential relaxation was not observed on the timescales measured. Tables (3) and (4) summarises the parameters derived from the pump-probe data.

Table (3): Parameters derived from exponential fitting of NADH pump-probe data

NADH (pD 7) vibration	Exponential time constant, τ_1 (ps)	Exponential amplitude, A_1 (Δ OD)	Exponential time constant, τ_2 (ps)	Exponential amplitude, A_1 (Δ OD)	**Relative amplitude of components (A_1/A_2)**
Nic. sym. dihydro C=C stretch (1688 cm^{-1})	2.6 +/- 0.2	3.6×10^{-5} +/- 1×10^{-6}			
Ad C-N/NH_2 + Nic C=O str (1625 cm^{-1})	2.3 +/- 0.1	1.29×10^{-4} +/- 5×10^{-6}	0.4 +/- 0.1	0.99×10^{-4} +/- 5×10^{-6}	1.3
Ad NH_2 + ring (1575 cm^{-1})	2.6 +/- 0.3	1.7×10^{-5} +/- 5×10^{-7}			
Ad ring (1482 cm^{-1})	2.3 +/- 0.2	2.3×10^{-5} +/- 8×10^{-7}			
Ad + Nic ring (1426 cm^{-1})	2.8 +/- 1.4	2.3×10^{-5} +/- 0.9×10^{-5}	0.7 +/- 0.3	3.7×10^{-5} +/- 0.9×10^{-5}	0.6

Table (4): Parameters derived from exponential fitting of NAD+ pump-probe data

NAD+ (pD 2) vibration	Exponential time constant, τ_1 (ps)	Exponential amplitude, A_1 (Δ OD)	Exponential time constant, τ_2 (ps)	Exponential amplitude, A_1 (Δ OD)	**Relative amplitude of components (A_1/A_2)**
Ad C-N/NH_2 + Nic C=O str (1666 cm^{-1})	1.3 +/- 0.1	8.80×10^{-4} +/- 1.5×10^{-5}	0.1 +/- 0.1	2.84×10^{-4} +/- 2.0×10^{-5}	3.1
Ad NH_2 + ring (1582 cm^{-1})	2.2 +/- 0.5	7.2×10^{-5} +/- 0.5×10^{-5}			
Ad ring(1508 cm^{-1})	2.4 +/- 1.3	7.6×10^{-5} +/- 0.72×10^{-5}			
Ad + Nic ring (1428 cm^{-1})	2.2 +/- 0.9	8.5×10^{-5} +/- 4.5×10^{-5}	0.6+/- 0.3	9.4×10^{-5} +/- 3.6×10^{-5}	0.9

The lifetimes of all measured vibrations fall within the range ~1-2.5 ps, though the weak signals of several modes give rise to relatively high uncertainties. Broadly, vibrational relaxation in NAD is seen to occur over a comparable, if not slightly longer, timescale to the Amide I and II vibrations, whose lifetimes are typically cited as ~1 ps[43–46]. Although rapid dynamics such as solvent fluctuation[47,48], hydrogen bond reconfiguration[49] and very fast conformational switches[50] occur around this timescale, the lifetime of the probe vibration is ideally longer than the timescale of the dynamics under investigation. In the case of NAD, the rapid decay of signal intensity would likely lead to poor sampling of even these ~1ps processes.

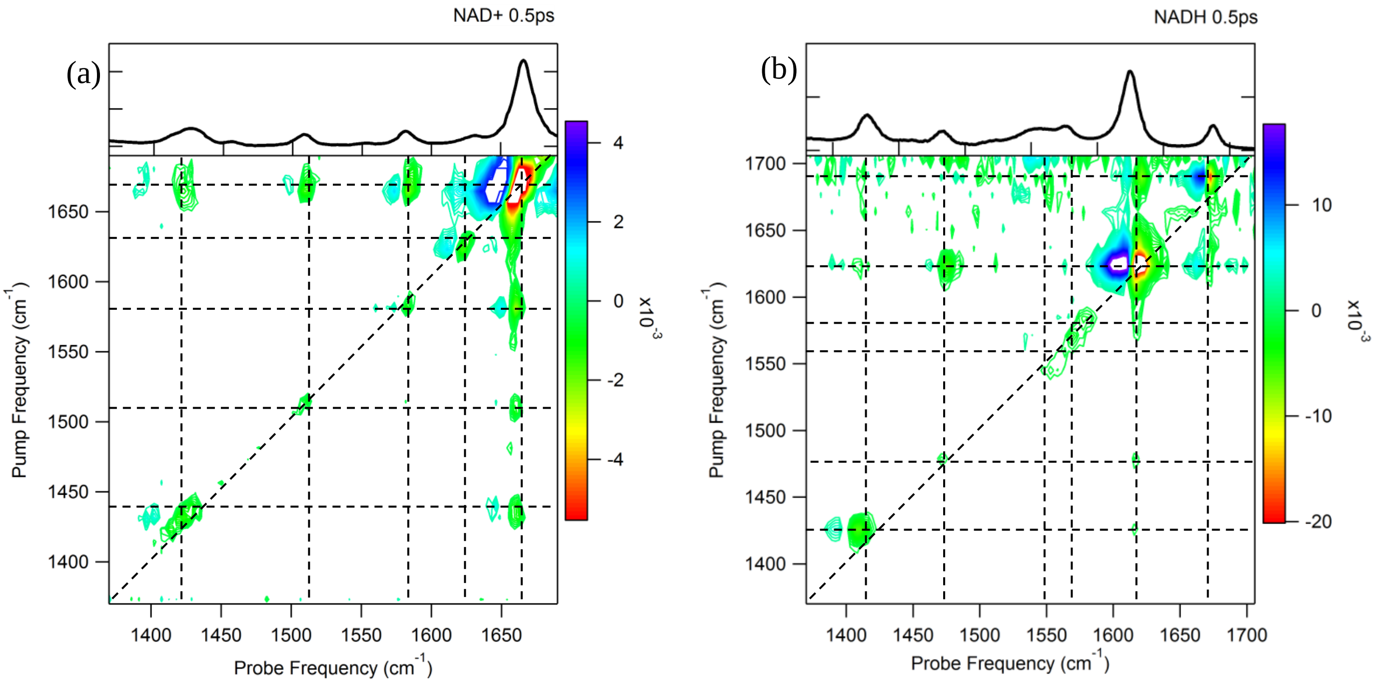


Figure (14): The 2DIR spectra of (a) NAD+ (pD 2.5) and (b) NADH (pD 7) obtained at waiting times of 0.5 ps. Dashed lines indicate the positions of peaks categorised previously. Intersections show potential locations for cross-peaks. The first (lowest) contour line represents +/-10 % of the maximum signal.

4.4.5 2DIR Spectroscopy

Representative 2D spectra of NAD+ and NADH are shown at a waiting time of 0.5 ps (Figure (14)). The NADH spectrum was obtained in neutral solution (pD 7) while the spectrum for NAD+ included here corresponds to the protonated form (pD 2.5). Inset above each 2D spectrum is the corresponding linear absorption profile of each molecule. Dashed lines are added to the spectra at the central frequencies of all assigned bands from the linear spectrum, along both the probe and pump frequency axes. These are included to emphasise potential coupling patterns between vibrational modes.

Vibrational coupling is manifest through sets of off-diagonal peaks. These "cross-peaks" occur at specific intersections of the pump and probe axes, the coordinate of which provides detail on the nature of the coupling[51]. For example, in NADH cross-peaks are observed at of (ω_{probe} =1482 cm^{-1}, ω_{pump} = 1624 cm^{-1}) and at the reciprocal coordinate of (1624 cm^{-1}, 1482 cm^{-1}). This demonstrates the coupling of the 1482 and 1624 cm^{-1} vibrations, whereby additional excitation pathways emerge from new Feynman pathways that can be made between the vibrational levels of these oscillators.

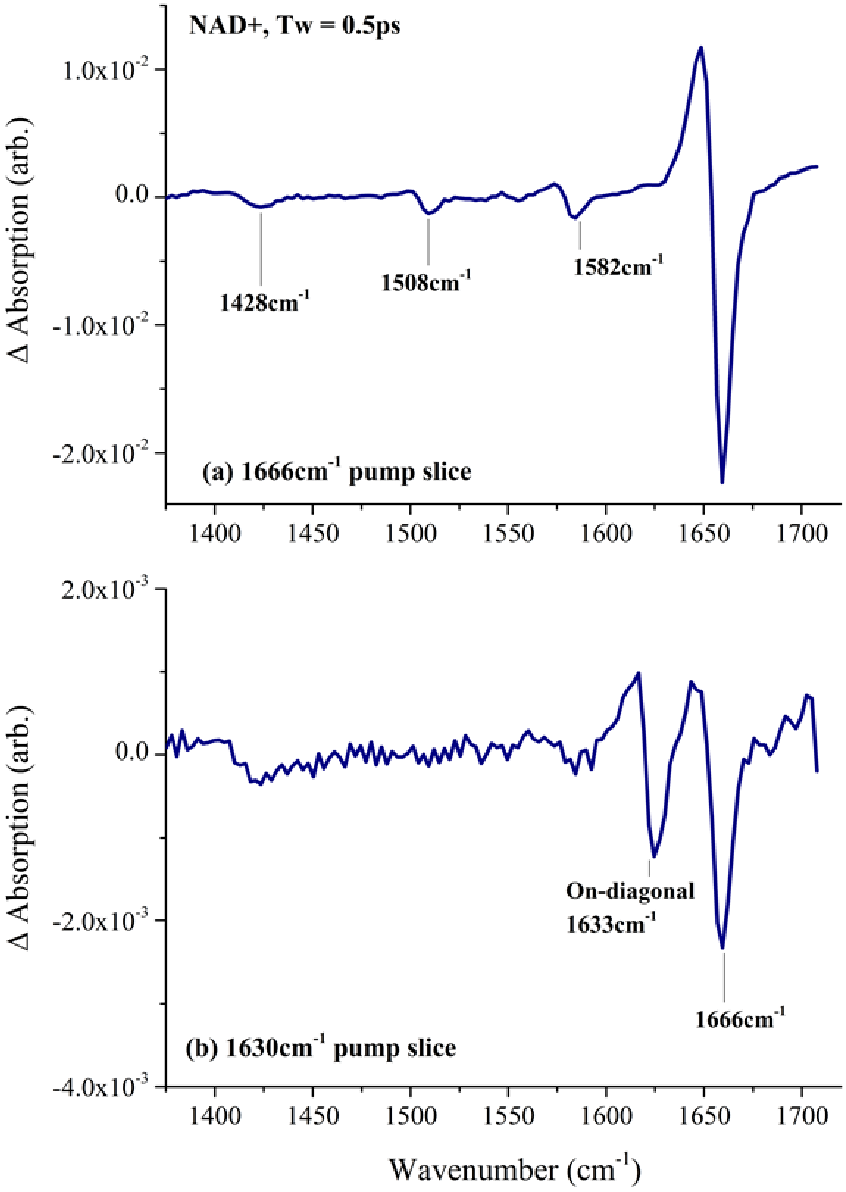


Figure (15): Cross-sections through the spectrum of NAD+ at pump frequencies of (a) 1666 cm^{-1} and (b) 1630 cm^{-1}

At a waiting time of 0.5 ps, NAD+ displays four sets of cross peaks that involve the 1666 cm^{-1} band, i.e. the on-diagonal peak at (1666 cm^{-1}, 1666 cm^{-1}), as shown in Figure (15)(a). Across this pump frequency, clear transitions can be seen at 1428, 1508, 1582 and 1633 cm^{-1}. The 1508 and 1582 cm^{-1} modes are unique to the adenine head group, while 1633 cm^{-1} originates from nicotinamide alone. The fourth cross-peak, at 1428 cm^{-1} contains contributions from both nucleobases. When considering the protonation state of adenine, the number of cross-peaks that appear at this pump frequency is not surprising; the 1666 cm^{-1} peak of NAD+ in this protonation state is composed of two bands: the nicotinamide C=O and the adenine C-N/NH$_2$. The off-diagonal signals observed from 1666 cm^{-1} are therefore attributable to a convolution of the couplings formed by each mode.

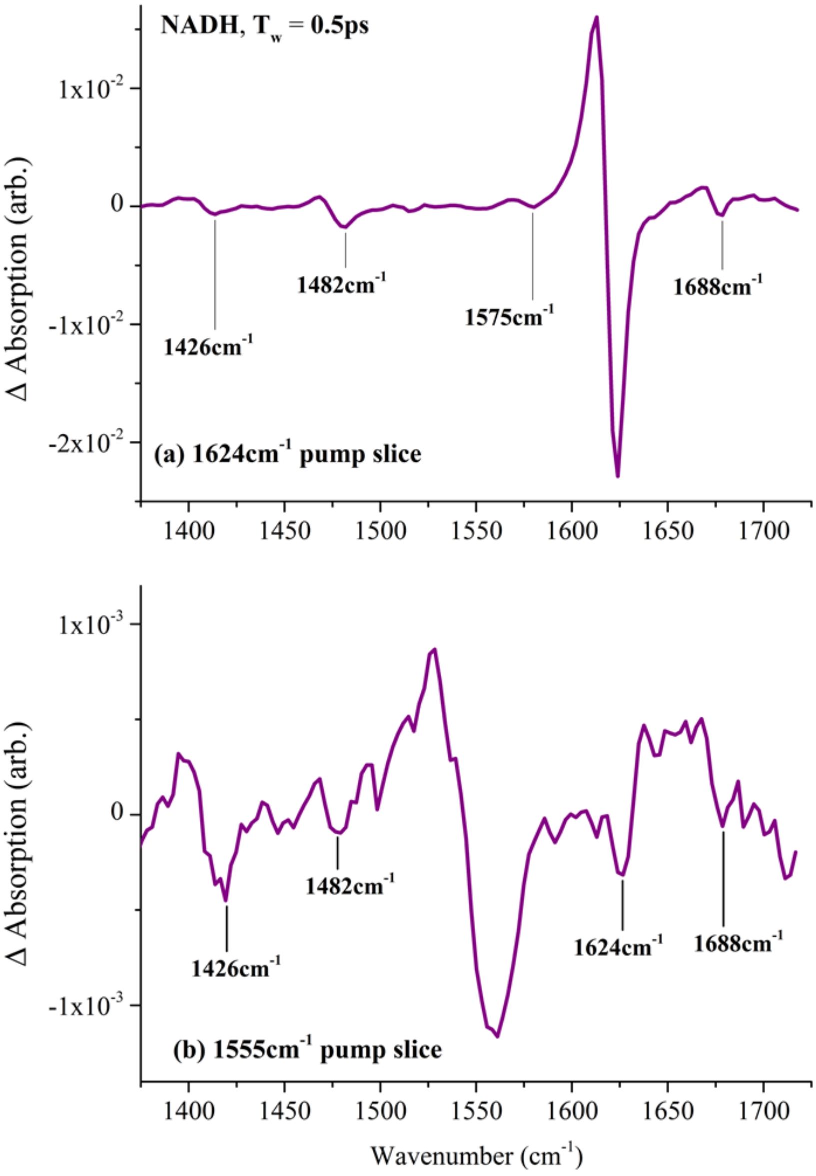


Figure (16):Cross-sections through the spectrum of NADH at pump frequencies of (a) 1624 cm^{-1} and (b) 1555 cm^{-1}. Marked peak-pairs identify the bleach transitions for off-diagonal features.

In NADH, extensive vibrational coupling is observed. The 1624 cm^{-1} band here also contains contributions from the nicotinamide C=O and adenine C-N/NH_2, and cross-peaks are resultantly observed at: 1426, 1482, 1575 and 1688 cm^{-1}, corresponding to vibrations of: nicotinamide/adenine, adenine, adenine and nicotinamide (Figure (16)(a)). Interestingly, from the pump slice corresponding to the asymmetric dihydro stretch (1555 cm^{-1}), cross-peaks can be observed to the adenine-exclusive band at 1482 cm^{-1} (Figure (16)(b)). It is possible that this coupling is detected as a result of overlap between the high frequency shoulder of the 1555 cm^{-1} peak and the adenine NH_2 (1575 cm^{-1}), however, this raises a vital issue regarding the spectroscopy of each coenzyme; it is possible to justify all observed adenine-nicotinamide coupling patterns as the result of adenine-adenine or nicotinamide-nicotinamide couplings that happen to coincide. However, it cannot be stated categorically that inter-nucleobase coupling does not contribute in any way to off-diagonal signals. The independent biological significance of the coenzyme fold notwithstanding, it becomes challenging to assign cross-peaks of both NAD+ and NADH without exact knowledge of their solution-phase conformations. Ultimately, it is difficult to establish a reference for intra-nucleobase vibrational couplings, from which to further isolate inter-base effects. To illustrate;

(1) NAD+, low pD; an “extended” molecular conformation is believed to exist, and thus the expectation of inter-base vibrational coupling is minimal. However, the adenine moiety is protonated, and so two IR bands are found at 1666 cm^{-1}. This is crucial, as cross-peak intensities are typically described as ratio of their corresponding on-diagonal peak intensities, which are convoluted in this case. Spectral overlap invariably occurs at 1428 cm^{-1}, and further, the excited state absorption band of a 1666 cm^{-1} mode overlaps with the nicotinamide NH_2 (1633 cm^{-1}). Many desirable reference points for intra-base coupling cannot be described quantitatively as a result of band overlap. It is fundamental to understand coupling

patterns of protonated NAD+, as this represents the only case where an extended coenzyme conformation is consistently reported.

(2) NAD+, neutral pD; the coenzyme is believed to exist in a dynamic equilibrium between folded and unfolded forms under these conditions, and thus interaction of the two head groups is plausible[18,30,31,32]. Though the deprotonated adenine C-N/NH_2 no longer obscures the nicotinamide C=O, it instead overlaps with the NH_2 bend of nicotinamide (1633 cm^{-1}). As the nicotinamide carbonyl and amide vibrations are coupled (as is evident from protonated NAD+ spectrum, Figure (15)(b)), this is likely to give rise to the appearance of adenine C-N/NH_2 interaction with the nicotinamide C=O. Genuine inter-base coupling between these bands, as a result of the molecular fold, is thus ambiguous.

(3) NADH, neutral pD; the coenzyme is potentially folded, from which inter-base vibrational coupling may occur. It is resultantly difficult to interpret the detected cross-peaks, as these may originate from the superposition of independent adenine and nicotinamide intra-base couplings, or from inter-base contact. It is likely that a significant degree of vibrational coupling occurs in the reduced nicotinamide moiety alone, such as that of symmetric and anti-symmetric dihydro ring modes, or that of the carbonyl and amide. There is also significant scope for interaction between nucleobase modes as, under these conditions, the adenine and nicotinamide rings may stack in a parallel, closely-packed fashion. An inter-ring separation of 3.9 Å, has been reported by Sarma *et al.*[52], while more recent MD studies by Tanner *et al*.[5] have predicted a separation of 5.2 Å. In either case, the range is sufficiently short to allow dipole-dipole interactions between the near-resonant C-N/NH_2 and nicotinamide C=O bands. Further opportunities for inter-base coupling occur from the near-resonances of NADH dihydro bands at 1426 and 1555 cm^{-1} with ring modes and the NH_2 of adenine, respectively. This gives rise to a potentially complex set of vibrational interactions in the reduced coenzyme.

Cross-peaks originating from coenzyme-protein interactions are thus likely to become convoluted in the underlying intra-molecular vibrational coupling of NAD. Quantitative analysis of these signals may resultantly prove challenging.

When applying the experimental considerations of a real 2DIR protein study however, intramolecular coupling in NAD may be somewhat advantageous from an alternate perspective. Assuming efficient, stoichiometric binding between a protein and the dinucleotide, the experimental concentration of NAD is limited by that of the protein, and is thus unlikely to exceed 10 mM (owing to solubility and stability constraints), *i.e.* one tenth of the concentration employed in this study. At 10 mM, the absorption bands of even a small protein are certain to overwhelm the NAD spectrum. However, owing to the comprehensive intramolecular couplings of the dinucleotide, it would become possible to identify the central frequencies of all NAD bands in the off-diagonal spectral region. This may prove useful for identifying changes in the linear absorption spectrum that occur through hydrogen bonding to the protein; changes that it would not be otherwise possible to observe. Further, the ability of a protein to perturb the existing intramolecular couplings of NAD may be capable of providing interesting qualitative information.

The viability of NAD as a probe of spectral diffusion is now considered. With the exception of the overlapped C-N/NH_2 and C=O bands, the vibrations are notably weak, and in the former case, vibrational relaxation is rapid (~0.6 ps). The dynamics observable through analysis of the C-N/NH_2 are therefore limited. The nicotinamide C=O has a significantly longer lifetime, and under specific conditions (NAD+, pD 7) it can be isolated from other dinucleotide absorption bands. However, the majority of protein and polypeptide species are likely to absorb light in the 1666 cm^{-1} region. Again, at an equimolar mixture of coenzyme-protein, it is unlikely that the nicotinamide C=O would be well-resolved unless protein binding gave rise to a significant shift in the carbonyl frequency through hydrogen bonding. This strategy would need to be evaluated on a case-specific basis, however.

The mode with the greatest potential for analysis is the 1688 cm^{-1} ring mode of NADH. There is reasonable separation of this symmetric C=C vibration from the majority of protein spectral features, which partly compensates for the moderate strength of the transition. Additionally, the narrow width of the band leads to a lower chance of overlap with higher frequency protein features, such as side chain C=O modes. Conversely, the study conducted by Jackson *et al.*[66], in addition to this data, shows the dihydro vibration to be invariant between H_2O, D_2O, methanol and DMSO. Sensitivity of this band to its chemical environment therefore appears to be low, which would limit spectral diffusion analysis. The antisymmetric, 1555 cm^{-1} ring mode appears to be more responsive to changes in environment (see Appendix (A.4.2)), though the high frequency shoulder of this band overlaps with the NH_2 mode of adenine (1575 cm^{-1}), and the ring mode is perhaps too broad to be discernible from complex protein absorptions.

Of the remaining NAD vibrations assigned in this study, none are particularly well-placed to act as probes of spectral diffusion. The adenine modes at 1455, 1482 and 1575 cm^{-1} are all strongly-linked, if not entirely attributable, to delocalised ring vibrations that are prone to the same lack of sensitivity as the NADH 1688 cm^{-1} mode. Additionally, these modes are likely to coincide with the Amide II band, in either its deuterated (~1460 -1490 cm^{-1}) or native (~1550 cm^{-1}) forms[53]. The adenine mode at 1426 cm^{-1} directly overlaps with a nicotinamide ring vibration, hence the usefulness of either mode is limited. Finally, the 1633 cm^{-1} of nicotinamide is only resolved in NAD+ at low solution pD, constraining the conditions under which it can be used. Additionally, these remaining modes are not strongly absorbing, and at the concentrations they would likely have to be used in (comparable to that of the protein under investigation) they would become a challenge to resolve adequately.

It is possible that coenzyme binding may cause detectable rearrangement of the target protein's Amide I band. From this, it would be possible to extract site-specific information from detailed Amide I band analysis, particularly where crystallography data is available. Though this approach is somewhat indirect, it could potentially give rise to significant changes in the protein absorption spectrum that are more readily observed than the coenzyme’s spectral features. Ideally, information could be extracted from both molecules, though some degree of structural modification may be required to fully harness the potential of NAD as a probe molecule.

4.5 Conclusions

A full assignment of NAD’s vibrational spectra has been constructed through a combination of FTIR, NMR and DFT calculations, summarising the effects of nicotinamide ring reduction and adenine protonation. Through pD dependent FTIR measurements, evidence was provided for the reassignment of bands in the reduced nicotinamide adenine dinucleotide spectrum. The inclusion of complementary DFT studies provided additional support to these assignments.

The potential for NAD as 2D-IR probe was examined. Ultimately, an ideal probe band is spectrally well-separated from other IR peaks, originates from a localised vibration, and has a long vibrational lifetime relative to the dynamics under investigation. No single band of the coenzyme spectra meets all of these criteria. The overlap of nicotinamide and adenine signatures was seen to create highly complex spectral features that may interfere with data extraction from a protein. Many vibrations were predicted to originate from delocalised ring modes, making them insensitive to their local environments. Further, the lifetimes of all vibrations were seen to be short. It was proposed that both NAD+ and NADH each possess one band that could be utilised as a short-lifetime probe, though only under highly specific conditions.

4.6 Future Work

The potential for NAD to function as a probe could be greatly enhanced through the use of synthesised analogues. Substitution of the nicotinamide C=O for an alternate functional group, such as an azide, would resolve the issue of band overlap with the adenine C-N/NH_2 stretch. Further, azide groups have proven to be excellent reporters of rapid solvent molecule dynamics in other 2DIR studies. Synthesis of an azido-NAD+ compound has indeed been performed[54], and recent IR categorisation of the modified coenzyme has shown that linear characteristics of the azide peak are highly sensitive to the local environments of various protein binding sites[55]. In the cases examined, azido-NAD+ is seen to maintain a comparable binding affinity to native NAD+, though critically, this would need to examined on a case by case basis, as azides groups have the potential to undergo "click" reactions that may denature the protein under investigation[56]. However, the relatively minor structural substitution made to produce azido-NAD+ exemplifies how the coenzyme's potential as a dynamic reporter can be fully exploited.

4.7 References

1. Belenky, P., Bogan, K. L. & Brenner, C. NAD+ metabolism in health and disease. *Trends Biochem. Sci.* **32,** 12–19 (2007).
2. Lin, S. J. & Guarente, L. Nicotinamide adenine dinucleotide, a metabolic regulator of transcription, longevity and disease. *Curr. Opin. Cell Biol.* **15,** 241–246 (2003).
3. Foster, J. W. & Moat, A. G. Nicotinamide adenine dinucleotide biosynthesis and pyridine nucleotide cycle metabolism in microbial systems. *Microbiol. Rev.* **44,** 83–105 (1980).
4. Smith, P. E. & Tanner, J. J. Molecular dynamics simulations of NAD + in solution. *J Am Chem Soc* **121,** 8637–8644 (1999).
5. Smith, P. E. & Tanner, J. J. Conformations of nicotinamide adenine dinucleotide (NAD(+)) in various environments. *J. Mol. Recognit.* **13,** 27–34 (2000).
6. Lin, H. Nicotinamide adenine dinucleotide: beyond a redox coenzyme. *Org. Biomol. Chem.* **5,** 2541–2554 (2007).
7. Yue, K. T. *et al.* Raman spectroscopy of oxidized and reduced nicotinamide adenine. *Biochemistry (Mosc.)* **25,** 4941–4947 (1986).
8. Pollak, N., Dölle, C. & Ziegler, M. The power to reduce: pyridine nucleotides–small molecules with a multitude of functions. *Biochem. J.* **402,** 205–218 (2007).
9. Iwaki, M., Cotton, N. P. J., Quirk, P. G., Rich, P. R. & Jackson, J. B. Molecular recognition between protein and nicotinamide dinucleotide in intact , proton-

translocating transhydrogenase studied by ATR-FTIR spectroscopy. *J. Am. Chem. Soc.* **126,** 517–523 (2006).

10. Berríos-Rivera, S. Metabolic engineering of escherichia coli: increase of NADH availability by overexpressing an NAD+-dependent formate dehydrogenase. *Metab. Eng.* **4,** 217–229 (2002).
11. Rossmann, M. G., Moras, D. & Olsen, K. W. Chemical and biological evolution of a nucleotide-binding protein. *Nature* **250,** 194–199 (1974).
12. Jauch, R., Humm, A., Huber, R. & Wahl, M. C. Structures of escherichia coli NAD synthetase with substrates and products reveal mechanistic rearrangements. *J. Biol. Chem.* **280,** 15131–15140 (2005).
13. Bottoms, C. A., Smith, P. E. & Tanner, J. J. A structurally conserved water molecule in rossmann dinucleotide-binding domains. *Protein Sci.* **11,** 2125–2137 (2009).
14. Bellamacina, C. R. The nicotinamide dinucleotide binding motif: a comparison of nucleotide binding proteins. *FASEB J.* **10,** 1257–1269 (1996).
15. Schmid, F.-X. Biological macromolecules: UV-visible spectrophotometry. *Encycl. Life Sci.* 1–4 (2001).
16. Griffith, P. R., Nye, M. J. & Alexander, J. C. Nuclear magnetic resonance and ultraviolet studies of the effect of p2H on reduced nicotinamide–adenine dinucleotide. *Can. J. Biochem.* **48,** 947–952 (1970).
17. Wilhelm, F. & Hirrlinger, J. The NAD+ /NADH redox state in astrocytes: independent control of the NAD+ and NADH content. *J. Neurosci. Res.* **89,** 1956–1964 (2011).
18. Catterall, W. A., Hollis, D. P. & Walter, C. F. Nuclear magnetic resonance study of the conformation of nicotinamide-adenine dinucleotide and reduced nicotinamide-adenine dinucleotide in solution. *Biochemistry (Mosc.)* **8,** 4032–4036 (1968).
19. Blumenstein, M. & Raftery, M. A. Natural abundance carbon-13 nuclear magnetic resonance spectra of nicotinamide adenine dinucleotide and related nucleotides. *Biochemistry (Mosc.)* **12,** 3585–3590 (1973).
20. Ragg, E. *et al.* ^{1}H, ^{13}C, ^{31}P-NMR studies and conformation analysis of NADP+ , NADPH coenzymes and of dimers from electrochemical reduction of NADP+. *Biochim. Biophys. Acta* **1076,** 49–60 (1991).
21. Patel, D. J. 220 MHz proton magnetic resonance spectrum of NADH. *Nature* **221,** 1239–1241 (1969).

22. Velick, S. F., Hayes, J. E. & Harting, J. The binding of diphosphopyridine nucleotide by glyceraldehyde-3-phosphate dehydrogenase. *J. Biol. Chem.* **203,** 527–544 (1953).
23. Eklund, H., Samama, J. P. & Jones, T. A. Crystallographic investigations of nicotinamide adenine dinucleotide binding to horse liver alcohol dehydrogenase. *Biochemistry (Mosc.)* **23,** 5982–5996 (1984).
24. Luisi, P. L., Baici, A., Bonner, F. J. & Aboderin, A. A. Relationship between fluorescence and conformation of epsilon NAD+ bound to dehydrogenases. *Biochemistry (Mosc.)* **14,** 362–368 (1975).
25. Lee, Y. M. *et al.* Conserved structural motif for recognizing nicotinamide adenine dinucleotide in poly(ADP-ribose) polymerases and ADP-ribosylating toxins: implications for structure-based drug design. *J. Med. Chem.* **53,** 4038–4049 (2010).
26. Tanner, J. J., Tu, S.-C., Barbour, L. J., Barnes, C. L. & Krause, K. L. Unusual folded conformation of nicotinamide adenine dinucleotide bound to flavin reductase P. *Protein Sci.* **8,** 1725–1732 (1999).
27. Pettegrew, J. W., Miles, D. W. & Eyring, H. Circular dichroism of adenosine dinucleotides. *Proc. Natl. Acad. Sci. U. S. A.* **74,** 1785–1788 (1977).
28. Weber, G. Intramolecular transfer of electronic energy in dihydro diphosphopyridine nucleotide. *Nature* **180,** 1409–1409 (1957).
29. Jardetzky, O. & Wade-Jardetzky, N. G. The conformation of pyridine dinucleotides in solution. *J. Biol. Chem.* **241,** 85–91 (1966).
30. Lee, C. Y., Eichner, R. D. & Kaplan, N. O. Conformations of diphosphopyridine coenzymes upon binding to dehydrogenases. *Proc. Natl. Acad. Sci.* **70,** 1593–1597 (1973).
31. Oppenheimer, N. J., Arnold, L. J. & Kaplan, N. O. A structure of pyridine nucleotides in solution. *Proc. Natl. Acad. Sci.* **68,** 3200–3205 (1971).
32. Zens, A. P. *et al.* Nuclear magnetic resonance studies on pyridine dinucleotides II: solution conformational dynamics of nicotinamide adenine dinucleotide and nicotinamide mononucleotide as viewed by proton T1 measurements. *J. Am. Chem. Soc.* **97,** 2850–2857 (1975).
33. Miles, D. W. & Urry, D. W. Reciprocal relations and proximity of bases in flavine adenine dinucleotide. *Biochemistry (Mosc.)* **7,** 2791–2799 (1968).
34. Frisch, M. *et al.* Gaussian 09. *Gaussian 09 Revis. B01 Gaussian Inc Wallingford CT* (2009).

35. Montgomery, J. A., Frisch, M. J., Ochterski, J. W. & Petersson, G. A. A complete basis set model chemistry VI: use of density functional geometries and frequencies. *J. Chem. Phys.* **110,** 2822–2827 (1999).

36. Wong, M. W. Vibrational frequency prediction using density functional theory. *Chem. Phys. Lett.* **256,** 391–399 (1996).

37. Andersson, M. & Uvdal, P. New scale factors for harmonic vibrational frequencies using the B3LYP density functional method with the triple-ζ basis set 6-311+ G (d, p). *J. Phys. Chem. A* **2,** 2937–2941 (2005).

38. Oppenheimer, N. J. & Kaplan, N. O. Structure of the primary acid rearrangement product of reduced nicotinamide adenine dinucleotide (NADH). *Biochemistry (Mosc.)* **13,** 4675–4685 (1974).

39. Prütz, W. A., Kissner, R., Koppenol, W. H. & Rüegger, H. On the irreversible destruction of reduced nicotinamide nucleotides by hypohalous acids. *Arch. Biochem. Biophys.* **380,** 181–191 (2000).

40. Johnson, S. L. & Polygena, T. T. Acid-catalysed hydration of reduced nicotinamide adenine dinucleotide and its analogues. *Biochemistry (Mosc.)* **16,** 1175–1183 (1977).

41. Wu, T., Wu, L. H., Knight, J. A. & Diethylaminoethyl, M. O. Stability of NADPH: effect of various factorson the kinetics of degradation. *Clin. Chem.* **32,** 314–319 (1986).

42. Burton, M. & Uni-, W. The reaction of Reduced Pyridine Nucleotides with Acid. *Arch. Biochem. Biophys.* 150–159 (1963).

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

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

45. DeFlores, L. P., Ganim, Z., Ackley, S. F., Chung, H. S. & Tokmakoff, A. The anharmonic vibrational potential and relaxation pathways of the amide I and II modes of n-methylacetamide. *J. Phys. Chem. B* **110,** 18973–18980 (2006).

46. Rubtsov, I. V., Wang, J. & Hochstrasser, R. M. Dual-frequency 2D-IR spectroscopy heterodyned photon echo of the peptide bond. *Proc. Natl. Acad. Sci.* **100,** 5601–5606 (2003).

47. Koziński, M., Garrett-Roe, S. & Hamm, P. Vibrational spectral diffusion of CN− in water. *Chem. Phys.* **341,** 5–10 (2007).

48. Yu, P., Yang, F., Zhao, J. & Wang, J. Hydration dynamics of cyanoferrate anions examined by ultrafast infrared spectroscopy. *J. Phys. Chem. B* **118,** 3104–3114 (2014).

49. Woutersen, S. *et al.* Peptide conformational heterogeneity revealed from nonlinear vibrational spectroscopy and molecular-dynamics simulations. *J. Chem. Phys.* **117,** 6833–6840 (2002).

50. Bagchi, S., Charnley, A. K., Smith III, A. B. & Hochstrasser, R. M. Equilibrium exchange processes of the aqueous tryptophan dipeptide. *J. Phys. Chem. B* **113,** 8412–8417 (2009).

51. Hamm, P. & Zanni, M. T. *Concepts and methods of 2D infrared spectroscopy*. (Cambridge University Press, 2011).

52. Sarma, R. H., Ross, V. & Kaplan, N. O. Investigation of the conformation of β-diphosphopyridine nucleotide (β-nicotinamide adenine dinucleotide) and pyridine dinucleotide analogs by proton magnetic resonance. *Biochemistry (Mosc.)* **7,** 3052–3062 (1968).

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

54. Hixson, S. S. & Hixson, S. H. Photochemical labelling of yeast alcohol dehydrogenase with an azide analog of NAD+. *Photochem. Photobiol.* **18,** 135–138 (1973).

55. Dutta, S., Cook, R. J., Houtman, J. C. D., Kohen, A. & Cheatum, C. M. Characterization of azido-NAD+ to assess its potential as a two-dimensional infrared probe of enzyme dynamics. *Anal. Biochem.* **407,** 241–246 (2010).

56. Moses, J. E. & Moorhouse, A. D., The growing applications of click chemistry, Chem. Soc. Rev., **36**, 1249-1262, (2007).

5. Extracting local vibrational dynamics from chaperone-protein complexes

5.1 Abstract

Many bacteria utilise macromolecular structures as vehicles for toxin transfer. Such “secretion systems” frequently show highly-conserved traits in the structures of their macromolecule-forming proteins, as well as in the general mechanisms of toxin secretion. The high degree of conservation observed between pathogen secretion mechanisms indicates a highly effective system that has required minimal adaption over time. However, slight qualitative deviations are consistently observed in the protein-protein interactions of a sub-set of human-opportunistic pathogens, suggesting that this discrepancy may hold mechanistic importance.

Proteins are the largest single biomolecules, and their size can readily impose limits to various measurement techniques. Consequently, protein-protein complexes are particularly challenging case-studies, though they have significant implications to health and disease. Here, 2DIR spectroscopy is applied to a somewhat simplified protein-polypeptide system that holds importance to bacterial toxin-transfer mechanisms.

With the objective of developing a reliable approach to the extraction of local dynamics from large, complex protein-protein systems, several experimental approaches are taken in this chapter to exploit the intrinsic vibrations of residues in a protein-protein binding site as IR probes.

5.2 Introduction

Polypeptide chains expressed *in vivo* are required to fold into their “native state” within crowded cellular environments[1,2]. The combination of interactions available to a protein *via* its primary structure do not exclusively result in this native conformation, and can readily produce “misfolded” states that are not only non-functional, but detrimental to the organism[3]. Many non-native protein conformations feature incomplete burial of hydrophobic side chains, which in turn promotes the aggregation of misfolded molecules through the hydrophobic effect[4]. As large proteins can require several hours to fold successfully[5], there is a significant risk of multiple amorphous proteins aggregating. To counter these effects, biological systems have evolved safeguard mechanisms, in which the class of proteins known as chaperones play a key role[1,6].

Chaperones are a ubiquitous family of proteins with diverse biological roles in protein-protein interaction[7–9]. As their substrates are exclusively other proteins, they do not necessarily possess well-defined active sites. Rather, interactions with a substrate can potentially be delocalised across multiple structural domains of the chaperone[1,6]. Chaperone proteins serve vital roles in the regulation of protein structure by forming complexes with unfolded or misfolded polypeptide chains.

Folded proteins typically possess hydrophilic, solvent-exposed surfaces, and hydrophobic interiors[10,11]. However, solvent exposure of nonpolar side chains during the folding process can drive polypeptides away from their native forms and into amorphous, aggregated structures. Chaperones inhibit aggregation by stabilising non-native conformations, allowing them to reach a functional, native state[1,9]. The chaperoning of substrate proteins is further believed to increase the favourability of mutated polypeptide sequences folding into a “native” conformation, thereby assisting biological evolution[1,12]. Chaperones involved with assisted-folding are often found to have further roles in protein disassembly, which likewise prevents the deleterious effects of protein aggregation[13]. The implications of chaperone action to health and disease are therefore vast.

Protein-chaperone systems present a significant challenge for many experimental techniques, as they fundamentally involve the interaction of two complex biomolecules, one of which undergoes a dynamic conformation transition while in complex with the chaperone. A technique such as X-ray crystallography is poorly-equipped to capture the dynamics that take place between binding and dissociation of the substrate protein, which ultimately prevents access to the most vital observations of chaperone interactions; conformation changes[14–16]. Likewise, a chaperone’s ability to produce a narrow distribution of conformations (the “native” state) from a broad catchment of initially misfolded states is of key interest, and the range of initial substrate conformations poses a challenge to crystallisation procedures[16]. Further, the amount of time required to extract and purify a chaperone complex after its expression is likely comparable to the length of time required for assisted folding to occur. For these reasons, structural information on chaperone systems has been limited to specific cases. Though it is possible to obtain dynamic information from these systems using NMR, the large scale of protein-chaperone complexes frequently exceeds the 40KDa "size limit" of this technique.

Motivation for unravelling the details of protein-chaperone interaction needs little exposition. Health conditions such as Parkinson's, Alzheimer's and Amyloidosis are associable to the aggregation of misfolded proteins into insoluble masses known as amyloids[3,4]. Reciprocally, many human-opportunistic pathogens utilise chaperone proteins to regulate the protein aggregation that underpins their toxification mechanisms[17–19].

The bacterium, *Pseudomonas aeruginosa (P. aeruginosa)* is a highly resilient, human-opportunistic pathogen that utilises a so-called type III secretion system (T3SS)[17,20–22] to attack target eukaryotic cells (Figure (1)(a)). Here, over 20 species of bacterial proteins assemble to form a macromolecular needle structure; a proteinaceous syringe that penetrates the membrane of target cells, allowing the transfer of toxic "effector" proteins directly into the target's cytoplasm. Critical to this mechanism is the "translocon" sub-structure; a pore-like construction at the needle-tip that permeates the membrane of the target cell.

(a)

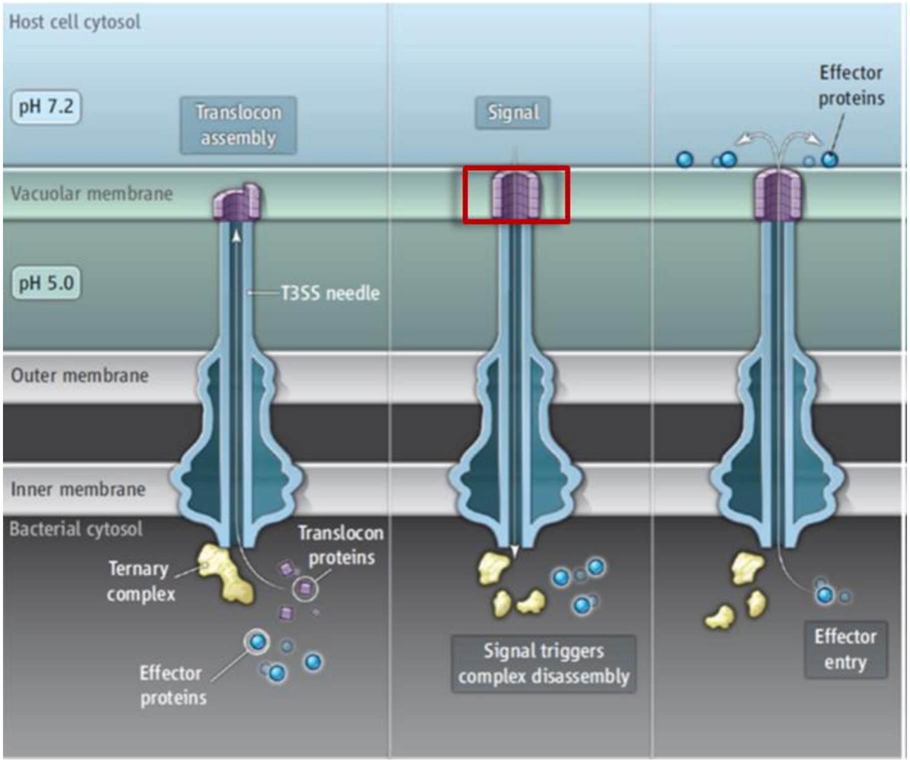
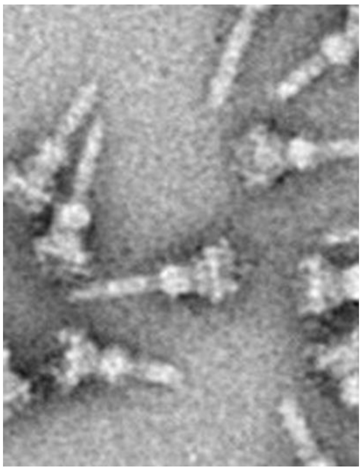


Figure (1): (a) Illustration of a type III secretion system[23]*, with the translocon pore outlined in red. (b) Cryo electron Microscopy image of T3SS needles from Salmonella typhimuriun*[20]*.*

(b)



A subtle discrepancy in the behaviour of T3SS has been proposed to occur exclusively in the toxification mechanism of human-opportunistic pathogens, a proposal that has recently received experimental verification *via* the study of *P. aeruginosa*[24,25]. In the vast majority of toxins, needle-forming proteins remain in chaperone-bound forms prior to macromolecule-assembly, with each species bound to a dedicated chaperone. However, in human opportunistic pathogens, a sub-class of these proteins shares a common chaperone, and further, associates to the same binding site[25]. The sub-class in question, the hydrophobic “translocators” are critical to translocon formation[26]. In the absence of their mutual chaperone, PcrH, each of the hydrophobic translocators, PopB and PopD, forms a ring-like octamer with 40 Å inner diameter and 80 Å outer diameter[21]. Mutation studies of PcrH have also revealed that inhibition of PcrH-translocator complex formation results in the rapid degradation of the bacterium[22,25]. Despite highly dissimilar amino acid sequences, the two hydrophobic translocators, PopB and PopD are each recognised by their chaperone, and their dissociation is similarly triggered at low pH[24]. This indicates a structural recognition mechanism that is vitally linked to the toxification mechanism and stability of PcrH. This result potentially extends to other human-opportunistic pathogens utilising T3SS, owing to the high degree of similarity between the needle-forming proteins of such bacteria. Indeed, the structure of PcrH has been seen to display high similarity with many other T3SS chaperones[27]. It includes three tetratricopeptide repeat (TPR) motifs, each of which consists of two α-helix structures aligned in an anti-parallel fashion. The TPR motif is commonly associated with protein-protein interactions[28], and has thus been well-categorised. For this reason, further information on the structural interactions of the hydrophobic translocators with this common chaperone motif is likely to prove insightful to the understanding of multiple systems and subsequently benefit drug-design[29].

At present, structural information on the complexes of hydrophobic translocators with their chaperone is limited to the *P. aeruginosa* system, chiefly due to the difficulty of crystallising hydrophobic proteins[15]. In this case of *P. aeruginosa*, crystallisation of a PopD-PcrH complex could only be performed by reducing the PopD structure to the ten-amino acid chain, $^{47\text{-}56}$PopD (Figure (2))[25]. Point-mutation studies have identified this short polypeptide as the key binding region of PopD. Indeed, binding of the translocator to PcrH is seen to be principally stabilised by three hydrophobic residues in peptide sequence; Val-49, Leu-51 and Ala-53. These residues are believed to be solvent-exposed, with the PopD-PcrH binding interaction allowing burial of the side chains between two helical regions of the chaperone; the so-called binding cleft. This marginally simplified system provides a model from which to derive broader understanding of the T3SS mechanism.

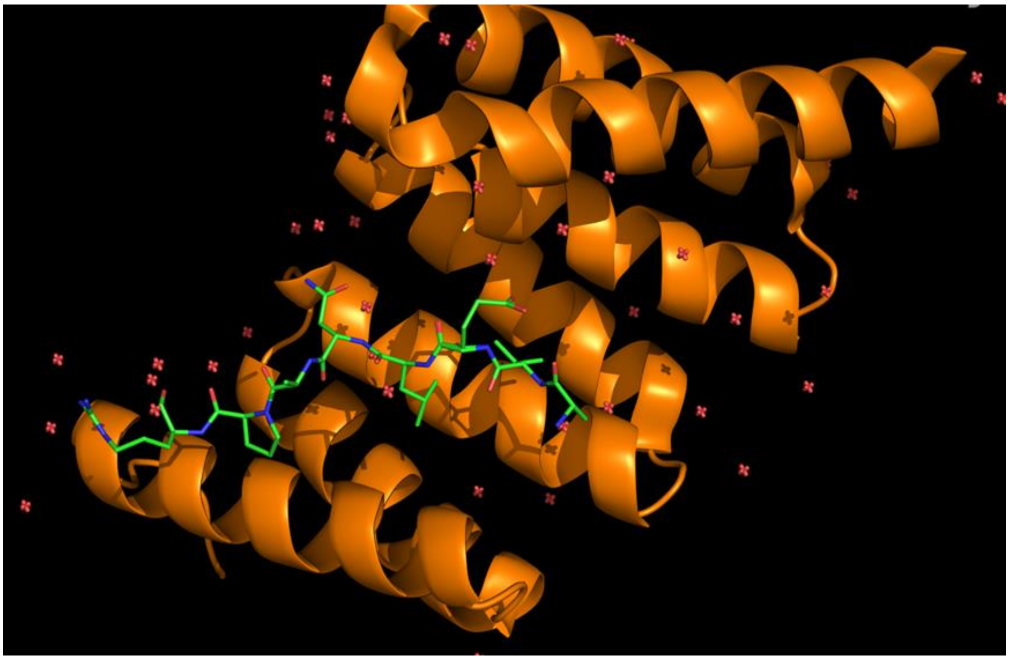


Figure (2): PcrH (orange) in complex with the PopD structural analogue, $^{47\text{-}56}$PopD (green). Crystal structure obtained by Job et al.[25] *(Protein databank (PDB) code: 2XCB).*

Vitally, substrate recognition mechanisms have been shown to originate from dynamic fluctuations in the conformation of local binding-site architecture[30–33], with individual substrates stabilising a particular conformational state. Such a recognition mechanism is likely to hold significance to other bacterial species with type III secretion systems. Such pathogens utilise highly conserved proteins in the construction of translocon systems, though they do so as a vehicle for the more diverse effector proteins[27].

The substitution of PopD for a relatively short peptide sequence offers perhaps the strongest opportunity to study a chaperone system using infrared spectroscopy. The “ligand” in this case is still a large molecule, likely with complex spectral features. However, each residue the translocator binding domain is known to occupy the binding site of PcrH, and thus any spectral changes observed upon complex formation can still be wholly attributed to interactions at this functionally-important region. This condensed analogue of PopD also permits the use of IR experimental strategies, such as isotope-labelling, that would be impractical for a full protein. Additionally, the ligand spectroscopy is somewhat simplified by this structural analogue, though it remains vastly more complex than either of the previously discussed case-studies, nitric oxide or nicotinamide adenine dinucleotide. Unlike other chaperone systems, the substrate in this case does not undergo continuous conformation transitions that would make measurement challenging. The temporal resolution of a technique such as 2DIR is not well-suited to the slow (ms-hours)[1,5] dynamics of such folding processes, though more rapid dynamics, such as those potentially associated with a structural recognition mechanism, may occur within fs-ps timescales[30–32].

In this investigation, 2DIR is applied to the complex of PcrH with ten-residue polypeptides resembling the binding domain of PopD, ^{47}DRVELNAPRQ^{56} (hereafter referred to as $^{47\text{-}56}$PopD). Several approaches are taken to the extraction information from this complex, with the objective of formulating a generic experimental approach that can be reused for multiple protein-protein systems. The success of these approaches is determined by their capacity to provide site-specific dynamics.

5.3 Experimental

5.3.1 Infrared spectroscopy

All linear IR spectra were obtained at room temperature, using either a Bruker Vertex 70 spectrometer (peptides) or a Nicolet IS-10 (proteins and protein complexes). Samples were housed between two CaF_2 windows separated by a 25 μm spacer. Protein concentrations varied throughout due to slight discrepancies in the concentration stage, and resultantly, the peptide concentration in complex samples also varied. PcrH was typically used ~50 mg/ml, or ~2.8 mM, however. Free peptide spectra were recorded at 10 mM. A path length of 25 μm was also employed in 2DIR measurements.

2DIR spectra were collected using the ULTRA laser system. Spectra were averaged over three separate measurements. The probe frequency spectra collected for each value of the interferometer stage position were averaged from 4000 laser shots in each measurement.

5.3.2 Synthesis of $^{47\text{-}56}$PopD peptides

Peptides were synthesised using a solid-phase, Fmoc protocol[34–36]. Further details of the synthesis, purification and analysis procedures are provided in Appendix sections (A.5.1), (A.5.2) and (A.5.3), respectively. Briefly, a Rink Amide 4-methylbenzhydrylamine (MBHA) resin was manually loaded with the N-terminus amino acid in the $^{47\text{-}56}$PopD sequence, Gln-56 (Q). Subsequent amino acids from the N to C termini were coupled in automated fashion using a Tribute peptide synthesiser, manufactured by Protein Technologies Inc. Isotopically-labelled residues were coupled manually.

The $^{47\text{-}56}$PopD peptides were purified using reverse-phase high-performance liquid chromatography (HPLC). Mass spectra were recorded for purified peptide fractions to confirm successful synthesis. These were acquired using a ThermoQuest Finnigan LC duo coupled to a Razel syringe pump.

5.3.3 Protein expression and purification

PcrH expression and purification was conducted by Dr Domenico Bellini *via* the following protocol:

The section of the *P. aeruginosa* CHA PcrH gene coding for amino acids 1-160 cloned into vector pETDuet-1 (Novagen) was kindly supplied by Andrea Dessen from IBS (Grenoble). The clone contained an additional thrombin site, allowing for cleavage of the His tag upon protein purification. PcrH (1-160) was expressed in *E. coli* BL21(DE3). Cultures were grown at 37 °C to A600 nm of 0.7 A.U., and expression of the recombinant proteins was induced by the addition of 1 mM isopropyl 1-thio-β-D-galactopyranoside for 3.5 h at 30 °C. Cells were lysed on ice by sonication in 25 mM Tris-HCl, pH 8.0, 0.2 M NaCl, 5% glycerol, in the presence of protease inhibitors (Roche Applied Science) and then centrifuged at 30,000g for 45 minutes. The supernatant was applied to a His Trap HP 5-ml column (GE Healthcare) washed with 25 mM Tris-HCl, pH 7.4, 0.2 M NaCl, 20 mM imidazole, 5 mM mercaptoethanol, and the bound proteins were eluted with a linear gradient of the same buffer containing 500 mM imidazole. Prior to thrombin cleavage, proteins were dialyzed into 25 mM Tris-HCl, pH 8.0, 0.1 M NaCl for 1 h. Cleaved products were loaded onto a His trap column to eliminate the un-cleaved fractions and subsequently onto a Superdex 200 column previously equilibrated in 20 mM Tris-HCl, pH 8.0, 0.2 M NaCl, 2 mM EDTA, and 3 mM dithiothreitol.

5.3.4 Isothermal Titration Calorimetry

ITC was used to determine the optimal concentrations for the IR by determining the binding constants of various synthesised peptide species with a target protein, PcrH. Again, further detail on this technique is provided in Appendix section (A.5.4).

Peptide-protein dissociations constants were measured using a MicroCal ITC 200. A 300 μM protein solution of volume 0.2 ml was stabilised at a temperature of 25 ˚C, and peptide from a 10 mM stock solution was injected into the cell in 2.43 μl volumes. The temperature stabilisation time between injections was set to 3 minutes. Binding was seen to be exothermic.

5.4 Results

The following results are presented in an alternate format to previous chapters, to reflect iterative redesign of the experimental approach over the course of the study. Periodic method refinement was necessary, as systems of this nature have not previously been examined using 2DIR, and so no generic protocol currently exists.

To date, the most closely-related investigations are those performed by Zanni *et al.*[37–39] and by Hochstrasser *et al.*[40,41] on amyloid fibril and transmembrane protein systems. In each case, the polypeptide chain under investigation was short enough to be synthesised, permitting the application of $^{13}C{=}^{18}O$ labels. Both the amyloid fibril and membrane protein were sufficiently large and well-ordered to exhibit highly delocalised Amide I vibrations, and as a result, systematic isotope-editing of the backbone C=O groups proved to be extremely effective in modulating transition dipole coupling across the secondary structure, from which a significant quantity of conformational detail could be inferred. This approach is not as viable for $^{47\text{-}56}$PopD, as the peptide is not observed to adopt secondary structure. As a result, the Amide I band likely to contain few dipole-dipole contributions, from which the spectral changes associated with isotope-labelling are expected to be minimal.

In each sub-experiment performed here, the required sensitivity could only practically be delivered by the state-of-the-art ULTRA spectrometer[42,43] at the Rutherford Appleton Laboratory. Owing to this, the time available for 2DIR measurements was highly restricted. Further, the lack of commercial availability of protein and peptide samples severely reduced the ease with which supporting FTIR studies could be performed.

5.4.1 Wild-type peptide and PcrH: FTIR

The linear spectrum of PcrH is displayed in Figure (3) alongside the linear spectrum of the wild-type peptide sequence, $^{47\text{-}56}$PopD.

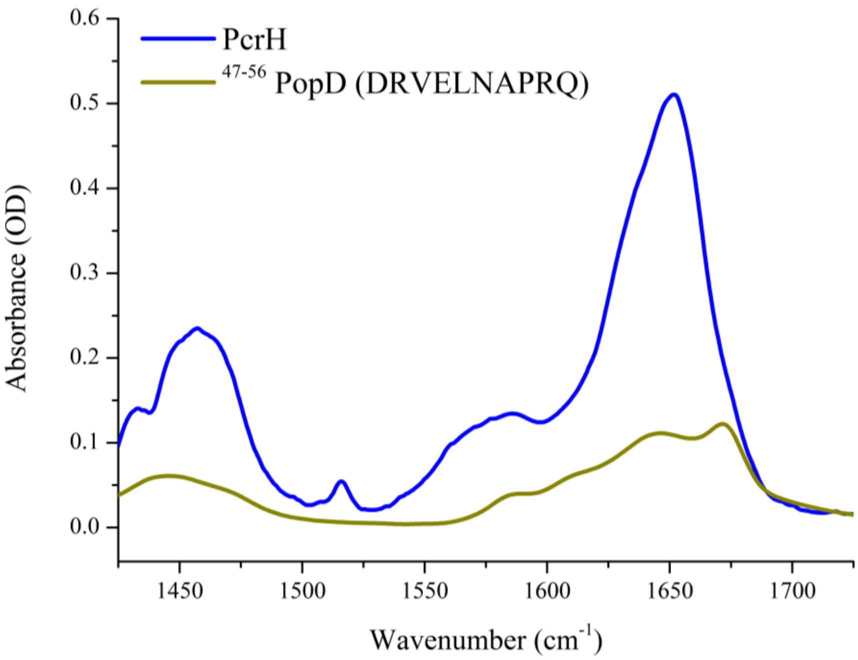


Figure (3): The linear IR spectra of PcrH (blue) and $^{47\text{-}56}$*PopD (gold) at equimolar concentration (5 mM).*

PcrH is composed of 7 α-helical domains interspersed by short turns, and the former dominates the protein’s IR absorption, giving rise to a strong Amide I band with central frequency ~1650 cm^{-1}. Other bands in the 1600-1700 cm^{-1} region are heavily obscured by spectral contributions from secondary structure, preventing the accurate assignment of further peak positions. Other characteristic protein features are observed at: 1460 and 1580 cm^{-1}; the deuterated Amide II and the carboxylate modes, respectively. As the secondary structure of PcrH is predominantly α-helical[25], it is not possible to isolate spectral contribution of the two helical domains to which $^{47\text{-}56}$PopD binds. The reportedly rigid structure of PcrH also reduces the likelihood of observing changes to the Amide I band upon binding of the peptide.

The spectrum of the wild-type peptide is less generic than that of the protein; four distinct bands are visible at 1586, 1610, 1645 and 1672 cm^{-1}. Though PopD is also predominantly α-helical[44], it is not clear whether the binding region natively adopts this structure. The crystallised complex of the isolated binding region reflects no secondary structure[25], and thus the Amide I band is unlikely to be well-defined through dipole-dipole interactions. The broad feature at 1645 cm^{-1} is therefore assigned to the Amide I of a disordered structure[45,46]. Detailed assignments of the absorptions of amino acid side chains suggest that the 1672 cm^{-1} likely originate from the C=O stretching modes of the asparagine, glutamine or glutamate residues within the sequence[47,48]. These designations are made with reference to the vibrational frequencies of free amino acids, though it is acknowledged that the potential for local environment to modify these vibrations is significant. Without computational support, it is challenging to anticipate the degree of similarity between the free side chain absorption spectrum and its protein counterpart, particularly when the polypeptide structure is unknown, as is the case for $^{47-56}$PopD. Assignments of the peptide's vibrational bands are thus somewhat speculative.

Without further modification of the peptide, the C=O stretching modes at 1672 cm^{-1} have the greatest potential as Infrared probes. Since only the binding region of PopD is employed in this study, information extracted from these modes is also specific to the protein binding site. It can be seen in Figure (3), however, that the Amide I of PcrH overlaps with the peptide 1672 cm^{-1}. Distinction of the latter mode therefore depends upon the concentration ratio between protein and peptide. Ideally, to extract information from the peptide, the experiment is optimised to achieve the highest concentration of bound peptide molecules. This must be determined from the dissociation constant of the system, which can be measured using isothermal titration calorimetry (ITC).

Preliminary FTIR tests of peptide solubility exposed potential constraints on the design of ITC experiments. At high concentrations of 25-30 mM, peptide solutions underwent rapid gelation. Gelled peptide samples displayed significantly modified IR characteristics from their solution-phase counterparts (Figure (4)). New absorption features were seen to appear ~1630 cm^{-1}. Although these vibrations could not be firmly assigned without detailed investigation, β-sheet formation is offered as a speculative explanation based on peptide self-assembly studies that have correlated solution-gel phase transitions with formation of this secondary structure type[49].

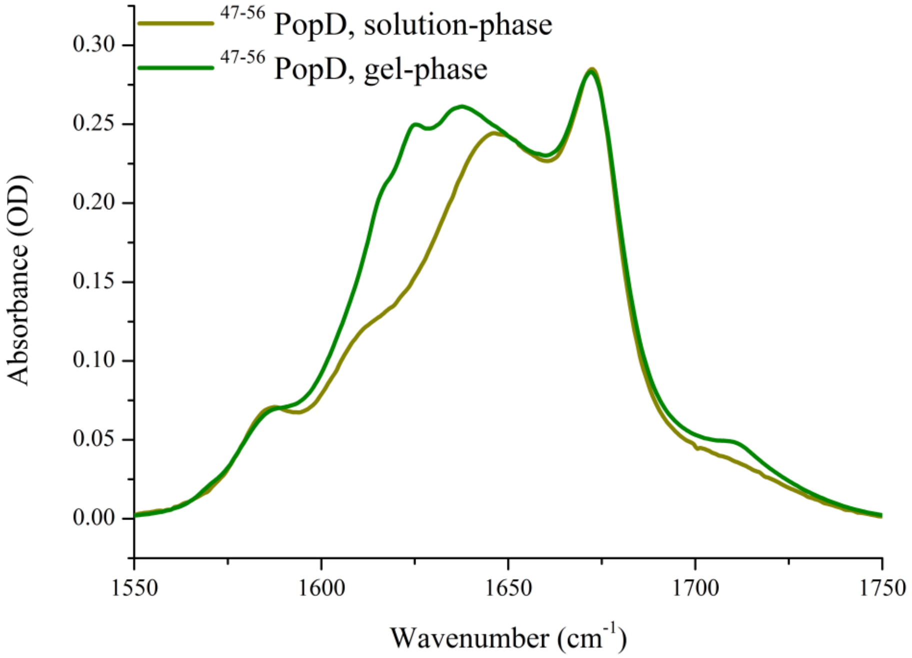


Figure (4): FTIR spectra of solution-phase $^{47\text{-}56}$*PopD (10mM) and gel-phase* $^{47\text{-}56}$*PopD (~25 mM). The latter spectrum is scaled to match the 1672 cm*$^{-1}$ *intensity of the former.*

No further efforts were made to establish the structures formed by peptide aggregation, as the self-assembly reactions of even simple peptides can generate highly complex macromolecular structures[49]. In the case of the ten-residue sequence used here, a number of possible inter-peptide interactions are likely possible.

At lower concentration (10 mM), small discrepancies between different peptide samples were also seen to exist, suggesting that small quantities of secondary structure may form significantly below 25 mM. The ability of the peptide to serve as a local probe is somewhat compromised by this, as the acid modes are likely to relay information from other peptides, as well as from the protein. Again, this effect can be minimised by limiting the quantity of unbound peptide in a peptide-protein sample, for which knowledge of the dissociation constant is a prerequisite. Further implications of peptide aggregation, with respect to both ITC and Infrared spectroscopy, are significant, and are discussed in more detail below.

Limited quantities of $^{47\text{-}56}$PopD were available to thoroughly investigate the source of limited spectral reproducibility. The effects of non-uniform TFA (trifluoroacetic acid) impurities in the lyophilised peptide were considered a possible source of uncertainty, though a pH dependence study of the peptide revealed few spectral changes in the range 4-10. Further, the purified peptide displayed no capacity to modify the pH of unbuffered deuterium oxide, suggesting that TFA impurity was negligible. Throughout this study, peptide spectra were observed to show optimal reproducibility in DMSO. Possible explanations for this are developed in later discussion.

5.4.2 Wild-type peptide and PcrH: ITC

Isothermal titration calorimetry was used to measure the dissociation constant of the PcrH-$^{47\text{-}56}$PopD complex, from which a sigmoidal titration curve (Figure (5)) was obtained. This “thermogram” was fitted assuming a single binding-site between the protein and peptide, yielding a dissociation constant of 2.4 mM.

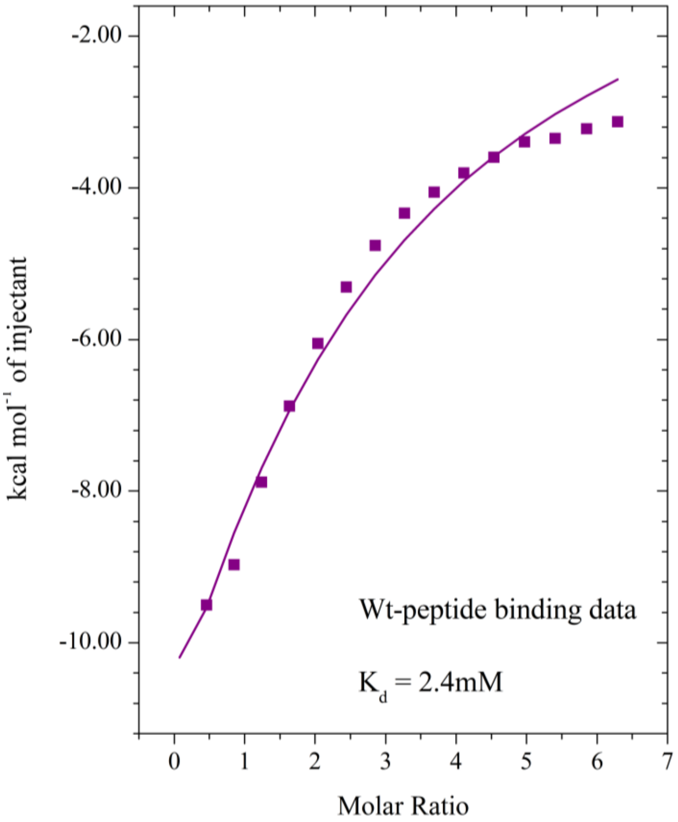


Figure (5): Titration curve of PcrH binding with $^{47\text{-}56}$*PopD, annotated with the dissociation constant extracted from the ITC experiment.*

This value reflects weak, stoichiometric binding between the molecules. The K_d derived from this experiment is likely to be an approximation, as saturation of the complex did not occur within the experimental range, likely affecting the fitting procedure used to determine K_d. However, the binding is still evidently weak, as indicated by the lack of complex saturation at a molar ratio of greater than 6:1 peptide: protein. Although the constraint of the binding stoichiometry to 1:1 resulted in a poor overlay between the fit-function and data points, in the case of weakly binding systems where complex formation does not saturate, accurate derivation of the binding stoichiometry from the data cannot be performed. As such the number of binding sites used in the fitting of all ITC data was fixed at 1, which reflects the observation of a single peptide binding site from X-ray crystallography. As a result of this weak binding and constrained model, the exact dissociation constant derived here, and in subsequent ITC measurements in this chapter, are treated as approximations only.

ITC is optimally accurate in the determination of dissociation of constants in the range of 1 nM-100 μM[50,51], and thus the reduction of uncertainty predicted through higher concentration repeats of the titration experiment was deemed disproportionate to the quantity of material required. Further, it was anticipated that aggregation and solubility of the peptide would become increasingly significant at higher concentration.

At the present concentration, peptide-peptide binding may still contribute in part to the heat release observed in the titration curve above. Accurate isolation of this contribution could not be performed without risking damage to the calorimeter, as the binding "saturation" in this experiment likely corresponds with the point of gel-formation. Additionally, the effective number of "binding sites" available in the peptide-peptide interaction likely presents an incredibly complex model for analysis.

It is noteworthy that Job *et al.* obtained the crystal structure of the PcrH-PopD complex from water: DMSO mixtures[25], suggesting that complex formation may be more favourable under these conditions. Under their protocol, the peptide is dissolved in pure DMSO prior to mixture with the protein. As DMSO is amphiphilic, it may limit hydrophobically-driven aggregation by effectively solvating both polar and nonpolar peptide residues.

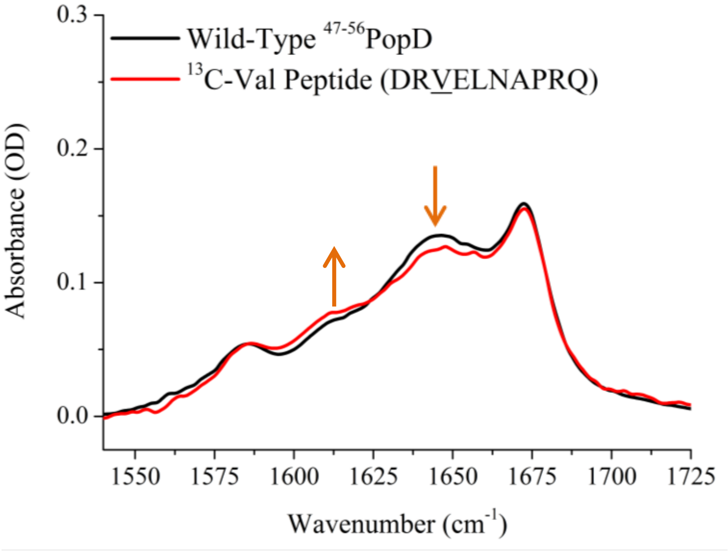
The viscosity of dimethyl sulfoxide poses a significant challenge to ITC, and can generate significant heat release when the cell and titrant DMSO concentrations are not precisely matched[52]. It was therefore not possible to use the protocol used by Job *et al* [25]. Binding measurements were therefore restricted to a non-optimal scenario from which the binding constant is something of an approximation. Regardless, weak binding reduces the practicality of the peptide's 1672 cm^{-1} mode as a local probe, as the concentration of PcrH must be significantly greater than that of the peptide to ensure that majority of peptide molecules are bound. At these relative concentrations, the peptide 1672 cm^{-1} mode cannot be resolved from the high frequency edge of PcrH's Amide I peak.

The protein-peptide system does not intrinsically lend itself to site-specific information, though this is an unsurprising result. Attempts were thus made to produce a spectrally-isolated reporter band through structural modification of $^{47-56}$PopD.

5.4.3 Isotope-labelled peptides and PcrH: FTIR

Isotope-labelling was employed to produce three additional peptide species. In each case, the backbone carbonyl of a single, different residue was substituted with a ^{13}C=O (i.e. ^{13}C=^{16}O) functional group, causing a downshift in the Amide I frequency for the selected residue. The hydrophobic residues: leucine, alanine and valine were selected for isotope-labelling due to their proposed significance in binding. Linear spectra of the three isotope-labelled peptides are each shown alongside that of the wild-type species in Figure (6).

(a)



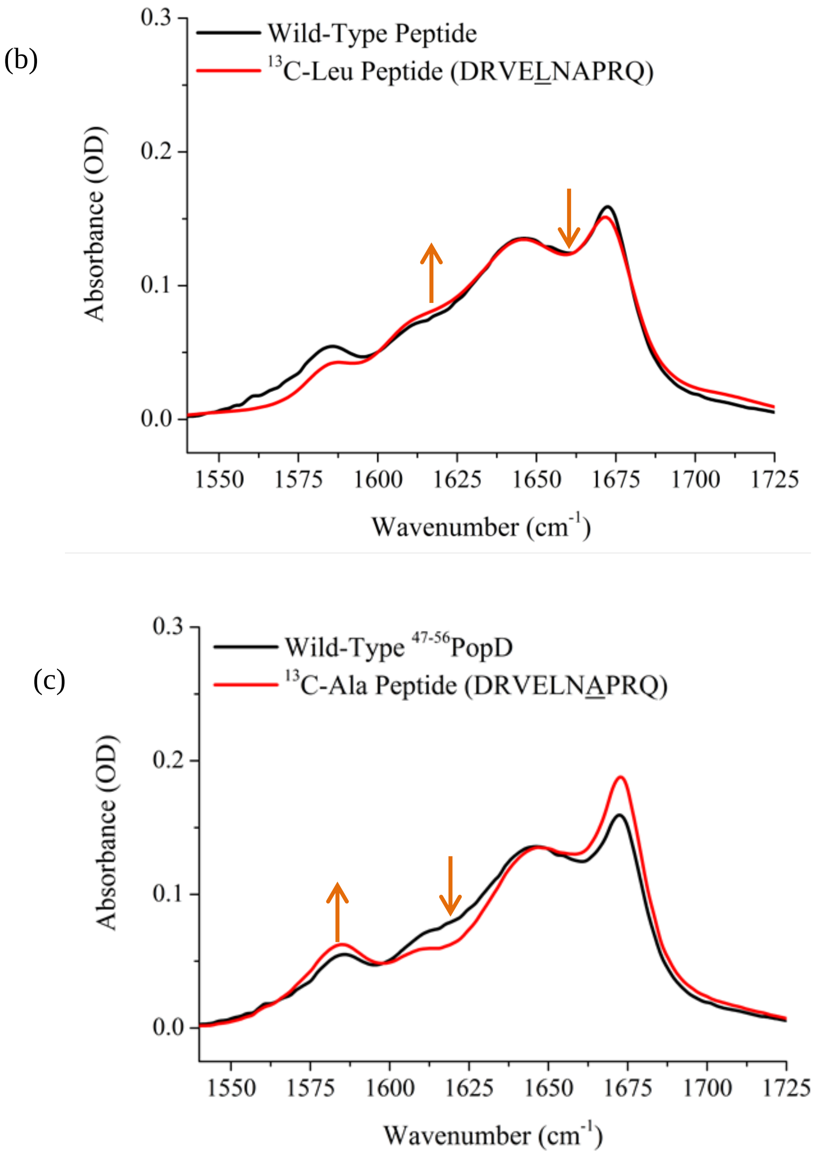


Figure (6): Comparison of the linear spectra of the wild-type $^{47\text{-}56}$PopD peptide (black), in relation to the isotopically-labelled peptides; (a) ^{13}C-Val, (b) ^{13}C-Leu and (c) ^{13C}-Ala

As the isotope label shifts a population of oscillators to a lower frequency, the creation of a ^{13}C=O mode will also be accompanied by depletion of the corresponding ^{12}C=O band. Through comparison of the wild-type and labelled peptides, it is possible to ascertain the unlabelled Amide I frequencies of the alanine, leucine and valine residues. Isotopic labelling of alanine reduces the peptide absorption ~ 1620 cm^{-1} and increase the absorption ~ 1585 cm^{-1}. Leucine and valine labels reduce the optical density ~ 1665 and 1650 cm^{-1}, increasing the optical density at ~1620 and 1610 cm^{-1}. Poor reproducibility of the peptide spectra creates some uncertainty in these assignments, as several other spectral changes are simultaneously observed between peptide species. The intensities ~1672, 1586 and 1570 cm^{-1} vary in ways that are not attributable to the isotope label. Reference measurements were thus performed in DMSO, supporting the assignments given above. The large variation in Amide I frequencies between each residue confirms the lack of secondary structure in $^{47\text{-}56}$PopD.

It can be seen that the unlabelled Amide I bands of both leucine and valine lie at relatively high frequency within the peptide spectrum. For this reason the modes are well-resolved from other peptide bands after labelling. The alanine-labelled variant (^{13}C-Ala-$^{47\text{-}56}$PopD) shows the most potential as an IR probe, though it is still far from ideal when considering the PcrH spectrum (Figure (8)). However, the down-shifted alanine mode has an advantage over the other vibrations considered here (*i.e.* the 1672 cm^{-1} peak, the leucine and valine variants), in that the new position of its Amide I lies at a region in which no other Amide I vibrations occur. In the PopD-PcrH system, the PopD binding region is enclosed within two helical regions of the chaperone, allowing the potential coupling of Amide I vibrations between the molecules. In a 2DIR spectrum, any cross-peaks between the peptide's alanine residue and PcrH would similarly be translated by isotope-labelling, potentially allowing them to be separated from other Amide I to Amide I couplings.

Considering the action of the alanine, valine and leucine side chains as hydrophobic "anchors" in the binding of PcrH, these side chains may appear the more logical choice for isotope-labelling. However, vibrations originating from aliphatic functional groups generally give rise to low molar extinction coefficients and weak absorption bands. Side chain modes are also unlikely to prove as sensitive to intermolecular coupling as the Amide I, owing to the large contribution on C=O stretching in the latter case.

5.4.4 Isotope-labelled peptides and PcrH: 2DIR

Spectra of PcrH, $^{47\text{-}56}$PopD and the wild-type complex are shown in Figure (7) at waiting times of 250 fs. Individual contributions to the protein spectrum are more clearly resolved here than in the linear absorption profile, allowing identification of bands at 1632 and 1645 cm^{-1}. The peptide's strong 1672 cm^{-1} band is discernible in the complex spectrum, though the remainder PcrH still contributes significantly to the optical density in this region.

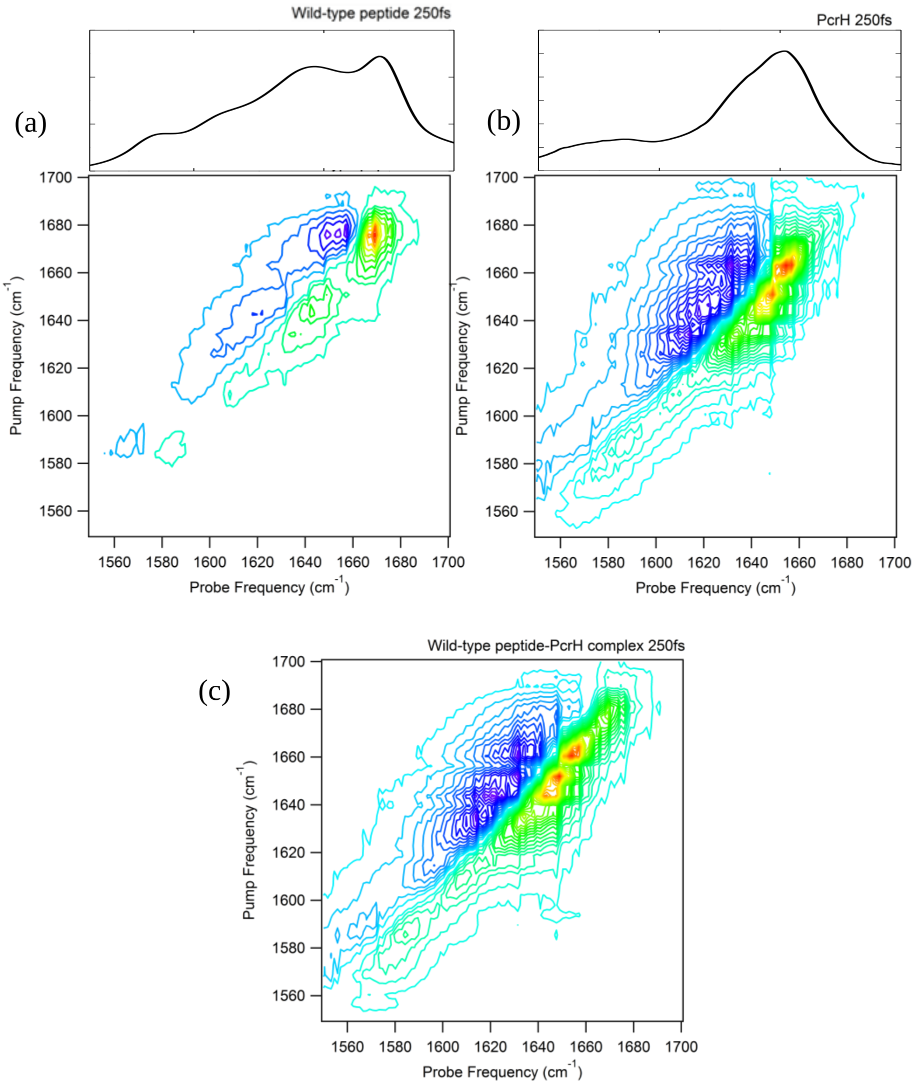


Figure (7): Representative 2DIR spectra of the (a) $^{47\text{-}56}$PopD, (b) PcrH and (c) the corresponding chaperone-peptide complex. All spectra are shown at the shortest waiting times of 250 fs, where the signal to noise ratio is greatest.

As a result of spectral convolution, the most effective analysis of this system ideally originates from new absorption information, i.e. spectral features of the complex that are not ascribable to a linear combination of the protein and peptide spectra. Substantial changes to the conformation of PcrH would likely give rise to pronounced changes in the Amide I. However, no such changes are observed here, consistent with circular dichroism measurements of the PcrH-PopD complex[44]. This diminishes the value of on-diagonal features in providing information specific to the chaperone complex. Off-diagonal features are not easily identified here. Without an isotope label, potential cross-peaks from the more intense Amide I of the peptide are likely to be obscured by the on-diagonal features. Indeed, pump axis frequency slices through wild-type complex reveal no resolvable off-diagonal peaks.

Figure (8) displays 2DIR spectra of ^{13}C-Ala-$^{47\text{-}56}$PopD and its complex with PcrH, at 250 fs. A slightly increased absorption of the complex at 1585 cm^{-1}, relative to the protein complex in Figure (7), reflects the isotope label. Several vibrations absorb light at this frequency, such as the 1580 cm^{-1} mode of the protein, the wild-type peptide's 1586 cm^{-1} band, and both the bound and unbound species of the labelled alanine residue. Vibrational relaxation data therefore remains unreliable.

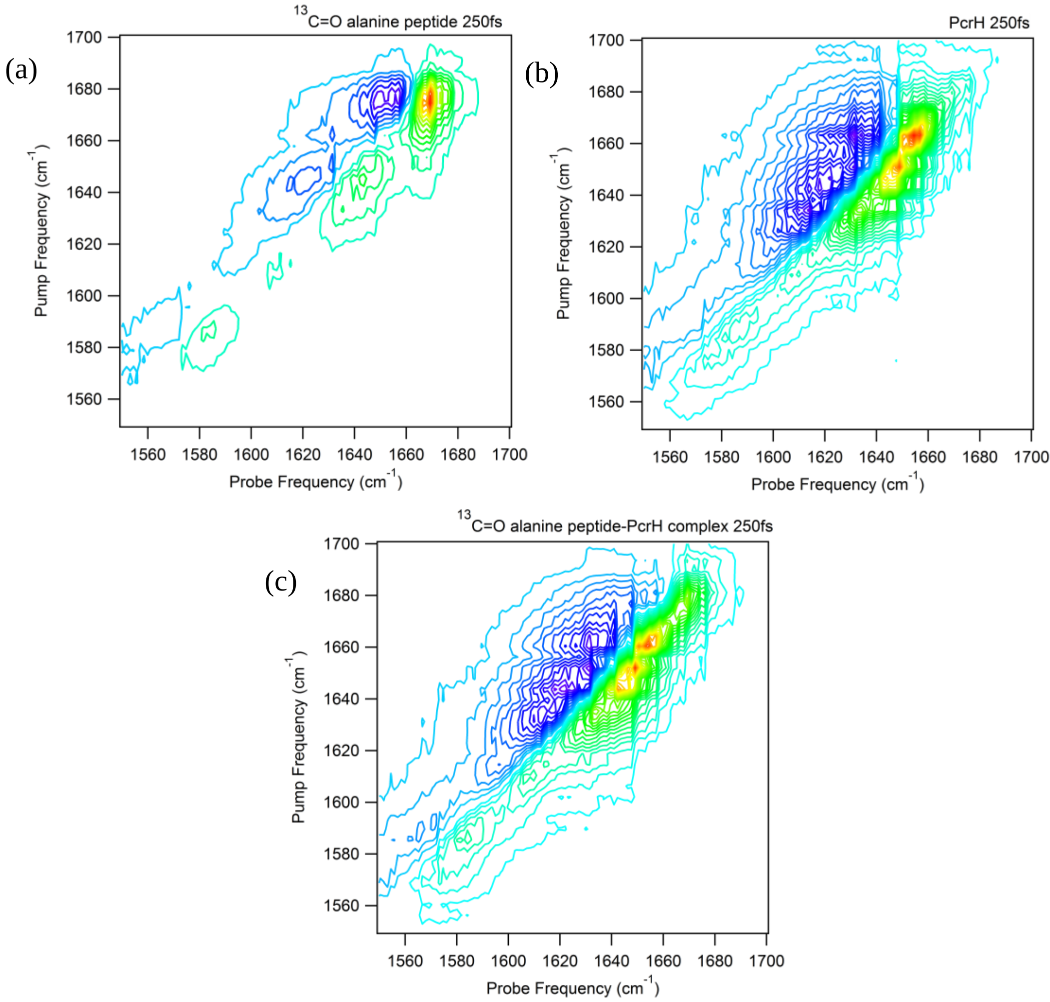


Figure (8): 2DIR spectra of (a) the isotopically-labelled peptide, ^{13}C-Ala $^{47-56}$PopD, (b) PcrH and the corresponding complex, all shown at waiting times of 250 fs.

Analysis of spectral diffusion is similarly affected by this convolution of bands, though in principle it is possible to subtract the spectral features of the overlapping modes using correctly-scaled reference spectra. However, correlation coefficients obtained after this subtraction had high associated uncertainty as a result of spectral lines from water vapour, and the on-diagonal effects of scattered light. Too few waiting times were measured to construct an approximate plot of spectral diffusion, and even at waiting times of 750 fs, the signal-to-noise ratio was seen to be too low to permit fitting, suggesting a short vibrational lifetime for the alanine mode.

Interestingly, off-diagonal signals were observed in pump slices of frequency ~1518 cm^{-1} though these pump slices are not included in the spectra above, owing to the fact that the probe detector was not configured to detect at these frequencies. Adjustment of the detector was not performed during the experiment as these weak off-diagonal signals were only recognised in retrospect, after a significant period of analysis.

The observed features differed prominently between PcrH, the wild-type peptide-PcrH complex and the isotope-labelled complex, as shown in Figure (9)(a). In free PcrH, off-diagonal signals occur at 1618 and 1650 cm^{-1}, the latter being far more intense. After inclusion of the wild-type peptide, the most intense off-diagonal feature appears at 1643 cm^{-1}, and additional features occur at 1578 and 1670 cm^{-1}. For the isotopically-labelled peptide, the off-diagonal features contain a peak at 1578 and 1643 cm^{-1}, though the 1670 cm^{-1} band is not observed.

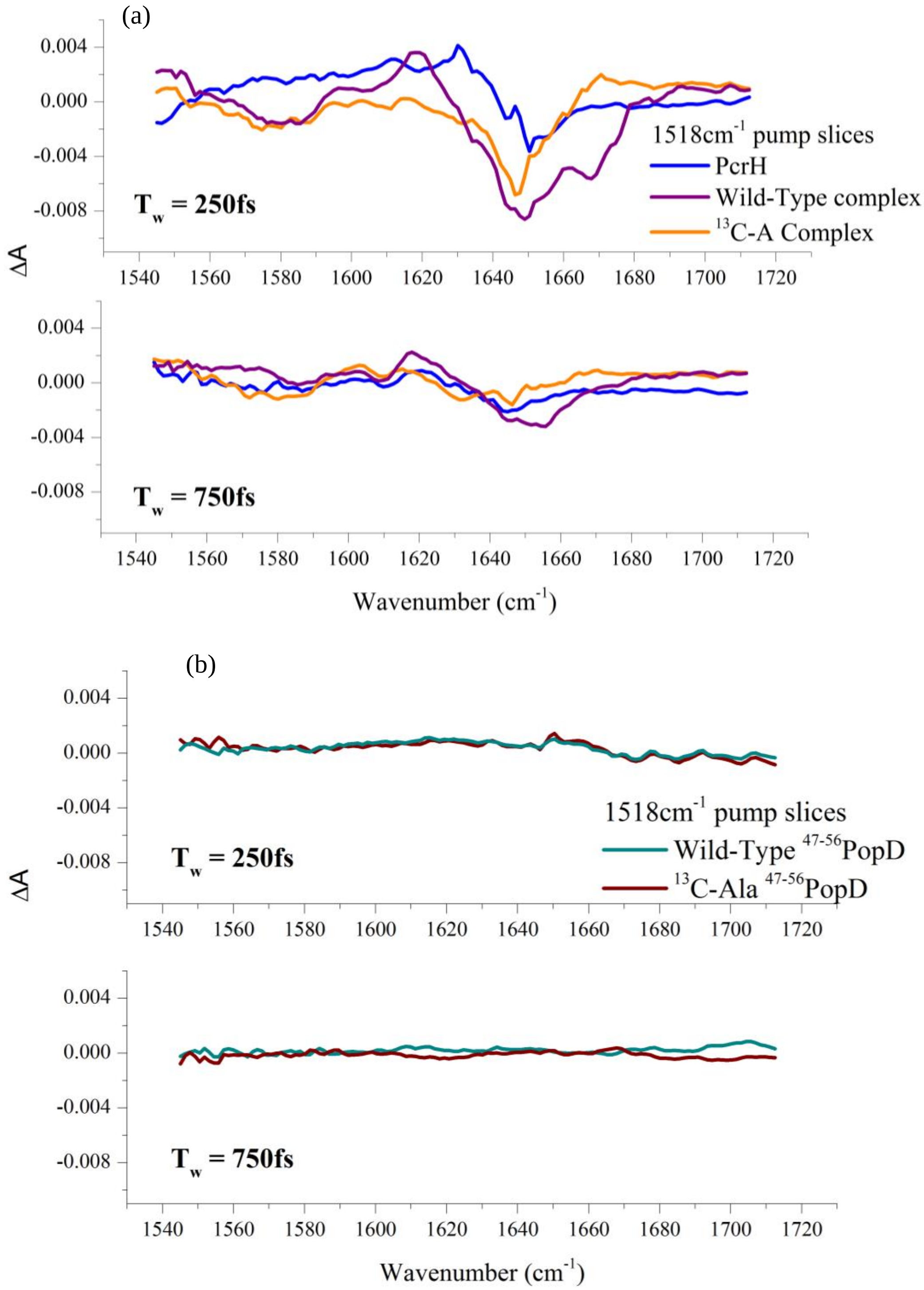


Figure (9): Pump-axis slices through the 2DIR spectra of: (a) free PcrH, the wild-type PcrH-peptide complex and the PcrH complex with ^{13}C-Ala $^{47\text{-}56}$PopD, and: (b) each of the free peptides, $^{47\text{-}56}$PopD and ^{13}C-Ala $^{47\text{-}56}$PopD. In each case, pump axis slices are shown at two waiting times.

The positions of off-diagonal peaks likely correspond to vibrations of carboxylate (1578 cm^{-1}), Amide I (1643 cm^{-1}) and side chain C=O stretches (1670 cm^{-1}) of the peptide[46]. Neither peptide species displays off-diagonal spectral features (Figure (14)(b)), suggesting that these peptide modes are excited through the vibration of a functional group of PcrH. Since the lower frequency cut-off of the probe detector in this study was ~1545 cm^{-1} it was not possible to resolve the corresponding on-diagonal 1518 cm^{-1} band. However, FTIR spectra of PcrH reveal a vibration at 1515 cm^{-1} (Figure (10)). This spectrally distinct peak is readily assignable to the symmetric C=C stretches of tyrosine rings, as studies by Barth[47] and Wolpert[48] conclude that other amino acid modes in the 1495- 1525 cm^{-1} region have negligibly low molar extinction coefficients. Assignment to the tyrosine ring mode was confirmed through measurement of the PcrH spectrum at pH 12.5. Deprotonation of the phenol ring in a free tyrosine molecule occurs with a pk$_a$ of 10.07, shifting the C=C stretching frequency to 1500 cm^{-1}. A similar response was observed in PcrH after pH increase; the 1515 cm^{-1} band is depleted and a new band is detected at 1501 cm^{-1} (Figure (10)).

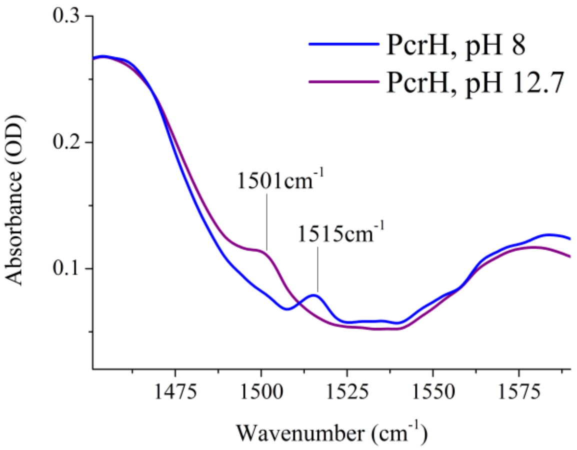


Figure (10): The tyrosine C=C stretching mode (1515 cm^{-1}/1501 cm^{-1}) of PcrH at pH 8 (blue) and pH 12.5 (purple).

Two tyrosine residues are located in the binding site of PcrH. Respectively, these residues; Tyr-40 and Tyr-47, are believed to form hydrogen bonds with the polar

PopD side chains Glu-50 and Asn-52[25] (Figure (11)). Coupling of the tyrosine OH groups with side chain C=O groups would give rise to an off-diagonal signal at 1670 cm^{-1}, as is observed. Sensitivity of the off-diagonal signals to the isotopic-labelling of alanine is explicable in the case that down-shifting of the alanine mode modifies dipole-dipole interactions in the complex, though interestingly, structural information obtained for the PopB-PcrH complex reveals direct hydrogen bonding between the backbone C=O of a proline residue and a PcrH tyrosine ring[24]. Structural comparison of the PopB and PopD chaperone complexes show that the H-bonded proline in PopB occupies a similar position to that of the labelled alanine residue in PopD. It is therefore possible that a tyrosine residue hydrogen bonds with both the backbone C=O of alanine and with the nearby Asn-52 side chain C=O, resulting in an isotope-sensitive coupling of the 1515 cm^{-1} vibration to 1670 cm^{-1}.

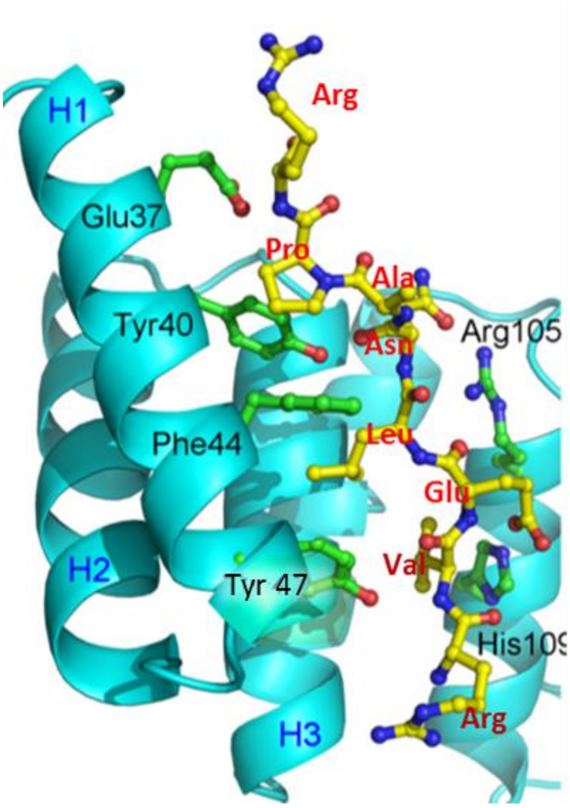


Figure (11): X-ray structure of $^{47\text{-}56}$*PopD within the PcrH binding domain. Peptide residues are labelled in red, PcrH residues are labelled in black. This figure has been modified from that published by Job et al.* [25]*; Tyr47 in the original publication was mislabelled as Thr47,* *though it is clear that a tyrosine ring is shown, and is further referred to in the article discussion*

Aspects of the off-diagonal information appear incongruous. If these signals are attributable to molecular coupling, then the spectral differences between the three 1518 cm^{-1} pump slices must be attributable to a maximum of two tyrosine moieties in the binding domain of PcrH (Tyr-40 and Tyr-47). The magnitude of these changes appears disproportionately large, assuming that all eight tyrosine rings have an initial contribution to the off-diagonal signal in free PcrH. The widths of each off-diagonal peak are also very large, approaching those of the linear spectrum. This is not necessarily unexpected, as each backbone C=O occupies only a relatively narrow position within the Amide I band, as demonstrated by the linear spectra of the three isotope-labelled peptides. To produce such broad off-diagonal features, the tyrosine modes would be required to couple with several backbone C=O groups as well as polar side chains of the peptide. In the absence of strong transition dipole coupling along the peptide backbone, this is unlikely to occur.

However, the shortcomings of isotope-labelling in this system warrant a revaluation of the experimental approach, and sufficient physical basis exists for the vibrational coupling of PcrH tyrosine residues with PopD. Indeed, mutation studies have revealed that the residue Tyr-40 may be linked with substrate binding affinity[22]. This may also indicate the role of Tyr-40 in the structural recognition mechanism of PcrH. Further, the tyrosine mode is intrinsically well-separated from other peptide absorptions, and is a native component of the protein. For this reason, the tyrosine mode may be capable of serving as a local probe without any structural modification of the system, ensuring that the native state of the complex is conserved.

5.4.5 Mutant peptides and PcrH: FTIR

Tyrosine coupling patterns in the wild-type complex appears to involve the peptide 1672 cm^{-1} C=O stretching modes, potentially through a hydrogen bonding interaction. It is unclear whether this interaction occurs from Tyr-40, Tyr-47 or a combination of the two. Each tyrosine residue is further capable of hydrogen bonding to a peptide side chain C=O. Tyr-40 may interact with the asparagine reside (Asn-52) of the peptide, while Tyr-47 may interact with the peptide's glutamine residue (Glu-50). In order to determine the residue-specific nature of the tyrosine coupling, two mutant species of the peptide were produced. By substituting the polar Asn and Glu residues with a hydrophobic alanine, hydrogen bonding to the tyrosine ring could be systematically restricted, giving further specificity to the resulting cross-peak patterns.

The term "mutant" is typically encountered in reference to full protein sequences, though it is adopted here due to the role of $^{47\text{-}56}$PopD as an analogue of PopD; implementing the same structural modifications described above, for PopD, would qualify as "mutation", and thus this nomenclature is extended to $^{47\text{-}56}$PopD.

	47	48	49	50	51	52	53	54	55	56
Wild-type	D	R	V	E	L	N	A	P	R	Q
E50A	D	R	V	A	L	N	A	P	R	Q
N52A	D	R	V	E	L	A	A	P	R	Q

Figure (12): The amino acid sequences of $^{47\text{-}56}$PopD and the two mutants, E50A and N52A, with the non-native residues marked in red.

The primary structure of each peptide mutant is depicted in Figure (12) with the numbering system derived from the complete PopD sequence. From this, mutant species are assigned the references E50A and N52A. Linear spectra of the mutants are shown in Figure (13), with the spectrum of the wild-type peptide included as a reference.

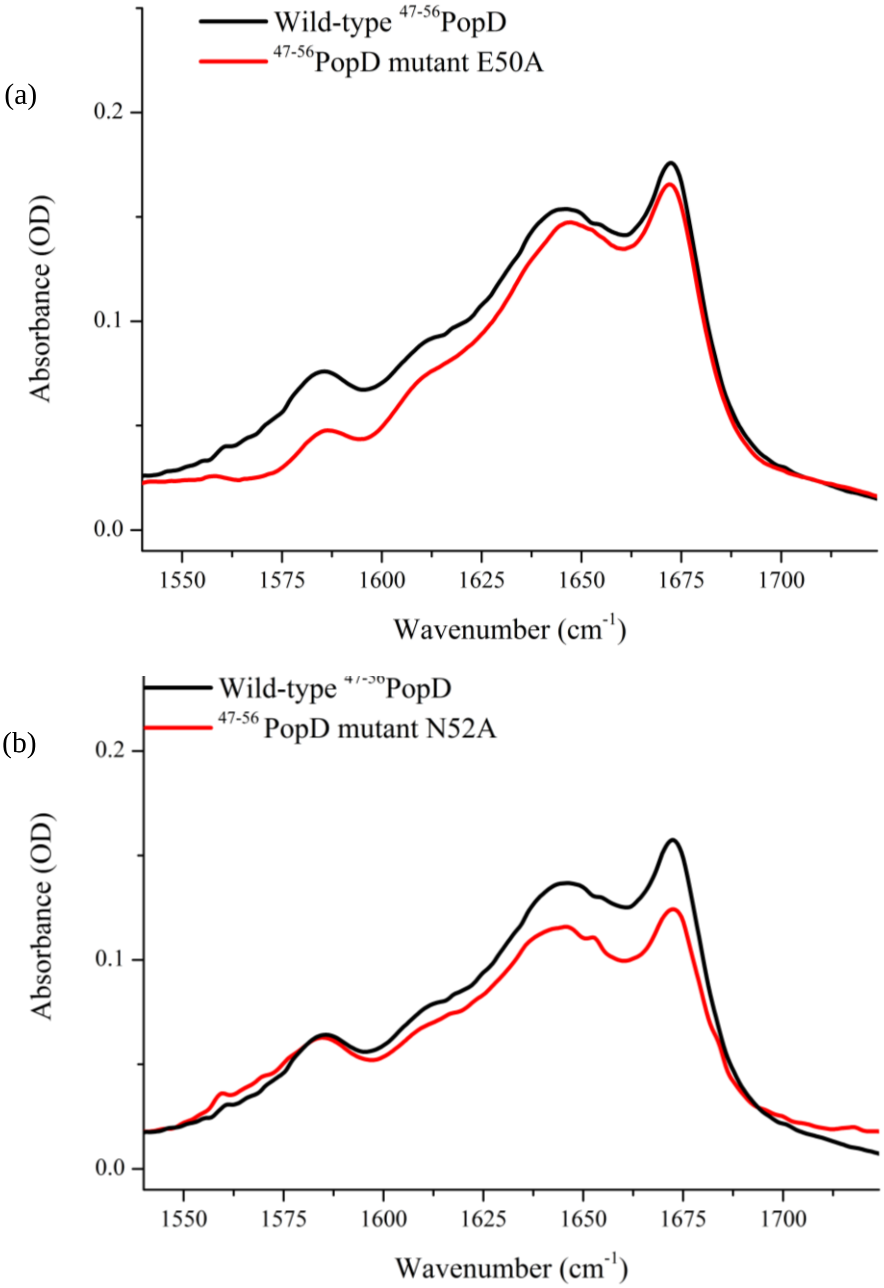


Figure (13): The linear IR spectrum of $^{47\text{-}56}$PopD, relative to that of (a) the E50A peptide mutant and (b) the N52A peptide mutant

Glutamic acid (E) exhibits strong absorption around 1572, 1609, and 1670 cm^{-1}. The vibrations of alanine are significantly weaker, and thus the E50A mutant has lower overall optical density than the wild-type peptide. Likewise, asparagine has strong bands around 1620 and 1680 cm^{-1}, contributing to a lower optical density upon replacement with alanine in N52A.

As the reporting tyrosine band is spectrally well-isolated, the overlapping PcrH and peptide bands in the 1580-1700 cm^{-1} range are no longer detrimental to 2DIR measurements. In reality, several side chain absorptions do occur close to 1515 cm^{-1}, such as the phenylalanine band at 1498 cm^{-1} or the alanine band at 1520 cm^{-1}, the molar extinction coefficients of these vibrations are very low (~50 M^{-1}cm^{-1}) in comparison to that of the tyrosine stretch (~500 M^{-1}cm^{-1}) to the point that they are unresolved at any PcrH concentration used[47,48].

5.4.6 Mutant peptides and PcrH: ITC

Binding of the E50A and N52A mutants to PcrH was measured using the same ITC parameters as those of the wild-type peptide. The resulting titration curves are shown in Figure (14). Dissociation constants of 0.8 mM and 1.3 mM were obtained for the E50A and N52A peptide, respectively.

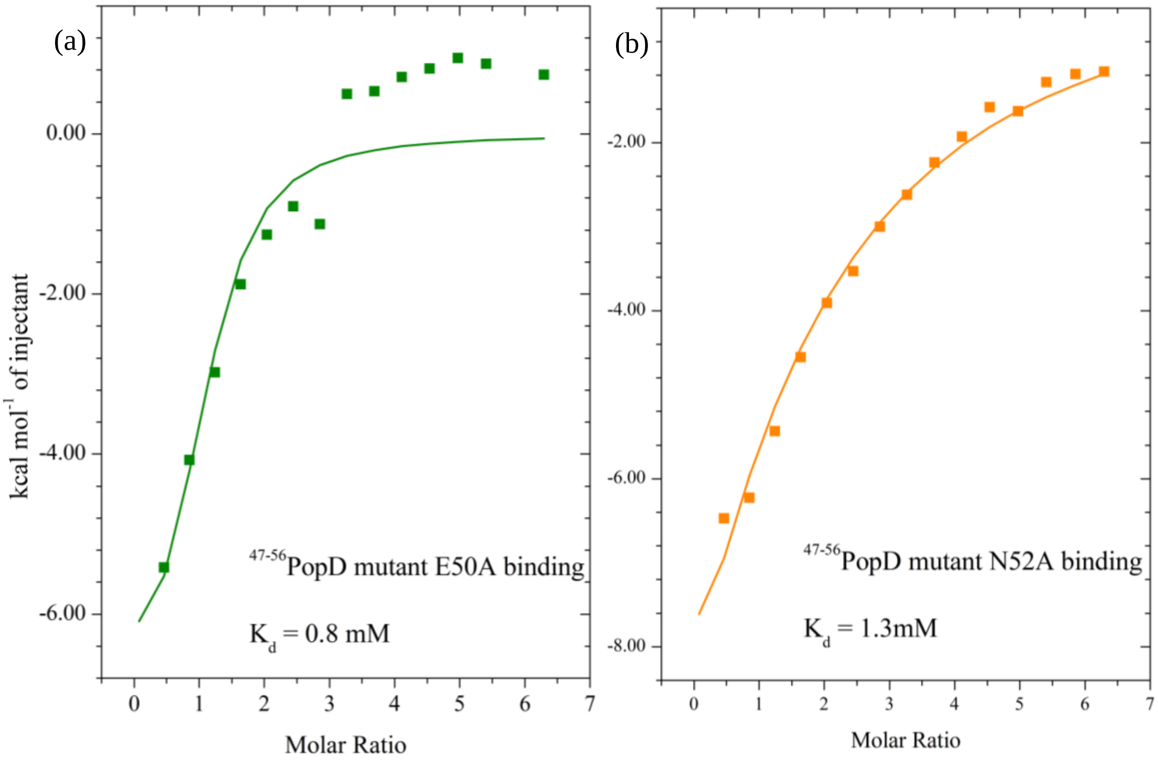


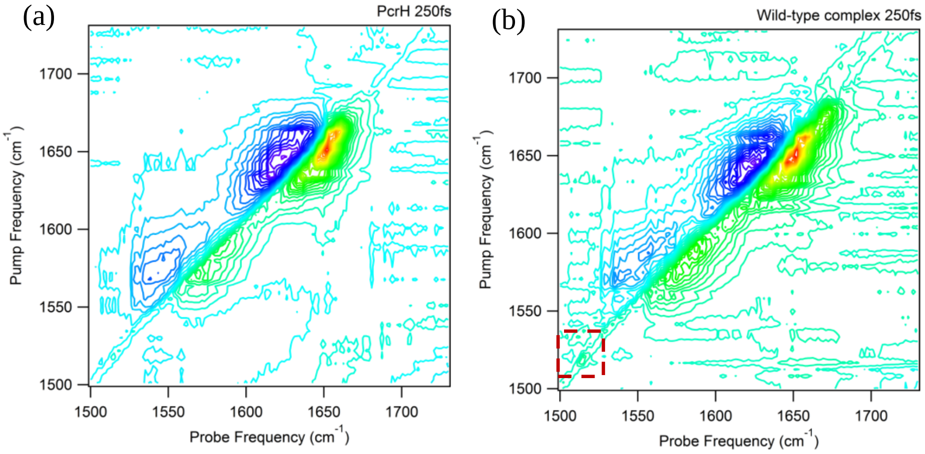
Figure (14): Titration curves for the binding interaction of PcrH with the mutant peptides (a) E50A and (b) N52A. Each plot is annotated with the extracted dissociation constant

It has been reported that the modification of polar residues in the PopD binding region does not affect complex formation with PcrH[25]. From these results, it is apparent that both the E50A and N52A mutants undergo a similarly weak binding with the protein as the wild-type peptide, though the lack of saturation in complex formation creates significant uncertainty in the retrieved K_d values. It therefore cannot be stated firmly that substitution of polar residues results has no effect whatsoever on the dissociation constant of the system, though it is clear that binding is not altogether inhibited by mutation.

With the objective of maximising the bound protein concentration for IR studies, the peptide K_d was taken as the largest value measured in the course of these ITC experiments, i.e. the 2.4 mM measured for the wild-type complex.

5.4.7 Mutant peptides and PcrH: 2DIR

2DIR spectra of PcrH and its complexes with wild-type, E50A and N52A peptides are shown in Figure (15) at waiting times of 250 fs. The on-diagonal C=C stretch of tyrosine appear at (ω_{probe} = 1515 cm^{-1}, ω_{pump} = 1515 cm^{-1}), and is faintly observable in all spectra excluding PcrH. A dashed red line is included in the wild-type complex spectrum to highlight the mode. The ESA of tyrosine is poorly defined however, and its Bleach overlaps with a pattern of on-diagonal scatter.



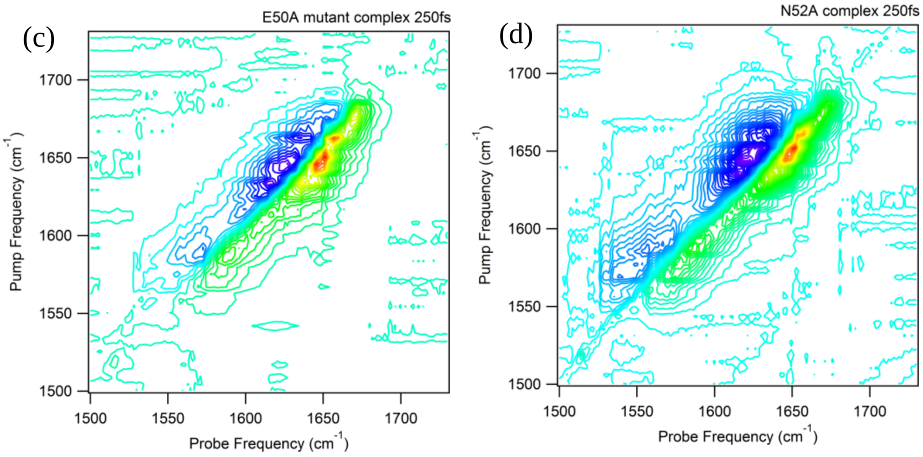


Figure (15): Representative 2DIR of (a) free PcrH, (b) the wild-type PcrH $^{47\text{-}56}$*PopD complex, (c) the PcrH-E50A mutant complex and (d) the PcrH-N52A mutant complex, all shown at 250 fs. The dashed red box in (b) marks the position of the relatively weak tyrosine mode.*

Scatter of this type often originates poorly dissolved or precipitated material in the laser path, which may scatter pump light into the probe beam, creating interference. As the pump beam is significantly more intense than the probe, even weakly-scattered light is detected in the probe spectrum. This manifests as an on-diagonal stripe as the probe light is interfered with at the frequency of the pump at that particular stage position, i.e. at $\omega_{\text{probe}} = \omega_{\text{pump}}$. This is most apparent in the wild-type complex spectrum in Figure (15).

The origin of the scattered light is significant in this case; unlike prior experiments, the stability of the protein here was seen to be low, causing significant precipitation after several hours. Protein precipitation is typically associated with aggregation, which becomes more likely at higher concentrations[53]. Under these conditions, further impetuses such as temperature or pH change can induce rapid sample degradation. High protein concentrations are required for IR spectroscopy, relative to other methods UV-Vis or NMR, and thus sample integrity may be temperamental.

The reason for precipitation in this experiment is unclear, though it indicates the potential existence of a major experimental obstacle for future studies. Meaningful study of PcrH stability could not be achieved with the limited quantity of remaining protein, or within the limited time available.

Solubility of the wild-type and N52A was further seen to be poor at the concentrations required to maximise the quantity of bound protein. For these reasons, the measurements were interrupted by the precipitation of the protein and a re-evaluation of the protocol. The E50A mutant complex was successfully measured at “high” concentration (2.5 mM), though spectra of the remaining three samples were obtained at lower PcrH concentrations (450 μM). The peptide concentration was proportionately reduced. The 2DIR spectra above are therefore obtained under two different sets of conditions, though the signal-to-noise ratio apparent in the spectrum contours does not immediately reflect the difference. This will be discussed in more detail later, however.

Pump frequency slices ~1515 cm^{-1} reveal a well-resolved on-diagonal tyrosine band with a sharp, narrow $\nu = 0\text{-}1$ at 1515 cm^{-1} and a broad, lower-intensity $\nu = 1\text{-}2$ transition at 1505 cm^{-1}. Figure (16) displays pump slices for each of the samples at 250 fs. The tyrosine mode remains clearly identifiable in low concentration spectra.

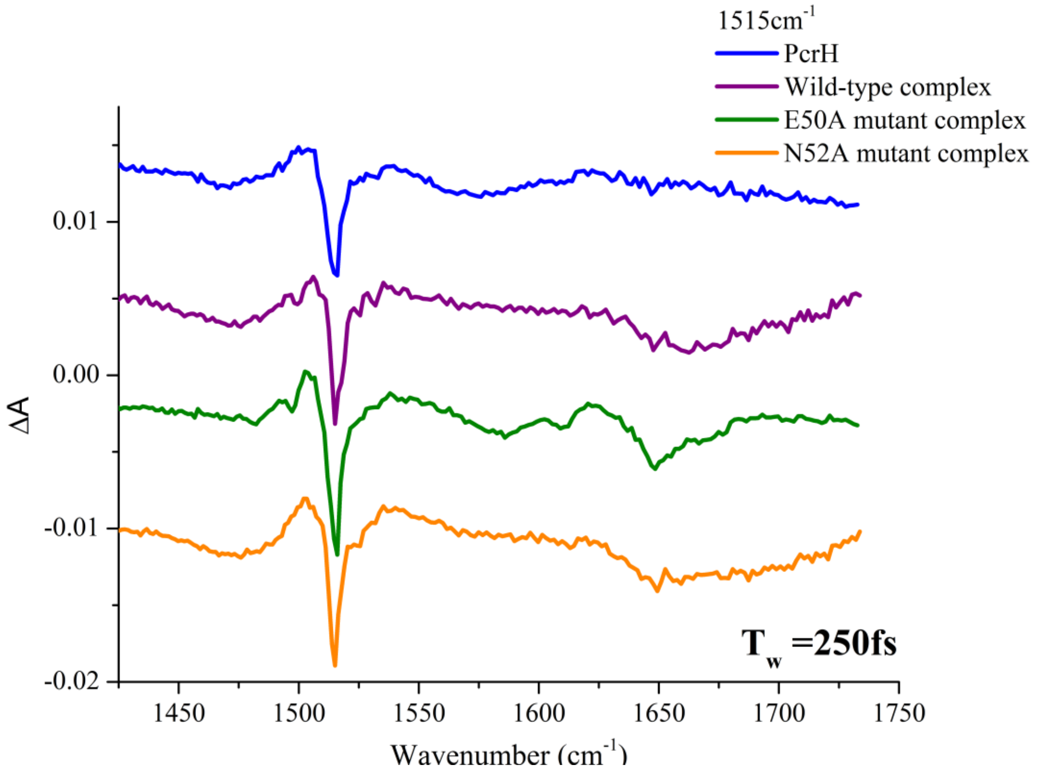
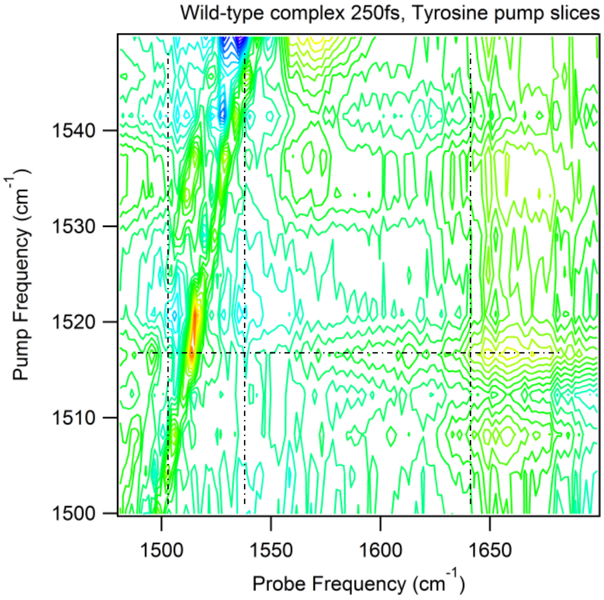


Figure (16): Pump-axis slices of PcrH, its wild-type complex and the two mutant complexes at 250 fs. Pump slices are shown at the tyrosine C=C stretch frequency of 1515 cm^{-1}

Each of the protein-peptide complexes has a weak, broad feature in the off-diagonal region, of central frequency ~1650 cm^{-1}. This is not seen in the spectrum of free PcrH, though a comparably broad feature ~1580 cm^{-1} appears in this pump slice. The E50A complex appears to have both of the described signals and another, smaller signal at 1610 cm^{-1}. Critically, outwith the E50A-protein spectrum, no clear ESA bands can be observed for any of these off-diagonal signals.

When taking into account the weak binding of peptide to protein, and the sub-population of tyrosine modes that should contribute to the mutation-sensitive cross-peaks, the off-diagonal signal sizes in the E50A spectrum appear disproportionately large. Although these features appear at the probe frequencies of genuine peaks, it was observed that these off-diagonal features could be seen across multiple pump slices far from the tyrosine band. These off-diagonal signals were observed across the full pump range of each spectrum, though they are most apparent at lower frequency owing to the low intensity of the on-diagonal signals. The signals were not of uniform intensity across the spectrum however, and exhibited maxima around the 1515 cm^{-1} pump frequency, as shown by Figure (17) (a). This gives the appearance of cross-peaks, though the signals are essentially vertical streaks emanating directly from real on-diagonal peaks. In certain cases, the off-diagonal signal bore close resemblance to an imprint of the linear absorption profile on the probe spectrum (Figure (17) (b)). It is possible that few of the off-diagonal features display ESA bands for this reason. This was not consistently observed however and evidently subject to variation from other unknown factors.

(a)



(b)

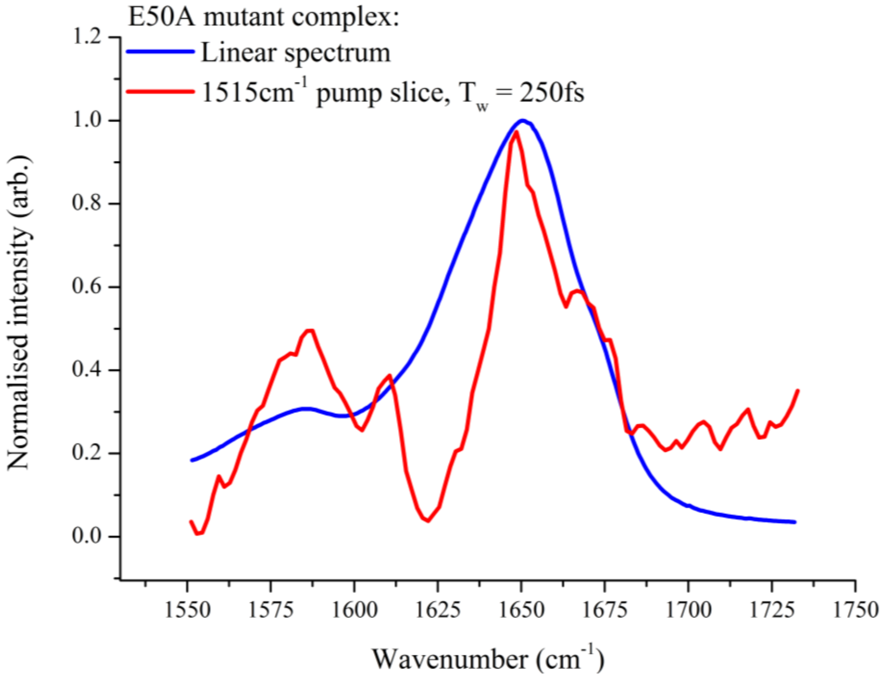


Figure (17): (a) A close-up of the 2DIR baseline for the wild-type complex. Truncation of the pump axis removed large on-diagonal signals, altering the contour scaling and allowing artefacts to be more clearly resolved. Dashed lines (vertical) separate regions of positive andnegative baseline amplitude and (horizontal) indicate an anomalous peak in the baseline near the 1515 cm^{-1} pump slice. (b) Comparison of the linear spectrum of the E50A mutant complex (blue) with the off-diagonal signal observed in the 1515 cm^{-1} pump slice (red) of the E50A complex 2DIR spectrum (at 250 fs).

Off-diagonal signals previously reported for the isotope-labelled complexes may also be a consequence of the same effect. This would justify the large, broad features observed in the pump axis frequency slices.

Genuine coupling patterns from the tyrosine ring mode are expected to be very weak, and would likely be convoluted in these artefacts. It is therefore not possible to state whether such vibrational coupling truly occurs, though certain aspects of the data are not wholly inconsistent with this hypothesis; off-diagonal maxima are observed near the 1515 cm^{-1} pump frequency slice, potentially as the result of a cross-peak underlying the artefact signal. Without a genuine off-diagonal feature, it is difficult to justify this occurrence.

Approximate proportionality between the artefact intensity and on-diagonal intensity, i.e. the OD of the Amide I (1650 cm^{-1}) and carboxylate peak (1580 cm^{-1}) appears to create a limitation for this system; the minimum off-diagonal signal size that can be resolved is dependent on the largest on diagonal signal. The combined Amide I bands of a protein and polypeptide gives an optical density at the upper measurable limits of 2DIR. Simultaneously, the tyrosine band lies at the opposite limit. This is not a common scenario in 2DIR studies, and could highlight a limitation in the method at present. With technological improvements however, the practicality of this approach may become more favourable.

5.5 Discussion and conclusions

Chaperone-protein systems are fundamentally difficult to simplify. In this case the substitution of PopD for a smaller structural analogue is possible, though many aspects of the resulting system seem disparate from the *in vivo* behaviour of the complex.

In its native state, PopD occupies a molten globular conformation[25,44]. Molten globule structures are generally classified as having secondary structure, but undefined, transient tertiary structures[54] that are not stabilised by particular side chain-to-side chain interactions that would otherwise define a “native” conformation. PopD has been shown to exhibit a chiefly α-helical secondary structure, and its poorly-defined tertiary conformation is believed to assist oligomerisation by allowing the intermolecular interactions of flexible, solvent-exposed hydrophobic residues. *In vivo*, PopD is expressed within bacteria cells in the presence of PcrH[21]. From this, the chaperone protein has the opportunity to pre-empt aggregation of the translocator proteins through early complex formation.

Both the PcrH-bound and oligomer states that PopD can occupy *in vivo* are stabilised through primarily hydrophobic interactions, and further, each state allows PopD to maintain an α-helical secondary structure. The fact that the dislocated binding domain of the translocator may still form a complex with PcrH strongly suggests that PopD’s hydrophobic binding residues, valine-49, leucine-51 and alanine-53 have exposed side chains. Unlike the full PopD sequence however, the sub-domain is unlikely to form the ring-like oligomer associated with the translocon. It is therefore proposed that in a solution-phase system containing only $^{47\text{-}56}$PopD, the hydrophobic force results in alternate oligomer or polymer structures that may be denatured with respect to the conformation required to bind with PcrH. This is supported by the linear spectrum of a gelled peptide sample, as shown in Figure (4). With respect to the solution-phase peptide, the gel-phase sample shows markedly different absorption bands. The central frequency of the Amide I is not observed at 1650 cm^{-1} after gelation, but rather occurs at 1634 cm^{-1}. Additional features are also observed after gel-formation, at 1616, 1624 and 1710 cm^{-1}. In D_2O, the Amide I frequency associated with β-sheet secondary structure lies in the range 1615-1638 cm^{-1}, with an average value of 1630 cm^{-1}. The spectral changes associated with the phase-change may therefore also indicate that the gelled structure is formed through hydrogen bonding with a β-sheet configuration.

Throughout these measurements, small variations in peptide spectra were commonly seen to occur from the same lyophilised sample. In particular, when normalised to the 1672 cm^{-1} band, the comparative intensity of the 1650 cm^{-1} peak was seen to vary, as was its central frequency. Likewise, the 1710 cm^{-1} amplitude was not consistent across measurements. This could indicate the formation of oligomers at far lower concentrations than required for complete gelation, with resultant spectral changes originating from the modification of hydrogen bond environments for both backbone and side chain vibrations. This aggregation may constitute a denaturing of the peptide with respect to chaperone-binding, with the resulting $^{47\text{-}56}$PopD structure or structures being poorly recognised by the PcrH binding site. This would also be consistent with the unexpectedly weak binding constant of the chaperone-peptide complex.

The low binding affinity of PcrH-peptide complexes is counter-intuitive to the natural role of PcrH as a chaperone protein. Further, the unfavourable binding casts uncertainty over whether the conformational dynamics of the chaperone binding site in the PcrH-$^{47\text{-}56}$PopD complex would be an accurate reflection of those that occur in the full complex, PcrH-PopD. Although no specific K_d has yet been published, PcrH has been described as a highly efficient binder of both PopB and PopD[21]. This observation is made based on the purification of complexes formed from bicistronic expression of translocator proteins and the chaperone in the cytoplasm of *E. Coli*. This co-expression is a better representation of the native protein conditions *in vivo*, though it involves synthesis of the full PopD sequence, preventing an X-ray crystal structure from being obtained.

A co-expressed complex is also problematic for Infrared Spectroscopy, as PopD is a large protein (~35 KDa, relative to ~18 KDa for PcrH) with an inherently complex, disordered structure, for which the solubility is likely poor. This may prevent the fully-structured complex from being concentrated to an easily measurable level. Further, *E coli.* would unable to express modified peptides, for example, azide, cyano or $^{13}C=^{18}O$ labelled, which limits flexibility in experimental design. Determination of the dissociation constant for a complex produced by co-expression is also not possible, though a method such as fluorescence spectroscopy may able to utilise the intrinsic fluorescence of tyrosine rings in the PcrH binding pocket to determine the relative concentrations of bound and unbound chaperone species, provided that fluorescent quenching occurs upon binding[55]. A fluorescence study of this type may also have an IR analogue in which the shifts of the tyrosine C=C peak centre reveal H-bond interactions that correspond to binding. The data acquired here appears to reflect very small (0.5 cm^{-1}) shifts in the tyrosine frequency between PcrH, and its wild-type, E50A and N52A complexes, though the effect may be associated to small variations in the intensity of water vapour peaks that lie close to the weak tyrosine mode. A more rigorous approach to the collection of FTIR spectra would permit the use of 2nd-derivative analysis of the spectra, potentially allowing a higher degree of sensitivity to sub-states within the tyrosine band[45,56,57]. Such an approach has proven successful in the study of tyrosine hydrogen-bonding configurations in biological case-studies[58,59].

DMSO was employed to obtain the X-ray structures of PcrH and PopD discussed here (PDB code 2XCB). The solvent may create a more favourable binding scenario, though, as discussed, this presents a challenge for isothermal titration calorimetry. It is also unclear whether DMSO alters the native conformation of the system, an aspect that may be difficult to determine *via* X-ray crystallography, as crystallisation of the complex may not be possible without the use of DMSO, owing to the molten globular structure of [47-56]PopD.

In other systems, ligand binding may cause significant conformational changes, with a resultant impact on the protein's Amide I band[60]. Such an approach is not practical here. Circular dichroism detects few changes associated with complex formation, such that the linear superposition of PcrH and PopD's CD spectra are almost indistinguishable from that of the complex[21]. This indicates that the secondary structures of both molecules are unchanged by binding. The unavoidably high concentrations of unbound protein and peptide would also diminish the intensity of Amide I changes. Further, the overlap of Amide I bands at ~1650 cm^{-1} would ultimately lead to difficulty in assigning small spectral changes to a particular molecule.

P. aeruginosa is a robust bacterium with high antibiotic resistance and the capacity to survive in both radioactive[61] and micro-gravity environments[62,63]. Despite this, a single point-mutation of the translocator chaperone, PcrH, inhibits complex formation with PopB and PopD, causing rapid degradation of the organism[22]. This indicates that the approach taken to maintain native characteristics of the system have been unsuccessful in this study, most likely due to the sequence of molecular interactions that occur during sample preparation; aggregation of the peptide is likely to preclude complex formation. In the native system, the order of these events must fundamentally be reversed. However, preservation of the native state and compatibility with IR methods are closely-linked here. A sample preparation method that more closely observes the biological conditions of the PcrH-PopD interaction is likely to give rise to complex formation with a far higher affinity, which provides a more favourable scenario for the development of new IR approaches.

5.6 Future work

Without further modification of the sample preparation protocol, unnatural amino acid labels offer the greatest potential for extraction of site-specific information. By modifying the alanine mutation carried by the E50A and N52A peptides to an azido or homo-alanine species, the spectral window of the probe can be shifted to a region where only the probe vibration occurs. This would eliminate a number of experimental constraints. Firstly, concentrations of the protein and peptide would no longer be limited by the Amide I modes, whose strong absorption can readily attenuate the laser pulse to such a low intensity that it cannot be reliably measured by the detector. The concentration of the system could therefore be raised. Secondly, at higher concentrations, the signal-to-noise ratio of peaks associated with bound or unbound forms of the peptide would become larger, allowing clearer distinction of each state. Thirdly, cyano and azido groups have demonstrated sizable shifts through changes in environment, perhaps allowing a clear separation of bound and unbound peptide species.

Polarisation-dependent 2DIR measurements are capable of altering the ratio of on-diagonal to off-diagonal intensities. By altering the relative polarisations of pulses within the 2DIR pulse sequence, the Feynman pathways explored by the molecular ensemble are altered. This can be employed to supress on-diagonal transitions while enhancing those that give rise to off-diagonal peaks. The use of tyrosine residues as local probes of the binding region is likely to become more favourable with this approach, as it effectively allows the enhancement of coupling patterns originating from the weak tyrosine vibration, while decreasing the intensity of Amide I modes.

Ultimately, optimisation of the binding affinity is desirable. This may be achievable through the inhibition of peptide aggregation. The use of small concentrations of DMSO may be beneficial in this capacity, and it is possible that a titration experiment using fluorescence spectroscopy or 2nd-derivative FTIR may be able to provide adequate information on complex formation, in place of ITC.

Substitution of the PopD analogue for the PopB binding domain may also lead to a decrease in peptide aggregation, thereby increasing the favourability of protein-binding. The translocon-chaperone interaction in *P. aeruginosa* is certainly not unique to this bacterium, and indeed appears analogous to other pathogens with comparable toxin secretion systems. Establishment of a successful experimental approach therefore sets a precedent for the investigation of many other protein-chaperone interactions. Further, it will constitute a major progression of 2DIR applications to biological systems, from studies of simple peptides structures to large protein-protein complexes.

5.7 Acknowledgments

Special thanks to our collaborators, Dr Martin Walsh and Dr Domenico Bellini of the Macromolecular Crystallography Group at the Diamond Light Source for their supervision on this project and for the preparation of all PcrH samples. Many thanks also to Morag Watson and Dr Craig Jamieson of the University of Strathclyde, for their instruction on peptide synthesis and for the use of their laboratory.

5.8 References

1. Hartl, F. U., Bracher, A. & Hayer-Hartl, M. Molecular chaperones in protein folding and proteostasis. *Nature* **475,** 324–332 (2011).
2. Frydman, J. Folding of newly translated protein in vivo: the chaperones. *Annu. Rev. Biochem.* **70,** 603–647 (2001).
3. Aguzzi, A. & O’Connor, T. Protein aggregation diseases: pathogenicity and therapeutic perspectives. *Nat. Rev. Drug Discov.* **9,** 237–248 (2010).
4. Stefani, M. Protein misfolding and aggregation: new examples in medicine and biology of the dark side of the protein world. *Biochim. Biophys. Acta* **1739,** 5–25 (2004).
5. Naganathan, A. N. & Muñoz, V. Scaling of folding times with protein Size. *J. Am. Chem. Soc.* **127,** 480–481 (2005).
6. Saibil, H. Chaperone machines for protein folding, unfolding and disaggregation. *Nat. Rev. Mol. Cell Biol.* **14,** 630–642 (2013).
7. Priya, S., Sharma, S. K. & Goloubinoff, P. Molecular chaperones as enzymes that catalytically unfold misfolded polypeptides. *FEBS Lett.* **587,** 1981–1987 (2013).
8. Whitley, D., Goldberg, S. P. & Jordan, W. D. Heat shock proteins: a review of the molecular chaperones. *J. Vasc. Surg.* **29,** 748–751 (1999).
9. Wyatt, A. R., Yerbury, J. J., Ecroyd, H. & Wilson, M. R. Extracellular chaperones and proteostasis. *Annu. Rev. Biochem.* **82,** 295–322 (2013).

10. Rose, G. D., Geselowitz, A. R., Lesser, G. J., Lee, R. H. & Zehfus, M. H. Hydrophobicity of amino acid residues in globular proteins. *Science* **229,** 834–838 (1985).

11. Muñoz, V., *Protein folding, misfolding and aggregation: classical themes and novel approaches*. (Royal Society of Chemistry, 2008).

12. Bross, P. *et al.* Protein misfolding and degradation in genetic diseases. *Hum. Mutat.* **14,** 186–198 (1999).

13. Wickner, S., Maurizi, M. R. & Gottesman, S. Posttranslational quality control: folding, refolding, and degrading proteins. *Science* **286,** 1888–1893 (1999).

14. Carpenter, E. P., Beis, K., Cameron, A. D. & Iwata, S. Overcoming the challenges of membrane protein crystallography. *Curr. Opin. Struct. Biol.* **18,** 581–586 (2008).

15. Ostermeier, C. & Michel, H. Crystallisation of membrane proteins. *Curr. Opin. Struct. Biol.* **7,** 697–701 (1997).

16. Fan, H. & Mark, A. E. Relative stability of protein structures determined by X-ray crystallography or NMR spectroscopy: a molecular dynamics simulation study. *Proteins* **53,** 111–120 (2003).

17. Ghosh, P. Process of protein transport by the type III secretion system. *Microbiol. Mol. Biol. Rev.* **68,** 771–795 (2004).

18. Tseng, T.-T., Tyler, B. M. & Setubal, J. C. Protein secretion systems in bacterial-host associations, and their description in the gene ontology. *BMC Microbiol.* **9,** 1–9 (2009).

19. Saier, M. H. Protein secretion systems in gram-negative bacteria. *Microbe* **1,** 414–419 (2006).

20. Schraidt, O. *et al.* Topology and organization of the salmonella typhimurium type III secretion needle complex components. *PLoS Pathog.* **6,** 3–12 (2010).

21. Schoehn, G. *et al.* Oligomerization of type III secretion proteins PopB and PopD precedes pore formation in Pseudomonas. *EMBO J.* **22,** 4957–4967 (2003).

22. Bröms, J. E., Edqvist, P. J., Forsberg, Å. & Francis, M. S. Tetratricopeptide repeats are essential for PcrH chaperone function in Pseudomonas aeruginosa type III secretion. *FEMS Microbiol. Lett.* **256,** 57–66 (2006).

23. Collier, R. J. Salmonella's safety catch. *Science* **328,** 981–982 (2010).

24. Discola, K. F. *et al.* Membrane and chaperone recognition by the major translocator protein popB of the type III secretion system of pseudomonas aeruginosa. *J. Biol. Chem.* **289,** 3591–3601 (2014).

25. Job, V., Matteï, P.-J., Lemaire, D., Attree, I. & Dessen, A. Structural basis of chaperone recognition of type III secretion system minor translocator proteins. *J. Biol. Chem.* **285,** 23224–23232 (2010).
26. Mueller, C. A., Broz, P. & Cornelis, G. R. The type III secretion system tip complex and translocon. *Mol. Microbiol.* **68,** 1085–1095 (2008).
27. Hueck, C. J. Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiol. Mol. Biol. Rev.* **62,** 379–433 (1998).
28. Blatch, G. L. & Lässle, M. The tetratricopeptide repeat: A structural motif mediating protein-protein interactions. *BioEssays* **21,** 932–939 (1999).
29. Cerveny, L. *et al.* Tetratricopeptide repeat motifs in the world of bacterial pathogens: role in virulence mechanisms. *Infect. Immun.* **81,** 629–635 (2013).
30. Thielges, M. C., Chung, J. K. & Fayer, M. D. Protein dynamics in cytochrome P450 molecular recognition and substrate specificity using 2D IR vibrational echo spectroscopy. *J. Am. Chem. Soc.* **133,** 3995–4004 (2011).
31. Fuchs, J. E., Grafenstein, S. von, Huber, R. G., Wallnoefer, H. G. & Liedl, K. R. Specificity of a protein-protein interface: Local dynamics direct substrate recognition of effector caspases: Local Dynamics Direct Specificity of Caspases. *Proteins Struct. Funct. Bioinforma.* **82,** 546–555 (2014).
32. Masterson, L. R. *et al.* Dynamics connect substrate recognition to catalysis in protein kinase A. *Nat. Chem. Biol.* **6,** 821–828 (2010).
33. Boehr, D. D., Nussinov, R. & Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* **5,** 789–796 (2009).
34. Chan, W. C. & White, P. D. *Fmoc solid phase peptide synthesis: a practical approach*. (Oxford University Press, 2000).
35. Dunn, B. M. & Pennington, M. W. *Peptide analysis protocols*. (Humana Press, 1994).
36. Pennington, M. W. & Dunn, B. M. *Peptide synthesis protocols*. (Humana Press, 1994).
37. Mukherjee, P., Kass, I., Arkin, I. T. & Zanni, M. T. Picosecond dynamics of a membrane protein revealed by 2D IR. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 3528–3533 (2006).
38. Wang, L. *et al.* 2DIR spectroscopy of human amylin fibrils reflects stable β-sheet structure. *J. Am. Chem. Soc.* **133,** 16062–16071 (2011).
39. Moran, S. D. *et al.* Two-dimensional IR spectroscopy and segmental ^{13}C labeling reveals the domain structure of human γD-crystallin amyloid fibrils. *Proc. Natl. Acad. Sci.* **109,** 3329–3334 (2012).

40. Falvo, C. *et al.* Frequency distribution of the amide-I vibration sorted by residues in amyloid fibrils revealed by 2D-IR measurements and simulations. *J. Phys. Chem. B* **116,** 3322–3330 (2012).

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

42. Greetham, G. M. *et al.* ULTRA laser system: a new dual-output 10 kHz Ti:sapphire amplifier with UV–IR generation for time-resolved spectroscopy. *Cent. Laser Facil. Annu. Rep.* 249–250 (2008).

43. Greetham, G. M. *et al.* ULTRA: a unique instrument for time-resolved spectroscopy. *Appl. Spectrosc.* **64,** 1311–1319 (2010).

44. Faudry, E., Job, V., Dessen, A., Attree, I. & Forge, V. Type III secretion system translocator has a molten globule conformation both in its free and chaperone-bound forms: Type III translocator folds into a molten globule. *FEBS J.* **274,** 3601–3610 (2007).

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

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

47. Barth, A. The infrared absorption of amino acid side chains. *Prog. Biophys. Mol. Biol.* **74,** 141–173 (2000).

48. Wolpert, M. & Hellwig, P. Infrared spectra and molar absorption coefficients of the 20 alpha amino acids in aqueous solutions in the spectral range from 1800 to 500 cm^{-1}. *Spectrochim. Acta. A. Mol. Biomol. Spectrosc.* **64,** 987–1001 (2006).

49. Fleming, S. et al. Assessing the utility of infrared spectroscopy as a structural diagnostic tool for β‑sheets in self-assembling aromatic peptide amphiphiles. *Langmuir*, **29**, 9510-9515 (2013)

50. Frederix, P. W. J. M. *et al.* Exploring the sequence space for (tri-)peptide self-assembly to design and discover new hydrogels. *Nat. Chem.* **7,** 30–37 (2014).

51. Krainer, G., Broecker, J., Vargas, C., Fanghänel, J. & Keller, S. Quantifying high-affinity binding of hydrophobic ligands by isothermal titration calorimetry. *Anal. Chem.* **84,** 10715–10722 (2012).

52. Wiseman, T., Williston, S., Brandts, J. F. & Lin, L. N. Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biochem.* **179,** 131–137 (1989).

53. Rajarathnam, K. & Rösgen, J. Isothermal titration calorimetry of membrane proteins — progress and challenges. *Biochim. Biophys. Acta* **1838,** 69–77 (2014).

54. Bondos, S. E. & Bicknell, A. Detection and prevention of protein aggregation before, during, and after purification. *Anal. Biochem.* **316,** 223–231 (2003).
55. Ohgushi, M. & Wada, A. ‘Molten-globule state’: a compact form of globular proteins with mobile side-chains. *FEBS Lett.* **164,** 21–24 (1983).
56. Jez, J. M., Schlegel, B. P. & Penning, T. M. Characterization of the substrate binding site in rat liver 3α-hydroxysteroid/dihydrodiol dehydrogenase: the roles of tryptophans in ligand binding and protein fluorescence. *J. Biol. Chem.* **271,** 30190–30198 (1996).
57. Kong, J. & Yu, S. Fourier transform infrared spectroscopic analysis of protein secondary structures. *Acta Biochim. Biophys. Sin.* **39,** 549–559 (2007).
58. Rieppo, L. *et al.* Application of second derivative spectroscopy for increasing molecular specificity of fourier transform infrared spectroscopic imaging of articular cartilage. *Osteoarthritis Cartilage* **20,** 451–459 (2012).
59. Sukumaran, S., Hauser, K., Rauscher, A. & Mäntele, W. Thermal stability of outer membrane protein porin from paracoccus denitrificans: FT-IR as a spectroscopic tool to study lipid–protein interaction. *FEBS Lett.* **579,** 2546–2550 (2005).
60. Fabian, H. *et al.* Impact of point mutations on the structure and thermal stability of ribonuclease T1 in aqueous solution probed by fourier transform infrared spectroscopy. *Biochemistry (Mosc.)* **33,** 10725–10730 (1994).
61. Barth, A. & Zscherp, C. Substrate binding and enzyme function investigated by infrared spectroscopy. *FEBS Lett.* **477,** 151–156 (2000).
62. Choudhary, S. & Sar, P. Uranium biomineralization by a metal resistant pseudomonas aeruginosa strain isolated from contaminated mine waste. *J. Hazard. Mater.* **186,** 336–343 (2011).
63. Kim, W. *et al.* Effect of spaceflight on pseudomonas aeruginosa final cell density is modulated by nutrient and oxygen availability. *BMC Microbiol.* **13,** 1–10 (2013).
64. Rahme, L. G. *et al.* Common virulence factors for bacterial pathogenicity in plants and animals. *Science* **268,** 1899–1902 (1995).

6. Conclusions and outlook

In Chapter (1) 2DIR spectroscopy was presented as a candidate for the extraction of potentially functional protein motions on the fs-ps timescale. The acquisition of rapid local dynamics, such as those that occur at protein active sites, was offered as a means of complementing X-ray and NMR, to expand the picture of biological structure and motion obtained through these powerful techniques. It was also proposed that the difficulty in obtaining vibrational data from large, complex protein structures can be circumvented through the use of effective probe or reporter molecules, and that these allow for the selection of information from a specific structural region. In the interest of preserving the protein native state, intrinsic vibrations were prioritised.

This thesis considered three general classes of protein-ligand complex, and the ease with which local vibrational dynamics can be obtained from utilising the bound substrate as probe.

In Chapter (3), the affinity of ferric haem groups for nitric oxide ligands was utilised to probe the active sites environments of two proteins, under different conditions. The incorporation of solvent isotope-dependence into this study provided a crucial point of discrimination for the solvent-linked dynamics of the haem cavity. The result of this study provided a succinct demonstration of the disparate motions that may occur between two enzymes which exhibit similarities in their static structures. In conjunction with 2DIR, the dynamic changes observed between H_2O and D_2O solvents were seen to be significant; not only revealing markedly different solvent dynamics between enzymes, but also demonstrating the role of water molecules in mediating ligand contact with the functional distal histidine. This strongly implicates water as a participant in the active cycles of each enzyme. Further support was given to this conclusion by the observed effect of substrate binding at the active site of horseradish peroxidase; water was expelled from the vicinity of the active site ligand, suggesting the importance of solvent in further reaction stages of the peroxidase enzyme. Though haem-ligand systems have been studied extensively by a number of techniques, the conclusions presented here are only accessible through an ultrafast technique such as 2DIR.

Chapter (4) explored the viability of a ubiquitous biological substrate to serve as a vibrational probe of complex proteins. The success of NAD as a reporter group has the potential to create a new, reliable experimental design that is applicable to many protein species without the need for any structural modifications that might detract from the biological significance of the study. Although the majority of intrinsic coenzyme vibrations fall short of the criteria for an ideal IR probe, some showed potential that could be explored as equipment sensitivity increases. For the present, out results confirm recent work by Cheatum *et al.* that show that only very small structural modifications would permit the powerful application of NAD in this fashion. Coenzyme structural analogues may therefore prove valuable in vastly extending the scope of 2DIR in protein studies. Though NAD was categorised in this study, a similar evaluation of FAD (flavin adenine dinucleotide) may reveal more favourable spectroscopy for protein-binding experiments in the IR.

The objective of Chapter (5) was to explore the prospects of 2DIR in application to a protein-chaperone complex. Categorisation of such a system is challenging *via* the overwhelmingly majority of measurement techniques, owing to the large molecules involved. Further difficulties arise in the Infrared from the frequently disordered structures of chaperone targets. The latter consideration reduces the practicality of isotope-labelling strategies that have been highly successful in the study of small, ordered peptide systems. It was concluded that membrane-like proteins with unstructured domains marginalise the spectral effects of isotope-labelling due to a lack of dipole-dipole interactions between Amide I oscillators. Further, the absence of secondary structure at the local region of interest, in this case a binding domain, renders isotope-editing temperamental, as the unlabelled frequencies of specific Amide I modes in the domain is highly variable. It was observed however, that the tyrosine side chain may be exploitable as a versatile probe group, applicable not only to this system, but in many other examples. Indeed, the C=C ring mode of tyrosine may prove to be one of few intrinsic vibrations to offer site-specific information. This possibility was further explored *via* “mutations” of the PopD analogue to manipulate potential H-bonding interactions in the binding site, though a potential technical limitation was observed during this study. Observation of the weak probe band appears to be somewhat limited by the co-occurrence of highly intense absorption features within the same spectral window. Significant analysis may be required to isolate the origin of this effect. Given the sensitivity required to perform the intended measurement, it was also necessary to acquire spectra using the ULTRA spectrometer, and so technical modification of the system is not possible in any case.

The tyrosine ring has been successfully employed as an intrinsic probe in linear IR spectroscopy, though its potential in 2DIR has not been categorised. It is concluded from this study that further study of this side chain vibration may reveal a powerful reporter group whose potential is determined by the structural location of the residue. This is an important result, as the effectiveness of this probe may be independent of the size of the molecule under investigation, pending the elimination of spectral artefacts. Should this side chain prove to be an exploitable probe, it is possible that tyrosine could be deployed at any point in a protein sequence through point-mutation providing that the native conformation of the protein is not adversely affected.

It was demonstrated that proteins exhibit functionally relevant dynamics on fs-ps timescales, and that 2DIR is well-placed to provide insight on these behaviours. It is also apparent that existing strategies of retrieving site-specific information, such as isotope-labelling or haem ligation may be exploited to a far greater extent using this recent method. Crucially however, 2DIR lends itself to novel method-development strategies that have not fully been explored or refined. There is therefore significant potential to broaden the range of biological systems to which 2DIR can successfully be applied, by expanding on the result obtained in these investigations.

Appendix

A.1 Supplementary information to Chapter (1)

A.1.1 Amino acid side chains

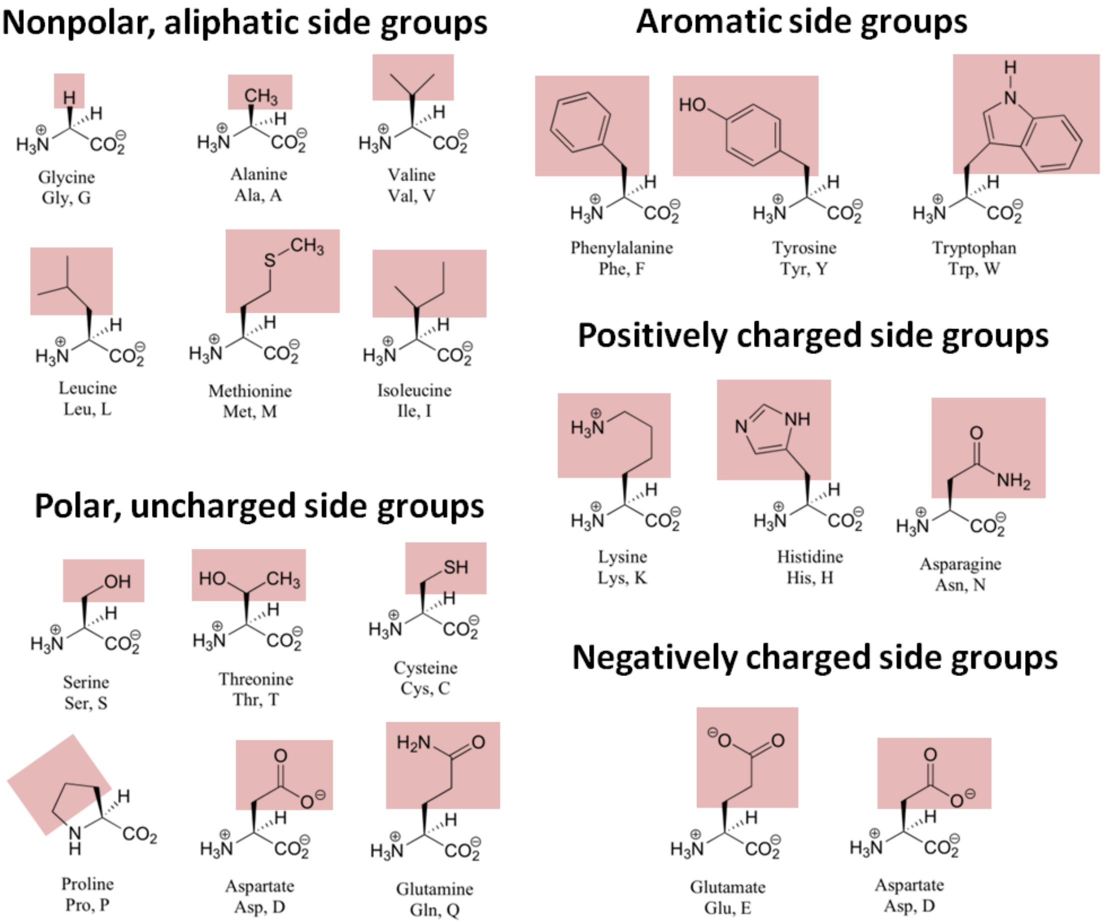


Figure (1): Structures of the twenty naturally-occurring amino acids, categorised by side chain properties. Side chains are shaded in red. Image obtained from ChemWiki†

A.1.2 Acquisition of 2DIR data: time-domain versus frequency-domain

While the data retrieved from both frequency and time-domain approaches to 2DIR has been demonstrated to contain equivalent information[1], the latter method requires additional considerations at an experimental level. The pump frequency axis in time-domain 2DIR is obtained through sampling of the pump auto-field correlation, and accordingly, the manner in which this signal is sampled has a direct effect on the quality of resultant 2DIR spectra. In order to produce data with high pump frequency resolution, a large range of t_1 values must be measured, usually with a very small sampling interval (<10 fs). For this reason, data acquisition in the time-domain is generally much slower than in frequency-domain 2DIR. Though, by consequence, higher pump axis resolutions can be obtained in the time-domain[2]. The following section (A.2.1) presents the correlation between resolution and auto-field correlation in a quantitative manner.

A.2 Supplementary information to Chapter (2)

A.2.1 Pump frequency resolution limits

The highest pump frequency that can be detected in Fourier transformation of the auto-field correlation is dictated by the Nyquist frequency[3]; a function of the sampling interval. Further, the spectral resolution of the pump axis is related to the maximum value of t_1, and thus, for high spectral resolution in the pump axis, a large number of points are typically collected for t_1.

Nyquist frequency and resolution limit

$$\frac{\omega_N}{2\pi} = \frac{1}{2\Delta t_1} \tag{1}$$

Where Δt is the sampling rate and ω_N is the Nyquist frequency; the highest frequency that can be recovered from the Fourier-transform.

$$\Delta\omega = \frac{1}{t_{1max}} \tag{2}$$

Where $\Delta\omega$ is the frequency resolution and t_{max} is the final time value in the sampling period. It is this principle that is exploited in the approach of zero-padding, as the sample range is doubled by the addition of $y = 0$ values to the end of the data range.

A.3 Supplementary information to Chapter (3)

A.3.1 UV-Vis spectral changes with HRP ligation

Ligand-haem binding is accompanied by changes in the UV-Vis absorption of the transition metal complex[4]. A Perkin-Elmer Lambda-2 spectrometer was therefore used to verify the binding of aromatic substrates and diatomic ligands to the enzyme horseradish peroxidase. Spectra were acquired at 0.12 mM enzyme, with 1.2 mM of the aromatic compound. Substrate binding converted the initial mixture of 5 and 6-coordinate metal complexes present in horseradish peroxidase to a purely 6-coordinate species, reportedly the consequence of a substrate-facilitated ligation of the haem group by a water molecule, generating a so-called 6-aquo complex[5]. Change of the coordination state was manifest as the loss of a shoulder peak at 383 nm, and the intensity increase of a primary 403 nm band[6]. Spectra were recorded between 250-750 nm, with a 100 μm.

A.3.2 Pump-probe residual

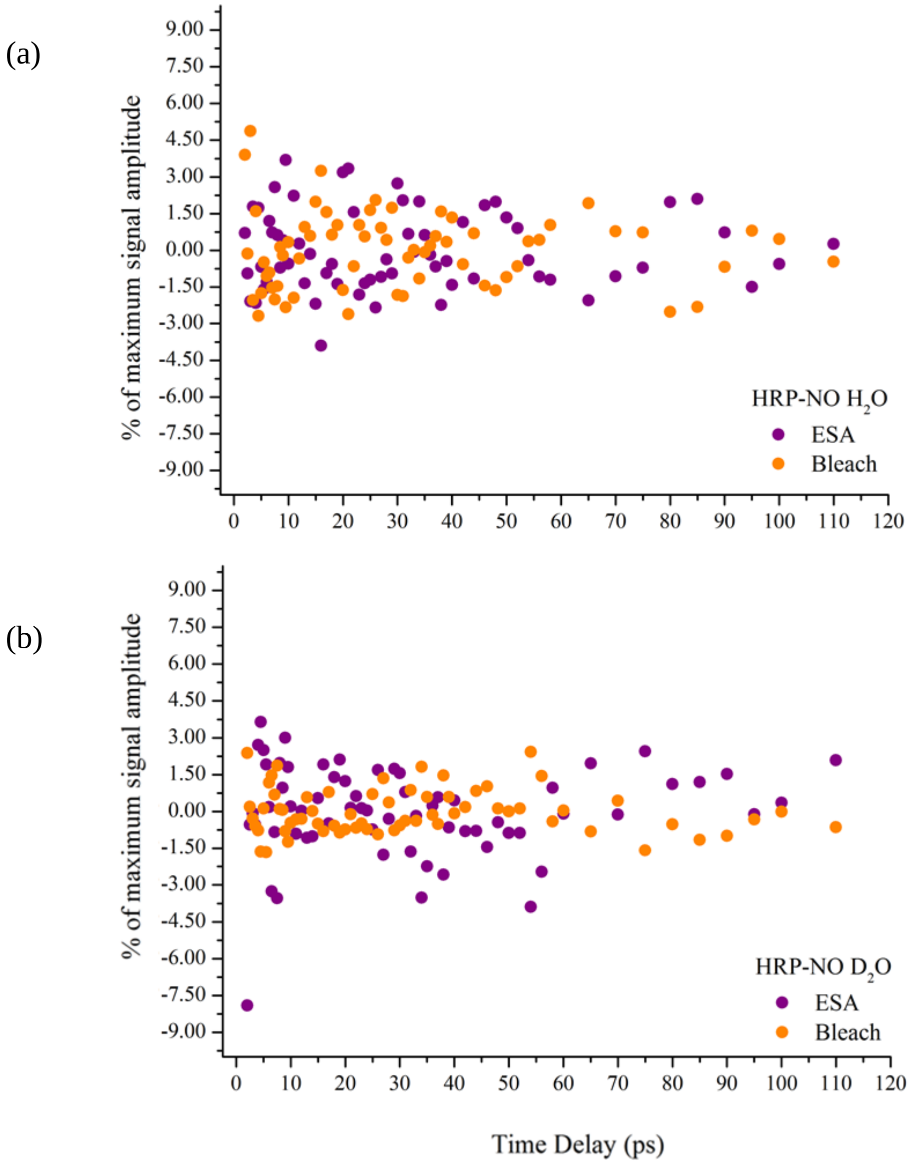


Figure (2): Residuals from the exponential fitting of the HRP-NO vibrational lifetime in (a) H_2O and (b) D_2O

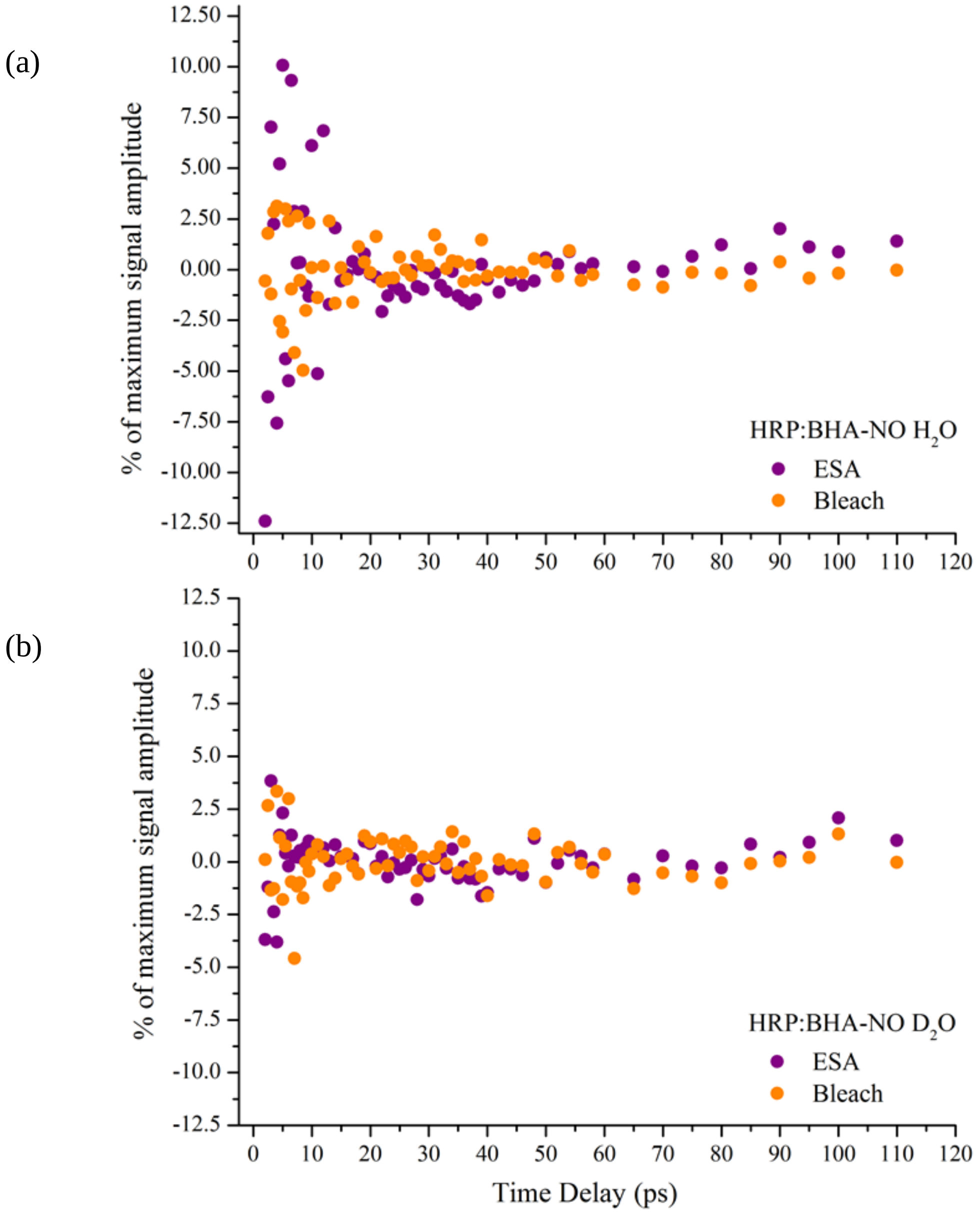


Figure (3): Residuals from the exponential fitting of the HRP:BHA-NO vibrational lifetime in (a) H2O and (b) D2O

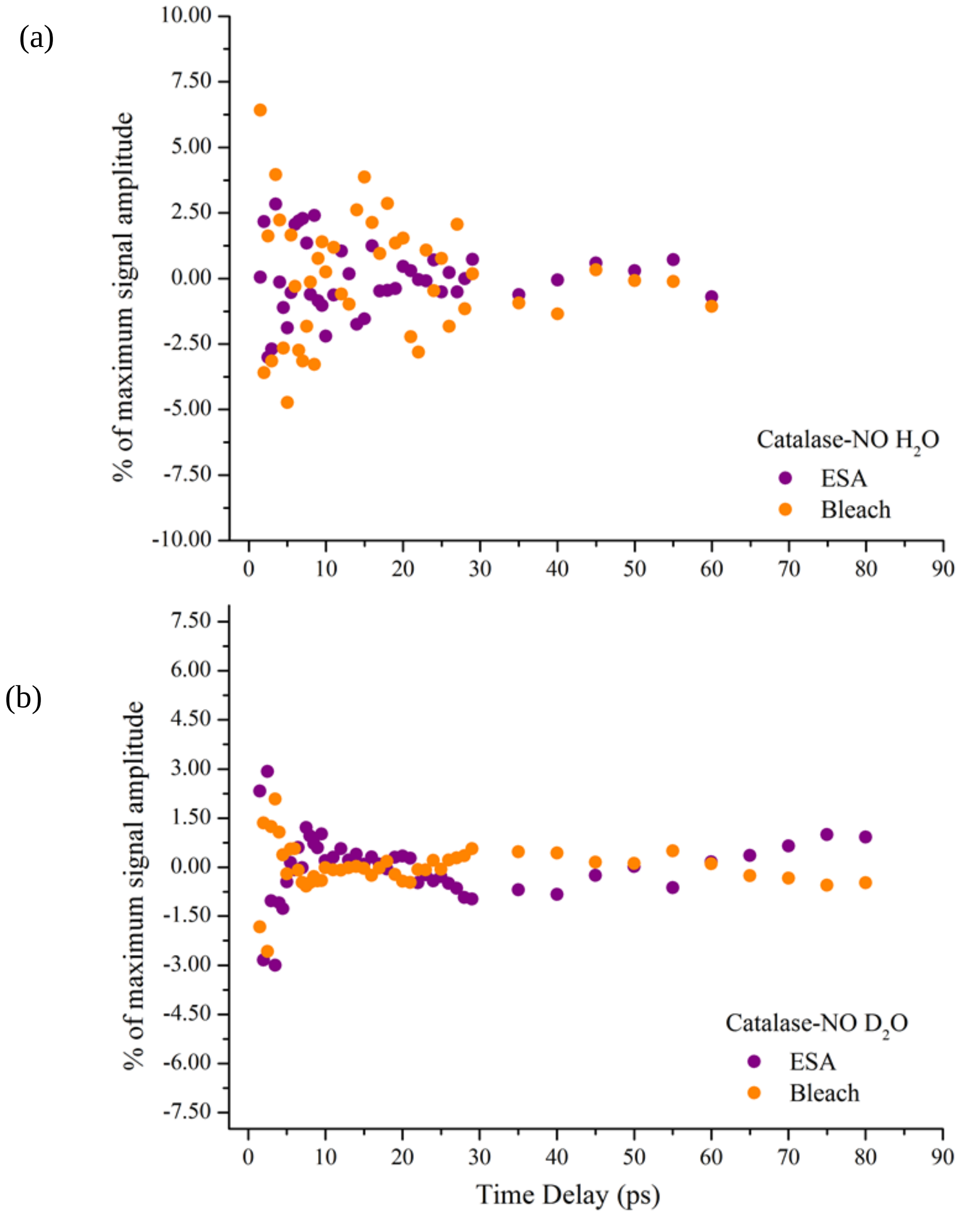


Figure (4): Residuals from the exponential fitting of the catalase-NO vibrational lifetime in (a) H2O and (b) D2O

A.3.3 Log-scale plots of spectral diffusion

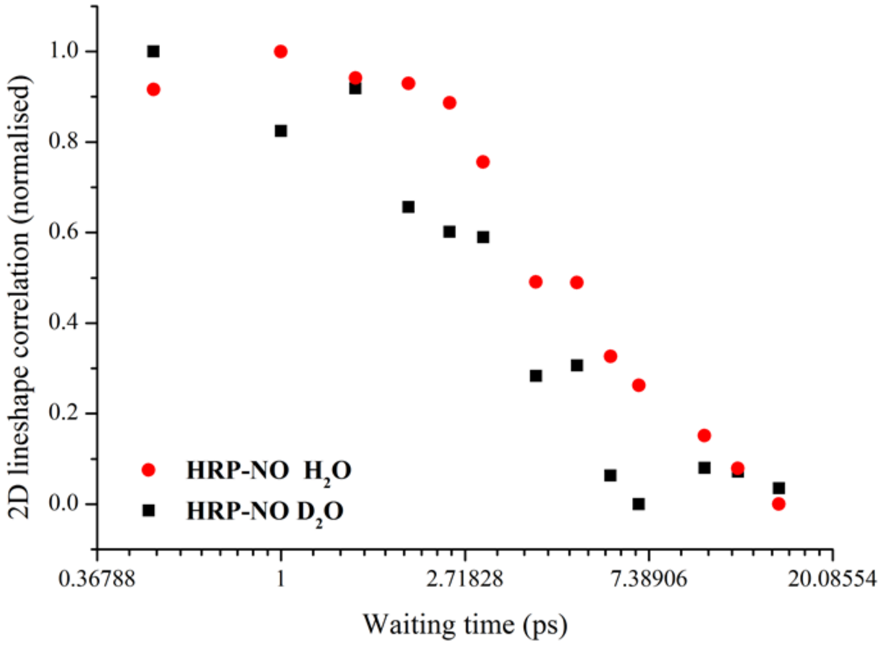


Figure (4): The spectral diffusion profile of HRP-NO (in both H_2O (red) and D_2O (black)) in lin-log format, i.e. a logarithmic x-axis and linear y-axis

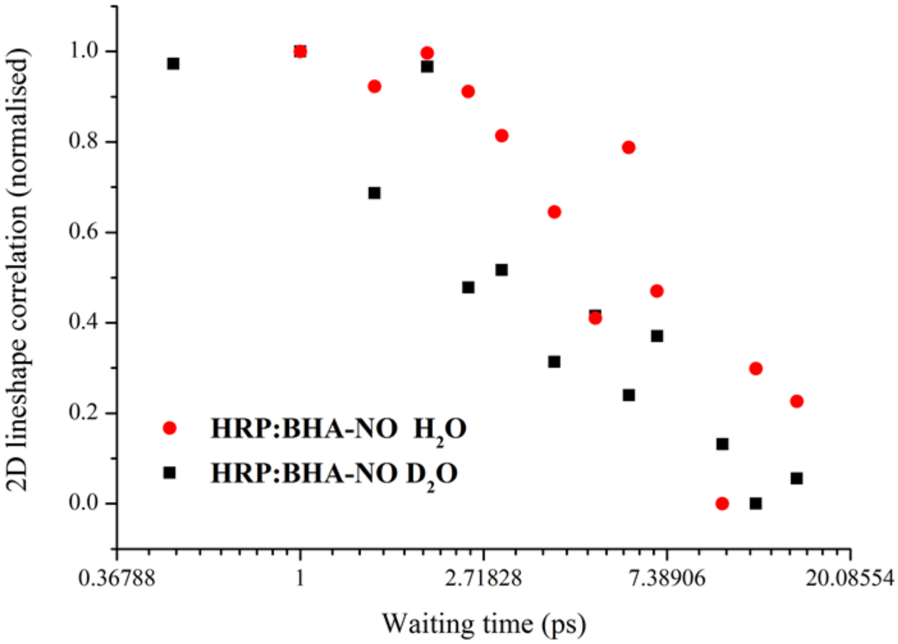


Figure (5): The spectral diffusion profile of HRP:BHA-NO (in both H_2O (red) and D_2O (black)) in lin-log format, i.e. a logarithmic x-axis and linear y-axis

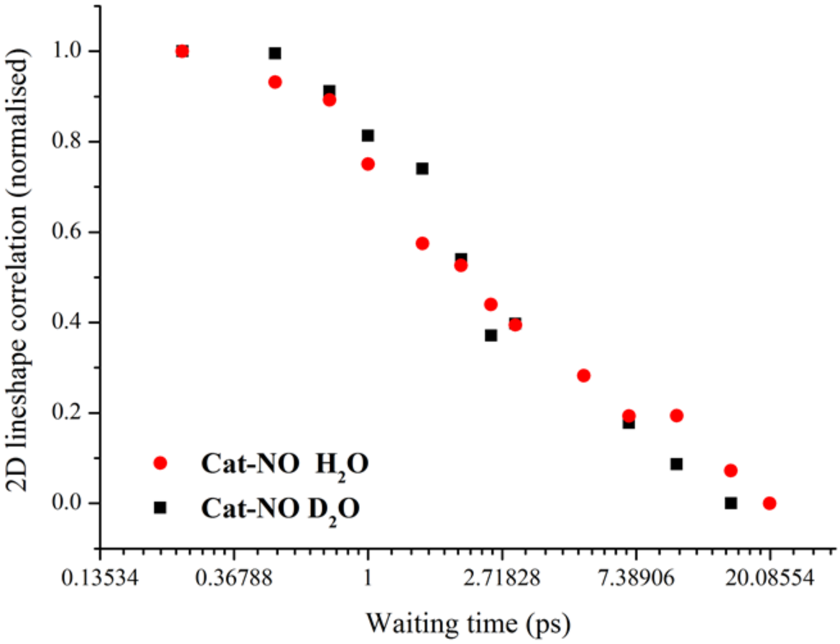


Figure (4): The spectral diffusion profile of catalase-NO (in both H_2O *(red) and* D_2O *(black)) in lin-log format, i.e. a logarithmic x-axis and linear y-axis*

A.3.4 Pump-probe spectroscopy of myoglobin-NO

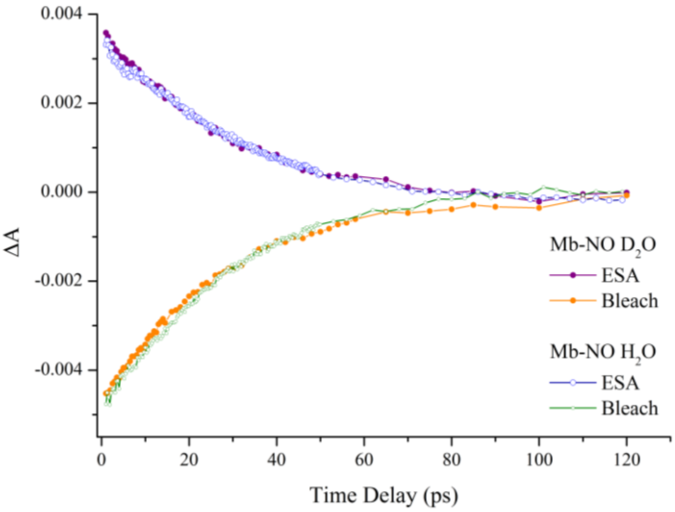


Figure (2): Pump-probe signals for Mb-NO, in both H2O and D2O. In either solvent, fitting of a single exponential function to the data gave a vibrational lifetime ~27 ps.

A.4 Supplementary information to Chapter (4)

A.4.1 ^{1}HNMR measurements (Experimental)

Proton nuclear magnetic resonance (^{1}H NMR) can be used to acquire valuable steady-state information on molecular structure[7]. This technique utilises radio-frequency pulses to measure the photon energy required to reverse the alignment of proton nuclear spin in the presence of an applied magnetic field, i.e. to switch the spin-state from +1/2 to -1/2. Reversal of the spin-state occurs when resonance is achieved between the radio-pulses and the energy difference, ΔE, between the two spin-states.

$$\Delta E = \hbar \gamma B_0 \tag{3}$$

Where B_0 is the strength of the external field, and γ is the "nuclear shielding". γ is determined by the specific proton environment, through factors such as electron density and inter-nuclear coupling. From this, ^{1}H NMR spectra relay the number of protons in a molecule, the coupling between adjacent protons, and the electron density of each ^{1}H nucleus, which in turn can be used to assign the functional group to which the proton belongs.

The pD and solvent dependence studies carried out in Chapter (4) followed protocols used by Catterall *et al.*[8] in an NMR investigation of the solution-phase conformation of NAD, under various conditions. As such, ^{1}H NMR data was acquired in parallel to the FTIR spectra included in Chapter (4). The same samples were used in each case, to verify complete consistency with published NMR results.

^{1}H NMR spectra were acquired with an Oxford Instruments NMR AS400 operating at 400 MHz. Samples were prepared from the same solutions as used for FTIR measurements to ensure consistency, and were measured within 24 hours of the equivalent IR scans.

A.4.2 IR detection of NAD conformations- extended (Results and discussion)

During the course of the experimental study presented in Chapter (4), efforts were made to establish an infrared “signature” for the folded and unfolded conformations of NAD. A detailed account of this sub-study is presented here.

Protonation of the free adenine or adenosine molecules triggers substantial changes in the vibrational spectra of each molecule. This effect is further seen to extend to the dinucleotides upon adenine group protonation. In the case of NAD+, where unfolding of the molecule is also believed to accompany protonation, it becomes unclear whether the former process contributes to observed spectral changes. However, it has been proposed that nonpolar solvent environments may also trigger the conformational transition of both coenzymes, from folded to unfolded. An FTIR analogue of experiments performed by Catterall[8] was conducted here to determine the effects of folding in the infrared. Coenzyme samples were thus prepared in mixtures of water and methanol, and vibrational spectra were acquired across a range of relative methanol concentrations, from 0-17 molar.

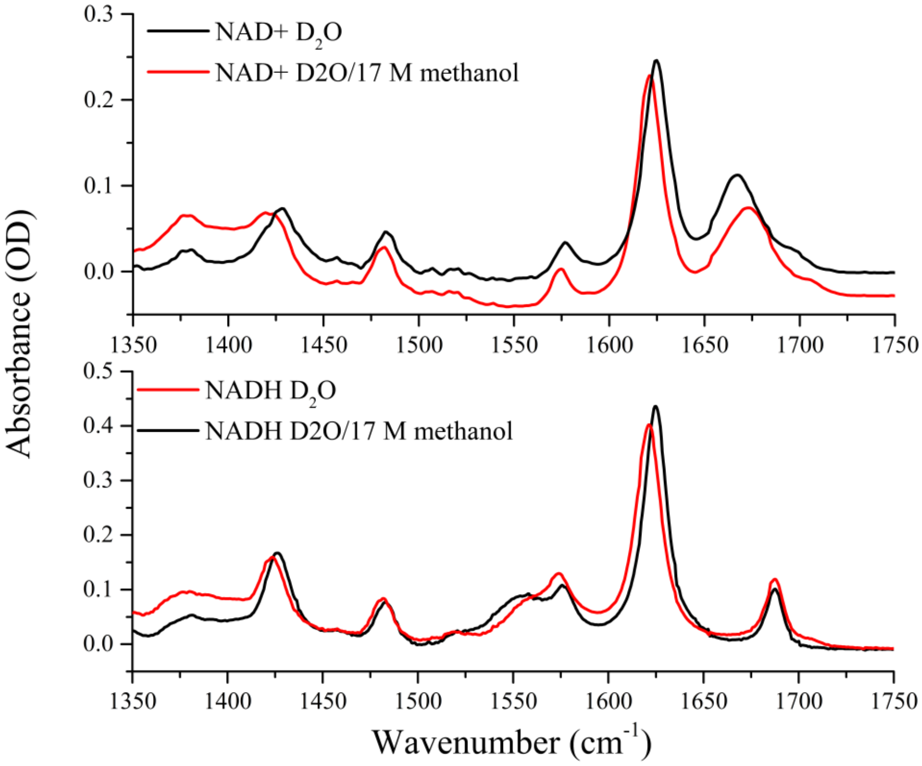


Figure (3): The spectra of NAD+ and NADH at neutral pD in water, and in a 30:70 solution of water: methanol D-4 (17 molar methanol concentration.)

Figure (3) compares the FTIR spectra of NAD+ and NADH in a 70% methanol (~17 M) and 30% water mixture to coenzyme spectra under neutral, aqueous conditions. The reduced coenzyme has been reported to show ^{1}H NMR chemical shift changes, ascribable to unfolding, at methanol concentrations of ~20% (~4 M), while the oxidised form shows similar changes (increasing chemical shift for all nicotinamide and adenine protons) at ~10% methanol (~2 M). The spectra at 70% methanol should therefore reflect fully unfolded conformations of each dinucleotide.

Increasing MeOD concentrations cause shifts of the central frequencies and FWHM of a number of bands in both NAD+ and NADH, though the observed dependences are very small in each case. For each "sensitive" mode, the final vibrational frequency at 70% MeOD is displaced by only 2-3 cm^{-1} relative to that in pure D_2O. The magnitude of this frequency change is comparable for each mode, though the direction of the shift varies between vibrations. Only the C=O stretch of NAD+ (1666 cm^{-1}) and dihydro ring (1555 cm^{-1}) mode of NADH can be seen to blue-shift with increasing methanol concentration, while the remainder are red-shifted.

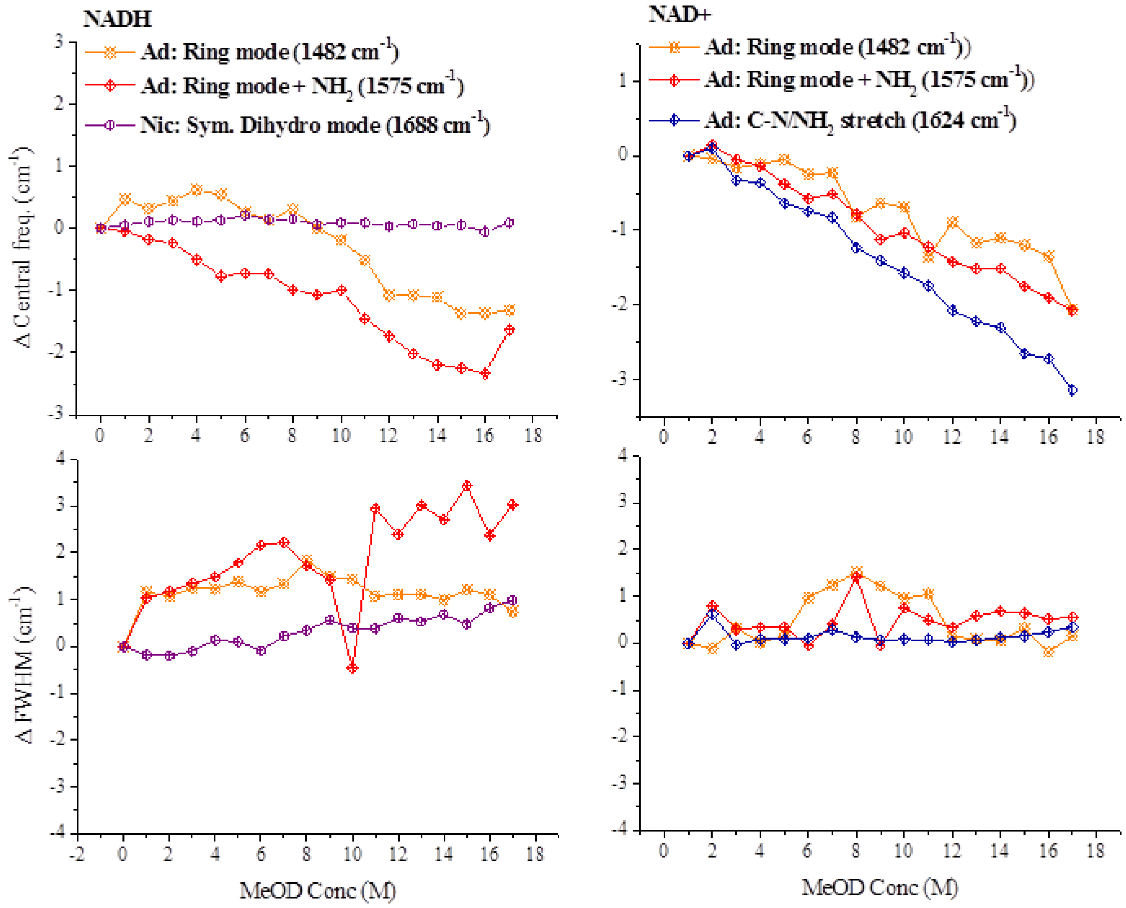


Figure (4): The change in peak positions (upper) and widths (lower) for NADH (left) and NAD+ (right) between 0-17 M methanol concentrations.

Each NAD band was fitted with a Gaussian profile in order to produce plots of peak position and FWHM as a function of MeOD concentration. Figure (4) shows representative results of this analysis; a number of peaks are omitted due to their convolution with other signals, leading to poor fitting. The correlation in each figure is essentially linear across the full range of the study (0-70% MeOD) for the vast majority of modes. Since the effect of methanol addition is generally small, when comparing solvent-dependent modes that are shared by NAD+ and NADH (such as the adenine mode at 1577 cm^{-1}) little distinction can be made between the observed responses. Although it seems unintuitive to expect different behaviour from a molecular group that is identical in either form of NAD, ^1H NMR data has identified different thresholds for unfolding with respect to the solvent polarity[8]. For NAD+, the chemical shifts of all nicotinamide and adenine protons show evidence of the unfolding process immediately after a 10% methanol volume is attained. Conversely, a volume of 20% is required to trigger the majority of changes in the NADH NMR spectrum. Further, the reported proton shifts for this experiment are not solely linear in response to the decreasing solution polarity; a mixture of linear and sigmoidal responses is observed.

Cumulatively, from these remarks on the NMR data, it can be inferred that the majority of IR spectral changes are attributable to the solvent gradient, rather than the unfolding process. In order to deconvolute the two contributions, a reference sample is required. Understanding the effect of MeOD on the NAD spectra without triggering unfolding, however, is problematic. It is debateable that an acidic solution of NAD+, when combined with a high MeOD concentration, would only experience spectral changes arising from the solvent, since the protonated NAD+ exists in an already unfolded state. This could, in principle, provide a firm baseline for the solvent-only response of certain vibrational modes in the coenzyme. However, the applicability of this reference sample is compromised by several factors already discussed, the foremost of which is band overlap. It is also debateable whether the vibrations of protonated adenine should respond identically to those of the deprotonated form.

An attempt was made to use NMN+ as a reference for the nictonamide vibrations of NAD+ alone. The mononucleotide form of the oxidised coenzyme is unable to fold or unfold, but retains a high degree of congruity with the NAD+ spectrum. However, through spectral overlap with adenine modes, only the 1666 cm^{-1} band can be used as a reference. Indeed, slight differences were observed between the carbonyl stretching modes of each NAD+ and NMN+, though these are discussed in more detail below. Although a parallel can be drawn between NMN+ and NAD+, analogous data cannot be provided for NADH. It therefore cannot be determined whether the blue-shift affecting the 1555 cm^{-1} dihydro-pyridine mode is linked to unfolding. It is certainly possible that the chemical environment of the 1555 cm^{-1} C=C stretching bands be comparable with those of the C2 and C4 protons; all of which are readily sensitive to unfolding under NMR measurement.

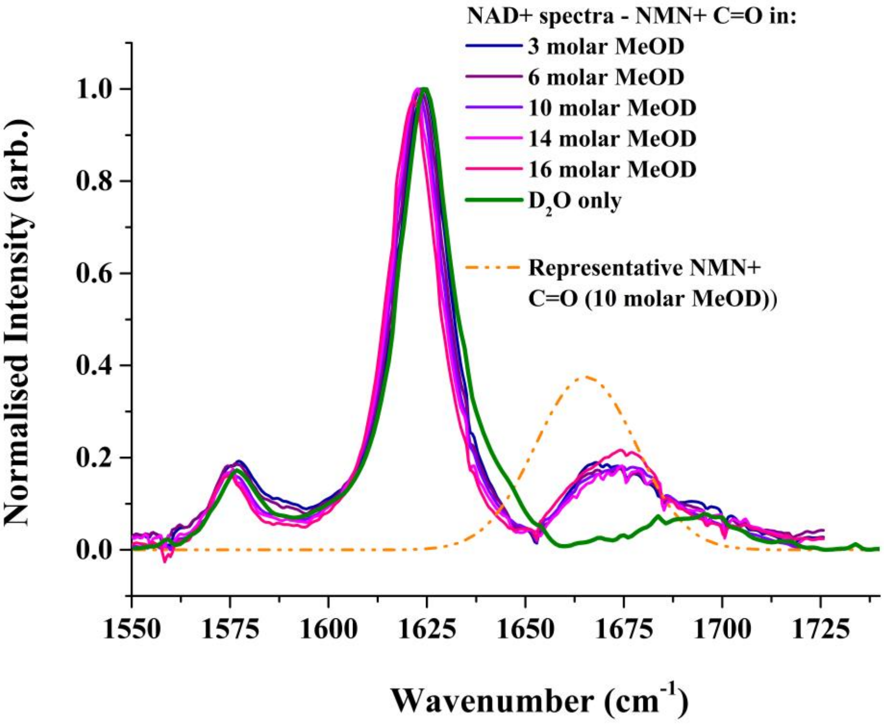


Figure (5): The residual spectra of NAD+ at various methanol concentrations after subtraction of an NMN+ C=O signal.

Poor fitting of the NAD+ 1666 cm^{-1} mode to a single Gaussian suggests the presence of at least one additional band towards the high frequency shoulder of the C=O absorption. Given the lack of competing vibrations predicted in this region, it is not unreasonable that these additional bands are also C=O stretching modes. In this case separate peaks may indicate splitting of the carbonyl oscillation into sub-states that originate, from different NAD+ conformations. In order to further analyse these hidden peaks, fitting of the NMN+ C=O was performed at a number of methanol concentrations. This band was well described by a single Gaussian function in each case. It was then assumed that the carbonyl peak of NMN+ is a strong reference for the C=O stretch of an extended NAD+ conformation, and that adequate subtraction of the single mononucleotide band from NAD+ will leave only hidden peaks that correspond to non-solvent-exposed sub-states.

The spectrum of NAD+ is shown at a range of MeOD concentrations following subtraction of an NMN+ carbonyl signal under the corresponding solvent conditions (Figure (5)). Two residual peaks are observed in the range 1660-1720 cm^{-1}: a band centred ~1670 cm^{-1}, of similar intensity to the 1577 cm^{-1} of adenine, and weak, broad band with central frequency ~1700 cm^{-1} Interestingly, little change seems to occur in either of the peaks as a result of the changing methanol concentration. Apparent differences to band shape are only seen at the highest measured value of MeOD concentration; at 16 M the 1670 cm^{-1} band arguably lies at a higher frequency, and shows greater intensity. However, the change does not occur gradually enough with decreasing solvent polarity to be associated with unfolding of NAD+. Moreover, the reduced solubility of NAD+ at high methanol concentrations casts doubt on whether the change is genuine.

Interestingly, the higher frequency residual band near 1700 cm^{-1} is also faintly observable in pure D_2O. The similar intensities of this mode between water and MeOD: D_2O mixtures of arbitrary concentration suggest that it is not linked to a conformational state of NAD+. Based on previous publications[9], it is possible that this 1700 cm^{-1} band belongs to the C=O stretching mode of NAD+ molecules with non-deuterated Amide groups[9], though this assignment is somewhat speculative.

The mode at 1670cm^{-1} is difficult to place; a methanol volume of ~15% (3M) appears adequate to both introduce and saturate this signal. Several strong vibrations linked to the pyrophosphate bridge are predicted through DFT calculation, with approximate, uncorrected frequencies in the range: 890-930 cm^{-1} (corrected ~ 860-900 cm^{-1}). The residual 1670 cm^{-1} band is possibly the result of overtone absorption. A ν = 0-2 excitation for an oscillator of fundamental frequency ~850cm^{-1} would result in overtone band ~1670 cm^{-1} if the anharmonicity of the oscillator was approximately 30 cm^{-1}. Though these bands are predicted for water-solvated NAD, an equivalent “overtone” is not seen under these aqueous conditions. Modes of the phosphate bridge may therefore differ in methanol to produce an overtone. The steric effect of solvent polarity on dinucleotide backbones has been briefly remarked upon[4], though not in enough detail to provide more than a tentative assignment.

However, since the band observed is invariant with solvent concentration, the assignment to a vibration of NAD seems logical. There is insufficient data to add weight to contrary explanations of this peak.

Ultimately, neither residual band shows any true correlation to the solvent polarity. Though the fixed positions of both bands could be taken as indicator of the C=O group that is screened from the solvent environment, one would expect intensity shifts as a result of the changing equilibrium between folded and unfolded forms of NAD+. The only firm statement that can be made from this study is that the C=O stretching mode of NAD+ seems predominantly solvent-exposed. The coenzyme may take a folded form in neutral, aqueous solution, but the lack of unknown absorption features after subtraction of the NMN+ C=O signal would imply that, even in a closed state, the NAD+ C=O is still in contact with the solvent.

A.5 Supplementary information to Chapter (5)

A.5.1 Synthesis of $^{47\text{-}56}$PopD peptides (Experimental)

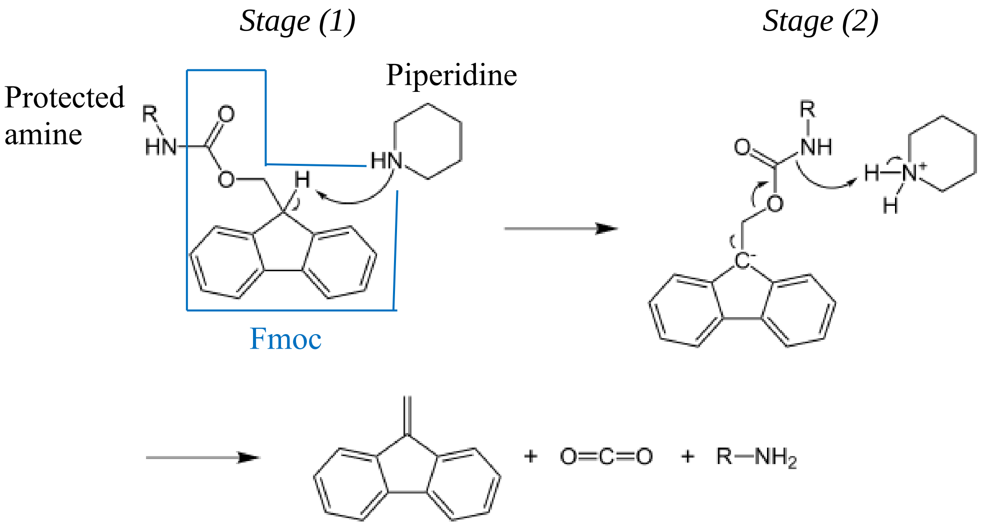
Peptides were synthesised using a solid-state protocol[11–13] in which the peptide sequence is constructed residue-by-residue while bound to a porous resin. The resin is composed of polystyrene chains, cross-linked with divinylbenzene (DVB) groups to create a mesh. Though the resulting polystyrene network is generally unreactive and insoluble, for the purposes of peptide synthesis, it is functionalised with "linker" groups that allow amino acids to anchor with the resin. The polypeptide therefore remains part of the solid-state polymer network, and the reactants are introduced in the solution-phase. Filtration of the by-products or residual reagents can be performed easily without the loss of product, and so the efficiency of solid-state synthesis is high.

In these syntheses, a rink amide resin was employed. This resin is functionalised with 4-methylbenzhydrylamine (MBHA) linkers that may bond covalently with the carboxy group of an amino acid by acting as an amino acid N-terminal. The polypeptide is therefore constructed from the C to N terminus, and the "last" amino acid in the desired sequence is held in place *via* a peptide bond formed with the resin linker group. In the sequence, DRVELNAPRQ, synthesis therefore begins with the "loading" of Gln (Q) onto the resin. This is conducted by the following procedure:

(1) Swelling: In order to use the resin as a reaction medium, it must be "swollen" using a solvent. This is achieved through the addition of dichloromethane (DCM) to the resin. DCM permeates the resin and solvates the polystyrene chains. However, since the cross-linked mesh is macroscopically insoluble, solvation increases the volume of the resin. Expansion of the mesh permits greater solvent access to the interior of the network, which crucially provides greater contact of the linker groups with solution-phase reactants.

In each synthesis reaction, 0.3 mmol of resin was used as the reaction medium.

(2) Deprotection: To prevent degradation or undesired reactions during the synthesis procedure, the N-termini and side chains of each amino acid are “protected” by large, stable chemical groups. A base-labile fluorenyl-9-methoxycarbonyl (Fmoc) group is used to protect peptide’s amine (NH_2) functional groups. Formation of peptide bonds must therefore be preceded by removal of the Fmoc group; a so-called deprotection step (Scheme (1)). This is also true of the first peptide bond that must be formed with the resin during loading, as the resin’s amine group is guarded by an Fmoc moiety. Deprotection is achieved by washing the resin with a solution of 20% piperidine and 80% dimethylformamide (DMF). Piperidine is a secondary amine that provides optimal deprotonation of the Fmoc fluorine ring, triggering the reaction shown in Scheme (1).



Scheme (1): The reaction of piperidine with an Fmoc protecting group to produce a deprotected amine (R-NH_2)

(3) Coupling: 1.5 mmol of the first amino acid (Fmoc-Gln) was dissolved in DMF. A five-times molar excess of peptide-to-resin is used in the initial loading phase as the ultimate reaction yield is directly linked to the quantity of peptide loaded. The coupling agent, HATU (shown in Figure (6)), is dissolved with the peptide at equal concentration, and the base, diisopropylethylamine (DIPEA), is added at double this concentration, i.e. 3 mmol. This solution is added to the deprotected resin, and the mixture is stirred continuously for four hours.

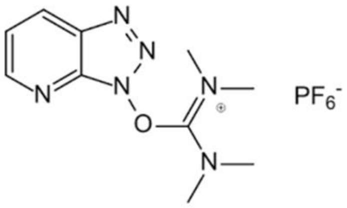
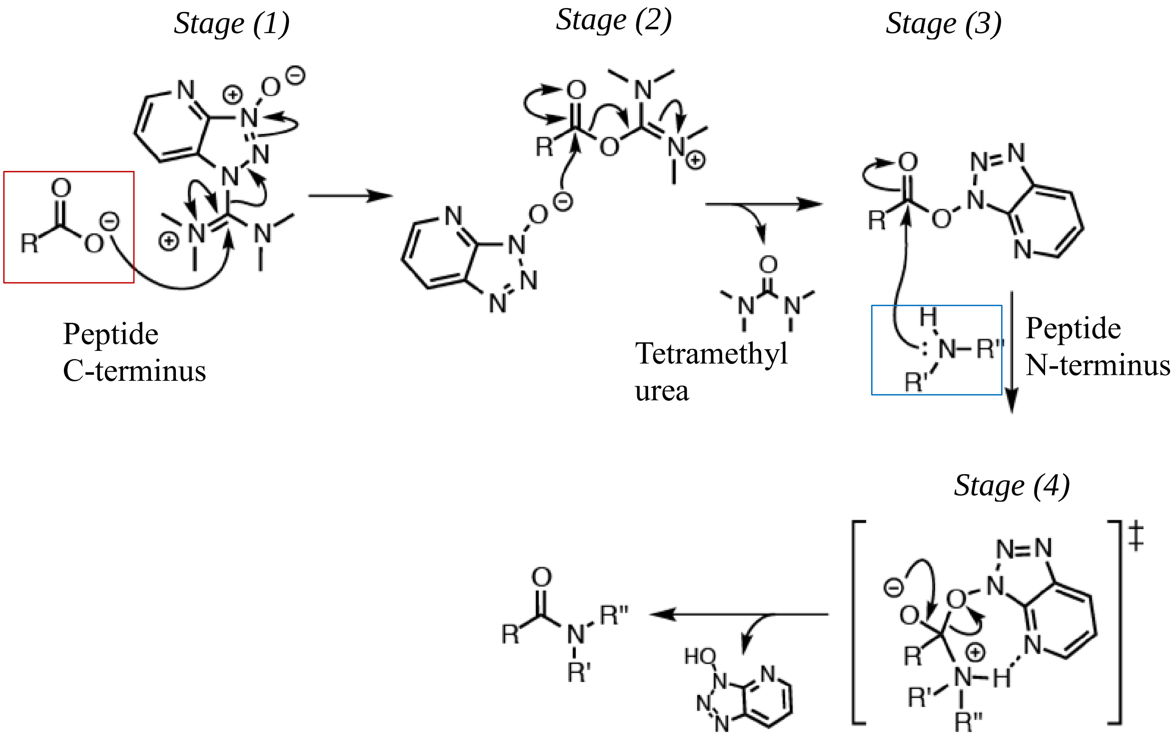


Figure (6): structure of the coupling agent, HATU

During the course of the reaction, DIPEA molecules deprotonate peptide carboxylic acids, allowing the coupling agent to form adducts with the resulting O^- group, generating an ester. The O- group electron is however, withdrawn across the HATU molecule, creating an unstable compound that rapidly degrades. After liberation of tetramethyl urea, the remaining HATU molecule (the pyridine-triazolo ring region) forms the compound shown in stage (3) of Scheme (2). Attack on this compound by an amine, such as that of the deprotected resin or amino acid, causes protonation of the singly-bonded ester oxygen, and the resulting OH joins the pyridine-triazolo leaving-group, leading to formation of a peptide bond.



Scheme (2): The peptide bond formation reaction, facilitated by HATU and DIPEA.

(4) Filtration: Excess reactants and by-products are washed from the resin using DMF, followed by DCM and diethyl ether.

By repeating steps (1) to (4), additional amino acids can be added to the resin-bound sequence. After loading of the resin however, the coupling of the further residues is conducted in an automated fashion using a peptide synthesiser. Firstly, the number of moles of peptide loaded onto the resin is calculated through a measurement of the UV-Vis absorption at 304 nm; the absorption peak of the Fmoc moiety. Dry peptide masses for the automated synthesis are calculated from the derived loading value. Each peptide is sealed in an individual cartridge with a molar equivalent of coupling agent. The following synthesis reaction steps ((1) to (4)) are fully automated. During the deprotection phase, the synthesiser monitors the optical density at 304 nm in order to ensure that efficient removal of Fmoc groups has been performed, increasing the reaction efficiency. Here, a Tribute peptide synthesiser was used, manufactured by Protein Technologies Inc.

After completion of the sequence, the peptide is cleaved from the resin through the addition of trifluoroacetic acid (TFA), which also removes the acid-labile protecting groups of the side chains. TFA is applied in a solution of 95% TFA, 2.5% H2O and 2.5% triisopropylsilane (TIS). During the cleavage phase, reactive cations are produced through side-chain deprotection. Undesired reactions between these liberated species are prevented by water and TIS, both of which serve as scavenger molecules, forming stable compounds with the cations. After ~4 hours, the cleavage mixture is filtered through a phase separator into cold (-20 ˚C) ether diethyl ether. Resin is removed by the phase separator, and the ether facilitates further separation of the peptide from by-products by solvating the cleavage scavengers whilst causing the peptide to precipitate. The resulting peptide solid is repeatedly washed in ether, yielding the crude peptide product.

Isotopically-labelled residues were not coupled automatically. Instead, a manual coupling phase was performed via an identical protocol to that of the resin loading procedure, though only a three times molar excess of the ^{13}C-containing amino acid was used (relative to the resin loading value), to reduce the quantity of materials required.

A.5.2 HPLC (Experimental)

Crude peptide products were purified *via* high performance liquid chromatography (HPLC)[14] in the reverse-phase. In this technique, a mixture of solute species is dissolved in a polar buffer and flowed through a channel packed with hydrophobic silica beads, referred to as a column. Polar solute molecules pass through the column, though nonpolar species with higher affinity for the column than the solvent become immobilised within the silica network. Each species of immobilised solute molecule will possess a different affinity for the column, which provides the means for separation of each species. Solvent is continuously flowed through the column from two sources, buffers A and B, with a controllable ratio of A to B. The buffers are chosen such that a gradually increasing ratio of B to A will produce a negative polarity gradient. Solute molecules in the column will become mobile when their affinity for the solvent becomes greater than that of the column. Each molecular species will therefore leave the column at a specific “elution” point that corresponds to a ratio of A and B.

The solvent’s optical density at programmable UV-Vis wavelengths is measured after the column, allowing re-solubilised products to be detected as they leave the column. When eluted compounds are detected, they are collected as “fractions”, ultimately allowing the various products from a reaction to be separated, thereby increasing the purity of each compound considerably.

Each of the synthesised peptide species was purified using a reverse-phase silica (C18) column, using water with 0.1% TFA (buffer A; polar) and acetonitrile with 0.1% TFA (buffer B; nonpolar). Crude reaction mixtures were injected in 450 μl volumes. The % of buffer B was increased from 5-60 over the course of 10 minutes, during which period the peptide was eluted. Eluted peptides gave rise to an absorption band at 214 nm, and automated fraction collection was triggered by an optical density at this wavelength that exceeded a specifiable threshold level.

Purified fractions were freeze-dried in order to obtain lyophilised peptide samples. Small amounts of each freeze-dried sample were re-solubilised in order to determine sample purity *via* analytical HPLC. Peptides with purity below 95% were re-purified using by repeating the HPLC protocol above.

A.5.3 Mass Spectrometry (Experimental)

Electrospray-ionisation mass spectrometry[15] was used to gauge the success of peptide synthesis through determination of the purified product's molecular weight. Dry peptides were dissolved in a mixture of H_2O and methanol (MeOH), and injected into the mass spectrometer. The inlet capillary for the solution terminates in a nozzle to which a strong electric (~4 kV) field is applied. A "Taylor Cone" forms in the nozzle, creating an aerosol spray from the injected solution. Solute molecules are ionised by the electric field.

Heating of the travelling droplets causes evaporation of the solvent, causing a reduction in droplet size. As this occurs, the like charges in the droplet become more densely-packed, allowing the influence of repulsion to overtake that of surface tension, causing desolvation of the solute molecules. Low concentrations of acetic acid in the initial solution leave the solute particles ionised after desolvation.

The ionised solute molecules are pumped into a high vacuum and contained in a quadrupole ion trap. Modulation of the applied electric field triggers the ejection of molecules from the trap based on the mass-to charge ratios (m/z) of each ion. Steady increase of the voltage applied to the mass trap therefore allows the separation and detection of ion across a large m/z range. Through this method, multiply charged ions can be produced, resulting in detected signals at m/2, m/3, etc.

Peptides were detected as positive ions using a ThermoQuest Finnigan LC duo mass spectrometer, coupled to a Razel syringe pump.

A.5.4 Isothermal Titration Calorimetry (ITC) (Experimental)

ITC[16] was used to determine the optimal concentrations for the IR by determining the binding constants of various synthesised peptide species with a target protein, PcrH.

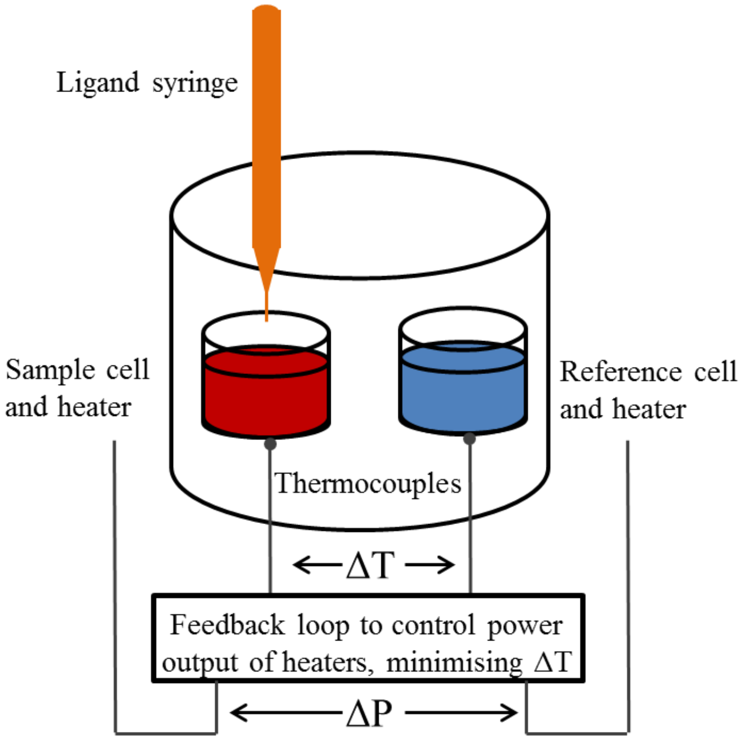


Figure (7): Schematic of an isothermal titration calorimeter.

In the technique of ITC, two fluid-containing cells (the sample and reference) are placed in contact with separate heating elements. An equivalent ambient temperature is maintained in each cell. The sample cell, however, contains a “target” molecule, in this case a protein, while the reference contains only buffer. Low volumes (2-3 µl) of a ligand solution are titrated into the sample cell, causing a deviation in the ambient temperature through an exothermic or endothermic binding process. The temperature change causes a difference in the power required to maintain thermal equilibrium between the sample and reference heaters, and this measured power difference can be integrated over the period of time required to reach the equilibrium temperature, yielding the additional energy exerted by the sample heater (in the case of an endothermic reaction) or the reference heater (in the case of an exothermic reaction). Further, by normalising this energy trace to the relative concentrations of protein and ligand, a “titration curve” can be extracted (Figure (8)(b)).

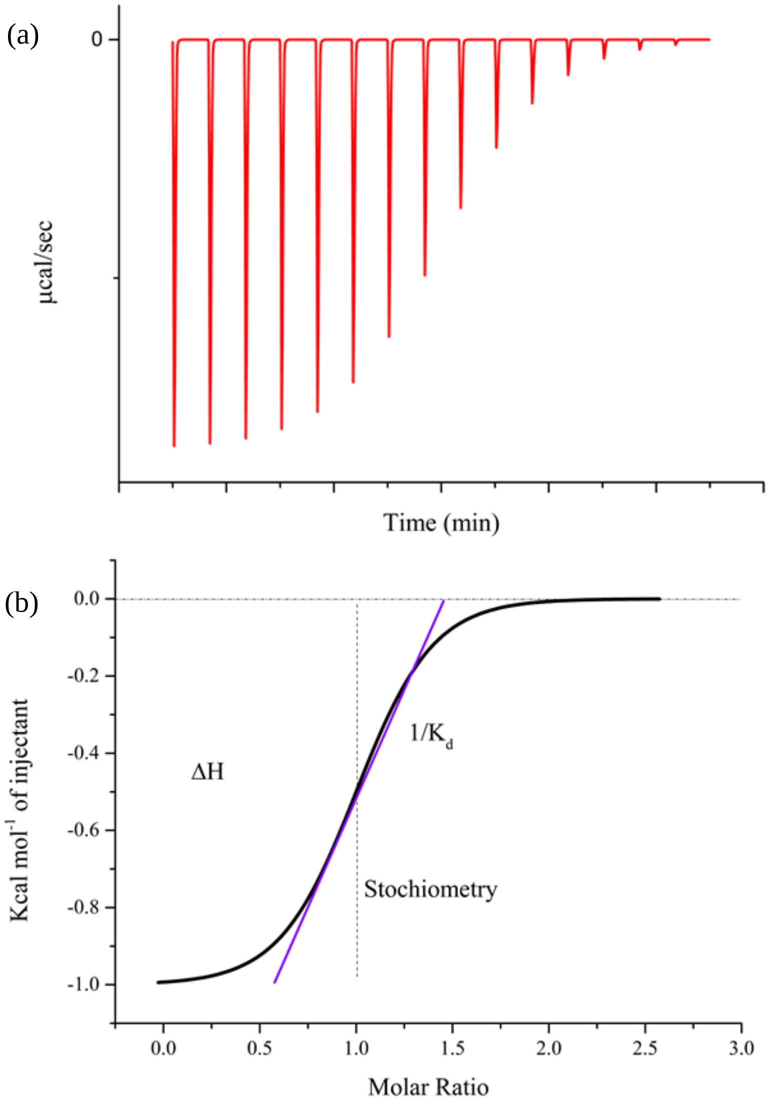


Figure (8): Schematic figures of the (a) raw data profile from an ITC experiment, showing spikes of inactivity in the calorimeter’s internal heater after each titration;

demonstrating an exothermic binding reaction, and (b) the resultant titration curve extracted from figure (a), annotated with a depiction of the dissociation constant.

Ligand and protein solutions are prepared such that, with the first injection, the relative concentration of ligand to protein is low, and the excess of protein allows the majority of ligands to bind, causing a relatively large temperature change. With each subsequent injection of the ligand solution, the fraction of unbound protein molecules in the cell decreases, and thus the temperature change also decreases. For an ideal ligand, i.e. a strongly-binding ligand that binds stoichiometrically (1:1), the protein will be in an overwhelmingly bound-state at an equimolar mixture of protein and ligand, and the temperature change upon further titration comes chiefly from dilution of the cell volume. In this case, the temperature change from binding quickly saturates. This scenario provides an ideal titration curve, such as that in Figure (8), from which important parameters may be extracted. The enthalpy, ΔH, describes the heat change in the system with 100% binding. Most significantly, however, it is possible to extract a quantitative measure of the intrinsic binding strength between the protein and ligand. The dissociation constant of the system (K_d), and its reciprocal, the association constant (K_a), describe the concentrations of free protein and ligand, relative to that of the protein-ligand complex.

$$K_d = \frac{1}{K_a} = \frac{[P][L]}{[C]} \tag{4}$$

Where [P] is the protein concentration, [L] is the ligand concentration and [C] the complex concentration, i.e. [PL]. K_d and K_a can be obtained from the rate of change of complex concentration (inferred from the change in heat release) as function of the free concentrations of protein and ligand (known experimental values).

For the purposes of the IR investigation in Chapter (5), knowledge of the dissociation constants of each protein-peptide complex was vital, not only to ascertain whether structural modifications of the peptide had an effect on the binding affinity, but also to the design of further measurements; alternate experimental approaches used in

Chapter (5) required either the maximisation of bound peptide or the maximisation of bound protein, whilst minimising the quantity of material used in either case.

Using the measured K_d, the concentration of the complex can be derived for any concentration of the free protein and ligand, allowing accurate experimental design; when [P] and [L] are expressed in terms of their total concentration ($[P_{total}] - [C]$) and ($[L_{total}]$-[C]), respectively, equation (2) becomes quadratic.

$$0= [C]^2+ ([P_{total}]+[L_{total}]+K_D)[C] + [P_{total}][L_{total}] \tag{5}$$

The resulting complex concentration can therefore be calculated using:

$$[C] = \frac{([P_{total}]+[L_{total}]+K_d) - \sqrt{([P_{total}]+[L_{total}]+K_d)^2 - 4([P_{total}][L_{total}])}}{2} \tag{6}$$

Peptide-protein binding constants were measured using a MicroCal ITC 200. A 300 μM protein solution of volume 0.2 ml was stabilised at a temperature of 25 ˚C, and peptide from a 10 mM stock solution was injected into the cell in 2.43 μl volumes. The temperature stabilisation time between injections was set to 3 minutes. Binding was seen to be exothermic.

Initially, binding constants for a particular protein-peptide system were obtained in both H_2O and D_2O in order to determine potential differences, though no significant variations were observed. With respect to ITC, solvent-mismatching between the cell and the titrant solution can create substantial baseline artefacts from dilution effects[17]. For example, when adding successive volumes of a titrant with 20% DMSO, 80% H_2O to a cell with 10% DMSO and 90% H_2O, the dilution effect is non-uniform across the experimental range, as the % DMSO in the cell increases. In these measurements, solvent-mismatching between H_2O and D_2O invariably occurs when D_2O is employed as the solvent, as protein molecules contain large numbers of exchangeable protons with which to slightly dilute D_2O. This is also true of the peptide, though the smaller size of this molecule lessens the effect. In an aqueous

(H_2O) system, the solvent is “pure”, and matched between the cell and titrant, providing a reference for the effects of H_2O/D_2O mismatching. Measurement in D_2O was preferable to maintain absolute consistency with IR samples.

A.6 References

1. Cervetto, V., Helbing, J., Bredenbeck, J. & Hamm, P. Double-resonance versus pulsed fourier transform two-dimensional infrared spectroscopy: an experimental and theoretical comparison. *J. Chem. Phys.* **121,** 5935–5942 (2004).
2. Deflores, L. P., Nicodemus, R. A. & Tokmakoff, A. Two-dimensional fourier transform spectroscopy in the pump-probe geometry. *Opt. Lett.* **32,** 2966–2968 (2007).
3. Hamm, P. & Zanni, M. T. *Concepts and methods of 2D infrared spectroscopy*. (Cambridge University Press, 2011).
4. Nienhaus, K. & Nienhaus, G. U. in *Probing heme protein-ligand interactions by UV/visible absorption spectroscopy* **305,** 215–241 (Humana Press, 2005).
5. Howes, B. D. *et al.* Mutation of residues critical for benzohydroxamic acid binding to horseradish peroxidase isoenzyme C. *Biopolymers* **62,** 261–267 (2001).
6. Heering, H. A., Smith, A. T. & Smulevich, G. Spectroscopic characterization of mutations at the Phe41 positionin the distal haem pocket of horseradish peroxidase C: structural and functional consequences. *Biochem J* **579,** 571–579 (2002).
7. Göbl, C. & Tjandra, N. Application of solution NMR spectroscopy to study protein dynamics. *Entropy* **14,** 581–598 (2012).
8. William A Catterall, Donald P Hollis, C. F. W. Nuclear magnetic resonance study of the conformation of nicotinamide-adenine dinucleotide and reduced nicotinamide-adenine dinucleotide in solution. *Biochemistry (Mosc.)* 4032–4036 (1968).
9. Iwaki, M., Cotton, N. P. J., Quirk, P. G., Rich, P. R. & Jackson, J. B. Molecular recognition between protein and nicotinamide dinucleotide in intact , proton-translocating transhydrogenase studied by ATR-FTIR spectroscopy. 517–523 (2006).
10. Pettegrew, J. W., Miles, D. W. & Eyring, H. Circular dichroism of adenosine dinucleotides. *Proc. Natl. Acad. Sci. U. S. A.* **74,** 1785–1788 (1977).
11. Chan, W. C. & White, P. D. *Fmoc solid phase peptide synthesis: a practical approach*. (Oxford University Press, 2000).

12. Dunn, B. M. & Pennington, M. W. *Peptide analysis protocols*. (Humana Press, 1994).
13. Pennington, M. W. & Dunn, B. M. *Peptide synthesis protocols*. (Humana Press. 1994).
14. Aguilar, M. I. *HPLC of peptides and proteins*. (Springer, 2004).
15. Ho, C. S. *et al*. Electrospray ionisation mass spectrometry: principles and clinical applications. *Clin. Biochem. Rev.* **24,** 3–12 (2003).
16. Leavitt, S. & Freire, E. Direct measurement of protein binding energetics by isothermal titration calorimetry. *Curr. Opin. Struct. Biol.* **11,** 560–566 (2001).
17. Rajarathnam, K. & Rösgen, J. Isothermal titration calorimetry of membrane proteins — Progress and challenges. *Biochim. Biophys. Acta* **1838,** 69–77 (2014).

† Image reference (Figure (1)):

http://chemwiki.ucdavis.edu/Wikitexts/UC_Davis/UCD_Chem_115_Lab_Manual/Lab_7%3A_Electrospray_Mass_Spectrometry, obtained 14/07/2015

Acknowledgements

Many thanks to my PhD supervisor, Dr Neil Hunt, for his input and guidance over the last few years, and more recently for his feedback throughout the progression of this thesis. Thanks to my past and present colleagues in the UCP group; Marco, Pim, Katrin, Danny, Gordon, Paul and Lennart. Many thanks also to John Revie for so frequently going out of his way for everyone in our research group.

Much of this work would have been unobtainable without the support of Tony Parker, Mike Towrie, Greg Greetham and Ian Clark at the Rutherford Appleton Laboratory, and your efforts are much appreciated. To Dom and Martin, I also offer my thanks for your commitment to such a challenging project.

I am very grateful for the work of our collaborators in SIPBS: Kirsty, Cesar, Paul and Nick, with whom the UCP group has accomplished much.

To our collaborators in Chemistry: Morag and Craig, I offer my sincere thanks for your time and patience.

Many thanks to my parents, my grandmother, and my brothers, Douglas, Robbie and Fergus, for their support during the writing of this thesis, and for their interest in my research.

I am very grateful for the support of my friends over the last several months; in particular those who offered to help me with proofreading, horribly misguided though that was.

Most of all, I would like to thank my fiancée, Haley, for her constant positivity and motivation during the writing of this thesis, and during my PhD.