Electron-Vibration-Vibration Two-Dimensional Infrared Spectroscopy as a structural probe of interactions in proteins and DNA

Hugh Richard Sowley

Doctor of Philosophy

Institute of Chemical Biology, Department of Chemistry

Imperial College London

Abstract

Modern structural biology has a number of powerful tools but despite this there are a number of problems in structural biology that these methods are unable to address. Some of these pertain to the need for large number for precise comparative structures for the drug discovery process. EVV 2DIR has the potential to fill some of these gaps, having the potential to determine molecular binding geometry. This thesis presents the first steps in exploring the potential of EVV 2DIR to be applied to the analysis of the structure of inhibitor- protein binding and presents the first EVV spectra of an inhibitor-protein complex.

For the inhibitor-protein complex studied, six vibrational couplings between seven vibrational modes were identified exclusively upon complex formation due to interactions between the two molecules. Experimental spectra were compared with *ab initio* calculations to assign these vibrations to specific motions on both the inhibitor and protein molecules. EVV 2DIR cross peaks can be sensitive to the geometry of the interacting groups which produce them. By measuring the spectra of the inhibitor-protein complex using two different polarisation schemes, quantitative comparison between calculated and experimental spectra was made possible. This allowed for the prospect of using calculation aided EVV 2DIR to determine the structure of protein-ligand complexes to be explored.

This thesis also presents the first EVV spectra of DNA. EVV 2DIR spectra were measured of duplex and G-quadruplex structures and compared with those of unstructured controls. In the absence of calculated spectra, assignments were made to some of the spectral features observed. EVV 2DIR was shown to be sensitive to the structural form of the DNA samples, containing cross peaks indicative of Watson-Crick base pairing, G-quadruplex formation and glycosidic bond conformation. The DNA spectra contained many unassigned peaks leaving open the possibility to assign many more structural indicators.

Declaration of originality

All of the work presented in this thesis is my own, aside for the following exceptions:

Shababa Selim collected the circular dichorism data presented in Chapter 3.

The calculated EVV 2DIR spectra presented in Chapter 4 were produced by the group of Wei Zhuang at The Dalian Institute of Chemical Physics and at The Fujian Institute of Research on the Structure of Matter at the Chinese Academy of Sciences.

Copyright Declaration

Acknowledgments

I would like to express my thanks to both of my academic supervisors David Klug and Keith Willison for both the opportunity to work on this project and also for their consistent support and advice throughout its duration. I would like to thank Geoff Holdgate and Astra Zeneca for their funding and supervision and for supplying the recombinant protein that was used for the main body of this work.

The calculations provided by Wei Zhuang and his group were incredibly valuable for making sense of the complex experimental data and I am very grateful to have been able to collaborate with them.

Finally I would like to express my gratitude to my friends and family, in particular to my parents Sheila and Keith for their continued encouragement and support.

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 Detecting and characterising biomolecular interactions and structures

The ability to detect biomolecular complexes, determine their geometries and characterise the interactions which give rise to them is crucial to our understanding of biology and our capability to intervene in it. Intramolecular interactions in biomolecules ultimately determine their complex structures, with their structures allowing specific intermolecular interactions and molecular recognitions to occur. Such specific intermolecular interactions govern all of biology. Within the human proteome, consisting of around 20,000 proteins, there are an estimated 650,000 distinct protein–protein interactions[1]. Over 4,000 of the human proteins are estimated to interact with DNA[2], with many of these responsible for regulating gene expression. Key also in the gene expression process is the specific interactions between nucleotide base pairs that allow for mutual recognition between them. Specific interactions between these base pairs facilitate the formation of several distinct DNA secondary structures, a couple of examples of which are considered later in this thesis, along with the preliminary results of using EVV 2DIR as a structural probe to distinguish them.

An enzyme's ability to recognise and selectively bind to its specific substrates is of course also mediated by a specific set of interactions between the two. In the same way, the vast majority of small molecule inhibitors' efficacy is based upon them having a specific structure capable of selectively forming distinct interactions with specific enzymes. It is therefore of great importance to be able to develop a precise understanding of the structure and nature of small molecule inhibitors' interactions with their target protein, in order to be able to intelligently develop effective therapeutics. This is not only true for inhibitor screening and hit validation, the process of determining if a small molecule is in fact binding to a target protein can also help lead optimisation[3].

Given the importance of knowing the structures of, and geometry of interaction between inhibitors and proteins, and also protein–protein interactions, numerous methods have been developed to determine such information:

X-ray crystallography is currently by far the most widely used and developed method for determining such structures with 120,000 crystal structures in the RCSB Protein Data Bank. There are large infrastructures in place allowing for high throughput structural determination; however there are numerous limitations to the technique, both practical and in terms of the information produced. For many purposes, protein crystals can act as an excellent model for the structure of a protein, however they represent an averaged un-hydrated crystal geometry, not the true solvated form. As a result of this, crystallography may miss any kind of protein dynamics which can be crucial to the mechanism of action of a given inhibitor. Water molecules frequently play important roles in facilitating inhibitor–protein interaction[4,5,6], but are often excluded from protein crystals and there can be difficulties differentiating between salts and the water molecules which have been included within the crystals[7]. In cases where water molecules are observed, it is not uncommon to see "blobs" of electron density which are modeled to be multiple water molecules, whose geometry is uncertain.[8] There are many natively unstructured proteins and membrane bound proteins which are not possible to crystalise[9]. The same is also true for dynamic regions of ordered protein such as the disordered loops which frequently play important regulatory roles in kinases[10]. Furthermore, X-ray crystallography requires relatively large amounts of protein for the production of crystals, making screening

large libraries of potential inhibitors protein expensive. Although it is still absolutely critical to biological sciences, a requirement for complementary techniques to crystallography has emerged.

Since the advent of protein crystallography, multidimensional NMR techniques have perhaps had the largest impact in terms of elucidating information about the molecular structure of biological systems. NMR techniques measure the parameter chemical shift which is highly sensitive to the local environment of the atom. As a result of this it is possible to determine which parts of which molecule are proximal to which parts of other molecules. It is easy to see how this could be incredibly valuable for determining if a ligand has bound to a protein, and if so in what geometry[11]. The relatively simple physics governing the spin couplings observed in multidimensional NMR make it now possible to computationally invert some spectrums to yield structures[12]. As these techniques are performed in solution, they are able to yield structural information about unstructured or membrane proteins which do not lend themselves well to crystallization[9]. Multidimensional NMR is not however, also without its limitations. Its low sensitivity results in a requirement of large amounts of material making it less suitable for screening large libraries of compounds. NMR is also restricted by its requirement for rotational averaging, which prevents it from being able to study dynamic processes occurring over short time scales. This results in a time averaged view of the system, which is inherently insensitive to potentially significant rapid motions[13]. There are also intrinsic limits on the size of protein systems which can be structurally assessed in this way. These limitations present an opportunity for the development of additional complementary techniques for the study of interactions between biomolecules and their structures.

The main body of this thesis describes the work done to develop EVV 2DIR as a tool for detecting the formation and determining the geometry of inhibitor-protein complexes by detecting coupled vibrations arising from interactions between the inhibitor and the protein binding site. A description of the EVV 2DIR technique is given in Chapter 2.

Other forms of coherent multidimensional spectroscopy have been applied to the problem of detecting and characterising intermolecular interaction of proteins with a

view to offer structural and dynamical information, complementary to that offered by the previously discussed techniques:

Two-dimensional vibrational photon echo spectroscopy has been used to study the complex of HIV-1 reverse transcriptase and its inhibitor rilpivirine. The fact that the inhibitor contains two nitrile arms, which have readily observable characteristic vibrational spectra, allowed them to be used as environment sensitive vibrational probes[14], showing different spectral properties in bulk water compared to when specifically bound to the enzyme. Using the nitrile arms as vibrational probes in this way, a hydrogen bond between one of the arms and a conserved water molecule was observed. This water molecule had not been included in previous crystal structures and its detection here demonstrates multidimensional vibrational spectroscopy's ability to provide valuable structural information about protein inhibitor interactions[15]. A similar approach was used by the same group to study a very similar inhibitor, also featuring nitrile arms for use as vibrational probes, in complex with HIV-1 reverse transcriptase. Spectral dynamics were observed which were attributed to vibrational couplings between protein vibrations and inhibitor vibrations.[16]

Vibrational probes have previously been artificially incorporated into biomolecules in order to allow for intermolecular binding to be detected and characterized by 2DIR spectroscopy[17]. By labeling individual residues of a target peptide of the PDZ2 domain of a human protein tyrosine phosphatase enzyme with an azide vibrational probe, the binding of the two could be detected by observing changes in the extinction coefficient of the spectral contributions of the probe.

As well as the use of vibrational labels such as the nitrile and azide groups used in the above examples, isotopic labels have also previously been employed for the use in 2DIR studies of biomolecular complexes. Doing so shifts the frequency of the labeled moieties away from the unlabeled equivalents, allowing them to be independently considered. By incorporating site specific heavy atom carbonyl groups into an influenza proton channel protein, the dynamics of the role of water in the binding of an inhibitor were elucidated using 2DIR[18].

At the time of writing, there has not been published accounts of the development of a multidimensional vibrational spectroscopy method to be used as a label free

structural probe for general protein–inhibitor complexes, without the need for coincidental vibrational probes. The main part of this thesis describes the work done to explore EVV 2DIR's capacity to achieve this.

1.2 FGFR1 and SU-5402

The first biological system that was studied using Electron-Vibration-Vibration Two-Dimensional Infrared Spectroscopy for this thesis was the complex of the kinase Fibroblast Growth Factor Receptor 1 (FGFR1) and its inhibitor SU-5402. The aim of this work was to demonstrate the technique's ability to detect the formation of inhibitor-protein complexes, characterise the interactions between the two and determine structural geometry. A brief overview of the biological significance of these molecules is provided below.

Protein kinases are enzymes which catalyse phosphorylation: the transfer of a phosphate group from an ATP molecule onto a specific amino acid residue on a substrate protein [19]. The human genome codes for 518 different protein kinases[20] that can been grouped into two major classifications, defined by the amino acid residues which they phosphorylate: tyrosine kinases and serine-threonine kinases[20]. Phosphorylation of a substrate protein may induce a conformational change resulting in modulation of its activity. Protein kinases modulating the activity of other protein kinases forms the backbone of cellular signal transduction pathways, a critical part of many cellular process[21]. Tyrosine kinases can then be classified as either receptor or non-receptor kinases. Receptor kinases are transmembrane proteins consisting of an intracellular kinase domain and an extracellular ligand binding domain. Extracellular ligand binding elicits activation of the kinase and allows for intracellular phosphorylation to occur. In this way, signaling pathways can be transduced across cell membranes.[22] Non-receptor proteins do not have an extracellular domain as they are located within the cell itself. The kinase studied in the work presented here, Fibroblast Growth Factor Receptor 1 (FGFR1), falls into the classification of a tyrosine receptor kinase. Note that FGFR1 construct studied here only includes the kinase

domain, not the transmembrane domain which will be largely unstructured if not embedded in a membrane.

There is a large degree of conservation of certain residues in the catalytic domain of eukaryotic protein kinases, and as a result of which many similar folds are observed along with a degree of structural conservation. This presents challenges when designing inhibitors to exclusively target specific kinases[23]. The catalytic domain of kinases can be considered as two lobes, termed the N and C-terminal lobes. The N-terminal lobes consist mostly of β-sheets with a single α-helix called the C-Helix, whilst the C-terminal lobe is comprised predominantly of α-helices. Between these two lobes, a deep cleft is formed, into which ATP binds. It is to this ATP binding site that SU-5402 binds, making it a competitive inhibitor. The structure of ATP and SU-5402 can be seen in Figure 1-1 and some degree of structural similarities can be seen between the two, allowing SU-5402 to be considered an ATP mimic.

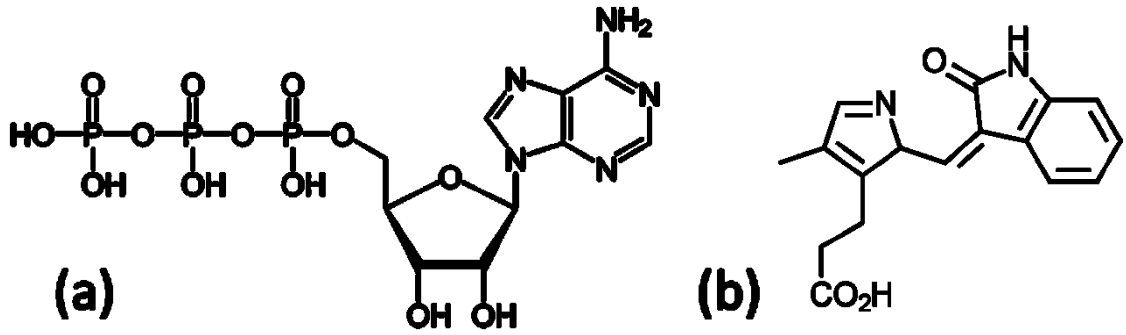

Figure 1-1 the structures of and similarities between (a) Adenosine Triphosphate (ATP) and (b) SU-5402

The crystal structure of the FGFR1 SU-5402 complex has been determined[24] and its cartoon representation showing secondary folds can be seen in Figure 1-2. It can be seen that SU-5402 has bound between the two main lobes of the kinase domain. Figure 1-3 shows the protein surface and illustrates how the ATP binding site, to which SU-5402 is bound, is exposed.

Figure 1-2 The cartoon representation of the crystal structure[24] of the FGFR1 SU-5402 complex. The SU-5402 molecule (orange) can be seen to be sat between the two main lobes of the protein fold (blue).

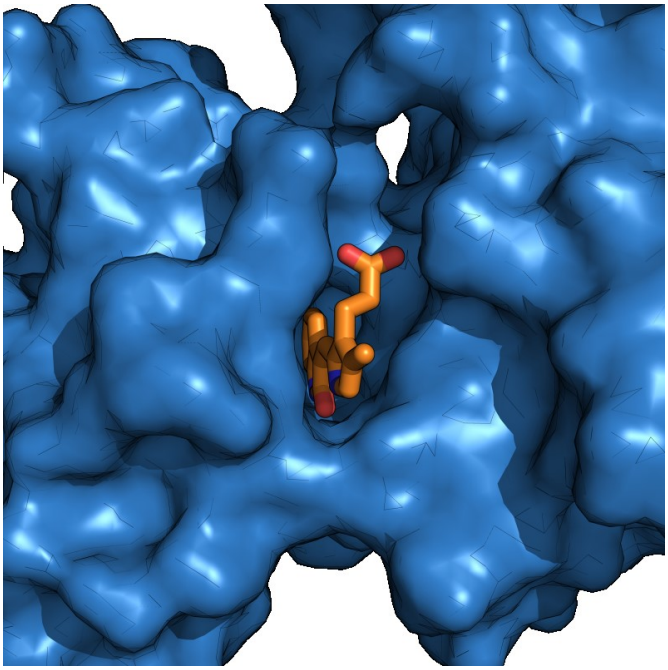

Figure 1-3 Depiction of the protein surface of the FGFR1 SU-5402 complex, showing SU-5402 (orange) to be sat within the exposed cleft of the ATP binding site.

FGFR1 is a member of the fibroblast growth factor receptor family, for which there are four distinct genes. There are however 48 distinct isoforms of FGFRs observed as a result of alternative splicing[25]. FGFR1 has presented itself as a drug target as it is implicated in a number of diseases and conditions including Alzheimer's disease[26,27], chronic obstructive pulmonary disease[28], inflammatory renal disease[29] and Parkinson's disease[30], to name but a few. Deregulation of FGFR activity results in abhorrent signaling which has been identified as a cause and hallmark of numerous human cancers[31]. As a result of this, many FGFR1 inhibitors have been developed, although the difficulties lie in developing inhibitors which target only FGFR1 due to considerable structural conservation between kinases[32]. Crystallography has been used as a method of tackling the issue of specific kinase selectivity in the kinase inhibitor design process; by comparing the structures of a non-selective inhibitor in complex with FGFR1 and those of non-target kinases, structurally informed refinements can be made to the inhibitor to yield selectivity[33].

In addition to these crystallographic studies, the FGFR1 protein has been studied using 2D NMR techniques. Its NMR spectra have been assigned and characterized for both the kinase domain[34] and the extracellular ligand binding domain[35]. This work was performed with the motivation of allowing for structural and dynamic changes upon inhibitor binding to be readily interpreted.

One of the ways in which the activity of protein kinases is modulated is by the position of the DFG motif of the activation loop of the kinase domain[36]. In the active state, the DFG motif is buried in the protein, forming what is known as the DFG-in state. Conversely, in the inactive state the DFG motif is rearranged such that it is exposed out of the protein, forming the DFG-out state. FGFR1 inhibitors have been developed which target both the DFG-in (type 1 inhibitors) and DFG-out states (type 2 inhibitors)[37]. By using both type 1 and type 2 FGFR1 inhibitors (including SU-5402) as chemical probes to selectively lock the protein in to either the DFG-in or out state the dynamics of the DFG-in to out flip have been studied using isothermal titration calorimetry and surface plasmon resonance[32]. The result of this work was the

suggestion that FGFR1 has a low propensity for the DFG-out form which potentially rationalizes why more type 1 FGFR1 inhibitors have been developed.

SU-5402, the inhibitor used in the studies presented here, is a type 1 kinase inhibitor, that is it locks the DFG loop in the "in" position. SU-5402 was initially used as a precursor compound in the development of SU-6668[38], but now is commonly used as a tool for studying the signaling pathways of FGFRs[31]. It has been shown to display pan-kinase inhibition, inhibiting several other kinases with sub μM kDs. As well as the crystal structure of the SU-5402 – FGFR1 complex being known[24], surface plasmon resonance and isothermal titration calorimetry have been employed to determine the thermodynamic parameters of SU-5402 – FGFR1 binding, with the kD of the interactions being found to be 76nM[32].

SU-5402 was primarily chosen for this work due to its strong visible absorption at 430 nm, as shown in its UV/ Vis spectrum shown in Figure 1-4. Compounds with strong visible absorptions give considerably enhanced EVV 2DIR signals; this is explained in 2

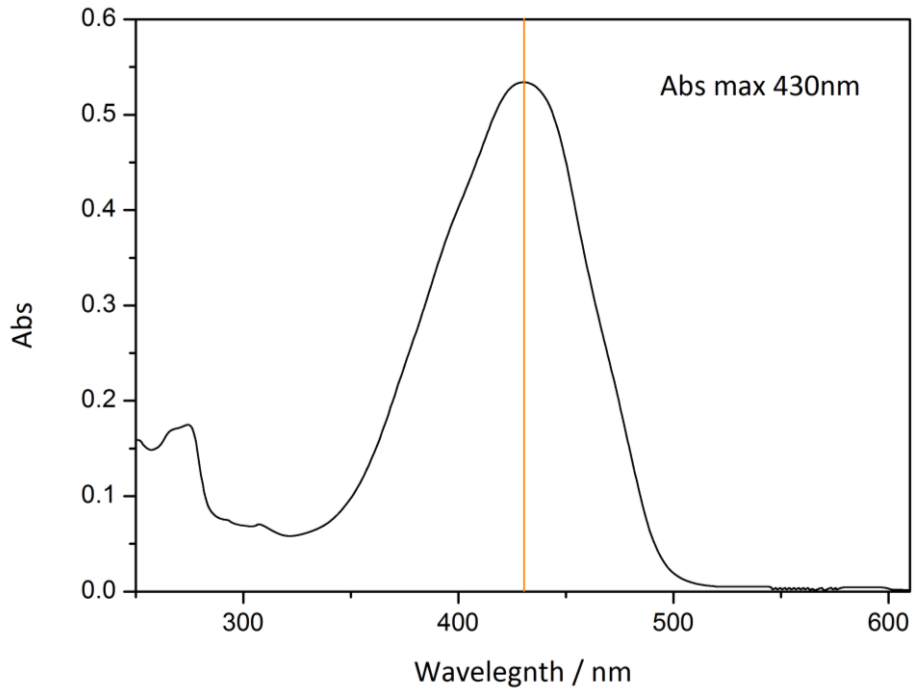

Figure 1-4 The UV/ Vis spectrum of SU-5402.

1.3 DNA structures

The latter part of this thesis presents the preliminary investigation into EVV 2DIR's sensitivity to different DNA secondary structures and its ability to detect and characterise the intra and intermolecular interactions which stabilize them. The two DNA secondary structures investigated were B-form duplex and G-quadruplex structures.

Deoxyribonucleic acid is found in all known living organisms and encodes the genetic information required for an organisms function, growth and reproduction. There are numerous different DNA conformations observed in nature, each stabilized through unique sets of inter and intra molecular interactions, including base to base hydrogen bonding and the interactions associated with base stacking[39]. The conformation of DNA most commonly thought of is the double-helix structure, formed between two strands with complementary sequence and famously first published by Watson and Crick in 1953[40]. In this duplex structure, a helical arrangement of nucleotide bases is observed, whilst the phosphodiester backbones run anti-parallel to one another. These double helix structures are facilitated by adenine and guanine binding to thymine and cytosine base pairs respectively, with the guanine-cytosine pairing stabilized by the formation of three hydrogen bonds and the adenine-thymine pairing stabilized by two, as illustrated in Figure 1-5.

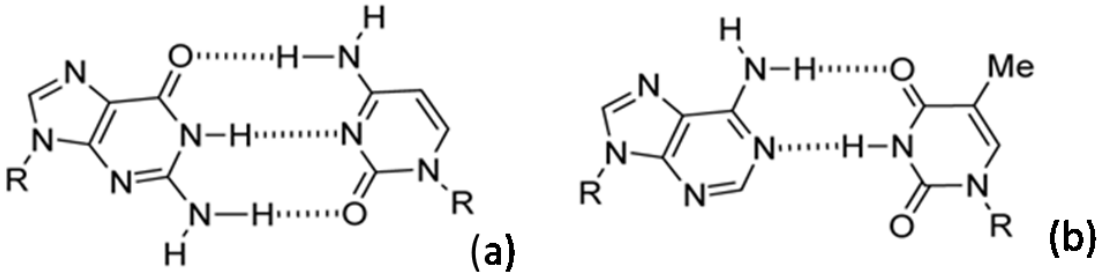

Figure 1-5 the hydrogen bond interactions between the (a) guanine-cytosine base paring (b) adenine-thymine base paring.

Other forms of 2DIR spectroscopy have been demonstrated to be sensitive to these Watson-Crick base pairings as well as the base stacking interactions which also stabilize the duplex forms:

The first reported application of ultrafast 2DIR spectroscopy to DNA was back in 2003[41]: vibrational couplings, corresponding to the intramolecular base pair interactions between guanine and cytosine carbonyl stretches were detected through the use of both pump probe and photon echo spectroscopies. Similar approaches have also been used to detect and characterise the N-H stretch in the guanine-cytosine base pair and the coupling between the N-H stretches and the carbonyl stretches as a result of the hydrogen bond between the two[42]. Ultrafast 2DIR spectroscopy has also been used to study the adenine-thymine Watson-Crick pairing[43,44]. In addition to the hydrogen bonding interactions between complementary bases, the interactions between stacked bases have also been detected using 2DIR techniques[45]. These base stacking interactions have shown to have a significant energetic contribution to the stabilisation of the duplex dimers[46]. As well as being used to study base-base interactions, 2DIR has been used to study the vibrational couplings between nucleotide bases and the sugar−phosphate backbone. This was achieved by pumping at the frequencies associated with base vibrations and probing at the frequencies associated with the backbone stretching region. Doing so, revealed duplex specific base-backbone couplings and also lead to a description of the dynamics of the intramolecular energy relaxation to the solvent[43].

This thesis presents two main sections of work: firstly to use EVV 2DIR to study a small molecule binding to large biomolecule and secondly to explore its sensitivity to DNA structure. An interesting application of pump-probe 2DIR has been reported which in a way combines these two elements. Using pump probe 2DIR, it was shown that the intercalating DNA stain Hoechst33258 induces structural changes in the duplex DNA minor groove to which it binds, with the relative orientation of the bases changing. The technique's sensitivity to hydrogen bonding made it possible to also see the exclusion of water molecules from the DNA minor groove, upon binding[47].

As well as investigating EVV 2DIR's ability to detect the formation of DNA duplexes, this thesis presents the preliminary investigation into its ability for detect the

formation of G-quadruplexes. G-quadruplexes are intramolecular DNA structures formed from π-stacked guanine tetrads, in guanine rich sequences of DNA. A guanine tetrad is a planar arrangement of four guanine bases stabilised by a cyclic system of eight inter-guanine hydrogen bonds[48] between the N_1-O_6 and N_2-N_7 with the O_6 atoms in guanine bases. This type of base pairing is called Hoogsteen base pairing and is not exclusive to homo-guanine interactions. A depiction of the guanine tetrad, with the Hoogsteen hydrogen bonds shown, can be seen in Figure 1-6.

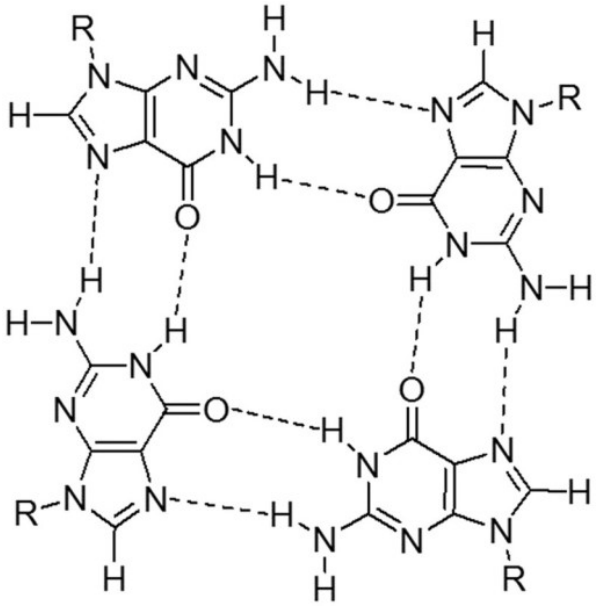

Figure 1-6 A guanine tetrad, showing the eight stabilising Hoogsteen pairing hydrogen bonds.

These guanine tetrads can then π stack on top of one another, stabilised by monovalent cations which sits in the central hollow between the tetrads. The cation mediated stacking of tetrads is depicted in Figure 1-7.

Figure 1-7 Schematic showing how the stacking of guanine tetrads is mediated by monovalent cations. The orange quadrilaterals represent the guanine bases of the tetrad and the blue circles represent the monovalent cations.

G-quadruplexes are most abundantly found in the guanine rich ends of eukaryotic chromosomes in the form of telomeres[49]. The primary function of telomeres is to protect chromosomes from recombination and becoming fused to one another. As DNA is replicated, the teleomeric DNA at chromosomal ends is shortened by between 50 - 100 bases per round of replication. This end replication effect has been shown to be one of the key contributors to the ageing process[50]. G-quadruplex structures are also key to the regulation of certain genes, with specific gene regulating proteins recognizing these structures[51].

The G-quadruplex forming DNA sequence used for the work described here was Myc2345, a guanine rich section of the promoter region of the c-Myc gene. The c-Myc protein, for which the gene encodes, is a nuclear phosphoprotein which plays a central role in cellular proliferation and cell growth. Abhorrent overexpression of this protein is implicated in many cancers[52]. Formation of G-quadruplexes in the promoter region, of which Myc-2345 is a part of, is responsible for around 80 % of the control of c-Myc expression[53].

The sequence of Myc2345 is given below:

$$\text{Myc2345} = \text{T}G_1\text{A}G_2G_3G_4\text{T}G_5G_6G_7G_8\text{A}G_9G_{10}G_{11}\text{T}G_{12}G_{13}G_{14}G_{15}\text{AA}$$

The 15 guanine bases, which are responsible for the quadruplex structure formation have been shown in red, and numbered with subscript numerals. These subscript numerals are labels for use with Figure 1-8, not an indicator of quantity. There are several different isoforms of G-quadruplex structures[54], all featuring the Hoogsteen facilitated guanine tetrads but only one has been considered in this work. Myc2345 forms an intramolecular parallel quadruplex. This structure of this is shown in Figure 1-8. Each of the orange quadrilaterals has been labeled with the specific guanine base from the Myc2345 sequence above. In this way it can be seen how the DNA strand folds up and down, parallel to the direction of tetrad stacking. Note that the first, 8th and last guanines in the sequence do not form part of the three tetrads.

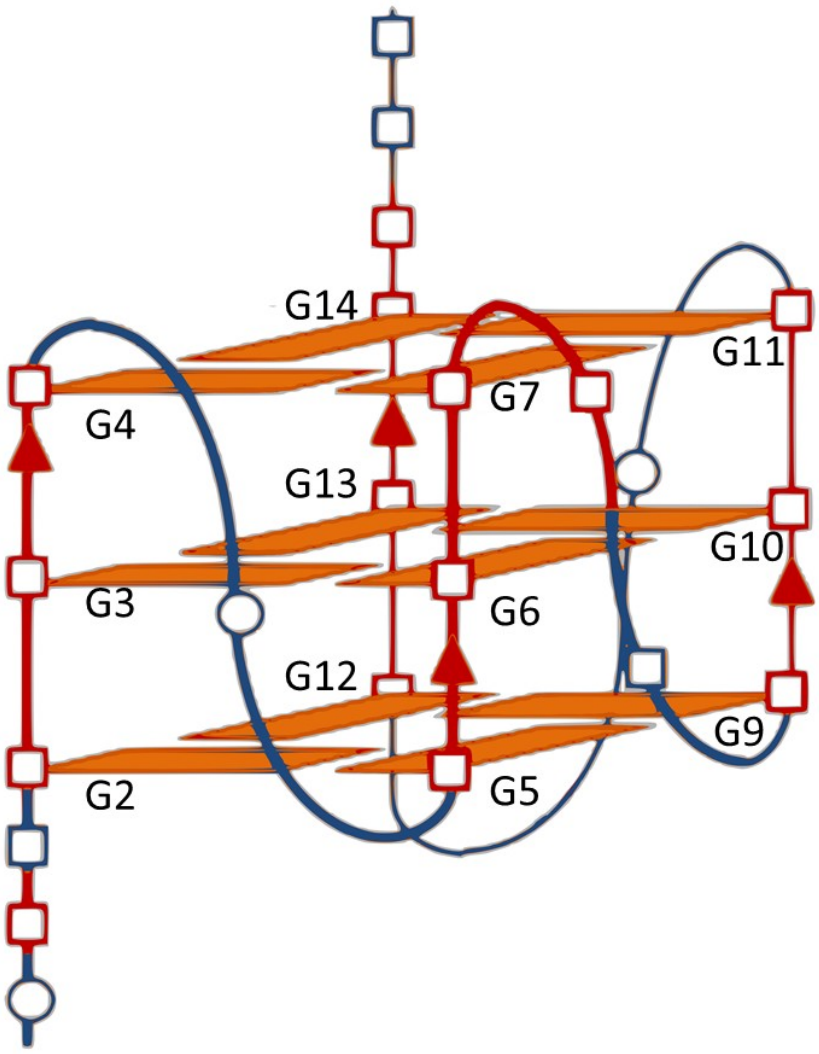

Figure 1-8 The structure of the parallel intramolecular G-quadruplex structure formed by Myc2345. The orange quadrilaterals represent the individual guanine bases which have been labeled in accordance with the labels in the sequences given above. The red and blue line shows how the sequence folds, with red indicating the sequence progressing upwards and blue downwards, from 5' to 3'. Stabilising monovalent cations have not been included. This structure was reported in reference [55].

The crystal structures of G-quadruplexes have been determined for numerous G-quadruplex forming sequences[56], however not for Myc2345. The structure of Myc2345 shown in Figure 1-8 was elucidated using NMR techniques[55]. Of course the same limitations of both techniques as discussed in the context of studying inhibitor-protein complexes apply here too.

G-quadruplexes have previously been investigated using vibrational spectroscopy, both experimentally and through calculations, with spectral differences observed upon quadruplex formation corresponding to the Hoogsteen base pairing - which is encouraging for EVV 2DIR's ability to detect such structures[57,58]. At the time of writing, there have been no published examples of the use 2DIR spectroscopy as a tool to study G-quadruplex structures.

2 Electron-Vibration-Vibration Two Dimension Infrared Spectroscopy

2.1 Introduction to vibrational spectroscopy

Molecules are not internally static and display vibrational motion of their nuclei relative to one another about an equilibrium position. These vibrations can be approximated classically, by thinking of the nuclei as masses connected by springs. Increasingly complex molecules will display increasingly complex relative motions of their nuclei, but these complex motions can be factorised into vibrational modes which can be thought of, for many purposes, as orthogonal and independent of one another. This treatment of masses on springs, vibrating along independent modes, yields harmonic motion of the nuclei and can be used as a model for considering linear infrared spectroscopy, but fails to explain the phenomena intrinsic to the technique

which forms the basis of this thesis. Linear infrared spectroscopy involves measuring the absorption as a function of frequency, of infrared radiation incident upon a sample. Absorption lines in a linear infrared spectrum correspond to excitation of specific vibrational modes and their frequency can be characteristic of specific groups of atoms, or functional groups. In order for a mode to be excited in this way, there must be a change in the dipole moment of the molecule associated with the nuclear motions of the mode. In this way, linear infrared spectroscopy has proved useful for identifying and characterising molecules and the functional groups which comprise them.

As the number of atoms in a molecule grows, so does the number of vibrational modes it can display, with the number of modes given by 3N-6, for nonlinear molecules, where N is the number of atom. As a result of this, increasingly large molecules display increasingly complex infrared spectra. The convoluted linear spectra of complex systems, along with its inability to distinguish coupled vibrational modes, which is not accounted for in these harmonic, masses on springs description, form part of the motivation of performing two dimensional infrared spectroscopy.

The following discussion of the harmonic oscillator model has been adapted form referenc [59]. The potential energy of a molecule increases as its nuclei are moved away from their equilibrium positions. When considering a diatomic molecule, the potential energy can be considered in terms of one quantity, *x*, the displacement of the two atoms away from their equilibrium position. For small values of *x,* the potential energy of the diatomic can be expressed as the first three terms of the Taylor series given in Equation 2-1:

$$V(x) = V_{(0)} + \left(\frac{dV}{dx}\right)_0 x + \frac{1}{2}\left(\frac{d^2V}{dx^2}\right)_0 x^2$$

Equation 2-1

Where V is potential energy and $V_{(0)}$ is the potential energy of the system at equilibrium nuclear displacemnet. The absolute potential energy of the diatomic is

not of interest here, only how it changes with nuclear displacement, so the first term with no x dependence can be ignored. The subscript zeroes denote that the derivatives are to be taken at the equilibrium position, so the second term, directly proportional to x, will be zero, given that the potential energy goes through a minimum at the equilibrium position. Equation 2-1 can therefore be reduced to give Equation 2-2

$$V(x) = \frac{1}{2}\left(\frac{d^2V}{dx^2}\right)_0 x^2$$

Equation 2-2

Equation 2-2 described a parabolic relationship between potential energy and displacement away from equilibrium position, yielding evenly spaced energy levels as shown in Figure 2-1

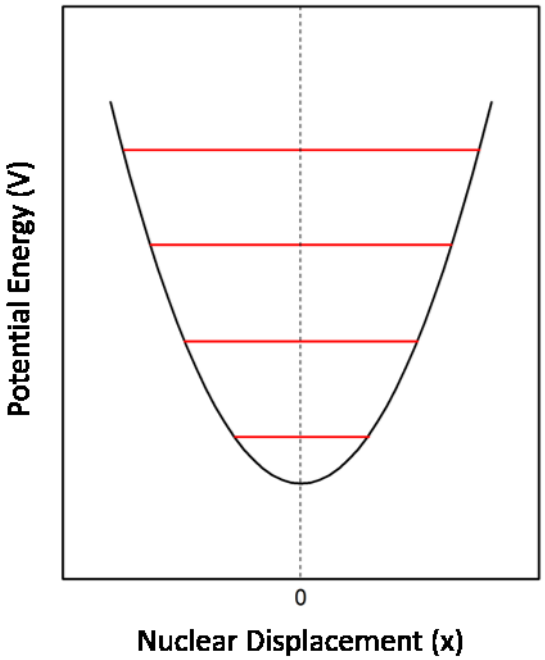

Figure 2-1 A harmonic energy well showing evenly spaced energy levels

A similar approximation can be used to describe the dependence of dipole moment of a diatomic on nuclear displacement of a diatomic, allowing for an understanding of the gross selection rule for vibrational transitions: To be IR active, and display an IR spectrum, a molecule must have a dipole moment that varies with extension.

The intensity of a transition from vibrational energy level to another (from v to v'), $\boldsymbol{\mu}_{v'v}$ is proportional to the square of the transition dipole moment, which is given by Equation 2-3:

$$\boldsymbol{\mu}_{v'v} = \langle v'|\boldsymbol{\mu}|v\rangle$$

Equation 2-3

Where $\boldsymbol{\mu}_{v'v}$is the transition dipole moment and $\boldsymbol{\mu}$ is the dipole operator. As the dipole moment of a molecule varies as a function of nuclear displacement, it can be expanded as a Taylor series with respect to nuclear displacement (when considering a diatomic model), x, to give Equation 2-4 below:

$$\boldsymbol{\mu} = \boldsymbol{\mu}_0 + \left(\frac{\mathrm{d}\boldsymbol{\mu}}{\mathrm{d}x}\right)_0 x + \left(\frac{\mathrm{d}\boldsymbol{\mu}^2}{dx^2}\right)_0 x^{\mathbf{2}} + \cdots$$

Equation 2-4

Where $\boldsymbol{\mu}_0$ is the dipole moment of the system at zero displacement. The elements of the transition matrix for transitions from levels with quantum numbers v to v' can therefore be given by substituting Equation 2-4 into Equation 2-3:

$$\langle v'|\boldsymbol{\mu}|v\rangle = \boldsymbol{\mu}_0\langle v'|v\rangle + \left(\frac{\mathrm{d}\boldsymbol{\mu}}{\mathrm{d}x}\right)_0 \langle v'|x|v\rangle + \left(\frac{\mathrm{d}\boldsymbol{\mu}^2}{dx^2}\right)_0 \langle v'|x^2|v\rangle + \cdots$$

Equation 2-5

The first term becomes zero when $v \neq v'$ due to the two states being orthogonal and for small nuclear displacements, such as those where the harmonic oscillator model is most appropriate, the quadratic and higher terms can be neglected, simplifying Equation 2-5 to Equation 2-6

$$\boldsymbol{\mu}_{v'v} = \left(\frac{\mathrm{d}\boldsymbol{\mu}}{\mathrm{d}x}\right)_0 \langle v'|x|v\rangle$$

Equation 2-6

Under these conditions it can be seen that the transition dipole moment is only non zero when the dipole moment varies as a function of nuclear displacement. This explains why homonuclear diatomics do not produce infrared spectra.

2.1.1 Anharmonicities and vibrational mode coupling

The harmonic description of the potential energy of a diatomic, given above, is only an approximation and real molecular systems display some degree of anharmonicity. This is particularly true at larger nuclear displacements seen at higher excited vibrational states. There are two types of anharmonicity that are exhibited by molecular systems in this context: Mechanical and electrical anharmonicity. The following discussion of these anharmonicities is an adapted version of that which can be found in reference [59].

Electrical anharmonicity is present when there is a nonlinear relationship between the dipole moment of the molecular system and the nuclear displacement away from the equilibrium position. This is contrary to a harmonic system in which the two are linearly related. The result of electrical anharmonicity is to change the intensity of vibrational transitions, but unlike mechanical anharmonicity, does not affect the energy level spacings and frequencies at which transitions occur.

To consider the effects of electrical anharmonicity on polyatomic molecules, we must use displacements of groups of atoms along normal mode coordinates instead of individual atomic positions (which were sufficient for the diatomic case). In this case we can rewrite Equation 2-4 in terms of motion along normal mode coordinates, *Q*, to give Equation 2-7.

$$\boldsymbol{\mu} = \boldsymbol{\mu_0} + \sum_i \left(\frac{\partial \boldsymbol{\mu}}{\partial Q_i}\right)_0 Q_i + \frac{1}{2}\sum_{i\,j} \left(\frac{\partial^2 \boldsymbol{\mu}}{\partial Q_i \partial Q_j}\right)_0 Q_i Q_j + \ldots \ldots$$

Equation 2-7

Where the summation it taken to be over all normal mode coordinates, *i*,*j*,*k* and so on. Electrical anharmonicity is present when there are non zero values of the cubic and higher terms of Equation 2-7 (those which were discarded in the harmonic diatomic model described above). As can be seen in Equation 2-7, there are now terms which are proportional to $Q_i Q_j$ (and higher). The presence of these quadratic (and higher) terms removes the orthogonality of the normal modes, giving non zero transition dipole moments for sates composed of the combination of two normal modes. These anharmonicities allow for the appearance of combination ($i \neq j$) and overtone ($i=j$) bands in the infrared spectra. That is the excitation of a mode at a frequency close to the sum of the frequency of two fundamental vibrational modes, or in the case of overtones, at multiples of the frequency of a fundamental mode.

Overtone and combination bands can also be present in the infrared spectra of molecules as a result of mechanical anharmonicities. Mechanical anharmonicity occurs when the potential energy surface of the system is not entirely parabolic. The result of this is that the restoring force, back to the equilibrium position, is no longer linear with displacement. This gives rise to an uneven spacing of energy levels, therefore changing their frequency compared to those given by the harmonic model.

In the same way that the Equation 2-4 was rewritten in terms of normal mode coordinates, we can rewrite Equation 2-1 to give an expression of potential energy as a function of normal mode coordinate:

$$V = V(0) + \sum_i \left(\frac{\partial \mathrm{V}}{\partial Q_i}\right)_0 Q_i + \frac{1}{2}\sum_{i\,j}\left(\frac{\partial^2 V}{\partial Q_i\,\partial Q_j}\right)_0 Q_i Q_j + \frac{1}{6}\sum_{i\,j\,k}\left(\frac{\partial^3 V}{\partial Q_i\,\partial Q_j\,\partial Q_k}\right)_0 Q_i Q_j Q_k \ldots \ldots$$

Equation 2-8

Mechanical anharmonicity presents itself in Equation 2-8 as non zero values in the cubic or higher terms. This again removes the independence of the normal modes, allowing mixing of their motion and the presence of overtone and combination bands

The presence of a combination band tell us that the fundamental modes which form it are coupled to one another in some way; that is the motion of one vibrational mode influences the motion of the other. Arising from mechanical anharmonicity, modes whose constituent atoms are covalently bonded together may be coupled as a result of sharing common atoms, whilst electrical anharmonicities may also couple modes, through space in the absence of a chemical bond. This idea of coupling does not commute with the simple normal mode picture in which the vibrational modes are entirely orthogonal to one another. Anharmonicity is required for molecular coupling to occur, however anharmonicity in a system does not mean that it must be coupled

Coupling of vibrational modes can occur either between vibrational modes of the same molecule (intramolecular) or between vibrational modes on different molecules (intermolecular). Intermolecular couplings can occur at varying strengths between different molecules in a variety of different setting, including non-covalently bound dimers, or between solute and solvent molecules. Intermolecular coupling are typically due to predominantly electrical coupling, due to the lack of covalent bond, however sufficiently strong electrical coupling can induce further mechanical couplings

2.1.2 Nonlinear polarisations

Central to spectroscopy – and the interaction of light and matter in general – is the idea that the electric component of an oscillating electromagnetic field, propagating

through a material, induces and oscillating polarisation in the charge density of said material. As with any oscillating charge, the oscillating polarisation then produces an electromagnetic field of its own. In majority of fundamental light matter interactions such as refraction and absorption, observed at low field strengths, the polarisation induced in a material is linear with respect to the electric field. This relationship is described by Equation 2-9 in which E is the electric field, P is the polarization and χ is the susceptibility[60]:

$$P = \chi E$$

Equation 2-9

The susceptibility term here is a descriptor of how readily the charges are redistributed in the material and is frequency dependent. It can be expressed in terms of electric permittivity as shown in Equation 2-10, where $\varepsilon(\omega)$ is the electric permittivity of the material at a given frequency ω, and ε_0 is the electric permittivity of free space.

$$\chi = \frac{\varepsilon(\omega)}{\varepsilon_0}$$

Equation 2-10

In this way, traditional spectroscopies such as infra-red absorption are described as linear spectroscopies.

In a situation in which the strength of the electric field incident on a material is greater than the electric field holding the electrons to the nuclei, the molecule can be ionized and electrons liberated from the molecule. Between the linear behavior observed at low field strengths and the behavior observed at field strengths sufficiently high to cause ionisation is an intermediate regime in which the induced polarisation is no longer linear with respect to the applied electric field. In order to access this intermediate regime, pulsed lasers are usually required over traditional sources, allowing for very strong electric fields to be applied for short durations of time, with high peak powers.

In this regime, the polarisation is best described as a Taylor series expansion of the input electric field, as shown in Equation 2-11[61].

$$P = \chi^{(1)}E + \chi^{(2)}EE + \chi^{(3)}EEE + \chi^{(4)}EEEE + \cdots$$

Equation 2-11

These higher terms in the expansion only become significant at higher field intensities.

Equation 2-11 shows for nonlinear polarisations, many susceptibilities are needed to describe the relationship between the input electric field and the induced polarisation. However, not all materials have non zero values of higher order susceptibilities and this can be seen when considering the even ordered contributions to the expansion shown in Equation 2-11. Even ordered terms in polynomials of this kind do not change sign when moving from positive to negative values of x, so in this way, for even ordered contributions to the nonlinear polarisation, the induced polarisation does not change sign upon reversing the input electric field. The implication of this is that it must be easier to rearrange the charge density in one direction than the other. Even ordered processes therefore only occur within materials with specific symmetry properties or at surface boundaries and explains the requirement for specific crystals for performing optical processes such as frequency doubling or sum frequency generation. Figure 2-2 shows a plot of first, second and third order contributions to nonlinear polarisation, as a function of electric field. It can be seen that the polarisation of the second order contribution does not change sign with reversal of E.

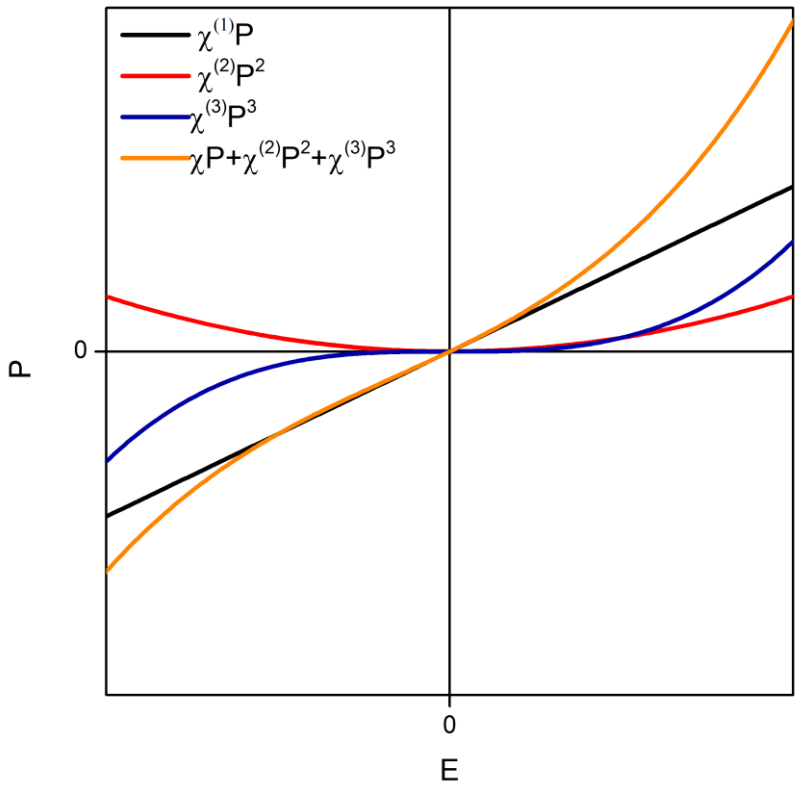

Figure 2-2 A plot of polarisation against electric field for linear ($\chi^{(1)}P$) and nonlinear regimes, adapted from reference[62].

For the case of a monochromatic electrometric wave, propagating through a material, if we consider the oscillating electric field component to be varying as $e^{i\omega t}$ where ω is the frequency of the wave, it can be seen that the nth order contribution to the nonlinear polarisation will oscillate at $n\omega$. For the example of a 2rd order process of frequency doubling, in an crystal of appropriate symmetry, the 2nd order contribution to polarisation will be proportional to $(e^{i\omega t})^2$, which is $e^{i2\omega t}$. The oscillating polarisation will therefore produce a new field at the 2nd harmonic of the input field. In the same way, in the case of the input field containing multiple frequencies, the sums and differences in the input frequencies will also appear in the induced polarisation and new fields will be launched at these frequencies.

It is the generation of new fields from inducing non-linear polarisations in this way that forms the basis of nonlinear spectroscopies. Nonlinear spectroscopies act to probe the frequency dependent susceptibility of a material and are named nth order

spectroscopies, where n is the order of susceptibility that is being probed and the order of the nonlinear polarisation induced. The processes that are driven in these techniques are referred to as (n+1) wave mixing techniques, again where n is the order of the nonlinear polarisation being induced. They are named in this way to account for the number of input frequencies, plus the new frequency created by the induced polarisation. Electron Vibration Vibration two dimensional infrared spectroscopy, the technique used for the work described in this thesis, is a four wave mixing process, probing the third order susceptibility of the materials of study.

2.1.3 Quantum mechanical coherences

The discussion of nonlinear optics up to this point has been entirely classical, with oscillating electric fields inducing oscillating polarisations in the charge density of a material. Although a detailed quantum mechanical description of nonlinear spectroscopy is not given here as it is not necessary for describing the work done, a brief overview of the idea of quantum mechanical coherences is provided as it provides a useful perspective for considering the processes involved and keeping track of the states accessed in EVV 2DIR.

The optically induced polarisations described above can be considered to be coherences, and it is these coherent states which give rise to the new radiated fields. A quantum mechanical coherence is a linear combination of two states, which can be created when an electric field perturbs a station state and mixes it with another[62]. The wavefunction of a quantum mechanical coherence, between two states, a and b can be given by Equation 2-12

$$\Psi(x,t) = c_a(t)\psi_a(x)e^{(i\omega_a t)} + c_b(t)\psi_b(x)e^{(i\omega_b t)}$$

Equation 2-12

Taking the squared modulus of this wavefunction gives us the probability density and shows us we have probabilities of finding the system in stationary populations of state a or b, given by $c_a(t)c_a^*(t)$ and $c_b(t)c_b^*(t)$, where the time dependent terms have canceled with their complex conjugates. We also see probabilities of finding the system in time dependent coherences of the two states, given by $c_a c_b^* e^{i(\omega_a-\omega_b)t}$ and $c_b(t)c_a^*(t)e^{i(\omega_b-\omega_a)t}$. These mixed states, or coherences can be seen to be oscillating between the two states at the difference in frequency between them. As the two states must have different dipole moments associated with them, the dipole moment of the system will be oscillating at this same rate, launching an electromagnetic field at this frequency.

2.2 Electron-Vibration-Vibration two dimensional infrared spectroscopy

In order to overcome the issue of spectral convolution in linear infrared spectroscopy, two dimensional infrared techniques have been developed. As described above, second order non-linear processes require a lack of inversion symmetry in order to occur. Therefore the lowest order nonlinear process that can be employed universally for a spectroscopy is third-order, driving four wave mixing processes. EVV 2DIR is a four wave mixing spectroscopy.

EVV 2DIR is a frequency domain experiment which uses two pulsed infra-red fields along with a pulsed visible field of sufficiently high intensity to induce nonlinear polarisation within a sample. Hereafter the two infrared pulses and visible pulse, along with their associated properties such as electric field, direction and frequency, shall be referred to by the Greek letters α, β and γ respectively, where β is the higher in frequency of the two infrared pulses. The two infrared fields, E_α and E_β, act to induce a resonant vibrational coherence within the sample, which is subsequently probed by the visible field Eγ, which induces a coherence with a non-resonant electronic state and drives the production of a four wave mixing signal. By producing the four wave mixing signal via an excited electronic state, a signal is produced in the visible

frequency range. All three of the fields take the form of narrow band, short pulses of duration 1.5ps and 25cm^{-1} in bandwidth (both values quoted as full-width half-maxima). The relatively long duration of the pulses, compared to those used in other ultrafast spectroscopic techniques, is required to achieve the narrow frequency profile of the pulses. This is necessary as EVV 2DIR is performed by scanning the frequency of the two infrared beams and detecting the intensity of a four wave mixing signal which is produced, exclusively when the two infrared beams are resonant with two coupled vibrational modes within the sample. The intensity of this FWM signal is then plotted as the z axis of a 3D graph, in which the two infrared frequencies form the x and y axis. This gives the advantage over other multidimensional infrared techniques of measuring a background free signal, which is zero when the infrared frequencies are off resonance, and does not require any dispersive elements or calibrated array detectors. Furthermore, detecting a visible signal (see Equation 2-13 below) as opposed to a signal in the infrared region allows for superior efficiency of detection. Spreading the spectral information over two dimensions and probing only coupled vibrational modes yields the benefit of considerable spectral decongestion, compared to linear infrared counterparts, as well of course, as information on the coupling between modes.

EVV 2DIR was originally prosed and implemented by Cho and Wright[63] and was initially named Doubly Vibrationally Enhanced two dimensional infrared spectroscopy. However, this name was changed to distinguish it from other techniques exploiting infrared double resonances.

2.2.1 Principles of EVV 2DIR

A coherence pathway describes the sequence of different coherences which are accessed across the course of a coherent nonlinear process, as each subsequent interaction between the matter and the fields occur. EVV 2DIR is a four wave mixing process, with three interactions of the input fields with the sample. There are 12 possible coherence pathways accessible with the three fields used for EVV 2DIR, each induce a third order ($\chi^{(3)}$) nonlinear polarisation. However there are only three

coherence pathways possible that induce a nonlinear polarisation oscillating at a frequency defined by the relationship of input frequencies, given in Equation 2-13

$$\omega_\delta = \omega_\beta - \omega_\alpha + \omega_\gamma$$

Equation 2-13

These are the three EVV 2DIR pathways and the evolution of their coherences, with each subsequent laser interaction is as follows:

i) EVV-IR 1 - gg → ga → ba → ea → aa

ii) EVV-IR 2 - gg → bg → ba → ea → aa

iii) EVV-Raman - gg → bg → ag → eg → gg

The three possible coherence pathways are depicted by the wave mixing energy level diagrams and corresponding double sided Feynman diagrams in Figure 2-3.

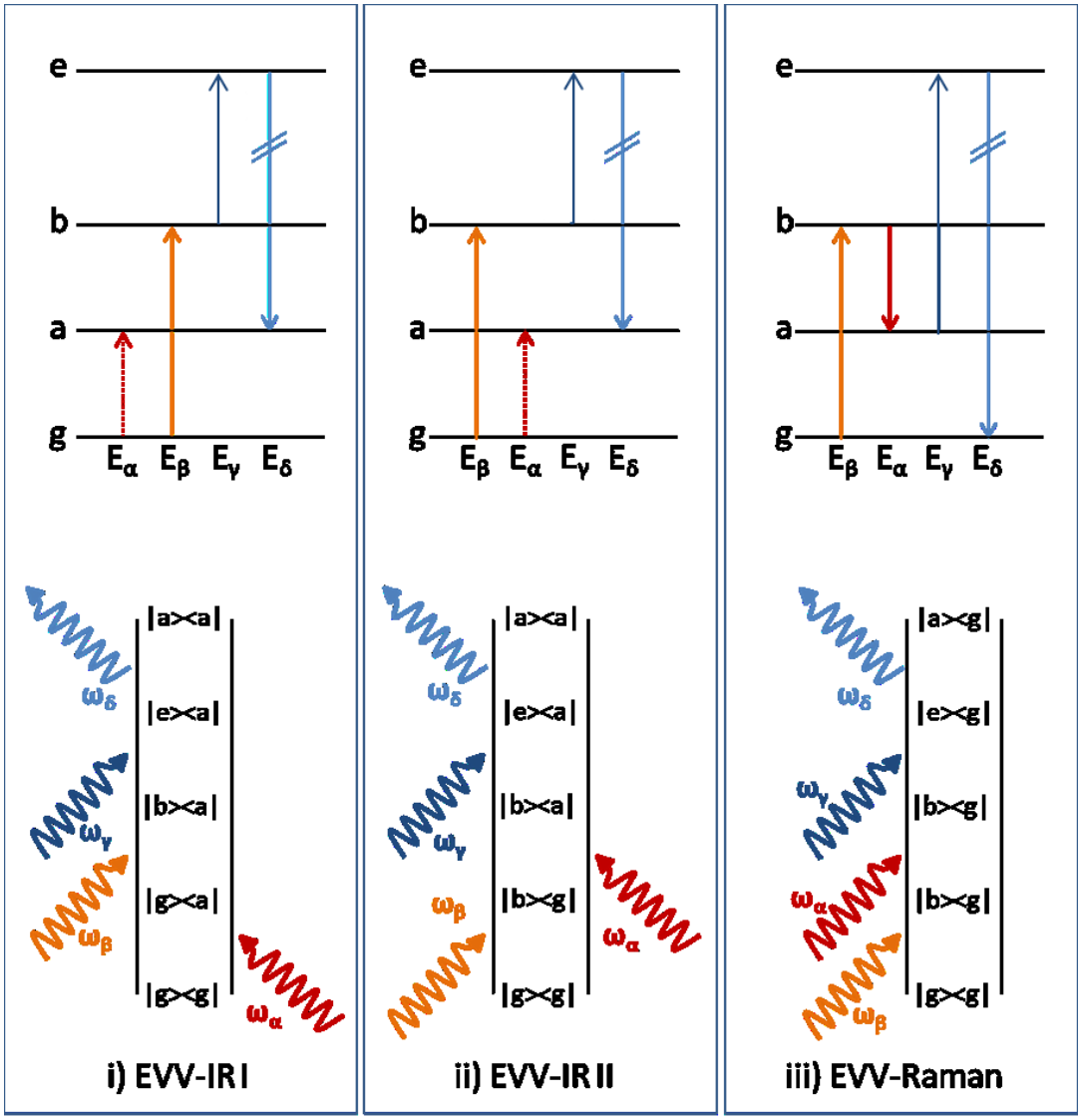

Figure 2-3 The wave mixing energy level diagrams (top) and corresponding double sided Feynman diagrams (bottom) describing the three possible EVV four wave mixing processes. The label g denotes the ground state, whilst a and b denote coupled vibrational states. The level labelled e is a non-resonant electronic state. In the wave mixing energy level diagrams, the order of the interactions in time is from left to right, whilst in the double sided Feynman diagrams it is from bottom to top. The dotted arrows in the WMEL diagram represent bra transitions whilst the solid arrows represent ket interactions, although the complex conjugates would also hold true for representing thesis processes.

The experiments performed for this thesis all used the EVV-IR1 pathway. As shown in Figure 2-3, the first light – matter interaction of the EVV-IR1 pathway, between the lower energy infrared pulse and the sample, creates a coherence between a vibrationally excited state, *a*, and the ground state, *g*. In the second interaction, the higher energy infrared pulse interacts with the other side of the wave function,

creating a new coherence between the vibrationally excited state, *a*, and another of higher energy, *b*, which will oscillate at the difference of the frequencies between the two, as described in Equation 2-12. The third interaction with the visible pulse creates the final coherence between the vibrationally excited state, *a*, and a virtual electronic excited state, *e*. This coherence oscillates at the frequency given by equation Equation 2-13, producing the EVV-2DIR signal, which is represented as the downwards arrow with two lines scored through it in the WLEM diagrams in Equation 2-2.

With all three of the input pulses coincident in time, all three of these pathways will take place (along with many other nonlinear non-EVV processes of different orders and involving different combinations of the input fields) but selection between them can be achieved by controlling the temporal ordering of the input pulses. It is not possible to exclusively select the EVV-IR 2 pathway or the EVV-Raman pathway in absence of one another, as both of them employ the pulse ordering β, α, γ. However this partial lack of pathway selectively was not an issue for the work described in this thesis as all spectra presented here were collected using the EVV-IR 1 pathway. The temporal pulse ordering used for accessing using the EVV-IR 1 pathway, along with the notation used for the two delays between the pulses is shown in Figure 2-4

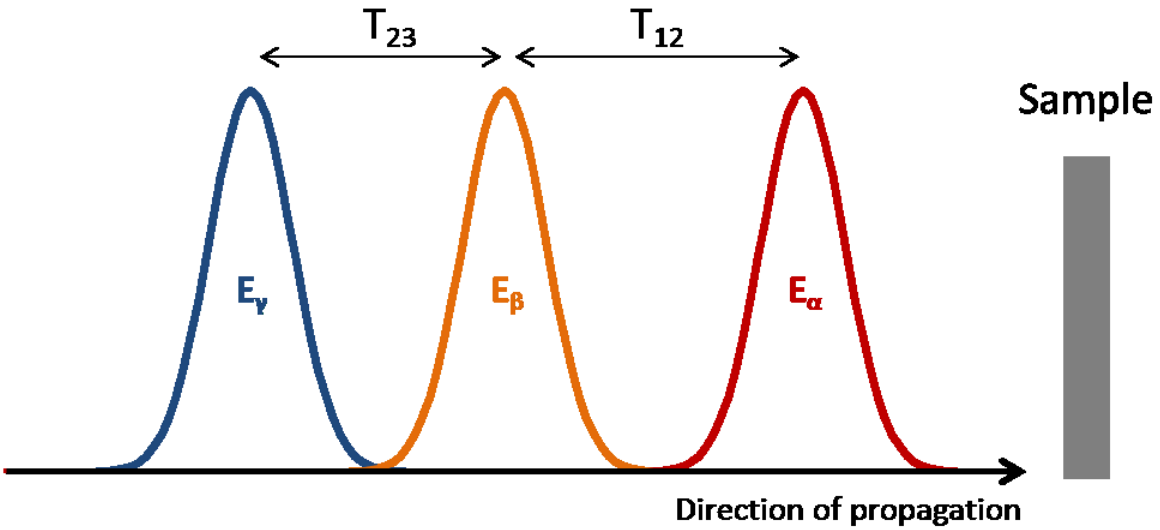

Figure 2-4 Diagram showing the temporal ordering of the input pulses used to select the EVV-IR 1 pathway. Pulse α arrives first, followed β after a delay, T_{12}, followed finally by γ after a second delay, T_{23}

Experimentally, it has been found that performing EVV 2DIR in this way yields cross peaks arising from one fundamental vibrational mode, *a*, coupled to a combination mode comprised of the same fundamental mode and one other, *a+b*. For these combination mode to be possible, mode *a* and *b* must be coupled. As a result of this it is possible to redraw the WMEL diagrams for the EVV processes to show coherences involving combination mode *a+b,* instead of the fundamental mode *b.* This produces the WMEL diagram seen in Figure 2-5.

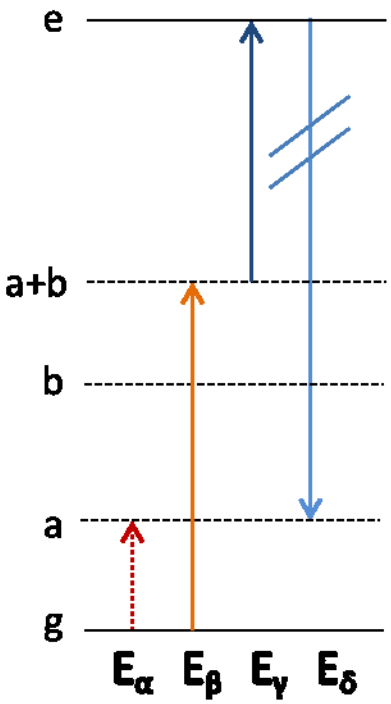

Figure 2-5 The WMEL diagram for the EVV-IR1 pathway, in which the resonant vibrational coherence being created is between a fundament vibrational mode, *a*, and a combination of *a* and another mode *b*

2.2.2 EVV 2DIR spectra

Because EVV 2DIR is measuring the couplings between pairs of two vibrational modes, via the coupling of the fundamental mode of one and the combination of the two, the cross peaks appear in the spectrum at $\omega_\alpha = \omega_a$ / $\omega_\beta = \omega_{a+b,}$ where ω_a and ω_{a+b} are the frequencies of the fundamental and combination modes, respectively. So contrary to other forms of 2D IR and 2D NMR, a peak at $\omega_\alpha = X$ / $\omega_\beta = Y$ does not simply correspond to a coupling of two modes at frequencies X and Y. As a result of this, cross peaks fall in vertical columns of common mode *a* and in diagonal lines of common mode *b.* A special case of peaks is also observed which arise from couplings between a fundamental mode and the overtone of this mode. These peaks will appear in the EVV 2DIR spectrum at frequencies of $\omega_\alpha = a$ cm^{-1} / $\omega_\beta = 2a$ cm^{-1}, although some anharmonic shift may be observed. These peaks therefore fall in a diagonal line of gradient 2:1. Figure 2-6 shows a schematic of an EVV 2DIR spectrum, with peaks arising from three coupled fundamental modes, labelled A, B and C. The columns of common mode *a* can be seen, along with the diagonal line of peaks of common mode *b.*

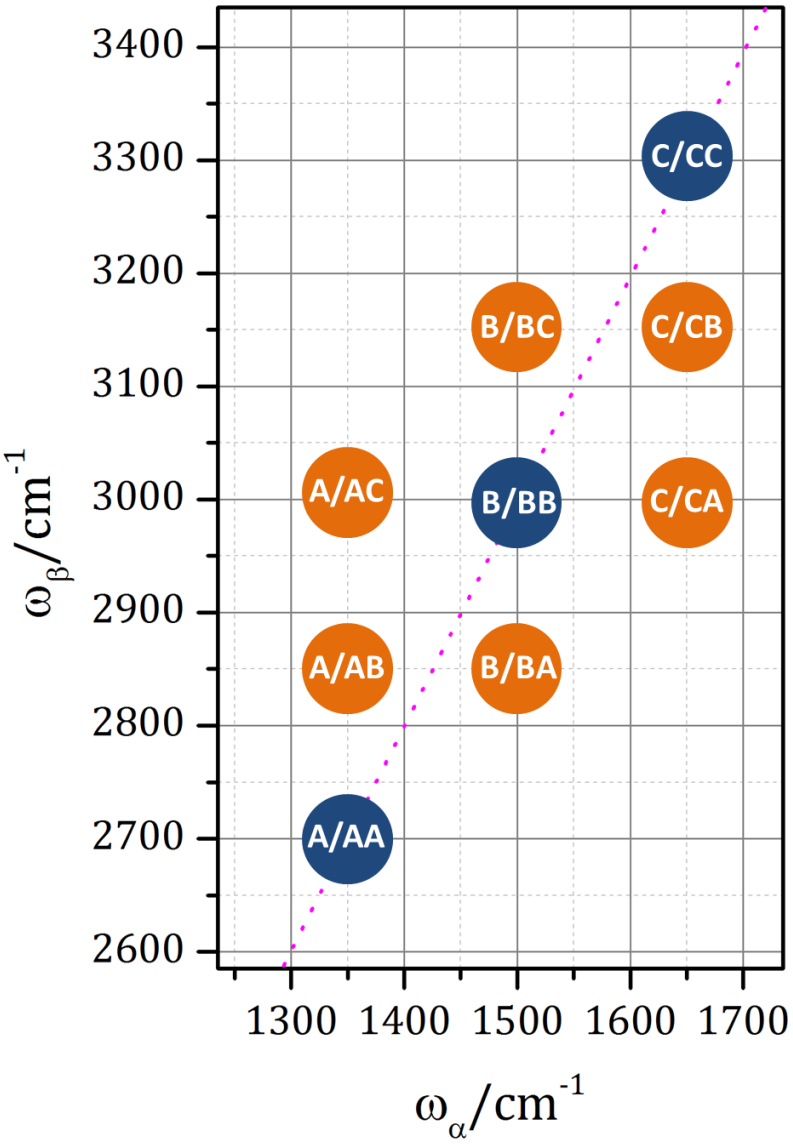

Figure 2-6 Schematic EVV 2DIR spectrum showing peaks arising from three coupled fundamental modes named A, B and C whose fundamental frequencies are 1350 cm^{-1}, 1500 cm^{-1} and 1650cm^{-1}. The peaks arising from overtone couplings are shown in blue, whilst the peak arising from couplings between the vibrational modes are shown in orange. A line of 2:1 gradient is shown in pink to illustrate the trend in locations of the overtone coupling peaks.

It is possible to "recast" EVV 2DIR spectra, and plot the signal intensities as a function of $\omega_\beta - \omega_\alpha$ against ω_α. After doing so the spectra resemble something more analogous to a 2D NMR spectrum, with diagonal self-couplings and off diagonal cross couplings between coupled modes. A schematic of the recast equivalent of the diagrammatic

spectrum shown in Figure 2-6 can be seen in Figure 2-7. After recasting in this way the frequency of the second vibrational mode in the combination mode is given by the Y axis, although this does not account for any anharmonic shift in frequency of the combination mode.

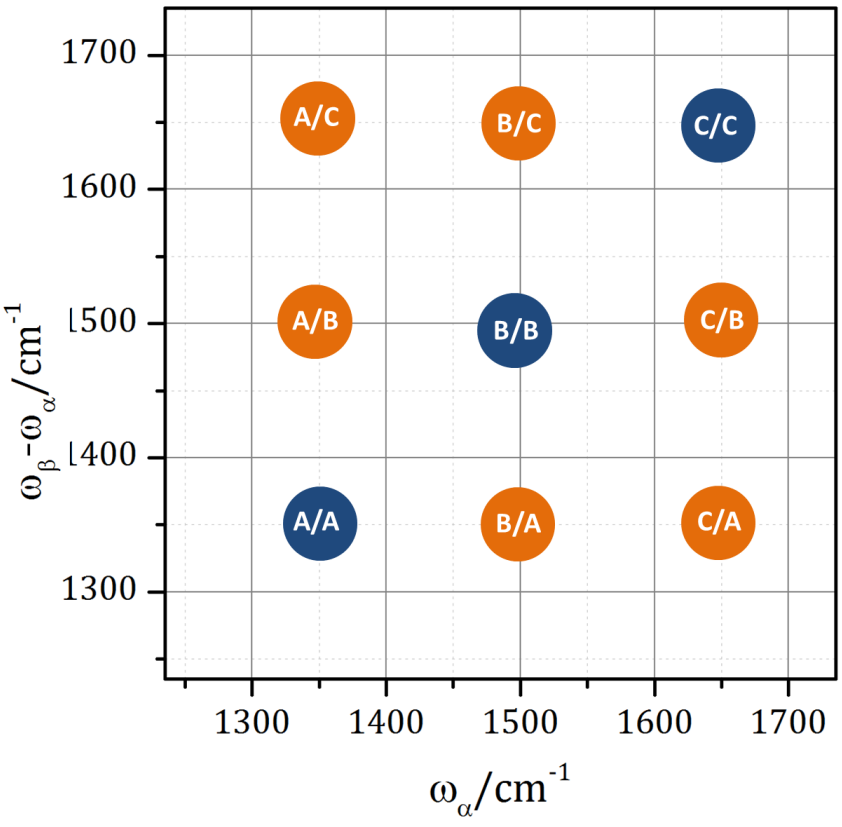

Figure 2-7 Schematic recast EVV 2DIR spectrum of the same model system described for Figure 2-6.

1.1.1. EVV 2DIR Phase matching

In order to maximise the intensity of the signals generated by the four wave mixing process, it is important to consider the relative geometry of the three input beams in order to meet the phase matching condition. At the individual photon level, the phase matching condition can thought of as a statement of conservation of momentum and energy of the input and output beams[64,65]. Through the EVV coherence pathway,

energy is clearly conserved, as shown in the energy level diagram, Figure 2-5, with the energy of the output beam related to the 3 input beams as in Equation 2-13

To meet the phase matching condition, and momentum to also be conserved, the vector sum of the wave vectors should be performed in the same way as with the energies above:

$$\vec{k}_\delta = \vec{k}_\gamma + \vec{k}_\beta - \vec{k}_\alpha$$

Equation 2-14

Where k_i is the wave vector of each experimental pulse and:

$$\overrightarrow{k_\iota} = \frac{\omega_i n_i}{c} \widehat{k_\iota}$$

Equation 2-15

In Equation 2.15, $\hat{k}$ is the unit vector giving the direction of propagation of the wave, c is the speed of light, ω_i is the frequency and n_i the refractive index at that frequency.

Unlike with the conservation of energy, it is possible for this condition to not be met. Equation 2-15 shows the wave vectors contain the direction of propagation, which are determined by the angles the beams are brought together at experimentally. The relative geometries between the input beams therefor need to be carefully considered to meet the condition described in Equation 2-14, else a phase mismatch will occur in which the wave vector of the produced signal is not the same as the nonlinear polarisation within the sample, which gave rise to it. This phase mismatch limits the length through the sample over which the four wave mixing signal can be produced before increasing interference occurs between the out of phase fields. The phase mismatch, Δk, is given by Equation 2-16, which shows it to be the difference between the wave vector of the actual signal and the wave vector of a signal that would conserve momentum:

$$\Delta\vec{k} = (\vec{k}_\gamma + \vec{k}_\beta - \vec{k}_\alpha) - \vec{k}_\delta$$

Equation 2-16

Figure 1-1 shows the vector sums associated with three different beam geometries, in which the vector length represents the beam energy or frequency: a) and b) show configurations where the phase matching condition has been met, and there is no phase mismatch, whilst c) shows a vector sum corresponding to an imperfect beam geometry in which a phase mismatch is present[66].

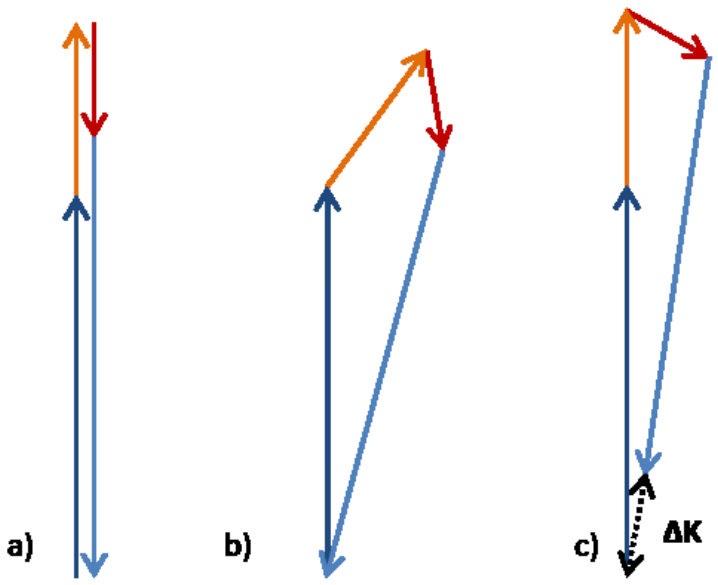

Figure 2-8 Vector diagrams showing a) collinear Δk=0 phase matching geometry, b) non-collinear Δk=0 phase matching geometry and c) non-collinear Δk≠0 phase matching geometry. The dark blue vector represents the 790nm beam wave vector, the orange and red vectors represent the 3000cm-1 and the 1500cm-1 centred beams wave vectors respectively and the light blue vector the signal beam wave vector. Vector lengths are not to scale

The EVV-2DIR experiment is typically performed by varying the infra-red frequencies across a range of around 800cm^{-1}. At constant directions of propagation the individual wave vectors will be varying linearly with frequency, meaning that a phase matching geometry designed to give $\Delta k=0$ at one set of input beam frequencies will give non-zero phase mismatches at other frequencies. Experimentally this is difficult to address so phase matching geometries are chosen such that $\Delta k=0$ at the centre of the experimental frequency range, to minimise the total phase mismatch across the entire experimental spectral range.

The above discussion of meeting the phase matching condition has only considered the energy and direction of propagation contributions to the wave vector, ω_i and $\hat{k}_i$. The third non constant term in Equation 2-15, n_i, the wavelength specific refractive index of the medium, also effects each of the wave vectors and therefor the degree of phase matching. The refractive index of a material is the ratio of the speed light

propagates though the medium to the speed light propagates in a vacuum so it follows that different speeds of propagation of the signal will lead to different phase mismatches. The refractive index of the sample materials can vary significantly across the spectral range, with the most rapid variations occurring around frequencies at which the sample absorbs. This, if unaddressed, would lead to significant changes in Δk around the regions of the spectrum which are of most interest. As the variation of n_i is complex, accounting for it on a frequency by frequency basis is not possible so the approach of tightly focusing the beams is used. This acts to bring in a wide range of input beam wave vectors, as the tight focus spreads the beam over a range direction of propagations but has the disadvantage of decreasing the volume of input beam overlap.

The beam geometry shown in Figure 2-8a is geometrically the simplest, where all three input beams are collinear along with the resultant signal. Practically however it is difficult or impossible to achieve as it would require the use of two dichroic mirrors which are able to distinguish the three input beams over the full frequency range of the experiment. This is perfectly feasible for transmitting the 790nm beams and reflecting the infrared beams (or *vice versa*), but optics to select between the two infrared beams over the range of the experiment, which are much closer in energy than they are to the 790nm, are not currently available.

The work described in this thesis was performed using a phase matching geometry such as that depicted in Figure 2.7b. There are a wide range of possible angles that can satisfy the phase matching conditions in such a non-co-linear way, varying from the collinear regime right up to very broad geometries. A vector diagram of a non-collinear, Δk=0, phase matching geometry with labelled angles can be seen in Figure 2-9 For any reasonable value of φ, a corresponding value of θ can be calculated through trigonometry.

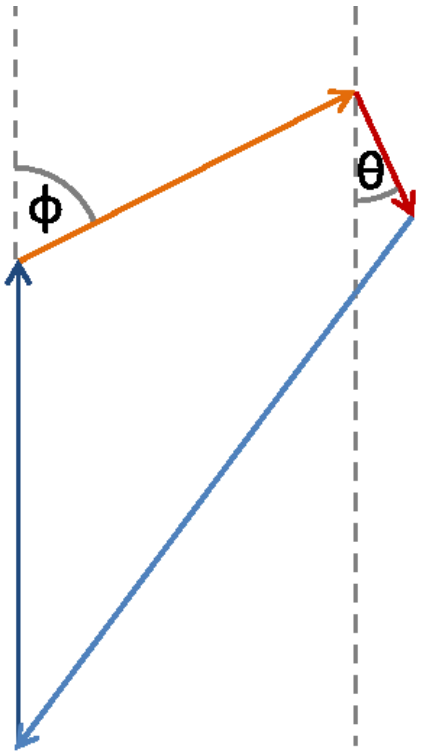

Figure 2-9: A vector diagram showing a non-collinear Δk=0 phase matching geometry. The dark blue vector represents the 790nm beam wave vector, the orange and red vectors represent the 3000cm^{-1} and the 1500cm^{-1} centered beams wave vectors respectively and the light blue vector the signal beam wave vector. φ and θ are the angles between the 3000cm^{-1} and the 1500cm^{-1} beams and the 790nm beam respectively. Vector lengths are not to scale

The disadvantage of using increasingly broad geometries, i.e. larger values of φ and θ, is that the volume of intersection of the three input beams, in which four wave mixing can occur, will become ever smaller, leading to less intense four wave mixing signals. Therefore the smallest corresponding values of φ and θ that are achievable experimentally were used and were calculated to be 2° and 2.42° respectively.

2.3 Applications of EVV 2DIR

2.3.1 Development and initial applications

EVV 2DIR was originally prosed and implemented by Cho and Wright[63] in 1999. The initial work performed by the Wright group was studies of small molecules as proof

of principle experiments. By recording the EVV spectra of acetonitrile, the spectral deconvolution offered by only producing signals from coupled vibrational modes was demonstrated; peaks arising from the fundamental C-H stretch, which typically dominate linear spectra were not observed in the spectra as they were not coupled to the C≡N mode[67]. By using mixtures of acetonitrile and deuterated acetonitrile as a simple model[68], it was further shown that cross peaks are not exhibited between non interacting molecules. The Wright group were also the first to report calculated EVV 2DIR spectra to accompany experimental spectra[69], again investigating the vibrational couplings in acetonitrile. The Klug group are the only group other than that of Wright to adopt the EVV 2DIR technique.

2.3.2 Application to proteins

EVV 2DIR spectroscopy has previously been applied in the Klug group to study proteins in several different ways. EVVs ability to decongest vibrational spectra and access a large frequency space make it a good tool for studying such large complex systems. The spectral contributions of several protein moieties including methylene groups, amide backbone and numerous amino acid side chains such as phenylalanine, tyrosine, and tryptophan have been characterised for EVV and are now readily identifiable in protein spectra. This allowed EVV 2DIR to be used to identify proteins, based on their unique "optical fingerprint" of a EVV cross peak intensity ratios[70,71]. EVV 2DIR has also been applied to the study of post translational modifications of proteins, both phosphorylation and nitration[72], with the degrees of modification able to be quantified.

EVVs ability to quantify and distinguish between proteins and their modified forms, has presented itself as a potentially powerful form of contrast for biological imaging[73]. By imaging biological samples numerous times with input pulse frequencies corresponding to different key cross peaks the spatial distribution of numerous different chemical species could be elucidated. EVV 2DIR has been demonstrated as an imaging tool by the Klug group on mouse kidney section and was able to differentiate between several different tissue types.[73]

2.3.3 Application to complex formation

The unique ability of EVV 2DIR to detect not only mechanical vibrational couplings but also purely electrical couplings enables it to detect intermolecular coupling even in the absence of a chemical bond. This can be pictured by considering a model of two proximal chemical groups, which are not covalently bonded to one another but are sufficiently close in space to interact. Each of the chemical groups has vibrational modes localised to their own atoms and any interaction between the two can only take place via through-space electrostatic interactions. This coupling can lead to EVV 2DIR cross peaks. This has previously been exploited to detect the formation of an intermolecular complex between benzonitrile and phenylacetylene using EVV 2DIR spectroscopy[74]. This was done by detecting presence of new cross peaks in the EVV 2DIR spectrum, corresponding to the coupling of vibrational modes of the benzonitrile with those of the phenylacetylene. These purely electrical couplings which are due to interactions of the dipole moments of the two groups will be dependent on the relative geometries of the interacting dipole moments. It was shown that the ratio of intensities of these cross peaks when measured in two different polarisation schemes (PPP with all input fields parallel to one another, and PPS in which the polarisation of the visible probe field has been rotated through 90°) can be used to calculate the angle between the dipole moment, if their interaction can be approximated to be dipolar.

2.3.4 Triple Resonance enhancement

EVV 2DIR can typically be considered to be a doubly resonant technique: The two infrared pulses induce double vibrational resonance coherence in the sample, whilst the interaction with the third pulse creates coherence with a non-resonant electronic state. This third interaction can be considered to be the coherent equivalent of half of a Raman scattering process. As a result of this, the electronic properties of the sample are also able to have an influence on the EVV signal. In the majority of cases, the molecules of the sample have no electronic transitions close in frequency to the 790nm visible pulse, therefore their electronic structure has no detectable impact on

the EVV signals and need not be considered. In the case of coloured compounds however, the visible pulse may be resonant, or close to resonant with an electronic transition in the molecule, and greatly enhanced signals can be observed when probing vibrational modes to which it is coupled. Under these conditions EVV 2DIR can be considered to be triply resonant.

This effect has been termed "triple resonance enhancement" and has previously been observed in the Klug group by Gardner and Fournier[73], in the retinal chromophore of bacteriorhodopsin. It was shown that a peak arising from a CH_3 moiety in retinals conjugated backbone appeared with 10^4 times higher signal intensity than a peak arising from a CH_3 moiety with no resonance enhancement. The UV/Vis spectrum of retinal shows its absorption maximum at 560nm, showing it is not necessary to achieve absolute resonance in order to observe triple resonance enhancement. Although near resonant transitions may yield resonance enhancement, the magnitude of the transition dipole of the electronic transition needs to be sufficient as well as its frequency being correct in order to observe a substantial enhancement.

Triple resonance enhancement in can be exploited to observe spectral features from compounds or moieties that without it, would be undetectable above a spectral background: For example detecting specifically the contributions from the retinal CH_3 moiety over the many hundreds of CH_3 moieties in the bacteriorhodopsin protein. This is a powerful tool for biological applications where molecules are typically very large, and compound of interest such as drugs or metabolites are typically comparatively small and rare.

The instrumentation used for the current implementation of EVV 2DIR (as described in section 3.1) uses a fixed frequency 790nm probe pulse. Because of this, only systems exhibiting electronic transitions sufficiently close to 790nm can yield resonantly enhanced signals. The use of a visible frequency optical parametric amplifier or other similar technology would enable the frequency of the visible probe pulse to be tuned, allowing for molecules or moieties with specific electronic properties to have their spectral features selectively enhanced.

The idea of resonance enhancement forms a key part of the experiments described in Chapter 4, in which the spectral features of a coloured drug molecule are enhanced

sufficiently so that they can be seen over the spectral background from a protein several hundred times larger.

3 Experimental Methods

3.1 EVV 2DIR Instrumentation

3.1.1 Overview of the laser system

This section provides an outline of how the three pulses required to perform an EVV 2DIR experiment are generated, controlled, synchronised, and how the signal is detected. It also discusses some of the theory behind the generation of the input pulse and signal, where it is of experimental significance.

In brief, a commercial Spectra Physics laser system is used to produce 1.5 ps 790 nm pulses of 25 nm bandwidth, 2 mJ pulse energy at 1 kHz. This output is split into three arms; two of the 790 nm arms are used to generate tunable infrared pulses using two optical parametric amplifiers (Spectra Physics OPA800C). The third arm remains unchanged at 790 nm and is brought together at the sample with the two infrared arms. Relative temporal control between the three pulses is achieved by the use of delay stages which act to shorten or lengthen the path lengths of the three arms such as to control the relative timings between the three pulses. The four wave mixing

signal is detected using a photomultiplier (Hamamatsu H7422P-50), after transmission through a filter stack designed to filter out the input beams and transmit only the four wave mixing signal, in the visible range of 675-715nm. A schematic diagram of the experimental instrumentation can be seen in Figure 3-1

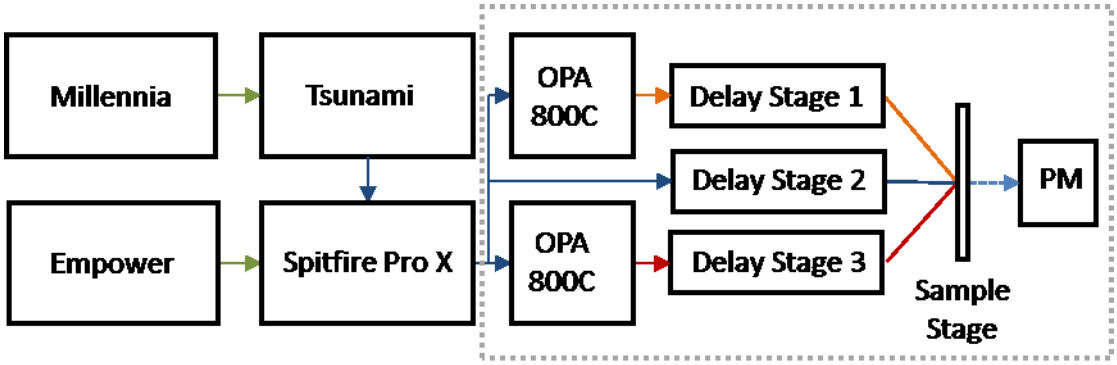

Figure 3-1 Schematic of the laser systems which comprise the EVV 2DIR spectrometer. The dotted line shows the parts of the experiment which are under nitrogen purge

The parts of the experiment which use infrared beams are enclosed and purged with nitrogen. This is shown in Figure 3-1 as the area enclosed by the dotted grey line. This prevents absorption of the infrared light by water molecules in the air, which would have the effect of attenuating the beam intensity and also distorting the temporal profile of the pulses.

3.1.2 Generation of picosecond 790 nm pulses.

The EVV 2DIR experiment, as implemented here, is driven by a commercial Spectra Physics laser system producing 1.5 ps, 790 nm, 2 mJ per pulse, 1 kHz output. In order to produce this output a Spectra Physics Tsunami oscillator is used, pumped by a Spectra Physics Millennia, to produce 100 fs 790 nm pulses at 100 MHz repetition rate. These 100 fs pulses are then temporally stretched and amplified using a Spectra Physics Spitfire Pro X regenerative amplifier, pumped by a Spectra Physics Empower

Pro. The regenerative amplifier temporally stretches the seed pulses before amplifying them to lower the peak power within the gain medium to below the damage threshold. After ejection from the gain medium, the amplified pulses are temporally compressed again. To further prevent damage to the gain medium (titanium sapphire crystal) in the amplifier, crystal is Peltier cooled to -10 °C. To prevent the formation of ice crystals on the cooled titanium sapphire crystal, a dry nitrogen purge is used to dehumidify the crystal housing.

3.1.3 Optical Parametric Amplifiers

In order to produce the tunable infrared picosecond pulses required to perform EVV 2DIR, the process of optical parametric generation is employed in two optical parametric amplifiers (Spectra Physics OPA800C). The principles of operation of the OPAs are outlined here as they give rise to effects which must be accounted for when performing EVV 2DIR experiments.

The process of generating picosecond tunable infrared output from picosecond 790 nm pulses, within the OPAs, can be thought of in 5 stages:

1. Generation of white light continuum
2. Amplification of white light continuum (optical parametric generation)
3. Frequency selection from the white light continuum
4. Amplification of selected frequency (optical parametric generation)
5. Conversion of amplified selected frequency into infrared by difference frequency mixing

1. Generation of white light continuum:

In order to convert the non-tunable 790 nm amplifier output to a tunable infrared output, white light is generated first. Doing so produces a wide range of visible wavelengths which can be selectively amplified to give the source of tunability. White light is generated in this case through supercontinuum generation, a complex process produced by several high order nonlinear optical processes such as self-phase modulation, self-focusing, Raman-shifting, and multi-photon processes[75]. Super continuum generation can be achieved by focusing a few μJ of ps laser pulse into a material such as quartz sapphire or specific liquids, to produce a broadband coherent output. In the case of the OPA800C, a proprietary solid material is used, which looks to be quartz. The 790 nm input to the OPA is split into three arms, one of which drives this super continuum generation process, whilst the other two are used as pump beams, as described below.

2. Amplification of white light continuum:

Next the white light continuum is amplified. This is done differently than the amplification achieved in a laser, in which a population inversion is required. Laser amplification would not be suitable for this multiwavelength application, as the energy levels of the gain medium need to be appropriately matched to the wavelength which is to be amplified. In the case of the OPA, optical parametric generation is used.

Optical parametric generation is a down converting nonlinear process in which a pump field (in this case the 790 nm regenerative amplifier output) is converted into two fields whose frequencies sum to produce the pump frequency, within an optically nonlinear material. These two new fields are referred to as the signal and the idler and their frequency relationship to the pump is given in Equation 3-1 and can be seen in Figure 3-2a.

$$\omega_{pump} = \omega_{signal} + \omega_{idler}$$

Equation 3-1

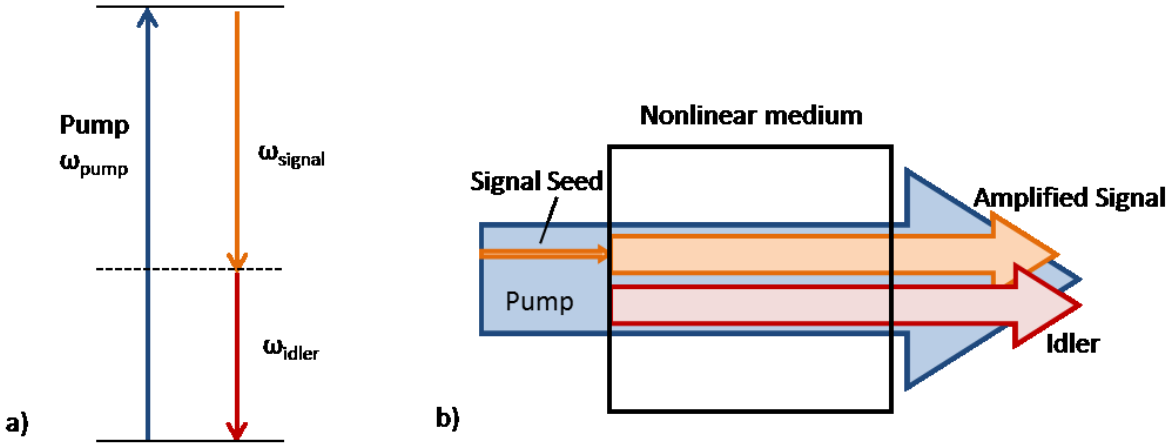

Figure 3-2 (a) the energy level diagram showing how frequency relationship between the pump and the resultant signal and idler for optical parametric amplification. (b) the geometry of the three beams involved in the seeded optical parametric amplification process.

Clearly there is a full range of possible signal and idler frequencies that would satisfy Equation 3-1, however these frequencies can be controlled through the introduction of a seed field which forces the signal to be produced at the frequency of the seed, along with an idler of corresponding frequency. This process is considered amplification as energy is being transferred from the pump beam frequency to the frequency of the initially much weaker seed.

In the case of the OPAs used here the white light continuum (which acts as the seed) is directed through a BBO crystal (barium borate), collinearly with one of the 790 nm pump beams. They are temporally overlapped by the use of a delay stage in the white light beam path and produce an amplified white light continuum.[76]

3. Frequency selection from the white light continuum:

The amplified white light is then directed onto a retroreflecting diffraction grating which acts to disperse the white light by its component frequencies and direct it back towards the BBO crystal. The diffraction grating is mounted in an actuator controlled mount, which allows the angle of the grating, relative to the incident amplified white light to be adjusted. In this way, selected frequencies out of the amplified continuum can be directed back to the BBO crystal, propagating through it in the opposite direction as the initial white light. This stage is the source of tunability of the final infrared outputs.

4. Amplification of selected frequency:

As the selected frequency from the amplified continuum passes through the BBO crystal for a second time it is overlapped, collinearly with the second pump beam. A variable delay stage is used to ensure temporal overlap whilst an adjustable telescope in the pump is used to optimise to pump spot size to that of the selected part of the continuum in the BBO. Here the selected frequency is amplified, again by optical parametric generation, with the selected frequency returning from the grating acting to seed the process. The angle of the BBO crystal is set such as to meet the phase matching condition and allow for maximum amplification. This produces two beams, the signal and idler, of bandwidth ~25 cm^{-1}. The achievable frequencies for these outputs, along with typical pulse energies at these frequencies can be seen in Figure 3-3.

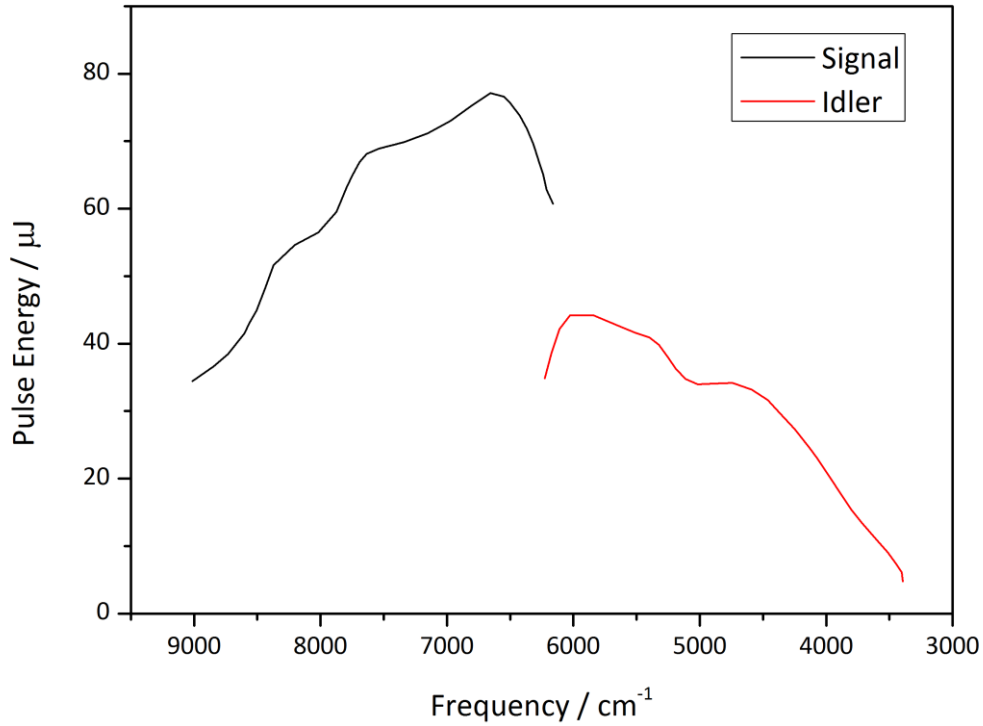

Figure 3-3 Typical pulse energies achievable for the signal and idler output of the OPA800C

5. Difference frequency mixing

Although there are 2 quantum and a few 1 quantum vibrational transitions in the available frequency range of the idler, the frequency ranges of interest for the work conducted here lie further into the infrared. To access these frequencies, in the 1250 – 3450 cm^{-1} range, a process called difference frequency mixing is employed. This process can be considered to be analogous to the optical parametric generation, in which the OPG signal forms the pump frequency and the OPG idler forms act as the new signal. This produces a new frequency at the difference of the OPG signal and idler, as described by Equation 3-2 and depicted in Figure 3-4a.

$$\omega_{DFM} = \omega_{signal} - \omega_{idler}$$

Equation 3-2

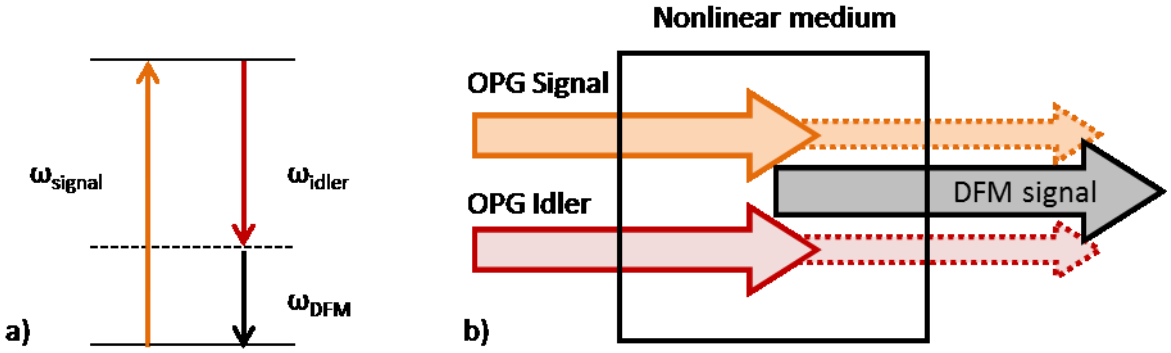

Figure 3-4 (a) the energy level diagram showing the frequency relationship between the signal from the optical parametric amplification process and the resultant idler and DFM signal, for the difference frequency mixing process. (b) the geometry of the three beams involved in the difference frequency mixing process.

This is done experimentally by focusing the signal and idler from the optical parametric generation into an $AgGaS_2$ crystal which acts as the nonlinear medium. The crystal is again mounted on an adjustable mount to allow its angle to be varied relative to the beams, allowing phase matching to be achieved. Given that the signal and idler are generated simultaneously, no delay stage is required to synchronise them. The DFM crystal (Difference Frequency Mixing) in the OPA800Cs rotates in plane with the table and must rotate across a range of about 70° for the OPA centred at 3000 cm^{-1}. This large angular variation with wavelength results in significant variable displacement in the DFM output signal, which must be accounted for experimentally. This is discussed in section 3.1.5. DFM signals can be generated in this way between 1000 cm^{-1} and 3450 cm^{-1} and the outputs are used for the EVV 2DIR experiments described here. The unwanted frequencies such as the OPG signal and idler are filtered out with an appropriate long pass filter.

3.1.4 Delivering the beams to the sample

Each of the three beam lines is guided using planar silver mirrors. Each beam line contains a telescope which is used to ensure the beams are collimated and size matched at the sample. An off axis parabolic mirror is used to bring all three of the beams together at a common focus at the sample. The condition for this common focus is that the three beams are parallel to one another as they strike the parabolic, which is shown in Figure 3-5.

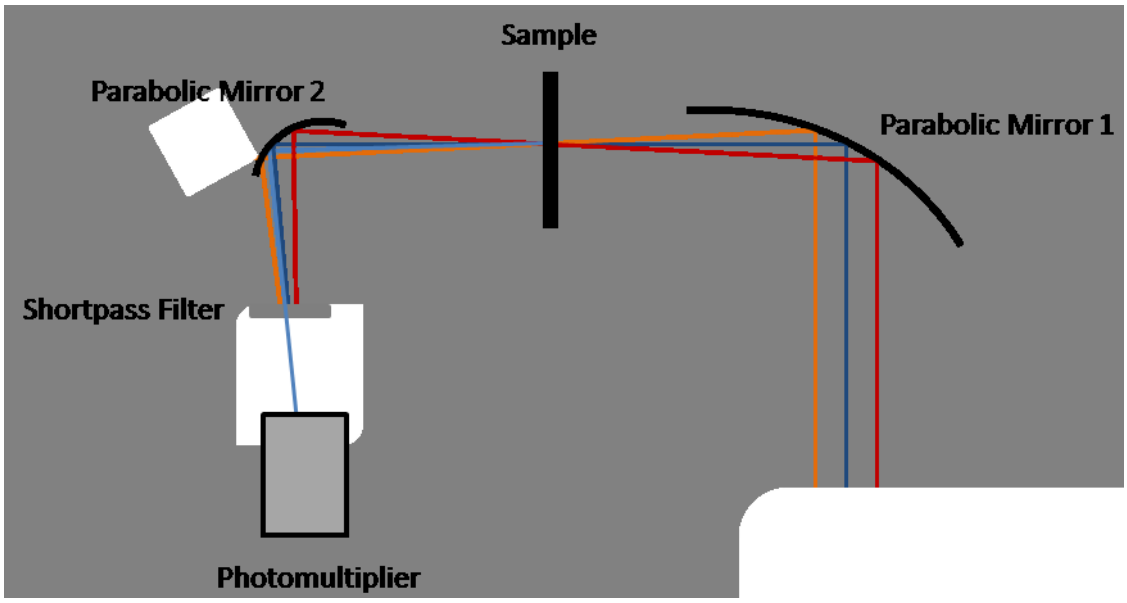

Figure 3-5 Schematic of the parabolic mirror bringing the three parallel beams to a common focus at the sample position. The generation of the signal is also shown. The signal is guided with a second parabolic mirror to the detector and the three input beams being removed before detection by a shortpass filter.

As the three input beams are focused at the sample position, the resultant signal, along with the residual input beams, will be divergent as they propagate away from the sample. In order to ensure signal light is not lost as a result of this, a second parabolic mirror is used to focus the signal onto the detector face.

3.1.5 Achieving temporal control

As described in section 3.1.3, the wide angular range over which the DFM crystal rotates to allow for phase matching results in considerable beam displacement of the OPA outputs. This affects the angles between the beams, as well as the temporal and spatial overlap of the beams at the sample position.

As the beam pointing changes with frequency, the pulses produced at one frequency will take a slightly different path to the sample to those at other frequencies, leading to different path lengths and therefore different pulse timings. If unaddressed, this would result in the relative timings between pulses to change across the spectral range of the experiment. Differences of up to 0.8 ps per spectral dimension have been observed, which is sufficient to give rise to considerable spectral artefacts, including large non-resonant background signals as a result of pulses being overlapped in time.

In order to compensate for this, a T0 calibration is performed. The T0 positions are the positions of each of the three delay stages which give temporal coincidence of the three beams at the central infrared frequencies – that is to make each of the three path lengths the same. These positions are found experimentally by scanning each of the three delay stages in turn and finding the positions which give maximum non-resonant four wave mixing signal in a sample of calcium fluoride. As this signal is non-resonant it only occurs when the three pulses are temporally coincident and therefore this optimisation of T0 positions rely on starting with some degree of temporal overlap of the three beams. In this case, temporal overlap of the 790 nm beam and an infrared beam can be achieved by scanning one in time relative to the other and looking for an up conversion signal at the sum of the two frequencies, generated in a BBO crystal.

Performing the T0 calibration involves measuring the T0 positions at a range of frequencies, across the experimental range for each of the OPA outputs, relative to the T0 positions at the central frequencies. By fitting a 3rd order polynomial to the resultant ΔT0/ frequency data, the T0 correction required to keep the pulses coincident at any frequency can be calculated. These corrections are applied

experimentally as spectra are collected. An example of a T0 calibration curve for the OPA centered at 3000 cm^{-1} can be seen in Figure 3-6.

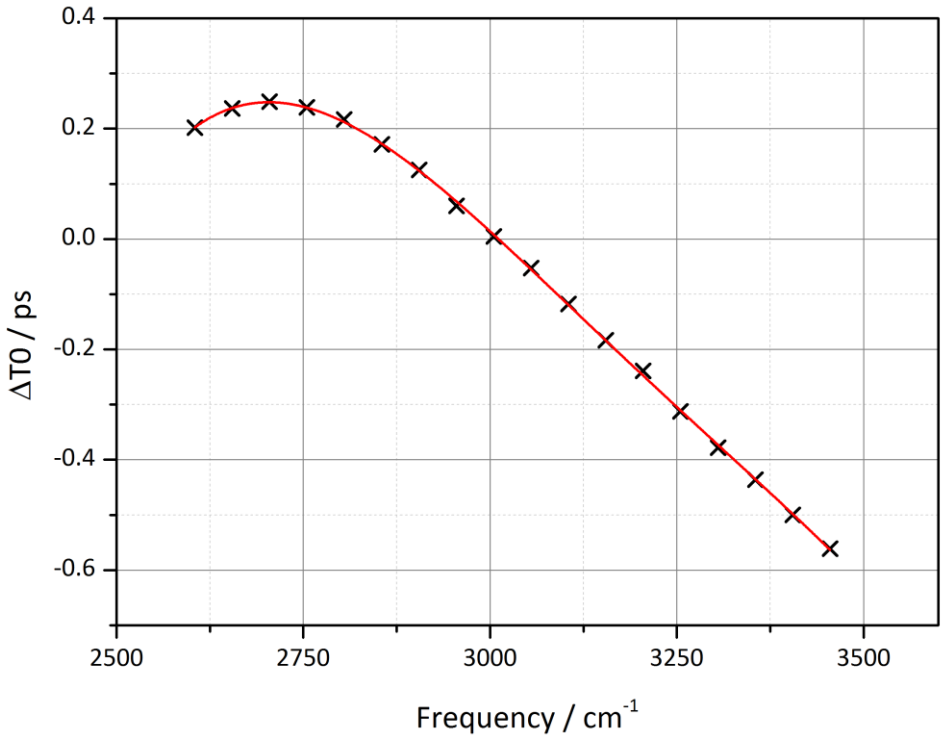

Figure 3-6: An example T0 calibration curve for the OPA centered at 3000cm^{-1} and its third order polynomial fit.

As well as affecting the angles between the beams as they come together at the sample position, the variation in beam pointing out of the OPAs, as a function of wavelength, also affects the temporal and spatial overlap of the beams.

3.1.6 Signal detection

The signal is guided to the Hamamatsu H7422P-50 photomultiplier detector face, with a parabolic mirror to compensate for the initial divergent nature of the signal, as shown in Figure 3-5. A set of three shortpass filters with their edge at 770 nm were used to remove the residual input beams. The voltage generated by the photomultiplier is passed to a gated box car averager, which integrates the signal voltage across a 2 ns window centered on the signal. This integrated value is passed on to the LabView software.

3.2 Processing EVV 2DIR data

In order to allow for quantitative analysis of EVV 2DIR spectra, the raw spectral data is first processed in a number of ways. This processing was implemented using a series of MATLAB scripts which were written for this purpose. The stages of this process are detailed below, with the order in which the stages are performed being crucial to processing the data correctly. The EVV 2DIR data in its raw form is collected as image-like matrices of signal intensities, with no frequency information. Once the data has been fully processed it is plotted as a spectrum and given frequency labels.

3.2.1 Smoothing

Firstly the raw data is smoothed using a nearest neighbor mean filter. This script acts to replace each element in the matrix with the mean of a 3x3 square of elements centered on the element in question. This is illustrated for two elements in Figure 3-7, where element 2,2 is replaced with the mean of the blue coloured cells and element 4,5 is replaced with the mean of the red coloured cells.

1,1	1,2	1,3	1,4	1,5	1,6
2,1	2,2	2,3	2,4	2,5	2,6
3,1	3,2	3,3	3,4	3,4	3,6
4,1	4,2	4,3	4,4	4,5	4,6
5,1	5,2	5,3	5,4	5,5	5,6

Figure 3-7 Depiction of the nearest neighbour smoothing filter.

This smoothing is useful as it acts to remove spectral noise introduced by laser instabilities. The OPAs produce output intensities with coefficients of variation between 0.5 and 1.5 % over 100 laser shots. They are also susceptible to intensity oscillations on the order of seconds in duration of magnitude of about 1 %. Smoothing the data as described here acts to considerably reduce the impact of these optical noises on the spectra.

The EVV 2DIR spectra discussed in this thesis were all collected with a frequency resolution of 5 cm^{-1}. Given that the bandwidth of both the infrared OPA outputs and the 790 nm visible beam are 25 cm^{-1}, measured at full width half maximum (five frequency domain pixels per dimension), smoothing in a nearest neighbour fashion

as described here using a 15 cm^{-1} square will not result in the loss of peak resolution. The spectral widths of the features probed in the experiments described here are also about 25 cm^{-1}, so again smoothing over this short range does not result in "over smoothing".

3.2.2 Infrared independent background subtraction

In order to be able to correctly consider peak intensities, peaks must be free of underlying background signals. The background signals to be removed can be categorised as either infrared dependent or infrared independent and are treated separately. Infrared independent background signals are subtracted immediately after smoothing, whilst any infrared dependent background, are removed after OPA energy normalisation (described below).

Infrared independent backgrounds are signals which are recorded during an experiment which are completely independent of the infrared frequencies which are being probed. They are present even when the infrared beams are blocked and arise from both optical and non-optical sources.

Non-optical sources are the signals recorded arising from the photomultiplier detector dark count voltage along with any constant electronic offset from zero which may be introduced by the signal collection electronics (gated integration unit and data acquisition board). This background can be recorded by measuring the reading given by the system with all 3 of the laser beams shuttered and may be negative.

There are two known sources of optical non-resonant background which are accounted for; both originate from the 790 nm visible beam. The photomultiplier has a stack of filters covering its face in order to only transmit the frequency range of the EVV signal. This filter stack however appears to be imperfect as a small amount of signal, proportional to the 790 nm intensity can be detected. This is true in both the absence and presence of a sample and is attributed to the leakage of 790 nm light through the filter stack.

Whilst studying molecules which exhibit peaks in their visible absorption spectra, those with peaks sufficiently close in frequency to the 790 nm visible beam (such as SU-5402), have shown signals with intensity proportional to both the concentration of the coloured compound and the 790 nm beam. A plot of infrared independent signal against SU-5402 concentration, measured with infrared beams shuttered, can be seen in Figure 3-8. The data displayed in Figure 3-8 is the mean of three measurements used in the SU-5402 concentration series work in Chapter 4. If the linear trend in Figure 3-8 were to be fitted with a straight line, the y-intercept would represent a mean non-optical infrared independent background. Given the proportionality of the intensity of these signals with compound concentration, these signals can be attributed to some kind of fluorescence process. Given that the 790 nm beam is already redder than the low frequency edge of the transmissive window of the filter stack, the signals must arise from multiphoton fluorescence processes.

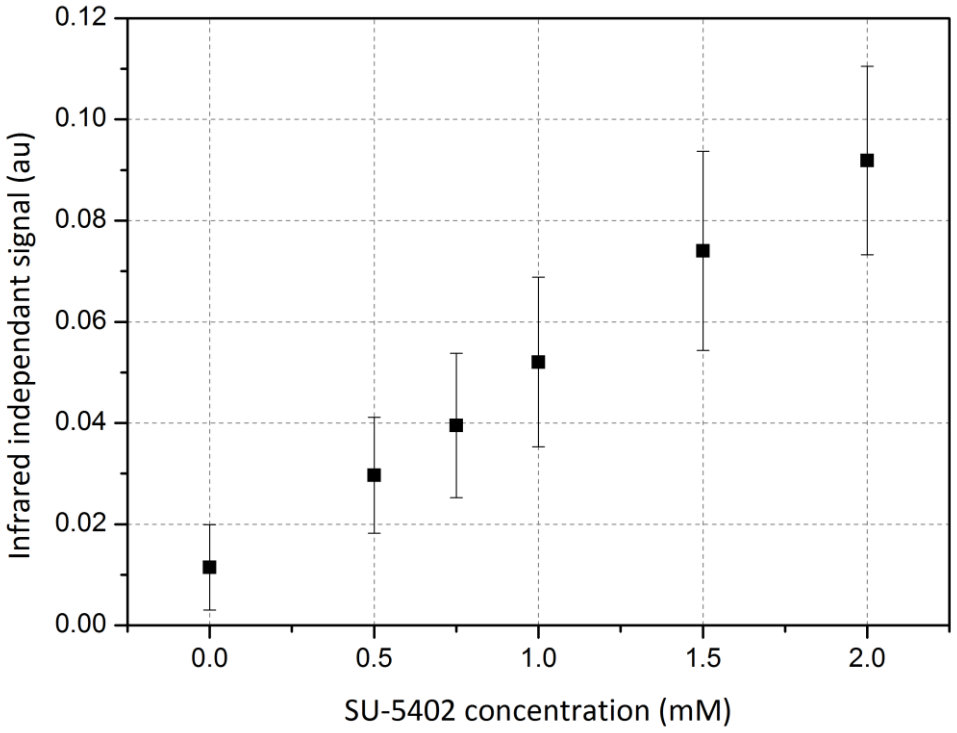

Figure 3-8 Plot of infrared independent signal against SU-5402 concentration

As these infrared independent signals are flat across the entire spectral range, the MATLAB script which performs this stage of the processing simply takes the signal measured for each sample, with the two infrared beams shuttered, and subtracts it from every element in the matrix.

3.2.3 Correction for variable infrared energy profiles

The use of two optical parametric amplifiers allows for the generation of two independently tunable infrared beams, with central frequencies of 1500 cm^{-1} and 3000 cm^{-1}. The pulse energies produced from the OPAs at these central frequencies are typically around 3 μJ and 8 μJ respectively, however the pulse energies produced are not flat as a function of wavelength. Example plots of pulse energy as a function of wavenumber for the two OPA outputs can be seen in Figure 3-9

As described previously in section 3.1.3, the beam pointing of the OPA outputs change slightly in the plane parallel to the table, as a function of output wavelength, owing to the changing angle of the DFM crystal. This would result in beam pointing changes at the sample position, resulting in variable degrees of overlap of the three experimental beams, if it were not for the implementation of the spatial filters. The use of spatial filters, reduces the complexity of the variable beam pointing and pulse energies problem to just a question of pulse energy at the sample, not varying pulse spatial overlap. By measuring the pulse energy as a function of wavenumber of the two OPAs, at the sample position, the energy available to drive the four wave mixing process in each of the beams can be recorded. These profiles can change slightly as a result of beam alignments so are recorded after each new beam alignment. Example plots of pulse energy as a function of wavenumber for the two OPAs, at the sample position can be seen in Figure 3-9.

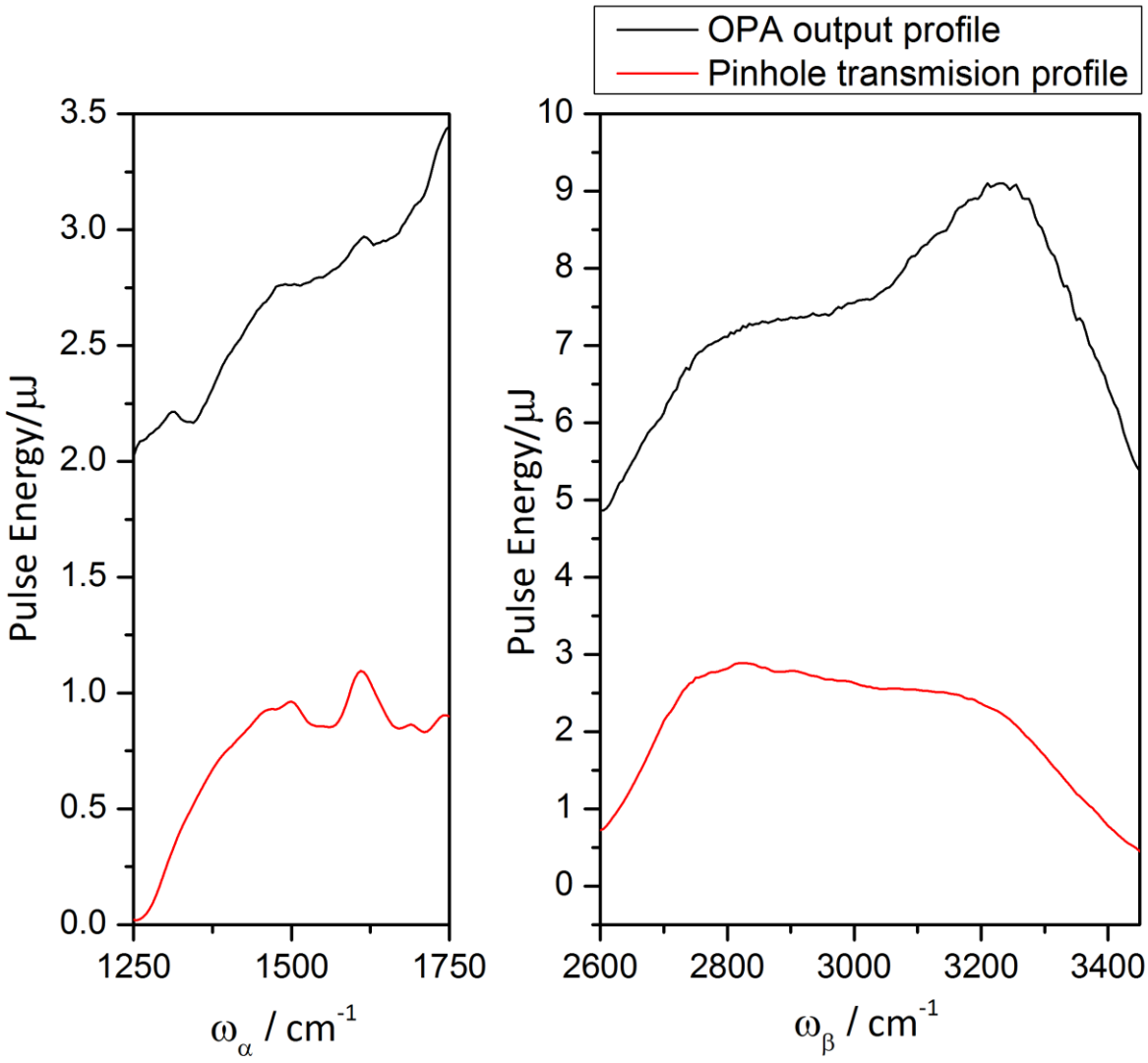

Figure 3-9 Typical pulse energies out of the OPA and at the sample position for both OPAs, as a function of frequency

As the EVV 2DIR signal intensity is proportional to the pulse energies of each of the three beams, experimental data will be inconsistent between experiments if the pulse energy profiles as a function of frequency are inconsistent. Attempting to reproduce identical pulse energy profiles to previous experiments by adjusting beam alignments would be incredibly impractical and difficult, so the experimental energy profiles are recorded at the sample position and accounted for in post processing.

A MATLAB script was written which would take the recorded pulse energy profiles and multiply them together to give a matrix of the same dimensions as the EVV 2DIR data, containing the product of the two infrared energies at each frequency coordinate. EVV 2DIR data which has already been processed by the previously

outlined procedures would then be divided by the OPA pulse energy profile matrix, corresponding to the conditions under which the data was corrected. This yields data which was now independent of variations in pulse energy as a function of frequency.

1.1.2. Infrared dependent background subtraction

After the data has been normalised with respect to input infrared energy, across the spectral range, any non-resonant background signal which is proportional to the infrared pulse energy is removed. This is done by a MATLAB script which takes the average value of a region of the spectrum with no peaks in it, then subtracts this value from every element in the data.

3.2.4 OPA frequency correction polystyrene

Before each new set of experiments the OPA output energies are measured as a function of frequency, then again after transmission through a NIST standard polystyrene film. From these two measurements a linear polystyrene absorption spectrum can be calculated for each OPA range. By comparing the frequencies of the peaks in this spectrum with those in the NIST standard polystyrene film spectrum, any frequency offsets can be calculated and accounted for in the EVV 2DIR spectra.

3.3 Preparation of hydrated gel spots

Water has strong absorption bands in the mid and near IR regions, some of which overlap with protein and DNA features[77], thus conducting EVV 2DIR experiments of protein or DNA samples containing bulk water would not be possible. EVV 2DIR spectra taken of dilute solutions would suffer from greatly reduced energy incident on the sample molecules at these frequencies, along with weak sample signal intensities due to the reduced concentration of sample molecules in the beam overlap volume. In order to circumvent this issue, whilst keeping some water in the samples to maintain biological relevancy, droplets of protein or DNA solution can be dried down to form gel-phases. This has previously been done by the Klug Group for protein samples at ambient conditions to yield of protein gel-phase "coffee-ring" structures.[72]

Coffee-ring refers to the structures formed when drops of solution or colloidal solution dry to leave ring like deposits of material on solid surfaces, as seen in spilt coffee[78]. This effect occurs due to the contact-line between the edge of the drop and the solid surface being pinned, preventing the radius of the drop moving inwards as the liquid evaporates (the mechanism of pinning is not fully understood but has been shown to be connected to surface inhomogeneity).[79,80] As the contact line is pinned and unable to move closer to the drop centre and the solution volume is decreasing, there is a solvent flux from the centre of the drop to the perimeter. This effect, in conjunction with increased rates of evaporation at the perimeter and the inclination to maintain a curved cap profile due to surface tension, results in deposition of solute in an annulus around the perimeter of the droplet.

Figure 3-10 shows an image of a coffee ring spot formed by allowing a 1 µl droplet of FGFR1 (1 mM, phosphate buffer 5 mM, NaCl 5 mM) to dry on a glass coverslip under ambient conditions, collected using an inverted bright field microscope. Considerably reduced buffer concentrations are used for this preparation, compared to those

typically used in biochemistry as higher salt concentrations would result in salt crystals forming in the protein spots, during drying.

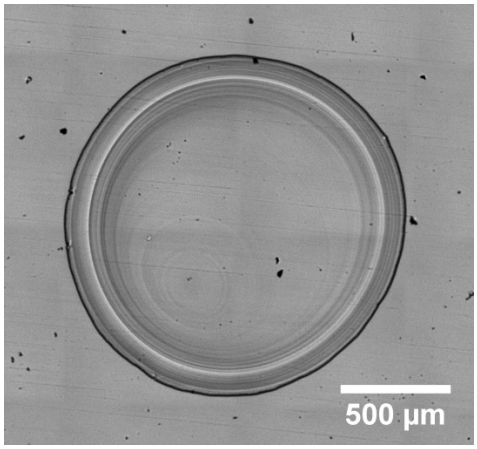

Figure 3-10 Bright field microscope image of a gel phase coffee ring structure, formed by drying protein solution at ambient humidity.

Although this methodology yields protein gel phases, it does suffer from a couple of limitations from an EVV 2DIR perspective: the rapidly changing curvature of the surface of the coffee rings is undesirable as any small movement of the beams across the sample surface, as a function of changing frequency or pulse timings, results in moving to a thinner or thicker region of the sample. This changes the number of molecules in the beam overlap volume and would lead to varying signal intensities. In addition, drying the samples at ambient conditions may lead to differing final humidities and sample thicknesses as a result of day to day variations in ambient humidity and temperature. Furthermore, introducing these gel phases to the extreme low humidity of the nitrogen purged laser system can eventually lead to cracking and crystallisation of the solute.

In order to negate these issues, it was found that by allowing 1 µl spots of protein solution to slowly dry at controlled high humidity, rounded, flat topped gel spots could be formed – presumably by considerably slowing the rate of evaporation, the changing gel spots are given time to equilibrate and redistribute, before becoming too thick and concentrated to do so. This solves both the problems of coffee ring

formation and of variable ambient conditions as controlled high humidities could be used in a temperature controlled (± 0.1 °C) laser laboratory.

In order to maintain controlled humidity conditions when conducting experiments, whilst protecting from the nitrogen purge, a simple humidity controlled sample cell was developed using commercially available parts and saturated salt solutions.

Saturated salt solutions, of sufficient volume, will maintain a fixed humidity within a closed volume by establishing an equilibrium between the solution and the water vapour within the volume. With a reservoir of undissolved salt, saturated salt solutions are able to adsorb or liberate large quantities of water from the local atmosphere, without changing the equilibrium humidity. Excess salt is required so that increased water content in the solution does not affect the concentration of the solution, in turn, impacting the equilibrium with the local atmosphere. By choosing salts of different hygroscopicity, different relative humidities can be achieved[81].

By mounting a 1" Ø circular coverslip in a 1" Ø lens mount followed by a ¼" thick 1" Ø rubber O-ring and a 2 mm thick 1" Ø CaF_2 window, a sealed volume could be created between the coverslip and the CaF_2 window. By depositing 1 µl protein spots along with spots of saturated salt solution onto the coverslips before sealing, the protein spots would dry slowly in the fixed humidity environment, forming flat topped hydrated gel spots. A schematic of this humidity controlled sample cell can be seen in Figure 3-11.

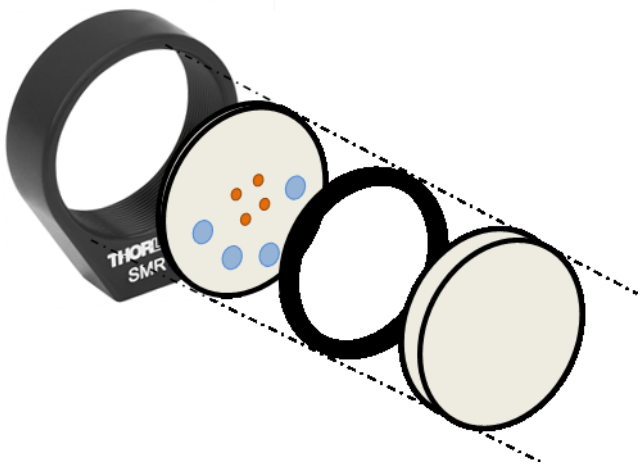

Figure 3-11 Schematic of the humidity controlled sample cells used for EVV 2DIR experiments. The first disk after the lens mount represents the thin glass coverslip with sample spots shown in orange and saturated salt solution spots shown in blue. The dark ring represents the rubber O-ring spacer, followed by the 2 mm CaF_2 window.

The calcium fluoride window forms the front of the sample cell, with the lasers traveling firstly through this side of the cell. CaF_2 was chosen for the front window as, unlike glass, it is transparent in the infrared region which is clearly essential for this application. Glass coverslips were chosen for use as the back window, with samples deposited onto it as it does not need to be transparent to infrared, only to visible light, due to the wavelength of the four wave mixing signal being 670 – 740 nm.

Five different saturated salt solutions were used to test the effect of humidity on the drying of 1 µl droplets of FGFR1 (1 mM, phosphate buffer 5 mM, NaCl 5 mM). Table 3.1 lists the salts and humidities used along with the observed results.

Salt	Relative humidity (% RH)	Outcome of FGFR1 gel spot
Potassium Nitrate	94.6	Wet Gel
Potassium Chloride	85.2	Wet Gel
Sodium Chloride	75.5	Ideal Gel
Magnesium Nitrate	54.4	Cracked Gel
Potassium Carbonate	43.2	Coffee Ring

Table 3.1The form of the FGFR1 gel spots produced at different relative humidities, along with the salt whose saturated solution was used to achieve that humidity.

Of the humidities investigated in Table 3.1, those greater than 75.5 %RH appeared upon first visual inspection to produce good quality, flat topped gel spots however when EVV 2DIR spectra were recorded it was found that the signal levels rapidly dropped off as a function of measurement time, disappearing to baseline over the course of a minute. By allowing the protein spots to remain at too high humidity they remained too well hydrated and soft. This resulted in the sample being forced away from the focus of the laser beams by either radiation pressure or a kind of local heating effect. By studying these overly hydrated spots using an inverted bright field microscope, markings could be seen across the surface of the spot where the laser had been incident upon them. An image of an FGFR1 spot at 85.2 %RH can be seen in Figure 3-12, with the perpendicular lines across the spot corresponding to the single axis at a time movement allowed by the sample stage.

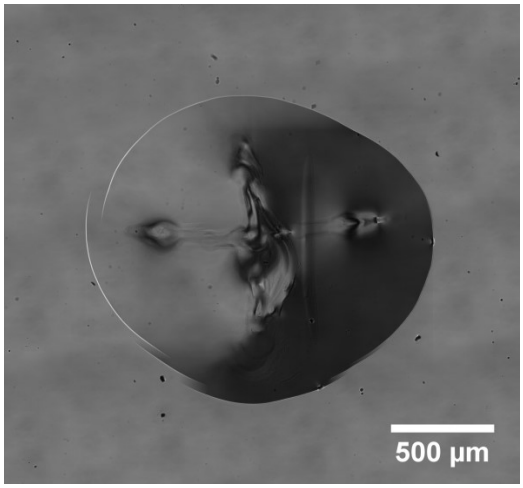

Figure 3-12 Bright field microscope image of a soft FGFR1 gel spot, which has been scored by the laser beams.

The use of saturated sodium chloride solution to produce an environment of 75.5 %RH yielded smooth, uniform, rounded gel spots which produced steady signals in the EVV 2DIR spectrometer. These spots did not mark under the laser, as can be seen in a microscope image of a 75.5 %RH spot in Figure 3-13

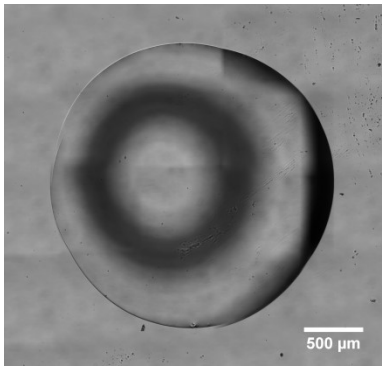

Figure 3-13 Bright field microscope image of an FGFR1 gel spot, formed at 75.5 %RH, which remains uncracked and unscored after time in the laser

A third type of non-coffee-ring gel spots were also observed by drying 1 µl droplets of FGFR1 (1 mM, phosphate buffer 5 mM, NaCl 5 mM) at 54.4 %RH. In this case the protein spots did not dry sufficiently quickly as to form coffee-rings, however continued to dry out beyond a well hydrated gel phase, resulting in a cracked topography. This cracking presumably comes as a result of the gel becoming too thick

to accommodate the decreasing spot volume without splitting. A microscope image of a cracked FGFR1 spot, dried at 54.4 %RH can be seen in Figure 3-14.

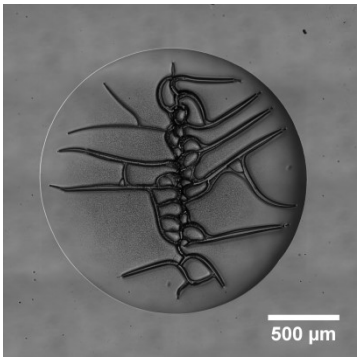

Figure 3-14 Bright field microscope image of a cracked FGFR1 gel spot, which was dried at 54.4 %RH.

The DNA studies described Chapter 5 were performed using DNA samples maintained in a gel phase. DNA samples in the gel phase had never previously been produced within the group so it was initially unclear how readily they could be formed. The same methodology was used to determine the optimal humidity conditions for producing DNA gel spots which was found to be 85.2 %RH, which was produced using saturated potassium chloride solution.

3.4 Sample Validation

The following section presents the work done to validate that the biomolecular samples being used for the EVV 2DIR studies were chemically sound and able to form the structures that were to be invested spectroscopically.

3.4.1 Validation of FGFR1 – SU-5402 Binding

Before looking for spectral signatures of FGFR1–SU-5402 complex formation, it was first necessary to confirm that the current samples of protein and inhibitor were in fact binding and forming said complex. In addition to identifying that binding was occurring, it was also necessary to determine the practical stoichiometry of the

binding process. The FGFR1 kinase domain construct used for this work contains a single binding site so the ideal stoichiometry for binding should be 1:1, SU-5402 to FGFR1. This 1:1 ratio, however, will not be observed if either the protein or inhibitor sample has undergone any kind of degradation and loss of activity. In order to determine this actual stoichiometry of binding, Isothermal Titration Calorimetry (ITC) was used.

Isothermal Titration Calorimetry is a technique which can be used to determine the binding constant, stoichiometry, enthalpy and entropy of complex formation. This is achieved by measuring the heat produced by the solution as one of the binding partners is titrated into the other[82,83].

Isothermal titration calorimeters consist of two cells, contained within an adiabatic jacket, such that no heat can be exchanged with the surroundings, or between the two cells. These two cells are the sample cell, in which the binding events occur between the molecules of interest, and a reference cell, which contains the equivalent solvent to that which the binding reaction is occurring in. One of the binding molecules is stored in the sample chamber and the other in an actuator controlled syringe capable of delivering precise amounts of material into the sample cell. The syringe needle rotates during operation and doubles as a stirrer due to the paddles featured at its tip. As the substrate in the syringe is titrated into the sample cell, heat is evolved and a heater is used to ensure that the solution in the reference cell remains at the same temperature as that in the sample cell. A schematic diagram of an ITC apparatus can be seen in Figure 3-15 By monitoring the work done necessary to keep the two cells at the same temperature, the heat evolved by every sequential addition of material can be determined. This allows for the production of a binding isotherm.

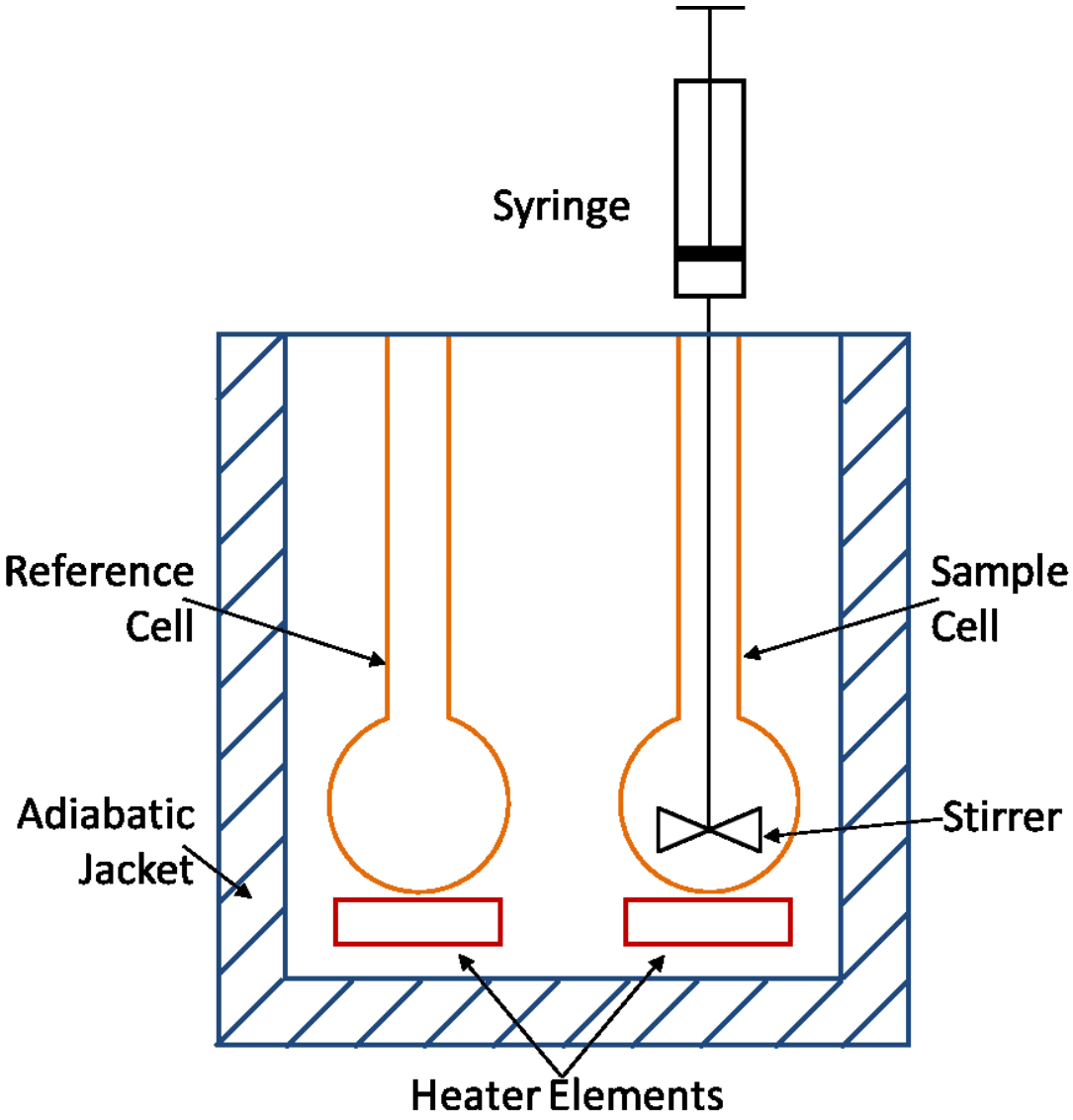

Figure 3-15 Schematic diagram of an isothermal titration calorimetry apparatus

To determine the percentage of drug and protein molecules that were functional and able to bind, a solution of 0.1 mM SU-5402 was titrated into a solution of 5 μM FGFR1. Both inhibitor and protein were in a buffer solution of 40mM HEPES, pH7.5, 200mM NaCl, 1mM TCEP containing 2% by volume DMSO. Serial additions of the SU-5402 solution were made up to a final molar ratio of SU-5402/ FGFR1 of 2:1. The measurement was repeated until three reasonable data sets were achieved. Many

attempts were plighted by air bubbled in the syringe or sample chamber. The plot of the mean result of these three experiments can be seen in Figure 3-16.

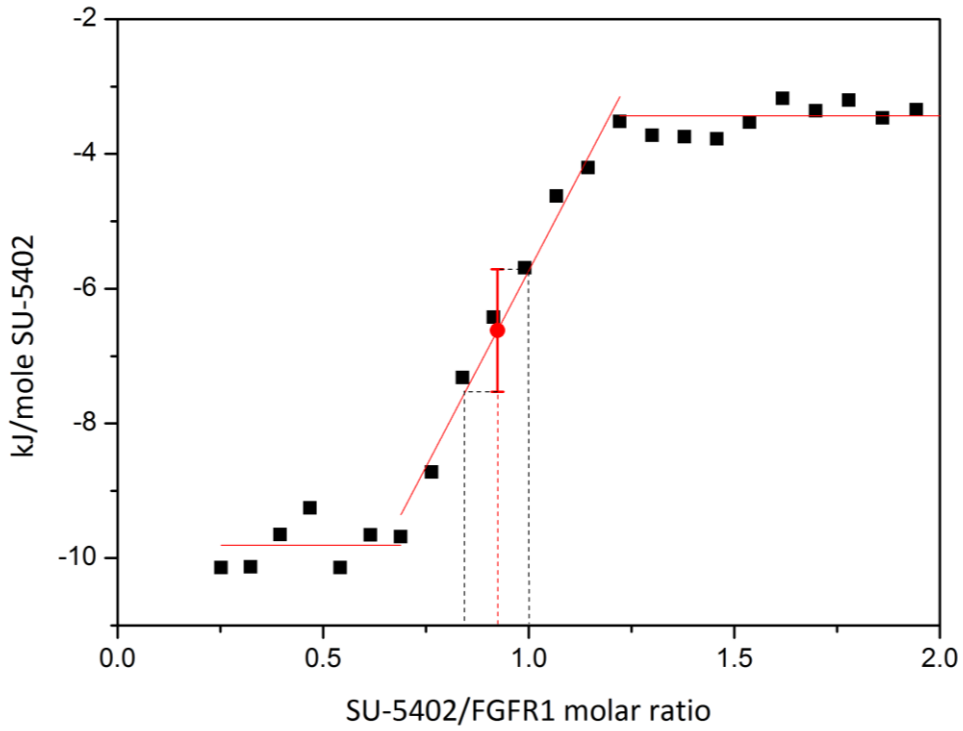

Figure 3-16 The ITC binding curve of additions of SU-5402 to FGFR1

The aim of these ITC experiment was purely to determine the effective stoichiometry of binding, and in turn understand to what degree the protein and inhibitor samples were active and able to bind to one another. For ITC measurements the reaction stoichiometry is determined from the titration equivalence point[84]. This is the point halfway up the titration curves, between the two flat regions representing the initial additions and point of binding site saturation. By fitting these two flat regions with lines of zero gradient (shown in Figure 3-16), the midpoint between them could be found. By fitting the middle section of the curve with a straight line(shown in Figure 3-16), the molar ratio of SU-5402 to FGFR1 at the point of equivalence was found to be 0.923. The average standard deviation of the three measurements at each of the points along this fitted region was 0.715 kJ/Mol. The combined standard deviation of the points form the fits of the flat regions of the binding curve was 0.208 kJ/Mol.

These variations combine to give a standard deviation in the molar ratio of the point of equivalence to be 0.079

It can be concluded from this analysis that 92.3±7.9% of the total protein and inhibitor content was able to bind in solution. This is perfectly tolerable for the EVV studies presented here and is in agreement, within error, to Astra Zeneca's determination of protein activity of close to 100%

3.4.2 Validation of G-quadruplex formation from Myc234 Sequence

Before looking for spectral signatures of the intermolecular interactions due to G-quadruplex formation, it was first necessary to confirm that the sequence and conditions being used did in fact result in the formation of G-quadruplexes. This was done using circular dichorism as quadruplex structures have been characterized using this approach previously in the literature. The circular dichorism measurements presented here were performed by Shababa Selim whilst she was a Masters student in the Klug group.

CD measurements were conducted with 250 μL Myc2345 solutions of concentration 20 μM with 200 μM KCl using an Applied Photophysics Chirascan instrument. The measurements were conducted in a 0.1 cm path length quartz cell using 1 nm/ s scan rate. The CD spectrum of the Myc2345 was taken five times averaged then smoothed using the 10-point adjacent average smoothing filter in Origin Pro 9.0. The mean CD spectrum of 20 μM Myc2345 + 200 μM KCl can be seen in Figure 3-17

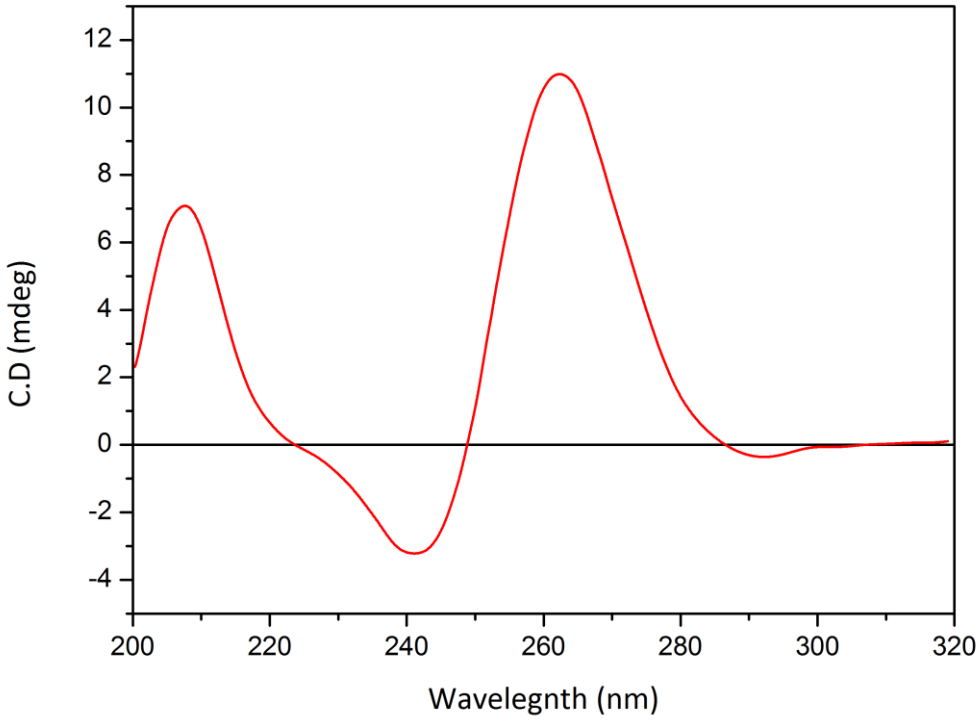

Figure 3-17 The circular dichorism spectrum of 20 μM Myc2345 + 200 μM KCl

Three peaks were observed in the CD spectrum shown in Figure 3-17. Two positive peaks can be seen at 210 nm and 263 nm and a negative peak can be seen at 243 nm. These three features have been reported to be characteristic of the parallel G-quadruplex structures which are expected to be formed by Myc2345 in the presence of KCl[85,86]. From this agreement with literature observations it can be concluded that the Myc2345 sequence forms quadruplex structures under the conditions of 10:1 KCl to DNA concentration. The results of the EVV 2DIR investigation of these structures can be seen in chapter 5.

4 Detecting Protein inhibitor interactions by EVV 2DIR

Section 2.3 described EVV 2DIR's ability to detect electrical, through-space vibrational couplings and how this has previously been exploited to detect intermolecular complex formation. This ability, combined with the technique's proven efficacy for studying large biomolecules, provides great potential for the study of the formation of many biomolecular complexes, particularly inhibitor–protein complexes, with potential applications in drug discovery. Upon inhibitor–protein complex formation, the inhibitor molecule is brought into a close configuration with the protein molecule. This will result in inhibitor vibrations being electrically, through space coupled to vibrations of the protein, giving rise to new EVV 2DIR peaks. Spectral signatures of inhibitor binding may also arise from phenomena such as, induced geometry changes of either molecule or exclusion of ordered water from the protein binding site.

EVV 2DIR can be considered as an optical analogue of 2D-NMR[87,88], a technique which has already been applied to the drug discovery process4. Contrary to 2D NMR, EVV 2DIR does not suffer from the need for relatively large masses of sample and intrinsic

limits to the molecular size of samples without heavy atom labelling. There have been previous examples of 2DIR studies of protein-inhibitor complexes however they have been largely limited to looking at carbonyl interactions.

The aims of the work described in this chapter were to identify and assign new spectral features arising from SU-5402 FGFR1 complex formation and to display EVV 2DIR's capacity as a tool to study the geometry of protein-inhibitor binding.

The approach used was to compare the spectral features of SU-5402 bound to FGFR1 with those of SU-5402 in the presence of the non-specifically binding protein BSA. This was done using difference spectra, produced by subtracting protein only spectra from protein + SU-5402 spectra, allowing for comparison of exclusively the SU-5402 contributions. Six binding dependent peaks were identified in this way and further confirmed to be binding specific by studying their intensity as a function of inhibitor to protein molar ratio. The binding sensitive peaks were assigned to vibrational modes with the aid of calculated spectra. Finally EVV spectra were recorded of the SU-5402 FGFR1 complex in two different polarisation schemes to explore how well EVV 2DIR spectroscopy can be used to determine the interaction geometry of inhibitor-protein complexes.

4.1 Experimental details and methodology

4.1.1 Sample composition

The FGFR1 samples used in the work described here were supplied by AstraZeneca as a construct of the kinase domain only, of length 334 amino acids. The protein was supplied at a concentration of 0.1 mM in a buffer of 40 mM HEPES pH 7.5, 200 mM NaCl, 1 mM TCEP. This buffer composition of relatively high buffer and salt concentration is typical for biochemical work, however is not compatible with the preparation of hydrated gel spots as described above. After drying, the gel spot volumes will be considerably less than those of the initial protein solution depositions, therefore the use of such high salt and buffer concentrations would result in gel spots approaching or exceeding salt saturation. In the event of salt precipitation,

the resultant crystals can cause the EVV beams and signals to scatter, as well as allow for other nonlinear processes to occur due to their symmetry properties. In the event of salt concentrations approaching saturation, the hygroscopicity of the salts can play a major role in final hydration state of the gel, preventing proper gel drying.

In addition to the high buffer and salt concentration, the protein itself in the samples, as supplied from AstraZeneca, is at too low a concentration (0.1 mM) to form gel spots of sufficient depth to give adequate EVV 2DIR signals. In order to increase the FGFR1 concentration and reduce the buffer concentration, the protein samples were buffer exchanged and concentrated using high molecular weight cut off centrifuge concentration columns. It was found that even after dilution, the hygroscopicity of HEPES made it unsuitable for forming gel spots, so the HEPES buffer was exchanged to a phosphate buffer. The final composition used for preparing protein gel spots was 1 mM protein, 5 mM phosphate buffer, 5 mM NaCl.

Many kinase inhibitors display low solubility in pure water. This can come as a result of part of the favourability of binding being due to the hydrophobic effect, and the targeting of hydrophobic pockets within kinases[10, 89]. This leads to the selection of hydrophobic molecules as inhibitor candidates. This presents itself as an issue when trying to create samples of relatively high (1 mM) FGFR1 concentration, with equal or higher concentrations of SU-5402 molecules. Fortunately, many hydrophobic small molecules are readily soluble in DMSO, which is miscible with water and has a similar volatility. This was exploited to produce FGFR1 SU-5402 solutions by adding 2% by volume additions of SU-5402 DMSO solutions. DMSO at this concentration is tolerable with respect to protein stability and chemistry.

The following four gel spots were produced, in accordance to the methodology in section 3.3 and maintained at 75.5 % relative humidity:

1) 1 mM FGFR1, 5 mM phosphate buffer, 5 mM NaCl, 2 %/ vol DMSO, pH 8
2) 1m M FGFR1, 1 mM SU-5402, 5 mM phosphate buffer, 5 mM NaCl, 2 %/ vol DMSO, pH 8
3) 1 mM BSA, 5 mM phosphate buffer, 5 mM NaCl, 2 %/ vol DMSO, pH 8
4) 1 mM BSA, 1 mM SU-5402, 5 mM phosphate buffer, 5 mM NaCl, 2 %/ vol DMSO, pH 8

4.1.2 Collection of EVV 2DIR Data

EVV 2DIR spectra of the four samples described above were recorded. Data was collected between the ranges ω_α = 1250 – 1750 cm^{-1} and ω_β = 2600 – 3450 cm^{-1} at 5 cm^{-1} resolution averaging 100 samples per point. All three beams were set to propagate with P polarisation and time delays of T_1 = 1.75 ps and T_2 = 1.0 ps were used to give a good compromise between minimising non-resonant background and maximising resonant signal size. The spectrum of each sample was recorded three times with the spectra off all four samples being recorded once before repeating all four again in the same order. This is to prevent any minor drift in the laser system over the ~20 hrs total data collection impacting the four samples differently. Each of the three repeats was recorded at slightly different positions on the gel spot to prevent any anomalous spot dominating the data collection.

The data was processed in MATLAB, as described in section 3.2. The individual repeats of each measurement were intensity normalised as discussed below.

4.1.3 Normalisation of EVV 2DIR spectra

The EVV 2DIR signal intensity is proportional to the square of the number of molecules in the focal volume of the three input beams. Therefore collecting spectra at different positions on a sample which is not completely uniform in thickness will result in different absolute signal magnitudes. In order to make the three repeats of each of the four samples comparable, they were normalised such that the major peak at ω_α = 1470 cm^{-1} / ω_β = 2945 cm^{-1} arising from protein methyl vibrations had an intensity of 1 – the intensity of each data point in the spectrum was divided by the intensity of the centre of the methyl peak. This allowed for the three spectra for each sample to be averaged and for the mean spectra for each of the four samples to be compared.

When deciding which protein peak to use for normalisation, there were two criteria considered:

1) The peak should be of high intensity to minimise the effect of any noise or artefacts.
2) The peak should be well resolved from any SU-5402 peaks.

Point 2 is significant as if a protein peak is chosen which is spectrally coincident with an SU-5402 peak, the normalisation of the protein + SU-5402 samples will result in lower intensity of protein features than would be correct.

If an appropriate protein peak has been chosen for normalisation, the ratio of this peak to other protein peaks (assuming they too are fully resolved from SU-5402 features) will be constant, for each protein, with and without the addition of SU-5402.

It was found that by normalising on the large protein methylene peak at ω_α = 1470 cm^{-1} / ω_β = 2945 cm^{-1}, the peak intensity of protein peaks which were well resolved from SU-5402 peak, were constant with and without the addition of SU-5402, to within the consistence of the individual measurements. This was true for both FGFR1 and BSA.

4.2 EVV 2DIR spectra of proteins and SU-5402

4.2.1 Protein Spectra

Figure 4-1 (a) and (b) show the processed (as described in section 3.2) EVV 2DIR spectra of FGFR1 and BSA, respectively. They each contain over 100 resolvable peaks of intensity above the 3σ level, where σ is noise, measured as the standard deviation of the signal free regions, over a dynamic range of over 400. As described in 2.3.2, proteins have previously been studied in the Klug group using EVV 2DIR, and assignments have been made to several protein peaks which are visible in the protein spectra presented here: the two largest spectral features seen at ω_α = 1470 cm^{-1} / ω_β = 2945 cm^{-1} and ω_α = 1640 cm^{-1} / ω_β = 3300 cm^{-1} correspond to methylene and amide vibrations respectively. Peaks arising from phenylalanine side chain vibrations can be seen at ω_α = 1470 cm^{-1} / ω_β = 3050 cm^{-1} and ω_α = 1480 cm^{-1}/ ω_β = 3070 cm^{-1}. An assigned tyrosine peak can also be seen at ω_α = 1530 cm^{-1} / ω_β = 3130 cm^{-1}. Further efforts to assign and characterise the protein spectra have not been made here, as the focus of this work was on the spectral features arising from inhibitor molecules, as opposed to those solely due to the protein.

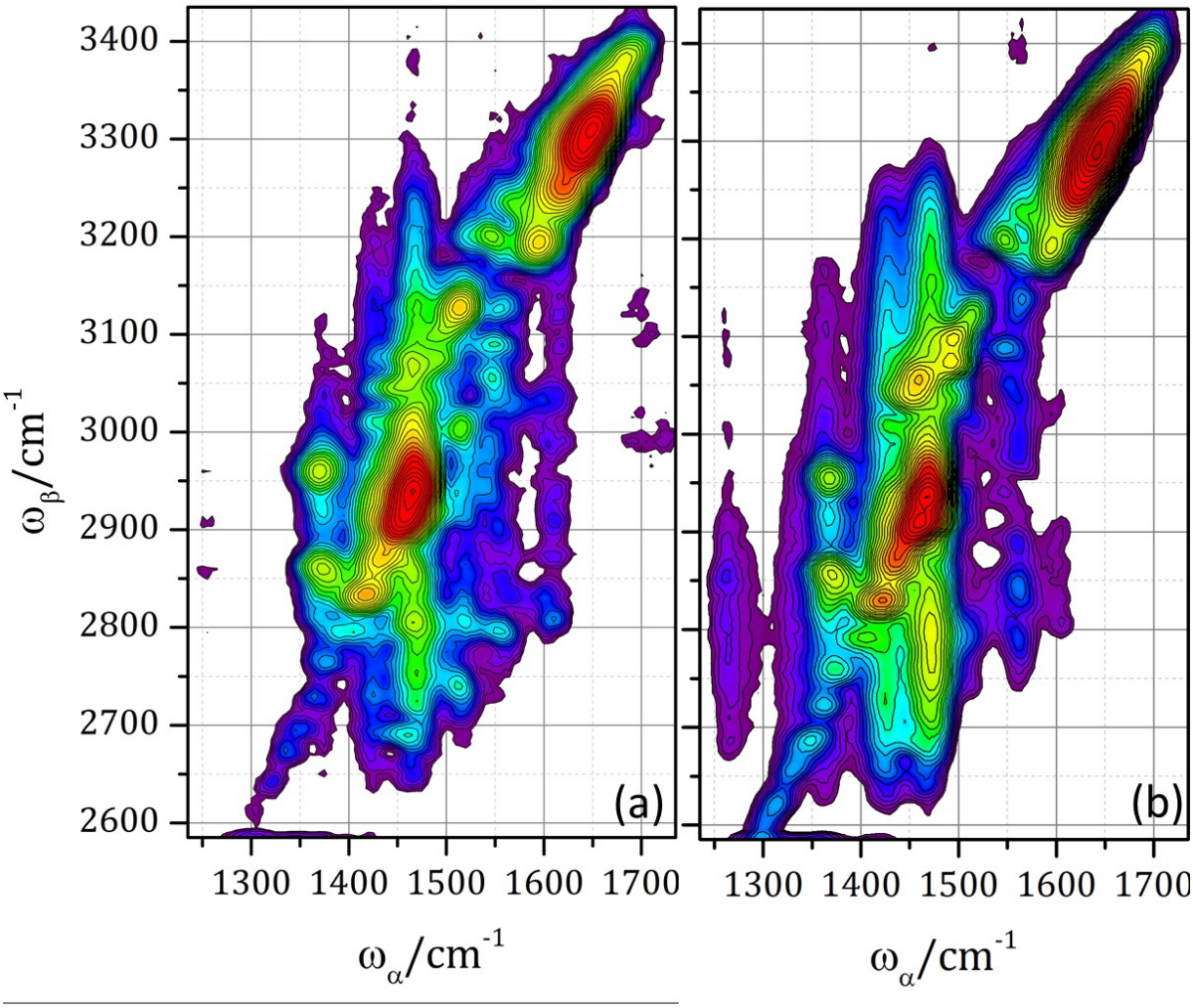

Figure 4-1 The EVV 2DIR spectra of (a) FGFR1 kinase domain, and (b) BSA. Spectra recorded at 5 cm^{-1} of gel phases held at 75 % RH.

Figure 4-2 (a) and (b) show the EVV 2DIR spectra of FGFR1 + SU-5402 and BSA + SU-5402, respectively, with both samples containing a 1:1 molar ratio of inhibitor to protein. By eye, there are significantly noticeable differences between these spectra and the spectra in Figure 4-1, with Figure 4-2 showing many new peaks which must correspond to vibrational couplings involving SU-5402. The fact that these new peaks are present at such high intensities demonstrates that resonance enhancement of the SU-5402 signals is occurring. There are nine phenylalanine residues in the FGFR1 kinase domain construct used in these samples, and the resultant peak at ω_α = 1470 cm^{-1} / ω_β = 3050 cm^{-1} is of comparable intensity to many of the SU-5402 peaks. In a

case where SU-5402 exhibits no resonance enhancement its peaks from aromatic vibrations could be expected to be of comparable intensity to a single phenylalanine reside. This would put them at around 1/ 81 of their current intensity, given that EVV signal increases as the square of the number of molecules in the beams. From this brief analysis is seems reasonable to estimate the degree of resonance enhancement observed to be around 50 to 100 fold. By considering only the spectra displayed in Figure 4-2 as they are, a SU-5402 feature exclusive to the binding protein spectra can be seen at $\omega_\alpha = 1660\ cm^{-1}$ / $\omega_\beta = 3240\ cm^{-1}$.

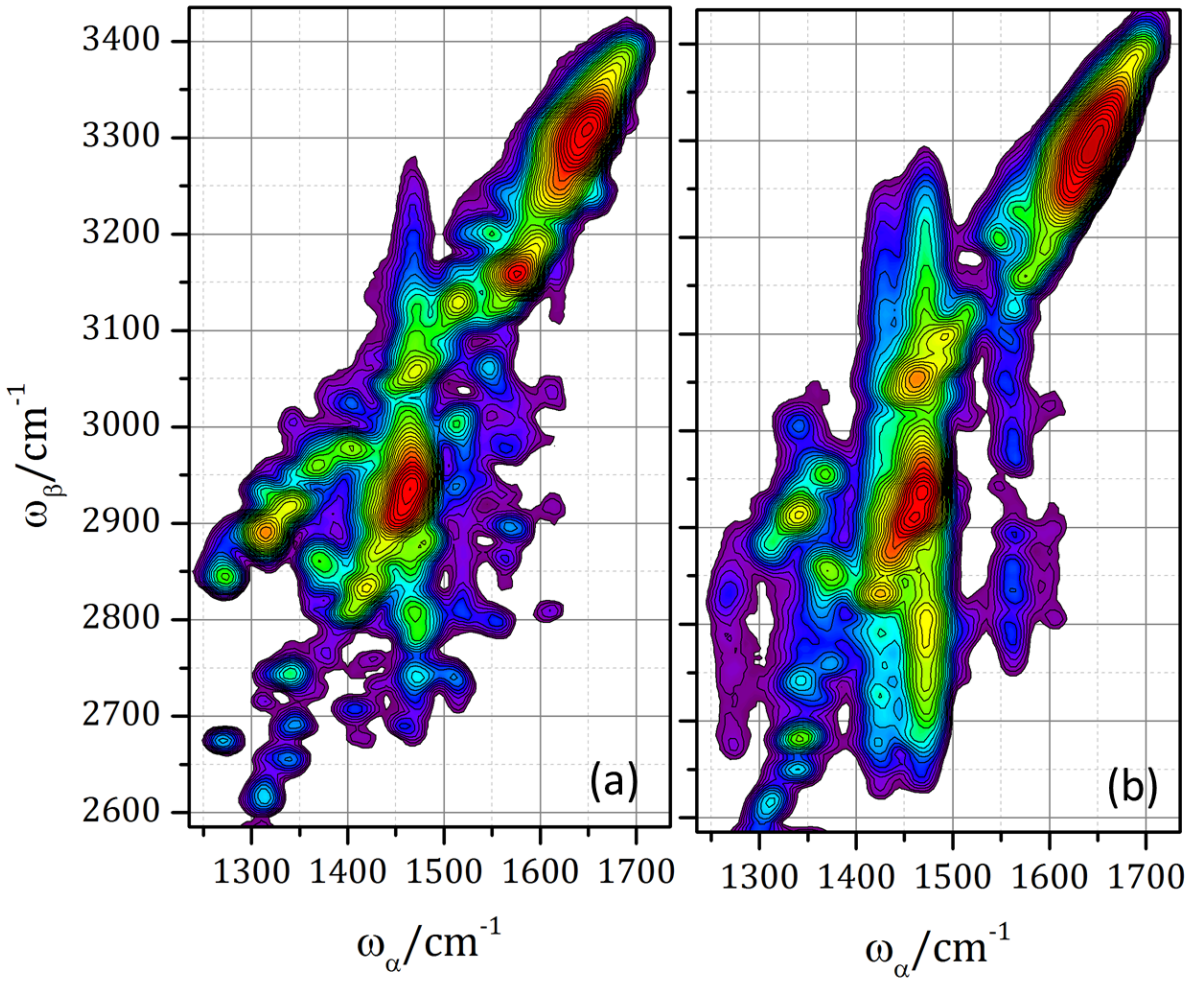

Figure 4-2 The EVV 2DIR spectra of (a) FGFR1 kinase domain + SU-5402 at 1:1 molar ratio, and (b) BSA SU-5402 at 1:1 molar ratio. Spectra recorded at 5 cm^{-1} of gel phases held at 75 % RH.

4.2.2 EVV 2DIR difference spectra and spectral signatures of SU-5402 binding

In many instances the spectrum of a complex sample may be recorded and the spectral contributions of just one of the components may be of interest in isolation. If the EVV spectrum of the complex sample consisting of molecules A and B is known and so is the spectrum of molecule A, the spectral contributions of A may be removed from the spectrum of A+B to leave only the spectral contributions of B (and the contribution of interactions between A and B). EVV 2DIR uses homodyne detection and as a result of this the EVV signal is proportional to the square of the number of molecules in the confocal volume of the three input beams as shown in Equation 4-1. As a result of this, a correct difference spectrum cannot be produced by taking the linear subtraction of the spectrum of B from the spectrum of A+B due to the cross terms shown in Equation 4-2

$$Signal_A \propto [A]^2$$

Equation 4-1

$$Signal_{AB} \propto ([A] + [B])^2 = [A]^2 + [B]^2 + 2[A][B]$$

Equation 4-2

Therefore in order to correctly calculate EVV 2DIR difference spectra, the square rooted intensities of the spectrum of A must be subtracted from the square rooted intensities of the spectrum of A+B. This yields the difference spectrum with square rooted intensities, which are linearly proportional to sample concentration. These can subsequently be squared again to make them comparable to the EVV spectra, as measured.

In order to allow for direct comparison and analysis of the SU-5402 contributions to the protein/ SU-5402 spectra shown in Figure 4-2, the protein contributions must first be removed. Doing so prevents the incomparable protein contributions convoluting the comparison and also may uncover SU-5402 peaks which would otherwise be obscured by larger, coincident protein peaks. These subtractions were made in accordance with the method described above; firstly square rooting the spectra, then subtracting the protein spectra from the protein and inhibitor spectra.

Figure 4-3 (a) and (b) show the difference spectra produced by subtracting the protein only contributions from the protein and SU-5402 spectra. These two difference spectra from FGFR1 and BSA show the spectral contributions of SU-5402 in a bound and nonspecifically bound form, respectively.

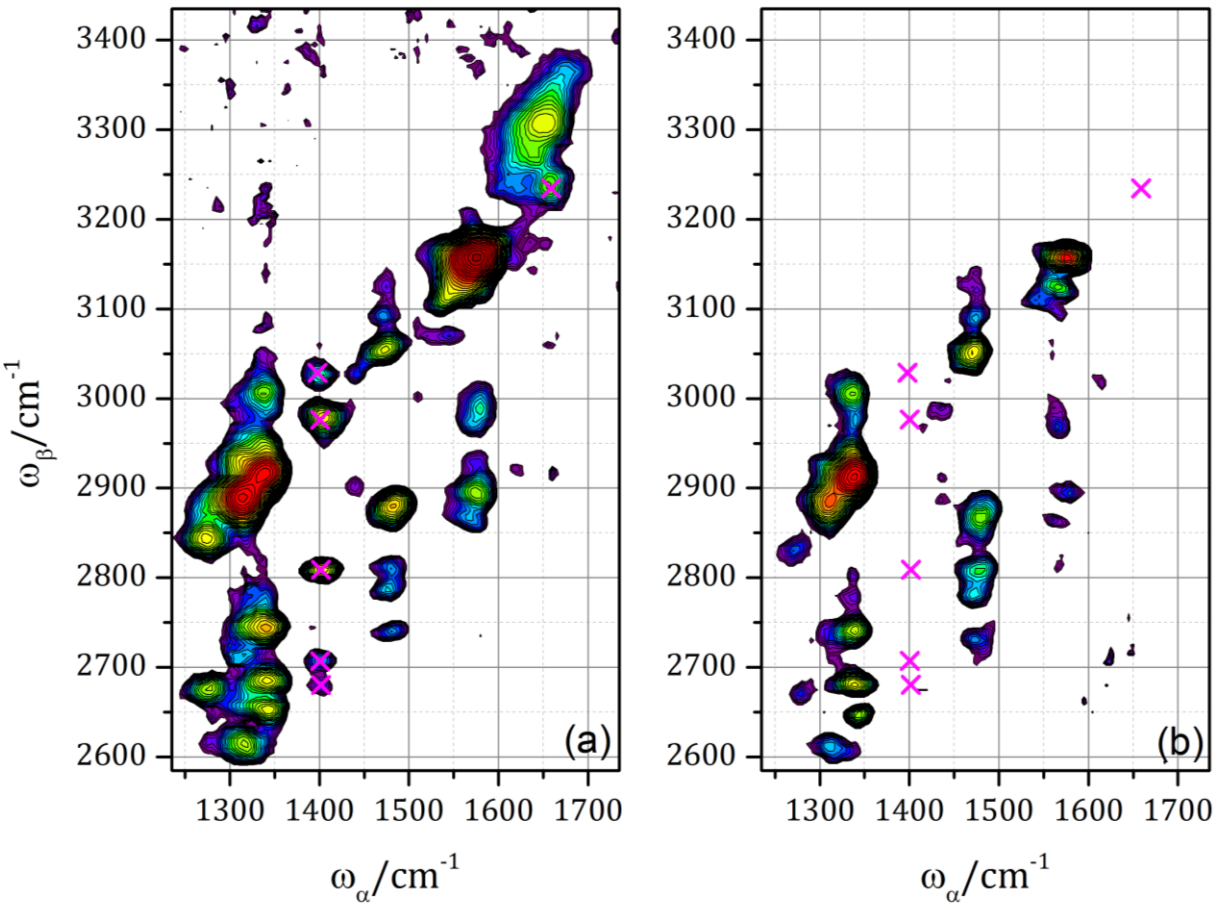

Figure 4-3 EVV 2DIR difference spectra of (a) SU-5402 + FGFR1 kinase domain and FGFR1 kinase domain showing the spectral contributions of bound SU-5402. (b) SU-5402 + BSA and BSA showing the spectral contributions of non-specifically bound SU-5402.

In the non-specifically bound difference spectra there are a total of 65 peaks of intensity above the 3σ level, where σ is noise, measured as the standard deviation of the signal free regions. Peaks above this level can be considered to have a 99.7% chance of being real spectral features.

All the peaks present in the unbound difference spectra are also present in the bound difference spectra with the addition of six more peaks at frequencies:

- $\omega_\alpha = 1400\ cm^{-1}$ / $\omega_\beta = 3030\ cm^{-1}$
- $\omega_\alpha = 1400\ cm^{-1}$ / $\omega_\beta = 2975\ cm^{-1}$
- $\omega_\alpha = 1400\ cm^{-1}$ / $\omega_\beta = 2810\ cm^{-1}$
- $\omega_\alpha = 1400\ cm^{-1}$ / $\omega_\beta = 2705\ cm^{-1}$
- $\omega_\alpha = 1400\ cm^{-1}$ / $\omega_\beta = 2680\ cm^{-1}$
- $\omega_\alpha = 1660\ cm^{-1}$ / $\omega_\beta = 3240\ cm^{-1}$

These new peaks have been marked with crosses in the bound difference spectrum, Figure 4-3a, and their locations have been marked with crosses in the unbound spectrum, Figure 4-3b, to emphasise their absence. Given that these peaks are only present in the bound spectra, it was hypothesised that they were due to new interactions between the inhibitor and the binding site, or new intramolecular interaction within the inhibitor, as a result of a geometry change upon binding.

In addition to these six new peaks, there is also a broad feature around $\omega_\alpha = 1650\ cm^{-1}$ / $\omega_\beta = 3300 cm^{-1}$, present exclusively in the specifically bound difference spectrum. This feature is at the frequency of the amide band in the protein spectra and assumes the same shape and can therefore be assigned to be an FGFR1 amide signal. The fact that an equivalent it is not observed in the BSA difference spectrum implies that it a real binding dependent signal, potentially arising due to conformational changes of part of the protein amide backbone, upon SU-5402 binding.

By looking at the bound and non-specifically bound difference spectra it can be seen that there is a large degree of similarity between the two and that the vast majority of

spectral features are present in both difference spectra. These can be assumed to be arising from pairs of coupled inhibitor vibrations which are not changing upon inhibitor binding.

4.3 Concentration dependence of inhibitor binding peaks

The fact that 6 peaks were present exclusively in the bound difference spectrum, suggested that they were due to binding specific interactions. In order to further confirm this to be the case, EVV 2DIR spectra were recorded of samples of SU-5402 and FGFR1 across a range of molar ratios, from no inhibitor content through to 2:1 SU-5402:FGFR1 molar ratio. The reason for doing so was to study how the intensity of these six hypothesised binding peaks varied as a function of SU-5402:FGFR1 molar ratio, compared to the 59 peaks that were present in both difference spectra. In the event that these six peaks were due to specific binding interactions between SU-5402 and the kinase binding site, their intensity should not increase upon increased SU-5402 concentration, above a 1:1 molar ratio of SU-5402:FGFR1. After a 1:1 molar ratio, all of the FGFR1 binding sites should be occupied and any excess of SU-5402 would remain unbound. It also follows that the square of the intensity of all 56 peaks present in both the bound and unbound difference spectra will increase linearly in intensity (EVV signal increases as the square of the number of molecules in the confocal volume) with continuing increases in SU-5402 concentrations, beyond a 1:1 molar ratio. These peaks that continue to increase in intensity after a 1:1 ratio will be arising from the coupling of two vibrational modes which are both centred on the inhibitor molecule and are unchanged by the process of binding.

4.3.1 2D Peak Fitting

As seen in the difference spectra, Figure 4-3 a and b, many of the peaks arising from SU-5402 are not fully resolved from one another and display some degree of overlap. These results in cases where the point of local maximum intensity is not at the central frequency of the feature giving rise to it and the individual intensity of the two features become convolved. This is exemplified in Figure 4-4 which shows a simulated cut through of an EVV spectrum in which two Gaussian features are shown alongside the sum of their contributions. It is clear that an increase in intensity of either of the features would result in the increase in total intensity at the central frequency of the other.

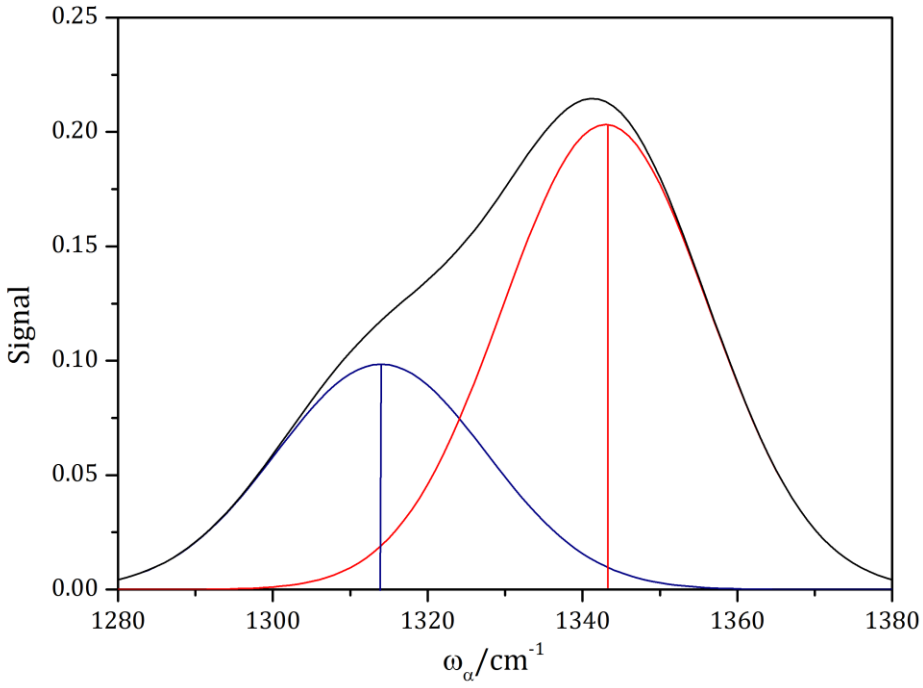

Figure 4-4. A simulated cut through of an EVV spectrum, biased on the FGFR1, SU-5402 difference spectrum at $\omega_\beta = 2745$ cm^{-1}. The red and blue curves describe two Gaussians with their centres marked and the black curve is their sum.

As a result of this, when considering how the intensity of the SU-5402 peaks behave as a function of SU-5402:FGFR1 molar ratio, the intensity of the spectrum at the apparent peak centre is a poor measure of the true intensity of the spectral feature.

In order to be able to correctly resolve the intensities and locations of these unresolved peaks, the experimental data can be fitted to a model containing as many terms as spectral features which are to be considered.

In order to study the SU-5402 peaks behaviour as a function of SU-5402:FGFR1 molar ratio, a fitting program was written in MATLAB, utilising MATLABs in built "fit" function which allows for a non-linear least squares analysis to be performed. Historically EVV 2DIR spectroscopy, as performed in the Klug group, has empirically given rise to Gaussian shaped features and this was observed again in the data presented in this thesis. For this reason, the MATLAB fitting program was written to fit the data with as many 2D Gaussians as were identifiable in the data.

Clearly, in order to fit an EVV 2DIR spectrum, the number of peaks to be fitted, along with start points for the frequency location, intensity and width parameters are needed. Previous versions of the fitting program attempted to identify peak locations by considering the first and second derivatives of the data, with respect to the two frequency dimensions. This approach, whilst partly successful was not wholly reliable as a single incorrectly identified peak could be sufficient to undermine large parts of the rest of the spectrums fitting. This approach was ultimately abandoned (until such a time that it can be worked on again) in favor of selecting the peak frequency position start points by hand. Slight errors in choice of peak position were not significant to the success of the fit as these only acted as starting positions, which could be varied by the program. The start points for intensity were taken to be the intensity of the data at the location start point and the width start points were taken to equivalent to the 25 cm^{-1} full width half max of the laser pulses. Upper and lower limits could be imposed on the values of all of the parameters: The intensity parameters were limited to be greater or equal to zero, the locations were allowed to vary by 5 cm^{-1} in each frequency dimension and the widths were limited to be greater than 18 cm^{-1} and less than 35 cm^{-1}. These restrictions were required as without them, physically impossible

parameter values were resultant on occasions, and peak locations would move to fit several Gaussians to single large features.

Very congested spectral regions such as the amide band of protein spectra clearly need to be fitted with many Gaussians, due to their shape, but are far too convoluted to produce any useful information by doing so. Nonetheless, it is still important to do so, such that their intensity is "accounted for" and does not affect the quality of fit for nearby peaks.

In order to study the SU-5402 peaks in the difference spectrum as a function of SU-5402:FGFR1 molar ratio, first the peaks' locations were determined for input into the fitting program. This was done using the mean 2:1 SU-5402 difference spectrum. This spectrum was chosen as it would offer the highest levels of signal but also as it may display features arising exclusively due to unbound inhibitor. These exclusively unbound peaks would only present themselves at molar ratios greater than 1:1. After identifying peak locations, the 2:1 difference spectrum was fitted as described above. The rest of the concentration series of mean difference spectra were then fitted using the peak locations and widths determined by the 2:1 fit, as starting points.

71 Gaussians were fitted to the SU-5402/ FGFR1 2:1 difference spectra. This included 65 peaks identifiable above the 3σ level and 6 Gaussians which were used to fit the broad amide band residual around ω_α = 1650 cm^{-1}/ ω_β = 3300 cm^{-1}. The initial fit to the 2:1 difference spectrum yielded an R^2 value of 0.972 and none of the fit parameters hit their restriction limits. This suggested appropriately chosen starting locations and widths were used. The result of this fitting can be seen in Figure 4-5.

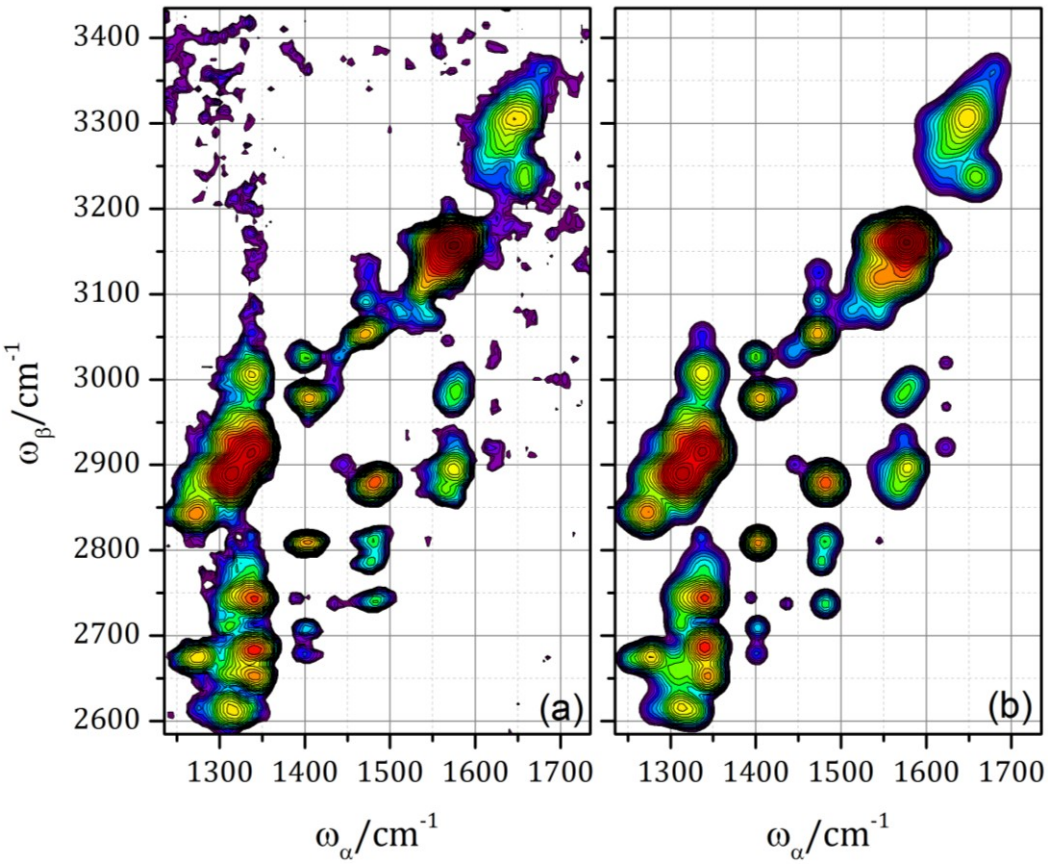

Figure 4-5 (a) shows the 2:1 SU-5402/ FGFR1 difference spectrum. (b) shows the fit to the 2:1 SU-5402/ FGFR1 difference spectrum using 71 two dimensional Gaussians

Figure 4-5 shows very good agreement between the experimental data and the fit to that data, given the complexity of the data. Although the quality of the fit is confirmed by the R^2 value, the representation given in Figure 4-5, whilst giving a good overview, does not allow for a close comparison of the two in the z dimension. One dimensional cut throughs of the experimental data and the fits at three different values of ω_α can be seen in Figure 4-6 Figure 4-7 Figure 4-8

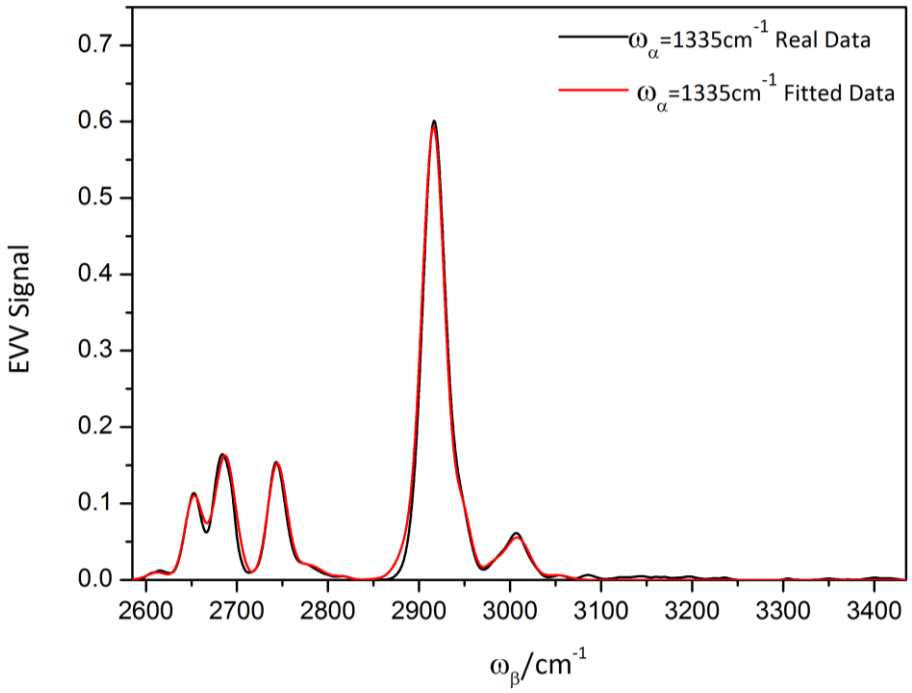

Figure 4-6 Cut throughs comparing the real and fitted EVV 2DIR spectrum at ω_α = 1335 cm^{-1} of the 2:1 SU-5402 FGFR1 difference spectra.

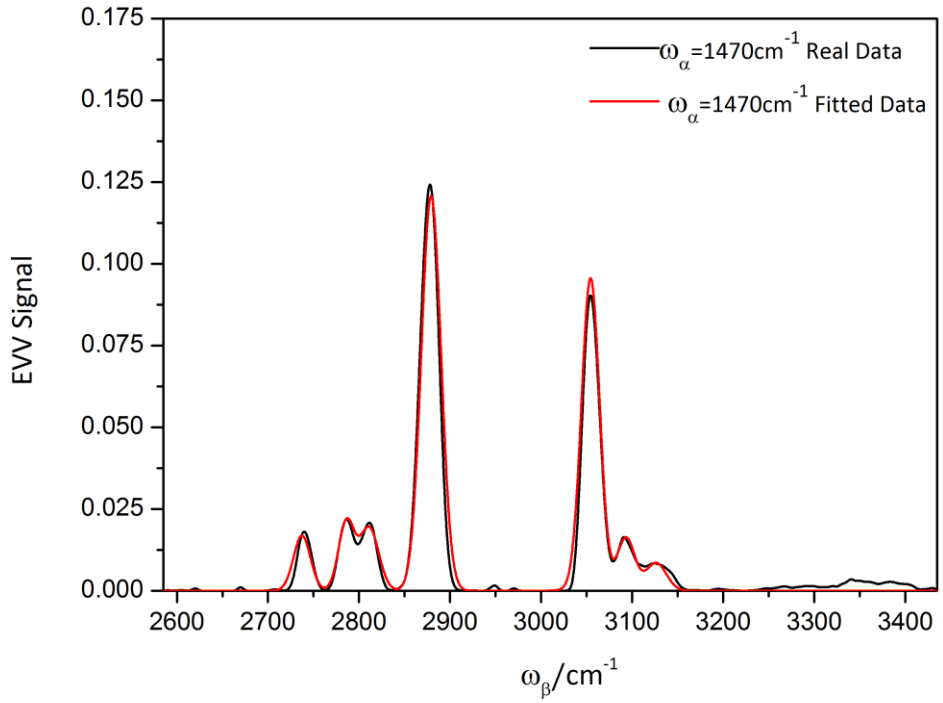

Figure 4-7 Cut throughs comparing the real and fitted EVV 2DIR spectrum at ω_α = 1470 cm^{-1} of the 2:1 SU-5402 FGFR1 difference spectra.

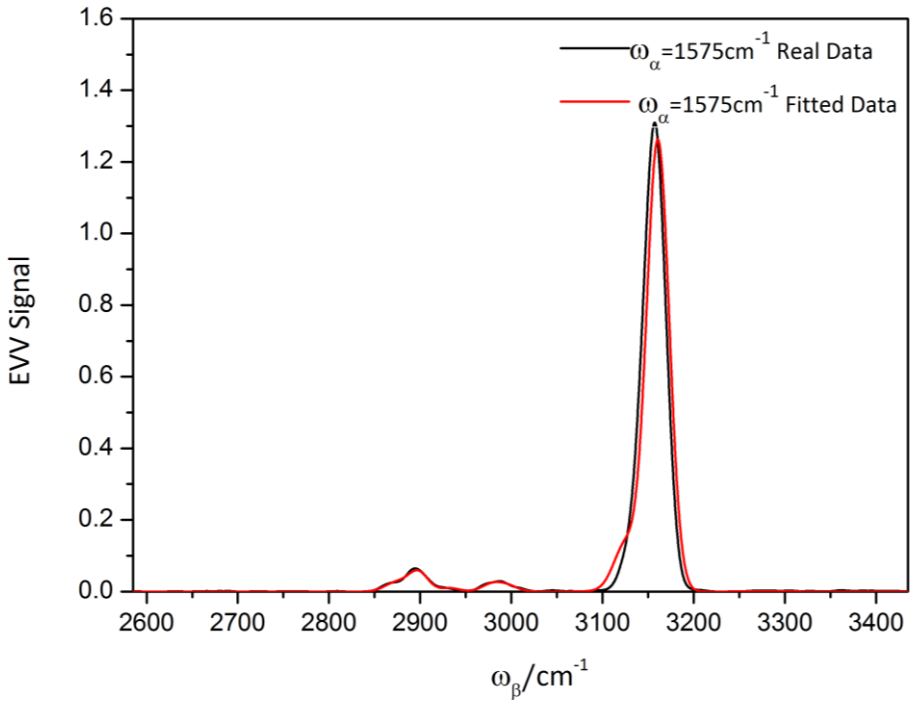

Figure 4-8 Cut throughs comparing the real and fitted EVV 2DIR spectrum at ω_α = 1575 cm^{-1} of the 2:1 SU-5402 FGFR1 difference spectra.

Figure 4-5 shows some small inconsistencies between the fit and experimental data with respect to the widths of asymmetric peaks, such as that at ω_α = 1470 cm^{-1}/ ω_β = 2740 cm^{-1}. Figure 4-7 however illustrates that, although in these asymmetric cases the fitted widths are not exact reproductions of the experimental widths, the intensities of the fit and experimental data agree.

After holding the position and width parameters constant, as determined by the 2:1 fitting, the other difference spectra in the concentration series were fitted. All the subsequent fits yielded R^2 values of 0.97, when rounded to two decimal places.

4.3.2 Concentration series sample preparation and data collection

The samples for the concentration series of SU-5402 were prepared according to the same methodology described in sections 4.1.1 and 4.1.2, to produce gel phase samples. The six samples produced were composed of 1 mM FGFR1, X mM SU-5402, 5 mM phosphate buffer, 5 mM NaCl, 2 %/ vol DMSO, where X assumed the values of 0 mM, 0.5 mM, 0.75 mM, 1 mM, 1.5 mM and 2 mM. Spectra were collected in the same way as those presented in section 4.2, across the same spectral range, again with three repeats of each measurement being made at different positions on the gel sample spot.

4.3.3 Peak responses

The square root of the fitted peak intensities of the six peaks that were present exclusively in the binding difference spectra and absent in the BSA control can be seen as a function of SU-5402:FGFR1 molar ratio Figure 4-9.

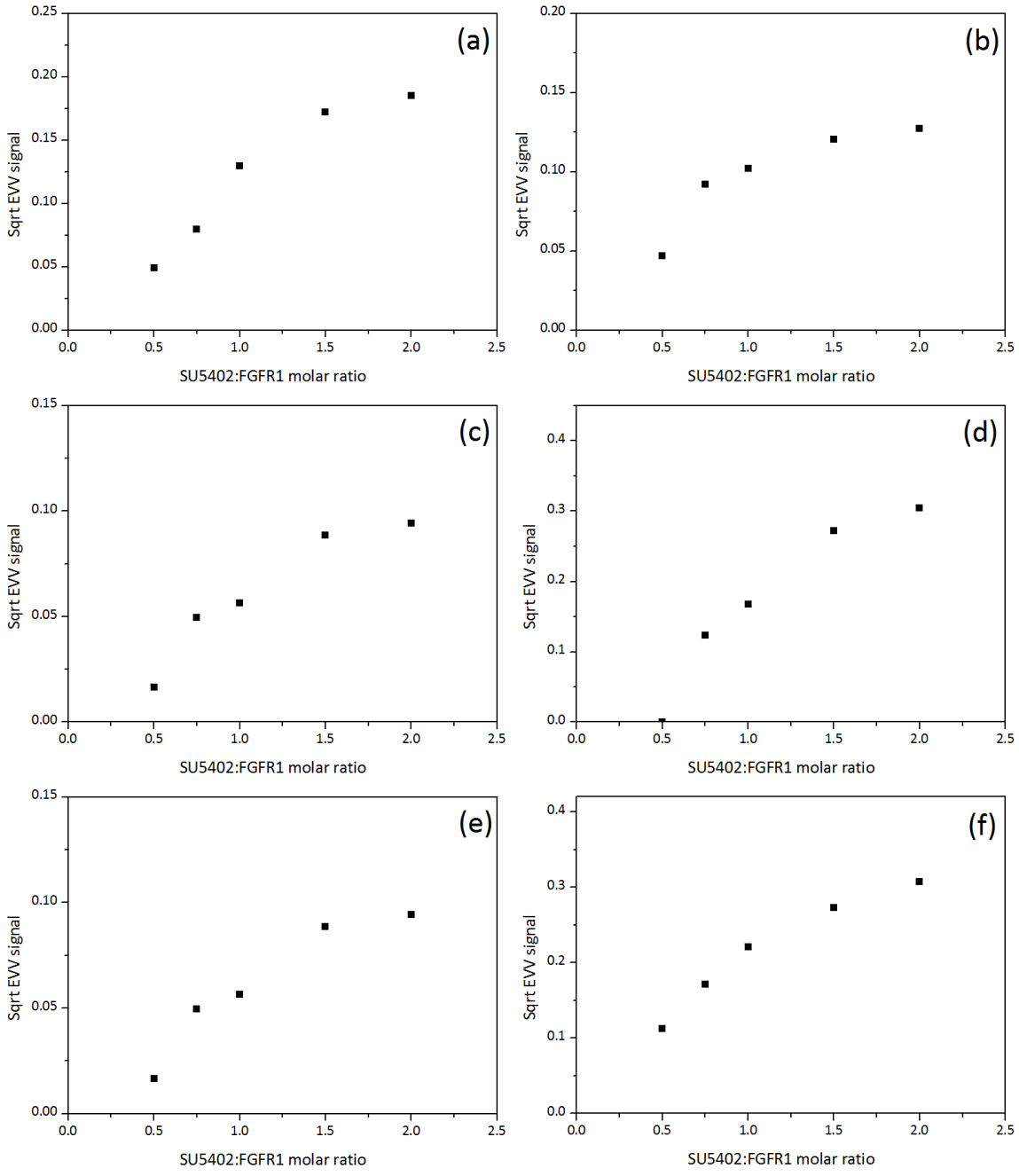

Figure 4-9 Plots showing the plateauing peak responses of the six binding dependent cross peaks as a function of SU-5402 FGFR1 molar ratio.

It was hypothesised that peaks arising exclusively from specifically bound SU-5402 would increase in intensity, up to a 1:1 molar ratio and then plateau. Figure 4-9 shows

a slowing in the rate of peak increase, with molar ratio above 1:1, but not a complete flattening.

Peaks present in both the bound and non-bound difference spectra were expected to display a linear increase in their square rooted intensity with increasing SU-5402:FGFR1 ratio. This was found to be the case for the binding independent peaks. As the number of observable peaks in the SU-5402/ FGFR1 difference spectra is so high, it would be tedious to depict all the non-binding peaks' linear responses to concentration. For contrast with the responses of the six binding sensitive peaks, the response of every tenth non-binding peak, when ordered by ω_α frequency, is shown in Figure 4-10. Included in this representative sample are peaks of both high and low intensity.

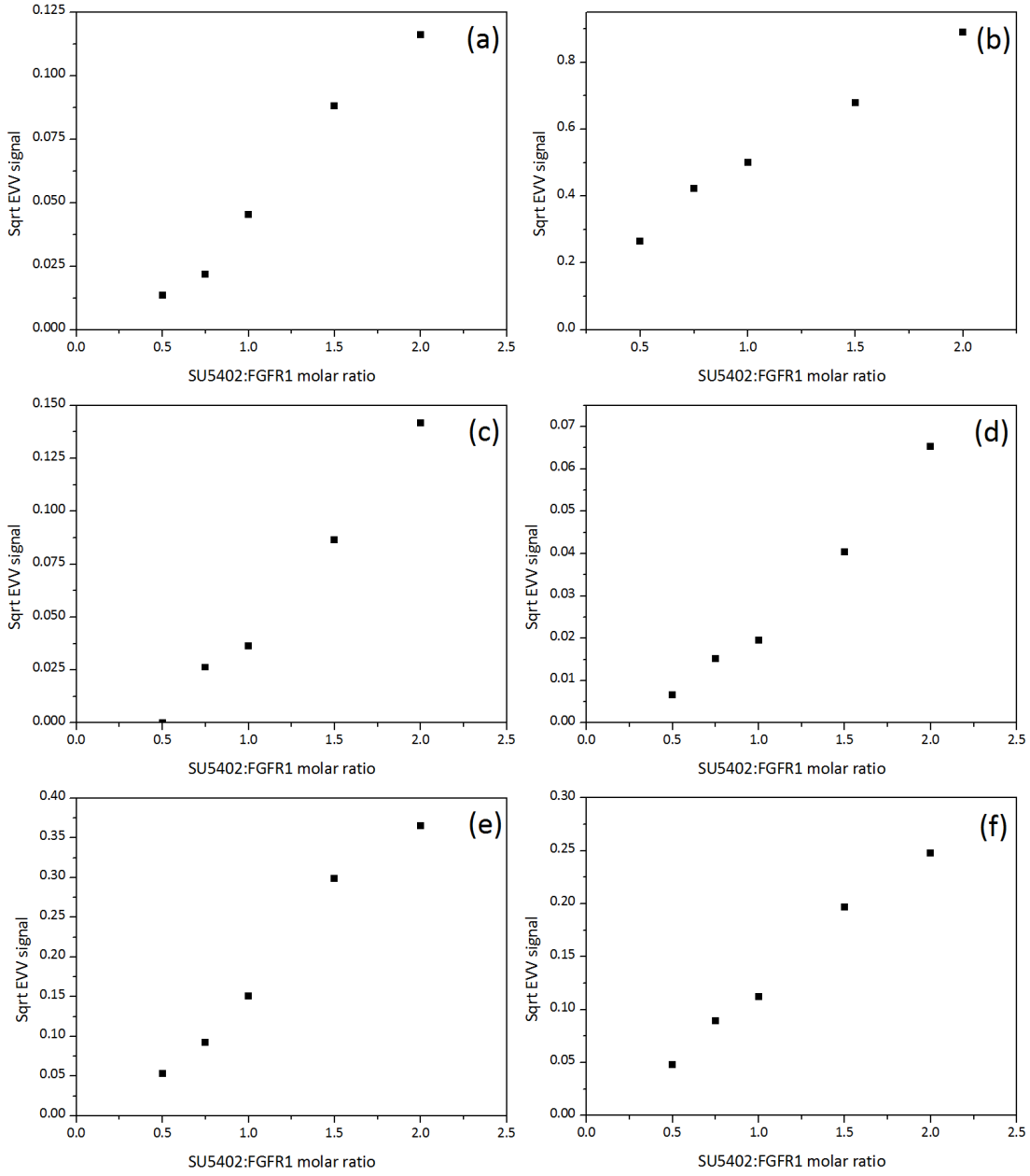

Figure 4-10 Plots showing the linear peak responses of a representative sample of six binding independent cross peaks as a function of SU-5402 FGFR1 molar ratio.

Comparing Figure 4-9 and Figure 4-10 it can be seen that the peaks exclusive to the bound difference spectra do display appreciable plateauing compared to the binding

independent peaks, though not an absolute flattening as expected with saturation of binding sites. This non-linear dependency on molar ratio, in addition to their absence in the non-binding BSA control, confirms their presence to be due to vibrational couplings present only when the SU-5402/ FGFR1 complex has been formed. Five of these six binding dependent peaks share a common ω_{α}. This suggests they arise from one vibrational mode coupled to five other different modes, in a way which is only possible upon complex formation.

The trend of the binding peaks continuing to increase in intensity after a 1:1 molar ratio (albeit more slowly) implies either an effective increase in protein concentration or an effective decrease in inhibitor concentration. An effective increase in protein concentration seems highly unlikely as there is no mechanism for this to occur. An effective decrease in inhibitor concentration however, could be achieved by at least two plausible mechanisms: photodamage to SU-5402 and inhomogeneous distribution of SU-5402 between solvents. As SU-5402 has a strong visible absorption close to the frequency of the visible field, it seems plausible that SU-5402 molecules could undergo photodamage, resulting in fragmentation or isomerisation, rendering them unable to complex with the protein binding site. As the samples will have already formed SU-5402/ FGFR1 complexes before exposure to the experimental fields, this process would have to occur on bound SU-5402 molecules, causing them to dissociate from the protein binding site and be replaced. Given the high concentration of the samples, diffusion times should not rule out this process. Fragmentation would result in the breaking of the conjugation system across the SU-5402 molecule, which gives rise to its strong visible absorption, and in turn its resonance enhancement. This would mean that the spectral features of photodamaged fragments would not appear above the protein background in the EVV 2DIR spectra.

SU-5402 displays a very low solubility in water and it was for this reason that high concentration solutions of SU-5402 in DMSO were added to aqueous FGFR1 solutions. DMSO has a slightly lower vapour pressure than water meaning that the final concentration of DMSO in the protein gel spots could be higher than the 2 % by volume of the solutions before drying to gel spots. The lower effective concentration of SU-5402 observed could be explained by a portion of SU-5402 molecules remaining

in a DMSO solvation shell preventing them from being able to take part in complex formation.

4.4 Binding peak assignments from calculated spectra

In order to assign the six binding sensitive cross peaks to the coupled pairs of vibrations which gave rise to them, comparison with calculated spectra was required. All of the calculated spectra presented and described in this thesis were calculated by the group of Wei Zhuang at The Dalian Institute of Chemical Physics and at The Fujian Institute of Research on the Structure of Matter at the Chinese Academy of Sciences. The details of the exact methods used to calculate these spectra are beyond the scope of this thesis. In brief the vibrational frequency calculations were carried out use density functional theory (DFT) with the 6-31g (d, p) basis set[90,91] in the GAUSSIAN 09 program[92]. The calculations of the EVV 2DIR peaks from the vibrations of the model were made in accordant to the method reported by Kwak et al[93]. A frequency scale factor of 0.96 was used to produce the peak frequencies seen in he calculated EVV 2DIR spectrum. The peaks were given a width of 25cm^{-1} FWHM to correspond to the similar widths seen experimentally.

Of course to calculate the EVV 2DIR spectra of any system, a chemical structure is required. For the calculations described here the co-crystal structure of the SU-5402/ FGFR1 complex was used. To perform the necessary quantum calculations on the entire protein crystal structure would be entirely unviable due to the length of time such a calculation would take. In order to minimise the complexity of the calculations, only the structure of the SU-5402 molecule, its surrounding protein atoms and a proximal water molecule were included. Five amino acid residues were chosen to be included in the calculation as they were sufficiently close to the SU-5402 molecule to experience vibrational couplings. In some cases only the proximal parts of the amino acid residues were included, with side chains pointing away from the inhibitor molecule being omitted. As well as being extremely computationally expensive, calculating the spectrum of the entire protein structure would be unnecessary, as this would be calculating the protein spectral features which have been removed when

calculating the difference spectra. A depiction of the structure used for calculating the FGFR1 bound SU-5402 spectrum can be seen in Figure 4-11, with SU-5402 carbon atoms shown in orange, and protein carbon atoms shown in pink. The inclusion of protein atoms will result in the calculation of binding independent protein-protein vibrational couplings, whose contributions would have been subtracted from the experimental data when the difference spectra were produced. However peaks arising from such couplings will likely not obscure the inhibitor peaks of interest as they will be considerably smaller than the inhibitor peaks as they will not be displaying resonance enhancement.

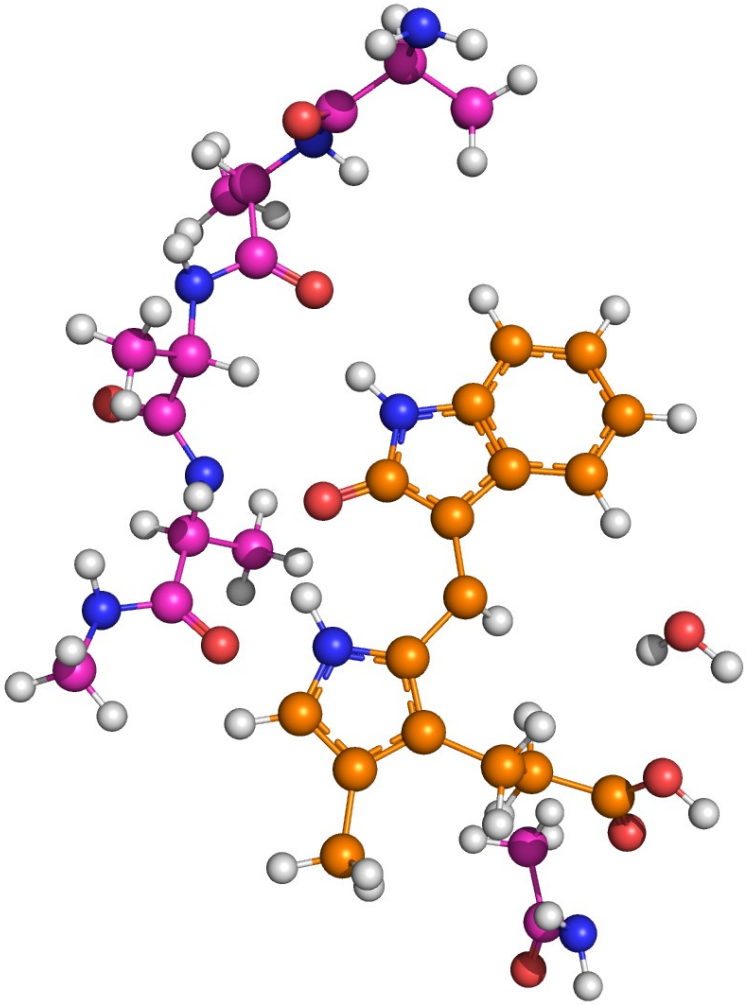

Figure 4-11 The geometry of the SU-5402 molecule along with the five partial amino acid residues and one water molecule used for calculating the EVV 2DIR spectrum of SU-5402 bound to FGFR1. The carbon atoms of the SU-5402 molecule are shown is orange whilst the included protein carbon atoms are shown in pink.

The results of these calculations can be seen in Figure 4-12a, whilst Figure 4-12b shows the experimental 1:1 SU-5402/ FGFR1 difference spectrum for comparison. In order to use the calculated spectrum to assign vibrational modes to the experimental data, it was first necessary to assign the experimental peaks onto their calculated equivalents. The assignment of experimental to calculated peaks can be seen in Figure 4-12, with the letter labels marking equivalents between the two. The assignments shown are not exhaustive as only the assignment of binding sensitive peaks is of

interest here. The process of assigning equivalent peaks in the calculated and experimental spectra shows very good agreement between the two, with many columns of common ω_α and diagonals of common ω_β - ω_α having clear equivalents.

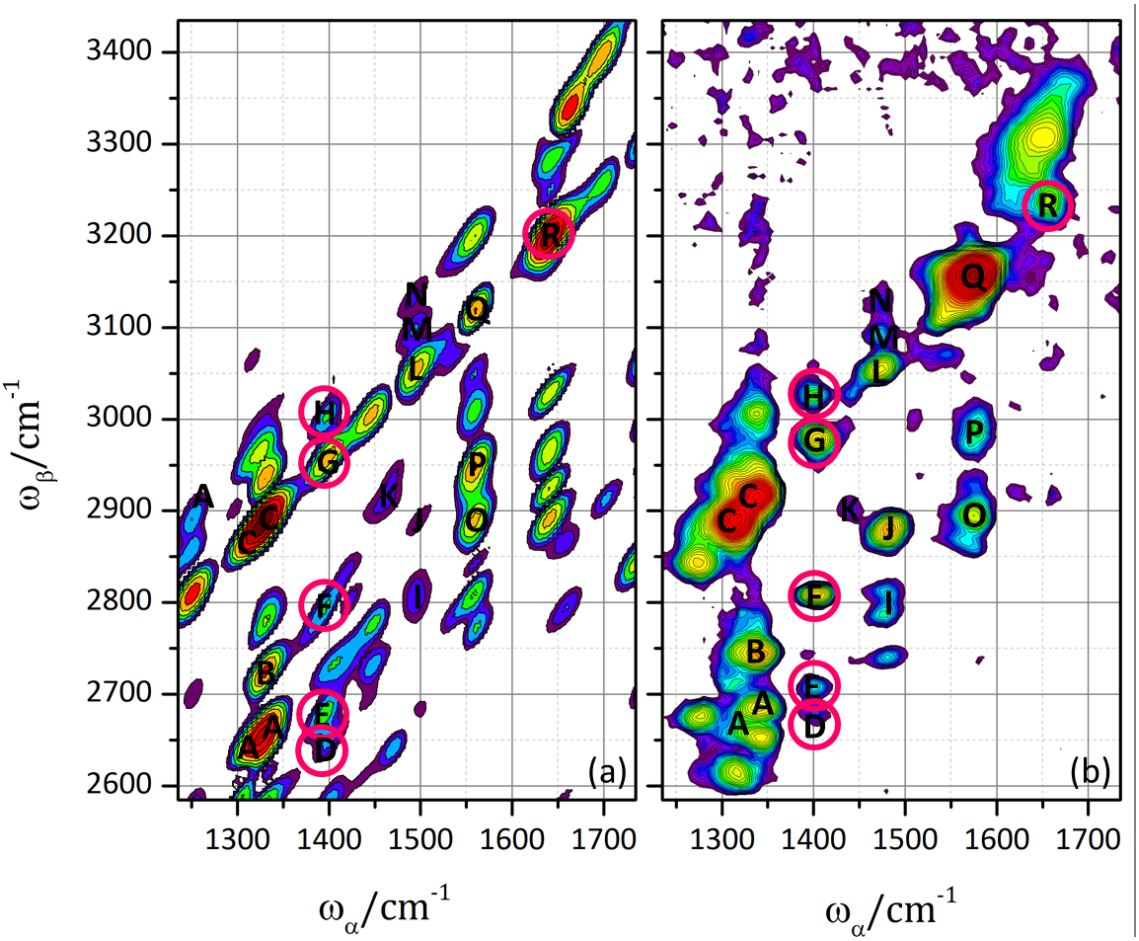

Figure 4-12 (a) the calculated EVV spectrum of SU-5402 and its proximal amino acid residues and water molecule. (b) the experimental 1:1 SU-5402/ FGFR1 difference spectrum for comparison. The letter labels show how the experimental peaks have been assigned onto the calculated peaks. Pink circles have been placed around the binding sensitive peaks.

There are however several peaks in the calculated spectrum which are not present in the experimental difference spectrum such as those seen between peaks P and Q and those below peak R. These extra peaks in the calculated spectrum can be accounted for even if it is assumed that the input geometry is a perfect representation of the experimental sample: the three pulses used in the EVV 2DIR experiment described here have a temporal full width half maximum of about 1.5 ps. Due to this, it is

necessary to have delays between the pulses of 1 – 2 ps to prevent temporal overlap of the pulses and suppress non-resonant signals. Each of the excited state coherences induced through the EVV coherence pathway has lifetime associated with it, with these coherences decaying away, exponentially over time. The delays between the input pulses mean that by the time a subsequent pulse arrives, many of the initially created coherences will no longer be present, resulting in weaker EVV signals being produced at longer delays. In this way, it can be seen that coherences between vibrational sates with very short lifetimes will give rise to much weaker spectral features and that coherences with lifetimes sufficiently short may not be visible at all. The way the calculations are performed does not consider any temporal factors and therefore shows peaks which may arise from states with extremely short lifetimes which are undetectable at the delays used in these experiments. For exactly the same reasons, the relative peak intensity's in the experimental and calculated spectra may also not agree.

By taking the Gaussian output file of the calculated spectrum and viewing it in GausView, it is possible to visualise the vibrational motions which are calculated to occur at given frequencies. By doing so, it was possible to see which specific motions on each atom contributed to the vibrational modes giving rise to the peaks. Seven distinct vibrational modes were observed to contributing to the six binding dependent peaks observed in the difference spectra.

Peaks D, E, F, G and H all fall on a common frequency of ω_α meaning one of the two coupled vibrational modes which give rise to the peaks is common to them all. The fitted experimental ω_α frequency of these five peaks was 1402.2 ± 1.8 cm^{-1} and ω_α component of these couplings was assigned to a calculated vibrational mode at 1401.3 cm^{-1} referred to from here on in as "Mode 1".

Peak	α Mode	β Mode
H	1	3
G	1	4
F	1	5
E	1	6
D	1	7
R	2	4

Table 4.1 The modes assigned to the α and β modes of the six binding dependent cross peaks

Mode 1:

As can be seen in Figure 4-13 all of the vibrational motion of Mode 1 lies on the atoms of the SU-5402 molecule. Motion is seen on all SU-5402 atoms, except for those of the propionic acid group. The contribution of largest amplitude to this mode is an asymmetric ring deformation on the pyrrole group comprised of atoms labelled 38 – 42. This mode forms the α mode of peaks D, E, F, G and H.

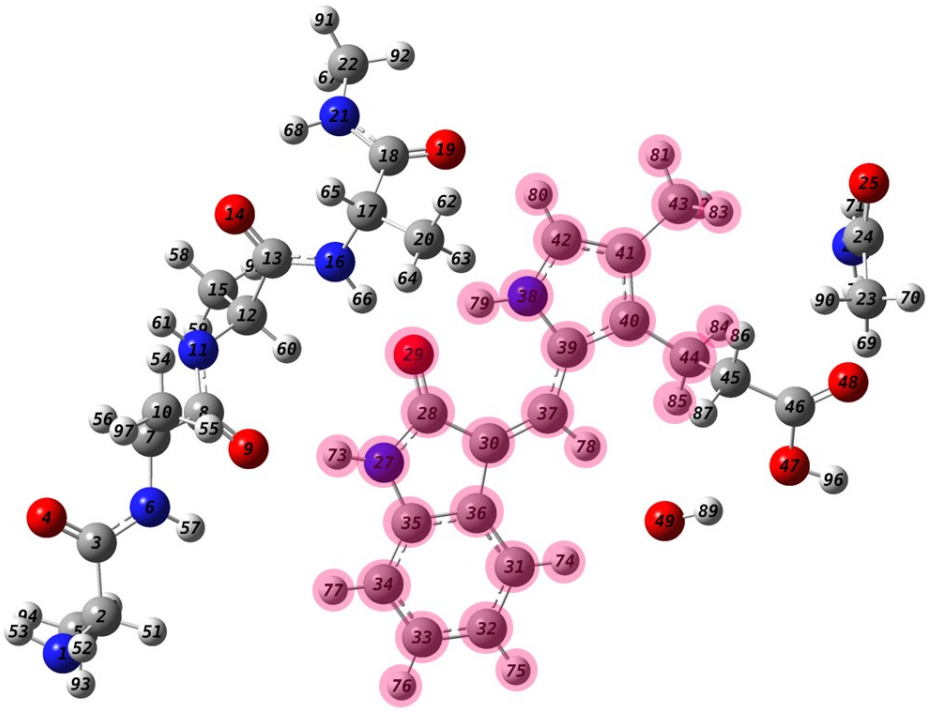

Figure 4-13 The structure of SU-5402 and FGFR1 binding site atoms used for calculating the EVV spectrum. The atoms showing motion in Mode 1 are highlighted in pink.

Mode 2:

As can be seen in Figure 4-14, the vibrational motion of Mode 2 is split between the SU-5402 molecule and the protein binding site. The SU-5402 molecule shows motion on the pyrrole group, the amide portion of the oxindole group, and the alkene carbon linking the two. The protein shows motion on the amide back bone groups formed of atoms labelled 14, 13, 16, and 66 and 9, 8, 11 and 61. The vibrational mode is likely conjugated between the SU-5402 molecule and the protein residues via hydrogen bonding interactions between the oxindole amide and the protein amide backbone. The contribution of largest amplitude to this mode is oxindole amide motion of the SU-5402 molecule. This mode forms the α mode of peak R.

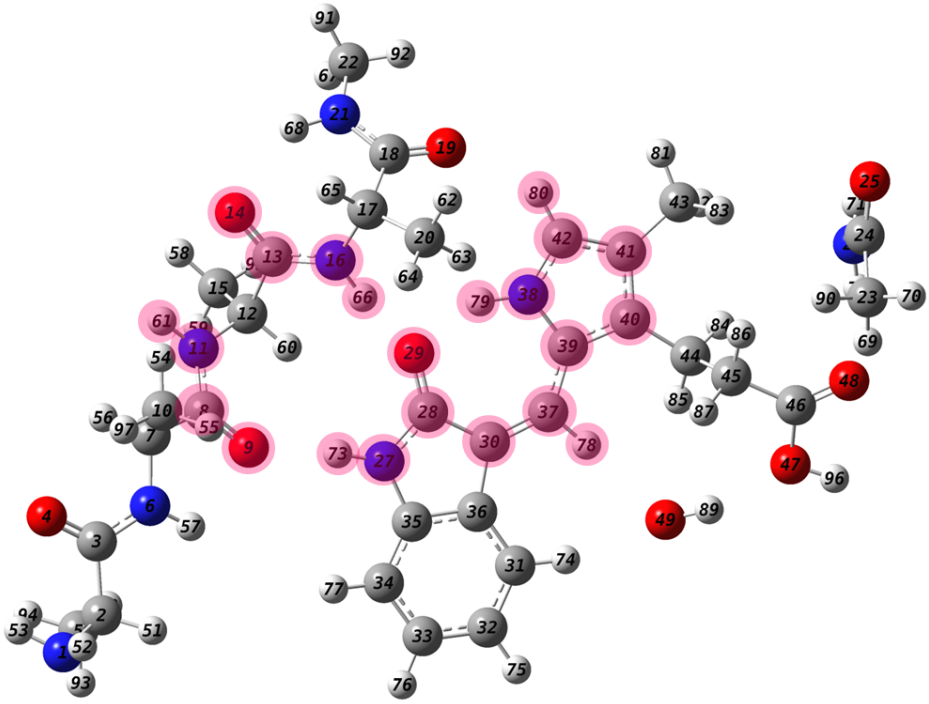

Figure 4-14 The structure of SU-5402 and FGFR1 binding site atoms used for calculating the EVV spectrum. The atoms showing motion in Mode 2 are highlighted in pink.

Mode 3:

As can be seen in Figure 4-15, the vibrational motion of Mode 3 is split between the SU-5402 molecule and the protein binding site. The SU-5402 molecule shows motion on the pyrrole group, the amide portion of the oxindole group, and the alkene carbon linking the two. The protein shows motion on the amide back bone groups formed of atoms labelled 14, 13, 16, and 66 and 9, 8, 11 and 61. The vibrational mode is likely conjugated between the SU-5402 molecule and the protein residues via hydrogen bonding interactions between the oxindole amide and the protein amide backbone. The contribution of largest amplitude to this mode is oxindole amide motion of the SU-5402 molecule. This mode forms the β mode of peak H.

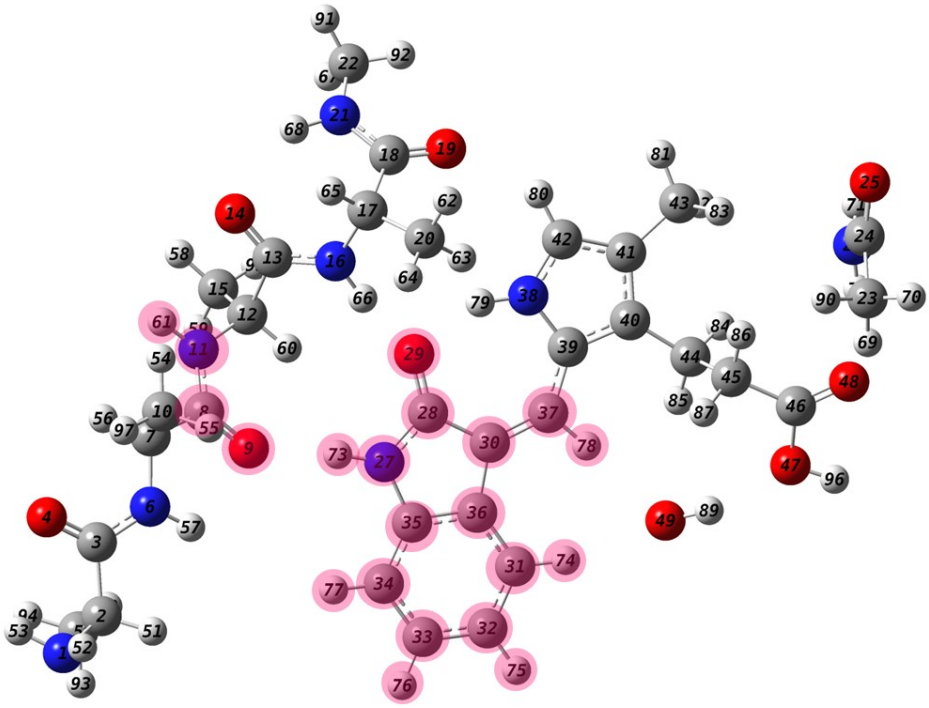

Figure 4-15 The structure of SU-5402 and FGFR1 binding site atoms used for calculating the EVV spectrum. The atoms showing motion in Mode 3 are highlighted in pink.

Mode 4:

As can be seen in Figure 4-16, the vibrational motion of Mode 4 is split between the SU-5402 molecule and the protein binding site. The SU-5402 molecule shows motion on the pyrrole group and its methyl, the amide containing five membered ring of the oxindole group, and the alkene carbon linking the two. The protein shows motion on the amide back bone groups formed of atoms labelled 14, 13, 16, and 66 and 18, 19, 21 and 68 and the methyl group of carbon 20. The vibrational mode is likely conjugated between the SU-5402 molecule and the protein residues via hydrogen bonding interactions between the oxindole amide and the protein amide backbone. The contribution of largest amplitude to this mode is oxindole amide motion of the SU-5402 molecule. This mode forms the β mode of peaks G and R.

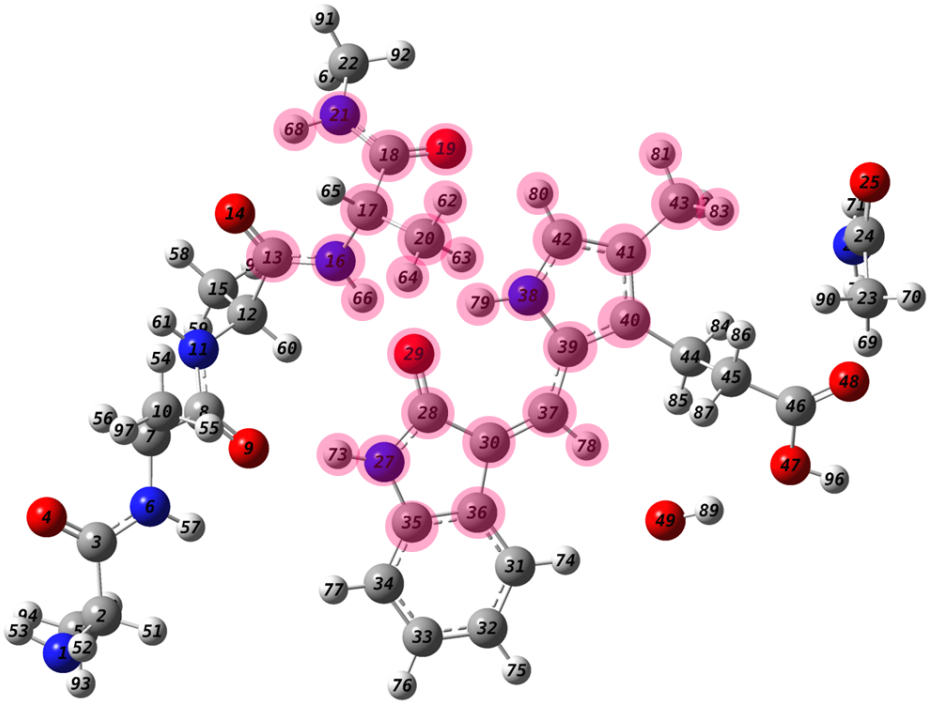

Figure 4-16 The structure of SU-5402 and FGFR1 binding site atoms used for calculating the EVV spectrum. The atoms showing motion in Mode 4 are highlighted in pink.

Mode 5:

As can be seen in Figure 4-17, the vibrational motion of Mode 5 is split between the SU-5402 molecule and the protein binding site. The SU-5402 molecule shows motion across all its atoms, except for those of the propionic acid group. The protein shows motion on parts of the amide back bone groups formed of atoms labelled 14, 13, 16, and 66 and 18, 19, 21 and 68 and the three methyl groups of carbons 15, 20 and 22. The vibrational mode is likely conjugated between the SU-5402 molecule and the protein residues via hydrogen bonding interactions between the oxindole amide and the protein amide backbone. The contribution of largest amplitude to this mode is oxindole amide motion of the SU-5402 molecule. This mode forms the β mode of peak F.

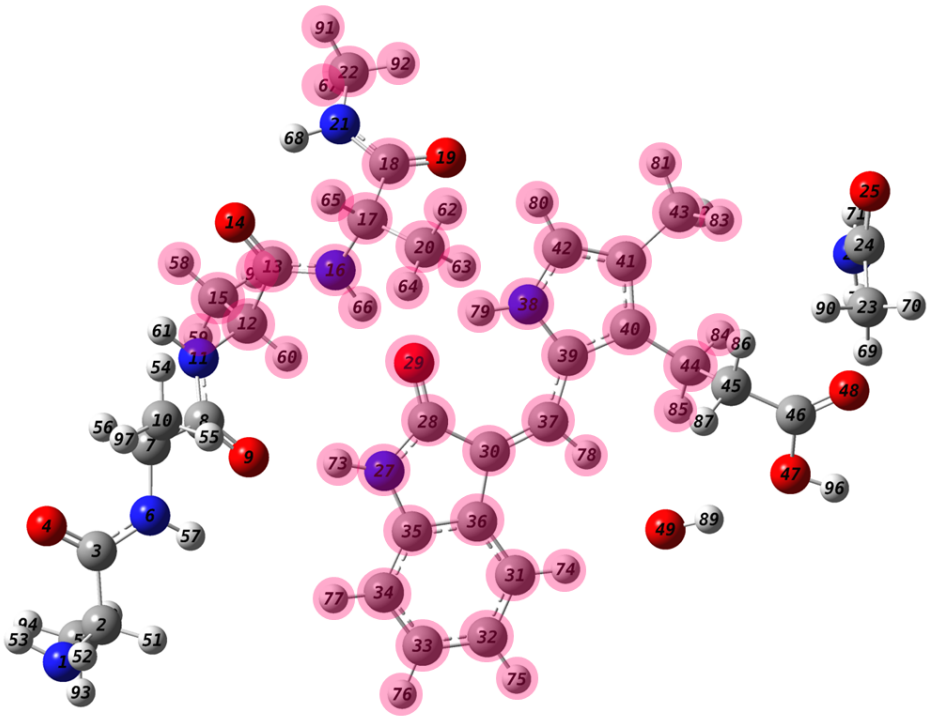

Figure 4-17 The structure of SU-5402 and FGFR1 binding site atoms used for calculating the EVV spectrum. The atoms showing motion in Mode 5 are highlighted in pink.

Mode 6:

As can be seen in Figure 4-18, the vibrational motion of Mode 6 is split between the SU-5402 molecule and the protein binding site. The SU-5402 molecule shows motion across all its atoms, except for those of the propionic acid group. The protein shows motion on the amide back bone groups formed of atoms labelled 5, 6, 7, 8 and 9 and the methyl group of carbon 2. Motion is also seen on the hydrogen atoms 66 and 60 The vibrational mode is likely conjugated between the SU-5402 molecule and the protein residues via hydrogen bonding interactions between the oxindole amide and the protein amide backbone. The contribution of largest amplitude to this mode is oxindole amide motion of the SU-5402 molecule. This mode forms the β mode of peak E.

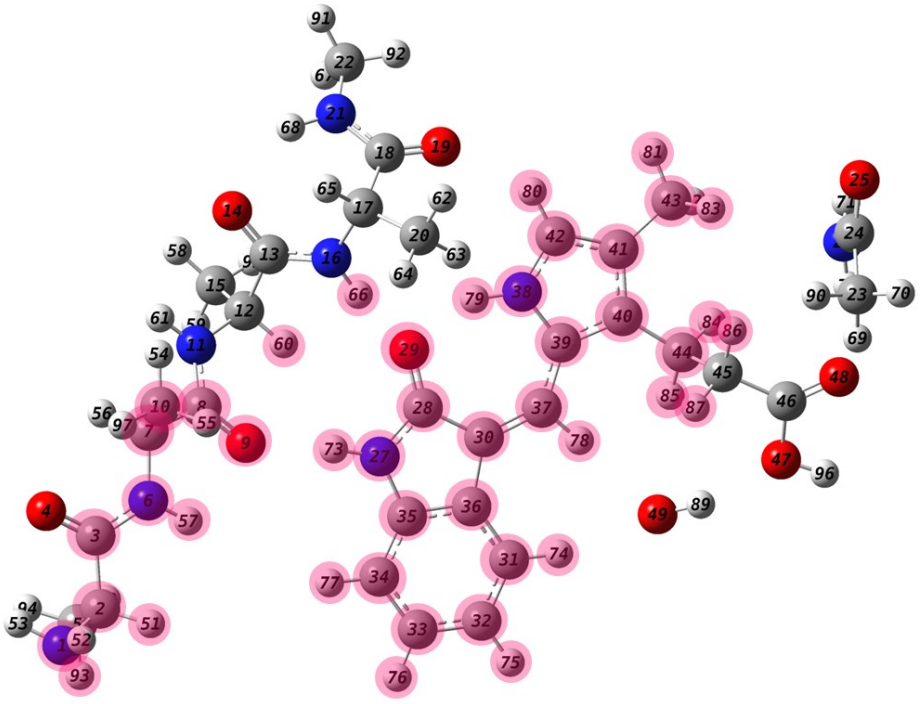

Figure 4-18 The structure of SU-5402 and FGFR1 binding site atoms used for calculating the EVV spectrum. The atoms showing motion in Mode 6 are highlighted in pink.

Mode 7:

As can be seen in Figure 4-19, the vibrational motion of Mode 7 is predominantly on the SU-5402 molecule with small motions on some protein residues. The SU-5402 molecule shows motion across all of its atoms, including the propionic acid group. The protein shows motion only on the amide back bone hydrogens labelled 60 and 61. The vibrational mode is likely conjugated between the SU-5402 molecule and the protein residues via hydrogen bonding interactions between the oxindole amide and the protein amide backbone. The contribution of largest amplitude to this mode is a ring deformation of the pyrrole group of the SU-5402 molecule. This mode forms the β mode of peak D.

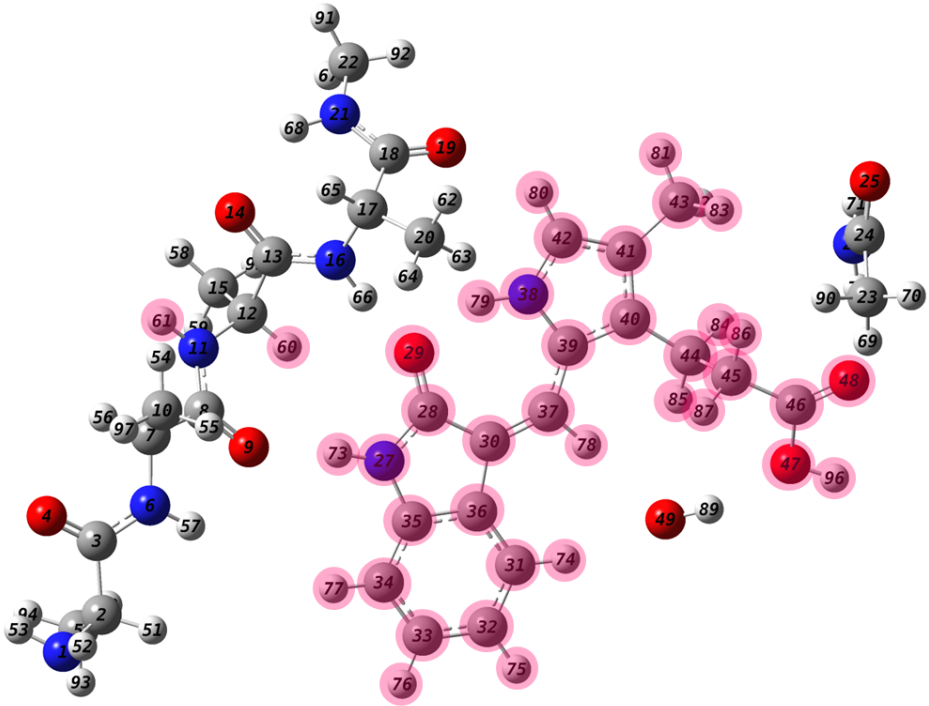

Figure 4-19 The structure of SU-5402 and FGFR1 binding site atoms used for calculating the EVV spectrum. The atoms showing motion in Mode 7 are highlighted in pink.

4.5 Determination of geometry of interaction

As discussed in section 2.3.2 it is possible in some cases to determine the angle between the two transition dipoles of two coupled vibrational modes, whose coupling gives rise to an EVV cross peak. This can be achieved exclusively for cases where the nature of the coupling between the two modes is predominantly electrical, in which case the coupling of the two modes can be approximated to dipole-dipole coupling. Using the dipole-dipole approximation the ratio of a given peak's intensity, when measured in the PPP and PPS polarisation schemes, can be related to the angle between the two transition dipole moments by the polarisability tensors and the dipole moments. There is currently no known equivalent process for through bond mechanically coupled modes, given the more complex physics governing the two.

The assignments made in section 4.4 suggest that the vibrational modes giving rise to the binding sensitive peaks of SU-5042 + FGFR1 are coupled mechanically. As a result of this, it would not be useful to apply the dipole-dipole treatment as it would yield incorrect angles. Although it is not possible to calculate the angle between the transition dipole moments for the binding sensitive peaks observed here, it is still possible to use them to determine if the geometry of interaction, between the SU-5402 molecule and its protein binding site is in agreement with the geometry used in the calculation.

Comparison of peak intensities between calculated and experimental EVV 2DIR spectra is not entirely reliable for a number of reasons. As discussed previously, experimental spectra are measured with delays between the three input pulses, in order to suppress non-resonant background signals. The coherences created have different lifetimes and decay over time at different rates during the delays between pulses. These lifetime effects are not accounted for in the calculation of spectra, therefore they cannot account for their effect on peak intensities.

Another reason for difficulty comparing peak intensities with calculation is the phase mismatch of the experiment changing as a function of frequency. The angles between the input beams required to give zero phase mismatch changes across the spectrum, not only due to the changing frequency of the beams, but due to the changing

refractive index of the material at these changing frequencies. Although steps have been taken to minimise it experimentally, the changing phase mismatch will result in peak intensities varying away from those calculated by differing amounts across the spectrum. For these reasons it is not possible, to compare the intensities of the binding peaks in the difference spectra to those from the calculation as a measure of the agreement between the experimental geometry and the geometry used in the calculation.

The ratio of a peak's intensity, when measured in the PPP polarisation scheme to that in the PPS polarisation scheme, is dependent on the relative geometry of the coupled vibrational modes. This ratio is also immune to all of the above difficulties of comparing experimental and calculated peak intensities – it will be constant for any set of inter-pulse delays or phase mismatch. For example, it is possible to achieve a set of angles between the beams that results in a phase mismatch corresponding to 50 % efficiency of signal production of a given peak. This 50 % loss of signal will be applied in both the PPP and PPS regimes, therefore unchanging the ratios of the two. As a result of this it is possible to record EVV 2DIR spectra of a sample in both polarisation schemes and compare the peak ratios to the peak ratios of calculated spectra. Doing so allows for direct comparison of experiment and calculation and can be used as a method of either verifying or invalidating the geometry used for the calculated spectra.

This section describes the work done to record EVV 2DIR spectra of the SU-5402 FGFR1 complex in both polarisation schemes allowing for the comparison of the two, with a view to geometric verification.

4.5.1 Polarisation ratio sample preparation and data collection

Gel spot samples of 1:1 molar ratio FGFR1:SU-5402 and of FGFR1 only were prepared as described above in section 3.3. The gel spots were produced from the following solutions

- 1 mM FGFR1, 1 mM SU-5402, 5 mM phosphate buffer, 5 mM NaCl, 2 %/ vol DMSO
- 1 mM FGFR1, 5 mM phosphate buffer, 5 mM NaCl, 2 %/ vol DMSO

In order to rotate the polarisation of the 790 nm ω_γ beam, a half wave plate mounted on a rotating actuator was installed in the ω_γ beam line. By measuring the transmission of the ω_γ beam through a polariser of known orientation, the actuator was calibrated such that positions corresponding to the half wave plate transmitting light with unchanged P polarisation, and for rotating its polarisation 90° to produce S polarised light were known. The pulse energy of the ω_γ beam remained unchanged upon polarisation rotation; however a small change in its timing relative to the two infrared pulses was observed which was accounted for before measurement. Detectable changes in beam pointing at the sample were not observed.

Three spectra were recorded of each of the two samples (FGFR1, with and without SU-5402) at both polarisations of ω_γ – a total of 12 measurements. The same spectral range and 5 cm^{-1} resolution was used as in the previously described FGFR1 SU-5402 experiments.

The data from previous experiments was intensity normalised after collection by dividing the spectrum by the intensity of the methylene peak at ω_α = 1470 cm^{-1} / ω_β = 2945 cm^{-1}. Doing this for the both PPP and the PPS data sets would of course destroy the PPS/ PPP ratio information that these experiments have been designed to determine. In order to prevent this, instead of measuring the six spectra of each of the two samples (3 x PPP, 3 x PPS) at six different positions across the gel spot, a PPP measurement was taken, followed by a PPS measurement at exactly the same position before moving to a new position for the next pair. Doing so allowed for the PPS/ PPP ratio for the methylene peak to be calculated for each of the pairs, and the average used to normalise the PPP and PPS spectra. The average methylene PPS/ PPP ratio was found to be 0.2477 ± 0.0191 (7.7 %). In this way, the PPP spectra were normalised such that their methylene peak intensities were 1, whilst the PPS methylene peak intensities were 0.2477, allowing them to be averaged and compared.

4.5.2 PPS/ PPP binding peak ratios

The three PPP and PPS spectra for both FGFR1 and FGFR1 + SU-5402 samples were processed as previously described, normalised on the methylene peak whilst taking account of the PPS/ PPP methylene average ratio and averaged. The PPP measurements of the same FGFR1 SU-5402 system in the same spectral range, described in section 4.1.2 were not included in this analysis to prevent any experimental difference in the alignment of the laser system leading to errors in the comparisons. The mean PPP FGFR1 only spectrum was subtracted from the mean PPP FGFR1 + SU-5402 spectrum, and the mean PPS FGFR1 only spectrum was subtracted from the mean PPS FGFR1 + SU-5402 spectrum, using the method described and used previously. The two resultant average difference spectra were then fitted using the same fitting procedure as previously described. The PPS/ PPP ratios of the fitted peak intensity of the six binding sensitive peaks can be seen below in Table 4.2 along with the PPS/ PPP ratios of their assigned peaks in the calculated spectra and the percentage difference between the two.

Peak	Experimental PPS/ PPP	Calculated PPS/ PPP	Difference between experiment and calculated ratios as a % of calculated
H	2.68358	0.208351	1188.012
G	0.385266	0.479252	19.61093
F	0.231835	0.634279	63.44905
E	0.592348	0.345739	71.32816
D	0.77034	0.264696	191.028
R	0.379955	0.765942	50.39377

Table 4.2 The ratio of fitted peak intensities of the mean PPS and PPP difference spectra compared to the ratios displayed in the calculated spectra.

As can be seen in Table 4.2, the experimental and calculated PPS/ PPP ratios, using the mean experimental difference spectra are not showing agreement. Before looking at possible spectroscopic reasons for this disagreement, it was first necessary to establish whether the variance of the data used in these calculations were sufficiently high to account for the disagreement observed.

To determine confidence of the PPP to PPS ratios of the binding peaks in the difference spectra, individual difference spectra were produced by subtracting the mean FGFR1

only spectra from each of the three individual FGFR1 + SU-5402 spectra for both PPP and PPS polarisations. The binding peaks of these individual difference spectra were fitted, as before, to determine the peak intensities in the individual spectra.

The fractional standard deviations of the intensities of these peaks must be added in quadrature to give the fractional standard deviation of their ratio[94], which are shown in Table 4.3:

Peak	**PPP SD %**	**PPS SD %**	**Propagated SD of ratio %**	**Difference between experiment and calculated ratios as a % of calculated**
H	12.06897	1.114689	12.12034	1188.012
G	16.884	1.610502	16.96064	19.61093
F	11.4094	8.538197	14.25045	63.44905
E	13.65915	5.802658	14.84059	71.32816
D	18.16881	34.55475	39.04019	191.028
R	0.173611	3.683036	3.687126	50.39377

Table 4.3 The percentage standard deviation of fitted peak intensities of the individual PPS and PPP difference spectra compared to the percentage difference between the experimental and calculated ratios.

The propagated standard deviations in the Table 4.3 show that even when accounting for the standard deviation of the individual measurements, the experimental PPS/ PPP ratios do not agree with those from the calculated spectra.

The above analysis has not accounted for the fact that each of the individual difference spectra was produced by subtracting an average of 3 measurements (mean FGFR1 only spectrum) from an individual measurement (individual FGFR1 + SU-5402 spectrum). The standard deviation of the individual FGFR1 only spectra, which compose the mean, must also be accounted for. To do so, the average standard deviations of 25 x 25 cm^{-1} area at each of the peak locations in the protein only spectra were calculated. These standard deviations should be added to those of the individual difference spectra peak intensities[94], then the relative standard deviations propagated to give those of the ratio, as above. These numbers are seen in that second Table 4.4:

Peak	PPP SD%	PPS SD%	Propagated for ratio %	Difference between experiment and calculated ratios as a % of calculated
H	94.23155	17.3783	95.82061	1082.461
G	32.57553	18.15602	37.29351	19.61093
F	36.48236	186.264	189.8031	63.44905
E	43.8052	23.58715	49.75187	71.32816
D	91.04543	72.20198	116.1998	191.028
R	52.60938	49.80875	72.44762	50.39377

Table 4.4 The percentage standard deviation of fitted peak intensities of the individual PPS and PPP difference spectra after accounting for the variation in protein only spectra, compared to the percentage difference between the experimental and calculated ratios. The rows corresponding to peaks whose propagated standard deviation of ratio is greater than the difference between the calculated and experimental ratio are highlighted in green.

As shown in Table 4.4, for three of the six binding sensitive peaks the experimental ratios of PPS/ PPP intensity agree with those of the calculated spectra, within the variance of the measurements used. If all six of the binding sensitive peaks showed such agreement it would imply that it may be possible to achieve experimental

agreement with the calculated ratios simply by repeating the same set of experiments many times over, to improve the statistics of the peak ratios. It would not even be necessary to repeat the entire measurements; simply the spectral regions in which the binding sensitive peaks occur, as their frequency locations are now known.

4.6 Conclusions and discussion

Six binding sensitive peaks have been identified and assigned to seven vibrational modes. All of the vibrational modes exhibited some atomic motion on the inhibitor molecule, leading to a more complicated coupling picture than was originally imagined of inhibitor only vibrations being through space coupled to protein only vibrations. That is not to say such vibrational couplings are not present in other frequency ranges.

It was initially a little surprising that only six binding specific peaks were observed when comparing the difference spectra of the bound and unbound inhibitor. This can partially be explained by considering the structure of SU-5402 and how it has a large conjugated system across the majority of the molecule, of alternating pi and sigma bonds. Such a structure prevents any real degree of geometric flexibility, with only the propionic acid side chain moiety able to rotate, relative to the rest of the molecule. The result of this is that to a large extent the geometry of the SU-5402 molecules is the same in solution or in the gel phase with a non-binding protein such as BSA, as it is when it has bound to SU-5402. This will cause many of the intramolecular vibrational couplings, within the SU-5402 molecule to remain unchanged upon complex formation, and the resultant EVV 2DIR cross peaks unchanged too. This also makes it unlikely that there will be binding specific peaks at other frequencies that are due to binding induced changes in SU-5402 geometry, however this may not be the case if the same methodology was applied to more flexible inhibitors.

Peaks H, E and D display different PPS/ PPP ratios from the calculated spectra and there are three possible implications of this:

a) The geometry used in the calculations, taken from the crystal structure is not correct.
b) There was a systematic error in the methodology used to produce the calculated spectra.
c) There was a systematic error in the collection or processing of the EVV 2DIR spectra.

To suggest that the published crystal structure of the FGFR1 SU-5402 complex is incorrect on the basis of the results presented here would be quite a bold statement. EVV 2DIR has never previously been shown to be able to determine geometries of interaction of such large molecules whereas crystallography has been established in this role for decades. That is certainly not to say that EVV 2DIR does not have the capacity to elucidate such information if performed appropriately, potentially succeeding in aspects that crystallography is incapable of, such as disordered protein domains or interactions with non-crystallographic water molecules. The following discussion will work on the assumption that the crystal structure geometry is correct.

The assumption will also be made that the methodology used to produce the calculated spectra was correct. The same approach has previously been shown to yield results consistent with experiments, albeit for smaller molecular systems. Furthermore as the exact details of the methodologies used are beyond the scope of this thesis so a rigorous analysis of potential sources of error would not be possible.

The most likely source of error in the EVV 2DIR methodology described here would be potential optical thickness of the samples in the infrared regions. At frequencies at which the samples are optically thick, all of the light incident on the sample will be absorbed in the initial depth of the sample, leaving no light to produce signals from subsequent depths. The effect of this is that peak intensities become distorted relative to one another; peaks in frequencies of high optical thickness will appear smaller relative to peaks at frequencies of lower optical thickness than they would in an optically thin sample. At first consideration, this may be thought not to affect the PPS/

PPP peak ratios, as optical thickness will affect both polarisation schemes equally; however this is not the case.

Optical thickness is not binary, with some spectral regions being more optically thick than others and all regions becoming more optically thick with increasing sample depth. In samples approaching optical thickness, small variations in sample depth will therefore result in different relative peak intensities, as will the addition of new molecules to the sample such as SU-5402. This variability of peak ratios as a function of sample thickness and composition can lead to imperfect subtractions in the production of the difference spectra. If the FGFR1 only spectrum was over or under subtracted from the FGFR1 + SU-5402 spectrum, the PPS/ PPP peak ratios would be incorrect, even if the over or under subtraction was consistent between polarisation regimes.

In order to account for this it should first be established whether or not the current methodology yields gel spots which are optically thick. That could be achieved by making sample gel spots, as previously described, on calcium fluoride windows or other infrared transparent material then measuring the infrared absorption spectra. This perhaps may be most conveniently achieved, given the small area of the gel spots, by placing an energy sensor behind the sample stage of the EVV 2DIR system and using the OPA outputs as a linear spectrometer. In the event that optically thick samples are observed, the same methodology of sample preparation can be used, however with lower concentration of the starting protein solution. This will of course sacrifice signal intensities but the issue associated with this can be resolved through increased signal averaging.

The spectral region used for this work was chosen simply as it was the spectral region which had previously been used within the group. This yielded the benefits of some known protein spectral assignments and established experimental protocol. The region had initially been selected by the group as it was predicted to contain the highest density of cross peaks, with the aim of yielding high information density, which was observed in the work described here. The calculated spectra of the SU-5402 FGFR1 complex show predicted peak locations across a much larger spectral range than that used here. The calculated spectrum over the larger spectral range can

be seen in Figure 4-20 alongside the calculated spectrum of all amino acid side chain groups (amino acid side chains groups calculated in vacuo).

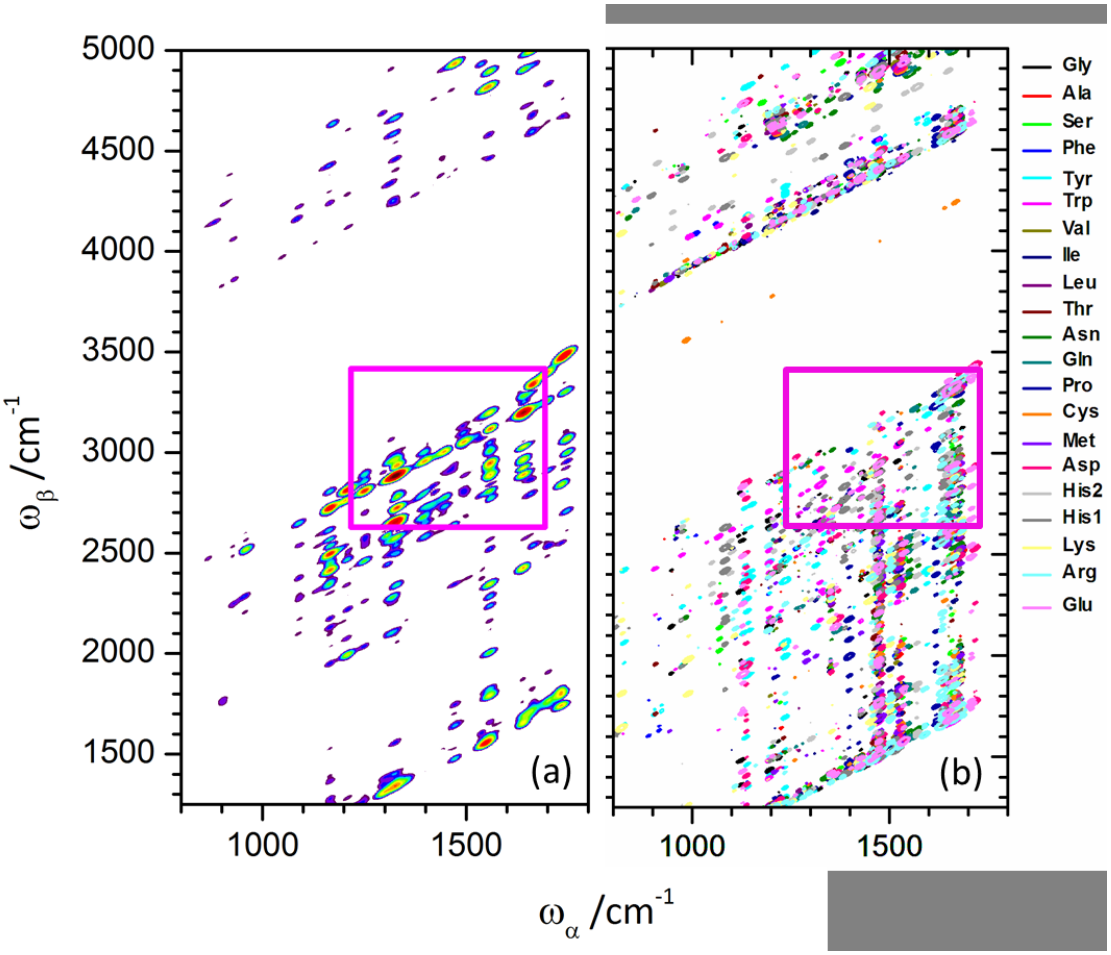

Figure 4-20 Calculated EVV 2DIR spectra over a wide spectral range for (a) the partial SU-5402 FGFR1 binding domain complex (b) all amino acid side chains, colour coded with key. The spectral range used for the experiments described here is shown in the pink rectangle.

Although it is not possible to extend the size of the spectral range of the EVV 2DIR system, due to the technical limitation of the OPAs used, it is possible to move the spectral range of the experiment to other spectral regions. Figure 4-20 shows much less congested spectral regions in the amino acid calculated spectrum at ω_β frequencies above 3700 cm^{-1}, which also contain SU-5402 FGFR1 complex peaks.

Another spectral range calculated to show reduced spectral congestion for the amino acid side chains, at higher ω_α frequencies can be seen in Figure 4-21

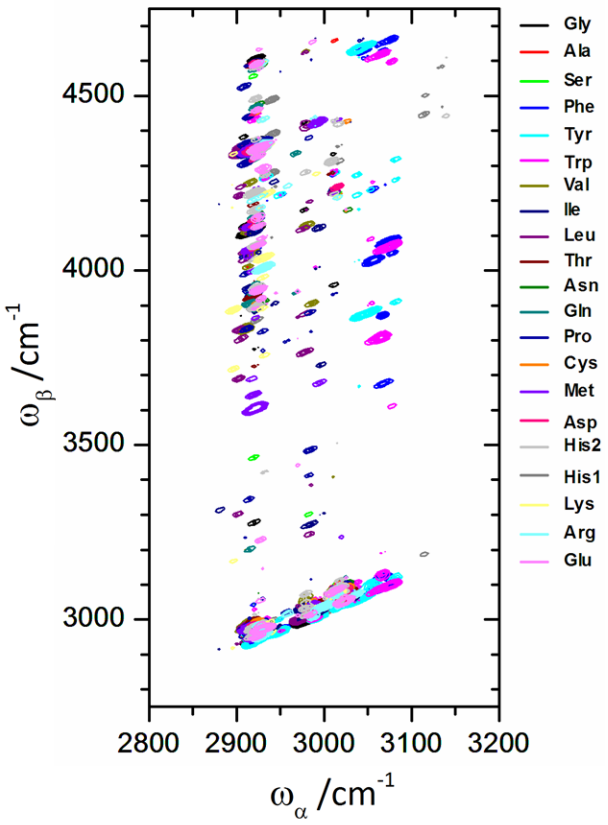

Figure 4-21 High ω_α calculated EVV 2DIR spectrum of all amino acid side chains, colour coded with key.

The higher ω_α spectra region again shows less congestion than the region used for the experiments described here. The equivalent spectral region has not been calculated for the SU-5402 FGFR1complex, but could easily be done so.

Repeating the experiments described here in one of this higher frequency spectral regions may yield fully resolved binding sensitive cross peaks, or at least cross peaks less convolved by protein spectra. Given the lack of congestion, it may be viable to fit both FGFR1 and SU-5402 peak in the same spectrum. This would allow for PPS/ PPP peak ratios to be determined without the errors associated with producing the difference spectra. This may result in spectra displaying agreement with the calculated spectra, validating EVV 2DIR's capacity to experimentally validate models of complex geometries.

The calculated spectra of the SU-5402 FGFR1 complex can also be used to guide the spectral region for further experiments. All of the binding sensitive peaks identified

here were assigned to complex vibrations spanning many atoms across both the protein and inhibitor. The calculated spectra predict more confined vibrational modes at higher frequencies, for example the two calculated modes depicted in Figure 4-22. The mode shown in pink, at 3250 cm^{-1} and the mode shown in green at 3313 cm^{-1} are confined to just a small number of atoms. In the event that these two modes give rise to an EVV cross peak, it would be entirely due to electrical coupling. Such a peak would allow for the validation of using the dipole-dipole approximation for determining angles in inhibitor binding geometries.

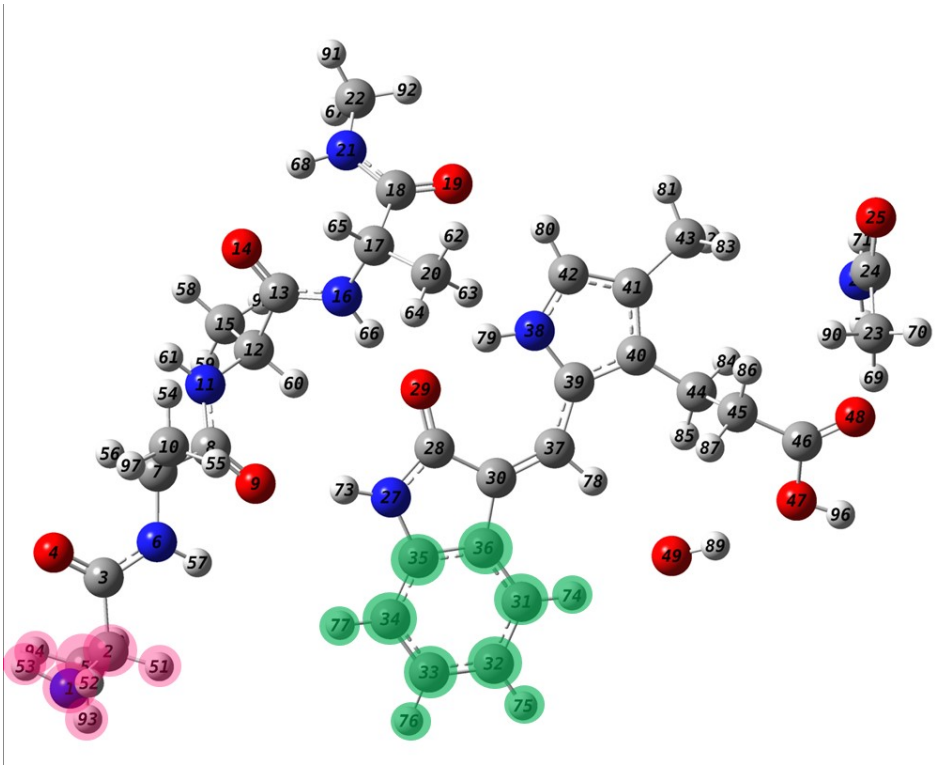

Figure 4-22 Two higher frequency modes (green mode at 3313 cm^{-1} and pink mode at 3250 cm^{-1}) predicated by the calculations, which show atomic motion confined to smaller parts of the system than those assigned to the binding dependent peaks observed.

Should the improvements to the methodology suggested above yield experimental PPS/ PPP ratios consistent with those in the calculated spectra, this would demonstrate EVV 2DIR's ability to verify the geometry used in such calculation to be

the same as the geometry of the actual SU-5402 FGFR1 complex. Not only this, but it would pave the way for EVV 2DIR to be used to determine the geometry of inhibitor protein complexes for which crystal structures are not already known, if the crystal structure of an apoprotein binding site has already been determined.

An initial informed guess of the interaction geometry between the known binding site and a novel inhibitor can be produced, based on a computationally cheap docking model for example. The EVV 2DIR spectrum of this initial geometry can then be calculated. The geometry of the computational model can then be iteratively improved such that its resultant calculated spectra become more similar to the experimental spectra. Once a geometry has been found whose calculated spectra are consistent with the experimental spectra, it can be said the actual geometry has been determined.

Using the exact same methods as were used to produce the calculated spectra shown here would be incredibly computationally expensive, with a full calculation being required for each new geometry iteration. Work is currently being done by collaborators to reduce the computational effort required to implement this progress, including improvements such as not recalculating geometry insensitive molecular contributions. If such a process is successfully implemented, the impact on the drug discovery process could be profound. Many of the limitations of crystallography could be overcome, such as the requirement of large amounts of protein and the growth of crystals. The requirement for an apo protein structure does of course mean that this method would not be suitable for studying protein inhibitor interactions for proteins who's structures are unknown.

5 Detecting intermolecular interaction in DNA structures using EVV 2DIR

Chapter four described how EVV 2DIR was used to detect and characterise intermolecular interactions in a protein-inhibitor complex. This was possible to do in a congested spectral region, without the need for extensive signal averaging, due to resonance enhancement of the colored inhibitor resulting in approximately 100 fold increase in the intensity of its spectral features, relative to the unenhanced protein. In this way, peaks from a hand full of interactions were visible over a background of many thousands of interactions. In contrast to a protein-inhibitor system, where each inhibitor-protein interaction is only seen once per bimolecular complex, intermolecular interactions between hybridised strands of DNA are conserved and repeated numerous times. For example two hybridised alternating guanine - cytosine 10mers will contain the same set of guanine-cytosine interactions 10 times, yielding EVV 2DIR signals arising from any guanine-cytosine vibrational couplings 100 times more intense than of a single pair. These repeating interactions should therefore be

readily detectable by EVV 2DIR, making DNA structures an ideal system to further develop EVV 2DIR as a tool for studying biomolecular interactions.

This chapter presents the first EVV 2DIR spectra that have been collected of DNA, and the preliminary investigation into EVV 2DIR's sensitivity to different DNA structural confirmations, namely B-form duplexs and G-quadruplexes. Upon formation of distinct 3-dimensional structures, new vibrational couplings will occur as a result of the new stabilising interactions. This gives the potential for the identification of new cross peaks in the EVV 2DIR spectra, as well as other spectral changed which may be indicative of secondary structure.

The work presented here forms only the very beginning into the investigation of EVV 2DIRs sensitivity as a structural probe for DNA. The experiments described were conducted only a couple of months before the writing of this thesis but should prove instructive for designing a set of future experiments to further the investigation of EVV 2DIRs suitability as a structural probe of DNA secondary structures.

The G-quadruplexes studied are those formed by the Myc2345 sequence in the presence of potassium ions, described in Chapter 1, and the B-form duplexes are those formed by the Myc2345 sequence and its complementary sequence, in the absence of potassium ions.

5.1 Justification of controls

When looking for new spectral features arising from molecular interactions, it can be useful to have a non-specifically interacting system to use as a comparative control, such as the non-binding BSA protein in Chapter 3. Other research groups studying DNA structures using multidimensional infrared spectroscopies have achieved such non-specifically interacting DNA control samples through the use of temperature controlled sample cells to melt the DNA structures at elevated temperatures[43]. This approach is perfectly suitable for studying liquid samples, however would not be possible for the gel phase samples used here for EVV 2DIR studies. Furthermore the

G-quadruplex structures formed from the Myc2345 sequence, as used here, are sufficiently stable that they do not melt below the boiling point of water.

The formation of both G-quadruplex structures and B-form duplex structures are DNA sequence dependent[95]: In the case of the G-quadruplex, there must be sufficient and appropriately spaced guanine bases to form stable tetrads capable of stacking, whilst duplexes are only formed when two sequences have sufficiently complementarity (or one sequence sufficiently self-complementarity in the case of homo-hybridisation) to form enough Watson-Crick base pairings to achieve stability. Therefore by altering the sequence of a given sample of DNA, whilst keeping the same total base content constant, (e.g. GCAT -> CTGA) it is possible to achieve control samples which do not form a specific structure. Care must be taken not to accidentally produce sequences with partial self-complementarity that can form homo-dimers

The sequence altered samples will contain exactly the same number of each nucleotide bases and therefore their spectral features arising from coupled modes within a given base, and those from the phosphate back bone groups should be identical. New intermolecular interactions associated with the formation of structures should therefore stand out as differences when comparing the two

Unfortunately, it is not only the intermolecular interactions that will differ between the samples and their corresponding controls. Any spectral contributions from interactions arising from base stacking, along the individual strands, will also be effected by sequence. This however does not undermine this approach, given the more dynamic behavior which is displayed by single stranded DNA which is not forming specific secondary structures[96]. These disordered single strands, which can be modeled as and described as flexible "worm-like" chains[97], adopt a variety of conformations in solution, which will no doubt be accompanied by a variety of base stacking and other intramolecular interactions. This distribution of conformations will presumably lead to a distribution of vibrational couplings and in turn, a distribution of low intensity EVV 2DIR features, which may be negligible in the experimental spectra. The specific interactions arising from the conformation of the secondary structure on the other hand, should give more intense specific features (all other things being equal) given the considerably reduced distribution of

confirmations. It is also not known how strongly these coupling will present themselves in EVV 2DIR spectra.

5.2 EVV 2DIR studies of B-Form DNA Duplexes

5.2.1 Sample preparation and data collection of DNA duplex

To compare the EVV 2DIR spectra of a B-form duplex formed from Myc2345 and its complementary strand and a control of two non-complementary strands, SS1 and SS2 (sequences provided in the appendix), two gel spot samples were prepared, as per the methodology described in section 3.3 for producing gel spots for EVV 2DIR. Two solutions were prepared:

- 1mM Myc2345, 1mM Myc2345C pH7
- 1mM SS1, 1mM SS2 pH7

These two solutions were annealed at 90° for 3 minutes and allowed to cool before depositing onto coverslips for gel formation. Drying 1ul spots of these solutions slowly at 85.2% relative humidity in the humidity controlled sample cell described in section 3.3 was found to produce flat topped gel spots, which remained stable in the beams over the course of an EVV 2DIR experiment.

Data was collected between the ranges ω_α=1235 – 1735cm^{-1} and ω_β= 2605 – 3455cm^{-1} at 5cm^{-1} resolution averaging 100 samples per point. All three beams were set to propagate with P polarisation and Time delays of T_1=1.75ps and T_2=1.0ps were used to give a good compromise between minimising non-resonant background and maximising resonant signal size. The spectrum of each sample was recorded three times, alternating between the two samples to prevent any minor drift in the laser system over the course of the data collection impacting the two samples differently. Each of the three repeats were recorded at slightly different position son the gel spot to prevent any anomalous spot dominating the data collection.

The data was processed in MATLAB, as described in section 3.2. The individual repeats of each measurement were intensity normalised on the largest spectral feature, observed at ω_α=1535cm^{-1} / ω_β=3010 cm^{-1} such that the repeats could be averaged. This served well to allow the individual repeats to be averaged, but does not necessarily allow for direct comparison of absolute peak intensities between the two data sets. For this to be possible, the two datasets must be normalised on a peak which is internally consistent between the two, as was done using the methylene peak for the protein studies of Chapter 4.

5.2.2 Spectra and discussion of Duplex Vs Single strand DNA

The mean processed EVV 2DIR spectra of the single stranded and duplex samples can be seen in Figure 5-1

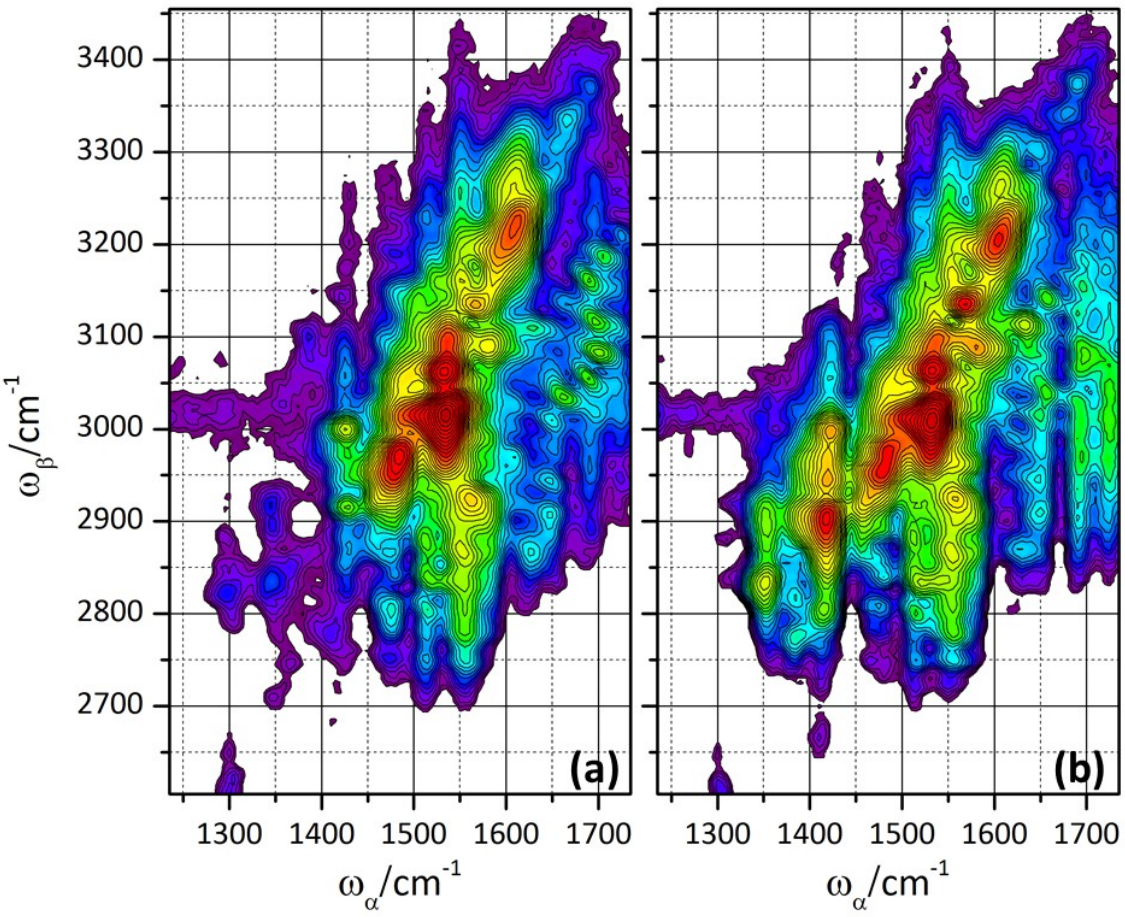

Figure 5-1 (a) The mean processed EVV 2DIR spectrum of non-complementary strands SS1 and SS2. (b) The mean processed EVV 2DIR spectrum of the duplex formed by complementary strands Myc2345 and Myc2345C

As can be seen in Figure 5-1 the EVV 2DIR spectra of DNA sequences are extremely complex. There are well over 100 peaks in each spectrum, with intensity above the 3σ level. On one had this is encouraging as it suggests a large information density, with a large number of vibrational couplings observed, however it makes the process of assigning and interpreting these first spectra a challenging one. At the time of writing, work is being done by collaborators to produced calculated EVV 2DIR spectra of a variety of model DNA systems. Unfortunately these calculations are due to arrive shortly after the submission this thesis so it will not be possible to use them to help assign and interpret this work. Preliminary interpretations and assignments have been provided here, based upon frequency assignments made to similar systems in the literature. Interpretation has also been made possible by comparing the spectrum of the single strands used for comparison with the duplex, with the spectrum of the

single stand used for comparison with the quadruplex forming sequence Myc2345, on the basis that the sequences used for comparison with Myc2345 does not contain cytosine.

As the spectra are so complex it was not possible to make assignments and interpretations across the entire spectrum, especially given the recency of the data. The discussion presented here represents just the preliminary analysis of these spectra. Initial interpretation was confined mainly to the range ω_α=1550-1735cm^{-1} / ω_β=2900- 3450 cm^{-1}. This spectral region contains many nucleotide base vibrations, according to assignments in the literature[98]. Figure 5-2 shows the spectra in this region. Upon first inspection most of the peaks are visible in both spectra but in some cases at different relative intensities. The extremely high density of peaks makes it difficult to distinguish between a single peak and several closely spaced peaks. Comparison with calculation should help make these matters clearer.

The peaks labelled A in Figure 5-2 have been assigned to the cytosine carbonyl stretch coupled to its overtone. The fundamental frequency of 1660cm^{-1} in the single stranded spectrum, and slightly down shifted frequency of 1650cm^{-1} in the duplex, is in agreement with the frequencies reported elsewhere[99] for this mode. Also used for this assignment was the relationship between the peaks marked A and the peaks marked B, which fall on common values of $\omega_{\alpha,}$ indicating they arise from another coupling of the same mode. Peaks B also down shifting again upon duplexation. These B peaks are noticeably absent in the spectra of Myc234 and its corresponding control (Figure 5-3), which do not contain cytosine, indicating they are present due to a coupling of a cytosine vibrational mode in ω_α=1660/1650cm^{-1}. By considering the difference between the ω_α and ω_β frequencies, it is possible to calculate the frequency of the fundamental mode which forms the other half of the combination mode probed at ω_β with the mode at ω_α. By doing so, it can be seen that the B peaks are due to a coupling between the modes at ω_α=1660/1650cm (assigned to cytosine carbonyl stretch) and a mode at around 1490cm^{-1}. In plane ring vibrations of cytosine have been previously assigned to vibrations at this frequency[99]. This coupling seems perfectly viable given that the two parts of the base concerned are directly bonded to one another. The overtone peak from this vibrational mode can be seen at ω_α=1490cm^{-1} / ω_β=2975 cm^{-1} in Figure 5-1

Peaks C and D have been assigned to the overtone of single stranded guanine carbonyl stretch and the duplex hybridized guanine carbonyl stretch respectively, with fundamental frequencies at 1675cm^{-1} and 1690cm^{-1}. These frequencies, and the shift observed upon duplex formation, is in agreement with assignments reported elsewhere[45,100]. Both duplex and single stranded peaks are present in both duplex and single stranded spectra. The implication of this in the case of the "single strand" control is that some degree of homo or hetero-dimerization was occurring. In the case of the duplex sample perhaps an excess of one of the two strands was present as a result of sample preparation errors. It is worth noting that linear infrared spectroscopy reports a decrease in peak intensity of the guanine carbonyl peak accompanying the upshift upon duplexation. If these relative peak intensities carry over to the EVV 2DIR spectra, a small amount of single strand in the duplex sample could potentially lead to a signal comparable in intensity to that of the majority duplex. It is difficult to talk about absolute peak intensity when comparing these two measurements as no peak is known to be consistent between the two which could be used for normalisation. Relatively though, the C/D peak ratio in the single stranded spectra is 1.7, whilst it is 0.7 in the duplex spectra. This difference in ratios is consistent with there being a higher proportion of duplex structures in the sample formed of complementary strands.

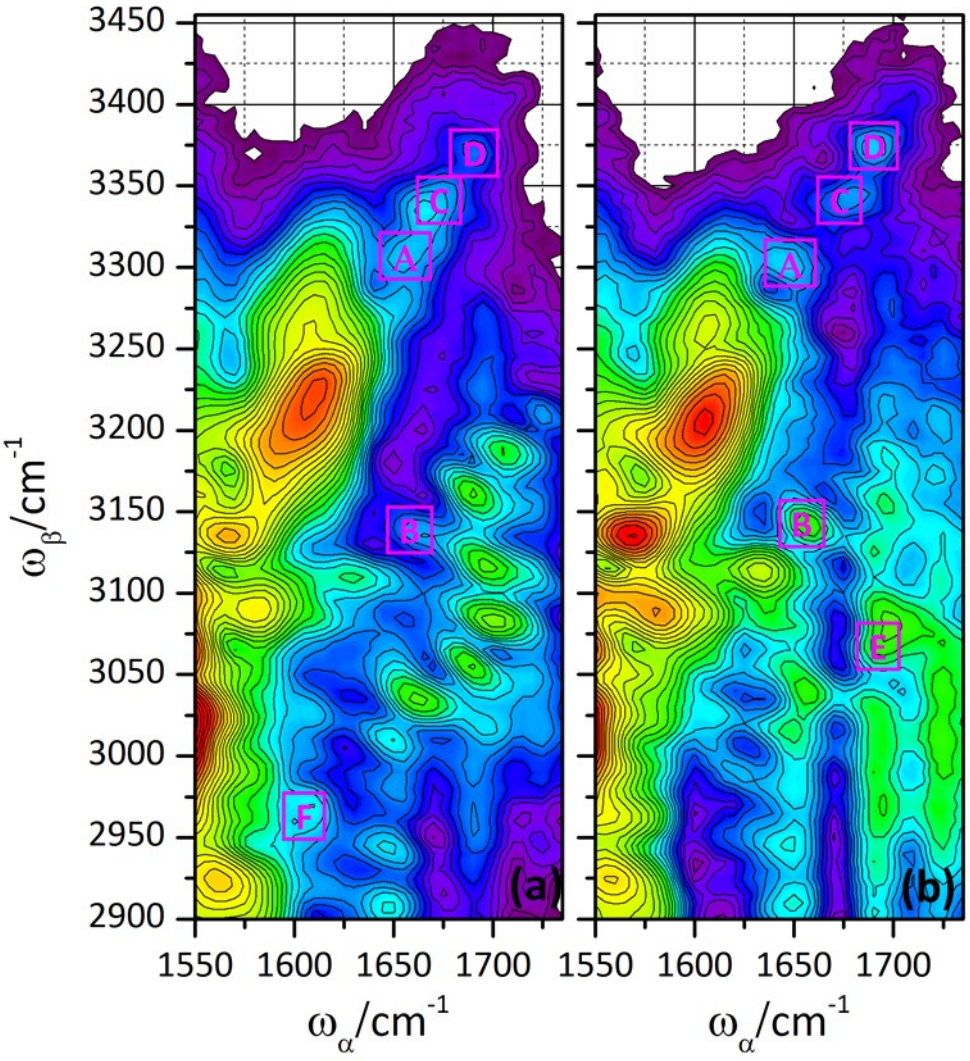

Figure 5-2 The base vibration regions of (a) The mean processed EVV 2DIR spectrum of non-complementary strands SS1 and SS2. (b) The mean processed EVV 2DIR spectrum of the duplex formed by complementary strands Myc2345 and Myc2345C

A clear addition to the duplex spectrum, which is absent or very weak in the single strand spectrum is peak E at ω_α=1690cm^{-1} / ω_β=3070 cm^{-1} which fills in the region of low intensity between the two peaks at ω_α=1685cm^{-1} / ω_β=3050 cm^{-1} and ω_α=1700cm^{-1} / ω_β=3085 cm^{-1}. Falling on the same ω_α frequency as peak D and being present only in the duplex spectrum, it makes sense to assign the mode in ω_α to be the

duplex hybridized guanine carbonyl stretch. The other fundamental mode must be at 1380 cm^{-1}. Vibrations at this frequency have previously been assigned to guanine in the anti-conformation, like which is seen in B-form duplex DNA[101]. The overtone peak of this anti-conformation mode can be seen in the duplex spectrum in Figure 5-1 at ω_α=1680cm^{-1} / ω_β=2760 cm^{-1} and is not noticeably present in the single strand spectrum

Only one peak has been identified to be present exclusively in the single stranded sample and has been marked at ω_α=1605cm^{-1} / ω_β=2960 cm^{-1} as peak F in Figure 5-2. No sensible assignment could be made for the fundamental mode at 1605cm^{-1}. The other component of the combination mode to which it is coupled is at 1355cm^{-1} and has previously been assigned as guanine in the syn-conformation[102]. It has been shown using NMR that a population of non-hybridised guanine will be 80% in the syn-conformation, in solution[103]. Assuming this assignment to be correct, this peaks absence in the duplex spectra is in agreement with the fact that B-form duplex DNA displays purines in the anti-conformation. The overtone peak of this syn-conformation mode can be seen if Figure 5-1 in the single strand spectrum and is again absent in the duplex spectrum.

The overtone of the unassigned fundamental mode at 1605cm^{-1} can be seen in Figure 5-1 at ω_α=1605cm^{-1} / ω_β=2960 cm^{-1} as a broad intense feature in both the single strand and duplex spectra. This is just one of many intense features in the spectrum which it has not been possible to assign using frequency assignments in the literature. The high density of peaks in the EVV spectra and also the high density of assigned frequencies in the literature make interpreting the spectra in this way both unreliable and incredibly slow. Furthermore by only using frequencies reported to be observed by other techniques, it limits the interpretation not only to the limitations of EVV 2DIR but also to the limitations of the techniques used to generate the assignments in the literature. Full assignments of these spectra will be made possible upon the arrival of calculated spectra. All of the preliminary asignmenst made here have been to vibrational modes of guanine and cytosine, It is unsurprising that the most noticeable differences between the two spectra originate from these bases as together they form 68% of the composition of the DNA sequences studied, Assuming equally sized spectra features from each of the bases, the contributions of guanine and cytosine

would be 4.5 times more intense than those of adenine and thymine in these spectra, based only on the composition of the sequences. Interestingly, preliminary calculations show the most intense base features of guanine and cytosine to be around an order of magnitude more intense than those of adenine and thymine. This approximate analysis suggests that it is unsurprising for the most readily assigned peaks to be arising from guanine and cytosine.

5.3 EVV 2DIR studies of G-Quadruplex Structures.

This section describes the preliminary work done to detect EVV 2DIR spectral features arising from the vibrational couplings present in intramolecular parallel G-quadruplexes. This was done by comparison of the spectra of Myc2345, in the presence of stabilising potassium ions, with the spectra of the DNA sequence named SS1, under the same conditions.

SS3 ("single strand 3" sequence provided in appendix A) was a sequence designed to contain the same number of each DNA base, but without any of the four adjacent guanine bases, required to make the guanine tetrads of a G-quadruplex structure. It was therefore used as a non-quadruplex forming control as justified above.

5.3.1 Sample preparation and data collection of Myc2345 quadruplexes

To compare the EVV 2DIR spectra of a quadruplex forming sequence and a non quadruplex forming control, two gel spot samples were prepared, as per the methodology described in section 3.3 for producing gel spots for EVV 2DIR. Two solutions were prepared:

- 1mM Myc2345 10mM KCl pH7
- 1mM SS1 10mM KCl pH7

These two solutions were annealed at 90° for 3 minutes and allowed to cool before depositing onto coverslips for gel formation. Drying 1ul spots of these solutions slowly at 85.2% relative humidity in the humidity controlled sample cell described in section 3.3 was found to produce flat topped gel spots, which remained stable in the beams over the course of an EVV 2DIR experiment.

The 1:10 ratio of DNA sequences to potassium ions is consistent with that validated by circular dichorism to produce G-quadruplex structures.

Data was collected between the ranges ω_α=1235 – 1735cm^{-1} and ω_β= 2605 – 3455cm^{-1} at 5cm^{-1} resolution averaging 100 samples per point. All three beams were set to propagate with P polarisation and Time delays of T_1=1.75ps and T_2=1.0ps were used to give a good compromise between minimising non-resonant background and maximising resonant signal size. The spectrum of each sample was recorded three times, alternating between the two samples to prevent any minor drift in the laser system over the course of the data collection impacting the two samples differently. Each of the three repeats were recorded at slightly different positions on the gel spot to prevent any anomalous spot dominating the data collection.

The data was processed in MATLAB, as described in section 3.2. The individual repeats of each measurement were intensity normalised on the largest spectral feature, observed at ω_α=1535cm^{-1} / ω_β=3010 cm^{-1} such that the repeats could be averaged.

5.3.2 EVV 2DIR Spectra of Myc2345 Quadruplex

The mean processed EVV 2DIR spectra of the single stranded and quadruplex forming Myc3245 sequence can be seen in Figure 5-3

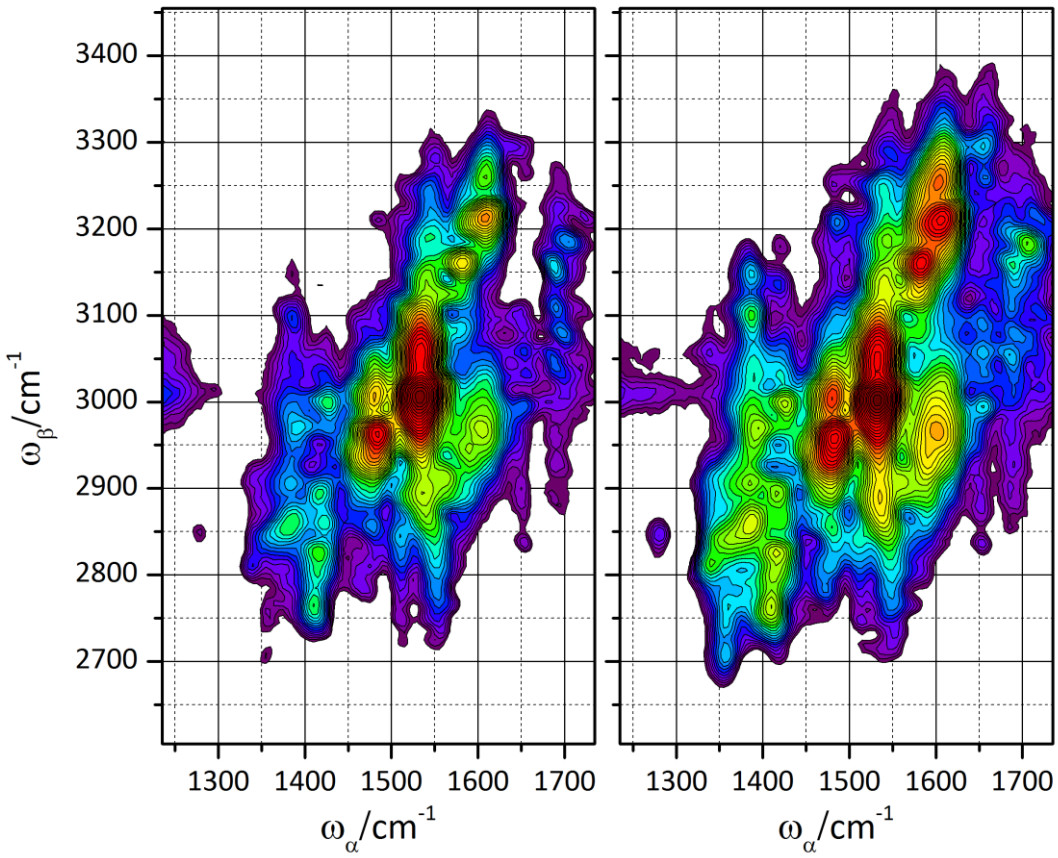

Figure 5-3 (a) The mean processed EVV 2DIR spectrum of the single strand SS3 in the presence of potassium ions. (b) The mean processed EVV 2DIR spectrum of quadruplex forming sequence Myc2345 in the presence of potassium ions.

Before considering the differences between the two spectra shown in Figure 5-3, it may be instructive to consider the differences between these spectra and those of the duplex and control shown described in section 5.2.

The high frequency overtone modes assigned to the guanine carbonyl stretches at ω_α=1675/1690cm^{-1} are not visible in the spectra shown in Figure 5-3. This seems surprising given the high guanine content of the Myc2345 sequence. The absence of

these overtone features can potentially be explained by considering the addition of potassium ions to the quadruplex forming sample and the control, to allow for stacking of guanine tetrads. It has been shown previously in the Klug group[104] that the presence of ions in a sample can cause the suppression of cross peaks arising from proximal moieties. These observations have been corroborated computationally, with calculations showing proximal ions affecting both the transition dipole of a probed coherence and also the electrical anharmonicities between the coupled modes. The effect of this is a reduction in EVV peak intensity. It was shown that a positive ion placed 5 Å from a methylene peak resulted in a 70% decrease in the transition dipole associated with the overtone peak. Considering this, it seems perfectly plausible that the 10:1 ratio of potassium ions to DNA molecules could result in the suppression of the guanine carbonyl stretch overtone peaks.

A second immediate difference between the spectra of Myc2345 and SS3 shown in Figure 5-3 and the duplex and single strand spectra shown in section 5.2 is the presence of the broad high intensity feature at around ω_α=1605 cm^{-1} / ω_β=2970 cm^{-1} in Figure 5-3. No sensible assignment was found for this frequency in the literature; however its overtone peak is strongly present in all of the DNA spectra presented here. The fact that this cross peak is present only in these Myc2345 and SS3 spectra suggests that it potentially originates from a structure mediated by the presence of potassium ions. Both Mcy2345 and SS3 sequences are guanine rich with adjacent guanine bases. This feature could potentially be due to guanine-guanine interactions which are not specific to the quadruplex structure[105]. This indicates that SS3 is potentially not a good control and may form at least some of the same interactions seen in G-quadruplexes.

In order to better see the differences between the spectra of the quadruplex forming Myc2345 sequence and the control sequence, SS3, spectra over a smaller spectral region are shown in Figure 5.4.

Three peaks have been identified in Figure 5.4 which appear exclusively in the quadruplex forming Myc2345 sample. The first of these peaks, marked A, can be seen at ω_α=1670cm^{-1} / ω_β=3135 cm^{-1}. Vibrations at 1670cm^{-1} have previously been assigned to the guanine carbonyl stretch as seen in G-quadruplexes, undergoing Hoogsteen hydrogen bonding, based on both calculated and experimental observations[57]. Peak A is due to the coupling of this mode to a combination of this mode and another at 1485cm^{-1}. This can be assigned as a guanine ring mode, based on assignments reported in the literature[99]. It seems perfectly plausible for a guanine ring mode to be coupled to its carbonyl stretch, given these two moieties are bound directly to one another. The peak marked B at ω_α=1485 cm^{-1} / ω_β=3135 cm^{-1} represents the coupling of the same two vibrational modes, however this time the guanine ring mode forms the fundamental mode in ω_α.

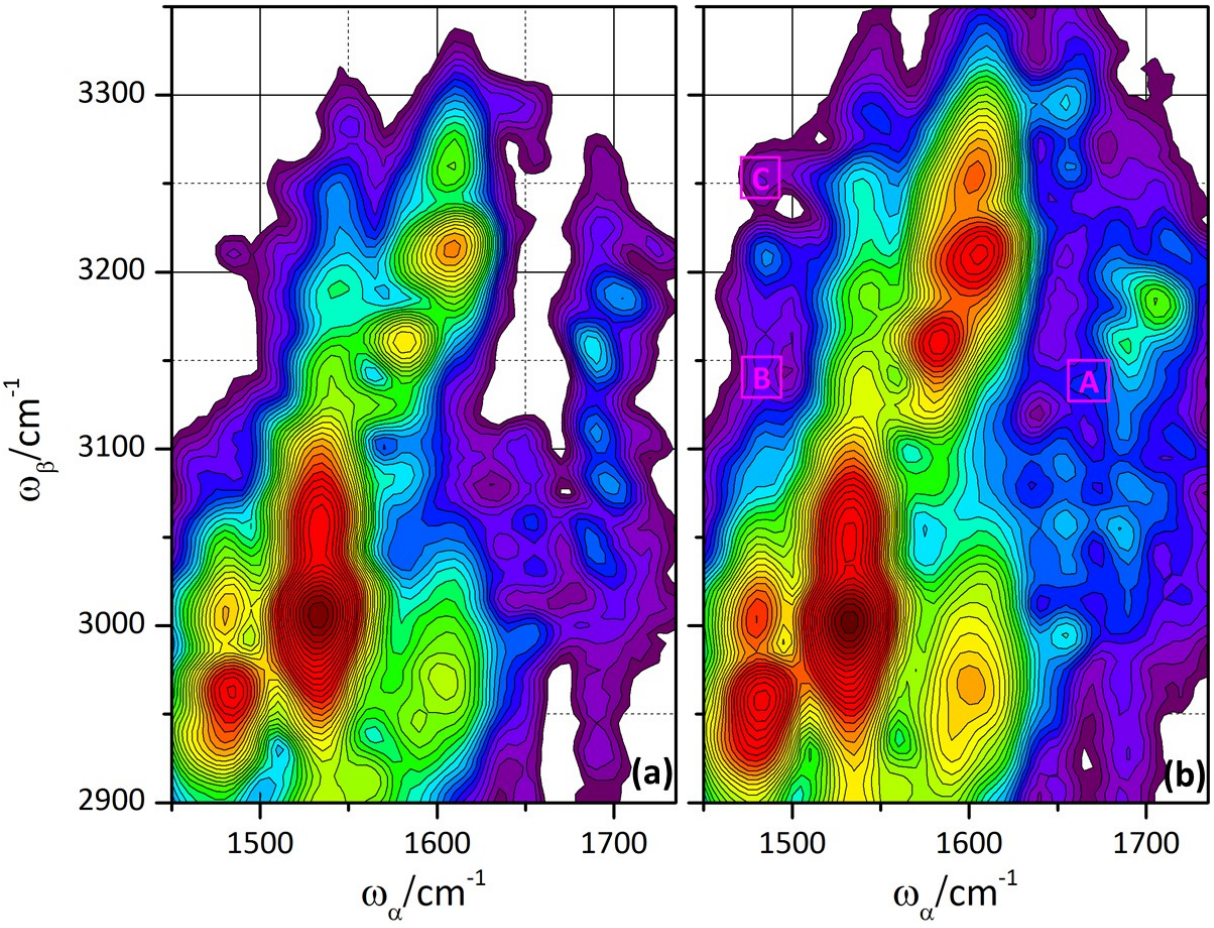

Figure 5-4 Confined spectral regions of (a) The mean processed EVV 2DIR spectrum of the single strand SS3 in the presence of potassium ions. (b) The mean processed EVV 2DIR spectrum of quadruplex forming sequence Myc2345 in the presence of potassium ions.

Peak C in Figure 5.4 is also present exclusively in the spectrum of the Myc2345 sequence. This represents peak arises again from the guanine ring mode at 1485 cm^{-1}, this time coupled to a mode at 1770 cm^{-1}. No reasonable assignment of this mode could be made from the literature. It seems likely, that this peak is due to a new vibrational coupling made possible as a result of the guanine ring being brought proximal to another vibrational mode, as a result of the quadruplex formation.

5.4 Conclusions

Peaks have been observed in the EVV 2DIR spectra whose assignments are consistent with them being indicative of duplex formation and quadruplex formation. In the event that the assignments made here are incorrect, some of these peaks must still be indicative of these secondary structures, as they are present exclusively in the structure forming samples.

From the measurements made of two single strands and the duplex forming complementary strands it was shown that shifts in carbonyl frequencies, corresponding to Watson Crick base parings are visible in the EVV 2DIR spectra. It was also shown that the EVV 2DIR spectrum is sensitive to the syn or anti-conformation of the guanine base, relative to the sugar rings of DNA. These spectral features, indicative of duplex formation, were assigned based on frequencies reported in the literature from linear spectroscopies, so ultimately just observing them in the EVV spectrum is not adding any information that could not have been extracted via linear techniques. It is the vibrational coupling information that is offered by EVV 2DIR that gives new information over these linear techniques. Three peaks were assigned to couplings between distinct vibrational modes (not overtones) although they were all couplings between vibrations on the same molecular groups. There are a great many more cross peaks reported in this data, which were unable to be assigned, of which some will no doubt be couplings between different parts of the individual DNA molecules, or between different DNA strands. Cross peaks corresponding to the hydrogen bonding between the carbonyl and amine groups of complementary Watson-Crick paired bases was not observed as the frequency of the amine vibrations is outside the spectral range studied here.

The EVV 2DIR spectra of quadruplex forming sequence Myc2345, displayed two cross peaks which were assigned to a mode specific to quadruplex formation, coupled to a guanine ring mode. The Myc2345 spectrum also showed a unique peak corresponding to the coupling of the guanine ring mode to another unknown mode.

The spectra showed a couple of suggestions that the non-structure-forming controls used were not entirely suitable for their purpose. Some intensity corresponding to

duplexed guanine was observed in the "single strand" control, whilst some unassigned non-overtone feature was present in both the guanine rich sequences used for the quadruplex study. The implication of this is that the control samples need to be redesigned for future works. It may be possible in the case of the duplex measurements, to measure the spectrum of the two single strands individually and add the two resultant spectra together. This combined spectrum could then be compared with the spectrum of the duplex sample.

Direct comparison of intensities of peaks between the measurements was not possible for the spectra presented here, as no peak was known to be consistent between the measurements to use for intensity normalization. Calculated spectra may serve as a guide as to which peaks may be suitable to use for intensity normalization. If this is not the case, a small quantity of an inert, spectrally simple internal standard could be added to each of the samples. If an appropriate additive was chosen, with a well resolved feature, the intensity of this feature could be used to normalize the data.

Once calculated spectra have been produced for these systems, full assignments will be made, making it possible to fully explore EVV 2DIRs true sensitivity to DNA secondary structures.

6 Conclusions and future work

This thesis presents the first examples of the application of EVV 2DIR to detect and characterise protein-inhibitor interactions and to study DNA secondary structures

The work presented demonstrates EVV 2DIRs ability to detect vibrational couplings arising due to an inhibitors interaction with its binding site. By comparison with calculated spectra these vibrational couplings could be assigned to specific vibrational modes of the inhibitor protein complex. It was shown that by considering the ratio of peaks intensity in two polarisation schemes, experimental spectra can be compared with calculated spectra as a method of validating the geometry of the inhibitor-protein complex used in the calculation. Although systematic errors arising from optically thick samples prevented good agreement between the experimental and calculated spectra, the good agreement between the two in terms of peak number and location suggests that improved experimental protocol may lead to agreement between the two. Experimental agreement with calculated spectra may be achieved by using optically thinner samples and potentially conducting the experiment in a less spectrally congested region, such as those described in section 4.6.

Achieving good agreement between experimental peak ratios and those calculated based on a crystal structure geometry would pave the way for using EVV 2DIR as a tool for determining the structure of inhibitor protein complexes. This could be done by combining calculations and experiment, by iteratively improving the geometry of a binding model such that its resultant calculated spectra converge upon the experimental spectra. This could be experimentally high throughput, requiring very small quantities of material, compared to current methods. The experiments described here used 30μg of protein in the gel spot, formed from 1μl of protein solution. Of this 30μg, only 75ng of protein would be in the 100μm wide focus of the beam during an experiment. With the relatively short time required to collect a spectrum, large numbers of compounds could be screened in quick succession. The spectra presented here each took 100 minutes to collect however this could be brought down significant by the use of currently available commercially lasers, with 100 times quicker repetition rate than the one used for this work.

There would be two requirements to be able to determine the structure of a protein inhibitor complex in this way: Firstly an apo crystal structure of the protein binding site that is representative of the structure in solution. Secondly (and more ambitiously) a computational method capable of systematically varying the geometry of its model such that the resultant PPP/PPS peak ratios approach those of the experimental spectrum. Importantly, this method would have to work on a usable timescale. The viability of such a method is to be explored by those specialist in the field and is far beyond the scope of this thesis.

Even without the ability to determine absolute geometry from an EVV spectrum using this iterative calculation approach, EVV 2DIR has been shown to have the potential to be a valuable tool for elucidating useful information about protein inhibitor interactions. All of the spectra presented in this thesis were collected at fixed time delays. It is possible that EVV 2DIR may be able to offer dynamical insights into inhibitor protein complexes, that may be of biological significance. This is something that would need exploring experimentally.

The first EVV 2DIR spectra of DNA to be recorded were presented in this thesis. The objective of the experiments was to determine EVV 2DIRs sensitivity to sequence

dependent secondary structures that DNA can form. Peaks were observed that were indicative of both duplex formation and quadruplex formation. As these were the first DNA spectra of their kind to be recorded, the peaks went largely unassigned. This was as a result of a lack of calculated spectra to make assignments from. The assignments that were made were made on the basis of observations in the linear spectroscopy literature. This resulted in the interpretation being limited to that which was already possible through linear techniques.

The DNA spectra showed an incredibly high peak density, with many nicely resolved features. This is encouraging for the prospect of using EVV 2DIR a a structural probe of DNA secondary structures, as many cross peaks means many possible opportunities to look for structural constraints. Although the features were not assigned, the arrival of calculated spectra will allow for full assignment and interpretation of these spectra. Once full interpretation is possible, there are a number of interesting experiments which will be possible. One of these would be the combination of the two parts of the work described in this thesis. G-quadruplexes have recently emerged as therapeutic targets and studies have shown that small molecules can be used to facilitate the formation of, and stabilize, G-quadruplex sturctures[106,107,108]. This can be done with the aim of altering the iterations with the proteins responsible for gene regulatory processes, which interact with them. It has been shown here that EVV 2DIR has the ability to detect interactions between inhibitors and protein binding sites, as well as its potential as a structural probe for DNA secondary structures. As a result of this it appears that EVV 2DIR may be a good tool for aiding the development of these new quadruplex targeting therapeutic small molecules. It would potentially be able to simultaneously report on the interactions between the small molecule and the quadruplex structure and well as the effect of small molecule binding on the stability and structure of the G-quadruplex structure itself. This is just one of many potential applications of EVV 2DIR, now that it is understood to be sensitive to interactions between biomolecules.

Bibliography

1. Stumpf, M. P. H. *et al.* Estimating the size of the human interactome. *PNAS* **105,** 6959–6964 (2008).

2. Hudson, W. H. & Ortlund, E. A. The structure, function and evolution of proteins that bind DNA and RNA. *Nat Rev Mol Cell Biol.* **15,** 749–760 (2014).

3. Anderson, A. C. The Process of Structure-Based Drug Design. *Cell Chem. Biol.* **10,** 787–797 (2003).

4. Gianti, E., Delemotte, L., Klein, M. L. & Carnevale, V. On the role of water density fluctuations in the inhibition of a proton channel. *PNAS* **113,** 8359–8368 (2016).

5. Bairagya, H. R., Mukhopadhyay, B. P. & Bhattacharya, S. Role of the conserved water molecules in the binding of inhibitor to IMPDH-II (human): A study on the water mimic inhibitor design. *J. Mol. Struct. THEOCHEM* **908,** 31–39 (2009).

6. Su, J., Zhang, T., Wang, P., Liu, F. & Tai, G. The water network in galectin-3 ligand binding site guides inhibitor design. *Acta Biochim. Biophys. Sin. (Shanghai).* **47,** 192–198 (2017).

7. Acharya, K. R. & Lloyd, M. D. The advantages and limitations of protein crystal structures. *Trends Pharmacol. Sci.* **26,** 10–14 (2005).

8. Zheng, H., Jing, H. & Zimmerman, M. D. The future of crystallography in drug discovery. *Expert Opin. Drug Discov.* **9,** 125–137 (2015).

9. Pellecchia, M. *et al.* Perspectives on NMR in drug discovery: a technique comes of age. *Nat. Rev. Drug Discov.* **7,** 738–745 (2008).

10. Eathiraj, S. *et al.* A novel mode of protein kinase inhibition exploiting hydrophobic motifs of autoinhibited kinases: discovery of ATP-independent inhibitors of fibroblast growth factor receptor. *J. Biol. Chem.* **286,** 20677–87 (2011).

11. Nilges, M. & Markwick, P. R. L. Structural Biology by NMR : Structure , Dynamics , and Interactions. *PLoS Comput. Biol.* **4,** (2008).

12. Schwieters, C. D., Kuszewski, J. J., Tjandra, N. & Clore, G. M. The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **160,** 65–73 (2003).

13. Hamm, P., Helbing, J. & Bredenbeck, J. Two-Dimensional Infrared Spectroscopy of Photoswitchable Peptides. *Annu. Rev. Phys. Chem.* **59,** 291–317 (2008).

14. McCoy, S. & Caughey, W. Infrared Studies of Azido, Cyano, and Other Derivatives of Metmyoglobin, Methemoglobin, and Hemins. *Biochemistry* **9,** 2387–2393 (1970).

15. 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).

16. Fang, C. *et al.* Two-dimensional infrared spectra reveal relaxation of the nonnucleoside inhibitor TMC278 complexed with HIV-1 reverse transcriptase. **105,** 1–6 (2007).

17. Bloem, R. *et al.* Ligand Binding Studied by 2D IR Spectroscopy Using the

Azidohomoalanine Label. **116,** 13705–13712 (2012).

18. Ghosh, A. *et al.* 2D IR spectroscopy reveals the role of water in the binding of channel-blocking drugs to the influenza M2 channel. *J. Chem. Phys.* **113,** 235105 1-9 (2014).

19. Ferrari, S. Protein kinases controlling the onset of mitosis. *Cell. Mol. Life Sci.* **63,** 781–95 (2006).

20. Manning, G., Whyte, D. B., Martinez, R., Hunter, T. & Sudarsanam, S. The protein kinase complement of the human genome. *Science* **298,** 1912–34 (2002).

21. Shchemelinin, I. Review Protein Kinases , Their Function and Implication in Cancer and Other Diseases. **101,** 81–100 (2006).

22. Yogi, A., O'Connor, S. E., Callera, G. E., Tostes, R. C. & Touyz, R. M. Receptor and nonreceptor tyrosine kinases in vascular biology of hypertension. *Curr. Opin. Nephrol. Hypertens.* **19,** 169–76 (2010).

23. Johnson, L. N. & Lewis, R. J. Structural basis for control by phosphorylation. *Chem. Rev.* **101,** 2209–42 (2001).

24. Rethinaswamy, A. *et al.* Structures of the Tyrosine Kinase Domain of Fibroblast Growth Factor Receptor in Complex with Inhibitors. *Science (80-.).* **276,** 955–960 (1997).

25. Galzie, Z., Kinsella, A. R. & Smith, J. A. Fibroblast growth factors and their receptors. **685,** 669–685 (1997).

26. Takami, K. *et al.* Fibroblast growth factor receptor-1 expression in the cortex and hippocampus in Alzheimer's disease. *Brain Res.* **802,** 89–97 (1998).

27. Ferrer, I. & Martí, E. Distribution of fibroblast growth factor receptor-1 (FGFR-1) and FGFR-3 in the hippocampus of patients with Alzheimer's disease. *Neurosci. Lett.* **240,** 139–142 (1998).

28. Kranenburg, A. R. *et al.* Chronic obstructive pulmonary disease is associated with enhanced bronchial expression of FGF-1, FGF-2, and FGFR-1. *J. Pathol.* **206,** 28–38 (2005).

29. Immunolocalization of fibroblast growth factor-1 (FGF-1), its inflammatory renal disease. **68,** 2621–2628 (2005).

30. Walker, D. G. *et al.* Immunohistochemical analyses of fibroblast growth factor receptor-1 in the human substantia nigra. Comparison between normal and Parkinson's disease cases. *Brain Res.* **794,** 181–187 (1998).

31. Knights, V. & Cook, S. J. De-regulated FGF receptors as therapeutic targets in cancer. *Pharmacol. Ther.* **125,** 105–17 (2010).

32. Klein, T. *et al.* Structural and dynamic insights into the energetics of activation loop rearrangement in FGFR1 kinase. *Nat. Commun.* **6,** 1–12 (2015).

33. Norman, R. a *et al.* Protein-ligand crystal structures can guide the design of selective inhibitors of the FGFR tyrosine kinase. *J. Med. Chem.* **55,** 5003–12 (2012).

34. Vajpai, N., Schott, A.-K., Vogtherr, M. & Breeze, A. L. NMR backbone assignments of the tyrosine kinase domain of human fibroblast growth factor receptor 1. *Biomol.*

NMR Assign. **8,** 85–88 (2014).

35. Kiselyov, V. V, Berezin, V., Bock, E. & Poulsen, F. M. H-1 and N-15 resonance assignment of the first module of FGFRl. *J. Biomol. NMR* **36,** 28 (2006).

36. Treiber, D. K. & Shah, N. P. Ins and outs of kinase DFG motifs. *Chem. Biol.* **20,** 745–746 (2013).

37. Tucker, J. A. *et al.* Structural insights into FGFR kinase isoform selectivity: Diverse binding modes of AZD4547 and ponatinib in complex with FGFR1 and FGFR4. *Structure* **22,** 1764–1774 (2014).

38. Laird, A. D. *et al.* SU6668 is a potent antiangiogenic and antitumor agent that induces regression of established tumors. *Cancer Res.* **60,** 4152–4160 (2000).

39. Hunter, C. Sequence Dependent DNA structure, the role of base stacking interactions. *J. Mol. Biol.* **230,** 1025–1054 (1993).

40. Watson, J. D. & Crick, F. H. C. Molecular structure of nucleic acids. *Nature* **171,** 737–738 (1953).

41. Krummel, A. T., Mukherjee, P. & Zanni, M. T. Inter and Intrastrand Vibrational Coupling in DNA Studied with Heterodyned 2D-IR Spectroscopy. *J. Phys. Chem. B* 9165–9169 (2003).

42. Fidder, H. *et al.* N − H Stretching Vibrations of Guanosine − Cytidine Base Pairs in Solution : Ultrafast Dynamics , Couplings , and Line Shapes. *J. Phys. Chem. A* **117,** 845–854 (2013).

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

44. Greve, C. *et al.* N − H Stretching Excitations in Adenosine-Thymidine Base Pairs in Solution : Pair Geometries , Infrared Line Shapes , and Ultrafast Vibrational Dynamics. *J. Phys. Chem. A* **117,** 594–606 (2013).

45. Krummel, A. T. & Zanni, M. T. DNA vibrational coupling revealed with two-dimensional infrared spectroscopy: Insight into why vibrational spectroscopy is sensitive to DNA structure. *J. Phys. Chem. B* **110,** 13991–14000 (2006).

46. Yakovchuk, P., Protozanova, E. & Frank-kamenetskii, M. D. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res.* **34,** 564–574 (2006).

47. Ramakers, L. A. I. *et al.* 2D-IR Spectroscopy Shows that Optimized DNA Minor Groove Binding of Hoechst33258 Follows an Induced Fit Model. *J. Phys. Chem. B* **121,** 1295–1303 (2017).

48. Neidle, S. & Balasubramanian, S. *Quadruplex Nucleic Acids*. (The Royal Society of Chemistry, 2006). doi:10.1039/9781847555298

49. Bochman, M. L., Paeschke, K. & Zakian, V. A. REVIEWS DNA secondary structures : stability and function of G - quadruplex structures. *Nat. Rev. Genet.* **13,** 770–780 (2012).

50. Neidle, S. & Parkinson, G. N. The structure of telomeric DNA. *Curr. Opin. Struct. Biol.* **13,** 275–283 (2003).

51. Rhodes, D. & Lipps, H. J. G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.* **43,** 8627–8637 (2017).

52. Donald M. Miller, Shelia D. Thomas, Ashraful Islam1, David Muench, and K. & Sedoris. c-Myc and Cancer Metabolism. *Clin. cancer Res.* **18,** 5546–5553 (2013).

53. Pecinka, P. & Kubista, M. DNA tetraplex formation in the control region of c- DNA tetraplex formation in the control region of c-myc. *Nucleic Acids Res.* **26,** 1167–1172 (1998).

54. Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K. & Neidle, S. Quadruplex DNA : sequence , topology and structure. *Nucleic Acids Res.* **34,** 5402–5415 (2017).

55. Tua, A., Modi, Y. S. & Patel, D. J. Propeller-Type Parallel-Stranded G-Quadruplexes in the Human c-myc Promoter. *J. Am. Chem. Soc.* **126,** 15021–15027 (2004).

56. Neidle, S. & Parkinson, G. N. Quadruplex DNA crystal structures and drug design. *Biochimie* **90,** 1184–1196 (2008).

57. Andrushchenko, V., Tsankov, D., Krasteva, M., Wieser, H. & Bouř, P. Spectroscopic detection of DNA quadruplexes by vibrational circular dichroism. *J. Am. Chem. Soc.* **133,** 15055–15064 (2011).

58. Guzm, M. R., Liquier, J., Brahmachari, S. K. & Taillandier, E. Characterization of parallel and antiparallel G-tetraplex structures by vibrational spectroscopy. *Spectrochim. Acta Part A* **64,** 495–503 (2006).

59. Atkins, P. W. & Friedman, R. S. Molecular Rotations and Vibrations. in *Molecular Quantum Mechanics* 344–376 (Oxford University Press, 1997).

60. New, G. Frequency mixing. in *Introduction to Nonlinear Optics* 19–44 (Cambridge University Press - M.U.A, 2011).

61. Römer, H. Foundations of Nonlinear Optics. in *Theoretical Optics* 83–105 (Wiley-VCH Verlag GmbH & Co. KGaA, 2005). doi:10.1002/3527604294.ch6

62. Wright, J. C. *Coherent multidimensional vibrational spectroscopy*. *International Reviews in Physical Chemistry* **21,** (2002).

63. Zhao, W. & Wright, J. C.ˆ for Doubly Vibrationally Enhanced Four Wave Mixing Spectroscopy Measurement of xˁ 3ˆ. 10–13 (1999).

64. Paschotta, R. Phase Matching. in *Encyclopedia of Laser Physics and Technology* (Wiley-VCH Verlag GmbH & Co. KGaA, 2008).

65. Milonni, P. W. & Eberly, J. H. Introduction to nonlinear optics. in *Laser Physics* 457–497 (John Wiley & Sons Inc, 2010).

66. Turnbull, M. T., Petrov, P. G., Embrey, C. S., Marino, A. M. & Boyer, V. Role of the phase-matching condition in nondegenerate four-wave mixing in hot vapors for the generation of squeezed states of light. *Phys. Rev. A - At. Mol. Opt. Phys.* **88,** 1–10 (2013).

67. Zhao, W. & Wright, J. C. Doubly vibrationally enhanced four wave mixing: the optical analog to 2D NMR. *Phys. Rev. Lett.* **84,** 1411–4 (2000).

68. Murdoch, K. M. *et al.* Isotope and mode selectivity in two-dimensional vibrational four wave mixing spectroscopy. *Chem. Phys. Lett.* **335,** 349–356 (2001).

69. Kwak, K., Cha, S., Cho, M. & Wright, J. C. Erratum: Vibrational interactions of acetonitrile: Doubly vibrationally resonant IR-IR-visible four-wave-mixing spectroscopy. *J. Chem. Phys.* **118,** 2968 (2003).

70. Fournier, F. *et al.* Optical fingerprinting of peptides using two-dimensional infrared spectroscopy: proof of principle. *Anal. Biochem.* **374,** 358–65 (2008).

71. Fournier, F. *et al.* Protein identification and quantification by two-dimensional infrared spectroscopy: implications for an all-optical proteomic platform. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 15352–7 (2008).

72. Rezende Valim, L. *et al.* Identification and Relative Quantification of Tyrosine Nitration in a Model Peptide Using Two-Dimensional Infrared Spectroscopy. *J. Phys. Chem. B* **118,** 12855–12864 (2014).

73. Fournier, F. *et al.* Biological and biomedical applications of two-dimensional vibrational spectroscopy: proteomics, imaging, and structural analysis. *Acc. Chem. Res.* **42,** 1322–31 (2009).

74. Guo, R. *et al.* Detection of complex formation and determination of intermolecular geometry through electrical anharmonic coupling of molecular vibrations using electron-vibration-vibration two-dimensional infrared spectroscopy. *Phys. Chem. Chem. Phys.* **11,** 8417–21 (2009).

75. Of, R. & Problems, T. Let there be white light : supercontinuum generation by ultrashort laser pulses. *Physics-Uspekhi* **49,** 605–627 (2006).

76. SpectraPhysics. OPA-800c User's Manual. (2003).

77. Haris, P. I. & Severcan, F. FTIR spectroscopic characterization of protein structure in aqueous and non-aqueous media. *J. Mol. Catal. - B Enzym.* **7,** 207–221 (1999).

78. Deegan, R. D. *et al.* Capillary flow as the cause of ring stains from dried liquid drops. *Nature* **389,** 827–829 (1997).

79. Zhang, J., Müller-Plathe, F. & Leroy, F. Pinning of the Contact Line during Evaporation on Heterogeneous Surfaces: Slowdown or Temporary Immobilization? Insights from a Nanoscale Study. *Langmuir* **31,** 7544–7552 (2015).

80. Wang, F.-C. & Wu, H.-A. Pinning and depinning mechanism of the contact line during evaporation of nano-droplets sessile on textured surfaces. *Soft Matter* **9,** 5703 (2013).

81. Greenspan, L. Humidity fixed points of binary saturated aqueous solutions. *J. Res. Natl. Bur. Stand. Sect. A Phys. Chem.* **81A,** 89 (1977).

82. Harding, S. E. & Chowdhry, B. Z. *Protein-ligand Interactions: Hydrodynamics and Calorimetry : a Practical Approach*. (Oxford University Press, 2001).

83. Freire, E., Obdulio, L. & Straume, M. Isothermal Titration Calorimetry. *Anal. Chem.* **62,** (1990).

84. Brown, A. Analysis of Cooperativity by Isothermal Titration Calorimetry. *Int. J. Mol. Sci.* **10,** 3457–3477 (2009).

85. Friedman, S. J. & Terentis, A. C. Analysis of G-quadruplex conformations using Raman and polarized Raman spectroscopy. *J. Raman Spectrosc.* **47,** 259–268

(2016).

86. Novy, J., Bohm, S. & Kralova, J. Formation and Temperature Stability of G-Quadruplex Structures Formation and Temperature Stability of G-Quadruplex Structures Studied Studied by Electronic and Vibrational Circular Dichroism by Electronic and Vibrational Circular Dichroism d Vibrational C. *Biopolymers* **89,** 144–152 (2008).

87. Scheurer, C. & Mukamel, S. Magnetic Resonance Analogies in Multidimensional Vibrational Spectroscopy. *Chem. Soc. japan* **75,** 989–999 (2002).

88. Scheurer, C. & Mukamel, S. Infrared analogs of heteronuclear nuclear magnetic resonance coherence transfer experiments in peptides Infrared analogs of heteronuclear nuclear magnetic resonance coherence transfer experiments in peptides. *J. Chem. Phys.* **116,** 6803–6816 (2002).

89. Liang, G., Liu, Z., Wu, J., Cai, Y. & Li, X. Anticancer molecules targeting fibroblast growth factor receptors. *Trends Pharmacol. Sci.* **33,** 531–41 (2012).

90. Becke, A. D. Density-functional thermochemistry . III . The role of exact exchange. *J. Chem. Phys.* **98,** 5648 (1993).

91. Lee, C., Hill, C. & Carolina, N. Development of the Colic-Salvetti correlation-energy formula into a functional of the electron density. *Physcial Rev. B* **37,** (1988).

92. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ort, J. B. F. Gaussian 09 Revision D.01. (2009).

93. Kwak, K., Cha, S., Cho, M. & Wright, J. C. Vibrational interactions of acetonitrile: Doubly vibrationally resonant IR–IR–visible four-wave-mixing spectroscopy. *J. Chem. Phys.* **117,** 5675 (2002).

94. Taylor, J. R. *Introduction To Error Analysis: The Study of Uncertainties in Physical Measurements*. (University Science Books, 1997).

95. Luthman, K. & Behe, J. Sequence Dependence of DNA Structure. *J. Biol. Chem.* **263,** 15535–15539 (1988).

96. Plumridge, A., Meisburger, S. P. & Pollack, L. Visualizing single-stranded nucleic acids in solution. *Nucleic Acids Res.* **45,** 1–13 (2017).

97. Murphy, M. C., Rasnik, I., Cheng, W., Lohman, T. M. & Ha, T. Probing Single-Stranded DNA Conformational Flexibility Using Fluorescence Spectroscopy. *Biophys. J.* **86,** 2530–2537 (2004).

98. Speedy, R. J. Kauzmann ' s paradox and the glass transition. *Biophys. Chem.* **105,** 411–420 (2003).

99. Tsuboi, M. Application of Infrared Spectroscopy to Structure Studies of Nucleic Acids. *Appl. Spectrosc. Rev.* **3,** 45–90 (1970).

100. Mantsch, H. H. & Chapman, D. *Infrared Spectroscopy of Biomolecules*. (Wiley, 1996).

101. Taillandier, E. [16] By. **211,** (1992).

102. Taillandier, E., Peticolas, W. L., Adam, S. & Huynh-Dinh, T. Polymorphism of the d(CCCGCGGG) dou-2 ble helix studies by FTIR spectroscopy. *Spectrochim. Acta Part A* **64,** 107–112 (1990).

103. Stolarski, R., Hagberg, C. & Shugar, D. Studies on the dynamic syn-anti equilibrium in purine nucleosides and nucleotides with the aid of 'H and 13C NMR spectroscopy. *Eur. J. Biochem.* **138,** 187–192 (1984).

104. Cheung, M. Electron-Vibration-Vibration Two-Dimensional Infrared Spectroscopy: A novel method for phosphorylation quantification. (Imperial College London, 2015).

105. Al-hashimi, H. M. A Historical Account of Hoogsteen Base-pairs in Duplex DNA. *Biopolymers* **99,** 1–21 (2014).

106. Han, H. & Hurley, L. H. G-quadruplex DNA : a potential target for anti-cancer drug design. *Trends Pharmacol. Sci.* **21,** 136–142 (2000).

107. Balasubramanian, S., Hurley, L. H. & Neidle, S. Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nat. Rev. Drug Discov.* **10,** 261–275 (2011).

108. Onel, B., Clement, L. & DanZhou, Y. DNA G-quadruplex and its potential as anticancer drug target. *Sci. China Chem.* **57,** 1605–1614 (2016).

Appendix A: DNA sequences used

Myc2345: TGAGGGTGGGGAGGGTGGGGAA

SS1: GGCCCTTACCGGGCATTCCGGG

SS2: GAGCTACAGAGCGCAGCTCTCG

SS3: TTCCCCACCCTCCCCACCCTCA

Appendix B: FGFR1 construct sequence

MHHHHHHGSTSLYKKAGSSENLYFQGAGVSEYELPEDPRWELPRDRLVLGKPLGEGAFG
QVVLAEAIGLDKDKPNRVTKVAVKMLKSDATEKDLSDLISEMEMMKMIGKHKNIINLLGA
CTQDGPLYVIVEYASKGNLREYLQARRPPGLEYSYNPSHNPEEQLSSKDLVSCAYQVARGM
EYLASKKCIHRDLAARNVLVTEDNVMKIADFGLARDIHHIDYYKKTTNGRLPVKWMAPE
ALFDRIYTHQSDVWSFGVLLWEIFTLGGSPYPGVPVEELFKLLKEGHRMDKPSNCTNELY
MMMRDCWHAVPSQRPTFKQLVEDLDRIVALTSNQE

Appendix C: EVV 2DIR spectrum fitting code.

This code is to be executed in MATLAB version R2014B or later. It fits 2D Gaussians to the EVV spectrum which should be saved as a .dat file named "datatofit". The start points for the fitting should be saved in a .dat file named "start_points" oragnised in four columns for the four Gaussian parameters ordered intensity, width, x centre, y centre.

```
parpool(16)

    data_raw = load('datatofit')
    start_points=load('start_points')

    start_points_temp=start_points



    for q=1:length(start_points_temp)

start_points_temp(q,1)=data_raw(round((start_points_temp(q,4))),((roun
d(start_points_temp(q,3)))))
    end

    ysub=(1:length(data_raw(:,1)));
    xsub=(1:length(data_raw(1,:)));


    %preparing surface data for using in fit function
    [xout, yout, zout]=prepareSurfaceData(xsub,ysub,data_raw) ;



    %creating  widthsigma starting values for input into equation
    for j = 1:length(start_points_temp)
        widthsigma(j) = 2;

    end

    %creating the start point data
    startpoints =
[start_points_temp(:,1)',widthsigma,start_points_temp(:,3)',start_poin
ts_temp(:,4)'];


  %create lower limits for parameters

    for m = 1:size(start_points,1)
        mina(m)=0;
        minx(m) = start_points_temp(m,3)-1;
        miny(m) = start_points_temp(m,4)-1;
        minwidthsigma(m)= 1.2
    end

    lowerbound = [mina,minwidthsigma,minx,miny];
```

```
    %create upper limits for parameters

    for p = 1:size(start_points,1)
        maxa(p)=2;
        maxx(p) = start_points_temp(p,3)+1;
        maxy(p) = start_points_temp(p,4)+1;
        maxwidthsigma(p)= 3.5
    end


    upperbound= [maxa,maxwidthsigma,maxx,maxy];

%create equation to be fitted - need  function "create_equation.m"

    created_equation = create_equation(size(start_points_temp));

    ft = fittype(created_equation,'numindep', 2);
    disp(ft)
    s = fitoptions('Method','NonlinearLeastSquares',...
                   'Startpoint',startpoints,...
                   'MaxIter',1000000,...
                   'MaxFunEvals',1000000,...
                   'TolX',(10^-10),...
                   'TolFun',(10^-10),...
                   'Lower',lowerbound,...
                   'Upper', upperbound,...
                   'Robust','LAR');

    [f,gof] = fit([xout,yout],zout, ft,s);


    zeval = feval(f,xout,yout);
    N = length(data_raw);
    cols = floor(length(zeval)/N);
    fitteddata = reshape(zeval(1:N*cols),[N,cols]);



        coeffvals = coeffvalues(f);

        cols = floor(length(coeffvals)/(length(coeffvals)/4));
%changed all 5s to 4s
        coeffvals_matrix =
reshape(coeffvals(1:length(coeffvals)/4*cols),[length(coeffvals)/4,col
s]); %changed all 5s to 4s




save('fitoutput','coeffvals_matrix','f','gof','fitteddata','data_raw')
;
```

This code is reliant on the function “create_equation.m” which is provided below:

```
function z = create_equation(num_gauss)

     z =  'a1*exp(-(x-x1).^2/(2*sigmax1.^2)-(y-
y1).^2/(2*sigmax1.^2))';
     if num_gauss > 1
         for i = 2:num_gauss
            z =  strcat(z,'+','a',num2str(i),'*exp(-(x-
x',num2str(i),').^2/(2*sigmax',num2str(i),'.^2)-(y-
y',num2str(i),').^2/(2*sigmax',num2str(i),'.^2))');
         end
     end

end
```