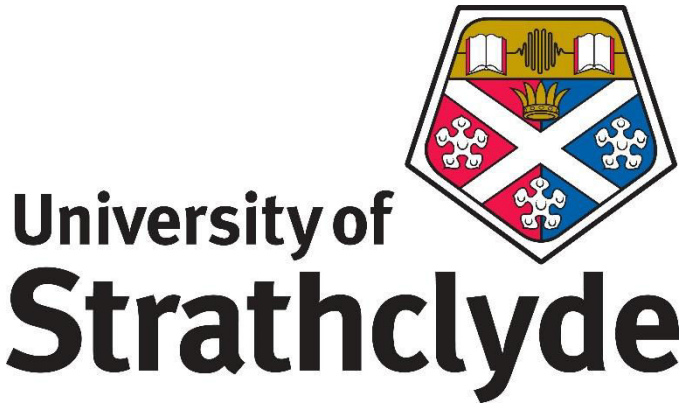
The Use of Terahertz Spectroscopy for Biomolecular Analysis



Thomas Harwood

Strathclyde Institute of Pharmacy & Biomedical Sciences

University of Strathclyde

This thesis is submitted for the degree of

Doctor of Philosophy

SIPBS 2016

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

Thomas Harwood

2016

Acknowledgements

There are many people I would like to thank for their help and support throughout my PhD. None more so than the late Dr Elizabeth Ellis, who was not only my PhD supervisor, but also my biochemistry lecturer as an undergraduate student. Her knowledge and passion helped me to understand as well as enjoy biochemistry, which ultimately played a part in my decision to undertake a PhD. I would also like to thank Dr Ellis for providing me with the opportunity to work on this project and having the confidence in me to generate my own ideas and work independently. Unfortunately Dr. Ellis was not able to see me complete my PhD, however, I hope she would be proud of the work we produced and I feel privileged to be her final student.

Additionally, I would like to acknowledge the other members of the Ellis lab, as well as the late Dr. Eve Lutz, for providing help and advice on experiments and making the long hours in University more enjoyable. I also want to express thanks to Prof. Gail McConnell for her role as acting supervisor at Strathclyde; organising my viva and ensuring I was prepared.

This work would not have been possible without the help of Dr. Adrian Lapthorn and Prof. Klaas Wynne at the University of Glasgow who saved my project when it was hanging in the balance. They not only supplied the equipment to carry out the measurements, allowed me access to their labs and resources, but also dedicated their time to help me through the latter stages of my PhD. I cannot begin to express how grateful I am for this. I specifically need to recognize Adrian for his help with the Beryllium experiments - which I think he instantly regretted. However, we can now look back and laugh…nervously.

Furthermore, I would also like to thank all the members (past and present) of the Ultrafast Chemical Physics group at the University of Glasgow for the discussions over countless cups of strong coffee, which helped me to understand the theory behind the laser systems used in this project. More so, it was refreshing to have others who

knew the problems that come with OKE measurements and working with lasers in general. It was good to vent our frustrations together and laugh (instead of cry) over difficult experiments!

One of the most enjoyable aspects of studying at SIPBS was the social aspect and I would like to acknowledge all the friends I have made over the years. The SIPBS parties, Friday pub lunches and nights out helped to keep us sane and we always had a lot of laughs.

Finally, I would never have made it this far without the continual support of my whole family, who, I'm sure, will be overjoyed that they can eventually stop asking, 'How much longer?'. In particular, my Mum, Dad and sister, who never stopped believing in me and provided both emotional and financial support to push me through until the end - I couldn't have done it without you. No doubt I have caused them undue stress over the last few years…but I'm sure they will take pride in the fact that they can eventually say their son and brother is a Doctor!

"A person starts to live when he can live outside himself."

- Albert Einstein

Contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

List of Tables

Abbreviations

AMP	Adenosine Monophosphate
ATP	Adenosine Triphosphate
ATR	Attenuated total reflectance
BSA	Bovine Serum Albumin
BWO	Backward Wave Oscillator
cAMP	Cyclic Adenosine Monophosphate
CAP	Catabolite Activator Protein
CD	Circular Dichroism
cGMP	Cyclic Guanosine Monophosphate
CMP	Cytidine Monophosphate
CRP	cAMP Receptor Protein
CW	Continuous Wave
Da	Daltons
DFT	Density Functional Theory
DHQD	Dehydroquinate Dehydratase
DLS	Depolarised Light Scattering
DNA	Deoxyribonucleic Acid
dNTP	Deoxynucleotide triphosphate
DSC	Differential Scanning Calorimetery
DTGS	deuterated tri-glycine
eV	Electron Volt
FAF	Fatty Acid Free

FIR	Far Infrared
FTIR	Fourier Transform Infrared Spectroscopy
FWHM	Full Width at Half Maximum
GaAs	Gallium Arsenide
GaP	Gallium Phosphide
GaP	Gallium Phosphide
GaSe	Gallium Selenide
GMP	Guanosine Monophosphate
GSH	Reduced Glutathione
GSSG	Oxidised Glutathione
GTP	Guanosine Triphosphate
HAS	Human Serum Albumin
HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HK	Histidine Kinase
INS	Inelastic Neutron Scattering
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IR	Infrared
K_D	Dissociation Constant
Lalba	α-lactalbumin
LB	Luria-Bertani
LINAC	Linear Partical Accelerator
M	Molar
MCPs	Methyl-accepting Chemotaxis Proteins
MG	Molten Globule
MOPS	3-(N-morpholino)propanesulfonic acid
MSG	Monosodium Glutamate
NAD	Nicotinamide adenine dinucleotide
NADPH	Nicotinamide adenine dinucleotide phosphate

NAG_3	N,N',N''-triacetylchitotriose
NC	Nitocellulose
NMP	Nucleotide Monophophate
NMR	Nuclear Magnetic Resonance
OD	Optical Density
OKE	Optical Kerr Effect
PC	Photoconductive
PDB	Protein Data Bank
PMR	Proton Magnetic Resonance
PSI	Pounds per Square Inch
PVDF	Polyvinylidene fluoride
qPCR	Quantitative Polymerase Chain Reaction
r.h.	Relative Humidity
RMS	Root Mean Square
RMSF	Root Mean Square Fluctuations
RNA	Ribonucleic Acid
RNAP	RNA polymerase holoenzyme
Rpm	Revolutions per Minute
SDL	Source Development Lab
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SS	Single Stranded
TB	Terrific Broth
THz	Terahertz
THz-TDS	Terahertz Time-Domain Spectroscopy
T_M	Melting Temperature
TMP	Thymidine Monophosphate
Tris	Tris(hydroxymethyl)aminomethane
UV	Ultraviolet

VDoS	Vibrational Density of States
ZeTe	Zinc Telluride

Abstract

Terahertz radiation occupies part of the electromagnetic spectrum between the mid-infrared and microwave bands (approximately 0.1 – 20 THz). Within this range, spectroscopic analysis of biological molecules can detect low-frequency vibrational modes that arise from intermolecular as well as intramolecular interactions and can provide significant information regarding structure and function. For example, the spectra of small biomolecules is dominated by individual modes and hydrogen bonds that can produce a distinctive spectral 'fingerprint', whereas the spectra of macromolecules, such as protein and DNA, can reveal information regarding conformation and dynamics. However, terahertz spectroscopy has habitually been underutilised as an investigative tool in biochemistry due to lack of terahertz sources and the strong absorbance of terahertz radiation by liquid water. The aim of this study was to bridge the gap between terahertz spectroscopy and biomolecular analysis by utilising three techniques to investigate a range of biomolecules of varying complexity. The bulk of this work employed terahertz Optical Kerr Effect (OKE) Spectroscopy as a novel means to study biomolecules in solution - This Raman based technique has the advantage that it does not suffer the same water absorption as conventional terahertz spectroscopies. We analysed a variety of small biomolecules, such as amino acids and nucleotides, and identified a variety of terahertz modes that are influenced by structure and interactions with the solvent. Additionally, we have identified spectral features in proteins molecules that are indicative of conformation and also identified novel vibrational modes that mediate protein-ligand binding in solution. Analysis of DNA and its components has also lead us to identify a multitude of previously undetected modes, some of which are attributed to phonons in the double helix and also hydrogen bonded interactions.

We also utilised two common methods of terahertz spectroscopy - Terahertz Time Domain Spectroscopy (THz-TDS) and Fourier Transform Infrared (FTIR) Spectroscopy - to create the first comprehensive database of small biomolecules in the terahertz range and also developed a novel membrane method for the simple and rapid analysis of biomolecules.

Chapter 1: Introduction

Chapter 1 Introduction

1.1 Terahertz Radiation

Terahertz radiation occupies part of the electromagnetic spectrum between the mid-infrared and microwave bands (Fig 1.1). There are numerous proposals of where the terahertz boundaries should be defined, however, it is reasonable to classify terahertz radiation as being in the region of 0.1 THz – 20 THz (wavelength (λ) 3 mm-15 μm or wavenumbers (ν) 3 cm^{-1} – 670 cm^{-1}). The terahertz region does not overlap with the visible spectrum and therefore cannot be seen by the naked eye, yet its warmth can be felt, similar to infrared radiation. Naturally occurring terahertz radiation is abundant and emitted by all objects above approximately 10 K, as part of black body radiation (Lee, 2009). The waves emitted from such sources are extremely weak, which is one of the reasons why the terahertz region is still the least investigated part of the electromagnetic spectrum. In addition, there have been technical difficulties in designing effective terahertz sources and detectors which led to this part of the spectrum being described as the 'terahertz gap'. However improvements in technology and significant research effort over the past 20 years has begun to bridge this gap.

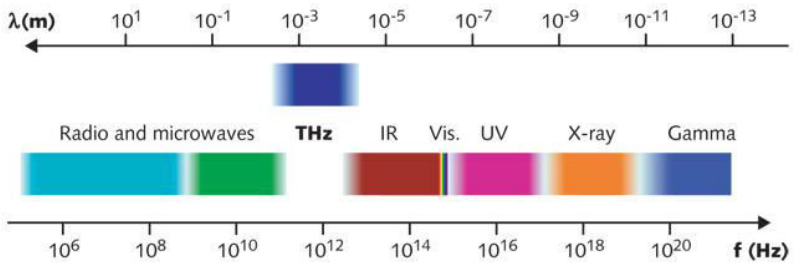


Figure 1.1: Electromagnetic Spectrum. *1 THz corresponds to: Period = 1 picosecond; Wavelength = 300 μm; Wavenumber = 33.3 cm^{-1}; Photon energy = 4.14 meV*

As terahertz borders microwave and infrared radiation in the electromagnetic spectrum, this results in terahertz waves sharing some useful properties with these frequency bands. For example, similar to microwave radiation, terahertz has the ability to pass through a number of materials including paper, wood, plastics and leather (Watanabe *et al.,* 2003), making it an attractive approach for security measures, such as scanning envelopes, packages and luggage at post offices or airports. In addition, like infrared, terahertz radiation is non-invasive and non-ionising making it useful for other security applications such as body scanners and also explosive detection (Foltynowicz *et al.,* 2006). The major shortcoming of terahertz radiation is due to its attenuation by water. This feature limits terahertz radiation's practicality in some areas: For example, long distance communications are almost impossible due to attenuation by atmospheric water. Nevertheless, terahertz radiation can be used at short distances of around 25 meters (Lee *et al.,* 2006) as a method for standoff detection in the military or as a tool for imaging concealed structures in the construction industry (Abina *et al.,* 2015)**.** In some cases, the capacity of water to absorb terahertz radiation can also be used as an advantage, especially in medical imaging where the difference in density and water content of tissues can be probed. However, due to the low penetrability of terahertz radiation into the skin, current uses are limited to burns diagnostics (Mittleman *et al.,* 1999) and epithelial cancers (Wallace *et al.,* 2004).

The ever-expanding number of applications being discovered for terahertz radiation has no doubt stimulated interest in this frequency range and research into producing and detecting terahertz radiation has increased vastly in recent years. Advances in optical design, brighter light sources and more sensitive detectors has allowed the field of submillimetre wavelength radiation to progress at a pace.

1.2 Terahertz Sources

The ability to generate low frequency terahertz radiation (0.1-3 THz) with sufficient power has traditionally been a major drawback in the development of terahertz technology. However, over the past two decades researchers have produced a number of terahertz sources including, continuous wave (CW) emitters, such as backward wave oscillators (BWOs). BWOs produce tuneable terahertz radiation utilising electrons and an applied DC voltage to transfer energy to an electromagnetic wave. The velocity of the electrons determines the frequency of the electromagnetic wave and therefore, by altering the DC voltage applied across the system, is it possible to tune the device over a range of frequencies (Lee, 2009). However, BWOs only have a spectral range covering tens of GHz (Kozlov & Volkov, 1998) and it would require multiple BWOs to cover the terahertz range. Other, less common, terahertz sources include P-type germanium laser (Bergner *et al.,* 2005) and terahertz gas lasers (Chantry, 1984), which have proved effective at generating high output powers in the terahertz frequency region. However, it was the team at the Source Development Lab (SDL) of Brookhaven National Laboratory that managed to generate the strongest terahertz radiation in the form of 0.3-ps, single-cycle pulse using a linear particle accelerator (LINAC). To generate pulsed terahertz radiation in this manner, electron bunches that are comparable to the terahertz wavelength are emitted into the LINAC and driven to a relativistic speed, followed by rapid deceleration to produce bremsstrahlung, or ‘braking radiation’. Alternatively, a magnetic field can be used to bend the electron bunch into a circular motion, producing ‘synchrotron radiation’.

Using this technique the team at SDL have generated terahertz radiation with pulse energy up to 100 μJoules (Shen *et al.,* 2007). Other studies have shown extraordinary ways of generating terahertz radiation, such as out of 'thin air' by using high-power laser pulses interacting with a photogenerated plasma (Hamster *et al.,* 1993) and another study has even shown that terahertz radiation can be produced when unpeeling adhesive tape (Horvat *et al.*, 2009).

1.3 Terahertz Spectroscopy

Few areas of terahertz research have received as much attention as 'Terahertz spectroscopy' due to the wealth of intriguing interactions that occur between radiation in this frequency range and matter in physical, chemical and biological systems. Terahertz radiation corresponds to processes occurring on a picosecond timescale and photon energies in the millielectronvolt (meV) range, meaning it interacts strongly with systems that have characteristic lifetimes and energetic transitions within this domain. Some of which include: weakly bound molecular crystals (Walther *et al.,* 2003; Allis *et al.,* 2007), phonons in crystalline solids (Schall *et al.,* 2001), relaxation dynamics in liquids (Rønne *et al.,* 1997; Arikawa *et al.,* 2008) and biological systems (He *et al.,* 2008; Markelz *et al.,* 2007). Therefore, it is not surprising that a spectroscopic tool with such remarkable scope has gained much attention and is now invaluable in many areas of research.

Traditionally, Fourier Transform infrared (FTIR) spectroscopy was usually the method of choice when it came to measuring samples above approximately 3 THz, due to the availability of commercial FTIR systems and their ability to accommodate and measure solid, liquid and gaseous samples with relative ease. Yet, below 3 THz the weak spectral intensity of common FTIR light sources does not lend itself to low frequency measurements. Arc lamps in combination with high sensitivity incoherent detectors, such as helium cooled bolometers, had to be employed to gain an adequate

signal to noise ratio to overcome the thermal effects of background radiation. However, other means of obtaining low frequency terahertz spectra are now available and decades of research has refined these techniques to produce high quality data, in a relatively simple manner. In addition, increased interest in terahertz spectroscopy has resulted in commercial terahertz spectrometers becoming available in recent years, allowing terahertz measurements to become accessible to a wider market. The background and theory of the terahertz systems used in this thesis are detailed below.

1.4 Terahertz Time Domain Spectroscopy

Advances in technology and optics have allowed numerous terahertz spectroscopy systems to be conceived and developed. However, none have made such a significant impact as the advent of Terahertz Time Domain Spectroscopy (THz-TDS), which initiated the dawn of a new era in Terahertz technology. Since the first pioneering work of Auston is the the 80's (Auston *et al.,* 1984) and subsequent work by Grischkowsky and co-workers in the early 90's (Katzenellenbogen & Grischkowsky, 1991) THz-TDS has grown at an astonishing rate, with academia, industry and also the military applying this technique to a variety of applications and samples, with varying degrees of success. The reason for THz-TDS's popularity is thanks to its relative simplicity in combination with room temperature operation, good bandwidth to cover terahertz frequencies (approximately 0.1-4 THz) and detection that allows reasonable signal to noise ratio while measuring both the amplitude and phase of the transient electric field - whereas standard techniques only measure amplitude. The amplitude and phase are directly related to the absorption coefficient and refractive index, which makes it possible to obtain these values with simple data analysis, giving THz-TDS a distinct advantage over conventional far-IR sources and detectors.

With few exceptions, almost all THz-TDS systems employ the same basic set-up, with variations on the method of terahertz generation and/or detection. Nevertheless, the

basic principles remain the same. Figure 1.2 shows a THz-TDS set-up at the most basic level - A femtosecond laser, centred at 800 nm, is used to generate ultrashort pulses of light, which is then split into the 'Pump Pulse' and 'Probe Pulse'. The pump pulse generally maintains around 90% of the light's energy and is used to generate the broadband terahertz pulse by interacting with an emitter. At the same time, the probe pulse is passed through a variable time delay, which then meets the terahertz pulse at the detector. As the terahertz pulse passes through a sample, there is a change in the electric field strength, which is then measured as an electrical signal when the terahertz pulse and probe pulse meet in the detector. By varying the time-delay of the shorter probe pulse, it is possible to scan the whole terahertz pulse and generate its electric field profile as a function of time. A Fourier transform can then be applied to the data to obtain a spectrum in the frequency-domain.

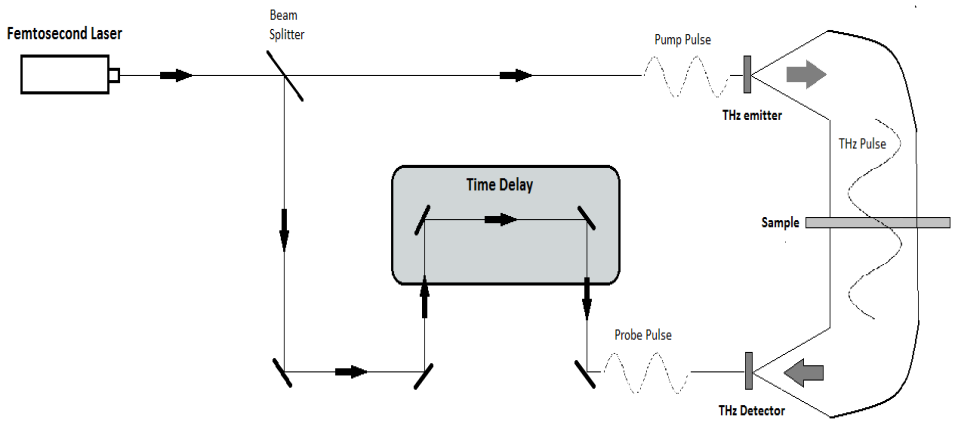


Figure 1.2: Schematic diagram of a typical THz-TDS set-up.

1.4.1 Generation and Detection

1.4.1.1 Photoconductive Antenna

The methods of generation and detection vary between systems. However, the two most common methods employed make use of photoconductive (PC) antennas or nonlinear crystals. A PC antenna consists of two metal electrodes deposited on a

semiconductor substrate and exploits the increased conductivity of semiconductors when they are exposed to light. To generate terahertz pulses, a electrical potential is applied between the electrodes, which accelerates free carriers (electrons-hole pairs) that have been generated by photons from the femtosecond laser hitting the semiconductor substrate (Fig. 1.3). The laser must produce photons with sufficiently large energy to overcome the bandgap of the semiconductor, which can be produced from a variety of materials, but gallium arsenide (GaAs) is most commonly used. When the bandgap is successfully overcome, broadband terahertz pulses are produced by the photo-induced current, arising from the acceleration and decay of the free carriers. There are a number of variations of PC antennas but the most commonly used is the 'Hertzian Dipole Antenna', which generates emissions of about 2 or 3 THz (Grischkowsky & Katzenellenbogen, 1991).

Photoconductive detection is similar to generation, however the bias electric field across the antenna is generated by the electric field of the terahertz pulse focused onto the antenna and not an external source. Therefore, by measuring the current generated across the antenna, the terahertz field strength and can obtained.

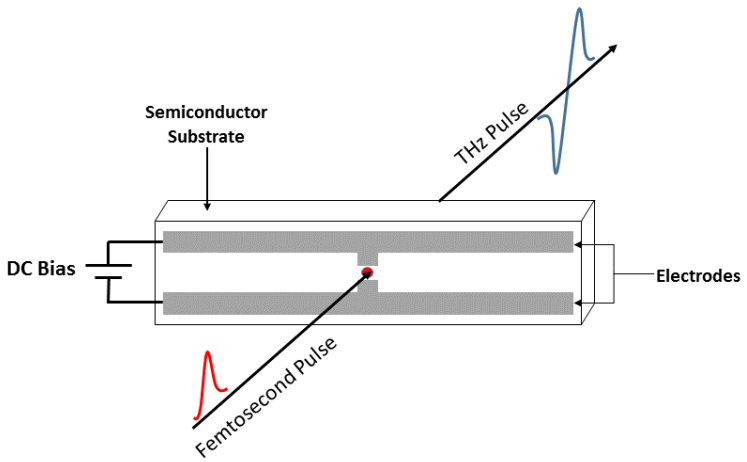


Figure 1.3: Photoconductive Antenna. *Terahertz pulse emission from a PC antenna excited by a femtosecond laser pulse*

1.4.1.2 Nonlinear Crystals

Optical rectification is a second-order nonlinear optical effect commonly used to generate terahertz radiation in nonlinear crystals. In linear optical processes, the electric dipole induced by the movements of electrons is proportional to the amplitude of the applied wave. However, if high intensity light, such as a femtosecond laser pulse, is applied to a material it can give rise to nonlinear optical phenomena, which induce nonlinear electron displacements from equilibrium. The media used are typically nonlinear crystals such as gallium phosphide (GaP), gallium selenide (GaSe) and most commonly zinc telluride (ZnTe). ZnTe has no inversion symmetry and Te has a higher electronegativity than Zn. This means that the electrons in the chemical bonds have a higher attraction towards Te, giving rise to asymmetric potential energy along the chemical bond, which increases nonlinear optical effects. These properties, along with being transparent, make these crystals ideal for the generation of Terahertz waves by optical rectification. Although the power produced by this method is lower than that produced using photoconductive antennas, its spectral content is generally broader (Bonvalet *et al.,* 1995).

A technique called ‘free space electro-optic sampling’ utilises nonlinear crystals as a means of terahertz detection, by measuring the field amplitude, which is proportional to the birefringence caused by the terahertz radiation interacting with the crystal. The terahertz electric field can therefore be detected by utilising a probe pulse and some common optics. In brief, ZnTe’s birefringence is changed by the terahertz pulse and is proportional to the strength of the electric field, which in turn changes the probe pulse from linearly to elliptically polarised. The probe beam is passed through a quarter-wave plate and a Wollaston prism to separate it into its horizontal and vertical polarised components before being received by a balanced detector (Fig. 1.4).

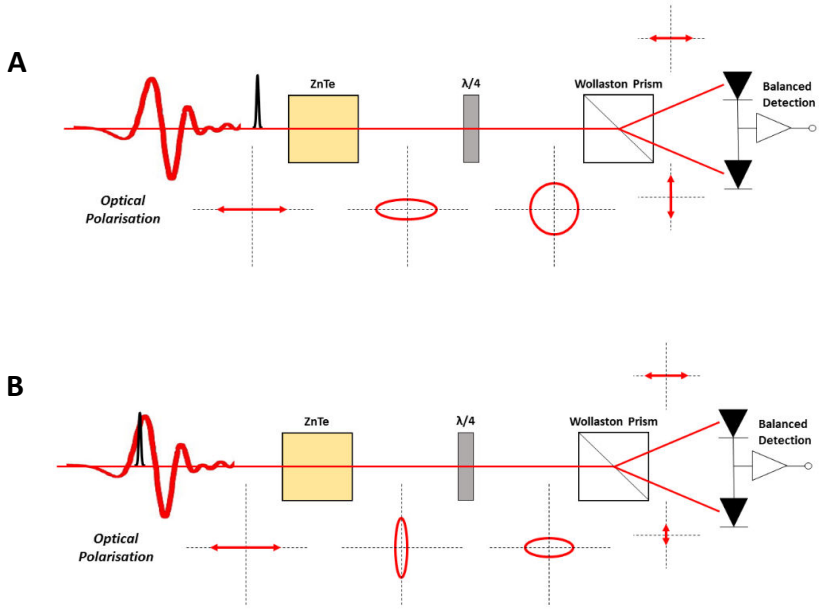


Figure 1.4: Electro-optic sampling used in THz-TDS. *(A) When there is a delay between the probe pulse and terahertz pulse the ZnTe crystal's birefringence will not change. The quarter-wave plate is set at such an angle that the resultant pulse will be circularly polarised and therefore have equal vertical and horizontal components, resulting in no signal. (B) The probe and terahertz pulses overlap in the ZnTe crystal, which results in a change in birefringence. The probe pulse becomes elliptically polarised and a signal is detected.*

1.5 Time-Resolved Spectroscopy

Ultrafast spectroscopy is one of the most popular and diverse methods to study the dynamics and interactions between molecules, with numerous different techniques being employed to measure a diversity of samples in areas of biology, chemistry, and physics. Popular ultrafast techniques include photon echo spectroscopy (Kurnit *et al.,* 1964; de Boeij *et al.,* 1998), fluorescence spectroscopy (Valeur, 2001; Lakowicz, 1999) and most commonly, pump-probe spectroscopy (Vanamerongen &

Vangrondelle, 1995; Henriksen & Engel, 2001). Pump-probe spectroscopy is the simplest experimental technique used to study ultrafast electronic dynamics. In this technique, an ultrashort laser pulse is split into two portions; a stronger beam (pump) is used to excite the sample, generating a non-equilibrium state, and a weaker beam (probe) is used to monitor the pump-induced changes in the optical constants (such as reflectivity or transmission) of the sample. Measuring the changes in the optical constants as a function of time delay between the arrival of pump and probe pulses yields information about the relaxation of electronic states in the sample.

1.6 Optical Kerr Effect Spectroscopy

One of the oldest and best-established methods for studying the ultrafast dynamics of liquids is Optical Kerr Effect (OKE) Spectroscopy due to its relative simplicity and high signal-to-noise ratio. Furthermore, OKE spectroscopy can provide a great deal of information on the liquid as it not only measures the orientational diffusion of the sample but can also detect Raman-active intramolecular and intermolecular modes. Therefore, it is not surprising that OKE spectroscopy has been utilised to improve our understanding of a number of simple liquids (Palese *et al.,* 1994) and is now being applied to ever more complex systems such as supercooled liquids (Torre *et al.,* 1998; Hinze *et al.,* 2000; Glorieux *et al.,* 2002), liquid crystals (Gottke *et al.,* 2002; Hunt & Meech, 2004), ionic liquids (Cang *et al.,* 2003; Shirota & Castner, 2005; Xiao *et al.,* 2006; Turton *et al.,* 2012) and most recently, as a tool to probe the dynamics of biological molecules (Turton *et al.,* 2014).

1.6.1 Origin of OKE Signal

Scottish physicist, John Kerr, first demonstrated the effect that bears his name in 1875 (Kerr, 1875). The so-called 'Kerr effect' becomes evident when an electric field is applied to an isotropic medium, inducing a change in birefringence, which is proportional to the square of the applied electric field, which is not to be confused with the similar Pockels effect, where a linear relationship is demonstrated between the electric field and induced birefringence. In the Kerr effect, the electric field interacts with the electric dipole moments of the molecules in the medium and causes them to become anisotropic, which leads to a difference in refractive indices parallel and perpendicular to the electric field.

The Optical Kerr Effect was first predicted by Buckingham (1956) and later observed by Mayer & Gires (1964) and Maker & Terhune (1964). The optical Kerr effect is only possible with intense light sources such as that from lasers, where a birefringence is induced using the light's own electric field. OKE spectroscopy is a third order non-linear spectroscopy and is possible using a pump-probe configuration. The origin of the OKE signal arises from a linearly polarised, intense, ultrafast laser pulse (pump pulse) interacting with an isotropic sample, which induces a dipole moment in the molecules, causing them to temporarily align with the polarisation of the pump pulse (Fig. 1.5). As the index of refraction of a liquid is determined by the polarisability of the molecules, the induced anisotropy results in a transient birefringence in the sample, which can then be monitored by the probe pulse in a time dependent manner. The probe pulse's plane of polarisation is set at a known angle to that of the pump pulse, so as the probe interacts with the birefringent medium, it will emerge from the sample elliptically polarised. In a typical configuration, the OKE signal is detected by placing a polariser in front of the detector at an angle perpendicular to the plane of polarisation of the probe pulse. Therefore, if birefringence is induced in the sample, a fraction of the pulse will reach the detector. By introducing a time-delay between the pump and probe pulses it is possible to measure the relaxation of the transient

birefringence of the sample, which reveals information about the microscopic dynamics.

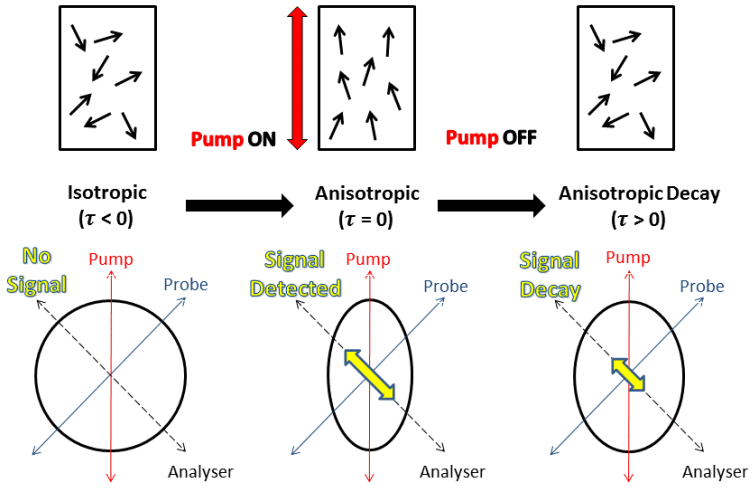


Figure 1.5: Principles of OKE spectroscopy. *The diagram illustrates the photoinduced birefringence and anisotropic decay that is measured using OKE spectroscopy.*

A large amount of information related to a liquid's dynamics can be extracted from the OKE spectrum: the intense pump pulse will induce a large, instantaneous electronic response originating from the distortion of the electron clouds within the sample and the molecules will also align to the electric field of the pump pulse followed by spontaneous rotational diffusion back to an isotropic equilibrium (Smith & Meech, 2002). The relaxation time of this process can be measured by monitoring the intensity of the probe beam in relation to the time delay of the initial excitation.

In an ultrafast OKE experiment, femtosecond laser pulses are used, which have a broadband optical spectrum and can measure over a large frequency range, typically several hundred wavenumbers. When these ultrafast laser pulses are employed, other processes have to be taken into account. For example, intermolecular interactions, including density fluctuations, librational motions and collisions between molecules can produce a transient anisotropy and contribute to the OKE signal. Intramolecular

resonances may also contribute to the OKE signal if they arise from Raman-active modes that are coherently excited and the frequencies, ω_1 and ω_2, lie within the bandwidth of the laser pulse (Fig. 1.6) (Kirkwood *et al.,* 2000).

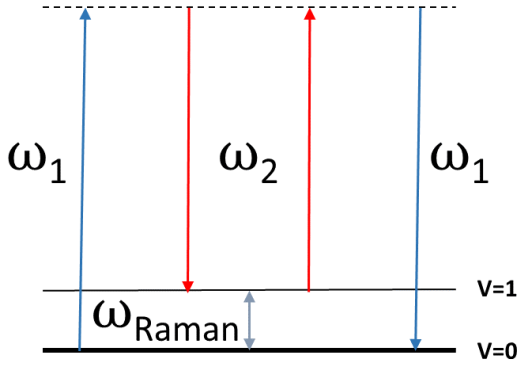


Figure 1.6: Energy level diagram for OKE*. ω_1 and ω_2 are contained in the spectrally broad pulse, which can excite Raman active modes, ω_{Raman}.*

1.7 Terahertz Spectroscopy of Biological Molecules

The understanding of conformation and flexibility of biomolecules is important for research, medicine and also the technological industry as the structure of protein and DNA can be indicative of disease state (Selkoe, 2004) and interactions between these molecules is commonly used in diagnostic tests (Hruschak, 1995). Biological molecules can often be characterised by their vibrational motions that occur at characteristic frequencies. Many of these vibrational modes occur in the picosecond range, which correspond to the time-scale occupied by terahertz radiation, making terahertz spectroscopy an attractive approach for the analysis of these molecules. Extensive analysis of biomolecules at infrared frequencies has been performed, which produce information regarding the motions of nuclei within the molecule (Mantsch & Chapman, 1996). At the near- and mid-infrared range, the bands correspond to stretching or bending motions of individual bonds in the molecule. However,

frequencies in the far-IR and terahertz range can analyse motions propagating over large sections of the molecular structure. Nevertheless, the size of the molecule has an influence on the spectra produced and hence the identification of features. For example, small biological molecules such as amino acids, nucleosides, and sugars tend to have easily detectable, isolated features. As the size of the molecule increases, the number of modes also increase and they overlap with increasing polymer length. When dealing with large, nonperiodic biological molecules such as DNA or proteins the spectra are much more difficult to interpret and may appear relatively featureless. All molecules possess intermolecular vibrational modes, which in the case of small biological molecules often causes the whole molecule to move in concert. In contrast, large biological molecules can also possess a large number of internal vibrational modes (intramolecular), with large groups of atoms moving within the molecule, requiring much more detailed and demanding analysis. Nevertheless, terahertz spectroscopy has shown promise as an effective method for studying DNA (Wittlin *et al.,* 1986), proteins (Markels *et al.,* 2000), cells (Liu *et al.,* 2006) and tissue (Mittleman *et al.,* 1999).

1.8 Aims

1.8.1 Optical Kerr Effect Spectroscopy as a Tool to Probe Biomolecular Dynamics

Terahertz measurements of biological samples have proved challenging due to the lack of powerful sources, detectors and the high attenuation of terahertz radiation by water. Successful measurements have been carried out using far-infrared and THz-TDS on crystalline biological samples (Korter *et al.,* 2006) or dry films (Laman *et al.,* 2008), however measurements on aqueous solutions have been largely unsuccessful. Water dynamics and the solvation layer surrounding biological macromolecules are essential for the correct structure and function of these molecules. Therefore, analysis

of these molecules in the absence of an aqueous environment calls into question the biological relevance of the detected modes. Elastic neutron scattering (Lerbret *et al.,* 2013; Roh *et al.,* 2006) and spontaneous Raman scattering (Genzel *et al.,* 1976; Urabe *et al.,* 1998) have provided some insight into the terahertz modes of biomolecules in aqueous solution, yet these techniques become unreliable at low frequencies (< 1 THz) due to the strong Rayleigh peak. OKE spectroscopy does not suffer from this drawback and can provide high quality data over a wide frequency range (0.3 – 20 THz). Furthermore, OKE spectroscopy has already been successfully applied to the analysis of protein solutions (Giraud *et al.,* 2003) and has shown promise in detection and differentiation of protein structural motifs (Hunt *et al.,* 2006). Therefore, in this thesis, we expand upon initial findings by applying OKE spectroscopy to probe the structural changes and dynamics of a variety of protein molecules undergoing conformational change. In addition, for the first time, we have applied OKE spectroscopy as a tool to measure the vibrational modes within different DNA molecules and monitored the difference between hybridised and denatured states.

1.8.2 Terahertz and Far Infrared Database of Small Biomolecules

Some of the earliest FTIR and THz-TDS studies on biological molecules were carried out on amino acids and nucleotides in the solid state, as these molecules represent protein and DNA at their fundamental level, respectively (Fischer *et al.,* 2002). Although previous studies have analysed these small biomolecules in detail, this topic has not been revisited in a number of years and there remain gaps in the literature were the far-IR and terahertz spectra of these molecules are absent. Furthermore, depending on the environment and physical state, the properties of these molecules will differ and influence the low frequency spectra. For example, in a charged medium, protonated or deprotonated forms of amino acids (zwitterions) occur, which alter the hydrogen bonding between these molecules and also surrounding water (Degtyarenko *et al.,* 2007). These changes are likely to influence the spectral features in the terahertz range as this region is most affected by intermolecular interactions. We therefore

hypothesis that protonation state should induce significant changes. In addition, amino acid zwitterions in a complex environment, such as a protein, heavily influence interactions and are important in many enzymatic reactions (Barth, 2000). Therefore, understanding the low frequency vibrations of amino acids and their zwitterionic forms is imperative for understanding protein reactivity and function.

Part of this thesis will focus on obtaining a complete database for the far infrared and terahertz spectra of small biomolecules. By obtaining the low frequency spectra of all 20 proteinogenic amino acids in different protonated states, we hope to elucidate how these individual subunits may contribute to the low frequency spectra of proteins. Similarly, we also aim to measure nucleotide components at varying levels of complexity in order to determine what structure and interactions are responsible for the overall DNA spectrum. Furthermore, normal mode and DFT calculations have proved successful at predicting the low frequency vibrations of biomolecules (Fischer *et al.,* 2002; Brandt *et al.,* 2008) and assigning spectral features. By generating a comprehensive low frequency database of small biomolecules we hope to provide the experimental data with which future theoretical work can be compared.

Chapter 2: Materials & Methods

Chapter 2 Materials & Methods

2.1 Materials

2.1.1 Bacterial Strains

E.coli BL21 (DE3):	F^- ompT gal dcm lon $hsdS_B(r_B^- m_B^-)$ λ(DE3)
E.coli BL21 (DE3) pLysS:	F^- ompT gal dcm lon $hsdS_B(r_B^- m_B^-)$ λ(DE3) pLysS(cm^R)
E.coli DH5α:	F^- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80d*lac*ZΔM15 Δ(*lac*ZYA-*arg*F)U169, hsdR17($r_K^- m_K^-$), λ–
E.coli NM522:	F^- endA+ $lacI^q$ (lacZ)M15 $proA^+$ B^+ /supE thi (lac-proAB) (hsdMS-(mcrB)5 (r_k^- mk^- McrBC$^-$)

2.1.2 *E. coli* Growth Media

2.1.2.1 Luria-Bertani (LB) Broth

0.5% (w/v) Yeast Extract, 0.5% (w/v) NaCl and 1% (w/v) tryptone, in distilled water and sterilised by autoclaving (Sambrook & Russell, 2001)

2.1.2.2 Luria-Bertani (LB) Agar

0.5% (w/v) Yeast Extract, 0.5% (w/v) NaCl, 1% (w/v) tryptone and 2% (w/v) agar in distilled water and sterilised by autoclaving (Sambrook & Russell, 2001).

2.1.2.3 Terrific Broth (TB)

2.4% (w/v) Yeast extract, 1.2% (w/v) Tryptone and 0.4% (v/v) Glycerol. Sterilise by autoclaving before adding 10% (v/v) sterile potassium phosphate solution (170 mM KH_2PO_4, 720 mM K_2HPO_4) (Sambrook & Russell, 2001).

2.1.3 Chemicals

Chemicals were bought from Sigma-Aldrich unless otherwise stated.

2.1.4 Antibiotics

Antibiotics were used for selective growth of *E. coli* and added to pre-sterilised growth media.

Ampicillin and kanamycin stocks were made at concentrations of 100 mg/ml and 30 mg/ml, respectively, in distilled water. Stock solutions were filter sterilised through 0.22 µl pore membranes and diluted to the appropriate concentration in growth media.

2.1.5 Isopropyl β-D-1-thiogalactopyranoside (IPTG)

Stock solutions of IPTG were made at a concentration of 1 M and sterilised through a 0.22 µm filter. Working concentrations were diluted 1:1000 from the stock as appropriate.

2.2 Methods

2.2.1 DNA Methods

2.2.1.1 General

2.2.1.1.1 DNA Analysis

Prior to all measurements, DNA samples were held at 95°C for 5 mins before being placed on ice. The sample was then slowly allowed to reach room temperature to promote specific base pairing.

2.2.1.1.2 Determination of DNA Concentration

The spectophotometer, Nanodrop^{TM}, was used to measure the concentration of nucleic acids. The sensitivity range for this equipment is between 2 ng and 3700 ng per 1 µl of solution (Labtech International, 2006). DNA has an absorption peak at 260 nm where an OD of 1 corresponds to approximately 50 µg/ml of double stranded DNA, 40 µg/ml for single-stranded DNA and RNA and 20 µg/ml for oligonucleotides. The purity of the DNA is estimated based on the A_{260}/A_{280} ratio, with non-proteinaceous preparations having a ratio between 1.8-2.0. The Nanodrop was first blanked using a single drop of dH_2O (1-5 µl) before analysing samples. Each sample was added to the detection area without introducing any air bubbles and concentration and wavelength recorded.

2.2.1.1.3 Restriction Digests

Restriction enzymes (New England Biolabs) are used to cut single and double stranded DNA at specific sites. They are used to modify and manipulate DNA for a number of biochemical tests. The appropriate enzyme was incubated with corresponding buffer, according to manufacturer's instructions, at 37°C for 1 hour and then digested fragments were separated using agarose gel electrophoresis.

2.2.1.1.4 Plasmid DNA Extraction

Plasmid DNA was extracted using mini- (Promega), midi- (Qiagen), maxi- (Sigma) or megaprep (Sigma) kits according to manufacturer's instructions.

2.2.1.1.5 DNA Precipitation

DNA was precipitated and resuspended in a smaller volume in order to concentrate the sample. 1/10 volume of sodium acetate (3 M, pH 5.2) was added to the sample along with 2.5-3.0 volumes of absolute ethanol and incubated on ice overnight, in a cold room. The mixture was centrifuged at >14,000 x g for 30 mins at 4°C and supernatant discarded. The pellet was rinsed numerous times with 70% Ethanol in order to minimise contamination and then centrifuged for 15 mins at >14,000 x g. Supernatant was discarded and the pellet allowed to air dry before re-suspending in dH_2O at desired concentration (Sambrook & Russell, 2001).

2.2.1.2 Agarose Gel Electrophoresis

Gel electrophoresis is a technique used to identify DNA or RNA fragments based on their size as the nucleic acids are separated by an applied electric field that moves the negatively charged nucleotides through the agarose gel. The bands can then be visualised and compared to known molecular weight standards.

2.2.1.2.1 Stock Solutions

2.2.1.2.1.1 TAE Buffer

Stock TAE buffer (50x) was made with 242 g Tris base, 57.1 ml glacial acetic acid and 100 ml of 0.5 M EDTA pH 8.0 then made up to 1 litre with distilled water. TAE buffer was used at 1x concentration and made by diluting 20 ml of 50x TAE buffer to 1 L in distilled water.

2.2.1.2.2 Protocol

1% w/v gels were made by dissolving agarose in 1x TAE at high temperature (in a microwave). Once the gel cooled to ~50°C GelRed Nucleic Acid Stain (Biotium) was added at a final concentration of 0.5 µg/ml and the gel poured into the gel caster with the comb inserted to form the wells. The gel was allowed to set for a minimum of 30 mins before the tank was filled with 1x TAE buffer and the comb removed. 6x DNA Loading dye (Biolabs) was added to the samples before loading and 1 kb or 100 bp ladders (Biolabs) where utilised alongside the samples to estimate the size of the fragments. All gels where run at 100V for 1 hour then the gel removed from the tank to allow the DNA bands to be visualised using UV radiation.

2.2.2 Protein Methods

2.2.2.1 General

2.2.2.1.1 Protein Dialysis

Slow buffer exchange was carried out using dialysis. The protein sample was placed inside dialysis tubing with the appropriate molecular weight cut off and clamped shut. The sample was placed into dialysis buffer 20 times the volume of the sample and stored at 4 °C overnight with slow mixing. During dialysis the buffer was replaced 2-4 times for stringent exchange.

2.2.2.1.2 Protein Concentration

Protein solutions were concentrated using Amicon Ultra-15 Centrifugal filter units with the appropriate molecular weight cut-off. The concentrator unit was first washed with buffer only, to remove any trace chemicals from the filter. The protein solutions were gradually added to the unit and spun in a centrifuge between 1000-5000 rpm, at room temperature, for no longer than 5 minutes. After each spin the solution was gently inverted to redistribute the protein throughout the solution and minimise the

formation of a gradient, which can lead to precipitation. More sample was added to the unit and the process repeated until the desired concentration was reached.

2.2.2.1.3 Bradford Assay

Protein standards were made with bovine serum albumin (BSA) at concentrations of 1, 2, 5, 10, 15 and 20 µg/ml in 800 µl distilled water and then brought up to 1 ml with 200 µl Coomassie Brilliant Blue G-250 and mixed well. The absorbance of the standards was measured at 595 nm and plotted as a standard curve against concentration. To determine the concentration of the protein solution, 5 µl of the protein was brought up to 800 µl with distilled water and 200 µl of Coomassie Brilliant Blue G-250 added. The absorbance was measured at 595 nm and the concentration of the unknown sample was determined from the standard curve.

2.2.2.2 Calculation of the Proportion of Protein-Ligand binding

If protein P binds to ligand L, then an equilibrium is formed:

$$P + L \rightleftharpoons PL$$

Where, P and L represent the free protein and ligand, respectively, and PL is the protein-ligand complex. For protein-ligand binding measurements we used protein to ligand ratios advised in the literature and confirmed full activation using dissociation constants (K_D) and experimental methods. The K_D can be defined as:

$$K_D = \frac{[P][L]}{[PL]} \quad (1)$$

As P and L represent *free* protein and ligand, respectively, *total* protein (P_0) and *total* ligand (L_0) can be calculated by:

$$P_0 = P + PL \quad (2)$$

$$L_0 = L + PL \quad (3)$$

Substitute *(2)* and *(3)* into equation *(1)*:

$$K_D = \frac{([P_0]-[PL])([L_0]-[PL])}{PL} \quad (4)$$

Which can be rearranged to the quadratic formula ($ax^2 + bx + c = 0$):

$$0 = [PL]^2 - ([P_0] + [L_0] + K_D)[PL] + [P_0][L_0]$$

Therefore:

$$[PL] = \frac{([P_0]+[L_0]+K_D) \pm \sqrt{([P_0]+[L_0]+K_D)^2 - 4[P_0][L_0]}}{2} \quad (5)$$

From equation *(5)* we calculated all proteins to be >95% bound by their ligands at the concentrations used in our measurements.

2.2.2.3 Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) is a technique used to separate proteins according to molecular weight. SDS is an anionic detergent that binds to denatured protein in a defined stoichiometry and β-mercaptoethanol reduces any disulfide bonds, thus aiding the denaturation process. The sample must be heated to 85°C in the presence of both reagents to fully prepare the sample. SDS gives an overall negative charge to the protein, proportional to the unit mass, so the migration of protein through the polyacrylamide matrix is solely determined by the molecular weight (Laemmle, 1970). Therefore, smaller proteins will travel further along the gel than larger proteins.

2.2.2.3.1 Solutions

2.2.2.3.1.1 Resolving Buffer (4x)

90.8 g of Tris base and 2 g of SDS was made up to 450 ml with distilled water and pH adjusted to 8.8 with concentrated HCl. The solution was then made up to 500 ml with distilled water and filtered through Whatman # 1 filter paper.

2.2.2.3.1.2 Stacking Buffer (4x)

15.14 g Tris base and 1 g SDS was made up to 200 ml with distilled water and pH adjusted to 6.8 with concentrated HCl. The solution was made up to 250 ml with distilled water and filtered through Whatman # 1 filter paper.

2.2.2.3.1.3 Running Buffer (10x)

The 10x running buffer stock was made using 30.28 g Tris base, 144 g glycine and 10 g SDS and made up to 1 litre with distilled water. 1x running buffer is used to run the gel and is prepared by diluting the 10x stock in distilled water.

2.2.2.3.1.4 Laemmli's Sample Buffer (2x)

3.13 ml of 1 M Tris-Cl (pH 6.8), 2 g SDS, 9 ml glycerol, 5 ml β-mercaptoethanol and 1 ml bromophenol blue (0.1%,v/v) was made up to 50 ml with distilled water.

2.2.2.3.1.5 Staining Solution

0.1% (w/v) Coomassie brilliant blue R250, 40% (v/v) ethanol and 10% (v/v) acetic acid.

2.2.2.3.1.6 De-staining Solution

40% (v/v) ethanol and 10% (v/v) acetic acid.

2.2.2.3.2 Protocol

2.2.2.3.2.1 Casting the Resolving Gel

Different concentrations of resolving gel were cast according to Table 2.1, which allows visualisation of proteins of varying molecular weights.

Table 2.1: SDS-PAGE Resolving Gel Recipe. *Higher percentage gels result in slower migration speed, hence, are better for resolving small MW proteins.*

	10 %	**12 %**	**15 %**
30% Acrylamide/Bis-acrylamide	6.7 ml	8 ml	10 ml
4 x Resolving buffer	5 ml	5 ml	5 ml
dH_2O	8.3 ml	7 ml	5 ml
Ammonium persulfate (100 mg/ml)	100 µl	100 µl	100 µl
TEMED	10 µl	10 µl	10 µl

Two glass plates were thoroughly cleaned with ethanol and distilled water before being clamped together with a rubber spacer to form a watertight seal. The resolving solution was transferred between the plates to polymerise and a thin layer of isopropanol added to remove any bubbles. After polymerisation (30-60 mins), the isopropanol was washed away with distilled water and any residual water removed.

2.2.2.3.2.2 Casting the Stacking Gel (5%)

1.64 ml 30% Acrylamide/bisacrylamide solution, 2.5 ml 4x stacking Buffer, 5.86 ml dH_2O, 60 µl ammonium persulphate (100 mg/ml) and 10 µl TEMED were mixed and poured on top of the resolving gel. A comb was inserted to create the wells and the solution allowed to polymerise. After polymerisation, the combs were removed and the plates and gel placed into an ATTO gel apparatus (AE6450, ATTO Corporation, Japan) and the tank filled with 1x Running buffer.

2.2.2.3.2.3 Sample Preparation and Loading

2x LSB was added to the sample in a 1:1 ratio before boiling at 100°C for 5 minutes. Samples were briefly centrifuged before loading into the wells alongside a known protein marker in order to estimate molecular weights. The gel was run at 125V and 200 mA for 1 hour.

2.2.2.3.2.4 Protein Visualisation

After running, the gels were removed from the electrophoresis apparatus, transferred into a dish containing staining solution and placed on a shaker unit with constant agitation until the gel was fully stained. The gel was removed from the stain and washed in water before being transferred into a dish containing de-staining solution. This was left until the protein bands became visible and the background stain was minimised.

2.2.2.4 Protein Overexpression

Plasmid constructs include a *T7* viral promoter, which allows inducible expression of the protein subcloned downstream. This mechanism exploits elements from the bacteriophage T7 and the *lac* operon to control expression of the T7 RNA polymerase (engineered into the *e. coli* genome) by addition of isopropyl β-D-1-thiogalactopyranoside (IPTG). IPTG is a non-hydrolysable analogue of lactose, which triggers transcription of the *lac* operon producing T7 RNA polymerase, this binds specifically to the T7 promotor on the plasmid DNA and transcribes the gene at a very high level. The gene of interest is therefore expressed at a very high level and in most cases produces a large quantity of the protein of interest.

2.2.2.4.1 General Methods

2.2.2.4.1.1 Preparing Competent Cells

E. coli cells were cultured on LB agar overnight at 37 °C, then a single cell was inoculated from the plate into LB broth and again incubated overnight at 37 °C with shaking (180 rmp). 1 ml of overnight culture was used to inoculate 50 ml of pre-warmed LB broth and incubated at 37 °C, 180 rpm until the optical density at 600 nm (OD_{600}) reached 0.4. The culture was then placed on ice for 15 minutes before being centrifuged at 4°C for 10 mins at 3,500 rpm and the supernatant discarded. The cell pellet was resuspended in 75 mM calcium chloride and incubated on ice for 20 minutes before centrifugation (4°C, 10 minutes at 3,500 rpm) and supernatant removed. The pellet was resuspended in 2 ml TFB2 (10 mM MOPS pH 7.0, 75 mM calcium chloride, 10 mM rubidium chloride and 15 % glycerol), then split into 200 µl aliquots and snap frozen in liquid nitrogen. The competent cells were stored at -80 °C.

2.2.2.4.1.2 Transformation

Transformation of competent *E.coli* cells with plasmid DNA was by the heat shock method. Plasmid DNA was added to 100 µl of ice cold TMC (10 mM Tris, pH 7.5, 10 mM MgCl, 10 mM CaCl) before mixing with 200 µl of competent cells and incubated on ice for 30 mins. Cells were then heat shocked for 90secs at 42°C and 1 ml of LB broth added to the sample and placed at 37°C for 2 hour to allow cells to recover and express the plasmid. The sample was centrifuged for 2 mins at 14,000xg and the cells resuspended in the residual supernatant. The cell suspension was then spread onto an LB agar plate containing the appropriate antibiotic and let to incubate at 37 °C overnight

2.2.2.4.2 CheY

The CheY expression vector, pKC1, (a kind gift from Prof Robert Bourret, University of North Carolina) was created by cloning the CheY gene into the *NdeI* and *BamHI* sites of the pET28a plasmid, producing a N-terminal His-tag and thrombin cleavage site.

2.2.2.4.2.1 Protein Induction and Cell Lysis

BL21 strains transformed with the CheY plasmid were grown from -80 °C glycerol stocks on LB agar plates supplied by kanamycin (30 µg/ml). An isolated colony was carefully picked to grow in 5 ml LB starter culture with kanamycin (30 µg/ml) in a 37 °C shaker overnight. The culture was transferred to a large flask containing 500 ml LB broth containing kanamycin (30 µg/ml) at 37°C in a shaking incubator and samples were taken every hour to check the growth by OD600 against LB media broth. When the OD_{600} = 0.4, expression was induced by adding IPTG (isopropyl β-D-1-thiogalactopyranoside) to the culture at a final concentration of 0.5 mM. The culture was transferred to room temperature and left for a further 8-10 hours in a

shaking incubator. Cells were harvested by centrifugation (5,000 x g for 10 mins at 4°C) and the pellet was retained and kept at -80°C until needed.

Cell pellets were re-suspended in 50 ml lysis buffer (50 mM Hepes, pH 7, 300 mM NaCl, 10 mM Imidazole) supplemented with protease inhibitor tablets (cOmplete ULTRA, EDTA free EASYpack, Sigma) and lysed by standard methods. All cellular debris was removed by centrifuging the sample twice at 60,000 x g for 30 mins at 4°C.

2.2.2.4.2.2 Nickel Affinity Purification

The supernatant from the cell pellet was loaded onto a 5 ml His-column pre-equilibrated with lysis buffer, at a flow rate of 0.33 ml/min. The column was washed with 10-20 column volumes of wash buffer (50 mM Hepes, pH 7, 300 mM NaCl, 20 mM Imidazole) to remove non-specific proteins before eluting CheY with an increasing gradient of elution buffer (50 mM Hepes, pH 7, 300 mM Nacl, 150 mM imidazole). Eluted fractions (1 ml volume) were analysed by absorbance at 280 nm to locate the CheY protein. The CheY containing fractions were pooled and incubated with 2 units of human thrombin per mg CheY in order to cleave the His-tag.

2.2.2.4.2.3 Size-exclusion Chromatography

Size-exclusion chromatography was employed to remove thrombin, the His-tag and also exchange CheY into storage buffer (50 mM Hepes, pH 7, 100 mM NaCl, 2 mM DTT and 10% (v/v) glycerol). The size-exclusion column (HiPrep 16/60 Sephacryl S-100 HR, GE healthcare) was first equilibrated with storage buffer and then a concentrated CheY solution was loaded at a flow rate of 0.5 ml/min. The final protein concentration was determined spectrophotometrically at 278 nm using a molecular extinction coefficient of 8,480 M^{-1} cm^{-1} and purity was determined by SDS-PAGE. CheY was used immediately or aliquots were flash frozen in liquid nitrogen and stored at -80°C.

2.2.2.4.3 Catabolite Activator Protein (CAP)

The CAP expression vector, pAKCRP-His6, (a kind gift from Prof. Charalampos Kalodimos, Rutgers University) was created from plasmid pAKCRP by use of site-directed mutagenesis to insert six His codons (CAC-CAC-CAC-CAC-CAC-CAC) after codon 209 of the cap gene. The His-tagged cap gene was cloned into the NdeI and HindIII sites of the pET21a plasmid (Kapanidis *et al.,* 2001).

2.2.2.4.3.1 Protein Induction and Cell Lysis

E. coli BL21 strains transformed with the CAP plasmid were grown from -80 °C glycerol stocks on LB agar plates supplied with 100 μg/ml ampicillin. An isolated colony was carefully picked to grow in 5 ml LB starter culture with ampicillin (100 μg/ml) in a 37 °C shaker overnight. The culture was transferred to a large flask containing 500 ml LB broth containing ampicillin (100 μg/ml) at 37 °C in a shaking incubator and samples were taken every hour to check the growth by OD600 against LB media broth. When the OD600 = 0.4, expression was induced by adding IPTG (isopropyl β-D-1-thiogalactopyranoside) to the culture at a final concentration of 1 mM and allowed to grow for a further 3 hours in a shaking incubator. Cells were harvested by centrifugation (5,000 x g for 10 mins at 4°C) and the pellet was retained and kept at -80 °C until needed.

Cell pellets were resuspended in 50 ml lysis buffer (20 mM Tris pH 7.9, 500 mM NaCl, 5 mM Imidazole and 1 mM β-mercaptoethanol) supplemented with protease inhibitor tablets (cOmplete ULTRA, EDTA free EASYpack, Sigma) and lysed by standard methods. All cellular debris was removed by centrifuging the sample twice at 60,000 x g for 30 mins at 4 °C.

2.2.2.4.3.2 Nickel Affinity Purification

The supernatant from the cell pellet was loaded onto a 5 ml His-column pre-equilibrated with lysis buffer, at a flow rate of 0.33 ml/min. The column was washed with 10-20 column volumes of wash buffer (20 mM Tris pH 7.9 , 500 mM NaCl, 20 mM Imidazole and 1 mM β-mercaptoethanol) to remove non-specific proteins before eluting CAP with an increasing gradient of elution buffer (20 mM Tris pH 7.9, 500 mM NaCl, 200 mM imidazole and 1 mM β-mercaptoethanol). Eluted fractions (1 ml volume) were analysed by absorbance at 280 nm to locate the CAP protein.

2.2.2.4.3.3 Size-Exclusion Chromatography

Gel filtration was employed to remove any additional impurities and imidazole from the buffer. The size-exclusion column (HiPrep 16/60 Sephacryl S-100 HR, GE healthcare) was first equilibrated with CAP buffer (20 mM Tris pH 7.9, 500 mM NaCl, and 1 mM β-mercaptoethanol) and then a concentrated CAP solution was loaded at a flow rate of 0.5 ml/min. The final protein concentration was determined spectrophotometrically at 278 nm using a molecular extinction coefficient of 19,940 M^{-1} cm^{-1} and purity was determined by SDS-PAGE.

2.2.3 OKE Measurements

2.2.3.1 Set-up

The components and experimental layout used for OKE spectroscopy vary little between laboratories. Like most ultrafast, pump/probe spectroscopies the source of the ultrafast pulses is usually a mode-locked Ti-sapphire laser and the remainder of the apparatus is composed of easily accessible optics and equipment. Our experimental approach for analysing biological molecules did not require any particular modifications to perform the measurements and the OKE setup is essentially standard: A Ti:sapphire laser (Coherent Micra) provided ~10 nJ pulses with a wavelength centred at 800 nm and a repetition rate of 82 MHz. The ultrafast pulse first passes through a Homosil prism pair compressor to pre-compensate for group velocity dispersion that causes the pulse to broaden as it travels through the optical path (Kafka & Baer; Reid & Wynne, 2000). We also equalised the optical material in each beam path to minimise dispersion and the pulse reaching the sample had a full-width at half maximum (FWHM) duration of ~21 fs. From the compressor, the beam is sent to a beam splitter, which reflects a small percentage of the light to be used as a probe beam and the majority of the light is transmitted as the intense pump beam, 10 % and 90 % respectively in our experimental set up. Either the pump or probe beam can be passed through an optical delay line (pump in our set-up), which is set on a translational stage with sub-fs resolution to control the pump-probe delay time in the sample. The probe beam is then passed through a half wave plate and a polariser that is set to pass light at 45 ° to the vertical, while the pump is passed through a polariser set to pass vertically polarised light. Both beams were co-focused onto the same spot of the sample by a 10 cm focal length achromatic lens before the pump beam is sent to a beam dump and the probe beam recollimated by a matching lens and sent through a quarter wave plate, Wollaston prism and a balanced photodiode detector (optical-heterodyne detection), which gives this system unparalleled signal-to-noise and dynamic range (Giraud et al., 2003). The output of

the detector is measured by a lock-in amplifier that is referenced to the difference of the pump and probe beams being mechanically chopped at rates of ~3 kHz in a ratio of 5:7 (Fig. 2.1). By using dual choppers to control the signal intensity we minimise spurious scattered pump light and hence improve signal to noise ratio. The frequency-domain OKE signal was derived from the time-domain measurement through a Fourier transform using a LabView program written by Dr David Turton. Solvent-free frequency domain spectra were obtained by manually subtracting the solvent spectrum from the sample spectrum after matching the spectra on water librations (>8 THz), as this band is scarcely affected by concentration and thermal variations (Sassi *et al.,* 2013).

OKE measurements were carried out in 1 mm or 2 mm path length Suprasil® quartz cuvettes. The cuvettes were extensively cleaned and dried in a vacuum chamber, to remove any residual liquid, before depositing the sample at the bottom of the cuvette using a syringe and needle. The cuvette was held in a close-fitting copper block and the temperature controlled by a pair of attached low-voltage electrical heaters (± 0.1 °C).

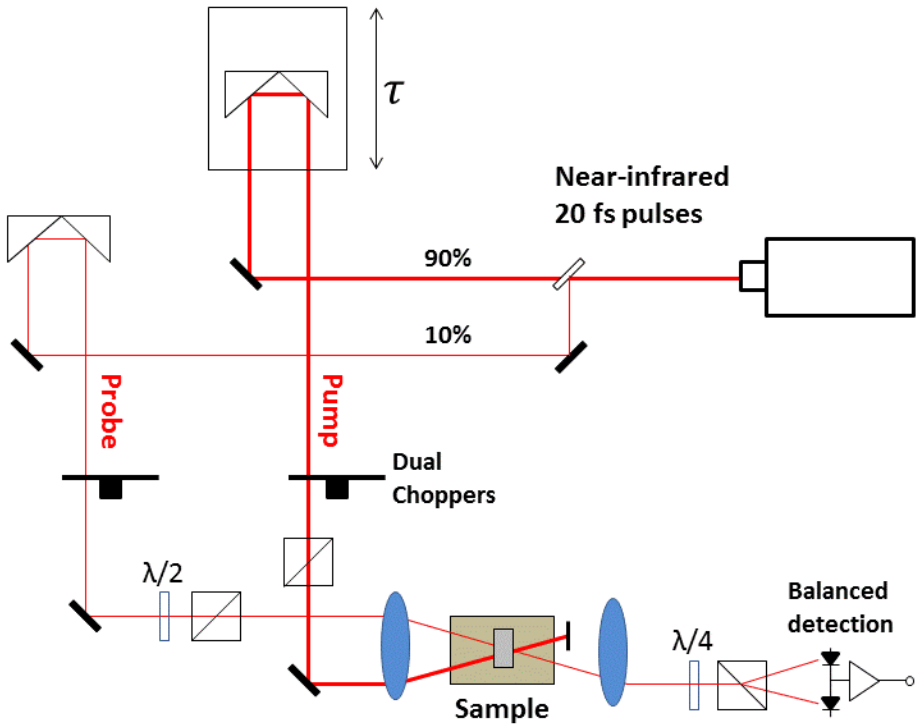


Figure 2.1: Schematic of the OKE system used for our biomolecular measurements.

2.2.3.2 Protein Preparation

Lyophilised proteins where dissolved in the appropriate buffer and then dialysed extensively with the same buffer in order to remove any impurities or chemical traces from the sample. Due to scattering effects, it was necessary to filter each sample (0.22 μm pore size) prior to measurements.

For comparisons of lyophilised proteins (Chapter 4) all proteins were measured in the appropriate buffer at a concentration of 250 mg/ml.

BSA – 100 mM sodium phosphate, pH 7.

Lysozyme – 100 mM sodium acetate, pH 5.4.

α-lactalbumin – 100 mM Tris, pH 8.

2.2.3.2.1 BSA Denaturation

The concentration was dropped to 100 mg/ml for temperature induced denaturation measurements of BSA as it was not possible to obtain reliable CD spectra at higher concentrations due to saturation of the detector. Therefore, both OKE and CD measurements were performed at this concentration for direct comparison.

2.2.3.2.2 Lalba Denaturation

Apo-lalba denaturation was measured under the same conditions as above (250 mg/ml in 100 mM Tris, pH 8). For holo-lalba denaturation, the stabilising calcium ion was included in the buffer by adding 20 mM calcium chloride to give a protein to calcium molar ratio of 1:1.

2.2.3.2.3 Phosphorylation of CheY

CheY OKE measurements were performed in 100 mM Tris (pH 7.4), 300 mM NaCl at a protein concentration of 80 mg/ml (~6 mM). NMR measurements were identical except 1 M deuterated Tris in D_2O solution (Sigma) was diluted to 100 mM dTris. Measurements were therefore carried out in 100 mM dTris pH 7.4, 300 mM NaCl, 10% D_2O.

The CheY cofactor, Mg^{2+}, was added to the sample either by dialysis or dissolving $MgCl_2$ into a dilute CheY sample at a protein to $MgCl_2$ molar ratio of 1:5 to ensure saturation. The sample was then concentrated to the appropriate concentration for OKE or NMR measurements using an Amicon Ultra filtration unit. If $MgCl_2$ was added directly to a concentrated CheY solution, precipitation would occur.

The formation of BeF_x in solution depends on the ratio of beryllium to free fluoride ions and only BeF_3^- is known to activate CheY (Cho *et al.,* 2000). Beryllium chloride and sodium fluoride were therefore added to CheY in the following molar ratios to form the active complex:

CheY : $BeCl_2$: NaF

1 : 4 : 25

Due to the toxicity of beryllium, $BeCl_2$ was added directly to the concentrated sample, followed by NaF and left to equilibrate at room temperature overnight.

2.2.3.2.4 Allosteric Activation of CAP

CAP measurements were performed in 50 mM sodium phosphate, pH 6, 500 mM KCl and 1 mM β-mercaptoethanol at a protein concentration of 15 mg/ml (300 µM). cAMP was added to dilute CAP in a protein to ligand molar ratio of 2:3 and then concentrated using an Amicon Ultra filtration unit.

2.2.3.2.5 Lysozyme-Inhibitor Complex

Lysozyme measurements were performed in 100 mM (pH 5.4) sodium acetate at a protein concentration of 250 mg/ml (~15 mM). At this pH, the dimerisation that occurs at higher pH (Sophianopoulos *et al.,* 1969) is avoided. Due to the stability of lysozyme, it was possible to add the inhibitor directly to the concentrated sample.

2.2.4 THz-TDS Measurements

2.2.4.1 The Experimental Apparatus Set-up

The methods of terahertz generation and detection for THz-TDS varies between laboratories. For our experiments, we employed a photoconductive antenna for terahertz generation and free space electro optic sampling for detection.

A Ti-sapphire laser (Coherent Legend) provided 2.5 mJ pulses with a wavelength centred at 800 nm and a repetition rate of 1 kHz. The pulse was first split into the pump (90%) and probe (10%) before the pump beam was passed through an optical delay line, which is set on a translational stage with sub-fs resolution to control the pump-probe delay time. Subsequently, the pump pulse was passed through a negative lens, collimated, and then focused through the photoconductive antenna to generate the terahertz pulse into the sample area. The terahertz pulse leaves the sample having experienced a change in electric field intensity and the beam is collimated by a parabolic mirror before a second parabolic mirror focuses the terahertz wave onto the ZnTe crystal. The probe pulse travels an optical path of equal distance to the pump and meets the terahertz pulse at the ZnTe crystal before passing through a quarter wave plate, Wollaston prism and a balanced photodiode detector to measure the terahertz signal (electro-optic sampling). The output of the detector is measured by a boxcar integrator that is referenced to the chopper at a rate of 500 Hz to minimise noise (Fig. 2.2). The frequency-domain signal is derived from the time-domain measurement through a Fourier transform using a LabView program written by Dr David Turton.

All membrane measurements were performed by orientating the custom holder (Section 2.2.5.2) so the beam passed through the appropriate well. Other wells were covered with aluminium foil, to attenuate terahertz radiation, ensuring the correct sample was measured. For pellet measurements, a standard FTIR pellet sample holder was adapted to take terahertz measurements.

For all measurements, we first recorded the signal of a corresponding blank before measuring the sample. The absorbance (A) was calculated using:

$$A = -\log_{10}\frac{I_0}{I}$$

Where I_0 is the incident intensity and I is the transmitted intensity.

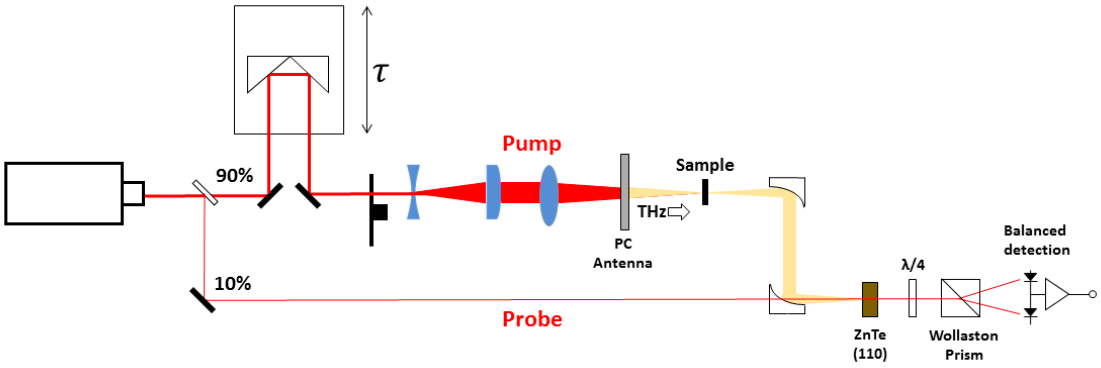


Figure 2.2: Schematic of the THz-TDS system used for our biomolecular measurements

2.2.4.2 Membrane Measurements

A sample holder was designed and manufactured to prepare and measure membrane samples. The holder consists of two austenite (316) stainless steel plates, drilled with 10 mm and 20 mm diameter holes for the sample wells. Corresponding holes were cut into two silicone rubber sheets (1.5 mm) and placed between the steel plates. To prepare a sample, a nitrocellulose membrane was sandwiched between the silicone sheets and screwed tight, using four M3 tapped holes in the steel plates, to create a watertight seal (Fig. 2.3). Prior to the samples, 200 µl of dH_2O was added to each well and dried in a desiccated atmosphere for 1 hour – this process is believed to equilibrate the membrane and avoid uneven distribution of the sample. For glutathione samples, 500 µl of a 250 mg/ml solution, in dH_2O, was pipetted into a 20 mm well and 500 µl of dH_2O only was pipetted into an adjacent well to be used as a blank. The membrane

and holder were then placed in a desiccated environment at room temperature overnight. To measure the terahertz signal, the membrane remained in the holder and was placed into the beam path.

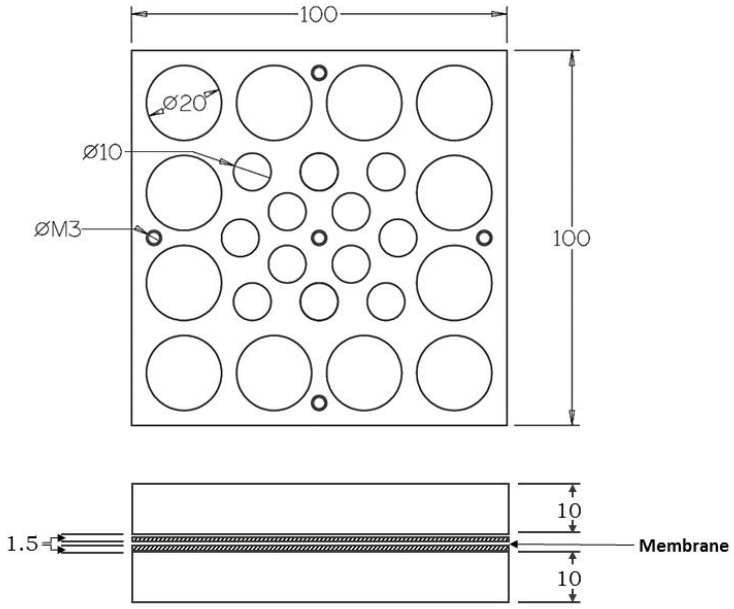


Figure 2.3: Membrane holder for THz-TDS measurements. *Plan view (top) and end elevation (bottom) of the holder. The hatched section of the end elevation represents the silicone sheets. All measurements are in millimetres.*

2.2.4.3 Polyethylene Pellets

Biological samples were mixed with high density polyethylene (Sigma) at ratios of 1:1 and 1:10 and subsequently ground with a mortar and pestle, ensuring particle size was smaller than the terahertz wavelengths to reduce scattering effects. Pellets where pressed in a 13 mm vacuum dye at minimum pressure (200 psi) for 10 minutes to minimise heating and remove moisture from the sample. Sample path length was dependent on mass and 1 mm and 3 mm pellets were pressed from 100 mg and 300 mg of mixture, respectively. Pure polyethylene pellets of 1 mm and 3 mm were also processed in the same way to be used as blanks.

2.2.5 Other Spectroscopies and Calorimetry

2.2.5.1 Fourier Transform Infrared Spectroscopy (FTIR)

Most organic molecules absorb light in the infrared region with the absorbance bands corresponding to specific bonds present in the molecule. FTIR covers an extremely broad frequency range and can therefore be used to analysis the properties of a substance over the entire infrared range. FTIR is more practical than older 'dispersive spectroscopy' techniques where monochromatic light was used to measure the sample at each wavelength, as FTIR utilises a broadband light source and Michelson interferometer to measure many infrared frequencies simultaneously. By first recording a background absorbance measurement, followed by a measurement with a sample in place, it is possible to obtain an absorbance spectrum that is unique to the molecule.

To analyse biological molecules a Bruker Vertex 70 FTIR spectrometer was utilised, which can produce radiation over a wide spectral range, from the terahertz (10 cm^{-1}) to visible (28,000 cm^{-1}). For far infrared and terahertz measurements, a globar or mercury arc lamp light source was utilised along with a silicon beam splitter and deuterated tri-glycine (DTGS) detector. Although this detector has the advantage of operating at room temperature, it has a low signal to noise ratio at low frequencies and we found it to become unreliable for our measurements below approximately 50-100 cm^{-1}. All measurements were performed using attenuated total reflectance (ATR), which utilises the property of total internal reflection to create an evanescent wave that propagates into the sample. Solid samples were placed directly onto the ATR diamond surface and screwed tight to ensure interaction between the sample and wave. Films were obtained by depositing 2-5 µl of a liquid samples directly onto the diamond surface and drying in a continuous flow of warm air. The instrument and sample chamber were purged with nitrogen prior to measurements to avoid contributions from H_2O and CO_2 and residual atmospheric absorbance was subtracted

by recording a background spectrum. Resolution was 4 cm^{-1} and 128 scans were accumulated for each spectrum. Spectra were smoothed with 5 points.

2.2.5.2 Circular Dichroism (CD) Spectroscopy

CD spectroscopy measures the difference in right- and left-handed circularly polarised light, which is influenced by chiral chromophores in the sample. CD is usually measured over a series of wavelengths in the ultraviolet (UV) region where it is useful for monitoring changes in protein structure. In the far-UV (<240 nm), the spectrum is principally due to absorption by peptide bonds and different types of secondary structure give rise to characteristic CD spectra in this range. For example, it is possible to estimate the α-helical content of proteins and peptides by monitoring the CD spectrum at 208 nm and 222 nm, which can be used as an indicator of protein denaturation (Krittanai & Johnson *et al.,* 1997). The tertiary structure of proteins can be monitored in the near-UV (260-320 nm) where the spectrum is sensitive to the arrangement of aromatic amino acids such as tryptophan, tyrosine and phenylalanine. Small variations in these residues can lead to differences in the CD spectrum that indicate a conformational change has occurred. Therefore, we utilised far-UV CD spectroscopy to monitor the thermal denaturation of BSA and near-UV spectroscopy to confirm the change in tertiary structure upon binding of an inhibitor to lysozyme.

All measurements were performed on a JASCO J-810 CD spectropolarimeter at the same concentrations and in identical buffers to our OKE measurements. Due to the high concentrations of protein used in these samples it was necessary to use demountable quarts cells (Hellma) with a short path length (10 μm) to overcome saturation of the detector. The demountable cells were first thoroughly cleaned with ethanol and distilled water and then dried before 2.5 μl of sample was deposited onto one of the cell windows and the other window carefully placed on top without the formation of bubbles. For measurements on lysozyme, the surface tension of the sample was adequate to hold the cell together but for BSA denaturation measurements the cell had to be sealed with electrical tape as evaporation became an issue at higher

temperatures. Native and inhibitor bound lysozyme samples were placed in the beam path and measured at room temperature. BSA thermal denaturation was measured from 5-80 °C (± 0.01 °C) in 5 °C increments using a Peltier device to control the temperature. The sample was held at each temperature for 5 minutes before proceeding with the measurement in order to allow the sample to equilibrate.

2.2.5.3 Nuclear Magnetic Resonance (NMR) Spectroscopy

NMR Spectroscopy is one of the only biophysical methods, which can provide high-resolution structures of biological molecules and their complexes at atomic resolution. NMR uses strong magnetic fields to probe the spin properties of atomic nuclei (*i.e.,* 1H, 13C or 15N), coupled with radio waves to promote transitions between nuclear energy levels. Two-dimensional NMR (2D NMR) is a common technique used to study protein structure and ligand binding as it produces information regarding each amino acid's location and also distances between residues. This leads to a specific 'fingerprint' for each protein that can be used to determine its structure.

CheY samples were prepared in 100 mM dTris pH 7.4, 300 mM NaCl, 10% D2O at 80 mg/ml in a total volume of 300 µL and placed within Shigemi 5 mm symmetrical microtubes matched for aqueous solutions. NMR spectra were acquired on a three channel Bruker AVANCE II+ NMR spectrometer operating at a proton resonance frequency of 600.13 MHz using a TBI [1H, 13C, Broadband] triple resonance probe head stabilised at a temperature of 298 K and equipped with an actively shielded gradient coil for delivering pulsed field-gradients. 1D 1H NMR spectra were acquired over a frequency width of 15 kHz (24.99 ppm) into 65K data points (acquisition time = 2.184 s) with 64 transients using an excitation sculpting routine to efficiently suppress the solvent resonance (Hwang & Shaka, 1995). 2D [1H, 15N] HSQC NMR spectra were acquired over an F2 (1H) frequency width of 9.6 kHz (16.022 ppm) into 2048 data points (acquisition time = 106.5 ms) and an F1 (15N) frequency width of 2.4 kHz (40 ppm) centred at an F1 offset of 120 ppm into 128 t1 increments for an acquisition time of 26.3 ms (F1 FID resolution = 38 Hz/pt). Data were typically

acquired with 384 transients per FID for a total acquisition time of 21 hours to provide good signal-to-noise data at natural abundance 15N levels. NMR spectra were acquired using a sensitivity improved echo/anti-echo pulse sequence according to Palmer *et al.,* 1991; Kay*, et al.* 1992 and Schleucher *et al.,* 1994.

All NMR data were acquired under TopSpin Version 2.1 running under Windows XP operating on a HP XW4300 workstation. All data were processed according to standard established Fourier transform protocols.

2.2.5.4 Differential Scanning Calorimetry (DSC)

DSC measures the heat effects associated with phase transitions and chemical reactions as a function of temperature by monitoring the difference in heat flow to the sample in comparison to a reference. DSC is commonly used to characterise the stability of biomolecules by measuring the heat change associated with the molecules' thermal denaturation (Privalov, 2015). Solutions of biomolecules can be measured with minimal preparation and used as a direct comparison to our OKE measurements. Therefore, we utilised DSC to determine the thermal denaturation temperatures of DNA oligonucleotides and α-lactalbumin in identical environments to our OKE samples (50 mg/ml of DNA in dH_2O and 250 mg/ml lalba in 100 mM (pH 8) Tris).

DSC measurements were carried out on a DSC Q100 (TA Instruments) in a similar manner for both DNA and protein samples. 5-10 µl of sample was loaded into an aluminium pan and hermetically sealed. The sample was loaded onto the heat pad alongside an empty, hermetically sealed, aluminium pan as a reference, then the sample chamber was closed and purged with nitrogen. Denaturation of the biomolecules was monitored between 30-95 °C at varying scanning speeds in order to determine accurate transition point.

2.2.5.5 LightCycler

Oligonucleotide melting temperatures (T_M) were determined using a Thermo Hybaid Express Thermal Cycler. SYBR green is pH sensitive and only fluoresces between pH 7.5 – 8. Therefore, the oligos were measured in a simple SYBR green buffer prepared by diluting SYBR Green I Nucleic Acid Gel Stain 10,000x (Thermo) to 1x in 50 mM (pH 7.6) sodium phosphate buffer. 20 µl of DNA sample was pipetted in a LightCycler capillary, centrifuged briefly, and placed into the LightCycler Sample Carousel. The fluorescence intensity was measured from 30-95 °C and the derivative plotted to determine the melting peak. Various scanning times were employed to determine accurate T_M.

Chapter 3: Protein Structure

Chapter 3 Protein Structure

Here, we analyse individual amino acids and small peptides at terahertz frequencies, in aqueous solution, utilising ultrafast OKE spectroscopy to elucidate how molecular size and structure influence the OKE spectrum. Amino acids can be hydrophilic or hydrophobic in nature and influence the dynamics of surrounding water molecules, which is an important aspect of maintaining protein structure (Bigelow, 1967). As OKE spectroscopy is sensitive to solvation dynamics, we aim to utilise this technique to analyse how the properties of different amino acids influence the solvation shell dynamics. Furthermore, by first analysing the fundamental units of protein molecules, we aim to establish the spectral profile of each amino acid and determine the contribution these molecules may give to an overall protein spectrum. Thereafter, we analyse three globular proteins - lysozyme, Bovine Serum Albumin (BSA) and α-lactalbumin (lalba) – under various conditions, in order to study differences in secondary and tertiary structure. Lysozyme and lalba have highly similar amino acid sequence and structure and are therefore useful to determine if it is possible to identify shared features in the OKE spectra, indicative of structural motifs. On the other hand, BSA is approximately 5 times larger and adopts a quite different tertiary structure, which can be used to establish if molecular weight and overall geometry largely influence the protein OKE spectrum. We also carry out temperature dependent measurements on BSA and lalba to determine if it is possible to follow thermally induced loss of ordered structure in the terahertz range using OKE spectroscopy.

A great deal of information regarding the dynamics of these biomolecules can be extracted from the OKE spectra: A typical MHz-THz OKE spectrum shows sharp intramolecular vibrations at high frequencies, similar to what can be detected in conventional FIR-FTIR spectroscopy. At lower frequencies librational modes can be seen, corresponding to the underdamped rocking motion of the molecules within their

solvent 'cage.' Below 1 THz β- and α-relaxations are detected and correspond to cage rattling motions and orientational diffusion, respectively (Fig. 3.1).

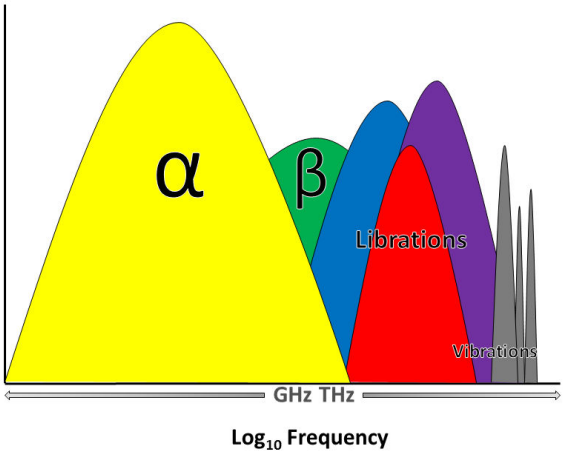


Figure 3.1: Typical OKE spectrum on a logarithmic frequency axis.

3.1 Aqueous measurements of small biomolecules in the terahertz range

Various terahertz studies have been carried out on amino acids in the solid state that have provided valuable information regarding the structure and interactions of these molecules in the crystalline phase. However, in order to simulate how these molecules interact in a biological environment requires aqueous measurements. The hydrophobicity of amino acid side chains are key to the correct folding and function of protein molecules and studies have shown that terahertz spectroscopy is sensitive to changes in the solvation dynamics of biomolecules (Heugen *et al.,* 2006; Born *et al.,* 2009; Whitmire *et al.,* 2003). Furthermore, it has been suggested that hydrophobic and hydrophilic groups change the hydrogen bonding dynamics in solution (Choudhury & Pettitt, 2005; Sterpone *et al.,* 2010; Pal *et al.,* 2002), yet it remains unclear whether the chemical properties of the molecules (*i.e,* charge, polarity *etc.*) or

steric hindrance plays the principal role in influencing solvation dynamics (Xu & Berne, 2001; Li *et al.,* 2007; Ball, 2008). Utilising Ultrafast OKE spectroscopy, we can probe the solvation dynamics of individual amino acids and elucidate how different side groups influence the hydrogen-bonding network of the aqueous solvent. In addition, by analysing the 20 common proteinogenic amino acids it may be possible to determine how each contributes to protein OKE spectra.

Due to the large difference in solubility of the amino acids and the low OKE signal often produced by biomolecules, we measured each amino acid at its saturation point in dH_2O to gain the best signal possible. All measurements were carried at the isoelectric point of the amino acids.

3.1.1 Water Spectrum

Water is usually the solvent of choice when studying biomolecules in order to recreate an *in vitro* environment. Water also displays unusual properties such as extensive hydrogen bonding and polarity, making it an interesting solvent to study in isolation. Consequently, the dynamics of pure water have been the subject of extensive analysis and the OKE spectrum is now reasonably well characterised (Fig. 3.2): two bands can readily be identified at ~1.8 THz (60 cm^{-1}) and ~5.4 THz (180 cm^{-1}) with the origin of the band at higher frequency largely being accepted as the longitudinal stretch of the hydrogen bond (Marti *et al.,* 1996). However, establishing the source of the band at ~1.8 THz has been the subject of much debate. It has been suggested that hydrogen bond bending of a O-O-O motif is responsible for this mode (Tsai & Wu, 2006) while, conversely, others have proposed that translational motions are responsible, irrespective of hydrogen bonding (Padro & Marti, 2003). In effect, it is likely that the combination of these factors influence this low frequency band.

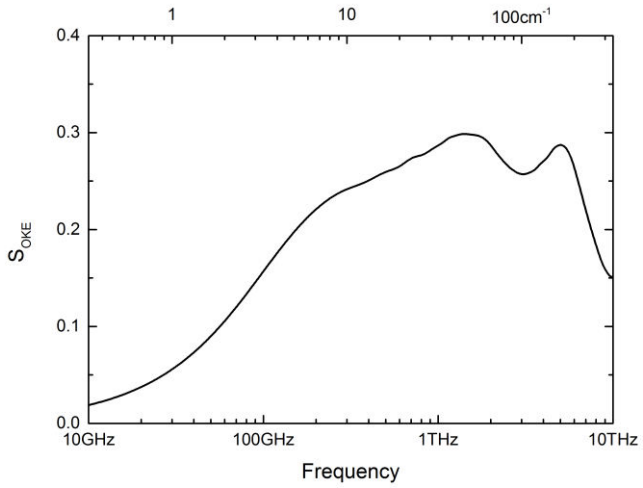


Figure 3.2: OKE spectra of pure water

3.1.2 Amino Acids & Peptides

We have documented the OKE spectra of all 20 proteinogenic amino acids and separated them into three categories: polar (hydrophilic) (Fig. 3.3), non-polar (hydrophobic) (Fig. 3.4) and charged (Fig. 3.5). However, many of these molecules do not produce a large OKE response due to limited solubility (Table. 3.1) or poor polarisability and their spectra do not change considerably from that of pure water. Most of the amino acids display the two common water modes and also an additional contribution between 1-3 THz, which originates from the molecule's librational motions. In the low frequency part of the spectrum we also detect the orientational diffusion, or α-relaxations, of the small biomolecules, which is why the amino acids display a greater amplitude than pure water in the frequency range. However, some of the amino acids measured show large changes in the OKE spectrum, compared to water, and we shall therefore concentrate on analysis of these molecules.

Table 3.1: Properties of amino acids

Symbol	Name	Mol. form	M_w	T_m (°C)	pK_a	pK_b	pK_c	pI	Solubility (mg/ml)
Ala	*L*-Alanine	$C_3H_7NO_2$	89.09	297	2.33	9.71	-	6.00	166.9
Arg	*L*-Arginine	$C_6H_{14}N_4O_2$	174.20	244	2.03	9.00	12.10	10.76	182.6
Asn	*L*-Asparagine	$C_4H_8N_2O_3$	132.12	235	2.16	8.73	-	5.41	25.1
Asp	*L*-Aspartic Acid	$C_4H_7NO_4$	133.10	270	1.95	9.66	3.71	2.77	5.04
Cys	*L*-Cysteine	$C_3H_7NO_2S$	121.16	240	1.91	10.28	8.14	5.07	250
Gln	*L*-Glutamine	$C_5H_{10}N_2O_3$	146.14	185	2.18	9.00	-	5.65	42.5
Glu	*L*-Glutamic Acid	$C_5H_9NO_4$	147.13	160	2.16	9.58	4.15	3.22	8.6
Gly	Glycine	$C_2H_5NO_2$	75.07	290	2.34	9.58	-	5.97	250.2
His	*L*-Histidine	$C_6H_9N_3O_2$	155.15	287	1.7	9.09	6.04	7.59	43.5
Ile	*L*-Isoleucine	$C_6H_{13}NO_2$	131.17	284	2.26	9.6	-	6.02	34.2
Leu	*L*-Leucine	$C_6H_{13}NO_2$	131.17	293	2.32	9.58	-	5.98	22.0
Lys	*L*-Lysine	$C_6H_{14}N_2O_2$	146.19	224	2.15	9.16	10.67	9.74	100
Met	*L*-Methionine	$C_5H_{11}NO_2S$	149.21	281	2.16	9.08	-	5.74	56
Phe	*L*-Phenylalanine	$C_9H_{11}NO_2$	165.19	283	2.18	9.09	-	5.48	27.9
Pro	*L*-Proline	$C_5H_9NO_2$	115.13	221	1.95	10.47	-	6.30	1622
Ser	*L*-Serine	$C_3H_7NO_3$	105.09	228	2.13	9.05	-	5.68	250
Thr	*L*-Threonine	$C_4H_9NO_3$	119.12	256	2.2	8.96	-	5.60	98.1
Trp	*L*-Tryptophan	$C_{11}H_{12}N_2O_2$	204.23	289	2.38	9.34	-	5.89	13.2
Tyr	*L*-Tyrosine	$C_9H_{11}NO_3$	181.19	343	2.24	9.04	10.10	5.66	0.46
Val	*L*-Valine	$C_5H_{11}NO_2$	117.15	315	2.27	9.52	-	5.98	88

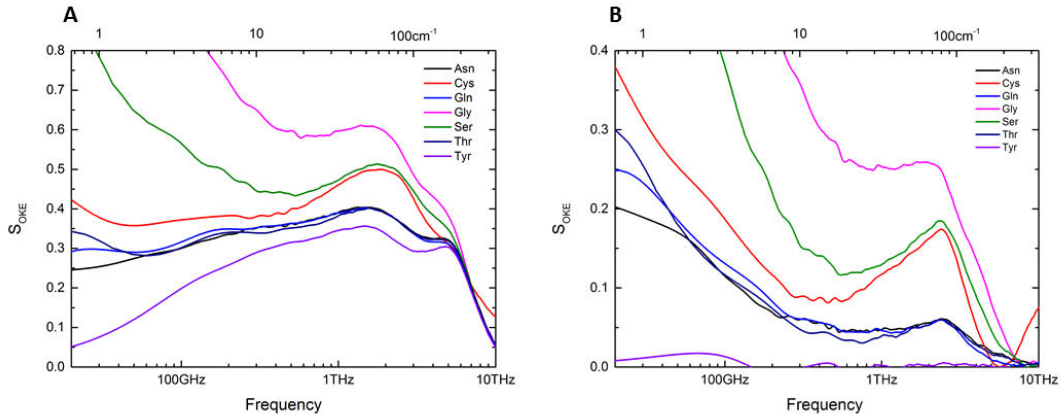


Figure 3.3: OKE spectra of polar (hydrophilic) amino acids. *(A) Solvated spectra (B) Solvent subtracted spectra exposing the amino acid contribution*

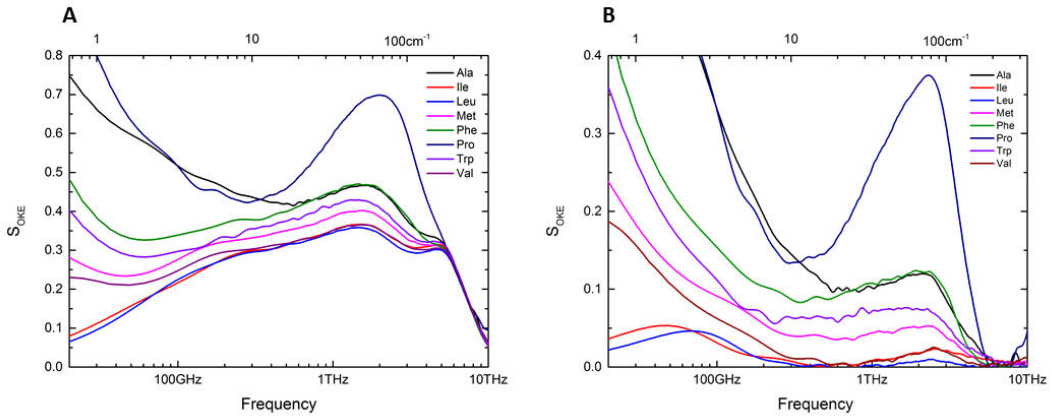


Figure 3.4: OKE spectra of non-polar (hydrophobic) amino acids. *(A) Solvated spectra (B) Solvent subtracted spectra exposing the amino acid contribution*

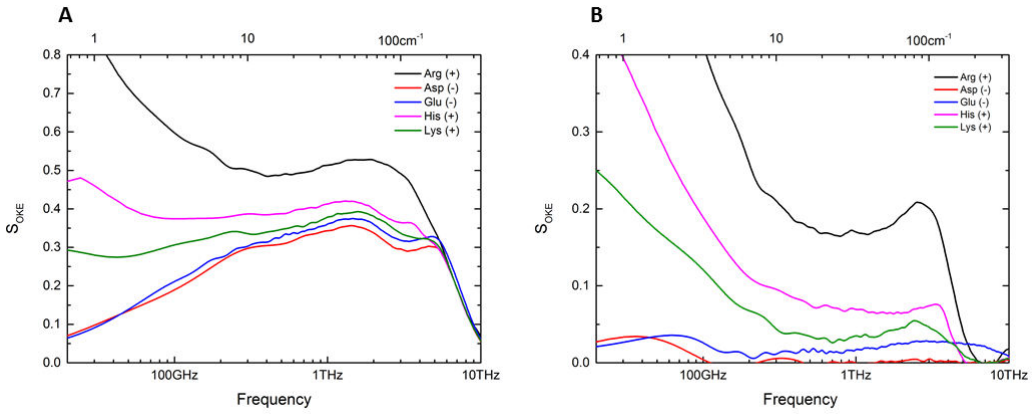


Figure 3.5: OKE spectra of charged amino acids. *(A) Solvated spectra (B) Solvent subtracted spectra exposing the amino acid contribution*

3.1.2.1 Alanine, Arginine & Phenylalanine

Alanine and arginine have a similar solubility, 167 mg/ml and 183 mg/ml respectively, but their large difference in mass means the molarity of alanine is almost double that of arginine (alanine = 1.87 M and arginine = 1.05 M). In contrast, phenylalanine is rather insoluble in water (27.9 mg/ml or 168.9 mM), yet upon analysis of the spectra both alanine and phenylalanine show a similar intensity of the librational band while arginine displays the strongest peak (Fig. 3.6). Although, phenylalanine is much less soluble than alanine, the addition of the benzyl group, makes phenylalanine much more polarisable due to the delocalised electrons in the benzyl ring, which demonstrates how the polarisability of the molecule effects the intensity of the OKE spectrum. Furthermore, the arginine solution has a lower molarity than alanine, yet the most pronounced changes are clearly present in the arginine spectrum, which displays a large increase in amplitude accompanied by a shift of the librational band from ~2.3 THz for alanine to ~2.8 THz for arginine. These changes can be attributed to arginine's unique structure; not only is arginine one of the largest amino acids but it also possesses a complex guanidinium group connected by a three-carbon aliphatic chain. The guanidinium ion has previously been

investigated using OKE spectroscopy and was shown to produce a large response (Turton *et al.,* 2013), which is attributed to the delocalised positive charge that makes the guanidinium group highly polarisable and anisotropic.

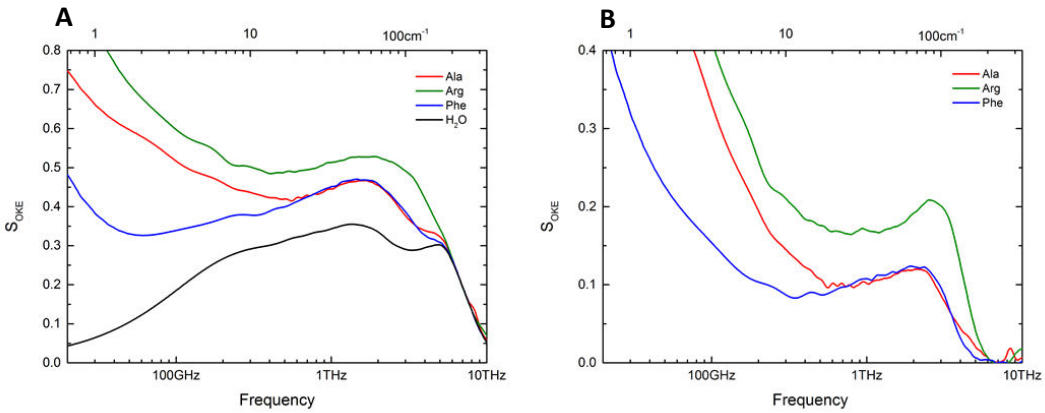


Figure 3.6: OKE spectra of Alanine, Arginine and Phenylalanine. *(A) Solvated spectra (B) Solvent subtracted spectra exposing the amino acid contribution*

The largest change detected in the arginine spectrum is the disappearance of the ~5.4 THz water H-bond stretching mode that is still clearly present in the alanine spectrum (Fig. 3.6). The polar nature of the guanidinium side group allows the formation of multiple hydrogen bonds and classifies arginine as a hydrophilic amino acid that is typically found on protein surfaces. These factors, in combination with arginine's large size, result in an extensive hydration shell surrounding the molecule, which would suggest disruption of the H-bonding network and may explain the absence of this mode from the arginine spectrum. We therefore carried out measurements on dilute arginine samples to see if these changes were concentration dependent, yet the librational band was not seen to shift in the dilute samples and the high frequency hydrogen-bond stretch of water is still disrupted at a concentration of 250 mM (Fig. 3.7). This would suggest that arginine, and the guanidinium group, have a large influence over the hydrogen bond network in water and may contribute significantly to the protein signal.

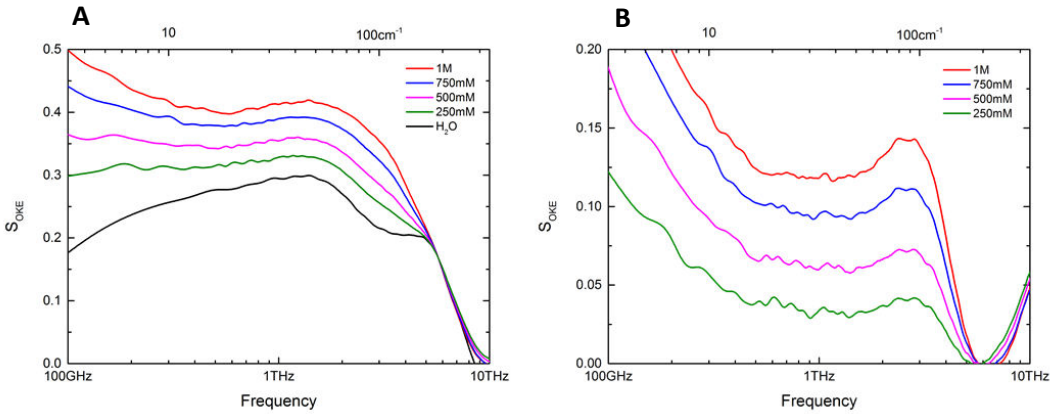


Figure 3.7: Concentration dependent OKE spectra of Arginine. *(A) Solvated spectra (B) Solvent subtracted spectra exposing the amino acid contribution*

3.1.2.2 Glycine, Serine & Cysteine

Glycine, serine and cysteine all have similar solubility in water yet their OKE spectra are quite different (Fig. 3.8). Serine and cysteine both display a sharper librational band at ~2.3 THz than most of the other amino acids, which may simply be a result of the more concentrated solutions giving a more intense peak. However, the glycine librational band is centred at the same frequency, yet is broader due to an additional contribution on the low frequency side, which can be attributed to the large α-relaxation band. That is, the orientational diffusion of the small glycine molecule will occur on a fast timescale, shifting this peak to higher frequencies compared to the larger serine and cysteine molecules. This, coupled with the high concentration result in a strong contribution from this band, so much so, that itbarron begins to overlap the librational band. The hydrogen bond stretch mode can be detected for glycine, serine and cysteine but the amplitude of this mode is different for each molecule. Serine and cysteine differ by a single atomic substitution, from a hydroxyl to a thiol group respectively. Previous studies have shown that vibrational coupling can occur and the strength of a hydrogen bond increases if additional hydrogen bonds are introduced into the system, which therefore increases the amplitude of the peak (Nielsen *et al.,* 2000). This is likely the case for serine as the additional hydrogen bonding from the

hydroxyl group couples with the hydrogen bond stretch of the water clusters, hence increasing the amplitude. Whereas, this mode in the cysteine spectrum does not differ from that of pure water. However, it is unclear why this band is particularly pronounced in the glycine spectrum.

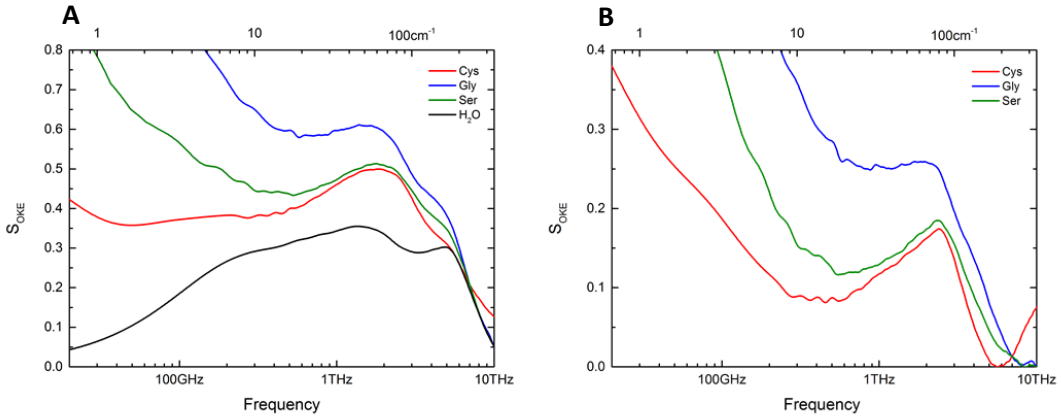


Figure 3.8: OKE spectra of Glycine, Serine and Cysteine. *(A) Solvated spectra (B) Solvent subtracted spectra exposing the amino acid contribution*

3.1.2.3 Glutathione

Measurements were also conducted on small peptides to determine how the OKE spectrum would change when analysing more complex molecules. The peptide glutathione (γ-glutamyl cysteinylglycine) exists in two states. In its reduced state, glutathione exists as a tripeptide but can also form a 'hexapeptide' upon oxidation through linking via a disulphide bond (Fig. 3.9). The results in Figure 3.10 show that both the reduced (GSH) and oxidised (GSSG) forms display an additional contribution between 1 THz and 3 THz, similar to the amino spectrum, which may be attributed to librations of the side chains on the peptides. However, previous calculations and Raman measurements performed on alanine polypeptides (23kDa) have assigned this band to localised backbone torsions (Lee & Krimm, 1998) in agreement with calculations on isolated reduced glutathione that assign modes in this region to torsions throughout the whole tripeptide (Wang *et al.,* 2005).

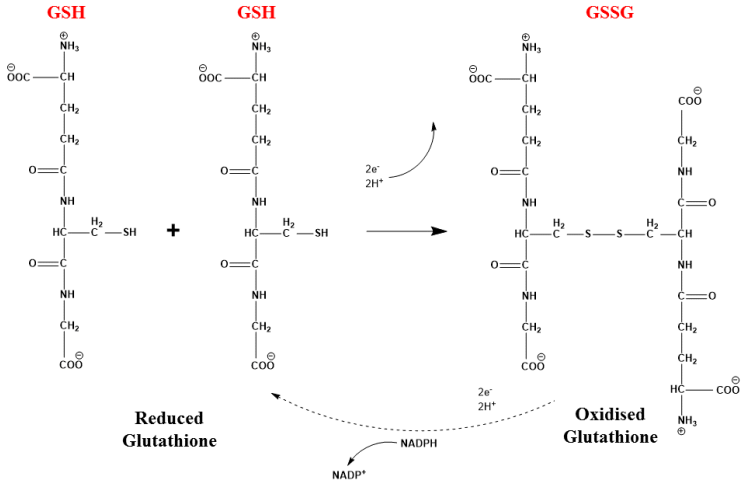


Figure 3.9: Comparison of GSH and GSSG structures. *The two molecules can be interconverted by oxidation or reduction reactions.*

Both glutathione samples were measured at a concentration of 150 g/L yet the reduced form displays a significantly higher amplitude than the oxidised form. Oxidised glutathione contains twice the number of amino acid residues than the reduced form, therefore the molar concentration is effectively halved. Consequently, the reduction in amplitude observed for the oxidised form can simply be explained by the reduced number of molecules in the sample. However, at low frequencies (<100 GHz) we detect the α-relaxation of the molecules, which may explain why the reduced form displays an increased signal in this region, as we detect the high frequency wing of the rotation of the smaller tripeptide.

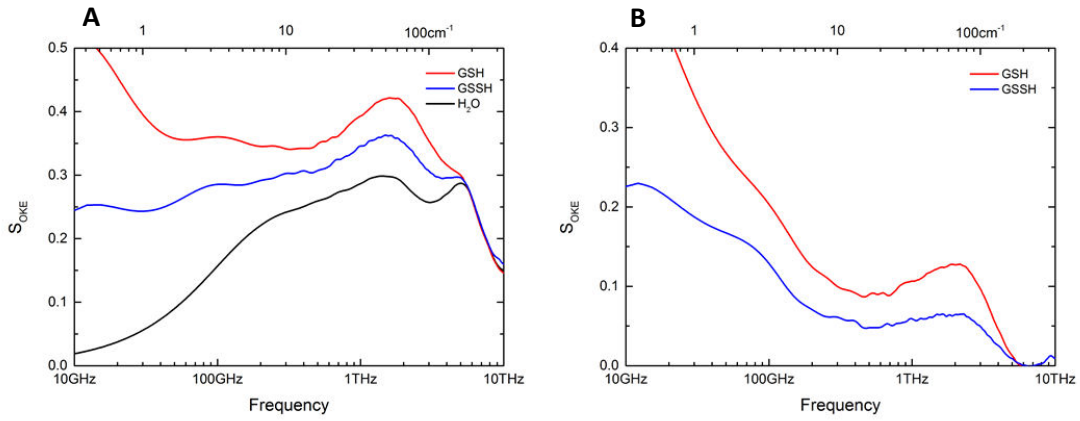


Figure 3.10: OKE spectra of Glutathione. *(A) Solvated spectra (B) Solvent subtracted spectra exposing the amino acid contribution*

3.2 Terahertz analysis of Protein Structure

Motions within protein molecules are known to occur over a wide range of timescales - from high frequency localised modes, to the very low frequency delocalised modes that occur in the terahertz range. These terahertz modes occur on a picosecond timescale and play an important role in protein structure and flexibility (Falconer & Markelz, 2012). For example, picosecond motions of α-helices are thought to be important in protein folding and enzyme activity (Barron *et al.,* 1997) and normal-mode calculations have been carried out to determine the frequency and character of these vibrational modes (Ma, 2005). In the enzyme lysozyme, this has led to the identification of a 3–4 cm^{-1} (~100 GHz) highly delocalised hinge-bending mode that opens and closes the binding cleft (Brooks & Karplus, 1985) and calculations performed on other enzymes have also shown the presence of collective modes around the active site (Levitt *et al.,* 1985). Furthermore, terahertz studies have demonstrated sensitivity to protein tertiary structure and changes in conformation have been detected in the very low frequency range. The conformational changes of two photoactive proteins, bacteriorhodopsin and photoactive yellow protein, have been

monitored using THz-TDS where light activation of the protein was detected by a simple change in absorbance (Whitmire *et al.,* 2003; George *et al.,* 2013). A similar change was also detected upon oxidation of the heme protein, myoglobin (Markelz *et al.,* 2008), where an increase in the overall terahertz dielectric response was observed and explained by the greater flexibility of the oxidised state in agreement with previous work carried out on similar heme proteins (Chen *et al.,* 2005.)

Here, we analyse protein standards in an attempt to elucidate the origin of terahertz protein modes and also study how the OKE terahertz spectrum is affected by changes in protein structure.

3.2.1 Lysozyme

Lysozyme is a small bacteriolytic enzyme found in a number of bodily fluids, including saliva, tears, mucus and milk. The enzyme's antibacterial activity stems from its ability to catalyse hydrolysis of the 1,4-beta-linkages between the N-acetylmuramic acid and N-acetyl-D-glucosamine residues in the bacterial peptidoglycan layer and it also displays chitinase activity by cleaving 1,4-N-acetylglucosamine linkages between chitin (Phillips *et al.,* 1967). One of the best sources of lysozyme is the egg white of birds, where lysozyme accounts for approximately 3.5% of the total egg white proteins (Abeyrathne *et al.,* 2013). The high lysozyme content and easy availability of hen eggs makes hen lysozyme a model system for many studies. It was involved in a number of scientific breakthroughs, including being the first enzyme to be have its crystal structure determined (Blake *et al.,* 1965) and also one of the first enzymes to be fully sequenced (Jolles *et al.,* 1963; Canfield*,* 1963), making it an ideal protein for OKE analysis.

3.2.1.1 Structure and Function

All lysozymes share a similar secondary and tertiary structure although amino acid sequence can vary between species. For example, hen lysozyme differs from that of human lysozyme by 51 out of 129 amino acids and also possesses an insertion between

position 47 and 48 (Mine *et al.,* 1999). Hen lysozyme has a molecular weight of 14.3 kDa, possesses six helix regions and displays high thermal stability due to the presence of four disulphide bridges present throughout the structure. The native enzyme folds into an ellipsoid shape with dimensions of approximately 4.5 x 3.0 x 3.0 nm (Cegielska-Radziejewska *et al.,* 2008) (Fig. 3.11).

Like most enzymes, factors such as pH and temperature also influence the structure and function of lysozyme in solution. Lysozyme usually exists as a reversible dimer at close to physiological pH (between pH 5 and pH 9) but is predominantly in a monomeric state at acidic pH (Sophianopoulos *et al.,* 1969). Furthermore, it has been shown that lysozyme exists in two states between 20°C and 30°C with a transition point at 25°C (Jolles & Jolles, 1984). NMR studies have shown that above this transition point lysozyme displays a decreased affinity for its inhibitor (Cozzone *et al.,* 1975), whereas below this temperature the same inhibitor is seen to be tightly bound to the enzyme (Jolles *et al.,* 1975).

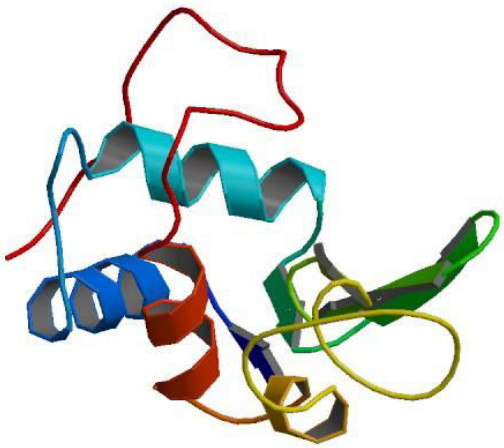


Figure 3.11: Ribbon diagram of lysozyme structure (PDB = 1LZA)

3.2.1.2 Results and Discussion

The solvent subtracted OKE spectra of all proteins display similar features, dominated by a strong peak at ~2.8 THz with a broad low frequency shoulder. For lysozyme, Dr David Turton fitted these features with a pair of Brownian oscillators, yielding the parameters $\omega_0 = 1.15$ THz and $\gamma = 0.52$ THz for the broad shoulder and $\omega_0 = 2.80$ THz and $\gamma = 1.24$ THz for the narrow peak (Fig. 3.12). As the damping (γ) is less than the frequency (ω_0), this implies that the vibrational mode involves an underdamped oscillator (Turton *et al.,* 2014).

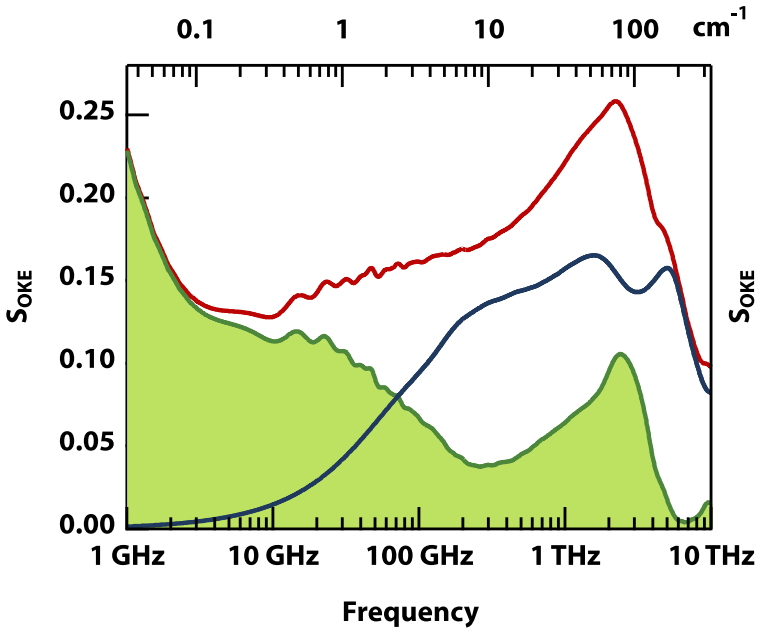


Figure 3.12: OKE spectrum of Lysozyme. *The figure shows the spectrum of the lysozyme solution at 15 mM (red), buffer solution (blue) and the solvent subtracted lysozyme contribution (green shading).*

As had been explained previously, for simple molecules, features in this region are usually attributed to librational modes. However, for larger molecules, such as proteins, it is suggested that this band is due to torsional and bending motions of the protein backbone with hindered rotations (Hunt *et al.,* 2004) and torsions (Yu., *et al.,*

2004) of the amino acid side chains. Yet, if the solvated-protein spectrum in the terahertz range were due to multiple localised modes, one would expect an inhomogeneous Gaussian distribution of local-mode frequencies—reflecting different local masses and forces—giving rise to a broad overdamped spectrum. Yet, this is not observed. An alternative explanation for the observed signal is librations of individual amino acids. Orientational motions of those moieties with the highest polarisability anisotropy produce the largest OKE signals and from our amino acid data it is apparent that tryptophan, with its planar fused-aromatic indole group, gives an unusually strong OKE response considering its low solubility (13.2 mg/ml). Coupled with the fact that lysozyme contains six tryptophan resides, three of which are located in the active site, we used this residue to test this hypothesis. In Fig. 3.13, the OKE spectrum of tryptophan in solution is shown at the same concentration as used for the lysozyme spectrum. This reveals a broad librational band at ~2 THz that is more than an order of magnitude weaker than the sharp 2.8 THz peak seen in lysozyme. Previous OKE studies of styrene and polystyrene have shown a reduction in the librational band strength by a factor of three upon polymerisation (Hunt *et al.,* 2004). Thus, the contribution of each individual tryptophan residue to the lysozyme spectrum is expected to be nearly two orders of magnitude weaker than the sharp 2.8 THz peak.

Having discounted the above possible explanations, the sharp terahertz peak observed appears to correspond to a single or a small number of delocalised modes involving motions of many residues. The number of residues involved in such motions is difficult to estimate, however, based solely on the measurements shown in Fig. 3.13, one can estimate that the delocalised mode is likely to involve at the very least a dozen residues.

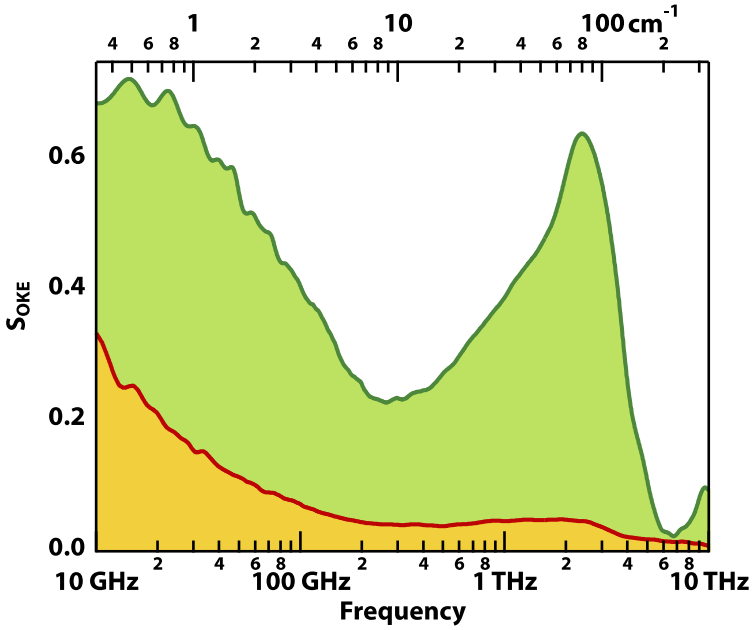


Figure 3.13: Comparison of the OKE spectra for lysozyme and tryptophan. *The solvent subtracted spectra of lysozyme (green) and tryptophan (red) both at a concentration of 15 mM.*

Dr Hans Senn carried out molecular dynamics simulations on lysozyme, with the principal aim of characterising the motions around 3 THz, that is, on a timescale of a few hundred femtoseconds. Fourier filtering (Sessions *et al.,* 1989) enabled us to analyse the motions in selected frequency windows. The root mean square fluctuations (RMSFs) of atomic positions for the unfiltered trajectories, which include all frequencies, show the expected pattern: surface residues are more mobile than residues in the core and side chains are more mobile than the backbone (Fig. 3.14a). Analysis of the motions in the 1.5–3.5 THz window shows a distinctly different pattern (Fig. 3.14b). There is no longer a 'cool' core and a 'warm' surface of the protein. Surface loops, which are very mobile overall, have lower mobility in this frequency range whereas side chains in the core, which are overall rather rigid, have RMSFs comparable to surface side chains. While side chains on the whole are more mobile than the backbone, both in the full spectrum and the 1.5–3.5 THz window, it is the high RMSFs of the terminal side-chain atoms that stand out in the 1.5–3.5 THz

window. Visual inspection of the filtered trajectories revealed that the motions in this frequency range are best described as side-chain librations or hindered rotations about terminal single bonds. These torsional oscillations result in large displacements for the terminal atoms most distant from the backbone; hence they appear 'hot' in Fig. 3.14b.

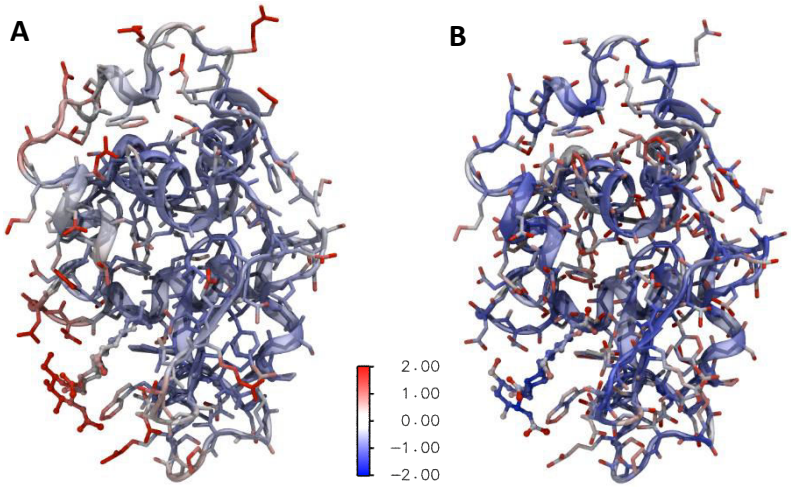


Figure 3.14: Motions of lysozyme from molecular dynamics simulations. *(A) Motions in the full spectral range from unfiltered trajectories (B) Motions in the 1.5 – 3.5 THz range from Fourier-filtered trajectories.*

The calculated vibrational density of states (VDoS) (Fig. 3.15) has a broad peak at ~2.5 THz, consistent with the experimental spectra. This feature is almost entirely due to contributions from the side chains while the backbone VDoS varies little in the 0.5–10 THz range. This finding agrees with the above characterisation that the motions around 2.5 THz are attributed to side-chain librations in the lysozyme protein. In comparing the VDoS to the OKE spectrum, it is important to remember that the VDoS includes all internal motions whereas OKE is sensitive only to those that influence the polarisability anisotropy. Therefore, it is not currently feasible to simulate the OKE spectrum for an entire solvated protein.

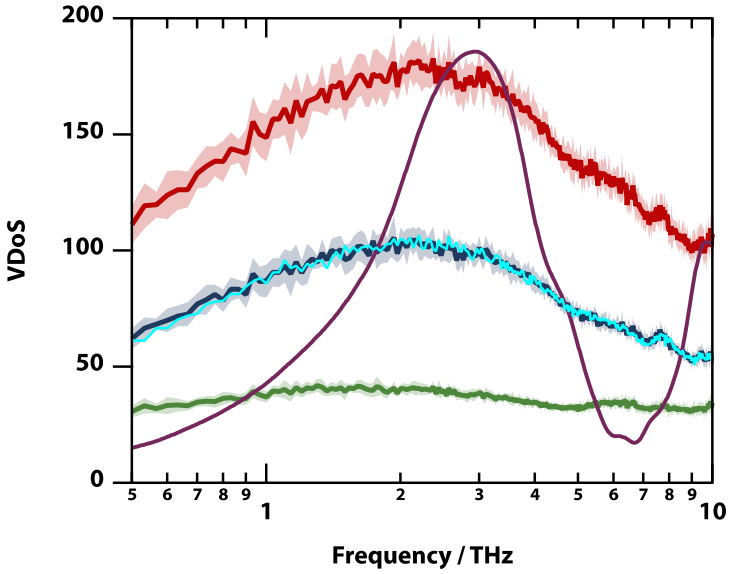


Figure 3.15: Vibrational Density of States from molecular dynamics. *The averaged vibrational density of states for lysozyme (red) with the contributions of the backbone (green) and side chains (blue). The experimental OKE spectrum (purple) is also shown.*

3.2.2 Analysis of Protein Tertiary Structure with OKE Spectroscopy

In order to analyse how protein secondary and tertiary structure influence the terahertz OKE spectrum we compared the lysozyme spectrum with a structurally homologous protein, α-lactalbumin, and also Bovine serum albumin (BSA), which has a considerably different structure.

3.2.2.1 α-lactalbumin

α-lactalbumin (lalba) plays an important role in lactose biosynthesis (Hill & Brew, 1975) and is an essential component of the milk enzyme lactose synthase. (Stanciuc & Rapeanu, 2010). Lalba forms the regulatory subunit of the lactose synthase heterodimer while β-1,4-galactosyltransferase forms the catalytic unit that converts UDP-galactose and glucose to lactose. Naturally, lalba is one of the most abundant whey proteins in milk (approximately 1.2g/L) and is easily isolated by

chromatography and precipitation/aggregations techniques (Kamau *et al.,* 2010). This availability, coupled with high solubility, have made it an attractive protein for structural studies and other novel uses for this protein have also been discovered. For example, as a monomer lalba has displayed properties of an effective bactericidal agent and a folding variant of lalba in human milk was shown to be effective against an antibiotic resistant strain of *Streptococcus pneumonia* (Hakansson *et al.,* 2000). Although the native structure does not display any antibacterial activity it appears that changes in conformation can lead to the exposure of bactericidal domains within the protein. These antibacterial domains or sequences have previously been observed by Pelligrini and co-workers (Pelligrini *et al.,* 1999), who discovered that proteolytic digestion of lalba by trypsin produced two peptides with bactericidal properties and another antibacterial peptide was also discovered upon proteolysis with chemotrypsin. It has also been shown that at physiological pH lalba associates with lysozyme to form a complex that attacks both gram-negative and gram-positive bacteria with enhanced efficiency (López Expósito & Recio*,* 2006).

3.2.2.1.1 Structure and Function

α-lactalbumin (lalba) is a small, calcium binding milk protein that is similar to Lysozyme in sequence size and dimensions, yet their functions are quite different. Most lalba proteins consist of 123 amino acids with a molecular weight of approximately 14,200Da, the exception being rat lalba, which contains an additional 17 C-terminal residues (Nitta & Sugai*,* 1989). In its native structure, lalba consists of two domains connected by a calcium binding loop held within a deep cleft. The larger domain is predominantly composed of alpha-helices: three major and two shorter helices, and the smaller domain consists of an antiparallel β-pleated sheet and short helix connected by a series of loops (Permyakov & Berliner*,* 2000) (Fig. 3.16).

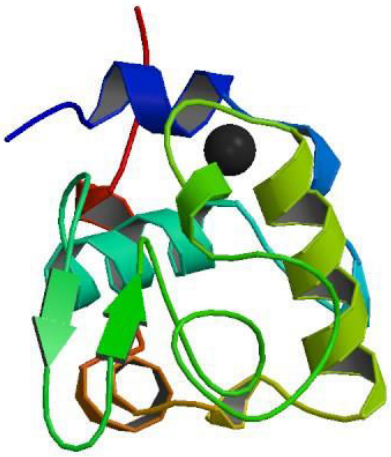


Figure 3.16: Ribbon diagram of calcium bound Lactablbumin (PDB = 1HFZ)

3.2.2.2 Bovine Serum Albumin

Albumins are a family of highly water soluble globular proteins. The most common of this family, serum albumin, is produced in the liver and readily found in milk and blood, where it accounts for approximately 1% and 5% of total protein, respectively (Schwenke *et al.,* 1996). The main function of blood serum albumin is to regulate oncotic pressure but it also acts as a carrier for a number of small molecules such as thyroid hormones, steroid and fatty acids (Baker, 1998). Bovine serum albumin (BSA) is one of the most widely studied, best characterised proteins and has found its way into a variety of applications in biochemistry as a protein standard and buffering agent due to its inert nature and its availability as a by-product of bovine blood from the cattle industry.

3.2.2.2.1 Structure and Function

BSA is originally composed of 607 amino acids before an 18 residue peptide is cut from the N-terminal region, followed by cleavage of further 4 amino acids to yield the mature protein (583a.a.). The native protein has a molecular weight of 66kDa and forms a ellipsoidal shape composed of 3 domains, each consisting of a large and small

double loop and a short and long connecting segment (Anand and Mukherjee, 2013) (Fig. 3.17). This arrangement is partially produced by a unique arrangement of nine disulphide loops created by 17 disulphide bonds. Yet, BSA also possesses another free thiol group, which allows it to form covalently linked dimers (Narazaki *et al.,* 1997). Serum albumins possess a large α-helix content, somewhere in the region of 70%, with the remainder of the protein consisting of random coils and turns. (Sjoholm and Ljungstedt, 1973; Reed *et al.,* 1975) - BSA does not possess any β-sheets (Carter *et al.,* 1989; Oberg, *et al.,* 2001).

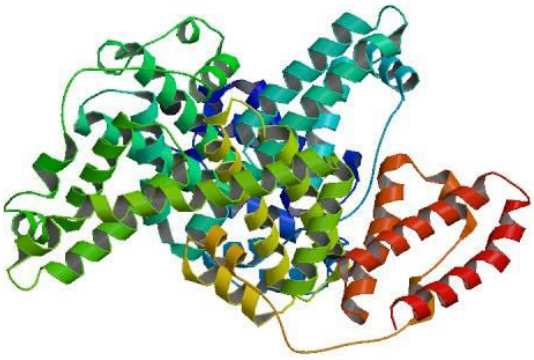


Figure 3.17: Ribbon diagram of BSA (PDB = 3V03)

3.2.2.3 Results and Discussion

The frequency of the strong peak at ~2.8 THz identified for lysozyme is almost identical for all proteins, yet there is a small but significant upshift in frequency of this feature in lalba compared to the other proteins (Fig. 3.18). Lalba is known to adopt a 'molten globule' state under the conditions of the experiment and the shift in this peak will be discussed later in this chapter (Section 3.2.3.2). In addition to this feature, variation in the low frequency shoulder of the 2.8 THz band between 300 GHz and 1.5 THz is more pronounced in the lysozyme spectrum than the other two proteins. This is surprising, as we would predict a similar spectrum between lysozyme and lalba due to the structural homology. However, it is possible that this band may be

influenced by a number of different low frequencies features. For example, we detect an increase in signal for all proteins below ~300 GHz, which agrees with Depolarised light scattering (DLS) measurements of lysozyme solutions that identified a feature between ~20 GHz and ~300GHz that was assigned to protein-induced solvent perturbations by comparison to inelastic neutron scattering (INS) (Perticaroli *et al.,* 2010). INS in mainly sensitive to the motions of hydrogen atoms within the sample, so it is possible to essentially eliminate the solvent from the spectrum by using D2O. The 20 GHz to 300 GHz feature that is present our OKE spectra is absent from the INS data, which shows that this feature must originate from water molecules directly influenced by the protein.

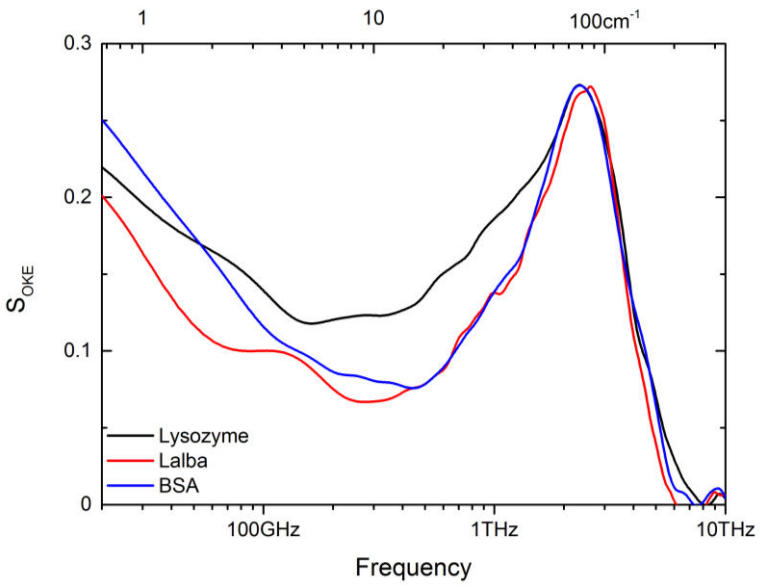


Figure 3.18: Solvent Subtracted OKE spectra of proteins*. Spectral intensities have been normalised to the ~2.8 THz peak.*

If the spectra below 300 GHz are influenced by the protein's solvation shells, then it would be logical to assume that surface hydrophobicity would influence this feature. However, there have been varying results as to whether hydrophobic or hydrophilic surfaces have the greatest influence over the solvent dynamics: Mid infrared (Rezus & Baker, 2007) and terahertz spectroscopy (Heyden & Havenith, 2010) studies agreed

that hydrophobic interactions impact more upon the water dynamics than hydrophilic ones, yet quasi-elastic scattering (Russo *et al.,* 2004) and MD calculations (Stirnemann *et al.,* 2010) would suggest otherwise. OKE studied on small solutes (Mazur *et al.,* 2009) and peptides (Mazur *et al.,* 2010) have shown that the retardation factor between hydrating and bulk water is larger for hydrophilic sites, confirming that hydrophilic interactions have a more prominent effect.

Early studies on protein hydrophobicity found BSA and lalba to be highly hydrophobic, whereas lysozyme displays hydrophilic properties (Bigelow 1967), in agreement with recent stalagmometric measurements (Amrhein *et al.,* 2015). Therefore, it is possible that the hydrophilic nature of lysozyme results in the solvation band moving to lower frequencies, causing the observed broadening of the ~2.8 THz feature on the low frequency side. Mazur *et al.* (2012) performed OKE measurements on lysozyme, BSA and trypsin samples at varying concentrations and found that the picosecond relaxation dynamics become slower with increasing concentration, caused by the solvent containing more 'hydrating water', forming a solvation shell around the protein, compared to 'bulk water', which presents relaxation dynamics on a faster timescale. We therefore carried out measurements on dilute lysozyme solutions in order to determine how protein concentration affects the ~2.8 THz low frequency shoulder (Fig. 3.19). A reduction was detected with decreasing concentrations, which agrees with Mazur *et al.* observations and confirms that the hydrophilic nature of lysozyme is responsible for this feature. At concentrations <100 mg/ml this feature becomes comparable in strength to the other proteins.

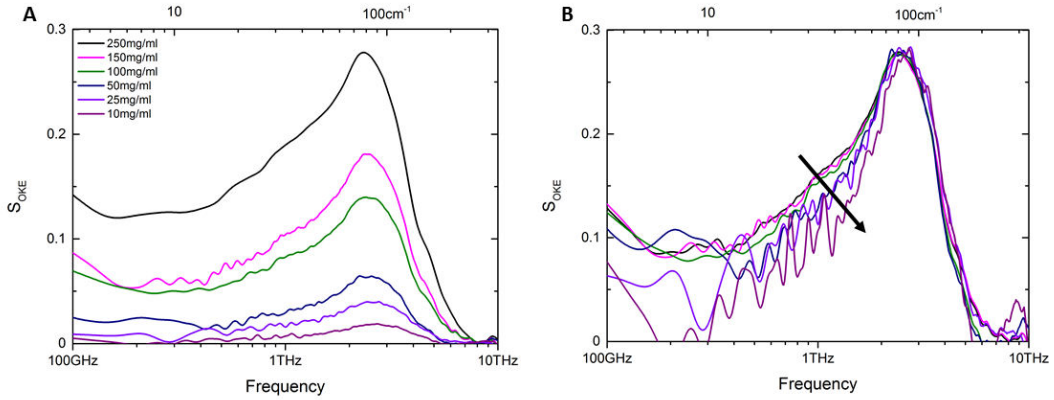


Figure 3.19: OKE spectra of lysozyme dilutions. *(A) Solvent subtracted spectra of lysozyme at varying concentrations (B) Dilutions normalised to the maximum concentration for direct comparison of the ~2.8 THz feature. The arrow shows the reduction in the low frequency shoulder with decreasing protein concentration.*

3.2.3 Temperature induced protein unfolding

3.2.3.1 Bovine Serum Albumin

Increasing the temperature of a protein will raise the kinetic energy until the molecule is working at optimum efficiency. Above this temperature, bonds and interactions begin to break down and quaternary, tertiary and secondary structure can be lost, but primary structure is retained. Occasionally proteins are capable of refolding into a functional state but, more commonly, the protein is irreversibly denatured. Of course, the stages of denaturation - and temperatures at, which they occur - vary widely between proteins and few have been studied as extensively as BSA.

Generally speaking, BSA undergoes two structural changes during heating: heating to approximately 65°C will result in changes that are largely reversible and can be considered the first stage. The second stage occurs upon subsequent heating, which results in irreversible denaturation, but does not necessarily result in complete loss of ordered structure (Kuznetsow *et al.,* 1975; Lin and Koenig, 1976; Oakes, 1976). Differential scanning calorimetry (DSC) measurements of 20% BSA solutions would

suggest that the onset of conformational change occurs at 58.1°C (Poole *et al.,* 1987) and denaturation follows at 64°C (Itoh *et al.,* 1976), which is in agreement with Ruegg *et al.* (1977) who used DSC to determine the denaturation temperature of BSA in simulated milk ultrafiltrate. However, others have reported a denaturation temperature of 70°C for fatty acid free BSA (FAF-BSA) (de Wit & Klarenbeek, 1984). When comparing BSA with and without fatty acids, Michnik (2003) also detected the same transition peak for FAF-BSA but identified an additional change at 56°C. Further studies have shown that the transitions taking place in serum albumin solutions are also concentration dependent and when heated above the transition stage, β-sheets appear to form. Using infrared spectroscopy techniques, β-sheet formation was detected in BSA solutions heated above 70°C (Lin & Koenig, 1976; Clark *et al.,* 1981) and human serum albumin (HSA) solutions above 65°C (Wetzel *et al.,* 1980). Wetzel *et al.,* (1980) identified a shoulder in the amide I band that would indicate the development of a β-sheet structure at concentrations of 50 mg/ml and 1.4 mg/ml, yet no shoulder was seen at 0.5 mg/ml, which would suggest that β-sheet formation is a concentration dependent process and involves intermolecular interactions. Other experiments have also looked at the concentration dependent effect of heat on serum albumin, with somewhat contradicting results. Ross and Shrake (1988) measured HSA over a concentration range of 1.43-74.1 mg/ml and found that higher concentrations had a stabilising effect, whereas measurements of BSA from 2-5 mg/ml show the opposite (Michnik, 2003).

3.2.3.1.1 Results and Discussion

Circular Dichroism (CD) spectroscopy was employed to monitor the changes in BSA solutions (100 mg/ml) as a function of temperature and the change at 222 nm used as an indicator of secondary structural modifications. At this frequency, CD is known to be sensitive to alterations in α-helix and β-sheet content (Kelly *et al.,* 2005), which can be used as a useful comparison to the OKE spectra in order to gain a better understanding of the dynamics taking place over the measured temperature range,

while also accommodating the high concentration of protein used in the corresponding OKE measurements.

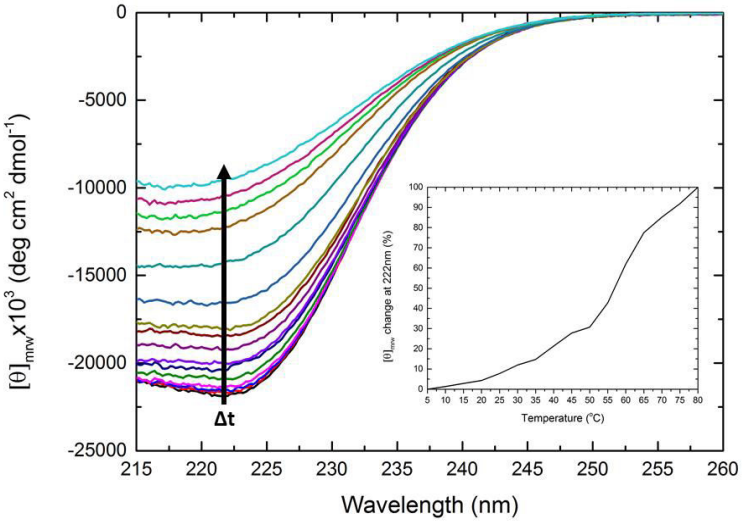


Figure 3.20: Thermal denaturation of BSA (100 mg/ml). *CD spectra of BSA at varying temperatures. The inset shows the percentage change of BSA at 222 nm over the temperature range measured.*

From the CD spectra (Fig. 3.20) it is clear that the largest changes in BSA structure occur between approximately 50-65°C, although there is still considerable difference between 25-55°C, which is the temperature range that our OKE data is shown (Fig. 3.21). OKE measurements were also recorded at temperatures >55°C, however it was not possible to gain reliable data at these temperatures due to scattering effects, and are therefore not included. The observed scattering is most likely due to aggregation of the sample at high temperatures and previous ultracentrifugation studies carried out on HSA have shown that the protein has a sedimentation coefficient is 4-4.5 S at neutral pH. However, when HSA is heated to temperatures more than 60°C and cooled down, the protein is more heterogeneous with sedimentation coefficients in the range of 26-36 S, which is indicative of high molecular weight complexes and hence, aggregation (Wetzel *et al.,* 1980). Furthermore, electron spin and fluorescence (Hull *et al.,* 1975) data show that residue Cys-34 is located within a pocket of HSA,

which was seen to unfold during thermal denaturation allowing intermolecular disulphide bonds to form and dimerisation to occur. These studies confirm why it was impossible to gain accurate measurements of BSA at high temperatures in our OKE experiments, since aggregates form and introduce scattering from the sample. However, it may be possible to reduce dimer formation by blocking the Cys-34 residue with cysteine, glutathione or other chemicals such as N-iodosuccinimide (Peters, 1985), yet the contribution from these molecules would have to be taken into account in the final OKE spectrum, complicating analysis. Therefore, this was not deemed necessary in our measurements.

The solvent subtracted OKE spectrum of BSA (Fig. 3.21) shows a significant increase in amplitude upon heating. From our CD measurements and previously studies (Lin and Koenig, 1976; Clark *et al.,* 1981), it is clear that the alpha helix content of BSA will change over the temperature range studied and could therefore be a contributing factor to the observed change in the OKE spectra. Previous studies have tried to understand the effect protein secondary structures have on the OKE signal using polypeptide chains as a simple model to produce α-helices and random coils (Hunt *et al.,* 2007). The authors showed that an α-helix conformation could be identified by a broad band at ~100 cm^{-1} (3 THz), whereas the peptide in a random coil configuration displayed a broad band at ~185 cm^{-1} (5.5 THz). Although we do not detect a pronounced band at 5.5 THz in our solvent subtracted protein spectra, initial comparisons suggest that the high frequency random coil mode is detected in our BSA measurements as a slight increase in signal with temperature, indicating a loss in ordered structure and increased random coil configuration. This increase in the 5.5 THz band is most prominent in the 55°C trace, which is consistent with our CD data and it is within the temperature range where a large conformational change is known to take place (Poole *et al.,* 1987; Michnik, 2003). Conversely, if this observation was correct, one would also expect to see a reduction in the signal at ~3 THz, which would represent a decrease in α-helical content. Yet this is not seen in our OKE BSA data and we therefore dismiss this change as an artefact. In fact, we detect an increase in

the ~3 THz peak, which could originate from a number of factors. For example, Hunt *et al.* (2007) also detected a change in amplitude of this peak when using temperature variations to switch poly-L-lysine between the α-helix (at 4°C) and random coil (at 35°C) configurations. At higher temperatures (random coil), the OKE spectrum of poly-L-lysine showed an increase in amplitude below 3 THz compared to the low temperature spectrum (α-helix), which is a similar trend seen for our BSA data. However, further temperature measurements revealed that heating poly-L-lysine caused increased Raman scattering below 3 THz, irrespective of conformation. Similarly, previous studies have also reported temperature-related increases in the low frequency density of states below 3 THz - the so-called 'Boson peak' region - which may account for the increased signal observed for BSA in this region (Paciaroni *et al.,* 1998; Caliskan *et al.,* 2004), which would suggest that the change in amplitude may not be representative of the protein conformational change. However, along with the increase in amplitude of the 3 THz peak, we also see a broadening of this feature on the low frequency side with increasing temperature. Hydrogen bonding modes are observed in this frequency region for pure water and solvent-peptide intermolecular hydrogen bonding can also be detected (Hunt *et al.,* 2005). It is therefore not unreasonable to assume that the hydrogen bonding network of the system may change as the protein undergoes a temperature-induced conformational change. Additionally, intramolecular backbone-torsional motions have also been assigned to bands in this frequency range during studies of polymer systems and in simulations of α-helices (Hunt *et al.,* 2004; Lee & Krimm, 1998). Therefore, as the protein unfolds and adopts an open conformation, resulting in more freedom and flexibility throughout the backbone, it is likely that these modes also contribute to the peak broadening.

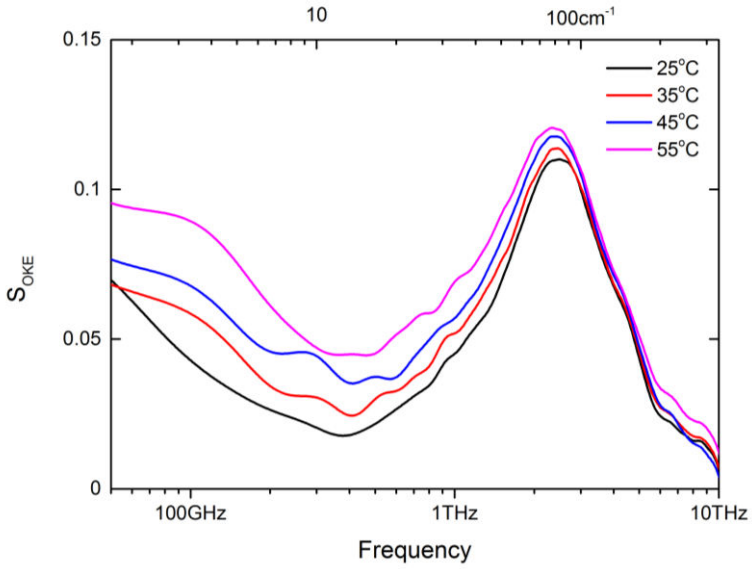


Figure 3.21: Solvent subtracted OKE spectra of BSA at increasing temperatures.

The low frequency region of the solvent subtracted spectra (<300 GHz) also shows an increase in amplitude with temperature and we have previously described this band as the water diffusing from the protein hydration layer. Measurements on carbohydrates (Paolantoni, *et al.,* 2009) and other small solutes (Mazur *et al.,* 2009) have found similar retardation factors to that of proteins, which would indicate that the water dynamics are not heavily influenced by the size or shape of the solute. Therefore, it is unlikely that we detect a conformational change in this region based solely on the unwinding of the protein. However, as BSA unfolds, the hydrophobicity of the protein will inevitably change, which may have an effect on the solvation layer and hence, OKE spectrum in this region. Alternatively, NMR (Halle & Davidovic, 2003) and DLS measurements (Dubin *et al.,* 1971) have shown that protein relaxation processes, or protein 'tumbling', occur on a timescale of around 5-10 ns, corresponding to a frequencies of tens of MHz. Although our data does not show these very low frequency processes, temperature induced intensity increases have been observed at these frequencies, associated with accelerated tumbling of the protein in solution in relation to reduced viscosity (Perticaroli *et al.,* 2010). As we have seen from our CD measurements and previously published data (Poole *et al.,* 1987; de Wit &

Klarenbeek, 1984; Lin and Koenig, 1976; Clark *et al.,* 1981), full denaturation of BSA does not occur under our OKE experimental conditions and it is unlikely that the degree of unfolding would significantly influence the hydrophobicity in this range. Therefore, it is likely that the change we detect >300 GHz is a consequence of increased protein tumbling that we cannot visualise due to the time scale of our measurements.

3.2.3.2 α-lactalbumin

Lalba has the surprising ability to bind a large number of cations (Mn^{2+}, Mg^{2+}, Na^{+} and K^{+}) that can all compete for the Ca^{2+} binding site (Permyakov *et al.,* 1985). Unlike the majority of calcium binding proteins, lalba does not possess a standard EF-hand motif, yet the primary calcium binding site is still located in a loop between two helices and the carboxylic groups of three Asp residues (82, 87 and 88) and two carbonyl groups on the protein backbone are responsible for the strong Ca^{2+} binding (Permyakov & Berliner, 2000) (For lalba structure, see Section 3.2.2.1.1, Fig. 3.15). Upon binding of Ca^{2+}, pronounced changes can be detected in the lalba structure and function, most notably, the increased stability of the protein's tertiary structure. Calcium depleted lalba is more sensitive to variations in pH and ionic strength (Stanciuc & Rapeanu, 2010), whereas the calcium loaded state is more resistant to denaturing agents (Permyakov *et al.,* 1985), high pressure (Dzwolak *et al.,* 1999) and is also structurally stable to more than 40°C higher than apo-lalba as shown by DSC measurements (Veprintsev *et al.,* 1997; Griko *et al.,* 1994).

Lalba can also adopt a molten globule (MG) state, corresponding to the protein retaining the majority of its secondary structure but displaying a dynamic tertiary structure (Dolgikh *et al.,* 1981). The molten globule state of lalba is the predominant configuration at acidic pH and also when the apo-state is exposed to low salt and elevated temperatures (Stanciuc & Rapeanu, 2010). Various studies have been carried out under these conditions in order to elucidate lalba's MG configuration. Dolgikh *et al.* (1985) was one of the first to identify this state using circular dichroism by

detecting secondary structures in the far UV but a featureless spectrum in the near UV, which would suggest a lack of fixed tertiary structure. X-ray scattering measurements subsequently identified the radius of Ca^{2+} loaded lalba to be 15.7 Å and the radius of the acid molten globule state to be 17.2Å (Kataoka *et al.,* 1997). From these studies it is believed that the MG structure of lalba retains a globular shape but becomes highly hydrated and expands from the native state (Kharakoz & Bychkova, 1997).

The MG state is possible due to a stabilising hydrophobic core, which allows the protein to retain a native-like structure (Wu & Kim, 1998). Furthermore, a helical domain is continually present in the molten globular state and appears to be protected from hydrogen exchange, which explains its stability in acidic environment (Schulman *et al.,* 1995). It has also been shown that the protein stability is entirely governed by the polypeptide sequence itself and not by other interactions such as disulphide bonds, which was apparent when lalba displayed a similar shape to the wild type at acidic pH after substituting its eight cysteine residues to alanine (Redfield *et al.,* 1999).

Previous groups have looked at the effects of protein denaturation in the terahertz range by changing the protein's environment - increasing temperature, pressure or adding a chemical denaturant. However, here it is possible to measure native and denatured states under identical conditions using a non-polarisable small ion, that is undetectable in OKE measurements, which allows us to eliminate the influence external factors may have on solvent or protein and carry out a direct comparison.

3.2.3.2.1 Results and Discussion

In a low salt environment, apo-lalba adopts the 'Molten Globule (MG)' state at room temperature (Permyakov *et al.,* 2000) and prior circular dichroism studies have confirmed that this is the prominent state under similar conditions used for our OKE measurements (Okazaki *et al.,* 1994). In our measurements we used a higher

concentration of Tris buffer than is typical for these experiments but the protein concentration is also significantly higher. High concentrations of Tris buffer have previously been shown to slightly stabilise and increase the denaturation temperature of apo-lalba (Griko *et al.,* 1999). However, this occurred at Tris concentrations of 300 mM and 500 mM using a protein concentration of 1 mg/ml, whereas our measurements use 100 mM Tris buffer and 250 mg/ml protein. Furthermore, we carried out DSC measurements (Fig. 3.22) to determine the denaturation temperatures, which match closely with previous results carried out in distilled water, at neutral pH (Relkin *et al.,* 1992). We can therefore be confident that the concentration of buffer and protein used in our measurements does not result in any significant stabilising effects and apo-lalba is in the MG state at 25°C. Moreover, the binding of the calcium ion increases the stability of the protein (Permyakov *et al.,* 2000) and holo-lalba is in a functional conformation at 25°C.

Our DSC measurements (Fig. 3.22) show that the apo-form of lactalbumin undergoes a thermal transition between 30-40°C, whereas the addition of the calcium ion dramatically increases the thermal stability of the protein by >25°C and the transition is now observed between 60-70°C. The temperature of these transitions agrees well with previously published data (Hendrix *et al.,* 2000; Relkin *et al.,* 1992) and the authors concluded that the transition was associated with the thermal denaturation of the protein and it has also been reported that MG transitions are not detectable using calorimetric studies (Yutani *et al.,* 1992). Hence, our OKE measurements of apo-lalba at 50°C display the spectrum of the fully denatured protein and the holo-form is in its functional state.

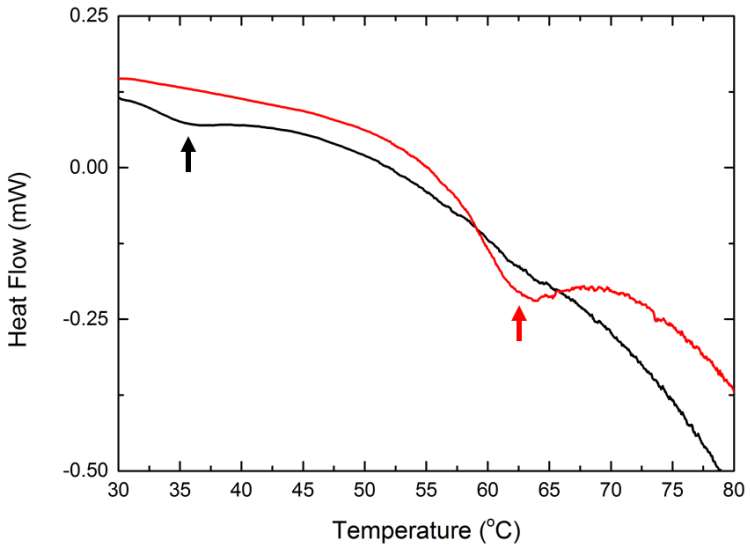


Figure 3.22: Thermal denaturation of lalba by DSC. *The black and red traces shows the thermal denaturation of apo-lalba and holo-lalba, respectively. The arrows indicate the point of transition.*

The most noticeable difference between the OKE spectra is the reduction in amplitude of the holo-state compared to apo-form (Fig. 3.23). That is, the calcium bound protein displays a weaker OKE signal at both 25°C and 50°C compared to the MG state and denatured protein, respectively. The OKE signal is influenced by the polarisability of the sample and as the amplitude change appears to be consistent across most of the spectrum, and not frequency specific, it is probable that this drop in amplitude is caused by a change in the overall polarisability of the protein in the apo- states. In fact, MD simulations have shown that greater flexibility in polypeptide chains causes an increase of atomic fluctuations, which can induce a larger polarisability and a higher scattering intensity (van Vlijmen & Karplus, 1999). Therefore, as the MG state adopts an 'open' conformation and the denatured protein lacks secondary or tertiary structure, this is likely the source of the increased signal.

The spectra of lalba at 25°C show a small shift of the boson peak to lower frequencies for the apo-form compared to the holo-form (Fig. 3.23). The molten globule state

retains its secondary structure but is swollen from the native state with a difference in radius of approximately 2Å (Kataoka *et al.,* 1997) and it is suggested that this alteration in the backbone structure of the protein likely induces the small shift. In the lowest frequency region (>300 GHz) of our OKE spectrum, apo-lalba also displays a slight increase in amplitude where the spectrum is sensitive to the protein hydration layer. The MG form is known to adopt a highly hydrated state compared to the active protein (Permyakov *et al.,* 2000), which could explain why we detect a change in this region of the spectrum.

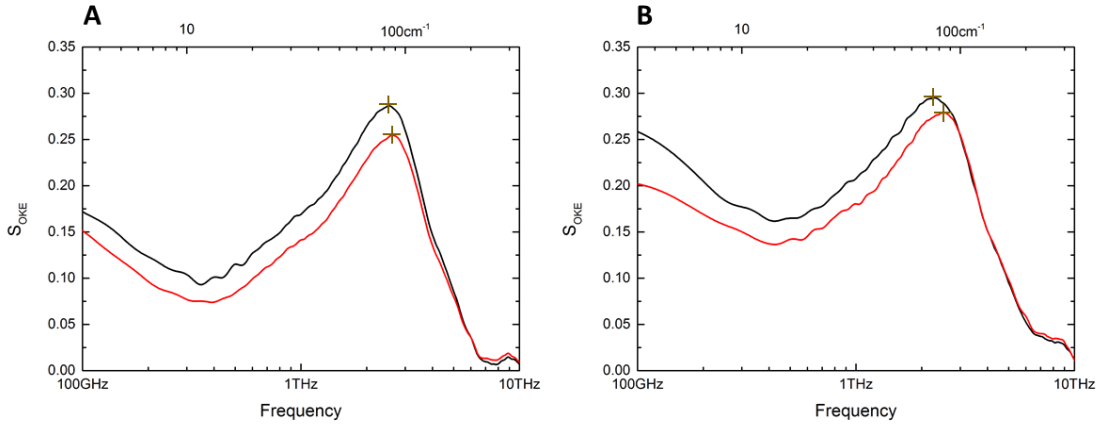


Figure 3.23: Solvent subtracted OKE spectra of apo- (Black trace) and holo-lalba (Red trace). *(A) Spectra at 25°C (B) Spectra at 50°C. The crosshairs highlight the shift of the 2.8 THz peak.*

Upon analysis of the samples measured at 50°C, we detect a more pronounced red shift of the boson peak (~2.8 THz) for the apo-protein compared to the holo-form than was seen at 25°C (Fig. 6). From the DSC data it is clear that apo-lalba is denatured at 50°C, whereas holo-lalba is still in a native conformation. Therefore, the shift in frequency must stem from the unfolded protein state. Sassi and co-workers (2013) measured the thermal denaturation of lysozyme from 25-85°C using depolarised light scattering (DLS) and although the protein spectrum appears to red-shift with increasing temperature, which is a similar observation to ours, the authors do not

discuss this. Temperature changes influence the hydrogen bond modes of the water only spectrum in this region, and therefore, the authors do not associate the shift with protein transitions. However, this does not explain the red-shift seen in our data as the temperature is constant for each sample.

Lerbret *et al.* (2012) carried out inelastic neutron scattering measurements of lysozyme solutions from 1-6 kbar and detected a blue shift in frequency with increasing pressure. They attributed this change to stiffening of the lysozyme residues due to the pressure induced loss of volume. Therefore, as our denatured sample shows a redshift, instead of a blueshift, it could be assumed that the residues of apo-lalba experience increased flexibility, which is consistent with the protein adopting an unfolded, open conformation. However, it is also worth noting that lysozyme does not denature under the conditions set by Lerbret *et al.* (2012) but requires a pressure of >8 kbar in D_2O (Hedoux *et al.,* 2011) to penetrate water molecules into the hydrophobic core of the protein (Hummer *et al.,* 1998). Yet, other Raman studies have looked at the denaturation of lysozyme at higher pressures and it appears that the ~2.8 THz peak continues to shift to higher frequencies after denaturation (Hedoux *et al.,* 2011). Therefore, in these measurements it is doubtful that the peak shift induced by increasing pressure is due to alterations within the secondary or tertiary protein structure and more likely to stem from the loss of volume surrounding individual residues. However, for our experiments, the conditions are identical and we can therefore be confident that the redshift of the 2.8 THz peak originates from the unfolding of the protein backbone.

3.3 Conclusions

Many groups have attempted to detect vibrational modes in proteins in the terahertz range. Raman and THz-TDS spectroscopy of crystalline lysozyme and other proteins have shown peaks in the 20–100 cm^{-1} (~1–3 THz) range (Acbas *et al.,* 2014; Urabe *et al.,* 1998; Perticaroli *et al.,* 2010). However, these are very sensitive to crystal packing and hydration levels and may be associated with phonon modes of the crystal lattice, rather than biologically relevant phenomena. Raman spectroscopy on solutions of lysozyme and other proteins has been able to show that some terahertz bands do exist but again it has been impossible to prove biochemical relevance (Genzel *et al.,* 1976; Hedoux *et al.,* 2006). Here we have shown that the vibrational spectrum of lysozyme in solution exhibits sharp peaks at 1.15 THz and 2.80 THz corresponding to two underdamped delocalised oscillators that involve the motion of at least a dozen residues in the protein. Analysis of the other proteins in this chapter is still on-going in the Wynne lab and we hope to assign specific parameters to the vibrational modes imminently. However, with the current data, we have shown that terahertz measurements of protein solutions can be used to successfully identify individual proteins by slight differences in their spectra. This may not accurately reflect the secondary and tertiary structure of these molecules as considerably different proteins display similar spectra, similar to previous OKE studies on a larger sample of proteins (Giraud *et al.,* 2003). Nevertheless, OKE terahertz spectroscopy is sensitive to other factors, such as surface hydrophobicity, and can also be used to monitor conformational changes within protein molecules as was shown by measurements on BSA and lalba where the changes are mainly linked to increased flexibility of the protein backbone.

Chapter 4: Protein-Ligand Binding

Chapter 4 Protein-Ligand Binding

There have been suggestions that terahertz frequency underdamped collective modes may direct protein dynamics along the correct path on a complex potential-energy surface (Deak *et al.,* 1998). This may be responsible for the transport of energy through the protein (Davydov, 1973), which may facilitate biological function through phonon-like modes (Liu *et al.,* 2008). Theoretical models of enzymes have suggested that concerted conformational fluctuations on a femtosecond to picosecond timescale promote catalysis by coupling to the reaction coordinate (Hay & Scrutton, 2012; Benkovic & Hammes-Schiffer, 2003; Agarwal, 2005; Yang & Bahar, 2005). This model is also supported by experiments in which the direct excitation of terahertz frequency modes with far-infrared light was shown to enhance the reaction rates in proteins (Austin *et al.,* 1989). However, the role of terahertz frequency vibrational modes in mediating efficient protein-ligand binding and biochemical reactions in general is not known. Therefore, we choose to monitor the terahertz spectra of three small molecule induced protein modifications that are central to the field of protein chemistry: Ligand binding to the active site was studied utilising lysozyme and its inhibitor N,N',N''-triacetylchitotriose (NAG_3), protein phosphorylation was analysed using the response regulator, CheY, and allosteric control was investigated using catabolite activator protein (CAP) and its ligand, cyclic adenosine monophosphate (cAMP).

4.1 Example 1: Lysozyme-Inhibitor Binding

4.1.1 Introduction

The structure and function of native lysozyme has been described previously (Section 3.2.1). However, the enzyme also has the ability to bind and hydrolyse different substrates, giving the enzyme its antibacterial properties. The active site of lysozyme lies in a deep cleft running between the enzyme's two domains where it can accommodate up to six units of a polymeric substrate in each of the six subsites, simply denoted: A, B, C, D, E and F (Blake *et al.,* 1967). A conformational change in the enzyme causes the substrate to adopt a strained conformation and distorts the bonding between the 4^{th} and 5^{th} unit, resulting in hydrolysis and cleavage (McKenzie & White FH*,* 1991). However, tri-saccharide molecules are also able to bind the A, B and C subsites of lysozyme's active site and induce a conformational change, yet as subsites D and E are not occupied, no hydrolysis can take place and the tri-saccharide remains bound as an efficient inhibitor (Jolles & Jolles*,* 1984). Therefore, we chose N,N',N''-triacetylchitotriose (NAG_3) as an inhibitor to induce a conformational change in lysozyme that could be measured using OKE spectroscopy.

4.1.2 Results and Discussion

Circular dichroism (CD) was employed to confirm the binding of NAG_3 to lysozyme and follow the conformational change. CD measures the differences in absorption of right- and left handed circularly polarised light, which in the near UV is sensitive to the orientation and interaction of aromatic amino acid chromophores. It is well documented that 3 tryptophan residues are present in the lysozyme active site and that all are influenced by binding of tri-saccharide ligands (Jollès & Jollès*,* 1984). However, it is Trp-108 that dominates the near-UV CD profile of lysozyme. Upon NAG_3 binding, an increase in signal is observed at the three peaks centred at 282 nm, 288 nm and 292 nm that confirms the conformational change (Fig. 4.1). This change

is influenced by a shift in Trp-108, resulting from an increased interaction with Glu-35 induced by NAG_3 binding (Teichberg *et al.,* 1970).

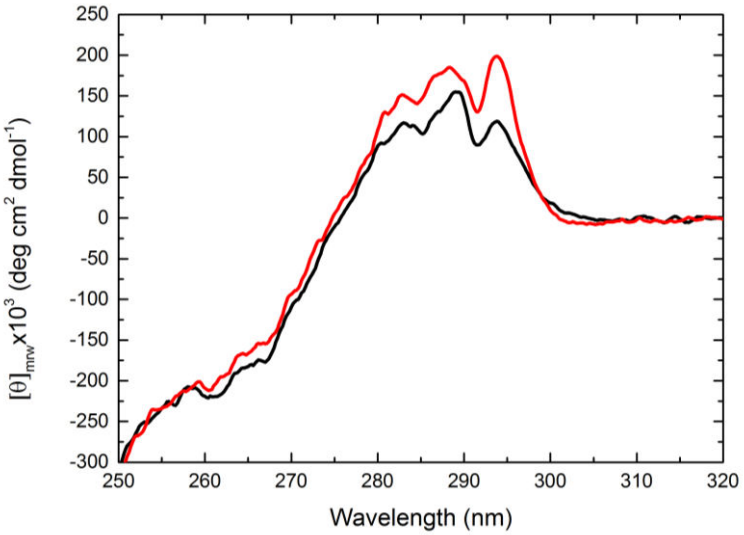


Figure 4.1: Near UV CD spectra of Lysozyme (black) and Lysozyme in presence of NAG_3 (red).

OKE spectra were obtained for native lysozyme and also after the addition of the inhibitor, NAG_3, at a protein-to-saccharide molar ratio of 1:1. Measurements for a 15 mM solution of NAG_3 alone showed that the weakly polarisable saccharide makes no measurable contribution. However, the OKE spectrum for lysozyme-NAG_3 shows a pronounced increase in intensity at around 3 THz compared to the native protein (Fig. 4.2). The difference spectrum shows an undulation, comprising an increase in intensity on the higher frequency side of the sharp lysozyme peak, accompanied by a weaker but significant decrease in intensity around 1 THz. From the broad bandwidth of our OKE measurement, it is evident that the spectrum at lower and higher frequencies is unperturbed. A simple interpretation of the undulation feature in the difference spectrum would be of a band shift from 1 to 3 THz. However, it has been shown that for OKE spectra, a band shift in isolation would give rise to a negative peak much larger in amplitude than a positive peak (Turton *et al.,* 2014), whereas here

we see the converse. Therefore, as the OKE signal is proportional to the product of the polarizability, it is likely that the anomalous strength of the peak at ~3 THz must be caused by an increase in the polarisability of the mode. Additionally, the difference spectrum can be fitted by a pair of Brownian oscillators (Fig. 4.2) yielding the parameters $\omega_0 = 1.29$ THz, $\gamma = 0.64$ THz for the negative mode and $\omega_0 = 2.89$ THz, $\gamma = 1.38$ THz for the positive mode. The amplitude of the positive mode is 7.5 times greater than that of the negative mode and since the damping factor γ is less than ω_0, the mode is significantly underdamped.

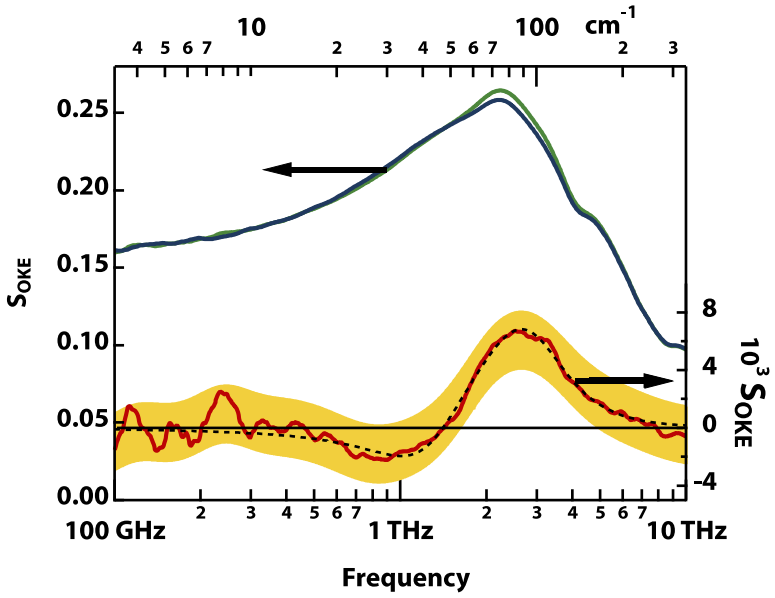


Figure 4.2: OKE spectra of lysozyme solutions. *Native lysozyme (blue trace) and in complex with NAG_3 (green trace). The difference spectrum (red trace) highlights the increase in intensity at ~3 THz with a weaker reduction in intensity at ~1 THz. The yellow region is the estimated uncertainty in the difference spectrum based on the repeatability of the measurements and the dashed trace is a fit by two brownian oscillators.*

From our OKE data we see that the undulating difference band is not inhomogeneously broadened into an overdamped lineshape, which implies that it cannot be caused by the sum of many small changes in many local modes. This agrees

with the X-ray crystal data of native and NAG_3-bound lysozyme as few localised changes are detected. Instead, we see a very small decrease in the average root-mean-square (RMS) atomic displacement upon binding, which suggests a small degree of stiffening overall (Kuhs, 1992). However, a few individual residues that are in contact with the NAG_3 ligand have significant displacement, in particular, Trp62 and Trp108 (Kuhs, 1992; Jollès & Jollès, 1984.) Yet, from our data on amino acids, which show a small OKE response and different lineshape at this concentration, we rule out that the undulation is caused by the change in environment of a single residue. Therefore, in an attempt to elucidate the origin of the OKE differences seen upon inhibitor binding, Dr Hans Senn compared root-mean-square-fluctuations (RMSFs) for simulations with and without bound NAG_3 (Fig. 4.3) and found no significant difference in the average RMSFs for several frequency windows spanning 0.5–5.5 THz. These results support the interpretation that the difference peak is due to changes in polarisability. Thus, the data prove that the undulation is caused by a delocalised mode of the protein blue-shifting accompanied by an increase in polarisability.

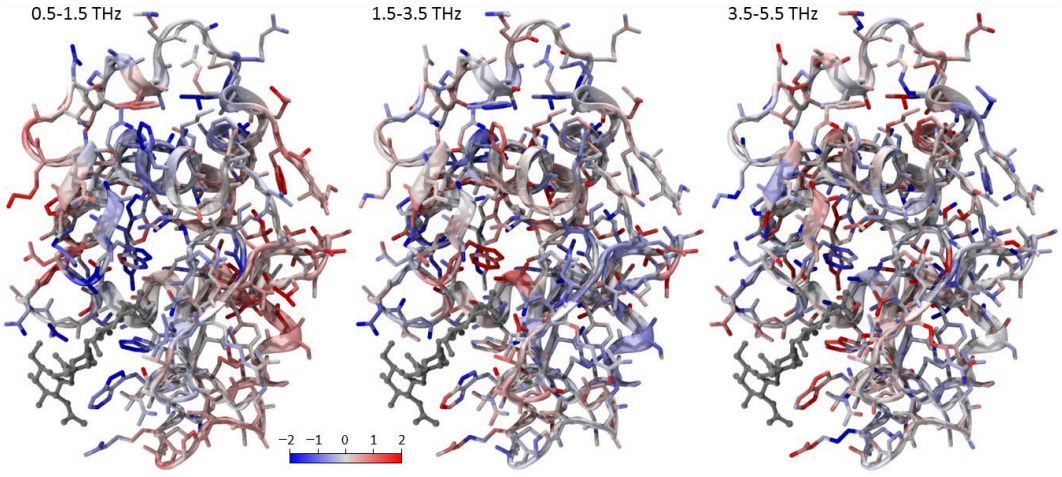


Figure 4.3: Normalised RMSF difference between free and NAG_3-bound lysozyme in three frequency windows.

4.2 Example 2: Phosphorylation of CheY

4.2.1 Introduction

Protein phosphorylation is an essential post-translational modification in a number of signal transduction pathways, regulating various processes throughout eukaryotic and prokaryotic cells by mediating enzyme activation and inhibition. The addition of the phosphate group to a polar or positively charged side chain of an amino acid can radically alter the protein chemistry, changing a previously hydrophobic portion into an extremely hydrophilic region, which can result in conformational changes that extend throughout the molecule. Therefore, phosphorylation usually occurs on serine, threonine, tyrosine and histidine residues in eukaryotes and controls a range of processes such as enzyme inhibition (Cole *et al.,* 2003), cell cycle/tumour suppression (Ashcroft *et al.,* 1999) and insulin signalling (van Weeren *et al.,* 1998). In prokaryotic cells, arginine, lysine and aspartate residues are also phosphorylated to control cellular processes. The best example of this is 'two-component signalling', where histidine or aspartate residues are usually involved (Chang *et al.,* 1998).

4.2.2 Two component systems

Two-component systems are generally simple pathways and signal transduction begins with an external stimuli, which initiates autophosphorylation of a histidine kinase (HK), where the phosphor-donor is a molecule of adenosine triphosphate (ATP). The phosphate group can then be transferred from the H-K to a 'response regulator', inducing a conformational change that leads to the appropriate response, which may be stimulation (or repression) of target genes or physical adaptation to the external environment.

Hundreds of two-component regulatory systems have been identified and described over the last decade (Chang *et al.,* 1998), yet few have received as much attention as the Che proteins responsible for chemotaxis in *E. coli* and these proteins are now one

of the model systems used to describe signal transduction in prokaryotes. The Che pathways begins when chemical gradients are sensed through trans-membrane receptors called methyl-accepting chemotaxis proteins (MCPs) and convey the information across the plasma membrane into the cytosol. The protein, CheW acts as an anchor to attach the histidine kinase, CheA, to the MCPs where it can autophosphorylate and in turn activate the response regulators, CheB and CheY, which accept the phosphate group from CheA. CheB is a methylesterase that removes methyl groups from the glutamate residues of the MCPs and along with CheR - a methyltransferase that adds methyl groups to the glutamate residues - controls the cell's response to the external stimuli as the level of methylation determines how the MCPs receive attractants and repellents (Fig. 4.4). CheY, in its phosphorylated state, induces cell tumbling by interacting directly with the flagellar switch protein, FliM, to change the direction of flagellum rotation from counter-clockwise to clockwise, allowing the bacterium to reorientate towards or away from an attractant or repellent, respectively. CheY is dephosphorylated by the phosphatase CheZ.

The importance of the CheY molecule in the Che pathway coupled with its relative simplicity has resulted in numerous studies investigating how the protein's structure (Cho *et al.,* 2000; Lee *et al.,* 2001) and dynamics (McDonald *et al.,* 2012; 2013) change between the native and phosphorylated states. Therefore, we choose to study the CheY protein for our OKE measurements as a model system of protein phosphorylation.

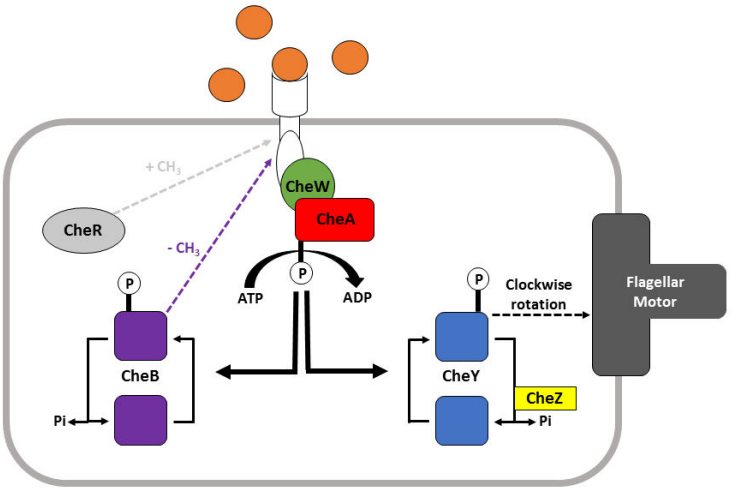


Figure 4.4: Signalling pathway of Che Proteins

4.2.3 Structure

CheY is small globular protein consisting of a single regulatory domain of 129 amino acids (approximately 14 kDa) and has served as a useful model for monitoring the structural changes that take place upon phosphorylation (McDonald *et al.,* 2013) as well as protein-protein interactions (Sagi *et al.,* 2003). The regulatory domain of CheY is folded into a doubly wound α/β topology with a central five-stranded parallel β-sheet and five α-helices (Fig. 4.5). Phosphorylation of CheY occurs on Asp57 but is dependent on the presence of a Mg^{2+} co-factor, which adopts an octahedral conformation and directly acts upon the side chain carboxylate oxygens of Asp57 and Asp13 as well as the backbone carbonyl oxygen on Asn59 (Stock, *et al.* 1993). Threonine87, alanine and lysine109 are also involved in phosphorylation and form hydrogen bonds with the phosphoryl group, initiating the conformational change in the protein (Cho *et al.,* 2000; Appleby & Bourret, 1998).

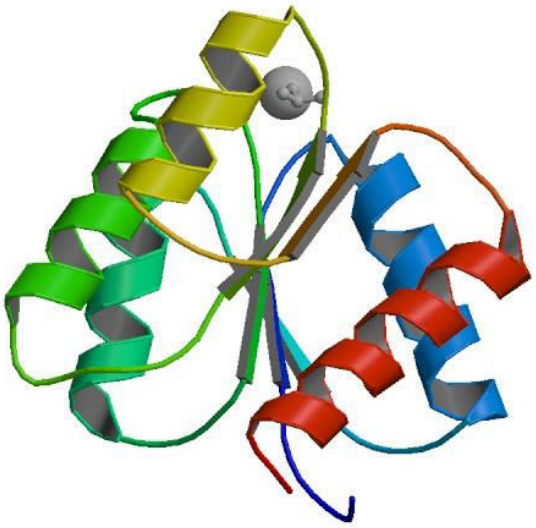


Figure 4.5: Ribbon diagram of phosphorylated CheY (PDB = 1FQW)

4.2.4 BeF_3^- as a phosphomimic

E. coli requires the ability to respond quickly to external stimuli and therefore possesses the ability to auto-dephosphorylate rapidly; the auto-dephosphorylation rate of CheY is 2.5 min^{-1} (Thomas *et al.,* 2008). This makes it extremely difficult to study the protein in its phosphorylated state. However, it has long been known that beryllofluoride forms a structure similar to phosphate and in 1987, Chambre and colleagues demonstrated that BeF_x could bind transducin by forming analogues of the γ-phosphate group of GTP (Bigay *et al.,* 1987). Furthermore, it has also been shown that $BeF_2.OH^-.H_2O$ and $BeF_3^-.H_2O$ could tightly bind to F-actin and microtubules (Combeau & Carlier, 1989). More recently, Yan and co-workers discovered that beryllofluoride can mimic the chemistry of aspartyl phosphate and bind irreversibly to the response regulator NtrC, as well as others, including CheY (Yan *et al.,* 1999). Although other metal complexes, such as orthovanadate and aluminofluorides (Fig. 4.6), have previously been used successfully as phosphomimics for other ATPase enzymes (Kleuss *et al.,* 1994), these analogs did not substantially stimulate the ATPase activity of response regulators such as NtrC (Yan *et al.* 1999).

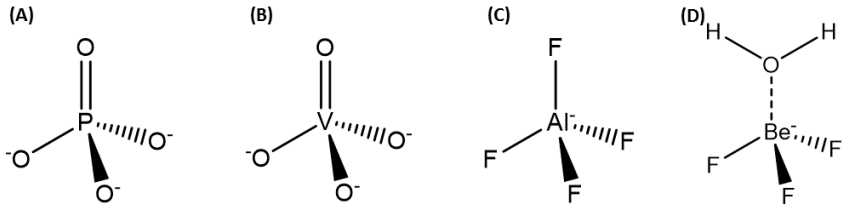


Figure 4.6: Structural analogs of the phosphate group. *(A) Phosphate (PO_4^{3-}) (B) Orthovanadate (VO_4^{3-}) (C) Aluminofluoride (AlF_4^-) (D) Beryllofluoride ($BeF_3^-.H_2O$)*

An unusual characteristic of beryllium is its tendency to form compounds in which it has an incomplete octet of valence electrons, which allows it to form a variety of different soluble ionic complexes when fluorine is present – BeF_x^{x-2} (where $x = 1 – 4$). Therefore, the predominant species in solution depends on the concentration of free fluoride ions. BeF_4^{2-} adheres to the octet rule, but BeF_2 and BeF_3^- do not, which leaves them with an electron deficient beryllium atom that can accept co-ordinate covalent bonds with water (fig. 4.6), or even the lone pair of electrons on an aspartate carboxylate group (fig. 4.7). Whatever the number of fluorides bound to a beryllium ion in solution, they will always form a strictly tetrahedral complex, with Be-F bond lengths of 1.55 Å, which is identical to a phosphate group. In theory, aluminofluorides can also adopt this tetrahedral complex, however it has been suggested that they may assume a trigonal bipyramid conformation that makes it an accurate analog of the 'transition state' of ATP hydrolysis in some enzymes (Fisher *et al.,* 1995) whereas beryllofluoride maintains its tetrahedral geometry and therefore maintains the enzyme in the 'ground state' resembling the Michaelis complex (Xue *et al.,* 1997). For activating response regulators, beryllofluoride in its BeF_3^- state is most efficient, either by adopting a $BeF_3^-.H_2O$ conformation and acting as a phophomimic (Fig. 4.6) or more likely, as a direct analog of aspartylphosphate – aspartyl.BeF_3^- (Fig. 4.7).

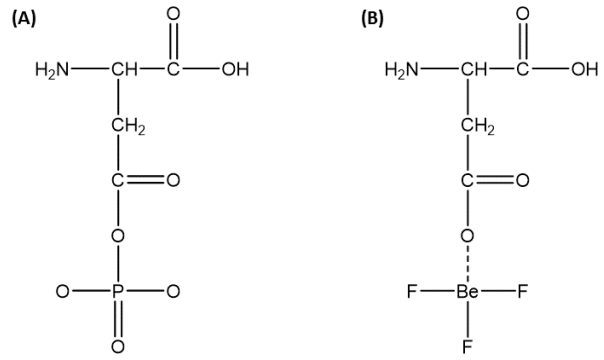


Figure 4.7: Binding of (A) phosphate and (B) BeF_3^- to an aspartate residue.

4.2.5 Results and Discussion

4.2.5.1 Sample Preparation

CheY was over expressed and purified utilising nickel affinity purification as the protein has been produced with an N-terminal His-tag (See section 2.2.3.4.3). The introduction of a tag does not introduce large modifications to the protein molecule (20 extra amino acids) and typically does not affect the protein's properties. In fact, it has been shown that the rates of autophosphorylation and autodephosphorylation are indistinguishable between wild-type and His-tagged CheY (Bourret *et al.,* 2010), which would suggest that the protein's chemistry is not changed by the addition of the tag. However, in order to maximise the OKE signal in our measurements it was necessary to obtain a high concentration of CheY, which highlighted that the solubility of the protein was an issue, as protein concentrations over 25 mg/ml regularly resulted in precipitation and it was not possible to obtain protein solutions above this concentration.

Although the poly histidine-tag does not appear to influence CheY's properties, it is also known that small tags can sometimes have a detrimental effect on protein solubility (Woestenenk *et al.,* 2004). The protein was cloned to include a specific cleavage site so the tag could be removed with the enzyme thrombin. We therefore

carried out the digestion at 4°C, 25°C and 37°C in order to determine the optimal conditions (Fig. 4.8). As is expected, cleavage was most efficient at 37°C but at this temperature the chances of protease degradation are also increased. At 4°C, CheY would be least effected by proteases, yet as large quantities of the protein were required to reach the desired concentrations for OKE measurements, it was also necessary to factor in cost. Therefore, it would be too expensive to acquire the large amounts of thrombin required to cleave CheY at this temperature. The poly histidine-tag is almost completely cleaved from CheY using 5 units of thrombin per 1 mg protein at 25°C (Fig. 4.8). At this ratio and temperature, the cost of thrombin is decreased and protease digestion is reduced. We therefore used these conditions in all subsequent purifications of CheY.

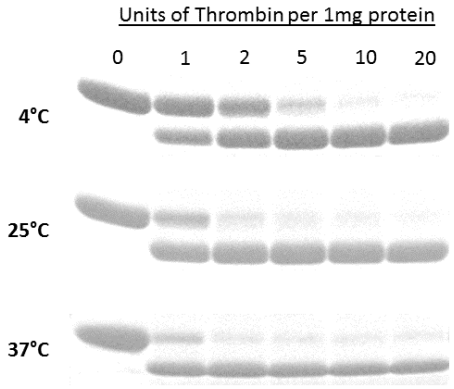


Figure 4.8: HIS-tag cleavage from CheY. *Removal of the tag results in faster migration speeds on the SDS-PAGE gel. Therefore, the lower bands represent the cleaved protein.*

Our hypothesis was confirmed and the solubility and stability of CheY increased significantly upon the removal of the his-Tag. However, it was necessary to perform size exclusion chromatography as an additional step in the purification process in order to remove the cleaved His-tag and thrombin from the sample. The final purity of CheY was thought to be >95% (Fig. 4.9).

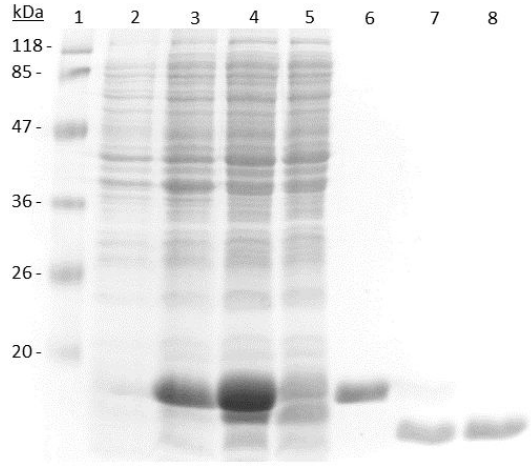


Figure 4.9: Stages of CheY purification. *Lanes: (1) EZ-RUN ladder (2) Before induction with IPTG (3) After induction with IPTG (4) Crude (5) Flow Through (6) Nickle affinity purification (7) Thrombin cleavage (8) Size exclusion chromatography.*

4.2.5.2 NMR Spectroscopy

The most appropriate technique to confirm the conformation change induced by BeF_3^- was NMR spectroscopy. NMR and OKE measurements were performed under identical conditions in order to eliminate any variations that may be induced due to changes in buffer or concentration of protein (80 mg/ml). The high protein concentration used in these measurements allowed us to carry out high quality, natural abundance ^{15}N-^{1}H HSQC data, which removed the need to isotopically label the protein. Our NMR data show characteristic regions of the HSQC spectra for apo-CheY, CheY in the presence of Mg^{2+} and BeF_3^- activated CheY. Fig. 4.10 shows the region of the spectrum containing the site of phosphorylation (D57), which is observed at a proton chemical shift of approximately 8.48 p.p.m. and a nitrogen shift of about 127.52 p.p.m. for apo-CheY. Although Mg^{2+} directly acts upon the side chain carboxylate oxygen of Asp57 (Stock *et al.,* 1993) it does not induce a large shift in the spectrum, whereas with BeF_3^- bound a large shift is observed. Additionally we detect a large shift in D64 and V108 from apo-CheY to the Mg^{2+} bound and BeF_3^- activated forms, which confirms the conformational change, as these residues are

located in close proximity to the β4–α4 loop that is known to be involved in CheY activation (McDonald *et al.,* 2012).

Fig. 4.11 shows the region including T87, which is located very close to the BeF_3^- binding site and forms a hydrogen bond between its hydroxyl group and the phosphate group upon phosphorylation (McDonald*,* 2013). We detect almost no difference between any of the residues observed in this region of the HSQC spectra between the apo- and Mg^{2+} bound forms but a large shift in T87 with BeF_3^- binding, showing it to be an effective phosphomimic that partakes in hydrogen bond formation. Fig. 4.12 and Fig. 4.13 also show characteristic regions of the CheY HSQC spectra and again we detect slight changes between apo-CheY and Mg^{2+} bound and much larger shifts in the BeF_3^- bound spectra. In particular, there are large shifts of V86 and I55 in Fig. 4.12 and Fig. 4.13 respectively, which is in good agreement with previous NMR measurements performed on CheY (Cho *et al.,* 2000; Hubbard *et al.,* 2003).

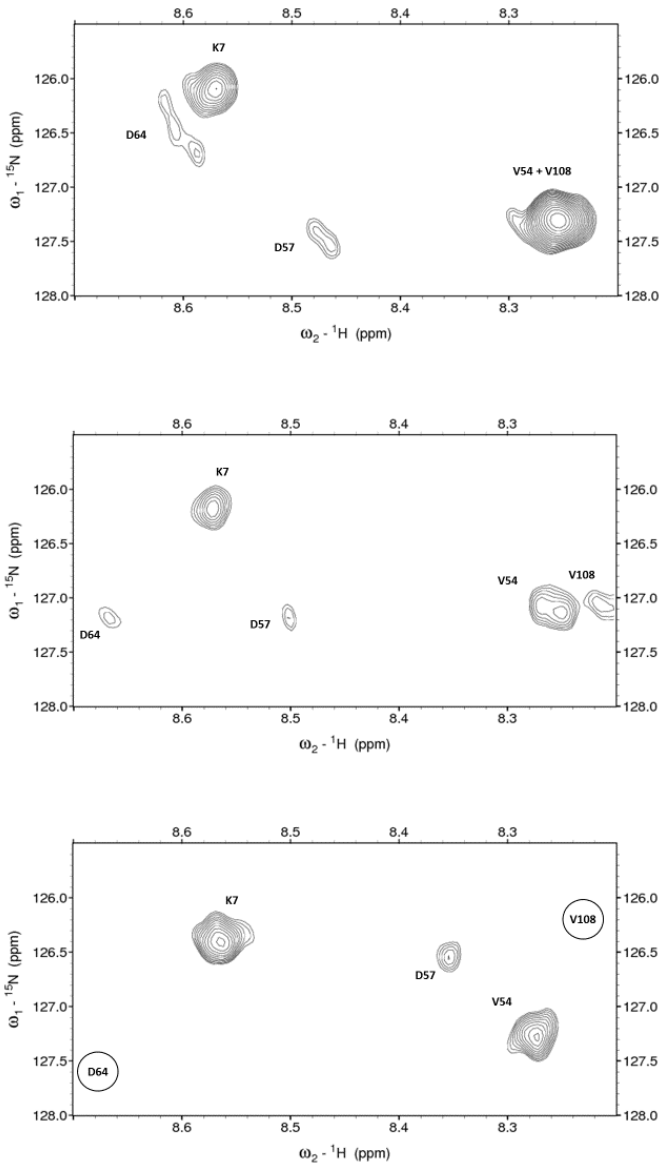


Figure 4.10: NMR HSQC spectral region including the site of phosphorylation (D57) for apo-CheY (top), CheY-Mg^{2+} (middle) and Chey-BeF_3^- (bottom). *Resonances are labelled with spectral assignments. Circle assignments indicate that the resonance has broadened beyond the viewing area. Square assignments indicate the resonance was not resolved in the individual measurement.*

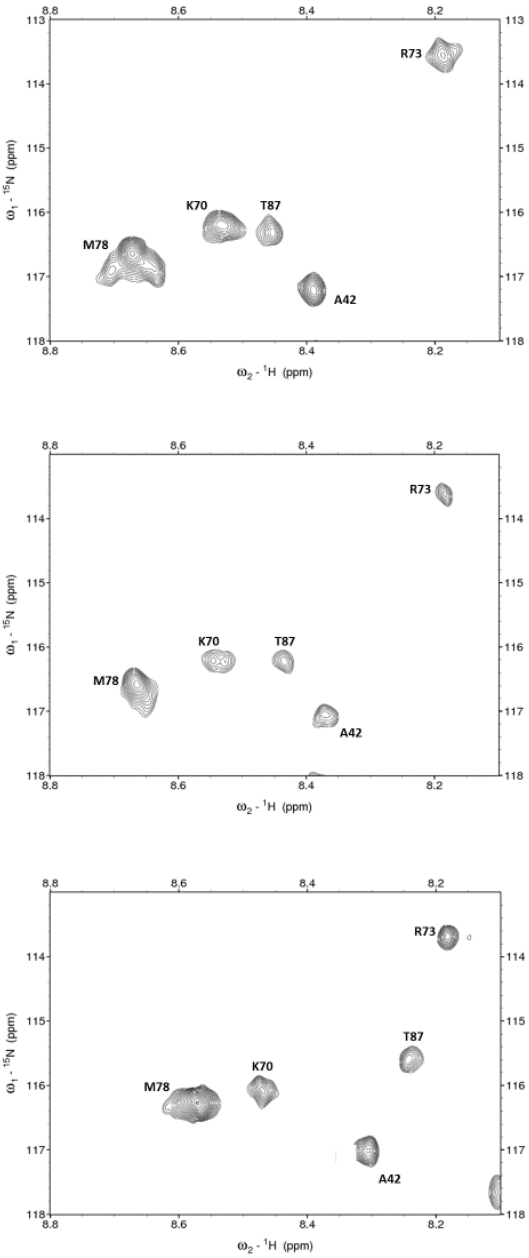


Figure 4.11: NMR HSQC spectra for apo-CheY (top), CheY-Mg^{2+} (middle) and Chey-BeF_3^- (bottom). *Resonances are labelled with spectral assignments. Circle assignments indicate that the resonance has broadened beyond the viewing area. Square assignments indicate the resonance was not resolved in the individual measurement.*

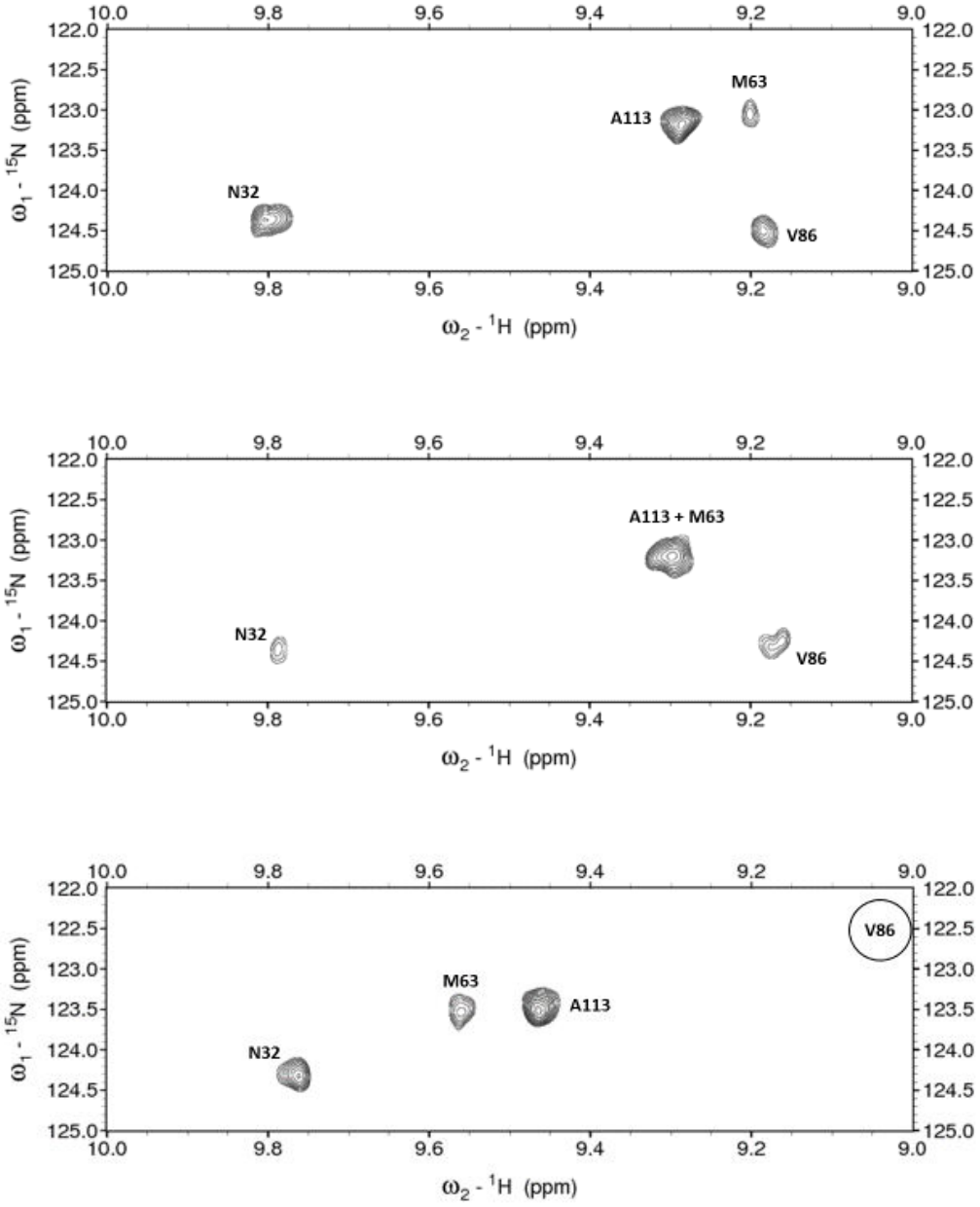


Figure 4.12: NMR HSQC spectra for apo-CheY (top), CheY-Mg^{2+} (middle) and Chey-BeF_3^- (bottom). *Resonances are labelled with spectral assignments. Circle assignments indicate that the resonance has broadened beyond the viewing area. Square assignments indicate the resonance was not resolved in the individual measurement.*

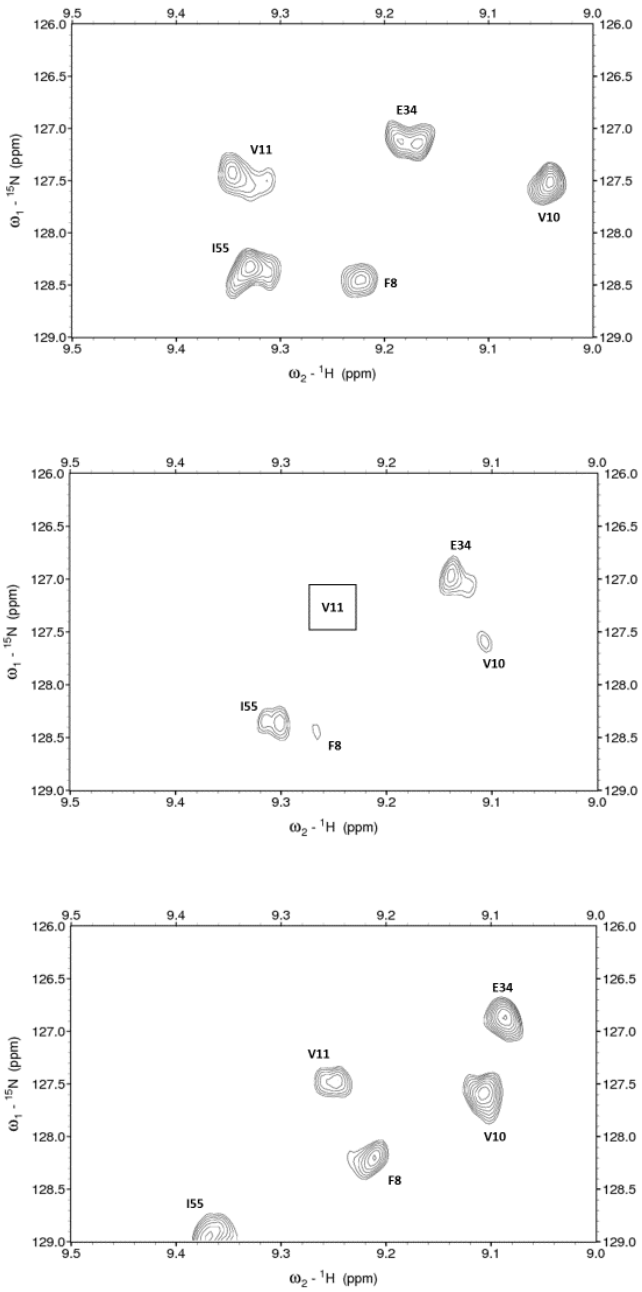


Figure 4.13: NMR HSQC spectra for apo-CheY (top), CheY-Mg^{2+} (middle) and Chey-BeF_3^- (bottom). *Resonances are labelled with spectral assignments. Circle assignments indicate that the resonance has broadened beyond the viewing area. Square assignments indicate the resonance was not resolved in the individual measurement.*

4.2.5.3 OKE Spectroscopy

Previous crystallography and NMR measurements have shown that a conformational change takes place when CheY binds the Mg^{2+} cofactor and subsequent changes follow upon phosphorylation (Cho *et al.,* 2000; Lee *et al.,* 2001). Our own NMR data confirmed that these changes occur under identical conditions used for OKE measurements, yet we were unable to observe any detectable change in the OKE spectra between the different proteins states, only very slight changes in amplitude over the whole spectrum (Fig. 4.14).

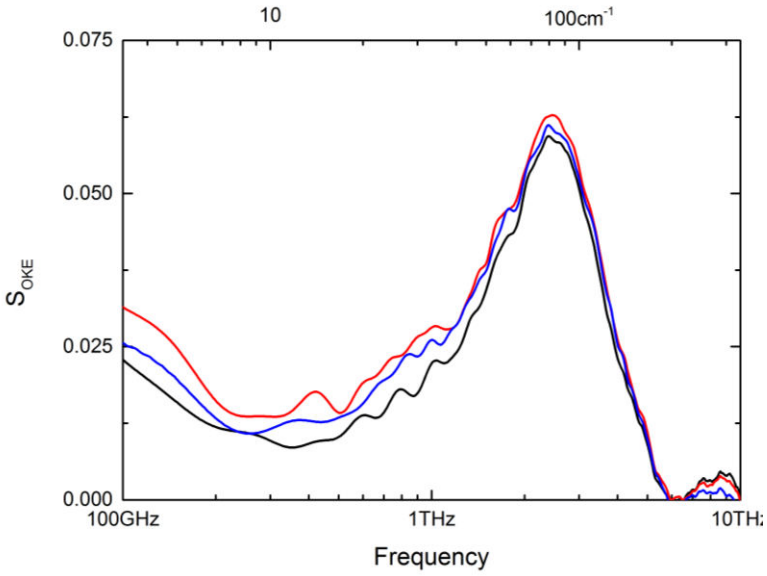


Figure 4.14: OKE spectra of apo-CheY (black), CheY-Mg^{2+} (red) and CheY-BeF_3^- (Blue)

Although conformational changes have previously been detected between apo-CheY and BeF_3^-—CheY, these changes are not vast and mainly localised around the active site and the allosteric pathway through the protein. For example, the largest change induced by phosphorylation is the rotation of Y106 to a buried position, facilitated by the motion of the β4-α4 loop consisting of residues 88-91 (Ma & Cui, 2007; Simonovic & Volz, 2001), which allows CheY binding to FliM. However, apart from

these changes, only very small variations are seen in the β5-α5 loops, side chain motions of K109 and F14 and the altered hydrogen bonding of some residues directly adjacent to the site of phosphorylation (Appleby & Bourret, 1998; Lee *et al.,* 2001). McDonald *et al.* (2013) used NMR spectroscopy to determine how these changes between apo-CheY and BeF_3^- – CheY influence the protein dynamics on both the microsecond-to-millisecond (μs-ms) and also the picosecond-to-nanosecond (ps-to-ns) timescales (corresponding to terahertz frequencies) and found limited changes in the picosecond backbone dynamics upon phosphorylation. The authors also measured the ps-to-ns motions of the side chains, which can reveal dynamics that are invisible to the backbone as these motions are more heterogeneous and sensitive to perturbation (Igumenova *et al.,* 2006). The largest change in the picosecond side chain dynamics was detected as a localised increase in rigidity around the phosphorylation site, but a slight increase in flexibility of the protein overall was also identified. Stiffening of the active site is to be expected as binding pockets typically become more rigid due to the formation of new bonds, and the increased flexibility detected elsewhere is thought to compensate for this entropic penalty (Lee *et al.,* 2000; Loh *et al.,* 2001).

CheY appears to be unique in its ability to restrict the fast dynamic changes to near the phosphorylation site as similar reports on the fast dynamics of catabolite activator protein (Tzeng & Kalodimos*,* 2012), calmodulin (Frederick *et al.,* 2007) and a PDZ domain (Petit *et al.,* 2009) described propagation throughout the molecule. We have previously shown that the OKE spectrum of lysozyme is sensitive to delocalised vibrations (Turton *et al.,* 2014) and will not detect single amino acid fluctuations and therefore, may not be sensitive to the localised changes seen in CheY. Why the fast dynamics of CheY are localised to the active site and along the path of allostery is not known, yet we suggest this is the reason why we detect similar terahertz spectra for the native and active protein.

4.3 Example 3: Terahertz Study of Signal Transduction

4.3.1 Introduction

The *Escherichia coli* transcriptional activator protein, Catabolite activator protein (CAP), also known as cAMP receptor protein (CRP), is one of the best examples of transcriptional regulation in prokaryotic cells. CAP activates transcription at more than 100 promoters in the presence of the allosteric effector, cAMP, which induces a conformational change to increase the protein's affinity for specific DNA sequences, which encourages binding and transcription by the RNA polymerase holoenzyme (RNAP) (Fig. 4.15). Transcriptional activation mediated by CAP is often perceived as a model system for structural and mechanistic studies owing to the relative simplicity of CAP mediated transcription - at the simplest promoters, CAP only requires two additional macromolecules (RNAP and DNA) to initiate transcription. This is simpler than most prokaryotic transcriptional activation and considerably less complex than eukaryotic regulation (usually requiring 10s of macromolecules to control transcription). In addition, it has been suggested that CAP displays a change in picosecond dynamics upon binding of cAMP (Tzeng & Kalodimos, 2012), making CAP an ideal protein for OKE analysis.

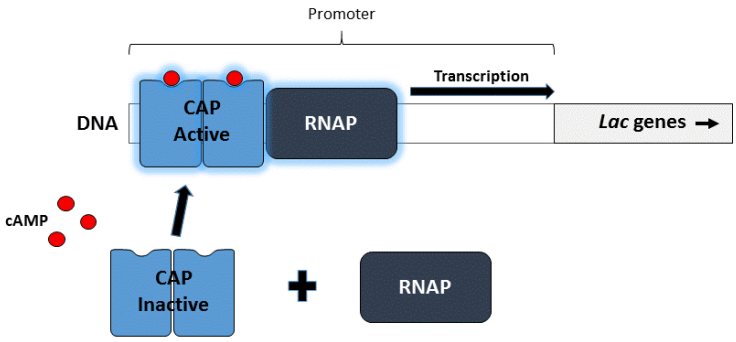


Figure 4.15: CAP mechanism of action

4.3.2 Structure

CAP is a homodimer of 47,238 Da with each subunit consisting of 209 amino acids and possessing two domains (Fig. 4.16). The larger N-terminal domain is composed of three α-helices and eight β-strands that form into a β barrel structure that is the site of cAMP binding in the primary '*anti'* conformation. This domain is also responsible for the subunit dimerisation and shows sequence homology to other cAMP binding proteins, including ion channels and protein kinases (Shabb & Corbin, 1992). The smaller C-terminal domain is responsible for DNA binding and possess three α-helices and four β-sheets, including a helix-turn-helix motif that is characteristic of DNA binding proteins and displays amino acid similarity to other regulatory proteins such as the *cro* repressor family (Ebright *et al.,* 1992). Each CAP subunit also possess a secondary cAMP binding site in the '*syn'* conformation, which is situated between the two domains.

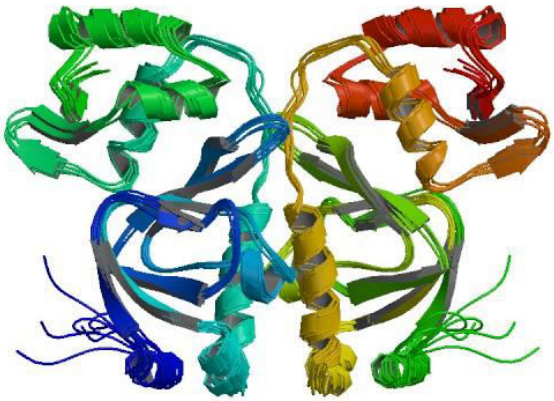


Figure 4.16: Ribbon diagram of CAP in the unlighted state (PDB = 2WC2)

4.3.3 Results and Discussion

4.3.3.1 Sample Preparation

Plasmid pAKCRP-His6 encodes CAP-His6 under the control of bacteriophage T7 gene 10 promoter. Plasmid pAKCRP-His6 was constructed from plasmid pAKCRP (Kapanidis *et al.,* 2001) by use of site-directed mutagenesis (Kunkel *et al.,* 1991) to insert six His codons after codon 209 of the crp gene. Therefore, the His-tag is located on the DNA binding domain and does not possess a cleavage site and cannot be removed from the protein. However, the tag does not alter protein conformation, function or hinder DNA binding in any way (Tzeng *et al.,* 2009; Tzeng at al 2013).

CAP-His6 was first purified by Ni^{2+}-affinity chromatography followed by size exclusion chromatography to remove any additional impurities and the final protein is thought to be >95% purity (Fig. 4.17). Apo-CAP was found to be fairly soluble and the concentration could be pushed close to 2 mM (90 mg/ml), which we have determined as adequate for detecting changes in protein dynamics with OKE spectroscopy. However, upon addition of cAMP the solubility was found to drop 6-fold to approximately 300 µM (15 mg/ml), which is at the limit for detecting protein changes using OKE spectroscopy. Nevertheless, we proceeded with measurements at

a concentration of 300 μM in the hope that the conformational changes occurring throughout CAP produce a large OKE response.

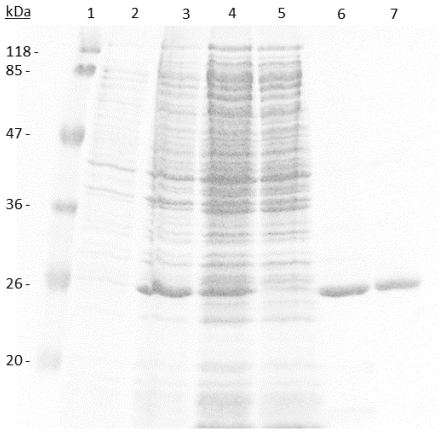


Figure 4.17: Stages of CAP purification. *Lanes: (1) EZ-RUN ladder (2) Before induction with IPTG (3) After induction with IPTG (4) Crude (5) Flow Through (6) Nickle affinity purification (7) Size exclusion chromatography.*

4.3.3.2 Conformational change confirmed by protease digestion

CAP contains two cAMP binding pockets on each subunit, *anti* and *syn,* and there are three cAMP-dependent conformers that exist: apo-CAP, which predominantly exists in solutions devoid of cAMP, CAP-$(cAMP)_2$, which dominates in solutions with micromolar concentrations of cAMP and CAP-$(cAMP)_4$, which is the dominant conformer in the presence of millimolar concentrations of cAMP (Takahashi *et al.,* 1989; Heyduk & Lee, 1989). The binding of cAMP takes place in a sequential manner and displays negative cooperativity. Firstly, the *anti* sites located on the N-terminal domain are occupied with a dissociation constant of 27.5 x 10^{-6} M^{-1} and once these sites are saturated, the *syn* sites on the C-terminal domain are loaded with a dissociation constant of 2.0 x 10^{-3} M^{-1} (Malecki *et al.,* 2000). Apo-CAP and CAP-$(cAMP)_2$ appear to be the biologically relevant forms and bind DNA with low and high affinity and specificity, respectively (Tutar, 2008a; 2008b). However, the

concentrations of cAMP required to achieve the CAP-$(cAMP)_4$ conformation are not typically found intracellularly (Epstein *et al.,* 1975) and DNA binding does not occur or is severely inhibited at these concentrations (Mukhopadhyay *et al.,* 1999).

The ligand, cAMP, contains a purine group in the form of adenine, which we would expect to give a large OKE response due its high polarisability. Therefore, it was necessary to determine the minimum concentration of cAMP that could be used to provide the fully active form of the protein while minimising any contribution to the OKE spectrum and also confirming the functionality of the purified protein. cAMP binding results in the reorientation of the CRP subunits and also reorientation of the domains within each subunit (Harmen, 2001), which makes apo-CAP and CAP-$(cAMP)_4$ resistant to proteases, whereas CAP-$(cAMP)_2$ is susceptible to digestion (Botsford & Harman, 1992; Kolb *et al.,* 1993). Therefore, we utilised protease digestion with α-chemotrypsin in order to determine the required cAMP concentration for each state. From our results we can see that at a CAP (monomer) to cAMP molar ratio of 1 : 1.5 the protease digestion appears to be close to maximum (Fig. 4.18), signifying that the *anti* sites are fully loaded and the protein is in its active form. Therefore, we used this ratio for OKE measurements.

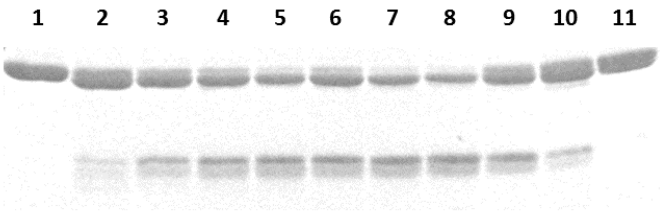


***Figure 4.18**: **CAP digestion with α-chemotrypsin.** The molar ratio of each well represents CAP (monomer) to cAMP. (1) No protease (2) 1 : 0 (3) 2 : 1 (4) 1 : 1 (5) 2 : 3 (6) 1 : 2 (7) 1 : 5 (8) 1 : 25 (9) 1 : 100 (10) 1 : 500 (11) 1 : 1000. The lower band shows digested fragments of the CAP protein. Therefore, the intensity of this band represents the degree of digestion and hence, conformational change.*

It is also interesting to note the extremely high concentrations of cAMP required to induce the *syn* conformation. Only when the molar concentration of cAMP was approximately 500 times higher than that of the CAP monomer did we begin to see the second conformational change that gives protease resistance (Fig. 4.18). Intracellular concentrations of cAMP never reach these levels (Epstein *et al.,* 1975) and it is therefore unknown why CAP possess the *syn* binding sites. It has been hypothesised that CAP may be stabilised in the presence of RNAP and therefore bind in the *syn* conformation at lower concentrations or CAP could possibly associate with adenylate cyclase, which may raise the cAMP concentration to high levels in the local environment (Kolb *et al.,* 1993). However, the high concentration of cAMP necessary to induce this second conformational change would mask the protein OKE signal and we therefore did not measure this change.

4.3.3.3 OKE Spectroscopy

The binding sites for the *anti* conformation are located 10 Å from the DNA-binding site, which is too far to directly influence the interaction between the protein and nucleic acids (Passner *et al.,* 2000) and the question still remains as to how the information is transferred from the N-terminal domain to the C-terminal domain. Two ways have been proposed of how the allosteric signal can be transmitted across the protein: the signal may be transmitted through the CRP hinge to the DNA binding domain or local changes in the N-terminal domain result in a modification of the interdomain interactions (Garges & Adhya, 1985; Weber & Steitz, 1987). These delocalised changes have been shown to occur on a picosecond timescale (Tzeng & Kalodimos, 2012) and from our previous measurements that show the ~3 THz region to be sensitive to delocalised motions in the enzyme lysozyme (Turton *et al.,* 2014), we would expect to detect the propagation of the allosteric signal in a similar frequency range for CAP. However, in our measurements no significant difference was observed between apo-CAP and CAP-$(cAMP)_2$ across the whole frequency range (Fig. 4.19).

Previous OKE studies (Hunt *et al.,* 2007) and our own measurements on protein denaturation have shown that OKE spectroscopy is sensitive to changes in protein secondary structure and circular dichroism measurements indicate that the binding of cAMP in the *anti* -conformation causes a 4% increase in the α-helical content of the N-terminal compared to the apo- form (Fic *et al.,* 2009). Comparisons between apo-CAP and CAP-$(cAMP)_2$ have found that the main differences occur on residues Gly71, Glu72, Arg82, Ser83, Arg123, Thr127 and Ser128 where this region is composed of mainly β structure for the cAMP loaded protein but α-helix or turn structure for apo-CAP (Fic *et al.,* 2009). Furthermore, the α-helix content on the surface of each subunit is also seen to change between the two forms (Dong *et al.,* 2002). However, the secondary structure changes we have previously detected with OKE spectroscopy were induced by denaturation and therefore occurred throughout the whole molecule. In comparison, the changes occurring in CAP are small and possibly beyond the detection capabilities of OKE spectroscopy.

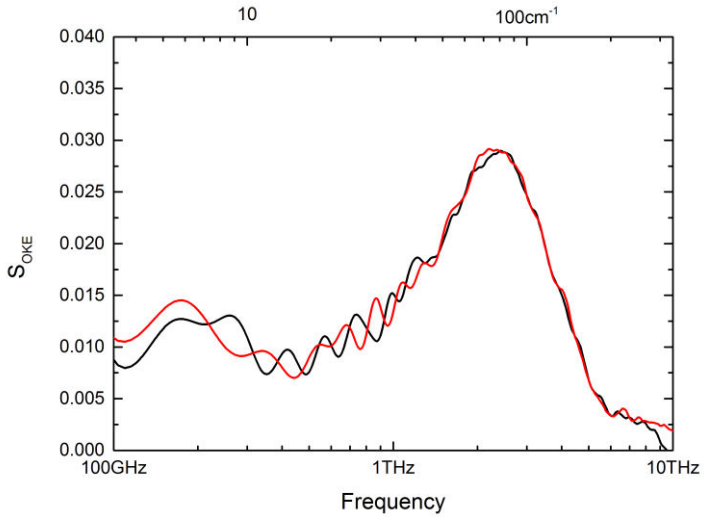


Figure 4.19: OKE spectra of apo-CAP (black) and CAP-$(cAMP)_2$ (red)

Larger changes in tertiary structure have also been detected between apo-CAP and CAP-$(cAMP)_2$ by fluorescence decay anisotropy and dynamic light scattering measurements (Blaszczyk *et al.,* 2001), which concluded that CAP-$(cAMP)_2$ adopts a more compact structure than apo-CAP. This was confirmed by more recent Forster resonance energy transfer studies that revealed cAMP binding causes a decrease in distance between Trp85 and Cys178, which would suggest a rearrangement of the N- and C-terminal domains, converting the protein from an open to closed conformation for apo-CAP to CAP-$(cAMP)_2$, respectively (Fic *et al.,* 2006; Polit *et al.,* 2003). This reduction in size results in increased tumbling of CAP-$(cAMP)_2$ compared to apo-CAP, which we would expect to influence the molecular ordination diffusion, which we may detect as a change in the low frequency region of our OKE spectrum, yet we see no measurable difference between the two forms.

Although there are considerable differences present between apo-CAP and CAP-$(cAMP)_2$ that we would hope to detect in our OKE measurements, the low protein concentration used in these measurements undoubtedly reduce our chances of detection. From previous measurements, we can see that the concentrations used here are at the very limit of what OKE spectroscopy can detect with regards to protein dynamics. Although it is difficult to predict the OKE signal each protein will produce, we can assume that the low concentration hindered our ability to detect significant changes between the CAP conformers.

4.4 Conclusions

We have chosen three model systems to characterise using OKE spectroscopy. Our measurements on lysozyme and NAG_3 have shown that terahertz OKE spectroscopy is sensitive to changes in protein dynamics. On addition of the inhibitor, a blue shift is observed from 1.29 to 2.89 THz accompanied by an anomalous strengthening of the high frequency mode, which must be due to an increase in the polarisability. As the difference is induced by the binding of NAG_3 to lysozyme, but is not present in the solution of NAG_3 only, the mode must be part of the binding coordinate and hence, for the first time, we have identified a protein terahertz mode of biochemical significance. The modes observed here in the ligand-bound complex corresponding to the binding coordinate are significantly underdamped, with the damping parameter γ about half the value of the frequency ω_0. It seems likely that proteins have evolved such underdamped vibrational motions by providing an interaction site, which is pre-organised (Roca *et al.,* 2006) in order to arrive at the optimal degree of damping for efficient biochemical reactivity and function.

We did not identify any significant changes upon analysis of the other model proteins. However, phosphorylation of CheY is seen to induce changes localised to the active site and allosteric pathway that occur mainly at low frequencies (*i.e.* MHz) (McDonald *et al.,* 2013). Therefore, we assume the dynamics of CheY are not accessible with our OKE measurements due to a lack of delocalised changes at terahertz frequencies. In contrast, CAP should be an ideal protein to study using terahertz spectroscopy as it is known to undergo large conformational changes and display dynamics on a picosecond timescale (Tzeng *et al.,* 2013). However, because of the signal produced by the solvent, the study of proteins in solution using OKE spectroscopy requires a high concentration of protein to detect significant changes. As a result, the low solubility of CAP hampered our measurements and no significant change could be detected.

Chapter 5:
DNA & Nucleic Acid Components

Chapter 5 DNA & Nucleic Acid Components

Deoxyribonucleic acid (DNA) is the hereditary macromolecule that holds genetic information and codes for all biological processes throughout the domains of life. It is therefore not surprising that there is considerable interest in studying the vibrational dynamics associated with these polymer molecules by spectroscopic means. Many spectroscopic studies (Maret *et al.,* 1979; Liu *et al.,* 1989; Weidlich *et al.,* 1990; Sarkar *et al.,* 1997) and theoretical (Mei *et al.,* 1981; Saxena *et al.,* 1991) work has been carried out in an attempt to comprehend the spectral data and assign the origins of the vibrational modes. In particular, the submillimetre-wave frequency regime (0.01 – 10 THz) is proposed to be rich in spectral features that stem from internal vibrations that are spread over large portions of the DNA chain, which may represent information related to the primary sequence of the DNA molecule (Saxen *et al.,* 1991; Van Zandt *et al.,* 1992).

Some of the earliest evidence that DNA could be analysed in the terahertz range was revealed by Wittlin *et al.* (1986), over the frequency range of 0.1-13.5 THz. They carried out research on highly orientated, dry films of lithium-DNA and sodium-DNA in the temperature range of 5-300K and identified 5 vibrational modes, including the lowest frequency modes for Li-DNA (1.35 THz) and Na-DNA (1.23 THz). Upon analysis, these modes were found to red shift upon hydration and a simple lattice dynamical model was used to explain the presence of these vibrations, and also the influence of hydration on these modes. This emphasised the potential of terahertz radiation for analysis of DNA and the following year Powell *et al.,* 1987 performed transmission measurements on poly(dA).poly(dT) and poly(dG).poly(dC) DNA and demonstrated that temperature, crystallinity and salting conditions have a pronounced effect on the spectra. The authors revealed four sharp bands are present for double

stranded helical poly(dA).poly(dT) at 63, 83, 100 and 110 cm^{-1} - at 7K under minimal salt and high polycrystalline conditions.

More recently, the advent of time-domain terahertz techniques and increased sensitivity of low frequency FTIR measurements have permitted various submillimetre investigations of polynucleotides. For example, THz-TDS imaging methods have been carried out on two single stranded artificial RNA chains composed of polyadenylic acid (poly-A) and polycytidylic acid (poly-C) (Fischer *et al.,* 2005). In this study, the absorption coefficient and refractive index of poly-A and poly-C, over the frequency range from 0.1 – 4.0 THz, were investigated by measuring thin films and pressed pellets. Poly-C was found to absorb more strongly and have a higher refractive index than poly-A, irrespective of sample preparation, and both samples displayed an unstructured absorption that increased with frequency. Similarly, Markelz *et al.* (2000) measured calf thymus DNA at room temperature, <5% relative humidity (r.h.) and observed a featureless absorbance spectrum that increased linearly with frequency. Fischer and co-workers argue that no spectrally resolvable resonances can be observed in the amorphous condensed phase due to the high density of IR-active modes and the high sensitivity of these modes to the local environment, which leads to the absorption profile simply following the spectral density of the vibrational modes. In addition, vibrational modes are not expected to display narrow linewidths in these samples due to damping caused by additional friction and intermolecular interactions from nearby neighbouring molecules. However, additional measurements carried out by Markelz *et al.* indicate that the hydration of DNA results in absorption features shifting to lower frequencies and becoming visible, which is consistent with a previous study that showed a Raman-active mode at 25 cm^{-1} to blueshift and broaden as DNA is dehydrated until the mode is virtually indiscernible at 0% relative humidity (Weidlich *et al.,* 1990). It is also worth noting that above 92% r.h. DNA forms the standard B conformation whereas between 45 % and 92 % r.h. it adopts an ordered helical A conformation and the conformation is disordered at lower r.h. (Lindsay *et al.,* 1988) (Fig. 5.1). Markelz may have detected this conformational change as the

spectrum of DNA taken at <5 % r.h. is distinctly different from that taken at 70 % r.h, while the spectrum of DNA at 43 % r.h – near the transition point – displays features from both.

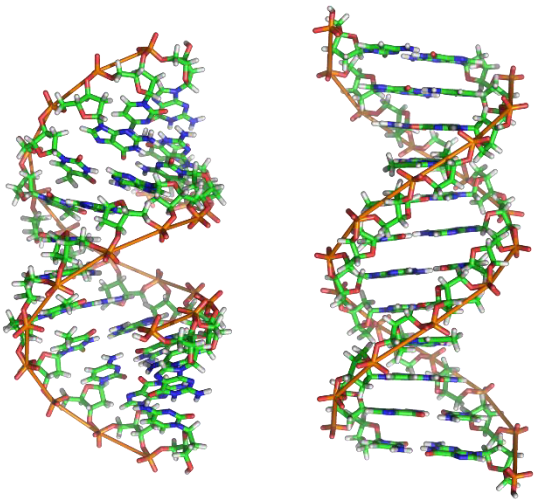


Figure 5.1: DNA Structures. *A-DNA (left) is the form of the molecule under dehydrated conditions and also in RNA-RNA and RNA-DNA complexes. It is shorter and wider with base pairs tilted relative to the helix axis compared to B-DNA (right), which is the form adopted at high water concentrations in DNA-DNA complexes.*

Globus and co-workers at the University of Virginia have perhaps carried out the most extensive experimental and theoretical research into the dynamics of DNA at terahertz frequencies, employing an IFS-66v FTIR equipped with a helium cooled Si-bolometer for high sensitivity measurements. They typically measure polynucleotides over a short frequency range (10-25 cm^{-1}) and claim to identify a high density of reproducible IR-active modes that are sensitive to a number of factors involving the sample, including; preparation techniques, thickness, orientation, water content and age. In early studies, various methods were used to prepare DNA samples of different thicknesses, which were measured as free-standing thin films or applied to near transparent substrate (Globus *et al.,* 2002) and displayed reproducible spectral features

that the authors attribute to multiple dielectric resonances - phonon modes. The same method was applied to single and double stranded artificial RNA molecules, with good agreement in absorption coefficient shown between theoretical and experimental data (Globus *et al.,* 2003). During ensuing studies, DNA was also measured in the liquid phase where similarities could be seen between the solid and liquid samples, although absorption is stronger in the liquid due to the water effect (Globus *et al.,* 2004). In fact, by dissolving DNA in water or buffer solution at a concentration of 50-200 mg/ml, they produced gels, which form an ordered liquid-crystalline phase that could be partially aligned by squeezing the gel through a syringe-needle onto the substrate in one direction. In agreement with a prediction presented by Froehlich (1983), they reported an increased intensity and sharper vibrational modes in the liquid environment compared to the solid phase (Globus *et al.,* 2006). The authors could not define the origin of the modes detected, but previous calculations have predicted that excitations throughout the DNA molecule contribute, for example: propeller-twist, hydrogen-bond breathing and base-roll and base-shift vibrational modes (Van Zandt & Saxena*,* 1989; Zhuang *et al.,* 1990). Even in recent studies, it is still difficult to successfully model DNA (or any macromolecules) in the terahertz range but it seems likely that molecular dynamics simulations show better correlation to experimental data than normal mode analysis as the former can better account for anharmonicity associated with low frequency modes (li *et al,* 2008).

Although there has been progress in the ability to analyse DNA molecules in the 'liquid phase' utilising standard THz-TDS and FTIR techniques, these measurements are typically performed on orientated DNA gels, which does not reflect a biological environment and also alter the conformation of the molecules. Furthermore, even under these conditions, limited vibrational modes have been identified and their origins unclear. Therefore, for the first time, we have carried out an extensive terahertz OKE study of DNA in solution. First, we analyse the components of DNA in isolation and also combined with their complimentary base pairs to determine if we can extract any features specific to the bases or if hydrogen bonding can be detected. After which

we analyse short oligonucleotides of repeating G-C or A-T bases, plasmid DNA and genomic DNA to establish if changes in sequence, length and structure can be detected in the terahertz range.

5.1 DNA Structure

5.1.1 Nucleic Acid Components

5.1.1.1 Nucleobases

Purine and pyrimidine rings that form the basis of nucleic acids have low solubility in water. Nucleobases have additional polar side groups that slightly increases their solubility, yet it was necessary to heat the samples to 50°C in order to reach a concentration of 50 mM for our measurements. Even at this low concentration the highly polarisable nucleobases give a significant OKE response and are distinct from the water spectrum (Fig. 5.2). However, the nucleobases all display very similar spectra except at low frequency where we detect an increase in signal for adenine and guanine compared to cytosine and thymine in the solvated spectrum. In the solvent-free spectrum this change becomes less pronounced but it is still present and may be explained by the increased polarisability of the purine compared to the pyrimidine rings (Riahi *et al.,* 2010). The librational band common to all small molecules (~2.5 THz) is most pronounced for adenine and least intense in the thymine spectrum, which might be surmised to stem from the differences in the purine and pyrimidine rings. However, Cytosine (pyrimidine) displays a larger intensity than guanine (purine) in the same region, which contradicts this hypothesis.

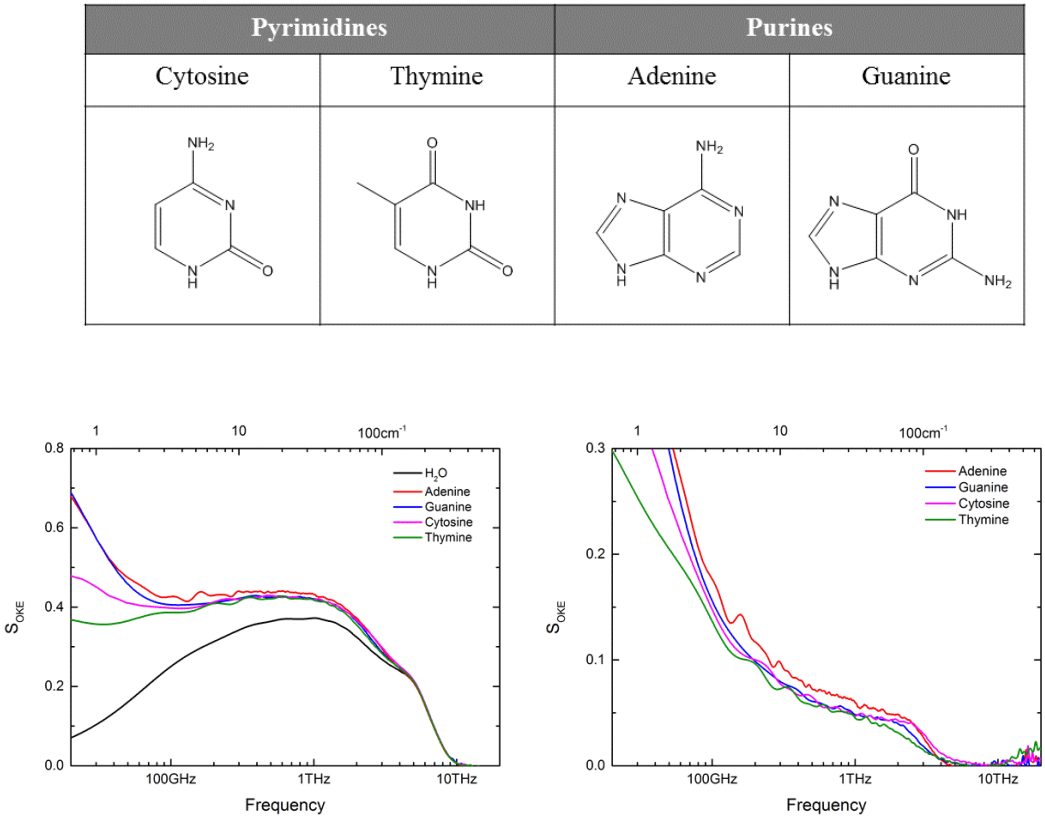


Figure 5.2: Structure and OKE spectra of solvated (left) and solvent subtracted (right) nucleobases

Due to the poor solubility of the nucleobases, it is unlikely that any significant interactions occur between the molecules at the concentration used in these measurements (50 mM). Hydrogen bonding and base stacking are essential interactions in the DNA double helix and H-bonding between bases is predicated to be visible in the terahertz range (Zhang *et al.,* 1998; Fischer *et al.,* 2002). Therefore, we utilised the more soluble nucleotide monophosphates (NMP's) in order to determine if we could detect any additional interactions by working at higher concentrations.

5.1.1.2 Nucleotides

Nucleotides are considerably more soluble than their bases as the ribose sugar presents numerous hydroxyl groups that can partake in hydrogen bonding and the phosphate groups adds a negative charge. We tested the least soluble NMP, GMP, to have a solubility limit of 1.6 M in pure water. Thereafter, we used this concentration to carry out OKE measurements on aqueous solutions of single NMP's and also mixtures of complimentary NMP's.

We measured AMP and TMP at 1.6 M and found that the spectra look very similar, dominated by a very broad band centred at ~1.5 THz (Fig. 5.3). The AMP spectrum is also significantly stronger than TMP in this region, in agreement with our nucleobase measurements. The librational band is usually detected at approximately ~2.5 THz in the OKE spectrum of small biomolecules, yet for AMP and TMP this has moved to lower frequencies. In addition, our previous measurements have shown that with increasing concentration the librational band typically sharpens, yet the feature remains broad even at the high concentrations used here. When we performed measurements on equimolar solutions of complimentary AMP and TMP the spectrum appears identical to the individual NMP's with an averaged intensity and no additional features (Fig. 5.3).

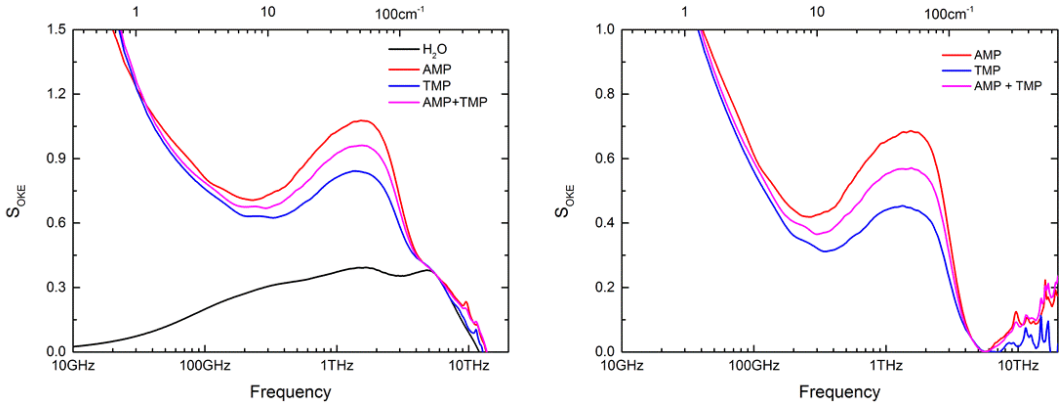


Figure 5.3: OKE spectra of solvated (left) and solvent subtracted (right) nucleotides. *AMP, TMP and an eqimolar concentration of both at 1.6 M*

The spectrum of CMP is more typical of what we would expect from the OKE spectrum of a small biomolecule at this concentration as it shows the librational band at ~2.5 THz as a sharp peak (Fig. 5.4). However, the spectrum of GMP is considerably different from the other NMP's and multiple new bands can be identified. For example, we detect the typical librational band as a strong peak at 2.3 THz but also additional shoulders on either side of this feature at 3.4 THz and 1.7 THz. At lower frequencies we also detect a very weak band at 0.7 THz and a strong feature centred at 0.35 THz (Fig. 5.4). However, when the CMP and GMP are combined in an equimolar ratio, the additional features detected for GMP disappear. This would suggest these features stem from GMP self-association and are concentration dependent, which was confirmed by performing measurements on 1 M GMP solutions (Fig. 5.5). Although the features are still present in the dilute sample, they are significantly weaker and there is an overall change in the spectral profile. To determine if the additional interactions seen for GMP were present for other NTPs at higher concentrations we carried out measurements on saturated solutions of AMP (3.0 M) and CMP (3.2 M). As expected, the overall intensity of the spectra increased for these molecules but no additional spectral features could be identified (Fig. 5.6).

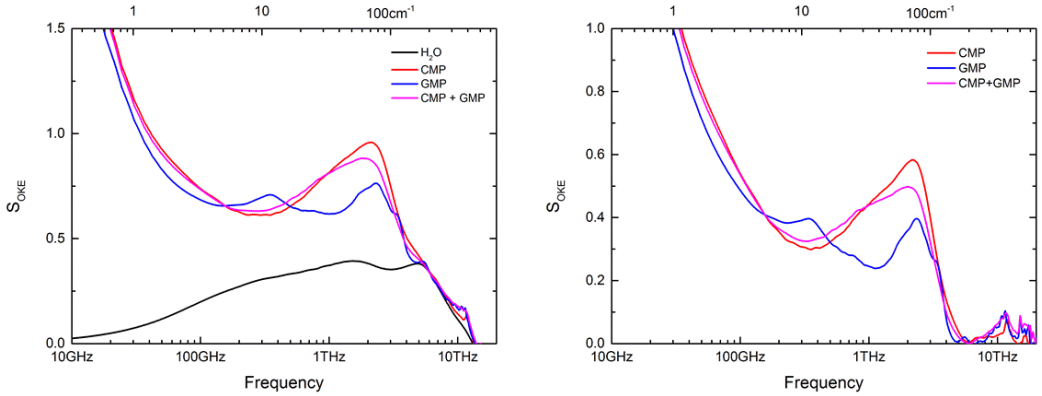


Figure 5.4: OKE spectra of solvated (left) and solvent subtracted (right) nucleotides. *CMP, GMP and an equimolar concentration of both at 1.6 M*

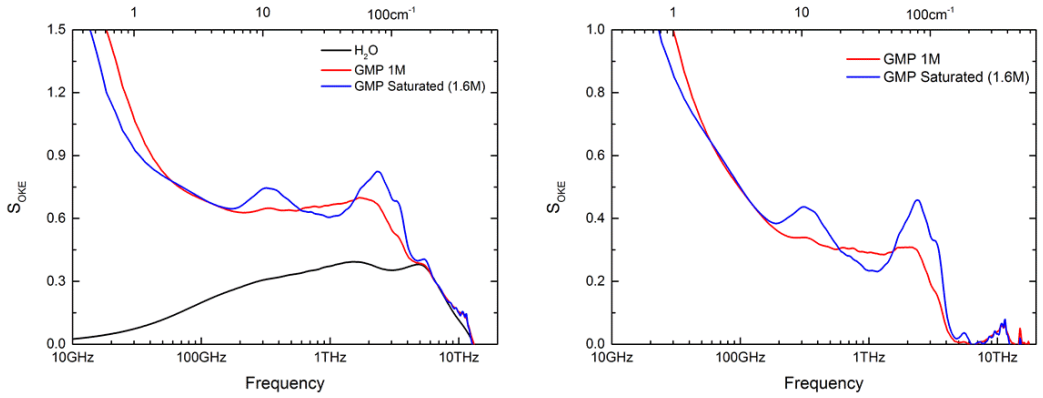


Figure 5.5: OKE spectra of solvated (left) and solvent subtracted (right) GMP. *GMP shows concentration dependent features at 1.6 M (saturated) that are lacking at lower concentrations*

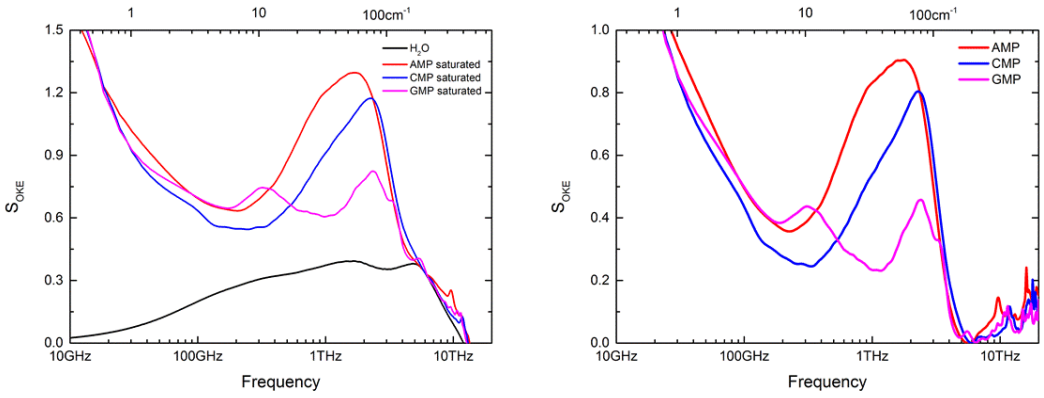


Figure 5.6: OKE spectra of solvated (left) and solvent subtracted (right) saturated aqueous nucleotide solutions. *AMP is saturated at a concentration of 3.0 M, CMP at 3.2 M and GMP at 1.6 M*

The large differences detected in the OKE spectrum of NTP's is unlike any changes we have detected for other biomolecules and can be, in part, attributed to these molecules ability to form multiple hydrogen bonds and also partake in π-π stacking. Hydrogen bonding is the dominant interaction between nucleobases in the gas phase (Yanson *et al.,* 1979) and also in organic solvents (Shoup *et al.,* 1966; Katz & Penman, 1966; Newmark & Cantor, 1968). However, NMR (Schweizer *et al.,* 1968), ultracentrifugation and UV (Solie & Schellman, 1968) data have shown that in aqueous solution π-π stacking is the main form of homo- and hetero- association between bases in agreement with molecular dynamics simulations (Pohorille *et al.,* 1984; Danilov *et al.,* 1997; Zendlova *et al.,* 2007). Proton magnetic resonance (PMR) spectroscopy has shown that at concentrations greater than 0.5 M, AMP displays stacking interactions in aqueous solution (Raszka & Kaplan, 1972), which suggests that our samples (1.6 M) contain soluble aggregates of AMP. This may explain the shift of the broad librational band for AMP compared to other small biomolecules as the large aggregates liberate at a lower frequency. PMR measurements of TMP have not been carried out but we would assume from the similarities in the OKE spectra that AMP and TMP display similar stacking properties. However, the association constant of purine-purine stacking is, on average, higher than that of pyrimidine-

pyrimidine (and purine-pyrimidine), which is largely governed by van der Waals forces, which increase with the dimensions of the aromatic portion of the interacting molecules (Ts'o, 1974; Nakano & Igarashi, 1970), hence it is surprising that AMP and TMP display such similar spectra. Nevertheless, the reduced stacking interactions of pyrimidines may explain why the CMP spectrum is comparable to that of other small biomolecules, such as amino acids, as the stacking interactions are reduced and aggregates do not readily form. GMP displays the strongest stacking interactions and aggregates can be detected at concentrations above 0.2 M, yet, the GMP aggregates are not fixed and undergo slow molecular exchange with free GMP In the solution (Raszka & Kaplan, 1972). This may explain why the GMP spectrum displays a sharp librational band at 2.3 THz and also a weaker shoulder at 1.7 THz, which represent librations of free GMP and GMP aggregates, respectively.

The strong band detected at 0.35 THz in the GMP spectrum is absent from the AMP spectrum and is therefore unlikely to stem from the stacking interactions and may possibly represent self-association hydrogen bonding. However, if this band were to originate from Watson-Crick hydrogen bonding, we would expect to detect this feature in the NTP mixtures, yet we do not. The fact that GMP possesses two hydrogen-bond donors, whereas the other NTPs only possess one may be a contributing factor to the additional feature detected in the GMP spectrum as hydrogen bonding may occur in different geometries between free monomers. Additionally, the amino group of GMP may promote additional self-association by hydrogen bonding between the stacked bases. Therefore, it is possible that the GMP 0.35 THz band does not represents Watson-Crick hydrogen-bonding in this case, but hydrogen-bonding between stacked molecules. Furthermore, studies have shown that when CMP is introduced into the solution it will intercalate into the GMP stack (Raszka & Kaplan, 1972), which would explain why this band is completely absent from the mixed solution, where CMP has disrupted the H-bonding between the GMP stack.

5.1.2 Oligonucleotides

Two sets of oligonucleotides were designed to be self-annealing; one composed of solely adenine and thymine (A-T's) residues and the other of only guanine and cytosine (G-C's) in the hope of determining if the different nucleotide structures can be detected using OKE spectroscopy and also if changes in hydrogen bond character can be identified – G-C base pairing forms three hydrogen bonds, whereas the A-T forms only two. The oligos were not designed as simple repeating units, but had a precise sequence in order to minimise mis-pairing between the bases and formation of 'overhangs' while still remaining self-annealing (Table. 5.1).

Table 5.1: Sequence of 20mer Oligonucleotides used for OKE analysis

Length (Bases)	Sequence (5' -> 3')
20	GGCGGCCCGCGCGGGCCGCC
20	AATAATTTATATAAATTATT

The OKE spectra for both oligonucleotides display similar characteristics, which may not be surprising as both chains are identical in length and expected to adopt the B-DNA structure in aqueous solution. In the 2-3 THz region the oligos display the 'boson peak' that is present in all biomolecular OKE measurements with the A-T oligo exhibiting an increase amplitude compared to G-C (Fig. 5.7). Interestingly this feature appears to display an unusual spectral profile at ~2.5 THz where we usually see a rounded or sharp feature, especially at higher concentrations. This feature can be detected in both oligo spectra and although it may appear subtle, it is reproduced in all measurements and is more pronounced for the A-T data. The spectral shape of the solvent subtracted spectra in this range are similar to the saturated GMP spectrum, yet we identify individual peaks for the oligonucleotides.

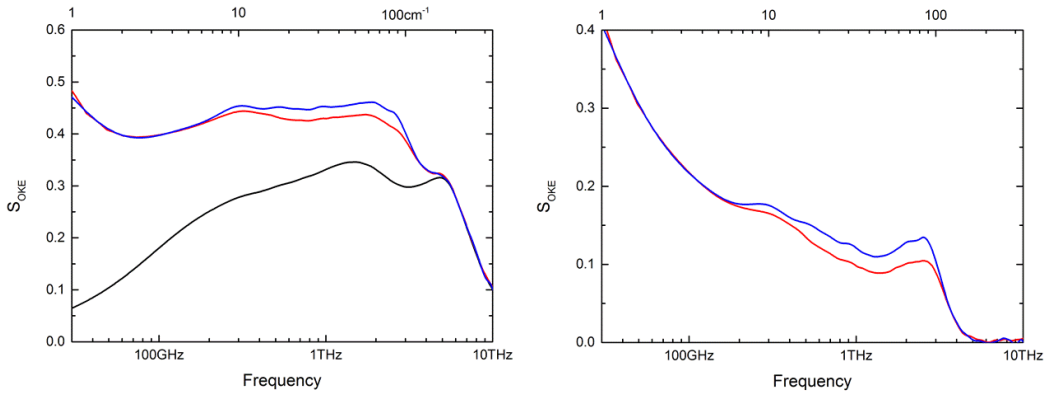


Figure 5.7: OKE spectra of solvated (left) and solvent subtracted (right) aqueous oligonucleotide solutions. *The oligos are 20mer repeating units of A-T's (blue trace) and G-C's (red trace) measured at a concentration of 10 mM (50 mg/ml).*

We carried out measurements on concentrated samples of the AT oligo and also subsequent dilutions, which indicated that this feature may stem from two peaks in close proximity, at approximately 2.5 THz and 2.9 THz, that cause the unusual shape of this feature (Fig. 5.8). We can assume that the 2.5 THz peak comes from the librational motions of the nucleotide sub units. In addition, it has long been suggested that phonon modes (lattice vibrations) may be responsible for spectral features in this range, originating from internal vibrations that are spread over sections of the DNA molecule, which may explain the mode at 2.9 THz (Woolard *et al.,* 2002). In fact, inelastic x-ray scattering measurements identified base-pair opening phonons in double stranded nucleic acids that coincide well with our OKE data (2.9 THz = 12 meV) (Krisch *et al.,* 2006).

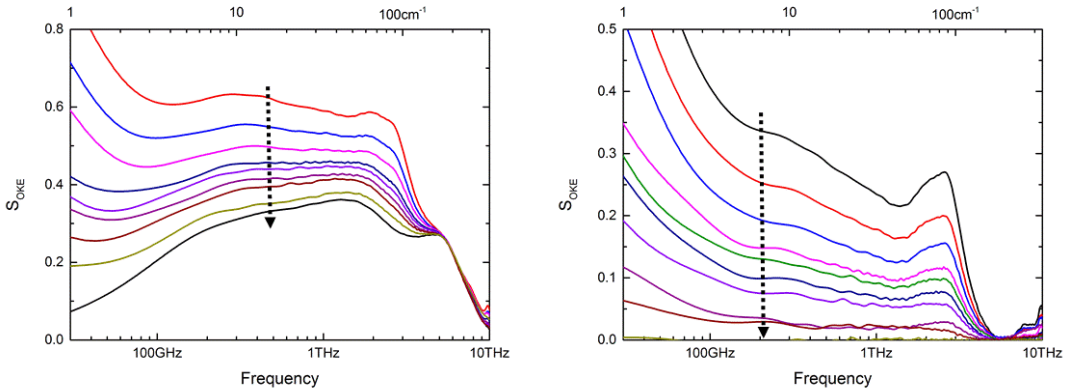


Figure 5.8: OKE spectra of solvated (left) and solvent subtracted (right) oligonucleotide dilutions*. The arrow indicates dilutions of the A-T oligo from 20 mM (100 mg/ml) to pure water.*

Interestingly, the band we detected at ~300 GHz at high concentrations of GMP is also apparent in both G-C and A-T spectra, although appears slightly more defined for the GC data. The origin of this mode cannot be attributed to phonons as the GMP mononucleotide displays a peak in the same position, demonstrating that this mode does not propagate over large sections of the oligonucleotide. Furthermore, base stacking interactions are important in maintaining the double helix structure of polynucleotides, yet our measurements on NTP's discounted this as the origin of the ~300 GHz feature. Numerous theoretical studies have predicted hydrogen bonding modes to be present in this low frequency terahertz region (Van Zandt and Saxena, 1989; Zhuang *et al.,* 1990) and we therefore assign the ~300 GHz band, in this case, to hydrogen bonding between base pairs of the oligonucleotides.

5.1.3 Plasmid & Genomic DNA

The measurements carried out on genomic and plasmid DNA had several drawbacks experimentally, that made it difficult to gather meaningful data. The large molecular structure of genomic DNA results in extremely viscous, gel-like solutions that made it difficult to reach the high concentrations required for OKE analysis, while

maintaining high optical quality for reliable measurements. This was overcome to some degree by extensive sonication of the sample in order to sheer the large genomic DNA into smaller fragments, yet it was not possible to reach concentrations above approximately 20 mg/ml. We measured the genomic DNA at this concentration and found the solvent subtracted spectrum to be featureless and display none of the bands displayed in the shorter oligonucleotide data (data not shown).

We also experienced concentration dependent issues with plasmid DNA and it was not possible to measure above approximately 2 mg/ml due to light scattering. However, unlike genomic DNA the issue was not due to viscosity as the sample appeared fluid. Instead, we believe that the scattered radiation is caused by the various supercoiled states adopted by the plasmid. Figure 5.9 shows that upon purification the plasmid adopts 6 different supercoiled states, which is shown from the multiple bands due to different migration speeds. We confirmed these bands to be supercoiled states of the same plasmid, and not genomic DNA contamination, by carrying out a restriction digest, which revealed a single band of high purity (Fig. 5.9). The various supercoiled states can result in a non-uniform distribution of molecules within the sample, leading to scattering and noise in the OKE trace. We also attempted to collect temperature dependent data of the plasmid (data not shown), but due to the low concentration used the spectrum is very similar to that of pure water. However, when held at 95°C the signal was seen to became significantly clearer due to the plasmid DNA denaturation and hence, removal of the supercoils, which proves that these structures were the source of the scattered radiation.

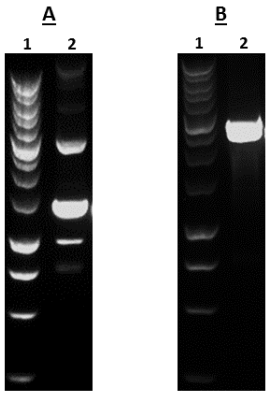


Figure 5.9: 1.5% agarose gel of pUC19 plasmid. *Lanes 1 and 2 show a 1kb ladder and the plasmid, respectively. (A) Displays the plasmid in numerous supercoiled states and (B) shows only one defined band after digestion, indicating pure plasmid.*

5.2 Label-free Nucleic Acid Hybridisation

It has been shown that terahertz radiation can detect various vibrational modes within the backbone of DNA and there have been a number of applications suggested for this technology. However, the possibility for label free detection of hybridisation has gained the most interest. Currently, hybridisation techniques are based on labelling of either the target or probe DNA, followed by fluorescence, radioactive, electrochemical or colorimetric detection. Although this method has proved to be extremely effective, the labelling of the DNA requires an additional preparation step, which takes skill, time and can complicate analysis. Furthermore, the addition of fluorescent tags can eventually introduce modifications in the DNA conformation, which can lower detection efficiency (Ozaki and McLaughlin, 1992), give rise to labelling yield fluctuations, fluorophore quenching and label degradation. These changes can affect the quantifiability of probe hybridisation (Zhu and Waggoner, 1997; Larramendy *et al.,* 1998; Zhu *et al.,* 1994), which is essential to process data accurately. For these reasons, a great deal of interest has been shown in terahertz

spectroscopy as a simple and reliable method to detect DNA hybridisation that does not suffer the drawbacks of current techniques. Most work in this area has been carried out by Bolivar and co-workers using THz-TDS, who have shown that there is a clear difference in the spectra produced between hybridised and denatured DNA. Hybridised samples were seen to have a significantly higher refractive index (Brucherseifer *et al.,* 2000), which is suggested to stem from the predicted phonon modes in double stranded DNA being absent in the single stranded sample. The group's final goal was to commercialise this research and create a ' THz Biochip' that would allow label-free analysis of DNA hybridisation (Bolivar *et al.,* 2004). However, the current method required significantly larger quantities of DNA (34 μg) than conventional detection techniques in order to make the sample sufficiently thick (30 μm) - for significant interaction length between the radiation and sample - and wide (2 mm) - for it to be broader than the terahertz wavelength to avoid scattering effects. In subsequent publications they have overcome these problems by using a novel waveguide approach to increase interaction length of the sample with terahertz waves. By combining this with resonators to amplify the DNA and terahertz wave interaction - subsequently amplifying the signal - they managed to reduce the amount of DNA needed to a level comparable to conventional experiments (~1 femtomole) and even displayed the potential for mutation detection (Nagel *et al.,* 2002a, 2002b, 2003). The commercialisation of this technique is now a competitive industry and 'Vibratees', has developed a different system for the analysis of DNA, and other biomolecules, utilising a novel CW, frequency-domain instrument (Globus *et al.,* 2013). Furthermore, without the use of waveguides the free space detection of single and double stranded DNA can be improved by using a metallic mesh in order to amplify the refractive index of the samples (Hasebe *et al.,* 2012).

In contrast to data presented by Bolivar and Nagel, Globus' FTIR measurements have shown that DNA in its denatured state generally displays a *greater* refractive index than the hybridised state (Globus *et al.,* 2004). In this work the authors also suggest that hybridised and denatured DNA can be distinguished by three absorption features,

and in later publications claim to identity an additional two vibrational modes (five total) that can be used to distinguish single stranded from double stranded DNA (Parthasarathy *et al.,* 2005). The additional vibrational modes and higher absorption detected for single stranded DNA was attributed to the increased flexibility of the molecule, allowing for more degrees of freedom. However, in this earlier work, only air dried samples displayed these specific features and were absent in the liquid phase. Yet subsequent work identified 4 four spectral features in gel samples that can be used to discriminate between the two states (Globus *et al.,* 2006).

Although a great deal of research has been carried on the terahertz dynamics of DNA hybridisation and a large number of vibrational modes predicted, there is little evidence to confirm the presence or frequency of these modes. In our initial DNA oligonucleotide measurements we observe features that we believe should only be present in the hybridised state, such as the long hypothesised 'phonon-like mode' that propagates along the DNA double helix, and therefore should be absent from the single stranded molecule, and also hydrogen bonding between the bases. Therefore, we aim to test our hypothesis by carrying out measurements on single and double stranded DNA oligonucleotides to ascertain how the hybridisation state influences the OKE spectrum.

5.2.1 Temperature Induced Denaturation

We first determined the melting temperature (T_m) of the oligos by means of differential scanning calorimetry (DSC) and SYBR green binding. DSC is used to measure the specific heat capacity of thermally induced transitions as a function of temperature and is commonly applied to DNA molecules as a measure of thermodynamic properties and hence, binding state (Privalov *et al.,* 2007). SYBR green is a cyanine dye that is used to detect the hybridisation state of DNA as the dye has a high affinity for the minor groove of double helix DNA but low affinity for single stranded DNA (Zipper *et al.,* 2004). The dye alone does not display fluorescence but when complexed with DNA absorbs blue light at 497 nm and emits

green light at 520 nm and therefore, green fluorescence is an indicator of hybridised DNA.

Our DSC and SYBR green measurements performed on the A-T oligonucleotide identified the melting temperature to be between 50 – 60°C (Fig. 5.10). For both measurements we detect only one well defined melting transition, which confirms that the DNA sequence was accurately designed as no additional melts were detected due to mis-pairing or hairpin loops. However, there are some discrepancy between the measurements as DSC measured the T_m to be 57°C whereas the SYBR green data shows a T_m of 51°C. This may be explained by the disruption of the base stacking interactions occurring first, causing unwinding of the helix and dissociation of SYBR green at slightly lower temperatures. In contrast, DSC measures the endothermic heat flow required to break the hydrogen bonds and is therefore somewhat more accurate. Nevertheless, both methods are in reasonable agreement.

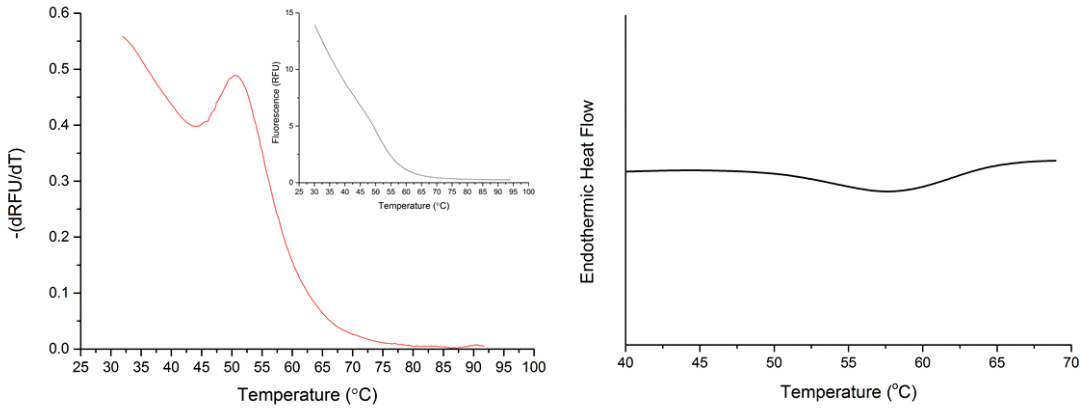


Figure 5.10: T_M of A-T oligonucleotide. *SYBR green measurements (left) show the derivative (main image) of the fluorescent intensity (inset) indicating a T_M at 51°C. DSC measurements (right) display an endothermic peak representing a T_M at 57°C.*

The same measurements were also performed on the G-C oligonucleotide in order to elucidate the melting temperature, yet we found this to be difficult utilising the current methods. Our DSC measurements did not show a defined peak but instead, a marginally negative sloping baseline (Fig. 5.11). Similarly, the derivative of the fluorescent intensity of SYBR green did not show a defined melting point but a gradual decrease in fluorescence. This would suggest that the G-C oligonucleotide does not dissociate in a near instantaneous manner like other short polynucleotides, but is a more gradual process. This observation is not surprising as G-C base pairing forms three hydrogen bonds, whereas A-T pairing forms only two, resulting in a large increase in stability over the whole oligo.

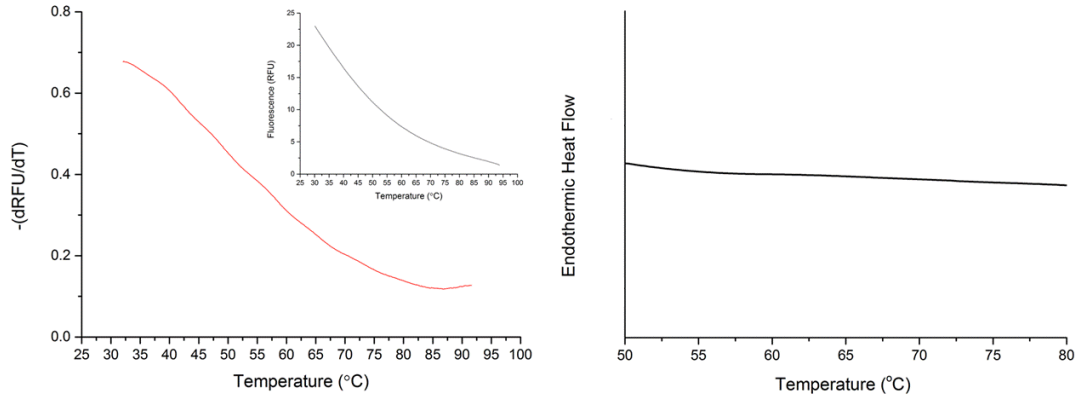


Figure 5.11: T_M of G-C oligonucleotide. *SYBR green measurements (left) show the derivative (main image) of the fluorescent intensity (inset) with no distinct peak. DSC measurements (right) also display no endothermic peak representing no defined T_M.*

The OKE spectra of both oligonucleotides display an increase in amplitude at frequencies below 2 THz as the samples are heated (Fig. 5.12), which can simply be explained by the increase in the molecular orientational diffusion band at very low frequencies as the viscosity is reduced due to heating. This increase in amplitude is relatively small and consistent between temperature points except between 50-60°C and 70-75°C for the A-T and G-C oligos, respectively, where we detect a sharp increase in the signal (Fig. 5.12). The observed jump in amplitude could be associated

with hyperchromicity, which is readily used to determine the hybridisation state of DNA with ultraviolet (UV) light - DNA bases absorb light in the 260 nm range and the hyperchromic effect occurs when DNA unwinds and the bases become unstacked, resulting in the absorption of more UV light. As the bases are no longer 'shielded' by the double helix, a marked increase in absorbance of 37% is detected and is commonly used to distinguish between hybridised and denatured states (Ho *et al.,* 1998). A similar trend could also be possible for our OKE data, however if the increase in signal were to originate from exposure of the base pairs, we would also expect a shift in the librational band as the bases gain a higher degree of freedom due to the removal of hydrogen bonds, yet this is not seen. Therefore, this effect is likely to again stem from the difference in the molecular orientational of the single and double stranded molecules. When the oligonucleotide adopts the structured double helix, the orientational diffusion, or 'tumbling', will be slow, yet when the DNA is denatured, the helix separates into two strands, which adopt a more compact random coiled conformation, increasing the orientational diffusion. Therefore, as this band is pushed to higher frequencies we detect the change as an increase in amplitude in the spectral range of our data.

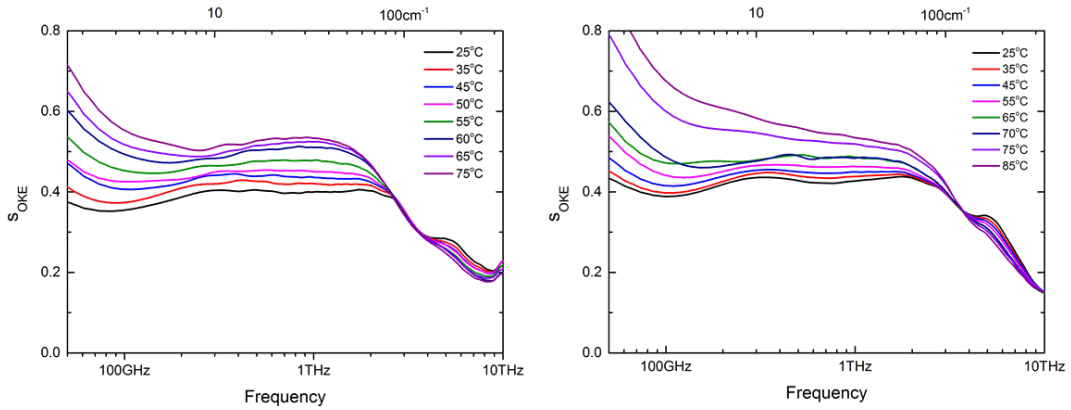


Figure 5.12: Solvated OKE spectra of oligonucleotide melting. *Temperature dependent changes in A-T oligonucleotide (left) and G-C oligonucleotide (right)*

Interestingly, the ~300 GHz band, which is more visible in the solvated spectrum (Fig. 5.12), becomes weaker with increasing temperature and also disappears completely in the same temperature range as we detect the jump in amplitude. Our NTP data has already established that the 300 GHz peak cannot originate from base stacking interactions, so we have assigned this mode to hydrogen bond effects, which is supported by the observations we have made in the current thermal measurements. The 300 GHz band is detected as a relatively strong peak at 25°C but as the temperature is increased we see a significant decrease in intensity until it disappears completely. Therefore, it would be logical to assume the reduction of this band is due to temperature effects on hydrogen bonding. This peak is more pronounced in the solvated spectra as the process of manually subtracting H_2O to produce the solvent subtracted spectrum (Fig. 5.13) introduces additional noise and masks this feature to some extent. To reliably subtract the H_2O spectrum, we found that the corresponding oligo and H_2O measurements must be performed consecutively in order to minimise any discrepancies in the spectrum induced by deviations in the laser. Unfortunately, it was not possible to collect the corresponding H_2O temperature measurements for the G-C data and the solvent subtracted spectrum is therefore not shown. However, the solvent subtracted spectrum of the A-T oligonucleotide (Fig. 5.13) helps to verify the presence of the phonon mode we previously identified as the feature at 2.9 THz. Again, we see a reduction in this band with increasing temperature until it is absent above 55°C. Theoretical work has predicted that DNA phonon modes in this frequency range originate from motions that are spread over one or more base pairs, are influenced by the weak hydrogen bonding and therefore are only present in the DNA double helix (Van Zandt & Saxena, 1992). This agrees well with our OKE data as the phonon mode is only visible at low temperatures when the oligo adopts the double helix conformation.

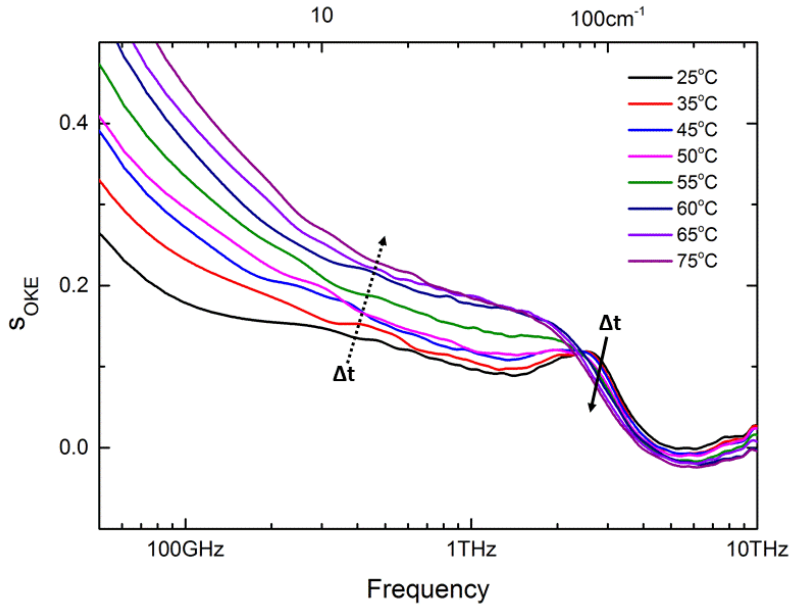


Figure 5.13: OKE spectrum of solvent subtracted A-T oligonucleotides. *Increased temperature is accompanied by a jump in amplitude (dashed arrow) and reduction of the phonon mode (solid arrow) at the T_M = 55°C*

Our OKE measurements on the A-T oligonucleotides have shown three separate changes in the spectra that occur at the T_m defined by our DSC and SYBR green measurements. We can therefore be confident that these changes in the terahertz spectrum are indicative of the DNA molecule transitioning from a hybridised to denatured state. Although we could not accurately define the T_M of the G-C oligonucleotide using DSC and SYBR green methods due to the increased stability of G-C base pairing, we believe that the changes observed in the G-C OKE signal between 70 - 75°C illustrate the denaturation of the GC oligo. To obtain reliable OKE data with high signal-to-noise ratio requires long scanning times in order to generate a sufficient number of traces. Therefore, the sample was held at the defined temperature points for a prolonged time and allows the gradual denaturation of the G-C oligo. However, calorimetric studies and SYBR green florescent measurements typically have fast scanning times, which may not allow the denaturation to proceed to completion, which we believe to be the case in this instance. Multiple DSC and

SYBR green measurements were performed at various scanning rates but the instruments were not capable of mimicking the time scales used for our OKE measurements. It is also noted that the fluorescent derivative does not reach zero even close to 95°C, which would suggest that the G-C oligo may not completely denature on the timescale of these experiments. However, by comparison with A-T oligo data, we are confident that our OKE spectrum shows the denaturation of the G-C oligo between 70 - 75°C.

5.2.2 Complimentary Oligonucleotides

It was possible to confirm the observations made with the 20mer, G-C and A-T, oligonucleotides while also eliminating any possible discrepancies due to heating effects by utilising 13mer and 17mer oligonucleotides that should only form a double helix complex (30mer) when dissolved together (Table.5.2).

Table 5.2: Sequence of Oligonucleotides used for OKE analysis

Length (Bases)	Sequence (5' -> 3')
13	CGAAAAATGTGAT
17	CTAGATCACATTTTTCG

The presence of the phonon mode is clearly displayed in the hybridised spectrum at 2.9 THz and the additional contribution of this mode can be detected in the solvent subtracted spectrum as a change in the overall spectral shape between the single stranded (SS) oligos and the hybridised complex (Fig. 5.14). Additionally, there is a large difference in amplitude between the SS oligos and complex in the 1 THz range, which agrees well our 20mer data and confirms this change to be indicative of hybridisation state. The 300 GHz band that we previously assigned to hydrogen bonding is clearly present in the complexed spectrum and completely absent from the 17mer spectrum, yet we also see a small contribution from this band in the 13mer

spectrum, which would suggest non-specific hydrogen bonding between 13mer strands. The sequences used for the 13mer and 17mer oligos was chosen to also be recognised by the CAP protein so that the DNA could be used in multiple experiments. Therefore, the sequence was not specifically designed to eliminate any self-annealing and it's possible that non-specific hydrogen bonding may occur. However, of the two oligos, the 17mer is more likely to partake in self-annealing as there are 4 complimentary base pairs on the 5' end to promote formation of the '30mer', yet there is no detection of the 300 GHz feature in the 17mer spectrum. Nevertheless, the 13mer spectrum, overall, displays features we now associate with single stranded DNA and we assume any non-specific hydrogen-bonding to be minimal.

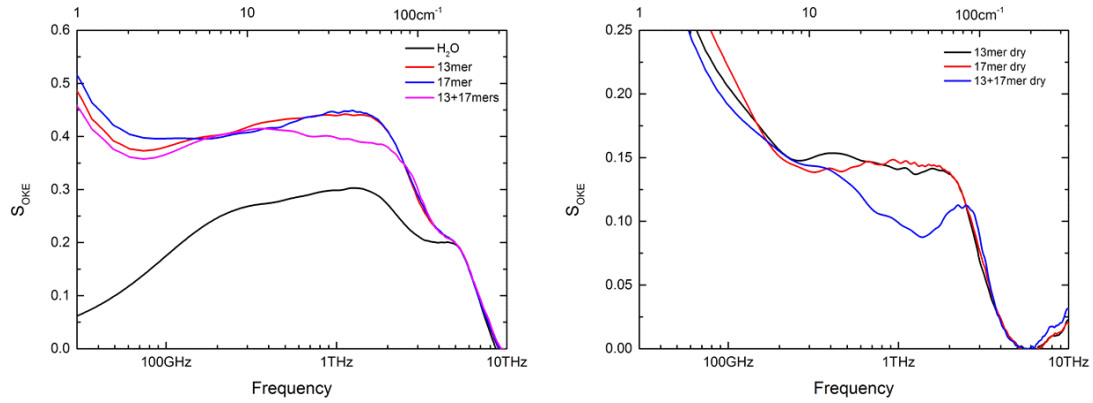


Figure 5.14: OKE spectra of solvated (left) and solvent subtracted (right) nucleobases. *13mer, 17mer and an equimolar concentration of both at 10 mM*

5.3 Conclusions

We have identified multiple vibrational modes in the terahertz spectrum associated with interactions between individual nucleotide components and also long chain oligonucleotides. We believe a novel mode detected at 0.35 THz indicates the presence of hydrogen bonding between nucleotides and a phonon mode is also detected at 2.9 THz in double stranded DNA. Both of these modes are indicative of DNA conformation and are accompanied by a jump in the amplitude of the OKE spectrum when transitioning from the hybridised to denatured state. The label free detection of DNA hybridisation in the terahertz range has gained much interest in recent years and we have identified novel modes that can be used to monitor the state of unmodified DNA molecules in solution. Further research focusing onto these modes may allow terahertz radiation to become a novel method for label free DNA analysis.

Chapter 6: Terahertz & Far-Infrared Database of Small Biomolecules

Chapter 6 Terahertz & Far-Infrared Database of Small Biomolecules

In this chapter we utilise two of the most common methods of terahertz spectroscopy, THz-TDS and FTIR, to analyse a variety of small biomolecules. The aim of this work was to generate a database of far-infrared and terahertz spectra of small biomolecules as data in this frequency range is lacking and a complete work of biologically relevant small molecules – amino acids, peptides and DNA components – would be beneficial for future research. Additionally, THz-TDS has shown potential for analysing a variety of biological molecules (Plusquellic *et al.,* 2008; Falconer *et al.,* 2012) but sample preparation has been a common problem. Therefore, we developed a simple and reliable membrane method to analyse these molecules, which could potentially be incorporated into a number of applications.

6.1 Terahertz-TDS Measurements

6.1.1 Sample Preparation of Biological Molecules

In order to analyse biomolecules in the terahertz range it is often necessary to use powder samples due to the high attenuation of terahertz waves by liquid water. Pressed pellets of polyethylene mixtures are commonly used and have been successful in determining the terahertz spectra of nucleotides, amino acids, proteins and DNA (Fischer *et al.,* 2002; Korter *et al.,* 2006; Markelz *et al.,* 2000). However, preparation of these samples is time consuming, especially if multiple measurements are to be performed under varying conditions (concentration, pH *etc.*). An alternative method that has also been successful in the analysis of DNA is the preparation of 'dry films', which is achieved by depositing a liquid sample onto a silicon or sapphire substrate

before drying in a desiccated atmosphere (Brucherseifer *et al.,* 2000). However, a common problem with this technique is the so-called 'coffee drop effect' where there is an uneven distribution of the molecules in the sample – more concentrated at edges compared to the centre.

A more recent technique that has shown promise for biological samples is the utilisation of charged membranes (Yoneyama *et al.,* 2008; Kikuchi *et al.,* 2009). This technique allows solutions of biomolecules to be deposited and dried with minimal preparation and multiple samples can be analysed rapidly. Furthermore, many biochemical applications utilise these membranes to fix and analyse biomolecules as, even when they are dried, some water is thought to remain bound which maintains the activity and structure of the molecule.

Some of the more common biochemical procedures that utilise these membranes are Western and Southern Blots for protein and DNA analysis, respectively. Both of these techniques require the binding of a tagged probe to the molecule of interest followed by fluorescent or colorimetric detection. Although these methods have proved to be extremely effective, the labelling of the probe molecules requires an additional preparation step, which takes skill, time and can complicate analysis. Previous studies have already shown that terahertz spectroscopy is sensitive to protein (Mickan *et al.,* 2002) and DNA binding (Bolivar *et al.,* 2002) and also capable of reliably measuring biomolecules quantitatively (Ueno*, et al.* 2006). Therefore, this gives us reason to believe that terahertz spectroscopy could potentially be applied to the detection of membrane bound biomolecules and possibly be incorporated into blotting protocols, removing the need for probe labelling or probes in general.

6.1.1.1 Polyethylene Pellets

Polyethylene pellets are a common method to analyse biomolecules in the terahertz range. Therefore, we used this method to gain low frequency data that could be incorporated into our terahertz database of small biomolecules and also as a

comparison to our novel membrane measurements. However, the manner in which polyethylene pellets are prepared can have an impact of the quality of the spectra. For obvious reasons, it is important to minimise moisture by preparing the pellets under vacuum and it is also advisable to use reduced pressure to limit transient heating of the sample. In addition, two of the most important factors that influence the resultant spectra are the ratio of polyethylene to sample and also pellet thickness. To gain accurate spectra we tested two ratios of sample to polyethylene (1:1 and 1:10) and two thicknesses (1 mm and 3 mm) with both the single amino acid histidine (Fig. 6.1) and the peptide glutathione (reduced) (Fig. 6.2).

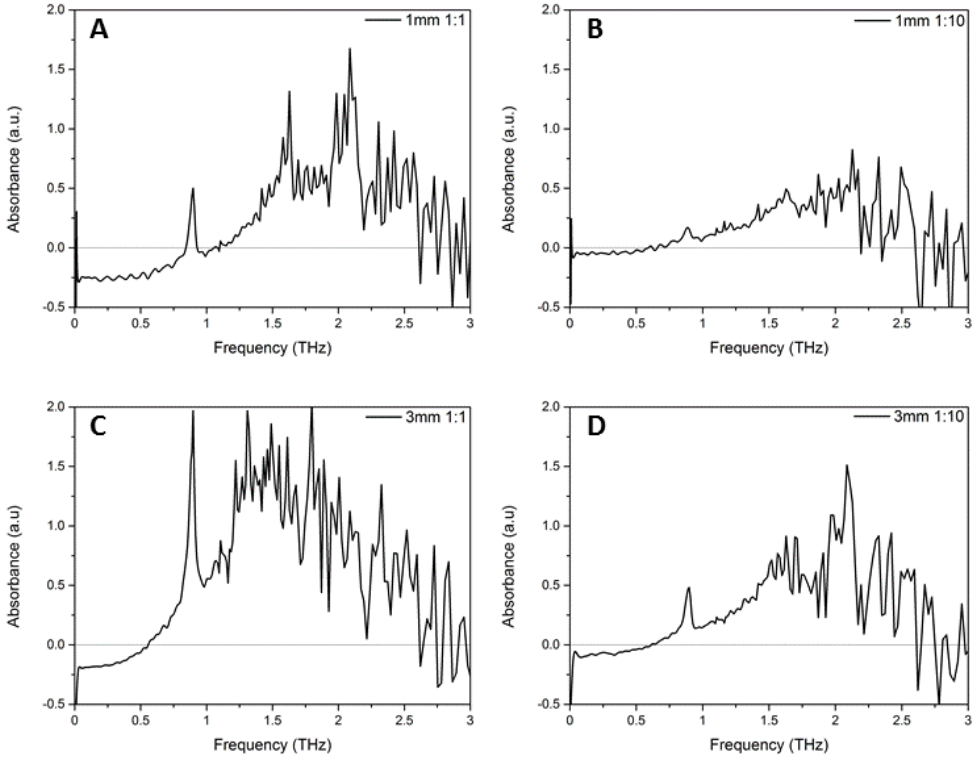


Figure 6.1: Raw THz-TDS spectra of Histidine : Polyethylene pellets. *(A) 1 mm sample in a 1:1 ratio (B) 1 mm sample in a 1:10 ratio (C) 3 mm sample in a 1:1 ratio (D) 3 mm sample in a 1:10 ratio. Histidine displays clear absorbance peaks at 0.9 THz, 1.6 THz and 2 THz.*

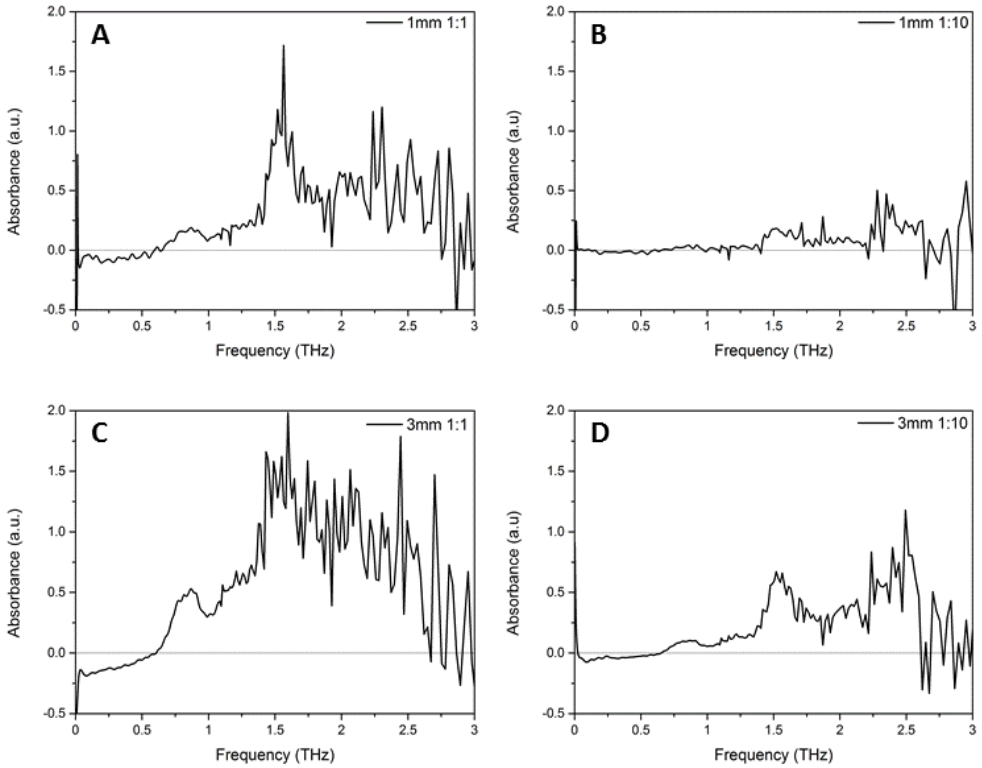


Figure 6.2: Raw THz-TDS spectra of GSH : Polyethylene pellets*. (A) 1 mm sample in a 1:1 ratio (B) 1 mm sample in a 1:10 ratio (C) 3 mm sample in a 1:1 ratio (D) 3 mm sample in a 1:10 ratio. GSH displays clear absorbance peaks at 0.8 THz and 1.5 THz.*

One of the spectral features that appears in our 1 mm samples is the present of oscillations in the low frequency part of the spectrum that is due to etalon artefacts. This feature is caused by reflections at the air/pellet interface and is dependent on the thickness and refractive index of the sample. As we can see in our 3 mm samples, these artefacts have been removed as the reflection now appears long after the main terahertz pulse and no longer occurs on a timescale that interferes with the data. The other feature that is clear from our measurements is the steep sloping baseline and additional noise seen at higher frequencies in the thicker pellets and pellets with an equal ratio of sample to polyethylene. This interference can mask spectral features and is caused by scattering effects due to inhomogeneities in the pellets and although

the samples were ground to ensure particle sizes smaller than the terahertz wavelengths, and hence, minimise this baseline offset arising from non-resonant light scattering, air can still be trapped in the matrix, which can attenuate the terahertz radiation by creating scattering interfaces. Therefore, although the Beer-Lambert relationship would suggest that thicker samples provide higher sensitivity, in practice, the signal-to-noise ratio is affected by attenuation caused by scattering and there is a practical limit to pellet size. From our measurements, we conclude that the 3 mm pellet in a 1:10 ratio is free from etalon artefacts but still produces a good signal-to-noise ratio and is therefore the most reliable spectrum.

6.1.1.1 A Membrane Method for Terahertz Spectroscopy

To perform membrane measurements on biomolecules we designed and built a sample holder that would allow simple application and drying of solutions onto the membrane, combined with compactness for easy measurements. The top and bottom plates are composed of austenite (316) stainless steel, which is extremely robust and inert so we can be confident of minimal contamination in our measurements. Wells of 10 mm and 20 mm diameter have been incorporated into the holder (Fig. 6.3) in order to accommodate different sample volumes and also to be used in various terahertz systems as the beam waste (ω_0) may vary according to:

$$NA \approx n\frac{D}{2f}$$

And

$$NA \simeq \frac{\lambda_0}{\pi\omega_0}$$

Where *NA* is the numerical aperture, *n* is the refractive index of the medium that the lens is working (*i.e,* air = 1), *D* is the diameter of lens, *f* is the focal length and λ_0 is the vacuum wavelength of light. Therefore, if the well is too narrow, the low frequency part of the spectrum may be compromised.

The membrane was sandwiched between two silicone sheets (containing wells that matched the holder) and screwed together to make a watertight seal so there was no cross contamination between wells (Fig. 6.3). The solution could then easily be deposited onto the membrane and dried before performing the measurement.

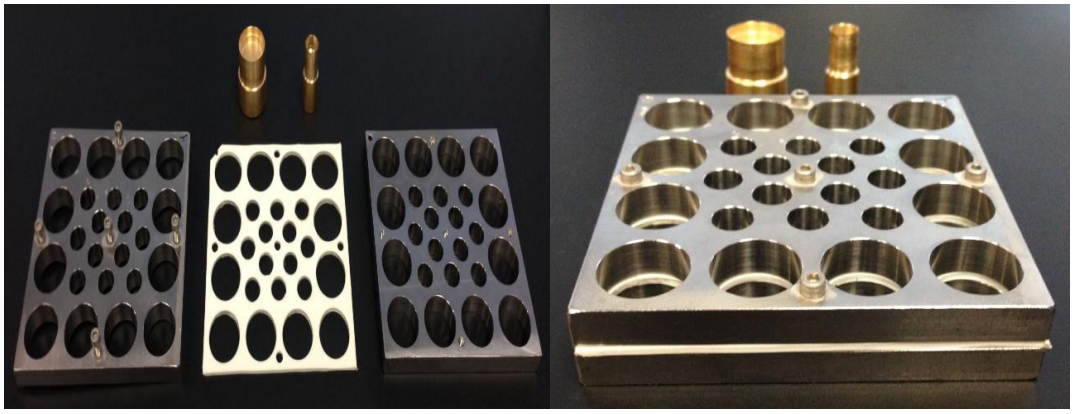


Figure 6.3: Terahertz spectroscopy holder for membrane measurements*. The holder dismantled (left) and assembled (right)*

We tested two commonly used protein binding membranes, polyvinylidene fluoride (PVDF) and nitrocellulose (NC), for attenuation and although both were equally transparent across the terahertz range, PVDF is very hydrophobic and requires soaking in methanol to 'activate' the membrane before binding biomolecules, whereas NC is hydrophilic and does not require any preparation. Therefore, we decided to proceed with NC for our measurements on biomolecules lest the addition of methanol complicated analysis.

We carried out membrane measurements on the antioxidant tri-peptide, reduced glutathione (GSH), as it has two easily identifiable peaks at 0.84 THz and 1.53 THz and is also highly soluble in water. Fig. 6.4 shows a comparison between the 3 mm, 1:10 ratio pellet and the membrane measurement. We see excellent agreement between both measurements and both GSH peaks are clearly present in the membrane measurement. Furthermore, the membrane appears to suffer less from etalon effects than thin pellets. Therefore, we have shown that the membrane method is an effective

tool to analyse biomolecules in the terahertz range. However, due to issues with our THz-TDS system, this research could not be continued and we were not able to carry out quantitative measurements.

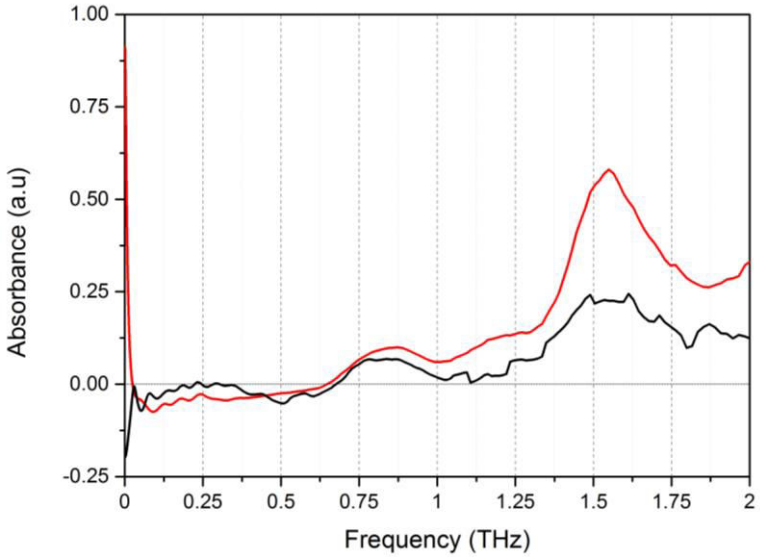


Figure 6.4: THz-TDS spectra of the GSH pellet (Red trace) compared to the membrane measurement (Black trace). *Spectra have been smoothed by 5 points.*

6.2 Amino Acid Zwitterions in the Far-Infrared

Amino acids are the building blocks of protein molecules and also form the basis of other essential biomolecules such as neurotransmitters and porphyrins in red blood cells. Due to their importance in biology, amino acids have been a common subject of spectroscopic study and their spectra are well characterised in the mid-infrared (Barth, 2000). However, since the first far-infrared investigation of amino acids in 1935 (Heintz, 1935) the spectra of these molecules has only been reported sporadically, with most studies focusing on the smallest amino acids, glycine (Gupta *et al.,* 1970; Machida, 1977) and alanine (Fukushima *et al.,* 1959; Suzuki *et al.,* 1959), and to our knowledge, a far-infrared study on all 20 α-amino acids is not available. Furthermore,

the protonation state of these molecules alter the structure and hydrogen bonding within these molecules and therefore, influence the far-infrared spectrum. However, limited far-infrared data on amino acid zwitterions is available (Trivella *et al.,* 2010).

Here, we present the first comprehensive study of all 20 α-amino acids, in their zwitterionic states, in the far infrared. We analysed amino acid powders without any further preparation and also dry films prepared for each amino acid at its isoelectric point, cationic and anionic states. Our initial intentions were to carry out the experimental work in conjunction with normal mode and Density Functional Theory (DFT) calculations to assign the vibrational modes. However, this was not possible in the current study and peaks are therefore assigned based on the modes available in literature and comparisons between different amino acids.

6.2.1 Glycine

Glycine is the smallest amino acid and is known to adopt three polymorphic states in its zwitterionic form - the α, β and γ forms (Fig. 6.5). By comparisons with previous spectroscopic data we believe the powder sample to mainly consist of the α polymorph(Chernobai et al., 2007; Trivella et al., 2010), whereas the dried glycine solutions should compose of mainly the β form (Trivella *et al.,* 2010). Both polymorphs adopt a monoclinic crystal structure but the α form adopts a $P2_1/n$ space group, whereas the β form adopts a $P2_1$. These differences are easily detectable in our experimental far infrared spectra.

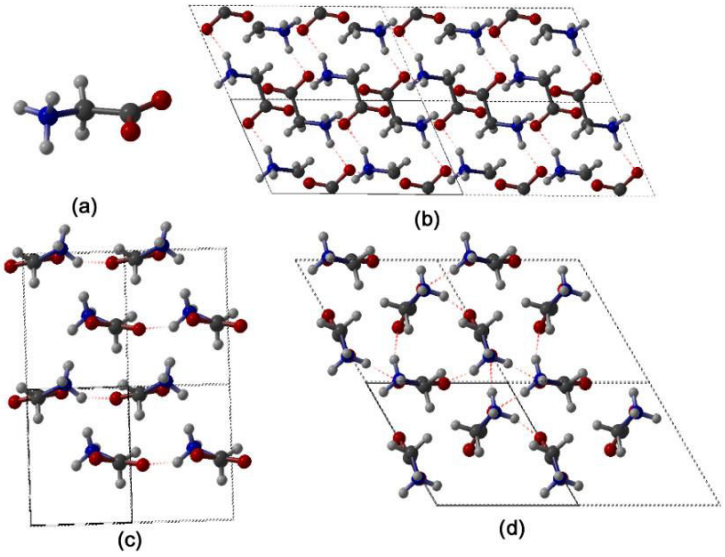


Figure 6.5: Glycine Polymorphs. *a) Shows glycine as a single molecule; b) represents the α-glycine crystal structure; c) β-glycine crystal structure; d) γ-glycine crystal structure.*

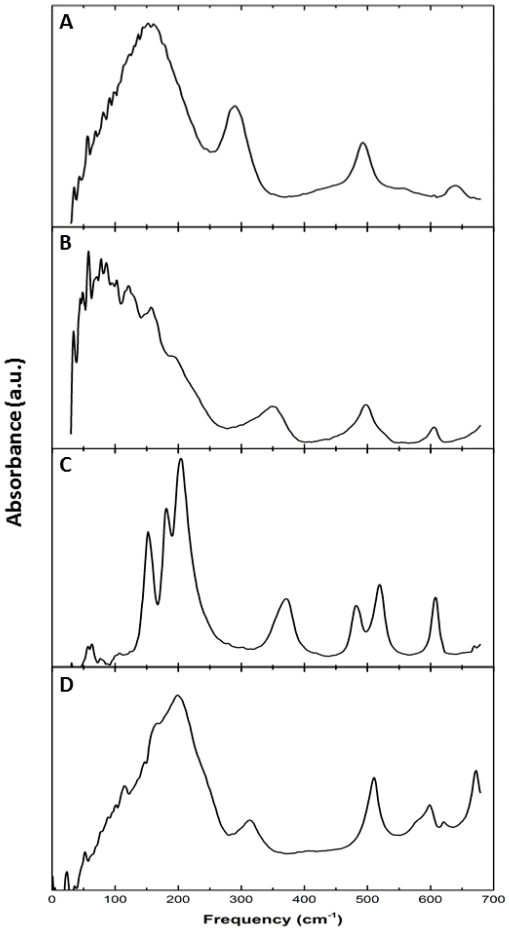
The far infrared spectra of glycine are shown in Figure 6.6. In the high frequency region of the spectra we detect a peak at 606 cm^{-1} for the solid form and 608 cm^{-1} for the pI sample. There is strong evidence that these modes originate from the COO^- wagging mode (Chowdhry et al., 2008; Goryainov et al., 1970), although it has also been suggested that this mode may be attributed to CC=O bending (Matei *et al*., 2005). The cationic and anionic forms also display modes in the >600 cm^{-1} region at 638 cm^{-1} and 672 cm^{-1}, respectively, which have both been assigned to the bending of the COO(H) group (Trivella *et al.,* 2010) with the protonation state influencing the frequency of this mode. For the anionic sample we also detected a peak at 599 cm^{-1}, which has previously not been assigned and a shoulder at 578 cm^{-1} that is assigned to COO^- wagging. However, we assign the stronger 599 cm^{-1} peak to this mode.

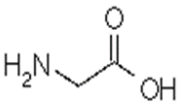
All glycine samples display a strong contribution in the ~500 cm^{-1} region and the difference between the zwitterionic forms is noteworthy. The powder sample shows a peak at 497 cm^{-1} with a shoulder on the high frequency side (520 cm^{-1}), whereas this

feature has split into two well resolved peaks at 519 cm^{-1} and 482 cm^{-1} for the pI sample. Trivella *et al.*, (2010) assigned these peaks to the NH_3^+ torsional mode and the COO^- rocking mode, respectively, based on the large blueshift of the NH_3^+ mode observed at low temperatures; this vibration is known to be temperature sensitive whereas the carboxylate vibration is not (Bandekar *et al.,* 1983; Feairheller *et al.,* 1971). For the cationic form we detected a peak at 493 cm^{-1} and at 511 cm^{-1} for the anionic. The cationic peak has simply been assigned to the COOH rocking mode, whereas for the anionic spectra, this region may be influenced by interactions involving the alkaline sodium cation (Trivella *et al.,* 2010). The band at 313 cm^{-1} detected in the anionic spectrum is assigned to a COO- rocking/ NCαC bending mode whereas the 199 cm^{-1} mode is again influenced by the cation, which becomes evident when using different alkali salts. For example, if KOH is used instead of NaOH a frequency shift of this mode is detected due to the increased mass of K^+ compared to Na^+ (Trivella *et al.,* 2010). The cationic sample displays a vibration at 290 cm^{-1}, which is assigned as the red-shifted 351 cm^{-1} and 371 cm^{-1} modes seen in the powder and pI samples, respectively, which is consigned to NCαC bending (Goryainov *et al.,* 2005; Herlinger *et al.,* 1970; Matei *et al.,* 2005). The remaining acidic and basic peaks detected in our far infrared measurements are regarded with a degree of uncertainty and no definitive assignment has yet been made.

The far infrared spectrum of the zwitterionic forms differs significantly in the region <200 cm^{-1} as the powder form exhibits a broad high intensity feature that contains multiple, poorly resolved modes, whereas for the pI sample, we see four sharp bands at 204 cm^{-1}, 181 cm^{-1}, 152 cm^{-1} and 61 cm^{-1}. Most of the modes in this low frequency region are assigned to lattice vibrations with the exception of the 158 cm^{-1} mode in our powder sample that is assigned to the COO- torsion, however, there is uncertainty whether this corresponds to the 181 cm^{-1} (Goryainov *et al.,* 2005) or 152 cm^{-1} (Trivella *et al.,* 2010) peak in the pI sample. There is also evidence that a hydrogen bond N-H···O elongation vibration may be detected in this frequency range (Feairheller *et al.,* 1971) and may contribute to the unassigned bands.

At the lowest frequencies, we measured the spectrum of glycine using a combination of FTIR and THz-TDS and identified peaks at 81 cm^{-1}, 68 cm^{-1}, 58/59 cm^{-1} and 51 cm^{-1}. Previous studies show some inconsistency in the identification of the glycine peaks in very low frequency range using different techniques. For example, using a novel waveguide approach, (Laman *et al.,* 2008) identified bands at 25 cm^{-1}, 42 cm^{-1}, 66 cm^{-1} and 95 cm^{-1} and upon cooling the band at 66 cm^{-1} (2 THz) split into two bands at 65 cm^{-1} and 70 cm^{-1}. Our THz-TDS measurements show some correlation with this study, but demonstrate better agreement with previous FTIR and THz-TDS measurements that identify vibrations at 68 cm^{-1} (Shot & Sievers, 1974) and 81 cm^{-1} (Shi & Wang, 2005) that match out measurements and are assigned to intermolecular vibrational modes mediated by hydrogen bonds.





Absorbance frequencies for Glycine (cm^{-1})

Cation	Powder	pI	Anion
639	-	-	672
-	606	608	599
-	520sh	519	578
493	497	482	511
290	351	371	313
-	226	-	-
-	194	204	199
155	158	181	166
-	121	152	-
-	-	-	115
-	87	-	-
-	77/81*	-	-
-	68*	-	-
-	58 /59*	61	-
-	51*	-	-

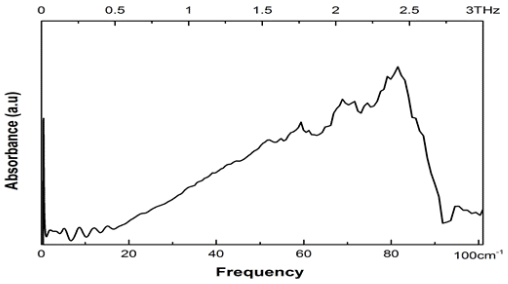


Figure 6.6: Far infrared spectra and absorbance frequencies of glycine zwitterions. *The figure shows (A) Cationic (B) Powder (C) pI and (D) Anion spectra. The spectrum below shows the low frequency THz-TDS spectrum of the powder sample. 'Sh' signifies a peak shoulder and an asterisk donates the THz-TDS absorbance frequencies.*

6.2.2 Alanine

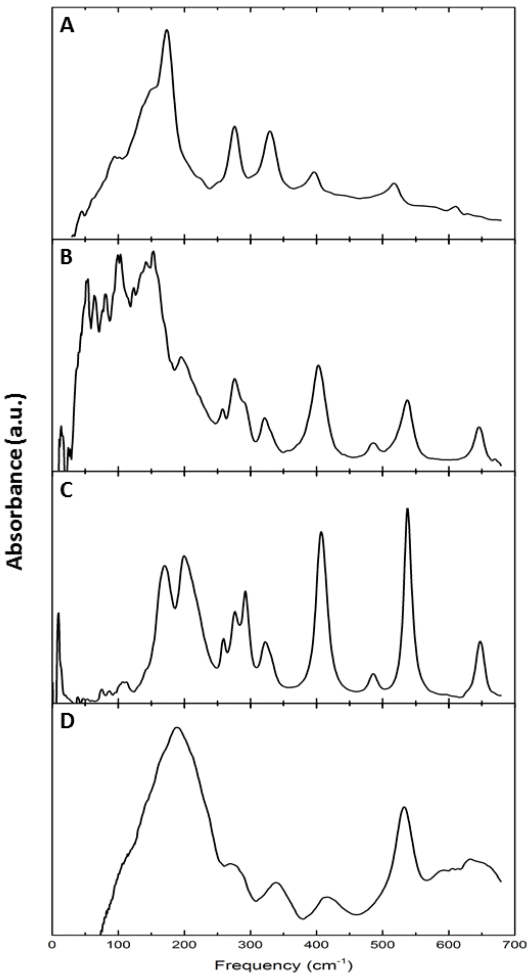
After glycine, alanine is the simplest amino acid as the α-carbon atom is bound by a simple methyl group (-CH_3) and it crystallises in an orthorhombic lattice. Although alanine and glycine are similar in structure, the increased spectral complexity of alanine becomes immediately apparent as the number of modes increases according to 3*N*-6 rule, where N is the number of atoms in the molecule.

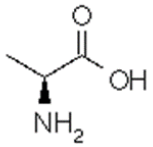
As alanine is known to adopt only one polymorphic state, our powder and film zwitterionic spectra show a high degree of similarity in the high frequency region of the spectra (Fig. 6.7). The COO- modes seen in the glycine spectra are shifted to higher frequencies for alanine and the band seen at 646 cm^{-1} in the powder is attributed to either the COO- wagging (Herlinger *et al.,* 1970; Wang *et al.,* 1971) or more recently, a COO- bending coupled with a CCαN bend (Trivella *et al.,* 2010). There is little debate that the origin of the next peak at 537 cm^{-1} is the COO- rocking mode and that the NH_3^+ torsion mode is represented by the weak band at 486 cm^{-1} (Matei *et al.,* 2005; Chowdhry *et al.,* 2008; Herlinger *et al.,* 1970; Wang *et al.,* 1971; Diem *et al.,* 1982). In this same frequency range we also detected independent peaks for the anionic (518 cm^{-1}) and cationic (532 cm^{-1}) samples, which have been tentatively assigned as the rocking mode of the carboxylic acid in its deprotonated or protonated state, respectively.

As we move into the mid-frequency range alanine is seen to possess considerable more peaks than was observed for glycine and these bands are thought to mainly represent CCαN deformations – equivalent to the 351 cm^{-1} peak observed for glycine. In the powder sample, the peak at 403 cm^{-1} can be assigned CCαN bending mode (Herlinger *et al.,* 1970; Wang *et al.,* 1971), the 321 cm^{-1} and 292 cm^{-1} bands are both CCαN bending/CO_2^- rocking coupled modes and the 276 cm^{-1} band assigned to CCαN bend (Trivella *et al.,* 2010). In contrast, the bands at 292 cm^{-1} and 276 cm^{-1} have also previously been assigned to CH_3 torsional modes (Diem *et al.,* 1982; Barthes *et al.*,

2002). Matei *et al,* describe the band at 258 cm^{-1} as the CO_2 torsion while Trivella *et al.* assign this peak to the CH_3 torsional mode.

The most pronounced differences between the powder and film zwitterionic samples become apparent below 250 cm^{-1} as this region of the spectrum is sensitive to the hydrogen bonding network, which will be effected by the changes in size and quality of the crystals in the different samples, as has previously been shown in other amino acid samples (Rungsawang *et al.*, 2006). The bands for both forms around 200 cm^{-1} can be assigned to lattice rotational modes (Trivella *et al.,* 2010) and the band at 164 cm^{-1} for the powder sample is assigned to the carboxylic acid torsion mode, corresponding to the 170 cm^{-1} peak in the pI film sample and the strong peaks at 174 cm^{-1} and 189 cm^{-1} for the cationic and anionic samples, respectively. The remaining bands in the very low frequency portion of the spectrum are likely to stem from lattice modes, however DFT calculations would suggest that a low frequency band caused by NCαC torsion should be visible at 62 cm^{-1}, that we were not able to detect in our FTIR measurements.





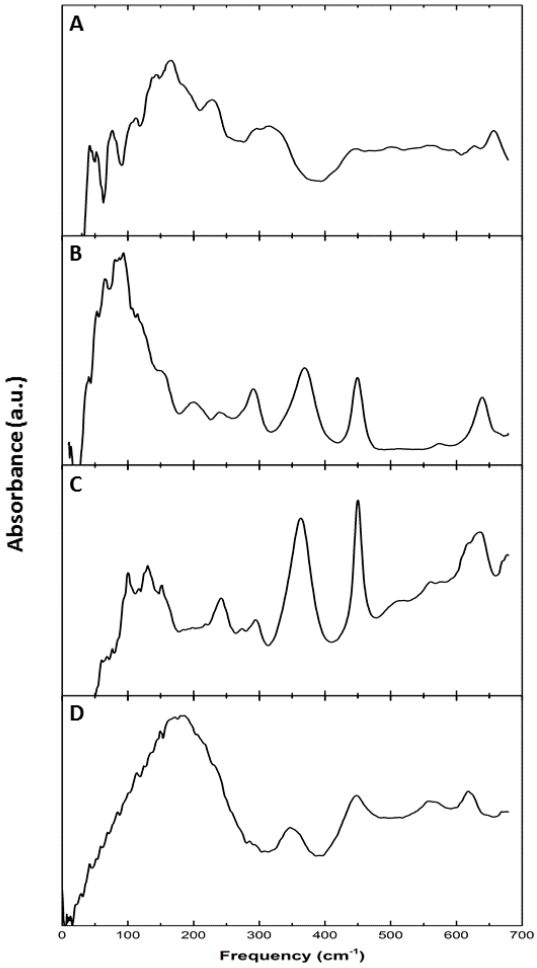
Absorbance frequencies for Alanine (cm^{-1})

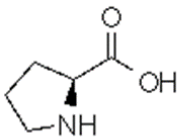
Cation	Powder	pI	Anion
610	646	648	633
518	537	538	532
-	486	487	-
396	403	407	417
-	334sh	330sh	-
329	321	322	338
-	292	292	-
276	276	277	270
-	258	259	-
-	195	200	-
174	164	170	189
-	152	-	-
146	-	-	-
97	101	109	114
-	82	-	-

Figure 6.7: Far infrared spectra and absorbance frequencies of alanine zwitterions. *The figure shows (A) Cationic (B) Powder (C) pI and (D) Anion spectra. 'Sh' signifies a peak shoulder.*

6.2.3 Proline

Proline is an atypical amino acid as its side group adopts a cyclic form, which is attached to a glycine molecule. In fact, proline is not technically an amino acid, but an 'imino acid' as it contains both imino and carboxyl functional groups. As proline is a small amino acid, one might expect the spectrum to be similar to that of glycine and alanine and by referring to these spectra in combination with the literature we attempt to assign some of the vibrational modes. Gupta *et al.,* (1973) proposed that the torsional mode of the proline ring should be detected between 527 cm^{-1} and 550 cm^{-1} based on calculations performed on proline oligomers. We detected a very weak band at 573 cm^{-1} in the proline spectrum that is absent from the glycine and alanine spectrum, which we therefore assign to the aforementioned vibration (Fig. 6.8). The COO- rocking/bending/wagging mode can be detected at 639 cm^{-1} and the bands at 201 cm^{-1} and 239 cm^{-1} are likely due to COO- torsional modes by comparing to alanine (Reva *et al.,* 1994).





Absorbance frequencies for Proline (cm^{-1})

Cation	Powder	pI	Anion
657	639	636	618
556	573	561	558
501	-	514	-
447	449	450	448
-	369	363	347
305	291	294	-
229	239	274	-
180	201	242	-
166	148	152	176
111	115	130	-
-	87	100	-
-	-	66	-

Figure 6.8: Far infrared spectra and absorbance frequencies of proline zwitterions. *The figure shows (A) Cationic (B) Powder (C) pI and (D) Anion spectra. 'Sh' signifies a peak shoulder.*

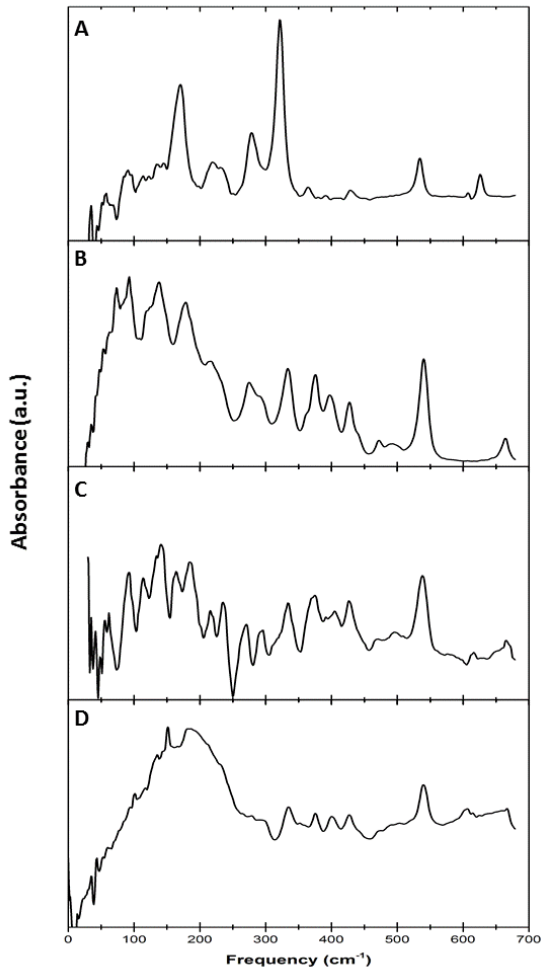
6.2.4 Valine, Leucine & Isoleucine

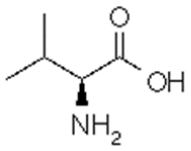
Valine, leucine and isoleucine differ from alanine by the addition of simple methyl groups at different positions on the side chain and all crystallise in a similar way – two molecules in the crystal asymmetric unit. This complicates mode assignments to the single molecules as this can lead to duplication of peaks in the far infrared spectrum. The sample preparation of these amino acids does not appear to affect the spectra and both powder and film samples show good agreement across the whole frequency range.

As we have seen, the high frequency region of the spectrum for simple amino acids is dominated by CO_2^- bending, rocking and wagging vibrations accompanied by the NH_3^+ torsional mode. Only one peak >600 cm^{-1} is detected for valine (Fig. 6.9), leucine (Fig. 6.10) and isoleucine (Fig. 6.11) and based on previous studies we would suggest that this mode can be recognised as the COO- bending (Trivella *et al.,* 2010) or wagging (Filho *et al.,* 2007; Herlinger *et al.,* 1970) mode. The single peak detected between 500-600 cm^{-1} for valine and leucine is assigned to COO- rocking, whereas a double peak is present for isoleucine at 556 cm^{-1} and 536 cm^{-1} and are assigned to the COO- rocking mode and the COO- coupled with bending or stretching around the α-carbon, respectively. The COO- rocking mode assignment remains consistent for the anionic and cationic samples as we see the strong, sharp peak in almost exactly the same position for valine and leucine, however the double peak for isoleucine becomes considerably less resolved for the basic sample and shows a larger shift to lower frequency in the acidic measurement than the other samples, yet our assignment remains the same. The position of the NH_3^+ mode for solid valine has been previously assigned to the peak at 472 cm^{-1} (Lima *et al.,* 2005), however, we detect a shared peak at 493 cm^{-1} for both valine and isoleucine that we assign to this mode in agreement with Trivella *et al* (2010). The author also assigned a very weak shoulder on the leucine spectrum, between 507-502 cm^{-1}, to the same mode, yet this shoulder was not detected in our spectrum.

In region of 250 cm^{-1}-500 cm^{-1} the zwitterionic forms of all these amino acids display a similar pattern dominated by 3 strong bands that have been shown to originate from skeletal deformations (Filho *et al.,* 2007; Herlinger *et al.,* 1970; Lima *et al.,* 2005; Pawlukojc *et al.,* 1995). However, the cationic spectra of these amino acids change considerably from the zwitterionic form, as the two higher frequency peaks minimise dramatically and a peak around 300 cm^{-1} intensifies.

In the region between 100 cm^{-1}-250 cm^{-1} the zwitterionic spectra mainly represent torsional modes and the torsion of the CO_2^- group contributes to the peaks detected at 293 cm^{-1}, 170 cm^{-1} and the shoulder at 180 cm^{-1} for valine, leucine and isoleucine, respectively. Again, the bands below 100 cm^{-1} are attributed to lattice modes (Husan *et al.,* 1984).

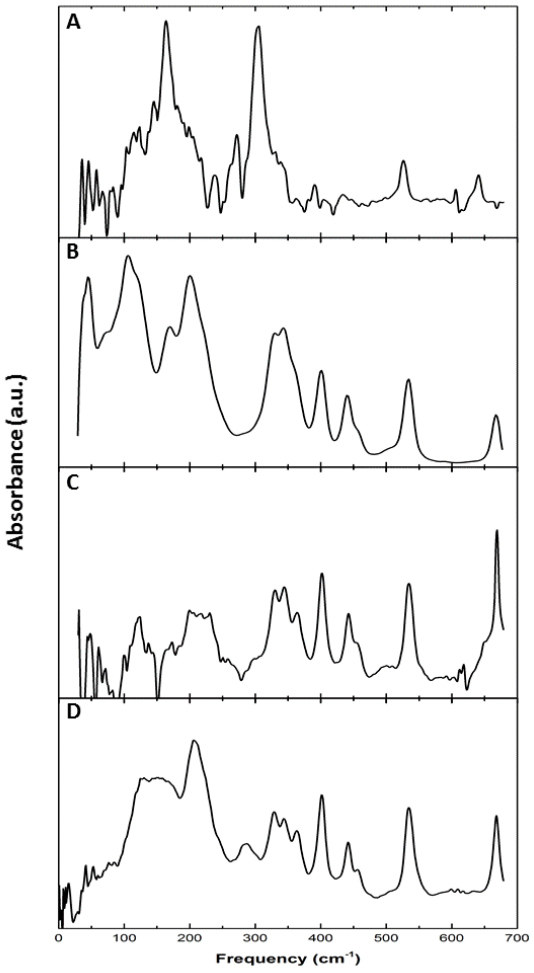


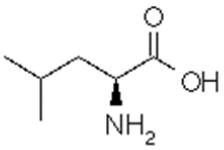


Absorbance frequencies for Valine (cm^{-1})

Cation	Powder	pI	Anion
626	665	-	668
535	540	538	540
-	493	495	489
-	472	472	473
-	441sh	-	-
427	427	428	428
392	397	400	401
-	375	372	376
365	362	-	-
322	334	333	335
-	293	296	298
280	275	-	-
224	217	230	233
170	178	182	183
138	138	140	153
-	120	-	-
92	93	92	93
-	73	-	-

Figure 6.9: Far infrared spectra and absorbance frequencies of valine zwitterions. *The figure shows (A) Cationic (B) Powder (C) pI and (D) Anion spectra. 'Sh' signifies a peak shoulder.*

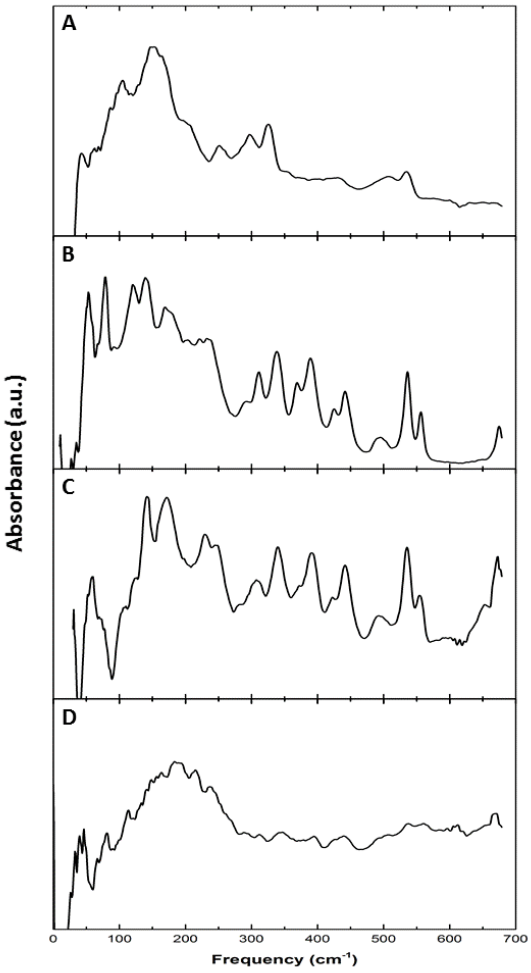


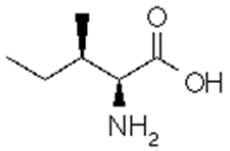


Absorbance frequencies for Leucine (cm^{-1})

Cation	Powder	pI	Anion
641	667	669	668
527	534	534	534
434	457	456	456
-	440	443	442
390	401	402	401
338	361	364	364
-	343	344	343
305	330	330	329
-	224	-	287
272	200	214	206
238	170	172	-
165	-	-	-
145	-	-	-
120	120	122	155
-	106	107	-

Figure 6.10: Far infrared spectra and absorbance frequencies of leucine zwitterions. *The figure shows (A) Cationic (B) Powder (C) pI and (D) Anion spectra. 'Sh' signifies a peak shoulder.*





Absorbance frequencies for Isoleucine (cm^{-1})

Cation	Powder	pI	Anion
-	675	670	667
-	-	654	-
536	556	555	566
-	536	535	536
508	493	493	504
435	442	441	441
-	425	423	423
-	390	390	396
-	369	370	-
327	338	340	344
298	311	307	315
253	293	292sh	292
210	232	247	237
-	221	229	217
-	203	-	-
160	170	171	-
-	180sh	-	185
-	139	141	-
107	121	120	-
-	79	-	-

Figure 6.11: Far infrared spectra and absorbance frequencies of isoleucine zwitterions. *The figure shows (A) Cationic (B) Powder (C) pI and (D) Anion spectra. 'Sh' signifies a peak shoulder.*

6.2.5 Serine, Cysteine & Threonine

Serine and threonine are similar to alanine but each possess a hydroxyl group on their side group and cysteine is identical in structure to serine but the hydroxyl group has been replaced by a thiol - therefore they differ by only a simple atomic substitution. Because these amino acids share a similar structure, one might expect a similar far infrared spectrum, which is the case for serine and cysteine, however, threonine displays a considerably different spectrum with changes in peak position and intensity. The powder zwitterionic forms of serine and cysteine display a similar spectral shape to alanine, which allows us to assign the vibrational modes by making comparisons and also referring to previously published work.

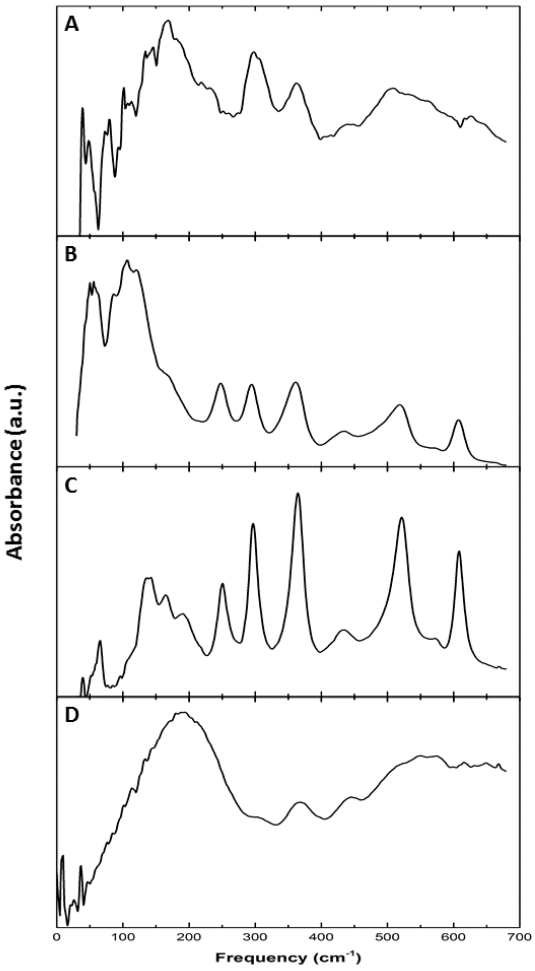
The pI samples for serine (Fig. 6.12) and threonine (Fig. 6.14) have good agreement with their powder spectra, especially in the high frequency region. However, it becomes immediately apparent that the cysteine pI sample (Fig. 6.13) is quite different from the powder form and notable differences are observed as a prominent doublet at 531 cm^{-1} and 513 cm^{-1} and a strong, narrow band at 164 cm^{-1} for the dry film. The changes between the cysteine zwitterionic spectra can be explained by the different crystal structures adopted by the molecules depending on the preparation conditions. For example, the solid L-cysteine·HCL form adopts an orthorhombic crystal with space group $P2_12_12_1$, (Kerr *et al.,* 1973; Kerr *et al.,* 1975; Moggach *et al.*, 2005) whereas the slow cooling of a warm cysteine solution produces a monoclinic crystal with space group $P2_1$ (Gorbitz & Dalhus, 1996). As our film samples are prepared by drying an amino acid solution onto the ATR crystal in a steady stream of warm air, we can assume the film sample had adopted the monoclinic form. On the other hand, serine and threonine are both known to form only an orthorhombic crystal with space group $P2_12_12_1$ under atmospheric pressure (Kistenmacher *et al.,* 1974; Janczak *et al.,* 1997; Moggach *et al.*, 2005), which explains the similarities between the zwitterionic forms. As might be expected, the acidic spectrum of cysteine is almost identical to the

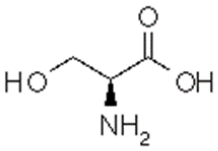
L-cysteine·HCL powder and therefore shares the same peak assignments whereas our basic data does not display well defined peaks.

We assign the highest frequency mode detected in our cysteine spectra, at 640 cm^{-1}, to the COO- bending mode based on the previous assignment of this mode to a 638 cm^{-1} peak for serine (Matei *et al.,* 2005). However, we did not detect this serine vibration in our measurements and, in fact, most theoretical and experimental studies carried out on serine in the far infrared range appear to display some inconsistency (Machida *et al.,* 1979; Pawlukojć *et al*., 2002). We believe the COO- rocking mode appears close to 600 cm^{-1} for serine and cysteine (608 cm^{-1} and 596 cm^{-1}, respectively) and is shifted to slightly lower frequencies for threonine (558 cm^{-1}). Matei *et al,* (2005) have assigned the NH_3^+ torsional mode to a reasonably well resolved band at 446 cm^{-1} for the threonine molecule, yet this assignment may need revaluated as recent calculations suggest that this mode is too weak to be detectable in experimental spectra (Quesada-Moreno *et al*., 2013). Nevertheless, we suggest that the small peaks detected at 435 cm^{-1} and 459 cm^{-1} for serine and cysteine, respectively, are also assigned to the NH_3^+ mode. However, Pawlukojć *et al,* (2002) suggested that a serine peak occurring at 430 cm^{-1} is caused by OH torsion coupled with a hydrogen bond torsion, which could therefore be the origin of these modes.

In the 200 cm^{-1}-400 cm^{-1} region serine and cysteine each display 3 peaks, one of which has been calculated to occur at 295 cm^{-1} in the serine spectrum and stems from the OH torsional mode (Machida *et al.,* 1979), which agrees well with our data. Furthermore, the shift of these cysteine peaks to lower frequencies, compared to serine, is thought to be caused by the increased atomic mass of the SH group compared to the OH group, which would agree with this assignment. In contrast, we have seen in the alanine spectrum that this frequency region is usually associated with backbone deformations and Pawlukojć *et al*, (2002) have assigned the serine 295 cm^{-1} mode to CCαN bending. Therefore, we view the bands in the 200-400 cm^{-1} region as combination of CCαN modes combined with the functional group.

The changes produced by the thiol vs hydroxyl groups become more obvious at low frequencies (<200 cm^{-1}) where hydrogen bond bending and crystal lattice vibrations are detected. As we have previously discussed, the crystal structures of serine and cysteine are significantly different and the difference in hydrogen bond character and strength between the thiol and hydroxyl group is easily demonstrated by the fact that H_2O exists as liquid at standard temperature and pressure, whereas H_2S is a gas. Therefore a combination of these factors, will give rise to a varying number of phonon modes and also frequency shifts of the internal molecular modes. Our low frequency THz-TDS measurements detected modes at 46 cm^{-1}, 56 cm^{-1}, 70 cm^{-1} and 81 cm^{-1} for cysteine, which have good agreement with previously published data (Korter *et al.,* 2006; Rungsawang *et al.,* 2006). Parker, (2013) assign the 56 cm^{-1} and 70 cm^{-1} modes to a liberation and torsion of the carboxylate group, respectively, based on Raman measurements (Minkov *et al*., 2010), whereas others have assigned the 56 cm^{-1} band to the carboxylate torsion (Rungsawang *et al.,* 2006). Additional measurements showed that the 56 cm^{-1} band did not change significantly when cooled to 77 K but all other modes in this region got stronger and shifted to higher frequencies, which would suggest hydrogen bonding. Therefore, we assign the 56 cm^{-1} mode to carboxylate torsion and the remaining bands to hydrogen bonding in agreement with Rungsawang *et al*, (2006).

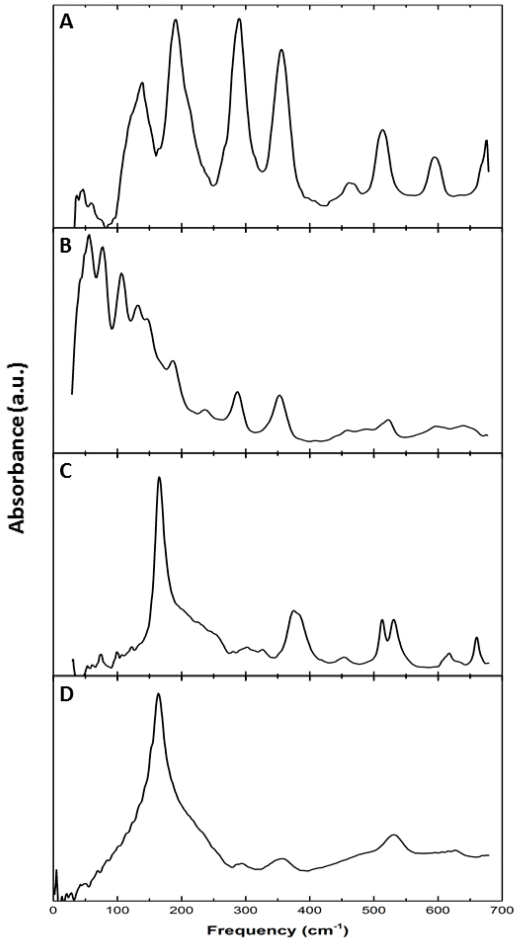


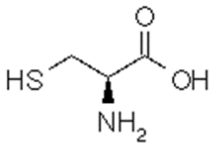


Absorbance frequencies for Serine (cm^{-1})

Cation	Powder	pI	Anion
612	608	609	-
-	-	572	-
506	519	522	523
435	435	434	440
363	361	365	365
298	295	297	-
-	248	251	-
-	-	191	-
166	169	165	191
-	121	137	-
-	106	-	-
-	86	-	-
-	62	66	-

Figure 6.12: Far infrared spectra and absorbance frequencies of serine zwitterions. *The figure shows (A) Cationic (B) Powder (C) pI and (D) Anion spectra. 'Sh' signifies a peak shoulder.*





Absorbance frequencies for Cysteine (cm^{-1})

Cation	Powder	pI	Anion
676	640	660	-
594	596	617	-
514	523	531	531
-	488	513	479
460	459	453	-
356	353	375	357
-	-	325	-
290	287	301	293
215	237	-	-
190	187	164	164
139	147	-	-
118	131	-	-
-	106	-	-
-	77/81*	-	-
-	70*	-	-
-	56/56*	-	-
-	46*	-	-

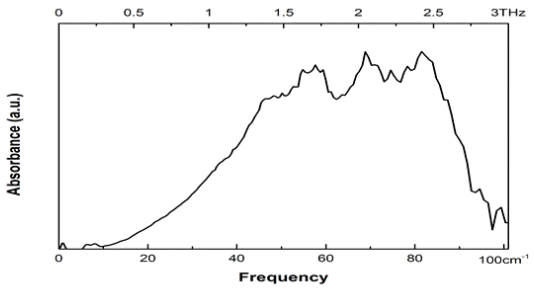
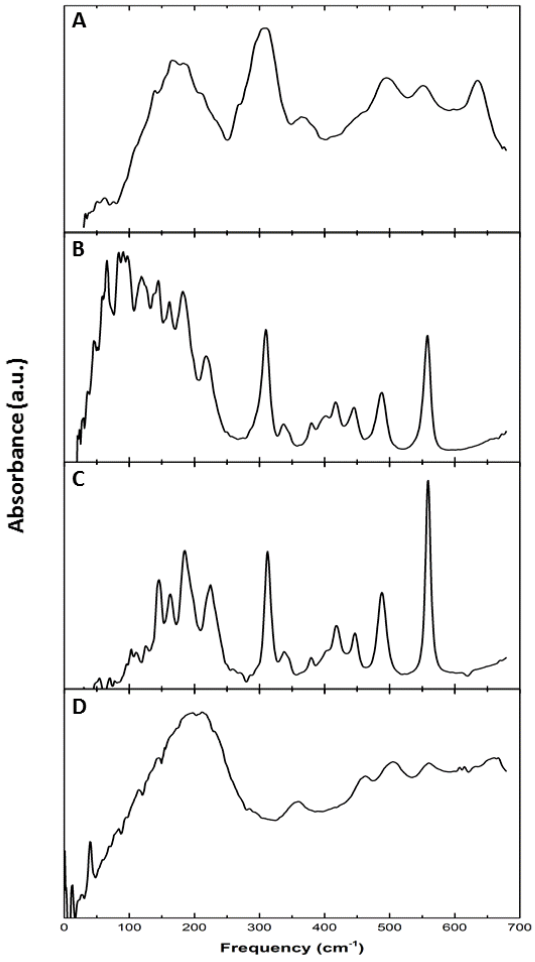
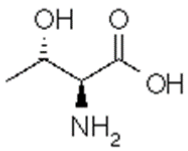


Figure 6.13: Far infrared spectra and absorbance frequencies of cysteine zwitterions. *The figure shows (A) Cationic (B) Powder (C) pI and (D) Anion spectra. The spectrum below shows the low frequency THz-TDS spectrum of the powder sample. 'Sh' signifies a peak shoulder and an asterisk donates the THz-TDS absorbance frequencies.*





Absorbance frequencies for Threonine (cm^{-1})

Cation	Powder	pI	Anion
635	-	-	-
552	558	559	560
496	488	488	505
-	446	447	463
-	417	418	-
-	402	401	-
365	379	379	-
-	337	338	-
310	310	312	360
-	218	225	-
176	182	185	198
-	162	163	-
-	145	146	-
-	119	125	-
-	95	-	-
-	85	-	-
-	69	-	-

Figure 6.14: Far infrared spectra and absorbance frequencies of threonine zwitterions. *The figure shows (A) Cationic (B) Powder (C) pI and (D) Anion spectra. 'Sh' signifies a peak shoulder.*

6.2.6 Asparagine & Glutamine

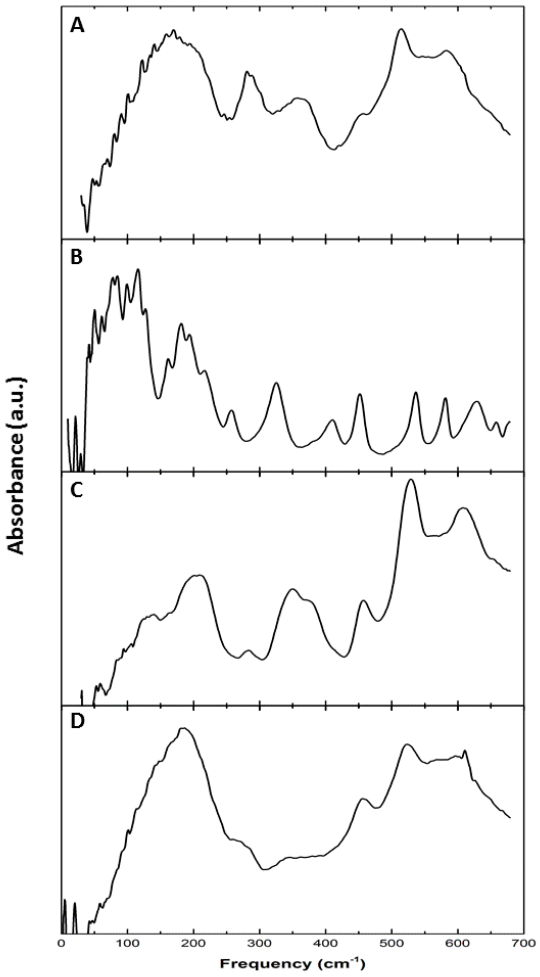
Asparagine and glutamine are considerably larger and more complex than the previous amino acids we have analysed as both possess a carboxamide as the side chain group and glutamine includes an additional methylene group in the side chain relative to asparagine. The amide group on the side chains of these amino acids has a high propensity to hydrogen bond, which makes these molecules important in protein chemistry and also makes analysis of the low frequency region of the FIR spectrum complicated. On first inspection, the powder spectra of asparagine (Fig. 6.15) and glutamine (Fig. 6.16) appear to share a similar spectral shape but there is some variation in peak frequencies.

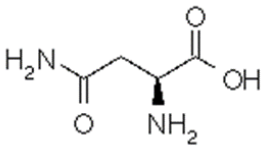
It is immediately apparent that the asparagine pI film sample is significantly different from the powder sample whereas both glutamine zwitterionic samples show very good agreement between peak frequencies; the reason for this being that asparagine can crystallise in two different forms. We assume our powder sample to be anhydrous, which means the molecule is crystallised in the monoclinic $P2_1$ symmetry, (Yamada *et al.,* 2007) whereas fast crystallisation of aqueous solutions results in asparagine monohydrate, which forms crystals with orthorhombic structure and $P2_12_12_1$ space group. (Verbist *et al.,* 1972). This presents significant differences between the backbone and side chain conformations between these crystals, which can be attributed to the different hydrogen bonding environment induced by the presence of the water molecules. Conversely, glutamine is only known to crystallise in the orthorhombic $P2_12_12_1$ symmetry.

The three peaks that are seen in both amino acid powder spectra between 550 cm^{-1}-660 cm^{-1} have previously been assigned, from highest to lowest frequencies, as CCαN out-of plane bending, COO- bending and NH_3^+ bending (Bellamy, 1975; Dhamelincourt & Ramirez, 1993). Both amino acids also share a common feature at 536 cm^{-1} and 537 cm^{-1} for asparagine and glutamine, respectively, which is thought to originate from the torsional mode from the NH_3^+ located on the side group (Casado

et al., 1995). Yet a similar mode it also present on other amino acids (*i.e,* glutamic acid and aspartic acid) that don't possess the amide group, which would make us reconsider this mode to be a COO- vibration. The torsional mode of the NH_3^+ group common to all amino acids was assigned to the glutamine peak at 453 cm^{-1} and asparagine peak at 452 cm^{-1} based of previous assignments of other amino acids (Bellamy*,* 1975; Susi *et al.,* 1983). However, Pawlukojc *et al* (2014) assign the 453 cm^{-1} glutamine peak to a bending, instead of torsional, mode involving the amide side group. Instead, they detected a peak at 441 cm^{-1} using inelastic neutron scattering that they believe to be the NH_3^+ torsion, but this mode was not detectable using FTIR spectroscopy. Nevertheless, further inelastic neutron scattering measurements of the salt of L-asparagine with picric acid, showed a dramatic shift in the NH_3^+ torsion mode from 494 cm^{-1} for the pure form, to 227 cm^{-1} for the cationic form (Pawlukojc *et al.,* 2014), which would suggest that the energy of the NH_3^+ torsion depends on the electron density in the amino acid. This is in good agreement with our data as the intensity of the peak at 456 cm^{-1} (NH_3^+ mode) in the pI sample shifts to 288 cm^{-1} in our cationic sample. Furthermore, a similar trend is also observed in our glutamine spectra, which makes us confident in our assignment of this mode. The basic spectra for both asparagine and glutamine show broad peaks in this region, with less intense features compared to the other samples, making mode assignment difficult.

In the low frequency region the COO- mode can be detected at 108 cm^{-1} and 116 cm^{-1} for glutamine and asparagine, respectively (Pawlukojc *et al.,* 2014). However, in the lowest frequency region, few general features can be extracted and each amino acid displays a distinct spectral profile. Both powder samples exhibit a spectral maximum at ~100 cm^{-1} that encompasses a wide band, formed by a number of overlapping lattice vibrations coupled with hydrogen bonds.

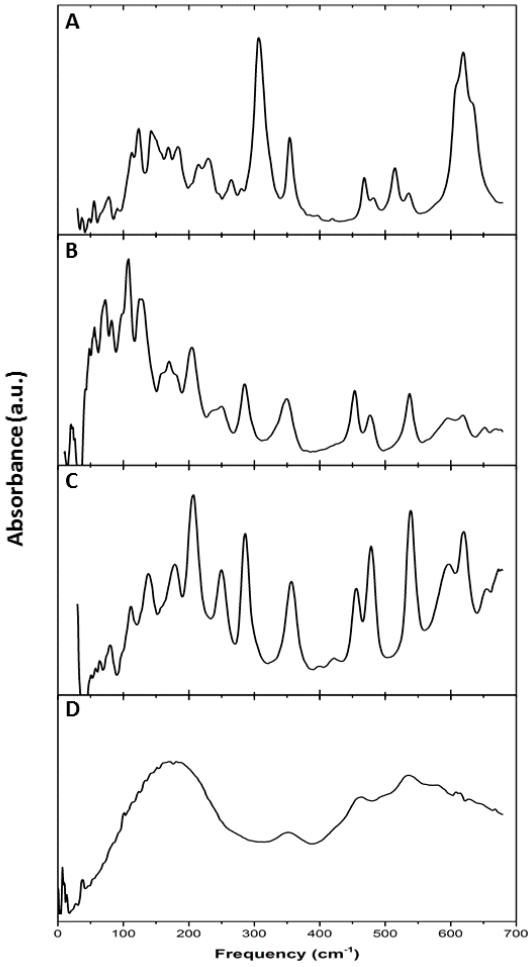


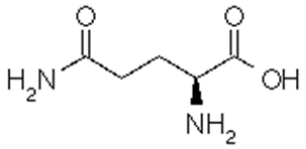


Absorbance frequencies for Asparagine (cm^{-1})

Cation	Powder	pI	Anion
-	658	-	-
583	631	608	-
549	581	-	-
514	536	529	523
451	452	456	456
-	410	-	-
-	380?	377	-
362	326	350	346
285	258	282	280
-	216	-	-
-	194	-	-
170	182	206	182
-	161	162	-
-	127	138	141
-	116	119	-
-	98	-	-
-	85	-	-
-	76	-	-

Figure 6.15: Far infrared spectra and absorbance frequencies of asparagine zwitterions. *The figure shows (A) Cationic (B) Powder (C) pI and (D) Anion spectra. 'Sh' signifies a peak shoulder.*





Absorbance frequencies for Glutamine (cm^{-1})

Cation	Powder	pI	Anion
-	669	-	580
634sh	653	654	-
619	619	620	-
607	596	596	-
535	-	-	-
515	537	538	536
481	476	478	493
468	453	455	464
420	-	420	-
398	-	399	-
354	349	357	352
307	285	286	-
280	-	-	-
265	250	250	-
230	235	-	-
215	204	205	-
183	181	179	175
169	170	169sh	-
143	159	-	-
123	128	138	-
113	108	111	-
-	97sh	-	-
77	82	-	-

Figure 6.16: Far infrared spectra and absorbance frequencies of glutamine zwitterions. *The figure shows (A) Cationic (B) Powder (C) pI and (D) Anion spectra. 'Sh' signifies a peak shoulder.*

6.2.7 Aspartic Acid & Glutamic Acid

The two acidic amino acids, aspartic acid and glutamic acid, are the only amino acids with a negative side group at neutral pH and play important roles in maintaining solubility and ionic character of proteins. Aspartic acid is an alanine molecule with an additional carboxylic acid and glutamic acid has one additional methylene group than aspartic acid.

We can see from our data that the different experimental conditions result in extremely different spectra being produced for both amino acids, which is likely due to the unusual crystallisation properties of these amino acid. Previous studies carried out on anhydrous L-aspartic acid have confirmed that it crystallises in the monoclinic $P2_1$ space group (Bendeif & Jelsch, 2007).

However, to our knowledge, no study has ever been carried out to elucidate if a monohydrate crystal structure can be formed under varying conditions. Based on our FIR data >250 cm^{-1} (Fig. 6.17), and by comparison with the similar amino acid, asparagine, we would suggest that aspartic acid adopts a monohydrate from with orthorhombic $P2_12_12_1$ symmetry when crystallised from an aqueous solution. Unlike aspartic acid, the production and crystallisation of L-glutamic acid has received great deal of attention as L-glutamic acid has a number of commercial applications, one of which being the use of the sodium salt, monosodium glutamate (MSG), as a flavour enhancer. The amino acid is known to form two polymorphs dependent on the crystallisation conditions; the metastable α form and the stable β form. Although both forms adopt an orthorhombic space group $P2_12_12_1$ they each have considerably different lattice parameters and contrasting morphologies (Kitamura *et al.,* 1989; Roelands *et al.,* 2007; Cashell, Corcoran, & Hodnett, 2003). When crystallisation occurs by cooling accompanied by agitation of the solution, the metastable α phase generally forms, that then transforms into the stable β form according to Ostwald's law of stages (Ferrari & Davey, 2004; Garti & Zour, 1997). As the α-form is not commercially available, our powder sample spectrum represents that of the β-form.

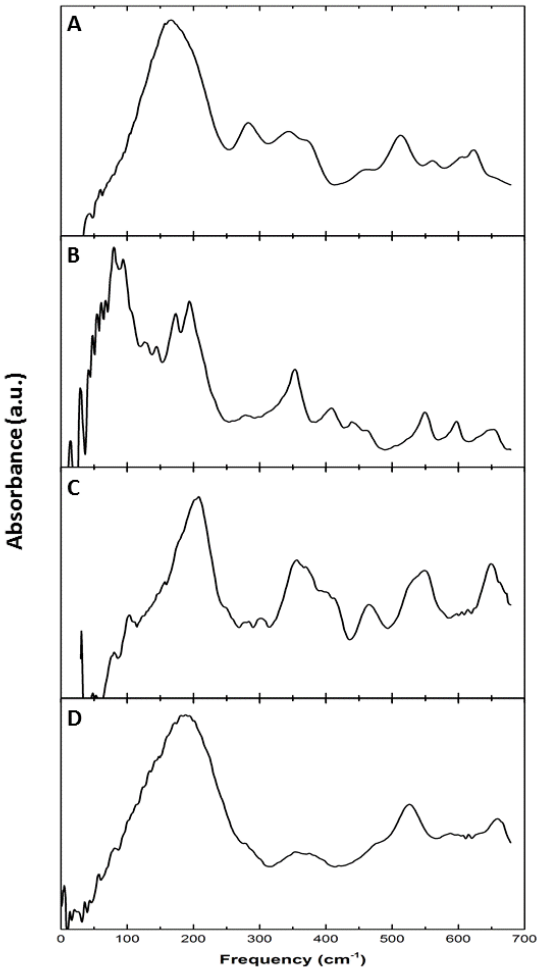
However, the pI dry film sample spectrum is completely different from that of the powder sample and is dominated by three peaks of similar intensity centred at 220 cm^{-1}, 357 cm^{-1} and 540 cm^{-1} (Fig. 6.18). To grow the correct polymorph, strict measures need to be implemented, where temperature and supersaturation have been shown to be important factors (Kitamura *et al.,* 1989). For example, crystals grown from solutions at <25°C mainly produce the α form (Redinha *et al*., 2013), whereas when the temperature is increased, the β-form begins to crystallise until approximately 45°C when the β form becomes prominent (Kitamura *et al.,* 1994; Garti & Zour, 1997). For our measurements, the solutions were dried onto the ATR diamond at ~30°C, which would suggest that our pI sample is a combination of both polymorphs and also explains the lack of any defined structure in our spectrum, which one might expect from a somewhat amorphous sample. The acid sample has a more defined structure, which would suggest that a single polymorph is present in this spectrum. The α form is known to crystallise from solution by utilising a pH-shift precipitation method, that involves adding HCl to acidify the solution (Roelands *et al.,* 2007), it is therefore possible that our cationic sample represents that of the α form. However, Sequeira *et al* (1972) determined that L-glutamic acid can also form a 3rd polymorph, L-glutamic acid·HCl, that can adopt a different structure. Therefore it is not known, which of these polymorphs are present in our acid spectrum.

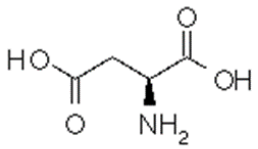
Although our far-IR spectra of both aspartic acid (Fig. 6.17) and glutamic acid (Fig 6.18) powders look somewhat different on first inspection, it is possible to extract common features from both. For example, >500 cm^{-1} we detect 3 bands for both powder samples and the higher frequency modes are thought to represent deformations of the COO- group (Navarrete *et al.,* 1994). Although these peaks vary in frequency and intensity, the COO- group is known to be unstable in amino acids and the frequency can shift between different molecules (Matei *et al.,* 2005). In the same range, a band detected at 597 cm^{-1} in the aspartic acid solid sample is completely absent in the dry film, which might be useful indicator of the polymorphic state. Ramirez *et al.,* (1995) assigned the band at 501 cm^{-1} to the OH torsional mode based

on normal mode calculations but from comparisons to other spectra we believe this mode is more likely to originate from the CCαN bending mode (Navarrete *et al.,* 1994). Matei *et al.* (2005) identified the NH_3^+ torsional mode as a weak shoulder at 463 cm^{-1} for glutamic acid and 449 cm^{-1} for aspartic acid, however we did not detect this glutamic acid mode in our spectrum and we identified the NH_3^+ torsion mode at 464 cm^{-1} for aspartic acid. This reasoning behind this assignment is the NH_3^+ peak in the powder sample, shifts dramatically to 282 cm^{-1} in the acidic sample, similar to what was seen for asparagine.

A prominent, broad feature centred at 350 cm^{-1} easily identifies the β form of glutamic acid and appears to be a tightly packed collection of backbone deformation modes that are somewhat poorly resolved. However, it is possible to assign the two most prominent modes at 396 cm^{-1} and 372 cm^{-1} to the CCαN torsion mode and CCαN bending respectively (Navarrete *et al.,* 1994). For the aspartic acid spectrum, the bands in this region are better resolved but presumably originate from similar vibrations and at lower frequencies the band we detected at 174 cm^{-1} can be assigned to COO- torsion (Navarrete *et al.,* 1994).

Unfortunately, our FTIR system is not reliable at very low frequencies, however we were able to obtain good low frequency data of solid glutamic acid utilising our THz-TDS system. Previous THz-TDS measurements have been carried out on glutamic acid with almost exactly the same sample preparation, yet have identified a variety of different peaks (Yan *et al.,* 2008; Taday *et al.,* 2003) and assignment of these modes have proved difficult (Lopez Navarrete *et al.,* 1995; Chhiba *et al.,* 1994; Ramirez and Lopez Navarrete*,* 1995). However, our data displays a very strong band at 41 cm^{-1} that we assign to a COO- torsion coupled with a CCαN torsion and the well resolved band at 82 cm^{-1} is likely to originate from skeleton torsion or torsion of the whole molecule (Ramirez *et al.,* 1995).

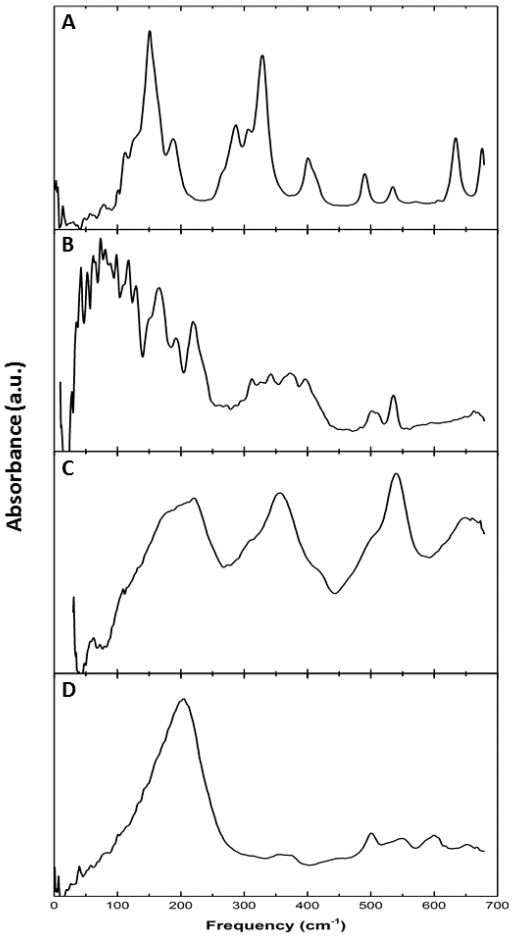


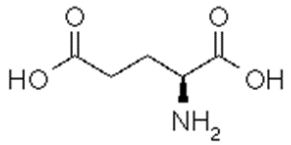


Absorbance frequencies for Aspartic Acid (cm^{-1})

Cation	Powder	pI	Anion
623	653	650	660
562	597	-	-
512	550	550	527
455	464	464	470
-	439	-	-
375	408	408	-
343	353	364	366
-	-	302	-
282	279	281	278
-	-	250	-
167	194	205	183
-	174	174	-
-	145	-	-
-	126	-	-
-	94	104	-
-	80	-	-

Figure 6.17: Far infrared spectra and absorbance frequencies of aspartic acid zwitterions. *The figure shows (A) Cationic (B) Powder (C) pI and (D) Anion spectra. 'Sh' signifies a peak shoulder.*





Absorbance frequencies for Glutamic Acid (cm^{-1})

Cation	Powder	pI	Anion
676	-	-	-
634	-	650	600
534	535	540	550
490	511	500	500
417sh	501	423	446
400	396	-	-
-	372	-	-
-	342	-	-
329	326	357	366
307	312	305	-
287	-	-	-
266	-	-	-
-	218	220	205
188	192	-	-
150	166	172	-
-	149	-	-
126	130	-	-
112	117	-	-
-	82*	-	-
-	64/68*	-	-
-	58*	-	-
-	41*	-	-

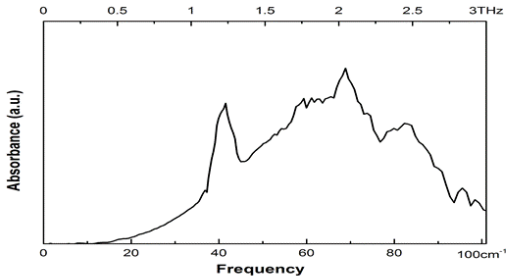


Figure 6.18: Far infrared spectra and absorbance frequencies of glutamic acid zwitterions. *The figure shows (A) Cationic (B) Powder (C) pI and (D) Anion spectra. The spectrum below shows the low frequency THz-TDS spectrum of the powder sample. 'Sh' signifies a peak shoulder and an asterisk donates the THz-TDS absorbance frequencies.*

6.2.8 Arginine, Lysine & Methionine

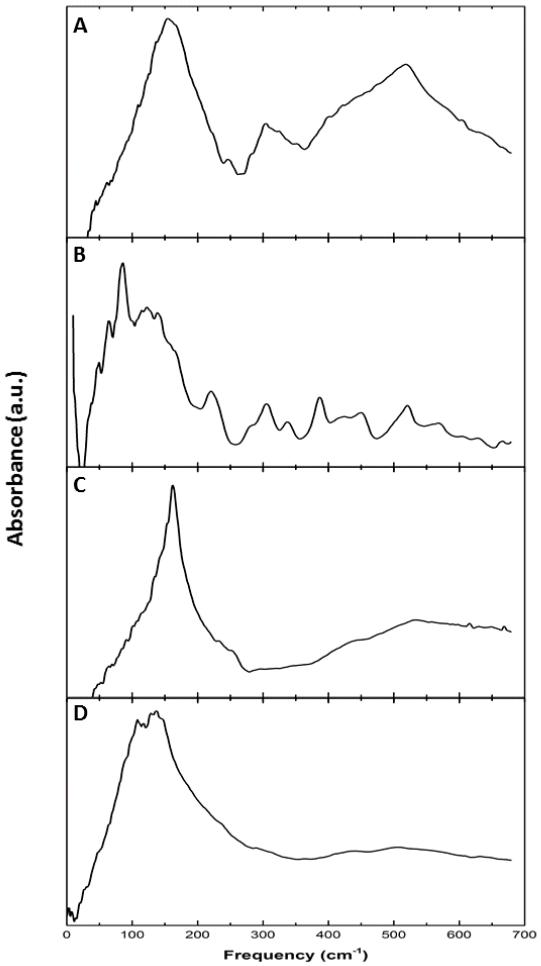
Arginine, lysine and methionine have been grouped together because they have the longest side chains of the amino acids. Arginine and lysine's side groups are both positively charged at physiological pH and lysine's side group is composed of a propylamine while arginine has a complex guanidinium group. Methionine is one of only two sulphur containing amino acids, however, unlike cysteine's thiol side group, methionine's thioether functional group does not usually play a catalytic role but more of a structural role in the protein interior.

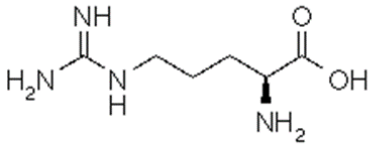
The powder spectrum represents arginine·HCL (Fig. 6.19), which adopts a P_{21} space group with two molecules in the asymmetric unit cell (Meera, Muralidharan, Dhanasekaran, Manyum, & Ramasamy, 2004). Arginine is a notoriously difficult molecule to crystallise, especially in its neutral form, and to date the only reported crystal structures containing neutral arginine are L-arginine dihydrate (Lehmann *et al.,* 1973), DL-arginine dihydrate (Suresh *et al.,* 1994) and DL-arginine monohydrate (Kingsford-Adaboh *et al.,* 2000). Therefore, our data representing the L-arginine molecule at the isoelectric point does not display many well resolved features due the lack of regular structure in the sample. The acidic arginine spectrum displays a more structured spectrum than the pI sample and also shares common features with the powder sample at 305 cm^{-1} and ~250 cm^{-1}. However, the other modes in the acidic spectrum are not well resolved, which would suggest that the sample has only achieved a semi-structured crystal, as to achieve the arginine·HCl P_{21} structure the crystal has to be grown in a 70°C ethanol solution (Meera *et al.,* 2004), which was not the case for our measurements. There is not a great change for lysine and methionine between the different samples, only that the peaks seems to red-shift slightly in the acidic sample.

Although arginine (Fig 6.19), lysine (Fig 6.20) and methionine (Fig 6.21) vary in functional group, a similar spectral shape is seen and common features can be deduced. At >550 cm^{-1}, the arginine spectrum differs considerably from the other two

amino acids and appears to display a collection of overlapping vibrations. As we have seen from our asparagine and glutamine measurements, this region is sensitive to motions of the amide side group. Therefore, we assign this region to the combined motions of the complex guanidinium side group. In the same region we assign the methionine peaks to COO- modes based on comparisons made with methionine derivatives (Grunenberg *et al.,* 1987) and from the same work we assign the methionine band at 447 cm^{-1} to the NH_3^+ torsion mode. As usual, in the 250 cm^{-1}- 450 cm^{-1} range, the bands are assigned to motions around the CCαN.

Data on these molecules in the low frequency range in lacking, however, the stand out feature in the arginine spectra is the strong band around 150 cm^{-1}, which may correspond to a Raman active mode detected at 163 cm^{-1} for the L-arginine·HCl·H_2O crystal and corresponds to wagging of the CCαN mode (Faria *et al.,* 2010).

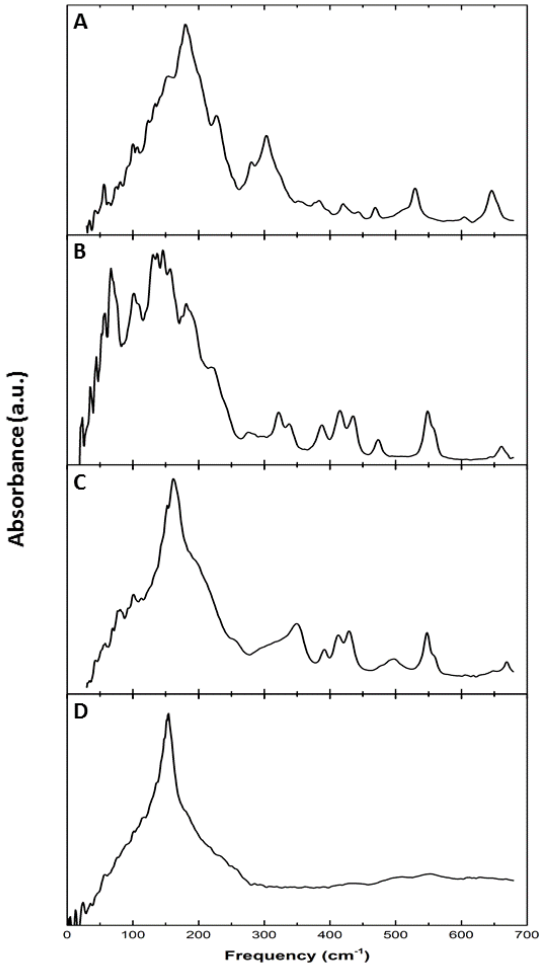


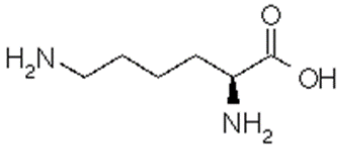


Absorbance frequencies for Arginine (cm^{-1})

Cation	Powder	pI	Anion
-	665	-	-
-	568	-	-
519	521	533	505
-	449	-	-
420	424	432	441
-	387	-	-
351	338	-	-
324	-	-	-
305	305	-	303
280sh	280sh	255sh	-
246	221	232	-
-	165sh	-	-
154	139	162	138
-	120	-	108
-	86	-	-

Figure 6.19: Far infrared spectra and absorbance frequencies of arginine zwitterions. *The figure shows (A) Cationic (B) Powder (C) pI and (D) Anion spectra. 'Sh' signifies a peak shoulder.*

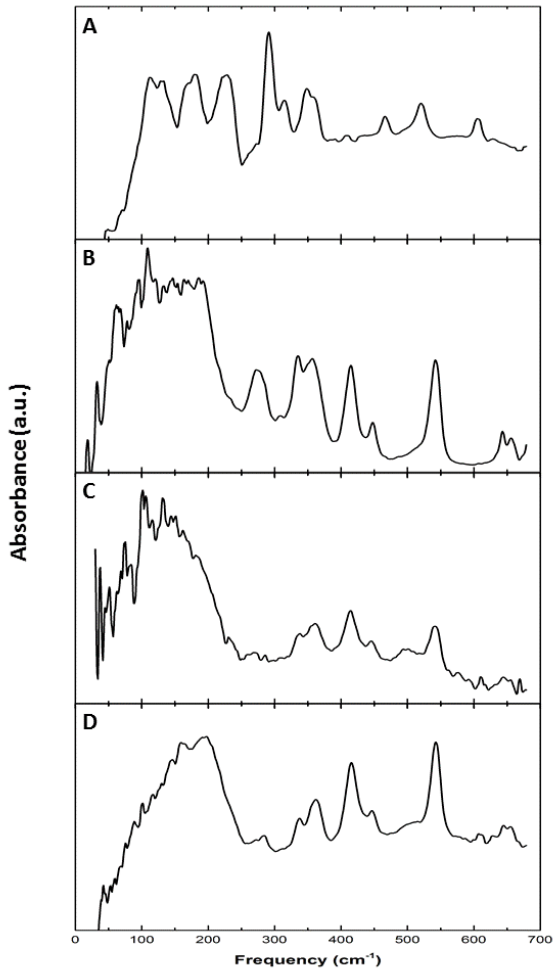


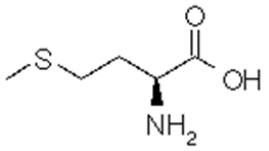


Absorbance frequencies for Lysine (cm^{-1})

Cation	Powder	pI	Anion
646	662	669	-
-	559sh	559sh	-
529	549	548	550
509	-	-	-
469	474	496	503
445	435	429	434
420	415	413	-
386	388	392	-
356	338	349	-
303	322	-	-
281	278	255	-
227	219	-	-
180	181	199	183
-	156	162	153
-	146	-	-
-	137	-	-
132	130	-	-
100	101	100	-
-	68	-	-

Figure 6.20: Far infrared spectra and absorbance frequencies of lysine zwitterions. *The figure shows (A) Cationic (B) Powder (C) pI and (D) Anion spectra. 'Sh' signifies a peak shoulder.*





Absorbance frequencies for Methionine (cm^{-1})

Acidic	Solid	pI	Basic
-	656	656	-
606	643	645	-
521	542	542	542
-	-	499	-
467	447	445	447
-	415	415	414
361	357	360	361
347	335	337	336
316	309	-	-
291	273	-	285
229	-	-	-
181	-	-	197
167	-	-	159
131	-	-	-
113	-	100	-

Figure 6.21: Far infrared spectra and absorbance frequencies of methionine zwitterions. *The figure shows (A) Cationic (B) Powder (C) pI and (D) Anion spectra. 'Sh' signifies a peak shoulder.*

6.2.9 Phenylalanine, Tyrosine, Tryptophan & Histidine

This group of amino acids are characterised by the presence of an aromatic ring in their side group. Phenylalanine is composed of an alanine molecule with a benzyl group attached and is the most hydrophobic of these amino acids, which is why it is almost always found buried within the protein core. Tyrosine, histidine and tryptophan are similar in structure to phenylalanine but tyrosine has an additional hydroxyl group attached to the end of the benzyl ring, while for histidine the phenyl group is substituted by an imidazole ring. In the case of tryptophan, the side group is an indole group that makes tryptophan the largest of the amino acids and it is often found on the protein surface due to the ability of indole N-H to donate a hydrogen bond.

Like L-cysteine, L-histidine is one of the amino acids that is known to crystallise in different forms: the stable A form, which has an orthorhombic space group $P2_12_12_1$ with 4 molecules per unit cell (Madden *et al.,* 1972) and the metastable B form that has a monoclinic space group $P2_1$ with 2 molecules per cell (Madden *et al.,* 1972.). It requires a high concentration of ethanol in the solvent in order to produce the metastable B form and we therefore assume that all our spectra represent samples in the stable A form (Kitamura *et al.,* 1994; Kitamura, 1993), which is confirmed by the similarity of peak frequencies between the powder and pI dry film samples.

The other amino acids also share good similarity between the powder and pI samples, although for phenylalanine an additional two bands appear at 275 cm^{-1} and 409 cm^{-1} in the pI measurement compared to the powder (Fig. 6.22). This is because phenylalanine is known to crystallise in both the anhydrous (powder) and also the monohydrate form (dry film) (Cuellar *et al.,* 2006).

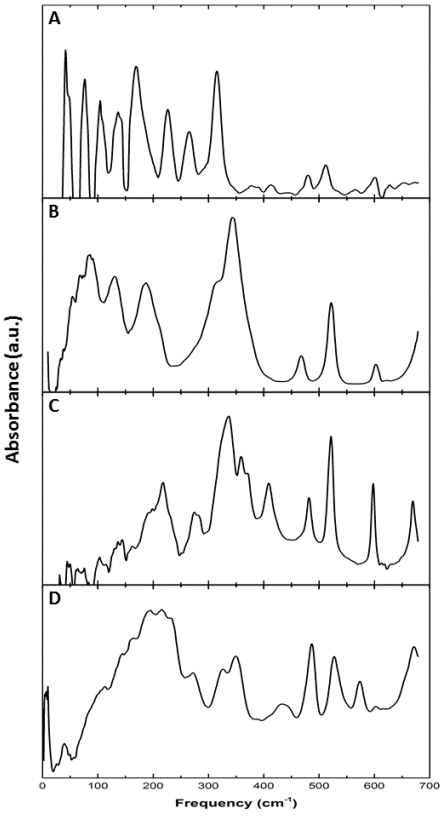
Although phenylalanine and tyrosine only differ by a single hydroxyl group, the powder spectrum changes enormously between the two molecules, with phenylalanine producing a reasonably simple spectrum (Fig. 6.22) and tyrosine displaying a complex spectrum encompassing a large number of sharp, well resolved

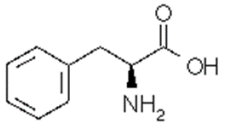
peaks (Fig. 6.23). By performing normal mode calculations on poly-L-tyrosine, Bahuguna *et al.* (1997) predicated a ring vibration should be present around 624 cm^{-1} and although we did not detect a band at this specific frequency, it could correspond to our tyrosine peak detected at 647 cm^{-1}. We also detected bands at 625 cm^{-1} and 627 cm^{-1} for tryptophan (Fig. 6.24) and histidine (Fig. 6.25), respectively, which have good agreement with Bahuguna's assignment. Another vibration of the benzene ring has been calculated to occur around 535 cm^{-1} (Grace *et al.,* 2002), which is in perfect agreement with a peak detected in histidine and could also correspond to the 523 cm^{-1} and 527 cm^{-1} peaks in phenylalanine and tryptophan. In fact, all of the aromatic group amino acids, excluding tryptophan, display a strong peak in this region, making us confident about the assignment. However, it is not surprising that tryptophan does not display this band, as it possess a much larger indole subgroup that would alter the vibration, and the increased complexity of this group becomes clear as we detect nine vibrations above 500 cm^{-1}, which is significantly more than the other aromatic amino acids.

In the 300-400 cm^{-1} region there is predicted to be three peaks for tyrosine that correspond to the ring-OH torsion, C-H in plane bending and CCαN bend mode (Grace *et al.,* 2002). In this region we detect a very strong band in the phenylalanine spectrum at 344 cm^{-1} with a shoulder on the low frequency side, which corresponds to the C-H plane bending and CCαN bending modes and is also in agreement with studies that have assigned the phenylalanine mode in this region to backbone bending (Xie, He, Miller, Sclavi, & Chance, 1999). Conversely, Raman studies (Grace *et al.,* 2002) have predicted a benzene-CH in plane bend to occur at 325 cm^{-1} for tyrosine, which could correspond to the shoulders observed at 318 cm^{-1} for tyrosine or 316 cm^{-1} for phenylalanine.

In the very low frequency region (>100 cm^{-1}) it is difficult to assign peaks, however, some low frequency peaks were reproducible in our FTIR spectra. For example, the lowest frequency peak detectable in our tyrosine spectrum, occurring at 33 cm^{-1}, has

previously been observed in various studies and assigned as the torsion of the entire side chain about the C-bond (Grace *et al.,* 2002). Furthermore, we were able to carry out THz-TDS measurements on solid histidine and detected the lowest frequency vibration as a strong band centred at 30 cm^{-1}, which has been assigned as a torsional vibrational mode (Rungsawang *et al.,* 2006).

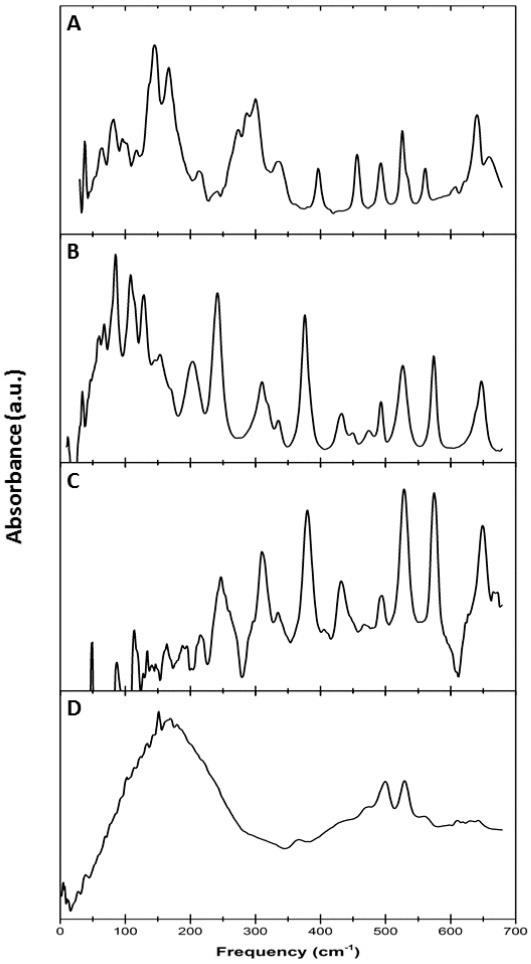


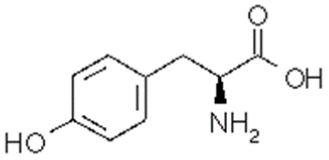


Absorbance frequencies for Phenylalanine (cm^{-1})

Cation	Powder	pI	Anion
-	-	669	670
-	604	598	574
512	523	522	527
479	469	482	487
415	-	409	434
391	-	370	-
380	-	359	349
315	344	337	326
265	316	275	269
226	-	-	232
169	187	218	215
-	-	193	195
135	129	138	165
105	85	103	-

Figure 6.22: Far infrared spectra and absorbance frequencies of phenylalanine zwitterions. *The figure shows (A) Cationic (B) Powder (C) pI and (D) Anion spectra. 'Sh' signifies a peak shoulder.*

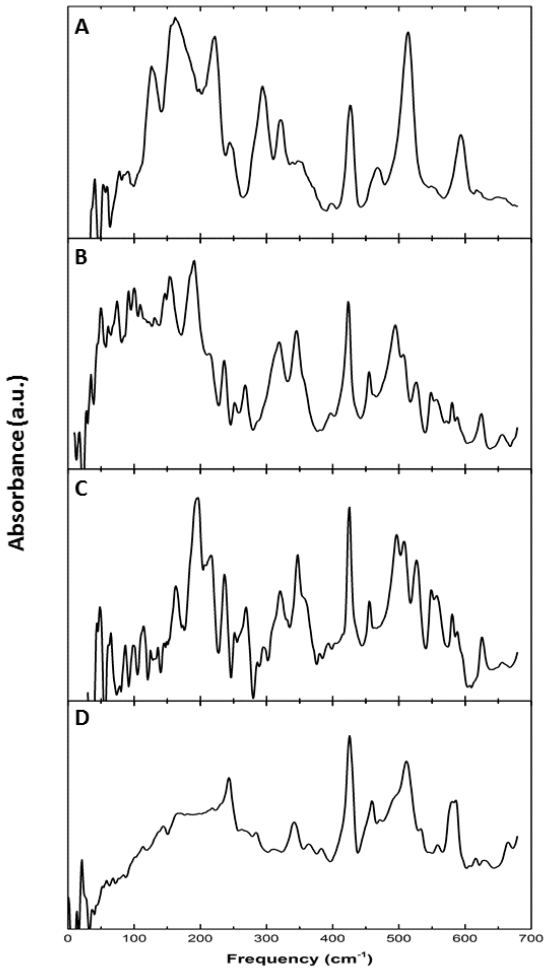


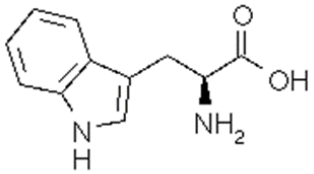


Absorbance frequencies for Tyrosine (cm^{-1})

Cation	Powder	pI	Anion
660	-	-	-
640	647	648	-
561	574	575	559
526	527	528	529
493	493	492	500
456	474	-	474
-	448	450	-
-	432	433	423
396	376	376	366
334	334	334	-
-	318	-	-
300	310	309	-
286	242	245	-
274	-	-	-
241	-	-	-
213	203	208	-
167	153	-	165
145	144	-	-
118	128	-	-
-	114sh	-	-
95	108	-	-
81	84	-	-
-	81	-	-
64	67	-	-
-	58	-	-
-	33	-	-

Figure 6.23: Far infrared spectra and absorbance frequencies of tyrosine zwitterions. *The figure shows (A) Cationic (B) Powder (C) pI and (D) Anion spectra. 'Sh' signifies a peak shoulder.*

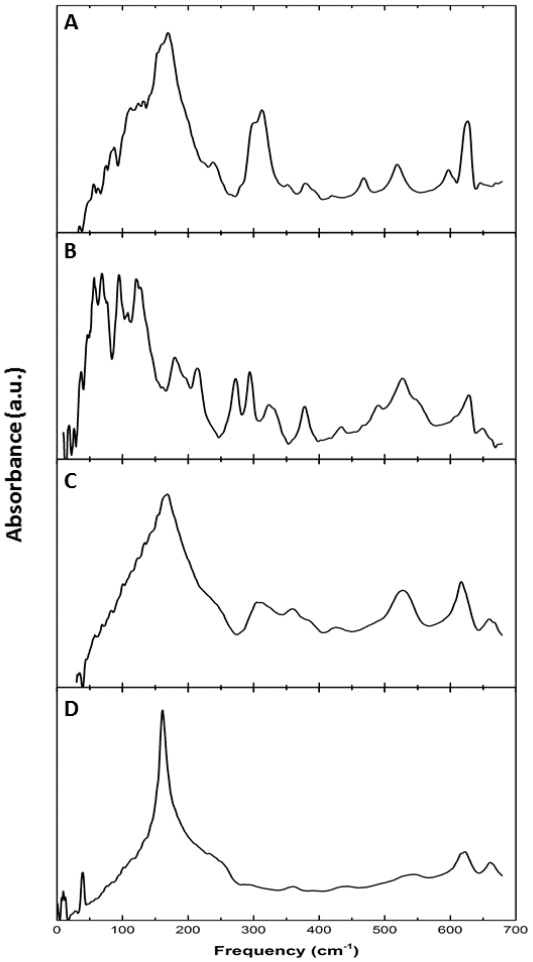


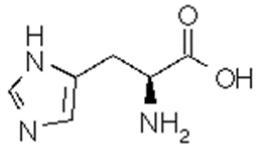


Absorbance frequencies for Tryptophan (cm^{-1})

Cation	Powder	pI	Anion
-	656	656	665
-	625	625	-
593	588	588	587
-	581	581	580
-	572	-	-
-	556	557	558
-	549	549	-
-	527	527	533
-	506	508	511
514	495	497	490
468	455	455	459
426	423	425	425
398	397	358	383
-	-	-	364
350	345	347	341
321	319	320	311
294	268	269	284
-	252	252	262
243	236	236	243
-	213	206	-
221	191	196	-
162	152	163	-
126	-	-	-
-	72	-	-

Figure 6.24: Far infrared spectra and absorbance frequencies of tryptophan zwitterions. *The figure shows (A) Cationic (B) Powder (C) pI and (D) Anion spectra. 'Sh' signifies a peak shoulder.*





Absorbance frequencies for Histidine (cm^{-1})

Cation	Powder	pI	Anion
-	650	660	662
625	627	616	622
599	-	-	-
-	535	-	-
520	527	528	546
468	491	484	-
-	433	425	436
391sh	-	385	-
381	378	360	360
350	325	308	291
310	295	-	-
297	272	-	-
238	214	227	231
-	192sh	-	-
-	181	-	-
165	124	168	161
108	94	-	-
-	79*	-	-
-	69 68*	-	-
-	54*	-	-
-	30*	-	39

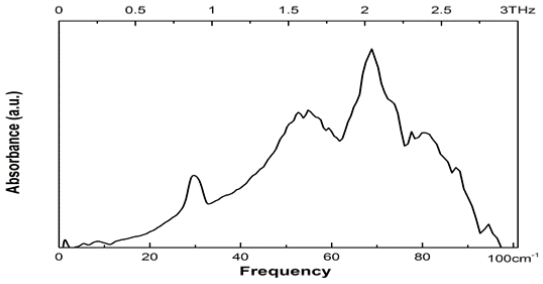


Figure 6.25: Far infrared spectra and absorbance frequencies of histidine zwitterions. *The figure shows (A) Cationic (B) Powder (C) pI and (D) Anion spectra. The spectrum below shows the low frequency THz-TDS spectrum of the powder sample. 'Sh' signifies a peak shoulder and an asterisk donates the THz-TDS absorbance frequencies.*

6.3 Peptides

Amino acids connect via a peptide bond to form polypeptides, which can be precursors to proteins or can act as signalling molecules, such as hormones (Siddle & Hutton, 1991). However, cysteine also has the potential to link via disulphide bonds, which play important roles in biochemistry. To expand upon our initial amino acid data we obtained the far infrared spectrum of two biologically relevant small biomolecules, cysteine and glutathione, in order to determine how disulphide bonds effect the far infrared spectrum and also elucidate the contribution individual amino acids give to a larger peptides spectrum.

6.3.1 Cystine

Cystine is formed by the joining of two cysteine molecules via oxidation of the thiol group, forming a covalent bond via disulphide linkage (Fig. 6.26). These bonds play an important role in the stabilisation and spatial structure of proteins and are therefore a good target to monitor protein behaviour. By analysing the simple cystine molecule we hope to gain a deeper understanding of these bonds in the far infrared region and elucidate how the far infrared spectrum changes from analysing small biomolecules, such as amino acids, to more complex molecules.

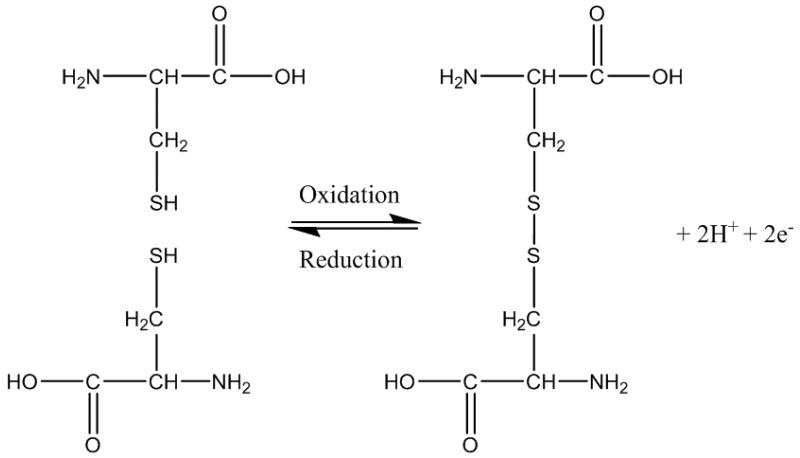


Figure 6.26: Comparison of cysteine (left) and cystine (right) structures. *The two molecules can be interconverted by oxidation or reduction reactions.*

By comparing cystine with its monomeric form, cysteine, we see a considerable difference in the spectra of these molecules (Fig. 6.27). As one would expect of a larger molecule, cystine displays more vibrations than cysteine, some of these stemming from the disulphide bond. It has been demonstrated that the dihedral angle of the disulphide bond effects the frequency of the S-S stretching vibration (Van Wart *et al.,* 1973), which can be used to help elucidate conformation in small biomolecules and structural changes in proteins. The disulphide bridge is known to adopt three basic conformations: gauche–gauche–gauche (+/-90° - +/-90° - +/-90°), gauche–gauche–trans (+/-90° - +/-90° - 180°) and trans–gauche–trans (180°, +/-90°, 180°), which are each associated with Raman lines at 510 cm^{-1}, 525 cm^{-1} and 540 cm^{-1}, respectively. In our cystine spectrum we detect a strong peak at 538 cm^{-1} and also a very weak peak at 500 cm^{-1}, which we assign to these stretching motions of the S-S bonds. Due to the difference in intensities of these peaks, we would assume the trans-gauche-trans conformation to be the dominant species in our sample with negligible amounts in the gauche–gauche–gauche conformation. Conversely, it has previously been suggested that the gauche–gauche–gauche conformation dominates in cystine (Van Wart *et al.,* 1973). Furthermore, studies carried out on cyclic model compounds (Weiss-Lopez *et al.,* 1986), diallyl trisulphide and allyl methyl disulphide (Devlin *et al.,* 1990), would

suggest that the bending mode of the disulphide bond is present in the 220 cm^{-1}-250 cm^{-1} range. Therefore, we assign the strong vibration detected at 230 cm^{-1} in our spectrum to the S-S bending mode, which is also in agreement with previous Raman measurements carried out on the cystine molecule. (Brandt *et al.,* 2008). Weiss-Lopez *et al*., (1986), also assigned peaks between 130 cm^{-1} - 70 cm^{-1} to the coupled bending and torsion motions of the disulphide bridge and it is in this region we detect a collection of vibrations centred at ~150 cm^{-1} that we assign to the same origin.

From our previous assignments of the monomeric cysteine amino acid, we would consider the cystine bands >600 cm^{-1} to stem from COO- modes and the peak detected at 451 cm^{-1} due to NH_3^+ torsion. The remainder of bands in the 200-400 cm^{-1} region - that have not been assigned to disulphide bond vibrations - should come from OH torsional modes. The changes in the spectrum of cystine compared to cysteine, not only originate from the different masses of the molecules but, as one would expect, cystine crystallises in a different form. The cysteine spectrum shown in Figure 6.27 represents the molecule crystallised in an orthorhombic $P2_12_12_1$ space group, whereas cysteine is known to crystallise in either a tetragonal phase $P4_1$ (Chaney *et al.,* 1974) or a hexagonal phase $P6_122$ (Dahaoui *et al.,* 1999), yet the S-S bond distance and angles do not differ significantly between these polymorphs. Clearly, these differences should have a profound effect on the low frequency region of our spectra. Below 100 cm^{-1} we only detect two bands for cystine at 95 cm^{-1} and 66 cm^{-1}, yet previous studies utilising Raman spectroscopy and THz-TDS have detected seven cystine modes in the same frequency range (Brandt *et al.,* 2008). The authors calculated normal modes of the isolated cystine molecule in the low frequency range that corresponded to torsions and deformations of the monomer, but could not assign these modes to the low frequency spectrum of the crystallised molecule. In addition, the combination of the disulphide bonds along with hydrogen bonding between the ammonium and carboxylate moieties (NH ··· O) leads to the formations of layers in the cystine crystal and DFT calculations have predicted that the lowest frequency

modes (<30 cm^{-1}) are a consequence of the relative motions of these hydrogen bonded layers (Brandt *et al.,* 2008).

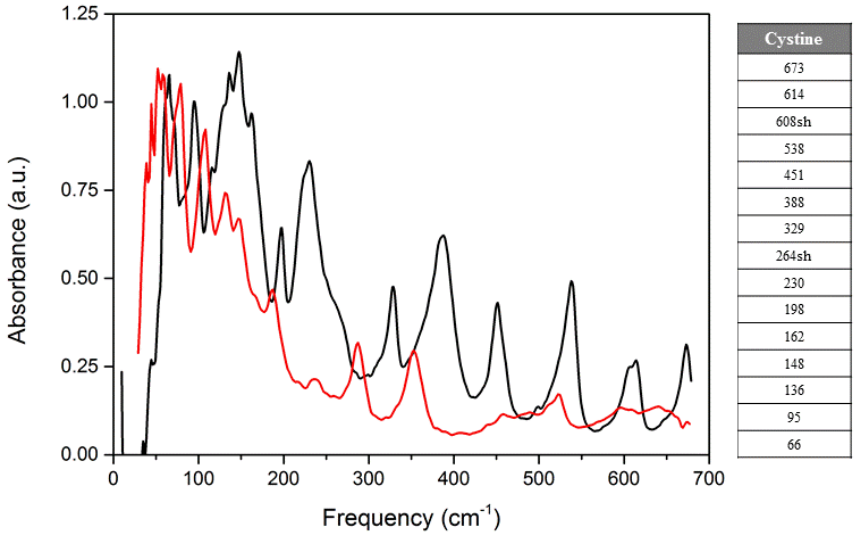


Figure 6.27: Far infrared spectra and absorbance frequencies (cm^{-1}) of cystine. *The spectrum of cysteine (red) has been included in the graph as a comparison to the cystine molecule (black). 'Sh' denotes a peak shoulder.*

6.3.2 Glutathione

Glutathione (GSH) is an antioxidant tripeptide of high biological significance as it plays a key role in many cellular process, such as protecting cells against oxidative stress, radiation and xenobiotics (Pompella *et al.,* 2003). The tripeptide is formed by a γ-peptide linkage between the carboxyl group of a glutamate and the amine group of a cysteine, followed by the addition of a glycine molecule to the carboxyl of the cysteine. The thiol group on the cysteine residue acts as an extremely effective reducing agent by serving as an electron donor to unstable molecules within the cell. However, by donating an electron, glutathione becomes reactive itself and must stabilise by binding with another reactive glutathione molecule via a disulphide bond, forming oxidised glutathione (GSSG) (Fig 6.28). This process can be reversed by the

enzyme, glutathione reductase in the presence of NADPH so glutathione can continue its antioxidant role.

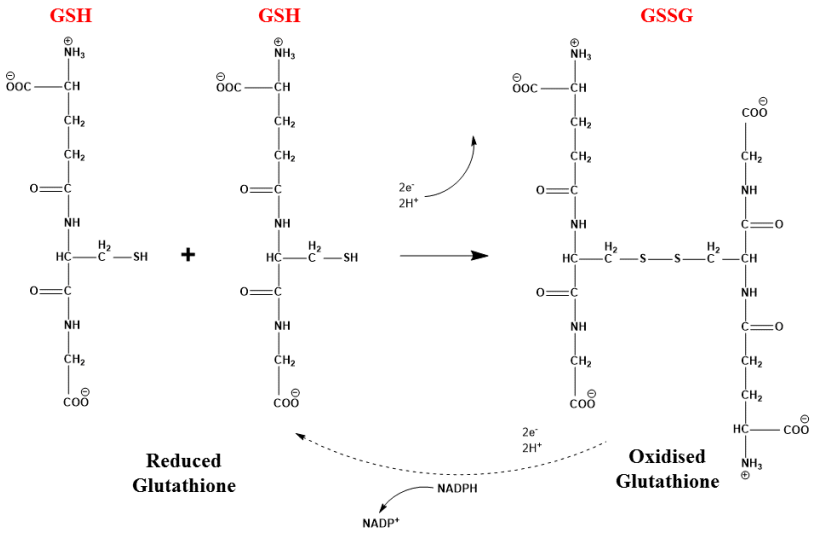


Figure 6.28: Comparison of GSH and GSSG structures. *The two molecules can be interconverted by oxidation or reduction reactions.*

By analysing GSH and GSSG we can not only observe the extent to which the far infrared spectrum changes from amino acids to short peptides but also elucidate the spectral signature of biologically relevant small molecules. As one would expect of a tripeptide, the far infrared spectrum of reduced glutathione displays considerably more peaks than was seen for the single amino acids. In contrast, the spectrum of oxidised glutathione is reasonably featureless compared to that of the reduced form (Fig 6.29), which is not surprised as spectral overlap becomes an issue for the larger molecules, conforming to the 3*N*-6 rule.

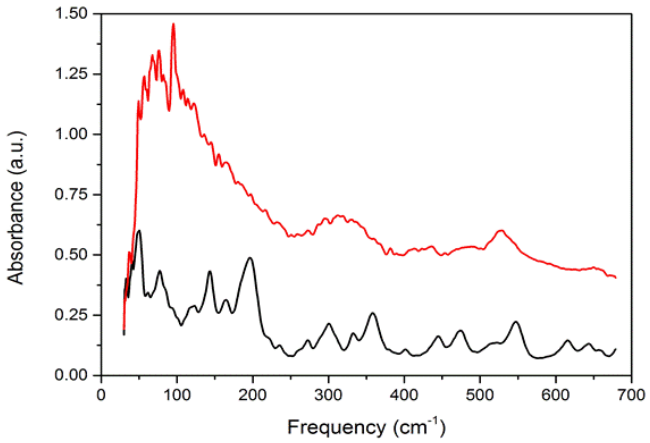
By comparing the monomeric units of the glutathione peptide we can extract information regarding the vibrations within the tripeptide. From our previous measurements on the amino acid constituents of glutathione, we can be confident that the vibrations >600 cm^{-1} in the reduced form are caused by bending and wagging motions of the COO- group in accordance with tip-enhanced Raman scattering

measurements performed on reduced glutathione (Deckert-Gaudig *et al.,* 2009). The feature detected in GSH between 500-550 cm^{-1}, with two peaks centred at 548 cm^{-1} and 523 cm^{-1} has previously been detected in both the glutamic acid and cysteine spectra at slightly lower frequencies, which would suggest that this mode stems from the amino acid elements and not the molecule as a whole. Unfortunately, we were not able to conclusively assign these modes in the amino acid spectrum due to differing assignments in the literature. For example Parker *et al.,* (2013) assign the cysteine vibrations in this region to CCαN bending whereas, Mink *et al.,* (2012) assign CO_2 scissoring. In the same frequency region, we detect a single peak at 526 cm^{-1} for GSSG which may be attributed to similar modes or, alternatively, an S-S stretching mode. As we have previously described for the cystine spectrum, this would suggest that the disulphide linkage for oxidised glutathione adopts a gauche–gauche–trans (+/-90° - +/-90° - 180°) conformation.

Usually, in the region between 425-500 cm^{-1} the vibrations are assigned to NH_3^+ torsions for single amino acids, but in the glutathione molecule, cysteine and glycine have lost their amine group due to the formation of the peptide bond. Therefore we tentatively assign the vibrations in this region to the glutamic acid amine torsion coupled with other modes. The glutathione peak detected at 358 cm^{-1} matched extremely well with bands detected in all of the amino acid monomers and it is in this region that we usually assign the OH torsional mode or deformations around the α-carbon.

The strongest peak seen in glutathione occurs at 197 cm^{-1} and is common to almost all amino acids, which we assign to COO- torsions. Below this frequency, we cannot assign vibrations based on comparisons to the single amino acids as differences in crystal structure and intermolecular vibrations need to be taken into account. Yet, DFT calculations have been carried out on the isolated GSH molecule at low frequencies and shown good agreement with experimental THz-TDS measurements (Wang *et al.,* 2005; Rutz *et al.,* 2005). In these studies, the experimental modes were slightly blue-

shifted compared to the calculations, which may be an influence of the crystal structure. However, as the overall shape and number of vibrations remains the same, the modes must originate from intramolecular modes.



Glutathione (reduced) (GSH)	Glutathione (oxidised) (GSSG)
658	-
644	-
615	-
548	-
523	526
474	482
445	-
402	-
358	-
333	322
300	-
273	-
261	-
235	-
221	-
197	-
165	-
144	-
123	-
97	95
85	-
77	-
62	-
58sh*	-
50/51*	-
40*	-
28*	-

Figure 6.29: Far infrared spectra and absorbance frequencies (cm^{-1}) of glutathione. *Comparison of the reduced (Black) and oxidised (Red) forms of glutathione. 'Sh' signifies a peak shoulder and an asterisk denotes THz-TDS absorbance frequencies.*

The low frequency spectrum of GSH was also obtained by THz-TDS and is shown earlier in this chapter in section 6.1. The molecule displays strong, well resolved peaks at 28 cm^{-1} (0.84 THz), 40 cm^{-1} (1.20 THz), 51 cm^{-1} (1.53 THz) and a weaker shoulder at 58 cm^{-1} (1.74 THz) (Fig. 3). These peaks showed excellent agreement with the literature and they have all been assigned to torsions within the glutathione molecule as a whole (Wang *et al.,* 2005). Unfortunately, we were not able to record the THz-TDS spectrum of oxidised glutathione but this has previously been obtained by

Yamamoto and co-workers from 1-3 THz (Yamamoto *et al.,* 2005). The spectrum was seen to be completely featureless, showing an increase in absorption with frequency, which is similar to the spectra of other large biomolecules in this frequency range.

6.4 Nucleic Acid Components and DNA

Due to the importance of DNA and RNA in all forms of life, a complete understanding of the physical properties of these molecules is imperative. However, similar to amino acids, a complete catalogue of the far infrared spectra of the constituents of DNA and RNA is lacking. Here we provide the far infrared spectra of DNA components, starting with the most fundamental units – nucleobases - before progressing to nucleosides, nucleoside monophosphates (NMP's) and deoxy nucleotide triphosphates (dNTP's) to determine how the increase in molecular complexity effects the far infrared spectrum. We finish by analysing the spectrum of long chain DNA molecules and identifying how individual nucleotides contribute to the spectrum. Normal mode and DFT calculations could not be applied to these molecules at present and the spectral peaks are assigned from modes in the literature.

6.4.1 Nucleobases

Nucleobases, or nitrogenous bases, are nitrogen containing compounds and a key component of DNA and RNA, as these molecules are responsible for the base pairing and stacking interactions that gives DNA its characteristic double helix structure. In addition, these molecules also play important roles in other biochemical processes such as cell signalling (cGMP and cAMP) and production of cofactors for enzymatic reactions (NAD and NADPH).

The nucleobases can be split into two categories based on their ring structure: The purines and pyrimidines (Fig. 6.30). Thymine, cytosine and the RNA base uracil are all derivatives of pyrimidine and are characterised by nitrogen's at positions 1 and 3

in the ring. Guanine and adenine are purine bases and consist of an imidazole ring fused to a pyrimidine ring. Due to the different structures and geometry of the double helix, guanine can only hydrogen bond to cytosine and thymine to adenine within the DNA molecule.

In order to understand the terahertz spectra of DNA macromolecules we began analysis at the most fundamental level with single nucleobases.

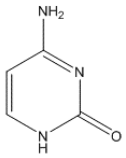
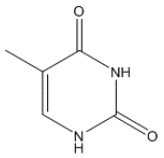
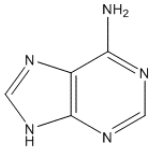
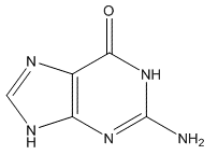
Pyrimidines		Purines	
Cytosine	Thymine	Adenine	Guanine
			

Figure 6.30: Structural comparison of the four DNA nucleobases

6.4.1.1 Pyrimidines

6.4.1.1.1 Thymine

The far infrared spectrum of thymine is shown in figure 6.31. The highest frequency mode detected in our thymine spectrum is a very weak band located at 617 cm^{-1} that is assigned to the in-phase CO bending mode, coupled with the C-CH_3 bend. An out of phase CO bending mode is also detected at 405 cm^{-1} in agreement with previously published work (Florian *et al.,* 1993). However the same mode has also been attributed to a band at 432 cm^{-1}, which corresponds to the peak at 429 cm^{-1} on our spectrum (Mathlouthi *et al.,* 1984)

Assignment of the NH modes for thymine has been carried out under a number of different experimental conditions. For examples, in Ar matrix-isolated thymine

(Nowak *et al.,* 1989; Graindourze *et al.,* 1990; Les *et al.,* 1992), NH wags are assigned to vibrations at 662 cm^{-1} and 545 cm^{-1} and although we don't detect these bands, these modes are expected to vary greatly in polycrystalline samples due to hydrogen bonding. A more recent study detected a NH mode at 560 cm^{-1} in a polycrystalline sample, which should correspond to our peak detected at 558 cm^{-1} (Florain *et al.,* 1993). However, vibrations of the pyrimidine ring are also expected in this region and the band at 558 cm^{-1}, along with the 473 cm^{-1} peak, can be assigned to bending of the ring, while the vibrations at 207 cm^{-1} and 429 cm^{-1} are caused by ring torsions (Zhang *et al.,* 1998). Although opinions are split over the assignment of these bands, from comparisons with polarisation studies we are confident that the high frequency in-plane bending mode is correctly assigned (Lewis *et al.,* 1984). In addition, the 207 cm^{-1} band has also previously been assigned to methyl torsions instead of ring torsions (Lagant *et al.,* 1992) but additional studies by Zhang *et al.,* disproved this assignment as the peak did not shift upon deuteration.

The two peaks detected at 321 cm^{-1} and 284 cm^{-1} are in perfect agreement with previous work and have been assigned to vibrations of the methyl group based on their shift to lower frequencies in deuterated samples (Zhang *et al.,* 1998). From previous calculations we would suggest that the higher frequency mode stems from wagging of the methyl group while the lower frequency mode is due to in-plane bending of the same group (Florian *et al.,* 1993; Les *et al.,* 1992; Rush *et al.,* 1995)

In the thymine crystal, two pairs of intermolecular hydrogen bonds are present and therefore we expect to see four bands in the low frequency region originating from these bonds. Similar peaks to those detected at 170 cm^{-1}, 97 cm^{-1} and 71 cm^{-1} have been observed in single crystals of methyluracil and methylthymine and been assigned to intramolecular stretching vibrations of hydrogen bonds (Harada *et al.,* 1970; Ozeki *et al.,* 1969) and subsequent calculations also support this assignment (Lagant *et al.,* 1992). Therefore, we assign the two lowest frequency modes at 97 cm^{-1} and 71 cm^{-1} to antisymmetric hydrogen bond stretching.

Our low frequency measurements do not possess the sensitivity of previous FTIR studies due to the lack of cryogenic detection and it was not possible to carry out THz-TDS measurements on these molecules at present. Nevertheless, we were able to detect a number of low frequency vibrations with the lowest detected at 28 cm^{-1}, which is good agreement with previous terahertz measurements of thymine dry films (Laman *et al.,* 2008). In the same study, they found that if the concentration of the thymine solution was lowered when producing the films, this feature disappears, which would suggest that this vibration is due to an external mode. In fact, *ab-initio* calculations would suggest that all low frequency (<100 cm^{-1}) modes in the thymine crystal will be external in nature (Jepsen *et al.,* 2007). This is in agreement with THz-TDS measurements and DFT calculations carried out by Fischer *et al.* (Fischer *et al.,* 2002) that assigned all the low frequency thymine modes to hydrogen bond vibrations.

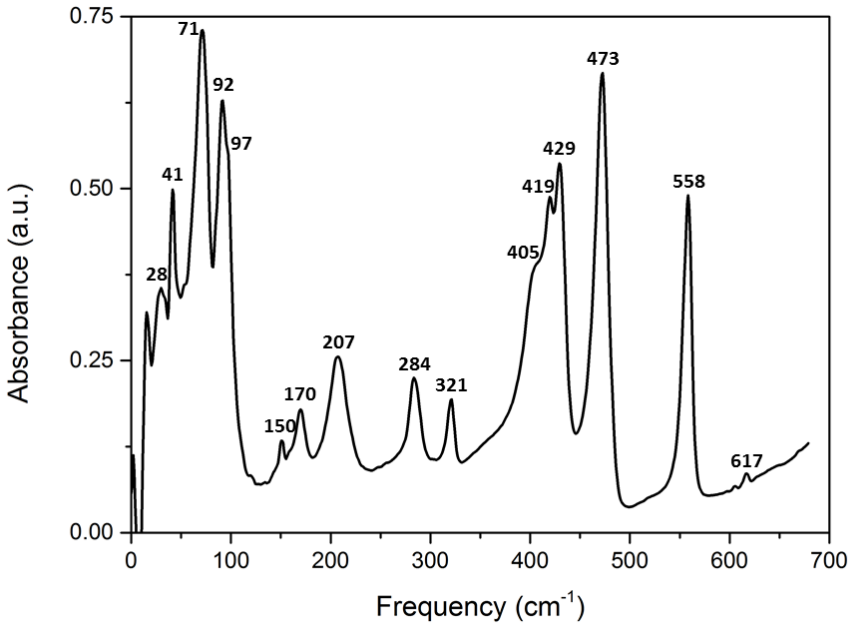


Figure 6.31: Far infrared spectra and absorbance frequencies of Thymine.

6.4.1.1.2 Cytosine

The cytosine FIR spectrum (Fig. 6.32) is considerably different from that of thymine. Although both nucleobases share the common pyrimidine ring, their side groups vary, and previous studies on thymine isotopomers have shown that small substitutions on the side groups can result in large changes in the FIR spectrum, especially in the very low frequency region, where hydrogen bond character heavily influences the spectrum (Zhang *et al.,* 1998). The highest frequency mode in our cytosine spectrum is detected at 597 cm^{-1}. Studies carried out on cytosine at 10 K have identified this peak at 613 cm^{-1} and shown considerable narrowing of the band upon cooling. In addition, this vibration, along with the peak at 531 cm^{-1}, are shown to red-shift considerably in deuterated cytosine, which makes us consider these bands as proton modes of either the imine or amine groups (Rozenberg *et al.,* 2004). Rozenberg and co-workers suggest that these modes are influenced by two different H-bonded protons of the amine group, whereas others merely assign the 531 cm^{-1} mode to NH_2 torsion and wagging modes, irrespective of H-bonding (Szczesniak *et al.,* 1988; Susi *et al.,* 1973).

Little research has been carried out on the low frequency modes of cytosine, however, our lowest frequency vibration detected at 51 cm^{-1} is in good agreement with previous FTIR (Beetz *et al.,* 1980) and THz-TDS studies (Fischer *et al.,* 2002) as are our other low frequency vibrations. From the previous theoretical calculations carried out on thymine, we would assume the modes in this low frequency region also stem from external vibrations.

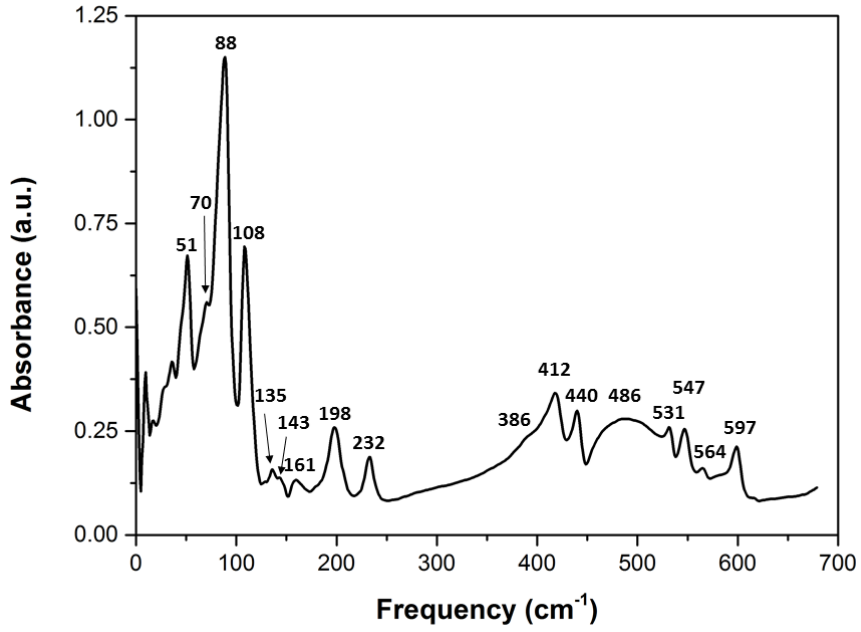


Figure 6.32: Far infrared spectra and absorbance frequencies of Cytosine.

6.4.1.2 Purines

6.4.1.2.1 Guanine

The guanine mode detected at 644 cm^{-1} is well documented and is ascribed to the in-plane stretching of the purine ring *i.e,* breathing mode (Delabar *et al.,* 1978; Majoube *et al.,* 1984; Shanmugasundaram *et al.,* 2009) and can be used as an indicator of DNA conformation by distinguishing between the B and Z forms (Lopez *et al.,* 2012). The next peak detected at 602 cm^{-1} in our spectrum (Fig. 6.33) is commonly assigned to NH_2 wagging motions (Florian *et al.,* 1993; Giese *et al.,* 2002; Giese *et al.,* 2002) however, recent normal coordinate analysis has assigned this vibration to the deformation of the imidazole ring and also assigned peaks at 501 cm^{-1} and 514 cm^{-1} to deformations of the pyrimidine ring (Lopes at el*,* 2012). From the same work, we assign our 241 cm^{-1} band to out-of-plane deformation of the pyrimidine ring, while at 159 cm^{-1}, the vibration stems from skeletal ring torsions. The mode detected at 397

cm^{-1} matches well with previous Raman measurements and is assigned to the synchronised deformation of the amine and carbonyl groups influenced by the hydrogen bonding in the crystal (Lopez *et al.,* 2012).

There has been a lack of measurements performed on guanine in the low frequency range but once again our FTIR measurements are in good agreement with Nishizawa *et al.* (2005), although we detected an additional band at 135 cm^{-1} that has previously not been recorded.

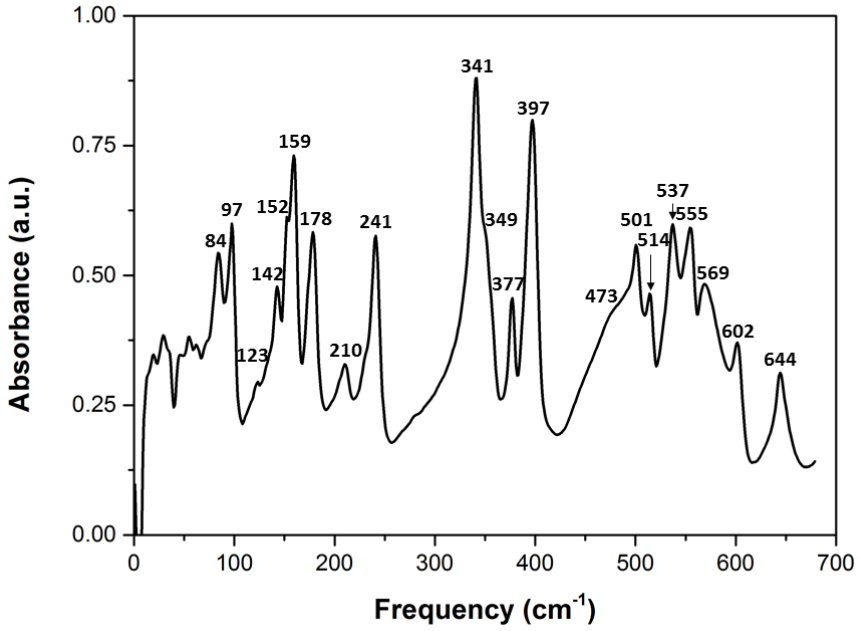


Figure 6.33: Far infrared spectra and absorbance frequencies of Guanine.

6.4.1.2.2 Adenine

The far-infrared spectrum of adenine is shown in Figure 6.34. The high frequency peak detected at 636 cm^{-1} is assigned to out-of-plane deformations of the imidazole ring and the peak below this at 617 cm^{-1} is ascribed to the in-plane deformation of the pyrimidine ring (Lopez *et al.,* 2013). The symmetric stretching of the Pyr/Im atoms, or breathing modes, are predicted to occur at higher frequencies than for guanine

(Lopez *et al.,* 2013) and cannot be seen in our measurements. Previous studies have detected two strong bands at 536 cm^{-1} and 552 cm^{-1} using inelastic neutron scattering, which were correlated with an infrared band centred at 542 cm^{-1}. With our far infrared ATR measurements we were able to detect this band at 541 cm^{-1} and also an additional weak shoulder, previously not detected, at 529 cm^{-1} that correlates well with DFT calculations that predict two band at 538 cm^{-1} and 527 cm^{-1} assigned to skeletal deformations involving the pyr ring (Lopes *et al.,* 2013). The band at 336 cm^{-1} has previously been detected in Raman measurements at 329 cm^{-1} and assigned to 'puckering' of the Pyr ring (Mohamed *et al.,* 2009). However, Lopez *et al*., (2013) assign this mode to the 'H-bond effect' on the displacement of the amine group, similar to the assignment they made of the guanine mode in their earlier work.

In the low frequency range, it is predicted that a butterfly-like motion of the Pyr and Im rings should be detected in the range between 229 cm^{-1} and 249 cm^{-1} (Lopez *et al.,* 2013). This correlates well with previous studies that detected a Raman active mode at 249 cm^{-1}, which they also assigned to this butterfly-like motion (Troullier *et al.,* 1993) and corresponds to the vibration we detected at 247 cm^{-1}. However, other groups have calculated this mode to occur at 200 cm^{-1} and assigned it to an experimental Raman band at 193 cm^{-1}, which could possibly correlate to our peak at 186 cm^{-1}.

Our low frequency adenine spectrum matches well with that of previous THz-TDS measurements (Fischer *et al.,* 2002) but other studies have identified additional bands at 85 cm^{-1}, 115 cm^{-1} and 130 cm^{-1} that we did not detect (Nishizawa *et al.,* 2005). Further studies have been carried out by Shen and co-workers on the temperature dependence of adenine and purines in the low frequency range and found that the vibrational modes shifted to higher frequencies and intensified as temperature was decreased. They performed molecular modelling based on DFT calculations on monomers, dimers and tetramers of adenine and found that all low frequency modes are intermolecular in nature and suggest that these vibrational modes are non-localised

but collective in nature *i.e,* phonon modes. In addition, the authors also suggest that the anharmonic components of the potentials may contribute to the temperature-dependent behaviour of these samples.

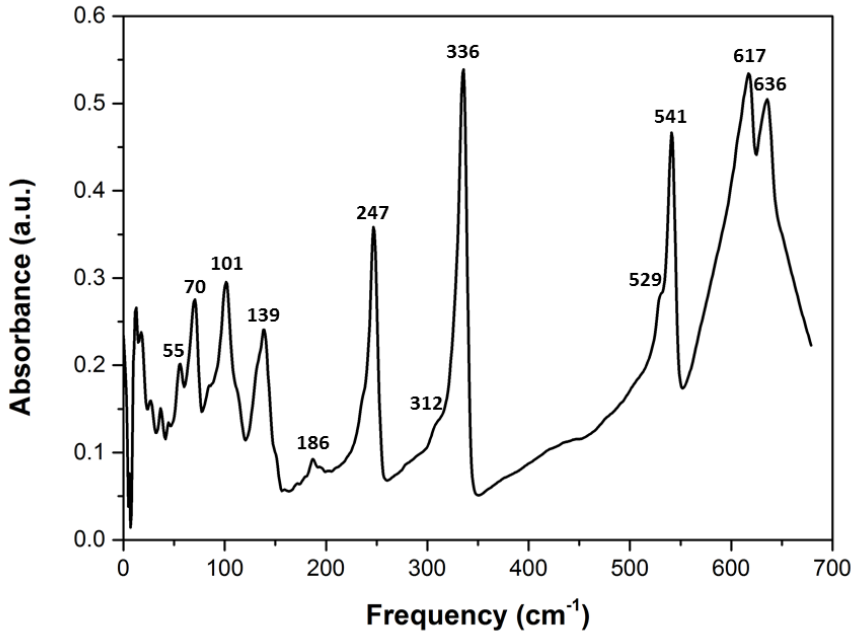


Figure 6.34: Far infrared spectra and absorbance frequencies of Adenine.

6.4.2 Nucleosides & Nucleotides

Nucleosides consist of a 5-carbon sugar ring, either ribose or deoxyribose, joined to one of the five nucleobases at the N1 position, whereas nucleotides share the same structure but also have one or more additional phosphate groups attached to the 5-carbon of the sugar ring (Fig. 6.35). The measurements presented in this section show the FIR spectra of nucleosides, with a ribose sugar, and nucleotides with a single phosphate attached *i.e,* nucleoside monophosphates (NMP's).

Although the phosphate group is small, consisting of only four atoms, the addition of this molecule makes a considerable difference to the FIR spectrum as can be seen in the nucleoside and nucleotide data. The nucleoside spectra display multiple peaks

with a distinctive spectral signature for each, whereas all of the nucleotides display a similar spectral profile dominated by a broad band centred at ~500 cm^{-1} and a steep increase in absorbance below ~200 cm^{-1}. These differences have previously been detected in terahertz measurements of nucleosides and NTPs (Nishizawa *et al.,* 2005). The authors propose that the changes detected between the spectra is due to the larger NTPs being more anisotropic than the nucleosides, making the forces between each molecules easier to distort, which results in the absorption bands broadening or splitting into weaker peaks.

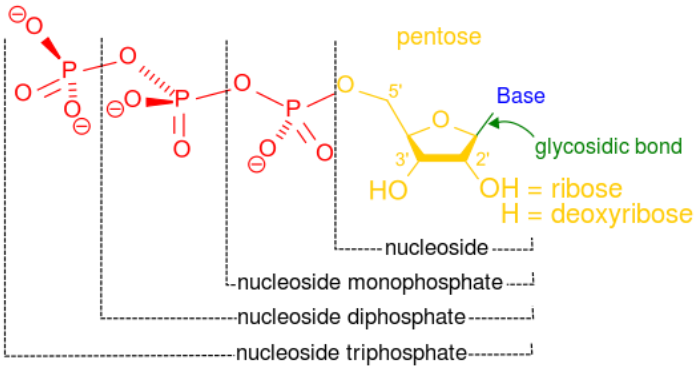


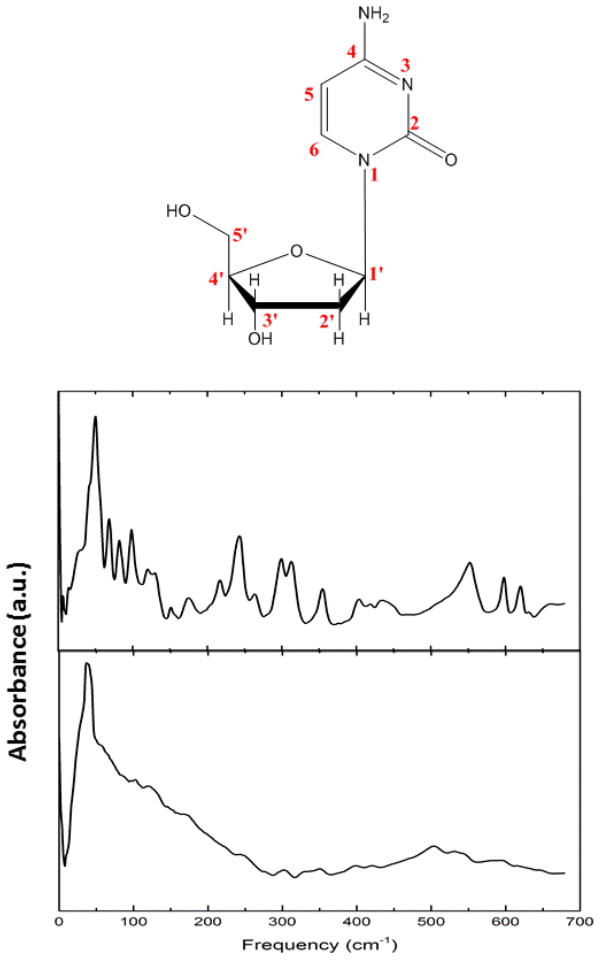
Figure 6.35: Structure of Nucleosides and Nucleotides. *Any nucleoside with one or more phosphates attached is termed as a nucleotide.*

6.4.2.1 Cytidine & Cytidine Monophosphate

The additional phosphate group added to the cytidine molecule makes analysis of the CMP spectrum difficult as the peaks are poorly resolved due to spectral overlap. However, we can identify some common features in both spectra (Fig. 6.36). For example, the highest frequency mode detected in our CMP spectrum occurs at 598 cm^{-1} and matches perfectly with a mode observed in our cytidine spectrum that has previously been assigned to ring deformations (Tsuboi *et al.,* 1987; Thomas *et al.,* 1993); however, other calculations of the cytidine molecule propose this mode to originate from the COO- bend (Sánchez-Cortés *et al.,* 1991). Furthermore, we have previously assigned a mode at this frequency to amine vibrations for the cytosine

nucleobase, which make the precise assignment of this vibration difficult. Another mode that is shared between CMP and cytidine is detected at 398 cm^{-1} and 403 cm^{-1}, respectively, and is assigned to C4N4 in-plane bending (Thomas *et al.,* 1993). Conversely, some of the modes detected in our cytidine spectrum, such as 553 cm^{-1}, 312 cm^{-1} and 264 cm^{-1} are not detected for CMP, which is surprising as these modes have previously been assigned to motions within the cytosine base, which should not be affected by the additional phosphate in CMP (Gavira *et al.,* 1997; Nishimura *et al.,* 1986). However, vibrations detected in the same frequency regions for deoxycytidine (which differs from cytidine by the removal of a single oxygen atom) have been assigned to vibrations of the ribose sugar, which is more likely to be influenced by the phosphate due to the direct ester bond. (Leulliot *et al.,* 1999; Nishimura *et al.,* 1986)

Our low frequency measurements match reasonably well with previous infrared and Raman measurements taken of crystalline cytidine, yet the authors merely assign everything below 200 cm^{-1} as either librational or translational motions (Lee *et al.,* 2001). However, earlier studies carried out on deoxycytidine would suggest that our peaks detected at 130 cm^{-1} and 121 cm^{-1} may be caused by the torsion of N1-C1' (Gaigeot *et al.,* 2000) and C4'C5', (Leulliot *et al.,* 1999) respectively. At the lowest frequencies detected, most studies suggest that these frequencies are librational in nature or influence by hydrogen bonding. However, DFT calculations carried out by Shishkin *et al.,* 2000 suggest that these modes represent the motion of the base unit and sugar units with respect to each other as rigid subsytems (Shishkin *et al.,* 2000).



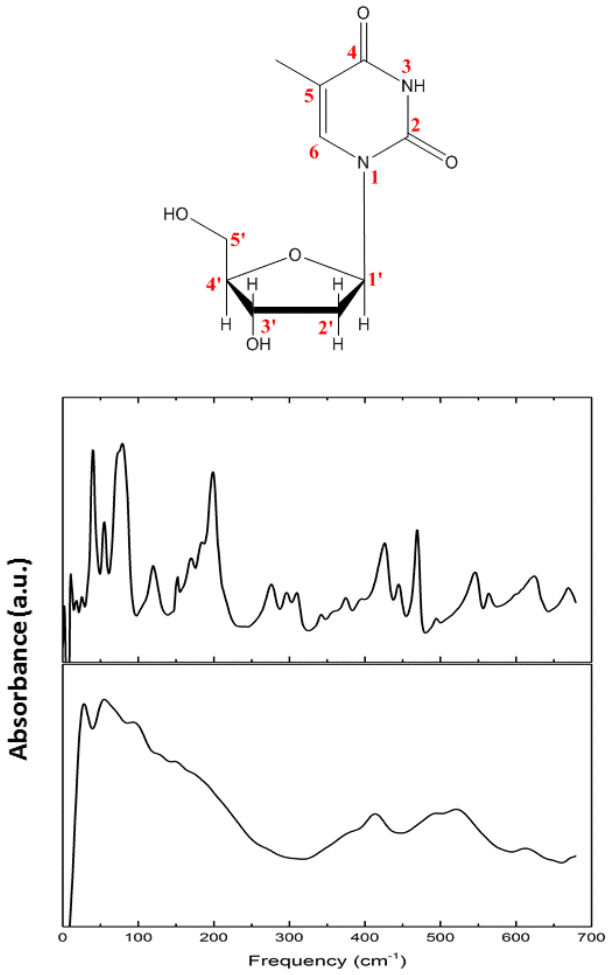
Cytidine	CMP
620	-
598	598
553	545
434	532
-	505
419	418
403	398
355	350
312	326
299	302
264	-
243	246
217	-
175	170
152	-
130	-
121	120
98	-
82	-
68	-
50	41
30	-

Figure 6.36: Far infrared spectra and absorbance frequencies (cm^{-1}) of Cytidine and CMP. *The upper spectrum shows cytidine and the lower shows CMP. The structure of cytidine and numbering of the atoms is also shown.*

6.4.2.2 Thymidine & Thymidine Monophosphate

The thymidine and TMP molecules have been somewhat neglected, and limited information exists in the literature on the low frequency vibrations of these molecules. However, Nishizawa *et al* (2005) have measured the low frequency spectrum of thymidine between 0.4-5.8 THz and identified a number of well resolved peaks. Our lowest frequency peaks detected at 40 cm^{-1}, 54 cm^{-1}, 79 cm^{-1} and 120 cm^{-1} (Fig. 6.37)

agree extremely well with the data collected by Nishizawa and co-workers, however they also detected additional low frequency modes at 44 cm^{-1}, 47 cm^{-1} and 56 cm^{-1} that we could not. Because of the close proximity of these modes, the resolution of our FTIR measurements is not adequate to distinguish between these peaks.



Thymidine	TMP
669	-
623	611
564	-
546	521
495	495
469	-
445	-
426	414
396	-
374	382
359	-
342	-
310	-
296	-
276	-
199	-
184	-
170	-
152	-
120	102
79	62
54	-
40	-

Figure 6.37: Far infrared spectra and absorbance frequencies (cm^{-1}) of Thymidine and TMP. *The upper spectrum shows thymidine and the lower shows TMP. The structure of cytidine and numbering of the atoms is also shown.*

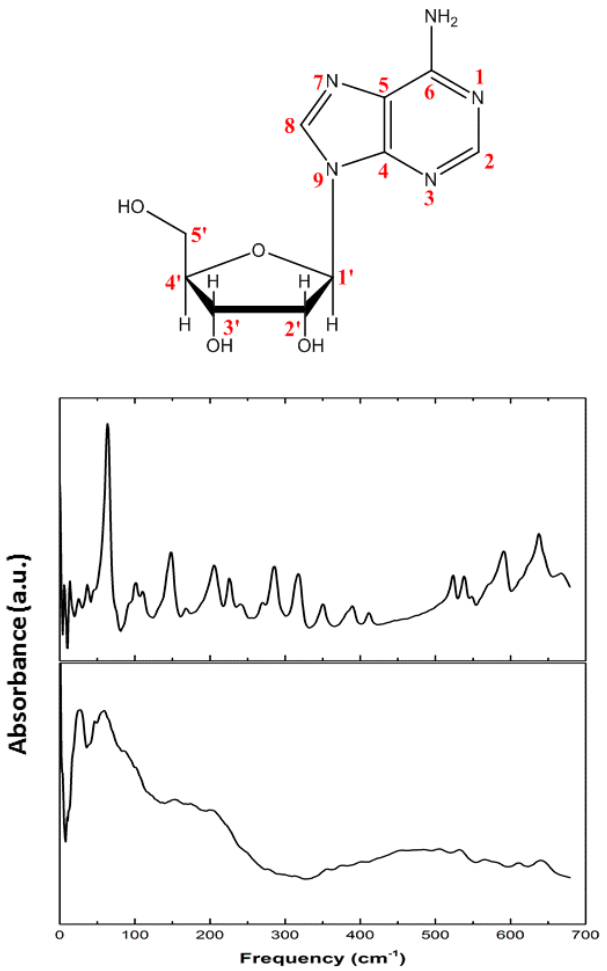
6.4.2.3 Adenosine & Adenosine Monophosphate

The far infrared spectra of adenosine and AMP are shown in Figure 6.38. The highest frequency peak detected for AMP at 640 cm^{-1} appears to have the same origin as the adenosine mode detected at 637 cm^{-1} and has been assigned to coupled deformations of C'C'O' + C'C'C' (Carmona *et al.,* 1990) in the ribose sugar. However, based on our previous assignment of adenine we assign this mode to deformations of the imidazole ring (Lopez *et al.,* 2013). Additionally, the AMP bands at 565 cm^{-1}, 534 cm^{-1} and 208 cm^{-1} may correspond to the 570 cm^{-1}, 538 cm^{-1} and 205 cm^{-1} peaks seen in adenosine. It has been suggested that the 538 cm^{-1} adenosine mode is again due to the ribose ring symmetric bend (Carmona *et al.,* 1990), yet Lord *et al,* (1967) assign this band to adenine skeletal ring deformation, which agrees well with our previous assignment of this mode for the adenine nucleobase. Immediately adjacent to this peak, we detected another adenosine vibration at 524 cm^{-1} that is absent from the AMP spectrum, but is probably masked by the very broad, almost featureless, band we detect in this region. Based on DFT calculations of the adenine nucleobase we assign this peak to skeletal deformations involving the pyrimidine ring (Lopes *et al.,* 2013)

In the 200 cm^{-1}-400 cm^{-1} region, we see a dense collection of peaks in the adenosine spectrum that was not previously detected in our adenine nucleobase data, which we therefore assume to stem from vibrations involving the ribose sugar. This was confirmed by studies carried out on deoxyadenosine and we assign the bands detected at 350 cm^{-1}, 285 cm^{-1} and 241 cm^{-1} to C1'–C2'–O2' in-plane bending, C1'–N9–C4–C8 out-of-plane wagging and again C1'–N9–C4–C8 out-of-plane wagging, respectively (Bailey*,* 1997)

Our adenosine data below 200 cm^{-1} is in very good agreement with previous low frequency measurements and the bands at 110 cm^{-1}, 101 cm^{-1}, 92 cm^{-1} and 64 cm^{-1} have previously been detected by a variety of techniques (Nishizawa *et al.,* 2005; Lee *et al.,* 2000, laman *et al.,* 2008) and are again assigned to lattice vibrations (Lee *et al.,* 2000). The lowest frequency mode is detected at 25 cm^{-1} and 26 cm^{-1} for adenosine

and AMP, respectively, and in this low frequency region the signal for AMP increases significantly, as is seen for all NTP's. Calculations on deoxyadenosine have predicted that this mode may be due to the rotation around the C1'-N9 glycosidic bond (Shishkin *et al.,* 2000), connecting the nucleobase and ribose sugar.

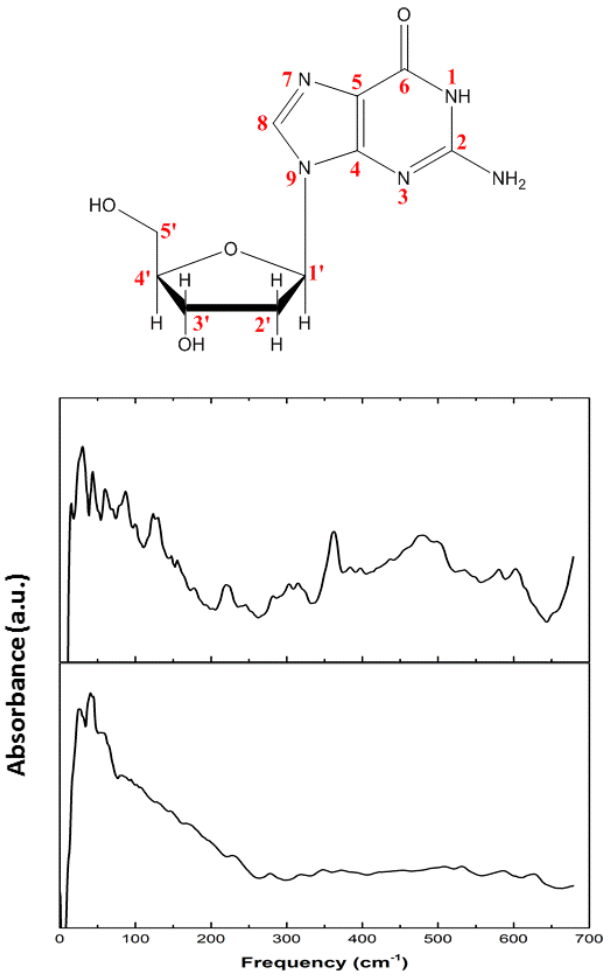


Adenosine	AMP
667	-
637	640
591	582
570	565
547	-
538	534
524	-
-	450
412	-
390	-
350	-
317	-
285	-
270	-
241	-
226	-
205	208
168	-
149	-
110	-
101	-
92	-
64	59
45	-
36	-
25	26

Figure 6.38: Far infrared spectra and absorbance frequencies (cm^{-1}) of adenosine and AMP. *The upper spectrum shows adenosine and the lower shows AMP. The structure of cytidine and numbering of the atoms is also shown.*

6.4.2.4 Guanosine & Guanosine Monophosphate

To date, limited studies have been carried out on the guanosine or GMP molecules. Below 200 cm^{-1} we detected a total of five vibration modes for both guanosine and GMP, yet there appears to be little correlation between frequencies (Fig 6.39). Early work by Urabe *et al.* (1987) identified the lowest frequency vibration in guanosine measurements at 20 cm^{-1} and hypothesised that this mode originated from base stacking interactions. This so-called 'S-Mode' has also been identified in DNA polymers and was previously been linked with motions of the DNA backbone (Urabe *et al.,* 1982; Urabe *et al.,* 1985) and water content (Demarco *et al.,* 1985). We did not detect this mode in our current measurements and, similarly, Nishizawa and co-workers were also unable to detect the S-mode utilising a gallium phosphide (GaP) terahertz wave generator. However, our guanosine measurements showed decidedly more inconsistencies in the low frequency region than our other nucleoside measurements and upon comparing our peak frequencies with those in the literature (Nishizawa *et al.,* 2005), we confirm the discrepancies in our data and therefore, cannot be confident in these particular assignments.



Guanosine	GMP
603	628
581	587
535	532
499	508
480	-
437	-
397	-
384	-
363	347
315	316
303	-
283	278
246	-
219	230
177	170
151	-
127	-
88	81
59	55

Figure 6.39: Far infrared spectra and absorbance frequencies (cm^{-1}) of guanosine and GMP. *The upper spectrum shows guanosine and the lower shows GMP. The structure of cytidine and numbering of the atoms is also shown.*

6.4.3 Deoxynucleotide Triphosphates

The measurements in this section have been performed on deoxynucleotide triphosphates (dNTP). The structures of these molecules is similar to those of the NMP molecules except an oxygen has been removed at the 2'-carbon of the sugar ring and an additional two phosphate molecules have been added. dNTP's are the precursors that create the DNA molecule; two of the phosphate molecules are cleaved to a monophosphate, and pyrophosphate eliminated, before the molecule is incorporated into the DNA chain via a phosphodiester bond.

All the measured dNTP's show a similar spectral profile dominated by two broad bands centred at ~475 cm^{-1} and ~180 cm^{-1} (Fig. 6.40). This spectral shape is more analogous to the far infrared spectrum obtained for long chain DNA molecules (Wittlin *et al.,* 1986; Woolard *et al.* 2002) than for the individual nucleic acid components we have previously measured. The same spectral shape was also seen in early studies performed on RNA samples where they assigned all modes above the ~300 cm^{-1} absorbance minima to ring vibrations within the subunits (Beets *et al.,* 1982). For the low frequency half of the spectrum, the signal is likely to come from weak interactions such as hydrogen bonding or possibly stacking of the nucleoside molecules (Woolard *et al.,* 2002).

As the dNTP molecules are considerably larger than the other nucleic acid components we have measured, due to the addition of the three phosphate groups, there will be considerably more normal modes of vibration governed by $3N$-6, where N is the number of atoms in the molecule, and we therefore expect some spectral overlap of peaks. Moreover, the additional phosphate groups increases the anisotropy within the molecule, which can reduce the peaks in the spectrum as we have seen for nucleosides and nucleotides. However, although the spectra of the dNTP molecules appear similar, each possess a number of individual peaks that gives each molecule its own spectral 'fingerprint' in the far infrared region. All of these distinct vibrations lie on the high frequency side of the ~300 cm^{-1} absorbance minima, which would

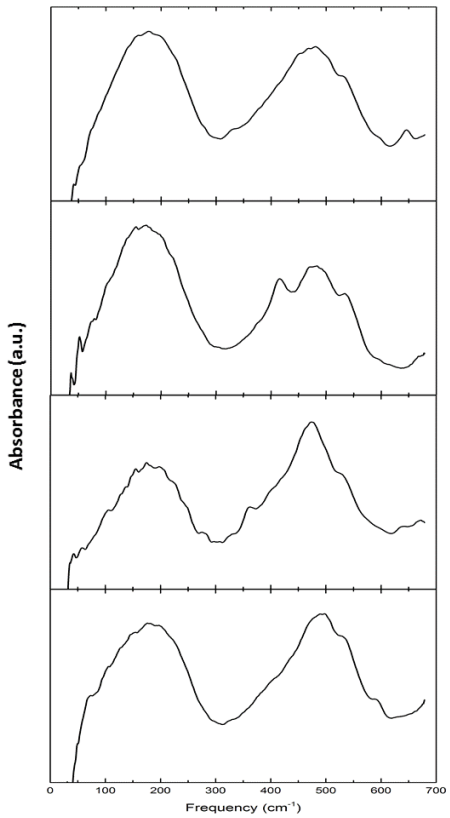
suggest that these modes all stem from intramolecular vibrations dependent on the nucleobase structure, as the rest of the molecule is identical for each sample.

Beginning in the high frequency range of the spectrum we detect a well resolved peak at 646 cm^{-1} for dATP that has previously been detected in our adenine nucleobase, -side and –tide spectra, which we assign to deformations of the imidazole ring in the adenine base. Additionally, dGTP, the other purine base, also displays a high frequency peak in the same region at 637 cm^{-1}, whereas the pyrimidines lack this feature. We can therefore be confident in our assignment of this mode to the imidazole ring. The adjacent weak band detected in dATP at 596 cm^{-1} is also present in our dCTP spectrum, as a much stronger peak at 593 cm^{-1} and is assigned to vibrations of the amine group based on previous cytosine assignments (Rozenberg *et al.,* 2004).

A common feature detected in all dNTP's is detected at 531 cm^{-1}, 532 cm^{-1}, 530 cm^{-1} and 533 cm^{-1} for dATP dTTP, dGTP and dCTP, respectively. Our previous measurements on simpler nucleic acid components have assigned this spectral region to either vibrations of the amine (Rozenberg *et al.,* 2004; Szczesniak *et al.,* 1988; Susi *et al.,* 1973) or the pyrimidine ring (Lopes *et al.,* 2013). However, as dTTP does not contain an amine group, but the peak is still present, this mode must originate from the pyrimidine ring.

The dTTP sample displays a strong feature at 417 cm^{-1} that is specific to this molecules and not seen in the other dNTP spectra. However, modes in this frequency region for thymine molecules have previous been assigned to either a CO deformation or ring torsions (Zhang *et al.,* 1998) and it is therefore unclear why this mode is only detected for dTTP and not the other dNTPs.

A lower frequency mode detected in dATP at 328 cm^{-1}, probably has a similar origin to the dGTP vibration detected at 370 cm^{-1}, yet we cannot make a definitive assignment of this mode due to contradicting predictions in the literature (Bailey *et al.,* 1997; Lopez *et al.,* 2012; Mohamed *et al.,* 2009).



dATP	dTTP	dGTP	dCTP
646	-	637	-
596	-	-	593
531	532	530	533
480	478	473	493
-	417	-	-
328	-	370	-
177	163	180	183

Figure 6.40: Far infrared spectra and absorbance frequencies of glycine dNTP's. *The spectra from top to bottom show the dATP, dTTP, aGTP and dCTP, respectively.*

6.4.4 Deoxyribonucleic Acid (DNA)

6.4.4.1 Analysis of DNA Species

The spectrum of long chain DNA is similar to the spectra we obtained for dNTP's and upon analysis we see that specific nucleotides independently contribute to the overall DNA spectrum. Salmon sperm (SS) DNA is commonly used in biochemical protocols, such as Southern blotting, due to its easy availability and low cost as a by-product of aquaculture (Emanuel & Chaikoff, 1953; Sambrook & Russell, 2001). For the same reasons, it is commonly used as a standard for spectroscopic investigation of long chain DNA molecules, which is why we choose to use this DNA as a model molecule in our measurements. In addition, we have also analysed a plasmid DNA as a

comparison. Plasmids are double stranded, short, circular DNA that are commonly found in bacterial cells, but may also be present in some archaea and eukaryotes. Plasmids replicate independently of chromosomal DNA and frequently carry additional genes that may aid in survival (*i.e.*, antibiotic resistance in bacteria) and are commonly used as cloning vectors for recombinant DNA in biochemistry labs and the pharmaceutical industry (Thomas & Summers, 2008). The difference in structure between SS and plasmid DNA makes these macromolecules an ideal comparison to elucidate if DNA structural changes can be detected at a macroscopic scale in the FIR. For this study, we choose to analyse the pUC19 plasmid cloning vector due to its availability, high-copy number and small size (2686 b.p).

Figure 6.41 shows the spectrum of SS and plasmid DNA in the 30-680 cm^{-1} range where both DNA molecules show identical spectra. The overall intensity of plasmid DNA is lower than SS, which can simply be explained by the different concentrations used – 2 mg/ml and 10 mg/ml, respectively – however, the frequency of all spectral features are in good agreement.

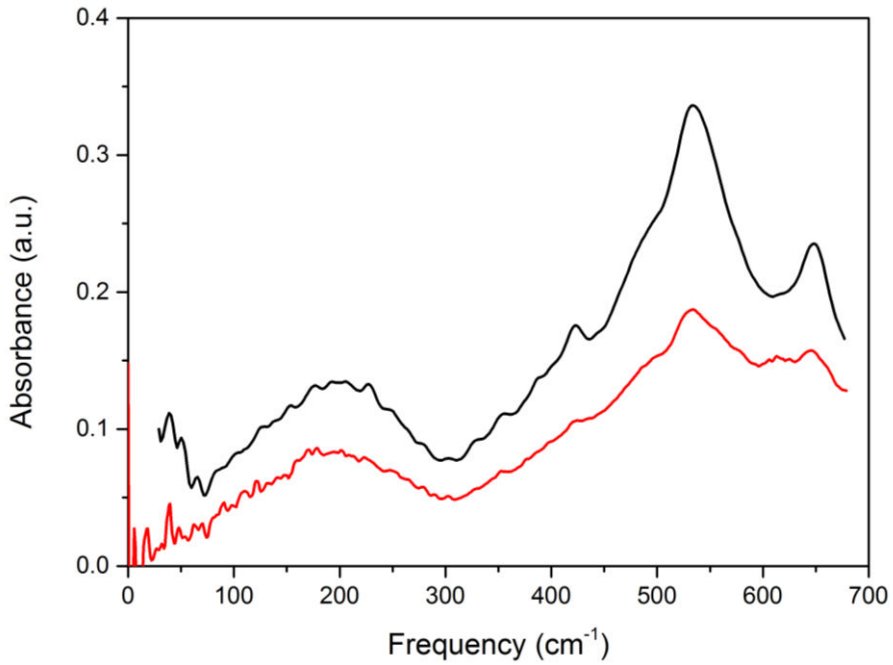


Figure 6.41: Far Infrared spectrum of Salmon Sperm DNA (Black) and Plasmid DNA (Red)

The spectra are dominated by two broad bands centred at 535 cm^{-1} and 200 cm^{-1}. In addition to these broad bands, we also detect smaller peaks at 648 cm^{-1}, 422 cm^{-1} and weak shoulder at 495 cm^{-1}, which is in good agreement with previous far infrared measurements performed on macromolecular DNA (Wittlin *et al.,* 1986; Woolard *et al.,* 2002). The highest frequency peak detected at 648 cm^{-1} is in near-perfect agreement with the purine vibrational mode detected for dATP and dGTP, which demonstrates that this peak in the DNA spectrum is specific to these bases. Furthermore, the 422 cm^{-1} DNA peak is also detected in dTTP, but no other dNTPS, making this nucleotide exclusively responsible for this mode in DNA molecules.

The largest peak detected in the SS DNA spectrum at 535 cm^{-1} is also detected in all dNTP's as a vibration of the pyrimidine ring. It is logical to assume that this DNA mode is from the same origin and the largest contribution to the DNA spectrum stems from the pyrimidine ring that is common to all nucleotides. However, the strongest peak detected in our dNTP measurements at ~475 cm^{-1} is only displayed as a weak

shoulder at 495 cm^{-1} in our DNA spectrum and must therefore originate from the triphosphate group. In the DNA chain, this group is cleaved to a monophosphate before forming the phosphodiesterase backbone of the DNA molecule, which explains the stark reduction of this band. Surrounding this feature are also a number of very weak oscillations that are thought to be soft phonon modes that are masked by the strong absorption from the 535 cm^{-1} band (Globus *et al.,* 2002). These phonon modes were found to be consistent between DNA species and therefore proven not to be artefacts. However, in accordance with our own data, it also demonstrates that this frequency region cannot be used to differentiate between different DNA molecules. However, we have revealed that certain peaks correspond to individual nucleotides and it may therefore be possible to determine a DNA chain's composition based on the intensity of these peaks. However, further research will need to be carried out in order to confirm this hypothesis.

Below the 300 cm^{-1} absorbance minima, the broad band is thought consist of a collection of intermolecular vibrations between base pairs and also phonon modes along the DNA backbone (Wittlin *et al.,* 1986; Woolard *et al.,* 2002). This broad, low intensity feature is to be expected in our relatively low resolution measurements as previous studies have shown that these modes produce fine spectral features in the very far infrared and require a resolution better that 0.5 cm^{-1}, which usually requires cryogenic detection (Globus *et al.,* 2002). Furthermore, the DNA chain represents an oscillator that is not periodic in structure. Hence, any oscillation contributing to an individual frequency is limited and detection of these modes by resonant absorption is therefore a weak process.

6.4.4.2 Analysis of DNA Hybridisation

To investigate if DNA hybridisation can be detected at terahertz and FIR frequencies we used a combination of THz-TDS and FTIR to scan the whole frequency range. We measured the FTIR spectrum of SS DNA in its hybridised and denatured states over the range 30 cm^{-1} – 680 cm^{-1} (1 – 20 THz) but were unable to identify any differences

between the two states (data not shown). However, this is not surprising, as we have seen from our analysis of the SS DNA spectrum, this range in mainly sensitive to intramolecular modes of the DNA nucleotides and would therefore not be effected by hybridisation state. Furthermore, previous studies have only identified differences between single and double stranded DNA below 2 THz (Bolivar *et al.,* 2002), with most being performed below 1 THz, (Nagel *et al.,* 2002a, 2002b, 2003; Globus *et al.,* 2004; Parthasarathy *et al.,* 2005), which is in the range where our FTIR measurements are unreliable. Therefore, we carried out THz-TDS measurements utilising our novel membrane method to determine if hybridisation could be detected at the very low frequencies. Figure 6.42 shows our raw THz-TDS data on hybridised and denatured SS DNA where there is a clear increase in absorbance given from the denatured sample above approximately 1 THz. Similar to our results, previous measurements have detected an increased absorption for denatured DNA compared to hybridised in the terahertz range and attributed this to the increased flexibility of the molecule, allowing for more degrees of freedom. Yet in those measurements the difference was also visible <1 THz (Globus *et al.,* 2004; Parthasarathy *et al.,* 2005), whereas this is not the case in our data. Furthermore, the authors claimed to identify a number of peaks that could distinguish between the hybridisation states, while we detect a reasonably featureless spectrum, with only three negative peaks detected at 1.11 THz, 1.16 THz and 1.87 THz for both states, which may be due to residual water absorption.

Due to the inconsistencies of our data with the literature, we cannot be confident that the changes we have detected are reliable and originating from the DNA molecules. Furthermore, the absorbance of our hybridised sample remains close to zero across the whole spectrum, which would suggest no terahertz attenuation. However, it is possible that the denatured sample holds more water molecules and the increased absorbance we see above 1 THz is due to water absorption, which could be an indirect way of measuring hybridisation state

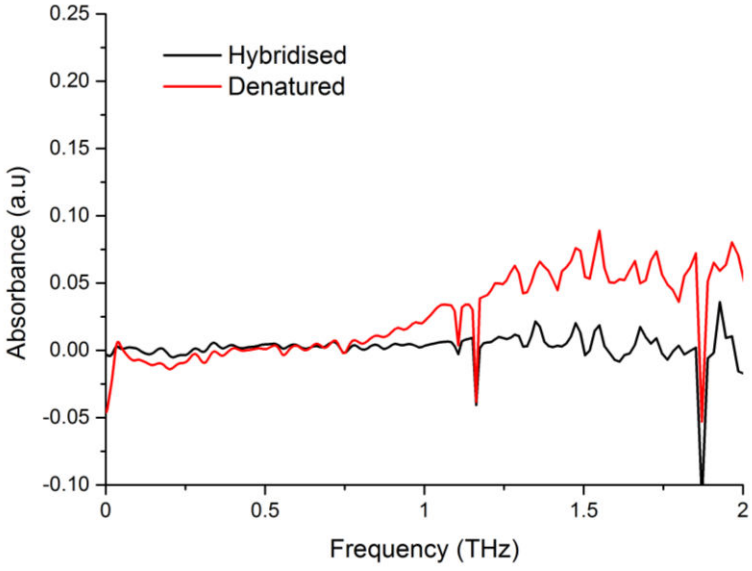


Figure 6.42: THz-TDS spectra of Hybridised (Black) and Denatured (Red) Salmon Sperm DNA. *For these measurements, 500 µl of 20 mg/ml DNA solutions were pipetted into a 20 mm diameter well and allowed to dry in a desiccated atmosphere overnight.*

6.5 Conclusions

We have developed a holder and membrane method for terahertz measurements that allows quick and simple analysis of biomolecules with minimum preparation and shown that the data is comparable to standard techniques. This method would allow rapid analysis of molecules under different conditions (*i.e,* amino acids zwitterions), which would otherwise require time consuming sample preparation. However, due to complications with our THz-TDS system, we were unable to continue with this work.

For the first time we have obtained the far infrared spectrum for all α-amino acids in their different protonated states. The far infrared spectrum of amino acids at their isoelectric point can generally be split into 5 sections (Fig. 6.43) with the high frequency portion dominated by CO_2 vibrations (500-700 cm^{-1}) and NH_2 modes (400-500 cm^{-1}). In the 300-400 cm^{-1} range, deformations around the α–carbon are a major contributor to the spectrum while between 200-300 cm^{-1} the origin of the modes varies and features can stem from the side group or CO_2 torsions. In the lowest frequency range the lattice modes and hydrogen bonding dominate the spectrum.

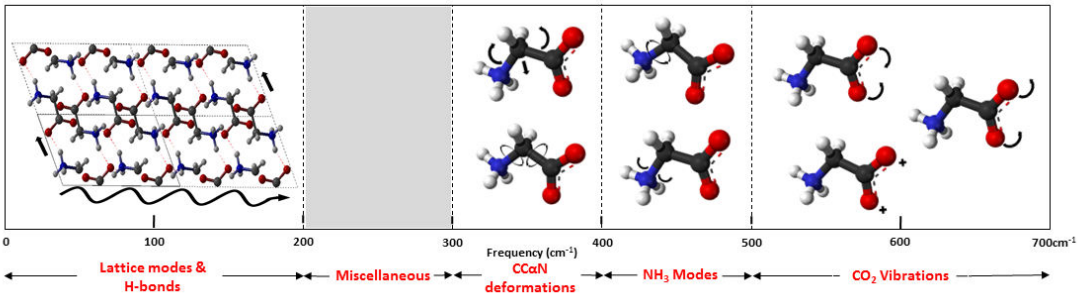


Figure 6.43: Characteristic regions of the amino acid far infrared spectrum.

We have also obtained the spectra of the amino acid zwitterions and shown considerable differences, dependent on the crystallisation of the molecule. Furthermore, the NH_2 mode for some amino acids is shown to red-shift dramatically in its protonated form. Disulphide bonds and short peptides were also analysed in the

far infrared utilising cysteine and glutathione where we were able to identify the conformation of these molecules based on their spectral profile around 500-550 cm^{-1} and also able to identify specific contributions to the glutathione spectrum from individual amino acids.

We have also analysed a variety of DNA and RNA components from the most fundamental level, through varying degrees of complexity, up to full chain DNA molecules. Due to the increased hydrogen bonding and complex crystal structures of nucleic acid components it is difficult to present a simplified profile of the far infrared spectrum for these molecules. However, it is possible to identify common vibrations such as the mode detected near 530 cm^{-1} for all the bases, which stems from vibrations of the pyrimidine ring. Additionally, an imidazole vibration that is specific to the purines occurs at ~650 cm^{-1}. By comparing some of these vibrations to the full length DNA spectrum, we have identified, for the first time, individual contributions specific bases make to the macromolecular DNA spectrum, most notably the 422 cm^{-1} DNA mode that stems solely from the thymine molecules in the chain.

Chapter 7:

Discussion &

Future Work

Chapter 7 Discussion & Future Work

The concept of using terahertz radiation as a means of detection and analysis has shown promise due to the transparency of many materials in this frequency range coupled with the nonionizing nature of this light. These properties have allowed terahertz radiation to be incorporated as an investigative tool in the military as a method of standoff detection (Lee *et al*., 2006) and well as in the security (Folyynowicz *et al.*, 2006) and construction (Abina *et al.,* 2015) industries to identify concealed materials. Terahertz radiation also corresponds to processes occurring on a picosecond timescale, which is the domain in which many biological processes are thought to occur (Falconer & Markelz, 2012; Hay & Scrutton, 2012; Benkovic & Hammes-Schiffer, 2003; Agarwal, 2005; Yang & Bahar, 2005), making it an attractive diagnostic tool for the field of biochemistry. Yet this application of terahertz is still in the early phases and further development is required. In this work, we have investigated this area by utilising THz-TDS to gather initial data to determine if terahertz can be incorporated into existing biochemical techniques, such as Southern and Western blots, in order to simplify detection and analysis. In addition, we have also generated a database of biomolecules at far-infrared and terahertz frequencies as a reference for future research, which we summaries later in this chapter. The major drawback of terahertz radiation in biological measurements is the high attenuation of this frequency range by liquid water and most measurements to date have been made on dry films and crystalline samples (Korter *et al.,* 2006; Laman *et al.,* 2008). The use of crystalline samples is adequate for studies aiming to incorporate terahertz into biochemical detection protocols as these conditions mimic sample preparation of current techniques. However, in order to assign biologically relevant terahertz modes, it is essential to carry out aqueous measurements that reflect the natural environment. OKE spectroscopy is a Raman technique that does not suffer the same water absorption as conventional terahertz spectroscopies and it has previously been

possible to generate high quality spectra of aqueous samples using this method (Giraud *et al.,* 2003; Turton *et al.,* 2014). In this thesis, utilising a combination of OKE, FTIR and THz-TDS, we have presented a comprehensive study of biomolecules in the terahertz range, identifying novel vibrational modes for simple small molecules as well as complex macromolecules, in both the liquid and solid phase.

7.1 Optical Kerr Effect Spectroscopy

7.1.1 Proteins

In agreement with early OKE studies carried out by Giruad *et al* (2003), our measurements on protein structure have revealed that the OKE terahertz spectra of all proteins share a similar spectral profile, irrespective of secondary structure. However, the improved signal to noise ratio and dynamic range of our current OKE system (Turton *et al.,* 2013) coupled with carefully selected model systems allowed us to identify significant changes that reflect proteins structure and chemistry. For example, we have obtained the terahertz spectrum of lysozyme solutions and identified 1.15 THz and 2.80 THz modes corresponding to two underdamped delocalised oscillators that involve the motion of at least a dozen residues. In subsequent ligand-binding measurements, we reveal changes in these modes, which are linked to the binding coordinate and hence of biochemical significance, proving, for the first time, that terahertz frequency vibrational modes mediate efficient protein-ligand binding in solution. For other proteins, such as α-lactalbumin (lalba) we also identify a shift in the ~2.8 THz mode when the protein adopts an open conformation or undergoes denaturation, in agreement with previous assignments of this band to delocalised backbone torsions (Giraud *et al.,* 2003; Hunt *et al*., 2007).

These initial measurements of lalba have encouraged further research and additional temperature dependent data is currently being collected in the hope of fitting parameters to these spectral features. In fact, the work in this thesis has stimulated

additional research into the dynamics of different proteins in the terahertz range and analysis of higher molecular weight proteins is currently on going. As discussed in this thesis (section 3.2.2.3), the hydration band of a protein can be detected below 300 GHz, yet this feature was usually obscured in our data by the fast tumbling (α-relaxation band, occurring at tens of megahertz) of the small proteins studied. By utilising larger proteins, we hope to push the α-relaxation to lower frequencies and fully expose the hydration band, which can be fitted with parameters to obtain information regarding the solvation dynamics of the protein. In addition, our ligand binding and phosphorylation experiments alongside successful two dimensional infrared (2D-IR) measurements on haem proteins (Hunt *et al.,* 2011) have led to oxidation and reduction reactions being studied in the model protein, Cytochrome C.

7.1.2 Nucleic Acids

Since the early work of Wittlin *et al* (1986) and Powel *et al* (1987), researchers have attempted to identify terahertz modes in DNA. Much of this research was aimed at identifying discerning features between hybridised and denatured states (Brucherseifer *et al.,* 2000; Globus *et al.,* 2004), while other studies focused on detecting the elusive 'phonon mode' that has previously been predicted with theoretical models (Van Zandt & Saxena*,* 1992; Woolard *et al.,* 2002). However, the majority of measurements have been performed on DNA dry films and the identification and assignment of low frequency features is lacking. Here, we have carried out one of the first terahertz studies of DNA and its components in solution, and identified a variety of novel vibrations that are linked to structure and intermolecular vibrations. Hydrogen bonding interactions are predicated to be visible in the terahertz range (Zhang *et al.,* 1998; Fischer *et al.,* 2002) and we have identified a novel mode at 0.35 THz in both saturated guanosine monophosphate (GMP) solutions and in double stranded DNA, which we attribute to hydrogen bonding. A novel mode was also identified at 2.9 THz in double stranded DNA samples, which we attribute to a delocalised DNA phonon mode, in good agreement with previous

inelastic x-ray scattering measurements (Krisch *et al.,* 2006). The presence of these modes were confirmed by carrying out hybridisation measurements, where we found both the phonon and hydrogen bonding modes to disappear in single stranded samples. Parameters are currently being fitted to these spectral features in conjunction with molecular dynamics simulations in order to confirm our observations and the origin of these features, while also tracking how these modes differ from a hybridised to denatured state.

This work has resulted in additional nucleic acid experiments being conducted in the Wynne lab with measurements being performed on different length oligonucleotides with varying sequences. With this data, it may be possible to determine if the number of base pairs directly influences the frequency of the phonon mode and the intensity of the hydrogen bonding mode. In addition, we hope to analyse RNA-RNA or DNA-RNA interactions, as these complexes are known to adopt the A-DNA configuration in solution, whereas our current data represents B-DNA (Cross *et al.,* 1997) (See Fig 5.1). Finally, the prospect of studying multi-stranded DNA (triple helix and quadruplexes) is intriguing, to determine how these structures influence the DNA terahertz vibrations.

Hopefully this data can be utilised in combination with other studies to drive further research into terahertz applications for DNA detection. For example, quantitative Polymerase Chain Reaction (qPCR) is a technique commonly used in the biological sciences that can detect the replication of a DNA product in 'real time' but requires non-specific, intercalating fluorescent dyes or sequence specific fluorescent probes (Garibyan & Avashia *et al.,* 2013). It has recently been shown that quantitative DNA measurements can be performed using THz-TDS with the potential of being applied to PCR analysis (Arora *et al.,* 2012) and further advances by Globus and co-workers (2006; 2013) on the analysis of DNA in the liquid phase have also demonstrated the potential of applying terahertz spectroscopy to aqueous DNA samples. In this work, we have identified the frequency of novel modes that determine between DNA in its

hybridised and denatured states. The next stage would be to focus on these modes and determine if quantitative data can be extrapolated from the features, which could potentially aid the development of label free DNA analysis.

7.2 Terahertz Time Domain Spectroscopy & FTIR

Considerable research has been carried out in an attempt to find applications for terahertz radiation in biochemical labs, yet this has been hampered by the high attenuation of terahertz waves by liquid water (Lee, 2009). To overcome this issue, we hypothesised that dry, charged membranes could be utilised to fix biological samples and therefore remove the water effect. The membranes commonly used in biochemical labs are made from polymer materials (Ho *et al.,* 1998), which were found to be highly transparent to terahertz light and we identified nitrocellulose as a suitable material for developing a novel membrane method for analysing small biomolecules. Unfortunately, this work could not be concluded due to complications with our THz-TDS system, yet the initial data looks promising and we were able to identify the low frequency spectral 'fingerprint' of small biomolecules that agrees well with previous measurements on crystalline samples (Wang *et al.,* 2005). Using this technique, it should simplify sample preparation in the terahertz range and allow us to complete the database of amino acid zwitterions by obtaining the very low frequency terahertz spectra.

Early work by Bolivar and co-workers (2004) was one of the first studies to identify the potential of applying terahertz to existing applications, for example, terahertz analysis of microarrays, which was coined the 'THz-Biochip'. Our aim was to progress from these studies and develop our membrane technique in order to incorporate terahertz radiation into existing methods to analyse biological macromolecules. Protein conformational changes (Whitmire *et al.,* 2003; George *et al.,* 2013) and DNA hybridisation state (Brucherseifer *et al.,* 2000; Bolivar *et al.,*

2004) have already been detected in the terahertz range, yet not with this method. Given that the nitrocellulose membrane we utilised for our experiments are already commonly used in biochemistry labs for detecting protein and DNA changes using fluorescent or colorimetric techniques (Ho *et al.,* 1998), applying terahertz spectroscopy to these methods would remove the need for complicated and time consuming detection. Therefore, to proceed with this work, quantitative measurements would need to be carried out in order to determine the detection limit of the system. Thereafter, we would look into integrating terahertz spectroscopy as a new detection method for common biochemical methods such as western or southern blots.

Previous studies have analysed the low frequency vibrations of small biomolecules in various states (Matei *et al.,* 2005; Trivella *et al.,* 2010), yet our FTIR measurements have provided the most extensive database of small biomolecules in the far infrared and terahertz range to date. Many theoretical studies have been performed on small biomolecules, however experimental data is required for meaningful comparison (Fischer *et al.,* 2002; Lopes *et al.,* 2012; 2013). Now high quality experimental spectra have been obtained, we provide a simple reference for future DFT and normal mode calculations, to provide a deeper understanding of the physical properties of these molecules and also an insight into their contribution to the overall protein or DNA spectrum. In fact, in our current measurements we have already identified individual nucleobase modes that contribute to the macromolecular DNA spectrum. By monitoring the intensity of these modes, it may be possible to determine the nucleobase composition, quantitatively, of unknown DNA sequences, which may be useful in identifying isochores – high GC content regions – (Bernardi, 2000) or identifying bacteria based on their nucleobase content (Wayne *et al.,* 1987).

References

Abeyrathne, E. D., Lee, H. Y., & Ahn, D. U. (2013) Egg white proteins and their potential use in food processing or as nutraceutical and pharmaceutical agents-a review. *Poult Sci,* **92**(12), 3292-3299.

Abina, A., Puc, U., Jeglič, A., & Zidanšek, A. (2014) Applications of Terahertz Spectroscopy in the Field of Construction and Building Materials. *Applied Spectroscopy Reviews*, **50**(4), 279–303.

Acbas, G., Niessen, K. A., Snell, E. H., & Markelz, A. G. (2014) Optical measurements of long-range protein vibrations. *Nat Commun*, **5**, 3076.

Agarwal P. K. (2005) Role of protein dynamics in reaction rate enhancement by enzymes. *J. Am. Chem. Soc.* **127**, 15248–15256.

Allis, D. G., Fedor, A. M., Korter, T. M., Bjarnason, J. E., & Brown, E. R. (2007) Assignment of the lowest-lying THz absorption signatures in biotin and lactose monohydrate by solid-state density functional theory. *Chemical Physics Letters*, **440**(4-6), 203–209.

Amrhein, S., Christin, B. K., Galm, L., Hubbuch, J. (2015) Non-invasive high throughput approach for protein hydrophobicity determination based on surface tension. *Biotechnol Bioeng,* **10,** 25677.

Anand, U., & Mukherjee, S. (2013) Binding, unfolding and refolding dynamics of serum albumins. *Biochimica et Biophysica Acta - General Subjects*, **1830**(12), 5394–5404.

Appleby, J. L., & Bourret, R. B. (1998) Proposed signal transduction role for conserved CheY residue Thr87, a member of the response regulator active-site quintet. *Journal of Bacteriology*, **180**(14), 3563–3569.

Arikawa, T., Nagai, M., & Tanaka, K. (2008) Characterizing hydration state in solution using terahertz time-domain attenuated total reflection spectroscopy. *Chemical Physics Letters*, **457**(1-3), 12–17.

Arnone, D. D. (2006) Terahertz pulsed imaging of human breast tumors. *Radiology,* **239**(2), 533-40.

Arora, A., Luong, T. Q., Krüger, M., Kim, Y. J., Nam, C. H., Manz, A., & Havenith, M. (2012) Terahertz-time domain spectroscopy for the detection of PCR amplified DNA in aqueous solution. *Analyst*, **137**(3), 575-9.

Ashcroft, M., Kubbutat, M. H., & Vousden, K. H. (1999) Regulation of p53 function and stability by phosphorylation. *Mol Cell Biol*, **19**(3), 1751–1758.

Austin, R., Roberson, M., & Mansky, P. (1989) Far-infrared perturbation of reaction rates in myoglobin at low temperatures. *Phys. Rev. Lett*, **62**, 1912–1915.

Auston, D. H., Cheung, K. P., Valdmanis, J. A., & Kleinman, D. A. (1984) Cherenkov radiation from femtosecond optical pulses in electro-optic media. *Physical Review Letters*, **53**(16), 1555–1558.

Bahuguna, G. P., Tandon, P., Gupta, V. D., Rastogi, S., & Mehrotra, C. (1997) Vibrational dynamics of poly(L-tyrosine). *Journal of Macromolecular Science*, **36**(4), 535-552.

Bailey, L.E., Navarro, R., Hernanz, A. (1997) Normal coordinate analysis and vibrational spectra of adenosine. *Biospectroscopy*, **3**, 47.

Baker, M. E. (1998) Albumin's role in steroid hormone action and the origins of vertebrates: Is albumin an essential protein? *FEBS Letters*, **439**(1-2), 9-12.

Ball, P. (2008) Water as an active constituent in cell biology. *Chem Rev,* **108**(1), 74-108.

Bandekar, J., Genzel, L., Kremer, F., & Santo, L. (1983) The temperature-dependence of the far-infrared spectra of L-alanine. *Spectrochim. Acta. Part A*, **39**, 357-366.

Barron, L. D., Hecht, L., & Wilson, G. (1997) The lubricant of life: A proposal that solvent water promotes extremely fast conformational fluctuations in mobile heteropolypeptide structure. *Biochemistry*, **36**(43), 13143-7.

Barth, A. (2000) The infrared absorption of amino acid side chains. *Prog. Biophys. Mol. Biol*, **74**(3-5), 141–73.

Barthes, M., Fahre Vik, A., Spire, A., Bordallo, H. N., & Eckert, J. (2002) Breathers or structural instability in solid L-alanine: A new IR and inelastic neutron scattering vibrational spectroscopic study. *Journal of Physical Chemistry A*, **106**(21), 5230–5241.

Beets, Jr. C. P., & Ascarelli, G. (1982) Far-infrared absorption of nucleotides and poly(I)·poly(C) RNA. *Biopolymers,* **21**, 1569-1586.

Beetz, C. P., & Ascarelli, G. (1980) The low frequency vibrations of pyrimidine and purine bases. *Spectrochimica Acta Part A: Molecular Spectroscopy*, **36**(3), 299–313.

Bellamy, L. J., (1975) The Infrared Spectra of Complex Molecules, *Chapman and Hall,* 3rd Ed.

Bendeif, E. E., & Jelsch, C. (2007) The experimental library multipolar atom model refinement of l-aspartic acid. *Acta Crystallographica Section C: Crystal Structure Communications*, **63**(6), 361–364.

Benkovic S., Hammes-Schiffer S. (2003) A perspective on enzyme catalysis. *Science,* **301***,* 1196–1202.

Bergner, A., Heugen, U., Bründermann, E., Schwaab, G., Havenith, M., Chamberlin, D. R., & Haller, E. E. (2005) New p-Ge THz laser spectrometer for the study of solutions: THz absorption spectroscopy of water. *Review of Scientific Instruments*, **76**(6).

Bernardi, G. (2000) Isochores and the evolutionary genomics of vertebrates, *Gene,* **241**(1), 3.

Bigay, J., Deterre, P., Pfister, C., & Chabre, M. (1987) Fluoride complexes of aluminium or beryllium act on G-proteins as reversibly bound analogues of the gamma phosphate of GTP. *The EMBO Journal*, **6**(10), 2907–2913.

Bigelow, C. C. (1967) On the average hydrophobicity of proteins and the relation between it and protein structure. *Journal of Theoretical Biology*, **16**(2), 187–211.

Birch, J. R. (1992) The far-infrared optical constants of polypropylene, PTFE and polystyrene. *Infrared physics,* **33,** 33-38.

Blake, C., Koenig, D. F., Mair, G. A., North, A. C. T., Phillips, D. C., & Sarma, V. R. (1965) Structure Of Hen Egg-White Lysozyme - A 3-Dimensional Fourier Synthesis At 2a Resolution. *Nature*, **206**(4986), 757–61.

Blake, C. C. F., Johnston, L. N., Mair, G. A., North, A. C. T., Philips, D. C., & Sarma, V. R. (1967) Crystallographic studies of activity of hen egg-white lysozyme. *Proc. Roy. Soc,* **B167**, 378-388.

Blaszczyk, U., Polit, A., Guz, A., & Wasylewski, Z. (2001). Interaction of cAMP receptor protein from Escherichia coli with cAMP and DNA studied by dynamic light

scattering and time-resolved fluorescence anisotropy methods. *Journal of Protein Chemistry*, **20**(8), 601–610.

Bolívar, H. P, Nagel, M., Richter, F., Brucherseifer, M., Kurz, H., Bosserhoff, A.,& Büttner, R. (2004). Label-free THz sensing of genetic sequences: towards 'THz biochips'. *Philos Trans A Math Phys Eng Sci*, **362**(1815), 323-33.

Bolivar, H. P., Brucherseifer, M., Nagel, M., Kurz, H., Bosserhoff, A., Büttner, R. (2002) Label-free probing of genes by time-domain terahertz sensing. *Phys Med Biol*, **47**(21), 3815-3821.

Bonvalet, A., Joffre, M., Martin, J. L., and Migus, A. (1995) Generation of ultrabroadband femtosecond pulses in the mid-infrared by optical rectification of 15 fs light pulses at 100 MHz repetition rate. *Appl. Phys. Lett*, **67**, 2907.

Borland, G., Smith, B. O., and Yarwood, J. (2009) EPAC proteins transduce diverse cellular actions of cAMP. *British Journal of Pharm,* **158,** 70-86.

Born, B., Kim, S. J., Ebbinghaus, S., Gruebele, M., & Havenith, M. (2009) The terahertz dance of water with the proteins: the effect of protein flexibility on the dynamical hydration shell of ubiquitin. *Faraday Discussions*, **141**, 161–173

Botsford, J. L., & Harman, J. G. (1992) Cyclic AMP in prokaryotes. *Microbiological Reviews*, **56**(1), 100–122.

Bourret, R. B., Thomas, S. A, Page, S. C., Creager-Allen, R. L., Moore, A. M., & Silversmith, R. E. (2010) Measurement of response regulator autodephosphorylation rates spanning six orders of magnitude. *Methods in Enzymology*, **471**(10), 89–114.

Brandt, N. N., Chikishev, A. Y., Kargovsky, A.V., Nazarov, M.M., Parashchuk, O.D., Sapozhnikov, D.A., Smirnova, I.N., Shkurinov, A.P., Sumbatyan N.V. (2008) Terahertz time-domain and Raman spectroscopy of the sulfur-containing peptide

dimers: Low-frequency markers of disulfide bridges. *Vibrational Spectroscopy,* **47**(1), 53-58.

Brooks, B., & Karplus, M. (1985). Normal modes for specific motions of macromolecules: application to the hinge-bending mode of lysozyme. *Proceedings of the National Academy of Sciences of the United States of America*, **82**(15), 4995–4999.

Brown, T. (2001) Southern Blotting and Related DNA Detection Techniques. *Encyclopaedia of life sciences.*

Brucherseifer, M., Nagel, M., Bolivar, P.H., Kurz. H., Bosserhoff, A., and Buttner, R. (2000) Label-free probing of the binding state of DNA by time-domain terahertz sensing. *Appl. Phy. Lett,* **77**(24), 4049-4051.

Buckingham, A. D. (1956) Birefringence Resulting from the Application of an Intense Beam of Light to an Isotropic. Medium. *Proc. Phys. Soc*, **B69**, 910.

Caliskan, G., Mechtani, D., Roh, J. H., Kisliuk, A., Sokolov, A. P., Azzam, S., Peral, I. (2004) Protein and solvent dynamics: How strongly are they coupled? *Journal of Chemical Physics*, **121**(4), 1978–1983.

Canfield, R. E. (1963) The amino acid sequence of egg white lysozyme. *J. Biol. Chem*, **238**, 2698–2707.

Cang, H., Li, J., Novikov, V. N., & Fayer, M. D. (2003) Dynamical signature of two "ideal glass transitions" in nematic liquid crystals. *Journal of Chemical Physics*, **119**(19), 10421–10427.

Carmona, P., & Molina, M. (1990) On the Raman spectra of mononucleotides. *Journal of Molecular Structure*, **219**, 323–328.

Carter, D. C., He, X. M., Munson, S. H., Twigg, P. D., Gernert, K. M., Broom, M. B., & Miller, T. Y. (1989) Three-dimensional structure of human serum albumin. *Science (New York, N.Y.)*, **244**(4909), 1195–1198.

Casado, J., López Navarrete, J. T., Ramírez, F. J. (1995) Force field and normal coordinate calculations of the amino acid L-asparagine. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, **51**(13)**,** 2347-2356.

Cashell, C., Corcoran, D., & Hodnett, B. K. (2003) Secondary nucleation of the beta-polymorph of L-glutamic acid on the surface of alpha-form crystals. *Chemical Communications (Cambridge, England)*, **3**, 374–375.

Chaney M.O., Steinrauf L.K., (1974) The Crystal and Molecular Structure of Tetragonal L-Cystine. *Acta Cryst*, **B30**, 711.

Chang, C., & Stewart, R. C. (1998). The two-component system. Regulation of diverse signaling pathways in prokaryotes and eukaryotes. *Plant Physiology*, **117**(3), 723–731.

Chantry, G. W. (1984) Long-wave optics: The science and technology of infrared and near-millimetre waves. Academic Press, Inc., London. **2**. Chap, 6, 574 – 576.

Chen Y.Z., Szabó A., Schroeter D.F., Powell J.W., Lee S. A., Prohofsky E. W., (1997) Effect of drug-binding-induced deformation on the vibrational spectrum of a DNA·daunomycin complex. *Phys. Rev*, **E 55**, 7414.

Chen, J.-Y., Knab, J. R., Cerne, J., and Markelz. A. G. (2005) Large oxidation dependence observed in terahertz dielectric response for cytochrome C. *Phys. Rev. E*, **72**(4).

Chernobai, G. B., Chesalov, Y. A., Burgina, E. B., Drebushchak, T. N., & Boldyreva, E. V. (2007) Temperature effects on the IR spectra of crystalliine amino acids,

dipeptides, and polyamino acids. I. Glycine. *Journal of Structural Chemistry*, **48**(2), 332–339.

Chhiba, M., Derreumaux, P., Vergoten, G. (1994) The Use of the Spasiba Spectroscopic Potential for Reproducing the Structures and Vibrational Frequencies of a Series of Acids - Acetic-Acid, Pivalic Acid, Succinic Acid, Adipic Acid And L-Glutamic Acid. *J. Mol. Struct,* **317**, 171-184.

Cho, H. S., Lee, S. Y., Yan, D., Pan, X., Parkinson, J. S., Kustu, S., Pelton, J. G. (2000). NMR structure of activated CheY. *Journal of Molecular Biology*, **297**(3), 543–551.

Choudhury M., Pettitt B.M., (2005) On the Mechanism of Hydrophobic Association of Nanoscopic Solutes. *J. Am. Chem. Soc,* **127** (10), 3556–3567.

Choudhury N., Pettitt B. M., (2005) Dynamics of water trapped between hydrophobic solutes. *J. Phys. Chem. B*, **109,** 6422–6429.

Chowdhry, B. Z., Dines, T. J., Jabeen, S., Withnall, R. (2008) Vibrational Spectra of α-Amino Acids in the Zwitterionic State in Aqueous Solution and the Solid State: DFT Calculations and the Influence of Hydrogen Bonding. *J. Phys. Chem,* **112**. 10333.

Christensen, A. E., Sleheim, F., de Rooij, J., Dremier, S., Schwede, F., Dao, K., Martinez, A., Maenhaut C., Bos J. L., Genieser, H. G., Doskeland, S. O. (2003) cAMP analog mapping of Epac1 and cAMP kinase. Discriminating analogs demonstrate that Epac and cAMP kinase act synergistically to promote PC-12 cell neurite extension. *J Biol Chem,* **278**, 35394–35402.

Clark, A. H., Saunderson, D. H., & Suggett, A. (1981) Infrared and laser-Raman spectroscopic studies of thermally-induced globular protein gels. *International Journal of Peptide and Protein Research*, **17**(3), 353–364.

Cole P. A, Shen K, Qiao Y, Wang D., (2003). Protein tyrosine kinases Src and Csk: a tail's tale. *Curr. Opin. Chem. Biol.* **7,** (5), 580–5.

Combeau, C., & Carlier, M. F. (1989) Characterization of the aluminum and beryllium fluoride species bound to F-actin and microtubules at the site of the γ-phosphate of the nucleotide. *Journal of Biological Chemistry*, **264**(32), 19017–19021.

Cozzone, P. J., Opella, S. J., Jardetzky, O., Berthou, J., & Jolles, P. (1975) Detection of new temperature-dependent conformational transition in lysozyme by carbon-13 nuclear magnetic resonance spectroscopy. *Proc. Natl. Acad. Sci*, **72**(6), 2095–2098.

Cross, C. W., Rice, J. S., & Gao, X. (1997) Solution structure of an RNA x DNA hybrid duplex containing a 3'-thioformacetal linker and an RNA A-tract. *Biochemistry*. 8;**36**(14), 4096-107.

Cuellar, M. C., Garcia, D. S., Straathof, A. J. J., Heijnen, J. J., & Wielen, L. A. M. Van Der. (2014). Crystallization of L-Phenylalanine anhydrate for product recovery during fermentation. *13th International Workshop on Industrial Crystallization.*

Dahaoui, S., Pichon-Pesme, V., Howard, J .A .K., Lecomte, C. (1999) CCD charge density study on crystals with large unit cell parameters: the case of hexagonal L-cystine par. *J. Phys. Chem*, **A103,** 6240-6250.

Danilov, V. I., Zheltovsky, N. V., Slyusarchuk, O. N., Poltev, V. I., & Alderfer, J. L. (1997) The study of the stability of Watson-Crick nucleic acid base pairs in water and dimethyl sulfoxide: computer simulation by the Monte Carlo method. *Journal of Biomolecular Structure & Dynamics*, **15**(1), 69–80.

Davydov, A. S. (1973) The theory of contraction of proteins under their excitation. *J. Theor. Biol,* **38**, 559–569.

De Boeij, W. P., Pshenichnikov, M. S., & Wiersma, D. A. (1998) Ultrafast solvation dynamics explored by femtosecond photon echo spectroscopies. *Annual Review of Physical Chemistry*, **49**, 99–123.

De Rooij, J., Zwartkruis, F. J. T., Verheijen, M. H. G., Cool, R. H., Nijman, S. M. B., Wittinghofer, A., and Bos, J. L. (1998) Epac is a Rap1 guanine-nucleotide-exchange factor directly activated by cyclic AMP. *Letters to Nature,* **396,** 474-477.

De Wit, J., and Klarenbeek, G. (1984) Effects of various heat treatments on structure and solubility of whey proteins. *J. Dairy Sci*, **67,** 2701-2710.

Deak, J., Chin, H., Lewis, C. & Miller, R., (1998) Ultrafast phase grating studies of hemeproteins: observation of the low-frequency modes directing functionally important protein motions. *J. Phys. Chem*, **B 102,** 6621–6634.

Deckert-Gaudig, T., Deckert, V. (2009) Journal of Raman Spectroscopy. *Tip-enhanced Raman Spectroscopy*, **40**(10), 1446–1451.

Degtyarenko, I. M., Jalkanen, K. J., Gurtovenko, A. A., & Nieminen, R. M. (2007) L-alanine in a droplet of water: A density-functional molecular dynamics study. *Journal of Physical Chemistry B*, **111**(16), 4227–4234.

Delabar, J.M., Majoube, M. (1978) Infrared and Raman spectroscopic study of 15-N and D-substituted guanines. *Spectrochim. Acta,* **34A,** 129.

Demarco C., Lindsay S.M., Pokorny M., Powell J., Rupprecht A., (1985) Interhelical effects on the low-frequency modes and phase transitions of Li- and Na-DNA. *Biopolymers***, 11**, 2035-40.

Destrade, C., Garrigou-Lagrange, C., Forel, M. T. (1971) Détermination du champ de force de valence de l'ion H3N+CH2COO- : Son apport dans l'étude de la conformation. *J. Mol. Struct,* **10**, 203.

Devlin M. T., Barany G., Levin I.W., (1990) Conformational properties of asymmetrically substituted mono-, di- and trisulfides: solid and liquid phase Raman spectra. *Journal of Molecular Structure*, **238,** 119-137.

Dhamelincourt, P., & Ramirez, F. J. (1993). Polarized micro-Raman and FT-IR spectra of L-glutamine. *Applied Spectroscopy*, **47**(4), 446–451.

Diem, M., Polavarapu, Oboodi, M., Nafie, L. A. (1982) Vibrational circular dichroism in amino acids and peptides. 4. Vibrational analysis, assignments, and solution-phase Raman spectra of deuterated isotopomers of alanine. *J. Am. Chem. Soc,* **104**, 3329.

Dolgikh, D. A, Gilmanshin, R. I., Brazhnikov, E. V, Bychkova, V. E., Semisotnov, G. V, Venyaminov S. Yu, & Ptitsyn, O. B. (1981) Alpha-Lactalbumin: compact state with fluctuating tertiary structure? *FEBS Letters*, **136**(2), 311–315.

Dolgikh, D. A., Abaturov, L. V., Bolotina, I. A., Brazhnikov, E. V., Bychkova, V. E., Gilmanshin, R. I., Lebedev, Y., Semisotnov, G. V., Tiktopulo, E. I., Ptitsyn, O. B. (1985) Compact state of a protein molecule with pronounced small-scale mobility: Bovine α-lactabumin. *Eur. Biophys*. J, **13,** 109–121.

Dong, A, Malecki, J. M., Lee, L., Carpenter, J. F., & Lee, J. C. (2002) Ligand-induced conformational and structural dynamics changes in Escherichia coli cyclic AMP receptor protein. *Biochemistry*, **41**(21), 6660–6667.

Doolittle, R. F. (1989) Redundancies in protein sequences, in Fasman, G. D., Prediction of Protein Structures and the Principles of Protein Conformation, *New York: Plenum*, 599–623

Dragoman, D., Dragoman, M. (2004) Terahertz fields and application. *Progress in Quantum Electronics*. **28**, 1–66.

Dubin S. B., Clark N. A., Benedek G. B., (1971) Measurement of the Rotational Diffusion Coefficient of Lysozyme by Depolarized Light Scattering: Configuration of Lysozyme in Solution. *J. Chem. Phys*. **54,** 5158-5164.

Dzwolak, W., Kato, M., Shimizu, A., & Taniguchi, Y. (1999) Fourier-transform infrared spectroscopy study of the pressure-induced changes in the structure of the bovine alpha-lactalbumin: the stabilizing role of the calcium ion. *Biochimica et Biophysica Acta*, **1433**(1-2), 45–55.

Ebright, Y. W., Chen, Y., Pendergrast, P. S., & Ebright, R. H. (1992). Incorporation of an EDTA-metal complex at a rationally selected site within a protein: application to EDTA-iron DNA affinity cleaving with catabolite gene activator protein (CAP) and Cro. *Biochemistry*, **31**(44), 10664–10670.

Emanuel, C. F., & Chaikoff, I. L. (1953) The large scale preparation of sodium desoxyribonucleate from ripe salmon testes. *J Biol Chem*, **203**(1), 167-71.

Epstein, W., Rothman-Denes, L. B., & Hesse, J. (1975) Adenosine 3':5'-cyclic monophosphate as mediator of catabolite repression in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, **72**(6), 2300–2304.

Façanha Filho, P. F., Freire, P. T. C., Lima, K. C. V., Mendes Filho, J., Melo, F. E. A., Pizani, P. S. (2008) High temperature Raman spectra of L-leucine. *Braz. J. Phyz.* **38**(1).

Falconer, R. J., & Markelz, A. G. (2012) Terahertz spectroscopic analysis of peptides and proteins. *Journal of Infrared, Millimeter, and Terahertz Waves*, **33**(10), 973–988.

Faria J. L. B., Freire P. T. C., Gonçalves, R. O., Melo F. E. A., Mendes Filho J., Lima, R. J. C. Moreno A. J. D. (2010) Polarized Raman spectra of L-arginine hydrochloride monohydrated single crystal. *Brazilian Journal of Physics*, **40**, 288-294.

Feairheller, W. R, Miller, Jr, J. T. (1971) The infrared spectra of amino acids and some simple polypeptides to 33 cm^{-1} at room and liquid nitrogen temperatures. *Appl. Spectrosc*, **25**, 175 – 181.

Ferrari, E. S., & Davey, R. J. (2004) Solution-mediated transformation of α to β L-glutamic acid: Rate enhancement due to secondary nucleation. In *Crystal Growth and Design*, **4**, 1061–1068.

Fic, E., Bonarek, P., Gorecki, a., Kedracka-Krok, S., Mikolajczak, J., Polit, a., Wasylewski, Z. (2009) CAMP receptor protein from escherichia coli as a model of signal transduction in proteins - A review. *Journal of Molecular Microbiology and Biotechnology*, **17**(1), 1–11.

Fic, E., Polit, A., & Wasylewski, Z. (2006) Kinetic and structural studies of the allosteric conformational changes induced by binding of cAMP to the cAMP receptor protein from Escherichia coli. *Biochemistry*, **45**(2), 373–380.

Filho, P. F. F., Freire, P. T. C., Lima, K. C. V., Filho, J. M., Melo, F. E. a., & Pizani, P. S. (2007) Raman spectra of L-leucine crystals, *Brazilian Journal of Physics,* **38**(1), 12.

Fischer, B. F., Hoffmann, M., Helm, H., Wilk, R., Rutz, F., Kleine-Ostmann, T., Koch, M., Jepsen, P. U. (2005) Terahertz time-domain spectroscopy and imaging of artificial *RNA. Optics Express,* **13**(14), 5205-5215.

Fischer, B. M., Walther, M., Jepsen, P. U. (2002) Far-infrared vibrational modes of DNA components studied by terahertz time-domain spectroscopy. *Phys. Med. Biol,* **47,** 3807–3814.

Fischer, B. M., Franz, M., and Abbott, D. (2006) THz spectroscopy as a versatile tool for investigating crystalline structures. *Infrared Millimeter Waves,* **14,** 362 – 362.

Fisher, A. J., Smith, C. A., Thoden, J. B., Smith, R., Sutoh, K., Holden, H. M., & Rayment, I. (1995) X-ray structures of the myosin motor domain of Dictyostelium discoideum complexed with MgADP.BeFx and MgADP.AlF4-. *Biochemistry*, **34**(28), 8960–8972.

Florian J. (1993) Vibrational motions of bases of nucleic acids by neutron inelastic scattering and resonance Raman spectroscopy. 1. Adenine and its deuterated species. *J Phys Chem,* **97**, 10889.

Florián, J., & Hrouda, V. (1993) Scaled quantum mechanical force fields and vibrational spectra of solid state nucleic acid constituents V: thymine and uracil. *Spectrochimica Acta Part A: Molecular Spectroscopy*, **49**(7), 921–938.

Foltynowicz, R. J., Allman, R. E., Zuckerman, E. (2006) Terahertz absorption measurement for gas-phase 2,4-dinitrotoluene from 0.05 THz to 2.7 THz. *Chem. Phys. Lett,* **431**, 34–38.

Frederick, K. K., Marlow, M. S., Valentine, K. G., & Wand, A. J. (2007) Conformational entropy in molecular recognition by proteins. *Nature*, **448**(7151), 325–329.

Fröhlich, H. (1985) Further evidence for coherent excitations in biological systems. *Phys. Letts. A*. **110.** 9.

Fukushima, K., Onishi, T., Shimanouchi, T., Mizushima, S. (1959) Assignment of vibration bands of dl-alanine. *Spectrochimica Acta,* **15,** 236-241.

Furberg, S., Petersen, C. S., Romming, C. (1965) A refinement of the crystal structure of cytidine. *Acta. Crystallogr.* **18**. 313.

Garibyan, L., & Avashia, N. (2013) Research Techniques Made Simple: Polymerase Chain Reaction (PCR). *J Invest Dermatol.* **133**(3), 6.

Gaigeot, M.-P., Leulliot, N., Ghomi, M., Jobic, H., Coulombeau, C., & Bouloussa, O. (2000) Analysis of the structural and vibrational properties of RNA building blocks by means of neutron inelastic scattering and density functional theory calculations. *Chemical Physics*, **261**(1-2), 217–237.

Garges, S., & Adhya, S. (1985) Sites of allosteric shift in the structure of the cyclic AMP receptor protein. *Cell*, **41**(3), 745–751.

Garti, N., & Zour, H. (1997) The effect of surfactants on the crystallization and polymorphic transformation of glutamic acid. *Journal of Crystal Growth*, **172**(3-4), 486–498.

Gavira, J. M., Campos, M., Diaz, G., Hernanz, A., & Navarro, R. (1997) Vibrational analysis and spectra of cytidine 3′-monophosphate (3′-CMP). *Vibrational Spectroscopy*, **15**, 1–16.

Genzel, L., Keilmann, F., Martin, T. P., Winterling, G., Yacoby, Y., Fröhlich, H., & Makinen, M. W. (1976) Low-frequency Raman spectra of lysozyme. *Biopolymers*, **15**(1), 219–225.

George D.K., Knab Joseph R., He Y., Kumauchi M., Birge Robert R., Hoff W.D., Markelz A.G. (2013) Photoactive yellow protein terahertz response: hydration, heating and intermediate states. *IEEE Trans. THz Sci,* **3**, 288–294.

Giese, B., & McNaughton, D. (2002). Density functional theoretical (DFT) and surface-enhanced Raman spectroscopic study of guanine and its alkylated derivatives. *Phys. Chem. Chem. Phys*, **4**(20), 5171–5182.

Giraud, G., & Wynne, K. (2003). A comparison of the low-frequency vibrational spectra of liquids obtained through infrared and Raman spectroscopies. *Journal of Chemical Physics*, **119**(22), 11753–11764.

Globus, T., Bykhovskaia, M., Woolard, D., and Gelmont, B. (2003) Sub-millimetre wave absorption spectra of artificial RNA molecules. *J. Phys. D: Appl. Phys,* **36**, 1314–1322.

Globus T., Moyer A.M., Gelmont B., Khromova T., Lvovska M.I., Sizov I., Ferrance J. (2013) Highly Resolved Sub-Terahertz Vibrational Spectroscopy of Biological Macromolecules and Cells. *IEEE Sensors Journal,* **13**(1), 72-79.

Globus, T., Parthasarathy, R., Khromova, T., Woolard, D., Swami, N., Gatesman, A. J., & Waldman, J. (2004) Optical characteristics of biological molecules in the Terahertz gap. *Analysis*, **5584**, 1–10.

Globus, T., Woolard, D., Crowe, T.W., Khromova, T., Gelmont, B. and Hessler J. (2006) Terahertz Fourier transform characterization of biological materials in a liquid phase. *Journal of Physics D*: *Applied Physics,* **39,** 3405-3413.

Globus, T., Woolard, D. L., Samuels, A. C., Gelmont, B. L., Hesler, J., Crowe, T. W., & Bykhovskaia, M. (2002) Submillimeter-wave Fourier transform spectroscopy of biological macromolecules. *Journal of Applied Physics*, **91**(9), 6105–6113.

Glorieux, C., Nelson, K. A., Hinze, G., Fayer, M. D. (2002) Thermal, structural and orientational relaxation of supercooled salol studied by polarization-dependent impulsive stimulated scattering. *J. Chem. Phys,* **116,** 3384-3395.

Görbitz, C. H., Dalhus, B. (1996) Redetermination of L-leucine at 120K. *Acta Cryst,* **52**, 1754-1756.

Goryainov, S. V., Kolesnik, E. N., Boldyreva, E. V. (2005) A reversible pressure-induced phase transition in β-glycine at 0.76 GPa. *Physica B,* **357,** 340.

Gottke, S. D., Brace, D. D., Cang, H., Bagchi, B., & Fayer, M. D. (2002) Liquid crystal dynamics in the isotropic phase. *Journal of Chemical Physics*, **116**(1), 360–367.

Grace, L. I., Cohen, R., Dunn, T. M., Lubman, D. M., & De Vries, M. S. (2002) The R2PI spectroscopy of tyrosine: A vibronic analysis. *Journal of Molecular Spectroscopy*, **215**(2), 204–219.

Graindourze, M., Grootaers, T., Smets, J., Zeegers-Huyskens, T., Maes G. (1991) FT-IR spectroscopic study of of uracil deriviatives and their hydrogen bonded complexes with proton donors. Hydrogen bonding of uracils with H_20 in Ar matrices. *J Mol. Struct,* **243**, 37-60.

Griko, Y. V, & Remeta, D. P. (1999) Energetics of solvent and ligand-induced conformational changes in alpha-lactalbumin. *Protein Science: A Publication of the Protein Society*, **8**(3), 554–561.

Griko, Y. V., Freire, E., & Privalov, P. L. (1994). Energetics of the alpha-lactalbumin states: a calorimetric and statistical thermodynamic study. *Biochemistry*, **33**(7), 1889–1899.

Grischkowsky, D., Keiding, S., Van Exter, M., and Fattinger, C. (1990) Far-infrared time-domain spectroscopy with terahertz beams of dielectrics and semiconductors. *J. Opt. Soc. Am,* **7,** 2006-2015.

Grischkowsky. D., and Katzenellenbogen, N. (1991) Femtosecond pulses of terahertz radiation: Physics and applications. *Picosecond Electronics and Optoelectronics*, **9**, 9-14.

Grunenberg, A., D. Bougeard. (1987) Vibrational spectra and conformational phase transition of crystalline L-methionine. *J. Mol. Struct,* **160**, 27–36.

Gu, P., Masahiko, T. (2005) Terahertz Radiation from Semiconductor Surfaces. *Topics Appl. Phys,* **97**, 63-97.

Gupta V. D., Singh R. D., Dwivedi A. M., (1973) Vibrational-Spectra and Dispersion Curves of Poly-L-Proline-II chain. *Biopolymers,* **12**(6), 1377-1385.

Gupta V. D., Singh R. D. (1970) Low frequency spectra of glycine. *Chemical Physics Letters.* **5**(4), 218-220.

Håkansson, A., Svensson, M., Mossberg, A. K., Sabharwal, H., Linse, S., Lazou, I., Svanborg, C. (2000) A folding variant of alpha-lactalbumin with bactericidal activity against Streptococcus pneumoniae. *Molecular Microbiology*, **35**(3), 589–600.

Halle, B., & Davidovic, M. (2003) Biomolecular hydration: from water dynamics to hydrodynamics. *Proceedings of the National Academy of Sciences of the United States of America*, **100**(21), 12135–12140.

Hamster, H., A. Sullivan, et al. (1993) Subpicosecond, electromagnetic pulses from intense laser-plasma interaction. *Phys Rev Lett,* **71**(17), 2725-2728.

Harada I., Lord R. C. (1970) Low-frequency infrared and Raman spectra of some adenine and uracil crystals. *Spectrochimica Acta Part A: Molecular Spectroscopy*, **26**(12), 2305-2318.

Harding, M. M., Long, H. A. (1968) The crystal and molecular structure of L-cysteine. *Acta Cryst,* **24,** 1096-1102.

Haring Bolivar, H. P., M. Nagel, et al. (2004) Label-free THz sensing of genetic sequences: towards 'THz biochips'. Philosophical Transactions of the Royal Society of London, Series A: Mathematical, *Physical and Engineering Sciences,* **362**(1815), 323-335.

Haring Bolivar, P., Brucherseifer, M., Nagel, M., Kurz, H., Bosserhoff, a, & Büttner, R. (2002) Label-free probing of genes by time-domain terahertz sensing. *Physics in Medicine and Biology*, **47**(21), 3815–3821.

Harman, J. G. (2001). Allosteric regulation of the cAMP receptor protein. *Biochimica et Biophysica Acta - Protein Structure and Molecular Enzymology*, **1547**(1), 1–17.

Hasebe, T., Kawabe, S., Matsui, H., Tabata, H. (2012) Metallic mesh-based terahertz biosensing of single- and double-stranded DNA. *J. Appl. Phys,* **112,** 702.

Hay, S., Scrutton, N. S. (2012) Good vibrations in enzyme-catalysed reactions. *Nat. Chem.* **29.** 4. 161-168.

He, Y., Ku, P. I., Knab, J. R., Chen, J. Y., & Markelz, A. G. (2008) Protein dynamical transition does not require protein structure. *Physical Review Letters*, **101**(17), 24.

Hédoux, A., Guinet, Y., & Paccou, L. (2011) Analysis of the mechanism of lysozyme pressure denaturation from raman spectroscopy investigations, and comparison with thermal denaturation. *Journal of Physical Chemistry B*, **115**(20), 6740–6748.

Hendrix, T., Griko, Y. V, & Privalov, P. L. (2000) A calorimetric study of the influence of calcium on the stability of bovine alpha-lactalbumin. *Biophysical Chemistry*, **84**(1), 27–34.

Henriksen, N. E., Engel, V. (2001) A Theoretical Analysis of Transient Signals and Their Relation to Nuclear Wave-Packet Motion. *Int. Rev. Phys. Chem*, **20**, 93.

Herlinger, A. W., Wenhold, S. L., & Long, T. V. (1970) Infrared spectra of amino acids and their metal complexes. II. Geometrical isomerism in bis (amino acidato) copper (II) complexes. *Journal of the American Chemical Society*, **92**(22), 6674–6681.

Heugen, U., Schwaab, G., Bründermann, E., Heyden, M., Yu, X., Leitner, D. M., & Havenith, M. (2006) Solute-induced retardation of water dynamics probed directly by terahertz spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America*, **103**(33), 12301–12306.

Heyden, M., & Havenith, M. (2010) Combining THz spectroscopy and MD simulations to study protein-hydration coupling. *Methods*, **52**(1), 74–83.

Heyduk, T., & Lee, J. C. (1989) Escherichia coli cAMP receptor protein: evidence for three protein conformational states with different promoter binding affinities. *Biochemistry*, **28**(17), 6914–6924.

Hill, R.L., and Brew, K. (1975) Lactose synthetase. *Adv. Enzymol. Relat. Areas Mol. Biol*, **43,** 411-490.

Hindle, F., Cuisset, A., Bocquet, R., and Mouret, G. (2007) Continuous-wave terahertz by photomixing: applications to gas phase pollutant detection and quantification, *CR Phys,* **9,** 262–275.

Hinze, G., Brace, D. D., Gottke, S. D., & Fayer, M. D. (2000) A detailed test of mode-coupling theory on all time scales: Time domain studies of structural relaxation in a supercooled liquid. *Journal of Chemical Physics*, **113**(9), 3723–3733.

Ho, P. S., van Holde, K. E., Johnson, W. C., & Shing, P. (1998). Principles of physical biochemistry. *Upper Saddle River*, N.J. Prentice-Hall.

Horvat, J., and Lewis, RA. (2009) Peeling adhesive tape emits electromagnetic radiation at terahertz frequencies. *Optics Letters*, **34**(14), 2195-2197.

Hruschak, J. (1995) Mosby's Diagnostic and Laboratory Test Reference. *AORN Journal*.

Hubbard, J. A., MacLachlan, L. K., King, G. W., Jones, J. J., Fosberry, A. P. (2003) Nuclear magnetic resonance spectroscopy reveals the functional state of the signalling protein CheY in vivo in Escherichia coli. *Mol Microbiol,* **49**(5), 1191-200.

Hull, H. H., Chang, R., & Kaplan, L. J. (1975) On the location of the sulfhydryl group in bovine plasma albumin. *Biochimica et Biophysica Acta*, **400**(1), 132–136.

Hummer, G., Garde, S., García, A. E., Paulaitis, M. E., & Pratt, L. R. (1998) The pressure dependence of hydrophobic interactions is consistent with the observed

pressure denaturation of proteins. *Proceedings of the National Academy of Sciences of the United States of America*, **95**(4), 1552–1555.

Hunt, N. T., Kattner, L., Shanks, R. P., & Wynne, K. (2007). The dynamics of water-protein interaction studied by ultrafast optical Kerr-effect spectroscopy. *Journal of the American Chemical Society*, **129**(11), 3168–3172.

Hunt, N. T., Meech, S. R. (2004) Orientational and interaction induced dynamics in the isotropic phase of a liquid crystal: polarization resolved ultrafast optical Kerr effect spectroscopy. *J. Chem. Phys,* **120**, 10828-10836.

Hunt, N. T., Turner, A. R., Wynne, K. (2005) Inter- and intramolecular hydrogen bonding in phenol derivatives: A model system for poly-L-tyrosine. *J. Phys. Chem. B*, **109**, 19008.

Hunt, N., Jaye, A., Hellman, A. & Meech, S. (2004) Ultrafast dynamics of styrene micro emulsions, polystyrene nano latexes, and structural analogues of polystyrene. *J. Phys. Chem,* **108**, 100–108.

Husan, S. K., Hasted, J. B., Rosen, D., Nicol, E., Birch, J. R. (1984) FIR spectra of saccharides and polysaccharides. *Infrared Physics*, **24**(2–3), 209-213.

Hwang, T. L., Shaka, A. J. (1995) Water Suppression That Works. Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients. *Journal of Magnetic Resonance*, **112**(2)**,** 275–279.

Igumenova, T. I., King Frederick, K., & Wand, A. J. (2006) Characterization of the fast dynamics of protein amino acid side chains using NMR relaxation in solution. *Chemical Reviews*, **106**, 1672–1699.

Itoh, T., Wada, Y., and Nakanishi, T. (1976) Differential Thermal Analysis of Milk proteins. *Agr. Biol. Chem*, **40**, 1083-1086.

Janczak, J., Zobel, D., & Luger, P. (1997) L-Threonine at 12 K. *Acta Cryst,* **C53***,* 1901-1904.

Jepsen, P. U., & Clark, S. J. (2007) Precise ab-initio prediction of terahertz vibrational modes in crystalline systems. *Chemical Physics Letters*, **442**(4-6), 275–280.

Jolles, J., Jauregui-Adell, J., Jolles, P. (1963) The chemical structure of hen's egg-white lysozyme: detailed study. *Biochim. Biophys. Acta,* **78**, 668–689.

Jollès, P., & Jollès, J. (1984) What's new in lysozyme research? - Always a model system, today as yesterday. *Molecular and Cellular Biochemistry*, **63**(2), 165–189.

Jolles, P., Saint-Blanchard, J., Allary, M., Perin, J. P., Cozzone, P. (1975) On the binding of N-acetylglucosamine and its short polymers to hen lysozyme at physiological temperature (40°C), *FEBS Lett*, **55**, 165–167.

Jordan, M., & Wurm, F. (2004) Transfection of adherent and suspended cells by calcium phosphate. *Methods,* **33,** 136-43.

Kafka, J. D. & Baer, T. (1987) Prism-pair dispersive delay lines in optical pulse compression. *Opt Lett,* **12**(6), 401-3.

Kamau, S. M., Cheison, S. C., Chen, W., Liu, X. M., & Lu, R. R. (2010) Alpha-lactalbumin: Its production technologies and bioactive peptides. *Comprehensive Reviews in Food Science and Food Safety*, **9**(2), 197–212.

Kapanidis, A. N., Ebright, Y. W., Ludescher, R. D., Chan, S., & Ebright, R. H. (2001) Mean DNA bend angle and distribution of DNA bend angles in the CAP-DNA complex in solution. *Journal of Molecular Biology*, **312**(3), 453–468.

Kataoka, M., Kuwajima, K., Tokunaga, F., & Goto, Y. (1997) Structural characterization of the molten globule of alpha-lactalbumin by solution X-ray scattering. *Protein Science : A Publication of the Protein Society*, **6**(2), 422–430.

Katz, L., & Penman, S. (1966) Association by hydrogen bonding of free nucleosides in non-aqueous solution. *J Mol Biol,* **15**(1), 220–231.

Katzenellenbogen, N., & Grischkowsky, D. (1991) Efficient generation of 380 fs pulses of THz radiation by ultrafast laser pulse excitation of a biased metal-semiconductor interface. *Applied Physics Letters*, **58**(3), 222–224.

Kay, L. E., Keifer, P., Saarinen, T. (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem,* **114**, 10663-5.

Kelly, S. M., Jess, T. J., & Price, N. C. (2005) How to study proteins by circular dichroism. *Biochimica et Biophysica Acta - Proteins and Proteomics*, **1751**(2), 119–139.

Kemp, M. (2006) Millimetre Wave and Terahertz Technology for the Detection of Concealed Threats – A Review. *Proc. of SPIE,* **6402.**

Kerr, K. A., Ashmore, J. P. (1973) Structure and conformation of orthorhombic L-cysteine. *ActaCryst,* **B29**, 2124-2127.

Kerr, J. (1875) A new relation between electricity and light: dielectrified media birefringent. *Phil. Mag*, **50**, 337.

Kerr, K. A., Ashmore, J. P., Koetzle, T. F. (1975) A neutron diffraction study of L-cysteine. *Acta Cryst,* **31,** 2022-2026.

Kharakoz, D. P., & Bychkova, V. E. (1997) Molten globule of human α-lactalbumin: Hydration, density, and compressibility of the interior. *Biochemistry*, **36**(7), 1882–1890.

Khawas. B. (1971) X-ray study of L-arginine HCl, L-cysteine, DL-lysine and DL-phenylalanine. *Acta Cryst,* **27,** 1571-1520.

Kikikuchi N., Tanno T., Watanabe M., Kurabayashi T. (2009) A Membrane Method for Terahertz Spectroscopy of Amino Acids. *Analytical Sciences,* **25**(3), 457-459.

Kingsford-Adaboh, R., Grosche, M., Dittrich, B., Luger, P. (2000) Arginine monohydrate at 100 K. *ActaCryst,* **C56**, 1274-1276.

Kirkwood, J. C., Ulness, D. J., and Albrecht, A. C., (2000) On the classification of the electric field spectroscopies: Application to Raman scattering. *J. phys. Chem. A*, **104**, 4167-4173.

Kirstenmacher, T. L., Rand, G. A., Marsh, R. E. (1974) Refinements of the crystal structure of DL-serine and anhydrous L- serine. *ActaCryst,* **B30**, 2573-2578.

Kitamura, M., Funahara, H. (1994) Effect of L- and D-Phenylalanine on Crystallization and Transtormation of L-Glutamic Acid Polymorphs. *J. Chem. Eng. Jpn,* **27**(1), 124-126.

Kitamura, M. (1989) Polymorphism in the crystallization of L-glutamic acid. *Journal of Crysta l Growth,* **96,** 541–546.

Kitamura, M. (1993) Crystallization behaviour and transformation kinetics of L-histidine polymorphs. *Chem. Eng. Jpn*, **26,** 303-307.

Kleuss, C., Raw, A. S., Lee, E., Sprang, S. R., & Gilman, A. G. (1994) Mechanism of GTP hydrolysis by G-protein alpha subunits. *Proceedings of the National Academy of Sciences of the United States of America*, **91**(21), 9828–9831.

Kolb, A., Busby, S., Buc, H., Garges, S., Adhya, S. (1993) Transcriptional regulation by cAMP and its receptor protein. *Annu Rev Biochem*, **63**, 749–795.

Korenstein-Ilan, A., Barbul, A., Hasin, P., Eliran, A., Gover, A., and Korenstein, R. (2008) Terahertz Radiation Increases Genomic Instability in Human Lymphocytes. *Radiation Research Society,* **170,** 224-234.

Korter, T. M., Balu, R., Campbell, M. B., Beard, M. C., Gregurick, S. K., Heilweil, E. J. (2006) Terahertz spectroscopy of solid serine and cysteine. *Chem. Phys. Lett*, **418**, 65–70.

Kozlov, G., & Volkov, A. (1998) Coherent source submillimeter wave spectroscopy. *Millimeter and Submillimeter Wave Spectroscopy of Solids,* **74**, 51–109.

Krisch, M., Mermet, A., Grimm, H., Forsyth, V. T., & Rupprecht, A. (2006) Phonon dispersion of oriented DNA by inelastic x-ray scattering. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, **73**(6), 1–10.

Krittanai, C., & Johnson, W. C. (1997) Correcting the circular dichroism spectra of peptides for contributions of absorbing side chains. *Analytical Biochemistry*, **253**(1), 57–64.

Kunkel, T. A., Bebenek, K., McClary, J. (1991) Efficient site-directed mutagenesis using uracil-containing DNA. *Methods Enzymol*, **204**, 125-39.

Kurnit, N. A., Abella, I. D., & Hartmann, S. R. (1964) Observation of a photon echo. *Physical Review Letters*, **13**(19), 567–568.

Kuznetsov, A. N., Ebert, B., Lassmann, G., & Shapiro, A. B. (1975) Adsorption of small molecules to bovine serum albumin studied by the spin-probe method. *Biochimica et Biophysica Acta*, **379**(1), 139–146.

Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**(5259), 680–685.

Lagant P., Elass A., Dauchez M., Vergoten G., Peticolas W.L. (1992) Normal coordinate treatment of 1-methyluracil in the crystal state: Use of the ultraviolet resonance Raman intensities to improve the vibrational force field. *Spectrochimica Acta Part A: Molecular Spectroscopy,* **48**(10), 1323–1333.

Lakowicz, J. R. (2006) Principles of fluorescence spectroscopy. *3rd Edition.*

Laman, N., Harsha, S. S., Grischkowsky, D., & Melinger, J. S. (2008) High-resolution waveguide THz spectroscopy of biological molecules. *Biophysical Journal*, **94**(3), 1010–1020.

Larramendy, M. L., W. El-Rifai, et al. (1998) Comparison of fluorescein isothiocyanate- and Texas red-conjugated nucleotides for direct labeling in comparative genomic hybridization. *Cytometry,* **31**(3), 174-9.

Lee, A. L., Kinnear, S. A., & Wand, A. J. (2000) Redistribution and loss of side chain entropy upon formation of a calmodulin-peptide complex. *Nature Structural Biology*, **7**(1), 72–77.

Lee, A. W. M., Qin, Q., Kumar, S., Williams, B. S., Hu, Q., & Reno, J. L (2006) Real-time terahertz imaging over a standoff distance (>25 meters). *Applied Physics Letters*, **89**(14). 141125.

Lee, S. A., Anderson, A., Smith, W., Griffey, R. H., & Mohan, V. (2000) Temperature-dependent Raman and infrared spectra of nucleosides . Part I — adenosine. *Journal of Raman Spectroscopy*, **896**, 891–896.

Lee, S. A., Li, J., Anderson, A., Smith, W., Griffey, R. H., & Mohan, V. (2001) Temperature-dependent Raman and infrared spectra of nucleosides. II – Cytidine. *Journal of Raman Spectroscopy*, **32**(9), 795–802.

Lee, S. Y., Cho, H. S., Pelton, J. G., Yan, D., Berry, E. A., & Wemmer, D. E. (2001) Crystal structure of activated CheY: Comparison with other activated receiver domains. *Journal of Biological Chemistry*, **276**(19), 16425–16431.

Lee, S.-H., & Krimm, S. (1998) Ab initio-based vibrational analysis of α-poly(L-alanine). *Biopolymers*, **46**(5), 283–317.

Lee, Y-S. (2009) Principles of Terahertz Science and Technology. *Springer Science+Business Media.*

Lehmann, M. S., Verbist, J. J., Hamilton, W. C., Koetzle, T. F. (1973) Precision neutron diffraction structure determination of protein and nucleic acid components. Part V. Crystal and molecular structure of the amino acid L-agrinine dihydrate. *J. Chem. Soc,* **2**, 133-137.

Lerbret, A., Hédoux, A., Annighöfer, B., & Bellissent-Funel, M.-C. (2012) Influence of pressure on the low-frequency vibrational modes of lysozyme and water: a complementary inelastic neutron scattering and molecular dynamics simulation study. *Proteins*, **81**(2), 326–40.

Leś, A., Adamowicz, L., Nowak, M. J., & Lapinski, L. (1992) The infrared spectra of matrix isolated uracil and thymine: An assignment based on new theoretical calculations. *Spectrochimica Acta Part A: Molecular Spectroscopy*, **48**, 1385–1395.

Leulliot, N., Ghomi, M., Jobic, H., Bouuloussa, O., Baumruk, V., & Coulombeau, C. (1999) Ground State Properties of the Nucleic Acid Constituents Studied by Density Functional Calculations. 2. Comparison between Calculated and Experimental Vibrational Spectra of Uridine and Cytidine. *J. Phys. Chem. B*, **103**, 10934–10944.

Levitt, M., Sander, C., & Stern, P. S. (1985) Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme. *Journal of Molecular Biology*, **181**(3), 423–447.

Lewis, T. P., Miles, H. T., Becker E. D. (1984) Infrared and raman-spectra and vibrational assignments for 1-methyluracil and isotopic derivatives. *Journal of Physical Chemistry*, **88**(15), 3253-3260.

Li, T., Hassanali, A. A., Kao, Y. T., Zhong, D., & Singer, S. J. (2007) Hydration dynamics and time scales of coupled water-protein fluctuations. *Journal of the American Chemical Society*, **129**(11), 3376–3382.

Lima, J. A., Freire, P. T. C., Lima, R. J. C., Moreno, A. J. D., Mendes Filho, J., & Melo, F. E. A. (2005) Raman scattering of L-valine crystals. *Journal of Raman Spectroscopy*, **36**(11), 1076–1081.

Lin, V. J., & Koenig, J. L. (1976) Raman studies of bovine serum albumin. *Biopolymers*, **15**(1), 203–218.

Lindsay, S. M., S. A. Lee, et al. (1988) The origin of the A to B transition in DNA fibers and films. *Biopolymers,* **27**(6), 1015-43.

Liu, C., Edwards, G. S., Morgan, S., Silberman, E. (1989) Low-frequency, Raman-active vibrational modes of poly(dA)·poly(dT). *Phys. Rev*, **A40**, 7394.

Liu, D., Chu, X. Q., Lagi, M., Zhang, Y., Fratini, E., Baglioni, P., Alatas, A., Said, A., Alp, E., Chen, S.H. (2008) Studies of phononlike low-energy excitations of protein molecules by inelastic X-ray scattering. *Phys. Rev. Lett*, **101**, 135501.

Liu, H-B., Plopper, G., Earley, S., Chen, Y., Ferguson, B., Zhang, X.-C. (2006) Sensing minute changes in the biological cell monolayers with THz differential time-domain spectroscopy. *Biosens. and Bioelec,* **22**(6), 1075-1080.

Liu, R., He, M., Su, R., Yu, Y., Qi, W., & He, Z. (2009) Insulin amyloid fibrillation studied by terahertz spectroscopy and other biophysical methods. *Biochem. and Biophys. Research. Comms,* **391**, 862-867.

Loh, A. P., Pawley, N., Nicholson, L. K., & Oswald, R. E. (2001) An increase in side chain entropy facilitates effector binding: NMR characterization of the side chain methyl group dynamics in Cdc42Hs. *Biochemistry*, **40**(15), 4590–4600.

Lopes, R. P., Marque,s M. P. M., Valero, R., Tomkinson, J., Batista de Carvalho, L. A. E. (2012) Guanine: A Combined Study Using Vibrational Spectroscopy and Theoretical Methods. *Spectroscopy: An International Journal*, **27**(5-6)**,** 273–292.

Lopes, R. P., Valero, R., Tomkinson, J., Marques, M. P. M., Batista de Carvalhoa, L. A. E. (2013) Applying vibrational spectroscopy to the study of nucleobases – adenine as a case-study. *New J. Chem*, **37**, 2691-2699.

López Expósito, I., & Recio, I. (2006) Antibacterial activity of peptides and folding variants from milk proteins. *International Dairy Journal*, **16**(11), 1294–1305.

Lopez Navarrete, J. T., Bencivenni, L., Ramondo, F., Hernandez, V., Ramirez, F. J. (1995) Structural and Spectroscopical Study of Glutamic-Acid in the Non-Zwitterionic Form. *J. Mol. Struct*, **330**, 261-266.

Lord, R. C., Thomas Jr, G. J. (1967) Raman Spectral Studies of Nucleic Acids and Related Molecules. I. Ribonucleic Acid Derivatives. *Spectrochim. Acta,* **23,** 2551.

Ma, J. (2005) Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. *Structure*, **13**, 373–380.

Ma, L., & Cui, Q. (2007) Activation mechanism of a signaling protein at atomic resolution from advanced computations. *Journal of the American Chemical Society*, **129**(33), 10261–10268.

Machida, K., Kagayama, A., Saito, Y., Kuroda, Y., & Uno, T. (1977) Vibrational spectra and intermolecular potential of the α-form crystal of glycine. *Spectrochimica Acta Part A: Molecular Spectroscopy*, **33**(5), 569–574.

Madden, J. J., McGandy, E. L., Seeman, N. C., (1972) The crystal structure of the orthorhombic for of L-(+)-histidine. *ActaCryst,* **B28**, 2377-2382.

Majoube, M. (1985) Guanine Residue : A Normal-Coordinate Analysis of the Vibrational Spectra. *Biopolymers*, **24**(6), 1075–1087.

Makareeva, E., Han, S., Vera, J. C., Sackett, D. L., Holmbeck, K., Phillips, C. L., Visse, R., Nagase, H., and Leikin, S. (2010) Carcinomas Contain a Matrix Metalloproteinase–Resistant Isoform of Type I Collagen Exerting Selective Support to Invasion. *Mol and Cell Path,* **70**, 4366-4374.

Maker, P. D., and Terhune, R. W. (1964) Intensity-Dependent Changes in the Refractive Index of Liquids. *Phys. Rev. Lett*, **12**, 507.

Małecki, J., Polit, A., & Wasylewski, Z. (2000) Kinetic studies of cAMP-induced allosteric changes in cyclic AMP receptor protein from Escherichia coli. *The Journal of Biological Chemistry*, **275**(12), 8480–8486.

Mantsch, H. (1996) Infrared Spectroscopy of Biomolecules. *Wiley-Liss.*

Maret, G., Boccara, N., Kiepenheuer, J. (1979) Biophysical effects of steady magnetic fields. *Proceedings in Physics*, **11**, 99-106.

Markelz, A. G. (2008) Terahertz dielectric sensitivity to biomolecular structure and function. *IEEE Journal on Selected Topics in Quantum Electronics*, **14**(1), 180–190.

Markelz, A. G., Knab, J. R., Chen, J. Y., & He, Y. (2007) Protein dynamical transition in terahertz dielectric response. *Chemical Physics Letters*, **442**(4-6), 413–417.

Markelz, A. G., Roitberg, A., and Heilweil, E. J. (2000) Pulsed Terahertz Spectroscopy of DNA, Bovine Serum Albumin and Collagen between 0.1 and 2.0 THz. *Chem. Phys. Lett*, **320**, 42–48.

Markelz, A. G., Whitmire, S., Hillebrecht, J., and Birge, R. (2002) THz time domain spectroscopy of biomolecular conformational modes. *Phys. Med. Biol,* **47,** 3797-3805.

Marti, J., Padro, J. A., Guardia, E. (1996) Molecular Dynamics Simulation of Liquid Water Along the Coexistence Curve: Hydrogen Bonds and Vibrational Spectra. *J. Chem. Phys,* **105**, 639-649.

Matei, A., Drichko, N., Gompf, B., Dressel, M. (2005) Far-infrared spectra of amino acids. *Chem. Phys,* **316,** 61.

Mathlouthi, M., Seuvre, A. M., Koenig, J. L. (1984) F.t.-i.r. and Laser-Raman spectra of adenine and adenosine. *Carbohydr. Res,* **131,** 1-15.

Mayer, G., and Gires, F. (1964) Action d'une onde lumineuse intense sur l'indice de réfraction des liquids. *C. R. Acad. Sci. (Paris)*, **258**, 2039.

Mazur, K., Heisler, I. A, & Meech, S. R. (2010) Ultrafast dynamics and hydrogen-bond structure in aqueous solutions of model peptides. *The Journal of Physical Chemistry. B*, **114**(32), 10684–10691.

Mazur, K., Heisler, I. A, & Meech, S. R. (2011) THz spectra and dynamics of aqueous solutions studied by the ultrafast optical Kerr effect. *The Journal of Physical Chemistry. B*, **115**(11), 2563–2573.

Mazur, K., Heisler, I. A, & Meech, S. R. (2012) Water dynamics at protein interfaces: ultrafast optical Kerr effect study. *The Journal of Physical Chemistry. A*, **116**(11), 2678–85.

McDonald, L. R., Boyer, J. A., & Lee, A. L. (2012) Segmental motions, not a two-state concerted switch, underlie allostery in CheY. *Structure*, **20**(8), 1363–1373.

McDonald, L. R., Whitley, M. J., Boyer, J. A., & Lee, A. L. (2013) Colocalization of fast and slow timescale dynamics in the allosteric signaling protein CheY. *Journal of Molecular Biology*, **425**(13), 2372–2381.

McKenzie, H. A., White, F. H. (1991) Lysozyme and alpha-lactalbumin: structure, function, and interrelationships. *Adv. Protein Chem*, **41**, 173–315.

Meera, K., Muralidharan, R., Dhanasekaran, R., Manyum, P., & Ramasamy, P. (2004) Growth of nonlinear optical material: L-arginine hydrochloride and its characterisation. *Journal of Crystal Growth*, **263**(1-4), 510–516.

Mei, W. N., Kohli, M., Prohofsky, E. W., Van Zandt, L. L. (1981) Acoustic modes and nonbonded interactions of the double helix. *Biopolymers,* **20,** 833.

Menikh, A., Mickan, S.P., MacColl, R., Mannella, C. A., and Zhang, X, C-. (2002) Amplified detection of avidin-biotin binding using terahertz wave technology. *Elsevier Science*.

Michnik, A. (2003) Thermal stability of bovine serum albumin DSC study. *Journal of Thermal Analysis and Calorimetry*, **71**, 509–519.

Mickan, S. P., Menikh, A., Liu, H., Mannella, C. A., MacColl, R., Abbott, D., Munch, J., and Zhang, X-C. (2002) Label-free bioaffinity detection using terahertz technology. *Phys. Med. Biol*, **47,** 3789–3795.

Mine, S., Tate, S., Ueda, T., Kainosho Imoto, M., (1999) Analysis of the Relationship Between Enzyme Activity and its Internal Motion using Nuclear Magnetic Resonance: 15N Relaxation Studies of Wild-type and Mutant Lysozyme. *J. Mol. Biol,* **286**, 1547-1565.

Minkabc, J., Hajbabc, L., Mihálya, J., Németha, C., Pálmaia, M., Sandströmc, M., (2012) Vibrational Spectroscopic Studies of Molecules with Biochemical Interest: *The Cysteine Zwitterion,* **47**(6)**,** 415-483.

Minkov, V. S., Goryainov, S. V., Boldyreva, E. V., & Görbitz, C. H. (2010) Raman study of pressure-induced phase transitions in crystals of orthorhombic and

monoclinic polymorphs of L-cysteine: dynamics of the side chain. *Journal of Raman Spectroscopy*, **41**(12), 1748–1758.

Mittleman, D. M., Gupta, M., Neelamani, R., Baraniuk, R. G., Rudd, J. V., and Koch, M. (1999) Recent advances in terahertz imaging. *Appl. Phys. B,* **68**(6), 1085-1094.

Moggach, S. A., Allan, D. R., Morrison, C. A., Parsons, S., & Sawyer, L. (2005) Effect of pressure on the crystal structure of L-serine-I and the crystal structure of L-serine-II at 5.4 GPa. *Acta Crystallographica Section B: Structural Science*, **61**(1), 58–68.

Moggach, S. A., Clark, S. J., & Parsons, S. (2005) L-Cysteine-l at 30 K. *Acta Crystallographica Section E: Structure Reports Online*, **61**(8).

Mohamed, T. A., Shabaan, I. A., Zoghaib, W. M., Husband, J., Farag, R. S., & Alajhaz, A. E. N. M. A. (2009) Tautomerism, normal coordinate analysis, vibrational assignments, calculated IR, Raman and NMR spectra of adenine. *Journal of Molecular Structure*, **938**(1-3), 263–276.

Moon, Y. U., Curtis, R. A., Anderson, C. O., Blanch, H. W., & Prausnitz, J. M. (2000) Protein—Protein Interactions in Aqueous Ammonium Sulfate Solutions. Lysozyme and Bovine Serum Albumin (BSA). *Journal of Solution Chemistry*, **29**(8), 699–718.

Moulton, P. F. (1985) Spectroscopic and laser characteristics of Ti:$A1_2O_3$. *J. Opt. Soc. Am. B,* **3**(1), 125-133.

Mucignat-caretta, C., Caretta, A. (1997) Binding of two fluorescent cAMP analogues to type I and II regulatory subunits of cAMP-dependent protein kinases. *Biochimica et Biophysica acta*, **1357**, 81-90.

Mukhopadhyay, J., Sur, R., & Parrack, P. (1999) Functional roles of the two cyclic AMP-dependent forms of cyclic AMP receptor protein from Escherichia coli. *FEBS Lett*, **453**(1-2), 215–218.

Nagai, M., Yada, H., Arikawa, T., and Tanaka, K. (2006) Terahertz time-domain attenuated total reflection spectroscopy in water and biological solution. *Int. J. Infrared Millimet. Waves,* **27,** 505-515.

Nagel, M., Bolivar, P. H., Brucherseifer, M., Kurz, H., Bosserhoff, A., and Buttner, R. (2002a) Integrated THz technology for label-free genetic diagnostics. *Appl. Phys. Lett*, **80**(1), 154-156.

Nagel, M., Bolivar, P. H., Brucherseifer, M., Kurz, H., Bosserhoff, A., & Büttner, R. (2002b) Integrated planar terahertz resonators for femtomolar sensitivity label-free detection of DNA hybridization. *Applied Optics*, **41**(10), 2074–2078.

Nagel, M., F. Richter, et al. (2003) A functionalized THz sensor for marker-free DNA analysis. *Phys Med Biol,* **48**(22), 3625-36.

Nakai, S. (1996) Food Proteins: Properties and Characterization. *Wiley.*

Nakano, N. I., and Igarashi, S. J. (1970) Molecular interactions of pyrimidines, purines, and some compounds in aqueous media. *Biochemistry,* **9**, 577–583.

Narazaki, R., Maruyama, T., Otagiri, M. (1997) Probing the cysteine 34 residue in human serum albumin using fluorescence techniques. *Biochim. Biophys. Acta,* **1138**, 275–281.

Navarrete, J. T. L., Hernandez, V., Ramirez, F. J. (1994) vibrational-spectra of [N-15] glutamic-acid and [H-2(4)] glutamic-acid. *Journal of Raman spectroscopy*, **25**, 861-867.

Newmark, R. A., Cantor, C. R. (1968) Nuclear magnetic resonance study of the interactions of guanosine and cytidine in dimethyl sulfoxide. *J Am Chem Soc,* **90**(18), 5010–5017.

Nielsen, O. F., Johansson, C., Christensen, D. H., Hvidt, S., Flink, J., Hoime Hansen, S., & Poulsen, F. (2000) Collective vibrational effects in hydrogen bonded liquid

amides and proteins studied by isotopic substitution. In *Journal of Molecular Structure,* **552**, 71–80.

Nishimura, Y., Tsuboi, M., Sato, T., and Aoki, K. (1986) Conformation-sensitive Raman lines of mononucleotides and their use in a structure analysis of polynucleotides: guanine and cytosine nucleotides. *J. Mol. Struct,* **146,** 123–153.

Nishizawa, J. I., Sasaki, T., Suto, K., Tanabe, T., Saito, K., Yamada, T., & Kimura, T. (2005) THz transmittance measurements of nucleobases and related molecules in the 0.4- to 5.8-THz region using a GaP THz wave generator. *Optics Communications*, **246**(1-3), 229–239.

Nishizawa, S., Sakai, K., Hangyo, M., Nagashima, T., Takeda, M.W., Tominaga, K., Oka, A., Tanaka, K ., and Morikawa, O. (2005) Terahertz Time-Domain Spectroscopy. *Topics Appl. Phys*, **97**, 203–271.

Nitta, K., & Sugai, S. (1989) The evolution of lysozyme and alpha-lactalbumin. *European Journal of Biochemistry / FEBS*, **182**(1), 111–118.

Nowak, M. J. (1989) IR matrix-isolation studies of nucleic acid constituents – The spectrum of monomeric thymine. *J. Molec. Struct,* **49,** 193:35.

Oakes, J. (1976) Thermally Denatured Proteins. *J. Chem. Soc. Faraday,* **I72**, 228-237.

Oberg, K. A., and Uversky, V. N. (2001) Secondary Structure of the Homologous Proteins, alpha-Fetoprotein and Serum Albumin, from their Circular Dichroism and Infrared Spectra. *Protein and Peptitide Letters***, 8**(4), 297-302.

Okazaki, A., Ikura, T., Nikaido, K., & Kuwajima, K. (1994) The chaperonin GroEL does not recognize apo-alpha-lactalbumin in the molten globule state. *Nature Structural Biology*, **1**(7), 439–446.

Olshevskaya, J. S., Ratushnyak, A. S., Petrov, A. K., Kozlov, A. S., Zapara, T. A. (2008) Effect of Terahertz Electromagnetic Waves on Neurons Syst*ems. IEEE region 8 sibircon,* **1,** 210-211.

Ozaki, H., and McLaughlin, L. W. (1992) The estimation of distances between specific backbone-labeled sites in DNA using fluorescence resonance energy transfer. *Nucleic Acids Res,* **20**(19), 5205–5214.

Ozeki, K., Sakabe, N.,Tanaka, J. (1969) The crystal structure of thymine. *Acta Cryst,* **B25**, 1038-1045.

Paciaroni, A., Bizzarri, A., & Cannistraro, S. (1998) Molecular-dynamics simulation evidences of a boson peak in protein hydration water. *Physical Review E*, **57**(6), R6277–R6280.

Padro, J. A., and Marti, J. (2003) An interpretation of the low-frequency spectrum of liquid water*, J. Chem. Phys,* **118**, 452-453.

Pal, S. K., Peon, J., & Zewail, A. H. (2002) Biological water at the protein surface: dynamical solvation probed directly with femtosecond resolution. *Proceedings of the National Academy of Sciences of the United States of America*, **99**(4), 1763–1768.

Palese, S., Schilling, L., Miller, D. R. J., Staver, R. P., Lotshaw, W. T. (1994) Femtosecond optical Kerr effect studies of water. *J. Phys. Chem*, **98**(25), 6308–6316.

Palmer, A. G., Cavanagh, J, Wright, P. E., Rance, M. (1991) Sensitivity improvement in proton-detected 2-dimensional heteronuclear correlation NMR spectroscopy. *J. Magn. Reson,* **93,** 1, 151-170.

Paolantoni, M., Cornez, L., Gallina, M. E., Sassi, P., Scarponi, F., Fioretto, D., & Morresi, A. (2009) Light scattering spectra of water in trehalose aqueous solutions: Evidence for two different solvent relaxation processes. *Journal of Physical Chemistry B*, **113**(22), 7874–7878.

Parker, S. F. (2013) Assignment of the vibrational spectrum of l-cysteine. *Chemical Physics*, **424**, 75–79.

Parthasarathy, R., Globus, T., Khromova, T., Swami, N., & Woolard, D. (2005) Dielectric properties of biological molecules in the Terahertz gap. *Applied Physics Letters*, **87**(11), 113901.

Passner, J. M., Schultz, S. C., & Steitz, T. A. (2000) Modeling the cAMP-induced allosteric transition using the crystal structure of CAP-cAMP at 2.1 A resolution. *Journal of Molecular Biology*, **304**(5), 847–859.

Pawlukojc, A., Bobrowicz, L., Natkaniec, I. (1995) The IINS spectroscopy of amino acids: l- and dl-valine. *Spectrochim. Acta,* **A51**, 303.

Pawlukojć, A., Hołderna-Natkaniec, K., Bator, G., & Natkaniec, I. (2014) INS, IR, RAMAN, 1H NMR and DFT investigations on dynamical properties of l-asparagine. *Vibrational Spectroscopy*, **72**, 1–7.

Pawlukojć, A., Leciejewicz, J., Tomkinson, J., & Parker, S. F. (2002) Neutron spectroscopic study of hydrogen bonding dynamics in l-serine. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, **58**(13), 2897–2904.

Pellegrini, A., Thomas, U., Bramaz, N., Hunziker, P., & Von Fellenberg, R. (1999) Isolation and identification of three bactericidal domains in the bovine α-lactalbumin molecule. *Biochimica et Biophysica Acta - General Subjects*, **1426**(3), 439–448.

Permyakov, E. A, & Berliner, L. J. (2000) alpha-Lactalbumin: structure and function. *FEBS Letters*, **473**(3), 269–274.

Permyakov, E. A., Kreimer, D. I., Kalinichenko, L. P., & Shnyrov, V. L. (1988) Interactions of calcium binding proteins, parvalbumin and alpha-lactalbumin, with

dipalmitoylphosphatidylcholine vesicles. *General Physiology and Biophysics*, **7**(1), 95–107.

Permyakov, E. A., Morozova, L. A., & Burstein, E. A. (1985) Cation binding effects on the pH, thermal and urea denaturation transitions in alpha-lactalbumin. *Biophysical Chemistry*, **21**(1), 21–31.

Permyakov, E. A., Reyzer, I. L., & Berliner, L. J. (1993) Effects of Zn(II) on galactosyltransferase activity. *Journal of Protein Chemistry*, **12**(5), 633–638.

Perticaroli, S., Comez, L., Paolantoni, M., Sassi, P., Lupi, L., Fioretto, D., Morresi, A. (2010) Broadband depolarized light scattering study of diluted protein aqueous solutions. *Journal of Physical Chemistry B*, **114**(24), 8262–8269.

Peters, T. (1985) Serum albumin. *Advances in Protein Chemistry*, **37**, 161–245.

Petit, C. M., Zhang, J., Sapienza, P. J., Fuentes, E. J., & Lee, A. L. (2009) Hidden dynamic allostery in a PDZ domain. *Proceedings of the National Academy of Sciences of the United States of America*, **106**(43), 18249–18254.

Phillips D.C. (1967) The Hen Egg-White Lysozyme Molecule. *Proc Natl Acad Sci USA,* **57**(3), 483–495.

Piesiewicz, R., Jansen, C., Wietzke, S., Mittleman, D., Koch, M., Kurner, T. (2007) Properties of Building and Plastic Materials in the THz Range. *Int. J. Infared. Milli Waves,* **28**, 363-371.

Plusquellic, D. F., Siegrist, K., Heilweil, E. J., Esenturk, O. (2007) Applications of Terahertz Spectroscopy in Biosystems. *Chem Phys Chem*, **8,** 2412 – 2431.

Pohorille, A., Pratt, L. R., Burt, S. K., & MacElroy, R. D. (1984) Solution influence on biomolecular equilibria: nucleic acid base associations. *J Bio Mol Struc Dynam*, **1**(5), 1257–80.

Polit, A., Blaszczyk, U., & Wasylewski, Z. (2003) Steady-state and time-resolved fluorescence studies of conformational changes induced by cyclic AMP and DNA binding to cyclic AMP receptor protein from Escherichia coli. *European Journal of Biochemistry*, **270**(7), 1413–1423.

Pompella, A., Visvikis, A., Paolicchi, A., Tata, V., Casini, A.F. (2003) The changing faces of glutathione, a cellular protagonist. *Biochemical Pharmacology,* **66**(8), 1499–503.

Poole, S., West, S. I., and Fry, J. C. (1987) Effects of basic proteins on the denaturation and heat-gelation of acidic proteins. *Food Hydrocolloids*, **1,** 301-316.

Popovych, N., Tzeng, S.-R., Tonelli, M., Ebright, R. H., & Kalodimos, C. G. (2009) Structural basis for cAMP-mediated allosteric control of the catabolite activator protein. *Proceedings of the National Academy of Sciences of the United States of America*, **106**(17), 6927–6932.

Powell, J. W., G. S. Edwards, et al. (1987). Investigation of far-infrared vibrational modes in polynucleotides. *Phys Rev A,* **35**(9), 3929-3939.

Privalov, P. L., & Dragan, A. I. (2015) Microcalorimetry of biological macromolecules. *Biophysical Chemistry*, **126**(1-3), 16–24.

Quesada-Moreno, M. M., Márquez-García, A. Á., Avilés-Moreno, J. R., & López-González, J. J. (2013) Conformational landscape of l-threonine in neutral, acid and basic solutions from vibrational circular dichroism spectroscopy and quantum chemical calculations. *Tetrahedron Asymmetry*, **24**(24), 1537–1547.

Ramirez, F. J. and Lopez Navarrete, J. T. (1995) Force-Field and Normal-Coordinate Calculations. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy,* **51**(2)**,** 293-302.

Raszka, M. and Kaplan, N.O. (1972) Association by hydrogen bonding of mononucleotides in aqueous solution. *Proc Natl Acad Sci USA,* **69**, 2025–2029.

Redfield, C., Schulman, B. A., Milhollen, M. A., Kim, P. S., & Dobson, C. M. (1999) Alpha-lactalbumin forms a compact molten globule in the absence of disulfide bonds. *Nature Structural Biology*, **6**(10), 948–952.

Redinha, J. S., Jesus, A. J. L., Pais, A. A. C. C., & Almeida, J. A. S. (2013) Crystallization : From the Conformer to the Crystal. *Advanced Topics on Crystal Growth*, 201–225.

Reed, R. G., Feldhoff, R. C., Clute, O. L., & Peters, T. (1975) Fragments of bovine serum albumin produced by limited proteolysis. Conformation and ligand binding. *Biochemistry*, **14**(21), 4578–4583.

Reid, G. D., Wynne, K. (2000) Ultrafast Laser Technology and Spectroscopy. *Encyclopedia of Analytical Chemistry R.A. Meyers (Ed.)*, **644**, 13670.

Relkin, P., Eynard, L., & Launay, B. (1992) Thermodynamic parameters of b-lactoglobulin and a-lactalbumin. A DSC study of denaturation by heating. *Thermochimica Acta*, **204**, 111–121.

Reva, I. D., Stepanian, S. G., Plokhotnichenko, A. M., Radchenko, E. D., Sheina, G. G., Blagoi, Yu. P. (1994) Infrared Matrix Isolation Studies of Amino Acids. Molecular Structure of Proline. *Journal of Molecular Structure*, **318**, 1-13.

Rezus, Y. L. A., & Bakker, H. J. (2007) Observation of immobilized water molecules around hydrophobic groups. *Physical Review Letters*, **99**(14). 8301.

Riahi, S., Eynollahi, S., Ganjali, M. R., & Norouzi, P. (2010) Electronic and Geometry Studies on Swainsonine , DNA Base Pairs and its Complex. *International Journal of electrochemical science*, **5**, 355–366.

Roca, M., Moliner, V., Tuñón, I., & Hynes, J. T. (2006) Coupling between protein and reaction dynamics in enzymatic processes: Application of Grote-Hynes theory to catechol O-methyltransferase. *Journal of the American Chemical Society*, **128**(18), 6186–6193.

Roelands, C. P. M., ter Horst, J. H., Kramer, H. J. M., Jansens, P. J. (2007) Precipitation mechanism of stable and metastable polymorphs of L-glutamic acid. *AI Ch E Journal,* **53**, 354–362.

Roh, J. H., Curtis, J. E., Azzam, S., Novikov, V. N., Peral, I., Chowdhuri, Z., Sokolov, A. P. (2006) Influence of hydration on the dynamics of lysozyme. *Biophysical Journal*, **91**(7), 2573–2588.

Rønne, C., Thrane, L., Åstrand, P. O., Wallqvist, A., Mikkelsen, K. V, & Keiding, S. ren R. (1997) Investigation of the temperature dependence of dielectric relaxation in liquid water by THz reflection spectroscopy and molecular dynamics simulation. *Journal of Chemical Physics*, **107**, 5319–5331.

Ross, P. D., & Shrake, a. (1988) Decrease in stability of human albumin with increase in protein concentration. *Journal of Biological Chemistry*, **263**(23), 11196–11202.

Rozenberg, M., Shoham, G., Reva, I., & Fausto, R. (2004) Low temperature FTIR spectroscopy and hydrogen bonding in cytosine polycrystals. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, **60**(1-2), 463–470.

Rozenberg, M., Shoham, G., Reva, I., Fausto, R. (2004) Low temperature FTIR spectroscopy and hydrogen bonding in cytosine polycrystals. *Spectrochimica Acta Part A*, **60**(1-2), 463-470.

Ruegg, M., Moor, U., and Blanc, B. (1977) A calorimetric study of the thermal denaturation of whey proteins in simulated milk ultrafiltrate. *J. Dairy Res*, **44,** 509-520.

Rungsawang, R., Ueno, Y., Tomita, I., & Ajito, K. (2006) Angle-dependent terahertz time-domain spectroscopy of amino acid single crystals. *The Journal of Physical Chemistry. B*, **110**(42), 21259–21263.

Rush, T., Peticolas, W. L. (1995) Ab initio transform calculation of resonance raman spectra of uracil, 1-methyluracil, and 5-methyluracil. *J. Phys. Chem,* **99**, 14647–14658.

Russo, D., Hura, G., & Head-Gordon, T. (2004) Hydration dynamics near a model protein surface. *Biophysical Journal*, **86**(3), 1852–1862.

Rutz, F., Wilkl, R., Kleine-Ostmannl, T., Grunenberg, J., Koch, M. (2005) THz vibrational spectra of selected tripeptides: experiments and theory. *Infrared and Millimeter Waves*, **12,** 737-738.

Sagi, Y., Khan, S., & Eisenbach, M. (2003) Binding of the chemotaxis response regulator CheY to the isolated, intact switch complex of the bacterial flagellar motor. Lack of cooperativity. *Journal of Biological Chemistry*, **278**(28), 25867–25871.

Sambrook, J., & Russell. D. W. (2001) Molecular Cloning, a Laboratory Manual. *Cold Spring Harbor Laboratory Press*, Cold Spring Harbor, NY

Sánchez-Cortés, S., Molina, A. M., García-Ramos, J. V., Carmona, P. (1991) Interactions of cytidine derivatives with metals as revealed by Surface Enhanced Raman Spectroscopy. *Journal of Raman Spectroscopy*, **22**, 819-824.

Sands, W. A., Woolson, H. D., Milne, G. R., Rutherford, C., Palmer, T. M. (2006) Exchange protein activated by cyclic AMP (Epac)-mediated induction of suppressor of cytokine signalling 3 (SOCS-3) in vascular endothelial cells. *Mol Cell Biol,* **26**, 6333–6346.

Sarkar, M., Dornberger, U., Rozners, E., Fritzsche, H., Strömbergand, R., Gräslund, A. (1997) FTIR Spectroscopic Studies of Oligonucleotides That Model a Triple-

Helical Domain in Self-Splicing Group I Introns. *Biochemistry,* **36**(49), 15463–15471.

Sassi, P., Perticaroli, S., Comez, L., Giugliarelli, A., Paolantoni, M., Fioretto, D., & Morresi, A. (2013) Volume properties and spectroscopy: A terahertz Raman investigation of hen egg white lysozyme. *The Journal of Chemical Physics*, **139**(22), 225101.

Saxena, V. K., Dorfman, B. H., Van Zandt, L. L. (1991) Identifying and interpreting spectral features of dissolved poly(dA)-poly(dT) DNA polymer in the high-microwave range. *Phys.* Rev, **A43**, 4510.

Schall, M., Walther, M., & Uhd Jepsen, P. (2001) Fundamental and second-order phonon processes in CdTe and ZnTe. *Physical Review B*, **64**, 094301.

Schleucher, J., Schwendinger, M., Sattler, M., Schmidt, P., Schedletzky, O., Glaser, S. J., Sørensen, O. W., Griesinger, C. J. (1994) A general enhancement scheme in heteronuclear multidimensional NMR employing pulsed field gradients. *Biomol NMR*, **4(2**), 301-6.

Schmidt, M., Sand, C., Jakobs, K. H., Michel, M. C., Weernink, P. A. (2007) Epac and the cardiovascular system. *Curr Opin Pharmacol,* **7**, 193–200.

Schulman, B. A., Redfield, C., Peng, Z. Y., Dobson, C. M., & Kim, P. S. (1995) Different subdomains are most protected from hydrogen exchange in the molten globule and native states of human alpha-lactalbumin. *Journal of Molecular Biology*, **253**(5), 651–657.

Schweizer, M. P., Broom, A. D., Ts'o, P. O. P., and Hollis, D. P. (1968) Studies of inter- and intramolecular interaction in monoucleotides by proton magnetic resonance. *J. Am. Chem. Soc,* **90**, 1042–1055.

Schwenke, K. D. (1996) Food Proteins. Properties and Characterization. *Food/Nahrung*, **40**(6), 348.

Selkoe, D. J. (2004) Cell biology of protein misfolding: the examples of Alzheimer's and Parkinson's diseases. *Nature Cell Biology*, **6**(11), 1054–1061.

Sequeira, A., Rajagopal, H., Chidambaram, R. (1972) A neutron diffraction study of the structure of L-glutamic acid.HCl. *Acta Cryst,* **B28,** 2514-2519.

Sessions, R. B., Dauber-Osguthorpe, P.,Osguthorpe, D. J. (1989) Filtering molecular dynamics trajectories to reveal low-frequency collective motions: phospholipase A2. *J Mol Biol,* **210**, 617-633.

Shabb, J. B., & Corbin, J. D. (1992) Cyclic nucleotide-binding domains in proteins having diverse functions. *Journal of Biological Chemistry*, **267**(9), 5723–5726.

Shanmugasundaram, M., & Puranik, M. (2009) Computational prediction of vibrational spectra of normal and modified DNA nucleobases. *Journal of Raman Spectroscopy*, **40**(12), 1726–1748.

Shen, Y., Watanabe, T., Arena, D. A., Kao, C. C., Murphy, J. B., Tsang, T. Y., Carr, G. L. (2007) Nonlinear cross-phase modulation with intense single-cycle terahertz pulses. *Physical Review Letters*, **99**(4), 1–4.

Shi, W., Y. J. Ding, et al. (2002) Efficient, tunable, and coherent 0.18-5.27-THz source based on GaSe crystal. *Opt Lett,* **27**(16), 1454-6.

Shi, Y., & Wang, L. (2005) Collective vibrational spectra of α- and γ-glycine studied by terahertz and Raman spectroscopy. *Journal of Physics D: Applied Physics*, **38**(19), 3741–3745.

Shirota, H., & Castner, E. W. (2005) Physical properties and intermolecular dynamics of an ionic liquid compared with its isoelectronic neutral binary solution. *Journal of Physical Chemistry A*, **109**(42), 9388–9392.

Shishkin, O. V., Pelmenschikov, A., Hovorun, D. M., Leszczynski. (2000) Theoretical analysis of low-lying vibrational modes of free canonical 2-deoxyribonucleosides. *Chem. Phys,* **260**, 317.

Shotts, W. J., Sievers, A. J. (1974) The far-infrared properties of polyamino acids. *Biopolymers*, **13,** 2593-2614.

Shoup, R. R, Miles, H. T, Becker, E. D. (1966) NMR evidence of specific base-pairing between purines and pyrimidines. *Biochem Biophys Res Commun,* **23**(2), 194–201.

Siddle, K., Hutton, J. C. (1990) Peptide Hormone Action: A Practical Approach. *IRL Press*.

Simonovic, M., & Volz, K. (2001) A Distinct Meta-active Conformation in the 1.1-A Resolution Structure of Wild-type ApoCheY. *Journal of Biological Chemistry*, **276**(31), 28637–28640.

Sjoholm, I., and Ljungstedt, I. (1973) Studies on the Tryptophan and Drug-binding Properties of Human Serum Albumin Fragments by affinity chromatography and circular dichroism measurements. *J. Biol. Chem*, **248**, 8434-8441.

Smith, N. A., & Meech, S. R. (2002) Optically-heterodyne-detected optical Kerr effect (OHD-OKE): Applications in condensed phase dynamics. *International Reviews in Physical Chemistry*, **21**(1), 75–100.

Solie, T. N., & Schellman, J. A. (1968) The interaction of nucleosides in aqueous solution. *Journal of Molecular Biology*, **33**(1), 61–77.

Sophianopoulos, A. J. (1969) Association sites of lysozyme in solution. I. The active site. *J Biol Chem*, **244**(12), 3188–3193.

Stanciuc, N. S. T. Ă., & Râpeanu, G. (2010) An overview of bovine α-lactalbumin structure and functionality. *Food Technology*, **34**(2), 82–93.

Stepanian, S. G., Reva, I. D., Radchenko, E. D., Rosado, M. T. S., Duarte, M. L. T. S., Fausto, R., Adamowicz, L. (1998) Matrix-Isolation infrared and theoretical studies of the glycine conformers. *J. Phys. Chem.* **102.** 1041-1054.

Sterpone, F., Stirnemann, G., Hynes, J. T., & Laage, D. (2010) Water hydrogen-bond dynamics around amino acids: The key role of hydrophilic hydrogen-bond acceptor groups. *Journal of Physical Chemistry B*, **114**(5), 2083–2089.

Stirnemann, G., Hynes, J. T., & Laage, D. (2010) Water hydrogen bond dynamics in aqueous solutions of amphiphiles. *Journal of Physical Chemistry B*, **114**(8), 3052–3059.

Stock, A. M., Martinez-Hackert, E., Rasmussen, B. F., West, A. H., Stock, J. B., Ringe, D., & Petsko, G. A. (1993) Structure of the Mg^{2+}-bound form of CheY and mechanism of phosphoryl transfer in bacterial chemotaxis. *Biochemistry*, **32**(49), 13375–13380.

Streetman, B., and Banerjee, S. (2000) Solid State Electronic Devices. *Prentice Hall. 5th ed.*

Suresh, S., Padmanabhan, S., Vijayan, M. (1994) X-ray studies on crystalline complexes involving amino acids and peptides. XXVII. Effect of chirality, specific interactions and characteristic aggregation patterns in the structures of arginine and its complexes with formic acid. *J Biomol Struct Dyn,* **11**(6), 1425-35.

Susi, H., Ard, J. S., & Purcell, J. M. (1973) Vibrational spectra of nucleic acid constituents—II Planar vibrations of cytosine. *Spectrochimica Acta Part A: Molecular Spectroscopy*, **29**(4), 725–733.

Susi, H., Byler, D. M., & Gerasimowicz, V. W. (1983) vibrational analysis of amino acids: cysteine , serine , beta-chloroalanin. *Journal of Molecular Structure*, **102**, 63–79.

Suzuki, S., Ohshima, T., Tamiya, N., Fukushima, K., Shimanouchi, T., Mizushima, S., (1959) Infrared spectra of deuterated α-amino acids $NH_3^+CDRCOO$: Assignment of the absorption bands of α-alanine. *Spectrochimica Acta,* **11**, 969-976.

Szczesniak, M., Szczepaniak, K., Kwiatkowski, J. S., KuBulat, K., & Person, W. B. (1988) Matrix isolation infrared studies of nucleic acid constituents. 5. Experimental matrix-isolation and theoretical ab initio SCF molecular orbital studies of the infrared spectra of cytosine monomers. *J. Am. Chem. Soc*, **110**(1), 8319–8330.

Taday, P. F., Bradley, I. V., & Arnone, D. D. (2003) Terahertz pulse spectroscopy of biological materials: L-glutamic acid. *Journal of Biological Physics*, **29**(2-3), 109–115.

Takahashi, M., Blazy, B., Baudras, A., & Hillen, W. (1989) Ligand-modulated binding of a gene regulatory protein to DNA. Quantitative analysis of cyclic-AMP induced binding of CRP from Escherichia coli to non-specific and specific DNA targets. *Journal of Molecular Biology*, **207**(4), 783–796.

Thomas, C. M, & Summers, D. (2008) Bacterial Plasmids. *Encyclopedia of Life Sciences*

Thomas, J. G. J., Tsuboi, M. (1993) Advances in Biophysical Chemistry, *Jai Press,* **vol. 3**, 1.

Thomas, S. A., Brewster, J. A., & Bourret, R. B. (2008) Two variable active site residues modulate response regulator phosphoryl group stability. *Molecular Microbiology*, **69**(2), 453–465.

Thrane, L., Jacobsen, R. H., Uhd Jepsen, P., Keiding, S. R. (1995) THz Reflection Spectroscopy of Liquid Water. *Chem. Phys. Letts,* **240,** 330-333.

Tonouchi, M. (2007) Cutting-edge terahertz technology. *Nature photonics*, **1**, 97-105.

Torre, R., Bartolini, P., Pick, R. M. (1998) Time-resolved optical Kerr effect in a fragile glass-forming liquid. *Phys. Rev. E*, **57**, 1912.

Trivella, A., Gaillard, T., Stote, R. H., & Hellwig, P. (2010) Far infrared spectra of solid state aliphatic amino acids in different protonation states. *Journal of Chemical Physics*, **132**(11), 1–10.

Troullier, N., & Martins, J. L. (1991) Efficient pseudopotentials for plane-wave calculations. II. Operators for fast iterative diagonalization. *Physical Review B*, **43**(11), 8861–8869.

Ts'o, P.O.P. (1974) Basic Principles in Nucleic Acid Chemistry. *Academic Press.*

Tsai, K. H., & Wu, T. M. (2006) Local structural effects on low-frequency vibrational spectrum of liquid water: The instantaneous-normal-mode analysis. *Chemical Physics Letters*, **417**(4-6), 389–394.

Tsuboi, M., Nishimura, Y., Hirakawa, A. Y., Peticolasin, W. L., Spiro, T. G. (1987) Biological Applications of Raman Spectroscopy. *John Wiley,* **2,** 3-109.

Turton, D., Harwood T., Lapthorn A., Ellis E., Wynne K., (2013) Ultrabroadband terahertz spectroscopies of biomolecules and water. *SPIE,* **8623**, 862303-1-7.

Turton, D. A., Senn, H. M., Harwood, T., Lapthorn, A. J., Ellis, E. M., & Wynne, K. (2014) Terahertz underdamped vibrational motion governs protein-ligand binding in solution. *Nature Communications*, **5**, 3999.

Turton, D. A., Sonnleitner, T., Ortner, A., Walther, M., Hefter, G., Seddon, K. R., Wynne, K. (2012) Structure and dynamics in protic ionic liquids: A combined optical Kerr-effect and dielectric relaxation spectroscopy study. *Faraday Discussions*, **154**, 145-153.

Tutar, Y. (2008a) Chemical linkage at allosteric activation of E. coli cAMP receptor protein. *Protein Journal*, **27**(1), 21–29.

Tutar, Y. (2008b) Syn, anti, and finally both conformations of cyclic AMP are involved in the CRP-dependent transcription initiation mechanism in E. coli lac operon. *Cell Biochemistry and Function*, **26**(4), 399-405.

Tzeng, S.-R., & Kalodimos, C. G. (2009) Dynamic activation of an allosteric regulatory protein. *Nature*, **462**(7271), 368–372.

Tzeng, S.-R., & Kalodimos, C. G. (2012) Protein activity regulation by conformational entropy. *Nature*, **488**, 236-240.

Tzeng, S.-R., & Kalodimos, C. G. (2013) Allosteric inhibition through suppression of transient conformational states. *Nature Chemical Biology*, **9**(7), 462–5.

Ueno, Y., Ajito, K. (2008) Analytical Terahertz Spectroscopy. *Analytical Sciences*, **24,** 185-191.

Urabe, H., Hayashi, H., Tominaga, Y., Nishimura, Y., Kubota, K., Tsuboi, M. (1985) Collective vibrational modes in molecular assembly of DNA and its application to biological systems-low frequency Raman spectroscopy. *J. Chem. Phys*, **82,** 531-5.

Urabe, H., Sugawara, Y., Tsukakoshi, M., Ikegami, A., Iwasaki, H., Kasuya, Y. (1987) Raman spectroscopic study on low-frequency collective modes in self-associates of guanosine monophosphates. *Biopolymers,* **26**(6)**,** 63–971.

Urabe, H., Tominaga, Y. (1982) Low-lying collective modes of DNA double helix by Raman spectroscopy. *Biopolymers,* **21**(12), 2477–81.

Urabe, H., Sugawara, Y., Ataka, M., & Rupprecht, A. (1998) Low-frequency Raman spectra of lysozyme crystals and oriented DNA films: dynamics of crystal water. *Biophysical Journal*, **74**(3), 1533–1540.

Valeur, B. (2001) Molecular Fluorescence - Principles and Applications. *Methods,* **(8)**.

Van Vlijmen, H. W. T. and Karplus, M. (1999) Analysis of calculated normal modes of a set of native and partially unfolded proteins. *J. Phys. Chem*, **B103,** 3009–3021.

Van Wart, H. E., Lewis, A., Scheraga, H. A., Saeva, F. D. (1973) Disulfide bond dihedral angles from Raman spectroscopy. *Proc Natl Acad Sci USA*, **70**(9), 2619-23.

Van Weeren, P. C., De Bruyn, K. M. T., De Vries-Smits, A. M. M., Van Lint, J., & Burgering, B. M. T. (1998) Essential role for protein kinase B (PKB) in insulin-induced glycogen synthase kinase 3 inactivation. Characterization of dominant-negative mutant of PKB. *Journal of Biological Chemistry*, **273**(21), 13150–13156.

Van Zandt, L. L. and Saxena. V. K. (1989) Millimeter-microwave spectrum of DNA: Six predictions for spectroscopy. *Phys Rev A,* **39**(5), 2672-2674.

Van Zandt, L. L., and Saxena, V. K. (1992) Structure & Functions, Volume 1: Nucleic Acids. *New York.*

Vanamerongen, H., Vangrondelle, R. (1995) Transient Absorption-Spectroscopy in Study of Processes and Dynamics in Biology. *Biochemical Spectroscopy*, **246**, 201-226.

Vargas-Luna, M., and Huerta-Franco, R. (2008) Biomedical Applications of Terahertz Spectroscopy: A Brief Review. *MEDICAL PHYSICS: Tenth Mexican Symposium on Medical Physics,* **1032**, 79-82.

Veprintsev, D. B., Permyakov, S. E., Permyakov, E. A., Rogov, V. V., Cawthern, K. M., & Berliner, L. J. (1997) Cooperative thermal transitions of bovine and human apo-α-lactalbumins: Evidence for a new intermediate state. *FEBS Letters*, **412**(3), 625–628.

Verbist, J. J., Lehmann, M. S., Koetzle, T. F., Hamilton, W. C. (1972) Precision neutron diffraction structure determination of protein and nucleic acid components. VI. The crystal and molecular structure of the amino acid L-asparagine monohydrate. *Acta Cryst***, B28,** 3006-3013.

Wallace, V. P., Fitzgerald, A. J., et al. (2004). Terahertz pulsed imaging of basal cell carcinoma ex vivo and in vivo. *Br J Dermatol,* **151**(2), 424-32.

Wallace, V. P., Wallace, V. P., Fitzgerald, A. J., Fitzgerald, A. J., Shankar, S., Shankar, S., Arnone, D. D. (2004) Dermatological Surgery and Lasers: Terahertz Pulsed Imaging of Basal Cell Carcinoma Ex Vivo and in Vivo. *British Journal of Dermatology*, **151**, 424–432.

Walther, M., Fischer, B. M., & Jepsen, P. U. (2003) Noncovalent intermolecular forces in polycrystalline and amorphous saccharides in the far infrared. *Chemical Physics*, **288**(2-3), 261–268.

Walther, M., Fischer, B., Schall, M., Helm, H., Uhd Jepsen, P. (2000) Far-infrared vibrational spectra of all-trans, 9-cis and 13-cis retinal measured by THz time-domain spectroscopy. *Chem. Phys. Lett,* **332**, 38-395.

Walther, M., Plochocka, P., Fischer, B., Helm, H., Jepsen, P. U. (2002) Collective Vibrational Modes in Biological Molecules Investigated by Terahertz Time-Domain Spectroscopy. *Biopolymers (Biospectroscopy),* **67**, 310.

Wang, C. H., & Storms, R. D. (1971) Temperature-Dependent Raman Study and Molecular Motion in L-Alanine Single Crystal. *J. Chem. Phys*, **55**, 3291-1971.

Wang, W., Yan, H., Yue, W., Zhao, G., Zhang, C., Liu, H., & Zhang, X. (2005) THz spectrum of reduced glutathione. *Science in China*, **48**(5), 585–592.

Watanabe, Y., Kawase, K., Ikari, T., Ito, H., Ishikawa, Y., and Minamide, H. (2003) Component spatial pattern analysis of chemicals using terahertz spectroscopic imaging. *Appl. Phys. Lett,* **83**, 800-802.

Watson, D. G., Sutor, D. J., Tollin, P. (1965) The crystal structure of deoxyadenosine monohydrate. Acta Crystallogr, **10**, 111-124.

Wayne, L. G., Brenner, D. J., Colwell, R. R., Grimont, P. A. D., Kandler, O., Krichevsky, M. I., Moore, L. H., Moore, W. E. C., Murray, R. G. E., Stackebrandt, E., Starr, M. P., Truper, H. G. (1987) Report of the Ad Hoc Committee on Reconciliation of Approaches to Bacterial Systematics. *Int J Syst Evol Microbiol,* **37,** 463-464.

Weber, I. T., & Steitz, T. A. (1987) Structure of a complex of catabolite gene activator protein and cyclic AMP refined at 2.5 A resolution. *Journal of Molecular Biology*, **198**(2), 311–326.

Weidlich, T., Lindsay, S. M., Rui, Q., Rupprecht, A., Peticolas, W. L., Thomas, G. A. (1990) A Raman study of low frequency intrahelical modes in A-, B-, and C-DNA. *J Biomol Struct Dyn*, **1**(1), 139-71.

Weining, W., Haitao, Y., Weiwei, Y., Guozhong, Z., Cunlin, Z., Haibo, L., Xicheng, Z. (2005) THz spectrum of reduced glutathione. *Physics and Astronomy*, **48**(5), 585-592.

Weiss-Lopez, B. E., Goodrow, M. H., Musker, W. K., Nash, C. P. (1986) Conformational dependence of the disulfide stretching frequency in cyclic model compounds. *J. Am. Chem. Soc,* **108** (6), 1271–1274.

Wetzel, R., Becker, M., Behlke, J., Billwitz, H., Böhm, S., Ebert, B., Lassmann, G. (1980) Temperature behaviour of human serum albumin. *European Journal of Biochemistry/FEBS*, **104**(2), 469–478.

Whitmire, S. E., Wolpert, D., Markelz, A. G., Hillebrecht, J. R., Galan, J., & Birge, R. R. (2003) Protein flexibility and conformational state: a comparison of collective vibrational modes of wild-type and D96N bacteriorhodopsin. *Biophysical Journal*, **85**(2), 1269–1277.

Wilson, E. B., & Decius, J. C. (1980) Molecular Vibrations: the Theory of Infrared and Raman Vibrational Spectra. *Dover: New York*.

Winer, K., & Cardona, M. (1987) Two-phonon frequency difference absorption in Si and Ge. *Solid state communications,* **64,** 1461-1464.

Wittlin, A., Genzel, L., Kremer, F., Haseler, S., and Poglitsch, A. (1986) Far-infrared spectroscopy on oriented films of dry and hydrated DNA. *Phys. Rev,* **34** (1), 493-500.

Woestenenk, E. A., Hammarström, M., Van Den Berg, S., Härd, T., & Berglund, H. (2004) His tag effect on solubility of human proteins produced in Escherichia coli: A

comparison between four expression vectors. *Journal of Structural and Functional Genomics*, **5**(3), 217–229.

Wong, H. C., Mao, J., Nguyen, J. T., Srinivas, S., Zhang, W., Liu, B., Li, L., Wu, D., and Zheng, J. (2000) Structural basis of the recognition of the Dishevelled DEP domain in the Wnt signalling pathway. *Nature America,* **7,** 1178-1184.

Woolard, D. L., Globus, T. R., Gelmont, B. L., Bykhovskaia, M., Samuels, a. C., Cookmeyer, D., Loerop, W. R. (2002) Submillimeter-wave phonon modes in DNA macromolecules. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, **65**(5), 051903.

Wu L. C., & Kim P. S. (1998) A specific hydrophobic core in the α-lactalbumin molten globule. *J. Mol. Biol,* **280,** 175–182.

Xiao, D., Rajian, J. R., Li, S., Bartsch, R. A., & Quitevis, E. L. (2006) Additivity in the optical Kerr effect spectra of binary ionic liquid mixtures: Implications for nanostructural organization. *Journal of Physical Chemistry B*, **110**(33), 16174–16178.

Xie, A., He, Q., Miller, L., Sclavi, B., & Chance, M. R. (1999) Low frequency vibrations of amino acid homopolymers observed by synchrotron far-IR absorption spectroscopy: Excited state effects dominate the temperature dependence of the spectra. *Biopolymers*, **49**(7), 591–603.

Xu, H., & Berne, B. J. (2001) Hydrogen-bond kinetics in the solvation shell of a polypeptide. *Journal of Physical Chemistry B*, **105**(48), 11929–11932.

Xu, J., K. W. Plaxco, et al. (2006) Collective dynamics of lysozyme in water: terahertz absorption spectroscopy and comparison with theory. *J Phys Chem, **B***110(47), 24255-9.

Xu, Y. W., Moréra, S., Janin, J., & Cherfils, J. (1997) AlF_3 mimics the transition state of protein phosphorylation in the crystal structure of nucleoside diphosphate kinase and MgADP. *Proceedings of the National Academy of Sciences of the United States of America*, **94**(8), 3579–3583.

Yamada, M., Hashizume, D., Shimizua, T., Yokoyama S. (2007) L- Asparagine. *Acta Cryst,* **E63,** 3802- 3803.

Yamamoto, K., Tominaga, K., Sasakawa, H., Tamura, A., Murakami, H., Ohtake, H., and Sarukara, N. (2005) Terahertz Time-Domain Spectroscopy of Amino Acids and Polypeptides. *Biophys, **J.***89(3), 22-24 .

Yan, D., Cho, H. S., Hastings, C. A, Igo, M. M., Lee, S. Y., Pelton, J. G., Kustu, S. (1999) Beryllofluoride mimics phosphorylation of NtrC and other bacterial response regulators. *Proceedings of the National Academy of Sciences of the United States of America*, **96**(26), 14789–14794.

Yang, L. W., Bahar I. (2005) Coupling between catalytic site and collective dynamics: a requirement for mechanochemical activity of enzymes. *Structure,* **13,** 893–904.

Yanson, I. K., Teplitsky, A. B., Sukhodub, L. F. (1979) Experimental studies of molecular interactions between nitrogen bases of nucleic acids. *Biopolymers,* **18**(5), 1149–1170.

Yoneyama, H., Yamashita, M., Kasai, S., Kawase, B. K., Ito, H., Ouchi, T. (2008) Membrane device for holding biomolecule samples for terahertz spectroscopy. *Optics Communications,* **281**(7), 1909–1913.

Yoo, S. J., Kim, J., Lee, C-S., Nam, Y. (2011) Simple and novel three dimensional neuronal cell culture using micro mesh scaffold. *Neurobiol,* **20,** 110-115.

Young, D. W., Wilson, H. R. (1975) The crystal and molecular structure of 2'-deoxycytidine. *Acta Crystallogr,* **31**, 961.

Yu, B., Zeng, F., Yang, Y., Xing, Q., Chechin, A., Xin, X., Alfano, R. R. (2004) Torsional vibrational modes of tryptophan studied by terahertz time-domain spectroscopy. *Biophysical Journal*, **86**(3), 1649–1654.

Yutani, K., Ogasahara, K., & Kuwajima, K. (1992) Absence of the thermal transition in apo-alpha-lactalbumin in the molten globule state. A study by differential scanning microcalorimetry. *Journal of Molecular Biology*, **228**(2), 347–350.

Zendlova, L., Hobza, P., Kabelac, M. (2007) Stability of Nucleic Acid Base Pairs in Organic Solvents: Molecular Dynamics, Molecular Dynamics/Quenching, and Correlated Ab Initio Study. *J. Phys. Chem. B*, **111**(10), 2591-2609.

Zhang, C., and Durbin, S. M. (2006) Hydration-induced far-infrared absorption increase in myoglobin. *J Phys Chem, **B******110**(46), 23607-13.

Zhang, C. F., Tarhan, E., Ramdas, A. K., Weiner, A. M., and Durbin, S. M. (2004) Broadened Far-Infrared Absorption Spectra for Hydrated and Dehydrated Myoglobin. *J. Phys. Chem, **B******108**(28), 10077–10082.

Zhang, P., Knowles, B. A., Goldstein, L. S. B., Hawley, R. S. (1990) A kinesin-like protein required for distributive chromosome segregation in Drosophila. *Cell*, **62**, 1053--1062.

Zhang, S. L., Michaelian, K. H., & Loppnow, G. R. (1998) Vibrational Spectra and Experimental Assignments of Thymine and Nine of Its Isotopomers. *The Journal of Physical Chemistry A*, **102**(2), 461–470.

Zhu, Z. and Waggoner, A. S. (1997) Molecular mechanism controlling the incorporation of fluorescent nucleotides into DNA by PCR. *Cytometry,* **28**(3), 206-11.

Zhu, Z., Chao, J., et al. (1994) Directly labeled DNA probes using fluorescent nucleotides with different length linkers. *Nucleic Acids Res,* **22**(16), 3418-22.

Zhuang, W., Feng, Y., et al. (1990) Self-consistent calculation of localized DNA vibrational properties at a double-helix-single-strand junction with anharmonic potential. *Phys Rev,* **A41**(12), 7033-7042.

Zipper, H., Brunner, H., Bernhagen, J., & Vitzthum, F. (2004) Investigations on DNA Intercalation and Surface Binding by SYBR Green I, its Structure Determination and Methodological Implications. *Nucleic Acids Research*, **32**(12), e103.