

CARDIFF UNIVERSITY
SCHOOL OF PHYSICS AND ASTRONOMY



**TOWARDS DISEASE DIAGNOSIS THROUGH
TERAHERTZ SPECTROSCOPY OF BIOLOGICAL
COMPONENTS AND TISSUE**

By

PERI VAUGHAN JONES

SEPTEMBER 2017

A THESIS SUBMITTED TO CARDIFF UNIVERSITY FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

**TOWARDS DISEASE DIAGNOSIS THROUGH TERAHERTZ SPECTROSCOPY
OF BIOLOGICAL COMPONENTS AND TISSUE**

ACKNOWLEDGEMENTS

It's hard to know where to start or where to end. I would not have a PhD without the amazing support and opportunity from my supervisor, Professor Carole Tucker. I met Carole for the first time in an open day at Cardiff University and she's been my tutor, mentor and friend ever since. She was and still is incredibly passionate, energetic and motivational and has been an absolute joy to work with for all these years. Through difficult times and the doubtful days of a PhD student's life, she has always been able to motivate and remind me of the advantages and positives of our research. Thank you Carole for all your help, you have definitely gone above and beyond any other "supervisor".

I've been fortunate to work with amazing people within the Astronomy Instrumentation Group. Firstly I would like to thank Professor Peter Ade for helping me and trusting me with all the equipment in the FTS lab. To this day, I still admire his no-nonsense attitude, impeccable dedication to physics, and ability to make a good cup of tea. I have learned a lot and laughed even more. I would also like to thank Dr Rashmi Sudiwala who has also spent hours trying to teach me LabView and helping me with random problems I had when trying to take measurements of chicken breast. I feel very fortunate to be part of such a warm and friendly group within the University and want thank you all for making my PhD years enjoyable.

I also need to give a massive thank you to all my wonderful friends at Cardiff University.

I have been lucky to have such an amazing group of people within the Physics

department. I am so glad that I started with an incredible year group and that you were all there for a cup of coffee or a pint/glass of Prosecco in the pub when I needed a break, even if they did become one too many! I would like to thank so many people but I want to give a special thank you to Tom Brien and Paul Moseley, who really supported me through my time of need. I cannot thank you enough for all your help before, during and after my accident and for all the coffees and pints that we've shared over the years...and hopefully many more in the future. To all of the Cardiff crew, thank you for making my time as a student an amazing life experience.

I want to thank my examiners, Dr Emyr Macdonald and Professor Vincent Wallace for dedicating their time to reading my thesis and for a weirdly enjoyable viva, despite being utterly nervous. You were both very calm and supportive throughout the examination and gave constructive feedback and corrections to my work. I am now starting to believe that my thesis is not too shabby after all.

Finally, I want to thank my family. Living on the opposite sides of the country isn't always easy. Mam, Dad, Jema, Ela, Lewis, Zack, Flo-rider, Taid El, Taid Now a Nain Jenny. Gallai ddim esponio pa mor lwcus ydwi i gal chi fel teulu. Er fy mod i yn gallu bod yn boen, ag yn aml iawn does ganddo chi ddim syniad am fy ngwaith, dachi i gyd dal yn gwranddo a mae eich cefnogaeth drost y blynyddoedd yn golygu bob dim. A special thank you to Mam and Dad: you both inspire me to be the best person I can be and without you two, I would not be where I am today.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iv
LIST OF TABLES	viii
LIST OF FIGURES	x
CHAPTER 1. INTRODUCTION	1
1.1 DNA and biological molecules	1
1.2 The Terahertz region	7
1.3 Outline of thesis	13
CHAPTER 2. MOLECULAR SPECTROSCOPY	15
2.1 Vibrational Spectroscopy	15
2.2 Rotational spectroscopy	22
2.2.1 Non-rigid rotator	24
2.3 Vibration-Rotation Spectroscopy	25
2.4 Polyatomic molecules	25
2.4.1 Fundamental vibration	26
2.4.2 Skeletal vibrations	27
2.4.3 Group vibrations	28
CHAPTER 3. FOURIER TRANSFORM THEORY AND SET UP	31
3.1 Fourier transform theory	31
3.2 Fast Fourier transform	32
3.3 Advantages	38
3.3.1 Throughput / Jaquinot advantage	38
3.3.2 Multiplex/Fellgett advantage	38
3.3.3 Connes advantage	39
3.3.4 Nyquist sampling	39
3.4 Martin-Puplett Interferometer	40
3.5 Portable FTS	42
3.6 Electronics	46
3.7 Detectors	47
3.7.1 Pyroelectric detector	47
3.7.2 Indium Antimonide (InSb)	47
3.7.3 Germanium	48
3.7.4 TES pulse tube (cryo-free)	48
3.8 Background measurements and Fourier Transform	49
CHAPTER 4. MOLECULAR MODELLING	51
4.1 Numerical approach	53
CHAPTER 5. RESULTS	60
5.1 Nucleobases and nucleosides	62

5.1.1	Nucleobases made using different solutions	62
5.1.2	Purines and Pyrimidines	86
5.1.3	Nucleobases and nucleosides	97
5.2	Simulation and experimental data of nucleobases	114
5.2.1	Guanine	115
5.2.2	Cytosine	117
5.2.3	Thymine	123
5.2.4	Adenine	125
5.2.5	Uracil	127
5.3	Terahertz spectroscopy of DNA	135
5.3.1	THz spectroscopy of human DNA and RNA	136
5.3.2	A comparison of the THz spectra of human and mouse DNA	139
5.3.3	THz spectroscopy of various colonies of mouse DNA	142
5.3.4	THz spectroscopy of mouse DNA after heating	144
5.3.5	THz spectroscopy of mouse DNA over time	146
5.3.6	Conclusion	151
5.4	THz spectra of liquid samples	153
5.4.1	Method	154
5.4.2	Results	159
5.4.3	Conclusion	168
5.5	THz data of tissue	171
5.5.1	Method	172
5.5.2	Results	180
5.5.3	Conclusion	189
CHAPTER 6.	DISCUSSION	191
6.1	Nucleobases and nucleosides	191
6.2	DNA&RNA	195
6.3	Solutions	198
6.4	Tissue	200
CHAPTER 7.	CONCLUSION	205
CHAPTER 8.	References	210

LIST OF TABLES

Table 2.1: Bond-stretches and their group wavenumber from common molecular groups	29
Table 3.1: Identification of the numbered items on the portable FTS diagram	43
Table 3.2: Quantitative differences between the detectors	49
Table 5.1: The concentration of each powdered nucleobase for each of the different liquid solutions	62
Table 5.2: Frequency of the resonances identified in the THz spectra of guanine using GaP wave generator by (Nishizawa J, et al., 2005) and samples made using DI water and IPA taken on the FTS presented in this thesis.	67
Table 5.3: Frequency of the resonances identified in the THz spectra of cytosine using GaP wave generator by (Nishizawa J, et al., 2005) and samples made using DI water and IPA taken on the FTS presented in this thesis.	73
Table 5.4: Frequency of the resonances identified in the THz spectra of thymine using GaP wave generator by (Nishizawa J, et al., 2005) and samples made using DI water and IPA taken on the FTS presented in this thesis	77
Table 5.5: Frequency of the resonances identified in the THz spectra of cytosine using GaP wave generator by (Nishizawa J, et al., 2005) and samples made using DI water and IPA taken on the FTS presented in this thesis	81
Table 5.6: The ratio of the amount of nucleoside powder to solvent for each nucleoside made into a solution using DI water and IPA	98
Table 5.7: Table of infrared and calculated wavenumbers for anhydrous guanine	102
Table 5.8: Comparison of the absorption frequency of literature IR spectra, THz experimental data and DFT simulated data of thymine	124
Table 5.9: Content of samples from Wales Cancer Bank	137
Table 5.10: Labelling and identification of the colony and mouse number for each DNA sample	140
Table 5.11: Identification of the colony and mouse number for each DNA sample	142
Table 5.12: Table showing the times and temperatures of each DNA sample	145

Table 5.13: Assignment of absorption features found using infrared and Raman spectroscopy.....	162
--	-----

LIST OF FIGURES

Figure 1-1: Molecular structure of ribose (left) and deoxyribose with the missing oxygen atom circled (right).	2
Figure 1-2: Structure of the DNA backbone showing the phosphate attached to the 5' carbon ring by the removal of an OH from the ribose and an H from the phosphate.....	2
Figure 2-1: A diagram of two atoms at either end of a spring as an example of a 'rigid rotator' model.....	15
Figure 2-2: Graph showing the harmonic and anharmonic oscillator and their energy levels. Blue dotted line shows the harmonic oscillator with equally spaced energy levels and the red line showing the Morse potential with narrower energy gaps for higher energy levels	20
Figure 2-3: Infrared spectra of 2 μm polypropylene (PP) sheet taken under vacuum and H_2O encased between two layers of 2 μm of PP taken within atmospheric conditions. Both spectra have been ratio-ed against the corresponding background.....	21
Figure 2-4: The three fundamental vibrations of H_2O a) Symmetric stretch 3651 cm^{-1} b) Symmetric bend 1595 cm^{-1} c) Antisymmetric stretch 3755 cm^{-1}	27
Figure 2-5: Branched and chained structures which may undergo skeletal vibrations.....	28
Figure 3-1: Interferograms for (a) monochromatic light and (b) two different frequencies	34
Figure 3-2: Box-car function with the Fourier transform showing the side ripples prominently.....	37
Figure 3-3: Triangular function showing the ripples being suppressed.....	37
Figure 3-4: Schematic diagram of the permanent Michelson interferometer in the FTS lab at Cardiff University	40
Figure 3-5: Images of the FTS taken when it was opened to show the internal structure. The images have been labelled from 1-5 corresponding to the sections in the schematic diagram in Figure 3.4.....	41
Figure 3-6: Bird's-eye view of the portable FTS including the dimensions in millimetres drawn on CAD	43
Figure 3-7: Set up of FTS in Cardiff for alignment and calibration	44

Figure 3-8: Schematic diagram of the polarisation of the light through the spectrometer	45
Figure 3-9: Schematic diagram of the FTS set up in the diagnostic room at Daresbury ..	46
Figure 3-10: Terahertz transmission of 2 μ m polypropylene (PP) from 20 to 620 cm^{-1} ...	50
Figure 5-1: A schematic depicting how bases (B) aggregate when placed in a solution which is represented as water molecules H ₂ O	63
Figure 5-2: THz spectra of guanine made using DI water, IPA, heavy water, HBSS and PBS from 20 to 620 cm^{-1}	65
Figure 5-3: Photograph of guanine made using DI water (left) and IPA (right)	66
Figure 5-4: (a) Optimised crystal cell structure of solid guanine calculated using LDA functional and PW methodology. The dashed lines along with the numbers represent the intermolecular H-bonding distance (measured in picometers). The distance between the two pairs of dimers is about 310.2 pm. (A) Oxygen atom from the upper guanine molecule. The remaining atoms were omitted. (b) Lateral view, along the b-axis, of the optimised unit cell (Lopes R. P, 2012).	69
Figure 5-5: THz spectra of cytosine made using DI water, IPA, heavy water, and PBS from 20 to 620 cm^{-1}	70
Figure 5-6: Photograph of cytosine made using DI water (left) and cytosine made using IPA (right).....	71
Figure 5-7: THz spectra of cytosine in IPA and DI water from 20 cm^{-1} to 320 cm^{-1} to show the main absorption features corresponding to the ones in Table 2	72
Figure 5-8: THz spectra of thymine made using DI water, IPA, heavy water, HBSS and PBS from 20 to 620 cm^{-1}	75
Figure 5-9: Photograph of thymine made using DI water (left) and made using IPA (right)	76
Figure 5-10: THz spectra of adenine made using DI water, IPA, heavy water, HBSS and PBS from 20 to 620 cm^{-1}	79
Figure 5-11: Photograph of adenine made using DI water (left) and adenine made using IPA (right).....	80
Figure 5-12: THz spectra for (a) protonated adenine and (b) deuterated adenine taken at 5K (red line) and 300K (grey line) with the predicted DFT features provided by the green lines taken by (Feng Zhang, 2014).	82
Figure 5-13: Molecular structure of 5 nucleobases: adenine, guanine, cytosine thymine and uracil.....	86

Figure 5-14: Molecular structure of adenine and guanine highlighting the imidazole ring	87
Figure 5-15: THz spectra of guanine and adenine in DI water from 20 to 620 cm^{-1}	88
Figure 5-16: THz spectra of guanine and adenine made using IPA from 20 to 620 cm^{-1}	89
Figure 5-17: Molecular structure of cytosine, thymine and uracil highlighting the main similarity between each molecule	90
Figure 5-18: THz spectra of cytosine, thymine and uracil made using DI water from 20 to 620 cm^{-1}	91
Figure 5-19: THz spectra of cytosine, thymine and uracil made using IPA from 20 to 620 cm^{-1}	92
Figure 5-20: Molecular structure of thymine (left) $\text{C}_5\text{N}_2\text{O}_2\text{H}_6$ and uracil (right) $\text{C}_4\text{N}_2\text{O}_2\text{H}_4$ with the red circles indicating the “O7” and “O8” positions.....	93
Figure 5-21: THz spectra of thymine and uracil made using DI water from 20 to 620 cm^{-1}	94
Figure 5-22: Molecular structure of the five nucleosides: guanosine, cytidine, thymidine, adenosine, and uridine.	97
Figure 5-23: THz spectra of glucose powder pressed into a 3.3 μm PP substrate and glucose film which was made by making a solution with DI water and left to dehydrate	99
Figure 5-24: THz spectra of guanine and guanosine made in DI water from 20 to 620 cm^{-1}	101
Figure 5-25: THz spectra of cytosine and cytidine, both made into a solution using DI water from 20 to 620 cm^{-1}	104
Figure 5-26: THz spectra of cytosine and cytidine made using IPA from 20 to 620 cm^{-1}	105
Figure 5-27: THz spectra of adenine and adenosine made using DI water from 20 to 620 cm^{-1}	107
Figure 5-28: THz spectra of adenine and adenosine made using IPA from 20 to 620 cm^{-1}	108
Figure 5-29: THz spectra of thymine and thymidine made using DI water from 20 to 620 cm^{-1}	110
Figure 5-30: THz spectra of thymine and thymidine made using IPA from 20 to 620 cm^{-1}	111

Figure 5-31: THz spectra of guanine made using DI water and IPA, the simulated data using DFT and scaled simulated data of guanine.	115
Figure 5-32: Two keto- tautomer of guanine that are the most common structure in anhydrous guanine. Structural representation of the calculated (DFT/B3LYP 6-31G**) four most stable tautomeric forms of guanine, in the gas phase, presented by (Lopes R. P, 2012)	116
Figure 5-33: Structure of the stacked sheets of the anhydrous guanine molecules viewed along the b-axis.	117
Figure 5-34: THz spectra of cytosine made using DI water and the simulated data of cytosine monohydrate	118
Figure 5-35: THz spectra of cytosine made using DI water and heavy water to show the agreement between the two samples	119
Figure 5-36: Structure of cytosine monohydrate as given by Jeffrey and Kinoshita. a) Shows the hydrogen bonding (dotted lines) along the a direction and b) shows the hydrogen bonding (dotted lines) viewed along the b-axis	120
Figure 5-37: THz spectra of cytosine made in IPA and the simulated data of cytosine alpha polymorph	121
Figure 5-38: Structure of anhydrous cytosine assuming that the molecules are in the amino form, viewed along the c-axis shown by (Marsh, 1963).....	122
Figure 5-39: Simulated THz spectra of thymine with the main absorption features identified and compared to experimental data of thymine made using DI water and IPA	123
Figure 5-40: THz spectra of adenine made in IPA and the simulated data of adenine form I	125
Figure 5-41: Structure of adenine hydrogen bonded via N-H---N to another adenine molecule given by Mahapatra et al. (Sudarshan Mahapatra, 2008).....	126
Figure 5-42: Packing diagram of adenine showing the hydrogen bonded molecules (dotted lines) viewed along the c-axis given my Mahapatra et al. (Sudarshan Mahapatra, 2008)	126
Figure 5-43: THz spectra of uracil made in IPA and the simulated data of uracil	127
Figure 5-44: a) initial assumption of the crystalline structure of uracil b) actual structure of crystalline uracil after refined electron-density maps in which one oxygen atom makes two hydrogen bonds with a nitrogen atom.....	128

Figure 5-45: Crystalline structure of uracil projected on to (0,0,1). The left side shows one layer of the hydrogen bonded molecules whereas the right side of the image shows the overlay of another layer	128
Figure 5-46: Resonance at 90 cm ⁻¹ . The atoms on “the left” with respect to the view of the reader will move upwards as the right atoms move downwards.	129
Figure 5-47: Resonance at 121 cm ⁻¹ . The atoms at "the back" with respect to the view of the reader, will move upward as the front move down.....	130
Figure 5-48: Resonance at 122 cm ⁻¹ . This is looking down the c-axis and showing the whole molecule twisting.	130
Figure 5-49: Resonance at 125 cm ⁻¹ . This is looking down the c-axis and showing the whole molecule shifting upwards and downwards.	131
Figure 5-50: Resonance at 418 cm ⁻¹ . This shows the bending of the hydrogen atoms at a much greater amplitude than the bending of the attached carbon atoms. As the hydrogen atom moves “upwards” the carbon atoms move “downwards”.....	132
Figure 5-51: Resonance at 433 cm ⁻¹ . The oxygen atoms (red) are symmetrically bending with respect to the nitrogen and hydrogen atoms along the a-axis.	132
Figure 5-52: Resonance at 545 cm ⁻¹ . The CO and NH molecule bend in and out as the remaining attaching atoms bend in the opposite direction.....	133
Figure 5-53: THz spectra of Wales Cancer Bank (WCB)-provided DNA and RNA samples after dehydration on 2μm PP from 20 to 580 cm ⁻¹	138
Figure 5-54: THz spectra of dried human DNA from Wales Cancer Bank and dried mouse DNA from biosciences department from 10 to 310 cm ⁻¹	141
Figure 5-55: THz spectra of mouse DNA from different colonies dried at room temperature and measured from 10 to 310 cm ⁻¹	143
Figure 5-56: THz spectra of mice DNA from the same colony (HPPT2) heated to temperatures of 70 ± 5 °C	145
Figure 5-57: THz Spectra of mouse DNA from different colonies from 10 to 310 cm ⁻¹ . AXN, KP, FAX, WKD, PAC, HTTF, MALC and LAV, were dried and left at room temperature for 2 days.	147
Figure 5-58: THz spectra of mice DNA from the same colony (HPPT2) measured between 10 and 310 cm ⁻¹ after heating and leaving at room temperature for a time of 5 days, except for ‘HPPT2 102’ sample which was only 1 day.	148

Figure 5-59: THz spectra of mice DNA from 10 to 310 cm^{-1} . The samples were from the same colony (HPPT2) and measured after heating and leaving at room temperature for a time of 7 days, except for the 'HPPT2 102' sample which was only 3 days.....	149
Figure 5-60: THz spectra of 'HPPT2 98' mouse DNA highlighting the difference in the absorption features over a period of 5 and 7 days, measured from 10 to 300 cm^{-1}	150
Figure 5-61: Pictures showing the CO ₂ canister next to the laboratory FTS. The green pipe feeds the gas into the sample and optics chamber from the canister. The red cryostat houses the cryogenic detector.	155
Figure 5-62: THz spectra of the sample chamber of the FTS after purging with CO ₂ for 15 minutes. Two spectra are shown, taken 12 minutes apart and indicate that the background spectrum is relatively stable.	156
Figure 5-63: Pictures of a) the initial sample holder with the liquid inside, b) the samples being looked at under a monochromatic light source	158
Figure 5-64: THz spectra of H ₂ O contained between two layers of 2 μm PP ratioed against the CO ₂ background.....	160
Figure 5-65: THz spectra of H ₂ O contained between two layers of 2 μm PP ratioed against the CO ₂ background and the spectra of two layers of 2 μm PP with no sample (air gap).....	161
Figure 5-66: Experimental absorption spectrum of liquid water from 269 to 298 K: (a) 269 K; (b) 277 K; (c) 298 K taken from (Woods & Wiedemann, 2004).....	163
Figure 5-67: THz spectra of liquid H ₂ O and D ₂ O	165
Figure 5-68: THz spectra of guanine in DI water solution compared to dried guanine made using DI water	167
Figure 5-69: Diagram of the experimental procedure for flattening the chicken breast.	175
Figure 5-70: Chicken liver and breast being prepared (left) and the chicken liver on a piece of foil being placed in to the isobutene for the snap-freezing (right).....	176
Figure 5-71: Picture of the cryostat microtome used for the cryosectioning of tissues..	177
Figure 5-72: The flat metal plates used to place the samples (left) and the embedding material that is used to hold the sample rigid and on the metal plate	177
Figure 5-73: The sample holder (left) and the sample holder with the tissue held in place in the opaque embedding material	178
Figure 5-74: Screen of the cryostat microtome showing the temperature within (-22) and the range of section thickness that can be made	179
Figure 5-75: Photograph of inside the cryostat.....	179

Figure 5-76: THz spectra of human skin and breast tissue embedded in wax.....	180
Figure 5-77: THz spectra of dry, flat chicken breast after being placed under weight and flattened within known shims while being left to dry in a vacuum oven	182
Figure 5-78: THz spectra of chicken breast at various thicknesses from 5 to 22 μm	184
Figure 5-79: THz spectra of chicken liver at various thicknesses between 5 and 20 μm and chicken breast at 22 μm for comparison	185
Figure 5-80: THz spectra of beef and pig kidney from 20 μm to 60 μm	187
Figure 5-81: THz spectra of pig kidney in red and pig liver from 20 μm to 60 μm	187
Figure 6-1: THz spectra of H ₂ O taken in a CO ₂ purged FTS and the THz spectra of chicken breast and liver taken under vacuum in the same FTS	201
Figure 6-2: THz spectra of dried chicken breast and liver, human DNA and mouse DNA from 20 to 300 cm^{-1}	202

“If nature has given you exceptional ability it would be a shame not to use it. You have to work on what you are given. Otherwise you will have achieved nothing in your life and wasted what you have in you.” Eddy Merckx

ABSTRACT

Using a Martin-Puplett interferometer and high sensitivity cryogenic detectors, a Fourier Transform Spectrometer (FTS) was used to investigate a range of biological samples in the Terahertz frequency region. Large biological molecules such as deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) and their constituent nucleobases and nucleosides: guanine, cytosine, adenine, thymine and uracil; guanosine, cytidine, adenosine, thymidine and uridine, have shown clear resonances within the region between 0.6 THz and 18 THz. Interest must be paid to the difference in the spectra of cytosine when the samples were prepared using different solvents, showing the potential and sensitivity of this tool for identifying small differences in the crystalline structure of biological samples.

Modelling the nucleobases using Density Functional Theory (DFT) provided confirmation with experimental data of the frequency location of the resonances, and identifies the nature of the resonances. Further experimental work is undertaken on samples of animal tissue and liquid solution (H_2O and D_2O) which renders difficulties in the analysis of the results due to the complexity of sample. This thesis provides a better understanding of the electromagnetic interaction of large biological molecules with THz radiation which will compliment and strengthen the future application of THz spectroscopy for detecting mutations that lead to cancer.

CHAPTER 1. INTRODUCTION

1.1 DNA and biological molecules

In April 1953, Watson and Crick (Watson & Crick, 1953) published their work on the structure of Deoxyribose Nucleic Acid (DNA). The double stranded DNA molecule was shown to have a double helix configuration. Further research since has shown that this is quite common in nature, for example in some protein structures (α -helix). Within the core of the DNA are four nucleobases which are hydrophobic. Each one is attached to a (deoxy) ribose sugar molecule which are hydrophilic and together make a unit called a nucleoside. For the DNA molecule specifically, it is more energetically favourable to be in this conformation due to the hydrophobicity of the bases, causing the double stranded molecule to twist and protect the centre.

RNA (ribonucleic acid) is just as important as DNA within the function of cells, tissue and organs. Whereas DNA contains genetic instructions for development and functionality of all living organisms, RNAs are vital for protein synthesis, translation and transmission of the genetic information. There are many different types of RNAs such as mRNA, rRNA and tRNA to name a few, responsible for protein synthesis (News Medical, 2013). RNA is most often found in a single stranded form. The main difference, as suggested in the name, is that the deoxyribose molecule has lost an oxygen atom from one of the OH as seen in Figure 1-1.

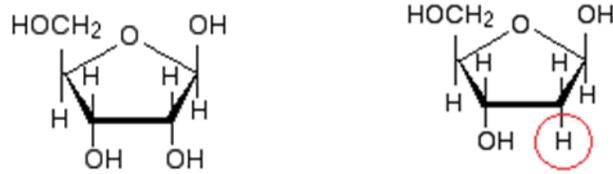


Figure 1-1: Molecular structure of ribose (left) and deoxyribose with the missing oxygen atom circled (right).

In the DNA strand, the nucleoside is bonded to a phosphate group by removing the $-OH$ from the 5' carbon atom and attaching the remaining $-CH_2$ with the oxygen atom in the phosphate molecule. This unit is now called a nucleotide. The phosphate molecule then attaches itself to another sugar molecule and builds into a strand making a phosphate-sugar backbone with the hydrophobic nucleobases at the centre as seen in Figure 1-2.

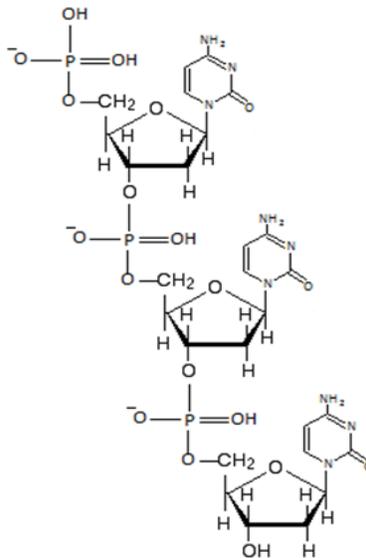


Figure 1-2: Structure of the DNA backbone showing the phosphate attached to the 5' carbon ring by the removal of an OH from the ribose and an H from the phosphate.

Watson and Crick discovered that the four nucleobases can only bond with specific complimentary partners such that cytosine would bond with guanine and adenine with thymine. They are joined together through hydrogen bonding and are stacked next to the adjacent pair when the DNA is twisted into the double helix shape.

DNA has millions of base pairs, with each double stranded DNA sectioned into genes which contains specific nucleobase sequencing which govern different functions or traits. In cell division, or mitosis, DNA replication will copy and create an identical copy of the mother genes by unwinding and splitting the DNA. During this process, mistakes or mutations often occur which are quickly repaired or in some cases destroyed. However, sometimes faulty genes can go unrecognised and will multiply at a much more aggressive rate. These mutations tend to lead into a high concentration of faulty cells and can form a tumour.

Tumours can be either benign or malignant (cancerous). A benign tumour grows slowly and does not necessarily spread to other parts of the body. They have similar cells to normal cells and will only cause a problem if they grow very large, press on or take up space close to a vital organs. Malignant tumours are the cancerous ones and tend to grow rapidly, can spread and damage surrounding tissues, and migrate to other tissues within the body via the blood stream. There are more than 200 different types of cancer, each one being very different and requiring tailored treatment which can be time consuming, difficult and expensive. Unfortunately, 1 in 2 people in the UK will get cancer in their lifetime, meaning treatment and care of cancer patients is costing even more time and money (Cancer Research UK, What is cancer?, 2014).

In the past few years, early diagnosis of cancer has been pushed to the forefront as a major priority, in an attempt to ‘beat cancer’. Cancer Research UK has made it one of their main priorities within their policy to enhance awareness campaigns, provide better screening facilities and diagnostic services. In their recent briefings it was stated that “it is concerning that in 2015-16, over 24,000 people waited longer than two months to start treatment, and this trend is not showing signs of improvement” (Cancer Research UK, 2017).

There are many different types of imaging systems spanning the electromagnetic spectrum which have a variety of usage depending on their characteristics. Some current screening and cancer imaging facilities in use are:

- Nuclear medicine imaging uses radiopharmaceuticals, which are radioactive sources attached to a targeted pharmaceutical, taken either through injection, orally, or inhaled. Once in the body, the radiopharmaceutical is detected using a gamma camera or a positron emission tomography (PET) scanner or probe. Currently, this is the most sensitive clinical imaging tool available according to (Fass, 2008). However, the use of radioactive sources poses an aspect of risk despite it having very few side effects as the dosage of radioactive substance is so low. Nuclear medicine imaging may not be able to identify small tumours and cannot tell the difference between benign and malignant tumours (Canadian Cancer Society, 2017).
- X-rays and Computed Tomography (CT) uses the absorption of x-ray radiation to image within the human body. An x-ray source and detector will be placed at

either side of the area of body that is of interest. CT can produce 3D images of the structure of the bones and surrounding tissue and organs by having a circular rotating frame around a patient lying in the centre (Medical Radiation, 2016). X-rays images will contain different structures which overlap e.g. rib cage over the lungs and heart. However, CT scans allow overlapping sections to be viewed individually, making the internal anatomy more apparent. Although CT is not as sensitive as PET imaging, it does have a better spatial and temporal resolution, giving a better contrast between tissue and bone of different densities (Fass, 2008). The contrast of the image depends on the dosage which depends on the area of interest i.e. more dense areas require higher dosages. But high dosages of x-rays can be harmful to existing healthy cells including DNA molecules which can lead to cancer.

- Magnetic Resonance Imaging (MRI) uses powerful magnets and radio waves to show high resolution images of soft tissue. The magnets are used to align the hydrogen nuclei's spin whereby the radio frequencies creates a varying magnetic field which flips the spin of the nuclei. When the radio frequencies are switched off, the spin of the nuclei return to their initial orientation, producing a radio signal which is measured with the detectors within the scanner. It can give better detail than CT scans and can be combined with CT images to identify tumours. There are variations in the basic concept which can give different images and information. Diffusion MRI is a way of measuring the flow of water molecules within the body and is often used to help diagnose strokes or tumours which can restrict the diffusion of water. Functional MRI is another branch whereby it is

used to look at the activity within the brain by measuring the blood flow in the brain (Lewis, 2014). Although this technique is not invasive or ionising like X-rays or CT, it is a very expensive technique and can have a long waiting list which is a big issue with early diagnosis (Cancer Research UK, Cancer Research UK, 2015).

- Ultrasound uses high frequency sound waves (above 20 KHz) and is based on the scattering, reflection and frequency shifts of these waves within a medium. It can also be sensitive to tissues' elasticity, and as cancer is less elastic than normal tissue, it has shown promising results when used to identify breast, prostate and liver cancer (Fass, 2008). Sonograms are the images created and are used for better detail of local soft tissue or fluid-filled cysts. However, this technique lacks detail such as whether a tumour is cancerous and cannot be used in certain parts e.g. lungs and bones (American Cancer Society , 2015).

Although these imaging techniques have been, and still are, successfully used for cancer diagnosis, they are still limited in detecting tumour volume and using it as a measurement of the progression of the cancer. Many of these techniques are combined to gain different biomarkers from different images. If there is a cause for concern from a scan, biopsies can be made which removes a small amount of the tissue. These can be examined by a pathologist through staining and looking under a microscope to identify the cancer cells and margins. This is currently the most effective way of diagnosing cancer but it is invasive and dependent on the judgement of the pathologists.

During the excision of cancerous tumours in surgery, there is a critical margin on the amount of tissue removed. It is important to remove all the cancerous tissue but it is also vital that the surrounding healthy tissue is not removed. Currently, there are no in-vivo equipment that can detect cancerous tissue or any potentially damaging cells within the vicinity that could develop into further complications after surgery. In breast cancer alone, it results in 20% of women returning for further surgery after the tumour has been removed (Jeevan, et al., 2008). Not only is this distressing for the patient but it could also be life threatening.

Therefore, there is a need for a diagnostic tool which;

- 1) Can be used *ex-* and *in- vivo* in identifying early onset stages of cancer, preferably on the molecular level
- 2) Can better understand the origins of the cancer through identifying and characterising the different types of cancers
- 3) Is non-ionising and painless for the patient.

Powerful tools are used today to diagnose cancer but there is still a lack of early diagnosis procedures that can detect cancer cells before they develop into any type of tumour without being invasive or ionizing. To that end and in recent years, an interest in the potential use of “Terahertz radiation” has taken off.

1.2 The Terahertz region

The terahertz (THz) region, broadly spanning from 1 to 0.016 mm (0.3 to 20 THz), has been vastly developed in recent years with applications including space, security, defence and medical industries. It bridges the gap between the microwave and mid-infrared region, and has been a widely researched area for the past couple of decades, but it still remains a complex and intriguing area for medical application and biophysics.

In 1968, Fröhlich (Frohlich, 1968) proposed that large molecules and long chain polymers have “extraordinary” dipole properties largely due to the hydrogen bonds. The interaction of external oscillations can propagate longitudinally resulting in electric modes at frequency ranges of $10^{11} - 10^{12} \text{ sec}^{-1}$. This hypothesis lends itself well for the use of Density Functional Theory (DFT) as a modelling tool for large molecules as explored in CHAPTER 4. Having low frequencies and low energies makes this region an attractive area for medical application due to it being safe and non-ionizing. This hypothesis welcomes the investigation of terahertz frequencies and large biological molecules.

Current THz probes and medical application:

Current research in the THz field predominantly involves the use of time-domain spectroscopy (TDS) such as THz pulsed spectroscopy (TPS) and THz pulsed imaging (TPI), to investigate different samples. Since the initial publication of THz transmitter and detectors in free space by (Smith, Auston, & Nuss, 1988), THz spectroscopy has come a long way with applications in many different fields such as material science, homeland security, defence, inspection and quality control and medical imaging and diagnosis.

The initial difficulty of THz spectroscopy was the production and detection of THz frequencies due to limitations in laser technology and detection systems. One method is photoconductive systems which use a femtosecond laser incident on a photoconducting material, often a semiconductor crystal such as gallium arsenide. The ultrafast pulsed laser (~100fs) will cause an excitation of photoconductive charge carriers which are subsequently separated and generates pulses of broadband terahertz radiation. According to (Nuss & Hu, 1995), they presented the first application of a THz time domain system to package inspection and chemical content.

The application of THz spectroscopy is rapidly growing, with many institutes and research groups working closely with medical practitioners and laboratories. A vast amount of research has been done using THz pulsed spectroscopy to look at differences in normal and diseased regions of human tissue. Early on (Schmuttenmaer & Kindt, 1996) looked at the properties of liquids using a TPI system which allowed them to measure the refractive index and absorption coefficient over the frequency ranges of 0.1 - 1.5 THz. This helped pave the way for better understanding between healthy tissue and tumour margins (Wallace, et al., 2004), (Wallace, et al., 2006) which showed good contrast between normal tissue and tumours. By using a portable TPI system (TeraView), they were able to explore the frequency range between 0.1 and 3 THz. The TPI system has the capabilities of obtaining the absorption coefficient and the refractive index of a material as well as depth information due to reflection of the THz pulses from within the sample. However, just like many time domain systems, they are limited to a short frequency range.

Other in vivo work has been done using a Teraview TPI system which was able to identify margins within freshly excised human breast cancer (Ashworth, et al., 2009). This technique exploits the high concentration of water in cancer cells which results in a greater contrast in the THz images. However, this technique is limited to the lower THz range between 0.4THz and 5THz, and it does not identify the origin of the cancer neither does it identify the stage of the cancer.

(Zhang , Zhao , Jin, Hou, & Jia, 2012) used a transmitted time domain spectroscopy (TDS) imaging system with a Ti: sapphire pulsed laser producing THz pulses with a time width of less than 100fs and a frequency range of 0.2 – 4THz to image healthy and cancerous mouse liver tissue. The absorption and imaging contrast between the healthy and cancerous tissue was better as frequency increased but they suffered as the signal-to-noise ratio (SNR) was very poor. It is admitted that “the absorption increased with the frequency increasing” but using their equipment “SNR was very poor such as the image at 2.69 THz”. Although there is evidence of image contrast, there is a lack of knowledge and scientific reasoning.

As some time domain THz systems have a high average power, there have been concerns regarding the damage and breakages to DNA from the intense light. (Titova L. V., et al., 2013) identified an increase in phosphorylated H2AX when THz radiation was incident on artificial human skin tissue which reconstructs normal skin tissue and keeps it metabolically active (Boelsma E, 2000). In the construction of chromosomes, DNA is wrapped around histones which are proteins, with H2A being a type of histone and H2AX being a subtype. When there are double stranded breaks within the DNA, the histones react by phosphorylation and changing its name to H2AX (Sedelnikova, Pilch,

Redon, & Bonner, 2003). It is important to recognise the increase in H2AX from the THz radiation, whereby the experiment was done with high powered, high intensity lasers. This could be avoided by using a low intensity source and low power output, avoiding any thermal heating or damage to the DNA.

Another interesting outcome from experimentation on biological molecules was achieved using Human fibroblasts (HFFF2) exposed to high intensity pulsed THz source in the range of 0.1-0.5 THz. The THz radiation was provided using the ENEA (ENEA, 2017) Compact Free Electron Laser. The investigation looked at the effects of high peak power, low average power, typically a few mW, on human fibroblasts. These are the cells that are responsible in synthesising extracellular matrix, collagen the structural framework for animal tissue and plays a critical role in wound healing. Different assays and test were undertaken to investigate the effects of the THz radiation on the fibroblasts samples. One was Cytokinesis-block micronucleus assay which consists of counting micronuclei in the binucleated (BN) cells. Binucleated cells are cells that contain more than one (usually two) nucleus in one cell. During mitosis (cell division) a cell doubles and divides its chromosomes into two daughter cells. But sometimes, the cell cannot split properly resulting in one cell with two nuclei. Micronuclei are small bodies that are formed during mitosis from lagging chromosomes. Sometimes, microtubules are not properly connected to the chromosomes which then move in opposite directions during “splitting” which result in chromosomes braking off and is enveloped as another nucleus in one of the daughter cells. Micronuclei are characterised in a cell that has some sort of DNA damage, therefore, micronuclei are commonly found in cancerous cells. Statistical results showed that there was an increase in micronuclei in the sample after radiation of THz compared

to untreated samples. However, this is not conclusive as there are many other possible triggers. The paper states that the human fibroblasts (HFFF2) were seeded into the petri dish 24 hours before THz exposure. This raises the question as to whether the fibroblasts were in normal conditions for 24 hours. Also, the cell culture was “fixed in freshly Carnoy modified solution (5:1 v/v methanol/acetic acid) and dyed with 4,6-diamodino-2 phenylindole” within their lab which also suggests that the samples were in alien environment.

Modelling work is used in looking at the breathing dynamics of DNA. (Alexandrov, Gelev, Bishop, Usheva, & Rasmussen, 2010) investigated how THz radiation can expose double stranded DNA to local strand separation using the Peyrard-Bishop-Dauxois (PBD) model. This model is the best known model for identifying DNA strand separation through “melting” and mechanical “unzipping” (Alexandrov, Bishop, & Usheva, 2008). They state that the physical nature of the interaction between DNA molecules and THz radiation is not known in detail. Their modelling investigation showed that a THz field may cause dynamical separation of the double stranded DNA, but that large breathing modes or fluctuations with significant amplitude did not occur and are very rare.

Furthermore, other evidence showing that THz radiation does not create any thermal effects were performed using a model based on Kirchhoff’s heat equation and the influence of continuous wave THz beam on water (Kristensen, Withayachumnankul, Jepsen , & Abbott, 2010). The THz effect on water showed that increasing frequency increases the temperature difference of the water sample. The temperature difference increased by 40% from 1.4K/mW to 2K/mW when the frequency is varied from 0.1THz to 10THz. However, this would only be the case for very high powered beams, with a

very small beam spot area for a long period of time which is not the case in any current THz devices. It was concluded that THz radiation is non-invasive when considering biological samples and would require much higher average power than the standard THz devices used today.

1.3 Outline of thesis

This thesis presents an underused technique to better understand the interaction between Terahertz frequencies and biological molecules in the attempt to build a THz probe for early diagnosis with spectral and spatial information. A Fourier Transform Spectrometer (FTS) based on a Martin-Puplett interferometer is used to gain spectroscopic data with a broadband THz frequency range and high signal-to-noise. CHAPTER 3 highlights the advantages of this technique as well as the set up for the portable FTS built for use at the THz line on ALICE (Accelerators and Lasers In Combines Experiments) free electron laser at Daresbury (Clarke, Dunning, Surman, & Thompson, 2010). This is the first of its kind and showcases the potential of the FTS being a portable device.

To understand the dynamics of DNA, there is a need to understand its components which is why in CHAPTER 5.1 the spectroscopic data of the nucleobases (guanine, cytosine, thymine, and adenine) are investigated. They were made into a solution using 5 different solvents to investigate the influence on solvents on the crystalline structure of the nucleobases. To build on the nucleobases, in section 5.1.3 the nucleosides (guanosine, cytidine, thymidine and adenosine) are also investigated in the solid state and compared to the spectra of their common nucleobase.

Healthy DNA and RNA samples were received from the Wales Cancer Bank, University Hospital Wales, Cardiff, to look at the potential difference in the spectra of double stranded DNA and single stranded RNA. Further investigation on mice DNA was undertaken to further identify key signatures in DNA molecules and to compare the THz spectra of human and mouse DNA.

To develop a THz probe that could be used *in-vivo*, there is a need to understand the spectra of biological samples in solution. Water is highly absorptive in the THz region, and the different concentrations of water in healthy and cancerous tissues are used by TDS systems to highlight the margins. However, there is a need to understand the spectral information which is why in section 5.4 the THz spectra of H₂O, D₂O and guanine in DI water were investigated.

Attaining healthy and cancerous samples were not possible due to time constraints and ethical reasons. However, samples of tissue from different animals were investigated in section 5.5. Various techniques in preserving and fixing tissues were investigated and compared with regards to the quality of the THz spectra.

This thesis provides an insight into the vast potential of Fourier Transform spectroscopy in the THz region. It gives an overview of its ability to identify fine spectral features in nucleobases and nucleosides, as well as the broad absorption features due to long range vibrational resonances of larger, more complicated, molecules.

CHAPTER 2. MOLECULAR SPECTROSCOPY

2.1 Vibrational Spectroscopy

To understand the absorption and transmission of electromagnetic waves through a medium or material, first consider a diatomic molecule as two masses on a spring seen in Figure 2-1.

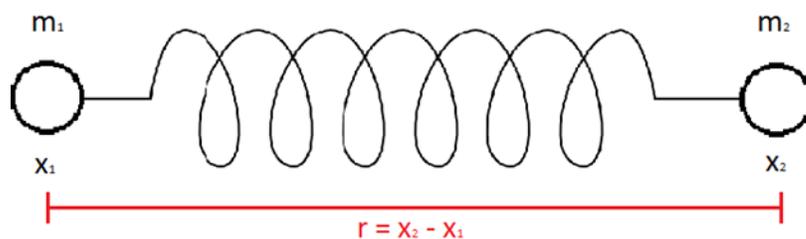


Figure 2-1: A diagram of two atoms at either end of a spring as an example of a ‘rigid rotator’ model

The molecule will undergo simple harmonic motion (SHM) from attractive and repulsive forces as well as other environmental influences. The force from the vibrating molecules can be written as;

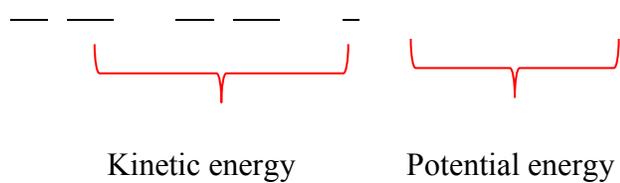
2-1

where μ is the reduced mass given by $\mu = \frac{m_1 m_2}{m_1 + m_2}$. If the general wave solution is applied, such that,

$$\psi = A \cos(\omega t - kx) \quad (2-6)$$

where the angular frequency is given by $\omega = \frac{2\pi\nu}{h}$, and applying the potential energy as stated in Equation 2-3, we can evaluate the vibrational energy as;

$$E = \frac{1}{2} \mu \dot{x}^2 + \frac{1}{2} k x^2 \quad (2-7)$$



However, classical views such as this one are not entirely correct when considering absorption of energy through spectroscopy. Therefore, there is a need to adopt a quantum mechanical approach. According to the quantum model, molecules can have discrete energy levels which involves having a non-zero ground level energy. The time-independent Schrodinger's equation is given by

$$-\frac{\hbar^2}{2\mu} \frac{d^2\psi}{dx^2} + V(x)\psi = E\psi \quad (2-8)$$

which can be simplified as $\frac{d^2\psi}{dx^2} + \frac{2\mu(E - V(x))}{\hbar^2}\psi = 0$ and using the reduced mass, μ .

2-9

The solution can be in the form of

2-10

where α is a constant and by substituting Equation 2-10 into Equation 2-9, it gives an expression for the Hermite differential equation. This equation can be satisfied by a set of functions given by the Hermite polynomials $H_v(\alpha x)$ where $(v=0, 1, 2, \dots)$, where each gives a solution to Equation 2-9. We finally end with the energy levels which have the discrete values given by Equation 2-11;

2-11

where v is the vibrational quantum number. We can also define the bandgap energy in terms of the wavenumber knowing that the energy of a photon is

which gives

2-12

so energy levels form a series going up in integers of $\frac{h\nu}{2}$, etc. which means that the zero-point energy at $v=0$ is not zero.

For various temperatures, the energy states of some atoms or molecules will change. The population of molecules at a certain energy level is given by the Boltzmann formula.

—

2-13

The more precise form of expressing the energies would be by the Morse potential;

2-14

where ϵ is the constant for a particular molecule and D_e is the dissociation energy, also given as the depth of the potential well. The Morse curve is shown in Figure 2-2, as the ‘anharmonic oscillator’.

When Equation 2-14 for the Morse potential is combined with the Schrodinger’s equation as a substitution for the energy, the vibrational energy levels can be better found for an anharmonic oscillator;

— —

2-15

where x_e is the anharmonicity term. Due to an anharmonicity term which is always positive and small (~ 0.01) the vibrational levels crowd more closely together with increasing vibrational quantum number, as seen in Figure 2-2.

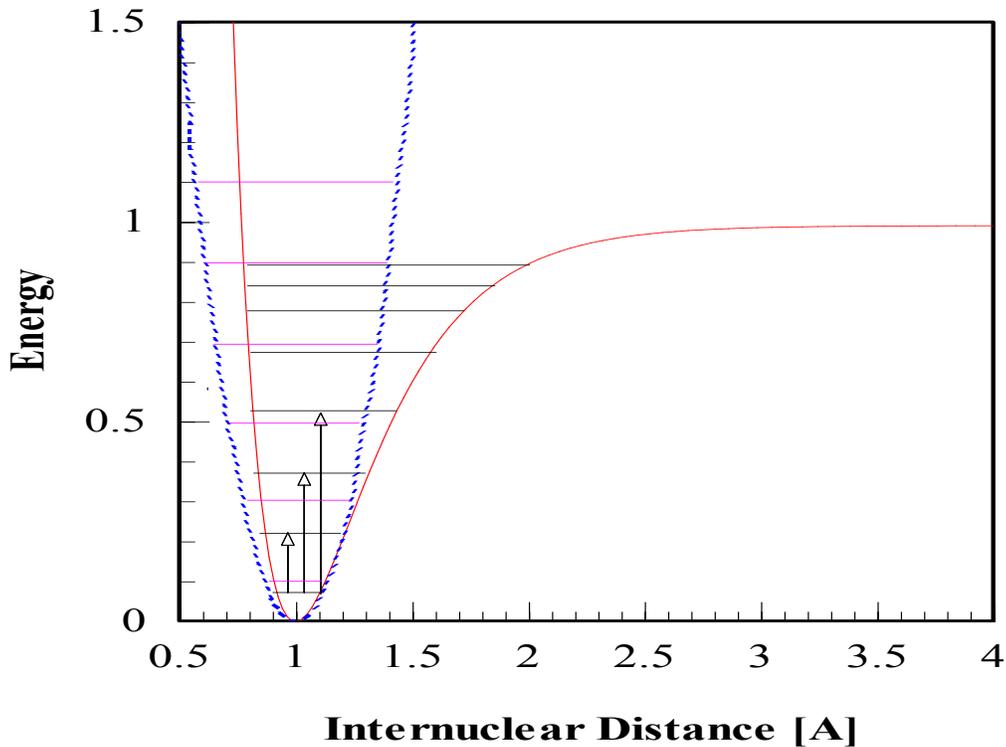


Figure 2-2: Graph showing the harmonic and anharmonic oscillator and their energy levels. Blue dotted line shows the harmonic oscillator with equally spaced energy levels and the red line showing the Morse potential with narrower energy gaps for higher energy levels

The transitional changes are represented as an absorption of the specific quanta of energy from the incident radiation. The frequency range for vibrational energy changes depends on the composition of gases and liquids. Hydrocarbons like polypropylene (PP), which is the substrate of choice within this thesis, and H₂O in liquid form are highly absorptive in this region as shown in Figure 2-3.

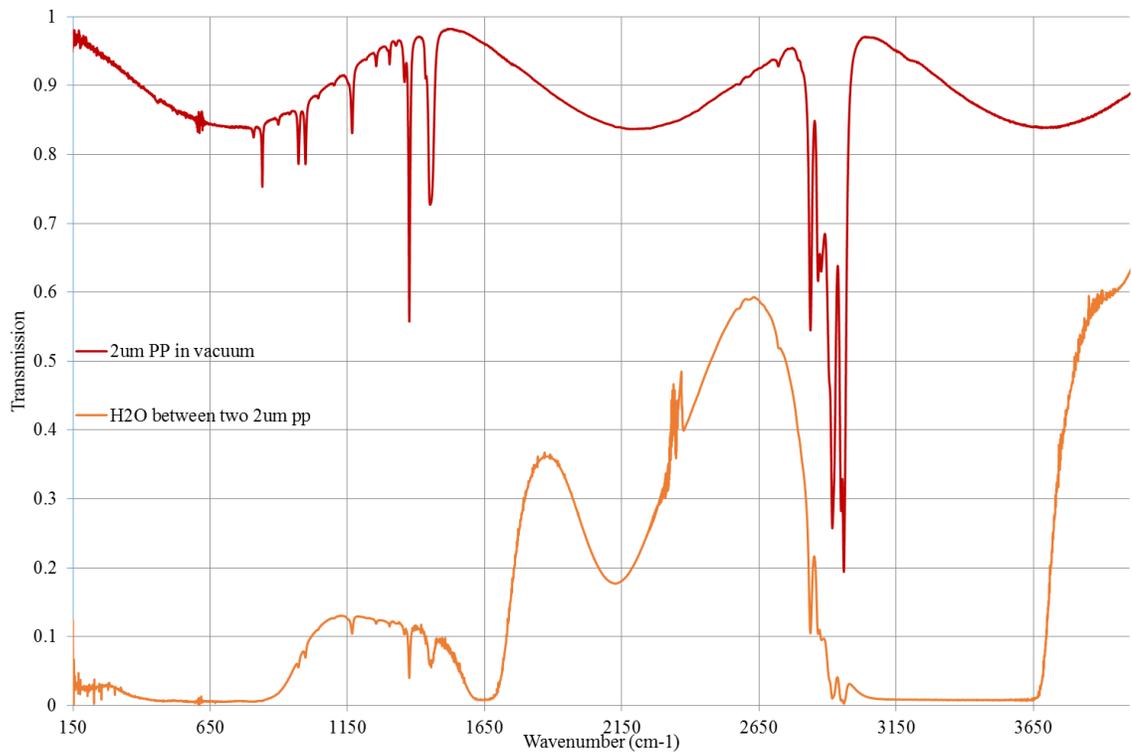


Figure 2-3: Infrared spectra of 2 μm polypropylene (PP) sheet taken under vacuum and H₂O encased between two layers of 2 μm of PP taken within atmospheric conditions. Both spectra have been ratio-ed against the corresponding background.

Figure 2-3 has a number of absorption features. It shows a very strong absorption from liquid H₂O, the most essential molecule for life on Earth. Infrared spectroscopy is an essential tool in identifying many essential molecules on Earth as well as other application such as astronomical applications in identifying whether an extrasolar planet (exoplanet) is habitable or not. However, with this strong absorption from water as well as the other components that make life on Earth, it is difficult to conduct any experiments on larger essential biological molecules.

2.2 Rotational spectroscopy

It has been shown that the vibrational energies of a molecule are quantized, but molecules can also possess rotational movement which is also has a defined energy. The simplest form is a rigid diatomic molecule with two atoms of masses m_1 and m_2 as seen in Figure 2-1. Consider a point between the two masses, at its centre of gravity, where the molecule rotates end-over-end. The moment of inertia about this point will be defined by

and as the molecule is rotating around its centre of gravity, the moments around this point will be _____, and the moment of inertia will be given by Equation 2-16.

$$\text{_____} \quad \text{2-16}$$

Just as for the vibrational energy states, here we also use the time-independent Schrodinger's equation but without the potential energy as we are assuming a rigid rotator.

$$\text{_____} \quad \text{2-17}$$

The angular momentum, L can be expressed as;

$$\text{_____} \quad \text{2-18}$$

where P_J is the angular momentum and $J = 0, 1, 2, \dots$ is the rotational quantum number and is always positive or zero. J represents the total angular momentum i.e. rotational spin as well as electron spin. This also has a quantization of its rotational angular momentum along its axis which is given by

where μ_B and obeys μ_B . The corresponding energies is given by Equation 2-19.

$$E_J = \frac{h^2}{8\pi^2 I} J(J+1) \quad \text{2-19}$$

If this is compared to the classical angular momentum equation,

$$L = I\omega \quad \text{2-20}$$

the classical angular momentum increases as the rotational quantum number increases. For longer wavelengths and far infrared spectroscopy, most of the data is measured in frequency or wavenumber, therefore Equation 2-20 must be divided by either h or hc .

$$\frac{E_J}{h} = \frac{h}{8\pi^2 I} J(J+1) \quad \text{2-21}$$

Using Equation 2-21, the allowed energy states can be determined, for example, at $J=0$ (lowest state), which means that the molecules are not rotating at all, but at $J=1$, which is the lowest angular momentum the molecule can have. To derive the selection rules between one state J' and another J'' , Equation 2-17 needs to be transformed into polar coordinates which will give the resulting wave solution. The results verify that, for this molecules, there is only a need to consider the transition in which J changes by one unit. Therefore the selection rules are which means the angular momentum can only be changed in units of .

2.2.1 Non-rigid rotator

For the rotational case, the molecules were assumed to be rigid which is not true as the line spacing will decrease with increasing J due to the stretches incurred in response to the centrifugal forces. In a similar way to the vibrational energy and the anharmonicity constant, there is a centrifugal distortion constant which is included in Equation 2-22 which gives;

2-22

However, diatomic molecules such as H_2 , N_2 , O_2 , etc. are said to be symmetric which means that they have no dipole moment which makes them infrared inactive. For a molecule to be infrared active, it needs to have a net change in the dipole moment so when it interacts with an electromagnetic wave at the same energy and frequency, the radiation will be absorbed, causing a change in the amplitude of the molecular vibration.

2.3 Vibration-Rotation Spectroscopy

Vibrational and rotational energies occur at different areas of the frequency spectrum. But according to the Born-Oppenheimer approximation, it is assumed that the total energy of a vibrating-rotating system is given by;

2-23

and by combining the rotational and vibrational energies we get;

2-24

However, the energy equation must also take into considerations effects of each component on each other and make adjustments accordingly. As the vibrational state changes, B will decrease as the vibrational quantum number increases because of the increase in the moment of inertia. Simplifying the equation leads to ignoring D (centrifugal distortion) and x_e (anharmonicity constant) as both are small and have little effect on the actual spectra.

2.4 Polyatomic molecules

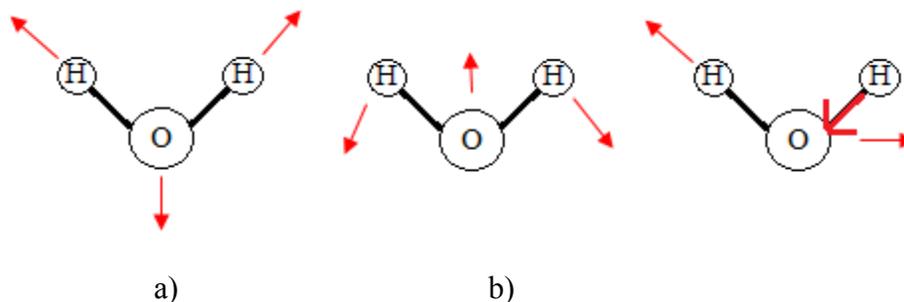
As we know that $\nu = \frac{1}{2\pi} \sqrt{\frac{k}{\mu}}$, we can see that increasing the mass will decrease the frequency of oscillations. Also, by increasing the bond strengths i.e. increasing the force constant k, we see that the oscillation frequency will increase.

2.4.1 Fundamental vibration

If we consider a molecule with N number of atoms, each atom can have three positional coordinates, giving the total number of coordinates as $3N$; it is said that the molecule has $3N$ degrees of freedom. The molecule is also free to move in three directions which means it has translational movement coordinates which uses three of the $3N$ degrees of freedom giving $3N - 3$. As well as this, non-linear molecules can have rotational movement about three perpendicular axes which leaves the molecule with $3N - 6$ degrees of freedom. For a linear molecule, this reduced to $3N - 5$ degrees of freedom due to it only having 2 rotational axes. Therefore, the molecule will have $3N - 6$ fundamental vibrations for a non-linear molecules and $3N - 5$ fundamental vibrations for a linear molecule.

Knowing that in an N -atomic molecule there will be $N - 1$ bonds (for an open chained molecule) between its atoms, then $N - 1$ of the vibrations are bond stretching motions. The other $2N - 5$ (non-linear) or $2N - 4$ (linear) are bond bending motions.

Considering H_2O is a non-linear triatomic molecule, it will have $3N - 6 = 3$ allowed vibrational modes as shown in Figure 2-4.



**Figure 2-4: The three fundamental vibrations of H_2O a) Symmetric stretch 3651 cm^{-1}
 b) Symmetric bend 1595 cm^{-1} c) Antisymmetric stretch 3755 cm^{-1}**

These motions are often called ‘normal modes’ of vibration which are defined as a motion in which all the atoms oscillate with the same frequency and pass through their equilibrium. The frequency of these are often quoted in wavenumbers, $\tilde{\nu}$, which is the inverse of the wavelength, and is measured in cm^{-1} as is shown in the caption for Figure 2-4. It is also easily related to the energy of a photon via Planck’s relation,

In complex or large molecules, all, or nearly all, of the atoms will undergo some kind of displacement. In some modes, all the atoms will undergo approximately the same displacement, whereas in other modes, small groups of atoms will undergo a larger or more vigorous displacement than others. These are usually split into two classes: skeletal and group vibrations.

2.4.2 Skeletal vibrations

Skeletal vibrations involve many of the atoms within the molecule and often arise from branched or chain structures as shown in Figure 2-5.

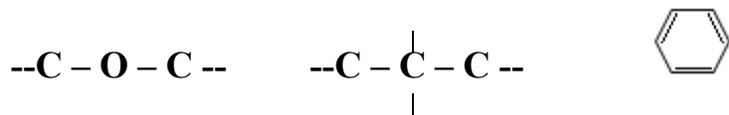


Figure 2-5: Branched and chained structures which may undergo skeletal vibrations

These vibrational modes will often be between 700 and 1400 cm^{-1} . They tend to give rise to multiple vibrational modes and hence several absorption bands which are difficult to individually extract. However, the whole complex of bands tends to be characteristic of a certain molecular structure, therefore this infrared region is often referred to as the “fingerprint” region (McCash, 1994).

2.4.3 Group vibrations

Consider an H_2O molecule modelled as 3 masses connected by a spring, with the O-H spring much stronger than the bond between the two H atoms. It is clear that, even with this simple model, if one of the nuclei is displaced more than the others, then the whole molecule will move in a very complicated motion. The form of the normal vibrations that contribute to the overall motion may be obtained from the knowledge of the bond lengths, angle and bond-stretching and bond-bending force constants. However, these calculations are complex when the molecule contains more than two atoms. The use of analytical and numerical calculations is visited in more detail in CHAPTER 4.

In general, a normal vibrational mode is linked to the movement of all the atoms in a molecule, but there can be localised movements in part of molecules. A common example is of a simple ethyl alcohol ($\text{CH}_3\text{CH}_2\text{OH}$) whereby the vibrational wavenumber of the O-H group is almost independent of the rest of the molecule. The hydrogen atom

acts as if it was attached to an infinite mass by a bond with a force constant that is typical for an OH bond. This collective motion is often referred as OH-stretching vibration and occurs at wavenumbers between 3590 to 3650 cm^{-1} , a range which depends on the different phases and a slight dependence on the part of the molecule in the immediate neighbourhood. Another group wavenumber would be the OH-bending at 1050 to 1200 cm^{-1} . These group vibrations are commonly found in the mid-infrared region as shown in a few examples in Table 2.1 (Hollas, 1995).

Table 2.1: Bond-stretches and their group wavenumber from common molecular groups

Bond-stretching	
Group	Wavenumber (cm^{-1})
$\equiv\text{C}-\text{H}$	3300
$-\text{O}-\text{H}$	3600
$-\text{C}\equiv\text{C}-$	2050
$\diagdown\text{C}=\text{C}\diagup$	1650
$\diagup\text{N}-\text{H}$	3350

As the molecule increases in size (greater number of atoms and hence greater volume occupied) understanding the collective vibrational and rotational modes becomes increasingly difficult. Infrared (IR) Spectroscopy has many uses for identification of different molecules e.g. IR spectroscopy of planets can identify the composition of their atmosphere. Whereas IR spectroscopy can tell us about small scale molecular species, microwave frequencies can probe the large scale structures pertinent to biological molecules. However, there is a need to understand larger molecules on Earth. Long chained polymers and large biomolecules will have vibrational, translational and rotational movements and there is a need to understand the combination of these. There is a large spectroscopic gap bridging the microwave region and the IR region which could hold even more information about the combinational vibrational, rotational and phonon modes. This is termed to be the ‘Terahertz region’ and is thought to have a vast amount of potential for spectroscopy of biological molecules.

In order to take experimental spectra and data, a spectrometer is used. There are two types which are in use; dispersive and Fourier Transform types. These days, most applications will make use of FTIR as we have the computing power to handle the data. Specifically, this thesis work has used a Fourier Transform Spectrometer based on a Martin-Puplett Interferometer for a low intensity THz source due to the advantage and access to highly sensitive detectors as detailed in CHAPTER 3.

Not only can the Fourier theory transform from time domain to frequency domain, it can also perform an inverse FT to be transformed back to its original domain. In processing from time into the frequency domain, the inverse FT is performed.

3.2 Fast Fourier transform

The Fast Fourier Transform (FFT) is the algorithm that computes the Discrete Fourier Transform (DFT) as explained above. It was the development of the FFT combined with advancements in computing power that is responsible for such rapid transformations (Series, 1965). This is an essential tool in signal and image processing and is a key component in using the Michelson or Martin-Puplett interferometer as described in section 3.4.

If we consider a simplified version of the Michelson interferometer, there is a source, one moveable, another fixed mirror, and a detector. For simplicity, consider a monochromatic light with a cosine function entering the interferometer with a beam divider at the centre which splits it to two beams of individual amplitude to give –

The intensity of the detected light is given by the combined amplitude of the light returning from both mirrors. The beams will have a different phase as a consequence of the moveable mirror. This is expressed as an imaginary exponential in the form expressed in Equation 3-4 where A is the total amplitude as a function of mirror distance.

3-4

And the corresponding intensity is given by the total amplitude multiplied by its complex conjugate as shown in Equation 3-5,

3-5

where Δr . Here, I_0 is the path difference, I_0 is the intensity at zero path difference (when both mirrors are at the same distance) and I is the intensity with respect to the mirror distance. It is expressed as a cosine function because the wave is symmetric. It can also be written as expressed in Equation 3-6.

3-6

If the source radiated light of a different wavelength at the same time, the interferogram would have a number of \cos^2 curves superimposed.

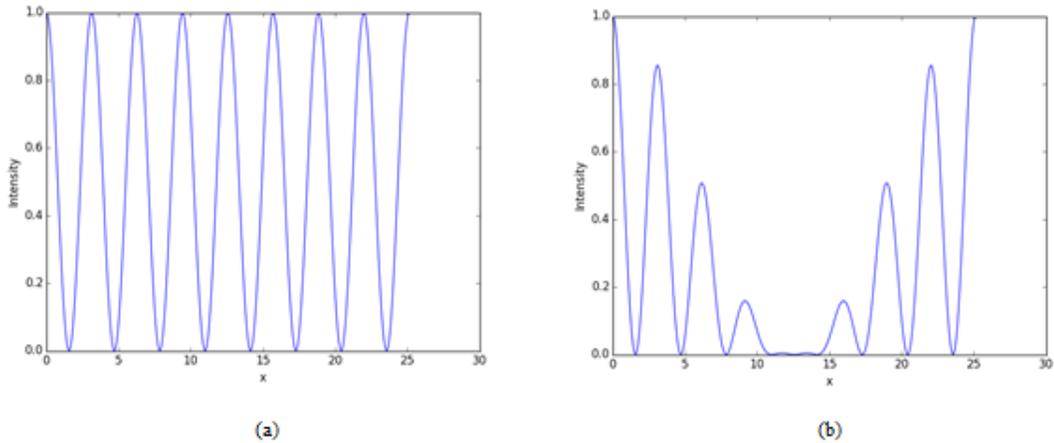


Figure 3-1: Interferograms for (a) monochromatic light and (b) two different frequencies

Figure 3-1 (a) shows a \cos^2 curve with the maximum when the path difference — — . However if there are two different frequencies, the resulting interferogram will look like that in Figure 3-1(b). This method of ‘visibly’ identifying the frequency was used by Michelson to determine hyperfine structures of fairly simple cases in the 1890s (Thorne, 1974) (Jacquinot, 1960). But as more frequencies are added as is in the case for a continuous light source, with difference phase, it is not possible to identify the frequencies without Fourier Transform.

In order to get the intensity with respect to the wavelength, the Fourier integral of Equation 3-6 is given by Equation 3-7.

3-7

This can be converted into an integral over the intensity with respect to the path difference giving;

3-8

Resolving power is an important factor and can be determined by the path difference of the instrument. If a monochromatic light at wavenumber k_l is used, then the interferogram is given by the cosine distribution much like Equation 3-5.

3-9

The Fourier Transform for a maximum path difference of Δ would be given by;

$$\int_{-\Delta}^{\Delta} \cos(k_l \Delta) \cos(k_l x) dx$$

3-10

Assuming that the wavelength is much smaller than the path difference, (which is almost always the case), then the first term is negligible and the equation becomes

This gives a central maximum at $k=k_l$, two zero points either side at $k=k_l \pm \Delta k$ and a resolving limit (from the FWHM) of Δk . However, the minima lie below zero and have a second maxima either side at 15% of the intensity of the central maxima which could be problematic when other frequencies are involved. Apodization of the function can solve this issue at the expense of the resolving power by converting Equation 3-11 into a sinc^2 form, which results in the first minima at $k=k_l \pm 2\Delta k$ and a resolving limit of $2\Delta k$.

When applying the Fourier Transform to the interferogram, the original integration limits suggest that the mirror movement will go from zero to infinity which is not possible. It is therefore limited from zero to the maximum path difference, $2L$. This looks like a top hat function and the Fourier Transform spectrum will show large ripples at its sides similar to that seen in Figure 3-2. However, these can be removed by issuing a triangular function which has a sinc^2 form and results in the first minima at $k=k_l \pm 2\Delta k$ and a resolving limit of $2\Delta k$. This removes the ripples at either side of the central maxima and the result is seen in Figure 3-3.

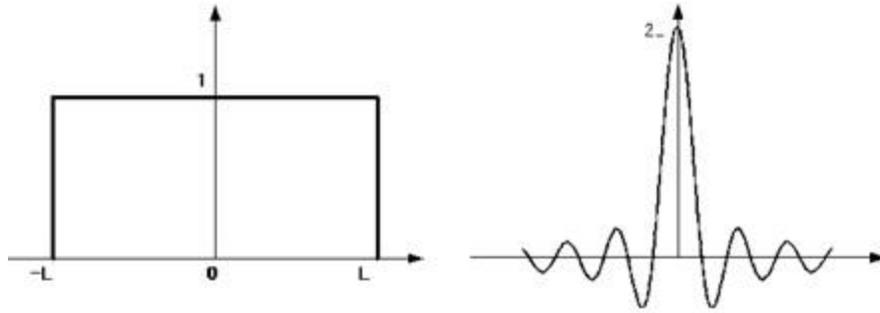


Figure 3-2: Box-car function with the Fourier transform showing the side ripples prominently

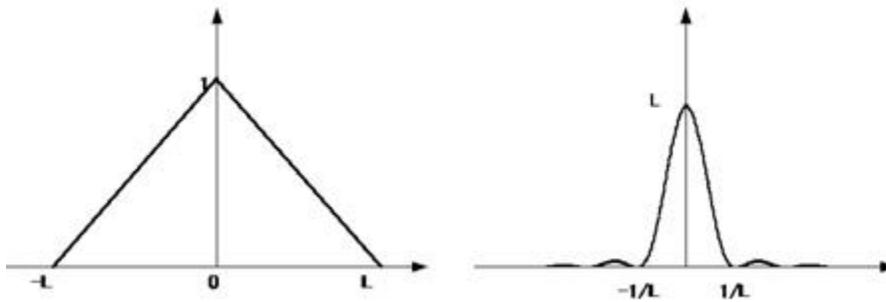


Figure 3-3: Triangular function showing the ripples being suppressed.

In order to have high resolution, the optical path difference must be large and is measured in discrete intervals rather than in one continuous function. If

where and are the maximum and minimum in the spectra, respectively, and is the resolution limit, then the sampling interval can be expressed as Equation 3-12.

$$\text{---} \quad \text{---} \quad \text{---}$$

3-12

3.3 Advantages

3.3.1 Throughput / Jaquinot advantage

Using a Fourier Transform Spectrometer means that the power and frequency from the source is continuous, passing through the interferometer and to the detector. Whereas in other diffracting spectrometers, the power from the source would be limited by the size of the grating. This increase in the incident power to the detector leads to a higher signal-to-noise ratio and it is easier to achieve high resolution. (Bell, 1972)

3.3.2 Multiplex/Fellgett advantage

As a FTS has a high throughput and contains all the wavelengths within each scan, these scans can be combined for signal averaging. In a grating spectrometer, each spectral element is recorded over a suitable integration time, t . If there are N elements, the total time to gather the full spectrum is Nt . However, within a FTS, each N element contains all the spectral information therefore, the integration time for the same signal is t and the entire scan will take only take time t . This improvement in the scanning speed improves the signal-to-noise ratio. Although most of the noise issues will arise from fluctuations within the detector, the noise equivalent power (NEP) is inversely proportional to the square root of the integration time. As the FTS takes the same time to scan all the elements as the grating does to scan each element, the NEP will be a factor of \sqrt{N} lower in the FTS improving the signal to noise ratio by \sqrt{N} . This is known as the Multiplex or Fellgett advantage. (Thorne, 1974)

3.3.3 Connes advantage

Many dispersive spectrometers use two spectra, usually a sample and a background or reference spectra. In using a FTS, it eliminates this by using a single beam where all the spectral elements are measured simultaneously as shown in the Jaquinot advantage. Connes advantage arises due to having this ability of knowing the frequency scale of the spectrum very accurately. The control of the wavelength domain through mirror movement means that it is possible to gain “mirror-movement averaging”, increasing the signal-to-noise of the spectrometer. It also allows very accurate spectral subtractions and ratioing any unwanted background spectral information.

3.3.4 Nyquist sampling

Nyquist sampling theorem/ Nyquist rate is the minimum amount of points that is needed to sample the signal. It should be at least twice the maximum frequency. It assures that there is a point at least twice in one cycle to provide a correct replication of the signal. The more the number of points, the better.

3.4 Martin-Puplett Interferometer

The Martin-Puplett interferometer is essentially the same as the Michelson interferometer except for the use of a wire grid polariser as a beam divider rather than a half-silvered mirror or a dielectric. This is advantageous as the modulation efficiency is higher and uniform over a broader spectral range and there is no need for a corrector plate (compensator). For a grid spacing d , the spectral range is from $0 \sim 1/4d$. Although the spectral content of the interferometer is determined by the other filters used for the desired spectral range. Figure 3-4 is the set-up of the interferometer used within the FTS lab at Cardiff University and uses a double sided polarising beam divider on a $4\mu\text{m}$ polypropylene substrate.

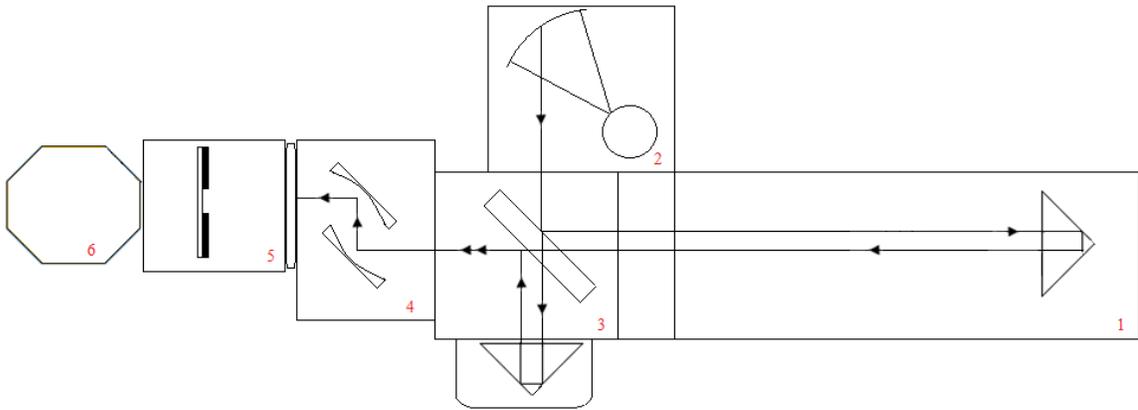


Figure 3-4: Schematic diagram of the permanent Michelson interferometer in the FTS lab at Cardiff University

The labels in Figure 3-4 correspond to the chambers shown in the photographs in Figure 3-5.



Figure 3-5: Images of the FTS taken when it was opened to show the internal structure. The images have been labelled from 1-5 corresponding to the sections in the schematic diagram in Figure 3.4.

The source is a Mercury arc lamp which peaks in the UV but the high intensity UV light is filtered using evaporate Ge (Germanium) on 4 μ m polypropylene, made at Cardiff. At the interface between the source box and the beam divider there is a polarising filter with the grids orientated horizontally. At the centre there are two wire grids made by photolithographically etching a 1 μ m period pattern onto each sides of a 0.7 μ m Mylar substrate making a double sided polarising beam divider. By making the beam divider double sided, the incident light is transmitted through the Mylar once for each arm and there is no need for a compensator. These were aligned at 45° to allow 50% transmission and 50% reflection to both arms of the interferometer. The reflected beam is directed towards the moveable rooftop mirror whereas the transmitted beam goes to the fixed rooftop mirror. The rooftop mirrors allow a full 90° rotation of the incident electric field vector at each mirror which then allows what was transmitted to be reflected at the beam divider and vice versa, regaining the full intensity of the beam. After being directed by a

pair of parabolic condensing mirrors, the beam exits through another polariser orientated with respect to the input polarisation and enters the sample chamber before entering the detector chamber. There is an option to further optically filter the beam by placing THz filters either in the sample chamber or at the detector. The detector output is connected to a lock-in amplifier which is used to electrically filter the signal and vary the signal amplification as and when needed; the output from this goes to a computer using data acquisition via LabView. The optical filters and detectors used within this spectrometer and in many other ground-, balloon- and space-based instruments, are made within the Astronomy Instrumentation Group (Peter A.R. Ade, 2006).

The high quality drive used for the moveable mirror is extremely accurate to allow very precise sampling measurements. The maximum optical path difference (OPD) allows for potentially high resolution spectroscopy. The quality of the mirrors used are also of a high standard for THz spectroscopy to allow for optical alignment using a HeNe laser. There are purchased from Symons mirror technology (Symons Mirror Technology).

3.5 Portable FTS

To better understand the applications and effects of high intensity THz sources, a portable FTS system was built to characterise the THz beam from the LINAC at the ALICE (Accelerators and Lasers In Combined Experiments) facility in Daresbury. This will give spectral information about the beam as well as characterising the full extent of the source.

The portable FTS was built in-house by the Astronomy Instrumentation Group at Cardiff University. The dimensions of the set up were limited to the size of the shielded case at

the exit of the THz line within the diagnostics room and are shown in millimetres in Figure 3-6.

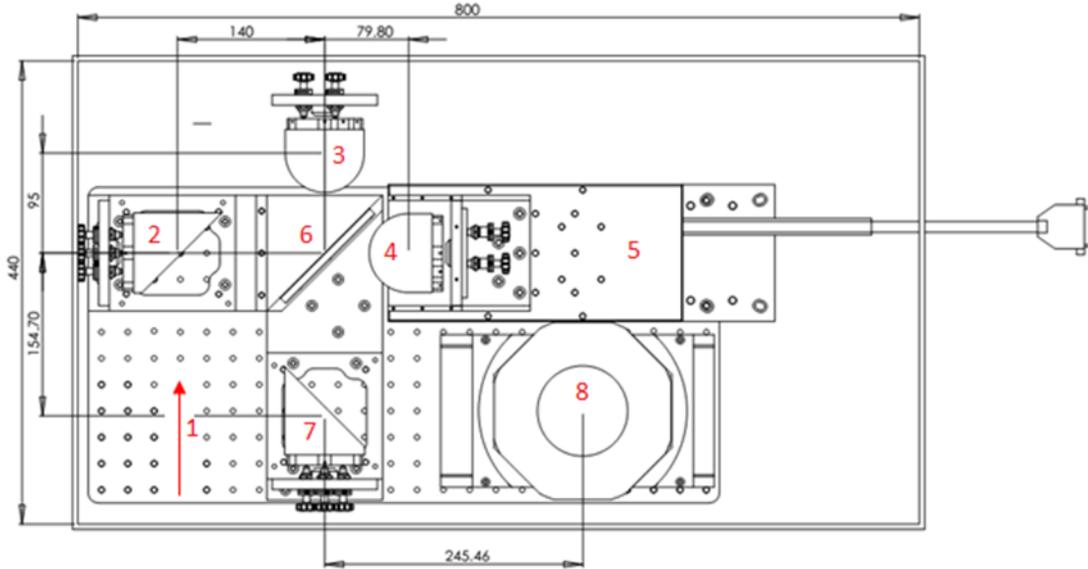


Figure 3-6: Bird's-eye view of the portable FTS including the dimensions in millimetres drawn on CAD

Highlighted in this image are the key components of the FTS and are described in Table 3.1.

Table 3.1: Identification of the numbered items on the portable FTS diagram

1	Source. At Daresbury this was the THz source from the LINAC therefore the set up would have been inside the metal casing within the lab. To calibrate and align the FTS, a black body source was used which was set at $V = 5.05V$ and $I = 2.97A$.
2	90° mirror.
3	This is the stationary rooftop mirror.
4	Moveable rooftop mirror.
5	Drive and the stage for the moveable mirror. It has a maximum distance (OPD) of

	20cm, giving it a resolution of 0.05 cm^{-1} (1.5 GHz).
6	Beam divider. This can be changed as the filters have been placed into mounts that screw perfectly into the holder. There are two – $5 \mu\text{m}$ Mylar dichroic or $2.5 \mu\text{m}$ double sided polarising beam divider.
7	90° mirror
8	2 detectors. Pyroelectric detector; two horns were tested (F/4.5 with a diameter of 10mm and F/3.5 with a diameter of 15mm but with a 10mm removable cap) and Indium Antimonide (InSb) hot-electron bolometer.

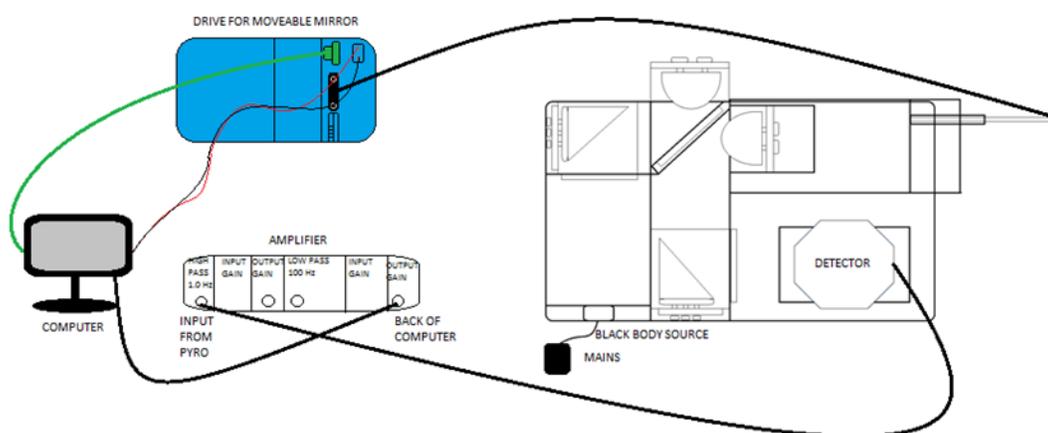


Figure 3-7: Set up of FTS in Cardiff for alignment and calibration

For calibration and alignment of the FTS at Cardiff, the set-up seen in Figure 3-7 was used. The source used was a filament black body source supplied by Interspectrum (IR source, 5.0V screened by a ceramic sleeve) (Interspectrum). At the input of the first mirror, there is a polarising filter with the grids orientated horizontally. After reflecting 90° from the first mirror, there is a double sided beam divider using $2.5\mu\text{m}$ Mylar. By making the beam divider double sided with grids on both sides, the incident light is transmitted through the Mylar once for each arm as seen in the diagram in Figure 3-8.

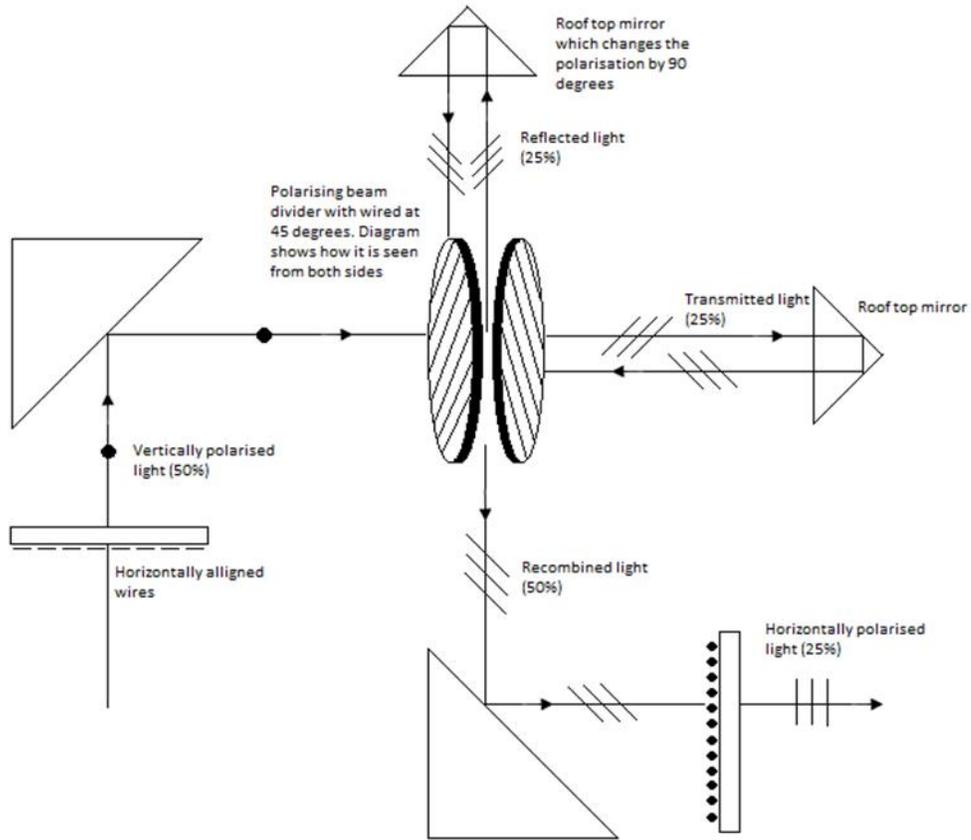


Figure 3-8: Schematic diagram of the polarisation of the light through the spectrometer

All the mirrors (including the rooftop mirrors) have three screws at the back to adjust the position of the mirrors and a laser was used to align all the components. The moveable rooftop mirror was controlled by the drive seen in Figure 3-7 and had a maximum optical path difference (OPD) of 20 cm. Once aligned, the laser was removed and the black body source was used. Using the pyroelectric detector meant that it had a lower signal to noise and a slower response compared to the InSb hot-electron bolometer. The Pyro was connected to the high pass input socket ‘A’ at the lock-in amplifier. The pyroelectric

detector has a Winston optical cone and two were tested, F/4.5 with a diameter of 10mm and F/3.5 with a diameter of 15mm but with a 10mm removable cap.

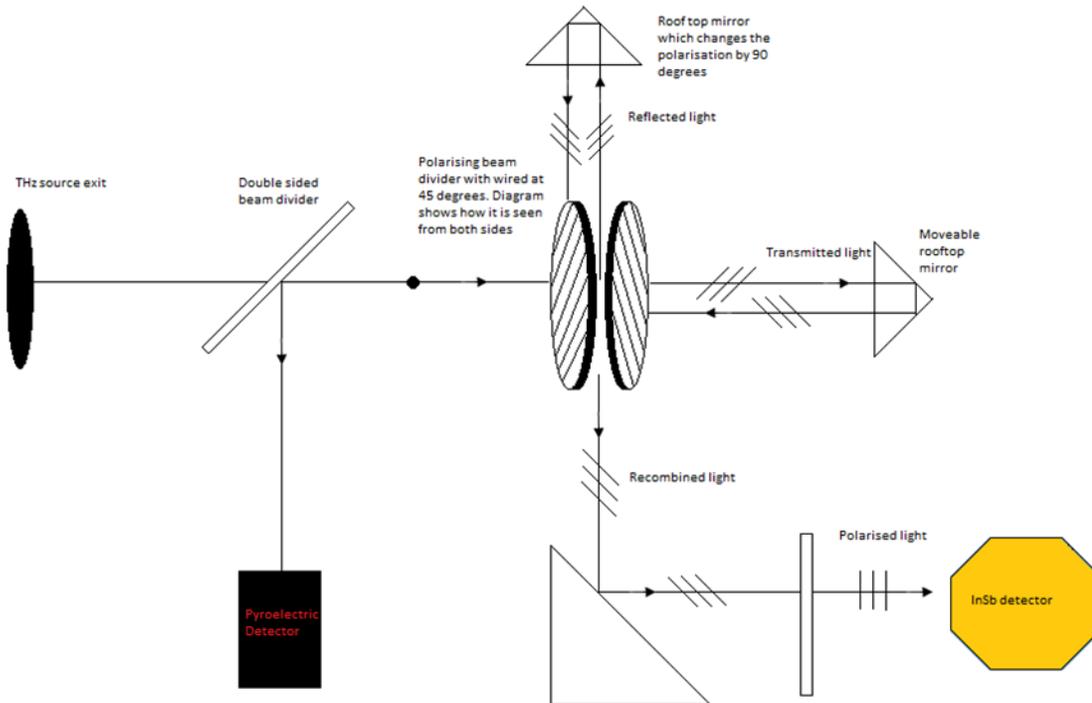


Figure 3-9: Schematic diagram of the FTS set up in the diagnostic room at Daresbury

3.6 Electronics

The Martin-Puplett FTS system is kept under vacuum, with one vacuum pump for the main body of the FTS including the source chamber, the centre beam divider chamber, the two arms, and another vacuum pump for the sample chamber although this is often vented and evacuated when changing the sample for the experiment. The portable FTS is open to atmospheric conditions as its prime purpose was the investigation of the THz beam at ALICE which had a low maximum frequency of 1 THz.

Control of the FTS is done via a computer with a LabView program which controls the movement of the drive and the acquisition of the detected signal. A calibration mode facility allows the user to see the interferogram, spectrum, and power from the sample and can be altered by changing the gain on the amplifier. Once the correct settings are made, the sample can be scanned and recorded. The raw interferogram is saved onto the hard drive and can be Fast Fourier Transformed on DOSbox (emulator programme or a host/command-line program) which has a pre-programmed code for the FFT.

3.7 Detectors

To be able to detect the lower THz frequencies, there is a need for sensitive detectors. A crucial component of the Cardiff FTS is the highly sensitive detector systems. It is specifically this that allows a Black Body source in a conventional FTS to be used for THz work. Different detectors were used over the course of the research, used especially for different reasons.

3.7.1 Pyroelectric detector

This was the detector of choice when using the Bio/portable-FTS within the Cardiff labs and at Daresbury. It was supplied by QMCI, UK (QMCI, 2009) and is operational at room temperature which eliminates any use of cryogenics. A pyroelectric detector's basic concept is that it has a pyroelectric crystal which is sensitive to polarisation changes when the temperature of the crystal changes the positions of the atoms. This polarisation change will give rise to a voltage difference across the crystal.

3.7.2 Indium Antimonide (InSb)

This is a high sensitivity InSb hot electron bolometer which operates at 4K or below. This detector was used for the characterisation of the THz beamline at ALICE in Daresbury as it has a better signal to noise than the pyroelectric detector. The disadvantage over the pyroelectric detector is that it only works at cryogenic temperatures resulting in cooling it down with liquid nitrogen and helium, and it significantly larger than the pyroelectric detector.

3.7.3 Germanium

Gallium doped germanium gives a better speed of response. This is used for higher frequencies than the InSb and operates at temperatures of 4K or below. This was the detector of choice in the FTS laboratory before the advancement of the superconducting detector (section 3.7.4). It was necessary to cool down with liquid nitrogen (77 K) and liquid helium (4 K).

3.7.4 TES pulse tube (cryo-free)

Transition Edge Sensor (TES) system takes advantage of the properties of superconductive niobium. This niobium thermistor is attached to a silicon nitride (SiN) absorber. The thermistor is held at cryogenic temperatures where it held between the normal and superconducting state i.e. a state of no resistance. This makes the resistance highly sensitive to small changes in temperature, making it an extremely sensitive thermometer and the ability to be sensitive to low energy (low frequency) signals. Some quantitative comparisons are shown in Table 3.2.

Table 3.2: Quantitative differences between the detectors

	Pyro	InSb	Ge	TES
System optical NEP	0.44 to 1.25 nW Hz ^{-1/2}	<1.5pW Hz ^{-1/2}	0.8pW Hz ^{-1/2}	<2pW Hz ^{-1/2}
Useful frequency range	0.3 to 20 THz	60 GHz to 1.5 THz	0.6 to 20 THz	0.1 to 20 THz

3.8 Background measurements and Fourier Transform

The “background” is the signal without any sample within the FTS, essentially just the beam. This will be used to ratio against the sample to remove any background noise. It would also be done if anything would be changed within the FTS set up, such as differently sized aperture.

Once the background scan has finished, this will produce an interferogram. To gain the Fourier Transform of this, the interferogram must be transformed using an in-house code written in Fortran. The background is Fourier Transformed and is saved for ratio-ing against the sample signal.

The sample is measured in the same way. The interferogram is transformed using the Fourier Transform code and saved. A different program is used to ratio the sample’s Fourier Transformed signal against the background’s Fourier Transformed signal. This removes any anomalies or noise that is found in the background and results in a spectrum of the sample itself.

In most cases, the sample is on a polypropylene substrate. The data for the substrate is taken in the same way as described previously, and the Fourier Transform of this is also taken. The spectrum for polypropylene in the FTS is shown in Figure 3-10.

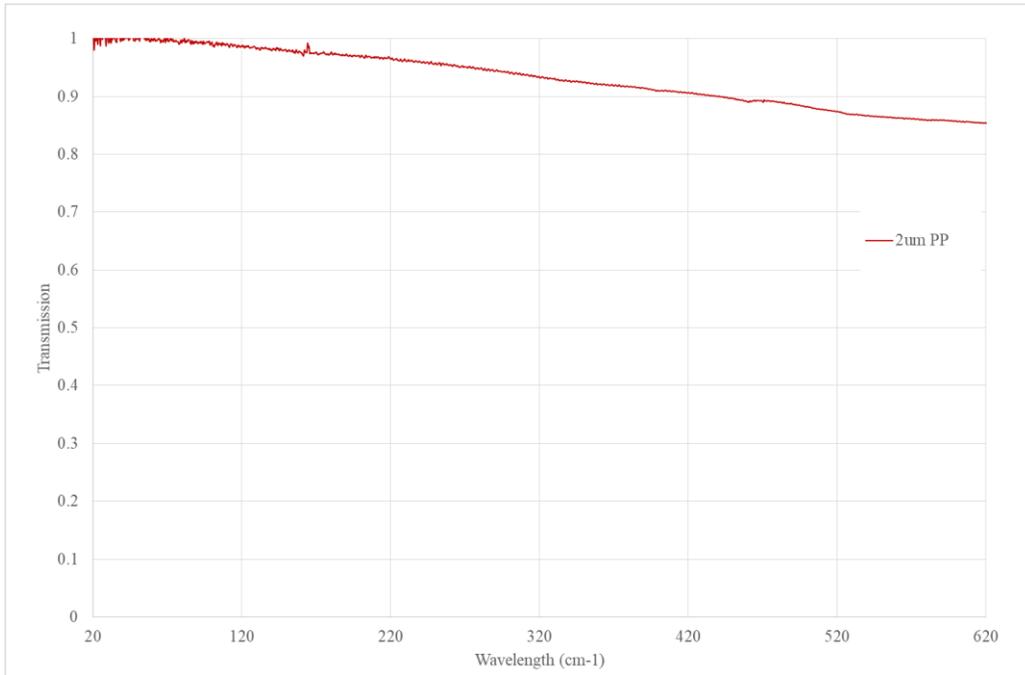


Figure 3-10: Terahertz transmission of 2 μm polypropylene (PP) from 20 to 620 cm^{-1}

The spectrum for 2 μm PP is ‘clean’, showing no interesting absorption features and only a decline in transmission towards the higher frequencies. The small amount of absorption is due to the Fabry-Pérot effect. This sinusoidal pattern represents the planar nature of the sample which acts as a Fabry-Pérot Etalon creating coherent waves when the sub-mm radiation is incident on the sample. However, at low frequency ranges such as the range used within the FTS, the sinusoidal patterns are over a wider range, resulting with a decrease in the material’s transmission.

CHAPTER 4. MOLECULAR MODELLING

While experimental scientists are working on learning about the structure and content of molecules through their electromagnetic interactions, there is a need for a complementary theoretical approach to model the physical basis of these effects. Computational chemists visualise molecules in terms of bond angles, bond lengths and dihedral lengths and work towards a structured molecule that acts just like a real one under the same conditions. Predictions can be made using quantum mechanics from first principles but this proves to be very difficult and very computationally demanding for many-bodied molecules.

Due to the complicated nature, computational theorists have taken the force field approach, where the electronic degrees of freedom are ignored and only the movement of the nuclei matter. As shown in CHAPTER 2 a molecule can have $3N-6$ ($3N-5$ for a linear molecule) degrees of freedom where N is the number of molecules. The Born-Oppenheimer approximation states that the electronic and nuclear motion can be separated on the assumption that (1) the electronic waveform depends on the nuclear position but not their velocities, as they move considerably slower compared to the electrons, (2) the rotation and vibration of the nuclear motions refer to the electrons' potential as being "smeared" as they move so quickly. (HyperPhysics, n.d.)). The force fields used have at least five constants that are measured by experiment or other measures. These include bond angle geometries such as bond stretching, compressing and

bond bending, dihedral motions and the non-bonded interactions, which are molecular force fields between one molecule to another which are usually Leonard-Jones 12-6 type potential energies (Hinchliffe, 2008). Each individual component is a term shown in Equation 4-1,

$$4-1$$

where E_{total} is the total Force Field energy, E_{stretch} is the energy due to the bond stretching, E_{bend} is the energy from bond bending, E_{torsion} is the torsional energy from the rotation around a bond, E_{vdw} and E_{elec} are the non-bonded terms (Van der Waals and electrostatic energy) and the E_{couple} is the coupling term between the three first terms. More information about each component and their function can be found in (Jensen, 2006).

Although force fields and molecular mechanics have been shown to give good descriptions of the potential energy density, it results in a macroscopic description, despite treating molecular spectroscopy in quantum mechanical terms. Despite this, the molecular mechanical force field approach has been used successfully when comparing to biological molecules (Tieleman, 2012).

In this chapter, I will give a brief overview of the current numerical methods employed to model and simulate the electromagnetic interactions with large biological molecules. This will be the approach of choice for the nucleobases and nucleosides that will be investigated and presented in CHAPTER 5.

4.1 Numerical approach

Density functional theory (DFT) is another way of calculating approximate solutions to a many-bodied system by using the electron density to investigate the structural, magnetic and electronic properties of molecules and materials.

D.R. Hartree (1920) assumed that the nucleus and electron of an atom formed a spherically symmetric field i.e. the charge distribution of the electron was used as the solution to the Schrödinger equation for the atom's potential energy (Ewels, 1997). By using this approximation, Hartree postulated the idea of the self-consistent model to give approximate solutions to the wave equation and solve this for a many-bodies problem. But later on (1930) Fock and Slater jointly pointed out that Hartree's solution forbade electrons in the same quantum state and had not considered quantum statistics i.e. that electrons with opposite spins can occupy the same state.

The combination of methods resulted in the Hartree-Fock (HF) method which is a way of approximating the wave function and energy of a many-bodied quantum system in a stationary state. It was essential to express the wave function of a multi-fermionic system that satisfies the anti-symmetry requirement, by using Slater's determinants. As electrons (fermions) have a spin of $\frac{1}{2}$ and an antisymmetric wave function, they must obey the Pauli's exclusion principle. As we are dealing with a multi-electron problem, Slater's determinant is an expression that will satisfy the requirements and consequences of the Pauli principle by changing the sign when two electrons (or fermions) occupy the same area. (For example, if two electrons were occupying the same state with the same spin, the wave function would be zero.)

The independent particle model assumes that the motion of one electron is independent of the motions of other electrons. One way of approximating the interactions between particles is using the HF theory by taking the average of all the interactions. Each electron is described as an orbital and the total wave function is given as the product of the orbitals. The shape of these orbitals gives the probability of finding an electron by including the attraction to the nuclei and repulsion of other electrons. As each electron has its own descriptive orbital, the HF equation for each electron has its own solution and must be solved iteratively. The resulting equations can then be written into a matrix eigenvalue problem and solved using the Fock-matrix diagonalization.

After evaluating the HF equations, the model can go two ways; in some cases an extra approximation can be added leading to semi-empirical methods; or it can be improved by adding more determinants, which will help converge the model towards an exact solution to the Schrödinger's equation.

Semi-empirical methods simplify the problem by neglecting the integrals involving more than two nuclei in the construction of the Fock matrix. Therefore the semi-empirical method relies on turning the remaining integrals into parameters and fitting them to experimental data. Although this is slightly analytical in its approach, it does produce good outcomes with accurate results and is computationally less demanding.

The HF method only deals with electron-electron interactions and does not take into account the correlation between electrons. To model the correlation effects, a multi-determinant solution must be used which is computationally much more demanding but

can give a systematic approach to the exact solution of the Schrödinger's equation (CHAPTER 4, (Jensen, 2006)).

Density functional theory (DFT) is similar in its approach to the HF model only it produces better approximations by modelling the many-bodied electron correlation by a function of the electron density. Ab initio (meaning from the beginning) methods such as energy correlation methods, try to derive solutions to the Schrödinger equations without any prior fitting parameters or information about the sample. This is possible but is incredibly difficult, time and computationally demanding. As X-ray crystallographic data is available for many crystalline samples (including nucleobases and nucleosides in their dehydrated state), it is possible to use parameters such as lattice dimensions, atomic positions and symmetry to help with the initial setup for the simulation.

The ground state electronic energy can be determined from the density of electrons and is given by E. B. Wilson (Harrison) who stated that;

- The integral of the density gives the number of electrons.
- The cusps in the density define the position of the nuclei.
- The height of the cusps define the corresponding nuclear charges.

Electron density is the square of the wave function integrated over $N-1$ electron coordinates. The energy of an electron or the system is a functional as it is dependent on the wave function or electron density which is in turn dependent on a function of the coordinates. Although this method does not take into consideration the electron-correlation and fails to calculate the exact solutions for the Schrödinger equations, the electron correlation is such a small fraction of the total energy that approximations for this term provide quite accurate results.

The main disadvantage with DFT is the inability to systematically improve the results and the failure to describe some features such as the Van der Waals interactions. This might be a cause for concern with regards to the reproducibility of the data, however, according to (King, Ouellette, & Korter, 2011), thorough analysis is needed to ensure that the calculated DFT data is reproducible. This is performed using the root-mean-squared deviations (RMSDs) of the calculated bond lengths instead of using the exact value given by the X-ray crystallographic data.

To enhance the approximations and solutions from DFT, the ‘implicit DFT’ is a theory that uses energy functions that are expressed in terms of the Kohn-Sham orbitals rather than the density orbitals from the electrons. Kohn-Sham equations are relevant as they are the set of Schrödinger equations for a fictitious system of non-interacting particles (electrons) that generate the same density as any given system of interacting particles. The equations are defined by a (modelled) local effective external potential from the non-interacting particles. (i.e. the Kohn-Sham equations are a set of Schrödinger equations that approximate the external potential effects on a specific electron from the other non-interacting electrons). This incorporates an approximation to the correlation effects.

The Local Density Approximation (LDA) is often used. This is a simpler model which takes a region where the density is only varying slightly and slowly so that the electron-correlation can be approximated for a uniform electron density. Gaussian (Frisch, et al., 2017) is a software for computational chemistry which uses Gaussian orbitals rather than Slater-type orbitals, as this was computationally less demanding back in the 1970s when it was first introduced by John Pople (Carnegie Mellon University, 2009) . The most recent version is Gaussian09.

Modelling and simulating large molecules, especially biological molecules that are largely influenced by their surroundings is very difficult to replicate. As computational power and technology is always developing and improving, better models and faster simulations have been made resulting in better comparison to experimental results. In 1970, (Susi H, 1970) investigated the properties of polycrystalline uracil, N,N-dideutero uracil, CC-dideutero uracil, and perdeutero uracil using Raman and IR spectroscopy. They were analysed using force field calculations.

According to (Susi H, 1970), with regards to the uracil molecules, zero order calculations with force constant values “reproduced the experimental frequencies within acceptable limits and lead to a reasonable potential energy distribution.” To help characterise the uracil modes, a potential energy distribution (PED) was calculated. This is used for modes which involve more than one internal displacement co-ordinate. Using this form allows classifying the vibrations generally but does not determine the form of the vibration using PED alone.

In more recent papers, investigations in nucleobases and biomolecular constituents have been shown to agree well with molecular mechanics and simulations. In 2002, far-infrared vibrational modes of DNA components were studied by THz time-domain spectroscopy (B M Fischer, 2002). DFT was employed for thymine using software package Gaussian 98. B M Fischer et al. calculated the absorption and index of refraction of thymine between 0.5 and 3.5 THz by assuming that each resonance was a damped oscillator and took the sum over all the different oscillation contributions. The position and intensity of each line was given from the DFT calculations, with a line width of 0.1THz chosen to best compare to their experimental data. Both sets of data were in

agreement but did incorporate many adjustments from experimental results which can be misleading. However, Fischer et al. only presented simulated data for the thymine molecule accounting for only one of the four (or five) of the nucleobases. Also, it was limited to a maximum frequency of 3.5 THz because of the TDS system that was used. The work presented in this thesis will cover a broader range from 0.6 to 20 THz with very high signal to noise at the lower frequencies.

Furthermore, (King, Ouellette, & Korter, 2011) investigated THz spectra and Solid-State DFT results of DNA nucleobases. Experimental data were taken using a pulsed Terahertz spectrometer. The samples were 1-methylthymine, 9-methyladenine and a co-crystal of both thymine and adenine in a 1:1 ratio. Once grown (left to evaporate at room temperature) the crystals were pressed into pellets with PTFE. Data for each sample was taken between 10 and 100 cm^{-1} (0.3 to 3 THz). It was compared to the theoretical results using DFT calculations performed on the CRYSTAL09 software package which is an ab initio computational tool for solid state chemistry and physics using initial atomic positions and lattice dimensions from the x-ray crystallographic data. Both experimental and simulated results were in agreement however, the spectral limitations of the THz spectrometer used can result in poor signal to noise and some missing absorption features as is explored in more detail in this thesis (section 5.1)

As well as nucleobases, some understanding of the dynamics of DNA and long chained polymers becomes complicated yet important. In the early 1990s, a model for the dynamical stretching of DNA was proposed by (Thierry Dauxois, 1993). They simplify the structure of a DNA molecule by representing the nucleosides as point masses along the DNA chain, only taking into account the transverse motions and that two neighboring

nucleotides on the same strand are connected by a harmonic potential. In this model, the structure is modelled as a harmonic oscillator but with “springs” attached to neighboring bases, as well as helicoidal springs at the points where the DNA coils. A good explanation of the model and its implications is given in (Zdravković, Satarić, & Tuszyński, 2004). It applies a Morse potential to the hydrogen bonding between base pairs and an inter-pair stacking potential. Since the bonding interactions of the sugar-phosphate backbone are much stronger than that of the base pairs, the degrees of freedom of the backbone are neglected. This method has resulted in good approximations (Alexandrov, Gelev, Bishop, Usheva, & Rasmussen, 2010) and (Swanson, 2011) for example. However, this model is treating the system as a classical harmonic oscillator. The boundary between classical and quantum systems and the correct approach as to which is used to model them is very abstract. This poses many complications in the application of simulations to large biological molecules such as nucleobases, nucleosides and DNA molecules which are investigated in this thesis. In section 5.2 I will present the results from the DFT applied to 5 nucleobases.

CHAPTER 5. RESULTS

It is proposed that THz frequencies can be used to probe the dynamics of DNA, however firstly it is important to better understand the subunits of DNA and how these lead to its basic structure. The single strands of DNA (ss-DNA) are attached through the hydrogen bonding between the complementary bases, adenine- thymine, cytosine- guanine to form the double stranded DNA (ds-DNA). These pairs will always bond to each other, with different sequences along the DNA providing specific genetic instructions for creating cells in the body. DNA is found in every cell within the body and is constantly undergoing some process of replication, but during this process spanning millions of subunits, replication errors can arise. Despite the human body being extremely efficient in repairing miscoded DNA, some mutations can be passed without recognition and repair leading to replication of an incorrect sequence into many daughter cells which can eventually form a cancerous tumour.

Ultimately, the sequencing and bonding of the bases can result in the mutation of DNA. To better understand the dynamics of DNA, THz spectroscopy with its sensitivity to long chain molecules and hydrogen bonds, has the capabilities of identifying characteristic features in the nucleobases and nucleosides which could potentially hold the answer to the early diagnosis of cancer.

In section 5.1.1 I will present the THz spectroscopy of the four nucleobases in solid form made using different solvents to investigate the solvation effects and the crystalline structure. Additionally, in section 5.1.2, attention is paid to the spectra for the bases with

similar structures, the purines and pyrimidines. To extend this work, the structure of nucleosides, which are the nucleobases attached to a 5-carbon sugar ring, are investigated in the solid state and are compared to the bases in section 5.1.3. These data represent a full spectral picture of solid state DNA base units at broader bandwidth, better spectral resolution and signal to noise than limited recent data elsewhere.

Crucially, data can be compared to simulations on the same units, using Density Functional Theory (DFT) modelling. The comparison of experimental results and modelled spectra are given in 5.2.

As the common form of DNA is either ss-DNA or RNA, or ds-DNA, it is essential that the THz spectra of DNA and RNA is investigated. By drying the DNA on a substrate the THz spectra of human and mouse DNA is presented in section 5.3.1. The difference between mouse and human DNA samples are investigated in section 5.3.2 with the results of a controlled set of mouse samples in 5.3.3 and the effects of heat and time on the dried mouse DNA in section 5.3.4 and 5.3.5. As DNA is naturally found hydrated within the body, in section 5.4 experiments were made to investigate the THz spectra of water (H_2O), heavy water (D_2O) for better identification and the nucleobase guanine in H_2O . Finally, in section 5.5, the investigation of animal tissue is performed to identify whether a THz probe could be developed that could potentially identify signatures within tissue *in vivo*.

5.1 Nucleobases and nucleosides

5.1.1 Nucleobases made using different solutions

5.1.1.1 Method

All the nucleobases were purchased from Acros Organics UK in powder form. All the bases were individually mixed with the solvent being investigated with the same ratio for each base. The amount of liquid solvent was consistent for all which was 10 ± 1 ml and was added to the specific weight of powder as shown in Table 5.1. Accurate to the weight of each base, as shown in Table 5.1, the solution concentrations were in the range of 12.49 ± 0.50 g/L to 17.47 ± 0.50 g/L. However it should be noted that the evaporation process used to produce solid samples for the different bases resulted in samples of variable optical depth.

Table 5.1: The concentration of each powdered nucleobase for each of the different liquid solutions

SAMPLE NAME	CONCENTRATION (± 0.50 g/L)				
	IPA	DI WATER	PBS	HBSS	HEAVY WATER
Guanine	12.49	12.98	13.06	13.01	13.01
Cytosine	13.16	13.20	13.19	13.28	13.25
Adenine	14.57	16.28	14.28	14.72	14.20
Thymine	17.22	17.38	17.47	17.36	17.25

The solvents used were: De-Ionized water (DI water); Isopropanol (IPA); Phosphate-buffered saline (PBS); Hank's balanced salt solution (HBSS); Deuterium Oxide (D₂O or heavy water). PBS and HBSS are often used by biologists as a cellular solvent to mimic the environment and pH within the body.

Water and IPA are polar molecules, whereas the DNA bases are have less polarity which means that they will not dissolve in the solution. This results in them aggregating, as it is more energetically favourable for the water molecules to bond to each other rather than bond, via van der Waals interactions, with the nucleobases. This true for all the nucleobases.

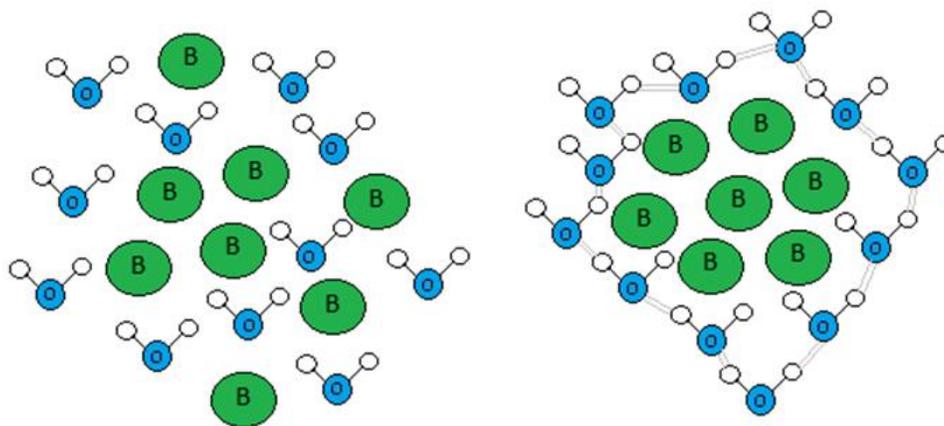


Figure 5-1: A schematic depicting how bases (B) aggregate when placed in a solution which is represented as water molecules H₂O

These potential variations in bonding sites and the presence of trapped solvents will affect crystalline structure which should be probed by THz spectroscopy. Samples were pipetted onto disks of 2µm thick polypropylene (PP). Once all the solvent had evaporated the powdered sample was left attached to the substrate. Once placed in the Fourier

Transform spectrometer (CHAPTER 3), the sample chamber was evacuated and the process of spectral analysis could begin. Each base was normalised against an instrument background and data are presented in the form of individual graphs split into further subsections to show the results of using different solvents for the same solute.

5.1.1.2 Guanine

Guanine made using the five solvents is shown in Figure 5-2 and the data is normalised against the spectrum of the background. Each sample is made so as to completely cover a 5mm aperture. The interferogram is sampled at every 1.6 μm and the resulting spectral resolution is $\text{---} = 0.25 \text{ cm}^{-1}$.

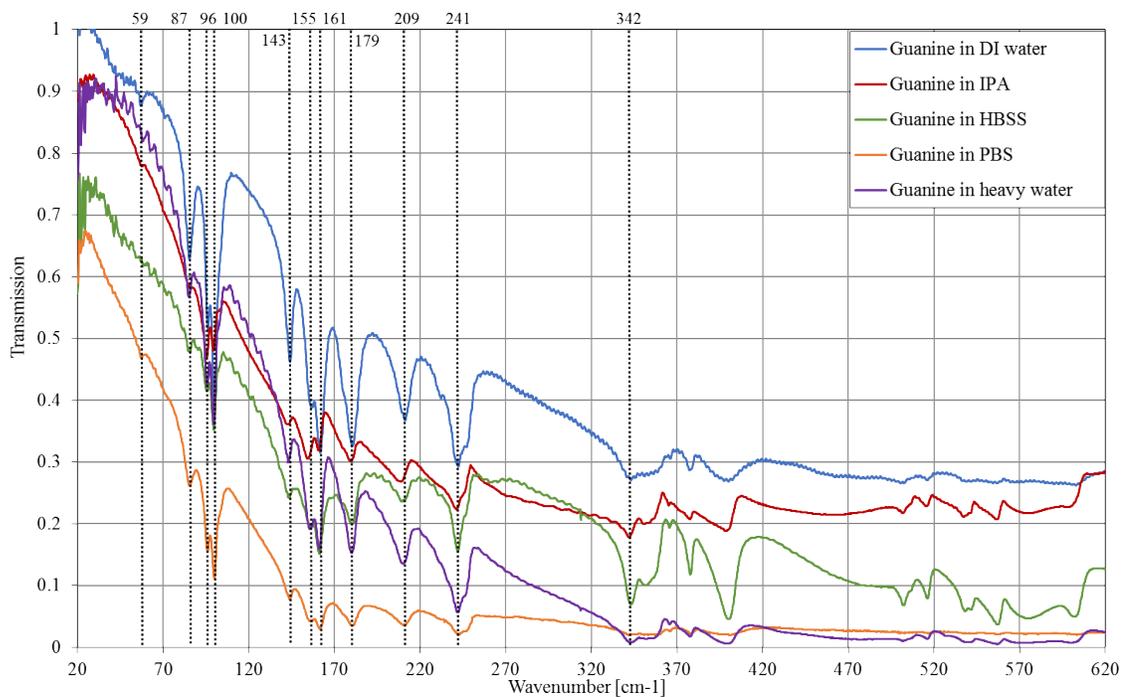


Figure 5-2: THz spectra of guanine made using DI water, IPA, heavy water, HBSS and PBS from 20 to 620 cm^{-1}

All the samples in Figure 5-2 show a decrease in transmission with increasing frequency, especially in the region above 270 cm^{-1} . This could be due to 1) sample thickness as guanine was difficult to dissolve and resulted in a white opaque sample seen in Figure 5-3 and/or 2) scattering effects due to the particle sizes. All the absorption features identified compare directly to each other within error of $\pm 2 \text{ cm}^{-1}$.



Figure 5-3: Photograph of guanine made using DI water (left) and IPA (right)

The rate at which the solutions evaporated varied, as well as the aggregation of the guanine molecules in different solutions. When the powder is added to the solvent, the hydrophobic nature of the molecules can cause a variable thickness and greater optical depth after evaporation.

(Nishizawa J, et al., 2005) measured THz transmission spectra of nucleobases based on difference frequency generation of widely tuneable THz waves using a Gallium Phosphide (GaP) crystal. This technique determined the absorption features of guanine as shown in Table 5.2.

Table 5.2: Frequency of the resonances identified in the THz spectra of guanine using GaP wave generator by (Nishizawa J, et al., 2005) and samples made using DI water and IPA taken on the FTS presented in this thesis.

GaP THz wave generator Wavenumber (cm ⁻¹)	Absorption features from FTS	
	DI water (cm ⁻¹)	IPA (cm ⁻¹)
	57.3	59.9
85.7	86.6	86.6
100.0	96.1	95.9
	99.9	99.9
143.7	142.9	142.9
	155.5	154.7
161.3	161.1	160.9
181.3	179.8	178.1
	208.7	209.2
	241.1	241.1

The frequencies of the absorption features noted on the right-hand side of Table 5.2 correspond to the features highlighted in Figure 5-2. Clearly there is good agreement with (Nishizawa J, et al., 2005) and the data presented here. The data presented by (Nishizawa J, et al., 2005) was also compared to time domain THz spectroscopy data taken by (Fischer, Walther, & Jepsen, 2002) which was in agreement. However, both were limited in their spectral resolution meaning that they were unable to identify the additional features found at 57.3, 96.1 and 155.5 (± 0.5) cm⁻¹. They were also bandwidth limited to a range between 0.5 to 4.0 THz (16.7 to 133.3 cm⁻¹) according to (Fischer, Walther, & Jepsen, 2002) and between 0.4 to 5.8 THz (13.3 cm⁻¹ and 193.3 cm⁻¹) according to

(Nishizawa J, et al., 2005) whereas the technique used in this work (see section (FTS)) can operate between 0.3 to 20 THz (10.0 to 666.7 cm^{-1}) with the capabilities of extending further into the mid infrared using a Bruker Vertex 70 (Bruker, 2017). Using (Nishizawa J, et al., 2005) method will only achieve 0.4 to 4.0 THz (13.3 to 133.3 cm^{-1}) due to significant losses in signal after 3.5THz as the thermal radiation from the pump laser is comparable to the THz power generated (Plusquellie D.F, et al., 2003). Although the frequency identification of the guanine absorption features have been made in these references, there was no attempt in either in identifying the resonances other than hydrogen bonds.

When the samples are made, some guanine molecules will bond to the solvent molecules through hydrogen bonding. When the samples are left to evaporate, it is possible that some molecules become trapped in the crystalline structure. An example of guanine structure is shown in Figure 5-4 from (Lopes R. P, 2012) using plane wave (PW) calculations to model the solid state conformation of guanine using density functional theory (DFT) methods within the Perdew-Zunger local density approximation (LDA). Assuming that the molecular structure of the solid sample obeys simple harmonic motion, any additional mass due to the solvent's molecules bonding to individual guanine molecules, can change the resonant frequency.

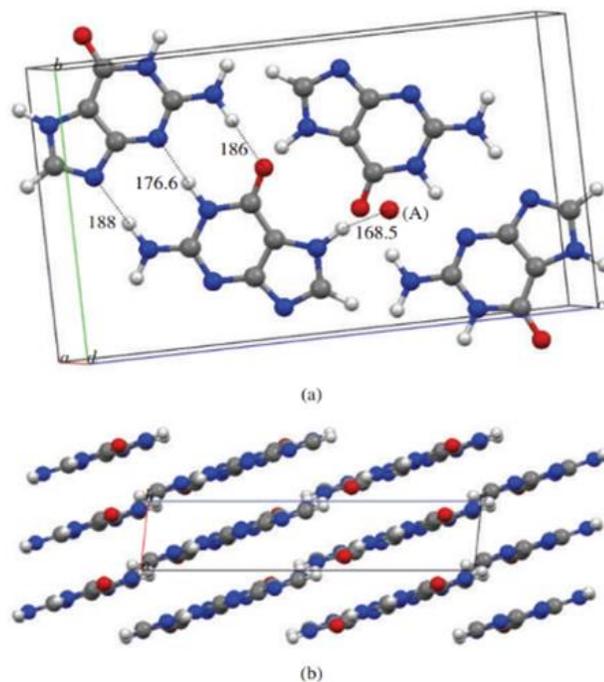


Figure 5-4: (a) Optimised crystal cell structure of solid guanine calculated using LDA functional and PW methodology. The dashed lines along with the numbers represent the intermolecular H-bonding distance (measured in picometers). The distance between the two pairs of dimers is about 310.2 pm. (A) Oxygen atom from the upper guanine molecule. The remaining atoms were omitted. (b) Lateral view, along the b-axis, of the optimised unit cell (Lopes R. P, 2012).

However, the absorption features for all the guanine samples occur at the same frequencies which suggest that there are no differences in the crystalline structure of the sample. The difference in the overall attenuation at higher frequencies due to the grain size aggregation suggests that different solvents can cause scattering effects and effect the overall optical density.

5.1.1.3 Cytosine

Guanine's complimentary base pair is cytosine and was made into solutions using the same solvents as shown in section 5.1.1.1. Unfortunately the "Cytosine in HBSS" sample became loose and repeatedly fell off within the vacuum chamber before any reasonable data could be taken. Figure 5-5 shows the spectral results for cytosine made using four solvents.

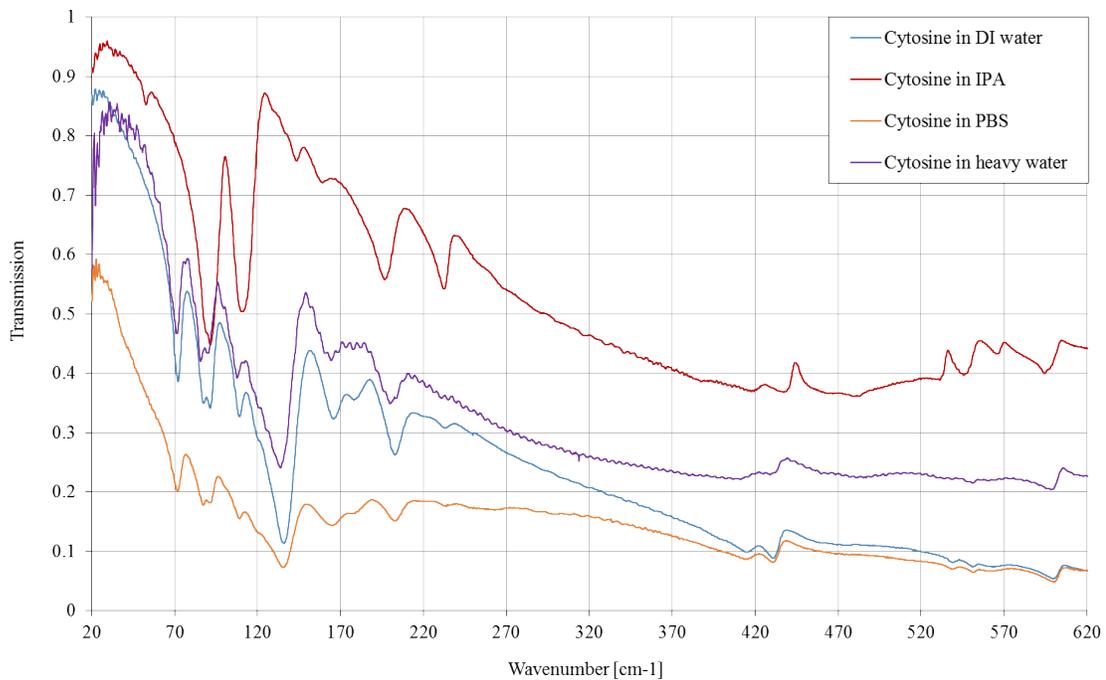


Figure 5-5: THz spectra of cytosine made using DI water, IPA, heavy water, and PBS from 20 to 620 cm^{-1}

Figure 5-5 has very interesting results. There are clear differences in the spectra for cytosine made using IPA compared to the other solvents. The picture in Figure 5-6 shows the difference in the structure of the samples after being made with DI water and IPA.



Figure 5-6: Photograph of cytosine made using DI water (left) and cytosine made using IPA (right)

With reference to Figure 5-5, there is a considerable difference in the spectrum of the cytosine made using IPA which suggests that the difference in the crystalline structure of the samples can affect the THz spectra. To better examine the differences between the features found in Figure 5-5, the frequency of the most apparent absorption features in the 20 cm^{-1} to 320 cm^{-1} region are shown in Figure 5-7.

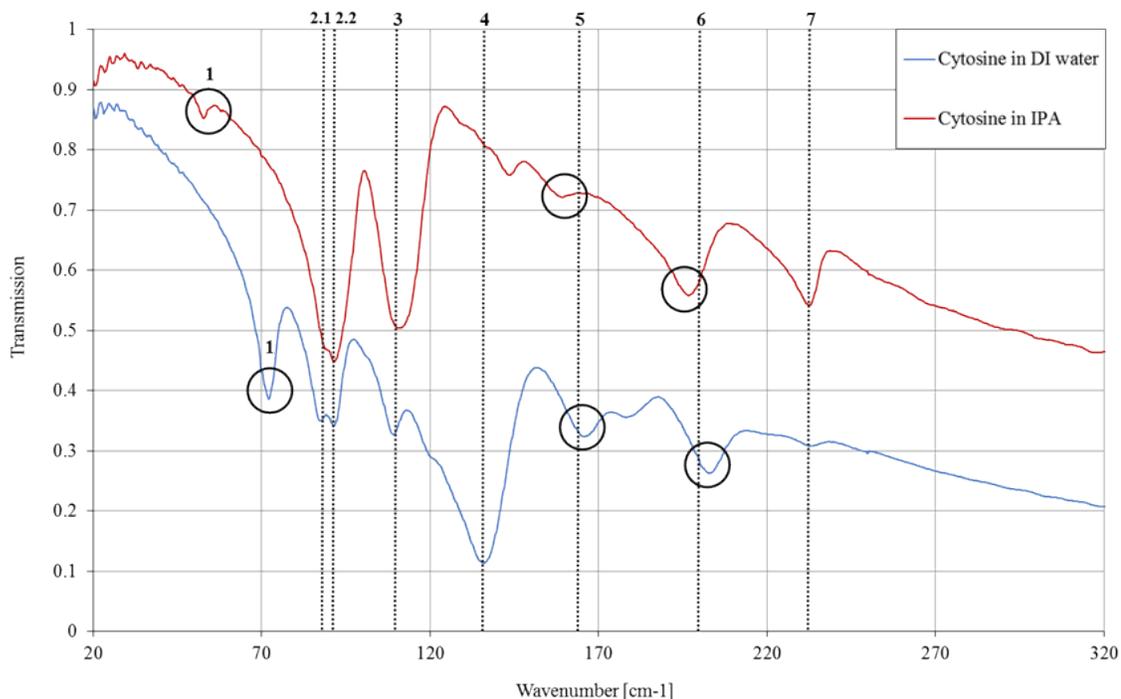


Figure 5-7: THz spectra of cytosine in IPA and DI water from 20 cm^{-1} to 320 cm^{-1} to show the main absorption features corresponding to the ones in Table 2

The frequency of the absorption features in the cytosine samples made using DI water, PBS and heavy water are in agreement within an error of $\pm 0.5 \text{ cm}^{-1}$. Therefore, only the frequency of the most prominent absorption features found in the “Cytosine in IPA” and “Cytosine in DI water” were tabulated. In Table 5.3 they have been labelled according to the corresponding features in Figure 5-7.

The labelling system (numbers 1 to 7) has been chosen to match the features that are assumed to correspond to each other. The strong absorption features labelled “2.1 and 2.2” for both spectra accounts for at least two absorption features which could be better defined if the data was taken at cryogenic temperatures.

Table 5.3: Frequency of the resonances identified in the THz spectra of cytosine using GaP wave generator by (Nishizawa J, et al., 2005) and samples made using DI water and IPA taken on the FTS presented in this thesis.

GaP THz wave generator	Absorption features from FTS			
	Wavenumber (cm ⁻¹)	Corresponding feature	Cytosine in DI water (cm ⁻¹)	Corresponding feature
53.33	1	72.07	1	52.67
	2 (1)	87.44	2 (1)	89.21
95.00	2 (2)	91.48	2 (2)	91.48
113.00	3	109.12	3	110.88
144.00	4	135.83	4	143.64
176.67	5	165.82	5	159.02
	6	202.87	6	196.57
	7	233.36	7	232.58

The absorption features identified by (Nishizawa J, et al., 2005) corresponds to the features found in “Cytosine in IPA” within error. This suggests that the samples made in the literature have the same crystalline structure as the samples made with IPA. Their sample was milled for 10 minutes with polyethylene (PE) powder and pressed into a 1 to 2 μm thick pellet with a 20 mm diameter. As the samples made using IPA instantaneously evaporate on the substrate, the structure of this sample must be the same as the pure crystalline structure of the cytosine without any further recrystallization.

The first absorption feature identified by (Nishizawa J, et al., 2005) at 53.3 cm⁻¹ corresponds to the initial feature found in “Cytosine in IPA” but it does not correspond to

any of the other samples made using DI water, PBS, HBSS and heavy water. However there are notable similarities in the resonance and/or shifts to the other features. In making the cytosine samples, the crystalline structure of cytosine made using IPA is different to the other solvents. (Barker & Marsh, 1964) confirmed that crystal structures of cytosine made using methanol solvents which were allowed to evaporate at room temperatures, have elongated needles in their structure, but even the smallest amounts of water present would produce the monohydrate crystalline form which was later determined by (Jeffrey & Kinoshita, 1968). Both cytosine monohydrate and anhydrous are looked in to more detail in CHAPTER 4.

5.1.1.4 Thymine

Thymine samples were made using the five solvents with concentrations described in section 5.1.1.1 and the spectra is shown in Figure 5-8.

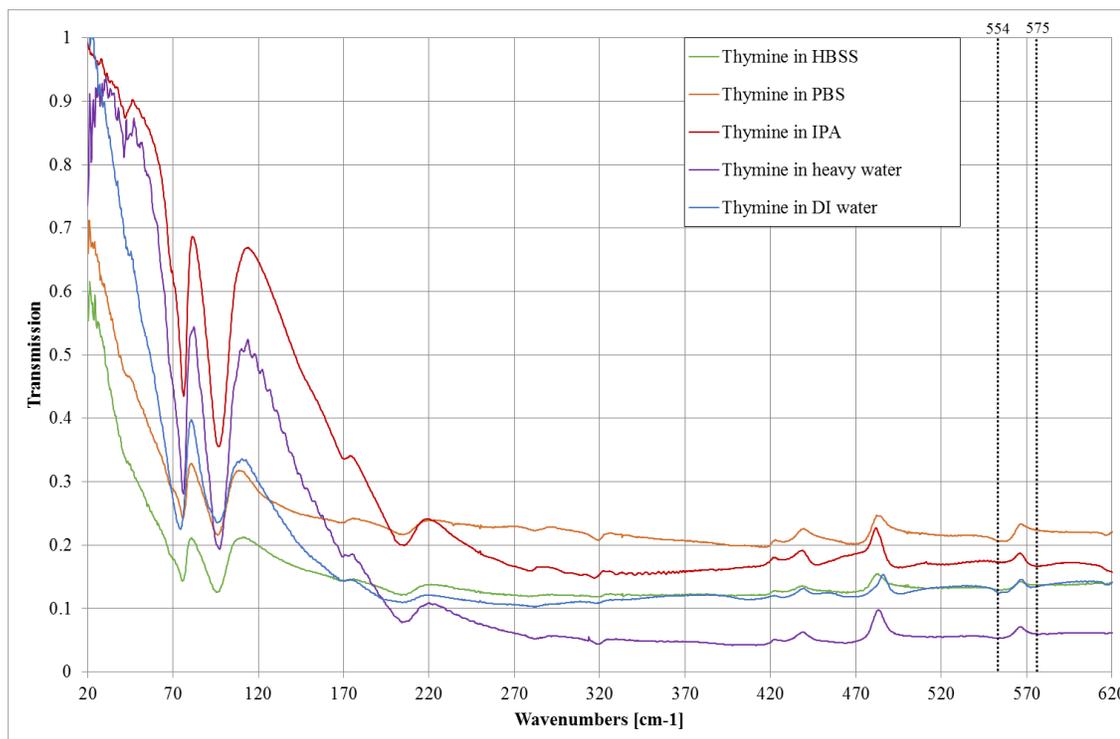


Figure 5-8: THz spectra of thymine made using DI water, IPA, heavy water, HBSS and PBS from 20 to 620 cm^{-1}

The spectra for all five thymine samples in Figure 5-8 display the same absorption features, with two significantly strong absorption features found at 75.1 cm^{-1} and 96.6 cm^{-1} . During the process of making the solutions, thymine showed interesting macroscopic crystalline characteristics for each different solution. This is shown in the pictures in Figure 5-9. When made with DI water, the crystals that formed on the

substrate were large and disordered which could disrupt the data through scattering. This can be the cause of the low transmission in each sample at higher frequencies.

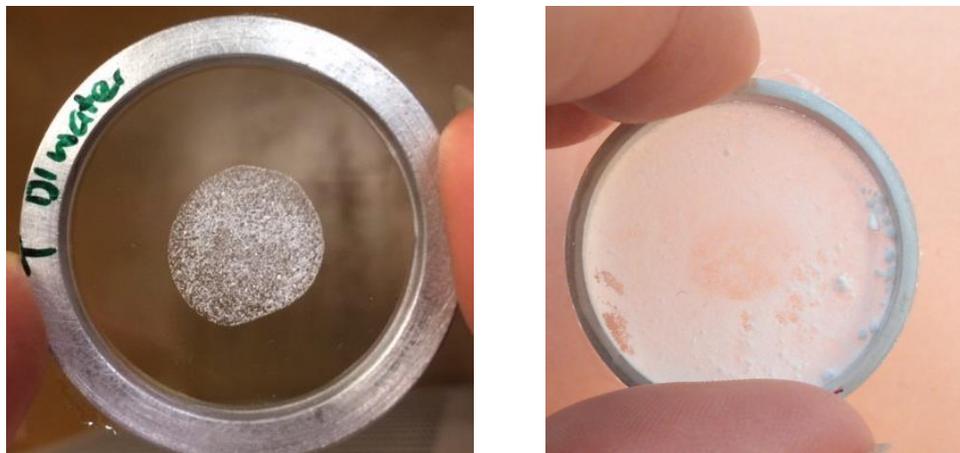


Figure 5-9: Photograph of thymine made using DI water (left) and made using IPA (right)

The thymine made using different solvents was compared using the frequency of the absorption features found in Figure 5-8. The best data sets to compare were the thymine in IPA and in DI water as the features were more prominent. This was predominantly done in the lower frequency region as the higher frequency features are ambiguous due to poor signal to noise ratio because of the scattering effects. Table 5.4 shows absorption frequencies found in the samples made using DI water and IPA.

Table 5.4: Frequency of the resonances identified in the THz spectra of thymine using GaP wave generator by (Nishizawa J, et al., 2005) and samples made using DI water and IPA taken on the FTS presented in this thesis

GaP THz wave generator	Absorption features from FTS	
	Thymine in DI water (cm ⁻¹)	Thymine in IPA (cm ⁻¹)
45.33	43.35	42.08
76.33	74.84	76.36
100.00	96.52	96.77
170.00	170.61	170.61
	192.28	205.39

Unlike the features seen in Table 5.3 for cytosine made using IPA and DI water, the features seen in Table 5.4 for thymine made using both solvents, do not differ by much. Interestingly, the features identified by (Nishizawa J, et al., 2005) in Table 5.4 matched very well to the features identified in Figure 5-8.

There has been a large amount of THz spectroscopy research on the structure of thymine and its isotopic derivatives (Zhang, Michaelian, & Lopnow, 1998) (Fischer, Walther, & Jepsen, 2002) (Laman, 2008). According to (C.P. Beetz Jr., 1980), the apparent bands at the higher frequencies (545 cm⁻¹ and 575 cm⁻¹) noted for thymine in Figure 5-8, are assigned to the in-plane, in-phase bending of the two oxygen molecules. These are not readily identifiable as all the samples have a low transmission at higher frequencies.

The samples from (C.P. Beetz Jr., 1980) were made by pressing the powder between polyethylene sheets or by making them into pellets by grinding the powdered samples

with potassium bromide (KBr). In our case, by making a solution with liquids and then letting it evaporate naturally at room temperature, the crystalline structure is likely to be that of the pressed form.

In an investigation into methylated nucleobases, (Tsuboi, 1967) found two vibrations at 245 cm^{-1} and 275 cm^{-1} for methyl-thymine which have been assigned to the CH_3 torsional mode and are assumed to correspond to the 285 cm^{-1} and 320 cm^{-1} features found in the thymine sample. The difference in the frequency of both features in both methyl-thymine and thymine (35 cm^{-1} and 30 cm^{-1} respectively) seem to be indicative of methyl torsional modes.

There has been a lot of research into the structure of thymine anhydrate and thymine monohydrate as the latter contains a water molecule and can disrupt the crystalline structure of the unit cell (K. Ozeki, 1968), (Gerdil, 1961). Both agree that each thymine molecule is attached to the next via hydrogen bonding between the N-H and the oxygen atom. From the literature, it can be assumed that, when the samples are made from solution, the thymine molecules bond together via hydrogen bonds connecting the NH of one thymine molecule to the oxygen molecule and forms an infinite chain (K. Ozeki, 1968). This might then prohibit the in-plane movement. It shows that different preparation techniques before taking data can affect the transmission spectra and certain modes as the crystalline structure can change.

5.1.1.5 Adenine

The data for adenine in 5 different solutions were also taken and are presented in Figure 5-10.

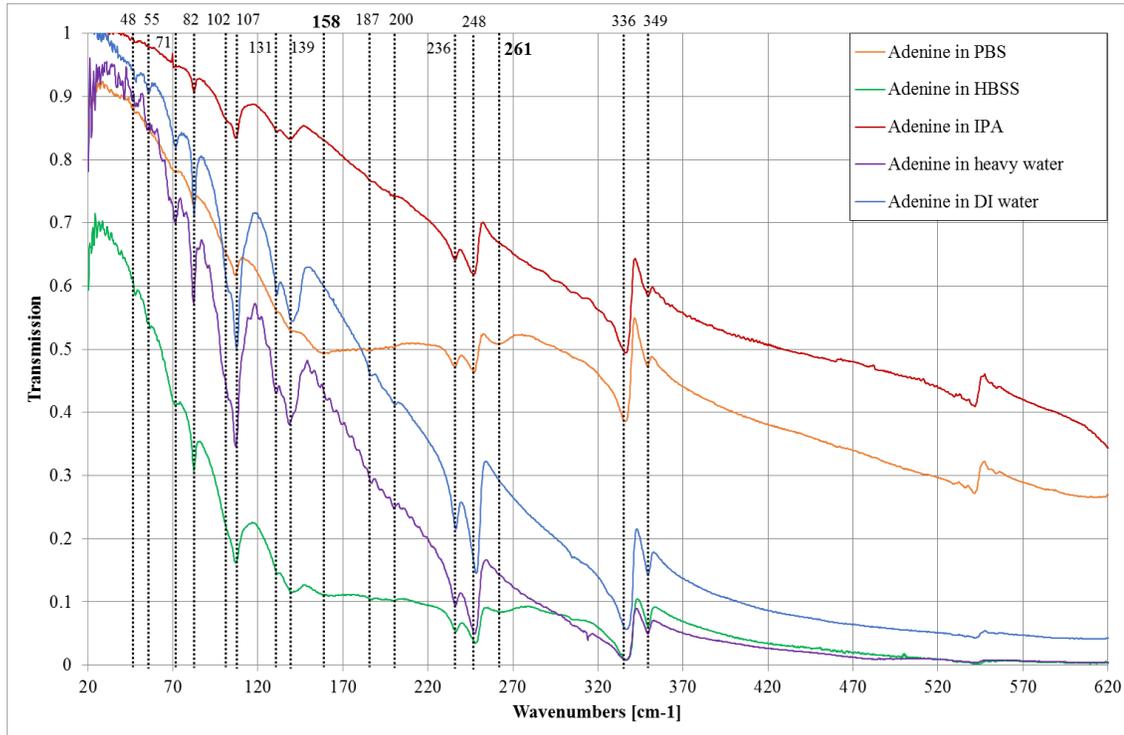


Figure 5-10: THz spectra of adenine made using DI water, IPA, heavy water, HBSS and PBS from 20 to 620 cm^{-1}

Interestingly, the spectral results for adenine made with the different solvents indicate similar features, suggesting that they have very similar crystalline structure. However, the difference in the surface tension of using DI water compared to IPA changes the concentration of the sample on the substrate as shown in Figure 5-11.



Figure 5-11: Photograph of adenine made using DI water (left) and adenine made using IPA (right)

Table 5.5 shows the absorption features found by (Nishizawa J, et al., 2005) and the absorption features presented in Figure 5-10.

Table 5.5: Frequency of the resonances identified in the THz spectra of cytosine using GaP wave generator by (Nishizawa J, et al., 2005) and samples made using DI water and IPA taken on the FTS presented in this thesis

GaP THz wave generator	Absorption features from FTS	
	DI water (cm ⁻¹)	IPA (cm ⁻¹)
	48.05	47.8
55.67	55.76	55.25
70.33	71.44	71.95
84.67	82.49	82.23
101.67	102.82	102.06
114.67	107.16	106.9
129.67	131.32	131.06
139.33	139.54	139.03
	186.83	187.08
	200.45	200.96
	236.43	235.91
	248.25	247.22
	336.91	336.65
	349.76	349.76

The absorption features found in adenine made using DI water and IPA are in agreement with each other and with the features identified by (Nishizawa J, et al., 2005). In Table 5.5, there are additional absorption features found in the data presented in Figure 5-10 at 48.0 cm⁻¹ and at higher frequencies beyond the limit of the equipment capabilities used by (Nishizawa J, et al., 2005). This highlights the sensitivity and spectral resolution of the THz FTS used in this thesis by identifying weak absorption features at low frequencies. It also further proves the advantages of the THz FTS by having a broad spectral range with high signal to noise.

Just as (Lopes R. P, 2012) had modelled the structure of guanine molecules, (Feng Zhang, 2014) has done the same for adenine by using the parameters of the unit cell. A terahertz time-domain spectroscopy system was used by (Feng Zhang, 2014) to investigate protonated adenine which was used without any purification and deuterated adenine which was recrystallised from hot deuterium oxide. The data shown in Figure 5-12 were taken at 5K and 300K and are shown in red and grey respectively, with (a) being the data for protonated adenine and (b) for deuterated adenine.

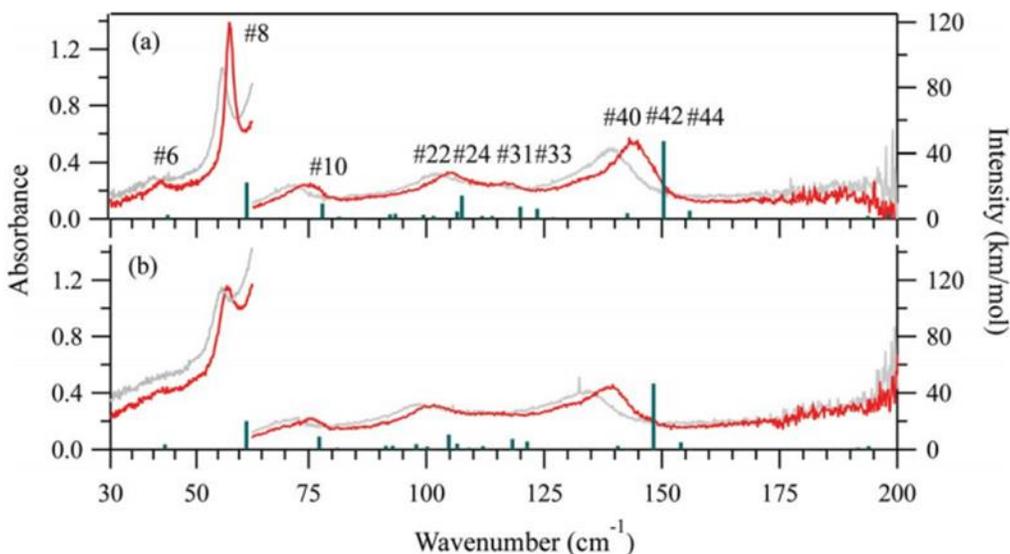


Figure 5-12: THz spectra for (a) protonated adenine and (b) deuterated adenine taken at 5K (red line) and 300K (grey line) with the predicted DFT features provided by the green lines taken by (Feng Zhang, 2014).

The difference in the temperature at which the data were taken produces a shift in the absorption features, with the data taken at 5K having slightly sharper features. Interestingly, this is not a distinct narrowing of features only a shift. At frequencies below 200 cm^{-1} , solid state vibrations are strongly influenced by the lattice dimensions of the unit cell which make them sensitive to change in temperature, whereas absorption features found in the region above 200 cm^{-1} due to internal or localised molecular vibrations will be less so. This is expected as at lower temperatures, the molecular vibrations will be reduced and result in less line broadening. As for the comparison between protonated and deuterated samples, there does seem to be a slight difference in the data between a) and b). It is mentioned that a deuterium substitution was confirmed by observing the loss of the O-H and N-H stretch bands from data taken by ATR –FTIR. This would suggest that the features would be shifted and different to that of the protonated adenine which is seen in Figure 5-12.

The main difference in the spectra in Figure 5-10 is a small but broad feature found in both the samples made with PBS and HBSS. This small feature is identified in bold at 158 cm^{-1} and 261 cm^{-1} and is not present in any of the other spectra. PBS and HBSS are solvents which contain growth hormones and other substances that are pH neutral but can help sustain cells or cultures outside the body. As the results show, THz spectroscopy is sensitive to very small changes in crystalline structure where one of the additions could have caused the feature found in the data. According to Thermo Fisher, HBSS can contain concentrations of 0.14g/L of calcium chloride, 0.10 g/L of magnesium chloride and sulphate, and 0.01 g/L of phenol red. However, this accounts for less than 1% compared to the concentration of the nucleobase samples. Unlike the work presented by

(Feng Zhang, 2014), the work presented here has no significant differences between adenine in DI water and heavy water which suggests that the spectra is of fully dehydrated crystalline adenine.

5.1.1.6 Conclusion

Using a THz FTS, spectral data of all four nucleobases have been presented. The samples were in solid state following evaporation from solutions using 5 different solvents. All the samples show unique THz transmission spectra with many sharp absorption features. The absorption features of all the samples were compared to THz transmission spectra taken from (Nishizawa J, et al., 2005) using a GaP THz wave generator between 0.4 to 5.8 THz (13.3 to 193.3 cm^{-1}) which were in agreement within error. However, many other absorption features were identified in this work due to the high sensitivity, high signal to noise and broad spectral range available using the FTS.

By using this technique, absorption features below 55 cm^{-1} that are unique to some nucleobases have been identified which has not been previously possible. In this region, unique features were found in guanine at 58.6 cm^{-1} and adenine at 48.0 cm^{-1} . Other features were identified which have not been identified before were found for guanine at 100.0 cm^{-1} and 155.09 cm^{-1} , cytosine at 88.32 cm^{-1} and adenine at 187.0 cm^{-1} as well as others at frequencies between 6.0 and 20.0 THz (200 to 600 cm^{-1}).

This work has indicated that samples of nucleobases made using different solvents can produce different THz spectra. THz spectra of cytosine has shown distinct differences between the samples made using DI water and IPA. This highlights 1) different solvents

will create a different crystalline structure 2) THz frequencies are sensitive to these small crystalline changes.

The spectra presented here for all the nucleobases made using different solvents have shown a decrease in transmission at frequencies above 170 cm^{-1} . When the hydrophobic nucleobases are placed into a solvent, they will aggregate and form large crystals after evaporation. Hence the data is obscured at these frequencies due to the effects of scattering.

5.1.2 Purines and Pyrimidines

The four nucleobases plus uracil (which replaces thymine in RNA) can be classified in two groups according to their molecular ring structure. The purines have a double ringed system as is the case for adenine and guanine. The pyrimidines have a single ringed system and consists of cytosine, thymine and uracil. Figure 5-13 shows the molecular structure of all the nucleobases including uracil (Mario Barbatti, 2010).

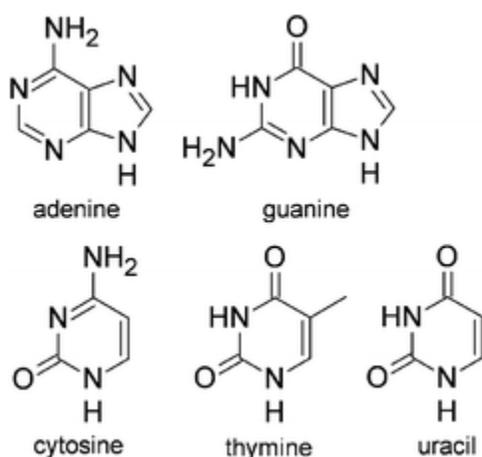


Figure 5-13: Molecular structure of 5 nucleobases: adenine, guanine, cytosine thymine and uracil.

THz spectroscopy has been performed for the purines and pyrimidines to investigate nucleobases of similar structure and compare their absorption features. The data presented here are the THz spectra of the nucleobases made using DI water and IPA in the same method as in section 5.1.3.1.

5.1.2.1 Purines

Figure 5-14 shows the molecular structure of guanine and adenine, the purines. As mentioned, these are double ringed systems which have an imidazole ring in common as highlighted in Figure 5-14. An imidazole ring is an aromatic heterocycle naturally having the formula $(CH)_2N(NH)CH$. When it becomes part of the double ringed system, the hydrogen atoms are lost and replaced the leaving the imidazole ring with the formula $C_3N(NH)$.

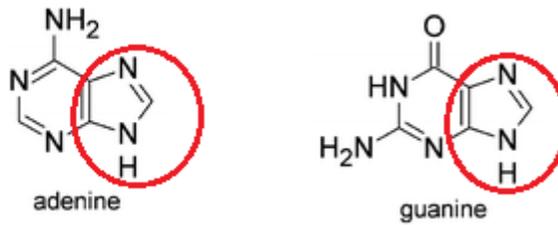


Figure 5-14: Molecular structure of adenine and guanine highlighting the imidazole ring

Figure 5-15 identifies the main absorption features in adenine and guanine made using DI water.

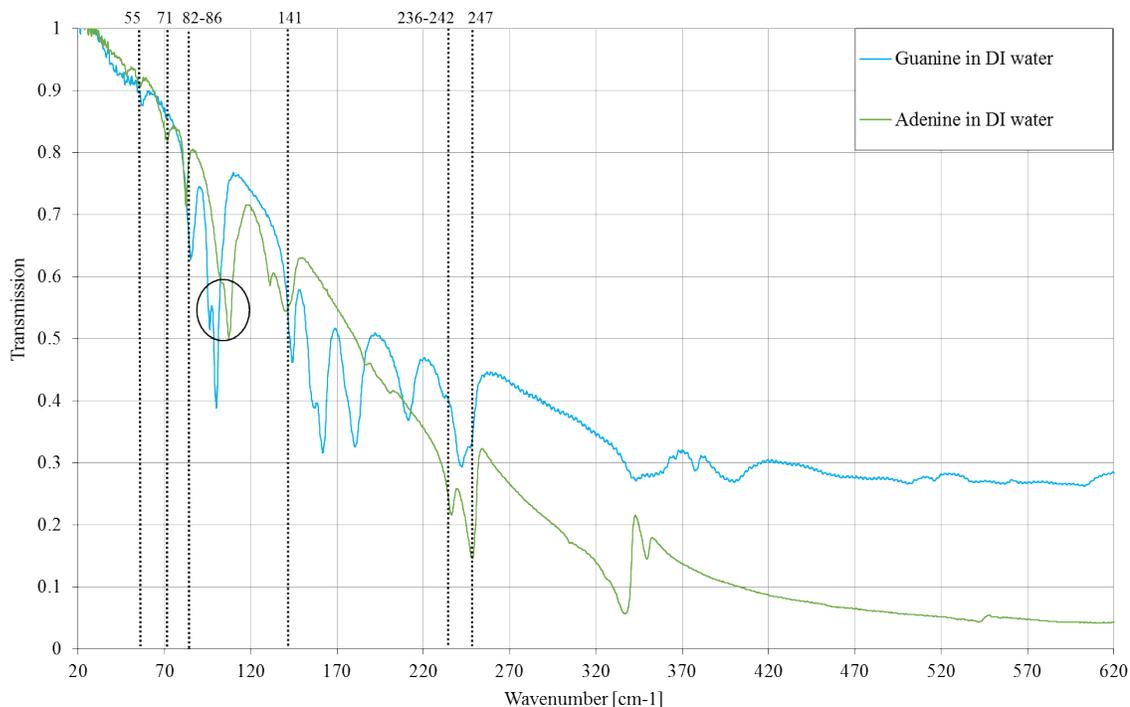


Figure 5-15: THz spectra of guanine and adenine in DI water from 20 to 620 cm^{-1}

Figure 5-15 shows that there are many common resonances between adenine and guanine. The features found at $55 \pm 1 \text{ cm}^{-1}$, $71 \pm 1 \text{ cm}^{-1}$, $84 \pm 2 \text{ cm}^{-1}$, $99 \pm 3 \text{ cm}^{-1}$ and $103 \pm 4 \text{ cm}^{-1}$ which are circled, $141 \pm 1 \text{ cm}^{-1}$, $239 \pm 3 \text{ cm}^{-1}$, and $247 \pm 1 \text{ cm}^{-1}$ are presumed to be common although shifted at some frequencies. As these features appear at similar frequencies and are of similar intensity, they have been assumed to be due to the imidazole ring which is the same for both adenine and guanine.

As well as these absorption features, there are also other similarities, such as the lack of any features in both samples in the region between 255 cm^{-1} and 320 cm^{-1} and the region between 410 cm^{-1} and 500 cm^{-1} . For comparison, the spectra of guanine and adenine made using IPA have been plotted in Figure 5-16.

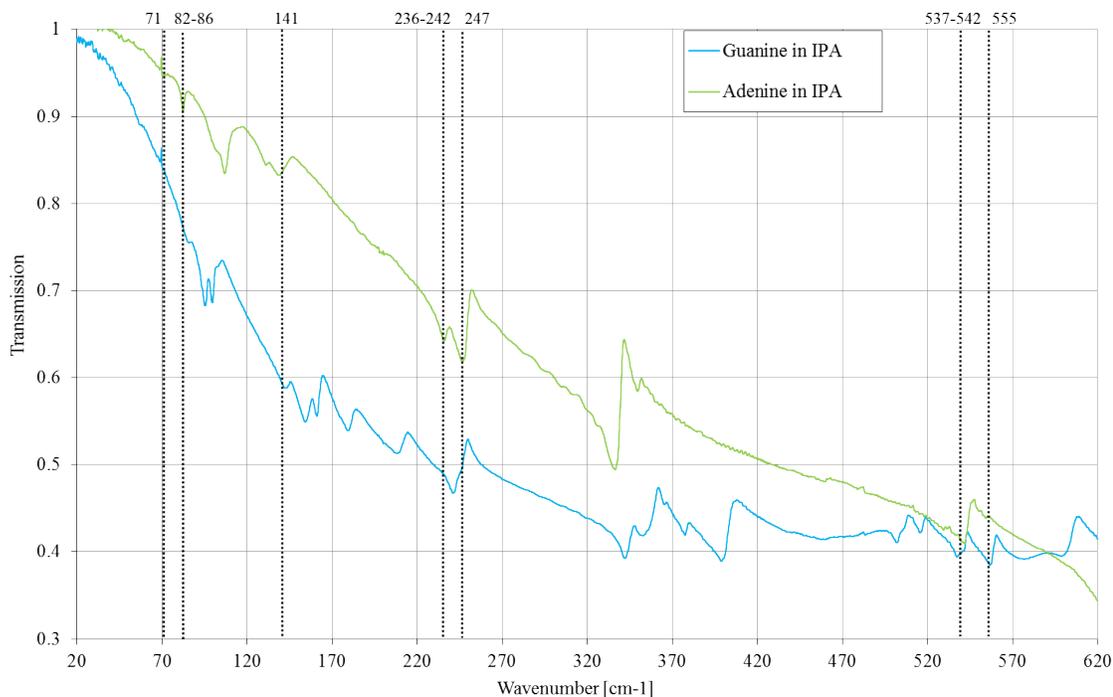


Figure 5-16: THz spectra of guanine and adenine made using IPA from 20 to 620 cm^{-1}

All the absorption features found in Figure 5-15 are also present in Figure 5-16 except for the ones identified at 537 – 542 cm^{-1} and 555 cm^{-1} . At these frequencies there are absorption features in both adenine and guanine which also suggests that they are linked to the imidazole ring.

Interestingly, there is a difference between adenine and guanine in both Figure 5-15 and Figure 5-16 in the region between 145 cm^{-1} and 220 cm^{-1} . Guanine has very strong absorption features here which highlights the sensitivity of THz spectroscopy to small differences in the molecular and crystalline structure of the samples despite them having similar atomic construction.

5.1.2.2 Pyrimidine

The molecular structures of cytosine, thymine and uracil are shown in Figure 5-17 highlighting the similarity between the pyrimidine molecules.

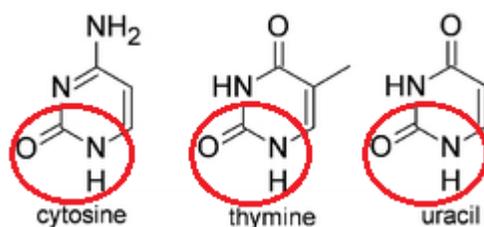


Figure 5-17: Molecular structure of cytosine, thymine and uracil highlighting the main similarity between each molecule

Clearly there are similarities in the molecular structure of cytosine, thymine and uracil, with the latter two being almost identical, except for an additional CH₂ in the thymine molecule.

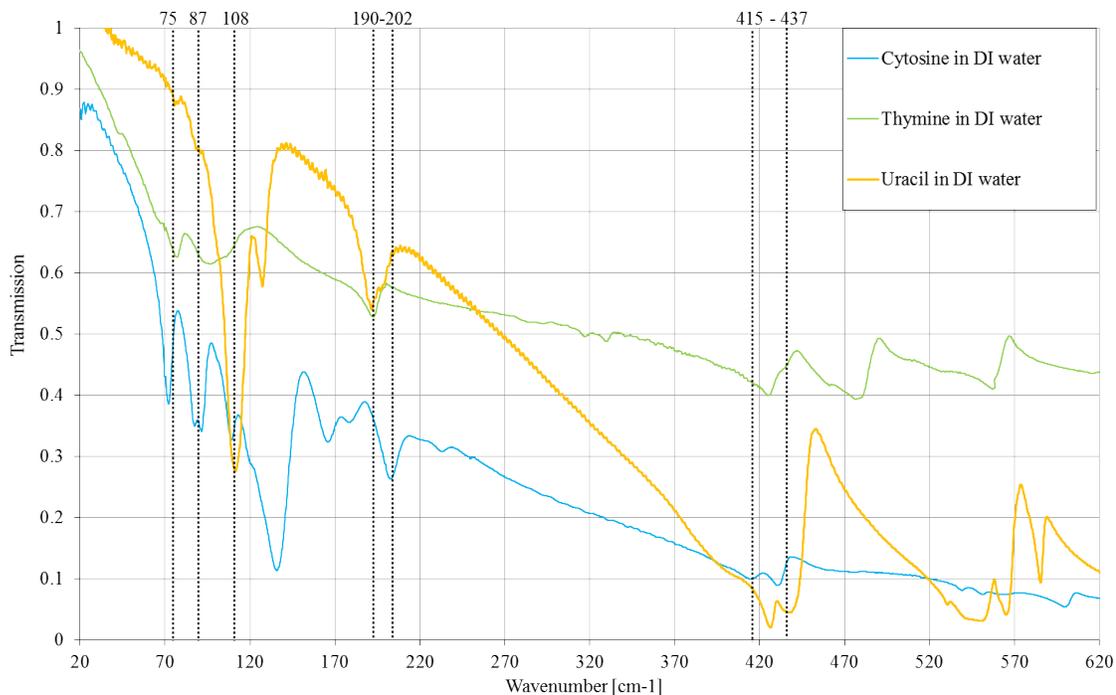


Figure 5-18: THz spectra of cytosine, thymine and uracil made using DI water from 20 to 620 cm⁻¹

Figure 5-18 has many common absorption features identified at 75, 87, 108, 190 – 202 and 415 – 437 cm⁻¹ with some frequency shifts. These could be presumed to be linked to the NH-C=O which is common in all the samples. When the structure of both molecules is looked at in more detail referring to Figure 5-17, in the cytosine molecule, the NH-C=O part is attached to an N=C-NH₂ molecule but in the thymine molecule it is attached to another NH-C=O. Both have similar molecular weight, calculated as 7.0×10^{-23} g for NH-C=O and 7.1×10^{-23} g for NH-C=O, which suggest that it cannot be a shift due to the change in mass as they are within < 2% identical.

Figure 5-19 shows the THz spectra of the pyrimidine samples made using IPA.

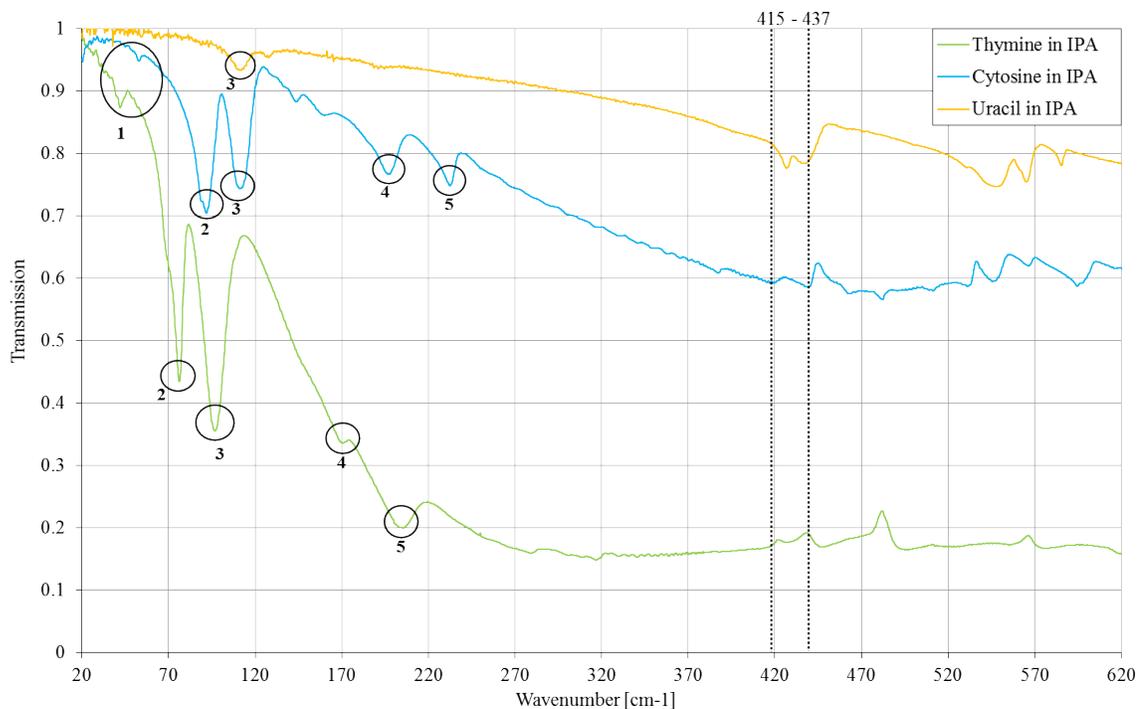


Figure 5-19: THz spectra of cytosine, thymine and uracil made using IPA from 20 to 620 cm^{-1}

In section 5.1.1.3, it was shown that the THz spectra of cytosine samples made using DI water and IPA resulted with different absorption features. With reference to Figure 5-19, it can be seen that there are similarities in the frequency of the absorption features, but there are greater shifts in the frequency between the three samples. The main similarities between the three samples have been circled and highlighted by using a labelling system to identify the features that could correspond to the same resonance. The THz spectra of the uracil sample made using IPA has a much higher transmission and weaker absorption features than the uracil sample made using DI water, however, they still generate absorption feature at the same frequencies.

At lower frequencies, the matching absorption features in cytosine and thymine are identified as circled at 42 cm^{-1} and 53 cm^{-1} , 76 cm^{-1} and 89 cm^{-1} , 96 cm^{-1} and 110 cm^{-1} , 170 cm^{-1} and 197 cm^{-1} (191 cm^{-1} in uracil), 205 cm^{-1} and 232 cm^{-1} and between 415 cm^{-1} and 437 cm^{-1} . Despite the difference in the THz spectra of cytosine made using DI water and IPA, it is clear that the region between 245 cm^{-1} and 400 cm^{-1} has no distinct absorption features in any of the samples.

In both Figure 5-18 and Figure 5-19, the uracil sample has a very similar THz spectrum to thymine. For closer scrutiny, the spectra for thymine and uracil have been compared in section 5.1.2.2.1.

5.1.2.2.1 *Thymine and Uracil*

Uracil is a product of spontaneous deamination¹ of cytosine, and in DNA this is corrected by the removal of uracil and hence why thymine is the nucleobase in DNA. However, in the structure of RNA, the thymine molecule is replaced by uracil as it is energetically less expensive to produce where quantity of bases is more important than lifespan² (Daily, 2017).

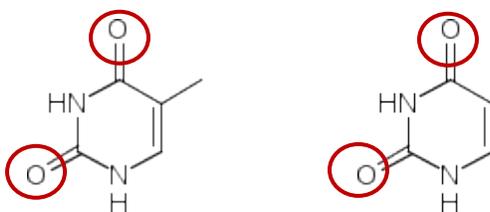


Figure 5-20: Molecular structure of thymine (left) $C_5N_2O_2H_6$ and uracil (right) $C_4N_2O_2H_4$ with the red circles indicating the “O7” and “O8” positions.

(C.P. Beetz Jr., 1980) compared uracil molecules to thymine, which have very similar structures as shown in Figure 5-20. The THz spectra of the samples made using DI water are presented in Figure 5-21.

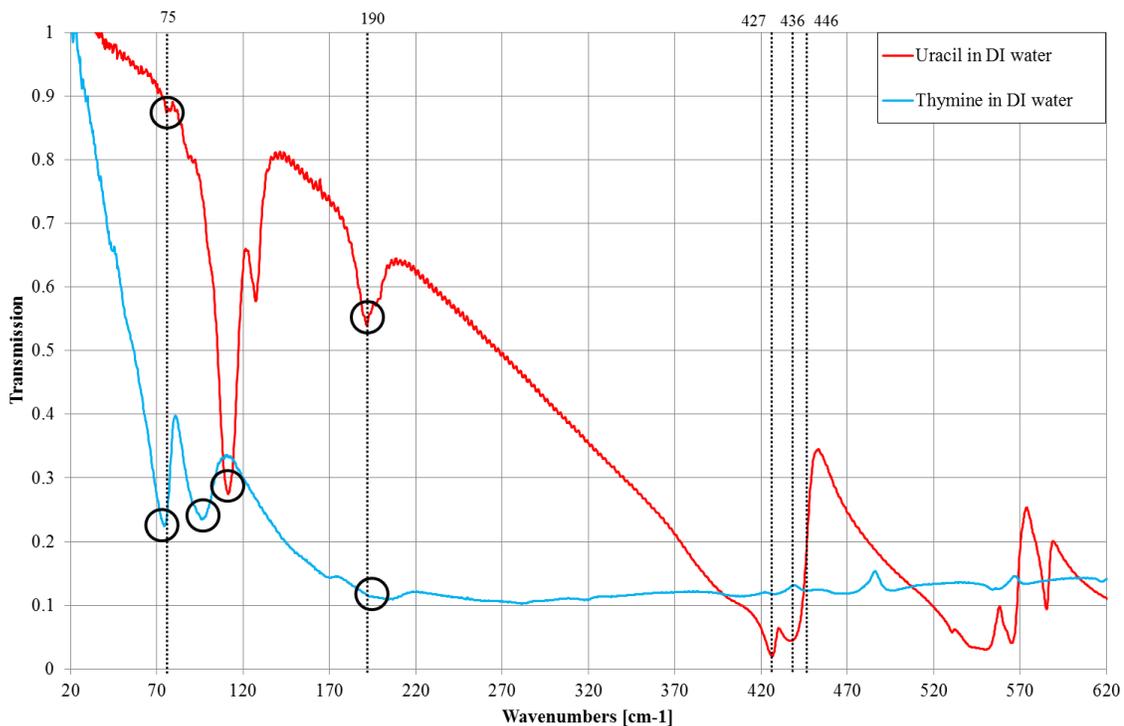


Figure 5-21: THz spectra of thymine and uracil made using DI water from 20 to 620 cm^{-1}

Highlighted in Figure 5-21 are the features at 75 cm^{-1} , 190 cm^{-1} , 427 cm^{-1} seen in both thymine and uracil, 108 cm^{-1} and 436 cm^{-1} in uracil only, 96 cm^{-1} and 446 cm^{-1} in the thymine only. The common absorption features are identified in the literature as a band at 433 cm^{-1} in thymine and 427 cm^{-1} in uracil and can be assigned to the out-of-plane wagging of the oxygen atoms identified in the molecular structure in Figure 5-20. The 206 cm^{-1} feature in thymine, found at 190 cm^{-1} in Figure 5-21, is probably the lowest

lying internal vibration which corresponds to the 190 cm^{-1} feature in the uracil molecule, and is thought to be an out-of-plane deformation.

When the spectra for the pyrimidines and the purines are compared there are some observational comparisons that can be made. Comparing the region between 270 cm^{-1} and 420 cm^{-1} in Figure 5-18 and Figure 5-19 which has no absorption features and the same region in Figure 5-15 and Figure 5-16, there are clearly some features unique to the purine samples. These features can most likely be assigned to the imidazole ring.

5.1.2.3 Conclusion

The THz spectra of five nucleobases have been analysed by their molecular structure grouping. The purine samples, made using DI water and IPA, had the same absorption features found at $55 \pm 1\text{ cm}^{-1}$, $71 \pm 1\text{ cm}^{-1}$, $84 \pm 2\text{ cm}^{-1}$, $99 \pm 3\text{ cm}^{-1}$, $103 \pm 4\text{ cm}^{-1}$, $141 \pm 1\text{ cm}^{-1}$, $239 \pm 3\text{ cm}^{-1}$, and $247 \pm 1\text{ cm}^{-1}$. As they are common to both cytosine and guanine they are presumed to be linked to the imidazole ring.

The pyrimidines made using DI water have common absorption features identified at 75, 87, 108, 190 – 202 and $415 - 437\text{ cm}^{-1}$ with some frequency shifts. Similar absorption features in cytosine and thymine made using IPA are identified at 42 cm^{-1} and 53 cm^{-1} , 76 cm^{-1} and 89 cm^{-1} , 96 cm^{-1} and 110 cm^{-1} , 170 cm^{-1} and 197 cm^{-1} , 205 cm^{-1} and 232 cm^{-1} and between 415 cm^{-1} and 437 cm^{-1} .

Comparing the spectra for the purines and pyrimidines showed a big difference in the region between 245 cm^{-1} and 400 cm^{-1} . There are no distinct absorption features in the pyrimidines whereas there are common absorption features in the purines at $339 \pm 3\text{ cm}^{-1}$

and $351 \pm 3 \text{ cm}^{-1}$. These absorption features are suggested to be linked to the features from the imidazole ring.

5.1.3 Nucleobases and nucleosides

As discussed in section 5.1.2, all five nucleobases have a ring structure, with adenine and guanine having a double ringed structure (purines) and cytosine and thymine only having a single ring (pyrimidines). The nucleosides are a combination of the individual base and a five-carbon sugar molecule shown in Figure 5-22. This type of bond is stronger than a hydrogen bond as it is bound via glycosidic linkage which is a type of covalent bond that joins a carbohydrate (sugar) molecule to the base molecule.

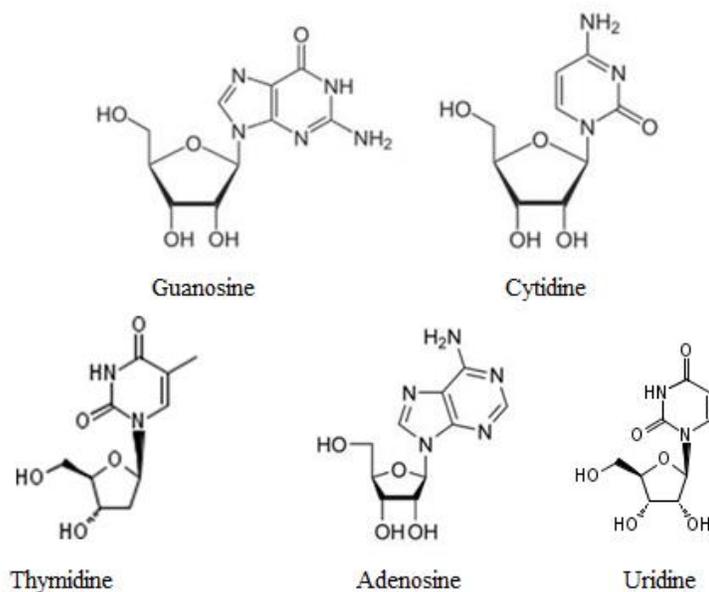


Figure 5-22: Molecular structure of the five nucleosides: guanosine, cytidine, thymidine, adenosine, and uridine.

This section presents the results of THz spectroscopy performed on all nucleobases and their complimentary nucleosides.

5.1.3.1 Method

The samples were made in the same method as explained in section 5.1.1.1 with the concentrations shown in Table 5.6. It includes the concentration of the samples made in DI water and IPA.

Table 5.6: The ratio of the amount of nucleoside powder to solvent for each nucleoside made into a solution using DI water and IPA

SAMPLE NAME	CONCENTRATION (± 0.50 g/L)	
	IPA	DI WATER
Guanosine	13.09	13.07
Cytidine	13.28	13.33
Adenosine	14.49	14.01
Thymidine	16.15	16.61
Uridine	14.86	15.01

Given that the structure of the nucleosides contains a 5-carbon sugar ring, it is important to investigate this as a separate entity as has been done for the nucleobases. Therefore, glucose has been added to the dataset, and compared to real results from literature.

5.1.3.2 The 5-carbon sugar ring

Glucose is an example of a carbohydrate and has the molecular formula $C_6H_{12}O_6$. The nucleosides investigated in this section are constructed of a nucleobase and a 5-carbon ring ribose with the formula $C_5H_{10}O_5$. It is important to investigate the THz spectra of glucose on its own to see whether there are any significant absorption features that are characteristic to the sugar ring. Figure 5-23 shows the THz spectra of both dried glucose pressed onto 3.3 μm PP and glucose film made by dissolving glucose powder in DI water and leaving to evaporate at 100°C on a PP substrate.

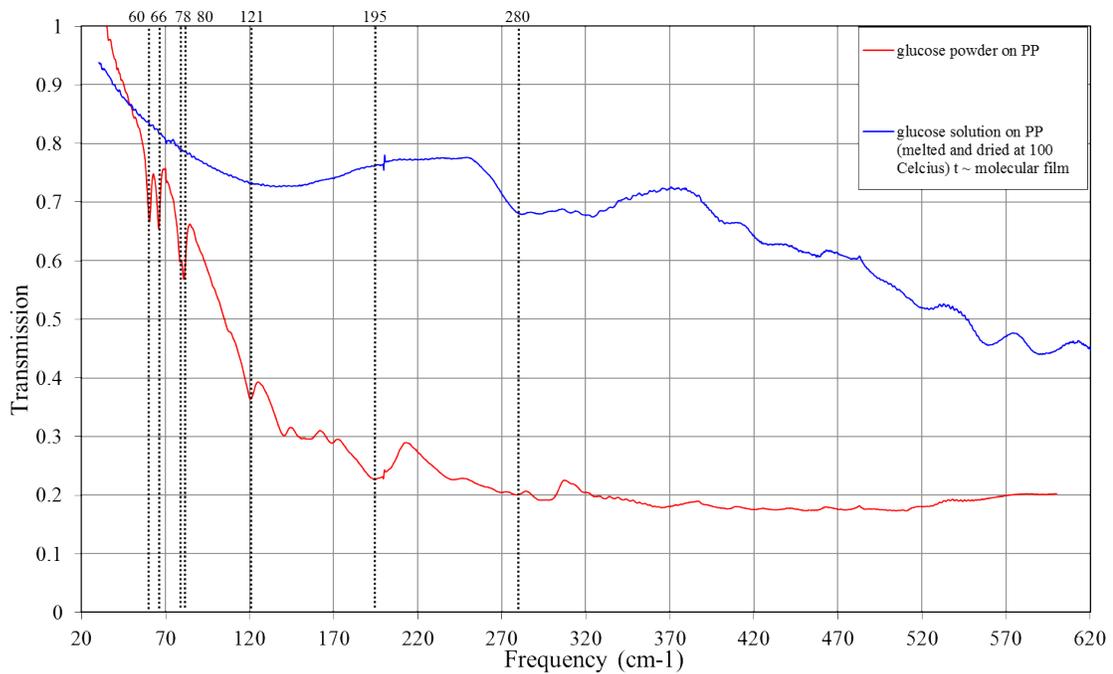


Figure 5-23: THz spectra of glucose powder pressed into a 3.3 μm PP substrate and glucose film which was made by making a solution with DI water and left to dehydrate

Clearly there is a difference between the spectra. Glucose is a polar molecule and hence is able to dissolve in water. When the solution is left to evaporate on the substrate, the resulting film will contain water and glucose molecules. This might explain the broad feature at 120 to 130 cm^{-1} which could be a hallmark for water absorption and is explored in more detail in section 5.4. (Liu & Zhang, 2006) probed the dehydration process of anhydrous and monohydrated D-glucose through THz absorption spectroscopy and found that the samples exhibited different absorption features. The monohydrate samples had absorption features at 60.7, 65.3 and 82.0 cm^{-1} whereas the anhydrous sample had absorption features at 43.0, 48.0, 70.0 and 86.7 cm^{-1} . With regards to the spectra of the glucose film, there is a small feature $70.0 \pm 1 \text{ cm}^{-1}$ but unclear as to whether it is a real absorption feature.

(Laman, 2008) conducted THz transmission spectroscopy on various biological molecules including D-glucose where they found three key absorption features at 63.0, 70.3 and 85.7 cm^{-1} similar to that found by (Liu & Zhang, 2006) which could correspond to the features identified in Figure 5-23 found at 60.5, 66.0 and 80.9 cm^{-1} . The additional small feature identified at 78.6 cm^{-1} could be due to the advantages of using a FTS, despite (Liu & Zhang, 2006) attempts to reduce inhomogeneous broadening by drop-casting to obtain an ordered polycrystalline film. However, as the results here and from (Liu & Zhang, 2006) show, formation of polycrystalline films of nucleosides and sugars are highly sensitive to environmental conditions.

From the previous sections on nucleobases (5.1.1) it is clear that all samples have very distinct absorption features at similar THz frequencies. This makes it difficult to identify any vibrational resonances that is from the sugar within the nucleoside.

5.1.3.3 Guanine and guanosine

The addition of the sugar carbon ring to the nucleosides increase the size of the molecule and would be expected to perturb the crystalline structure. This structural change is clearly visible through the THz spectra seen in Figure 5-24.

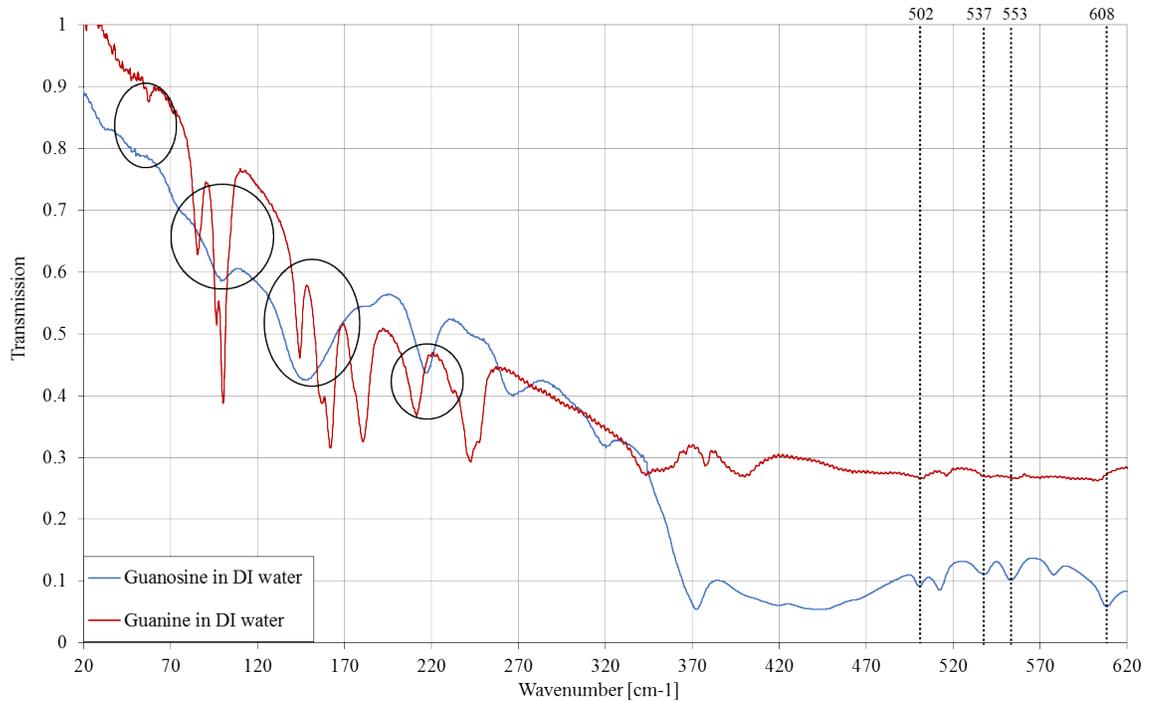


Figure 5-24: THz spectra of guanine and guanosine made in DI water from 20 to 620 cm^{-1}

The THz spectra for guanine and guanosine made using DI water presented in Figure 5-24 show resonances in the same region but with the absorption features for the guanosine seemingly broader and/or shifted in frequency.

Although, it is interesting to note that the features found by the guanine sample seems to be sharper and highly absorptive, both samples still exhibit common absorption features.

There is no literature work on the low frequency THz spectra of guanosine, however there is some work on the mid-infrared region. Table 5.7 shows some mid infrared spectroscopy data for anhydrous guanine from (Lopes R. P, 2012) which was taken using a “Bruker Optics Vertex 70 Fourier Transform Infrared (FTIR)” spectrometer and also shows data calculated using Density Functional Theory (DFT) using the plane-wave method, as explained in CHAPTER 4. In the “approximate description” the letters correspond to the atoms in the molecule, the numbers corresponds to the location of the atoms, the Δ representing the in-plane ring deformation of skeletal atoms and the Γ representing the out-of-plane ring deformations of skeletal atoms.

Table 5.7: Table of infrared and calculated wavenumbers for anhydrous guanine

FT-IR (cm ⁻¹)	Calculated plane-wave method (cm ⁻¹)	Approximate description
503	494, 506, 512, 519	$\Delta(N1-C6-C5) + \Delta(C2-N3-C4); \Delta(N3-C2-C1) + \Delta(C4-C5-C6)$
540 - 557	538, 542, 560, 563	$\Delta(C2-N1-C6) + \Delta(N3-C4-C5)$
604	614, 615, 616	$\Gamma(C8-N7-C5) - \Gamma(C4-N9-C8)$
644	643, 647, 648, 649	Pyr + Im ring breathing

The lowest frequency and the first absorption feature in Table 5.7 was at 503 cm⁻¹ which corresponds to the feature identified in this work at 502 cm⁻¹ in Figure 5-24 as well as the 540cm⁻¹ and 557 cm⁻¹ features which can be identified at 537 cm⁻¹ and 553 cm⁻¹. The 503 cm⁻¹ can be ascribed to in-plane deformation of the atoms in the *Pyr* (Pyrimidine) ring.

For the 604 cm^{-1} feature noted in the “FT-IR” column, there is a corresponding weak absorption feature in the spectra in Figure 5-24 at 608 cm^{-1} . This feature is assigned to the out-of-plane vibration. The plane-wave calculated method analysis led to the feature being due to the *Im* (Imidazole) ring deformation. Also, (Lopes R. P, 2012) noted that the feature could imply “displacement of (N7)H and (C8)H hydrogen atoms.” As the THz spectra of guanosine is very similar to guanine in this region, it is assumed that the assignment of the features found for guanine could be the same for guanosine. For a better understanding there is a need for DFT modelling of the guanosine molecule which is explored in more detail in section 5.2.

5.1.3.4 Cytosine and cytidine

The difference in the structure of cytosine and cytidine is explored in this section with THz spectroscopy performed on both nucleobase and nucleoside made using DI water and IPA. As discussed in detail in section 5.1.1.3, cytosine showed a different THz spectrum for samples made using DI water and IPA. Figure 5-25 shows the spectra for cytosine and cytidine when made using DI water.

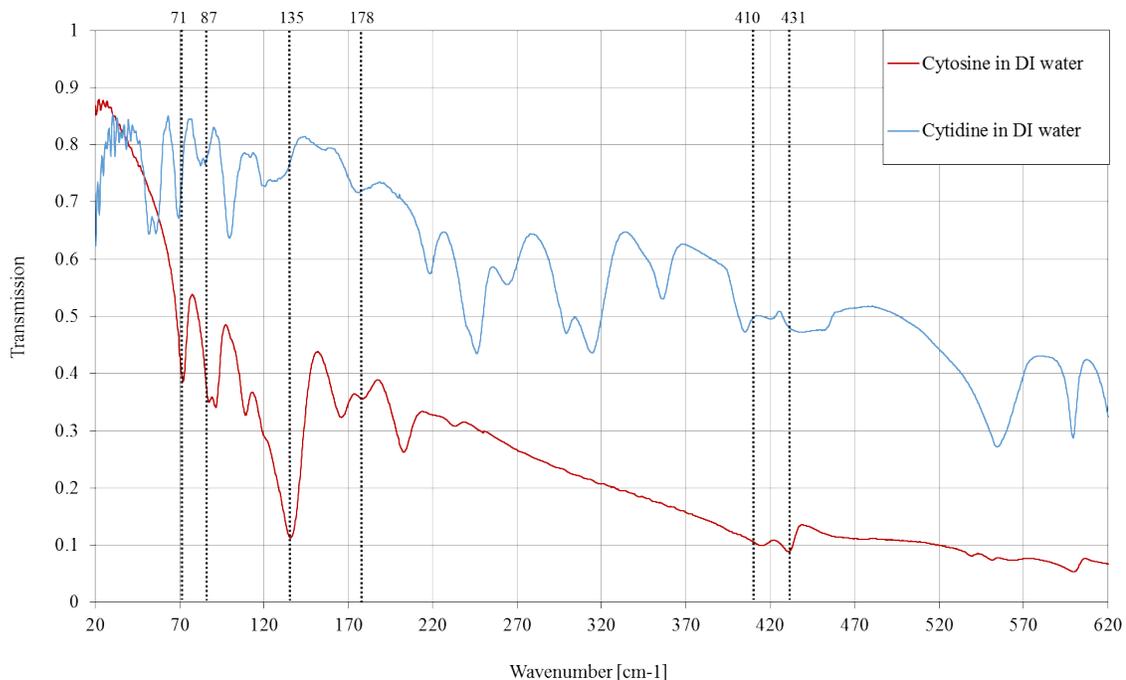


Figure 5-25: THz spectra of cytosine and cytidine, both made into a solution using DI water from 20 to 620 cm^{-1}

There are some common features between the cytosine and cytidine spectra highlighted in Figure 5-25 at $71 \pm 2 \text{ cm}^{-1}$, $87 \pm 3 \text{ cm}^{-1}$, $135 \pm 2 \text{ cm}^{-1}$, $178 \pm 3 \text{ cm}^{-1}$, $410 \pm 4 \text{ cm}^{-1}$ and $431 \pm 2 \text{ cm}^{-1}$. Unlike the comparison of guanine and guanosine in section 5.1.3.3, the absorption features found in both cytosine and cytidine are sharp and have similar line width.

Another interesting region is between 241 cm^{-1} and 400 cm^{-1} where the spectrum for the cytidine molecule has many absorption features whereas the cytosine has none. This region's absorption features can be associated with either the glycosidic link between the nucleobase and the carbon sugar ring, and/or vibrational or phonon vibrations from the five-carbon sugar ring only.

According to (C.P. Beetz Jr., 1980), similarities in the absorption spectrum of the 3 nucleosides (uridine, cytosine and inosine) between $250\text{ cm}^{-1} - 400\text{ cm}^{-1}$ shows that the features in this region are not characteristic to the bases. They are associated with the furanose³ ribose ring. The cytidine absorption bands were identified at 248, 264, 300, 317, 346, 354 and 378 cm^{-1} and are well matched to the absorption features present in the cytidine spectra in Figure 5-25. Samples of cytosine and cytidine made using IPA were also compared in Figure 5-26.

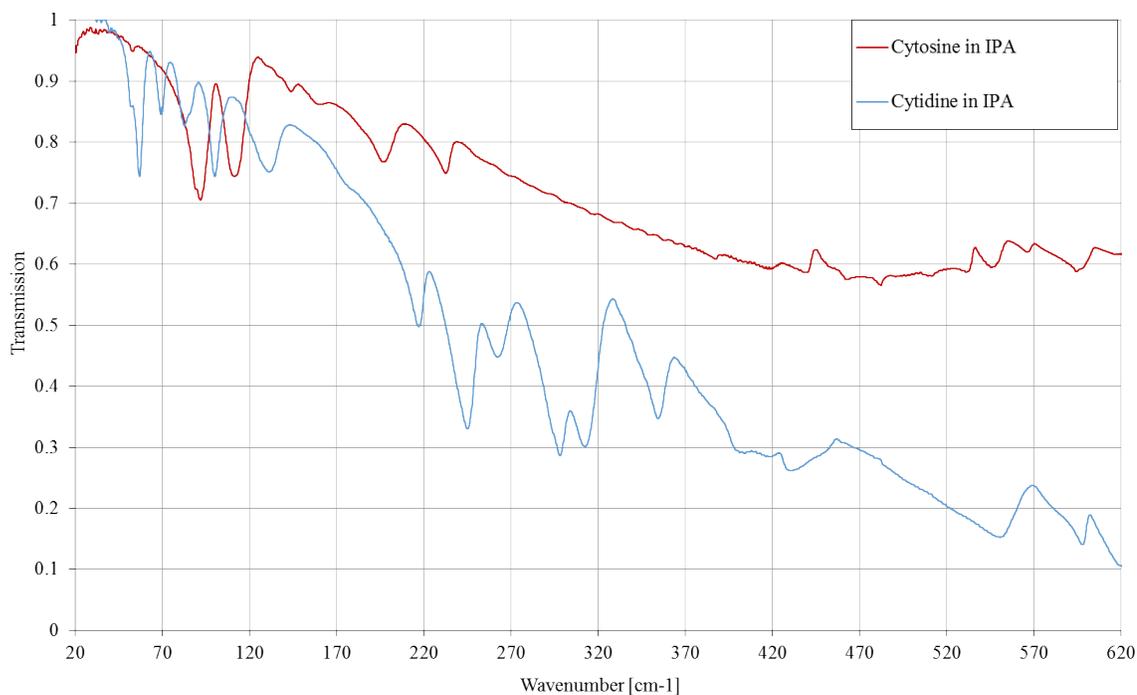


Figure 5-26: THz spectra of cytosine and cytidine made using IPA from 20 to 620 cm^{-1}

³ A furanose is a collective term for carbohydrates that have a chemical structure that includes a five-membered ring system consisting of four carbon atoms and one oxygen atom

The comparison of cytosine and cytidine made using IPA show there are less similarities compared to the samples made using DI water although the samples of cytidine made using DI water and IPA generate the same THz spectra. This suggests that the sample of cytidine made using DI water or IPA potentially forms a similar crystalline structure to that of the cytosine in the monohydrate form.

However, both Figure 5-25 and Figure 5-26 have a region between 241 cm^{-1} and 400 cm^{-1} where the spectrum for the cytidine molecule has many absorption features but the cytosine made in DI water and IPA has none. This further enhances the suggestion that the absorption features here are unique to the cytidine molecules.

5.1.3.5 Adenine and adenosine

Here the structure of adenine and its nucleoside adenosine is investigated. Figure 5-27 shows the THz spectra of both made using DI water.

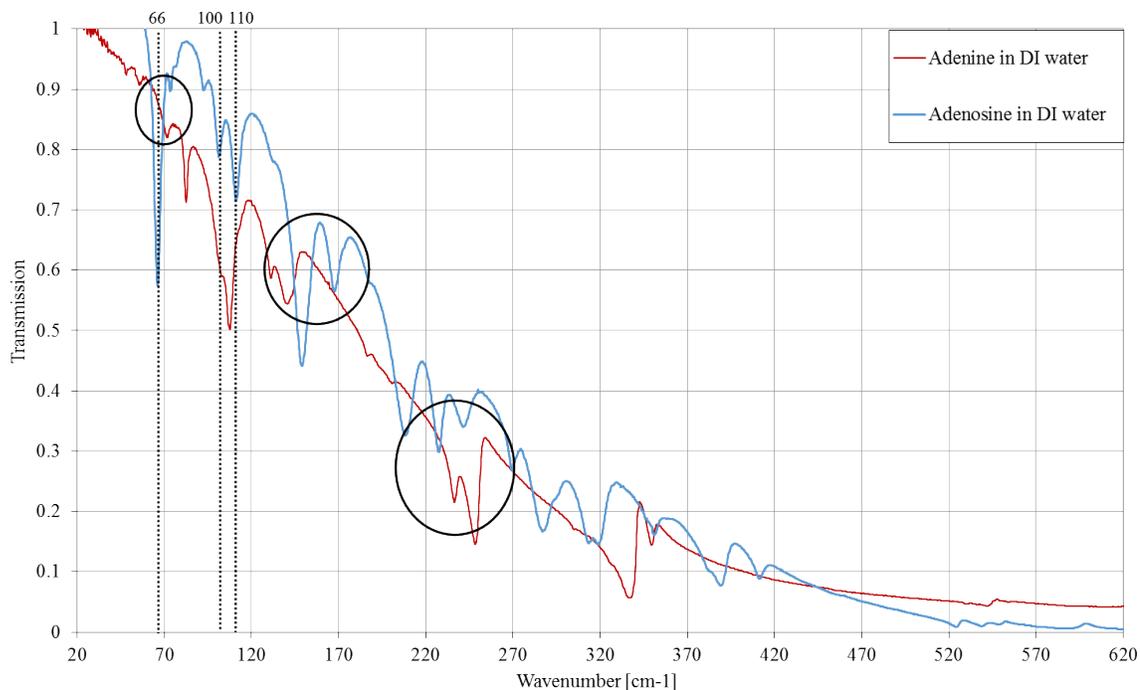


Figure 5-27: THz spectra of adenine and adenosine made using DI water from 20 to 620 cm^{-1}

The THz spectra of adenine and adenosine have a few common absorption features identified in Figure 5-27. In the paper by (Nishizawa J, et al., 2005), adenosine samples were investigated up to 5 THz using a pulsed laser and were found to have absorption features at the same frequencies as shown in Figure 5-27. They were identified at 65 cm^{-1} , 100 cm^{-1} and 110 cm^{-1} . When understanding the mechanism of the molecules, (Lee S.A., 2000) has assigned the same absorption features to lattice modes whereas (Bailey L.E., 1997) associated the 65-70 cm^{-1} to the torsion of the ribose residue.

Further THz spectroscopy studies were made on adenosine by (Laman, 2008) which found significant sharpening of the absorption features when cooled to 77K. The absorption features were found at 25.7, 65.3, 73.7, 92.3, 101.0 and 111.0 cm^{-1} . The data

presented in Figure 5-27 is limited at lower frequencies therefore it is not possible to identify the 25.7 cm^{-1} but the other absorption feature identified are well matched.

Adenosine exhibits strong vibrations between the adenine and ribose addition, mediated by the glycosidic bond, in the low-frequency region, and these vibrations were strongly coupled with the intermolecular vibrations (Feng Zhang, 2014). For comparison, the THz spectra of adenine and adenosine made using IPA are shown in Figure 5-28.

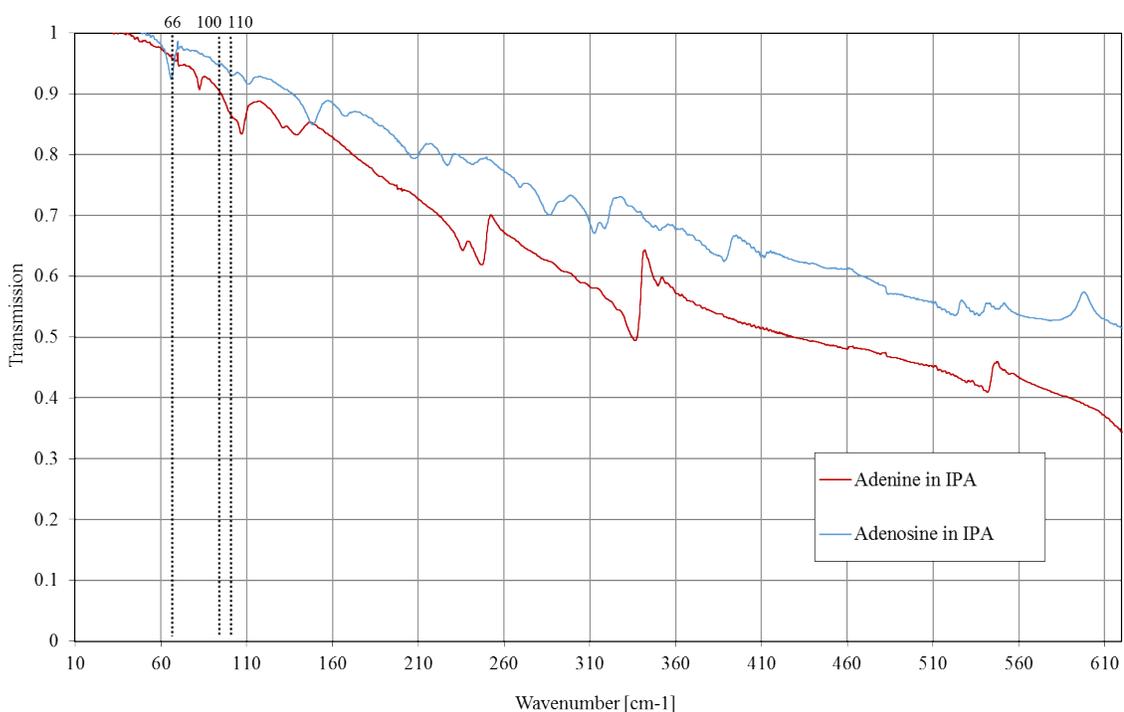


Figure 5-28: THz spectra of adenine and adenosine made using IPA from 20 to 620 cm^{-1}

Although the intensity of the absorption feature of the samples made using DI water and IPA are not the same, their frequencies are and these are identified in Figure 5-28. The samples of adenine and adenosine made using IPA have a larger distribution on the

substrate meaning that the local concentration is less than the sample made using DI water. However, there are no differences between the frequencies of the absorption features which means that the crystalline structure of the samples is the same.

5.1.3.6 Thymine and thymidine

The THz spectra of thymine and thymidine were investigated. These were also made using DI water and IPA.

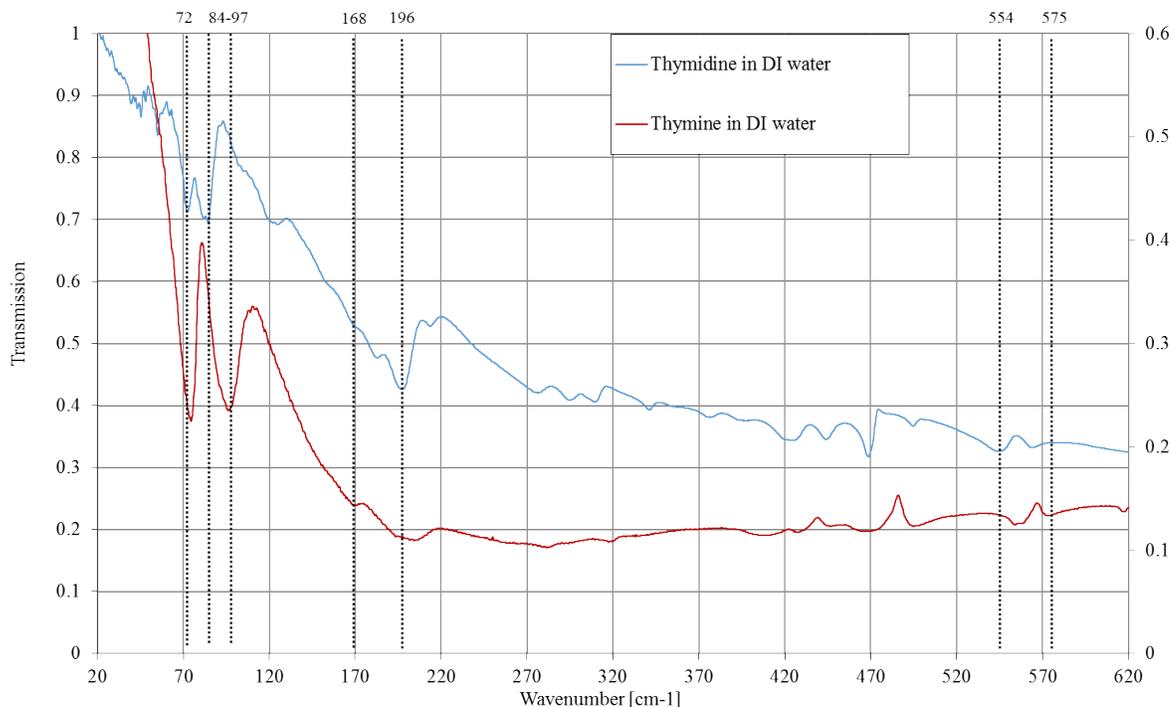


Figure 5-29: THz spectra of thymine and thymidine made using DI water from 20 to 620 cm^{-1}

The thymine and thymidine spectra in Figure 5-29 are similar with the exception of a few absorption features. This is unusual due to the difference in the structure for both thymine and thymidine.

According to (Laman, 2008) absorption peaks for thymine were found at 28 cm^{-1} , 33 cm^{-1} , 68 cm^{-1} and 83 cm^{-1} at 77K. There is an absorption feature that can be identified at $43.0 \pm 2 \text{ cm}^{-1}$ in Figure 5-29 for thymidine but as the thymine sample is placed on a secondary axis for a better comparison, the feature is not visible. However there are absorption features at 74.0 and 97.3 cm^{-1} for thymine and at 72.0 and 84.0 cm^{-1} for thymidine which may correspond to the same resonances as found by (Laman, 2008). As well as this, it was reported that the two lower frequency peaks seem to “split” when the samples are

cooled and the splitting is due to red and blue shifting of the transmitted radiation. By cooling the sample, the intermolecular potential and the anharmonicity of the modes are much steeper which in turn causes the blue shifting of the absorption peak. The red shifting effect on the neighbouring peak is more difficult to predict but is thought to be due to the “Van der Waals forces interacting with hydrogen bonds”.

Samples of thymine and thymidine were also made using IPA and are shown in Figure 5-30.

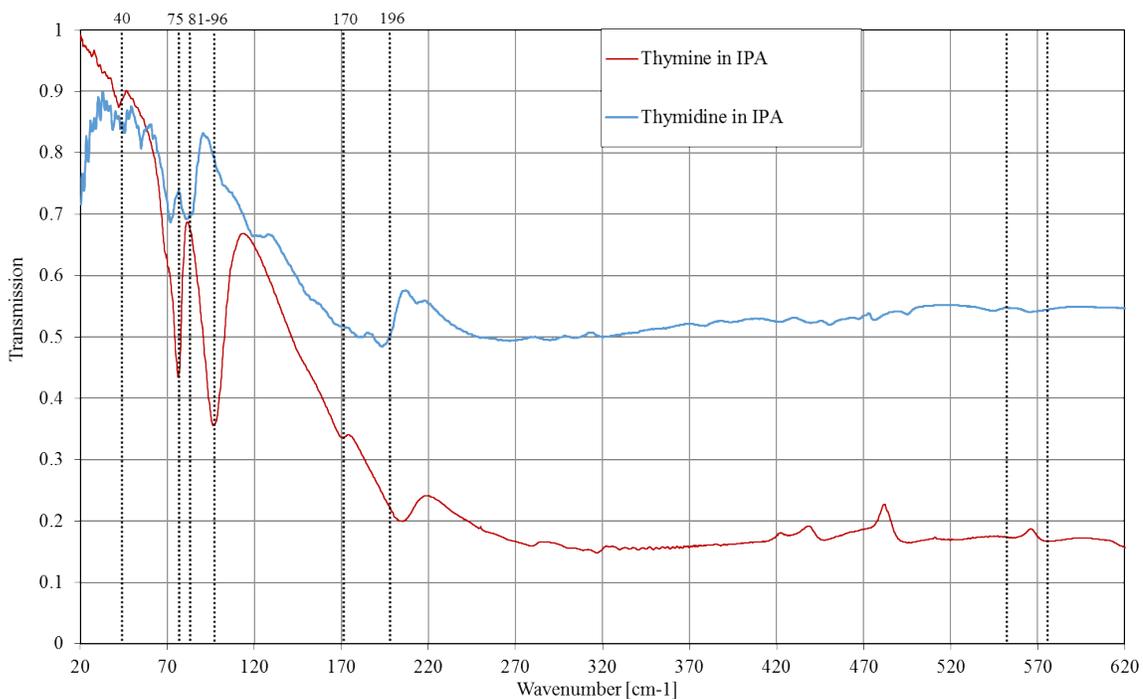


Figure 5-30: THz spectra of thymine and thymidine made using IPA from 20 to 620 cm^{-1}

Highlighted are the same absorption features as seen in the samples made using DI water. This shows that the crystalline structure of thymine and thymidine samples were not

influenced by the different solvents. Figure 5-30 also identifies the absorption feature at $42.0 \pm 2 \text{ cm}^{-1}$ which is present for both thymine and thymidine.

5.1.3.7 Conclusion

In this section, the THz spectra of nucleosides were compared to their complimentary nucleobases. Three of the nucleosides (cytosine, thymidine and adenosine) were compared to literature data taken using a TDS system.

The addition of a sugar ring perturbs the THz spectra therefore an attempt was made in section 5.1.3.2 to identify the THz spectra of glucose in two different forms; pressed into polypropylene and a solution made using DI water and left to evaporate resulting in a thin film. These two methods resulted in different THz spectra because of the difference in the structure of monohydrated and anhydrous glucose. The formation of polycrystalline films of nucleosides and sugars are highly sensitive to environmental conditions, therefore it is not possible to extract unique features from glucose in the nucleoside spectra.

Despite the difference in molecular structure, there were many absorption features that both the nucleosides and nucleobases shared. They predominantly occurred between 40 and 250 cm^{-1} where the spectra of every sample had many absorption features.

Absorption features within the 250 and 400 cm^{-1} region can be associated with the ribose ring and not with the nucleobase. This is highlighted in the THz spectra of cytosine and cytidine which shows the lack of any absorption features in the cytosine sample but a number of features in the cytidine sample.

The nucleosides explored here had the same absorption features for samples made using DI water and IPA. The cytosine samples made using DI water and IPA were compared to the cytidine and it was seen that the absorption features present in the cytosine sample made using DI water was better matched to the cytidine spectra.

THz spectra of guanine and guanosine is unique to this thesis and section 5.2 investigates the identification of the resonances found in these samples in more detail.

5.2 Simulation and experimental data of nucleobases

Density functional theory (DFT) is a way of calculating approximate solutions to a many-bodied systems by using the electron density to investigate the structural, magnetic and electronic properties of molecules and materials.

Hartree assumed that the nucleus and electron of an atom formed a spherically symmetric field i.e. the charge distribution of the electron was used as the solution to the Schrodinger equation for the atom's potential energy.

One way of approximating the interactions between particles is using the Hartree-Fock (HF) theory by taking the average of all the interactions. Each electron is described as an orbital and the total wave function is given as the product of the orbitals.

Density functional theory (DFT) is similar in its approach to the HF model, however it produces better approximations by modelling the many-bodied electron correlation as a function of the electron density.

Ab initio (meaning from the beginning) methods such as energy correlation methods, try to derive solutions to the Schrodinger's equations without any prior fitting parameters or information about the sample. This is possible but is incredibly difficult, time and computationally demanding. As x-ray crystallographic data is available for many crystalline samples such as nucleobases and nucleosides in their dehydrated state, it is possible to use parameters such as lattice dimensions, atomic positions and symmetry to help with the initial setup for the simulation.

5.2.1 Guanine

Simulated data was performed for guanine alpha (Guille 2006 KEMDOW) which suggests it is the crystalline structure of anhydrous guanine given by (Clegg, 2006). The capital letters in the files are references to the original X-ray data sets from the Cambridge Structural Database (<https://www.ccdc.cam.ac.uk/>). It is said that the molecules are linked to each other by N – H···N and N – H···O hydrogen bonds to form sheets, between which there are stacking interactions.

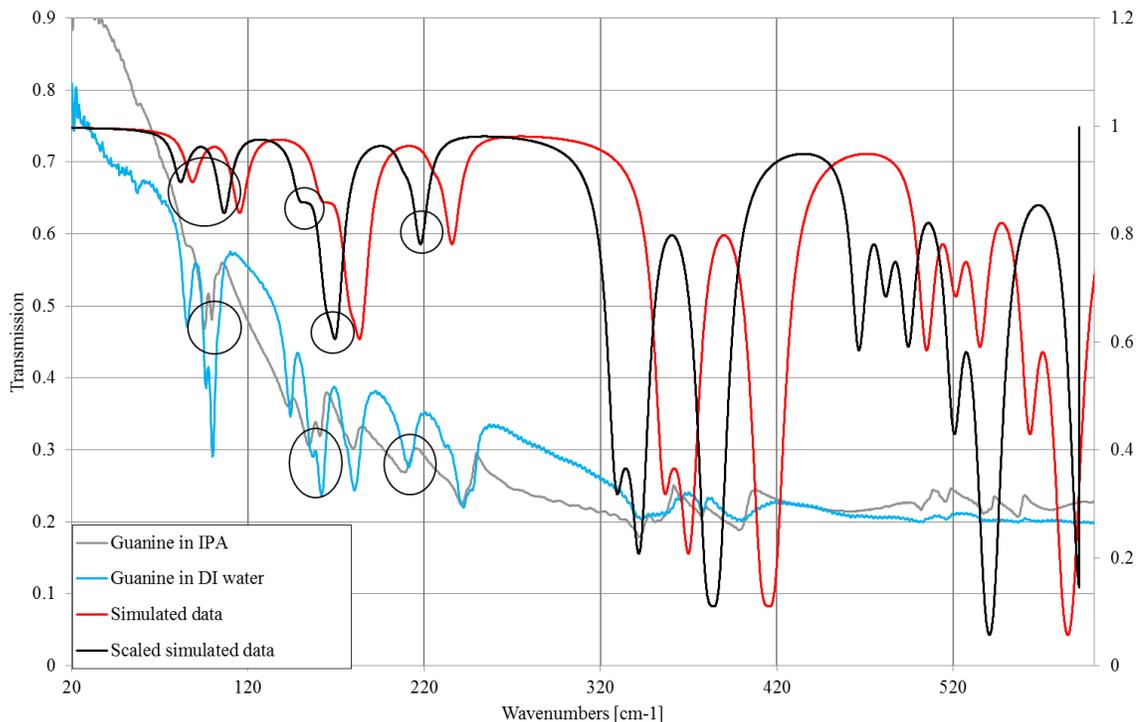


Figure 5-31: THz spectra of guanine made using DI water and IPA, the simulated data using DFT and scaled simulated data of guanine.

The data presented here in Figure 5-31 show the experimental data from guanine made using IPA and simulated data from the DFT modelling. Across the whole spectrum there are some similarities that can be extracted. However, the scattering effects in the experimental data cause some problems in identifying these features at frequencies above 308 cm^{-1} . After scaling the simulated data, some absorption features below 308 cm^{-1} were in better agreement with the experimental data despite the absorption features above 308 cm^{-1} being in better agreement before scaling.

It is reassuring to see a similar trend in the overall spectrum such as the features found between $70 - 260\text{ cm}^{-1}$, $332 - 430\text{ cm}^{-1}$ and $495 - 620\text{ cm}^{-1}$ in both experimental and simulated data.

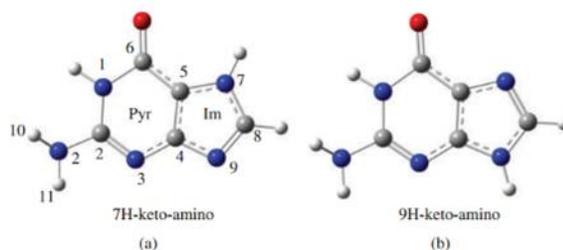


Figure 5-32: Two keto- tautomer of guanine that are the most common structure in anhydrous guanine. Structural representation of the calculated (DFT/B3LYP 6-31G**) four most stable tautomeric forms of guanine, in the gas phase, presented by (Lopes R. P, 2012)

According to (Clegg, 2006) the anhydrous guanine base will favour the 7H-keto-amino structure as seen in Figure 5-32. They will form stacked sheets along the b-plane as seen in Figure 5-33 . The lowest energy conformation calculated for isolated guanine, at the

DFT B3LYP/6-31G level, is the 7H-keto-amino tautomer represented in Figure 5-32 (a) (R. Pedro Lopes, 2012).

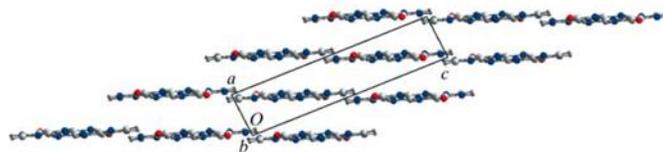


Figure 5-33: Structure of the stacked sheets of the anhydrous guanine molecules viewed along the b-axis.

5.2.2 Cytosine

Molecular modelling of cytosine was performed for two different crystalline forms. From section (1.1.1.nucleobases in solution) it is clear that the dehydrated cytosine samples had significant differences when made using DI water and IPA.

5.2.2.1 Cytosine monohydrate

For cytosine made with DI water, the simulations were based on the monohydrate structure of cytosine given by (Carven & McClure, 1973). This is based on the monohydrate work of (Kinoshita, 1961) using photographic intensity data.

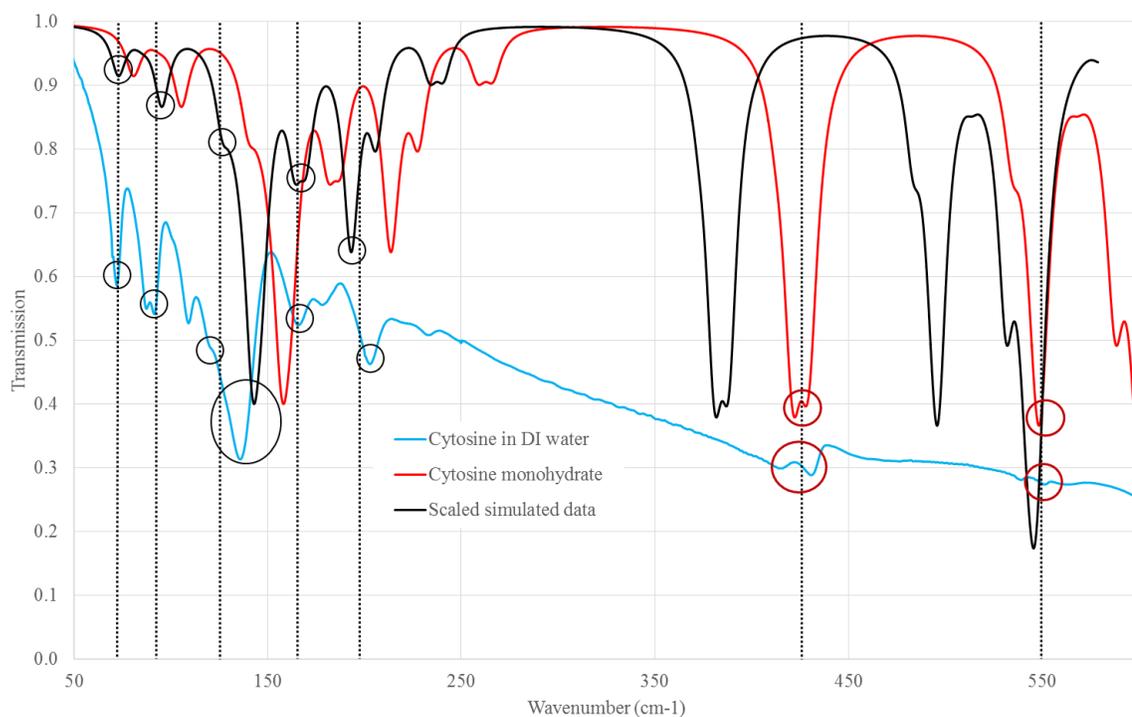


Figure 5-34: THz spectra of cytosine made using DI water and the simulated data of cytosine monohydrate

The minima of each feature for the experimental and the simulated results were plotted on a straight line graph to calculate the scaling factor, which was calculated to be 0.9041. At lower frequencies, the absorption features match well with that found in the real experimental data, however, they do not match well at higher frequencies. At higher frequencies, the original simulated data is better suited to the features found in the experimental data. Figure 5-34 shows the similarities between the experimental data and the scaled simulated spectra below 350 cm^{-1} and the similarities between the experimental data and the original simulated data above 350 cm^{-1} .

The authors in (Kinoshita, 1961) seem to suggest that the cytosine molecules are hydrogen bonded in parallel ribbons which are linked into a three-dimensional structure through hydrogen bonds from the water molecules. If this is true, then the spectra should look different for cytosine monohydrate made using distilled water (H₂O) and deuterium oxide (D₂O). This does not coincide with the results gained from the experimental data. Data for cytosine made using water and heavy water were left to evaporate in the same conditions and is shown in Figure 5-35. Spectroscopy performed using the same FTS on the same date resulted in identical spectra, within error.

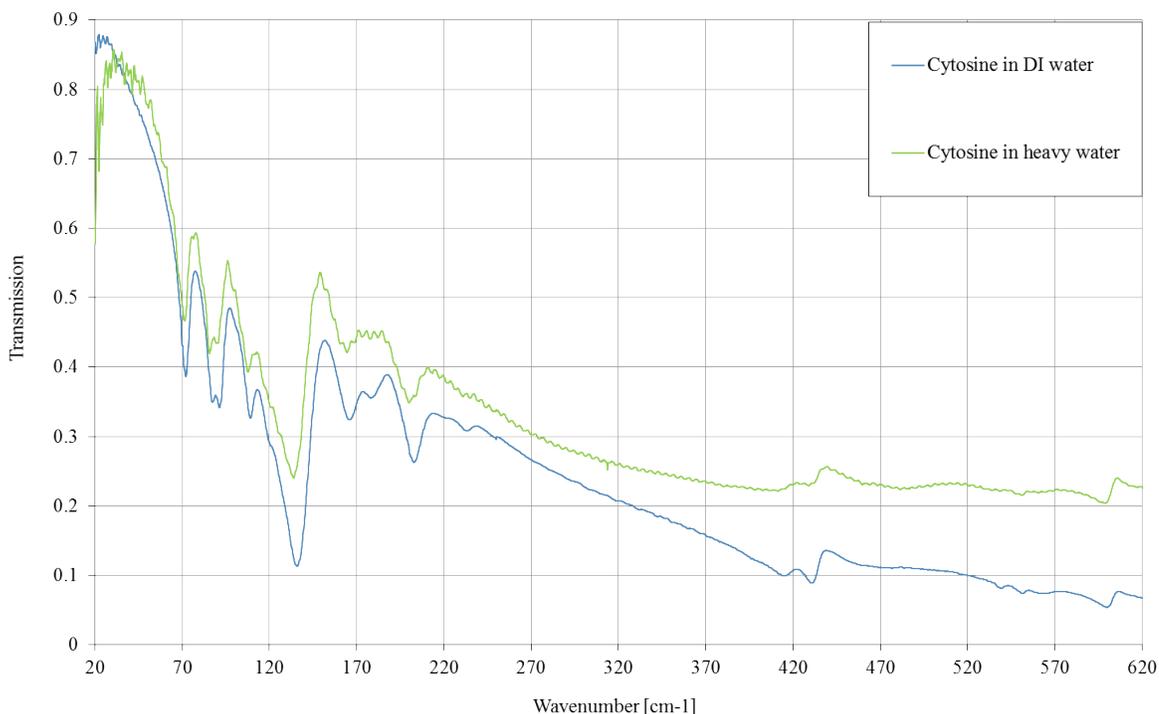


Figure 5-35: THz spectra of cytosine made using DI water and heavy water to show the agreement between the two samples

The crystalline structure seems to be in the form of “ribbons” extending out along the *b* direction. These ribbons are linked into a 3-D structure through hydrogen bonding of water molecules. Each water molecule is hydrogen bonded to two oxygen and one nitrogen atom in three adjacent ribbons as shown in Figure 5-36.

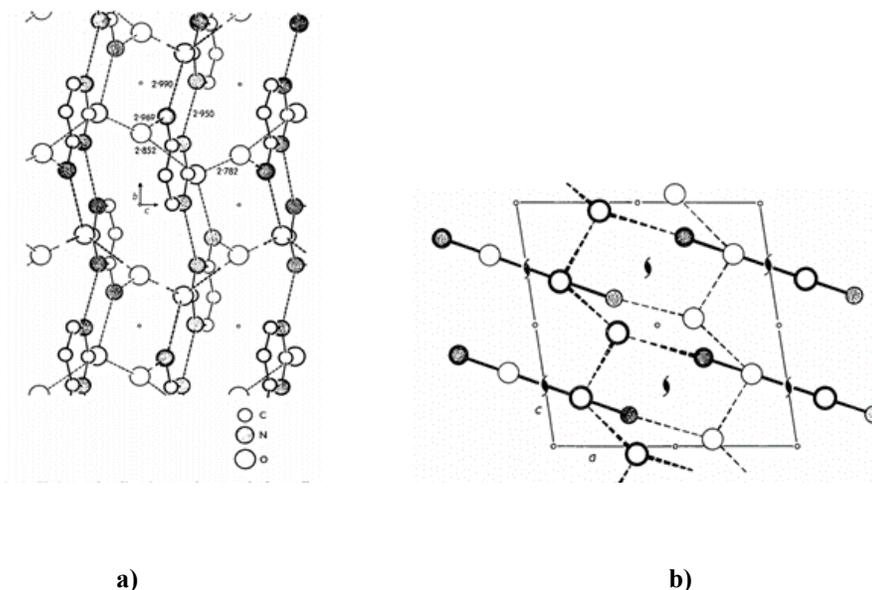


Figure 5-36: Structure of cytosine monohydrate as given by Jeffrey and Kinoshita. a) Shows the hydrogen bonding (dotted lines) along the *a* direction and b) shows the hydrogen bonding (dotted lines) viewed along the *b*-axis

5.2.2.2 Cytosine alpha polymorph

For cytosine made with IPA, the simulations were based on the cytosine alpha polymorph. The simulated data was quoted to have formed the structure shown by

(Craven, 1973) which is a more recent investigation of cytosine and its monohydrate. This work builds on the work by (Kinoshita, 1961) and (Marsh, 1963).

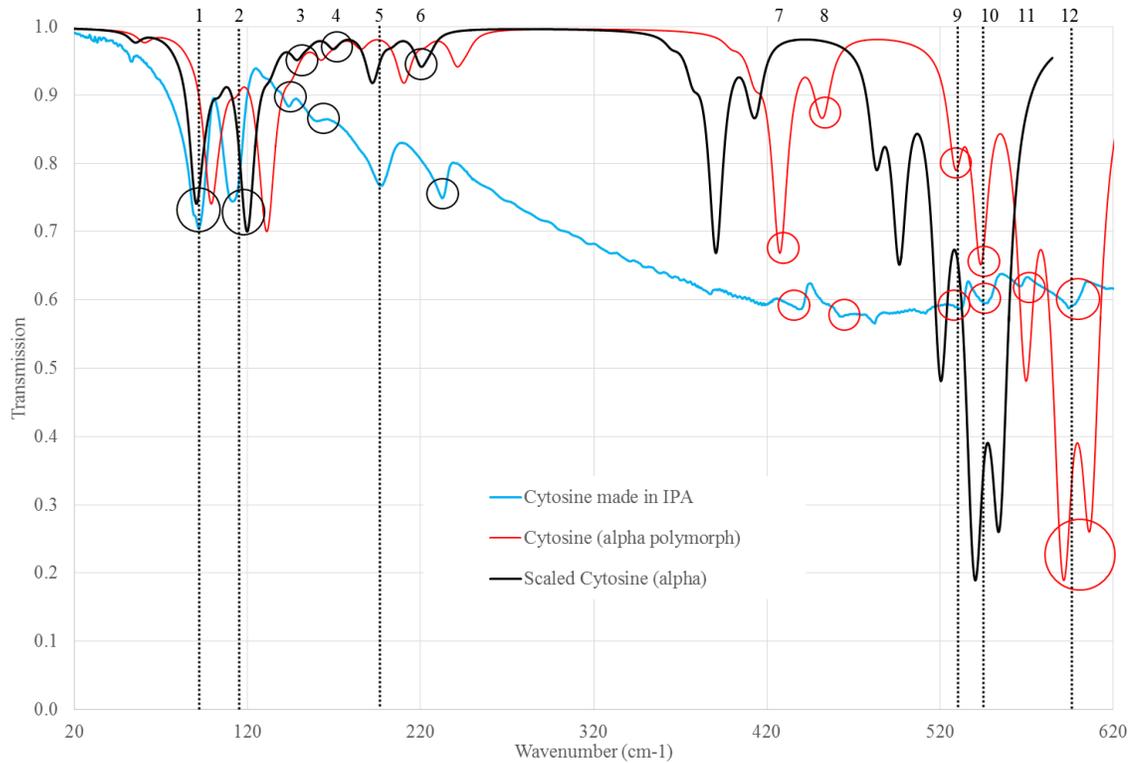


Figure 5-37: THz spectra of cytosine made in IPA and the simulated data of cytosine alpha polymorph

In Figure 5-37, the scaled simulated data match the experimental data well at lower frequencies but at higher frequencies it is difficult to match any features from the simulated data (original and scaled) to the small features found in the experimental data. The scaling factor used was 0.9712.

For the simulated cytosine (alpha polymorph) it is not clear what is meant by this conformation and it proved impossible to obtain a specific paper on the alpha polymorph

of cytosine. However, this is assumed to be the preferred structure of anhydrous cytosine when made using methanol solutions according to (Marsh, 1963).

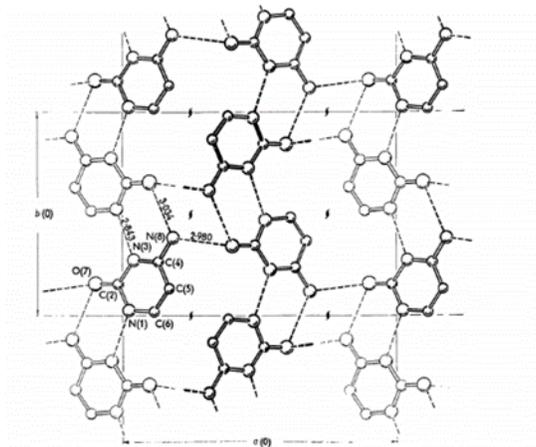


Figure 5-38: Structure of anhydrous cytosine assuming that the molecules are in the amino form, viewed along the c-axis shown by (Marsh, 1963)

Figure 5-38 shows the structure of the anhydrous cytosine as seen along the c-axis with the dashed lines representing the hydrogen bonds that have formed at all three available proton sites; two from the amino nitrogen (N8) to the oxygen (O7) atoms and one from another nitrogen (N1) to nitrogen (N3) atoms.

The difference in the bond distances of monohydrate cytosine and anhydrous cytosine, including cytosine-5-acetic acid, which is not relevant for this work, is shown in (Marsh, 1963).

5.2.3 Thymine

The simulated data for thymine was in very good agreement with the experimental data such that scaling was not needed. The simulated data is placed on a secondary axis for better comparison to the experimental data. The data for thymine made using DI water and IPA have been plotted for comparison in Figure 5-39.

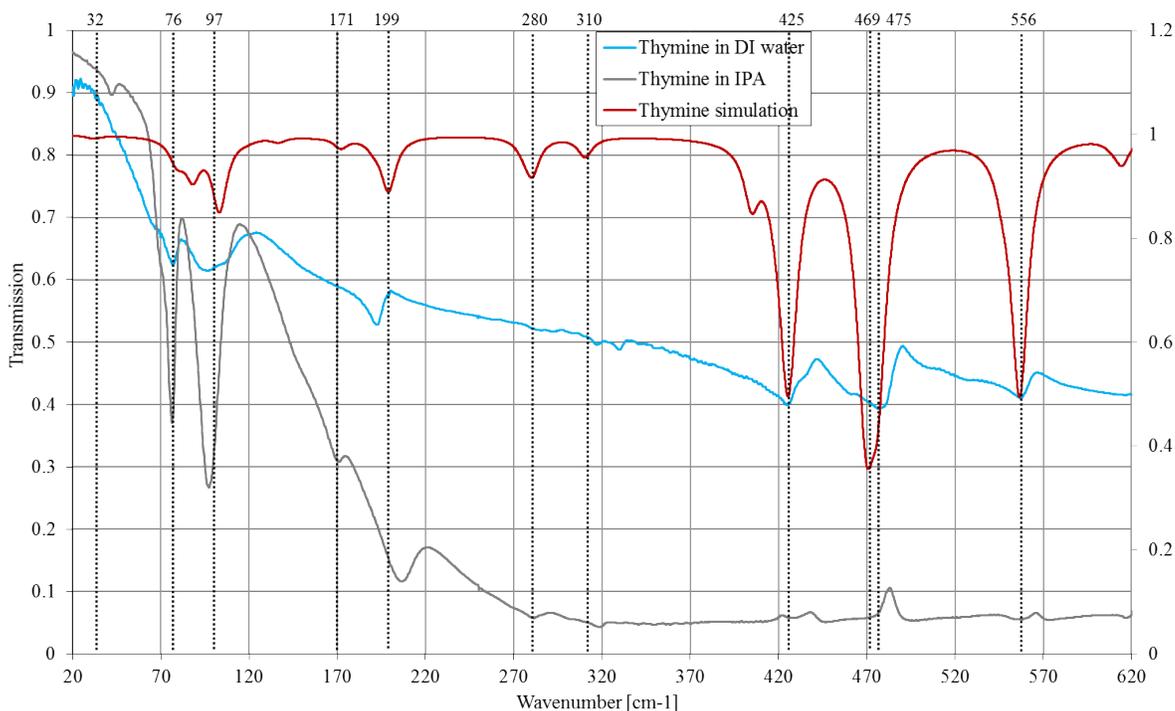


Figure 5-39: Simulated THz spectra of thymine with the main absorption features identified and compared to experimental data of thymine made using DI water and IPA

According to (Fischer, Walther, & Jepsen, 2002), who also use DFT simulations to compare the absorption features found in thymine crystals using THz TDS system, the four lowest-frequency absorption features arise from intermolecular motions in the form of out-of-plane and in-plane vibrations of the hydrogen bonds. However, they state that

the numerical procedure presented in their work cannot be a definitive assignment of the experimental data. Another comparison is made to the IR data by (Zhang, Michaelian, & Loppnow, 1998) which show good agreement to the experimental data as well as the DFT simulated data highlighted in Figure 5-39 and are compared in Table 5.8

Table 5.8: Comparison of the absorption frequency of literature IR spectra, THz experimental data and DFT simulated data of thymine

Natural abundance Thymine (cm-1)	Association	Thymine made with IPA (cm-1)	DFT Simulated thymine (cm-1)
209	Torsion (ring)	205.39	199.0106
284	Deformation (C—CH ₃)	281.50	280.3835
321	Wag/out-of-plane deformation (C—CH ₃)	318.8	310.4195
403	Deformation (CO)		404.7879
432	Torsion (ring)	425.403	425.2498
		461.4413	469.9477
475	Deformation (ring)	476.8143	475.9081
559	Deformation (ring)	557	556.681

(Zhang, Michaelian, & Loppnow, 1998) have also assigned each absorption feature with a corresponding mode. It highlights that the modes identified at these frequencies are intermolecular interactions between the hydrogen bonded molecules.

5.2.4 Adenine

Simulation of adenine were made on the structure of “Adenine form I” according to the work of (Sudarshan Mahapatra, 2008).

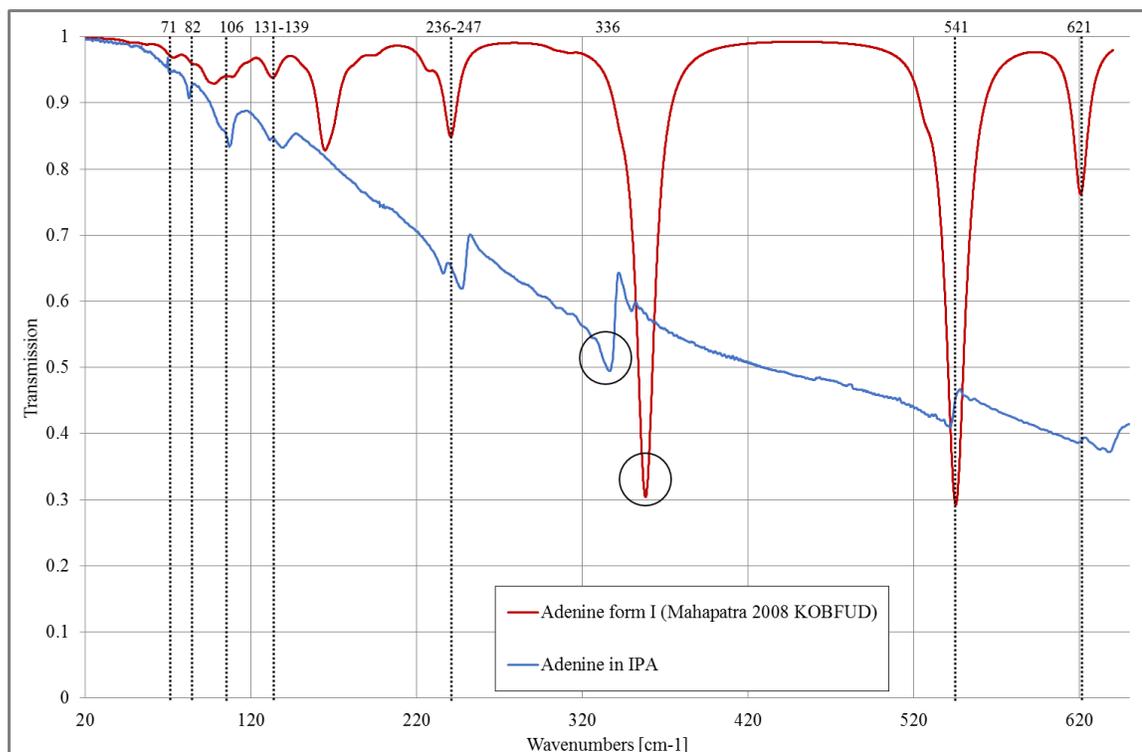


Figure 5-40: THz spectra of adenine made in IPA and the simulated data of adenine form I

According to (Sudarshan Mahapatra, 2008) “Attempts to crystallize anhydrous adenine always resulted in the formation of the trihydrate. The structure of adenine trihydrate results in stacks of adenine molecules interspersed with two layers of water molecules. The chains and water sheets are connected via N-H \cdots O and O-H \cdots O hydrogen bonds, and it is of interest to note that there is no N-H \cdots N hydrogen bond.”

This is the similar case for guanine monohydrate, as also being a purine molecule, therefore the double ringed nature of their molecular structure could cause this water-connecting sheets to form.

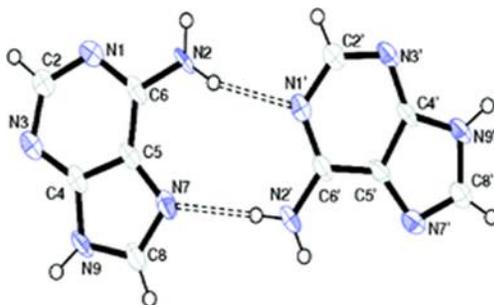


Figure 5-41: Structure of adenine hydrogen bonded via N-H---N to another adenine molecule given by Mahapatra et al. (Sudarshan Mahapatra, 2008)

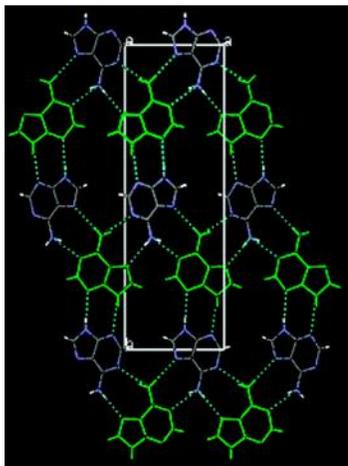


Figure 5-42: Packing diagram of adenine showing the hydrogen bonded molecules (dotted lines) viewed along the c-axis given by Mahapatra et al. (Sudarshan Mahapatra, 2008)

5.2.5 Uracil

Molecularly, uracil is the simplest molecule of the nucleobases and has the least complicated spectrum.

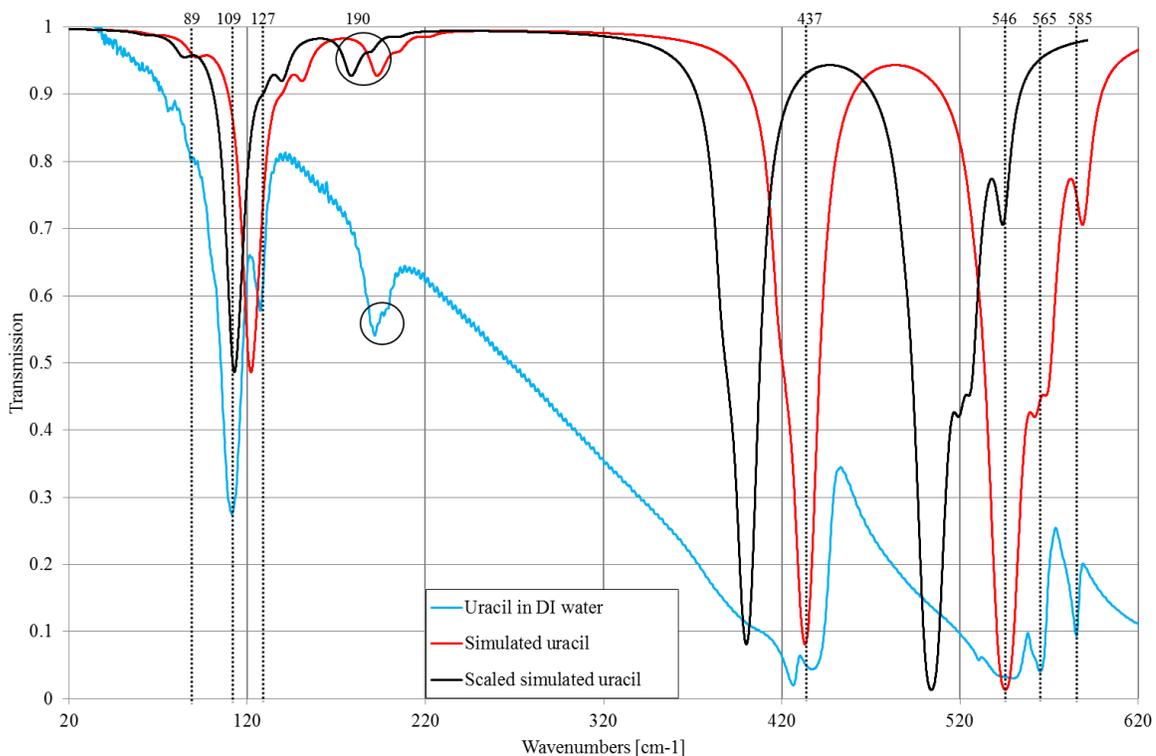


Figure 5-43: THz spectra of uracil made in IPA and the simulated data of uracil

The simulated data seem to match the experimental data well as seen in Figure 5-43. By scaling the spectrum, the lower frequency absorption features match the experimental data better. The scaling factor used was 0.9239 but as seen in the other spectra, at higher frequencies, the original simulated data matches the experimental data better.

Crystallographic experiments on the structure of uracil have established that the molecules arrange themselves into parallel sheets parallel to the c-axis.

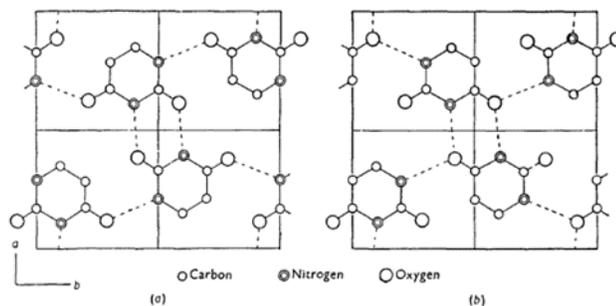


Figure 5-44: a) initial assumption of the crystalline structure of uracil b) actual structure of crystalline uracil after refined electron-density maps in which one oxygen atom makes two hydrogen bonds with a nitrogen atom

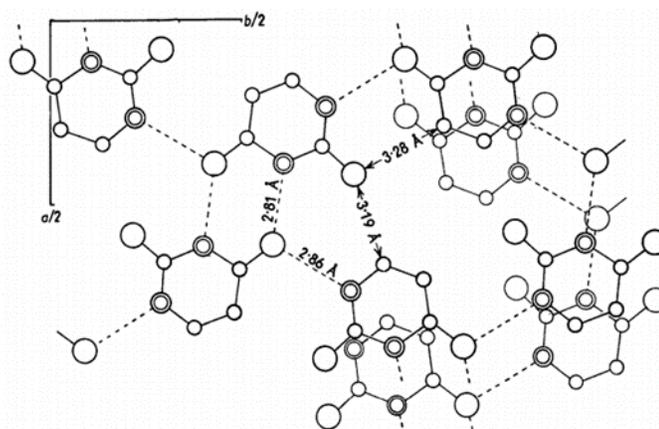


Figure 5-45: Crystalline structure of uracil projected on to (0,0,1). The left side shows one layer of the hydrogen bonded molecules whereas the right side of the image shows the overlay of another layer

As can be seen in both Figure 5-44 and Figure 5-45, each molecule has an oxygen atom which has two hydrogen bonds to a nitrogen atom. To complement the simulated data and to aid visualisation of the resonances within the samples, animations were created of the movements at particular frequencies. These animations are frozen in time in the

following figures (5-46 to 5-52) with each one relating to a different resonance in the experimental uracil spectrum. The formula for uracil is $C_4H_4N_2O_2$; the atomic species are displayed in different colour in the animations whereby the grey atoms represent the carbon (C_4), the blue atoms represent the nitrogen (N_2), the red atoms represent the oxygen (O_2) and the small white atoms represent the hydrogen (H_4) species. Figure 5-46 and Figure 5-47 attempt to illustrate the molecules “wobble” whereby one side of the molecule will move in the opposite direction to the other side.

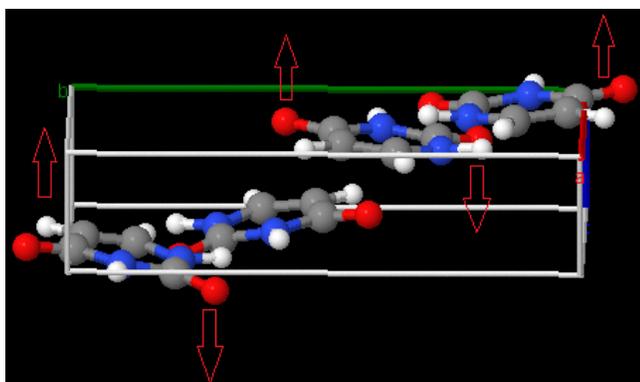


Figure 5-46: Resonance at 90 cm^{-1} . The atoms on “the left” with respect to the view of the reader will move upwards as the right atoms move downwards.

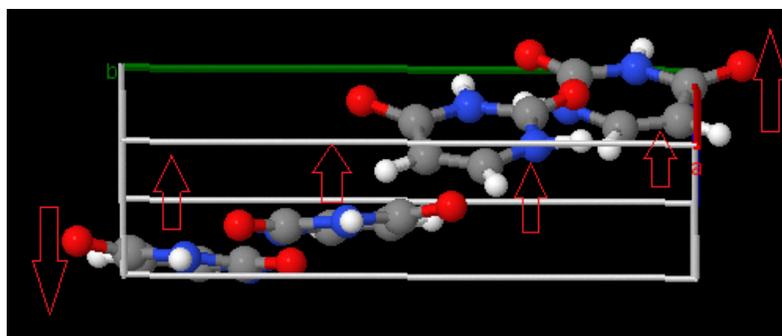


Figure 5-47: Resonance at 121 cm^{-1} . The atoms at "the back" with respect to the view of the reader, will move upward as the front move down.

The images found in Figure 5-48 and Figure 5-49 are shown along the c-axis and identifies the twisting and shifting of the molecules.

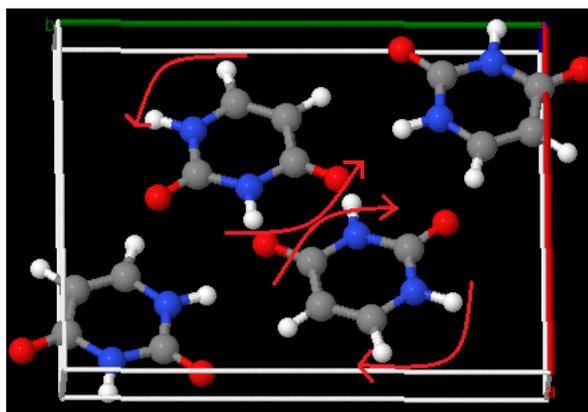


Figure 5-48: Resonance at 122 cm^{-1} . This is looking down the c-axis and showing the whole molecule twisting.

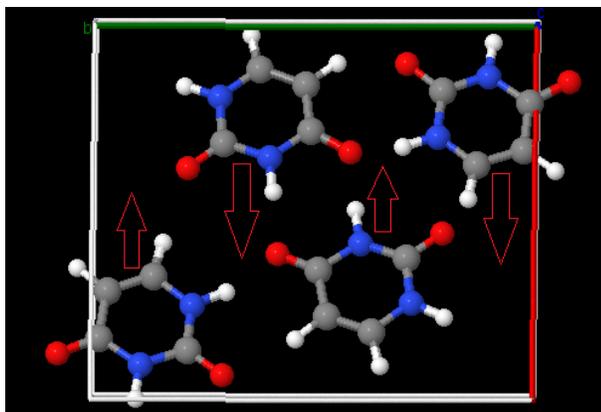


Figure 5-49: Resonance at 125 cm^{-1} . This is looking down the c -axis and showing the whole molecule shifting upwards and downwards.

The animations for the higher frequency resonances have also been represented in this way, however, the motion of the atoms and molecules become increasingly more difficult to demonstrate in 2D images. Fig shows the 418 cm^{-1} resonance, which shows the bending of hydrogen atoms and a small bend in the collective carbon atoms attached either side.

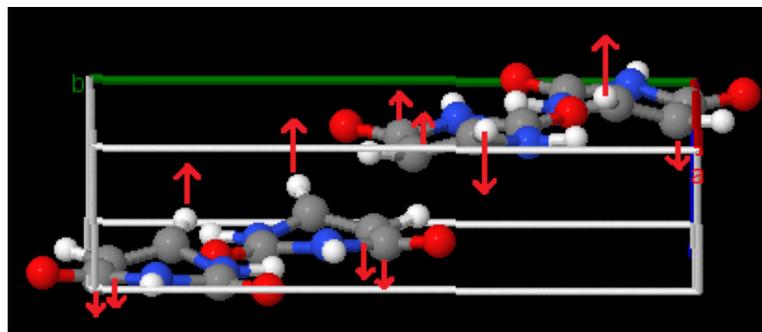


Figure 5-50: Resonance at 418 cm^{-1} . This shows the bending of the hydrogen atoms at a much greater amplitude than the bending of the attached carbon atoms. As the hydrogen atom moves “upwards” the carbon atoms move “downwards”.

The resonances at 433 cm^{-1} show some symmetric bending of the two oxygen atoms with respect to the NH atom along the a-axis as shown in Figure 5-51.

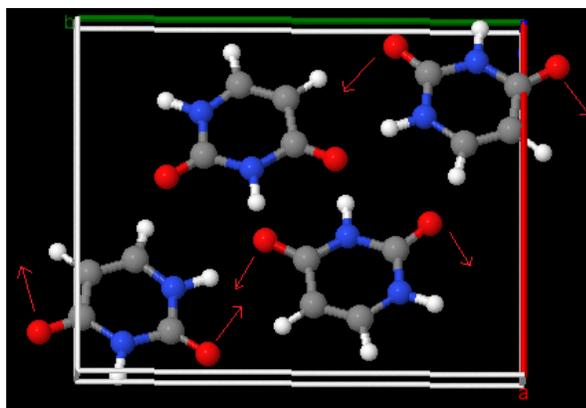


Figure 5-51: Resonance at 433 cm^{-1} . The oxygen atoms (red) are symmetrically bending with respect to the nitrogen and hydrogen atoms along the a-axis.

Finally, Figure 5-52 represents the vibrational modes at 545 cm^{-1} where the molecule is shown to asymmetrically stretch along the a-axis as viewed along the c-axis.

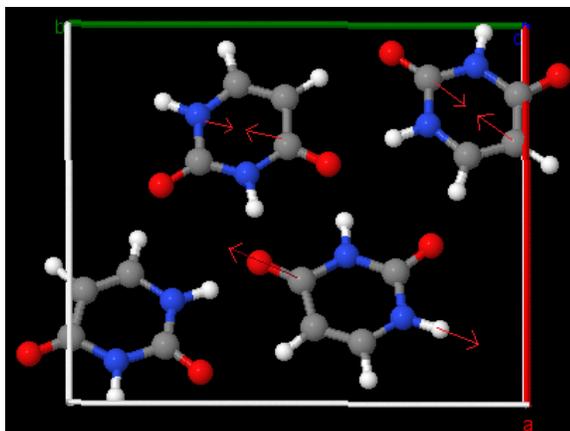


Figure 5-52: Resonance at 545 cm⁻¹. The CO and NH molecule bend in and out as the remaining attaching atoms bend in the opposite direction.

The nucleosides, thymidine, cytidine and uridine were also simulated and compared to the experimental results. These are more complicated structures as shown in section 5.1.3, and demonstrate a more complicated and less complimentary spectra. Their analysis is still in progress.

It is important to state that modern DFT approximations are quite accurately predicting the solid state molecular motions and that there is an encouraging coincidence between predicted and measured resonances. However, deviations in resonant frequencies will arise due to incorrect assumptions about bond length, intra and intermolecular distances, temperatures and how these scale with frequency.

These animations for uracil present an insight into the motion of the molecules individually and their movement as a unit. They clearly indicate that this frequency region (0.6 to 18 THz) is essential in better understanding the movement of larger molecules, in particular nucleobases and nucleosides. If this region can be sensitive

enough to show strong absorption features from crystalline structures, there is also potential of being able to probe deeper into larger biological molecules. Section 5.3 presents the results from the application of this technique of Terahertz spectroscopy to DNA.

5.3 Terahertz spectroscopy of DNA

DNA is a very robust polymer and is routinely dehydrated and rehydrated for biological experiments, such as polymer reaction chain (PCR) methods. This method is used to speed up the DNA replication process to obtain a higher concentration of the same DNA chain. Some infrared spectroscopy techniques have been used to elucidate whether DNA has denatured or changed its conformation during rehydrating the sample (Wood, 2016). However, this has not been extended to the THz region, which is essential for understanding the long-range conformational changes due to the method's sensitivity to the hydrogen bonding.

A gene is a length of DNA coded for a particular job or purpose, e.g. a specific gene will be responsible for the production of the protein insulin. It was predicted that humans would have up to 100,000 or more genes as we are assumed to be the most complex creature that we know. However, during the genome project, it was found that humans have approximately 20,500 protein-coding genes⁴ (Institute, 2016). This only accounts for a small percentage of the entire genome (~1.5%) with the other ~98.5% being thought to be non-coding molecules such as DNA, RNA and other regulatory molecules.

To investigate the potential of THz spectroscopy in identifying changes in the conformation of DNA, samples of human DNA and RNA were obtained from the Wales Cancer Bank (WCB) and samples of mouse DNA were kindly provided by the Stem Cell Research Institute, Cardiff University. All the sample solutions were left to evaporate on 2µm of polypropylene (PP) and investigated using the Fourier Transform Spectrometer

⁴ <https://www.genome.gov/12011238/an-overview-of-the-human-genome-project/>

(FTS) (see CHAPTER 3 for more details) where the samples were held in vacuum. The results from the investigation are separated into five sections presenting the THz spectra of: human DNA and RNA (5.3.1); human and mouse DNA (5.3.2); different colonies of mouse DNA (5.3.3); heating effects on mouse DNA (5.3.4) and time effect of one colony of mouse DNA (5.3.5).

5.3.1 THz spectroscopy of human DNA and RNA

DNA strands contain all the information that an organism needs to function, but ribonucleic acids, RNA, are the molecules that extract and transfer the genetic codes for the creation of proteins. DNA is a double stranded molecule formed into a tight helix through the hydrogen bonding of the nitrogenous base pairs. RNA is a single stranded molecule with a uracil base instead of a thymine base molecule.

DNA tends to hold its conformation even after being dehydrated but RNA is structurally variable as it has many different geometries depending on its function (Difflen.com, 2017). Due to RNA being present in most organic materials and living organisms, it is difficult to handle RNA and keep it in its natural state. RNA is very complex and can sometimes form into a secondary structure but more often than not, when it is dehydrated it tends to break down.

The samples were received in individual aliquots of DNA and RNA in a buffered solution labelled as shown in Table 5.9. The concentrations were unknown.

Table 5.9: Content of samples from Wales Cancer Bank

SAMPLE LABELLING	TISSUE	BARCODE	DIAGNOSIS
WCB DNA sample 1	Kidney	RWMBV0000123FN 1ADS19	Papillary renal cell carcinoma
WCB DNA sample 3	Kidney	RWMBV0000123FN 1ARS3	Papillary renal cell carcinoma
WCB RNA sample 2	Kidney	RWMBV0000138FN 1ADS8	Clear cell carcinoma
WCB RNA sample 4	Kidney	RWMBV0000138FN 1ARS8	Clear cell carcinoma

Identification from the barcode; ADS was representative of DNA and ARS represented RNA. The samples were all from nominally healthy tissue but, as the WCB do not have healthy controls, they use non-involved tissue from a cancer patient i.e. an area of tissue that has been removed from a cancer patient but has no cancer present as determined by a pathologist. The “diagnosis” column is the difference between the cancers of the patient. Figure 5-53 presents the THz spectra of the dehydrated human DNA and RNA samples.

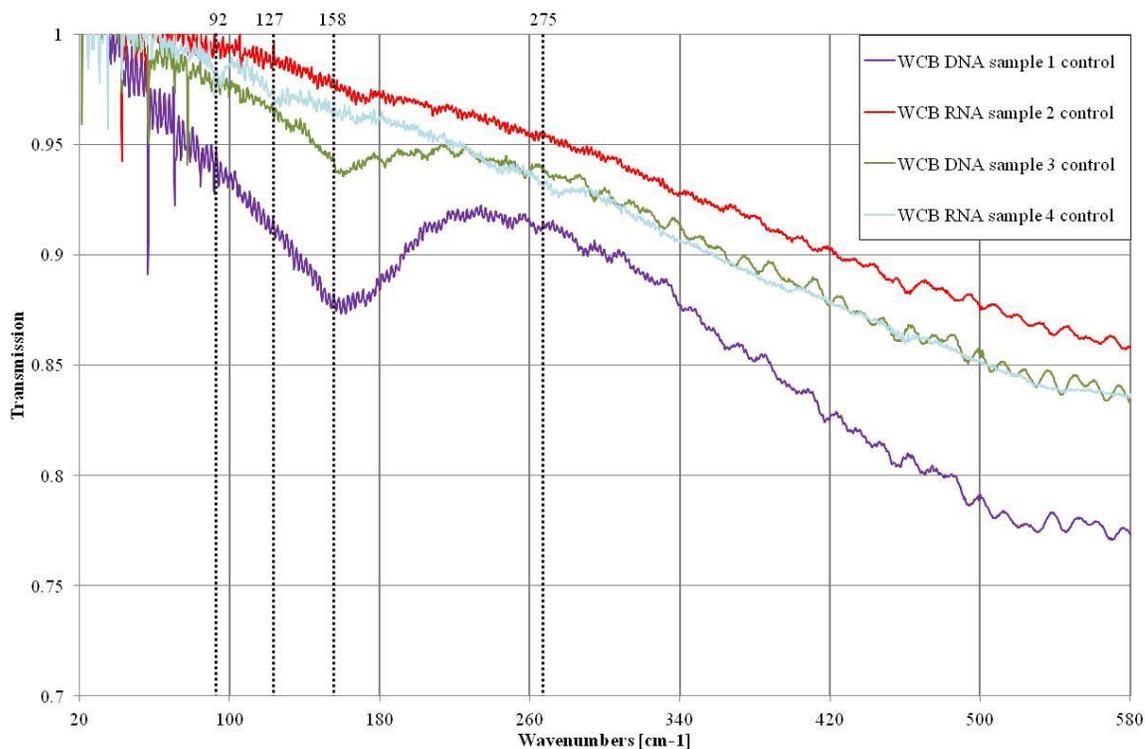


Figure 5-53: THz spectra of Wales Cancer Bank (WCB)-provided DNA and RNA samples after dehydration on 2µm PP from 20 to 580 cm⁻¹

Figure 5-53 shows the spectra of the four samples identified by number in accordance to Table 5.9 (sample 1 and 3 = DNA, sample 2 and 4 = RNA). There is a common broad absorption feature in the DNA samples at 158 cm⁻¹ which is not present in the RNA samples. For the RNA sample, the spectra show a few unique absorption features. These are highlighted in Figure 5-53 at 92 cm⁻¹, 127 cm⁻¹ and 275 cm⁻¹. These are very small but are significant enough to be present in all the RNA samples but not in the DNA. The spectra of all the samples show a small fringing effect induced by small motions of the aperture in the FTS. It is also further heightened by changing the scale of the transmission axis.

The absorption feature at 158 cm^{-1} in Figure 5-53 could be an indication towards the collective vibrational modes from hydrogen bonds within the molecule which is lacking in RNA due to its unstable nature when dehydrated.

Investigating dehydrated human RNA and DNA in the THz proves to be very difficult without prior knowledge of other factors such as its conformation in the dehydrated state, the nucleic acids sequencing, its length, etc. as they tend to be very sensitive to their surroundings and can break down.

5.3.2 A comparison of the THz spectra of human and mouse DNA

Generally, mouse and humans share the same sets of genes. Most of the genes found in one species are found in another or one that is closely related (Importance of mouse genome, 2010). However, as reported in Nature (Feng Yue, 2014), the authors found many variations in the DNA and gene expression between the mouse and human such as the activity and regulation of genes within the immune system and some metabolic processes.

All the DNA samples received from the Stem Cell Research Institute were in a solution of Ultrapure water. Ultrapure water is achieved by removing any solids, salts or dissolved gas. The samples were all prepared in the Bioscience laboratory of Cardiff University and were kept in a dark fridge within hermetically sealed containers. The samples are identified in Table 5.10.

Table 5.10: Labelling and identification of the colony and mouse number for each DNA sample

COLONY	MOUSE NUMBER
WKD	238
FAX	221
PAC	1395

The different letter labelling for the mouse DNA corresponds to the colony the mouse came from and the following number is the ID of the specific mouse. In general, the colonies will contain the same assortment of genetic modifications. Figure 5-54 shows the THz spectra of the human DNA shown in section 5.3.1 and the mouse DNA.

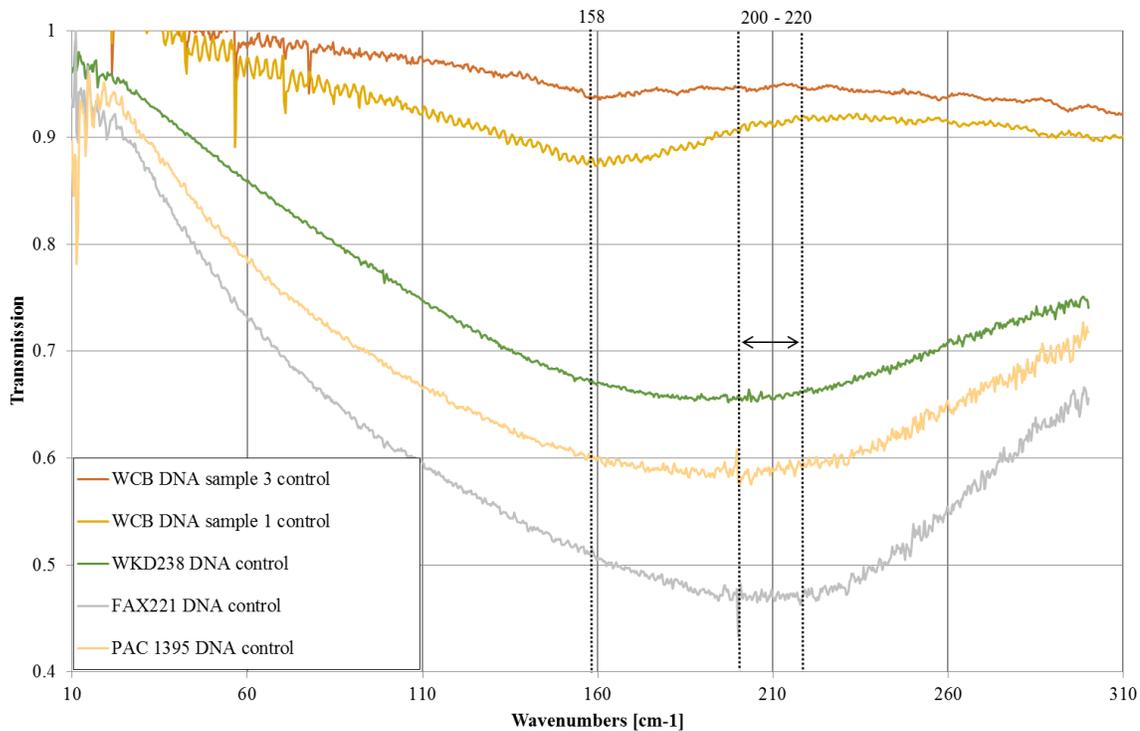


Figure 5-54: THz spectra of dried human DNA from Wales Cancer Bank and dried mouse DNA from biosciences department from 10 to 310 cm^{-1}

From Figure 5-54 there is a difference between human DNA and the mouse DNA. The spectra for the mouse DNA has a common absorption feature at $200 \text{ cm}^{-1} - 220 \text{ cm}^{-1}$. However, the human DNA has its absorption feature at 158 cm^{-1} .

It is interesting to notice the similarities between the mice DNA considering that they came from different colonies. However, the difference in the broad absorption feature between the human and mice DNA proposes that there might be a THz sensitivity to DNA from different organisms. However, this is only a speculation as it is not possible to identify what these broad features are attributed to. Both human and mouse DNA were donated to this research, but unfortunately the specifics of the samples were not available.

Therefore, the difference in the THz spectra presented here could be representative of variations within the samples such as: length of the DNA strands, the nucleic acids sequencing, and potentially, the conformation of the sample once dehydrated. The nucleobases and nucleosides have a unique THz spectrum for different conformations in the solid state (section 5.1) due to the hydrogen bonding. The broad absorption feature here could be a combination of resonances from the nucleobases and nucleosides. This would require further investigation with better control of the samples. An attempt at controlling some factors was achieved in the following sections.

5.3.3 THz spectroscopy of various colonies of mouse DNA

For a better understanding of the THz spectra of mice DNA, there was a need to control some variables. Eight aliquots of mice DNA in solution, each from different colonies were left on PP substrates for approximately 24 hours to dry in the same environment. The concentrations of these samples were unknown, but the identification of each sample is shown in Table 5.11.

Table 5.11: Identification of the colony and mouse number for each DNA sample

COLONY	MOUSE NUMBER
AXN	334
KP	77
FAX	221
WKD	212
PAC	1395
HTF	1412
MALC	1396
LAV	1570

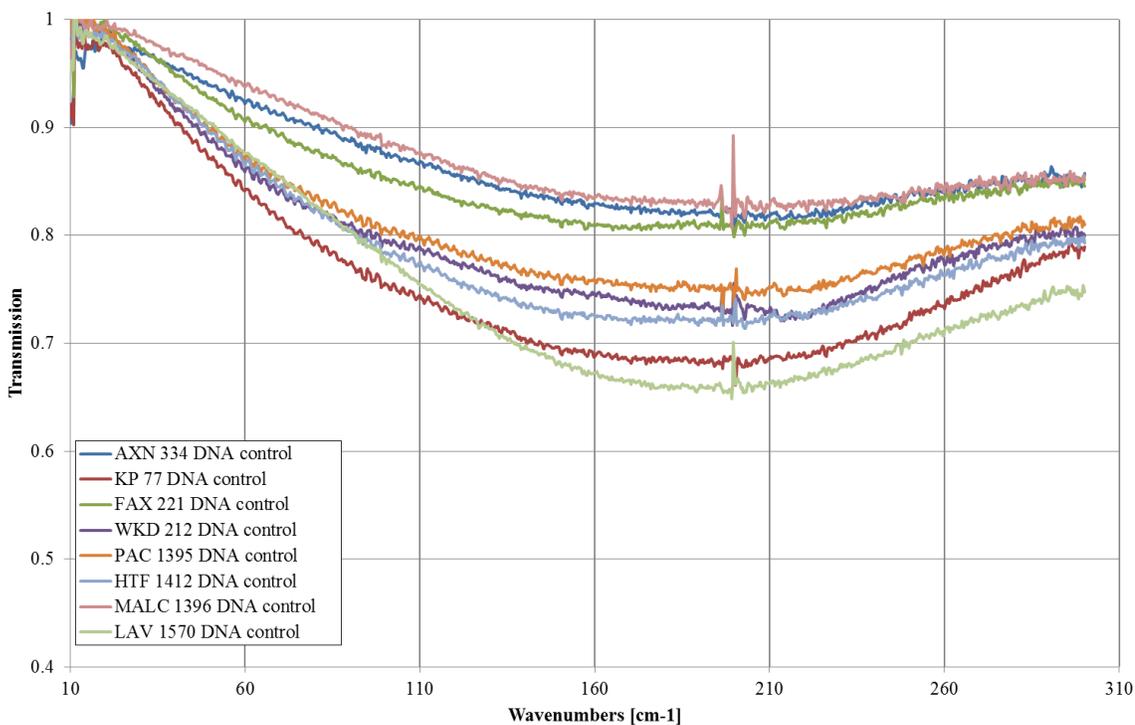


Figure 5-55: THz spectra of mouse DNA from different colonies dried at room temperature and measured from 10 to 310 cm^{-1}

Figure 5-55 shows the spectra of all the different mouse DNA samples with the common absorption feature visible between 200 cm^{-1} and 220 cm^{-1} . The variation in the absolute absorption will be due to unquantified differences in the sample concentration and/or the alignment of that sample with the aperture within the FTS.

The repeatability of the absorption feature in Figure 5-55 suggests that, either the DNA present in the samples have the same conformation, and/or this broad absorption feature is due to the accumulation of thousands of hydrogen bonding sites within the DNA molecule. The results in Figure 5-55 have been treated identically within the process of dehydration and collecting the spectroscopic data. It was not clear what the length of the

DNA strands were, the concentration of DNA within the solution, the nucleic acid sequence, etc. but it is interesting to note the similarities between the spectra of dehydrated DNA from individual mice.

5.3.4 THz spectroscopy of mouse DNA after heating

Double stranded DNA can be exposed to heat in order to artificially to uncoil the molecule for transcription and replication processes. This laboratory technique is called polymerase chain reaction (PCR) and is often used to accelerate DNA replication. The ‘melting’ temperatures, T_m , of DNA, which is the temperature at which the DNA strands are in a single-stranded conformation, are in the region between 65-90 °C (Fang Liu, 2007). This temperature varies depending on different parameters, such as the specific nucleotide sequence and the length of the DNA strand. To investigate the effects of heating, mouse DNA was heated at temperatures between 70 ± 5 °C⁵.

Table 5.12 shows the amount of time they were left at temperature. The “method” column is numbered as they had slightly different methods; 1) each samples were placed on a PP substrate and left in an oven for the time specified, 2) the DNA samples were left in the solution inside their tubes and were placed in a water bath for the time period specified.

⁵ These temperatures were also chosen for this thesis as these were the temperatures that resulted in an interesting THz spectra.

Table 5.12: Table showing the times and temperatures of each DNA sample

SAMPLE	TIME (minutes)	METHOD
HPPT2 93	38	1
HPPT2 98	69	1
HPPT2 99	35	1
HPPT2 102	58	2
HPPT2 89	57	1

The data shown in Figure 5-56 is of mouse DNA from the same colony, HPPT2.

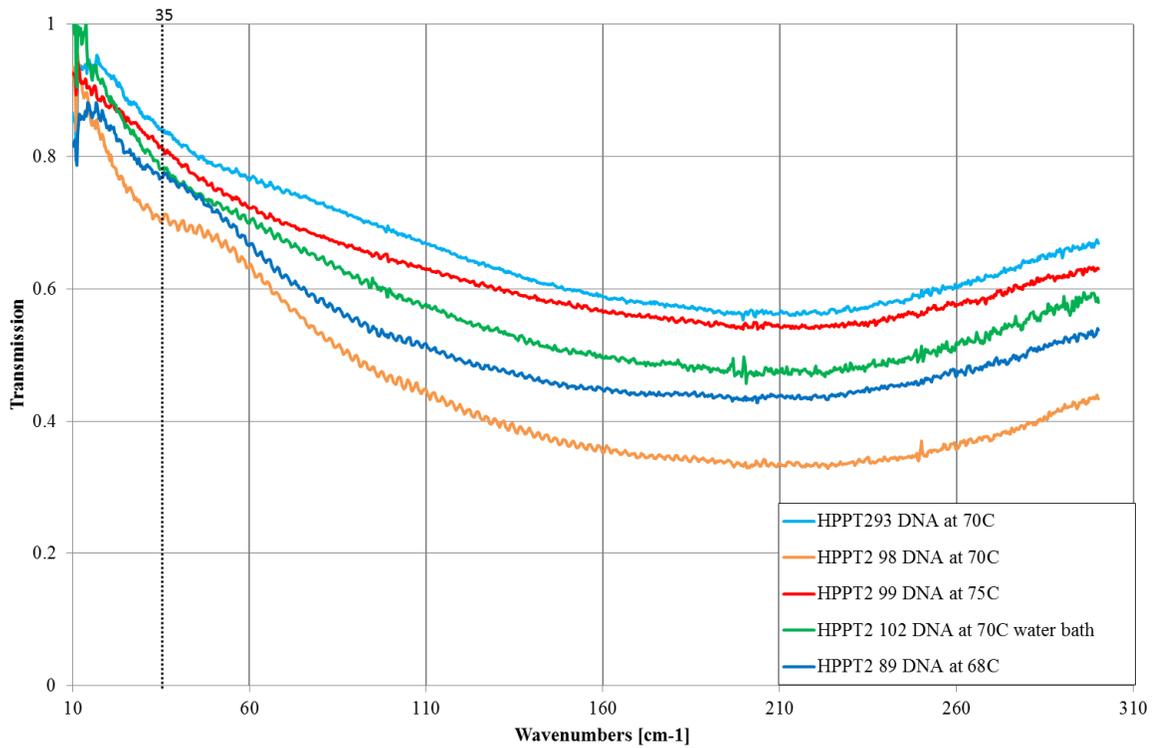


Figure 5-56: THz spectra of mice DNA from the same colony (HPPT2) heated to temperatures of 70 ± 5 °C

Figure 5-56 shows a broad absorption feature centred at 210 cm^{-1} that is a constant in all mice DNA samples, but there is also a smaller feature emerging at 35 cm^{-1} . This could represent a change in the DNA structure, potentially a signature of single stranded DNA, after heating. As noted, the samples were donated in solution form using Ultrapure water therefore it eliminates the suggestion that any salts or other solids could have influenced the absorption features within the spectra of the DNA.

However, the number of unknown factors with the DNA samples means that it is still not conclusive that this additional absorption feature at 35 cm^{-1} is due to a change in the DNA conformation. Not having the information such as the concentration of DNA strands, their length, the nucleic acid content, etc. means that it is not possible to calculate the correct melting temperature, hence why a range of temperatures were tested.

5.3.5 THz spectroscopy of mouse DNA over time

It was assumed that once DNA is dehydrated, it would remain in the same conformation. However, there is no evidence to confirm this as dehydrated DNA has not been widely researched. Therefore it was decided to perform a series of spectroscopic runs over a timescale to check any ‘aging’ effects. Figure 5-57 presents the spectra of the mice DNA samples identified in section 5.3.3 after leaving at room temperature for 2 days.

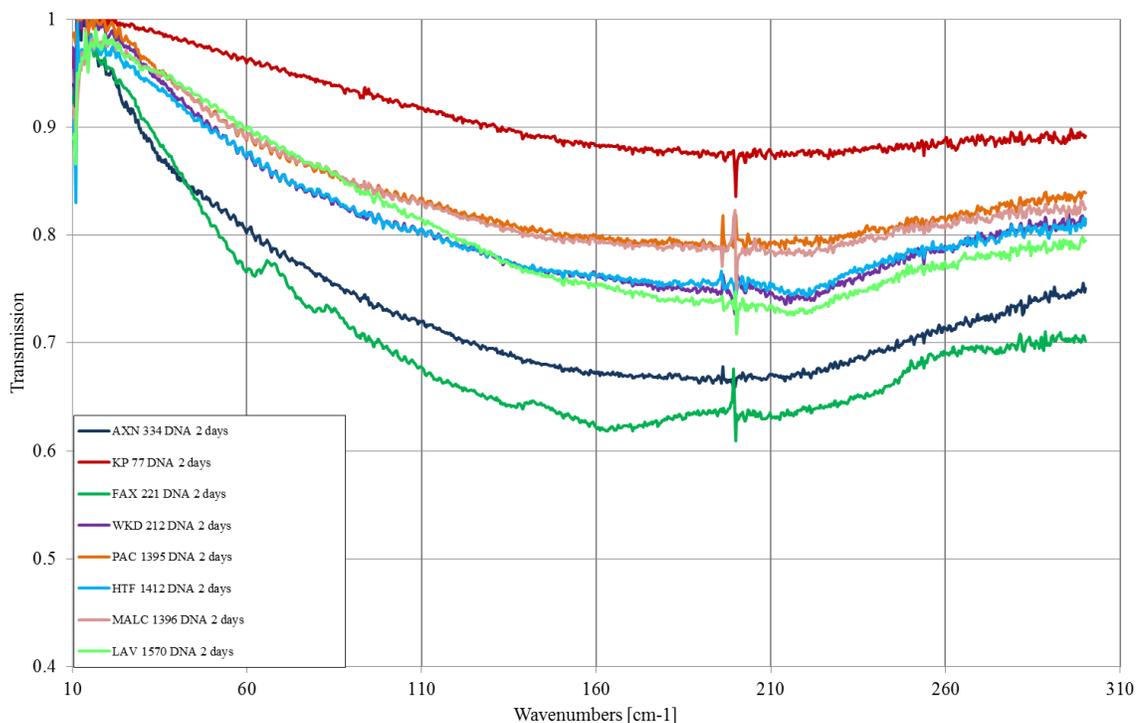


Figure 5-57: THz Spectra of mouse DNA from different colonies from 10 to 310 cm^{-1} . AXN, KP, FAX, WKD, PAC, HTTF, MALC and LAV, were dried and left at room temperature for 2 days.

Most of the samples in Figure 5-57 seem to possess the same broad absorption feature between 180 cm^{-1} and 210 cm^{-1} . The sample labelled “FAX 221 DNA 2 days” has smaller features at 62 cm^{-1} , 80 cm^{-1} and 137 cm^{-1} . Interestingly, 4 samples: “LAV 1570 2 days”, “WKD 212 2 days”, “HTF1412 2 days” and “MALC 1396 2 days”, have a small feature at 217 cm^{-1} . This suggests that over a time of 2 days, the samples have changed resulting in a difference and/or addition in an absorption feature in the THz spectra. All samples were made and kept in the same lab under the same conditions. However, these are DNA samples from different colonies of mice, which makes it difficult to predict what these features represent.

To further extend this research into the time effect of dehydrated DNA, the spectra of the DNA samples from the same colony that was heated and presented in section (5.3.4) was also repeated after a time of 1, 5, 3 and 7 days and presented in Figure 5-58 and Figure 5-59.

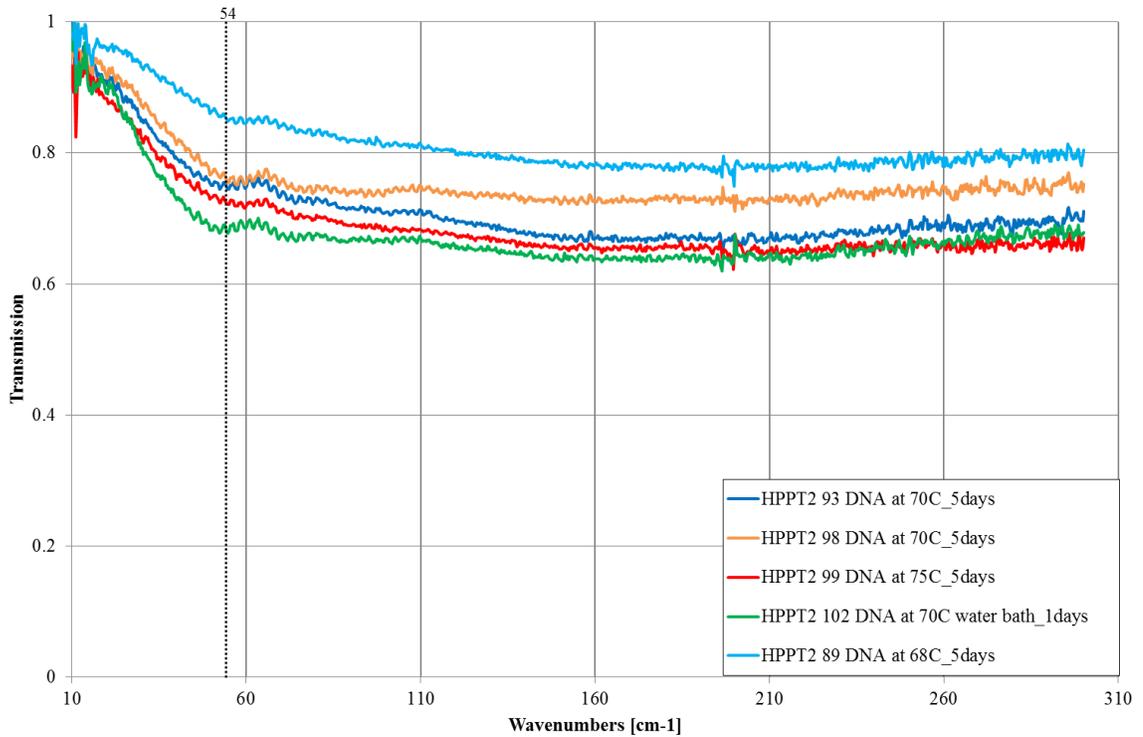


Figure 5-58: THz spectra of mice DNA from the same colony (HPPT2) measured between 10 and 310 cm^{-1} after heating and leaving at room temperature for a time of 5 days, except for 'HPPT2 102' sample which was only 1 day.

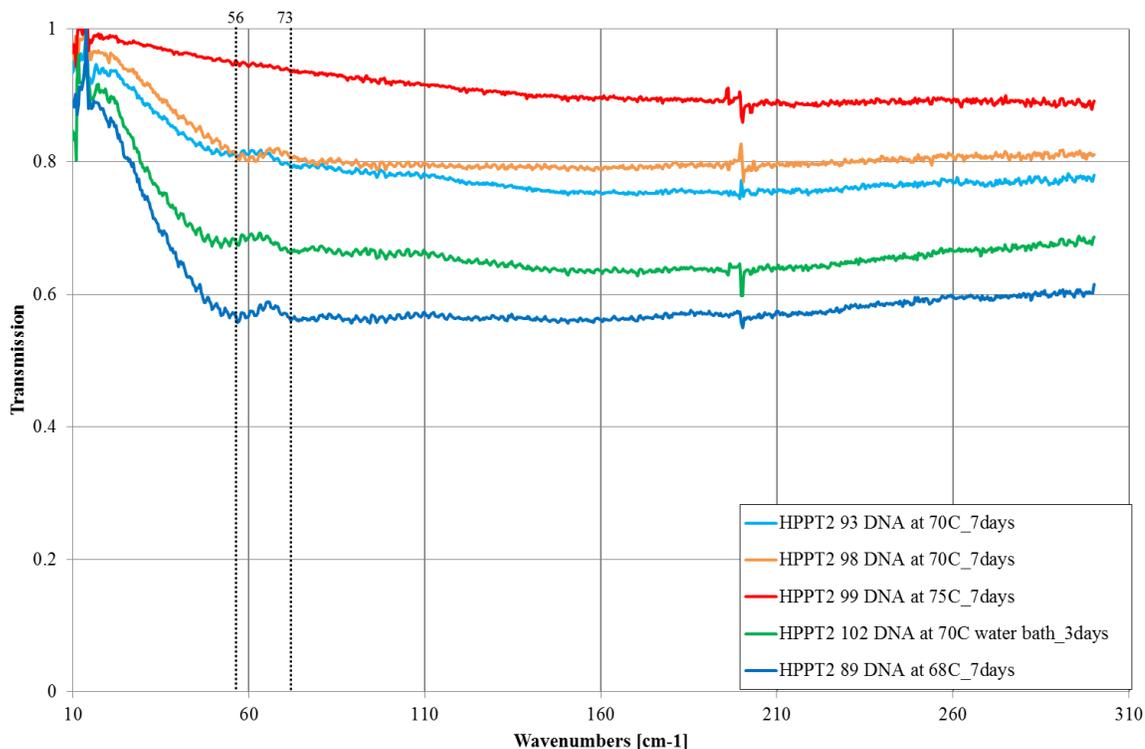


Figure 5-59: THz spectra of mice DNA from 10 to 310 cm^{-1} . The samples were from the same colony (HPPT2) and measured after heating and leaving at room temperature for a time of 7 days, except for the ‘HPPT2 102’ sample which was only 3 days

Clearly there is a difference between the initial set of data at time = 0 days shown in Figure 5-56 when the first set of data was taken, and the data shown in Figure 5-58 and Figure 5-59. The main broad feature present in all DNA samples between 180 cm^{-1} and 210 cm^{-1} has disappeared over time. There are many uncontrollable factors involved in these measurements such as environmental factors, variation of concentration across the different samples, alignment differences, etc. But the lack of the broad absorption feature means that there must be some changes in the conformation of the DNA over time. For

clarification, Figure 5-60 illustrated the effects of time on the THz spectra of the same DNA sample.

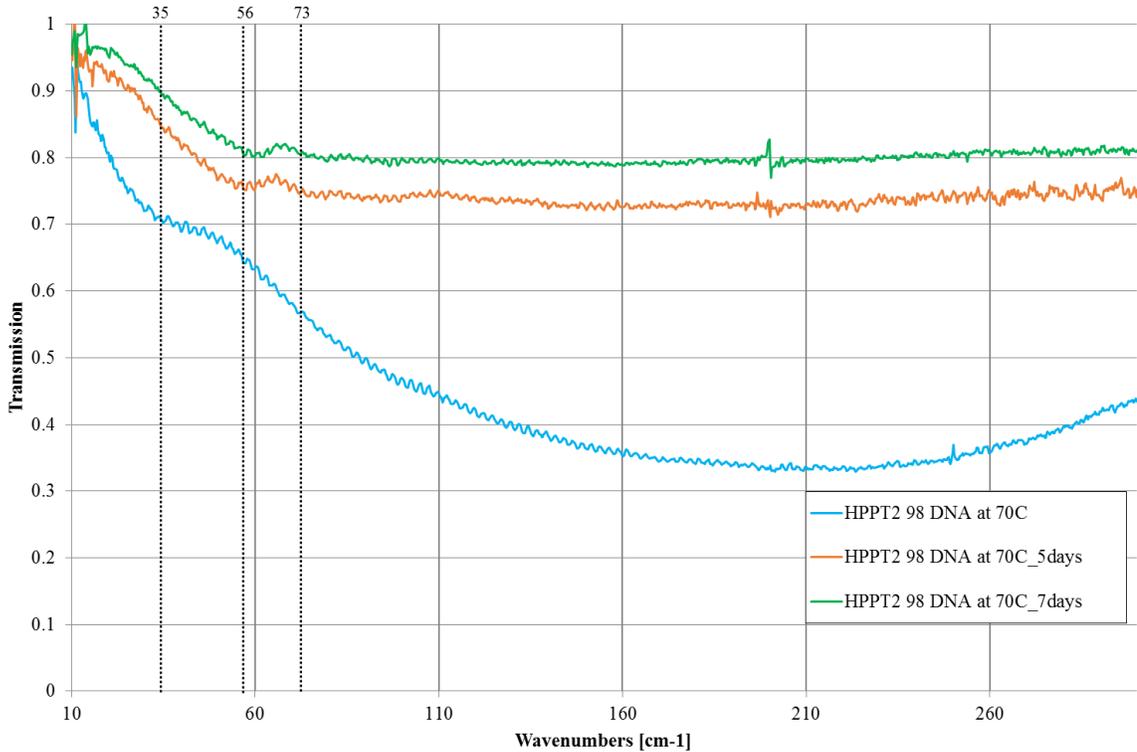


Figure 5-60: THz spectra of 'HPPT2 98' mouse DNA highlighting the difference in the absorption features over a period of 5 and 7 days, measured from 10 to 300 cm⁻¹

As well as the broad absorption feature at 210 cm⁻¹ there is an addition of a smaller feature seen in all the samples. This could represent a change in the DNA structure, potentially a signature of single stranded DNA, after heating. This feature at 35 cm⁻¹ in the Figure 5-56 has shifted to higher frequencies, 54 cm⁻¹ after 5 days and 56 cm⁻¹ after 7 days, within error.

5.3.6 Conclusion

This investigation in the structure of dried DNA and RNA has posed a lot of questions. The comparison of the THz spectra of human DNA and RNA in section (1.2.1) has identified an absorption feature in the DNA at 158 cm^{-1} which is not in the spectra for RNA. It suggests that this broad absorption feature is a superposition of many resonances due to the hydrogen bonding between the nitrogenous bases at the centre of the double helix.

Although mice and humans share similarities in their genome, it appeared according to section 5.3.2 that the THz spectra of human DNA were different to mouse DNA. Both human and mice DNA have a broad absorption feature but the absorption feature identified in the human DNA was at 158 cm^{-1} whereas the absorption feature for the mice DNA was centred at 210 cm^{-1} . However, without extraction of identical lengths of DNA strands it is difficult to be conclusive.

A collection of mice DNA was investigated to identify whether there were any spectral differences in the THz region between the DNA of mice from different colonies. In section 5.3.2, Figure 5-55 presented the spectra for 8 different DNA samples which had a consistent broad absorption feature centred around 210 cm^{-1} .

Heating DNA is often used in polymerase chain reaction (PCR) technique which is a laboratory technique used to artificially replicate DNA molecules. Temperatures between $65\text{-}90^\circ\text{C}$ are used depending on the specific parameters of the DNA. As the providers of the samples did not know the specific parameters of the DNA, a range of temperatures were tried. Section 0 presents the results of heating the DNA samples to $70 \pm 5^\circ\text{C}$. This

resulted in an additional feature in some of the samples at 35 cm^{-1} which could be the result of strand separation or conformational changes in the DNA. However, this is not conclusive as other factors such as the DNA length and nucleic base sequencing also have a big influence.

The investigation into the effect of time over a period of a couple of days has proved that the features for heated samples have changed over time. The broad feature at 210 cm^{-1} is not present in the samples after a period of 5 days and again after 7 days. However the feature that is found at 35 cm^{-1} in Figure 5-57 is present but at 54 cm^{-1} after 5 days and 57 cm^{-1} after 7 days. It is unclear what is changing over time as there are many uncontrollable factors such as the length of the DNA strand, the complimentary base sequence and the conformation of the DNA strands after dehydration, but further investigations with different control elements would provide more information.

5.4 THz spectra of liquid samples

Using molecular dynamic simulations (Errington & Debenedetti, 2001) have been able to identify the relationship between translational and orientational order in liquid water through hydrogen bonds. They were able to identify the orientation of individual molecules as well as their relative separation. However, combining the contributions from individual molecules for a better understanding of the translational and orientational order of water in bulk still proves to be extremely complicated.

THz frequencies are able to excite collective motions of solvents such as water due to the energies involved with hydrogen bonding between the H₂O molecules. This is a very powerful tool but it is also a detriment for research in THz spectroscopy of biological molecules. Real life replication of biological molecules such as nucleobases, nucleosides, and DNA are all naturally found within a solution, and the human body is composed on average of 60% water (Nestle-Waters, 2017).

Simulations have been performed to model the effect of water on biological molecules such as DNA and proteins by using THz frequencies. (Heyden, et al., 2010) investigated the THz spectrum of liquid water from first principles using simulations in time and space. This led to the conclusion that, despite the advantages of mid-infrared spectroscopy in the identification of the intramolecular motions in the finger print region, the THz region provides very broad features due to the very complex intermolecular motions that are happening on a picosecond timescale. Despite its issues with respect to the absorption of THz radiation by water, THz spectroscopy has led to the discoveries of solvation shells surrounding biological molecules within solution and how the solvents

can potentially govern the molecular conformations and structures of some biological molecules (Steinke, Gillams, Pardo, Lorenz, & McLain, 2016).

However, there are still many unknowns about the dynamics of biological molecules within solvents. THz spectroscopy has shown promising results on different molecules, but the majority of this work has been based on simulations and time-domain spectroscopy which is used to probe the refractive index and absorption coefficient of the medium. In this section, the Fourier Transform Spectrometer (FTS) (see CHAPTER 3 for more details) is used to investigate the THz spectra of water, heavy water and nucleobases in water.

5.4.1 Method

Here, a method was found to purge the sample chamber of the FTS of atmospheric water vapour and to allow solutions to be placed in it.

5.4.1.1 Chamber purge

The sample chamber and the beam condensing chamber were vented and a thin green pipe was placed at the bottom, being careful not to obscure any of the optics. This was connected to a CO₂ canister which was held in place on a trolley and can be seen in the images in Figure 5-61.



Figure 5-61: Pictures showing the CO₂ canister next to the laboratory FTS. The green pipe feeds the gas into the sample and optics chamber from the canister. The red cryostat houses the cryogenic detector.

The chambers were left to purge before two sets of data were taken over a 12 minute period and were compared to test the stability of the gas in Figure 5-62.

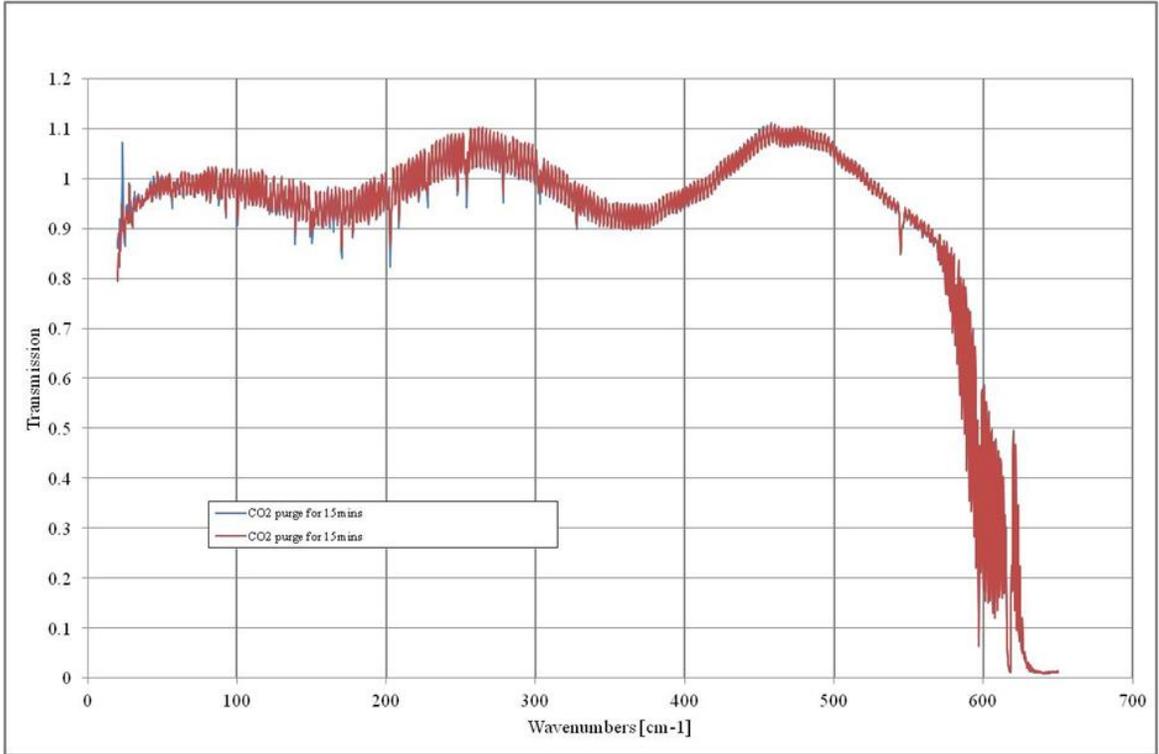


Figure 5-62: THz spectra of the sample chamber of the FTS after purging with CO₂ for 15 minutes. Two spectra are shown, taken 12 minutes apart and indicate that the background spectrum is relatively stable.

Figure 5-62 shows good agreement between the two spectra, with both having high transmission. The transmission above 1 is because the optics within the FTS are originally aligned within vacuum conditions. At atmospheric pressure, or when the chambers are being purged with other gases, the optics within the FTS will be slightly misaligned to the original. This shows the sensitivity of the FTS to very small changes and the accuracy to which it has been designed under vacuum.

When the CO₂ pipe was removed and the lid was placed on top to close the chambers, the CO₂ amount was stable. After being left for 30 minutes, the spectra looked the same.

Conventionally a frequent choice for purging is Nitrogen (N_2). When this procedure was repeated for N_2 gas over a period of 1 minute, the atmospheric absorption lines were reappearing as the nitrogen gas leaked. CO_2 is heavier than N_2 such that it fills the chambers and remains within the chamber better than the N_2 .

5.4.1.2 Sample holders

Designing sample holders for liquid samples proved to be difficult as the windows of the holder needed to be transparent, rigid and have a cumulatively low optical path length. The sample holder was made from two layers of $2\mu\text{m}$ of polypropylene (PP) adhered to two identically sized metal rings, with the liquid in between. However, initial results showed that there were little or no transmission through the samples therefore there was a need for making them thinner. Figure 5-63 a) shows the first sample holder with two metal rings of the same size which was too thick for any signal to be transmitted, b) shows the modified sample holder where one smaller disk of polypropylene was placed on top of a larger area disk and held under a monochromatic light to show the liquid sample within the centre. By placing one smaller disk of polypropylene onto a larger area polypropylene disk with a small droplet of the liquid, it proved to be more efficient at distributing and containing the liquid without it moving too much. Also, the smaller PP disk has a thin layer of glue around the outer edges of the rings to help close the two disks together. Despite this improvement, calculating the thickness of the liquid sample was very difficult.



a)



b)

Figure 5-63: Pictures of a) the initial sample holder with the liquid inside, b) the samples being looked at under a monochromatic light source

5.4.1.3 Consistency and reliability

Part of the difficulty with this experiment was consistency and repeatability. There is a high chance of contamination from oil and dust when making the sample and handling the holder. These examples can affect the spectra by potentially having a signature absorption due to contamination and scattering from relatively large particles of dust. They could also affect the interface between the surface of the metal ring and PP.

5.4.2 Results

Samples of pure DI water, cytosine in DI water, guanine in DI water and thymine in DI water were tested. The bases were mixed with the DI water and were made to the same ratio as has been used for making dried samples. Being hydrophobic, the bases do not dissolve well in water leaving optically opaque samples. It also meant that some droplets contained a higher concentration of solute than others.

5.4.2.1 THz spectroscopy of liquid H₂O

Figure 5-64 identifies the THz spectra for the liquid samples and Figure 5-65 compares the spectra of the same liquid samples and the sample holder without any liquids. Each sample was produced using a unique sample holder to prevent cross-contamination of the bases.

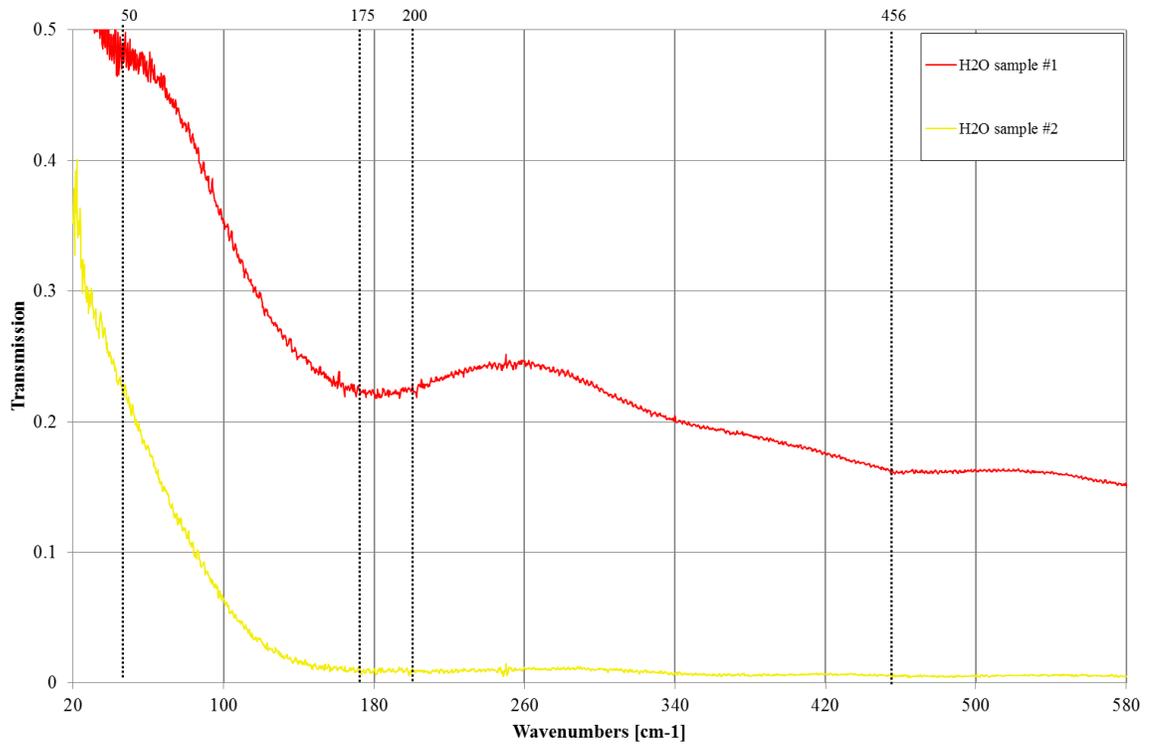


Figure 5-64: THz spectra of H₂O contained between two layers of 2µm PP ratioed against the CO₂ background

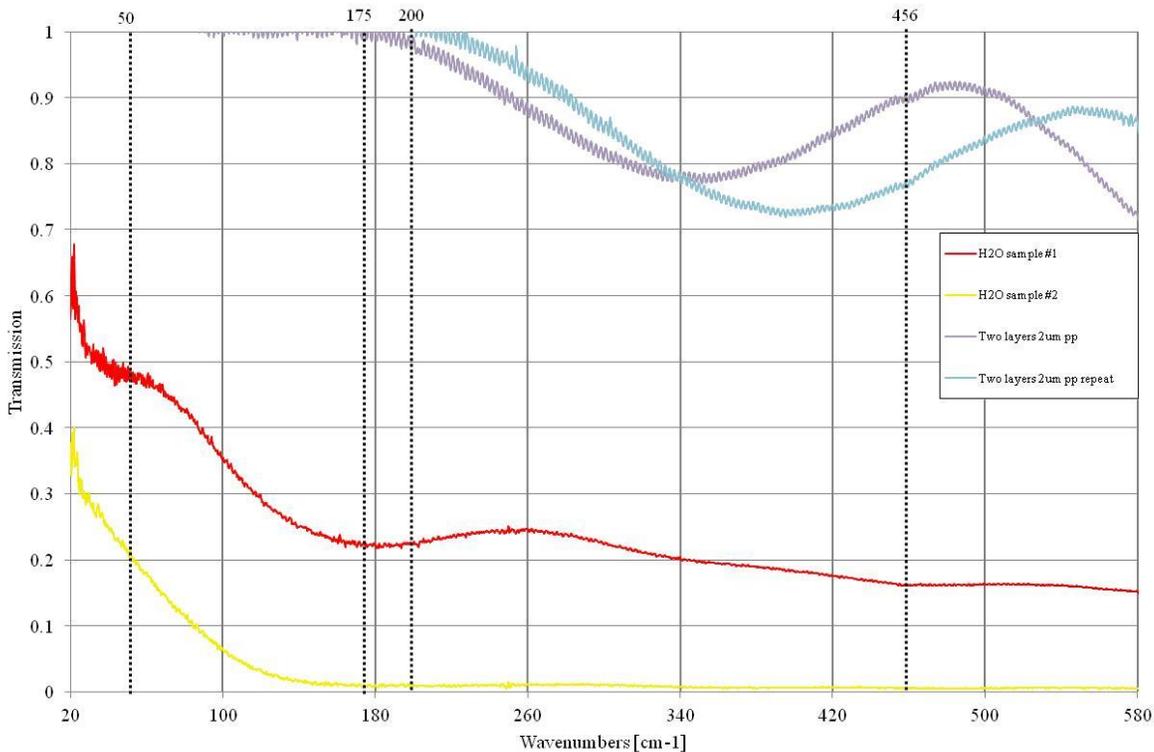


Figure 5-65: THz spectra of H₂O contained between two layers of 2µm PP ratioed against the CO₂ background and the spectra of two layers of 2µm PP with no sample (air gap)

Both Figure 5-64 and Figure 5-65 show a common broad absorption feature at around 175 to 200 cm⁻¹ and a low transmission towards 550 cm⁻¹. The thickness of the two PP sheets and the air gap in between can be determined through the Fabry P erot effect observed in the spectra for the “Two layers of PP” in Figure 5-65. These are calculated to be $8.0 \pm 1.0\mu\text{m}$. Clearly there is a difference in the spectra, hence a difference in the optical thickness of the empty PP sample containers which reinstates the difficulty in controlling the thickness of the samples and their holders.

5.4.2.1.1 Analysis from 20 to 100 cm⁻¹ (0.6 to 3.0 THz)

Identified in Figure 5-64 and Figure 5-65 is a potential absorption feature at 50 cm⁻¹. Martin Chaplin (Chaplin, 2016) has accumulated the work of researchers over the past decades, trying to better understand H₂O, one of the simplest compounds in the universe and the most essential for life on Earth. Table 5.13 shows the identification of low frequency vibrational modes given by (Chaplin, 2016).

Table 5.13: Assignment of absorption features found using infrared and Raman spectroscopy

Wavenumber (cm⁻¹)	Assignment
50	Hydrogen bond bend
183.4	Hydrogen bond stretch
395.5	L1, libration
686.3	L2, libration

Table 5.13 identifies the 50 cm⁻¹ absorption feature to be hydrogen bond bending which is plausible as it is present in all the samples in Figure 5-64 and Figure 5-65.

(Woods & Wiedemann, 2004) have performed far infrared spectroscopic work on liquid water using the Stanford University Short Intense Electron Source (SUNSHINE) and found prominent absorption bands at 40, 55, 65, 80, and 90 cm⁻¹ at 269 K and were found to be less prominent as temperatures increased, shown in Figure 5-66.

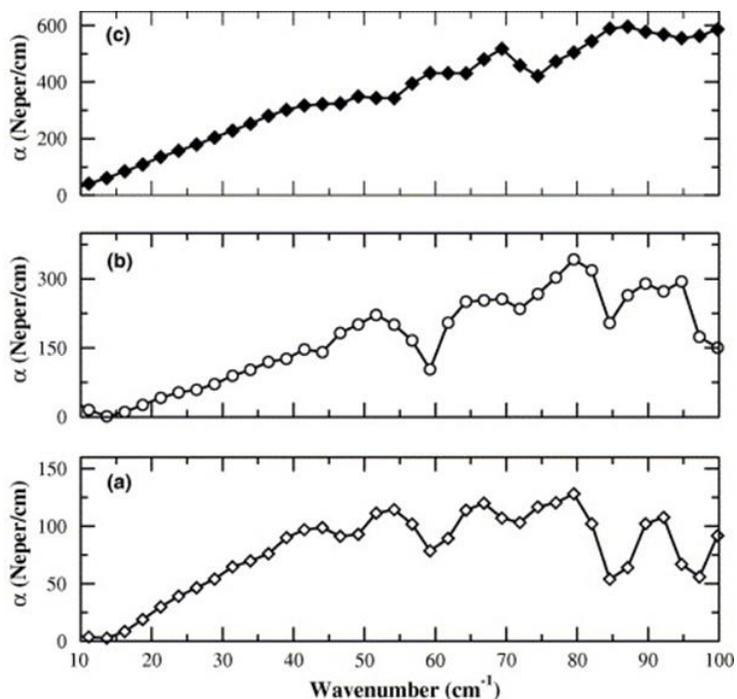


Figure 5-66: Experimental absorption spectrum of liquid water from 269 to 298 K: (a) 269 K; (b) 277 K; (c) 298 K taken from (Woods & Wiedemann, 2004)

Clearly the features in Figure 5-66(C) have broadened with the loss of a few prominent features. The spectra in Figure 5-64 and Figure 5-65 were taken at room temperatures (25 °C, 298 K), but the absorption features might sharpen if taken at lower temperatures as done by (Woods & Wiedemann, 2004).

Other techniques such as far infrared vibrational rotational tunnelling (VRT) spectroscopy measurements have identified low frequency torsional motions involving the flipping of free hydrogen atoms in isolated water clusters (Saykally, 2001). These motions are observed in the $<100 \text{ cm}^{-1}$ region of the far infrared spectrum but it is yet unclear whether these types of intermolecular vibrations might also have a significant contribution in bulk water.

5.4.2.1.2 Analysis from 100 to 600 cm⁻¹ (3.0 to 18.0 THz)

In liquid water, H₂O molecules are attracted to each other which mean that they have a preferred orientation and cannot freely rotate. Therefore the molecule undergoes libration movements where the molecule can “sway” backwards and forwards. These are seen in infrared spectroscopy but can be better identified using THz spectroscopy as they have a lower frequency than other vibrational movements.

With reference to Table 5.13, L1 is a minor band and L2 is a major band. According to (Walrafen , 1990), the intermolecular vibrational modes belonging to frequencies above 300 cm⁻¹ stem from librational (i.e., hindered rotational) motion due to hydrogen bonds. Cluster vibrations involving combinations of OH---O hydrogen bond stretching and bending are responsible for the translational absorption band at 200 cm⁻¹ (Gaiduk & Vij, 2001).

Other work has been undertaken on the THz spectroscopy of liquid water from first principles using correlations in time and space using ab initio molecular dynamics simulations of the infrared spectrum of water (Heyden M, 2010). The 200 cm⁻¹ absorption feature is a reoccurring theme, with (Heyden M, 2010) demonstrating that it is dominated by particle motion within the first solvation shell. Through further analysis they confirm that this band is dominated by stretching motions of hydrogen bonds.

The librations depend on the moment of inertia of the molecule which can change in the case of replacing hydrogen with deuterium causing the frequencies to shift down (Chaplin, 2016). According to Chaplin, “the librations depend on the moments of inertia such that the almost doubling of the moments of inertia of D₂O, relative to H₂O, reduces

the frequency by about a factor of $\sqrt{2}$.” However, (Chaplin, 2016) does continue with explaining that these differences between the spectra of H₂O and D₂O might be very small as these involve network movements along linear or near-linear hydrogen bonds.

5.4.2.2 Comparison of the THz spectra of liquid H₂O and D₂O

To investigate the 200 cm⁻¹ feature found in the liquid samples and whether any differences can be found between H₂O and D₂O, liquid samples of D₂O were made in the same way as outlined in section 5.4.2.1 and probed in the far infrared as shown in Figure 5-67.

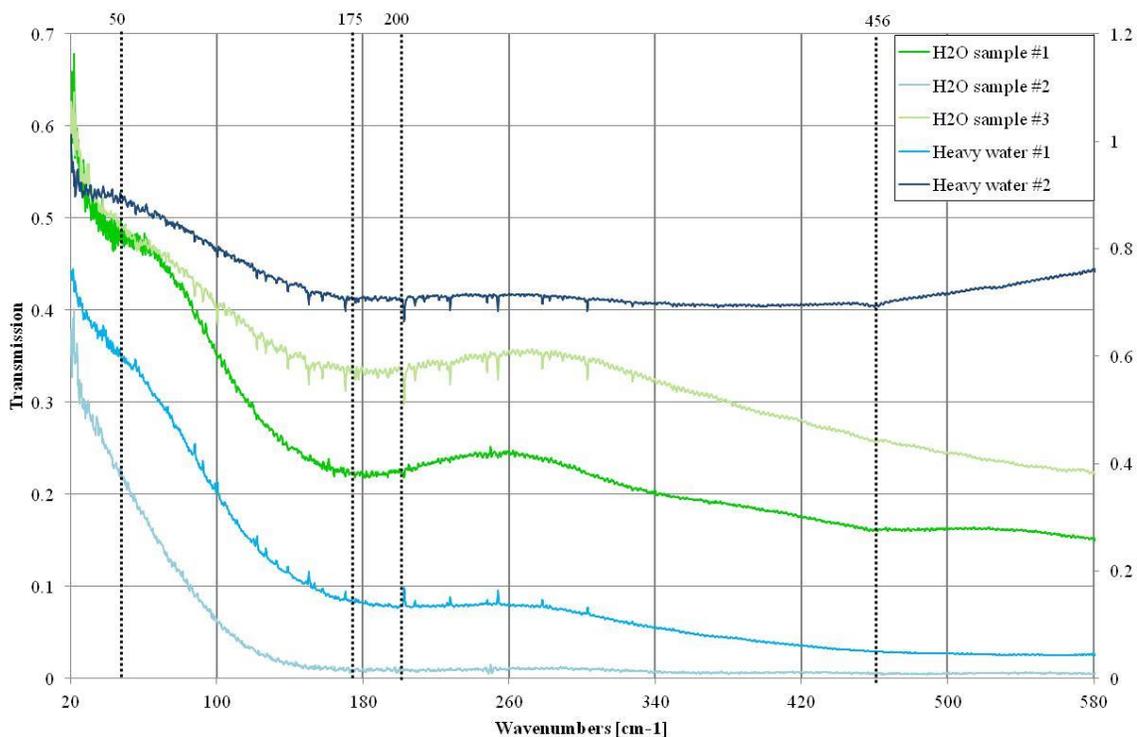


Figure 5-67: THz spectra of liquid H₂O and D₂O

The spectrum for the “heavy water #2” was placed on a secondary axis for a better comparison of all five spectra. The small noise features found in all the spectra are due to issues when ratioing against the CO₂ background. The amount of CO₂ in the sample chamber varied slightly throughout the experiment because of disruptions when swapping samples. Background data were taken often throughout the experiment but there is clearly some difficulty in consistency.

For all five spectra seen in Figure 5-67, there is clearly a feature centred at 180 cm⁻¹ which better supports the assumption that this is translational vibrations involving OH---O hydrogen bond stretching and bending. The hydrogen bonding and structure of H₂O and D₂O have been investigated through x-ray diffraction, neutron diffraction and molecular modelling from different publications with all suggesting slightly different predictions for the bond strength and length. According to (Benmore, 2008), D₂O is more structured than H₂O but the hydrogen bond length in H₂O is ~4% shorter than D₂O but are less linear. This suggests that the weakening in the length of the bond is compensated by the strength of the straightening and generally, the hydrogen bonding in both is the same.

5.4.2.3 THz spectroscopy of guanine solution using H₂O

As shown in other research and in the spectra presented in Figure 5-64 and Figure 5-67, water is highly absorptive. There has been some attempt at understanding the dynamics of large biological molecules in water, especially the solvation effects (Keutsch & Saykally, 2001), (Nagai, Yada, Arikawa, & Tanaka, 2006) (Kim & Born, 2008) (Choe,

Lademann, & Darvin, 2016). In this work the THz spectroscopy of guanine in DI water was studied.

The method was the same as in section (5.4.1 method). Figure 5-68 is the THz spectra of guanine in a solution using DI water in the sample holder described in section (5.4.1.2sample holder) and the spectra of guanine made in to a solution using DI water and left to evaporate on 2 μ m PP.

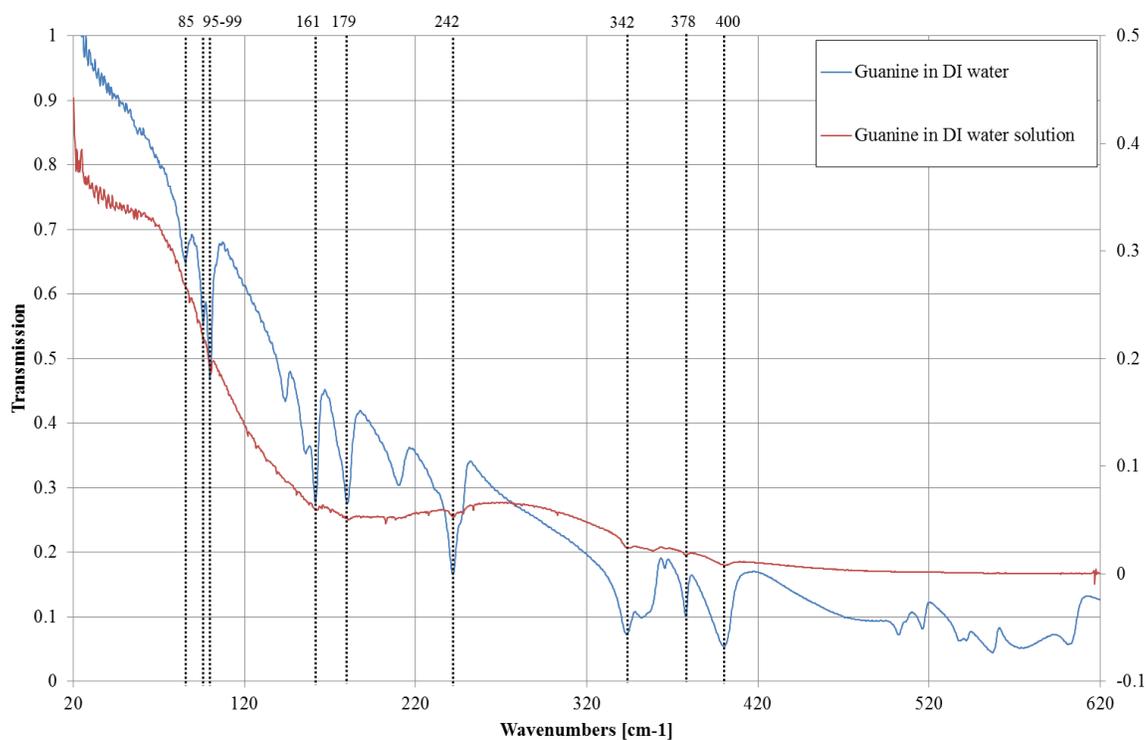


Figure 5-68: THz spectra of guanine in DI water solution compared to dried guanine made using DI water

In Figure 5-68, there are multiple absorption features which are common to both samples. Absorption features are present at 85, 95, 99, 161, 179, 242, 342, 378 and 400 cm^{-1} for both spectra but are very much weaker in the spectra of guanine in water solution. The

spectrum for the liquid solution sample also possess the translational absorption feature at 200 cm^{-1} due to the hydrogen bond stretching and bending. It also has a feature at 50 cm^{-1} which has been identified as hydrogen bond bending.

Some attempt at understanding the mechanisms of hydrogen bonds in water have been made by (Chandra, Nguyen, Uchimar, & Zeegers-Huyskens, 2000) suggesting that guanine molecules occur as a complex mixture of unusual tautomeric forms in aqueous solutions, depending on hydration degree. This is an important factor to consider because if the structure of the hydrogen bonded guanine molecules change, then the THz spectrum could potentially change.

As shown in section 5.1, each nucleobase and nucleoside has a unique THz spectrum. The results presented in Figure 5-68 might suggest that the characteristic features found for nucleobases can be identified in its solvated form. However, it is not clear from Figure 5-68 whether this is due to the inability of guanine to dissolve in water, resulting in the aggregation of undissolved guanine molecules which in turn could be forming a crystalline structure similar to the solid state.

5.4.3 Conclusion

To better understand the interaction of water in the liquid state with THz frequencies and the interaction of biological molecules in a liquid solution, THz spectroscopy was undertaken using the FTS. CO_2 proved to be the most efficient gas for purging the sample chamber as it was heavier than N_2 and has a consistent transmission from 20 to 600 cm^{-1} .

To be able to probe liquid samples in the THz region, it was necessary to create sample holders with high transmission in the 20 to 600 cm^{-1} region and have a short path length. The most efficient sample holder was shown in section 5.4.1.2 whereby two 2 μm PP substrates of different diameters were used as windows and were pressed to hold the liquid sample between the sheets by using the surface tension of water and a small amount of adhesive glue around the outer region of the metal rings.

The THz spectra of water shown in Figure 5-64 and Figure 5-65 resulted in a broad absorption feature between 175 and 200 cm^{-1} . There also were small features at 50 cm^{-1} and 456 cm^{-1} . These features were identified with comparison to previous literature work and confirmed that the 200 cm^{-1} was a translational absorption feature due to hydrogen bond stretching and bending. The smaller 50 cm^{-1} was attributed to hydrogen bond bending and higher frequency absorption features were attributed to librations, or “swaying” of the water molecules.

It was said that the librations depend on the moment of inertia of the molecule which can change in the case of replacing hydrogen with deuterium causing the frequencies to shift down (Chaplin, 2016). THz spectra of D_2O was presented in Figure 5-67 and compared to the THz spectra of H_2O . Both samples possessed the broad absorption feature at 175 to 200 cm^{-1} , the small feature at 50 cm^{-1} and at higher frequencies around 456 cm^{-1} . This better supports the assumption that the broad absorption feature at 175 to 200 cm^{-1} is translational vibrations involving OH---O hydrogen bond stretching and bending and that the replacement of a hydrogen with a deuterium does not change the THz spectra of liquid water.

As biological molecules within the body are naturally within a solvent and that the human body is on average 60% H₂O, it was necessary to investigate whether it was possible to identify any THz signature of biological molecules within a solvent. The THz spectra of the guanine in water solution does possess the broad absorption features significant to H₂O at 50 cm⁻¹ and 175 – 200 cm⁻¹, but it also has weak fine lines which correspond to the absorption feature found in the dried guanine. This provides some evidence that THz spectroscopy could be able to identify the characteristic absorption feature of nucleobases when in a liquid solution, providing that the concentration of the powder is high enough.

5.5 THz data of tissue

The development of THz technology has allowed for easier and better spectrometers to be used and applied to different samples. Many of these tend to be THz time-domain spectrometers (TDS) such as the range of spectrometers and probes available from TeraView (TeraView, 2016). Since becoming more widely available, THz spectroscopy of biological samples such as different types of tissues have gained a lot of interest, especially with regards to cancer detection and early diagnosis. Animal tissue is often used such as pork and rat tissue, as used by (He, Azad, Ye, & Zhang, 2006), to better understand the difference in the THz spectra of different organisms and between fat, muscle and skin. The absorption coefficient varies for different samples depending on their water content. This technique takes advantage of the high absorption of water which is the basis of how THz spectroscopy has been used to identify cancer margins in samples of tissue (Ashworth, et al., 2009).

As shown in section 5.1, 5.3 and 5.4, terahertz spectroscopy of biological components such as nucleobases, nucleosides and DNA have very interesting THz signatures which provide information about the conformation of the molecules. These are the building blocks of larger structures such as cells and tissues. Therefore, there is a need to better understand the difference between tissue types. Although THz TDS has provided an insight into the application of THz spectroscopy on indentifying between animal fat, muscle and skin, and cancer margins within human tissue samples, there is still a lack of understanding of the structural difference and whether there is any way of characterizing

different tissue types. In this section, different tissue types from human and animal has been investigated using the Fourier Transform Spectrometer (FTS) outlined in CHAPTER 3. This section will investigate the different methods of preparing tissue samples and the resulting spectra.

5.5.1 Method

Various methods of fresh tissue extraction and stabilisation are employed, many of which may hinder the use of THz spectroscopy. These are identified and explored in section 5.5.1.1, 5.5.1.2, and 5.5.1.3.

5.5.1.1 Formalin Fixed Paraffin Embedded (FFPE)

To preserve fresh tissue and use it to identify cancer margins through histology, there is a need to fix and embed the tissue. Sections of patient's tissue are brought to the histology laboratory for examination. The samples will contain both healthy and cancerous tissue as the precise margin is not clear. A core needle biopsy⁶ is performed by taking small cylinders (or cores) from the cancerous tissue. These were placed on a sheet of foil while the rest of the tissue is placed in 10% neutral buffered formalin for 48 hours to be fixed. Formalin is used to cross-link the nitrogen and some other atoms in the proteins. A methylene bridge $-(CH_2)-$ is formed between adjacent proteins or proteins that are in very close proximity. Other molecules such as carbohydrates, lipids and nucleic acids, are said to be trapped in a matrix of cross-linked proteins.⁷

⁶ <http://www.cancer.org/treatment/understandingyourdiagnosis/examsandtestdescriptions/forwomenfacingabreastbiopsy/breast-biopsy-biopsy-types>

⁷ <http://synapses.clm.utexas.edu/lab/howto/cross-linking%20fixatives.pdf>

After 48 hours of being in formalin, the tissue is much firmer and can be cut into smaller sections. These sections are then treated separately and were placed into a “thermos scientific” machine which dehydrates them using alcohol and xylene. The aim is to remove all the water from within the cells so that it can be replaced by wax to preserve the shape of the sample. This means that the tissue is now fixed and is unable to degrade or “rot”.

Once the samples has been fully dehydrated, a small metal tray is filled with molten wax which is liquid and transparent at 60°C. The tissue is placed into the wax and a cassette which contains holes, is placed on top as a lid. The container is then filled with more molten wax until it is full. The cassette for each sample has a specific bar code which labels the tissue. The tissue sample in the molten wax is then left at room temperature to solidify where it is then ready to be sectioned.

For histological purpose, the wax usually needs to be cool and harden so that the wax can be cut smoothly. But for thicker samples i.e. thicker than 10µm, the wax needs to be slightly warm and soft to avoid the wax becoming too brittle when being cut. The sectioning is performed using a microtome and by moving a blade across the samples to slice them according to the thickness specified on the machine. The microtome used to slice the tissue in wax was successful in slicing samples of thickness 20µm, 30µm, 40µm and 50µm. The tissue might vary in density and may contain holes which mean that the thickness is not consistent throughout the sample.

The sliced samples are then placed into a water bath at 50°C. The warm water is essential to soften the wax to ensure that it adheres to the substrate. The substrate used was 2.8µm

polypropylene (PP) and the samples were human breast and skin tissue. For histological purposes, the breast tissue was stained blue as it signifies the orientation and position of this specific tissue sample with respect to the whole breast.

5.5.1.2 Lyophilizing

To investigate the potential of THz spectroscopy in the identification of tissues, there is a need for fresh tissue samples at an optimal thickness. However, fresh tissue will contain a large amount of water which is problematic when the sample chamber is under vacuum, and as shown in section 5.4, water is highly absorptive in the THz frequencies.

Lyophilising is a preserving technique using snap freezing of a sample of tissue and then placing it in a high vacuum, low temperature environment to remove the water. According to (Png, et al., 2008), this is the most effective way of taking THz measurements of tissue. To investigate further, chicken breast were bought from a local supermarket, coarsely sectioned and these slices were placed under pressure to render them into thin planar samples. Figure 5-69 is a diagram of the set up used to compress the chicken breast.

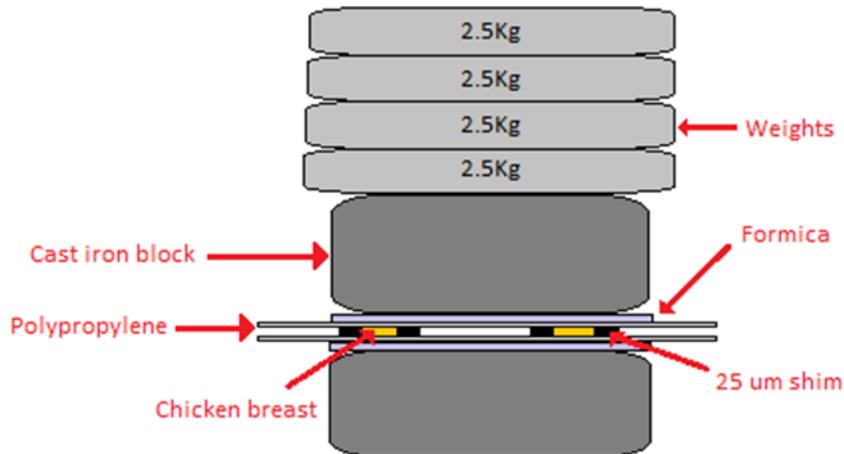


Figure 5-69: Diagram of the experimental procedure for flattening the chicken breast

Sections of chicken breast were placed between two polypropylene substrates with shims of known thickness, using a total of 10Kg weights and left in a vacuum oven for 60 hours. The samples were then sufficiently flat and dry for investigating using the FTS. Shims of 25 μm thick were used but when the tissue was removed from the vacuum oven, it was difficult to evaluate the exact thickness. Once the data for the samples were collected, the thickness of the samples was measured using a micrometer, which showed that “Chicken breast 1” and “Chicken breast 3” were $90.0 \pm 10.0 \mu\text{m}$ and the other two (2 and 4) were $70.0 \pm 10.0 \mu\text{m}$.

5.5.1.3 Frozen section procedure

5.5.1.3.1 University Wales Hospital microtome and procedure

It was not possible to get human tissue so chicken breast was bought from a supermarket and chicken liver was kindly donated by a butchers in Cardiff market. The liver and breast was cut into small 1 cm^2 cubes, placed onto small pieces of foil before being

placed in liquid isobutene to be snap frozen. They were then placed into a polystyrene box filled with dry ice to keep cold. The process can be seen in Figure 5-70.



Figure 5-70: Chicken liver and breast being prepared (left) and the chicken liver on a piece of foil being placed in to the isobutene for the snap-freezing (right).

The samples were then placed in labelled plastic tubes and are left in the freezer at $-80 \pm 3^{\circ}\text{C}$ until required. The samples in the freezer could be thawed to temperatures of $-22 \pm 2^{\circ}\text{C}$ in the cryosection machine, HYRAX C20⁸, as shown in Figure 5-71.

⁸ Similar to the machine used at the Heath hospital
<http://www.bu.edu/becf/downloads/BioInterface%20Technologies/HM%20525%20Manual.pdf>



Figure 5-71: Picture of the cryostat microtome used for the cryosectioning of tissues

The tissue adhered to a metal disk using OCT embedding matrix glue. When placed into its position in the cryostat, the glue, which is transparent and liquid at room temperature, turns opaque and freezes rapidly which ensures minimal damage from crystal formation and reinforces the tissue. This material can be easily removed once it returns to room temperature. The metal holders and embedding material are shown in Figure 5-72.



Figure 5-72: The flat metal plates used to place the samples (left) and the embedding material that is used to hold the sample rigid and on the metal plate

Once the sample and glue have solidified, they can be positioned ready to be cut as shown in Figure 5-73.

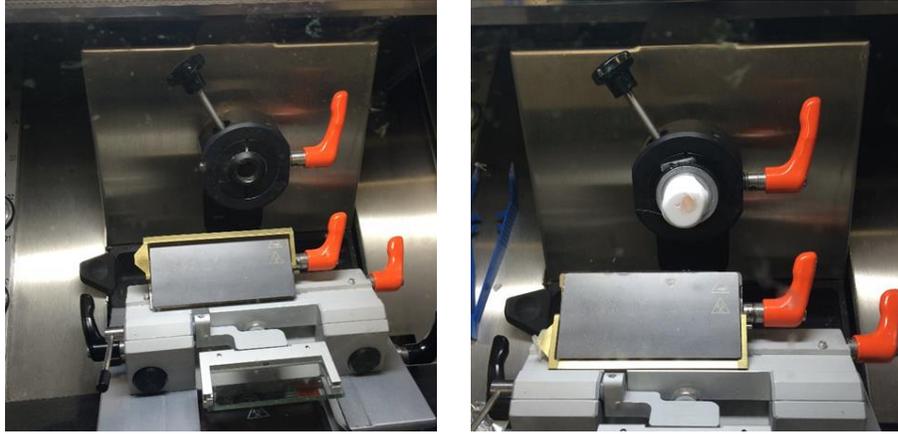


Figure 5-73: The sample holder (left) and the sample holder with the tissue held in place in the opaque embedding material

For histology purposes, once sliced, a warm (room temperature) glass slide is then placed on top of the sectioned sample and immediately sticks to the slide with the glue turning transparent again. The substrate used for THz spectroscopy purposes was PP. This was difficult as PP is a poor conductor which caused some difficulty in getting the substrate sufficiently warm to thaw the glue and allow it to stick. The maximum thickness possible using the cryostat was 22 μm with a range of thicknesses including 5, 10, 20 and 22 μm .



Figure 5-74: Screen of the cryostat microtome showing the temperature within (-22) and the range of section thickness that can be made

5.5.1.3.2 Cardiff University Biosciences laboratory microtome and procedure

The maximum thickness available from the WCB was 22 μm as noted, but there was a similar cryostat in the bioscience department at Cardiff University which was able to cut thicker samples of tissue. Figure 5-75 shows the image and functionality of the cryostat used.



Figure 5-75: Photograph of inside the cryostat

The samples were kept on dry ice and embedded using the OCT embedding matrix before placing in the cryostat to be sliced.

5.5.2 Results

The results presented here are the spectra of different tissue samples that were provided using the techniques outline in section 5.5.1.1, 5.5.1.2, and 5.5.1.3.

5.5.2.1 *Formalin fixed paraffin embedded tissue samples*

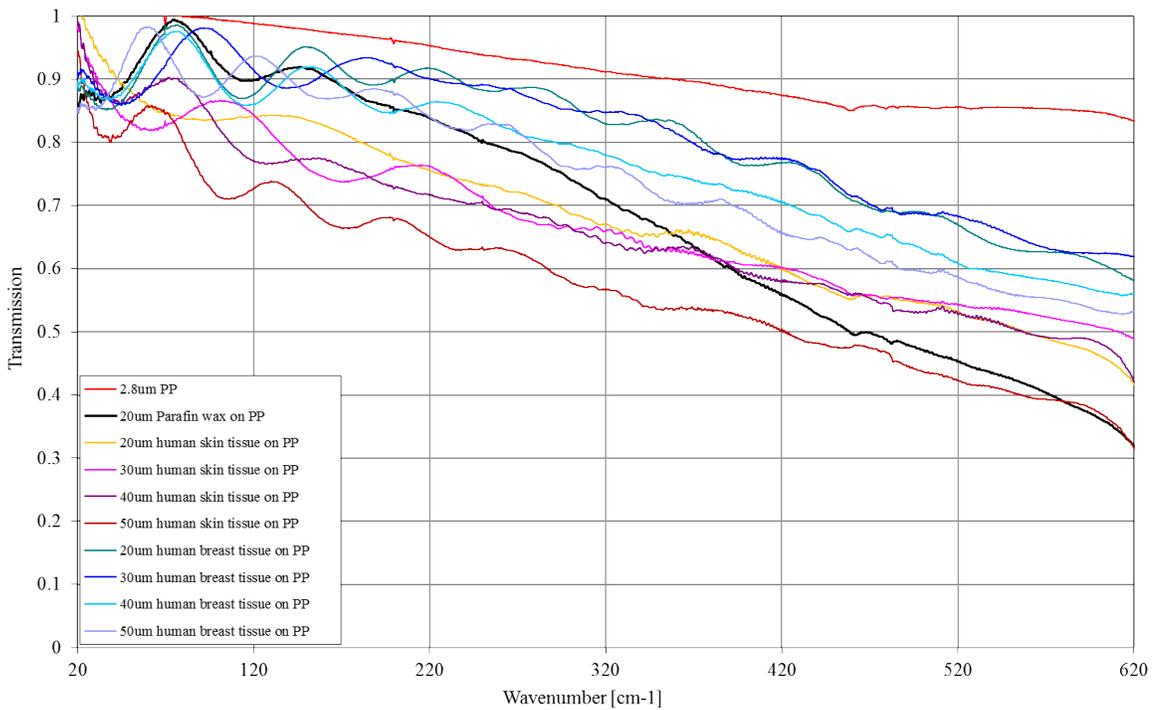


Figure 5-76: THz spectra of human skin and breast tissue embedded in wax

Figure 5-76 is the transmission spectra for various thicknesses of human skin and breast tissue. However, the inconsistency in the thickness is clear to see from the spectra as the spectra of the 20 μ m human breast tissue display very similar transmission to that of the 30 μ m sample. Inconsistencies can also be seen in the 40 μ m and 50 μ m human breast tissue.

The ripples in the spectra found in Figure 5-76 are due to the Fabry Perot effect of the planar dielectric sample, similar to polypropylene. This effect can be useful in determining the refractive index or thickness of the wax, but as the thickness of the tissue is not uniform and consistent with the surrounding wax, it was not possible to model the spectra and remove its features.

5.5.2.2 Lyophilised tissue samples

Samples of chicken breast were lyophilized and investigated under vacuum using the FTS.

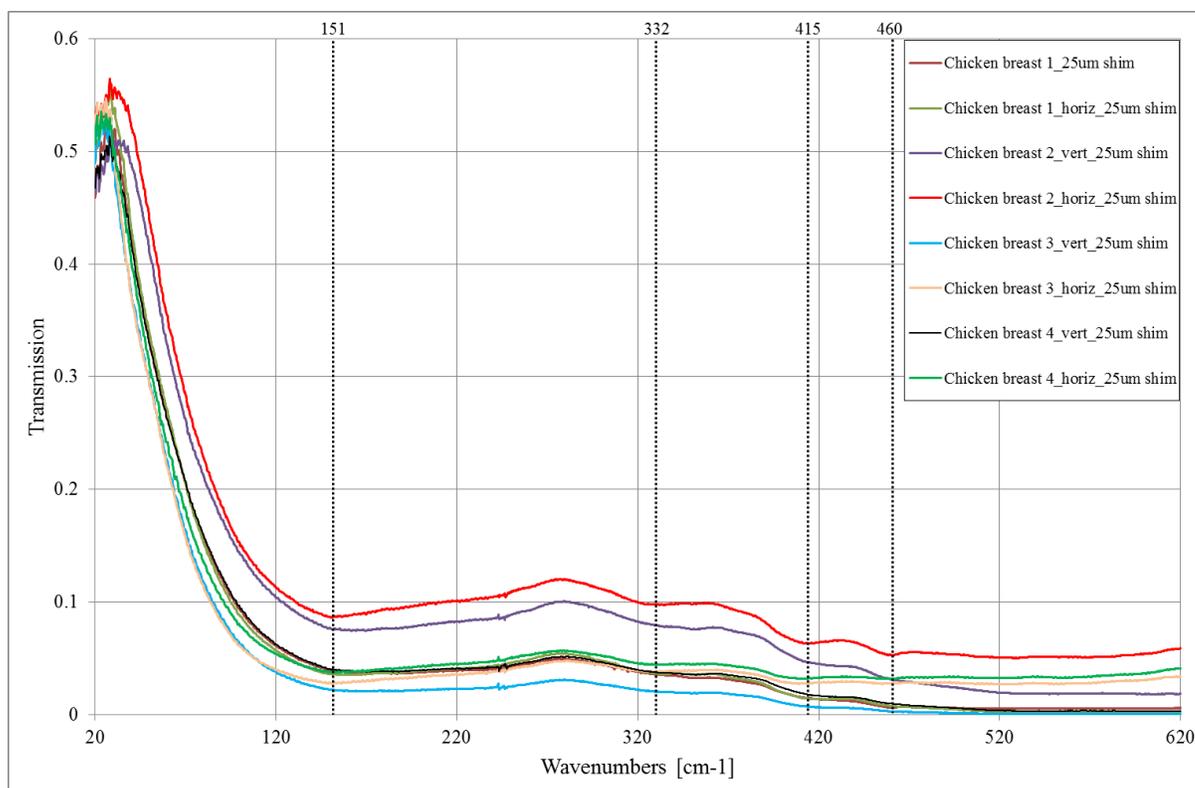


Figure 5-77: THz spectra of dry, flat chicken breast after being placed under weight and flattened within known shims while being left to dry in a vacuum oven

It is important to note that all the samples in Figure 5-77 are measured without any substrate. The spectra in Figure 5-77 shows 4 samples of chicken breast numbered and labelled as “vert” or “horiz”. The tissues had linear fibres which could potentially have affected the data as the FTS is a polarised Martin-Puplett interferometer. The spectra for all the samples were consistent which means there are no polarisation effects. Identified in Figure 5-77 are four absorption features common to each sample at 151, 332, 415 and 460 cm^{-1} .

There has been some attempt at learning more about tissues from THz-TDS by extracting the refractive index and absorption coefficient of the sample (Ashworth, et al., 2009) but

this only determined the composition of the sample in terms of healthy fat, healthy fibrous and cancerous through the water content. Another method used THz pulsed imaging (TPI) systems by differentiating between tissue types and identifying cancer margins in single samples (Woodward, et al., 2003). However, these systems exploit the fact that water concentrations within cancer cells are high in comparison to healthy cells which will give a greater image contrast. These methods are very efficient at probing the composition of the tissue thickness and water content of samples (Pickwell, Cole, Fitzgerald, Pepper, & Wallace, 2004) but do not provide the signal to noise and the spectroscopic information across a broad range of frequencies (Han, et al., 2001). This does not answer the questions of how the cancer was formed and what is different between the healthy and damaged tissue. It is also unclear as to which stage the cancer is in which is essential when early diagnosis can stop the cancer from spreading.

5.5.2.3 Cryosectioning tissue samples

Two microtomes, from the University Hospital of Wales and Cardiff University Biosciences Laboratory, were used to cryosection samples of chicken breast and liver, pig kidney and liver, and beef.

5.5.2.3.1 Chicken liver and breast from 5 to 22 μm

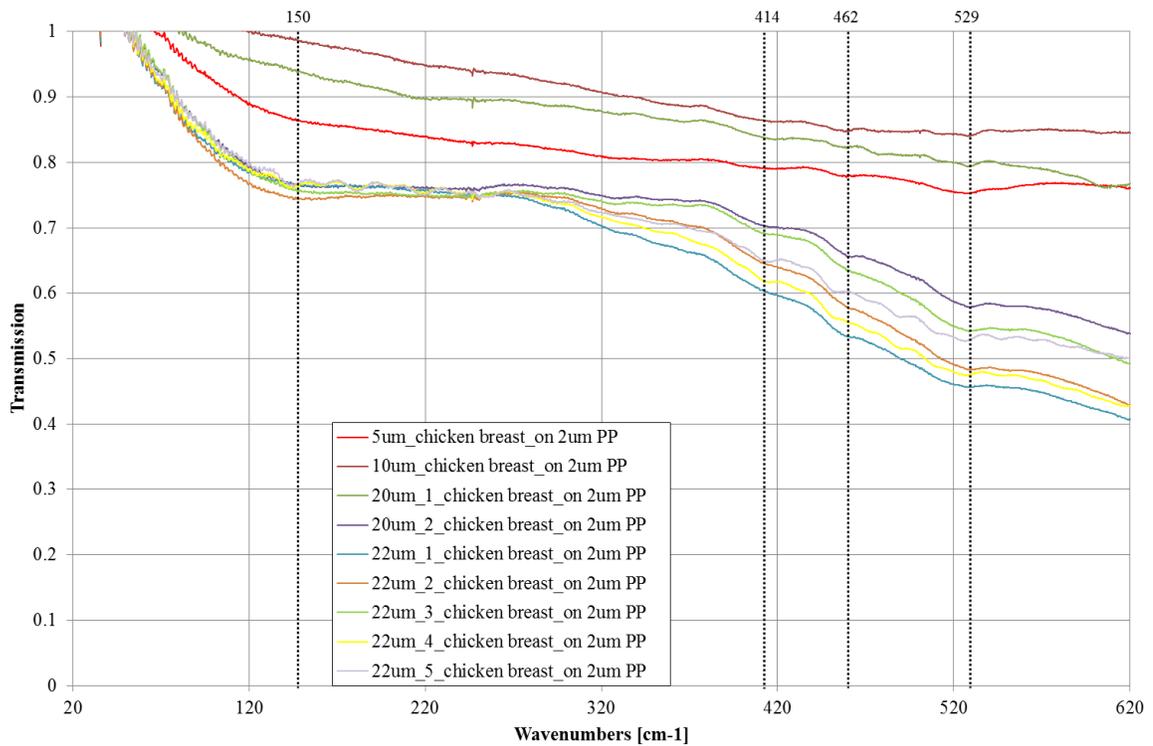


Figure 5-78: THz spectra of chicken breast at various thicknesses from 5 to 22 μm

Figure 5-78 highlights the most prominent absorption features in each of the spectra for chicken breast. These are found at approximately 150 cm^{-1} , 414 cm^{-1} , 462 cm^{-1} and 529 cm^{-1} . As a comparison, data were also taken for chicken liver which was difficult to align in the spectrometer as liver can be quite porous.

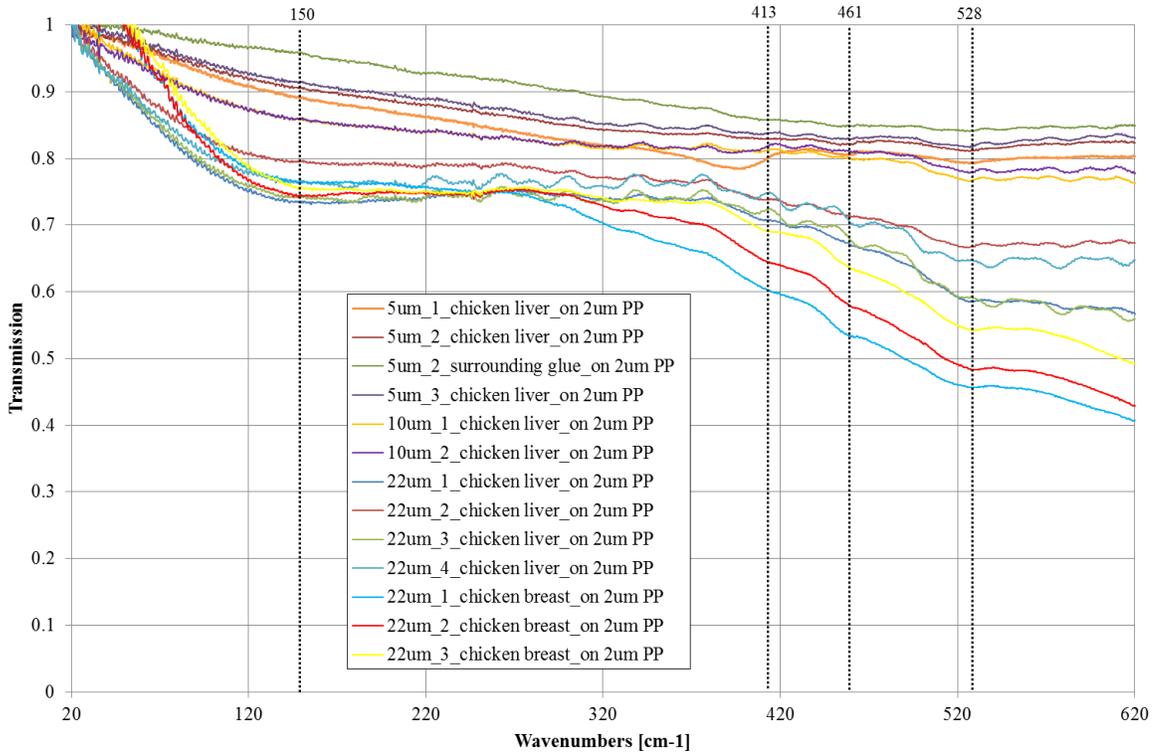


Figure 5-79: THz spectra of chicken liver at various thicknesses between 5 and 20 μm and chicken breast at 22 μm for comparison

Figure 5-79 highlights the main absorption features found in chicken liver at different thicknesses. These are identified at 150 cm^{-1} , 413 cm^{-1} , 461 cm^{-1} and 528 cm^{-1} . The numbering system for all the samples were for identification as each sample was offloaded onto individual substrate rings. The samples with the same thickness are identical, but there can be some discrepancies within the sample which are unable to be controlled. Clearly, the spectra for chicken breast and liver are identical with absorption features at the same frequencies within error.

Having broad features such as the one centred at 150 cm^{-1} makes it difficult to determine what the main resonant contributions are. There are likely to be a number of smaller absorption features within the broad feature from different proteins configurations. Therefore, there could be a contribution from all the nucleobases within this feature which result in a cumulative broad feature. Comparing different tissue types is important as they contain different protein structures.

5.5.2.3.2 Beef muscle, pig liver and kidney from 20 to 60 μm

Samples of beef, pig liver and kidney were investigated using the microtome at the Cardiff University Biosciences Laboratory. The following graphs are plotted separately for ease of comparison as there were many samples.

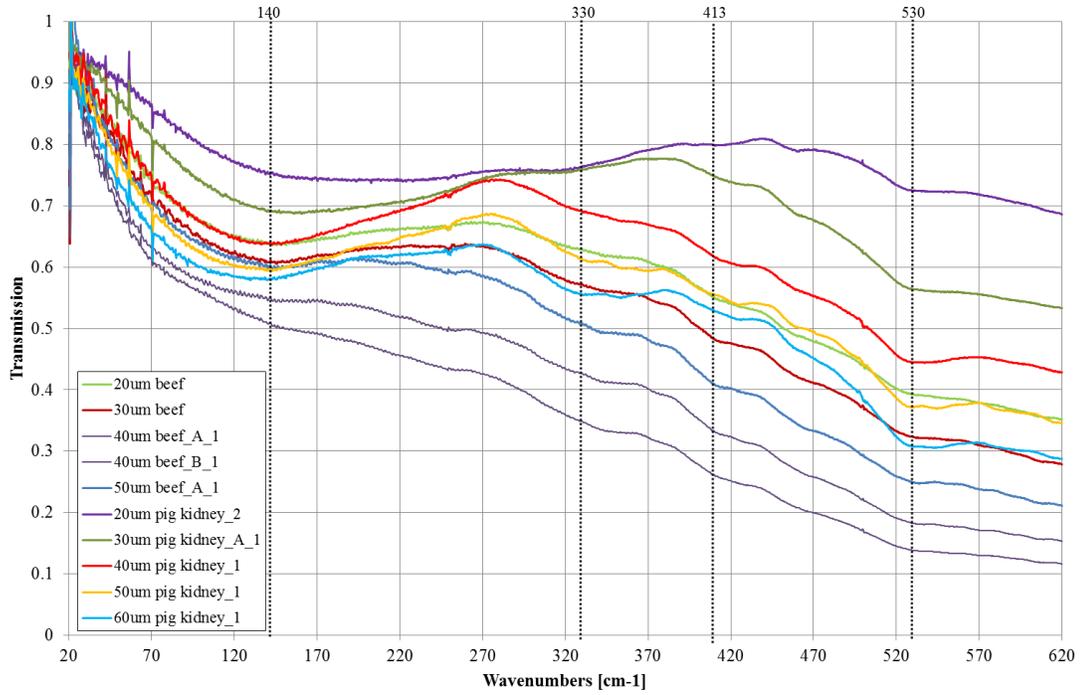


Figure 5-80: THz spectra of beef and pig kidney from 20 μm to 60 μm

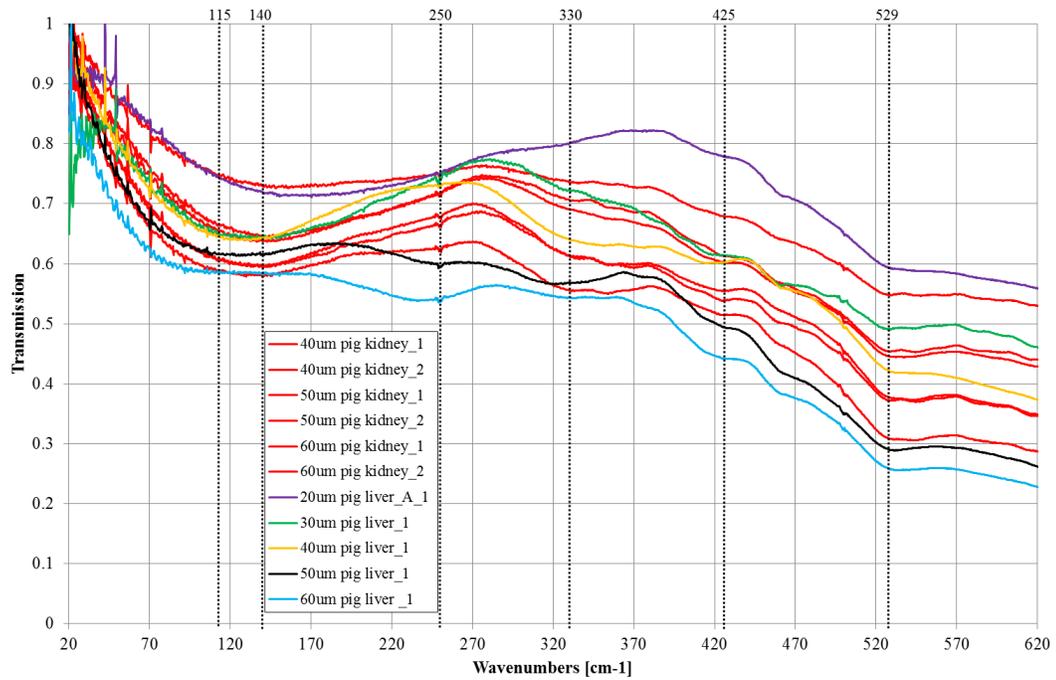


Figure 5-81: THz spectra of pig kidney in red and pig liver from 20 μm to 60 μm

5.5.2.3.2.1 20 to 300 cm⁻¹ (0.6 to 9 THz)

In the region below 300 cm⁻¹, all the tissue samples in Figure 5-80 and Figure 5-81 have a broad absorption feature which is centred around 140 cm⁻¹. However, in Figure 5-81, the samples of pig liver have the same broad absorption feature only at lower frequencies and is identified at 115 cm⁻¹. Figure 5-81 also highlights an absorption feature at 250 cm⁻¹ which is unique to and most prominent in the 40, 50 and 60 μm samples of pig liver.

5.5.2.3.2.2 300 to 600 cm⁻¹ (9 to 18 THz)

The features above 300 cm⁻¹ are common in all the samples across Figure 5-80 and Figure 5-81. Although the features are small and broad, which makes it difficult to identify and characterise them, the beef, pig kidney and pig liver have the same absorption features at 330, 413/425 and 530 cm⁻¹.

The thicker samples present more attenuation across all frequencies which is expected for denser samples. The absorption graphs shown by (Ashworth, et al., 2009) have a small difference between fibrous tissue and cancerous tissue which they say is a contribution from absorption and scattering but that the cancer samples will have a higher attenuation due to structural changes which, in turn, changes scattering properties.

Terahertz imaging and spectroscopy of tissue samples is still an ongoing area of research. Current work has concentrated on the contrast in images of tissue samples due to the higher concentration of water in the cancerous area. Some attempt at THz spectroscopy of animal tissue has been done using THz-TDS of pork skin, fat and tissue and rat skin and tissue (Hea, Azad, Ye, & Zhang, 2006). But there were clear issues with high

absorption of the transmitted THz source due to the high water content, which meant that these signals were multiplied by 700 to compare to the other spectra. For higher water content samples (i.e. in skin and lean tissue) there was a higher absorption at higher frequency, so beyond 0.7 THz (23 cm^{-1}) the samples were opaque.

5.5.3 Conclusion

Terahertz spectroscopy of tissue is limited due to the presence of water in biological tissue which drives research towards paraffin embedded samples. However, these samples do not work well for THz spectroscopy because the samples contain too much wax which acts like a Fabry Perot resulting in suppression of the absorption features due to the tissue itself.

The data presented in this thesis has shown that THz spectroscopy using an FTS can work in identifying absorption features unique to tissue samples. All the tissue samples were lean tissue (i.e. no skin or fat) and all have a broad absorption feature at 140 cm^{-1} which could be a superposition of resonances from the protein strands.

From section 5.4, there was a water absorption feature identified at 200 cm^{-1} which is not the same as the feature found in tissue. Therefore, water contributions can be ruled out of these tissue samples.

Tissue samples are a complex combination of water, proteins, cells, DNAs, and other minerals and molecules found within living organisms. The samples of tissue investigated might potentially possess a unique absorption feature due to the protein structures. There are some differences between samples of tissue from different sources, but more research

would have to done to identify whether THz can be sensitive to small changes in the tissue due to cancer.

CHAPTER 6. DISCUSSION

In this thesis, the potential of THz spectroscopy to detect the first mutations that lead to cancer has been investigated. In order to build a better picture of the electromagnetic interaction of large biological molecules with THz radiation, long chain strands of DNA and RNA have been investigated as well as various tissue samples. This work has then been complimented by a more detailed analysis of the sub units of DNA; the nucleobases and nucleosides and a comparison made between experimental data and DFT based modelling.

6.1 Nucleobases and nucleosides

Each nucleobase; guanine, cytosine, adenine, thymine, and uracil, was investigated using a Fourier Transform Spectrometer (FTS) in the range of 0.6 to 18 THz (20 to 600 cm^{-1}). Four of the main nucleobases in DNA were made into solutions using DI water, IPA, D₂O (heavy water), Hank's Buffered Salt Solution (HBSS) and Phosphate Buffered Saline (PBS) and left to evaporate at room temperature on 2 μm of polypropylene (PP).

Each nucleobase demonstrated a unique THz spectrum with a number of strong and sharp absorption features. Interestingly, the THz spectra of guanine, thymine and adenine showed the same absorption features irrespective of the solvent used, but cytosine had a different THz spectrum for the sample made using IPA compared to the other solvents. The different features were prominent throughout the range from 20 to 300 cm^{-1} (0.6 to 10 THz). The difference between the spectra was attributed to the difference in crystalline

structure of the sample. The hydrophobicity of the nucleobases resulted in aggregation within the solvent, causing areas of high concentration and granularity within the solution. DI water, D₂O, HBSS and PBS have high surface tension unlike IPA. The samples made using IPA had an even distribution on the substrate and evaporated quickly. This would have a large effect on the crystalline structure and resulted in a difference in the THz spectra.

DNA and its components would naturally be surrounded by water within cells, but it is necessary to investigate the difference in the solvation effects of the nucleobases and the resulting effects on its solid-state structure after evaporating. DI water was used as the purest form of H₂O, but the HBSS and PBS were also a necessary solvent to investigate as these contain traces of other substances that can be found naturally within cells. However, despite the difference in the THz spectrum of cytosine made using IPA, the results showed no effects on the individual molecular structure of the nucleobases.

To further investigate the structure of the samples, attention was paid to the spectra of bases with similar molecular structure; the purines and pyrimidines. Guanine and adenine (purines), made using DI water and IPA, had similar absorption features identified at 55 cm⁻¹, 71 cm⁻¹, 84 cm⁻¹, 99 cm⁻¹, 103 cm⁻¹, 141 cm⁻¹, 239 cm⁻¹, and 247 cm⁻¹. They have similar crystalline structures and possess similar hydrogen bonding sites. Cytosine and thymine (pyrimidine), made using DI water, showed common absorption features (identified at 75, 87, 108, 190 – 202 and 415 – 437 cm⁻¹) but displayed some frequency shifts. The spectra for cytosine and thymine made using IPA also demonstrated similar resonances but there were greater frequency shifts between the assumed common absorption features.

Comparison of the spectra of purine and pyrimidines showed a big difference in the region between 245 cm^{-1} and 400 cm^{-1} . There were no distinct absorption features in the pyrimidines whereas there are common resonances in the purines at $339 \pm 3\text{ cm}^{-1}$ and $351 \pm 3\text{ cm}^{-1}$. These absorption features are suggested to be the identification of the features from the imidazole ring, which is only present in the purines.

The nucleosides also presented a set of unique THz spectra with many strong absorption features. Guanosine, cytidine, thymidine and adenosine were made into a solution using DI water and IPA and compared to their nucleobase counterpart. The addition of a 5-carbon sugar ring perturbs the hydrogen bonding of the molecules and results in a distinct difference in the THz absorption therefore, an attempt was made to extract absorption features that were related to a glucose molecule. The glucose sample indeed showed absorption features in the THz region but it was not possible to match these to the spectra of the nucleosides as the THz spectra for pure glucose demonstrated a different conformation and crystalline structure to that when part of a nucleoside.

Within the structure of the nucleosides, the 5-carbon sugar ring is attached via a glycosidic linkage. A common region of investigation was between 250 to 400 cm^{-1} due to the presence of absorption features in the nucleosides samples but the lack of any features in the nucleobase samples. According to (C.P. Beetz Jr., 1980), it was noted that any absorption features within the 250 and 400 cm^{-1} region can be associated with the ribose ring and not with the base. The results in this thesis demonstrated that there was a clear difference within the 250 and 400 cm^{-1} region for each nucleoside and their corresponding nucleobase, which further supports that the absorption features here are associated with the ribose ring.

From comparing the THz spectra of all the nucleobases and nucleosides, it seems that the nucleosides, the structurally larger molecules, result in a spectrum with a higher number of strong absorption features. The increased number of hydrogen bonding sites in a nucleoside molecule will result in a larger number of absorption features which supports the results shown in this thesis. This is important in the understanding of the interaction between the THz frequencies and large biological molecules. It suggests that there is a link between the size of the molecular sub unit and the number of absorption features, which could pose an issue when investigating a branch of these sub units or the formation of a long chained molecule such as DNA or RNA.

To better understand the origin of the THz absorption features of nucleobases, Density Functional Theory (DFT) simulations were calculated for guanine, cytosine, thymine, adenine and uracil. The simulations were based on some initial conditions, including bond lengths and angles, from crystallographic data. For cytosine, a simulation was run for a monohydrated and anhydrous sample in order to probe the different THz spectra achieved for samples made using DI water and IPA. This suggests that the THz spectra of cytosine made using DI water resulted in a crystalline structure which included some H₂O molecules.

When the simulated spectra were compared to the experimental results, it was clear that there was an agreement between the two. However, there was a clear issue at lower frequencies, with ‘common’ absorption features showing a frequency shift to higher energies. The simulated data was scaled in frequency leading to a better agreement with the experimental spectra at lower frequencies. However, the absorption features above 200 cm⁻¹ showed a better correlation to the original simulated data. For simplicity, the

modelling assumes 0 K thermal environments. Clearly this is not the case for the experimental results. At frequencies below 200 cm^{-1} , solid-state vibrations are influenced by the lattice dimensions of the unit cell resulting in a temperature dependent frequency shift. At frequencies above 200 cm^{-1} , the absorption features are influenced by internal or molecular vibrations which are less temperature sensitive.

Simulations were also run for thymidine, uridine and cytidine. As these are larger molecules, the simulations take a greater amount of computational power and time to run. The resulting comparison between the simulations and experimental results showed a poorer correlation between the absorption features. This work is still ongoing with the aim of having a complete set of experimental and simulated results for all 5 nucleosides.

DNA strands with thousands of nucleobases contain hundreds of thousands of hydrogen bonds, which will potentially give a THz absorption that includes hundreds of individual resonances. They will all contribute to the extraordinary dipole properties due to the hydrogen bonds which interact through long-range Coulomb forces across the DNA strands. To investigate this better, the THz spectra of RNA and DNA molecules were investigated.

6.2 DNA&RNA

The THz spectra of dried human DNA and RNA were taken using the FTS between 20 and 600 cm^{-1} (0.6 to 20 THz). The comparison of the THz spectra of human DNA and RNA in section (1.2.1) has identified an absorption feature in the DNA centred at 158 cm^{-1} which is not in the spectra for RNA. It suggests that this broad absorption feature

could be a collection of many resonances due to the hydrogen bonding between the nitrogenous bases at the centre of the double helix.

To further investigate, the THz spectra of dried mice DNA were investigated. Human beings and mice share a similar genome, however, the absorption feature identified in the human DNA was at 158 cm^{-1} whereas the absorption feature for the mice DNA was centred at 210 cm^{-1} . This could suggest that THz frequencies are sensitive to DNA from different species, or could be that the DNA has been extracted in a different way.

A collection of mice DNA was investigated to identify whether there were any spectral differences between the DNA of mice from different colonies. Figure 5-55 presented the spectra for 8 different DNA samples which had a consistent broad absorption feature centred around 210 cm^{-1} .

Heat is often used to separate the double helix of a DNA molecule into single stranded DNA for replication purposes. Section (5.3.4) presents the results of heating the DNA samples to $70 \pm 5^\circ\text{C}$. This resulted in an additional feature in some of the samples at 35 cm^{-1} which could be the result of strand separation or conformational changes in the DNA. However, this is not conclusive as other factors such as the DNA length and nucleic base sequencing also have a big influence.

All the DNA samples were measured after dehydration on a $2\text{ }\mu\text{m}$ PP substrate, but the THz spectra of the same samples after a period of 5 and then 7 days suggested that the DNA might have changed over time. The broad feature at 210 cm^{-1} is not present in the samples after a period of 5 days and again after 7 days. However the feature found at 35 cm^{-1} is present but shifted to a frequency of 54 cm^{-1} after 5 days and 57 cm^{-1} after 7 days.

It is unclear what was changing over time as there are many uncontrollable factors such as the length of the DNA strand, the complementary base sequence and the conformation of the DNA strands after dehydration.

This work highlights that there is a need for further investigations with a variety of controlled elements to better understand what the absorption features correspond to and whether THz spectroscopy can be used to identify changes in healthy and cancerous DNA. There are many different factors that can be investigated:

- The conformation of DNA and RNA molecules after dehydration. As shown in this thesis, there is change in the THz spectra of dried DNA over time. This would need further time controlling experiments to evaluate the rate at which DNA changes, what the conformational changes are, and whether there is a stability point for the conformations of dried DNA.
- The length of the DNA and RNA samples were unknown which also increases the difficulty in understanding the absorption features from the THz spectra. Knowing the length of the DNA and RNA would help with calculating the energies associated with the bonding of the double helix. It could potentially change the conformation of the DNA and RNA after dehydration which could change the THz spectrum. As the number of hydrogen bonds are decreased (due to a smaller number of base pairs in a smaller strand of DNA) would the number of resonances decrease? Would THz frequencies using the technique highlighted in this thesis, show individual resonances?
- Control over the sequence of base pairs. Different base pairs have different energies for example, guanine and cytosine have 3 hydrogen bonding sites which

results in a stronger bond and higher denaturing (uncoiling) temperatures, compared to adenine and thymine. This would help in calculating the correct “melting” temperature of the DNA and potentially result in a double- and single-stranded molecule from the same source, giving better reliability and consistency.

6.3 Solutions

All the biological samples investigated in sections 5.1 and 5.3 of this thesis have been dehydrated and the data taken in vacuum. Nucleobases, nucleosides, RNA and DNA are naturally found within the nucleus of cells and are surrounded by liquid water. However, water is highly absorptive in the THz region due to the high number of hydrogen bonds. Many have attempted to understand the nature of water in the THz region using THz-Time Domain Spectroscopy (Rønne, Åstrand, & Keid, 1999) (Bergner, Heugen, Brundermann, Schwaab, & Havenith, 2005). Due to the limited frequency range and the relatively low signal to noise of these systems, work has proven very difficult. Many have attempted to simulate the THz response of water and the effect of solvation on biological molecules (Heyden M, 2010). A short piece of work was undertaken here, using an FTS, to determine a broad THz spectrum of water.

A sample holder was designed and manufactured allowing liquid samples to be measured in the FTS, and the FTS was modified to allow a CO₂ purge. The resulting THz spectra of DI water showed a broad absorption feature at 175 to 200 cm⁻¹. There were also small features at 50 cm⁻¹ and 456 cm⁻¹. These features were compared to previous literature which confirmed that the 200 cm⁻¹ was a translational absorption feature due to hydrogen bond stretching and bending. The smaller 50 cm⁻¹ was attributed to hydrogen bond

bending and higher frequency absorption features were attributed to librations, or “swaying” of the water molecules (Chaplin, 2016).

The THz spectrum of heavy water was investigated as some literature had suggested that the THz absorption of D₂O would be different to the THz spectra of H₂O (Chaplin, 2016). The spectra for H₂O and D₂O possessed the same broad absorption feature at 175 to 200 cm⁻¹, the small feature at 50 cm⁻¹ and at higher frequencies around 456 cm⁻¹. This better supports the assumption that the broad absorption feature at 175 to 200 cm⁻¹ arises due to translational vibrations involving OH---O hydrogen bond stretching and bending and that the replacement of a hydrogen with a deuterium does not change the THz spectra of liquid water.

To investigate the effects of water on nucleobases, the THz spectra of guanine in DI water solution were measured. A high concentration of guanine powder was needed but as guanine is hydrophobic, it does not dissolve in water, resulting in an optically opaque sample. The THz spectrum of the guanine in liquid solution did possess the broad absorption features significant to H₂O at 50 cm⁻¹ and 175 – 200 cm⁻¹, but it also had weak features which corresponding to the absorption features found in the dried guanine. This provides some evidence that THz spectroscopy could be able to identify the characteristic absorption features of nucleobases when in a liquid solution, providing that the concentration of the powder is high enough. Although, the aggregation of guanine molecules in solution due to their hydrophobicity, could cause them to form a similar crystalline structure to that found in the dried solid state structure, which is not a true replication of biological molecules within living organisms.

6.4 Tissue

Tissue samples are a complicated structure including proteins, cells, DNA and other nutrients and microscopic components.

Several methods of preparing tissue samples have been investigated and are detailed in section 5.5.1. Formalin Fixed Paraffin Embedded (FFPE) tissue is often the most commonly used technique for histological purposes. It is one of the most common forms of identification of cancer margins within tissue samples. However, it is not a viable option for investigations in THz spectroscopy. According to literature, lyophilisation is the best way of probing tissue through using THz spectroscopy (Png, et al., 2008). This method was performed on several samples of chicken breast but there were discrepancies in the sample thickness and difficulties in achieving uniform samples. The most reliable method found for this work was cryosectioning tissue samples.

Samples of chicken liver and breast were sectioned to a maximum thickness of 22 μm and a common broad absorption feature was identified at 150 cm^{-1} . Smaller features were present at 413, 461 and 530 cm^{-1} which seemed to be unique to the chicken breast tissue samples.

Samples of beef and pig kidney and liver also had a broad absorption feature centred around 140 cm^{-1} and a smaller broad absorption feature at 530 cm^{-1} . The beef sample showed an overall decrease in the transmission due to the density of muscle compared to other organs. The pig kidney and liver had smaller absorption features identified at 250, 330, 425 and 530 cm^{-1} .

The samples were all dehydrated and measured under vacuum but comparing to data from section 5.4, the water absorption feature identified at 200 cm^{-1} was not the same as the broad absorption feature found in the tissue samples at around 140 cm^{-1} as is shown in Figure 6-1. Therefore, water contributions can be ruled out of these tissue samples.

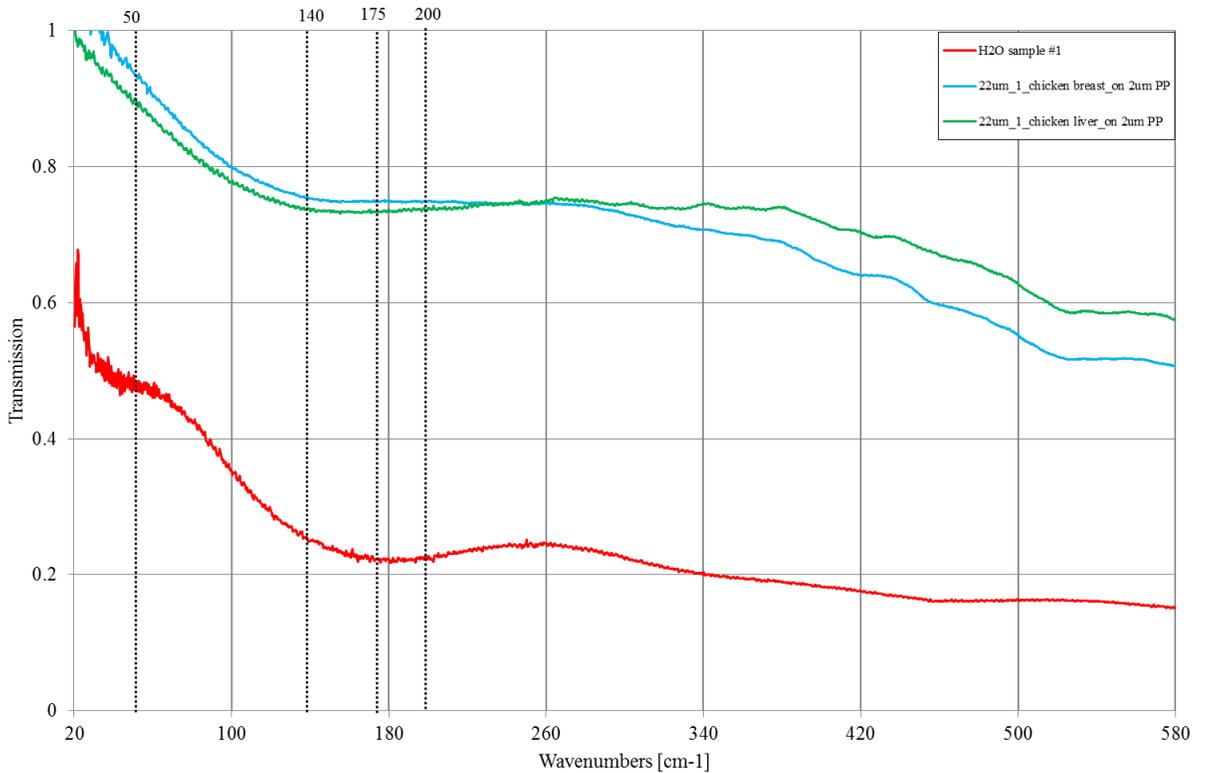


Figure 6-1: THz spectra of H₂O taken in a CO₂ purged FTS and the THz spectra of chicken breast and liver taken under vacuum in the same FTS

In section 5.3, there is a broad absorption feature in human DNA at 160 cm^{-1} and in the mouse DNA between 180 and 200 cm^{-1} . By comparing to the THz spectra of chicken breast and liver to the spectra of the dried DNA samples in Figure 6-2, there is no clear correlation between the DNA spectra and the tissue spectra, suggesting that the individual

components i.e. DNA, RNA and tissue, can be individually identified through their THz spectra.

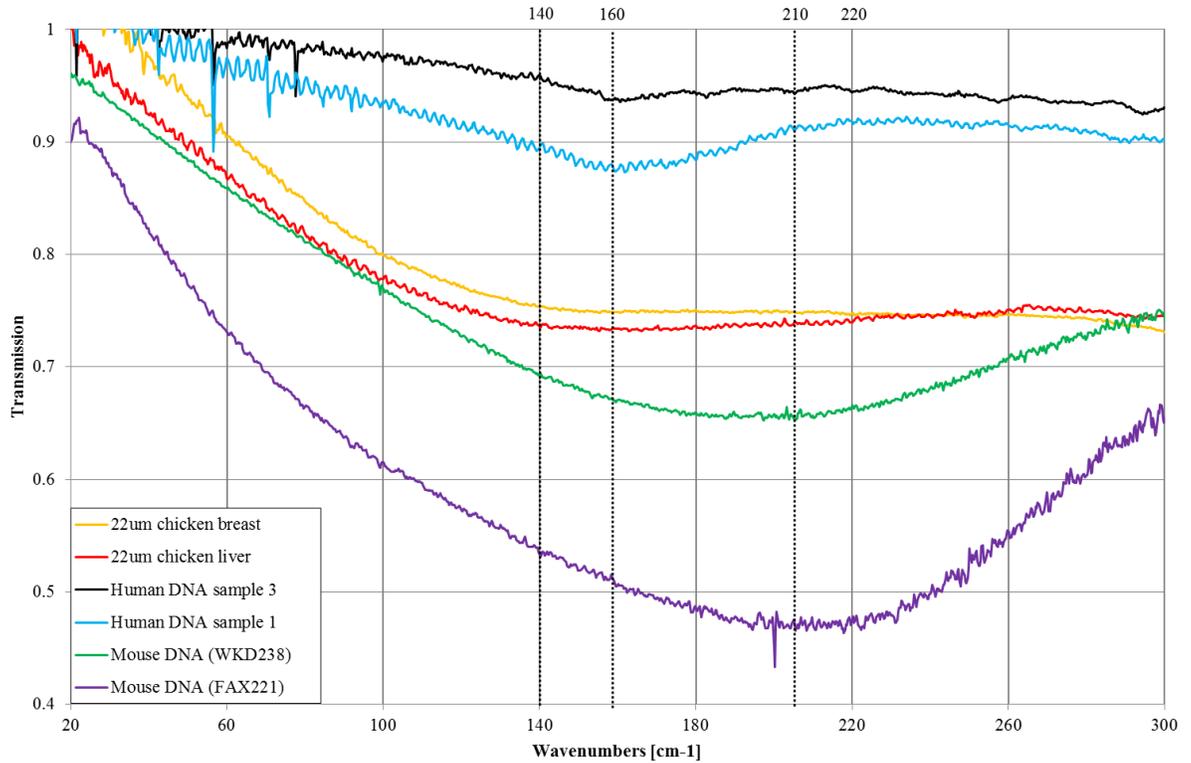


Figure 6-2: THz spectra of dried chicken breast and liver, human DNA and mouse DNA from 20 to 300 cm⁻¹

The absorption feature exhibited by the tissue samples is much broader than that for the DNA. Just as the THz spectra of DNA might contain a superposition of many sharp resonances due to the hydrogen bonds between the bases, the THz spectra of tissues is suspected to be made up of a collection of resonances including those due to the DNA molecules.

The results in this thesis suggest:

- Terahertz spectroscopy can be used to identify small changes in biological molecules. In the solid state, the results show that as the molecular size of the samples increase, there is an increase in the number of resonances. This means that when DNA and RNA are investigated, the broad absorption feature present in the spectrum for DNA could be a superposition of all the absorption features present in the individual constituents.
- THz spectroscopy is very sensitive to hydrogen bonding, which poses a problem when investigating the properties of biological molecules in their natural habitat; water. However, initial investigation shown in this thesis has provided some insight in the identification of guanine molecules in liquid H₂O. Further investigations will continue with other nucleobases, nucleosides, DNA and RNA within a solution made with H₂O, only to be aware that the concentration of the solute must be high.
- The potential of THz spectroscopy as a tool for identifying early onset stages of cancer is unclear. The investigation of varying types of normal animal tissue in this thesis proved that identification of different tissue types could be done using THz spectroscopy. However, the method that was implemented meant that it was not a replication of the samples' natural state. There is need for further investigations of fresh tissue samples and measuring them using the FTS in a similar manner to that used for the solutions, shown in section 5.4.1. Other potential avenues of investigation using the FTS would be to determine the THz spectra of all the other constituents such as amino acids, proteins, enzymes, etc.

which could also contribute to the superposition of resonances that are presumed to be contributing to the broad absorption features.

CHAPTER 7. CONCLUSION

By comparing the THz spectra of the different components shown in this thesis, it has been shown that THz spectroscopy of biological molecules is a potentially useful and interesting tool. However, the extraction of tissue, measurement of solutions and the complexity of the molecules renders analysis complications. A combined approach of experimentally determined resonances and modelling has shown early success and will be required for full extraction of structural understanding in the future.

Nucleobases and nucleosides are a key component of DNA and RNA, and have shown interesting characteristic features in the region of 20 to 600 cm^{-1} (0.6 to 20 THz). This work has provided evidence that there are unique absorption features across the frequency range which have not been seen before due to the limited frequency ranges and relatively low signal to noise inherent to other THz techniques. Depending on the method and solvent used to create dry samples of nucleobases and nucleosides, the THz spectra have proven sensitive to small changes in the crystalline structure, as shown in the difference between cytosine made using IPA and DI water.

The difference in cytosine made using DI water and IPA was further investigated by comparing experimental results with DFT simulations. Spectra were produced for cytosine monohydrate and anhydrate which suggested that when the cytosine samples were made using DI water, some H_2O molecules has been contained in the crystalline structure.

It was apparent that common absorption features between the experimental and simulated spectra of nucleobases had a greater frequency shift at frequencies below 200 cm^{-1} . This was resolved by scaling the simulated data. At these frequencies, the solid-state vibrations are strongly influenced by lattice dimensions within their unit cell meaning that the absorption features in this region will be temperature dependent. As the simulated data is based on samples at 0K for simplicity, there is a shift between the simulated and experimental absorption features. Sensible future direction will be to take experimental data at cryogenic temperatures. This is indeed possible at Cardiff University within the FTS laboratory with the addition of a 4K detection system whereby the samples would sit within windows manufactured within the detector system and maintaining the same constant temperature. However, the data produced using DFT simulations in this thesis are in very good agreement to the experimental results, including the two different spectra that were produced from cytosine being made using DI water and IPA.

Comparing the THz spectra of human DNA and RNA showed that DNA had a THz absorption feature at 158 cm^{-1} whereas RNA was apparently featureless. Human and mouse DNA samples were compared which showed a difference in the frequency of their broad absorption features with mouse DNA samples having an absorption feature centred at 210 cm^{-1} . However, this work is not conclusive as there are a number of uncontrollable factors that could influence the THz spectra of the samples such as the effects of dehydration on the conformation of RNA and DNA, the nucleic acid sequence, the length of each RNA and DNA strand, etc. To further understand DNA and RNA in the THz frequency region, there is a need to understand the conformation of DNA and RNA molecules after dehydration. As shown in this thesis, there is change in the THz spectra

of dried DNA over time. As the length of the DNA and RNA samples were unknown, having control over the number of base pairs and the length of the molecules would help with calculating the energies associated with the bonding of the double helix. Furthermore, having an understanding of the sequence of base pairs would further specify the energies associated with unravelling the double helix.

This investigation has proven that H₂O and D₂O are highly absorptive in the THz region due to the high density of hydrogen bonds. The future potential of THz spectroscopy of biological molecules is difficult to predict. However, this thesis has provided further investigation of the THz spectroscopy of solutions which has been proven to be difficult using other conventional TDS systems. It has provided a better understanding of the interaction between THz frequencies and liquid solutions which is essential in developing this technique towards *ex-* and *in-vivo* investigation of large biological molecules.

Tissue samples from animal organs and muscle, prepared by cryosectioning, were investigated. Investigating chicken breast and liver showed that both possessed a broad absorption features at 150 cm⁻¹. Smaller features were present at 413, 461 and 530 cm⁻¹ which seemed to be unique to the chicken breast tissue samples. The investigation of varying types of normal animal tissue in this thesis proved that identification of different tissue types could be done using THz spectroscopy. However, there is need for further investigations of fresh tissue samples and measuring them using the FTS in a similar manner to that used for the solutions, shown in section 5.4.1.

Comparing the THz spectra of the chicken breast and liver to THz spectra of H₂O showed that both possessed a broad absorption feature but at different frequencies suggesting that the THz absorption feature found in the tissue is unique to the tissue fibres rather than any water. Further comparison of the THz spectra of the chicken breast and liver to the THz spectra of dried DNA showed that they all had broad absorption features at different frequencies. Therefore, the broad THz absorption feature in the spectra of tissues has been assumed to be superposition of resonances due to the DNA molecules among other molecules. However this would require looking further at the THz spectra of all the other constituents such as amino acids, proteins, enzymes, etc.

This thesis has presented promising results on the advantage of THz spectroscopy using a unique and underutilised technique with world leading equipment which provides excellent signal to noise, broadband frequency range and other advantages that are not achievable using other conventional manufactured THz devices. It has proven that THz frequencies are useful in identifying small changes in the crystalline structure of solid structures which could have many advantages and application in pharmaceutical and drug industries. It has been shown that THz spectra of biological molecules within their natural habitat in solution will be difficult to conduct as water molecules are highly absorptive in the THz frequency region. However, THz spectra of dried samples of human and mouse RNA and DNA has unravelled a vast amount of interesting results including a significant difference in the spectra of human RNA and DNA, and the spectra of human DNA and mouse DNA. The difference between human and mouse DNA suggests that THz spectroscopy could potentially be used to identify between different species. There is a potential of learning more about the changes in the conformation of dried DNA and RNA

over a time which could unravel more information about the solvation effects and the assistance of water on the conformation of these molecules and others such as proteins. This thesis has provided a better understanding of the electromagnetic interaction of large biological molecules with THz radiation which will compliment and strengthen the future application of THz spectroscopy for detecting mutations that lead to cancer.

CHAPTER 8. REFERENCES

- Alexandrov, B. S., Gelev, V., Bishop, A. R., Usheva, A., & Rasmussen, K. (2010). DNA Breathing Dynamics in the Presence of a Terahertz Field. *Phys. Lett A.*, 374-384.
- Alexandrov, B., Bishop, A., & Usheva, A. (2008). The Peyrard-Bishop-Dauxois Model of DNA Dynamics. *APS*.
- American Cancer Society . (2015, November 30). *American Cancer Society* . Retrieved from Treatment and support/Ultrasound for cancer: <https://www.cancer.org/treatment/understanding-your-diagnosis/tests/ultrasound-for-cancer.html>
- Ashworth, P. C., Pickwell-MacPherson, E., Provenzano, E., Pinder, S. E., Purushotham, A. D., Pepper, M., & Wallace, V. P. (2009). Terahertz pulsed spectroscopy of freshly excised human breast cancer . *Optics Express*, 12444-12454.
- Auston, D. H., & Nuss, M. C. (1988). Electrooptical generation and detection of femtosecond electrical transients. *IEEE Journal of Quantum Electronics*, 184-197.
- B M Fischer, M. W. (2002). Far-infrared vibrational modes of DNA components. *Physics in medicine and biology* , 21(47), 3807-3814.
- B.S. Alexandrov, V. G. (2010). DNA breathing dynamics in the presence of a terahertz field. *Physics letter A*, 374, 1214-1217.

- Bailey L.E., N. R. (1997). Normal coordinate analysis and vibrational spectra of adenosine. *Biospectroscopy*, 47-59.
- Barker , D. L., & Marsh, R. E. (1964). The crystal structure of cytosine. *Acta Crystallographica* , 1581-1587.
- Becke, A. D. (1993). A new mixing of Hartree–Fock and local density-functional theories. *J. Chem. Phys.* (98).
- Bell, R. J. (1972). *Introductory Fourier Transform Spectroscopy*. New York: Academic Press, Inc. .
- Benmore, A. K. (2008). Quantum Differences between Heavy and Light Water. *Physical Review Letters*, 101(6).
- Bergner, A., Heugen, U., Brundermann, E., Schwaab , G., & Havenith, M. (2005). New p-Ge THz laser spectrometer for the study of solutions: THz absorption spectroscopy of water . *AIP*.
- Boelsma E, G. S. (2000). Characterization and comparison of reconstructed skin models: morphological and immunohistochemical evaluation. *Acta Derm. Venereol*, 82-88.
- Bruker. (2017). *VERTEX 70 overview*. Retrieved from Bruker : <https://www.bruker.com/products/infrared-near-infrared-and-raman-spectroscopy/ft-ir-research-spectrometers/vertex-series/vertex-7070v/overview.html>

- C.P. Beetz Jr., G. A. (1980). The low frequency vibrations of pyrimidine and purine bases. *Spectrochimica Acta Part A: Molecular Spectroscopy*, 299-313.
- Canadian Cancer Society. (2017, April 27). *Nuclear Medicine Imaging*. Retrieved from Canadian Cancer Society: <http://www.cancer.ca/en/cancer-information/diagnosis-and-treatment/tests-and-procedures/nuclear-medicine-imaging/?region=on>
- Cancer Research UK. (2014, October 14). *What is cancer?* Retrieved from Cancer Research UK: <http://www.cancerresearchuk.org/about-cancer/what-is-cancer>
- Cancer Research UK. (2015, April 25). *Cancer Research UK*. Retrieved from About cancer/MRI scan: <http://www.cancerresearchuk.org/about-cancer/cancer-in-general/tests/mri-scan>
- Cancer Research UK. (2017, April 25). *Our policy on early diagnosis*. Retrieved from Cancer Research UK: <http://www.cancerresearchuk.org/about-us/we-develop-policy/our-policy-on-early-diagnosis>
- Carnegie Mellon University. (2009, September 29). *About John A. Pople* . Retrieved from The John A. Pople lecture in theoretical and computational chemistry: <http://www.chem.cmu.edu/pople/pople.html>
- Carven, R. J., & McClure, B. M. (1973). New Investigations of Cytosine and Its Monohydrate. *Acta Crystallographica*, 1234-1238.

- Chandra, A. K., Nguyen, M. T., Uchimaru, T., & Zeegers-Huyskens, T. (2000). DFT study of the interaction between guanine and water. *Journal of molecular structure*, 61-66.
- Chaplin, M. (2016, December 16). *Water vibrational spectrum*. Retrieved from Water structure and science : http://www1.lsbu.ac.uk/water/water_vibrational_spectrum.html#ta
- Choe, C., Lademann, J., & Darvin, M. (2016). Depth profiles of hydrogen bound water molecule types and their relation to lipid and protein interaction in the human stratum corneum in vivo. *Royal Society of Chemistry*, 6329-6337.
- Clarke, J., Dunning, D., Surman, M., & Thompson, N. (2010, March 22). *Science and Technology Facilities Council - ASTeC*. Retrieved from ALICE free electron laser: <https://www.astec.stfc.ac.uk/Pages/ALICE-free-electron-laser.aspx>
- Clegg, K. G. (2006). Anhydrous guanine: a synchrotron. *Acta Crystallographica Section C*, 62, 515-517.
- Craven, R. M. (1973). New Investigations of Cytosine and Its Monohydrate . *Acta Crystallographica Section B*, 29, 1234-1238.
- Daily, S. (2017, March 07). *RNA*. Retrieved from Science Daily: <https://www.sciencedaily.com/terms/rna.htm>

- Dhillon, S. S., Vitiello, M. S., Linfield, E. H., Davies, A. G., Hoffman, M., Booske, J., . . . Johnston, M. (2017). The 2017 terahertz science and technology roadmap. *Journal of Physics D: Applied Physics*.
- Diffen.com. (2017, March 12). *DNA vs. RNA*. Retrieved from Diffen.com: http://www.diffen.com/difference/DNA_vs_RNA
- ENEA. (2017, May 02). *Frascati Research Centre* . Retrieved from ENEA: <http://www.enea.it/it/centro-ricerche-frascati>
- Errington, J. R., & Debenedetti, P. G. (2001). Relationship between structural order and the anomalies of liquid water. *Nature*, *409*, 318-321.
- Ewels, C. (1997, November 13). *Hartree Equations*. Retrieved from ewels: <http://www.ewels.info/science/publications/thesis/html/node10.html>
- Fang Liu, E. T.-K. (2007). The Human Genomic Melting Map. *Computational biology*, 874-886.
- Fass, L. (2008). Imaging and cancer: A review. *Molecular Oncology* *2*, 115-152.
- Feng Yue, Y. C. (2014). A comparative encyclopedia of DNA elements in the mouse genome. *Nature*, 355-364.
- Feng Zhang, O. K.-i.-W. (2014). Analysis of vibrational spectra of solid-state adenine and adenosine in the terahertz region. *RSC Adv.*, 269-278.

- Fischer, B. M., Walther, M., & Jepsen, P. U. (2002). Far-infrared vibrational modes of DNA components studied by terahertz time-domain spectroscopy. *Physics in medicine and biology*, 3807-3814.
- Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., . . . al., e. (2017, January 05). *Density Functional Theory (DFT) methods*. Retrieved from Gaussian: <http://gaussian.com/dft/>
- Frohlich, H. (1968). Long-range coherence and energy storage in biological systems. *International Journal of Quantum Chemistry*, 641-649.
- Fröhlich, H. (1972). Selective long range dispersion forces between large systems. *Physics letters A*, 153-154.
- Gaiduk, V. I., & Vij, J. K. (2001). The concept of two stochastic processes in liquid water and analytical theory of the complex permittivity in the wavenumber range 0-1000cm⁻¹. *Physical Chemistry Chemical Physics*, 5173-5181.
- Gerdil, R. (1961). The Crystal Structure of Thymine Monohydrate. *Acta Crystallographica* , 333-344.
- Han, P. Y., Tani, M., Usami, M., Kono, S., Kersting, R., & Zhang, X.-C. (2001). A direct comparison between terahertz time-domain spectroscopy and far-infrared Fourier transform spectroscopy. *Journal of Applied Physics*, 2357-2359.
- Harrison, N. M. (n.d.). *An Introduction to Density Functional*. London and Warrington. Retrieved from http://www.ch.ic.ac.uk/harrison/Teaching/DFT_NATO.pdf

- He, M., Azad, A., Ye, S., & Zhang, W. (2006). Far-infrared signature of animal tissues characterized by terahertz time-domain spectroscopy. *Optics Communications*, 389-392.
- Hea, M., Azad, A. K., Ye, S., & Zhang, W. (2006). Far-infrared signature of animal tissues characterized by terahertz time-domain spectroscopy. *Optics Communications*, 389-392.
- Heyden M, S. J. (2010). Dissecting the THz spectrum of liquid water from first principles via correlations in time and space. *PNAS*, 12068-12073.
- Heyden, M., Sun, J., Funker, S., Mathias, G., Forbert, H., Havenith, M., & Marx, D. (2010). Dissecting the THz spectrum of liquid water from first principles via correlations in time and space. *PNAS*, 12068-12073.
- Hinchliffe, A. (2008). *Molecular Modelling for beginners*. London: John Wiley and Sons Ltd.
- Hoffman, M. (2006). *Novel Techniques in THz Time-Domain-Spectroscopy*.
- Hollas, J. M. (1995). *Modern Spectroscopy*. Chichester, England : John Wiley and Sons, Ltd. .
- HyperPhysics. (n.d.). *The Born-Oppenheimer Approximation*. (HyperPhysics) Retrieved November 15, 2016, from <http://hyperphysics.phy-astr.gsu.edu/hbase/molecule/bornop.html>

Importance of mouse genome. (2010, July 23). Retrieved from National Human Genome Research Institute : <https://www.genome.gov/10001345/importance-of-mouse-genome/>

Institute, N. H. (2016, May 11). *An Overview of the Human Genome Project.* Retrieved from National Human Genome Research Institute : <https://www.genome.gov/12011238/an-overview-of-the-human-genome-project/>

Interspectrum. (2016). *IR Sources.* (Interspectrum FTIR spectroscopy instruments & accessories) Retrieved October 2016, from Interspectrum: <http://interspectrum.ee/ir-sources/>

Jacquinet, P. (1960). New developments in interference spectroscopy. *Reports on Progress in Physics*, 23, 268-310.

Jeevan, R., Cromwell, D. A., Trivella, M., Lawrence, G., Kearins, O., Pereira, J., . . . Van der Meulen , J. H. (2008). Reoperation rates after breast conserving surgery for breast cancer among women in England: restrospective study of hosptial episode statistics. *British Medical Journal.*

Jeffrey , G. A., & Kinoshita, Y. (1968). The crystal structure of cytosine monohydrate . *Acta Crystallographica* , 20-28.

Jensen, F. (2006). *Introduction to Computational Chemistry.* Chinchester: Wiley .

K. Ozeki, N. S. (1968). The Crystal Structure of Thymine. *Acta. crystallographica Section B*, 1038-1045.

- Keutsch , F., & Saykally , R. (2001). Water clusters: untangling the mysteries of the liquid, one molecule at a time. *PNAS*, 10533-10540.
- Kim, S. J., & Born, B. P. (2008). Real-Time Detection of Protein–Water Dynamics upon Protein Folding by Terahertz Absorption Spectroscopy. *Angewandte Chemie International Editioion*, 6486-6489.
- King, M. D., Ouellette, W., & Korter, T. M. (2011). Noncovalent Interactions in Paired DNA Nucleobases Investigated by Terahertz Spectroscopy and Solid-State Density Functional Theory. *J. Phys. Chem*, 9467–9478.
- Kinoshita, G. J. (1961). The Crystal Structure of Cytosine Monohydrate . *Acta Crystallographica*, 16(20), 20-28.
- Korter, M. D. (2011). Noncovalent Interactions between Modified Cytosine and Guanine DNA Base Pair Mimics Investigated by Terahertz Spectroscopy and Solid-State Density Functional Theory. *J. Phys. Chem*, 14391–14396.
- Kristensen, T. T., Withayachumnankul, W., Jepsen , P. U., & Abbott, D. (2010). Modelling terahertz heating effects on water. *Optics Express*, 4727-4739.
- Laman, N. S. (2008). High-Resolution Waveguide THz Spectroscopy of Biological Molecules. *Biophysical Journal*, 1010-1020.
- Lee S.A., A. A. (2000). Temperature-dependent Raman and infrared spectra of nucleosides. Part I - Adenosine . *Journal of Raman Spectroscopy* , 891-896.

- Lewis, T. (2014, December 5). *What is an MRI?* Retrieved from LiveScience:
<https://www.livescience.com/39074-what-is-an-mri.html>
- Liu, H.-B., & Zhang, X.-C. (2006). Dehydration kinetics of D-glucose monohydrate studied using THz time-domain spectroscopy. *Chemical Physics Letters*, 229-233.
- Lopes R. P, M. M. (2012). Guanine: A Combined Study Using Vibrational Spectroscopy and Theoretical Methods. *Spectroscopy: An International Journal Volume 27*, 273–292 .
- Mario Barbatti, A. J. (2010). The UV absorption of nucleobases: semi-classical ab initio spectra. *Physical Chemistry Chemical Physics*, 4959-4967.
- Marsh, D. L. (1963). The crystal structure of cytosine. *Acta Crystallographica*, 17, 1581-1587.
- McCash, C. N. (1994). *Fundamentals of molecular spectroscopy*. Berkshire, England : McGraw-Hill Book Company Europe .
- Medical Radiation. (2016, October 6). *Computed Tomography (CT)*. Retrieved from Understanding Medical Radiation: <https://www.medicalradiation.com/types-of-medical-imaging/imaging-using-x-rays/computed-tomography-ct/>
- Nagai, M., Yada, H., Arikawa, T., & Tanaka, K. (2006). Terahertz time-domain attenuated total reflection spectroscopy in water and biological solution. *International Journal of Infrared and Millimeter Waves*, 505-515.

- Nakajima, S., Hoshina, H., Yamashita, M., & Otani, C. (2007). Terahertz imaging diagnostics of cancer tissues with a chemometrics technique. *Applied Physics Letters*.
- Nestle-Waters. (2017, April 11). *How much water is in the human body*. Retrieved from Nestle waters: <http://www.nestle-waters.com/healthy-hydration/water-body>
- News Medical. (2013, August 8). *What is RNA?* Retrieved from News-Medical: <http://www.news-medical.net/life-sciences/What-is-RNA.aspx>
- Nishizawa J, S. T., Nishizawa, J., Sasaki, T., Suto , K., Tanabe, T., & Saito, K. (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*, 229-239.
- Nuss, M. C., & Hu, B. B. (1995). Imaging with terahertz waves. *Optics Letters*, 1716-1718.
- Peter A.R. Ade, G. P. (2006). A Review of Metal Mesh Filters. *SPIE*, 6275. doi:10.1117/12.673162
- Pickwell, E., Cole, B. E., Fitzgerald, A. J., Pepper, M., & Wallace, V. P. (2004). In vivo study of human skin using pulsed terahertz radiation. *Physics in Medicine and Biology* , 1595-1607.

- Plusquellie D.F, K. T., R.J, B., R.C, B., C.R, H., A. R, W., & J. L, D. (2003). Continuous-wave terahertz spectroscopy of plasmas and biomolecules. *Int. journal of high speed elec. and syst.*
- Png, G. M., Choiq, J. W., Ng, B. W.-H., Mickan, S. P., Abbott, D., & Zhang, X.-C. (2008). The impact of hydration changes in fresh bio-tissue on THz spectroscopic measurements. *Physics in Medicine and Biology*, 3501-3517.
- QMCI. (2009, November). *Terahertz pyroelectric detectors*. (QMCI) Retrieved November 01, 2016, from http://www.terahertz.co.uk/index.php?option=com_content&view=article&id=236&Itemid=530
- R. Pedro Lopes, M. P. (2012). Guanine: A Combined Study Using Vibrational Spectroscopy and Theoretical Methods. *Spectroscopy: An International Journal*, 27, 273-292.
- Rønne, C., Åstrand, P.-O., & Keid, S. R. (1999). THz spectroscopy of liquid H₂O and D₂O. *Physical Review Letters*.
- Saykally, F. N. (2001). Water clusters: Untangling the mysteries of the liquid, one molecule at a time. *PNAS*, 98(19), 10533-10540.
- Schmittenmaer, C. A., & Kindt, J. T. (1996). Far-Infrared Dielectric Properties of Polar Liquids Probed by Femtosecond Terahertz Pulse. *J. Phys. Chem*, 10373-10379.

- Sedelnikova, O. A., Pilch, D. R., Redon, C., & Bonner, W. R. (2003). Histone H2AX in DNA Damage and Repair . *Cancer biology and therapy* , 233-235.
- Series, A. A. (1965). James W. Cooley and John W. Tukey . *Math. Comput.* , 19, 297-301.
- Smith, P. R., Auston, D. H., & Nuss, M. C. (1988). Subpicosecond photoconducting dipole antennas. *IEEE Journal of Quantum Electronics*, 255-260.
- Steinke, N., Gillams, R. J., Pardo, L. C., Lorenz, C. D., & McLain, S. E. (2016). Atomic scale insights into urea-peptide interactions in solution. *Physical Chemistry Chemical Physics*, 3862-3870.
- Sudarshan Mahapatra, S. K. (2008). Anhydrous Adenine: Crystallization, Structure, and Correlation with Other Nucleobases. *Crystal growth and design*, 8(4), 1223-1225.
- Susi H, A. J. (1970). Vibrational spectra of nucleic acid constituents –I. *Spectrochimica Acta*, 27A, 1549-1562.
- Swanson, E. S. (2011). Modeling DNA response to terahertz radiation. *Physical Review*, 83(4).
- Symons Mirror Technology*. (2017, November 24). Retrieved from Symons Mirror Technology: <http://www.symonsmirror.com/>
- TeraView. (2016). *TeraView*. Retrieved from Terahertz Applications: <http://www.teraview.com/applications/index.html>

- Thierry Dauxois, M. P. (1993). Thermodynamics of a nonlinear model for DNA denaturation . *Physica D*, 66, 35-42.
- Thorne, A. P. (1974). *Spectrophysics*. London: Chapman and Hall Ltd.
- Tieleman, L. M. (2012). Force field for classical molecular dynamics. *Biomolecular Simulations* , 924, 197-213.
- Titova , L. V., Ayesheshim , A. K., Golubov, A., Fogen, D. R., Rodriguez-Juarez, R., Hegmann, F., & Kovalchuk, O. (2013). “Intense THz pulses cause H2AX phosphorylation and activate DNA damage response in human skin tissue”. *Biomedical Optics Express*, 559-568.
- Titova, L., Ayesheshim, A., Golubov, A., Rodriguez-Juarez, R., Woycicki, R., Hegmann, F., & Kovalchuk, O. (2013). Intense THz pulses down-regulate genes associated with skin cancer and psoriasis: a new therapeutic avenue? *Scientific Reports*.
- Tsuboi, Y. K. (1967). Intra-red absorption spectra of the single crystals of 1-methyl-thymine, 9-methyladenine and their 1:1 complex. *Spectrochimica Acta Part A: Molecular Spectroscopy*, 23(4), 969-983.
- Wallace, V. P., Fitzgerald, A. J., Pickwell, E., Pye, R. J., Taday, P. F., Flanagan, N., & Ha, T. (2006). Terahertz Pulsed Spectroscopy of Human Basal Cell Carcinoma. *Applied Spectroscopy* , 1127-1133.

- Wallace, V. P., Fitzgerald, A. J., Shankar, S., Flanagan, N., Pye, R., Cluff, J., & Arnone, D. D. (2004). Terahertz pulsed imaging of basal cell carcinoma ex vivo and in vivo. *British Journal of Dermatology*, 424-432.
- Walrafen, G. E. (1990). Raman spectrum of water: transverse and longitudinal acoustic modes below ~ 300 cm^{-1} and optic modes above ~ 300 cm^{-1} . *The journal of physical chemistry*, 2237-2239.
- Watson, J. D., & Crick, F. H. (1953). Molecular structure of nucleic acids: A structure for Deoxyribose Nucleic Acid. *Nature*, 737-738.
- Wood, B. R. (2016). The importance of hydration and DNA conformation in interpreting infrared spectra of cells and tissues. *Chem. Soc. Rev.*, 1980-1998.
- Woods, K. N., & Wiedemann, H. (2004). The relationship between dynamics and structure in the far infrared absorption spectrum of liquid water. *Chemical Physics Letters*, 393(1-3), 159-165.
- Woodward, R. M., Wallace, V. P., Pye, R. J., Cole, B. E., Arnone, D. D., Linfield, E. H., & Pepper, M. (2003). Terahertz Pulse Imaging of ex vivo Basal Cell Carcinoma. *Journal of Investigative Dermatology*, 72-78.
- Zdravković, S., Satarić, M., & Tuszyński, J. (2004). Biophysical Implications of the Peyrard-Bishop-Dauxois Model of DNA Dynamics. *ASP*, 1, 171-181.

Zhang , C. H., Zhao , G. F., Jin, B. B., Hou, Y. Y., & Jia, H. (2012). Terahertz Imaging on Subcutaneous Tissues and Liver Inflamed by Liver Cancer Cells. *Terahertz science and technology*.

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* , 461-470.