



# Exploring the Dynamic Landscape of Small Molecule-DNA Binding

A thesis presented for the degree of  
Doctor of Philosophy  
In the Faculty of Science  
Of the University of Strathclyde

*by*

Lennart Alexander Ingmar Ramakers

Strathclyde, The Department of Physics

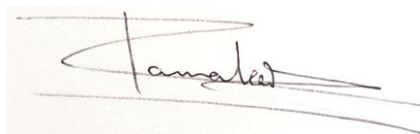
September, 2017

## Declaration of Authenticity and Author's Rights

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

The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by University of Strathclyde Regulation 3.50. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

Signed:

A handwritten signature in black ink, appearing to read 'Tanaka', is written over a horizontal line. The signature is stylized and includes a large, sweeping flourish above the name.

Date: 30<sup>th</sup> of May 2018

## Table of Contents

Declaration of Authenticity and Author's Rights	1
Acknowledgements	5
Abbreviations	6
Abstract	8
Chapter 1: General Introduction	10
1.1 DNA	11
1.2 Minor Groove Hydration and Binding	12
1.3 Melting Mechanisms	15
1.4 IR Spectroscopy	16
1.5 Natural and Non-natural IR Probes	22
1.6 2D-IR of Small Molecule-DNA Binding	26
1.7 References	28
Chapter 2: Experimental Methods and Materials	36
2.1 Material and Sample Preparation	37
2.2 UV-visible Absorbance Spectroscopy	38
2.3 Fluorescence Emission Spectroscopy	38
2.4 FT-IR Spectroscopy	38
2.5 Strathclyde IR Pump-Probe Spectroscopy	39
2.6 Strathclyde and ULTRA 2D-IR Spectrometers	41
2.7 LIFETIME 2D-IR Spectrometer	43
2.8 Reference	46
Chapter 3: Characterisation of the Asymmetric Azide Stretch as a Non-Natural IR Probe: Benzyl Azide	47
3.1 Abstract	48
3.2 Introduction	49
3.3 Results and Discussion	51
3.4 Conclusions	75
3.5 Experimental Methods	76
3.6 Appendix	78

3.7 References	82
Chapter 4: IR Spectroscopy of mixed DNA duplexes containing A-tract and Alternating A-T Sequences	87
4.1 Abstract	88
4.2 Introduction	89
4.3 Results and Discussion	92
4.4 Conclusions	111
4.5 Experimental Methods	112
4.6 Appendix	114
4.7 References	116
Chapter 5: 2D-IR Spectroscopy Reveals Optimised DNA Hoechst 33258 Binding Follows an Induced Fit Model	122
5.1 Abstract	123
5.2 Introduction	124
5.3 Results and Discussion	126
5.4 Conclusions	137
5.5 Experimental Methods	138
5.6 Appendix	139
5.7 References	141
Chapter 6: Probing Minor Groove Binding and Hydration via a Novel Azido-Derivative of Hoechst 33258	146
6.1 Abstract	147
6.2 Introduction	148
6.3 Results and Discussion	150
6.4 Conclusions	170
6.5 Experimental Methods	171
6.6 Addendum	173
6.7 Appendix	176
6.8 References	182



Chapter 7: Impact of Minor Groove Binding on the Thermal Denaturation of DNA Sequences	187
7.1 Abstract	188
7.2 Introduction	189
7.3 Results and Discussion	190
7.4 Conclusions	205
7.5 Experimental Methods	205
7.6 References	206
Chapter 8: General Conclusions	208
8.1 Overall Conclusions	209
8.2 Further Work and Recommendations	212

## Acknowledgements

I would like to start these acknowledgements by thanking my PhD supervisor, **Neil Hunt**, for giving me the opportunity to undertake this interesting and challenging research topic. I would like to also thank **Neil** for his enthusiasm, invaluable insights and support throughout the entire project. Thank you for all of your help, patience with explaining ultrafast optics and for taking the time to read all of my abstracts, reports and most importantly my thesis.

I would also like to thank my 2<sup>nd</sup> supervisor, **Glenn Burley**. His enthusiasm and support helped me to get to grips with the more biological aspects of my PhD and encouraging me to see the biological relevance of my results. I would like to thank my colleagues in the Hunt group, namely **Gordon, Danny, Niall, Robby** and **Lucy** as well as **John** and **Tilemachos**, it was a pleasure to work or relax in a pub with all of you over the past 4 years.

I would also like to thank all of the members of the Molecular Structure and Dynamics Group, at the STFC Central Laser Facility in the Rutherford Appleton Laboratory, for all of their help. In particular, I would like to thank both **Paul Donaldson** and **Greg Greetham** for all of their help with the ULTRA and LIFETIME spectrometers. Without all of their hard work and dedication much of the presented work in this thesis would not have been possible. I am still convinced that both of you are “laser wizards”. Additionally, I would like to thank **Tony Parker**, without whose brilliant insights I would not have been able to get as much out of my PhD as I did.

Another person who I would like to take this opportunity to thank is **Paul Lane**. Even though we only got to work together in the laser lab for a relatively short time, your brilliant sense of humour, help with coding “the Behemoth” and knowledge made every day in the lab a lot more enjoyable. I’m sure you’ll probably not be surprised to learn I still often catch myself thinking about a certain song about San Diego...

**René**, my new brother-in-law, I always enjoyed our lunches. They were a welcome break from my PhD and it was always nice to be able to talk to someone in Dutch during the day. Additionally, your excellent sense of humour helped to keep me from becoming too serious and stressed.

**Lieve Britta**, I would also like to take a moment to thank you for all the support that you gave me throughout my PhD. I appreciate all the time you took to listen to me as well as for sharing your tips and tricks on how to survive life as a PhD student. I still remember the week that we got the opportunity to work together in the same lab as one of the best weeks of my entire PhD. Unfortunately, I am not allowed to pick anyone to be a “paranimf” for me during my viva. However, I just wanted to take this opportunity to say that you would have been my first choice.

Finally, I would like to finish these acknowledgements by taking a moment to thank both my **parents** for their unwavering support throughout the 4 years of my PhD (Dank je wel). I would like to thank them for always taking the time to listen to me talk about my research and also reminding me to make sure I also took some time off from the PhD to enjoy my time in Glasgow.

Thanks for all the help and the fond memories, **Lennart**

## Abbreviations

**DNA** – Deoxyribonucleic acid

**ds-DNA** – Double stranded deoxyribonucleic acid

**ss-DNA** – Single stranded deoxyribonucleic acid

**T** – Thymine DNA base

**A** – Adenine DNA base

**G** – Guanine DNA base

**C** – Cytosine DNA base

**W-C hydrogen bonds** – Watson Crick hydrogen bonds

**H-bonds** – Hydrogen bonds

**T<sub>w</sub>** – Waiting time

**τ** – Coherence time

**β** – Coupling Constant between two vibrational modes

**ESA** – Excited State Absorbance

**μ** – Electrostatic potential of the local environment

**DFT** – Density Functional Theory

**B3-LYP** – Becke, 3-parameter, Lee-Yang-Parr functional

**FT-IR** – Fourier-Transform Infrared Spectroscopy

**PCA** – Principal Component Analysis

**2D-IR** – Two Dimensional Infrared Spectroscopy

**FFCF** – Frequency-frequency correlation function

**OD** – Optical density

**δP** – Hansen Solubility Parameter due to polar interactions

**δH** – Hansen Solubility Parameter due to hydrogen bonding interactions

**A<sub>3</sub>T<sub>3</sub>** – A-tract DNA sequence

**(AT)<sub>3</sub>** – Alternating AT DNA sequence

**AT<sub>25</sub>** – Base-paired T<sub>2</sub> C=O stretch

**G<sub>5</sub>C<sub>5</sub>(-)** – GC C=O antisymmetric stretch

**AT<sub>45</sub>** – Base-paired T<sub>4</sub> C=O stretch

**G<sub>S</sub>C<sub>S</sub>(+)** – GC C=O symmetric stretch

**T<sub>R</sub>** – T ring vibration

**A<sub>R1</sub>T** – Coupled AT ring I vibration

**G<sub>S</sub>C<sub>R</sub>** – C ring mode + G C=O

**G<sub>R</sub>C<sub>S</sub>** – G ring mode + C C=O

**A<sub>R2</sub>T** – Coupled AT ring II vibration

**T<sub>2</sub>** – Carbonyl at the number 2 position in the Thymine base ring

**T<sub>4</sub>** – Carbonyl at the number 4 position in the Thymine base ring

**T<sub>m</sub>** – DNA melting temperature

**H33258** – Hoechst33258 binder

**H-A<sub>3</sub>T<sub>3</sub>** – Complex formed between the A-tract DNA sequence and the Hoechst33258 binder

**H-(AT)<sub>3</sub>** – Complex formed between the alternating AT DNA sequence and the Hoechst33258 binder

**N<sub>3</sub>-bBI** – A *de novo* azido-Hoechst33258 derivative.

**α** – benzimidazole-amide dihedral angle (specifically related to the *de novo* azido-Hoechst33258 derivative)

**N<sub>3</sub>-A<sub>3</sub>T<sub>3</sub>** – Complex formed between the A-tract DNA sequence and the *de novo* azido-Hoechst33258 derivative.

**N<sub>3</sub>-(AT)<sub>3</sub>** – Complex formed between the alternating AT DNA sequence and the *de novo* azido-Hoechst33258 derivative.

## Abstract

*Deoxyribonucleic acid (DNA) is a fundamental component in all living organisms found in Nature. Although both the double-helix structure of this biomolecule and the fact that it stores all the genetic information required by the organism to survive are well understood there are still aspects related to this molecule which are not as clear. Particularly the interactions underpinning the formation of complexes between other molecules, such as proteins, and DNA remain unclear. One of these is the process underlying the formation of small molecule-DNA complexes. Such complexes are known to form via interactions between these small molecules and the DNA minor groove, it has proven to be complicated to develop a set of rational rules for the synthesis of binders to target particular DNA sequences. Such rules are complicated by the complex molecular environment found within the DNA minor groove as well as the role of the spine of hydration found within this groove. Another aspect of the behaviour of DNA which has attracted a lot of attention is related to the mechanism underlying the melting of short DNA sequence. It is important to gain a better understanding of these mechanisms as this process is thought to be important in DNA transcription, replication and repair as well as being important for the application and development of DNA scaffolds. Additionally, it is thought that understanding the impact of binding on this transition could be used to gain further insights into the interactions underpinning these complexes.*

*Here, FT-IR, ultrafast IR pump-probe and two-dimensional infrared (2D-IR) spectroscopy have been applied to investigate the interactions underpinning the formation of small molecule-DNA. Utilising a specifically designed minor groove binding ligand, incorporating an azide moiety as a non-natural IR probe, the questions outlined above were addressed. The spectroscopy of the azides and the DNA bases were initially studied separately, in order to gain the understanding of the spectroscopy of these vibrational modes necessary to maximise the information extracted from the DNA complexes. From the initial investigation of the asymmetric azide stretch of benzyl azide, a model compound, it was found that this mode could be used to determine the electrostatic potential and hydrogen bonding strength of the local environment. In the case of the DNA base modes, it was found that both changes in the structure of the duplex, due to alterations in the sequence and the binding of an archetypical minor groove binder, Hoechst 33258, could be reliably extracted. For the binding of Hoechst 33258, these modes reveal details about the interactions underpinning the formation of complexes with both its target and a sub-optimal DNA sequences. The insights gained were*

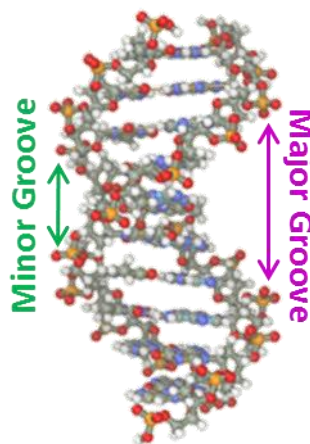
*then brought together to study the complexes formed between DNA and the specially-designed ligand. This allowed both the interactions underlying these complexes and the nature of the water in the minor groove to be explored. Finally, these new spectroscopic methods were then utilised to begin to reveal details of the melting transition of these DNA duplexes and the impact on melting of ligand binding.*

# General Introduction

Chapter 1

## 1.1 DNA

The majority of organisms found in nature utilise deoxyribonucleic acid (DNA) to encode and store the genetic information vital to survival.<sup>1</sup> Encoded using the cytosine (C), guanine (G), thymine (T) and adenine (A) bases, a dimer of the DNA strand and its complementary counterpart form a dimer and adopt an anti-parallel helix, the iconic structure of all DNA oligonucleotides. Although DNA oligonucleotides are known to be able to exist in three distinct structural conformers, often referred to as the *A*, *B* and *Z*-forms. The main difference between these different conformers is that the double helix in the *Z*-form is left-handed whereas for the *A* and *B*-forms the double helix is right-handed. The *A* and *B*-forms represent different structural conformers for DNA as they exhibit different helical parameters (eg. helical diameter). It is known that these different conformers are found under specific conditions where *A*-DNA is found when the duplex is strongly dehydrated, *Z*-form is adopted in the presence of high salt concentrations and finally in the physiological conditions found within living cells the *B*-form is dominant. Due to its dominance in the normal conditions under which all biologically interesting DNA processes occur, the *B*-form is the most widely studied conformer.<sup>2,3</sup> The anti-parallel helix structure of a short DNA oligonucleotide (sequence: d(GGAAATTGTC)<sub>2</sub>) is shown in figure 1.1.

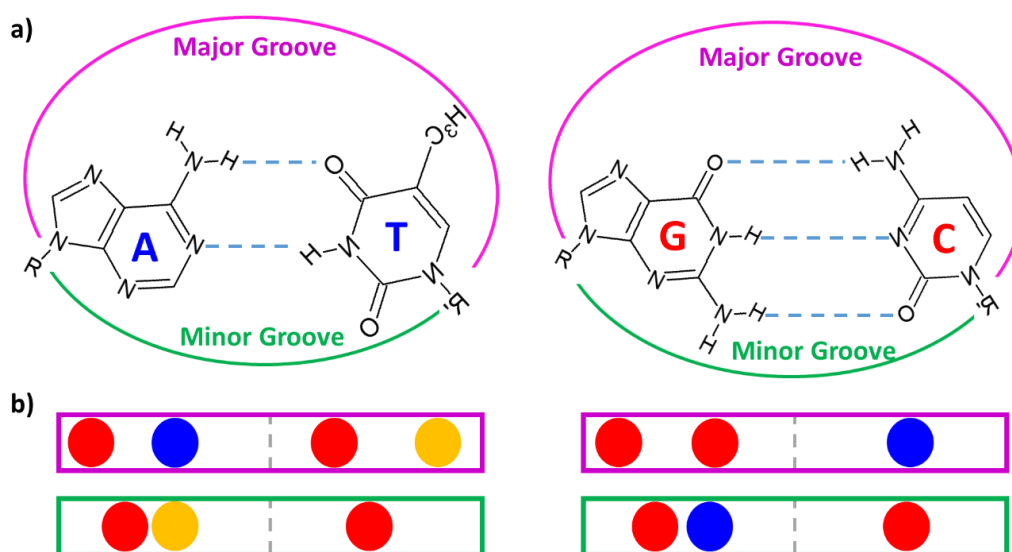


**Figure 1.1:** Diagram of the *B*-conformer of the DNA duplex, with the minor and major groove indicated by green and purple arrows respectively.

The two strands in the *B*-form DNA duplex are held together by the Watson-Crick hydrogen bonds, formed between the individual bases in each of the strands, leading to the formation of the classical double helix. As both strands within the duplex adhere to right-handed helices, this affects the overall topology of the molecule resulting in the formation of two helical grooves. One of these grooves is wide and is often referred to as the major groove



(Fig.1.1, purple arrow), while the other, referred to as the minor groove (Fig.1.1, purple arrow), is relatively narrow. This particular topology of the B-DNA double helix plays an important role in the biological processes involving DNA such as replication and transcription.



**Figure 1.2:** a) The minor and major groove accessibility of the AT and GC base pairs within B-DNA, and b) the possible interactions; H-bond acceptor (red), H-bond donor (blue) and aliphatic hydrogen (yellow), between the AT and GC base pairs via the major (purple) and minor (green) grooves.<sup>2,3</sup>

These important biological processes are often underpinned by sequence-specific interactions between DNA and other molecules, mediated via the major or minor grooves. These interactions are fundamental to many processes in occurring within living cells and can be split into a few distinct groups. Proteins tend to interact via the DNA major groove,<sup>4,5</sup> whereas small molecules tend to interact either via the minor groove or by intercalation between the DNA bases.<sup>6,7</sup> Molecular recognition of double-stranded DNA (dsDNA) sequences by transcription factors for example is essential for the initiation of transcription.<sup>8</sup> This identification of the DNA sequence by other molecules and proteins is thought to be mediated via interactions with the different moieties of the individual bases which are located in the major and minor grooves of the duplex (Fig.1.2).

## 1.2 Minor Groove Hydration and Binding

The DNA minor groove represents a highly complex molecular environment. This complexity of the molecular environment is due to differing contributions that the different parts of the duplex provide. The first of these contributions is from the moieties of the DNA bases located within the minor groove. As shown in Fig.1.2.(b), these moieties are different for the AT and

GC base pairs, however it is noted that they fall into broadly three different categories. These categories are aliphatic hydrogens, H-bonding accepting and H-bond donating groups. Together these different DNA base moieties located within the minor groove mean that the floor of the groove is covered in a highly sequence dependent mixed array of these different contacts. The second contribution to the molecular environment originates from the walls of the DNA minor groove. These walls are formed from the deoxyribose sugar moieties within the DNA duplex and are known to be highly hydrophobic, leading to the walls of the groove being highly hydrophobic. Finally, it has been found that the AT-rich region<sup>9</sup> of the DNA sequence are noted to have a negative potential, whereas stretches rich in GC bases are known to have a more positive potential.<sup>10</sup> Together these features mean that the minor groove has a large range of contrasting contributions from the different parts of the DNA oligonucleotide, as well as being influenced by the exact DNA sequence.

In addition to this basic complexity of molecular environment found within the minor groove, in all biologically relevant environments the DNA duplex is hydrated.<sup>11,12</sup> The hydration of DNA and specifically the minor groove has been studied using a range of different techniques including X-ray crystallography,<sup>13,14</sup> NMR,<sup>15</sup> dielectric relaxation<sup>16</sup> thermodynamics<sup>17,18</sup> and simulations.<sup>19,20,21</sup> This hydration is vital to the function of all biomolecules, which have been noted to undergo a loss of physiological function upon dehydration.<sup>22</sup> The complexity of the molecular environments observed at the exposed surfaces of both DNA and proteins means that the exact role of the hydration water in the function of these molecules has remained difficult to fully understand. The role of this hydration water enhancing the stability of a large variety of proteins has led to the development of the so-called “iceberg model”.<sup>23</sup> In this model a layer of tightly bound water is found at the surface of the protein and within this model this layer supports the structure of the biomolecule without hindering any of the structural changes required for its physiological function. In the case of DNA,<sup>24,25,26</sup> it is thought that this water plays an important role in the structure, conformation, function and the molecular recognition of DNA by other molecules. In DNA both the major and minor grooves are known to contain a number of water molecules. In the major groove these molecules form a monolayer at the bottom of the groove whereas within the minor groove these water molecules are often found to form highly ordered and rigid structures, often referred to as the minor groove spine of hydration.<sup>13,14</sup> The contrasting behaviours of the water present in these two grooves within the DNA duplex has led to a great deal of interest in the exact role of these water molecules. Acoustic and densimetric studies have indicated

that a portion of these hydration waters are released upon the formation of DNA complexes suggesting that they are involved in the molecular recognition of the DNA sequence.<sup>27,28</sup> Recently, due to the increasing interest in molecules which interact with DNA through the minor groove the dynamics and nature of the minor groove spine of hydration have been probed by the incorporation of fluorescent dyes into the DNA duplex.<sup>29</sup> While some of these dyes were unable to report on the nature of the hydration within the minor groove,<sup>29,30</sup> the most successful one, utilising Hoechst 33258 bound in the DNA minor groove, reported reduced dynamics of water bound in the minor groove.<sup>31</sup>

The small molecules which bind into the minor groove of the DNA duplex are often referred to as minor groove binders. Most currently known minor groove binding molecules are based on a range of different molecular motifs.<sup>32</sup> Recently, these minor groove binders have been widely studied as sequence-selective probes of the DNA minor groove as well as the discovery that these molecules exhibit therapeutic activity.<sup>33,34,35</sup> These binders are normally designed in such a way as to complement both the overall helical shape of the groove and as many of the different features within the minor groove as possible.<sup>32</sup> This means that most of these binders contain a combination of aromatic moieties, which can form hydrophobic interactions between the ligand and the walls of the DNA minor grooves, and H-bonding moieties which can form H-bonds to the moieties located on the floor of the minor groove. A few notable examples of these binders are molecules based on the archetypal netropsin,<sup>36</sup> distamycin,<sup>37</sup> polyamides,<sup>38,39</sup> lexitropsins,<sup>40</sup> *bis*-phenylamidine<sup>9</sup> and *bis*-benzimidazole<sup>41</sup> motifs. The minor groove binding molecules based on these different motifs are noted to be highly sequence specific, however so far no specific guidelines for the rational design of minor groove binders to control this sequence specificity have been reported.<sup>32</sup> This is thought to be due to the complex nature of the molecular environment present within the minor groove as well as the large variety of interactions formed between the ligand and the duplex upon the formation of the complex. Finally, the development of these design principles is further complicated by the changes in the DNA hydration associated with the formation of these small molecule-DNA complexes.<sup>27,28</sup> Each of these basic motifs bind to the DNA duplex in a different way with a subset of the binders even forming a 2:1 molecular stoichiometry with the duplex.<sup>38,40</sup> Hoechst33258,<sup>42</sup> a member of the *bis*-benzimidazole family of binders, represents a widely studied archetypal binder. Hence in this thesis we explore the interaction of this binder with both its target A-tract sequence and a sub-optimal alternating AT sequence in order to gain insights into both the formation of these complexes.

Additionally, a novel azido-derivative is studied, with a view to using the incorporation of the non-natural IR probe to study the minor groove and the nature of the hydration within it.

### 1.3 Melting Mechanisms

Despite the exploration of the changes in the structural aspects within the DNA duplex due to the presence of an A-tract via NMR,<sup>43,44,45,46</sup> crystallography,<sup>47,48,49</sup> computational<sup>50,51,52</sup> and theoretical studies,<sup>53,54,55,56</sup> the impact of these sequences on the thermal denaturation mechanism of short lengths of DNA remains relatively unclear. The mechanism by which the strands dissociate from the duplex are important in a range of biological processes as well as being important for the application of DNA scaffolds for nano-devices and self-assembly.<sup>57,58</sup> A range of studies including both computational<sup>59,60</sup> and experimental techniques<sup>57,61,62,63</sup> have been carried out on the denaturation and reassembly of short DNA oligonucleotides and hairpins. These studies highlight the importance of the exact DNA sequence as well as the length of the DNA oligonucleotide on the denaturation mechanism. However, these studies do not consider the changes which occur to the denaturation process when an A-tract is replaced with its alternating AT sequence counterpart.

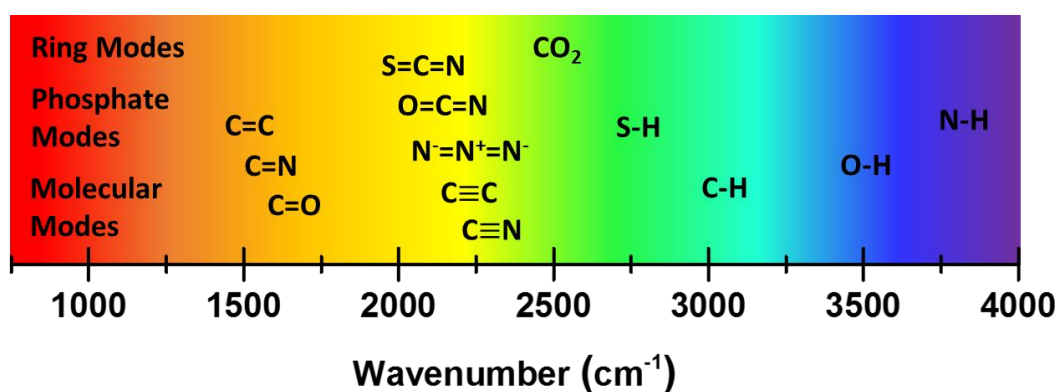
Similarly, very little is known about the impact of the presence of minor groove binders on the mechanism of the melting of these short DNA oligonucleotides. Recently a single-molecule AFM study was carried out which determined that interactions formed between the archetypal minor groove binder H33258 and an A-tract sequence increased the overall duplex rigidity.<sup>64</sup> This increase in rigidity was assigned to the bifurcated H-bonds formed between the ligand and the duplex. However, the exact contributions of the H-bonding interactions and the ligand-duplex hydrophobic contacts, within the complex, to the stabilisation of the melting temperature has remained unclear.<sup>17,18,41</sup>

Two dimensional infrared spectroscopy (2D-IR)<sup>72,73,74</sup> has proved to be a versatile methodology to probe both dynamics and structural aspects of the DNA macromolecule due to its high temporal resolution and ability to probe the complex intermolecular couplings within DNA.<sup>65,66,67,68</sup> Recently 2D-IR has been used to unravel the melting mechanisms and understand the impact of the melting on the structural dynamics of short DNA oligonucleotides.<sup>69,70</sup> One of these studies has led to valuable insights into the impact of the alterations in the position of the GC bases within the sequence on the overall melting mechanism.<sup>69</sup> While the other demonstrated that melting induced more rapid spectral

diffusion, as well as indicating the water in the spine of hydration is dynamically restricted in ds-DNA.<sup>70</sup>

## 1.4 IR Spectroscopy<sup>71</sup>

Electromagnetic radiation can be used to probe and explore the different energy levels present within molecules. Specifically, the electronic energy levels can be explored utilising UV-visible radiation, vibrational energy levels via infrared (IR) radiation, within the mid-IR frequency range, and rotational energy levels via microwaves. All of these different spectroscopic methods yield useful information about the molecular system being studied, with IR spectroscopy revealing the different chemical moieties within the structure, as these vibrate at different frequencies. An overview of the mid-IR frequency range and the approximate positions of a range of common groups and moieties is summarised in figure 1.3.



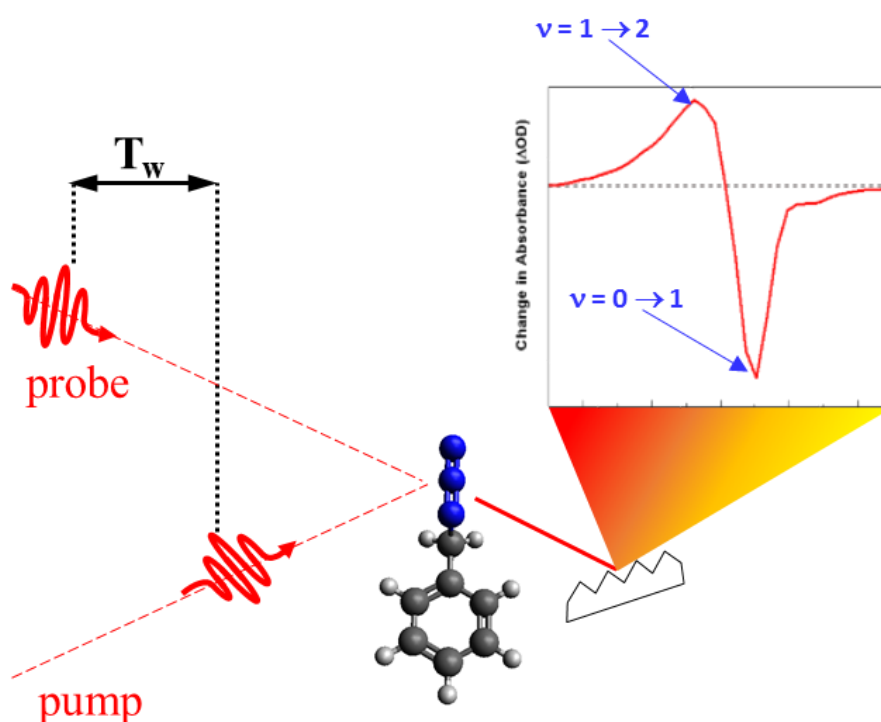
**Figure 1.3:** Illustration of the Mid-IR range from the low frequency (red) to the high frequency (purple) end with well-known vibrational modes positioned at their approximate wavenumber. The colours are included to indicate the high and low frequency ends of the mid-IR spectral range.

At the low frequency end of the IR spectrum (800 – 1500 cm<sup>-1</sup>) there are a number of low energy vibrational modes associated with molecule wide warping of the molecular structure (molecular modes), phosphate stretches, as well as a plethora of different ring modes and C-H bends. This results in this low energy part of the spectrum often being referred to as the fingerprint region, as these different low energy global modes tend to be broadly similar for molecules within the same molecular family. The (1500 – 1800 cm<sup>-1</sup>) and the (2550 – 4000 cm<sup>-1</sup>) spectral windows contain the stretch modes associated with double bonds and bonds to hydrogen atoms, respectively. These vibrational modes are commonly found in a wide range of chemical systems and solvents, including many naturally occurring IR modes seen in

biological systems. Finally there is a narrow spectral window ( $1800 - 2550 \text{ cm}^{-1}$ ) which is associated with the stretching modes of cyanates, thiocyanates, nitriles, alkynes and azides.

Linear IR spectra for molecular systems are often obtained by measuring the absorbance of the sample at a range of different mid-IR frequencies and plotting the results as a function of the wavenumber. These spectra yield a series of separate distinct peaks which can then be used to identify the moieties present in the molecular system. This ability to uniquely identify different chemical moieties means that IR spectroscopy was originally used, alongside other spectroscopy techniques, in organic chemistry to quantify the products of synthetic work.

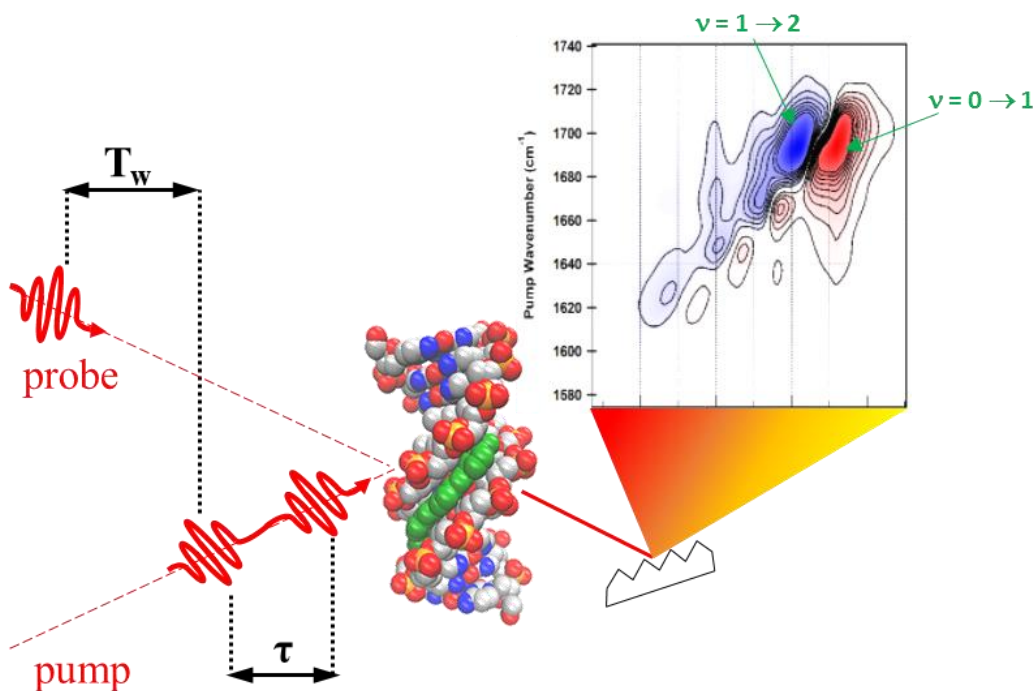
The development of ultrafast laser, capable of producing sub-femtosecond pulses, has led to the development of a range of different ultrafast pump-probe spectroscopies including IR pump-probe methods. Based on the original principle first developed in the 1980's a wide range of pump-probe spectroscopic methods were developed, allowing the ultrafast dynamics of molecular systems to be probed with a sub-picosecond temporal resolution. Among these different pump-probe methods, IR pump-probe allows the vibrational dynamics of different modes to be explored. The methodology used to carry out these measurements is shown in figure 1.4.



**Figure 1.4:** Schematic illustration of an Ultrafast IR pump-probe measurement.  $T_w$ , the waiting time, is the time between the arrival of the pump and probe pulses at the sample.

In ultrafast IR pump-probe spectroscopy, initially a sub-picosecond IR pump pulse is allowed to interact with the sample being studied. This interaction results in a  $v = 0 \rightarrow 1$  transition, exciting a sizable portion of the molecules and leading to an increase in the 1<sup>st</sup> excited state of the vibrational modes, with frequencies within the bandwidth of the pump pulse. After a variable waiting time ( $T_w$ ) the probe pulse is allowed to interact with the sample. The interaction between the probe pulse and the sample results in either a  $v = 1 \rightarrow 2$  transition or a  $v = 1 \rightarrow 0$  transition of the molecules. Chopping the pump pulse and measuring the probe pulse with and without the initial excitation of the vibrational modes allows the pump-probe response of the vibrational modes to be measured. As the time ( $T_w$ ) between the arrival of the pump and the probe pulses at the sample increases the molecules initially excited by the pump pulse undergo vibrational relaxation resulting in a decrease in the population of the 1<sup>st</sup> excited state of the vibrational mode and leading to a decrease in the intensity of the pump-probe response of the sample. Scanning the variable waiting time ( $T_w$ ) between the pump and probe pulses allows the vibrational lifetime of the vibrational modes to be determined.

In addition to ultrafast IR pump-probe spectroscopy, the ability to produce sub-picosecond mid-IR pulses lead to the development of two-dimensional infrared spectroscopy (2D-IR).<sup>72,73,74</sup> This vibrational analogue of 2D-NMR spectroscopy can be thought of as an extension of the pump-probe technique where the signal is dispersed over an additional vibrational pump axis. This can be achieved via two separate methodologies. One of these methodologies, initially developed by Hochstrasser,<sup>75</sup> involves utilising a simple pump-probe spectrometer, where the pump pulse is passed through a Fabry-Pérot interferometer to produce a narrow band pulse, scanning the centre frequency of the pump-pulse and plotting the measured pump-probe spectrum as a function of the pump frequency produced a 2D-IR spectrum. However, this approach leads to a temporal resolution in the picosecond range meaning any dynamics occurring in the sub-picosecond time domain are lost. Alternatively, and more commonly is an approach referred to as Fourier-Transform 2D-IR (FT-2D-IR). The pulse sequence utilised in such a FT-2D-IR spectroscopy is shown in figure 1.5.



**Figure 1.5:** Schematic illustration of a FT-2D-IR measurement.

In two-dimensional IR (2D-IR) spectroscopy the sample is irradiated with a sequence of three separate IR pulses. This pulse sequence is illustrated in figure 1.5. Basically this sequence consists of three separate pulses where the first and second pulse are separated by a time  $\tau$  and the second and third pulse are separated by a time  $T_w$ . Finally, it is noted that the time between the arrival of the final pulse at the sample and the production of the sample are separated by a time  $t$ .

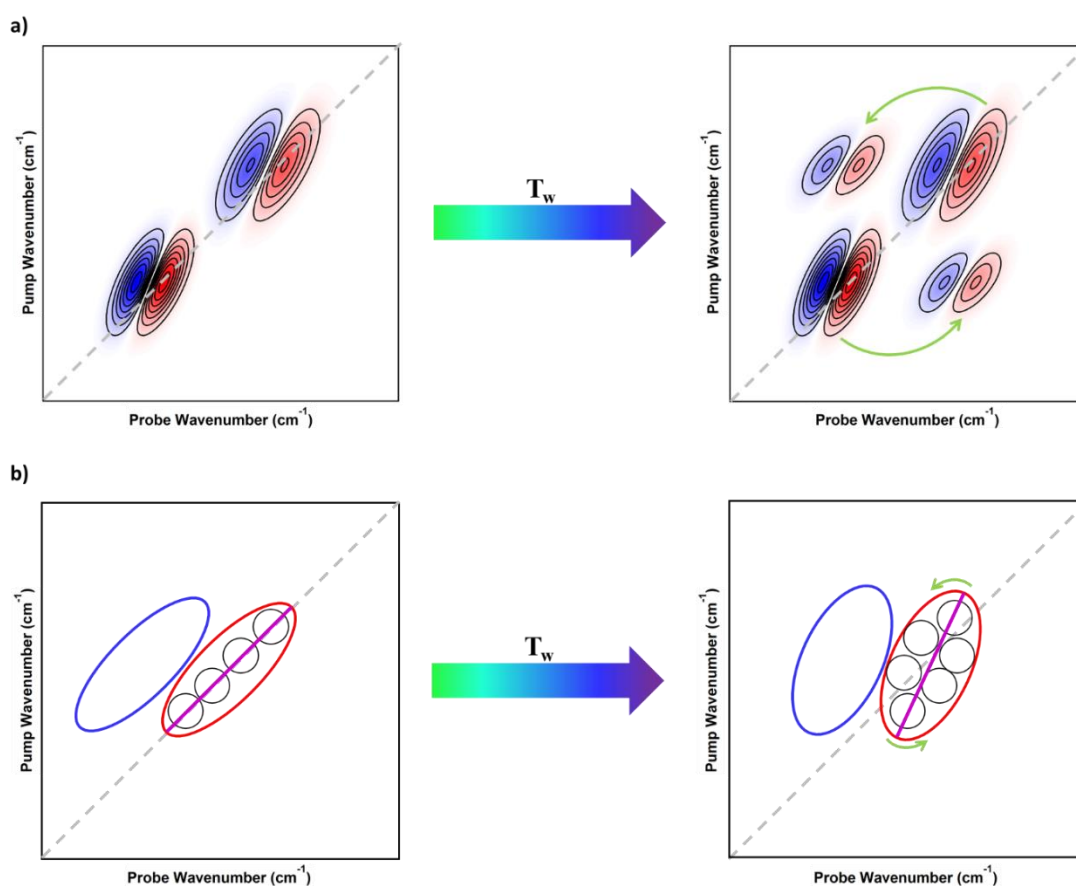
The interactions between the first pulse to arrive at the sample and the molecules within the sample, which are noted to be at thermodynamic equilibrium before this first field-matter interaction, results in the molecular system evolving to a coherence. This coherence oscillates during the time delay  $\tau$  between the arrival of the first and second pulse at the sample. The arrival of the second pulse results in another field-matter interaction taking place and brings the system to a population on either the first excited or the ground vibrational state of the vibrational mode being studied. Finally, after a waiting time of  $T_w$  the third pulse arrives at the sample and the interaction between the sample and the pulse drives the system into a second coherence, resulting in the production of a 2D-IR signal ( $S_{2D}(t, T_w, \tau)$ ), often referred to as a photon echo, after a further time,  $t$ . This signal produced by the interaction of this pulse sequence with the sample is noted to be generated by the third-order nonlinear polarisation of the sample and is produced in the phase-matching



direction given by  $\mathbf{k}_s = -\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3$ . For the pseudo-pump probe geometry of the 2D-IR spectrometers used here, this means that the produced signal is collinear with the third pulse, resulting in the signal being superimposed onto the residual light of the third pulse in the original sequence. This is referred to as a self-heterodyned signal and allows phase information about the signal to be recovered within the detector via a spectral interferometry method.<sup>76</sup> Finally the 2D-IR spectrum ( $S_{2D}(\omega_t, T_w, \omega_\tau)$ ) is generated by performing two Fourier transforms, one with respect to  $t$  and the other with respect to  $\tau$ . The Fourier transform with respect to  $t$  is noted to be carried out by the diffraction grating within the detectors used in the 2D-IR spectrometers and the Fourier transform with respect to  $\tau$  is carried out numerically during data analysis.

In a 2D-IR spectrum, each peak observed in the FT-IR spectrum of the sample, gives rise to a negative feature (red) located on the 2D-IR spectrum diagonal. These are assigned to the respective  $\nu = 0 \rightarrow 1$  transitions, each with an accompanying, positive (blue),  $\nu = 1 \rightarrow 2$  peak shifted to lower probe frequency by the anharmonicity of the bonding potential. Any peaks observed in the off-diagonal region between these diagonal peaks on the spectrum indicate coupling or chemical exchange. Essentially the resulting spectrum can be thought of as a 2D correlation map of the vibrational frequency of the modes within the samples. This means that the utilization of 2D-IR allows vibrational coupling, chemical exchange and spectral diffusion of the vibrational modes within the system to be explored in addition to the 2D-vibrational relaxation of the vibrations.

The presence of coupling or the chemical exchange within the studied molecular system results in the appearance of off-diagonal features between the two diagonal peaks in the 2D-IR spectrum. In the case of coupling these cross-peaks are not waiting-time dependent meaning that the cross-peaks arising from coupling are always present in the 2D-IR spectrum. In contrast to this cross-peaks originating from chemical exchange appear as the waiting time between the arrival of the pump pulses and the probe pulses at the sample increases, as shown in Fig.1.6.(a).



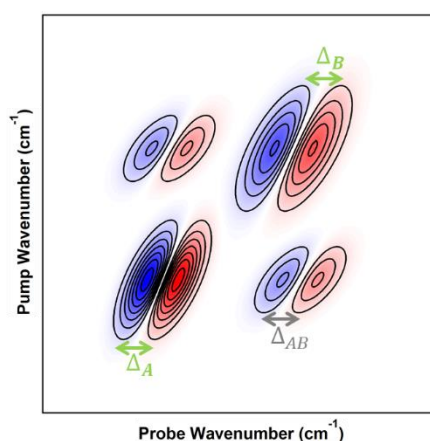
**Figure 1.6:** Illustration of changes observed in 2D-IR spectra as the waiting time increases, a) the appearance of cross-peaks due to coupling or chemical exchange and b) spectral diffusion.

Considering 2D-IR spectra as a 2D vibrational correlation map, changes within the spectra can be associated with changes in the molecular system being studied. The appearance of cross-peaks in the off-diagonal region can then be explained by a transfer of vibrational energy between two vibrational modes due to the presence of H-bonding, dipole-dipole interactions or coupling within the molecular system. Alternatively, the appearance of these cross-peaks can be explained by the chemical exchange of the molecules between different distinct local environments. In addition to the appearance of features in the off-diagonal region as the waiting time increases it is also noted that there are often changes to the appearance of the line-shapes of the diagonal features.

It is often found that the peaks corresponding to the  $\nu = 0 \rightarrow 1$  transition (red) are initially strongly elongated along the spectrum diagonal. As the waiting time increases and the interactions between the molecular and its local environment rearrange, the correlation between the pump and probe frequencies decreases. This change in the correlation results in the line-shape becoming less diagonally elongated. This process is referred to as spectral

diffusion and it can be quantified by a range of different metrics including the slope of the nodal line, the eccentricity and the central line slope (Fig.1.6.(b), purple line)<sup>77,78</sup> reflecting the dynamics associated with the frequency-frequency correlation function (FFCF) of the vibrational mode.

Finally, the 2D-IR spectrum of the molecule can also be used to determine both the projection angles between the transition dipole moments and coupling constants between separate vibrational modes. The projection angle between the transition dipole moments can be extracted by comparing the intensities of the cross-peaks as the polarisation of the pump and probe pulses are changed from a parallel to a perpendicular polarisation configuration.



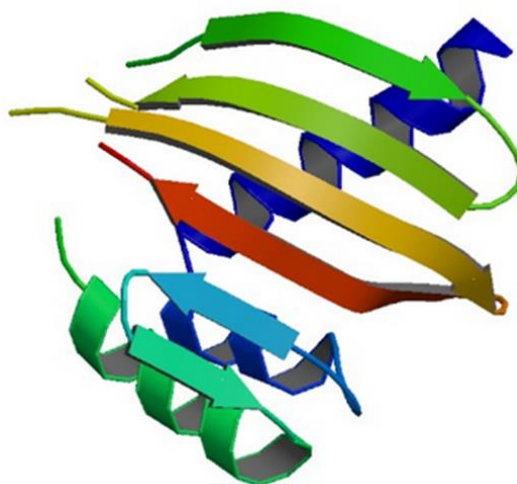
**Figure 1.7:** Model 2D-IR spectrum indicating the parameters required to calculate the coupling constant ( $\beta$ ) between two vibrational modes.

The coupling constant between two separate vibrational modes can be extracted by determining the anharmonicities of both modes and the associated cross-peak, as indicated in figure 1.7. Together all of these aspects indicate that 2D-IR spectroscopy is information rich with respect to both molecular structure and ultrafast dynamics, demonstrating the versatility of this technique to study a wide range of different molecular systems. Often these techniques are implemented to study both the naturally occurring IR modes within systems as well as a variety of different non-natural probes which can be incorporated into systems to gain new insights. Recently these techniques and IR probes are being used to gain a new understanding into biomolecule and their physiological functions.

### 1.5 Natural and Non-Natural IR Probes

In order to be able to study specific regions of interesting biological molecules using spectroscopic methods, probes are often implemented to extract the maximum amount of information from these methods.<sup>79,80</sup> Sometimes these probes are naturally occurring groups

in the molecule while others are non-naturally occurring functional groups. For example in the IR spectroscopy of proteins or peptides, these naturally occurring groups could be the amide links of the protein<sup>81,82</sup> or the thiol group of cysteine.<sup>83</sup> Such naturally occurring IR probes have been used to study a variety of different aspects of proteins and peptides including investigating their secondary structures, solvation dynamics, local electrostatics and vibration couplings present within these biomolecules.<sup>82,84</sup>



**Figure 1.8:** Diagram illustrating the different aspects of protein secondary structures which can be explored and probed via the naturally occurring amide I mode.<sup>85</sup>

In addition to these naturally occurring IR probes, the incorporation of non-natural probes can also be used to study a variety of different aspects of such systems. Examples of non-natural IR probes are the incorporation of azido<sup>86</sup> (label located at  $\sim 2110\text{ cm}^{-1}$ ) or cyano<sup>87,88</sup> (label located at  $\sim 2240\text{ cm}^{-1}$ ) amino acid derivatives into the peptide or protein. Each of these methods has distinct advantages and disadvantages associated with them. For the naturally occurring probes the advantages include the fact that they do not perturb the structure or function of the system being studied. However their vibrational bands are often found in congested regions of the spectrum making it difficult to extract environmental information from these vibrations.<sup>79</sup> The non-natural probes have the advantage that they are highly sensitive to the local environment, intense and in an uncongested spectral window.<sup>79</sup> However their introduction can perturb the physical system being studied and it can be difficult to successfully introduce these probes.

The guiding principles in the design and utilisation of non-natural probes are that these probes must be as intense as possible and cause minimal perturbation to the physical

system.<sup>89</sup> Popular non-natural IR probes include nitriles<sup>90</sup>, cyanates<sup>91</sup>, thiocyanates<sup>92</sup> and azides.<sup>93</sup> Such probe moieties have been successfully used in recent biological studies,<sup>94,95,96</sup> providing new insights into the studied systems. These studies use standard linear IR techniques, such as FT-IR, and non-linear techniques, such as femtosecond IR pump-probe and 2D-IR.

Recently a multitude of studies into using the azide ion and organic azides as non-natural probes have been carried out.<sup>97,98,99</sup> The studies have focused on these azides probes as the asymmetric azide stretch is known to occur at approximately 2100 cm<sup>-1</sup> for both the organic and ionic azides. None of the natural vibrations of proteins, peptides or DNA occupy this region of the IR spectrum and water only has a small combination band in this region.<sup>79</sup> This means that the asymmetric azide vibrations occupy a transparent window in many biological systems. The fact that the asymmetric azide vibration has a large extinction coefficient and a high sensitivity to its local electrostatic environment means that it is ideally suited to being used as a non-natural IR probe.<sup>100</sup>

The development of azido amino acid derivatives, such as  $\beta$ -azidoalanine, has led to a plethora of IR studies being carried out with azido probes.<sup>101,102</sup> Two general methods of integrating these probes into the biological systems have been developed. One method involves substituting natural subunits with their azide derivatives. For proteins azido amino acids have been developed and for DNA azido nucleic acids have been produced. In one recent study  $\beta$ -azidoalanine was incorporated into amyloid peptide and it was shown to be an excellent reporter of the local environment inside the aggregate formed by these peptides.<sup>103</sup> Additionally, this study showed that this probe blue-shifts as its local environment becomes increasingly protic. This shows that this azido amino acid derivative could be potentially incorporated into a protein and be used to observe the folding and unfolding of the protein in question. Another noteworthy study implementing this methodology of probe introduction was carried out by Ye *et al.*<sup>104</sup> In this study a transmembrane receptor protein called rhodopsin which detects light was studied using *p*-azido-*L*-phenylalanine as an IR probe. This receptor protein has been extensively studied as it translates the external light stimulus into a chemical message the cells read and pass on to nerve cells. Several studies on this receptor have revealed that photo-excitation induces a series of structural changes causing the receptor to pass through several spectroscopically distinct intermediates.<sup>105</sup> It has been shown that these structural changes are the concerted

movement of a bundle of transmembrane  $\alpha$ -helices which create a pocket on the inside of the cell allowing a signalling protein to bind to the receptor.<sup>106,107</sup> However the temporal sequence in which this concerted movement occurs remained unclear. In their study Ye *et al.*<sup>104</sup> introduced their probe at several different positions in the bundle of  $\alpha$ -helices. Studying these probes using standard and time resolved FT-IR the temporal sequence of the structural changes was deduced. This study demonstrates the versatility and effectiveness of using azido IR probes to gain a deeper understanding of how biological molecules carry out their specific function.

The other method of introducing probes involves synthesising azido derivatives of the ligand that bind to the protein in question. This methodology has the potential to allow protein-ligand interactions to be studied. This methodology has already been successfully implemented in several studies.<sup>108</sup> In a study by Bloem *et al.*<sup>109</sup> the binding of a peptide sequence to a protein, known as PDZ2 was investigated using such azido derivatives of the peptide sequence. By using 2DIR and incorporating azidohomoalanine at several different positions in the peptide ligand they were able to explore the local environment at several different positions in the active site of PDZ2. Another interesting study which utilised this methodology was recently carried out by Dutta *et al.*<sup>110</sup> In this work an azide derivative of NAD<sup>+</sup> was developed. This is especially interesting as NAD<sup>+</sup> is known to act as a cofactor for a large number of different proteins and so a single probe can be used to measure many different systems.<sup>110</sup> The probe was characterised using both pump-probe and 2DIR methods and it was found to be an excellent probe of the local dynamics around the azide probe. However during this study it was noted that its characterisation was complicated by the presence of accidental Fermi resonances.

Fermi resonances occur as a result of coupling between a fundamental mode and a combination or overtone band of a molecule which occurs when the energy of the modes are similar to each other.<sup>111</sup> Azides are known to be especially vulnerable to accidental Fermi resonances as the transparent region of the spectra that the azide peak occupies contains many such overtone and combination bands. This is due to the large number of molecular vibrations found in the fingerprint region of the IR spectrum.<sup>112</sup> The presence of Fermi resonance greatly complicates the application of these vibrations as probes of the local environment as there are many more parameters to consider when characterising the probes, such as the different environments modulating the strength of the intramolecular

coupling. Recent Lipkin *et al.*<sup>112</sup> carried out a study aimed at eliminating the Fermi resonance by adjusting the energy of the asymmetric azide stretch or the combination/overtone bands involved in the Fermi resonance. The adjustments in the energy of these modes were altered by introducing isotopic labels. This work was carried out on 3-azidopyridine, a probe molecule which has been found to be sensitive to binding interactions between the environment and the nitrogen in the pyridine ring.<sup>113</sup> In this study it was found that isotopic labelling of the azide group could greatly reduce or remove the Fermi resonances. The most effective of these labelling techniques was changing the central nitrogen or final two nitrogen atoms to nitrogen-15 atoms.

In the current study, we are interested in using both the naturally occurring DNA base modes as well as a non-natural azide probe to gain insights into DNA and the complexes it forms with small molecules.

### 1.6 2D-IR of Small Molecule-DNA Binding

The aim of this thesis is to gain a deeper understanding of the mechanism underpinning the formation of such DNA complexes. In addition to this primary aim, this thesis also aims to explore the mechanism underlying the melting transition of two different short DNA oligonucleotides and the impact of binding on this melting mechanism. Here, to achieve these aims 2D-IR is utilised on both the naturally occurring DNA base modes as well as a non-natural azide probe, incorporated into a novel minor groove binding ligand.

In the first experimental chapter, IR spectroscopic techniques are used to understand how the asymmetric azide stretch, of an archetypal aromatic azide, is altered as its local environment changes. These changes in the local environment around the azide probe are achieved by altering the solvent used to prepare the benzyl azide samples. Altering the solvent in a systematic way allowed the changes in the asymmetric azide to be used to generate calibration curves for different aspects of the local environment. This characterisation is carried out to allow such an azide moiety to be incorporated into a novel DNA minor groove binder, allowing the environment within the minor groove to be studied and gain an insight into the biological water within the minor groove to be studied.

Following on from this, in the second experimental chapter contained within this work the DNA base modes of two short DNA oligonucleotides are studied and compared. This chapter functions as an initial exploration of the information which can be extracted from these naturally occurring base modes. The utilisation of 2D-IR difference spectroscopy allows the

structural alterations within the DNA duplex due to a change in the central AT sequence of the duplex from an A-tract to an alternating AT motif. The results of the changes in the 2D-IR spectra upon these changes was determined to be due to a loss in a highly-ordered region of the AT propeller twists and due to a loss of structure in the minor groove spine of hydration. Fully consistent with the difference observed by NMR spectroscopy carried out on similar DNA sequences. Finally, in this chapter the application of a two state model and 2D-IR difference spectra to the FT-IR and 2D-IR spectra observed for each sequence as they melt has allowed details of the mechanism underlying this transition to be explored.

The results from these first two chapters are combined and utilised to study the complexes formed between the archetypal *bis*-benzimidazole minor groove binder, Hoechst 33258, and a novel azido-derivative of this ligand. This has allowed differing aspects of the complexes formed between *bis*-benzimidazole ligands and DNA oligonucleotides containing either an A-tract or an alternating AT motif to be explored. The insights previously gained from the unbound DNA chapter allow the mechanisms underlying the binding interactions to be explored via changes in the naturally occurring DNA base modes. These changes reveal that these *bis*-benzimidazole ligands bind to their target A-tract sequence via an induced fit mechanism whereas binding to the sub-optimal alternating AT motif was found to follow a rigid body type interaction for Hoechst 33258 and a non-specific interaction for its azido-derivative. In addition to highlighting the differences in the binding interactions for the target and sub-optimal sequences, it also reveals the additional benefit of the induced fit interaction by demonstrating that the addition of a hydrophilic azide moiety to the ligand does not alter the observed binding interaction for the A-tract, but is observed to drastically alter the rigid body interaction seen in the sub-optimal binding scenario. The study of the asymmetric azide stretch, utilising the information gained from the first experimental chapter presented in this thesis, reveals the nature of the water within the DNA minor groove. These results indicate that the dynamics of the water within the minor groove are significantly slowed when compared to the dynamics in solution. However, it is noted that these results do not suggest that these water molecules are frozen in place as is often expected for such biological water molecules.

In the final experimental chapter, an attempt is made to use FT-IR and 2D-IR spectroscopy to gain an insight into the impact of binding on the melting mechanism of the two DNA sequences studied in this thesis. Through a combination of the techniques employed in the



previous experimental chapters the FT-IR and 2D-IR spectra of these DNA complexes as the temperature of the sample increases suggests that binding does not alter the overall melting mechanism of the sequences. It is noted that the only impact appears to be an increase in the temperature at which the sequences starts to undergo melting. Interestingly the results of this chapter also seem to suggest that the majority of the stabilisation of the duplex upon binding is as a result of the hydrophobic contacts formed between the ligand and the side of the minor groove, fully consistent with previous thermodynamic studies of such DNA complexes.

## 1.7 References

---

- (1) Watson, J. D.; Crick, F. H. C. Molecular Structure of Nucleic Acids, *Nature*, **1953**, 4356, 737.
- (2) Goodsell, D. (2001). Molecule of the Month: DNA. DOI: 10.2210/rcsb\_pdb/mom\_2001\_11.
- (3) Wolfram Saenger (1994) *Principles of Nucleic Acid Structure* (Springer-Verlag, New York).
- (4) Pavletich, N. P.; Pabo, C. O. Crystal Structure of a Five-Finger GLI-DNA Complex: New Perspectives on Zinc Fingers. *Science*, **1993**, 261, 1701-1707.
- (5) Mamoon, N. M.; Song, Y.; Wellman, S. E. Histone H1<sup>0</sup> and its Carboxyl-Terminal Domain Bind in the Major Groove of DNA, *Biochemistry*, **2002**, 41, 9222-9228.
- (6) Liu, H.-K.; Sadler, P. J. Metal Complexes as DNA Intercalators, *Accounts of Chemical Research*, **2011**, 44, 349-359.
- (7) Bailly, C.; Chaires, J. B. Sequence-Specific DNA Minor Groove Binders. Design and Synthesis of Netropsin and Distamycin Analogues. *Bioconjugate Chem.* **1998**, 9, 513-538.
- (8) Richard E. Dickerson (1983) The DNA Helix and How it is Read. *Scientific American* **249** (December), pp. 94-111.
- (9) Edwards, K. J.; Brown, D. G., Spink, N.; Skelly, J. V.; Neidle, S.; Molecular structure of the B-DNA dodecamer d(CGCAAATTTGCG)<sub>2</sub>. An examination of propeller twist and minor-groove water structure at 2.2 Å resolution., *J. Mol. Biol.*, **1992**, 226, 1161-1173.
- (10) Heinemann, U.; Alings, C.; Crystallographic study of one turn of G/C-rich B-DNA., *J. Mol. Biol.*, **1989**, 210, 369-381.
- (11) Zhong, D.; Pal, S. K.; Zewail, A. H.; Biological water: A critique, *Chemical Physics Letters*, **2011**, 503, 1-11.
- (12) Pal, S. K.; Zewail, A. H.; Dynamics of Water in Biological Recognition, *Chem. Rev.*, **2004**, 104, 2099-2123.
- (13) Drew, H. R.; Dickerson, R. E.; Structure of a B-DNA dodecamer. III. Geometry of hydration., *J. Mol. Biol.*, **1981**, 151, 535-556.

- 
- (14) Liepinsh, E.; Otting, G.; Wüthrich, K.; NMR observation of individual molecular of hydration water bound to DNA duplexes: direct evidence for a spine of hydration water present in aqueous solution., *Nucleic Acids Research*, **1992**, *20*, 6549-6553.
- (15) Halle, B.; Denisov, V. P.; Water and monovalent ions in the minor groove of B-DNA oligonucleotides as seen by NMR. *Biopolymers*, **1998**, *48*, 210-233.
- (16) Umehara, T.; Kuwabara, S.; Mashimo, S.; Yagihara, S.; Dielectric study on hydration of B-, A-, and Z-DNA., *Biopolymers*, **1990**, *30*, 649-656.
- (17) Haq, I. Part II: The Thermodynamics of Drug-Bipolymer Interaction, Thermodynamics of Drug-DNA Interactions. *Archives of Biochemistry and Biophysics*, **2002**, *403*, 1-15.
- (18) Haq, I.; Ladburry, J. E.; Chowdhry, B. Z.; Jenkins, T. C.; Chairs, J. B. Specific Binding of Hoechst33258 to the d(CGCAAATTTGCG)<sub>2</sub> Duplex: Calorimetric and Spectroscopic Studies. *J. Mol. Biol.* **1997**, *271*, 244-257.
- (19) Schneider, B.; Cohen, D. M.; Schleifer, L.; Srinivasan, A. R.; Olson, W. K.; Berman, H. M.; A systematic method for studying the spatial distribution of water molecules around nucleic acid bases., *J. Biophys.*, **1993**, *65*, 2291-2303.
- (20) Feig, M.; Pettitt, B. M.; A molecular simulation picture of DNA hydration around A- and B-DNA., *Biopolymers*, **1998**, *48*, 199-209.
- (21) Duan, Y.; Wilkosz, P.; Crowley, M.; Rosenberg, J. M.; Molecular dynamics simulation study of DNA dodecamer d(CGCGAATTCGCG) in solution: conformation and hydration., *J. Mol. Biol.*, **1997**, *272*, 553-572.
- (22) Rupley, J. A.; Careri, G.; Protein hydration and function., *Adv. Protein Chem.*, **1991**, *41*, 37-172.
- (23) Kauzmann, W.; Some Factors in the Interpretation of Protein Denaturation., *Adv. Protein Chem.*, **1959**, *14*, 1-63.
- (24) Pal, S. K.; Zhao, L.; Xia, T.; Zewail, A. H.; Site- and sequence-selective ultrafast hydration of DNA, *Proc. Natl. Acad. Sci. USA*, **2003**, *26*, 13746-13751.
- (25) Pal, S. K.; Zewail, A. H.; Dynamics of Water in Biological Recognition, *Chem. Rev.*, **2004**, *104*, 2099-2123.
- (26) Zhong, D.; Pal, S. K.; Zewail, A. H.; Biological water: A critique, *Chemical Physics Letters*, **2011**, *503*, 1-11.
- (27) Chalikian, T. V.; Plum, G. E.; Sarvazyan, A. P.; Breslauer, K. J.; Influence of drug binding on DNA hydration: acoustic and densimetric characterization of netropsin binding to the poly(dAdT).poly(dAdT) and poly(dA).poly(dT) duplexes and the poly(dT).poly(dA).poly(dT) triplex at 25 degrees C. *Biochemistry*, **1994**, *33*, 8629-8640.
- (28) Chalikian, T. V.; Völker, J.; Srinivasan, A. R.; Olson, W. K.; Breslauer, K. J.; The hydration of nucleic acid duplexes as assessed by a combination of volumetric and structural techniques., *Biopolymers*, **1999**, *5*, 459-471.
- (29) Brauns, E. B.; Madaras, M. L.; Coleman, R. S.; Murphy, C. J.; Berg, M. A.; Measurement of Local DNA reorganisation on the Picosecond and Nanosecond Time Scales, *J. Am. Chem. Soc.*, **1999**, *121*, 11644-11649.

- 
- (30) Hess, S.; Davis, W. B.; Voityuk, A. A.; Rosch, N.; Michel-Beyerle, M. E.; Ernstring, N. P.; Kovalenko, S. A.; Lustres, J. L. P.; Excited-State Photophysics of an Acridine Derivative Selectively Intercalated in Duplex DNA, *Chem. Phys. Chem.*, **2002**, *3*, 452-455.
- (31) Pal, S. K.; Zhao, L.; Zewail, A. H.; Water at DNA surfaces: Ultrafast dynamics in minor groove recognition, *Proc. Natl. Acad. Sci. U.S.A.*, **2003**, *100*, 8113-8118.
- (32) Neidle, S.; DNA minor-groove recognition by small molecules, *Nat. Prod. Rep.*, **2001**, *18*, 291-309.
- (33) White, C. M.; Heidenreich, O.; Nordheim, A.; Beerman, T. A. Evaluation of the Effectiveness of DNA-Binding Drugs to Inhibit Transcription Using the c-fos Serum Response Element as a Target, *Biochemistry*, **2000**, *39*, 12262-12273.
- (34) Zhang, X.; Kiechle, F. Hoechst33342-Induced Apoptosis is Associated with Decreased Immunoreactive Topoisomerase I and Topoisomerase I-DNA Complex Formation, *Annals of Clinical & Laboratory Science*, **2001**, *31*, 187-198.0
- (35) Bellowini, M.; Moncollin, V.; D'Incalci, M.; Mongelli, N.; Mantovani, R. Distamycin A and Tallimustine Inhibit TBP Binding and Basal *In Vitro* Transcription, *Nucleic Acids Research*, **1995**, *23*, 1657-1663.
- (36) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E.; Binding of an antitumor drug to DNA, Netropsin and C-G-C-G-A-A-T-T-BrC-G-C-G., *J. Mol. Biol.*, **1985**, *183*, 553-563.
- (37) Coll, M.; Frederick, C. A.; Wang, A. H. J.; Rich, A.; A bifurcated hydrogen-bonded conformation in the d(AT) base pairs of the DNA dodecamer d(CGCAAATTTGCG) and its complex with distamycin, *Proc. Natl. Acad. Sci. USA*, **1987**, *84*, 8385.
- (38) Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Dervan, P. B.; Antiparallel side-by-side dimeric motif for sequence-specific recognition in the minor groove of DNA by the designed peptide 1-methylimidazole-2-carboxamide netropsin., *Proc. Natl. Acad. Sci. USA*, **1992**, *89*, 7586-7590.
- (39) Cho, J. Y.; Parks, M. E.; Dervan, P. B.; Cyclic polyamides for recognition in the minor groove of DNA., *Proc. Natl. Acad. Sci. USA*, **1995**, *92*, 10389-10392.
- (40) Kopka, M. L.; Goodsell, D. S.; Han, G. W.; Chiu, T. K.; Lown, J. W.; Dickerson, R. E.; Defining GC-specificity in the minor groove: side-by-side binding of the di-imidazole lexitropsin to C-A-T-G-G-C-C-A-T-G. *Structure*, **1997**, *15*, 1033-1046.
- (41) Jin, R.; Breslauer, K. J.; Characterization of the minor groove environment in a drug-DNA complex: bisbenzimidazole bound to the poly[d(AT)].poly[d(AT)]duplex., *Proc. Natl. Acad. Sci. U.S.A.*, **1988**, *85*, 8939-8942.
- (42) Downs, T. R.; Wilfinger, W. W. Fluorometric Qualification of DNA in Cells and Tissue. *Anal. Biochem.* **1983**, *131*, 538-547.
- (43) Kintanar, A.; Klevit, R. E.; Reid, B. R.; Two-dimensional NMR investigation of a bent DNA fragment: assignment of the proton resonances and preliminary structure analysis. *Nucleic Acids Research*, **1987**, *15*, 5845.
- (44) Katahira, M.; Sugeta, H.; Kyogoku, Y.; Fujii, S.; Fujisawa, R.; Tomita, K.; One- and two-dimensional NMR studies on the conformation of DNA containing the oligo(dA)oligo(dT) tract. *Nucleic Acids Research*, **1988**, *16*, 8619.

- 
- (45) Nadeau, J. G.; Crothers, D. M.; Structural basis for DNA bending. *Proc. Natl. Acad. Sci. USA*, **1989**, *86*, 2622.
- (46) Moe, J. G.; Russo, I. M.; Proton exchange and base-pair opening kinetics in 5'-d(CGCGAATTCGCG)-3' and related dodecamers. *Nucleic Acids Research*, **1990**, *18*, 821.
- (47) Shatzky-Schwartz, M.; Arbuckle, N. D.; Eisenstein, M.; Rabinovich, D.; Bareket-Samish, A.; Haran, T. E.; Luisi, B. F.; Shakked, Z.; X-ray and solution studies of DNA oligomers and implications for the structural basis of A-tract-dependent curvature. *J. Mol. Biol.*, **1997**, *267*, 595.
- (48) DiGabriele, A. D.; Steitz, T. A.; A DNA dodecamer containing an adenine tract crystallizes in a unique lattice and exhibits a new bend. *J. Mol. Biol.*, **1993**, *231*, 1024.
- (49) DiGabriele, A. D.; Sanderson, M. R.; Steitz, T. A.; Crystal lattice is important in determining the bend of a DNA dodecamer containing an adenine tract. *Proc. Natl. Acad. Sci. USA*, **1989**, *86*, 1816.
- (50) Young, M. A.; Beveridge, D. L.; Molecular dynamics simulations of an oligonucleotide duplex with adenine tracts phased by a full helix turn. *J. Mol. Biol.*, **1998**, *281*, 675.
- (51) Beveridge, D. L.; Dixit, S. B.; Barreiro, G.; Thayer, K. M.; Molecular dynamics simulations of DNA curvature and flexibility: helix phasing and premelting. *Biopolymers*, **2004**, *73*, 380.
- (52) Strahs, D.; Schlick, T.; A-tract bending: insights into experimental structures by computational models. *J. Mol. Biol.*, **2000**, *301*, 643.
- (53) Ulanovsky, L. E.; Trifonov, E. N.; Estimation of wedge components in curved DNA. *Nature*, **1987**, *326*, 720.
- (54) De Santis, P.; Palleschi, A.; Savino, M.; Scipioni, A.; Validity of the nearest-neighbor approximation in the evaluation of the electrophoretic manifestations of DNA curvature. *Biochemistry*, **1990**, *29*, 9269.
- (55) Crothers, D. M.; Haran, T. E.; Nadeau, J. G.; Intrinsically bent DNA. *Journal of Biological Chemistry*, **1990**, *265*, 7093.
- (56) Manning, G. S.; The persistence length of DNA is reached from the persistence length of its null isomer through an internal electrostatic stretching force. *Biophysical Journal*, **2006**, *91*, 1.
- (57) Burmistrova, A.; Gabelica, V.; Duwea, A.-S.; De Pauw, E.; Ion Mobility Spectrometry Reveals Duplex DNA Dissociation Intermediates. *J. Am. Soc. Mass. Spectrom.*, **2013**, *24*, 1777.
- (58) Alberts, B.; DNA replication and recombination, *Nature*, **2003**, *421*, 431-435.
- (59) Perez, A.; Orozco, M.; Real-Time Atomistic Description of DNA Unfolding. *Angew. Chem. Int. Ed.*, **2010**, *49*, 4805.
- (60) Wong, K.-Y.; Pettitt, B. M.; The Pathway of Oligomeric DNA Melting Investigated by Molecular Dynamics Simulations. *Biophysical Journal*, **2008**, *95*, 5618.
- (61) Ma, H.; Wan, C.; Wu, A.; Zewail, A. H.; DNA folding and melting observed in real time redefine the energy landscape. *Proc. Natl. Acad. Sci. USA*, **2007**, *104*, 712.
- (62) Jung, J.; Van Orden, A.; A Three-State Mechanism for DNA Hairpin Folding Characterised by Multiparameter Fluctuation Spectroscopy. *J. Am. Chem. Soc.*, **2006**, *128*, 1240.

- 
- (63) Shen, Y.; Kuznetsov, S. V.; Ansari, A.; Loop Dependence of the Dynamics of DNA Hairpins, *J. Phys. Chem. B*, **2001**, *105*, 12202.
- (64) Burmistrova, A.; Fresch, B.; Sluysmans, D.; De Pauw, E.; Remacle, F.; Duwez, A.-S.; Force measurements reveal how small binders perturb the dissociation mechanisms of DNA duplex sequences, *Nanoscale*, **2016**, *8*, 11718-11726.
- (65) Peng, C. S.; Jones, K. C.; Tokmakoff, A. Anharmonic Vibrational Modes of Nucleic Acid Bases Revealed by 2D IR Spectroscopy, *J. Am. Chem. Soc.*, **2011**, *133*, 15650-15660.
- (66) Yang, M.; Szyz, L.; Elsaesser, T. Femtosecond Two-Dimensional Infrared Spectroscopy of Adenine-Thymine Base Pairs in DNA Oligomers, *J. Phys. Chem. B*, **2011**, *115*, 1262-1267.
- (67) Greve, C.; Elsaesser, T. Ultrafast Two-Dimensional Infrared Spectroscopy of Guanine-Cytosine Base Pairs in DNA Oligomers, *J. Phys. Chem. B*, **2013**, *117*, 14009-14017.
- (68) Krummel, A. T.; Zanni, M. T. DNA Vibrational Coupling Revealed with Two-Dimensional Infrared Spectroscopy: Insight into Why Vibrational Spectroscopy is Sensitive to DNA Structure, *J. Phys. Chem. B*, **2006**, *110*, 13991-14000.
- (69) Sanstead, P. J.; Stevenson, P.; Tokmakoff, A. Sequence-Dependent Mechanism of DNA Oligonucleotide Dehybridization Resolved through Infrared Spectroscopy. *J. Am. Chem. Soc.*, **2016**, *138*, 11792-11801.
- (70) Hithell, G.; Gonzalez-Jimenez, M. Greetham, G. M.; Donaldson, P. M.; Towrie, M.; Parker, A. W.; Burley, G. A.; Wynne, K.; Hunt, N. T.; Ultrafast 2D-IR and Optical Kerr Effect Spectroscopy Reveal the Impact of Duplex Melting on the Structural Dynamics of DNA, *Phys. Chem. Chem. Phys.*, **2017**, DOI: 10.1039/C7CP00054E.
- (71) Hamm, P.; Zanni, M. T.; Concepts and Methods of 2D Infrared Spectroscopy, Cambridge University Press, Cambridge, 2011.
- (72) Hunt, N. T. 2D-IR spectroscopy: Ultrafast Insights into Biomolecule Structure and Function, *Chem. Soc. Rev.*, **2009**, *38*, 1837-1848.
- (73) Park, S.; Kwak, K.; Fayer, M. D. Ultrafast 2D-IR Vibrational Echo Spectroscopy: A Probe of Molecular Dynamics, *Laser Phys. Lett.*, **2007**, *10*, 704-718.
- (74) Baiz, C. R.; Mcrobbie, P. L.; Anna, J. M.; Geva, E.; Kubarych, K. J. Two-Dimensional Infrared Spectroscopy of Metal Carbonyls, *Accounts of Chemical Research*, **2009**, *42*, 1395-1404.
- (75) Hamm, P.; Lim, M.; Hochstrasser, R. M.; Structure of the Amide I Band of Peptides Measured by Femtosecond Nonlinear-Infrared Spectroscopy, *J. Phys. Chem. B.*, **1998**, *102*, 6123-6138.
- (76) Lepetit, L.; Chériaux, G.; Joffre, M.; Linear techniques of phase measurement by femtosecond spectral interferometry for applications in spectroscopy., *J. Opt. Soc. Am. B.*, **1995**, *12*, 2467-2474.
- (77) Kwak, K.; Park, S.; Finkelstein, I. J.; Fayer, M. D.; Frequency-frequency correlation functions and apodization in two-dimensional infrared vibrational echo spectroscopy: a new approach., *J. Chem. Phys.*, **2007**, *127*, 124503.
- (78) Fenn, E. E.; Fayer M. D.; Extracting 2D IR frequency-frequency correlation function from two component systems., *J. Chem. Phys.*, **2011**, *135*, 074502.

- 
- (79) Kim, H.; Cho, M.; Infrared Probes for Studying the Structure and Dynamics of Biomolecules, *Chem. Rev.*, **2013**, 113, 5817-5847.
- (80) Waegele, M. M.; Culik, R. M.; Gai, F.; Site-Specific Spectroscopic Reporters of the Local Electric Field, Hydration, Structure, and Dynamics of Biomolecules, *J. Phys. Chem. Lett.*, **2011**, 2, 2598-2609.
- (81) Bandaria, J. N.; Dutta, S.; Hill, S. E.; Kohen, A.; Cheatum, C. M.; Fast Enzyme Dynamics at the Active Site of Formate Dehydrogenase, *J. Am. Chem. Soc.*, **2008**, 130, 22-23.
- (82) Cho, M.; Coherent Two-Dimensional Optical Spectroscopy, *Chem. Rev.*, **2008**, 108, 1331-1418.
- (83) Kozinski, M.; Garrett-Roe, S.; Hamm, P.; 2D-IR Spectroscopy of the Sulfhydryl Band of Cysteines in the Hydrophobic Core of Proteins, *J. Phys. Chem. B*, **2008**, 112, 7645-7650.
- (84) Hamm, P.; Zanni, M. T.; Concepts and Methods of 2D Infrared Spectroscopy, Cambridge University Press, Cambridge, 2011.
- (85) Marcos, E.; Basanta, B.; Chidyausiku, T. M.; Tang, Y.; Oberdorfer, G.; Lui, G.; Swapna, G. V.; Guan, R.; Silva, D. A.; Dou, J.; Pereira, J. H.; Xiao, R.; Sankaran, B.; Zwart, P. H.; Montelione, G. T.; Baker, D.; Principles for designing proteins with cavities formed by curved beta sheets., *Science*, **2017**, 355, 201-206.
- (86) Lee, K. K.; Park, K. H.; Joo, C.; Kwon, H. J.; Han, H.; Ha, J. H.; Park, S.; Cho, M.; Ultrafast internal rotational dynamics of the azido group in (4S)-azidoproline: Chemical exchange 2DIR spectroscopic investigations *Chem. Phys.*, **2012**, 396, 23-29.
- (87) Waegele, M. M.; Tucker, M. J.; Gai, F.; 5-Cyanotryptophan as an infrared probe of local hydration status of proteins, *Chem. Phys. Lett.*, **2009**, 478, 249-253.
- (88) Huang, C. Y.; Wang, T.; Gai, F.; Temperature dependence of the CN stretching vibration of a nitrile-derivatized phenylalanine in water, *Chem. Phys. Lett.* **2003**, 371, 731-738.
- (89) Getahun, Z.; Huang, C. J.; Wang, T.; DeLeon, B.; DeGrado, W. F.; Gai, F.; Using Nitrile-Derivatized Amino Acids as Infrared Probes of Local Environment, *J. Am. Chem. Soc.* **2003**, 125, 405-411.
- (90) Hu, W.; Webb, L. J.; Direct Measurement of the Membrane Dipole Field in Bicelles Using Vibrational Stark Effect Spectroscopy, *J. Phys. Chem. Lett.* **2011**, 2, 1925-1930.
- (91) Tucker, M. J.; Kim, K. S.; Hochstrasser, R. M.; 2D IR photon echo study of the anharmonic coupling in the OCN region of phenyl cyanate, *Chem. Phys. Lett.* **2009**, 470, 80-84.
- (92) Stafford, A. J.; Ensign, D. L.; Webb, L. J.; Vibrational Stark Effect Spectroscopy at the Interface of Ras and Rap1A Bound to the Ras Binding Domain of RalGDS Reveals an Electrostatic Mechanism for Protein-Protein Interaction, *J. Phys. Chem. B* **2010**, 114, 15331-15344.
- (93) Thielges, M. C.; Axup, J. Y.; Wong, D.; Lee, H. S.; Chung, J. K.; Schultz, P. G.; Fayer, M. D.; Two-Dimensional IR Spectroscopy of Protein Dynamics Using Two Vibrational Labels: A Site-Specific Genetically Encoded Unnatural Amino Acid and an Active Site Ligand, *J. Phys. Chem. B* **2011**, 115, 11294-11304.
- (94) Inouye, H.; Gleason, K. A.; Zhang, D.; Decatur, S. M.; Kirschner, D. A.; Differential effects of phe19 and phe20 on fibril formation by amyloidogenic peptide A $\beta$ 16-22 (Ac-KLVFFAE-NH<sub>2</sub>), *Proteins: Struct., Funct., Bioinf.* **2010**, 78, 2306-2321.

- 
- (95) Bischak, C. G.; Longhi, S.; Snead, D. M.; Costanzo, S.; Terrer, E.; Londergan, C. H. *Biophys. J.* **2010**, *99*, 1676.
- (96) Mukherjee, S.; Chowdhury, P.; DeGrado, W. F.; Gai, F.; Site-Specific Hydration Status of an Amphipathic Peptide in AOT Reverse Micelles, *Langmuir*, **2007**, *23*, 11174-11179.
- (97) Ohta, K.; Tayama, J.; Tominaga, K.; Ultrafast vibrational dynamics of SCN<sup>-</sup> and N<sub>3</sub><sup>-</sup> in polar solvents studied by nonlinear infrared spectroscopy, *Phys. Chem. Chem. Phys.* **2012**, *14*, 10455-10465.
- (98) Hamm, P.; Lim, M.; Hochstrasser, R. M.; Non-Markovian Dynamics of the Vibrations of Ions in Water from Femtosecond Infrared Three-Pulse Photon Echoes, *Phys. Rev. Lett.* **1998**, *81*, 5326.
- (99) Choi, J. H.; Raleigh, D.; Cho, M.; Azido Homoalanine is a Useful Infrared Probe for Monitoring Local Electrostatics and Side-Chain Solvation in Proteins, *J. Phys. Chem.* **2011**, *2*, 2158-2162.
- (100) Gai, X. S.; Coutifaris, B. A.; Brewer, S. H.; Fenlon, E. E.; A direct comparison of azide and nitrile vibrational probes, *Phys. Chem. Chem. Phys.* **2011**, *13*, 5926-5930.
- (101) Dutta, S.; Li, Y. L.; Rock, W.; Houtman, J. C. D.; Kohen, A.; Cheatum, C. M.; 3-Picolyl Azide Adenine Dinucleotide as a Probe of Femtosecond to Picosecond Enzyme Dynamics, *J. Phys. Chem. B*, **2012**, *116*, 542-548.
- (102) Tucker, M. J.; Gai, X. S.; Fenlon, E. E.; Brewer, S. H.; Hochstrasser, R. M.; 2D IR photon echo of azido-probes for biomolecular dynamics, *Phys. Chem. Chem. Phys.* **2011**, *13*, 2237-2241.
- (103) Oh, K. I.; Lee, J. H.; Joo, C.; Han, H.; Cho, M.;  $\beta$ -Azidoalanine as an IR Probe: Application to Amyloid A $\beta$ (16-22) Aggregation, *J. Phys. Chem. B* **2008**, *112*, 10352-10357.
- (104) Ye, S.; Zaitseva, E.; Caltabiano, G.; Schertler, G. F. X.; Sakmar, T. P.; Deupi, X.; Vogel, R.; Tracking G-protein-coupled receptor activation using genetically encoded infrared probes, *Nature*, **2010**, *464*, 1386-1389.
- (105) Hofmann, K. P.; Scheerer, P.; Choe, H. W.; Park, J. H.; Heck, M.; Ernst, O. P.; A G protein-coupled receptor at work: the rhodopsin model., *Trends Biochem. Sci.*, **2009**, *34*, 540-552.
- (106) Farrens, D. L.; Altenbach, C.; Yang, K.; Hubbell, W. L.; Khorana, H. G. *Science*, **1996**, *274*, 497.
- (107) Scheerer, P.; Park, J. H.; Hildebrand, P. W.; Kim, Y. J.; Krauss, N.; Choe, H. W.; Hofmann, K. P.; Ernst, O. P.; Crystal structure of opsin in its G-protein-interacting conformation., *Nature*, **2008**, *455*, 497-502.
- (108) Bazewicz, C. G.; Liskov, M. T.; Hines, K. J.; Brewer, S. H.; Sensitive, Site-Specific, and Stable Vibrational Probe of Local Protein Environments: 4-Azidomethyl-L-Phenylalanine, *J. Chem. Phys. B*, **2013**, *117*, 8987-8993.
- (109) Bloem, R.; Koziol, K.; Waldauer, S. A.; Buchli, B.; Walser, R.; Samatanga, B.; Jelesarov, I.; Hamm, P.; Ligand Binding Studied by 2D IR Spectroscopy Using the Azidohomoalanine Label, *J. Phys. Chem. B*, **2012**, *116*, 13705-13712.
- (110) Dutta, S.; Rock, W.; Cook, R. J.; Kohen, A.; Cheatum, C. M.; Two-dimensional infrared spectroscopy of azido-nicotinamide adenine dinucleotide in water., *J. Chem. Phys.*, **2011**, *135*, 055106.

---

(111) Hamm, P.; Zanni, M. in *Concepts and Methods of 2D Infrared Spectroscopy*, Cambridge University Press, Cambridge, **2011**, pg. 140-142

(112) Lipkin, J. S.; Song, R.; Fenlon, E. E.; Brewer, S. H.; Modulating Accidental Fermi Resonance: What a Difference a Neutron Makes, *J. Phys. Chem. Lett.* **2011**, 2, 1672-1676.

(113) Nydegger, M. W.; Dutta, S.; Cheatum C. M.; *2D IR Study of 3-Azidopyridine as a Potential Spectroscopic Reporter of Protonation State*, *J. Chem. Phys.*, **2010**, 133, 134506-134508.



# Experimental Methods and Materials

Chapter 2

This chapter outlines the sample preparation, the different types of standard spectroscopic techniques and describes the advanced IR spectrometers used throughout this thesis. These are intended as a reference for all subsequent results chapters and further details, such as the exact concentrations of the samples used, are given at the end of each result chapter.

## 2.1 Materials and Sample Preparation

### ***Benzyl Azide Samples (Sample Preparation for Chapter 3):***

All of the solvents used were purchased from Sigma Aldrich. Benzyl azide was synthesised using a substitution reaction between 1 equivalent benzyl bromide and 1.5 equivalents of sodium azide. The reaction was carried out in a 4:1 mixture (by volume) of acetone and water, the reaction was heated overnight and subsequently extracted with diethyl ether three times, dried over sodium sulphate and then the solvent was removed (under reduced pressure) using a rotary evaporator.<sup>114</sup> All solvents were used without further purification. Solutions of benzyl azide were prepared in a wide range of protic and aprotic solvents. Once prepared the benzyl azides were thoroughly shaken for at least 10 minutes.

### ***DNA Samples (Sample Preparation for Chapters 4, 5, 6 and 7):***

Lyophilised, salt-free DNA oligonucleotides were obtained from Eurofins; Hoechst 33258 (H33258), D<sub>2</sub>O, dimethyl sulfoxide (DMSO), NaCl, monobasic and dibasic sodium phosphate were obtained from Sigma-Aldrich. A *de novo* azido-derivative of H33258 (N<sub>3</sub>-bBI) was synthesized and purified by Dr. J. J. May (Burley Group)<sup>115</sup>. All chemicals were used without further purification. All samples were prepared in a pD (or pH) 7 phosphate buffer solution with an overall ionic strength of 200 mM.

This buffer solution was prepared by mixing 423  $\mu$ L of a 1 M monobasic phosphate stock solution with 577  $\mu$ L of a 1 M dibasic phosphate stock solution, and then diluting the overall mixture to a total volume of 10 mL. Finally 58.4 mg of NaCl was added to the phosphate buffer solution to increase the overall ionic strength to 200 mM.

The DNA oligonucleotides were made up to 10 mM stocks in the D<sub>2</sub>O phosphate buffer solution. For the A<sub>3</sub>T<sub>3</sub> (5'-GGAAATTTGC-3') sequence these stocks were diluted to a final concentration of 2.5 mM and for the (AT)<sub>3</sub> (5'-GGATATATGC-3') sequence these stocks were diluted to a final concentration of 5 mM. The prepared samples were then annealed at 90 °C for 10 minutes. All the other concentrations used were prepared by serial dilution from the original A<sub>3</sub>T<sub>3</sub> and (AT)<sub>3</sub> samples.

For the Hoechst33258:DNA complexes a 100 mM stock solution of the Hoechst33258 (H33258) ligand was prepared in DMSO. All samples of this complex were prepared using pD 7 phosphate buffer solution to a final duplex:ligand molar ratio of 1:1 and annealed at 90 °C for 10 minutes.

For the N<sub>3</sub>-bBI:DNA complexes a 100 mM stock solution of the N<sub>3</sub>-bBI ligand was prepared in DMSO. All samples of this complex were prepared using pD7 phosphate buffer solution to a final duplex:ligand ratio of 1:1 and annealed at 90 °C for 10 minutes.

## 2.2 UV-visible Absorption Spectroscopy

For all the UV-visible absorption spectroscopy experiments the samples were held in a demountable transmission cell utilising CaF<sub>2</sub> windows separated by a 10 μm polytetrafluoroethylene (PTFE) spacer. UV-visible spectroscopy experiments were carried out using a Perkin-Elmer Lambda 25 at a resolution of 1 nm with samples at concentrations defined in the relevant results chapter. These measurements were repeated at 1 μM (samples prepared via serial dilution to ensure accuracy) to provide duplex melting temperature data at this concentration to compare with the fluorescence measurements.

## 2.3 Fluorescence Emission Spectroscopy

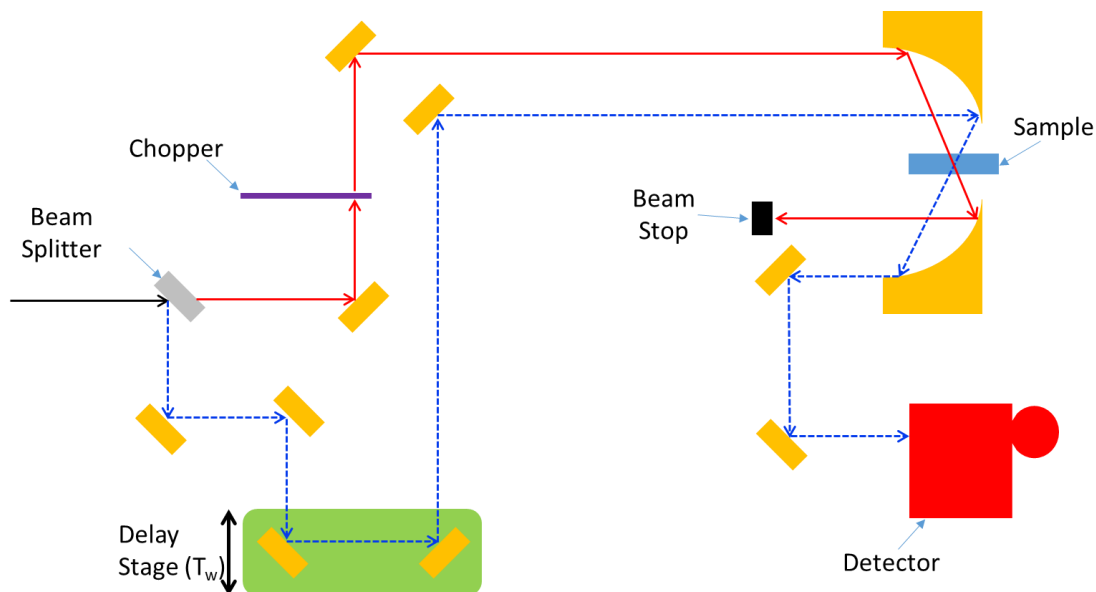
The fluorescence emission spectroscopy experiments were carried out using a Horiba Fluorolog2 at a resolution of 1 nm. The samples were excited at 350 nm and the emitted fluorescence was recorded from 380 nm – 600 nm. The 1 μM sample was diluted from the original samples (concentrations of 2.5 mM (A<sub>3</sub>T<sub>3</sub> duplex/H-A<sub>3</sub>T<sub>3</sub> complex/N<sub>3</sub>-A<sub>3</sub>T<sub>3</sub> complex) or 5 mM ((AT)<sub>3</sub> duplex/H-(AT)<sub>3</sub> complex/N<sub>3</sub>-(AT)<sub>3</sub> complex)) via serial dilution to ensure accuracy. The samples were held in a quartz cuvette with a path length of 1 cm.

## 2.4 FT-IR Spectroscopy

The FT-IR spectroscopy experiments were carried out using a Bruker Vertex 70 spectrometer at a resolution of 1 cm<sup>-1</sup>. For all IR measurements, samples were held between two CaF<sub>2</sub> windows separated by a polytetrafluoroethylene spacer of 50 μm thickness at concentrations defined in the relevant results chapter. Each FT-IR spectrum is composed of the average of 25 individual scans, and for each sample an FT-IR spectrum of the solvent used in the sample is also recorded to yield a solvent spectrum (average of 25 scans). Before each FT-IR spectrum is measured a background scan was performed, this background scan is also composed of an average of 25 individual measurements.

## 2.5 Strathclyde IR Pump-probe Spectrometer

A diagram of the optical pathways in the Strathclyde ultrafast IR pump-probe setup is shown in Figure 2.1.



**Figure 2.1:** Schematic of the optical setup of the Strathclyde Ultrafast IR pump-probe spectrometer utilised. In this schematic the probe path is indicated by dashed blue arrows and the pump path is indicated by red arrows. The reference path is omitted for clarity.

The Strathclyde IR pump-probe spectrometer utilises a Ti:Sapphire oscillator (Coherent Micra) to seed a regenerative amplifier (Coherent Legend Elite) to generate a pulse train consisting of 35 fs pulses at a repetition rate of 1 kHz. The peak power of these pulses was measured to be in the range 2.5 W - 2.7 W. Using a beam splitter 1 W of each of the pulses was directed into an optical parametric amplifier (OPA), setup to convert the wavelength of these pulses to the mid-IR frequency range. The resulting train of IR pulses was observed to have a peak power of 40 mW - 50 mW.

These pulses were then directed into the optical paths of the IR pump-probe spectrometer setup shown in figure 2.1. Briefly, in the spectrometer a beam splitter was utilised to split each of the IR pulses into a pump pulse (containing 90% of the original power), a probe pulse (containing 5% of the original power, via the front reflection of the beam splitter) and a reference pulse (at 5% of the original power, via the back reflection of the beam splitter). The path of the reference pulse is omitted from the schematic of the optical paths in figure 2.1 for clarity, however this pulse is directed into a separate detector and allows the effect of shot-to-shot fluctuations in the pulse power to be removed from the measured pump-probe

spectrum. After the beam splitter the probe pulse is then directed onto a variable optical delay stage, allowing the arrival of this pulse at the sample to be delayed by a time referred to as the waiting time ( $T_w$ ). Whereas the pump pulse travels along a simple optical path containing a chopper rotating at 500 Hz (triggered by the laser), leading to every other pump pulse being blocked. At the end of the optical paths both the pump and the probe pulses are directed onto a parabolic mirror and focused onto the sample. After the sample the pulses are re-collimated using a second parabolic mirror and the remaining pump pulse is blocked by a beam stop whereas the probe pulse is directed into a detector containing a diffraction grating (centred at 4.5  $\mu\text{m}$ ) and a 64-pixel liquid nitrogen cooled MCT (mercury cadmium telluride) array. Scanning  $T_w$  then allows an IR pump-probe spectrum to be obtained via equation (1).

$$S(T_w) = -\log\left(\frac{I_{probe, on}}{I_{probe, off}}\right) \dots (1)$$

Where  $S(T_w)$  is the pump-probe spectrum at a waiting time of  $T_w$ ,  $I_{probe, on}$  the intensity of the probe pulse when the sample has been irradiated by the pump pulse and  $I_{probe, off}$  is the intensity of the probe pulse when the sample has not been irradiated by the pump pulse.

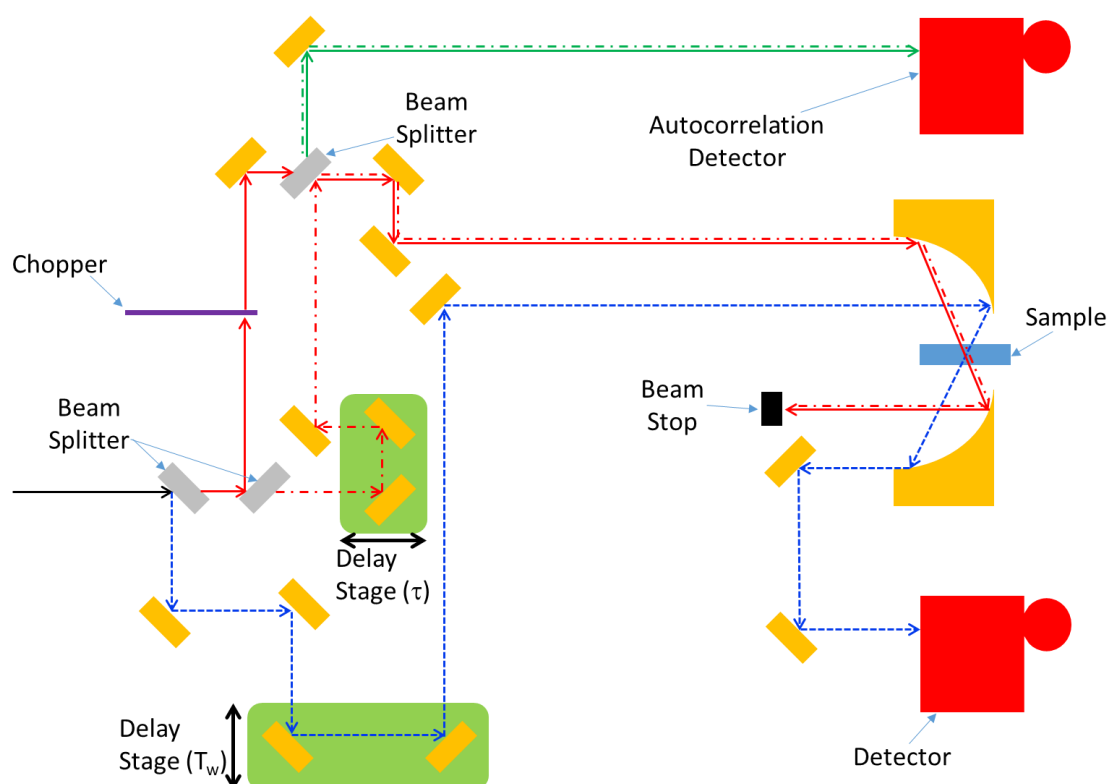
Note as the pump-probe signal is emitted from the sample in a collinear direction to the residual probe pulse, the signal is self-heterodyned allowing the phase information contained within the signal to be recovered via a spectral interferometry method.<sup>116</sup> A range of pump-probe spectra were collected at a series of waiting time ( $T_w$ ) to allow the vibrational dynamics of the mode to be extracted. In the Strathclyde IR pump-probe spectrometer this was achieved by positioning the optical delay stage in the probe path at each different waiting time and measuring the resultant pump-probe at a 1 second acquisition time (1000 shot average). Once the pump-probe spectrum at each of the selected waiting times has been measured, the scan is repeated to obtain 12 scans of the pump-probe spectra in total. Finally it is noted that in the Strathclyde Ultrafast IR pump-probe spectrometer the pump-probe spectrum was measured in both the ZZZZ and ZZYY polarisation geometries.

Once the data has been measured each of the pump-probe spectra in each separate scan is baseline corrected, using a specially designed LabView program, by selecting the baseline pixels and fitting a polynomial baseline. This baseline correction and averaging procedure is carried out for both the pump-probe spectrum measured in the ZZZZ geometry and the ZZYY geometries. Finally, the spectra at the ZZZZ and ZZYY polarisation geometries were used to

generate the isotropic pump-probe spectra. After this baseline subtraction the data is used to produce an average pump-probe spectrum and the spectrum at each waiting time is fitted with a set of Gaussian peaks, based on the fitting of the FT-IR spectrum. The change in the intensity of these peaks was plotted against the waiting time allowing the vibrational lifetime(s) of the mode to be extracted.

## 2.6 Strathclyde and ULTRA 2D-IR spectrometers

A diagram of the optical pathways utilised in both the Strathclyde and ULTRA 2D-IR<sup>117,118</sup> spectrometers is shown in figure 2.2. While the first spectrometer is located at the University of Strathclyde, the ULTRA system is based at the STFC Central Laser Facility in the Rutherford Appleton Laboratory.



**Figure 2.2:** Schematic of the optical setup of the Strathclyde and ULTRA 2D-IR spectrometer utilised.<sup>117,118</sup> In this schematic the probe path is indicated by dashed blue arrows, the pump path is indicated by red arrows and the autocorrelation path is indicated by green arrows. The reference path is omitted for clarity.

As for the Strathclyde IR pump-probe spectrometer, the pulse trains for the Strathclyde and ULTRA 2D-IR spectrometers are generated utilising Ti:Sapphire oscillators and regenerative amplifiers. For the Strathclyde system this is achieved using the same laser system as described in for the ultrafast IR pump-probe spectrometer. In the case of the ULTRA

spectrometer these sub-50fs pulse trains are produced using a Ti:Sapphire oscillator (Thales laser) to seed a regenerative amplifier, generating a pulse with a repetition rate of 10 kHz.

The pulse trains produced in both of these systems are then passed through an OPA allowing the wavelength of the pulses to be increase to 4.5  $\mu\text{m}$  in the case of the Strathclyde system (allowing the 2D-IR spectra of the asymmetric stretch of benzyl azide to be measured) and 6  $\mu\text{m}$  in the case of the ULTRA 2D-IR spectrometer (allowing the 2D-IR spectra of the DNA base modes to be measured).

After the OPA in both of the Strathclyde and ULTRA systems the IR pulses pass through a beam splitter. Both of the 2D-IR spectrometers utilise the same probe path as previously described in the IR pump-probe spectrometer, however the pump path now contains an additional interferometer. In the 2D-IR spectrometers the pump pulse is passed through a Mach-Zehnder interferometer adapted using a variable optical delay stage in one of the arms and a chopper (set to half the repetition rate of the laser system) in the other arm (Figure 2.2). This results in the production of two collinear pump pulses with a variable time delay,  $\tau$ , between them. This delay is referred to as the coherence time. From the final beam splitter in the interferometer one of the pulse pairs is directed onto the autocorrelation path (green, Fig.2.2) and an autocorrelation signal between the two pump pulses is recorded. The other pair of collinear pump pulses and the probe pulse are then directed onto a parabolic mirror and focused onto the sample following a pseudo-pump probe geometry. After the sample the pulses are re-collimated using a second parabolic mirror and pump pulses are blocked using a beam stop. The signal plus the residual probe light are directed into a detector unit utilising a diffraction grating and a liquid nitrogen cooled MCT array. (For the Strathclyde system there are 64 pixels in this array and for the ULTRA system there are 128 pixels in this array).

During the measurement of the 2D-IR spectrum the optical delay stage between the two pump pulses and the probe pulse remains fixed at a single waiting time. The signal is then detected as the coherence time ( $\tau$ ) between the two pump pulses is scanned. In all of the measurements presented in the relevant results chapters this stage was scanned from a temporal delay of -0.4 ps to 4 ps in 3 fs steps. At each of the chosen coherence times the signal generated by the interaction between the pulse sequence and the sample was acquired for 0.4 seconds, and the resultant average signal obtained during the signal was recorded. This entire procedure was then repeated to obtain 6 individual spectra for the

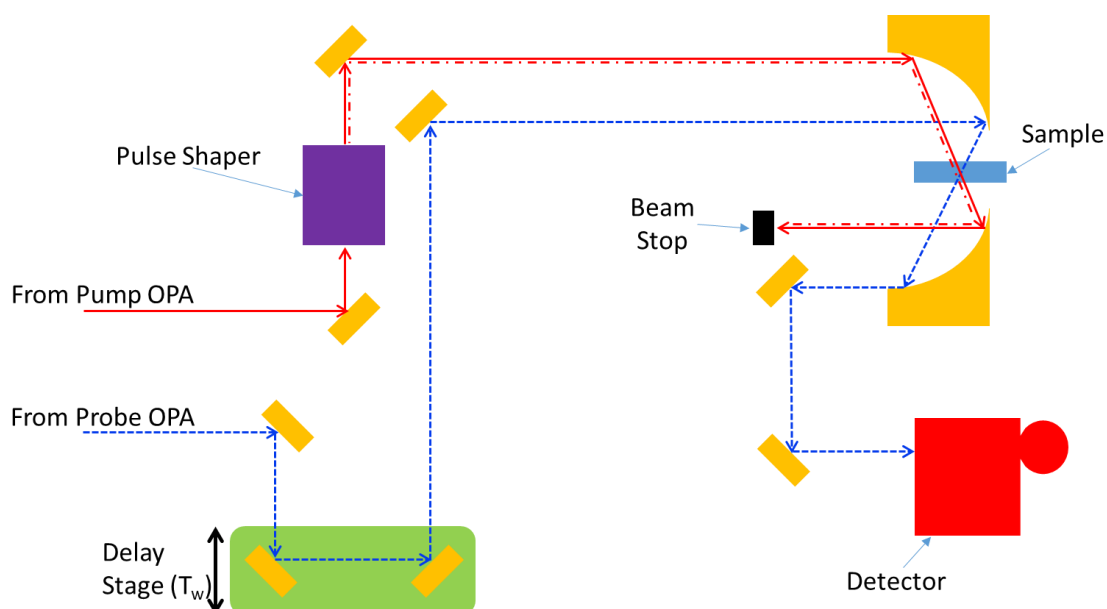
ULTRA spectrometer and 20 individual spectra for the Strathclyde spectrometer. During each scan the autocorrelation signal between the two pump pulses was also recorded.

After the measurement the individual spectra and autocorrelations are averaged. The average autocorrelation is then scanned to obtain an initial idea of the exact time zero between the two pump pulses. This initial guess for time zero is then used to carry out a Fourier transform, with respect to the coherence time, on the average spectrum to obtain an initial 2D-IR spectrum. In the final stage of the analysis program the phasing of the 2D-IR spectrum is corrected by adjusting the exact time zero between the two pump pulses using the slice projection theorem.<sup>119,120</sup> In this process the time zero is adjusted manually to until the best overlap between the pump-probe spectrum of the sample and the sum of the pump slices of the 2D-IR spectrum is obtained. Once this overlap has been achieved the 2D-IR spectrum is known to be correctly phased. Upon the completion of this phasing, the final 2D-IR spectrum is produced. This entire procedure is repeated for each desired waiting time. All measurements on the DNA base modes were performed in the ZZZZ polarisation geometry. All measurements on the asymmetric azide stretch were performed in the both the ZZZZ and ZZYY polarisation geometries.

## 2.7 LIFETIME 2D-IR Spectrometer<sup>121,122,123</sup>

A diagram of the optical pathways utilised in both the LIFETIME 2D-IR spectrometer is shown in figure 2.3.

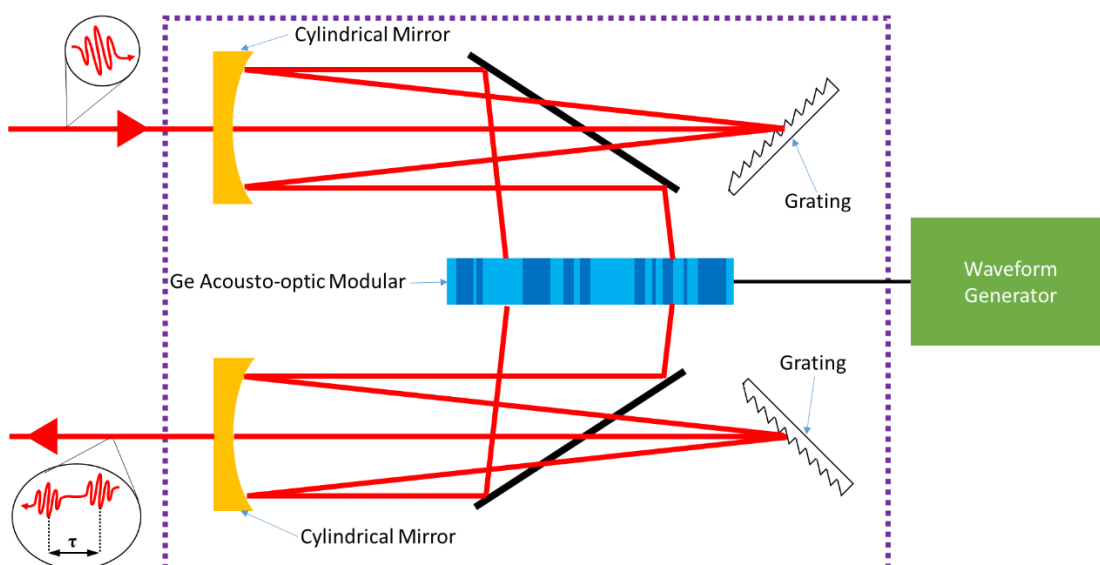




**Figure 2.3:** Schematic of the optical setup of the LIFETIME 2D-IR spectrometer utilised. In this schematic the probe path is indicated by dashed blue arrows and the pump path is indicated by red arrows.

In contrast to the Strathclyde and ULTRA 2D-IR systems the pulse train in the LIFETIME 2D-IR spectrometer utilises a pair of 100 kHz Yb:KGW amplified laser systems coupled to pulse-pickers to produce a pulse train at a repetition rate of 100 kHz. Using a pair of OPAs the wavelength of these pulses was altered to produce pulses centred at approximately 4.8  $\mu\text{m}$ , with a temporal duration of  $\sim 300$  fs.

One of these pulse trains is then directed into the probe optical path of the LIFETIME system. As in both the Strathclyde and ULTRA 2D-IR spectrometers the probe path contains a single optical delay stage, allowing the arrival of the probe pulse at the sample to be delayed by a variable amount of time,  $T_w$ . The other pulse train is directed onto the pump path of the spectrometer. This path contains an IR pulse shaper,<sup>124</sup> allowing the overall shape of the pump pulse to be arbitrarily altered. A diagram of the IR pulse shaper present in the LIFETIME 2D-IR spectrometer is shown in figure 2.4.



**Figure 2.4:** Schematic of the IR pulse shaper used in the LIFETIME 2D-IR spectrometer.<sup>124</sup> Inserts demonstrating that the pulse shaper can be used to produce two collinear pulses separated by a variable time delay  $\tau$ , from a single input pulse.

After the pulse enters the IR pulse shaper it is directed onto a diffraction grating and cylindrical mirror pair leading to the different frequencies present in the pulse to become spatially separated and the overall beam to be collimated. This collimated beam is then passed through a germanium acousto-optic modulator. This modulator is coupled to a waveform generator which allows the collimated wave-front to be re-shaped. After the modulator a further cylindrical mirror and grating to reconstitute the reshaped IR pulse. Using an appropriate series of arbitrary waveforms for the Ge modulator allows the pulse shaper to produce a series of collinear pairs of pump pulses at a number of different temporal delays ( $\tau$ ). Therefore in the LIFETIME spectrometer the IR pulse shaper is used to simulate the effect of the modified Mach-Zehnder interferometer present in the Strathclyde and ULTRA 2D-IR spectrometers.

As in the ULTRA and Strathclyde 2D-IR spectrometer systems the shaped pump pulse and probe pulse are then directed onto a pair of parabolic mirrors allowing the two pulses to be focused onto the sample. After the sample the pump pulse is blocked via a beam stopper and whereas the residual probe light and signal are directed into a detector consisting of a diffraction grating and a 128 pixel liquid nitrogen cooled MCT detector. By scanning the value of the coherence time ( $\tau$ ) between the two pump pulses (by utilising the appropriate waveform for the Ge acousto-optic modulator in the IR pulse shaper) at a signal waiting time yields a 2D-IR spectrum.

In the case of the LIFETIME 2D-IR spectrometer the signal generated at each of the waveforms used in the IR pulse shaper was collected for 100 s, this was then repeated 20 times for all of the loaded waveforms and an average spectrum was generated. It is important to note that the LIFETIME spectrometer was set up to perform phase cycling to suppress sample scatter throughout all the measurements. A Fourier transform with respect to the coherence time was then performed using the time zero set by the IR pulse shaper to produce the final 2D-IR spectrum. This procedure was repeated for each waiting time. All measurements were performed in the ZZZZ polarisation geometry and for the measurements of the asymmetric azide stretch background measurements were performed to allow the water librations to be subtracted from the obtained data.

## 2.8 References

---

- (114) Synthetic work carried out by Dr. J. J. May, Burley Group.
- (115) Doctoral Thesis by Dr. J. J. May, submitted 30<sup>th</sup> September 2016.
- (116) Lepetit, L.; Chériaux, G.; Joffre, M.; Linear techniques of phase measurement by femtosecond spectral interferometry for applications in spectroscopy., *J. Opt. Soc. Am. B.*, **1995**, *12*, 2467-2474.
- (117) Shaw, D. J.; Adamczyk, K.; Frederix, P. W. J. M.; Simpson, N.; Robb, K.; Greetham, G. M.; Towrie, M.; Parker, A. W.; Hoskisson, P. A.; Hunt, N. T. Multidimensional Infrared Spectroscopy Reveals the Vibrational and Solvational Dynamics of Isoniazid, *J. Chem. Phys.*, **2015**, *142*, 212401.
- (118) Greetham, G. M.; Burgos, P.; Cao, Q.; Clark, I. P.; Codd, P. S.; Farrow, R. C.; George, M. W.; Kogimtzis, M.; Matousek, P.; Parker, A. W.; *et al.* M. ULTRA: A Unique Instrument for Time-Resolved Spectroscopy. *Applied Spectroscopy*, **2010**, *12*, 1311-1319.
- (119) Hamm, P.; Zanni, M.; in *Concepts and Methods of 2D Infrared Spectroscopy*, Cambridge University Press, Cambridge, **2011**, pg. 84-87.
- (120) Gallagher Faeder, S. M.; Jonas, D. M.; Two-dimensional electronic correlation and relaxation spectra: Theory and model calculations., *J. Phys. Chem. A.*, **1999**, *103*, 10489-10505.
- (121) Donaldson, P. M.; Greetham, G. M.; Shaw, D. J.; Parker, A. W.; Towrie, M.; A 100 kHz Pulse Shaping 2D-IR Spectrometer based on Yb:KGW Amplifiers., *Opt. Lett.*, **2016**, submitted.
- (122) Greetham, G. M.; Donaldson, P. M.; Nation, C.; Sazanovich, I. V.; Clark, I. P.; Shaw, D. J.; Parker, A. W.; Towrie, M.; A 100 kHz Time-Resolved Multiple-Probe Femtosecond to Second IR Absorption Spectrometer., *Applied Spectroscopy*, **2016**.
- (123) Hithell, G.; Shaw, D. J.; Donaldson, P. M.; Greetham, G. M.; Towrie, M.; Burley, G. A.; Parker, W. A.; Hunt, N. T. Long-Range Vibrational Dynamics Are Directed by Watson-Crick Base Pairing in Duplex DNA. *J. Phys. Chem. B*, **2016**, *120*, 4009-4018.
- (124) Hamm, P.; Zanni, M.; in *Concepts and Methods of 2D Infrared Spectroscopy*, Cambridge University Press, Cambridge, **2011**, pg. 199-201.

# Characterisation of the Asymmetric Azide Stretch as a Non-Natural IR Probe: Benzyl Azide

Chapter 3

### 3.1 Abstract

*Molecular environments present within bio-macromolecules and their complexes provide important contributions to their overall entropy and enthalpy. In order to gain a better understanding of the complex behaviour of these systems these local environment need to be understood at a more fundamental level. One approach which allows both the nature and dynamics of these local environments to be explored is the implementation of non-natural IR probes. Here, we investigate the alterations in the characteristic asymmetric azide stretching mode of benzyl azide, an archetypal non-natural azide probe, to changes in the electrostatic potential and hydrogen bonding strength of its local molecular environment. The results of FT-IR, ultrafast IR pump-probe and 2D-IR spectroscopies showed that the line-shape of the asymmetric azide stretch of benzyl azide is complicated by the presence of two Fermi resonances. Alterations in the electrostatic potential of the local molecular environment were found to alter the position of the modes underlying the Fermi resonances, changing the strength of the coupling to these modes, and led to a shorter vibrational lifetime of the azide mode. Changes in the environmental hydrogen bonding strength resulted in a shift in the position of the azide stretch, alterations in the lifetimes of the Fermi resonances and alters the coupling between the high-frequency shoulder and the azide mode. Finally the coupling constant between the low-frequency shoulder and the azide mode is determined to be sensitive to the exact nature of its molecular environment. Overall this investigation demonstrates the sensitivity of benzyl azide for its local environment and illustrates that the response of both the asymmetric azide stretch and its Fermi resonances can be used to gain a more fundamental understanding of molecular environments.*

## 3.2 Introduction

In order to be able to study specific regions of interesting biological molecules using spectroscopic methods, molecular probes are often implemented.<sup>125,126</sup> Sometimes these probes are naturally occurring groups in the molecule while others are non-naturally occurring functional groups. For example in the IR spectroscopy of proteins or peptides, these naturally occurring groups could be the amide links of the protein<sup>127,128</sup> or the thiol group of cysteine.<sup>129</sup> Examples of non-natural IR probes are the incorporation of azido<sup>130</sup> or cyano<sup>131,132</sup> amino acid derivatives into the peptide or protein. Each of these methods has distinct advantages and disadvantages associated with them. For the naturally occurring probes the advantages include the fact that they do not perturb the structure or function of the system being studied. However their vibrational bands are often found in congested regions of the spectrum making it difficult to extract environmental information from these vibrations.<sup>79</sup> The non-natural probes have the advantage that they are highly sensitive to the local environment, intense and in an uncongested spectral window.<sup>79</sup> However their introduction can perturb the physical system being studied and it can be difficult to successfully introduce these probes.

The guiding principles in the design and utilisation of non-natural probes are that these probes must be as intense as possible and cause minimal perturbation to the physical system.<sup>89</sup> Popular non-natural IR probes include nitriles,<sup>90</sup> cyanates,<sup>133</sup> thiocyanates<sup>134</sup> and azides.<sup>93</sup> Such probe moieties have been successfully used in recent biological studies,<sup>96,135,136</sup> providing new insights into the studied systems. These studies use standard linear IR techniques, such as FT-IR, and non-linear techniques, such as femtosecond IR pump-probe and 2D-IR.

Recently a multitude of studies into using the azide ion and organic azides as non-natural probes have been carried out.<sup>99,137,138</sup> The studies have focused on these azide probes as the asymmetric azide stretch is known to occur at approximately  $2100\text{ cm}^{-1}$  for both the organic and ionic azides. None of the natural vibrations of proteins, peptides or DNA occupy this region of the IR spectrum and water only has a small combination band in this region.<sup>79</sup> This means that the asymmetric azide vibrations occupy a transparent window in many biological systems. Together with the fact that the asymmetric azide vibration has a large extinction coefficient and is highly sensitive to its local electrostatic environment means that it is ideally suited to being used as a non-natural IR probe.<sup>139</sup>

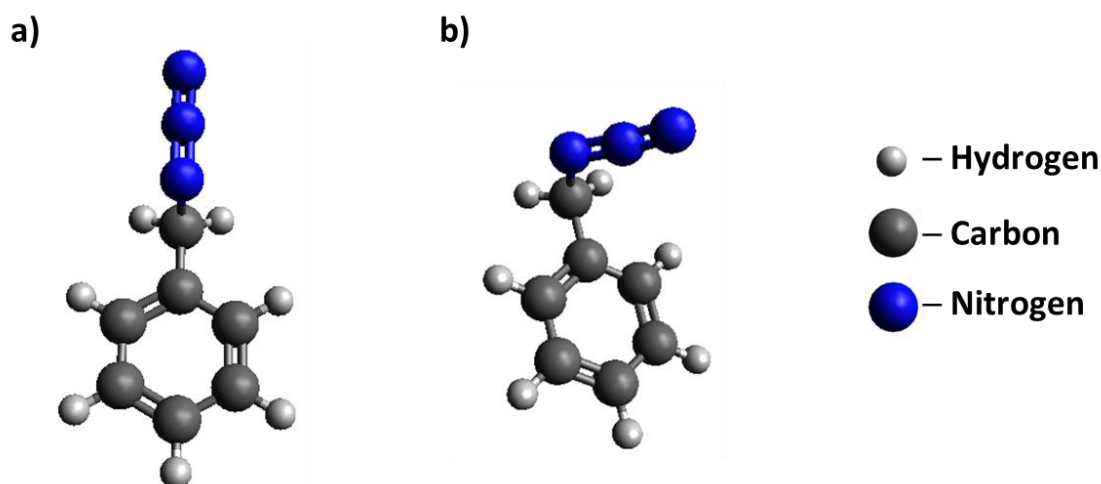
The development of azido amino acid derivatives, such as  $\beta$ -azidoalanine, has led to a plethora of IR studies being carried out with azido probes.<sup>140,141</sup> Two general methods of integrating these probes into the biological systems have been developed. One method involves substituting natural subunits with their azide derivatives. This methodology has been successfully utilised to study the internal environment of the of amyloid peptide aggregates<sup>142</sup> and understand the temporal order of the spectrally distinct structural intermediates that rhodopsin,<sup>143,144,145</sup> a transmembrane light receptor, passes through upon external stimulation.<sup>146</sup> The other method of introducing probes involves synthesising azido derivatives of the ligand that bind to the protein in question. This methodology has the potential to allow the interactions within biomacromolecular complexes to be studied.<sup>108,109</sup> Notably, this methodology has additionally been used to develop non-natural probes, such as an azido-NAD<sup>+</sup> cofactor derivative,<sup>147</sup> allowing a single probe to be used to probe the binding sites of a wide-range of different proteins.

Here, a range of IR spectroscopies are utilised to characterise the response of the asymmetric azide stretch of an archetypal azide, of which benzyl azide is our chosen exemplar, to changes in its local environment. The FT-IR spectroscopy of the asymmetric azide stretch of this azide was determined to have a complicated line-shape including both a low and high frequency shoulder. The positions and intensities of these shoulders were found to be strongly correlated to the electrostatic potential ( $\mu_{\text{Solvent}}$ ) of the local environment surrounding the probe. Whereas the position of the central azide stretch was found to be characteristic of the hydrogen bonding strength ( $\delta H_{\text{Solvent}}$ )<sup>148</sup> of the local environment. Additionally the vibrational relaxation of the asymmetric azide stretch was found to significantly slowed upon decreasing local electrostatic potential whereas changes in the hydrogen bonding strength of the local environment resulting in an increase and decrease in the lifetimes of the low and high frequency shoulders respectively. Finally the 2D-IR response of the asymmetric azide stretch of benzyl azide allowed the shoulders to be identified as Fermi resonances,<sup>149,150,151</sup> and it was noted that the alterations in the exact nature of the local environment led to changes in the coupling constants observed between these Fermi resonances and the central azide stretch. This demonstrates the sensitivity of benzyl azide to act as a non-natural probe of the local molecular environment. This work is carried out with a view to utilising the asymmetric azide stretch as a non-natural IR probe to interrogate the molecular environment in more complex systems, such as small molecule-DNA complexes.

### 3.3 Results and Discussion

#### 3.3.1 Quantum Chemistry Calculations

*Ab initio* chemical calculations allow a variety of information about molecules to be accessed by the application of quantum mechanics. Here, Density Functional Theory (DFT) is used to calculate the optimised geometry, the potential energy surface and simulate the IR spectrum of benzyl azide. This methodology calculates the electron density of the molecule by assuming that there exists a one-to-one correlation between the energy of the molecule and the electron density, as shown by Hohenburg and Kohn.<sup>152</sup> This electron density function is then used to calculate all the other properties of the molecule being studied.<sup>153</sup> All of the calculations performed in this chapter were carried out using the 6-311+G(d,p) basis set and implementing the DFT B3-LYP functional, using the Gaussian09 software.<sup>154</sup> In order to understand the IR spectroscopy and gain an initial idea of the responsiveness of the asymmetric azide stretch of benzyl azide as a non-natural probe quantum calculations were performed. Utilising these calculations on benzyl azide and its ground state revealed the presence of two distinct stable rotamers. The structural difference between these two distinct rotamers is the position of the azide moiety of the molecule. The optimised structural geometry of these two rotamers of benzyl azide are shown in figure 3.1.

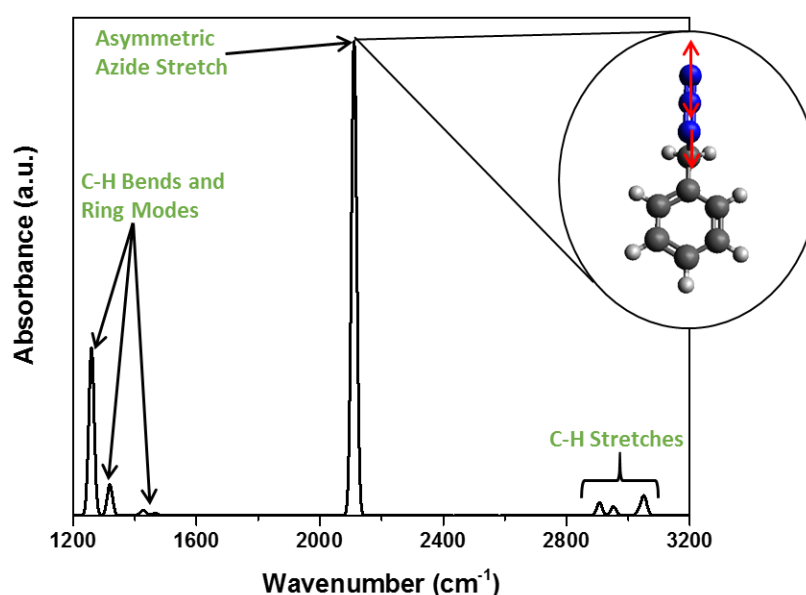


**Figure 3.1:** Geometry optimised structures of two different rotamers of benzyl azide referred to as the a) open rotamer and b) closed rotamer.<sup>154</sup>

One of the rotamers, referred to as the open rotamer (Fig. 3.1.(a)), has the azide moiety positioned at  $118^\circ$  to the benzene ring. This means that the azide moiety is positioned pointing away from the ring and into the environment surrounding the molecule. In the other rotamer, referred to as the closed rotamer (Fig. 3.1.(b)), the azide moiety essentially points back towards the benzene ring and as such would be expected to be partially shielded from



the environment surrounding the molecule by the rest of the molecule. From these calculations it was noted that the energy difference between these two rotamers was  $7.9 \text{ kJ mol}^{-1}$ , and the energy barrier associated with the interconversion between these two forms was determined to be  $11.3 \text{ kJ mol}^{-1}$  (with respect to the lowest energy rotamer). Overall this means that at room temperature there is no interconversion between these two different forms as the energy barrier is greater than the available thermal energy ( $2.5 \text{ kJ mol}^{-1}$  at  $298.15 \text{ K}$ ). Using the calculated energy difference between the open and closed rotamers their relative abundances were calculated to be 96% and 4%, respectively at room temperature, assuming a Boltzmann distribution.<sup>155</sup> Using the individual simulated IR spectra of the open and closed conformers as well as their relative abundances at room temperature, the overall IR spectrum of benzyl azide was generated.



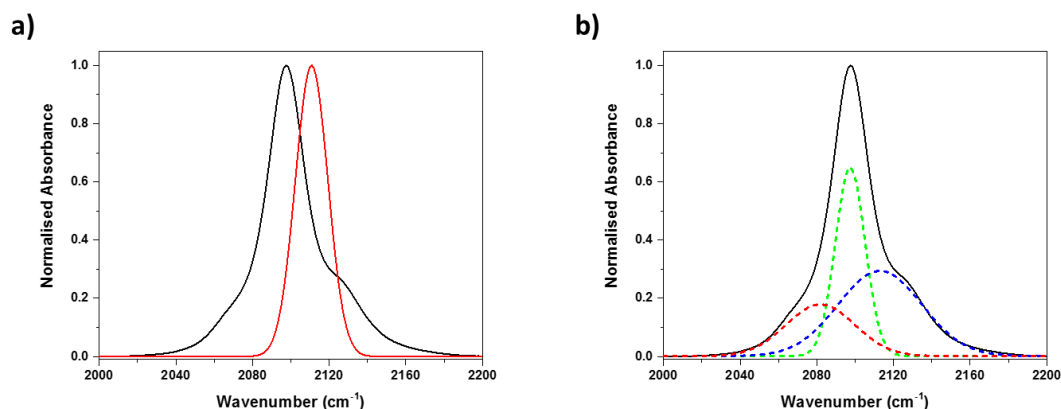
**Figure 3.2:** Simulated and assigned IR spectrum of benzyl azide in DMSO (simulated utilising the Polarizable Continuum Model (PCM)), insert shows the asymmetric azide stretching mode of the dominant rotamer at room temperature.<sup>154</sup>

The simulated and assigned IR spectrum of benzyl azide is shown in figure 3.2. This simulated IR spectrum is dominated by the modes arising from the open rotamer, due to the low relative abundance of the closed rotamer at room temperature. Interestingly, from these calculations it is noted that the predicted vibrational frequency of the asymmetric azide stretch is unaffected by the different molecular geometries seen in the open and closed rotamers. Additionally, the asymmetric azide stretch in these two rotamers are noted to have

very similar dipole moments. From these calculations it can be seen that below  $1600\text{ cm}^{-1}$  the spectrum is predicted to be dominated by various C-H bending, wagging, rocking and scissor modes as well as various ring modes. Above  $2800\text{ cm}^{-1}$  there are a few weak absorbance peaks arising from C-H stretching modes. The  $1660 - 2800\text{ cm}^{-1}$  spectral window is found to contain only one single peak arising from the asymmetric azide stretching mode. The isolation of this mode demonstrates that any changes in the position or line-shape of this asymmetric azide stretch must be due to changes in the local environment around the azide moiety. These calculations show that benzyl azide is an excellent candidate to function as a non-natural IR probe.

### 3.3.2 FT-IR Spectroscopy

A comparison between the experimentally measured (Fig. 3.3.(a), black) and simulated (Fig. 3.3.(a), red) IR spectra of the asymmetric azide stretching mode of benzyl azide is presented in figure 3.3.(a).



**Figure 3.3:** a) Comparison between the experimentally measured (black) and simulated<sup>154</sup> (red, corrected using a basis set specific precomputed vibrational scaling factor)<sup>156</sup> FT-IR spectrum of the asymmetric azide stretching mode of benzyl azide and b) fit of the experimentally measured asymmetric azide stretch of benzyl azide. All presented spectra were measured/simulated in DMSO utilising PCM.

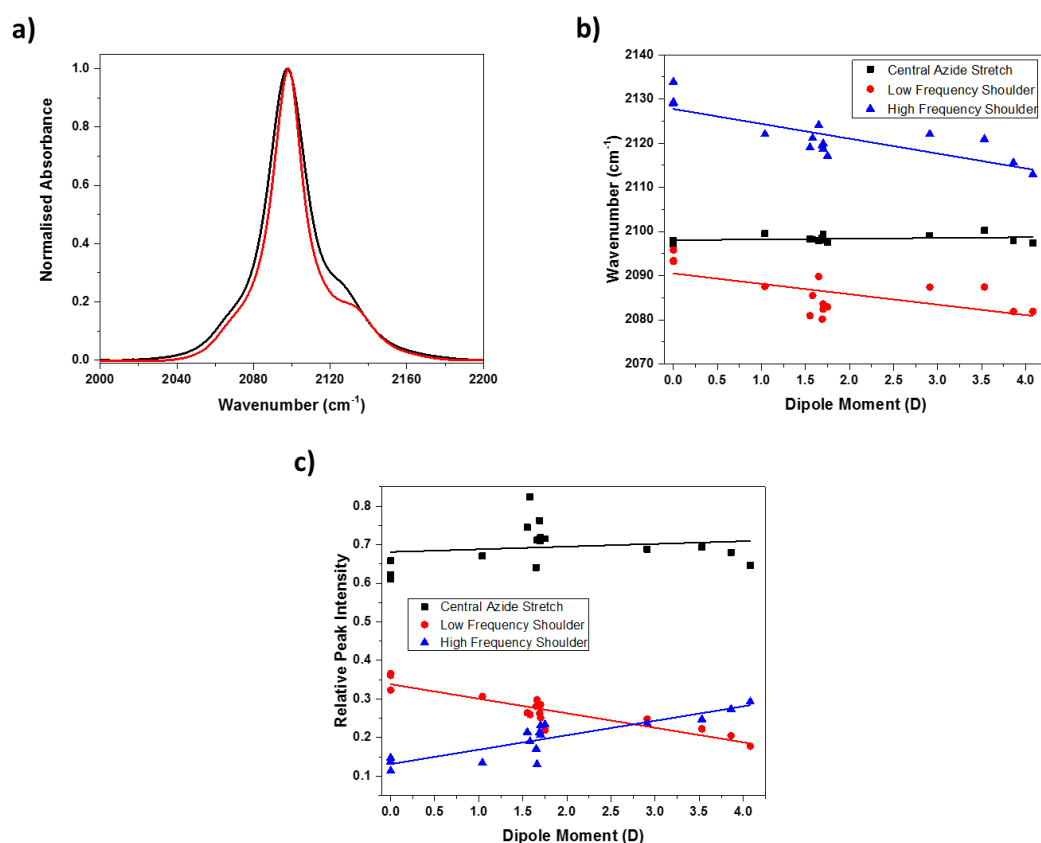
The simulated IR spectrum of benzyl azide predicts that the asymmetric azide stretching mode of this molecule consists of a single Gaussian peak centred at  $2110\text{ cm}^{-1}$ . However the experimentally measured IR spectrum of this mode has a rather distinct line-shape, including a sharp central peak and two shoulders. Fitting the IR spectrum of the azide mode using three Gaussian peaks (Fig. 3.3.(b)) shows that one of these shoulders is shifted to lower frequency by approximately  $20\text{ cm}^{-1}$  and the other shoulder is shifted to higher frequency by approximately  $15\text{ cm}^{-1}$ , when compared to the position of the sharp central peak. Due to the

fact that the 2000–2200  $\text{cm}^{-1}$  spectral window only contains the responses of a few moieties (*e.g.* azide, nitrile, ynamine and thiocyanate moieties) and benzyl azide only contains a single azide moiety it is thought that the sharp central peak can be assigned as the asymmetric azide stretch, as this peak is the dominant feature in FT-IR spectrum. While it has been noted in previous IR studies characterising azides as non-natural IR probes that they often have complicated line-shapes the appearance of these additional shoulders is expected to complicate the characterisation of this particular azide, as well as adding an additional layer of complexity to the IR spectroscopy. The initial problem is that it is difficult to assign these peaks, as it is possible that they arise from a variety of sources including inter-molecular interactions, overtones of vibrational modes, combination band of vibrational modes or Fermi resonances. Fermi resonances are known to be a highly specific example of intramolecular coupling where a normally weak overtone or combination happens to be resonant with a nearby fundamental mode, allowing the weak band to become brighter by harnessing intensity from the fundamental mode.<sup>149</sup> It is noted that there are a few notable examples of Fermi resonances complicating the modes occurring in the 2000  $\text{cm}^{-1}$  – 2200  $\text{cm}^{-1}$  spectral window.<sup>157, 158, 159</sup>

While it not possible to exclude the possibility of the shoulders being caused by the presence of overtones, combination bands or Fermi resonances without the investigation of the IR response of a range of isotopically labelled benzyl azide derivatives, it is possible to determine if one or both of the shoulders are caused by inter-molecular stacking interactions between multiple benzyl azide molecules in solution by decreasing the sample concentration. However it was found that when the concentration of benzyl azide was reduced from 250 mM to 1 mM the FT-IR line-shape was unaffected (Fig. A3.1). This strongly suggests that the shoulders observed in the spectrum of the asymmetric azide stretch are not caused by inter-molecular interactions, such as  $\pi$ -stacking interactions between the benzene rings in two separate molecules, as the 250-fold decrease in the concentration would be expected to lead to the reduction of the intensity of the shoulders as the individual molecules are less likely to be in close proximity to each other. This means that it is not possible to assign these shoulders utilising only the FT-IR spectroscopy of benzyl azide.

To start the characterisation of the asymmetric azide stretch in an archetypal aromatic azide to its local environment the solvent surrounding the benzyl azide molecules was altered. This allows certain physical parameters of the local environment to be changed, one such

parameter is the electrostatic potential that the azide moiety experiences, a computational study has predicted that the asymmetric azide stretch is sensitive to such changes in the local environment. The electrostatic potential of the local environment can be altered by changing the dipole moment of the solvent. The overall response of the asymmetric azide stretch to a change in the dipole moment of its local environment can be seen by comparing the absorbance band in heptane (Fig. 3.4.(a), red), which has no dipole moment,<sup>160</sup> and in DMSO (Fig. 3.4.(a), black), which has a dipole moment of 4.1 D.<sup>160</sup> Comparing these responses it can be seen that increasing the dipole moment of the solvent, used to prepare the sample, leads to changes in the entire line-shape of the IR absorbance. Most noticeable is the increase of the intensity of the high frequency shoulder with increasing solvent dipole moment. In addition to this, the low frequency shoulder seems to have shifted to an even lower frequency, resulting in an apparent increase in the width of the central azide stretch. Due the appearance of differences in the line-shape the asymmetric azide stretch was measured in a range of solvents with different dipole moments, and the observed bands were fitted using three Gaussian peaks, as shown in Fig. 3.3.(b). The peak positions and relative intensities obtained by this fitting procedure were plotted against the dipole moment of the solvent used in the measurements as shown in figure 3.4.



**Figure 3.4:** a) FT-IR spectra of the asymmetric azide stretch of benzyl azide dissolved in DMSO (black) and heptane (red), b) position of the central azide stretch (black), low (red) and high (blue) frequency shoulders as the dipole moment of the local environment increases; c) relative intensities of the central azide stretch (black), low (red) and high (blue) frequency shoulders as the dipole moment of the local environment increases.

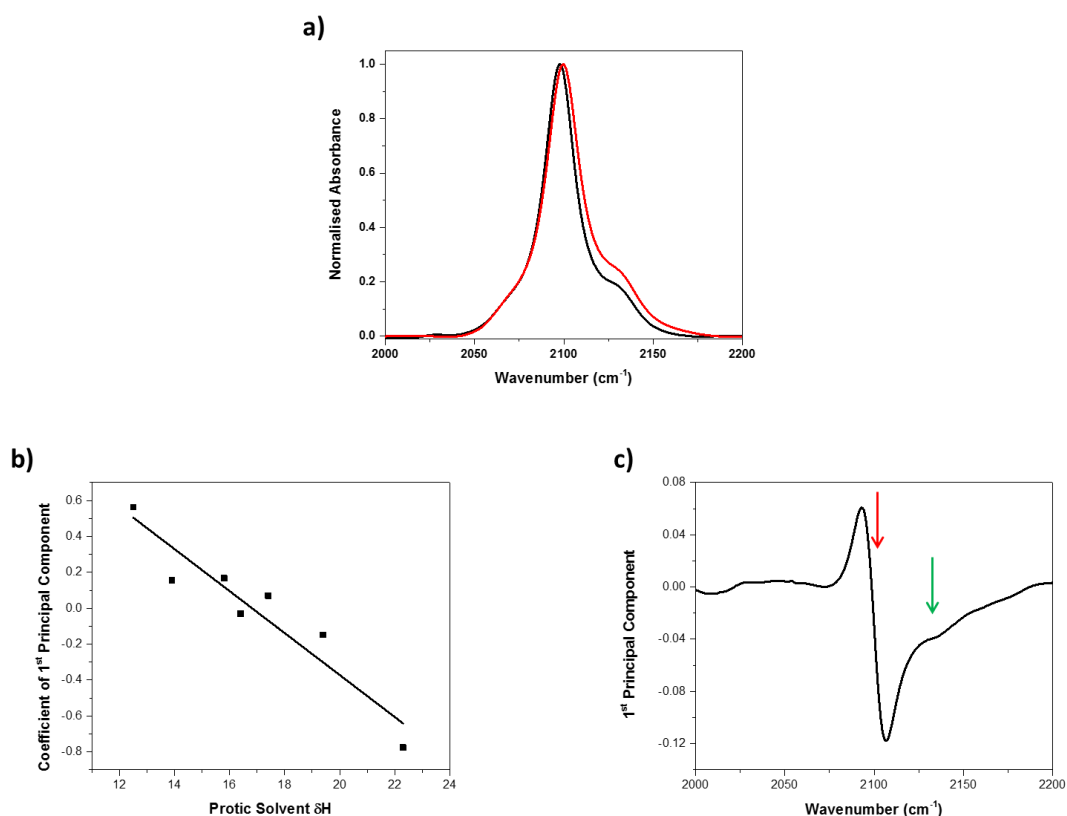
The positions (Fig. 3.4.(b)) and relative intensities (Fig. 3.4.(c)) of the central azide stretch (black), high frequency (blue) and low frequency shoulders (red) as the dipole moment of the surrounding local environment is gradually increased highlights how this local parameter affects each component of the absorbance band. In the case of the central azide peak, altering the dipole moment of the local environment appears to have no significant impact on the position of this mode (Fig. 3.4.(b), black). In addition to this it can be seen that there is no clear correlation between the relative intensity of the central azide peak and the dipole moment of the local environment (Fig. 3.4.(c)). This lack of any solvatochromic response of the asymmetric azide stretching mode of benzyl azide as the dipole moment of the solvent is altered demonstrates one of the differences between aromatic and aliphatic azides as a non-natural IR probe, as the latter group has been found to undergo a significant amount of solvatochromism as the electrostatic environment around the azide moiety is changed. In addition to highlighting this key difference, this also illustrates that the changes in the entire

line-shape of the azide absorbance band of benzyl azide as the dipole moment of the environment is altered is driven by changes in the low and high frequency shoulders.

In the case of the low frequency shoulder of the azide absorption band, as the dipole moment of the surrounding environment increases the shoulder shifts to lower frequency (Fig. 3.4.(b), red) and the relative intensity of the shoulder decreases (Fig. 3.4.(b), red). For the high frequency shoulder of the azide absorption band is found to shift to lower frequency (Fig. 3.4.(b), blue) and undergoes a hyperchromic shift (Fig. 3.4.(c), blue) as the local environment dipole moment increases. These changes in the shoulder of the absorption band alongside the lack of any changes in the central azide peak can explain the changes in to the entire line-shape due to changes in the dipole moment of the local environment. The shift to lower frequency of both shoulders leads to the appearance of the central azide stretch, and the change in the intensity of the of the high frequency shoulder results in the relatively large increase in this part of the azide absorption band of as the local environments dipole moment increases. This means that using the linear IR response of this azide moiety to determine the dipole moments of its local environment is possible but not trivial.

In addition to the importance of the electrostatic potential in the environment surrounding the azide probe, hydrogen bonding also plays an important role in biological systems. The detection of the hydrogen bonding strength of the local environment by non-natural IR probes has been widely studied as a result. However obtaining an accurate quantitative measure of the hydrogen bonding strength for a wide range of local environments is difficult. The most effective measure of this property of the local environment is to consider its hydrogen bonding solubility parameter ( $\delta H$ ), a measure of the stabilisation energy provided to the environment by its ability to form hydrogen bonds.<sup>148</sup> The overall response of the azide absorption band to a change in the hydrogen bonding strength of its local environment can be seen by comparing the absorbance band in 1-hexanol (Fig. 3.5.(a), black), which has a  $\delta H$  value of 12.5, and in methanol (Fig. 3.5.(a), red), which has a  $\delta H$  value of 22.3.<sup>148</sup> Comparing the azide absorption band of benzyl azide in these two environments shows that as the hydrogen bonding strength of the surrounding environment increases the central azide peak undergoes a small  $2\text{ cm}^{-1}$  shift to higher frequency and the high frequency shoulder undergoes a hyperchromic shift. Interestingly the low frequency shoulder appears to be unaffected by this change in hydrogen bonding strength. Due the appearance of differences in the line-shape the asymmetric azide stretch was measured in a range of solvents with

different hydrogen bonding strengths, and the observed bands were fitted using three Gaussian peaks, as shown in Fig. 3.3.(b). The variation in the relative intensities and positions of these three Gaussian peaks are shown in the appendix (Fig. A3.2(a) and A3.2(b), respectively). These changes in the peak properties show no definitive trend with increasing hydrogen bonding strength in its local environment even though there are clear changes in the line-shape of the absorbance. This highlighting the difficulty associated with extracting data accurately by fitting the azide absorption of benzyl azide. In order to determine whether it is possible to extract the hydrogen bonding strength of the local environment surrounding the azide moiety principal component analysis was applied to the azide absorption band as the hydrogen bonding strength is increased. The results of this analysis are shown in figures 3.5.(b) and 3.5.(c).

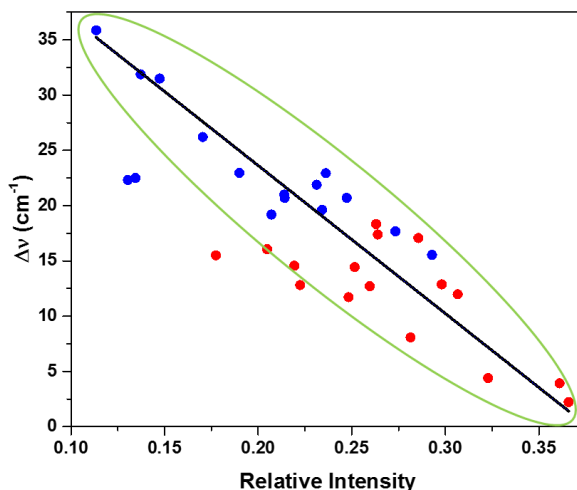


**Figure 3.5:** a) FT-IR spectra of the asymmetric azide stretch of benzyl azide dissolved in 1-hexanol (black,  $\delta H = 12.5$ ) and methanol (red,  $\delta H = 22.3$ ), b) coefficient of the 1<sup>st</sup> principal component as the hydrogen bonding strength of the local environment increases and c) 1<sup>st</sup> Principal component extracted from the principal component analysis carried out on the asymmetric azide stretch as the hydrogen bonding strength of the local environment increases.

In principal component analysis the covariance matrix between all of the selected spectra is calculated. The eigenvectors and eigenvalues are then determined for this covariance matrix. Each of the eigenvectors is a principal component of the data and its associated eigenvalue describes the percentage of the overall variance, within the dataset, described by the principal component. In this way the changes in the spectrum originating from different changes in the system (e.g. the electrostatic potential and hydrogen bonding strength of the local environment) are separated into the different principal components. The coefficient of the 1<sup>st</sup> principal component, isolated via the principal component analysis implementing the covariance matrix methodology, decreases linearly as the hydrogen bonding strength of the surrounding environment is increased (Fig. 3.5.(b)). This linear variation of this coefficient means that it is possible to extract information about the strength of the hydrogen bonds formed in the local environment ( $\delta H$  value)<sup>148</sup> surrounding the azide moiety. In order to gain a deeper understanding of the nature of the changes detected in the azide absorbance band as the hydrogen bonding strength is increased the 1<sup>st</sup> principal component is isolated. This component is shown in Fig. 3.5.(c) and it is noted to contain two features. One of these features, near 2100 cm<sup>-1</sup> is found to consist of a positive peak located to the high frequency side of a negative peak (indicated by a red arrow, Fig. 3.5.(c)). This is indicative of a shift to higher frequency of the central peak in the azide absorption band. The other feature is a negative peak located near 2120 cm<sup>-1</sup>, and this is thought to be consistent with a change in intensity of the high frequency shoulder of the azide absorbance (indicated by a green arrow, Fig. 3.5.(c)). Combining these features present in the isolated 1<sup>st</sup> principal component and the coefficients of this component required to reproduce the azide band at each  $\delta H$  value of the local environment,<sup>148</sup> shows that this component describes a shift to higher frequency of the central peak and an increase in the high frequency shoulder of the azide band. This is fully consistent with the changes to the line-shape of the azide absorbance found by the comparison of the spectrum in the presence of low (Fig. 3.5.(a), black) and high (Fig. 3.5.(a), red) hydrogen bonding strength in the surrounding environment as well as the changes observed as the hydrogen bonding strength of the environment is gradually increased (Fig. A3.3). It is interesting to note that it is only possible to extract a trend from the azide absorbance when the changes in the central band and the high frequency shoulder are tracked simultaneously. Again this shows that it is possible to extract information about the hydrogen bonding strength of the local environment from the FT-IR spectrum of benzyl azide, but again it is found that this is not a trivial process.



As it is noted that the changes in the relative intensities of the shoulders appear to decrease, as the spectral distance to the central peak increases. In order to explore this apparent correlation between the spectral distance to the asymmetric azide stretch and the relative intensities of the two shoulders is further explored in figure 3.6.



**Figure 3.6:** Correlation between the position of both the high (blue) and low frequency (red) shoulders and the distance between the shoulders and the central peak ( $\Delta\nu$ ).

A plot of the relative intensity of the shoulders against the spectral distance to the central peak in the azide absorbance reveals that for both of these shoulders these two properties are negatively correlated, even though it is noted that there are a few outliers (data points located outside the green ellipse, Fig. 3.6). The calculated negative correlation between these two properties was determined to be -0.85, showing that this correlation is significant. This dependency of the relative intensity of both shoulders on the spectral distance to the central peak in the absorbance band is suggestive of coupling between the modes responsible for the appearance of the shoulders and the asymmetric azide stretch (central peak in the absorbance band).

A coupling between these modes and the asymmetric azide stretch is thought to explain the observed behaviours of the relative intensities of the shoulders as the electrostatic potential of the local environment is increased by altering the solvent dipole moment. As the electrostatic potential is increased the vibrational modes of the benzyl azide molecules are affected. From the shift to lower frequency of both of the shoulders suggests that the modes, giving rise to the overtones or combination bands underlying the shoulders, undergo a

vibrational solvatochromic shift due to the change in the electrostatic potential of the local environment. These shifts changes the spectral distance between these, overtones or combination bands, and the asymmetric azide to increase and decrease for the low and high frequency shoulders respectively. The observed changes in the relative intensities of the shoulders can then be explained as these shifts enhancing the coupling between the azide mode and the high frequency shoulder as well as decreasing the coupling between the low frequency shoulder and the azide mode.

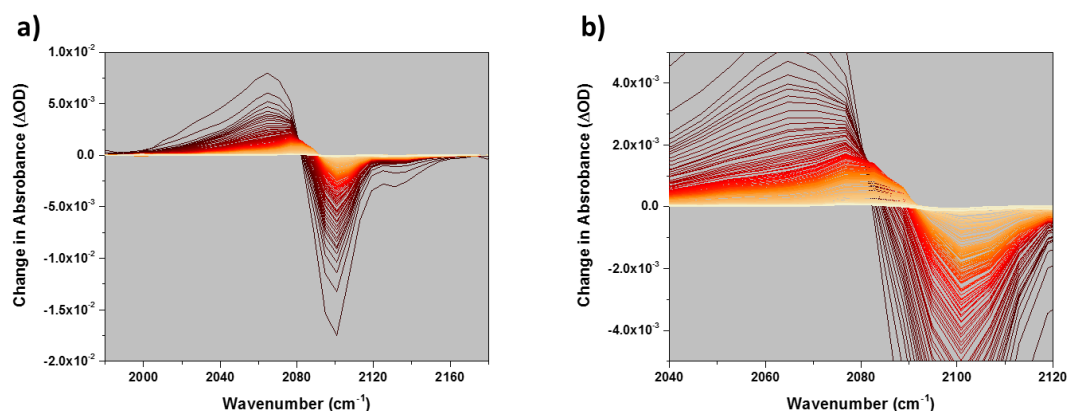
In the case of the change in the hydrogen bonding strength of the surrounding environment it is thought that the changes observed in the overall line-shape of the absorbance band can also be explained by the presence of coupling between the asymmetric azide stretch and the shoulders. In the case of increasing hydrogen bonding strength it is thought that the central peak in the absorbance undergoes a slight shift to higher frequency and the high frequency shoulder shifts to lower frequency. This causes the observed increase in the relative intensity of the high frequency shoulder, however the small shift of the central peak is thought to have only a minimal effect on the relative intensity of the low frequency shoulder. It is interesting to note that it is not easily possible to extract these changes in the line-shape of the azide absorbance band, as the hydrogen bonding strength is increased, by simply fitting Gaussian peaks to this absorbance. Finally it is also noted that only the mode responsible for the overtone or combination band underlying the high frequency shoulder is affected by the increase in hydrogen bonding strength of the local environment.

### **3.3.3 Ultrafast IR Pump-Probe Spectroscopy**

Based on the above it was determined that the FT-IR spectrum of the asymmetric azide stretch of benzyl azide can be used to extract information about the electrostatic potential and hydrogen bonding strength of its local environment. The reliance on the shoulders to extract both parameters means this is a non-trivial task when both the electrostatic and hydrogen bonding strength of the local environment change simultaneously. In order to further characterise the response of the azide absorbance of benzyl azide to its local environment ultrafast IR pump-probe spectroscopy was utilised.

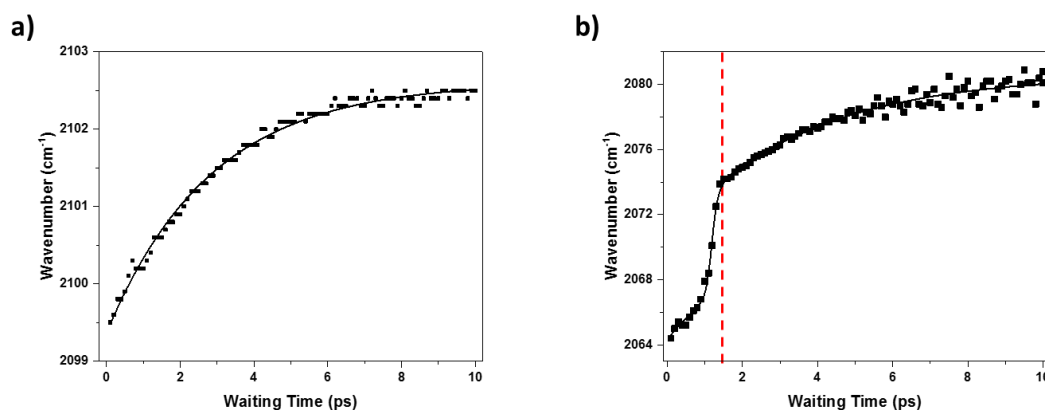
In an ultrafast IR pump-probe spectrum each peak observed in the FT-IR spectrum gives rise to a negative peak located at the same frequency as the peak in the FT-IR spectrum and a positive peak shifted to lower probe frequency by the anharmonicity of the bonding potential. The negative peak often referred to as the bleach, arises from the  $\nu = 0 \rightarrow 1$

transition, and the positive peak often referred to as the excited state absorbance (ESA), arises from the  $\nu = 1 \rightarrow 2$  transition. For the azide absorption band of benzyl azide this means that the overall observed ESA and bleach have contributions from the central peak and both shoulders seen in the FT-IR spectrum. The isotropic IR pump-probe spectrum, calculated from the spectra measured in the ZZZZ and ZZZY polarisation geometries<sup>161,162</sup> to eliminate the effects of the rotation of the molecules in the solution, of benzyl azide in DMSO as the waiting time is increased from 100 fs (black) to 50 ps (white) are shown in figure 3.7.



**Figure 3.7:** Isotropic pump-probe spectra of the asymmetric azide stretch of benzyl azide in DMSO over a range of waiting times from 0 ps to 30 ps are shown a) over the entire probe window and b) over a restricted probe window to highlight the temporal evolution of the pump-probe line-shape.

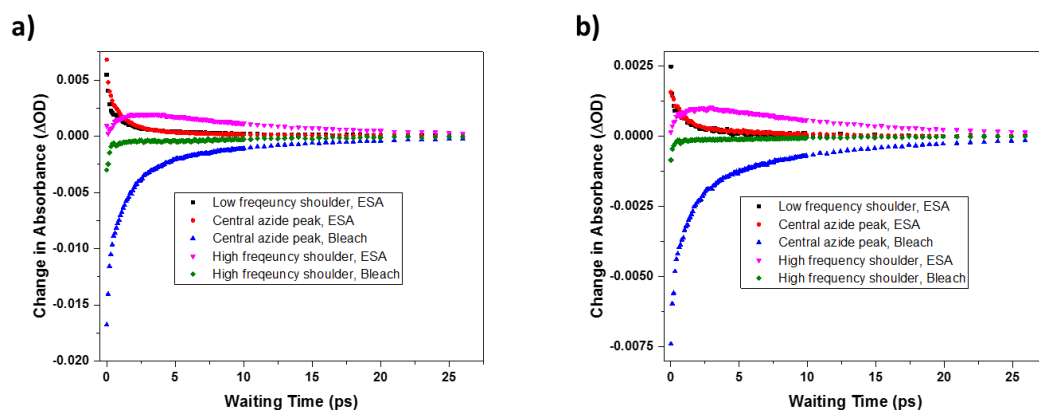
The presence of the two shoulders in the azide absorbance of benzyl azide generate a complicated line-shape for the pump-probe response due to the overlapping ESA and bleaches of the different components of the azide absorption band (Fig 3.7.(a), black). The complexity of the line-shape of the pump-probe spectrum of the azide absorbance band is further increased by the non-trivial evolution of this line-shape with increasing waiting time. It is suspected that this observed temporal evolution of the pump-probe line-shape is caused by the central peak and the shoulders of the azide absorbance band exhibiting different vibrational lifetimes. It is expected that these different lifetimes result in changes to both the minimum of the bleach and the maximum of the ESA as the waiting time gradually increases. The position of the bleach (Fig. 3.8.(a)) and ESA (Fig. 3.8.(b)) of the isotropic pump-probe signal were tracked over a range of waiting times, this analysis aided in the formation of a fitting scheme for these spectra to be developed.



**Figure 3.8:** Temporal evolution of the a) intensity minimum of the bleach and b) intensity maximum of the excited state absorbance of the pump-probe signal of benzyl azide in DMSO.

The positions of the bleach (Fig. 3.8.(a)) and the ESA (Fig. 3.8.(b)) are found to shift to higher frequency as the waiting time between the arrival of the pump and probe pulses at the sample increases. For the bleach, the increase in the frequency is found to shift by approximately  $3 \text{ cm}^{-1}$ . This variation is determined to follow an exponential decay exhibiting a decay rate of 2.8 ps. The change in the frequency and the decay rate can only be fully explained if the lifetime of the high-frequency shoulder of the azide absorption band is significantly longer than the lifetime of the central azide peak. The changes in the position of the ESA (Fig. 3.8.(b)) of the azide mode are found to be more distinct and complicated than the changes in the position of the bleach (Fig. 3.8.(a)) of the response. However, it is noted that the overall change in the position of the ESA is a  $\sim 15 \text{ cm}^{-1}$  shift to higher frequency. At waiting times less than 2 ps the temporal evolution of the position of the ESA can be best described as a sigmoidal curve, suggesting that at approximately 1.2 ps there is a distinct change in the low frequency side of the line-shape of the ESA. Beyond the first 2 ps, the trend of the temporal evolution of the ESA changes to an exponential decay, with a decay rate of 3.6 ps. Together the temporal evolution of the bleach and the ESA position suggest that the different contributions to the azide absorption band have different lifetimes. The initially step change in the position of the ESA position suggest that the low-frequency shoulder has the shortest lifetime, the rapid decay of this component of the ESA of the response leading to the step change in the overall position of the ESA within the first 2 ps. Additionally the long waiting time the shift to higher position of the positions of both the bleach and the ESA positions of the, with decay rates suggests that the lifetime of the high-frequency shoulder is approximately 3-4 ps longer than the lifetime of the central azide peak of the azide absorption band.

These relatively large differences in the lifetimes of the different components which contribute to the absorbance band of the asymmetric azide stretch means that any physically acceptable fitting model must be able to account for each of the components of the azide band. The complex temporal evolution of the position of the ESA of the pump-probe response seems to demonstrate that this part of the line-shape is dominated by the central peak and low-frequency shoulder of the azide absorption band at short waiting times ( $< 2$  ps) whereas at longer waiting times the presence of the high-frequency shoulder becomes increasingly dominant. An initial fitting model of the pump-probe response of the asymmetric azide stretch of benzyl azide consisting of four Gaussian peaks, modelling the low-frequency and central peak contributions to the ESA as well as the central peak and high-frequency contributions to the bleach of the signal. This initial model was successful at recreating the pump-probe response for short waiting times but failed to accurately recreate the line-shape of the ESA as the waiting time increased beyond  $\sim 2$  ps. This failure of the initial model was addressed by introducing an additional Gaussian peak in the ESA to account for the high-frequency shoulder contribution to this part of the pump-probe response. The overall fits produced by this final model at waiting times of 100 fs, 5 ps, 10 ps and 20 ps are shown in Fig.A3.4, and this shows that the finalised model remains successful even as the line-shape of the spectrum changes with waiting time. This model shows that the temporal evolution of the pump-probe line-shape of the asymmetric azide stretch can be attributed to the different contributions changing at different rates. The lifetime data from each of the components of the ESA and bleach of the pump-probe response of the asymmetric azide stretch of benzyl azide in typical aprotic and protic environments, DMSO and 1-butanol, are shown in figure 3.9.



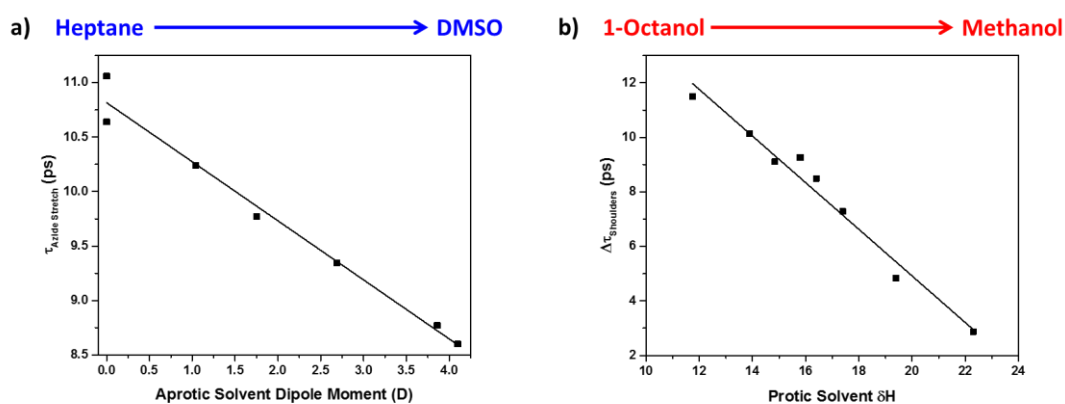
**Figure 3.9:** Isotropic lifetime data of the azide absorption band of benzyl azide in a) DMSO and b) 1-butanol extracted from fitting the ZZZZ and ZZYY pump-probe spectra.

Initially the temporal evolution of the intensities of the components of the ESA and bleach of the isotropic pump-probe response of the azide mode in a typical aprotic (represented by DMSO, Fig. 3.9.(a)) and protic (represented by 1-butanol, Fig. 3.9.(b)) environment appear to be very similar. In both of these types of environments many of the contributions to the azide mode appear to be biexponential in nature. This biexponential nature can be most clearly seen in the decay data of the azide stretch bleach due to the initial rapid decrease in intensity which transitions into a more gradual decay in the signal. One notable exception in both the protic and aprotic environments is the behaviour of the high-frequency shoulder (Figs 3.9.(a) and 3.9.(b), pink) contribution to the ESA of the signal. This contribution initially appears to increase and then it decreases exponentially. This behaviour is seen in both of the protic and aprotic environments and can also be observed by comparing the model fit to the pump-probe signal at a waiting times of 100 fs (Fig.A3.4.(a)) and 5 ps (Fig.A3.4.(b)). However it is thought that this is an artefact introduced by the model, caused by the fact that the ESA contribution of the blue-shifted shoulder strongly overlaps with the bleach of the azide stretch causing it to be obscured at short delay times. This overlap would give the appearance that the initial intensity of the blue-shifted shoulder is much smaller than its actual value due to the large intensity of the azide stretch bleach. This idea is further supported by the fact that these experiments were carried out using broadband pulses. The width of these pulses is known to be approximately  $300\text{ cm}^{-1}$  and so they are wider than the signal meaning all transitions should be excited by the same pulse.

Even though the vibrational dynamics of the benzyl azide are very similar in both aprotic and protic environments it was found that the decay of the azide stretch contribution to the ESA

decayed via a single exponential. This behaviour was only observed for data recorded in protic solvents and so it is possible that the ultrafast dynamics of this probe could also reveal whether the local environment is hydrophobic or hydrophilic. However, this difference is only subtle and it is thought it would not be possible to determine the exact nature of the environment in a more complicated case where there are both hydrophilic and hydrophobic components.

The vibrational relaxation of the components of the azide absorption of benzyl azide as the electrostatic potential or the hydrogen bonding strength of the local environment around the azide moieties were altered. Changes in the electrostatic potential or hydrogen bonding strength of the local environment results in changes in all of the lifetimes of each of the contributions to the azide absorbance band. In the case of the changes in the vibrational lifetimes of the low and high frequency shoulders of the azide absorbance do not vary in a way which allows the electrostatic potential of the local environment to be determined, this is shown in Fig.A3.5. However, it was noted that the lifetime of the central peak of the azide absorption band is found to decrease linearly with increasing dipole moment, and so increasing electrostatic potential, of the local environment surrounding the azide moiety of benzyl azide (Fig. 3.10.(a)).



**Figure 3.10:** a) Change in the lifetime of the asymmetric azide stretch as the dipole moment of the local environment is gradually increased and b) the difference between the lifetimes of the high and low frequency shoulders as the hydrogen bonding strength of the local environment gradually increases.

This linear decrease in the lifetime of the central peak of the azide absorbance is consistent with a previous study demonstrating that increasing the electrostatic potential in the local environment led to a decrease in the vibrational lifetime of a thiocyanate probe.<sup>163</sup> Additionally the relatively large change in the vibrational lifetime of the asymmetric azide

stretch of benzyl azide as the electrostatic potential of an aprotic environment changes demonstrates that this IR probe is usefully for exploring the properties of such aprotic environments.

Increasing the hydrogen bonding strength of the local environment, by increasing the  $\delta H$  value of the solvent,<sup>148</sup> results in linear changes in the vibrational lifetimes of all of the contributions to the azide absorption band of benzyl azide (Appendix, Fig.A3.6). In the case of the central peak it is noted that this decreases linearly with increasing hydrogen bonding strength of the surrounding local environment (Fig.A3.6, black). It is noted that this trend, in the lifetime of the central peak, is similar to the trend observed for the change in the electrostatic potential of the local environment in Fig. 3.10.(a). Together with the fact that solvents which exhibit a greater hydrogen bonding strength are known to have a larger dipole moment it is thought that the observed change in the central peak is due to the local electrostatic potential in the protic environment. For the low and the high frequency shoulders of the azide mode as the hydrogen bonding strength of the local environment increases the lifetimes of these shoulders increase (Fig.A3.6, red) and decrease (Fig.A3.6, blue), respectively. Unlike the lifetime of the central peak of the azide absorbance band, the lifetimes of the low and high frequency shoulders that contribute to this band do not vary linearly with increasing dipole moment (Fig.A3.5.(a), red and blue) and so the linear variation of the lifetimes of these components are thought to be caused by the increase in hydrogen bonding strength of the local environment. It is interesting to note that as the hydrogen bonding strength of the local environment increases the lifetime of the high-frequency shoulder increases whereas the lifetime of the low-frequency shoulder decreases. This suggests that the two modes underlying overtones or combination bands which result in these shoulders are affected in very different ways by the formation of hydrogen bonding interactions. It is thought that the formation of hydrogen bonds to the benzyl azide molecule can be used to explain the opposing behaviours of these two components by understanding how the formation of these bonds affect the two mechanisms by which vibrational mode undergo relaxation after excitation, namely by the transfer of energy to the low energy solvent bath modes or to the low energy molecular modes.

In the case of the high-frequency shoulder the decrease in the lifetime is thought to be consistent with the formation of direct hydrogen bonds to the underlying mode. It is expected that such interactions between the solute and solvent primarily enhances the

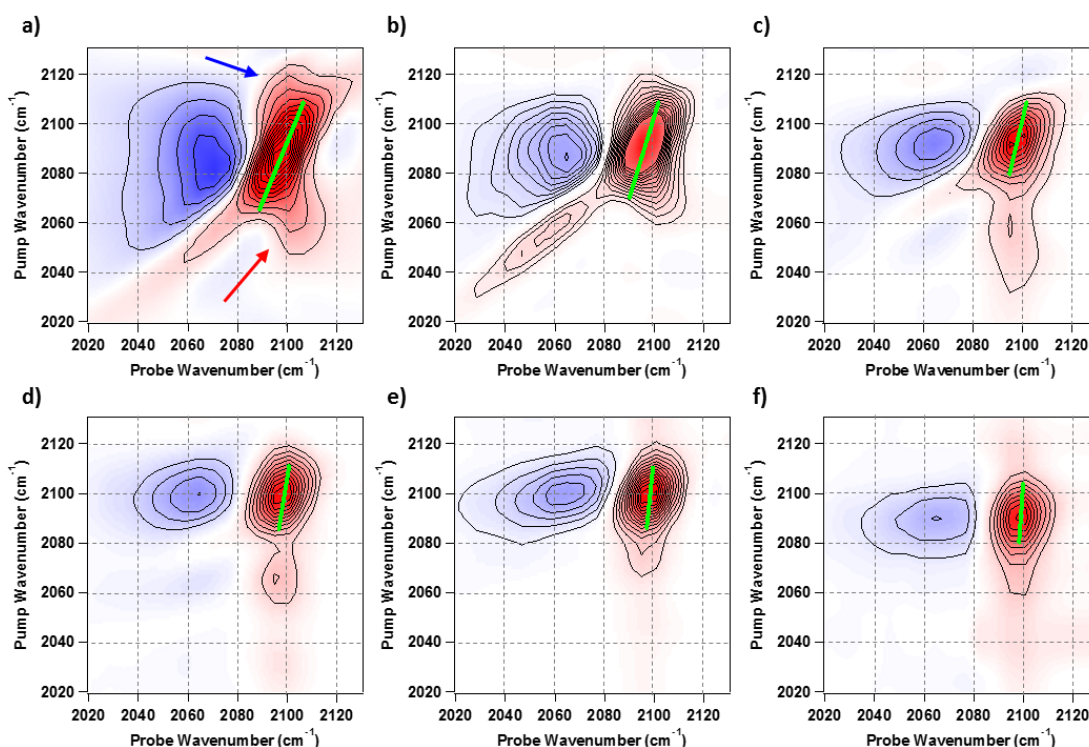


efficiency of the transfer of the energy from this mode to the solvent bath modes, increasing the rate of relaxation and so decreasing the lifetime of the vibrational mode.<sup>164</sup> Therefore the stronger the hydrogen bonds formed the more rapid the vibrational relaxation of the vibrational mode. In contrast to this it is thought that the effect of the formation of hydrogen bonding interactions to benzyl azide result in an increase in the lifetime of the low frequency shoulder. It is thought that this is due to the formation of these hydrogen bond interactions resulting in the shifting of the low energy molecular bath modes, which often involve relatively large contributions from the ring modes, decreasing the efficiency of the energy transfer mechanism to these bath modes thus increasing this vibrational lifetime. Finally even though it was noted that the changes in the lifetimes of the low- and high-frequency shoulders of the azide band were linearly correlated with the change in the hydrogen bonding strength of the local environment it was noted that this correlation was improved by utilising the difference between the lifetimes of these two shoulders (Fig. 3.10.(b)). Demonstrating that the pump-probe response of the asymmetric stretch of the benzyl azide can be used to extract information on the electrostatic potential and the hydrogen bonding strength of the local environment around benzyl azide, however it is noted that this requires prior knowledge of the nature of the environment. In order to try and expand upon this and determine a methodology that can be utilised to extract information about the local environment without the need for prior knowledge the azide absorbance of benzyl azide was explored using 2D-IR spectroscopy.

#### **3.3.4 2D-IR Spectroscopy**

In 2D-IR spectroscopy, each peak observed in the FTIR spectrum gives rise to a negative feature (red) located on the 2D-IR spectrum diagonal. These are assigned to the respective  $\nu = 0 \rightarrow 1$  transitions, each with an accompanying, positive (blue),  $\nu = 1 \rightarrow 2$  peak shifted by bonding potential anharmonicity to lower probe frequency. Any peaks occurring in the off-diagonal region result from a range of different effects including coupling between two vibrational modes, exchange between two different isomeric forms or the exchange between different environments present in the sample.<sup>165</sup> Finally it is noted that the polarisation of the pulses within the 2D-IR pulse sequence can be altered from a completely parallel configuration (the ZZZZ polarisation) to the perpendicular polarisation (the ZZZY polarisation) which is known preferentially enhance the off-diagonal features within the 2D-IR spectrum.<sup>166</sup>

The 2D-IR spectrum of the asymmetric azide stretch, of benzyl azide, recorded at a range of different waiting times are presented in Fig. 3.11. At short waiting times (250 fs) the 2D-IR spectrum is observed to contain diagonal features associated with the central peak as well as the low-frequency and high-frequency shoulders (Appendix, Fig. A3.7) previously identified in the azide absorption band of benzyl azide. In addition to these diagonal features the 2D-IR spectrum of benzyl azide contains two distinct off-diagonal cross-peaks, indicated by arrows in Fig. 3.11.(a).

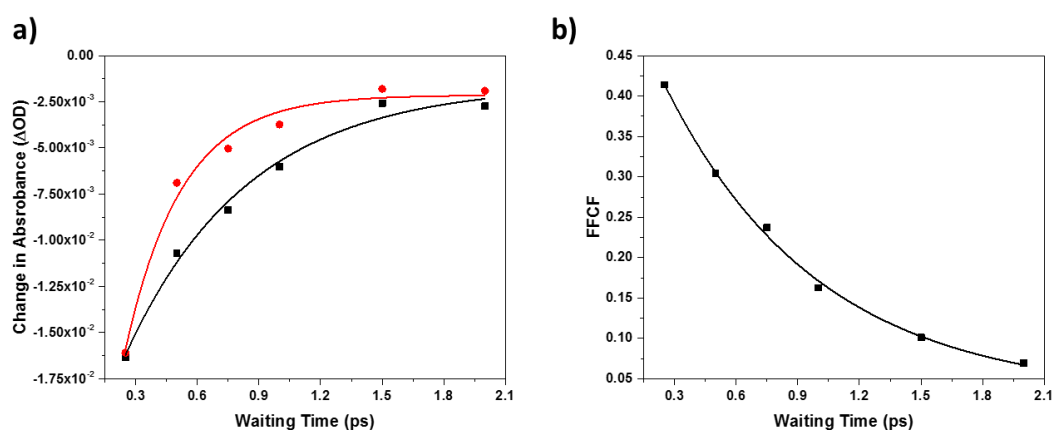


**Figure 3.11:** Waiting time study of benzyl azide in 1-hexanol measured at a) 250 fs, b) 500 fs, c) 750 fs, d) 1000 fs, e) 1500 fs and f) 2000 fs. Central lines indicated in green.

The presence of these cross-peaks, in this 2D-IR spectrum, indicates the presence of coupling or chemical exchange between the modes observed on the spectrum diagonal. Due to the appearance of these features at such early waiting times in the 2D-IR spectrum of the asymmetric azide stretch of benzyl azide it is thought that these features are consistent with the low-frequency and high-frequency shoulders arising from the presence of Fermi resonances between the asymmetric azide stretch and the overtones or combination bands of other molecular vibrational modes. This identification of the origins of the low-frequency and high-frequency shoulders as Fermi resonances is supported by the fact that the intensities of the cross-peaks, located utilising the peaks positions obtained by fitting the FT-

IR spectra of the samples (section 3.3.2, pg. 53), only decrease as the waiting time increases (Fig.3.12.(a)).

As the waiting time of the spectrum is increased the overall line-shape of the 2D-IR spectrum changes, these changes are driven by both homogenous broadening of the line-shape resulting in an increase in the slope of the central line of the spectrum (indicated in green, Fig.3.11.(a)-(f)) and the differences in the lifetimes of the different contributions to the azide absorbance. The overall result of these changes is found to yield a 2D-IR spectrum at long waiting times ( $\geq 1500$  fs) which has a distinctly circular line-shape (Fig.3.11.(f)).

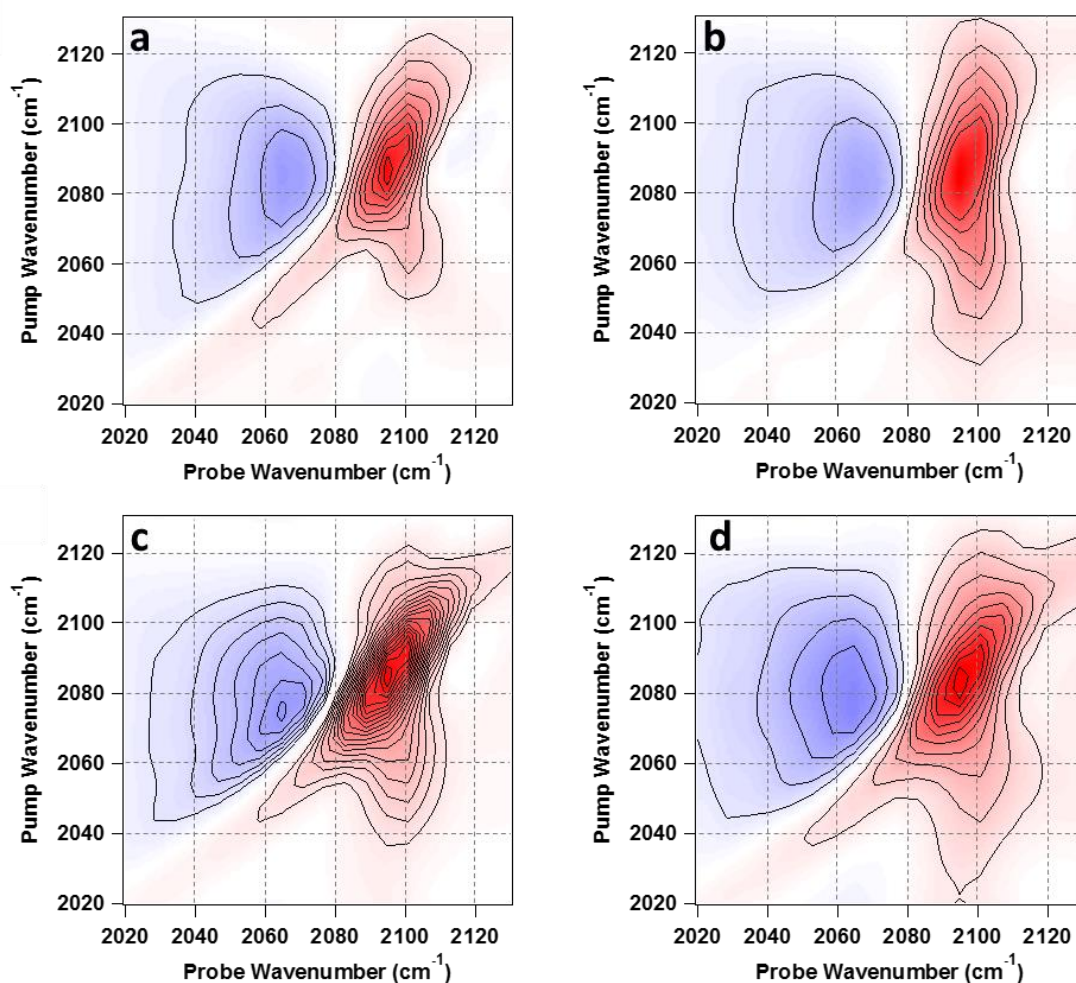


**Figure 3.12:** a) Intensities of the cross-peaks to the low-frequency (black) and high frequency (red) shoulders and b) Spectral diffusion dynamics of benzyl azide in 1-Hexanol presented via the frequency-frequency correlation function, extracted via the central line slope methodology.

The homogenous broadening of the line-shape of the 2D-IR spectrum can be tracked using the central line slope methodology<sup>167,168</sup> allowing the spectral diffusion dynamics of the asymmetric azide stretch to be extracted (Fig.3.12.(b)). Altering the nature of the local environment surrounding benzyl azide was found to change the diffusion dynamics of the asymmetric azide stretch. The different diffusion dynamics of the asymmetric azide stretch are summarised in Table A3.1, in the appendix. In the case of the protic solvents it can be seen that as the environment surrounding the solvent changes from 1-hexanol, which exhibits a relatively weak hydrogen bonding strength (with a  $\delta H$  value of 12.5), to methanol, which exhibits relatively strong hydrogen bonding strength (with a  $\delta H$  value of 22.3), the rate of the observed diffusion dynamics is found to change from  $\sim 750$  fs to  $\sim 390$  fs.<sup>148</sup> It is noted that as the hydrogen bonding strength of the environment increases, the viscosity of the local environment actually decreases. It is thought that this decrease in the viscosity of the local

environment is consistent with the increased rate of the dynamics observed by the asymmetric azide stretch, consistent with the increased rate at which hydrogen bonding networks can rearrange at lower viscosities. In the case of the aprotic environments it is found that as the electrostatic potential of the local environment altered by increasing the solvent dipole moment from 0 D (heptane) to 4.1 D (DMSO) the diffusion dynamics decrease from  $\sim 2.3$  ps to  $\sim 510$  fs. It is noted that in the case of the aprotic environment the change in the viscosity of the local environment is not correlated to the change in the observed diffusion rate. It is noted that as the electrostatic potential of the local environment slowly decreases the dynamics, observed by the azide probe, slow down. This trend is essentially linear with the solvent dipole moment, it is thought that this linear trend continues until the electrostatic potential of the local environment vanishes, when the solvent dipole moment becomes zero (heptane) at which the dynamics rapidly increases to  $\sim 2.3$  ps. It is thought that the initial decrease in rate the dynamics as the electrostatic potential decreases is due to the weaker interactions between the azide moiety and the local environment. The rapid increase as the electrostatic potential of the local environment vanishes is thought to be consistent with the natural spectral diffusion rate of the azide moiety due to the very weak dispersive forces underpinning the interactions between the azide moiety and the environment.

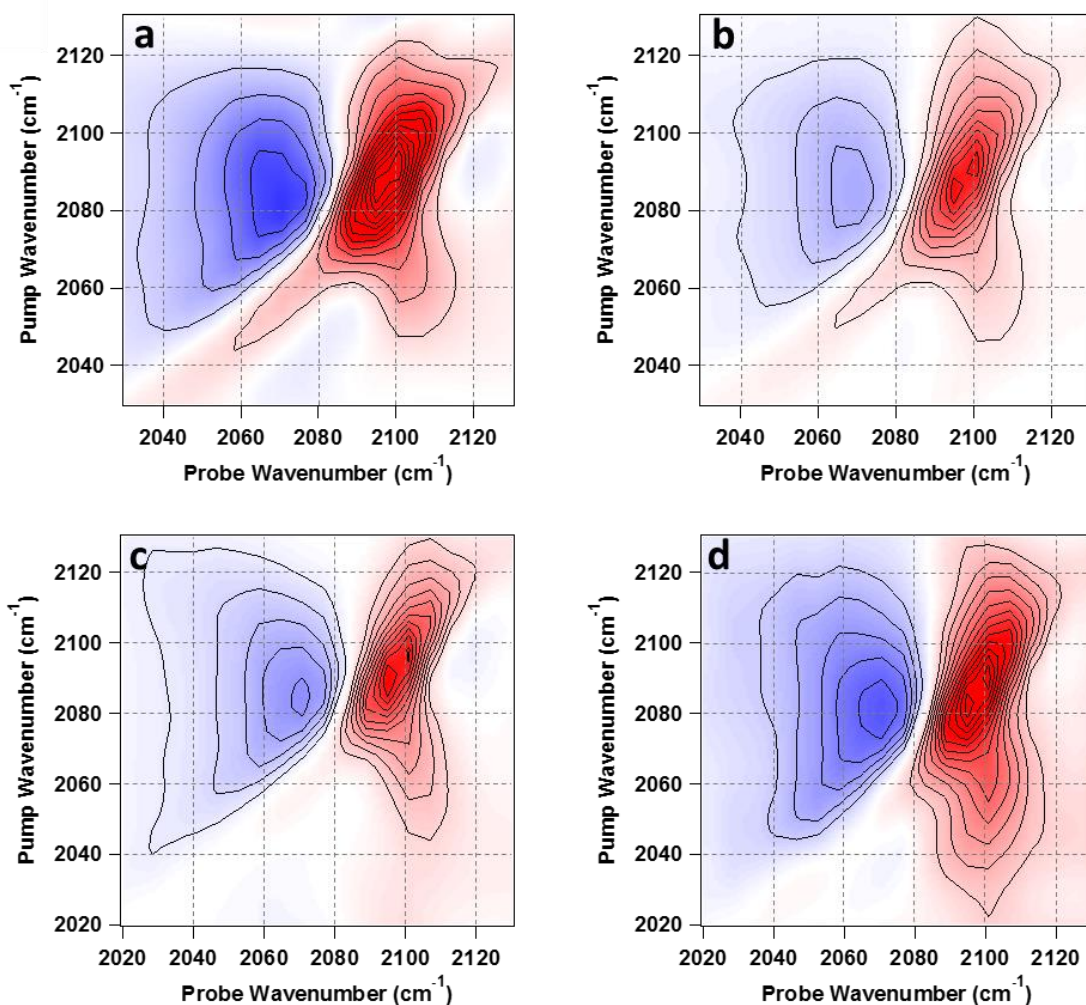
In addition to the changes seen in the diffusion dynamics of the asymmetric azide stretch upon the alteration of the local environment results in changes in the overall line-shape. In the case of protic solvents these changes are demonstrated in Fig.3.13.



**Figure 3.13:** Graphs showing the 2DIR spectrum of Benzyl Azide in Heptane a) ZZZZ polarisation and b) ZZY Y polarisation; Benzyl Azide in DMSO c) ZZZZ polarisation and d) ZZY Y polarisation. All spectra are measured at a waiting time of 250 fs.

It is noted that the overall line-shape changes when the electrostatic potential of the local environment is increased by changing the solvent surround the benzyl azide molecules from heptane (Fig.3.13.(a)&3.13.(b)), with no dipole moment, to DMSO (Fig.3.13.(c)&3.13.(d)), known to exhibit a dipole moment of 4.1 D. These changes can be particularly dominant in the spectra obtained using the perpendicular pulse configuration (Figs.3.13.(b)&(d)). It is noted that as the electrostatic potential of the surrounding environment increases the positions and intensities of these cross-peaks change leading to the observed changes in the global line-shape.

In the case of protic environments the changes to the 2D-IR line-shape as the hydrogen bonding strength of the local environment increases are shown in Fig.3.14.



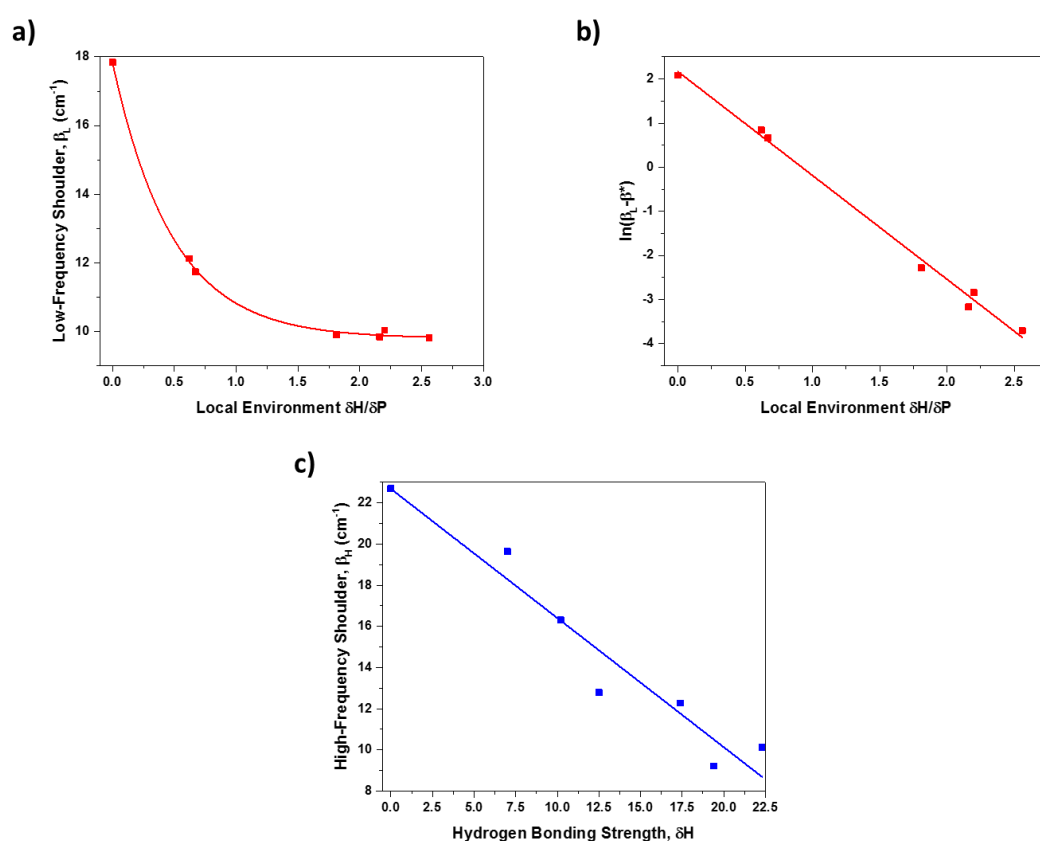
**Figure 3.14:** Graphs showing the 2DIR spectrum of Benzyl Azide in 1-Hexanol a) ZZZZ polarisation and b) ZZZY polarisation; Benzyl Azide in Methanol c) ZZZZ polarisation and d) ZZZY polarisation. All spectra are measured at a waiting time of 250 fs.

As with the response of the 2D-IR spectrum of the asymmetric azide stretch to the electrostatic potential of the local environment, it was noted that the 2D-IR response was also altered by changes in the hydrogen bonding strength of its local environment. As the strength of the hydrogen bonding present in the local environment is increased from those present in 1-hexanol (Figs.3.14.(a)&(b)), with a  $\delta H$  value of 12.5, to those present in methanol (Figs.3.14.(c)&(d)), with a  $\delta H$  value of 22.3, it is noted that majority of the changes to the line-shape involve changes in the cross-peaks.<sup>148</sup> As seen with previously, as the local electrostatic potential was changed, the changes in the line-shape with changing hydrogen bonding strength are driven by alterations in the positions and intensities of the cross-peaks present in the spectrum. It is noted that changes in the cross peaks appear to dominate the



alterations in the line-shape of the asymmetric azide stretch as the nature of the local environment changes.

In 2D-IR spectrum the intensities of such cross peaks between the two separate modes are linked to the dot product between the transition dipole moments of the coupled vibrational modes. However in this case it is thought that as the local environment will not have a large effect on the transition dipole moments of the vibrational modes associated with benzyl azide. This suggests that the alteration in the apparent intensities of the cross-peaks as the interactions with the local environment change are a result of changes in the anharmonicity of the cross-peaks. This indicates that the interactions with the local environment results in changes to the strength of the coupling<sup>169,170</sup> between the asymmetric azide stretch and its two Fermi resonances.



**Figure 3.15:** a) Coupling constant ( $\beta_L$ ) between the asymmetric azide stretch and low-frequency shoulder as the local environment is altered, b) extracted linear calibration curve based on this coupling constant and c) Coupling constant ( $\beta_H$ ) between the asymmetric azide stretch and high-frequency shoulder as the hydrogen bonding strength of the environment increases.

The coupling constants between the asymmetric azide stretch and its two Fermi resonances were extracted from the 2D-IR spectra utilising a methodology derived in a previous study<sup>171</sup> the results of this process are shown in Fig.3.15. In this study, the pump slices of the 2D-IR spectra used to calculate these coupling constants are located utilising the peaks positions obtained by fitting the FT-IR spectra of the samples (section 3.3.2, pg. 53). In the case of the low-frequency shoulders the coupling constant ( $\beta_L$ ) is found to vary exponentially with changes in the overall nature of the local environment (Fig.3.15.(a)). These overall changes in the local environment take into account the different features of the molecular structure of the environment, including but not limited to the hydrogen bonding strength, the electrostatic potential, dielectric constant, the presence of hydrogen accepting and donating groups. All of these features can be accurately encapsulated by the ratio of the enthalpic contributions to the environment from H-bonding type interactions and from electrostatic interactions  $\left(\frac{\delta H}{\delta P}\right)$ .<sup>148</sup> The linearization of this exponential variation (Fig.3.15.(b)) of the coupling between the low-frequency shoulder and the asymmetric azide stretch can therefore be used to extract the exact molecular nature of the local environment surrounding the azide moiety and allowing it to be compared to a particular molecule or molecular family (*e.g.* esters, amides or nitriles). Finally in the case of the coupling ( $\beta_H$ ) between the high-frequency shoulder and the asymmetric azide stretch was found to decrease linearly with increasing hydrogen bonding strength of the local environment (Fig.3.15.(c)).<sup>148</sup> This again demonstrates that the high-frequency is more strongly affected by the formation of hydrogen bonds between benzyl azide and its local environment.

### 3.4 Conclusions

In conclusion, we have characterised the response of the asymmetric azide stretch of benzyl azide to changes in its local molecular environment. The complex line-shape of the azide mode was determined to result from the presence of two Fermi resonances, one centred at a slightly higher frequency than the azide mode and the other centred at a slightly lower frequency than the azide mode. Increases in the electrostatic potential of the local environment results in shifts to lower frequencies of both Fermi resonances, additionally a decrease in the vibrational lifetime of the asymmetric azide stretch was observed. Increases in the hydrogen bonding strength of the local environment resulted in a shift to higher frequency of the asymmetric azide stretch as well as increasing the lifetime of the low-frequency shoulder and decreasing the lifetime of the high frequency shoulder. Additionally



the 2D-IR response of the asymmetric azide stretch illustrated that the coupling constant between the low-frequency shoulder ( $\beta_L$ ) and the azide mode varied exponentially with changes in the nature of the molecular environment, whereas the coupling constant between the high-frequency shoulder ( $\beta_H$ ) and the azide mode was found to decrease linearly with increasing hydrogen bonding strength. All of these changes in the response of the asymmetric azide stretch of benzyl azide could be used to extract relevant molecular information about the local environment, including both the electrostatic potential and the hydrogen bonding strength. The responses of the Fermi resonances as the local environment changes suggested that only the mode underlying the high-shoulder participates in hydrogen bonds. In combination with the simulated IR spectrum this suggests the low-frequency and high-frequency shoulders are the result of a Fermi resonance between the asymmetric azide stretch and overtones of a  $\text{CH}_2$  scissor mode of a ring mode respectively. Overall, the results of the characterisation of the asymmetric azide stretch response of benzyl azide demonstrate that such a moiety is well suited to act as a non-natural IR probe.

## 3.5 Experimental

### 3.5.1 Materials

All of the solvents used were purchased from Sigma Aldrich. Benzyl azide was synthesised using a substitution reaction between 1 equivalent benzyl bromide and 1.5 equivalents of sodium azide. The reaction was carried out in a 4:1 blend of acetone and water, the reaction was heated overnight and subsequently extracted with diethyl ether three times, dried over sodium sulphate and then the solvent was removed (under reduced pressure) using a rotary evaporator.<sup>172</sup> All solvents were used without further purification. Solutions of benzyl azide were prepared in a wide range of protic and aprotic solvents. Once prepared the benzyl azides were thoroughly shaken for at least 10 minutes.

### 3.5.2 IR Spectroscopy

All of the FT-IR experiments were carried out using a Bruker Vertex 70 at a resolution of  $1\text{ cm}^{-1}$ . The samples were held in a demountable Harrick cell utilising  $\text{CaF}_2$  windows and a  $50\text{ }\mu\text{m}$  Teflon spacer. All of the ultrafast pump-probe measurements were carried out using the pump-probe optical paths in the Strathclyde FT-2D-IR spectrometer.<sup>173</sup> All of the 2D-IR measurements were performed using the 2D-IR optical paths in the Strathclyde FT-2D-IR spectrometer. The mid-IR pulses produced by the Strathclyde spectrometer were centred at  $2100\text{ cm}^{-1}$ , with a duration of  $\sim 100\text{ fs}$  at a repetition rate of  $1\text{ kHz}$ .

For the pump-probe measurements the mid-IR pulses, generated using an OPA, are used to generate both the pump and the probe pulses. A variable temporal delay,  $T_w$ , between the pump and probe pulses was then generated using an optical delay stage. The vibrational relaxation of the pump-probe response was then measured, utilising a MCT detector, by scanning  $T_w$ .

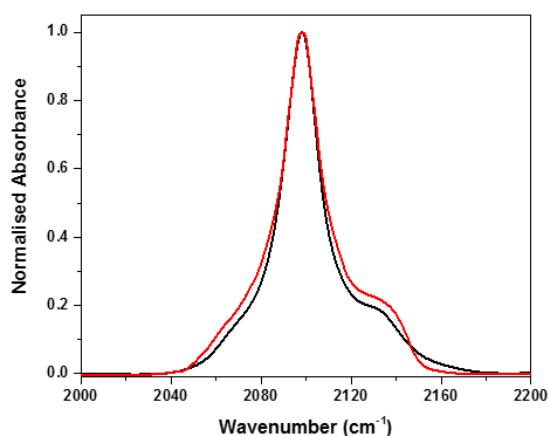
For the 2D-IR spectra, the delays between the pump and probe pulses are still generated using an optical delay stage. The pump pulses are passed through a Mach-Zehnder interferometer to generate the two collinear pump pulses separated by a variable time delay,  $\tau$ , allowing a pseudo-pump probe geometry to be utilised.<sup>174</sup> The 2D-IR data sets were then measured for a range of different waiting times by scanning  $\tau$  at each waiting time. Spectra were then generated by performing a Fourier transform along  $\tau$ .

All of the benzyl azide samples for the FT-IR spectroscopy were prepared at 250 mM, the sample measured at 1 mM was prepared from the original sample via serial dilution. All of the benzyl azide samples for the IR pump-probe and the 2D-IR spectroscopy were prepared at 100 mM.

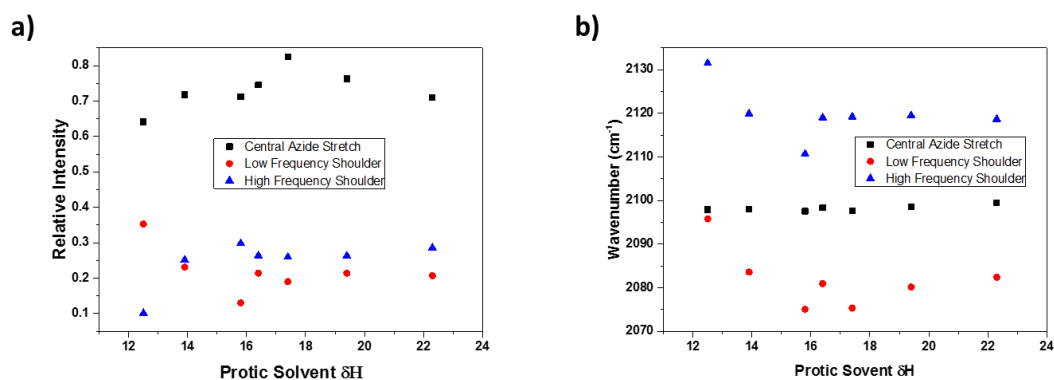
### 3.5.3 Quantum Chemistry Calculations

All of the *ab initio* calculations presented in this chapter were carried out using the Gaussian09 software package. Geometry optimisations and frequency calculations were explored for the benzyl azide molecule. All of these calculations were carried out using the 6-311+G(d,p) basis set and implementing the hybrid-DFT B3LYP functional.<sup>154</sup>

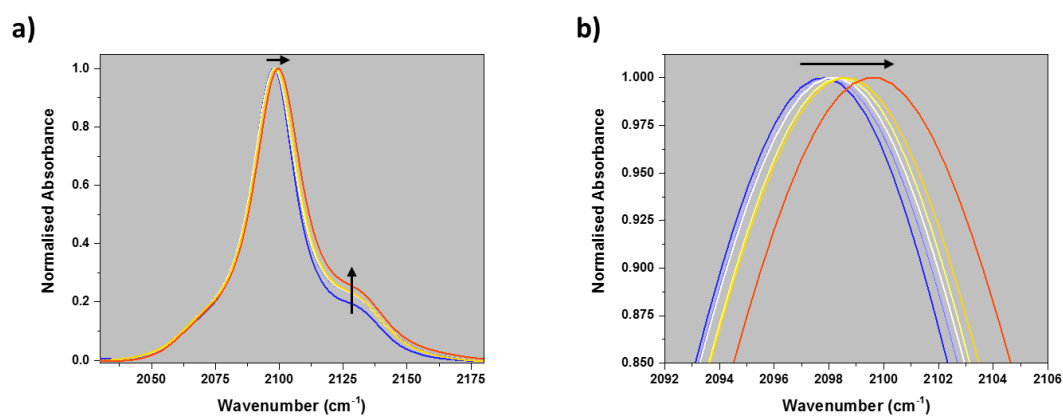
### 3.6 Appendix



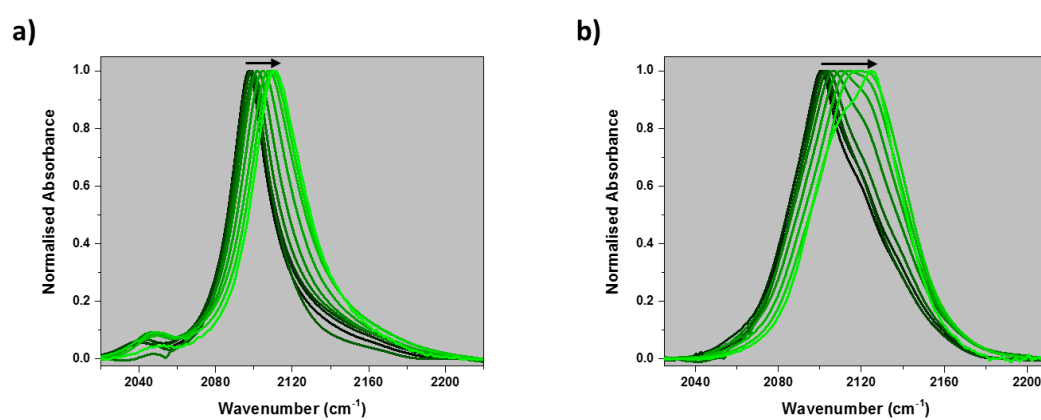
**Figure A3.1:** FT-IR Spectra of the asymmetric azide stretch of benzyl azide measured at 250 mM (black) and 1 mM (red) in hexanol.



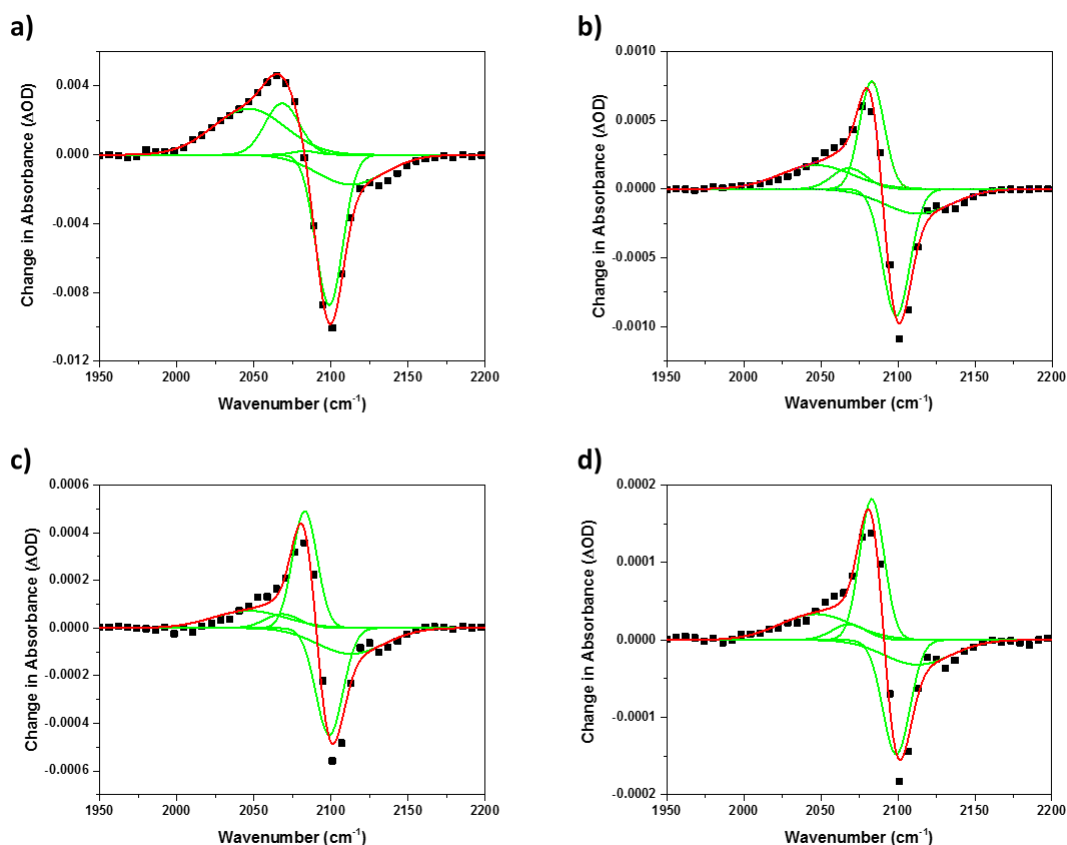
**Figure A3.2:** a) position of the central azide stretch (black), low (red) and high (blue) frequency shoulders as the hydrogen bonding strength of the local environment increases and b) relative intensities of the central azide stretch (black), low (red) and high (blue) frequency shoulders as the hydrogen bonding strength of the local environment increases.



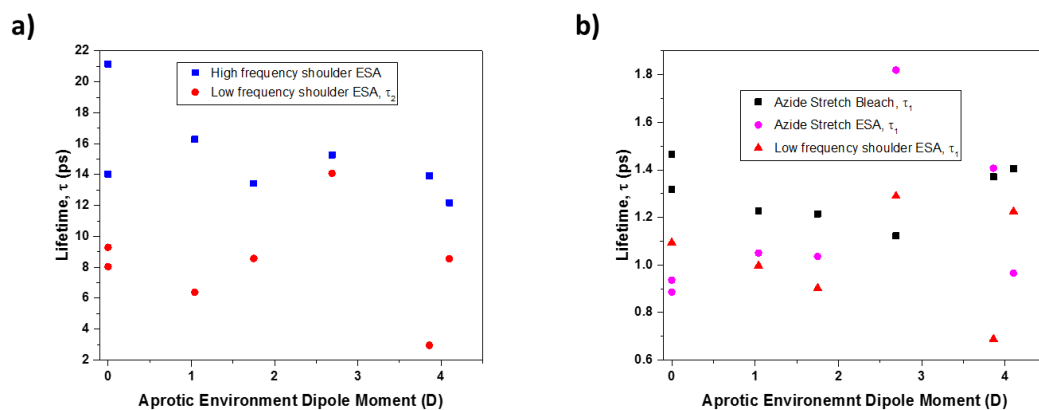
**Figure A3.3:** a) Azide Absorption band of benzyl azide as the hydrogen bonding strength of the local environment increases and b) Asymmetric azide stretch of benzyl azide as the hydrogen bonding strength of the local environment increases.



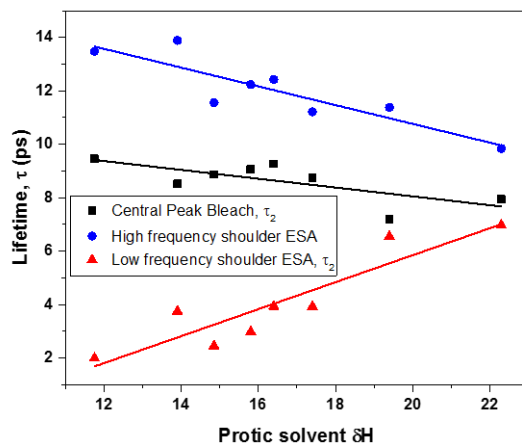
**Figure A3.4:** Overall response of the asymmetric azide stretch of two archetypal aliphatic azides, a) 3-azido-propan-amine and b) 3-azido-propanoic acid, as the hydrogen bonding strength of the local environment is increased.



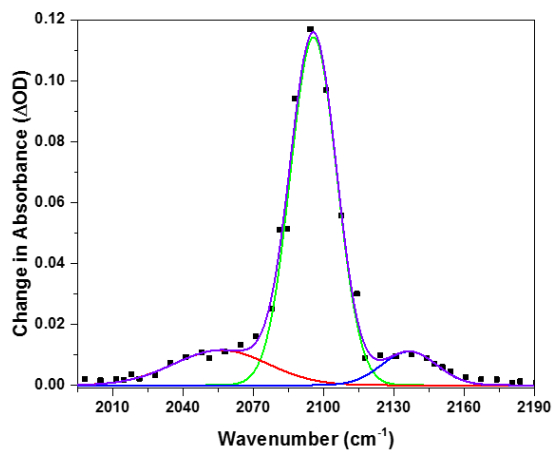
**Figure A3.5:** Fit applied to the ultrafast IR pump-probe spectrum of the azide absorbance band of benzyl azide in DMSO measured at waiting times of a) 100 fs, b) 5 ps, c) 10 ps and d) 20 ps. The individual Gaussian peaks are represented by the green curves, the overall fit is represented by the red curve and the experimentally recorded data points are black.



**Figure A3.6:** Lifetimes of a) the ESA of the high frequency shoulder (blue) and low frequency shoulder ESA,  $\tau_2$  (red) as the dipole moment of the local environment increases; lifetimes of a) the bleach of the azide stretch (black), ESA of the azide stretch (pink) and low frequency shoulder ESA,  $\tau_1$  (red) as the dipole moment of the local environment increases.



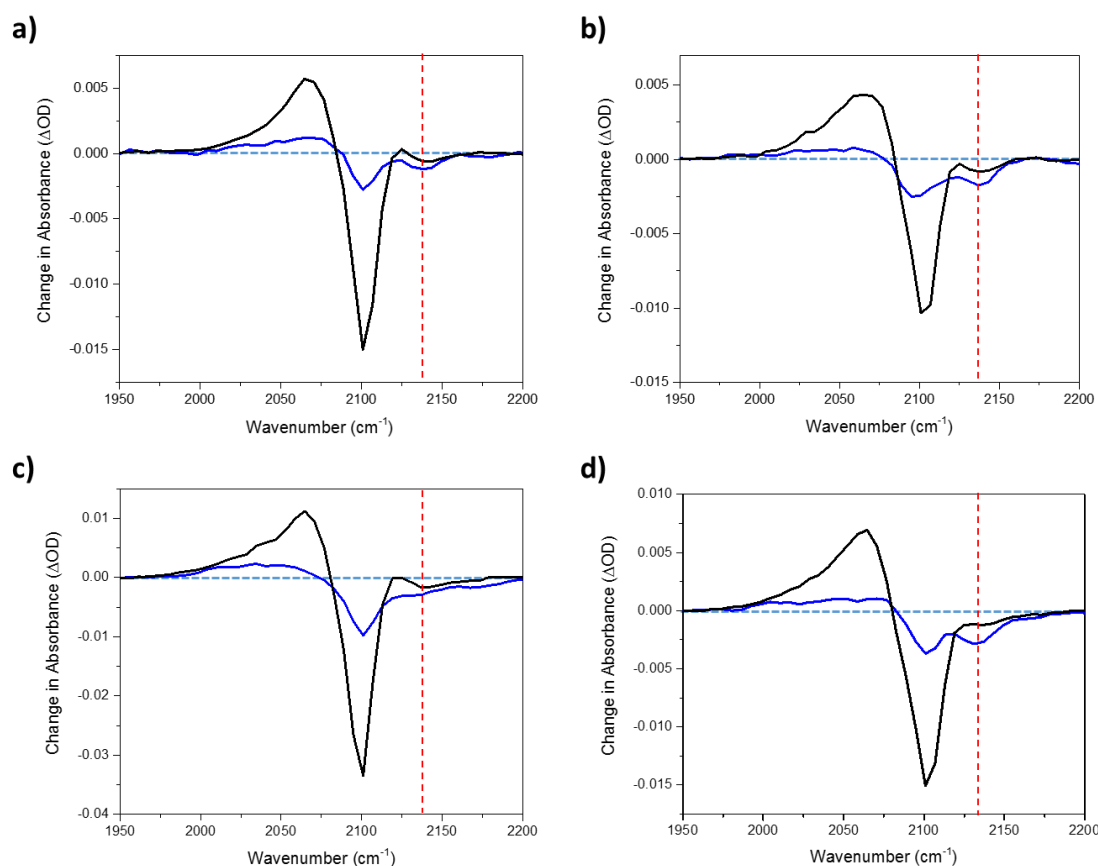
**Figure A3.7:** Lifetimes of the central peak bleach  $\tau_2$  (black), high frequency shoulder ESA (red) and low frequency shoulder  $\tau_2$  (blue) as the hydrogen bonding strength of the local environment increases.



**Figure A3.8:** 2D-IR spectrum diagonal of benzyl azide (black) in 1-Hexanol demonstrating the presence of the central peak (green) as well as the low-frequency (red) and the high-frequency (blue) shoulders. Overall fit shown in violet.

**Table A3.1:** Spectral Diffusion Dynamics of the Azide Stretch in Different Environments

Protic Environment		Aprotic Environment	
Solvent ( $\delta H^{148}$ , Viscosity <sup>175</sup> )	Diffusion Rate (fs)	Solvent ( $\mu^{160}$ , Viscosity <sup>175</sup> )	Diffusion Rate (fs)
1-Hexanol (12.5, 4.59)	748	Heptane (0, 0.386)	2261
1-Propanol (17.4, 1.95)	494	Chloroform-d (1.09, 0.563)	756
Ethanol (19.4, 1.09)	612	Acetone (2.91, 0.316)	583
Methanol (22.3, 0.553)	394	DMSO (4.1, 1.996)	513



**Figure A3.9:** Pump slices through the central azide stretch (black) and the high frequency shoulder (blue) of the 2DIR spectra of benzyl azide measured in the XXYY configuration in a) 1-Hexanol, b) Methanol, c) Heptane and d) DMSO.

### 3.7 References

(125) Kim, H.; Cho, M.; Infrared Probes for Studying the Structure and Dynamics of Biomolecules, *Chem. Rev.*, **2013**, 113, 5817-5847.

(126) Waegele, M. M.; Culik, R. M.; Gai, F.; Site-Specific Spectroscopic Reporters of the Local Electric Field, Hydration, Structure, and Dynamics of Biomolecules, *J. Phys. Chem. Lett.*, **2011**, 2, 2598-2609.

- 
- (127) Bandaria, J. N.; Dutta, S.; Hill, S. E.; Kohen, A.; Cheatum, C. M.; Fast Enzyme Dynamics at the Active Site of Formate Dehydrogenase, *J. Am. Chem. Soc.*, **2008**, *130*, 22-23.
- (128) Cho, M.; Coherent Two-Dimensional Optical Spectroscopy, *Chem. Rev.*, **2008**, *108*, 1331-1418.
- (129) Kozinski, M.; Garrett-Roe, S.; Hamm, P.; 2D-IR Spectroscopy of the Sulfhydryl Band of Cysteines in the Hydrophobic Core of Proteins, *J. Phys. Chem. B*, **2008**, *112*, 7645-7650.
- (130) Lee, K. K.; Park, K. H.; Joo, C.; Kwon, H. J.; Han, H.; Ha, J. H.; Park, S.; Cho, M.; Ultrafast internal rotational dynamics of the azido group in (4S)-azidoproline: Chemical exchange 2DIR spectroscopic investigations *Chem. Phys.*, **2012**, *396*, 23-29.
- (131) Waegele, M. M.; Tucker, M. J.; Gai, F.; 5-Cyanotryptophan as an infrared probe of local hydration status of proteins, *Chem. Phys. Lett.*, **2009**, *478*, 249-253.
- (132) Huang, C. Y.; Wang, T.; Gai, F.; Temperature dependence of the CN stretching vibration of a nitrile-derivatized phenylalanine in water, *Chem. Phys. Lett.* **2003**, *371*, 731-738.
- (133) Tucker, M. J.; Kim, K. S.; Hochstrasser, R. M.; 2D IR photon echo study of the anharmonic coupling in the OCN region of phenyl cyanate, *Chem. Phys. Lett.* **2009**, *470*, 80-84.
- (134) Stafford, A. J.; Ensign, D. L.; Webb, L. J.; Vibrational Stark Effect Spectroscopy at the Interface of Ras and Rap1A Bound to the Ras Binding Domain of RalGDS Reveals an Electrostatic Mechanism for Protein-Protein Interaction, *J. Phys. Chem. B* **2010**, *114*, 15331-15344.
- (135) Inouye, H.; Gleason, K. A.; Zhang, D.; Decatur, S. M.; Kirschner, D. A.; Differential effects of phe19 and phe20 on fibril formation by amyloidogenic peptide A $\beta$ 16-22 (Ac-KLVFFAE-NH<sub>2</sub>), *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 2306-2321.
- (136) Bischak, C. G.; Longhi, S.; Snead, D. M.; Costanzo, S.; Terrer, E.; Londergan, C. H. *Biophys. J.* **2010**, *99*, 1676.
- (137) Ohta, K.; Tayama, J.; Tominaga, K.; Ultrafast vibrational dynamics of SCN<sup>-</sup> and N<sub>3</sub><sup>-</sup> in polar solvents studied by nonlinear infrared spectroscopy, *Phys. Chem. Chem. Phys.* **2012**, *14*, 10455-10465.
- (138) Hamm, P.; Lim, M.; Hochstrasser, R. M.; Non-Markovian Dynamics of the Vibrations of Ions in Water from Femtosecond Infrared Three-Pulse Photon Echoes, *Phys. Rev. Lett.* **1998**, *81*, 5326.
- (139) Gai, X. S.; Coutifaris, B. A.; Brewer, S. H.; Fenlon, E. E.; A direct comparison of azide and nitrile vibrational probes, *Phys. Chem. Chem. Phys.* **2011**, *13*, 5926-5930.
- (140) Dutta, S.; Li, Y. L.; Rock, W.; Houtman, J. C. D.; Kohen, A.; Cheatum, C. M.; 3-Picolyl Azide Adenine Dinucleotide as a Probe of Femtosecond to Picosecond Enzyme Dynamics, *J. Phys. Chem. B*, **2012**, *116*, 542-548.
- (141) Tucker, M. J.; Gai, X. S.; Fenlon, E. E.; Brewer, S. H.; Hochstrasser, R. M.; 2D IR photon echo of azido-probes for biomolecular dynamics, *Phys. Chem. Chem. Phys.* **2011**, *13*, 2237-2241.
- (142) Oh, K. I.; Lee, J. H.; Joo, C.; Han, H.; Cho, M.;  $\beta$ -Azidoalanine as an IR Probe: Application to Amyloid A $\beta$ (16-22) Aggregation, *J. Phys. Chem. B* **2008**, *112*, 10352-10357.



---

(143) Hofmann, K. P.; Scheerer, P.; Choe, H. W.; Park, J. H.; Heck, M.; Ernst, O. P.; A G protein-coupled receptor at work: the rhodopsin model., *Trends Biochem. Sci.*, **2009**, *34*, 540-552.

(144) Farrens, D. L.; Altenbach, C.; Yang, K.; Hubbell, W. L.; Khorana, H. G. *Science*, **1996**, *274*, 497.

(145) Scheerer, P.; Park, J. H.; Hildebrand, P. W.; Kim, Y. J.; Krauss, N.; Choe, H. W.; Hofmann, K. P.; Ernst, O. P.; Crystal structure of opsin in its G-protein-interacting conformation., *Nature*, **2008**, *455*, 497-502.

(146) Ye, S.; Zaitseva, E.; Caltabiano, G.; Schertler, G. F. X.; Sakmar T. P.; Deupi X.; Vogel, R.; Tracking G-protein-coupled receptor activation using genetically encoded infrared probes, *Nature*, **2010**, *464*, 1386-1389.

(147) Dutta S.; Rock W.; Cook, R. J.; Kohen, A.; Cheatum C. M.; Two-dimensional infrared spectroscopy of azido-nicotinamide adenine dinucleotide in water., *J. Chem. Phys.*, **2011**, *135*, 055106.

(148) Hansen, C. M.; in *Hansen Solubility Parameters A User's Handbook*, CRC Press, London, **2000**, pg. 1-24.

(149) Hamm, P.; Zanni, M. in *Concepts and Methods of 2D Infrared Spectroscopy*, Cambridge University Press, Cambridge, **2011**, pg. 140-142

(150) Lipkin, J. S.; Song, R.; Fenlon, E. E.; Brewer, S. H.; Modulating Accidental Fermi Resonance: What a Difference a Neutron Makes, *J. Phys. Chem. Lett.* **2011**, *2*, 1672-1676.

(151) Nydegger, M. W.; Dutta, S.; Cheatum C. M.; *2D IR Study of 3-Azidopyridine as a Potential Spectroscopic Reporter of Protonation State*, *J. Chem. Phys.*, **2010**, *133*, 134506-134508.

(152) Hohenberg, P.; Kohn, W.; Inhomogeneous Electron Gas, *Phys. Rev.*, **1964**, *136*, B864.

(153) Jensen, F.; *Introduction to Computational Chemistry (Second Edition)*, John Wiley & Sons Ltd., Chichester, **2007**, pg. 232-267.

(154) Gaussian 09, Revision D.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2013.

$$(155) \quad P(E_i, T) = \frac{e^{(-E_i/RT)}}{\sum_i (e^{(-E_i/RT)})}$$

Where  $P(E_i, T)$  is the relative abundance of the state of energy  $E_i$  at temperature  $T$ .  $R$  is the universal gas constant.

(156) Computational Chemistry Comparison and Benchmark DataBase (CCCBDB), Release 18, October 2016, NIST Standard Reference Database 101 (III.B.3.a.(XIII.C.1.)), <http://cccbdb.nist.gov/vibscalejust.asp>, 24<sup>th</sup> September 2017 at 15:30

---

(157) Hamm, P.; Zanni, M. in *Concepts and Methods of 2D Infrared Spectroscopy*, Cambridge University Press, Cambridge, **2011**, pg. 140-142

(158) Lipkin, J. S.; Song, R.; Fenlon, E. E.; Brewer, S. H.; Modulating Accidental Fermi Resonance: What a Difference a Neutron Makes, *J. Phys. Chem. Lett.* **2011**, *2*, 1672-1676.

(159) Nydegger, M. W.; Dutta, S.; Cheatum C. M.; *2D IR Study of 3-Azidopyridine as a Potential Spectroscopic Reporter of Protonation State*, *J. Chem. Phys.*, **2010**, *133*, 134506-134508.

(160) Lide, D. R.; *CRC Handbook of Chemistry and Physics: A Ready Reference Book of Chemical and Physical Data (79<sup>th</sup> Edition)*, **1999**, CRC Press, London, Sections 9-42 to 9-50.

$$(161) \quad S_{\text{Isotropic}} = (S_{\text{ZZZZ}} + 2S_{\text{ZZYY}})/3$$

Where  $S_{\text{Isotropic}}$  is the isotropic pump-probe signal,  $S_{\text{ZZZZ}}$  is the pump-probe signal measured in the ZZZZ polarisation geometry and  $S_{\text{ZZYY}}$  is the pump-probe signal measured in the ZZYY polarisation geometry.

(162) Hamm, P.; Zanni, M.; in *Concepts and Methods of 2D Infrared Spectroscopy*, Cambridge University Press, Cambridge, **2011**, pg. 88-108

(163) van Wilderen, L. J. G. W.; Kern-Michler, D.; Müller-Werkmeister, H. M.; Bredenbeck, J.; Vibrational dynamics and solvatochromism of the label SCN in various solvents and hemoglobin by time dependent IR and 2D-IR spectroscopy, *Phys. Chem. Chem. Phys.*, **2014**, *16*, 19643

(164) Ohta, K.; Tominaga, K.; Vibrational population relaxation of hydrogen-bonded phenol complexes in solution: Investigation by ultrafast infrared pump-probe spectroscopy, *Chemical Physics*, **2007**, *341*, 310-319.

(165) Hamm, P.; Zanni, M.; in *Concepts and Methods of 2D Infrared Spectroscopy*, Cambridge University Press, Cambridge, **2011**, pg. 13

(166) Hamm, P.; Zanni, M.; in *Concepts and Methods of 2D Infrared Spectroscopy*, Cambridge University Press, Cambridge, **2011**, pg. 88-108.

(167) Kwak, K.; Park, S.; Finkelstein, I. J.; Fayer, M. D.; Frequency-frequency correlation functions and apodization in two-dimensional infrared vibrational echo spectroscopy: a new approach., *J. Chem. Phys.*, **2007**, *127*, 124503.

(168) Fenn, E. E.; Fayer M. D.; Extracting 2D IR frequency-frequency correlation function from two component systems., *J. Chem. Phys.*, **2011**, *135*, 074502.

(169) Hamm, P.; Zanni, M.; in *Concepts and Methods of 2D Infrared Spectroscopy*, Cambridge University Press, Cambridge, **2011**, pg. 109-126.

(170) Rubtsov, I. V.; Hochstrasser, R. M.; Vibrational Dynamics, Mode Coupling, and Structural Constraints for Acetylproline-NH<sub>2</sub>, *J. Phys. Chem. B*, **2002**, *106*, 9165-9171.

$$(171) \quad \beta_{ij} = \sqrt{\left(\Delta_{ij} \left( (\omega_i - \omega_j + \Delta_{ii})(\omega_i - \omega_j - \Delta_{ij}) \right) / 2(\Delta_{ii} + \Delta_{jj}) \right)}$$

Where  $\beta_{ij}$  is the coupling constant between two vibrational modes,  $\omega_i$  is the position of the first vibrational mode,  $\omega_j$  is the position of the second vibrational mode,  $\Delta_{ii}$  is the anharmonicity of the first vibrational mode,  $\Delta_{jj}$  is the anharmonicity of the second vibrational mode and  $\Delta_{ij}$  is the anharmonicity of the cross peak between the two modes.

---

(172) Synthetic work carried out by Dr. J. J. May, Burley Group.

(173) 
$$S_{Iso}^{PP}(T_w, \omega) = S_{XXXX}^{PP}(T_w, \omega) + 2S_{XXYY}^{PP}(T_w, \omega)$$

Where  $T_w$  is the pump-probe waiting time,  $\omega$  is the probe frequency,  $S_{XXXX}^{PP}(T_w, \omega)$  is the XXXX pump-probe spectrum,  $S_{XXYY}^{PP}(T_w, \omega)$  is the XXYY pump-probe spectrum and  $S_{Iso}^{PP}(T_w, \omega)$  is the isotropic pump-probe spectrum.

(174) Deflores, L. P.; Nicodemus, R. A.; Tokmakoff, A.; Two Dimensional Fourier Transform Spectroscopy in the Pump-probe Geometry., *Opt. Lett.*, **2007**, 32, 2966–2968.

(175) Lide, D. R.; *CRC Handbook of Chemistry and Physics: A Ready Reference Book of Chemical and Physical Data (79<sup>th</sup> Edition)*, **1999**, CRC Press, London, Sections 6-170 to 6-174.

IR Spectroscopy of mixed DNA duplexes  
containing A-tract and Alternating A-T  
Sequences

Chapter 4

## 4.1 Abstract

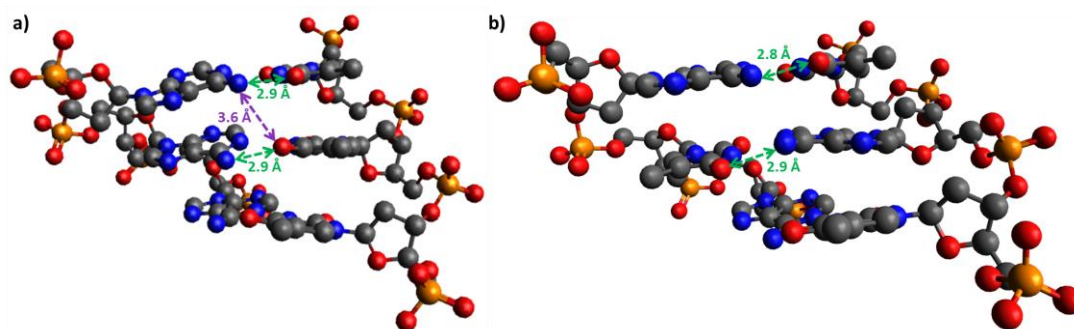
*Alterations in the order of an AT-rich sequence, from an A-tract to an alternating AT motif, within a DNA oligonucleotide leads to structural perturbations and changes in the physical properties of the double helix. Changing the AT-rich region of the oligonucleotide from an A-tract to an alternating AT sequence is also thought to lead to a loss of inter-strand bifurcated H-bonds within the DNA duplex along with changes in the population and order of the minor groove spine of hydration. In this chapter, the structural changes due to such an alteration of an AT-rich region of the DNA duplex are investigated and alongside this the effect of such a change in the sequence on the thermal denaturation of the duplex is explored. Results from 2D-IR spectroscopy, which is sensitive to H-bonding and structural perturbations, alongside results from UV-visible and FT-IR spectroscopy shows that replacing an A-tract with an alternating AT sequence results in a loss of a region of highly ordered AT propeller twists as well as changes in the order of the minor groove spine of hydration. Furthermore these results indicate that the thermal denaturation of the studied DNA sequences follow an unzipping mechanism, where the differences in the observed transition temperatures and intermediate states are a direct consequence of the change in the AT sequence. Finally, the utility of 2D-IR spectroscopy to study DNA structural changes is explored and its capabilities to simplify the quantification of solution phase DNA oligonucleotides is demonstrated.*

## 4.2 Introduction

Deoxyribonucleic acid (DNA) is the fundamental repository of genetic information for all organisms found in nature.<sup>176</sup> This information is encoded by differing combinations of four different nucleic acids, namely adenine (A), thymine (T), cytosine (C) and guanine (G), which are held in an anti-parallel double helix. DNA can be structurally altered by changing the environmental conditions,<sup>177,178,179,180</sup> which causes a change in the conformer of the anti-parallel double helix, leading to large structural perturbations. However, it has been noted that the exact structure of DNA oligonucleotides depends on the sequence of the DNA bases and that changes in the sequence leads to smaller structural perturbations.<sup>181,182</sup> This was first observed by the differing mobility's of DNA fragments in gel electrophoresis experiments.<sup>183</sup> These experiments lead to the understanding that the presence of repeating in-phase A·T base pairs lead to the bending of the helical axis of the DNA strand towards the minor groove of the duplex, leading to a low gel mobility.<sup>184,185,186,187</sup> Such DNA sequences are more commonly referred to as A-tracts and are separated into two broad subsets.<sup>188</sup> These subsets are known as the symmetric A-tracts, defined as sequences of the form 5'-A<sub>n</sub>T<sub>n</sub> where  $n \geq 2$ , and asymmetric A-tracts, defined as sequences of the form 5'-A<sub>n</sub> where  $n \geq 4$ . In addition to the interesting structural perturbations that these sequences impart on the DNA duplex it has been found that these A-tracts play an important role in the regulation of vital biological processes involving DNA,<sup>189,190,191</sup> and represent the preferred binding sites of a subset of minor groove binding molecules.<sup>192,193,194</sup>

Due to the prevalence and importance of these A-tract sequences in nature, a multitude of NMR,<sup>195,196,197,198</sup> crystallography,<sup>199,200,201</sup> computational<sup>202,203,204</sup> and theoretical studies<sup>205,206,207,208</sup> have been carried out in order to gain a deeper understanding of the impact of the A-tract sequence on all of the structural aspects of the DNA duplex and develop a set of guiding principles focused on revealing the relation between structure and sequence. Such studies have reveal that the base roll, tilt and propeller twist are all altered in the A-tract sequences when compared to both more generic and alternating A-T sequences (*e.g.* sequences of the form 5'-(AT)<sub>n</sub>,  $n \geq 2$ ).<sup>52,56,209</sup> Of particular interest was the observation of the strong effect of the sequence on the propeller twist, as it was postulated that the size and apparent order of the propeller twists of the A·T base pairs within all A-tracts facilitates the formation of inter-strand bifurcated hydrogen bonds mediated through the major groove.<sup>210</sup>

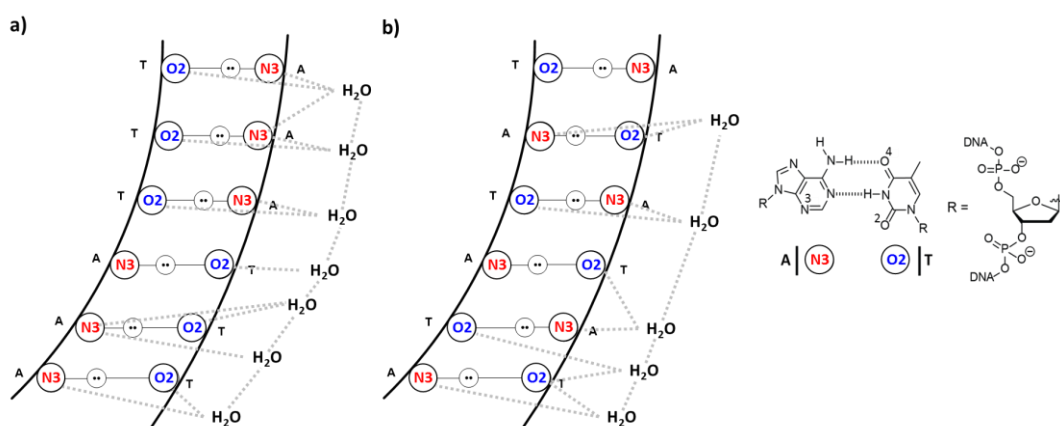
A schematic representation of such a three-centre hydrogen bond within a symmetric A-tract sequence is shown in figure 4.1.



**Figure 4.1:** Schematic representation of a) a three base pair long stretch of an A-tract duplex (sequence: 5'-CGCAAATTTGCG-3', PDB ID: 1S2R<sup>211</sup>) and b) a three base pair long stretch of an alternating duplex (sequence: 5'-CGCATATATGCG-3', PDB ID: 1DN9<sup>181</sup>).

The proposed bifurcated hydrogen bonds formed in A-tract sequences are thought to occur between the A amine moiety and the T<sub>4</sub> carbonyl of the next A-T base pair in the sequence mediated through the major groove of the DNA duplex, as indicated by a purple arrow in Fig. 4.1.(a). In contrast to this in the alternating A-T sequence the A amine moiety and the T<sub>4</sub> carbonyls of the next A-T base pairs in the sequence are on the same strand (Fig. 4.1.(b)) meaning that it is not possible to form the same inter-strand hydrogen bonding. These additional interactions between the two strands in the duplex are thought to increase the stability and rigidity of the A-tract sequence. Additionally it has been proposed that the presence of these interactions also enhance the geometries between the bases in A-tract DNA leading to stronger base stacking interactions. A multitude of studies utilising X-ray crystallography,<sup>47,212,213</sup> Raman<sup>214,215</sup> and CD<sup>216</sup> spectroscopy have been used to investigate both symmetric and asymmetric A-tract sequences and have determined that such bifurcated hydrogen bonds are indeed present in many A-tract sequences.

In addition to the differences in the structures of the A-tract and alternating A-T sequences another important aspect of AT-rich regions of the DNA duplexes is the minor groove spine of hydration.<sup>217,218</sup> Due to the narrow widths associated with the minor grooves of AT-rich regions these spines of water molecules are highly ordered and found to interact both with the A(N<sub>3</sub>) position and the T<sub>2</sub> carbonyl. Such spines of hydration have been observed using both NMR<sup>219,220</sup> and X-ray crystallography<sup>181,209</sup> and schematic representations of these spines in both the an A-tract (5'-A<sub>3</sub>T<sub>3</sub>) and alternating A-T (5'-(AT)<sub>3</sub>) sequence are reproduced in figure 4.2.



**Figure 4.2:** Schematic representation of a) spine of hydration in the minor groove of an A-tract duplex (sequence: 5'-CGCAAATTTGCG-3', PDB ID: 1D65<sup>209</sup>) and b) spine of hydration in the minor groove of an alternating duplex (sequence: 5'-CGCATATATGCG-3', PDB ID: 1DN9<sup>181</sup>). The dotted grey lines are included to give an idea of the order within the spine of hydration for each duplex.

From the schematic representations it can be seen that the spine of hydration imparts enthalpic and entropic contributions to the stabilities of the both of the sequences. The fact that such spines of hydration have been observed in both A-tract and alternating A-T sequences has led to the suggestion that this means that all AT-rich sequences are equally stabilised by the presence of these water molecules.<sup>221,222</sup>

Despite the exploration of the changes in the structural aspects within the DNA duplex due to the presence of an A-tract, the impact of these sequences on the thermal denaturation mechanism of short lengths of DNA remains relatively unclear. The mechanism by which the strands dissociate from the duplex are important in a range of biological processes as well as being important for the application of DNA scaffolds for nano-devices and self-assembly. A range of studies including both computational<sup>223,224</sup> and experimental techniques<sup>225,226,227,228</sup> have been carried out on the denaturation and reassembly of short DNA oligonucleotides and DNA hairpins. These studies highlight the important of the exact DNA sequence as well as the length of the DNA oligonucleotide on the denaturation mechanism. However these studies do not consider the changes which occur to the denaturation process when an A-tract is replaced with its alternating A-T sequence counterpart.

Further study is thus required to develop a clear picture of the important aspects of the structural differences and changes in the spines of hydration between an A-tract sequence, of which  $d(\text{GGAAATTTGC})_2$ ,  $(\text{A}_3\text{T}_3)$  is our chosen exemplar and an alternating A-T sequence,  $d(\text{GGATATATGC})_2$ ,  $((\text{AT})_3)$ . In addition to this, further study is required to gain a clear picture



of the impact this alteration has on the duplex dissociation mechanism of short DNA oligonucleotides.

It is believed that two dimensional infrared spectroscopy (2D-IR)<sup>229,230,231</sup> has sufficient sensitivity to differentiate the structural details that are central to the differences between DNA oligonucleotides containing an A-tract and alternating A–T sequences. 2D-IR spectroscopy offers the advantage of being able to probe the structure, intermolecular interactions and hydration by directly measuring the lineshapes, coupling and dynamics of vibrational modes.<sup>229,230,231</sup> The application of 2D-IR has recently extended our understanding of the nature of vibrational modes of DNA bases,<sup>232</sup> Watson-Crick (W-C) base pairing<sup>233,234</sup> and base stacking<sup>235</sup> as well as revealing energy transfer mechanisms between bases and backbone<sup>236,237</sup> and the interaction of water with DNA.<sup>238</sup> Most recently, temperature-jump 2D-IR reported sequence-dependent pre-melting dynamics.<sup>239</sup>

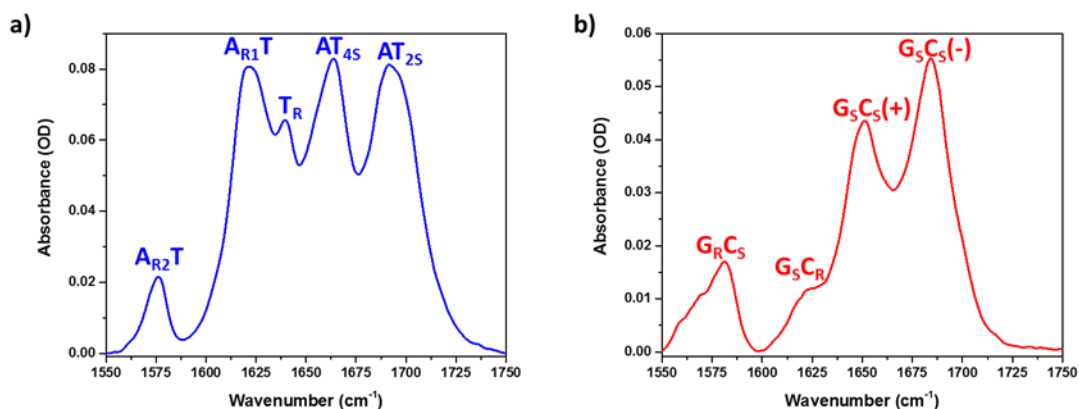
Here, the distinct difference between the structure of the symmetric A<sub>3</sub>T<sub>3</sub> A-tract and the (AT)<sub>3</sub> alternating A–T sequence is shown to be due to the loss of an ordered propeller twist arrangement of the base pairs present in the A-tract sequence. This results from the presence of three-centred hydrogen bonds within the AT-rich section of the DNA duplex which is only possible within the A-tract sequence. The presence of the unique structure of the symmetric A-tract sequence also governs the formation of a more highly ordered and populated minor groove spine of hydration than seen in the alternating A–T sequence. Overall, it is shown that these two effects contribute equally to the enhancement of the stability and so the rigidity of the A-tract sequence when compared to the alternating A–T sequence. In addition to this further analysis of the FT-IR and 2D-IR spectroscopy of these sequences, as the temperature was increased, suggested that both duplexes melted via an unzipping mechanism. In addition to these changes the role of the decreased order of the spine of hydration associated with the alternating A–T sequence also contributes to the lower overall melting temperature of this sequence.

## 4.3 Results and Discussion

### 4.3.1 Difference Spectroscopy

The FTIR spectra of the A<sub>3</sub>T<sub>3</sub> and (AT)<sub>3</sub> DNA sequences in the 1550 – 1700 cm<sup>-1</sup> region are shown in Fig.4.4(a). Five peaks are observed in the spectrum of the A<sub>3</sub>T<sub>3</sub> sequence (Fig.4.4, black) centred at 1576 cm<sup>-1</sup>, 1622 cm<sup>-1</sup>, 1646 cm<sup>-1</sup>, 1666 cm<sup>-1</sup> and 1692 cm<sup>-1</sup>. Similarly five peaks are observed in the spectrum of the (AT)<sub>3</sub> sequence (Fig.4.4, red) centred at 1577 cm<sup>-1</sup>

<sup>1</sup>, 1622 cm<sup>-1</sup>, 1650 cm<sup>-1</sup>, 1664 cm<sup>-1</sup> and 1689 cm<sup>-1</sup>. This spectral region features partially overlapping contributions from both GC and AT base pairs, but the peaks in the A<sub>3</sub>T<sub>3</sub> and (AT)<sub>3</sub> DNA sequences in Fig.4.4 can be assigned through reference to spectra of DNA duplexes containing exclusively GC or AT base pairs (Fig.4.3) and previous 2D-IR and computational studies.<sup>232,235,236,240</sup>



**Figure 4.3:** Assigned FT-IR spectra of a) AT only duplex (sequence: 5'-ATTATTATTATATTA-3') and b) GC only duplex (sequence: 5'-GCCGCCGCCG-3'). All data recorded at 20 °C.

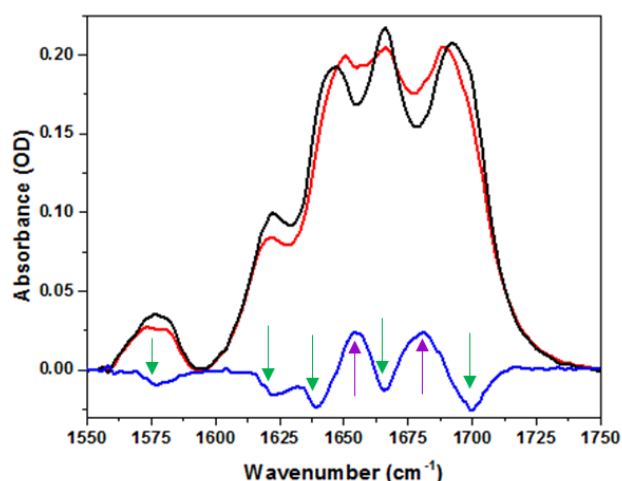
It is clear from the close agreement between peak positions in the AT spectrum and both the A<sub>3</sub>T<sub>3</sub> and (AT)<sub>3</sub> sequences that the *ds*-DNA spectra are dominated by the AT modes. By contrast the GC peaks occur at different frequencies and are generally weaker than the AT modes. The assignments are summarized in Table 4.1 and agree well with computational predictions of deuterated B-DNA spectra.<sup>240</sup>

**Table 4.1:** Assignment of peaks in the IR spectra of A<sub>3</sub>T<sub>3</sub> and (AT)<sub>3</sub> DNA duplexes, showing comparison with AT and GC-only sequences.

Assignment	Position (cm <sup>-1</sup> )				Description
	A <sub>3</sub> T <sub>3</sub>	(AT) <sub>3</sub>	AT <sup>a</sup>	GC <sup>b</sup>	
AT <sub>2S</sub>	1692	1689	1692		Base-paired T <sub>2</sub> C=O stretch
G <sub>S</sub> C <sub>S</sub> (-)				1684	GC C=O antisymmetric stretch
AT <sub>4S</sub>	1666	1666	1664		Base-paired T <sub>4</sub> C=O stretch
G <sub>S</sub> C <sub>S</sub> (+)				1651	GC C=O symmetric stretch
T <sub>R</sub>	1646	1650	1640		T ring vibration
A <sub>R1</sub> T			1622 (s)		Coupled AT ring I vibration/ C ring mode + G C=O
G <sub>S</sub> C <sub>R</sub>	1622	1622		1622 (w)	
G <sub>R</sub> C <sub>S</sub>				1582 (w)	G ring mode + C C=O/ Coupled AT ring II vibration
A <sub>R2</sub> T	1576	1577	1576 (s)		

<sup>a</sup> obtained from the sequence: 5'-ATTATTATTATATTA-3' (Fig 4.3.(a)); <sup>b</sup> obtained from the sequence: 5'-GCCGCCGCCG-3' (Fig 4.3.(b)); modes marked (s) are those which contribute strongly to the overall spectrum and modes marked (w) are those which contribute weakly to the overall spectrum.

The main difference between the FTIR spectra for the A<sub>3</sub>T<sub>3</sub> and (AT)<sub>3</sub> sequences are relatively small with the alteration of the AT part of the sequence causing a 3 cm<sup>-1</sup> shift to lower frequency of the AT<sub>2S</sub> mode and a 4 cm<sup>-1</sup> shift to higher frequency in the T<sub>R</sub> mode. Using linear regression to overlap the FTIR spectra of these two sequences allows the difference spectrum to be obtained and the differences to be explored (Fig. 4.4).

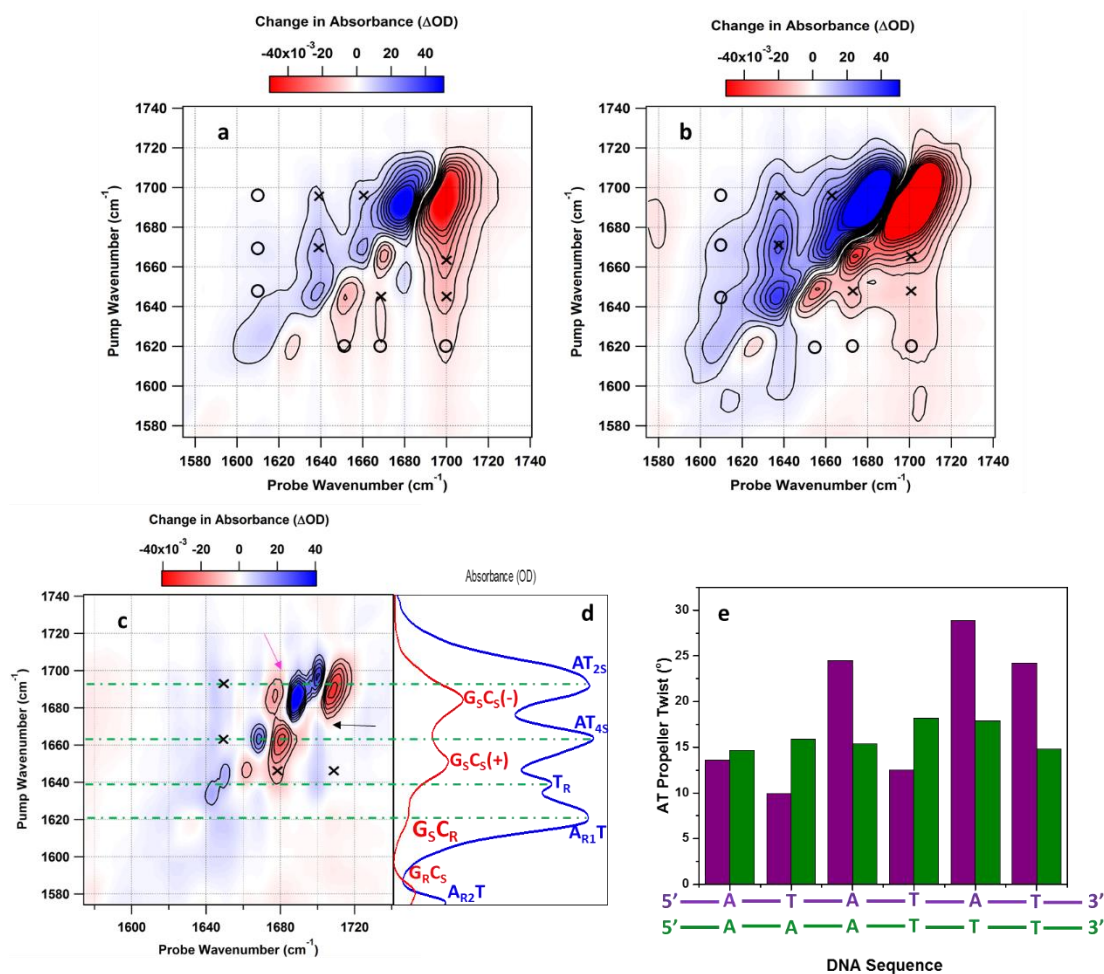


**Figure 4.4:** FT-IR spectrum of the  $A_3T_3$  sequence (black), the  $(AT)_3$  sequence (red) and the difference spectrum between these two sequences (blue). All data recorded at 20 °C.

The FT-IR difference spectra between the  $(AT)_3$  and  $A_3T_3$  duplexes (blue, Fig. 4.4) shows a series of negative peaks (green arrows) and two positive peaks (purple arrows). The positions of the negative peaks are found to be at 1575  $\text{cm}^{-1}$ , 1622  $\text{cm}^{-1}$ , 1640  $\text{cm}^{-1}$  and 1700  $\text{cm}^{-1}$  and the position of the positive peaks are determined to be 1650  $\text{cm}^{-1}$  and 1680  $\text{cm}^{-1}$ . Again contrasting these positions with the peak positions in the FTIR spectra of exclusively AT and GC sequences (Fig.4.3) shows the positions of the negative peaks in the difference spectrum are consistent with AT base modes, whereas the positions of the positive peaks are consistent with GC base modes. These distinct changes in the spectrum due to small alterations in the DNA sequence demonstrates the information rich nature of the FTIR spectroscopy of DNA. However due to the larger degree of congestion in this spectral window it is difficult to fully understand these changes and definitively assign them to particular base modes. This spectral congestion can be unraveled by 2D-IR methods, which report vibrational coupling patterns in the off-diagonal region of the spectrum.

The 2D-IR spectrum of the  $A_3T_3$  and  $(AT)_3$  sequences are shown in Figs 4.4.(a) and 4.4.(b). Each peak observed in the FTIR spectrum of each of these sequences gives rise to a negative feature located on the 2D-IR spectrum diagonal. These are assigned to the respective  $\nu = 0 \rightarrow 1$  transitions, each with an accompanying, positive,  $\nu = 1 \rightarrow 2$  peak shifted by  $\sim 10 \text{ cm}^{-1}$  to lower probe frequency.<sup>236,239,240,241,242,243</sup> Peaks located in the off-diagonal region primarily indicate the presence of coupling between diagonal vibrational modes, though a small contribution is expected from energy transfer due to the fast ( $\sim 650$  fs) vibrational relaxation

of the base vibrational modes.<sup>236</sup> The 2D-IR spectra of the  $A_3T_3$  and  $(AT)_3$  duplexes contain two groups of off-diagonal peaks, these are indicated in Figs 4.5.(a) and 4.5.(b). Those due to intra-base coupling between the  $AT_{2S}$ ,  $AT_{4S}$  and  $T_R$  modes that are primarily located on the thymine base are indicated by crosses. Other off-diagonal peaks (Figs 4.5.(a) and 4.5.(b), open circles) indicate inter-base coupling between the adenine-based  $A_{R1}T$  mode and the three thymine modes induced by Watson-Crick H-bonding.

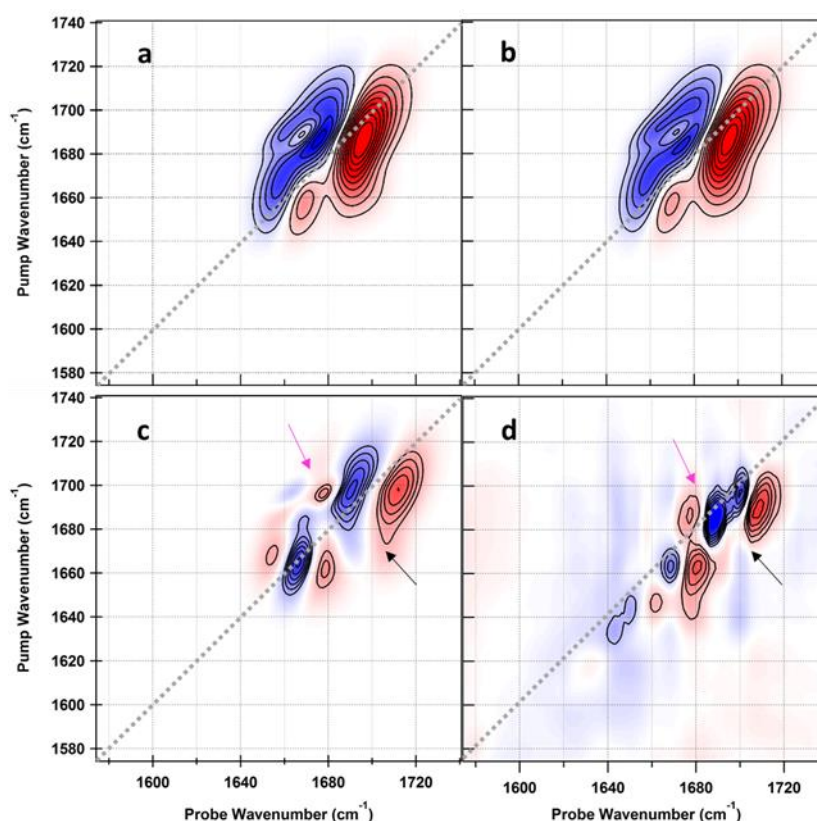


**Figure 4.5:** a) 2D-IR spectrum of  $A_3T_3$  duplex, b) 2D-IR spectrum of  $(AT)_3$  duplex, c) 2D-IR difference spectrum due to change in AT region sequence, d) assigned AT only (blue) and GC only (red) duplex FTIR spectrum and e) AT propeller twists of  $A_3T_3$  (green, PDB ID: 1S2R<sup>211</sup>) and  $(AT)_3$  (purple, PDB ID: 1DN9<sup>181</sup>).

A difference 2D-IR spectrum shows the impact of the change in the AT region between the  $A_3T_3$  and  $(AT)_3$  sequences (Fig.4.5.(c)). A comparison of the peak positions in the 2D-IR difference spectrum with the linear AT and GC spectra (Fig.4.5.(d)) suggests that the change in the AT sequence primarily affects the  $AT_{2S}$  and  $AT_{4S}$  vibrational modes. Other smaller

features appear in the 2D-IR difference spectrum that originate from changes in the  $T_R$  mode and related off-diagonal peaks (crosses Fig.4.5.(c)).

The overlapping peaks in the IR spectrum of the DNA bases mean that the 2D-IR off-diagonal peaks are essential for clear differentiation between changes affecting the AT and GC base pairs. The peaks in the difference FT-IR and 2D-IR spectra focus on the  $AT_{25}$  and  $AT_{45}$  modes, but to rule out contributions from GC modes and to quantify the changes, we employed a model 2D-IR spectrum (Fig.4.6) constructed from 2D Gaussian lineshapes, which simulate the coupled peaks in the 2D-IR spectra of the DNA duplexes (Figs 4.6.(a) and 4.6.(b)). A variety of spectral changes were modelled and the difference



**Figure 4.6:** Simulated 2D-IR spectra of the coupled  $AT_{25}$  and  $AT_{45}$  modes before (a) and after (b) a shift of a subset of the bands to higher frequency. (c) shows the simulated difference 2D-IR spectrum that results from a subtraction of (a) from (b). (d) Shows the difference 2D-IR spectrum obtained following change in A–T sequence between  $A_3T_3$  and  $(AT)_3$ . Pink and black arrows point to off-diagonal features linking the two modes in experimental and simulated spectra.

spectra simulated (Appendix (Fig.A4.1)). Even though simulated spectra including shifts in both the AT and GC modes were considered, the best agreement with the experimental 2D-IR difference spectrum (Fig.4.6.(d)) was obtained by shifting a subset of the  $AT_{25}$  and  $AT_{45}$

peaks to higher frequency (Fig.4.6.(c)). Crucially, this simulation recreated the off-diagonal structure of the difference spectrum as well as the diagonal peaks. Additionally, it is noted that this is consistent with the data from the X-ray crystallography which suggested that the structure of the GC-ends of the duplex were similar in both duplexes.<sup>181,211</sup> The results are summarized in Table 4.2, which show that when the AT sequence is altered from the A-tract form (A<sub>3</sub>T<sub>3</sub>) to the alternating arrangement ((AT)<sub>3</sub>) ~25% of the AT<sub>25</sub> mode is shifted ~6 cm<sup>-1</sup> to higher frequency while ~27% of the AT<sub>45</sub> mode is shifted by ~11 cm<sup>-1</sup>.

**Table 4.2:** Changes to the IR response of due to AT sequence alteration between A<sub>3</sub>T<sub>3</sub> and (AT)<sub>3</sub>.

	<b>Changes due to Ordered to Disordered Phase Transition</b>
<b>AT<sub>25</sub> Shift (cm<sup>-1</sup>)</b>	6.3
<b>AT<sub>25</sub> Shifted Subset Size (%)</b>	<b>25.4</b>
<b>Change in T<sub>2</sub> H-bond Strength (kJmol<sup>-1</sup>)<sup>a</sup></b>	-2.8
<b>AT<sub>45</sub> Shift (cm<sup>-1</sup>)</b>	10.7
<b>AT<sub>45</sub> Shifted Subset Size (%)</b>	<b>26.7</b>
<b>Change in T<sub>4</sub> H-bond Strength (kJmol<sup>-1</sup>)<sup>a</sup></b>	-4.8

<sup>a</sup>Values calculated using the C=O bond energy (~ 743 kJmol<sup>-1</sup>)<sup>215</sup> via the methodology as outlined in ref. 244 using equation ref. 245.

Based on these models, we assign the peaks appearing in the FT-IR and 2D-IR difference spectra to a shift to higher frequency of a portion of the AT<sub>25</sub> and AT<sub>45</sub> base vibrations due to the alterations in the sequence in the AT region of the duplex. No contributions from the diagonal or off-diagonal peaks attributable to the GC region of the duplex were observed, showing that the GC regions of the DNA duplex are not affected by the alteration in the AT region between the two sequences.

Shifts of the AT<sub>25</sub> and AT<sub>45</sub> modes to higher frequency are consistent with a decrease in the strength of the H-bonds formed to the T<sub>2</sub> and T<sub>4</sub> carbonyl groups upon AT sequence alteration of 2.8 and 4.8 kJmol<sup>-1</sup>, respectively.<sup>215,244,245</sup> It is important to note that, although the two modes exhibit coupling, it has been demonstrated that this is weak and that a local mode picture, treating the two carbonyls as separate modes is appropriate.<sup>232</sup> The smaller change

in the  $T_R$  vibrational mode is likely to be a result of changes to the two T carbonyls since their stretches are known to contribute to this mode.<sup>240</sup>

The  $T_2$  carbonyl points directly into the minor groove of the DNA duplex and therefore it is known that it is not directly involved in the Watson-Crick base pairing hydrogen bonds. Crystallographic<sup>181,209</sup> and NMR<sup>219,220</sup> studies of DNA duplexes demonstrate that a spine of hydration is contained within the minor groove of all DNA duplexes containing AT base pairs. The exact structure of the spine of hydration is dependent on the DNA sequence<sup>181,209</sup> however these water molecules are known to form hydrogen bonds to the  $T_2$  carbonyls. The shift to higher frequency of a portion of the  $T_2$  carbonyls suggests that the alteration of the AT region of the DNA duplex causes a change in the spine of hydration. Specifically the shift to higher frequency of these  $T_2$  carbonyls is consistent with a decrease in the strength of the hydrogen bonds formed to these carbonyl groups in the alternating  $((AT)_3)$  compared to the A-tract  $(A_3T_3)$  sequence. This decrease in the strength of the hydrogen bonding to the  $T_2$  carbonyls suggest that the spine of hydration in the A-tract sequence  $(A_3T_3)$  is more ordered, allowing stronger H-bonding interactions to be formed, than the spine within the alternating sequence  $((AT)_3)$ . This is consistent with crystallographic studies carried out on sequences, which contain these two differing AT regions. These studies show that the spine of hydration in the A-tract sequence contains more water molecules and the structure of the spine has a higher degree of order than its counterpart in the alternating sequence (Fig.4.2).<sup>181,209</sup>

The  $T_4$  carbonyl is directly involved in the Watson-Crick hydrogen bonds that form between the AT base pairs. The difference 2D-IR spectrum between the A-tract and alternating sequences show that the change in the AT region between these two sequences causes a portion of the  $T_4$  carbonyls to undergo a shift to higher frequency in the alternating sequence. This shift of the  $T_4$  carbonyls shows that the Watson-Crick hydrogen bonds to these groups are weaker in the  $(AT)_3$  sequence. This suggests that there are structural differences between the  $(AT)_3$  and  $A_3T_3$  duplex. The two major structural changes that can cause the Watson-Crick hydrogen bonds to weaken are a base buckle and propeller twist.<sup>241</sup> An increase in the base buckle angle will reduce the strength of the W-C H-bonds, but a computational study predicted that the  $T_4$  carbonyl is not strongly affected by this until an extreme angle of buckle ( $\sim 50^\circ$ ) is reached.<sup>241</sup> This is not consistent with the crystallographic studies that have been carried out on DNA duplexes containing  $(AT)_3$  and  $A_3T_3$  sequences.<sup>181,209</sup> An increase in the propeller twist will cause an increase in the length of the Watson-Crick H-bond involving the

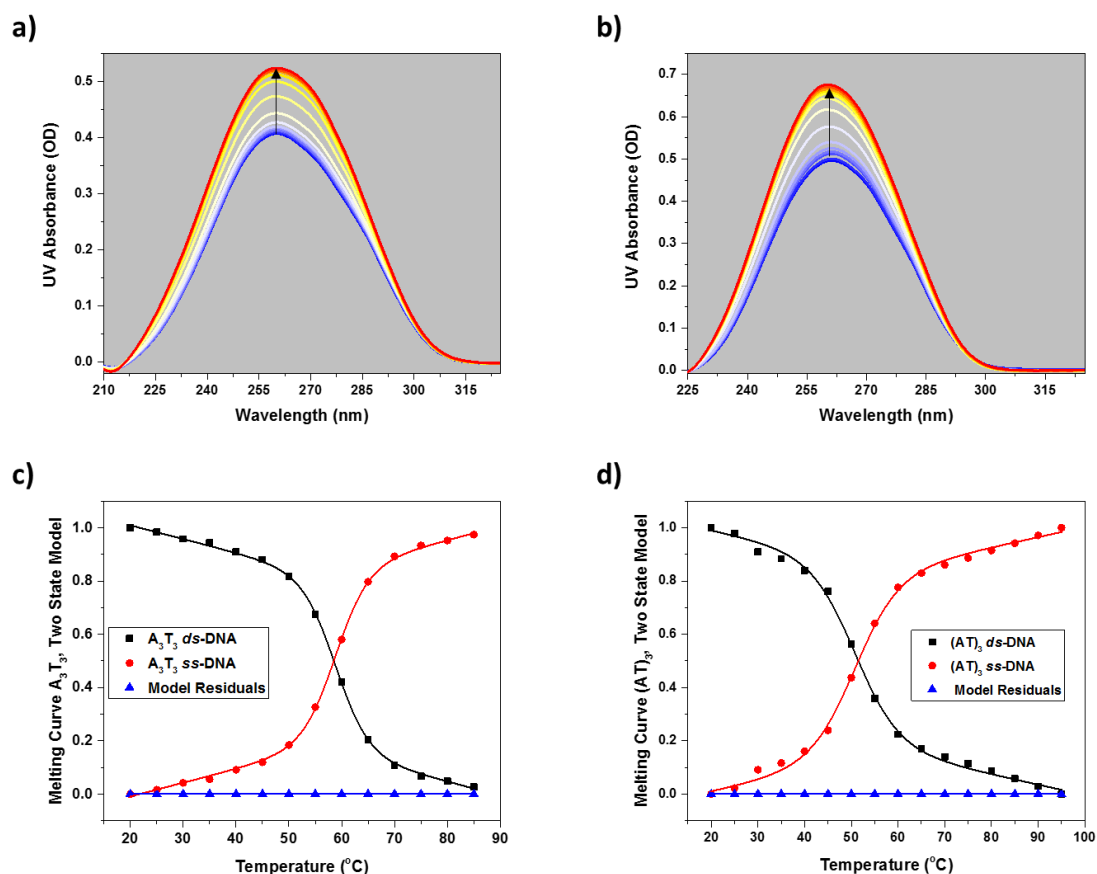


T<sub>4</sub> carbonyl, decreasing its strength. The increase of the propeller twist of the AT base pairs would not cause a major change in the structure of the DNA duplex and so this change is consistent with both the presented spectroscopic data and with X-ray crystallography<sup>181,209</sup> studies which have reported changes in the AT base pair propeller twists. The AT base pair propeller twists of the duplex containing the A<sub>3</sub>T<sub>3</sub> sequence<sup>209</sup> (green, Fig.4.4(e)) and of the duplex containing the (AT)<sub>3</sub> sequence<sup>181</sup> (purple, Fig.4.4(e)) are shown in a bar-graph (Fig.4.4(e)). This bar-graph shows that in the A<sub>3</sub>T<sub>3</sub> sequence the propeller twists are similar and highly ordered, whereas in the (AT)<sub>3</sub> sequence the propeller twists differ along the sequence and there is a large degree of variability in the size of the twist. In particular there are three AT base pairs in the (AT)<sub>3</sub> sequence have a significantly larger propeller twist when compared to the A<sub>3</sub>T<sub>3</sub> sequence. This is consistent with our observations of a subset of T<sub>4</sub> related modes shifting to higher frequency upon sequence alteration. The differences in the AT propeller twists does not affect the GC base pairs, which is also entirely consistent with our spectroscopic data.

#### 4.3.2 Thermal DNA Denaturation

##### *Duplex*

The melting transitions of short DNA oligonucleotides are often thought to occur via a simple two state model.<sup>226,246</sup> In this model it is assumed that at each temperature the sample contains a mixture of *ds*-DNA and *ss*-DNA. Applying this model to a spectroscopic study assumes that the spectrum of the sample at each temperature can be represented by a linear combination of the *ds*-DNA and *ss*-DNA spectra. Additionally, it is assumed that there are no contributions to the overall spectrum of the sample due to interactions between the *ds*-DNA and *ss*-DNA strands. Applying this model to the UV-visible spectra of uncomplexed A-tract (A<sub>3</sub>T<sub>3</sub>) and alternating AT ((AT)<sub>3</sub>) DNA sequences yields simple melting curves (Fig.4.7).

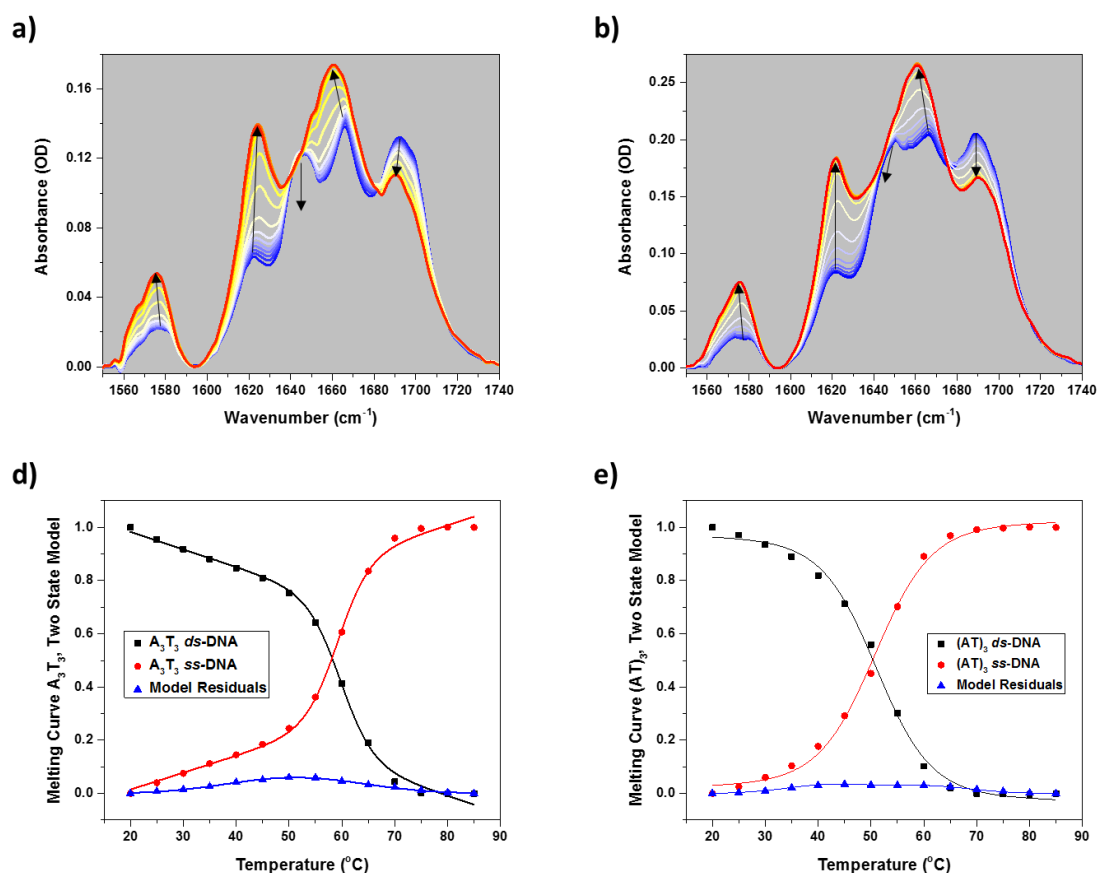


**Figure 4.7:** UV-visible absorbance spectra of the a)  $A_3T_3$  and b)  $(AT)_3$  sequences at 5 °C intervals between 20 – 95 °C; Melting curves for the c)  $A_3T_3$  and d)  $(AT)_3$  sequences extracted by the application of a two state model to the UV-visible spectra. The residuals of the two state model are shown in blue.

These melting curves are found to show the usual sigmoidal behaviour and fitting these curves yields melting temperatures of 59 °C for the  $A_3T_3$  sequence and 51 °C for the  $(AT)_3$  sequence. The model residuals obtained upon the application of the two state model to the UV-visible spectra of these two sequences (blue, Figs. 4.7.(c)&7.(d)) were found to be vanishingly small and randomly distributed. Together these two features of the model residuals show that the UV-visible spectra suggest that the melting of these two mixed A-T/G-C sequences indeed occurs via the simple two state mechanism. However, it is noted that the UV-visible absorbance of the two different sequences (Figs. 4.7.(a)&(b)) are very similar, showing a single absorbance peak centred at ~260 nm.

In contrast to this, the FT-IR spectra of the  $A_3T_3$  and  $(AT)_3$  sequences are noted to contain several different peaks arising from the different vibrational DNA base modes (assigned in

Table 4.1). The FT-IR spectra for the A-tract and alternating AT sequences, as the sample temperature increases, are shown in Figs. 4.8.(a)&(b).

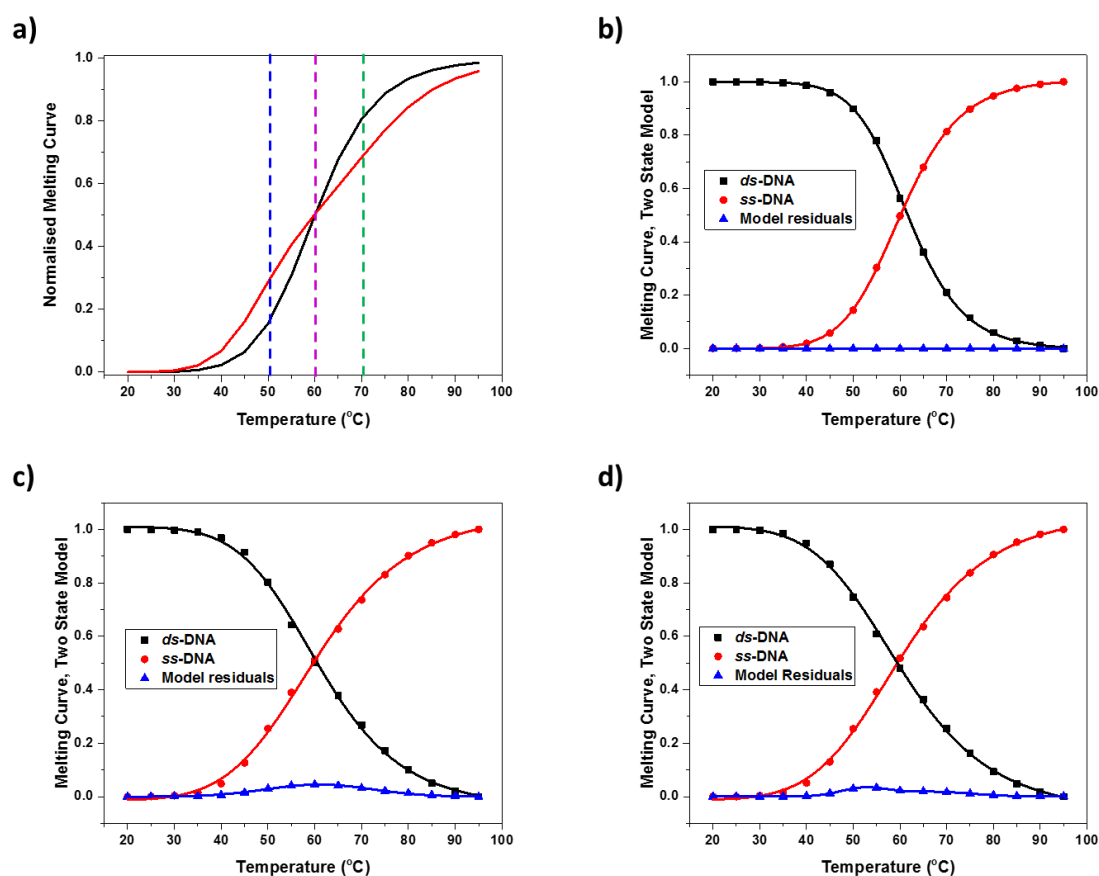


**Figure 4.8:** FT-IR spectra of the a)  $A_3T_3$  and b)  $(AT)_3$  sequences at 5 °C intervals between 20 – 95 °C; c) two state model coefficients of ss (red) and ds (black)  $A_3T_3$  FT-IR spectra against temperature, and d) two state model coefficients of ss (red) and ds (black)  $(AT)_3$  FT-IR spectra against temperature. The residuals of the two state model are shown in blue.

Comparing the high (Figs. 4.8.(a)&(b), red) and low (Figs. 4.8.(a)&(b), blue) temperature spectra for the  $A_3T_3$  and  $(AT)_3$  sequences highlights the changes that the DNA spectrum undergoes upon melting. The most prominent change in these spectra is the reduction of the number of resolvable peaks from five in the *ds*-DNA spectra (Figs. 4.8.(a)&(b), blue) to only four in the *ss*-DNA spectra (Figs. 4.8.(a)&(b), red). These changes in the base vibrational modes are thought to be largely driven by the loss of the W-C hydrogen bonds upon melting. However, additional smaller changes in the FT-IR spectra are thought to originate from the other processes occurring upon the melting of the duplexes such as the loss of base stacking and the loss of the spine of hydration as well as the solvation of the DNA bases in the single strands.

Applying the two state model to the FT-IR spectra of the two sequences again yields simple melting curves (Figs. 4.8.(c)&(d)). These curves are sigmoidal in nature and yield melting temperatures of 60 °C and 51 °C for the A-tract and alternating AT sequences, respectively, which is fully consistent with the melting temperatures extracted from the UV-visible spectra. However, the application of the two state model to the FT-IR spectra is found to result in small and non-randomly distributed residuals. These features of the residuals suggest that the changes in the FT-IR spectra and therefore the melting of the mixed A-T/G-C DNA duplexes cannot be explained by a simple two state model.

In order to understand gain a deeper understanding of the appearance of these model residuals the response of the two state model to different melting mechanisms of the duplex was explored using model spectra. These model spectra were produced using the double stranded and single stranded FT-IR spectra of exclusively AT and GC containing duplexes. It is hoped that the analysis of these artificially produced FT-IR spectra, based on three distinct melting mechanisms, will allow the mechanism underlying the melt of the DNA sequences to be determined from the distinct residuals produced upon the two state model analysis of the artificial spectra. The first model assumes that both the AT and GC base pairs melt simultaneously (black, Fig.4.9.(a)), and is primarily included to act as a negative control. The second model proposes that the GC base pairs melt first (referred to as unzipping) (red, Fig.4.9.(a)) and finally the last model explores the case in which the AT base pairs melt before the GC base pairs (referred to as bubble formation) (red, Fig.4.9.(a)). Here, the unzipping and bubble formation models of the melt were chosen as these processes were previously identified, in a recent mass spectroscopy study,<sup>225</sup> as possible dissociation mechanisms for mixed AT/GC DNA sequences.

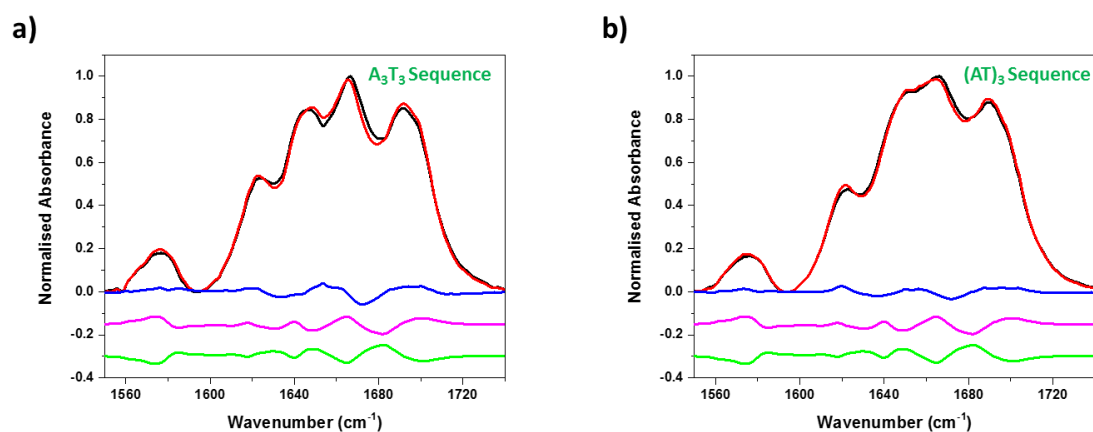


**Figure 4.9:** a) Possible melting curves for the  $A_3T_3$  and  $(AT)_3$  sequences in the case where both the base pairs melt simultaneously (black) or melt at separate temperatures (red). The melting curves of the two state model applied to the synthetic FT-IR spectrum of a duplex where b) both base pairs melt simultaneously, c) GC base pairs melt first and d) AT base pairs melt first.

The results of the application of the two state model to the synthetic data modelling the simultaneous, unzipping and bubble formation mechanisms underlying the melting transition within the synthetic FT-IR data are shown in figure 4.9. In the case of the simultaneous melting of the AT and GC bases (Fig. 4.9.(b)), the two state model extracts the usual sigmoidal melting curves from the FT-IR spectra, without any residuals. This is fully consistent with the melting of both DNA types of bases simultaneously, which would indeed lead to the duplex melt following a two state model as seen for the exclusively AT DNA sequence (Fig.A4.2). Although the application of the two state model to the synthetic spectra where one of the base pairs melts first (Fig. 4.9.(c)&(d)) results in melting curves being extracted from the data, in both of these cases the model produces small but non-randomly distributed residuals (Fig. 4.9.(c)&(d), blue). The presence of these structured residuals in the case where one of the DNA bases melt first is consistent with the results observed when the

two state model is applied to the  $A_3T_3$  and  $(AT)_3$  FT-IR spectra. This suggests that these two mixed AT/GC sequences melt via either the unzipping or bubble formation mechanisms.

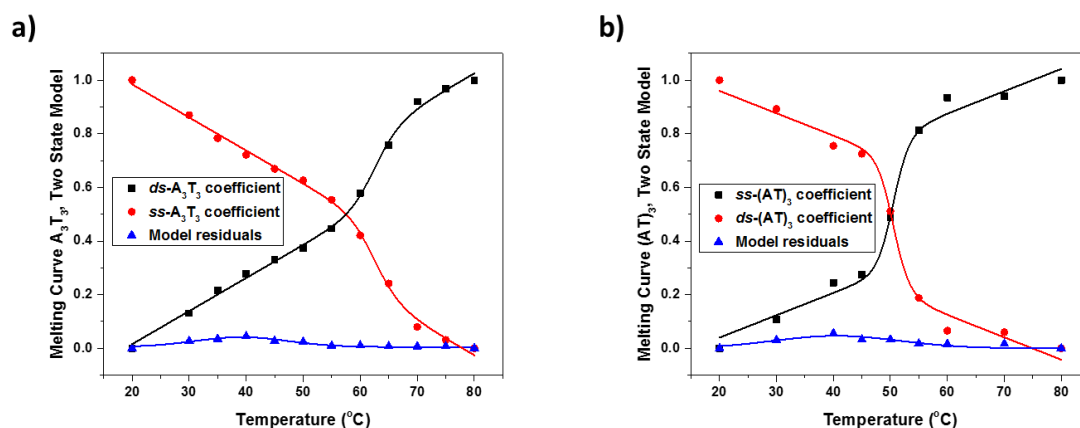
In order to determine which of the two proposed melting mechanisms underlie the thermal denaturation of the  $A_3T_3$  and  $(AT)_3$  sequences the spectral structure of the residuals associated with the synthetic data following these proposed methods are compared to the experimentally observed residuals (Figs. 4.10.(a) & (b)).



**Figure 4.10:** Spectral structure of the two state model residuals (blue) when applied to a) the  $A_3T_3$  FT-IR spectra at 50 °C, b) the  $A_3T_3$  FT-IR spectra at 45 °C. The unzipping at 50 °C and bubble formation model spectra residuals at 50 °C are shown in pink and green (offset).

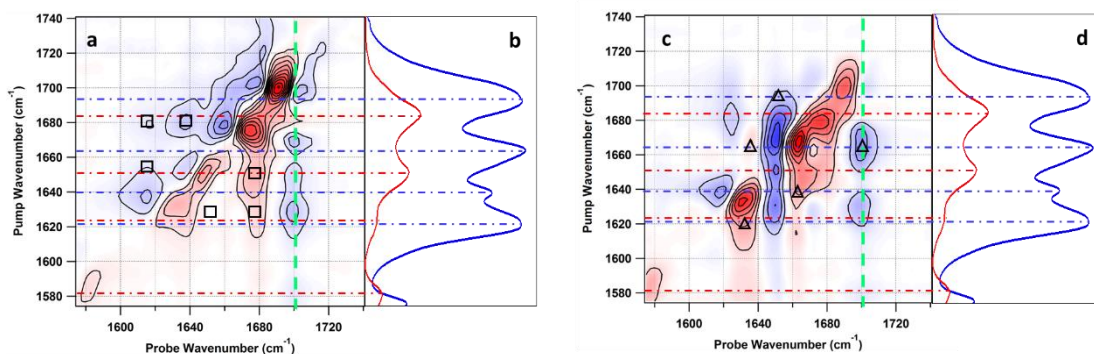
The spectral structure of the residuals obtained from the synthetic FT-IR spectra of duplexes melting via the unzipping and bubble formation mechanisms are found to consist of a series of positive and negative peaks (Figs. 4.10.(a)&(b), blue). Interestingly the patterns of the residuals for the two different mechanisms are found to be identical but completely out of phase with each other. Comparing the spectral structure of the residuals of these two mechanism to those obtained from the  $A_3T_3$  and  $(AT)_3$  sequences suggests that these duplexes melt via the unzipping mechanism.

The melting of the  $A_3T_3$  and  $(AT)_3$  sequences were further investigated by 2D-IR spectroscopy. The melting curves, extracted utilising the two state model, for the  $A_3T_3$  and  $(AT)_3$  sequences are shown in Fig. 4.11.(a)&(b). For the  $A_3T_3$  sequence the melting curve yielded a melting temperature of 64 °C. The melting temperature determined for the  $(AT)_3$  complex was found to be 52 °C.



**Figure 4.11:** Melting curve, extracted utilising the two state model, from the 2D-IR spectra of the a)  $A_3T_3$  and b)  $(AT)_3$  sequences.

The melting curves obtained from the 2D-IR spectra of the  $A_3T_3$  and  $(AT)_3$  sequences (Figs. 4.11.(a)&(b)), obtained via the two state model, lead to higher temperatures being observed for the two sequences. In addition to the slightly higher melting temperatures both melting curves are found to contain a significant linear slope. The difference in the overall appearance of these melting curves cannot be revealed by the two state model, therefore temperature-induced 2D-IR difference spectra are generated to identify the physical changes leading to the initial sloped section and the sigmoidal step change of these melting curves.



**Figure 4.12:** Temperature-induced 2D-IR difference spectrum for the  $A_3T_3$  sequence a) observed for an increase in temperature from 20 °C to 50 °C and c) observed for an increase in temperature from 55 °C to 80 °C. Plots b) and d) are FT-IR spectra of GC and AT-only DNA sequences to show the relative (weighted per base) magnitudes of peaks.

The temperature-induced 2D-IR difference spectrum for the  $A_3T_3$  sequence when the sample is heated from 20 °C to 50 °C is shown in Fig.4.12.(a). By comparison with exclusively AT and GC FT-IR spectra (Fig.4.12.(b)), features located at approximately 1580  $\text{cm}^{-1}$ , 1625  $\text{cm}^{-1}$ , 1650  $\text{cm}^{-1}$  and 1675  $\text{cm}^{-1}$  on the spectrum diagonal are can be assigned as the  $G_{RC_S}$ ,  $G_{SC_R}$ ,  $G_{SC_S}(+)$

and  $G_5C_5(-)$  vibrational modes, respectively. These assignments are supported by the distinct GC cross-peak pattern seen between these features in the difference spectrum (Fig.4.12.(a), squares). Finally the changes observed at  $1700\text{ cm}^{-1}$  on the spectrum diagonal is assigned to the  $AT_{2s}$  mode, this assignment is supported by the appearance of cross-peaks to the positions of other AT modes (Fig.4.12.(b), green dashed line). The assignments of the features in this temperature-induced difference spectrum are further supported by difference pump slices at the  $G_5C_5(-)$  (red, Fig.A4.4.(a)) and  $AT_{2s}$  (blue, Fig.A4.4.(a)) modes, which are found to only contain cross-peaks at other *ds*-GC or *ds*-AT positions, respectively.

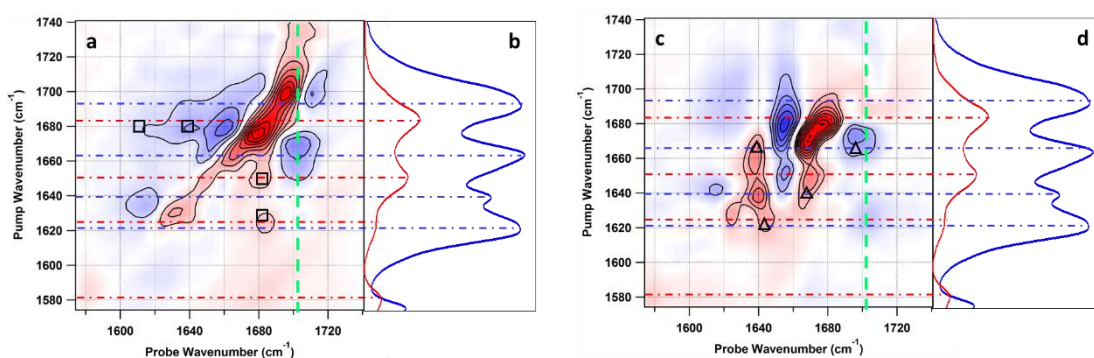
The temperature-induced changes when the sequence is then heated beyond  $50\text{ }^\circ\text{C}$  to  $80\text{ }^\circ\text{C}$  are illustrated by the 2D-IR difference spectrum shown in Fig.4.12.(c). There are found to be four features located at approximately  $1640\text{ cm}^{-1}$ ,  $1662\text{ cm}^{-1}$ ,  $1675\text{ cm}^{-1}$  and  $1700\text{ cm}^{-1}$ . Comparison to the FT-IR spectra of exclusively AT and GC sequences leads these features arising from changes in the  $T_R$ ,  $AT_{4s}$ ,  $G_5C_5(-)$  and  $AT_{2s}$  modes, respectively. The assignment of the majority of the features arising from changes in the AT modes is supported by the presence of cross-peaks between these features (triangles, Fig.4.12.(c)). Finally, it is noted that even though there is a contribution to the high temperature difference spectrum from the  $G_5C_5(-)$  mode, it is clearly dominated by changes in the AT base modes. Here, these assignments for the temperature-induced difference spectrum are further supported by difference pump slices at the  $G_5C_5(-)$  (red, Fig.A4.4.(b)) and  $AT_{2s}$  (blue, Fig.A4.4.(b)) modes, which are found to only contain cross-peaks at other *ds*-GC or *ds*-AT positions, respectively.

Broadly speaking it is noted that there are two different types of features present in all of the temperature-induced 2D-IR difference spectra. One type corresponding to a shift in frequency, indicated by three stripes of alternating sign, and the other due to a change in intensity of the mode, indicated by a simple peak pair. In the case of the  $A_3T_3$  sequence the appearance of the peak pairs assigned to *ds*-DNA GC modes in the low temperature induced 2D-IR difference spectrum suggest that there is a substantial decrease in the intensities of these modes as the temperature is increased to  $50\text{ }^\circ\text{C}$ . Additionally the absence of similar decreases in the intensities of the *ds*-DNA AT modes suggests that the first process identified in the melting curve of the  $A_3T_3$  duplex is due to the melting of the GC ends of the duplex. Furthermore the dominance of *ds*-AT modes in the high temperature-induced difference spectrum, indicating that these modes are predominantly lost as the temperature is increased beyond  $50\text{ }^\circ\text{C}$ , lends additional support to the hypothesis that the GC base pairs



melt before the AT base pairs in the  $A_3T_3$  sequence. The shift to higher frequency of the  $AT_{25}$  mode is interesting as the  $T_2$  carbonyl, the dominant contribution to this mode, is not involved in the W-C H-bonding but rather points into the minor groove forming H-bonds to the spine of hydration. This shift to higher frequency therefore suggests that as the temperature increases the H-bonds between these carbonyls and the spine of hydration weaken, showing this mode can be used to directly probe changes in the DNA minor groove.

For the  $(AT)_3$  sequence the temperature-induced changes in the spectrum after the initial process are shown in the difference 2D-IR spectrum at 40 °C (Fig.4.13.(a)). There are three dominant features located at 1625  $cm^{-1}$ , 1675  $cm^{-1}$  and 1700  $cm^{-1}$ , alongside two weaker ones located at 1650  $cm^{-1}$  and 1666  $cm^{-1}$  on the spectrum diagonal. The dominant changes can be assigned, by comparisons to exclusively AT and CG FT-IR spectra (Fig.4.13.(b)), as the  $G_5C_R$ ,  $G_5C_5(-)$  and  $AT_{25}$  modes, respectively whereas the weaker features arise from the  $G_5C_5(+)$  and  $AT_{45}$  modes respectively. The assignment of the changes arising from the GC modes are supported by the cross-peaks between these features (Fig.4.13.(a), squares). The assignment of the feature at 1700  $cm^{-1}$  as arising from the  $AT_{25}$  mode is also supported by the appearance of cross-peaks to positions consistent with other AT modes (Fig.4.13.(a), green dashed line). As for the temperature induced spectra for the  $A_3T_3$  sequence, these assignments are further supported by difference pump slices at the  $G_5C_5(-)$  (red, Fig.A4.5.(a)) and  $AT_{25}$  (blue, Fig.A4.5.(a)) modes, which are found to only contain cross-peaks at other  $ds$ -GC or  $ds$ -AT positions, respectively.



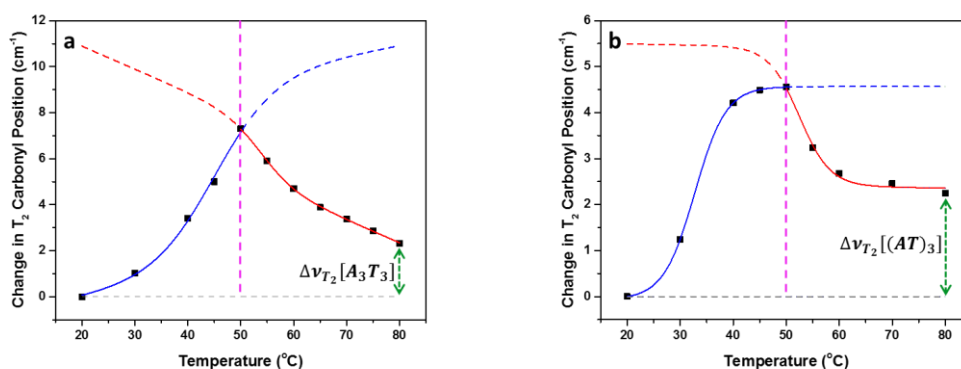
**Figure 4.13:** Temperature-induced 2D-IR difference spectrum for the  $(AT)_3$  sequence a) observed for an increase in temperature from 20 °C to 40 °C and c) observed for an increase in temperature from 45 °C to 80 °C. Plots b) and d) are FT-IR spectra of GC and AT-only DNA sequences to show the relative (weighted per base) magnitudes of peaks.

For the  $(AT)_3$  sequence the high temperature 2D-IR difference spectrum is shown in fig.4.13.(c). As the temperature of the sample is increased beyond 40 °C and the duplex fully melts, there are five distinct features near 1625  $\text{cm}^{-1}$ , 1640  $\text{cm}^{-1}$ , 1665  $\text{cm}^{-1}$  and 1675  $\text{cm}^{-1}$ . Comparing the positions of these features to the FT-IR spectra of exclusively AT and GC sequences shows that these features in the high temperature difference spectrum arise from changes in the  $A_{R1T}$ ,  $T_R$ ,  $AT_{4S}$  and  $G_5C_5(-)$  modes respectively, as the temperature increases beyond 40 °C. The assignment of the majority of the features arising from changes in the AT modes is supported by the presence of cross-peaks between these features (triangles, Fig.4.13.(c)). As with the high temperature  $A_3T_3$  difference spectrum, it is noted that even though there is a contribution to the high temperature difference spectrum from the  $G_5C_5(-)$  mode, it is clearly dominated by changes in the AT base modes. Again, these assignments are further supported by difference pump slices at the  $G_5C_5(-)$  (red, Fig.A4.5.(b)) and  $AT_{2S}$  (blue, Fig.A4.5.(b)) modes, which are found to only contain cross-peaks at other *ds*-GC or *ds*-AT positions, respectively.

In the temperature-induced 2D-IR difference spectrum of the  $(AT)_3$  sequence there are three peak pairs found to correspond to *ds*-DNA GC vibrational modes (Fig.4.13.(a)). The appearance of these features indicate a significant loss in the intensity of the GC *ds*-DNA vibrational modes as the temperature is increased to 40 °C. As with the A-tract sequence this demonstrates that the initial slope of the melting curve originates from the melting of the GC ends of the duplex. The presence of a weak peak pair arising from the  $AT_{4S}$  mode suggests that at 40 °C there is a small decrease in the intensity of the *ds*-DNA AT modes, this is consistent with the start of a melting process of the central AT base pairs of the duplex thought to underlie the sigmoidal step within the melting curve. As previously noted for the  $A_3T_3$  sequence, the features observed in the high temperature 2D-IR difference spectrum of the  $(AT)_3$  sequence is dominated by features assigned to *ds*-AT modes. These features are consistent with the loss of these modes suggesting that the central AT base pairs of the  $(AT)_3$  sequence melt at a high temperature than the GC-ends of the duplex. Finally, it is also noted that as the temperature increased to 40 °C the  $AT_{2S}$  mode of the  $(AT)_3$  sequence is also seen to undergo a shift to higher frequency. This is consistent with the weakening of the hydrogen bonds formed between these carbonyls and the spine of hydration. This illustrates the melting of both the  $A_3T_3$  and  $(AT)_3$  DNA sequences follow an unzipping mechanism, where the GC ends of the duplex undergoes a melting transition, before the central AT base pairs, leading to duplex fraying and at a higher temperature unzipping.

### Minor Groove

The  $AT_{25}$  mode of both the  $A_3T_3$  and  $(AT)_3$  duplexes was found to shift to higher frequency with increasing temperature, consistent with the weakening of the H-bonds between the  $T_2$  carbonyls and the spine of hydration. Allowing the changes in the minor groove to be investigated as the temperature increases to be probed via the  $AT_{25}$  mode. The changes in the position of the  $T_2$  carbonyl stretching mode as the temperature is increased for both sequences was extracted from the temperature-induced 2D-IR difference spectra over the experimental temperature range are shown in figure 4.14.



**Figure 4.14:** Change in the position of the  $T_2$  carbonyl stretch of the a)  $A_3T_3$  sequence and b)  $(AT)_3$  sequence with increasing temperature.

As the temperature increases the  $AT_{25}$  mode in both sequences are initially undergo a shift to higher frequency (Figs.4.14.(a)&(b), blue). However, at 50 °C there is an abrupt change in the behaviour and the mode starts to shift back to lower frequency (Figs.4.14.(a)&(b), red). However the overall change in the position of the  $T_2$  carbonyls,  $\Delta\nu_{T_2}[A_3T_3]$  and  $\Delta\nu_{T_2}[(AT)_3]$  are noted to be 2.33  $\text{cm}^{-1}$  and 2.25  $\text{cm}^{-1}$ , respectively, consistent with the overall shift to higher frequency of the  $T_2$  carbonyls observed in the FT-IR spectrum of an exclusively AT sequence upon melting (Fig.4.3.(a)&A4.3.(a)). The change in the direction of the shift in the position of the  $T_2$  carbonyls at 50 °C is indicative of two different processes occurring as the DNA duplex melts. At low temperatures the dominant process involves a decrease in H-bonding strength whereas at temperatures above 50 °C the dominant process involves an increase in H-bonding strength. Fitting the changes in the position of the  $T_2$  carbonyls as the temperature increases using a piecewise function consisting of two sigmoid curves yields transition temperatures for these two different processes. For the  $A_3T_3$  sequence these transition temperatures were found to be 46 °C and 54 °C, respectively, whereas for the  $(AT)_3$  sequence these two transitions are centred at 33 °C and 51 °C, respectively.

Comparing the two sequences it is interesting to note that the temperature at which the  $T_2$  carbonyls are exposed to the surrounding solvent is almost identical in the  $A_3T_3$  (54 °C) and  $(AT)_3$  (51 °C) sequences shows that this process is largely unaffected by replacing the A-tract sequence with its alternating counterpart. However it can also be noted that the maximum shift to higher frequency of the  $T_2$  carbonyls is greater for the  $A_3T_3$  sequence (7.3  $\text{cm}^{-1}$ ) than for the  $(AT)_3$  sequence (4.6  $\text{cm}^{-1}$ ). This suggests that the interaction between the sequence and the spine of hydration is greater for the  $A_3T_3$  sequence and it is thought that this may be a reflection of the combination of the larger number of water molecules and the greater degree of order in the spine of hydration associated with this sequence.

The shift to higher frequency as the temperature is initially increased is fully consistent with a decrease in the strength of the H-bonds present between the  $T_2$  carbonyls in the DNA duplex and the spine of hydration. It is expected that upon melting the spine of hydration is lost and so it is proposed that this process drives the initial increase in frequency of the  $AT_{25}$  mode. Interestingly, in both the  $A_3T_3$  and  $(AT)_3$  sequences it is noted that the transition temperature at which the spine of hydration is lost are almost identical to the melting temperature of the GC ends of the duplex. It is thought that this indicates that the fraying of the ends of the DNA duplex triggers the loss of the spine of hydration.

The change in the direction of the shift of the  $T_2$  carbonyls after the spine of hydration has been lost, when the temperature increases beyond 50 °C, is thought to be consistent with the DNA duplexes unwinding. The unwinding of the double helix of the duplex is thought to expose the  $T_2$  carbonyls to the molecular environment surrounding the duplexes allowing new H-bonds to be formed and so explaining the change in the direction of the shift. Here, it is observed that the sequence of the central AT region of the duplexes effects the overall mechanism of the melting transition. In the case of the A-tract ( $A_3T_3$ ) sequence the unwinding of the duplex occurs at least 10 °C before the central AT base pairs melt, resulting in the existence of an unwound intermediate stage as this duplex melts. In the alternating ( $(AT)_3$ ) sequence the unwinding of the duplex happens as the central AT base pairs melt, this concerted unwinding and melting means that the alteration of the sequence has resulted in the loss of a stable unwound intermediate in the melting mechanism of the duplex.

#### 4.4 Conclusions

Overall the results show the utility of 2D-IR spectroscopy as a technique for the investigation of DNA sequences and the changes that an alteration in the sequence can impart on both the

structure of the duplex and the mechanism by which it dissociates as it undergoes thermal denaturation. Using the DNA base vibrational modes as natural IR probes, difference spectroscopy illuminates that when the A-tract within the DNA oligonucleotide is replaced with its alternating counterpart there is a  $4.8 \text{ kJmol}^{-1}$  decrease in the strength of Watson-Crick hydrogen bonds involving the T<sub>4</sub> carbonyl. This is accompanied by a loss in the number of hydrogen bonds between the spine of hydration and the T<sub>2</sub> carbonyls. These changes in these two vibrational modes show that the alteration of the order of the AT region of the DNA oligonucleotide leads to the loss of three-centred inter-strand hydrogen bonds as well as a decrease in both the order and number of water molecules present in the minor groove spine of hydration. Additionally this highlights the sensitivity of 2D-IR spectroscopy to perturbations in the structure of DNA duplexes and alterations in the hydrogen bonding strengths. The extension of the 2D-IR study to the thermal denaturation of the two DNA sequences allows the mechanism underlying this transition to be studied. This study determined that for both of these short oligonucleotides this follows an unzipping mechanism. This mechanism was determined to consist of the fraying of the GC ends of the duplexes. The fraying of the GC ends and the loss of the minor groove spine of hydration were found to occur within the same temperature range, suggesting that the fraying of the duplexes leads to the loss of the spine of hydration for both sequences. In the A<sub>3</sub>T<sub>3</sub> sequence it was noted that this loss of the spine of hydration was followed by the unwinding of the duplex, occurring at approximately 54 °C. The melting mechanism of the A<sub>3</sub>T<sub>3</sub> sequence then terminates with the loss of the W-C hydrogen bonds between the AT base pairs at 64 °C, completing the unzipping of the strands. In contrast to this it was determined that following the loss of the spine of hydration the (AT)<sub>3</sub> sequence unwinds and the AT W-C hydrogen bonds are lost in a concerted unwinding/unzipping process which occurs over the temperature range of 50 °C to 55 °C. These mechanisms highlight the change in the melting mechanism caused by the presence of an A-tract which introduces an unwound duplex intermediate into this mechanism and so delaying the final unzipping of the strands. This further demonstrates the versatility of 2D-IR as a technique to gain a more fundamental understanding of important DNA processes.

## 4.5 Methods

### 4.5.1 Materials

The lyophilised, salt-free DNA oligonucleotides were obtained from Eurofins and used without further purification. All other chemicals used were obtained from Sigma-Aldrich and

were used without further purification. All the samples used in the UV-visible and IR experiments were prepared in pD 7 phosphate buffer solution at a range of concentrations. The DNA oligonucleotides were made up to 10mM stocks in the D2O phosphate buffer solution and diluted to 2.5mM. The prepared samples were then annealed at 90 °C for 10 minutes. All the other concentrations used were prepared by serial dilution from the original 2.5mM samples.

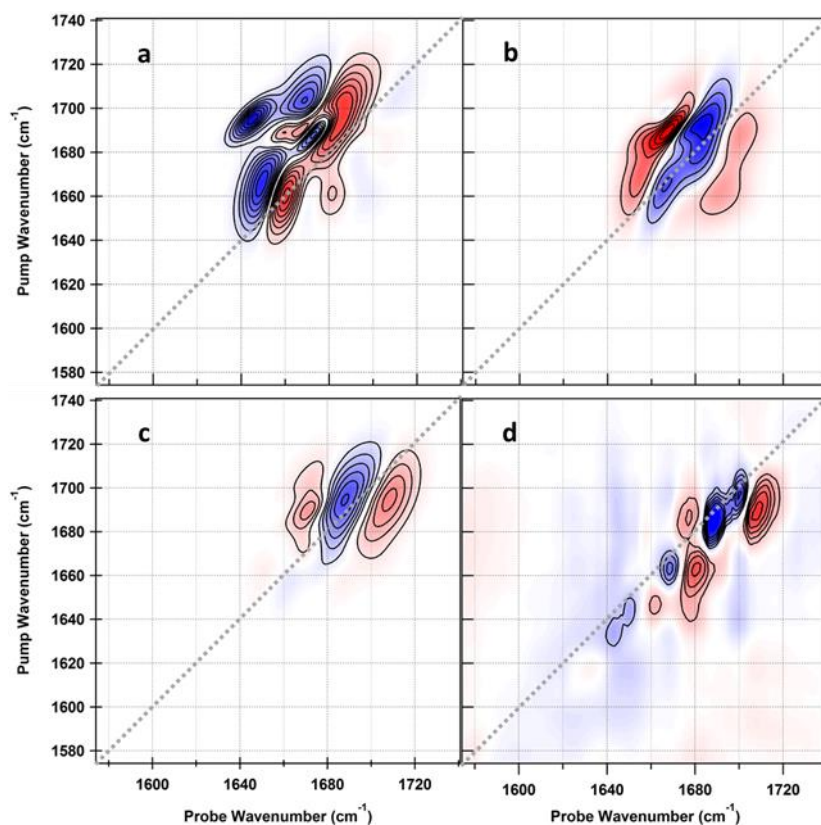
#### **4.5.2 FT-IR and UV-visible Spectroscopy**

All of the FTIR experiments were carried out using a Bruker Vertex 70 spectrometer at a resolution of 1  $\text{cm}^{-1}$ . The samples were held in a demountable Harrick cell utilising  $\text{CaF}_2$  windows and a 50  $\mu\text{m}$  polytetrafluoroethylene spacer. All of the UV/visible measurements were carried out using a Perkin Elmer Lambda 25 UV-visible spectrometer at a resolution of 1 nm. All the UV-vis measurements were carried out using a demountable Harrick cell utilising  $\text{CaF}_2$  windows and polytetrafluoroethylene spacer (the spacer thickness adjusted for the sample concentration). All of these measurements were carried out at a series of temperature within the experimental temperature range of 20°C – 90°C.

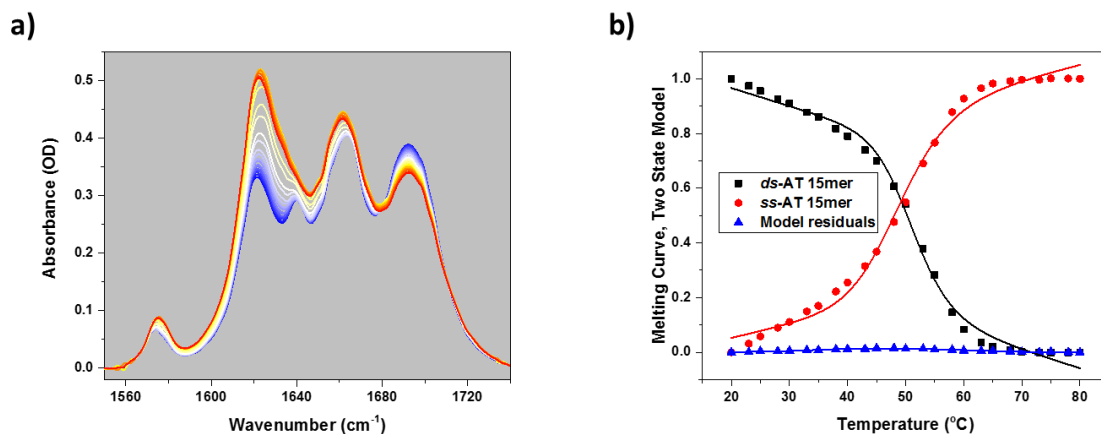
#### **4.5.3 2D-IR Spectroscopy**

2D-IR spectra were collected using the ULTRA FT-2D-IR spectrometer as previously described.<sup>247,248</sup> This employs three mid-infrared pulses (two ‘pump,’ one ‘probe’) arranged in a pseudo pump-probe beam geometry. The pulses used had a temporal duration of ~ 100fs; a centre frequency of 1650  $\text{cm}^{-1}$  and a bandwidth of ~300  $\text{cm}^{-1}$ , at a repetition rate of 10 kHz.

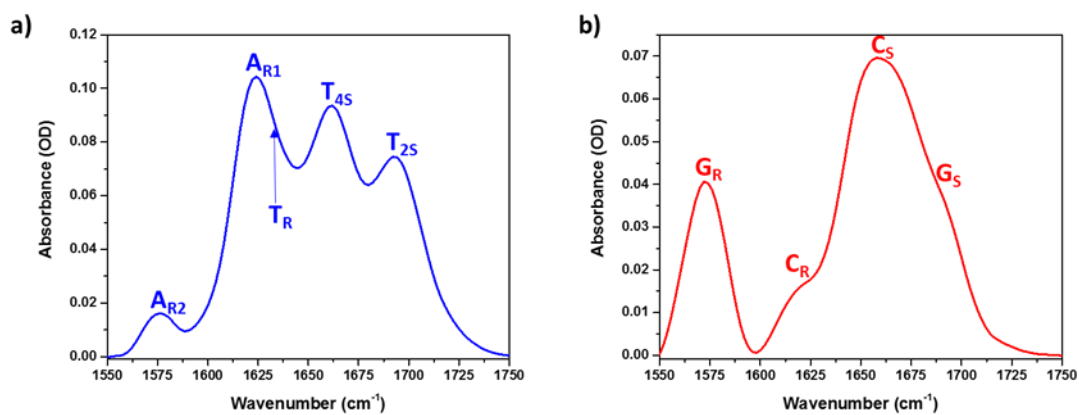
## 4.6 Appendix



**Figure A4.1:** a) Model difference 2D-IR spectra due to an increase in mode anharmonicity, b) Model difference 2DIR spectra due to a decrease in mode anharmonicity, c) Model difference 2D-IR spectra due to a global blue shift of the modes and d) Experimentally observed difference 2D-IR due to alteration in the AT sequence (included to demonstrate the departure of these models from the experimentally observed data).

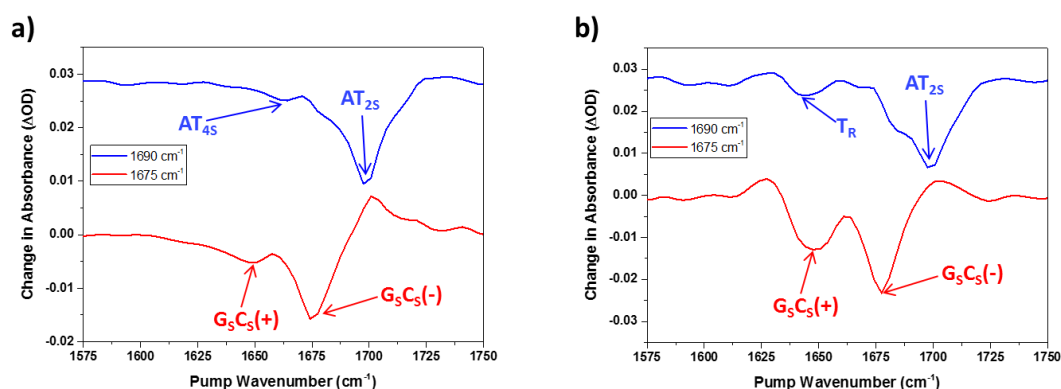


**Figure A4.2:** FT-IR spectra of the a) AT 15-mer between 20 – 95 °C and the b) melting curves for the AT 15-mer sequence extracted by the application of a two state model to the FT-IR spectra, residuals of the model are shown in blue.

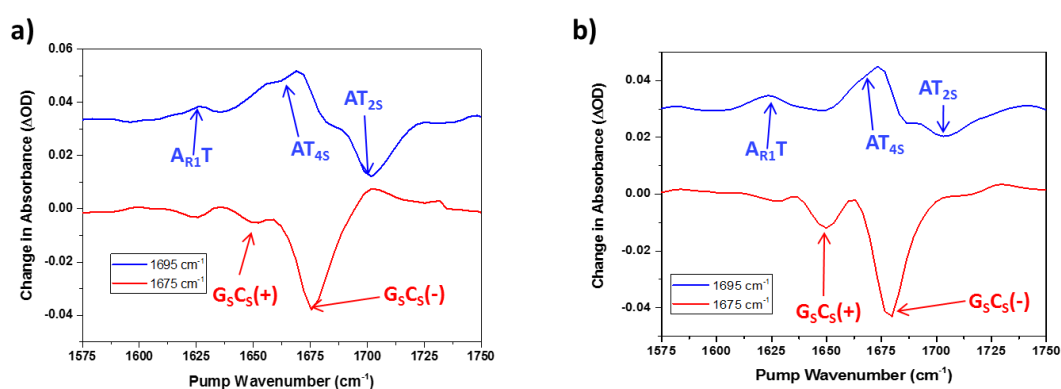


**Figure A4.3:** Assigned FTIR spectra of a) AT only single strand (sequence: 5'-ATTATTATTATTA-3') and b) GC only single strand (sequence: 5'-GCCGCCGCCG-3'). All data recorded at 90 °C.





**Figure A4.4:** Assigned probe slices at the  $AT_{2s}$  (blue, offset by 0.03) and  $G_5C_5(-)$  (red) mode positions of the a) 20°C to 50°C and b) 55°C to 80°C temperature-induced 2D-IR difference spectra of the  $A_3T_3$  sequence.



**Figure A4.5:** Assigned probe slices at the  $AT_{2s}$  (blue, offset by 0.03) and  $G_5C_5(-)$  (red) mode positions of the a) 20°C to 40°C and b) 45°C to 80°C temperature-induced 2D-IR difference spectra of the  $(AT)_3$  sequence.

## 4.7 References

- (176) Watson, J. D.; Crick, F. H. C. Molecular Structure of Nucleic Acids, *Nature*, **1953**, 4356, 737.
- (177) Franklin, R. E.; Gosling, R. G. The Structure of Sodium Thymonucleate Fibres. I. The influence of Water Content, *Acta. Cryst.*, **1953**, 6, 673.
- (178) Cooper, P. J.; Hamilton, L. D. The A-B conformational change in the sodium salt of DNA, *J. Mol. Bio.*, **1966**, 2, 562.
- (179) Bradbury, E. M.; Price, W. C.; Wilkinson, G. R. Infrared studies of molecular configurations of DNA, *J. Mol. Bio.*, **1961**, 3, 301.
- (180) Wood, B. R.; The importance of hydration and DNA conformation in interpreting infrared spectra of cells and tissues, *Chem. Soc. Rev.*, **2016**, 45, 1980.
- (181) Yoon, C.; Privé, G. G.; Goodsell, D. S.; Dickerson, R. E.; Structure of an alternating-B DNA helix and its relationship to A-tract DNA. *Proc. Natl. Acad. Sci. USA*, **1988**, 85, 6332.

- 
- (182) Dickerson, R. E.; DNA structure from A to Z., *Methods in Enzymology*, **1992**, 211, 67.
- (183) Marini, J. C.; Levene, S. D.; Crothers, D. M.; Englund, P. T.; Bent helical structure in kinoplast DNA. *Proc. Natl. Acad. Sci. USA*, **1988**, 79, 7664.
- (184) Levene, S. D.; Wu, H. M.; Crothers, D. M.; Bending and flexibility of kinetoplast DNA. *Biochemistry*, **1986**, 25, 3988.
- (185) Koo, H. S.; Drak, J.; Rice, J. A.; Crothers, D. M.; Determination of the extent of DNA bending by an adenine-thymine tract. *Biochemistry*, **1990**, 229, 4227.
- (186) Calladine, C. R.; Drew, H. R.; McCall, M. J.; The intrinsic curvature of DNA in solution. *J. Mol. Biol.*, **1988**, 201, 127.
- (187) Ulanovsky, L.; Bodner, M.; Trifonov, E. N.; Choder, M.; Curved DNA: design, synthesis and circularization. *Proc. Natl. Acad. Sci. USA*, **1986**, 83, 862.
- (188) Haran, T. E.; Mohanty, U.; The unique structure of A-tracts and intrinsic DNA bending, *Quarterly Reviews of Biophysics*, **2009**, 42, 41.
- (189) Perez-Martin, J.; Rojo, F.; De Lorenzo, V.; Promoters responsive to DNA bending: a common theme in prokaryotic gene expression. *Microbiology Reviews*, **1994**, 58, 268.
- (190) Hagerman, P. J.; Sequence-directed curvature of DNA. *Annual Reviews in Biochemistry*, **1990**, 59, 755.
- (191) Gimenes, F.; Takeda, K. I.; Fiorini, A.; Gouveia, F. S.; Fernandez, M. A.; Intrinsically bent DNA in replication origins and gene promoters. *Genetics and Molecular Research*, **2008**, 7, 549.
- (192) Abu-Daya, A.; Brown, P. M.; Fox, K. R. DNA Sequence Preferences of Several AT-Selective Minor Groove Binding Ligands, *Nucleic Acids Research*, **1995**, 23, 3385.
- (193) Drobyshev, A. L.; Zasedatelev, A. S.; Yershov, G. M.; Mirzabekov, A. D. Massive Parallel Analysis of DNA-Hoechst33258 Binding Specificity with a Generic Oligodeoxyribonucleotide Microchip. *Nucleic Acids Research*, **1999**, 27, 4100.
- (194) Breusegem, S. Y.; Clegg, R. M.; Loontjens, F. G. Base-Sequence Specificity of Hoechst33258 and DAPI Binding to Five (A/T)<sub>4</sub> DNA Sites with Kinetic Evidence for More Than One High-Affinity Hoechst33258-AATT Complex, *J. Mol. Biol.*, **2002**, 315, 1049.
- (195) Kintanar, A.; Klevit, R. E.; Reid, B. R.; Two-dimensional NMR investigation of a bent DNA fragment: assignment of the proton resonances and preliminary structure analysis. *Nucleic Acids Research*, **1987**, 15, 5845.
- (196) Katahira, M.; Sugeta, H.; Kyogoku, Y.; Fujii, S.; Fujisawa, R.; Tomita, K.; One- and two-dimensional NMR studies on the conformation of DNA containing the oligo(dA)oligo(dT) tract. *Nucleic Acids Research*, **1988**, 16, 8619.
- (197) Nadeau, J. G.; Crothers, D. M.; Structural basis for DNA bending. *Proc. Natl. Acad. Sci. USA*, **1989**, 86, 2622.
- (198) Moe, J. G.; Russo, I. M.; Proton exchange and base-pair opening kinetics in 5'-d(CGCGAATTCGCG)-3' and related dodecamers. *Nucleic Acids Research*, **1990**, 18, 821.

- 
- (199) Shatzky-Schwartz, M.; Arbuckle, N. D.; Eisenstein, M.; Rabinovich, D.; Bareket-Samish, A.; Haran, T. E.; Luisi, B. F.; Shakked, Z.; X-ray and solution studies of DNA oligomers and implications for the structural basis of A-tract-dependent curvature. *J. Mol. Biol.*, **1997**, *267*, 595.
- (200) DiGabriele, A. D.; Steitz, T. A.; A DNA dodecamer containing an adenine tract crystallizes in a unique lattice and exhibits a new bend. *J. Mol. Biol.*, **1993**, *231*, 1024.
- (201) DiGabriele, A. D.; Sanderson, M. R.; Steitz, T. A.; Crystal lattice is important in determining the bend of a DNA dodecamer containing an adenine tract. *Proc. Natl. Acad. Sci. USA*, **1989**, *86*, 1816.
- (202) Young, M. A.; Beveridge, D. L.; Molecular dynamics simulations of an oligonucleotide duplex with adenine tracts phased by a full helix turn. *J. Mol. Biol.*, **1998**, *281*, 675.
- (203) Beveridge, D. L.; Dixit, S. B.; Barreiro, G.; Thayer, K. M.; Molecular dynamics simulations of DNA curvature and flexibility: helix phasing and premelting. *Biopolymers*, **2004**, *73*, 380.
- (204) Strahs, D.; Schlick, T.; A-tract bending: insights into experimental structures by computational models. *J. Mol. Biol.*, **2000**, *301*, 643.
- (205) Ulanovsky, L. E.; Trifonov, E. N.; Estimation of wedge components in curved DNA. *Nature*, **1987**, *326*, 720.
- (206) De Santis, P.; Palleschi, A.; Savino, M.; Scipioni, A.; Validity of the nearest-neighbor approximation in the evaluation of the electrophoretic manifestations of DNA curvature. *Biochemistry*, **1990**, *29*, 9269.
- (207) Crothers, D. M.; Haran, T. E.; Nadeau, J. G.; Intrinsically bent DNA. *Journal of Biological Chemistry*, **1990**, *265*, 7093.
- (208) Manning, G. S.; The persistence length of DNA is reached from the persistence length of its null isomer through an internal electrostatic stretching force. *Biophysical Journal*, **2006**, *91*, 1.
- (209) Edwards, K. J.; Brown, D. G.; Spink, N.; Skelly, J. V.; Neidle, S.; Molecular structure of the B-DNA dodecamer d(CGCAAATTTGCG)<sub>2</sub>. An examination of propeller twist and minor-groove water structure at 2.2 Å resolution. *J. Mol. Biol.*, **1992**, *226*, 1161.
- (210) Fritsch, V.; Ravishanker, G.; Beveridge, D. L.; Westhof, E.; Molecular dynamics of poly(dA).poly(dT): comparisons between implicit and explicit solvent representations. *Biopolymers*, **1993**, *33*, 1537.
- (211) Woods, K. K.; Maehigashi, T.; Howerton, S. B.; Sines, C. C.; Tannenbaum, S. T.; Williams, L. D.; High-Resolution Structure of an Extended A-Tract: [d(CGCAAATTTGCG)]<sub>2</sub>. *J. Am. Chem. Soc.*, **2004**, *126*, 15330.
- (212) Coll, M.; Frederick, C. A.; Wang, A. H.; Rich, A.; A bifurcated hydrogen-bonded conformation in the d(A.T) base pairs of the DNA dodecamer d(CGCAAATTTGCG) and its complex with distamycin. *Proc. Natl. Acad. Sci. USA*, **1987**, *84*, 8385.
- (213) Premilat, S.; Albiser, G.; X-ray fibre diffraction study of an elevated temperature structure of poly(dA).poly(dT). *J. Mol. Biol.*, **1997**, *274*, 64.
- (214) Chan, S. S.; Austin, R. H.; Mukerji, I.; Spiro, T. G.; Temperature-Dependent Ultraviolet Resonance Raman Spectroscopy of the Premelting State of dA\*dT DNA. *Biophysical Journal*, **1997**, *72*, 1512.

- 
- (215) Mukerji, I.; Williams, A. P.; UV Resonance Raman and Circular Dichroism Studies of a DNA Duplex Containing an A<sub>3</sub>T<sub>3</sub> Tract: Evidence for a Premelting Transition and Three-Centred H-Bonds. *Biochemistry*, **2002**, *41*, 69.
- (216) Park, Y. W.; Breslauer, K. J.; A spectroscopic and calorimetric study of the melting behaviours of a 'bent' and a 'normal' DNA duplex: [d(GA4T4C)]<sub>2</sub> versus [d(GT4A4C)]<sub>2</sub>. *Proc. Natl. Acad. Sci. USA*, **1991**, *88*, 1551.
- (217) Drew, H. R.; Dickerson, R. E.; Structure of a B-DNA dodecamer. III. Geometry of hydration. *J. Mol. Biol.*, **1981**, *151*, 535.
- (218) Kopka, M. L.; Fratini, A. V.; Drew, H. R.; Dickerson, R. E.; Ordered water structure around a B-DNA dodecamer, a quantitative study. *J. Mol. Biol.*, **1983**, *163*, 129.
- (219) Kubinec, M. G.; Wemmer, D. E.; NMR Evidence for DNA Bound Water in Solution. *J. Am. Chem. Soc.*, **1992**, *114*, 8739.
- (220) Liepinsh, E.; Otting, G.; Wuthrich, K.; NMR observation of individual molecules of hydration water bound to DNA duplexes: direct evidence for a spine of hydration water present in aqueous solution. *Nucleic Acids Research*, **1992**, *20*, 6549.
- (221) Johannesson, H.; Halle, B.; Minor groove hydration of DNA in solution: dependence on base composition and sequence. *J. Am. Chem. Soc.*, **1998**, *120*, 6859.
- (222) Phan, A. T.; Leroy, J. L.; Gueron, M.; Determination of the residence time of water molecules hydrating B'-DNA and B-DNA by one-dimensional zero-enhancement nuclear Overhauser effect spectroscopy. *J. Mol. Biol.*, **1999**, *286*, 505.
- (223) Perez, A.; Orozco, M.; Real-Time Atomistic Description of DNA Unfolding. *Angew. Chem. Int. Ed.*, **2010**, *49*, 4805.
- (224) Wong, K.-Y.; Pettitt, B. M.; The Pathway of Oligomeric DNA Melting Investigated by Molecular Dynamics Simulations. *Biophysical Journal*, **2008**, *95*, 5618.
- (225) Burmistrova, A.; Gabelica, V.; Duwea, A.-S.; De Pauw, E.; Ion Mobility Spectrometry Reveals Duplex DNA Dissociation Intermediates. *J. Am. Soc. Mass. Spectrom.*, **2013**, *24*, 1777.
- (226) Ma, H.; Wan, C.; Wu, A.; Zewail, A. H.; DNA folding and melting observed in real time redefine the energy landscape. *Proc. Natl. Acad. Sci. USA*, **2007**, *104*, 712.
- (227) Jung, J.; Van Orden, A.; A Three-State Mechanism for DNA Hairpin Folding Characterised by Multiparameter Fluctuation Spectroscopy. *J. Am. Chem. Soc.*, **2006**, *128*, 1240.
- (228) Shen, Y.; Kuznetsov, S. V.; Ansari, A.; Loop Dependence of the Dynamics of DNA Hairpins. *J. Phys. Chem. B*, **2001**, *105*, 12202.
- (229) Hunt, N. T. 2D-IR spectroscopy: Ultrafast Insights into Biomolecule Structure and Function, *Chem. Soc. Rev.*, **2009**, *38*, 1837.
- (230) Park, S.; Kwak, K.; Fayer, M. D. Ultrafast 2D-IR Vibrational Echo Spectroscopy: A Probe of Molecular Dynamics, *Laser Phys. Lett.*, **2007**, *10*, 704.
- (231) Baiz, C. R.; McRobbie, P. L.; Anna, J. M.; Geva, E.; Kubarych, K. J. Two-Dimensional Infrared Spectroscopy of Metal Carbonyls, *Accounts of Chemical Research*, **2009**, *42*, 1395.

---

(232) Peng, C. S.; Jones, K. C.; Tokmakoff, A. Anharmonic Vibrational Modes of Nucleic Acid Bases Revealed by 2D IR Spectroscopy, *J. Am. Chem. Soc.*, **2011**, *133*, 15650.

(233) Yang, M.; Szyk, L.; Elsaesser, T. Femtosecond Two-Dimensional Infrared Spectroscopy of Adenine-Thymine Base Pairs in DNA Oligomers, *J. Phys. Chem. B*, **2011**, *115*, 1262.

(234) Greve, C.; Elsaesser, T. Ultrafast Two-Dimensional Infrared Spectroscopy of Guanine-Cytosine Base Pairs in DNA Oligomers, *J. Phys. Chem. B*, **2013**, *117*, 14009.

(235) Krummel, A. T.; Zanni, M. T. DNA Vibrational Coupling Revealed with Two-Dimensional Infrared Spectroscopy: Insight into Why Vibrational Spectroscopy is Sensitive to DNA Structure, *J. Phys. Chem. B*, **2006**, *110*, 13991.

(236) Hithell, G.; Shaw, D. J.; Donaldson, P. M.; Greetham, G. M.; Towrie, M.; Burley, G. A.; Parker, W. A.; Hunt, N. T. Long-Range Vibrational Dynamics Are Directed by Watson-Crick Base Pairing in Duplex DNA. *J. Phys. Chem. B*, **2016**, *120*, 4009.

(237) Guchhait, B.; Liu, Y.; Siebert, T.; Elsaesser, T. Ultrafast Vibrational Dynamics of the DNA Backbone at Different Hydration Levels Mapped by Two-Dimensional Infrared Spectroscopy, *Structural Dynamics*, **2016**, *3*, 043202.

(238) Yang, M.; Szyk, L.; Elsaesser, T. Decelerated Water Dynamics and Vibrational Couplings of Hydrated DNA Mapped by Two-Dimensional Infrared Spectroscopy, *J. Phys. Chem. B*, **2011**, *115*, 13093.

(239) Sanstead, P. J.; Stevenson, P.; Tokmakoff, A. Sequence-Dependent Mechanism of DNA Oligonucleotide Dehybridization Resolved through Infrared Spectroscopy. *J. Am. Chem. Soc.*, **2016**, *138*, 11792.

(240) Lee, C.; Cho, M. Vibrational Dynamics of DNA. II. Deuterium Exchange Effects and Simulated IR Absorption Spectra, *J. Chem. Phys.*, **2006**, *125*, 114509

(241) Lee, C.; Park, K.-H.; Cho, M. Vibrational Dynamics of DNA. I. Vibrational Basis Modes and Couplings, *J. Chem. Phys.*, **2006**, *125*, 114508.

(242) Lee, C.; Park, K.-H.; Kim, J.-A.; Hahn, S.; Cho, M. Vibrational Dynamics of DNA. III. Molecular Dynamics Simulations of DNA in Water and Theoretical Calculations of the Two-Dimensional Vibrational Spectra, *J. Chem. Phys.*, **2006**, *125*, 114510.

(243) Lee, C.; Cho, M. Vibrational Dynamics of DNA. IV. Vibrational Spectroscopic Characteristics of A-, B-, and Z-Form DNAs, *J. Chem. Phys.*, **2007**, *126*, 145102.

(244) Thorogood, H.; Waters, T. R.; Parker, A. W.; Wharton, C. W.; Connolly, B. A. Resonance Raman Spectroscopy of 4-Thiothymidine and Oligodeoxynucleotides Containing This Base Both Free in Solution and Bound to the Restriction Endonuclease *EcoRV*. *Biochemistry*, **1996**, *35*, 8723-8733.

$$(245) \quad \Delta\Delta H = \left(\frac{\Delta\nu}{\nu}\right) U_{Bond}$$

where  $\Delta\Delta H$  is the change in the hydrogen bond strength,  $\Delta\nu$  is the shift in the frequency of the vibrational mode,  $\nu$  is the original frequency of the vibrational mode and  $U_{Bond}$  is the bond energy of the affected moiety. Note this equation is only valid if  $\Delta\nu$  is small compared to  $U_{Bond}$ .

(246) Tinoco, I.; Sauer, K.; Wang, J. C.; Puglisi, J. D.; *Physical Chemistry: Principles and Applications in Biological Sciences*, Prentice-Hall, Englewood Cliffs, NJ, **2002**

---

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

(248) Greetham, G. M.; Burgos, P.; Cao, Q.; Clark, I. P.; Codd, P. S.; Farrow, R. C.; George, M. W.; Kogimtzis, M.; Matousek, P.; Parker, A. W.; *et al.* M. ULTRA: A Unique Instrument for Time-Resolved Spectroscopy. *Applied Spectroscopy*, **2010**, *12*, 1311.

# 2D-IR Spectroscopy Reveals Optimised DNA Hoechst 33258 Binding Follows an Induced Fit Model

Chapter 5

This Chapter is based on:

Ramakers, L. A. I.; Hithell, G.; May, J. J.; Greetham, G. M.; Donaldson, P. M.; Towrie, M.; Parker, A. W.; Burley, G. A.; Hunt, N. T.; 2D-IR Spectroscopy Shows that Optimized DNA Minor Groove Binding of Hoechst33258 Follows an Induced Fit Model, *J. Phys. Chem. B.*, **2017**, *121*, 1295-1303.

## 5.1 Abstract

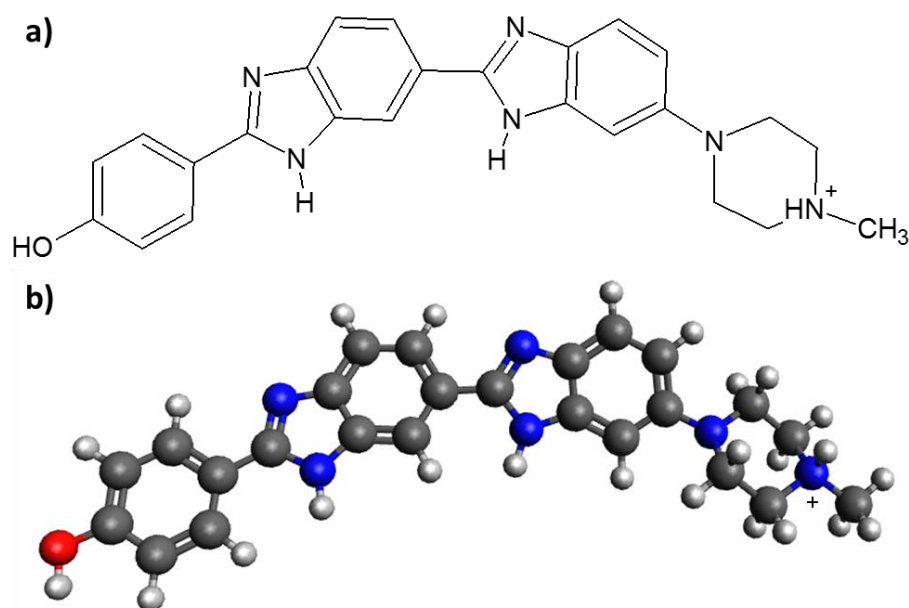
*The induced fit binding model describes a conformational change occurring when a small molecule binds to its biomacromolecular target. The result is enhanced non-covalent interactions between ligand and biomolecule. Induced fit is well-established for small molecule-protein interactions, but its relevance to small molecule-DNA binding is less clear. We investigate the molecular determinants of Hoechst33258 binding to its preferred A-tract sequence relative to a sub-optimal alternating A-T sequence. Results from 2-dimensional infrared spectroscopy, which is sensitive to H-bonding and molecular structure changes, show that Hoechst33258 binding results in loss of minor groove spine of hydration in both sequences, but an additional perturbation of the base propeller twists occurs in the A-tract binding region. This induced fit maximizes favourable ligand-DNA enthalpic contributions in the optimal binding case and demonstrates that controlling the molecular details that induce subtle changes in DNA structure may hold the key to designing next-generation DNA-binding molecules.*



## 5.2 Introduction

Deoxyribonucleic acid (DNA) is the fundamental repository of genetic information used in the majority of organisms in nature.<sup>249</sup> A structural hallmark of DNA is the anti-parallel helix where the major and minor grooves provide sites for sequence-selective binding of proteins and small molecules. Molecular recognition of double-stranded DNA (dsDNA) sequences by transcription factors for example is essential for the initiation of transcription. Furthermore, DNA-binding small molecules such as minor groove binders (MGBs) can perturb various processes associated with gene expression,<sup>250</sup> making them excellent candidates for the design of sequence-selective probes of DNA function in cells and potentially novel therapeutics.<sup>251,252,253</sup> A comprehensive set of guiding principles for the rational design of MGBs to target DNA sequences in a highly sequence-selective manner has yet to emerge however, as a consequence of the complex combination of competing enthalpic and entropic contributions from H-bonding, van der Waals forces and changes in hydration of the DNA and ligand.

An archetypal exemplar of this problem is the *bis*-benzimidazole family of ligands, including Hoechst33258 (Fig.5.1), which show preferential binding to AT-rich dsDNA. Spectroscopic and DNA-footprinting studies reveal that Hoechst33258 binds to dsDNA in a 1:1 stoichiometry and exhibits a 22-fold binding preference for A-tract dsDNA (*e.g.*, 5'-A<sub>3</sub>T<sub>3</sub>) over an alternating sequence (*e.g.*, 5'-ATATAT).<sup>254,255,256,257</sup> The current reasoning for this sequence preference is that the A-tracts possess a narrower minor groove with a well-defined spine of hydration. The release of this water upon ligand binding is linked to the observed entropic driving force for Hoechst33258 binding.<sup>258,259</sup> In contrast, crystallography suggests that the hydration of the AT-minor groove is not as strongly ordered.<sup>260,261,262</sup>



**Figure 5.1:** a) The chemical structure and b) optimized Molecular Geometry<sup>154</sup> of Hoechst 33258

Overall, H33258 binding has been found to be endothermic, but favourable enthalpic contributions arise from hydrophobic interactions between the minor groove wall and the ligand. Direct H-bonding interactions between the ligand and dsDNA are not believed to be an important stabilising factor.<sup>263,264,265,266,267,268,269,270,271</sup> This balance in favour of entropic contributions has led to widespread claims that complex formation follows a ‘rigid body’ model with no significant structural change to the dsDNA upon ligand binding. This however neglects changes in base orientation upon binding revealed by crystallography<sup>263-267</sup> and NMR<sup>268-271</sup> studies. Indeed, NMR has revealed that the minor groove of A-tract dsDNA narrows upon ligand binding, rather than being inherently narrow, prompting suggestion of a significant contribution from an induced fit mechanism,<sup>272</sup> but unfortunately a direct comparison of optimal and sub-optimal binding was not performed.

The induced fit process, in which the biomolecule changes structure to better accommodate the ligand, is often found in binding to proteins (*e.g.* as seen in heat shock protein (Hsp)90 inhibitors),<sup>273</sup> but rarely invoked in DNA binding. Further study is thus required to develop a clear picture of the molecular determinants controlling the sequence selectivity of ligand binding to the minor groove of target dsDNA sequences. We therefore employ two-dimensional infrared (2D-IR) spectroscopy<sup>274,275,276</sup> to probe the binding of Hoechst33258 (H33258)<sup>277</sup> to a preferred A-tract sequence, d(GGAAATTTGC)<sub>2</sub>, (A<sub>3</sub>T<sub>3</sub>), and a sub-optimal binding sequence d(GGATATATGC)<sub>2</sub>, (AT)<sub>3</sub>.

2D-IR offers the advantage of being able to probe structure, intermolecular interactions and hydration by directly measuring the coupling and dynamics of vibrational modes.<sup>274-276</sup> It has extended our understanding of the nature of the vibrational modes of DNA bases<sup>278</sup>, Watson-Crick (W-C) base pairing<sup>279,280</sup> and base stacking,<sup>281</sup> as well as revealing energy transfer mechanisms between bases and backbone<sup>282,283</sup> and the interactions of water with DNA.<sup>284</sup> Most recently, temperature-jump 2D-IR reported sequence-dependent pre-melting dynamics.<sup>285</sup>

Here, we show that the distinct difference between H33258 binding to A-tract and alternating sequence DNA is the loss of the ordered propeller twist arrangement of the base pairs present in the uncomplexed A-tract. This results in a subtle but distinct conformational change within the minor groove to accommodate the ligand, maximising favourable enthalpic contributions to binding. The uniquely ordered A-tract sequence also governs the formation of the spine of hydration and so its removal contributes to the entropic gain from releasing hydration water. Although the results for binding to the alternating sequence are similar, the emphasis is shifted toward loss of the spine of hydration, suggesting that the induced fit effect is crucial to the observed difference in binding affinities.

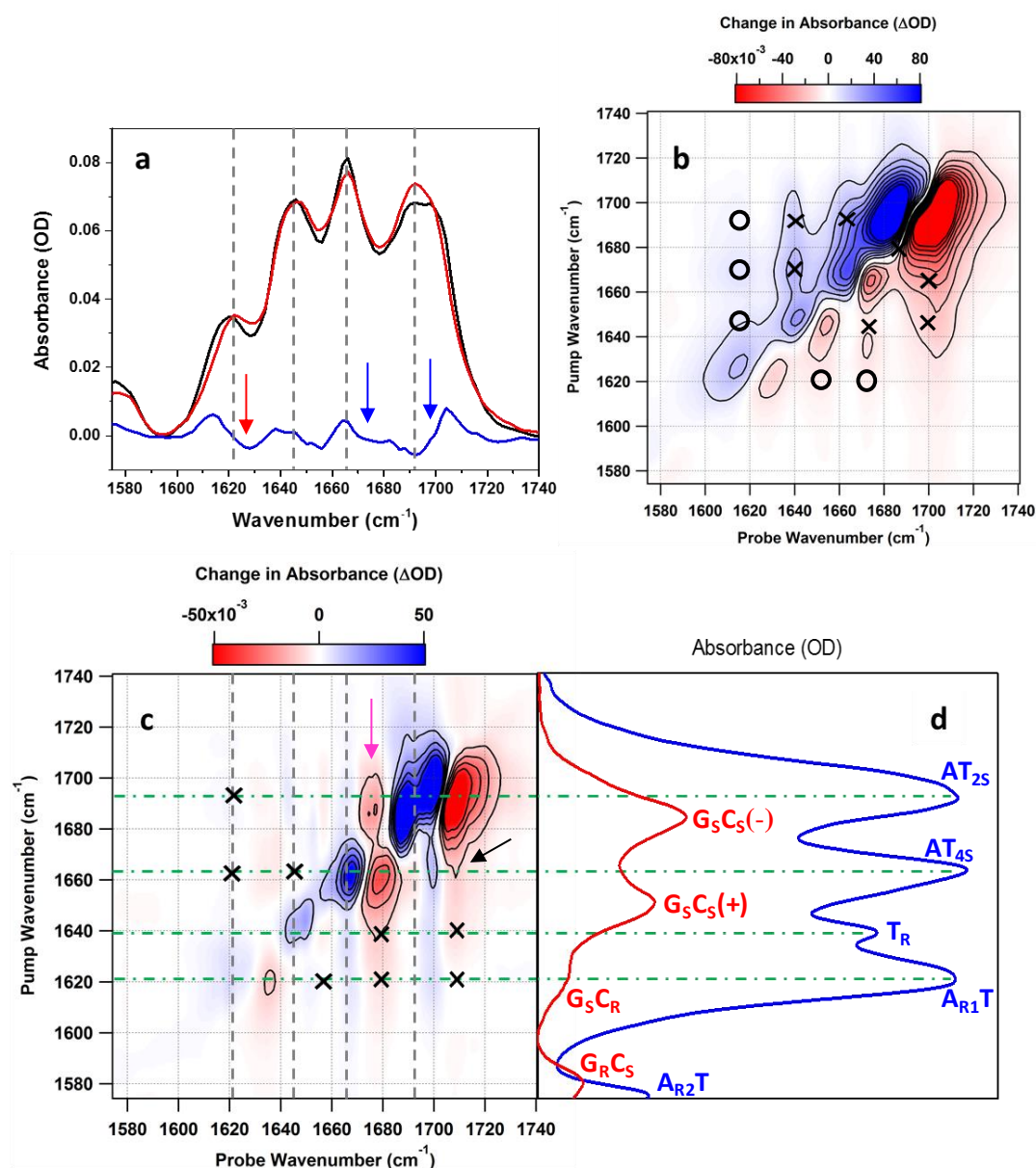
## 5.3 Results and Discussion

### 5.3.1 Difference Spectroscopy

The FTIR spectra of the  $A_3T_3$  DNA sequence and its complex with H33258 ( $H-A_3T_3$ ) are shown in Fig.5.2(a). The corresponding spectra of  $(AT)_3$  and  $H-(AT)_3$  are shown in Fig.5.3(a). The IR spectra of both DNA sequences show four peaks (Fig.5.2(a), Fig.5.3(a), red), these are indicated by grey dashed lines in the figures and listed in Table 5.1. The results are consistent with previous studies and peaks are assigned by reference to spectra of DNA duplexes containing exclusively GC or AT base pairs, shown in Figs.5.2(d) and 5.3(d).<sup>278,281,282,286</sup> The assignments are summarized in Table 5.1. The peaks in the spectra of the two DNA sequences are dominated by AT modes, though contributions from GC base pairs, which are less intense, do influence the magnitude of the absorption between the peaks. This spectral congestion can be unraveled by 2D-IR methods, which report vibrational coupling patterns in the off-diagonal region of the spectrum.

The changes to the FTIR spectra upon formation of the  $H-A_3T_3$  and  $H-(AT)_3$  complexes (Fig.5.2(a), Fig.5.3(a), black) are displayed via difference spectra (Fig.5.2(a), Fig.5.3(a), blue).

Since H33258 exhibits no vibrational modes in this spectral region (Fig.A5.2), these are assigned to modifications of the DNA upon binding.



**Figure 5.2:** a) FTIR spectra of A<sub>3</sub>T<sub>3</sub> DNA with (black) and without (red) H33258. Blue spectrum shows the binding-induced difference FTIR spectrum. b) 2D-IR spectrum of H-A<sub>3</sub>T<sub>3</sub> complex. Crosses mark off-diagonal peaks assigned to coupling of modes primarily located on T-base. Circles show off-diagonal peaks assigned to coupling of modes on A and T bases induced by W-C base pairing. c) Ligand binding induced 2D-IR difference spectrum. Crosses show locations of small off-diagonal features. d) FTIR spectra of GC and AT-only DNA sequences to show the relative (weighted per base) magnitudes of peaks and peak positions to aid assignment of the IR spectrum of the A<sub>3</sub>T<sub>3</sub> sequence.

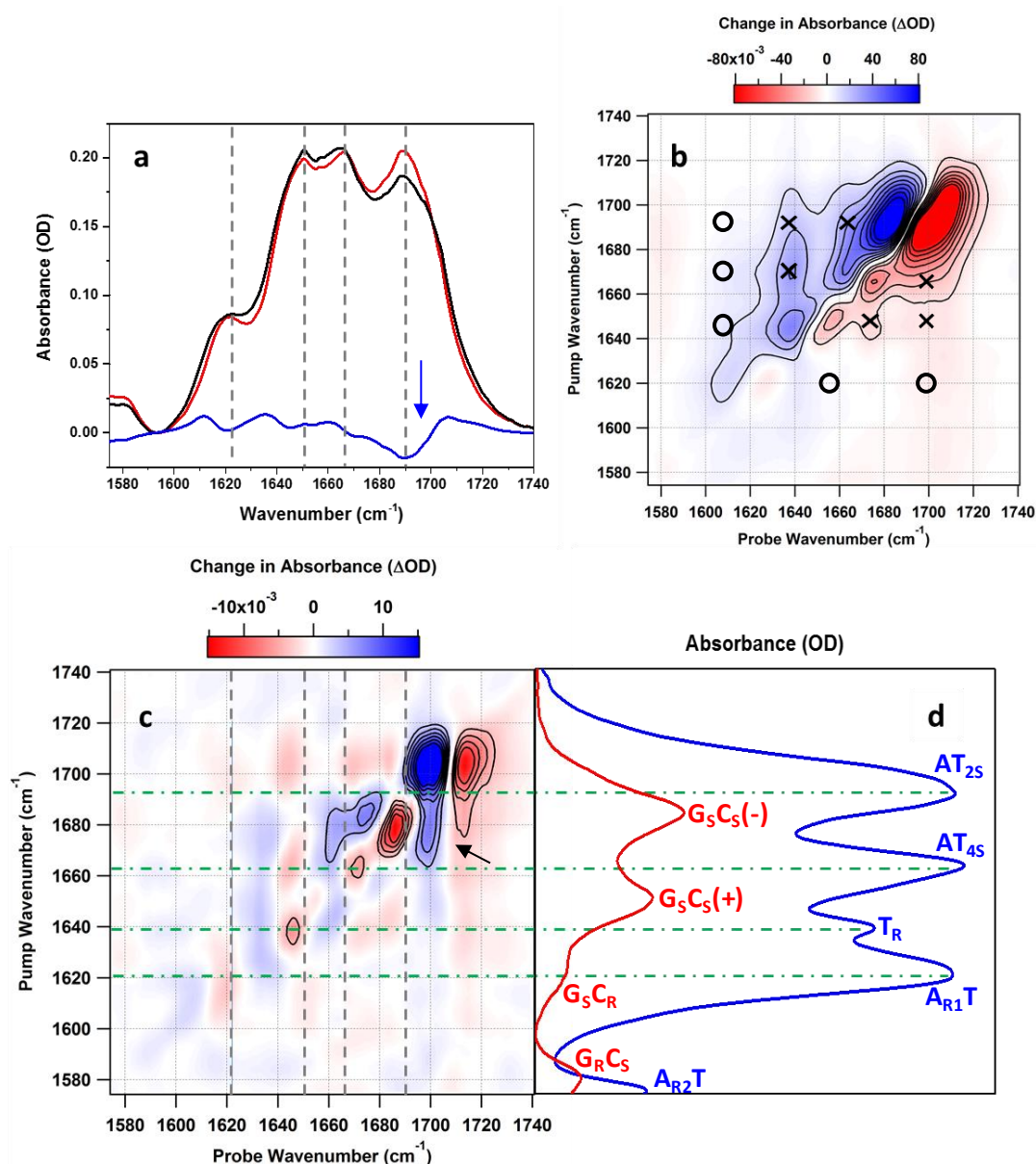
**Table 5.1:** Assignment of peaks in the IR spectra of A<sub>3</sub>T<sub>3</sub> and (AT)<sub>3</sub> DNA duplexes, showing comparison with AT and GC-only sequences.

Assignment	Position (cm <sup>-1</sup> )				Description
	A <sub>3</sub> T <sub>3</sub>	(AT) <sub>3</sub>	AT <sup>a</sup>	GC <sup>b</sup>	
AT <sub>2s</sub>	1692	1689	1692		Base-paired T <sub>2</sub> C=O stretch
G <sub>s</sub> C <sub>s</sub> (-)				1684	GC C=O antisymmetric stretch
AT <sub>4s</sub>	1666	1666	1664		Base-paired T <sub>4</sub> C=O stretch
G <sub>s</sub> C <sub>s</sub> (+)				1651	GC C=O symmetric stretch
T <sub>R</sub>	1646	1650	1640		T ring vibration
A <sub>R1</sub> T	1622	1622	1622 (s)		Coupled AT ring vibration/ C ring mode + G C=O
G <sub>s</sub> C <sub>R</sub>				1622 (w)	

<sup>a</sup>obtained from the sequence: 5'-ATTATTATTATATTA-3' (Fig A5.1(a)); <sup>b</sup> obtained from the sequence: 5'-GCCGCCGCCG-3' (Fig A5.1(b)); modes marked (s) are those which contribute strongly to the overall spectrum and modes marked (w) are those which contribute weakly to the overall spectrum.

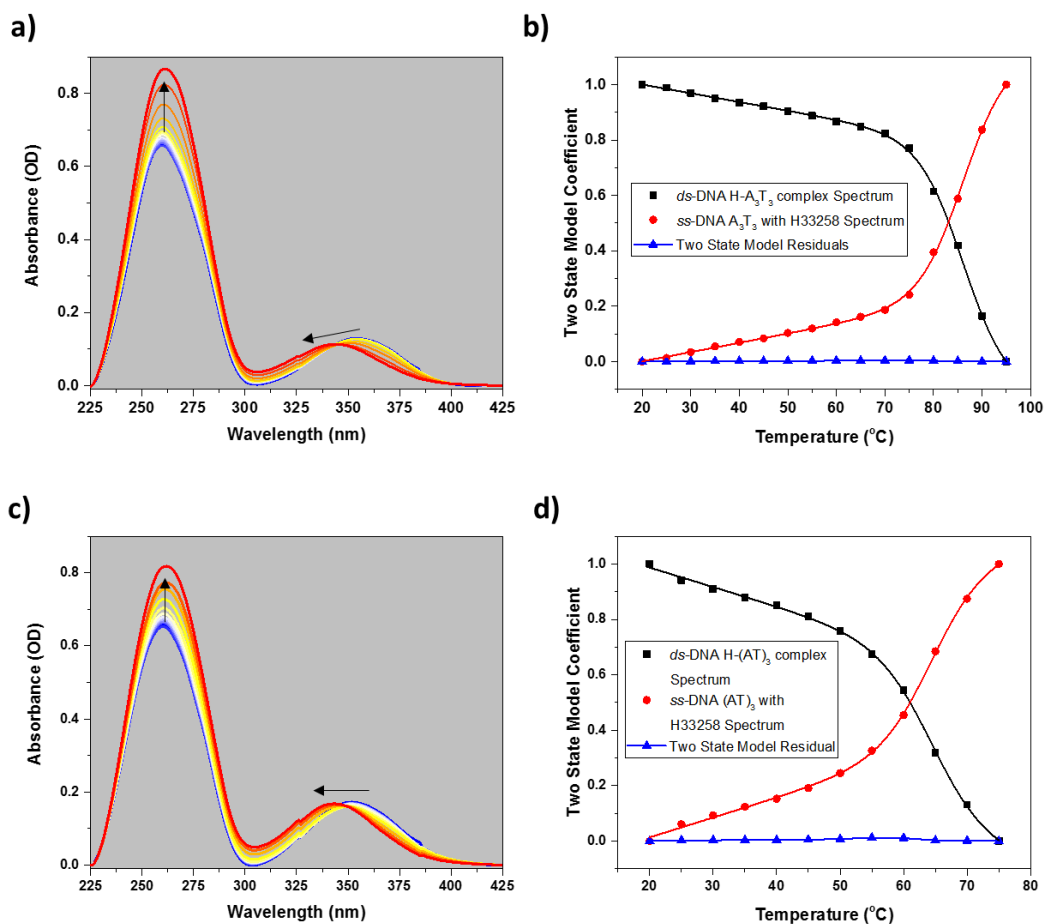
The H-A<sub>3</sub>T<sub>3</sub> sequence shows three main features in the difference IR spectrum (Fig.5.2(a)). Two of these, near 1665 cm<sup>-1</sup> and 1700 cm<sup>-1</sup> (blue arrows), consist of a negative peak located to the lower frequency side of a positive peak, suggesting that the AT<sub>4s</sub> and AT<sub>2s</sub> modes shift to higher wavenumber. The third feature (red arrow) has reversed positive/negative contributions, consistent with a shift of the A<sub>R1</sub>T mode to lower frequency.

The features in the FTIR difference spectrum obtained for the H-(AT)<sub>3</sub> complex (Fig.5.3(a)) are smaller and less well-defined than those for the complex with the A-tract sequence (Fig.5.2(a)). The shift of the AT<sub>2s</sub> peak at 1690 cm<sup>-1</sup> to higher frequency (blue arrow), as observed for the A<sub>3</sub>T<sub>3</sub> sequence, is still detected, but the other features are less apparent.



**Figure 5.3:** a) FTIR spectra of  $(AT)_3$  DNA with (black) and without (red) H33258. Blue spectrum shows the binding-induced difference FTIR spectrum. b) 2D-IR spectrum of H- $(AT)_3$  complex. Crosses mark off-diagonal peaks assigned to coupling of modes primarily located on T-base. Circles show off-diagonal peaks assigned to coupling of modes on A and T bases induced by W-C base pairing. c) Ligand binding induced 2D-IR difference spectrum. d) FTIR spectra of GC and AT-only DNA sequences to show the relative (weighted per base) magnitudes of peaks and peak positions to aid assignment of the IR spectrum of the  $(AT)_3$  sequence.

Due to the modest impact of the binding on the on the IR spectra, these H33258:DNA complexes were further studied using UV-visible spectroscopy. The UV-visible spectra of the H- $A_3T_3$  and H- $(AT)_3$  complexes as the temperature of the sample is increased are shown in Figs. 5.4.(a)&(c), respectively.



**Figure 5.4:** a) UV-visible spectra of the H- $A_3T_3$  complex at 5 °C intervals between 20–95 °C, b) melting curves of the H- $A_3T_3$  complex extracted by the application of the two state model, c) UV-visible spectra of the H- $(AT)_3$  complex at 5 °C intervals between 20–75 °C and d) melting curves of the H- $(AT)_3$  complex extracted by the application of the two state model. The residuals of the two state model are shown in blue in c) and d).

As the temperature increases the UV-visible spectra of these complexes change. These changes are found to consist of a hyperchromic shift in the 260 nm peak and a shift to higher frequency of the 355 nm peak (indicated by arrows, Figs.5.4.(a)&(c)). The application of the two state model (described previously in chapter 4) to these UV-visible spectra yields melting curves for each of the complexes (Fig. 5.4.(b)&(d)). Fitting these melting curves, using a sigmoidal function, yields melting temperatures of 84 °C and 67 °C for the H- $A_3T_3$  and H- $(AT)_3$  complexes, respectively. Comparing these melting temperature to those obtained for the unbound sequences (60 °C and 51 °C for the  $A_3T_3$  and  $(AT)_3$  sequences respectively, chapter 4), shows ligand-induced stabilization of the melting temperatures of the DNA duplexes by 24 °C and 16 °C for the  $A_3T_3$  and  $(AT)_3$  sequences, respectively, confirming H33258 binding. This stabilization of the melting temperatures upon binding was also verified by fluorescence



emission spectroscopy carried out on the H-A<sub>3</sub>T<sub>3</sub> and H-(AT)<sub>3</sub> complexes (Fig.A5.3). The reduced stabilization for the alternating sequence is consistent with the reported 22-fold preference of H33258 for binding to A-tracts.<sup>256</sup>

The 2D-IR spectrum of the H-A<sub>3</sub>T<sub>3</sub> complex is shown in Fig.5.2(b). Each peak observed in the FTIR spectrum gives rise to a negative feature located on the 2D-IR spectrum diagonal. These are assigned to the respective  $\nu = 0 \rightarrow 1$  transitions, each with an accompanying, positive,  $\nu = 1 \rightarrow 2$  peak shifted by  $\sim 10 \text{ cm}^{-1}$  to lower probe frequency.<sup>282,285,286,287,288,289</sup> Peaks located in the off-diagonal region primarily indicate the presence of coupling between diagonal vibrational modes, though a small contribution is expected from energy transfer due to the fast ( $\sim 650 \text{ fs}$ ) vibrational relaxation of the base vibrational modes.<sup>282</sup> Two groups of off-diagonal peaks are indicated in Fig.5.2(b). Those due to intra-base coupling between the AT<sub>25</sub>, AT<sub>45</sub> and T<sub>R</sub> modes that are primarily located on the thymine base are marked with crosses. Other off-diagonal peaks (Fig 5.2(b), open circles) indicate inter-base coupling between the adenine-based A<sub>R1</sub>T mode and the three thymine modes that is induced by Watson-Crick H-bonding.

A difference 2D-IR spectrum shows the impact of binding H33258 (Fig.5.2(c)). A comparison of the peak positions in the 2D-IR difference spectrum with the linear AT and GC spectra (Fig.5.2(d)) suggests that H33258 binding primarily affects the AT<sub>25</sub> and AT<sub>45</sub> vibrational modes of the A<sub>3</sub>T<sub>3</sub> duplex. Other, smaller, features appear in the 2D-IR difference spectrum that originate from changes in the A<sub>R1</sub>T modes, consistent with the FTIR spectrum, but also the T<sub>R</sub> mode and related off-diagonal peaks (crosses Fig.5.2(c)).

The 2D-IR spectrum of the H-(AT)<sub>3</sub> complex and the binding-induced 2D-IR difference spectrum for the (AT)<sub>3</sub> sequence are shown in Figs.5.3(b) and (c), respectively. The major changes upon binding affect the AT<sub>25</sub> and AT<sub>45</sub> modes and the overall pattern of peaks in the 2D-IR difference spectrum is similar to that for the A-tract sequence, although the AT<sub>25</sub> feature is clearly dominant.

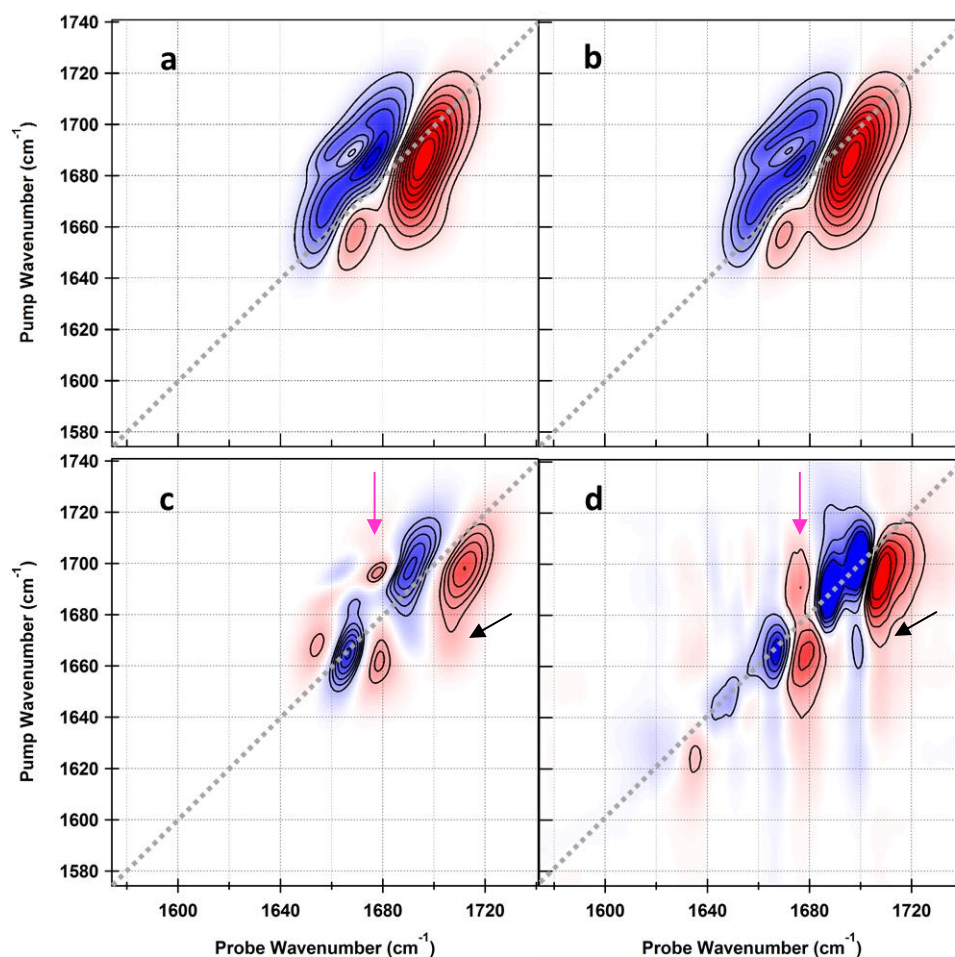


**Table 5.2:** Changes to the IR response of A<sub>3</sub>T<sub>3</sub> and (AT)<sub>3</sub> DNA duplexes due to H33258 binding.

Changes due to H33258 Binding	Host DNA Duplex	
	A <sub>3</sub> T <sub>3</sub>	(AT) <sub>3</sub>
AT <sub>2s</sub> Shift (cm <sup>-1</sup> )	6.4	15.1
AT <sub>2s</sub> Shifted Subset Size (%)	<b>20.3</b>	<b>11.3</b>
Change in T <sub>2</sub> H-bond Strength (kJmol <sup>-1</sup> ) <sup>a</sup>	-2.8	-6.6
AT <sub>4s</sub> Shift (cm <sup>-1</sup> )	8.7	15.6
AT <sub>4s</sub> Shifted Subset Size (%)	<b>9.6</b>	<b>1.7</b>
Change in T <sub>4</sub> H-bond Strength (kJmol <sup>-1</sup> ) <sup>a</sup>	-3.9	-6.9

<sup>a</sup>Values calculated using the C=O bond energy (~743 kJmol<sup>-1</sup>)<sup>290</sup> via the methodology as outlined in ref. 291 using equation ref. 292.

The overlapping peaks in the IR spectrum of the DNA bases mean that the 2D-IR off-diagonal peaks are essential for clear differentiation between changes affecting the AT and GC base pairs (e.g. arrows Figs 5.2(c)&5.3(c)). The peaks in the difference FT-IR and 2D-IR spectra focus on the AT<sub>2s</sub> and AT<sub>4s</sub> modes, but to rule out contributions from GC modes and to quantify the changes, we employed a model 2D-IR spectrum (Fig.5.5) constructed from 2D Gaussian lineshapes, which simulate the coupled peaks in the 2D-IR spectra of the DNA duplexes (Fig.5.4(a) and 5.4(b)). A variety of spectral changes were modelled and the difference spectra simulated (Fig.A5.4).



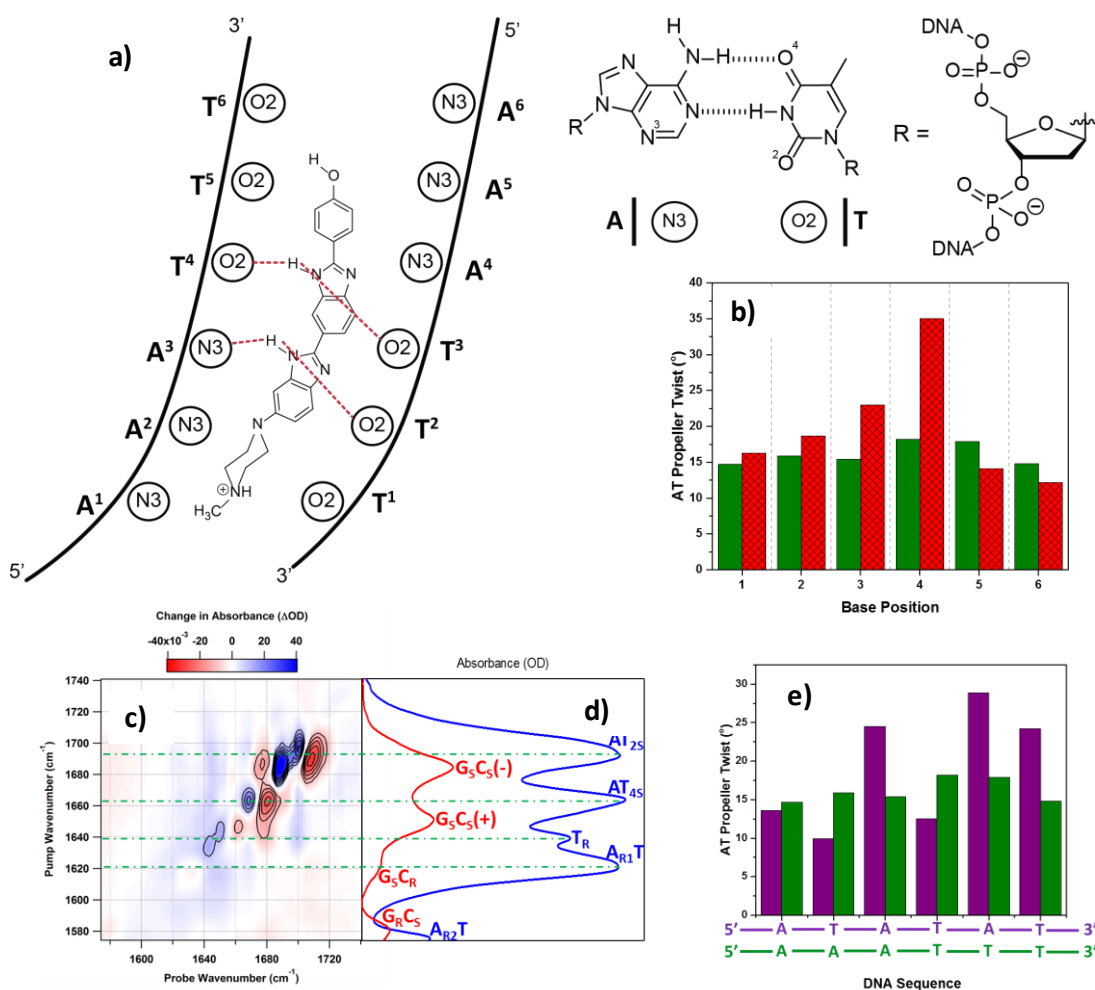
**Figure 5.5:** Simulated 2D-IR spectra of the coupled  $AT_{25}$  and  $AT_{45}$  transitions before (a) and after (b) a shift of a subset of the bands to higher frequency. (c) shows the simulated difference 2D-IR spectrum that results from a subtraction of (a) from (b). (d) Shows the difference 2D-IR spectrum obtained following formation of the  $H-A_3T_3$  complex. Pink and black arrows point to off-diagonal features linking the two modes in experimental and simulated spectra.

The best agreement with the experimental 2D-IR difference spectra (Figs.5.2(c)&5.3(c)) was obtained by shifting a subset of the  $AT_{25}$  and  $AT_{45}$  peaks to higher frequency (Fig.5.5(c)). Crucially, this simulation recreated the off-diagonal structure of the difference spectrum as well as the diagonal peaks (see arrows Fig.5.5(c&d)). The results are summarized in Table 5.2, which show that  $\sim 20\%$  of the  $AT_{25}$  mode of the  $A_3T_3$  (A-tract) sequence is shifted  $\sim 6\text{ cm}^{-1}$  to higher frequency while  $\sim 10\%$  of the  $AT_{45}$  mode shifts by  $\sim 9\text{ cm}^{-1}$ . For the alternating sequence, the shift to higher frequency is  $\sim 7\text{-}9\text{ cm}^{-1}$  larger for both modes, but the fraction of the carbonyls affected is smaller. Particularly noteworthy is the very small (2%) fraction of the  $AT_{45}$  peak that is shifted whereas the 11% change in the  $AT_{25}$  mode is more comparable with the  $A_3T_3$  sequence.

Based on these models, we assign the peaks appearing in the FT-IR and 2D-IR difference spectra to a binding-induced shift to higher frequency of a portion of the AT<sub>25</sub> and AT<sub>45</sub> base vibrations. No contributions from diagonal or off-diagonal peaks attributable to GC were observed, showing that the binding interactions of H33258 are localized to the AT region of the duplex.

Shifts of the AT<sub>25</sub> and AT<sub>45</sub> modes of the A-tract sequence to higher frequency are consistent with a decrease in the strength of H-bonds formed to the T<sub>2</sub> and T<sub>4</sub> carbonyl groups upon H33258 binding of 2.8 and 3.9 kJmol<sup>-1</sup>, respectively.<sup>290-292</sup> It is important to note that, although the two modes exhibit coupling, it has been demonstrated that this is weak and that a local mode picture, treating the two carbonyls as separate modes is appropriate.<sup>278</sup> The smaller changes in the A<sub>R1</sub>T and T<sub>R</sub> vibrational modes are likely to be a result of changes to the two T carbonyls since their stretches are known to contribute to these modes.<sup>286</sup>

The T<sub>2</sub> carbonyls point directly into the minor groove of the DNA duplex and do not engage in W-C H-bonding. Water molecules are present in the minor groove as a spine of hydration and form H-bonds to the T<sub>2</sub> carbonyl.<sup>258,259</sup> The reduction in strength of these H-bonds to some of the T bases in the sequence is consistent with structural studies showing that the bis-benzimidazole of H33258 displaces water molecules from the minor groove.<sup>258,259</sup> Although the ligand has been shown to replace the H-bonds to the T<sub>2</sub> carbonyls, our results show that the overall H-bond strength is lower than that formed by the water, consistent with evidence that the interaction between H33258 and DNA is not stabilized to any great degree by H-bonding.<sup>259</sup>



**Figure 5.6:** a) Diagram illustrating the binding of H-A<sub>3</sub>T<sub>3</sub> from X-ray crystallography.<sup>266</sup> Pink lines show H-bonds between ligand and DNA. b) AT propeller twists of the free (green bars, PDB ID: 1S2R<sup>260</sup>) and the complexed A<sub>3</sub>T<sub>3</sub> (red bars, PDB ID: 296D<sup>266</sup>). c) 2D-IR difference spectrum between (AT)<sub>3</sub> and A<sub>3</sub>T<sub>3</sub> duplex d) FTIR spectra of AT only (blue) and GC only (red) DNA sequences e) AT propeller twists of A<sub>3</sub>T<sub>3</sub> (green, PDB ID: 1S2R<sup>260</sup>) and (AT)<sub>3</sub> (purple, PDB ID: 1DN9<sup>261</sup>) base sequences.

The frequency shift of the T<sub>4</sub> carbonyl, which is involved in the W-C interaction, shows that binding is also accompanied by a structural change affecting the DNA bases. The two major structural changes that can weaken the W-C hydrogen bond are a buckle and propeller twist.<sup>286</sup> An increase in the base buckle angle will reduce the strength of the W-C H-bonds, but a computational study predicted that the T<sub>4</sub> carbonyl is not strongly affected by this until an extreme angle of buckle (~50°) is reached.<sup>286</sup> This is not consistent with the relatively minor structural changes detected by NMR<sup>268-271</sup> or crystallography<sup>263-267</sup> on H33258 binding to the A<sub>3</sub>T<sub>3</sub> duplex.<sup>263</sup> An increase in the propeller twist will cause an increase in the length of the W-C hydrogen bond to the T<sub>4</sub> carbonyl, decreasing its strength. This is consistent both

with the spectroscopic data and previous structural studies using 2D-NMR<sup>268-271</sup> and X-ray crystallography,<sup>263-267</sup> which have reported changes in propeller twist. Crystallographic results of the structure of the H-A<sub>3</sub>T<sub>3</sub> complex are reproduced in (Fig.5.6(a)) alongside a bar-graph showing perturbations in the propeller twist in the H-A<sub>3</sub>T<sub>3</sub> complex (Fig.5.6(b), red bars) relative to the uncomplexed DNA (Fig.5.6(b), green). H33258 forms two bifurcated H-bonds to T<sub>2</sub> carbonyls (Fig.5.6(a), pink lines) and accommodating these interactions increases the propeller twist in some of the AT base pairs with one particularly strongly affected. This is consistent with our observations of a subset of T<sub>4</sub>-related modes shifting to higher frequency upon ligand binding. The propeller twisting does not affect the GC base pairs, which is also entirely consistent with our spectroscopic data.

Comparing the A-tract sequence with the alternating sequence, it is clear that the disruption to the spine of hydration, affecting the T<sub>2</sub> carbonyl is still present in the data. The smaller shifted subset size and the larger frequency shift observed for the T<sub>2</sub> carbonyl in the (AT)<sub>3</sub> sequence supports the conclusion of an X-ray crystallography<sup>267</sup> study of Hoechst33258 binding to a (AT)<sub>2</sub> sequence, which indicated the presence of only one bifurcated H-bond as opposed to two in the complex with A<sub>3</sub>T<sub>3</sub>. However, the impact upon the propeller twist, while present, is markedly reduced. This is shown by the relatively small size (1.7%) of the shifted subset in the (AT)<sub>3</sub> duplex showing that this does not undergo significant structural changes upon binding compared to the A<sub>3</sub>T<sub>3</sub> (9.7%) sequence. This is a major difference between the two strands and may be relevant to the differences in binding affinity. It is clear from thermodynamic measurements that H33258 binding is entropically-driven; this is attributed to the release of water from the spine of hydration. Our data reflects this and the minor groove of the A-tract sequence is known to contain a more ordered spine of hydration than the alternating sequence.<sup>260-262</sup> Many reports assume that conformational changes in the DNA accompanying this binding are negligible, citing a 'lock and key' or 'rigid body' interaction.<sup>258,259,293,294,295</sup> The enthalpy of binding has been measured to be endothermic overall, but the greatest favourable contribution to binding has been assigned to hydrophobic contacts between the DNA groove and the ligand.<sup>259,295</sup> A-tract DNA is unique in that it features a very regular set of propeller twists (Fig 5.6) due to the formation of 3-centre H-bonds, leading to a rigid and ordered section of the helix.<sup>290</sup> The indications from our data are that, rather than this unique structural feature being beneficial to ligand binding in itself, the disruption of the A-tract structure leads to a closer interaction between ligand and DNA than is possible in the alternating sequence. While the change in macroscopic conformation

of the DNA is small and unlikely to outweigh the entropic benefits of releasing hydration water, our results indicate that binding to A-tract DNA is better described by an induced fit type model, where the DNA structure changes to accommodate the ligand. By contrast, binding to a sub-optimal sequence ((AT)<sub>3</sub>) proceeds without significant change to the DNA structure, which lacks the enthalpic benefit of the improved DNA-ligand interactions and also does not gain from any entropic benefit from disrupting the ordered propeller twists. A structure-related narrowing of the minor groove and restriction of dynamical motion upon ligand binding to A-tract DNA has been suggested in simulations but was stated to be insignificant.<sup>290,294</sup> In contrast, our results support conclusions drawn from NMR studies, proposing that induced fit is an important aspect of optimized Hoechst binding.<sup>272</sup>

Further evidence that the difference 2D-IR spectrum observed on the formation of the H-A<sub>3</sub>T<sub>3</sub> complex results from an order-disorder change in the base orientation within the AT region of the duplex arises from a difference 2D-IR spectrum between the uncomplexed A<sub>3</sub>T<sub>3</sub> and (AT)<sub>3</sub> sequences (Fig.5.6.(c)). These structures differ mainly through the three-centered H-bonds and ordered propeller twists of the A<sub>3</sub>T<sub>3</sub> sequence (Fig.5.6(e), green) in contrast to the alternating sequence (Fig.5.6.(e), purple).<sup>261</sup> The difference 2D-IR spectrum (Fig.5.6.(c)) is remarkably similar to that arising from formation of the H-A<sub>3</sub>T<sub>3</sub> complex; the AT<sub>25</sub> and AT<sub>45</sub> modes shift by 6.3 cm<sup>-1</sup> and 10.7 cm<sup>-1</sup>. This observation supports the hypothesis that the ligand binding induces changes in the propeller twists of the A<sub>3</sub>T<sub>3</sub> sequence. It is also noticeable that the T<sub>2</sub> mode shifts between the two uncomplexed sequences. This suggests a weakening of the hydrogen bonds to the T<sub>2</sub> in the alternating sequence versus the A-tract, consistent with the more ordered spine of hydration (and so stronger H-bonds) found in the minor groove of the latter sequence.<sup>260,262</sup> This further indicates that an entropic gain may arise from losing the spine of hydration as well as the ordered propeller twist caused by induced fit of the ligand.

## 5.4 Conclusion

In conclusion, 2D-IR shows that binding H33258 to two DNA sequences leads to shifts in vibrational modes associated with specific AT base pairs that are due to the loss of the spine of hydration and formation of direct H-bonding between DNA and ligand as well as alterations in the propeller twist induced by the ligand locating in the minor groove. Comparison of binding to A-tract and alternating DNA sequences revealed that binding to A<sub>3</sub>T<sub>3</sub> results in loss of the ordered propeller twist arrangement of bases found in the

uncomplexed DNA. This is not replicated to the same extent in the alternating sequence and we propose that these structural changes constitute an induced fit type interaction that facilitates superior accommodation of H33258 and increased hydrophobic interactions between ligand and DNA. This contradicts current pictures which treat H33258 binding as a rigid body interaction and complements the entropic release of water from the minor groove. Finally, the results fully demonstrate 2D-IR capabilities to simplify quantification of solution phase DNA-binding.

## 5.5 Methods

### 5.5.1 Materials

The lyophilised, salt-free DNA oligonucleotides were obtained from Eurofins; Hoechst33258, D<sub>2</sub>O, DMSO, monobasic and dibasic sodium phosphate were obtained from Sigma-Aldrich. All chemicals were used without further purification. All the samples used in the experiments were prepared in pH 7 phosphate buffer solution. All of the DNA oligonucleotides were dissolved to produce 10 mM stocks and a 100 mM stock of Hoechst33258 was prepared using DMSO. Using these stocks all samples were prepared using pH7 phosphate buffer solution to a final duplex:H33258 ratio of 1:1 and annealed at 90 °C for 10 minutes.

### 5.5.2 UV-visible Spectroscopy

The samples were held in a demountable Harrick cell utilising CaF<sub>2</sub> windows and a 50 µm polytetrafluoroethylene spacer for all the IR experiments. FTIR experiments were carried out using a Perkin-Elmer Lambda 25 at a resolution of 1 nm with sample concentrations of 2.5 mM (A<sub>3</sub>T<sub>3</sub> duplex/H-A<sub>3</sub>T<sub>3</sub> complex) or 5 mM ((AT)<sub>3</sub> duplex/H-(AT)<sub>3</sub> complex). These measurements were repeated at 1 µM (samples prepared via serial dilution to ensure accuracy) to provide duplex melting temperature data at this concentration to compare with the fluorescence measurements.

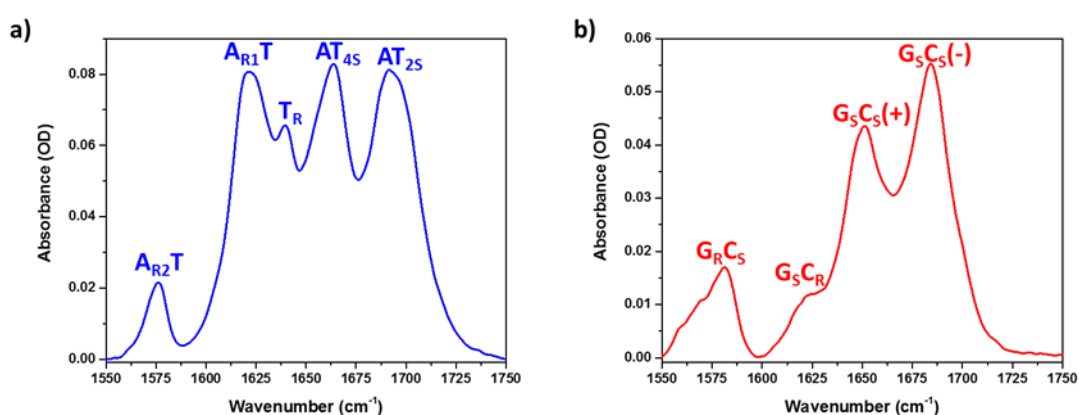
### 5.5.3 Fluorescence Spectroscopy

The 1 µM sample was diluted from the original samples (concentrations of 2.5 mM (A<sub>3</sub>T<sub>3</sub> duplex/H-A<sub>3</sub>T<sub>3</sub> complex) or 5 mM ((AT)<sub>3</sub> duplex/H-(AT)<sub>3</sub> complex)) via serial dilution to ensure accuracy. The samples were held in a quartz cuvette with a path length of 1 cm. The fluorescence experiments were carried out using a Horiba Fluorolog2 at a resolution of 1 nm. The samples were excited at 350 nm and the emitted fluorescence was recorded from 380 nm – 600 nm.

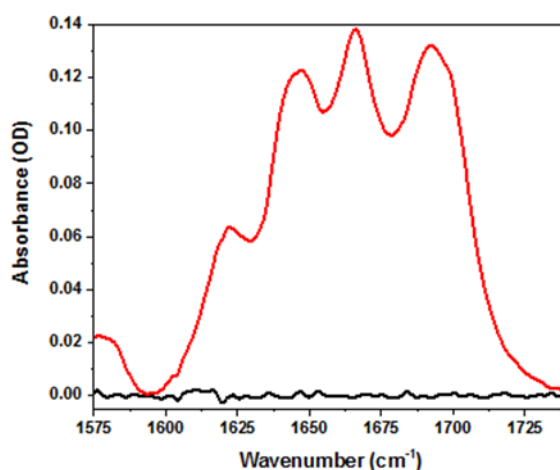
### 5.5.4 2D-IR and FT-IR Spectroscopy

The samples were held in a demountable Harrick cell utilising CaF<sub>2</sub> windows and a 50  $\mu\text{m}$  polytetrafluoroethylene spacer for all the IR experiments. FTIR experiments were carried out using a Bruker Vertex 70 at a resolution of 1  $\text{cm}^{-1}$  with sample concentrations of 2.5 mM (A<sub>3</sub>T<sub>3</sub> duplex/H-A<sub>3</sub>T<sub>3</sub> complex) or 5 mM ((AT)<sub>3</sub> duplex/H-(AT)<sub>3</sub> complex). 2D-IR spectra were collected using the ULTRA FT-2D-IR spectrometer.<sup>296,297</sup> The IR pulses used had a temporal duration of  $\sim 100\text{fs}$ ; a center frequency of 1650  $\text{cm}^{-1}$  and a bandwidth of  $\sim 300\text{cm}^{-1}$ , at a repetition rate of 10 kHz. FT-2D-IR measurements were carried out at concentrations of 1.25 mM (A<sub>3</sub>T<sub>3</sub> duplex/H-A<sub>3</sub>T<sub>3</sub> complex) or 2.5 mM ((AT)<sub>3</sub> duplex/H-(AT)<sub>3</sub> complex).

## 5.6 Appendix

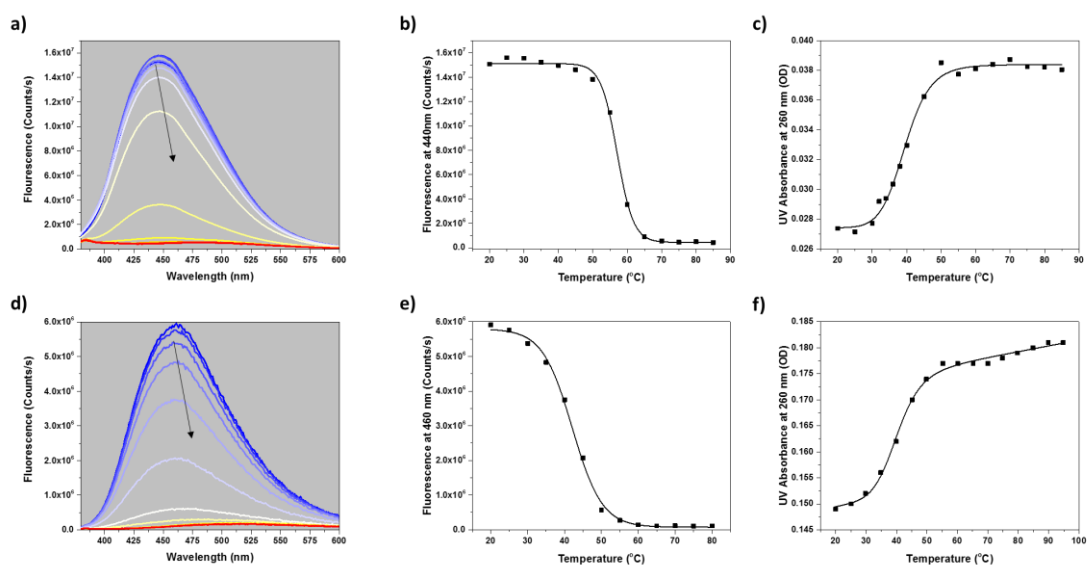


**Figure A5.1:** Assigned FTIR spectra of a) AT only duplex (sequence: 5'-ATTATTATTATATTA-3') and b) GC only duplex (sequence: 5'-GCCGCCGCCG-3'). All data recorded at 20 °C.

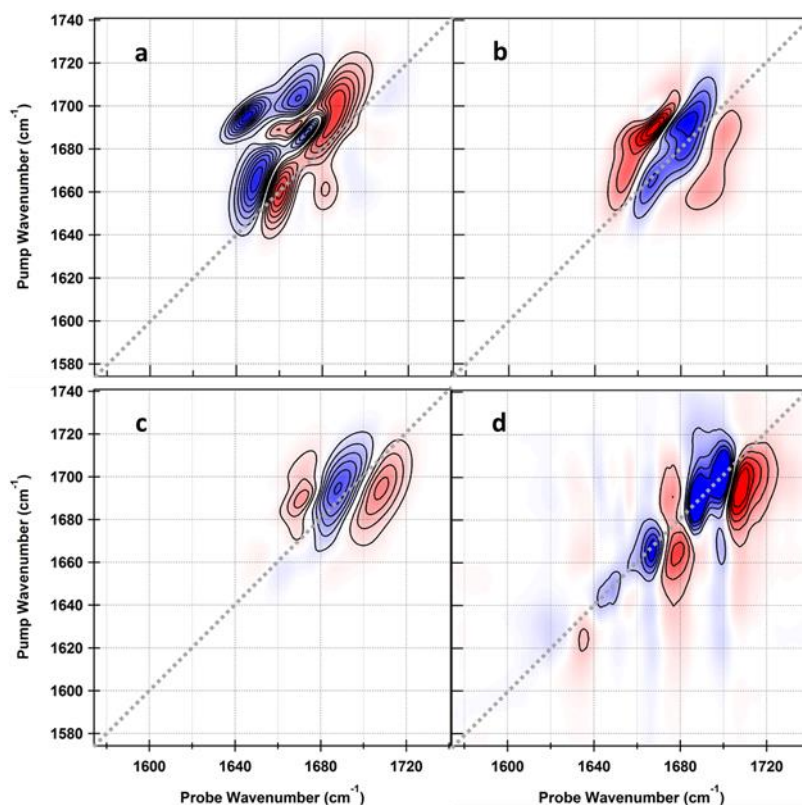


**Figure A5.2:** FTIR spectra of A<sub>3</sub>T<sub>3</sub> duplex (red) only and Hoechst33258 (black) only. All presented data was measured at a concentration of 2.5 mM.





**Figure A5.3:** a) Fluorescence Emission spectra of the H-A<sub>3</sub>T<sub>3</sub> complex at 5 °C intervals between 20– 85 °C, b) thermal profile of the emitted fluorescence of the H-A<sub>3</sub>T<sub>3</sub> complex, c) Melting curve of the uncomplexed A<sub>3</sub>T<sub>3</sub> sequence derived from the hyperchromic shift of the UV-visible DNA absorbance band, d) Fluorescence Emission spectra of the H-(AT)<sub>3</sub> complex at 5 °C intervals between 20– 80 °C, e) thermal profile of the emitted fluorescence of the H-(AT)<sub>3</sub> complex and f) Melting curve of the uncomplexed (AT)<sub>3</sub> sequence derived from the hyperchromic shift of the UV-visible DNA absorbance band.



**Figure A5.4:** a) Model difference 2D-IR spectra due to an increase in mode anharmonicity, b) Model difference 2DIR spectra due to a decrease in mode anharmonicity, c) Model difference 2D-IR spectra due to a global blue shift of the modes and d) Experimentally observed difference 2D-IR due to H33258 binding (included to demonstrate the departure of these models from the experimentally observed data).

## 5.7 References

(249) Watson, J. D.; Crick, F. H. C. Molecular Structure of Nucleic Acids, *Nature*, **1953**, 4356, 737.

(250) Neidle, S. DNA Minor-Groove Recognition by Small Molecules. *Nat. Prod. Rep.* **2001**, 18, 291-309.

(251) White, C. M.; Heidenreich, O.; Nordheim, A.; Beerman, T. A. Evaluation of the Effectiveness of DNA-Binding Drugs to Inhibit Transcription Using the c-fos Serum Response Element as a Target, *Biochemistry*, **2000**, 39, 12262-12273.

(252) Zhang, X.; Kiechle, F. Hoechst33342-Induced Apoptosis is Associated with Decreased Immunoreactive Topoisomerase I and Topoisomerase I-DNA Complex Formation, *Annals of Clinical & Laboratory Science*, **2001**, 31, 187-198.0

(253) Bellorini, M.; Moncollin, V.; D'Incalci, M.; Mongelli, N.; Mantovani, R. Distamycin A and Tallimustine Inhibit TBP Binding and Basal *In Vitro* Transcription, *Nucleic Acids Research*, **1995**, 23, 1657-1663.

---

(254) Harshman, K. D.; Dervan, P. B. Molecular Recognition of B-DNA by Hoechst33258. *Nucleic Acids Research* **1985**, *13*, 4825-4835.

(255) Drobyshev, A. L.; Zasedatelev, A. S.; Yershov, G. M.; Mirzabekov, A. D. Massive Parallel Analysis of DNA-Hoechst33258 Binding Specificity with a Generic Oligodeoxyribonucleotide Microchip. *Nucleic Acids Research*, **1999**, *27*, 4100-4105.

(256) Breusegem, S. Y.; Clegg, R. M.; Loontjens, F. G. Base-Sequence Specificity of Hoechst33258 and DAPI Binding to Five (A/T)<sub>4</sub> DNA Sites with Kinetic Evidence for More Than One High-Affinity Hoechst33258-AATT Complex, *J. Mol. Biol.*, **2002**, *315*, 1049-1061.

(257) Abu-Daya, A.; Brown, P. M.; Fox, K. R. DNA Sequence Preferences of Several AT-Selective Minor Groove Binding Ligands, *Nucleic Acids Research*, **1995**, *23*, 3385-3392.

(258) Haq, I. Part II: The Thermodynamics of Drug-Bipolymer Interaction, Thermodynamics of Drug-DNA Interactions. *Archives of Biochemistry and Biophysics*, **2002**, *403*, 1-15.

(259) Haq, I.; Ladburry, J. E.; Chowdhry, B. Z.; Jenkins, T. C.; Chairs, J. B. Specific Binding of Hoechst33258 to the d(CGCAAATTTGCG)<sub>2</sub> Duplex: Calorimetric and Spectroscopic Studies. *J. Mol. Biol.* **1997**, *271*, 244-257.

(260) Woods, K. K.; Maehigashi, T.; Howerton, S. B.; Sines, C. C.; Tannenbaum, S. T.; Williams, L. D.; High-Resolution Structure of an Extended A-Tract: [d(CGCAAATTTGCG)]<sub>2</sub>, *J. Am. Chem. Soc.*, **2004**, *126*, 15330-15331.

(261) Yoon, C.; Privé, G. G.; Goodsell, D. S.; Dickerson, R. E. Structure of an Alternating-B DNA Helix and its Relationship to A-tract DNA, *Proc. Natl. Acad. Sci. USA*, **1988**, *85*, 6332-6336.

(262) Edwards, K. J.; Brown, D. G.; Spink, N.; Neidle, S. Molecular Structure of the B-DNA Dodecamer d(CGCAAATTTGCG)<sub>2</sub>. An examination of Propeller Twist and Minor-Groove Water Structure at 2.2 Å Resolution, *J. Mol. Biol.*, **1992**, *226*, 1161-1173.

(263) Teng, M.; Usman, N.; Frederick, C. A., Wang, A. H.-J. The Molecular Structure of the Complex of Hoechst33258 and the DNA Dodecamer d(CGCGAATTCGCG). *Nucleic Acids Research*, **1988**, *16*, 2671-2690.

(264) Quintana, J. R.; Lipanov, A. A.; Dickerson, R. E. Low-Temperature Crystallographic Analyses of the Binding of Hoechst33258 to the Double-Helical DNA Dodecamer C-G-C-G-A-A-T-T-C-G-C-G, *Biochemistry*, **1991**, *30*, 10294-10306.

(265) Vega, M. C.; Sáez, I. G.; Aymami, J.; Erita, R.; van der Marel, G. A.; van Boom, J. H.; Rich, A.; Coll, M. Three-Dimensional Crystal Structure of the A-tract DNA Dodecamer d(CGCAAATTTGCG) Complexed with the Minor-Groove-Binding Drug Hoechst33258, *Eur. J. Biochem.*, **1994**, *222*, 721-726.

(266) Spink, N.; Brown, D. G.; Skelly, J. V.; Neidle, S. Sequence-Dependent Effects in Drug-DNA Interaction: The Crystal Structure of Hoechst33258 Bound to the d(CGCAAATTTGCG)<sub>2</sub> Duplex, *Nucleic Acids Research*, **1994**, *22*, 1607-1612.

(267) Carrondo, M. A. A. F.; De C. T.; Coll, M.; Aymami, J.; Wang, A. H.-J.; van der Marel, G. A.; van Boom, J. H.; Rich, A. Binding of Hoechst Dye to d(CGCGATATCGCG) and its Influence on the Conformation of the DNA Fragment. *Biochemistry*, **1989**, *28*, 7849-7859.

---

(268) Parkinson, J. A.; Barber, J.; Douglas, K. T.; Rosamond, J.; Sharples, D. Minor-Groove Recognition of the Self-Complementary Duplex d(CGCGAATTCGCG)<sub>2</sub> by Hoechst33258: A High-Field NMR Study. *Biochemistry*, **1990**, *29*, 10181-10190.

(269) Fede, A.; Labhardt, A.; Bannwarth, W.; Leupin, W. Dynamics and Binding Mode of Hoechst33258 to d(GTGGAATTCAC)<sub>2</sub> in the 1:1 Solution Complex as Determined by Two-Dimensional <sup>1</sup>H NMR. *Biochemistry*, **1991**, *30*, 11377-11388.

(270) Higgins, L. D.; Searle, M. S. Site-Specificity of Bis-Benzimidazole Hoechst33258 in A-tract Recognition of the DNA Dodecamer Duplex d(GCAAATTTTGC)<sub>2</sub>. *Chem. Commun.*, **1999**, *18*, 1861-1862.

(271) Parkinson, J. A.; Ebrahimi, S. E.; McKie, J. H.; Douglas, K. T. Molecular Design of DNA-Directed Ligands with Specific Interactions: Solution NMR Studies of the Interaction of a m-Hydroxy Analog of Hoechst33258 with d(CGCGAATTCGCG)<sub>2</sub>. *Biochemistry*, **1994**, *33*, 8442-8452.

(272) Bostock-Smith, C. E.; Harris, S. A.; Laughton, C. A.; Searle, M. S. Induced Fit DNA Recognition by a Minor Groove Binding Analogue of Hoechst33258: Fluctuations in DNA A-tract Structure Investigated by NMR and Molecular Dynamics Simulations, *Nucleic Acids Research*, **2001**, *29*, 693-702.

(273) Gooljarsingh, L. T.; Fernanades, C.; Yan, K.; Zhang, H.; Grooms, M.; Johanson, K.; Sinnamon, R. H.; Kirkpatrick, R. B.; Kerrigan, J.; Lewis, T.; Arnone, M.; King, A. J.; Lai, Z.; Copeland, R. A.; Tummino, P. J.; A biochemical rationale for the anticancer effects of Hsp90 inhibitors: Slow, tight binding inhibition by geldanamycin and its analogues. *Proc. Natl. Acad. Sci. U. S. A.*, **2006**, *20* (103), 7625-7630.

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

(275) Park, S.; Kwak, K.; Fayer, M. D. Ultrafast 2D-IR Vibrational Echo Spectroscopy: A Probe of Molecular Dynamics, *Laser Phys. Lett.*, **2007**, *10*, 704-718.

(276) Baiz, C. R.; Mcrobbie, P. L.; Anna, J. M.; Geva, E.; Kubarych, K. J. Two-Dimensional Infrared Spectroscopy of Metal Carbonyls, *Accounts of Chemical Research*, **2009**, *42*, 1395-1404.

(277) Downs, T. R.; Wilfinger, W. W. Fluorometric Qualification of DNA in Cells and Tissue. *Anal. Biochem.* **1983**, *131*, 538-547.

(278) Peng, C. S.; Jones, K. C.; Tokmakoff, A. Anharmonic Vibrational Modes of Nucleic Acid Bases Revealed by 2D IR Spectroscopy, *J. Am. Chem. Soc.*, **2011**, *133*, 15650-15660.

(279) Yang, M.; Szyk, L.; Elsaesser, T. Femtosecond Two-Dimensional Infrared Spectroscopy of Adenine-Thymine Base Pairs in DNA Oligomers, *J. Phys. Chem. B*, **2011**, *115*, 1262-1267.

(280) Greve, C.; Elsaesser, T. Ultrafast Two-Dimensional Infrared Spectroscopy of Guanine-Cytosine Base Pairs in DNA Oligomers, *J. Phys. Chem. B*, **2013**, *117*, 14009-14017.

(281) Krummel, A. T.; Zanni, M. T. DNA Vibrational Coupling Revealed with Two-Dimensional Infrared Spectroscopy: Insight into Why Vibrational Spectroscopy is Sensitive to DNA Structure, *J. Phys. Chem. B*, **2006**, *110*, 13991-14000.

(282) Hithell, G.; Shaw, D. J.; Donaldson, P. M.; Greetham, G. M.; Towrie, M.; Burley, G. A.; Parker, W. A.; Hunt, N. T. Long-Range Vibrational Dynamics Are Directed by Watson-Crick Base Pairing in Duplex DNA. *J. Phys. Chem. B*, **2016**, *120*, 4009-4018.

- 
- (283) Guchhait, B.; Liu, Y.; Siebert, T.; Elsaesser, T. Ultrafast Vibrational Dynamics of the DNA Backbone at Different Hydration Levels Mapped by Two-Dimensional Infrared Spectroscopy, *Structural Dynamics*, **2016**, *3*, 043202-043217.
- (284) Yang, M.; Szyc, L.; Elsaesser, T. Decelerated Water Dynamics and Vibrational Couplings of Hydrated DNA Mapped by Two-Dimensional Infrared Spectroscopy, *J. Phys. Chem. B*, **2011**, *115*, 13093-13100.
- (285) Sanstead, P. J.; Stevenson, P.; Tokmakoff, A. Sequence-Dependent Mechanism of DNA Oligonucleotide Dehybridization Resolved through Infrared Spectroscopy. *J. Am. Chem. Soc.*, **2016**, *138*, 11792-11801.
- (286) Lee, C.; Cho, M. Vibrational Dynamics of DNA. II. Deuterium Exchange Effects and Simulated IR Absorption Spectra, *J. Chem. Phys.*, **2006**, *125*, 114509.
- (287) Lee, C.; Park, K.-H.; Cho, M. Vibrational Dynamics of DNA. I. Vibrational Basis Modes and Couplings, *J. Chem. Phys.*, **2006**, *125*, 114508.
- (288) Lee, C.; Park, K.-H.; Kim, J.-A.; Hahn, S.; Cho, M. Vibrational Dynamics of DNA. III. Molecular Dynamics Simulations of DNA in Water and Theoretical Calculations of the Two-Dimensional Vibrational Spectra, *J. Chem. Phys.*, **2006**, *125*, 114510.
- (289) Lee, C.; Cho, M. Vibrational Dynamics of DNA. IV. Vibrational Spectroscopic Characteristics of A-, B-, and Z-Form DNAs, *J. Chem. Phys.*, **2007**, *126*, 145102.
- (290) Mukerji, I.; Williams, A. P. UV Resonance Raman and Circular Dichroism Studies of a DNA Duplex Containing an A<sub>3</sub>T<sub>3</sub> Tract: Evidence for a Premelting Transition and Three-Centered H-Bonds, *Biochemistry*, **2002**, *41*, 69-77.
- (291) Thorogood, H.; Waters, T. R.; Parker, A. W.; Wharton, C. W.; Connolly, B. A. Resonance Raman Spectroscopy of 4-Thiothymidine and Oligodeoxynucleotides Containing This Base Both Free in Solution and Bound to the Restriction Endonuclease EcoRV. *Biochemistry*, **1996**, *35*, 8723-8733.
- (292) 
$$\Delta\Delta H = \left(\frac{\Delta\nu}{\nu}\right) U_{Bond}$$
where  $\Delta\Delta H$  is the change in the hydrogen bond strength,  $\Delta\nu$  is the shift in the frequency of the vibrational mode,  $\nu$  is the original frequency of the vibrational mode and  $U_{Bond}$  is the bond energy of the affected moiety. Note this equation is only valid if  $\Delta\nu$  is small compared to  $U_{Bond}$ .
- (293) Kiser, J. R.; Monk, R. W.; Smalls, R. L.; Petty, J. T. Hydration Changes in the Association of Hoechst33258 with DNA. *Biochemistry*, **2005**, *44*, 16988-16997.
- (294) Fresch, B.; Remacle, F. Atomistic Account of Structural and Dynamical Changes Induced by Small Binders in the Double Helix of a Short DNA, *Phys. Chem. Chem. Phys.*, **2014**, *16*, 14070-14082.
- (295) Han, F.; Taulier, N.; Chalikian, T. V. Association of the Minor Groove Binding Drug Hoechst33258 with d(CGCGAATTCGCG)<sub>2</sub>: Volumetric, Calorimetric, and Spectroscopic Characterizations, *Biochemistry*, **2005**, *44*, 9785-9794.
- (296) Shaw, D. J.; Adamczyk, K.; Frederix, P. W. J. M.; Simpson, N.; Robb, K.; Greetham, G. M.; Towrie, M.; Parker, A. W.; Hoskisson, P. A.; Hunt, N. T. Multidimensional Infrared Spectroscopy Reveals the Vibrational and Solvational Dynamics of Isoniazid, *J. Chem. Phys.*, **2015**, *142*, 212401.

---

(297) Greetham, G. M.; Burgos, P.; Cao, Q.; Clark, I. P.; Codd, P. S.; Farrow, R. C.; George, M. W.; Kogimtzis, M.; Matousek, P.; Parker, A. W.; *et al.* M. ULTRA: A Unique Instrument for Time-Resolved Spectroscopy. *Applied Spectroscopy*, **2010**, *12*, 1311-1319.

# Probing Minor Groove Binding and Hydration via a Novel Azido-Derivative of Hoechst 33258

Chapter 6

## 6.1 Abstract

*Designing sequence-selective molecular probes of DNA-ligand interactions requires accurate control of enthalpic contributions from DNA-ligand contacts and entropic effects relating to the dynamic molecular environment of the minor groove. Combining a de novo-designed bis-benzimidazole ligand featuring an azido moiety with NMR and 2D-IR spectroscopy has facilitated simultaneous interrogation of both the impact of ligand binding on the molecular structure of DNA and the dynamic environment of the bound probe molecule. Binding of the probe to 10-mer ds-DNA featuring an A-tract (AAATTT) motif leads to a large stabilisation of the melting temperature, while NMR and 2D-IR spectroscopy reveal sequence selective binding via an induced fit interaction. The probe is sited in an environment featuring interactions with solvent molecules albeit exhibiting significantly restricted dynamics relative to bulk water. Upon binding to a ds-DNA sequence featuring an alternating ATATAT motif, the melting temperature stabilisation is significantly reduced and 2D-IR spectroscopy indicates a loss of the specific induced fit interaction while the probe resides in an environment with H-bonding and dynamics strongly reminiscent of bulk water. Together these results allow calibration of the enthalpic benefits arising from induced fit binding and evaluate the local impact upon sequence selectivity of incorporating an azido moiety into a minor groove binding ligand.*

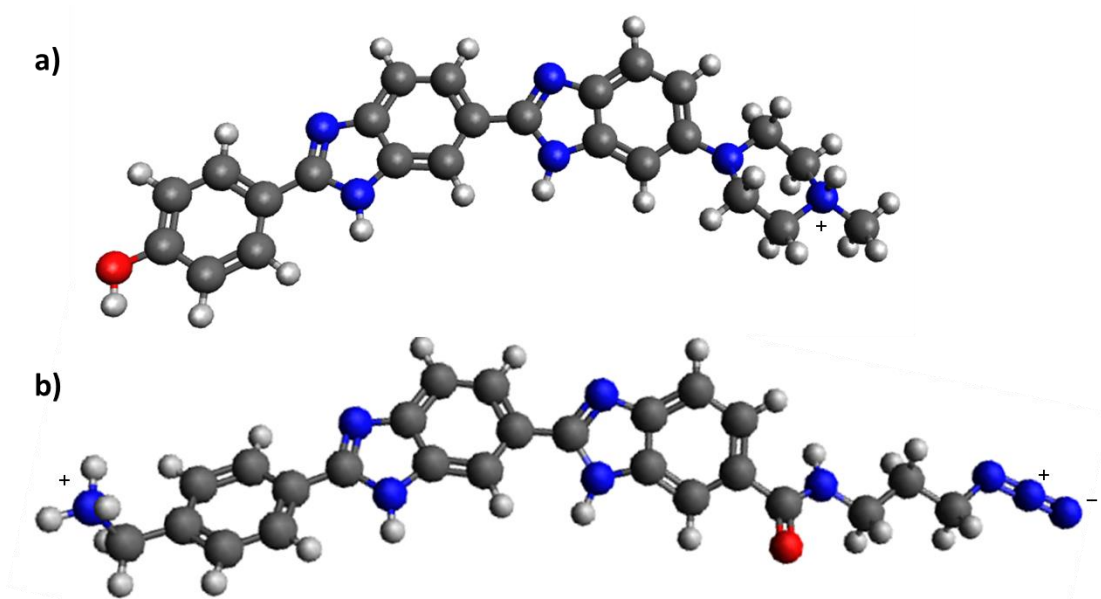


## 6.2 Introduction

Minor groove binding to DNA is a widely studied topic,<sup>298</sup> due to the potential of these minor groove binding molecules to act as sequence-sensitive probes of DNA function in cells and their potential therapeutic activities.<sup>299,300,301</sup> The construction of any scheme to rationally design a minor groove binder is complicated by the plethora of competing factors, which contribute to both the changes in enthalpy and entropy to be balanced upon binding. These factors include but are not limited to the loss of the spine of hydration, the formation of ligand-duplex H-bonds, the formation of favourable hydrophobic contacts between the ligand and the duplex and the loss of rotational degrees of freedom of the ligand upon binding. This means that we need to gain a more fundamental understanding of aspects of the molecular interactions such as H-bonding patterns, hydrophobic interactions as well as the shape of both the minor groove and the ligand. Additionally it is important to understand attributes of the molecular environment including the loss of the spine of hydration and the relative solubility of the ligand versus the bound state. Currently the molecular interactions formed between the ligand and DNA duplex have been widely studied using both NMR spectroscopies<sup>302,303,304,305,306,307</sup> and X-ray crystallography.<sup>308,309,310,311,312,313,314,315</sup> In addition to this, a computational<sup>316</sup> and a fluorescence<sup>317</sup> study have allowed the molecular environment that the ligand encounters upon binding to be explored.

Two dimensional spectroscopy (2D-IR)<sup>318,319,320</sup> has proved to be a sensitive probe of dynamics and structural aspects of the DNA macromolecule via a combination of inherently high time resolution and the ability to measure complex intermolecular vibrational couplings induced by duplex formation.<sup>321,322,323,324</sup> In chapter 5, 2D-IR spectroscopy was used to probe the molecular interactions underpinning the binding of H33258, an archetypal minor groove binder, to both its preferred A-tract sequence and a sub-optimal alternating A–T sequence.<sup>325</sup> In this chapter it was demonstrated that 2D-IR can differentiate optimal induced fit binding from sub-optimal rigid body binding, where the formation of the complex just leads to the interruption of the minor groove spine of hydration. However, the 2D-IR spectroscopy of the DNA modes only reports on aspects of the molecular interactions of DNA binding. In the current study we will explore the possibility of extending 2D-IR spectroscopy to explore aspects of the molecular environment which are also important when considering DNA minor groove binding.

It is often recognized that natural IR probe modes within biological systems are located in highly congested regions of the IR spectrum.<sup>326</sup> This can present challenges when attempts are made to probe the nature of the molecular environments found within these systems, and in order to overcome such issues it is often possible to introduce non-natural IR probe moiety into these systems. There are a multitude of non-natural IR probes (e.g. nitriles,<sup>327</sup> cyanates,<sup>328</sup> thiocyanates<sup>329</sup> and azides<sup>330</sup>) which have been used to successfully study a range of different biological systems.<sup>331,332,333</sup> Among these different non-natural IR probes, the aliphatic azides have been shown to be highly sensitive to hydrogen bonding. More specifically it has been predicted by a recent computational study that such azide moieties are actually highly sensitive to changes in the geometry of the hydrogen bonds formed to the probe.<sup>334</sup> This special behaviour of the aliphatic azides is thought to be ideally suited to the study of the minor groove of DNA. We show that the incorporation of a non-natural azide probe into a ligand based on a head-to-tail *bis*-benzimidazole-type<sup>298</sup> backbone ( $N_3$ -bBI) allows us to simultaneously explore the structural and dynamic effects of binding via the DNA and the non-natural probe, as well as evaluating the effect of the incorporation of an azide group on the induced fit and rigid body interaction mechanisms.



**Figure 6.1:** Optimised geometry of the a) H33258 and b)  $N_3$ -bBI ligands obtained utilising the hybrid-DFT B3-LYP functional with a 6-311G(d,p) basis set.<sup>154</sup>

Since the synthesis of the original *bis*-benzimidazole minor groove binder, H33258,<sup>335</sup> these pharmacophores have been found to exhibit interesting biological activities.<sup>336,337,338</sup> A wide range of analogues of H33258 have been studied to gain a more fundamental understanding

of the nature of these binding interactions.<sup>339,340,341</sup> Unfortunately, any definitive trends between structural change and the sequence-selectivity have remained elusive.<sup>342</sup>

N<sub>3</sub>-bBI (Fig.6.1) is structurally similar to H33258 with the most prominent difference being the replacement of the piperazine ring by an aliphatic azide label, allowing the effect of an additional hydrophilic group on the binding to be investigated. We compare the binding of N<sub>3</sub>-bBI to a sequence featuring an A-tract motif, d(GGAAATTTGC)<sub>2</sub>, (A<sub>3</sub>T<sub>3</sub>), with a sequence containing an alternating A–T motif, d(GGATATATGC)<sub>2</sub>, (AT)<sub>3</sub>. H33258 has been shown to show higher affinity for A-tract versus alternating sequences<sup>343</sup> and to bind via different mechanisms (chapter 5) allowing direct evaluation of the impact of the azide label.

2D-IR spectroscopy shows that N<sub>3</sub>-bBI participates in an induced fit to the A-tract sequence, similar to H33258, but exhibits non-specific binding to the alternating A–T sequence whereas H33258 bound with reduced affinity via a rigid body interaction. Upon binding to the A-tract sequence the azide probe encounters a unique H-bonding environment with significantly restricted dynamics whereas binding to the alternating sequence results in an azide environment strongly reminiscent of bulk water. The melting temperature stabilisation caused by the N<sub>3</sub>-bBI ligand in the alternating sequence was significantly lower than observed for H33258, suggesting that it is more selective for the A-tract sequence, implying that the benefits of induced fit binding outweigh any negative impact of azide incorporation, in contrast to the rigid-body binding mechanism.

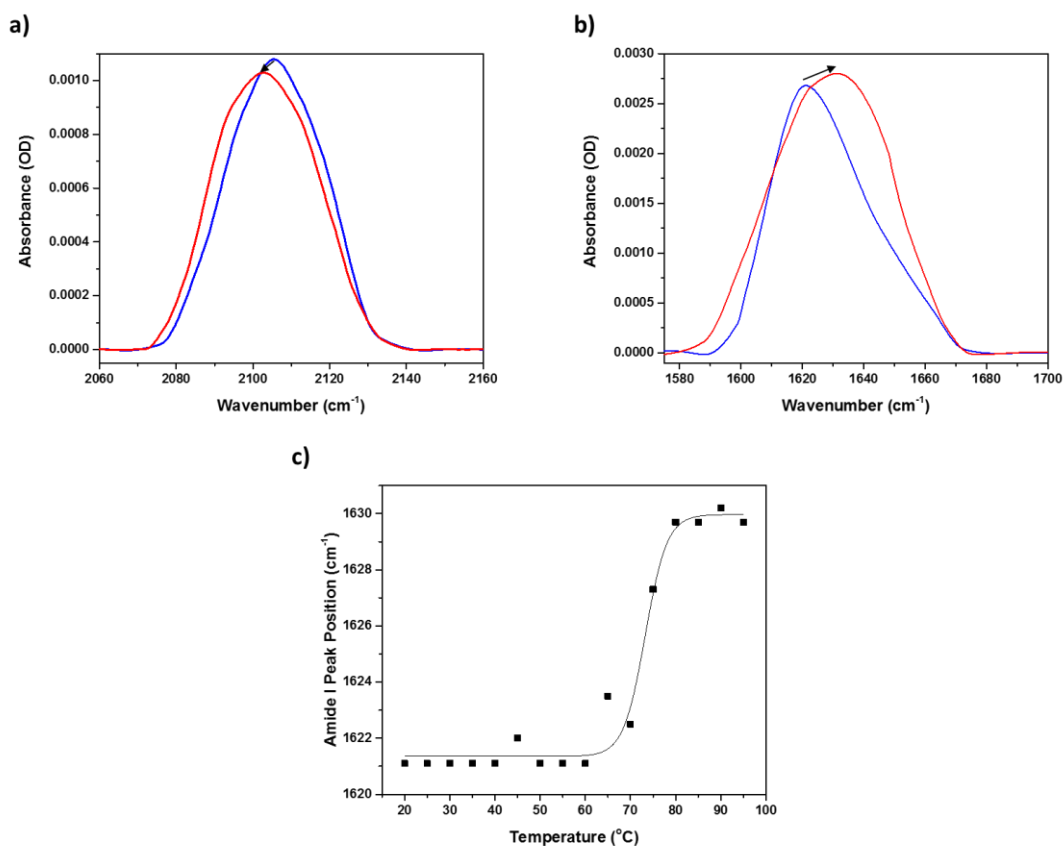
## 6.3 Results and Discussion

### Difference Spectroscopy

#### 6.3.1 Ligand Characterisation

In order to gain a better understanding of the similarities and differences between the geometries of H33258 and N<sub>3</sub>-bBI ligands a series of quantum chemical computations were carried out. The calculations, performed on the *de novo* N<sub>3</sub>-bBI ligand, predicted that the FT-IR spectrum contains a range of low frequency ring modes, an amide I mode, predicted to be centred at ~1656 cm<sup>-1</sup>, and an asymmetric azide stretch located at ~2105 cm<sup>-1</sup> (Fig.A6.1). It is important to understand the spectra of the amide I and asymmetric azide stretch, as the amide I mode is predicted to be located in the same spectral window as the DNA base vibrational modes and the azide mode will function as a non-natural IR probe for the N<sub>3</sub>-bBI:DNA complexes. The FT-IR spectra of the asymmetric azide stretch and amide I mode of

$N_3$ -bBI in a pD7 phosphate buffer solution at 20 °C (blue) and 95 °C (red) are shown in figure 6.2.



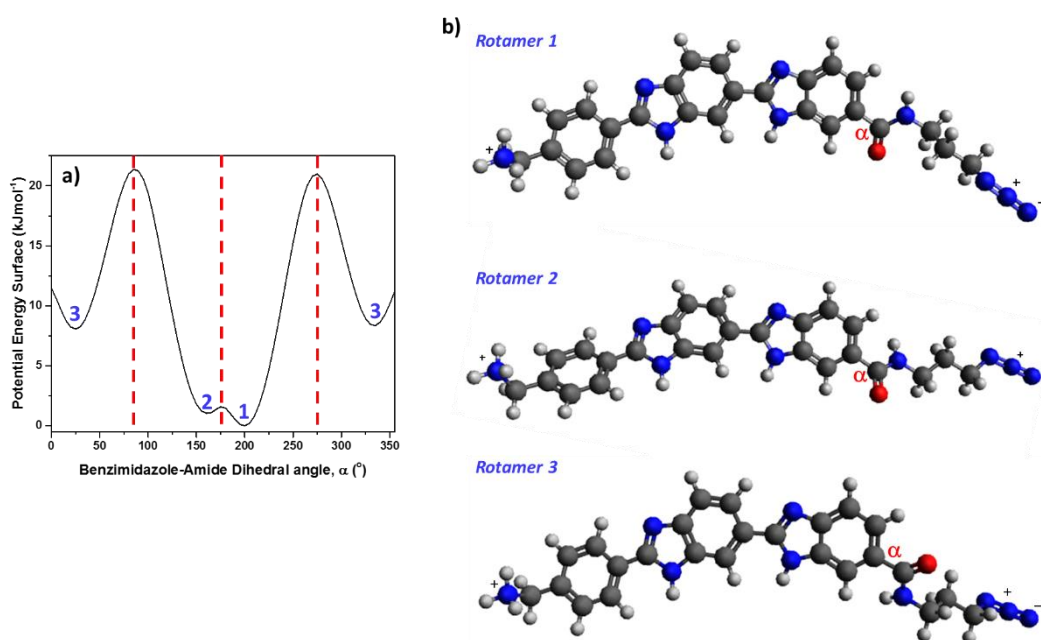
**Figure 6.2:** a) FT-IR spectra of the azido-Hoechst asymmetric azide stretch measured at 20 °C (blue) and 95 °C (red), b) FT-IR spectra of the azido-Hoechst amide I mode measured at 20 °C (blue) and 95 °C (red), and c) the thermal profile of the position of the absorbance maximum of the amide I vibrational mode of azido-Hoechst in pD7 phosphate buffer solution.

The asymmetric azide stretch of  $N_3$ -bBI is found to consist of a single peak, centred at 2105 cm<sup>-1</sup> at 20 °C (Fig.6.2.(a), blue) and centred at 2102 cm<sup>-1</sup> at 95 °C (Fig.6.2.(a), red). The observed shift to lower frequency is noted to be consistent with the changes observed for other azides, in aqueous environments, as the sample temperature increases. Overall the simplicity of the line-shape of the asymmetric azide stretch of  $N_3$ -bBI shows the utility of this vibrational mode to act as a non-natural IR probe of the molecular environment. In contrast to the simplicity of the asymmetric azide stretch, the amide I mode of  $N_3$ -bBI at 20 °C (Fig.6.2.(b), blue) found to consist of two peaks centred at 1620 cm<sup>-1</sup> and 1633 cm<sup>-1</sup> respectively. As the temperature is increased it is noted that the peak centred at 1620 cm<sup>-1</sup> decreases in intensity and the peak centred at 1633 cm<sup>-1</sup> increases in intensity. Furthermore

tracking the peak position of the amide I band as the sample temperature is increased yields a sigmoidal curve with a transition temperature of 73 °C.

Interestingly the change in the amide I mode of  $N_3$ -*b*BI as the temperature is increased is noted to coincide with the disappearance of an aggregate in the sample. Further it was noted that the amount of aggregate in the sample could be controlled by changing the pD value of the surrounding environment (Fig.A6.1). Increasing the pD of the solution leading to a larger amount of aggregate being observed in the sample resulting in the 1620  $\text{cm}^{-1}$  becoming the dominant contribution to the amide I band (Fig.A6.1, pink) whereas decreasing the pD of the solution decreased the amount of aggregate observed in the sample resulting in the 1633  $\text{cm}^{-1}$  contribution dominating the amide I mode (Fig.A6.1, red). This is consistent with  $N_3$ -*b*BI being only partially soluble at pD7 leading to the presence of two distinct environments, as seen from the amide I mode (Fig.6.2.(b), blue), with the dissolution of the  $N_3$ -*b*BI aggregate at high temperatures leading to the observed changes in the amide I mode at pD7 as the temperature is increased to 95 °C. Finally it is noted that the sensitivity of the solubility to the pD value of the surrounding solution is consistent with the protonation state of the benzyl amine moiety of the ligand. As the pD increases the benzyl amine moiety becomes deprotonated removing the positive charge on this moiety. The loss of this charge decreases the solubility of the ligand leading to the formation of an aggregate. However, in the subsequent binding experiments the sample is annealed at 90 °C for approximately 10 minutes to attempt to overcome the preference of this ligand to aggregate.

In addition to the characterisation of the physical properties of the ligand in solution the overall shape of the  $N_3$ -*b*BI ligand was investigated using quantum chemical calculations.<sup>154</sup> The most important aspect of the ligand shape is the overall curvature, which ideally is iso-helical with the target DNA sequence. One of the most important contributions to the geometry of the  $N_3$ -*b*BI ligand was found to be the shape of the flexible N-(3-azidopropyl)formamide moiety, and the calculation carried out on the ligand demonstrate that the geometry of this moiety is linked to the benzimidazole-amide dihedral angle ( $\alpha$ ). Rotating the benzimidazole-amide dihedral angle ( $\alpha$ ) reveals three stable rotamers of  $N_3$ -*b*BI, these rotamers are shown in figure 6.3, alongside the potential energy surface obtained by rotating  $\alpha$ .

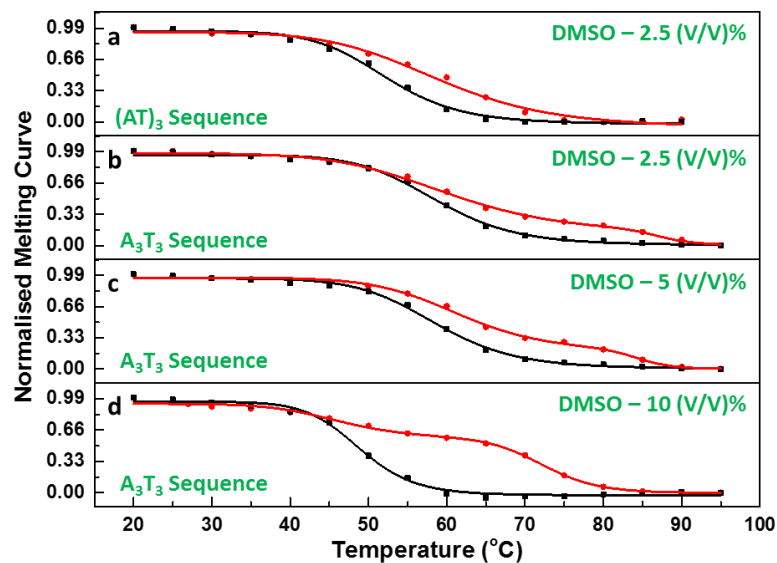


**Figure 6.3:** a) Calculated potential energy surface for the benzimidazole-amide dihedral angle ( $\alpha$ ) and b) the optimised geometries of the identified rotamers. All calculations performed utilising the hybrid-DFT B3LYP functional with a 6-311G(d,p) basis set.<sup>154</sup>

The calculated potential energy surface obtained from the rotation of  $\alpha$ , is shown in Fig.6.3.(a). From the calculated potential energy surface the relative abundances of the three rotamers present at 20 °C were determined to be 59.1% (rotamer 1), 38.6% (rotamer 2) and 2.3% (rotamer 3).<sup>344</sup> Furthermore it was noted that at 20 °C the thermal energy available to the system is 1.65 kJmol<sup>-1</sup>, which is sufficient to allow thermal interconversion between rotamers 1 and 2. The comparison between the curvatures of these two dominant interconverting rotamers and the archetypal H33258 ligand shows that these calculation predict that the iso-helical shape of these two ligands are relatively similar (Figs.A6.3.(b)&(c)), broadly suggesting that there should be similarities in the binding geometries of these two different *bis*-benzimidazole ligands.

### 6.3.2 Binding Characterisation

UV-visible absorption spectroscopy of a sample containing a 1:1 mixture of N<sub>3</sub>-*b*BI and (AT)<sub>3</sub> sequence yielded a melting curve that could be well-represented by a sigmoidal function with a T<sub>m</sub> of 59 °C (Fig.6.4.(a), red). Comparing the melting curve of the complex formed between N<sub>3</sub>-*b*BI and the (AT)<sub>3</sub> sequence to the melting curve of the uncomplexed DNA sequence (Fig.6.4.(a), back) shows that N<sub>3</sub>-*b*BI binding results in a T<sub>m</sub> stabilisation of 7 °C.



**Figure 6.4:** Normalised UV-visible melting curves of the DNA without (black) and with (red) the  $N_3$ -bBI ligand for a) the  $(AT)_3$  sequence with 2.5 (V/V)% DMSO, b) the  $A_3T_3$  sequence with 2.5 (V/V)% DMSO, c) the  $A_3T_3$  sequence with 5 (V/V)% DMSO and d) the  $A_3T_3$  sequence with 10 (V/V)% DMSO.

In contrast, measurements of a sample containing a 1:1 mixture of the  $N_3$ -bBI ligand and the  $A_3T_3$  dsDNA sequence produced a more complex melting curve (Fig.6.4.(b), red). Fitting of this profile required a function that was the sum of two sigmoidal functions with melting temperatures of 60 °C and 84 °C. The former value was in close agreement with that obtained for the uncomplexed  $A_3T_3$  (Fig.6.4.(b), black) dsDNA sequence, the latter, 24 °C higher, is attributed to formation of the  $N_3$ -bBI: $A_3T_3$  complex. The stabilisation of the melting temperature of the bound  $A_3T_3$  duplexes in this sample is noted to be identical to stabilisation observed for the binding of H33258 to the A-tract sequence (chapter 5). It was observed that  $N_3$ -bBI was less soluble than in the presence of the  $(AT)_3$  sequence, but that incorporation of small amounts of DMSO increased the solubility. Repeating the UV-vis melting experiments in the presence of increasing amounts of DMSO (from 2.5 (v/v)% to 10 (v/v)%) led to a progressive increase in the amplitude of the sigmoidal function with the higher melting temperature (Figs.6.4.(b)-(d), red). This is assigned to the DMSO increasing the solubility of the ligand such that a greater fraction of the DNA formed a complex with a higher melting temperature than the uncomplexed DNA. This demonstrates that the  $N_3$ -bBI ligand stabilising the A-tract to a much greater extent than the alternating sequence. In order to avoid perturbing the structure of the  $A_3T_3$  duplex (SI, Fig.A6.4) all subsequent datasets including

the A-tract DNA sequence were obtained at 2.5 (V/V)% DMSO and so the bound fraction of dsDNA within the A<sub>3</sub>T<sub>3</sub> samples is 23%.

#### A-tract binding

#### **Infrared spectroscopy – DNA Base Modes**

The FT-IR spectra of the A<sub>3</sub>T<sub>3</sub> DNA sequence and its complex with N<sub>3</sub>-bBI (N<sub>3</sub>-A<sub>3</sub>T<sub>3</sub>) are shown in Fig.6.5.(a). The IR spectrum of the A<sub>3</sub>T<sub>3</sub> sequence shows four peaks (Fig.6.3.(a), red), these are indicated by grey dashed lines in the figure and listed in Table 6.1. The results are consistent with previous studies and the peaks are assigned by reference to spectra of DNA duplexes containing exclusively GC or AT base pairs, shown in Figs.6.5.(d) and A6.5.<sup>345,346,347,348</sup> The assignments are summarized in Table 6.1.

**Table 6.1:** Assignment of peaks in the IR spectra of A<sub>3</sub>T<sub>3</sub> and (AT)<sub>3</sub> DNA duplexes, showing comparison with AT and GC-only sequences.

Assignment	Position (cm <sup>-1</sup> )				Description
	A <sub>3</sub> T <sub>3</sub>	(AT) <sub>3</sub>	AT <sup>a</sup>	GC <sup>b</sup>	
AT <sub>2S</sub>	1692	1689	1692		T <sub>2</sub> C=O stretch
G <sub>S</sub> C <sub>S</sub> (-)				1684	GC C=O antisymmetric stretch
AT <sub>4S</sub>	1666	1666	1664		Base-paired T <sub>4</sub> C=O stretch
G <sub>S</sub> C <sub>S</sub> (+)				1651	GC C=O symmetric stretch
T <sub>R</sub>	1646	1650	1640		T ring vibration
A <sub>R1</sub> T	1622	1622	1622 (s)		Coupled AT ring vibration/
G <sub>S</sub> C <sub>R</sub>				1622 (w)	C ring mode + G C=O

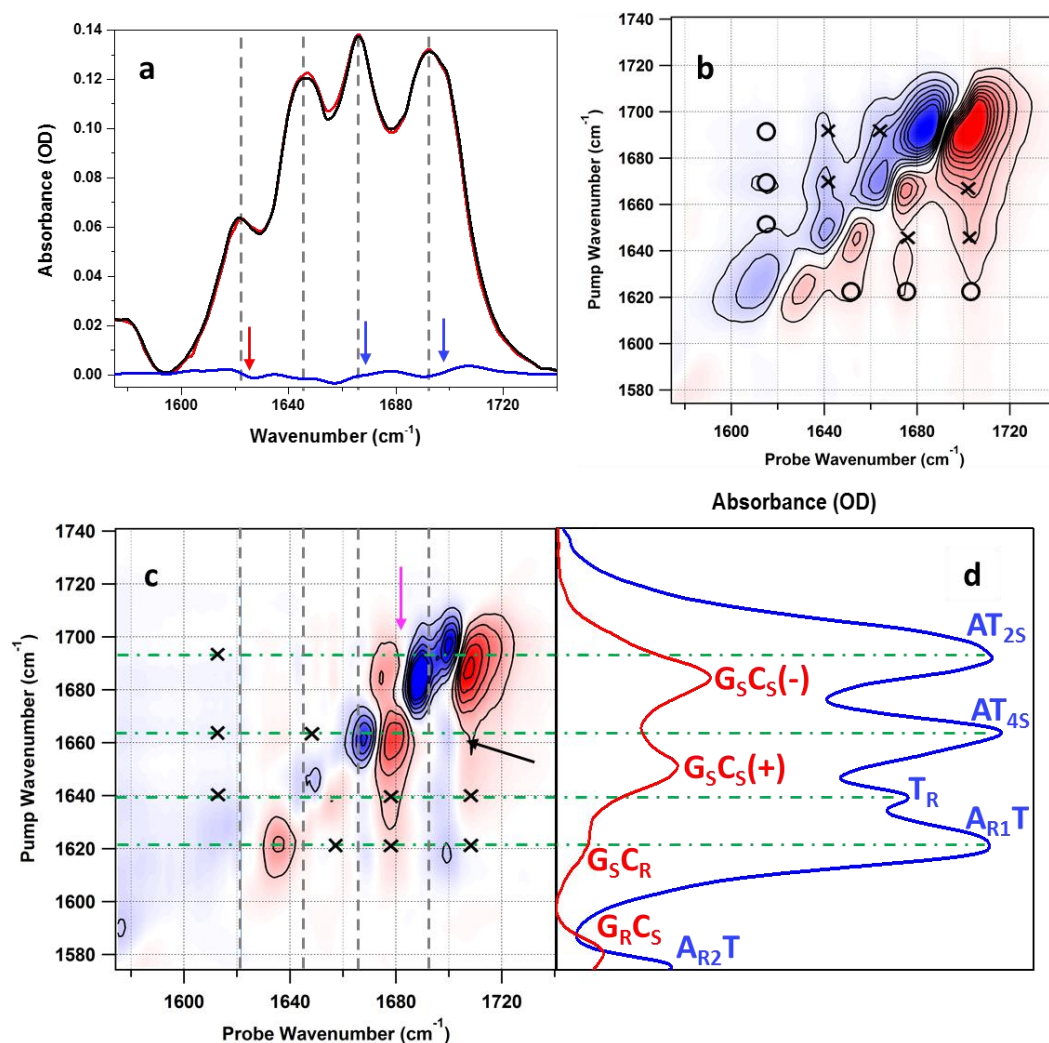
<sup>a</sup> obtained from the sequence: 5'-ATTATTATTATTA-3' (SI, Fig A6.5(a)); <sup>b</sup> obtained from the sequence: 5'-GCCGCCGCCG-3' (SI, Fig A6.5(b)); modes marked (s) are those which contribute strongly to the overall spectrum and modes marked (w) are those which contribute weakly to the overall spectrum.

Both AT and GC vibrational modes contribute to the FT-IR spectrum of the A<sub>3</sub>T<sub>3</sub> sequence, with the GC modes mainly influencing the intensities between the observed peaks.<sup>325</sup> The presence of distinct cross peaks patterns between the vibrational modes, accessed by 2D-IR methods, allow changes in AT and GC vibrational modes to be separated from each other unravelling the congestion of this region.

The binding-induced changes to the FT-IR spectrum upon the formation of the N<sub>3</sub>-A<sub>3</sub>T<sub>3</sub> complex (Fig. 6.5.(a), black) are extracted via a difference spectrum (Fig. 6.5.(a), blue). The



$N_3$ -bBI ligand is noted to exhibit an amide I mode located at  $1622\text{ cm}^{-1}$  (SI, Fig.A6.7) but, at the concentrations measured, its contribution is negligible (3–5%).



**Figure 6.5:** a) FT-IR spectra of  $A_3T_3$  DNA with (black) and without (red)  $N_3$ -bBI. Blue spectrum shows the binding-induced difference FT-IR spectrum (complex-free sequence). b) 2D-IR spectrum of  $N_3$ - $A_3T_3$  complex. Crosses mark off-diagonal peaks assigned to coupling of modes primarily located on T-base. Circles show off-diagonal peaks assigned to coupling of modes on A and T bases induced by W-C base pairing. c) Ligand binding induced 2D-IR difference spectrum. Crosses show locations of small off-diagonal features. d) FT-IR spectra of GC and AT-only DNA sequences to show the relative (weighted per base) magnitudes of peaks and peak positions to aid assignment of the IR spectrum of the  $A_3T_3$  sequence, see also Fig A6.5).

The  $N_3$ - $A_3T_3$  complex shows three main features in the difference IR spectrum (Fig.6.5.(a), blue). Two of these features, near  $1665\text{ cm}^{-1}$  and  $1700\text{ cm}^{-1}$  (blue arrows), consist of a positive peak located to the higher frequency side of a negative peak, suggesting that the  $AT_{4S}$  and  $AT_{2S}$  modes shift to a higher wavenumber upon binding. The third feature is located at

approximately  $1625\text{ cm}^{-1}$  (red arrow), and in this case it consists of a positive peak located to the lower frequency side of a negative peak. This is consistent with a shift to lower wavenumber of the  $A_{R1}T$  mode.

The 2D-IR spectrum of the  $N_3\text{-}A_3T_3$  complex is shown in Fig.6.5.(b). Each peak observed in the FT-IR spectrum gives rise to a negative feature (red) located on the 2D-IR spectrum diagonal. These are assigned to the respective  $\nu = 0 \rightarrow 1$  transitions, each with an accompanying, positive (blue),  $\nu = 1 \rightarrow 2$  peak shifted to lower probe frequency by the anharmonicity of the vibrational mode. Peaks located in the off-diagonal region of the 2D-IR spectrum primarily indicate the presence of coupling between diagonal vibrational modes, though a small contribution is expected from energy transfer due to the fast ( $\sim 650$  fs) vibrational relaxation of the base vibrational modes.<sup>347</sup> Off-diagonal peaks present in the 2D-IR spectra are separated into two broad subsets; those due to intra-base coupling between the  $AT_{2S}$ ,  $AT_{4S}$  and  $T_R$  modes that are primarily located on the thymine base (Fig.6.5.(b), crosses) and those due to inter-base coupling between the adenine-based  $A_{R1}T$  mode and the three thymine modes that is induced by Watson-Crick H-bonding (Fig.6.5.(b), circles).

A binding-induced 2D-IR difference spectrum illustrating the modifications to the DNA vibrational modes of the  $A_3T_3$  sequence upon the binding of  $N_3\text{-}bBI$  is shown in Fig.6.5.(c). Comparing the positions of the features in this binding-induced 2D-IR difference spectrum (Fig.6.3.(c)) with the assigned FT-IR spectra of GC and AT only DNA sequences (Fig.6.5.(d)) demonstrates that the interaction between the  $N_3\text{-}bBI$  and the  $A_3T_3$  duplex primarily affects the  $AT_{4S}$  and  $AT_{2S}$  vibrational modes. Additional, smaller, features in this difference spectrum are assigned as the  $T_R$  and  $A_{R1}T$  modes, which is broadly consistent with those observed in the FT-IR difference spectrum. The assignments of the binding-induced 2D-IR difference spectrum are supported by the appearance of distinct AT cross-peak patterns (crosses, Fig.6.5.(c)).

Considering the results of the FT-IR and 2D-IR experiments together the modifications upon formation of the  $N_3\text{-}A_3T_3$  complex (Fig.6.5.(a), blue) are almost identical to the pattern observed upon the formation of the H33258- $A_3T_3$  complex.<sup>325</sup> The only difference appears to be the overall intensity of the features in the difference spectrum which are noted to be more intense for the H33258- $A_3T_3$  complex (see pg. 114, chapter 5). This is consistent with the lower bound population in the  $N_3\text{-}A_3T_3$  system (23%) owing to the reduced solubility of  $N_3\text{-}bBI$  under the conditions studied.

Overall the shifts in the  $AT_{25}$  and  $AT_{45}$  modes of the  $A_3T_3$  sequence to higher frequencies upon the formation of the  $N_3-A_3T_3$  complex (Fig.6.5.(c)) is consistent with a decrease in strength of the hydrogen bonds formed to the  $T_2$  and  $T_4$  carbonyls. Within the A-tract sequence the minor groove contains an ordered spine of hydration and additional three-centred H-bonds between sequential base pairs leads to a region of highly order AT propeller twists. Upon binding the  $N_3$ -*b*BI ligand displaces a section of the spine of hydration and forms bifurcated H-bonds to the  $T_2$  carbonyls. These bifurcated H-bonds are weaker than the original H-bonds from the spine of hydration leading to a shift to higher frequency of a subset of the  $AT_{25}$  mode. Additionally the formation of the bifurcated ligand-duplex H-bonds results in the increase in the twist of a subset of the AT base pairs, lengthening the  $T_4$  W-C H-bonds for these base pairs. This leads to a decrease in the strength of these H-bonds which would be reflected in the spectral signature as a shift to higher frequency of a subset of the  $AT_{45}$  mode.<sup>325</sup>

In the previous study, on H33258 complexes, a model was developed to quantify the changes in the  $AT_{25}$  and  $AT_{45}$  modes upon the formation of the DNA complex. This model fitted the  $AT_{25}$  and  $AT_{45}$  peaks in the 2D-IR to 2D-Gaussians and modelled the shift of the diagonal and cross peaks associated with these modes to allow the population and shift of these modes to be extracted from the binding-induced difference spectrum.<sup>325</sup> The best agreement with the experimental 2D-IR difference spectra (Fig.6.5.(c)) was obtained by shifting a subset of the  $AT_{25}$  and  $AT_{45}$  peaks to higher frequency (Fig.A6.7). Crucially, this simulation recreated the off-diagonal structure of the difference spectrum as well as the diagonal peaks (arrows Figs.A6.7(c)&(d)). The results of the model are shown in Table 6.2; alongside the results of the binding-induced changes to the  $A_3T_3$  duplex upon the formation of the H33258:DNA complex.

**Table 6.2:** Changes to the IR response of A<sub>3</sub>T<sub>3</sub> DNA duplex due to N<sub>3</sub>-bBI and H33258 binding.

Binding-Induced Changes	Minor Groove Binder	
	N <sub>3</sub> -bBI	H33258 <sup>b</sup>
AT <sub>25</sub> Shift (cm <sup>-1</sup> )	5.5	6.4
AT <sub>25</sub> Shifted Subset Size (%)	<b>4.3 (18.8)</b>	<b>20.3</b>
Change in T <sub>2</sub> H-bond Strength (kJmol <sup>-1</sup> ) <sup>a</sup>	-2.4	-2.8
AT <sub>45</sub> Shift (cm <sup>-1</sup> )	8.4	8.7
AT <sub>45</sub> Shifted Subset Size (%)	<b>3.4 (14.6)</b>	<b>9.6</b>
Change in T <sub>4</sub> H-bond Strength (kJmol <sup>-1</sup> ) <sup>a</sup>	-3.7	-3.9

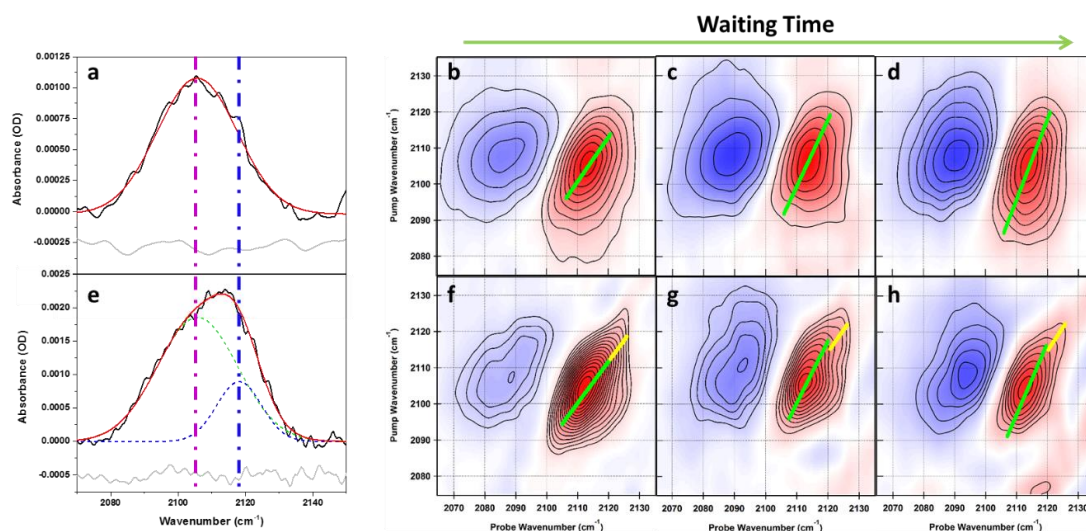
<sup>a</sup>Values calculated using the C=O bond energy (~743 kJmol<sup>-1</sup>)<sup>349</sup> via the methodology as outlined in ref.350 using equation ref.351. <sup>b</sup>Values for H33258 binding-induced differences from ref.325.

Based on this model the assignment of the peaks, observed in both the FT-IR and 2D-IR difference spectra upon the formation of the N<sub>3</sub>-A<sub>3</sub>T<sub>3</sub> complex, can be confirmed as resulting from changes to the A–T base modes of the A<sub>3</sub>T<sub>3</sub> duplex. The significantly smaller sizes of the shifted subsets extracted using the model for the N<sub>3</sub>-bBI ligand, is consistent with a bound population of 23%. Once this lower population is taken into account the sizes of the shifted subsets of the T<sub>4</sub> and T<sub>2</sub> carbonyls (Table 6.2, red) upon N<sub>3</sub>-bBI binding match those observed from the interactions in the H33258:A<sub>3</sub>T<sub>3</sub> complex. For the N<sub>3</sub>-A<sub>3</sub>T<sub>3</sub> complex the model determined that ~19% of the AT<sub>25</sub> mode of the A<sub>3</sub>T<sub>3</sub> (A-tract) sequence is shifted ~6 cm<sup>-1</sup> to higher frequency while ~15% of the AT<sub>45</sub> mode shifts by ~8 cm<sup>-1</sup>. Additionally the absence of any GC features, on the diagonal or in the off diagonal region, in the 2D-IR difference spectrum show that the interactions between N<sub>3</sub>-bBI and the DNA duplex are localized to the A-tract. The similarity between the size of the shifted portions of the AT<sub>25</sub> and AT<sub>45</sub> vibrational modes when H33258 and N<sub>3</sub>-bBI bind, are consistent with N<sub>3</sub>-bBI participating in a similar induced fit mechanism with the A<sub>3</sub>T<sub>3</sub> sequence. This is also found to be entirely consistent with the results of the NMR spectroscopy of the N<sub>3</sub>-A<sub>3</sub>T<sub>3</sub> system.

#### ***Infrared spectroscopy – azido label***

The asymmetric azide stretch of the unbound N<sub>3</sub>-bBI ligand was found to consist of a single peak centred at 2106 cm<sup>-1</sup> (Fig.6.6.(a)). Upon binding to the A<sub>3</sub>T<sub>3</sub> sequence a high frequency component centred at 2118 cm<sup>-1</sup> (Fig.6.6.(e), blue) was observed. Fitting the absorbance

band to Gaussian line-shape functions shows that this new high frequency component constitutes 21% of the azide absorbance band.



**Figure 6.6:** a) FTIR spectra of the asymmetric azide stretch of unbound  $N_3$ -bBI ligand (black) at 20 °C, overall fit shown in red with residuals (grey, offset by  $-2.5 \times 10^{-4}$  OD), 2D-IR spectra of the asymmetric azide stretch of the unbound  $N_3$ -bBI ligand at 20 °C recorded at waiting times of b) 300 fs, c) 500 fs and d) 700 fs, central line slopes shown in green. e) FTIR spectra of the asymmetric azide stretch of the  $N_3$ - $A_3T_3$  complex (black) at 20 °C, showing an unbound contribution (green) and a bound contribution (blue), overall fit shown in red with residuals (grey, offset by  $-5 \times 10^{-4}$  OD), 2D-IR spectra of the asymmetric azide stretch of the  $N_3$ - $A_3T_3$  complex at 20 °C recorded at waiting times of f) 300 fs, g) 500 fs and h) 700 fs, central line slopes shown in green and yellow.

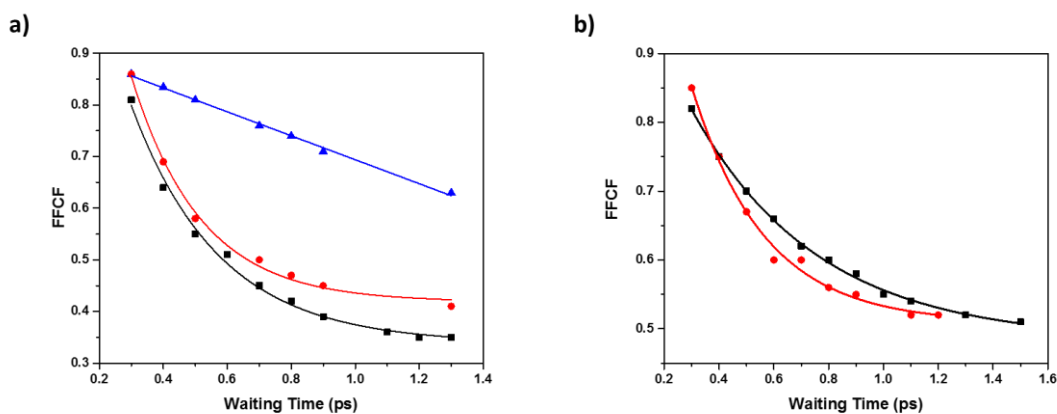
As seen for the DNA 2D-IR spectra, each mode observed in the FT-IR spectrum contributes a peak pair, with a negative (red,  $\nu = 0 \rightarrow 1$ ) peak located on the spectrum diagonal and a positive (blue,  $\nu = 1 \rightarrow 2$ ) peak shifted to lower probe frequency by the anharmonicity of the bonding potential. The 2D-IR spectra, of the asymmetric azide stretch, of the unbound  $N_3$ -bBI ligand recorded at waiting times of 300 fs, 500 fs and 700 fs are presented in Figs.6.6.(b),(c)&(d), respectively. The peak corresponding to the  $\nu = 0 \rightarrow 1$  transition (red) is initially strongly elongated along the spectrum diagonal, indicating a contribution from inhomogeneous broadening. As the waiting time increases the peak exhibits the effects of spectral diffusion (chapter 1), becoming less diagonally elongated. This change in shape can be quantified via the Centre Line Slope (Figs.6.6.(b),(c)&(d), green line)<sup>352,353</sup> reflecting the dynamics associated with the frequency-frequency correlation function (FFCF) of the asymmetric stretch of the azide moiety. These measurements allow the azide to act as a reporter of the dynamics of its local molecular environment. In the case of the unbound  $N_3$ -

*b*BI ligand the FFCF (Fig.6.7.(a), black), extracted via the Centre Line Slope methodology, yielded a spectral diffusion rate of  $247(\pm 0.5)$  fs.

Upon binding to the  $A_3T_3$  sequence, the additional component on the high frequency side of the asymmetric azide stretch of the  $N_3$ -*b*BI ligand is also visible in the 2D-IR spectrum (Figs.6.6.(f),(g)&(h)). This is consistent with the FT-IR data and it is noted that there are no off-diagonal component is observed linking them.

Obtaining 2D-IR spectra of the  $N_3$ - $A_3T_3$  complex at a range of waiting times shows that both components undergo spectral diffusion with different dynamics. The central line slopes for the two components that contribute to the azide band of the  $N_3$ - $A_3T_3$  complex, shown in Figs.6.6.(f),(g)&(h), demonstrate that the unbound (green) and bound (yellow) components of the band diffuse at rates of  $218(\pm 0.4)$  fs (Fig.6.7.(a), red) and  $\sim 3.5$  ps (Fig.6.7.(a), blue) respectively, causing a discontinuity in the central line to appear at longer waiting times.

Tracking these changes in the central line slopes observed in the 2D-IR spectra of the unbound  $N_3$ -*b*BI and the  $N_3$ - $A_3T_3$  complex allows the diffusion dynamics of both molecular environments to be determined. These diffusion dynamics of the asymmetric azide stretch with and without the  $A_3T_3$  sequence, at both 20 °C (Fig.6.6) and 80 °C (SI, Fig.A6.8), are presented in Fig.6.7.



**Figure 6.7:** The spectral diffusion dynamics of the asymmetric azide stretch of the  $N_3$ -*b*BI without (black) and the unbound (red) and bound (blue)  $N_3$ -*b*BI with the  $A_3T_3$  sequence at a) 20 °C and b) 80 °C.

The slow spectral diffusion of the high frequency component is quoted as an approximate value as the shorter ( $\sim 1$  ps) lifetime of the azide probe means it is ill-defined. At 80 °C, after the  $N_3$ - $A_3T_3$  complex has undergone dissociation, the spectral diffusion dynamics in the

sample without (Fig.6.7.(b), black) and with (Fig.6.7.(b), red) the *ss*-DNA were found to be 459( $\pm$ 0.5) fs and 480( $\pm$ 0.6) fs respectively.

A computational study, on the impact of different H-bonding interactions on the peak position of the azide mode, predicted all H-bonds of  $\sigma$ -symmetry induce shifts to higher frequency while those of  $\pi$ -symmetry induce shifts to lower frequency.<sup>334</sup> Therefore the exact peak position is determined by the geometry of the H-bonding network surrounding the azide. This predicted sensitivity is reflected in the 5.6  $\text{cm}^{-1}$  shift to lower frequency of the peak, caused by the changes in the geometry of the local H-bonding network upon the aggregation of the  $\text{N}_3$ -*b*BI ligand (SI, Fig.A6.9).

The high frequency component, in the FT-IR and 2D-IR spectra, upon binding represents a 21% sub-set of all the  $\text{N}_3$ -*b*BI molecules present. Additionally the absence of cross-peaks between these two sub-ensembles shows that there is no coupling or exchange between these two groups, therefore the high-frequency is assigned to the bound portion of the  $\text{N}_3$ -*b*BI ligand present in this system (Fig.6.4.(b), red). The position of the bound component is consistent with the transfer of the azide moiety into the minor groove leading to the loss of its original hydration shell. Once located within the minor groove hydrogen bonds can only form to the terminal nitrogen atom of the azide moiety, leading to a shift to higher frequency as this atom is predicted to be more strongly affected by  $\sigma$ -H-bonds than  $\pi$ -H-bonds.<sup>334</sup> This is further supported by the  $\sim$ 12  $\text{cm}^{-1}$  decrease in the width of the new component when compared to the width of the unbound azide mode, suggesting the  $\text{N}_3$ -*b*BI molecules contributing to this component experience a reduced number of different possible H-bonding networks.

The spectral diffusion dynamics of the probe in water (Fig.6.7.(a), black) matches those of the unbound segment of the azide band observed for the  $\text{N}_3$ - $\text{A}_3\text{T}_3$  complex (Fig.6.7.(a), red), further supporting this assignment, while those of the additional high frequency shoulder are significantly slower (Fig.6.7.(a), blue).<sup>316</sup> The change in the centre line slope observed for the bound contribution to the azide band of the  $\text{N}_3$ - $\text{A}_3\text{T}_3$  complex appears to decrease linearly as the waiting time increases. This is consistent with dynamics that are slow compared to the lifetime of the azide probe ( $\sim$ 1 ps), which is consistent with the  $\sim$ 3.5 ps diffusion dynamics extracted from this data. It is noted that the dynamics of this portion of bound  $\text{N}_3$ -*b*BI is remarkably slow compared to the  $\sim$ 0.2 ps dynamics in the free solution. It is thought that the probe does not directly probe the DNA duplex, as it is known to show dynamics on a  $\sim$ 20 ps



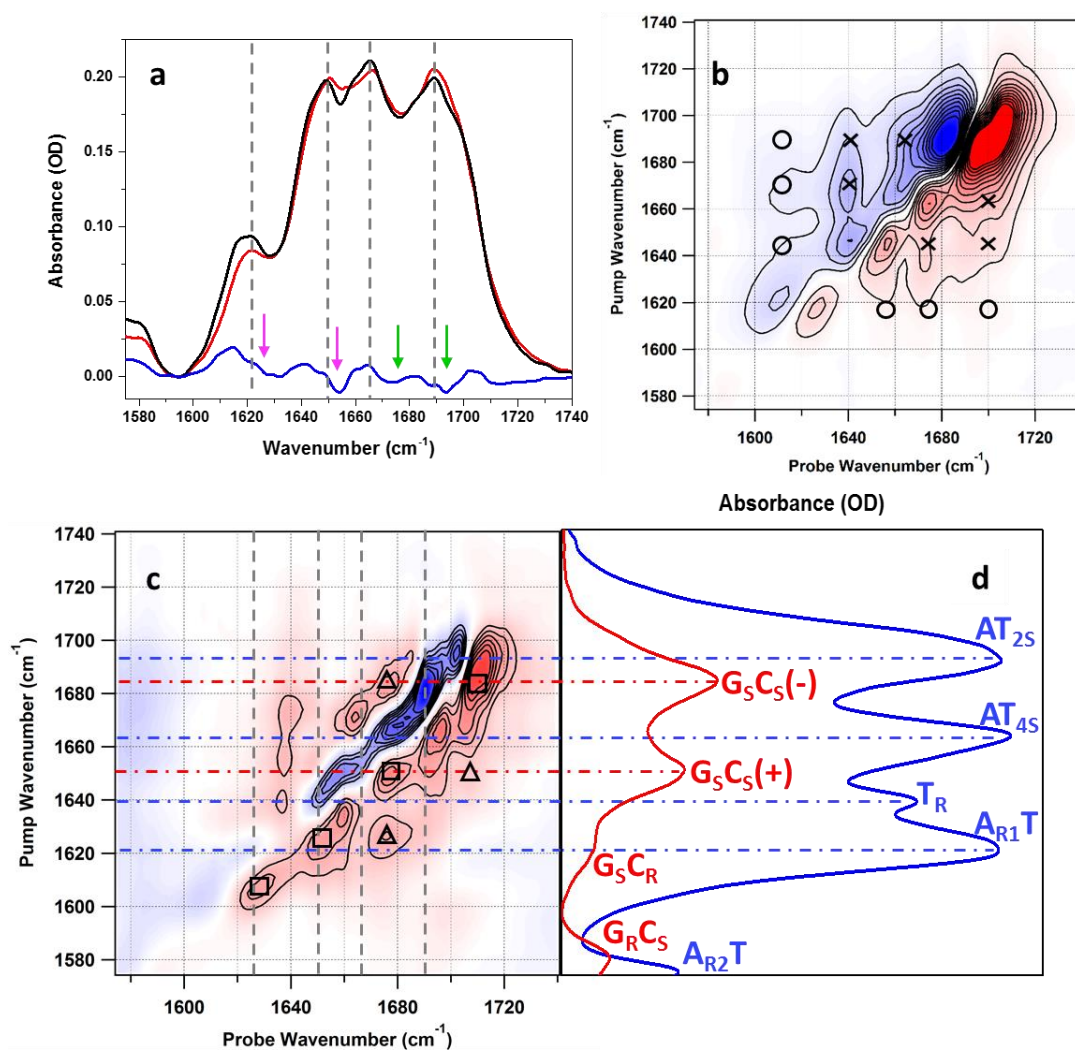
timescale.<sup>316,317</sup> The unique hydrogen bonding patterns suggest that the probe is in contact with water molecules so we conclude that the probe is partially solvent exposed. The strongly restricted dynamics demonstrate that the water probed by the azide is constrained by the minor groove, consistent with the label probing the remaining spine of hydration.

#### Alternating A–T sequence binding

#### **Infrared spectroscopy – DNA Base Modes**

The indications from the UV-visible study are that the interactions between the N<sub>3</sub>-bBI ligand and the (AT)<sub>3</sub> sequence leads to a ~3.5-fold lower stabilisation of T<sub>m</sub> when compared to the N<sub>3</sub>-A<sub>3</sub>T<sub>3</sub> complex. For H33258, the interactions between the ligand and the (AT)<sub>3</sub> sequence follow a rigid body model, still consistent with an AT-only interaction, but this mechanism only leads to a 1.5-fold lower stabilisation of T<sub>m</sub> (16 °C versus 24 °C).<sup>325</sup> This comparison suggests that the binding of N<sub>3</sub>-bBI to the alternating sequence is significantly less than the rigid body interaction seen for H33258. The molecular interactions underpinning the N<sub>3</sub>-(AT)<sub>3</sub> complex can be investigated via the DNA base modes. The FT-IR spectra of the (AT)<sub>3</sub> sequence and its complex with N<sub>3</sub>-bBI (N<sub>3</sub>-(AT)<sub>3</sub>) are presented in Fig.6.8.(a). As seen in the IR spectra of the A-tract sequence, the spectra of the alternating sequence contains four distinct peaks, indicated by dashed grey lines Fig.6.8.(a). The assignment of these peaks are summarised in Table 6.1.





**Figure 6.8:** a) FT-IR spectra of  $(AT)_3$  DNA with (black) and without (red)  $N_3$ -bBI. Blue spectrum shows the binding-induced difference FT-IR spectrum (complex-free sequence). b) 2D-IR spectrum of  $N_3$ - $(AT)_3$  complex. Crosses mark off-diagonal peaks assigned to coupling of modes primarily located on T-base. Circles show off-diagonal peaks assigned to coupling of modes on A and T bases induced by W-C base pairing. c) Ligand binding induced 2D-IR difference spectrum. d) FT-IR spectra of GC and AT-only DNA sequences to show the relative (weighted per base) magnitudes of peaks and peak positions to aid assignment of the IR spectrum of the  $(AT)_3$  sequence.

The IR difference spectrum associated with the formation of the  $N_3$ - $(AT)_3$  complex (Fig.6.8.(a), blue) is different from the difference spectrum associated with the formation of the  $N_3$ - $A_3T_3$  complex (Fig.6.5.(a), blue). While it contains some features reminiscent of those seen in the binding-induced difference spectrum of the  $N_3$ - $A_3T_3$  complex (Fig.6.5.(a), blue) associated with shifts to higher frequency of the  $AT_{2S}$  and  $AT_{4S}$  modes (green arrows), there are additional features, near  $1650\text{ cm}^{-1}$  and  $1622\text{ cm}^{-1}$  which are more consistent with modifications to the GC modes upon binding (pink arrows).

The binding-induced 2D-IR difference spectrum highlighting the changes to the DNA base modes of the (AT)<sub>3</sub> sequence upon binding is found to contain a multitude of separate features (Fig.6.8.(c)). The number of changes and their positions when compared to the assigned linear IR spectrum of DNA oligonucleotides containing exclusively GC and AT base pairs (Fig.6.8.(d)) demonstrates that upon binding to the (AT)<sub>3</sub> sequence N<sub>3</sub>-bBI causes perturbations to both AT and GC base vibrational modes. These additional GC diagonal and cross-peaks observed in the binding-induced difference spectrum are highlighted by squares and triangles respectively (Fig.6.8.(c)). This reflects both the complexity and congestion of the FT-IR binding-induced difference spectrum (Fig.6.8.(a), blue). From the 2D-IR difference spectrum it is observed that the AT<sub>2S</sub>, AT<sub>4S</sub> and G<sub>S</sub>C<sub>S</sub>(+) vibrational modes undergo the largest alterations upon binding. Additional, smaller, features associated with perturbations to the T<sub>R</sub>, A<sub>R1</sub>T and G<sub>R</sub>C<sub>S</sub> vibrational modes are also observed.

The congestion of this spectral window means that it can be complicated to verify changes in all of these modes accurately. The most dominant features observed in the 2D-IR difference spectrum are noted to be located in the 1645–1710 cm<sup>-1</sup> region of the DNA base modes. To verify the contribution of AT and GC modes to the changes seen in this region of the DNA base mode spectrum, we employ a model 2D-IR spectrum (SI, Fig.A6.10) constructed from 2D-Gaussian line-shapes, which simulate the coupled peaks in the 2D-IR spectra of the DNA duplexes (SI, Fig.A6.10.(a)&(b)). The best agreement with the experimental 2D-IR binding-induced difference spectrum, observed for the N<sub>3</sub>-(AT)<sub>3</sub> complex, was obtained by shifting a subset of the AT<sub>2S</sub>, G<sub>S</sub>C<sub>S</sub>(-), AT<sub>4S</sub> and G<sub>S</sub>C<sub>S</sub>(+) modes to higher frequency (SI, Fig.A6.10.(c)).

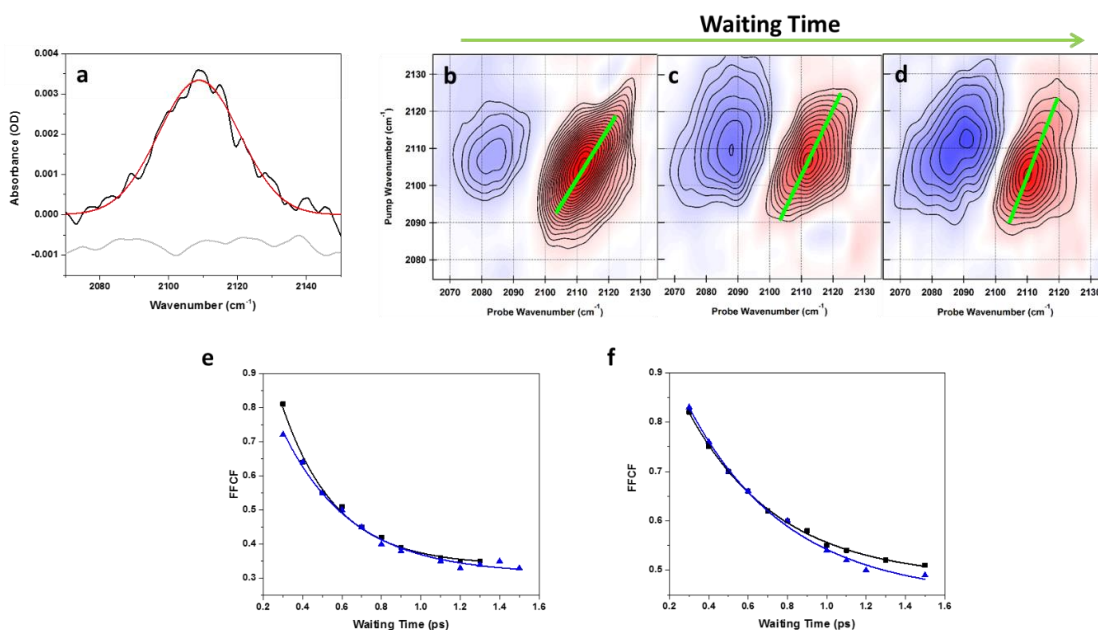
Despite the model not exactly reproducing the fine details of the off-diagonal region of the experimentally observed 2D-IR difference spectrum, this model does reproduce the overall appearance of the changes observed (SI, Figs.A6.10.(c)&(d)). Crucially it is noted that this is only achieved if the same portion of the G<sub>S</sub>C<sub>S</sub>(+) and G<sub>S</sub>C<sub>S</sub>(-) vibrational modes undergo similar shifts to higher frequency. The complexity of the spectral signature of the interactions between the N<sub>3</sub>-bBI ligand and the (AT)<sub>3</sub> sequence means that it is difficult to use this type of model to accurately quantify the underlying modifications. It is thought that the appearance of the smaller T<sub>R</sub> and A<sub>R1</sub>T features are the result of the shifts of the two T carbonyls as these moieties contribute to both of these ring modes. Finally the small feature assigned as the

$G_{RC_5}$  mode is thought to be a result of the shift of the C carbonyl upon the formation of the  $N_3$ -(AT)<sub>3</sub> complex.

Overall these changes can be used to understand the molecular determinants of the interactions within the  $N_3$ -(AT)<sub>3</sub> complex. Particularly noteworthy is the very small (~2%) fraction of the AT<sub>4S</sub> peak that is shifted, this is thought to be consistent with the DNA duplex itself behaving as a rigid body, as seen in a previous study of the complex formed between the (AT)<sub>3</sub> sequence and H33258.<sup>325</sup> The shift of the AT<sub>2S</sub> mode to higher frequency indicates that upon binding the strength of the H-bonds formed to the T<sub>2</sub> carbonyl, known to point directly into the minor groove, decreases. This is consistent with the loss of the spine of hydration upon binding and the formation of hydrogen bonds between these carbonyls and the ligand. Finally the shift of a portion of the  $G_5C_5(+)$  and  $G_5C_5(-)$  modes are thought to be consistent with a decrease in the strength of the hydrogen bonds formed to the C carbonyl, which participates in the GC W-C H-bond located on the floor of the minor groove. This is thought to suggest that upon binding hydration water are also expelled from one of the GC ends of the (AT)<sub>3</sub> sequence. The involvement of the C carbonyl in both the  $G_5C_5(+)$  and  $G_5C_5(-)$  modes explains why both of these modes are perturbed by binding. Finally the larger shift observed for the GC modes when compared to the AT<sub>2S</sub> mode upon binding is thought to indicate that after binding the ligand does not form hydrogen bonds to the GC base pairs. Together these changes are thought to indicate that the formation of the complex between the  $N_3$ -bBI and the (AT)<sub>3</sub> sequence largely follows a non-specific interaction mechanism where the DNA behaves as a rigid body.

#### ***Infrared spectroscopy – azido label***

Upon the formation of the  $N_3$ -(AT)<sub>3</sub> complex the asymmetric azide stretch (Fig.6.9.(a)) is noted to only undergo a subtle 3 cm<sup>-1</sup> shift to higher frequency. This much more subtle change is significantly different from the appearance of a high-frequency shoulder observed upon binding to the A-tract sequence. This reflects the differences in the binding interactions between the  $N_3$ -bBI ligand and these two different sequences.



**Figure 6.9:** a) FTIR spectra of the asymmetric azide stretch of the  $N_3$ -(AT) $_3$  complex (black) at 20 °C, overall fit shown in red with residuals (grey, offset by  $-5 \times 10^{-4}$  OD), 2D-IR spectra of the asymmetric azide stretch of the  $N_3$ -(AT) $_3$  complex at 20 °C recorded at waiting times of b) 300 fs, c) 500 fs and d) 700 fs, central line slopes shown in green. The spectral diffusion dynamics of the asymmetric azide stretch of the  $N_3$ -bBI ligand without (black) and with (blue) the (AT) $_3$  sequence at e) 20 °C and f) 80 °C.

The 2D-IR spectra, of the asymmetric azide stretch, for the  $N_3$ -(AT) $_3$  complex, at different waiting times, are shown in Figs.6.9.(b),(c)&(d). As the waiting time increases the peak, corresponding to the  $\nu = 0 \rightarrow 1$  transition (red), undergoes spectral diffusion resulting in the Centre Line Slope (Figs.6.9.(b),(c)&(d), green lines) becoming more vertical with increasing  $T_w$ , reflecting the dynamics associated with the FFCF of the azido moiety within the  $N_3$ -(AT) $_3$  complex. Results of the Centre Line Slope analysis shows that the spectral diffusion dynamics of the asymmetric azide stretch of the  $N_3$ -(AT) $_3$  complex at 20 °C is  $351(\pm 0.6)$  fs (Fig.6.9.(e), blue). Upon heating to 80 °C it is observed that the spectral diffusion dynamics of the azido-label increase to  $507(\pm 0.5)$  fs (Fig.6.9.(e), blue). This demonstrates that at 80 °C the ligand is fully dissociated from the duplex and its behaviour becomes similar to the unbound ligand.

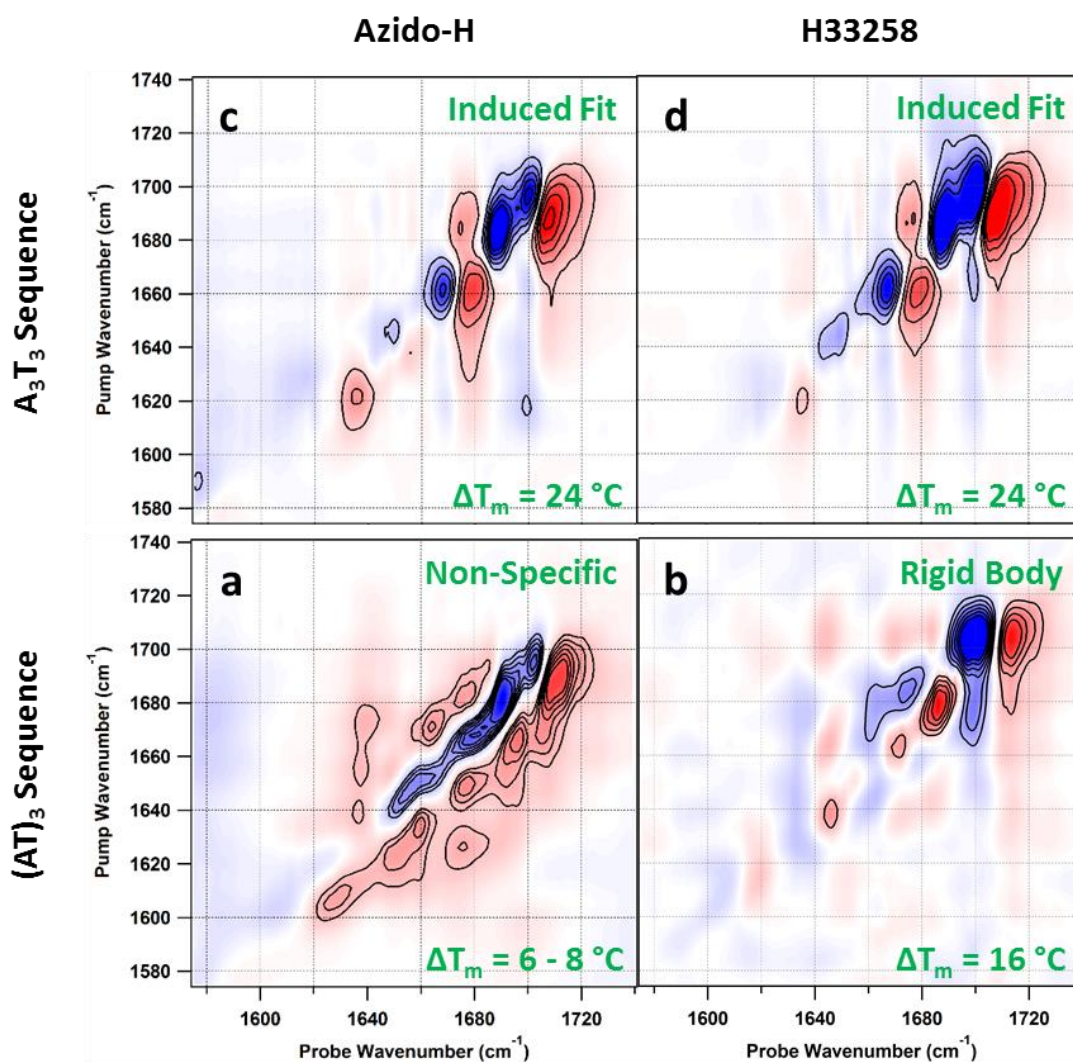
The  $3 \text{ cm}^{-1}$  shift to higher frequency in the peak position, observed in the IR spectrum upon binding, is consistent with the dissolution of the  $N_3$ -bBI aggregate known to form at pD7 (SI, Fig.A6.9). Together with the changes seen for the azide mode of the  $N_3$ -A $_3$ T $_3$  complex (Fig.6.6.(e)), this seems to suggest that upon binding to the (AT) $_3$  sequence the azide does not reside in the DNA minor groove in the same manner.

It is noted that the spectral diffusion dynamics of the azide mode upon the formation of the  $N_3$ -(AT)<sub>3</sub> complex slows by ~100 fs. In a previous study<sup>67</sup> on the spectral diffusion of DNA NH modes, it was reported that the diffusion dynamics slowed upon a decrease in the humidity of the sample showing these dynamics are linked to the number of water molecules surrounding the mode. Here, the increase in the diffusion dynamics upon binding suggests that the azide moiety does not reside in the minor groove but in an area with a reduced number of water molecules compared to the bulk solution. It is noted that any proposed binding interaction between  $N_3$ -bBI and the (AT)<sub>3</sub> sequence must satisfy the constraints imposed by both results of the changes in the natural IR DNA probes and the non-natural azide probe upon the formation of the  $N_3$ -(AT)<sub>3</sub> complex. Thus we propose that the weaker non-specific binding interaction between  $N_3$ -bBI and the (AT)<sub>3</sub> sequence leads to the azide moiety residing outside the minor groove of the duplex. This would seem to explain the spectral diffusion dynamics of the azide mode in the  $N_3$ -(AT)<sub>3</sub> complex which would be probing the slightly more structured water molecules located at the interface between the aqueous solution and the (AT)<sub>3</sub> duplex.

#### *Molecular Determinants of Sequence Selectivity*

To date the investigations of minor groove binding has shown that H33258 and  $N_3$ -bBI both bind to A-tract DNA via an induced fit mechanism, however these ligands respond very differently to an alternating AT sequence. H33258 binds with a large stabilisation of  $T_m$  (16 °C, versus 24 °C for A-tract binding) and a lower binding affinity leading to descriptions of 'sub-optimal' binding. Spectroscopically this was shown to arise from a binding mode more reminiscent of a rigid body interaction, lacking the induced fit component of the A-tract binding.  $N_3$ -bBI shows a marked drop in duplex stabilisation (7 °C) upon binding to the alternating sequence and does not demonstrate spectroscopically similar signature to H33258 suggesting a different, presumably poorer quality of binding. The spectral signatures of these interactions are summarized in figure 6.10.





**Figure 6.10:** Spectral signatures of molecular interactions within the a)  $N_3$ - $A_3T_3$ , b)  $H$ - $A_3T_3$ ,<sup>325</sup> c)  $N_3$ - $(AT)_3$  and d)  $H$ - $(AT)_3$ <sup>325</sup> complexes with the identified mechanisms and stabilisations shown in green.

The binding-induced 2D-IR difference spectra for the interactions between the A-tract ( $A_3T_3$ ) sequence and the  $N_3$ - $b$ BI and H33258 ligands are shown in Figs.6.10.(a)&(b). In addition to the similarity in the stabilisation (24 °C) of the sequence by the ligands it is noted that the spectral signatures of these interactions are remarkably similar. In contrast to this the binding-induced 2D-IR difference spectra seen for the interactions formed between the alternating ( $(AT)_3$ ) sequence and both of these *bis*-benzimidazole ligands are significantly different (Figs.6.10.(c)&(d)) leading to a 9 °C lower  $T_m$  stabilisation of the  $(AT)_3$  sequence upon the formation of the  $N_3$ - $(AT)_3$  complex. This difference is primarily due to contribution of GC modes to the difference spectrum upon the formation of the  $N_3$ - $(AT)_3$  complex indicating that the interactions between the  $N_3$ - $b$ BI ligand and the  $(AT)_3$  duplex are not localised to the AT region, as seen in for the H33258: $(AT)_3$  complex.

The large degree of similarity in the spectral signatures (Table 6.2) of the binding interactions seen for the  $N_3$ -*b*BI and H33258<sup>325</sup> ligands and the A-tract sequence shows that the formation of both of these DNA complexes follows the induced fit mechanism. This indicates that the effect of the introduction of the hydrophilic azide probe into the *bis*-benzimidazole ligand ( $N_3$ -*b*BI) on these interactions is negligible. However the differences seen in the spectral signatures of the binding interactions between these two ligands and the alternating sequence suggest that in the case of the H33258 ligand the complex follows a rigid body mechanism, whereas the  $N_3$ -*b*BI ligand participates in a non-specific interaction with this sequence. Based on this, in the case of the alternating sequence the addition of the hydrophilic azido probe results in a loss of the sequences-specificity of the *bis*-benzimidazole ligand. Together the changes to the spectral signatures due to the addition of the azido-label to the *bis*-benzimidazole ligand ( $N_3$ -*b*BI) highlights key differences between the induced fit and rigid body interactions. The hydrophilic nature of the azide moiety is thought to result in an additional unfavourable contribution to the enthalpic change upon binding into the hydrophobic minor groove. In the case of binding to the A-tract sequence this unfavourable contribution, due to the azide, is overcome by the favourable enthalpic contribution due to the enhanced interactions provided by the induced fit mechanism. In contrast to this the weaker rigid body interaction underlying the formation of complexes between the *bis*-benzimidazole ligands and the alternating sequence, does not lead to an enhancement in the intermolecular interactions between the ligand and the duplex, leading to a smaller favourable enthalpic contribution. It is thought that this smaller enthalpic contribution from the rigid body interaction is not sufficient to overcome the destabilizing effect of the azido probe causing the ligand to interact in a non-specific manner and lose its sequence-selectivity.

## 6.4 Conclusion

In conclusion, we have designed a new ligand ( $N_3$ -*b*BI) that is found to be more selective for A-tract DNA than H33258. Characterisation of the binding to the A-tract and alternating sequences shows that the stabilisation changes from 24 °C to 7 °C (versus 24 °C to 16 °C for H33258). NMR and 2D-IR spectroscopy show that A-tract binding of the ligand is similar to the induced fit model derived for H33258 while binding to the alternating sequence is non-specific for the  $N_3$ -*b*BI ligand, involving both AT and GC base pairs. The azide probe shows mobile water near the azide except when buried in the minor groove, as seen in the  $N_3$ -A<sub>3</sub>T<sub>3</sub> complex. Within the minor groove the azide encounters a unique H-bonding environment with exceptionally slow dynamics, suggesting the probe is exploring the remaining spine of

hydration within the A<sub>3</sub>T<sub>3</sub> minor groove. Overall we show that the impact of the azide moiety is not enough to overcome the enthalpic benefits of induced fit binding but the rigid body interaction of *bis*-benzimidazole binding to alternating sequences is significantly perturbed by the azide.

## 6.5 Methods

### 6.5.1 Materials

The lyophilised, salt-free DNA oligonucleotides were obtained from Eurofins; D<sub>2</sub>O, DMSO, monobasic and dibasic sodium phosphate were obtained from Sigma-Aldrich. The Hoechst 33258 azide derivative (N3-bHoechst) was synthesized and purified by John J. May working in the Burley Group at Strathclyde University. All chemicals were used without further purification. All the samples used in the experiments were prepared in pD 7 phosphate buffer solution. All of the DNA oligonucleotides were dissolved to produce 10 mM stocks and a 100 mM stock of N3-bHoechst was prepared using DMSO. Using these stocks all samples were prepared using pD7 phosphate buffer solution to a final duplex: N3-bHoechst ratio of 1:1 and annealed at 90 °C for 10 minutes.

### 6.5.2 UV-visible Spectroscopy

The samples were held in a demountable Harrick cell utilising CaF<sub>2</sub> windows and a 50 μm polytetrafluoroethylene spacer for all the IR experiments. FTIR experiments were carried out using a Perkin-Elmer Lambda 25 at a resolution of 1 nm with sample concentrations of 2.5 mM (A<sub>3</sub>T<sub>3</sub> duplex/aH-A<sub>3</sub>T<sub>3</sub> complex) or 5 mM ((AT)<sub>3</sub> duplex/aH-(AT)<sub>3</sub> complex). These measurements were repeated at 1 μM (samples prepared via serial dilution to ensure accuracy) to provide duplex melting temperature data at this concentration to compare with the fluorescence measurements.

### 6.5.3 Fluorescence Spectroscopy

The 1 μM sample was diluted from the original samples (concentrations of 2.5 mM (A<sub>3</sub>T<sub>3</sub> duplex/aH-A<sub>3</sub>T<sub>3</sub> complex) or 5 mM ((AT)<sub>3</sub> duplex/aH-(AT)<sub>3</sub> complex)) via serial dilution to ensure accuracy. The samples were held in a quartz cuvette with a path length of 1 cm. The fluorescence experiments were carried out using a Horiba Fluorolog2 at a resolution of 1 nm. The samples were excited at 365 nm and the emitted fluorescence was recorded from 380 nm – 600 nm.

### 6.5.4 2D-IR and FT-IR Spectroscopy

The samples were held in a demountable Harrick cell utilising CaF<sub>2</sub> windows and a 50 μm polytetrafluoroethylene spacer for all the IR experiments. FTIR experiments were carried out



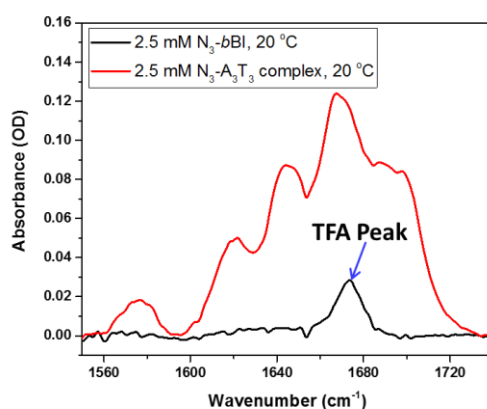
using a Bruker Vertex 70 at a resolution of  $1\text{ cm}^{-1}$  with sample concentrations of 2.5 mM ( $A_3T_3$  duplex/aH- $A_3T_3$  complex) or 5 mM ( $(AT)_3$  duplex/aH- $(AT)_3$  complex). 2D-IR spectra, of the DNA base vibrations, were collected using the ULTRA FT-2D-IR spectrometer.<sup>48,49</sup> The IR pulses used had a temporal duration of  $\sim 100\text{fs}$ ; a center frequency of  $1650\text{ cm}^{-1}$  and a bandwidth of  $\sim 300\text{ cm}^{-1}$ , at a repetition rate of 10 kHz. FT-2D-IR measurements were carried out at concentrations of 1.25 mM ( $A_3T_3$  duplex/*azido*H- $A_3T_3$  complex) or 2.5 mM ( $(AT)_3$  duplex/*azido*H- $(AT)_3$  complex). 2D-IR spectra of the asymmetric azide stretch were collected using the LIFETIME FT-2D-IR spectrometer,<sup>354,355</sup> as previously described.<sup>356</sup> Briefly, the IR pulses used has a temporal duration of  $\sim 300\text{fs}$ ; a center frequency of  $2100\text{ cm}^{-1}$ , at a repetition rate of 100 kHz. An IR pulse shaper (Phasetch) was utilized to generate the two collinear pump pulses separated by a variable time delay,  $\tau$ . The waiting time,  $T_w$ , between the pump and probe pulses was then set using an optical delay after the pulse shaper. The 2D-IR data sets were then measured for a range of  $T_w$  values by scanning  $\tau$  at each waiting time. Spectra were then generated by performing a Fourier transform along  $\tau$ . Phase cycling was also employed in these measurements to suppress the scatter originating from the samples. FT-2D-IR measurements of the azide mode were carried out at a concentration of 5 mM for both the unbound  $N_3$ -*b*BI and the two  $N_3$ -*b*BI:DNA complexes.

#### 6.5.5 Gaussian09 Calculations

All of the *ab initio* calculations presented in this chapter were carried out using the Gaussian09 software package. Geometry optimisations, frequency calculations and the potential energy surface were explored for the H33258 and  $N_3$ -*b*BI ligands. All of these calculations were carried out using the 6-311G(d,p) basis set and implementing the hybrid-DFT B3LYP functional.<sup>154</sup>

## 6.6 Addendum

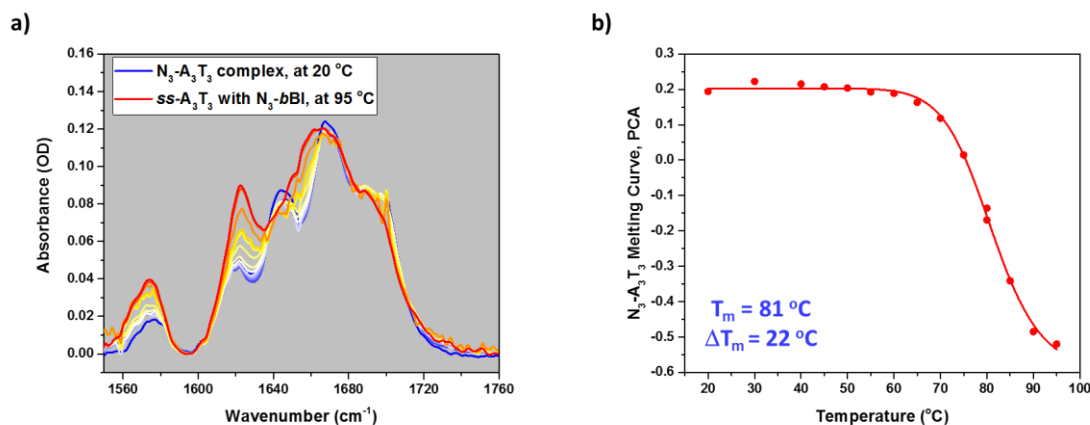
At the date of submission, the 28<sup>th</sup> of August 2017, the specially synthesised *de novo* N<sub>3</sub>-bBI ligand explored in this chapter was thought to be highly pure. However subsequent experimental work carried out for the preparation of a paper, in the Burley group at Strathclyde, has revealed that the purity of the ligand explored in the previous experimental chapter was in fact only 75 - 80% pure. However, it is thought that the impurity in the original sample of the N<sub>3</sub>-bBI binder did not bind to the DNA duplexes and so it is thought that the presence of this impurity should not alter the results in the previous chapter associated with the bound ligands. In order to address this a high purity sample of the N<sub>3</sub>-bBI ligand was synthesised and the key FT-IR measurements for the N<sub>3</sub>-A<sub>3</sub>T<sub>3</sub> complex were repeated to evaluate the effects of the presence of the impurity on the spectroscopy. The amide I region of the FT-IR spectra of the purified N<sub>3</sub>-bBI (black) and its N<sub>3</sub>-A<sub>3</sub>T<sub>3</sub> complex (red) recorded at 20 °C, are shown in Fig 6.11.



**Figure 6.11:** FT-IR spectra of the amide I region of the uncomplexed N<sub>3</sub>-bBI ligand (black) and its N<sub>3</sub>-A<sub>3</sub>T<sub>3</sub> complex (red) recorded at 20 °C.

The purified version of the N<sub>3</sub>-bBI ligand used in these measurements was known to be a trifluoroacetic (TFA) acid salt, this is supported by the appearance of a sharp TFA peak in the amide I region of the FT-IR spectrum of the unbound ligand (black, Fig.6.11). Beyond the appearance of this TFA peak the absorbance of the purified N<sub>3</sub>-bBI ligand is still very small in this region (3-5%), fully consistent with results in the previous experimental chapter. Unfortunately, the presence of the TFA in the sample means that upon binding the spectrum of the appearance of the DNA modes (red, Fig.6.11) is altered by this mode and so it is not possible to generate a binding induced difference spectrum for the N<sub>3</sub>-A<sub>3</sub>T<sub>3</sub> complex form using the purified form of the N<sub>3</sub>-bBI ligand. Therefore, the thermal denaturation of the N<sub>3</sub>-

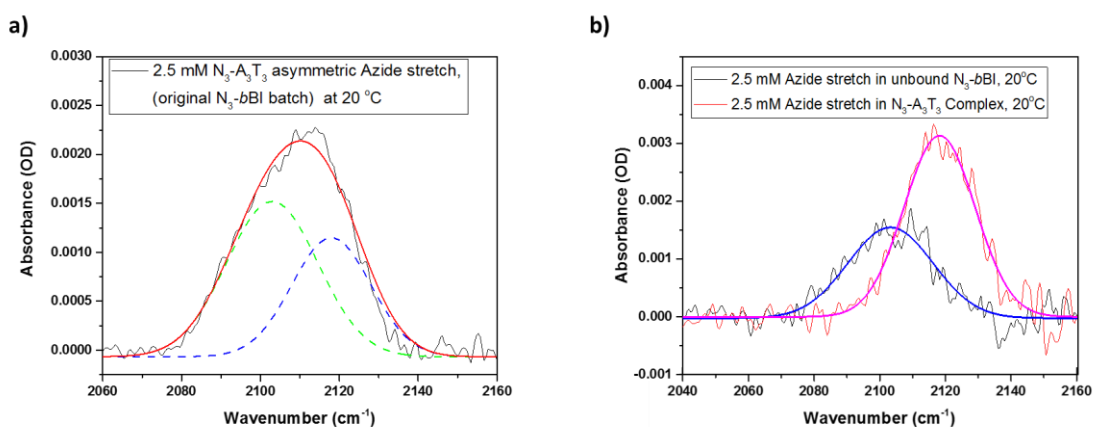
$A_3T_3$  complex was investigated to determine the effect of the binding interactions between the duplex and the ligand on the melting temperature. The FT-IR spectra of the DNA base modes and extracted melting curve of the complex are shown in Figs. Add.6.12.(a)&(b) respectively.



**Figure 6.12:** FT-IR spectra of the a)  $N_3-A_3T_3$  complex at 5 °C intervals between 20 – 95 °C and b) the melting curve of the  $N_3-A_3T_3$  complex (extracted via PCA analysis).

As the temperature of the  $N_3-A_3T_3$  complex is increased the FT-IR spectra of the DNA base modes (Fig.6.12.(a)) undergo a number of different changes. The most prominent change in the spectrum is an overall reduction of the number of resolvable peaks from five, at low temperatures, to four at high temperatures, as observed for the uncomplexed  $A_3T_3$  sequence in chapter 4. This change is consistent with the melting of the DNA duplex. Analysing these spectra, via PCA, yielded a melting curve of the  $N_3-A_3T_3$  complex that could be well-represented by a sigmoidal function with a  $T_m$  of 81 °C (Fig.6.12.(b), red). This reveals that the binding interactions formed between the  $A_3T_3$  duplex and the  $N_3$ -bBI ligand of 22 °C and this is noted to be identical to stabilisation observed for the binding of H33258 to the A-tract sequence (chapter 5). This indicates that the binding interactions in the  $N_3$ -bBI: $A_3T_3$  and H33258: $A_3T_3$  complexes are similar. Additionally, it is noted that when the purified  $N_3$ -bBI ligand is used to prepare the  $N_3-A_3T_3$  complex a clear solution was formed, this is consistent with the melting curve of the complex being represented by a single sigmoid indicating all of the  $A_3T_3$  duplexes are bound.

Finally, the changes in the asymmetric azide stretch of the purified  $N_3$ -bBI ligand upon binding are investigated to determine if this was affected by the presence of the impurity in the original  $N_3$ -bBI ligand, these FT-IR spectra are shown in Fig.6.13.

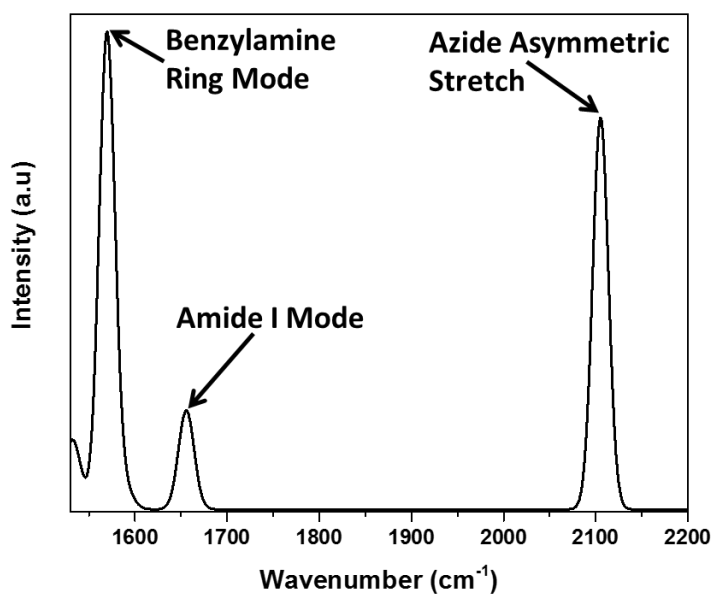


**Figure 6.13:** FT-IR spectra of the asymmetric azide stretch of a) the  $N_3$ - $A_3T_3$  complex (impure ligand); b) the  $N_3$ - $A_3T_3$  and  $N_3$ - $b$ BI ligand (purified ligand).

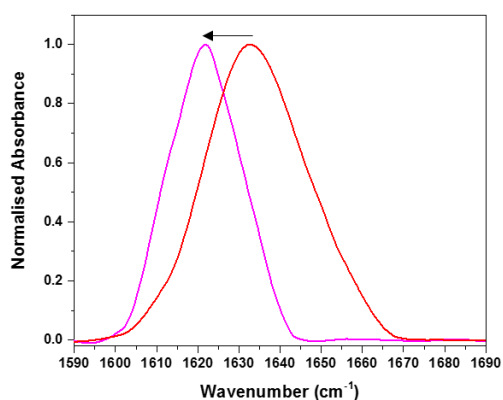
In the previous experimental chapter it was noted that upon binding to the  $A_3T_3$  sequence the asymmetric azide stretch of the  $N_3$ - $b$ BI ligand (Fig.6.13.(a)) gained a high frequency component centred at  $2118\text{ cm}^{-1}$ . For the purified  $N_3$ - $b$ BI ligand the asymmetric azide stretch of the unbound ligand was found to consist of a single peak centred at  $2103\text{ cm}^{-1}$  (blue, Fig.6.13.(b)). Upon binding to the  $A_3T_3$  sequence the entire peak was found to be shifted to high frequency by  $15\text{ cm}^{-1}$ , yielding a single peak centred at  $2118\text{ cm}^{-1}$  (red, Fig.6.13.(b)). The shift of the asymmetric azide stretch upon binding is noted to be fully consistent with the high frequency component gained upon the binding of the impure ligand in the previous experimental chapter. Furthermore, the shift of the entire azide mode of the purified  $N_3$ - $b$ BI ligand upon binding to the  $A_3T_3$  sequence is consistent with all of the ligands being bound, consistent with the observation of a single sigmoidal melting curve for the  $N_3$ - $A_3T_3$  complex (Fig.6.12.(b)).

Overall, these results of the spectroscopy of the complex formed between the purified  $N_3$ - $b$ BI ligand and the  $A_3T_3$  sequence indicate that the presence of the impurity only seems to have resulted in a decrease in the solubility of the ligand. This is thought to have led to the observation in the original chapter that only 21% of the duplexes were bound. However, the interactions between the  $N_3$ - $b$ BI ligand and the DNA sequences and the changes in the asymmetric azide stretch of the ligand were unaffected by the presence of the impurity. Therefore, it can be concluded from this addendum that broadly speaking the presence of the impurity in the original  $N_3$ - $b$ BI ligand did not affect the complexes formed in the previous experimental chapter.

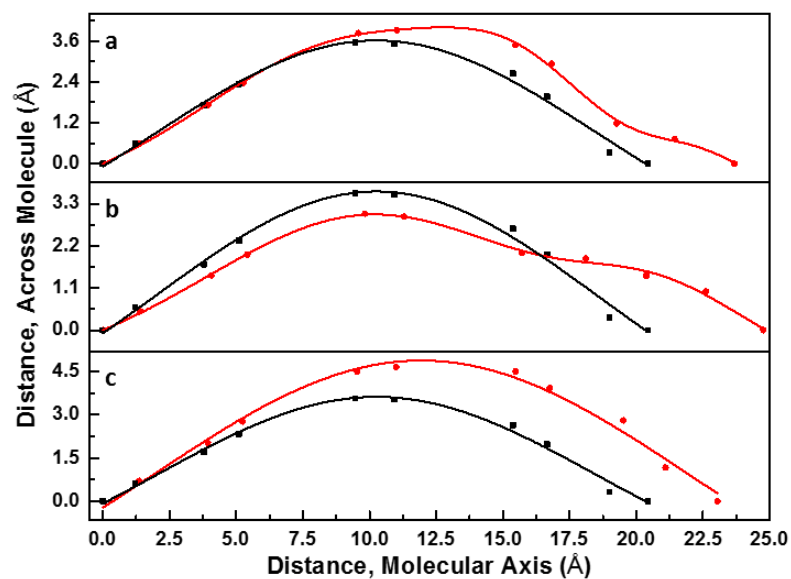
## 6.7 Appendix



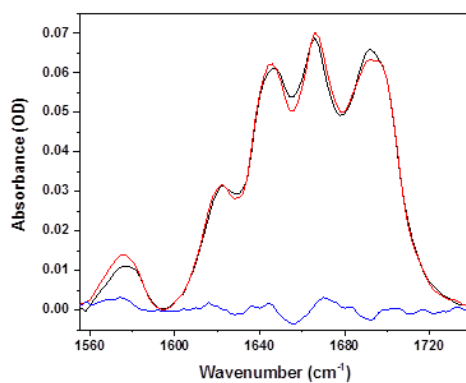
**Figure A6.1:** Simulated and assigned IR spectrum generated from a B3LYP/6-311G(d,p) Gaussian09<sup>154</sup> calculation performed on the deuterated N<sub>3</sub>-bBI ligand.



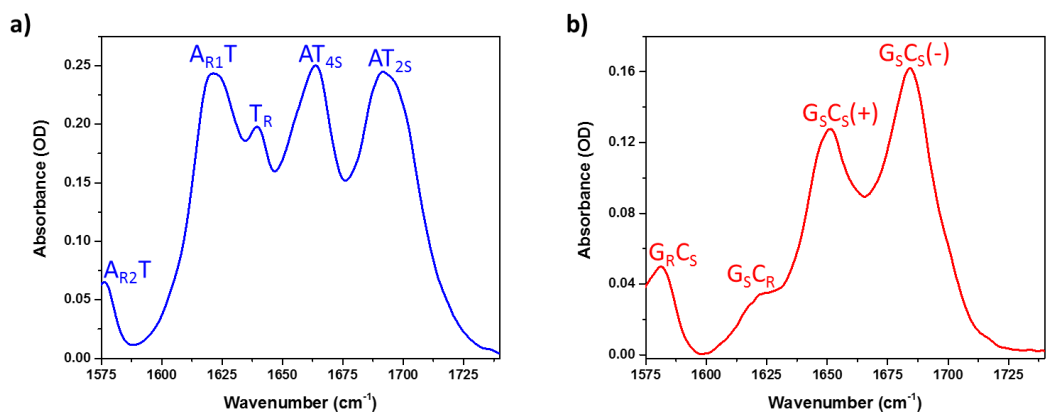
**Figure A6.2:** FT-IR spectra of the amide I vibrational mode of N<sub>3</sub>-bBI measured at pD 5 (red) and pD 9.5 (magenta). Measured at an ionic strength of 200 mM.



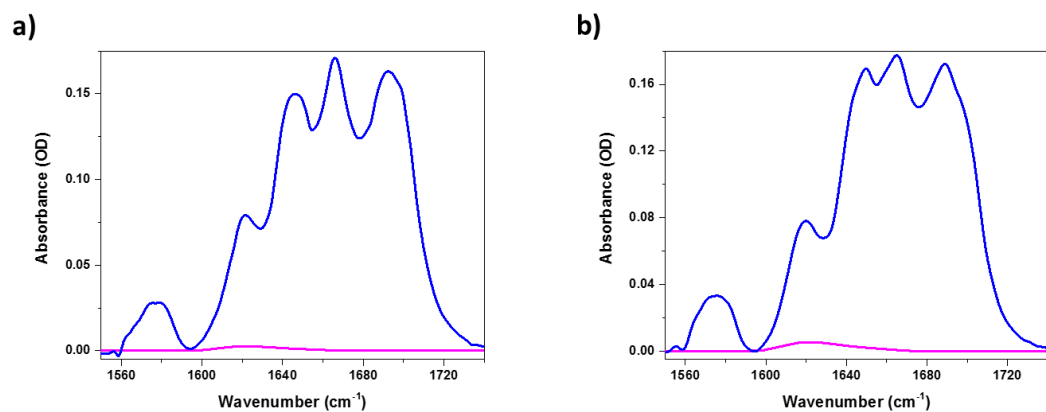
**Figure A6.3:** Calculated curvature of H33258 (black) and a) rotamer 3 (red), b) rotamer 2 (red) and c) rotamer 1 (red) of N<sub>3</sub>-bBI. Calculated from the optimised geometries of both ligands performed utilising the hybrid-DFT B3LYP functional with a 6-311G(d,p) basis set.<sup>154</sup>



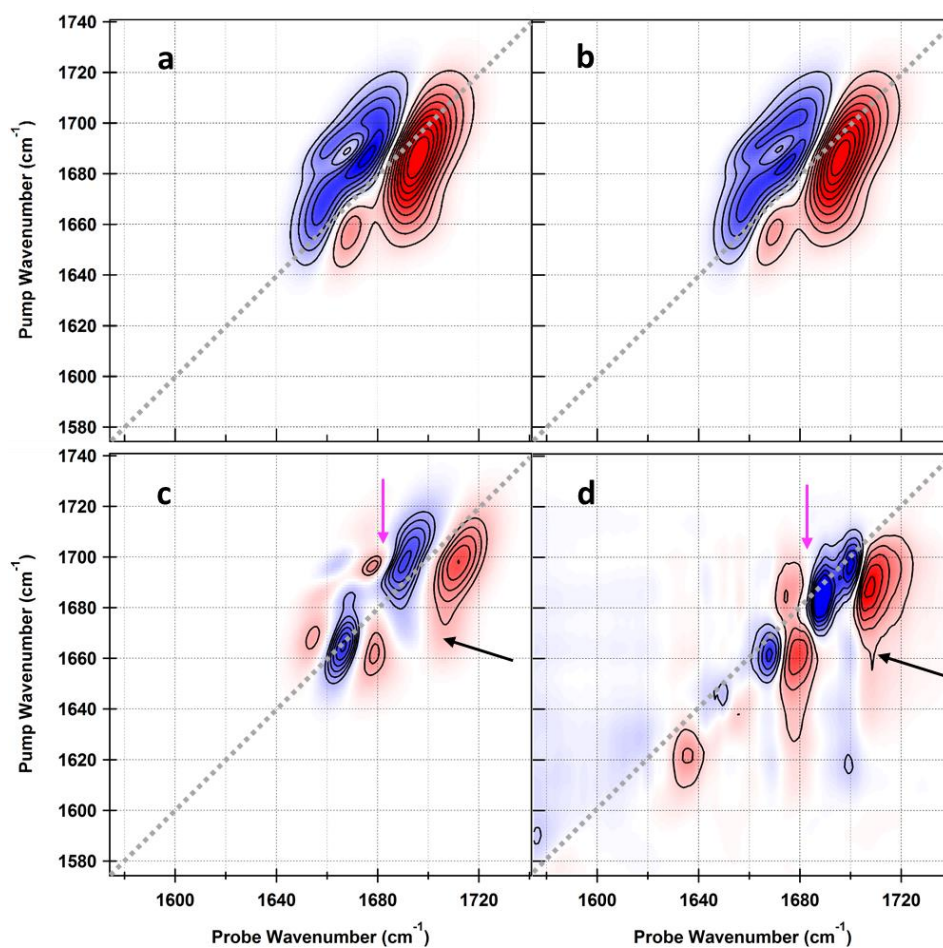
**Figure A6.4:** FT-IR spectra of the DNA base modes of the A<sub>3</sub>T<sub>3</sub> sequence without DMSO (black) and with 10 (V/V)% DMSO (red). Blue spectrum shows the DMSO-induced difference FTIR spectrum.



**Figure A6.5:** Assigned FTIR spectra of a) AT only duplex (sequence: 5'-ATTATTATTATATTA-3') and b) GC only duplex (sequence: 5'-GCCGCCGCCG-3').

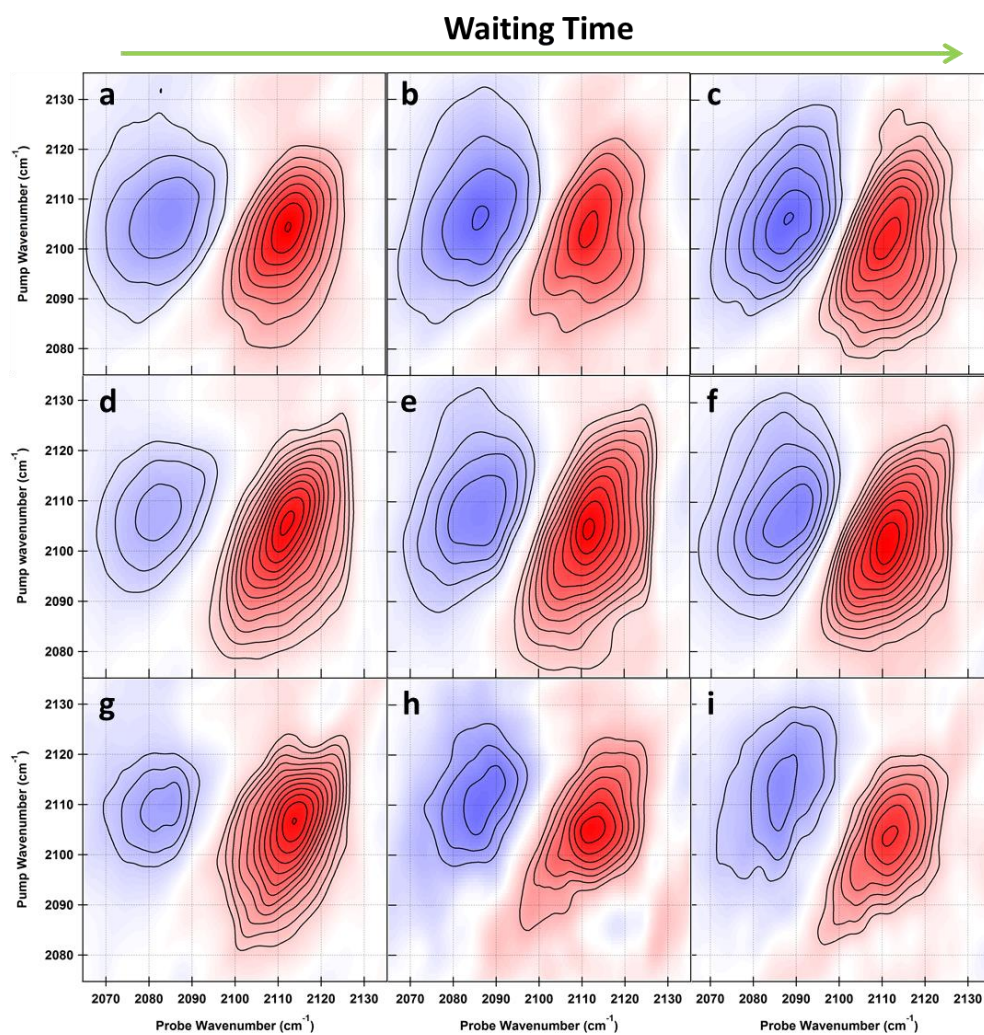


**Figure A6.6:** FTIR spectra of a)  $A_3T_3$  duplex only (blue) and  $N_3$ -bBI only (pink), b)  $(AT)_3$  duplex only (blue) and  $N_3$ -bBI only (pink). All spectra recorded at 20 °C.

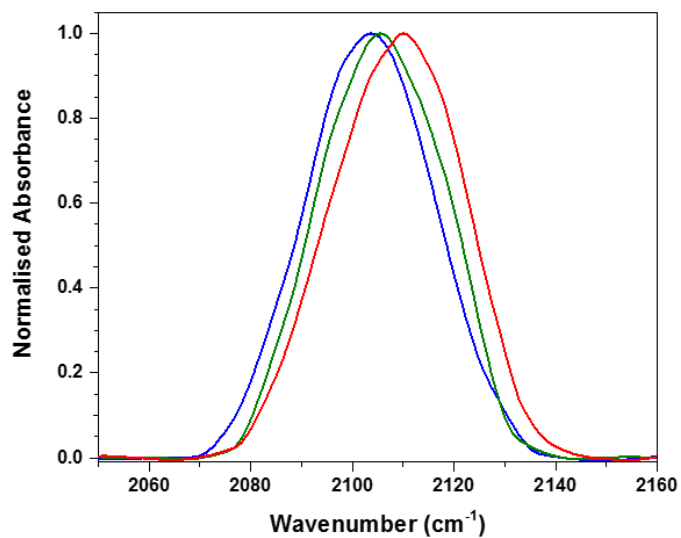


**Figure A6.7:** Simulated 2D-IR spectra of the coupled  $AT_{25}$  and  $AT_{45}$  transitions before (a) and after (b) a shift of a subset of the bands to higher frequency. (c) shows the simulated difference 2D-IR spectrum that results from a subtraction of (a) from (b). (d) Shows the experimentally observed difference 2D-IR spectrum obtained following formation of the  $N_3$ - $A_3T_3$  complex. Pink and black arrows point to off-diagonal features linking the two modes in experimental and simulated spectra.

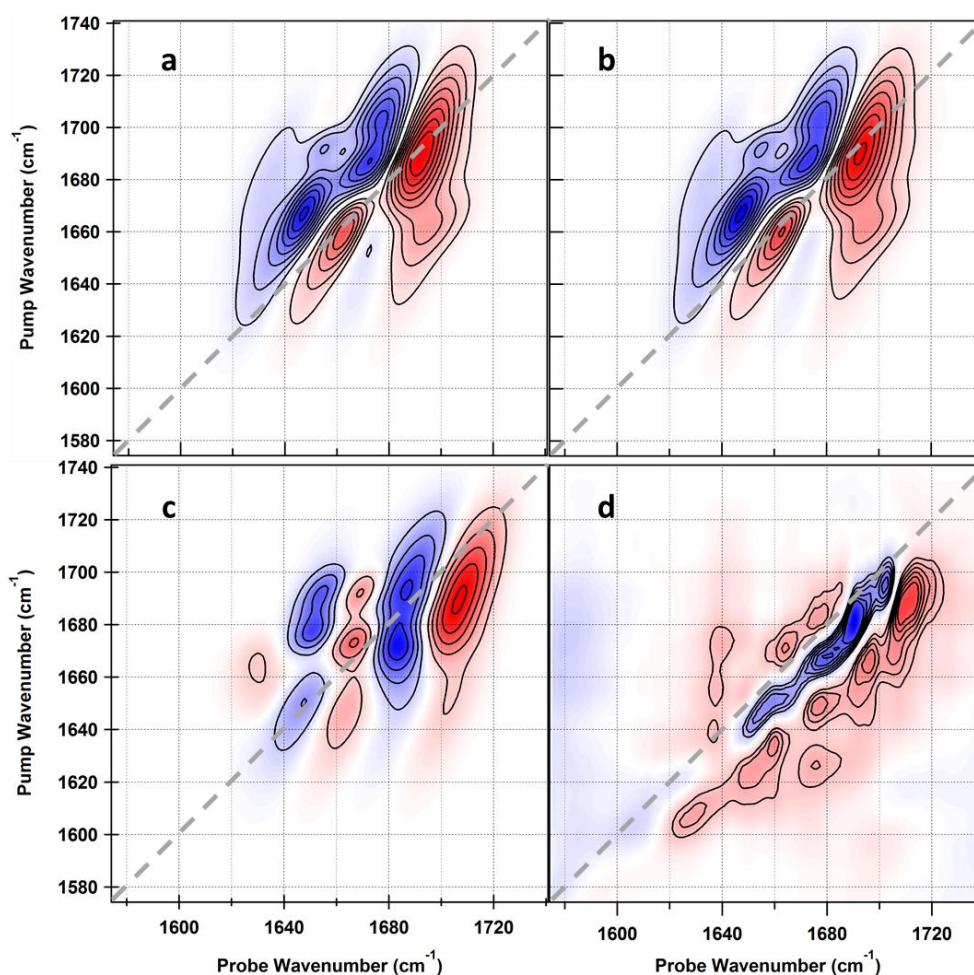




**Figure A6.8:** 2D-IR spectra of the asymmetric azide stretch of the  $N_3$ -bBI ligand in solution at 80 °C (top row) recorded at waiting times of a) 300fs, b) 500fs and c) 700fs. The  $N_3$ -bBI ligand in the  $N_3$ - $A_3T_3$  complex at 80 °C (middle row) recorded at waiting times of d) 300fs, e) 500fs and f) 700fs. The  $N_3$ -bBI ligand in the  $N_3$ - $(AT)_3$  complex at 80 °C (bottom row) recorded at waiting times of g) 300fs, h) 500fs and i) 700fs.



**Figure A6.9:** Normalized FTIR spectra of the asymmetric azide stretching mode of  $N_3$ -bBI in an aqueous solution (red, pD 5), partially dissolved (green, pD 7) and suspended in an aqueous environment (blue, pD 9.5).



**Figure A6.10:** Simulated 2D-IR spectra of the coupled  $AT_{2s}$ ,  $G_5C_5(-)$ ,  $AT_{4s}$  and  $G_5C_5(+)$  transitions before (a) and after (b) a shift of a subset of the  $AT_{2s}$ ,  $G_5C_5(-)$  and  $G_5C_5(+)$  bands to higher frequency. (c) shows the simulated difference 2D-IR spectrum that results from a subtraction of (a) from (b). (d) Shows the difference 2D-IR spectrum obtained following formation of the  $N_3-(AT)_3$  complex.

## 6.8 References

(298) Neidle, S. DNA Minor-Groove Recognition by Small Molecules. *Nat. Prod. Rep.* **2001**, *18*, 291-309.

(299) White, C. M.; Heidenreich, O.; Nordheim, A.; Beerman, T. A. Evaluation of the Effectiveness of DNA-Binding Drugs to Inhibit Transcription Using the c-fos Serum Response Element as a Target, *Biochemistry*, **2000**, *39*, 12262-12273.

(300) Zhang, X.; Kiechle, F. Hoechst33342-Induced Apoptosis is Associated with Decreased Immunoreactive Topoisomerase I and Topoisomerase I-DNA Complex Formation, *Annals of Clinical & Laboratory Science*, **2001**, *31*, 187-198

(301) Bellorini, M.; Moncollin, V.; D'Incalci, M.; Mongelli, N.; Mantovani, R. Distamycin A and Tallimustine Inhibit TBP Binding and Basal *In Vitro* Transcription, *Nucleic Acids Research*, **1995**, *23*, 1657-1663.

- 
- (302) Rydzewski, J. M.; Leupin, W.; Chazin, W.; Sequence-dependent variation in DNA minor groove width dictates orientational preference of Hoechst 33258 in A-tract recognition: solution NMR structure of the 2:1 complex with d(CTTTTGCAAAAG)<sub>2</sub>, *Nucleic Acids Res.*, **2000**, *28*, 728-735.
- (303) Pelton, J. G.; Wemmer, D. E.; Structural characterization of a 2:1 distamycin A.d(CGCAAATTGGC) complex by two-dimensional NMR., *Proc. Natl. Acad. Sci. U.S.A.*, **1989**, *86*, 5723-5727.
- (304) Parkinson, J. A.; Barber, J.; Douglas, K. T.; Rosamond, J.; Sharples, D. Minor-Groove Recognition of the Self-Complementary Duplex d(CGCGAATTCGCG)<sub>2</sub> by Hoechst33258: A High-Field NMR Study. *Biochemistry*, **1990**, *29*, 10181-10190.
- (305) Fede, A.; Labhardt, A.; Bannwarth, W.; Leupin, W. Dynamics and Binding Mode of Hoechst33258 to d(GTGGAAATCCAC)<sub>2</sub> in the 1:1 Solution Complex as Determined by Two-Dimensional <sup>1</sup>H NMR. *Biochemistry*, **1991**, *30*, 11377-11388.
- (306) Higgins, L. D.; Searle, M. S. Site-Specificity of Bis-Benzimidazole Hoechst33258 in A-tract Recognition of the DNA Dodecamer Duplex d(GCAAAATTTGCG)<sub>2</sub>. *Chem. Commun.*, **1999**, *18*, 1861-1862.
- (307) Parkinson, J. A.; Ebrahimi, S. E.; McKie, J. H.; Douglas, K. T. Molecular Design of DNA-Directed Ligands with Specific Interactions: Solution NMR Studies of the Interaction of a m-Hydroxy Analog of Hoechst33258 with d(CGCGAATTCGCG)<sub>2</sub>, *Biochemistry*, **1994**, *33*, 8442-8452.
- (308) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E.; Binding of an antitumor drug to DNA, Netropsin and C-G-C-G-A-A-T-T-BrC-G-C-G., *J. Mol. Biol.*, **1985**, *183*, 553-563.
- (309) Brown, D. G.; Sanderson M. R.; Garman, E.; Neidle, S.; Crystal structure of a berenil-d(CGCAAATTTGCG) complex: An example of drug-DNA recognition based on sequence-dependent structural features, *J. Mol. Biol.*, **1992**, *226*, 481-490.
- (310) Coll, M.; Frederick, C. A.; Wang, A. H. J.; Rich, A.; A bifurcated hydrogen-bonded conformation in the d(AT) base pairs of the DNA dodecamer d(CGCAAATTTGCG) and its complex with distamycin, *Proc. Natl. Acad. Sci. U.S.A.*, **1987**, *84*, 8385.
- (311) Teng, M.; Usman, N.; Frederick, C. A., Wang, A. H.-J. The Molecular Structure of the Complex of Hoechst33258 and the DNA Dodecamer d(CGCGAATTCGCG). *Nucleic Acids Research*, **1988**, *16*, 2671-2690.
- (312) Quintana, J. R.; Lipanov, A. A.; Dickerson, R. E. Low-Temperature Crystallographic Analyses of the Binding of Hoechst33258 to the Double-Helical DNA Dodecamer C-G-C-G-A-A-T-T-C-G-C-G, *Biochemistry*, **1991**, *30*, 10294-10306
- (313) Vega, M. C.; Sáez, I. G.; Aymami, J.; Erita, R.; van der Marel, G. A.; van Boom, J. H.; Rich, A.; Coll, M. Three-Dimensional Crystal Structure of the A-tract DNA Dodecamer d(CGCAAATTTGCG) Complexed with the Minor-Groove-Binding Drug Hoechst33258, *Eur. J. Biochem.*, **1994**, *222*, 721-726.
- (314) Spink, N.; Brown, D. G.; Skelly J. V.; Neidle, S. Sequence-Dependent Effects in Drug-DNA Interaction: The Crystal Structure of Hoechst33258 Bound to the d(CGCAAATTTGCG)<sub>2</sub> Duplex, *Nucleic Acids Research*, **1994**, *22*, 1607-1612.
- (315) Carrondo, M. A. A. F.; De C. T.; Coll, M.; Aymami, J.; Wang, A. H.-J.; van der Marel, G. A.; van Boom, J. H.; Rich, A. Binding of Hoechst Dye to d(CGCGATATCGCG) and its Influence on the Conformation of the DNA Fragment. *Biochemistry*, **1989**, *28*, 7849-7859.

- 
- (316) Furse, K. E.; Corcelli, S. A.; The Dynamics of Water at DNA Interfaces: Computational Studies of Hoechst 33258 Bound to DNA, *J. Am. Chem. Soc.*, **2008**, *130*, 13103-13109.
- (317) Sajadi, M.; Furse, K. E.; Zhang, X.-X.; Dehmel, L.; Kovalenko, S. A.; Corcelli, S. A.; Ernsting, N. P.; Detection of DNA-Ligand Binding Oscillations by Stokes-Shift Measurements, *Angew. Chem. Int. Ed.*, **2011**, *50*, 9501-9505.
- (318) Hunt, N. T. 2D-IR spectroscopy: Ultrafast Insights into Biomolecule Structure and Function, *Chem. Soc. Rev.*, **2009**, *38*, 1837-1848.
- (319) Park, S.; Kwak, K.; Fayer, M. D. Ultrafast 2D-IR Vibrational Echo Spectroscopy: A Probe of Molecular Dynamics, *Laser Phys. Lett.*, **2007**, *10*, 704-718.
- (320) Baiz, C. R.; McRobbie, P. L.; Anna, J. M.; Geva, E.; Kubarych, K. J. Two-Dimensional Infrared Spectroscopy of Metal Carbonyls, *Accounts of Chemical Research*, **2009**, *42*, 1395-1404.
- (321) Peng, C. S.; Jones, K. C.; Tokmakoff, A. Anharmonic Vibrational Modes of Nucleic Acid Bases Revealed by 2D IR Spectroscopy, *J. Am. Chem. Soc.*, **2011**, *133*, 15650-15660.
- (322) Yang, M.; Szyz, L.; Elsaesser, T. Femtosecond Two-Dimensional Infrared Spectroscopy of Adenine-Thymine Base Pairs in DNA Oligomers, *J. Phys. Chem. B*, **2011**, *115*, 1262-1267.
- (323) Greve, C.; Elsaesser, T. Ultrafast Two-Dimensional Infrared Spectroscopy of Guanine-Cytosine Base Pairs in DNA Oligomers, *J. Phys. Chem. B*, **2013**, *117*, 14009-14017.
- (324) Krummel, A. T.; Zanni, M. T. DNA Vibrational Coupling Revealed with Two-Dimensional Infrared Spectroscopy: Insight into Why Vibrational Spectroscopy is Sensitive to DNA Structure, *J. Phys. Chem. B*, **2006**, *110*, 13991-14000.
- (325) Ramakers, L. A. I.; Hithell, G.; May, J. J.; Greetham, G. M.; Donaldson, P. M.; Towrie, M.; Parker, A. W.; Burley, G. A.; Hunt, N. T.; 2D-IR Spectroscopy Shows that Optimised DNA Minor Groove Binding of Hoechst33258 Follows an Induced Fit Model, *J. Phys. Chem. B*, **2017**, *121*, 1295-1303.
- (326) Kim, H.; Cho, M.; Infrared Probes for Studying the Structure and Dynamics of Biomolecules, *Chem. Rev.* **2013**, *113*, 5817-5847.
- (327) Hu, W.; Webb, L. J.; Direct Measurement of the Membrane Dipole Field in Bicelles Using Vibrational Stark Effect Spectroscopy, *J. Phys. Chem. Lett.* **2011**, *2*, 1925-1930.
- (328) Tucker, M. J.; Kim, K. S.; Hochstrasser, R. M.; 2D IR photon echo study of the anharmonic coupling in the OCN region of phenyl cyanate, *Chem. Phys. Lett.* **2009**, *470*, 80-84.
- (329) Stafford, A. J.; Ensign, D. L.; Webb, L. J.; Vibrational Stark Effect Spectroscopy at the Interface of Ras and Rap1A Bound to the Ras Binding Domain of RalGDS Reveals an Electrostatic Mechanism for Protein-Protein Interaction, *J. Phys. Chem. B*, **2010**, *114*, 15331-15344.
- (330) Thielges, M. C.; Axup, J. Y.; Wong, D.; Lee, H. S.; Chung, J. K.; Schultz, P. G.; Fayer, M. D.; Two-Dimensional IR Spectroscopy of Protein Dynamics Using Two Vibrational Labels: A Site-Specific Genetically Encoded Unnatural Amino Acid and an Active Site Ligand, *J. Phys. Chem. B*, **2011**, *115*, 11294-11304.
- (331) Inouye, H.; Gleason, K. A.; Zhang, D.; Decatur, S. M.; Kirschner, D. A.; Differential effects of phe19 and phe20 on fibril formation by amyloidogenic peptide A $\beta$ 16-22 (Ac-KLVFFAE-NH<sub>2</sub>), *Proteins: Struct., Funct., Bioinf.*, **2010**, *78*, 2306-2321.

---

(332) Bischak, C. G.; Longhi, S.; Snead, D. M.; Costanzo, S.; Terrer, E.; Londergan, C. H.; Probing structural transitions in the intrinsically disordered C-terminal domain of the measles virus nucleoprotein by vibrational spectroscopy of cyanylated cysteines., *Biophys. J.*, **2010**, *99*, 1676-1683.

(333) Mukherjee, S.; Chowdhury, P.; DeGrado, W. F.; Gai, F.; Site-Specific Hydration Status of an Amphipathic Peptide in AOT Reverse Micelles, *Langmuir*, **2007**, *23*, 11174-11179.

(334) Choi, J.-H.; Oh, K.-I.; Cho, M.; Azido-derivatized compounds as IR probes of local electrostatic environment: Theoretical studies, *J. Chem. Phys.*, **2008**, *129*, 174512.

(335) Downs, T. R.; Wilfinger, W. W. Fluorometric Qualification of DNA in Cells and Tissue. *Anal. Biochem.* **1983**, *131*, 538-547.

(336) White, C. M.; Heidenreich, O.; Nordheim, A.; Beerman, T. A. Evaluation of the Effectiveness of DNA-Binding Drugs to Inhibit Transcription Using the c-fos Serum Response Element as a Target, *Biochemistry*, **2000**, *39*, 12262-12273.

(337) Zhang, X.; Kiechle, F. Hoechst33342-Induced Apoptosis is Associated with Decreased Immunoreactive Topoisomerase I and Topoisomerase I-DNA Complex Formation, *Annals of Clinical & Laboratory Science*, **2001**, *31*, 187-198.

(338) Bellorini, M.; Moncollin, V.; D'Incalci, M.; Mongelli, N.; Mantovani, R. Distamycin A and Tallimustine Inhibit TBP Binding and Basal *In Vitro* Transcription, *Nucleic Acids Research*, **1995**, *23*, 1657-1663.

(339) Beerman, T. A.; McHugh, M. M.; Sigmund, R.; Lown, J. W.; Rao, K. E.; Bathini, Y.; *BioChim. Biophys. Acta.*, **1992**, *1131*, 53.

(340) Lombardy, R. L.; Tanious, F. A.; Ramachandran, K.; Tidewell, R. R.; Wilson, W. D.; *J. Med. Chem.*, **1996**, *39*, 1453.

(341) Pilch, D. S.; Xu, Z.; Sun, Q.; LaVoie, E. J.; Lui, L. F.; Geacintov, N. E.; Breslauer, K. J.; *Drug Des. Discov.*, **1996**, *13*, 115.

(342) Kim, J. S.; Sun, Q.; Yu, C.; Liu, A.; Liu, L. F.; LaVoie, E. J.; *Bioorg. Med. Chem. Lett.*, **1998**, *6*, 163.

(343) Breusegem, S. Y.; Clegg, R. M.; Loontjens, F. G. Base-Sequence Specificity of Hoechst33258 and DAPI Binding to Five (A/T)<sub>4</sub> DNA Sites with Kinetic Evidence for More Than One High-Affinity Hoechst33258-AATT Complex, *J. Mol. Biol.*, **2002**, *315*, 1049-1061.

(344) 
$$P(E_i, T) = \frac{e^{(-E_i/RT)}}{\sum_i (e^{(-E_i/RT)})}$$

Where  $P(E_i, T)$  is the relative abundance of the state of energy  $E_i$  at temperature  $T$ .

(345) Peng, C. S.; Jones, K. C.; Tokmakoff, A. Anharmonic Vibrational Modes of Nucleic Acid Bases Revealed by 2D IR Spectroscopy, *J. Am. Chem. Soc.*, **2011**, *133*, 15650-15660.

(346) Krummel, A. T.; Zanni, M. T. DNA Vibrational Coupling Revealed with Two-Dimensional Infrared Spectroscopy: Insight into Why Vibrational Spectroscopy is Sensitive to DNA Structure, *J. Phys. Chem. B*, **2006**, *110*, 13991-14000.

---

(347) Hithell, G.; Shaw, D. J.; Donaldson, P. M.; Greetham, G. M.; Towrie, M.; Burley, G. A.; Parker, W. A.; Hunt, N. T. Long-Range Vibrational Dynamics Are Directed by Watson-Crick Base Pairing in Duplex DNA. *J. Phys. Chem. B*, **2016**, *120*, 4009-4018.

(348) Lee, C.; Cho, M. Vibrational Dynamics of DNA. II. Deuterium Exchange Effects and Simulated IR Absorption Spectra, *J. Chem. Phys.*, **2006**, *125*, 114509.

(349) Mukerji, I.; Williams, A. P. UV Resonance Raman and Circular Dichroism Studies of a DNA Duplex Containing an A<sub>3</sub>T<sub>3</sub> Tract: Evidence for a Premelting Transition and Three-Centered H-Bonds, *Biochemistry*, **2002**, *41*, 69-77.

(350) Thorogood, H.; Waters, T. R.; Parker, A. W.; Wharton, C. W.; Connolly, B. A. Resonance Raman Spectroscopy of 4-Thiothymidine and Oligodeoxynucleotides Containing This Base Both Free in Solution and Bound to the Restriction Endonuclease *EcoRV*. *Biochemistry*, **1996**, *35*, 8723-8733.

$$(351) \quad \Delta\Delta H = \left(\frac{\Delta\nu}{\nu}\right) U_{Bond}$$

where  $\Delta\Delta H$  is the change in the hydrogen bond strength,  $\Delta\nu$  is the shift in the frequency of the vibrational mode,  $\nu$  is the original frequency of the vibrational mode and  $U_{Bond}$  is the bond energy of the affected moiety. Note this equation is only valid if  $\Delta\nu$  is small compared to  $U_{Bond}$ .

(352) Kwak, K.; Park, S.; Finkelstein, I. J.; Fayer, M. D.; Frequency-frequency correlation functions and apodization in two-dimensional infrared vibrational echo spectroscopy: a new approach., *J. Chem. Phys.*, **2007**, *127*, 124503.

(353) Fenn, E. E.; Fayer M. D.; Extracting 2D IR frequency-frequency correlation function from two component systems., *J. Chem. Phys.*, **2011**, *135*, 074502.

(354) Donaldson, P. M.; Greetham, G. M.; Shaw, D. J.; Parker, A. W.; Towrie, M.; A 100 kHz Pulse Shaping 2D-IR Spectrometer based on Yb:KGW Amplifiers., *Opt. Lett.*, **2016**, submitted.

(355) Greetham, G. M.; Donaldson, P. M.; Nation, C.; Sazanovich, I. V.; Clark, I. P.; Shaw, D. J.; Parker, A. W.; Towrie, M.; A 100 kHz Time-Resolved Multiple-Probe Femtosecond to Second IR Absorption Spectrometer., *Applied Spectroscopy*, **2016**.

(356) Hithell, G.; Shaw, D. J.; Donaldson, P. M.; Greetham, G. M.; Towrie, M.; Burley, G. A.; Parker, W. A.; Hunt, N. T. Long-Range Vibrational Dynamics Are Directed by Watson-Crick Base Pairing in Duplex DNA. *J. Phys. Chem. B*, **2016**, *120*, 4009-4018.

# Impact of Minor Groove Binding on the Thermal Denaturation of DNA Sequences

Chapter 7



## 7.1 Abstract

*In chapter 4, FT-IR and 2D-IR spectroscopy have been used to gain insights into the mechanisms underlying the melting transition of an A-tract (5'-A<sub>3</sub>T<sub>3</sub>) and alternating AT (5'-ATATAT) DNA sequences. Additionally these techniques have been successfully employed to gain a deeper understanding of the interactions underpinning the formation of complexes between two different bis-benzimidazole ligands and both the A-tract and alternating AT DNA sequences. However, so far very little is known about the impact of binding on the melting mechanism. In this chapter, the impact of two bis-benzimidazole minor groove binders on the melting mechanisms of an A-tract (5'-A<sub>3</sub>T<sub>3</sub>) and alternating AT (5'-ATATAT) DNA sequences is investigated. Results from both FT-IR and 2-dimensional infrared spectroscopy indicate that binding does not fundamentally alter the melting mechanism of either of the DNA sequences explored. In fact the only impact of such binding interactions was found to be an increase in the temperature at which the sequences start to melt, not altering the end-fraying mechanism observed for the unbound sequences (chapter 4). Interestingly, in addition to showing that the binding does not fundamentally alter the melting mechanism, the temperature induced shifts in the T<sub>2</sub> carbonyls indicate that the increase in the melting temperature is predominantly due to the ligand-DNA hydrophobic interactions formed upon binding.*

## 7.2 Introduction

Very little is known about the impact of the presence of minor groove binders on the mechanism of the melting of these short DNA oligonucleotides. Recently a single-molecule AFM study was carried out which determined that interactions formed between the archetypal minor groove binder H33258 and an A-tract sequence increased the overall duplex rigidity.<sup>357</sup> This increase in rigidity was assigned to the bifurcated H-bonds formed between the ligand and the duplex. However, the exact contributions of the H-bonding interactions and the ligand-duplex hydrophobic contacts, within the complex, to the stabilisation of the melting temperature has remained unclear.<sup>358,359,360</sup>

Two dimensional infrared spectroscopy (2D-IR)<sup>361,362,363</sup> has proved to be a versatile methodology to probe both dynamics and structural aspects of the DNA macromolecule due to its high temporal resolution and ability to probe the complex intermolecular couplings within DNA.<sup>364,365,366,367</sup> Recently 2D-IR has been used to unravel the melting mechanisms and understand the impact of the melting on the structural dynamics of short DNA oligonucleotides.<sup>368,369</sup> One of these studies has led to valuable insights into the impact of the alterations in the position of the GC bases within the sequence on the overall melting mechanism.<sup>69</sup> While the other demonstrated that melting induced more rapid spectral diffusion, as well as indicating the water in the spine of hydration is dynamically restricted in *ds*-DNA.<sup>70</sup>

In chapters 4, 5 and 6, both FT-IR and 2D-IR spectroscopy have been used to explore different aspects of the behaviour of short DNA oligonucleotides.<sup>370,371</sup> This has led to a deeper understanding of the mechanisms underlying the melting of such sequences and the interactions underpinning the binding of an archetypal minor groove binder. Here an attempt is made to utilise IR spectroscopy to gain further insights into the impact of these binding interactions on the melting mechanism of these sequences. In order to gain these insights the changes in the FT-IR and 2D-IR spectra of the base modes of the H-(AT)<sub>3</sub> complex are studied. The interactions underlying this complex were found to follow a “rigid body” model (Chapter 5), the fact that the (AT)<sub>3</sub> duplex doesn’t experience any significant structural perturbation allows the impact of the binding on the melting mechanism to be explored. This exploration leads to the hypothesis that the binding of H33258 does not alter the overall steps in the melting of the DNA duplex, instead the interactions between the ligand and the DNA duplex result in the delayed onset of the melting transition. Furthermore the impact of

binding on the melting transition of the  $A_3T_3$  sequence supports this hypothesis, however this analysis is further complicated by the fact that the “induced fit” binding interactions in this case result in a perturbation of the  $A_3T_3$  sequence (Chapter 5).

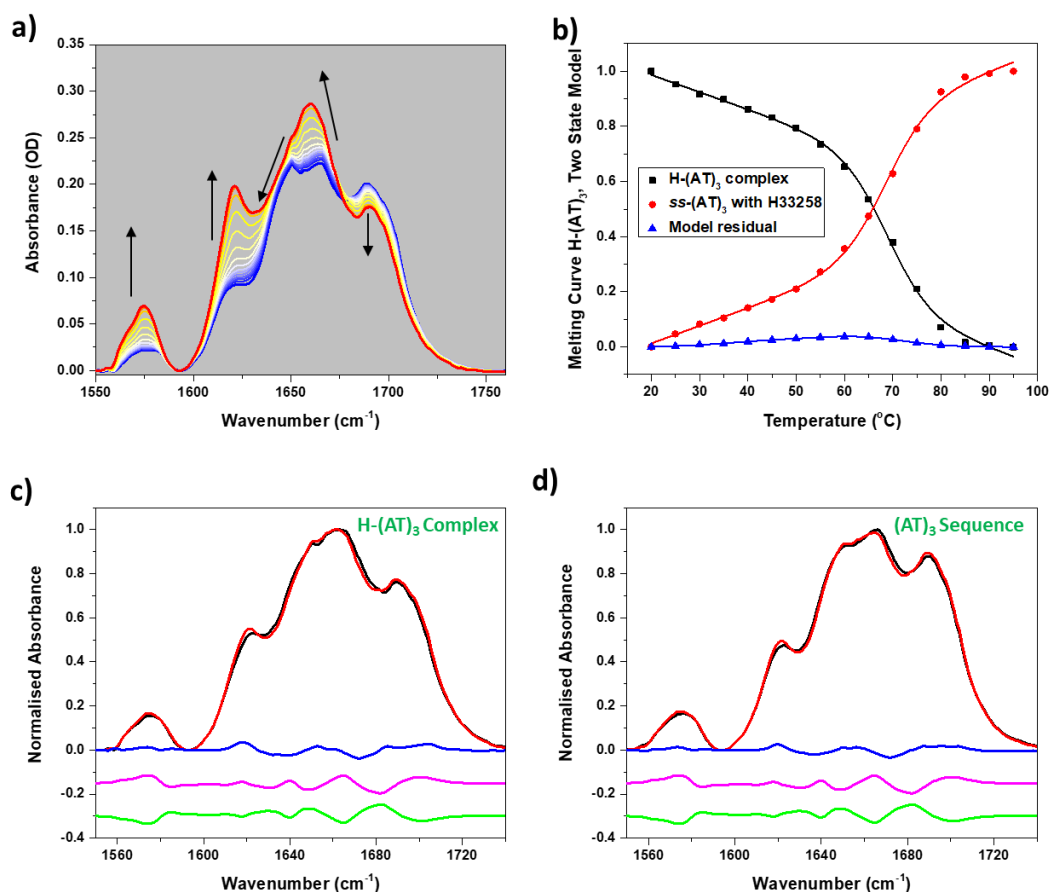
## 7.3 Results and Discussion

The structure of this chapter is as follows. The first section will focus on the impact of the binding of H33258 to the alternating AT  $((AT)_3)$  sequence, as in chapter 5 the interactions underlying this sequence were noted to follow a rigid body mechanism. This presents the simplest possible binding scenario as these interactions do not alter the structure of the duplex meaning any changes in the FT-IR and 2D-IR spectra of the DNA base modes is only due to the increase in temperature. After this the impact of the binding of H33258 to the A-tract ( $A_3T_3$ ) sequence is considered. In chapter 5 it was found that the binding interactions underlying this complex followed an induced fit mechanism, leading to a perturbation of the duplex structure. This presents a more challenging scenario, and it presents a good test of the hypothesis developed for the previous example. Finally, the complexes formed between the azido-derivative of the H33258 ligand and these two sequences are briefly considered and the challenges in the analysis of this data will be discussed.

### 7.3.1 H- $(AT)_3$ complex

#### **DNA Base Modes**

When the melting of the alternating sequence was studied in chapter 4, it was shown that as the temperature increases the FT-IR spectra of the base modes undergoes a series of changes. These changes are also observed in the FT-IR spectrum of the H- $(AT)_3$  complex as this melts (Fig.7.1.(a)). For the H- $(AT)_3$  complex the most prominent change is an overall reduction in the number of resolvable peaks from five to four. The spectral changes are noted to be identical to those observed for the uncomplexed DNA (chapter 4) and so are consistent with the melting of the complex.



**Figure 7.1:** FT-IR spectra of the a)  $H-(AT)_3$  complexes at 5 °C intervals between 20 – 95 °C; b) two state model coefficients of ss (red) and ds (black)  $H-(AT)_3$  FT-IR spectra against temperature. Spectral structure of the two state model residuals (blue) when applied to c) the  $H-(AT)_3$  FT-IR spectra at 65 °C and d)  $(AT)_3$  sequence FT-IR spectra at 45 °C. The characteristic unzipping and bubble formation model residuals are shown in pink and green respectively in these plots.

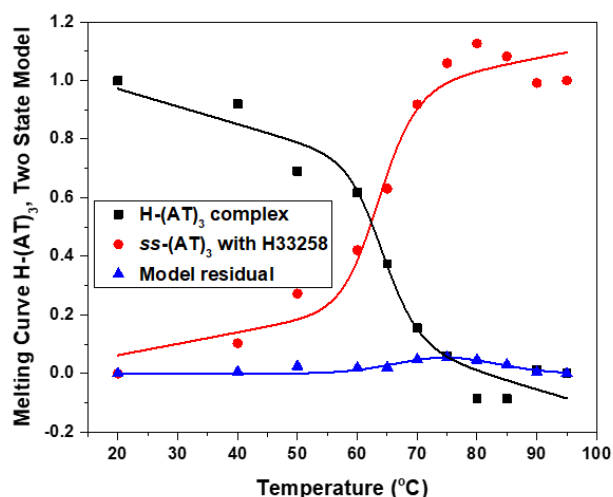
Application of the two state model to the IR spectra obtained upon the heating of the complex also shows similar results to the free DNA sequence. Specifically sigmoidal DNA melting curves are obtained (Fig.7.1.(b)). Fitting a sigmoid to the melting curve yields a melting temperature for the  $H-(AT)_3$  complex of 68 °C. This shows that the formation of the  $H-(AT)_3$  increases the melting temperature of the alternating sequence by 16 °C, as observed in Chapter 5. The application of the two state model to the  $H-(AT)_3$  complex results in small but non-randomly distributed residuals (Fig.7.1.(b), blue), as observed for the unbound sequence in Chapter 4. As for the unbound sequence the thermal profile of these residuals can be represented by a sum of two Gaussian peaks. The peak of the residuals is found to be at approximately 60 °C, approximately 15 °C higher than the peak in the thermal profile of the residuals seen for the uncomplexed sequence. The presence of these structured residuals

suggests that the melting mechanism of this complex cannot be fully explained via a two state model.

The spectral structure of the model residuals were compared with the characteristic residuals produced by the duplexes melting via either an unzipping (pink, Figs.7.1.(c)&(d)) or bubble formation mechanism (green, Figs.7.1.(c)&(d)). Here, the unzipping mechanism refers to the scenario in which the ends of the duplex melt before the middle of the sequence, leading to the duplex unzipping as it melts. Whereas, in the bubble formation mechanism the middle of the duplex start to melt before the ends leading to the formation of a melted bubble within the duplex during the melt. Comparison between the experimentally observed residuals for the H-(AT)<sub>3</sub> complex (blue, Fig.7.1.(c)) and those produced by the two possible melting mechanisms suggests that this complex melts via an unzipping mechanism. Furthermore, the large degree of similarity between the experimentally observed residuals of the H-(AT)<sub>3</sub> complex (blue, Fig.7.1.(c)) and unbound (AT)<sub>3</sub> sequence (blue, Fig.7.1.(d)) means that these results indicate that the binding of H33258 does not alter the overall melting mechanism of the (AT)<sub>3</sub> duplex. Instead the results of the FT-IR spectroscopy on the DNA base modes of the H-(AT)<sub>3</sub> complex suggests that the binding interactions underpinning the complex simply delay the onset of the mechanism by 16 °C.

As discussed in the previous DNA chapters (Chapters 4, 5 and 6) both GC and AT base modes contribute to the entire FT-IR spectrum. As a consequence of this it is difficult to determine which changes in the overall spectrum are due to alterations in the GC or AT regions of the duplex. However the distinct GC and AT cross peak patterns observed in the off-diagonal regions of the 2D-IR spectra of DNA sequences allows the changes in the overall spectrum to be assigned to changes in either the GC or AT sections of the duplex. Therefore to further explore the impact of H33258 binding on the melting mechanism of the (AT)<sub>3</sub> sequence are explored via 2D-IR spectra of the complex.

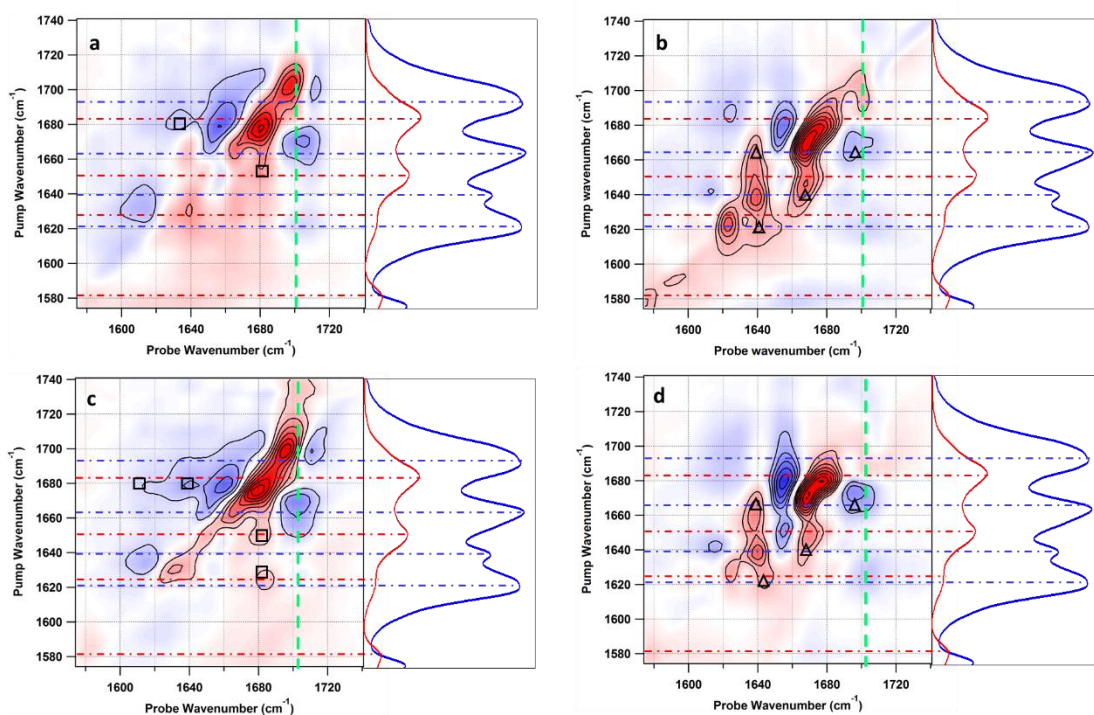
The DNA melting curves, extracted utilising the two state model, for the H-(AT)<sub>3</sub> complex is shown in Fig.7.2. For the H-(AT)<sub>3</sub> complex this analysis also yields sloped melting curves, containing a single sigmoidal step, in agreement with those extracted from the FT-IR spectra of the complex. For the H-(AT)<sub>3</sub> complex (Fig.7.2.(a)) fitting this melting curve to a sigmoid yields a melting temperature of 64 °C. This melting temperature is similar to those obtained from the FT-IR spectra of the base modes of these complexes and show that the binding leads to an increase in the melting temperature of 14 °C.



**Figure 7.2:** Fitted Melting curves (black and red) and model residuals (blue), extracted utilising the two state model, from the 2D-IR spectra of the H-(AT)<sub>3</sub> complex

The only overall change in the melting curves of the H-(AT)<sub>3</sub> complex (Fig.7.2) is noted to be the overall stabilisation of the melting temperature compared to the unbound duplex (Chapter 3). The model residuals are noted to be small but non-randomly distributed (blue, Fig.7.2) again suggesting that the two state model cannot fully describe the melting mechanism. To further investigate the changes occurring during the melt, temperature-induced 2D-IR difference spectra were generated for the H-(AT)<sub>3</sub> complex.

Two 2D-IR difference spectra are analysed, one of these difference spectra was generated to reveal the initial changes in the experimentally measured spectrum underlying the initial slope of the extracted melting curves (Fig.7.2) from 20 °C to 65 °C and the second was generated to gain insights into the changes leading to the sigmoidal step observed in the H-(AT)<sub>3</sub> melting curve from 65 °C to 90 °C. These temperature-induced 2D-IR difference spectra for the H-(AT)<sub>3</sub> and uncomplexed (AT)<sub>3</sub> sequence are shown in Fig 7.3.



**Figure 7.3:** Temperature-induced 2D-IR difference spectrum for the H-(AT)<sub>3</sub> complex a) observed for an increase in temperature from 20 °C to 65 °C and b) observed for an increase in temperature from 65 °C to 90 °C. Temperature-induced 2D-IR difference spectrum for the uncomplexed (AT)<sub>3</sub> sequence c) observed for an increase in temperature from 20 °C to 50 °C and d) observed for an increase in temperature from 50 °C to 90 °C.

Both of the temperature-induced 2D-IR difference spectra of the H-(AT)<sub>3</sub> complex (Figs.7.3.(a)&(b)) are found to be very similar to the 2D-IR difference spectra obtained upon heating the uncomplexed (AT)<sub>3</sub> sequence (Figs.7.3.(c)&(d)). Comparing these temperature induced 2D-IR difference spectra it can be seen that these are broadly similar for the H-(AT)<sub>3</sub> complex and the uncomplexed (AT)<sub>3</sub> sequence. Before the melt, the temperature induced 2D-IR difference spectra of the complex (Fig.7.3.(a)) and the unbound sequence (Fig.7.3.(c)) both contain features at 1650 cm<sup>-1</sup>, 1675 cm<sup>-1</sup> and 1700 cm<sup>-1</sup> on the spectrum diagonal. After the melt, the temperature induced 2D-IR difference spectra of the complex (Fig.7.3.(b)) and unbound sequence (Fig.7.3.(d)) are both noted to contain diagonal features located at approximately 1620 cm<sup>-1</sup>, 1640 cm<sup>-1</sup>, 1660 cm<sup>-1</sup> and 1680 cm<sup>-1</sup>. From these 2D-IR difference spectra it can be seen that the initial changes in the 2D-IR spectrum of the base modes of the H-(AT)<sub>3</sub> complex (Fig.7.3.(a)) are associated with decreases in the intensity of the G<sub>5</sub>C<sub>5</sub>(+) (1650 cm<sup>-1</sup>) and G<sub>5</sub>C<sub>5</sub>(-) (1675 cm<sup>-1</sup>) modes as well as a shift to higher frequency of the AT<sub>25</sub> (1700 cm<sup>-1</sup>) mode, similar to the changes observed in the uncomplexed (AT)<sub>3</sub> sequence (Fig.7.3.(c)). The assignment of the features associated with the loss in intensity of the ds-GC

modes is supported by the appearance of the associated cross peaks in the difference spectrum, indicated with open squares on Figs.7.3.(a)&(c). Finally the assignment of the feature at  $1700\text{ cm}^{-1}$  on the spectrum diagonal is supported by the appearance of features suggesting a shift to higher frequency of cross peaks between this feature and positions associated with other *ds*-AT base modes (green dashed line, Fig.7.3.(a)&(c)). The loss of intensity of these *ds*-GC modes suggests that the temperature of the complex is raised from  $20\text{ }^{\circ}\text{C}$  to  $65\text{ }^{\circ}\text{C}$  the GC-ends of the  $(\text{AT})_3$  sequence starts to melt, as observed for the uncomplexed sequence in Chapter 4.

It is noted that the temperature-induced difference spectra of the H- $(\text{AT})_3$  complex, before the melt, contains features associated with a temperature-induced shift to higher frequency of the  $\text{AT}_{25}$  mode, as observed for the uncomplexed sequence in chapter 4. This change in the frequency of the  $\text{AT}_{25}$  mode is associated with a decrease in the strength of the H-bonds formed to the  $\text{T}_2$  carbonyls. In the complex these carbonyls are known to be involved in the H-bonding interactions formed between the DNA sequence and H33258, therefore it is thought that this mode can be used to allow the role of the H-bonding in the stabilisation of the DNA duplex to be explored. This idea will be revisited in the following section.

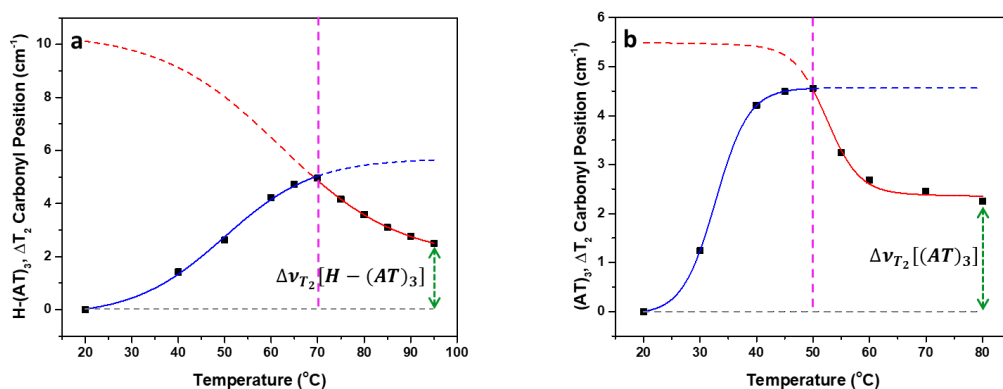
Finally, the temperature-induced difference spectrum showing the changes underlying the sigmoidal step in the melting curve (Fig.7.3.(b)) extracted from the 2D-IR spectra of the H- $(\text{AT})_3$  complex (Fig.7.2) are found to be primarily associated with the loss in intensity of the  $\text{A}_{\text{R1T}}$  ( $1620\text{ cm}^{-1}$ ),  $\text{T}_{\text{R}}$  ( $1640\text{ cm}^{-1}$ ) and  $\text{AT}_{45}$  ( $1660\text{ cm}^{-1}$ ) modes. Again these assignments are supported by the presence of cross peaks between these features appearing the difference spectra (triangles, Fig.7.3.(b)&(d)). Finally the feature located at approximately  $1680\text{ cm}^{-1}$  in these difference spectrum is thought to be due to a loss in the intensity of the  $\text{G}_5\text{C}_5(-)$  mode. Overall it is thought that as the temperature of the complex is increased from  $65\text{ }^{\circ}\text{C}$  to  $90\text{ }^{\circ}\text{C}$  the central AT region of the complex melts. It is thought that the contributions of the  $\text{G}_5\text{C}_5(-)$  mode to this temperature-induced difference spectrum are due to the final stage of the melting of the GC-ends of the duplex. This is fully consistent with the hypothesis that the binding of H33258 to the  $(\text{AT})_3$  sequence does not change the melting mechanism, but leads to the overall melt occurring at a higher temperature.

### ***DNA Minor Groove***

The temperature-induced 2D-IR difference spectrum when the H- $(\text{AT})_3$  complex was heated from  $20\text{ }^{\circ}\text{C}$  to  $65\text{ }^{\circ}\text{C}$  (Fig.7.3.(a)) was found to consist primarily of features arising from the



loss of intensity of *ds*-GC modes, however in addition to these features it was noted that there was a feature present in this difference spectrum associated with a temperature-induced shift to higher frequency of the AT<sub>25</sub> mode. As noted for the uncomplexed (AT)<sub>3</sub> sequence this is consistent with a loss in the strength of the H-bonds formed to the T<sub>2</sub> carbonyls (Chapter 4). As these carbonyls are positioned in the minor groove this mode should be a very sensitive reporter of both ligand binding, in the DNA complex, and hydration, in the uncomplexed sequence. It is thought that the changes in the position of this mode will allow the changes in the ligand H-bonding interaction to be extracted for the complex, in a similar manner to the extraction of the changes in hydration of the unbound sequence in chapter 4. The change in the position of the AT<sub>25</sub> mode in the H-(AT)<sub>3</sub> complex and uncomplexed (AT)<sub>3</sub> sequence as the sample temperature is increased are shown in Fig.7.4.



**Figure 7.4:** Change in the position of the T<sub>2</sub> carbonyl stretch of the a) H-(AT)<sub>3</sub> complex and b) (AT)<sub>3</sub> sequence with increasing temperature.

The position of the AT<sub>25</sub> mode, in both the H-(AT)<sub>3</sub> (Fig.7.4.(a)) and uncomplexed (AT)<sub>3</sub> sequence (Fig.7.4.(b)), initially increases as the temperature increases, however as the temperature increases beyond the extracted melting temperature of the sample the shift of this mode is seen to abruptly change direction (Figs.7.4.(a)&(b)). The overall change in the position of the T<sub>2</sub> carbonyls,  $\Delta v_{T_2}[H - (AT)_3]$  and  $\Delta v_{T_2}[(AT)_3]$  are noted to be 2.5 cm<sup>-1</sup> and 2.25 cm<sup>-1</sup> respectively, consistent with the overall shift to higher frequency of the T<sub>2</sub> carbonyls observed in the FT-IR spectrum of an exclusively AT sequence upon melting (Chapter 4). The abrupt change in the direction of the shift for the AT<sub>25</sub> modes of both the complex and unbound sequence, is indicative of two different processes occurring as the temperature is increased. At low temperatures the dominant process involves a decrease in the strength of the hydrogen bonds formed to the T<sub>2</sub> carbonyls, in the complex it is thought

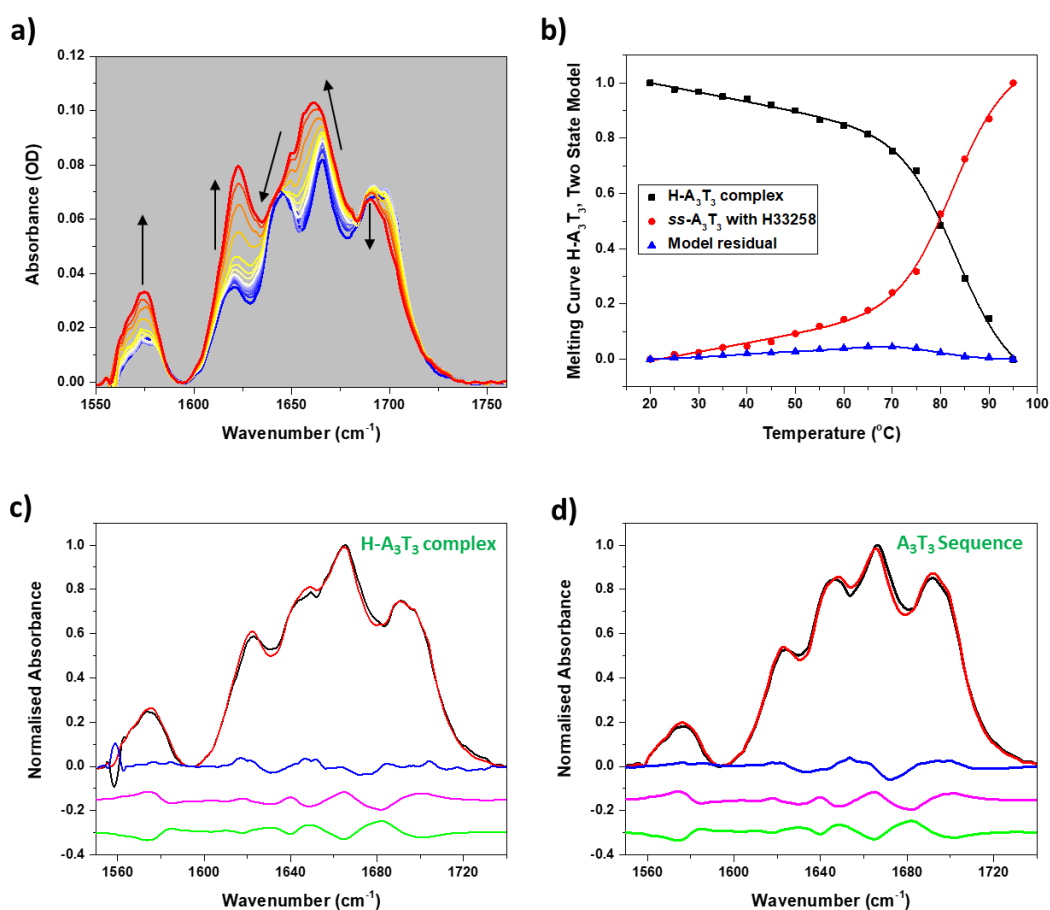
that this is consistent with the loss of the bifurcated hydrogen bonds between the duplex and H33258 as the complex melts. At temperature increases beyond the observed melting temperature of the sample the change in the direction of the shift indicates the increase in the strength of hydrogen bonds formed to the T<sub>2</sub> carbonyls, consistent with the unwinding and solvation of these moieties after H33258 dissociates. Fitting the changes in the position of the T<sub>2</sub> carbonyls as the temperature increases using a piecewise function consisting of two sigmoid curves yields transition temperatures for these two different processes. For the H-(AT)<sub>3</sub> complex these transition temperatures were found to be 50 °C and 62 °C, respectively, whereas for the uncomplexed (AT)<sub>3</sub> sequence these two transitions are centred at 33 °C and 51 °C, respectively.

As seen in both the FT-IR and 2D-IR spectroscopy, the impact of the binding of H33258 to the (AT)<sub>3</sub> sequence does not fundamentally change the melting mechanism of the (AT)<sub>3</sub> duplex. Overall, as the temperature is increased and the complex melts a loss in the GC base pairing is seen to occur below the sigmoidal step within the melting curve, and a loss of the AT base pairing is seen to occur above it. This is spectroscopically identical to the unbound (AT)<sub>3</sub> duplex but shifted up in temperature by 16 °C. This means that the melting mechanism is the same but the onset of the melt is delayed due to the binding interactions formed between the duplex and the ligand. The frequency shifts in the T<sub>2</sub> carbonyls shows that the ligand H-bonds start to weaken before the ligand dissociates from the minor groove of the complex. This suggests that the formation of the H-bonds between the ligand and the duplex are not directly responsible for the stabilisation of the duplex. This is consistent with the previous thermodynamic studies which indicated that the stabilisation of the melting temperature of the duplex is primarily due to the hydrophobic contacts between the ligand and the sides of the DNA minor groove.<sup>17,18</sup>

### 7.3.2 H-A<sub>3</sub>T<sub>3</sub> complex

#### **DNA Base Modes**

As the temperature of the H-A<sub>3</sub>T<sub>3</sub> complex is increased the FT-IR spectra of the DNA base modes (Fig.7.5.(a)) undergo a number of different changes. The most prominent change in the spectrum is an overall reduction of the number of resolvable peaks from five, at low temperatures, to four at high temperatures, as observed for the uncomplexed A<sub>3</sub>T<sub>3</sub> sequence in chapter 4.

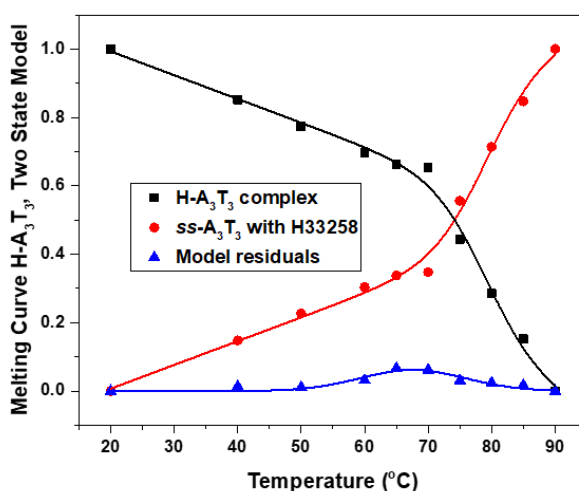


**Figure 7.5:** FT-IR spectra of the a) H-A<sub>3</sub>T<sub>3</sub> complex at 5 °C intervals between 20 – 95 °C; b) two state model coefficients of ss (red) and ds (black) H-A<sub>3</sub>T<sub>3</sub> FT-IR spectra against temperature. Spectral structure of the two state model residuals (blue) when applied to c) the H-A<sub>3</sub>T<sub>3</sub> FT-IR spectra at 80 °C and d) A<sub>3</sub>T<sub>3</sub> sequence FT-IR spectra at 50 °C. The characteristic unzipping and bubble formation model residuals are shown in pink and green respectively in these plots.

The DNA melting curves, obtained by the application of the two state model to the FT-IR spectra of the H-A<sub>3</sub>T<sub>3</sub> complex, are shown in figure 7.5.(b). Fitting these to sigmoidal curves yields a melting temperature for the complex of 83 °C, showing that the melting temperature is stabilised by 24 °C upon binding. It is noted that the application of this model to the FT-IR spectra results in small but non-randomly distributed residuals (blue, Fig.7.5.(b)), showing that the two state model cannot fully explain the melting mechanism of the complex. In order to understand the differences between the actual melting mechanism of the H-A<sub>3</sub>T<sub>3</sub> complex and the two state model the spectral structure of the residuals (blue) near the observed melting temperature of the complex was explored (Fig.7.5.(c)). Comparing the spectral structure of these residuals with the characteristic residuals of the duplex melting via either an unzipping (pink, Figs.7.5.(c)&(d)) or bubble formation mechanism (green, Figs.7.5.(c)&(d)) shows that the experimentally observed residuals are more consistent with the duplex

melting via the unzipping mechanism. This conclusion is further supported by the similarity between the spectral structure of the residuals of the two state model observed for the H-A<sub>3</sub>T<sub>3</sub> complex (Fig.7.5.(c), blue) and the uncomplexed A<sub>3</sub>T<sub>3</sub> sequence (Fig.7.5.(d), blue). This is also consistent with the hypothesis developed for the H-(AT)<sub>3</sub> complex, suggesting binding H33258 to such DNA sequences only results in an elevation of the onset temperature of the melting transition without altering the underlying mechanism.

The melting of the H-A<sub>3</sub>T<sub>3</sub> complex was further studied using 2D-IR spectroscopy, as previously done for both the uncomplexed sequence (Chapter 4) and H-(AT)<sub>3</sub> complex. Applying the two state model to the 2D-IR spectra of this complex yields sigmoidal melting curves, shown in Fig.7.6. Fitting these melting curves yields a melting temperature of 83 °C for the complex, consistent with the 24 °C stabilisation in the melting temperature expected for the H33258:A<sub>3</sub>T<sub>3</sub> complex (Chapter 5).

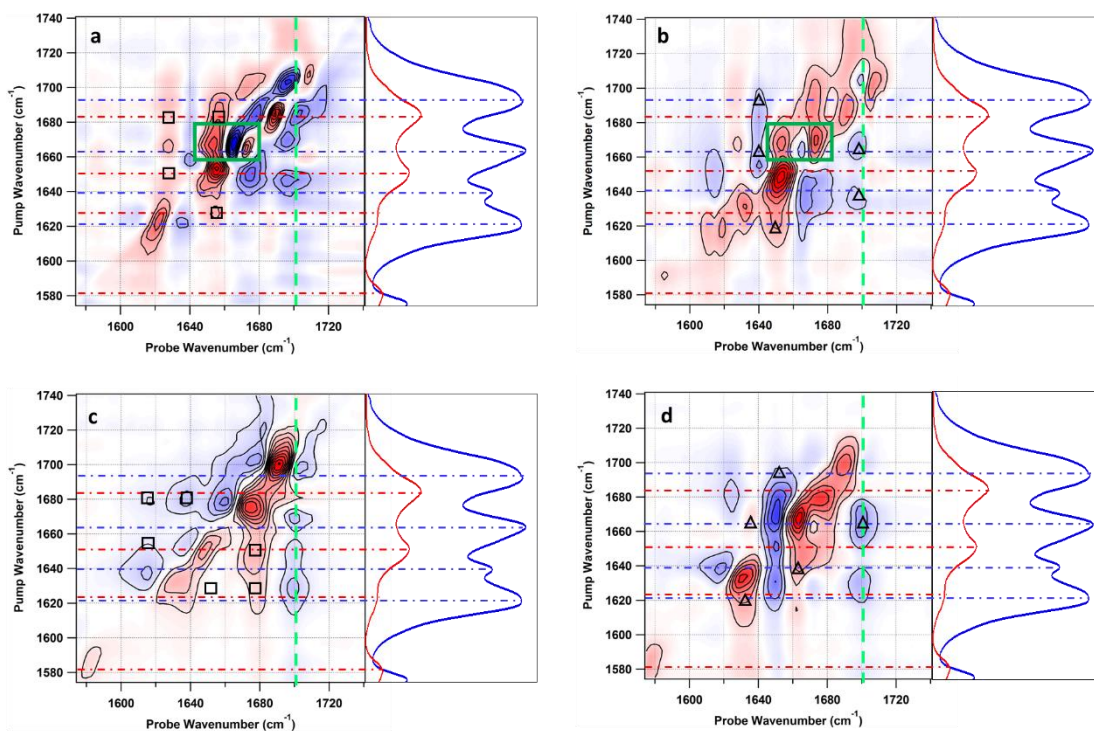


**Figure 7.6:** Melting curve, extracted utilising the two state model, from the 2D-IR spectra of the H-A<sub>3</sub>T<sub>3</sub> complex.

The melting curve obtained when the 2D-IR spectra of the H-A<sub>3</sub>T<sub>3</sub> complex (Fig.7.6) are found to be similar to those obtained from the analysis of the FT-IR spectroscopy. Fitting these melting curves, to sigmoidal curves, yields a melting temperature for the complex of 80 °C, which is noted to be broadly similar to the melting temperature obtained from the analysis of the FT-IR spectra. Overall it is noted even though the binding of the H33258 ligand to the A<sub>3</sub>T<sub>3</sub> sequence results in a 24 °C increase in the melting temperature, it does not result in any

other significant changes in the IR spectroscopy seen as the temperature of the sample increases and the complex melts.

The application of the two state model to the 2D-IR spectroscopy of the H-A<sub>3</sub>T<sub>3</sub> complex is found to produce small but non-randomly distributed residuals (blue, Fig.7.6). The thermal profile of these residuals can be represented by a Gaussian peak centred at approximately 70 °C. The presence of these residuals suggest that the two state model cannot fully describe the melting mechanism underlying the melting transition of the H-A<sub>3</sub>T<sub>3</sub> complex. In order to understand these differences between the two state model and the actual melting mechanism underlying this complex temperature induced difference spectra were generated for the H-A<sub>3</sub>T<sub>3</sub> complex. As for the unbound A<sub>3</sub>T<sub>3</sub> sequence two temperature-induced 2D-IR difference spectra are analysed, one of these difference spectra was generated to reveal the initial changes in the spectrum underlying the initial slope of the extracted melting curves (Fig.7.6) from 20 °C to 50 °C and the second was generated to gain insights into the changes leading to the sigmoidal step observed in the H-A<sub>3</sub>T<sub>3</sub> melting curve from 65 °C to 90 °C. These temperature-induced 2D-IR difference spectra for the H-A<sub>3</sub>T<sub>3</sub> and compared to those obtained for the uncomplexed A<sub>3</sub>T<sub>3</sub> sequence are shown in Fig 7.7.



**Figure 7.7:** Temperature-induced 2D-IR difference spectrum for the H-A<sub>3</sub>T<sub>3</sub> complex a) observed for an increase in temperature from 20 °C to 50 °C and b) observed for an increase in temperature from 65 °C to 90 °C. Temperature-induced 2D-IR difference spectrum for the uncomplexed A<sub>3</sub>T<sub>3</sub> sequence a) observed for an increase in temperature from 20 °C to 50 °C and b) observed for an increase in temperature from 50 °C to 90 °C.

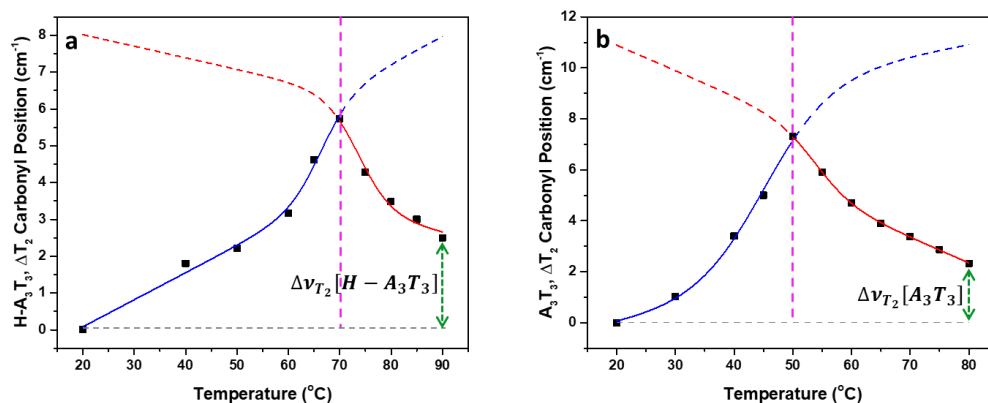
Both the temperature-induced difference spectra for the H-A<sub>3</sub>T<sub>3</sub> complex (Figs.7.7.(a)&(b)) and the unbound A<sub>3</sub>T<sub>3</sub> sequence (Fig.7.7.(c)&(d)) are noted to be broadly similar. The temperature-induced difference spectrum for the changes underlying the initial slope of the melting curve was found to be largely due to the loss of intensity of *ds*-GC modes in the uncomplexed A<sub>3</sub>T<sub>3</sub> sequence (Fig.7.7.(c)). These assignments were found to be supported by the appearance of the cross-peaks (squares, Fig.7.7.(c)). These GC features are also found to be present in the H-A<sub>3</sub>T<sub>3</sub> temperature-induced difference spectrum (Fig.7.7.(a)). Similarly in the temperature-induced difference spectra of the H-A<sub>3</sub>T<sub>3</sub> complex (Fig.7.7.(b)) and A<sub>3</sub>T<sub>3</sub> sequence (Fig.7.7.(d)) focused on the changes occurring during the sigmoidal step observed in the extracted melting curves. Here, the comparison between the differences seen in the H-A<sub>3</sub>T<sub>3</sub> complex and the unbound duplex illustrate that the majority of the changes in the base modes of these spectra arise due to the loss in intensity of *ds*-AT modes. Again these assignments are supported by the presence of cross-peaks in the off-diagonal region (triangles, Figs.7.7.(b)&(d)). These results support the hypothesis that the only impact of

H33258 binding on the melting mechanism of the  $A_3T_3$  sequence is an increase in the temperature at which the duplex starts to melt. This is again fully consistent with the fact that the binding interaction are highly localised to the AT-rich region of the duplex (Chapter 5).

Even though it is noteworthy how similar the temperature-induced difference spectra of the unbound  $A_3T_3$  sequence and H- $A_3T_3$  complex, there is one notable difference. This difference is found in the temperature-induced difference spectrum for the changes in the DNA base modes when the temperature of the complex is increased to 50 °C. This feature, indicated with a green box in Fig.7.7.(a), is thought to be consistent with a shift to higher temperature of the  $AT_{45}$  mode. It is suspected that this changes is due to a relaxation of the induced fit geometry, which the  $A_3T_3$  duplex assumes upon binding (Chapter 5), as the temperature increases and the ligand-DNA H-bonds start to become weaker.

### ***DNA Minor Groove***

The exploration of the temperature-induced 2D-IR spectra of the H-(AT)<sub>3</sub> (Fig.7.3.(a)) and H- $A_3T_3$  (Fig.7.7.(a)) complexes revealed that increasing the temperature to 50 °C results in the  $AT_{25}$  mode shifting to higher frequency. This is thought to be consistent with a decrease in the strength of the hydrogen bonds formed to the  $T_2$  carbonyls from H33258, demonstrating that the changes in these interactions to be studied as a function of temperature. The changes in the position of the  $T_2$  carbonyl stretching mode as the temperature is increased for both the H- $A_3T_3$  complex and the unbound  $A_3T_3$  sequence were extracted from the temperature-induced 2D-IR difference spectra over the experimental temperature range are shown in figure 7.8.



**Figure 7.8:** Change in the position of the T<sub>2</sub> carbonyl stretch of the a) H-A<sub>3</sub>T<sub>3</sub> complex and b) A<sub>3</sub>T<sub>3</sub> sequence with increasing temperature.

As noted in chapter 4, the changes in the position of the T<sub>2</sub> carbonyl as the temperature of the sample is increased is observed to be consistent in the uncomplexed A<sub>3</sub>T<sub>3</sub> sequence (Fig.7.8.(b)) with the loss of the spine of hydration as the melting transition starts and the solvation of the carbonyl moieties as the duplex unwinds and these groups becomes exposed to the surrounding aqueous environment (Chapter 4). In the uncomplexed A<sub>3</sub>T<sub>3</sub> sequence these transitions were determined to occur at 46 °C and 54 °C, respectively (Chapter 4). As was seen for the H-(AT)<sub>3</sub> complex, upon the binding of H33258 to the A<sub>3</sub>T<sub>3</sub> sequence both of these transition temperatures were found to increase to 66 °C and 74 °C, respectively. This increase of 20 °C is broadly consistent with the stabilisation of the duplex due to binding. Additionally it was noted that the overall shift in the position of the T<sub>2</sub> carbonyls,  $\Delta\nu_{T_2}[H - A_3T_3]$  and  $\Delta\nu_{T_2}[A_3T_3]$  are noted to be 2.6 cm<sup>-1</sup> and 2.3 cm<sup>-1</sup>, respectively, consistent with the overall shift to higher frequency of the T<sub>2</sub> carbonyls observed in the FT-IR spectrum of an exclusively AT sequence upon melting.

For the H-A<sub>3</sub>T<sub>3</sub> complex, it is noted that the initial transition is due to the weakening of the ligand-DNA hydrogen bonds formed upon binding instead of the loss of the minor groove spine of hydration. Here, it is interesting to note that this process occurs 20 °C before the ligand is known to dissociate from the duplex. This further supports the idea that the hydrogen bonding formed between the ligand and the sequence do not contribute significantly to the stabilisation of the duplex. This again suggests that the stabilisation of the duplex is largely due to the hydrophobic contacts between the ligand and the walls of the DNA minor groove, and is fully consistent with previous thermodynamic studies carried out on these H33258:DNA complexes.<sup>17,18</sup> It is noted that the formation of these hydrophobic



interactions stabilising the duplex and preventing the sequence from unwinding are also thought to be consistent with the observed increase in duplex rigidity noted in the single molecule AFM study.<sup>64</sup>

In the case of the H-A<sub>3</sub>T<sub>3</sub> complex, the results of the FT-IR and 2D-IR spectroscopy show that the impact of the binding of H33258 does not fundamentally change the melting mechanism of the DNA duplex. Rather it appears that the binding interactions result in the delay of the onset of the melting transition of the duplex by 24 °C. Broadly speaking, when the complex melts a loss in the GC base pairing is seen to occur below the sigmoidal step in the melting curve and the loss of the AT base pairs above. Alongside this loss of the GC base pairing the shifts to lower and higher frequencies of the AT<sub>45</sub> and AT<sub>25</sub> modes respectively, are thought to be due to the relaxation of the induced fit geometry of the bound duplex, as the ligand H-bonds start to weaken with increasing temperature. Finally the T<sub>2</sub> carbonyls are found to report on the changes in the ligand H-bonding as the temperature increases, revealing that these interactions start to weaken before the complex melts. This is found to be in agreement with thermodynamic studies of such complexes suggesting that the stabilisation of the duplex is primarily due to the hydrophobic interactions between the walls of the minor groove and the ligand.<sup>17,18</sup> Overall, it is noted that the melting mechanism is not fundamentally altered by the binding of H33258 with the only change in the spectroscopy resulting from the relaxation of the induced fit geometry of the bound duplex.

### 7.3.3 N<sub>3</sub>-bBI complexes

As seen for the H33258:DNA complexes it is thought that a similar analysis of the impact on the melting mechanisms of the (AT)<sub>3</sub> and A<sub>3</sub>T<sub>3</sub> sequences due to the binding of the de novo N<sub>3</sub>-bBI could be carried out. These complexes are thought to be much more challenging to analyse as in the case of the A<sub>3</sub>T<sub>3</sub> sequence only 23% of the duplexes are bound (Chapter 6) lead to the temperature-induced spectra being a combination of the changes occurring the unbound and bound duplexes as the temperature of the sample is increased. It is noted that due to the fact that this complex follows the same induced fit binding mechanism as seen in the H-A<sub>3</sub>T<sub>3</sub> complex, it is possible that the result of such an analysis would show that the impact of the binding of the N<sub>3</sub>-bBI ligand only results in an increase in the onset temperature of the melting mechanism. In the case of the (AT)<sub>3</sub> sequence the binding of the N<sub>3</sub>-bBI ligand is thought to follow a non-specific interaction thought to be present within the complex (Chapter 6). Such a non-specific binding interaction is thought to result in the ligands being

bound to the DNA sequence at a multitude of different positions. The decrease in the strength of H-bonds between the ligand and the (AT)<sub>3</sub> sequence at such a large of different positions throughout the duplex as well as the changes in the duplexes due to the loss of the Watson-Crick H-bonding contributing to any temperature-induced 2D-IR difference spectrum for this complex, making it challenging to understand these changes. For these reasons the impact of this ligand on the melting of these duplexes is not explored here.

## 7.4 Conclusions

In conclusion, we have revealed that the formation of complexes between these two different DNA sequences with the H33258 ligand does not fundamentally change the mechanism underlying the melting transition of the duplex. Rather the only impact of the binding is an increase in the temperature at which the bound duplex starts to melt (by 16 °C and 24 °C for the (AT)<sub>3</sub> and A<sub>3</sub>T<sub>3</sub> sequences, respectively). Interestingly in these H33258:DNA complexes studying the AT<sub>25</sub> mode reveals that the ligand-DNA H-bonds are lost at much lower temperatures than the observed dissociation temperatures for these complexes, providing further evidence that the ligand-duplex H-bonds within the complex do not significantly contribute to the overall stabilisation of the complex. Together these results indicate that the increase in the melting temperature upon binding is predominantly due to the ligand-duplex hydrophobic interactions.

## 7.5 Experimental

### *2D-IR and FT-IR Spectroscopy*

Lyophilised, salt-free DNA oligonucleotides were obtained from Eurofins; H33258, D<sub>2</sub>O, DMSO, monobasic and dibasic sodium phosphate were obtained from Sigma-Aldrich. N<sub>3</sub>-bBI was synthesized and purified by Dr. J. J. May (Burley Group). All chemicals were used without further purification. All samples were prepared using pD7 phosphate buffer solution to a final duplex:ligand ratio of 1:1 and annealed at 90 °C for 10 minutes. For all IR measurements, samples were held between two CaF<sub>2</sub> windows separated by a polytetrafluoroethylene spacer of 50 μm thickness in a temperature controlled Harrick cell. FTIR measurements were carried out using a Bruker Vertex 70 spectrometer at a resolution of 1 cm<sup>-1</sup> with sample concentrations of 2.5 mM (A<sub>3</sub>T<sub>3</sub> duplex/H-A<sub>3</sub>T<sub>3</sub> complex/N<sub>3</sub>-A<sub>3</sub>T<sub>3</sub> complex) or 5 mM ((AT)<sub>3</sub> duplex/H-(AT)<sub>3</sub> complex/N<sub>3</sub>-(AT)<sub>3</sub> complex). 2D-IR spectra were collected using the ULTRA FT-2D-IR spectrometer.<sup>372,373</sup> The IR pulses used had a temporal duration of ~ 100fs; a center frequency of 1650 cm<sup>-1</sup> and a bandwidth of ~300 cm<sup>-1</sup>, at a repetition rate of 10 kHz. FT-2D-

IR measurements were carried out at concentrations of 1.25 mM ( $A_3T_3$  duplex/ $H-A_3T_3$  complex/ $N_3-A_3T_3$  complex) or 2.5 mM ( $(AT)_3$  duplex/ $H-(AT)_3$  complex/ $N_3-(AT)_3$  complex).

### UV-visible Spectroscopy

The samples were held in a demountable Harrick cell utilising  $CaF_2$  windows and a 50  $\mu$ m polytetrafluoroethylene spacer for all the UV-visible measurements. UV-visible measurements were carried out using a Perkin-Elmer Lambda 25 at a resolution of 1 nm with sample concentrations of 2.5 mM ( $H-A_3T_3$  / $N_3-A_3T_3$  complex) or 5 mM ( $H-(AT)_3$  / $N_3-(AT)_3$  complex).

## 7.6 References

- 
- (357) Burmistrova, A.; Fresch, B.; Sluysmans, D.; De Pauw, E.; Remacle, F.; Duwez, A.-S.; Force measurements reveal how small binders perturb the dissociation mechanisms of DNA duplex sequences, *Nanoscale*, **2016**, *8*, 11718-11726.
- (358) Haq, I. Part II: The Thermodynamics of Drug-Bipolymer Interaction, Thermodynamics of Drug-DNA Interactions. *Archives of Biochemistry and Biophysics*, **2002**, *403*, 1-15.
- (359) Haq, I.; Ladburry, J. E.; Chowdhry, B. Z.; Jenkins, T. C.; Chairs, J. B. Specific Binding of Hoechst33258 to the  $d(CGCAAATTTGCG)_2$  Duplex: Calorimetric and Spectroscopic Studies. *J. Mol. Biol.* **1997**, *271*, 244-257.
- (360) Jin, R.; Breslauer, K. J.; Characterization of the minor groove environment in a drug-DNA complex: bisbenzimidazole bound to the poly[d(AT)].poly[d(AT)]duplex., *Proc. Natl. Acad. Sci. U.S.A.*, **1988**, *85*, 8939-8942.
- (361) Hunt, N. T. 2D-IR spectroscopy: Ultrafast Insights into Biomolecule Structure and Function, *Chem. Soc. Rev.*, **2009**, *38*, 1837-1848.
- (362) Park, S.; Kwak, K.; Fayer, M. D. Ultrafast 2D-IR Vibrational Echo Spectroscopy: A Probe of Molecular Dynamics, *Laser Phys. Lett.*, **2007**, *10*, 704-718.
- (363) Baiz, C. R.; McRobbie, P. L.; Anna, J. M.; Geva, E.; Kubarych, K. J. Two-Dimensional Infrared Spectroscopy of Metal Carbonyls, *Accounts of Chemical Research*, **2009**, *42*, 1395-1404.
- (364) Peng, C. S.; Jones, K. C.; Tokmakoff, A. Anharmonic Vibrational Modes of Nucleic Acid Bases Revealed by 2D IR Spectroscopy, *J. Am. Chem. Soc.*, **2011**, *133*, 15650-15660.
- (365) Yang, M.; Szyk, L.; Elsaesser, T. Femtosecond Two-Dimensional Infrared Spectroscopy of Adenine-Thymine Base Pairs in DNA Oligomers, *J. Phys. Chem. B*, **2011**, *115*, 1262-1267.
- (366) Greve, C.; Elsaesser, T. Ultrafast Two-Dimensional Infrared Spectroscopy of Guanine-Cytosine Base Pairs in DNA Oligomers, *J. Phys. Chem. B*, **2013**, *117*, 14009-14017.
- (367) Krummel, A. T.; Zanni, M. T. DNA Vibrational Coupling Revealed with Two-Dimensional Infrared Spectroscopy: Insight into Why Vibrational Spectroscopy is Sensitive to DNA Structure, *J. Phys. Chem. B*, **2006**, *110*, 13991-14000.

- 
- (368) Sanstead, P. J.; Stevenson, P.; Tokmakoff, A. Sequence-Dependent Mechanism of DNA Oligonucleotide Dehybridization Resolved through Infrared Spectroscopy. *J. Am. Chem. Soc.*, **2016**, *138*, 11792-11801.
- (369) Hithell, G.; Gonzalez-Jimenez, M. Greetham, G. M.; Donaldson, P. M.; Towrie, M.; Parker, A. W.; Burley, G. A.; Wynne, K.; Hunt, N. T.; Ultrafast 2D-IR and Optical Kerr Effect Spectroscopy Reveal the Impact of Duplex Melting on the Structural Dynamics of DNA, *Phys. Chem. Chem. Phys.*, **2017**, DOI: 10.1039/C7CP00054E.
- (370) Ramakers, L. A. I.; Hithell, G.; May, J. J.; Greetham, G. M.; Donaldson, P. M.; Towrie, M.; Parker, A. W.; Burley, G. A.; Hunt, N. T.; 2D-IR Spectroscopy Shows that Optimised DNA Minor Groove Binding of Hoechst33258 Follows an Induced Fit Model, *J. Phys. Chem. B.*, **2017**, *121*, 1295-1303.
- (371) Ramakers, L. A. I.; May, J. J.; Hithell, G.; Greetham, G. M.; Donaldson, P. M.; Towrie, M.; Parker, A. W.; Burley, G. A.; Hunt, N. T.; Separating Structural and Dynamic Factors in DNA-ligand Binding (*in preparation*)
- (372) Shaw, D. J.; Adamczyk, K., Frederix, P. W. J. M.; Simpson, N.; Robb, K.; Greetham, G. M.; Towrie, M.; Parker, A. W.; Hoskisson, P. A.; Hunt, N. T. Multidimensional Infrared Spectroscopy Reveals the Vibrational and Solvational Dynamics of Isoniazid, *J. Chem. Phys.*, **2015**, *142*, 212401.
- (373) Greetham, G. M.; Burgos, P.; Cao, Q.; Clark, I. P.; Codd, P. S.; Farrow, R. C.; George, M. W.; Kogimtzis, M.; Matousek, P.; Parker, A. W.; *et al.* M. ULTRA: A Unique Instrument for Time-Resolved Spectroscopy. *Applied Spectroscopy*, **2010**, *12*, 1311-1319.

---

## General Conclusions

Chapter 8

---

In this thesis the exploration of the versatility of 2D-IR to study aspects of DNA, including the melting mechanisms and binding interactions between DNA and small molecules, was approached by characterising both a non-natural asymmetric azide stretch and the DNA base modes. After this initial characterisation these modes were applied to explore and gain insights into the melting of two short DNA oligonucleotides, the formation of complexes between DNA duplexes and two different *bis*-benzimidazole ligands, probe the nature of the DNA minor groove environment that these ligands experience upon binding and finally the impact of these ligands on the melting mechanisms of these two short DNA oligonucleotides. Here, the conclusions of these experimental chapters are summarised and the recommendations for further work are outlined.

## 8.1 Overall Conclusions

In Chapter 3, the response of the asymmetric azide stretch of benzyl azide to changes in its local environment was characterised. The absorption band of this asymmetric azide stretch was found to be complicated due to the presence of two accidental Fermi resonances, one centred at a slightly higher frequency than the azide mode and the other centred at a slightly lower frequency than the azide mode. Increasing the electrostatic potential of the local environment results in shifts to lower frequencies of both Fermi resonances, additionally a decrease in the vibrational lifetime of the asymmetric azide stretch was observed. Increases in the hydrogen bonding strength of the local environment resulted in a shift to higher frequency of the asymmetric azide stretch as well as increasing the lifetime of the low-frequency shoulder and decreasing the lifetime of the high frequency shoulder. Additionally the 2D-IR response of the asymmetric azide stretch illustrated that the coupling constant between the low-frequency shoulder ( $\beta_L$ ) and the azide mode varied exponentially with changes in the nature of the molecular environment, whereas the coupling constant between the high-frequency shoulder ( $\beta_H$ ) and the azide mode was found to decrease linearly with increasing hydrogen bonding strength. All of these changes in the response of the asymmetric azide stretch of benzyl azide could be used to extract relevant molecular information about the local environment, including both the electrostatic potential and the hydrogen bonding strength. The responses of the Fermi resonances as the local environment changes suggested that only the mode underlying the high-shoulder participates in hydrogen bonds. In combination with the simulated IR spectrum this suggests the low-frequency and high-frequency shoulders are the result of a Fermi resonance between the asymmetric azide

---

stretch and overtones of a CH<sub>2</sub> scissor mode of a ring mode respectively. Overall, the results of the characterisation of the asymmetric azide stretch response of benzyl azide demonstrate that such a moiety is well suited to act as a non-natural IR probe.

In addition to the characterisation of the non-natural asymmetric azide stretch, the DNA base modes and how these modes change upon changes in the DNA structure due to both changes in the sequence and as the duplexes undergo a melting transition were explored in Chapter 4. Overall the results show the utility of 2D-IR spectroscopy as a technique for the investigation of DNA sequences and the changes that an alteration in the sequence can impart on both the structure of the duplex and the mechanism by which it dissociates as it undergoes thermal denaturation. Using the DNA base vibrational modes as natural IR probes, difference spectroscopy illuminates that when the A-tract within the DNA oligonucleotide is replaced with its alternating counterpart there is a 4.8 kJmol<sup>-1</sup> decrease in the strength of Watson-Crick hydrogen bonds involving the T<sub>4</sub> carbonyl. This is accompanied by a loss in the number of hydrogen bonds between the spine of hydration and the T<sub>2</sub> carbonyls. These changes in these two vibrational modes show that the alteration of the order of the AT region of the DNA oligonucleotide leads to the loss of three-centred inter-strand hydrogen bonds as well as a decrease in both the order and number of water molecules present in the minor groove spine of hydration. Additionally this highlights the sensitivity of 2D-IR spectroscopy to perturbations in the structure of DNA duplexes and alterations in the hydrogen bonding strengths. The extension of the 2D-IR study to the thermal denaturation of the two DNA sequences allows the mechanism underlying this transition to be studied. This study determined that for both of these short oligonucleotides this follows an unzipping mechanism. This mechanism was determined to consist of the fraying of the GC ends of the duplexes. The fraying of the GC ends and the loss of the minor groove spine of hydration were found to occur within the same temperature range, suggesting that the fraying of the duplexes leads to the loss of the spine of hydration for both sequences. In the A<sub>3</sub>T<sub>3</sub> sequence it was noted that this loss of the spine of hydration was followed by the unwinding of the duplex, occurring at approximately 54 °C. The melting mechanism of the A<sub>3</sub>T<sub>3</sub> sequence then terminates with the loss of the W-C hydrogen bonds between the AT base pairs at 64 °C, completing the unzipping of the strands. In contrast to this it was determined that following the loss of the spine of hydration the (AT)<sub>3</sub> sequence unwinds and the AT W-C hydrogen bonds are lost in a concerted unwinding/unzipping process which occurs over the temperature range of 50 °C to 55 °C. These mechanisms highlight the change in the melting

---

mechanism caused by the presence of an A-tract which introduces an unwound duplex intermediate into this mechanism and so delaying the final unzipping of the strands. This further demonstrates the versatility of 2D-IR as a technique to gain a more fundamental understanding of important DNA processes.

Taken together the conclusions from Chapters 3 and 4 illustrate that together the IR spectroscopies of the asymmetric azide stretch and the DNA base modes can be used to study the local environment surrounding the azide moiety and structural changes within the DNA duplex. In the following two experimental chapters the characterisation of these vibrational mode obtained in these chapters is then used to study the complex formed between two short DNA oligonucleotides and *bis*-benzimidazole minor groove binders. In the first of these experimental chapters, chapter 5, FT-IR and 2D-IR spectroscopy of the DNA base vibrational modes was used to explore the binding interactions within the complexes between H33258 and two DNA sequences. In chapter 5, 2D-IR shows that binding H33258 to two DNA sequences leads to shifts in vibrational modes associated with specific AT base pairs that are due to the loss of the spine of hydration and formation of direct H-bonding between DNA and ligand as well as alterations in the propeller twist induced by the ligand locating in the minor groove. Comparison of binding to A-tract and alternating DNA sequences revealed that binding to A<sub>3</sub>T<sub>3</sub> results in loss of the ordered propeller twist arrangement of bases found in the uncomplexed DNA. This is not replicated to the same extent in the alternating sequence and we propose that these structural changes constitute an induced fit type interaction that facilitates superior accommodation of H33258 and increased hydrophobic interactions between ligand and DNA. This contradicts current pictures which treat H33258 binding as a rigid body interaction and complements the entropic release of water from the minor groove. Finally, the results fully demonstrate 2D-IR capabilities to simplify quantification of solution phase DNA-binding.

Building from the conclusions the experiments carried out on the binding interactions between two short DNA oligonucleotides and the archetypal *bis*-benzimidazole minor groove binder, H33258, the same analysis was carried out on the complexes formed between these DNA oligonucleotides and a specially synthesized azide derivative of H33258. In addition to analysis the utilisation of the azide derivative allowed the minor groove environment that such a ligand encounters upon binding to be explored. In this experimental chapter it was found that the new ligand (N<sub>3</sub>-*b*BI) is found to be more selective for A-tract DNA than H33258.



---

Characterisation of the binding to the A-tract and alternating sequences shows that the stabilisation changes from 24 °C to 7 °C (versus 24 °C to 16 °C for H33258). NMR and 2D-IR spectroscopy show that A-tract binding of the ligand is similar to the induced fit model derived for H33258 while binding to the alternating sequence is non-specific for the N<sub>3</sub>-bBI ligand, involving both AT and GC base pairs. The azide probe shows mobile water near the azide except when buried in the minor groove, as seen in the N<sub>3</sub>-A<sub>3</sub>T<sub>3</sub> complex. Within the minor groove the azide encounters a unique H-bonding environment with exceptionally slow dynamics, suggesting the probe is exploring the remaining spine of hydration within the A<sub>3</sub>T<sub>3</sub> minor groove. Overall we show that the impact of the azide moiety is not enough to overcome the enthalpic benefits of induced fit binding but the rigid body interaction of *bis*-benzimidazole binding to alternating sequences is significantly perturbed by the azide.

In the final experimental chapter presented as part of this thesis all of the previous analysis is applied to the melting of the bound DNA sequences is explored. In this experimental chapter, we have revealed that the formation of complexes between these two different DNA sequences with the H33258 ligand does not fundamentally change the mechanism underlying the melting transition of these duplexes. Rather the only impact of the binding is an increase in the temperature at which the bound duplex starts to melt (by 16 °C and 24 °C for the (AT)<sub>3</sub> and A<sub>3</sub>T<sub>3</sub> sequences respectively). Interestingly in these H33258:DNA complexes studying the AT<sub>25</sub> mode reveals that the ligand-DNA H-bonds are lost at much lower temperatures than the observed dissociation temperatures for these complexes, providing further evidence that the ligand-duplex H-bonds within the complex do not significantly contribute to the overall stabilisation of the complex. Together these results indicate that the increase in the melting temperature upon binding is predominantly due to the ligand-duplex hydrophobic interactions.

## 8.2 Further Work and Recommendations

In this thesis 2D-IR spectroscopy has been applied to the DNA modes and the asymmetric azide stretch in order to gain insight into the melting mechanism of DNA sequences, the formation of the DNA:*bis*-benzimidazole complexes and the impact of these ligands on the melting transition. Moving beyond the results explored and discussed in this thesis there are several recommendations for further work in this area. The first of these recommendations is centred on the work carried out using the non-natural IR probe. In this thesis the asymmetric azide stretch was chosen as it is known to be located within the spectral clear

---

window for biological systems (1800 – 2300 cm<sup>-1</sup>), however it was found that the lifetime of this mode was rather short and so here it is recommended that further work could be carried out utilising another non-natural mode which exhibits a much longer vibrational mode (*e.g.* a thiocyanate moiety). Carrying out further work on such a derivative of the H33258 ligand would be interesting as it would allow the dynamics of environment which the ligand encounters within the DNA minor groove to be explored for a longer period of time and potentially might allow the reduced dynamics of the remaining spine of hydration to be more fully studied.

In chapters 5 and 6, the DNA base modes were used to study the interactions between the H33258 and N<sub>3</sub>-bBI ligands and two different DNA sequences. One of these sequences (the A-tract sequence) represented the preferred binding target of these ligands whereas the other (the alternating AT sequence) was known to be a sub-optimal binding target. Contrasting the binding of these two different ligands showed that the addition of a hydrophilic azide moiety to the archetypal ligand resulted in an enhanced selectivity for the A-tract sequence. This highlights that the generation of 2D-IR difference spectra can be used to assess how minor alterations of the ligand structure affect the sequence selectivity of these small molecules. Therefore, it is proposed that it would be interesting to carry on this particular path by carrying out a series of small structural alterations and utilise 2D-IR spectroscopy to determine changes in the sequence selectivity. It is hoped that if these structural alterations are done in a systematic way that it might be possible to use the gathered information to extract a series of rational design rules for such minor groove binders.

In addition to being used to probe the interactions formed within the DNA complexes, the DNA base modes were also used to extract the melting mechanism of both uncomplexed and complexed DNA sequences. Within this exploration the T<sub>2</sub> carbonyl was utilised as a probe of the changes in the minor groove of the DNA sequences as the melting transition occurred. However, it was noted that the spectral region in which the DNA base modes are found is heavily congested. In particular the AT<sub>2</sub> and G<sub>5</sub>C<sub>5</sub>(-) modes are noted to be difficult to separate from each other, meaning that it is difficult to extract data from the T<sub>2</sub> carbonyls. Here it is recommended that further work should be carried out for both the complexed and uncomplexed DNA sequences where the carbonyls on the C and the G bases are isotopically labelled. This labelling will result in this vibrational mode shifting to a lower frequency,

---

allowing the AT<sub>2</sub> mode to be isolated and be more thoroughly studied as these complexes and duplexes melt. It is hoped that this will allow the changes in the minor groove to be more closely studied and would also allow spectral diffusion within the minor groove to be studied in mixed DNA sequences and their complexes for the first time at a range of temperatures. Additionally, the labelling of the C and G carbonyls should result in the G<sub>5</sub>C<sub>5</sub>(+) mode shifting down to lower frequency. Assuming a large enough shift in these modes is obtained it is thought that it should be possible to track the changes in the T<sub>4</sub> carbonyl (*e.g.* the AT Watson Crick H-bonding) and the G/C carbonyls (*e.g.* the GC Watson Crick H-bonding) as the temperature is increased. In both the complexed and uncomplexed sequences this labelling would allow the melting mechanism to be directly probed and in the case of the induced fit mechanism proposed for the H-A<sub>3</sub>T<sub>3</sub> and N<sub>3</sub>-A<sub>3</sub>T<sub>3</sub> complexes the isolation of the T<sub>4</sub> carbonyl would allow the relaxation of this DNA fit geometry (as proposed in chapter 7) to be studied directly. Finally, it is recommended that this work is expanded to further DNA sequences and other archetypal minor groove binders (*e.g.* polyamides) to build a dataset which might allow the development of general rational design principals for tailored sequence-selective minor groove binders.