Investigating the binding mode of azide labelled derivatives of Hoechst 33258 by NMR, UV-Vis and IR spectroscopy.

John May

PhD Thesis

Date: 30 September 2016

Academic Supervisors: Dr. Glenn. A. Burley and Dr. Neil. T. Hunt

Declaration of Authenticity

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

Signed: ______________________ (John J. May)

Date: 30th September 2016.

Acknowledgments

I would firstly like to thank the University of Strathclyde for their generosity in providing a scholarship and the opportunity to pursue a PhD in the University.

I am grateful to Dr. Glenn Burley and Dr. Neil Hunt for firstly offering me the opportunity to study this project, and would like to acknowledge them for their support, advice, suggestions and patience. Special thanks go to Lennart Ramakers and Gordon Hithell within the Hunt group for firstly their help with getting me up to speed with the acquisition and interpretation of solution-phase infrared spectra – a technique I was not familiar with upon commencement of this project. I would also like to thank Lennart for supplying infrared data through the course of this project. I would also like to thank Dr John Parkinson for his help with the acquisition of NMR data through the course of this project.

I would also like to thank the technical staff at the University of Strathclyde for their help with the use of the analytical and other instruments in the department, especially Mr Craig Irving, Mrs Patricia Keating and Mr Gavin Bain. This project would be impossible without their help.

I would like to thank my family who without their help and support, completion of this thesis would not have been possible. A special thank you goes to Dr Ioana Sovago for all her help, love, support and help with drawing crystal structures through the most difficult of times through my PhD.

I would also like to apologise to the unfortunates who had the bad luck to share a lab or office with me over the course of my PhD and as such, had to put up with my terrible banter. Sorry.

Abstract

This thesis describes attempts to develop a method for studying the dynamics of dsDNA: small molecule ligand association and dissociation through 2D-IR spectroscopy. The strategy employed was to synthesise azide-bearing derivatives of the archetypal minor groove binding ligand, Hoechst 33258, with the azide acting as a reporter functional group.

Chapter 1 describes both the state of the art, and also what is unknown regarding the processes of association and dissociation of small molecules from dsDNA, and how this project aims to gain a fuller understanding of the dynamics of association between dsDNA and a small molecule MGB. The clinical significance of minor groove binders is also discussed.

Chapter 2 describes the synthesis of two azide-bearing derivatives of H33258 through an amide coupling strategy, and the significant problems encountered with the purification of these compounds. Future alternative pathways to these compounds are proposed. An investigation into the spectroscopic properties of the azide functional groups in the free compounds is also presented.

Chapter 3 describes the investigation of the utility of the azide functional group through a comparative study of the thermal dissociation of these compounds from dsDNA. It was found that the utility of the azide was dependent on the position of the azide within the molecule, and is sensitive to the changes in solvation of the minor groove of dsDNA. One compound in particular exhibited a marked change in both the shape and the intensity of the azide absorption band and was commensurate with the melting temperature, T_m of the complex.

Chapter 4 describes the investigation into the structural origins of the marked difference in the azide absorption band of one MGB when in complex with a dsDNA oligomer through structural characterization of this complex by NMR spectroscopy. It was found that the azide functional group is in close proximity to the exocyclic amine of guanosine, and this specific interaction is proposed to give rise to the observed changes in the azide absorption band. The orientation of the molecule was found to be opposite to that reported for H33258, a reason why this is the case is also proposed.

Abbreviations

Abbreviation	Meaning
DNA	deoxyribonucleic acid
ds	Double-Stranded
MGB	Minor Groove Binder
ss	Single-Stranded
A	Adenosine
G	Guanosine
T	Thymidine
C	Cytidine
U	Uridine
RNA	Ribonucleic acid
mRNA	messenger RNA
HDAC	Histone Deacetylase protein
TBP	TATA-box Binding Protein
DBD	DNA-binding Domain
RNAP II	RNA Polymerase II
pTEFb	positive transcription elongation factor
EXAFS	Extended X-ray Absorption Fine Structure
Arg	Arginine
Asn	Asparagine
NpN	Phosphate base step, N = A, G, T, C, U
Ile	Isoleucine
K	Equilibrium binding constant
k	Rate constant

Abbreviation	Meaning
G	Gibbs free energy
H	Enthalpy
S	Entropy
R	Universal Gas Constant
T	Temperature
ITC	Isothermal Titration Calorimetry
EDTA	Ethylene Diamine Tetraacetic Acid
C_p	Heat Capacity at Constant Pressure
UV	Ultraviolet
HOMO	Highest Occupied Molecular Orbital
LUMO	Lowest Unoccupied Molecular Orbital
A	Absorbance
ϵ	Molar Extinction Coefficient
l	Path length
c	Concentration
$I_{(0)}$	Intensity (of incident beam)
T_m	Melting Temperature of dsDNA
NMR	Nucleic Magnetic Resonance
IR	Infrared
COSY	Correlation Spectroscopy
TOCSY	Total Correlation Spectroscopy
NOESY	Nuclear Overhauser Effect Spectroscopy
b.p.	base pair

Abbreviation	**Meaning**
ppm	parts per million
nOe	nuclear Overhauser effect
V	Vibrational energy
k_f	Force constant
x	Internuclear distance from equilibrium
c	Speed of light
μ	reduced mass
h	Planck's constant
D_e	Potential minimum
$\widetilde{G}$	Allowed Absorption Energies
$\widetilde{v}$	Vibrational Spectroscopic term
x_e	Anharmonicity constant
τ	Coherence time
T_w	Waiting time
NAD	Nicotinamide dinucleotide
DIEA	Diisopropylethylamine
eq.	equivalents
HATU	1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxide hexafluorophosphate
Boc_2O	Di*tert*-butyl pyrocarbonate
MeOH	Methanol

Abbreviation	Meaning
DMF	dimethyl formamide
Boc-	*tert*-butyl carbamate
DCM	Dichloromethane
DMAPA	3,3-dimethylaminopropylamine
RP-HPLC	Reversed-phase High performance liquid chromatography
RF	Radio Frequency
Im	Imidazole
FT-IR	Fourier Transform- Infrared
Fmoc-	9-fluorenylmethoxycarbonyl
EtOH	ethanol
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
pK_b	Base dissociation constant
pK_a	Acid dissociation constant
DCC	Dicyclohexylcarbodiimide
DCU	Dicyclohexylurea
IPA	Isopropyl alcohol
DMSO	Dimethyl sulfoxide
ATR	Attenuated Total Reflection
ESI	Electrospray Ionisation
MALDI-TOF	Matrix-assisted laser desorption ionisation - Time of flight

Abbreviation	Meaning
R_f	Retardation Factor
MS	mass spectrometry
δ	Chemical Shift
λ	wavelength
ODN	Oligodeoxyribonucleotide
UV-Vis	Ultraviolet-Visible

Table of Contents **Page Number**

1. Introduction.

1.1. Structure and Function of DNA:

Deoxyribonucleic acid (DNA) is found in almost every type of living cell, and acts as the repository of genetic information necessary to regulate every cellular process. This information is firstly transcribed into a corresponding RNA template, and after several processing events, mature mRNA is transported out of the nucleus into the cytoplasm, where protein synthesis takes place, known as translation (Fig. 1.1).[1] The process of gene expression resulting in the synthesis of proteins can be exogenously regulated by small molecules that either bind noncovalently or covalently with select subsets of duplex DNA sequences.[2] The aim of this project is to study the molecular processes that govern non-covalent interactions between small molecules and duplex DNA. The ability to selectively and reliably regulate gene expression by small molecules has shown great promise as both anticancer[3] and antibacterial agents,[4] and some compounds have found use as therapeutic agents. One class of small molecules can regulate gene expression by binding double-stranded DNA (dsDNA) in the minor groove (referred to minor groove binders, or MGBs) and show particular promise as medical and biological tools.[1] However, the precise dynamic mechanisms that govern the interactions between dsDNA and an MGB are not well understood because these interactions occur on picosecond timescales, too quickly to be studied by traditional biochemical and spectroscopic techniques.[5] Ultrafast two-dimensional IR spectroscopy will be used to probe the dynamics of MGB-DNA interactions in real time.[6] The long-term goal of this work is to achieve a fuller understanding of the fundamental dynamics of DNA-ligand interactions which will enable the rational design of more effective DNA binding therapeutics.

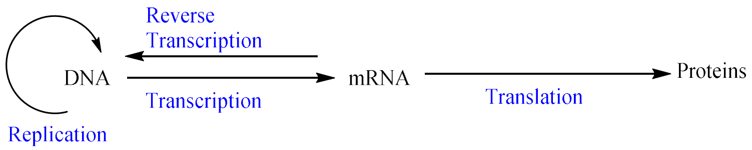


Figure 1.1. The central dogma of molecular biology.

1.1.1. Structure of single-stranded DNA (ssDNA).

The primary structure of ssDNA is a linear polymer of small heterocyclic aromatic molecules, referred to as nucleotides. There are four nucleotides present in DNA; these are Adenosine, **1.1**, (A), Guanosine **1.2** (G), Thymidine **1.3** (T), and cytidine **1.5** (C) (Fig. 1.2) Uridine, **1.4**, is found exclusively in RNA.[7] Nucleosides (Fig. 1.2) are formed when a nucleotide is covalently bonded to a molecule of ribose. Nucleosides are covalently linked through phosphodiester bonds between the 3’-OH of one nucleoside to the 5’-OH of the next.

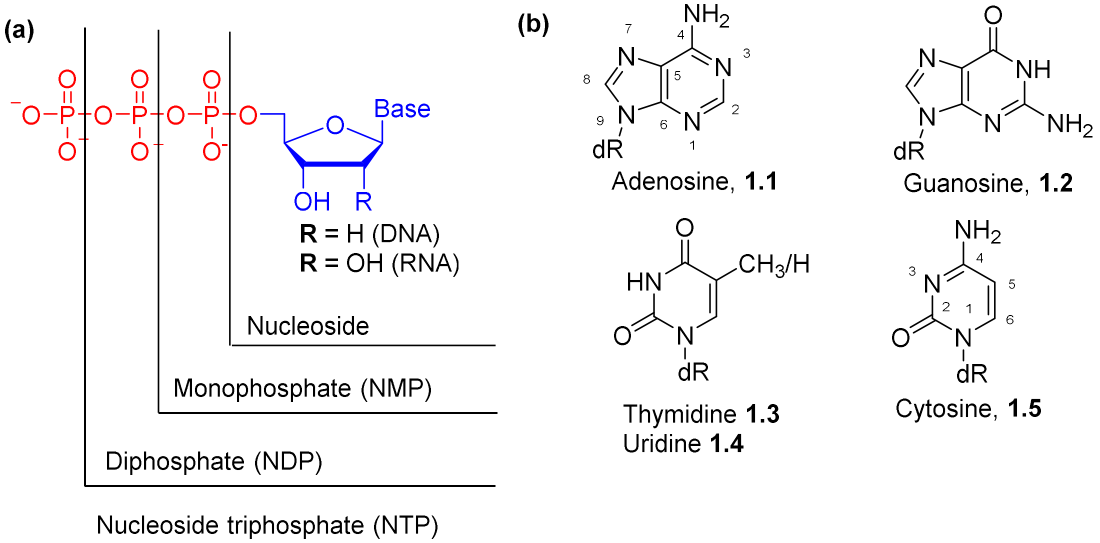


Figure 1.2. (a) General structure of a nucleoside.[155] **(b)** Structures of the four natural DNA and RNA bases. Purine and pyrimidine numbering schemes are shown. dR refers to the sugar deoxyribose (blue).[155] Reprinted by permission from Macmillan Publishers Ltd.: Nature Nanotechnology © 2008.

Biosynthesis of DNA is driven by the exergonic hydrolysis of the triphosphate bond (Fig. 1.2) to inorganic pyrophosphate and the nucleoside monophosphate, which is incorporated into the nascent ssDNA strand.[8]

1.1.2. Structure of double-stranded DNA (dsDNA)

Cellular DNA exists as an antiparallel double helix, (Figs. 1.3, 1.4) a structure first proposed by Watson and Crick from X-ray diffraction data in 1953.[9] The most dominant of the non-covalent interactions that lead to duplex formation are Watson-Crick base pairings. These hydrogen bond networks will pair adenosine to thymidine (uridine in RNA) and guanosine to cytidine (Fig. 1.3). Adenosine forms two hydrogen bonds to thymidine, whereas guanosine forms three hydrogen bonds to cytidine. The heteroatoms in the nucleobases act as H-bond donors, and the electropositive H-atoms act as acceptors (Fig. 1.3). Formation of this hydrogen-bonded complex typically leads to formation of a B-type duplex.

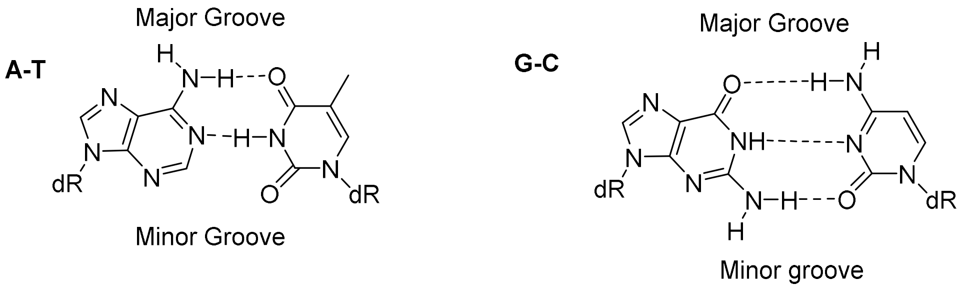


Figure 1.3: Watson-Crick base-pairing between two strands of ssDNA.[52]

1.1.3. Structure of duplex DNA: A-type, B-type, and Z-type DNA.

Three major helical conformations exist for duplex DNA, A, B, and Z-DNA. B-DNA is the conformation of DNA predominantly observed in nature. B-DNA is defined as a right-handed double-helix which has one helical turn every 10 nucleotides, for a distance of 36 Å.[8] An eleventh nucleotide, referred to as an internucleotide, also exists between each turn of the helix (Fig. 1.4). The hydrophobic bases stack on top of each other, facing away from the solvent, whereas the hydrophilic phosphodiester backbone points out towards the solvent.[7] This is an entropic effect, as solvation of the relatively hydrophobic bases is disfavoured with respect to aggregation of water and stacking of the bases.[7] A further enthalpic benefit toward formation of dsDNA comes from π-stacking interactions between adjacent bases, giving rise to formation of the familiar double-helical shape.

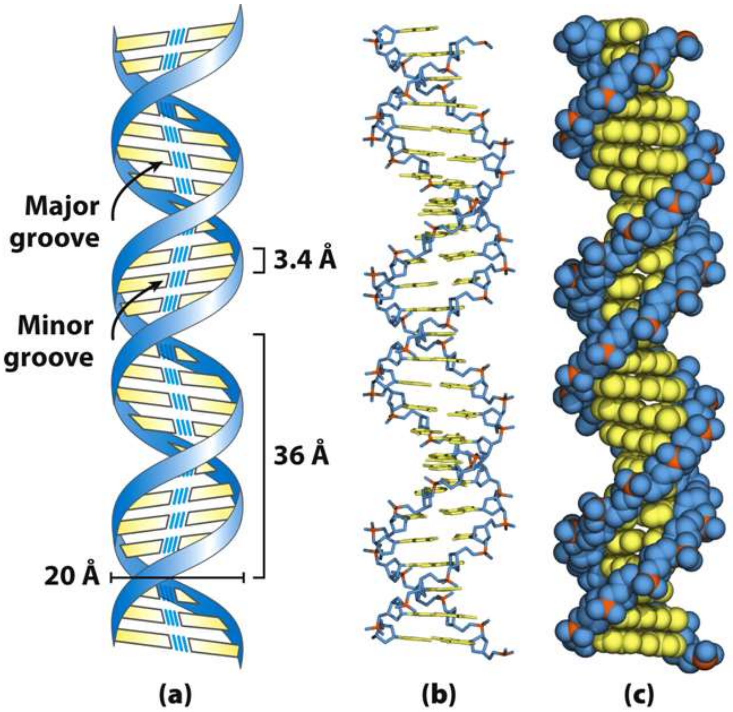


Figure 1.4. Structure of idealised B-DNA. (a) Ribbon model of a typical B-DNA duplex, indicating the length of a nucleotide and a helical turn. (b) Skeletal model of dsDNA, the phosphate backbone is in blue and red; the DNA bases are in yellow. (c) Space filling model of the same duplex, colour code is as for (b). From: Lehninger Principles of Biochemistry 6E, by David L. Nelson, *et al,* Copyright 2013 by W.H. Freeman and Company. Used by Permission of the publisher.[156]

In aqueous solution the B-DNA conformation (Fig. 1.4) is favoured.[7] Formation of B-DNA is strongly exergonic; helix formation is enthalpically-driven, due to phosphodiester-solvent hydrogen bond formation, π-stacking between adjacent bases and Watson-Crick base pairing.[7] A negative entropy change is observed upon duplexation of two complementary strands of dsDNA, and hence, duplexation becomes disfavoured at higher temperatures. The temperature at which this happens is referred to as the melting temperature (T_m) and is a property specific to a given sequence of dsDNA, and complexes thereof.[7]

No global structural ordering of DNA is observed when a strand of DNA is dissolved in an organic solvent or when relative humidity is below 75%. When the relative humidity reaches approximately 75-80%, the A-form of dsDNA is observed (Fig. 1.5). A-DNA has 11bp per turn and 2.9 base pairs per turn of the helix; it resembles a compressed B-DNA helix (Fig. 1.5).[10,11] Due to the lower relative humidity, less extensive base stacking is observed (Fig. 1.5).[12]

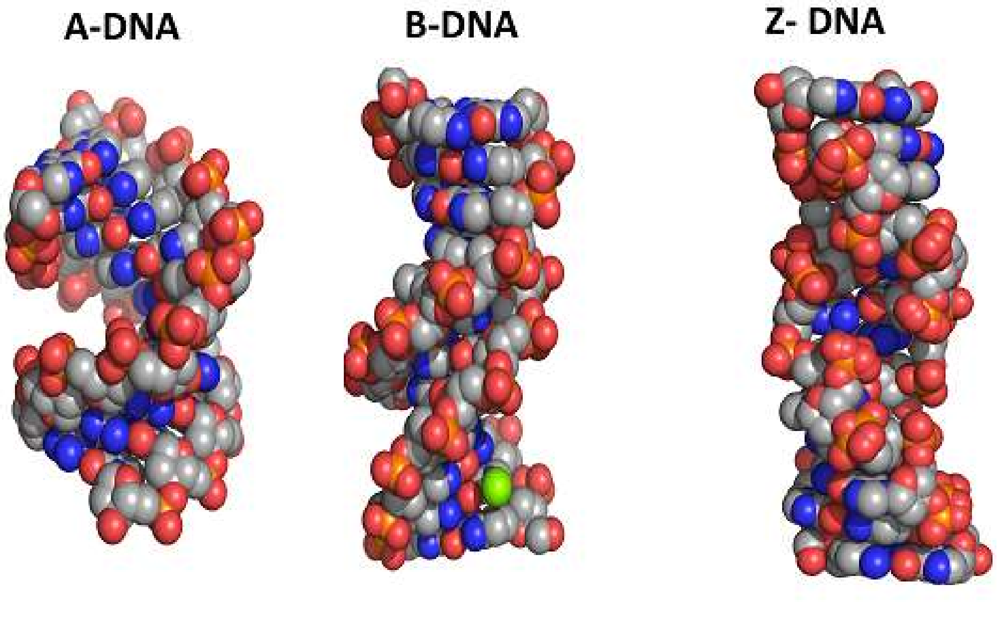


Figure 1.5: Crystal structures of the three helical forms of duplex DNA, from left: A-DNA (PDBID: 4IZQ)[12], B-DNA (4OCB)[14], Z-DNA (4C64)[157]. Grey spheres represent carbon atoms, red spheres represent oxygen, orange phosphorous, and blue, nitrogen. The green sphere in (b) represents an Mg^{2+} cation.

Another structure of dsDNA that can be formed is Z-DNA. This form is radically different from the other two structures of DNA, and is only seen in purine-pyrimidine repeats, especially G-C. Z-DNA is a left-handed helix, with 12bp per turn,[11] and is read antisense, i.e. in the 3′ to 5′ direction. Moreover, the phosphate backbone in Z-DNA proceeds in a zigzag fashion, unlike A and B-DNA which wind around each other in straight lines, forming the helix.

1.1.4. The role of the solvent and metal cations in dsDNA structure.

The polyanionic nature of the phosphodiester backbone indicates that dissolved metal cations should have a substantial effect on preservation of DNA structure. Privé *et al.* showed that the presence of counterions stabilize the structure of duplex DNA.[13] Mg^{2+} and Na^{+} cations octahedrally co-ordinate to 6 water molecules and maintain the structure of the tightly bound, largely hexahedral co-ordination of water molecules in the minor groove of dsDNA. Structural changes in the crystal structure of B-type dsDNA are observed when water was gradually added to dried fibres of DNA.[13] Long-range ordering of duplex DNA was not seen below 75% relative humidity. A-form dsDNA was observed to predominate between 75-85% relative humidity and undergoes a structural transition to the more familiar B-form when relative humidity was increased to greater than 85%.[7]

1.1.5. Structure of the grooves of B-DNA helices.

B-DNA has two unsymmetrical grooves between the phosphodiester backbone, referred to as the major and minor grooves (Figs. 1.4, 1.6). The major groove is 22 Å in width, and is more open to solvent, and the edges of the bases are more open to the environment. The minor groove of B-DNA is narrower, 12 Å in width, and consequently is much less hydrated than bases in the major groove. The difference in hydration of the grooves is also reflected in the pattern of hydration of the major and minor grooves. The greater solvent accessibility of the major groove leads to a hydration pattern that reflects the hydrogen bonding surface of the

DNA duplex. The minor groove is markedly different. There exists a hydration shell within dsDNA, which exhibits an IR spectroscopic signature more similar to that of solid than of liquid water.[7] These water molecules are intrinsic to the structure of dsDNA.[13] High-resolution X-ray diffraction studies on a B-DNA oligomer has shown that water is immobilised in a structured manner within the minor groove.[14] There exist two classes of structured water molecule, defined by the local bonding environment experienced by the water molecule. Class 1 water molecules are tightly bound deep inside the minor groove, and are tetrahedrally bound (Fig 1.6).[14] There are 11-12 molecules of water per nucleotide bound the minor groove of DNA, two to the ionic phosphates, four to the phosphodiester oxygen atoms, and a further four to the sugar oxygen atoms. The other two water molecules are bound to the hydrogen bond donors on the bases.[15] Class 2 water molecules are less tightly bound than Class 1 water molecules, and consist of 8-9 water molecules hydrogen-bonded to the class 1 water molecules. The spectroscopic properties of these water molecules are similar to those of bulk water,[7] and are consequently less involved in defining the shape of dsDNA.

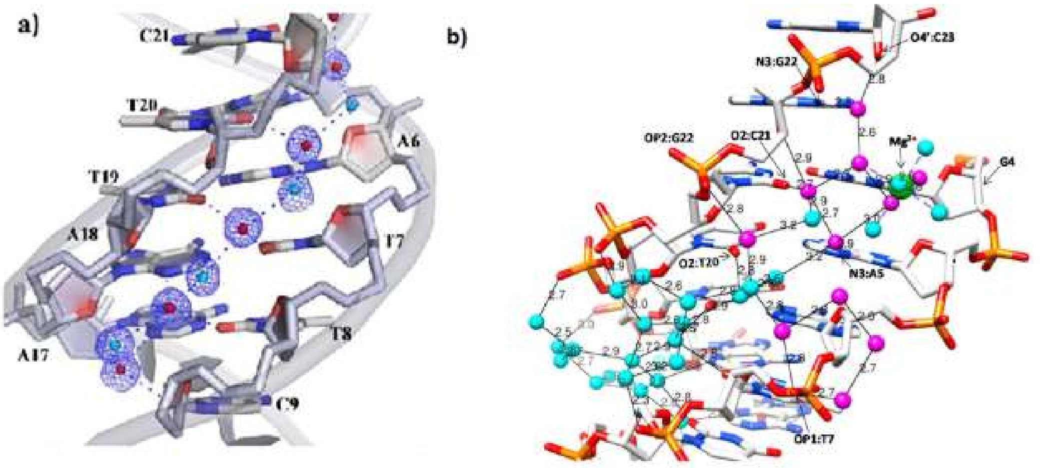


Figure 1.6 (a) Schematic showing the class I bound water molecules tightly bound by dsDNA.[14] **(b)** Skeletal model of class 1 (cyan) and class 2 (mauve) water molecules hydrogen bonded in the minor groove of duplex DNA. One Mg^{2+} (green) is also present. Adapted and reproduced with permission from Ref. 14. Copyright 2013 American Chemical Society.[14]

Class 1 water molecules in the sphere of hydration directly bridge the adenine N3 and the thymine O2, forming a “spine of hydration” through the minor groove in A-T rich regions.[16]

1.2. Molecular recognition of dsDNA

1.2.1. The process of transcription.

Transcription refers to the synthesis of strands of messenger RNA (mRNA), from a template strand of dsDNA, which is then used for the synthesis of proteins by a process known as translation. As translation is an RNA-, not DNA-based process and also this process occurs in the ribosome not the nucleus, this process falls outside the subject of this thesis and hence will not be described further.[7] Another fundamental cellular process is replication, by which dsDNA is copied during cell division. The DNA-binding proteins used in replication are structurally similar to transcription factors and for brevity this process will not be described in this thesis.[8]

Synthesis of mRNA is achieved through the interaction of DNA with three different functions (Fig. 1.7). These are:

1. Formation and maintenance of an open chromatin structure,
2. Assembly of the preinitiation complex,
3. Elongation of the nascent mRNA strand.[17]

The first step of transcription is the effect of HDAC (histone deacetylase) proteins; these proteins acetylate the lysine residues of chromatin, hence removing the charge-charge interactions between the positively charged lysine residues and the polyanionic DNA backbone. This allows for binding of the TATA-box binding protein (TBP) to its consensus binding site, the TATA box, the binding mode of which is discussed later in this chapter (*Section 1.2.4*).

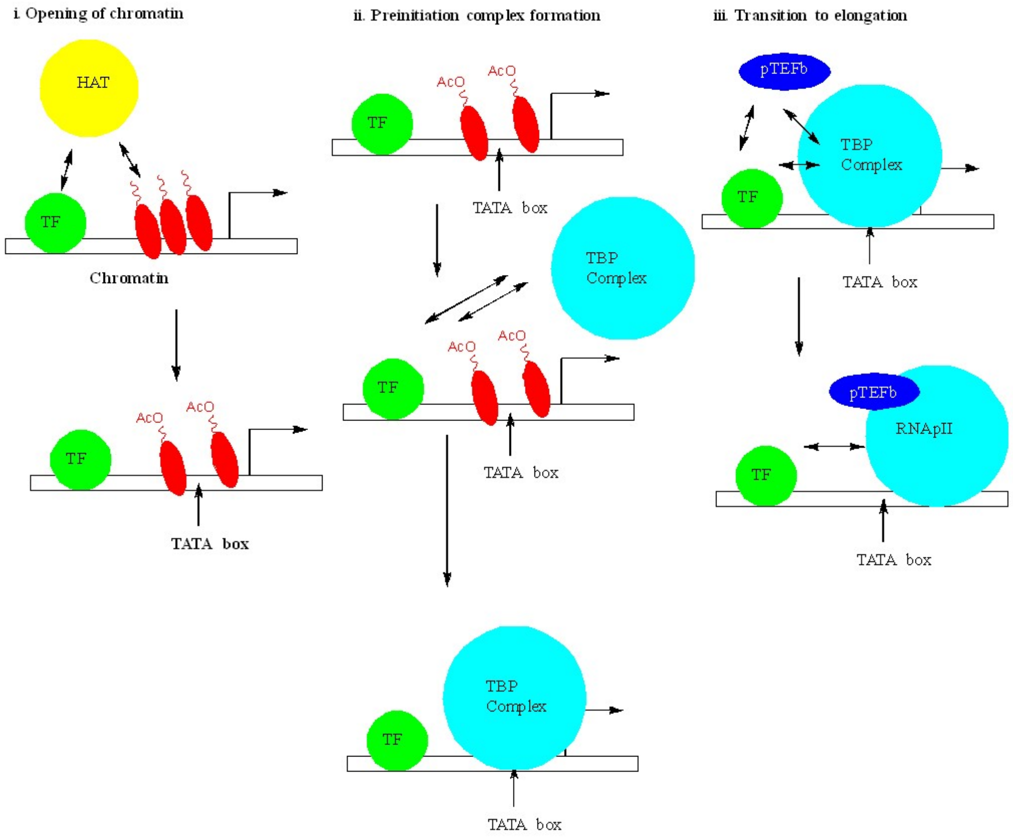


Figure 1.7: Overview of the steps involved in transcription.[17]

Protein-protein interactions between the TBP and a transcription factor, forming a complex known as the pre-initiation complex, and finally the elongation comes from interactions between the transcription factor, RNA polymerase II (RNAP II), and an enzyme such as pTEFb kinase.[17] For transcription to function correctly, the three steps of the transcription process must be controlled in a site-specific manner. The critical first step is recognition by transcription factors of its consensus binding site, typically 4 b.p. in length. 4 base pair DNA-binding domains (DBDs) do not represent a unique binding site in a genome, which typically for mammals consists of billions of base pairs. DNA binding proteins have evolved several strategies to confer selectivity. Some are simple, for example dimerisation, but others are more nuanced, for example, shape recognition either due to inherent DNA structure or induced by partial intercalation. The equilibrium structures of 1500 of these complexes have been characterized, and the strategies employed by these proteins have been explained.[18] What is less well understood, however, are the molecular mechanisms by which association takes place; whether the shape changes in the DNA duplex upon binding of a protein is due to transient conformational changes in the structure of dsDNA, or the approaching ligand induces a change in the structure of dsDNA through a noncovalent interaction and the conformation of the dsDNA: ligand conformation is fixed through binding to other residues in the DBD.

1.2.2: Recognition of DNA by proteins in the major groove: Zinc fingers

Regions of proteins that form contacts with the major groove of dsDNA form two major groups: zinc fingers and leucine zippers.[19] Zinc fingers occur widely in nature, and regions coding for zinc fingers have been found in 3% of human DNA.[20,21] Proteolytic digestion of the protein shows that the active site contains repeating compact domains of 3kDa molecular weight.[22] One Zn^{2+} ion is coordinated to two cysteine and two histidine residues. Further evidence for this observation comes from EXAFS data.[23] The 3kDa regions are roughly 30 amino acids in length, 25 of these residues fold around the metal centre forming a hydrophobic core stabilising the compact finger motif (Fig. 1.10). These fingers are also rich in basic and polar residues, suggesting that this is the part of the DNA that binds to the polyanionic surface of the major groove of DNA.[21]

Binding of proteins to the major groove of dsDNA is a dynamic process which causes the structure of B-DNA to deviate from the idealised shape. The ability of dsDNA to change shape from idealised B-DNA is dependent on base sequence, and hence, recognition of the shape of a strand of dsDNA is one of two main mechanisms governing specified DNA-protein interactions,[24] the other being a direct read-out of the H-bond profile of base-pairs, known as base-readout.[19]

Base readout is defined as the direct interactions between the edges of the bases of dsDNA and the residues of DNA-binding domains. An archetypal example is shown by the zinc finger protein Zif268 binding to its target sequence (Fig. 1.8 a).[19] The shape of dsDNA in complex with Zif268 results in widening of the major groove[25] which exploits both hydrogen bonding and hydrophobic contacts. Hydrogen bonding is the more dominant interaction between major groove binding proteins and DNA, and a high degree of sequence selectivity is attained when the hydrogen bond donor binds to two acceptors.[19] Immobilised water molecules mediate the hydrogen bond environment in the major groove, as they reflect the positions of the H-bond donors and acceptors at the base edges.[19] Hydrogen bonds are used more often in the recognition of purines, while hydrophobic contacts are seen more in the recognition of pyrimidines, reflected by the amino acid residues that make the contacts (Fig. 1.8a).[19,25]

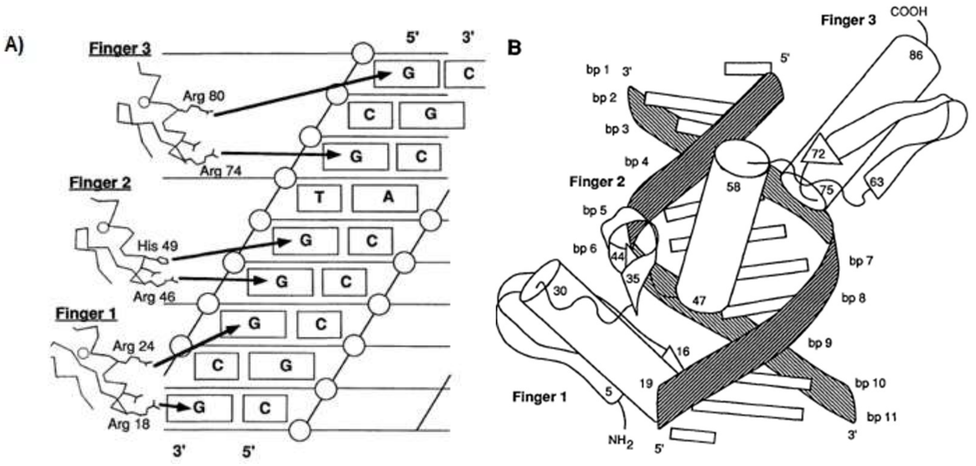


Figure 1.8. (a) Contacts between Zif268 and a target strand of DNA. **(b)** Schematic of the Zif268:dsDNA complex Each of the cylinders represent the structure shown in Fig. 1.10.[25] Reprinted by permission from Macmillan Publishers Ltd.: Nature © 1991.

Zif268 contains three “fingers”, each of which bind to a specified sequence of duplex DNA. The binding as shown in Fig. 1.8 b is complementary to the shape of the major groove, which is similar to that in idealised B-DNA, implying that the folding of the DBD adapts to the shape of dsDNA and not the opposite.[25] Eleven hydrogen bonds are formed between guanine residues and the fingers and the DNA read in an antiparallel fashion relative to the direction of protein synthesis (Fig. 1.8a). In each of the three fingers, an arginine residue forms a specific bidentate hydrogen bond with the N7 and O4 of guanosine (Figs. 1.8 b, 1.9). A further arginine residue makes a nonspecific hydrogen bond to the phosphodiester backbone, as does one of the histidine amino acids bound to the metal centre (Fig. 1.10). No contacts are made with the cytidine base on the complementary strand.[25]

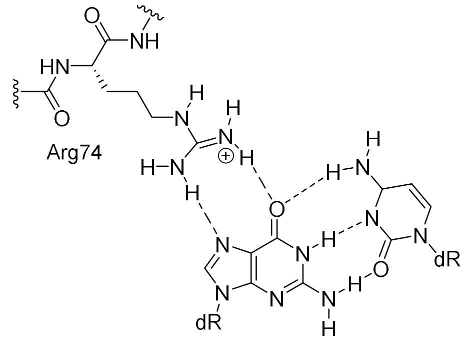


Figure 1.9: Illustration of the bidentate H-bond between a G-C base pair and the Arg74 residue of Zif268. Binding only occurs on the strand bearing the guanosine residue.[158]

The role of the Zn^{2+} ion is also key to the function of the protein as tetrahedral geometry caused by co-ordination of 2 cysteine and 2 histidine residues around the zinc centre fixes the helix-turn-helix motif, thus presenting the contacting residues to the surface of the dsDNA.[21] The Zn^{2+} ion has also been observed to mediate binding to dsDNA.[21] The mode of binding of zinc fingers to dsDNA is a 1:1 interaction between a G nucleotide on a single strand and a single finger (Fig. 1.8 a). Each finger recognises 3 base pairs (Fig. 1.8 a) by forming a bidentate hydrogen bond to a guanosine nucleotide. The non-contacting base pairs stabilise the formation of the bound complex through vertical π-stacking interactions. The specificity of a zinc finger to its target strand increases as the number of fingers is increased.[21] The identity of the bases outside the direct binding site is also important, as they create the complementary binding surface between the protein and the dsDNA.

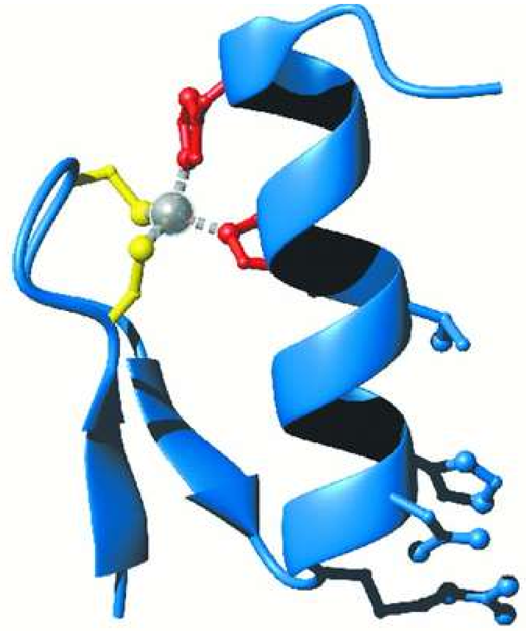


Figure 1.10. The role of Zn^{2+} (silver) in stabilising protein structure in Finger 2 of Zif268. Two cysteine (yellow) and two histidine residues (red) act as co-ordinating ligands. The α-helix and the DNA-contacting residues are shown in blue.[159]

One example of this is the zinc-finger type protein, TTK, isolated from *Drosophila* which requires a highly deformable ATA motif[26] to bind correctly to DNA. In contrast to Zif268, the dsDNA target sequence actively changes conformation when the protein is bound to dsDNA, presumably due to a strain energy benefit.[21] This shows that both shape- and base-specific interactions are influential in controlling the specificity of dsDNA recognition.

1.2.3. Recognition of DNA by proteins in the major groove: Leucine zippers.

Leucine zippers have the biological role of mediating cellular stress responses, more common in plants than animals, although it is hypothesised that the diverse cellular functions attributed to the action of leucine zippers on dsDNA are due to heterodimerisation. Leucine zippers have DBDs 60-80 residues in length with highly conserved leucine residues every seven residues, and a DNA-contacting basic region 16 residues in length. The structure of the dimer is an amphipathic helix with 2 turns per leucine heptad.[27] The basic region resembles that seen for zinc fingers, and forms contacts with the major groove of dsDNA in much the same way as the α-helices in zinc fingers do.[28] The noncontacting helix has a role in recognising the minor groove and facilitates formation of the active binding conformer of the protein.

The contacting helix of leucine zippers recognises the major groove of a G-C rich region of dsDNA, for example the sterol regulatory element, namely ATCACCCCAC (Fig. 1.11).[28]

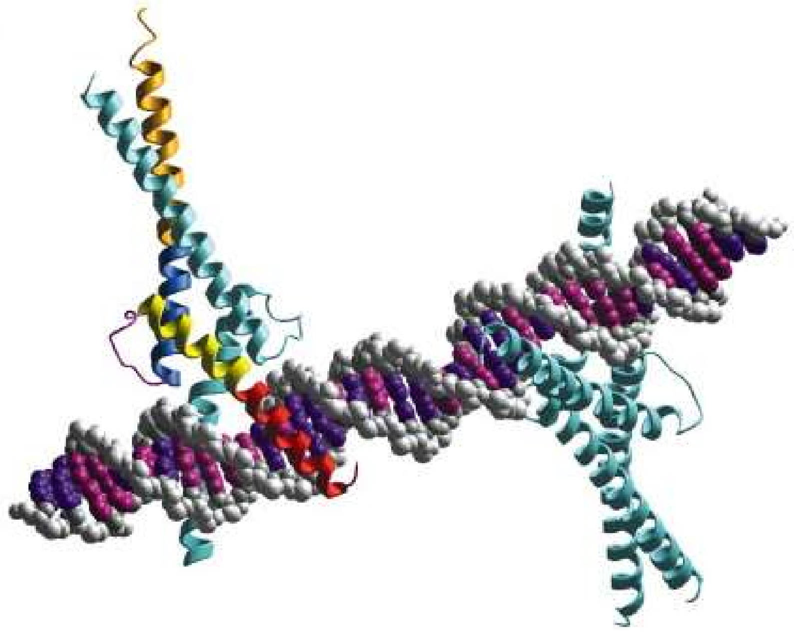


Figure 1.11. Crystal structure of the sterol regulatory element in complex with dsDNA.[28]

1.2.4 Interactions of DBDs in the minor groove.

Proteins with A-T rich consensus binding sites tend to recognise the minor groove of dsDNA. One important example is the TATA-box binding proteins (TBP). Unlike major groove binding peptides, the recognition element of the TATA-box is a β-sheet, as opposed to the α-helices present in both zinc fingers and leucine zippers.[29] Extensive deformation of the DNA duplex is observed upon binding of the ligand (Fig. 1.12).[30] This deformation weakens the interstrand hydrogen bonds, facilitating strand separation and allowing mRNA synthesis to occur.[29]

Furthermore, in opposition to major groove contacting proteins, the only hydrogen bonding contacts between the DNA and the protein are observed between the edges of the β-sheet and the phosphate backbone of the DNA.[30] These hydrogen bonding interactions are inherently nonspecific, and hence, proteins recognise the minor groove of DNA by a different mechanism than the major groove.

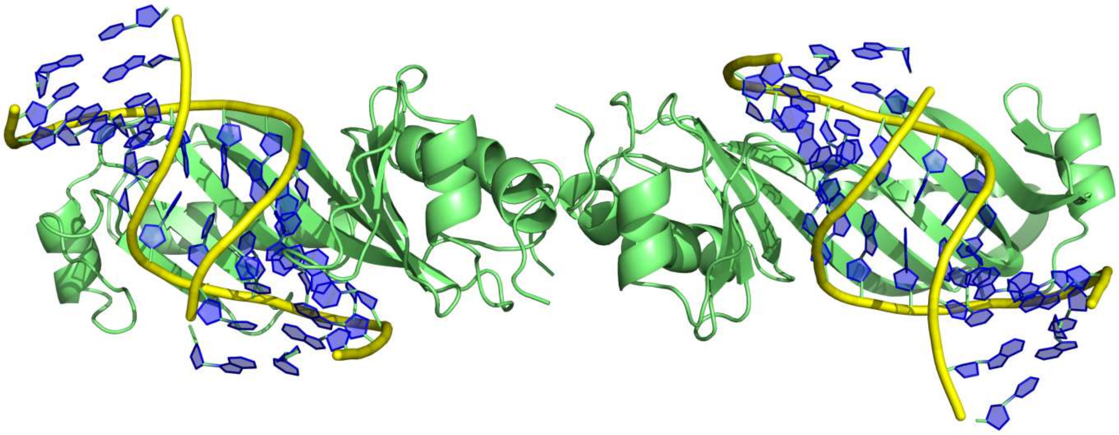


Figure 1.12. Crystal structure of a TATA-box binding protein in complex with minor groove of dsDNA. Note the deformation of the DNA from B-DNA. (PDB ID: 1VTO)[30]

1.3. Origins of sequence specificity in the formation of protein-dsDNA complexes.

1.3.1. Lack of a simple code in DNA-protein recognition.

Footprinting studies (*Section 1.5.1*) have shown that a transcription factor will bind multiple short binding sites in genomic DNA, for example, two of the zinc fingers in the GATA3 protein will recognise multiple GATA sites, whereas the finger on the N-terminus prefers to bind a GATC sequence.[31] This lack of a 1:1 correspondence between a DNA-binding motif and its consensus binding sequence of dsDNA implies that other structural factors in both DNA-binding proteins lead to increased binding specificity.[18] Two means of recognition have been shown to exist between proteins and DNA duplexes: base and shape readout (Fig. 1.13).

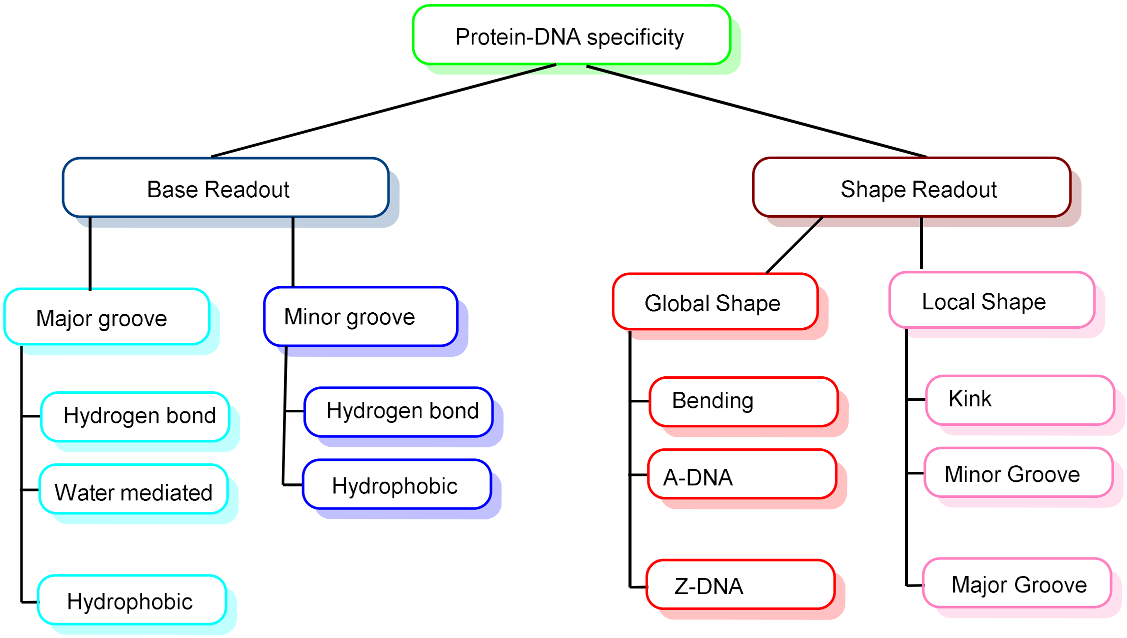


Figure 1.13. Summary of the factors governing protein-DNA binding.[19]

1.3.2: Base readout

Protein contacts with the major groove are mediated by either hydrogen bonding, or hydrophobic contacts between the protein recognition element and the exposed edges of the bases in the major groove (*Section 1.2*). Base readout is more established than shape readout, due to its easy identification by DNase I footprinting and X-ray crystallography. Hydrogen bonding is seen frequently in major groove binding proteins due to the more exposed binding surface of the groove, leading to a unique binding surface defined by base sequence.[19] The water molecules hydrogen bonded to the base edges also exhibit a pattern reflecting the edges of the bases; this property is also exploited by binding proteins. One example of water-mediated hydrogen bonding is the Trp repressor (Fig. 1.12) which in addition to a bidentate hydrogen bond between an Arg residue and each guanosine of the 5’-CTAG-3’ half binding site, shows extensive water- and ion-mediated contacts between the protein and the major groove.[32,19]

The zinc fingers of the GATA3 transcription factor also recognise the GATA binding site through hydrophobic contacts in the major groove (Fig. 1.14).[31] In addition to the bidentate binding interaction between Arg276 and guanosine, there also exists a bond between the thymidine O4 and Arg 276. A hydrogen bond also exists between Asn286 and the adjacent thymidine and adenine of the binding site. The hydrophobic pocket formed by the interaction of Arg 276 and Asn 286 recognises the 5-methyl group of thymidine. Hence, a unique motif for the recognition of a 5'-GATA-3' sequence is achieved.[33]

Although the binding interaction between the two amino acids and the four nucleotides is specific, it is not unique within the genome - a protein needs to recognise more than 12 nucleotides for a binding site to be functionally unique. Nature achieves this through dimerisation of DNA-binding proteins. In the case of the full GATA3 protein, co-operative binding exists between both fingers – leading to an increased affinity for the consensus binding sequence of 5'-ATCWGATA-3' (where W = A or T). The basic region located at the C-terminus of the protein reads an A-T rich minor groove immediately adjacent to the binding site, hence increasing its specificity (Fig. 1.14).[33]

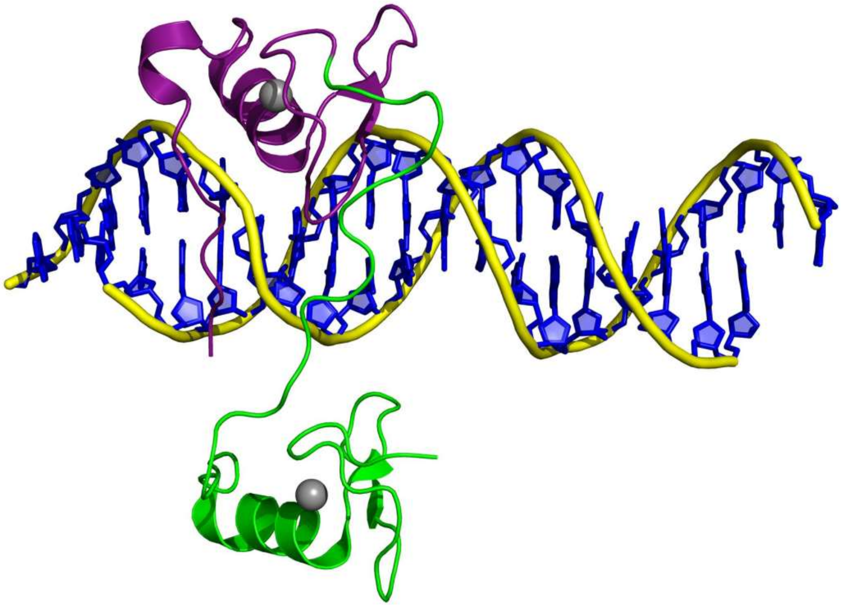


Figure 1.14: Crystal structure of the GATA3-DNA complex. (PDB ID: 4HC9).[33]

1.3.3. Shape (Indirect) Readout

As shown in Fig. 1.14, the basic C-tail of the GATA3 transcription factor reads the minor groove immediately adjacent to the binding site. However, as shown in Fig. 1.6, the structure of the minor groove is markedly different than the major groove, and is not suited to the formation of specific hydrogen bonding interactions. It has been demonstrated by the Rohs group that deviations in local and global DNA shape from idealised B-DNA (Fig. 1.4) can be "read" by incoming proteins based on their physical and chemical properties.[24] Shape variations in DNA duplexes are both common and predictable,[34,35] and are a direct result of base sequence. Hence, even without direct contacts between bases and incoming proteins, specific interactions between DNA and a ligand are possible.[19]

The most common structural variation read by proteins is a narrow minor groove exhibited by A-T tracts. The narrower and deeper minor groove, as well as the neutral thymidine methyl group in the groove, leads to an increased negative electrostatic potential on the surface of the DNA duplex (Fig. 1.15), which can be read by incoming arginine residues.[19] As shown in Fig. 1.14, there is an interplay between base and shape readout in the majority of reported protein-

DNA complexes.[18] In addition to narrow minor grooves, two other local structural variations exist, bends and kinks.[19] Both are defined as a structural variation from linear dsDNA towards a bent conformation, but the defining feature of a kink is a structural deviation over 1 base pair, whereas a bend occurs over a longer range (Fig. 1.15).

Bent DNA is frequently seen in A-tracts, where conformationally flexible A-T base-pairs are found. This is sequence-dependent; significant bending is observed for ApA and ApT steps, but not for TpA steps.[36] Bent DNA has also been shown to be a recognition element for the human papillomavirus E2 (Fig. 1.15 a).[37]

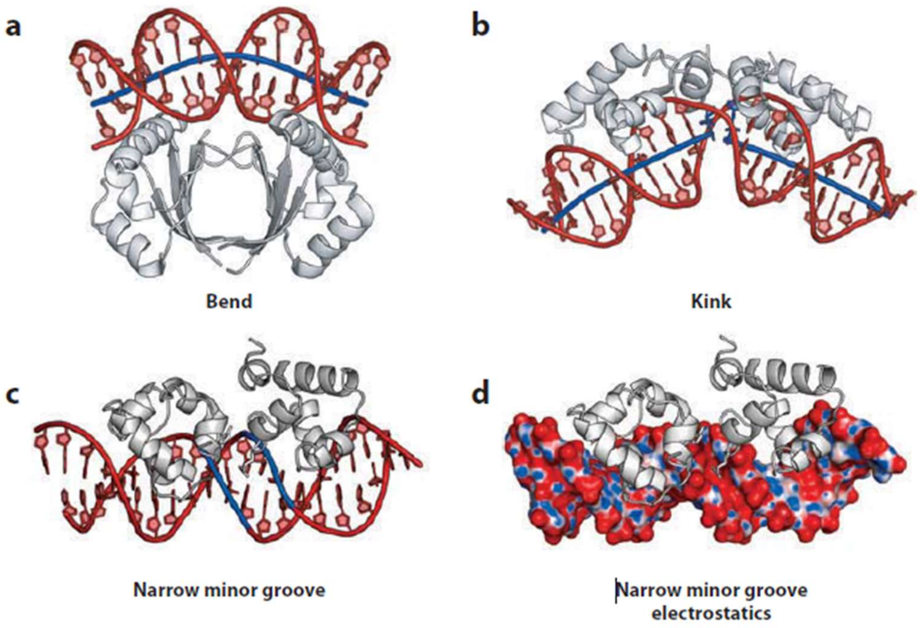


Figure 1.15 (a) An example of bent DNA shape readout – HPV-18 E2 (PDB: 1JJ4). **(b)** An example of a DNA kink induced by intercalation of leucine residues across a CpG step – the Lac repressor (PDB: 2KEI). **(c)** The narrow minor groove (blue) is recognised by incoming arginine residues – Phage 434 (PDB: 2OR1). **(d)** The enhanced negative electrostatic potential of the narrow minor groove of the Phage 434: dsDNA complex is shown in red.[24,19]

DNA kinking is frequently seen at pyrimidine-purine steps, due to weaker π-stacking interactions between the bases.[38] The weakest stacking interaction is seen between TpA steps, and the presence of these steps can be recognised by intercalating side-chains, for example in the Lac repressor[19] (Fig. 1.15 b) or in the high-mobility group protein, HMBG1 (Fig. 1.16).[18,39] Three residues were shown to intercalate across the duplex, Ile13, Phe97, and Ile116, which caused unwinding of the duplex, leading to the "kinked" structure shown in Fig. 1.16. The narrowed minor groove shape caused by the unwinding is then recognised by the arginine residues on the linker region primarily by electrostatic attraction, although some nonspecific hydrogen bonding is also present.[39] Unusually for a protein-DNA complex, HMBC1 recognises its target sequence of dsDNA entirely by shape. It is more common that shape and base factors act together for a given protein to find its unique target DNA sequence.

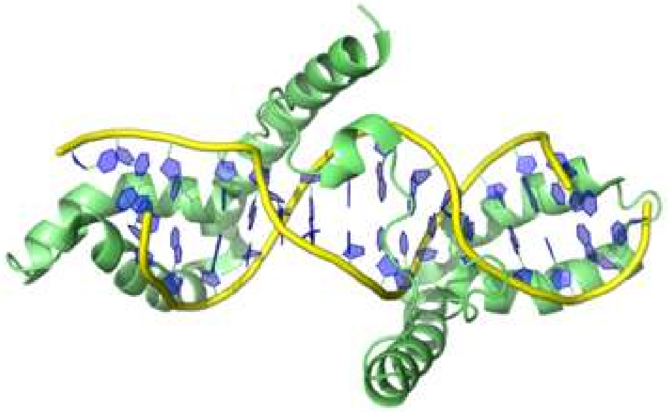
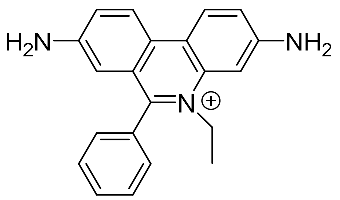


Figure 1.16: NMR structure of the HMBG1 protein (PDBID: 2GZK).[39]

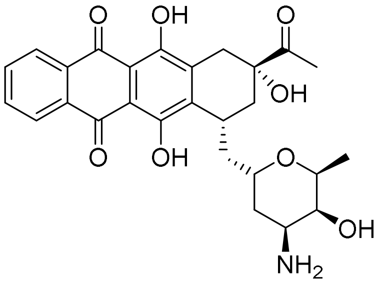
1.4: Small molecules that bind dsDNA.

1.4.1. Intercalation.

Intercalating small molecules are planar molecules containing fused aromatic rings, and are frequently charged. Two examples of intercalating molecules are the synthetic DNA stain ethidium bromide and the anthracycline anticancer agent daunomycin (Fig. 1.17).



Ethidium Bromide, **1.6**



Daunomycin, **1.7**

Figure 1.17: Examples of DNA intercalating molecules.[7]

As observed with intercalating proteins, intercalators insert between a pyrimidine-purine base step, causing both unwinding of the DNA helix and the effective length of the strand of DNA increases by approximately one base-pair. Substantial conformational changes in the shape of DNA are observed (Fig. 1.18). The mechanism of intercalation involves attraction of the cationic small molecule to the polyanionic surface of DNA, followed by hydrophobic interactions between the heteroaromatic bases and the intercalator, leading to insertion of the intercalator into the DNA double helix. This causes conformational changes in dsDNA; unwinding occurs and the length of the dsDNA strand increases by approximately 3.4Å (the width of one base-pair, or the intercalating molecule). Many DNA intercalating molecules are carcinogenic in that they can inhibit normal cell function.[40] X-ray crystallographic studies of **1.7** in complex with d(GCATGC)$_2$ shows intercalation of **1.7** preferentially across trinucleotide sequences 5'-GCA-3' (Fig. 1.18).[41,42] The overall structure of the complex is a distorted B-DNA helix, unwound by 36^o with the two central rings of the ligand intercalated into the structure The length of the duplex increases by approximately 3.4 Å – close to the length of one base pair. The class I bound water molecules remain in the minor groove. The close contacts between the Na$^+$ cation, a solvated water and the methoxy moiety lead to a ternary complex, important in its biological activity.

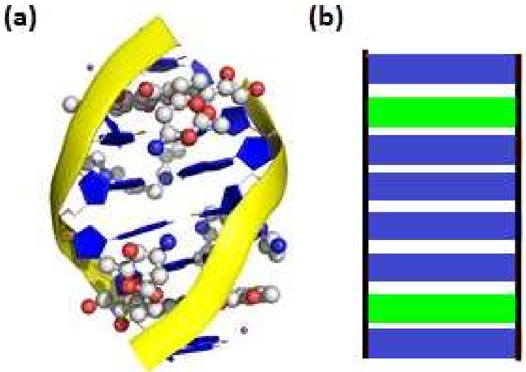
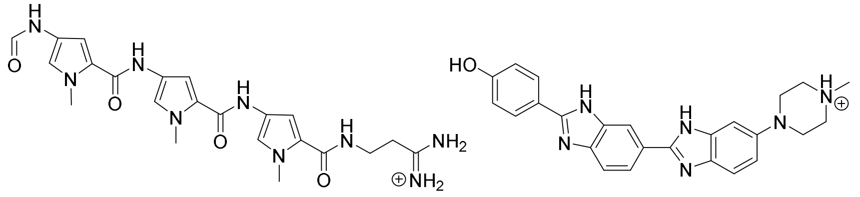


Figure 1.18. **(a)** Crystal structure of daunomycin **1.7** in complex with a dsDNA hexamer. Two molecules of spermidine are also present (PDB ID: 1D11) **(b)** Schematic showing the intercalation binding mode between **1.7 (**green) and the dsDNA hexamer (blue).[41]

The rest of the molecules are weakly bound, suggesting the presence of the Class II hydration sphere. Hence, the observed structural changes of the **1.7**-dsDNA complex from idealised B-DNA prevent the binding of transcription factors to dsDNA and hence provide the mechanism of daunomycin's biological activity.[41] The base pair immediately adjacent to the intercalating ligand is almost completely shielded from hydration by the aminosugar group. Stronger hydrogen bonds are seen between the bases and the ligand compared to the phosphate backbone, suggesting that the mechanism of intercalation involves the intercalating molecule "forcing" its way into the helix and the driving force for the formation of the intercalated complex is hydrogen bonding between nucleotides and the ligand.[41] Subsequent NMR and computational studies have confirmed this.[42]

1.4.2: Minor groove binders.

Minor groove binders such as the non-ribosomal peptide distamycin A **1.8** and the fluorescent DNA stain, Hoechst 33258 **1.9** (Fig. 1.19) bind reversibly to A-T rich sequences in the minor groove of DNA.[7] The driving force for this interaction has been ascribed to hydrophobic interactions between the pyrrole rings and C2 of adenosine; they can approach each other due to the complementary shape of **1.8** and the minor groove. A further energetic contribution comes from the hydrogen bond contacts between the amide protons and the hydrogen bond acceptors in A-T base pairs, i.e. direct base readout.[7] A schematic showing the hydrogen bond contacts formed between **1.8** and dsDNA with 2:1 and 1:1 stoichiometry are shown in Fig. 1.20.[43]



Distamycin A, **1.8**

Hoechst 33258, **1.9**

Figure 1.19: Structures of Distamycin A **1.8** and Hoechst 33258 **1.9**

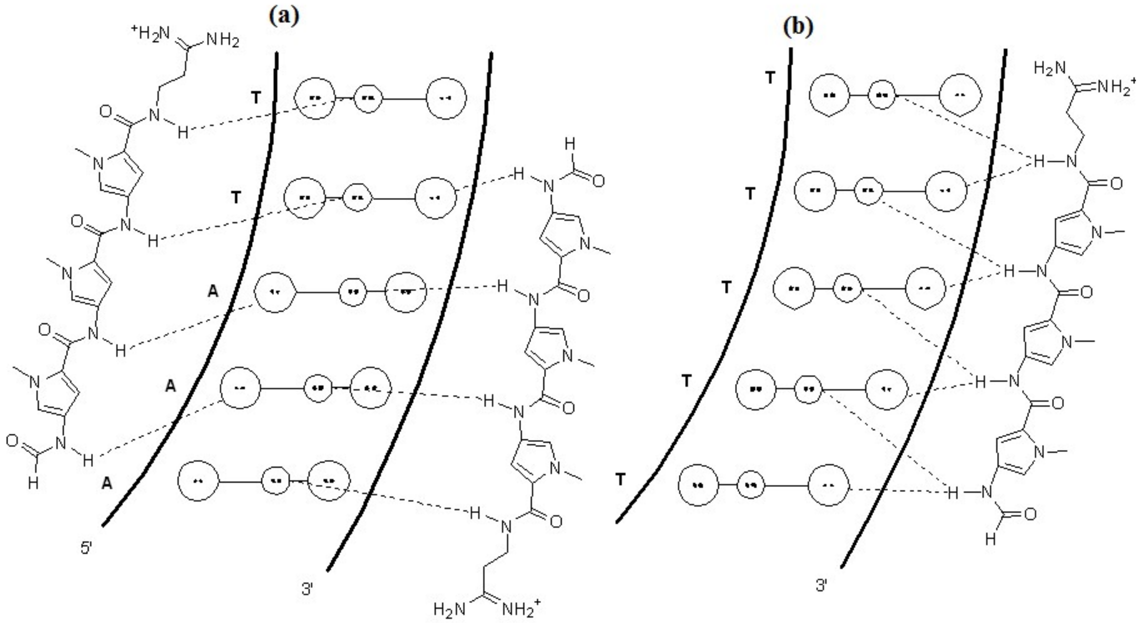


Figure 1.20. Schematic showing the contacts between **1.8** and duplex DNA with **(a)** 2:1 and **(b)** 1:1 stoichiometry.[160]

Distamycin A can bind in a 2:1 antiparallel stoichiometry to A-T rich sequences of dsDNA (Fig. 1.20).[43] Structural analysis of the complex by X-ray crystallography showed that the minor groove of DNA expands considerably to accommodate the two guest molecules (Fig. 1.21 a).[44] A 1:1 binding mode of distamycin A with dsDNA has also been reported to A_5 tracts of dsDNA (Fig. 1.21 b).[45] The hydrogen bonds in the case of the 1:1 complex are bifurcated, whereas in the 2:1 complex, hydrogen bonds are made between one molecule of **1.8** and one strand of the DNA duplex.[44,45]

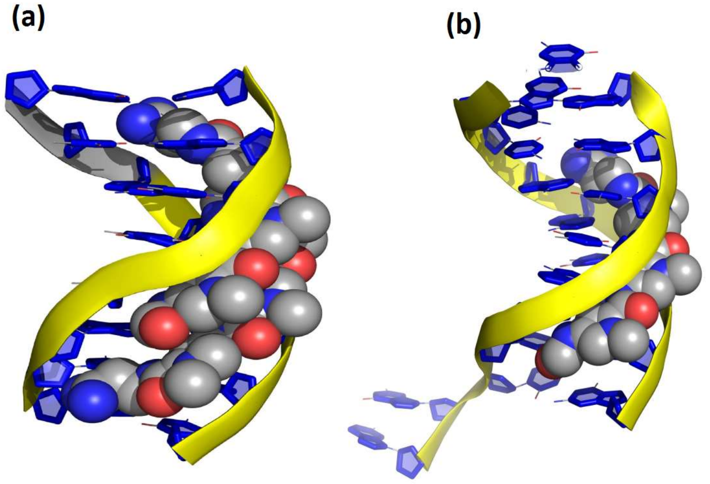


Figure 1.21: **(a)** Crystal structure of a 2:1 complex between **1.8** and dsDNA.[44] **(b)** A 1:1 complex between **1.8** and dsDNA.[45]

Although distamycin favours A-T tracts of dsDNA, its preferred sequence was found to be 5'-ATAAT-3' through a fluorescence displacement assay of **1.6**.[46] Subsequent studies by circular dichroism and isothermal titration calorimetry suggested an important role for the terminal adenosine and thymidine bases in the binding site in determination of the formation of a 1:1 or a 2:1 binding mode of the distamycin.[47] Hydrogen bonds with four of the five bases in the binding site are observed in both of the 1:1 and 2:1 complexes, and binding occurs in a "slipped" conformation (i.e. the base pairs recognised by the two molecules overlap, but are not the same).[44] Hydrogen bonds are formed between the amide protons of distamycin and the O2 and N3 atoms of thymidine and adenosine respectively, indicating that the selectivity for A-T tracts observed for distamycin is at least in part governed by base readout. However, as with minor groove binding proteins, indirect readout is also present.

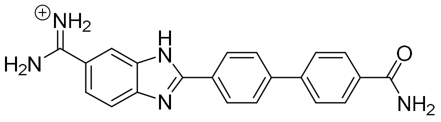
Binding of **1.8** to dsDNA is a stepwise process, where the second molecule of distamycin recognises the distorted 1:1 complex, and the presence of the flanking regions promotes this unwinding process.[48] Furthermore, the association constants of K_1 are lower than the typical association constant of a transcription factor (typically 10^9 or greater) and will hence be displaced by an incoming transcription factor. Hence, for **1.8** to exhibit its biological function of disruption of transcription, the formation of a 2:1 complex (in which the full binding constant is effectively the product of K_1 and K_2) is necessary to effectively inhibit transcription. These observations show that "latent specificity" is present in both small molecules and proteins that bind dsDNA, and hence, as with proteins, both base and shape readout are present in the determination of a complex formed between **1.8** and dsDNA.[49]

The kinetics of the distamycin-DNA interaction have also been quantitatively measured using a quartz crystal microbalance, which allows for measurement of molecular interactions as a mass change.[50] The k_{on} (rate constant) was found to be $72 \pm 12 \times 10^3$ $M^{-1}s^{-1}$ and k_{off} was found to be 0.047 ± 0.012 s^{-1}, which is comparable to that of **1.7** (k_{on} $94 \pm 3.8 \times 10^3$ $M^{-1}s^{-1}$ k_{off} was found to be 0.046 ± 0.002 s^{-1}). However, it has been shown that the k_{on} for **1.7** is uncharacteristically fast for an intercalator, an observation explained by a fast interaction of the aminosugar side-chain in **1.7**. Generally, groove binders form complexes faster than intercalators due to the time-consuming unwinding steps that lead to the formation of the complex.[51]

This data also shows that the rate-limiting step in the formation of DNA-ligand complexes is the conformational change in DNA observed between binding of the first incoming molecule of **1.8**, and the formation of the distorted 1:1 complex recognised by the second molecule of **1.8**.[50] Studying the dynamics of association of a ligand to its consensus DNA binding site will furnish further information on both the conformational change observed in dsDNA and the importance of flanking groups (and hence, the flexibility of the binding site) in determining the binding of a ligand, allowing for the design of more selective and better binding therapeutics.

1.4.3. Minor groove binders comprising linked heterocycles.

A second class of MGB has been studied extensively, the benzimidazole class, an exemplar of which is Hoechst 33258 **1.9**. Structurally, they are characterised as lined aromatic units with a concave structure complementary to that of the minor groove of B-DNA. Positively charged functional groups on one, or both ends of the molecule are also necessary for three reasons. Firstly, they recognise the enhanced negative electrostatic potential exhibited by the narrow minor grooves of A-T tracts. Secondly, they form a water-mediated hydrogen bond with the N3 of adenosine, and thirdly, they confer water solubility to a quite hydrophobic core of the molecule. One further compound, structurally related compound to **1.9** is the synthetic ligand DB1963 **1.10** (Fig. 1.22).[14]



1.10

Figure 1.22: Structure of the MGB DB1963.

Inspection of the structure of **1.10**: dsDNA complex shows that unlike **1.8** and **1.9**, there exists no specific hydrogen bonding pattern for recognition by dsDNA. Moreover, crystallographic studies of **1.10** in complex with dsDNA have shown that any hydrogen bonding interactions are mediated by a conserved water cluster, which resembles the class I bound water molecules in free DNA (Fig. 1.23).[14] As described in section 1.1.5, the hydrogen bonding pattern of bound water molecules in the minor groove of dsDNA do not reflect the edges of the bases themselves,[52] and hence, the origins of specificity in binding must be different than that observed for **1.9.** It is thought that non-polyamide MGBs recognise their consensus sequences of dsDNA through shape readout, and the structural feature recognised by dsDNA are the significantly more narrow minor groove characteristic to A-T rich dsDNA.[24,35]

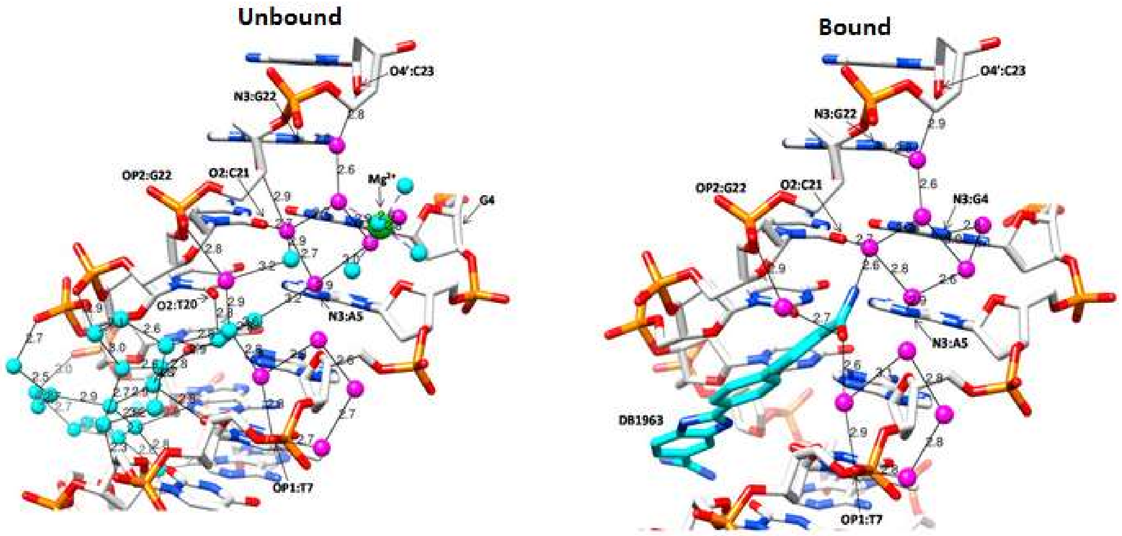


Figure 1.23: Crystal structure of an A-T tract of free dsDNA, and with one molecule of **1.10** bound. The conserved class I water cluster is shown in maroon, whereas the displaced water molecules are shown in cyan. Reproduced with permission from Reference 14. Copyright American Chemical Society.[14]

1.4.4. Thermodynamics of dsDNA-MGB association.

The formation and stability of complexes of both small molecules and proteins with dsDNA is governed by the free energy of binding, described by equations (Eq. 1.1) and (Eq. 1.2).

$$\Delta G^{\circ} = \Delta H^{\circ} - T\Delta S^{\circ} \qquad \text{(Eq. 1.1)}$$

$$\Delta G^{\circ} = -RT\ln K \qquad \text{(Eq. 1.2)}$$

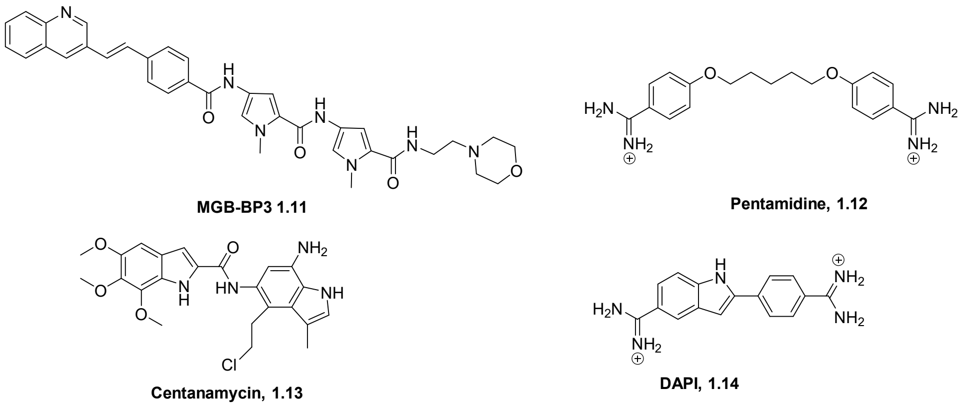
A negative value of ΔG° shows that formation of the complex is spontaneous, while the magnitude of ΔG° directly influences the binding constant, K, and hence the degree of stabilisation of a given complex, relative to the free duplex. The most common method used for directly studying the thermodynamics of formation of a DNA: ligand complex is isothermal titration calorimetry (ITC) which measures heat given to or taken from the surroundings, thus directly measuring ΔH°. The process can be repeated at different temperatures to obtain a figure for ΔS°, through solving a set of simultaneous equations. K can then be calculated from this data by solving equation 1.2. Binding can be both an enthalpically-driven i.e. the formation of a hydrogen-bonded complex – implying the origins of formation of the complex is driven by base readout or entropically driven i.e. driven by the positive entropy change from extrusion of water from the complex and Van der Waals' interactions between the DNA and the ligand.

Unsurprisingly, given the difference in their binding modes, compound **1.8** has a markedly different thermodynamic profile for binding of both **1.9** and **1.10** to dsDNA which are somewhat similar to each other.[53] As expected by the formation of a hydrogen-bonded complex, the enthalpy change is strongly negative: -12.3 kcal.mol^{-1} for the first molecule, and -18.8 kcal.mol^{-1} for the second.[54] This overrides the negative entropy change observed upon binding of both distamycin molecules (-7.5 cal. K^{-1} and -34.47 cal. K^{-1}) leading to a free energy change of -10.1 and -8.7 kcal. mol^{-1} for binding of the first and second equivalent of **1.8** respectively.[54] Van der Waals interactions between the 1:1 **1.8**: DNA complex and the second incoming equivalent of **1.8** cause the increased negative observed enthalpy change, offsetting the more negative entropy change caused by immobilisation of water molecules.[54]

The main contribution of **1.10** binding to A-T rich regions of dsDNA to the ΔG° value comes from the large positive entropy change (29.3 cal.mol^{-1}K^{-1}) observed upon binding. The enthalpy change, while negative (-1.7 kcal.mol^{-1}. K^{-1}), is not large enough to explain the sub-micromolar dissociation constant observed for **1.10**. When multiplied by the temperature (293 K), the TΔS° contribution to ΔG° becomes -8.7 kcal.mol^{-1}K^{-1}, 80% of the total free energy contribution to the observed K_a of **1.10** to dsDNA.[14] Major contributions to the positive entropy change include solvent reorganisation (i.e. returning of the cyan water molecules in Fig. 1.23 to the bulk solvent), as well as conformational changes in dsDNA.[19] Understanding the conformational changes in dsDNA upon binding of a ligand by studying the dynamics of a dsDNA-ligand mixture will allow for a deeper understanding of the physical phenomena underlying the thermodynamics of formation of these complexes.

1.4.5. Clinical Significance of MGBs

MGBs present a family of DNA-binding compounds that have considerable therapeutic potential as members of this class of compounds exhibit both anticancer[55] and anti-infective properties.[56] Exemplars of clinically relevant MGBs **1.11-1.14** are shown in Fig. 1.24. However, despite the considerable therapeutic potential of MGBs as anticancer and anti-infective agents, comparatively few of these compounds have progressed to clinical use. Exceptions do exist, **1.12**[57] is currently being used as a treatment for leishmaniosis and **1.13** **(**Fig. 1.24) is currently being employed as a as a treatment for malaria.[58] Efforts towards the rational design of MGBs as therapeutic agents have been hampered by, among other reasons, a poor understanding of the molecular processes by which a given MGB interacts with dsDNA.[59]


MGB-BP3 1.11

Pentamidine, 1.12

Centanamycin, 1.13

DAPI, 1.14

Figure 1.24: Examples of clinically relevant MGBs, showing their structural diversity.[56]

1.5: Techniques used to study dsDNA: MGB complexes

1.5.1. DNase I footprinting

Biochemical methods have been used by research groups studying MGBs in order to establish their binding affinity and sequence selectvity.[2] One such method is DNase I footprinting, in which a DNA-ligand complex is digested using the nonspecific endonuclease DNase I.[60] The MGB masks the DNA from the endonuclease, leaving a gap, referred to as a "footprint" after gel electrophoresis. This directly and unambiguously gives the binding site for a given binding ligand (Fig. 1.25). Such a technique is useful for both minor and major groove binders and the technique is effective for any length of targeted DNA. The DNA sequence under investigation can be radiolabelled with ^{32}P, and the positions of all the fractions on the electrophoresis plate can be shown quickly and easily. Moreover the binding constant can be estimated from the concentration of masked DNA.[7] The main disadvantage of this technique is that in the case of high-affinity binders i.e. less than 1 nM dissociation constant, higher concentrations of DNA are required in order to attain enough unbound DNA for meaningful analysis. This leads to a greater dilution error and effects of the higher concentrations, for example aggregation, become significant.

A related technique is to induce chemical cleavage of DNA by making a derivative of the ligand with a moiety that will cleave duplex DNA at either end of the strand, one used by Dervan and co-workers is an iron (II) EDTA complex.[61] This works similarly to a restriction nuclease enzyme in that it cleaves DNA selectively via recognition of the target sequence of DNA. Oxidative cleavage of dsDNA, in the presence of atmospheric oxygen and dithiothreitol occurs selectively at the binding site of the ligand.

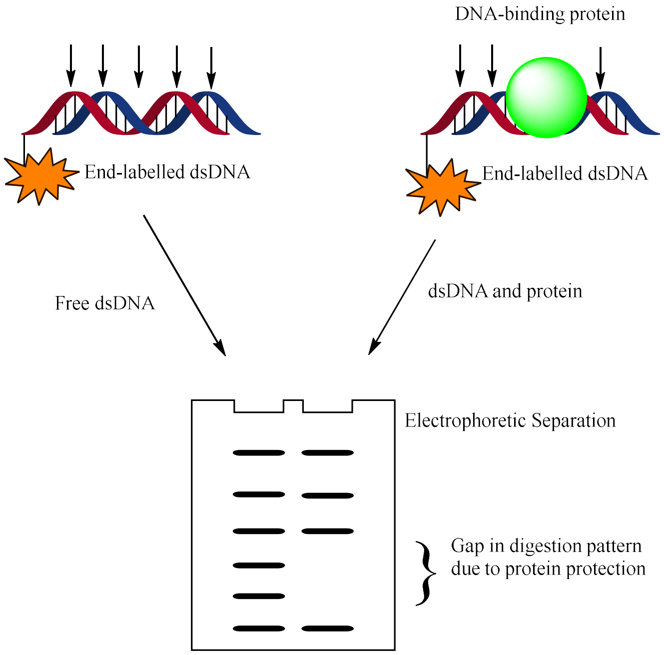


Figure 1.25: Schematic of a DNase I footprinting experiment.[156]

The digested duplex DNA strand is then analysed through gel electrophoresis, and the resultant strands are sequenced chemically. From the difference in lengths of DNA strand it is possible to deduce the binding site from the sites cleaved by Fe (II). The MGB “guides” the Fe (II) to the ends of the DNA, where cleavage occurs, an example of this system is compound **1.15** (Fig. 1.26 a). The information gained from this experiment is the identification of the binding site, as well as any mismatched sites as less common fragments. Qualitative judgments can be made to the specificity of a given ligand for a given sequence of dsDNA from the approximate ratio of the “target” and “mismatch” binding sites. These techniques however, are unsuited to the aim of this project, as they give no information on studying the dynamics of the DNA-ligand complex.

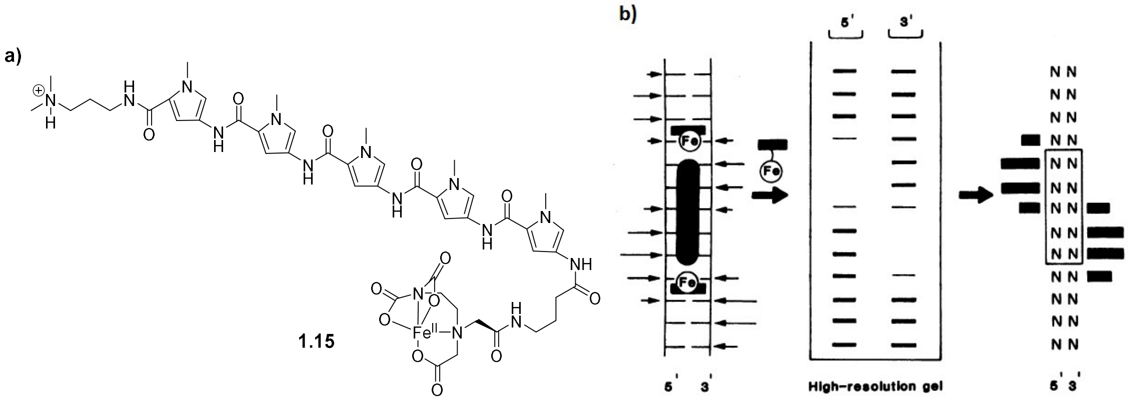


Figure 1.26: a) MGB derivatised with an Fe(II).EDTA complex, (b) Schematic representation cleavage patterns exhibited by complexation of **1.15** with dsDNA.[161] Reprinted by permission from Macmillan Publishers Ltd.: Nature © 1986.

This works opposite to footprinting in that the electrophoresis assay shows the binding site from cleaved DNA, not uncleaved DNA, as per DNase I fingerprinting. Binding constants can also be derived by titrating concentrations of ligand with respect to the target strand of DNA.[1] The equilibrium constant estimated can then be linked to the free energy change of the binding process shown in Eq. 1.2.[62]

1.5.2. Isothermal titration calorimetry.

The titration technique is commonly twinned with isothermal microcalorimetry, to determine the heat dissipated to the surroundings upon binding. Given that ligand-DNA interactions are reversible and involve noncovalent interactions, the enthalpy change for the reaction is small and hence, sensitive equipment is required for this technique. The temperature change of the surroundings is directly related to the enthalpy change of the reaction shown in equation 1.3. C_p defines the heat capacity of the surroundings.

$$\Delta H = C_p \, \Delta T \qquad \text{(Eq. 1.3)}$$

The entropy change of the reaction can now be calculated from equation 1.1. The thermodynamic parameters of DNA-MGB association are now known and conclusions can be drawn from the numbers of the three parameters, ΔG°, ΔH°, and ΔS°. If ΔG° is negative, the reaction is spontaneous and will occur at the temperature quoted, and *vice versa*. This, however, gives no indication of the kinetics of the reaction. For example, iron oxidation is thermodynamically favourable but reaction proceeds only slowly.[63] If ΔH° is negative, the reaction is exothermic and heat is lost to the surroundings. If it is positive, the reaction is endothermic and heat is taken in from the surroundings. Enthalpy alone does not determine the spontaneity of the reaction, as a high enough positive entropy change will drive the reaction at high temperatures. ΔS° is defined as a measure of the order of molecules in a system. Even if ΔH° is positive, chemical reactions can still be spontaneous, provided the entropy change is positive and the temperature is high enough i.e. $T\Delta S > \Delta H$. (Eq. 1.1). Conversely, processes where $\Delta S > 0$ can be still spontaneous provided they are sufficiently exothermic, and the temperature is low enough.

Biochemical methods are very useful for determining the specificity (if any) of a given ligand for the protein as well as derivation of thermodynamic data of ligand-DNA association. However, these techniques give no indication of the dynamics of the process, nor does it show how the ligand binds to DNA, nor any changes that occur in the structure of the DNA-ligand complex upon binding. X-ray crystallography provides high-resolution structural information on a ligand-DNA complex.

1.5.3. X-ray crystallography

Since the first report of the double-helical structure of DNA in 1953,[9] X-ray crystallography has been the technique of choice for studying the behaviour of duplex DNA when the system is perturbed, for example upon the binding of a ligand and dehydration.[7] This technique has been used to elucidate the binding site of **1.8** to a DNA oligomer, a 5'-AATT-3' site in the minor groove, with 2:1 stoichiometry (Fig. 1.21 b).[64]

X-ray crystallography gives a direct, visual representation of every atom in a macromolecule or a hydrogen-bonded complex. Although hydrogen atoms cannot be observed, inferences can be made about which atoms form hydrogen bonds by comparing the distances between the donor and acceptor, and inserting the hydrogen atom in the idealised position. If they are closer than expected, then it can be assumed a hydrogen bond is formed.

There also exist disadvantages with the use of crystallography as a technique. Firstly, because a crystal of DNA in the solid state does not necessarily reflect the nature of DNA in solution, and consequently care must be taken in the crystallisation process to ensure a solid-state structure representative of the solution-phase structure is obtained.[7] Moreover, X-ray crystallography gives no information on the dynamics of DNA-ligand binding; crystal structures only show the equilibrium structure of the complex. Since DNA is a dynamic structure, crystallography gives incomplete information on the molecular processes of and MGB: dsDNA complex, nor does it yield information as to how an MGB finds its sequence and hence is not a useful technique for studying these processes although it gives useful, complementary structural information to the more suited spectroscopic techniques discussed in the next section.

1.5.4. Ultraviolet Spectroscopy

DNA absorbs ultraviolet light at a maximum wavelength of 260 nm, due to the aromatic nature of the nucleobases. The absorption of UV light is accompanied by the promotion of an electron from the HOMO to the LUMO of the irradiated π-system. This method can be used to measure the concentration of DNA in a stock solution of such a low concentration that weighing the solid prior to dissolution becomes impractical. The concentration varies directly with the absorbance and are related by the extinction coefficient (Eq. 1.4).[62] The extinction coefficient of a single-stranded DNA oligomer is dependent on the length of the strand and its composition, and one method for approximation of this parameter is addition of extinction coefficients of the component dinucleotides in a given sequence of DNA, and subtracting the sum of the extinction coefficients of the nucleotides.[7]

$$A = \varepsilon lc \qquad \text{(Eq. 1.4 b)}$$

$$A = \log_{10}\frac{I}{I_0} \qquad \text{(Eq. 1.4 b)}$$

One frequently used application of UV spectroscopy is to study the thermal dissociation of dsDNA, referred to as melting. The reason why a melting curve is obtained (Fig. 1.27) comes from the hypochromicity exhibited by dsDNA compared to ssDNA. Hypochromicity is defined as the decrease in absorption of electromagnetic radiation of a particular species. When dsDNA is irradiated with UV radiation, the transition dipole moment caused by excitation of an electron induces a smaller transition dipole moment in the opposite strand, hence causing the hypochromicity. Upon melting of the duplex the transition dipole moment induced across a base pair is no longer present, and hypochromicity is no longer observed. .[7] When the double helix structure is denatured by heating referred to as "melting", the hydrogen-bonded double-helical structure breaks down and due to thermal motion the bases move away from each other. Consequently, because fewer strands of DNA are Watson-Crick base paired, less hypochromicity is observed, and this can be followed by UV spectroscopy. This gives rise to a characteristic sigmoid curve (Fig. 1.27). The temperature at which half of the DNA has melted, T_m, is characteristic of the duplex, and is sequence dependent.[7]

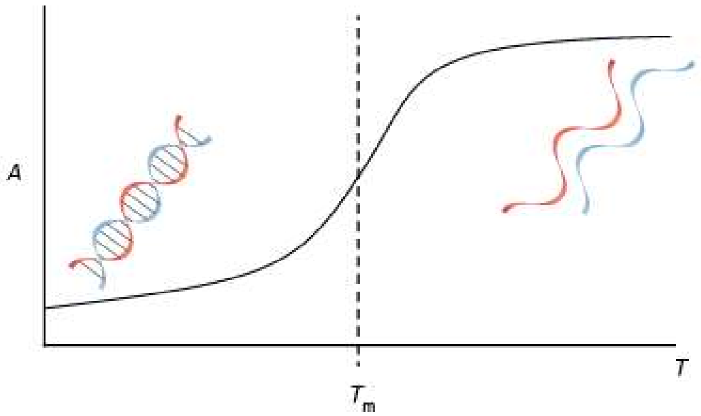


Figure 1.27: A typical melting curve, with graphical representations of dsDNA and melted DNA.

Given that the mechanism of melting of dsDNA comes from breaking the intermolecular H-bonds in a DNA duplex, the exact temperature at which a given length of dsDNA "melts" is dependent on the G-C content of DNA as these two bases form a stronger base pair with 3 hydrogen bonds (Fig. 1.3).[8] The absorbance/temperature plot for a given strand of duplex DNA is sigmoidal in shape. The initial increase in absorbance is due to the greater thermal motion of duplex DNA without dissociation of the two strands, and the consequent loss of hypochromicity.[7] As the strands begin to pull apart the gradient of the curve increases and reaches a maximum when half of the strands are dissociated. The curve then becomes flat when both the strands are fully dissociated. The melting temperature, T_m, is defined as the temperature where exactly half of the DNA has melted. Melting curves can be used to qualitatively investigate the stabilising effect of a particular ligand on duplex DNA. Because the structure of duplex DNA is maintained by the ligand, melting does not occur until the ligand is fully dissociated from the DNA, and consequently the graph is shifted to the right, i.e. T_m increases. UV spectroscopy only identifies T_m, it gives no indication of how a given duplex melts i.e. which base pairs dissociate first, when a given ligand dissociates, or over which timescales a DNA duplex melts. A time-resolved spectroscopic method is required for studying these complexes; the techniques most suited for these experiments are NMR and IR spectroscopy.

1.5.5: Nuclear Magnetic Resonance Spectroscopy (NMR):

UV spectroscopic techniques show the temperature range over which the DNA melts, as well as the stabilising effect the ligand has on the duplex. However, UV-Vis spectroscopy gives no information on the dynamics of the process of association, or dissociation of a ligand, nor does it provide any information as to the structure of the complex. To probe the dynamics of a chemical process, a different spectroscopic process is required. Two processes are used for biomolecular applications: NMR, which is used routinely, and less often, IR spectroscopy. A prerequisite of studying the dynamics of a biomolecule in an NMR experiment is knowledge of the structure of both the DNA and the ligand, typically obtained *via* X-ray crystallography or a separate NMR experiment. Two-dimensional techniques, as well as a high-field spectrometer are also required as the overlapping peaks of the bases can be difficult to resolve on a lower-field instrument. The main nuclei studied in analyses of DNA dynamics are ^{1}H and ^{31}P because of their high natural abundance, and their different positions in a given nucleotide.[7] Pulsed 2D-NMR techniques are frequently used to elucidate the structure of a small molecule in complex with a dsDNA oligodeoxyribonucleotide, the three techniques routinely used are COSY, TOCSY and NOESY.

COSY and TOCSY are often used to analyse nuclear spins from the same residue through correlation. The pulse sequence for a COSY experiment is shown in Fig. 1.28, a TOCSY experiment uses a similar pulse sequence, but repeated 90° pulses are used in order to create isotopic mixing, leading to longer range correlations.[7] For dsDNA applications, the C1′ H-atom (Fig. 1.2) of any nucleoside has a characteristic chemical shift (5.0–6.3 ppm)[65] and hence the positions of the sugar protons can be derived from the correlations, although this technique is only useful for small oligonucleotides (< 12 b.p.), due to signal degeneracy.[7]

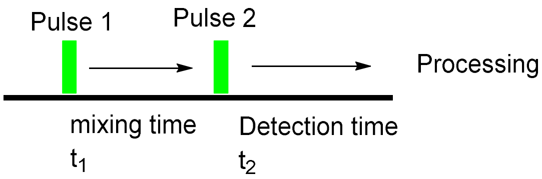


Figure 1.28. Schematic of the COSY pulse sequence – Pulse 2 is repeated over a period of time (usually 70 ms) for a TOCSY experiment.[162]

The COSY effect comes from the lengthening of the spin-spin relaxation time (t_2) (the time required for magnetised spins to return to equilibrium following a 90° pulse) spins separated by few chemical bonds compared to spins further away. COSY can also be used for elucidation of base sequence as the sugar protons are connected to the next base by the phosphodiester.[7] In addition, the design of a TOCSY experiment, essentially a COSY experiment with more 90° pulses between t_1 and t_2 removes the effects of chemical shifts and allows for correlation of all spins in an oligonucleotide, with a positive relationship between the strength of a correlation and the distance between two nuclei.

In the case of NOESY experiments, the pulse sequence is shown in Figure 1.29. The pulse sequence is similar to COSY and TOCSY, but an extra RF pulse, polarised perpendicular to both the initial pulse and the magnetic field vector is used to induce the NOESY effect.

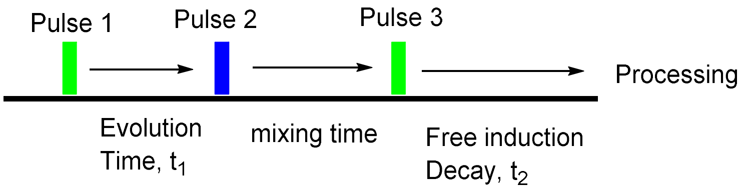


Figure 1.29. Schematic of the NOESY experiment. The perpendicular pulse is highlighted blue.[162]

NOESY spectroscopy makes use of the nuclear Overhauser effect in which two spins close together in space can transfer energy as a means of decay.[62] A cross-peak is seen only if the spins are less than 5Å away, as the effect varies inversely with the sixth power of the distance of the two peaks. nOe effects are also dependent on the size of the solute, and the viscosity and temperature of the solution.[65,7] NOESY experiments are good for investigating the conformation of DNA, especially through the sugar pucker. Changes in DNA structure will allow an NMR experiment to determine whether a DNA-binding small molecule intercalates, or binds the minor groove. Furthermore, for a C2'-endo sugar (B-DNA), in which the C3′ of ribose is in the same plane as the other carbon atoms (Fig. 1.30), a nOe correlation exists between H1′ and both H2′/H2″ (Fig. 1.2) and correlation between H1′/H3′ is not seen, and *vice versa* for a C3′-endo pucker (A-DNA). Hence, changes in the hydration sphere of a DNA duplex can be followed by the nOe.

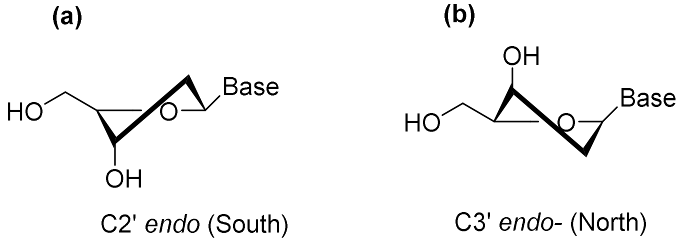


Figure 1.30: **(a)**, C2'-endo and **(b)**, C3'-endo sugar puckers.[163]

Furthermore, in B-DNA there exists a nOe correlation between H1′ and H8 (Fig. 1.1) of purine bases (A,G), and H6 of the pyrimidines (C,U,T), and hence, lengths of B-DNA can be sequenced from an nOe experiment (Section 4.1).[65] This allows conformational changes in DNA due to ligand binding to be identified. A-form DNA can be positively identified from the NOESY data from the glycosidic bond angle (between the base and the sugar), and defined as either anti- or syn- to H1′.[7] If the sugar pucker has been well-defined from a previous experiment the intensities of the nOe will show the sugar pucker. Repeated nOe data shows B-DNA, and similar nOe correlations can be seen for A-DNA.

Much structural information can be obtained through carrying out NMR experiments but the main advantage NMR has over the other techniques is the ability to investigate dynamics in solution. Over millisecond timescales, NMR can identify interactions between ligands and DNA as well as conformational changes, shown by changes in chemical shifts. For slow rates of proton exchange, saturation transfer can be used.[7] This detects protons in different chemical environments, and the intensity is inversely related to the spin-lattice relaxation time, T_1.[7] Several transitions can be seen and this method has been used to probe the activity of minor groove binding ligands, such as distamycin A **1.8**.[66] By this method, the 1:1 and 2:1 overlapping binding modes of distamycin to dsDNA were established, as well as the opening of the minor groove to accommodate the second equivalent of distamycin (Fig. 1.21 a). For slightly faster exchange rates, (10^4-10^8 s^{-1}) a method involving the exchange of imino protons has been developed. The imino protons are used as a probe, because they resonate in a silent area of the spectrum (12-14 ppm), and are not solvent accessible in B-DNA.[7] They act as an indirect measure of fluctuations in the chemical environment brought on by ligand binding due to "opening" of the DNA structure due to local fluctuations in the structure. The imino proton is then free to exchange with the solvent, and the rate at which this happens can be measured by NMR, by measuring the integral of the imino peak, as exchange occurs with D_2O the intensity drops.[67] For more rapid exchange processes, the effect of T_1, defined as the time taken for perturbed system to return to equilibrium must be observed.[7]

The limitations of NMR are that the timescales are confined to the spectrometer frequency and the nucleus being studied. For ^{1}H NMR a shift of 1 ppm over the course of an experiment using a 500 MHz spectrometer can resolve events only the timescale of 2 ms.[7] For processes that exhibit faster kinetics, such as probing hydrogen bond formation and dissociation processes in ligand-DNA association, a faster timescale is required.[59] Time-based IR spectroscopic measurements have been shown to work on femtosecond timescales, and hence is the dominant technique used in the course of this project.

1.5.6. Infrared Spectroscopy.

Vibrational spectroscopy involves changes in the vibrational state of a molecule upon absorption of a photon of a particular frequency. This leads to a slight increase in bond length of the excited functional group, and increased molecular motion. The behaviour of a diatomic molecule upon absorption of low energy radiation and hence a small perturbation of the system can be approximated to that of a parabola, whose function is expressed in Eq. 1.5, and gives the separation between two vibrational energy levels.

$$V= \frac{1}{2}k_f x^2 \qquad \text{(Eq. 1.5)}$$

The energy required to excite a particular bond is given in equation 6, and is related to Hooke's Law.[62]

$$V = \frac{1}{2\pi c}\sqrt{\frac{k_f}{\mu}} \qquad \text{(Eq. 1.6)}$$

The force constant, k_f, is related to the curvature of the potential energy of the bond close to equilibrium bond length, the steeper the curve, the stronger the bond. The global selection rule for infrared absorption is dependent on a change in the overall electric dipole. The actual selection rule is $\Delta V = \pm 1$, and thus every molecule with a heteroatom bond will contain at least one infrared active bond, homonuclear diatomic stretching vibrations do not change the overall dipole moment and hence do not generate or absorb radiation. These vibrations are referred to as infrared inactive.[62] The dominant transition in a molecule is from $v = 0$ to $v = 1$, and this is referred to as the fundamental vibration. This comes from the width of the Boltzmann distribution of vibrational energy being approximately 200cm^{-1} at room temperature,[62] and as the energy required to obtain higher vibrational states is generally far higher than this, the ground-state (i.e. v = 0) is almost exclusively populated.

The harmonic approximation implies that bond strength does not weaken upon absorption of higher energy photons, and consequently that bonds will not dissociate. Hence, a correction term must be introduced to correct for the convergence of energy levels as v increases and hence the equations for V and G change, (Eq. 1.7 and 1.8). This is the expression that describes the familiar Morse potential energy surface.

$$V = hcD_e \ \{1\text{-}e^{-a(x)}\}^2 \qquad \text{(Eq. 1.7 a)}$$

$$a = \left(\frac{m_{eff}\omega^2}{2hcD_e}\right) \qquad \text{(Eq. 1.7 b)}$$

Consequently, the vibrational spectroscopic term for an anharmonic oscillator also changes and is shown in equation 1.8b.

$$\tilde{G}(v) = \left(v+\tfrac{1}{2}\right)\tilde{v} - \left(v+\tfrac{1}{2}\right)^2 x_e \tilde{v} \qquad \text{(Eq. 1.8 a)}$$

$$x_e = \frac{a^2\hbar}{2m_{eff}\omega} = \frac{\tilde{v}}{4\breve{D}_e} \qquad \text{(Eq. 1.8 b)}$$

Hence, because of the convergence of the energy levels and the inclusion of a dissociation energy term, there are only a finite number of values that v can take (0,1,2...v_{max}). The deviations from the harmonic oscillator model for a chemical bond are referred to as anharmonicity.[62] Anharmonicity also allows for the presence of overtones (transition from v = 0 to {v + >1}), despite the transition being forbidden. This is due to the selection rule being derived from the harmonic oscillator model, and therefore the selection rule is not rigorously true for anharmonicsystems.[62]

1.6: Two-dimensional IR spectroscopy and biomolecular dynamics

1.6.1. Theory and experimental method

2D-IR spectroscopy can be broadly thought of as analogous to 2D-NMR spectroscopy, in that pulses of IR, instead of a magnetic field and RF radiation, are used to induce a coherence and then probe the structural characteristics of a sample.[68,69] A typical time-domain experiment involves two narrowband "pump" pulses separated by the coherence time, τ (Fig. 1.31).[70] This pulse sequence generates firstly a vibrational coherence, and the second pulse creates a population of excited vibrational states.[70] A third pulse, referred to as the "probe" pulse is then applied after the waiting time (T_w, Fig. 1.31), which gives rise to an oscillating dipole. The oscillating dipole relaxes through radiation of an infrared photon, which after heterodyning to a reference pulse to recover phase information, gives rise to the raw data, a time-domain interferogram.

Fourier transform of this interferogram with reference to the frequencies of the first pump-pulse and the probe pulse gives the 2D-IR spectrum as a contour plot, similar to 2D-NMR experiments (*Section 1.5.4*). The off-diagonal cross-peaks in the contour plot from an experiment such as in Fig. 1.31 come from vibrational transition dipole moments (i.e. excited functional groups) close together in space, which can transfer energy between each other, broadly similar to a NOESY experiment.[70,71,72]

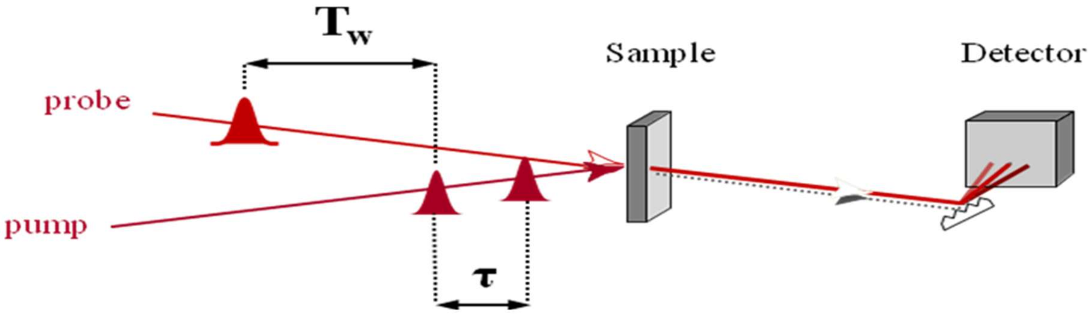


Figure 1.31: Schematic of a 2D-IR experimental setup.[164]

The waiting time, T_w, (analogous to T_2 in 2D-NMR) must be shorter than the experimentally-derived relaxation time of the functional group in question, and are therefore generally limited to a few picoseconds.[59,69] However, it has been previously shown that key biomolecular processes do occur on the picoseconds timescale, and hence 2D- IR is uniquely suited to probing biomolecular processes, for example the association of a ligand to dsDNA.[73]

1.6.2. 2D-IR investigations of the dynamics of dsDNA.

In recent years, 2D-IR has proven its value in investigation of the nonequilibrium dynamics of base monophosphates,[74] the solvation dynamics of water at the surface of dsDNA,[75] and the nonequilibrium dynamics of dsDNA oligomers containing both G-C[72] and A-T base pairs.[71] These studies have shown that 2D-IR is a useful tool for studying the dynamics of dsDNA, and the key role of water molecules governing the binding of both proteins and small molecules to dsDNA. At the time of writing, no such experiment has been reported on a complex of a dsDNA oligomer with a small molecule. The work presented in this thesis describes efforts undertaken towards understanding the dynamics, both of dsDNA and of the associated water molecules in determination of the process of association and dissociation of MGBs to dsDNA. To aid in the interpretation of the spectra, the MGBs will have to be outfitted with a reporter functional group with a fundamental IR absorption band removed from the absorption frequencies of dsDNA, akin to using a fluorophore for single-molecule fluorescence experiments.

Section 1.6.3: Use of vibrational probes

The most studied groups for suitability as use as infrared probes include nitriles[76], azides[77,78] and carbonyls, specifically the amide I band for protein IR spectroscopy (Fig. 1.34).[79] Ideally, an infrared probe will have the following properties:[78]

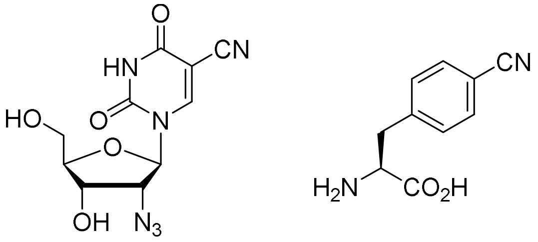
i. A fundamental absorption frequency distinct from those present in the underivatised molecule.

ii. A narrow and intense absorption band. This simplifies interpretation of the spectrum, and minimises the concentration of sample required for procurement of good data.

iii. Small size to minimise perturbation of structure from the natural substrate, a problem with the use of fluorophores.

iv. Sensitivity to changes in local environment, allowing investigation of the local environment via IR spectroscopy.

v. Chemical stability and ease of synthesis. Ideally, this should be achieved in one step from the molecule under study, or an immediate precursor, so no new and difficult synthetic routes will have to be sought to obtain a derivatised molecule.



1.16 **1.17**

Figure 1.34: An analogue of 2'-deoxyuridine containing both azide and nitrile functional groups e.g. **1.16**,[165] and (L)-para-cyano phenylalanine **1.17.**

Ultrafast spectroscopy has previously been used to study the dynamics of noncovalent interactions between proteins,[80] and also with proteins and small moleucles.[81] The probes used to date have been intrinsic to the protein or small molecule being studied due to synthetic tractability and strong absorption profile. The interactions of heme proteins at equilibrium have been studied,[6] as well as the inhibition of catalase by nitric oxide,[82] giving new and powerful insights into how these compounds behave in solution, and the implications for their biological action. These investigations have shown that biomolecules in solution behave differently to that of the crystal structure and that small structural changes are crucial for understanding the activity of these proteins The conformations of smaller biomolecules such as NAD have also recently been studied.[81] The dynamics of dsDNA have not extensively been studied to date, either of a native duplex, or of dsDNA in complex with a ligand. Studying the dynamics of DNA will yield insights into how MGBs associate with dsDNA, and given that many MGBs are small molecules, the incorporation of azide probes into these compounds should be synthetically tractable. Similar insights into how these compounds interact with dsDNA will allow for the design of programmable MGBs with anticancer or antibacterial properties.

1.6.4. Previous attempts to study dsDNA: MGB dynamics by other techniques.

To date, the only fruitful investigations of dsDNA:MGB dynamics have involved nanosecond-scale simulations, and lacked structural resolution.[5] Because of the fast association and very slow dissociation kinetics of formation of the complex, the microsecond timescale resolution typically available to NMR experiments cannot show the subtle conformational changes in DNA structure, and time-averaged and broadened signals are observed.[83] However, time-averaged signals do suggest that dynamic structural changes in dsDNA do exist, and are essential for accommodation of the MGB. Furthermore, the inherent flexibility and the narrow minor groove of A-T tracts of dsDNA have also shown that these sequences are “preformed” for ligand recognition, indicating that small molecule ligands find their target sequence by shape readout.[84]

The overall objective of this project is to acquire a deeper understanding of the molecular processes that govern the formation of dsDNA: MGB complexes. It is envisioned that the knowledge obtained from this work will inform better simulations of the motions of dsDNA in the presence of an incoming MGB, and hence, the design of better DNA binding therapeutic agents.

1.7: Hypothesis

The hypothesis that will be tested over the course of this work is the shape-selective mechanism of binding of an MGB (e.g. **1.9**) will predominantly determine of the recognition of target dsDNA sequence. As observed with the rate-determining step for binding of **1.8**[48] and the lower ΔG° value observed for **1.9**[83] in complex with a suboptimal binding sequence, the conformational changes that occur in a representative DNA duplex to accommodate the ligand is the key determinant favouring complex formation.

Due to the entropic thermodynamic driving force of binding of **1.9** to dsDNA, the mechanism of melting of a given MGB: dsDNA complex is hypothesised to occur through dissociation of the ligand, followed by rapid melting of the duplex (Fig 1.35).[19] Hence, if the equilibrium binding constant, K, is governed by the enthalpy change, ΔH° (Eqs. 1.1 and 1.2) upon binding of an MGB to a dsDNA oligomer *i.e.* base readout, the energy required to thermally break the hydrogen-bonded complex will be lower than that required to break the hydrogen bonds at the end of the DNA oligomer, which due to co-operativity effects, will be weaker than the hydrogen bonds formed at the middle of the duplex.[7] Consequently, the DNA oligomer will melt at the ends first by a “peeling” mechanism (Fig. 1.35) leading to complete dissociation of the complex.

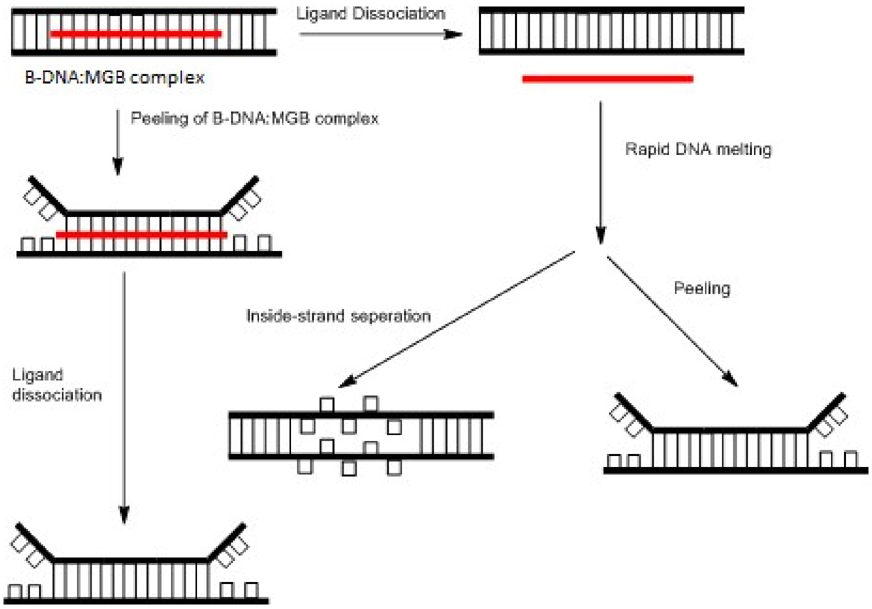


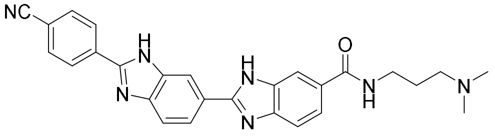
Figure 1.35. Mechanisms by which a dsDNA: MGB complex can melt. DNA shown in black, ligand in red.[143]

This can be related to the Gibbs free energy equations (1.1) and (1.2) (*Section 1.4*). However, if the thermodynamics of MGB: dsDNA ligand formation is governed by the positive entropy change generated upon formation of the complex *i.e.* shape-readout, as the temperature increases ΔG° increases until upon a temperature at which the free energy change of complex formation becomes positive, and the ligand dissociates. As the temperature of the complex is now above the melting temperature of the free duplex, rapid melting of the dsDNA oligomer occurs. This can either occur by a "peeling" or inside-strand separation (bubbling) mechanism (Fig. 1.38), dependent on base sequence. As enthalpy values are usually much higher than entropy values, it follows that a base readout mechanism (enthalpy-driven) is less sensitive to temperature than a shape readout mechanism (entropy-driven) *(Section 1.4.4)*. Hence, an MGB: dsDNA complex which associates by a direct readout mechanism will stabilise the complex to such a degree that the DNA duplex will start to peel and the melting DNA duplex causes the ligand to dissociate. If, however, the thermodynamics of dsDNA: MGB association are governed by indirect readout, ligand dissociation should occur first, followed by rapid melting of dsDNA. 2D-IR spectroscopy can discriminate between these two mechanisms through correlation of the vibrational spectra of dsDNA, melted DNA, and the DNA:MGB complex.[59] 2D-IR spectroscopy will be used to probe the dissociation of a complex of dsDNA and an MGB with a reporter functional group. Consequently, using derivatives of **1.9** outfitted with azide functional groups will allow for direct observation of a melting dsDNA: MGB complex by 2D-IR.

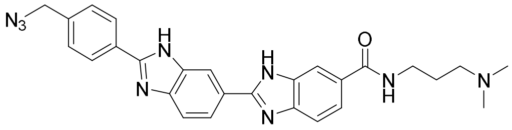
1.8. Objectives.

The objective of this project is to interrogate the binding mode of an infrared-labelled MGB: dsDNA complex by 2D-IR. The specific aims are threefold:

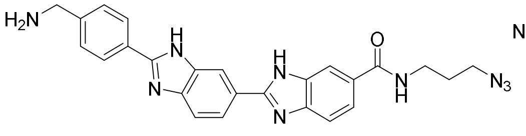
i. To synthesise azide-bearing MGB analogues of **1.9**, **1.18 – 1.21** (Fig. 1.36)
ii. Interrogation of the binding mode of MGB: dsDNA complexes by a combination of electronic and vibrational spectroscopy.
iii. Structural characterisation of the binding mode of MGBs by NMR spectroscopy.



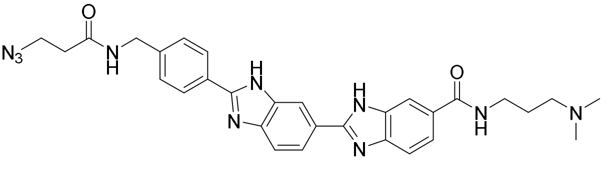
1.18



1.19



1.20



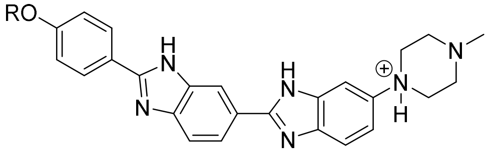
1.21

Figure 1.36. The target molecules to be synthesised over the course of this project.

2. Synthesis of azide-labelled derivatives of Hoechst 33258.

2.1. The binding mode of Hoechst 33258 to dsDNA

Hoechst 33258, (**1.9**, Fig. 2.1), is the archetypal minor groove binding ligand. It consists of three linked aromatic moieties: two benzimidazoles and a phenyl ring. Upon binding to A-T rich regions of dsDNA, its fluorescence quantum yield increases sixty-fold (from 0.01 to 0.6)[85]. Due to this property, compound **1.9** is used extensively as a chromosomal stain for fluorescence microscopy.[86] Other derivatives of **1.9** are known, for example Hoechst 33342 **2.1**, the ethoxy derivative of **1.9**, which has been previously shown to be more cell-permeable than **1.9**.[87] Hoechst 33258 and its derivatives due to their easy availability have served as model compounds for studying complexes of small molecules with dsDNA.



R = H, **1.9**
R = Et, **2.1**

Fig. 2.1: Hoechst 33258, **1.9**, and Hoechst 33342, **2.1**

Compound **1.9** and derivatives thereof have the following structural features shared by small molecule MGBs *(Sections 1.4.2-1.4.3)*. The concave structure of **1.9** is complementary to the curvature of the minor groove of dsDNA. **1.9** is also a planar molecule, which also aids in the recognition of the target sequence of dsDNA.[88] The cationic piperazinium moiety confers water solubility to **1.9** and crystallographic evidence exists for an electrostatic interaction between the piperazinium group and the anionic phosphate backbone.[64]

Compound **1.9** binds in a 1:1 stoichiometry to A-tract dsDNA, although it does not differentiate significantly between the exact sequence of the tract, **1.9** will bind to any sequence of four to six consecutive A-T base pairs with high affinity (Table 2.1).[89] A distinct preference for ApT over TpA base steps has also been described;[90] this is due to the large positive roll angle observed for TpA steps, leading to a wider minor groove, and hence a lower free energy of binding for **1.9**. Conversely, ApT steps exhibit a similar negative roll angle, causing narrowing of the minor groove.[91] A co-operative 2:1 stoichiometry binding mode similar to distamycin A (Fig. 1.21) has not been observed to exist, although a non-cooperative consecutive binding mode for 2 equivalents of **1.9** with A_4T_4 tracts of dsDNA has been observed.[92]

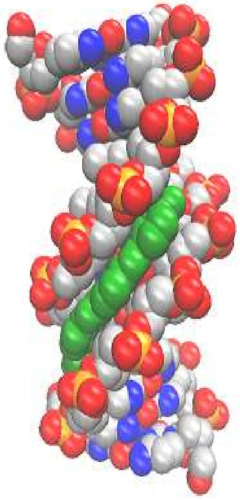


Figure 2.2: Crystal structure of dsDNA with one molecule of **1.9** bound. (PDB ID: 264D)[64]

Sequence	K_d (M)
5'-AAATTT-3'	$(3.5 \pm 1.5) \times 10^8$
5'-ATATAT-3'	$(5.0 \pm 2.0) \times 10^7$

Table 2.1: Dissociation constants of an A-tract and an AT repeat of dsDNA with **1.9**.[166,49]

Other binding modes have been observed by both mass spectrometry and ITC, these include nonspecific partial intercalation of **1.9** across a base pair,[53] "laddering" of **1.9**, templated by ssDNA,[93] and electrostatic interactions between the phosphate backbone of dsDNA and the positively charged piperazino group of **1.9**.[94] However, these binding modes do not compete with the binding mode shown in Fig. 2.2, and are only observed under nonphysiological conditions, or when the molar ratio of **1.9** greatly exceeds the number of equivalents of binding sites in the dsDNA strand.

2.1.1. Thermodynamics of formation of complexes of H33258 with DNA.

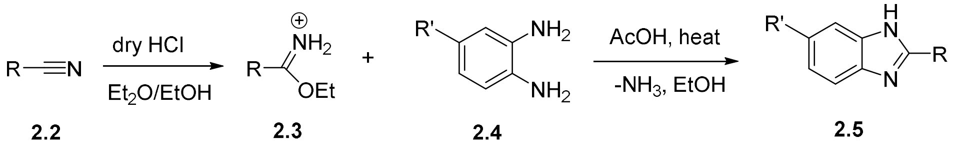
The large K_d values shown in Table 2.1 indicate a strongly negative free energy change upon duplexation of -11.3 kcal.mol^{-1}.[95] Isothermal titration calorimetry studies have also shown that formation of complexes between **1.9** and dsDNA is entropically driven (consistent with shape readout) rather than enthalpy (base readout) (*Section 1.3*).[19] There exists some debate on whether the association between **1.9** and dsDNA is an exothermic[5,96] or an endothermic[97,53] process, although the absolute consensus value of the enthalpy change is in the range ±10 kcal.mol^{-1}. This indicates that the formation of a hydrogen-bonded complex (an enthalpic process) is not the main driving force for the formation of the complex, rather a positive entropy change coming firstly from removal of the class II bound water molecules to the bulk solvent.[14] Hydrophobic burial of **1.9** into the nonpolar minor groove was also observed, as evidenced by a negative ΔC_p (-256 ± 10 cal.K^{-1}mol^{-1}).[96] The heat capacity change is consistent with returning of bound water to the bulk solvent. Conformational changes of either the ligand or the duplex DNA couple could explain the change in heat capacity, but no significant conformational changes are exhibited by the complex compared to free dsDNA (Fig. 2.2).[96] Addition of osmolytes to a solution of the complex of **1.9** with dsDNA have also shown that despite the dehydration observed by calorimetry, more water molecules are associated with the **1.9** complex compared to a free duplex. It is likely that these water molecules are located outside of the complex, and consequently have more rotational and translational degrees of freedom compared to bound water free dsDNA – and hence a positive entropy change is observed. Despite the negative enthalpy change associated with formation of a hydrogen bonded complex between the dsDNA and incoming **1.9**, this is offset by the positive enthalpy change associated with the removal of the 60 ± 13 class II structural water molecules associated with the duplex to the bulk solvent (Fig 1.24).[96] The debate within the literature is not that the formation of a

MGB: dsDNA complex is driven by entropy or enthalpy, the debate is that whether the endothermic removal of the class II water molecules is completely offset by the formation of hydrogen bonds between **1.9** and the edges of the DNA bases.

Structural investigations of complexes of **1.9** with dsDNA have shown that **1.9** does not directly recognise the hydrogen bonding pattern of the minor groove; it selects its target sequence of dsDNA by shape readout, namely the narrow minor grooves exhibited by A-tracts.[35] NMR and X-ray crystallographic studies of the **1.9**: dsDNA complex have shown that all hydrogen bond contacts between **1.9** and DNA are either with the phosphate backbone, water mediated or a bifurcated hydrogen bond across an A-T base pair.[98,99] As presented in *Section 1.1.5*, the bound water molecules making up the spine of hydration do not represent the edges of the bases in the DNA sequence,[52] and hence, these hydrogen bond contacts are, by definition, nonspecific. Less well-understood however, are the dynamics of association between **1.9** and dsDNA, namely the supramolecular interactions between dsDNA and an approaching ligand leading to formation of a complex. The work carried out over the course of this project will contribute to a fuller understanding of the mechanism of action of MGBs in complex with dsDNA, thus hopefully leading to the design of better DNA-binding therapeutics.

2.2. Previous Syntheses of Hoechst 33258 derivatives

The synthesis of Hoechst 33258 was first described by Löwe *et al.* in 1974,[100] and the key synthetic steps were reported as two consecutive Pinner-type cyclisations to furnish both benzimidazole rings (Scheme 2.1). The nitrile **2.2** is protonated by the HCl gas, and undergoes nucleophilic attack by ethanol to form the iminium salt **2.3**. The salt **2.3** can then undergo condensation with diamine **2.4** to furnish the 2,6-disubstituted benzimidazole **2.5**. The Hoechst core is generally built from right to left as shown in scheme 2.1, with sequential Pinner reactions forming the two benzimidazole rings.



Scheme 2.1: Pinner-type cyclisation forming a 2,6-disubstituted benzimidazoles. Yields typically 65-70%, dependent on substituents.[100,103]

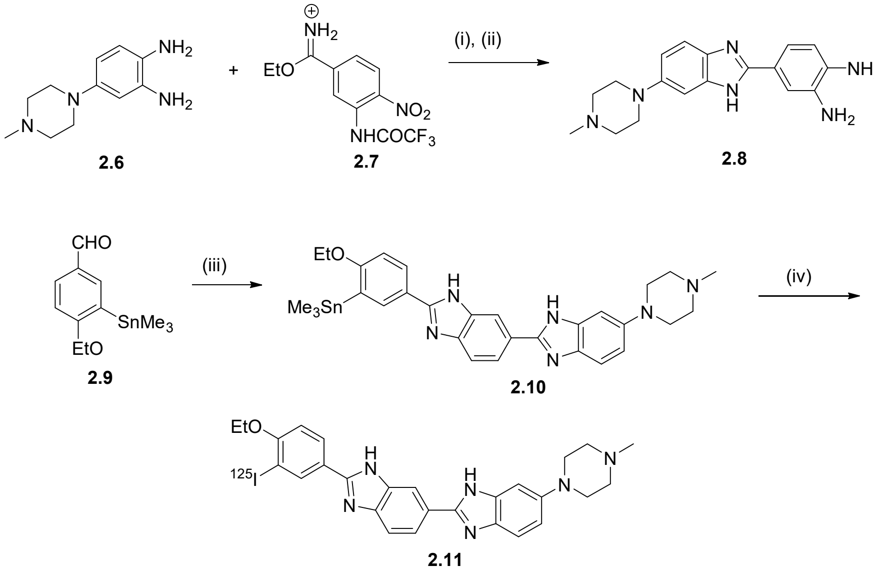
This synthetic route has proven modular, leading to easy derivatisation of the Hoechst core – and hence, derivatives of **1.9** and **2.1** have been prepared with either radioactive or DNA alkylating moieties for investigation of derivatives of **1.9** and **2.1** as possible anticancer agents.

2.2.1. Radiolabelled derivatives of H33258.

Given the combination of favourable properties of **1.9** as a potential drug candidate,[101] its water solubility, ease of large-scale synthesis and selective, highly exergonic binding mode to A-T tracts of dsDNA, Hoechst 33258 has been investigated as a possible anticancer agent. It was found that although **1.9** and **2.1** have shown some activity against both helminths and cancer cell lines, they have proven only poorly cytotoxic.[102,87] While this leads to its important uses in molecular biology, the poor toxicity observed for Hoechst dyes renders it as a poor drug candidate, although the modular synthetic route to **1.9** allows for easy derivatisation to analogues bearing radioisotopes, for example ^{125}I.

One such example, a derivative of **1.9** was achieved by the Kassis group and reported in 1996.[103] A derivative of **1.9** was synthesized in a manner like that described by Löwe, and is described in Scheme 2.2.[103] The dimethylaminopiperazinyl compound **2.6** was cyclised with the Pinner salt **2.7** in acetic anhydride and after a Pd/C catalysed hydrogenation to furnish **2.8**. Finally, a Traube reaction with the aldehyde **2.9** in nitrobenzene/toluene furnished **2.10**. Compound **2.11** could be efficiently prepared from **2.10** using radioactive NaI, lactoperoxidase and small amounts of H_2O_2.[103] The presence of the bulky iodine atom had a slight negative effect on the binding constant of association of radiolabelled H33258 ($K_a = 2.7 \pm 0.1 \times 10^7$ M$^-$

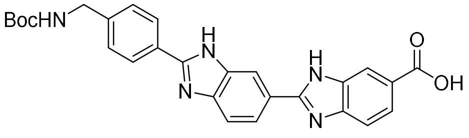
^{1}) and showed some toxicity against two tumour cell lines. More importantly, the drug did not cross the blood-brain barrier, and no deiodination was observed.[103] This indicates that localisation of any possible chemotherapy agent is theoretically possible, and a fuller understanding of the mechanism of binding of these compounds will allow for the design of cytotoxic derivatives of **1.9** with optimised binding and cytotoxic profiles.



Scheme 2.2: Synthesis of radiolabelled **1.9**, **2.11** highlighting the benzimidazole forming steps. *Conditions and reagents:* (i) **2.6**, **2.7** (both 1.0 eq.), Ac_2O, 55 °C 12 h, (ii) Pd/C 10 wt%, EtOH, 1.25 h, (iii) **2.8**, **2.9** (both 1.0 eq.) $PhNO_2$/PhMe (4:1) 100 °C, 24-36 h. (iv) $Na^{125}I$ (2 mC_i) in 0.1 M NaOH, 1 μL lactoperoxidase (4 mg. mL^{-1}), 10 μL H_2O_2 (0.06 vol), pH 4.9 buffer, 5 min, radiochemical yield 80%.

2.3. Aims of Chapter 2.

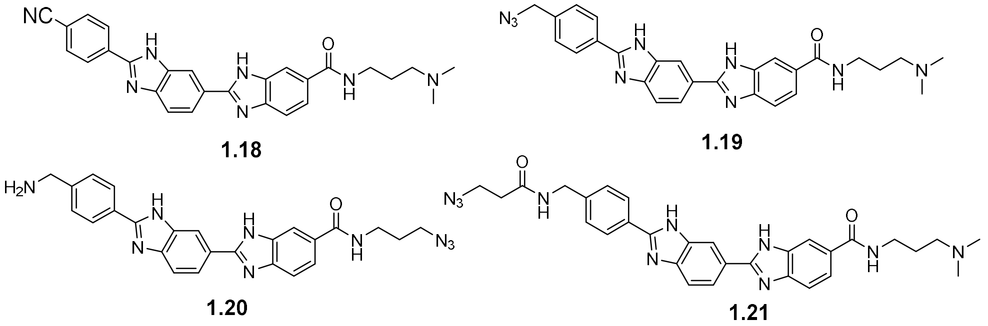
The derivative of **1.9** selected for derivatisation is the previously described Boc-protected amino acid derivative **2.12** (Figure 2.3).[104]



2.12

Figure 2.3. The Boc-protected Hoechst amino acid **2.12**

Compound **2.12** previously been incorporated into peptides (Scheme 2.4), and its binding mode to A-T tracts of dsDNA has previously been described (*Section 4.1*).[94] Provided that this compound can be synthesised on a sufficiently large scale, easy derivatisation to the target compounds **1.18-1.21 (**Figure 2.4) should be possible. Hence, dynamic data on both sides of the Hoechst core can be obtained. Furthermore, the nitrile in **1.18** is reduced to the amine through the synthesis of **2.12**,[104] so this should allow for the facile synthesis of a nitrile-bearing derivative of **1.9** and hence comparison of azides and nitriles as IR probes can be carried out.



1.18

1.19

1.20

1.21

Figure 2.4. The target compounds **1.18-1.21.**

The aims of chapter 2 are twofold:

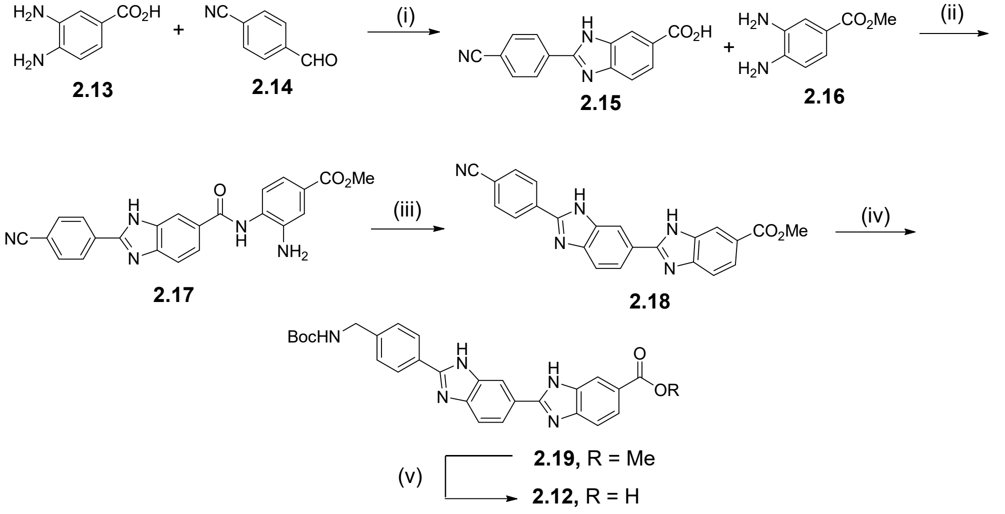
i. To synthesise the four IR-probe bearing compounds **1.18-1.21** from **2.12**.

ii. To evaluate the IR signal in the azide/nitrile region in the aforementioned compounds for use in 2D-IR experiments.

2.4. Synthesis of the Hoechst Amino Acid Core (2.12).

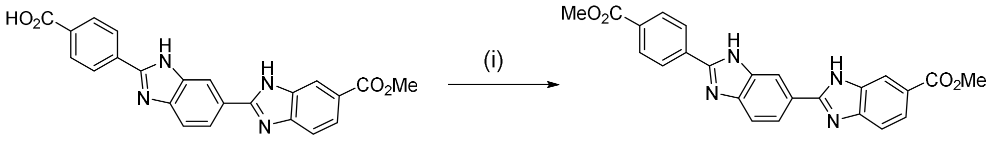
The synthetic route taken towards compound **2.12** is shown in Scheme 2.3, in a manner first reported by Behrens *et al.* and was synthesized in our hands on a 5.5 g scale in 31% overall yield.[104] A Traube reaction between **2.13** and **2.14** in nitrobenzene/DMF afforded the cyano acid **2.15** in 70 % isolated yield. A HATU- mediated amide coupling of **2.15** with methyl-3,4-diaminobenzoate afforded the *ortho*-amino amide **2.17** in 81% yield. An acid-catalysed

Pinner-type ring closing reaction then furnished the Hoechst-type cyano-ester **2.18** in 75% yield.



Scheme 2.3: Synthesis of compound **2.12.** *Conditions and reagents*: (i) $PhNO_2$ 150ºC**,** 24hrs, 70%. (ii), HATU (1.0 eq.), DIEA (2.5 eq.) then methyl-3,4-diaminobenzoate (1.0 eq.) 81%. (iii) $PhNO_2$, c.H_2SO_4 cat. 75%. (iv) $NaBH_4$ (7.0eq.), Boc_2O (2.0eq.), $NiCl_2.6H_2O$ (0.1eq) 75%. (v) LiOH (aq. 1.0M, 5eq.)/MeOH 1:1 v/v 50ºC, 3hrs 96%.

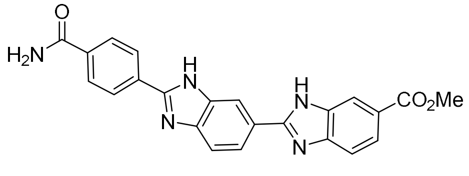
The original preparation as outlined by Behrens *et al.* towards **2.18** indicated the use of 2 equivalents of concentrated sulfuric acid. Under these conditions ring-closing did occur, but this was accompanied by a second side-reaction in which the nitrile of **2.18** was observed to efficiently hydrolyse to the carboxylic acid **2.20**. The side-reaction was confirmed through precipitation of the crude carboxylic acid from the recrystallisation mixture and derivatisation by synthesis of the methyl ester **2.21** (Scheme 2.4). Compound **2.21** was purified by rerystallisation from DMF/water and was recovered in 82% yield. It was found that reducing the number of equivalents of H_2SO_4 to a catalytic amount resulted in minimal hydrolysis, furnishing the desired product **2.18** in 75% yield with no noticeable decline in the rate of ring closure. Although the use of acid was necessary to effect ring-closure, it was found that better yields of **2.18** were obtained when a smaller volume of H_2SO_4 was used.



2.20 **2.21**

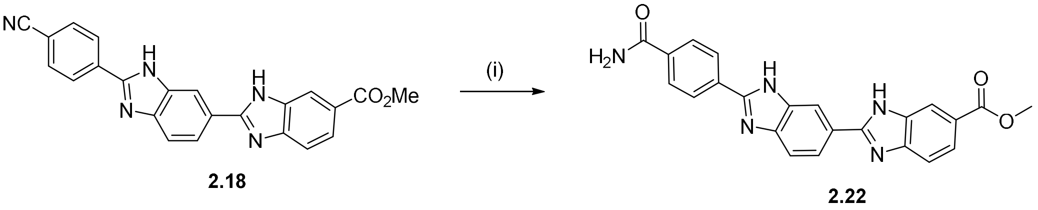
Scheme 2.4: Synthesis of the diester **2.21**. *Conditions and reagents:* (i) $SOCl_2$ (1.5 eq.), MeOH (solvent), reflux, 3 h 82%.

Reduction of **2.18** to furnish **2.19**, originally reported by Behrens *et al.* to proceed by a one-pot Pd-catalysed hydrogenation and Boc- protection was found to proceed only slowly, with only trace amounts of product detected after 48 h under 40 atm pressure of H_2 by mass spectrometry. An alternative nickel boride mediated reduction first reported by Caddick *et al.*[105] was found to work although yields were variable due to the need to fastidiously dry the solvent (methanol) prior to use. When freshly dried solvent is used, the Boc-protected amine **2.19** was obtained in 75% isolated yield. When the methanol contained trace amounts of water however, a competing hydrolysis reaction, forming significant amounts of the formamide **2.22** (Fig. 2.5), was observed by ^{1}H NMR and mass spectrometry. Compound **2.22** was not isolated however, as the two compounds **2.19** and **2.22** shared the same R_f (0.4, 9:1 DCM: MeOH) could not be separated by column chromatography, or by recrystallisation. However, a derivative of **2.22**, **2.24** was prepared and fully characterized (*Section 2.4*). **2.22** could be hydrolysed to the carboxylic acid using hot aqueous acid, which could be efficiently separated by extraction from alkaline aqueous solution to furnish pure **2.22**. Once the pure amino-ester **2.19** had been synthesized on scale, a facile LiOH-mediated alkaline hydrolysis procedure furnished the Boc-protected amino acid **2.17** in 96% isolated yield.



2.22

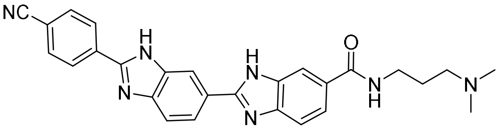
Figure 2.5: Structure of the side-reaction product **2.22.**



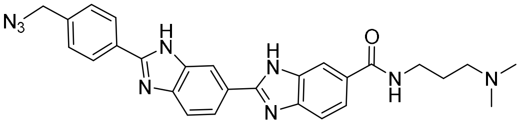
Scheme 2.5: Observed side-reaction forming the undesired formamide **2.22**. *Conditions and reagents:* (i) $NaBH_4$ (7.0 eq.), reagent grade methanol, 0°C – r.t. 16h, 75% (HPLC). **2.22** not isolated, detected by MS.

2.5. Attempts towards the synthesis of IR-labelled compounds 1.18 and 1.19.

Once the synthesis of **2.12** had been achieved the first two target compounds were identified, the nitrile **1.18** and the azide **1.19** (Figure 2.5), both derivatives of the previously described diamine **2.24** (Scheme 2.6).[104]



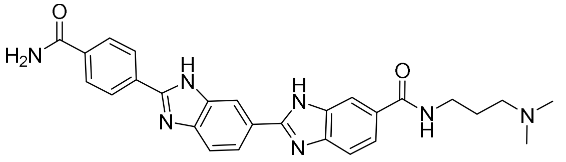
1.18



1.19

Figure 2.6: The first two target labelled compounds, **1.18** and **1.19**

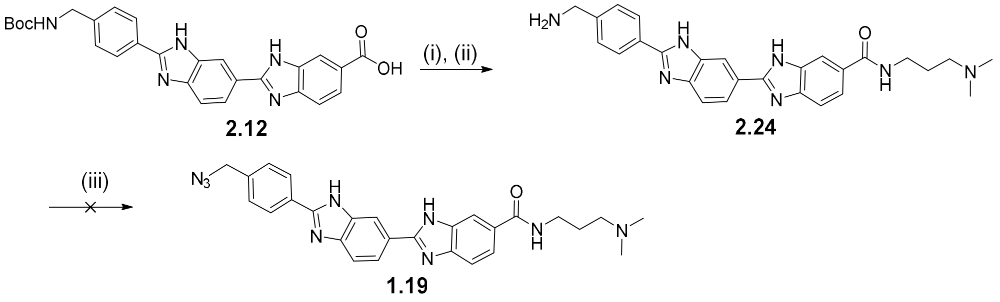
Compound **2.24** was prepared from **2.12** through saponification and then a HATU-mediated amide coupling with 3,3-dimethylaminopropylamine (DMAPA) (Scheme 2.6). Again, it was found that under our LiOH-mediated saponification conditions, base-mediated hydrolysis of the nitrile group in **2.18** to the corresponding formamide **2.22** (Scheme 2.5) was observed, and again, the two compounds could not be efficiently separated from each other. However, it was found that when the mixture was coupled to DMAPA furnishing the two compounds **1.18** and **2.23**, the two compounds were isolated by RP-HPLC (Appendix 7.8). It may have been possible, by utilising milder saponification conditions, for example KOH, to obtain the cyano-acid **2.21** in greater purity, but parallel FT-IR and pump-probe experiments on the model compound benzonitrile have shown that nitriles are quite poor 2D-IR probes (Section 2.8), and consequently and further attempts to synthesise **1.18** were abandoned in favour of synthesis of the azide **1.19**.



2.23

Figure 2.7: Structure of the formamide **2.23**

Toward the synthesis of the azide **1.19**, a Cu(II) catalysed diazotransfer protocol using imidazole-1-sulfonyl azide hydrochloride was chosen (Scheme 2.6).[106] Boc-protected amino acid **2.12** was firstly coupled to DMAPA *via* a HATU-mediated amide coupling reaction, followed by acid deprotection, furnishing the diamine product **2.24** in 62 % yield.[104]



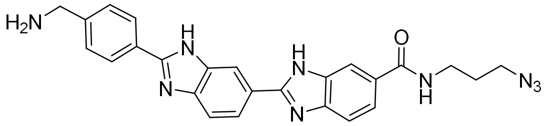
Scheme 2.6: Attempted synthesis of the azide-bearing Hoechst **1.19**. *Conditions and reagents:* (i) HATU (1.0 eq.), DIEA (6.0 eq.), DMAPA (8.0 eq.), DMF, r.t. 2 hr, not isolated. (ii) HCl (4 M in dioxane), r.t. 30 minutes, 62% (relative to **2.18**). (iii) Im-SO_2N_3.HCl (1.2 eq.), K_2CO_3 (4 eq.), $CuSO_4$ (0.1 eq.) MeOH, r.t. 16 h, no reaction.

The product **2.24** was then subjected to the diazotransfer conditions but no azide formation was detected by either FT-IR or mass spectrometry.

It was hypothesised that poor solubility of **2.24** in methanol may have caused this lack of reaction, and to prove this hypothesis, the reaction was carried out on freshly-deprotected **2.19** which was soluble in methanol. As before, no reaction was observed. It is believed that given the diazotransfer protocol has been successful on benzylamines, furnishing benzyl azide,[106] that the origin of the lack of observed reaction between Im-SO_2N_3 and **2.24** is similar to that observed when hydrogenation of **2.18** to **2.19** was observed – namely that some kind of catalyst poisoning by the benzimidazoles had occurred. This disappointing result led to alternative target compounds being sought.

2.6. Synthesis of the Hoechst azide 1.20

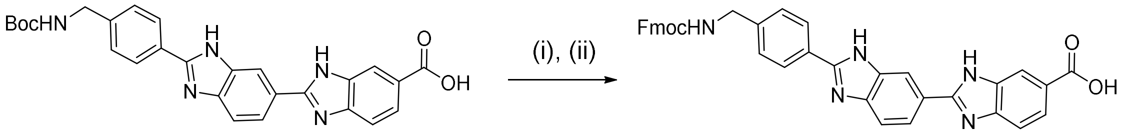
After the failure of the diazotransfer reaction, it was hypothesized that the acid **2.12** could be efficiently coupled to 3-azidopropylamine by a HATU-mediated amide coupling reaction, followed by Boc- deprotection to furnish the target compound **1.20** (Fig. 2.8).



1.20

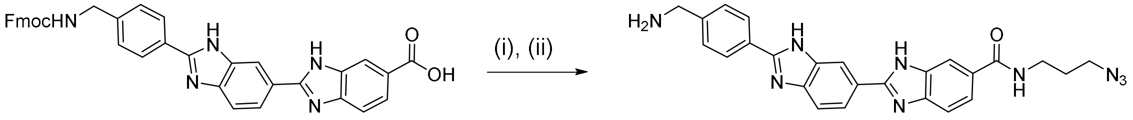
Figure 2.8: Structure of the target compound **1.20**.

The short-chain azide was prepared from the analogous bromide and sodium azide by an S_N2 reaction in quantitative yield. To protect the azide from potential acid-mediated decomposition,[107] the Fmoc-derivative of **2.12**, **2.25** was prepared by acid-deprotection of the Boc- group present in the free carboxylic acid followed by Fmoc protection using conditions first described by Carpino in 59% overall yield and in good purity after a recrystallisation from EtOH/H_2O (Scheme 2.7).[108]



Scheme 2.7: Synthesis of compound **2.25**. *Conditions and reagents:* (i): HCl (4 M in dioxane) x.s. r.t. 30 min. (ii) Fmoc-Cl (1.5 eq.), $NaHCO_3$ (sat, aq.), dioxane, r.t. 16 hrs, 59% (over 2 steps).

With **2.25** in hand, the azide **1.20** was prepared from **2.25** by a HATU mediated amide coupling to 3-azidopropionic acid in the same manner as for the synthesis of **2.24** (Scheme 2.9). After deprotection by piperidine, the desired product **1.20** was detected by MS. The purity of crude **1.20** was 72% according to analytical RP-HPLC, with the major impurities were shown by mass spectrometry to come from the Fmoc- protecting group (Appendix 7.9).



2.25 **1.20**

Scheme 2.8. Synthesis of compound **1.20**. *Conditions and reagents: (*i) HATU (1.0 eq), DIPEA (6.0 eq.) DMF r.t. 0.25 h, then 3-azdopropylamine (8.0 eq.) DMF, r.t. 2 h. (ii) piperidine (20 % v/v in water) xs r.t. 30 min, 2 %.

However, it was found that upon attempted purification by RP-HPLC, the product peak was observed to broaden significantly into multiple peaks, confirmed by successive runs of the fractions mixed with each other (Appendix 7.9). This was thought to be due to aggregation of **1.20** in solution, a finding which has some precedent in the literature for **1.9**.[109] Despite changing the HPLC conditions (changing the pH, the column temperature or using a C8 HPLC column) similar results were observed upon purification of the compound. To protect the sensitive azide group from decomposition upon lyophilisation of the fractions, the acidic TFA buffer mixture –which gave the best results from the HPLC- was neutralised with triethylamine. Although triethylammonium trifluoroacetate is non-volatile under these conditions, it was thought that buffer exchange through a short plug of C18 RP-silica would efficiently remove the triethylammonium trifluoroacetate salts from purified **1.20**. Unfortunately, **1.20** was observed to streak down the reversed-phase column, and multiple rounds of chromatographic purification were required to remove the trifluoroacetate salts. As a consequence of this difficult and inefficient purification procedure, the yield obtained of the purified product was 2%. Moreover, a strong peak in the FT-IR at 1675cm^{-1} associated with the presence of trifluoroacetate counterions associated with the primary amine in **1.20** (Fig.

2.9) led to the abandonment of this method as a means to obtain pure **1.20**, as this would interfere with the ability to interpret the FT-IR spectrum of **1.20** in complex with dsDNA (Section 3.4). Hence, alternative purification conditions without the use of chromatography were sought.

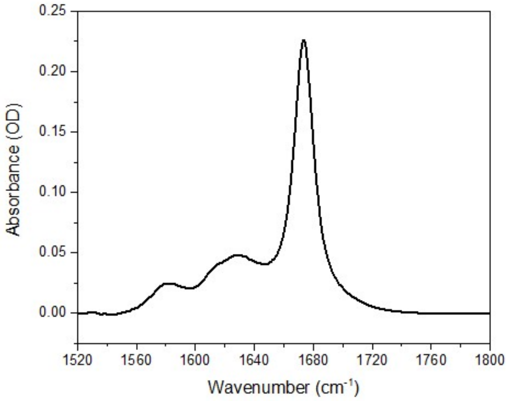


Figure 2.9. FT-IR spectrum of HPLC-purified **1.20** at 20mM in DMSO. The peak at 1675cm^{-1} is the stretching vibration of the carbonyl of TFA.

Because of the failures of HPLC as a means to purify **1.20**, recrystallisation as a purification method was investigated. Unfortunately, it proved impossible to separate pure **1.20** from the by-products from the Fmoc-deprotection step. From this result, it was decided to return to the Boc- protected amino acid **2.12,** of which the deprotection by-products are all volatile compounds. To ascertain whether the azide group, as previously thought, was prone to decomposition in the presence of strong acids, crude **1.20** was stirred in 4 M solution of HCl in dioxane overnight (Scheme 2.9). The azide IR signal showed no signs of decomposition over this period (Fig. 2.10), and this allowed for the purification of **1.20** by both selective precipitation and recrystallisation of the hydrochloride salt (Scheme 2.9). Using the Boc-deprotection method, 1.34 g of Hoechst azide **1.20** was prepared in 93% yield directly from **2.12**, and could be efficiently purified by recrystallisation from methanol.

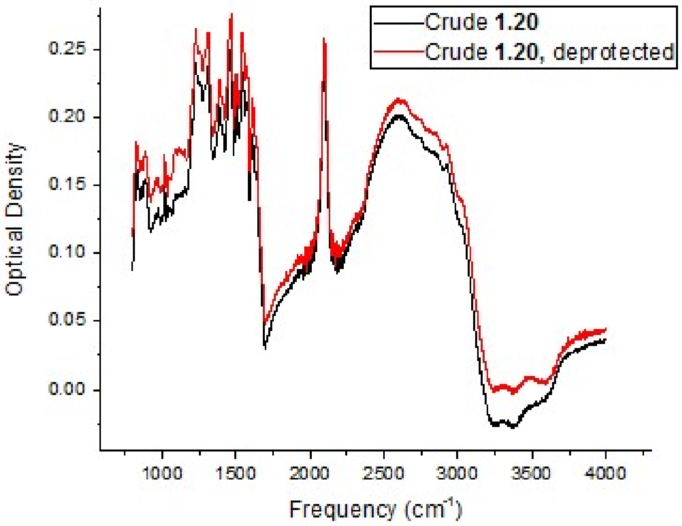
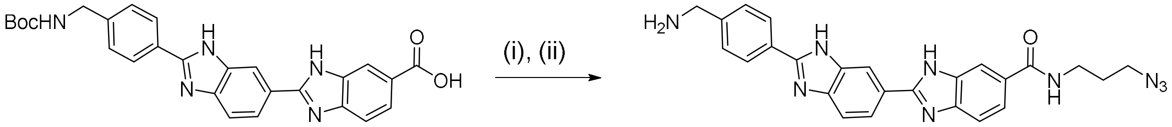


Figure 2.10. ATR FT-IR spectrum (atmospheric water subtracted) of crude **1.20** before (black) and after (red) 16 h of stirring in 4M HCl dioxane solution.



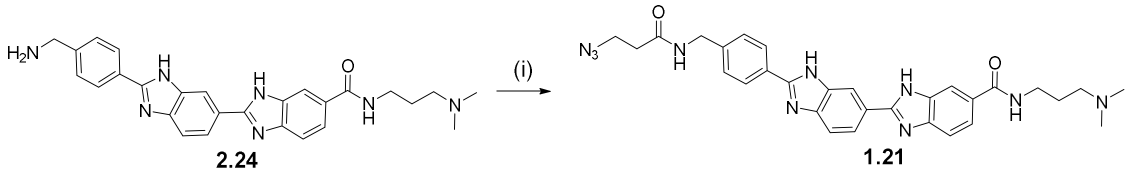
2.12 **1.20**

Scheme 2.9. Synthesis of compound **1.20**. *Conditions and reagents:* (i) HATU (1.0 eq.) DIPEA (6.0 eq.) DMF, 0.25 h, r.t. then 3-azidopropylamine (8.0 eq.), DMF, r.t. 2 h. (ii) HCl (4 M in dioxane), r.t. 0.5 h, 93% (over 2 steps)

2.7. Synthesis of the IR-labelled derivative of Hoechst 33258, 1.21.

*2.7.1. HATU-mediated amide coupling of **2.24** to **1.21***

With **2.24** in hand, attention turned to synthesis of the Hoechst azide **1.21** (Scheme 2.10). The synthetic approach firstly taken was, similarly to that pursued for **1.20**, a HATU-mediated coupling between the diamine **2.24** and 3-azidopropionic acid (Scheme 2.11)

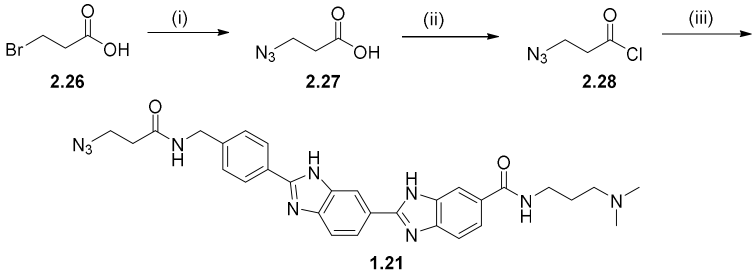


2.24 **1.21**

Scheme 2.10: Synthesis of the Hoechst azide **1.21**. *Conditions and reagents:* (i) 3-azidopropionic acid (3.0 eq.), HATU (2.5 eq.), DIEA (6.0 eq.) DMF r.t. 0.5 h, then **2.24** (1.0 eq.), DIPEA (6.0 eq.) DMF, r.t. 16 h, 2 % overall.

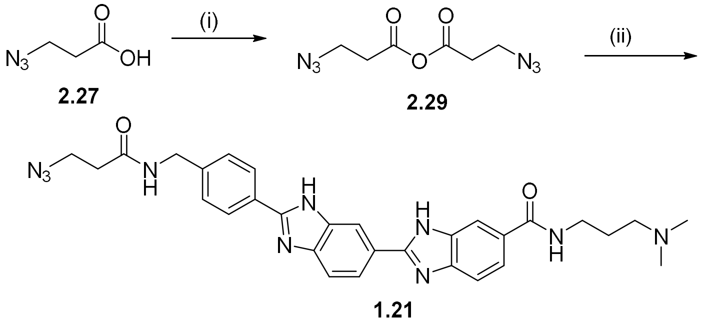
Similar problems of aggregation of **1.21** during HPLC purification led to poor product recovery leading to an isolated yield again of 2%, despite the crude purity of **1.21** being 79% and quantitative conversion of **2.24** was observed (Appendix 7.10). Because of the inherent problems of HPLC purification of **1.21**, a different method of purification of the hydrochloride salt of **1.21,** similarly to **1.20** was sought. Given that the reaction depicted in Scheme 2.10 went to completion, the reaction was repeated, and then recrystallisation conditions were sought, without success. The problem seemed to be an inability to remove the tetramethylurea by-product from the reaction mixture according to elemental analysis and ^{1}H NMR (Appendix 7.12). The similar solubility of tetramethyluronium chloride to **1.21** in a wide range of solvents frustrated any efforts to isolate **1.21** by recrystallisation.

To attempt to solve this problem, 3-azidopropionic chloride **2.28** (Scheme 2.11) was prepared, and then reacted with **2.24** in dry THF in the presence of triethylamine to act as an acid scavenger.[110] This reaction also went to completion by crude ^{1}H-NMR (Appendix 7.11), and a short column of C18 RP silica was used to separate the triethylammonium hydrochloride salt from **1.21**. However, as with **1.20** extensive streaking through the silica was observed, with some **1.21** eluting with the $Et_3N.HCl$ with the solvent front, some **1.21** eluting as expected, and some not eluting at all, even when pure acetonitrile was used as the elution buffer. Changing the base from triethylamine to pyridine was tried as the hydrochloride salt of pyridine was soluble in ethanol, whereas the hydrochloride salt of **1.21** is only poorly soluble. Hence, if **1.21** could be synthesized using pyridine as a base, crude **1.21** could be purified by recrystallisation from ethanol. However, very poor conversion of **2.24** to **1.21** was observed by mass spectrometry. This is due to the differing pK_b values of pyridine (5.25)[111] compared to benzylamine (4.66),[112] the parent structure of **2.24**. Consequently, the HCl generated by the reaction protonated the primary amine forming the hydrochloride salt, and hence pyridine was not found to be suitable as an acid scavenger for the synthesis of **1.21**. Because of these disappointing results, attempts towards the synthesis of **1.21** *via* an acid chloride method were abandoned in favour of another method.



Scheme 2.11. Synthesis of **1.21** from 3-azidopropionyl chloride. *Conditions and reagents:* (i) NaN_3 (1.5 eq.) H_2O, reflux, 16 h, quant. (ii) Oxalyl Chloride (1.2 eq.), THF, r.t. 4 h, not isolated. (iii) **2.31** (1.0 eq.), **2.28** (0.51 M in THF, 10.0 eq.) THF, Et_3N, r.t. 16 h, 12%.

Amide bonds can, in addition to activated triazole esters and acid chlorides, also be prepared from amines and acid anhydrides. It was found that the anhydride of **2.27**, compound **2.29** (Scheme 2.12) could be efficiently prepared using DCC in dry DCM.[113]



Scheme 2.12. Attempted preparation of **1.21** from **2.24**. *Conditions and reagents:* (i) **2.27** (1.0 eq.), DCC (0.45 eq.) DCM, r.t. 4 h, **2.34** not isolated. (ii) **2.29** (0.5 M in DCM, 10.0 eq.) **2.24** (1.0 eq.), full conversion observed by HPLC, MS, 5 % overall yield upon purification.

To prevent the presence of free DCC in the reaction mixture, thus leading to the guanidinylation of **2.24**, DCC was kept as the limiting reagent. Although this reaction worked, LC-MS analysis of the crude mixture indicated the presence of DCU in the mixture, despite the precipitated DCU being filtered prior to addition of the crude anhydride to a solution of **2.24.** As such, due to the aforementioned problems with chromatographic purification of **1.21**, again only a 5 % yield was observed.

A final route tried towards **1.21** was to use a catalytic amount of boric acid as an amide coupling reagent, a reaction first reported by Tang in 2005.[114] However, it was found that the starting material **2.24** was insoluble in the solvent (toluene), and consequently no reaction was observed after 16 h under Dean-Stark conditions, and the starting material was recovered quantitatively. Consequently, to date, no satisfactory synthetic pathway to **1.21** has been developed, although small amounts have been prepared, and the behaviour of **1.21** in complex with dsDNA is reported in Chapter 3.

2.8. IR Characterisation of the azide band of compounds 1.20 and 1.21.

The majority of the infrared spectroscopic experiments discussed hereafter in this section form part of a collaborative effort that links the work in this thesis with the PhD work of Lennart Ramakers of the Department of Physics at the University of Strathclyde. Where the data presented is a result of this collaboration it is clearly referenced.

The azide region of the IR spectrum of free **1.20** and **1.21** at 20 mM in ethanol (**1.20**) and phosphate buffer (**1.21**) was investigated in order to understand the differences between the azide signals of the short-chain azide-bearing precursors and the full compounds (Fig. 2.11). The positions of the two azide peaks were very similar to the short-chain precursors – indicating that the presence of the Hoechst backbone has little effect on the position of the azide.

Interestingly, the two azide peak shapes were noticeably different; the peak of **1.20** exhibited a simple Gaussian shape, whereas the peak of **1.21** exhibited a more complex trapezoid shape. Very similar peak shapes were observed for both **1.20** and **1.21** and the azide-bearing precursors 3-azidopropylamine and **2.27**. This is further evidence that the azide functional group is sufficiently remote from the other functional groups in the molecule such that the presence of the Hoechst backbone does not significantly affect the azide signal. The reason for the change in peak shape is currently unclear; there is presumably some interaction between the azide and the carbonyl of the carboxylic acid/amide. Interestingly, **1.20**, despite possessing a carbonyl group, shows no analogous interaction. Future experiments will investigate the origin of the change in the shape of the azide peaks in 3-azidopropionic acid and 3-azidopropylamine.

No distinct shoulder peaks were observed. This lack of shoulder peaks, which were observed in benzyl azide (Fig. 2.11), indicated a lack of Fermi resonances in the azide IR spectrum,[115] significantly simplifying analysis of the complexes. Furthermore, changes in the shape of the complex should come entirely from the azide interacting with the dsDNA.

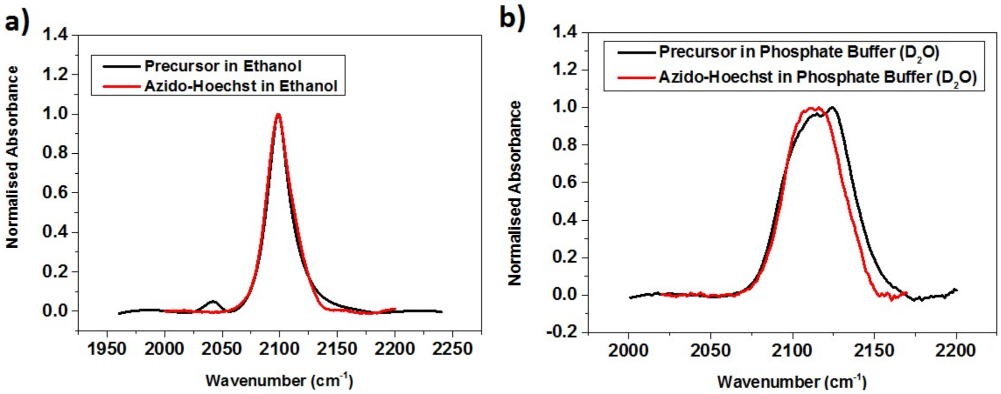


Figure 2.11. **(a)** The azide region of the IR of **1.20**, compared to 3-azidopropylamine **(b)** The azide region of **1.21** compared to **2.27**.[118]

When the azide absorption bands for compounds **1.20** and **1.21** were compared to benzyl azide (a model compound for the original target **1.19**, which contains a pronounced shoulder on the blue side of the peak (Fig. 2.12). This peak was identified as the first overtone of the in-plane aromatic C-H bending vibration, which due to quantum mechanical mixing with the azide stretching vibration (a Fermi resonance) its absorbance is enhanced.[115,116] The presence of this additional shoulder peak firstly complicates interpretation of the spectrum, as two vibrations contribute to this peak, and secondly, the observed extinction coefficient is diminished by the presence of the Fermi resonance. These observations, as well as the synthetic intractability of **1.19** have vindicated the choice to pursue **1.20** and **1.21** as the compounds for this study.

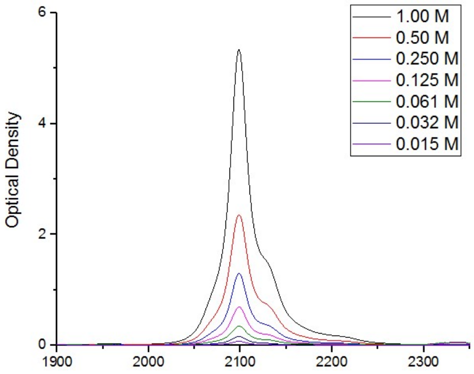


Figure 2.12. IR spectrum of the azide stretching vibration of benzyl azide at concentrations from 1 M to 0.015 M in IPA.

Nonetheless, the extinction coefficient for benzyl azide was found to be 1044 $M^{-1}cm^{-1}$and hence, data shows that azides are visible in protic solvents at a concentration of 10-15 mM, and these concentrations are required to obtain a sufficiently strong azide IR spectrum to carry out the DNA melting experiments.

*2.8.1: Comparison of the azide and nitrile peaks in analogues of **1. 9.***.

Although a satisfactory synthetic route to **1.18** has to date not been found, nitrile IR vibration of the model compound benzonitrile was compared to the short-chain precursor azides, 3-

azidopropylamine and 3-azidopropanoic acid. The short-chain azides serve as useful approximations of the azide signals of the **1.20** and **1.21**, and hence studying the azide signals of these two compounds should help to understand the behaviour of the peak in complex with dsDNA. Secondly, a comparison of the azide and nitrile peaks will serve to understand the different responses of the azide and nitrile to changing solvent conditions. Thirdly, the model compounds are soluble in a greater number of solvents than either **1.20** or **1.21**, leading to a fuller understanding of the azide response in a variety of solvent environments.

A comparison of the azide peaks of 50 mM 3-azidopropylamine and 3-azidopropionic acid in isopropyl alcohol with benzonitrile (Fig. 2.13) showed that molar extinction coefficients of the azide peak (1050 $M^{-1}cm^{-1}$) was five times larger than the nitrile (200 $M^{-1}cm^{-1}$). The nitrile peak reaches a maximum at 2250 cm^{-1}, a difference of 150 cm^{-1} from the azides which come at 2098 for **1.20** and 2104 cm^{-1} for **1.21** respectively. Hence, azides are better suited as vibrational probes than nitriles. Furthermore, no prominent Fermi resonances are visible in the two short-chain azides when compared to benzyl azide, the peak broadening is thought to be an interaction between the azide and the solvent, or the amide group.

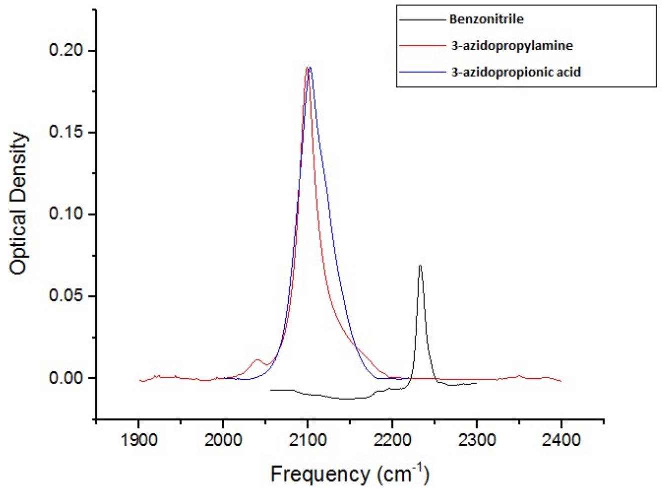


Figure 2.13. Comparison of the azide and nitrile signals of benzonitrile, 3-azidopropionic acid and 3-azidopropylamine at 50 mM in isopropyl alcohol.

The five-fold drop in sensitivity between the azide and nitrile response explains the reason why pursuit of the synthesis of **1.18** was abandoned in favour of **1.20** and **1.21,** as the concentration of complex required to observe a satisfactory signal-to-noise ratio for the nitrile response is prohibitively high. The azide responses of two short-chain azide compounds were also measured in a variety of solvents in order to investigate the changes in the azide region of the spectrum (Fig. 2.14). Protic solvents were chosen, as although the minor groove of dsDNA is less polar than bulk water, the presence of the spine of hydration means that the overall solvent environment of the minor groove is polar and protic.

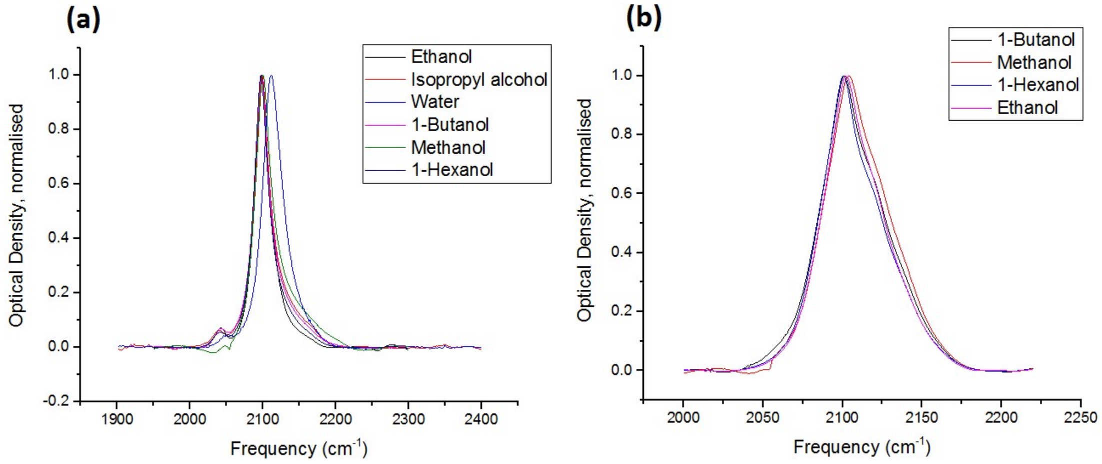


Figure 2.14. (a) FT-IR spectrum of 3-azidopropylamine in various solvents. **(b)** FT-IR spectrum of 3-azidopropionic acid in various protic solvents. Water has been deliberately omitted.[118]

The peak position changed to higher frequency with increasing solvent polarity, indicating that azides are sensitive to changes in solvent polarity. This observation indicates that merely a change in solvent polarity would be experienced by the azide, leading to a detectable change in the signal. Furthermore, as the azide functional group is capable of hydrogen bonding, if a specific hydrogen bond between the edge of a base and the azide of **1.20** or **1.21** is formed, this should be detectable in its IR response.

The origins of the changes in the spectrum were also investigated by plotting the azide peak position against the various Hansen solubility parameters.[117] It was found that the origin of the change in the azide peak is due to the change in the hydrogen bonding parameter (Fig. 2.15). A good shallow exponential fit for the change in the hydrogen bonding parameter with peak position was found When progressively more water was added to a methanolic solution of 3-azidopropionic acid, the shape of the azide IR absorption band changed from a Gaussian peak resembling the azide IR absorption band of 3-azidopropylamine to resemble the trapezoid shape observed for **1.21** (Fig. 2.16).

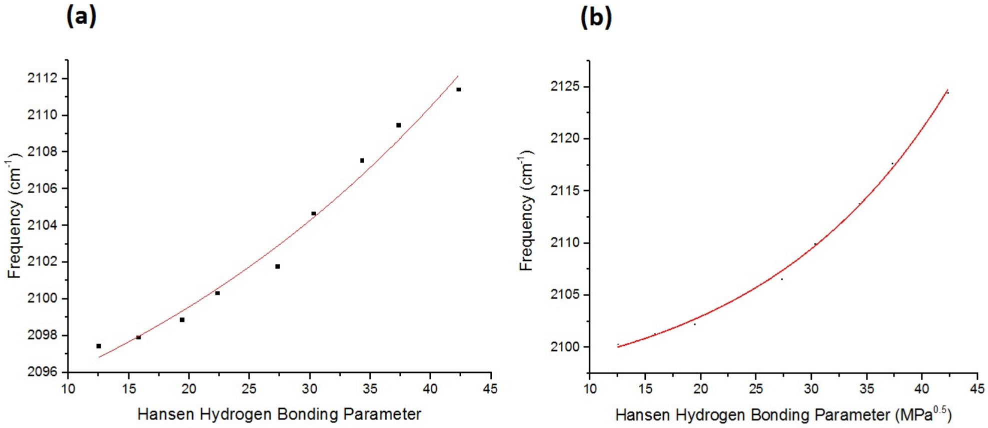


Figure 2.15. **(a)** Change in the azide spectrum of 3-azidopropylamine with the Hansen hydrogen bonding parameter. $R^2 = 0.997$. **(b)** Change in the peak position observed for 3-azidopropionic acid. $R^2 = 0.998$.

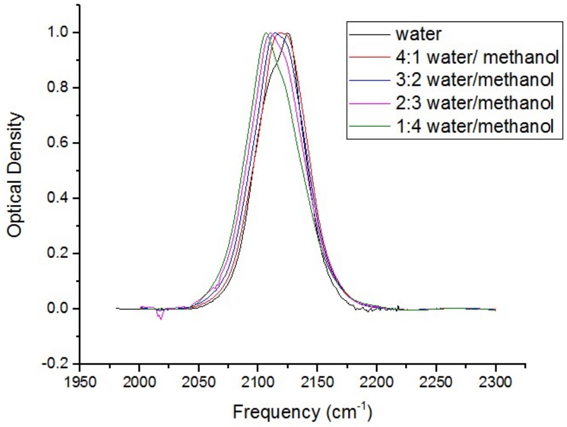


Figure 2.16. FT-IR spectra of 3-azidopropionic acid in blends of methanol/water.[118] The trapezoid shape is visible in a 4:1 volumetric ratio of water/methanol.

The trapezoid peak shape observed for **1.21** is present at the 4:1 volumetric ratio (red line, Fig. 2.16). In pure water, an "inversion" of the peak shape observed in Fig. 2.11 (b) is observed, and is speculatively due to an interaction between water and the azide that is not present with organic solvent. The reasons for this are currently unclear.[118]

*2.8.2: Pump-probe IR spectrum of **1.20.***

Finally, to find the vibrational excitation lifetime of **1.20** in solution, pump-probe spectra of a 20 mM solution of **1.20** in a 4:1 volumetric mixture of DMSO and water were carried out (Fig. 2.17).[118] Diagonal peaks corresponding to the azide of **1.20** in an excited state were found after a waiting time of 1 ps. This showed that azides can be used as vibrational probes with sub-picosecond time resolution. Hence, **1.20** is suitable for use as a probe for our investigations of the dynamics of dissociation of this compound from dsDNA on a picosecond timescale.

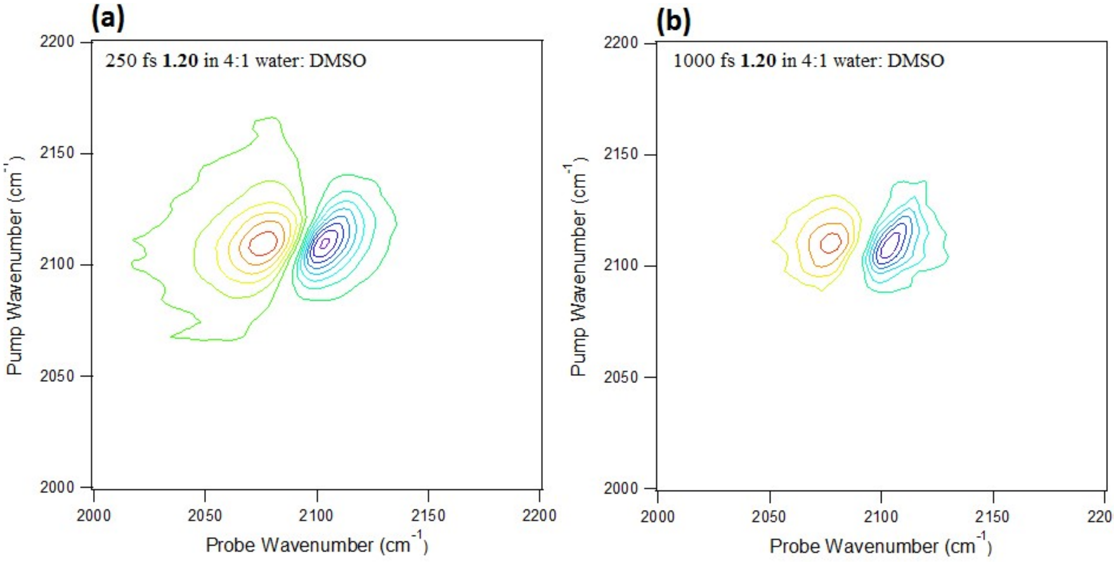


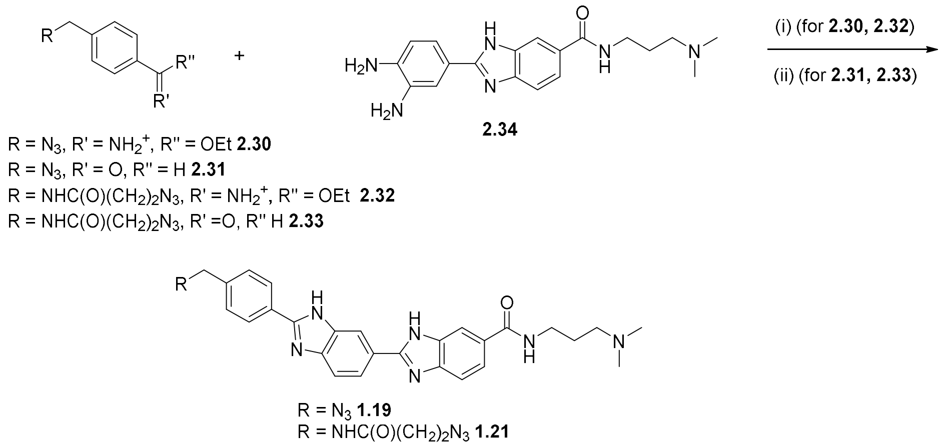
Figure 2.17. **(a)** Pump-probe spectrum of **1.20** in 4:1 (v/v) DMSO: water with 250 fs waiting time. **(b)** Pump-probe spectrum of the same sample with a waiting time of 1000 fs.[118]

2.9. Conclusions and Future work.

Two azide labelled derivatives of **1.9**, have been synthesised, compounds **1.20** and **1.21** Compound **1.20** could be synthesised directly from **2.12** on a 1.5 g scale in 93% yield, and in sufficient purity for spectroscopic experiments with no recourse to chromatography necessary. No analogous procedure was found for **1.21**, and hence, the yields of **1.21** obtained were much lower (12%).

The IR spectrum of the azide signals of **1.20** and **1.21** were evaluated and compared to the precursors, 3-azidopropionic acid and 3-azidopropylamine. No significant change in either absorbance or peak shape was observed between the precursors and full **1.20** and **1.21** – hence any changes in the azide IR spectrum seen for **1.20** and **1.21** are due to interaction of the ligand with dsDNA, and not due to vibrational coupling with the ligand. Hydrogen bonding is responsible for changes in the azide signal in response to changes in the solvent properties, and hence azides are suitable functional groups as IR probes. 2D-IR experiments on compound **1.20** have shown that the azide has a vibrational lifetime of approximately 2 ps, and hence sub-picosecond time resolution through 2D-IR is possible.

Future work will focus on developing a robust synthesis of the MGB **1.21**. From the methods employed through the course of this project, a HATU-mediated coupling appeared to work more reliably compared to the other methods used, and the main impurity observed was the tetramethyluronium by-product. If recrystallisation conditions could be found such that the free amine of **1.21** could be efficiently separated from the by-product, washing and then protonation should allow access to gram-scale synthesis of **1.21**, similarly to **1.20**. Also, further investigations into the unusual peak shape exhibited by **1.21** are necessary to understand the origins of the peak shape. Further investigation of derivatives of 3-azidopropionic acid will be necessary.



Scheme 2.13. Proposed future routes to **1.19** and **1.21.** *Reagents and conditions:* (i) **2.30** or **2.32**, **2.34** Ac$_2$O, 55°C 16 h,[103] or (ii) **2.31**, or **2.33** and **2.39**, PhNO$_2$ 150 °C, 16 h.[119]

Alternatively, synthesis of **1.19** and **1.21** could possibly be achieved by either Pinner cyclisation using the iminium salts **2.30** and **2.37** or direct condensation of aldehydes **2.31** and **2.33** with the diamine **2.34** akin to the methods described in Section 2.1.[103,119]

This method for preparation of **1.21** was not investigated through the course of this project as firstly it was thought that the high temperatures, or acidic conditions necessary to affect these transformations would cause either thermal[120] or acid decomposition[121] of the sensitive azide group, and secondly although the purification of **1.21** proved difficult and inefficient, the compounds could be efficiently prepared. However, these methods have the distinct advantage of being two component mixtures of which one (**2.30** – **2.34**) is inherently hydrophobic (the Pinner salts upon hydrolysis, forming an ester) thus greatly simplifying purification, provided full conversion of **2.34** occurs. Moreover, by this proposed method, **2.18** can, in theory be efficiently prepared through condensation of **2.14** with **2.34**, allowing for the direct comparison of the azide and nitrile signals of IR-labelled small molecules in complex with a dsDNA oligomer.

2.10. Experimental

2.10.1. General experimental techniques and procedures.

All reagents and solvents were purchased from commercial sources and, unless stated, were used without any further purification. Where indicated, dry solvents were prepared through distillation over 3 Å molecular sieves, or dried through a column of basic alumina.

2.10.1.1. NMR spectroscopy

NMR spectra recorded in deuterated solvents at 400, 500 or 600 MHz for ^{1}H NMR and 100, 125 and 150 MHz for ^{13}C NMR on Bruker AV3 spectrometers. Spectra were analysed using the Bruker TOPSPIN with chemical shifts quoted in ppm relative to the residual solvent signal as indicated in the text. Where necessary, full assignment of the one-dimensional spectra was carried out using COSY, HMBC, HSQC and NOESY experiments on the aforementioned spectrometers.

2.10.1.2. Mass spectrometry

Electrospray ionisation (ESI) mass spectra were acquired on a Finnigan LCQ-DUO mass spectrometer. Matrix-assisted laser desorption ionisation - Time of flight (MALDI-TOF) were acquired on a Shimadzu KRATOS mass spectrometer

2.10.1.3. Infrared Spectroscopy

Attenuated total reflection (ATR) FT-IR spectra were collected on a Shimadzu FT-IR spectrophotometer outfitted with an ATR attachment and processed using Shimadzu IRSolution software. Only peaks exhibiting strong absorbance, or relevant to particular functional groups are reported.

Solution-phase FT-IR spectra were recorded on a Bruker Vertex 80 spectrometer using Bruker OPUS 6.0 software and analysed using the OriginPro 2015 software. 50 μL of given sample was encased between two CaF_2 plates fitted a 50 μm polyethylene spacer.

2.10.1.4. UV-Vis spectroscopy

UV-Vis spectra were collected on either a Shimadzu UV-1800 spectrophotometer or a Thermo Scientific Nanodrop spectrophotometer, Spectra were processed using OriginPro 2015 software.

2.10.1.5. TLC experiments

Reported R_f values were obtained using aluminium-backed Merck Millipore 60G F_{254} silica gel of layer thickness 200 µm. Identification of the spots was identified by UV irradiation at either 254 or 350 nm wavelength, and stained with an appropriate staining mixture.

2.10.1.6. HPLC experiments

Analytical HPLC experiments were performed on a Dionex UltiMATE 3000 system using a C18 Phenomenex Aeris 3.6 µm wide pore column with a length of 250 mm and internal diameter of 4.6 mm. Either water (Buffer A) /acetonitrile (Buffer B) or 0.1% v/v water (Buffer C)/ 0.1% TFA v/v acetonitrile (Buffer D) were used as buffer systems, are detailed within the characterisation data. All retention times are reported using a solvent gradient of 5 - 90% buffer B in buffer A or buffer D in buffer C, over a run time of 18 minutes.

Semipreparative HPLC was performed on a Dionex UltiMATE 3000 system using either a Dionex Kinetex 5 µm C18 column of length 150 mm and internal diameter 21.2 mm (flow rate 9.0 mL/min) or an ACE 5 µm column of length 250 mm with internal diameter 10 mm (flow rate 4.0 mL/min). Elution buffers used were identical to those described for the analytical HPLC experiments.

2.10.1.7. Melting point determination

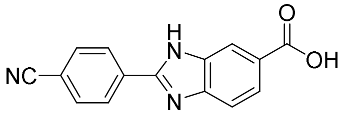
Melting points for all samples were collected using a Griffin melting point apparatus and are reported without correction. Melting points for compounds containing azides were not collected due to risk of explosion.

2.10.1.8: Elemental Analysis

Elemental analysis was carried out by the University of Strathclyde Microanalysis service, using a PerkinElmer elemental analyser.

2.10.2. Synthetic procedures

2-(4-cyanophenyl)-1H-benzo[d]imidazole-6-carboxylic acid (**2.15**) [104]



2.15

To a stirred solution of **2.13** (4.50 g, 39 mmol, 1.0 eq.) in nitrobenzene (120 mL) was added **2.14** (5.21 g, 39 mmol, 1.0 eq.), and the mixture was heated to 150 °C. After 1 h, upon observation of a turbid solution, 25 mL of DMF was added to redissolve the precipitate. The mixture was then stirred at 150 °C for a further 6 hours until the reaction was complete, as observed by TLC. The crude reaction mixture was then allowed to cool to room temperature and then added to 500 mL cold diethyl ether, at which point an off-white precipitate was observed. The precipitate was collected by suction filtration, and the filter cake washed five times with 100 mL of hexane, until a fine off-white powder was obtained and no yellow colour was observed in the filtrate. The crude product thus obtained was then recrystallised from ethanol/water to afford **2.15** (7.18 g, 27.3 mmol, 70%) as an off-white solid.

^{1}H NMR (400 MHz, DMSO, δ= 2.50 ppm): 13.46 (1H, bs), 12.83 (1H, bs) 8.35 (2H, d, *J* = 7.2 Hz)), 8.22 (1H, d, *J* = 7.2 Hz), 8.05 (1H, d, *J* = 8.0 Hz), 7.87 (1H, d, *J* = 8.4 Hz), 7.70 (1H, d, *J* = 8.4 Hz).

^{13}C NMR (100 MHz, DMSO, δ = 40.0 ppm): 168.2, 135.7, 134.2, 133.5, 130.4, 127.8, 123.8, 119.0, 113.0.

FT-IR (ATR): 2980, 2228, 1670, 1608, 1425, 1325, 1303 cm^{-1}

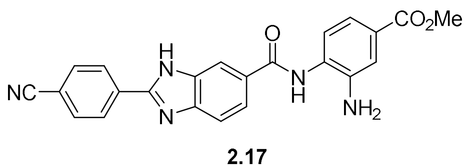
ESI-MS[M+H]$^+$: 264 m/z.

Melting point: > 300 °C.

R_f: 0.3 (7:2:1 DCM:MeOH:Et_3N)

Elemental Analysis: Calc.: 68.44% H 3.45% N 15.96%. Found. C 68.10% H 3.50% N 15.75%

Methyl 3-amino-4-(2-(4-cyanophenyl)-1H-benzo[d]imidazole-6-carboxamido)benzoate (**2.17**)[104]



2.17

To a stirred solution of compound **2.15** (7.0 g, 27 mmol, 1.0 eq.) in peptide grade DMF (120 mL) was added HATU (10.26 g, 27 mmol, 1.0 eq.) and DIEA (10.0 mL, 8.66 g, 67.5 mmol, 2.5 eq.). The reaction mixture was stirred at r.t. for 10 minutes before **2.16** (4.9 g, 27 mmol, 1.0 eq.) was added. The mixture was then stirred at r.t for 24 h. The reaction mixture was then added to 500 mL of cold water precipitating the crude product as a brown solid. The precipitate was collected by suction filtration and washed five times with 100 mL of water until no yellow colour was observed in the filtrate. The wet brown solid thus obtained was then suspended in 25 mL water and lyophilised to dryness to afford 2.17 (9.2 g, 22.3 mmol, 81%) as finely divided light brown solid. A mixture of two regioisomers was obtained, however, no attempts were undertaken to isolate the regioisomers.

^{1}H NMR (500MHz, DMSO, δ=2.50 ppm): 13.49 (1H, bs), 9.75 (1H, bs) 8.38 (2H, d, *J* = 8.2 Hz), 8.06 (d, 2H, *J* = 8.0 Hz), 8.01 (1H, s), 7.91 (1H, d, *J* = 8.4 Hz), 7.84 (1H, s), 7.78 (1H, d, *J* = 8.0 Hz), 7.58 (d, 1H, *J* = 7.6 Hz), 6.80 (1H, d, *J* = 7.6 Hz), 5.86 (2H, s), 3.83 (3H,s).

^{13}C NMR (125MHz, DMSO δ = 40.0ppm): 166.5, 148.9, 141.0, 136.3, 134.3, 133.5, 133.3, 133.2, 130.3, 129.4, 128.7, 127.7, 127.5, 122.5, 119.3, 118.9, 116.8, 115.1, 113.2, 112.8, 111.8, 52.0.

FT-IR (ATR): 3399, 2226, 1768, 1716, 1433 cm^{-1}.

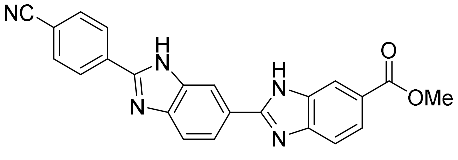
ESI-MS $[M+H]^+$:412 m/z.

Melting Point: 215 – 218 °C.

$R_{f:}$ 0.4 (9:1 DCM: MeOH)

Elemental Analysis: Calc.: C 67.15%, H 4.17% N 17.02%. Found: C 66.60% H 4.29% N 16.55%.

Methyl 2'-(4-cyanophenyl)-1H,3'H-[2,5'-bibenzo[d]imidazole]-6-carboxylate (**2.18)**[104]



2.18

To a stirred suspension of compound **2.17** (9.0 g 22 mmol, 1.0 eq.) in 125 mL nitrobenzene was added 2 drops of concentrated H_2SO_4 and then heated to 190 °C for 6 h at which point TLC analysis ($R_f = 0.45$, 9:1 DCM/MeOH) showed complete conversion of **2.17**. The crude product was then separated from the bulk solvent through addition of petroleum ether, furnishing an oily brown crude product. Further triturations of petroleum ether precipitated crude **2.18** as a fine brown powder, which was then collected by vacuum filtration. The filter cake was then washed 5 times with 100 mL of hexane upon which no yellow colour was observed in the filtrate. Crude **2.18** was then purified by recrystallisation from a mixture of hot DMF (125 mL), saturated Na_2CO_3 (50 mL), and water (150 mL). The precipitate which formed upon cooling of the mixture was then collected by suction filtration, and the filter cake washed three times with water. The wet solid obtained was then suspended in 25 mL water and then lyophilised to afford pure **2.18** (6.50g, 77%) as a finely divided brown powder.

^{1}H NMR (500MHz, DMSO $\delta = 2.50$ ppm): 8.44 (1H, s), 8.38 (2H, d, $J = 8.0$ Hz), 8.17 (1H, s), 8.13 (1H, d, $J = 8.2$ Hz), 8.02 (2H, d, $J = 8.0$ Hz), 7.82 (1H, d, $J = 8.1$ Hz), 7.79 (1H, d, 8.2 Hz) 7.66 (1H, s, $J = 8.2$ Hz), 3.88 (3H, s)

^{13}C NMR (125MHz, DMSO δ = 40.0 ppm): 166.2, 153.2, 153.0, 132.8, 132.3, 130.0, 129.9, 128.2, 127.0, 126.8, 126.1, 124.3, 122.3, 116.0, 114.2, 51.8.

FT-IR (ATR): 2226, 1697, 1541, 1431, 1302 cm^{-1}

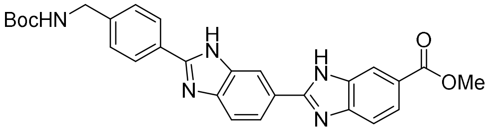
ESI-MS [M+H]: 394 m/z

Melting Point: 250 – 255 °C

Elemental Analysis: Calc.: C 70.22% H 3.84% N 17.80%. Found: C 70.55% H 3.90% N 17.99%

R$_f$: 0.45 (9:1 DCM: MeOH)

Methyl 2'-(4-(((tert-butoxycarbonyl)amino)methyl)phenyl)-1H,3'H-[2,5'-bibenzo[d]imidazole]-6-carboxylate **(2.19)**[105]



2.19

To a stirred suspension of compound **2.18** (6.25 g, 16.5 mmol, 1.0 eq.) in 250 mL dry methanol under a stream of argon was added $NiCl_2.6H_2O$ (0.41 g, 1.65 mmol, 0.1 eq), and di*tert*-butyl pyrocarbonate (7.65 mL,7.24 g, 33 mmol, 2.0 eq.), and the mixture was cooled to 0 °C. Solid $NaBH_4$ (4.14 g, 7.0 eq.) was added to the mixture in portions. The reaction was both exothermic and evolved copious amounts of hydrogen. Upon addition of the final portion of $NaBH_4$, the mixture was then allowed to slowly warm to room temperature. After 6 h at room temperature, TLC analysis (R_f = 0.4, 9:1 DCM: MeOH) showed complete conversion to the product, which gave a positive ninhydrin test. Diethylenetriamine (1.80 mL, 1.70 g, 16.5 mmol, 1.0 eq.) was then added to the reaction mixture and allowed to stir for a further 0.5 h. The crude mixture was then diluted with 200 mL ethanol and directly filtered through a pad of Celite. The mixture was then concentrated *in vacuo* and the residue redissolved in 100mL n-butanol. The mixture was then washed twice with saturated aqueous $NaHCO_3$ (2 × 100 mL), once with 100 mL brine and 3 × 100 mL water, dried over Na_2SO_4 and concentrated *in vacuo*. The residue was purified *via* recrystallisation from ethanol/water to afford **2.19** (6.10 g, 12.2 mmol, 74%) as an off-white solid**.**

^{1}H NMR (400MHz, DMSO δ= 2.50 ppm): 8.46 (1H, s), 8.41 (1H, s), 8.39 (1H, d, *J* = 8.2 Hz), 8.19 (2H, s), 8.16 (1H, s), 8.11 (1H, d, *J* = 8.2 Hz), 8.02 (1H, d, *J* = 7.4 Hz), 7.83 (1H, d, *J* = 7.4 Hz), 7.78 (1H, d, *J* = 7.4 Hz), 7.76 (1H, d, *J* = 8.2 Hz), 7.43 (2H, d, *J* = 7.8 Hz), 4.22 (2H, d, *J* = 4.6 Hz), 3.88 (3H, s), 1.41 (9H, s).

^{13}C NMR (100MHz, DMSO δ= 40.0 ppm): 167.9, 167.2, 156.6, 155.5, 153.7, 135.1, 133.4, 129.4, 128.7, 128.0, 127.7, 127.1, 124.7, 124.1, 123.6, 121.8, 119.1, 78.5, 52.7, 28.8

FT-IR (ATR): 1747, 1713, 1455, 1300, 1222, 1190 cm^{-1}.

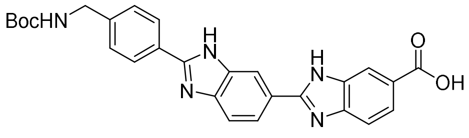
ESI-MS [M+H]: 498 m/z.

Melting Point: 195 – 200 °C (decomposes)

R_f = 0.4 (9:1 DCM: MeOH)

Elemental Analysis: Calc.: C 67.59% H 5.47% N 14.08%. Found: C 67.31% H 5.52% N 14.00%

2'-(4-(((tert-butoxycarbonyl)amino)methyl)phenyl)-1H,3'H-[2,5'-bibenzo[d]imidazole]-6-carboxylic acid **(2.12)**[104]



2.12

To a solution of compound **2.19** (5.9 g, 11.8 mmol) in 60 mL methanol was added 60 mL of a 1M aqueous LiOH solution (1.44 g LiOH, 5 eq.) and then heated to reflux for 2 h. The resultant clear solution was then cooled to room temperature and the MeOH removed *in vacuo*. The resultant aqueous solution was then acidified to pH 3 by dropwise addition of 1M aqueous HCl. The resultant pale brown precipitate was isolated by centrifugation, and the supernatant was decanted. The pellet thus obtained was then washed three times with pH 3 water, with centrifugation and decantation of the supernatant. The brown pellet obtained was then taken up in a 50 °C, 5% v/v ethanol/water mixture and subsequently lyophilised to afford pure **2.12** (5.50 g, 11.4 mmol, 96%) as a pale brown powder.

^{1}H NMR (500 MHz, DMSO, δ=2.50 ppm): 8.80 (1H, s), 8.66 (1H, br s), 8.41 (2H, d, *J* = 7.2 Hz), 8.35 (1H, d, *J* = 7.2 Hz), 8.31 (1H, s), 8.02 (1H, d, *J* = 6.4 Hz), 7.95 (1H, d, *J* = 6.7 Hz), 7.84 (1H, d, *J* = 7.0 Hz), 7.74 (1H, d, *J* = 6.4 Hz), 7.49 (2H, d, *J* = 7.1 Hz), 4.23 (2H, s), 1.41 (9H, s).

^{13}C NMR (125 MHz, DMSO, δ=40.0 ppm): 167.5, 156.4, 152.4, 136.6, 133.5, 128.8, 128.1, 127.7, 126.0, 124.1, 78.4, 43.9, 28.8. (Not all signals reported due to degeneracy and poor signal strength)

FT-IR: (ATR): 2768, 1692, 1620, 1573, 1390, 1231 cm^{-1}.

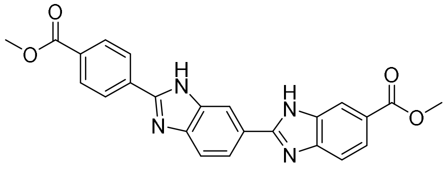
ESI-MS $[M+H]^+$: 484 m/z.

Melting Point: > 300 °C

Elemental Analysis: Calc.: 67.07%, 5.21%, 14.48%. Found: C 67.0%, H 5.35%, N 14.4 %

R_f: 0.25 (7:2:1 DCM: MeOH: Et_3N).

Methyl 2'-(4-(methoxycarbonyl)phenyl)-1H,3'H-[2,5'-bibenzo[d]imidazole]-6-carboxylate **(2.25)**



2.21

To a suspension of hydrolysed **2.18** (3.30 g, 7.9 mmol) **(**recovered through acidification of the aqueous fraction of the recrystallisation mixture) in 150 mL dry methanol was added $SOCl_2$ (0.86 mL, 1.41 g, 11.85 mmol), and the mixture heated to reflux for 3 h until a clear solution was observed, and full conversion was observed by TLC. The solvent and excess $SOCl_2$ were removed *in vacuo* and the dark solid residue was recrystallised from DMF/water to afford pure **2.21** (2.65 g, 6.47 mmol 82%) as a dark brown crystalline solid.

^{1}H NMR (500 MHz DMSO δ = 2.50 ppm): 8.28 (2H, d, *J* = 7.6 Hz), 8.15 (1H, s), 8.11 (2H, d, *J* = 8.8 Hz), 7.86 (2H, d, *J* = 8.8 Hz), 7.70 (1H, d, *J* = 8.4 Hz), 3.87 (3H, s), 3.86 (3H, s)

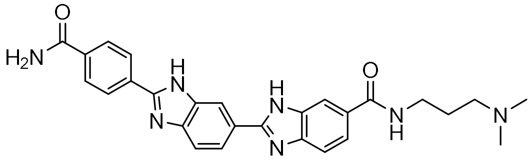
^{13}C NMR (500 MHz DMSO δ = 40.0 ppm): 167.0 166.2, 152.8, 133.3, 130.3, 127.6, 124.7, 124.5, 117.7, 115.4, 52.8, 52.6.

FT-IR (ATR): 1720, 1612, 1433, 1240, 1295 cm^{-1}.

ESI-MS [M - 2OAc + 3H]$^+$: 311 m/z.

R_f = 0.5 (9:1 DCM: MeOH).

Characterisation data of methyl 2'-(4-carbamoylphenyl)-1H,3'H-[2,5'-bibenzo[d]imidazole]-6-carboxylate (**2.23**)



2.23

^{1}H NMR (600 MHz, DMSO δ = 2.50 ppm): 9.62 (1H, br s), 9.22 (1H, br s), 8.52 (1H, s), 8.32 (2H, d, *J* = 8.0 Hz), 8.21 (1H, s), 8.16 (1H, dd, *J* = 8.3, 1.6 Hz), 8.13 (1H, br s), 8.09 (2H, d, *J* = 8.5 Hz), 7.88 (2H, d, *J* = 9.0 Hz), 7.76 (1H, d, *J* = 9.0 Hz), 7.51 (1H, br s), 3.39 (2H, q, *J* = 7.2 Hz), 3.13-3.16 (2H, m), 2.82 (3H, s), 2.81 (3H, s), 1.94 (2H, quin, *J* = 7.2 Hz).

^{13}C NMR (150 MHz, DMSO, δ = 40.0 ppm): 167.8, 167.3, 159.2, 159.0, 158.7, 153.8, 153.2, 145.8, 142.2, 141.1, 136.3, 132.1, 130.0, 128.7, 127.6, 127.2, 125.6, 123.1, 122.7, 120.3, 117.9, 116.0, 58.4, 46.3, 42.8, 37.0, 25.3.

FT-IR (ATR): 3165.2, 1666.5, 1587.4, 1548.8, 1311.5 cm^{-1}.

ESI-MS [M+H]: 482 m/z

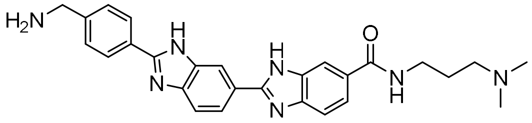
RP-HPLC: (5 – 90% buffer D in buffer C) t_r: 16.8 min.

Elemental Analysis: Calc. C 47.77% H 5.47% N 11.82%. Found: C 47.80%, H 5.75%, N 11.65% (consistent with ditrifluoroacetate salt, pentahydrate)

UV-Vis: λ_{max} = 340 nm (DMSO)

Molar extinction coefficient = 24000 $M^{-1}cm^{-1}$ (DMSO)

2'-(4-(aminomethyl)phenyl)-N-(3-(dimethylamino)propyl)-1H,3'H-[2,5'-bibenzo[d]imidazole]-6-carboxamide (**2.24**)[104]



2.24

To a stirred solution of **2.12** (2.50 g, 5.17 mmol) in 10 mL peptide grade DMF, was added HATU (1.96 g 5.17 mmol, 1.0 eq.) and DIPEA (2.50 ml, 2.00 g, 15 mmol, 3.0 eq.) and the mixture was stirred at r.t. for 30 minutes. DMAPA (2.50 ml, 3.13 g, 30.6 mmol, 5.9 eq.) was then added to the mixture and stirred at r.t for 2 h until full conversion of the starting material was confirmed by TLC (negative bromocresol green test, loss of spot at R_f = 0.25, 7:2:1 DCM: MeOH: Et_3N). The mixture was concentrated *in vacuo*. The dark, viscous residue was then taken up in 40 mL of saturated aqueous $NaHCO_3$ solution, and the pH was adjusted to 11 with 10 mL of 1M aqueous NaOH solution. The aqueous layer was then extracted three times with 50 mL of n-butanol and the combined organic layers, was filtered. The resultant clear yellow solution was then dried over Na_2SO_4, filtered, and concentrated to dryness *in vacuo.* The residue was then taken up in 50 mL of aqueous 2M HCl and then heated to reflux for 90 minutes until a clear solution was observed. The solution was then concentrated *in vacuo* and the crude product obtained was purified by recrystallisation from hot ethanol to afford pure **2.24** (1.50 g, 3.2 mmol, 62%) as an off-white solid. Purer samples for spectroscopic experiments were obtained by preparative RP-HPLC.

^{1}H NMR (400 MHz, DMSO δ = 2.50 ppm): 10.92 (1H, br s), 9.12 (1H, d, *J* = 5.4 Hz) 8.96 (1H, d, *J* = 6.2 Hz), 8.86 (1H, br s), 8.57-8.48 (3H, m), 8.37 (1H, s), 8.14 (1H, d, *J* = 9.5 Hz), 8.08 (1H, dt, *J* = 7.8, 2.3 Hz), 7.98 (1H, d, *J* = 5.5 Hz), 7.96 (1H, d, *J* = 5.3 Hz), 7.87 (1H, dd, *J* = 8.5, 3.0 Hz), 7.79 (1H, d, *J* = 8.5 Hz) 4.12 (2H, d, *J* = 6.0 Hz), 3.38 (2H, q, *J* = 6.2 Hz), 3.12 (2H, q , *J* = 4.3 Hz), 2.74 (3H, s), 2.73 (3H, s), 2.00 (2H, quin, *J* = 6.6 Hz).

^{13}C NMR (100 MHz, DMSO δ = 40.0 ppm): 166.2, 153.1, 151.2, 132.0, 130.3, 128.7, 128.0, 124.5, 122.3, 116.6, 116.1, 116.0, 114.0, 113.9, 112.0, 54.9, 42.4, 37.1, 24.6. (Only strong well resolved peaks are reported, signal degeneracy also observed).

FT-IR (ATR): 2969, 2638, 1701, 1625, 1468, 1219 cm^{-1}

ESI-MS [M+H]$^+$: 468

RP-HPLC (5 – 90% buffer D in buffer C) t$_r$: 10.4 min.

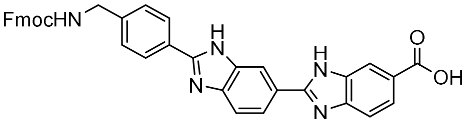
Elemental analysis: Calc. 51.43% H 6.55%, N 15.55%. Found: C 51.40%, H 6.35%, N 15.1% (Consistent with the dihydrochloride pentahydrate)

UV-Vis: λ_{max} = 332 nm

Molar Extinction Coefficient: 33000 M^{-1}cm^{-1} (water)

HPLC: t$_r$ (5-90% MeCN/water, 0.1% TFA over 18 minutes): 10.7 min.

2'-(4-(((((9H-fluoren-9-yl)methoxy)carbonyl)amino)methyl)phenyl)-1H,3'H-[2,5'-bibenzo[d]imidazole]-6-carboxylic acid (**2.25**)[108]



2.25

5ml neat TFA was added to compound **2.12** (1.00 g, 2.06 mmol), and the resultant solution was stirred for 90 minutes until no trace of starting material was observed by TLC (R_f = 0.0, 7:2:1 DCM: MeOH: Et_3N). The TFA was removed *in vacuo*, and the residue taken up in 50 mL water. The solution was raised to approximately 8 though the addition of 2 mL saturated aqueous $NaHCO_3$. A solution of Fmoc-Cl (0.80 g, 3.1 mmol, 1.50 eq.) in 50 mL of 1,4-dioxane was then added. The mixture was then stirred at room temperature for 16 hours. The dioxane was then removed *in vacuo* and the resulting mixture adjusted to pH 3 by dropwise addition of 1M HCl. The resultant precipitate was then collected by filtration, and the filtrate extracted twice with 50 mL n-butanol. The filter cake was redissolved in n-butanol, dried over Na_2SO_4 and then evaporated *in vacuo*. The residue was then recrystallised from ethanol/water to afford pure **2.25** (0.73 g, 59%) as an off-white to pale brown solid.

^{1}H NMR (500 MHz DMSO δ = 2.50ppm): 13.24 (1H, br s), 8.49 (1H, s), 8.44 (1H, d, *J* = 10.4 Hz), 8.41 (2H, d, *J* = 6.8 Hz), 8.18 (1H, d, *J* = 7.8 Hz), 7.92-7.80 (8H, m), 7.72 (2H, d, J = 8.5 Hz), 7.42 (2H, d, *J* = 8.4 Hz), 7.35 (2H, d, *J* = 8.0 Hz), 4.39 (2H, d, *J* = 7.4 Hz), 4.29 (2H, d, *J* = 7.7 Hz), 4.26 (1H, d, *J* = 6.8Hz).

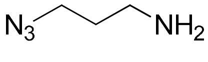
^{13}C NMR (500 MHz DMSO δ = 40.0ppm): 167.7, 154.0, 153.8, 153.5, 153.1, 145.7, 144.3, 133.6, 125.7, 125.2, 125.1, 127.9, 127.6, 127.5, 127.3, 126.2, 126.1, 125.7, 125.6, 125.3, 125.1, 125.0, 122.9, 122.8, 120.6, 120.3, 119.5, 116.7, 114.8, 113.0, 65.9, 64.3, 50.6, 47.3, 44.1.

ESI-MS $[M+H]^+$: 606 m/z

R_f : 0.4 (7:2:1 DCM:MeOH:Et_3N).

Elemental Analysis: Calc.: C 73.32% H 4.46% N 11.56% C 73.32%, H 4.46%, N 11.50%

3-azidopropylamine[122]



To a stirred solution of 3-bromopropylamine hydrobromide (10.0 g, 46 mmol, 1.0 eq.) in 100 mL water was added sodium azide (4.50 g, 69 mmol, 1.5 eq.) and the mixture heated to reflux for 12 h. The mixture was then cooled to 0 °C. Solid sodium hydroxide pellets were added to a final concentration of 1M, and the previously clear solution became cloudy. The mixture was then extracted with diethyl ether (3 × 100 mL). The combined organic layer was then dried over Na_2SO_4, filtered and then concentrated *in vacuo*. The pure product was obtained as yellow oil in quantitative yield.

^{1}H-NMR ($CDCl_3$, $\delta = 7.26$ ppm): 2.59 (2H, t, $J = 6.5$ Hz), 1.90 (2H, s), 1.77 (2H, t, $J = 6.5$ Hz), 1.59 (2H, quin, $J = 7.5$ Hz)

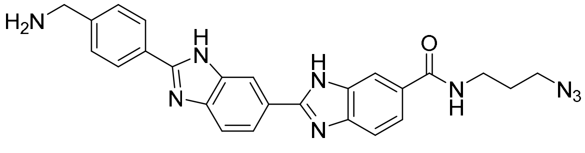
^{13}C-NMR ($CDCl_3$, $\delta = 77.0$ ppm): 49.0, 39.2, 32.6

FT-IR (ATR): 2091, 1662, 1258cm$^{-1.}$

ESI-MS $[M+H]^+$: 101 m/z

2'-(4-(aminomethyl)phenyl)-N-(3-azidopropyl)-1H,3'H-[2,5'-bibenzo[d]imidazole]-6-carboxamide (**1.20**):[104]

Method A (from 2.25):



1.20

To a stirred solution of compound **2.25** (350 mg, 0.58 mmol) in 5 mL peptide grade DMF was added HATU (220 mg, 0.58 mmol, 1.0 eq.) and DIPEA (0.6 mL, 0.45 g, 3.48 mmol, 6.0 eq.), and the compound stirred at room temperature for 30 minutes. 3-azidopropylamine (580 mg, 5.80 mmol, 10 eq) was then added and the mixture was then stirred at room temperature for 16 hours. 50 mL of a 20% v/v aqueous solution of piperidine was then added and the mixture was stirred at room temperature for 15 minutes. The crude reaction mixture was then extracted three times with n-butanol (3 × 50 mL). The combined organic phases were then dried over Na_2SO_4, and then concentrated *in vacuo* to furnish crude **1.20,** which was then purified by RP-HPLC (mobile phase: 5 – 90% buffer D in buffer C). The fractions containing **1.20** were then neutralised using triethylamine, and then lyophilised. The residual oil was then desalted through a short column of C18 RP silica to furnish pure **1.20** (4.7 mg, 10.1 µmol 1.74%) as a pale yellow solid.

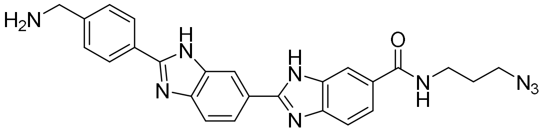
^{1}H NMR (400MHz CD_3OD δ = 4.87, 3.34 ppm): 8.40 (1H, s), 8.16 (1H, s), 8.14 (1H, s), 8.07 (1H, dd, *J* = 7.0, 2.0 Hz), 7.78 (2H, dd, *J* = 7.5, 1.5 Hz), 3.96 (2H, br s), 3.54 (2H, t, *J* = 6.5 Hz), 3.48 (2H, t, *J* = 6.5 Hz), 1.96 (1H, quin, *J* = 6.5 Hz).

FT-IR (ATR): 2098.6, 1672.3, 1201.7, 1175.6, 1127.6cm^{-1}

Retention time (5 – 90% buffer D in buffer C): 12.1 minutes

ESI-MS [M+H]$^+$: 466

Method B (from 2.12):



1.20

To a stirred solution of **2.12** (1.50 g, 3.1 mmol) in 10 ml peptide grade DMF was added HATU (1.18 g, 3.1 mmol, 1.0 eq.), and DIPEA (2.0 ml, 1.5 g, 4.0 eq.) and the mixture stirred at room temperature for 30 minutes. 3-azidopropylamine (2.50 g, 8.0 eq.) was added to the mixture and then the mixture was stirred overnight at room temperature. Once the reaction was complete according to bromocresol green test, the mixture was concentrated *in vacuo* as far as possible, and the oily dark residue taken up in a pH 3 solution of HCl in water. The brown precipitate that formed was then collected by centrifugation, the supernatant decanted and the pellet was then thoroughly washed four times with pH 3 water, with centrifugation and decanting of the supernatant for each wash. The pellet was then taken up in 5 mL of a 4 M solution of HCl in 1,4, -dioxane, and stirred at r.t. for 30 minutes. The mixture was then precipitated through addition of 50 mL diethyl ether and crude **1.20** was collected by centrifugation. The mixture was washed twice more with diethyl ether with centrifugation and decanting of the supernatant for each wash. Pure **1.20** (1.34 g, 2.88 mmol, 93%) was obtained by recrystallisation from methanol.

^{1}H NMR (400 MHz, DMSO δ = 2.50 ppm): 8.89 (2H, s), 8.66 (3H, s), 8.45 (2H, d, *J* = 7.6 Hz), 8.45 (1H, d, *J* = 7.6 Hz), 8.31 (1H, s), 8.03 (1H, d, *J* = 8.1 Hz), 8.01 (1H, d, *J* = 7.6 Hz), 7.87 (1H, d, *J* = 7.6 Hz), 7.78 (2H, d, *J* = 8.5 Hz), 4.14 (2H, q, *J* = 5.2 Hz), 3.45 (2H, q, *J* = 6.5 Hz), 3.38 (2H, t, *J* = 7.5 Hz), 1.83 (2H, quin, *J* = 7.5 Hz) (one benzimidazole resonance missing).

^{13}C NMR (100 MHz, DMSO δ = 40.0 ppm): 166.0, 153.0, 151.1, 138.6, 134.8, 132.9, 132.3, 130.1, 128.6, 128.4, 125.2, 124.5, 116.1, 114.0, 113.8, 49.1, 42.3, 37.4, 28.8 (some resonances missing due to degeneracy and weak signals of quaternary atoms).

FT-IR (ATR): 2924, 2100, 1614, 1442, 1139, 1109 cm^{-1}

MALDI-MS [M+H]$^+$: 465

ESI-MS [M+H]$^+$: 465

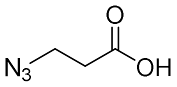
HPLC: t$_r$ (5 – 90% buffer D in buffer C): 12.1 minutes

Elemental Analysis: Calc.: C 50.72% H 5.79% N 21.29%. Found: C 51.99%, H 6.04% N 19.58% (consistent with the hydrochloride pentahydrate, some decomposition of the azide observed).

UV-Vis: λ_{max} = 334 nm

Molar Extinction coefficient: 24500 M^{-1}cm^{-1} (DMSO)

3-azidopropionic acid (**2.27**)[110]



2.27

To a stirred solution of 10 g 3-bromopropionic acid (10.0 g, 65 mmol) in 100ml water was added sodium azide (6.37g 98 mmol, 1.5 eq.) and the mixture was heated to reflux for 12 hours. The mixture was then acidified to pH 1 with 4 M aqueous HCl, accompanied by a colour change from orange to pale yellow, upon which the mixture extracted with diethyl ether (5 x 100 mL).

The combined fractions were then dried over Na_2SO_4, filtered, and then concentrated *in vacuo*. The residue, which contained traces of Et_2O by ^1H NMR, was then dried *in vacuo* overnight, and the yellow oil obtained (7.4 g, quantitative yield) was observed to be pure **2.27**.

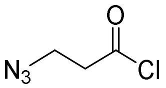
^1H NMR (500MHz, $CDCl_3$, $\delta = 7.26$ ppm): 3.59 (2H, t, $J = 6.5$ Hz), 2.64 (2H, t, $J = 6.5$ Hz)

^{13}C NMR (125MHz, $CDCl_3$, $\delta = 77.0$ ppm): 176.8, 46.4, 33.8.

FT-IR (ATR): 2098, 1712, 1240, 1176 cm^{-1}.

ESI-MS $[M+H^+]$: 116

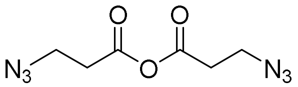
3-azidopropionyl chloride (**2.37**)[123].



2.28

To a stirred solution of 3-azidopropionic acid (1.00g, 8.7mmol) in 17 mL dichloromethane was added oxalyl chloride (0.9 ml, 1.32 g, 10.7 mol, 1.2 eq), and the mixture stirred at room temperature for 6 hours. The resultant solution thus obtained was then used in the next step without any purification as a solution of the product with an assumed concentration of 0.51 M.

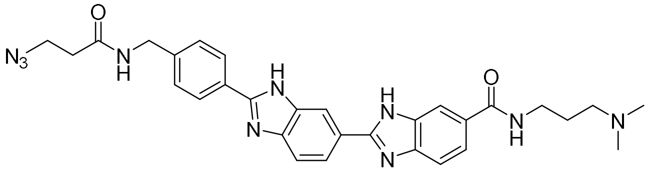
3-azidopropionic anhydride **(2.29)**



2.29

To a stirred solution of DCC (1.00 g, 4.85 mmol, 0.45 eq.) in dry DCM (11.0 mL) was added 3-azidopropionic acid (1.25 g 10.6 mmol, 1.0 eq.) and the mixture stirred at room temperature for 4 hours. Upon completion, the now cloudy mixture was filtered through a pad of Celite to remove the white precipitate that formed. The resultant pale yellow solution was dried by blowing argon through the solution for 30 minutes and stored under argon without any further purification.

2'-(4-((3-azidopropanamido)methyl)phenyl)-N-(3-(dimethylamino)propyl)-1H,3'H-[2,5'-bibenzo[d]imidazole]-6-carboxamide **(1.21):**



1.21

Method A (from 2.24 by HATU-mediated coupling):

To a stirred solution of **2.24** (200 mg, 1.65 mmol, 3.0 eq.) in 2ml dry DMF, was added HATU (420 mg, 1.1 mmol, 2.0 eq.) and DIPEA (0.5 mL, 0.625 g, 4.8 mmol, 8 eq.). The mixture was then stirred vigorously at room temperature for 30 minutes upon which **2.24** (250 mg, 0.55 mmol, 1.0 eq.) was added. The mixture was stirred at room temperature until full conversion of **2.24** was observed by Kaiser test. Once complete, the residue was then concentrated *in vacuo* as far as possible. The residue was then taken up in 50 mL of saturated aqueous $NaHCO_3$, adjusted to pH 11 with 1 M aqueous NaOH, and the mixture extracted three times with 50 mL n-butanol. The collected organic layer was then dried over Na_2SO_4 and filtered. The resultant orange solution was then concentrated *in vacuo* and the residue taken up in a 1:1 v/v mixture of water and acetonitrile. Crude **1.21** was then purified by RP-HPLC using a TFA peptide buffer system (5-90% MeCN over 30 minutes). Fractions confirmed to contain **1.21** by ESI-MS were then buffered by addition of triethylamine, lyophilised to dryness, and the oily residue desalted through a short C18 RP silica plug to furnish **1.21** (3.60 mg, 6.37 μmol, 1.2%) as a hygroscopic orange solid.

Characterisation as per Method C.

Method B (from 2.24 and 2.28):

To a flame-dried three-necked flask equipped with a stirrer bar containing a stirred suspension of **2.24** (100 mg, 0.21 mmol) in 10 mL dry THF was added dry triethylamine (400 mg, 4.2 mmol, 20.0 eq.) and a solution of **2.28** (0.51 M, 4.2 mL, 10.0 eq.) and the mixture stirred at room temperature until reaction was complete (negative Kaiser test). Upon completion, any remaining acid chloride was quenched through addition of ethanol, and the mixture concentrated to dryness *in vacuo*. The crude solid thus obtained was taken up in 10 mL saturated aqueous $NaHCO_3$ and adjusted to pH 11 through addition of 1M aqueous NaOH, and then extracted 3 times with 10 mL n-butanol. The collected organic layer was then filtered, dried over Na_2SO_4 and concentrated to dryness *in vacuo*. The residue was then taken up in 10 mL ethanol, and precipitated as a solid through dropwise addition of 4 M HCl in dioxane. The precipitate thus obtained was then isolated by centrifugation, and washed twice with DCM, with centrifugation and decanting of the supernatant. The pellet thus obtained was then dissolved in 5 mL water and then lyophilised to dryness to obtain crude **1.21,** which was observed by ^{1}H NMR to contain approximately 3 eq. of triethylamine hydrochloride. This crude mixture was then purified by repeated desalting procedures through a short column of C18 RP-silica to furnish pure **1.21** (15.0 mg, 25.2 μmol, 12%) as an orange solid.

Method C (from 2.24 and 2.29):

To a stirred solution of deprotonated **2.24** (200 mg, 0.43 mmol, 1.0 eq.) in 10 ml dry DMF was added a solution of **2.29** in DCM (0.5 M, 9 mL, 4.3 mmol 10.0 eq.) and the mixture stirred at room temperature for 16 h. Once confirmed complete by negative ninhydrin test, the mixture was then concentrated *in vacuo*. The residue was taken up in saturated aqueous $NaHCO_3$ (25 mL) and then 5 mL 1 M NaOH was added. The aqueous layer was then extracted twice with 25 mL n-butanol. The separated aqueous layer was then dried over Na_2SO_4 and concentrated *in vacuo*. The residue, which contained dicyclohexylurea and other impurities was then purified as per method A to furnish **1.21** (11.5 mg, 20 μmol, 4.7 %) as a hygroscopic orange solid.

^{1}H NMR (500 MHz, DMSO δ = 2.50 ppm): 10.97 (1H, s), 9.11 (1H, br s), 8.97 (1H, br s), 8.86 (1H, t, *J* = 6.0 Hz) 8.61 (1H, d, *J* = 8.4 Hz) 8.56 (1H, d, *J* = 8.4 Hz) 8.52 (1H, d, *J* = 8.0 Hz), 8.46 (1H, d, *J* = 7.8 Hz), 8.37 (1H, s), 8.08-8.13 (1H, m), 8.00 (1H, d, *J* = 8.2 Hz), 7.86 (1H, d, *J* = 8.4 Hz), 7.66 (1H, d, *J* = 8.4 Hz) 7.54 (1H, d, *J* = 8.0 Hz), 4.66 (1H, br s) 4.40 (1H, q, *J* = 5.5 Hz), 3.57 (2H, t, *J* = 6.3 Hz), 3.39 (2H, q, *J* = 6.3 Hz), 3.11 (2H, t, *J* = 5.5 Hz), 2.74 (3H, s), 2.73 (3H, s), 2.52 (2H, t, *J* = 6.6 Hz), 2.00 (2H, quin, *J* = 7.0 Hz).

FT-IR (ATR): 2924, 2100, 1614, 1442, 1139, 1109 cm^{-1}.

Retention time (Buffer D in buffer C): 13.1 minutes

ESI-MS = 565 [M+H]$^+$.

UV-Vis: λ_{max} = 334 nm

Extinction coefficient: 30500 M^{-1}cm^{-1} (water)

Retention time (0.1 % TFA in H_2O/MeCN, 5-90 % MeCN 18 min gradient): 13.2 min.

Elemental Analysis: Calc.: C 50.64% H 5.79% N 17.89%. Found: C: 50.39%, H 5.35%, N 16.69% (consistent with the trifluoroacetate salt pentahydrate).

3. Analysis of thermal melting of an A_3T_3 dsDNA tract in complex with compounds 1.9, 1.20 and 1.21.

3.1 Introduction

*3.1.1: Current state-of-the-art in the analysis of the dynamics of H33258***:** *dsDNA complexes.*

Although the thermodynamics of formation of complexes of Hoechst 33258 **1.9** (Fig. 3.1) with A-T tracts of dsDNA are well understood, *i.e.* the main driving force is a large positive entropy change,[53,124] the molecular processes by which the complex forms is much less well-understood.

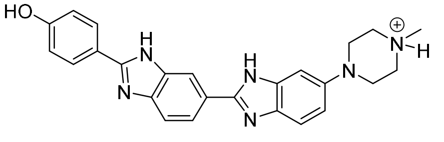

1.9

Figure 3.1: Structure of Hoechst 33258, **1.9**

One of the main reasons why this is the case is that the standard spectroscopic techniques used to study the dynamics of biomolecular processes lack the time resolution (in the case of NMR), or structural resolution (in the case of fluorescence) necessary to investigate the dynamics of this process (Section 1.5.3).[7] The kinetics of complex formation between **1.9** and an $A_{30}T_{30}$ tract of dsDNA has been shown to be extremely fast, with a k_{on} of 220 ± 65 x 10^3 s^{-1}.[50] As such, studies of the dynamics of biomolecular processes have largely been limited to MD simulations and calculations. The infrared spectroscopy work described in part in this thesis represents the first direct observation of the mechanism of thermal dissociation of a bisbenzimidazole: dsDNA complex, with a view to looking at the association dynamics of this process.

3.1.2. Aims.

The aims of this chapter are:

i. To study the melting of dsDNA into ssDNA by both UV-Vis and IR spectroscopy.

ii. The UV and IR melting behaviour of **1.9** in complex with dsDNA will be studied by UV-Vis and IR spectroscopy. This data will serve as a comparison to ensure that the data obtained for **1.20** and **1.21** is similar enough to that obtained for **1.9.**

iii. To study the melting behaviour of **1.20** and **1.21** in complex with dsDNA by UV-Vis and IR spectroscopy in both the 1575 – 1700 cm^{-1} and azide regions in order to ascertain whether the azide functional group serves as a useful probe for IR studies on ligand-dsDNA complexes.

*3.1.3. Fluorescence studies of the **1.9**: dsDNA complex.*

Fluorescence studies of free dsDNA have shown evidence of an intermediate state between melted and duplex DNA; that there exists a transient state in which the DNA no longer exists in a duplex, nor is it fully melted, rather it is characterised by transient duplexation and mispairing as the bases attempt to form Watson-Crick base pairs with each other. Hence, DNA melting is not a simple two-state model, as would be expected by inspection of a simple UV-Vis melting curve.[125] Further computational studies have shown that in high-salt concentration solutions, an intermediate state exists between base paired and melted DNA upon heating of the sample. The characteristic feature of this state is weak and nonspecific base pairing between the two strands. Full melting is observed when the mixture is heated further – i.e. to the top of the sigmoid curve (Fig 1.27).[126]

The fluorescence of Hoechst 33258 increases sixtyfold in complex with dsDNA compared to free solution, and hence fluorescence spectroscopy has been used to study the kinetics of association and dissociation of the ligand to dsDNA.[127] Unlike the hypothesised linear two-state model in which association of **1.9** with dsDNA causes a change in the conformation of DNA (Fig 3.2), a more nuanced picture of binding was observed.[90] The rate constant k_1

represents the rate of formation of a complex of dsDNA and an incoming ligand, and k_2 represents a rearrangement step in which the dsDNA changes conformation to accommodate the ligand in its optimal binding site. In the case of **1.9**, this conformational change comes from "slipping" of the ligand from the AATT binding site to accommodate the bulky piperazino moiety – consistent with the different binding sites observed from X-ray structures.[64,128] This appears to be related to the shape of the dsDNA, if a TpG or TpC base step is present, the piperazino moiety sits across the GC base pair, leading to an apparent 5'-AATTG-3' binding site.

$$\text{(a)}\qquad \text{DNA + ligand} \underset{(k_{-1})}{\overset{(k_{+1})}{\rightleftharpoons}} \text{Complex} \underset{(k_{-2})}{\overset{(k_{+2})}{\rightleftharpoons}} \text{Rearranged complex}$$

$$\text{(b)}\qquad K_a = K_1(1+K_2)$$

Figure 3.2. (a) The classical two-state model governing formation of **1.9**: dsDNA complexes.[90] **(b)** The overall association constant for the process.

If a TpA base step is present, the wide minor groove exhibited by this base step accommodates the piperazino group, and the apparent binding site becomes 5'-AATTA-3'.[90]

Although evidence exists for this mechanism, observed fluorescence spectra also showed evidence for a parallel single-step association mechanism between **1.9** and dsDNA (Fig. 3.2).[90] The optimised complex and suboptimal complexes are mutually exclusive, and the mechanism of complex formation can be thought of as the ligand "seeking" its optimal binding site, and weak association and dissociation occurs until the optimal binding site is found.[90] Elements from both of these proposed reactions were observed from the fluorescence spectra of the **1.9:** dsDNA complex, and neither of these models proved sufficient to explain the binding behaviour of **1.9**.[90]

$$\text{(a)}\quad \text{DNA + ligand} \underset{k_{-a}}{\overset{k_{+a}}{\rightleftharpoons}} \text{Suboptimal complex}$$

$$\text{DNA + ligand} \underset{k_{-a}}{\overset{k_{+a}}{\rightleftharpoons}} \text{Optimal complex}$$

$$\text{(b)}\ K_a = K_A + K_B$$

Figure 3.3: (a) The single step model of ligand: dsDNA association. (b) The relationship of this model to the overall binding constant.[90]

Although fluorescence has some utility in understanding the binding behaviour of **1.9** to A-T tracts of dsDNA, it involves electronic excitation of the whole molecule, and hence although it has the temporal resolution necessary to probe the dynamics of **1.9**: dsDNA formation, it lacks the structural resolution necessary for probing the changes in dsDNA structure, and hence to what extent conformational changes in dsDNA are as a determinant of the binding behaviour of **1.9** and its analogues. With the 2D-IR approach taken by this project, it will prove possible, by understanding the changes in the IR spectrum of dsDNA, understand the conformational changes of dsDNA upon binding of a small molecule ligand. Moreover, circular dichroism and fluorescence studies on both duplex and hairpin DNA have shown that A-T tracts are unique in that a premelting complex occurs where the oligomer is not fully melted, nor is it recognisably dsDNA.[129,130] it has been hypothesised that these "premelted complexes" present unique binding sites for either proteins or small molecules. The greater structural resolution of 2D-IR spectroscopy should allow for some structural characterisation of this "premelted complex".

3.1.4. Molecular Dynamics simulations of the H33258: dsDNA complex.

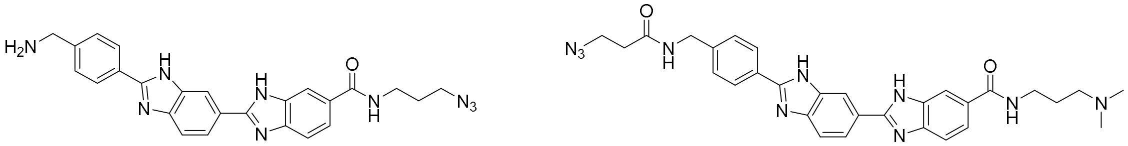
The dynamics of the hydration of complexes of **1.9** and dsDNA is key to understanding the dynamics of complex formation and have been previously studied by calculation and time resolved fluorescence spectroscopy. In addition to the observation that the thermodynamic driving force governing binding of **1.9** to dsDNA is dependent on the extrusion of water from the groove to the bulk solvent, it has been shown that the fluorescence lifetime of both bound and free **1.9** is dependent on the solvation of the chromophore.[131] It was shown that, although there is an increase in fluorescence lifetime consistent with the change in solvent environment from water to the minor groove, the magnitude of the change would suggest that the ligand is still heavily hydrated by weakly bound water molecules. Another longer contribution to the fluorescence lifetime comes from conformational reorganisation of the DNA duplex to accommodate **1.9**.[131] These observations serve as evidence for a rigid binding mode of a heavily hydrated complex with water-mediated hydrogen bonding between **1.9** and dsDNA. Molecular dynamics simulations based on this finding have shown that the solvation of **1.9** is not markedly different from that in free solution, and the enhanced fluorescence lifetime can be wholly attributed to interactions between the duplex and the ligand.[132]

Molecular dynamics simulations have also been used to model the behaviour of an A-tract of DNA upon binding of **1.9**.[84] The simulation showed that the energy minimum is the formation of narrow B-type DNA tracts such as that shown in Figure 2.2. The preference of **1.9** for A-T tracts as opposed to A-T repeats, T-A tracts, or especially, T-A repeats was explained by the characteristically high twist angle of TpA steps leading into widening of the minor groove, greater hydration of the groove, and hence dramatically reduced binding affinity.[84] Conversely, A-T tracts exhibit narrow minor grooves at equilibrium,[35] and can be considered to be a "pre-arranged" binding site for the incoming ligand.[5] A-T repeats however, although they exhibit a minor groove of a similar width to idealised B-DNA at equilibrium,[35] are made up of the most conformationally flexible of any two base-pair steps[19] and as such, can arrange into conformations that mirror both that of an A-tract and consecutive ApT steps.[5]

Consequently, there is evidence for both a prearranged "lock-and-key" for A-T tracts, and an "induced-fit", in which dehydration of the groove causes narrowing for A-T repeats, model for the binding of **1.9** to dsDNA for an A-tract, or A-T repeats, respectively. The 2D-IR approach taken in the course of this project will serve to give a fuller picture of the binding mechanism of compounds similar to **1.9** – allowing for the design of new DNA binding therapeutics. This approach will - like fluorescence anisotropy - indicate the binding state and local solvation dynamics through labelling of **1.9** with an infrared active probe, and secondly through correlation of the vibrations associated with the DNA duplex, direct observation of the time-resolved picture of the association and dissociation mechanisms of **1.9** from dsDNA can be obtained.

3.1.5. IR studies of dsDNA.

Two studies on the IR signatures of DNA bases have served as models for interpreting the IR spectra of complexes of **1.20** and **1.21** (Fig. 3.4) in complex with dsDNA: a review of experimental studies reported by Gräslund *et al.*,[133] and a thorough computational study by Cho *et al.*.[134,135,136,137]



1.20 **1.21**

Figure 3.4. Structures of azides **1.20** and **1.21**.

The review published by Gräslund[133] served as a library for initial assignment of the one-dimensional IR spectra through recording and assignment of the IR spectra of DNA oligomers containing exclusively G-C or A-T base pairs in D_2O. Works by Cho *et al.*, calculated the effects of base pairing,[134] solvent deuteration,[135] simulation of the dynamics of dsDNA systems and the 2D-IR spectra of dsDNA, and the IR signatures of A-, B-, and Z-form dsDNA.[137] The vibrational modes related to the bases (Fig. 3.5) are coupled to each other through the base pairs, and hence, changes in the IR spectrum of dsDNA, and by extension complexes with small molecules, are indicative of transition to a melted state.[138]

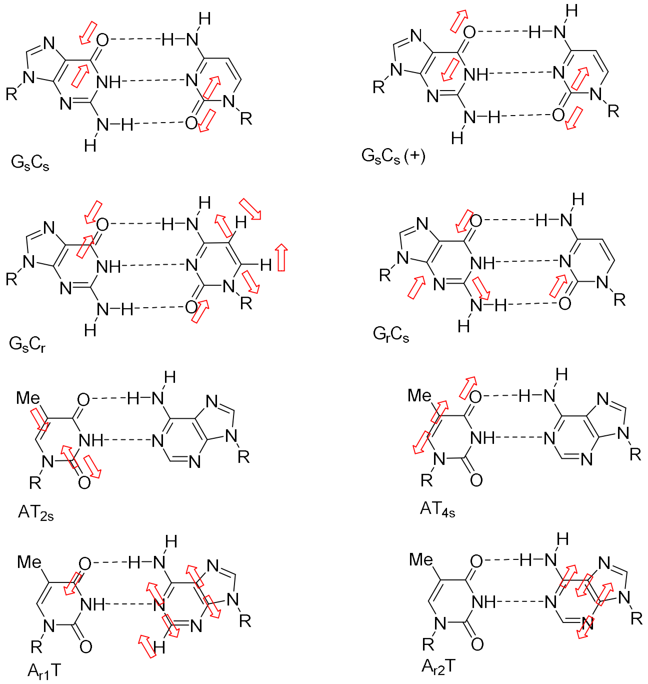


Figure 3.5. The eight basis vibrational modes of DNA bases. The nomenclature of the vibrations are indicated below, and the bonds comprising the vibrations are shown with red arrows.[135]

Vibration (GC pair)	**Frequency (cm^{-1})**	**Vibration (AT pair)**	**Frequency (cm^{-1})**
G_s	1694.7	T_{2s}	1699.3
C_s	1646.1	T_{4s}	1650.0
C_r	1624.4	A_{r1}	1627.0
G_r	1576.6	A_{r2}	1554.4
		T_r	1627.0

Table 3.1. Calculated frequencies of deuterated base pairs associated with 5 D_2O molecules.[135]

Wavenumbers (cm^{-1})	**Assignment**	**Fig 3.5 Assignment**	**Comments**
1688-1691	T ss/ds	$A_{r1}T_{2s}$ (dsDNA) T_{2s} (ssDNA)	C2=O2 stretch of thymidine. Increases in intensity upon duplexation.
1689-1678	G ds	G_sC_s (dsDNA)	C6=O6 stretch of Guanosine paired to Cytosine carbonyl
1673-1660	G ss/	G_s (ssDNA)	C6=O6 stretch of G
1671-1655	T ss /	T_{4s} (ssDNA)	C4=O4 of T ss. Less intense in dsDNA.
1655-1679	C ss/ds	G_sC_s (+)	C carbonyl Increases in intensity upon duplexation
1645-1641	T dsC ds	T_r, G_sC_r	In plane ring vibration
1632-1622	A ss/ds	$A_{r1}T_r$ (dsDNA) A_{r1} (ssDNA)	Decreases in intensity upon duplexation.

			Contribution from cytosine here too.
1590-1575	G ss/ds,	G_rC_s (dsDNA)	C=N ring vibrations
		G_r (ssDNA)	Decreases in intensity upon duplexation

Table 3.2. Experimental observations of the melting behaviour of the 1550-1700 cm^{-1} region of DNA.[133]

The basis vibrational modes of the vibrations shown in Fig. 3.5 are shown in Table 3.1. The frequencies of the vibrational modes shown in Tables 3.1 and 3.2 serve as a guide for assignment of the IR spectra of complexes of **1.20** and **1.21** with dsDNA. Mapping the changes in the IR spectrum of these complexes will firstly indicate the melting behaviour of the duplex, and secondly correlation of the changes in the azide region of the IR spectrum with the changes observed in the DNA region will prove the efficacy of the use of the azide as a reporter functional group. In the case of the Gräslund study, the melting behaviour of the free dsDNA was recorded in order to study the different behaviour of the vibrational modes of dsDNA and ssDNA – i.e. the behaviour of DNA as it melts. The results are summarised in Table 3.2. The data in Table 3.2 provides a useful means for interpreting the changes in the IR spectrum upon melting of dsDNA. Some differences exist between the Cho and Gräslund studies; these are likely due to discrepancies between calculation and assignment of the experimental data. Where possible, the Cho assignments are also shown in Table 3.2.

3.2 DNase I footprinting analysis.

To find the consensus binding sequence of **1.20** and **1.21**, and to compare to that of **1.9**, a DNase I footprinting experiment was carried out on **1.20** and **2.24** by Professor K. R. Fox at the University of Southampton.[139]

For each sequence of dsDNA tested (in which all possible dsDNA hexamers were represented) the gaps in the digestion patters pointed to a clear preference for 5'-AAATTT-3' sequences over all others, notably 5'-ATATAT-3' was observed (Figs. 3.6-3.8).[139] As such, the DNA oligomers chosen over the course of this project all contained the consensus sequence 5'-AAATTT-3'. The footprinting results showed that **1.20** and **2.24** have an almost identical consensus binding sequence to **1.9** (5'-AATT-3')[89], although the presence of the amide functional groups increase the length of the consensus binding site by two base pairs. Gaps in the digestion pattern corresponding to other A-T tracts of four and six base pairs are also evident from the digestion patterns. As expected, T-A tracts are absent.

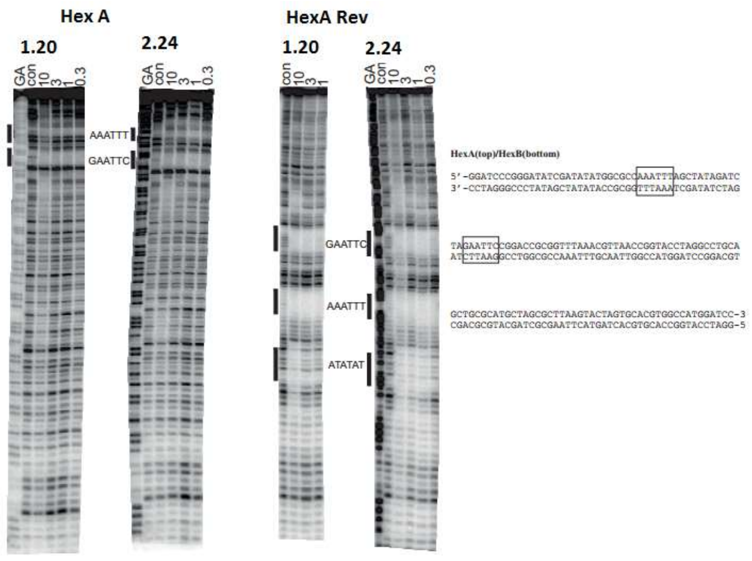


Figure 3.6. DNase I footprinting results of the HexA dsDNA fragment (sequences shown.)[139]

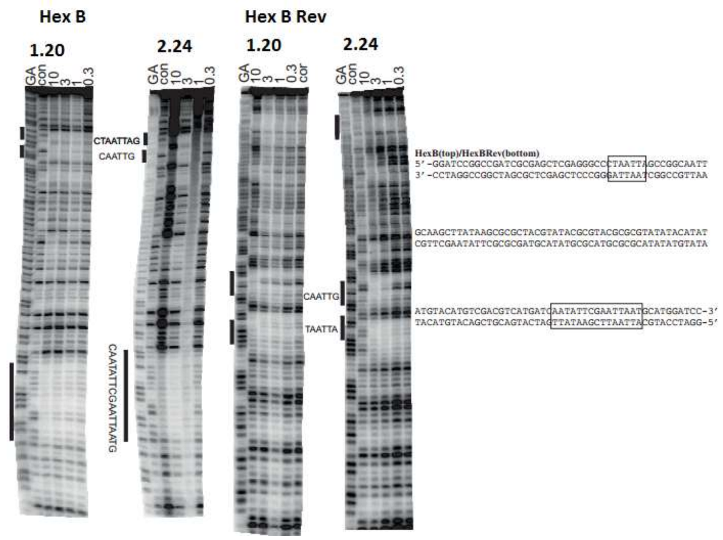


Figure 3.7. DNase I footprinting results of the HexB dsDNA fragment (sequences shown.)[139]

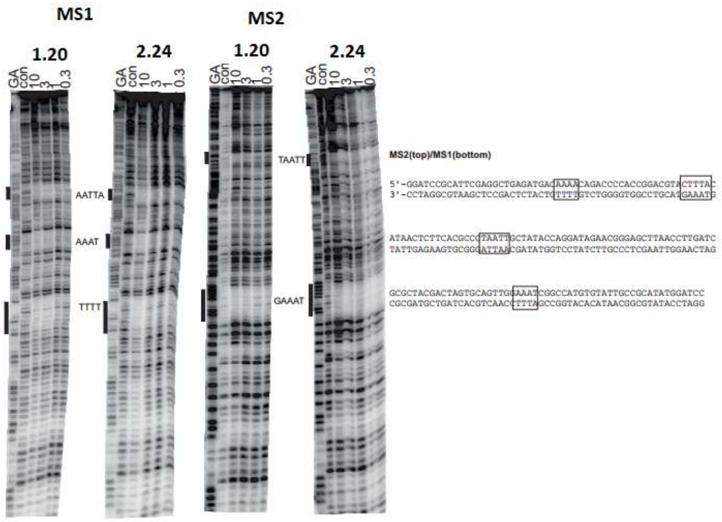


Figure 3.8. DNase I footprinting results of the MS1 and MS2 dsDNA fragments (sequences shown.)[139]

3.3. Analysis of the melting behaviour of dodecamer and decamer dsDNA.

The majority of the infrared spectroscopic experiments discussed hereafter in this chapter form part of a collaborative effort that links the work in this thesis with the PhD work of Lennart Ramakers of the Department of Physics at the University of Strathclyde. Where the data presented is a result of this collaboration it is clearly referenced.

3.3.1: UV and IR melts of free dodecamer dsDNA.

The DNA oligomer initially used in this study, ODN 3.1, is shown in Fig. 3.9.

5'-GCCAAATTTGCG-3'
3'-CGGTTTAAACGC-5'

Figure 3.9. Structure of ODN 3.1. The second strand is complementary to that depicted.

The melting temperature of ODN 3.1 was measured by UV spectroscopy, and the melting temperature was observed to be 51 °C. (Fig. 3.10).

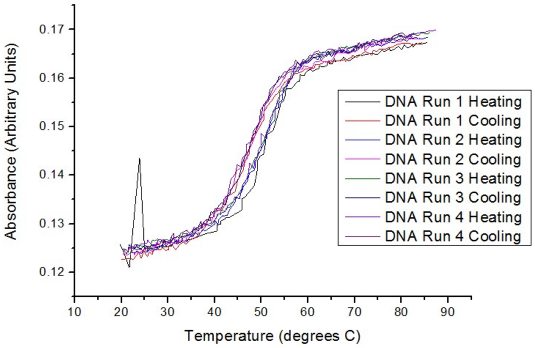


Figure 3.10. Melting curve of ODN1 at 1μM. The spike in Run 1 is an artefact of the spectrometer.

The IR melting curve in D_2O is shown in Figure 3.11. Four peaks were identified from the 1550 – 1700 cm^{-1} region of the spectrum at 1570, 1620, 1660 and 1690 cm^{-1}. These vibrations are dominated by the G_rC_s, $A_{r1}T_r$, $T_{4s}C_s$, and T_{2s} vibrations respectively, although contributions from other base pairs are undoubtedly present.[136] Fig. 3.11 also shows the changes in the spectrum upon melting of the DNA duplex.

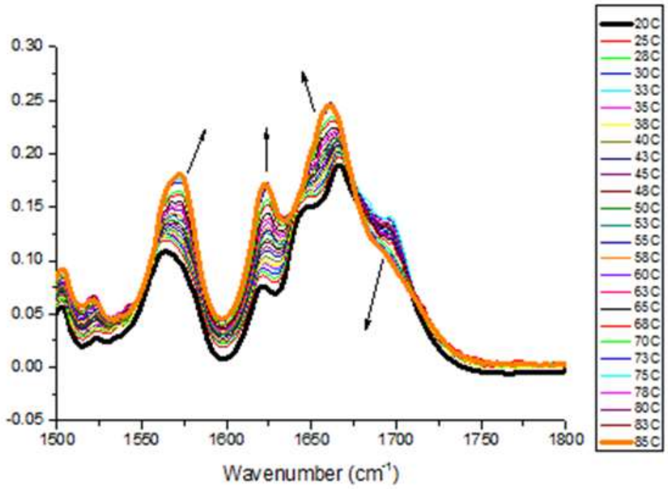


Figure 3.11. DNA melting curve for free ODN 3.1 at 5 mM. Arrows indicate the changes observed in the spectrum with increasing temperature.

To understand the changes in the spectrum, a difference spectrum of the IR spectrum of ODN 3.1 was generated through subtraction of the observed DNA spectrum from the spectrum collected for dsDNA at 32 °C. The results are shown in Fig. 3.12.

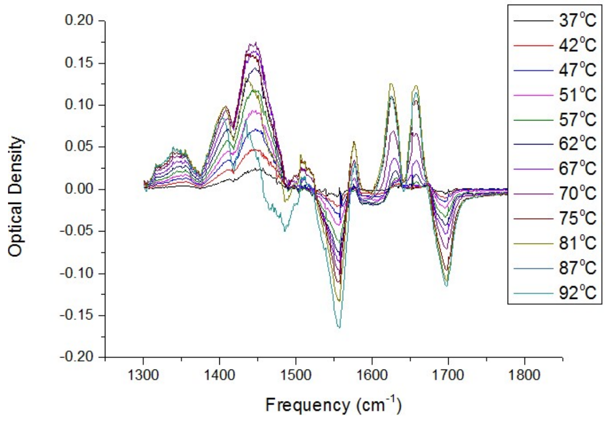


Figure 3.12. Difference spectrum of ODN 3.1.

The difference spectrum reflects the changes observed in Fig. 3.11. As predicted in Table 3.2, there is an increase in the peaks at 1620 and 1660 cm^{-1} and a decrease at 1690 cm^{-1}. In contrast to Table 3.2, there appears to be a decrease in the intensity of the peak at 1570 cm^{-1}, this is likely due to a blue shift in the peak position upon melting, as evidenced by the presence of the positive peak at ~1590 cm^{-1}. This blue shift is due to the breaking of the Watson-Crick base pairing upon melting, the lack of hydrogen bonding due to base-pairing strengthens the C=N bonds, and hence the vibrational frequency increases. Moreover, the presence of the peak centred at 1450 cm^{-1} (due to the presence of phosphate salts coming from the buffer) could complicate interpretation of the IR absorption band at 1570 cm^{-1}. Consequently, all further investigations of the dsDNA reported in this chapter have focused on the 1600 – 1700 cm^{-1} region.

The next step in the analysis of the IR spectrum of ODN 3.1 was to fit the spectrum to a series of Gaussian peaks corresponding to the appropriate IR bands in the spectrum. An example of the fitting is shown in Appendix 7.13. The strategy employed was to take the spectrum at 51 °C, at which the presence of melted and duplex DNA is approximately equal and fit the spectrum to five Gaussian peaks at 1570 cm^{-1}, 1620 cm^{-1}, 1645 cm^{-1}, 1665 cm^{-1} and 1690 cm^{-1}. Once this has been completed, the spectra at lower temperatures were analysed in the same manner, thus essentially fitting the peaks for dsDNA. Finally, the peaks were fitted at higher temperatures, thus the contribution to the IR spectrum of melting dsDNA from melted, ssDNA could be evaluated. The changes in peak height between the two purine ring vibrations was analysed, and a lineshape resembling a combination of a straight line and a sigmoid was found (Fig. 3.13). The turning point for the sigmoidal section of the curve was observed to be 68 °C. The discrepancy in T_m observed when the technique was changed from UV to IR spectroscopy is due to the fact that a more concentrated sample was used for the IR melting experiments – a fuller discussion of this phenomenon can be found in *Section 3.7*.[7]

Although a sigmoid profile similar to that observed for the UV melting curve was observed and hence melting was observed, it was thought that the melting temperature was too high to allow for fitting of a full sigmoidal curve for a ODN 3.1 complex with a stabilising ligand.

Consequently the decision was taken to use a 10-mer duplex, ODN 3.2 (Fig. 3.14 a) which maintains the A_3T_3 binding site, but the terminal GC base pairs are missing .

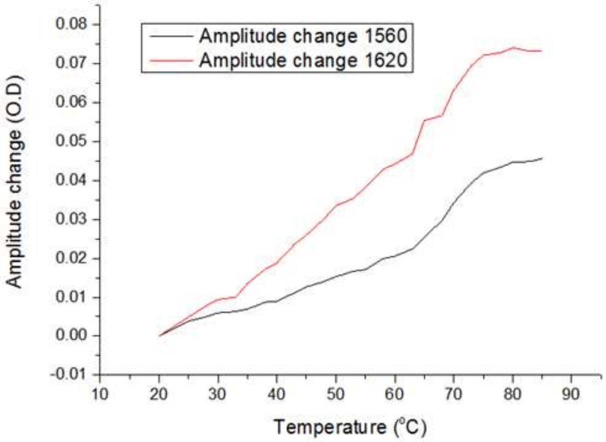


Figure 3.13. Changes in peak amplitude of the guanosine (1560 cm^{-1}) and adenosine (1620 cm^{-1}) purine ring vibrations (Fig. 3.12).

3.3.2 Analysis of the thermal melts of decamer dsDNA by UV and IR specroscopy.

ODN 3.2 exhibited a melting temperature of 35 °C at 1 µM concentration (Fig. 3.14 b).

(a) 5'-GCAAATTTCG-3'
3'-CGTTTAAAGC-3'

(b)

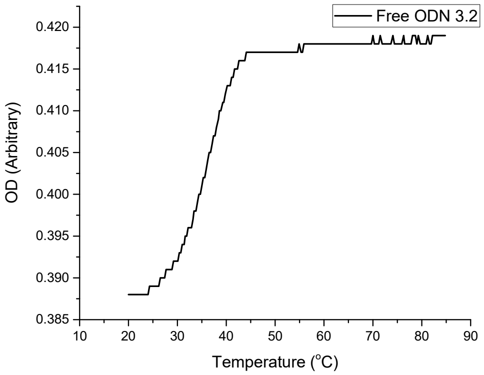


Figure 3.14. (a) Sequence of ODN 3.2. **(b)** UV melting curve of ODN 3.2, T_m = 35 °C. (1 µM concentration)

The IR melts of free ODN 3.2 were carried out at 2.5 mM concentration and the results are shown in Fig. 3.15. A similar melting profile to that observed for ODN 3.1 was found.

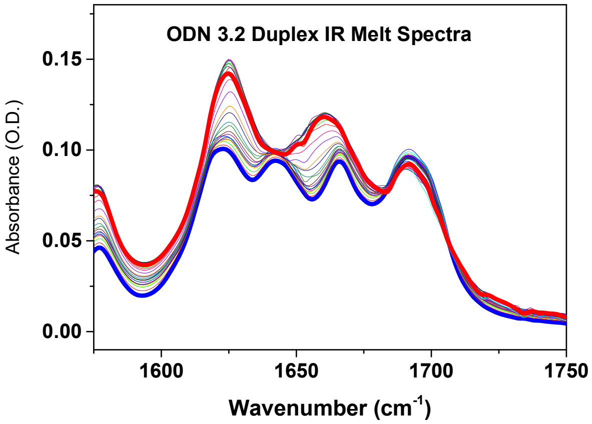


Figure 3.15. Changes in the IR spectrum of free ODN 3.2 as a function of temperature in the region between 1600 and 1700 cm^{-1} The total duplex concentration was 2.5 mM.[118]

The IR spectra of free ODN 3.2 at both 20 °C and 80 °C with are shown in Figure 3.16.

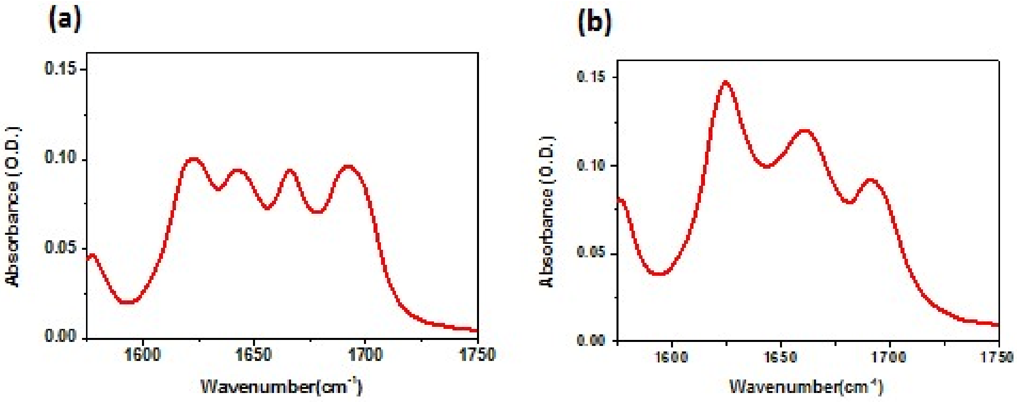


Figure 3.16. (a) The IR spectrum of ODN 3.2 at 20 °C between 1600 and 1700 cm^{-1}. **(b)** The same sample at 80°C.[118]

From inspection of Figure 3.16, changes in the spectrum reflect those expected in Tables 3.1 and 3.2. An increase in the IR response at 1620 cm^{-1} was observed. A decrease in the absorption band of the peak at 1692 cm^{-1} (G_s and T_{2s} stretching vibrations) was also observed signifiying melting of the A-T base pairs of ODN 3.2. The fates of the peaks at 1645 cm^{-1} and

1665 cm^{-1} are more complex due to the redshift of the G_s carbonyl stretching vibration in dsDNA to that observed in ssDNA (Table 3.2). A blue shift was observed for the 1645 cm^{-1} band, coming from breaking of the Watson-Crick base pairing increasing the electron density on the C=C and C=N bonds of the pyrimidine rings of the thymidine and cytosine bases, increasing the bond strength and hence, the absorption frequency. Because of this redshift and the blueshift in the C_s carbonyl stretching vibration observed upon melting, the behaviour of these two peaks is more difficult to understand due to these two peaks coming together upon melting of ODN 3.2. The contributions of these peaks could be deconvoluted through fitting of these vibrations to four Gaussian peaks as per Appendix 7.13. It was found that plotting the changes in peak height from that observed in Fig. 3.16 was the best indicator of dsDNA melting (Fig. 3.17).

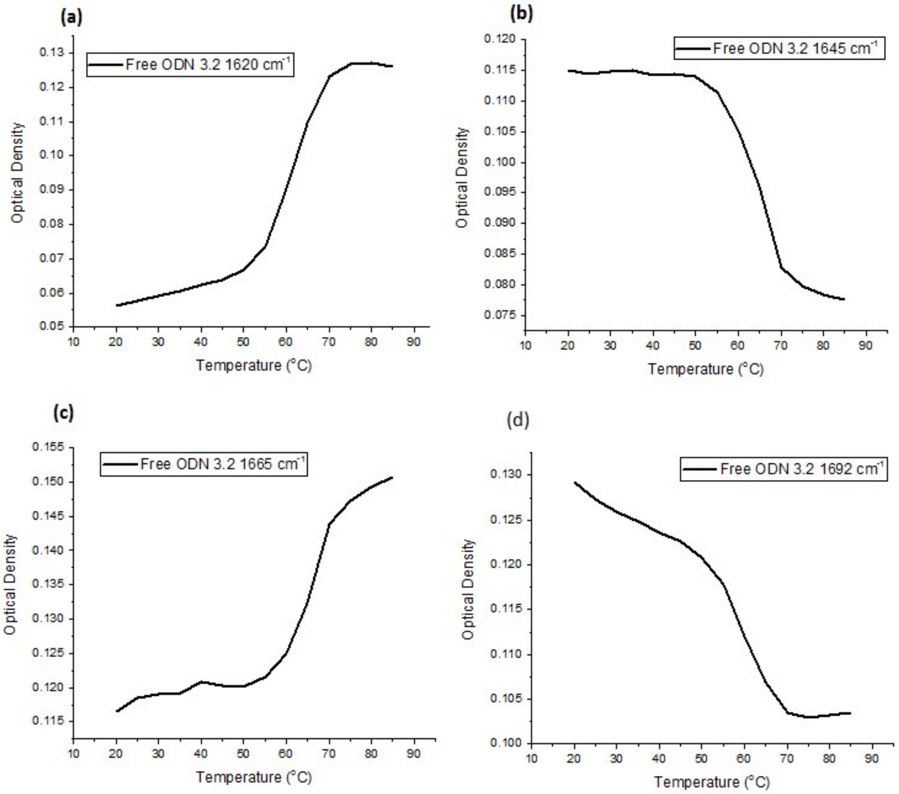


Figure 3.17. Comparison of the changes in peak height of the IR spectrum of free ODN 3.2 as a function of temperature.**(a)** 1620 cm^{-1}, **(b)** 1645 cm^{-1} **(c)** 1665 cm^{-1} **(d)** 1692 cm^{-1}.[118]

The observed melting temperatures of ODN 3.2 are summarised in Table 3.3. The melting temperatures of ODN 3.2 by IR are significantly higher than observed by UV-Vis spectroscopy (58 – 66 °C). This is due to the significantly higher concentration of sample required for a good signal to noise ratio for IR spectroscopy as opposed to UV spectroscopy (2.5 mM for IR, 1 μM for UV). It is well-known that T_m is dependent on both salt and oligomer concentration,[140] and is a consequence of the thermodynamics of association of ssDNA to form dsDNA (*Section 3.7.1*).[141,142]

Fig 3.14 shows that the melting behaviour of free ODN 3.2 is somewhat complex, and the four peaks each have different values of T_m (Table 3.3, Fig. 3.18). As such, to get a single melting temperature from this data set, principal component analysis was carried out on ODN 3.2 (Fig. 3.18). The melting temperature was observed to be 59 °C, 24 °C higher than at 1 μM.

Peak Centre (cm^{-1})	**T_m (°C)**
1620	61
1645	65
1665	66
1692	58

Table 3.3. Melting temperatures of free ODN 3.2, found from differentiation of the sigmoid curves.

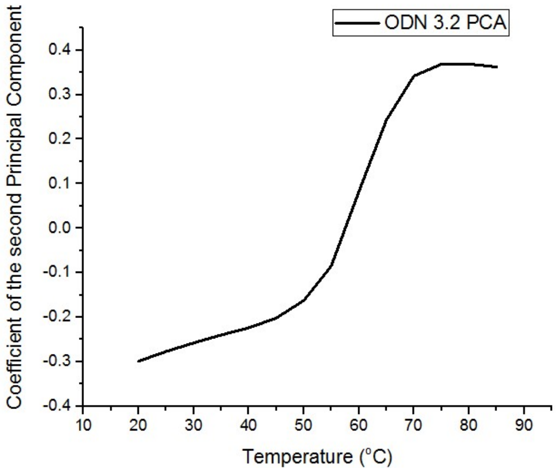


Figure 3.18 Principal Component Analysis of the of the IR melt of ODN 3.2.[118]

From the data in Figures 3.17-3.18, and Tables 3.2 and 3.3 it is tempting to suggest that as the peaks at 1665 and 1645 cm^{-1} in the IR spectrum are dominated by G-C base pair vibrational modes, and those at 1623 and 1692 cm^{-1} are dominated by A-T vibrational modes, that ODN 3.2 melts in the central A-T tract first, i.e. by a "bubbling" mechanism (Fig. 1.35).[143] However, although this may be the case, this data is far from conclusive, as minor contributions from G-C base pairs exist in the peaks at 1623 cm^{-1} and 1692 cm^{-1} and *vice versa*. Other reasons, for example base stacking interactions may be a cause of the varied values of T_m for each band. Time-resolved correlation of these vibrational modes of dsDNA by 2D-IR will unambiguously show the mechanism by which ODN 3.2 melts.

3.3.3: Comparison of the UV melts at high and low concentration.

To prove that the higher concentration of ODN 3.2 in the sample was the root cause of the observed discrapancy in T_m between the UV and IR melting experiments, UV melting experiments were carried out on free ODN 3.2 at concentrations ranging from 1 μM to 2500 μM. The results are shown in Fig. 3.19.

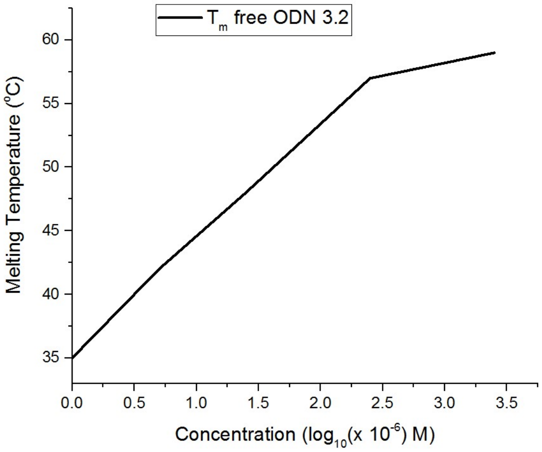


Figure 3.19. Variance in T_m with increasing concentrations of ODN 3.2.[118]

Fig. 3.19 shows that T_m follows a logarithmic distribution, reaching an asymptotic maximum at 59 °C. This finding proves that the discrepancy in the observed T_m between the UV-Vis experiments and the IR melting experiments is solely due to the increased concentration of ODN 3.2 in the sample. It also appears that 59 °C is the maximum melting temperature of free ODN 3.2 – whether the presence of an MGB increases this maximum will be investigated.

3.4. Analysis of the melt of 1.9: ODN 3.2 complex by UV-Vis and IR spectroscopy.

*3.4.1. UV-Vis spectroscopy of the **1.9**: ODN 3.2 complex at 1 μM.*

To show that the binding of azide-labelled compounds **1.20** and **1.21** to A-T tracts of dsDNA are similar to that observed for **1.9**, the melting behaviour of complexes of **1.9** in complex with dsDNA were investigated. The UV-Vis melt of **1.9** in complex with ODN 3.2 and ODN 3.1 are shown in Figs. 3.20 and 3.21 respectively.

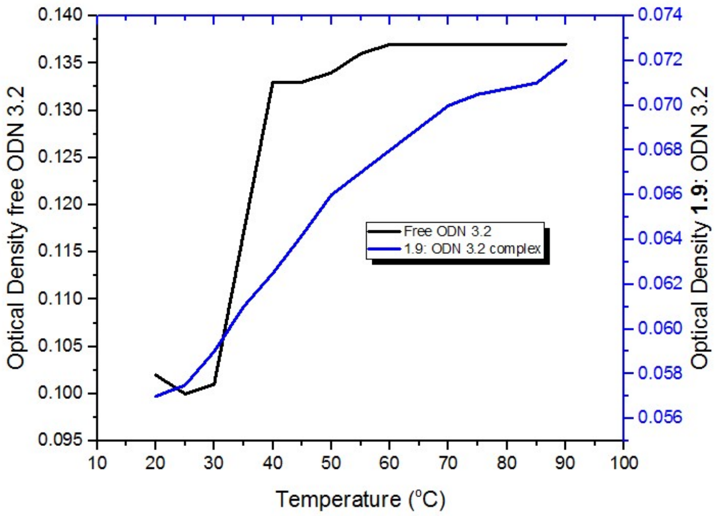


Figure 3.20. UV melting curves at 1 μM of (black) free ODN 3.2, and (blue) the **1.9**: ODN 3.2 complex..[118]

The melting temperature for the **1.9**: ODN 3.2 complex was observed to be 48 °C, presenting a stabilisation (ΔT_m) of 13 °C, indicative of significant stabilisation when **1.9** is in complex with ODN 3.2. A similar ΔT_m was observed when for the complex of **1.9** with ODN 3.1 (Fig. 3.21), the stabilisation was found to be 11 °C (T_m = 62 °C). Despite the lower ΔT_m observed for ODN 3.1 than for ODN 3.2, the similar values for ΔT_m showed that changing from the original dodecamer in ODN 3.1 to the decamer ODN 3.2 had little effect on the binding strength of **3.1**, and hence, provided the target binding site remains the same. This vindicated the decision to change the oligonucleotide from ODN 3.1 and ODN 3.2. The origin of the 2 °C discrepancy in the ΔT_m of the two complexes of **1.9** with ODN 3.1 and ODN 3.2 is likely to be due to photodegradation of the sample of **1.9** used in the experiment with ODN 3.1. **1.9** is known to be photosensitive, and although stock solutions were kept away from light at -20 °C, the sample was likely to have aged somewhat. This was confirmed by a UV-Vis experiment of this sample of **1.9** and a new sample of **1.9** which showed a reduced value of the molar extinction coefficient (data not shown).

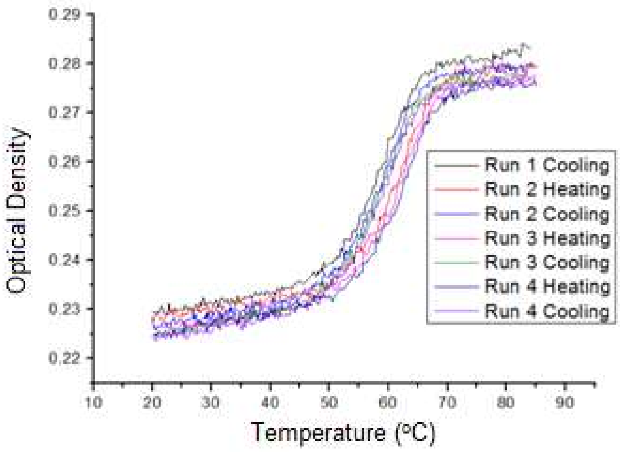


Figure 3.21. UV-Vis melt of the ODN 3.1:**1.9** complex at 1 μM.

3.4.2. High concentration UV-Vis spectroscopic measurements.

Given the observed increase in T_m with increasing concentration of ODN 3.2, the melting behaviour of the **1.9**: ODN 3.2 complex at 2.5 mM concentration was investigated by UV-Vis spectroscopy, in order to obtain a ΔT_m at a concentration necessary for DNA melts by IR spectroscopy to be carried out.

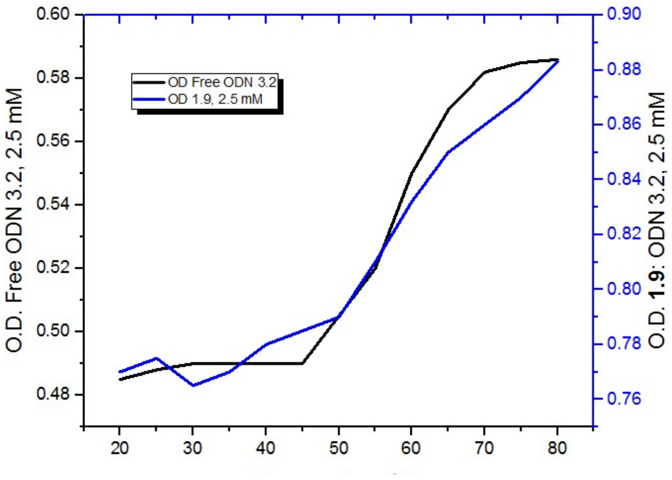


Figure 3.22. Comparison of the melting temperature of free ODN 3.2 (black) and its complex with **1.9** (blue). .[118]

The melting temperature for the **1.9**:ODN 3.2 complex was 59 °C - exactly that found for the free duplex. Extensive structural characterisation of the binding mode of complexes of **1.9** with dsDNA oligomers have shown that a single, high-affinity minor groove binding mode does exist at high concentration,[144,99] and that consequently the lack of observed stabilisation of the **1.9**: ODN 3.2 complex compared to free ODN 3.2 is presumably related to the thermodynamic properties of dsDNA melting observed for free ODN 3.2 (Fig. 3.19). The observed lack of stabilisation observed through the course of these melting experiments is not important for the purposes of this project; all that is sufficient isformation of a well-defined complex between ODN 3.2 and a given ligand as well as the formation of sigmoidal melting curves upon spectroscopic investigation of these complexes, such that they can be investigated fully.

*3.4.3. IR melting experiments on the **1.9**: ODN 3.2 complex.*

Melting experiments were carried out on the 1:1 stoichiometric complex of **1.9** and ODN 3.2 at 2.5 mM using identical conditions to those used for free ODN 3.2 (Fig. 3.23). The four peaks were present as for free ODN 3.2 (*Section 3.3.2*) and the basic shape of the spectrum was identical to free ODN 3.2, indicating that the presence of the ligand does not significantly alter the Watson-Crick base pairing, as expected from structural studies of this complex.[64]

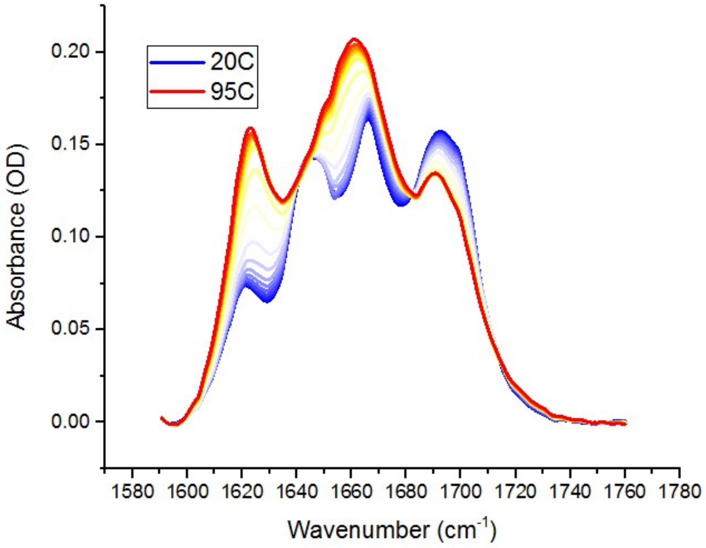


Figure 3.23. The IR spectrum of the **1.9**:ODN 3.2 complex.[118]

Analysis of the changes in peak heights of the complex were also carried out leading to a similar profile for free ODN 3.2, indicating no change in the fundamental melting mechanism (Fig. 3.24). Sigmoidal profiles were obtained for the peaks at 1620, 1645, and 1692 cm^{-1}. The peak at 1665 cm^{-1} exhibited a linear change with increasing temperature, this is due to an overlapping vibration from the phenolic stretching vibration **1.9** which obscured the sigmoid change in the peak, and hence only a linear profile was observed. The melting temperatures of the sigmoid profiles are summarised in Table 3.4.

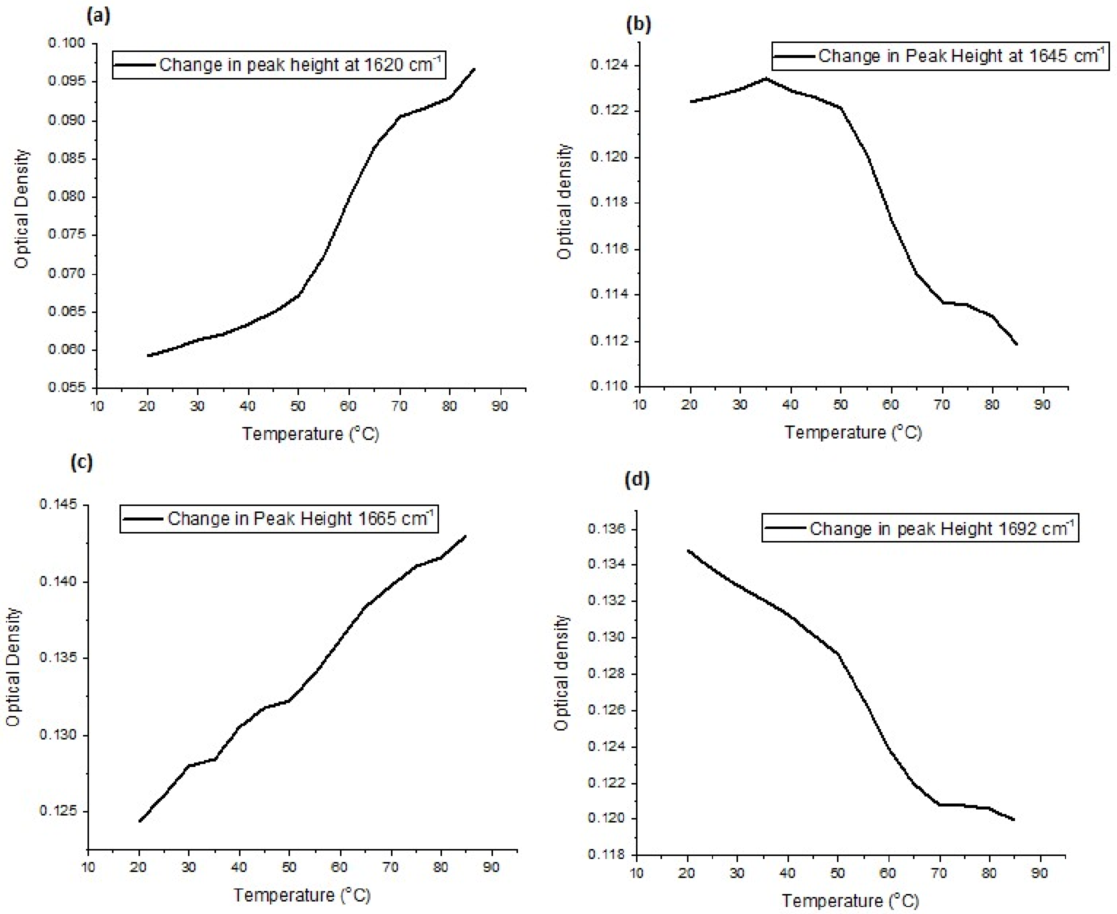


Figure 3.24. Changes in the peak heights of the **1.9**:ODN 3.2 spectrum at **(a)** 1620 cm^{-1} **(b)** 1645 cm^{-1} **(c)** 1665 cm^{-1}, **(d)** 1692 cm^{-1}.[118]

The observed melting temperatures for each of the individual peaks showed somewhat lower melting temperatures than for free ODN 3.2 by 4-5 °C, which seems at first glance a troubling finding, because as shown in Figs. 3.20 – 3.21, **1.9** is a stabilising ligand. However, the analysis of this spectrum was markedly more difficult than for free ODN 3.2, undoubtedly due to the presence of vibrational modes of the ligand, some of which do come underneath the peaks of the DNA spectrum, and could not be easily subtracted, as was the case for solvent peaks.

Peak Centre (cm^{-1})	T$_m$ (°C)
1620	59
1645	59
1665	-
1692	54

Table 3.4. Melting temperatures observed for the **1.9**: ODN 3.2 complex.

To obtain a single value for T$_m$ from the dataset shown in Fig. 3.23, principal component analysis was carried out on the spectrum of the **1.9**:ODN 3.2 complex (Fig. 3.25). When compared to the PCA of the IR melt of free ODN 3.2 (Fig. 3.25), little difference in the melting temperature was observed. This observation proved that the observed discrepancy in the IR melting temperatures of free ODN 3.2 and the **1.9**: ODN 3.2 complex was due to a difficulty in effectively fitting a sigmoid profile to the raw melting curves.

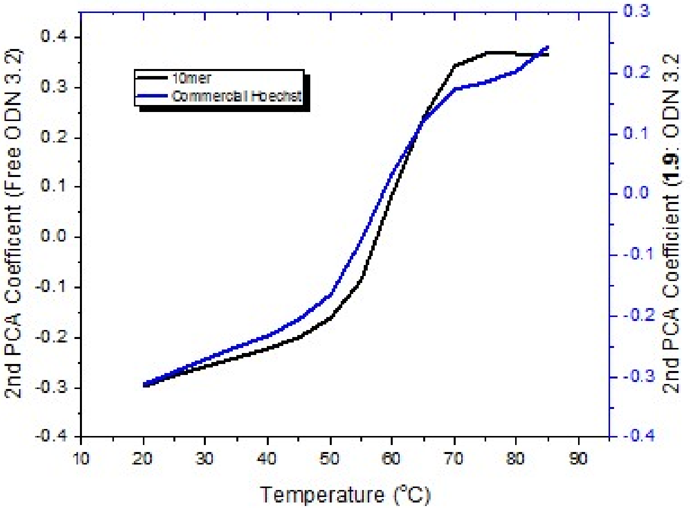


Figure 3.25. Comparison of the PCA of free ODN 3.2 (black) and the **1.9**:ODN 3.2 complex.[118]

Fig 3.25 shows unambiguously that monitoring the changes in peak height brought about by melting works as an indicator of the melting of ODN 3.2 even in the presence of a small molecule ligand. However, it cannot from the DNA region of the spectrum alone, indicate whether the ligand is bound to dsDNA or not, and as such, the mechanism of melting (Fig. 1.35) cannot effectively be determined. Hence, this data set acquired for the **1.9**:dsDNA complex shows the need for the synthesis of IR-labelled derivatives of **1.9**, **1.20** and **1.21** (Fig. 3.4).

3.5. Spectroscopic characterisation of the melting behaviour of the 1.20: ODN 3.2 complex.

*Section 3.5.1. Proof of a minor groove binding mode for **1.20***

To prove that **1.20** exhibited a minor groove binding mode similar to that observed for **1.9** at high concentration, fluorescence experiments were carried out at 2.5 mM. A large increase in the fluorescence of **1.20** was observed, similar to that observed for **1.9** (Fig. 3.26). This is consistent with a minor groove binding mode similar to that observed in Fig. 2.2, different complexes of **1.9** with dsDNA do not exhibit as pronounced an increase in fluorescence as the canonical minor groove mode.[124,53]

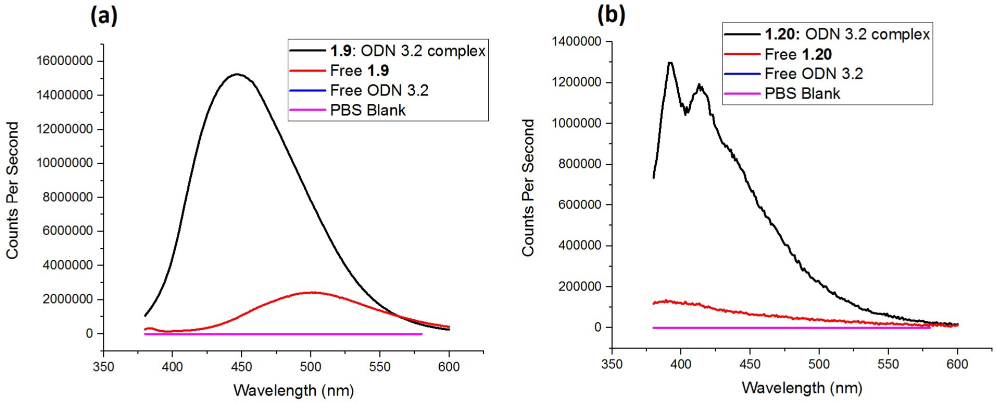


Figure 3.26. (a) Fluorescence spectrum of **1:9** in complex with ODN 3.2 compared to the free compounds. **(b)**. The same spectrum as for the **1.20**: ODN 3.2 complex.

The shapes of the spectra of the **1.9** and **1.20** complexes are very different in nature; this could be due to structural differences between the two complexes (*Chapter 4*). The reason for this difference was not pursued further, however.

Section 3.5.2. Low concentration UV-Vis spectroscopy.

Having achieved the synthesis of **1.20** (*Section 2.6*), the melting behaviour of these complexes was investigated by UV-Vis and IR spectroscopy. The UV-Vis spectrum showed that **1.20** stabilised ODN 3.2 by 9 °C (44 °C for the complex against 35 °C for free ODN 3.2), 2 °C lower than found for **1.9** (Fig. 3.27). This is preseumably due to the loss of the electrostatic interaction between the cationic piperazino group in **1.9** when formally substitued for the azide group in **1.20**. Although there is a primary ammonium group in **1.20**, any electrostatic or hydrogen bonding interactions of this group with ODN 3.2 does not fully compensate for the loss of the tertiary amino- group present in **1.9**. Nonetheless, **1.20** is a stabilising ligand, and can hence be carried forward into the other spectroscopic experiments.

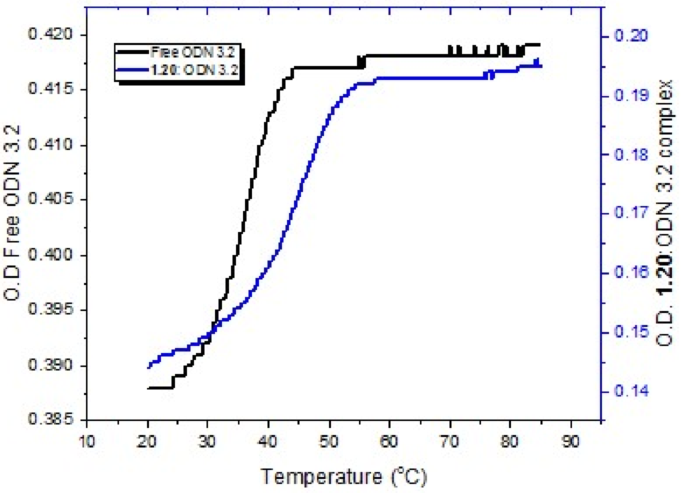


Figure 3.27. UV melting curves of ODN 3.2 (black) and **1.20**: ODN 3.2 complex (blue).

Section 3.5.3. High concentration UV-Vis spectroscopy.

To investigate whether, as in the case of **1.9**, that no stabilisation of ODN 3.2 occurs at high concentration, and also to obtain a value for T_m for the IR experiments, a UV-Vis melting experiment was carried out at 2.5 mM concentration (Fig. 3.28).

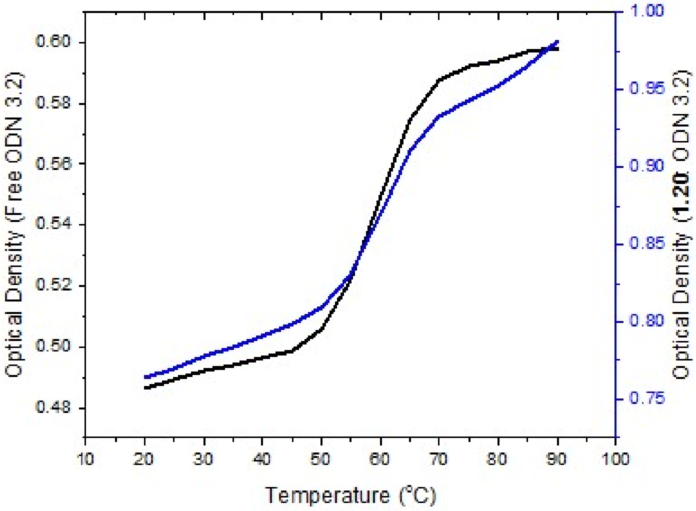


Figure 3.28. Comparison of the melt of the **1.20**: ODN 3.2 complex (blue) with free ODN 3.2 (black). $\Delta T_m = 0$ °C.

As expected, a T_m of 59 °C was found at high concentration, identical to that observed for the **1.9**: ODN 3.2 complex (59 °C). This observation is further evidence that the thermodynamics of ODN 3.2 duplexation, which depend in part on total oligomer concentration, dominate the melting behaviour of the complex of **1.20** in complex with ODN 3.2. To investigate the melting behaviour of this complex, UV-Vis melting experiments were carried out at a range of concentrations from 1 μM to 2500 μM (Fig. 3.29). The differences followed a logarithmic trend with stabilisation only observed at very low concentrations (1-5 μM), and the maximum for both the free oligomer and the complex reached an asymptotic maximum approaching 59 °C. As such, this data provides evidence that the complex of **1.20** with ODN 3.2 exhibits similar melting behaviour to complexes of **1.9** with ODN 3.2, although the smaller ΔT_m at 1 μM pointed towards a weaker association constant for **1.20** compared to **1.9**. Hence, from this data, a similar binding mode for **1.20** to ODN 3.2 can be inferred.

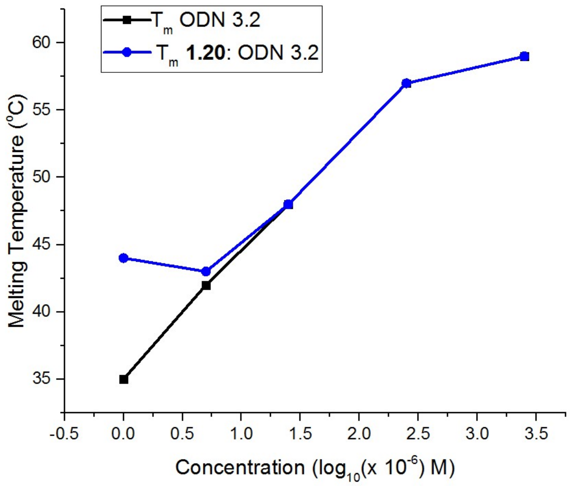


Figure 3.29. Variation in T_m with increasing **1.20**: ODN 3.2 concentration compared to free ODN 3.2. Stabilisation was only observed at concentrations below 25 μM.[118]

Elucidation of the binding mode of **1.20** to ODN 3.2 was found by NMR (Chapter 4) and was shown to bind in the minor groove of an A_3T_3 tract of dsDNA, and hence as with the **1.9**: ODN 3.2 complex, the origins of the lack of obserbaed stabilisation at high concentration is likely to lie in the thermodynamics of dsDNA melting.[142]

*3.5.4. IR melting analysis of the 1600-1700 cm^{-1} region of the **1.20**:ODN 3.2 complex*

Having established that firstly, forms a highly fluorescent complex at high concentration with ODN 3.2 (*Section 3.5.1*), and **1.20** stabilises ODN 3.2 at low concentration (*Section 3.5.2*). pointing towards a minor groove binding mode, attention turned to investigation of the of the complex formed between ODN 3.2 and **1.20** by IR spectroscopy (Fig. 3.30). By the same method described for free ODN 3.2 and its complex with **1.9**, the changes in peak height were modelled as previously (Fig. 3.30). Surprisingly, unlike observed for **1.9**, sigmoidal peaks were found for all four of the four vibrations at 1620, 1645, 1665, and 1692 cm^{-1}. The observed melting temperatures are summarised in Table 3.5.

The data in Table 3.5 shows that there is little difference between the T_m observed for both the IR and UV melts; the T_m data for both the UV and IR melting experiments are both 59 °C.

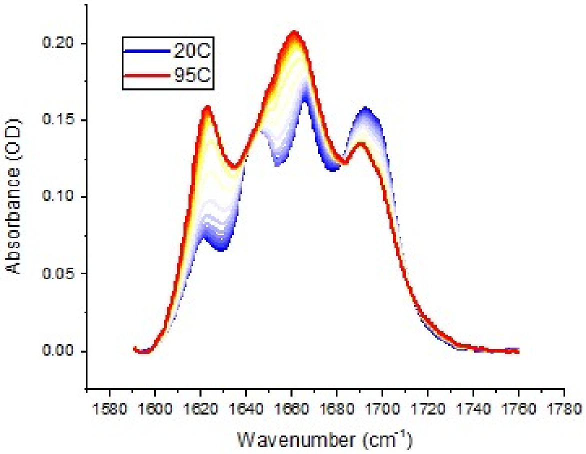


Figure 3.30. Changes in the IR spectrum of the **1.20**: ODN 3.2 complex in the 1600-1700cm^{-1} region as a function of increasing temperature from 20 °C (blue) to 95 °C (red).[118]

Peak Centre (cm^{-1})	T_m (°C)
1620	60
1645	63
1665	67
1692	57

Table 3.5. Observed T_m for the four peaks observed in the **1.20**: ODN 3.2 complex at 2,5 mM.

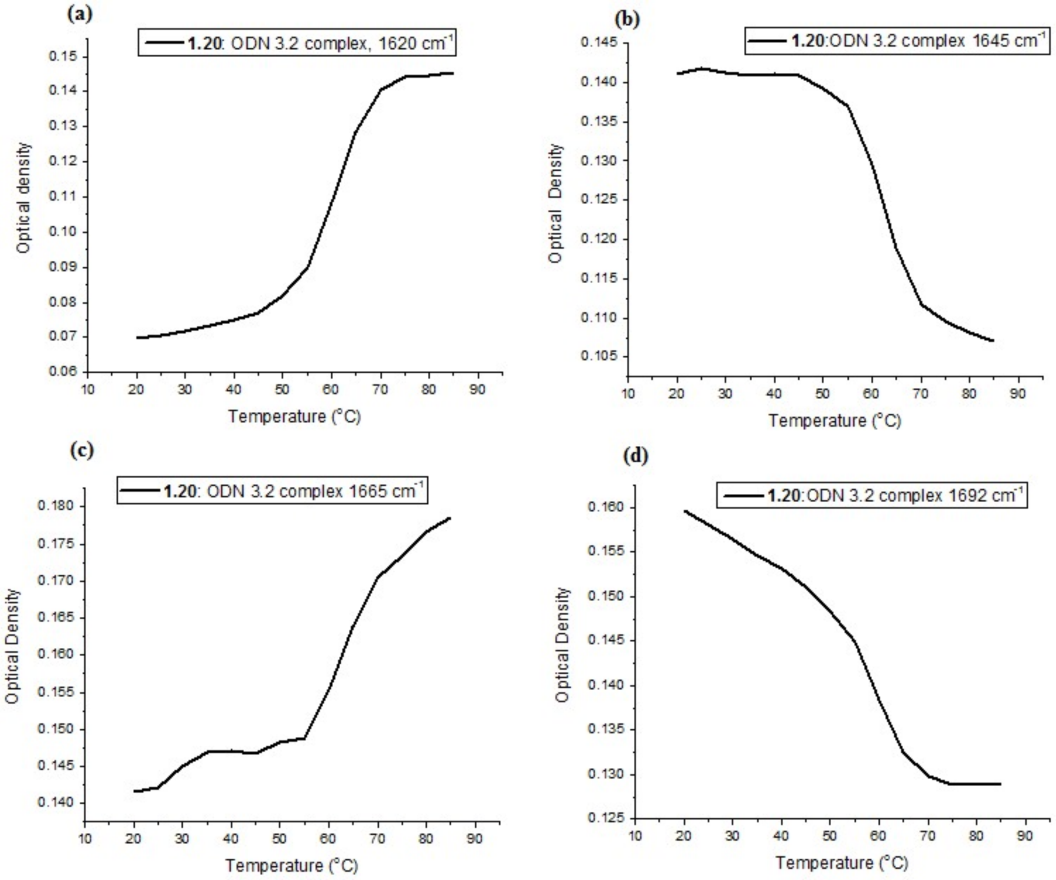


Figure 3.31. Changes in the peak height of the **1.20**:ODN 3.2 complex at **(a)** 1620 cm^{-1} **(b)** 1645 cm^{-1} **(c)** 1665 cm^{-1} **(d)** 1692 cm^{-1}.[118]

In contrast to the IR spectrum observed for the **1.9:** ODN.3.2 complex, a sigmoidal profile could be observed for the peak at 1665 cm^{-1}, presumably due to the subsititution of the phenol moiety in **1.9** for the benzylamine in **1.20.**

To prove the correspondence of the IR melts of the **1.20** complex to that observed for free ODN 3.2, PCA was carried out on the spectrum of the complex (Fig. 3.32). An almost exact correspondence between the PCA of the melting curves of ODN 3.2 and its complex with **1.20** was observed. Hence, the changes in the IR spectrum of the **1.20**: ODN 3.2 complex are sensitive to the decoupling of vibrational modes caused by DNA melting, and although not entirly unaffected by the presence of the ligand, these perturbations of the spectrum do not affect the characteristic shape of the 1600-1700 cm^{-1} region of the DNA IR spectrum.

Consequently, this region of the spectrum can be used as a test for whether the azide probe installed on **1.20** is sensitive to the changes in the solvent environment caused by melting of the **1.20**: ODN 3.2 complex.

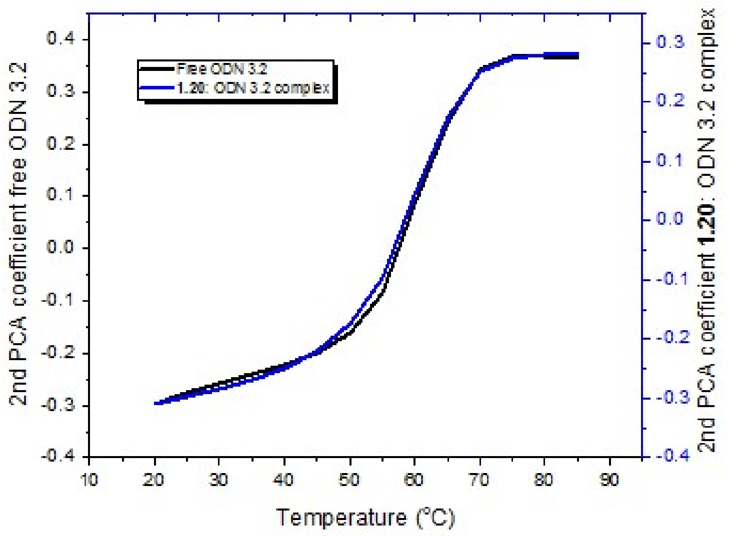


Figure 3.32. PCA of the 1600-1700 cm^{-1} region of the free ODN 3.2 melt (black) and that for the **1.20**: ODN 3.2 complex (blue). The T_m for both was observed to be 59 °C.[118]

*3.5.5 Analysis of the azide region of the **1.20**:ODN 3.2 melt.*

The azide peak of free **1.20** was observed to come at 2098.5 cm^{-1}, a transparent region of the DNA IR spectrum. Hence it is suited to act as a reporter functional group for **1.20** binding to ODN 3.2, provided it is sensitive to melting of ODN 3.2. To test whether the azide in **1.20** acts as a reporter functional group, the changes in the azide region as well as in the 1600-1700 cm^{-1} regions were investigated with increasing temperature. The difference between the spectra at 20 °C and 80 °C was striking (Fig. 3.33) – while the spectrum at 80 °C resembled that of free **1.20**, an extra peak rising at lower frequency to the central peak was observed at 20 °C (Fig 3.33). Moreover, the overall absorbance of the azide absorption band increased almost twofold (Fig. 3.33). To obtain a reasonable signal to noise ratio for the complex, the sample concentration had to be raised from 2.5 mM to 15 mM. However, at this concentration, precipitation of the complex was observed, which was remedied by increasing the ionic strength of the solution from 100 mM to 1 M. No change in T_m was observed upon either changing the concentration or ionic strength of the solution.

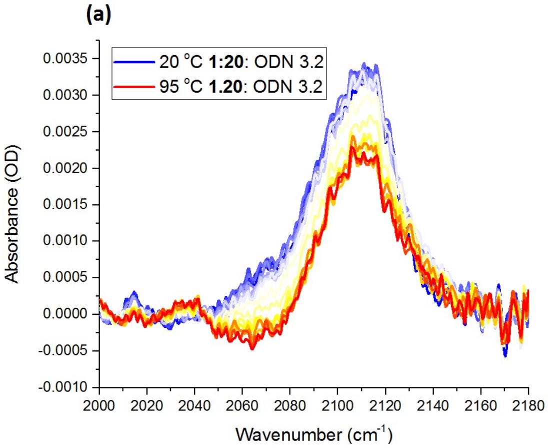


Figure 3.33. **(a)** Changes in the azide IR spectrum of the **1.20**: ODN 3.2 complex as a function of temperature, 15 mM concentration **(b)** Free **1.20** in 4:1 water: DMSO solution, 20 mM concentration.[118]

At intermediate temperatures, two striking changes in the spectrum at around the T_m were observed. Firstly, the IR response was significantly muted due to ligand dissociation from the complex. The increase in solvation of the azide due to breakdown of the complex was the cause of this dampening effect, reducing the observed signal by half. Secondly, as opposed to the free azide in solution which can be fitted to two Gaussian peaks, when **1.20** is in complex with ODN 3.2 a third peak emerges at the low frequency end of the spectrum, at 2069 cm^{-1}. The absorbance of this peak was observed to decrease to zero upon heating to above T_m. Gratifyingly, analysis of the peak heights of the azide region also exhibited sigmoidal changes around T_m, indicating that the azide peak – in the case of the **1.20**: ODN 3.2 complex - is sensitive to dsDNA melting (Fig. 3.34). Principal component analysis of the central azide signal at 2097 cm^{-1} also supports this observation, exibiting a sigmoidal change at 60 °C, very similar to that observed for the 1600 – 1700 cm^{-1} region of the spectrum (Fig. 3.35). Consequently, from this data, the azide region of the **1.20**: ODN 3.2 complex is sensitive to melting of the complex, and the azide functional group does act as a reporter functional group of melting of the **1.20**:ODN 3.2 complex, and can be used in future 2D-IR experiments.

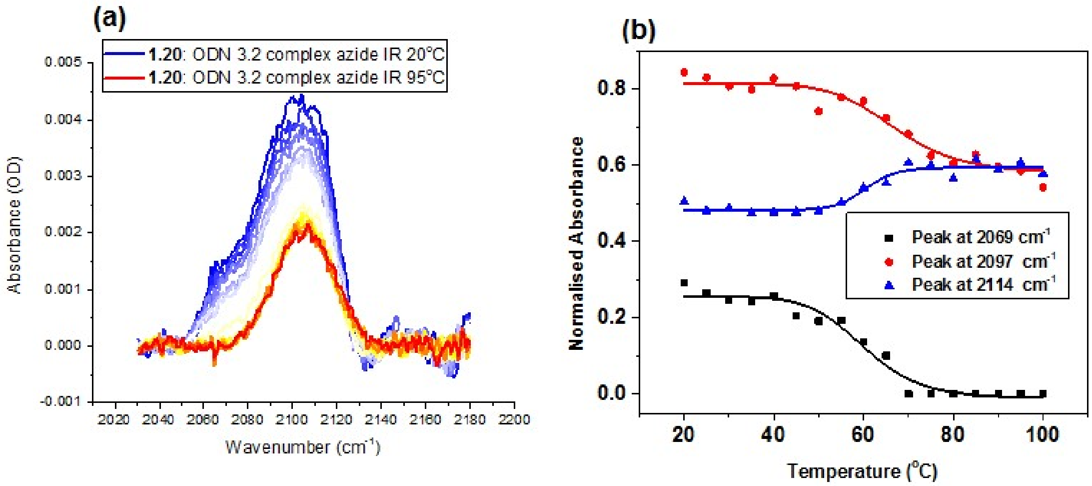


Figure 3.34 (a) Changes in the azide region of the IR spectrum. **(b)** Observed changes in the three components of the azide peak as a function of temperature.[118] To aid in the visualisation of the spectra, the absorbances in **(b)** have been normalised.

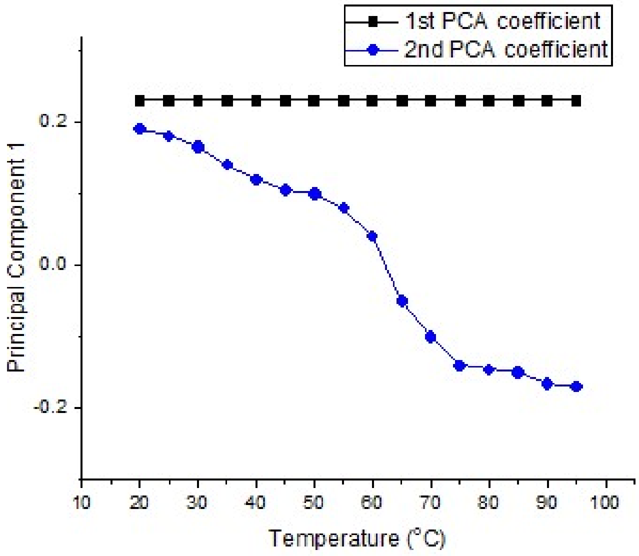


Figure 3.35. PCA of the changes in the azide peak in Fig. 3.34 (b).[118]

3.6 Analysis of the melt of the 1.21: ODN 3.2 melt by UV and IR spectroscopy.

3.6.1. UV-Vis melting experiments at 1 µM concentration.

Although there existed significant problems with the synthesis and purification of **1.21** (*Section 2.5*) small amounts of the compound were synthesised suitable for use in melting experiments in complex with ODN 3.2. The T_m of this complex was 51 °C and ΔT_m, hence was16 °C at 1 µM concentration (Fig. 3.36). The higher value of ΔT_m can be explained by firstly the tertiary amine in **1.21** interacting with the phosphate backbone and the amide functional group forming a bifurcated hydrogen bond across an A-T base pair, similarly to distamycin.[66,94]

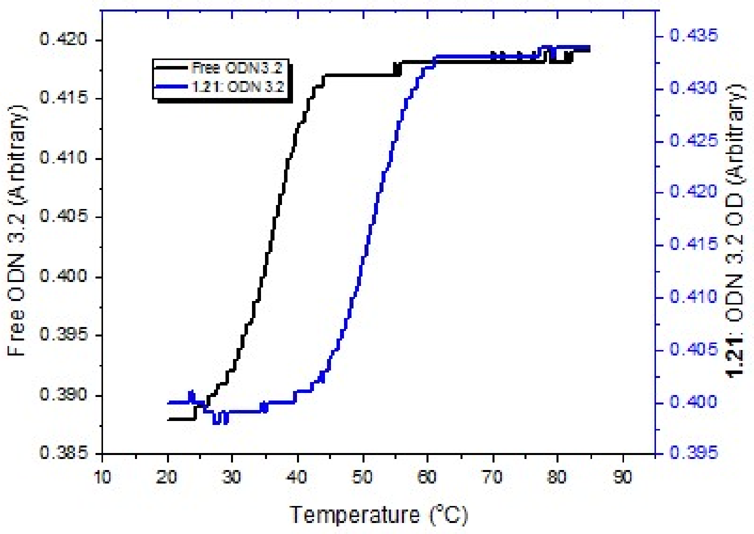


Figure 3.36. UV melt of the **1.21**: ODN 3.2 complex (blue) at 1 µM concentration compared to free ODN 3.2.

*3.6.2. UV melting experiments on the **1.21**: ODN 3.2 complex at 2.5 mM.*

As observed for both compounds **1.9** and **1.20**, there was no observed stabilisation of ODN 3.2 in complex with **1.21**, at 2.5 mM concentration. The observed T_m of the **1.21**: ODN 3.2 complex was 59 °C (Fig. 3.37). Hence, at high complex concentration, none of the three ligands tested exhibited any thermal stabilisation relative to free ODN 3.2, and hence the IR melting curves would also be expected to not exhibit any stabilisation of the duplex. The similarity between the UV melting curves for the three ligands would point toward a similar binding mode for the three ligands, although it is certainly not definitive evidence.

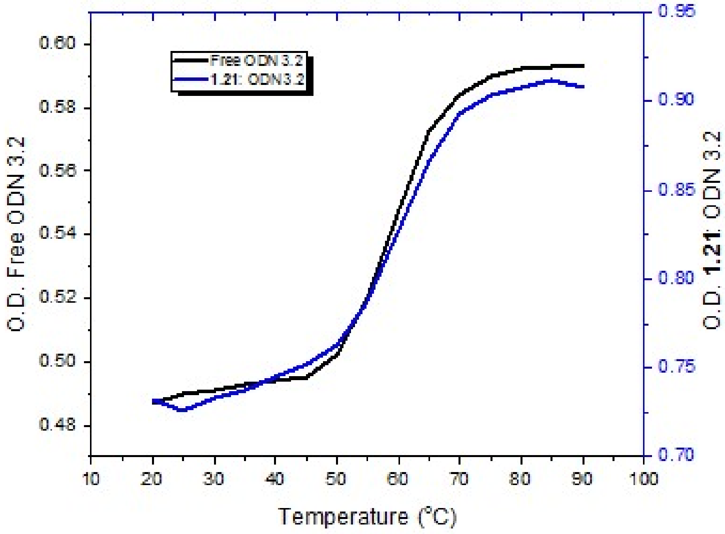


Figure 3.37. UV-Vis melting curves of both free ODN 3.2 (black) and the **1.21**: ODN 3.2 complex at 2.5 mM. [118]

*3.6.3. Analysis of the 1600-1700 cm^{-1} region of the IR spectrum of the **1.21**:ODN 3.2 complex.*

The IR spectrum of the **1.21**: ODN 3.2 complex was monitored with increasing temperature, and the four peaks at 1620, 1645, 1665 and 1692 cm^{-1} were monitored (Fig. 3.38). The changes in peak heights were fitted, and the changes in peak height were reported (Fig. 3.39).

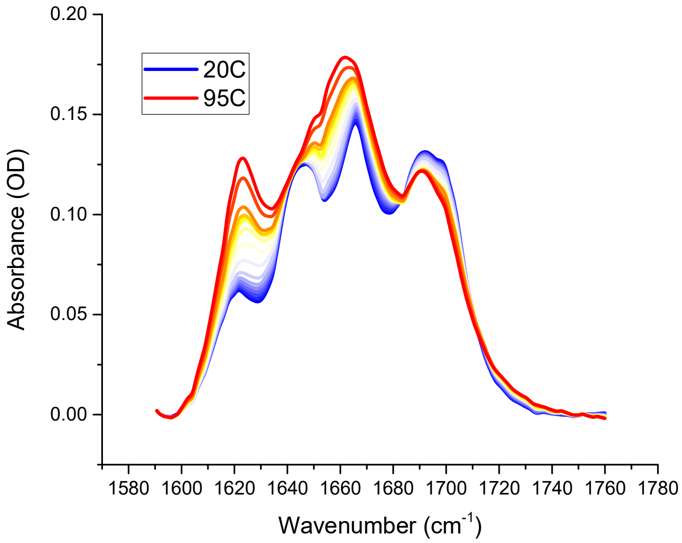


Figure 3.38. IR spectrum of the **1.21**: ODN 3.2 complex between 20 °C (blue) and 95 °C (red).[118]

Fig. 3.39 shows changes consistent with that observed for free ODN 3.2, the **1.9**: ODN 3.2 complex and the **1.20**: ODN 3.2 complex. The changes in peak height exhibit sigmoidal behaviour, similar to those observed for **1.9** and **1.20** (Fig. 3.39). The melting temperatures observed in Fig. 3.41 and summarised in Table 3.6 are also not significantly different from those observed for **1.9**, **1.20** and **1.21**. This observation showed that the presence of a Hoechst-type ligand, despite the structual diversity of the end groups of **1.9**, **1.20** and **1.21**, does not affect the melting behaviour of ODN 3.2 at high concentrations, only the melting temperature at low complex concentration.

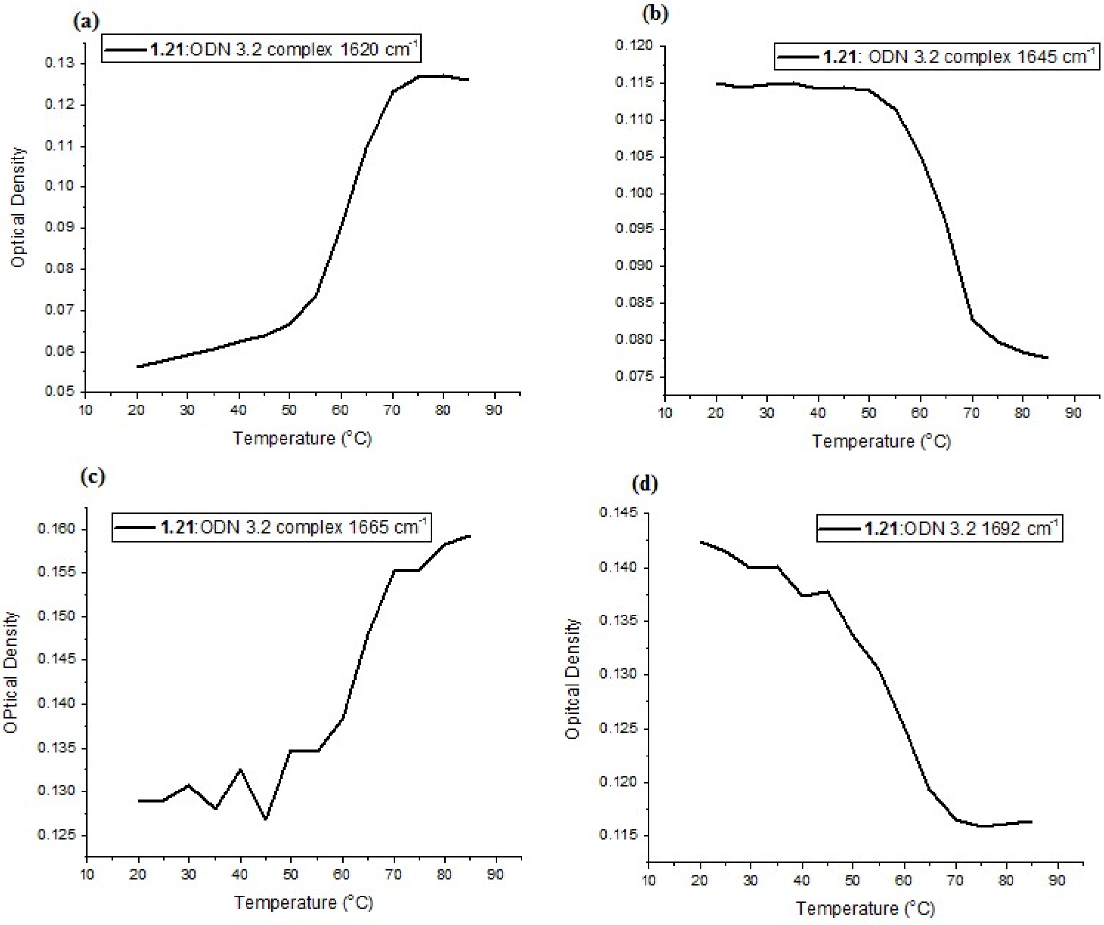


Figure 3.39. Changes in peak height of the **1.21**: ODN 3.2 complex with increasing temperature at **(a)** 1620 cm^{-1}, **(b)** 1645 cm^{-1}, **(c)** 1665 cm^{-1}, **(d)** 1692 cm^{-1}. Melting temperatures are listed in Table 3.6.[118]

Peak Centre (cm^{-1})	T$_m$ (°C)
1620	61
1645	64
1665	66
1692	57

Table 3.6. Observed T$_m$ for the **1.21**: ODN 3.2 complex at the four frequencies identified.[118]

To serve as a direct comparison to the UV-Vis melts, PCA was carried out on the IR melting dataset of the **1.21**: ODN 3.2 complex to furnish a single melting curve (Fig. 3.40). As expected, the observed T_m was identical (59 °C).

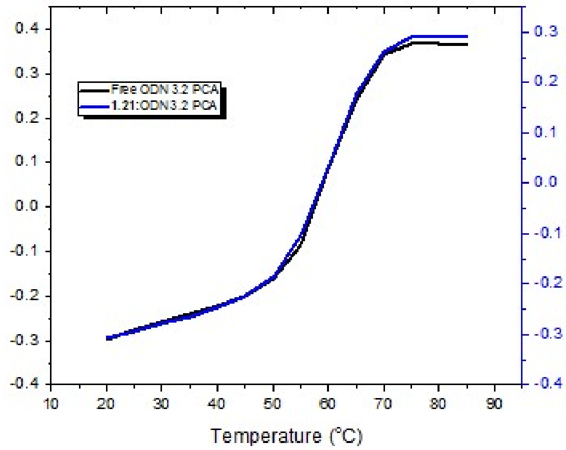


Figure 3.40. Principal component analysis of free ODN 3.2 (black) and the **1.21**:ODN 3.2 complex (blue).[118]

*3.6.4. Analysis of the Azide region of the **1.21**:ODN 3.2 complex*

The changes in the azide response of the **1.21**: ODN 3.2 complex with temperature were monitored by IR spectroscopy (Fig. 3.41). Narrowing of the azide peak of **1.21** upon complexation with ODN 3.2 was observed. The shape of the peak for the **1.21** complex was noticeably different to that observed for the **1.20** peak. Unlike the **1.20** complex, the **1.21** complex exhibited no shoulder peaks in the complex compared to free solution. Upon heating of the **1.21**:ODN 3.2 complex past the T_m, little change in the shape of the spectrum was observed (Fig. 3.42), except for broadening of the peak at high temperatures.

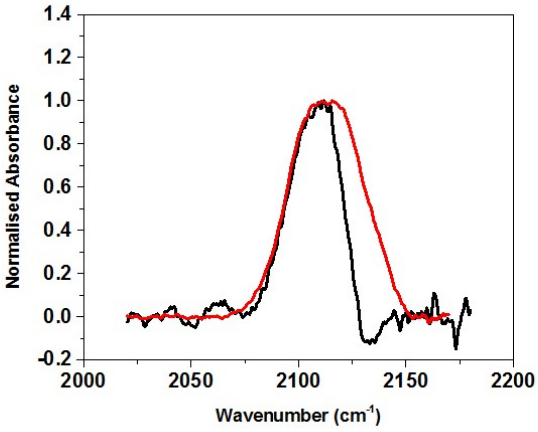


Figure 3.41. Comparison between the azide spectra of free **1.21** (red) and the **1.21**: ODN 3.2 complex (black) Absorbances have been normalised.[118]

With increasing temperature, broadening of the azide peak was observed consistent with a reversion in peak shape to that observed for the free azide spectrum (Fig. 3.42). However, even at 95 °C the azide response was not equivalent to that observed for free **1.21**, indicating that some interaction between **1.21** and ssDNA remains. Changes in the azide spectra could be modelled by fitting the complex spectrum at 20 °C to one Gaussian peak, and modelling the change in the spectrum as a second peak growing in as a blue shoulder (Fig. 3.43). The growth of this second peak, unlike the azide in **1.20**, was found to grow linearly with temperature.

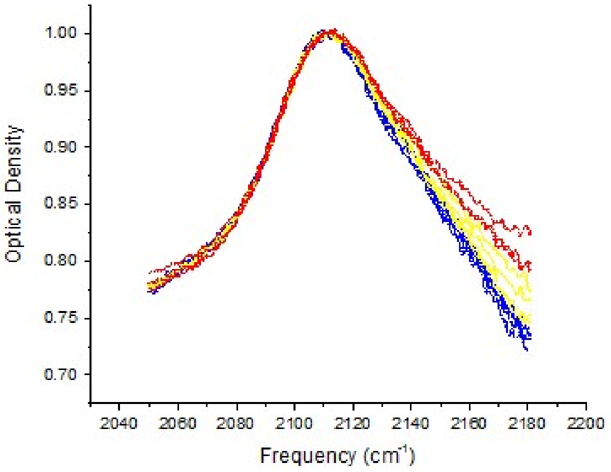


Figure 3.42. Changes in the peak shape of the azide spectrum with increasing temperature. Absorbances have been normalised [118]

This would suggest that the azide in **1.21** is in a completely different chemical environment to that experience by the azide in **1.20.** The data also suggests that the azide in **1.21** may possibly be interacting with melted dsDNA in a way that **1.20** is not.

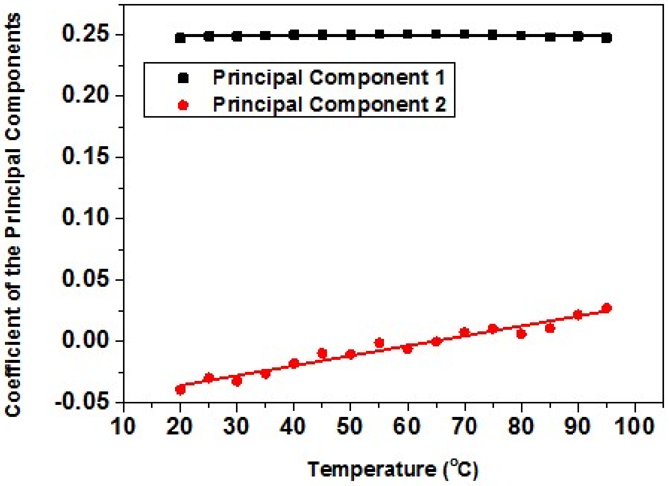


Figure 3.43. Changes in the azide absorbance band of the **1.21**:ODN 3.2 complex with increasing temperature, analysed by PCA.[118]

The linear progression of the PCA of the azide region of the **1.21**: ODN 3.2 complex suggests either a different binding environment, or alternatively the azide in **1.21** is not sensitive to ODN 3.2 melting. Without structural characterisation of this complex, it is impossible to determine the origins of the very different behaviour of the two azides of **1.20** and **1.21.**

3.7. Discussion

*3.7.1. Thermodynamics of melting of ODN 3.2 and the complexes with **1.9**, **1.20** and **1.21***

The thermodynamic parameters governing dsDNA formation can be rationalised through calculation of the thermodynamic parameters governing dsDNA formation from equation 3.1 (see Appendix 7.14 for derivation).[142,145]

$$\Delta H^{\circ} = T_m\left[\Delta S^{\circ} - R\ln\left(\frac{0.5}{0.25\left(\frac{C_T}{2}\right)}\right)\right] \qquad 3.1$$

From the melting temperatures of free ODN 3.2 at 5 μM and 1 μM (315 K and 308 K respectively), the enthalpy and entropy changes associated with duplexation of two strands of ssDNA can be calculated, and upon substitution of the values of C_T and T_m found, the value of $\Delta H°$ was found to be -193.9 $kJ.mol^{-1}$, and the value of $\Delta S°$ was found to be -516.2 $J.K^{-1}.mol^{-1}$, typical values for a decamer duplex.[142] The signs of $\Delta H°$ and $\Delta S°$ are consistent with an entropically-driven melting process, in that duplex formation becomes less spontaneous as temperature increases, giving rise to the sigmoidal curves seen for the melting of both free ODN 3.2 and the three complexes discussed in this chapter.[142] Back substitution into the observed values for T_m at higher concentrations cause the relationship in 3.1 to break down, consistent with the logarithmic distribution observed for T_m at higher concentrations. Although the figures obtained for free ODN 3.2 serve as an indication of the thermodynamics of duplexation, a more accurate ITC experiment will be required in order to obtain a fuller understanding of the thermodynamic forces driving duplex melting.

In the case of the complexes of ODN 3.2 with the compounds **1.9**, **1.20**, and **1.21** the degree of stabilisation is dependent on both entropic (due to removal of bound water) and enthalpic (formation of hydrogen bonds between dsDNA and the ligand) contributions to ΔT_m (*Section 3.1*). The thermodynamics of association of **1.9** to and A_3T_3 tract of dsDNA have previously been investigated, and the value of $\Delta G°$ was found to be - 9.94 ± 0.06 $kcal.mol^{-1}$ at 25°C, and binding was found to be endothermic (+ 4.3 ± 0.01 $kcal.mol^{-1}$),[53] indicating that the hydrogen bonding interactions of the benzimidazole moieties and the electrostatic interaction of the piperazino group to the backbone are more than offset by the energy required to remove bound water molecules from the binding surface. This removal of water from the interface also gives rise to the large entropy change obsered upon binding of **1.9** to cognate sequences of dsDNA ($\Delta S° = 40 \pm 2.0$ $cal.mol^{-1}$). This positive entropy change is the cause of the difference in ΔT_m observed.

In the case of **1.20** and **1.21**, the presence of amide bonds, which as in the case of **1.8**, are known to form strong hydrogen bonds with the minor groove of dsDNA[48] may increase the enthalpic contribution to formation of its complex with ODN 3.2 such that formation of the complex becomes enthalpically and entropically favourable. The presence of the amide may explain why binding of **1.20**, despite the loss of the tertairy amine cation from **1.9** still exhibits significant ΔT_m at low oligomer concentration, as well as the increase in observed ΔT_m for the **1.21**: ODN 3.2 complex when compared to **1.9**. As with free ODN 3.2, ITC experiments are required to deconvolute the contributions of enthalpy and entropy to the binding free energy of the complexes of **1.20** and **1.21** to ODN 3.2

3.7.2. Comparative analysis of the UV-Vis and IR melting results.

As shown in the relevant sections, an almost perfect corespondence between the observed T_m by UV-Vis spectroscopy and the 1600-1700 cm^{-1} region of the IR spectrum was observed. This shows that the origins of the changes in the IR spectrum are similar to the origins of hypochromicity of the UV spectrum of dsDNA: Watson-Crick base pairing. Hence, the sigmoidal shapes of the curves accurately reflect the relative populations of melted and duplexed DNA of the sample at a given temperature. Unfortunately, however, due to the lower extinction coefficent of a molecular vibration compared to that observed for UV excitiation of ODN 3.2, higher sample concentrations are necessary. Consequently, studying the effects of the small molecule ligands **1.9**, **1.20** and **1.21** on dsDNA melting by IR spectroscopy proved impossible as no stabilisation effect was observed at the concentration required for an adequate signal to noise ratio. Consequently, 2D-IR is required in order to draw conclusions about the mechanism of melting.

Given that IR spectroscopy has greater structural resolution than either UV-Vis or fluorescence spectroscopy, it is a reasonable assumption to make that hypothesis for the mechanism of melting can be drawn from the IR data set. Unfortunately, given the similarity of the functional groups – and hence, the vibrational modes - of the bases no one band in the 1600 – 1700 cm^{-1} region is represented by solely G-C or A-T base pairs (Fig. 3.5, Tables 3.1 and 3.2).[137] Although the band at 1620 cm^{-1} is almost solely composed of the A_{r1}-T ring mode, a small contribution from the analogous vibration G_sC_r is also present. The observed IR spectrum of ODN 3.2 can be thought of as a hybrid of the calculated spectra for GC and AT tracts of B-DNA (Fig. 3.13).[137] Hence, the global conformation of the complexes of **1.9**, **1.20** and **1.21** with ODN 3.2 is unmistakably B-DNA.

Consequently, in the 1600 – 1700 cm^{-1} region of the IR spectrum, there exist no unique markers of each base pair, and hence 2D-IR is required to probe the relative populations of coupled and decoupled vibrational modes to understand the mechanism of melting. What this data has shown is that due to the identical structural feature of dsDNA directly measured by both UV and IR spectroscopy (Watson-Crick base pairing), the value of T_m obtained by these complexes corresponded exactly. Unfortunately however, due to the aforementioned signal to noise ratio issues, investigation of the three ODN 3.2: **1.9**, **1.20** and **1.21** complexes at a concentration in which stabilisation was observed by UV-Vis spectroscopy proved impossible. Again, 2D-IR spectroscopy will be necessary to understand the dynamics of the process of dissociation of the ligand upon duplex melting.

*3.7.3. Comparative analysis of the azide responses of the **1.20**:ODN 3.2 and **1.21**:ODN 3.2 complexes.*

Unlike the 1600 – 1700 cm^{-1} region of the IR spectrum of the two complexes, the azides in **1.20** and **1.21** exhibited significantly different behaviour. Given that the 1600 – 1700 cm^{-1}

region of the spectrum exhibited almost identical behaviour at 2.5 mM this finding was surprising. The azide in the **1.20**: ODN 3.2 complex is likely to penetrate deeply into the minor groove, evidenced by the presence of the new shoulder peak when **1.20** is in complex with ODN 3.2 compared to either free **1.20** or when its complex with ODN 3.2 is fully melted (Fig. 3.34). As such, it can be hypothesised that this azide when bound is deeply buried in the minor groove, there exists a specific interaction with one of the functional groups located in the floor of the minor groove of ODN 3.2, presumably with the guanosine amine group of the GC base pair immediately adjacent to the A_3T_3 tract of ODN 3.2. A structural study is required to confirm this (Chapter 4). On the other hand, the azide of **1.21** does not exhibit the same behaviour. Although significant narrowing of the absorption band was observed upon binding of **1.21** to ODN 3.2. indicating that the azide in the **1.21**: ODN 3.2 complex is in a significantly different solvent environment to that observed for **1.20,** consistent with the placement of the azide functional group in a different part of the Hoechst core. The azide of **1.21** is likely to be buried within minor groove like **1.20**, although in a completely different environment. This would explain the lack of observed specific interaction as for **1.20**. More likely, the narrowing of the azide absorption band is due to lower hydration of the azide of **1.21** in the poorly hydrated A-tract of ODN 3.2. Hence, this data shows that judicious placement of the azide functional group allows for investigation of a localised region of a given solvent environment of dsDNA oligomer by 2D-IR, provided the structure of the complex is known.

The sigmoidal profiles observed for the changes in the azide absorption band for the **1.20**: ODN 3.2 complex provides further evidence for a specific interaction with the minor groove. A close correlation between the T_m found for the UV-Vis, and the IR regions investigated of the **1.20**: ODN 3.2 complex shows that the azide in **1.20** is sensitive to dsDNA melting, specifically the breakdown of the double-helical structure. In contrast, **1.21** exhibits a linear profile, and as such is much less sensitive to melting of the DNA duplex. This suggests the azide is not buried in the minor groove as in **1.20**, and the azide absorption band is solely responsive to the hydration of the azide group. This may be a worthwhile finding in itself, however due to the difficulties in the preparation of pure **1.21** (*Section 2.6*) the differences in the azide response could not be investigated further. Unlike for **1.20**, the structure of the complex of **1.21** with an A_3T_3 tract of dsDNA has not been investigated over the course of this project. A structure of this complex will have to be obtained in order to understand the different environments between the two azides in complex with dsDNA.

3.8. Summary

The thermal decomposition of complexes of two azide-bearing derivatives of **1.9**, the synthesis of which is discussed in Chapter 2, have been investigated by UV-Vis and IR spectroscopy, and compared to **1.9**. All three compounds significantly increased the T_m of the dsDNA oligomer at 1 μM concentration, and although evidence was found for complex formation at higher concentrations, no stabilisation was observed.

Melting profiles for the four complexes were investigated by FT-IR spectroscopy, with one exception due to overlap from an absorption band from **1.9,** showed sigmoidal profiles with T_m close to that found by UV-Vis spectroscopy. When the IR spectrum was processed by PCA, an almost exact correlation between the IR and UV-Vis melts was also found, indicating that the 1600-1700 cm^{-1} region of dsDNA serves as a good indicator of the relative populations of melted and duplex DNA in ODN 3.2, which bodes well for future 2D-IR expeeriments.

Analysis of the azide absorption band of the **1.20** and **1.21**:

ODN 3.2 complexes yielded different results for the two complexes. The azide of **1.20** is buried deep into the minor groove of dsDNA and interacts with the functional groups of dsDNA at the floor of the groove, evidenced by the presence of an extra band in the IR spectrum. This extra band was observed to disappear upon melting of the duplex, and hence in the case of **1.20**, the azide peak serves as a good reporter functional group for the melting of **1.20**: ODN 3.2 complexes. In case of **1.21** however, the azide served as quite a poor probe functional group as a linear, not sigmoidal change in the spectrum was observed, although a significant change in the shape of the absorption band of the azide response of **1.21** was observed in complex with ODN 3.2. Furthermore, the azide response at 95 °C was similar, but not identical to that of free **1.21**. Further investigation of this observation is required.

3.8.1. Future work.

Future work will focus on using another method for analysis of UV melting curves. The UV-Vis melting profiles of complexes of ODN 3.2 and either **1.20** or **1.21** can be expressed as a fraction of melted dsDNA instead of absorbance, and from this curve, expressions for equilibrium constant can then be derived in a manner similar to Appendix 7.14.[145] From the equilibrium constant, values for ΔH° and ΔS° can then be calculated through construction of a van t'Hoff plot.[145] Given that the process of melting is trimolecular, the expression for the equilibrium constant for the complexes of ODN 3.2 with **1.20** and **1.21** can be expressed in a similar manner (Eq. 3.2).[145] Given an equilibrium constant can be derived, it follows that firstly the thermodynamic parameters governing formation of the complex can be calculated using a van t'Hoff plot, and secondly, through comparative analysis of the thermodynamics of association of free ODN 3.2 with either **1.20** or **1.21** can be obtained. The advantage of this process over the one used in this thesis is that the thermodynamic parameters can be derived from a single concentration, thus avoiding the problems encountered due to lack of stabilisation when incresing the concentration of the complexes (Fig 3.29).

2D-IR investigations of the **1.20**: ODN 3.2 complex will be carried out in order to understand the dynamics of melting of this duplex, asthis work has shown that the azide in **1.20** is a suitable reporter functional group for such experiments. Further investigation of the **1.21** complex is required, as the azide exhibited different melting behaviour as the temperature increased – although this is contingent on the development of a more efficient synthesis and purification protocol for **1.21**.

Structural characterisation of the **1.20**: ODN 3.2 complex is now necessary and was carried out by NMR (Chapter 4). To understand the structural basis of the differences between the two azide responses in the **1.20** and **1.21**: ODN 3.2 complexes, structural studies on the **1.21**: ODN 3.2 complex will have to be carried out. Unfortunately, the problems with the purification of **1.21** have precluded this experiment from being carried out in the course of this project.

3.9. Experimental

3.9.1 Preparation of free dsDNA samples:

All DNA oligomers were purchased purified from Eurogentec and used without any further purification, sequences used are indicated in the main text. Stock solutions of low concentration (< 1 mM) were prepared by dissolution of the ssDNA in water and the concentration determined from the UV absorbance at 260 nm. Preparation of spectroscopic samples were carried out through pipetting an appropriate volume of the stock solutions of each into a 1.5 mL Eppendorf centrifuge tube, followed by addition 200 μL of a 500 mM stock solution of pH 7.0 phosphate buffer for a final phosphate concentration of 20mM and NaCl concentration of 100 mM. The mixture was then diluted to a final volume of 1 mL with Millipore water.

For higher concentration samples, solid samples of ssDNA were added to a 1.5 mL Eppendorf centrifuge tube; phosphate buffer was added to the mixture and diluted to an appropriate volume as for the lower concentration samples.

For the IR experiments, the same procedure was carried out as above, but the samples were prepared in D_2O buffer, as the base vibrations of interest lay underneath the water bending vibration. For the azide absorption band measurements, the buffer concentration was increased to 1M NaCl, 100 mM NaH_2PO_4 and 100 mM Na_2HPO_4 through direct addition of buffer to a weighed sample of DNA and appropriate ligand.

*3.9.2. Preparation of samples of complexes of dsDNA with compounds **1.9, 1.20, 1.21***

Compound **1.9** was purchased from Sigma-Aldrich and used without further purification, and a molar extinction coefficient of 42000 $M^{-1}cm^{-1}$ was used to calculate concentrations of solution. Compounds **1.20** and **1.21** were prepared as described in Chapter 2 and molar extinction coefficients for solutions were used as found. Stock solutions of the ligands **1.20** and **1.21** were prepared in DMSO to a concentration of 50 mM. Samples of dsDNA were prepared as Section 3.9.1. An appropriate volume of the ligand was added to the sample such that a complex of 1:1 stoichiometry was obtained. The sample was then annealed by heating to 80 °C and then cooling to room temperature before measurement.

3.9.3. Spectroscopic Data collection and processing.

UV-Vis spectra were recorded at 260 nm wavelength on either a PerkinElmer Lambda 25 spectrophotometer equipped with a variable temperature Harrick cell accurate to ±1 °C. The path length was varied between 6.25 μm and 50 μm as appropriate. Measurements were taken at 5 °C intervals from 20 °C to 80 °C and repeat measurements were taken until measurements agreed with each other. Alternatively, a Shimadzu UV-1800 spectrometer outfitted with a Shimadzu TMSPC-8 temperature controller was used, with a temperature ramp of 0.5 °C per minute from 20 °C to 80 °C. A Shimadzu cell of path length 10 mm was used.

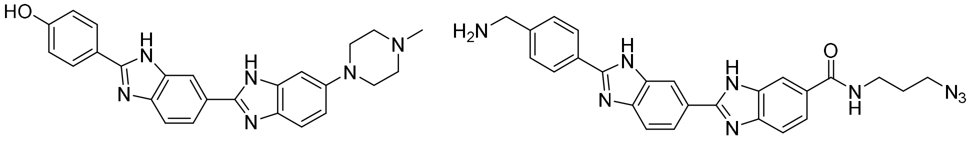
Fluorescence spectra were collected between 350 and 600 nm wavelength at 2.5 mM concentration on a Perkin-Elmer instrument equipped with a Harrick cell of path length 6.25 μM. The raw data were processed using the Origin 2016 software package.

FT-IR spectra were recorded on a Bruker Vertex 80 spectrometer equipped with a Harrick temperature controlled cell, accurate to ±1 °C. Samples were placed onto CaF_2 discs prior to placement in the Harrick cell. Spectra were taken at 2-3 °C intervals and were recorded after no further change was observed in the IR spectrum, when the system was at equilibrium. All data collected was processed using the Origin 2016 software package.

4. Structural Characterisation of a Hoechst 33258 analogue in complex with a dsDNA dodecamer.

4.1 Introduction.

NMR spectroscopy has proven invaluable as a tool for investigating both biomolecular structure and dynamics of DNA: ligand complexes in solution. Although it lacks the structural resolution of X-ray crystallography, it is a solution-phase technique that negates any effects that dehydration of a dsDNA: ligand complex as well as crystal packing can have on a DNA: ligand complex (*Section 4.1.3*).[146] This chapter highlights the techniques used and the structural information gained from NMR studies of complexes of H33258, **1.9** (Fig. 4.1) with a dodecamer sequence of dsDNA. The overarching aim of this chapter is to characterize the structure of the interaction of **1.20** (Fig. 4.1) in complex with a dsDNA oligomer, and to find the structural basis for the changes in the azide response of the **1.20**: ODN 3.2 complex compared to free **1.20**.



1.9

1.20

Figure 4.1. Structure of Hoechst 33258 **1.9** and the analogue **1.20**.

*4.1.1: Previous Studies of **1.9**: dsDNA complexes by NMR.*

Since the first reports of structural studies of dsDNA structures by NMR in the early 1980s, much work regarding the structure and dynamics of free DNA duplexes and in complex with proteins/peptides, other oligonucloetides (forming a triple helix) and small molecules are known.[7] One of the first NMR structures of a dsDNA oligomer in complex with a small molecule was with **1.9**.[99]

Since this first report, NMR studies of complexes of H33258 with both A-tracts and A-T repeats of dsDNA have been published investigating the high-affinity minor groove binding mode, and are more fully discussed in *Section 2.1*. One such study of the structure of two molecules of **1.9** bound to a dsDNA oligomer containing two non-consecutive A-T tracts is shown in Fig. 4.2.[91] The NMR structure mirrors that observed for the X-ray structure solved by Spink *et. al.* (Fig. 2.2).[64]

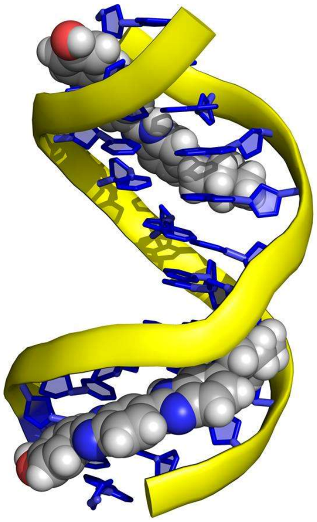
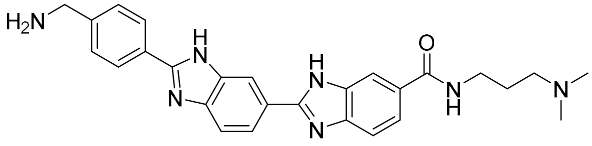


Figure 4.2 NMR structure of d(CTTTTGCAAAAG)$_2$ in complex with two molecules of H33258. PDBID: 1QSX.[91]

The 2 molecules of **1.9** are bound to the two A-tracts of DNA, and MD simulations revealed structural characteristics typical of complexes of **1.9** and dsDNA, as well as evidence of co-operative binding of 2 molecules of **1.9** where an oligomer contains 2 binding sites.[84] Firstly, binding of **1.9** to dsDNA is driven by a positive entropy change, and is attracted to the narrow minor groove, and enhanced negative electrostatic potential of A-T tracts. Shape complementarity between A-T tracts and H33258 appears important for complex formation.[84] Binding to sequences of dsDNA containing G-C base pairs is disfavoured for two reasons: firstly, sequences containing G-C base pairs have curvatures not complementary to the shape of **1.9** (Fig 4.2). Secondly, the C2-amino group of guanosine bases protrudes into the minor groove, leading to an unfavourable steric clash between the G-bases and the incoming **1.9** molecule. Sequences of dsDNA containing G-C base pairs do not exhibit the same conformational flexibility as A-T base pairs, and consequently, are less able to assume the optimal narrow minor groove conformation associated with H33258 binding. This is due to the large helical twists observed for pyrimidine-purine base steps (TpG, CpA).[84] For this reason, the steric bulk of the piperazine moiety of **1.9** causes a shift in the binding mode from AATT to AATTG, which exhibits a slightly wider minor groove.

This is in contrast to distamycin and netropsin, which have less bulky amidine tail-groups, which can comfortably be accommodated by A-tracts.[147] Derivatives of **1.9** with less bulky tail-groups, such as the diamine **2.24** (Fig. 4.3), are easily accommodated by both AAATTT and AAAAA tracts of dsDNA.[94]



2.24

Figure 4.3. Structure of the Hoechst derivative **2.24**

4.1.2. Aims

The objective of the work described in this chapter is to elucidate the binding mode of the azide-bearing derivative **1.20**, (Fig. 4.1). The specific aims are twofold:

i. To determine the binding mode of **1.20** to its target dsDNA sequence.

ii. To identify the location of the azide in **1.20** in complex with dsDNA, and hence, deduce the local solvent/hydrogen bonding environment of the azide when in complex with dsDNA. (*Section 3.5.4)*.

*4.1.3. The binding mode of Hoechst analogue **2.24** to dsDNA as established by NMR.*

The binding mode of **2.24** to varying sequences of dsDNA has been solved previously,[94] and the binding mode of **2.24** to dsDNA is largely independent of base sequence, save for the need for a region of at least 4 consecutive A-T base pairs.[89] The key enthalpic binding interaction is a bifurcated hydrogen bond between the benzimidazole NH atoms and the central base pair of the A-T tract (Fig. 4.4),[99] and the main entropic driving force is the dehydration of the DNA oligomer upon formation of the complex (*Section 3.1*).[131]

The hydrogen bonding interaction of **2.24** is not specific to any base step, and consequently, Hoechst 33258 will bind to any A-T tract of firstly, sufficient length of 4 base pairs, and secondly, with the ability to exhibit a narrow minor groove conformation. For this reason, **2.24** will bind to ApT steps but cannot bind TpA steps. This fact, coupled with the strong ion-dipole interaction formed between the piperazino- group and the phosphate backbone of dsDNA, leads to the origins of the parallel (i.e. from 5'- to 3'-) directionality of H33258 binding to dsDNA. In the case of a longer A-T tract than shown in Fig. 4.4, for example A_3T_3, the binding mode of **1.9** is directed towards the 3'-end of the minor groove, in order to accommodate the bulky piperazino group (Fig. 4.5).[128]

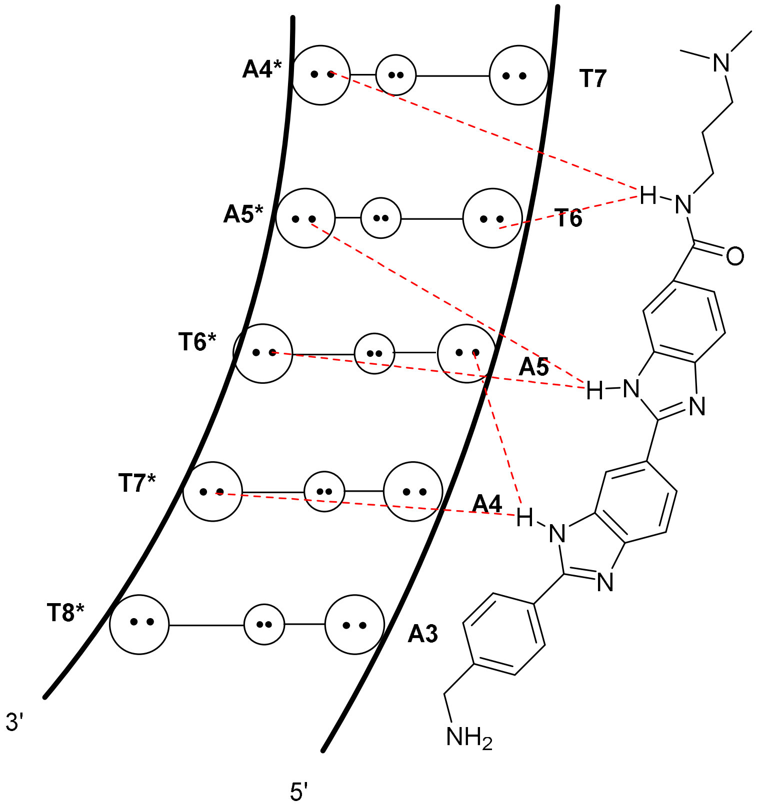


Figure 4.4. Diagram indicating the key hydrogen bonding interactions between **2.24** and the dsDNA oligomer, d(GCAAATTTGC)$_2$.[99] The 5 base pair binding site is shown.

When a less bulky aliphatic tertiary amine is used instead of N-methylpiperazine, as in the diamine **2.24** (Fig. 4.3), the same binding mode shown in Fig. 4.4 is still observed, but the bifurcated hydrogen bonds are between the A-T base pairs in the middle of the A-T tract. Less bulky tail-groups can be effectively accommodated in narrow minor grooves, and consequently, the bifurcated hydrogen bonds are formed between the benzimidazole NH groups and the central A-T base pair. The amide bond present in **2.24** also forms a hydrogen bond across the next base pair, in a manner similar to that observed for distamycin.[94]

The binding modes of **1.9** and **2.24** are subtly different, induced by both base sequence and the steric bulk of the cationic group. The cationic group experiences an electrostatic interaction with the narrowest point of the minor groove, where the minor groove is especially narrow, in the case of A, or A_3T_3 tracts, the bulky piperazine group of **1.9** is distorted away from the narrowest part of the minor groove and towards the centre of the minor groove, subtly altering the binding site. The aim of the work described in this chapter is to elucidate the binding mode of **1.20**, in order to understand the effect that changing the location of the cationic group from ring 3 to ring 1 (Fig. 4.6), and secondly to locate the azide functional group, in order to understand the structural basis for the origin of the red shoulder peak, as well as the increase in absorbance of the azide absorption band in the IR spectrum of the **1.20**: ODN 3.2 complex.

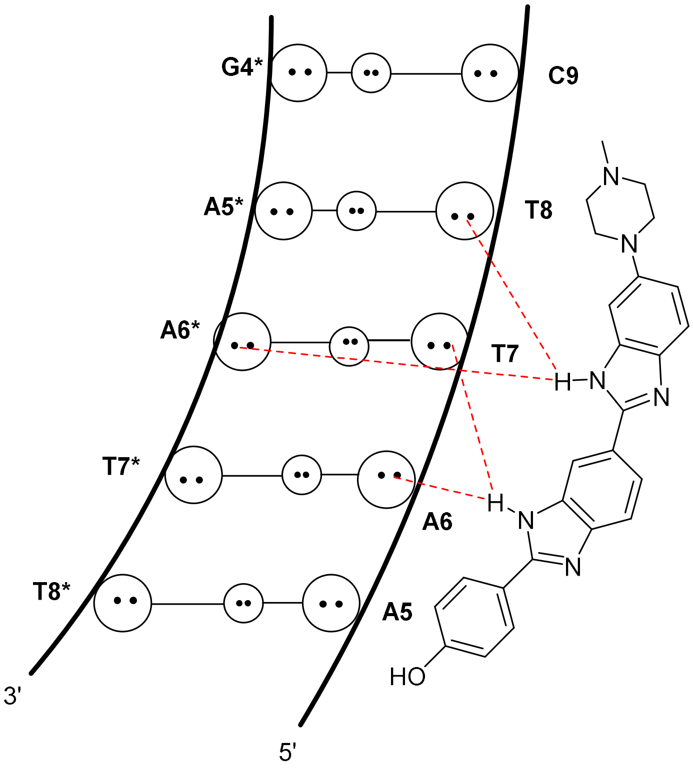


Figure 4.5. Diagram showing the hydrogen bonding interactions between **1.9** and an A_2T_2 tract of dsDNA.[94] The reported hydrogen bonding interaction is shown. The minor groove in this dodecamer is suitably wide enough to accommodate the piperazine group in the TpC and GpA steps.

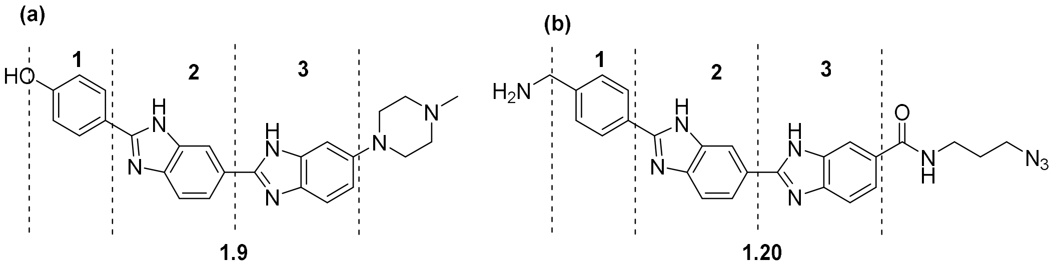


Figure 4.6. **(a)** Numbering of the rings of **1.9**. **(b)** Numbering of the rings of **1.20**.

4.1.4. Choice of NMR as a technique

Two techniques are mainly used for structure determination of small molecules in complex with biomolecules: NMR and X-ray crystallography. NMR was chosen as a method of structural characterization for the following reasons:

i. NMR is a solution-phase technique, and consequently distortion of the structure due to dehydration and crystal packing effects cannot be observed as has been previously seen for **1.9**.[64, 99]

ii. Crystallisation of complexes of dsDNA with small molecules is a difficult process. The much simpler sample preparation for an NMR study therefore makes structure characterization by NMR more attractive.

iii. Precipitation of the complex of **1.20** with dsDNA was observed, further complicating growth of single crystals.

Although crystallography gives a more direct visualization of the complex (NMR requires *in silico* processing to obtain precise atomic co-ordinates), as shown earlier in this section, high-resolution structures can be obtained from NMR studies.

4.1.5. NMR techniques used for structural characterization of dsDNA: ligand complexes.

Three main 2D-NMR techniques are predominantly used to assign the ^{1}H spectrum of dsDNA oligomers; these are NOESY, COSY and TOCSY (*Section 1.5.4*). The most important of these is NOESY, as assignment of interbase correlations (hereafter referred to as the "spectroscopic walk") is only possible using the NOESY data (Fig. 4.10).[65] Strong nOe signals exist between these two resonances on adjacent bases, and hence assignment of this region allows for the identification of two resonances on each base, from which the other resonances can be filled in.[65]

COSY and TOCSY are both through-bond 2D-NMR techniques that help with unambiguous assignment of the ^{1}H-NMR. COSY cross-peaks come from energy relaxation between two resonances separated by three bonds. From this spectrum, any ambiguity between NOESY assignments can be cleared up. This is especially useful in the case of strong ligand binding, in which conformational changes in the DNA duplex can occur, leading to vastly reduced nOe signals compared to the free duplex.[7] TOCSY gives information in much the same way as COSY, but cross-peaks are observed for full spin systems, not just adjacent protons as observed for COSY. A TOCSY experiment can clear up ambiguity in assignments of protons with similar chemical shifts from different bases, as can happen often with sugar protons.[65]

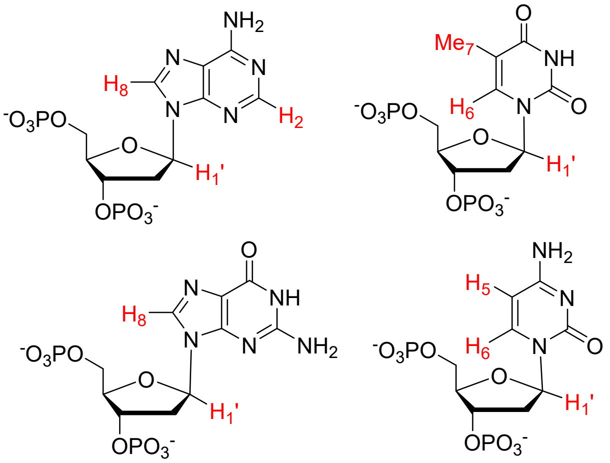


Figure 4.7. Structures of the four DNA nucleotides with the key nonexchangeable protons highlighted in red. Exchangeable protons are shown in black. For clarity, the H2’-H5’ protons are not shown.

4.2 NMR assignment of free ODN 4.1.

4.2.1: Assignment of the fingerprint of the dodecamer DNA duplex ODN 4.1.

The ^{1}H-NMR spectrum of free ODN 4.1 (Fig. 4.8) in H_2O/D_2O solution is shown in Fig. 4.9. ODN 4.1 was chosen as it is a self-complementary analogue of ODN 3.2, containing the consensus binding site of 5’-AATT-3’ as shown by DNase I footprinting (*Section 3.2*). Significant stabilisation of ODN 3.2 was also observed at 1 μM concentration. On this basis, ODN 4.1 was chosen.

5'- G1 C2 G3 A4 A5 A6 T7 T8 T9 C10 G11 C12 -3'
3'- C12* G11* C10* T9* T8* T7* A6* A5* A4* G3* C2* G1* -5'

Figure 4.8: Structure of ODN 4.1. Strand 2 is marked with an asterisk throughout.

The peaks in the spectrum can be broken down into four approximate and distinct groups: the imino region (11 - 14 ppm), the aromatic region (7 - 8 ppm), the H1' and CH5 region (5 - 6 ppm) the H3' region (4.5 - 5 ppm), the H4' region (4-4.5 ppm), the H5' region (3.8 - 4.3 ppm) the H2' region (1.5 – 2.5 ppm), and the thymidine methyl region (1 - 1.5 ppm).

For a meaningful comparison of the binding mode of **1.20** to dsDNA, it proved necessary to assign the spectrum of dsDNA, and then obtain an energy-minimised structure. To aid in the assignment of the free DNA spectrum, a chemical shift predictor was used to calculate the approximate chemical shifts of the duplex (Appendix 7.15).[148,149] The NOESY spectrum was recorded and assigned in both D_2O and 9:1 v/v H_2O/D_2O and a table of peak assignments created (Appendix 7.16). Assignment of the spectrum was carried out from the NOESY spectrum using the "sequential walk" technique (Fig. 4.10).[65] The fingerprint region of a DNA duplex is referred to as the region between 5-8 ppm in which all the base aromatic and sugar H1' correlations exist. However, it is necessary to find a starting point for assignment of the spectrum, usually the protons of the free 5' atom of the first base.

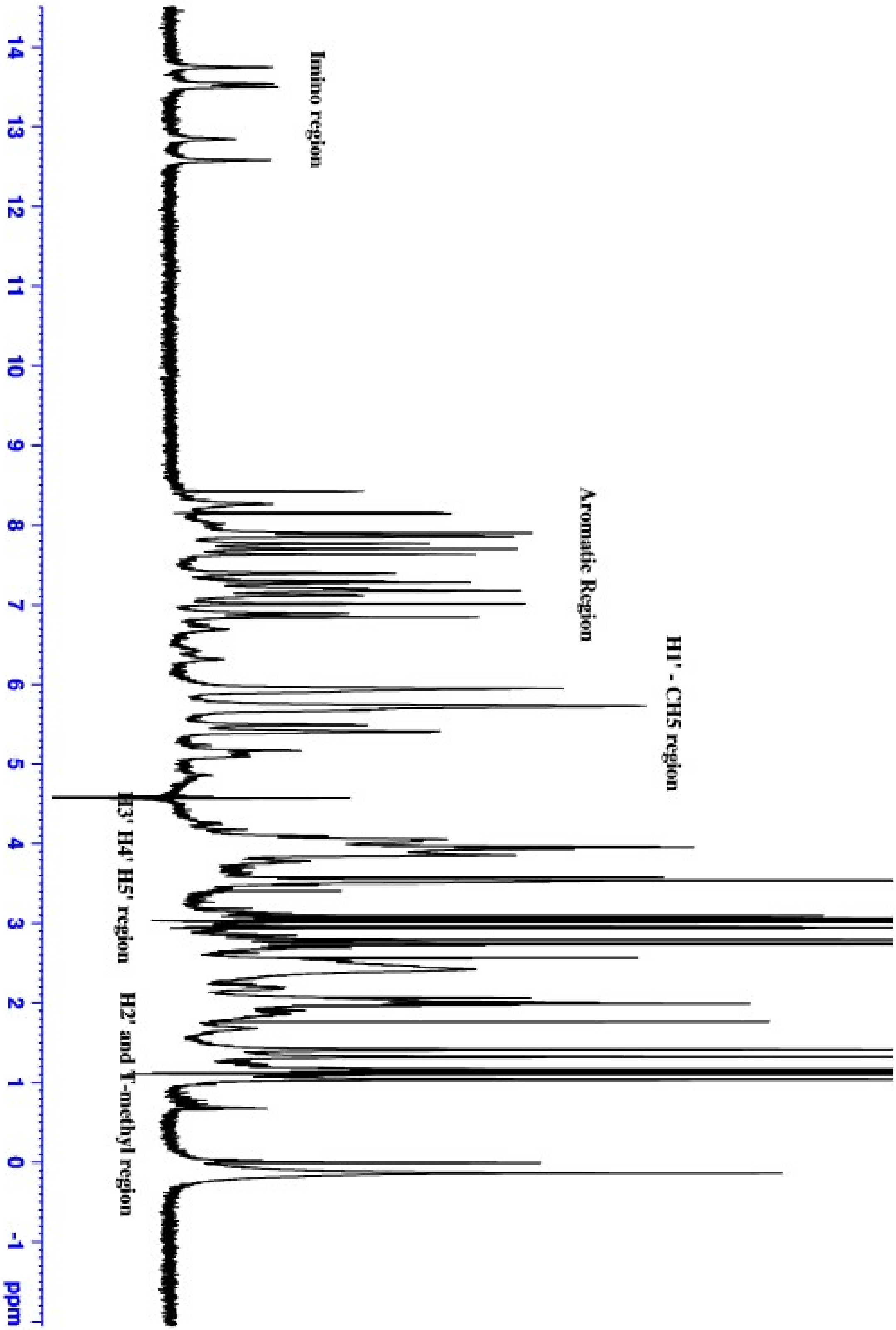


Figure 4.9. ^{1}H NMR of ODN 4.1

The peaks of the G1H5'/H5'' proton (denoted as a pseudoproton Q5' due to its identical chemical shift) have a distinct chemical shift and serve as a starting point for assignment of the NOESY spectrum. Starting from the G1Q5'-H8 nOe correlation of the G1 nucleotide, the walk proceeds through H1' proton of the same base (Fig. 4.11). NOESY correlations between the H1' proton of G1 and the H5 and H6 protons of C2 could then be observed, and so on as per Fig. 4.11. This technique allows for the assignment of the both the aromatic and H1' sugar protons of dsDNA, and consequently the other peaks corresponding to the other sugar protons can be filled in.

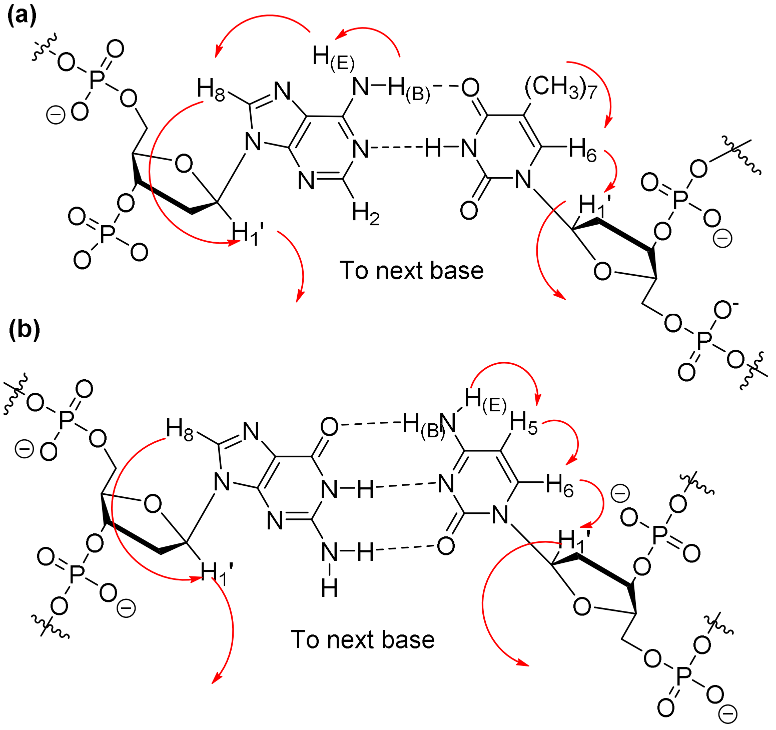


Figure 4.10. "Sequential walk" down the major groove of **(a)** an A-T and **(b)** a G-C base pair.

The assignments of the first six bases are shown in Fig. 4.11. The peaks of the first three bases, G1, C2, and G3 proved simple as the chemical shifts of these protons were different enough to each other such that assignment of these peaks was facile. The assignments of the A-T tract however, proved more difficult because the protons are in very similar chemical environments, they have very similar chemical shifts and signal degeneracy complicates assignment of the peaks and consequently, the signal resolution is poor. The technique employed was to work backwards from the T7C7 proton, find the protons from A6 and then find correlations for A4 from G3. By a process of elimination, therefore, the A5 protons could be assigned (Fig. 4.11).

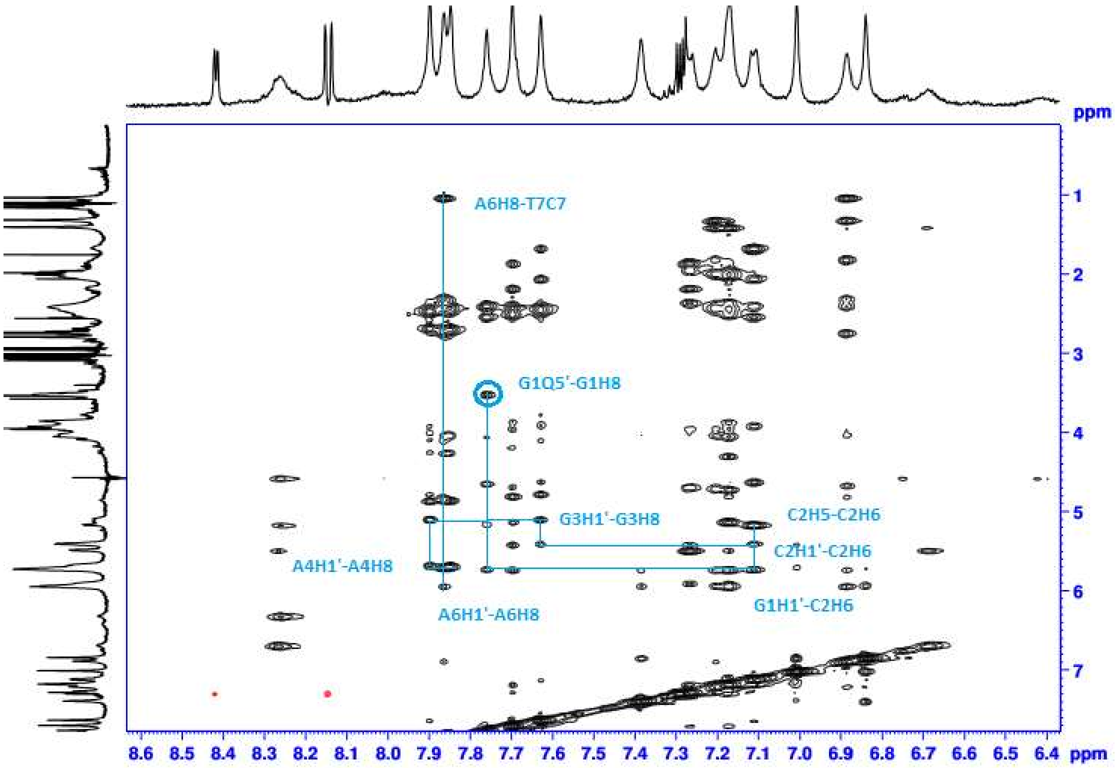


Figure 4.11. Assignment of the NOESY spectrum of the first six bases of ODN 4.1, G1 C2 G3 A4 A5 A6.

Correct assignment of the A_3T_3 tract was essential as the hypothesized binding site of the ligand **1.20** is across the middle of the A-T tract, and the perturbations of chemical shift caused by binding of **1.20** are expected to be localised across these base pairs.

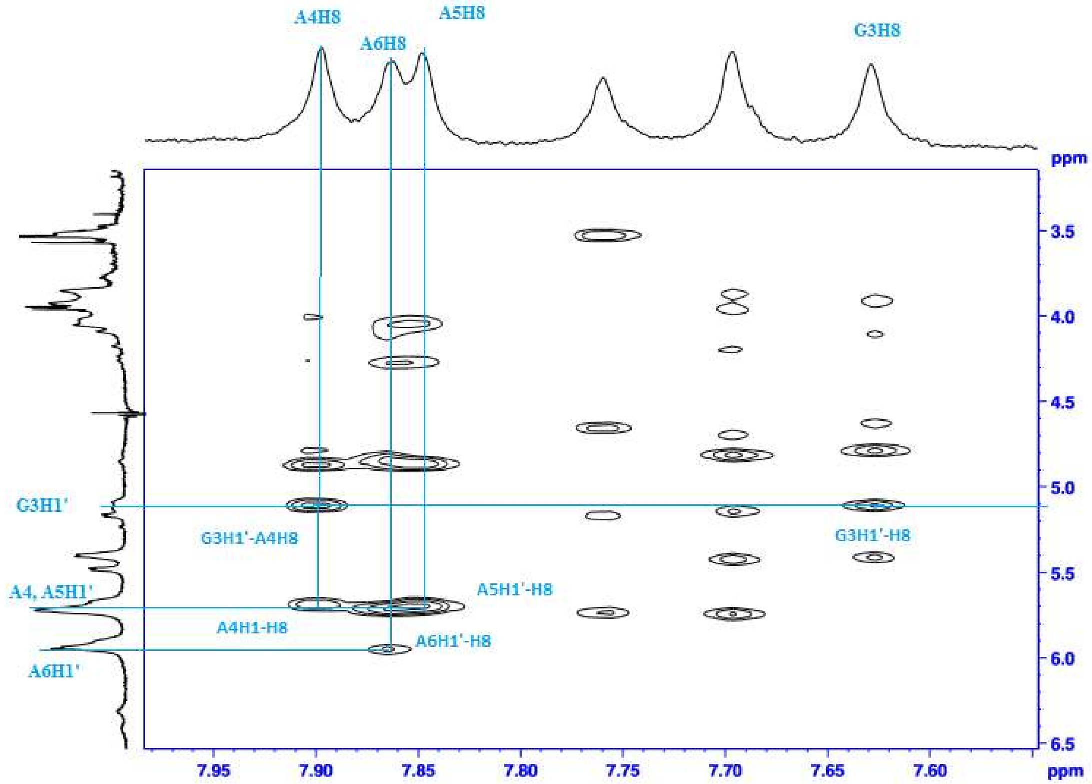


Fig. 4.12. Assignment of the G3, A4, A5, A6 nucleotides of ODN 4.1.

Fig. 4.11 shows a definite correlation exists between the H8 proton of H6 and the C7 methyl group of T7. From here, the three thymidine H6 and C7 protons can be assigned (Figs. 4.13, 4.14). Assignment of the H6 and C7 peaks proved facile due to good signal resolution (Fig. 4.13). However, the H1’ peaks were more difficult to assign, and required assignment of the aromatic peaks of the C10-C12 tract in order to unambiguously assign the H1’ peaks of the T7-T9 bases. This signal degeneracy, as also observed for the A-tract, is a product of the chemical environment of the protons, and is therefore unavoidable without changing the A_3T_3 tract (Appendix 7.15). The H1’ of T7 was distinct enough from both T8 and T9 and could be assigned directly from its correlation to T7H6, whereas T8 and T9 were found to have very similar chemical shifts, and careful inspection of the NOESY spectrum was essential to assign these peaks (Fig. 4.13).

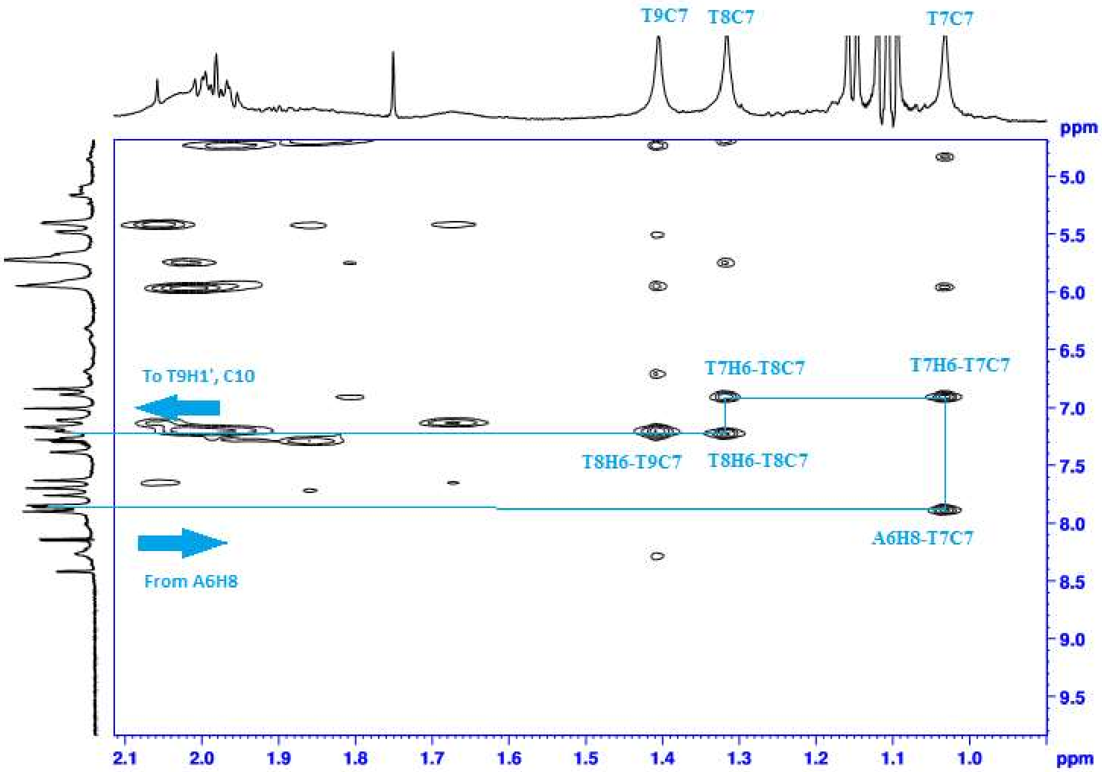


Figure 4.13: Assignment of the T7-T9 methyl peaks of ODN 4.1.

They could however, be distinguished from correlations (or lack thereof) to C10H5. Assignment of C10-C12 proved as facile as for G1-G3 (Fig. 4.14).

Once the spectroscopic walk was finished, the H2' – H5' sugar protons could then be filled in from the NOESY and TOCSY correlations from H1'. The complete assignment is shown in Appendix 7.16. The one base in which no H5' could be determined was C10. This is presumably due to a very similar chemical shift value for H4' and H5'/H5'' in C10 as predicted (Appendix 7.15), and hence the correlation contour between these two protons was indistinguishable from the diagonal of the NOESY plot.

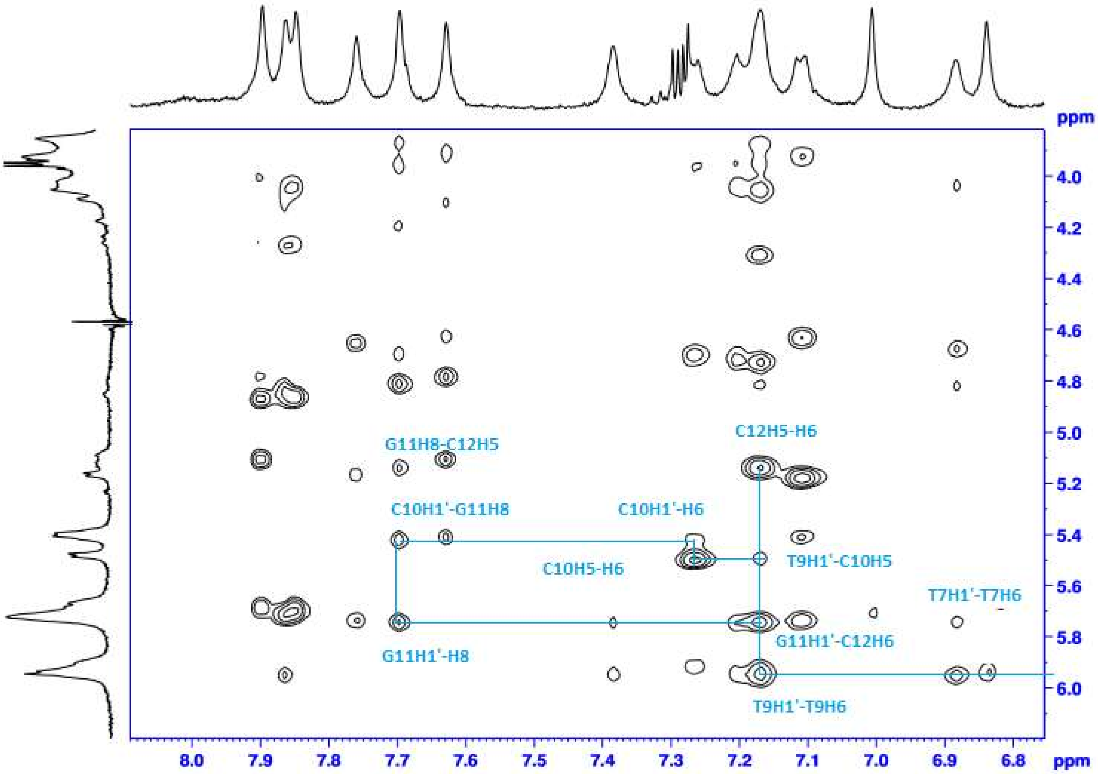


Figure 4.14. Assignment of the final three bases of ODN 4.1. Relevant peaks not in the "walk" are circled and labelled. The T9H1' and C12H1' had almost identical chemical shifts, and are not distinguished here. Their assignments are reported in Appendix 7.15.

4.2.2: Assignment of the minor groove of ODN 4.1.

With the assignment of the fingerprint region and the sugar protons in hand, it became possible to assign the protons in the minor groove, namely the thymidine imino protons and the AH2 protons. Three imino protons exist for each of the guanosine and thymidine protons, but only five peaks are visible in the ^{1}H spectrum (Fig 4.15) – this was found to be due to signal degeneracy of two guanosine imino protons (Appendix 7.16). Assignment of the NOESY spectrum of these peaks involved a somewhat different approach to that described in Section 4.4.1. The correlations observed in these peaks are horizontal (*i.e.* between 2 Watson-Crick base-paired strands) rather than vertical (down a strand of ssDNA) (Fig. 4.16).

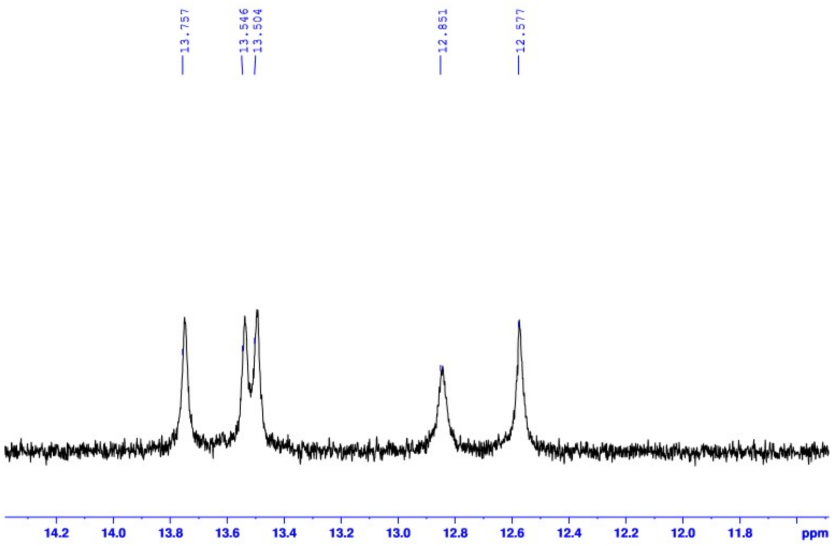


Figure 4.15. Imino region of the ^{1}H NMR spectrum of ODN 4.1

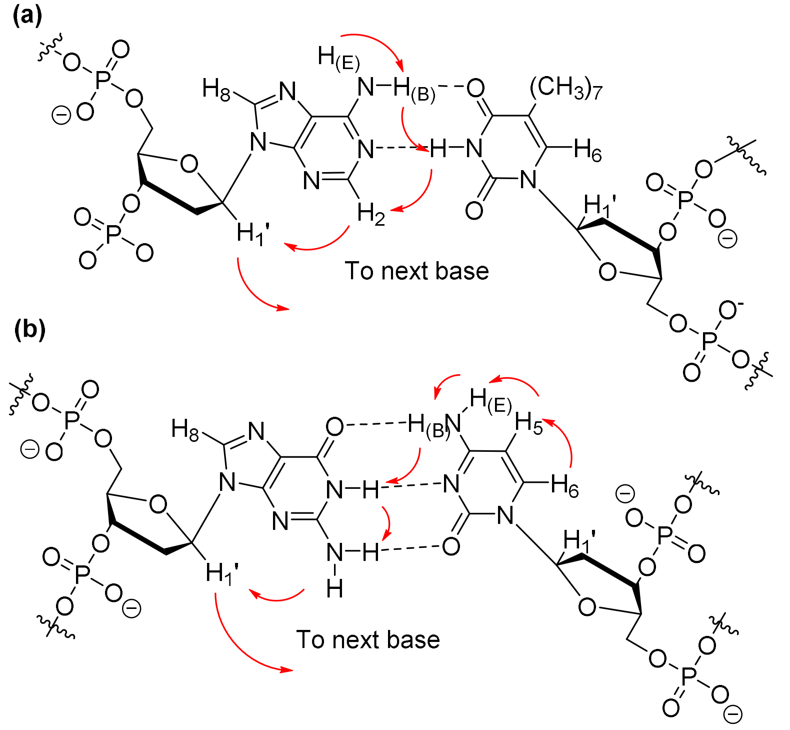


Figure 4.16. Horizontal minor groove nOe correlations in (a) and A-T and (b) a G-C base pair.[65]

A correlation exists between cytosine H5 and the amine protons, which due to base pairing, have very different chemical shifts (Appendix 7.16) (Fig. 4.16). Hence, to start the assignment of this part of the spectrum is to assign the cytosine amine protons from the H5 and H6 correlations. The amine protons from C2 and C10 could be assigned unambiguously, but signal overlap between the resonances of the three cytosine bases prevented unambiguous assignment of the C12 amino protons. This proved not to be important; however, as the three guanosine imino protons could be unambiguously assigned from these two imino protons (Appendix 7.16).

Upon assignment of the guanosine imino protons, the final peaks to assign were the adenosine H2 and thymidine imino protons. These protons were assigned through inspection of the imino-imino resonances, and the adenosine H2 protons were assigned through the minor groove "walk" depicted in Fig. 4.16. The correlations are shown in Fig. 4.17. The assignment of free ODN 4.1 was then completed, and this assignment served as a guide to assigning the complex of **1.20** with ODN 4.1.

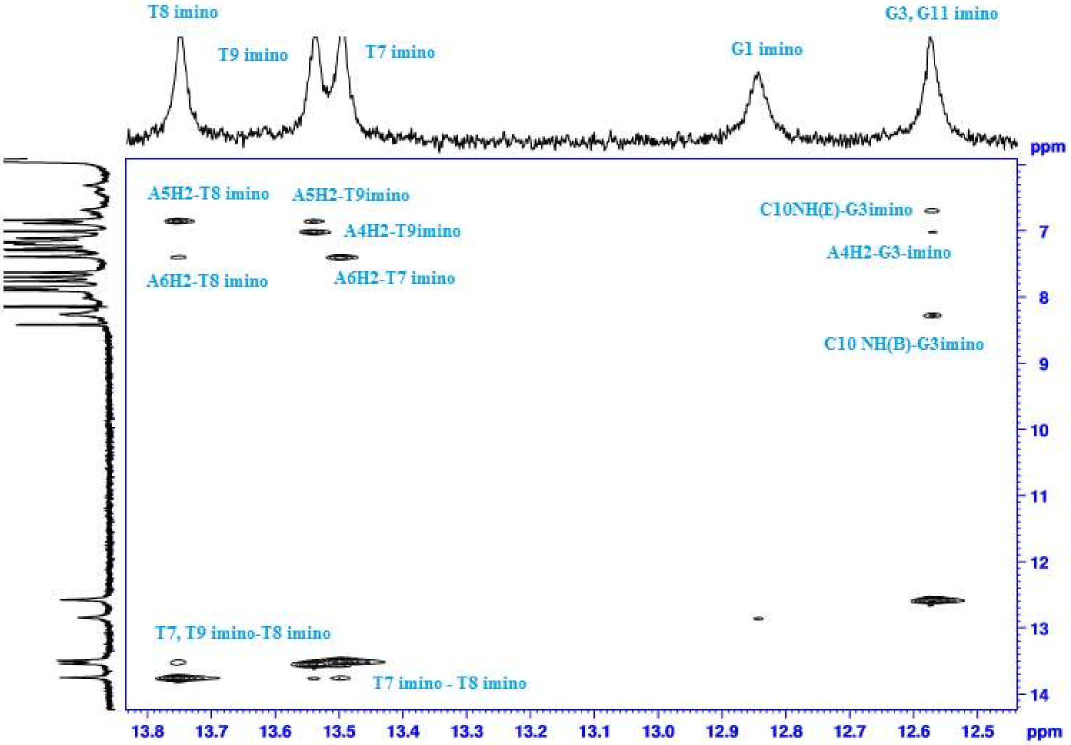


Figure 4.17. Assignment of the minor groove of free ODN 4.1 (correlations to H1' protons not shown).

4.2.3. Summary

The ODN 4.1 spectrum was fully assigned by NMR using the "spectroscopic walk" technique. The spectroscopic walk technique is only possible in the assignment of dsDNA oligomers if the oligomer in question is unambiguously in the B-form (Fig. 1.5), as the observed nOe patterns for A- and Z-form dsDNA are very different, and require other techniques in order to assign these oligomers fully.[65] This assignment of ODN 4.1 serves as a platform for understanding the conformational changes in ODN 4.1 due to binding of a molecule of **1.20** to ODN 4.1 through chemical shift perturbations.

4.3 Assignment of the NMR spectrum of the 1.20: ODN 4.1 complex

*4.3.1. Assessment of the **1.20**: ODN 4.1 complex by ^{1}H-NMR*

The full ^{1}H NMR spectrum of the complex is shown in Fig. 4.18. A well-resolved ^{1}H-NMR spectrum was observed, and hence a well-defined complex was formed. If this was not the case, and either multiple, or a low affinity complex was present, then essentially two compounds would be present in solution. This would result in a poorly-resolved spectrum making assignment extremely difficult, even impossible.[65,99]

The **1.20**: ODN 4.1 complex was analysed at a variety of temperatures from 293 – 308 K, and was found that the spectrum at 303 K gave the best resolved data, and hence all data shown in this section comes from this data set.

Figures 4.19 and 4.20 show comparisons between the imino region and the thymidine methyl region of the complex, indicating that breaking of the dyadic symmetry in the thymidine region has occurred, as the three thymidine imino peaks integrating for two protons are now represented by six resonances integrating for one proton each (Fig. 4.19). The guanosine imino peaks are largely unaffected save for a downfield chemical shift, indicating that these

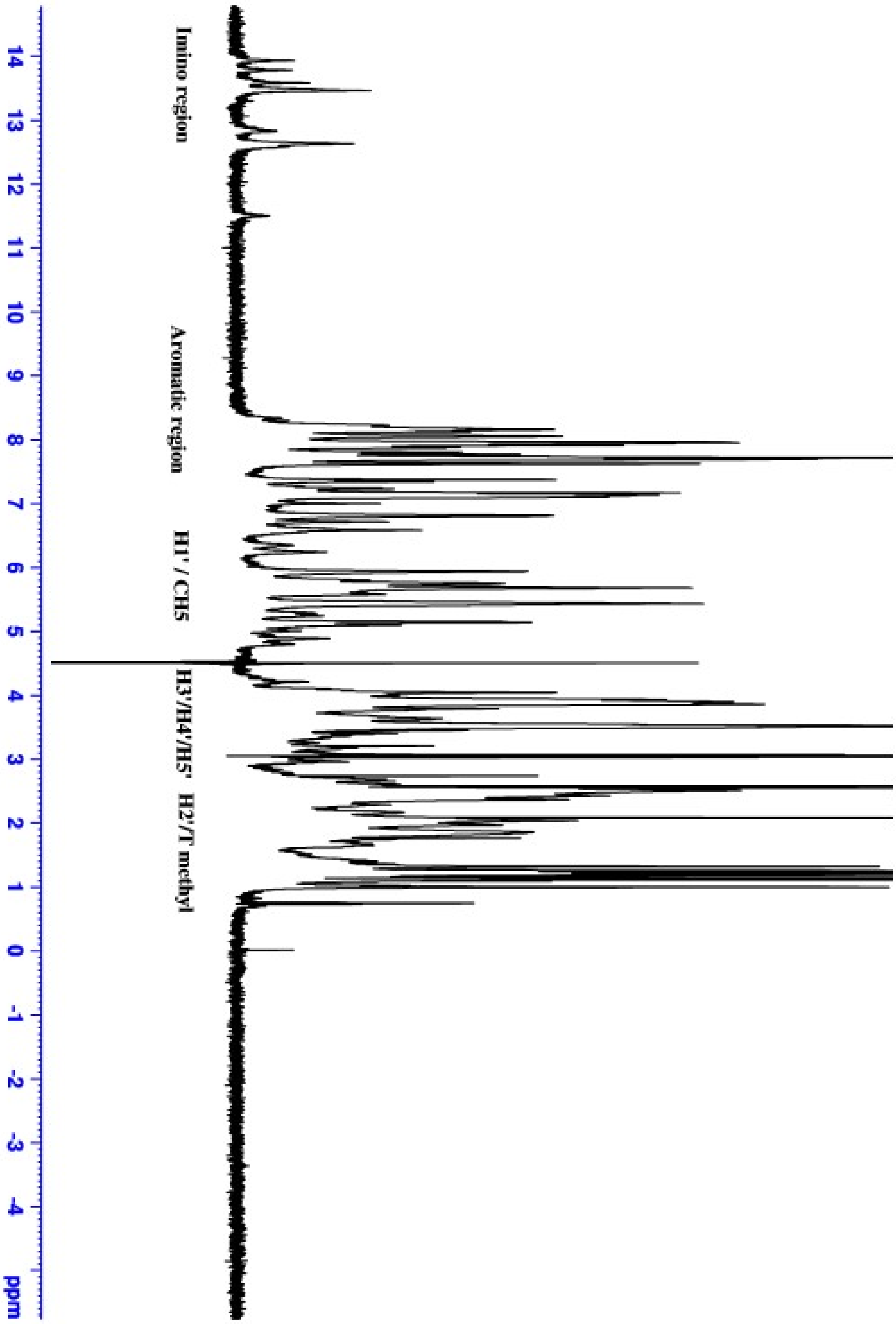


Figure 4.18. ^{1}H-NMR of the **1.20**: ODN 4.1 complex in 9:1 v/v H_2O/D_2O

protons are still equivalent. This indicates binding of **1.20** occurs across the A_3T_3 tract of ODN 4.1. Breaking of the symmetry axis is also seen in the thymidine methyl protons in the complex compared to free ODN 4.1. The thymidine methyl protons are equivalent on each strand, but binding of **1.20** breaks this symmetry – leading to the formation of six distinct methyl peaks in the ^{1}H-NMR spectrum (Fig. 4.20).

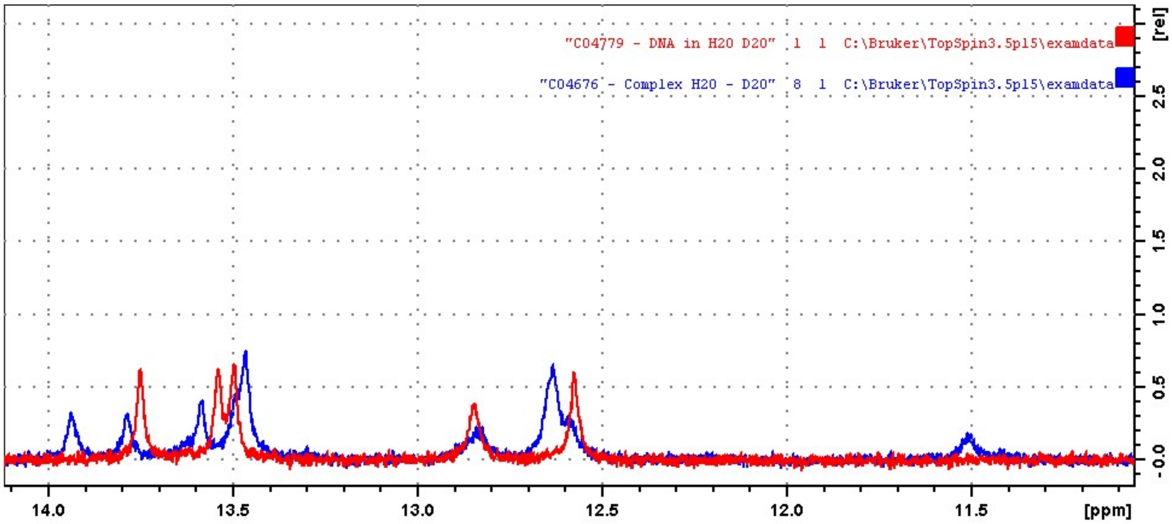


Figure 4.19. Expansion of the ^{1}H-NMR spectrum of the imino region of free ODN 4.1 (red) and the complex of **1.20** with ODN 4.1 (blue). The benzimidazole protons of **1.20** at 12.59 and 11.50 ppm are also present.

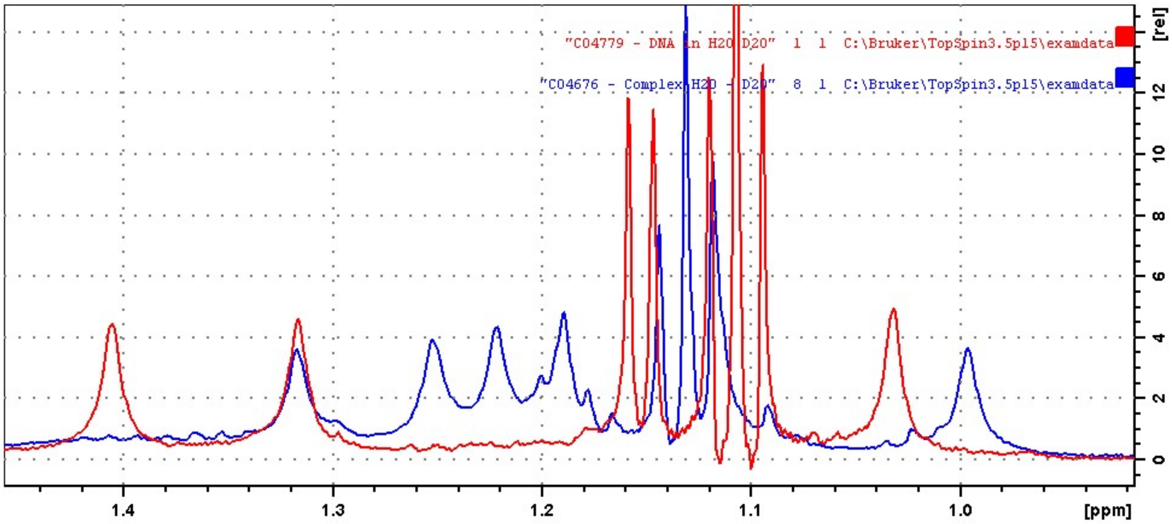


Figure 4.20. Expansion of the ^{1}H-NMR spectrum of the thymidine methyl region of free ODN 4.1 (red) and in complex with 1.20 (blue). The large triplet present at 1.13 ppm is residual HPLC buffer.

4.3.2. Assignment of the H6/H8 and H1' protons of the ***1.20****: ODN 4.1 complex.*

The strategy employed for assignment of the spectrum is the same as for assignment of free ODN 4.1 in that the starting point is the Q5' pseudoproton in the first guanosine nucleotide (Fig. 4.21).

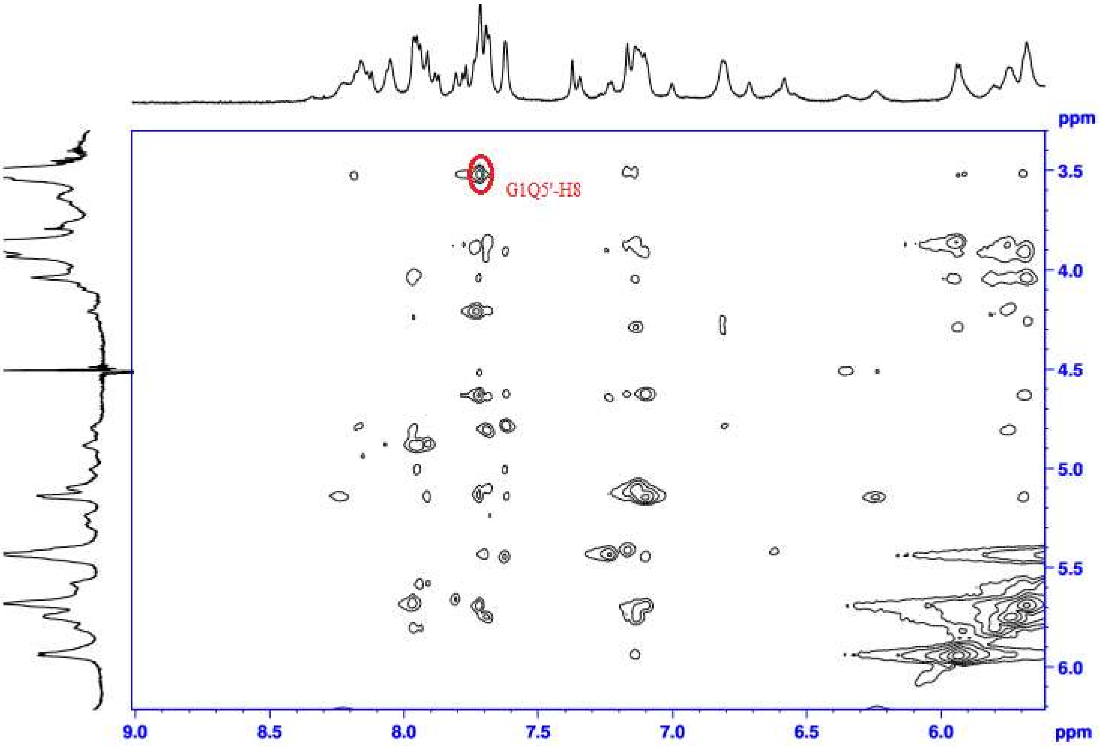


Figure 4.21. Fingerprint region of the NOESY spectrum of the ODN 4.1:**1.20** complex. The start point of the spectroscopic walk (G1Q5'-G1H8) is highlighted.

There was little perturbation of chemical shifts of the G1-G3 protons of the **1.20**: ODN 4.1 from that observed for free ODN 4.1 (*Section 4.2.1*). This suggested that **1.20** binds ODN 4.1 across the central A-T tract. Assignment of the A-tract required careful inspection of the spectrum, because, as with the imino and methyl groups of the paired thymidine bases, binding of **1.20** into the A-tract has broken the dyadic symmetry of the DNA duplex, leading to six peaks representing the adenosine H8 protons. These two spin systems could be distinguished from each other as the H1' atoms of G3 and G3* nucleotides have subtly different chemical shifts to one another – perhaps a clue as to the presence of the azide functional group. The "sequential walk" for the first strand is shown in Fig. 4.22, and the same walk for the second strand is shown in Fig. 4.23.

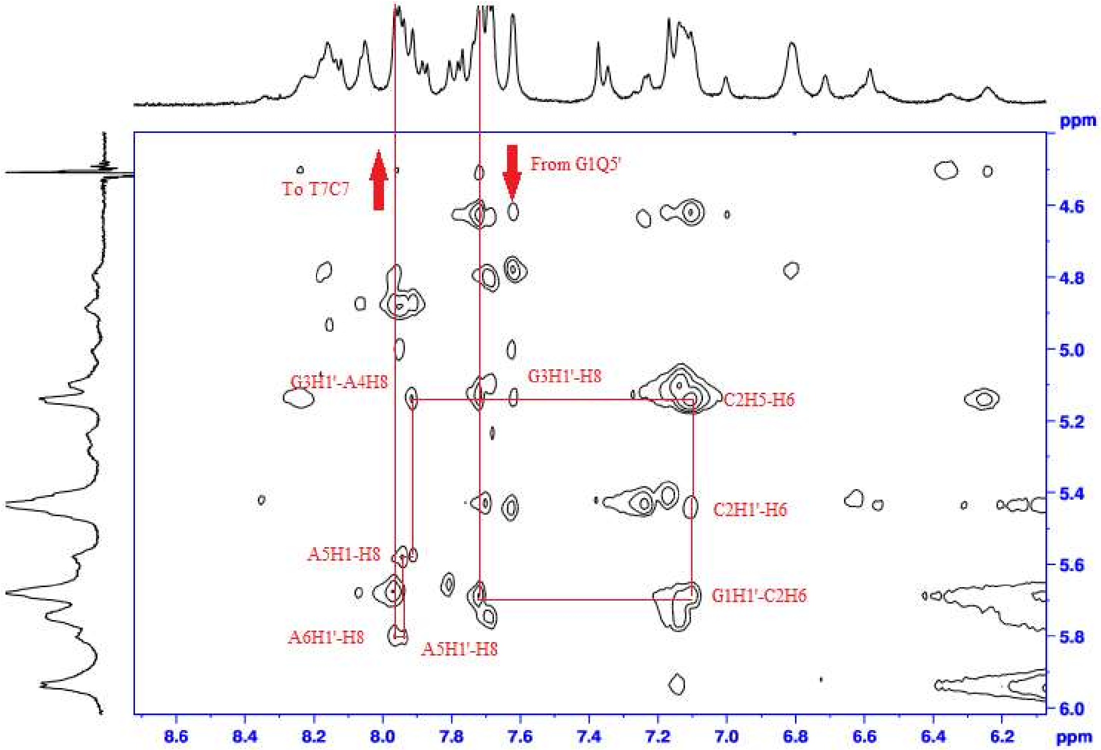


Figure 4.22. The G1-A6 sequential walk. The G3*H1' peak is at 5.00 ppm (not marked).

Unexpectedly, the perturbation of the A-tract was not well-resolved; however, using the strategy employed for solving the A-tract of free ODN 4.1 (using the T7 methyl correlation to find A6) the two A-tracts could be effectively resolved (Figs. 4.22 and 4.23). The chemical shifts for A5* and A6*H1' of the complex were similar to free ODN 4.1 (Appendix 7.17). As expected for the binding of Hoechst-type compounds, the most significant perturbations of chemical shift were found in the T-tract of the complex (Appendix 7.17).

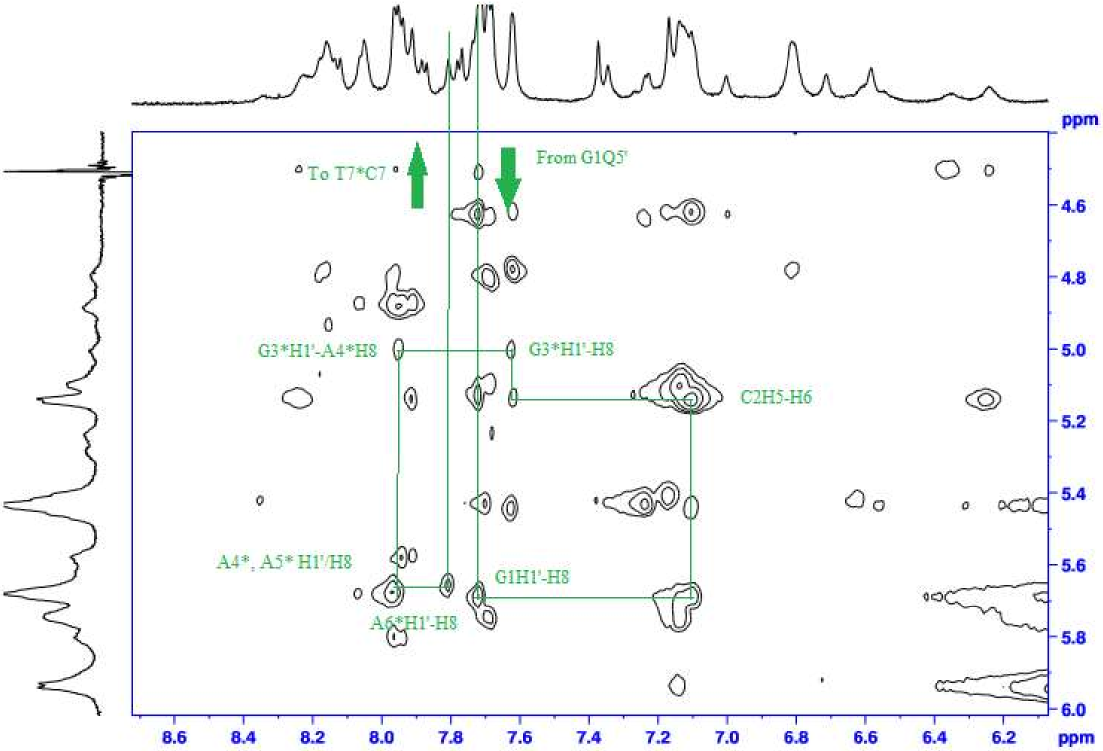


Figure 4.23. The G1*-A6* spectroscopic walk.

T7 and T7* could be easily assigned and distinguished from one another, as the methyl and H6 protons had chemical shifts well-resolved from one another (Fig. 4.24). In the second strand, the chemical shifts of T8* and T9* can be considered identical to one another, and to further complicate matters, are identical to that observed for T8H6. Assignment of these different peaks was difficult and assignment of these peaks involved careful inspection of the correlations to the resonances belonging to both T7* and C10*. Assignment of these peaks was possible (Fig. 4.24).

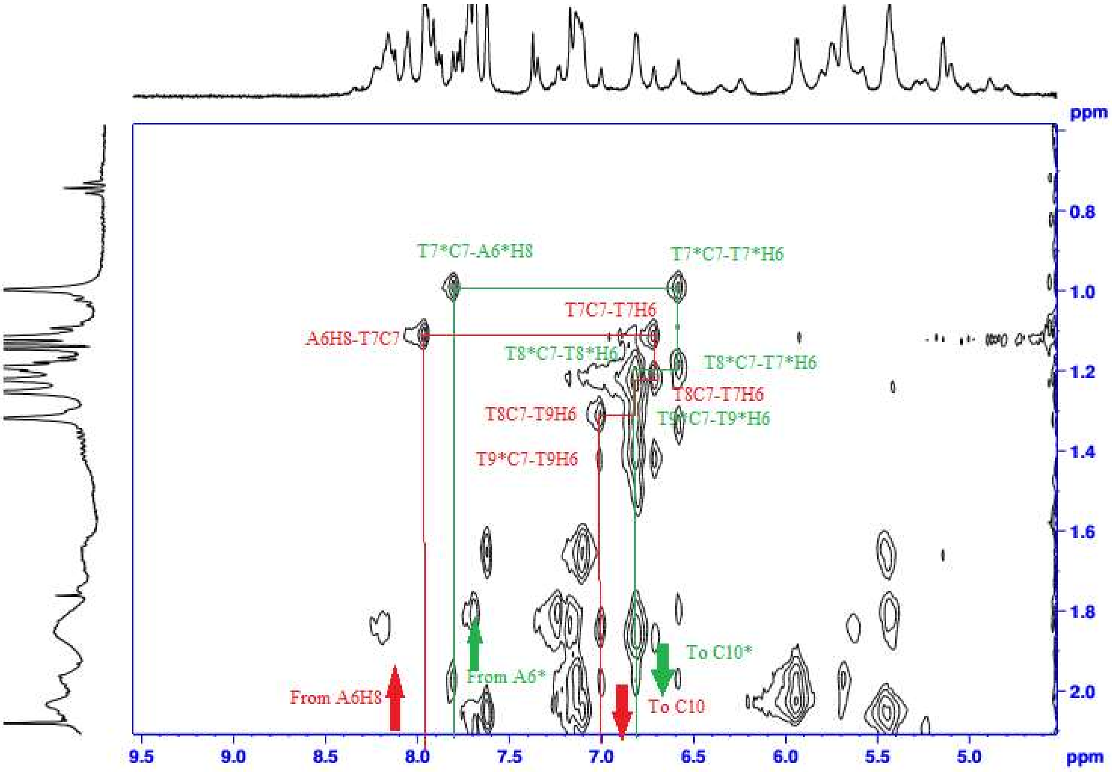


Figure 4.24. The T7-T9 H6 and C7 correlations (red) and the T7*-T9* correlations (green).

The thymidine H1' protons proved difficult to assign, as only weak nOe correlations were observed for the H6-H1' protons in the **1.20**: ODN 4.1 complex compared to free ODN 4.1. This is likely due to rotation about the glycosidic bond, suggesting a subtle conformational change from B-DNA has occurred upon binding of **1.20**, as has been observed for both **1.9** and **2.24**.[94,144,84] As such, a slight rotation of the glycosidic bond is enough to increase the distances between the H6 and H1' nuclei such that the two nuclei are sufficiently separated that the NOE effect is less pronounced. Strong H6/H8-H1' NOE correlations are observed when nucleosides are in both the *anti-* (found in A-and B- DNA) and in the *syn-* (in Z-DNA) conformations, and are lower when an intermediate conformation is observed (Fig. 4.25).[7]

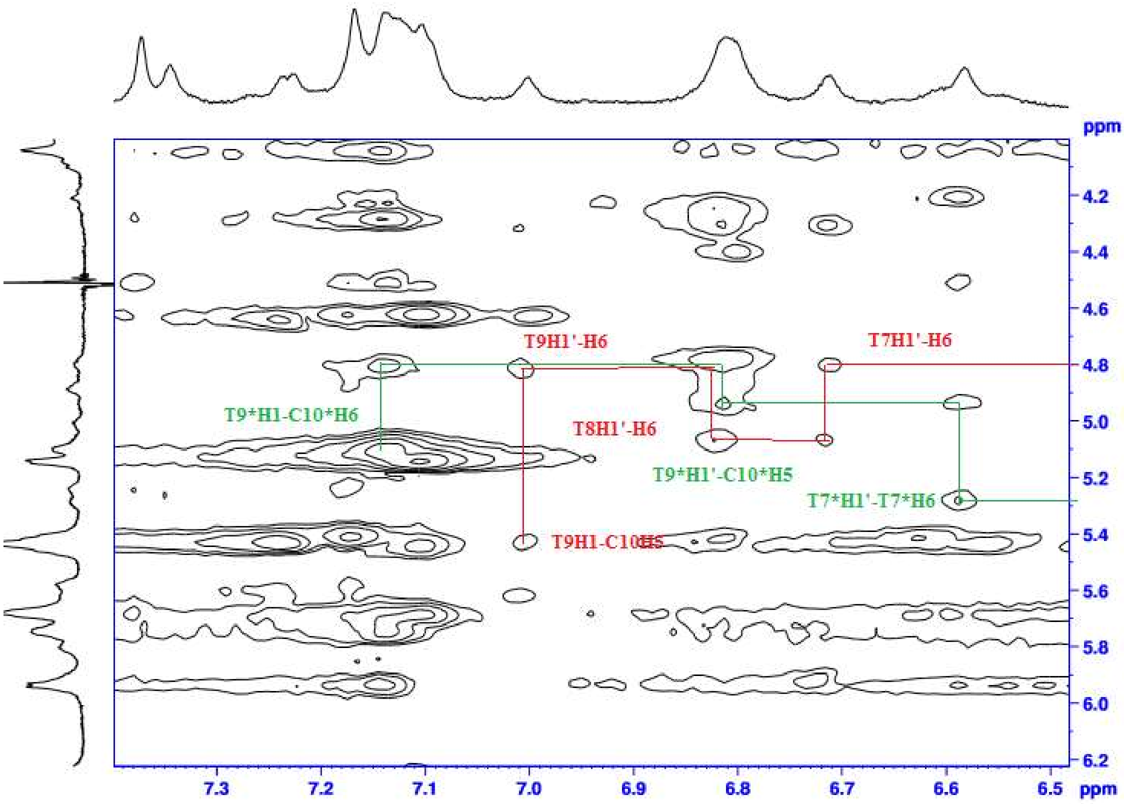


Figure 4.25. Assignment of the H6-H1' correlations in the T-tract of both strands from the NOESY spectrum. Some horizontal correlations are visible, but not shown.

This conformational shift is most likely to be a consequence of ligand binding and the thermodynamic cost of this conformational change could be a cause of the lower observed ΔT_m of **1.20** compared to **1.9**. (*Sections 3.3-3.4*).

The observed low nOe for the H6-H1' correlation in the T-tract also made continuation of the spectroscopic walk more difficult – and required the omission of the H1' resonances. Fortunately, a correlation between the T9H6 protons and both of the aromatic protons of C10 could be found, and hence, the protons corresponding to the C10 base could be assigned (Fig 4.26). As with the H1' proton of G3 (Strand 1) and G3* (Strand 2) (Fig. 4.8), the two aromatic protons and the H1' proton of C10 and C10* were perturbed with respect to each other, indicating that the ligand is interacting not only with the T-tract, but with the G-C base pair immediately adjacent to the T-tract.

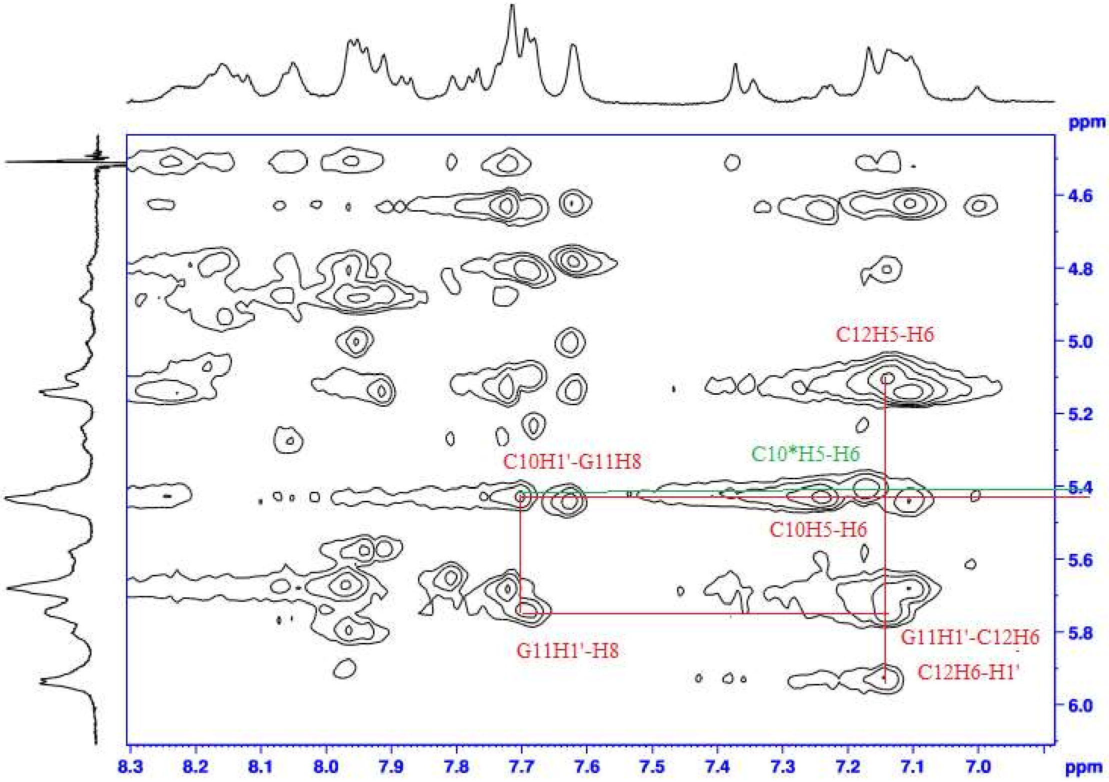


Figure 4.26. nOe assignment of T9-C12, and T9*-C12. T9 and T9*H1' have been omitted for clarity. The G11 and C12 bases of the complex were found to be equivalent.

The G11 and C12 bases showed no perturbation between each of the two strands and insignificant perturbation from free ODN 4.1, indicating that there is little effect on these two bases coming from binding of the ligand – further evidence that the A_3T_3 tract is being bound by the ligand.

Having completed the spectroscopic walk for both strands, the other sugar protons could now be assigned. As with free ODN 4.1, this was achieved by the correlation of the H1' protons to other sugar protons by TOCSY or from nOe correlations from the H1' proton by NOESY. (Appendix 7.17). Due to extensive peak overlap, very few of the H5'/H5'' protons of the complex were unambiguously assigned thus preventing the assignment of the H4’-H5’ correlations.

4.3.3. Assignment of the dsDNA minor groove protons.

Assignment of the protons in the minor groove of the **1.20**: ODN 4.1 complex was complicated by not only breaking of the dyadic symmetry of the duplex, but also by the presence of the ligand protons, which come between 7.5 and 8.9 ppm in free DMSO solution (Appendix 7.6). Parkinson *et al.* reported that there is a lack of nOe correlations between the imino protons of thymidine bases and ligand protons, they do exist between the H2 protons of adenosine and the H1’ protons of both adenosine and thymidine protons, allowing for unambiguous assignment of the adenosine H2 protons.[99] Moreover, the presence of the extra benzimidazole resonances at 11.5 and 12.6 ppm also complicated assignment of the spectrum. From close examination of previous NMR studies of bisbenzimidazole ligands in complex with dsDNA, it was found that no close contacts are observed between the **1.20** and the thymidine imino protons of ODN 4.1. As such, it was possible to discriminate the adenosine H2 protons from the ligand protons through correlation to the T imino protons (Fig. 4.16). Using the starting point of the previously assigned C2 and C10* amino protons, a minor groove “spectroscopic walk” was carried out as depicted in Fig. 4.17. The G3 and G11 imino protons can then be assigned, and then by a process of elimination the G1 protons could then be assigned at 12.84 ppm. This solvent-exposed proton shows no correlations and its integration is somewhat lower than for the other two bases – presumably some hydrogen-deuterium exchange occurs.

With the guanosine imino protons assigned, the spectroscopic walk was continued into the adenosine H2 and the thymidine imino protons. Breaking of the dyadic symmetry of ODN 4.1 caused by binding of **1.20** doubles the number of resonances in the spectrum (Fig. 4.18) and horizontal as well as vertical cross-peaks are also observed due to the narrow A-tract exhibited by ODN 4.1.[150] Consequently, care had to be taken to identify the strand on which the proton resides. With the assignment of the guanosine imino protons in hand, it was then possible to assign the imino protons from both the 5'- and 3'- end of the duplex (Fig. 4.28). The thymidine H1' resonances in the complex were difficult to unambiguously assign to either the thymidine imino or the adenosine H2 resonances due to a weak NOE signal. As such, they are not shown in Fig. 4.28. The remaining unassigned protons in the 10-14 ppm region at 11.51 and 12.58 ppm were, by a process of elimination and literature precedent, unambiguously assigned as the two benzimidazole protons.[99]

*4.3.4. Assignment of the protons of **1.20** in complex with ODN 4.1.*

Having assigned the H2 and imino protons of the complex of **1.20** with dsDNA, it was now possible to assign the protons belonging to **1.20**. With all other protons in the aromatic region of the spectrum now assigned (Fig. 4.18, Appendix 7.17), the remaining protons therefore belong to the ligand, and can be assigned. The NOESY spectrum of free **1.20** (Appendix 7.6) has three distinct spin systems, corresponding to the three aromatic moieties (Fig. 4.28 b). The three aromatic moieties are distant from one another, and from the benzimidazole protons. To obtain the orientation of binding of **1.20** the two key protons to assign were the benzylic protons H_1 and the amide proton H_A, (Fig. 4.28 a) which have NOE correlations to each of the rings, and have chemical shifts distinct from the rest of the molecule (Appendix 7.6).

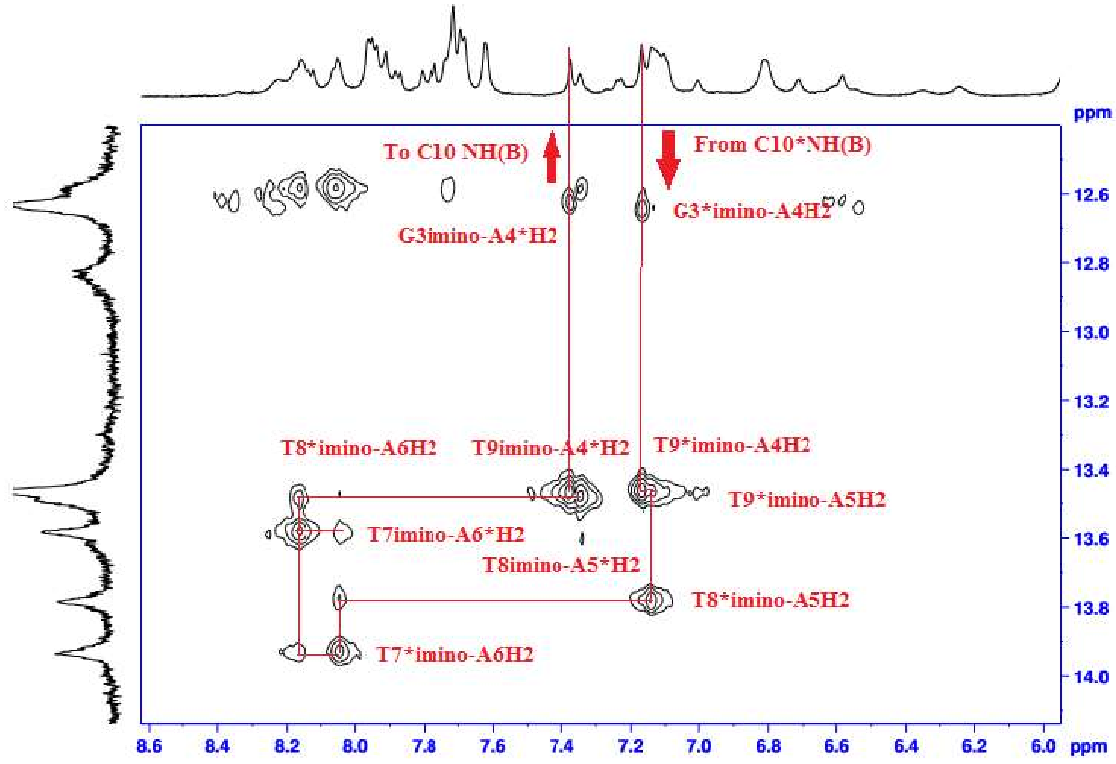


Figure 4.27. Assignment of the adenosine H2 and imino protons. One benzimidazole proton is visible at 12.58 ppm.

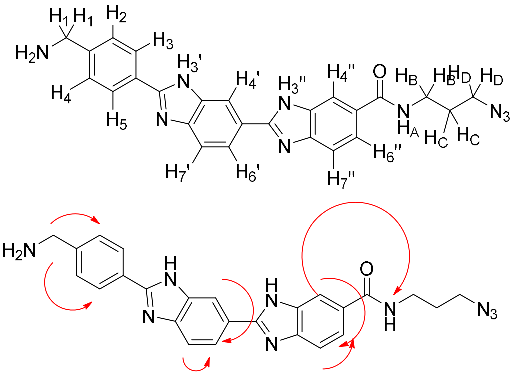
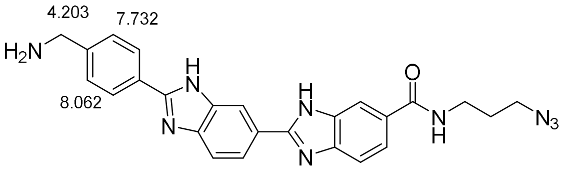


Figure 4.28. **(a)** Nomenclature of the protons in free **1.20**. **(b)** nOe correlations present in **1.20** (*Appendix 7.6*).

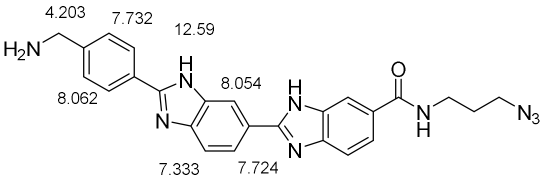
The first ligand protons assigned were the two benzimidazole protons NH3', and NH3'' at 11.51 and 12.59 ppm. Strong nOe correlations between the peak at 11.50 to one at 8.15 ppm and a corresponding correlation exists between the peak at 12.58 ppm with one at 8.05 ppm (Fig 4.30). The identical nOe volume of these correlations indicated that these resonances belonged to the H_4 protons of **1.20** (Fig. 4.28 b), and were assigned as such. Furthermore, H_1 could be assigned to be 4.203 ppm by a process of elimination from the fingerprint region, as the other aromatic peaks with correlations to peaks between 4.0 and 4.5 ppm had been assigned as belonging to ODN 4.1. Analysis of the NOESY and TOCSY spectrum around the peak allowed for assignment of the H_2 – H_5 protons (Fig. 4.29). As per literature precedent, the H_2 and H_3 protons and the H_4 and H_5 protons were found to have identical chemical shifts.[99]



1.20

Figure 4.29. Assignment of the benzylic protons and aromatic ring 1 (Fig. 4.6) of **1.20**

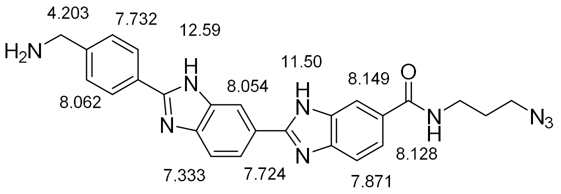
A weak correlation between the peak at 7.732 ppm and 12.59 ppm was observed, allowing for assignment of the H_3’ proton to the peak observed at 12.59 ppm. A strong nOe correlation between the H_3’ proton and the hitherto unassigned peak at 8.054 ppm was also observed. Hence, the peak at 8.054 ppm was assigned to the H_4’ proton (Fig. 4.30). Analysis of the nOe correlations between the H_4’ protons allowed for assignment of the H_6’ and H_7’ protons of the leftmost benzimidazole ring at 7.724 and 7.333 ppm respectively (Fig. 4.30).



1.20

Figure 4.30. Assignment of rings 1 and 2 (Fig. 4.6) of **1.20**.

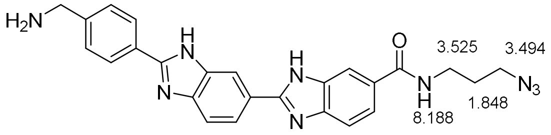
By a process of elimination, the peak at 11.50 ppm was assigned to H_3’’ (Fig. 4.31). By an identical process used with the previous ring, the H_4’’ proton was then assigned to 8.15 ppm. Two coupled protons were observed at chemical shifts of 8.13 ppm and 7.87 ppm, and although difficult due to signal degeneracy, a correlation between the peak at 8.13 and 8.15 ppm could be observed. Consequently, these resonances could be assigned to H_6'' and H_7'' respectively (Fig. 4.31).



1.20

Figure 4.31. Assignment of rings 1-3 of **1.20** in complex with ODN 4.1.

Assignment of the amide proton H_A was achieved through both identification of its TOCSY and nOe correlation with the H_4'' and H_6'' protons and comparison with the spectra in D_2O, which due to hydrogen/deuterium exchange, are absent from the D_2O spectrum. This proton was hence assigned to the peak at 8.19 ppm. A strong correlation to the peak at observed to a peak at 3.53 ppm and a weaker one was observed to a peak at 1.85 ppm, indicating that these resonances belonged to the H_B and H_C protons respectively. These resonances were similar to that observed for free **1.20** (3.45 and 1.83 respectively), and hence served as further evidence that these assignments were correct. Inspection of the correlations of these protons allowed for assignment of the final resonance, H_D at 3.49 ppm (Fig. 4.32).



1.20

Figure 4.32. Assignment of the alkyl tail of **1.20**. The aromatic assignments are as Fig. 4.31.

*4.3.5. Intermolecular nOe correlations between **1.20** and ODN 4.1.*

Once the assignments of all the chemical shifts of ODN 4.1 and **1.20** were complete, assignment of the binding mode of **1.20** to ODN 4.1 was possible. Definite nOe correlations were observed between ring 1 (Fig. 4.6) and surprisingly, A5*H2 which suggested that **1.20** binds with opposite directionality to both **1.9**[99] and **2.24**.[94] This would suggest that the primary amine in **1.20** is binding in a commensurate mode to the piperazine **1.9** and the tertiary amine in **2.24**. This observation showed that the main determinant of the orientation of Hoechst-type ligand binding to dsDNA is the position of the primary positive charge, not shape complementarity between the Hoechst core and the minor groove. This loss of shape complementarity is likely to be the origin of the lower observed ΔT_m of the **1.20**: ODN 3.2 complex when compared to both **1.9** and **1.21**.

Further evidence for this reversed binding mode comes from the nOe correlations between the benzimidazole protons and the centre of the A_3T_3 region. Contacts between H_3' (ring 2) and both A5* and A6* were observed, with A6* being significantly stronger. A similar pattern was observed between ring 3 and base A6. The amide proton H_A exhibited close nOe contacts between the A6 and A5 bases, and the tail showed correlations between A5 and A4. The observed intermolecular nOe correlations between **1.20** and ODN 4.1 are shown in Table 4.1 and Fig. 4.33.

These observations imply that binding of **1.20** to ODN 4.1 places the azide group between the A4 and G3 bases, and a specific interaction between the azide of **1.20** and the C2 amino group of G3 (located in the minor groove) is likely to be the origin of the changes in the azide IR band of **1.20** in complex with ODN 3.2.

Proton in 1.20	**Correlation to ODN 4.1**
H_2, H_3	A5*H2 (m), A5*H3’ (w)
H_4, H_5	A5*H2 (s), A5*H1’ (w), T7*H1’ (w)
H_3’	A6*H2 (m), A5*H2 (w)
H_4’	T7*H1’ (w)
H_3’’	A6H2 (m)
H_4’’	A6H2 (s), T7H1’ (w), T8*H1’ (m)
H_6’’	A6H2 (s)
H_A	T7*H1’ (w), T8*H1’ (w)
H_B	A5H2 (m)
H_C	A4H2 (s), A5H2 (s)
H_D	A4H2 (m), A5H2 (m)

Table 4.1. Intermolecular nOe effects observed for the **1.20**: ODN 4.1 complex. Indications of the strengths of the correlations are also given, (s) = strong > 1.5×10^6, (m) = medium $9 \times 10^5 < x < 1.5 \times 10^6$, (w) = weak, < 9×10^5 (units arbitrary). Resonances with no observed intermolecular crosspeaks are not listed.

4.4. Discussion

*4.4.1. Analysis of chemical shift perturbations of the **1.20**: ODN 4.1 complex.*

Analysis of chemical shift perturbations gives an indication of which protons are close in space to the ligand, and furthermore the relative orientation in space of a given ligand. Given the majority dsDNA-interacting small molecules are planar, aromatic compounds, they exhibit the ring currents characteristic of aromatic systems, and induce either an upfield or downfield shift, dependent on whether the proton in question is coplanar with the ring (shielded), or sits above or below the ring (deshielded).

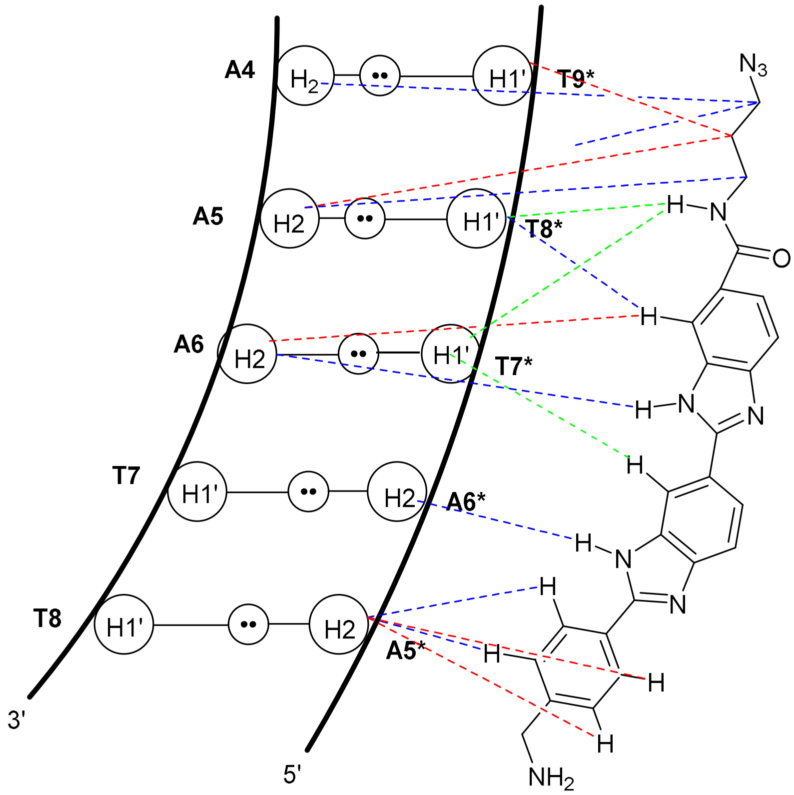


Figure 4.33. Diagram showing the observed intermolecular nOe observed for the **1.20**: ODN 4.1 complex. Strong nOe are depicted in red, medium in blue and weak are in green.

In the case of **1.9**, the H1' atoms of the A-T tract sit coplanar to **1.9** and are shielded while the H2 of adenines sit perpendicular to the ring, and are consequently deshielded (Fig 4.34). Some information about the binding mode can therefore, by comparison to previous work, give clues as to the binding mode of compound **1.20**.

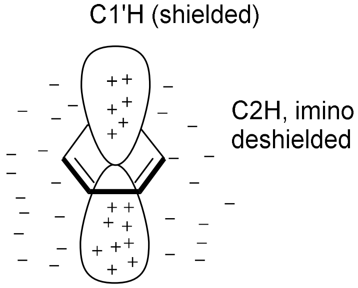


Figure 4.34. Pictorial representation of the local magnetic field perturbation induced by the ring currents due to the aromatic rings of **1.20**. Imino and adenosine C2H protons sit parallel to the ring and are deshielded, while C1'H protons sit perpendicular and are shielded.

Chemical shift perturbations for the H6/H8 protons, the H1', and H4' have been previously reported with an A_3T_3 tract with compound **2.24**, and comparison of the assignment of **1.20** to **2.24** should provide a good, if inexact idea of the binding mode of **1.20**.[94] The pattern of the H6/H8 perturbations of the complexes of **1.20** and **2.24** are in good agreement, showing that despite the observed differences in the binding mode of the ligand from the assignment (*Section 4.3.5*), the minor groove binding mode is largely retained (Fig. 4.35). The base pairs are numbered from the 5'- end as per Fig. 4.36.

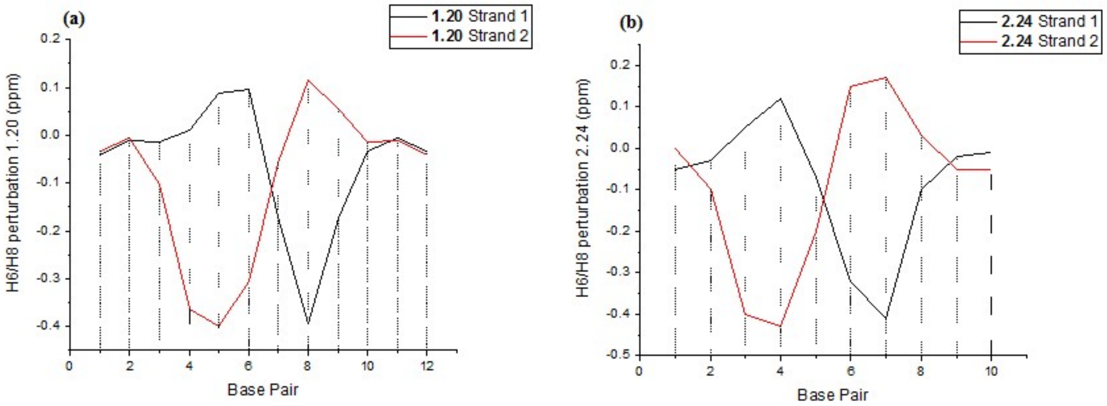


Figure 4.35. (a) H6/H8 Chemical shift perturbations found for the **1.20**:ODN 4.1 complex **(b)** Reported H6/H8 chemical shifts for the **2.24** complex by Bunkenborg *et al.*[94]

A similar pattern was observed for the H1' protons, and as shown in Fig. 4.34 the local ring currents exhibited by the aromatic rings of approaching **1.20** are the cause of the observed perturbations in the chemical shifts around the binding site.

Bunkenborg (2002) *1 2 3 4 5 6 7 8 9 10*
This work **1 2 3 4 5 6 7 8 9 10 11 12**
5'- G C G A A A T T T C G C - 3'
3'- C G C T T T A A A G C G - 5'

Figure 4.36. Base pair nomenclature in this section.

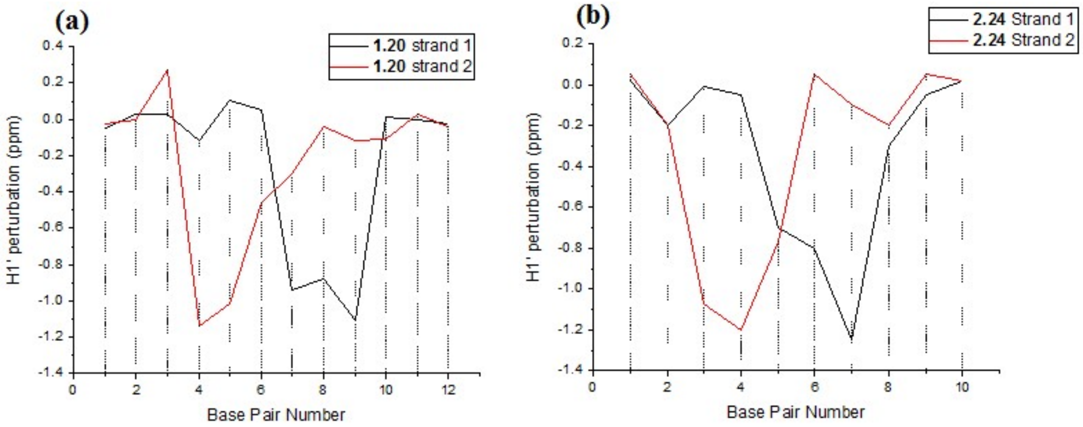


Figure 4.37. (a) Observed H1' perturbations for the **1.20**: ODN 4.1 complex. **(b)** Reported chemical shifts for the **2.24** complex.[94]

The stark perturbations observed for the **1.20**: ODN 4.1 complex (> 1 ppm in the thymidine bases) are typical for binding of bisbenzimidazole ligands to A-T tracts and are explained by the shielding effect of the local ring currents induced by binding of **1.20** (Fig. 4.37). Similar perturbations were reported for the **1.9** complex by Parkinson *et al.*[99] which were indicative of a minor groove binding mode to ODN 4.1. Perturbations were observed across all six base pairs of the A-T tract indicating that despite the antiparallel binding mode of **1.20** to the A_3T_3 tract, the minor groove binding mode is largely maintained. Given that A-T tracts of dsDNA are relatively conformationally rigid this finding points toward a two-state binding model (*Section 3.1)*, in that binding of the ligand causes a rearrangement of the conformation of ODN 4.1 to accommodate the approaching ligand. Due to the unusual structure and binding mode of **1.20**, however, this observation cannot be extended to other Hoechst-type ligands, or other sequences of dsDNA, although this observation may prove useful in interpreting any future 2D-IR experiments. Similar shielding effects between the **1.20** and **2.24** complexes were also observed for the H4' protons (Fig. 4.38). However, the chemical shift perturbations of the T8 and T9 H4' protons in this instance were more pronounced (> -2 ppm), consistent with the T7 H4' protons reported by Bunkenborg *et al.* (Fig. 4.38 b).[94]

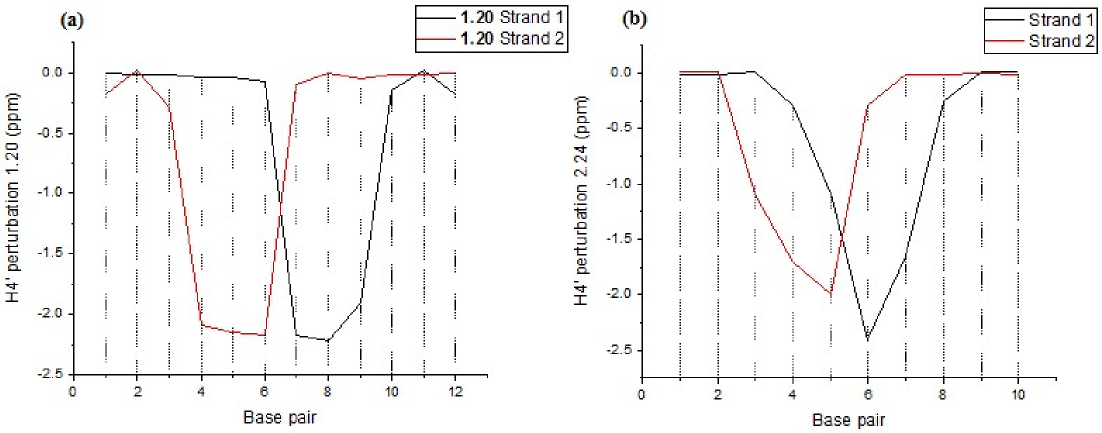


Figure 4.38. **(a)** Perturbation of the H4' protons in the **1.20**: ODN 4.1 complex. **(b)** Reported perturbations for the **2.24** complex reported by Bunkenborg *et al*.[94]

Analysis of the perturbations of the H4' protons showed that the binding mode of **1.20** to ODN 4.1 is somewhat different to that observed of **2.24**. From this data, all of the thymidine H4' protons in both strands are more shielded by the rings of **1.20** (Fig. 4.34) when compared to **2.24**. This observation therefore provides clues as to the structure of the **1.20**: ODN 4.1 complex.

The H2'/H2'' and H3' perturbations for **2.24** in complex with an A_3T_3 tract of dsDNA were not reported. They were, however, quoted by Parkinson *et al.* for the complex of **1.9** with the self-complementary dsDNA oligomer, ODN 4.2 (Fig. 4.39).[99]

5'-GCGCAATTCGCG-3'
3'-CGCGTTAAGCGC-5'

Figure 4.39. Structure of ODN 4.2.

Although the binding site is A_2T_2 not A_3T_3, the main binding site is preserved, and comparison of the shifts observed for the complex of **1.9** to ODN 4.2, which has been crystallised (Fig. 2.2) to that observed for **1.20** to ODN 4.1 – and may provide some clues as to the binding mode. Again, the patterns of the binding mode are somewhat similar, despite the change in the structure of both the compound and the oligomer (Fig. 4.40).

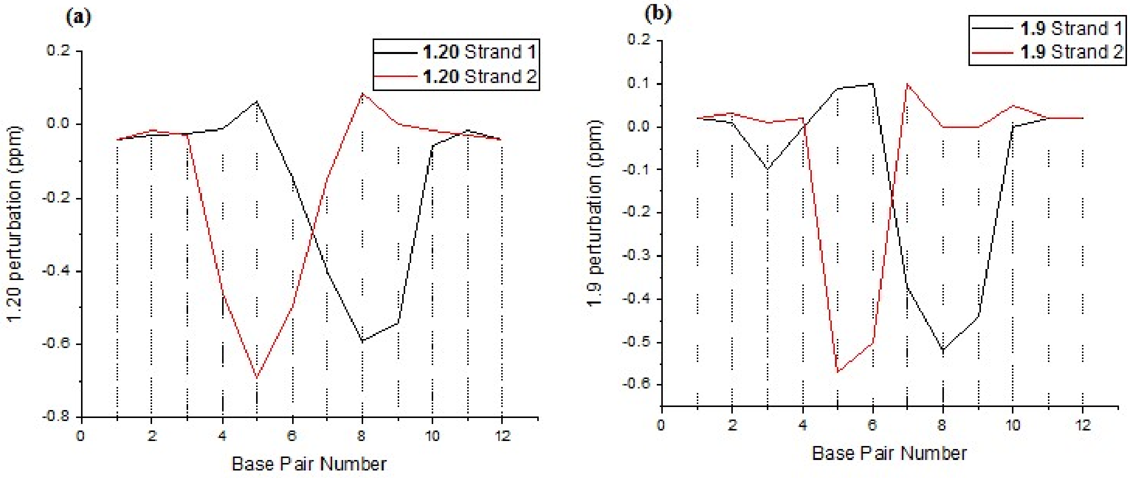


Figure 4.40. **(a)** Chemical shift perturbation of the H2’ protons in the **1.20**: ODN 4.1 complex. **(b)** H2’ perturbations in the **1.9**: ODN 3.2 complex

Similar perturbations were observed in the **1.20**: ODN 4.1 complex about the A-T base pairs compared to the **1.9**: ODN 4.2 complex, indicating a broadly similar binding mode, although **1.20**, like **2.24** spans 5-6 base pairs, instead of 4 for **1.9**. Given that the H2’ and H2’’ are on the same carbon atom, similar perturbations were unsurprisingly observed for the H2’’ protons (Fig. 4.41).

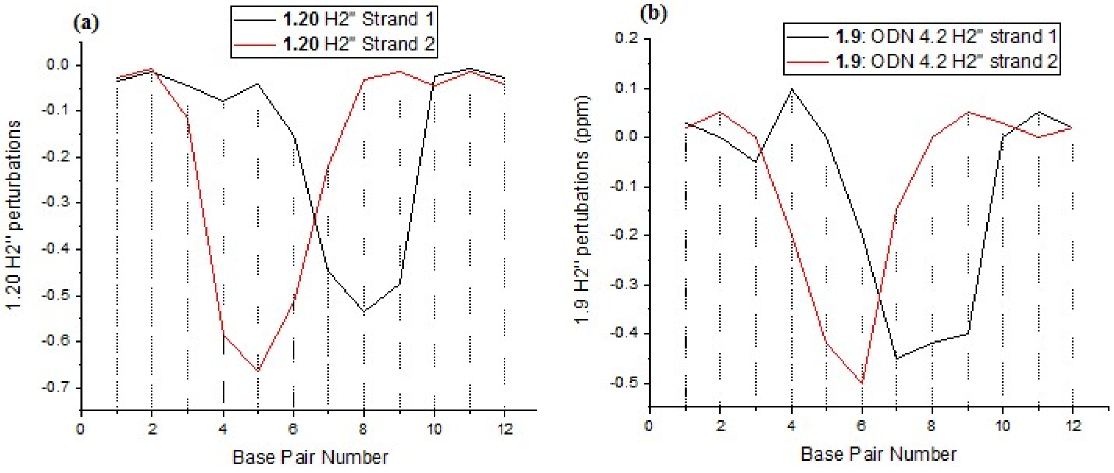


Figure 4.41. **(a)** H2’’ chemical shift perturbations observed for the **1.20**: ODN 4.1 complex. **(b)** H2’’ chemical shift perturbations reported for the **1.9**:ODN 4.2 complex.[99]

Again, the observed chemical shift perturbations for the H2'' protons exhibit a pattern consistent with a minor groove binding mode, although the values of the chemical shifts are different, this is likely due to subtle differences between the structures of the **1.20**: ODN 4.1 complex and the **1.9**: ODN 4.2 complex, whether this is due to the differences in the structure between **1.20** and **1.9** or the different binding sites between ODN 4.1 (A_3T_3) and ODN 4.2 (A_2T_2) is unclear.

Similar perturbation patterns were observed for the H3' protons of the **1.20**: ODN 4.1 complex compared to the **1.9**: ODN 4.2 complex (Fig. 4.42).

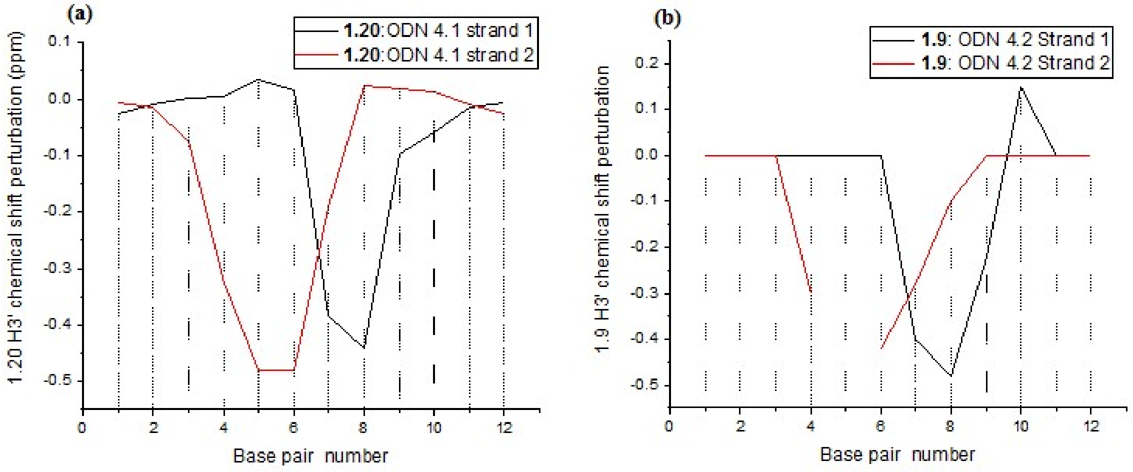


Figure 4.42. Comparison of chemical shift perturbation of H3' protons of the **1.20:** ODN 4.1 complex compared to the **1.9**: ODN 4.2 complex (one perturbation was not reported).[99]

The perturbation patterns observed were broadly similar, although a difference at the 3'- end of ODN 4.2 caused by binding of **1.9** was reported by Parkinson *et al.,* presumably due to a conformational shift in ODN 4.2 to accommodate the bulky piperazino moiety. This observation indicated that the azide in **1.20** is likely to be easily accommodated in the minor groove of ODN 4.1, similarly to the tertiary amine in **2.24**.[94]

*4.4.2. Evidence for a single binding mode of **1.20** to ODN 4.1.*

The one-dimensional NMR spectrum of the **1.20**: ODN 4.1 complex showed resonance doubling in the region G3-C10, consistent with a minor groove binding mode of an unsymmetrical ligand across these six base pairs. Although no intermolecular nOe have been characterized for the A4*-T9 binding mode, it is likely that the primary amine is interacting with ODN 4.1 in this region. Furthermore, **1.20** was observed to form a well-defined complex by NMR – if **1.20** formed a second binding mode with similar affinity to a minor groove binding mode, a different pattern to resonance doubling would be observed in the imino and thymidine methyl regions. Consequently, at high salt: ligand ratios, the minor groove binding mode described in *Section 4.3.4* is the sole binding mode observed under normal conditions.

*4.4.3 Comparison of the binding modes of **1.20** and **2.24***

Given the different directionalities of the binding modes of **1.20** and **2.24** it is striking that the chemical shift perturbations of **1.20** and **2.24** are similar, and are consistent with a similar binding mode. Further evidence for this comes from the intermolecular nOe patterns (Figs. 4.33 and 4.43), which suggest that bifurcated hydrogen bonds are formed between the central three base pairs of ODN 4.1, although in different directions – the amide points toward the 3'-end of the duplex in the **2.24** complex and the 5'- end of the duplex in the **1.20**: ODN 4.1 complex. Given the ability of the primary amine to form an electrostatic interaction with the narrow minor groove of the 3'- end of the A-T tract, it seems somewhat surprising that a single binding mode was observed for **2.24**.[94]

This observation can be explained through the fact that the primary amine in **1.20** (and **2.24**) is more solvated than either the tertiary amine in **2.24** or the piperazine in **1.9**.[151] Hence, desolvation of the cation in order to enter the minor groove is more thermodynamically unfavourable than for the tertiary amine in **2.24**. The significantly higher thermodynamic cost of desolvation of primary amines compared to more diffuse cations such as a guanidine or a tertiary amine has been previously proposed as a reason why arginine residues are almost exclusively found in the DBD of minor-groove contacting proteins.[151] The thermodynamic cost of desolvation is likely to be a factor in the lower ΔT_m observed for the **1.20**: ODN 3.2 complex compared to either **1.9** or **1.21**.

It was reported by Bunkenborg *et. al.* that **2.24** fitted neatly into the centre of the minor groove of the A_3T_3 tract pointing towards a "lock-and-key" mechanism (Fig. 4.43).

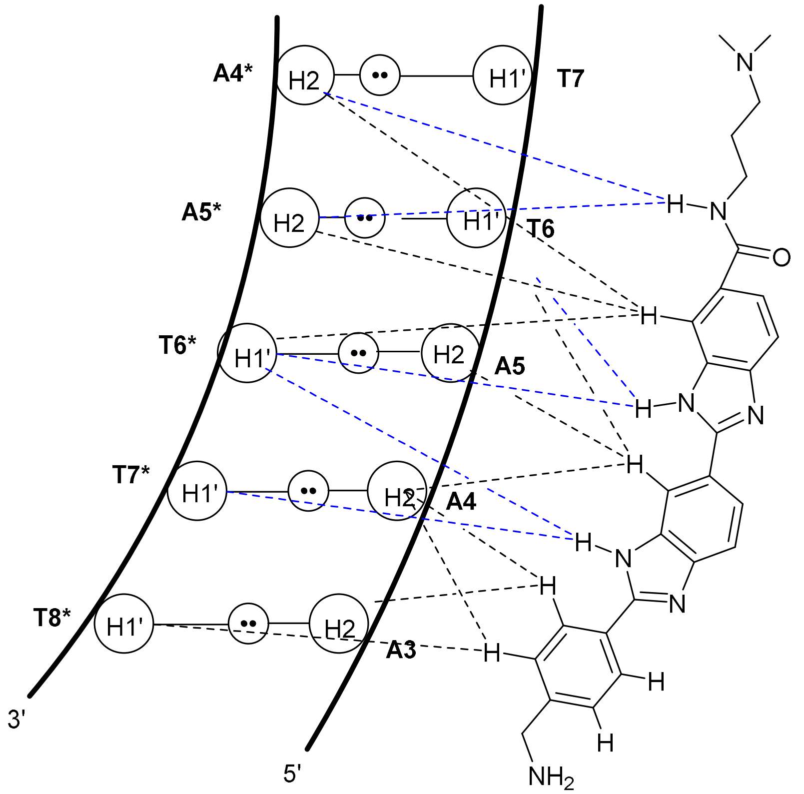


Figure 4.43. Binding mode of **2.24** to an A_3T_3 tract of dsDNA. Close nOe contacts are represented by dashed black lines, bifurcated hydrogen bonds in blue. The relative strengths are not indicated.[94]

This means that in that the shape of **2.24** is complementary to that of the minor groove of dsDNA, and the binding surface of A-T tracts can be considered "pre-formed" for an incoming Hoechst ligand.[84,94] Given the observed antiparallel binding mode of **1.20** to ODN 4.1, it stands to reason that the shape complementarity of the Hoechst ligand to the minor groove of B- type dsDNA is also lost. Hence, in order to accommodate a molecule of **1.20** in its minor groove, a significant conformational shift in ODN 4.1 is required. This observation pointed towards a two-state model (Fig. 3.3) for the binding of **1.20** to ODN 4.1 in that a weak binding interaction between **1.20** and ODN 4.1 is sufficient to induce a change in the conformation of ODN 4.1 such that stronger intermolecular contacts were possible. The conformational change in ODN 4.1 is almost certainly a cause of the lower observed ΔT_m for the **1.20**: ODN 3.2 complex when compared to **1.9** or **1.21**. The conformational change in ODN 4.1 is also likely to be responsible for the observed precipitation of the **1.20**: ODN 4.1 at low buffer: complex concentration ratio.

*4.4.4. The optimum binding site of **1.20** to dsDNA.*

Referring back to the DNase I footprinting data, the digestion patterns of **1.20** and **2.24** are identical, and are comprised of consecutive runs of at least four A-T base pairs (Figs. 3.6 – 3.8).[139] As per **2.24**, both 5'-AAAA-3' and 5'-AATT-3' are preferred binding modes over either 5'-ATATAT-3' or 5'-TTAA-3' runs. This data showed that the preferred binding site of **1.20** is still AATT over TTAA or ATAT, despite its reversed orientation in the minor groove relative to either **1.9** or **2.24**. The consensus binding sequence is therefore maintaned, and the lower value of ΔT_m obtained for **1.20** indicated that it was a lower affinity ligand than either **1.9** or **1.21**.

The binding mode of **1.20** to ODN 4.1 described in this chapter showed that the origins of the lower ΔT_m of **1.20** to ODN 3.2 (*Section 3.5*) is due to this observed reversed orientation, and not, as first thought, a result of the removal of the cationic moiety from ring 3. The reversed orientation retains the electrostatic interaction between **1.20** and the narrow minor groove presented by ODN 4.1, and this is likely to be the origin of the reversed orientation of **1.20** in the minor groove of ODN 4.1 compared to both **1.9** and **2.24**. As such, the origin of the lower ΔT_m of **1.20** compared to **1.9** or **1.21** is due to the fact that ODN 4.1 undegoes a conformational shift in order to accommodate a molecule of **1.20**, and this incurs a significant thermodynamic cost.

4.5 Conclusions and future work

4.5.1. Conclusions

The structure of **1.20** in complex with an A-T tract has been solved by NMR. Observation of intermolecular nOe crosspeaks have indicated that although **1.20** clearly sits across the centre of the A_3T_3 tract of ODN 4.1, which differs from the orientation of the molecule previously reported for compounds **1.9** and **2.24**. Significant conformational changes in the minor groove were required in order to accommodate the binding ligand. The H_D protons of **1.20** have also been shown to sit above the terminal A4 residue of ODN 4.1, indicating that the azide is in close proximity to the exocyclic amine of guanosine residue G3, and this interaction is the origin of the observed changes in the azide response of **1.20** in complex with ODN 3.2 when compared to the free ligand. The consensus binding sequence of **1.20** was also compared to the DNase I footprinting work presented in *section 3.2* and was observed to be identical to that of **2.24**. This conformational change has also been posited as the origin of precipitation of the complex at high complex: salt concentration ratios.

4.5.2. Future work

A three-dimensional model of the **1.20**: ODN 4.1 complex will be determined in order to gain a deeper understanding of the three-dimensional structure. Future work will then be directed toward understanding the dynamic changes of **1.20** and **1.21** in complex with target, as well as sub-optimal dsDNA sequences, most notably 5’-ATATAT-3’, and 5’-AATT-3’, as well as the other binding sequences suggested by the digestion pattern in Figs. 3.6 – 3.8.

4.6. Experimental

4.6.1. Synthesis and purification of ODN 4.1.

The self-complementary DNA oligonucleotide d(GCGAAATTTCGC)$_2$ (ODN 4.1) was synthesized using standard solid-phase protocols on a 20 µM scale, using reagents and columns purchased from LINK Technologies Incorporated. All solvents were purchased from Sigma-Aldrich, and used without further purification, furnishing the DNA oligomer in 70 % crude purity. The DNA oligomers were purified by RP-HPLC on a Dionex ULTIMate 3000 system equipped with a Phenomenex Clarity column, 250 mm in length, with 5 µm pore size and 10.0 mm internal diameter. The elution buffer used was 100 mM triethylammonium acetate in water (buffer A) and 100 mM triethylammonium acetate in 80/20% v/v acetonitrile/water (Buffer B), and the proportion of solvent B increased from 5% to 40% over 22 minutes. The crude mixture was then desalted using a GE Healthcare NMP-25 column and lyophilized to dryness. The desalting procedure was then repeated as necessary until the molar ratio of residual buffer salts to dsDNA was lower than 0.5:1 by ^{1}H-NMR. 8.5 mg of sample was obtained (23.3 % overall yield) and was characterized by MALDI, analytical RP-HPLC and ^{1}H-NMR

4.6.2. Preparation of samples for NMR.

NMR samples were prepared through dissolution of the dsDNA into a 100 mM phosphate buffer at pH 7.0 in either D_2O (99.9 %, Sigma, Aldrich) or a mixture of 90% Millipore H_2O/ 10% D_2O, for a final concentration of 2 mM dsDNA (4mM ssDNA) of volume 550 µL. The mixture was then annealed at 80 °C and allowed to cool to room temperature before the clear solution was introduced to the NMR tube by pipette. The solvent system was changed from

D_2O to H_2O or *vice versa* through lyophilisation of the sample, and redissolution of the residue in the desired solvent.

A 100 mM solution of the ligand **1.20** in DMSO-d^6 was prepared, and then added to the sample of dsDNA in H_2O/D_2O, in approximately 25 mol% aliquots and then annealed at 80 °C until a clear brown solution was observed. The mixture was checked by ^{1}H-NMR to confirm formation of the full complex, and then lyophilized to remove the deuterated DMSO. The solid obtained was then resuspended in 550 μL of either D_2O or 90% H_2O/10% D_2O as required. To change the solvent, the sample was washed into a falcon tube with a little Millipore water, lyophilised to dryness and then redissolved in the desired solvent. An internal NMR standard (2, 2, -dimethyl-2-silapentane-5-sulfonate, DSS) was added prior to the collection of all spectra, and all chemical shifts are reported relative to this standard.

4.6.3. NMR experiments.

All NMR experiments were performed on a Bruker AVANCE III 600 MHz spectrometer with a mixing time of 150 ms. NOESY experiments on the free DNA duplex were carried out at 298 K and 303 K for the complex. A spectral width of 6000 Hz (10 ppm) was used for experiments in D_2O and 12000 Hz (20 ppm) for the experiments in H_2O were used. 4096 complex points in t_2 were used for a total of 512 t_1 experiments with 64 scans at the given temperature. Spectra were collected in H_2O using a WATERGATE-NOESY pulse sequence.[152]

TOCSY experiments were carried out with T_w of 70 and 20 ms using a total of 256 scans (128 dummy scans and 128 scans). A spectral width of 12 ppm was used (7200 Hz) and no isotopic labelling was employed.

Once processed, the processed spectrum was analysed using the SPARKY program (University of California, San Francisco).[153]

5. General conclusions and Future Directions.

5.1 General conclusions.

Through the course of this project, two derivatives of Hoechst 33258 (**1.9**), bearing an azide functional group to act as an IR reporter has been synthesised, compounds **1.20** and **1.21**. Thermal melting experiments were carried out on the two compounds in complex with dsDNA. The azide absorption band of **1.20** was markedly different to **1.21**, whereas the 1600 – 1700 cm^{-1} regions for the two compounds was very similar. This observation showed that the melting mechanisms were very similar. The T_m for the IR melting experiments corresponded exactly to the T_m as investigated by UV-Vis spectroscopy, this observation showed that both the UV-Vis and IR melts were sensitive to the same structural feature of dsDNA, namely Watson-Crick base pairing.

The azide absorption bands of the **1.20**: ODN 3.2 and the **1.21**: ODN 3.2 complexes were markedly different: the **1.20**: ODN 3.2 azide band exhibited an increase in absorption, and the presence of a red shoulder peak, indicative of a hydrogen bonding interaction with the minor groove of ODN 3.2. Further evidence of this came from structure elucidation of the **1.20**: ODN 4.1 complex by NMR. Analysis of the NOESY correlations indicated that the azide sits at the 5'-end of the minor groove, and is in close proximity to the G3-C10* base pair of ODN 4.1 (Fig. 4.18). As such, it can be expected that there is a hydrogen bonding interaction between the exocyclic amine of G3 and the azide of **1.20**. The change in the peak shape and absorbance of the azide vibration exhibited a sigmoidal peak profile commensurate with the T_m obtained for both UV-Vis and the 1600 – 1700 cm^{-1} region of the IR spectrum.

The azide peak profile of the **1.21**: ODN 3.2 complex was very different, the peak shape did narrow when in complex with ODN 3.2 compared to free **1.21**, there was no dramatic change in peak shape as observed for the **1.20** complex. Moreover, unlike **1.20** the azide absorption band of the **1.21**: ODN 3.2 complex did not exhibit a sigmoidal profile upon melting. The reasons for this are unclear, as significant problems existed with the purification of sufficient amounts of **1.21** suitable for spectroscopic experiments, thus limiting the amount of data that could be collected from this complex. Consequently, direct comparison of the environment experienced by the azides of **1.20** and **1.21** was impossible, and furthermore, fewer spectroscopic experiments could be carried out on the **1.21**: ODN 3.2 complex.

5.2. Future directions.

Firstly, in order to serve as a direct comparison of the local environments of the azides in **1.20** and **1.21**, a satisfactory synthetic route towards pure **1.21** has to be developed, investigating the proposed methods in Scheme 2.13. Once this compound has been prepared, it should be subjected to the same tests as described in Chapter 3, in order to provide a full data set of UV and IR experiments, leading into 2D-IR experiments. A structure of **1.21** in complex with ODN 4.1 should also be obtained, either by NMR or X-ray crystallography to provide a structural explanation for the different behaviour of the azide functional group in **1.21** compared to **1.20**. The UV melting curves should be inspected closely in order to extract both the enthalpic and entropic contributions to the binding interaction of **1.20** and **1.21** to dsDNA.[145]

Once a full data set has been acquired for both **1.20** and **1.21**, the same set of experiments described in Chapters 3 and 4 could be carried out for other binding sites to dsDNA, for example 5'-ATATAT-3' and other combinations thereof, of which **1.20** is known to bind (Figs. 3.6 – 3.8). Comparing the dynamics of association of **1.20** and **1.21** to various sequences A-T rich of dsDNA will provide a fuller picture of the molecular processes governing association of these two ligands to A-T rich binding sites, and hopefully, the data will inform on the design of better DNA-binding therapeutics.

The techniques described in this thesis could be extended to other classes of MGBs for example a derivative of distamycin A (**1.8**) **5.1** (Fig. 5.1). Analysis of the 2D-IR spectrum of **5.1** in complex with 5'-ATAAT-3', 5'-AAATTT-3' – with which **1.8** forms a 2:1 complex and 5'-AAAAA-3' with which **1.8** forms a 1:1 complex (Fig. 1.20).[43]

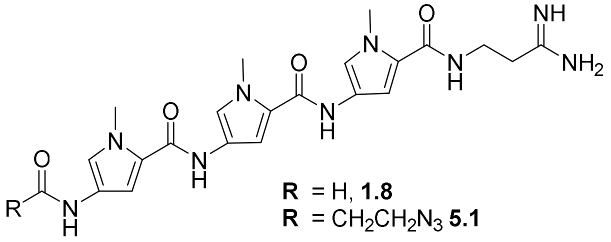


Figure 5.1. Proposed azide-labelled derivative of **1.8**, **5.1**.

Analysis of the binding dynamics of **5.1** in complex with dsDNA will firstly provide insights into the factors that determine the stoichiometry of binding of **5.1** to dsDNA. Secondly, comparative analysis of the dynamics of the interaction between dsDNA and both **5.1**, **1.20**, and **1.21** will allow for the design of MGBs with optimised thermodynamic profiles, with optimised enthalpic and entropic contributions. This same approach could also be extended to the Dervan pyrrole-imidazole hairpin polyamides (Fig. 5.2) with azides located on either the N-terminus (**5.2**), or the hairpin "turn" (**5.3**).[154]

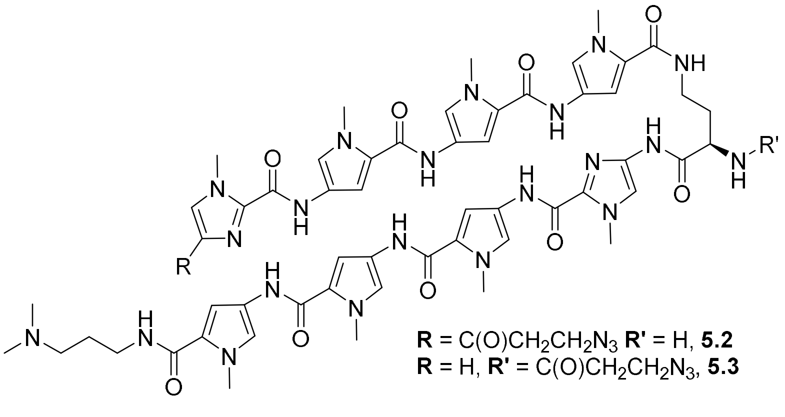


Figure 5.2. Proposed structures of Dervan polyamides labelled with azides on the N-terminus (**5.2**) and on the hairpin turn motif (**5.3**)

6. References

(1) Dervan, P. B.; Doss, R. M.; Marques, M. M. Programmable DNA Binding Oligomers for Control of Transcription. *Curr. Med. Chem.* **2005**, *5* (4), 373–387.

(2) Dervan, P. B.; Edelson, B. S. Recognition of the DNA Minor Groove by Pyrrole-Imidazole Polyamides. *Curr. Opin. Struct. Biol.* **2003**, *13* (3), 284–299.

(3) Chenoweth, D. M.; Harki, D. A.; Phillips, J. W.; Dose, C.; Dervan, P. B. Cyclic Pyrrole - Imidazole Polyamides Targeted to the Androgen Response Element Cyclic Pyrrole-Imidazole Polyamides Targeted to the Androgen Response Element. *J. Am. Chem. Soc.* **2009**, *131*.

(4) Khalaf, A. I.; Bourdin, C.; Breen, D.; Donoghue, G.; Scott, F. J.; Suckling, C. J.; Macmillan, D.; Clements, C.; Fox, K.; Sekibo, D. A. T. Design, Synthesis and Antibacterial Activity of Minor Groove Binders: The Role of Non-Cationic Tail Groups. *Eur. J. Med. Chem.* **2012**, *56*, 39–47.

(5) Bostock-Smith, C. E.; Harris, S. A.; Laughton, C. A.; Searle, M. S. Induced Fit DNA Recognition by a Minor Groove Binding Analogue of Hoechst 33258: Fluctuations in DNA A Tract Structure Investigated by NMR and Molecular Dynamics Simulations. *Nucleic Acids Res.* **2001**, *29* (3), 693–702.

(6) Adamczyk, K.; Candelaresi, M.; Kania, R.; Robb, K.; Bellota-Anton, C.; Greetham, G. M.; Pollard, M. R.; Towrie, M.; Parker, A. W.; Hoskisson, P. A.; et al. The Effect of Point Mutation on the Equilibrium Structural Fluctuations of Ferric Myoglobin. *Phys. Chem. Chem. Phys.* **2012**, *14* (20), 7411–7419.

(7) Bloomfield, V. A.; Crothers, D. M.; Tinoco, I. *Nucleic Acids: Structures, Properties and Functions*, 1st Editio.; University Science Books, 2000.

(8) Alberts, B. *Molecular Biology of the Cell*.

(9) Watson, J. .; Crick, F. H. . Molecular Structure for Nucleic Acids: A Structure for Deoxyribonucleic Acids. *Nature* **1953**, *171* (4356), 737–738.

(10) Wang, A. H.-J.; Quigley, G. J.; Kolpak, F. J.; Crawford, J. L.; van Boom, J. H.; van der Marel, G.; Rich, A. Molecular Structure of a Left-Handed Double Helical DNA Fragment at Atomic Resolution. *Nature* **1979**, *282* (5740), 680–686.

(11) Dickerson, R. E.; Drew, H. R.; Conner, B. N.; Wing, R. M.; Fratini, A. V; Kopka, M. L. The Anatomy of A-, B-, and Z-DNA. *Science (80-.).* **1982**, *216* (4545), 475–485.

(12) Mandal, P. K.; Venkadesh, S.; Gautham, N. Structure of the Tetradecanucleotide d(CCCCGGTACCGGGG)$_2$ as an A-DNA Duplex. *Acta Crystallogr. Sect. F* **2012**, *68* (4), 393–399.

(13) Privé, G. G.; Yanagi, K.; Dickerson, R. E. Structure of the B-DNA Decamer d(CCAACGTTGG)$_2$ and Comparison with Isomorphous Decamers d(CCAAGATTGG)$_2$ and d(CCAGGCCTGG)$_2$. *J. Mol. Biol.* **1991**, *217* (1), 177–199.

(14) Wei, D.; Wilson, W. D.; Neidle, S. Small-Molecule Binding to the DNA Minor Groove Is Mediated by a Conserved Water Cluster. *J. Am. Chem. Soc.* **2013**, *135* (4), 1369–1377.

(15) Falk, M.; Hartman, K. A.; Lord, R. C. Hydration of Deoxyribonucleic Acid. II. An Infrared

Study. *J. Am. Chem. Soc.* **1963**, *85* (4), 387–391.

(16) Edwards, K. J.; Brown, D. G.; Spink, N.; Skelly, J. V.; Neidle, S. Molecular Structure of the B-DNA Dodecamer d(CGCAAATTTGCG)$_2$ An Examination of Propeller Twist and Minor-Groove Water Structure at 2·2Åresolution. *J. Mol. Biol.* **1992**, *226* (4), 1161–1173.

(17) Frietze, S.; Farnham, P. J. Transcription Factor Effector Domains. In *A Handbook Of Transcription Factors*; Hughes, T. R., Ed.; Springer, 2011; pp 261–278.

(18) Slattery, M.; Zhou, T.; Yang, L.; Machado, A. C. D.; Gordân, R.; Rohs, R. Absence of a Simple Code: How Transcription Factors Read the Genome. *Trends Biochem. Sci.* **2014**, *39* (9), 381–399.

(19) Rohs, R.; Jin, X.; West, S. M.; Joshi, R.; Honig, B.; Mann, R. S. Origins of Specificity in Protein-DNA Recognition. *Annu. Rev. Biochem.* **2010**, *79*, 233–269.

(20) Bateman, A.; Birney, E.; Cerruti, L.; Durbin, R.; Etwiller, L.; Et.al. The Pfam Protein Families Database. *Nucleic Acids Res.* **2002**, *30*, 276–280.

(21) Klug, A. The Discovery of Zinc Fingers and Their Applications in Gene Regulation and Genome Manipulation. *Annu. Rev. Biochem.* **2010**, *79* (213–231).

(22) Miller, J.; McLachlan, A. D.; Klug, A. Repetitive Zinc-Binding Domains in the Protein Transcription Factor III A from Xenopus Oocytes. *J. Embryol.* **1985**, *4*, 1609–1614.

(23) Ginsberg, A. M.; King, B. O.; Roeder, R. G. Xenopus 5S Gene Transcription Factor, TFIIIA: Characterisation of a cDNA Clone and Measurement of RNA Levels throughout Development. *Cell* **1984**, *39*, 479–489.

(24) Rohs, R.; West, S. M.; Sosinsky, A.; Liu, P.; Mann, R. S.; Honig, B. The Role of DNA in Shape in Protein-DNA Recognition. *Nature* **2009**, *461*, 1248–1253.

(25) Paveltich, N. A.; Pabo, C. O. Zinc Finger-DNA Recognition: Crystal Structure of a Zif268-DNA Complex at 2.1 Angstrom Resolution. *Science.* **1991**, *252*, 809–816.

(26) Klug, A.; Jack, A.; Viswamitra, M. A.; Kennard, O.; Shakked, Z.; Steitz, T. A. A Hypothesis on a Specific Sequence-Dependent Conformation of DNA and Its Relation to the Binding of the Lac-Repressor Protein. *J. Mol. Biol.* **1979**, *131* (4), 669–680.

(27) Zhang, Y. P.; Zho, J. H.; Wang, L. Roles of the bZIP Gene Family in Rice. *Genet. Mol. Res.* **2014**, *13* (2), 3025–3036.

(28) Parraga, A.; Bellsolell, L.; Ferré-D'Amaré, A. R.; Burley, S. K. Co-Crystal Structure of Sterol Regulatory Element Binding Protein 1a at 2.3 Angstrom Resolution. *Structure* **1998**, *6* (5), 661–672.

(29) Burley, S. K. The TATA Box Binding Protein. *Curr. Opin. Struct. Biol.* **1996**, *6* (1), 69–75.

(30) Kim, J. L.; Nikolov, D. B.; Burley, S. K. Co-Crystal Structure of TBP Recognizing the Minor Groove of a TATA Element. *Nature* **1993**, *365* (6446), 520–527.

(31) Chen, Y.; Bates, D. L.; Dey, R.; Chen, P.-H.; Machado, A. C. D.; Laird-Offringa, I. A.; Rohs, R.; Chen, L. DNA Binding by GATA Transcription Factor Suggests Mechanisms of DNA

Looping and Long-Range Gene Regulation. *Cell Rep.* **2015**, *2* (5), 1197–1206.

(32) Otwinowski, Z.; Schevitz, R. W.; Zhang, R.-G.; Lawson, C. L.; Joachimiak, A.; Marmorstein, R. Q.; Luisi, B. F.; Sigler, P. B. Crystal Structure of Trp Represser/operator Complex at Atomic Resolution. *Nature* **1988**, *335* (6188), 321–329.

(33) Chen, Y.; Bates, D. L.; Dey, R.; Chen, P.-H.; Machado, A. C. D.; Laird-Offringa, I. A.; Rohs, R.; Chen, L. DNA Binding by GATA Transcription Factor Suggests Mechanisms of DNA Looping and Long-Range Gene Regulation. *Cell Rep.* **2012**, *2* (5), 1197–1206.

(34) Lazarovici, A.; Zhou, T.; Shafer, A.; Dantas Machado, A. C.; Riley, T. R.; Sandstrom, R.; Sabo, P. J.; Lu, Y.; Rohs, R.; Stamatoyannopoulos, J. A.; et al. Probing DNA Shape and Methylation State on a Genomic Scale with DNase I. *Proc. Natl. Acad. Sci.* **2013**, *110* (16), 6376–6381.

(35) Zhou, T.; Yang, L.; Lu, Y.; Dror, I.; Dantas Machado, A. C.; Ghane, T.; Di Felice, R.; Rohs, R. DNAshape: A Method for the High-Throughput Prediction of DNA Structural Features on a Genomic Scale. *Nucleic Acids Res.* **2013**, *41* (W1), W56–W62.

(36) Olson, W. K.; Gorin, A. A.; Lu, X. J.; Hock, L. .; Zhurkin, V. B. DNA-Sequence Dependent Deformability Deduced from Protein-DNA Crystal Complexes. *Proc. Natl. Acad. Sci.* **1998**, *95*, 11163–11198.

(37) Rohs, R.; Sklenar, H.; Shakked, Z. Structural and Energetic Origins of Sequence-Specific DNA Bending: Monte Carlo Simulations of Papillomavirus E2-DNA Binding Sites. *Structure* **2005**, *13*, 1499–1509.

(38) Shakked, Z.; Rabinovich, D. The Effect of the Base Sequence on the Fine Structure of the DNA Double Helix. *Prog. Biophys. Mol. Biol.* **1986**, *47* (3), 159–195.

(39) Stott, K.; Tang, G. S. F.; Lee, K.-B.; Thomas, J. O. Structure of a Complex of Tandem HMG Boxes and DNA. *J. Mol. Biol.* **2006**, *360* (1), 90–104.

(40) Duskova, K.; Sierra, S.; Fernandez, M.-J.; Gude, L.; Lorente, A. Synthesis and DNA Intercalation of Ethylenediamine Platinum (II) Complexes Linked to DNA Intercalants. *Bioorg. Med. Chem.* **2012**, *20*, 7112–7118.

(41) Wang, A. H. J.; Ughetto, G.; Quigley, G. J.; Rich, A. Interactions between an Anthracycline Antibiotic and DNA: Molecular Structure of Daunomycin Complexed to d(CGTACG)$_2$ at 1.2 Angstrom Resolution. *Biochemistry* **1987**, *26* (4), 1152–1163.

(42) Searle, M. S.; Hall, J. G.; Denny, W. A.; Wakelin, L. P. G. NMR Studies of the Interaction of the Antibiotic Nogalamycin with the Hexadeoxyribonucleotide Duplex d(5'-GCATGC-3')$_2$. *Biochemistry* **1988**, *27* (12), 4340–4349.

(43) Gottesfeld, J. M.; Melander, C.; Suto, R. K.; Raviol, H.; Luger, K.; Dervan, P. B. Sequence-Specific Recognition of DNA in the Nucleosome by Pyrrole-Imidazole Polyamides. *J. Mol. Biol.* **2001**, *309* (3), 615–629.

(44) Mitra, S. N.; Wahl, M. C.; Sundaralingam, M. Structure of the Side-by-Side Binding of Distamycin to d(GTATATC)$_2$. *Acta Crystallogr. Sect. D* **1999**, *55*, 602–609.

(45) Uytterhoven, K.; Sponer, J.; Van Meervelt, L. Two 1:1 Binding Modes for Ditamycin in the Minor Groove of d(GGCCAATTGG)$_2$. *Eur. J. Biochem.* **2002**, *269*, 2868–2877.

(46) Boger, D. L.; Fink, B. E.; Brunette, S. R.; Tse, W. C.; Hedrick, M. P. A Simple, High-Resolution Method for Establishing DNA Binding Affinity and Sequence Selectivity. *J. Am. Chem. Soc.* **2001**, *123* (25), 5878–5891.

(47) Asagi, M.; Toyama, A.; Takeuchi, H. Binding Affinity and Mode of Distamycin A with A/T Stretches in Double-Stranded DNA: Importance of the Terminal A/T Residues. *Biophys. Chem.* **2010**, *149* (1–2), 34–39.

(48) Asagi, M.; Toyama, A.; Takeuchi, H. Binding Affinity and Mode of Distamycin A with A/T Stretches in Double-Stranded DNA: Importance of the Terminal A/T Residues. *Biophys. Chem.* **2010**, *149*, 34–39.

(49) Hampshire, A. J.; Fox, K. R. The Effects of Local DNA Sequence on the Interaction of Ligands with Their Preferred Binding Sites. *Biochimie* **2008**, *90* (7), 988–998.

(50) Furusawa, H.; Nakayama, H.; Funasaki, M.; Okahata, Y. Kinetic Characterization of Small DNA-Binding Molecules Interacting with a DNA Strand on a Quartz Crystal Microbalance. *Anal. Biochem.* **2016**, *492*, 34–42.

(51) Hoheisel, J. D.; Helwa, R. Analysis of DNA–protein Interactions: From Nitrocellulose Filter Binding Assays to Microarray Studies. *Anal. Bioanal. Chem.* **2010**, No. 398, 2551–2561.

(52) Nguyen, B.; Neidle, S.; Wilson, W. D. A Role for Water Molecules in DNA−Ligand Minor Groove Recognition. *Acc. Chem. Res.* **2009**, *42* (1), 11–21.

(53) Haq, I.; Ladbury, J. E.; Chowdhry, B. Z.; Jenkins, T. C.; Chaires, J. B. Specific Binding of Hoechst 33258 to the d(CGCAAATTTGCG)$_2$ Duplex: Calorimetric and Spectroscopic Studies. *J. Mol. Biol.* **1997**, *271* (2), 244–257.

(54) Rentzeperis, D.; Marky, L. A.; Dwyer, T. J.; Geierstanger, B. H.; Pelton, J. G.; Wemmer, D. E. Interaction of Minor Groove Ligands to an AAATT/AATTT Site: Correlation of Thermodynamic Characterization and Solution Structure. *Biochemistry* **1995**, *34* (9), 2937–2945.

(55) Cai, X.; Jr., P. J. G.; Hoff, D. D. Von. DNA Minor Groove Binders: Back in the Groove. *Cancer Treat. Rev.* **2009**, *35* (5), 437–450.

(56) Barrett, M. P.; Gemmell, C. G.; Suckling, C. J. Minor Groove Binders as Anti-Infective Agents. *Pharmacol. Ther.* **2013**, *139* (1), 12–23.

(57) Yanow, S. K.; Purcell, L. A.; Pradel, G.; Sato, A.; Rodriguez, A.; Lee, M.; Spithill, T. W. Potent Antimalarial and Transmission-Blocking Activities of Centanamycin, a Novel DNA-Binding Agent. *J. Infect. Dis.* **2008**, *197* (4), 527–534.

(58) Van Der Meide, W. F.; Sabajo, L. O. A.; Jensema, A. J.; Peekel, I.; Faber, W. R.; Schallig, H. D. F. H.; Fat, R. F. M. L. A. Evaluation of Treatment with Pentamidine for Cutaneous Leishmaniasis in Suriname. *Int. J. Dermatol.* **2009**, *48* (1), 52–58.

(59) Hunt, N. T. 2D-IR Spectroscopy: Ultrafast Insights into Biomolecule Structure and Function. *Chem. Soc. Rev.* **2009**, *38* (7), 1837–1848.

(60) Dey, B.; Thukral, S.; Krishnan, S.; Chakrobarty, M.; Gupta, S.; Manghani, C.; Rani, V. DNA-Protein Interactions: Methods for Detection and Analysis. *Mol. Cell. Biochem.* **2012**, *365*, 279–299.

(61) Schultz, P. G.; Dervan, P. B. Sequence-Specific Double-Strand Cleavage of DNA by Penta N-Methylpyrrolecarboxamide-EDTA.Fe(II). *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 6834–6837.

(62) Atkins, P.; De Paula, J. *Physical Chemistry, 9th Edition*, 9th Editio.; Oxford University Press: Oxford, 2010.

(63) Stumm, W.; Lee, G. F. Oxygenation of Ferrous Iron. *Ind. Eng. Chem.* **1961**, *53* (2), 143–146.

(64) Spink, N.; Brown, D. G.; Skelly, J. V; Neidle, S. Sequence-Dependent Effects in Drug-DNA Interaction: The Crystal Structure of Hoechst 33258 Bound to the d(CGCAAATTTGCG)$_2$ Duplex. *Nucleic Acids Res.* **1994**, *22* (9), 1607–1612.

(65) Brkljaca, R.; Urban, S.; Hollingsworth, K.; Turnbull, W. B.; Parkinson, J. A. *Modern NMR Techniques for Synthetic Chemistry*, 1st ed.; Fisher, J., Ed.; London, 2014.

(66) Pelton, J. G.; Wemmer, D. E. Structural Characterization of a 2:1 Distamycin A: d(CGCAAATTGGC)$_2$ Complex by Two-Dimensional NMR. *Proc. Natl. Acad. Sci.* **1989**, *86* (15), 5723–5727.

(67) Leroy, J. L.; Kochoyan, M.; Huynh-Dinh, T.; Guéron, M. Characterization of Base-Pair Opening in Deoxynucleotide Duplexes Using Catalyzed Exchange of the Imino Proton. *J. Mol. Biol.* **1988**, *200* (2), 223–238.

(68) Cho, M. Coherent Two-Dimensional Optical Spectroscopy. *Chem. Rev.* **2008**, *108* (4), 1331–1418.

(69) Baiz, C. R.; Reppert, M.; Tokmakoff, A. An Introduction to Protein 2D IR Spectroscopy. In *Ultrafast Infrared Spectroscopy*; 2013; pp 361–404.

(70) Zanni, M. T.; Hochstrasser, R. M. Two-Dimensional Infrared Spectroscopy: A Promising New Method for the Time Resolution of Structures. *Curr. Opin. Struct. Biol.* **2001**, *11* (5), 516–522.

(71) Hithell, G.; Shaw, D. J.; Donaldson, P. M.; Greetham, G. M.; Towrie, M.; Burley, G. A.; Parker, A. W.; 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.

(72) 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* (28), 13991–14000.

(73) Ganim, Z.; Chung, H. S.; Smith, A. W.; Deflores, L. P.; Jones, K. C.; Tokmakoff, A. Amide I Two-Dimensional Infrared Spectroscopy of Proteins. *Acc. Chem. Res.* **2008**, *41* (3), 432–441.

(74) 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* (39), 15650–15660.

(75) Siebert, T.; Guchhait, B.; Liu, Y.; Costard, R.; Elsaesser, T. Anharmonic Backbone Vibrations in Ultrafast Processes at the DNA-Water Interface. *J. Phys. Chem. B* **2015**, *119*, 9670–9677.

(76) Suydam, I. T.; Boxer, S. G. Vibrational Stark Effects Calibrate the Sensitivity of Vibrational Probes for Electric Fields in Proteins. *Biochemistry* **2003**, *42*, 12050–12055.

(77) Nyedegger, M. W.; Dutta, S.; Cheatum, C. M. Two-Dimensional Infrared Study of 3-Azidopyridine as a Potential Spectroscopic Reporter of Protonation State. *J. Chem. Phys.* **2010**, No. 133, 134506.

(78) 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* (6), 2237–2241.

(79) 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* (31), 6123–6138.

(80) Adamczyk, K.; Candelaresi, M.; Robb, K.; Gumiero, A.; Walsh, M. A.; Parker, A. W.; Hoskisson, P. A.; Tucker, N. P.; Hunt, N. T. Measuring Protein Dynamics with Ultrafast Two-Dimensional Infrared Spectroscopy. *Meas. Sci. Technol.* **2012**, *23* (6), 62001.

(81) Simpson, N.; Shaw, D. J.; Frederix, P. W. J. M.; Gillies, A. H.; Adamczyk, K.; Greetham, G. M.; Towrie, M.; Parker, A. W.; Hoskisson, P. A.; Hunt, N. T. Infrared Spectroscopy of Nicotinamide Adenine Dinucleotides in One and Two Dimensions. *J. Phys. Chem. B* **2013**, *117* (51), 16468–16478.

(82) Candelaresi, M.; Gumiero, A.; Adamczyk, K.; Robb, K.; Bellota-Anton, C.; Sangal, V.; Munnoch, J.; Greetham, G. M.; Towrie, M.; Hoskisson, P. A.; et al. A Structural and Dynamic Investigation of the Inhibition of Catalase by Nitric Oxide. *Org. Biomol. Chem.* **2013**, *11* (44), 7778–7788.

(83) Embrey, K. J.; Searle, M. S.; Craik, D. J. Interaction of Hoechst 33258 with the Minor Groove of the A + T-Rich DNA Duplex d(GGTAATTACC)$_2$ Studied in Solution by NMR Spectroscopy. *Eur. J. Biochem.* **1993**, *211* (3), 437–447.

(84) Bostock-Smith, C. E.; Laughton, C. A.; Searle, M. S. Solution Structure and Dynamics of the A-T Tract DNA Decamer Duplex d(GGTAATTACC)$_2$: Implications for Recognition by Minor Groove Binding Drugs. *Biochem. J.* **1999**, *342* (Pt 1), 125–132.

(85) Latt, S. A.; Wohlleb, J. C. Optical Studies of the Interaction of 33258 Hoechst with DNA, Chromatin, and Metaphase Chromosomes. *Chromosoma* **1975**, *52* (4), 297–316.

(86) Purschke, M.; Rubio, N.; Held, K. D.; Redmond, R. W. Phototoxicity of Hoechst 33342 in Time-Lapse Fluorescence Microscopy. *Photochem. Photobiol. Sci.* **2010**, *9*, 1634–1639.

(87) Lalande, M. E.; Ling, V.; Miller, R. G. Hoechst 33342 Dye Uptake as a Probe of Membrane Permeability Changes in Mammalian Cells. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78* (1), 363–367.

(88) Praveen Reddy, B. S.; Murari Sondhi, S.; Lown, J. W. Synthetic DNA Minor Groove Binding Drugs. *Pharmacol. Ther.* **1999**, *84*, 1–111.

(89) Abu-Daya, A.; Brown, P. M.; Fox, K. R. DNA Sequences of of Several AT-Selective Minor Groove Binding Ligands. *Nucleic Acids Res.* **1995**, *23*, 3385–3392.

(90) Breusegem, S. Y.; Clegg, R. M.; Loontiens, F. G. Base-Sequence Specificity of Hoechst 33258 and DAPI Binding to Five (A/T)$_4$ dsDNA Sites with Kinetic Evidence for More than One High-Affinity Hoechst 33258-AATT Complex. *J. Mol. Biol.* **2002**, *315* (5), 1049–1061.

(91) Gavathiotis, E.; Sharman, G. J.; Searle, M. S. 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)$_2$. *Nucleic Acids Res.* **2000**, *28* (3), 728–735.

(92) Fornander, L. H.; Wu, L.; Billeter, M.; Lincoln, P.; Nordén, B. Minor-Groove Binding Drugs: Where Is the Second Hoechst 33258 Molecule? *J. Phys. Chem. B* **2013**, *117* (19), 5820–5830.

(93) Guan, Y.; Shi, R.; Li, X.; Zhao, M.; Li, Y. Multiple Binding Modes for Dicationic Hoechst 33258 to DNA. *J. Phys. Chem. B* **2007**, *111* (25), 7336–7344.

(94) Bunkenborg, J.; Behrens, C.; Jacobsen, J. P. NMR Characterization of the DNA Binding Properties of a Novel Hoechst 33258 Analogue Peptide Building Block. *Bioconjug. Chem.* **2002**, *13* (5), 927–936.

(95) Bostock-Smith, C. .; Searle, M. S. DNA Minor Groove Recognition by Bis-Benzimidazole Analogues of Hoechst 33258: Insights into Structure-DNA Affinity Relationships by Fluorescence Titration Measurements. *Nucleic Acids Res.* **1999**, *27* (7), 1619–1624.

(96) Kiser, J. R.; Monk, R. W.; Smalls, R. L.; Petty, J. T. Hydration Changes in the Association of Hoechst 33258 with DNA. *Biochemistry* **2005**, *44*, 16988–16997.

(97) Han, F.; Taulier, N.; Chalikian, T. V. Association of the Minor Groove Binding Drug Hoechst 33258 with d(CGCGAATTCGCG)$_2$: Volumetric, Calorimetric, and Spectroscopic Characterizations†. *Biochemistry* **2005**, *44* (28), 9785–9794.

(98) Bostock-Smith, C. E.; Searle, M. S. DNA Minor Groove Recognition by Bis-Benzimidazole Analogues of Hoechst 33258: Insights into structure—DNA Affinity Relationships Assessed by Fluorescence Titration Measurements. *Nucleic Acids Res.* **1999**, *27* (7), 1619–1624.

(99) Parkinson, J. A.; Barber, J.; Douglas, K. T.; Rosamond, J.; Sharples, D. Minor-Groove Recognition of the Self-Complementary Duplex d(CGCGAATTCGCG)$_2$ by Hoechst 33258: A High-Field NMR Study. *Biochemistry* **1990**, *29* (44), 10181–10190.

(100) Loewe, H.; Urbanisch, J. Basic Substituted 2,6-Benzimidazole Derivatives, a Novel Series of Substances with Chemortherapeutic Activity. *Arzneinmittel-Forschnung* **1974**, *24*, 1927–1933.

(101) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv. Drug Deliv. Rev.* **2001**, *46* (1–3), 3–26.

(102) Patel, S. R.; Kvols, L. K.; Rubin, J.; O'Connell, M. J.; Edmonson, J. H.; Ames, M. M.; Kovach, J. S. Phase I--II Study of Pibenzimol Hydrochloride (NSC 322921) in Advanced Pancreatic

Carcinoma. *Invest. New Drugs* **1991**, *9* (1), 53–57.

(103) Harapanhalli, R. S.; McLaughlin, L. W.; Howell, R. W.; Rao, D. V; Adelstein, S. J.; Kassis, A. I. (125I/127I) IodoHoechst 33342: Synthesis, DNA Binding, and Biodistribution. *J. Med. Chem.* **1996**, *39* (24), 4804–4809.

(104) Behrens, C.; Harrit, N.; Nielsen, P. E. Synthesis of a Hoechst 32258 Analogue Amino Acid Building Block for Direct Incorporation of a Fluorescent, High-Affinity DNA Binding Motif into Peptides. *Bioconjug. Chem.* **2001**, *12* (6), 1021–1027.

(105) Caddick, S.; Judd, D. B.; Lewis, A. K. d. K.; Reich, M. T.; Williams, M. R. . A Generic Approach for the Catalytic Reduction of Nitriles. *Tetrahedron* **2003**, *59* (29), 5417–5423.

(106) Goddard-Borger, E. D.; Stick, R. V. An Efficient, Inexpensive, and Shelf-Stable Diazotransfer Reagent: Imidazole-1-Sulfonyl Azide Hydrochloride. *Org. Lett.* **2007**, *9* (19), 3797–3800.

(107) Gassman, P. G.; Cryberg, R. L. Chemistry of Nitrenium Ions. IX. Discrete Existence of Singlet and Triplet Nitrenium Ions. *J. Am. Chem. Soc.* **1969**, *91* (18), 5176–5177.

(108) Carpino, L. A.; Han, G. Y. 9-Fluorenylmethoxycarbonyl Function, a New Base-Sensitive Amino-Protecting Group. *J. Am. Chem. Soc.* **1970**, *92* (19), 5748–5749.

(109) Kaushik, M.; Kukreti, S. Temperature Induced Hyperchromism Exhibited by Hoechst 33258: Evidence of Drug Aggregation from UV-Melting Method. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2003**, *59* (13), 3123–3129.

(110) Lee, J.-E.; Ahn, E.; Bak, J. M.; Jung, S.-H.; Park, J. M.; Kim, B.-S.; Lee, H. Polymeric Micelles Based on Photocleavable Linkers Tethered with a Model Drug. *Polymer (Guildf).* **2014**, *55* (6), 1436–1442.

(111) Linnell, R. Notes- Dissociation Constants of 2-Substituted Pyridines. *J. Org. Chem.* **1960**, *25* (2), 290.

(112) Hall Jr., H. K. Correlation of the Base Strengths of Amines. *J. Am. Chem. Soc.* **1957**, *79* (20), 5441–5444.

(113) Yang, C.; Mi, J.; Feng, Y.; Ngo, L.; Gao, T.; Yan, L.; Zheng, Y. G. Labeling Lysine Acetyltransferase Substrates with Engineered Enzymes and Functionalized Cofactor Surrogates. *J. Am. Chem. Soc.* **2013**, *135* (21), 7791–7794.

(114) Tang, P.; Krause, H.; Fürstner, A. Boric Acid Catalysed Amide Formation from Carboxylic Acids and Amines: N-Benzyl-4-Phenylbutyramine. *Org. Synth.* **2005**, *81*, 262–265.

(115) 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.

(116) Nakamoto, K. *Infrared and Raman Spectra of Inorganic and Coordination Compounds: Theory and Applications in Inorganic Chemistry (Volume A)*; 2009.

(117) Hansen, C. The Three Dimensional Solubility Parameter and Solvent Diffusion Coefficient and Their Importance in Surface Coating Formulation., Technical University of Denmark, 1967.

(118) Ramakers, L. A. I. PhD Thesis in Preparation, 2016.

(119) Ranjan, N.; Fulcrand, G.; King, A.; Brown, J.; Jiang, X.; Leng, F.; Arya, D. P. Selective Inhibition of Bacterial Topoisomerase I by Alkynyl-Bisbenzimidazoles. *Medchemcomm* **2014**, *5* (6), 816–825.

(120) Pinto, R. M.; Guerra, M.; Copeland, G.; Olariu, R. I.; Rodrigues, P.; Barros, M. T.; Costa, M. L.; Dias, A. A. The Mechanism of Pyrolysis of Benzyl Azide: Spectroscopic Evidence for Benzenemethanimine Formation. *J. Phys. Chem. A* **2015**, *119* (18), 4118–4126.

(121) L'Abbe, G. Decomposition and Addition Reactions of Organic Azides. *Chem. Rev.* **1968**, *1* (1), 345–363.

(122) Landi, F.; Johansson, C. M.; Campopiano, D. J.; Hulme, A. N. Synthesis and application of a new cleavable linker for "click"-based affinity chromatography. *Org. Biomol. Chem.* **2010**, *8* (1), 56-59.

(123) Srinivasan, R.; Tan, L. P.; Wu, H.; Yang, P.-Y.; Kalesh, K. A.; Yao, S. Q. High-Throughput Discovery of Mycobacterium Tuberculosis Protein Tyrosine Phosphatase B (MptpB) Inhibitors Using Click Chemistry. *Org. Lett.* **2009,** *11,* (22), 5102 - 5105.

(124) Buurma, N. J.; Haq, I. Calorimetric and Spectroscopic Studies of Hoechst 33258: Self-Association and Binding to Non-Cognate DNA. *J. Mol. Biol.* **2008**, *381* (3), 607–621.

(125) Jung, J.; Van Orden, A. A Three-State Mechanism for DNA Hairpin Folding Characterized by Multiparameter Fluorescence Fluctuation Spectroscopy. *J. Am. Chem. Soc.* **2006**, *128* (4), 1240–1249.

(126) Izanloo, C.; Parsafar, G. A.; Abroshan, H.; Akbarzadeh, H. Denaturation of Drew–Dickerson DNA in a High Salt Concentration Medium: Molecular Dynamics Simulations. *J. Comput. Chem.* **2011**, *32* (16), 3354–3361.

(127) Adhikary, A.; Buschmann, V.; Muller, C.; Sauer, M. Ensemble and Single-Molecule Fluorescence Spectroscopic Study of the Binding Modes of the Bis-Benzimidazole Derivative Hoechst 33258 with DNA. *Nucleic Acids Res.* **2003**, *31* (8), 2178–2186.

(128) Vega, M. C.; García Sáez, I.; Aymamí, J.; Eritja, 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)$_2$ Complexed with the Minor-Groove-Binding Drug Hoechst 33258. *Eur. J. Biochem.* **1994**, *222* (3), 721–726.

(129) Chan, S. S.; Breslauer, K. J.; Hogan, M. E.; Kessler, D. J.; Austin, R. H.; Ojemann, J.; Passner, J. M.; Wiles, N. C. Physical Studies of DNA Premelting Equilibria in Duplexes with and without Homo dA-dT Tracts: Correlations with DNA Bending. *Biochemistry* **1990**, *29* (26), 6161–6171.

(130) 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. U.S.A* **2007**, *104* (3), 712–716.

(131) 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* (14), 8113–8118.

(132) 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* (39), 13103–13109.

(133) Banyay, M.; Sarkar, M.; Gräslund, A. A Library of IR Bands of Nucleic Acids in Solution. *Biophys. Chem.* **2003**, *104* (2), 477–488.

(134) Lee, C.; Park, K.-H.; Cho, M. Vibrational Dynamics of DNA. I. Vibrational Basis Modes and Couplings. *J. Chem. Phys.* **2006**, *125* (11).

(135) Lee, C.; Cho, M. Vibrational Dynamics of DNA. II. Deuterium Exchange Effects and Simulated IR Absorption Spectra. *J. Chem. Phys.* **2006**, *125* (11).

(136) 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* (11).

(137) Lee, C.; Cho, M. Vibrational Dynamics of DNA: IV. Vibrational Spectroscopic Characteristics of A-, B-, and Z-Form DNA's. *J. Chem. Phys.* **2007**, *126* (14).

(138) Adnet, F.; Liquier, J.; Taillandier, E.; Singh, M. P.; Rao, K. E.; Lown, J. W. FTIR Study of Specific Binding Interactions Between DNA Minor Groove Binding Ligands and Polynucleotides. *J. Biomol. Struct. Dyn.* **1992**, *10* (3), 565–575.

(139) Fox, K. R. *Unpublished Work*; 2016.

(140) Fuchs, J.; Fiche, J.-B.; Buhot, A.; Calemczuk, R.; Livache, T. Salt Concentration Effects on Equilibrium Melting Curves from DNA Microarrays. *Biophys. J.* **2010**, *99* (6), 1886–1895.

(141) Ozel, A. B.; Srivannavit, O.; Rouillard, J.-M.; Gulari, E. Target Concentration Dependence of DNA Melting Temperature on Oligonucleotide Microarrays. *Biotechnol. Prog.* **2012**, *28* (2), 556–566.

(142) Aboul-ela, F.; Koh, D.; Tinoco, I. Base-Base Mismatches. Thermodynamics of Double Helix Formation for dCAAAXAAAAG + dCTTTYTTTG (X, Y = A,C,G,T). *Nucleic Acids Res.* **1985**, *13* (13), 4811–4824.

(143) Zhang, X.; Chen, H.; Le, S.; Rouzina, I.; Doyle, P. S.; Yan, J. Revealing the Competition between Peeled ssDNA, Melting Bubbles, and S-DNA during DNA Overstretching by Single-Molecule Calorimetry. *Proc. Natl. Acad. Sci.* **2013**, *110* (10), 3865–3870.

(144) D. Higgins, L.; S. Searle, M. Site-Specificity of Bis-Benzimidazole Hoechst 33258 in A-Tract Recognition of the DNA Dodecamer Duplex d(GCAAAATTTTGC)$_2$. *Chem. Commun.* **1999**, No. 18, 1861–1862.

(145) Marky, L. A.; Breslauer, K. J. Calculating Thermodynamic Data For Transitions of Any Molecularity from Equilibrium Melting Cruves. *Biopolymers* **1987**, *26*, 1601–1620.

(146) Pellecchia, M.; Sem, D. S.; Wuthrich, K. NMR in Drug Discovery. *Nat Rev Drug Discov* **2002**, *1* (3), 211–219.

(147) Chen, X.; Ramakrishnan, B.; Sundaralingam, M. Crystal Structures of the Side-by-Side Binding of Distamycin to AT-Containing DNA Octamers d(ICITACIC)$_2$ and d(ICATATIC)$_2$. *J. Mol. Biol.* **1997**, *267* (5), 1157–1170.

(148) Lam, S. L. DSHIFT: A Web Server for Predicting DNA Chemical Shifts. *Nucleic Acids Res.* **2007**, *35* (suppl 2), W713–W717.

(149) Altona, C.; Faber, D. H.; Hoekzema, A. J. A. W. Double-Helical DNA 1H Chemical Shifts: An Accurate and Balanced Predictive Empirical Scheme. *Magn. Reson. Chem.* **2000**, *38* (2), 95–107.

(150) Katahira, M.; Sugeta, H.; Kyogoku, Y. A New Model for the Bending of DNAs Containing the oligo(dA) Tracts Based on NMR Observations. *Nucleic Acids Res.* **1990**, *18* (3), 613–618.

(151) Deng, Z.; Wang, Q.; Liu, Z.; Zhang, M.; Machado, A. C. D.; Chiu, T.-P.; Feng, C.; Zhang, Q.; Yu, L.; Qi, L.; et al. Mechanistic Insights into Metal Ion Activation and Operator Recognition by the Ferric Uptake Regulator. *Nat Commun* **2015**, *6*.

(152) Piotto, M.; Saudek, V.; Sklenar, V. Gradient-Tailored Excitation for Single-Quantum NMR Spectroscopy of Aqueous Solutions. *J. Biomol. NMR* **1992**, *2*, 661–665.

(153) Goddard, T. D.; Kneller, D. G. SPARKY 3. University of California: San Francisco 2004.

(154) Dervan, P. B.; Ross, R. M.; Marques, M. M. Programmable DNA Binding Oligomers for Control of Transcription. *Curr. Opin. Med. Chem.* **2005**, *5*, 373–387.

(155) Berti, N.; Burley, G. A. Nucleic Acid and Nucleotide Mediated Synthesis of Inorganic Nanoparticles. *Nat. Nanotechnol.* **2008**, *3*, 81–87.

(156) Nelson, D. L.; Cox, M. M.; Lehninger, A. L. *Lehninger Principles of Biochemistry*, 6th editio.; Macmillan Learning: New York, 2013.

(157) Chatake, T.; Sunami, T. Direct Interactions between Z-DNA and Alkaline Earth Cations, Discovered in the Presence of High Concentrations of $MgCl_2$ and $CaCl_2$. *J. Inorg. Biochem.* **2013**, *124* (0), 15–25.

(158) Peisach, E.; Pabo, C. O. Constraints for Zinc Finger Linker Design as Inferred from X-Ray Crystal Structure of Tandem Zif268–DNA Complexes. *J. Mol. Biol.* **2003**, *330* (1), 1–7.

(159) Wolfe, S. A.; Nekludova, L.; Pabo, C. O. DNA Recognition by Cys_2His_2 Zinc Finger Proteins. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29* (1), 183–212.

(160) Dervan, P. B. DNA-Binding Pyrrole and Imidazole Polyamide Derivatives. WO 1998049142 A1, 1998.

(161) Dervan, P. B. Design of Sequence-Specific DNA-Binding Molecules. *Science (80-.).* **1986**, *232* (4749), 464–471.

(162) Friebolin, H. *Basic One- and Two-Dimensional NMR Spectroscopy*, 5th ed.; WILEY-V C H VERLAG GMBH, 2010.

(163) Cheatham, T. Force Field of the Sugar Pucker.

(164) Hunt, N. T. Unpublished Work, 2016.

(165) 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* (13), 5926–5930.

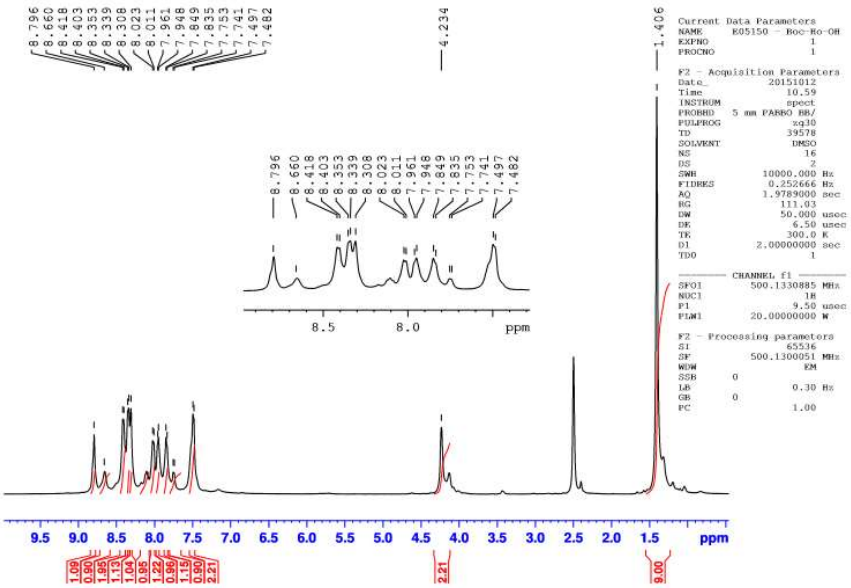
(166) Loontiens, F. G.; Regenfuss, P.; Zechel, A.; Dumortier, L.; Clegg, R. M. Binding Characteristics

of Hoechst 33258 with Calf Thymus DNA, poly[d(A-T)] and d(CCGGAATTCCGG)$_2$: Multiple Stoichiometries and Determination of Tight Binding with a Wide Spectrum of Site Affinities. *Biochemistry* **1990**, *29* (38), 9029–9039.

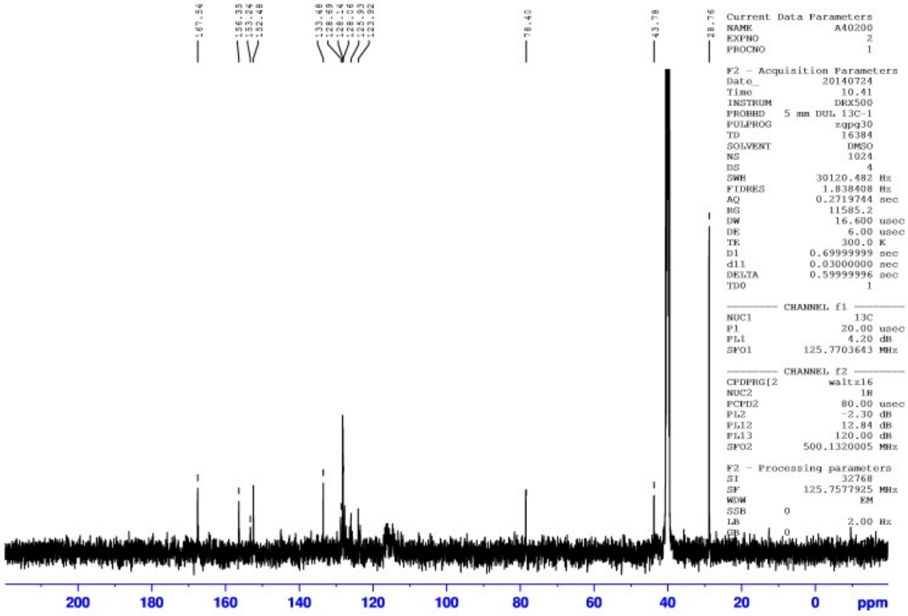
7. Appendices

7.1 – Characterisation data of novel and significant compounds.

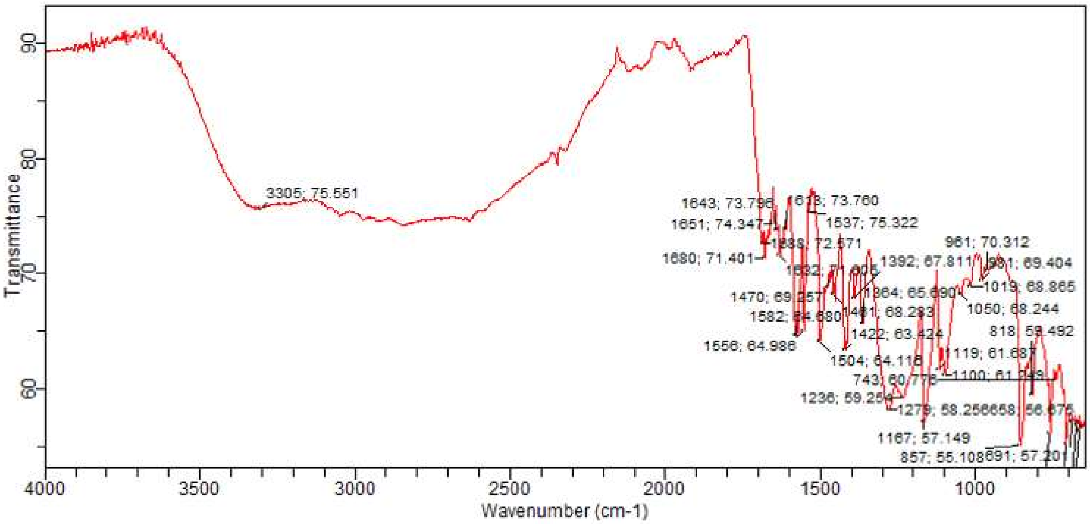
*7.1 Compound **2.12** characterisation.*



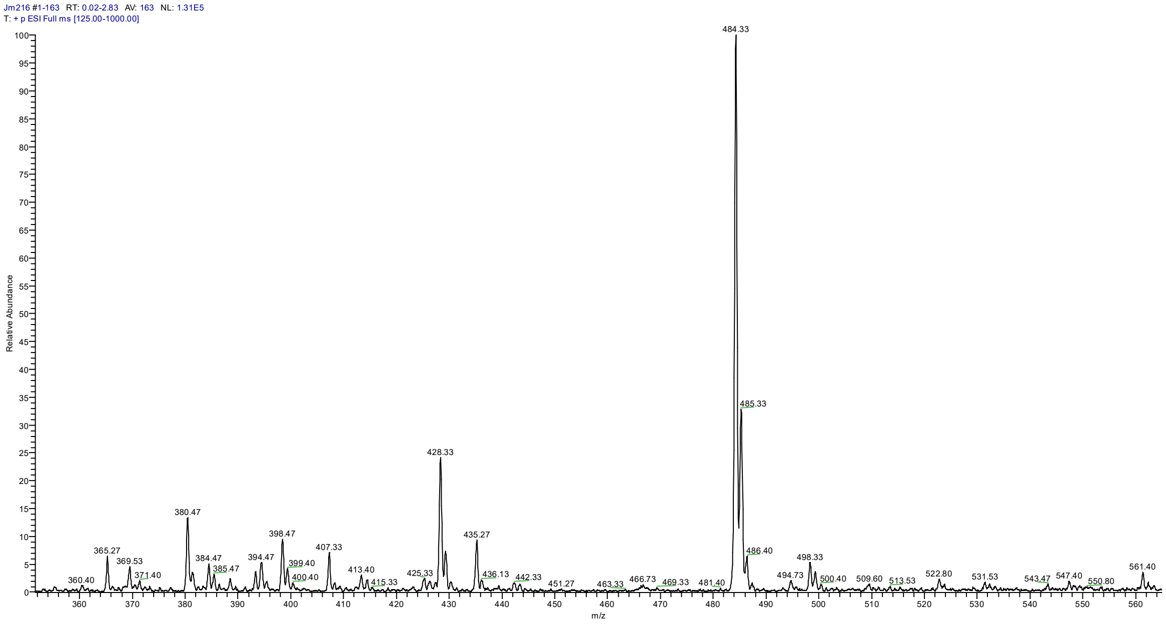
^{1}H NMR spectrum of compound **2.12**



^{13}C NMR of compound **2.12**

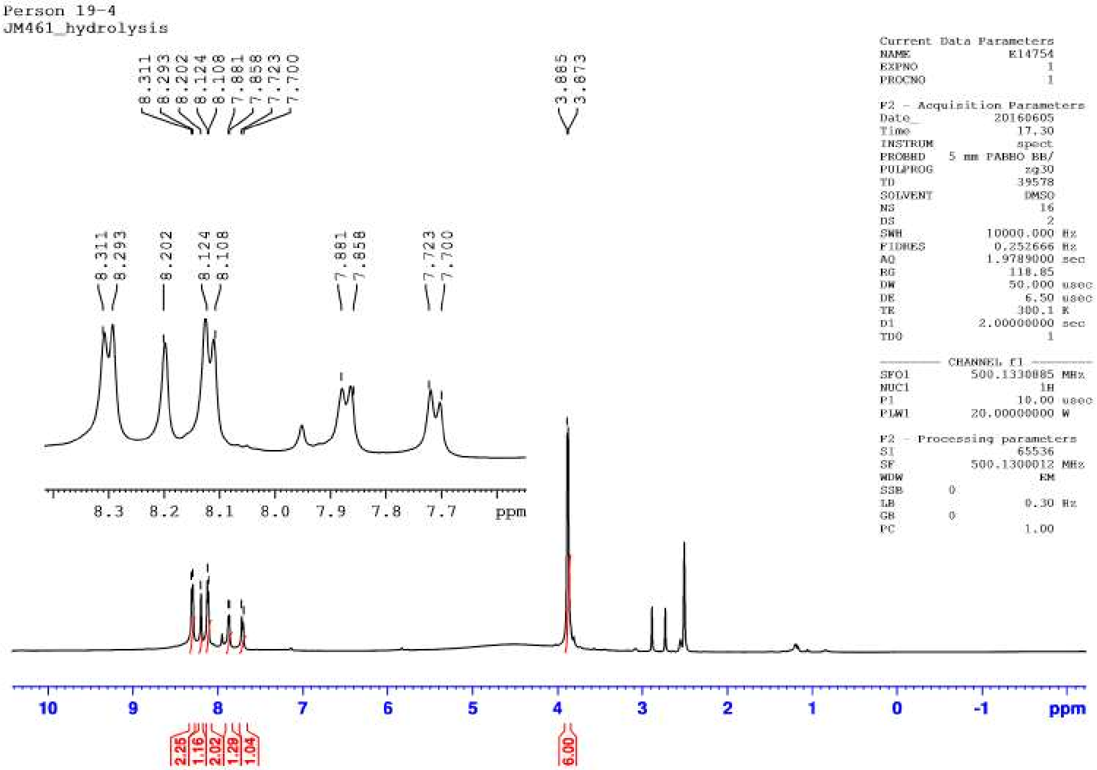


FT-IR spectrum of compound **2.12**

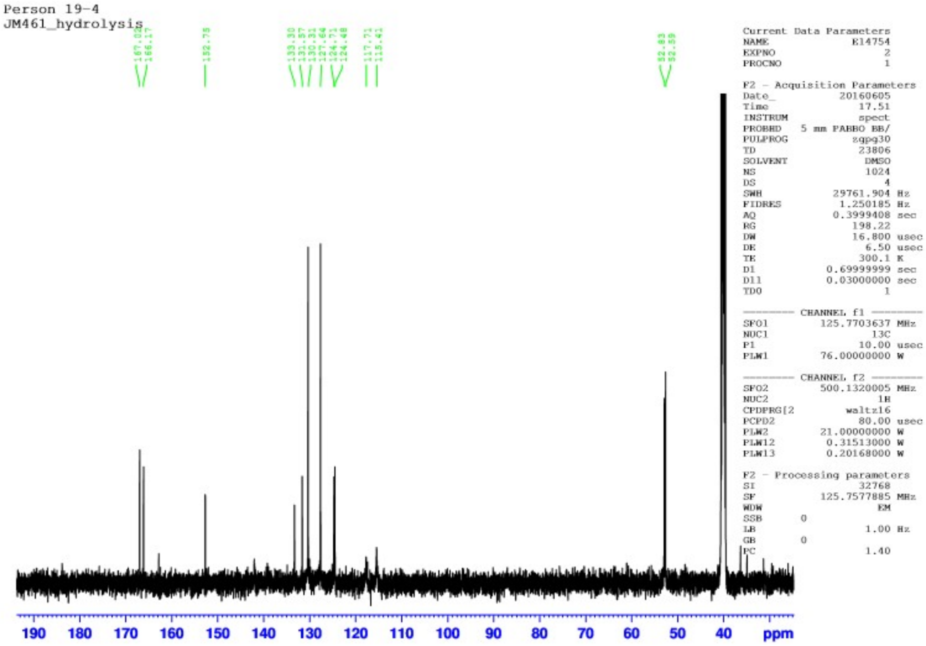


ESI-MS of compound **2.12**. Main peak is at 484 $[M+H]^+$.

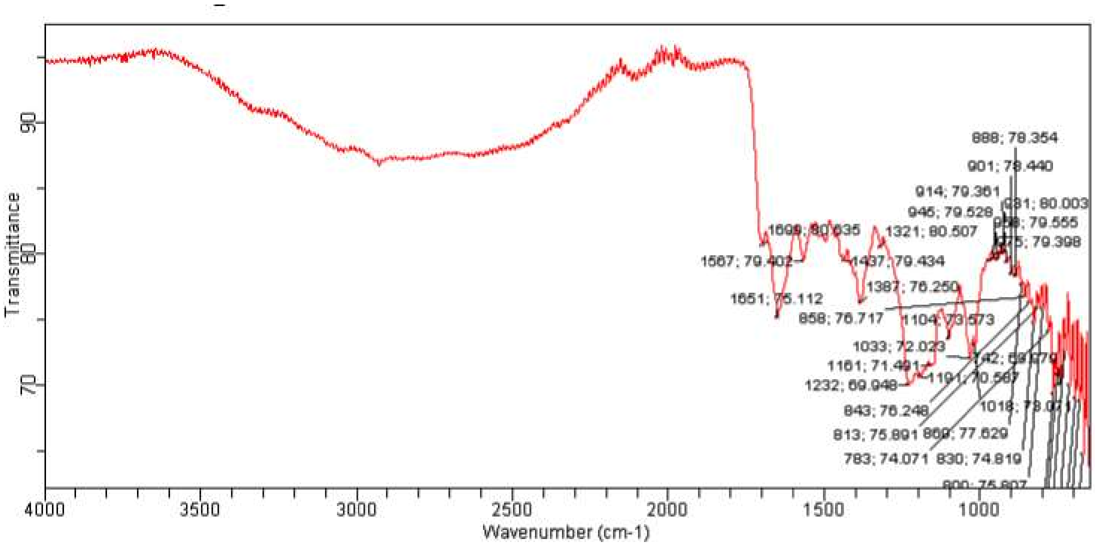
*7.2. Compound **2.22** characterisation*



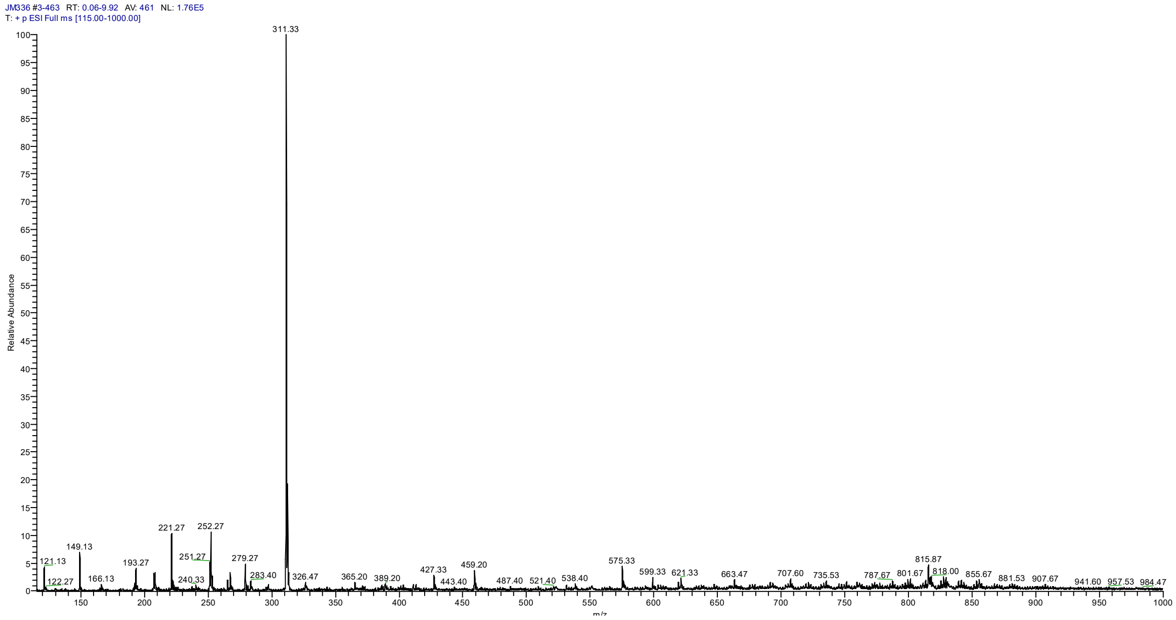
^{1}H NMR of **2.22.** Some residual DMF is present.



^{13}C spectrum of **2.22**

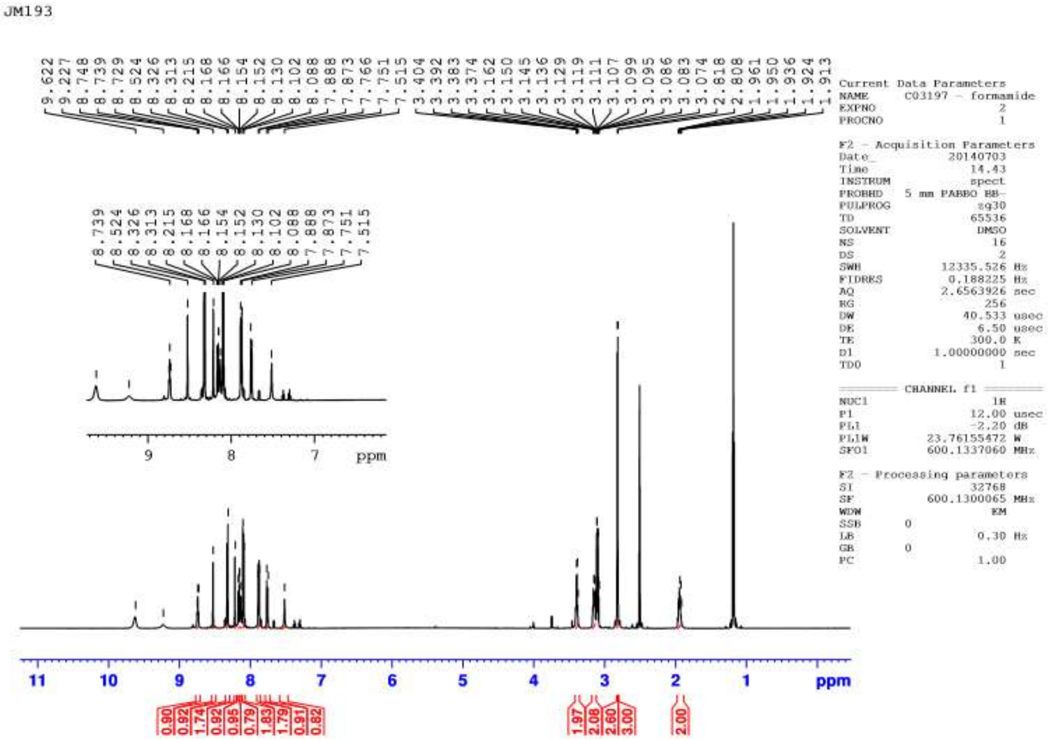


FT-IR spectrum of **2.22.**



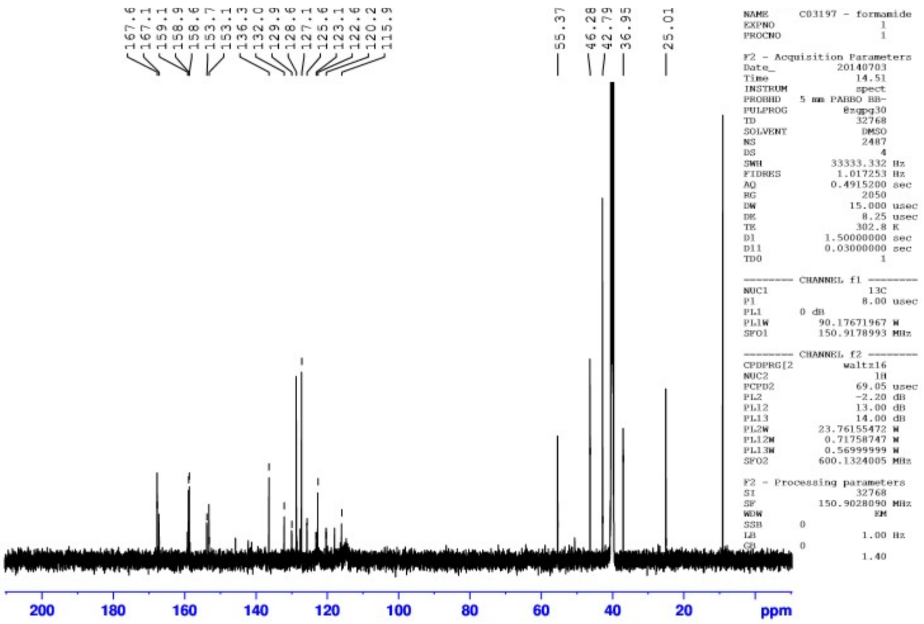
ESI-MS of **2.22**. Main peak at 311 is $[M\text{-}2OAc+3H]^+$

*7.3. Compound **2.23** spectra*



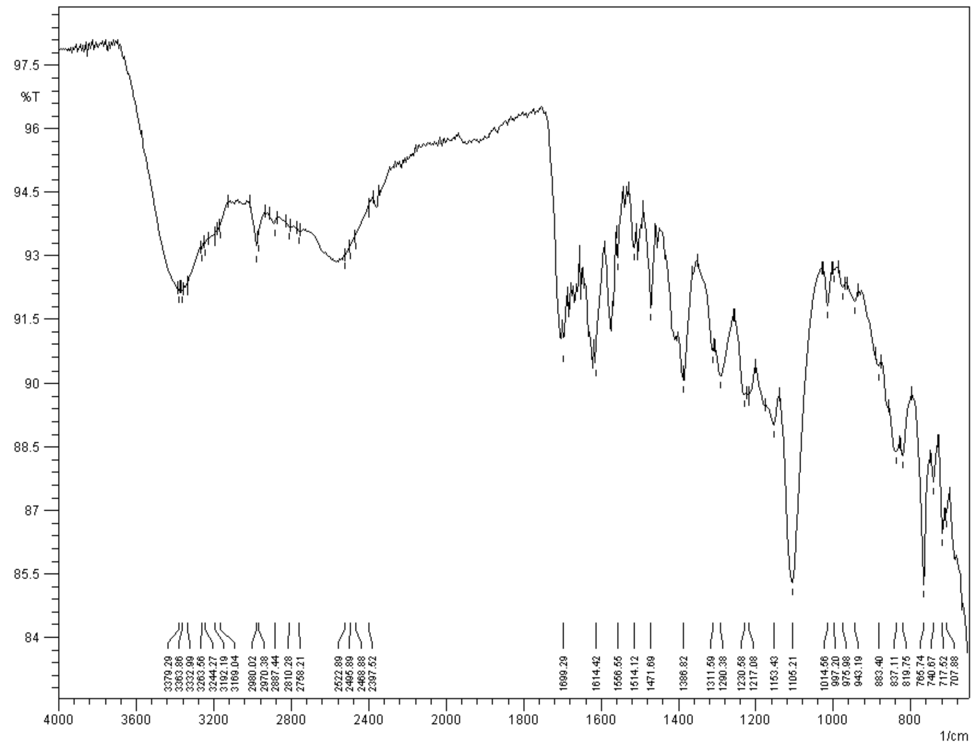
```
Current Data Parameters
NAME       C03197 - formamide
EXPNO                   2
PROCNO                  1

F2 - Acquisition Parameters
Date_            20140703
Time                14.43
INSTRUM             spect
PROBHD   5 mm PABBO BB-
PULPROG              zg30
TD                  65536
SOLVENT              DMSO
NS                     16
DS                      2
SWH             12335.526 Hz
FIDRES           0.188225 Hz
AQ              2.6563926 sec
RG                    256
DW                 40.533 usec
DE                   6.50 usec
TE                  300.0 K
D1             1.00000000 sec
TD0                     1

======== CHANNEL f1 ========
NUC1                   1H
P1                  12.00 usec
PL1                 -2.20 dB
PL1W          23.76155472 W
SFO1          600.1337060 MHz

F2 - Processing parameters
SI                  32768
SF            600.1300065 MHz
WDW                    EM
SSB      0
LB                   0.30 Hz
GB       0
PC                   1.00
```

^{1}H NMR of **2.23.** Some residual buffer is present.

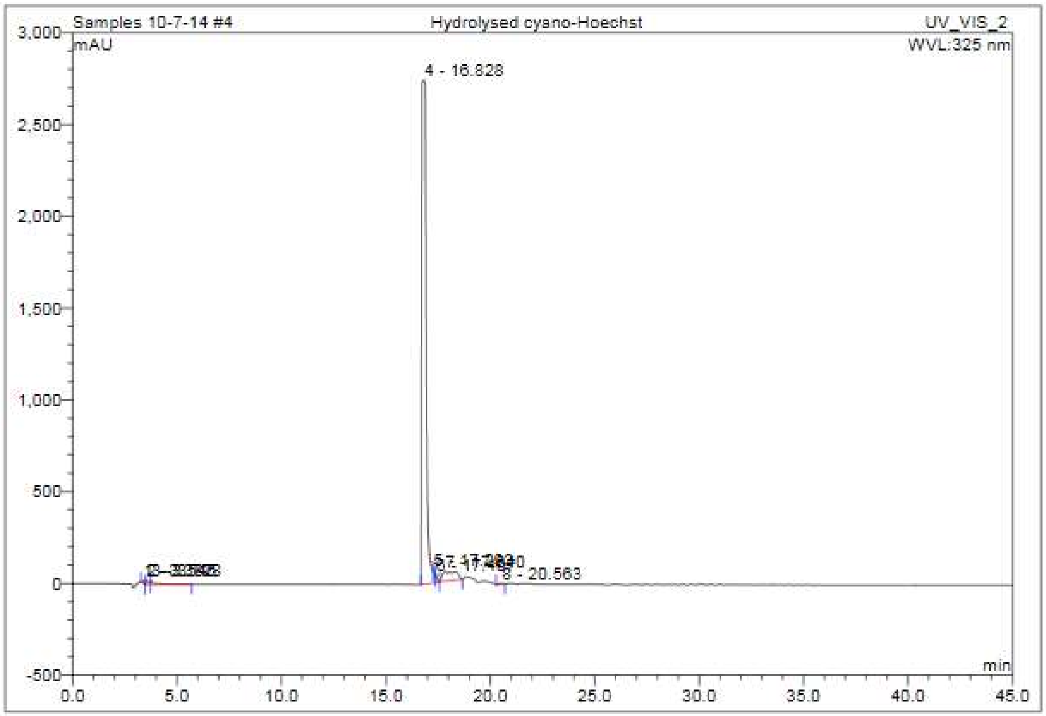


```
NAME       C03197 - formamide
EXPNO                   1
PROCNO                  1

F2 - Acquisition Parameters
Date_            20140703
Time                14.51
INSTRUM             spect
PROBHD   5 mm PABBO BB-
PULPROG           @zgpg30
TD                  32768
SOLVENT              DMSO
NS                   2487
DS                      4
SWH             33333.332 Hz
FIDRES           1.017253 Hz
AQ              0.4915200 sec
RG                   2050
DW                 15.000 usec
DE                   8.25 usec
TE                  302.8 K
D1             1.50000000 sec
D11            0.03000000 sec
TD0                     1

-------- CHANNEL f1 --------
NUC1                  13C
P1                   8.00 usec
PL1      0 dB
PL1W          90.17671967 W
SFO1          150.9178993 MHz

-------- CHANNEL f2 --------
CPDPRG[2          waltz16
NUC2                   1H
PCPD2               69.05 usec
PL2                 -2.20 dB
PL12                13.00 dB
PL13                14.00 dB
PL2W          23.76155472 W
PL12W          0.71758747 W
PL13W          0.56999999 W
SFO2          600.1324005 MHz

F2 - Processing parameters
SI                  32768
SF            150.9028090 MHz
WDW                    EM
SSB      0
LB                   1.00 Hz
GB       0
                     1.40
```

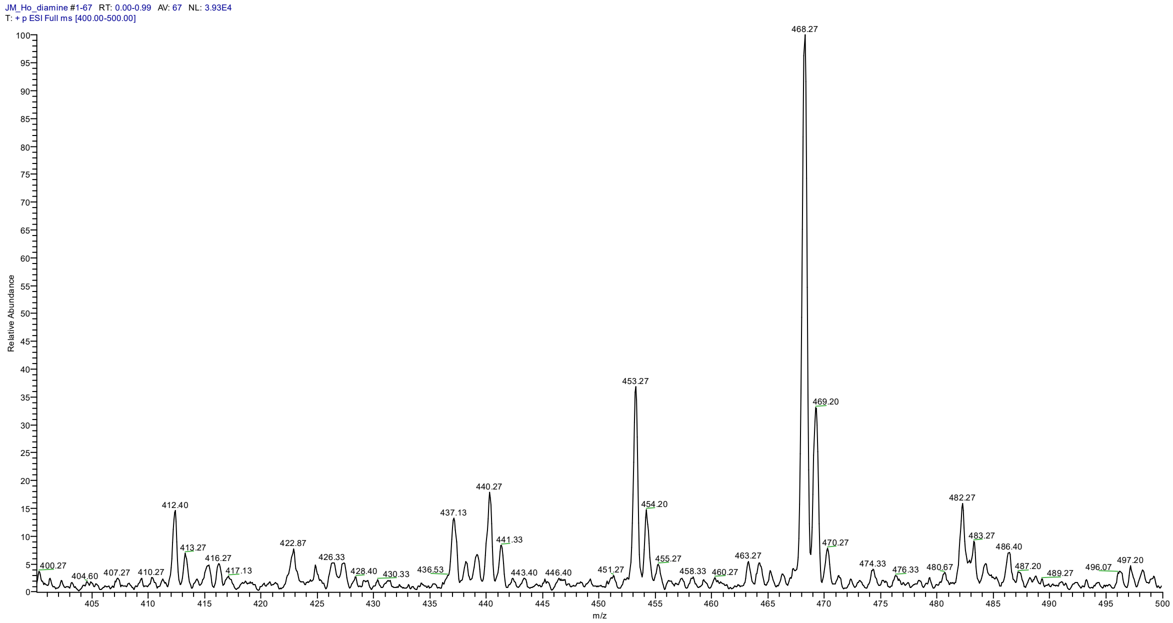
^{13}C NMR of **2.23**. Some buffer is visible also.



FT-IR of compound **2.23**

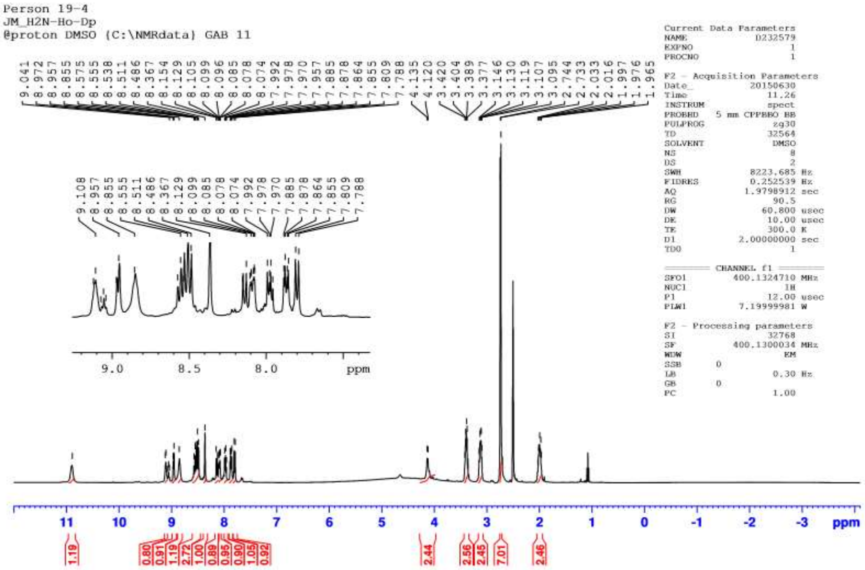


Analytical HPLC trace of compound **2.23**



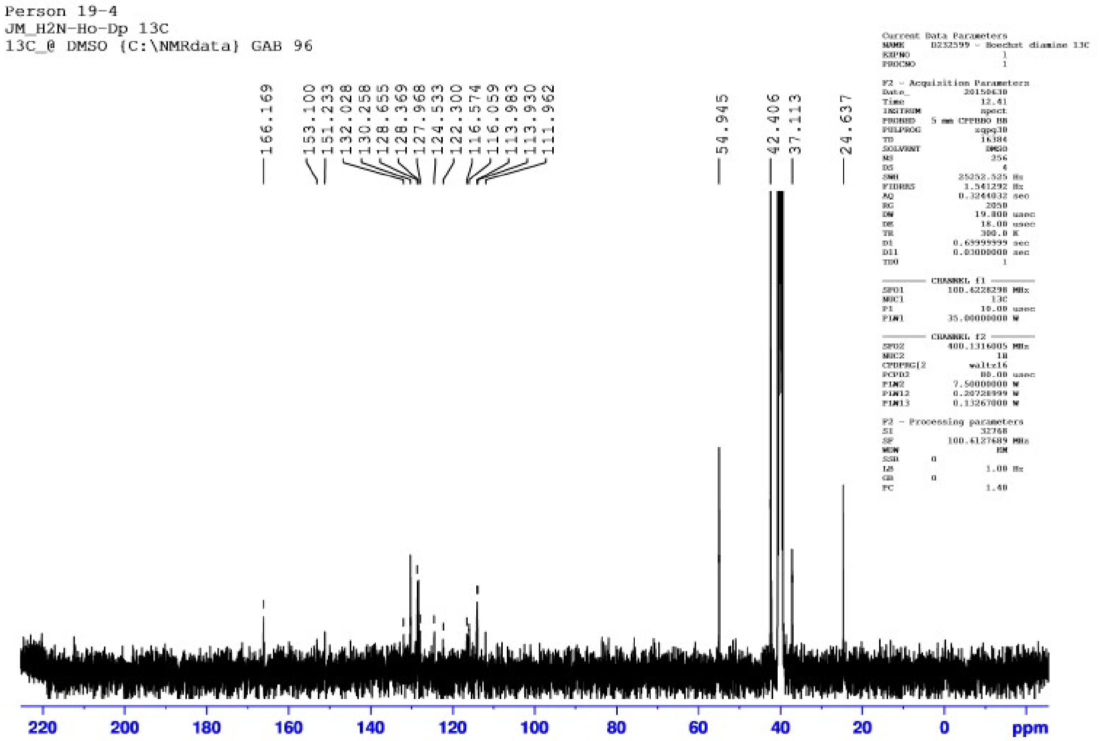
ESI-MS of compound **2.23**. The main peak is at 482 $[M+H]^+$. A second fragment peak is at 453.

7.4. *Compound **2.24** characterisation.*

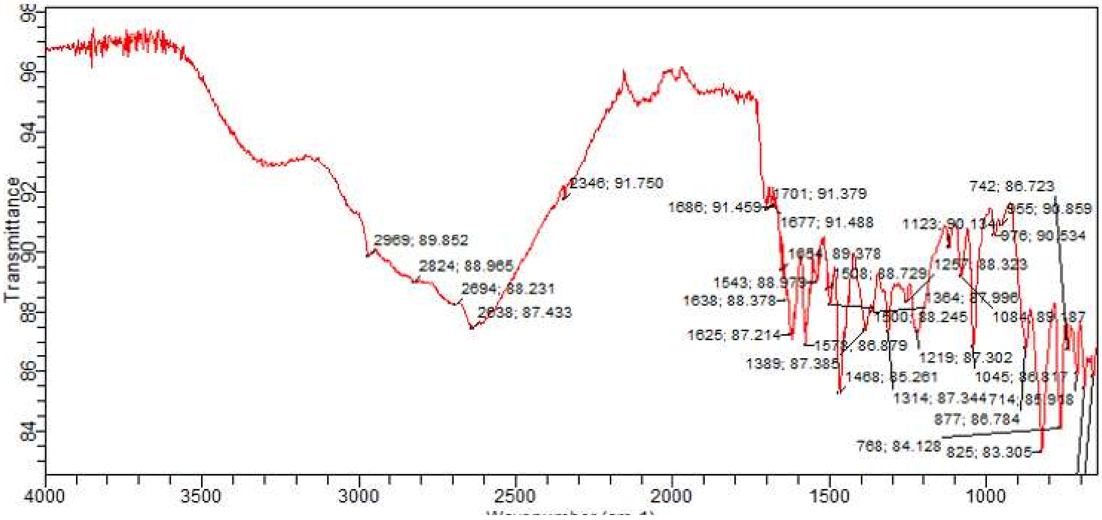


```
Current Data Parameters
NAME            D232579
EXPNO                 1
PROCNO                1

F2 - Acquisition Parameters
Date_          20150630
Time              11.26
INSTRUM           spect
PROBHD   5 mm CPPBBO BB
PULPROG            zg30
TD                32564
SOLVENT            DMSO
NS                    8
DS                    2
SWH            8223.685 Hz
FIDRES         0.252539 Hz
AQ            1.9798912 sec
RG                 90.5
DW               60.800 usec
DE                10.00 usec
TE                300.0 K
D1           2.00000000 sec
TD0                   1

======== CHANNEL f1 ========
SFO1        400.1324710 MHz
NUC1                 1H
P1                12.00 usec
PLW1         7.19999981 W

F2 - Processing parameters
SI                32768
SF          400.1300034 MHz
WDW                  EM
SSB      0
LB                 0.30 Hz
GB       0
PC                 1.00
```

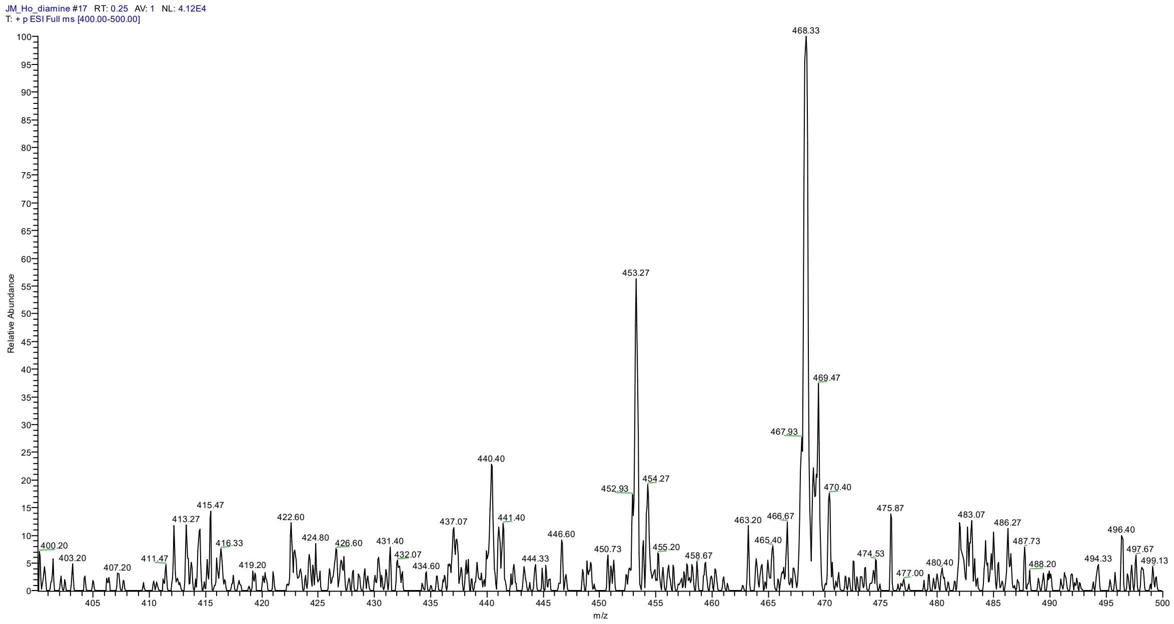
^{1}H NMR of compound **2.24**.



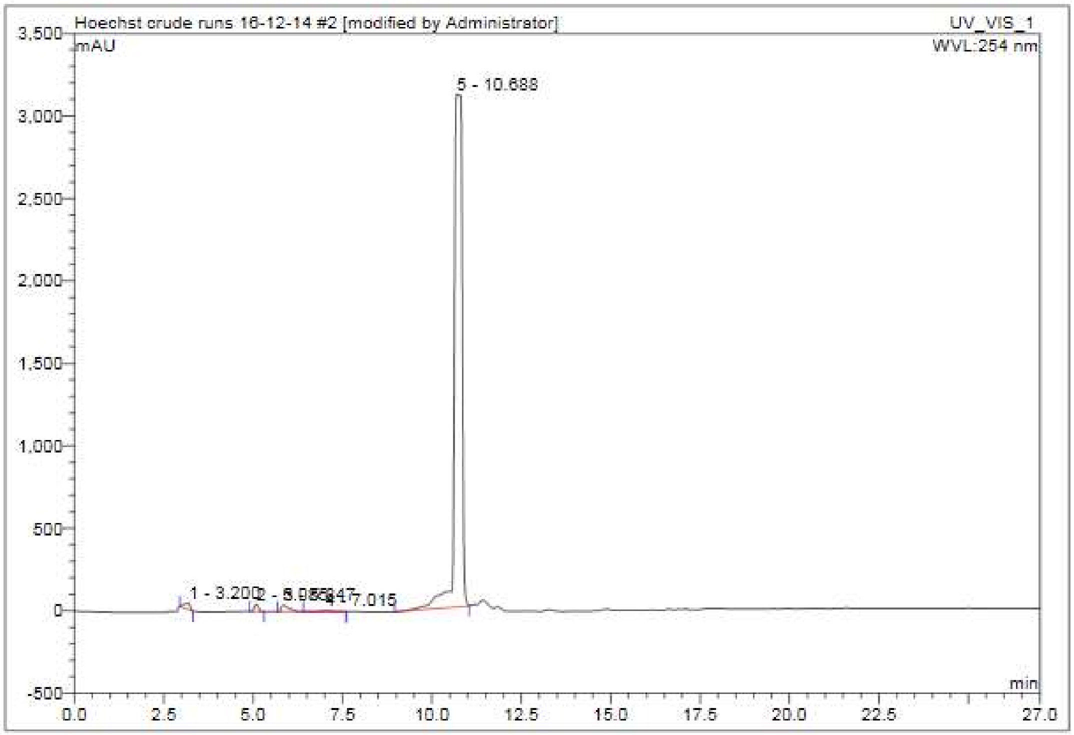
^{13}C NMR of **2.24**. Poor solubility of the substrate in DMSO was the origin of the poor signal to noise ratio.



FT-IR of **2.24**

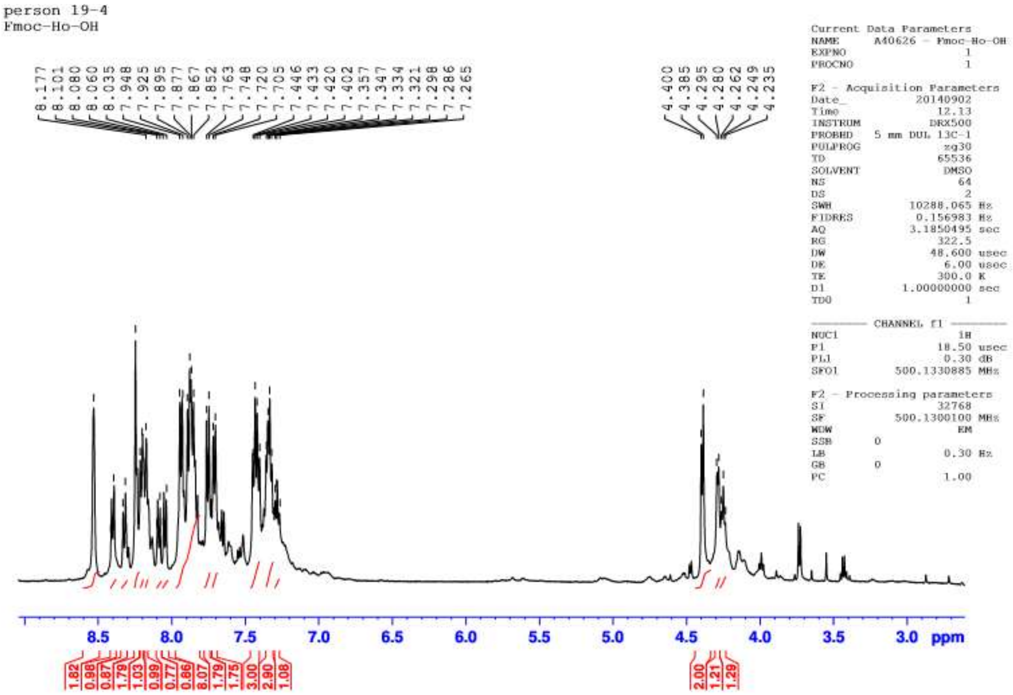


ESI-MS of **2.24.** Main peak at 468 is $[M+H]^+$. The second peak at m/z = 453 is a fragment missing one methyl group.

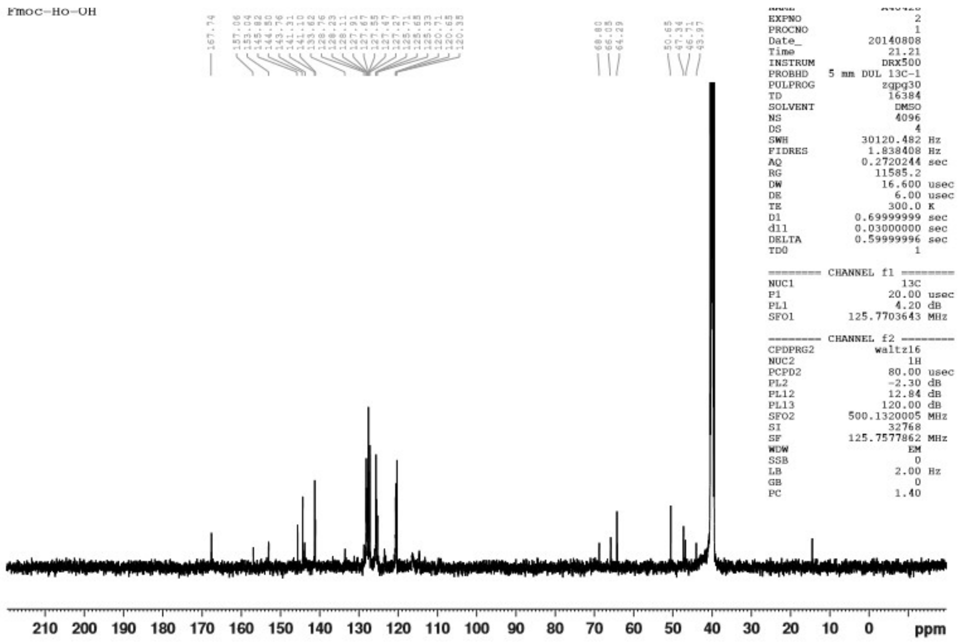


Analytical HPLC trace of purified **2.24**. 18 minute run, eluent 5-90% MeCN in water, 0.1% TFA.

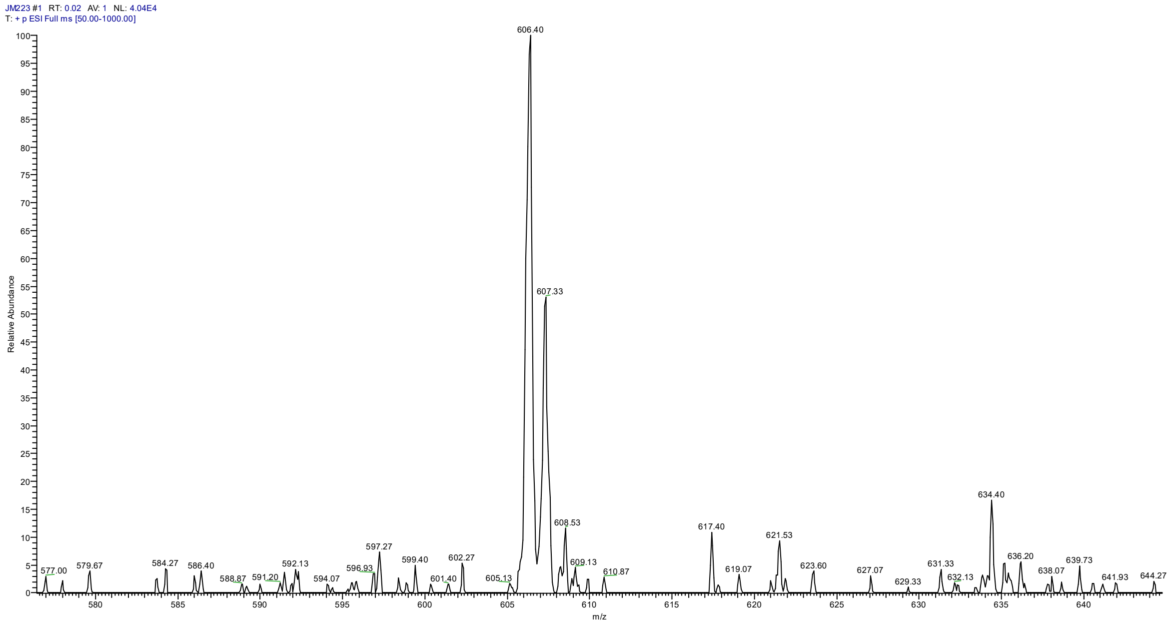
7.5. Compound ***2.25*** characterisation data.



^{1}H NMR of compound **2.25**

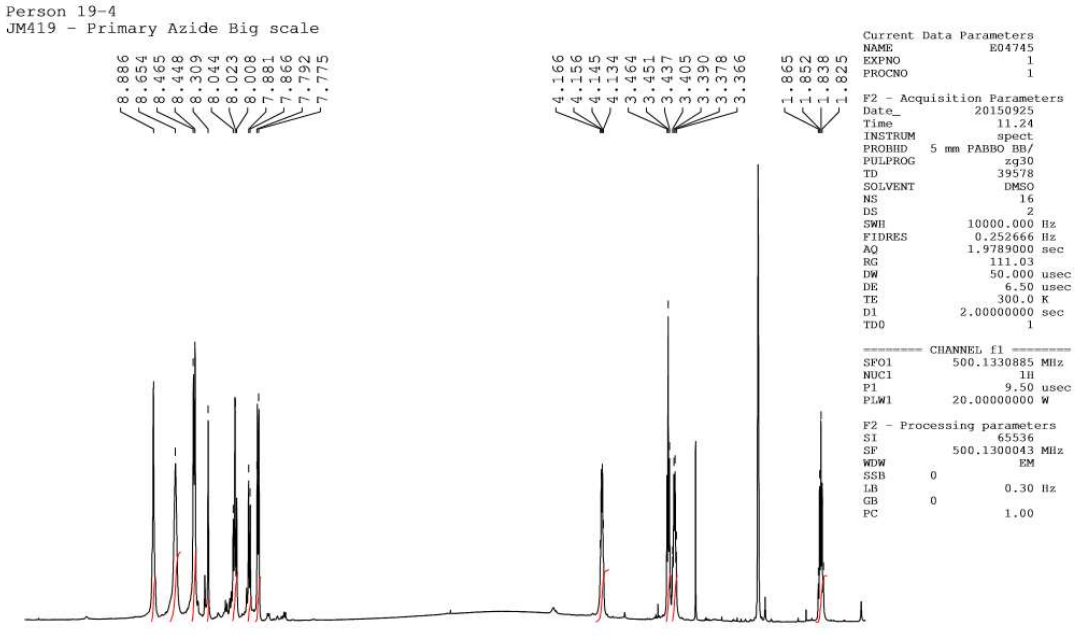


^{13}C NMR of compound **2.25**



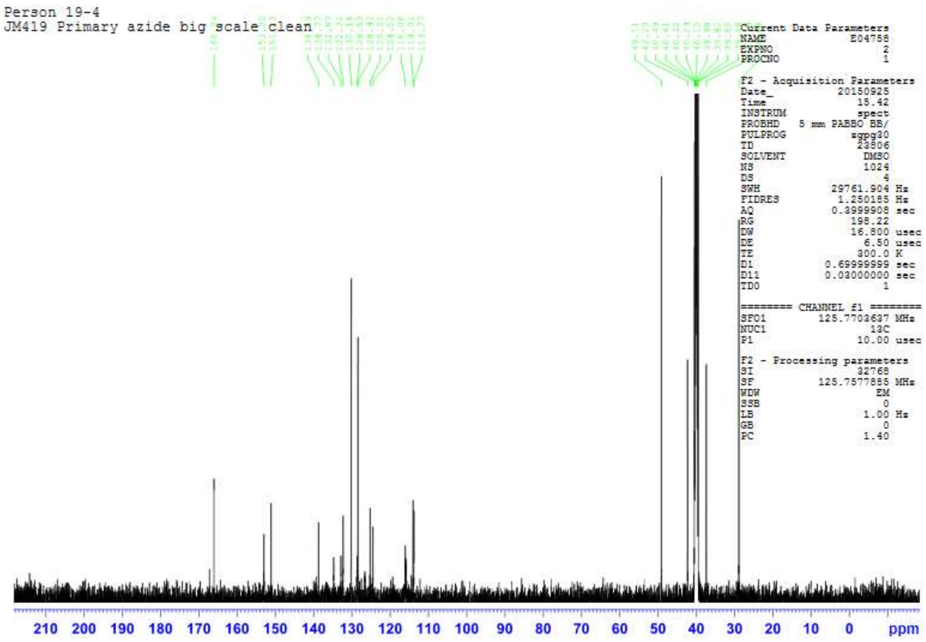
ESI-MS of **2.25.** Main peak at 606 is $[M+H]^+$

7.6 Compund **1.20** characterisation data

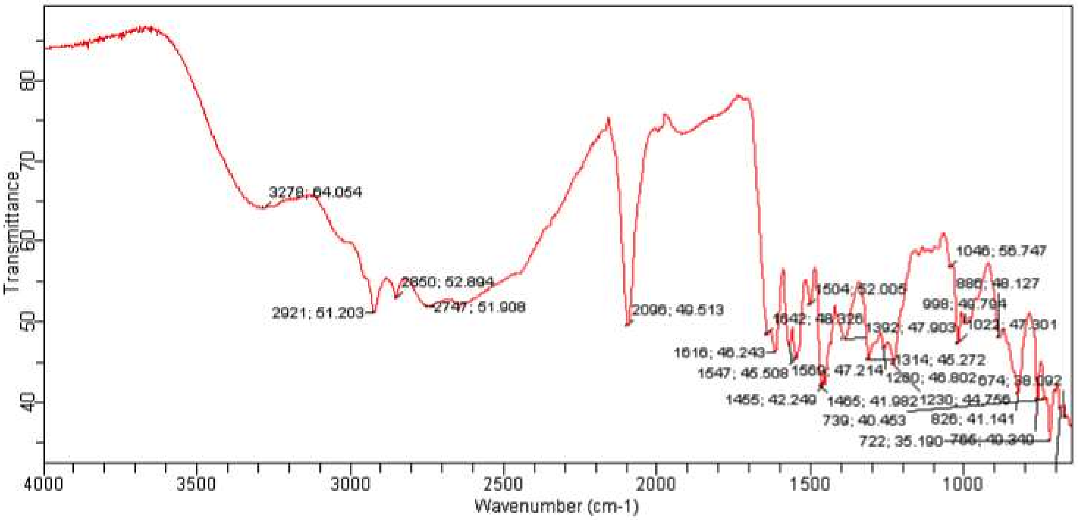


```
Current Data Parameters
NAME            E04745
EXPNO                1
PROCNO               1

F2 - Acquisition Parameters
Date_         20150925
Time             11.24
INSTRUM          spect
PROBHD   5 mm PABBO BB/
PULPROG           zg30
TD               39578
SOLVENT           DMSO
NS                  16
DS                   2
SWH          10000.000 Hz
FIDRES        0.252666 Hz
AQ           1.9789000 sec
RG              111.03
DW              50.000 usec
DE                6.50 usec
TE               300.0 K
D1          2.00000000 sec
TD0                  1

======== CHANNEL f1 ========
SFO1       500.1330885 MHz
NUC1                1H
P1                9.50 usec
PLW1       20.00000000 W

F2 - Processing parameters
SI               65536
SF         500.1300043 MHz
WDW                 EM
SSB      0
LB                0.30 Hz
GB       0
PC                1.00
```

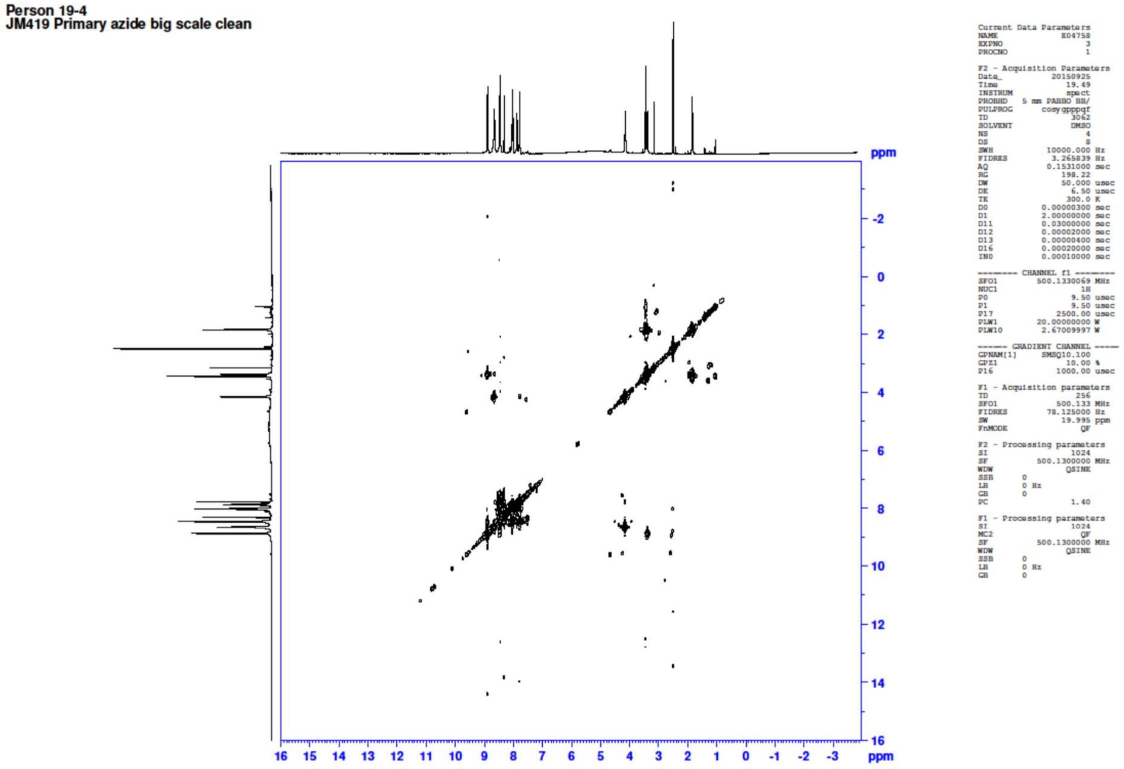
^{1}H spectrum of **1.20**



^{13}C spectrum of **1.20**

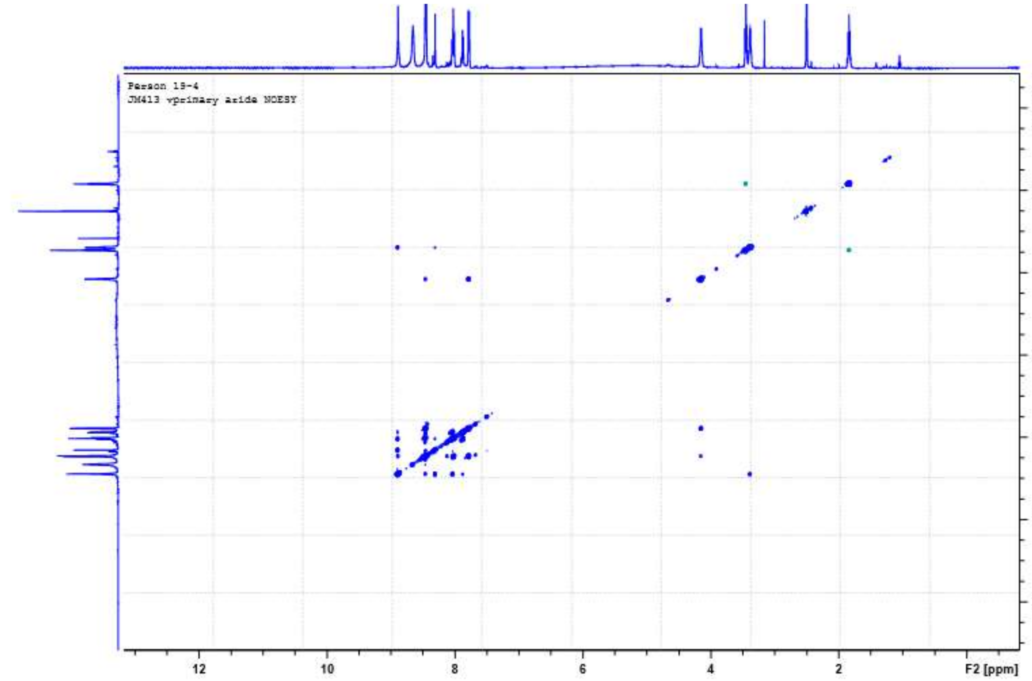


FT-IR spectrum of **1.20**

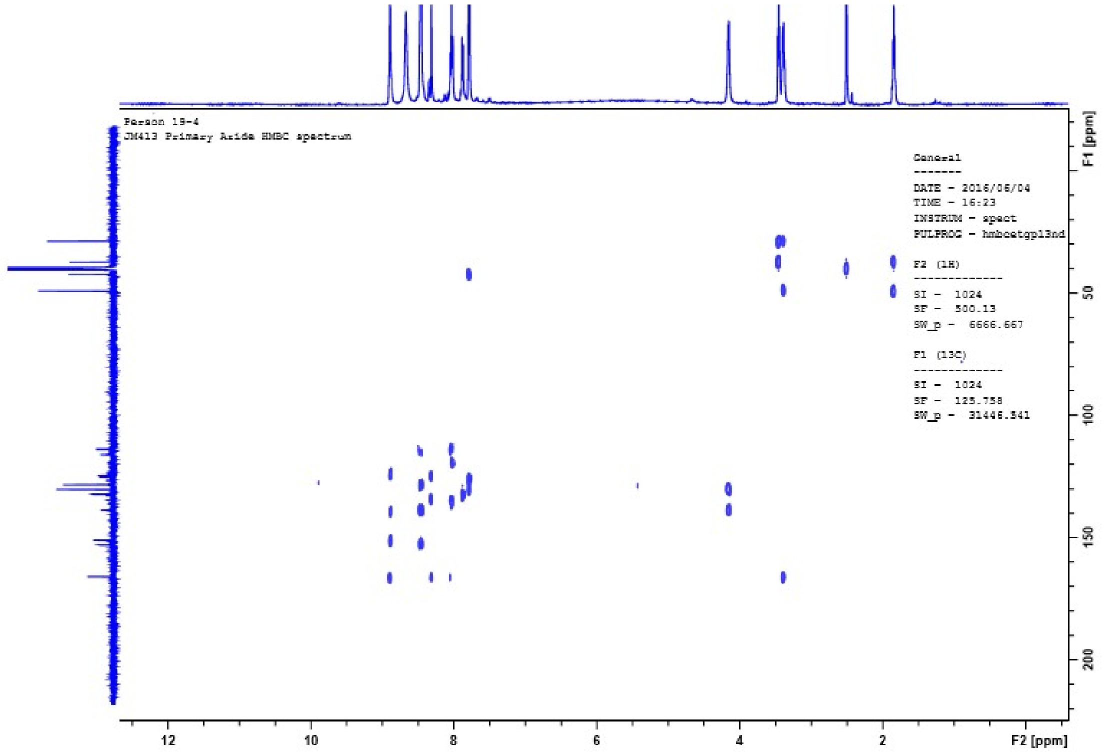


```
Current Data Parameters
NAME               K04758
EXPNO                   3
PROCNO                  1

F2 - Acquisition Parameters
Date_            20150925
Time                19.49
INSTRUM             spect
PROBHD    5 mm PABBO BB/
PULPROG        cosygpppqf
TD                   3062
SOLVENT              DMSO
NS                      4
DS                      8
SWH             10000.000 Hz
FIDRES           3.265839 Hz
AQ              0.1531000 sec
RG                 198.22
DW                 50.000 usec
DE                   6.50 usec
TE                  300.0 K
D0             0.00000300 sec
D1             2.00000000 sec
D11            0.03000000 sec
D12            0.00002000 sec
D13            0.00000400 sec
D16            0.00020000 sec
IN0            0.00010000 sec

======== CHANNEL f1 ========
SFO1          500.1330069 MHz
NUC1                   1H
P0                   9.50 usec
P1                   9.50 usec
P17               2500.00 usec
PLW1          20.00000000 W
PLW10          2.67009997 W

====== GRADIENT CHANNEL =====
GPNAM[1]        SMSQ10.100
GPZ1                10.00 %
P16               1000.00 usec

F1 - Acquisition parameters
TD                    256
SFO1              500.133 MHz
FIDRES          78.125000 Hz
SW                 19.995 ppm
FnMODE                 QF

F2 - Processing parameters
SI                   1024
SF            500.1300000 MHz
WDW                 QSINE
SSB       0
LB        0 Hz
GB        0
PC                   1.40

F1 - Processing parameters
SI                   1024
MC2                    QF
SF            500.1300000 MHz
WDW                 QSINE
SSB       0
LB        0 Hz
GB        0
```

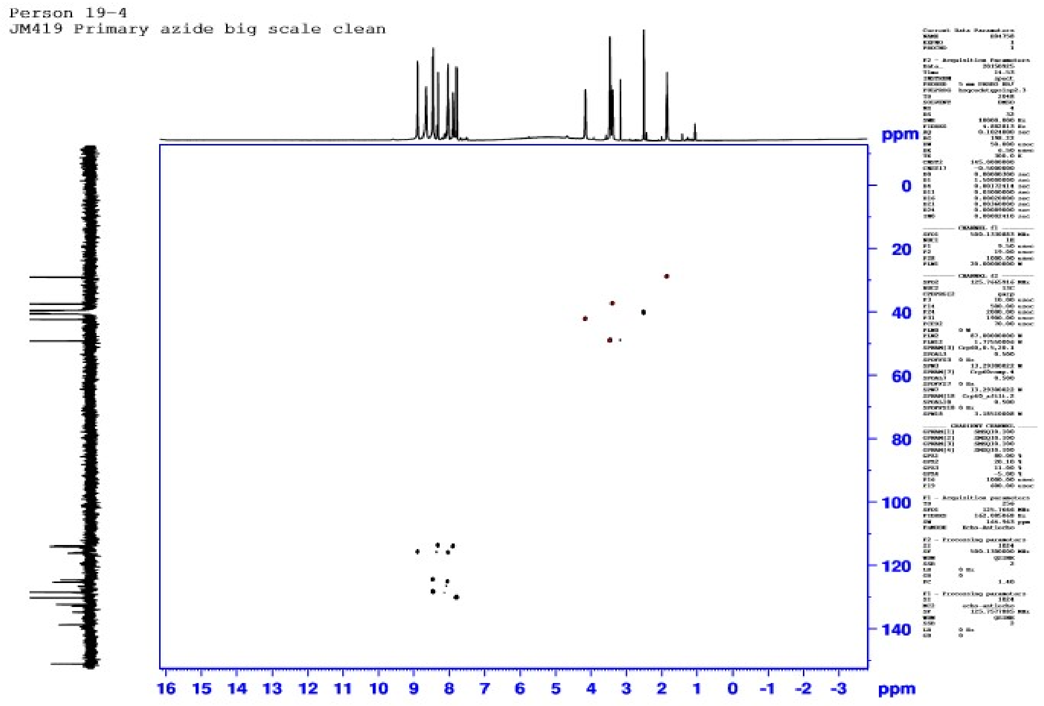
^{1}H-COSY spectrum of **1.20.**



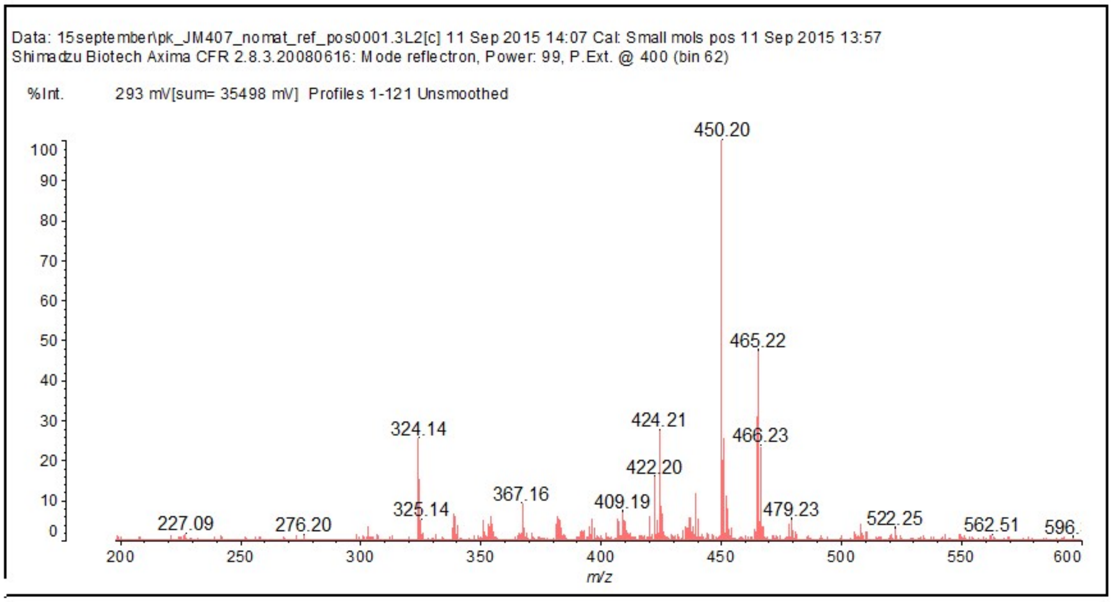
NOESY spectrum for compound **1.20**



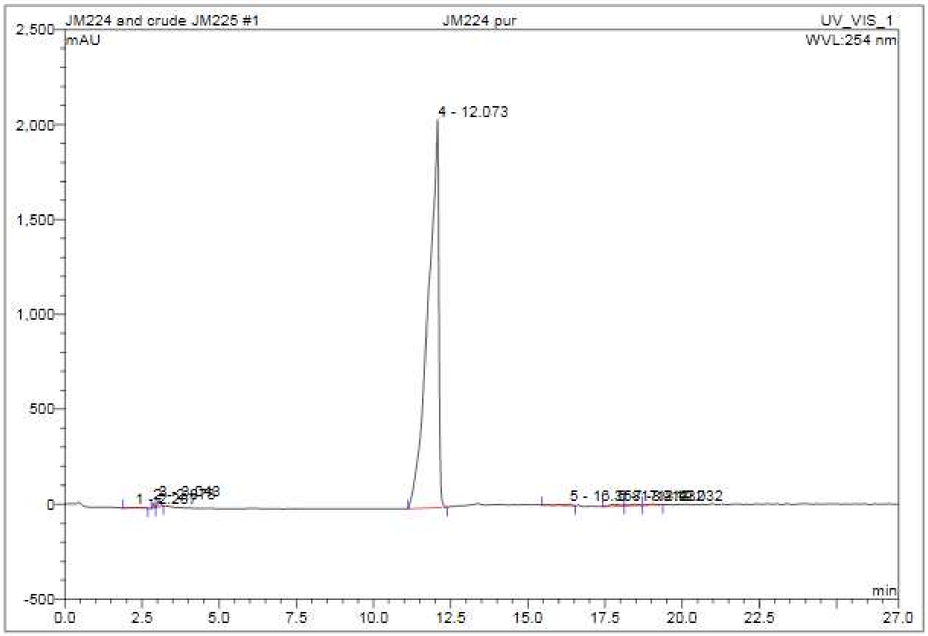
HMBC spectrum of compound **1.20**



HSQC spectrum of **1.20**

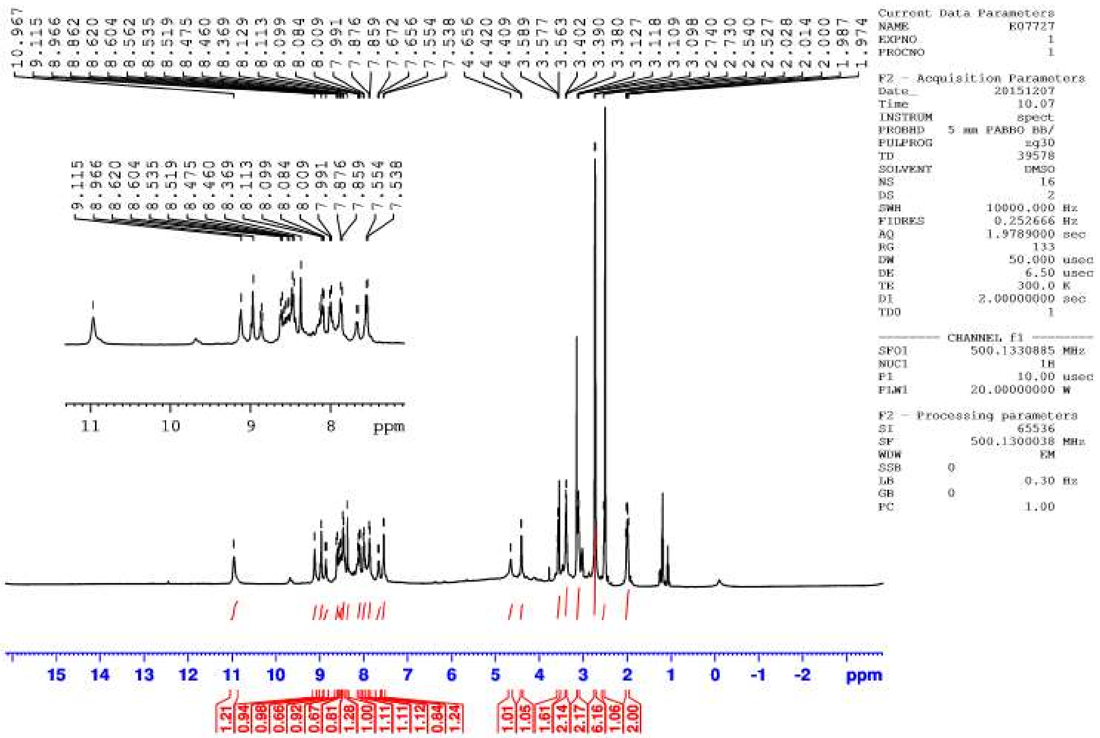


MALDI-MS of **1.20**. Peak at 465 $[M+H]^+$ is observed. The main peak at 450 is a fragment, as are 324 and 424.



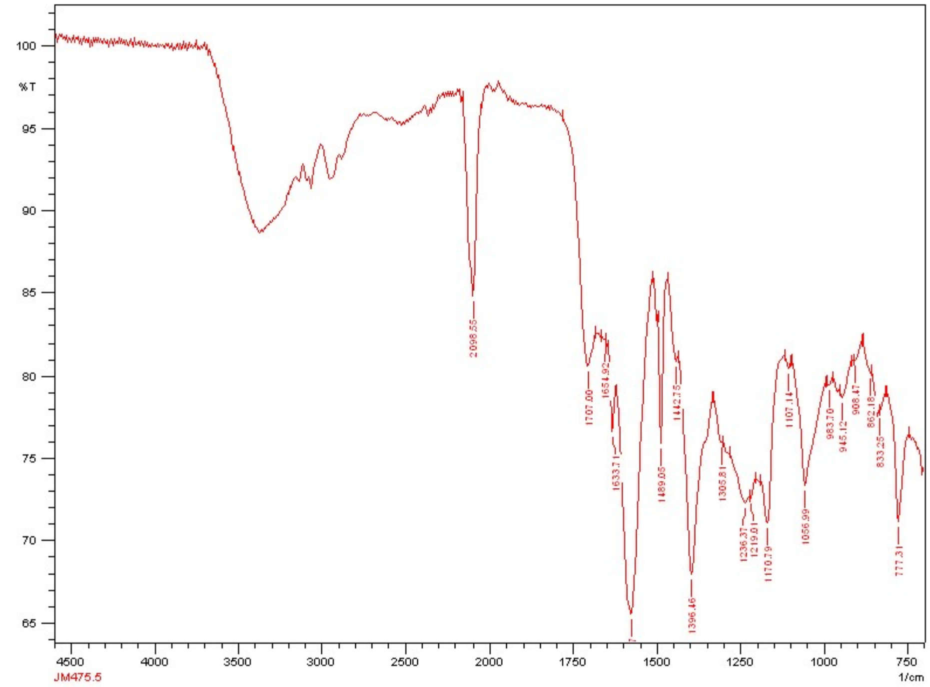
Analytical HPLC trace of compound **1.20.**

*7.7. Compound **1.21** characterisation*

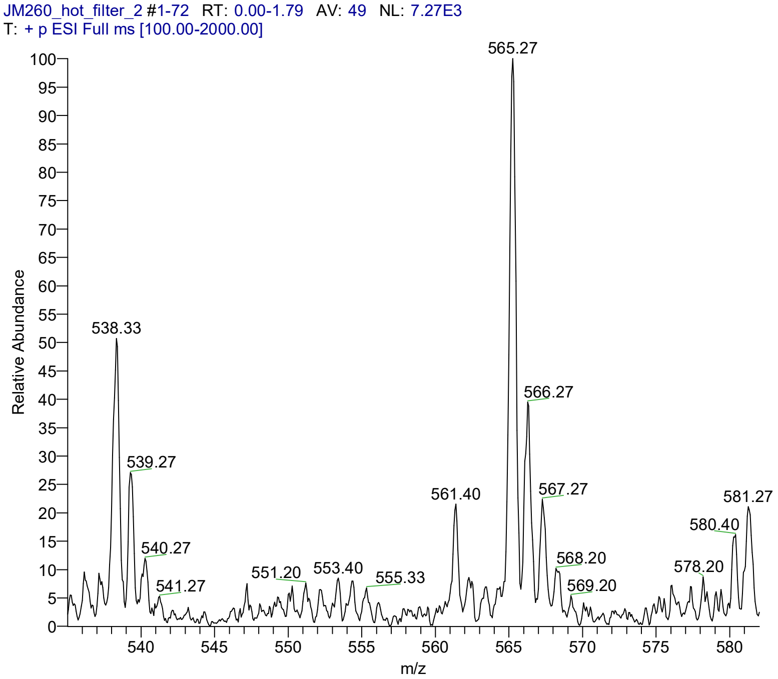


```
Current Data Parameters
NAME                E07727
EXPNO                    1
PROCNO                   1

F2 - Acquisition Parameters
Date_             20151207
Time                 10.07
INSTRUM              spect
PROBHD   5 mm PABBO BB/
PULPROG               zg30
TD                   39578
SOLVENT               DMSO
NS                      16
DS                       2
SWH          10000.000 Hz
FIDRES        0.252666 Hz
AQ           1.9789000 sec
RG                     133
DW              50.000 usec
DE               6.50 usec
TE               300.0 K
D1          2.00000000 sec
TD0                      1

======== CHANNEL f1 ========
SFO1       500.1330885 MHz
NUC1                    1H
P1               10.00 usec
PLW1       20.00000000 W

F2 - Processing parameters
SI                   65536
SF         500.1300038 MHz
WDW                     EM
SSB        0
LB                0.30 Hz
GB         0
PC                    1.00
```

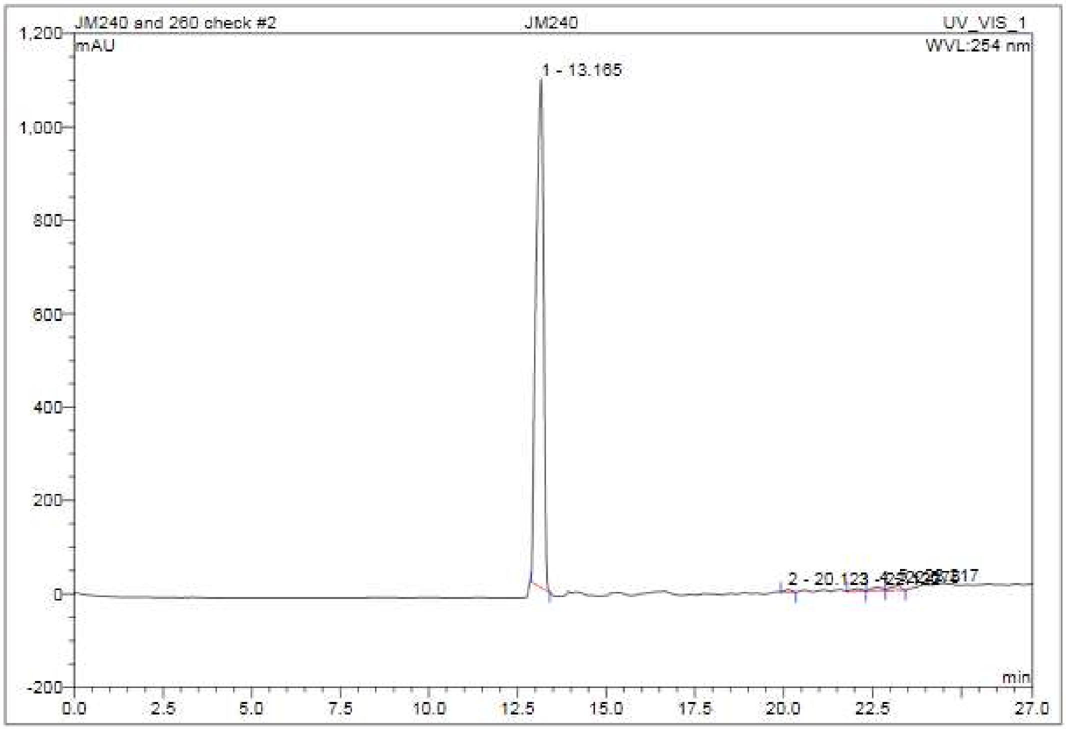
^{1}H NMR of compound **1.21**



FT-IR spectrum of **1.21**

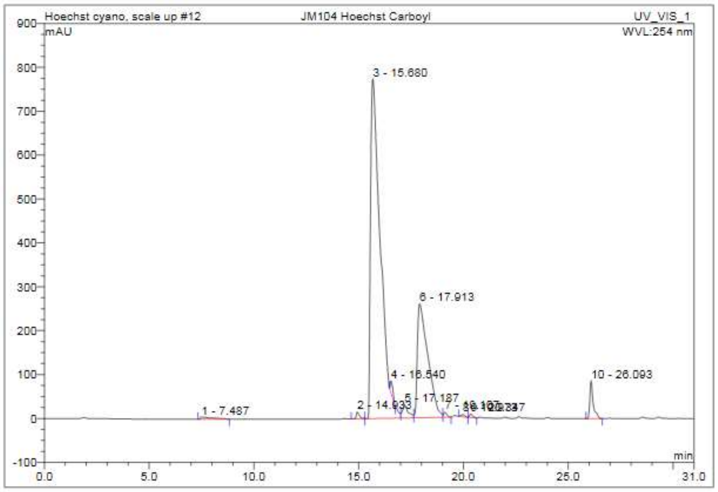


ESI-MS of **2.34**. The peak at 565 is [M+H]. The peak at 538 was shown to be a fragment by MS/MS.



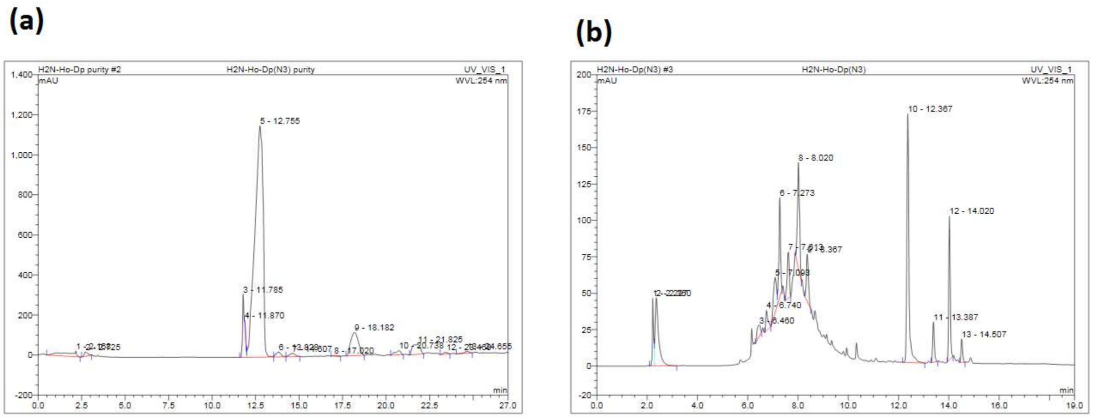
HPLC trace of compound **1.21**

*7.8. Preparative HPLC trace of **2.23** and **1.18***



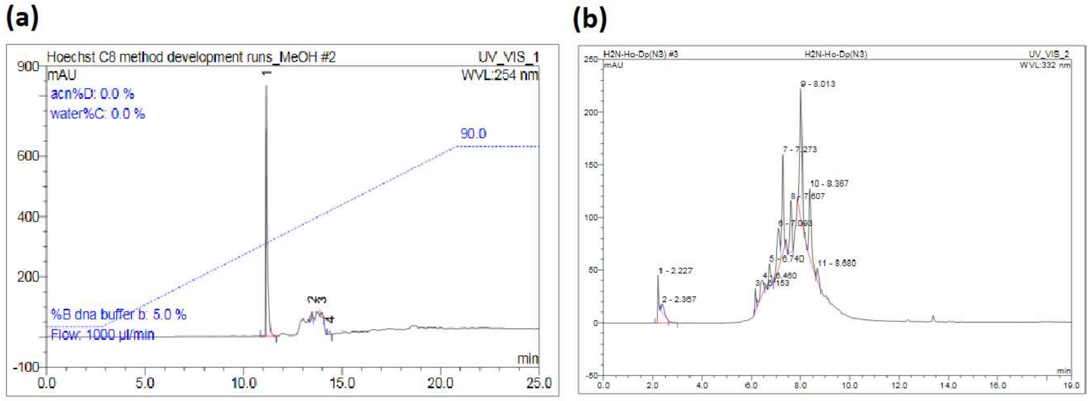
Preparative HPLC trace of crude **1.18.** The peak at 15.68 minutes is **2.23**, the peak at 17.91 minutes was **1.18**. Not enough **1.18** was recovered for further characterisation/ experiments.

*7.9: HPLC traces of crude **1.20** prepared by method A.*



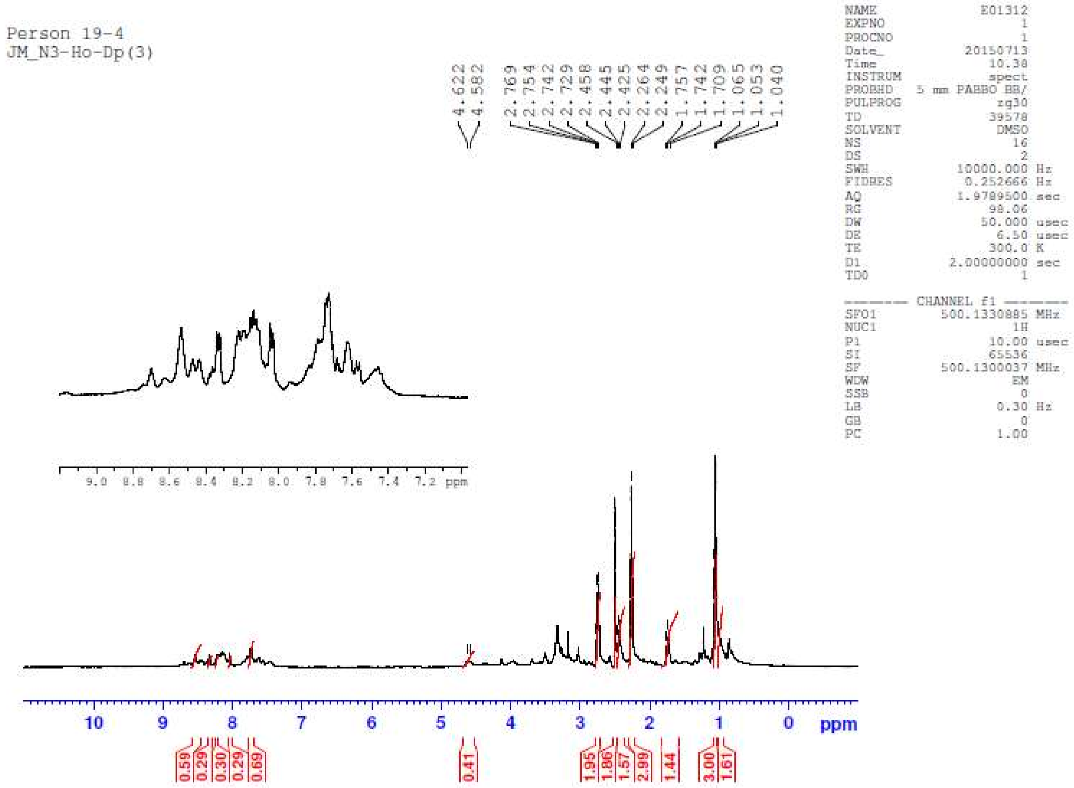
(a) Analytical trace of crude **1.20**. **(b)** Preparative trace of the same sample.

*7.10. HPLC traces of crude **1.21** prepared by Method A.*

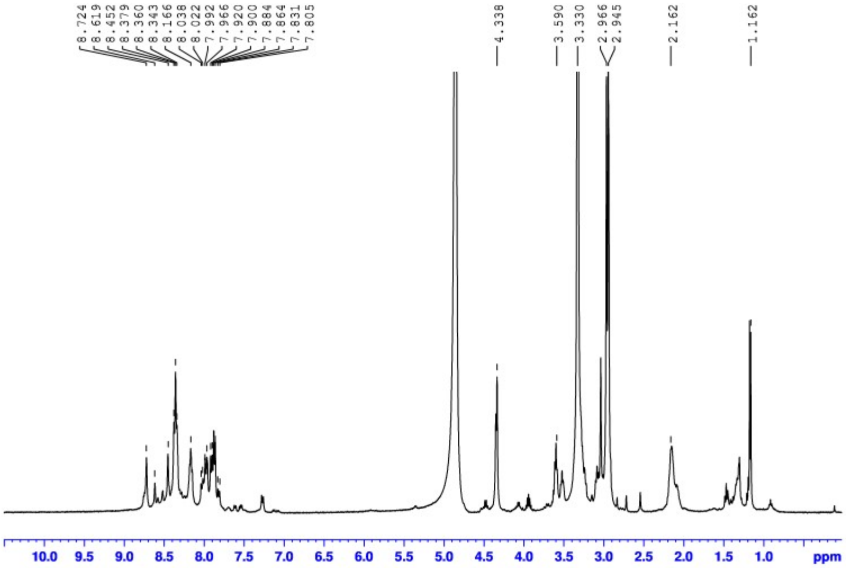


(a) Analytical trace of crude **1.21**. **(b)** Preparative trace of the same sample.

*7.11 Crude ^{1}H – NMR of **1.21** with triethylammonium hydrochloride*

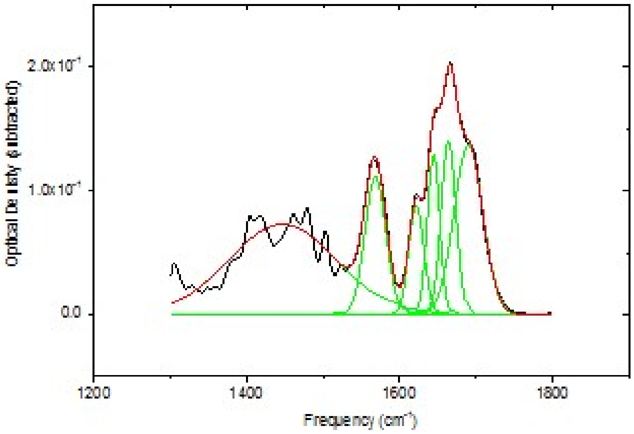


*7.12: Crude **1.21** containing tetramethylurea.*



Crude **1.21** in CD_3OD. The tetramethyluronium species is present at 3.30 ppm.

7.13. Example of fitting of free ODN 3.2 at 25 °C.



The fit of the raw spectrum of ODN 3.2 at 25 °C. The 4 Gaussian peaks between 1600-1700 cm^{-1} and the curved baseline subtraction are visible.

7.14. Derivation of Equation 3.1[145]

$$\text{ssDNA 1} + \text{ssDNA 2} \rightleftharpoons \text{dsDNA}$$

$$\mathrm{K} = \frac{[\text{dsDNA}]}{[\text{ssDNA 1}][\text{ssDNA 2}]}$$

$$\text{and } [\text{ssDNA 1}] = [\text{ssDNA 2}], \text{then}$$

$$\mathrm{K} = \frac{[\text{dsDNA}]}{[\text{ssDNA}]^2}$$

$$\text{At } \mathrm{T}_{\mathrm{m},} [\text{dsDNA}] = [\text{ssDNA}] = 0.5\ \mathrm{C}_\mathrm{T}, \text{then}$$

$$\mathrm{K} = \frac{0.5\ \frac{\mathrm{C_T}}{2}}{(0.5\ \frac{\mathrm{C_T}}{2})^2}\ \text{, which simplifies to:}$$

$$\mathrm{K} = \frac{0.5}{0.25\ \frac{\mathrm{C_T}}{2}}$$

$$-\mathrm{RT_m} \ln\left(\frac{0.5}{0.25\ \frac{\mathrm{C_T}}{2}}\right) = \Delta\mathrm{H}^\circ - \mathrm{T_m}\Delta\mathrm{S}^\circ$$

$$\mathrm{T_m}\Delta\mathrm{S}^\circ - \mathrm{RT_m} \ln\left(\frac{0.5}{0.25\ \frac{\mathrm{C_T}}{2}}\right) = \Delta\mathrm{H}^\circ$$

$$\mathrm{T_m}\ (\Delta\mathrm{S}^\circ - \mathrm{RT_m} \ln\left(\frac{0.5}{0.25\ \frac{\mathrm{C_T}}{2}}\right) = \Delta\mathrm{H}^\circ \qquad 3.1$$

Type equation here.

7.15. Predicted Chemical Shifts of ODN 4.1[148]

Base	H6/H8	H2	H5	TH7(Me)	CNH(B)	CNH(E)	Imino
G1	7.92						12.99
C2	7.41		5.40		8.40	6.51	
G3	7.86						12.74
A4	8.08	7.24					
A5	8.05	7.05					
A6	8.12	7.58					
T7	7.19			1.35			13.67
T8	7.44			1.59			13.92
T9	7.41			1.58			13.76
C10	7.45		5.61		8.49	6.92	
G11	7.87						13.13
C12	7.40		5.39		8.16	6.50	

Base	H1’	H2’	H2’’	H3’	H4’	H5’	H5’’
G1	5.93	2.62	2.73	4.83	4.22	3.73	3.73
C2	5.72	1.99	2.37	4.79	4.14	4.17	4.11
G3	5.53	2.66	2.74	4.99	4.31	4.08	4.01
A4	5.93	2.64	2.88	5.04	4.41	4.12	4.14
A5	5.81	2.60	2.84	5.03	4.41	4.23	4.18
A6	6.12	2.54	2.88	4.99	4.44	4.28	4.22
T7	5.95	2.00	2.53	4.84	4.22	4.32	4.20
T8	6.09	2.15	2.62	4.89	4.21	4.21	4.15
T9	6.07	2.17	2.55	4.89	4.20	4.18	4.09
C10	5.66	1.99	2.35	4.85	4.11	4.09	4.08
G11	5.94	2.63	2.72	4.98	4.35	4.10	4.03
C12	6.17	2.16	2.21	4.48	4.04	4.25	4.10

Predicted chemical shifts of ODN 4.1 using the DSHIFT program.[148]

7.16. Full assignment of free ODN 4.1

Base	H6/H8	H2	H5	TH7(Me)	CNH(B)	CNH(E)	Imino
G1	7.757						12.84
C2	7.110		5.180		8.256	6.318	
G3	7.630						12.57
A4	7.896	7.006					
A5	7.850	6.835					
A6	7.865	7.400					
T7	6.884			1.048			13.50
T8	7.207			1.331			13.75
T9	7.170			1.421			13.53
C10	7.267		5.494		8.265	6.689	
G11	7.695						12.59
C12	7.172		5.140		n.d	n.d	

Base	H1’	H2’	H2’’	H3’	H4’	H5’	H5’’
G1	5.733	2.411	2.554	4.653	4.065	3.522	3.522
C2	5.410	1.688	2.067	4.632	3.921	3.921	3.894
G3	5.109	2.425	2.486	4.786	4.108	3.873	3.769
A4	5.691	2.471	2.690	4.869	4.261	3.998	3.925
A5	5.694	2.430	2.709	4.852	4.278	4.201	4.071
A6	5.949	2.422	2.748	4.821	4.302	4.101	4.045
T7	5.740	1.820	2.322	4.687	4.146	4.118	n.d
T8	5.946	1.999	2.400	4.716	4.042	4.050	3.955
T9	5.918	1.965	2.455	4.724	4.076	4.046	3.952
C10	5.425	1.872	2.192	4.698	4.041	n.d	n.d
G11	5.745	2.433	2.517	4.812	4.183	3.995	3.857
C12	5.950	2.002	2.058	4.288	4.040	4.260	4.260

Full assignment of free ODN 4.1

*7.17. Assignment of the **1.20**: ODN 4.1 complex.*

Base	H6/H8	H2	H5	TH7(Me)	CNH(B)	CNH(E)	Imino
G1	7.716						12.84
C2	7.099		5.140		8.225	6.241	
G3	7.617						12.65
A4	7.907	7.164					
A5	7.938	7.141					
A6	7.962	8.041					
T7	6.714			1.113			13.58
T8	6.813			1.216			13.49
T9	6.996			1.311			13.47
C10	7.234		5.436		8.234	6.542	
G11	7.688						12.65
C12	7.138		5.105		8.229	6.545	

Base	H6/H8	H2	H5	TH7(Me)	CNH(B)	CNH(E)	Imino
G1*	7.716						12.84
C2*	7.099		5.140		8.221	6.241	
G3*	7.617						12.63
A4*	7.950	7.375					
A5*	7.966	7.343					
A6*	7.806	8.160					
T7*	6.579			0.9922			13.93
T8*	6.807			1.180			13.78
T9*	6.805			1.254			13.46
C10*	7.165		5.401		8.342	6.607	
G11*	7.688						12.65
C12*	7.138		5.105		8.236	6.350	

Assignment of the aromatic protons of both strands of the **1.20**: ODN 4.1 complex. The sugar assignments are overleaf.

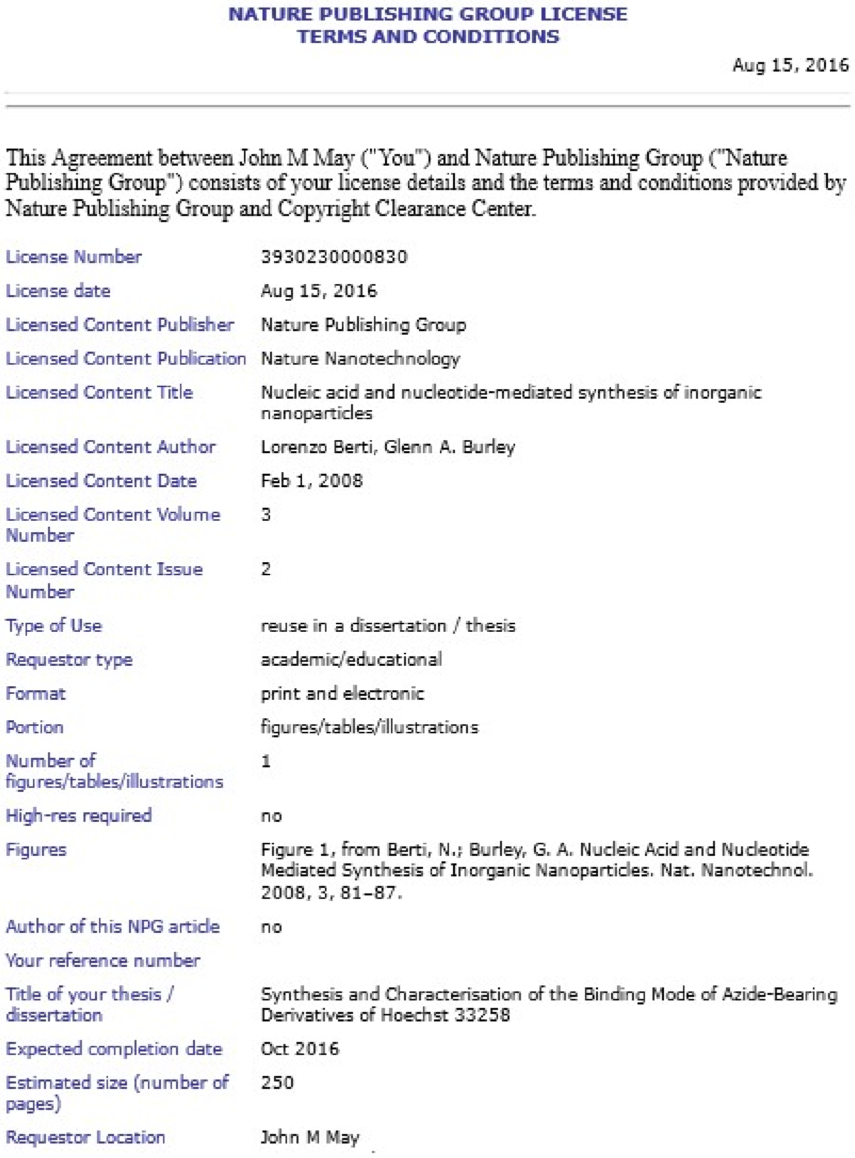
Base	H1’	H2’	H2’’	H3’	H4’	H5’	H5’’
G1	5.685	2.368	2.521	4.626	4.014	3.517	3.517
C2	5.440	1.658	2.052	4.622	3.906	n.d.	n.d.
G3	5.137	2.401	2.442	4.785	4.622	4.094	3.772
A4	5.572	2.460	2.611	4.874	4.216	3.996	3.908
A5	5.800	2.497	2.668	4.885	4.237	n.d.	n.d.
A6	5.899	2.274	2.596	4.805	4.232	4.023	4.023
T7	4.800	1.423	1.875	4.304	1.978	n.d.	n.d.
T8	5.070	1.409	1.864	4.275	1.829	n.d.	n.d.
T9	4.630	1.853	1.981	4.626	2.170	n.d.	n.d.
C10	5.440	1.813	2.168	4.638	3.901	n.d.	n.d.
G11	5.744	2.418	2.510	4.798	4.197	3.935	3.853
C12	5.926	1.963	2.030	4.281	3.859	4.055	4.055

Base	H1’	H2’	H2’’	H3’	H4’	H5’	H5’’
G1*	5.690	2.370	2.512	4.626	4.514	3.517	3.517
C2*	5.436	1.658	2.053	4.622	3.906	n.d.	n.d.
G3*	5.001	2.410	2.442	4.798	4.622	4.094	3.722
A4*	5.576	2.473	2.676	4.886	4.214	n.d.	n.d
A5*	5.656	2.515	2.679	4.874	4.260	4.015	3.918
A6*	5.652	2.272	2.529	4.629	4.201	3.901	3.883
T7*	5.277	1.326	1.806	4.206	1.974	n.d.	n.d.
T8*	4.930	1.308	1.737	4.236	1.898	n.d.	n.d.
T9*	4.779	1.507	1.867	4.398	1.984	n.d.	n.d.
C10*	5.695	1.845	2.075	4.622	3.809	n.d	n.d.
G11*	5.744	2.418	2.510	4.798	4.197	3.935	3.893
C12*	5.926	1.963	2.030	4.281	3.859	4.055	4.055

Assignment of the sugar protons of the **1.20**: ODN 4.1 complex.

7.18 Copyright permissions.

7.18.1 Copyright permission for Figure 1.2



NATURE PUBLISHING GROUP LICENSE
TERMS AND CONDITIONS

Aug 15, 2016

This Agreement between John M May ("You") and Nature Publishing Group ("Nature Publishing Group") consists of your license details and the terms and conditions provided by Nature Publishing Group and Copyright Clearance Center.

License Number	3930230000830
License date	Aug 15, 2016
Licensed Content Publisher	Nature Publishing Group
Licensed Content Publication	Nature Nanotechnology
Licensed Content Title	Nucleic acid and nucleotide-mediated synthesis of inorganic nanoparticles
Licensed Content Author	Lorenzo Berti, Glenn A. Burley
Licensed Content Date	Feb 1, 2008
Licensed Content Volume Number	3
Licensed Content Issue Number	2
Type of Use	reuse in a dissertation / thesis
Requestor type	academic/educational
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
High-res required	no
Figures	Figure 1, from Berti, N.; Burley, G. A. Nucleic Acid and Nucleotide Mediated Synthesis of Inorganic Nanoparticles. Nat. Nanotechnol. 2008, 3, 81–87.
Author of this NPG article	no
Your reference number	
Title of your thesis / dissertation	Synthesis and Characterisation of the Binding Mode of Azide-Bearing Derivatives of Hoechst 33258
Expected completion date	Oct 2016
Estimated size (number of pages)	250
Requestor Location	John M May

7.18.2. Copyright Permission for Figure 1.4.

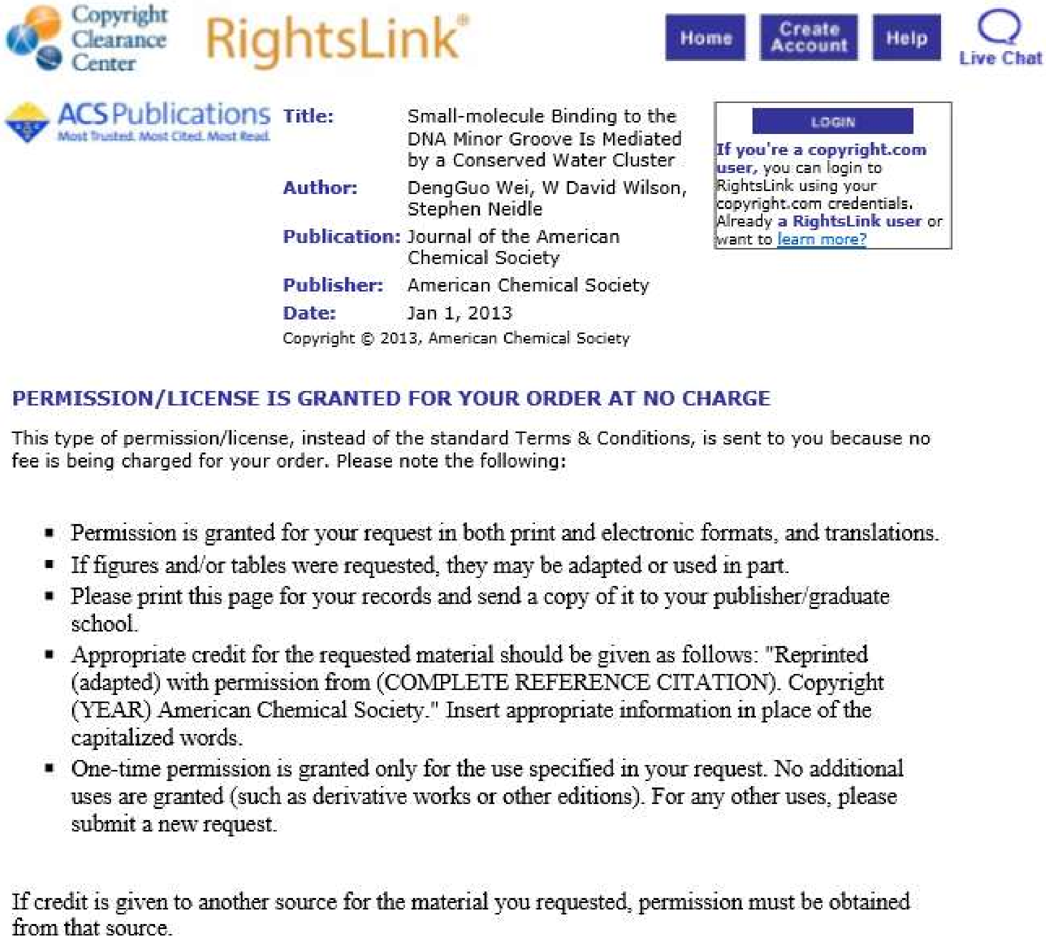
Dear Mr. May,

Thank you for your request. You may use this figure for your thesis.

Please use the correct credit in close proximity to the figure:

From: Lehninger Principles of Biochemistry 6E, by David L. Nelson, et al, Copyright 2013 by W.H. Freeman and Company. Used by Permission of the publisher

7.18.3 Copyright permission for Figures 1.6 and 1.23



Title:	Small-molecule Binding to the DNA Minor Groove Is Mediated by a Conserved Water Cluster
Author:	DengGuo Wei, W David Wilson, Stephen Neidle
Publication:	Journal of the American Chemical Society
Publisher:	American Chemical Society
Date:	Jan 1, 2013

Copyright © 2013, American Chemical Society

If you're a copyright.com user, you can login to RightsLink using your copyright.com credentials. Already **a RightsLink user** or want to learn more?

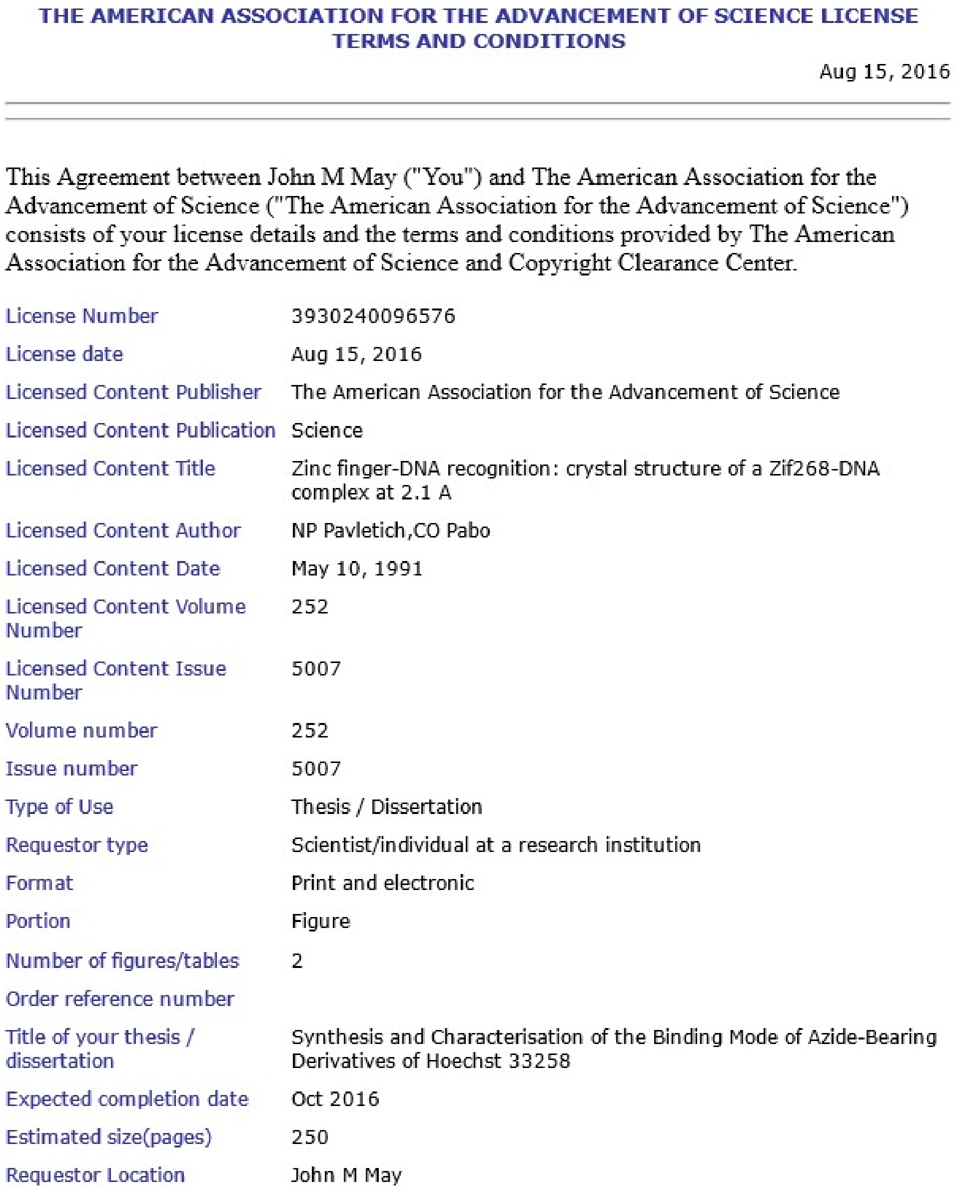
PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

If credit is given to another source for the material you requested, permission must be obtained from that source.

7.18.4. Copyright permission for Figure 1.8



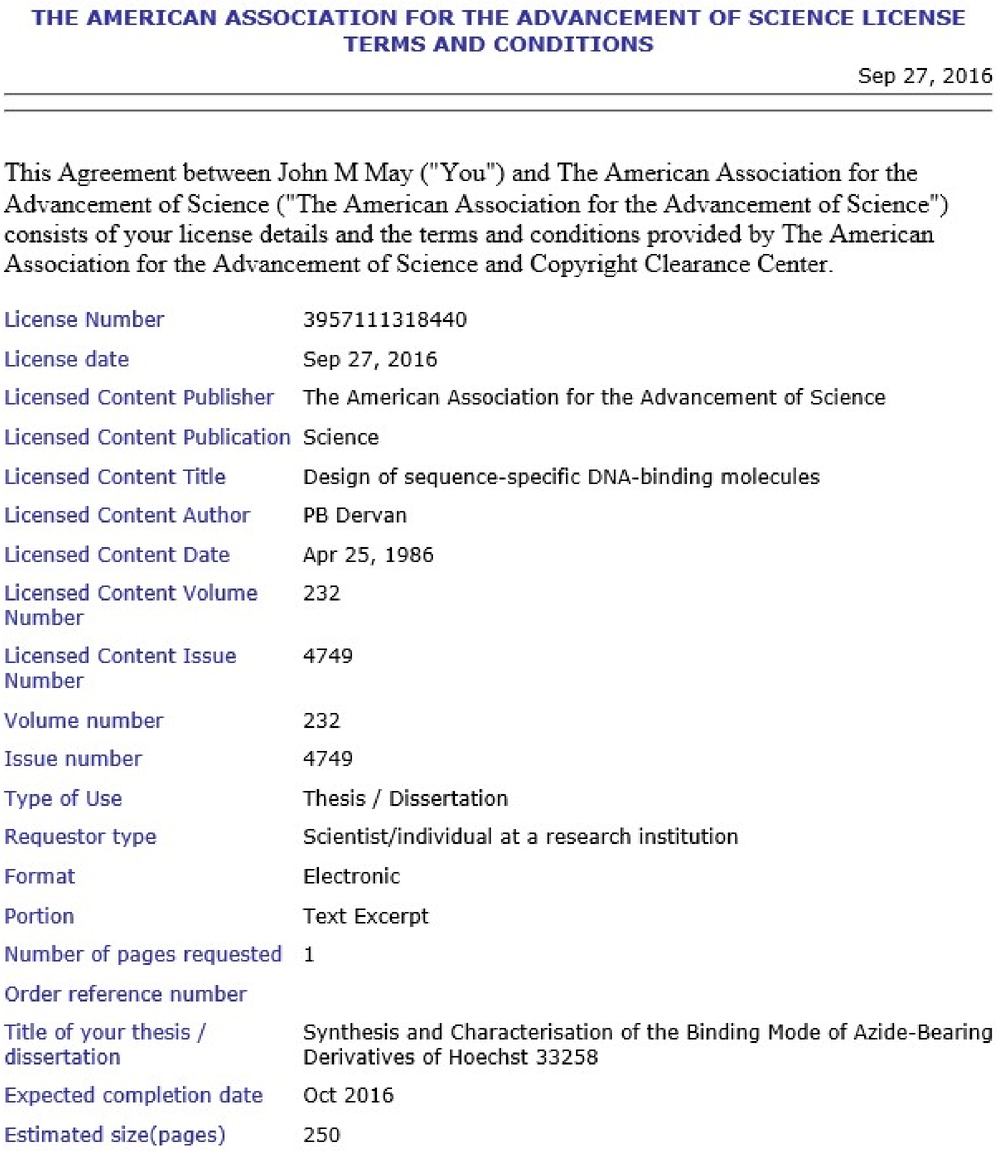
THE AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE LICENSE TERMS AND CONDITIONS

Aug 15, 2016

This Agreement between John M May ("You") and The American Association for the Advancement of Science ("The American Association for the Advancement of Science") consists of your license details and the terms and conditions provided by The American Association for the Advancement of Science and Copyright Clearance Center.

License Number	3930240096576
License date	Aug 15, 2016
Licensed Content Publisher	The American Association for the Advancement of Science
Licensed Content Publication	Science
Licensed Content Title	Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A
Licensed Content Author	NP Pavletich,CO Pabo
Licensed Content Date	May 10, 1991
Licensed Content Volume Number	252
Licensed Content Issue Number	5007
Volume number	252
Issue number	5007
Type of Use	Thesis / Dissertation
Requestor type	Scientist/individual at a research institution
Format	Print and electronic
Portion	Figure
Number of figures/tables	2
Order reference number	
Title of your thesis / dissertation	Synthesis and Characterisation of the Binding Mode of Azide-Bearing Derivatives of Hoechst 33258
Expected completion date	Oct 2016
Estimated size(pages)	250
Requestor Location	John M May

7.18.5 Copyright permission for Figure 1.26



THE AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE LICENSE TERMS AND CONDITIONS

Sep 27, 2016

This Agreement between John M May ("You") and The American Association for the Advancement of Science ("The American Association for the Advancement of Science") consists of your license details and the terms and conditions provided by The American Association for the Advancement of Science and Copyright Clearance Center.

License Number	3957111318440
License date	Sep 27, 2016
Licensed Content Publisher	The American Association for the Advancement of Science
Licensed Content Publication	Science
Licensed Content Title	Design of sequence-specific DNA-binding molecules
Licensed Content Author	PB Dervan
Licensed Content Date	Apr 25, 1986
Licensed Content Volume Number	232
Licensed Content Issue Number	4749
Volume number	232
Issue number	4749
Type of Use	Thesis / Dissertation
Requestor type	Scientist/individual at a research institution
Format	Electronic
Portion	Text Excerpt
Number of pages requested	1
Order reference number	
Title of your thesis / dissertation	Synthesis and Characterisation of the Binding Mode of Azide-Bearing Derivatives of Hoechst 33258
Expected completion date	Oct 2016
Estimated size(pages)	250