



# **Expanding the Heterocyclic Repertoire of DNA Minor Groove Binding Polyamides**

---

**Khalid Aman**

Thesis submitted for the degree of  
Master of Philosophy


8 January, 2019

**Academic Supervisor:** Professor G. A. Burley

### **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.

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

Signed:  (Khalid Aman)

Date: 8 January, 2019

# Contents

<b>Acknowledgements</b> .....	6
<b>Abstract</b> .....	7
<b>Abbreviations</b> .....	9
<b>Chapter 1</b> Introduction.....	14
1.1 Genes and the structure of DNA .....	15
1.2 Regulation of Transcriptional Initiation and Elongation .....	20
1.3 Exogenous Regulation of Transcription .....	22
1.4 DNA Recognition by Minor Groove Binders .....	23
1.5 Polyamides as a New Class of Programmable Minor Groove Binders.....	28
1.6 Nuclear Receptors as points of control for gene expression .....	28
1.7 Prostate cancer and the androgen receptor (AR).....	30
1.8 Polyamides as down-regulators of AR gene expression .....	32
1.9 Accessibility of Nuclear Chromatin.....	33
1.10 Py-Im polyamide Design Features and Pairing Rules.....	34
1.11 Thiazotropins.....	39
1.12 Influence of steric bulk on the polyamide-dsDNA binding interaction .....	41
1.13 Hypothesis.....	42
1.14 Project Aims and Objectives.....	43
<b>Chapter 2</b> Synthesis of an 8-ring hairpin polyamide incorporating a <i>N</i> -terminal <sup>1</sup> Pr-Im building block .....	45
2.1 Synthesis of polyamide building blocks .....	46
2.2 Solid Phase Synthesis of polyamides.....	49
2.2.1 Solid phase support .....	51
2.3 Aim and objectives of this chapter.....	53
2.4 Results and discussion .....	53
2.4.1 Synthesis of Py and Im monomers.....	53
2.4.2 Synthesis of the <i>N</i> -isopropylimidazole monomer.....	54
2.4.3 Solid phase synthesis of 8-ring hairpin polyamides.....	55
2.5 Summary .....	58
2.6 Experimental .....	59
2.6.1 General protocols .....	59
2.6.2 Synthetic procedures .....	60
2.6.3 Solid Phase Synthesis.....	63

<b>Chapter 3</b> Profiling the dsDNA binding kinetics of an 8-ring hairpin polyamide incorporating a <i>N</i> -terminal <sup>1</sup> Pr-Im.....	67
3.1 Kinetic rate theory of ligand-dsDNA binding.....	69
3.2 Rates of binding for comparison.....	71
3.2.1 Experimental analysis of PA-dsDNA interactions.....	71
3.3 Aim and Objectives.....	74
3.4 Results.....	75
3.4.1 Determination of DNA binding profile of PA4 and comparison with PA1-3.....	75
3.4.2 UV melting analysis.....	75
3.4.3 switchSENSE® fluorescence melting analysis.....	77
3.4.4 switchSENSE® association/dissociation kinetics measurements of polyamide-dsDNA complexes .....	79
3.5 Summary.....	83
3.6 Experimental .....	85
3.6.1 Pre-hybridization of DNA probes for kinetics experiments.....	85
3.6.2 Kinetics determination .....	86
3.6.3 Melting temperature determinations .....	87
<b>Chapter 4</b> Structural determination of polyamide-dsDNA complexes by NMR-restrained molecular dynamics simulations.....	89
4.1 NMR theory and MD theory.....	90
4.1.1 NMR first principles .....	90
4.2 Molecular Dynamics .....	93
4.2.1 Molecular mechanics of biological complexes from first principles .....	93
4.2.2 Previous structural work on polyamide-dsDNA complexes .....	96
4.3 Aim and objectives .....	99
4.4 NMR structure results and discussion.....	99
4.4.1 Workflow strategy for NMR-restrained MD simulations .....	99
4.4.2 Sample preparation .....	100
4.4.3 Acquisition parameters .....	104
4.4.4 Analysis of the NOESY spectra with SPARKY .....	106
4.4.5 MD simulations.....	113
4.5 Summary.....	118
4.6 Experimental .....	119
4.6.1 Preliminary Molecular Dynamics .....	119
4.6.2 NMR Distance Restraints.....	120
4.6.3 NMR-restrained molecular dynamics simulations.....	120
<b>Chapter 5</b> Future Work.....	122

<b>References</b> .....	126
<b>Appendix</b> .....	130
1 HPLC and LC-MS method parameters.....	132
Semi-preparative HPLC method.....	132
LC-MS method.....	132
Analytical HPLC generic methods.....	133
2 Spectroscopic characterization of compounds described in Chapter 2.....	134
3 NMR chemical shifts of PA·ODN4 complex.....	148
4 Profiling of the binding kinetics.....	154
5 Analysis of NMR-derived MD structure.....	159

## Acknowledgements

I would like to thank the University of Strathclyde for their generosity, facility support, and the opportunity to pursue an MPhil degree.

I would like to thank Prof. Glenn A. Burley for giving me the opportunity to work on this challenging but important project and all his support throughout the past year. I would also like to thank those who contributed to this project. In particular, I would like to thank Dr John A. Parkinson for help with navigating the complexities of nucleic acid NMR studies, including the acquisition of the high-quality NOESY data set for structural determination, Thomas Welte from Dynamic Biosensors for developing the switchSENSE® assay and analyzing the samples, and Dr Giacomo Padroni for his help with getting me up to speed with the synthetic protocols as well as strong support throughout the project.

I would also like to acknowledge present and past members (2017/2018) of the Burley group.

Of course, I would also like to thank my parents, Sayed and Bernadette, for being there and supporting me throughout my life. In both their direct and indirect support, they have taught me more about life than I can ever thank them for.

Lastly, I have to thank my fiancée Karen. She stood by my side throughout this endeavour, through the amazing, the challenging, and the downright difficult times. She provided helpful feedback and challenged me. If anyone should benefit from the work described in this thesis, they owe thanks to her.

## Abstract

Exogenous control of gene expression is an essential goal to better understand and ultimately treat diseases which involve dysregulation of transcription. In this effort, Pyrrole-Imidazole (Py-Im) hairpin polyamides (PAs) are minor groove binding (MGB) organic small molecules that demonstrate a versatile capability to recognize, bind, and alter transcription in vitro and, more recently, in vivo.<sup>[1]</sup> Despite their utility for distinguishing G-C/C-G base pairs, there are several limitations which impede the utilization of polyamides on scale, a major hurdle being that current monomeric pairings do not distinguish a T-A base pair from an A-T. To break this degeneracy, novel building blocks have been sought after to both improve selectivity as well as tune the pharmacokinetic properties of polyamides. Using an 8-ring hairpin scaffold that has demonstrated activity against prostate cancer, the impact of an *N*-isopropylimidazole monomer (<sup>i</sup>Pr-Im) in the terminal position is reported and a comparison made with the *N*-methylimidazole (Im) analogue, along with thiazole analogues studied by Padroni *et al.* (isopropyl thiazole, <sup>i</sup>Pr-Nt, and methyl thiazole, Me-Nt)<sup>[2]</sup>. UV absorption / fluorescence melt analyses and detailed interrogation of the binding kinetics suggest a strong affinity of hairpin polyamides incorporating the terminal <sup>i</sup>Pr-Im unit, comparable with Im. The data gathered further suggest that the <sup>i</sup>Pr-Im unit itself maintains specificity for G over C, A, or T. The switchSENSE® data also provided insight into the kinetic basis of the selectivity differences, showing a difference between the imidazole analogues and the thiazole analogues. A three-dimensional NMR structure of the <sup>i</sup>Pr-Im-containing polyamide in complex with a 12-mer sequence of double-stranded DNA containing the targeted androgen response element 5'-WWGWWCW-3' (W = A/T) revealed an enhanced major groove compression over the previously reported Im analogue, but also a deeper minor groove penetration than previously reported <sup>i</sup>Pr-Nt polyamide.<sup>[2]</sup>

Chapter 1 introduces the regulation of gene expression using MGB small molecules.

Chapter 2 describes the synthesis of hairpin polyamides, starting with the building blocks and then the solid phase synthetic protocol for constructing 8-ring hairpin polyamides.

Chapter 3 showcases the thermal UV melt data, and examines the polyamide binding kinetics with DNA duplexes. Further insight into the binding was obtained with switchSENSE® technology, which allowed the generation of kinetic rate maps. The data outlined in this chapter enabled the profiling of the kinetic parameters that govern PA-dsDNA binding.

Chapter 4 provides a structural view of a novel <sup>i</sup>Pr-Im-containing polyamide-DNA complex, highlighting the closer proximity of the N-terminal <sup>i</sup>Pr-Im N2 to the exocyclic amine of G5 compared with the N-terminal <sup>i</sup>Pr-Nt analogue. Furthermore, the 3D NMR-restrained molecular dynamics (MD) structure indicated the major groove compression was greater for polyamide with the terminal <sup>i</sup>Pr-Im monomer compared with the Me-Im analogue, highlighting the importance of a seemingly subtle change in steric bulk facing away from the minor groove for augmenting structural distortion of the DNA duplex.

Chapter 5 takes the key findings of this thesis and uses them as a lens for looking to the future direction that this research may take. Namely, the potential of this novel <sup>i</sup>Pr-Im monomer unit to replace imidazole as a G-selective residue that can induce greater structural distortion of the DNA duplex by polyamides.

[1] A. A. Kurmis, F. Yang, T. R. Welch, N. G. Nickols, P. B. Dervan, *Cancer Res.* **2017**, *77*, 2207–2212.

[2] G. Padroni, J. A. Parkinson, K. R. Fox, G. A. Burley, *Nucleic Acids Res.* **2018**, *46*, 42–53.



## Abbreviations

AR	Androgen receptor
ARE	Androgen Response Element
B <sub>0</sub>	Magnetic field
Boc	<i>tert</i> -Butyl carbamate
Boc-β-Ala-OH	<b>Boc-β-alanine</b>
Boc-Im-OH	4-(( <i>tert</i> -butoxycarbonyl)amino)-1-methyl-1H-imidazole-2-carboxylic acid
Boc-Py-OH	4-(( <i>tert</i> -butoxycarbonyl)amino)-1-methyl-1H-pyrrole-2-carboxylic acid
Boc-γ-turn-OH	<b>4-((<i>tert</i>-Butoxycarbonylamino)butyric acid</b>
Bn	Benzyl
bp	base pair
CD	Circular dichroism
CDI	1,1'-Carbonyldiimidazole
COSY	Correlation spectroscopy
CTD	C-terminal repeat domain
Cryo-EM	Cryo-electron microscopy
D <sub>2</sub> O	Deuterium oxide
DBD	DNA Binding Domain
DCC	<i>N,N'</i> -Dicyclohexylcarbodiimide
DCM	Dichloromethane
DCU	Dicyclohexylurea
DLS	Dynamic light scattering
DIPEA	Diisopropylethylamine
DMF	Dimethylformamide

DMSO	Dimethyl sulfoxide
DMSO-d <sub>6</sub>	Deuterated Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
Dp	Dimethylaminopropylamine
DQF-COSY	Double-quantum filtered Correlation Spectroscopy
dsDNA	Double-stranded DNA
DPPA	Diphenylphosphoryl azide
EC	Elongation Complex
EDC	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide
ER	Estrogen Receptor
ESI	Electrospray Ionisation
ESP	Electrostatic surface potential
EtOAc	Ethyl Acetate
EtOH	Ethanol
FID	Free Induction Decay
Fmoc	Fluorenylmethyloxycarbonyl protecting group
Fr	Furan building block
GABA	$\gamma$ -aminobutyric acid
HAT	Histone acetylase
HATU	1-[Bis(dimethylamino)methylene]-1 <i>H</i> -1,2,3-triazolo[4,5- <i>b</i> ]pyridinium 3-oxid hexafluorophosphate, <i>N</i> -[(Dimethylamino)-1 <i>H</i> -1,2,3-triazolo-[4,5- <i>b</i> ]pyridin-1-ylmethylene]- <i>N</i> -methylmethanaminium hexafluorophosphate <i>N</i> -oxide
HBTU	2-(1 <i>H</i> -benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HDAC	Histone deacetylase

HFIP	Hexafluoro-2-propanol
HOAt	1-Hydroxy-7-azabenzotriazole
HOBt	1-Hydroxy-benzotriazole
Hp	Hydroxypyrrole building block
HPLC	High pressure liquid chromatography
HSP	Heat shock protein
HSQC	Heteronuclear single coherence spectroscopy
Ht	Hydroxythiophene building block
ICD	Induced CD signal
Im	Imidazole building block
IPA	Isophthalic acid
ISPA	Isolated spin pair approximation
ITC	Isothermal Titration Calorimetry
$K_D$	Equilibrium Dissociation Constant
$k_{off}$	Dissociation rate constant
$k_{on}$	Association rate constant
LBD	Ligand Binding Domain
LC-MS	Liquid chromatography-mass spectrometry
LNA	Locked nucleic acid
LNCaP	Lymph Node Carcinoma of the Prostate
LogD	Distribution coefficient
MARDIGRAS	Matrix Analysis of Relaxation for Discerning the Geometry of an Aqueous Structure
MD	Molecular dynamics
MeCN	Acetonitrile
MeOH	Methanol

MGB	Minor Groove Binder
mRNA	Messenger RNA
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect spectroscopy
NR	Nuclear Receptor
NRPS	Non-Ribosomal peptidase
Nt	4-carboxylic thiazole building block
NTP	Nucleoside-5'-triphosphate
ODN	Oligodeoxyribonucleotide
PA	Polyamide
PAM	4-(Hydroxymethyl)phenylacetamidomethyl-polystyrene
PIC	Pre-initiation complex
PME	Particle-Mesh Ewald
PNA	Peptide Nucleic Acid
PSA	Prostate-Specific Antigen
Py	Pyrrole building block
PyBOP	Benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate
Pz	Pyrazole building block
rmsd	Root-mean-square deviation
RNA	Ribonucleic acid
RNAP	RNA polymerase
RP-HPLC	Reverse Phase High Pressure Liquid Chromatography
SPR	Surface Plasmon Resonance
SPS	Solid Phase Synthesis

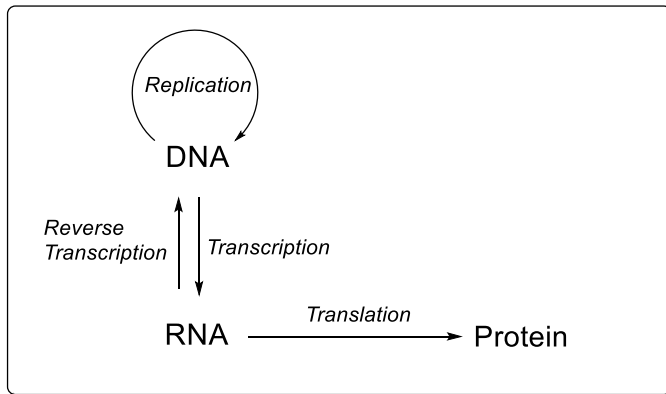
ssDNA	Single-stranded DNA
TBD	TATA-box Binding Domain
TEA	Triethylamine
TF	Transcription Factor
TFIIA	Transcription factor IIA
TFIIB	Transcription factor IIB
TFIID	Transcription factor IID
TFIIE	Transcription factor IIE
TFIIF	Transcription factor IIF
TFIIH	Transcription factor IIH
TFA	Trifluoroacetic acid
TFAA	Trifluoroacetic anhydride
Th	5-carboxylic thiazole building block
THF	Tetrahydrofuran
$T_m$	Melting temperature
TOCSY	Total correlation spectroscopy
TPW	TFA/phenol/water
TSS	Transcription Starting Site
UV	Ultraviolet

## **Chapter 1**

### **Introduction**

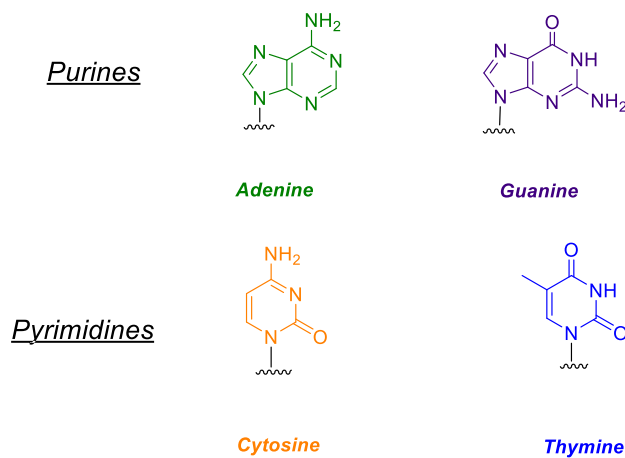
## 1.1 Genes and the structure of DNA

Biological processes that give rise to life are not random but coded by sequences of DNA known as *genes*. Current understanding places DNA's sole purpose as Nature's genetic repository. It is the reading of this genetic code by RNA polymerases that transcribe it into ribonucleic acid, or RNA, forming the critical first step in gene expression (Fig. 1.1.1).<sup>[3]</sup>



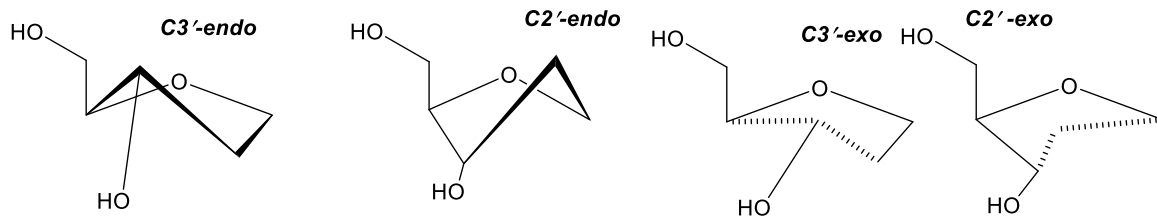
**Figure 1.1.1.** Central dogma of protein expression.

The primary structure of DNA consists of a linear chain of four nucleotides: adenine, guanine, cytosine, or thymine (A, G, C, and T, respectively) connected by a phosphodiester backbone. Nucleotides are monomeric units comprising three structural components: a deoxyribose sugar, a nitrogenous base, and a phosphate group. The four nitrogenous bases of DNA are from two categories, the purines and the pyrimidines (fig 1.1.2).<sup>[3]</sup>



**Figure 1.1.2.** Schematic of the four nucleobases A, G, C, and T that comprise DNA.

The sugar component in DNA is the pentose 2'-deoxy-D-ribose; and differs from RNA D-riboses by the absence of a hydroxyl group at the 2' position. As the ribose, which is in the  $\beta$ -furanose form, is not a planar molecule, it can be in one of 4 different 'pucker' conformations in the DNA backbone (Fig. 1.1.3).

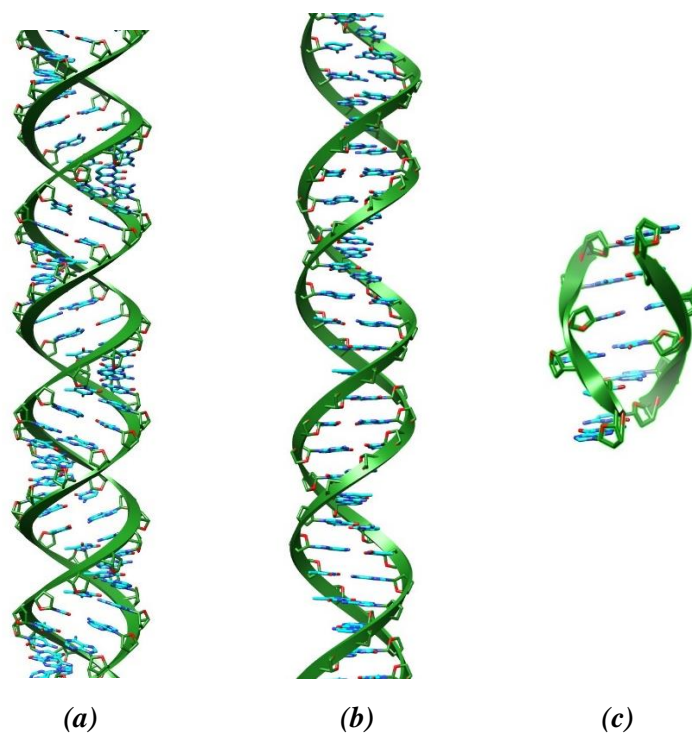


**Figure 1.1.3.** 2'-deoxyribose sugar 'pucker' conformations.

The phosphodiester group links the sugars to construct the backbone of the DNA strand with 5'-3' directionality (fig. 1.1.5(a)). The four nucleobases of DNA are linked covalently to the phosphodiester backbone through the anomeric carbon of their respective 2'-deoxy-D-ribose as shown in fig. 1.1.5(a).

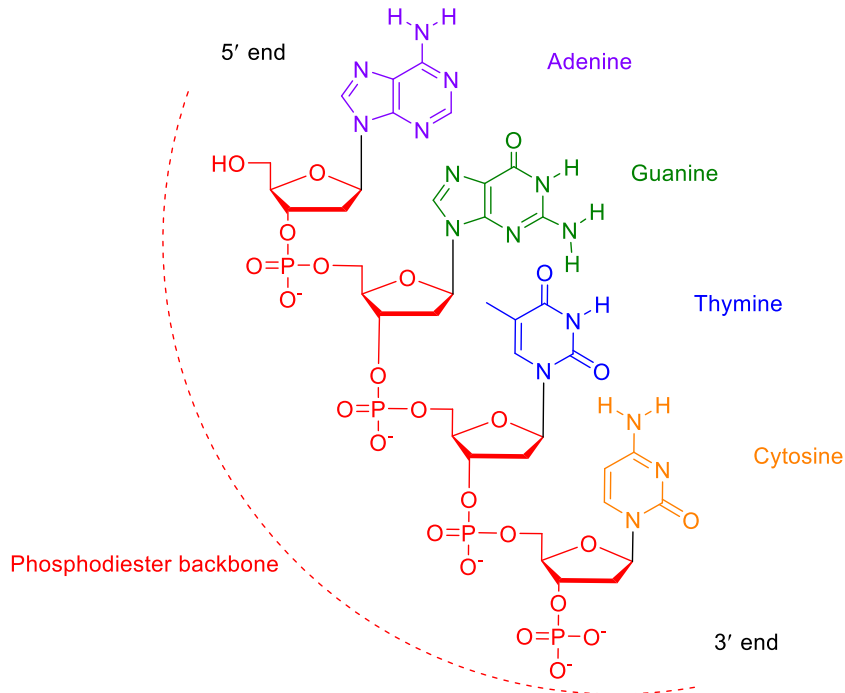
Double-stranded DNA (dsDNA) forms a complex quaternary structure from two molecules of single-stranded DNA (ssDNA). Requiring immense scientific effort to decipher, the true structure of the most stable and therefore the predominant form under physiological conditions, designated B-DNA, was solved by Watson and Crick in 1953.<sup>[3]</sup> They found that not only did A and T always pair up through hydrogen bonding interactions (as did G and C), but the two strands wound in a right handed double helix, with the negatively-charged phosphates necessarily orientated away from each other.<sup>[3]</sup> There have also been observed rarer variant forms of DNA (A and Z, fig. 1.1.4).<sup>[3]</sup>



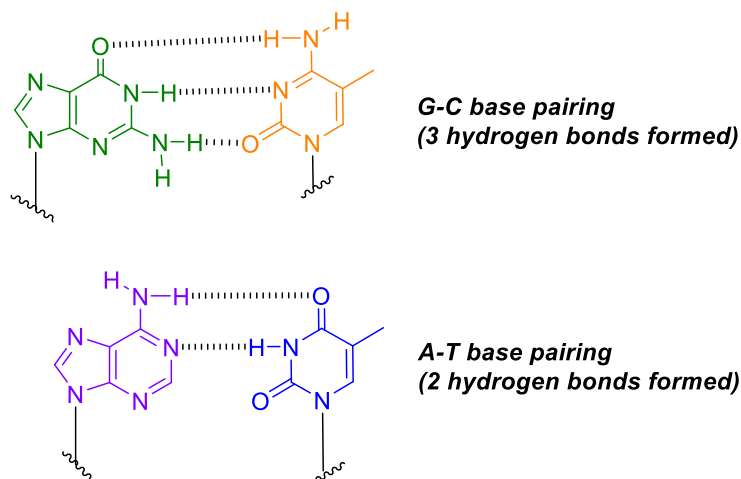


**Figure 1.1.4.** 3D structures of the different forms of DNA: (a) A-DNA, (b) B-DNA, (c) X-ray structure of Z-DNA (PDB ID 1W0E)<sup>[4]</sup>.

(a)



(b)

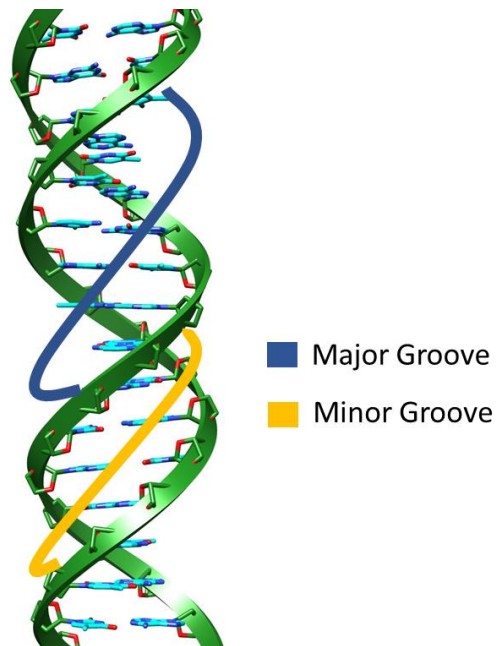


**Figure 1.15.** Primary structure of DNA (a) and nucleobase pairings through hydrogen bonding interactions (b).

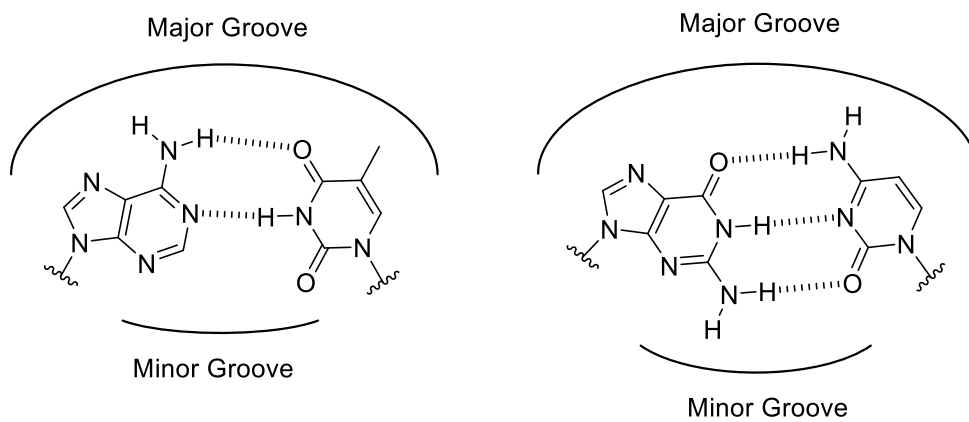
From the primary sequence of single-stranded DNA (ssDNA), a right handed double helix, or *duplex*, arises. This duplex is formed from the anti-parallel orientation of ssDNA to form the hydrogen bonding network and increase the base-stacking interactions between the nucleobases. These base-stacking interactions are the most stabilizing force of the dsDNA duplex.<sup>[3,5]</sup>

From examination of B-DNA structure, it is evident that there are two distinct regions in the helix that allow for access to the nucleobase pairs. These are termed the *major groove* and the *minor groove*.<sup>[3,6]</sup>

(a)



(b)



**Figure 1.1.6.** (a) The major and minor grooves of the B-DNA double-helix; (b) the nucleobase orientations presented from each groove to proteins or other DNA recognition molecules.<sup>[3]</sup>

The recognition of a specific portion of the nucleobase letter code, termed the *promoter* sequence, in DNA by RNA polymerases (RNAP) is where the process of transcription begins.<sup>[3]</sup>

## 1.2 Regulation of Transcriptional Initiation and Elongation

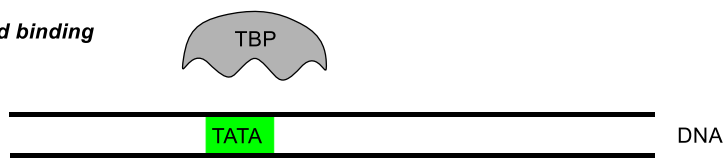
For the instructions coded within a genetic sequence to be carried out, transcription is the vital first step wherein enzymes known as RNA polymerases unwind the dsDNA, with one strand acting as the template for the synthesis of a new single strand of RNA. There are three types of these polymerases found in eukaryotes: RNA Pol I, responsible for rRNA synthesis, RNA Pol II for mRNA synthesis, and RNA Pol III which generates tRNA along with 5SRNA and other smaller RNAs.<sup>[7]</sup> Once complete, the new strand of RNA can then be used in specific expression pathways, dependent on its type.<sup>[3]</sup>

RNA polymerase II is vital for genetic expression in eukaryotic cells. In order for transcription to take place, RNA Pol II is assembled as a complex consisting of 12 subunits, beginning with the TATA-binding protein (TBP) drawing to the promoter site. This is followed by recruitment of additional transcription factors (TFIIB, TFIIE, TFIIF, TFIIH) along with Pol II to generate the ‘closed’ complex (fig. 1.2.1), also known as the pre-initiation complex (PIC).<sup>[3]</sup>

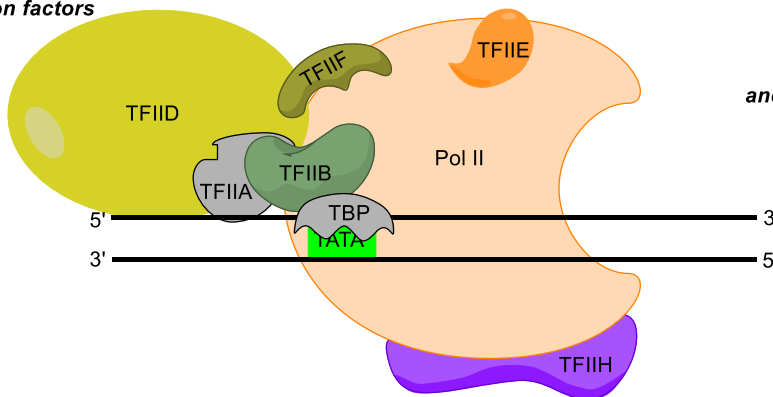
The DNA is unwound, transforming the complex into an ‘open’ state. Thereafter, phosphorylation of Pol II and initiation take place, followed by the release of several transcription factors and association of elongation factors, including P-TEFb (to inhibit pausing of Pol II), furthering along transcription. During the elongation stage, Pol II reads along the template strand, synthesizing the complementary nucleotides to create a new strand of mRNA in the 5’ to 3’ direction (strand read in reverse, 3’ to 5’).<sup>[3]</sup>

Termination occurs and the newly synthesized mRNA is released once the transcript synthesis is complete, with the RNA Pol II having reached a stop codon (sequence signaling the polymerase to halt transcription). This results in Pol II release and dephosphorylation, with the mRNA sent off to direct protein synthesis after splicing in a maturation stage.<sup>[3]</sup>

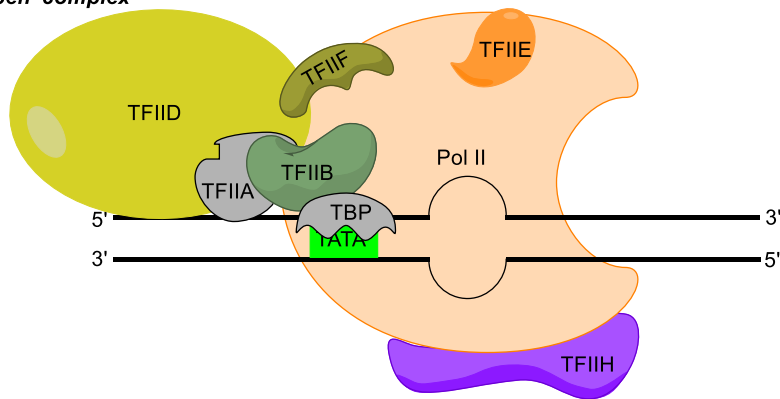
1. TBP recognition and binding to TATA sequence



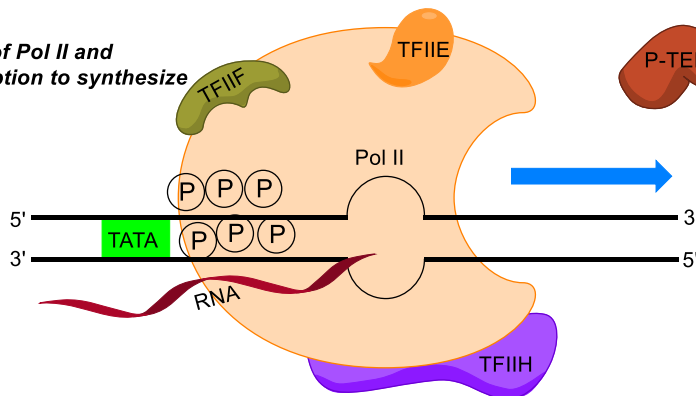
2. Recruitment of additional transcription factors



3. Unwinding of DNA to produce 'open' complex



4. Phosphorylation of Pol II and initiation of transcription to synthesize new RNA

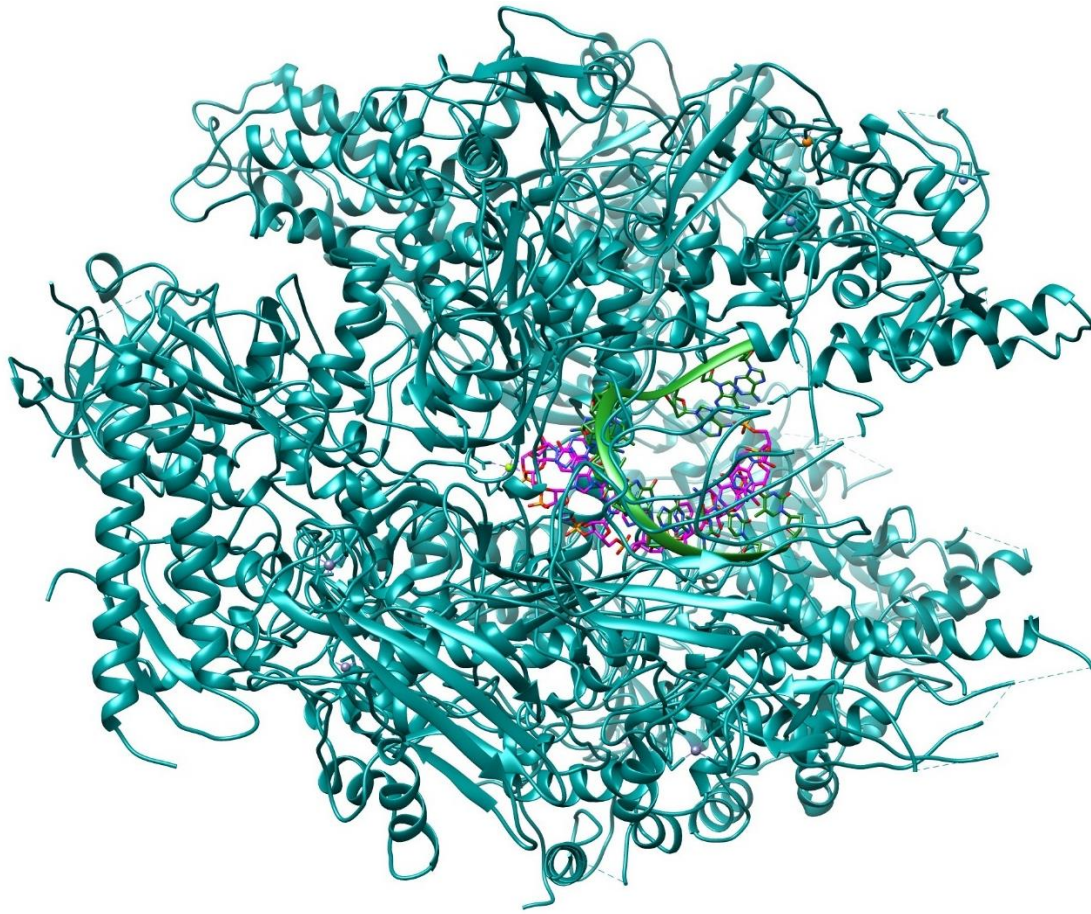


7. RNA Pol II release and dephosphorylation

6. Termination and RNA release

5. Elongation

Figure 1.2.1. Schematic of events outlining the DNA transcription process.<sup>[3]</sup>



**Figure 1.2.2.** Crystal structure of RNA Polymerase II elongation complex (cyan; PDB ID 1I6H).<sup>[8]</sup> DNA template is green and RNA synthesized is in magenta.

### 1.3 Exogenous Regulation of Transcription

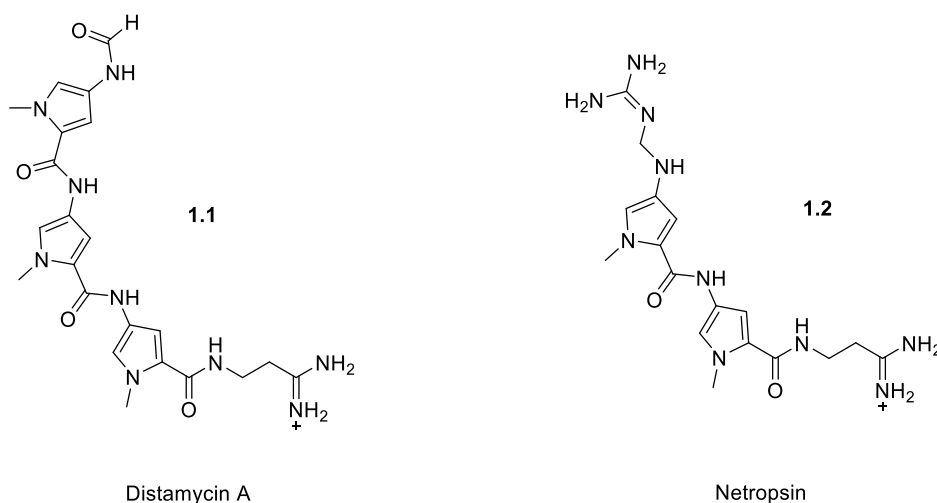
Since the subsequent steps of translation and protein synthesis are highly energy intensive, transcription is naturally an effective target for regulation of protein expression. Regulation is mediated by DNA-binding proteins, and for RNA polymerases this often involves interference with the promoter interaction, either directly or allosterically.<sup>[3]</sup> DNA-binding proteins that effect regulation target specific sequences, with an affinity that is generally  $10^4$  to  $10^6$  times higher than for other sequences (extremely selective for target sequences). These regulatory proteins can access the major and minor grooves of the dsDNA duplex in eukaryotes, even though it is buried deep within the convoluted heterochromatin structure. This is possible due to transcriptional activation signaling many conformational changes in the chromatin, known as chromatin remodeling, thereby opening up access to the DNA.<sup>[9]</sup>

Additionally, DNA-binding proteins known as transactivators and coactivators facilitate transcription factor association and themselves are points of regulatory control, responding in kind to the dynamic cellular environment.<sup>[3]</sup> A key capability of some DNA-binding transactivators is their strong affinity for a specific target site in the dsDNA sequence, despite the highly inaccessible condensed state of the chromatin. This provides a precedent for the potential access of single molecules that, should the affinity and selectivity be high enough, are able to reach and bind to a target site of the dsDNA duplex.<sup>[3,9]</sup>

This thesis explores one such strategy for a type of single molecule that can be utilized for the downregulation of a particular subset of eukaryotic genes. In particular, the focus of this work will be towards an application for the androgen response element of prostate cancer (CaP).

## 1.4 DNA Recognition by Minor Groove Binders

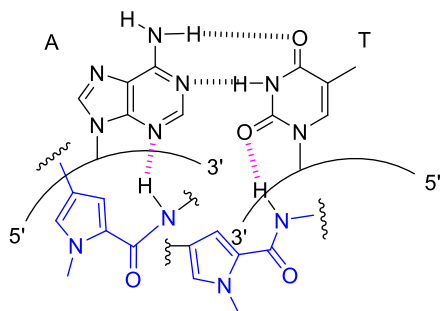
The discovery of the natural products Distamycin A (isolated from *Streptomyces distallicus*) and netropsin (isolated from *Streptomyces netropsis*) shed light on a new class of small molecule capable of recognition and binding to A/T-rich sequences in the minor groove of B-form DNA (structures shown below in fig. 1.4.1).<sup>[10]</sup>



**Figure 1.4.1.** Natural products Distamycin A and Netropsin share common structural features.

The structure of **1.1** and **1.2** consists of a formamide-amidine (Distamycin A) or a guanidine-amidine (Netropsin) head-tail design linked by repeated N-methylpyrrole units through amide bonds.

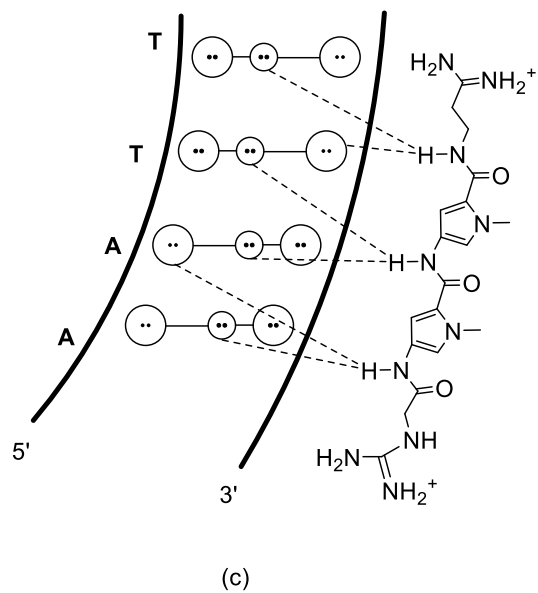
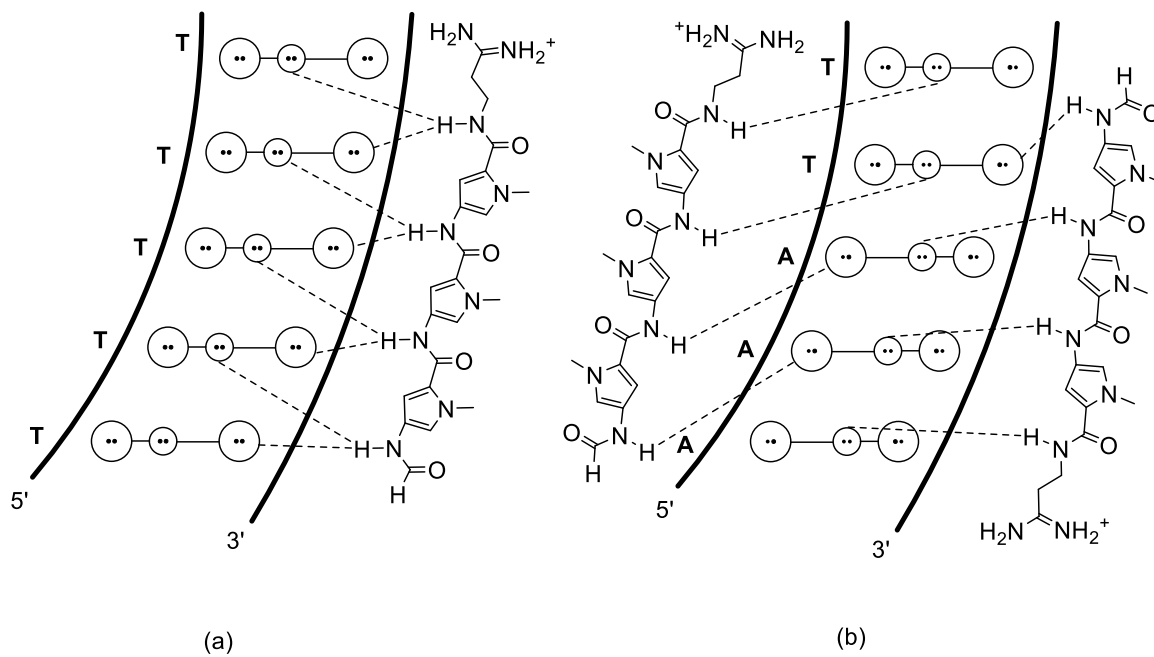
The key interaction that affords recognition of the DNA base pairs is hydrogen bonding of the amide protons by the thymine O2 and the adenine N3 atoms, as depicted in fig. 1.4.2.<sup>[11]</sup>



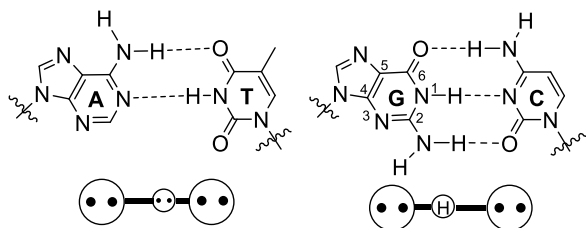
**Figure 1.4.2.** Recognition interactions between the amide protons of the natural products and A-T base pairs of dsDNA are through hydrogen bonding contacts.

The binding mode of Distamycin A can be 2:1 or 1:1 with respect to dsDNA.<sup>[12]</sup> In the case of 2:1 binding, the amide bonds of one molecule overlap the pyrrole rings of another, thereby affording recognition of a total seven base pairs (in contrast to the 4 base pairs in the 1:1 binding mode). Netropsin has only ever been reported in the 1:1 binding mode.<sup>[12,13]</sup>

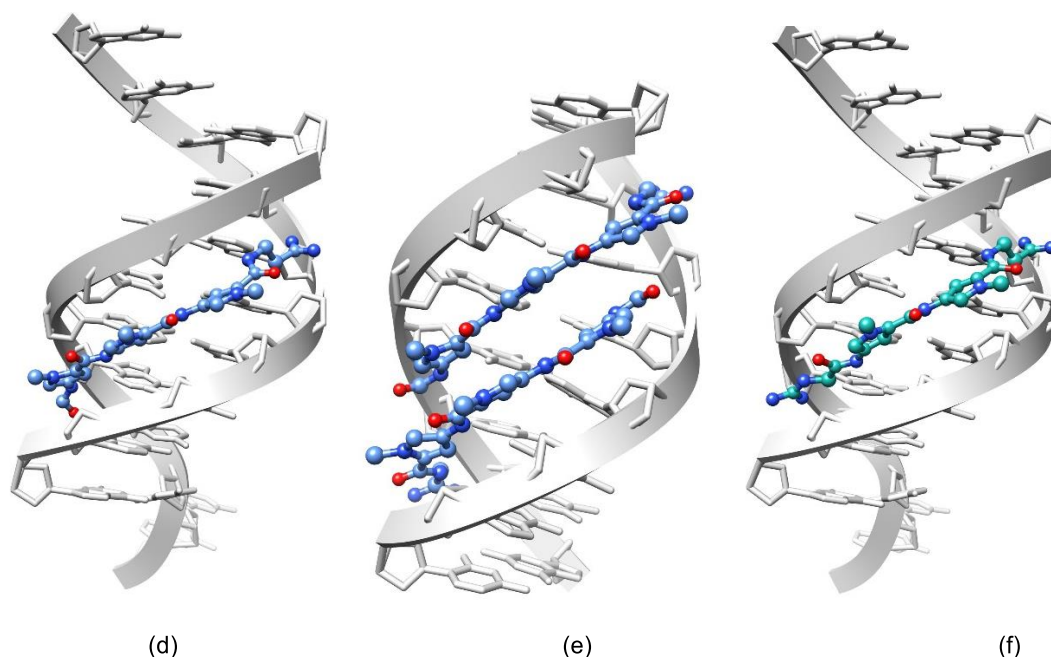




Base pairing schematic legend:



(Figure caption is on the next page).



**Figure 1.4.3.** (a) Observed hydrogen bonding pattern of Distamycin A in (a) 1:1 and (b) 2:1; Netropsin was only reported in 1:1 (c); Crystal structures of (d) 1:1 (PDB ID 1JTL)<sup>[14]</sup> and (e) 2:1 (PDB ID 378D)<sup>[15]</sup> binding of Distamycin A and (f) 1:1 binding of Netropsin (PDB ID 1Z8V)<sup>[16]</sup> to dsDNA.<sup>[17]</sup>

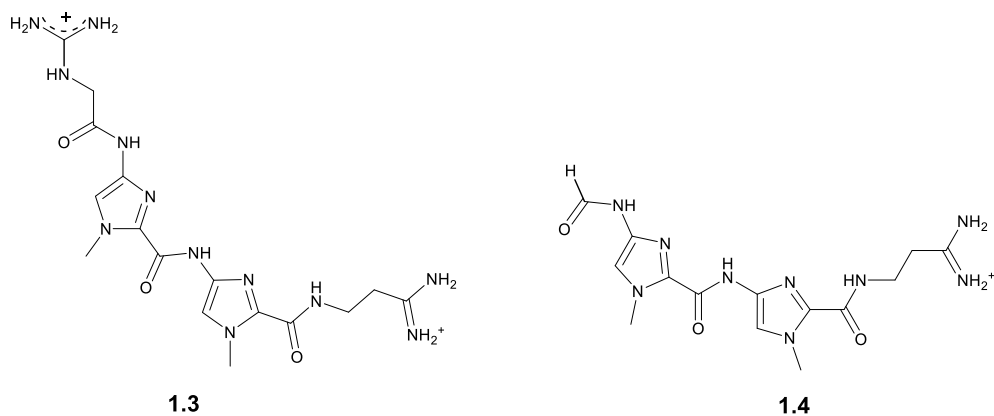
With a more complete physical description paramount to this effort, it was an X-ray structure of the Netropsin-dsDNA complex obtained by Richard Dickerson that made clear how the ligand fit into the minor groove: the amide bonds directed towards the floor, thereby allowing for the hydrogen bonding contacts outlined above in figure 1.4.3.<sup>[11,18]</sup> Furthermore, the curvature of distamycin A matched the register of the dsDNA minor groove. Based on this observed recognition of the Watson-Crick base pairs within the minor groove of B-DNA, scientific efforts have focused on understanding and exploiting the dsDNA-binding capability of these polyamide natural products in the design of small molecules to target specific base pairs.

The initial designs of synthetic minor groove binding polyamides, or *lexitropsins*, involved the replacement of the *N*-methylpyrrole monomer unit with an *N*-methylimidazole.

### *Lexitropsins*

With inspiration from the natural products Distamycin A and Netropsin, an innovative class of small molecules known as the *lexitropsins* were developed with the intent of targeting specific sequences of

DNA. A breakthrough in the recognition capability was achieved by Lown *et al.*, replacing the pyrroles of netropsin with imidazole analogues.<sup>[18-20]</sup> Effective discrimination of G/C from A/T was achieved by replacing the guanidinium functional unit of the dicationic Im-lexitropsin **1.3** with a formyl group, as in **1.4** (fig. 1.4.4).<sup>[18,19]</sup>



**Figure 1.4.4.** Lexitropsins integrating imidazole units as dicationic (**1.3**) or monocationic (**1.4**).

Kopka *et al.* also reported that this single atom modification allowed for the recognition of the N2 of a guanosine nucleotide through a hydrogen bonding interaction between the exocyclic amine of the targeted guanine and N3 of the imidazole residue.<sup>[18]</sup>

## 1.5 Polyamides as a New Class of Programmable Minor Groove Binders

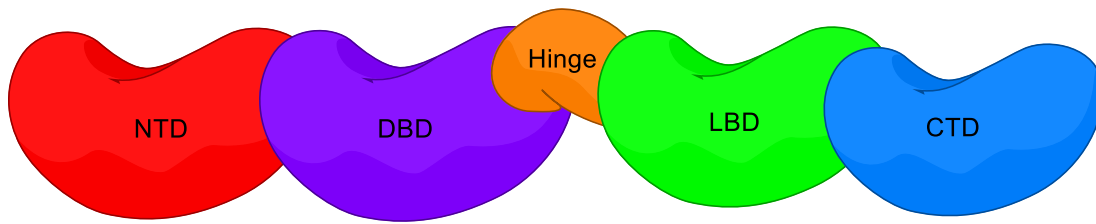
Taking inspiration from the natural minor groove binders, the need for more selective and programmable analogs drove the design of polyamide architectures wherein aromatic heterocycles (Im and Py) were connected in sequence by peptide bonds. Increased recognition selectivity and higher affinity was obtained when pairing heterocycles through alkyl tethers, which allowed for 2:1 binding recognition in a hairpin.<sup>[21-23]</sup>

It has been shown that for a select sequence, polyamides can bind with nanomolar affinity, in turn disrupting transcription factor-dsDNA binding interactions of specific genes.<sup>[22,24]</sup> This attribute allows for the deployment of polyamides to regulate gene transcription, making them a potential new class of precision therapeutics. Based on some success *in vitro*, early *in vivo* studies have so far suggested that an 8-ring hairpin polyamide could out-compete the transcription factor TFIID from binding to RNA-pol III, and in doing so halt the transcription of 5S RNA in cultured *Xenopus* kidney cells.<sup>[25]</sup> However, as subsequent chapters will outline, there are several limitations to the widespread development of polyamides as a class of effective therapeutics.

## 1.6 Nuclear Receptors as points of control for gene expression

Capable of high-affinity binding to dsDNA, nuclear receptors (NRs) are a family of transcription factors that, upon binding of a ligand (*e.g.*, steroid hormones, such as 5 $\alpha$ -dihydrotestosterone) can initiate transcription of a target subset of genes. This key ligand-protein interaction is what affords a high level of regulation, rendering the activity of NRs to be interlinked with stimuli (the ligands) from the cellular environment.<sup>[3]</sup> NRs are ideal points for fine-tuning gene expression, as demonstrated by their proliferation as targets for drug development efforts towards cardiovascular disease, diabetes, reproductive disorders, and various cancers.<sup>[9]</sup>

The basic structure of nuclear receptors, which can be mono-, di-, or trimers, or heterodimers, consists of an N-terminal domain (NTD), a DNA-binding domain (DBD), a ligand binding domain (LBD), a 'hinge' linker to tether the DBD to the LBD, and, in some NRs, a C-terminal domain.<sup>[3,9]</sup>



**Figure 1.6.1.** Nuclear receptor domains.

The NTD exhibits promoter and cell-specific activity, thereby playing a role in specificity of action among receptor isoforms. The modulatory domain is the target for phosphorylation mediated by different signaling pathways.

The DBD consists of 9 Cys residues among others that are highly conserved. Two Zinc-fingers span ca. 60-70 amino acids (AA's) and a C-terminal extension, containing the T and A boxes. The Cys residues coordinate to the zinc ions in a tetrahedral fashion, with the complex folding together. One of the two  $\alpha$ -helices at the core of the DBD binds to the B-DNA major groove, and the other sits orthogonally to the first.<sup>[9]</sup> The hinge is a relatively variable domain, allowing for a flexible tethering of the DBD and LBD domains. The hinge domain also bears some nuclear localization activity.<sup>[9]</sup>

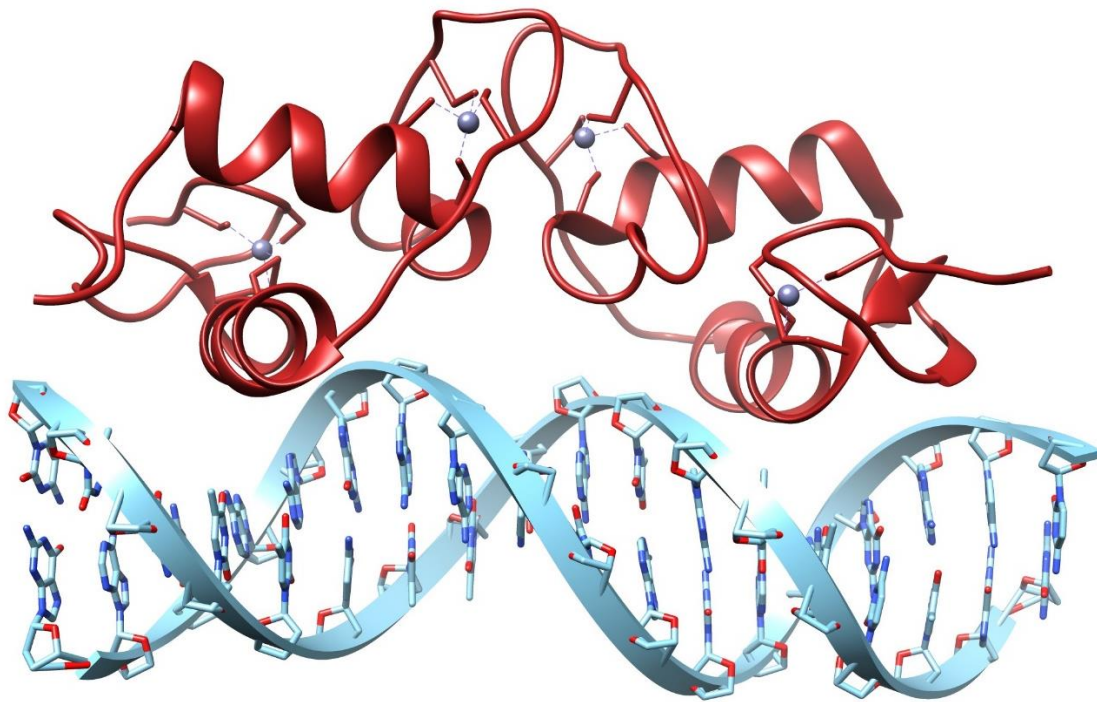
The LBD has several functions, including ligand binding and regulation of transcriptional activity, mediating homo- and hetero-dimerization, and interaction with HSP-90 proteins, among other roles. The LBD is formed by 12 well-conserved  $\alpha$ -helices, 11 fixed and one with mobility. They fold in a fashion that creates a 3-layered network, where the core is situated between two other layers to present a cavity for the ligand to slot into. Although there are variations on this architecture, the overall shape of the LBD is well-conserved across the NR superfamily. It is also important to note that the binding of a ligand to the NR induces a conformational shift to a more compact form.<sup>[9]</sup>

Upon activation, nuclear receptors demonstrate recognition of specific DNA sequences, termed *response elements*. NRs are also capable of binding with high affinity for regulating transcription, which is achieved through modifications to the chromatin structure through various mechanisms, one of which involves the recruitment of histone acetylase (HAT).<sup>[9]</sup> The androgen receptor, like most steroid hormone receptors, binds to palindromes of 5'-AGAACA-3', and this is its *response element*. The

binding is regulated by the two main androgenic hormones dihydrotestosterone (DHT) and testosterone.<sup>[26]</sup>

## 1.7 Prostate cancer and the androgen receptor (AR).

The androgen receptor (NR3C4), a member of the steroid hormone receptor subfamily, is a nuclear pore receptor that is reliant on 5 $\alpha$ -dihydrotestosterone or testosterone binding for its activation as a transcription factor, controlling the expression of many target genes.<sup>[26]</sup> Like the rest of the subfamily, it comprises a conserved three-domain structure (fig. 1.7.1): the N-terminal domain (NTD), the DNA-binding domain (DBD), and the C-terminal ligand binding domain (LBD).



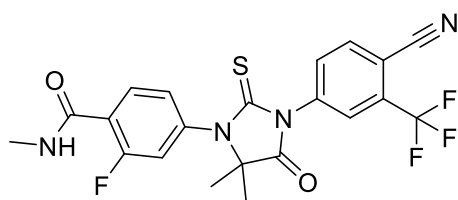
**Figure 1.7.1.** Crystal structure of the androgen receptor (in red) and Androgen Response Element (ARE) sequence (blue) (PDB ID 1R4I).<sup>[27]</sup>

Ultimately, prostate cancer cells undergo higher proliferation than cell death, and the AR controls this ratio.<sup>[28]</sup> Complete removal of the androgen receptor (via excision of the testes) has been shown by Huggins *et al.* to lead to regression of prostate cancer.<sup>[29]</sup> This suggested that control of prostate cancer is possible with removal of the androgen receptor, as could less invasively be accomplished via its biomolecular down-regulation.

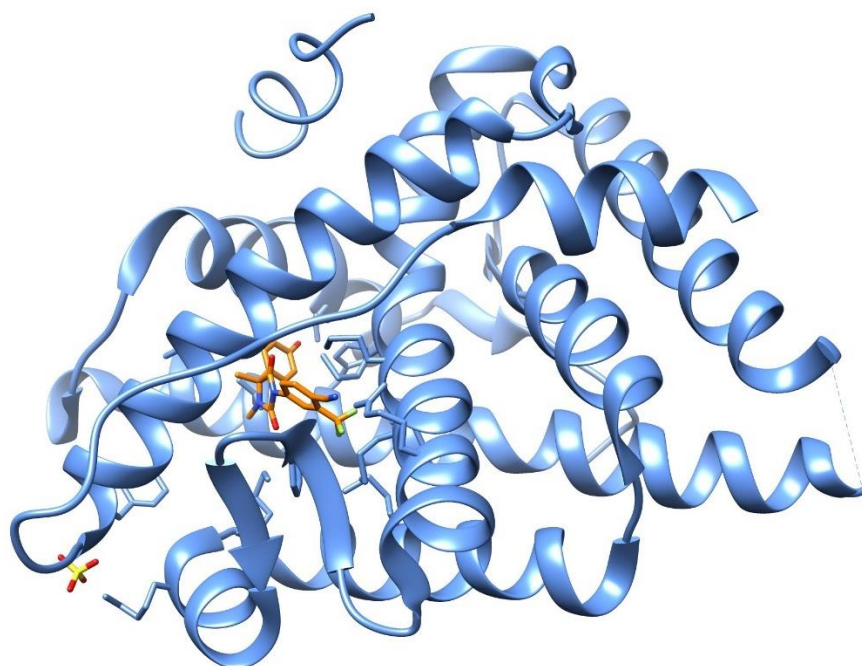
Androgen deprivation therapy (ADT) involves blocking of the production or effects of testosterone in addition to other hormones, and as a result the AR complex is effectively inactivated. In conjunction with radiotherapy, ADT can be effective at controlling levels of cancerous cell proliferation, eventually leading to remission.<sup>[30]</sup>

Enzalutamide, a small molecule that binds to the AR-ligand binding domain (ABD) and downregulates transcription, is the current standard of care (SOC).<sup>[31]</sup>

(a)



(b)



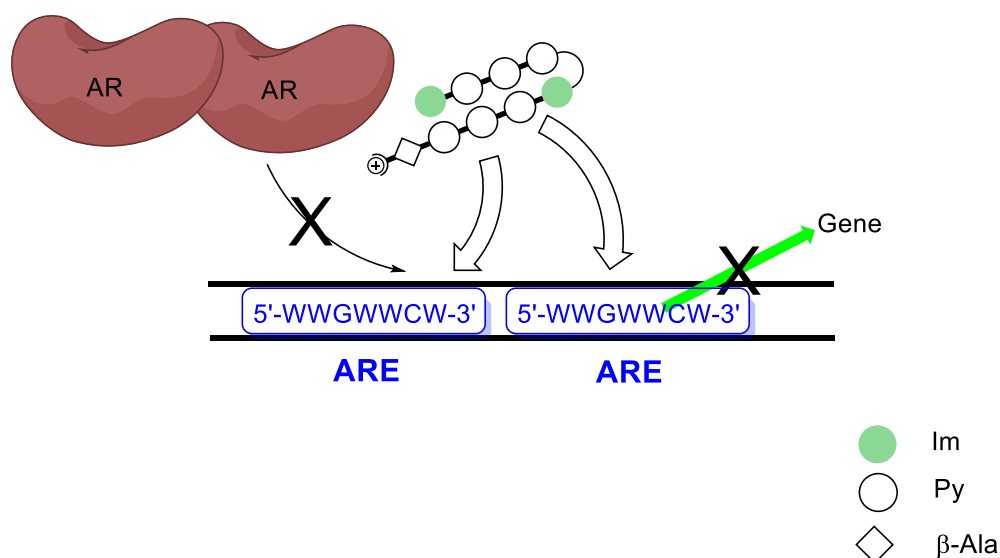
**Figure 1.7.2.** (a) Structure of enzalutamide; (b) crystal structure of AR LBD (blue) complexed with coactivator peptide (*orange*) (PDB ID 3V49)<sup>[31]</sup>, indicating the binding pocket where the enzalutamide inhibitor would sit.

Despite the potential of these treatments for dealing with prostate cancer, this remains an unmet medical need. In 2018, it is estimated that new diagnoses will reach 164,690 Americans, and about 29,430 deaths from prostate cancer, costing the healthcare system in the US alone \$11.85 billion.<sup>[32,33]</sup>

One key reason for this is the proliferation of a hormone refractory version of prostate cancer that involves genes under the control of the androgen receptor. This version of prostate cancer is a splice variant, potentially arising from the genomic rearrangement of the AR gene and/or alternative splicing of the AR pre-mRNA.<sup>[34]</sup> The existence of castration-resistant prostate cancer (CRPC) signaled the necessity for a new strategic approach, and the emergence of resistance towards enzalutamide further exacerbated this cancer's proliferation.<sup>[35]</sup>

## 1.8 Polyamides as down-regulators of AR gene expression

The androgen response element (ARE) is a consensus sequence 5'-WWGWWCW-3' (W=A/T) and is the site of AR binding. It is a dimeric sequence and the androgen receptor (AR) binds as a dimer enabling the transcription of a subset of genes for proliferation. Due to its highly conserved nature, the ARE is a viable target for disruption of the nuclear receptors' binding and is the logical focus for the design of a DNA minor groove binding agent. To this end, a variety of hairpin and cyclic polyamides have been explored as potential transcription blockers.<sup>[1]</sup>

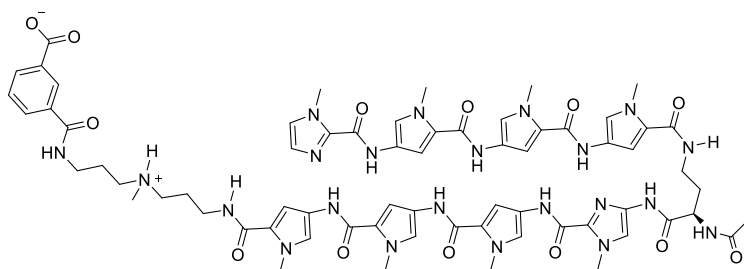


**Figure 1.8.1.** Schematic of ARE downregulation by polyamides.

Most recently, the discovery of ARE-1, an 8-ring hairpin polyamide (fig. 1.8.2), was found to be effective against enzalutamide-resistant VCaP and LREX' CaP xenograft and cell cultures. Dervan *et al.* found that long-term treatment with ARE-1 led to a substantial reduction in transcription of the



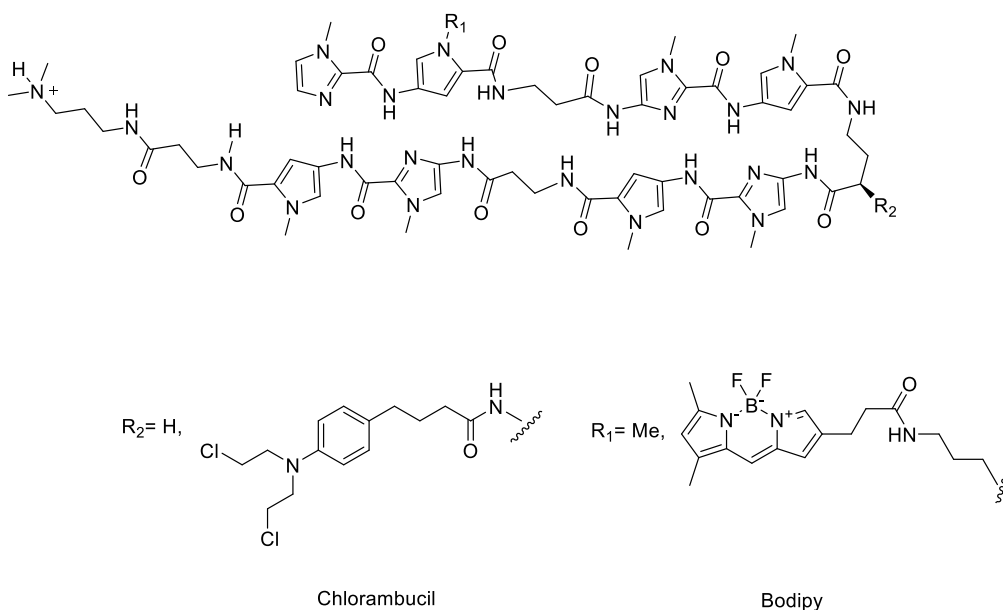
androgen receptor, as revealed in a head to head study with enzalutamide.<sup>[1]</sup> That study revealed that ARE-1 was significantly more efficacious.



**Figure 1.8.2.** Structure of ARE-1 polyamide.<sup>[1]</sup>

## 1.9 Accessibility of Nuclear Chromatin

Each human cell contains more than 2 m of DNA, and the packaging of this material within the nucleus requires very tight compaction.<sup>[3]</sup> For a small molecule, such as a polyamide, to bind to a target sequence in the dsDNA minor groove, the most obvious issue is that of access: can the polyamide get to its target sequence and bind with not only the desired affinity, but also the required specificity? Seeking an answer to that crucial question, Joel Gottesfeld in collaboration with Peter Dervan published in 2003 that an 8-ring hairpin polyamide, with either a Chlorambucil or Bodipy covalently attached to the free amine of the GABA turn unit, did indeed reach its target sequence, binding with a high degree of affinity and perhaps surprising specificity.<sup>[36]</sup>

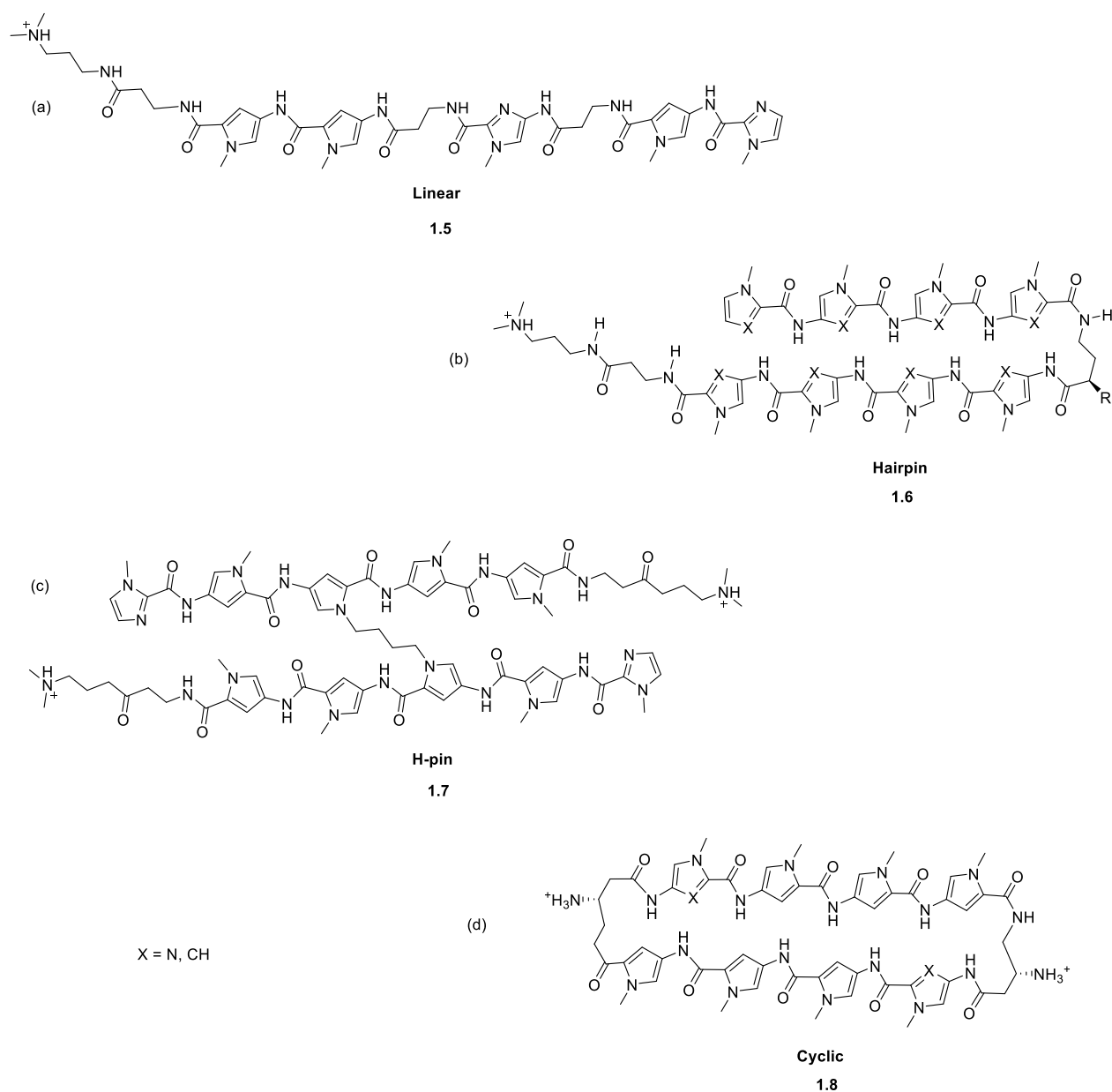


**Figure 1.9.1.** Chemical structures of chlorambucil and bodipy for PA modification for nuclear chromatin localization study.<sup>[36]</sup>

This study demonstrated that within the cellular environment, the hairpin polyamides could get to where they needed to be effective as exogenous gene regulators. As is the case with nature's challenges, there have been many more examples where the polyamides could not get to their target site, owing to poor cellular uptake as well as non-productive residence in the cytoplasm.<sup>[37]</sup>

## **1.10 Py-Im polyamide Design Features and Pairing Rules**

The simplest structure for a polyamide is a linear format, where the linking of the Py/Im units requires periodic  $\beta$ -alanine interspaces for maintaining a proper fit to the curvature of the helix for extended sequence recognition. The hairpin design (fig. 1.10.1), developed by Peter Dervan is the current "state of the art" and is the focus of a large effort towards practical applications due its superior binding affinity and specificity as compared to other alternative architectures of polyamides studied (H-pin and cyclic designs).<sup>[38]</sup>

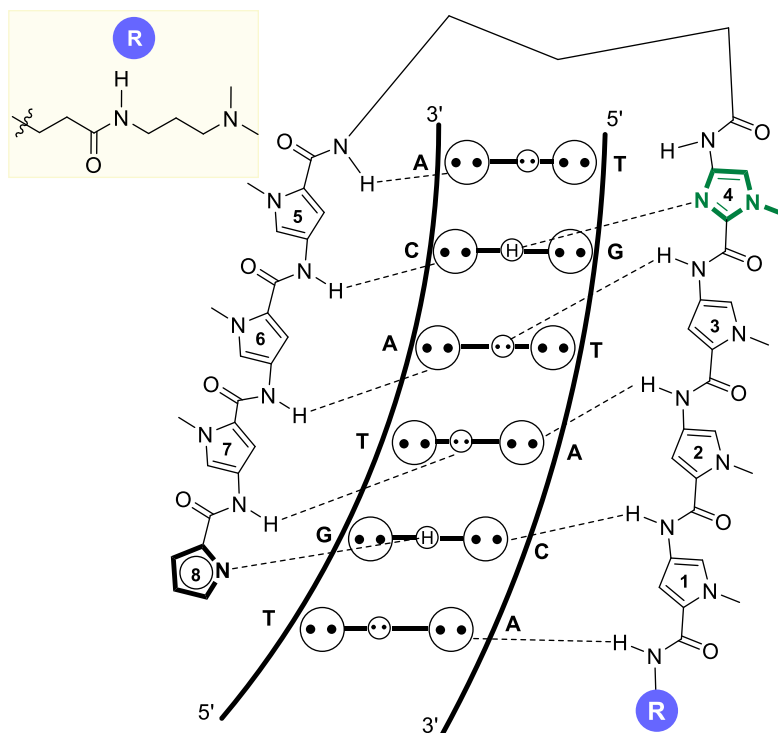


**Figure 1.10.1.** Polyamide architectures.

Among the designs studied, two are the most promising. The first is the linear format (fig. 1.10.1, **1.5**), where interspersed  $\beta$ -alanine residues are required for “resetting” the curvature, thereby maintaining a proper fit to the DNA minor groove. The second is the hairpin scaffold, which has two sets of amino acids linked by a GABA unit to allow for locking of the Py/Im pairings.<sup>[38]</sup> Prior the conception of the hairpin polyamide design, Peter Dervan at the California Institute of Technology discovered that binding rules existed for polyamides that formed hydrogen bonding contacts in recognition of specific nucleotide base pairs.<sup>[39,40]</sup>

These pairing ‘rules’ are outlined here:

- **Py/Im or Im/Py pairings can discriminate between C-G and G-C, respectively**
- **Py/Py is degenerate for A-T and T-A**
- **Hp/Py can discriminate T-A over A-T**



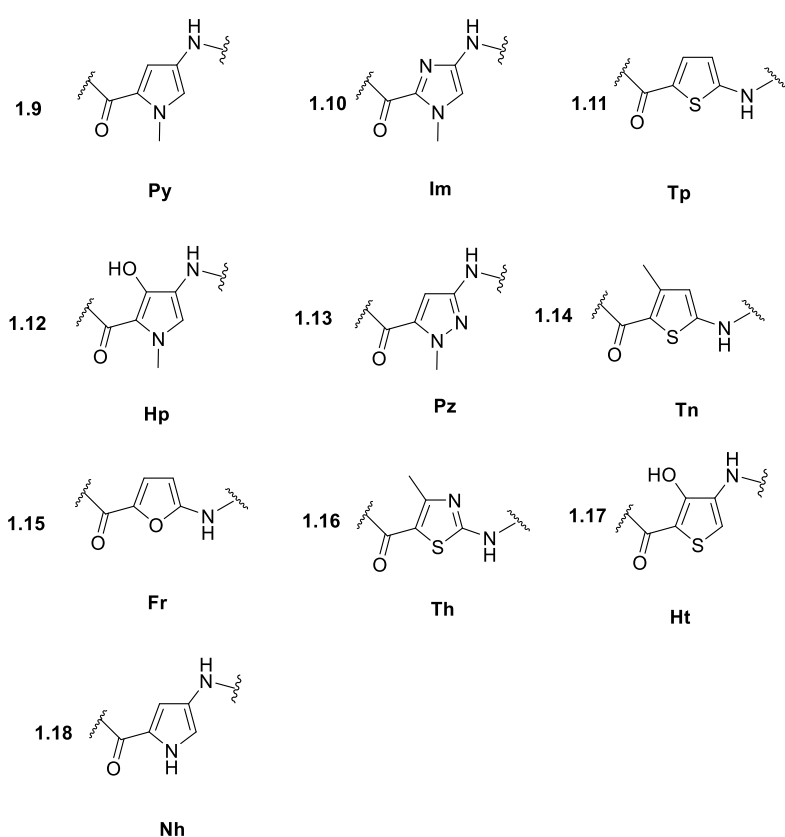
**Figure 1.10.2.** 8-ring hairpin polyamide-dsDNA schematic depicting the hydrogen bonding behind the pairing rules.

The hydroxypyrrole (Hp) residue designed, proposed, and studied extensively by Dervan’s group, initially showed promise in targeting the T nucleobase.<sup>[41,42]</sup> With the pyrrole shape maintaining the curvature of the polyamide, the C3 hydroxyl group added a steric clash with the larger A compared with T, giving good selectivity. Furthermore, there was an additional hydrogen bonding contact made.

Despite the promise of the hydroxypyrrole (Hp) unit, there existed a major complication. While the Hp residue was found to be able to distinguish T-A from A-T when paired opposite Py, the unit itself was found unstable in the presence of acid or free radicals.<sup>[43]</sup> This instability has been potentially attributed to the Hp’s electron-rich nature. Ultimately, this lack of stability has limited the utilization of these

polyamides in a biological context, re-establishing the need for a polyamide design to break the A-T/T-A degeneracy.<sup>[41,42,44]</sup>

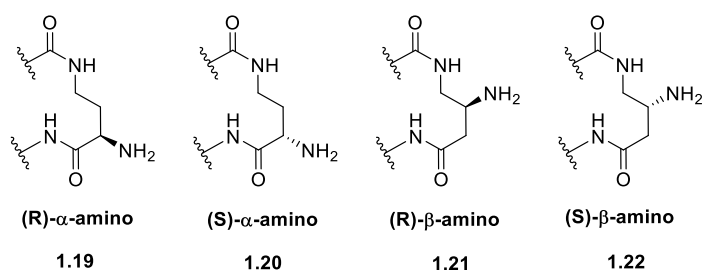
Since the inherent limitation caused by the AT/TA pairing degeneracy, scientific efforts were focused on identification of alternative heterocycles with the intent of finding a more robust yet sufficiently selective unit. At the time of the writing of this thesis, that objective has not been achieved, despite the screening of various heterocyclic monomers (fig. 1.10.3), some of which have abolished binding entirely.<sup>[45]</sup>



**Figure 1.10.3.** Standard Py, Im heterocyclic units (1.9 and 1.10, respectively) alongside the variety of explored heterocyclic monomers.<sup>[45]</sup>

The hairpin design of polyamides was found to enhance the binding affinity, limiting the degrees of freedom of the pyrrole-imidazole ring pairs and preventing them from slipping.<sup>[46]</sup> The current state-of-the-art scaffold covalently tethers two 4-ring Py/Im sequences by a  $\gamma$ -amino butyric acid (“ $\gamma$ -turn”) linkage (C-terminus of one PA to N-terminus of the other), which itself has shown a propensity for A-T/T-A pairings over G-C/C-G.<sup>[47]</sup>

Dervan also found that modification at the  $\gamma$ -turn could impact the binding affinity of polyamides for dsDNA. With two of the aliphatic positions on the GABA turn ( $\alpha$  and  $\beta$ ) amenable to substitution, detailed studies revealed that installation of a charged amino group at the alpha or beta position enhanced the binding affinity.<sup>[48]</sup>

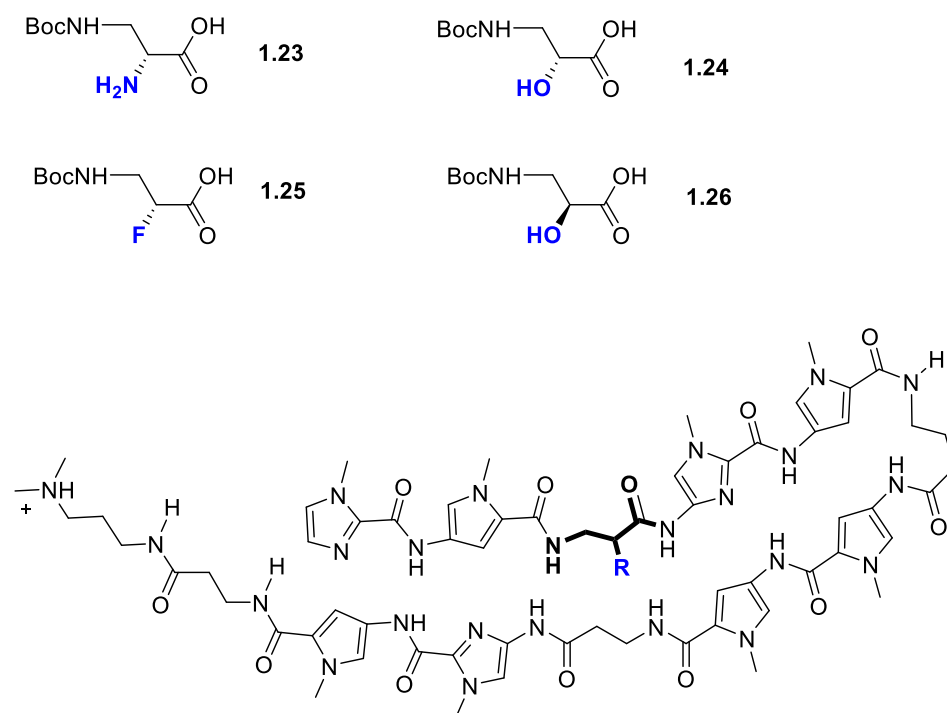


**Figure 1.10.4.** Modifications to the GABA ( $\gamma$ ) turn unit in hairpin polyamides.<sup>[48]</sup>

Further modification of the  $\gamma$ -turn was undertaken with the intent of improving cellular uptake. Introduction of various aryl functionality onto the  $\beta$ -amino position of the GABA turn found that placing a simple phenyl via an amide linkage promoted both cell uptake and nuclear concentration.<sup>[49]</sup> The analog of an 8-ring hairpin polyamide with the aryl functional group at the  $\beta$ -position required the use of MG132 protease inhibitor to tame cytotoxicity when studied in LNCaP cells.<sup>[49]</sup> Although the initial cell uptake studies with the aryl modifications recorded absolute values that were deemed likely inaccurate due to the incorporation of FITC, that study did also bring to light yet another major concern for polyamides: aggregation.<sup>[49]</sup> At higher concentrations, polyamide nuclear uptake dropped substantially, likely due to aggregate formation at concentrations  $\geq 1 \mu\text{M}$ .<sup>[50]</sup> It was found that use of 2-hydroxypropyl- $\beta$ -cyclodextrin aided with the solubility of polyamides (hairpin and cyclic scaffolds studied).<sup>[50]</sup> That said, the tendency of polyamides to aggregate under higher concentrations is a concern that needs to be resolved.

Increasing the number of paired rings that constitute a hairpin polyamide beyond 4 (8 rings total) changes the structure's curvature to a point where the polyamide ligand no longer fits well within the B-DNA minor groove, as indicated by a substantial drop in the affinity.<sup>[51,52]</sup> This brought about another key modification that was found to impact the structure of hairpin polyamides: the  $\beta$ -alanine unit. The added flexibility of this monomer relaxes the curvature of the polyamide, allowing for the linear design

(shown above in fig. 1.10.1) to bind with strong affinity.<sup>[51,53]</sup> Recognition of eleven base pairs was thus achieved with the insertion of a  $\beta$ -Ala residue.<sup>[51]</sup> There have also been modifications to this monomer within the context of the hairpin architecture, where various substitutions have been implemented in efforts to address the A-T/T-A degeneracy and where larger polyamides require the incorporation of  $\beta$ -Ala residues (shown in fig. 1.10.5).<sup>[54]</sup>

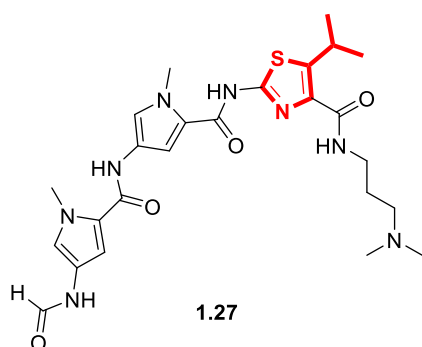


**Figure 1.10.5.** Modifications to the  $\alpha$ -position of  $\beta$ -Ala monomer unit and incorporation into a hairpin scaffold.<sup>[54]</sup>

The experiment found that the (S)-isoserine derivative (**1.26**) was 7 times more selective for T over A, as compared to the essentially non-selective  $\beta/\beta$  pairing.<sup>[54]</sup>

## 1.11 Thiazotropins

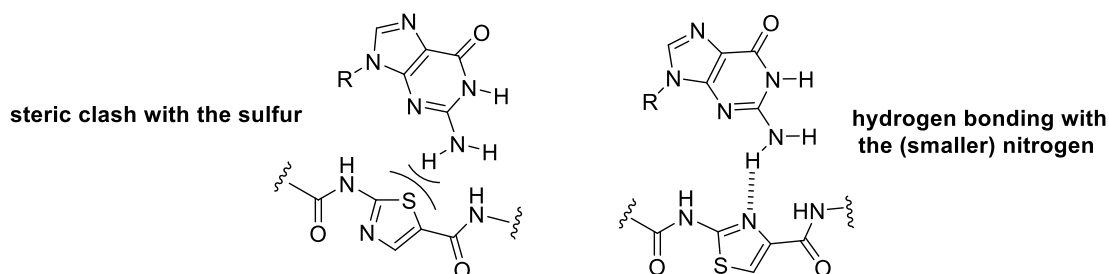
*Thiazotropins* were developed as an alternative to the standard lexitropin design with the incorporation of a thiazole unit (fig.1.11.1).<sup>[55]</sup> The thiazole unit can be good for pharmacokinetics due to the increased hydrophobicity of the sulfur atom potentially enhancing cell uptake.



**Figure 1.11.1.** Thiazotropsin A.

Based on the initial work of Lown to install a thiazole moiety in the polyamide scaffold, Nguyen placed the unit internally into the hairpin structure, but this resulted in poor binding affinity.<sup>[55,56]</sup> Lown found this was due to the improper orientation, and so reformatted the sulfur atom to the opposite face (away from the minor groove floor). Dervan's work to install the thiazole in between two  $\beta$ -Ala residues found this worked for the linear design.<sup>[45]</sup>

It was important that the sulfur atom, due to its bulkier size, was oriented away from the minor groove since the steric clash with even the smaller T nucleotide lowered the binding affinity.<sup>[57]</sup>

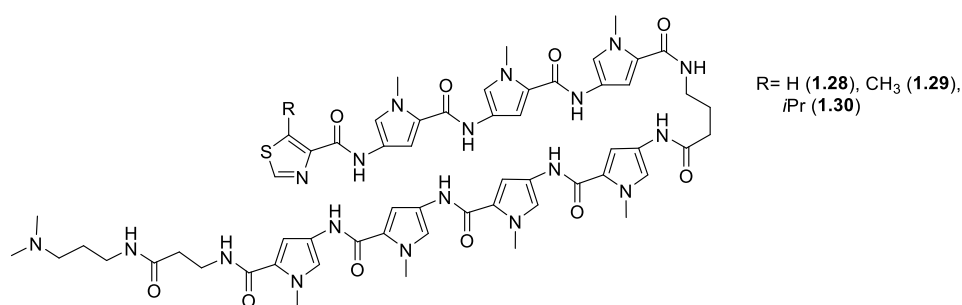


**Figure 1.11.2.** Steric clash between the sulfur atom with the exocyclic amine of a G-nucleotide necessitates synthetic re-orientation of the thiazole unit for enhanced binding affinity.

The thiazotropsin are an interesting subclass of the lexitropsins that have demonstrated a capability as MGBs. The thiazole unit confers added hydrophobicity for improved cell permeation, which also adds interactions with the DNA minor groove.<sup>[17,58,59]</sup> A consequence of the bulkier sulfur, they have been shown to bind in a 2:1 “staggered” fashion, with the heterocycle of one ligand placed over the amide bond of the other ligand when bound in the minor groove.<sup>[58,59]</sup>



Inspired by the thiazotropins synthesized and evaluated by the Suckling group, the Burley group explored the incorporation of an alkyl thiazole moiety into an 8-ring hairpin polyamide. This resulted in an enhancement of the dsDNA binding affinity along with significantly increased structural distortion in which a ‘pinching’ of the target sequence in the minor groove induced significant compression of the major groove.<sup>[2,57]</sup>

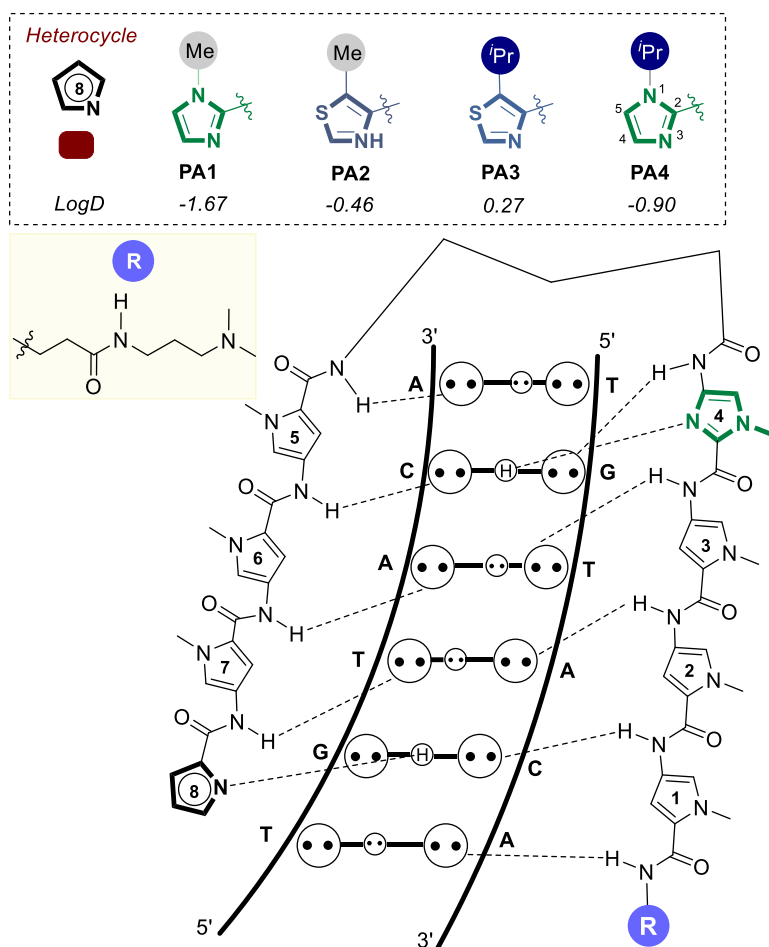


**Figure 1.11.3.** 8-ring hairpin polyamide incorporating a terminal thiazole unit with/without alkyl substituents facing away from the minor groove were synthesized and evaluated by the Burley group.<sup>[2]</sup>

Despite the lower selectivity of the thiazole units (G/T degeneracy), the 3D NMR structure indicated that there was more substantial major groove compression induced by the binding of the polyamide to the B-DNA minor groove than was observed with the N-terminal methylimidazole unit. Furthermore, the bulkier isopropyl group increased the compression further than the methyl substituent on the terminal thiazole. This trend suggests that there could be substantial modulation of the duplex distortion without the sulfur atom, just by bulking up the alkyl moiety on the polyamide facing away from the minor groove.<sup>[2]</sup>

## 1.12 Influence of steric bulk on the polyamide-dsDNA binding interaction

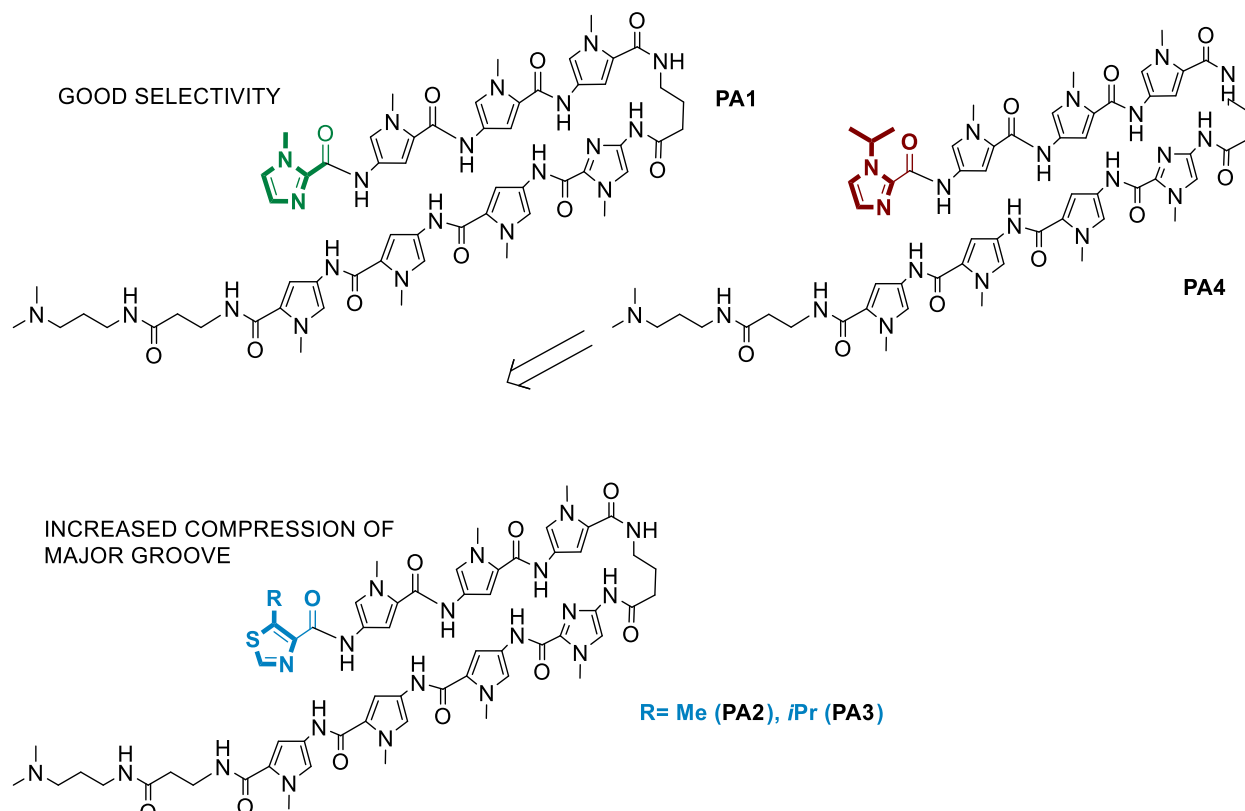
The work on the Nt-PA analogues hinted at the potential implications on the binding interaction by seemingly subtle changes in steric bulk of the non-minor groove facing side of the polyamide. The extra 2 methyl groups of the isopropyl substituent resulted in substantial major groove compression. Furthermore, the theoretically calculated distribution coefficient, logD, values for the hairpin PAs suggest that the isopropyl unit may confer significant hydrophobicity to the molecule, allowing for the modulation of pharmacokinetic properties.



**Figure 1.12.1.** Impact of changing the N-terminal substituent on the theoretically calculated logD values for the hairpin polyamides (PA1-4) to be evaluated and compared (higher value indicates increased hydrophobicity; logD calculated through chemicalize software from ChemAxon at chemicalize.com).

### 1.13 Hypothesis

From the understanding that size and fit are critical factors to forging a strong affinity relationship between the DNA minor groove and hairpin polyamides, and taking a sterics-focused approach, it is hypothesized that increasing the steric bulk of the typical N-terminal Me-Im unit of an 8-ring hairpin polyamide to an *i*Pr-Im unit would augment the structural distortion whilst retaining a high degree of binding affinity and selectivity. This modification would also increase the hydrophobicity of the polyamide itself, which could potentially improve cell uptake. In turn, we expect that this subtle, 2-carbon extension to the 8-ring hairpin scaffold will strike an appropriate balance between hydrophobicity, DNA-binding affinity and sequence selectivity between the thiazole PAs and the well-established Im-PAs.

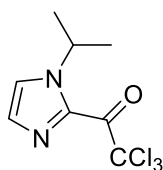


**Figure 1.13.1.** 8-ring hairpin polyamides for study. **PA4** is anticipated to exhibit binding behavior between **PA1** and **PA2/3**.

## 1.14 Project Aims and Objectives

The specific aims of this project are to:

- Synthesise the novel <sup>*i*</sup>Pr-Im building block that can be efficiently installed in the polyamide scaffold



1.31

**Figure 1.14.1.** <sup>*i*</sup>Pr-Im-COCCl<sub>3</sub> building block for incorporation into the N-terminal (ring 8) position of an 8-ring hairpin polyamide.

- Determine the impact to dsDNA binding affinity and selectivity of the <sup>*i*</sup>Pr-Im building block when incorporated at the N-terminal position of an 8-ring hairpin polyamide (**PA4**) targeting the ARE sequence

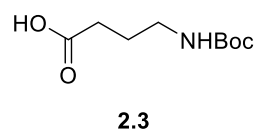
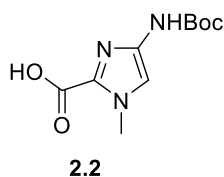
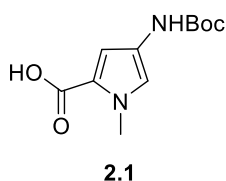
- Determine the kinetics governing the PA4-dsDNA binding interaction and comparison with PA1-dsDNA, PA2-dsDNA and PA3-dsDNA analogues
- Determine a 3D NMR structure of the PA4-dsDNA complex to provide insight into the solution-phase binding interaction upon added alkyl bulk from the isopropyl substituent in the N-terminal monomer unit

## **Chapter 2**

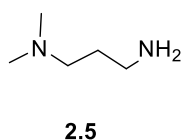
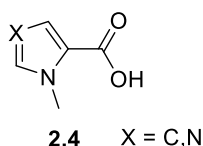
**Synthesis of an 8-ring hairpin polyamide incorporating a *N*-terminal *i*Pr-Im building block**

The synthesis of the standard 8-ring hairpin polyamides takes a modular approach in which the assembly of a variety of monomeric building blocks through amide bond linkages is conducted via well-established reactions. The typical building blocks that have successfully been used are N-methylimidazole (Im), N-methylpyrrole (Py), beta-alanine, gamma-amino butyric acid (GABA, or  $\gamma$ -turn), as shown in fig 2.1 below.

#### Internal building blocks



#### Terminal building blocks

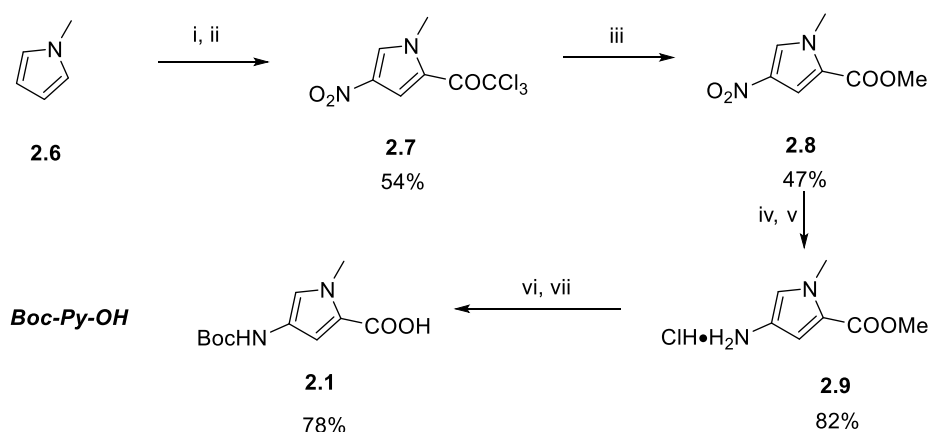


**Figure 2.1.** Standard building blocks of hairpin polyamide synthesis.

The assembly of polyamides from these monomeric building blocks is conducive to their programmability, requiring only modification to the coupling conditions or incorporation of a dimer for overcoming more problematic couplings.

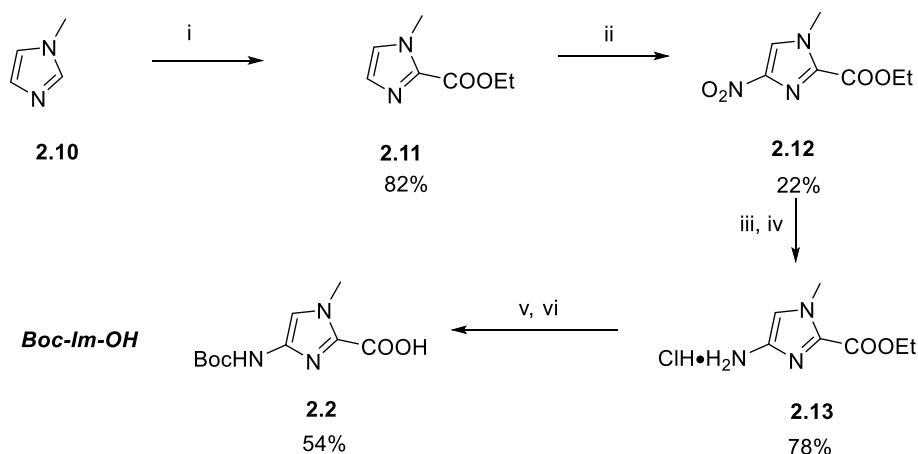
## 2.1 Synthesis of polyamide building blocks

Baird *et al.* optimized the synthesis of the Py and Im monomers from the commercially available starting materials **2.6** and **2.10**, respectively.<sup>[60]</sup>



**Scheme 2.1.** Synthesis of Boc-Py-COOH (**2.18**). Reagents and conditions: (i) trichloroacetyl chloride, rt, 3 h (ii) 90% nitric acid, Ac<sub>2</sub>O, -40 °C, 4 h (iii) NaOMe, MeOH, rt, 2 h (iv) H<sub>2(g)</sub> (1 atm), 10% Pd/C, rt, 48 h (v) HCl (vi) 10% Na<sub>2</sub>CO<sub>3(aq)</sub>, Boc<sub>2</sub>O, rt, 3 h (vii) 2 M NaOH<sub>(aq)</sub>, 60 °C, 6 h.<sup>[60]</sup>

The Im building block was synthesized in a very similar route to the Py monomer, but using slightly different reagents, namely ethyl chloroformate to form an ethyl ester directly.

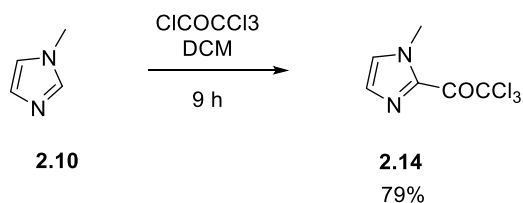


**Scheme 2.2.** Synthesis of Boc-Im-COOH (**2.23**). Reagents and conditions: (i) ethyl chloroformate, TEA, MeCN, -20 °C, 36 h (ii) conc. H<sub>2</sub>SO<sub>4(aq)</sub>, 90% nitric acid, 0 °C, 50 min (iii) H<sub>2(g)</sub> (1 atm), 10% Pd/C, rt, 48 h (iv) HCl (v) DIPEA, Boc<sub>2</sub>O, 60 °C, DMF, 18 h (vi) 1 M NaOH<sub>(aq)</sub>, 60 °C, 3 h.<sup>[60]</sup>

The synthesis of these monomer units was completed on a substantially larger scale (50 g), and avoiding column chromatography. The nitration was the most problematic, providing a mere 22% yield of the desired isomer in synthesis of Boc-Im-OH.<sup>[60]</sup>

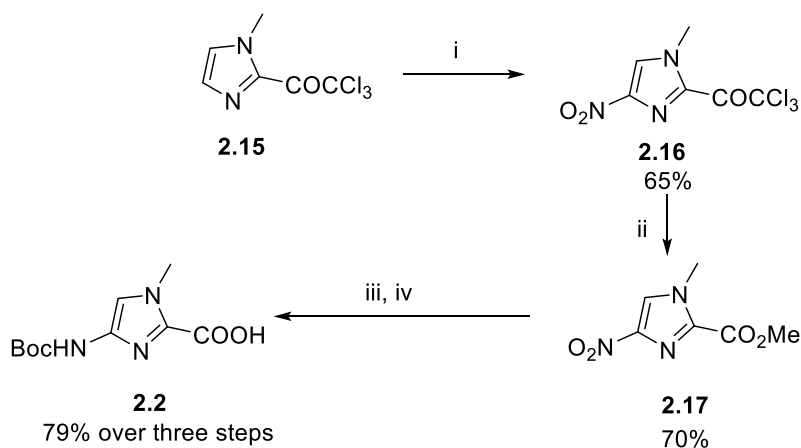
#### Synthesis of *N*-isopropylimidazole

Taking inspiration from the synthetic route used to synthesize the Py and Im monomers by Jaramillo *et al.*, one potential expedient synthesis of the *i*Pr-Im-COCCl<sub>3</sub> building block was in a single step from the commercially available 1-isopropylimidazole, treating with trichloroacetylchloride.<sup>[61]</sup>



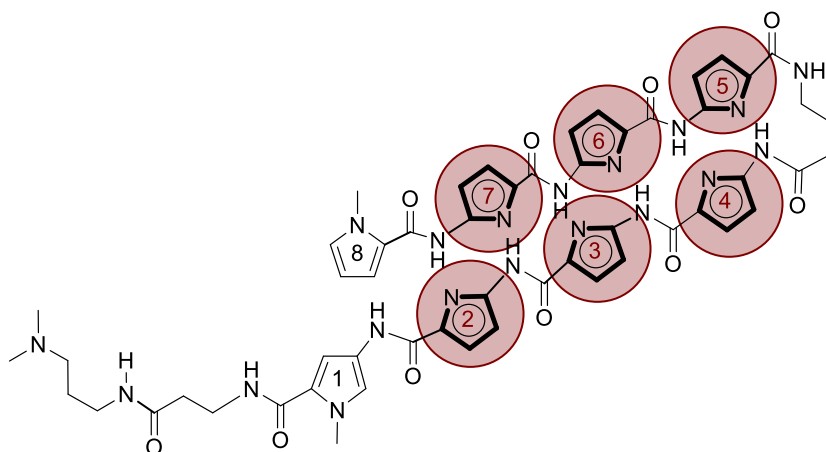
**Scheme 2.3.** Jaramillo *et al.* synthesis of **2.14**.

In that reaction outlined in scheme 2.3, the authors found that the product was stable enough to isolate via column chromatography of the TEA-quenched, filtered, and concentrated reaction mixture. The *i*Pr-Im-COCCl<sub>3</sub> unit could then be taken forward to generate the Boc-protected amino acid for placement internally in the polyamide at ring positions 1 through 7 (fig. 2.2), following the synthetic route outlined in scheme 2.4.



**Scheme 2.4.** Jaramillo *et al.* synthetic route to internal BocNH-Im-COOH building block **2.2**. Reagents and conditions: (i) HNO<sub>3</sub>, H<sub>2</sub>SO<sub>4(aq)</sub>, Ac<sub>2</sub>O, 0 °C to rt, 24 h (ii) NaOMe, MeOH, 3 h (iii) Boc<sub>2</sub>O, 10% Pd/C, H<sub>2(g)</sub>, MeOH, 28 h (iv) KO<sup>t</sup>Bu, H<sub>2</sub>O, THF, rt, 1 h.<sup>[61]</sup>



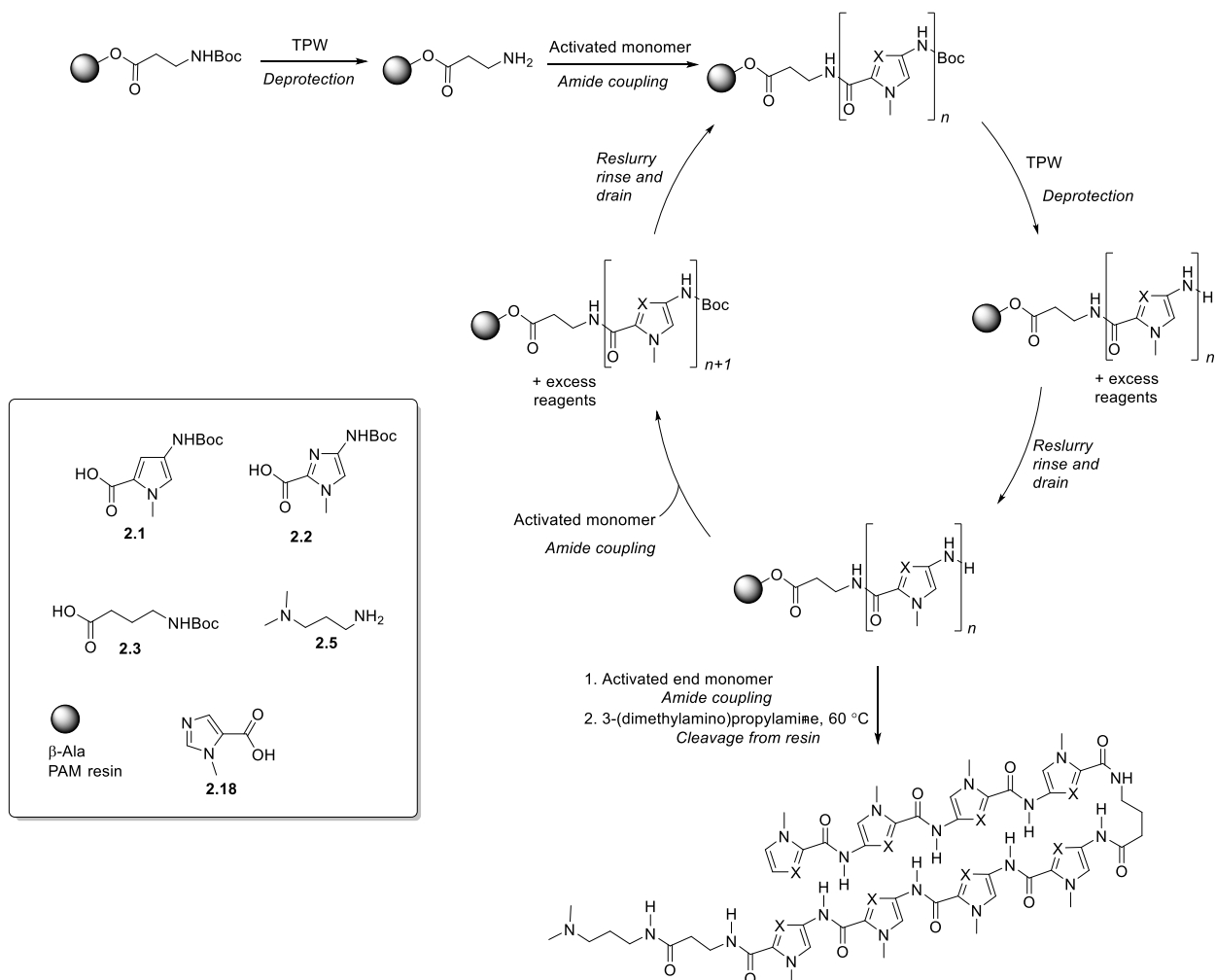


**Figure 2.2.** Potential internal positions for **Pr-Im** building block.

## 2.2 Solid Phase Synthesis of polyamides

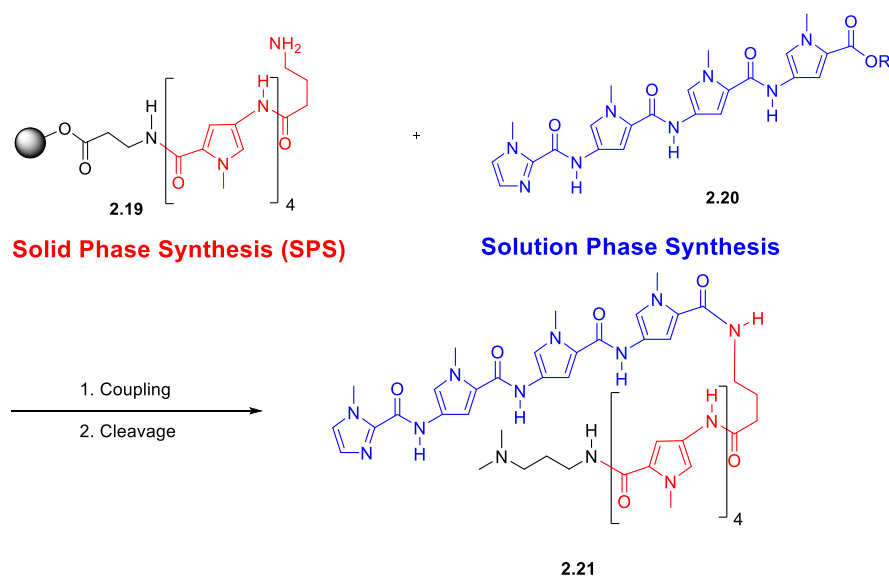
A tremendous breakthrough in the synthesis of polyamides was borrowed from peptide synthesis when Dervan and Baird employed the solid phase strategy.<sup>[60]</sup> This methodology enabled access to larger polyamides of added complexity, mitigating the challenges associated with solution phase syntheses (numerous isolations, purifications, etc.). Furthermore, this increased the efficiency of the process, reducing synthesis time from months to days.<sup>[38,60]</sup> Baird *et al.* also developed both the Boc-based  $\beta$ -Ala-PAM and, later, the Fmoc-based  $\beta$ -Ala-Wang resin chemistries.<sup>[60,62]</sup>

In the protocol using the  $\beta$ -Ala-PAM resin, the Boc-protecting group was cleaved immediately prior to the addition of the pre-activated Py/Im monomer or dimer acid.<sup>[60]</sup> The remaining core of the polyamide scaffold was built from sequential coupling reactions to install Py/Im/ $\beta$ -Alanine/GABA units. Once the terminal monomer unit was in place, the polyamide could be cleaved from the resin support with an aliphatic amine at elevated temperature (60-70 °C). 3-dimethylaminopropylamine was typically used for cleavage, although bis-(dimethylaminopropyl)amine (shorthand **DD**) has also been used where a further amide linkage for additional modification is required.



**Scheme 2.5.** Solid phase synthesis cycle of hairpin polyamides.

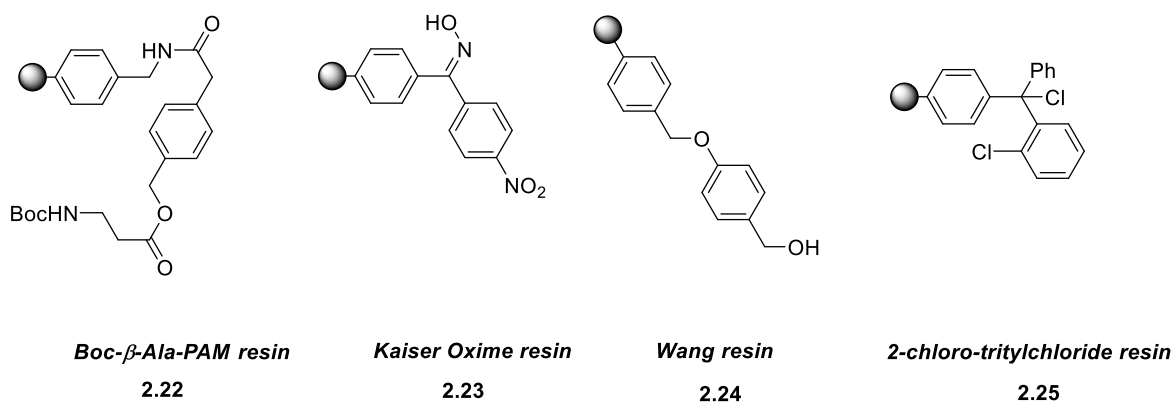
The Burley group found that the poor coupling efficiencies observed after the  $\gamma$ -turn unit installation warranted a rethinking of the synthesis methods to enhance the robustness. This led to the development of a convergent synthesis strategy, in which the second half of the rings required are synthesized in solution phase as a tetramer, and then placed in a single coupling step to the resin-adhered polyamide via coupling with the deprotected amine of the GABA ( $\gamma$ -turn) unit. The solution and solid phase syntheses of polyamides in tandem improved the overall yield and crude purity.<sup>[63]</sup>



**Scheme 2.6.** Convergent synthesis of an 8-ring hairpin PA by the Burley group.<sup>[63]</sup>

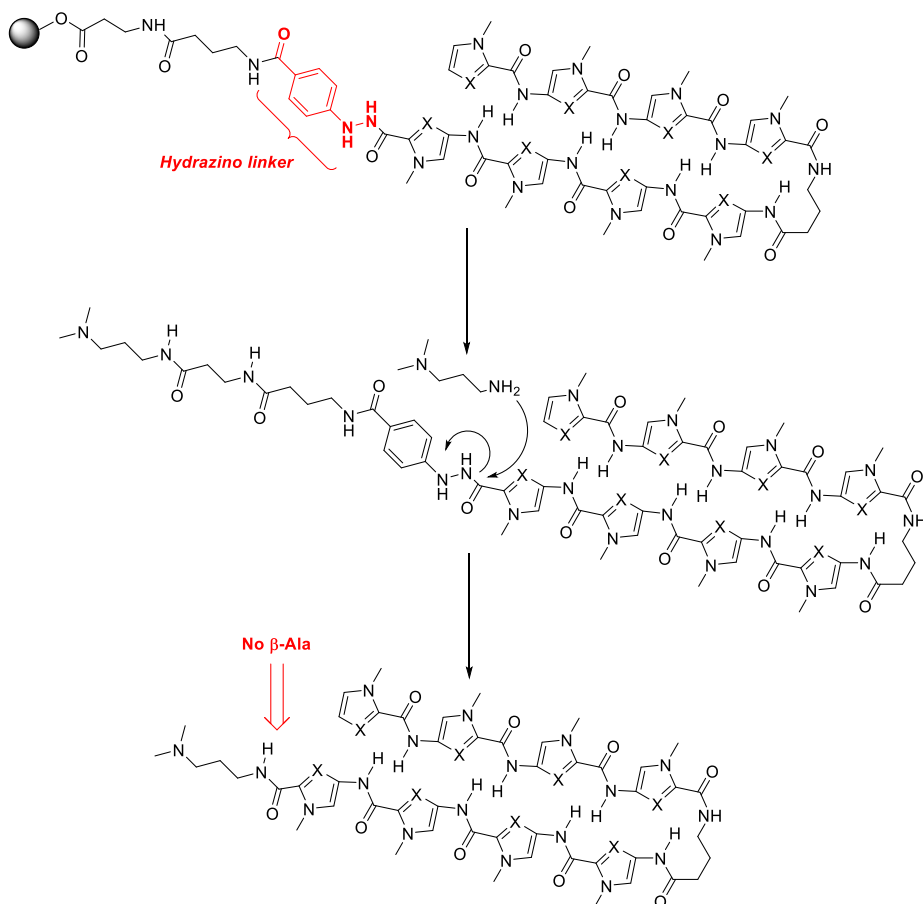
### 2.2.1 Solid phase support

The choice of which solid support/resin to employ in the synthesis of polyamides has important implications for the reagents and conditions that will be used. Peter Dervan and Eldon Baird developed the solid phase synthetic technology using Boc- $\beta$ -Ala-PAM due to its resistance to acid and adequate thermal stability.<sup>[60,62]</sup> The incorporation of a  $\beta$ -alanine residue at the C-terminus of the polyamide is a consequence of using this resin. The solid phase protocol involves repeated cycles of deprotection by acid treatment (TFA), followed by coupling with either HATU or DCC/HOAt in the presence of Hunigs base (DIPEA). Using the Kaiser oxime resin allows for excision of the T/A-selective  $\beta$ -alanine present at the C-terminus. Using a base-mediated deprotection strategy, the Fmoc- $\beta$ -Ala-Wang resin has also been used for machine-assisted syntheses, affording equivalent or better yields and purities to Boc-based chemistry.<sup>[62]</sup>



**Figure 2.3.** Types of solid phase resins used for PA synthesis.

In cases of the Boc-β-Ala-PAM resin where excision of the β-alanine unit was desired, the Burley group found that a newly developed hydrazino linker allowed for the β-Ala excision during the cleavage step.<sup>[63]</sup>



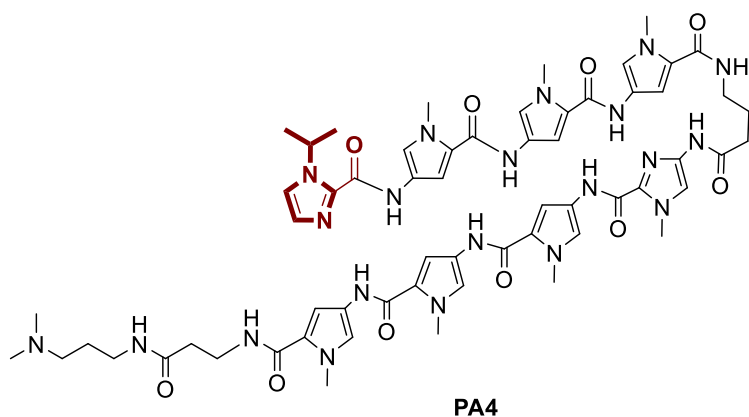
**Scheme 2.7.** Proposed mechanism for excision of β-Ala subunit from polyamide during cleavage.

## 2.3 Aim and objectives of this chapter

The versatility of the solid phase synthesis of polyamides provides access to a variety of Py/Im sequences, dictated by the target DNA sequence. The programmability of this modular approach was ideal for synthesizing a polyamide that targets the ARE sequence 5'-WGWWCW-3' (W=A/T), as demonstrated by Dervan.<sup>[1]</sup>

This chapter will describe the synthesis of the building blocks used to construct hairpin polyamide **PA4**.

Also detailed herein is the solid phase synthetic procedure for synthesizing **PA4**.

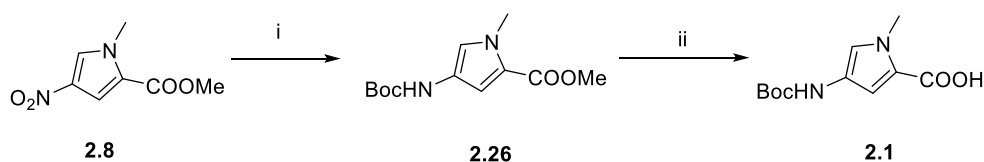


**Figure 2.4.** 8-ring hairpin polyamide **PA4** targeted for synthesis.

## 2.4 Results and discussion

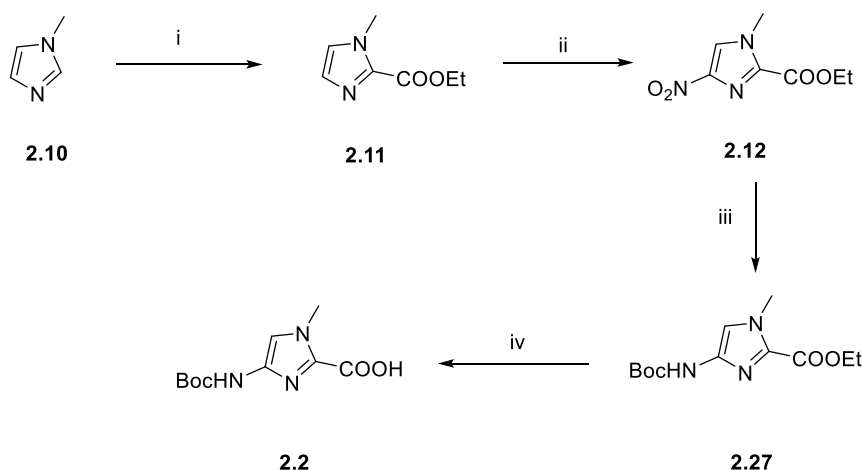
### 2.4.1 Synthesis of Py and Im monomers

The standard *N*-methylpyrrole (Py) and *N*-methylimidazole (Im) building blocks used in standard hairpin polyamides were synthesized by literature methods in moderate to good yields.<sup>[2,60,61]</sup> The *N*-methylpyrrole building block synthesis starts from the commercially available methyl-4-nitro-1-methylpyrrole-2-carboxylate (**2.8**) to which a one-pot reduction and Boc-protection of the free amine was applied. The isolated intermediate **2.26** was then saponified to give the corresponding carboxylic acid in moderate yield (60-70%).<sup>[60,61]</sup>



**Scheme 2.8.** Synthetic route to Py monomer **2.1**. Reagents and conditions: (i)  $\text{Boc}_2\text{O}$  (1.2 equiv.), 10% Pd/C,  $\text{H}_{2(\text{g})}$  atm. 55 psi, 20 °C, 3 h; (ii) NaOH 1 M, 65 °C, 3 h.

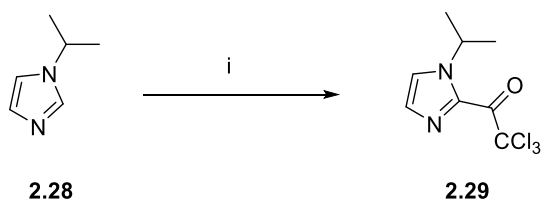
The synthesis of the *N*-methylimidazole carboxylic acid building block **2.2** was carried out also following a literature reported protocol, starting from *N*-methylimidazole **2.10**. Installation of the ethyl carboxylate onto the C2 position using ethyl chloroformate in the presence of excess triethylamine formed **2.11** in 49% yield. The problematic nitration of the C4 position of **2.11** was carried out, and found not to be chemoselective (twice as much C5 nitration observed). The advanced intermediate **2.12** was then taken through a one-pot reduction and Boc-protection, followed by a saponification of the ethyl ester to yield the desired carboxylic acid monomer **2.2**, designated Im.<sup>[2,60,61]</sup>



**Scheme 2.9.** Synthetic route to Py monomer **2.1**. Reagents and conditions: (a)  $\text{ClCOOEt}$  (1.3 equiv.),  $\text{Et}_3\text{N}$ , 0 °C, 12 h; (b)  $\text{HNO}_3$  99% (3.7 equiv.), TFAA, 0 °C, 12 h. (c)  $\text{Boc}_2\text{O}$  (1.2 equiv.), 10% Pd/C,  $\text{H}_{2(\text{g})}$  atm. 55 psi, 20 °C, 3 h; (d) LiOH 1 M, 60 °C, 4 h.

## 2.4.2 Synthesis of the *N*-isopropylimidazole monomer

Starting from the commercially available *N*-isopropylimidazole (**2.28**), the use of trichloroacetylchloride chemoselectively acylated the C-2 position of the heterocycle. This provided the terminal monomer building block **2.29**, designated <sup>i</sup>Pr-Im-COCCl<sub>3</sub>.<sup>[61]</sup>

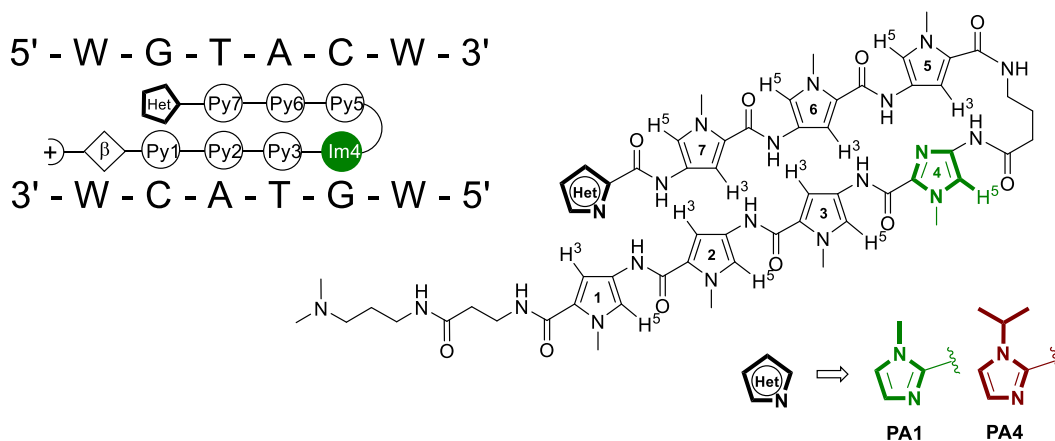


**Scheme 2.10.** Synthesis of *i*Pr-Im-COCCl<sub>3</sub> monomer. Conditions: (i) trichloroacetylchloride (1.05 equiv), triethylamine (1 equiv), 20 °C, 17 h.

**2.29** was flash purified through silica gel chromatography, yielding a white crystalline solid in 47-50% yield.

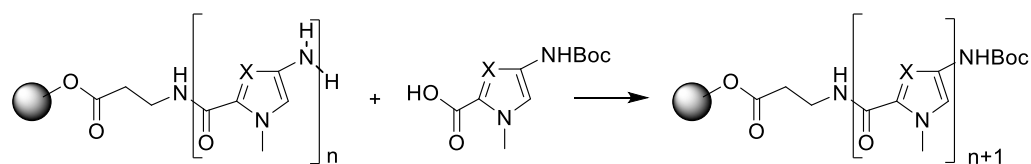
### 2.4.3 Solid phase synthesis of 8-ring hairpin polyamides

Targeting the ARE sequence 5'-WGWCW-3', the intended polyamide architecture for synthesis is shown in fig. 2.5. The N-terminal position is most easily amenable to alteration, and so this is where the novel *i*Pr-Im monomer unit would be installed (**PA4**), to compare head to head with Me-Im (**PA1**).



**Figure 2.5.** Schematic of 8-ring hairpin polyamide to be synthesized; the modification is at the N-terminal position.

Owing to the specific requirements of the intended coupling at each stage of the synthesis, the times optimized in the Burley laboratory were followed as a starting point, as outlined here in table 2.1.

**Table 2.1.** Coupling conditions for solid phase synthesis of polyamides with  $\beta$ -Ala-PAM resin.

Monomer unit to couple	Reactive amine on solid support	Coupling reagents	Reaction time h
Boc-Py-OH	NH <sub>2</sub> -Py	HATU	2
Boc- $\beta$ -Ala-OH	NH <sub>2</sub> -Py	HATU	2
Boc- $\gamma$ -turn-OH	NH <sub>2</sub> -Py	HATU	2
Boc-Im-OH	NH <sub>2</sub> -Py	DCC/HOAt	6
Boc-Im-OH	NH <sub>2</sub> -Im	DCC/HOAt	6
Boc- $\beta$ -Ala-OH	NH <sub>2</sub> -Im	DCC/HOAt	16
Boc- $\gamma$ -Turn	NH <sub>2</sub> -Im	DCC/HOAt	16
Boc-Py-OH	NH <sub>2</sub> -Im	BTC	3
Im-COOH	NH <sub>2</sub> -Py	DCC/HOAt	6
Boc-Py-OH	NH <sub>2</sub> - $\gamma$	HATU	2

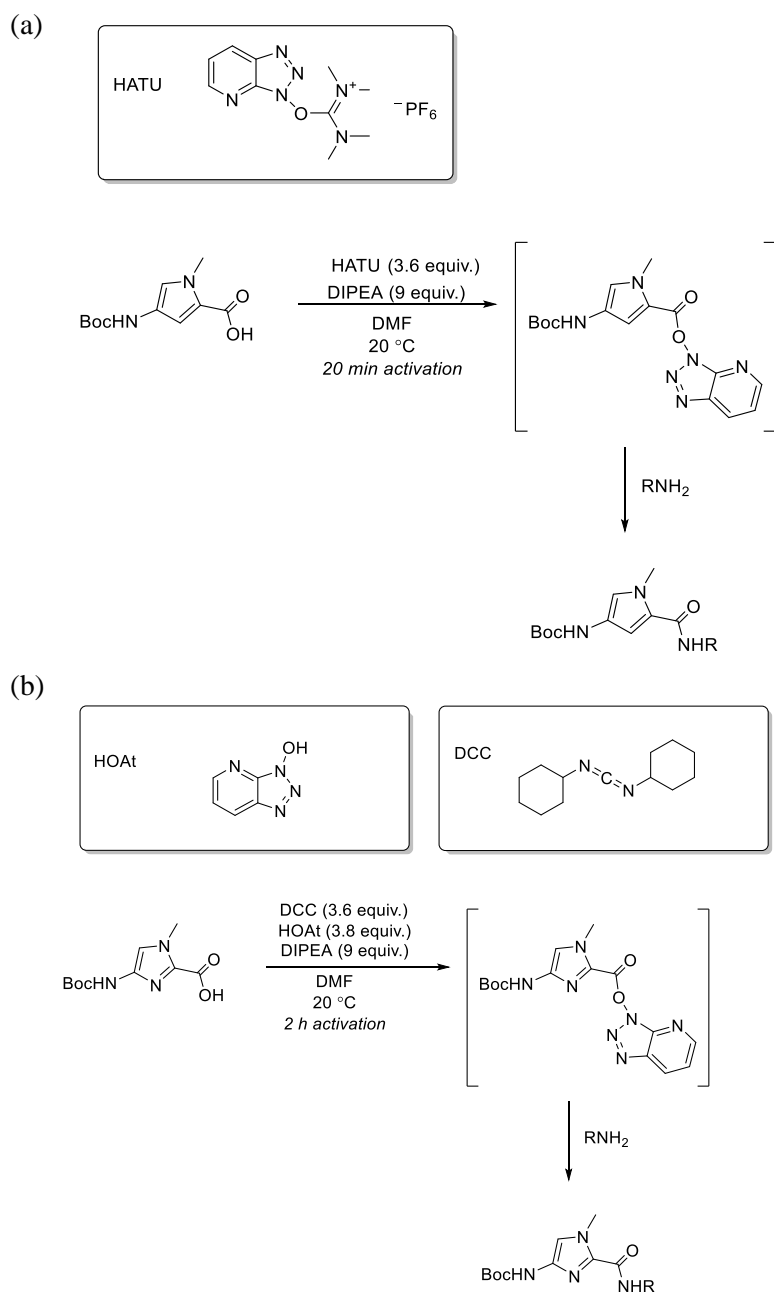
Starting with the Boc- $\beta$ -Ala-PAM resin, the cycles of deprotection were accomplished with the use of a 92.5% TFA/5% phenol/2.5% water mixture (TPW). This provided ample time to ensure the complete deprotection of the Boc protecting group. After each deprotection step during the synthesis, the resin was washed with a sequence of 4 x 1 min x 2 mL DMF to completely remove the acidic residues left over which would hinder the subsequent coupling (premature deprotection of the Boc of the incoming monomer).

The subsequent coupling of three consecutive Py units onto the solid support was achieved with HATU/DIPEA for activation of the BocNH-Py-COOH building block. This activation was typically given 20-30 minutes, usually accompanied by a change in the physical appearance of the mixture (darkens).

After each coupling, the resin was washed with 4 x 1 min x 2 mL DMF, 4 x 1 min x 2 mL DCM, and followed by another 4 x 1 min x 2 mL DMF prior to the next deprotection step.



Previously reported failed coupling reactions with the use of HATU for activating the acid of the Im monomer necessitated the use of DCC/HOAt. After a 2 h activation, filtration of the resulting slurry to remove the urea by-product DCU and addition of DIPEA to the filtrate containing the activated acid was performed prior to coupling with the PA on resin. This coupling, although reported to be complete in 6 h, was typically given 16 h (overnight) to ensure the coupling went to completion.



R = Polyamide on resin after deprotection with TFA/Phenol/Water

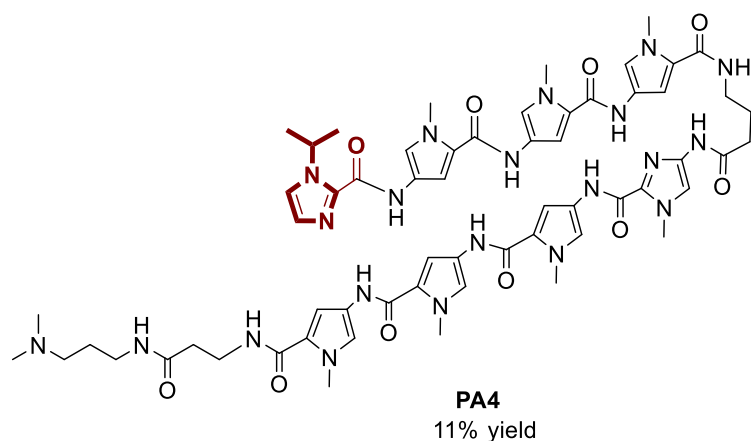
**Scheme 2.11.** Activation intermediates for amide bond synthesis using (a) HATU and (b) DCC/HOAt.

The stability of the Boc-protected amine of the Im monomer required an extended deprotection time, with a wash sequence of 2 mL x 1 min, 2 mL x **10 min**, 2 mL x **10 min**.

The coupling of the gamma-amino butyric acid (GABA) turn unit was achieved with the same DCC/HOAt method used for the Im unit, and again allowed to couple overnight (16 h) to ensure completion.

The next three pyrrole units were installed using the HATU/DIPEA conditions from earlier in the synthesis. This second half of the polyamide was more challenging, with an increase in impurities (typically truncated products) as each unit was installed.

After the N-terminal Im unit was coupled with the standard DCC/HOAt protocol, the resin was rinsed with DMF, DCM and methanol, and the polyamide cleaved with 3-dimethylaminopropylamine at 60 – 70 °C, thereby completing the synthesis of the standard hairpin **PA1** as a control. The N-terminal *i*Pr-Im analogue was synthesized likewise, using the trichloroacetylated monomer **2.29**, which proved stable enough to handle and store for repeat use. After cleavage, the crude polyamide cleavage reaction mixture was purified via semi-preparative HPLC.



**Figure 2.6.** 8-ring hairpin polyamide synthesized on solid phase resin, cleaved, and purified in this work.

## 2.5 Summary

This chapter detailed the synthesis of the 8-ring hairpin polyamide PA4 studied in the DNA binding interactions outlined in chapters 3-4. The development of the solid phase synthetic methodology by

Dervan and Baird was crucial to the timely synthesis of the PA series. The coupling conditions followed were based on previous optimizations within the Burley lab, and were found reasonably successful synthesizing **PA4** in 11% yield and with 95% AN HPLC purity after purification. The synthesis of the <sup>i</sup>Pr-Im monomer **2.29** itself followed a literature reported protocol for the similar Im monomer using trichloroacetyl chloride.<sup>[61]</sup> The coupling of **2.29** onto the penultimate Py unit of the polyamide on resin was achieved in 16 h at room temperature with 9 equiv DIPEA.

## 2.6 Experimental

### 2.6.1 General protocols

Reagents and solvents from commercial sources were used without further purification. Solvents were of peptide/HPLC-grade. Solvents were not dried, but used as is from supplier, including the use of anhydrous DMF. Solid phase experiments were conducted in air, and the building block syntheses were conducted in inert atmosphere.

#### *NMR Spectroscopy*

NMR data was collected on one of two instruments, a Bruker 400 UltraShield™ B-ACS spectrometer and a Bruker Avance 500 spectrometer. Chemical shifts were referenced to TMS and are reported in parts-per-million (ppm). Coupling constants are stated in Hertz (Hz). All NMR data was processed using the Topspin NMR software package. Proton and carbon NMR peaks were assigned based on a combination of Correlation Spectroscopy (COSY), Heteronuclear Single Quantum Coherence (HSQC), Heteronuclear Multiple-Bond Correlation Spectroscopy (HMBC), and Nuclear Overhauser Effect Spectroscopy (NOESY).

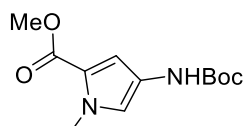
#### *Liquid Chromatography-Mass Spectrometry (LC-MS) / High Performance Liquid Chromatography (HPLC)*

LC-MS data was collected with Agilent 1200 series instrument with a quadrupole mass detector, with electrospray ionization (ESI), or run by Swansea University's HRMS facility. Semi-preparative HPLC

was performed for the isolation of polyamides using a Dionex UltiMate 3000 series system (see Appendix for method and column parameters).

## 2.6.2 Synthetic procedures

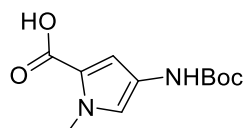
### *Methyl 4-((tert-butoxycarbonyl)amino)-1-methyl-1H-pyrrole-2-carboxylate (2.26)*



A solution of methyl-4-nitro-1-methylpyrrole-2-carboxylate, (**2.8**, 2.5 g, 13.6 mmol), di-tert-butyl dicarbonate (3.6 g, 16.5 mol), and a catalytic amount of 10 wt% Pd/C in 50 mL EtOAc/MeOH (equal parts by volume) was magnetically stirred for 24 h at ambient (19-25 °C) under H<sub>2(g)</sub> atmosphere. The reaction mixture was then concentrated under vacuum and recrystallized from a DCM/petroleum ether mixture to give the methyl carboxylate title compound as an off-white solid (2.22 g, 64% yield).

<sup>1</sup>H NMR (CDCl<sub>3</sub>, ppm, 400 MHz) δ: 1.5 (s, 9H, *H*-Boc), 3.8 (s, 3H, O-CH<sub>3</sub>), 3.9 (s, 3H, N-CH<sub>3</sub>), 6.3 (s, 1H, *Py-H*), 6.65 (s, 1H, Boc-NH), 7.08 (s, 1H, *Py-H*)

<sup>13</sup>C NMR (CDCl<sub>3</sub>, ppm, 126 MHz) δ: 28.4, 36.7, 51.0, 80.2, 107.8, 119.8, 122.1, 153.2, 161.5



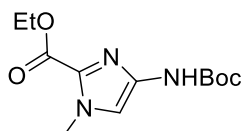
### *4-((tert-butoxycarbonyl)amino)-1-methyl-1H-pyrrole-2-carboxylic acid (2.1)*

To a solution of **2.26** (2.22 g, 8.7 mmol) in 8 mL methanol was charged a 1 M aqueous NaOH solution (8.4 mL). The resulting mixture was heated to 70 °C for 3.5 h. The reaction mixture was then concentrated under reduced pressure and the pale orange/yellow residue retaken up in water. The pH was then adjusted to 3 using 10% HCl<sub>(aq)</sub> solution, upon which a precipitate formed. The product was then extracted with 3 x 30 mL ethyl acetate and the organic phases combined, dried over anhydrous Na<sub>2</sub>SO<sub>4</sub> and distilled to dryness under reduced pressure. The resulting residue was taken into a minimal amount of petroleum ether and sonicated to induce crystallization. The petroleum ether was at that point

removed under vacuum distillation to give the carboxylic acid as a yellow-tinted solid (1.27 g, 61% yield).

$^1\text{H}$  NMR (DMSO- $d_6$ , ppm, 400 MHz)  $\delta$ : 7.20 (d, 1H, Py-*H*), 6.73 (s, 1H, Boc-NH), 6.25 (s, 1H, Py-*H*), 3.90 (s, 3H, N- $\text{CH}_3$ ), 1.50 (s, 9H, *H*-Boc)

$^{13}\text{C}$  NMR (DMSO- $d_6$ , ppm, 126 MHz)  $\delta$ : 165.6, 153.2, 122.5, 121.2, 109.3, 80.3, 36.9

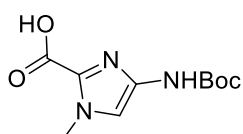


*Ethyl 4-((tert-butoxycarbonyl)amino)-1-methyl-1H-imidazole-2-carboxylate (2.27)*

To a solution of **2.12** (2.80 g, 14.0 mmol) and (Boc) $_2$ O (3.70 g, 17.0 mmol) in MeOH (60 mL) was added a catalytic amount of 10 wt% Pd/C. The mixture was stirred at room temperature under a hydrogen atmosphere for 17 h. The mixture was filtered through celite and the filtrate distilled at reduced pressure to yield the crude product as an orange solid. **2.27** was isolated after recrystallisation from DCM/hexane as a pale yellow solid (3.02 g, 80 %).

$^1\text{H}$  NMR (CDCl $_3$ , ppm, 500 MHz)  $\delta$ : 1.33 (t, 3H,  $J = 7$  Hz, O- $\text{CH}_2\text{CH}_3$ ), 1.43 (s, 9H, *H*-Boc), 3.90 (s, 3H, Im- $\text{CH}_3$ ), 4.33 (q, 2H,  $J = 7$  Hz, O- $\text{CH}_2\text{CH}_3$ ), 6.85 (s, 1H, CONH), 7.15 (s, 1H, *Im-H*).

$^{13}\text{C}$  NMR (DMSO- $d_6$ , ppm, 126 MHz)  $\delta$ : 14.4, 28.2, 35.9, 61.4, 80.8, 113.2, 131.5, 137.8, 152.6, 158.8



*4-((tert-butoxycarbonyl)amino)-1-methyl-1H-imidazole-2-carboxylic acid (2.2)*

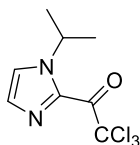
The ethyl ester **2.27** (3.0 g, 11.2 mmol) was dissolved in a solution of 50% MeOH / 50% 1 M LiOH $_{(aq)}$  (60 mL, 60 mmol). The solution was stirred for 2 h at ambient temperature (19-25 °C), monitoring the reaction progress with TLC analysis. The methyl alcohol was removed under reduced pressure and the aqueous solution adjusted to pH 3 with 1 M HCl $_{(aq)}$ . A white precipitate formed at this point, which was washed by pH 3 water, centrifuging, and removal of the supernatant (repeated 3 times). The purified

product **2.2** was then lyophilized, leaving an off-white solid (1.40 g, 54 %).

$^1\text{H}$  NMR (DMSO- $d_6$ , ppm, 500 MHz)  $\delta$ : 1.45 (s, 9H, *H*-Boc), 3.88 (s, 3H, Im- $\text{CH}_3$ ), 7.06 (s, 1H, *Im-H*), 9.35 (s, 1H, CONH)

$^{13}\text{C}$  NMR (DMSO- $d_6$ , ppm, 126 MHz)  $\delta$ : 28.4, 35.9, 80.4, 113.1, 133.4, 136.9, 153.0, 160.2

*2-trichloroacetyl-N-isopropyl imidazole* (**Pr-Im-COCCl<sub>3</sub>**, **2.29**)



To a solution of trichloroacetyl chloride (5.5 g, 30.2 mmol) in 8 mL methylene chloride was charged a solution of *N*-isopropylimidazole (**2.28**, 3.3 g, 30.0 mmol) in 15 mL methylene chloride over a period of 2.5 h via syringe pump at ambient (19-25 °C). The resulting solution was stirred with a magnetic stir bar at ambient for 17 h, at which point the reaction mixture was cooled to 5 °C and triethylamine was added, forming a precipitate (TEA•HCl). The slurry was allowed to warm to ambient and stirred for a further 2 h prior to filtration through a 55 mm diameter Buchner funnel lined with filter paper. The filtrate was distilled to dryness under reduced pressure and the oily concentrate flash-purified through normal phase silica gel chromatography (5:1 DCM/petroleum ether,  $R_f = 0.57$ ) to give **2.29** as an off-white solid (3.60 g, 47% yield). The monomer was stored in a freezer (-15 to -20 °C) until required for coupling.

Melting point = 79-80 °C

$^1\text{H}$  NMR (CDCl<sub>3</sub> 500 MHz)  $\delta$ : 1.45 (d, 6H,  $J = 6.7$  Hz), 5.34 (sep, 1H,  $J = 6.7$  Hz), 7.31 (s, 2H)

$^{13}\text{C}$  NMR (CDCl<sub>3</sub> 126 MHz)  $\delta$ : 172.3, 135.4, 130.9, 123.0, 95.3, 50.4, 23.6

IR (neat)  $\nu_{\text{max}}$ : 1690  $\text{cm}^{-1}$ , 1385  $\text{cm}^{-1}$ , 1364  $\text{cm}^{-1}$

ESI (+ve mode) HRMS calculated for  $[\text{M}+\text{H}]^+$  254.9857; found 254.9853.

### 2.6.3 Solid Phase Synthesis

#### *General Coupling Protocol*

Polyamide **PA4** was prepared via solid phase support utilizing the Boc-protected  $\beta$ -Ala-PAM resin available from Sigma Aldrich (~ 0.5 mmol/g).

The sequence for a typical coupling started with swelling of the resin for a minimum 1 h in  $\text{CH}_2\text{Cl}_2$ . The Boc-group was deprotected using a 92.5% TFA / 5% Phenol / 2.5% water mixture (1 x 1 min, 2 x 3 min). A solution of the desired building block monomer (Py, Im, iPrIm, GABA,  $\beta$ -Ala; 4 equiv) was activated with either HATU (3.6 equiv) and DIPEA (9 equiv) for 20 minutes in DMF (1-2 mL) or, in the case of Im or GABA building blocks, with HOAt (3.8 equiv) and DCC (3.6 equiv) in DMF for 2 h. In the latter case, a filtration to remove the DCU byproduct is warranted due to its gel-like nature. The solution of activated monomer was then transferred to the resin for a specified agitation period at ambient temperature (19-25 °C).

Coupling of the novel **2.29** onto the terminal position of the standard 8-ring hairpin polyamide was completed by 16 h agitation at ambient temperature (19-25 °C) in the presence of a large excess (9 equiv) DIPEA.

#### *General cleavage protocol*

Cleavage of the polyamides from the resin support was achieved by heating the resin in neat 3-dimethylaminopropylamine at 60-70 °C for 15-20 h. The slurry was at that point filtered to remove the resin, rinsed with 1 mL methanol, and the resulting filtrate purified through semi-preparative HPLC. The collected fractions were then distilled at reduced pressure to remove acetonitrile, retaken in Milli-Q water, and freeze-dried to give the off-white to yellow to brown to red colored polyamide.

#### *Polyamide synthetic protocols*

##### *Synthesis of PA4*

**PA4** was prepared starting from the Boc- $\beta$ -Ala-PAM resin (~0.5 mmol/g loading, supplier specified). After 1 h swelling of the resin in methylene chloride, deprotection of the Boc group was achieved with

a TPW mixture (92.5% TFA, 5% phenol, 2.5% water) – 1 x 1 min + 2 x 3 min for Py and turn units, 1 x 1min + 2 x 10 min for Im unit.

#### *Py/Py and GABA/Py coupling*

*Pre-activation of monomer unit: 4.0 mol equiv of Py with 3.6 equiv HATU in 1.5 mL anhydrous DMF. 9 equiv DIPEA added and the mixture agitated for 20 minutes at ambient (19-25 °C).*

After deprotection, the resin was washed with DMF (4 x 1 mL x 1 min), rinsed with 1.5 mL anhydrous DMF. The pre-activated mixture of Py (4 mol equiv) in 1.5 mL anhydrous DMF with 3.6 equiv HATU and 9 equiv DIPEA was charged to the resin. The resulting slurry was allowed to agitate for not less than 2 h at ambient (19-25 °C).

#### *Py/Im, Im/GABA coupling*

*Pre-activation of monomer unit: 4.0 mol equiv of Im or GABA with 3.6 equiv DCC, 3.8 equiv HOAt in 1.5 mL anhydrous DMF for 2 h, filtration of DCU precipitate, and addition of 9 equiv DIPEA prior to coupling.*

After deprotection, the resin was washed with DMF (4 x 1 mL x 1 min), followed by rinsing with 1.5 mL anhydrous DMF. The pre-activated mixture of Im or GABA (4 mol equiv) was charged to the resin and allowed to agitate for a minimum 12 h at ambient (19-25 °C).

#### *N-methylimidazole N-terminal coupling*

After deprotection, the resin was washed with DMF (4 x 1 mL x 1 min), rinsed with 1.5 mL anhydrous DMF, at which point a solution of the DCC/HOAt activated monomer Im (after filtration of the DCU) in 1.5 mL anhydrous DMF with 9 equiv DIPEA was charged to the resin and the resulting mixture agitated at ambient (19-25 °C) for 16 h.

#### *N-isopropylimidazole N-terminal coupling*



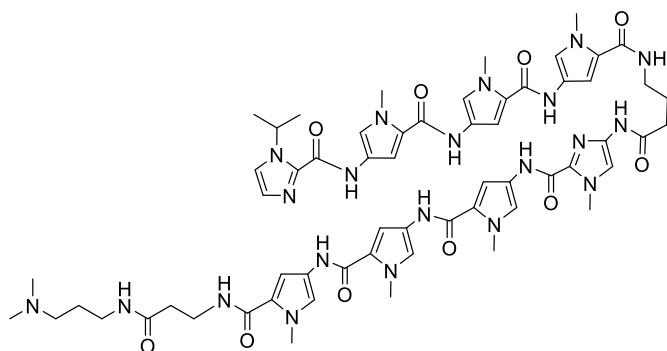
After deprotection, the resin was washed with DMF (4 x 1 mL x 1 min), rinsed with 1.5 mL anhydrous DMF, at which point a solution of the <sup>i</sup>PrIm-COCCl<sub>3</sub> monomer in 1.5 mL anhydrous DMF with 9 equiv DIPEA was charged to the resin and the resulting mixture agitated at ambient (19-25 °C) for 16 h.

*Protocol for cleavage of polyamide from resin*

Polyamide covalently attached to resin was agitated in 1.5 mL of 3-dimethylaminopropylamine at 70 °C for 16 h. After cooling to ambient (19-25 °C), the resin was filtered off and washed with 2 x 0.5 mL methanol. The resulting solution was purified through semi-prep reverse-phase HPLC (Appendix 1.1.1).

*Compound data*

**Polyamide PA4**



*Coupling sequence:*

1 h swelling of β-Ala-PAM resin, (TPW, **Py**/HATU/DIPEA) x3, (TPW, **Im**/HOAt/DCC/DIPEA) x1, (TPW, **GABA** /HOAt/DCC/DIPEA) x1, (TPW, **Py**/HATU/DIPEA) x 3, (TPW, <sup>i</sup>**Pr-Im**/DIPEA) x 1

*Isolated yield (from starting Boc-β-Ala PAM resin): 11%*

*Characterization data:*

<sup>1</sup>H NMR (d<sub>6</sub>-DMSO 500 MHz) δ: 1.42 (d, 6H, J=6.65 Hz), 1.71-1.83 (m, 4H), 2.34-2.38 (m, 4H), 2.62-2.65 (m, 2H), 2.74 (s, 3H), 2.75 (s, 3H), 2.99-3.03 (m, 1H), 3.10-3.14 (m, 1H), 3.19-3.23 (m, 1H), 3.80 (s, 6H), 3.85-3.86 (m, 12H), 3.96 (s, 3H), 5.59-5.65 (sep, 1H, J = 6.65 Hz), 6.88 (s, 1H), 6.90 (s, 1H), 7.07-7.08 (m, 3H), 7.16-7.17 (m, 4H), 7.22 (s, 1H), 7.23 (s, 1H), 7.27 (s, 1H), 7.30 (s, 1H), 7.46 (s, 1H),

7.62 (s, 1H), 8.01-8.03 (m, 2H), 9.18 (s, 1H), 9.88 (s, 1H), 9.89 (s, 1H), 9.93 (s, 1H), 9.94 (s, 1H), 9.96 (s, 1H) 10.24 (s, 1H), 10.43 (s, 1H)

HPLC purity (310 nm): 95.7% [retention time: 21.66 min]

ESI-HRMS: 1250.6059 [1250.6079 calculated for M+H C<sub>58</sub>H<sub>72</sub>N<sub>21</sub>O<sub>10</sub><sup>+</sup>]

## **Chapter 3**

**Profiling the dsDNA binding kinetics of an 8-ring hairpin polyamide incorporating a *N*-terminal <sup>i</sup>Pr-Im**

Scientific study towards an understanding of MGB binding mechanics required methods for objective interrogation of the interaction between a ligand (MGB) and its substrate (dsDNA). Dervan *et al.* demonstrated the enthalpic discrimination of particular DNA sequences by a 6-ring hairpin polyamide.<sup>[22]</sup> That work showcased several methods for the determination of binding free energy. Among the methods available for interrogation is the application of UV thermal melting analysis, a common technique for determining the stability of a dsDNA duplex, which provides a fast, qualitative screening of the binding stabilization.

Due to the structural perturbation of a B-DNA duplex typically observed upon polyamide binding, circular dichroism (CD) spectropolarimetry provided another means to qualitatively assess the binding interaction by generating conformational data. Dervan *et al.* found that binding of a 6-ring hairpin polyamide to the DNA duplex induced chirality in the hairpin polyamide studied. Furthermore, the CD spectra were in agreement with the thermal UV melting analysis in that particular study.<sup>[22]</sup>

DNase I footprinting, a technique that involves the digestion of a sequence by DNase I enzyme, determines the sequences that the polyamide has bound to and thereby prevented digestion.<sup>[64]</sup> This method primarily evaluates the sequence selectivity of polyamides, and can also give a qualitative assessment of the binding affinity. Despite these methods being good for *qualitatively* assessing the binding interaction, a more *quantitative* interrogation is required for a deeper understanding of this very specific type of binding interaction.

What is perhaps considered the gold standard of thermodynamics measurements of MGB interactions with dsDNA, isothermal titration calorimetry (ITC) directly measures the enthalpy change,  $\Delta H$ , from which the free energy of binding,  $\Delta G_B$ , can be calculated. Going a step further, the entropic contribution can also be determined through equation 3.1.<sup>[22]</sup>

$$\Delta G_B = \Delta H - T\Delta S \quad \text{Equation 3.1}$$

The detailed kinetic information from ITC, together with the thermal UV melt stabilization and CD spectra, complete the binding profile of a polyamide for dsDNA sequences. This was detailed in a

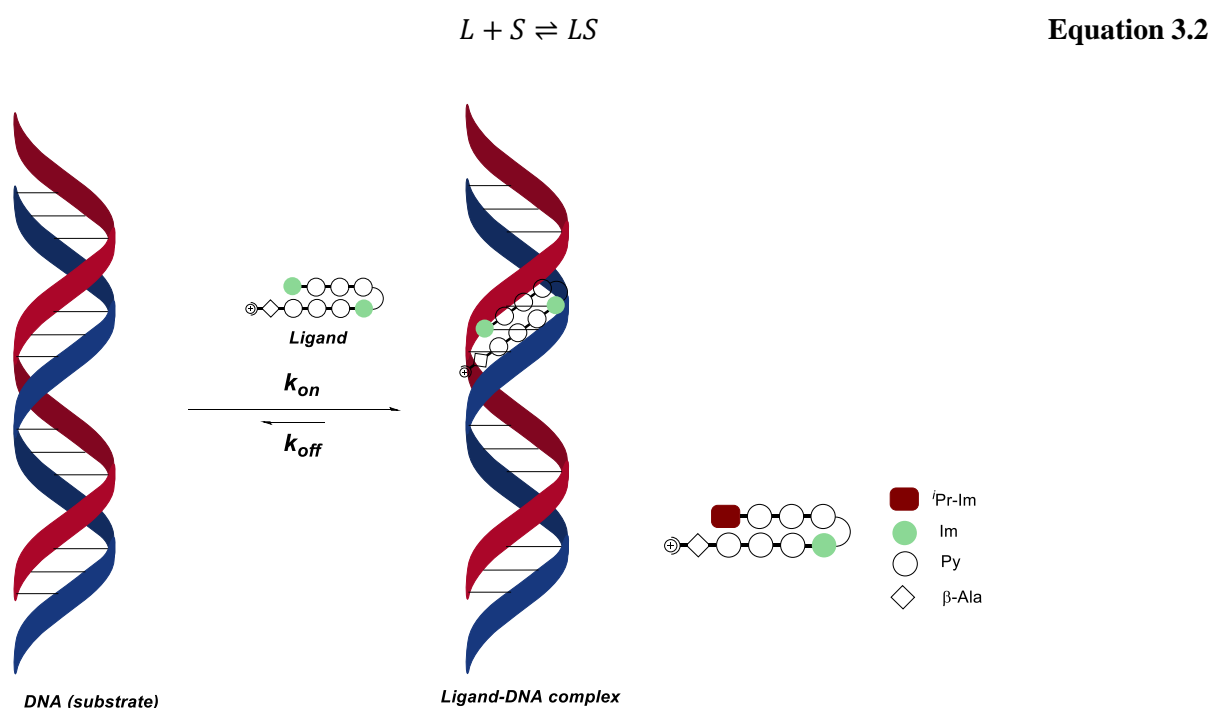
landmark 1996 paper by Dervan in which he addresses our limited knowledge at the time of the forces governing binding affinity and specificity of the hairpin polyamides.<sup>[22]</sup>

To provide further insight into the observed binding affinity, the kinetics of the binding interaction require even more extensive interrogation, and there are few methods that enable this readily. Surface plasmon resonance (SPR) is one option that provides  $k_{on}$  and  $k_{off}$  data, which together can be used to quantify the binding affinity,  $K_D$ .

Another technology, and the one utilized in the work of this thesis, is the switchSENSE® biosensor platform from Dynamic Biosensors, which is a very similar technique to SPR but measures the fluorescence of labelled DNA strands immobilized on a surface instead of the refractive index caused by a mass change.

### 3.1 Kinetic rate theory of ligand-dsDNA binding

For any given interaction between a ligand  $L$  and its target substrate  $S$ , there exists an equilibrium between the bound and unbound states.



**Figure 3.1.** Equilibrium of ligand (hairpin polyamide) association to a substrate (dsDNA) forming the ligand-substrate (polyamide-dsDNA) complex.

The formation of ligand-DNA complex is described by time-dependent rate equations.

Forward rate equation: 
$$\frac{d[LS]}{dt} = k_{on}[L][S] \quad \text{Equation 3.3}$$

Backward rate equation: 
$$\frac{d[LS]}{dt} = k_{off}[LS] \quad \text{Equation 3.4}$$

At the equilibrium state: 
$$K_A = \frac{[LS]}{[L][S]} = \frac{k_{on}}{k_{off}} \quad \text{Equation 3.5}$$

And the dissociation constant  $K_D$  also referred to as the *binding affinity* of a substrate (DNA) for its ligand, is inversely related to the  $K_A$ .

$$K_D = \frac{1}{K_A} \quad \text{Equation 3.6}$$

A balance between the  $k_{on}$  and  $k_{off}$  rates describes the true binding interaction of a ligand for its substrate. In particular, a higher  $k_{on}$  and a lower  $k_{off}$  indicate a strong binding interaction, with a higher  $K_A$ . Conversely, a lower  $K_D$  characterizes a higher binding affinity.<sup>[65,66]</sup>

In order to determine the binding affinity and the  $k_{on}$  and  $k_{off}$  rate constants for polyamides, surface plasmon resonance (SPR) can be used as was shown extensively by the Wilson lab.<sup>[67]</sup>

### *The problem of aggregation*

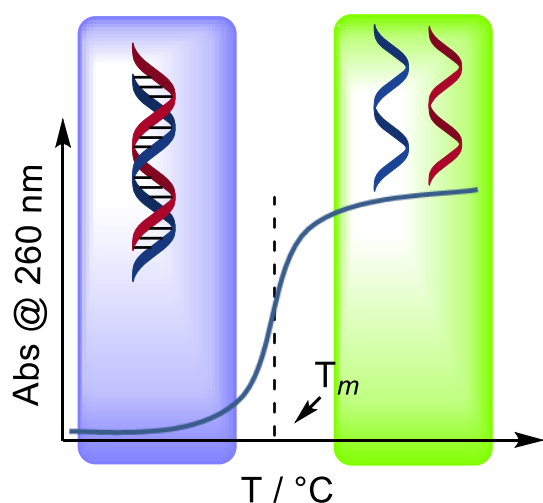
Py-Im hairpin polyamides have demonstrated a propensity for aggregation, and this may hinder the accuracy of measured kinetic data.<sup>[50,68,69]</sup> Potentially due to hydrogen bonding and  $\pi$ - $\pi$  stacking of the heterocycles between polyamides, the PA-PA interactions can manifest themselves as reduced solubility with time. Going to higher concentrations, Wang *et al.* in the Wilson lab found that  $> 5 \mu\text{M}$  was too high and at this concentration ITC was unsuccessful, likely due to the aggregation of 8-ring hairpin polyamides.<sup>[68]</sup> Aggregation thus poses a challenge to the quantification of PA-DNA binding, and caution must be exercised on the timeframe of making solutions to analyzing them.

## 3.2 Rates of binding for comparison

### 3.2.1 Experimental analysis of PA-dsDNA interactions

#### *UV melting analysis*

An expedient method for the qualitative comparison of binding affinities between polyamides for sequences of dsDNA is the thermal UV melt assay.<sup>[22,38]</sup> Due to an effect known as hypochromicity (fig. 3.2), the absorption of UV light is reduced when the DNA forms a double helix and is increased when the DNA is in single-stranded form.<sup>[3]</sup> Exploiting this phenomenon, the UV melt technique involves the slow heating (*e.g.* 1 °C/min) of a sample of dsDNA with/without PA in the presence of a buffer system and the monitoring of the UV absorbance at 260 nm (wavelength for max absorption of DNA). A polyamide binding to a dsDNA duplex will stabilize it. Conceptually, this means that a PA bound to a particular sequence will elevate the melting temperature,  $T_m$ , observed, the degree of stabilization correlating with  $\Delta T_m$ .



**Figure 3.2.** Hypochromicity allows for the UV melting temperature analysis of dsDNA. The x-axis value at the inflection point is designated as the melting temperature of a duplex.

UV melting analysis offers a quick screening for potential binding of a particular polyamide to a given sequence, and is conducted at 1  $\mu\text{M}$  concentration (minimal sample required).<sup>[22,70]</sup> The data only gives qualitative analysis of binding affinity, and the actual binding itself is potentially biased with the temperature elevation inherent to this method. Nevertheless, the technique has seen widespread use

throughout polyamide-DNA investigations, making it a vital screening tool for assessing the selectivity and qualitative affinity early on. However, another technology is required for determination of detailed binding information.

#### *Surface plasmon resonance for interrogation of binding kinetic parameters*

Surface plasmon resonance (SPR) is an optical detection platform that allows real-time and label-free characterization of kinetics and steady-state affinity properties of biomolecular interactions. For a two-component system, typically one of the binding partners is immobilized on the sensor surface, and its interaction with the mobile binding partner (present in solution) is monitored via measurement of the refractive index, RI. The observed change in the RI correlates to mass concentration near the surface with a response output recorded as resonance units (R.U.).<sup>[71,72]</sup>

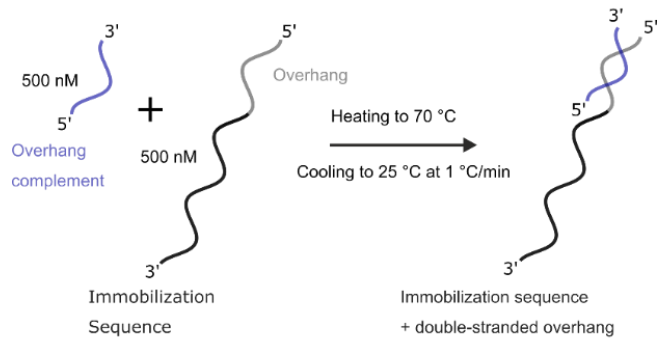
SPR as a method for interrogating polyamide-dsDNA binding interactions was developed mainly by David Wilson using Biacore instrumentation. To DNA sequences labelled on the 5' end with Biotin for immobilization onto a streptavidin-coated chip (Biacore SA) are flowed solutions of known concentrations of polyamides, and the binding data analyzed via a two-site model.<sup>[67]</sup> This technique provides limited data on the dissociation kinetics of polyamide-dsDNA complexes, and an additional and orthogonal examination of the binding dynamics are required, so this method was forgone in light of a relatively new technique, utilizing Dynamic Biosensors' switchSENSE® system.

#### *switchSENSE® technology for determination of the binding kinetic parameters*

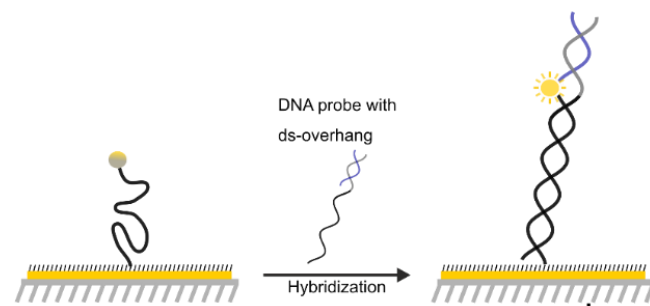
A quick and reliable method for the interrogation of polyamide-dsDNA binding interactions was developed by Dynamic Biosensors (Germany).<sup>[65,73]</sup> Operating the switchSENSE® sensor in *static mode*, the technique utilizes a pre-hybridized DNA probe with an immobilization strand and double-stranded overhang sequence to which the polyamide is targeted. Pre-functionalization of the chip's surface is achieved with attaching DNA containing a (proprietary) fluorophore. After immobilization of the DNA probe to the surface, the binding of the polyamide quenches the fluorophore to different extents and at different rates, dependent on the affinity of the PA for the particular sequence studied.



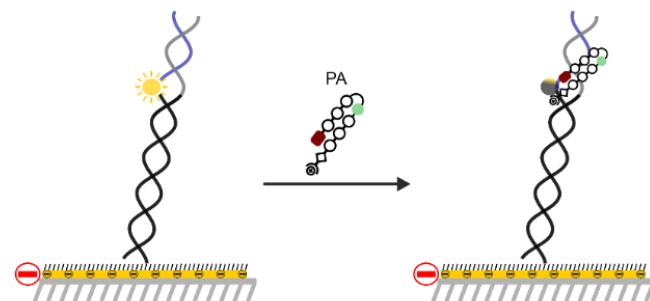
### DNA probe pre-hybridization



### DNA probe immobilization

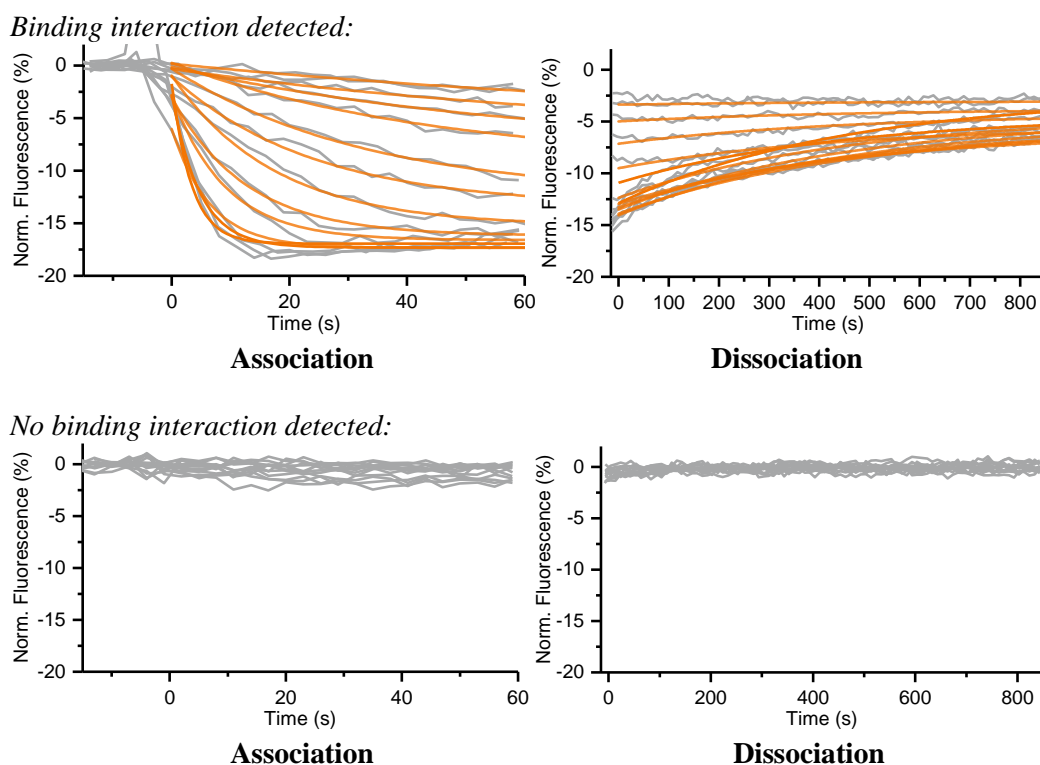


### PA binding



**Figure 3.3.** Static mode switchSENSE® experimental setup for determination the binding profile of PA1-4 for ODN1-3 duplexes.

The output of the interaction is vastly different for a binding vs. a nonbinding interaction, as shown in fig. 3.4.



**Figure 3.4.** Typical association/dissociation curves generated for polyamide binding with dsDNA via the switchSENSE® experimental setup.

From the fitting of the data to appropriate kinetics equations governing the binding of ligands and substrates, the  $k_{on}$  and  $k_{off}$  values can be determined.

A major benefit of the switchSENSE® method is that it is inherently label-free, implying that the effect of the polyamide is measured, as compared FITC-labelled polyamides (or other fluorophore-tagged counterparts) for kinetics measurements, making for difficulty in decoupling of the tag's impact to the binding data collected. A further benefit is that it is performed at room temperature, which removes inherent bias that heating the sample would impart to the binding interaction.<sup>[73,74]</sup>

### 3.3 Aim and Objectives

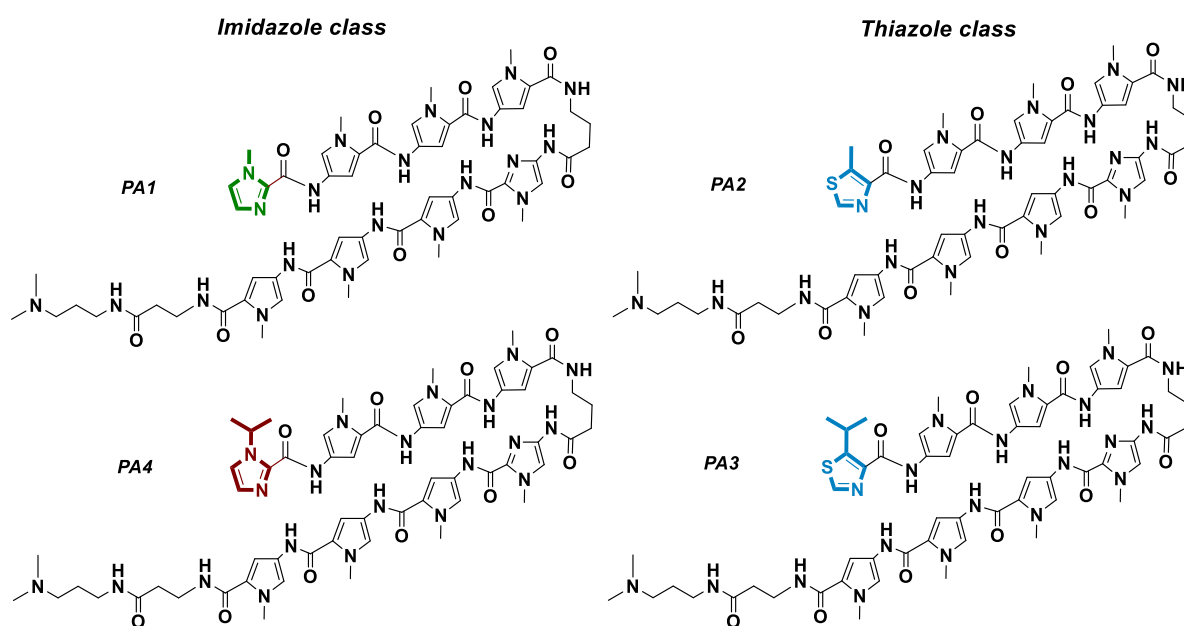
This chapter aims to profile the binding interaction of the 8-ring hairpin polyamide **PA4** with various oligodeoxynucleotide (ODN) sequences. This involves the determination of the binding affinity through UV-melting experiments, which will also be indicative of the selectivity due to changes at the N-terminus of an 8-ring hairpin PA. Furthermore, the switchSENSE® experiments will interrogate the binding kinetics, yielding the  $K_D$ ,  $k_{on}$ , and  $k_{off}$  values for both target and mismatch ODN sequences. The

generation of kinetic rate maps will aim to provide insight into the binding interaction to understand the differences in binding affinity and selectivity observed and a comparative analysis of four polyamides, PA1 through PA4, will be made.

## 3.4 Results

### 3.4.1 Determination of DNA binding profile of PA4 and comparison with PA1-3

The binding capability of four 8-ring hairpin polyamides was studied for the target ARE sequence 5'-ATGTACT-3' using a combination of UV and fluorescence-based methods. Profiling of the kinetic rates through switchSENSE® technology provided insight into the observed binding affinities, allowing for categorization of the polyamides into two separate classes, based on the N-terminal monomer.



**Figure 3.5** 8-ring hairpin polyamides evaluated for binding to dsDNA.

### 3.4.2 UV melting analysis

UV melting analysis was performed on PA4 on 8 sequences of dsDNA, varying the nucleotide recognized at the N-terminal position from G to T, A and C. The <sup>i</sup>Pr-Im unit in the terminal position (PA4; paired across from a Py) upheld the strong selectivity for G-C over the other nucleotide base pairs. The data is shown below in table 3.1.

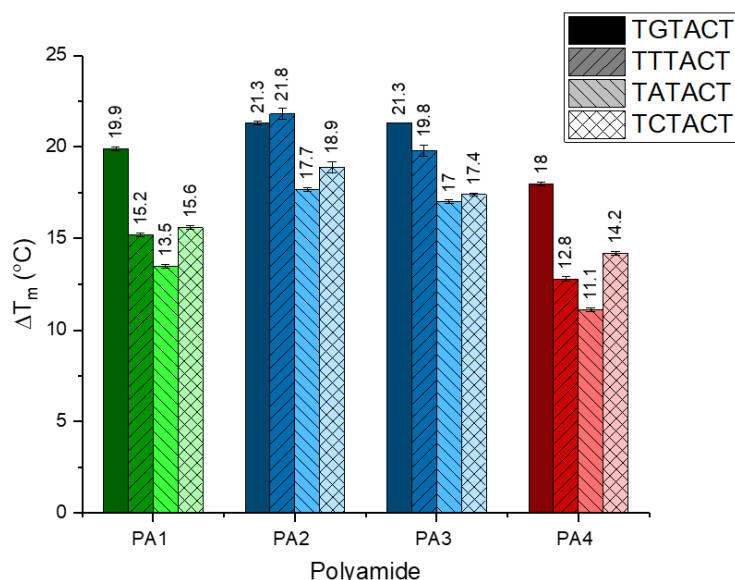
**Table 3.1.** dsDNA  $T_m$  analysis of free DNA with and without **PA4** for target sequence as well as single point mutations (shown is average value from 3 melting measurements).

	5'-CGATG <u>T</u> ACTAGC-3'		5'-CGAT <u>T</u> TACTAGC-3'	
	$T_m$ (°C)	$\Delta T_m$ (°C)	$T_m$ (°C)	$\Delta T_m$ (°C)
Free DNA	46.7 ± 0.1	-	44.0 ± 0.1	-
DNA + PA4	64.6 ± 0.1	18.0 ± 0.1	56.8 ± 0.1	12.8 ± 0.1

	5'-CGATC <u>T</u> ACTAGC-3'		5'-CGAT <u>A</u> TACTAGC-3'	
	$T_m$ (°C)	$\Delta T_m$ (°C)	$T_m$ (°C)	$\Delta T_m$ (°C)
Free DNA	45.3 ± 0.1	-	41.5 ± 0.1	-
DNA + PA4	59.6 ± 0.1	14.2 ± 0.1	52.7 ± 0.1	11.1 ± 0.1

A comparative analysis of **PA4** with **PA1-PA3** (fig. 3.6; **PA1-3** data reproduced here from G. Padroni et al.<sup>[2]</sup>) indicated a similar binding selectivity between **PA1** and **PA4**, with most duplex stabilization observed for the target sequence TGTACT and least thermal stabilization for the TCTACT mismatch. In comparison with the thiazoles **PA2-PA3**, the sequence discrimination of **PA4** more closely resembles the N-Me Im analogue **PA1**.



**Figure 3.6.** DNA UV melt stabilization of **PA1-4** for target dsDNA sequence and single-point mutations (sequence: 5'-CGATXTACATGC-3', where **X** = G/C/A/T; data shown for **PA1-PA3** was reproduced here for comparative analysis, first reported in NAR by G.Padroni<sup>[2]</sup> for the identical sequences studied here; error bars calculated from the standard deviation of three independent measurements. Underlined sequence corresponds to the PA binding site).


The binding affinity of **PA4** was comparable to the control **PA1** for the target sequence, and showed similar thermal discrimination of the mismatch sequences. The thiazole analogues **PA2-3** demonstrated slightly higher affinity for all sequences studied, but lack the discrimination of G-C from T-A when paired across a Py unit, which limited the utility of these monomers.

### 3.4.3 switchSENSE® fluorescence melting analysis

For the switchSENSE® analyses of **PA1-4**, the DNA sequences chosen included the target ARE sequence (**ODN1**), a single point mismatch (**ODN2**) and a two-point mismatch (**ODN3**). **ODN1-3** included the stabilization sequence GCGATT (required to ensure a stable duplex is maintained and avoid dissociation of the base pairs over time) alongside the immobilization sequence that is the complement of the DNA immobilized on the biosensor surface and immobilizes the sequences prior to analysis (table 3.1).

**Table 3.2.** DNA sequences immobilized for switchSENSE® kinetics /  $T_m$  experiments stabilization sequence in blue, PA binding site (minor groove) in red, immobilization sequence in black).

DNA sequences studied	
Target sequence ( <b>ODN1</b> )	5'-GCGATTATGTACTATCA-3'
Single mismatch ( <b>ODN2</b> )	5'-GCGATTATGCATTATCA-3'
Two-letter mismatch ( <b>ODN3</b> )	5'-GCGATTATTTACTATCA-3'

  
 Polyamide binding site

switchSENSE® analysis of the melting stabilization induced by the polyamide binding was determined by operation of the instrument in static mode. The trend in the thermal stabilization agreed with that obtained from the UV-melting analysis. **PA1** showed a strong affinity for the target sequence 5'-GTAC-3', with a melting stabilization of 5.12 °C. **PA1** binding to the mismatch sequence 5'-TTAC-3' showed a modest stabilization of 1.97 °C, indicating the unit is more selective for G over T. The GCAT mismatch sequence was much less tolerated, and no melting stabilization was detected.

For **PA4**, the stabilization of the target sequence was slightly higher than the methyl-Im analogue, with an increase of the melting temperature by 6.78 °C. The selectivity was confirmed with negligible (0.71

°C) stabilization of the 5'-TTAC-3' mismatch, and no thermal stabilization of the 5'-GCAT-3' sequence.

**Table 3.3.** Fluorescence UV melting temperatures determination via switchSENSE®.

Sample	$\Delta T_m$ at 20 nM (°C)	Error $\Delta T_m$ ( $\pm$ °C)
PA1 + ODN1	5.12	0.17
PA1 + ODN2	1.97	0.13
PA1 + ODN3	0.00	0.29

Sample	$\Delta T_m$ at 20 nM (°C)	Error $\Delta T_m$ ( $\pm$ °C)
PA2 + ODN1	6.60	0.53
PA2 + ODN2	5.93	0.40
PA2 + ODN3	2.87	1.37

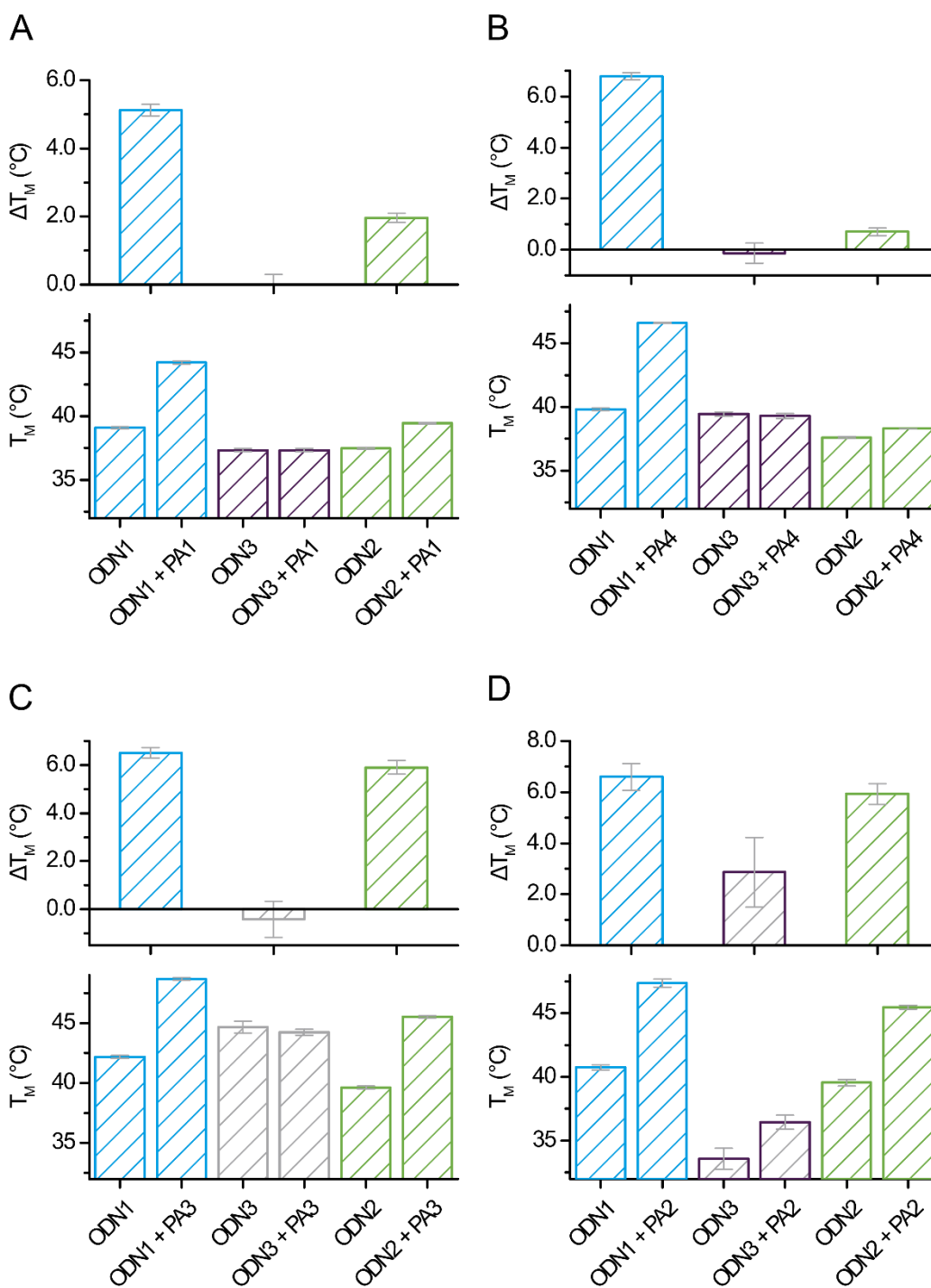
  

Sample	$\Delta T_m$ at 20 nM (°C)	Error $\Delta T_m$ ( $\pm$ °C)
PA3 + ODN1	6.50	0.22
PA3 + ODN2	5.81	0.23
PA3 + ODN3	-0.41	0.75

Sample	$\Delta T_m$ at 20 nM (°C)	Error $\Delta T_m$ ( $\pm$ °C)
PA4 + ODN1	6.78	0.14
PA4 + ODN2	0.71	0.08
PA4 + ODN3	-0.13	0.39

Comparative analysis of the fluorescence melt stabilization of **ODN1-3** (fig. 3.7) indicated that the N-terminal imidazole polyamides **PA1** and **PA4** have a strong selectivity preference for **ODN1** (the target sequence) over the 1 and 2 base pair mismatches (**ODN2** and **ODN3**, respectively). Furthermore, the melting stabilization indicated that **PA1** and **PA4** are similar in both the degree of stabilization as well as the selectivity, with **PA4** slightly demonstrating a slightly higher melting stabilization for **ODN1** and negligible stabilization of the **ODN2** single base pair mismatch.



**Figure 3.7.** Fluorescence melting temperature comparisons of PA1-4 with ODN1-3.

### 3.4.4 switchSENSE® association/dissociation kinetics measurements of polyamide-dsDNA complexes

switchSENSE® analysis of the polyamide-dsDNA binding interactions was carried out isothermally (25 °C), and the rate constants determined through the appropriate equations. The picomolar binding

affinity of **PA1** and **PA4** (table 3.4) indicate that an imidazole-based monomer in the N-terminal position confers a higher affinity to the binding interaction compared with alkyl thiazole PA analogues **PA2-3**. Kinetically, the basis of this trend is due to the faster  $k_{on}$  of the PAs with N-terminal imidazole groups (fig. 3.5), which is on the order of  $10^6$  to  $10^7$ , roughly 10 times higher than for the thiazole-containing PAs (magnitude on order of  $10^5$ ). The  $k_{off}$  values are around the same order of magnitude for all four PAs studied,  $10^{-3}$  for the target and  $10^{-2}$  for the mismatch sequences studied. As a result, the affinity,  $K_D$ , of the imidazole-analogues is in the picomolar range, 254 pM for **PA1** and 188 pM **PA4**. The  $K_D$  is significantly higher for the alkyl thiazole analogues, in the low nanomolar magnitude, indicating a worse affinity of these compounds for their target dsDNA sequence. Furthermore, **PA2-3** lack discrimination of a G-C base pair from a T-A base pair at the N-terminal position of the PA.

**Table 3.4.**  $K_D$  from switchSENSE<sup>®</sup> experiments.

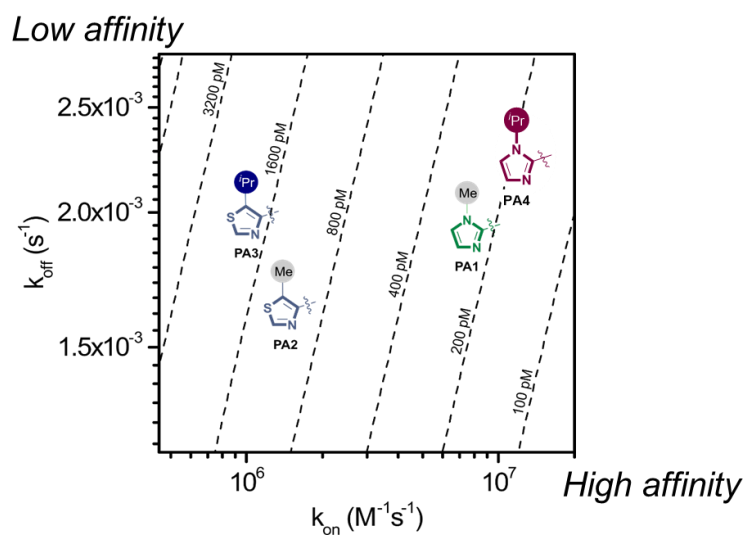
$K_D$ (pM)	PA1	PA2	PA3	PA4
<sup>5</sup> GCG ATT AT <b>G</b> TA <b>C</b> TAT--- CGC TAA TA <b>C</b> AT <b>G</b> ATA <sup>5</sup> <b>ODN1</b>	254 ± 8	1170 ± 70	1970 ± 240	188 ± 5
<sup>5</sup> GCG ATT AT <b>T</b> <b>T</b> AC TAT--- CGC TAA TA <b>A</b> <b>A</b> TGATA <sup>5</sup> <b>ODN2</b>	1320 ± 70	1250 ± 110	2880 ± 440	967 ± 35
<sup>5</sup> GCG ATT AT <b>G</b> <b>C</b> AT TAT--- CGC TAA TA <b>C</b> <b>G</b> TA ATA <sup>5</sup> <b>ODN3</b>	ND	15400 ± 7700	ND	1100 ± 100

The rate map shown in fig. 3.8 provides an intuitive view of the relationship between  $k_{on}$  and  $k_{off}$ , enabling the classification of these polyamides into different binding regimes.

The rate values are substantially higher for the polyamide with the N-terminal thiazole that bears the isopropyl substituent vs. the methyl substituent, implying that the subtle change in this N-terminal heterocycle (a mere 2-carbon extension) had a measurable impact on the binding interaction. This also was observed for the methyl and isopropyl analogues of the N-terminal imidazole polyamides. The rate map provides insight into the dissociation and association rates' contribution to the observed binding

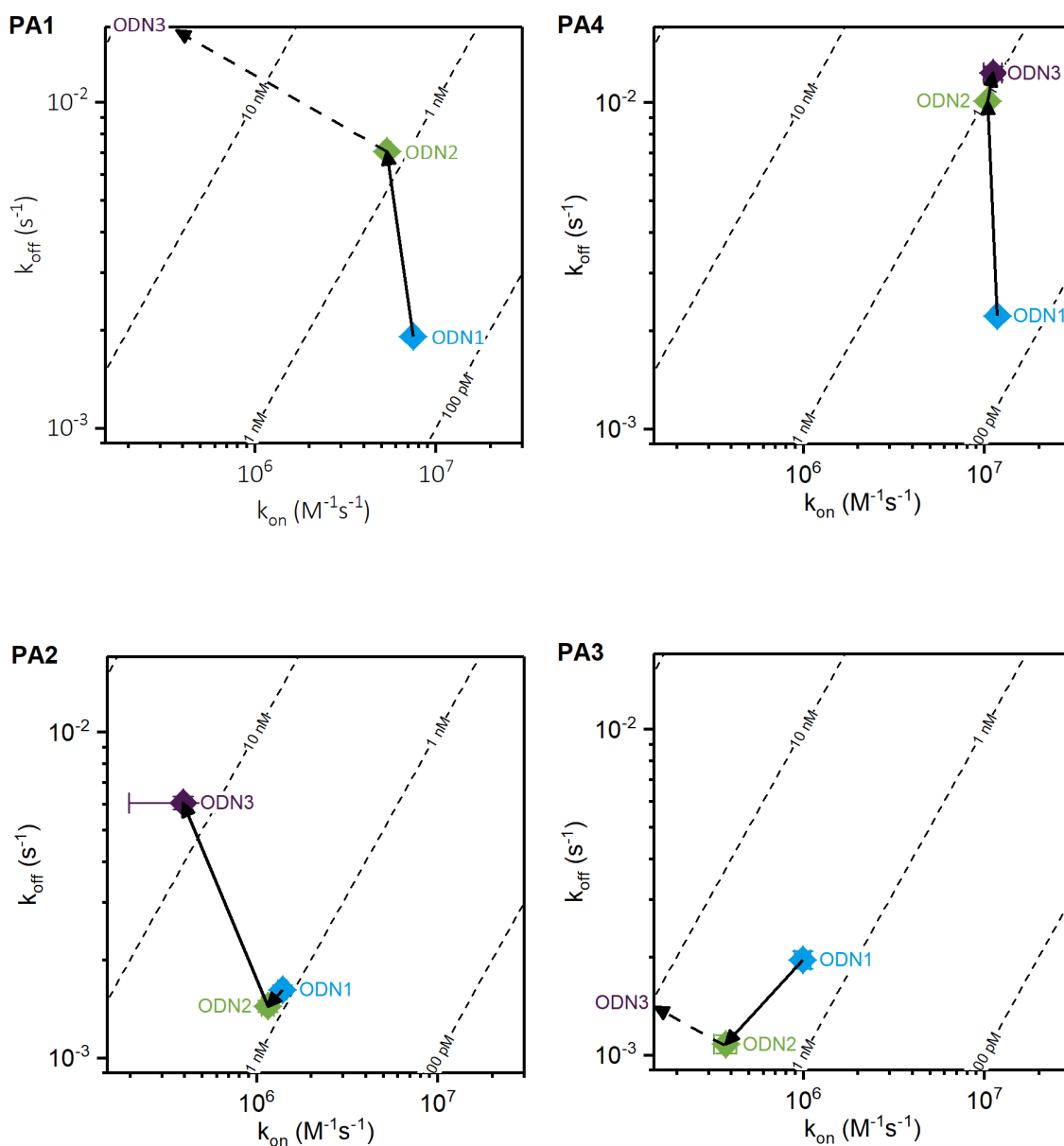


affinity, indicating that although the  $k_{off}$  rates are similar, the difference in binding is attributable to faster  $k_{on}$  rates for **PA1** and **PA4**.



**Figure 3.8.** Kinetic rate map of **PA1-4** for the target dsDNA sequence 5'-ATGTACT-3' (**ODN1**).

The binding association/dissociation constants were further analyzed in the rate maps for each polyamide with each ODN sequence (fig. 3.9). The differences in the selectivity of each polyamide for the target sequence ODN1 compared with the mismatch sequences ODN2 and ODN3 revealed some interesting trends that appear specific for each polyamide to each of the DNA sequences studied.



**Figure 3.9.** Kinetic rate maps of **PA1-4** for the target and mismatch sequences **ODN1-3**.

**PA1** and **PA4** show strong selectivity (fig. 3.9), with the target sequence **ODN1** well separated from the mismatches. By comparison, the thiazole analogues **PA2** and **PA3** have similar affinities for both the target sequence and the single-letter mismatch **ODN2**, with **PA3** showing a slightly slower  $k_{off}$ .

The association/dissociation rates show a very interesting trend for **PA4**: it appears that, much like **PA1**, the  $k_{on}$  is fast for all the sequences studied, and it is the dissociation rate  $k_{off}$  that varies and forms the basis of the discrimination of the target sequence from the mismatches.

For **PA3**, there is a slight difference in the affinity for the target **ODN1** and the single letter mismatch **ODN2**, due to the faster association rate of **PA3** with **ODN1** over **ODN2** (table 3.5).

**PA2** displays fascinating kinetics compared with the others for **ODN1-3**. For the target **ODN1** and the single letter mismatch **ODN2**, the association and dissociation rates are similar. This is in full agreement with the thermal UV and fluorescence melt data, which indicated equal thermal stabilization for both G and T in the position targeted by the N-terminal heterocycle. For the two-letter mismatch sequence **ODN3**, the lower binding affinity is due to a combination of both slower association and faster dissociation rates (table 3.5).

**Table 3.5.**  $K_D$  and rate constants from switchSENSE<sup>®</sup> experiments.

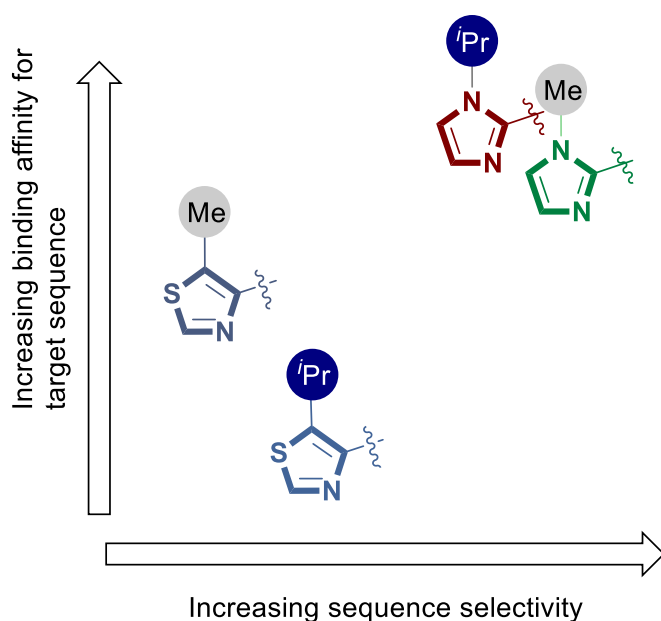
<b>PA1</b>	Sample	$k_{ON}$ ( $M^{-1}s^{-1}$ )	Error $k_{ON}$ ( $M^{-1}s^{-1}$ )	$k_{OFF}$ ( $s^{-1}$ )	Error $k_{OFF}$ ( $s^{-1}$ )	$K_D$ (pM)
	PA1 + ODN1	<b><math>7.49 \times 10^6</math></b>	$2.10 \times 10^5$	$1.91 \times 10^{-3}$	$3.00 \times 10^{-5}$	<b>254</b>
	PA1 + ODN2	$5.37 \times 10^6$	$2.80 \times 10^5$	$7.07 \times 10^{-3}$	$1.00 \times 10^{-4}$	1320
	PA1 + ODN3	No interaction	No interaction	No interaction	No interaction	No interaction
<b>PA2</b>	Sample	$k_{ON}$ ( $M^{-1}s^{-1}$ )	Error $k_{ON}$ ( $M^{-1}s^{-1}$ )	$k_{OFF}$ ( $s^{-1}$ )	Error $k_{OFF}$ ( $s^{-1}$ )	$K_D$ (pM)
	PA2 + ODN1	<b><math>1.39 \times 10^6</math></b>	$8.00 \times 10^4$	$1.62 \times 10^{-3}$	$5.00 \times 10^{-5}$	<b>1170</b>
	PA2 + ODN2	$1.15 \times 10^6$	$9.00 \times 10^4$	$1.44 \times 10^{-3}$	$6.00 \times 10^{-5}$	1250
	PA2 + ODN3	$3.93 \times 10^5$	$1.96 \times 10^5$	$6.07 \times 10^{-3}$	$2.70 \times 10^{-4}$	15400
<b>PA3</b>	Sample	$k_{ON}$ ( $M^{-1}s^{-1}$ )	Error $k_{ON}$ ( $M^{-1}s^{-1}$ )	$k_{OFF}$ ( $s^{-1}$ )	Error $k_{OFF}$ ( $s^{-1}$ )	$K_D$ (pM)
	PA3 + ODN1	<b><math>9.92 \times 10^5</math></b>	$1.01 \times 10^5$	$1.96 \times 10^{-3}$	$1.20 \times 10^{-4}$	<b>1970</b>
	PA3 + ODN2	$3.74 \times 10^5$	$5.20 \times 10^4$	$1.08 \times 10^{-3}$	$7.00 \times 10^{-5}$	2880
	PA3 + ODN3	No Interaction	No Interaction	No Interaction	No Interaction	No Interaction
<b>PA4</b>	Sample	$k_{ON}$ ( $M^{-1}s^{-1}$ )	Error $k_{ON}$ ( $M^{-1}s^{-1}$ )	$k_{OFF}$ ( $s^{-1}$ )	Error $k_{OFF}$ ( $s^{-1}$ )	$K_D$ (pM)
	PA4 + ODN1	<b><math>1.18 \times 10^7</math></b>	$3.00 \times 10^5$	$2.22 \times 10^{-3}$	$3.00 \times 10^{-5}$	<b>188</b>
	PA4 + ODN2	$1.04 \times 10^7$	$4.00 \times 10^5$	$1.01 \times 10^{-2}$	$1.00 \times 10^{-4}$	967
	PA4 + ODN3	$1.12 \times 10^7$	$1.30 \times 10^6$	$1.23 \times 10^{-2}$	$6.00 \times 10^{-4}$	1100

### 3.5 Summary

This chapter focused on profiling the polyamide-dsDNA kinetics and binding interaction. To accomplish this, it was first necessary to qualitatively screen the polyamides by UV melting analysis

over different sequences alongside a suitable control, which was elected to be the 8-ring hairpin polyamide **PA1**, originally designed by the Dervan lab for targeting the ARE consensus sequence.<sup>[1]</sup> This rapid screening provided promising preliminary data, highlighting a comparable binding affinity and selectivity for the polyamide containing the novel <sup>i</sup>Pr-Im monomer in the N-terminal position (ring position 8, **PA4**). A more detailed analysis of the binding kinetics was determined through the use of switchSENSE® biosensor technology in collaboration with Thomas Welte (Dynamic Biosensors), providing insight into association/dissociation rate constants as they relate to the observed binding affinities. The binding affinity for the target sequence 5'-ATGTACT-3' (**ODN 1**) of **PA4** was 188 pM, which was comparable to the 254 pM measured for the N-methyl analogue **PA1**. Mapping of the kinetic rates showed that although the  $k_{off}$  for all PAs tested were of similar magnitude, the imidazole PAs demonstrated both stronger affinity and selectivity compared to the N-terminal alkyl thiazole PAs. The kinetic basis for the sequence discrimination of the polyamides analyzed varied between each polyamide for **ODN1-3**. **PA4** displayed a very similar association rate for the target **ODN1** and the single and two-letter mismatch sequences (**ODN2** and **ODN3**, respectively), but its selectivity came from the substantially faster  $k_{off}$  for the mismatches. Importantly, it has been revealed by this work that the kinetic basis for the selectivity differences varies between imidazole PAs, where the dissociation gives rise to different selectivity, and the thiazole PAs, where both the slower association and faster dissociation rates factor into the selectivity differences observed.

Ultimately, the data provided by the UV and switchSENSE® methods suggest that **PA4** strikes a fine balance between increasing steric bulk and maintenance of both affinity and specificity. Analysis of the binding kinetics of **PA4** found it comparable to the control **PA1**, which places both the polyamide as well as the novel <sup>i</sup>Pr-Im building block as viable candidates for practical use in research.



**Figure 3.10.** Trends observed from kinetics experiments.

## 3.6 Experimental

### 3.6.1 Pre-hybridization of DNA probes for kinetics experiments

DNA oligodeoxyribonucleotides were produced containing, at the 3'-end, a generic 48-mer switchSENSE immobilization sequence (5' - ATC AGC GTT CGA TGC TTC CGA CTA ATC AGC CAT ATC AGC TTA CGA CTA- 3'), complementary to the tethered DNA on the switchSENSE chip, and the corresponding target sequence as an overhang at the 5'-end. Short oligos complementary to the individual overhang sequences were also synthesized. All DNA oligos used for switchSENSE experiments were synthesized by biomers.net GmbH (Ulm, Germany).

To yield DNA probes with double-stranded overhang regions, the immobilization strand and the corresponding overhang complement strand were mixed at an equal concentration (500 nM) in PE140 buffer (10 mM NaPi, pH 7.4, 140 mM NaCl, 50  $\mu$ M EDTA, 50  $\mu$ M EGTA, 0.05 % Tween 20) and heated to 70°C. For optimal annealing, the DNA solution mix was then slowly cooled to room temperature at a rate of 1 °C/min.

For immobilization of DNA probes on switchSENSE chips, the pre-hybridized DNA probes were hybridized to the surface grafted complement of the immobilization sequence using an inbuilt routine of the DRX instrument.

### 3.6.2 Kinetics determination

Binding experiments of polyamides to DNA probes were carried out using a DRX switchSENSE platform on an MPC2-48-2-Y1-S sensor chip (Dynamic Biosensors GmbH, Martinsried, Germany). PE140 buffer (10 mM NaPi, pH 7.4, 140 mM NaCl, 50  $\mu$ M EDTA, 50  $\mu$ M EGTA, 0.05 % Tween 20) served as running buffer for the interaction experiments and for hybridization and on-chip immobilization of the DNA nano levers.

The association and the dissociation kinetics of the polyamides to the DNA probes was measured under a flow rate of 1000  $\mu$ l/min polyamide solution and buffer, respectively. Binding traces, corresponding to the absolute fluorescence intensity readout, were recorded using the instrument's **static measurement mode**. In this measurement mode, a constant mild negative potential is applied to the measurement electrode, which results in an upright orientation of the DNA nano levers. Every set of association and dissociation experimental cycles in the polyamide concentration series was referenced to a buffer blank injection recorded prior to the polyamide injection. The sensor surface was regenerated after every dissociation measurement, by treatment with a basic regeneration solution (Dynamic Biosensors GmbH, Martinsried, Germany), allowing for the immobilization of fresh DNA probes for the subsequent measurements.

For data analysis, the kinetic data sets were processed using switchANALYSIS software (Dynamic Biosensors GmbH, Martinsried, Germany) and finalized plots were generated with Origin2015® software (Additive GmbH, Friedrichsdorf, Germany). The kinetic rate constants were determined with the following equations:

$$\text{Dissociation : } y(t) = y_0 + A_{off} * e^{-\frac{t}{\tau_{off}}} \quad \text{Equation 3.7}$$

Where  $k_{off} = \frac{1}{\tau_{off}}$

Association : 
$$y(t) = y_0 + A_{on} * e^{-t*(c*k_{on}+k_{off})}$$
 **Equation 3.8**

$y$  = sensor signal

$y_0$  =  $y$  offset

$A_{off}$  = fit amplitude of dissociation fit

$t$  = time;  $\tau_{off}$  = dissociation time constant

$k_{off}$  = dissociation rate constant

$A_{on}$  = fit amplitude of association fit

$c$  = concentration

$k_{on}$  = association rate constant.

Equilibrium Dissociation constant: 
$$K_D = \frac{k_{off}}{k_{on}}$$
 **Equation 3.9**

### 3.6.3 Melting temperature determinations

#### *UV absorption melting analysis*

Melting temperature analyses were performed on a Shimadzu UV spectrometer UV 1800 equipped with a temperature-controlled cell holder (cell path length = 1 cm). An aqueous solution of phosphate buffer (consisting of 10 mM monobasic phosphate, 10 mM dibasic phosphate, 100 mM NaCl, at pH 7.0) was used. The corresponding duplex DNA and the PA ligands were mixed at a stoichiometry of 1:2 (DNA:PA) providing a final dsDNA concentration of 1  $\mu$ M for each experiment. Denaturing profiles were recorded at 260 nm from 20 to 90  $^{\circ}$ C, with a heating rate of 1  $^{\circ}$ C/min. The reported melting temperatures were defined as averages of the maximum of first derivative plots of three independent denaturing profiles.

#### *Fluorescence melting by switchSENSE<sup>®</sup>*

Melting experiments were carried out using a DRX switchSENSE platform on a MPC2-48-2-Y1-S sensor chip (Dynamic Biosensors GmbH, Martinsried, Germany). PE140 buffer (10 mM NaPi, pH 7.4, 140 mM NaCl, 50  $\mu$ M EDTA, 50  $\mu$ M EGTA, 0.05 % Tween 20) served as running buffer for the melting experiments and on-chip immobilization of the pre-hybridized DNA probes.

Pre-hybridized DNA probes, both in the absence and in the presence of 20 nM of polyamide, were immobilized on the switchSENSE chip and heated in PE140 buffer to 60  $^{\circ}$ C at a rate of 5  $^{\circ}$ C/min, under

a buffer flow rate of 5  $\mu\text{l}/\text{min}$ . The melting curves were recorded in the static measurement mode and referenced to the temperature-dependent fluorescence intensity change of the DNA-attached dye.

The melting curves were referenced to the bare DNA measurements using the switchANALYSIS software (Dynamic Biosensors GmbH, Martinsried, Germany). Final data processing was carried out using Origin2015® software (Additive GmbH, Friedrichsdorf, Germany). For better visibility of the thermal shift, the referenced data was normalized. The melting temperatures were extracted from the referenced and normalized data by global non-linear regression using a Boltzman equation:

$$y = \frac{A_1 - A_2}{1 + e^{(T - T_M)/dx}} + A_2 \quad \text{Equation 3.10}$$

$y$  = sensor signal

$A_1$  = sensor signal start level

$A_2$  = sensor signal saturation level

$T$  = Temperature

$T_M$  = Melting Temperature

To quantify the degree of thermal stabilization, the thermal shift ( $\Delta T_M$ ) was calculated from the melting temperatures in presence of polyamide and absence of polyamide.

$$\Delta T_M = T_{M,20 \text{ nM}} - T_{M,0 \text{ nM}} \quad \text{Equation 3.11}$$

$T_{M,20 \text{ nM}}$  = Melting Temperature at 20 nM polyamide concentration

$T_{M,0 \text{ nM}}$  = Melting Temperature at 0 nM polyamide concentration



## **Chapter 4**

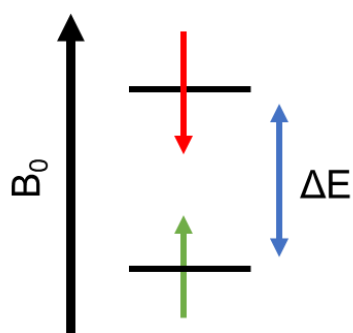
### **Structural determination of polyamide-dsDNA complexes by NMR-restrained molecular dynamics simulations**

The first X-ray crystal structure of netropsin was published in 1985 by the Dickerson lab for the target sequence 5'-CGCGAAATTCGCG-3', revealing the intricate fit of the natural product into the minor groove to displace the water spline and form new hydrogen bonding contacts.<sup>[11]</sup> Since then, numerous X-ray crystal structures have been obtained and studied to understand the finer details of MGB binding interactions.<sup>[21,75]</sup> In a breakthrough achievement of NMR capability, a study led by Peter Dervan and David Wemmer in 2004 focused on the solution phase structure determination of cyclic PA-dsDNA complex using [<sup>1</sup>H-<sup>1</sup>H] NOESY NMR data.<sup>[76]</sup> The high scientific value attributed to this method is that it incorporates solution phase data in phosphate buffer at pH 7.4, which more accurately depicts the environment of biological systems. This NMR-guided structural analysis of the polyamide-dsDNA complexes in a buffered aqueous system was selected as the method of choice for understanding the basis of the binding differences between **PA1-PA4** observed in kinetic experiments described in chapter 3.<sup>[2]</sup>

## 4.1 NMR theory and MD theory

### 4.1.1 NMR first principles

As well articulated by P.J. Hore, the analogy of a nucleus' spin as being a proficient spy on the atoms' environment is a good one.<sup>[77]</sup> It is the magnetic attributes of the nuclei in an atom that enable NMR analysis. More fundamentally, the exposure of spin-1/2 nuclei (<sup>1</sup>H, <sup>13</sup>C, <sup>15</sup>N) to an external magnetic field ( $B_0$ ) orients the spins to align either parallel or antiparallel. This splits the two states into discrete energy levels that bear a magnetic quantum number,  $m$ .<sup>[77]</sup>



**Figure 4.1.** Energy level diagram for the hydrogen nucleus in a magnetic field.

The energy levels are filled according to the Boltzmann distribution, with a slightly larger population residing in the lower energy level.<sup>[77]</sup>

$$\frac{N_{\text{antiparallel}}}{N_{\text{parallel}}} = e^{-\Delta E/k_B T} \quad \text{Equation 4.1}$$

In a standard NMR experiment, the 1D spectrum is obtained by measuring the free-induction decay (FID). This response is produced when the radiofrequency (RF) pulse sequence provides the sample the energy necessary for a ‘spin-flip’ from parallel (lower energy) to anti-parallel (higher energy), equalizing the population states. That activation energy required for the spin-flip to occur is a function of the nucleus’ gyromagnetic ratio ( $\gamma$ ) and the magnetic field strength ( $B_0$ ).<sup>[77]</sup>

$$\Delta E = \frac{\gamma h B_0}{2\pi} \quad \text{Equation 4.2}$$

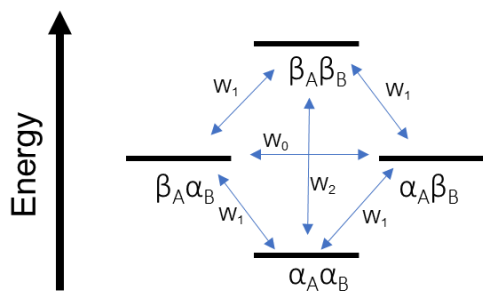
After the RF excitation pulse, the nuclei are allowed to relax back to the ground state energy levels. This relaxation is accompanied by a voltage response in which the FID is obtained. A Fourier Transform (FT) algorithm is then applied to this FID to deconvolute the data into the standard 1D NMR spectra familiar to most.<sup>[77]</sup>

The different chemical environments of the nuclei in a sample give rise to detectable changes in the resonance frequencies, measured as a *chemical shift* from a reference signal.<sup>[77]</sup>

### 2D [<sup>1</sup>H-<sup>1</sup>H] NOESY NMR

Among the various 2D methods which are available to help assign the spectral data and determine aspects of the 3D structure of a molecule, the [<sup>1</sup>H-<sup>1</sup>H] NOESY experiment is of considerable utility as it provides useful information on the spatial distances between certain atoms.<sup>[77]</sup>

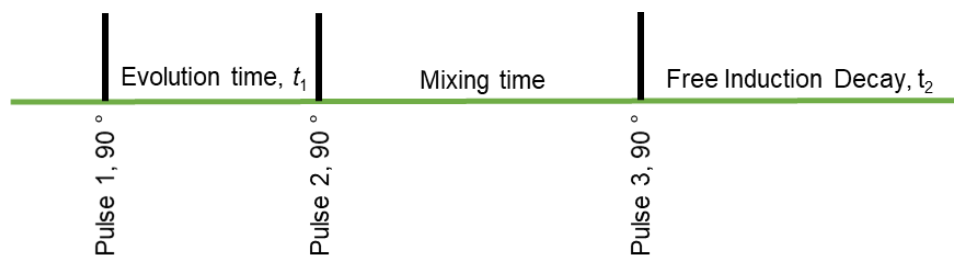
The spins of two nuclei can be coupled via two mechanisms, through-bond *scalar coupling* or through-space *dipolar coupling*. The Nuclear Overhauser Effect, or NOE, is a phenomenon in which the dipolar coupling of two nuclei (in the current instance, this is proton-proton dipolar coupling) generates four potential energy levels as a result of the spin-lattice relaxation pathways available.<sup>[77]</sup>



**Figure 4.2.** Energy level diagram for a pair of spin-1/2 nuclei A and B.

The cross-relaxation mechanism allows for the determination of the proximity of two protons, which is ultimately manifested in the intensity observed for the NOE correlation cross-peak.<sup>[77]</sup>

The pulse sequence, as for most 2D NMR experiments, is unique for the NOE set up.

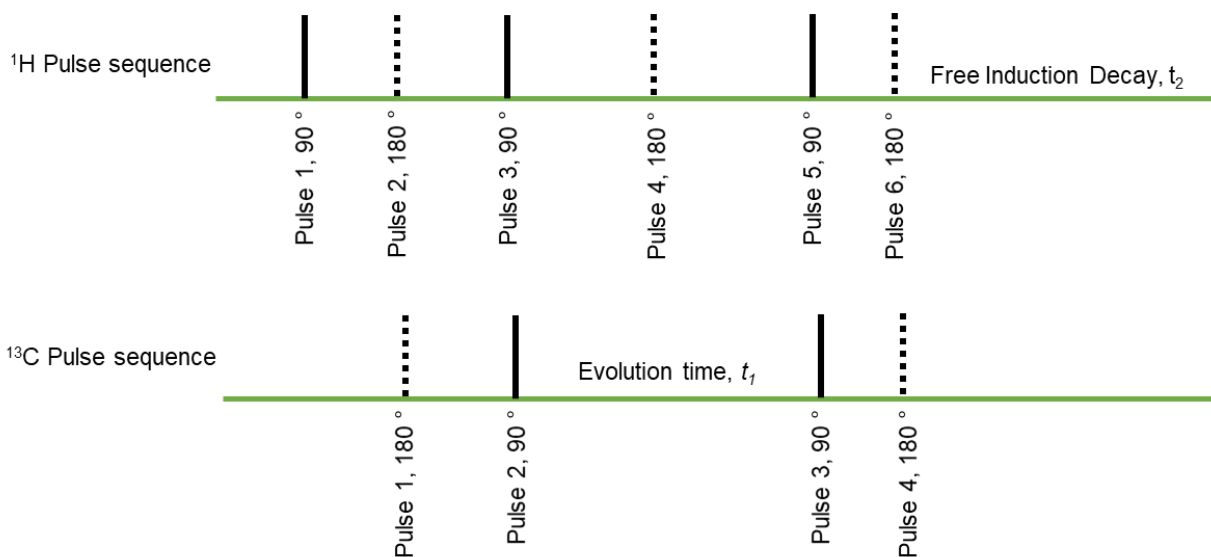


**Figure 4.3.** Pulse sequence for a standard NOESY experiment.

### 2D [ $^1\text{H}$ - $^{13}\text{C}$ ] HSQC NMR

The 2D Heteronuclear Single-Quantum Correlation spectroscopy (HSQC) NMR experiment invokes the use of a specialized pulse sequence to obtain information on the heteronuclear correlations between protons and other nuclei to which they are directly attached, which for the purpose of this thesis is  $^{13}\text{C}$ . It is a scalar coupling, and becomes useful when looking for confirmation of assignments in either 1D spectrum, helping to trace the carbon skeleton frame of a molecule/complex.<sup>[77]</sup>

The pulse sequence is depicted below in fig. 4.4.



**Figure 4.4.** HSQC pulse sequence for 2D [ $^1\text{H}$ - $^{13}\text{C}$ ] NMR.

## 4.2 Molecular Dynamics

As quantum mechanical models are far too resource-intensive to calculate for biomolecules, the use of molecular dynamics fills an important void for practical structural determination. Molecular dynamics (MD) simulations are computer simulations that study the physics of a time-evolving system of atoms and/or molecules.<sup>[78]</sup> The value of the theory lies in the fact that the calculations can be performed on biomolecules and complexes in the solution phase (i.e. buffered water) and provide information on the dynamics of the true environment. This key benefit over X-ray crystal diffraction data collected avoids the bias introduced by the removal of the complex or biomolecule in question from the aqueous environment. Crystallization and removal of solvent from the system can substantially affect any molecule's form, bringing into question the validity of the crystal structure.<sup>[19,31,78,79]</sup>

### 4.2.1 Molecular mechanics of biological complexes from first principles

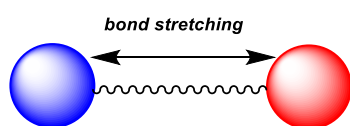
The molecular dynamics of a system of atoms, in our case the polyamide-dsDNA complex in aqueous buffer, is a computer simulation that describes the evolution of the physical motions over time. In this way, it is perhaps incorrect to narrow the result of these simulations to a single structure, since the value of these calculations lies in the frame by frame video. Nevertheless, with averaging over an ensemble of clusters one can obtain a reasonable structure depicting a most probable conformation.

The equations of motion that govern the atoms' motions are derived from the potential energy of the atoms and bonds, including the Van der Waals interactions present throughout the system. In the case of large systems of biomolecular interactions, the complexity is too large to reasonably be modelled with a quantum mechanics approach, and thus a molecular mechanics approach is more appropriate.

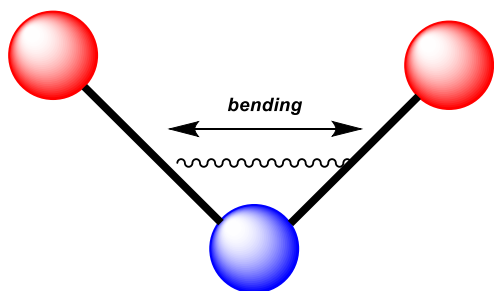
### *Choosing the force field*

Confident reliance on the parameters/constants, or *force field*, employed in the molecular mechanics set of potential energy equations is vital for an accurate representation of the system of interest.<sup>[79,80]</sup> For this study, the force field chosen was AMBER, or Assisted Model Building with Energy Refinement. The equations specific to the AMBER force field are outlined here:

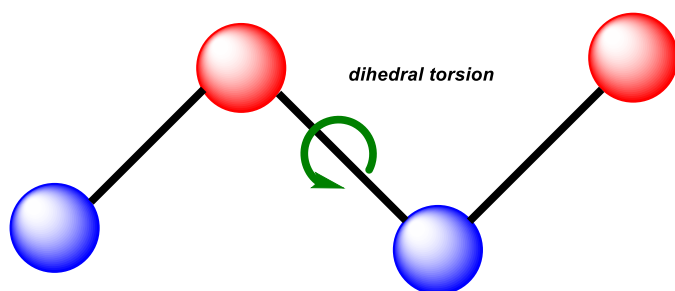
$$E = \Sigma_{bonds} E_{stretch} + \Sigma_{angles} E_{bend} + \Sigma_{dihedrals} E_{torsion} + \Sigma_{pairs} E_{nonbond} \quad \text{Equation 4.3}$$



$$\Delta E_{stretch} = k_{stretch}(l - l_{eqm})^2 \quad \text{Equation 4.4}$$

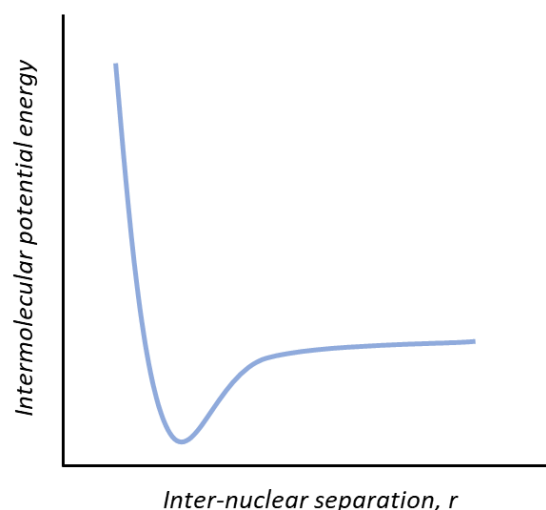


$$\Delta E_{bend} = k_{bend}(\theta - \theta_{eqm})^2 \quad \text{Equation 4.5}$$



$$\Delta E_{torsion} = k_{eqm} + \Sigma_{r=1}^n k_r [1 + \cos(r\theta)] \quad \text{Equation 4.6}$$

### Van der Waals (Lennard-Jones potential)



**Figure 4.5.** Lennard-Jones potential.

$$\Delta E_{nonbond} = k_{nonbond} \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right] \quad \text{Equation 4.7}$$



$$\Delta E_{nonbond} = \frac{q_1 q_2}{\epsilon r_{12}} \quad \text{Equation 4.8}$$

The Lennard-Jones potential considers the London dispersion forces that govern the complexes' intermolecular interactions.

#### Empirical input

Of course, without experimental data to refine it, the MD calculated model would only be theoretical. The placement of NMR-restraints on the inter-atomic distances allow for a more accurate depiction of the true system. This can be achieved by the integration of NOE cross peaks from the NMR NOESY spectra at different mixing times. There exists an inverse relationship between the NOE signal intensity  $\sigma$  and the inter-nuclear distance  $r$ .

$$\sigma \propto \frac{1}{r^6} \quad \text{Equation 4.9}$$

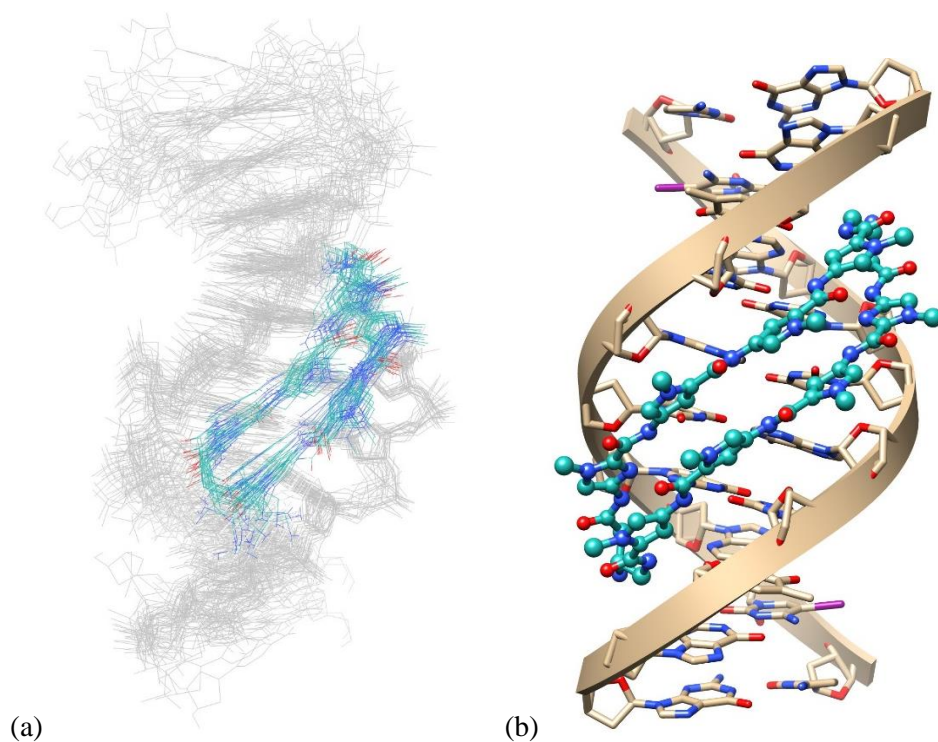
Due to perturbation from the molecular environment, as well as spin diffusion effects, the NOESY intensities can be unreliable in the complexed ligand-DNA system based on the isolated spin pair approximation (ISPA). To mitigate these errors, the use of MARDIGRAS, or Matrix Analysis of Relaxation for Discerning the Geometry of an Aqueous Structure is required.<sup>[81]</sup> This software algorithm takes the initial pre-refined model and calculates a matrix of mixing coefficients, which is then refined with the experimental intensities, and the process re-iterated to reduce the error until experimental data agrees with the theoretical model. The output from the MARDIGRAS algorithm is a set of proton-proton distances that can then be used as restraints for the AMBER MD production runs. This approach requires some prior knowledge of the system, however, such as a starting X-ray crystal structure or using a computationally-derived model.<sup>[79,81]</sup>

#### **4.2.2 Previous structural work on polyamide-dsDNA complexes**

The NMR structure of a 6-ring hairpin polyamide in complex with dsDNA was first published in a landmark paper by David Wemmer, where the titration of a polyamide to the DNA duplex provided evidence that the polyamide did indeed bind in the hairpin conformation based on the NOE cross-peak correlations.<sup>[82]</sup> Furthermore, the study revealed tremendous insight on the binding of the imidazole N2 to the exocyclic amine, with the observation of a substantial chemical shift of the proton not involved in the Watson-Crick base pair.<sup>[82]</sup>

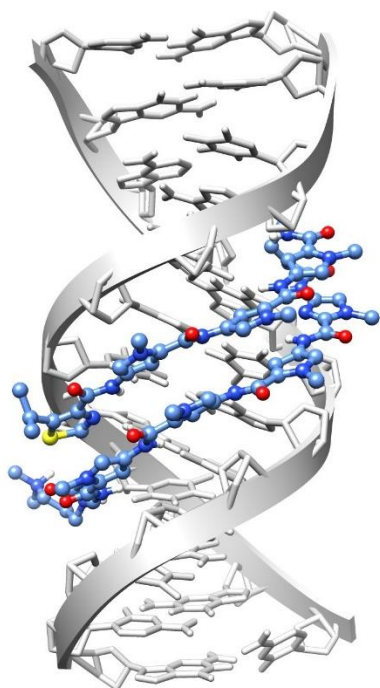
Wemmer and Zhang *et al.* pioneered the effort for structural determination of a 6-ring cyclic polyamide•dsDNA complex by NMR-restrained molecular dynamics.<sup>[76]</sup> The NMR experiments of that work demonstrated consistency with the then-understood binding site of the polyamides, with good Van der Waals contacts and strong hydrogen bonding between the polyamide and nucleobases of the DNA. That work outlined the protocol for determination of the structure from NOESY data.<sup>[76]</sup> It also hinted at some structural distortion of the DNA in the binding site, later verified by the crystal structure of an 8-ring cyclic polyamide•dsDNA complex determined by Chenoweth *et al.* (fig 4.6). The effect of polyamide binding on DNA distortion was manifested in a widening of the minor groove and a compression of the major groove.<sup>[21,75]</sup>





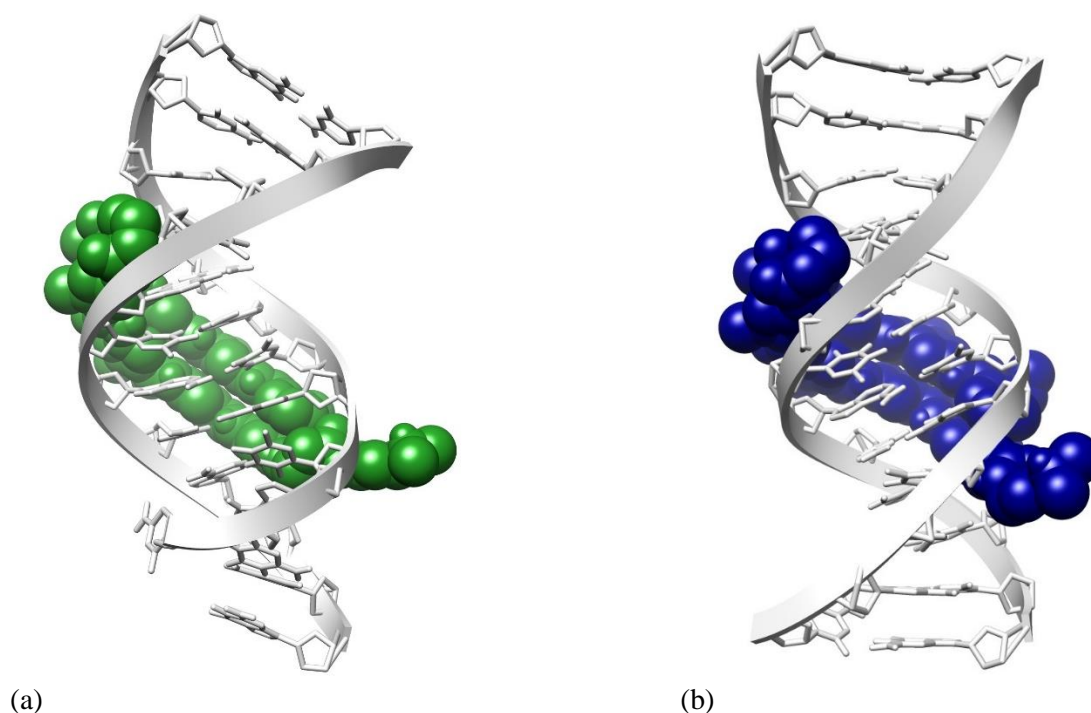
**Figure 4.6.** (a) Ensemble of cluster averages from NMR-restrained MD simulation of a 6-ring cyclic PA•dsDNA and (b) crystal structure of an 8-ring cyclic PA•dsDNA complex (PD ID 1PQQ and 3OMJ, respectively; structures are reproduced with UCSF Chimera ver 1.12 using coordinates of the respective PDB file).<sup>[21,76]</sup>

More recently, Padroni *et al.* determined the NMR structure of several 8-ring hairpin polyamides in complex with dsDNA, accomplished with NOE-restrained molecular dynamics simulations with AMBER.<sup>[2]</sup> The polyamides **PA1-3** were studied and the resulting structures were compared to the structure of the free ODN4 duplex.



**Figure 4.7.** PA3•ODN4 minimized average structure from minimized average clusters derived from NMR-restrained MD simulations (PDB ID 5ODM).<sup>[2]</sup>

That structural study shed light on the influence of the N-terminal monomer unit, where the N-alkyl thiazoles were found to induce a greater compression of the major groove by a ‘pinching’ of the duplex at the target site 5’-ATGTACA-3’. Furthermore, the N-terminal isopropyl thiazole analogue was found to induce even greater compression relative to the N-terminal methyl thiazole, indicating that the influence of a seemingly subtle change in steric bulk facing away from the minor groove could affect the binding interaction significantly.<sup>[2]</sup>



**Figure 4.8.** Major groove compression by (a) N-methylimidazole and (b) N-isopropylthiazole polyamides (PDB ID 5OE1, 5ODM for **PA1•ODN4** and **PA3•ODN4**, respectively).<sup>[2]</sup>

To enable a direct comparison to previously published structures, and since the N-terminal position of the 8-ring hairpin polyamide is where the novel 'Pr-Im monomer was installed, the oligodeoxynucleotide dodecamer sequence 5'-CGATGTACATCG-3' was chosen as the target substrate to which PA4 would be bound.

### 4.3 Aim and objectives

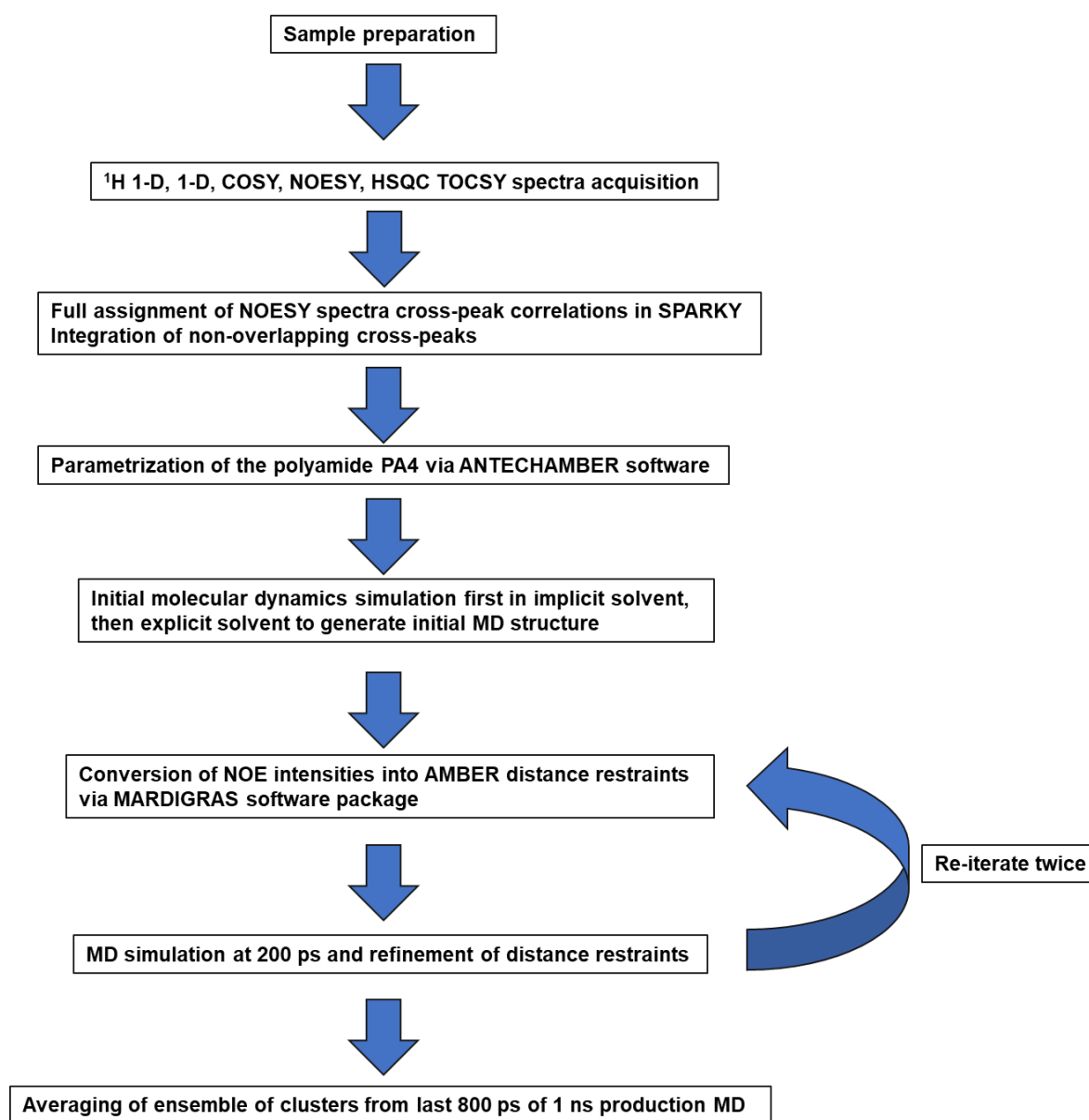
This chapter details the NMR structural determination of the **PA4•ODN4** complex. A description of the sample preparation and the acquisition parameters for the NOESY data set is detailed along with the assignment protocol of the NOE cross-peak correlations. The protocol for NMR-restrained molecular dynamics simulations is also described and the calculated DNA structural parameters are discussed.

## 4.4 NMR structure results and discussion

### 4.4.1 Workflow strategy for NMR-restrained MD simulations

The synthesized hairpin polyamide PA4 previously evaluated for its binding affinity and its selectivity as discussed in chapter 3 indicated it was a reasonably capable ligand for a structural study. To obtain

further insight, the protocol followed for the structural determination of the **PA4•ODN4** complex, based on the work of Zhang *et al.* and Padroni *et al.*, is outlined in fig. 4.9.<sup>[2,76]</sup>



**Figure 4.9.** Workflow for structural determination of PA4•ODN4 complex.

#### 4.4.2 Sample preparation

In preparation for the collection of accurate solution NMR data of biomolecules, and in particular DNA, there are several concerns for chemists that need to be carefully considered. The first is maintenance of an accurate pH, and when handling DNA, the buffering of the pH between 7.0-7.4 is vital for ensuring a stable double-helix characteristic of B-DNA. This also enables the observation of exchangeable proton

resonances (OH, NH) when H<sub>2</sub>O is used as the primary solvent. Phosphate buffer is typically used for this purpose. The dependence on data requirements dictates the solvent to use for the samples. 90% H<sub>2</sub>O /10% D<sub>2</sub>O is used as the primary solvent for obtaining information on exchangeable protons, such as the imino resonances that appear in the 12-14 ppm region of the spectrum (arising from Watson-Crick base-pairing).<sup>[2]</sup>

The solvent choice is also dictated by the hardware present in the set up. The capability for pulsed-field gradients suggests the use of 90% H<sub>2</sub>O / 10% D<sub>2</sub>O for observation of the imino resonances and confirming the assembly of the DNA duplex structure.

Use of 99% D<sub>2</sub>O affords several advantages and this is the solvent in which the 2D experiments should be run. The advantages include among them a reduction in complexity of <sup>1</sup>H NMR data, along with higher digital resolution of 2D and higher dimension NMR experiments.

The ability to lyophilize samples of DNA and re-dissolve in a different solvent while maintaining the integrity of the samples supports the practicality of collecting multiple sets of data, each of which can aid the analysis by providing additional information.

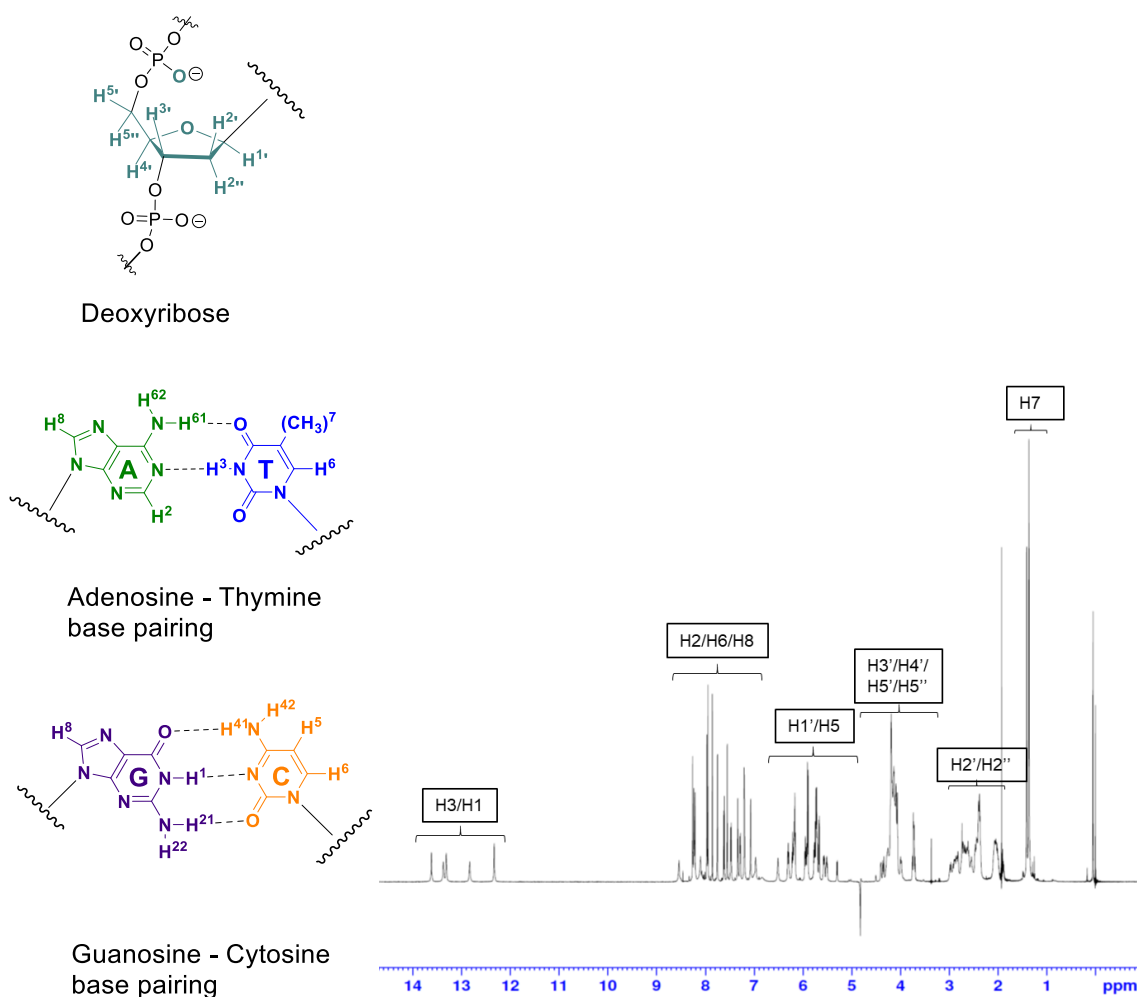
### *Resonance Assignment*

Due to the large number of protons in even a dodecamer sequence of dsDNA, there is an expected overlapping of a high percentage of the peaks in a standard 1D <sup>1</sup>H NMR experiment. This makes the confident assignment of the signals a near-impossible task with a mere analysis of the spectrum alone, necessitating the need for 2D NMR experiments, including [<sup>1</sup>H-<sup>1</sup>H]-NOESY, COSY, HSQC, and TOCSY.

The <sup>1</sup>H NMR signals of DNA fall into distinct regions, characterized by type of proton (table 4.1 on next page).

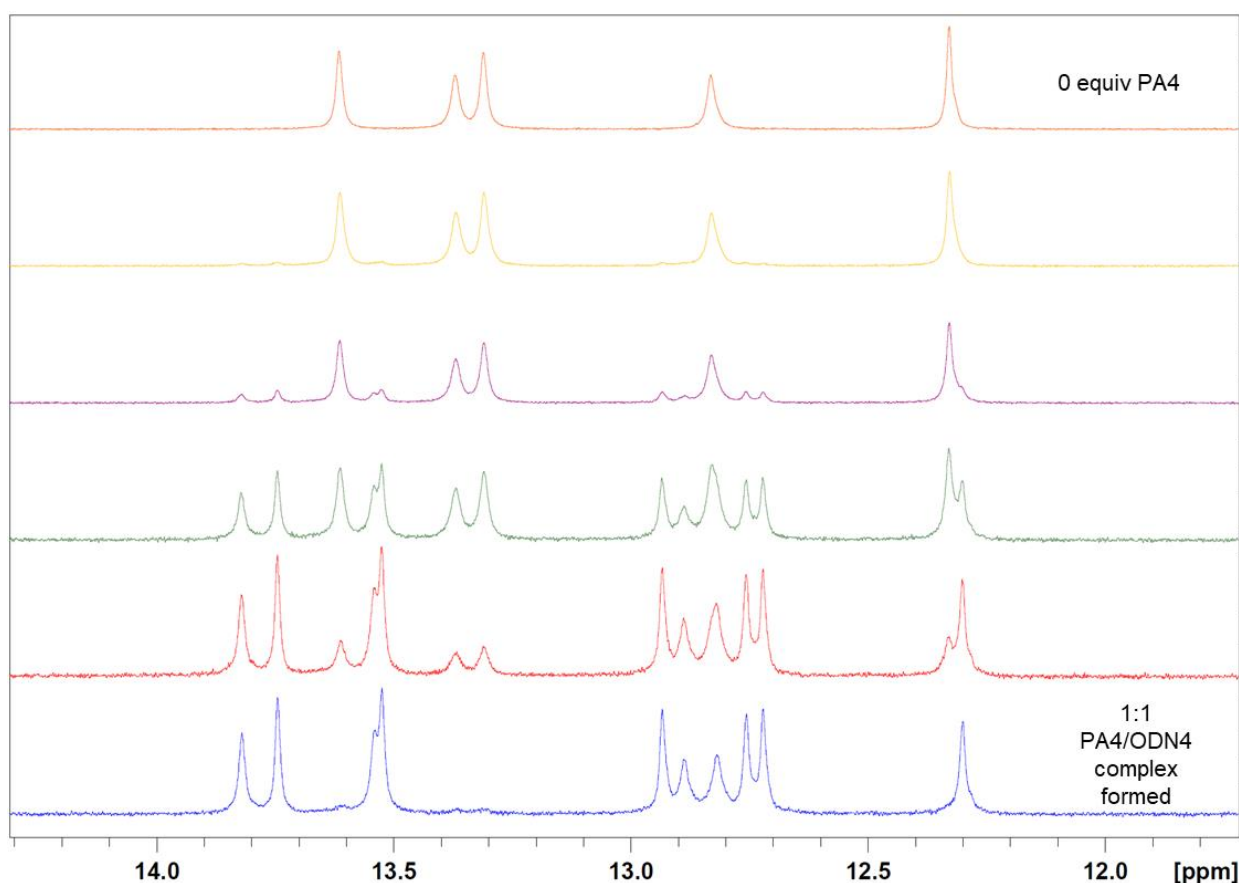
**Table 4.1.** Chemical shift regions of typical DNA protons.

<i>DNA proton</i>	<i>Chemical shift region (ppm)</i>
Imino of nucleobases	12 - 14
Aromatic nucleobases	7 - 8.5
Cytosine H5	5.5 - 6
Sugar (deoxyribose)	2 - 6



**Figure 4.10.** Numbering designation for DNA peaks (left), with the  $^1\text{H}$  NMR spectrum of ODN4 shown to the right.

For this work, a sample of HPLC purified self-complementary  $d(\text{CGATGTACATCG})_2$  (**ODN4**) was purchased from Eurogentec and dissolved in 600  $\mu\text{L}$  of  $\text{H}_2\text{O}$  buffered with 100 mM phosphate at pH = 7.4 containing 0.1 mM deuterated trimethylsilylpropionic acid sodium salt- $d_4$  and lyophilized. The sample was re-dissolved in 600  $\mu\text{L}$  of 90%  $\text{H}_2\text{O}$  with 10%  $\text{D}_2\text{O}$ . A 5 mm high-precision NMR tube was used for ODN4 data acquisition. A concentrated solution of **PA4** (15 mM in Milli-Q water) was titrated into the ODN4 solution in small aliquots. After each addition of polyamide to the buffered solution containing **ODN4**, 1D  $^1\text{H}$  NMR spectra were acquired until the end point of the titration was reached as determined by inspection of the imino proton chemical shift region. After acquisition of the NOESY spectrum in 10%  $\text{D}_2\text{O}$  / 90%  $\text{H}_2\text{O}$ , the sample was lyophilized and re-dissolved in 99%  $\text{D}_2\text{O}$  to complete the acquisition of the full data set (HSQC, TOCSY, and NOESY at 4 mixing times).<sup>[2]</sup>



**Figure 4.11.** NMR titration of **PA4** aqueous solution (15 mM) into **ODN4** (1.5 mM solution in 90%  $\text{H}_2\text{O}$ /10%  $\text{D}_2\text{O}$ , with pH 7.4 phosphate buffer).

The 1D  $^1\text{H}$  NMR spectrum provides an initial indication of binding taking place during the titration, where certain imino resonances present in the dsDNA duplex become shifted, as shown in fig. 4.11.

The distinct region of the imino resonances is indicative that a binding event has taken place, with the shift of peaks evident upon titration of the polyamide. This ensured the formation of a 1:1 (PA:dsDNA) complex. Fig. 4.20(b) traces the region of **PA4** binding by indicating that the most significantly shifted protons of the complex reside in the central portion of the **ODN4** sequence, 5'-G<sub>5</sub>T<sub>6</sub>A<sub>7</sub>C<sub>8</sub>-3'.

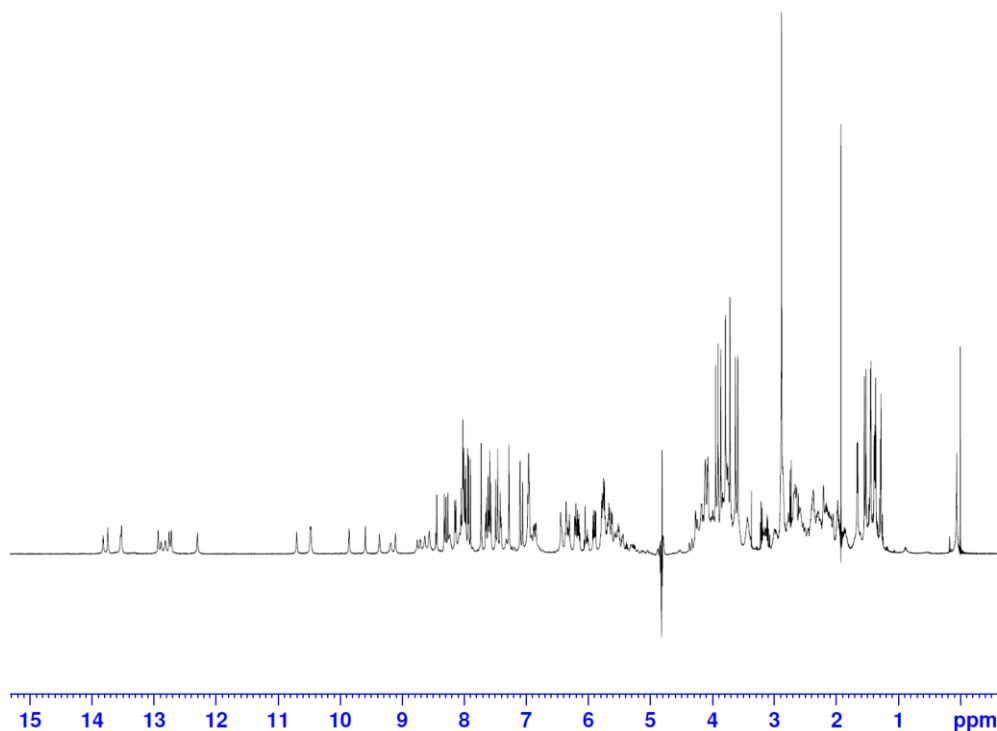
#### 4.4.3 Acquisition parameters

NMR data for polyamide-DNA complex was collected on a Bruker AVANCE-II<sup>+</sup> NMR spectrometer operating at a magnetic field of 14.1 T (600.13 MHz for <sup>1</sup>H resonance). Software for running the NMR experiment, including data collection, was TopSpin (version 3.5 patch level 5, Bruker, Reinstetten, Germany) running on a Hewlett Packard Z420 workstation under Windows Professional version 7 (Microsoft Inc.). A triple-resonance probe head [TBI-z] equipped with actively shielded z-gradient coil for delivery of pulsed field gradients was manually tuned for a [<sup>1</sup>H, <sup>13</sup>C, and <sup>31</sup>P] frequency configuration on each sample studied to allow uninterrupted data acquisition to progress using all channels in their turn. The probe temperature was maintained at 298 K in all instances. One-dimensional (1D) <sup>1</sup>H NMR data were acquired using either pre-saturation (pulse program zggppr) or double pulsed field gradient spin-echo (pulse program zgesgp) to eliminate the residual solvent resonance. Data were typically acquired for 90% H<sub>2</sub>O samples with between 64 and 256 transients into 32K data points (acquisition time: 1.09 s) over a frequency width equivalent to 20.0276 ppm centered at  $\delta^1\text{H} = 4.692$  ppm. Hard proton 90° pulses were typically calibrated at 9.6W with an average value of 9.7  $\mu\text{s}$  across the samples studied. Data acquired for D<sub>2</sub>O-dissolved samples were acquired over a frequency width equivalent to 12.0166 ppm (acquisition time: 2.27 s).

Two-dimensional (2D) NMR data sets were acquired as follows for samples dissolved in D<sub>2</sub>O. Proton-only data were acquired over frequency widths equivalent to 12 ppm (7194 Hz) in both F2 and F1 centered at an offset of  $\delta^1\text{H} = 4.702$  ppm. 2D DQF-COSY data (pulse program: cosydfphpr) were acquired with 16 transients for each of 1024 States - TPPI t1 increments for acquisition times (aq) of 285 ms ( $\omega_2$ ) and aqmax = 71 ms ( $\omega_1$ ) and a total data accumulation time of 10 hrs. 2D TOCSY data (pulse program: dipsi2phpr) were acquired into 2048 data points ( $\omega_2$  acquisition time = 142 ms) with eight transients for each of 512 States - TPPI t1 increments (aqmax  $\omega_1 = 35.4$  ms) using a spin-lock



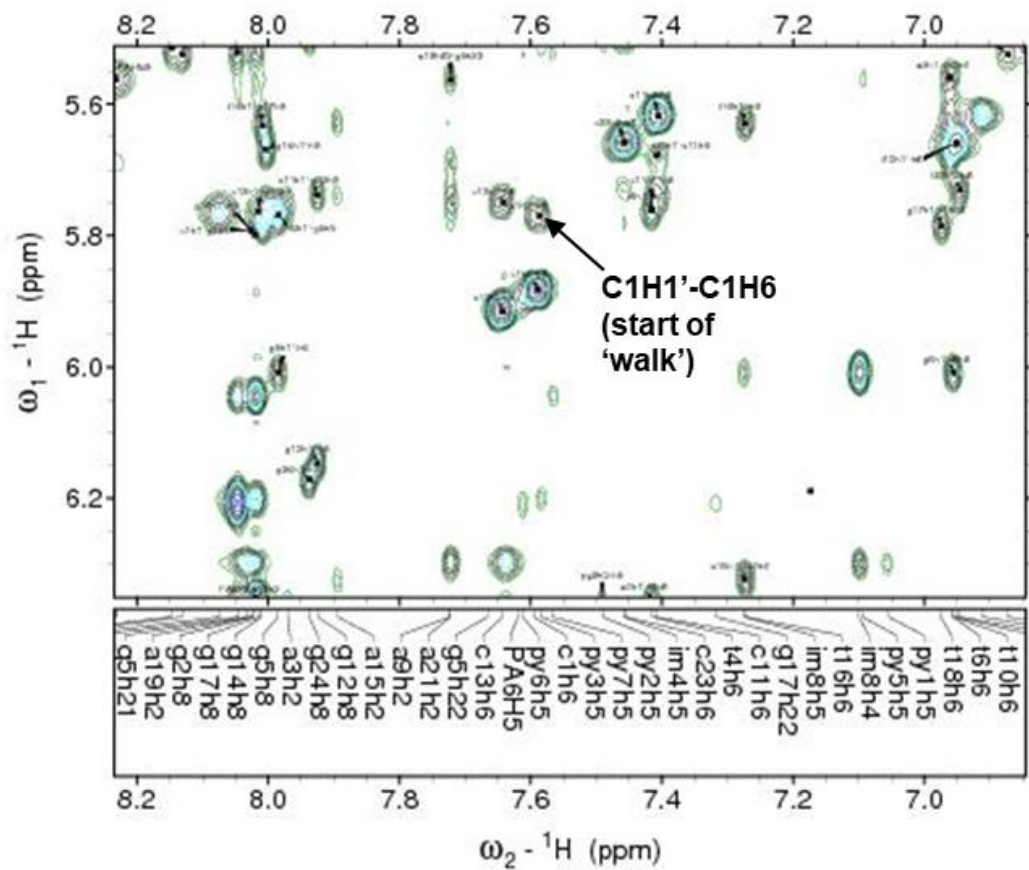
time of 70 ms. The same conditions were used to acquire 2D NOESY data (pulse program: noesyphpr) with mixing times ranging from 100 ms to 250 ms (four mixing times total). 2D [<sup>1</sup>H, <sup>13</sup>C] correlation data were acquired using echo/anti-echo-TPPI data acquisition schemes with sensitivity improvement, with (pulse program: hsqcedetgpsisp. 2) and without (pulse program: hsqcetgpsisp.2) DEPT-editing and without/with non-uniform sampling (NUS), respectively. NUS data were acquired with 25% sampling of 512 t1 increments into 2048 data points for frequency widths equivalent to 10 ppm (F2) and 170 ppm (F1) with 128 transients per t1 increment for a total data accumulation time of 10 h. DEPT-edited HSQC data were acquired with a traditional acquisition mode using the same conditions and 64 transients per t1 increment for a total data accumulation time of 20 hrs. For samples dissolved in 90% H<sub>2</sub>O, 2D NOESY data were acquired with excitation sculpting for solvent suppression (pulse program: noesyegpph) using State-TPPI for 1024 t1 increments and 4096 data points over F2 and F1 frequency widths equivalent to 25 ppm centered at the solvent frequency resonance. Data were acquired with 16 transients per t1 increment with a relaxation delay of 7 s between transients for a total data accumulation time of 33 h. Mixing times were defined as 180 and 250 ms. All data were processed according to established NMR data processing protocols and transferred into SPARKY software for data assignment and reduction in preparation for molecular structure calculations.



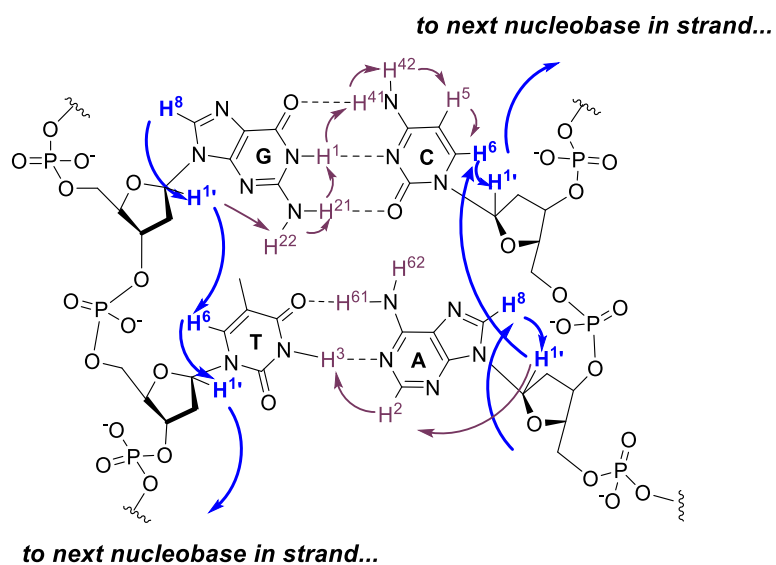
**Figure 4.12.** 1D  $^1\text{H}$  NMR spectrum for **PA4•ODN4** complex in 90 %  $\text{H}_2\text{O}$  / 10%  $\text{D}_2\text{O}$ .

#### 4.4.4 Analysis of the NOESY spectra with SPARKY

To enable the utilization of the spectral intensities from 2D NOESY data, the assignment of as many correlations as possible is required. This is a challenging task, owing to the potential overlap of a significant proportion of the signals, as well as the potential of incorrect assignments. An effective strategy was developed and implemented for this purpose, where the NOESY contacts of the DNA nucleobases were assigned first, starting with the exchangeable protons (in the 90%  $\text{H}_2\text{O}$ /10%  $\text{D}_2\text{O}$  spectrum). As a starting point, the first nucleotide in the sequence is cytosine C1 which displays a distinct and identifiable C1H1'-C1H6 cross-peak correlation, which served as the starting point for 'walking' of the strands.

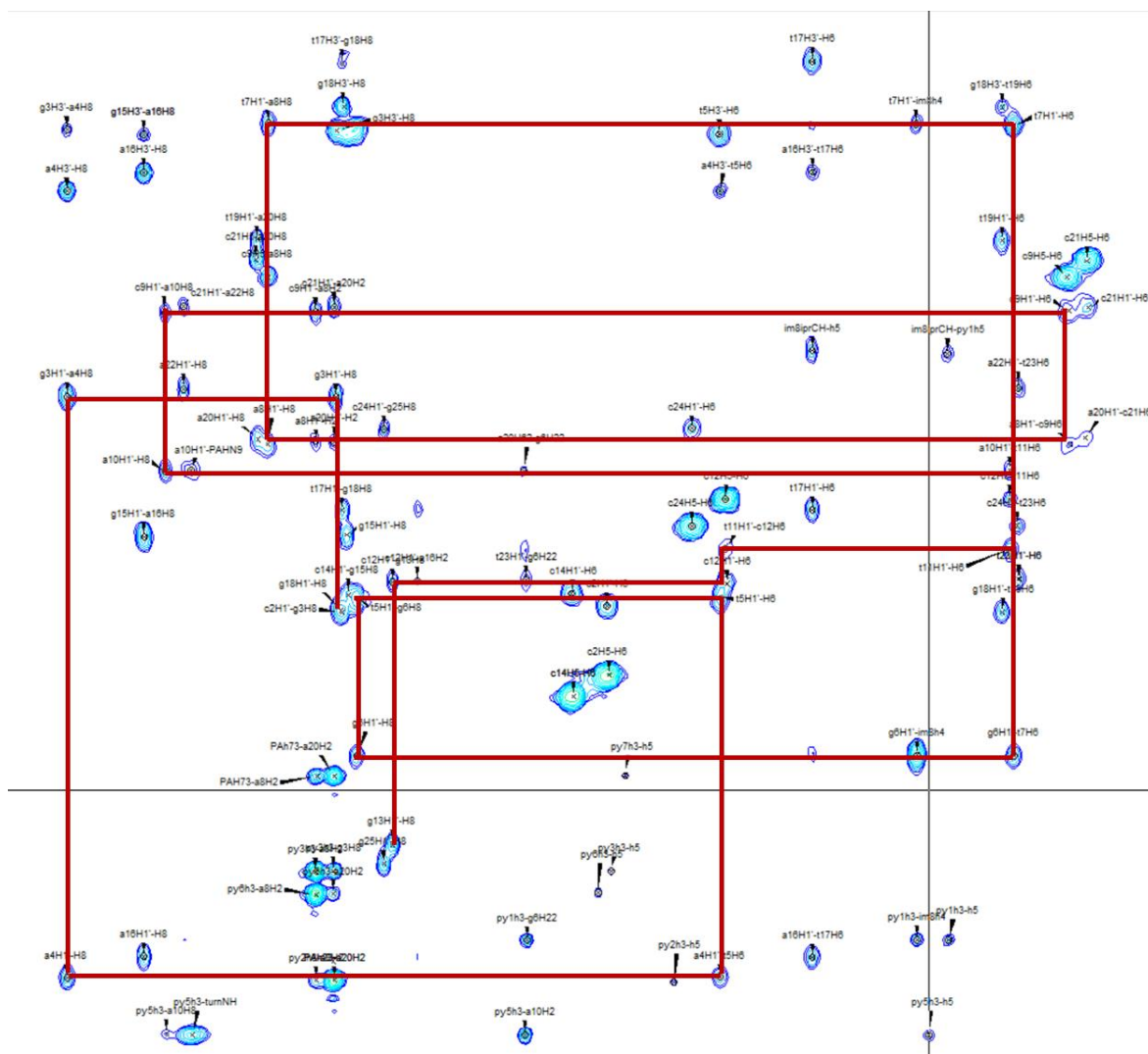


**Figure 4.13.** Zoom-in of NOESY spectral region for PA4•ODN4 complex showing the cytosine C1 correlations between base and sugar H1' protons (clearly defined and separate from other signals) for beginning the DNA 'walk' (250 ms mixing time in 90% H<sub>2</sub>O / 10% D<sub>2</sub>O).



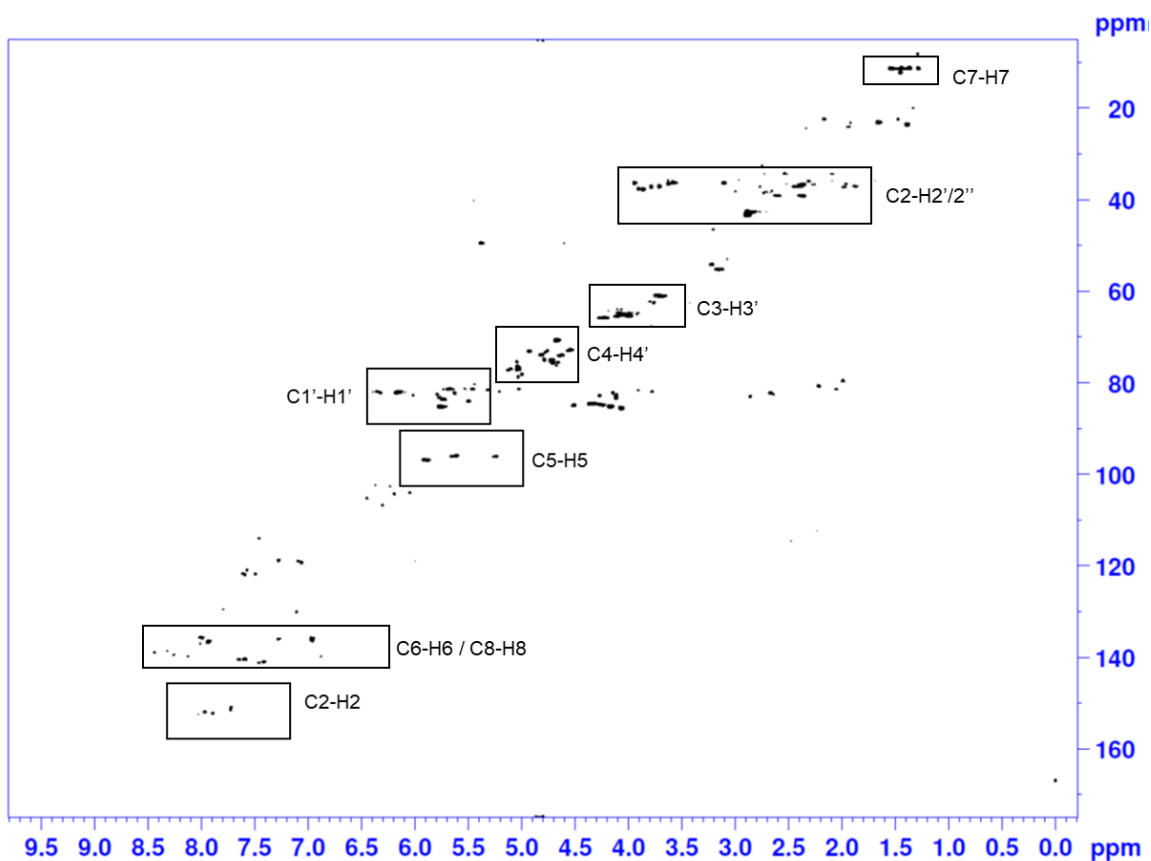
**Figure 4.14.** NOE assignment strategy: starting with the sugar H1' to nucleobase protons.

The 'strand-walking' strategy depicted in fig. 4.14 allows for the starting point of the assignments and from there the 2D NOESY correlations of most nucleobase protons were ascertained (C1 to G12 'walk' shown in fig. 4.15 below).



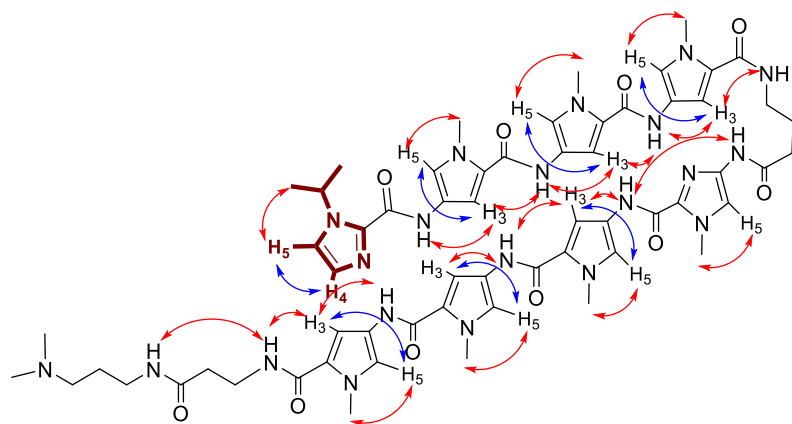
**Figure 4.15.** NOE ‘walk’ of strand C1-G12 for PA4•ODN4 complex obtained from NOESY spectrum at 250 ms in 90% H<sub>2</sub>O/10% D<sub>2</sub>O.

With the majority of nucleobase protons assigned, the next stage was the assignment of the deoxyribose protons of the DNA backbone. These assignments were made more confidently with the help of the [<sup>1</sup>H-<sup>13</sup>C] HSQC data set, wherein the different sugar carbons are located in distinct regions of the 2D map.

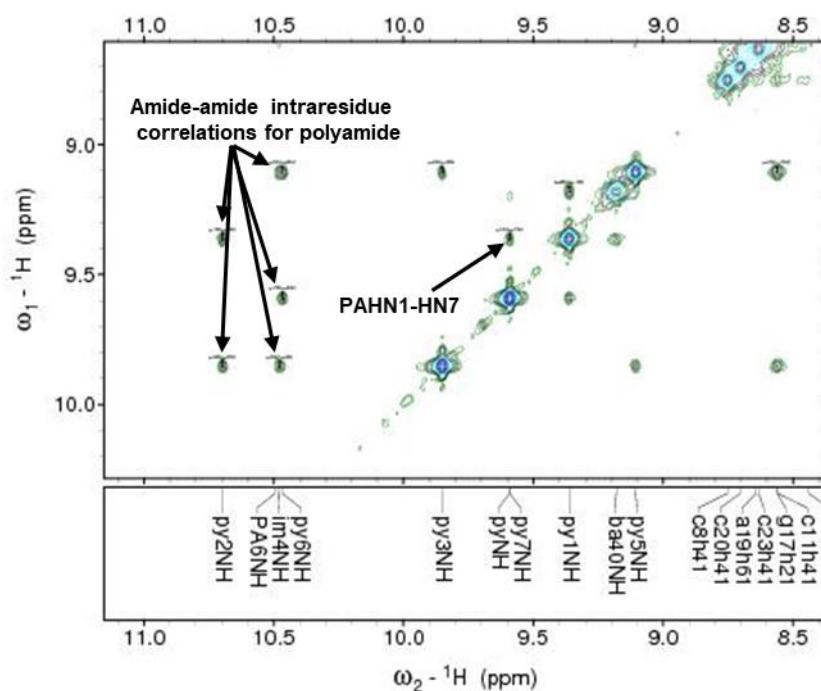


**Figure 4.16.** [ $^1\text{H}$ - $^{13}\text{C}$ ] HSQC spectrum showing the distinct DNA deoxyribose regions of **PA4•ODN1** complex to support the assignment of the skeleton.

After the assignment of the sugar protons in the backbone, the polyamide was separately assigned walking along the residues starting from the N-terminal position. This was aided with the assignment of the TOCSY data set, which provided correlations between the H3-H5 protons of the pyrrole units. Fortunately, there were the very clearly distinguished protons of the isopropyl group, and they served as a convenient starting point.

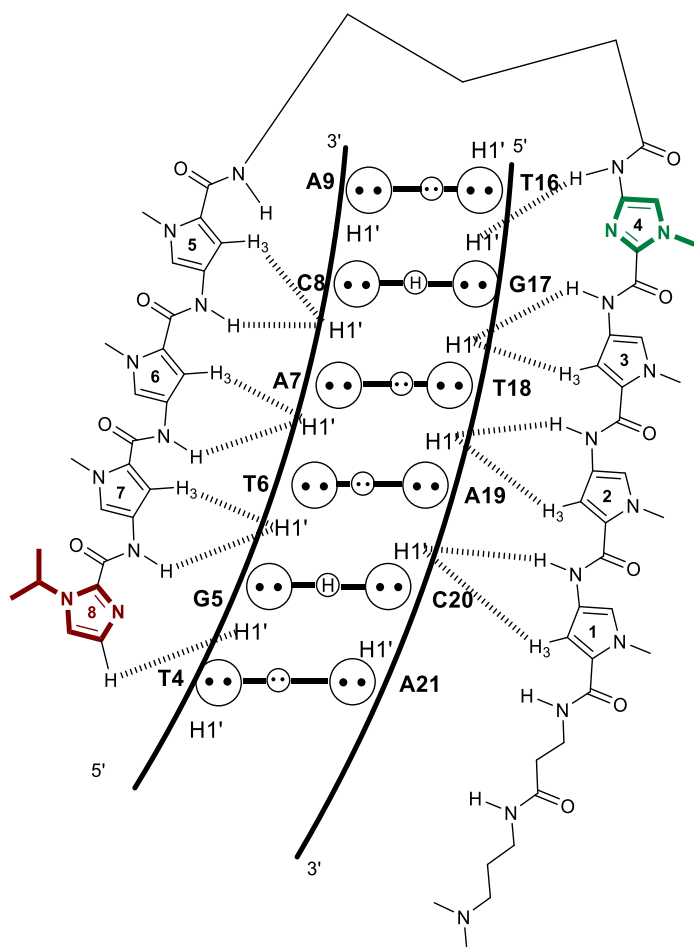


**Figure 4.17.**  $^1\text{H}$  NMR 2D NOESY and TOCSY correlations used for PA4 proton assignment. Red arrows indicate observed NOESY cross correlations; blue arrows indicate both NOESY and TOCSY observed correlations.

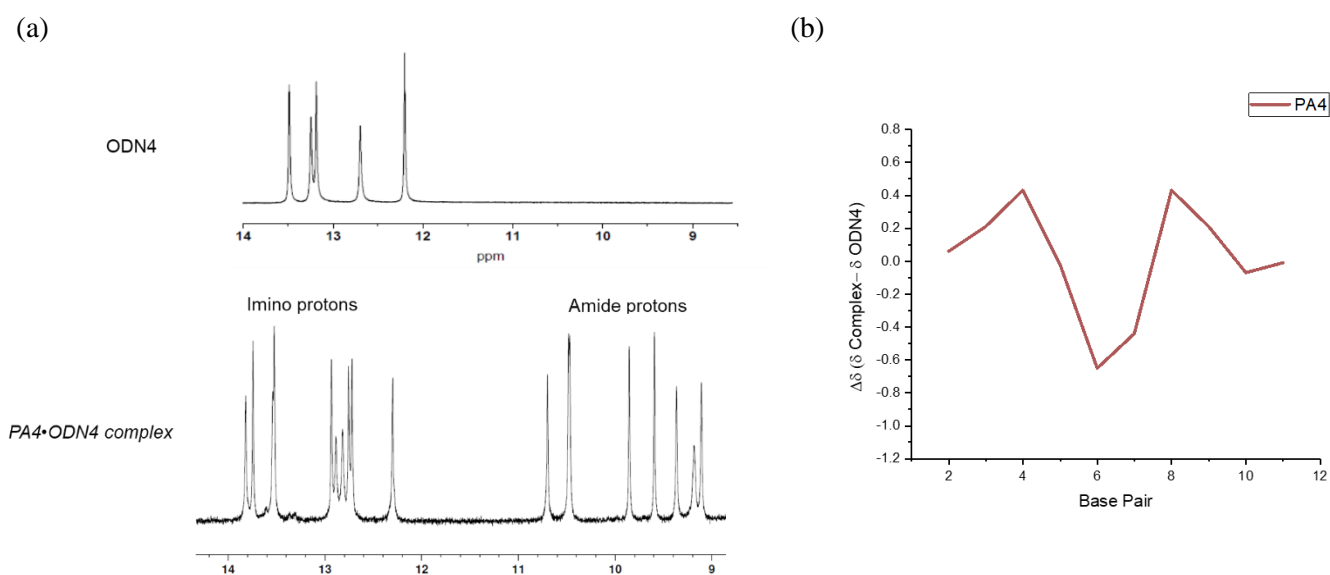


**Figure 4.18.** PA4 intra-amide correlations present in PA4•ODN4 complex ('walk' of polyamide starts here; 250 ms mixing time in 90%  $\text{H}_2\text{O}$  / 10%  $\text{D}_2\text{O}$ ).

With the assignment of the majority of the protons of the DNA and the polyamide aromatic/alkyl shifts as separate compounds, the next stage was the polyamide-DNA inter-residue correlation assignments. The methyl protons of the polyamide were quite helpful, as they were relatively easy to decipher from the NOESY dataset and in close proximity with both the pyrrole/imidazole aromatic protons as well as the sugar protons of the nucleotides.



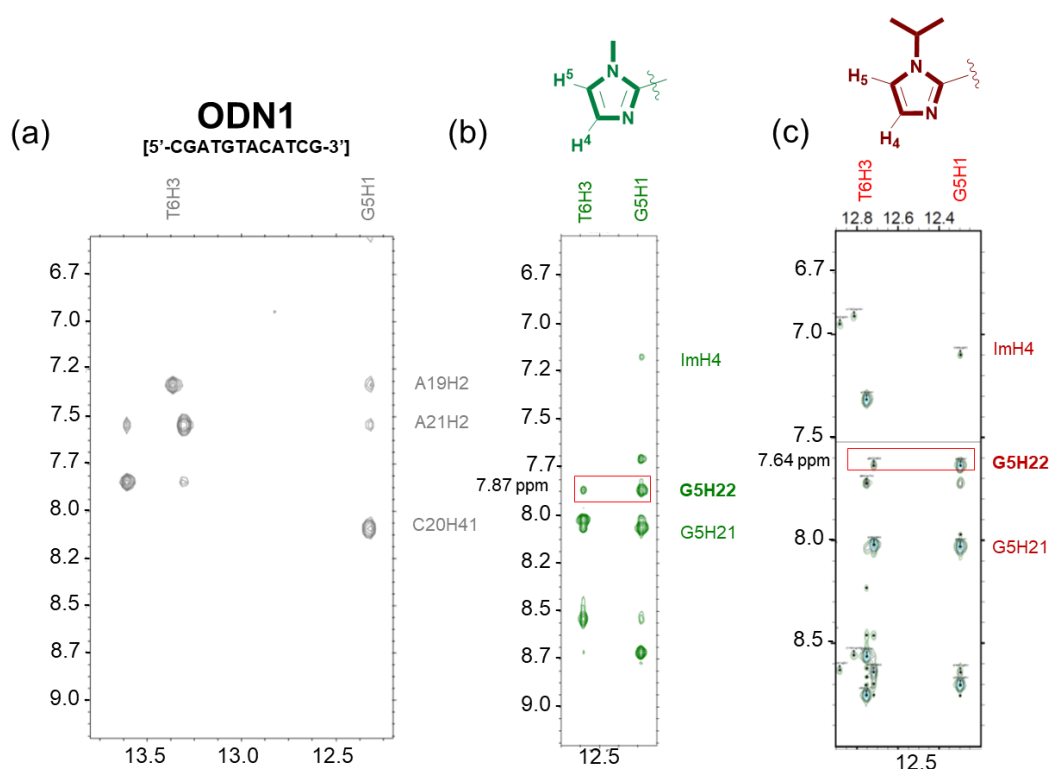
**Figure 4.19.** Schematic depicting the inter-residue NOE correlations observed for the PA4•ODN4 complex.



**Figure 4.20.** (a) 1D  $^1\text{H}$ -NMR imino-amide chemical shift region of free ODN4 and PA4•ODN4 complex (b) chemical shift perturbation plot of imino resonances.



Studying the G5N2 proton not involved in base-pairing, the chemical shift conveys important information about the strength of the hydrogen bond as it relates to the presence of the methyl or isopropyl functionality. **PA4•ODN4** has a less deshielded value of 7.64 ppm, compared with the 7.87 ppm shift for **PA1•ODN4** (fig. 4.21). This suggests that the hydrogen bond is slightly weaker in the *i*Pr-Im analogue, but substantially stronger than the thiazole analogues (**PA2-3**), which displayed a chemical shift of 6.76 ppm and 6.59 ppm for the methyl thiazole and isopropyl thiazole N-terminal units, respectively.<sup>[2]</sup>



**Figure 4.21.** NMR strip plots of 2D <sup>1</sup>H NOESY spectra of (a) **ODN4**, (b) **PA1•ODN4** and (c) **PA4•ODN4**, showing key NOE correlations/shifts (strip plots (a) and (b) are reproduced here from G. Padroni *et al.* for a comparative analysis with PA4<sup>[2]</sup>).

#### 4.4.5 MD simulations

Analysis of the NOE data provided intensities for key correlations that were transformed into distance restraints through cycles of MARDIGRAS, refining the AMBER MD model for the **PA4•ODN4** complex. The statistical output of the production MD run is shown below in table 4.2.

**Table 4.2.** Statistics and output of NMR restraints used for MD model determination of **PA4•ODN4**.

Entry	Total Restrains Applied	DNA	Polyamide	Inter- residue	Rmsd* (bond)	Rmsd* (angle)	Penalty* (Kcal)
<b>PA4•ODN4</b>	310	181	77	52	0.0073	4.439	280.86

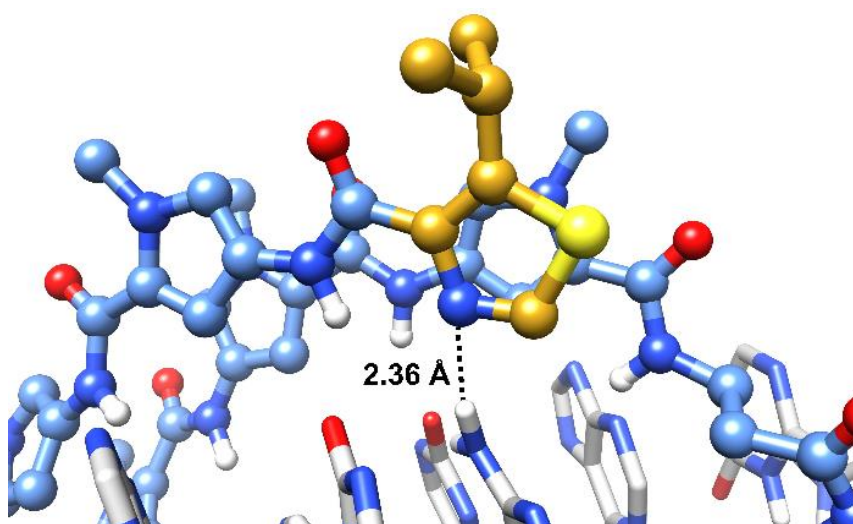
\*output values from the last frame of 1 ns MD simulation



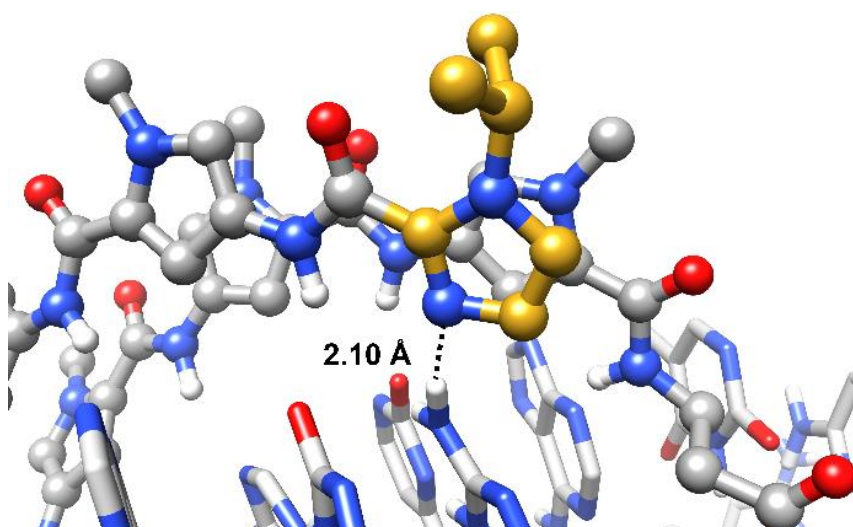
**Figure 4.22.** Ensemble of conformations of **PA4•ODN4** complex obtained from clustering of the last 800 ps of the 1 ns molecular dynamics production run.

After 1 ns MD production simulation, the average ensemble MD structure comprised of the average clusters (fig. 4.22) of the **PA4•ODN4** complex indicated that the expected nucleotide region of the DNA, 5'-GTAC-3', was where the expected ring pairings were aligned. The proximity of the hydrogen of the G5N2 exocyclic amine to the <sup>i</sup>Pr-Im N3, which has a free lone pair of electrons available for hydrogen bonding, was similar to the **PA1•ODN4** complex, with values of 2.10 Å and 2.01 Å, respectively (fig. 4.23). This is in contrast to the **PA3•ODN4** complex, where the distance under consideration here was 2.36 Å, significantly larger.

(a)



(b)

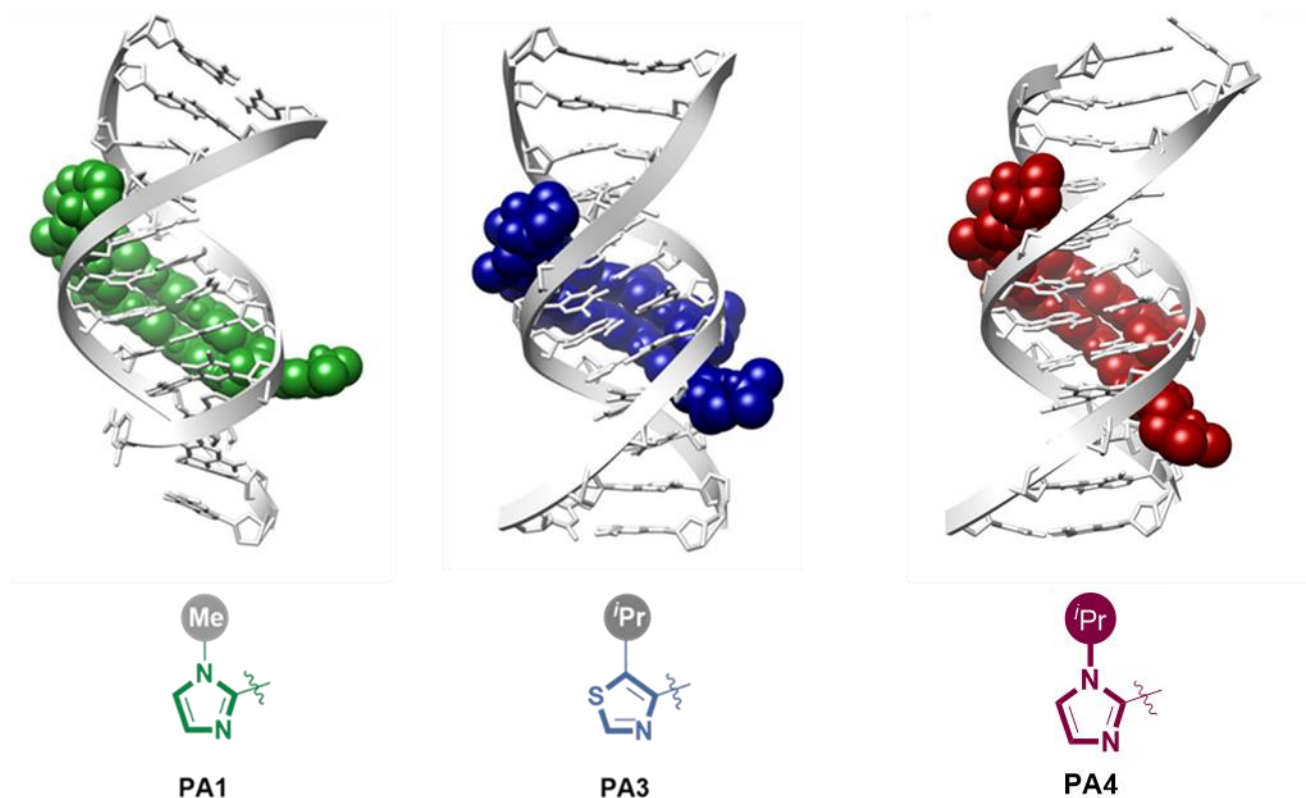


**Figure 4.23.** Proximity of the minor groove-facing (a) N3 of *iPr-Nt* and (b) N3 of *iPr-Im* to the exocyclic amine (N2) of guanine G5 of **ODN4** (structures were produced through UCSF Chimera software, using the average MD structure of the ensemble of cluster averages; **PA3**/*iPr-Nt* was reproduced through the software using the published structure from PDB ID 5ODM<sup>[2]</sup>).

This increased proximity for hydrogen bonding in **PA4**•**ODN4** compared with **PA3**•**ODN4** is likely to play a role in the enhanced selectivity observed at that N-terminal position. Whether this is due to the

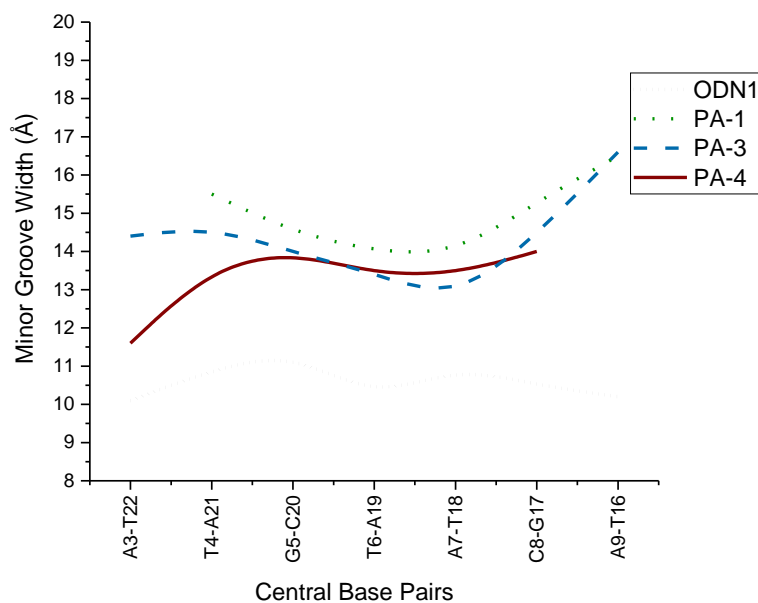
steric bulk of the sulfur atom prohibiting a strong hydrogen bond or a combination with the different electronic character inherent to the thiazole heterocycle itself is undetermined.

There was also significant structural distortion of the duplex due to a compression of the major groove by the steric bulk of the isopropyl. The compression was similar to that observed by the isopropyl thiazole analogue.

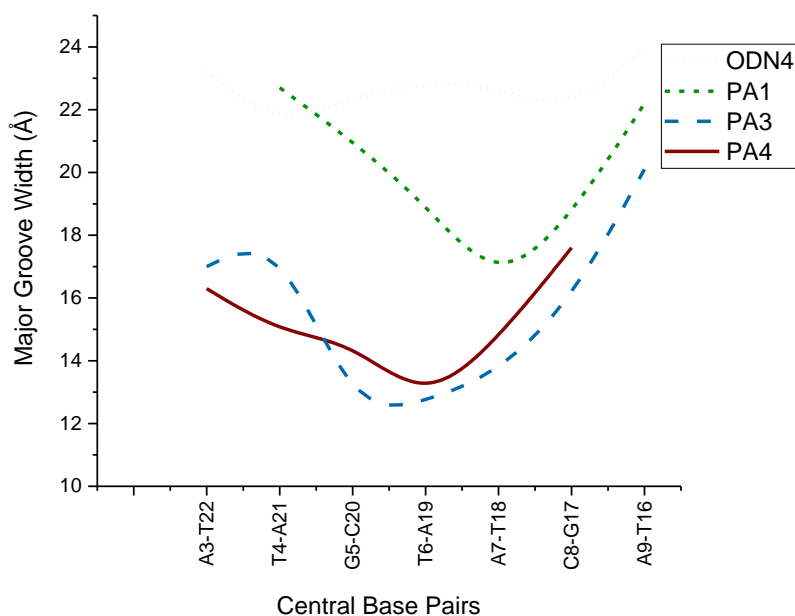


**Figure 4.24.** Major groove compression observed for **PA1**, **PA3**, and **PA4** in complex with **ODN4** (structure for PA4 was produced through UCSF Chimera software, using the average MD structure of the ensemble of cluster averages; **PA1/PA3** were reproduced through UCSF Chimera using the published structures from PDB ID 5OE1 for **PA1**, 5ODM for **PA3**<sup>[2]</sup>).

(a)



(b)

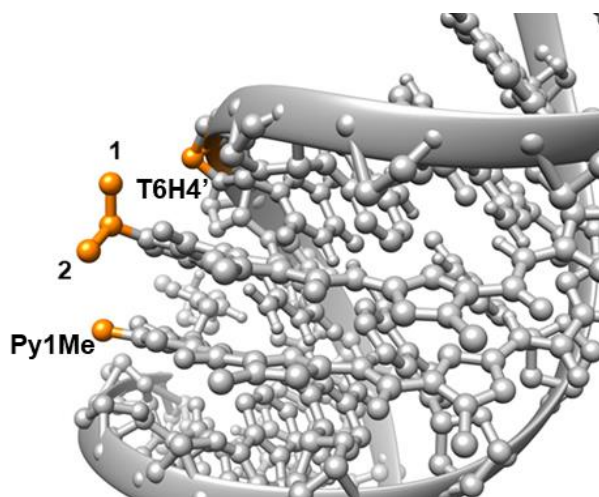


**Figure 4.25.** Major and minor groove width of **PA1•ODN4**, **PA3•ODN3**, and **PA4•ODN4** complexes (values were calculated through x3DNA using the average structure from the ensemble of average clusters of the last 800 ps of a 1 ns production MD simulation (values for **ODN4**, **PA1**, and **PA3** were reproduced here for a comparative analysis, and are taken from G. Padroni, *et al.*<sup>[2]</sup>).

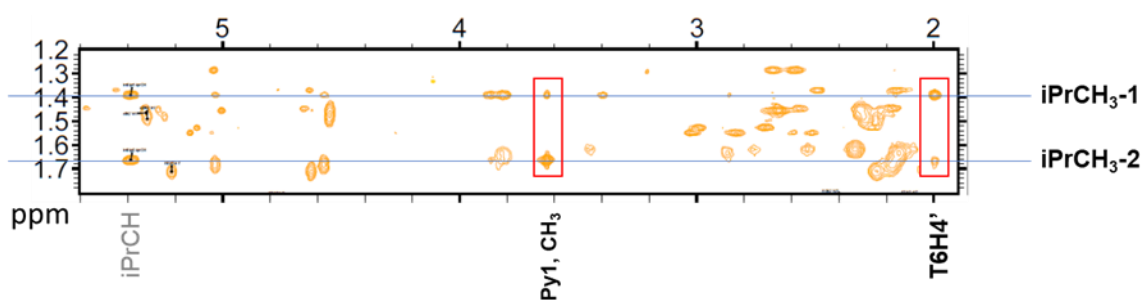
The structural distortion of the B-DNA duplex was greater with the 'Pr-Im unit compared with the methyl analogue, confirming the hypothesis that this subtle change in the substituent facing away from the minor groove can impart significant structural distortion. The major groove was more compressed

by the increased steric bulk from the isopropyl unit, which sits in an orientation that puts one methyl in close proximity to the Py1-Me and the other to the T6H4' of the DNA backbone (fig. 4.26).

(a)



(b)



**Figure 4.26.** (a) MD average structure of **PA4•ODN4** complex, showing the orientation of the isopropyl group in the minor groove; (b) key NOE cross-correlations providing evidence for this orientation (250 ms mixing time in 99% D<sub>2</sub>O).

## 4.5 Summary

This chapter detailed the 2D NMR analysis of the **PA4•ODN4** complex, utilizing the NOESY intensity signals as distance restraints for the 1 ns MD-simulation. The atomic coordinates used for the 3D modelling for analysis of parameters were obtained from averaging the resulting ensemble of 11 cluster averages. Close analysis of the average ensemble structure from the last 800 ps of a 1 ns production MD run indicated the nitrogen N3 of the <sup>3</sup>Pr-Im monomer unit has a comparable hydrogen bond strength to the Me-Im analogue. Furthermore, the steric bulk of the isopropyl substituent compressed the major groove of the B-DNA duplex ODN4. Overall, this structural study provided insight into the observed

binding affinity and selectivity outlined in chapter 3, and provides a framework for the detailed analysis of future PA•dsDNA complexes, perhaps those that may incorporate mutations to the sequence analyzed in this study. Lastly, the NMR-derived structure provides further evidence to support the hypothesis that the fine balance between steric bulk for duplex distortion and minor groove penetration depth is met with the 'Pr-Im in the N-terminal position of an 8-ring hairpin polyamide.

## 4.6 Experimental

### 4.6.1 Preliminary Molecular Dynamics

Preliminary molecular dynamics for distance restraint calculations of the **PA4•ODN4** complex was run in explicit solvent using the SANDER module of AmberTools16.<sup>[83]</sup> The starting model of **ODN4** was generated using the UCSF CHIMERA 1.12 program.<sup>[84]</sup> The initial structure of the **PA4•ODN4** complex was generated based on an X-ray-derived structure of a cyclic polyamide•dsDNA complex as the starting structure (PDB ID 3OMJ).<sup>[85]</sup> Input files (prmtop and inpcrd) were generated using the nucleic acid force field ff99bsc1 and GAFF for **ODN4** and ligands, respectively. 21 Na<sup>+</sup> ions for the complex were added to neutralize the negative charge of the phosphate groups. An octahedral box with outer edges of approximately 10.0 Å of TIP3P water was placed around the structures. A 10 Å cut-off for non-bonded interactions was applied with the Particle Mesh Ewald (PME) method to account for long range electrostatic interactions. Initial minimizations of 1000 steps (500 steps of steepest descent and 500 steps of conjugate gradient minimization) were performed keeping the solute fixed with a 500 kcal/mol•Å force constant. The entire system was then minimized over 2500 steps. A first MD run of 10000 steps was performed (20 ps) using the SHAKE algorithm. The initial temperature of 0 K was brought gradually up to 300 K and kept constant using Langevin dynamics and keeping the solute fixed with weak restraints (10 kcal/mol•Å). The RST file was then used for a second MD run of 50000 steps (100 ps) removing the restraints. A snapshot was recorded every 100 steps to generate the trajectory file. RST files were checked for general consistency with the NOE data and used for distance calculations.

#### 4.6.2 NMR Distance Restraints

In all cases, the distance restraints were obtained from 2D [<sup>1</sup>H-<sup>1</sup>H] NOESY (250, 200, 150, and 100 ms mixing times). NMR-FAM SPARKY was used for the assignment process.<sup>[86]</sup> Non-overlapping peaks were integrated using the Gaussian fit to generate the intensity file for MARDIGRAS.<sup>[81]</sup> Through CORMA.IN, the RST files obtained from preliminary molecular dynamics were converted into pdb format to be used with MARDIGRAS (using an isotropic model and relative error of 10%). The correlation times used were roughly established in each case checking MARDIGRAS calculation outputs from 1 ns to 8 ns. Absolute unnormalized noise was defined as a fraction of the smallest peak that could be integrated in the spectrum. The methyl jump 3 model was used. MARDIGRAS was run for 50 cycles. The resulting “.dst” files were compiled with RAND-RESTR to generate an average distance file. These coordinates were converted to restraints for input to AMBER using M2AHOMO and filtering out selected out-of-range distances. The force constants applied were 10 kcal/mol•Å<sup>2</sup> for lower and upper bounds. The width of the parabola was set to be 2 Å in all cases. Calculations were repeated at the beginning of each molecular dynamics cycle as reported below.

#### 4.6.3 NMR-restrained molecular dynamics simulations

Initial molecular dynamics were run using the Generalized-Born implicit solvent model. After initial minimization (500 steps, 250 steepest descent, 250 conjugated gradient), 100 ps MD were run at constant temperature. NMR restraints were gradually increased over the first 20 ps and kept constant for the remaining 80 ps. The output files were error checked, and the RST file was used for a second cycle of MARDIGRAS distance calculations. The new restraints were applied this time using the explicit solvent model and following the same protocol as above for the preliminary molecular dynamics. The RST file was then used as for the last cycle of MARDIGRAS and the calculated restraints were applied for the third MD simulation. A production run of 1 ns was obtained. Clustering of the last 800 ps was carried out using UCSF Chimera 1.12 to produce 11 representative structures which were each minimized and deposited in the Protein Data Bank (PDB ID **6GZ7**). An average minimized structure was obtained from the entire ensemble of (minimized) clusters using the CPPTRAJ module of AmberTools16. Utilizing this average minimized structure, and removing the polyamide



through Chimera, DNA structural groove analysis along with helical parameters were calculated using both Curves+ and X3DNA software to complement each other.<sup>[87,88]</sup> Bond distances, calculations, and the generation of structural figures were performed on the average minimized structure using UCSF Chimera.

## **Chapter 5**

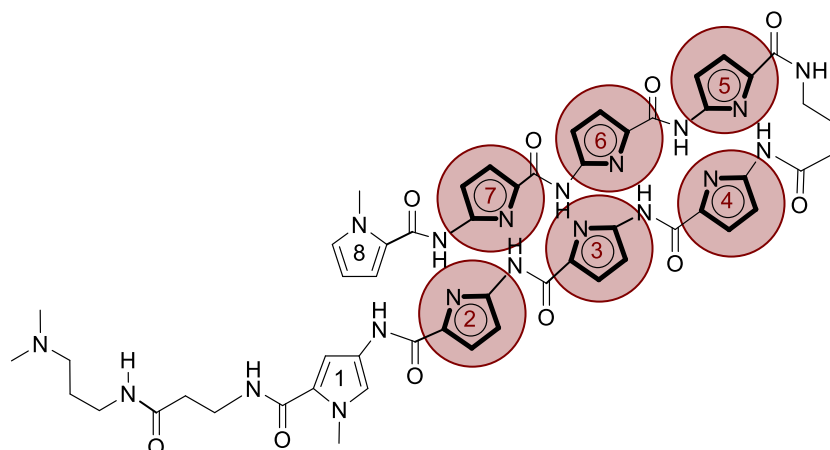
### **Future Work**

This thesis explored the synthesis of an 8-ring hairpin polyamide incorporating the novel *i*Pr-Im monomer unit in the N-terminal position (**PA4**) for targeting the ARE sequence 5'-WWGWWCW-3'. Analysis of the binding kinetics revealed that **PA4** bound with similar affinity and selectivity to the methyl imidazole analogue **PA1** for the target sequence. The switchSENSE® data provided further insight into the binding differences, with a rate map of **PA1-4** binding affinities separating the polyamides into two classes, the imidazole class and the thiazole class (classification based on the N-terminal heterocycle). The imidazole class of polyamides was revealed to have different binding affinities for the target sequence **ODN1** compared with the N-terminal thiazole-containing **PA2** and **PA3**, primarily owed to a faster association rate ( $k_{on}$ ) despite very similar dissociation ( $k_{off}$ ) rates. The selectivity observed for the target sequence over the mismatches varied for each polyamide, with **PA4** demonstrating the higher dissociation rates among the four polyamides tested for the mismatched sequences.

Determination of the binding kinetics remains a vital part of the workflow towards designing and developing better minor groove binding agents. The capability of a label-free method such as was demonstrated with the switchSENSE® biochip has proven utility for this end, and it would become a valuable asset for future ligand studies. The insight provided by the switchSENSE® tech enabled the work of this thesis, showing a key feature of the N-terminal imidazole class, **PA1-2** having a higher binding affinity due to a faster  $k_{on}$  rate. Going forward, future efforts should focus on the collection and comparison of the kinetic data from the switchSENSE® platform to the kinetic data from the Biacore SPR platform. These two methods vary in their reporter (fluorescence quenching vs. refractive index from mass change) and thus a systematic comparison should be conducted.

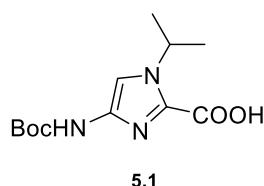
A 3D NMR structure obtained through molecular dynamics with the AMBER16 toolkit indicated that the novel *i*Pr-Im monomer induced a greater structural distortion compared to the Me-Im monomer. This is manifested in a greater compression of the major groove of **ODN4** by **PA4** compared with **PA1**. Furthermore, the NMR structure highlighted a key difference that is likely a factor in the increased selectivity for **PA4** over **PA3** with a deeper penetration depth and consequentially a stronger hydrogen bond between the exocyclic amine of guanine (G5H22) and N2 of the *i*Pr-Im monomer. From this fine

balance of steric bulk (inducing compression) and minor groove penetration (for good selectivity), this <sup>i</sup>Pr-Im monomer serves potential for replacement of the Im monomer in the N-terminal position of hairpin polyamides going forward. Further work is needed to understand if this monomer can be ubiquitous in the effects of its placement, with seven other potential ring positions available for installation (fig. 5.1).



**Figure 5.1.** Potential internal positions for <sup>i</sup>Pr-Im building block.

This would necessitate the development of a synthesis for the internal block of <sup>i</sup>Pr-Im (**5.1**), which may be accomplished using the route explored by Jaramillo *et al.* for the Me-Im analogue.<sup>[61]</sup>



**Figure 5.2.** Internal isopropyl imidazole amino acid building block (BocNH-<sup>i</sup>Pr-Im-COOH, **5.1**) for installation at internal positions of a polyamide.

Additionally, future work would need to focus on determining the impact of the <sup>i</sup>Pr-Im monomer to the pharmacokinetic properties of polyamides. The added alkyl bulk from the isopropyl was enough to substantially influence the hydrophobicity of the polyamide, with a theoretical logD calculated to be -0.90 (calculated by Chemicalize software from ChemAxon at chemicalize.com), indicating it lies between **PA2** and **PA3** based on their theoretical logD values. This added hydrophobic character may

enhance cellular uptake, necessitating the systematic evaluation of these compounds in cell permeability studies.

A critical assessment of the relationship between major groove distortion and *in vivo* potency also needs to be conducted, along with a cytotoxicity evaluation. The data collected would comprise a significant step forward for these <sup>i</sup>Pr-Im-containing hairpin polyamides as modulators of gene expression.

## References

- [1] A. A. Kurmis, F. Yang, T. R. Welch, N. G. Nickols, P. B. Dervan, *Cancer Res.* **2017**, *77*, 2207–2212.
- [2] G. Padroni, J. A. Parkinson, K. R. Fox, G. A. Burley, *Nucleic Acids Res.* **2018**, *46*, 42–53.
- [3] A. L. Lehninger, D. L. Nelson, M. M. Cox, *Lehninger Principles of Biochemistry*, W.H. Freeman, New York, **2013**.
- [4] T. Chatake, I. Tanaka, H. Umino, S. Arai, N. Niimura, *Acta Crystallogr. D. Biol. Crystallogr.* **2005**, *61*, 1088–1098.
- [5] P. Yakovchuk, E. Protozanova, M. D. Frank-Kamenetskii, *Nucleic Acids Res.* **2006**, *34*, 564–574.
- [6] J. D. Watson, F. H. C. Crick, *Nature* **1953**, *171*, 737.
- [7] S. Hahn, *Nat. Struct. Mol. Biol.* **2004**, *11*, 394.
- [8] A. L. Gnat, P. Cramer, J. Fu, D. A. Bushnell, R. D. Kornberg, *Science* **2001**, *292*, 1876 LP-1882.
- [9] A. Aranda, A. Pascual, *Physiol. Rev.* **2001**, *81*, 1269–1304.
- [10] F. Arcamone, S. Penco, P. Orezzi, V. Nicolella, A. Pirelli, *Nature* **1964**, *203*, 1064.
- [11] M. L. Kopka, C. Yoon, D. Goodsell, P. Pjura, R. E. Dickerson, *Proc. Natl. Acad. Sci.* **1985**, *82*, 1376 LP-1380.
- [12] J. G. Pelton, D. E. Wemmer, *Proc. Natl. Acad. Sci. U. S. A.* **1989**, *86*, 5723–5727.
- [13] J. G. Pelton, D. E. Wemmer, *J. Biomol. Struct. Dyn.* **1990**, *8*, 81–97.
- [14] K. Uytterhoeven, J. Sponer, L. Van Meervelt, *Eur. J. Biochem.* **2003**, *269*, 2868–2877.
- [15] X. Chen, B. Ramakrishnan, S. T. Rao, M. Sundaralingam, *Nat. Struct. Biol.* **1994**, *1*, 169.
- [16] K. Van Hecke, P. C. Nam, M. T. Nguyen, L. Van Meervelt, *FEBS J.* **2005**, *272*, 3531–3541.
- [17] J. M. Withers, G. Padroni, S. M. Pauff, A. W. Clark, S. P. Mackay, G. A. Burley, *DNA Minor Groove Binders as Therapeutic Agents, Comprehensive Supramolecular Chemistry II* Elsevier Inc., **2017**.
- [18] M. L. Kopka, D. S. Goodsell, G. W. Han, T. K. Chiu, J. W. Lown, R. E. Dickerson, *Structure* **1997**, *5*, 1033–1046.
- [19] K. Kissinger, K. Krowicki, J. C. Dabrowiak, J. W. Lown, *Biochemistry* **1987**, *26*, 5590–5595.
- [20] D. S. Goodsell, H. L. Ng, M. L. Kopka, J. W. Lown, R. E. Dickerson, *Biochemistry* **1995**, *34*, 16654–16661.
- [21] D. M. Chenoweth, P. B. Dervan, *J. Am. Chem. Soc.* **2010**, *132*, 14521–14529.
- [22] D. S. Pilch, N. Poklar, C. A. Gelfand, S. M. Law, K. J. Breslauer, E. E. Baird, P. B. Dervan, *Biochemistry* **1996**, *35*, 8306–8311.
- [23] P. B. Dervan, B. S. Edelson, *Curr. Opin. Struct. Biol.* **2003**, *13*(3), 284–299.
- [24] W. A. Greenberg, E. E. Baird, P. B. Dervan, *Chem. – A Eur. J.* **1998**, *4*, 796–805.

- [25] J. M. Gottesfeld, L. Neely, J. W. Trauger, E. E. Baird, P. B. Dervan, *Nature* **1997**, 387, 202.
- [26] H.-J. Jin, J. Kim, J. Yu, *Transl. Androl. Urol.* **2013**, 2, 158–177.
- [27] P. L. Shaffer, A. Jivan, D. E. Dollins, F. Claessens, D. T. Gewirth, *Proc. Natl. Acad. Sci. U. S. A.* **2004**, 101, 4758 LP-4763.
- [28] O. L. Zegarra-Moro, L. J. Schmidt, H. Huang, D. J. Tindall, *Cancer Res.* **2002**, 62, 1008 LP-1013.
- [29] C. Huggins, C. V Hodges, *Cancer Res.* **1941**, 1, 293 LP-297.
- [30] M. A. Perlmutter, H. Lepor, *Rev. Urol.* **2007**, 9, S3–S8.
- [31] N. Liu, W. Zhou, Y. Guo, J. Wang, W. Fu, H. Sun, D. Li, M. Duan, T. Hou, *J. Chem. Inf. Model.* **2018**, 58, 1652–1661.
- [32] R. L. Siegel, K. D. Miller, A. Jemal, *CA. Cancer J. Clin.* **2018**, 68, 7–30.
- [33] *American Cancer Society. Facts & Figures 2018. American Cancer Society. Atlanta, Ga. 2018, 2018.*
- [34] C. C. T. Sprenger, S. R. Plymate, *Horm. Cancer* **2014**, 5, 207–217.
- [35] R. Leibowitz–Amit, A. M. Joshua, *Curr. Oncol.* **2012**, 19, S22–S31.
- [36] B. Dudouet, R. Burnett, L. A. Dickinson, M. R. Wood, C. Melander, J. M. Belitsky, B. Edelson, N. Wurtz, C. Briehn, P. B. Dervan, et al., *Chem. Biol.* **2003**, 10, 859–867.
- [37] S.-Y. Chiang, R. W. Bürlü, C. C. Benz, L. Gawron, G. K. Scott, P. B. Dervan, T. A. Beerman, *J. Biol. Chem.* **2000**, 275, 24246–24254.
- [38] P. B. Dervan, *Bioorg. Med. Chem.* **2001**, 9, 2215–2235.
- [39] W. S. Wade, M. Mrksich, P. B. Dervan, *J. Am. Chem. Soc.* **1992**, 114, 8783–8794.
- [40] M. Mrksich, W. S. Wade, T. J. Dwyer, B. H. Geierstanger, D. E. Wemmer, P. B. Dervan, *Proc. Natl. Acad. Sci.* **1992**, 89, 7586 LP-7590.
- [41] C. L. Kielkopf, S. White, J. W. Szewczyk, J. M. Turner, E. E. Baird, P. B. Dervan, D. C. Rees, *Science* **1998**, 282, 111 LP-115.
- [42] C. L. Kielkopf, R. E. Bremer, S. White, J. W. Szewczyk, J. M. Turner, E. E. Baird, P. B. Dervan, D. C. Rees, *J. Mol. Biol.* **2000**, 295, 557–567.
- [43] M. A. Marques, R. M. Doss, S. Foister, P. B. Dervan, *J. Am. Chem. Soc.* **2004**, 126, 10339–10349.
- [44] A. R. Urbach, J. W. Szewczyk, S. White, J. M. Turner, E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc.* **1999**, 121, 11621–11629.
- [45] M. A. Marques, R. M. Doss, A. R. Urbach, P. B. Dervan, *Helv. Chim. Acta* **2003**, 85, 4485–4517.
- [46] J. W. Trauger, E. E. Baird, M. Mrksich, P. B. Dervan, *J. Am. Chem. Soc.* **1996**, 118, 6160–6166.
- [47] M. Mrksich, M. E. Parks, P. B. Dervan, *J. Am. Chem. Soc.* **1994**, 116, 7983–7988.
- [48] C. Dose, M. E. Farkas, D. M. Chenoweth, P. B. Dervan, *J. Am. Chem. Soc.* **2008**, 130, 6859–6866.
- [49] J. L. Meier, D. C. Montgomery, P. B. Dervan, *Nucleic Acids Res.* **2012**, 40, 2345–2356.

- [50] A. E. Hargrove, J. A. Raskatov, J. L. Meier, D. C. Montgomery, P. B. Dervan, *J. Med. Chem.* **2012**, *55*, 5425-5432.
- [51] S. E. Swalley, E. E. Baird, P. B. Dervan, *Chem. - A Eur. J.* **1997**, *3*, 1600–1607.
- [52] J. J. Kelly, E. E. Baird, P. B. Dervan, *Proc. Natl. Acad. Sci.* **1996**, *93*, 6981 LP-6985.
- [53] R. Burnett, C. Melander, J. W. Puckett, L. S. Son, R. D. Wells, P. B. Dervan, J. M. Gottesfeld, *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 11497–11502.
- [54] P. E. Floreancig, S. E. Swalley, J. W. Trauger, P. B. Dervan, *J. Am. Chem. Soc.* **2000**, *122*, 6342–6350.
- [55] S. K. Sharma, M. Tandon, J. W. Lown, *J. Org. Chem.* **2000**, *65*, 1102–1107.
- [56] D. H. Nguyen, J. W. Szewczyk, E. E. Baird, P. B. Dervan, *Bioorg. Med. Chem.* **2001**, *9*, 7–17.
- [57] C. J. Suckling, *Chem. Biol. Inter.* **2015**, *5(3)*, 166–174.
- [58] M.-V. Salvia, F. Addison, H. Y. Alniss, N. J. Buurma, A. I. Khalaf, S. P. Mackay, N. G. Anthony, C. J. Suckling, M. P. Evstigneev, A. H. Santiago, et al., *Biophys. Chem.* **2013**, *179*, 1–11.
- [59] H. Y. Alniss, M.-V. Salvia, M. Sadikov, I. Golovchenko, N. G. Anthony, A. I. Khalaf, S. P. MacKay, C. J. Suckling, J. A. Parkinson, *ChemBioChem* **2014**, *15*, 1978–1990.
- [60] E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc.* **1996**, *118*, 6141–6146.
- [61] D. Jaramillo, Q. Liu, J. Aldrich-Wright, Y. Tor, *J. Org. Chem.* **2004**, *69(23)*, 8151-8153.
- [62] N. R. Wurtz, J. M. Turner, E. E. Baird, P. B. Dervan, *Org. Lett.* **2001**, *3*, 1201–1203.
- [63] A. J. Fallows, I. Singh, R. Dondi, P. M. Cullis, G. A. Burley, *Org. Lett.* **2014**, *16*, 4654–4657.
- [64] M. Brenowitz, D. F. Seneor, M. A. Shea, G. K. B. T.-M. in E. Ackers, in *Enzym. Struct. Part K*, Academic Press, **1986**, 32–181.
- [65] J. Knezevic, A. Langer, P. A. Hampel, W. Kaiser, R. Strasser, U. Rant, *J. Am. Chem. Soc.* **2012**, *134*, 15225–15228.
- [66] J. Atkins, P.; De Paula, *Physical Chemistry*, W.H. Freeman, **2010**.
- [67] J. P. Ramos, B. Babu, S. Chavda, Y. Liu, A. Plaunt, A. Ferguson, M. Savagian, M. Lee, S. Tzou, S. Lin, et al., *Biopolymers* **2013**, *99*, 497-507.
- [68] S. Wang, R. Nanjunda, K. Aston, J. K. Bashkin, W. D. Wilson, *Biochemistry* **2012**, *51*, 9796–9806.
- [69] S. Wang, K. Aston, K. J. Koeller, G. D. Harris, N. P. Rath, J. K. Bashkin, W. D. Wilson, *Org. Biomol. Chem.* **2014**, *12*, 7523–7536.
- [70] W. D. Wilson, F. A. Tanious, M. Fernandez-Saiz, C. T. Rigl, *Evaluation of drug-nucleic acid interactions by thermal melting curves*. in: *Methods in Molecular Biology, vol. 90: Drug-DNA Interaction Protocols* (Ed.: K.R. Fox), Humana Press, Totowa, NJ, **1997**, pp. 219–240.
- [71] Y. Liu, W. D. Wilson, *Quantitative Analysis of Small Molecule–Nucleic Acid Interactions with a Biosensor Surface and Surface Plasmon Resonance Detection*. in: *Methods in Molecular Biology, vol. 613: Drug-DNA Interaction Protocols* (Ed.: K.R. Fox), Humana Press, Totowa, NJ, **2010**, pp. 1–23.
- [72] B. Nguyen, F. A. Tanious, W. D. Wilson, *Methods* **2007**, *42*, 150–161.
- [73] A. Cléry, T. J. M. Sohler, T. Welte, A. Langer, F. H. T. Allain, *Methods* **2017**, *118–119*, 137–



145.

- [74] J. L. Meier, A. S. Yu, I. Korf, D. J. Segal, P. B. Dervan, *J. Am. Chem. Soc.* **2012**, *134*(42), 17814-17822.
- [75] D. M. Chenoweth, P. B. Dervan, *Proc. Natl. Acad. Sci.* **2009**, *106*, 13175 LP-13179.
- [76] Q. Zhang, T. J. Dwyer, V. Tsui, D. A. Case, J. Cho, P. B. Dervan, D. E. Wemmer, *J. Am. Chem. Soc.* **2004**, *126*, 7958–7966.
- [77] Peter J. Hore, *Nuclear Magnetic Resonance*, Oxford Science Publications, **1995**.
- [78] B. J. Alder, T. E. Wainwright, *J. Chem. Phys.* **1959**, *31*, 459–466.
- [79] M. De Vivo, M. Masetti, G. Bottegoni, A. Cavalli, *J. Med. Chem.* **2016**, *59*, 4035–4061.
- [80] E. G. Lewars, *Computational Chemistry, 2nd Ed.*, Springer, **2011**.
- [81] B. A. Borgias, T. L. James, *J. Magn. Reson.* **1990**, *87*, 475–487.
- [82] R. P. L. de Clairac, B. H. Geierstanger, M. Mrksich, P. B. Dervan, D. E. Wemmer, *J. Am. Chem. Soc.* **1997**, *119*, 7909–7916.
- [83] D. A. Case, D. S. Cerutti, T. E. I. Cheatham, T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Goetz, D. Greene, N. Homeyer, et al., *Univ. California, San Fr.* **2017**, AMBER 2017, University of California, San Francisco.
- [84] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *J. Comput. Chem.* **2004**, *25*, 1605–1612.
- [85] D. M. Chenoweth, P. B. Dervan, *J. Am. Chem. Soc.* **2010**, *132*, 14521–14529.
- [86] W. Lee, M. Tonelli, J. L. Markley, *Bioinformatics* **2015**, *31*, 1325–1327.
- [87] C. Blanchet, M. Pasi, K. Zakrzewska, R. Lavery, *Nucleic Acids Res.* **2011**, *39*, W68–W73.
- [88] X.-J. Lu, W. K. Olson, *Nucleic Acids Res.* **2003**, *31*, 5108–5121.
- [89] P. Iyer, A. Srinivasan, S. K. Singh, G. P. Mascara, S. Zayitova, B. Sidone, E. Fouquerel, D. Svilar, R. W. Sobol, M. S. Bobola, et al., *Chem. Res. Toxicol.* **2013**, *26*, 156–168.
- [90] M. El Hassan, C. Calladine, *J. Mol. Biol.* **1998**, *282*, 331–343.

## **Appendix**

## Contents

<a href="#">Appendix</a> .....	130
<a href="#">1 HPLC and LC-MS method parameters</a> .....	132
<a href="#">Semi-preparative HPLC method</a> .....	132
<a href="#">LC-MS method</a> .....	132
<a href="#">Analytical HPLC generic methods</a> .....	133
<a href="#">2 Spectroscopic characterization of compounds described in Chapter 2</a> .....	134
<a href="#">3 NMR chemical shifts of PA·ODN4 complex</a> .....	148
<a href="#">4 Profiling of the binding kinetics</a> .....	154
<a href="#">5 Analysis of NMR-derived MD structure</a> .....	159

## 1 HPLC and LC-MS method parameters

### Semi-preparative HPLC method

Column specification: Kinetex 5  $\mu$ m C18 100A, 150x 21.2mm

Column temperature: 19-25 °C (uncontrolled)

Mobile phase A: 0.1% v/v TFA in water

Mobile phase B: 0.1% v/v TFA in acetonitrile

Flow rate: 9.0 mL/min

Gradient profile:

Time (mins)	%A	%B
0	90	10
5	90	10
25	40	60
28	10	90
30	10	90
31	90	10
37	90	10

UV Detection: 310 nm

### LC-MS method

Column Specification: Zorbax 45mm x 150mm C18 column at 40 °C

Column Temperature: 40 °C

Mobile Phase A: 5 mM ammonium acetate in water.

Mobile Phase B: 5 mM ammonium acetate in acetonitrile.

The gradient employed was:

Time (min)	Flow Rate (ml/min)	% A	% B
0	1.0	95	5
1.48	1.0	95	5
8.5	1.0	0	100
13.5	1.0	0	100
16.5	1.0	95	5
18	1.0	95	5

UV: 254 nm

MS: Agilent Quadrupole

Ionisation mode: Positive and negative electrospray.

Scan Range: 100 to 2000 AMU positive, 120-2000 AMU negative.

## Analytical HPLC generic methods

Column specification: Aeris 3.6  $\mu\text{m}$  WIDEPOR $\text{E}$  XB-C18, 250 x 4.6 mm

Column Temperature: 25  $^{\circ}\text{C}$

Mobile Phase A: 0.1% v / v TFA in water

Mobile Phase B: 0.1% v / v TFA in acetonitrile

Flow Rate: 0.5 mL/min

Method A

Gradient profile:

Time (mins)	%A	%B
0	85	15
5	85	15
25	40	60
30	40	60
31	10	90
36	10	90
37	85	15
42	85	15

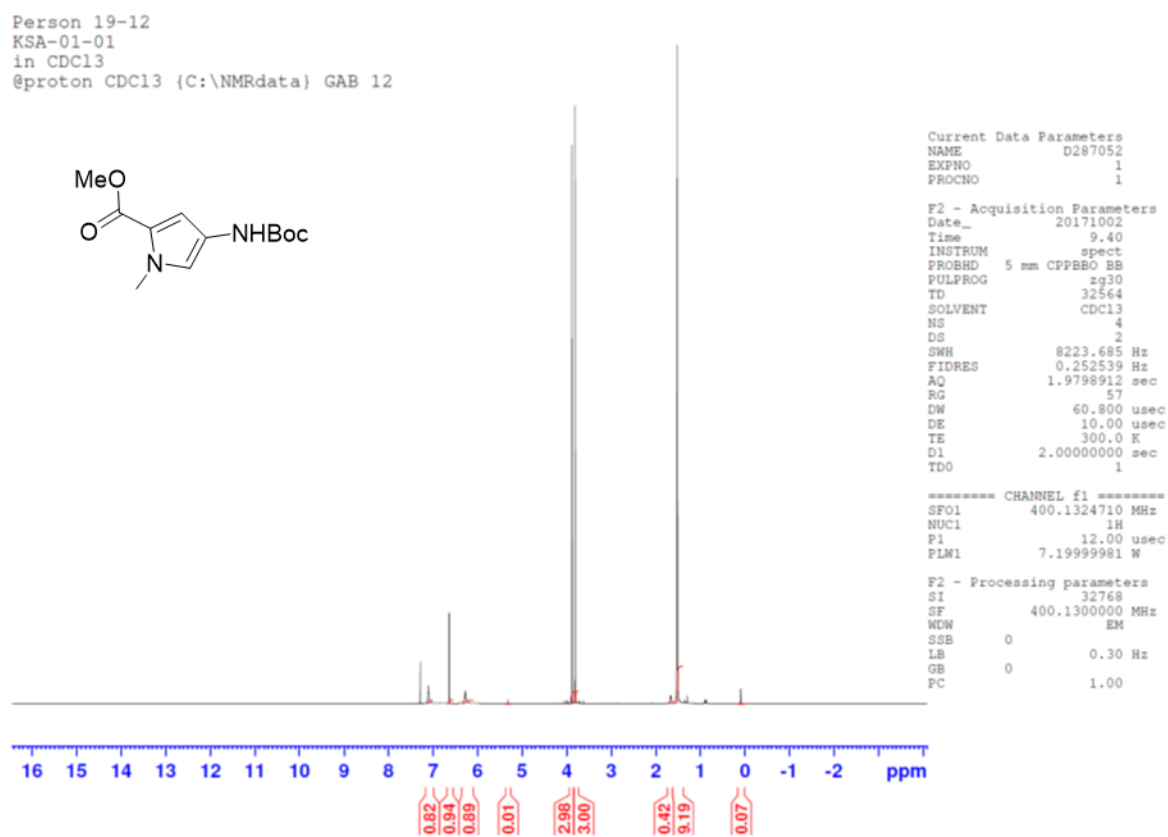
Method B

Gradient profile:

Time (mins)	%A	%B
0	85	15
5	85	15
25	10	90
28	10	90
29	10	90
30	85	15
37	85	15

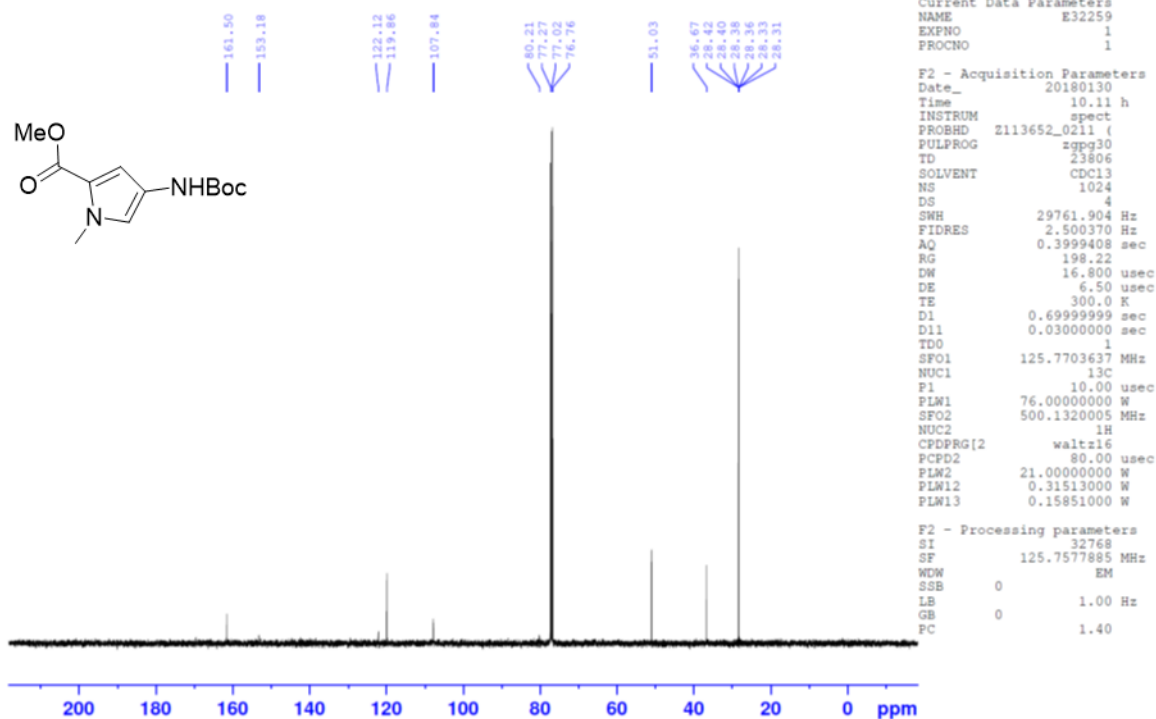
## 2 Spectroscopic characterization of compounds described in Chapter 2

Methyl 4-((tert-butoxycarbonyl)amino)-1-methyl-1H-pyrrole-2-carboxylate (**2.26**)<sup>[89]</sup>



Appendix 1. <sup>1</sup>H-NMR of **2.26**.

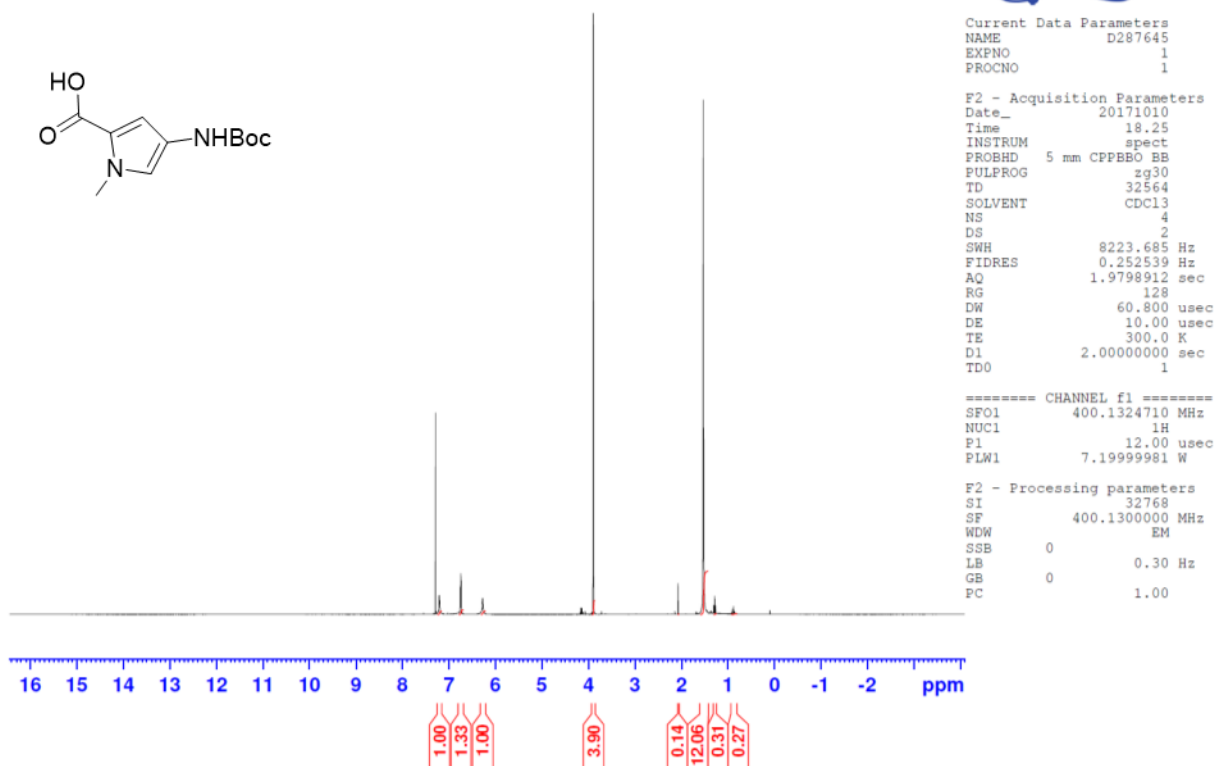
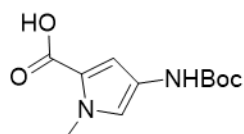
Person 19-4  
 KSA-01-01  
 in CDCl<sub>3</sub>  
 methyl ester of Py monomer block  
 KSA 29-Jan-2018



Appendix 2. <sup>13</sup>C-NMR of 2.26.

4-((tert-butoxycarbonyl)amino)-1-methyl-1H-pyrrole-2-carboxylic acid (**2.1**)<sup>[89]</sup>

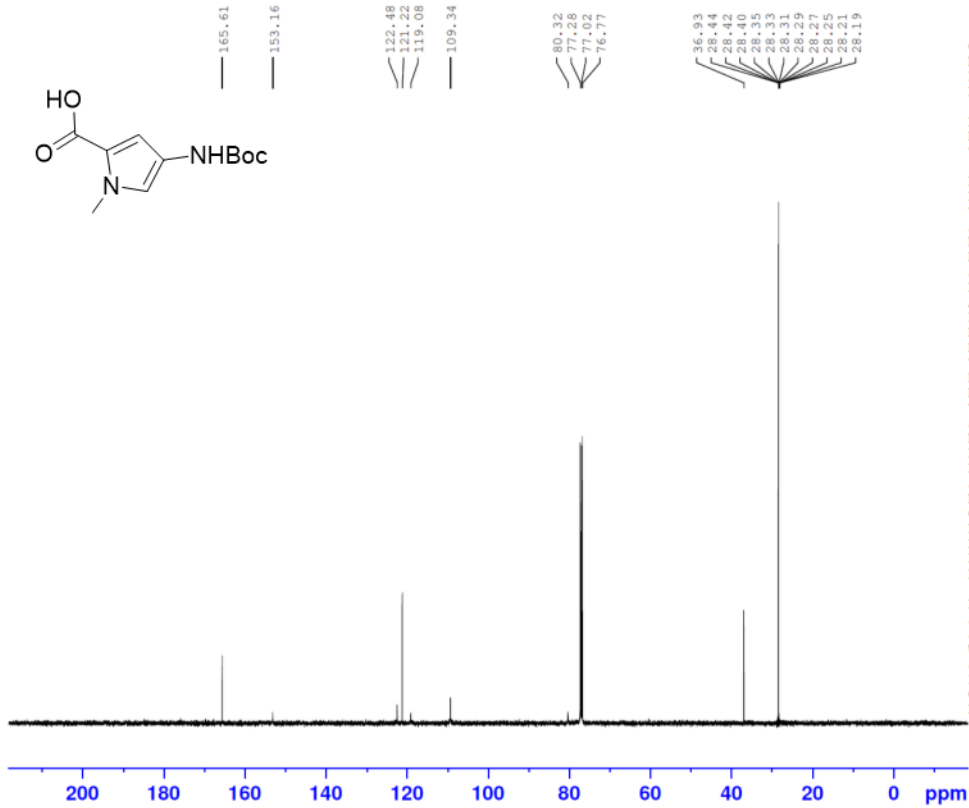
Person 19-4  
KSA-01-03  
in CDC13  
@proton CDC13 {C:\NMRdata} GAB 25



Appendix 3. <sup>1</sup>H NMR of **2.1**.



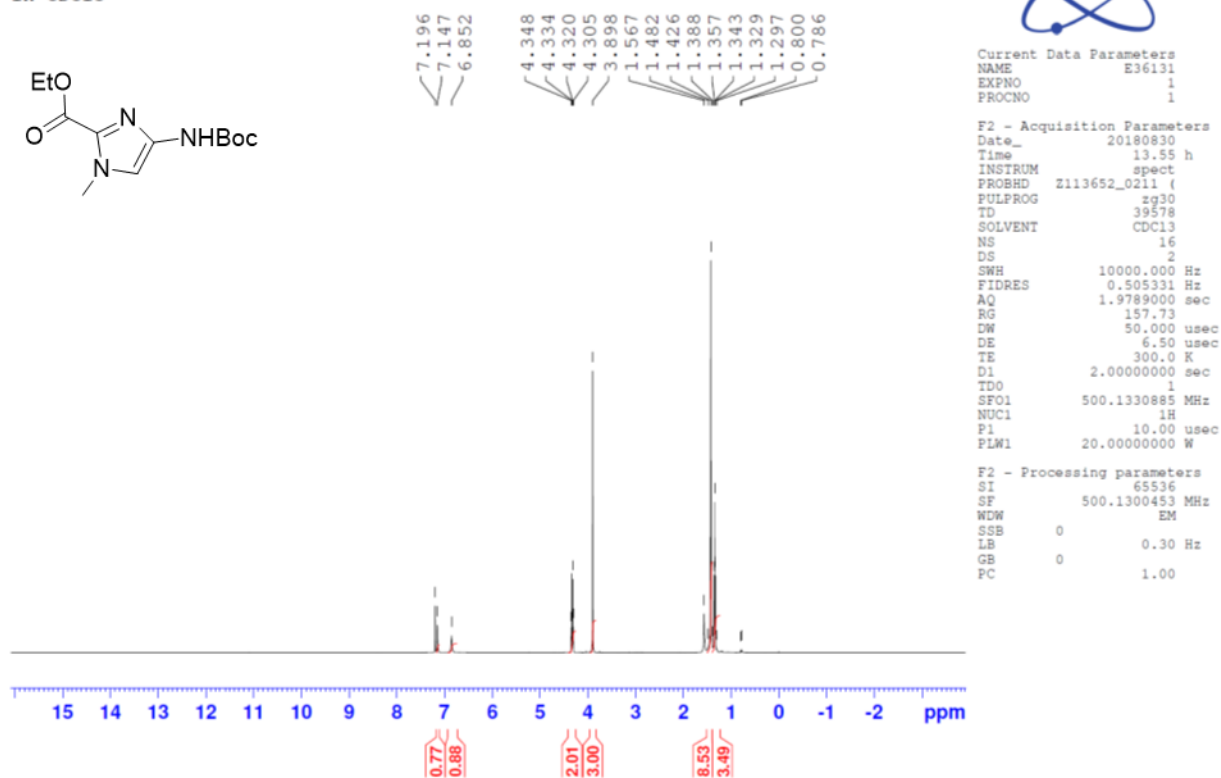
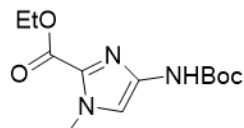
Person 19-4  
 KSA-01-03  
 Boc-Py-COOH  
 in CDCl3  
 KSA 08-Nov-2017



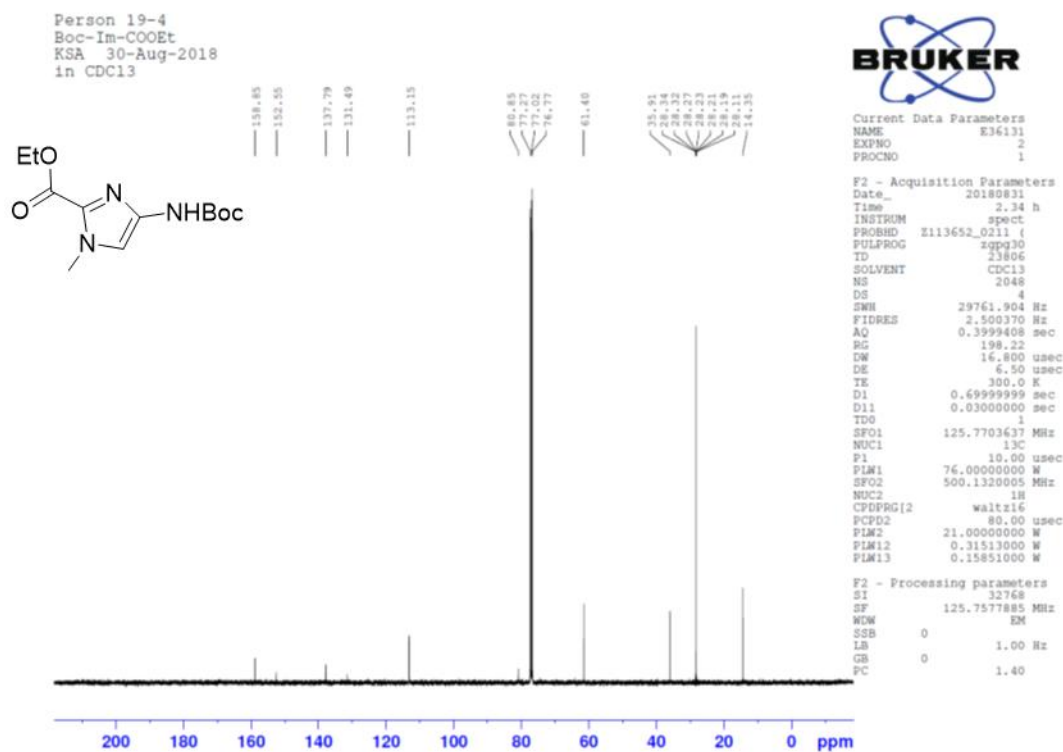
Appendix 4. <sup>13</sup>C NMR of 2.1.

Ethyl 4-((tert-butoxycarbonyl)amino)-1-methyl-1H-imidazole-2-carboxylate (**2.27**) [61]

Person 19-4  
Boc-Im-COOEt  
KSA 30-Aug-2018  
in CDCl3

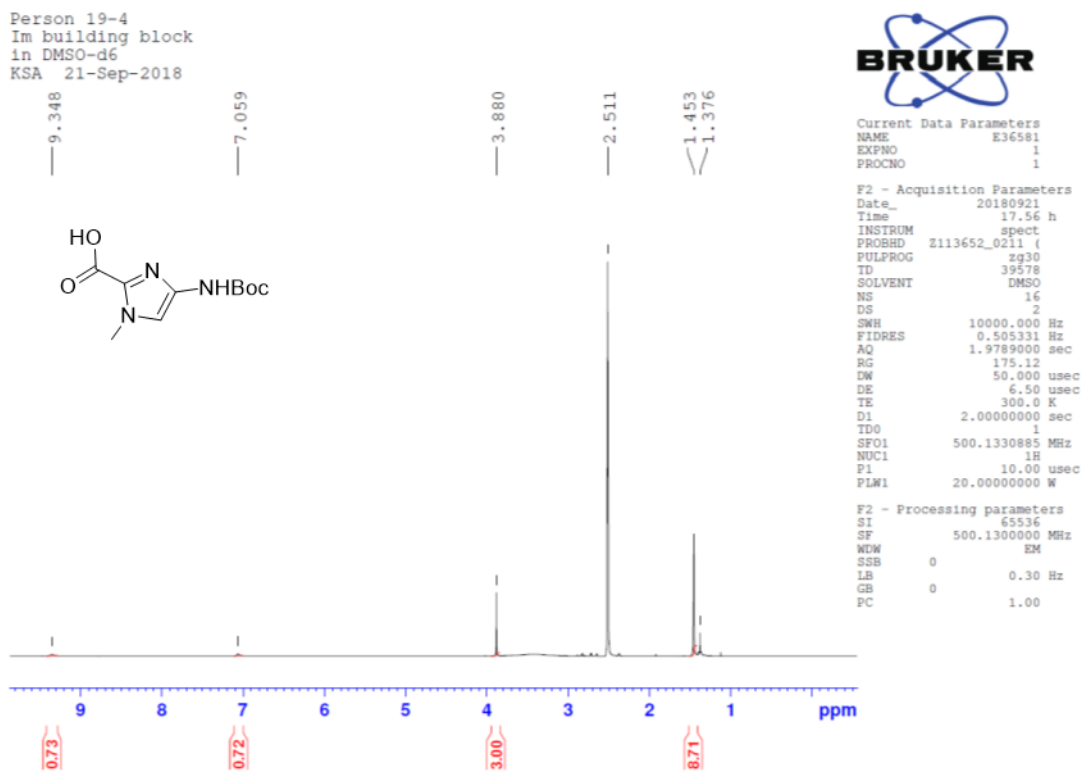


Appendix 5. <sup>1</sup>H NMR of **2.27**.



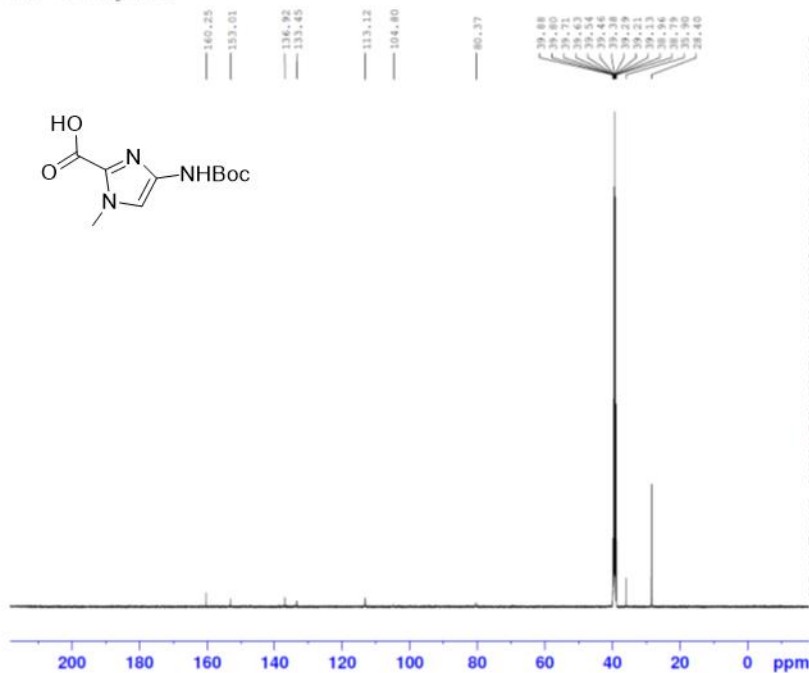
Appendix 6.  $^{13}\text{C}$  NMR of 2.27.

4-((tert-butoxycarbonyl)amino)-1-methyl-1H-imidazole-2-carboxylic acid (2.2) <sup>[61]</sup>



Appendix 7.  $^1\text{H}$  NMR of 2.2.

Person 19-4  
Boc-Im-COOH  
in DMSO  
KSA 30-Aug-2018



Current Data Parameters  
NAME E36132  
EXPNO 2  
PROCNO 1

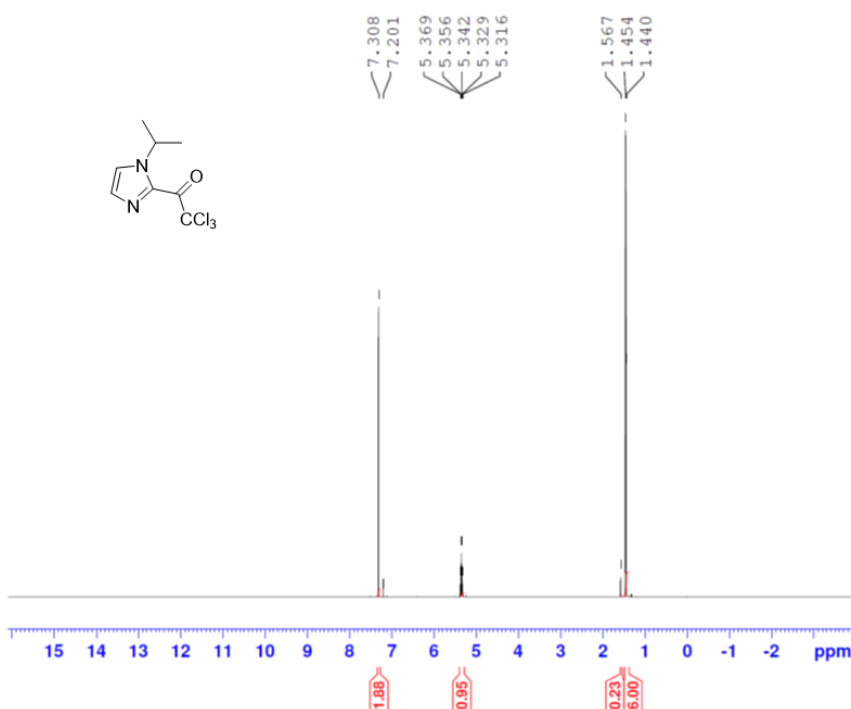
F2 - Acquisition Parameters  
Date\_ 20180831  
Time 3.21 h  
INSTRUM spect  
PROBHD Z113652\_0211 ( )  
PULPROG zgpg30  
TD 23806  
SOLVENT DMSO  
NS 2048  
DS 4  
SWH 29761.904 Hz  
FIDRES 2.500370 Hz  
AQ 0.3999408 sec  
RG 198.22  
DW 16.800 usec  
DE 6.50 usec  
TE 300.0 K  
D1 0.69999999 sec  
D11 0.03000000 sec  
TD0 1  
SFO1 125.7703437 MHz  
NUC1 13C  
P1 10.00 usec  
PLW1 76.00000000 W  
SFO2 500.1320005 MHz  
NUC2 1H  
PCPD2 waltz16  
PCPD2 80.00 usec  
PLW2 21.00000000 W  
PLW12 0.31513000 W  
PLW13 0.15851000 W

F2 - Processing parameters  
SI 32768  
SF 125.7577885 MHz  
MW EM  
SSB 0  
LB 1.00 Hz  
GB 0  
PC 1.40

Appendix 8. <sup>13</sup>C NMR of 2.2.

2-trichloroacetyl-N-isopropylimidazole (2.29) [61]

Person 19-4  
2-trichloroacetyl N-isopropylimidazole  
in CDCl<sub>3</sub>  
KSA-01-24



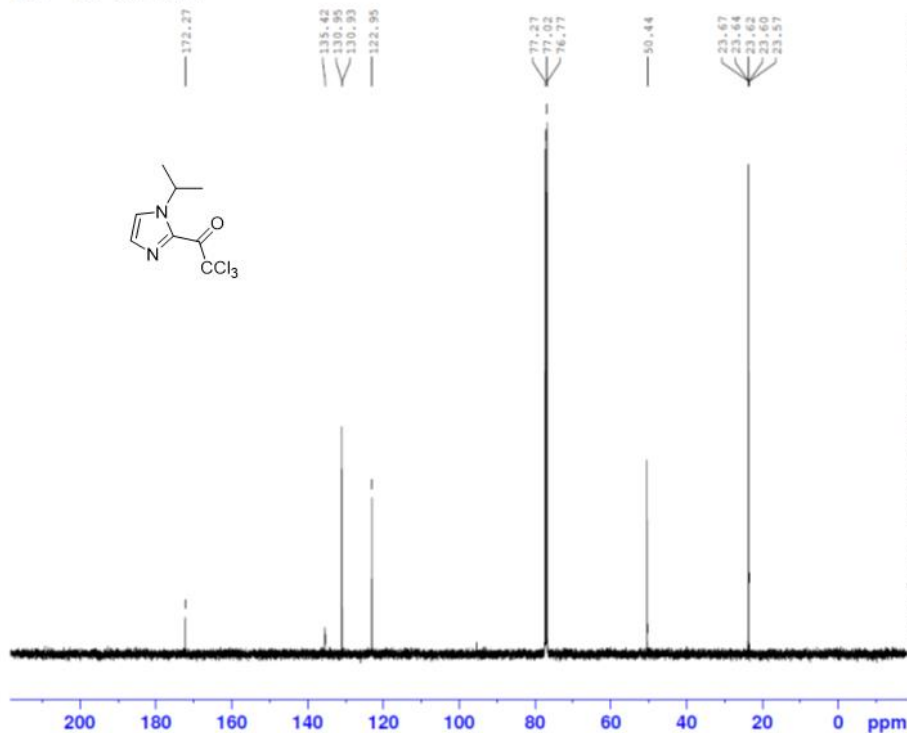
Current Data Parameters  
NAME E32069  
EXPNO 1  
PROCNO 1

F2 - Acquisition Parameters  
Date\_ 20180119  
Time 14.09 h  
INSTRUM spect  
PROBHD z113652\_0211 (4  
PULPROG zg30  
TD 39578  
SOLVENT CDC13  
NS 16  
DS 2  
SWH 10000.000 Hz  
FIDRES 0.505331 Hz  
AQ 1.9789000 sec  
RG 111.03  
DW 50.000 usec  
DE 6.50 usec  
TE 300.0 K  
D1 2.0000000 sec  
TD0 1  
SFO1 500.1330885 MHz  
NUC1 1H  
P1 10.00 usec  
PLW1 20.00000000 W

F2 - Processing parameters  
SI 65536  
SF 500.1300426 MHz  
WDW EM  
SSB 0  
LB 0.30 Hz  
GB 0  
PC 1.00

Appendix 9. <sup>1</sup>H NMR of 2.29.

Person 19-4  
KSA-01-18  
product  
in CDCl<sub>3</sub>  
KSA 29-Nov-2017



Current Data Parameters  
NAME E31304  
EXPNO 2  
PROCNO 1

F2 - Acquisition Parameters  
Date\_ 20171129  
Time 18.25 h  
INSTRUM spect  
PROBHD z8438\_0002 (DU  
PULPROG zgpg30  
TD 23806  
SOLVENT CDC13  
NS 1024  
DS 4  
SWH 29761.904 Hz  
FIDRES 2.500370 Hz  
AQ 0.3999408 sec  
RG 198.22  
DW 16.800 usec  
DE 6.50 usec  
TE 300.0 K  
D1 0.69999999 sec  
D11 0.03000000 sec  
TD0 1  
SFO1 125.7703637 MHz  
NUC1 13C  
P1 8.50 usec  
PLW1 50.00000000 W  
SFO2 500.1320005 MHz  
NUC2 1H  
PCPDPRG2 waltz16  
PCPD2 80.00 usec  
PLW2 20.00000000 W  
PLW12 0.41328001 W  
PLW13 0.20788001 W

F2 - Processing parameters  
SI 32768  
SF 125.7577885 MHz  
WDW EM  
SSB 0  
LB 1.00 Hz  
GB 0  
PC 1.40

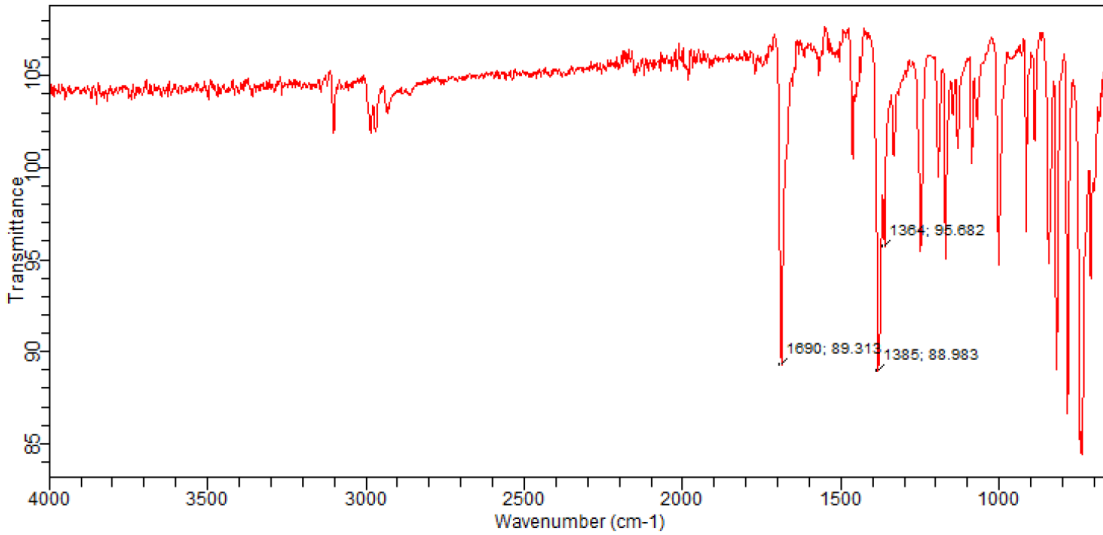
Appendix 10. <sup>13</sup>C NMR of 2.29.



# Agilent Technologies

Sample ID: KSA-01-24  
Sample Scans: 32  
Background Scans: 32  
Resolution: 4 cm<sup>-1</sup>  
System Status: Good  
File Location: C:\Program Files (x86)\Agilent\MicroLab PC\Results\ATR 32 4cm\KSA-01-24\_2018-04-03T16-43-30.a2r

Method Name: ATR 32 4cm  
User: STUDENT  
Date/Time: 03/04/2018 16:41:55  
Range: 4,000.00 - 650.00  
Apodization: Happ-Genzel



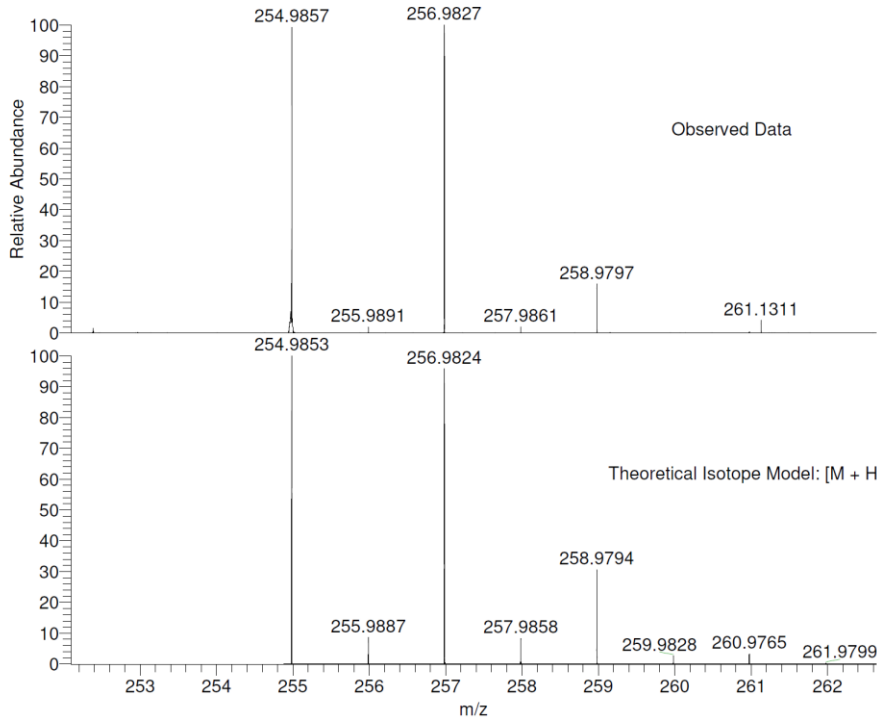
## Appendix 11. ATR-IR spectrum of 2.29.

KSA-01-24  
(DCM)/MeOH + NH<sub>4</sub>OAc  
C<sub>8</sub>H<sub>9</sub>Cl<sub>3</sub>N<sub>2</sub>O

EPSRC National Facility Swansea  
LTQ Orbitrap XL

STRBUR  
10/04/2018 13:12:38

SM: 7G



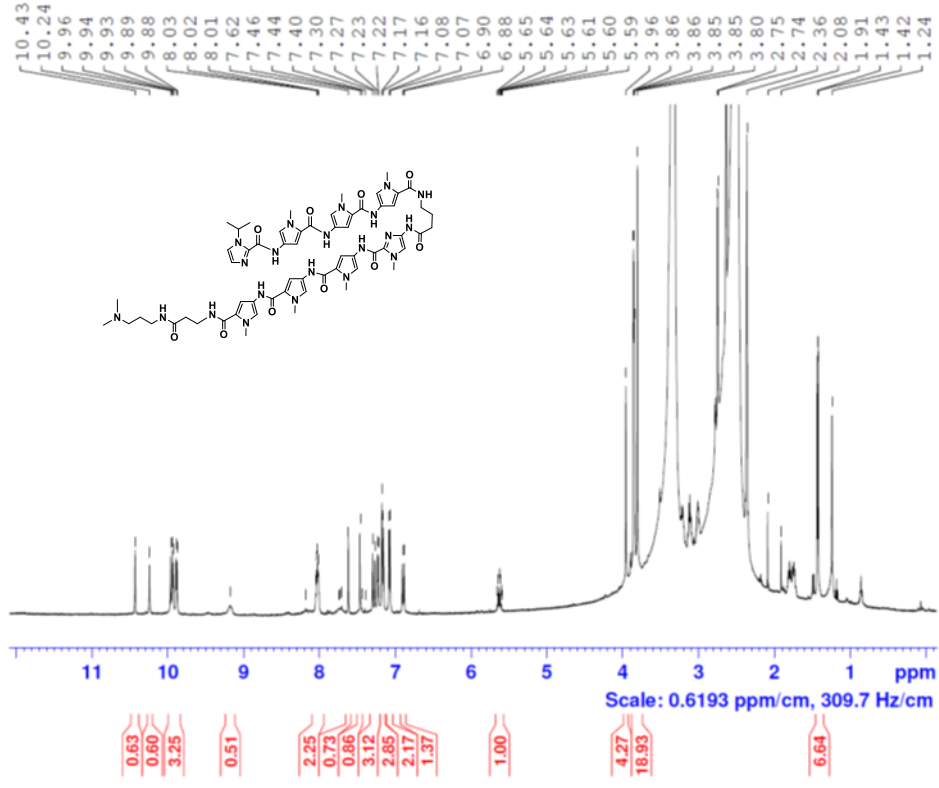
NL:  
1.62E5  
STRBUR\_M74CW\_44129#42-  
55 RT: 0.74-1.04 AV: 12 T:  
FTMS + p NSI Full ms  
[120.00-1935.00]

NL:  
9.27E3  
C<sub>8</sub>H<sub>9</sub>Cl<sub>3</sub>N<sub>2</sub>OH:  
C<sub>8</sub>H<sub>10</sub>Cl<sub>3</sub>N<sub>2</sub>O<sub>1</sub>  
p (gss, s/p:40) Chrg 1  
R: 100000 Res. Pwr. @FWHM

## Appendix 12. HRMS of 2.29.

# PA4

Person 19-4  
 KSA-01-26  
 polyamide with iPrIm terminal  
 in DMSO-d6

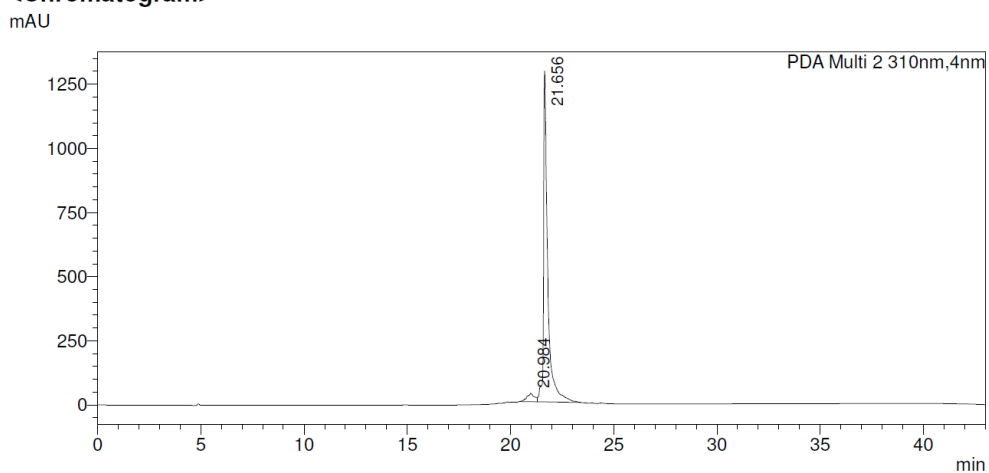


Appendix 9. <sup>1</sup>H NMR spectrum of PA4.

## <Sample Information>

Sample Name : KSA\_01\_20\_A\_II  
 Sample ID : KSA\_01\_20\_A\_II  
 Data Filename : polyamide with isopropyl\_II\_15122017\_005.lcd  
 Method Filename : PA Method trial.lcm  
 Batch Filename : polyamide with isopropyl\_II.lcb  
 Vial # : 1-2 Sample Type : Unknown  
 Injection Volume : 50 uL  
 Date Acquired : 15/12/2017 17:19:00 Acquired by : Shimadzu  
 Date Processed : 15/12/2017 18:02:02 Processed by : Shimadzu

## <Chromatogram>



## <Peak Table>

PDA Ch2 310nm

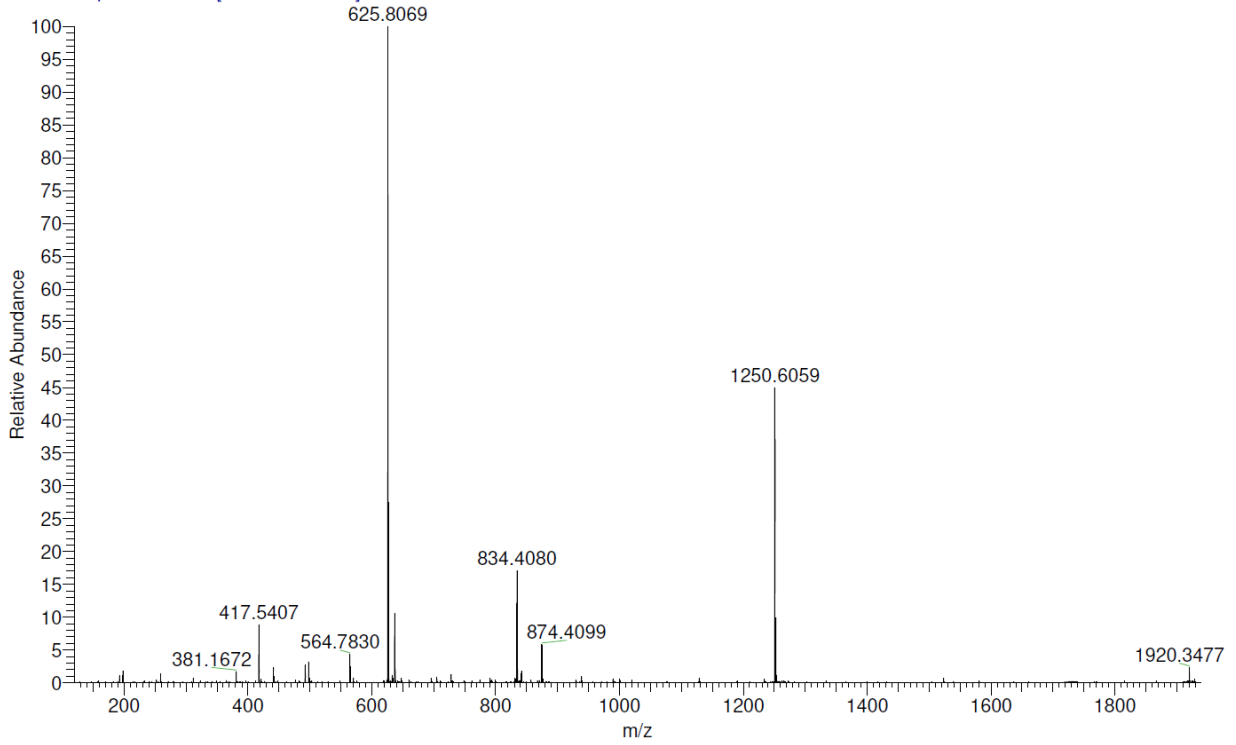
Peak#	Ret. Time	Area	Height	Conc.	Unit	Mark	Name
1	20.984	805317	33204	4.286		M	
2	21.656	17983828	1289934	95.714		V M	
Total		18789144	1323138				

**Appendix 10. HPLC chromatogram of PA4.**





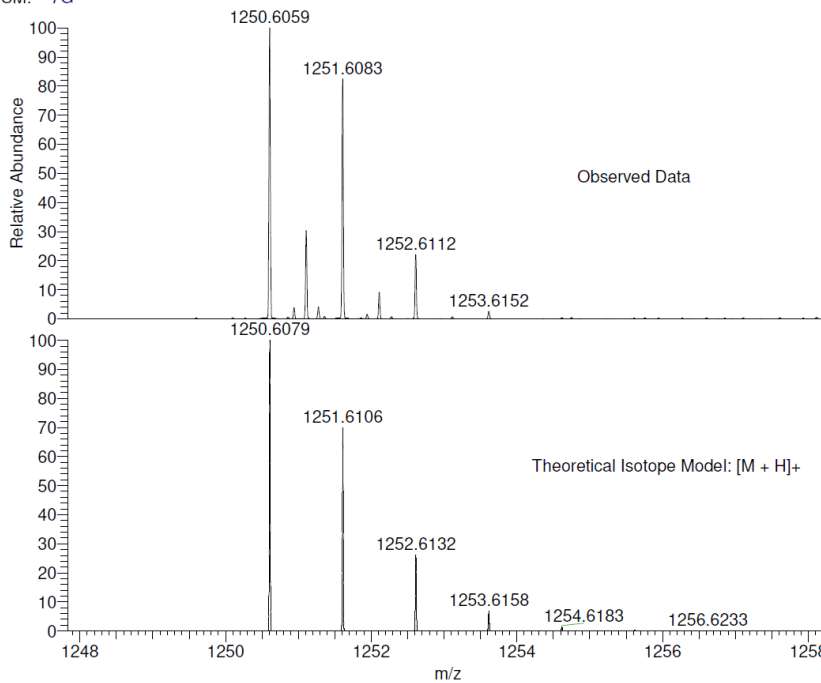
STRBUR\_M743H\_44126 #40-54 RT: 0.74-1.04 AV: 12 SM: 7G NL: 4.79E7  
T: FTMS + p NSI Full ms [120.00-1935.00]



KSA-01-26  
(MeOH)/MeOH+NH4OAc  
C60H75N21O10  
SM: 7G

EPSRC National Facility Swansea  
LTQ Orbitrap XL

STRBUR  
10/04/2018 13:21:21

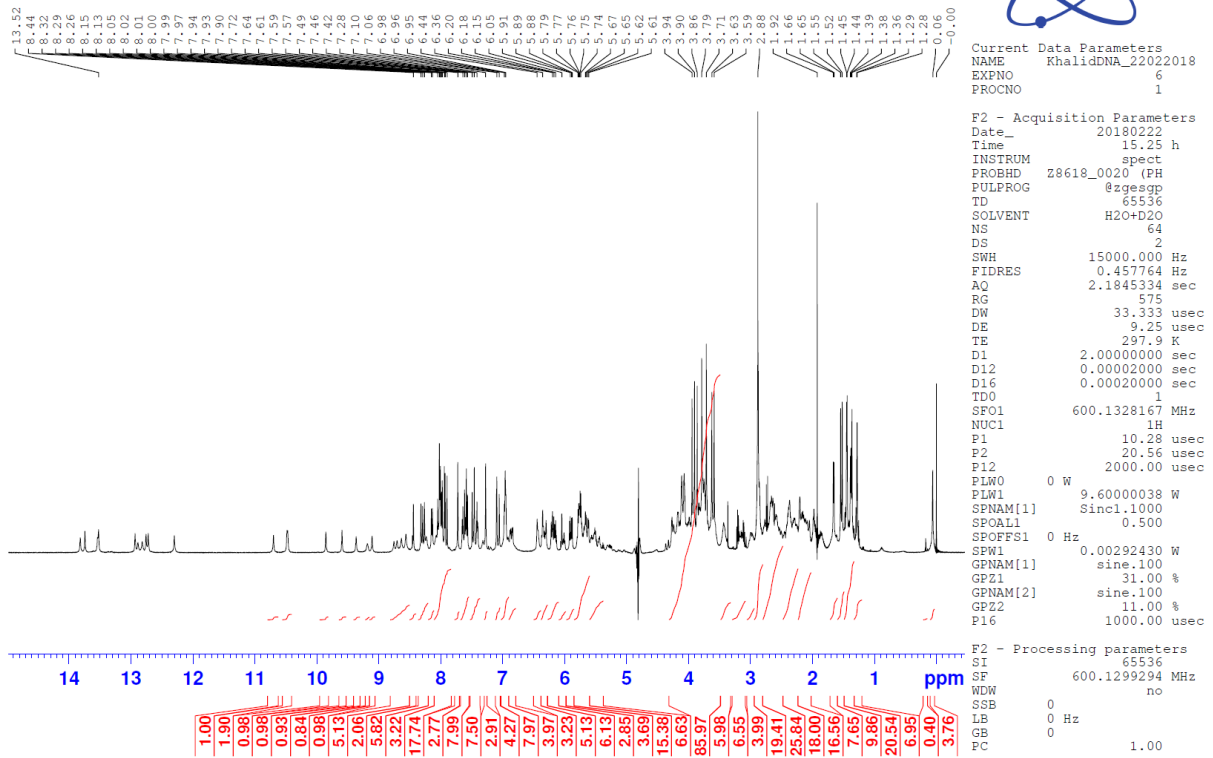


NL:  
2.15E7  
STRBUR\_M743H\_44126#40-  
54 RT: 0.74-1.04 AV: 12 T:  
FTMS + p NSI Full ms  
[120.00-1935.00]

NL:  
1.10E4  
C<sub>60</sub>H<sub>75</sub>N<sub>21</sub>O<sub>10</sub>H:  
C<sub>60</sub>H<sub>76</sub>N<sub>21</sub>O<sub>10</sub>  
p (gss, s /p:40) Chrg 1  
R: 100000 Res. .Pwr . @FWHM

## Appendix 16. HRMS of PA4.

d(CGATGTACATCG) 1.5 mM in 90% H<sub>2</sub>O / 10% D<sub>2</sub>O @ 298 K  
 1H NMR Spectrum at 600 MHz  
 Start of Titration - 5th Point 115 uL of 15 mM solution of polyamide



Appendix 17. <sup>1</sup>H NMR spectrum of PA4•ODN4 complex in 90% H<sub>2</sub>O/10% D<sub>2</sub>O.

### 3 NMR chemical shifts of PA·ODN4 complex

**Appendix 18.** <sup>1</sup>H sugar chemical shifts of free ODN4 (H5' and H5'' were not stereochemically assigned. “'” and “''” refer to the most upfield and downfield signal, respectively).

ODN4	H1'	H2'	H2''	H3'	H4'	H5'	H5''
C1	5.714	2.382	1.886	4.697	4.064	3.724	3.724
G2	5.561	2.732	2.821	5.016	4.336	3.986	4.101
A3	6.292	2.681	2.968	5.046	4.501	4.237	4.237
T4	5.757	2.065	2.446	4.868	4.184	4.188	4.284
G5	5.902	2.547	2.717	4.913	4.36	4.16	4.16
T6	5.705	2.083	2.432	4.881	4.205	n.d	n.d
A7	6.168	2.67	2.865	5.021	4.406	4.107	4.17
C8	5.506	2.059	2.382	4.802	4.183	4.172	4.247
A9	6.207	2.625	2.91	4.98	4.397	n.d	n.d
T10	5.943	2.009	2.412	4.848	4.172	n.d	n.d
C11	5.716	2.015	2.381	4.844	4.129	n.d	n.d
G12	6.151	2.614	2.47	4.694	4.2	4.174	4.174

**Appendix 19.** <sup>1</sup>H base chemical shifts of free ODN4.

ODN4	H1	H2	H3	H41	H42	H5	H6	H7	H8
C1	-	-	-	6.962	n.d	5.871	7.601	-	-
G2	12.82	-	-	-	-	-	-	-	7.958
A3	-	7.849	-	-	-	-	-	-	8.255
T4	-	-	13.31	-	-	-	7.068	1.406	-
G5	12.32	-	-	-	-	-	-	-	7.751
T6	-	-	13.37	-	-	-	7.19	1.359	-
A7	-	7.335	-	-	-	-	-	-	8.233
C8	-	-	-	6.506	8.094	5.295	7.285	-	-
A9	-	7.549	-	-	-	-	-	-	8.208
T10	-	-	13.61	-	-	-	7.196	1.363	-
C11	-	-	-	6.95	8.528	5.658	7.465	-	-
G12	n.d	-	-	-	-	-	-	-	7.932

**Appendix 20.** <sup>13</sup>C base chemical shifts of free ODN4.

ODN4	C1'	C2	C2'	C3'	C4'	C5	C5'	C6	C7	C8
C1	85.23	-	37.27	75.11	85.53	96.8	n.d	140.4	-	-
G2	81.35	-	37.36	76.92	84.65	-	65.54	-	-	135.5
A3	82.31	152.2	38.46	76.81	84.89	-	63.92	-	-	138.6
T4	82.07	-	36.59	74.94	82.56	-	n.d	135.3	11.43	-
G5	82.01	-	38.35	76.08	84.12	-	n.d	-	-	135.5

<b>T6</b>	82.61	-	36.73	75.04	83.1	-	n.d	135.8	11.38	-
<b>A7</b>	82.09	151.9	38.04	76.57	84.49	-	n.d	-	-	138.9
<b>C8</b>	83.23	-	37.23	73.9	n.d.	95.27	n.d	139.5	-	-
<b>A9</b>	82.73	151.6	38.52	76.41	84.72	-	n.d	-	-	138.8
<b>T10</b>	82.41	-	36.52	74.09	82.87	-	n.d	135.8	11.39	-
<b>C11</b>	83.82	-	36.59	74.05	85.45	96.03	n.d	141.1	-	-
<b>G12</b>	82.05	-	39.15	70.79	85.28	-	n.d	-	-	136.5

**Appendix 21.** <sup>1</sup>H nucleotide base chemical shifts of **PA4•ODN4** complex.

<b>ODN4</b>	<b>H1</b>	<b>H2</b>	<b>H21</b>	<b>H22</b>	<b>H3</b>	<b>H41</b>	<b>H42</b>	<b>H5</b>	<b>H6</b>	<b>H61</b>	<b>H62</b>	<b>H7</b>	<b>H8</b>
<b>C1</b>	-	-	-	-	-	n.d.	n.d.	5.90	7.60	-	-	-	-
<b>G2</b>	12.88	-	-	-	-	-	-	-	-	-	-	-	8.03
<b>A3</b>	-	7.98	-	-	-	-	-	-	-	n.d.	n.d.	-	8.45
<b>T4</b>	-	-	-	-	13.74	-	-	-	7.43	-	-	1.55	-
<b>G5</b>	12.30	-	8.03	7.64	-	-	-	-	-	-	-	-	8.00
<b>T6</b>	-	-	-	-	12.72	-	-	-	6.97	-	-	1.29	-
<b>A7</b>	-	8.05	-	-	-	-	-	-	-	8.46	5.59	-	8.14
<b>C8</b>	-	-	-	-	-	8.75	6.43	5.27	6.88	-	-	-	-
<b>A9</b>	-	7.72	-	-	-	-	-	-	-	n.d.	n.d.	-	8.30
<b>T10</b>	-	-	-	-	13.54	-	-	-	6.97	-	-	1.45	-
<b>C11</b>	-	-	-	-	-	8.56	6.91	5.62	7.42	-	-	-	-
<b>G12</b>	n.d.	-	n.d.	n.d.	-	-	-	-	-	-	-	-	7.94
<b>C13</b>	-	-	-	-	-	n.d.	n.d.	5.93	7.66	-	-	-	-
<b>G14</b>	12.81	-	n.d.	n.d.	-	-	-	-	-	-	-	-	8.01
<b>A15</b>	-	7.90	-	-	-	-	-	-	-	n.d.	n.d.	-	8.33
<b>T16</b>	-	-	-	-	13.52	-	-	-	7.28	-	-	1.53	-
<b>G17</b>	12.75	-	8.56	7.32	-	-	-	-	-	-	-	-	8.02
<b>T18</b>	-	-	-	-	12.93	-	-	-	6.98	-	-	1.45	-
<b>A19</b>	-	8.03	-	-	-	-	-	-	-	8.64	5.57	-	8.15
<b>C20</b>	-	-	-	-	-	8.70	6.43	5.25	6.85	-	-	-	-
<b>A21</b>	-	7.73	-	-	-	-	-	-	-	n.d.	n.d.	-	8.27
<b>T22</b>	-	-	-	-	13.82	-	-	-	6.96	-	-	1.37	-
<b>C23</b>	-	-	-	-	-	8.63	6.95	5.66	7.47	-	-	-	-
<b>G24</b>	n.d.	-	n.d.	n.d.	-	-	-	-	-	-	-	-	7.95

**Appendix 22.** <sup>1</sup>H sugar backbone chemical shifts of **PA4•ODN4** complex.

<b>ODN4</b>	<b>H1'</b>	<b>H2'</b>	<b>H2''</b>	<b>H3'</b>	<b>H4'</b>	<b>H5'</b>	<b>H5''</b>
<b>C1</b>	5.790	1.886	2.385	4.717	4.077	3.708	3.743
<b>G2</b>	5.460	2.869	2.777	5.043	4.329	n.d.	n.d.
<b>A3</b>	6.367	2.834	2.834	5.137	4.525	n.d.	n.d.

<b>T4</b>	5.775	2.518	2.590	5.047	4.275	n.d.	n.d.
<b>G5</b>	6.020	2.587	2.674	5.035	4.243	n.d.	n.d.
<b>T6</b>	5.032	1.683	2.174	4.572	1.995	n.d.	n.d.
<b>A7</b>	5.533	2.213	2.701	4.673	2.863	n.d.	n.d.
<b>C8</b>	5.325	1.457	2.311	4.547	2.223	n.d.	n.d.
<b>A9</b>	5.574	2.175	2.571	4.663	2.675	n.d.	n.d.
<b>T10</b>	5.699	1.869	2.306	4.765	3.917	n.d.	n.d.
<b>C11</b>	5.748	1.970	2.369	4.824	4.116	n.d.	n.d.
<b>G12</b>	6.160	2.369	2.601	4.681	4.181	n.d.	n.d.
<b>C13</b>	5.767	1.988	2.441	4.732	4.077	n.d.	n.d.
<b>G14</b>	5.678	2.775	2.893	5.049	4.372	n.d.	n.d.
<b>A15</b>	6.335	2.712	2.993	5.109	4.512	n.d.	n.d.
<b>T16</b>	5.636	2.259	2.385	4.934	4.148	n.d.	n.d.
<b>G17</b>	5.796	2.654	2.654	5.005	4.273	n.d.	n.d.
<b>T18</b>	5.215	1.711	2.242	4.627	2.062	n.d.	n.d.
<b>A19</b>	5.526	2.191	2.712	4.667	2.645	n.d.	n.d.
<b>C20</b>	5.319	1.486	2.294	4.549	2.297	n.d.	n.d.
<b>A21</b>	5.446	2.157	2.488	4.634	2.682	n.d.	n.d.
<b>T22</b>	5.745	1.837	2.103	4.771	3.785	n.d.	n.d.
<b>C23</b>	5.510	2.095	2.318	4.781	4.117	n.d.	n.d.
<b>G24</b>	6.189	2.403	2.629	4.684	4.171	n.d.	n.d.

**Appendix 23.** <sup>1</sup>H PA4 chemical shifts of PA4•ODN4 complex (' and '' refer to the most upfield and the most downfield signal, respectively).

PA4	H2'	H2''	H3	H3'	H3''	H4	H4'	H4''	H5	CH <sub>3</sub>	NH	iPr1Me	iPr2Me	iPrCH
<b>Dp</b>	3.14	3.18	-	1.95	2.13	-	2.97	3.63	-	-	8.08	-	-	-
<b>β-Ala</b>	1.63	2.15	-	2.88	3.82	-	-	-	-	-	9.18	-	-	-
<b>Py1</b>	-	-	6.30	-	-	-	-	-	7.07	3.63	9.36	-	-	-
<b>Py2</b>	-	-	6.35	-	-	-	-	-	7.50	3.72	10.70	-	-	-
<b>Py3</b>	-	-	6.20	-	-	-	-	-	7.60	3.87	9.85	-	-	-
<b>Im4</b>	-	-	-	-	-	-	-	-	7.46	3.95	10.48	-	-	-
<b>Turn</b>	2.10	2.53	-	1.62	2.33	-	2.75	3.44	-	-	8.24	-	-	-
<b>Py5</b>	-	-	6.46	-	-	-	-	-	7.10	3.59	9.11	-	-	-
<b>Py6</b>	-	-	6.24	-	-	-	-	-	7.62	3.79	10.47	-	-	-
<b>Py7</b>	-	-	6.05	-	-	-	-	-	7.58	3.91	9.59	-	-	-
<b>Im8</b>	-	-	-	-	-	7.10	-	-	7.28	-	-	1.39	1.66	5.39

**Appendix 24.** <sup>13</sup>C PA4 chemical shifts of PA4•ODN4 complex.

PA4	C3	C4	C5	N-CH <sub>3</sub>	iPr1Me	iPr2Me	iPrCH
<b>Py1</b>	106.8	-	119.4	36.6	-	-	-
<b>Py2</b>	102.5	-	121.8	37.1	-	-	-

<b>Py3</b>	104.4	-	-	37.8	-	-	-
<b>Im4</b>	-	-	114.1	36.4	-	-	-
<b>Py5</b>	105.3	-	119.1	36.3	-	-	-
<b>Py6</b>	102.7	-	121.7	37.2	-	-	-
<b>Py7</b>	104.1	-	121.0	37.7	-	-	-
<b>Im8</b>	-	130.1	118.9	-	23.6	23.2	49.5

**Appendix 25.**  $^{13}\text{C}$  ODN4 chemical shifts of PA4•ODN4 complex.

ODN4	C1'	C2	C3'	C4'	C5	C6	C7	C8
<b>C1</b>	82.55	-	75.7	85.6	96.9	140.5	-	-
<b>G2</b>	81.42	-	76.4	84.7	-	-	-	135.7
<b>A3</b>	82.00	152.0	77.3	85.0	-	-	-	139.0
<b>T4</b>	83.36	-	75.5	82.8	-	136.2	-	-
<b>G5</b>	82.82	-	78.8	85.0	-	-	-	135.9
<b>T6</b>	81.42	-	73.1	79.8	-	136.5	-	-
<b>A7</b>	81.27	152.0	76.3	83.0	-	-	-	139.8
<b>C8</b>	81.60	-	72.9	80.6	96.4	139.8	-	-
<b>A9</b>	80.61	151.5	75.6	82.3	-	-	-	139.5
<b>T10</b>	81.48	-	73.0	81.7	-	135.9	-	-
<b>C11</b>	83.64	-	73.9	83.4	96.0	141.1	-	-
<b>G12</b>	82.06	-	70.8	85.3	-	-	-	136.6
<b>C13</b>	85.27	-	75.1	85.5	96.9	140.5	-	-
<b>G14</b>	81.38	-	77.2	84.7	-	-	-	137.0
<b>A15</b>	82.22	152.3	77.0	84.9	-	-	-	138.7
<b>T16</b>	82.30	-	73.2	82.2	-	136.1	-	-
<b>G17</b>	82.60	-	78.2	84.8	-	-	-	135.7
<b>T18</b>	82.00	-	74.0	81.4	-	136.3	-	-
<b>A19</b>	81.4	152.6	75.6	82.6	-	-	-	139.6
<b>C20</b>	81.55	-	72.9	80.8	96.1	139.7	-	-
<b>A21</b>	80.35	150.9	74.1	82.3	-	-	-	139.5
<b>T22</b>	81.5	-	75.0	82.0	-	135.9	-	-
<b>C23</b>	84.0	-	73.3	82.7	96.1	141.2	-	-
<b>G24</b>	82.2	-	70.9	85.3	-	-	-	135.7







## 4 Profiling of the binding kinetics

**Appendix 28.**  $K_D$  and rate constants from switchSENSE<sup>®</sup> experiments, along with melt temperatures recorded.

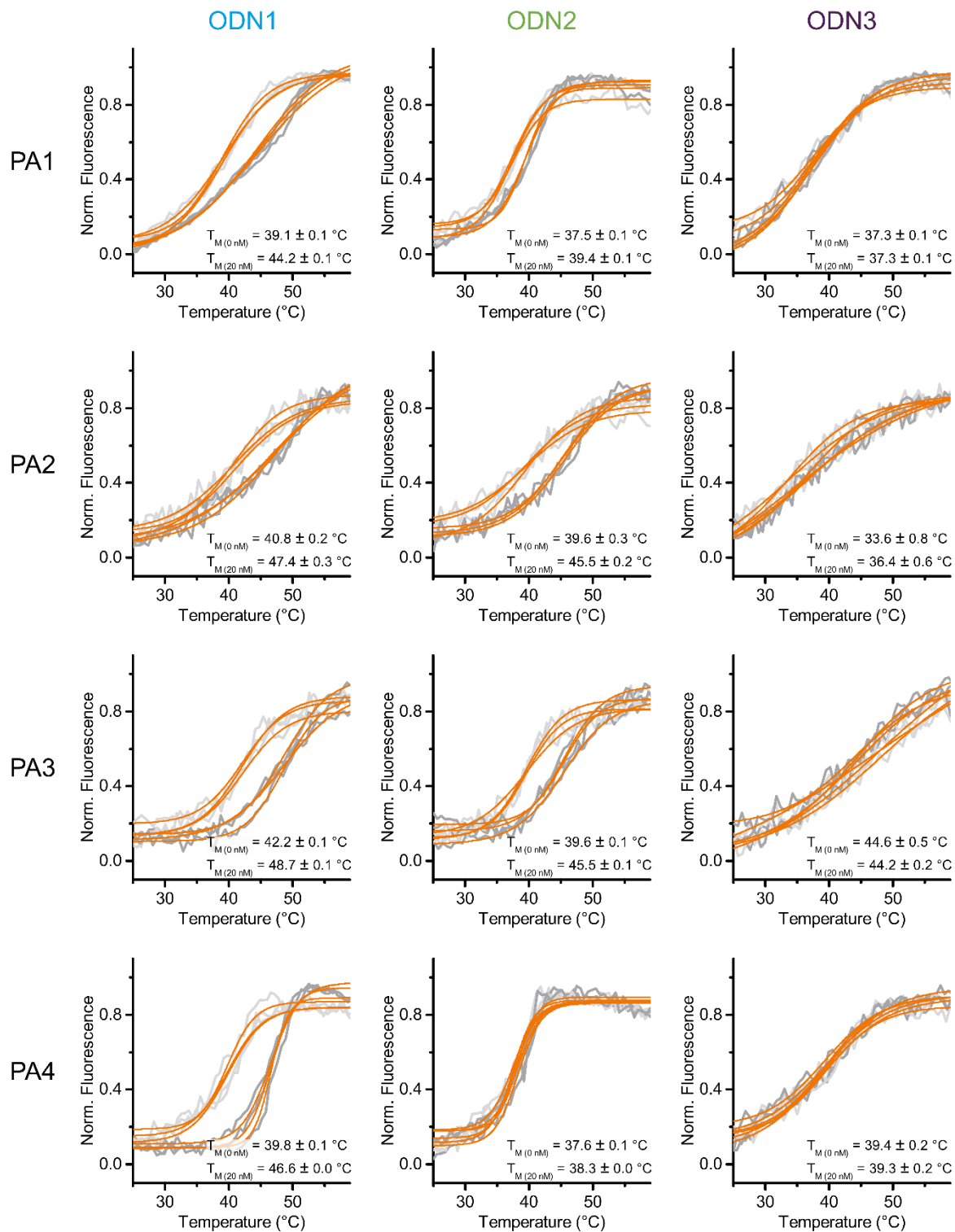
Sample	$k_{ON}$ ( $M^{-1}s^{-1}$ )	Error $k_{ON}$ ( $M^{-1}s^{-1}$ )	$k_{OFF}$ ( $s^{-1}$ )	Error $k_{OFF}$ ( $s^{-1}$ )	$K_D$ (pM)	Error (pM)	$K_D$	$\Delta T_m$ at 20 nM ( $^{\circ}C$ )	Error $\Delta T_m$ ( $^{\circ}C$ )
PA1 + ODN1	$7.49 \times 10^6$	$2.10 \times 10^5$	$1.91 \times 10^{-3}$	$3.00 \times 10^{-5}$	254	8		5.12	0.17
PA1 + ODN2	$5.37 \times 10^6$	$2.80 \times 10^5$	$7.07 \times 10^{-3}$	$1.00 \times 10^{-4}$	1320	70		1.97	0.13
PA1 + ODN3	No interaction	No interaction	No interaction	No interaction	No interaction	No interaction		0.00	0.29

Sample	$k_{ON}$ ( $M^{-1}s^{-1}$ )	Error $k_{ON}$ ( $M^{-1}s^{-1}$ )	$k_{OFF}$ ( $s^{-1}$ )	Error $k_{OFF}$ ( $s^{-1}$ )	$K_D$ (pM)	Error (pM)	$K_D$	$\Delta T_m$ at 20 nM ( $^{\circ}C$ )	Error $\Delta T_m$ ( $^{\circ}C$ )
PA2 + ODN1	$1.39 \times 10^6$	$8.00 \times 10^4$	$1.62 \times 10^{-3}$	$5.00 \times 10^{-5}$	1170	70		6.60	0.53
PA2 + ODN2	$1.15 \times 10^6$	$9.00 \times 10^4$	$1.44 \times 10^{-3}$	$6.00 \times 10^{-5}$	1250	110		5.93	0.40
PA2 + ODN3	$3.93 \times 10^5$	$1.96 \times 10^5$	$6.07 \times 10^{-3}$	$2.70 \times 10^{-4}$	15400	7700		2.87	1.37

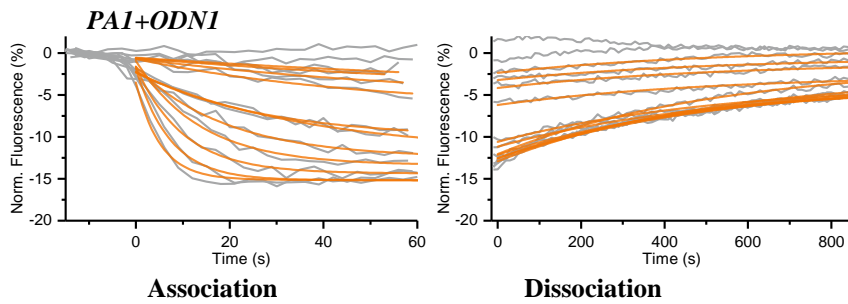
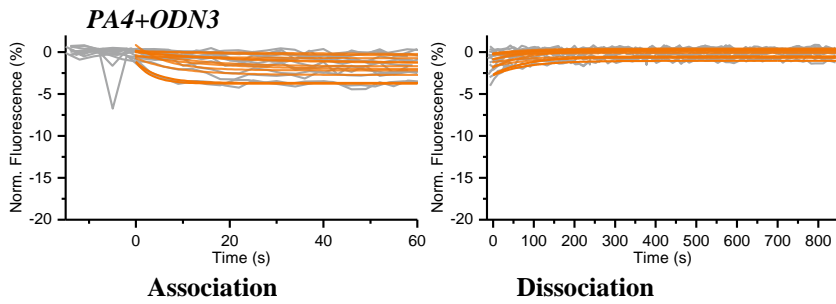
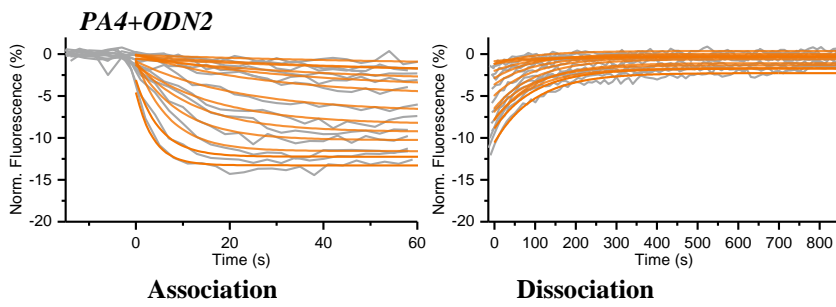
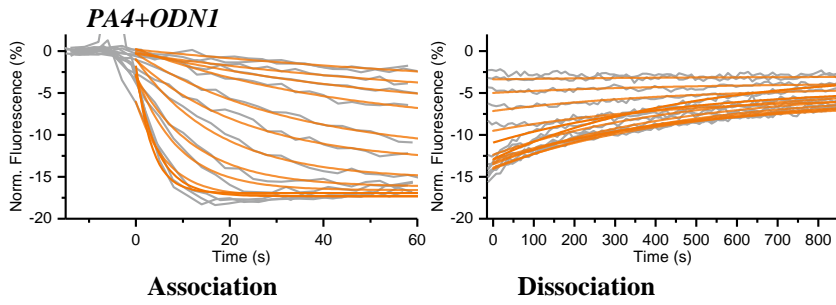
Sample	$k_{ON}$ ( $M^{-1}s^{-1}$ )	Error $k_{ON}$ ( $M^{-1}s^{-1}$ )	$k_{OFF}$ ( $s^{-1}$ )	Error $k_{OFF}$ ( $s^{-1}$ )	$K_D$ (pM)	Error (pM)	$K_D$	$\Delta T_m$ at 20 nM ( $^{\circ}C$ )	Error $\Delta T_m$ ( $^{\circ}C$ )
PA3 + ODN1	$9.92 \times 10^5$	$1.01 \times 10^5$	$1.96 \times 10^{-3}$	$1.20 \times 10^{-4}$	1970	240		6.50	0.22
PA3 + ODN2*	$3.74 \times 10^5$	$5.20 \times 10^4$	$1.08 \times 10^{-3}$	$7.00 \times 10^{-5}$	2880	440		5.81	0.23
PA3 + ODN3	No Interaction	No Interaction	No Interaction	No Interaction	No Interaction	No Interaction		-0.41	0.75

Sample	$k_{ON}$ ( $M^{-1}s^{-1}$ )	Error $k_{ON}$ ( $M^{-1}s^{-1}$ )	$k_{OFF}$ ( $s^{-1}$ )	Error $k_{OFF}$ ( $s^{-1}$ )	$K_D$ (pM)	Error (pM)	$K_D$	$\Delta T_m$ at 20 nM ( $^{\circ}C$ )	Error $\Delta T_m$ ( $^{\circ}C$ )
PA4 + ODN1	$1.18 \times 10^7$	$3.00 \times 10^5$	$2.22 \times 10^{-3}$	$3.00 \times 10^{-5}$	188	5		6.78	0.14
PA4 + ODN2	$1.04 \times 10^7$	$4.00 \times 10^5$	$1.01 \times 10^{-2}$	$1.00 \times 10^{-4}$	967	35		0.71	0.08
PA4 + ODN3	$1.12 \times 10^7$	$1.30 \times 10^6$	$1.23 \times 10^{-2}$	$6.00 \times 10^{-4}$	1100	100		-0.13	0.39

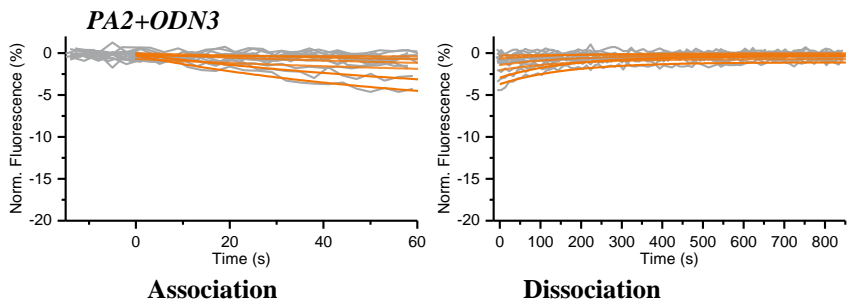
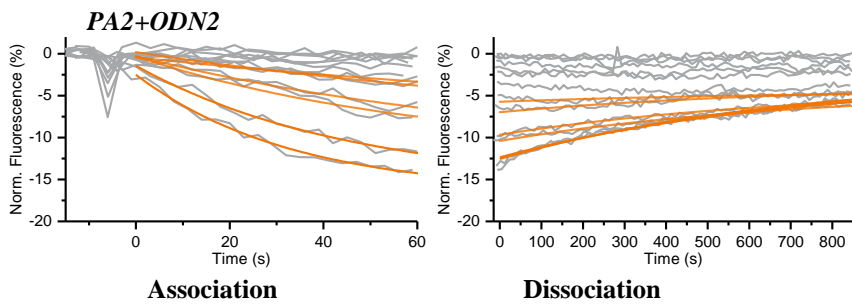
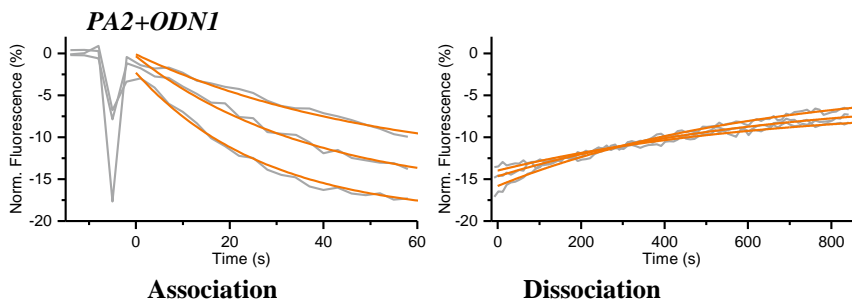
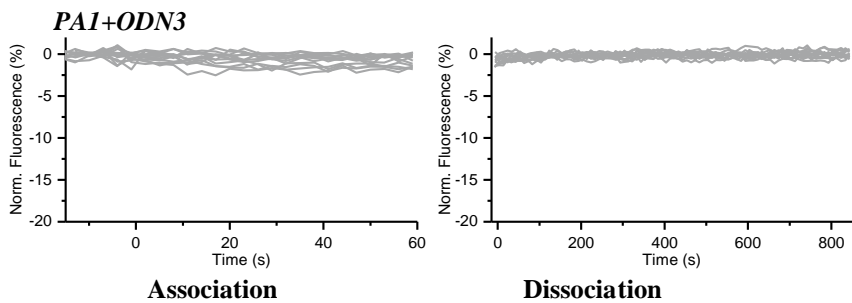
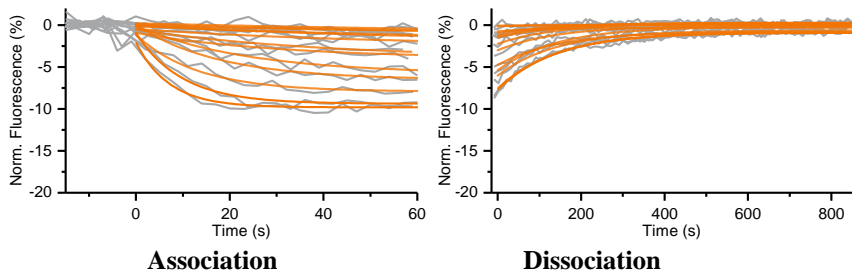
\* For PA3 + ODN2, an extended concentration regime was employed (up to 49 nM) as, despite evident association of PA3 to ODN2, the global exponential fit did not converge for the concentration regime used in the standard protocol

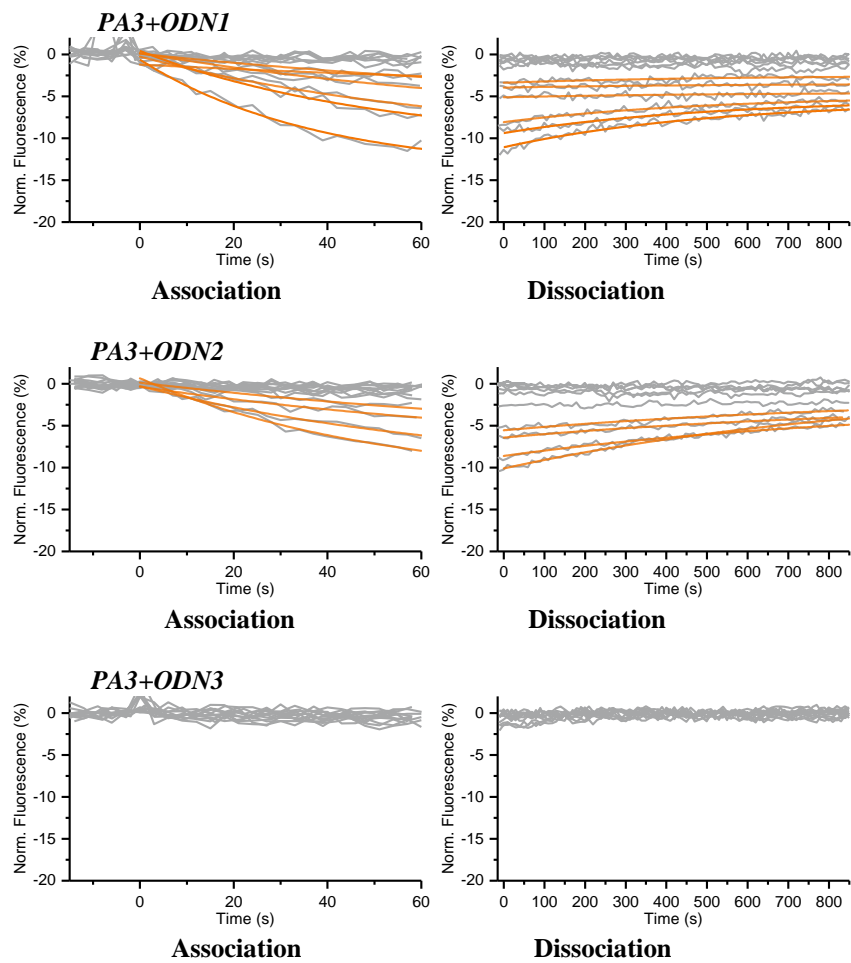


**Appendix 29.** Fluorescence melting curves of PA1-4 for ODN1-3 ( $\Delta T_m = T_{m, \text{complex}} - T_{m, \text{ODN1-3}}$ ; error bars were calculated from 3 independent runs each with 0 nM and 20 nM PA1-4).



*PA1+ODN2*





**Appendix 30.** Association and dissociation fluorescence response curves for **PA1-4** with **ODN1-3** at PA concentrations ranging from 450 pM to 25.5 nM (up to 49 nM in **PA3-ODN2** case).

## 5 Analysis of NMR-derived MD structure

**Appendix 31. PA4•ODN4** complex minor and major groove widths; direct P-P distances and refined P-P distances, which take into account the directions of the sugar-phosphate backbones (values were calculated from a minimized average structure obtained from the clustered conformations through X3DNA; subtracting 5.8 Å from the values shown will account for the Van der Waals radii of the phosphate groups, for comparison with FreeHelix and Curves)<sup>[90]</sup>.

	Minor Groove		Major Groove	
	P-P	Refined	P-P	Refined
1 CG/CG	---	---	---	---
2 GA/TC	---	---	---	---
3 AT/AT	11.6	---	16.3	---
4 TG/CA	13.6	13.6	15.7	14.9
5 GT/AC	14	14	15.1	14.6
6 TA/TA	13.5	13.4	12.7	12.6
7 AC/GT	13.7	13.6	14.7	14.7
8 CA/TG	14	---	17.6	---
9 AT/AT	---	---	---	---
10 TC/GA	---	---	---	---

**Appendix 32. PA4•ODN4** complex axis parameters (calculated values were calculated from a minimized average structure obtained from the clustered conformations using Curves+).

BP-Axis	Xdisp	Ydisp	Incline	Tip	Ax-bend
1) G2-C23	-1.47	0.97	2.9	10.5	---
2) A3-T22	-0.57	1.61	6	4.4	2.1
3) T4-A21	0.68	1.27	10	2.9	2.2
4) G5-C20	0.35	1.24	4.2	8.9	1.8
5) T6-A19	0.54	0.14	5.6	2	1.8
6) A7-T18	0.55	-0.79	5.3	3.5	2
7) C8-G17	0.2	-1.48	6.2	-4	1.2
8) A9-T16	0.48	-0.95	6.5	-2.7	0.6
9) T10-A15	0.08	-1.73	5.6	0.3	0.5
10) C11-G14	-1.71	-1.56	5.6	-2.9	0.6
Average:	-0.09	-0.13	5.8	2.3	Total bend = <b>11.8°</b> (1-10)

**Appendix 33. PA4•ODN4** complex local base pair parameters (calculated values were calculated from a minimized average structure obtained from the clustered conformations through X3DNA).

Base pair	Shear	Stretch	Stagger	Buckle	Propeller	Opening
1 C-G	0.3	-0.15	-0.53	25.7	-6.02	-1.19
2 G-C	-0.51	-0.25	-0.24	-1.49	-9.97	-0.02

3 A-T	0.12	-0.04	0.07	4	-9.08	2.29
4 T-A	-0.29	-0.02	-0.07	15.25	-16.81	1.41
5 G-C	0	-0.09	0.07	-1.26	-0.13	-3.88
6 T-A	-0.3	0.11	0.09	6.06	-15.05	-13.63
7 A-T	-0.03	-0.06	0.16	1.89	-21.63	-9.23
8 C-G	-0.34	-0.07	0.06	-0.17	-6.88	-2.55
9 A-T	0.26	-0.07	0.24	1.45	-8.08	1.75
10 T-A	-0.15	-0.04	-0.06	12.03	6.8	-1.18
11 C-G	0.4	-0.21	-0.52	20.41	-6.14	0.04
~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
ave.	-0.05	-0.08	-0.07	7.62	-8.45	-2.38
s.d.	0.3	0.1	0.26	9.37	7.78	4.93

**Appendix 34. PA4•ODN4 complex local base pair parameters** (calculated values were calculated from a minimized average structure obtained from the clustered conformations through X3DNA).

Step	Shift	Slide	Rise	Tilt	Roll	Twist
1 CG/CG	-1.39	-1.4	4.09	-9.35	5.95	31.98
2 GA/TC	-0.48	-0.42	3.18	-3.7	-2.15	40.99
3 AT/AT	0.03	-0.48	2.98	-0.57	2.25	25.57
4 TG/CA	-2.1	0.52	3.4	-9.95	13.29	42.25
5 GT/AC	-1.03	-0.06	2.99	-1.74	-1.08	30.81
6 TA/TA	0.01	0.47	3.14	-1.27	6	36.14
7 AC/GT	0.75	0.52	3.14	2.59	-2.81	34.77
8 CA/TG	1.9	1.14	3.13	2.07	6.65	41.84
9 AT/AT	0.81	-0.67	3.28	0.03	4.83	24.8
10 TC/GA	-0.26	-0.6	3.18	1.97	0.44	35.71
~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
ave.	-0.17	-0.1	3.25	-1.99	3.34	34.49
s.d.	1.16	0.76	0.32	4.48	4.98	6.27

**Appendix 35. Phosphate backbone angle values of PA4•ODN4 complex** (calculated values were calculated from a minimized average structure obtained from the clustered conformations through X3DNA).

**Strand I**

Base	alpha	beta	gamma	delta	epsilon	zeta	chi
1 C	---	---	56.9	138.5	-151.5	-130	-119.5
2 G	-77.6	-177.9	50.1	144.6	-177.4	-98.1	-127.6
3 A	-68.5	-168.1	52.9	141	-170.3	-90.8	-106.1
4 T	-73.9	169.2	53.9	139.1	-96	179	-95.9
5 G	-79.3	173.3	27.8	144.2	-179.4	-93.1	-102.1
6 T	-64.3	169.7	58.5	108.1	-175.3	-89.7	-117.3
7 A	-63.2	166.2	57.4	111.2	-176.9	-87.9	-116
8 C	-59.6	166.9	58	121.4	-166.9	-87.5	-110.5



9 A	-65.6	159.8	59.1	114.6	-171.1	-88.3	-111.9
10 T	-65.6	168.9	51	136.2	-131.2	-167.5	-103.8
11 C	-66.4	144.4	54.6	138.3	-168.7	-113	-117

**Strand II**

<b>Base</b>	<b>alpha</b>	<b>beta</b>	<b>gamma</b>	<b>delta</b>	<b>epsilon</b>	<b>zeta</b>	<b>chi</b>
1 G	-70.3	-178.8	51.4	122.7	---	---	-123
2 C	-66.8	176.7	59	142.6	-179.4	-89	-109.5
3 T	-60.2	165	66.5	120.3	-177	-91.5	-129
4 A	-69.1	156.8	57.1	88	176.2	-80.5	-131
5 C	-59.3	168.7	57.4	122.3	-165.8	-87	-115.2
6 A	-64.8	165.9	58.5	118.1	-179.3	-89	-114.7
7 T	-68.2	163.7	60.8	106.4	-173.6	-86.8	-117.7
8 G	-69	167.8	50.9	142.4	-174.3	-101.7	-94.8
9 T	-71.5	167.5	52.9	141.7	-129.4	-166.9	-104.8
10 A	-81.4	175.8	48.6	134.2	-148.5	-139.9	-111.2
11 G	-68.4	-170.6	58.8	138.7	-158	-77.4	-150