



Expanding the DNA binding repertoire of Pyrrole-Imidazole polyamides

Giacomo Padroni

Thesis submitted for the degree of

Doctor of Philosophy

29th March 2018

Academic Supervisor: Dr Glenn 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: _____ (Giacomo Padroni)

Date: 29th March 2018

Contents

Acknowledgements.....	8
Abstract.....	9
Abbreviations.....	11
Chapter 1.....	17
Introduction.....	17
1.1 Significance of molecular recognition of nucleic acids.....	18
1.2 Structure of DNA.....	19
1.3 Molecular recognition of dsDNA.....	22
1.3.1 Structural motifs of DNA binding proteins.....	24
1.4 Transcription of nucleosomal DNA.....	26
1.5 RNA polymerase.....	27
1.6 Regulation of transcriptional elongation.....	29
1.6.1 Structure and function of nuclear receptors.....	31
1.7 Exogenous regulation of gene expression.....	34
1.7.1 Triplex-forming oligonucleotide and peptide nucleic acids.....	35
1.7.2 Minor groove binders as regulators of gene expression.....	38
1.8 Downregulation of genes controlled by the Androgen Receptor using polyamides..	51
1.9 Current limitations of Py-Im polyamide designs targeting the AR.....	56
1.10 Expanding heterocyclic diversity of Py-Im polyamides.....	59

1.10.1	Thiazole containing MGBs	61
1.11	Hypothesis.....	65
1.12	Aims and Objectives	66
Chapter 2	68
Synthesis of thiazole-containing polyamides	68
2.1	Synthetic strategies for the preparation of polyamides	69
2.1.1	Biosynthesis of distamycin A and netropsin.....	69
2.1.2	Chemical synthesis of distamycin A.....	71
2.2	Solid phase synthesis of Py-Im polyamides.....	72
2.2.1	Synthesis of monomer building blocks.....	73
2.2.2	Choice of the support in solid phase synthesis of polyamides.....	76
2.2.3	Limitations of solid phase synthesis methods.....	79
2.3	Incorporation of Nt in MGBs.....	84
2.4	Aims and objectives of this chapter	86
2.5	Results and discussion	87
2.5.1	Synthesis of Boc-protected Py and Im amino acid monomers.....	87
2.5.2	Optimisation of standard solid phase synthesis protocols.....	88
2.5.3	Preparation of polyamides incorporating Nt at the N-terminus (position 8)..	93
2.5.4	Attempts to prepare polyamides containing Nt at an internal position (position 4)	98
2.5.5	Preparation of polyamides incorporating Nt at an internal position (position 6)	104

2.5.6	Preparation of Im containing polyamides	109
2.6	Summary	110
2.7	Experimental	111
2.7.1	General procedure	111
2.7.2	NMR spectroscopy.....	111
2.7.3	HPLC experiments.....	112
2.7.4	IR data and melting points	112
2.7.5	Mass spectrometry	112
2.7.6	Synthetic procedures for small molecules.....	112
2.7.7	General procedure for synthesis of polyamides	131
Chapter 3	137
Assessing the DNA binding profile of Nt-containing polyamides	137
3.1	Determination of MGBs interaction with dsDNA	138
3.2	Techniques used to study polyamide·dsDNA complexes.....	139
3.2.1	UV melting analysis.....	139
3.2.2	DNase I footprinting	141
3.2.3	Isothermal titration calorimetry (ITC)	144
3.2.4	Determining the aggregation properties of MGBs in aqueous solution.....	146
3.2.5	Surface plasmon resonance	148
3.2.6	SwitchSENSE® method to determine kinetic parameters of protein·DNA interactions.....	150
3.2.7	Circular Dichroism.....	151

3.3	Aims and objectives of this chapter	153
3.4	Results and Discussion	155
3.4.1	Binding characteristics of polyamides containing Nt in position 8	155
3.4.2	Binding characteristics of polyamides containing Nt in position 6	173
3.5	Summary	176
3.6	Experimental	179
3.6.1	UV melting analysis.....	179
3.6.2	DNase I Footprinting	179
3.6.3	Circular Dichroism.....	180
Chapter 4	181
Structural basis of Nt-containing polyamides binding to their target dsDNA	181
4.1	NMR structural determination of MGB·dsDNA complexes	182
4.1.1	Principles of Nuclear Magnetic Resonance	182
4.1.2	1D NMR studies of ligand-dsDNA interactions	185
4.1.3	2D-NMR and structural assignment of dsDNA	188
4.1.4	Principles of molecular dynamics and NMR-structural determination.....	193
4.2	Aims and objectives of this chapter	196
4.3	Results and Discussion	197
4.3.1	NMR assignment of ODN1.....	197
4.3.2	Titration of polyamides into a dsDNA solution.....	201
4.3.3	Assignment of PA·ODN1 complexes	202
4.3.4	Chemical shift analysis of binding mode of PA·ODN1 complexes.....	207

4.3.5	Comparative nOe analysis of the polyamides binding mode when in complex with ODN1	211
4.3.6	NMR-restrained molecular dynamics of PA·ODN1 complexes.....	213
4.3.7	Rationalisation of the differences in duplex distortion of Nt-containing polyamides bound to the target ODN1.....	218
4.4	Summary	220
4.5	Experimental.....	220
Chapter 5	226
Future directions towards the development of polyamides as gene regulatory agents		226
References.....		230
Appendix.....		251

Acknowledgements

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

I would like to thank Dr Glenn Burley for giving me the opportunity to work on this project and the support throughout the years. I would also like to thank every person, who contributed to this project. In particular, I would like to thank Prof. Keith Fox for the DNase footprinting, Dr Thomas Welte for developing the SwitchSENSE assay, Dr John Parkinson for introducing me to NMR spectroscopy of nucleic acids, the acquisition of the data and the helpful suggestions; and Khalid Aman for the immense synthetic effort.

I would also like to acknowledge present and past members (2014-2018) of the Burley group. In particular, I would like to thank Olivia for the infinite help in and outside the lab, Steven for the “open-minding” conversations and for proof-reading this document, and finally, Iain for making coffee.

I would like to thank my parents for the unconditional support in all my life choices. Finally, some special thanks go to Josephine for the love and support demonstrated every moment of the last two years.

Abstract

Pyrrole-Imidazole (Py-Im) hairpin polyamides are a programmable class of compounds that bind in the minor groove of double-stranded DNA (dsDNA). The hairpin conformation enables sufficient flexibility to form a side-by-side arrangement within the minor groove of target dsDNA sequences. An Im-Py pair discriminates G over C, A and T, whereas a Py-Py pair binds A and T over C and G. The possibility to target specific dsDNA sequences enables polyamides to modulate gene expression by the disruption of transcription factors such as the androgen receptor, overexpressed in prostate cancer. A current limitation of the biological applications of Py-Im polyamides is their variable cell permeability profile. Im-rich polyamides tend to have reduced cellular and nuclear uptake in eukaryotic cells. Although modifications to the periphery of the polyamide scaffold have been explored at length, there has been limited studies on tuning the physicochemical properties of the G-recognising Im unit.

This thesis describes the preparation and the evaluation of the binding mode of hairpin polyamides with Im units replaced by a series of 5-substituted 2-amino-4-carboxylic thiazole (Nt) building blocks.

Chapter 1 introduces the endogenous and exogenous DNA recognition by protein and small molecules, the modulation of gene expression using Py-Im polyamides and the limitations of this approach.

Chapter 2 describes the preparation of the building blocks and the optimisation of the solid phase synthesis of hairpin polyamides containing Nt units. These optimised conditions were used for the preparation of a suite of 8-ring polyamides incorporating Nt in different positions of the polyamide scaffold.

Chapter 3 describes the evaluation of the binding mode of Nt-containing polyamides using a combination of biophysical and biochemical methods. The 5-isopropyl substituted Nt (^{iPr}Nt) building block was identified as an alternative to Im for targeting G when incorporated at the N-terminus of the polyamide scaffold.

Chapter 4 describes the structural characterisation of three selected polyamides binding to their dsDNA target sequence using NMR spectroscopy and molecular dynamics. This study reveals that the polyamide containing ^{iPr}Nt at the N-terminus induces a greater compression of the major groove compared to the Im-containing analogue.

Finally, Chapter 5 reflects on the work of this thesis and offers some perspective of the future development of Py-Im polyamides as gene expression modulators.

Abbreviations

AP1	Activator protein 1
AR	Androgen receptor
ARE	Androgen responsive element
B ₀	NMR magnetic field
Boc	<i>tert</i> -Butyl carbamate
Boc-β-Ala-OH	Boc-β-alanine
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	4-(<i>tert</i> -Butoxycarbonylamino)butyric acid
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 ₂ 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 ₆	Deuterated Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
Dp	Dimethylaminopropylamine
DQF-COSY	Double-quantum filtered COSY
dsDNA	Double-stranded DNA
DPPA	Diphenylphosphoryl azide
EC	Elongation complex
EDC	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide
ER	Estrogen receptor
ERG	ETS-related gene
ESI	Electrospray Ionisation
ESP	Electrostatic surface potential
EtOAc	Ethyl Acetate
EtOH	Ethanol
ETS	Erythroblast transformation-specific
FID	Free Induction Decay
Fmoc	Fluorenylmethyloxycarbonyl protecting group
Fr	Furan building block
GABA	γ -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 rates for discerning geometry in solution
MD	Molecular dynamics
MeCN	Acetonitrile
MeOH	Methanol
MGB	Minor groove binder
MOM	2' Methoxy-ethyl ether
mRNA	RNA messenger
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
PMO	Phosphorodiamidate Morpholino Oligomer
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
RU	Resonance Unit
SAM	S-adenosyl methionine
SPR	Surface Plasmonic 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
TFMSA	Trifluoromethanesulfonic acid
TFO	Triplex-forming oligonucleotide

Th	5-carboxylic thiazole building block
THF	Tetrahydrofuran
TIPS	Triisopropylsilyl ether
T_m	Melting temperature
TMPRSS2	Transmembrane protease, serine 2
Tn	Methyl Thiophene building block
TOCSY	Total correlation spectroscopy
Tp	Thiophene building block
Tp-Cl	3-chloro thiophene building block
Tp-OMe	3-methoxy thiophene building block
TPW	TFA/phenol/water cleavage cocktail
TSS	Transcription Starting Site
UV	Ultraviolet
VCaP	Vertebral cancer of the prostate

Chapter 1

Introduction

1.1 Significance of molecular recognition of nucleic acids

The functions and processes of living organisms are almost exclusively regulated by selective interactions between two or more molecular components.¹ The ability to understand and modulate these interactions provides a means to control biological activities, which can be related to a disease. Small- to medium-sized synthetic ligands are often used both as tools for understanding molecular processes and as therapeutic agents to correct pathological states.²⁻⁴ These ligands can be rationally designed taking into account the forces involved in the binding. The ensemble of non-covalent interactions used by synthetic ligands for the selective recognition of macromolecules includes hydrogen bonding, van der Waals, dipole-dipole and dipole-induced-dipole interactions as well as hydrophobic interactions. The surrounding aqueous environment also influences the binding with water molecules capable of contributing both enthalpically (*i.e.*, bridging hydrogen bonds) and entropically (*i.e.*, by the displacement of ordered hydration states).⁵ Nature has often inspired chemists to design ligands that provide a complementary match of the shape of the target macromolecules. In the case of dsDNA, the natural products netropsin and distamycin A represent the architectural starting point for the development of programmable compounds for the control of transcription and replication.⁶⁻⁸ The molecular design of these ligands has been enormously enhanced by the advent of techniques for structural visualisation such as X-ray crystallography, nuclear magnetic resonance and more recently cryo-electron microscopy.⁹⁻¹⁵ This complete set of tools allowed the generation of pyrrole-imidazole (Py-Im) polyamides, a class of compounds with exquisite binding affinity and selectivity for dsDNA.

By virtue of the ability to synthetically program the target sequence, Py-Im polyamides offer an appealing alternative to oligonucleotide strategies for controlling gene expression.^{2,16} One of the remaining challenges for the development of the ideal DNA binders is identifying

optimal drug-properties in order to ensure cellular and nuclear permeation, appropriate tissue distribution and reduced toxicity.

1.2 Structure of DNA

DNA is an oligomer composed of monomers called nucleotides. Each nucleotide is comprised of a phosphate group, a 2'-deoxyribose sugar and a nitrogenous base (Figure 1.1).¹⁷ The four bases in DNA are the purines (guanine (G) and adenine (A)), and the pyrimidines (cytosine (C) and thymine (T)). Although a variety of higher order DNA structures are possible, the most common in biological systems is a right-handed helix, called B-DNA. This helical structure is composed of two antiparallel strands, each characterised by a sugar-phosphate backbone, in which the 2'-deoxyribose is linked to the phosphates in both the 5' and 3' positions. The nitrogenous base is covalently attached to the 2'-deoxyribose at the C1' via an N-glycosidic bond (Figure 1.1a). Bases of opposite strands pair with one another by hydrogen bonding that occurs between a purine and a pyrimidine. Watson-Crick pairing involves A pairing with T, and C pairing with G (Figure 1.1b).¹⁸

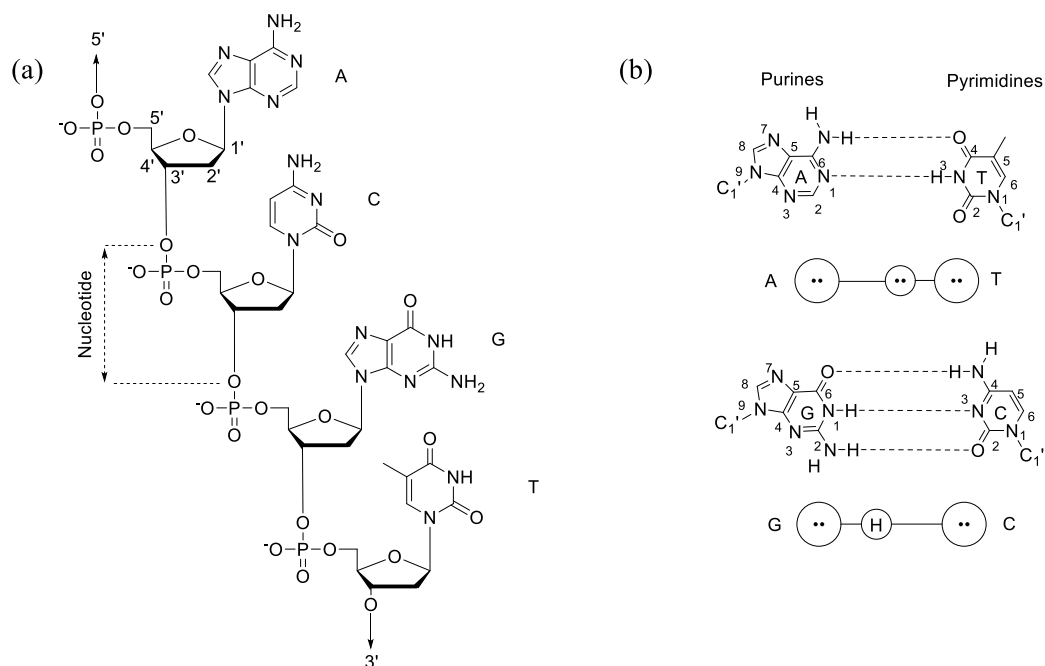


Figure 1.1 (a) Schematic representation of the primary structure of DNA. (b) Overview of Watson-Crick base pairing.

The right-handed helical conformation of dsDNA is a result of non-covalent interactions, which optimise the stacking of the hydrophobic base pairs, thereby reducing solvent exposure.¹⁹ B-DNA has a complete helical turn every 10 nucleotides with the strands assuming an asymmetric arrangement, resulting in differently sized grooves, termed the major and minor grooves (Figure 1.2a).²⁰ Other structures such as A-DNA and Z-DNA are formed under different hydration states or upon interaction with proteins through conformational rearrangement of the flexible sugar-phosphate backbone (Figure 1.2b-c).^{21,22} For example, in an ideal A-DNA, the preferential pucker of the deoxyribose is C3' endo (north) opposite to the C2' endo normally observed for B-DNA (south) (Figure 1.3).²³ This results in a reduced twist between the base pairs of A-DNA, which is more compact compared to B-DNA with a complete helical turn every 11 base pairs. A structural hallmark of A-DNA is a base pair inclination of approximately 20° relative to the DNA axis, which induces a narrower major groove and wider minor groove.²⁴ A very similar conformation is normally observed in RNA

double helices, where the backbone flexibility is limited by the presence of an hydroxyl group at C2', which locks the sugar pucker in a C3' endo conformation.²² DNA can also assume a left-handed helical conformation called Z-DNA.^{25,26} This is mostly limited to dsDNA sequences containing alternating purines and pyrimidines tracts and is characterised by a zigzag conformation of the backbone, which results from close proximity of the phosphates groups. For this reason, high ionic strength and protein binding are generally necessary to induce Z-DNA and compensate for electrostatic repulsion.

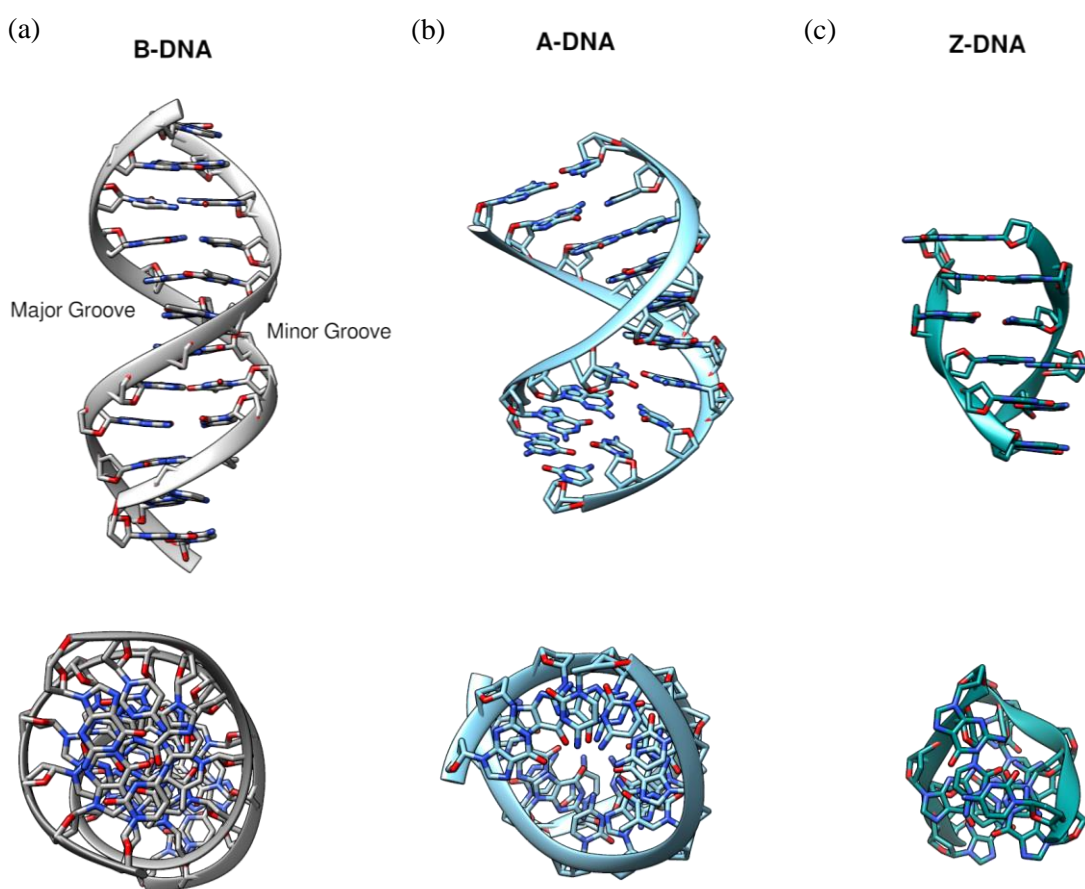


Figure 1.2 Lateral and top view of dsDNA structures: (a) B-DNA (x-ray, PDB 4C64),²⁷ (b) A-DNA (x-ray, PDB 41ZQ), (c) Z-DNA (x-ray, PDB 1W0E).²⁸

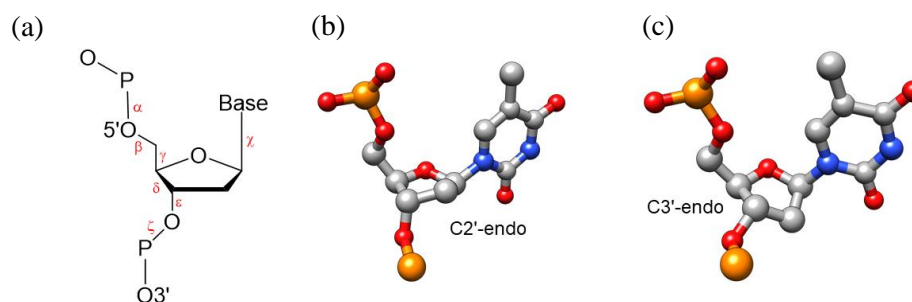


Figure 1.3 (a) DNA sugar torsion angles labelling convention. (b) Schematic representation of the C2' endo (south) and (c) C3' endo (north) sugar pucker conformations found in B and A- DNA, respectively.

1.3 Molecular recognition of dsDNA

The human genome is composed of 3.2 billion base pairs, encoding all the information necessary to maintain the normal function of the human body.²⁹ Gene expression is the process to generate a final protein product, which is achieved via transcription of DNA to RNA followed by translation of this RNA intermediate into protein. Fundamental for this process is the selective recognition of dsDNA sequences by DNA binding/processing proteins such as enzymes and transcription factors. Protein-DNA interactions can occur within the major or the minor groove of dsDNA as well as through electrostatic interactions with the phosphodiester backbone.¹

Nature achieves selective binding to dsDNA through a combination of base and shape readout.¹ Base readout consists of the selective recognition of the base edges and is driven by an interplay of hydrophobic contacts and hydrogen bonding (Figure 1.4-5).¹ The unique acceptor/donor pattern of each base in the major groove is normally utilised for sequence-specific base readout by proteins.¹ The minor groove lacks this exclusive recognition pattern, therefore interactions are generally determined by shape readout or by contact with the hydration spine present in certain regions of the narrow groove.

Shape readout is based on the structural deviations from the ideal B-DNA assumed by the double helix locally or globally. Increased narrowing of the minor groove over a particular sequence of DNA represents one of the main examples of dsDNA binding based on shape readout.^{1,30-33} The negative electrostatic potential of the minor groove in A/T regions of DNA derives from shape-induced electrostatic focusing and is generally exploited for specific recognition of this sequence by proteins through positively charged amino acid residues such as arginine and lysine.¹ Others shape variations such as kinks (locally) and bending (globally) are used for specific recognition of dsDNA by proteins.¹

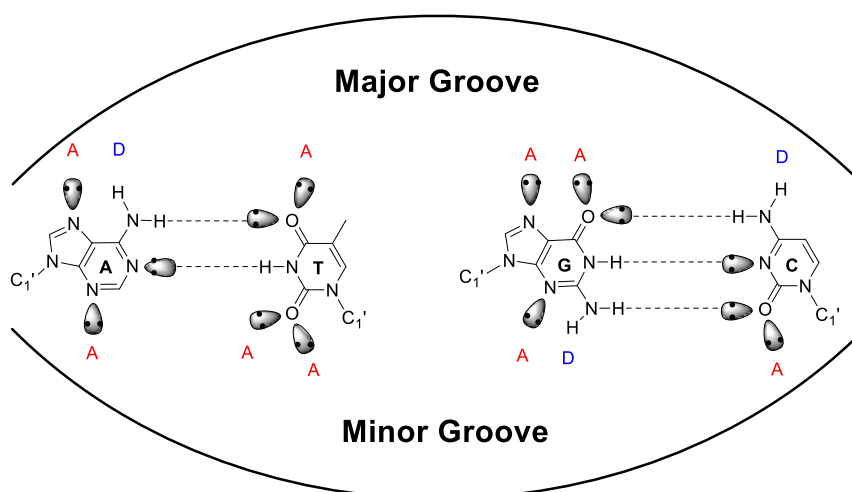


Figure 1.4 Schematic representation of the hydrogen bond acceptor/donor patterns of the major and minor groove edges of Watson-Crick base pairs.

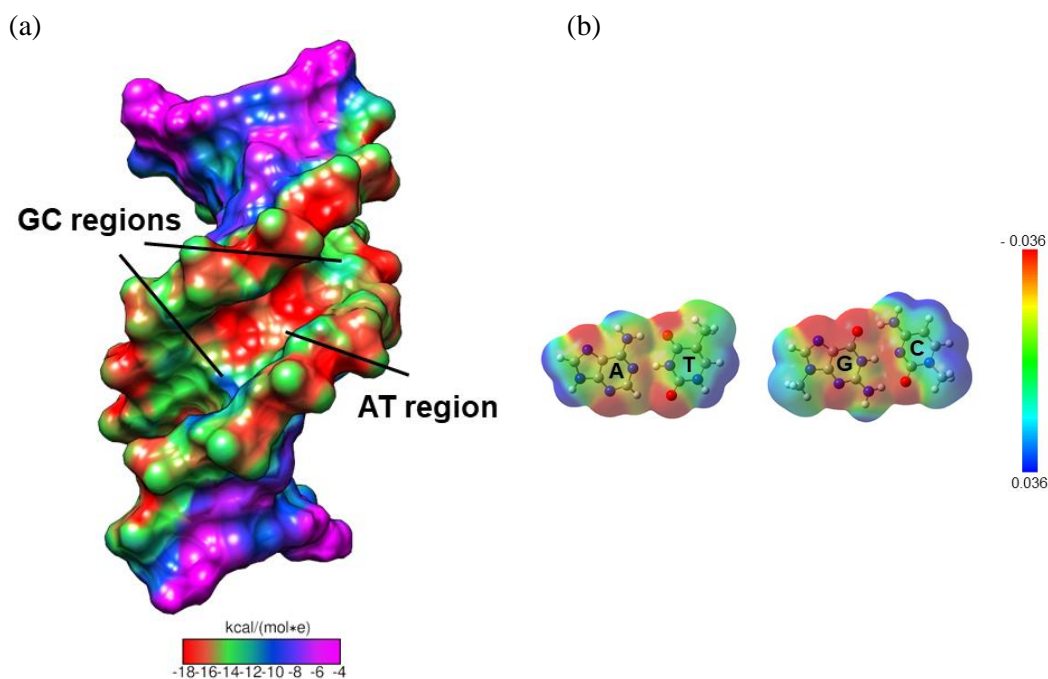


Figure 1.5 (a) Coulombic electrostatic surface potential (ESP) of the B-DNA dodecamer solved in Chapter 4 of this thesis. (b) ESP of Watson-Crick base pairs.

1.3.1 Structural motifs of DNA binding proteins

Proteins recognising dsDNA sequences are often characterised by common structural features in the DNA binding domain (DBD) such as zinc fingers, leucine zippers and TATA-box binding proteins (TBP).³⁴ Zinc fingers are motifs composed of 20-30 amino acid residues, which typically fold in a $\beta\beta\alpha$ motif (Figure 1.6). The structure is stabilised by histidine and cysteine residues, which act as coordinating sites for Zn^{2+} . Generally, a 3-base pair sequence is recognised by particular amino acid residues located on the top of the α -helix through hydrogen bonding and hydrophobic interactions with the dsDNA major groove. Zinc fingers are used as 3-base pair recognition modules that can be combined to provide selectivity for longer sequences.^{35,36}

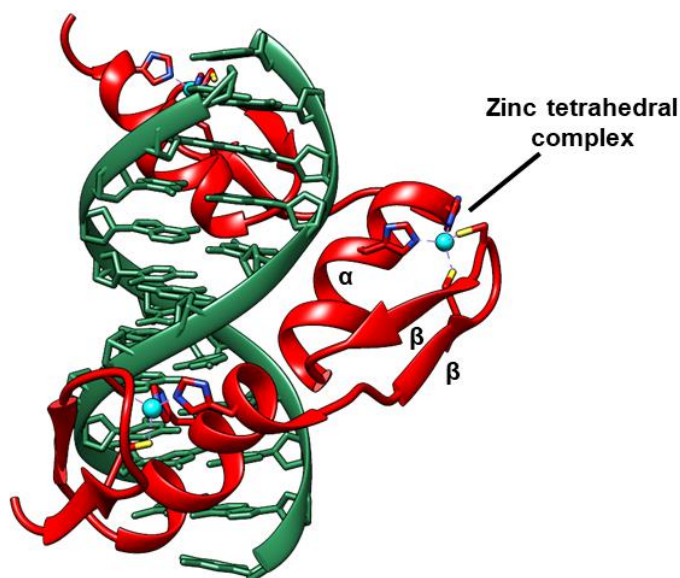


Figure 1.6 Crystal structure of a classical zinc finger interacting with the major groove of dsDNA (PDB 1ZAA).³⁶

Leucine zippers are dsDNA-binding dimeric motifs present in several transcription factors such as Activator Protein 1 (AP1).³⁷ They interact with the dsDNA major groove via their N-terminal amino acids. The 60-80 amino acid residues of two α -helical monomers are generally composed of 3-4 repeating units, formed by 7 consecutive leucine residues (Figure 1.7). The scissor shape of the dimeric leucine zipper is mediated by hydrophobic contacts between non-polar amino acid residues of each monomer in which the driving force of dimerisation is the reduction of solvent exposure.³⁸



Figure 1.7 Crystal structure of a leucine zipper bound to dsDNA (PDB 1YSA).³⁸

In contrast to the dsDNA binding of zinc fingers and leucine zippers, TBPs recognise dsDNA through shape readout of the narrow minor groove of A/T regions. Further interaction involves nonspecific hydrogen bonding between the β -sheet portion of the protein domain and the DNA phosphate backbone. As will be discussed in Section 1.6, this class of proteins is essential for transcriptional initiation as they induce large DNA distortions, which result in weaker inter-strand interactions leading to dsDNA unwinding and facilitating mRNA elongation.⁴

1.4 Transcription of nucleosomal DNA

Transcription is the process that generates a cognate sequence of mRNA from a corresponding ssDNA template. The sequence information in mRNA is then translated into a primary protein sequence by the ribosomes.³⁹ In eukaryotic cells, the initial step of transcription involves the unwinding of the nucleosome complexes formed by DNA and histones. In this particular state, 146 base pairs of dsDNA are wrapped around the nucleosome core (composed of 8 histone proteins: H2A, H2B, H3 and H4) involving roughly 120 DNA-protein interactions, which include electrostatic interactions and hydrogen bonding between the phosphate backbone of

DNA and the amides of the proteins (Figure 1.8).⁴⁰ This strong contact is also reinforced by a suite of interactions mediated by water molecules that compensate for the local sequence-dependent dsDNA distortion, and by the shape readout through highly conserved arginine residues that insert in 12 of the 14 minor groove portions facing toward the protein surface. Nucleosomes are connected via an 80-base pair linker DNA and stabilised by other proteins such as histone H1.^{41,42} Unwrapping this structure represents the first step toward transcription initiation. Enzymes such as histones acetyltransferases (HATs) and histone deacetylases (HDACs) add or remove acetyl groups from the amino moieties of lysine residue side chains, thereby modulating these protein-dsDNA electrostatic interactions.

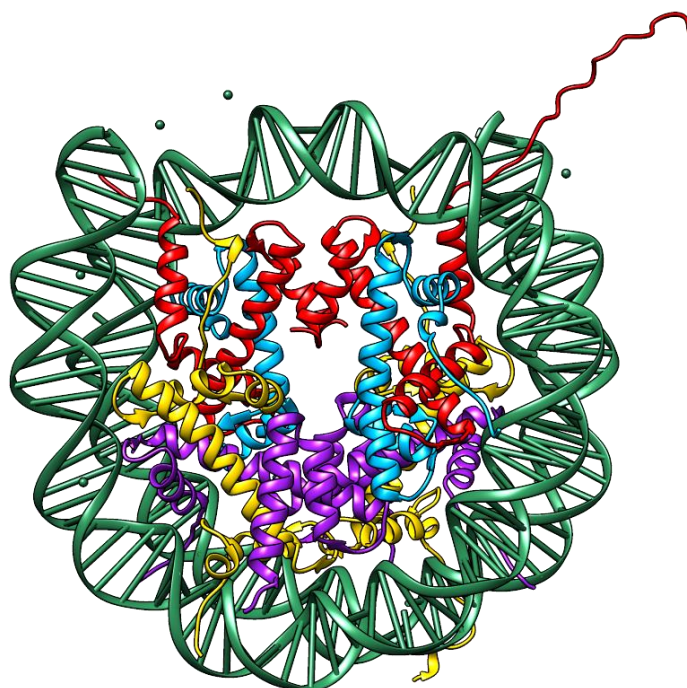


Figure 1.8 Crystal structure of the nucleosomal DNA (green) organised around the octameric histonic complex (red: H3, blue: H4, yellow: H2A and purple: H2B) (PDB 1AOI).⁴³

1.5 RNA polymerase

RNA production from decondensed dsDNA is mediated by RNA polymerase. Three main

types of RNA polymerases are found in mammals: Pol I, II and III, responsible for the production of ribosomal, messenger and transfer RNA, respectively.^{44,45} Although the structures and molecular mechanisms of RNA polymerases are still not completely understood, the advent of cryo-EM is now enhancing our knowledge of these molecular machines.⁴⁶ The core enzyme is formed by 10 protein subunits (Rbp) and is complemented by a heterodimeric complex formed of Rbp4 and Rbp7 in Pol II (Figure 1.9). From a structural point of view, positive amino acid residues of Rpb1 and Rpb2 subunits form the cleft and active site of the polymerase with the other subunits arranged at the periphery of the complex. Rpb1 represents the mobile component of the Pol II clamp, with its motion being highly regulated by the neighbouring switch subunits. As confirmed by X-ray and cryo-EM studies, the Rbp4 and Rbp7 complex forms a dynamic link between the clamp portion and the C-terminal repeat domain (CTD), whose phosphorylation state regulates the recruitment of essential nuclear proteins associated with mRNA biogenesis.

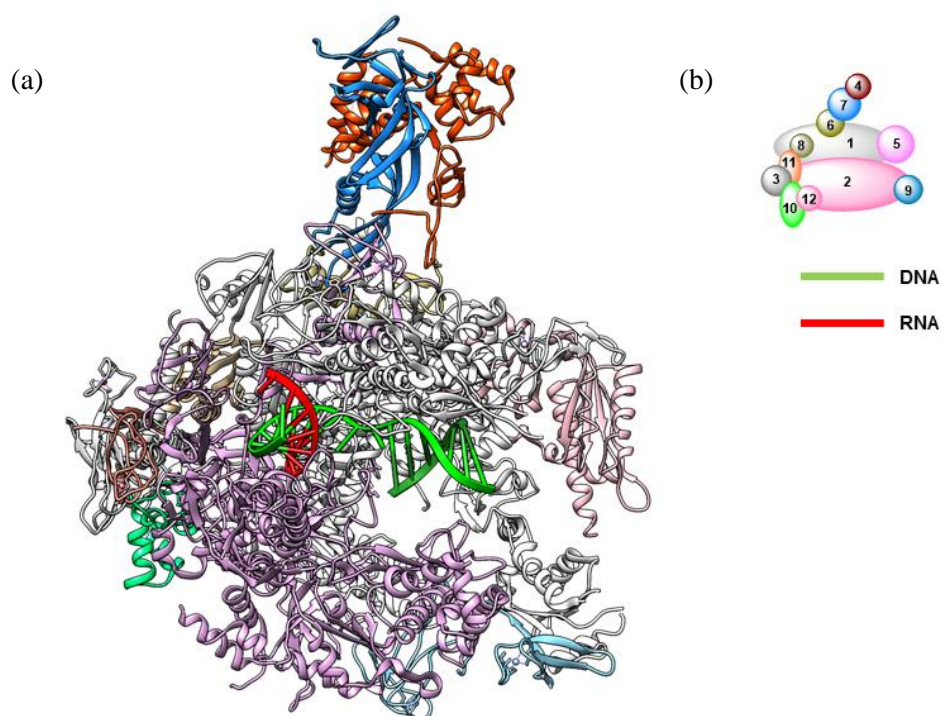


Figure 1.9 (a) Crystal structure of the RNA pol II transcriptional elongation complex (PDB 1I6H).⁴⁷ (b) Schematic representation of RNA Pol II Rbp subunits.

1.6 Regulation of transcriptional elongation

The first step of transcription is the formation of the pre-initiation complex (PIC), which is composed of several transcription factors interacting within close proximity of the transcription start site (TSS).⁴⁸ Other protein factors such as co-activators and co-repressors are involved generally in binding to distal portions of DNA relative to the TSS. The importance of each component and the chronological order of the PIC assembly have been established with a combination of X-ray, NMR and cross-linking experiments.⁴⁶

The first stage of PIC formation is the binding of the TBP portions of Transcription Factor IID (TFIID) to TATA elements located between 31 and 23 base pairs upstream from the TSS.⁴⁸

The concave β -sheet portions of the TBP subunit of TFIID is quasi-symmetrical and interacts with an 8 A/T base pair region in the minor groove (Figure 1.10). The phenylalanine residues of TBP intercalate into the 5' end of the TATA region, inducing a perturbation of the standard B-DNA double-helix and slightly unwinding the DNA. The resulting DNA bending and widening of the minor groove allows interactions between the β -sheet portions of TBP and the central portion of the TATA box.

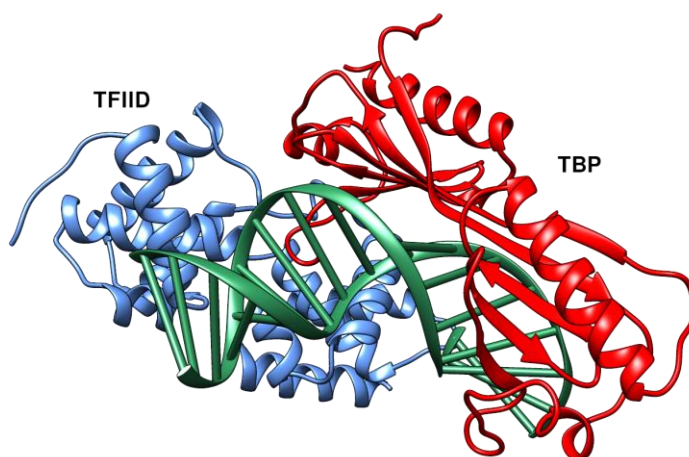


Figure 1.10 Crystal structure of TBP-TFIID in complex with dsDNA (PDB 1AIS).⁴⁹

The next step of PIC assembly is the formation of a quaternary protein-DNA complex composed of TFIID-DNA-TFIIB-TFIIA. TFIIA and TFIIB bind on the downstream side of the complex and presumably act as a bridge between TBP, Pol II and other transcription factors, providing the directionality of transcription (Figure 1.11a). Further transcription factors, such as TFIIE, TFIIIF and TFIIH, stabilise the PIC. TFIIIF forms a stable complex with Pol II prior to binding to the TFIID-DNA-TFIIB-TFIIA quaternary complex, while TFIIE and TFIIH complete the formation of the PIC and respectively function as the ATPase and helicase, required for melting of the dsDNA template. TFIIH also promotes phosphorylation of Pol II, which regulates the switch to the elongation phase.⁴⁶

Following formation of the PIC, RNA Pol II is released from the PIC and traverses the unwound DNA along the single stranded template, synthesising complementary mRNA. This nascent strand is then elongated with the sequential addition of complementary nucleoside triphosphates (NTPs). The reaction progresses as NTPs enter the polymerase active site and elongate the growing RNA strand on the 3' end, releasing pyrophosphate (Figure 1.11b). The mechanism of specific recognition and nucleotide incorporation is mediated by two metal ions present in the active sites. One of the metal ions (A) is fixed on the growing RNA strand, while the second (B) is mobile and free to recognise the incoming NTP. It is hypothesised that base pairing of the complementary NTP with the DNA template would induce a conformational change to trigger the loop to close the active site. The RNA polymerase then proceeds along the DNA strand repeating this incorporation process until it encounters an elongation stop site.^{45,50,51}

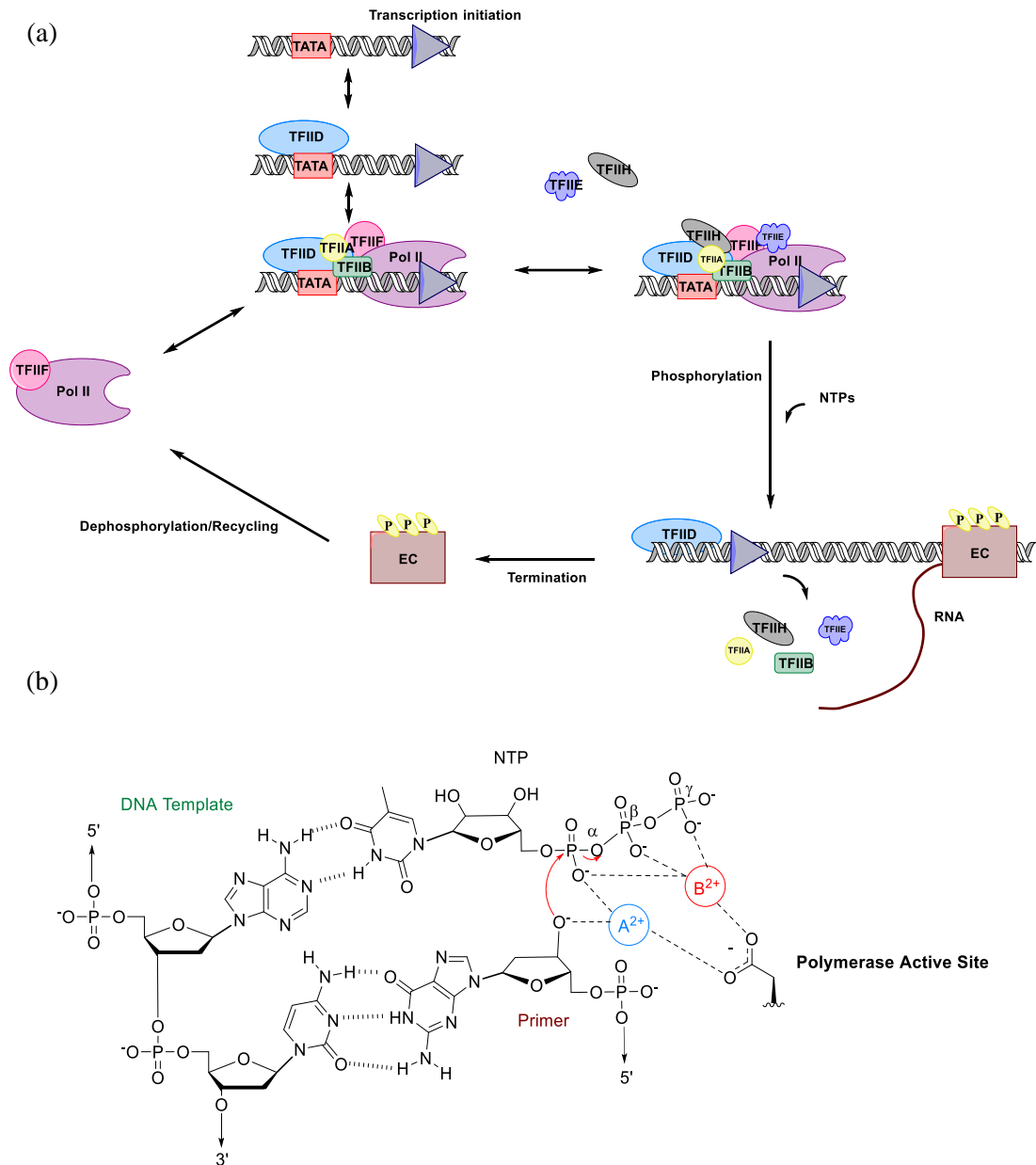


Figure 1.11 (a) Schematic representation of eukaryotic transcription.⁴⁶ (b) Schematic representation of the two cation-mediated mechanism of RNA elongation.⁵²

1.6.1 Structure and function of nuclear receptors

Nuclear receptors (NRs) are a class of transcription factors essential for responding to physiological stimuli.^{48,53–55} There are 48 different NRs, divided into seven primary sub-types according to structure, ligands, cellular localisation and mode of binding to dsDNA.⁵⁶ The

entirety of the hormone receptors are localised in the cytoplasm in the inactive state and recognise DNA responsive sequences as dimers (either homo or hetero).

NRs are classified into sub-families according to the degree of homology between distinct sub-regions (Figure 1.12). The N-terminal domain is generally composed of regions A/B, which are normally responsible for hormone-independent functions and its composition varies between NR sub-families. The DNA binding domain (DBD) is found in the highly conserved region C of the N-terminal domain, which contains cysteines and basic amino acids in fixed positions that coordinate Zn^{2+} (see 1.3.1).⁵⁵ Region D is a variable hinge, which plays a role in ligand recognition in certain NRs. Sub-region E in the C-terminal portion forms the ligand binding domain (LBD), while F is found only in certain NRs and its role is not completely understood.^{54,55}

The function of NRs is regulated by the binding of selective ligands, such as hormones, which induce a conformational change to activate dsDNA binding.⁵⁴ LBDs are generally composed of 11 fixed α -helices that form a globular hydrophobic core. A twelfth mobile helix functions as a lid, which is kept in a suitable conformation for co-activator recognition by the binding of agonist ligands (Figure 1.13). A stable hydrophobic groove within the LBD is recognised by co-activators through a highly conserved motif, LXXLL (where L= leucine and X = any amino acid). This motif, known as the NR box, is essential for mediating gene expression control. The NR box perfectly fits the LBD groove through a combination of hydrophobic interactions between leucine and the hydrophobic core of the LBD, and via hydrogen bonding between glutamic acid and lysine residues of helix 12 (H12) and helix 3 (H3) that clamp the flanking region of the LXXLL motif in the optimal binding orientation.⁵⁶ NR conformational changes, specific to each NR sub-type, mediate intricate paths of receptor activation for nuclear permeation, DNA recognition and transcription of a selected pool of genes (Figure 1.12b).⁵⁵ Due to the importance of NRs in physiological regulation, several small molecules have been designed to pharmacologically regulate their activity. While agonist ligands mimic

the action of the natural hormones, antagonists recognise the LBD inducing different conformational changes, which prevent co-activator recognition (Figure 1.13).

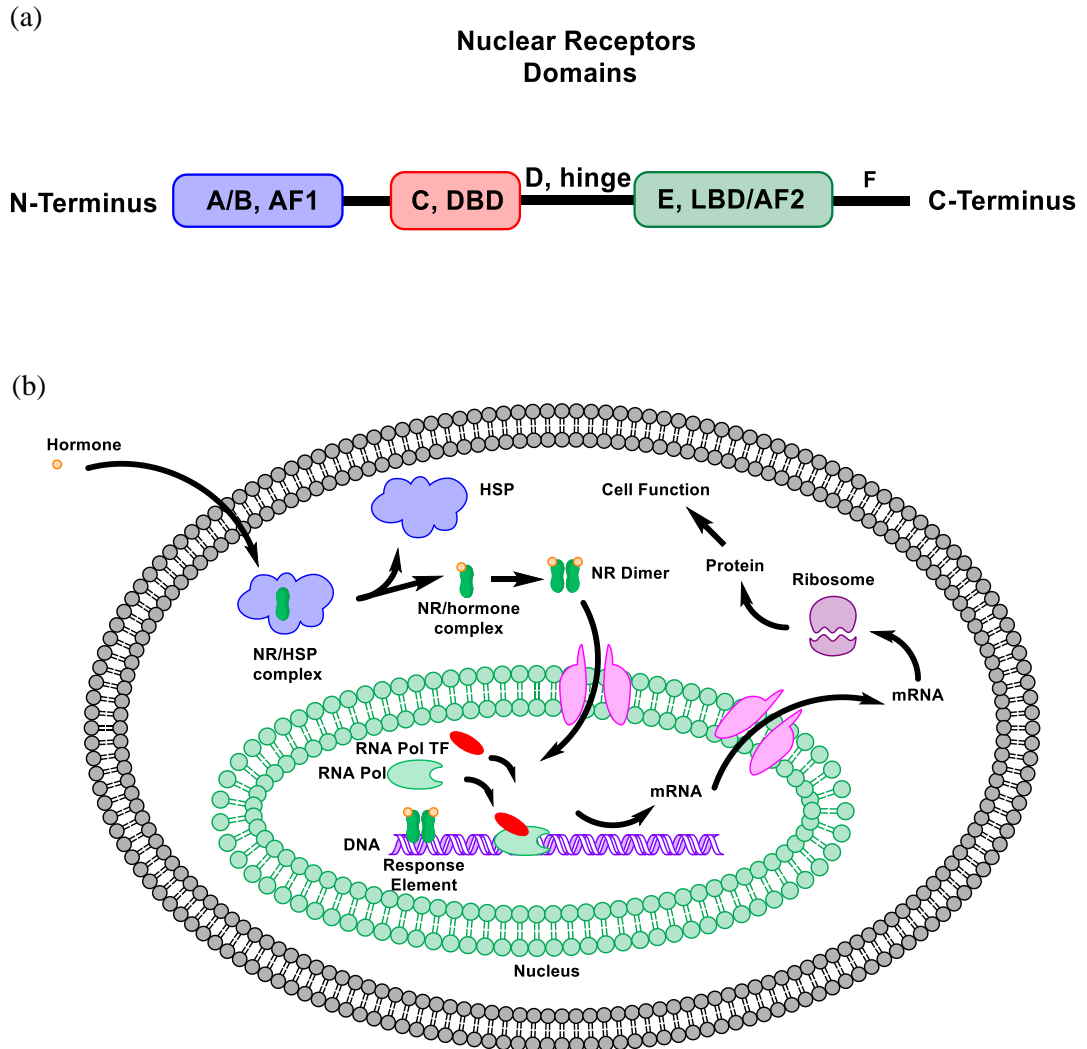


Figure 1.12 (a) Schematic representation of NR domains. (b) Signal cascade of hormone NRs (HSP = heat shock protein).⁵⁵

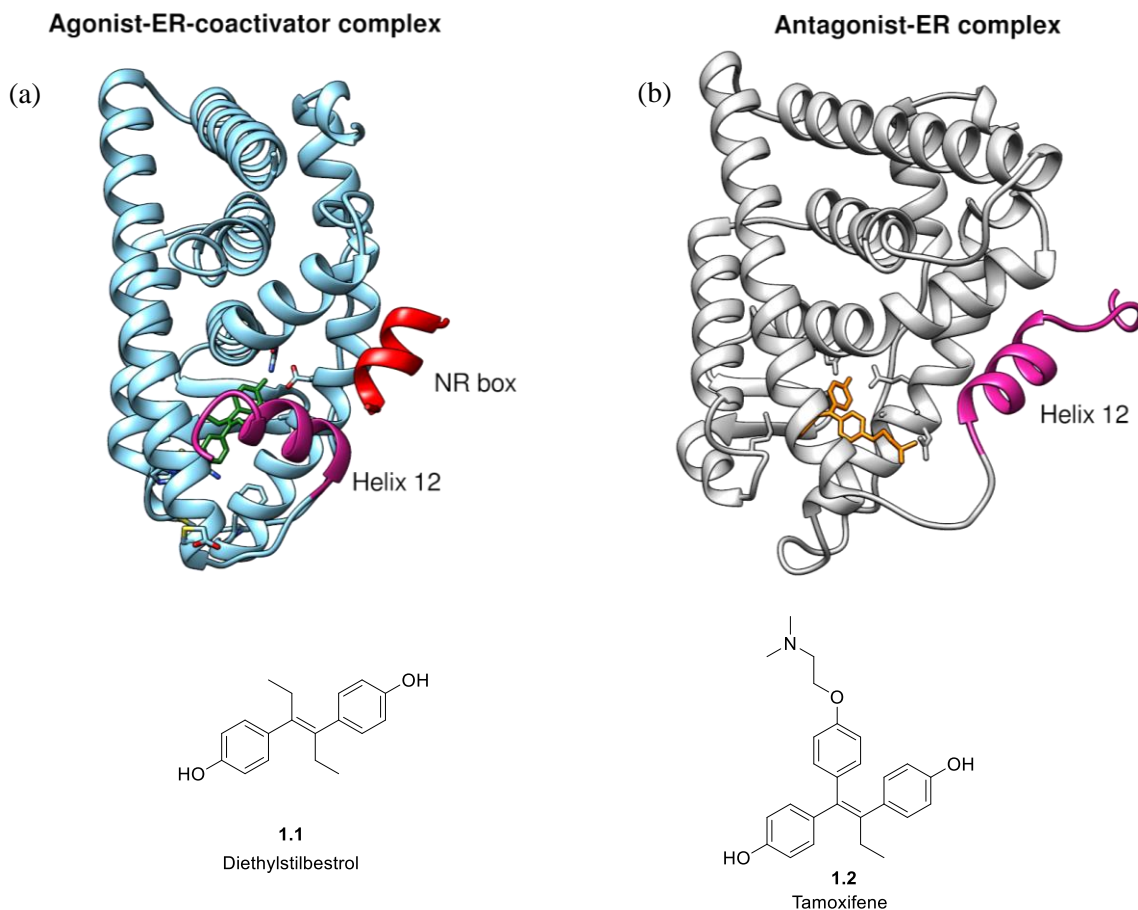


Figure 1.13 Crystal structure of the estrogen receptor (ER) bound by agonist and antagonist small molecules. (a) The activator diethylstilbestrol (green) orients H12 (magenta) in the “open” position. Allowing the LBD to bind the NR box (red) (PDB 3ERD).⁵⁷ (b) The antagonist 4-hydroxy-tamoxifene (orange) orients H12 in the “closed” conformation, preventing interaction with NR boxes (PDB 3ERT).⁵⁷

1.7 Exogenous regulation of gene expression

Compounds that selectively bind to specific DNA sequences can silence the expression of genes responsible for pathologies ranging from microbial infection to cancer.^{9,16,58–62} This thesis will focus on strategies involving sequence-selective DNA recognition to downregulate the transcription of a subset of eukaryotic genes.

1.7.1 Triplex-forming oligonucleotide and peptide nucleic acids

The idea of using oligomeric nucleic acids to control gene expression is based on complementary Watson-Crick and/or Hoogsteen base pairing.⁶³ At the transcriptional level, recognition of double-helix B-DNA by TFOs takes place in the major groove exploiting the hydrogen bond acceptor/donor pattern of the Hoogsteen base pair face, to form a stable triplex helix that blocks transcription elongation or transcription factor recognition (Figure 1.14).⁵⁸ Generally, TFOs are homopurine or homopyrimidine strands composed of 20-40 nucleotides.⁵⁸ Although TFOs exhibit a high degree of sequence selectivity, a biologically significant limitation of this approach is the restriction to targeting only polypurine regions at physiological conditions via antiparallel-oriented binding. In this binding mode, TFOs form antiparallel Hoogsteen base pairs with purines of the DNA target sequence (in particular G recognises G:C Watson-Crick base pairs and A recognises A:T) (Figure 1.15).⁵⁸ The parallel-oriented binding for polypyrimidine recognition is limited by the necessity of a positive charge on the N3 of C, which is not protonated at physiological pH (Figure 1.15b).

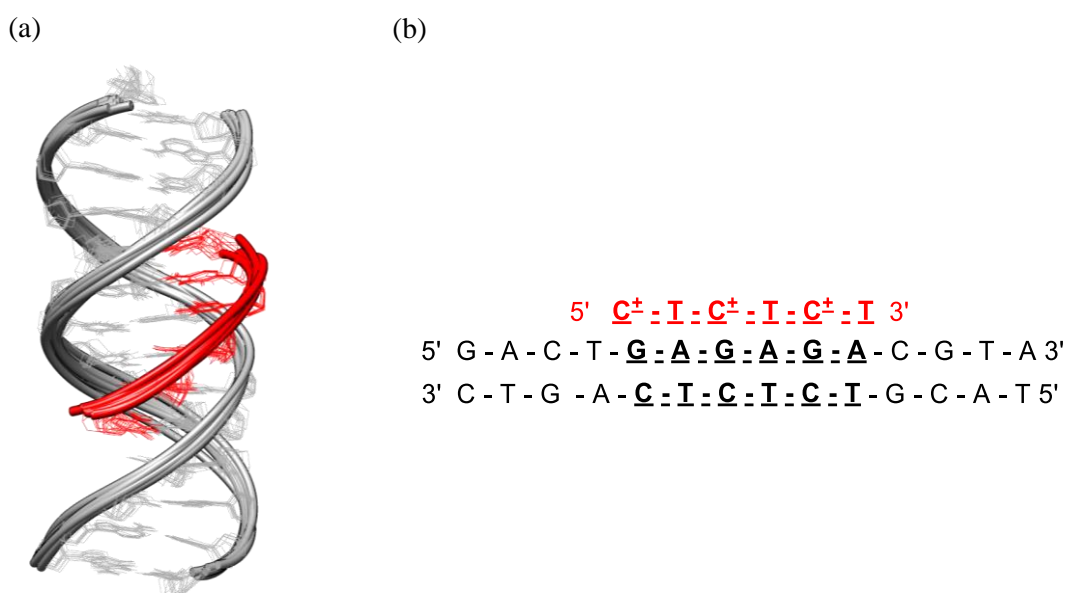
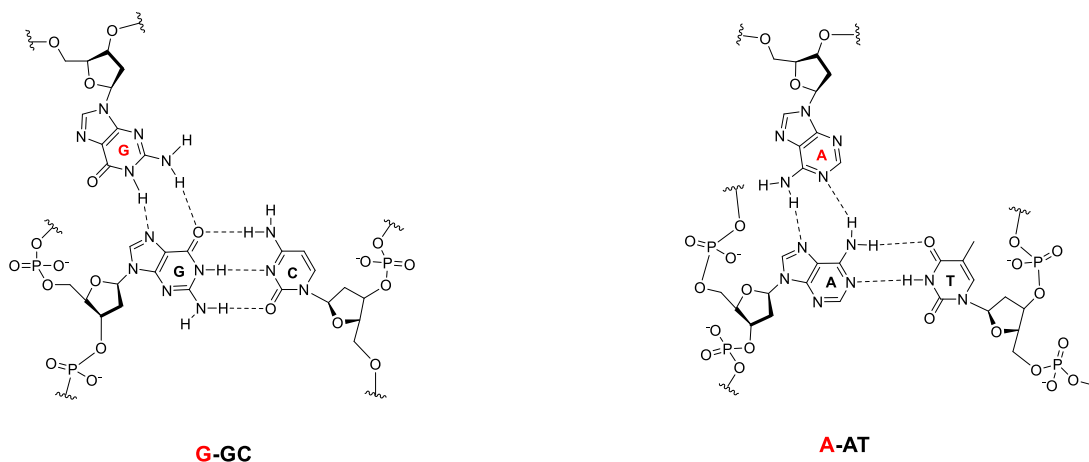


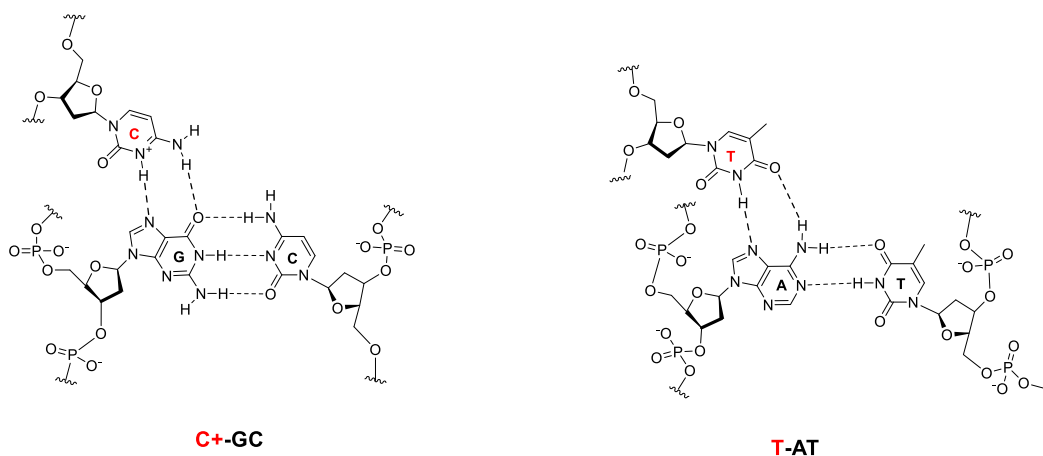
Figure 1.14 (a) NMR-derived structures and (b) sequence of TFO:DNA triplex at pH 5 (PDB 1BWG).⁶⁴

Antiparallel Purine-directed Hoogsteen Base Pairing

(a)



(b)

Parallel Pyrimidine-directed Hoogsteen Base Pairing**Figure 1.15** Schematic representation of (a) purine and (b) pyrimidine Hoogsteen base pairing.

Other limiting factors to their biological application include poor DNA binding affinity due to electrostatic repulsion between the phosphate negative charges, poor cellular uptake and susceptibility to enzymatic digestion.⁶⁵ General strategies adopted to overcome these limitations focus on varying the structural features of both the sugar and phosphate portions

of the backbone (Figure 1.16). Phosphate modifications include phosphorothioate, in which one oxygen is replaced with sulphur; and phosphoramidate, in which a bridging oxygen is replaced by a nitrogen. Common sugar modifications consist of locked nucleic acids (LNA), in which an oxymethylene linker connects C2' and C5', or modification at the 2' position with O-methoxy-ethyl derivatives (MOE).

More drastic modifications, deviating away from the natural nucleic structure, include phosphoramidate morpholino oligonucleotides (PMOs) and peptide nucleic acids (PNAs).^{59,66,67} In PNAs, the sugar-phosphate backbone is replaced by aminoethyl glycine units linked via amide bonds. A carboxymethylene group connects this backbone to the natural nucleobases, which drive the sequence-selective DNA recognition (Figure 1.16). PNAs are resistant to peptidase and nuclease degradation, and the absence of negative charges in the backbone enhances the affinity for DNA. Although several additional structural modifications have been explored at length, the use of PNAs is limited by poor cellular and nuclear uptake and conjugation to cell penetrating peptides and cationic lipids is generally required to overcome this problem.^{68–70}

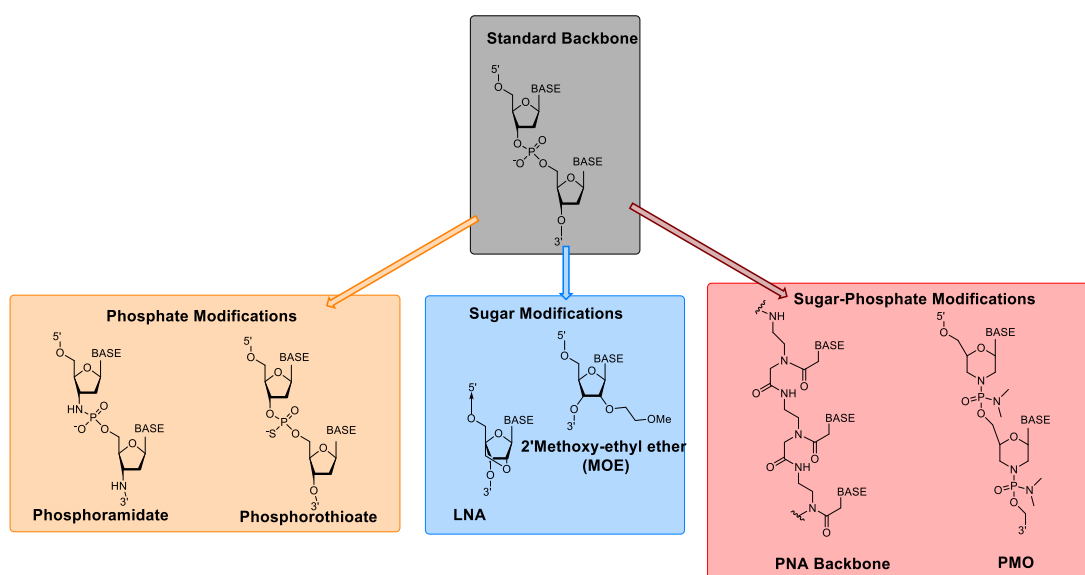


Figure 1.16 Schematic representation of sugar-phosphate backbone modifications.

1.7.2 Minor groove binders as regulators of gene expression

In addition to major groove dsDNA recognition by large proteins, Nature has also developed small ligands that interact with the minor groove of dsDNA. The development of therapeutically relevant compounds began with the isolation of distamycin A (**1.3**) and netropsin (**1.4**) from *Streptomyces distallicus* and *netropsis*, respectively (Figure 1.17).^{6,7} Distamycin A consists of an N-formylated methyl tripyrrole amide chain that is terminated with an ethylenamidino moiety (tail), while netropsin has only two pyrrole rings with the formyl group replaced by a more complex N-carbamimidoylglycyl amino moiety. Their recognition of the DNA minor groove is enthalpically-driven by non-covalent interactions such as hydrogen bonding, van der Waals interactions, electrostatic contacts between their positive charges and the phosphate backbone; as well as by shape recognition, with the concave curvature of the compounds matching the helicity of the minor groove (Figure 1.18).⁷¹ The displacement of the ordered hydration spine of the minor groove provides an additional entropic contribution to dsDNA binding, which selectively occurs at A/T-rich regions of DNA.^{5,72,73}

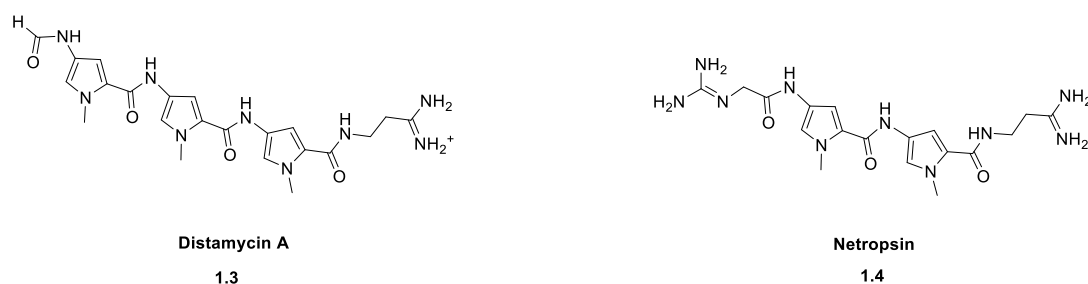


Figure 1.17 Structures of distamycin A and netropsin.

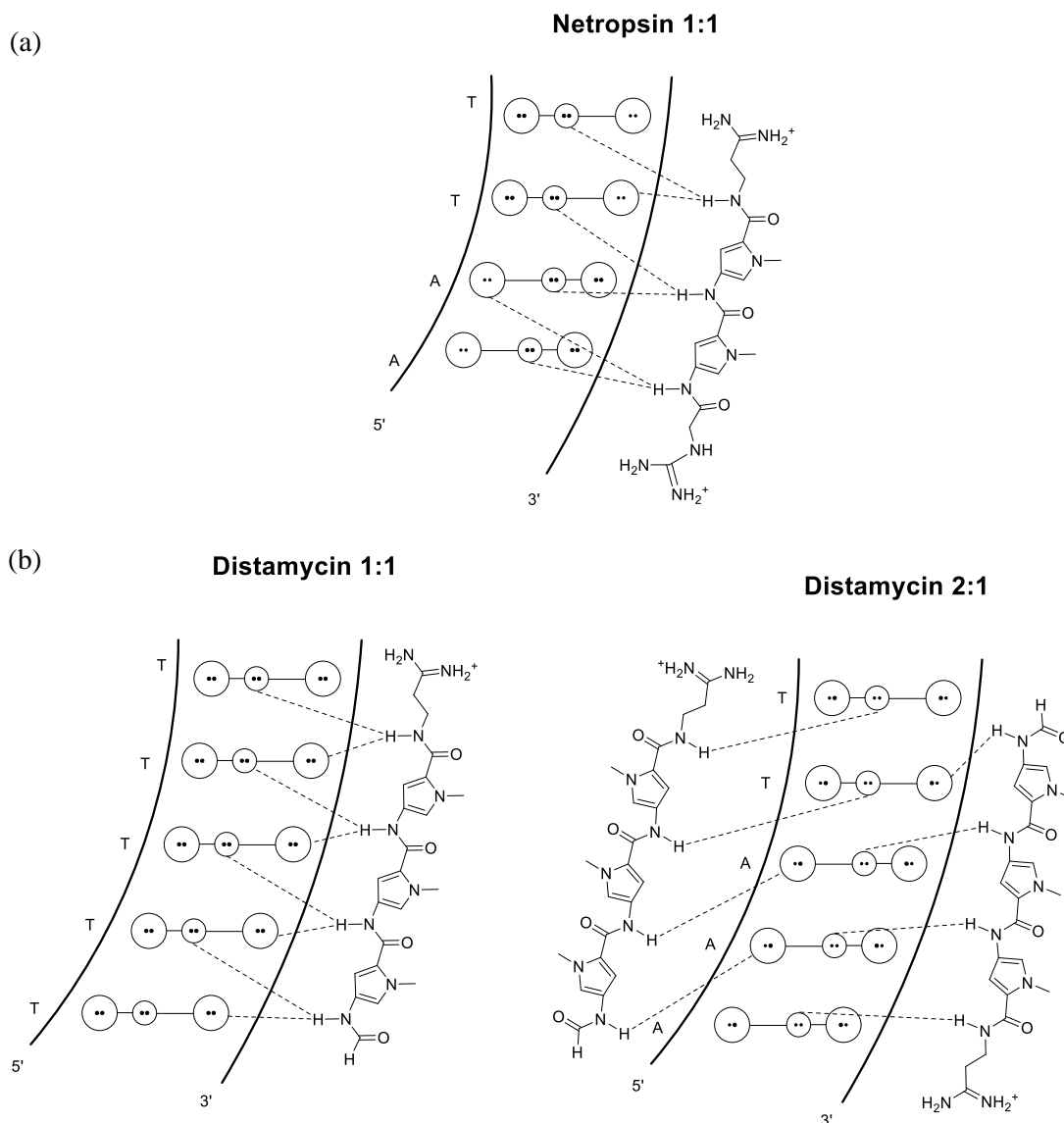


Figure 1.18 Schematic representation of the binding modes of (a) netropsin and (b) distamycin with dsDNA.

The structural basis of this selectivity profile arises from the topological and molecular differences between G/C and A/T regions in the minor groove. The presence of the exocyclic amine of G(N2) at the floor of the groove results in an unfavourable steric clash with the Py (C3)(H3) bonds of the ligands, and widens the groove, reducing both the negative electrostatic potential well and van der Waals contacts between the wall of the groove and the aromatic rings of the ligands. Distamycin forms a stable complex with A/T regions of dsDNA either in

1:1 or 2:1 stoichiometry, depending on the DNA sequence (Figure 1.18-19).^{9,74-77} In both cases, the amide protons of distamycin form hydrogen bonds with T(O2) and A(N3).

X-ray and NMR studies of **1.3** indicate that a larger perturbation of the minor groove is induced by the accommodation of a second distamycin molecule, which forms a non-covalent dimeric aggregate with head-to-tail side-by-side orientation.⁹ The repulsive forces between the positive charges of the tail induce a staggered conformation, in which the aromatic rings of one ligand overlap with the amide bonds of the opposite molecule (Figure 1.19). Although there is no clear explanation for the preference of these two binding modes, a variety of biophysical approaches has suggested higher binding affinity for 1:1 binding compared to 2:1 and identified 5'ATAAT as the favoured sequence.^{76,77} Taken collectively, these data suggest that the local deformation of the target dsDNA sequence strongly influences the distamycin binding mode. In contrast with the multiple binding modes of distamycin, only 1:1 binding modes are observed for the doubly positive charged netropsin.^{73,78} Despite the limitation caused by the degeneracy of the hydrogen bond donor/acceptor pattern of A/T regions, which complicates A over T discrimination, small molecule minor groove binders (MGBs) represent an attractive system to target DNA sequences for therapeutic applications. In particular, their relatively small size presents an appealing starting point for the development of cell permeable DNA ligands, which target relatively large DNA sequences within the genome.³

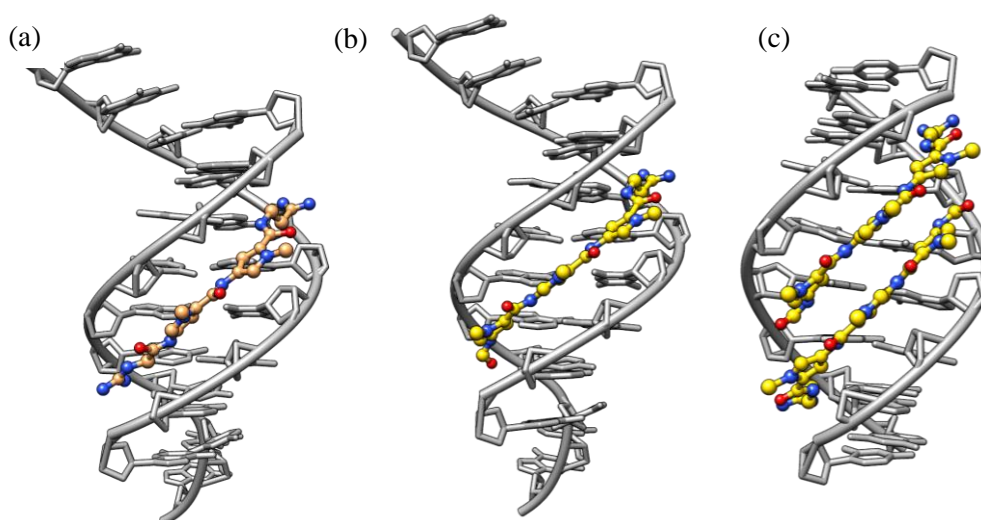


Figure 1.19 Crystal structures of (a) netropsin (salmon)(PDB 1Z8V)⁷⁹ and distamycin A (yellow) in (b) 1:1 (PDB 1JTL)⁸⁰ and (c) 2:1 (PDB 378D)⁸¹ binding modes within the B-DNA minor groove.

1.7.2.1 Lexitropsins

Lexitropsins are a class of synthetic MGBs, based on the natural products distamycin and netropsin, with the ability to specifically read different sequences of DNA.⁸² Initial development of the lexitropsin design was mainly focused on expanding the sequence discrimination and the length of recognition. Replacement of pyrrole (Py) building blocks with alternative five-membered aromatic rings, taking into account the topological differences of the minor groove, represented the main strategy to achieve sequence discrimination.⁸³ The first step towards a complete set of binding rules was pioneered by the Lown group through the incorporation of imidazole (Im) units into the netropsin scaffold as a means to discriminate G/C from A/T regions (Figure 1.20).^{71,83–85} Although early thermodynamic studies revealed that G/C regions of DNA were tolerated, enhanced selective binding was only observed when replacing the guanidinium group of the head with a formyl moiety, as for distamycin (Figure 1.20).⁸⁵ The reduced cationic character, which non-specifically directed the binding towards the more electron-rich A/T regions, along with the hydrogen bonding between Im(N3) and G(N2), were able to drive the discrimination of G/C base pairs from A/T DNA tracts.^{84,86} Further refinement of G/C recognition was achieved by introducing Im units into analogues

capable of binding to dsDNA in a 2:1 stoichiometry (Figure 1.21). In fact, by exploiting the head-to-tail side-by-side antiparallel orientation assumed by the ligands inside the minor groove and the hydrogen bond between Im(N3) and G(N2), it was possible to design small ligands capable of discriminating GC over CG, AT and TA.⁸³

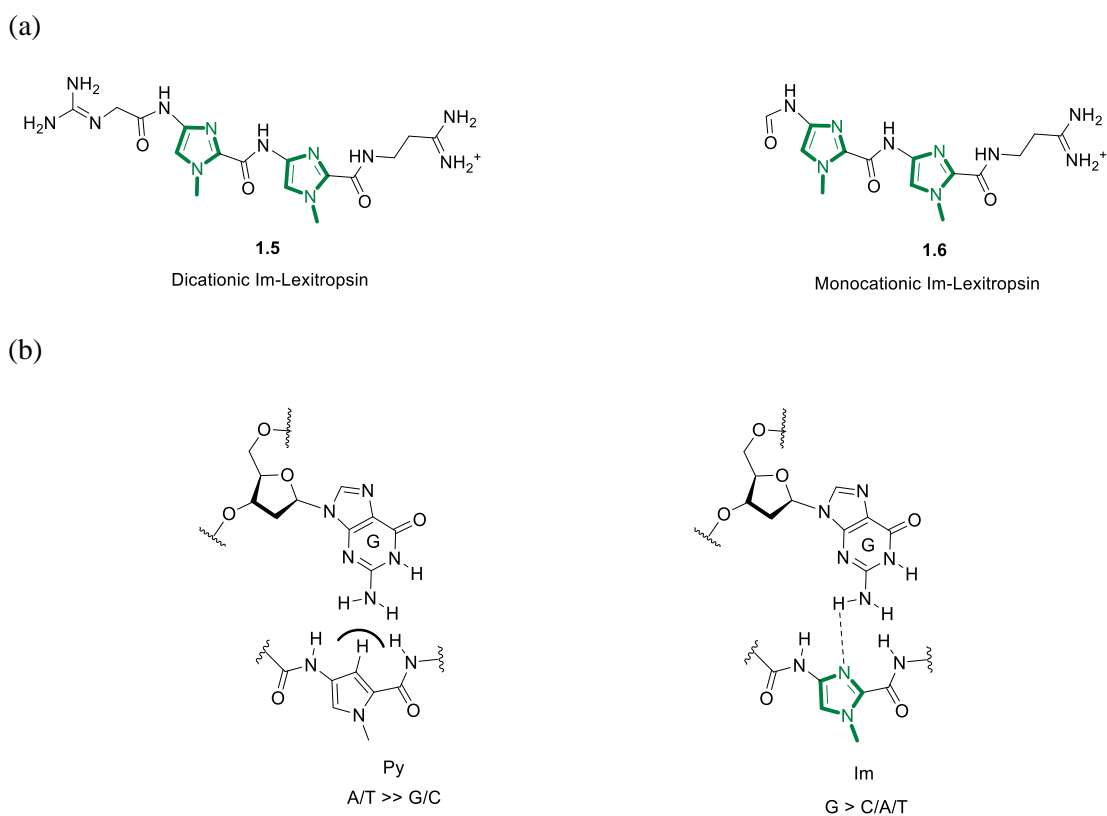


Figure 1.20 (a) Structures of the first examples of Im-containing lexitropsins (**1.5-6**). (b) Schematic representation of interaction of Py and Im with G.

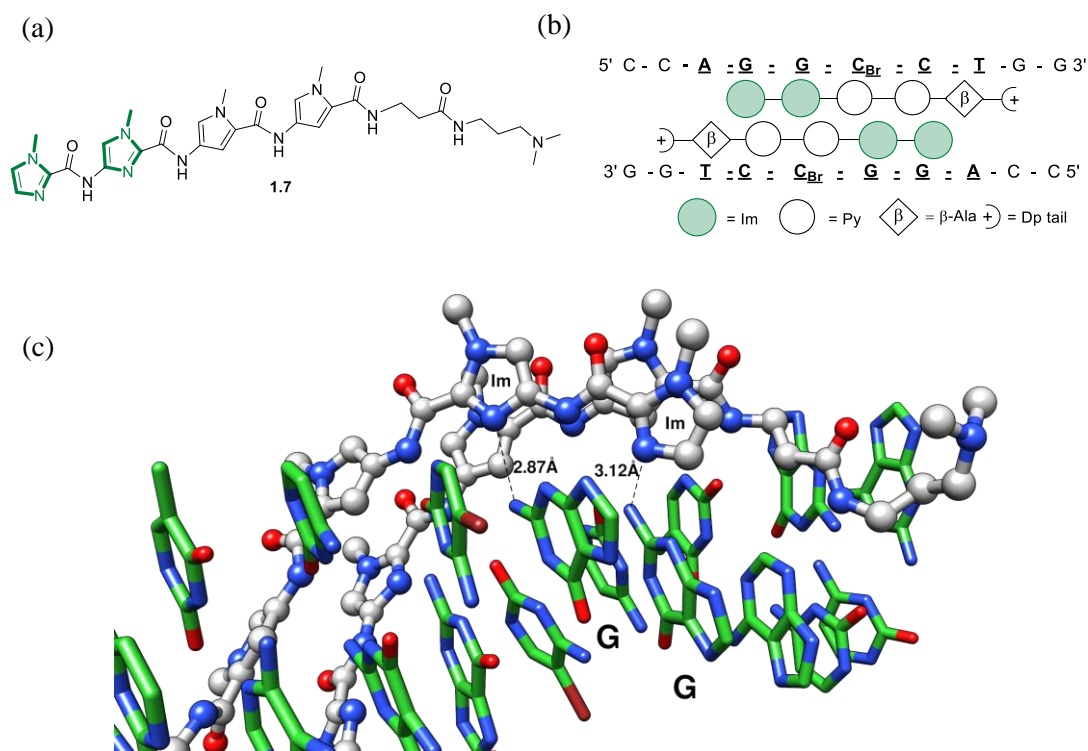


Figure 1.21 (a) Structure of an Im-containing MGB (**1.7**). (b) Schematic representation of **1.7** bound to its dsDNA target sequence (Dp = dimethylaminopropylamine C_{Br} = 5-bromocytosine). (c) Crystal structure highlighting the hydrogen bonding between Im and G(N2) in **1.7**:dsDNA 2:1 complex (PDB 365D).⁸⁶

The lack of a clear asymmetric topology between A and T minor groove edges, both of which possess the ability to accept hydrogen bonds through A(N3) and T(O2), resulted in a more challenging design for MGBs able to discriminate AT over TA base pairs. However, a subtle asymmetric cleft generated by the A(H2) in the minor groove was exploited by Dervan and co-workers, leading to the identification of the hydroxypyrrole (Hp) building block, which selectively recognised T over A and completed the set of binding rules (Figure 1.22).⁸⁷⁻⁸⁹ As confirmed by crystallographic studies of a 4-ring MGB containing Hp binding to dsDNA in a 2:1 stoichiometry, the steric clash between A(H2) and Hp(OH3) along with the hydrogen bonding between Hp(OH3) and T(O2) orients Hp over T. This binding behaviour was further confirmed in other MGBs scaffolds. However, the lower binding affinity, which resulted from incorporation of one or more Hp units, limited the widespread use of this building block.

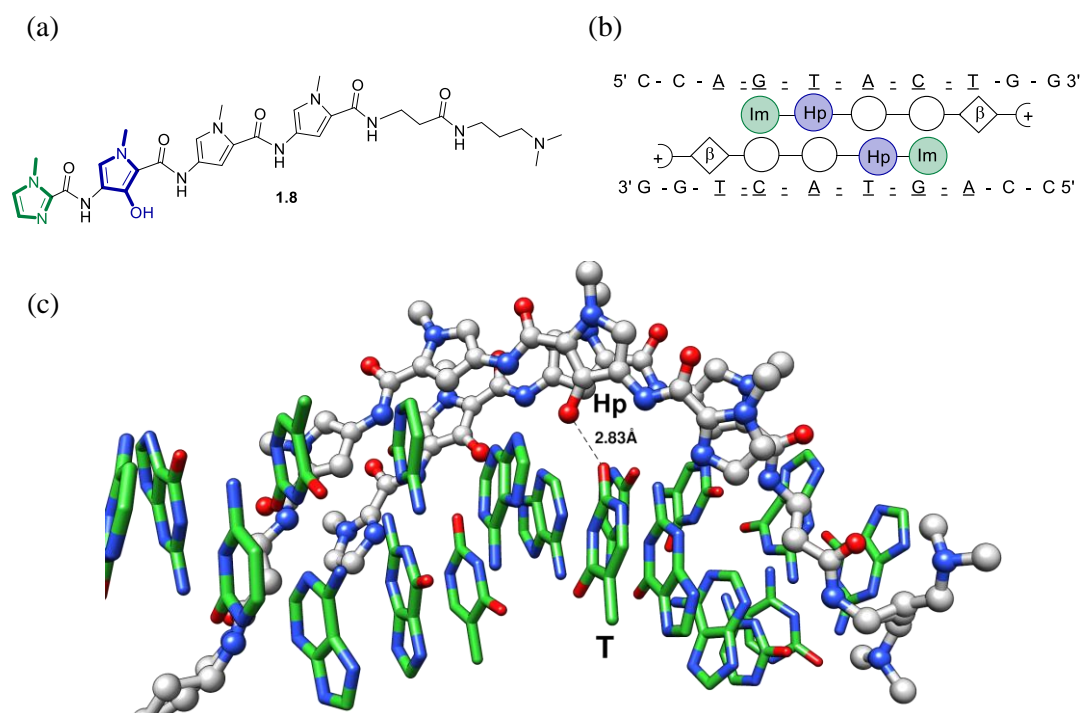


Figure 1.22 (a) Structure of a Hp-containing MGB (**1.8**). (b) Schematic representation of **1.8** bound to its dsDNA target sequence. (c) Crystal structure highlighting the hydrogen bonding between Hp and T(O2) in a **1.8**:dsDNA 2:1 complex (PDB 407D).⁹⁰

In summary, the binding rules for sequence recognition in a side-by-side orientation are: Im/Py pairs with GC over CG, AT and TA; Hp/Py pairs with TA over AT, GC and CG; and Py/Py preferentially pairs with TA or AT over CG and GC (Table 1.1).

Table 1.1 Binding rules of MGBs binding in an antiparallel side-by-side conformation.

Pair	G·C	C·G	A·T	T·A
Py-Py	-	-	+	+
Py-Im	-	+	-	-
Im-Py	+	-	-	-
Hp-Py	-	-	-	+
Py-Hp	-	-	+	-

In parallel with the design of the binding rules, several works focused on expanding the length of the target dsDNA sequence. A combination of biophysical and structural studies found that

a maximum of five consecutive aromatic rings could retain high binding affinity as the more pronounced curvature of longer MGBs loses the track of the minor groove register.^{2,86} One solution to this limitation came from the introduction of flexible β -alanine units, which reset the curvature of the MGBs and allow for the targeting of DNA base sequences of up to 16 contiguous base pairs.⁹¹⁻⁹³ Nevertheless, long compounds containing a high β -alanine content bind preferentially via a 1:1 recognition motif.^{94,95}

Although a clear explanation for this behaviour has not yet been given, a critical role seems to be played by the base pair composition of the target sequence. In particular, it was suggested that the strength of the base stacking in DNA strands containing contiguous purine residues limits the flexibility of the double helix, which less favourably accommodates two molecules of ligand.⁹⁵ The main consequence of this binding behaviour is the limited selectivity, which results from the loss of the antiparallel side-by-side orientation. Here, the selective recognition of G was complicated by the equal possibility of Im(N3) of hydrogen bonding with G(N2) of both GC and CG base pair, which is roughly located in the middle of the groove. NMR studies of a linear MGB (**1.9**) binding to d(CCAAAGAGAAGCG)·d(CGCTTCTCTTTGG) suggested that the binding orientation *N* to *C* aligned with 3' to 5' of the purine rich strand is concomitantly driven by the formation of a bifurcated hydrogen bond between the amide protons of the ligand and A(N3) and T(O2) along with the hydrogen bonding between Im(N3) and G(N2) (Figure 1.23).⁹⁵ Despite the lower selectivity, MGBs binding in 1:1 motif have found widespread use in targeting -GAA- repeats, which are involved in genetic diseases, such as Friedreich's ataxia.^{92,93}

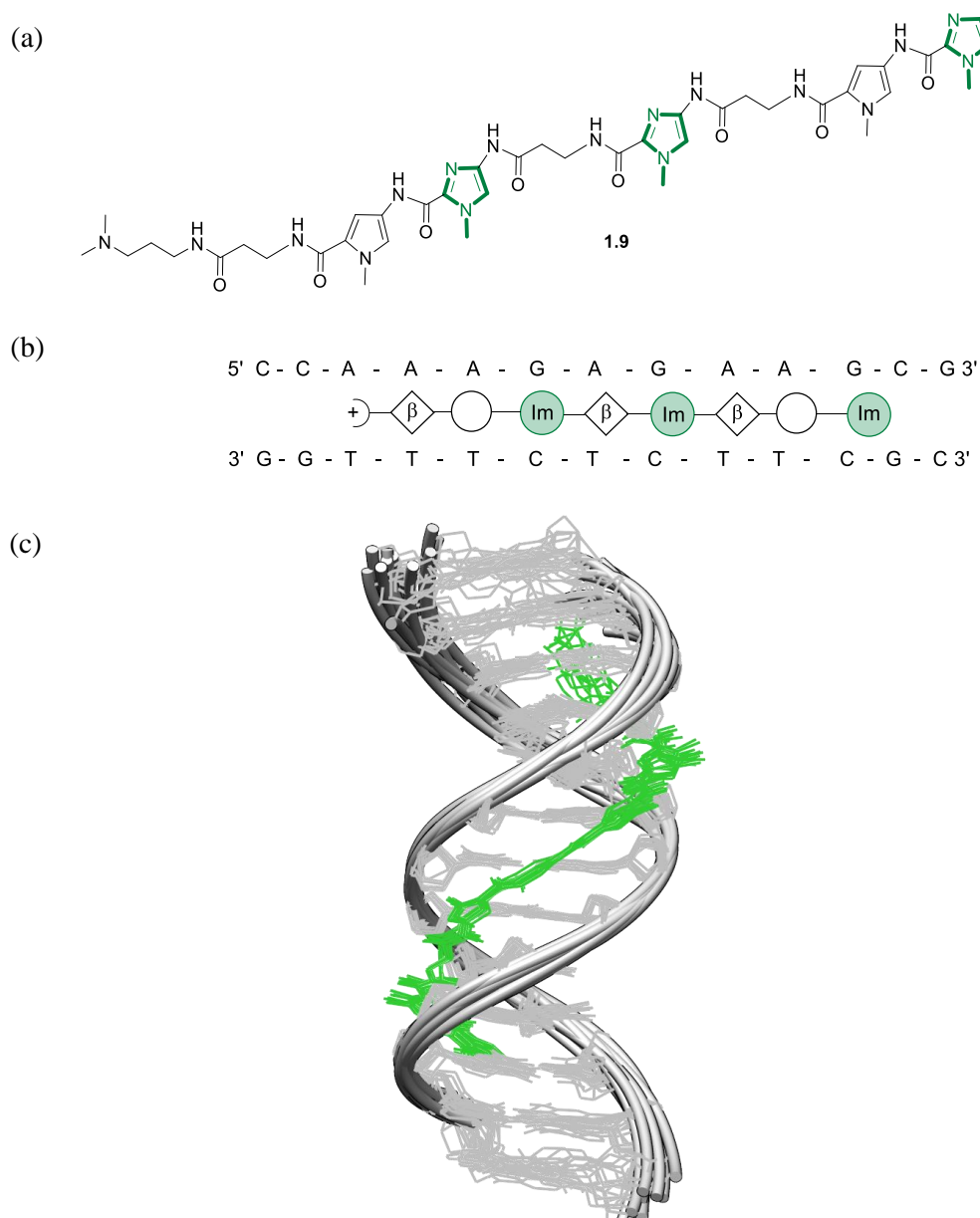


Figure 1.23 (a) Structure of a linear polyamide (**1.9**) targeting 9 base pair of dsDNA. (b) Schematic representation of the interaction between **1.9** and its dsDNA target sequence. (c) NMR-derived structures of **1.9** in complex with the target sequence (PDB 1LEJ).⁹⁵

1.7.2.2 DNA-binding polyamides

Although MGBs containing Im units binding in a 2:1 stoichiometry displayed enhanced sequence selectivity, DNA recognition was limited to palindromic sequences (Figure 1.24a).³

The subsequent evolution in the design of selective MGBs consisted of tethered analogues, in

which asymmetric strands of aromatic rings were linked through flexible chains to maintain the side-by-side disposition, necessary for the selectivity.³ The Dervan and Lown groups pioneered the H-pin motif (Figure 1.24b), which is characterised by the covalent linkage of two Py-Im polyamides chains through a four-carbon tether between the two nitrogen atoms of a Py/Py pair, which resulted in higher binding affinity and specificity compared to the related unlinked analogue.^{71,96–100} Increased sequence specificity was achieved through the development of a hairpin motif (Figure 1.24c), in which six or eight aromatic rings are spaced out by a flexible linker, binding in a hairpin conformation within the groove.¹⁰¹ Several studies identified γ -aminobutyric acid as the optimal linker to allow the folding of the aromatic rings into a hairpin conformation.⁹⁴

Due to the enhanced binding affinity (25 to 150-fold), specificity (1 to 20-fold) and synthetic tractability compared to H-pin polyamides, hairpin polyamides have been extensively explored since first reported in 1994.¹⁰² DNase I footprinting and NMR experiments indicated that hairpin polyamides predominately bind to dsDNA in a *N* to *C*, 5' to 3' orientation (Figure 1.25).⁹⁴ In contrast to the “slipped” side-by-side orientation observed for 2:1 MGBs, NMR structural studies using a six-ring polyamide suggested that, in the hairpin conformation, the aromatic rings of the opposite sides are stacked upon each other.¹² This increased side-by-side arrangement was considered responsible for the enhanced sequence discrimination.¹² This observation inspired the design of cyclic polyamides, in which the covalent link between *C* and *N*-termini reduces the possibility of a “slipped” disposition of the aromatic rings (Figure 1.24d).^{103–105} Although early NMR and thermodynamic studies on a six-ring cyclic polyamide confirmed the expected side-by-side binding conformation and revealed enhanced binding affinity compared to the relative hairpin polyamide, the loss of specificity towards a single mismatch target sequence as well as challenging synthetic routes to access cyclic analogues discouraged further developments until the early 2000s.¹⁰⁶

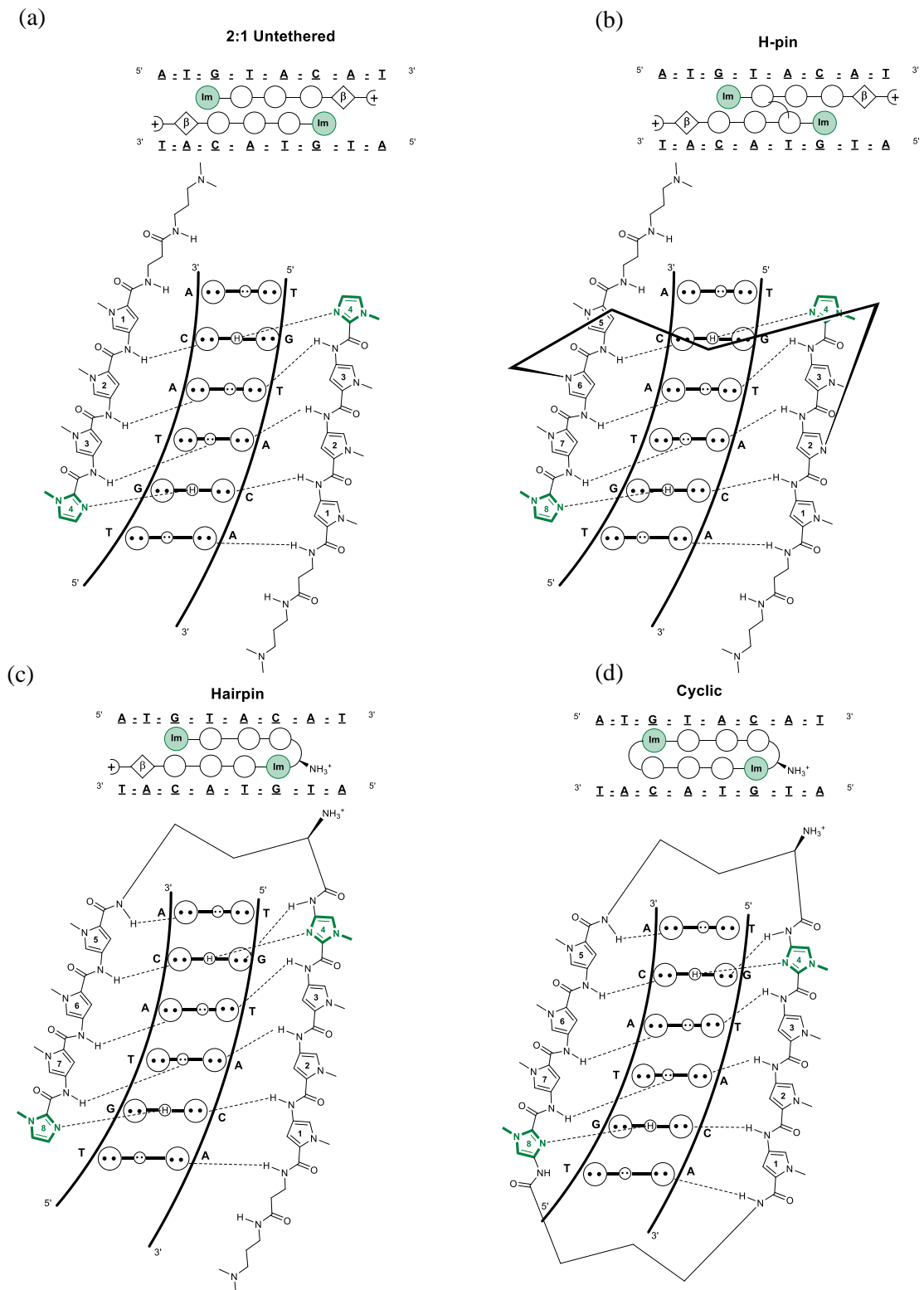


Figure 1.24 Schematic representation of (a) untethered, (b) H-pin, (c) hairpin and (d) cyclic polyamides interacting with dsDNA.

A significant improvement in dsDNA binding affinity and sequence selectivity emerged from the introduction of a chiral amine on the flexible linker for both hairpin and cyclic polyamides (Figure 1.26).^{94,107} The presence of the extra cationic charge increased the strength of the electrostatic interactions with the negatively charged phosphate backbone of the DNA and improved the general water solubility. Both the position and stereochemistry of the amine moiety also expanded the possibility of controlling the general binding orientation of the polyamides. Melting temperature studies suggested that the installation of the amine moiety in the α -position of the turn with an *R* configuration favoured the *N* to *C* 5' to 3' binding orientation.¹⁰⁸ Furthermore, molecular modelling of the hairpin polyamide in complex with a dsDNA sequence indicated that the *R* configuration in the α -position was the optimal fit for the minor groove, whereas the *S* configuration would clash against the DNA wall. Additionally, chiral amino substitutions in the β -position of the turn showed better affinity than the α -substitutions with slightly higher affinity for the *R* enantiomer as the amine moiety points up and out of the minor groove.

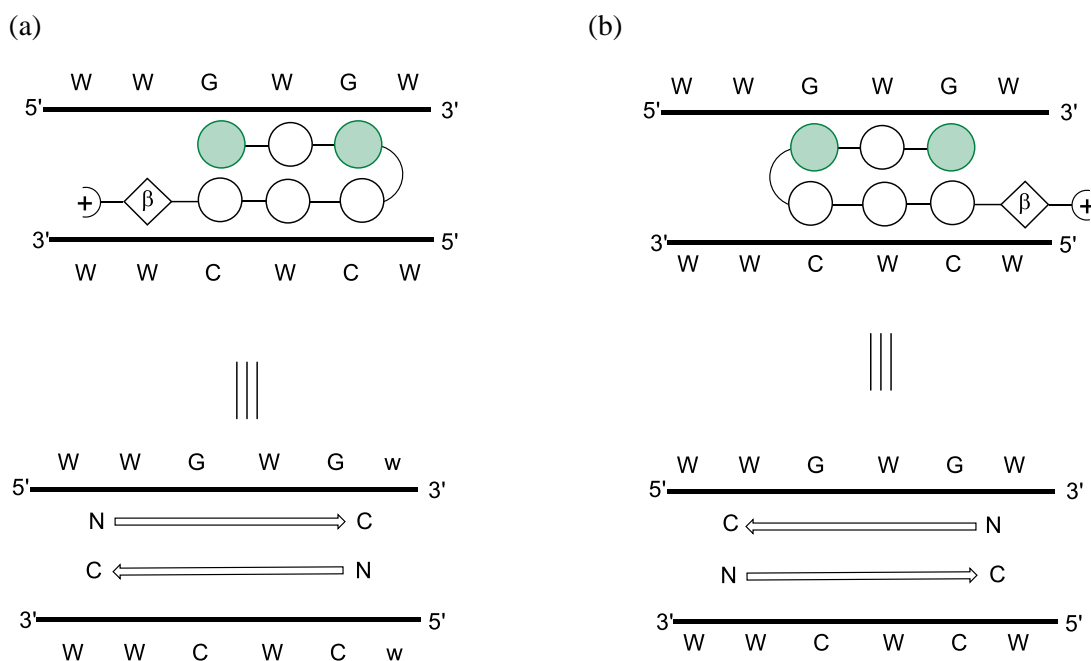


Figure 1.25 Schematic representation of the (a) forward and (b) reverse binding of polyamide (W=A/T).

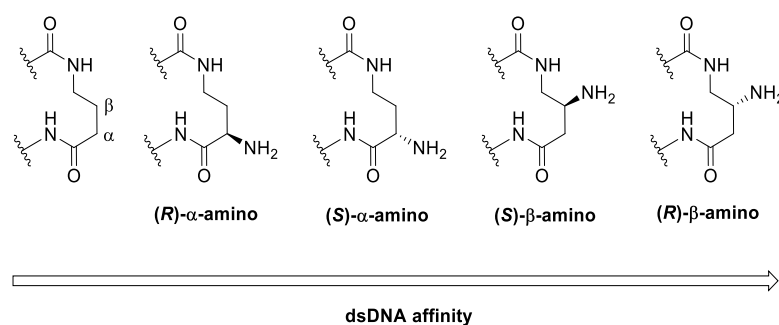


Figure 1.26 Structures of proposed modifications at the γ -turn highlighting enhanced dsDNA binding affinity.¹⁰⁸

1.7.2.3 Polyamides targeting longer dsDNA sequences

A clear disadvantage of 6-8 ring hairpin and cyclic polyamides over oligonucleotide strategies is the ability to recognise only relatively short dsDNA sequences ranging from 5-7 base pairs, which would appear several times throughout the entire genome. With the aim of addressing this limitation, larger polyamide scaffolds have been developed by the Dervan and Sugiyama groups.^{100,109–113} Common strategies adopted include the insertion of flexible β -alanine units, which reset the curvature of the ligand and overcome the 5 consecutive aromatic ring limit and chemical ligation of hairpin and/or cyclic units, which can occur both through turn-turn or turn-C terminus conjugation (Figure 1.27).¹¹⁰ However, limited synthetic tractability and variable cell uptake properties pose challenges in their development as gene regulating agents.

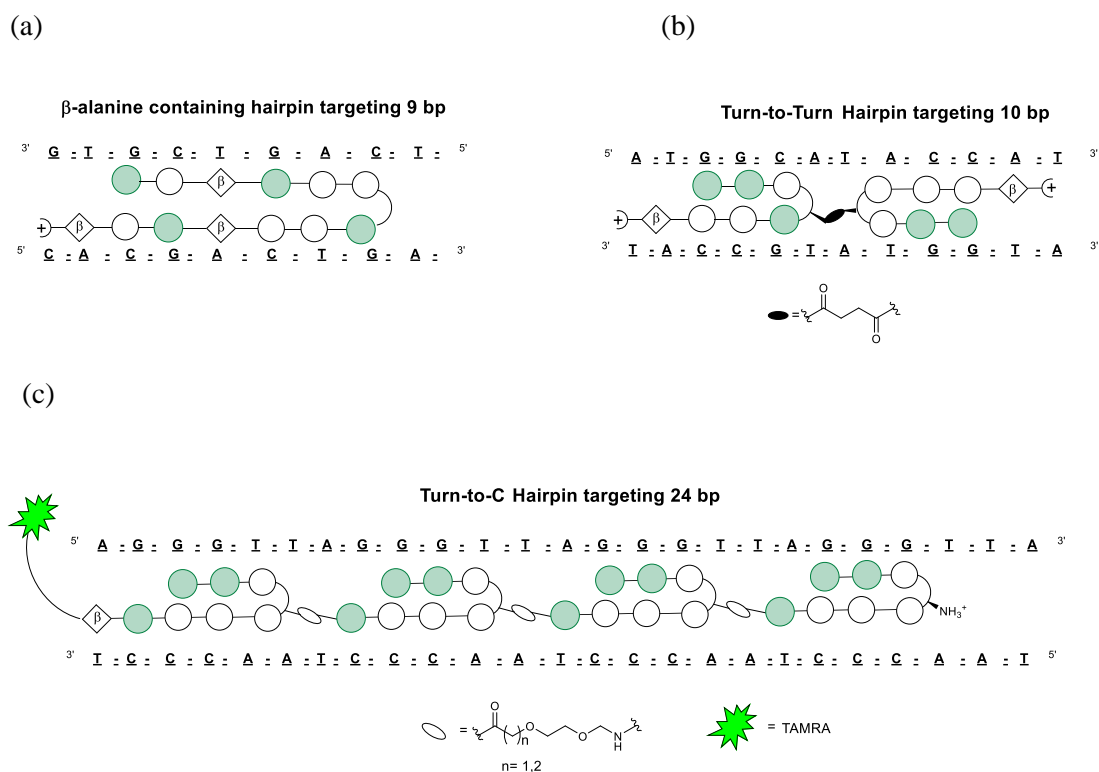


Figure 1.27 Schematic representation of (a) a polyamide containing β -alanine units targeting 9 bp.¹¹⁴ (b) turn-to-turn conjugate polyamide targeting 10 bp.¹¹⁵ (c) turn-to-C conjugate polyamide targeting 24 bp.¹¹⁰

1.8 Downregulation of genes controlled by the Androgen Receptor using polyamides

The high dsDNA binding affinity and sequence selectivity of polyamides have resulted in their development as antitumor agents for prostate cancer chemotherapy. Exploiting established pairing rules, a set of hairpin and cyclic polyamides were designed to bind to DNA sequences associated with the responsive elements of androgen, estrogen and glucocorticoid NRs.^{116–118} As the scope of this work mainly addresses the treatment of prostate cancer, the description of biological and therapeutic effects is limited to compounds targeting the androgen response element (ARE).

The androgen receptor (AR) is a hormone NR that plays a crucial role in the regulation of the development of male physiological processes and reproductive system, and the disruption of its normal activity is often related to the onset of prostate cancer.^{119–121} The AR is characterised by a C-terminal ligand binding domain (LBD), composed of twelve α -helices surrounding a hydrophobic ligand binding pocket. The binding of hormones, such as testosterone and dihydrotestosterone (DHT), induces conformational changes in the AR, which then recruits the binding of co-factors, which in turn mediates the binding of the N-terminal AR domain to DNA. The AR DNA-binding domain (DBD) includes two zinc fingers that recognise the ARE sequence, which consists of minor deviations from the general palindromic sequence 5'-AGAACTnnnTGTTCT (n = A,T,C or G) (Figure 1.29).¹²² The AR is involved in the regulation of prostate specific antigen (PSA) through interactions with proximal (140 bp apart) and distal (thousands of bp apart) DNA regions relative to the TSS, which results in recruitment of TBP, RNA Pol II and other regulatory factors. Generally, prostate cancer cells are dependent on over-activated AR-mediated signals (Figure 1.29). Evidence suggests that the chromosomal rearrangement of the AR responsive gene TMPRSS2 with the oncogene ERG is one of the most frequent mutations (50% of cases) associated with the pathology.^{123,124}

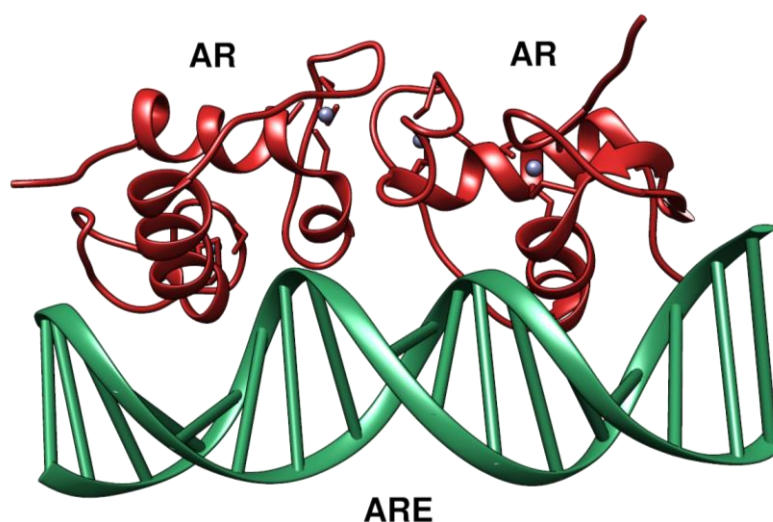


Figure 1.28 Crystal structure of a dimeric DBD-AR complex binding to the palindromic ARE sequence (PDB 1R4I).¹²⁵

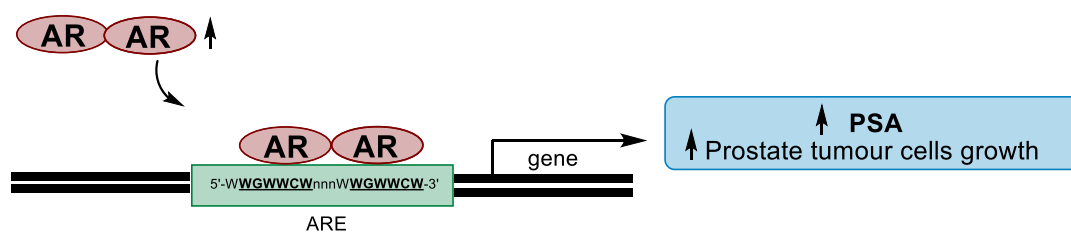


Figure 1.29 Schematic representation of transcriptional activation controlled by the AR binding to the ARE.

The current frontline therapy for prostate cancer consists of androgen deprivation leading to a stall in cancer cell growth and potential remission. However, the cancer often re-emerges in a hormone refractory form.^{126,127} Indeed, the amplification of genes under the control of the AR maintains a central role in the refractory form of the tumour. The transition between the hormone dependent and independent forms is often related to the splicing variant AR-V7 that generates AR active isoforms lacking the LBD.¹²⁸ New therapeutic approaches are therefore required to address the lack of efficacy of current drugs, particularly for the treatment of the hormone refractory form of the disease. The inhibition of dsDNA binding of the AR is extremely attractive as only few mutations have been identified in the DBD of the AR in prostate cancer.¹²⁹ For this reason, a series of hairpin and cyclic polyamides have been developed with the aim of disrupting the DNA-binding action of the AR, thereby inhibiting the first step of the cancer cascade (Figure 1.30).^{117,130–132}

Polyamides **1.10** and **1.11** were designed to recognise the 5'-WGWWCW (W = A or T) sequence, which forms part of the larger ARE sequence 5'-AGAACTnnnTGTTCT. The binding selectivity of **1.10** and **1.11** was confirmed through DNase I footprinting experiments and UV melting stabilisation.¹¹⁷ The suggested mechanism of AR-DNA binding inhibition relies on the structural perturbations induced on the DNA double-helix by the binding of the polyamides. This hypothesis was confirmed by X-ray crystallography studies on cyclic polyamides targeting two different dsDNA sequences. In both cases, the formation of the

complex resulted in modest widening of the minor groove accompanied by increased bending of the duplex axis (Figure 1.30c).^{132,133} A narrowing of the major groove ($>4\text{\AA}$) was observed over the central base pairs of the target sequence recognised by the TFs (Figure 1.30d-e). This mechanism of action was indirectly confirmed by 30-40% reduction in DHT-induced PSA expression in LNCaP prostate cancer cell lines.¹³⁴ However, treatment with Py-Im polyamides also resulted in accumulation of the tumour suppressor p53, suggesting that multiple pathways are affected by polyamide binding.¹³⁵ It was also shown that polyamide **1.10** effectively reduced global RNA Pol II TSS occupancy in LNCaP cells co-treated with DHT, and induced degradation of the RNA Pol II Rbp1 subunit. Further studies on VCaP cells, which contain the TMPRSS2-ERG fusion, showed that polyamide treatment also reduced DNA damage, resulting from overexpression of the oncogene ERG, and perturbed DNA binding recognition and activity of topoisomerases I and II, causing stalling of RNA Pol II.^{131,135} In addition, polyamide occupancy of the target dsDNA sequence effectively blocks RNA Pol II read-through and might also play a role in the mechanism of action.¹³⁶ The lack of DNA damage observed after treatment with Py-Im polyamides represents an advantage compared to current genotoxic treatments for prostate cancer and is a promising step towards anticancer agents that do not affect the health of non-tumour tissues.

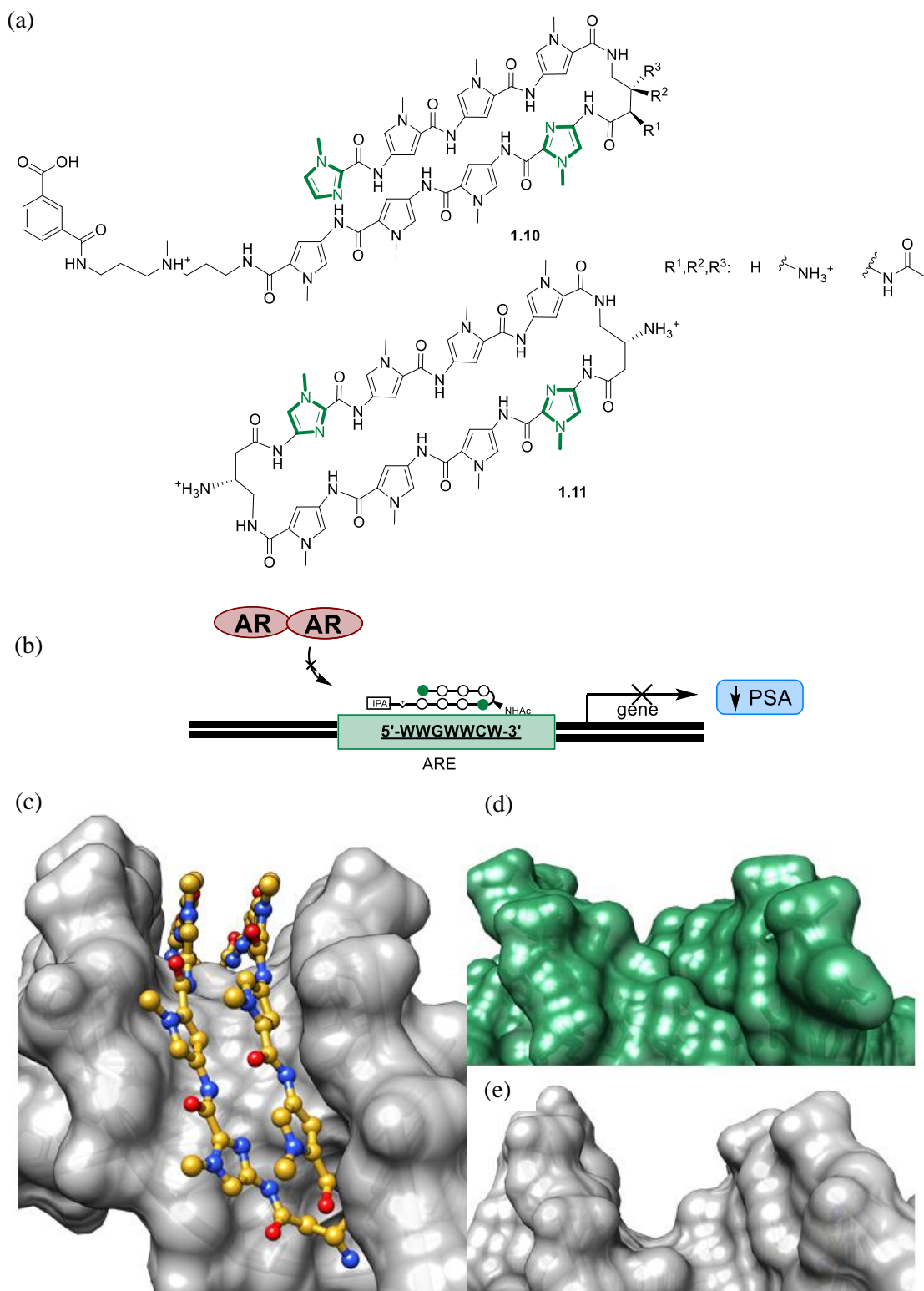


Figure 1.30 (a) Structures of polyamides **1.10-11** targeting the ARE sequence 5'WGWWCW (W= A/T). (b) Schematic representation of the putative mechanism of action of polyamides recognising the ARE sequence. (c) Crystal structure of **1.11** in complex with d(CCATGTACGG)₂ DNA highlighting the widening of the minor groove. Details of the major groove of (d) native DNA fragment (PDB 1D8G)¹³⁷ and (e) DNA bound by **1.11**(PDB 3OMJ).¹³²

1.9 Current limitations of Py-Im polyamide designs targeting the AR

Py-Im polyamides targeting the ARE (5'WGWWCW) showed different levels of antitumour activity based on cyclic and hairpin conformations as well as different substitutions at the C-terminus and γ -turn.^{130,138,139} This alteration in the activity profiles was attributed to differences in cellular permeation.^{130,139} Thus, one of the most challenging aspects in the design of Py-Im polyamides has been the prediction of cellular permeability, which relates to their overall pharmacokinetic properties.¹⁴⁰ Several studies on polyamides labelled with a fluorophore at the C-terminus have revealed that the cellular uptake mechanism is not always clear and is potentially biased by the fluorophore moiety required for visualisation.^{141–144} This was further complicated by differences observed between different fixed and living cell lines. In general polyamides accumulated in the cytoplasm followed by low retention in the nucleus.¹⁴¹ Increasing the number of Im units in the polyamide scaffold decreased cellular and/or nuclear permeation and enhanced toxicity; this made the recognition of G-rich regions of DNA extremely challenging *in vivo*.¹⁴¹

Many modifications on the C-terminus have been investigated to ameliorate the low nuclear uptake (Figure 1.31). For example, Jacobs and Dervan proposed an isophthalic group (IPA) linked through an amide or an oxime to a C3 aliphatic linker to target PSA expression in LNCaP.¹³⁰ While the length of the aliphatic linker did not influence the polyamide's properties, the introduction of the oxime group further inhibited PSA expression, suggesting improved cellular uptake of the compounds. The role of the IPA modification on tissue distribution was evaluated using isotopically labelled (¹⁴C) polyamides.¹⁴⁵ After repeated administrations, polyamides were found to preferentially accumulate in LNCaP tumour cells in comparison to host tissue such as liver, lungs and kidneys.

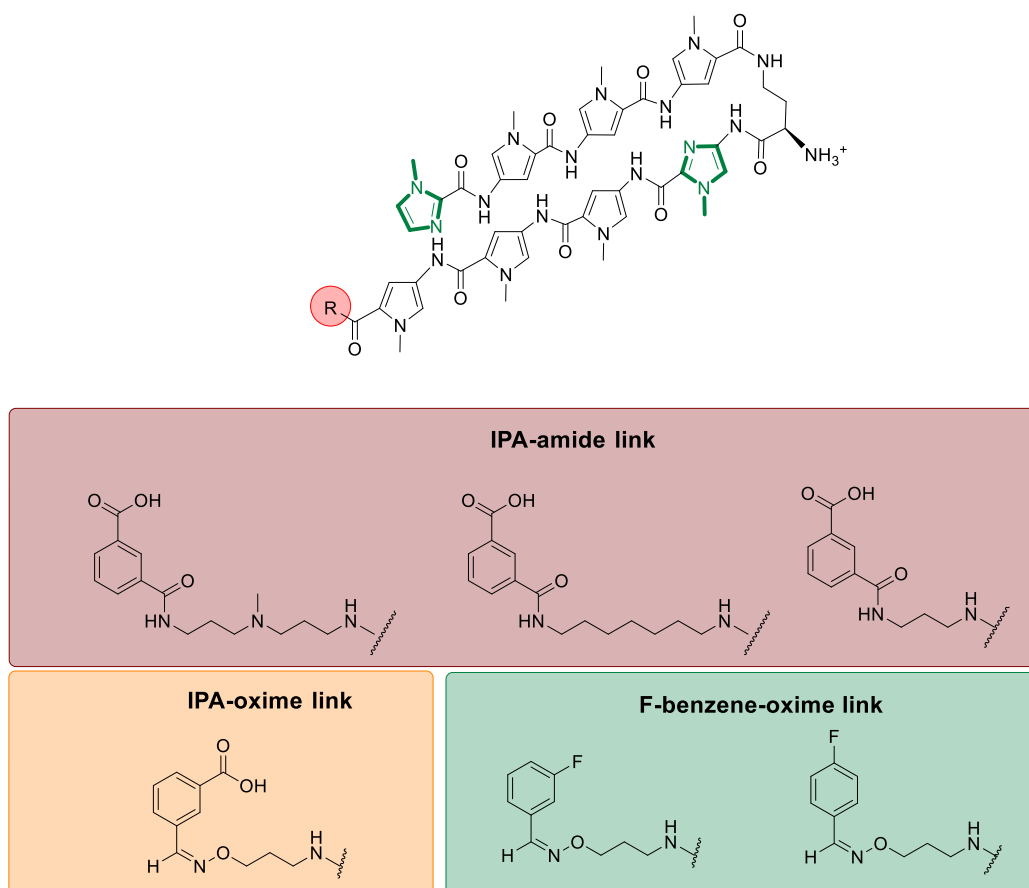


Figure 1.31 Structures of C-terminal modifications in hairpin polyamides targeting the ARE dsDNA sequence.

The same study also indicated that γ -turn substitution influences polyamide activity and distribution. In particular, polyamides with (*R*)- α -amino substitution accumulated less in host tissues such as liver, kidneys and lungs, compared to the acetylated parent compounds.¹³⁸ The effect of several turn modifications both in the α and β positions of the γ -turn was evaluated for cytotoxicity and nuclear permeation on LNCaP cells and animal toxicity *in vivo* (Figure 1.32). Interestingly, large aromatic modifications with an (*R*)- β -orientation resulted in inhibition of LNCaP cell growth at a lower concentration compared to the parent (*R*)- β -amino compounds. Confocal microscopic analysis of analogous FITC-conjugated polyamides suggested better nuclear permeation, presumably induced by the increased lipophilicity of the aromatic function at the γ -turn.¹⁴² However, the development of additional modifications with

(*R*)- β -amino- γ -turn orientation was discouraged by further studies, which compared the animal toxicity profiles induced by hairpin polyamides with (*R*)- β -amino- γ -turn and (*R*)- α -amino- γ -turn. For example, polyamides with acetylated (*R*)- α -amino- γ -turn resulted in no evident toxicity in rodents, whilst retaining activity towards LNCaP xenografts. In contrast, acetylated (*R*)- β -amino- γ -turn analogues caused the death of the animal at the same concentration.

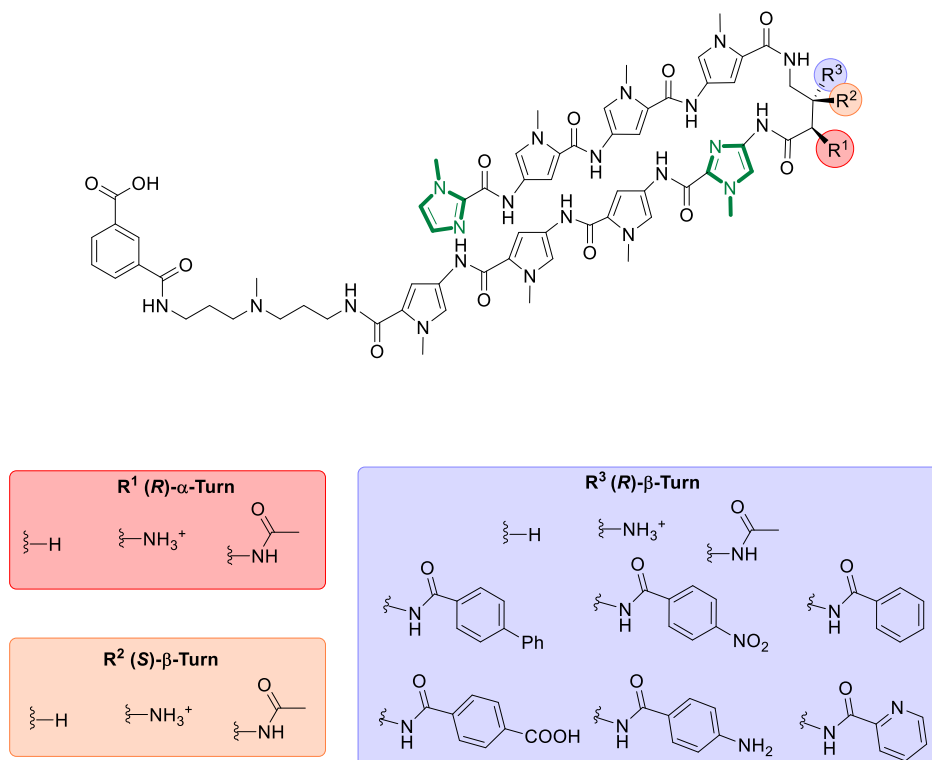


Figure 1.32 Structures of γ -turn modifications introduced in hairpin polyamides targeting the ARE dsDNA sequence.

To date, the relative effects of all of these modifications have not been directly compared, therefore there is no comprehensive set of rules for designing optimal compounds with enhanced inhibitory activity of PSA expression, cellular uptake and tumour-specific tissue localisation *in vivo*. Our current knowledge suggests that modifications at the (*S*) and (*R*)- β -amino- γ -turn improve the DNA binding affinity, but also result in increased *in vivo* toxicity compared to polyamides with the (*R*)- α -amino- γ -turn substitution. Enhanced cellular uptake was observed by introducing IPA at the C-terminus of the polyamide and increasing

hydrophobicity, through γ -turn substitution and decreasing the Im content.

1.10 Expanding heterocyclic diversity of Py-Im polyamides

All current strategies adopted to address the limiting factor of Py-Im polyamides have focused on further functionalisation of the polyamide core. This increases the molecular weight, complicates synthetic routes, and occasionally results in an unclear structure-activity relationship profile.^{138–140,143,146} The possibility of modulating the hydrophobic character of the polyamide core by exploiting different 5-member heterocyclic rings represents an alternative approach for addressing these limitations. At present, the investigation of novel building blocks has focused on the identification of alternatives to Hp for discrimination of T and A nucleotides.^{147,148} Aromatic rings such as furan (Fr), pyrazole (Pz), thiophene (Tn, Tp and Ht) and thiazole (Th and Nt) have been studied either in hairpin or 1:1 linear motifs (Figure 1.33). Calculations of the binding affinities for dsDNA suggested distinct behaviours when building blocks were introduced into different polyamide structures (Table 1.2). Incorporation of alternative building blocks in the hairpin motif has been complicated by the increased steric occupancy generated by the side-by-side arrangement of the aromatic rings in the minor groove. For example, building blocks such as Fr and Th, which report good affinity in the 1:1 motif, completely lost affinity if inserted into the hairpin motif.¹⁴⁷ The position of the building blocks in the scaffold also influenced binding affinity and selectivity. The majority of building block investigations have been limited to internal positions of the hairpin and linear polyamide scaffolds, where heterocycles are linked at both the extremities. In this case, the natural curvature of the rings can have a strong influence on the ability of the ligand to match the shape of the minor groove. Interestingly, less is known about substitutions at the N-terminal position of polyamides, where the heterocycles are linked only at the carboxylic acid and are therefore better able to adapt to the shape of the minor groove. For example, in contrast to thiophene rings incorporated into internal positions, 3-methoxy and chloro thiophenes (Tp-

OMe and Tp-Cl) have been reported to discriminate T over A, when incorporated at the N-terminus of 8-ring hairpin polyamides (Figure 1.34 and Table 1.3).¹⁴⁹

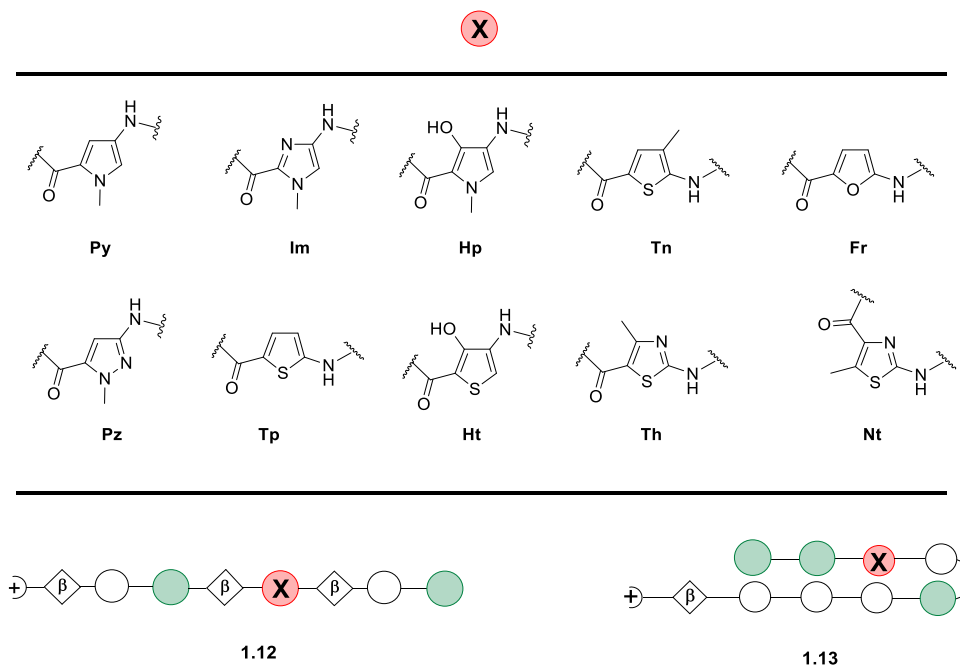


Figure 1.33 Structures of the heterocycles inserted into 1:1 motif /hairpin polyamides.

Table 1.2 Qualitative binding analysis of novel building blocks in 1:1 motif and hairpin polyamides.¹⁴⁷

Building block	1:1 Motif				Hairpin			
	A T	T A	C G	C G	A T	T A	C G	C G
Py	+	+	-	-	+	+	-	-
Im	+	+	+	+	-	-	-	+
Fr	+	+	+	+	-	-	-	-
Pz	/	/	/	/	+	+	-	-
Hp	+	+	-	-	-	+	-	-
Tn	+	+	-	-	+	+	-	-
Tp	/	/	/	/	+	+	-	-
Ht	+	+	-	-	-	-	-	-
Th	+	+	-	-	-	-	-	-
Nt	+	+	+	+	/	/	/	/

+/- affinity/no affinity, / no data acquired

Based on K_a calculated from DNase I footprinting experiment.

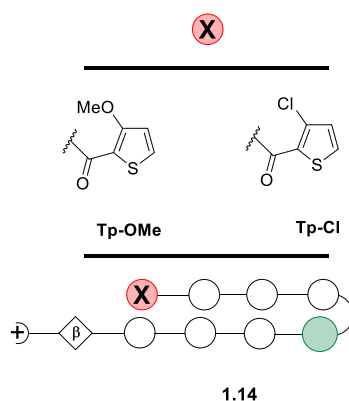


Figure 1.34 Structures of heterocycles installed at the N-terminal position of the polyamide.

Table 1.3 Qualitative binding analysis of polyamides containing Tp-OMe or Tp-Cl at the N-terminal position.

Building block	Hairpin			
	A T	T A	C G	C G
Tp-OMe	-	+	-	-
Tp-Cl	-	+	-	-

+/- affinity/no affinity.

Based on K_a calculated from DNase I footprinting experiment.

1.10.1 Thiazole containing MGBs

The 1,3-substitution pattern of thiazole aromatic rings is similar to imidazole and provides an attractive alternative for the selective recognition of G.¹⁵⁰ As for the imidazole unit, thiazoles contain a hydrogen bond acceptor (either S or N), which can be used to select G through H bonding with G(N2) in the minor groove. As a result of the lower dipole moment, thiazoles are considered intrinsically more hydrophobic than imidazole.¹⁵¹ The higher pK_a of N3 protonated conjugate acid of the imidazole ring (7.10) compared to the thiazole analogue (2.52) indicates that Im(N3) is significantly more basic.¹⁵¹

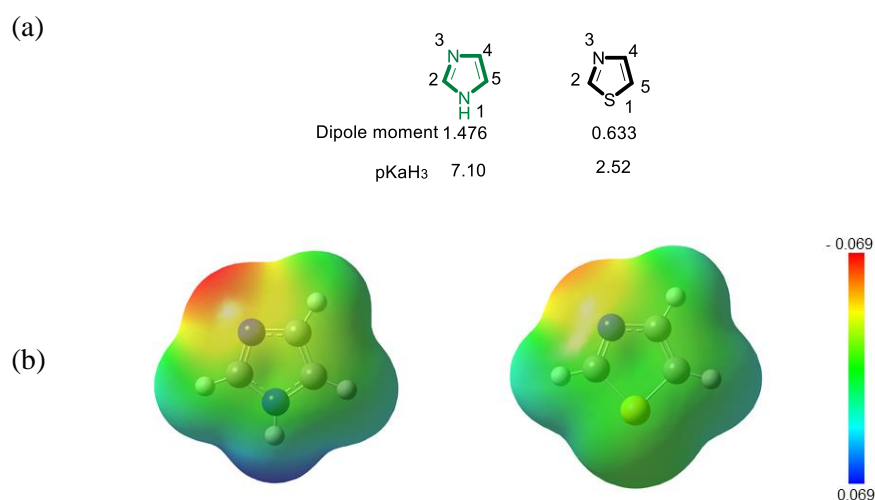


Figure 1.35 (a) Comparison of the physical properties of an imidazole and thiazole aromatic ring.¹⁵¹ (b) ESP of imidazole and thiazole rings.

The incorporation of thiazole building blocks into small untethered MGBs was pioneered by the Lown group.^{97,152–154} Compounds incorporating 2-amino-5-carboxylate thiazole (Th) resulted in weaker DNA binding, which strictly recognised AT sequences (Figure 1.36). Molecular modelling studies suggested that the sulphur atom was directed toward the floor of the minor groove. Despite the potential of H-bonding, it was suggested that the large dimension of the sulphur generated a steric clash against G(N2) located in the minor groove (Figure 1.37).

Nguyen *et al.* reported the insertion of a 4-methyl-Th building block in an 8-ring hairpin polyamide with the aim of creating a new building block to discriminate T over A (Figure 1.36).¹⁵⁵ It was hypothesised that the steric hindrance of the sulphur would induce a steric clash with the aromatic proton A(H2), biasing the binding towards T. However, a decreased binding affinity for dsDNA was observed when 4-methyl-Th was incorporated into hairpin polyamides.¹⁴⁷

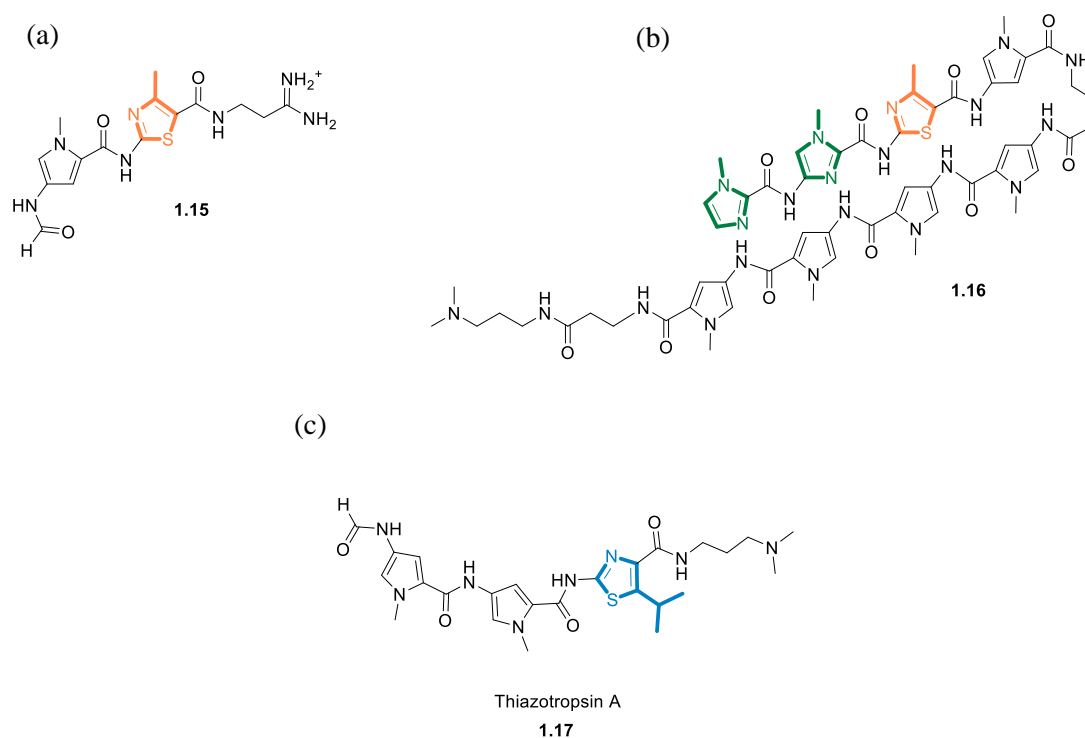


Figure 1.36 (a) Structure of the first MGB containing a Th **1.15**. (b) Structure of 8-ring hairpin polyamide incorporating Th **1.16**.¹⁵⁵ (c) Structure of thiazotropsin A **1.17**.¹⁵⁶

Exploration of 2-amino-4-carboxylate thiazoles (Nt) was limited to untethered, linear 1:1 motifs and H-pin polyamides.¹⁴⁷ Interestingly, changing the orientation of the carboxylic acid substitution to position 4 of the thiazole ring resulted in a different binding behaviour. In fact, Nt tolerated G/C regions with a recognition profile similar to an imidazole building block, suggesting that the nitrogen (N3) was facing into the minor groove (Figure 1.36).

Nt building blocks in small untethered MGBs have been largely explored by the Suckling group with the aim of expanding their antibacterial properties.^{157–162} Nt-containing MGBs (named thiazotropsins) showed an enhanced biological effect towards Gram negative bacteria compared to analogues lacking Nt. Initially, the increased cellular permeability, which correlated with the enhanced hydrophobicity of the MGB, was suggested as the main explanation of this phenomenon.¹⁵⁹ It was also demonstrated that Nt inhibits the efflux pump mechanism, which is adopted by bacteria as a method of resistance.¹⁵⁹ Furthermore,

thiazotropins displayed decreased non-specific toxicity compared to the Im-containing MGBs, which was attributed to the lower polarity of the Nt building block. The Suckling group also explored different substitutions (*i.e.*, isopropyl and isopentyl) at position 5 of Nt as an additional means to modulate MGBs hydrophobicity. The analogue thiazotropin A exhibited a binding preference for 5'ACTAGT, confirming G-recognition of the Nt building block (Figure 1.37).¹⁵⁶ The DNA binding mode of thiazotropin A (**1.17**) was investigated by NMR, which suggested that the head-to-tail staggered dimeric disposition was adopted within the minor groove (Figure 1.38).^{82,157,163,164} Thus, the hydrophobic character and DNA recognition properties of Nt rings are potentially attractive variables for addressing the limited pharmacokinetics of hairpin polyamides.

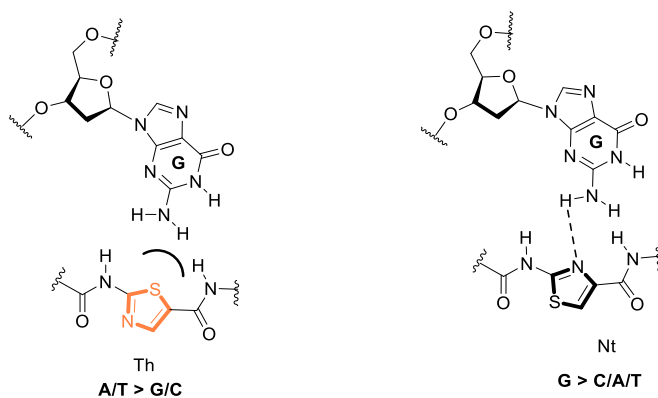


Figure 1.37 (a) Schematic representation of Th and Nt interactions with G.

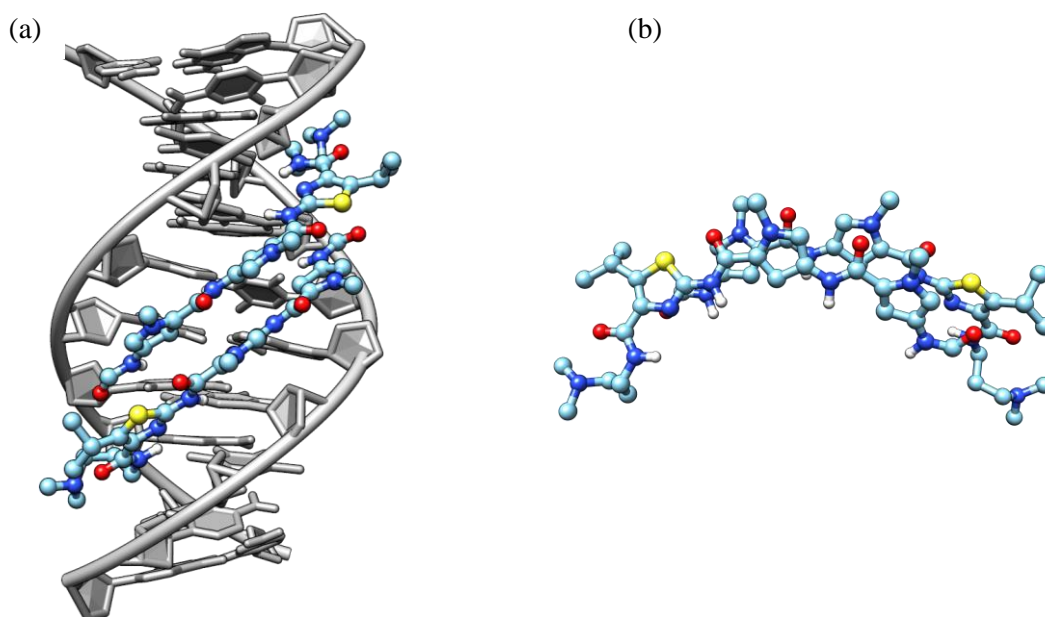


Figure 1.38 (a) NMR-derived structure of thiazotropsin A interacting with dsDNA containing the target 5'ACTAGT. (b) Relationship between molecules of thiazotropsin A assuming a staggered dimeric conformation (PDB 1RMX).⁸²

1.11 Hypothesis

The hypothesis to be tested in this project is that the insertion of the Nt building block in representative hairpin polyamides will retain binding selectivity for G in a target dsDNA sequence. The assumption is that Nt will adopt a similar binding profile to Im, but will increase the general hydrophobicity, offering opportunities to improve cell permeation in biological systems. This hypothesis is further corroborated by the predicted logD value of polyamides incorporating Nt, which also suggested the influence of 5-alkyl substitutions on the overall polyamide lipophilicity (Figure 1.39).

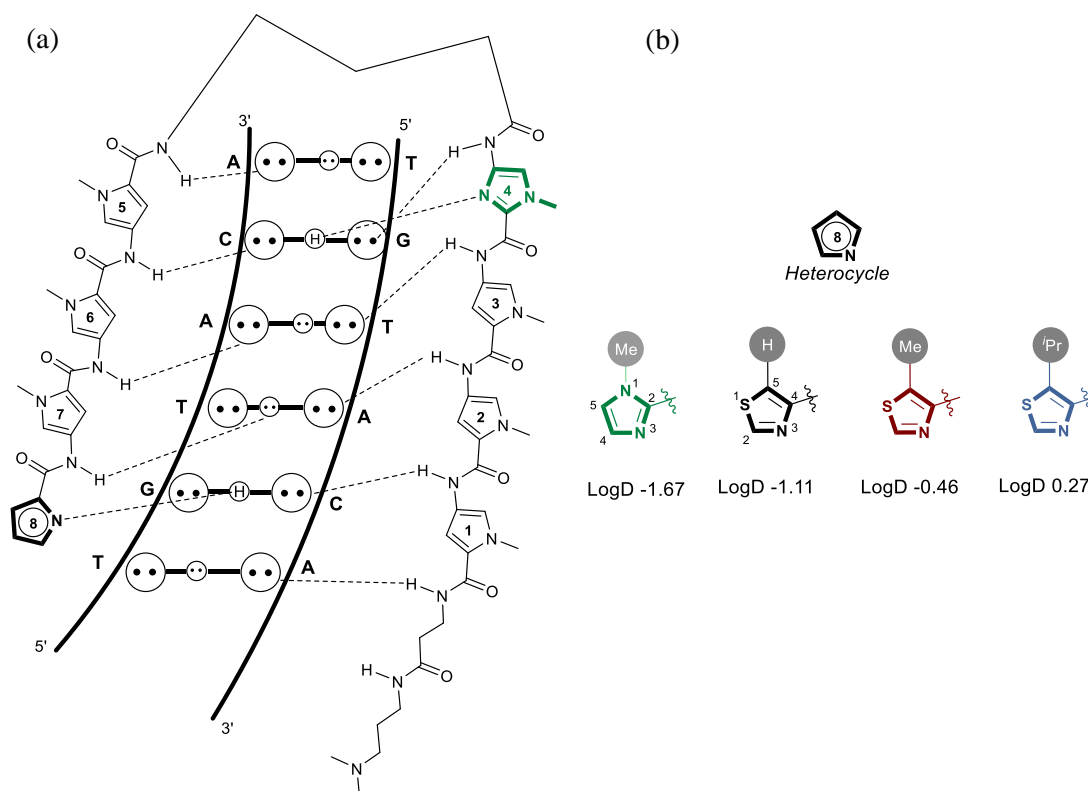


Figure 1.39 (a) Schematic representation of polyamide:DNA interactions. (b) LogD values at pH 7.4 of polyamides incorporating different heterocycles at position 8, predicted using ChemAxon.

Structural characterisation of Nt-containing polyamides in complex with a target dsDNA sequence will give insight into the perturbation of the DNA double helix induced by systematic modification of the ligands, thereby providing the basis for the rational design of a new generation of polyamides that disrupt the AR binding to its target ARE sequence.

1.12 Aims and Objectives

The specific aims of this project are as follows:

- (i) Synthesis of Nt amino acid building blocks incorporating R groups in the 5-position (Figure 1.40, Chapter 2).

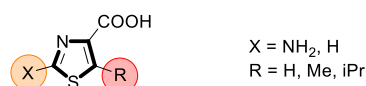


Figure 1.40 General structure of Nt building block.

- (ii) Incorporation of Nt building blocks into defined positions of an 8-ring polyamide scaffold by solid phase synthesis (Figure 1.41, Chapter 2).

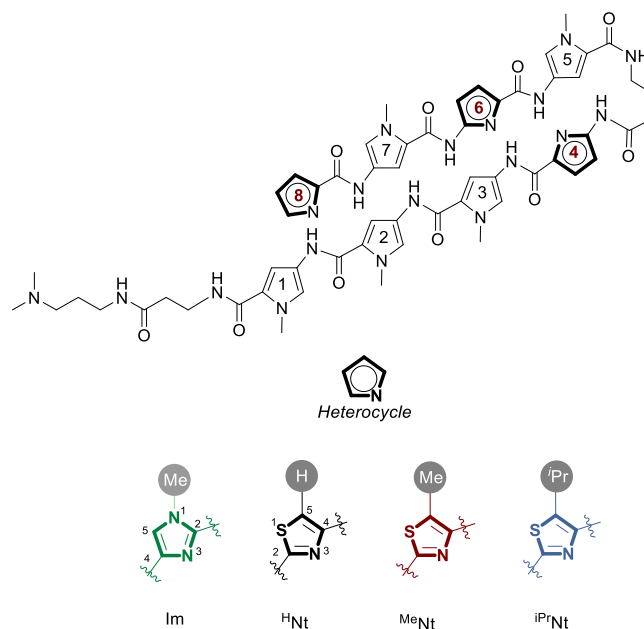


Figure 1.41 (a) Positions of the polyamide scaffold proposed for incorporation of Nt building blocks. (b) Heterocycles investigated in this work.

- (iii) Comparison of the dsDNA binding properties of Nt-containing polyamides with cognate Im-containing analogues (Chapter 3).
- (iv) Structural characterisation of Nt-containing polyamides in complex with target dsDNA sequences using a combination of NMR spectroscopy and molecular dynamics (Chapter 4).

Chapter 2

Synthesis of thiazole-containing polyamides

2.1 Synthetic strategies for the preparation of polyamides

2.1.1 Biosynthesis of distamycin A and netropsin

After their first isolation in the 1960s, distamycin A **1.3** and netropsin **1.4** received widespread interest for their antiviral and antibacterial properties.^{7,165–167} Recent work by Hao *et al.* and Vingadassalon *et al.* identified the gene clusters responsible for the production of distamycin A and netropsin in *S. ambofaciens* (Figure 2.1).^{168–170} The monomeric acetylated pyrrole **2.1** was recognised as the common precursor for the biosynthesis of both the natural products and a variety of oligopyrrole derivatives with hybrid structures between distamycin and netropsin. Even though the mechanism used to regulate the number of rings incorporated into the compounds has not been elucidated, a number of non-ribosomal peptide synthetases (NRPS) is presumed to be responsible for stepwise incorporation of the monomeric unit **2.1** (Scheme 2.1).^{169,170} In the case of netropsin, the guanidine-containing group **2.9** is presumably inserted after incorporation of two pyrrole rings (**2.11**), followed by formation of **2.12** upon NRPS-mediated addition of the amidine tail **2.5**. In the case of distamycin, incorporation of three pyrrole building blocks is followed by removal of the acetyl group of the third pyrrole by amidohydrolase *dst24* (**2.3**), formylation by the formyltransferase *dst26* and incorporation of the amidine tail to obtain **2.6**. The final step for both products is methylation of the pyrroles by the methyltransferase *dst15* using *S*-adenosyl methionine (SAM) as a co-factor.

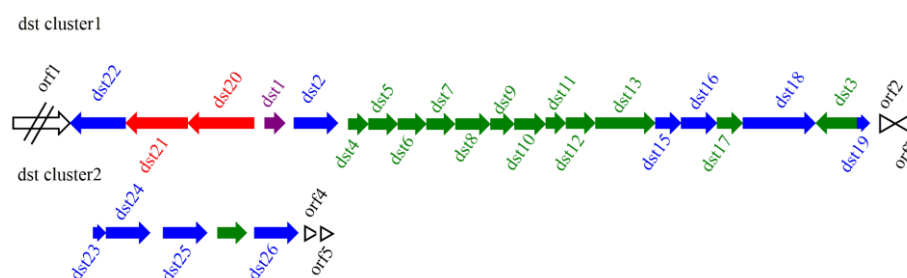
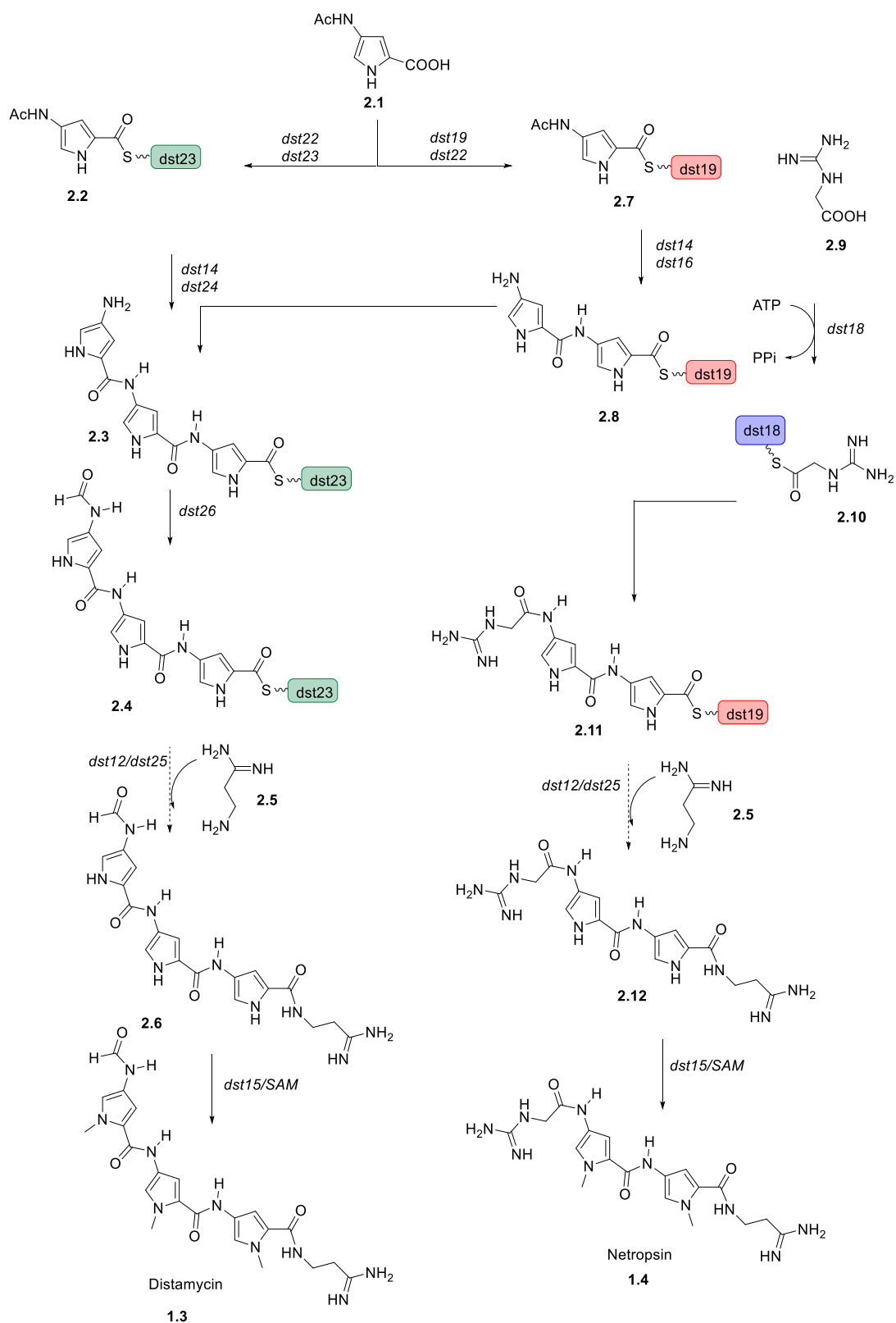
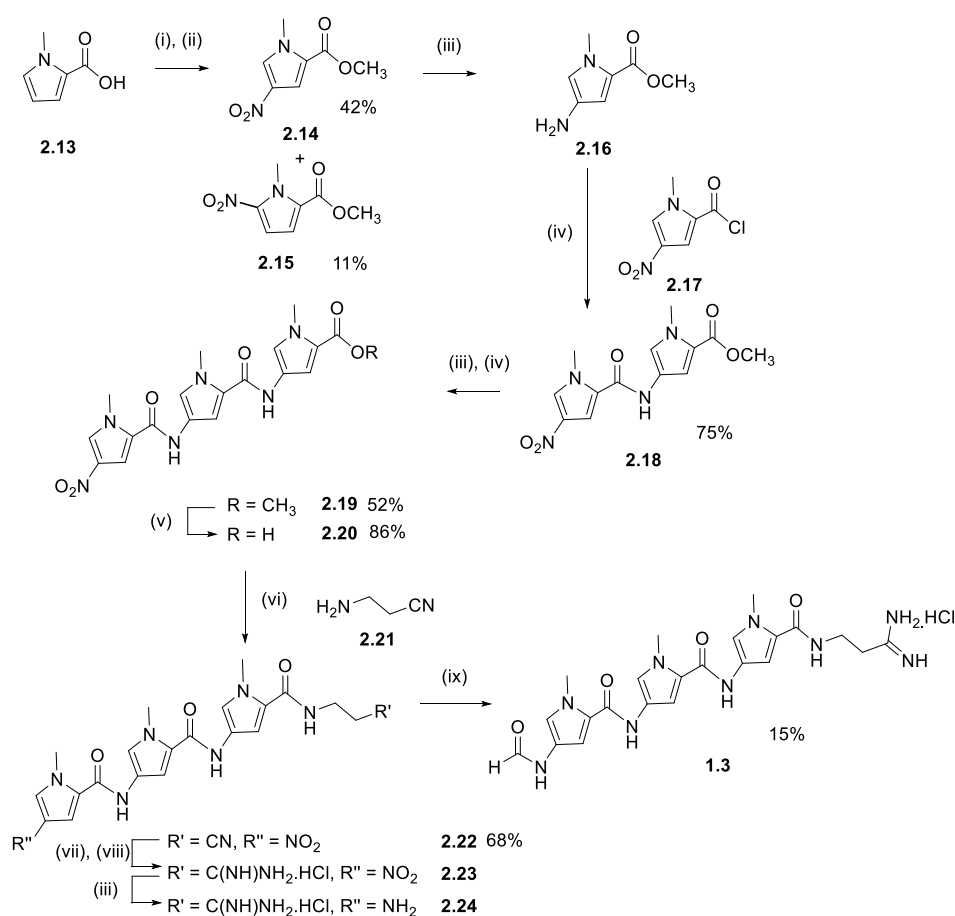


Figure 2.1 Schematic representation of the gene clusters involved in distamycin A biosynthesis (white arrow = gene delimiting the cluster, red arrow = putative resistant gene, purple arrow = putative regulatory gene, green arrow = putative precursor biosynthetic gene, blue arrow = putative assembly gene).¹⁷⁰



2.1.2 Chemical synthesis of distamycin A

The antimicrobial activities demonstrated by netropsin and distamycin A prompted the development of synthetic routes to access libraries of analogues with potentially enhanced biological activities.^{7,165–167} The modular incorporation of aromatic monomers (*e.g.*, **2.1**) used naturally by *S. ambofaciens* was anticipated by chemists to be amenable for the design of suitable pathways for chemical synthesis. The first synthesis of distamycin was reported by Arcamone *et al.* in 1964 and more recently developed by Bialer *et al.*, where both involved iterative amide coupling of monomeric units, followed by nitro reductions (Scheme 2.2).¹⁷¹



Scheme 2.2 Synthesis used by Bialer *et al.* in the total synthesis of distamycin.¹⁷¹ *Reagents and conditions:* (i) HNO_3 , $(\text{CH}_3\text{CO})_2\text{O}$, $-25\text{ }^\circ\text{C}$, 30 min. (ii) H_2SO_4 , MeOH Δ , 12 h. (iii) Pd/C, H_2 (1 atm). (iv) NaHCO_3 , **2.17**, 16 h. (v) NaOH, EtOH, Δ . (vi) DCC, HOBT, **2.21**, rt, 16 h. (vii) HCl. (viii) NH_3 , EtOH. (ix) HCO_2H , DCC, rt, 16 h.

Nitration of the commercially available carboxylic acid **2.13** using concentrated nitric acid in acetic anhydride, followed by esterification afforded **2.14**. A limitation of this method was the generation of the undesired 5-nitro derivative **2.15**, which required purification by column chromatography to afford the desired compound **2.14** in 42% yield. After reduction of the nitro group using Pd-catalysed hydrogenation, the resulting amine **2.16** was coupled to the acyl chloride **2.17**, which was obtained by saponification of **2.14** and treatment with thionyl chloride, yielding dimer **2.18**. This cycle of hydrogenation/coupling was then repeated to give **2.19**, which was itself hydrolysed and coupled with β -aminopropionitrile (**2.21**) to generate **2.22**. The nitrile group of **2.22** was then converted to amidine **2.23** after a stepwise treatment with HCl and ammonia. Hydrogenation of the nitro derivative **2.23**, followed by formylation using formic acid, afforded distamycin A (**1.3**), which was purified via column chromatography. Although limited by a poor-yielding step (*i.e.*, formylation) and successive chromatography purification, this method provided the basis of modern combinatorial approaches currently used for the synthesis of MGBs analogues.^{8,97,162,172–177} A major improvement in MGB synthesis came from the Dervan laboratory, which optimised the synthesis of the monomeric units and established a solid phase synthetic method for the synthesis of Py-Im polyamides.^{101,178}

2.2 Solid phase synthesis of Py-Im polyamides

The requirement for repetitive coupling of building blocks, linked by amide bonds, inspired the Dervan group to develop solid phase methodologies based on cognate peptide approaches. First developed by Merrifield in 1963, solid phase synthesis uses a solid support to facilitate the preparation of long polymeric chains that can be cleaved from the support after completion.¹⁷⁹ Monomeric units are sequentially added to the support without further purification apart from washing cycles that leave the desired products immobilised to the support. The relevant benefits are the enhanced coupling efficiency that derives from the

ability to use excess reagents, and reduced synthesis time. As for standard peptide synthesis, the elongation of the polyamides involves formation of amide bonds between an amine, which is fixed on the solid support, and an activated carboxylic acid, generally used in excess (Scheme 2.3). The amine group of monomer units is protected with either base or acid labile groups according to the nature of the solid support used for the synthesis (see section 2.2.2). After completion, the immobilised polyamide is cleaved off the resin using an appropriate cleaving agent (normally *N,N*-dimethylaminopropylamine (**2.30**) or bis-(3-aminopropyl)-methylamine (**2.31**)) and purified by RP-HPLC.

2.2.1 Synthesis of monomer building blocks

Facilitating practical synthetic routes of MGBs is the availability of large quantities of building blocks (*i.e.*, **2.25-29**), utilised for both solution and solid phase synthesis. From the original synthesis of distamycin A reported by Bialer *et al.*, numerous strategies for multi-gram scale syntheses have been developed from readily available starting materials using a limited number of steps.^{172,178,180} The following sections report the details of the synthesis of the most common monomers Boc-Py-OH (**2.25**) and Boc-Im-OH (**2.26**) (Figure 2.2).

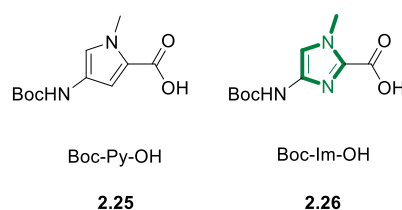
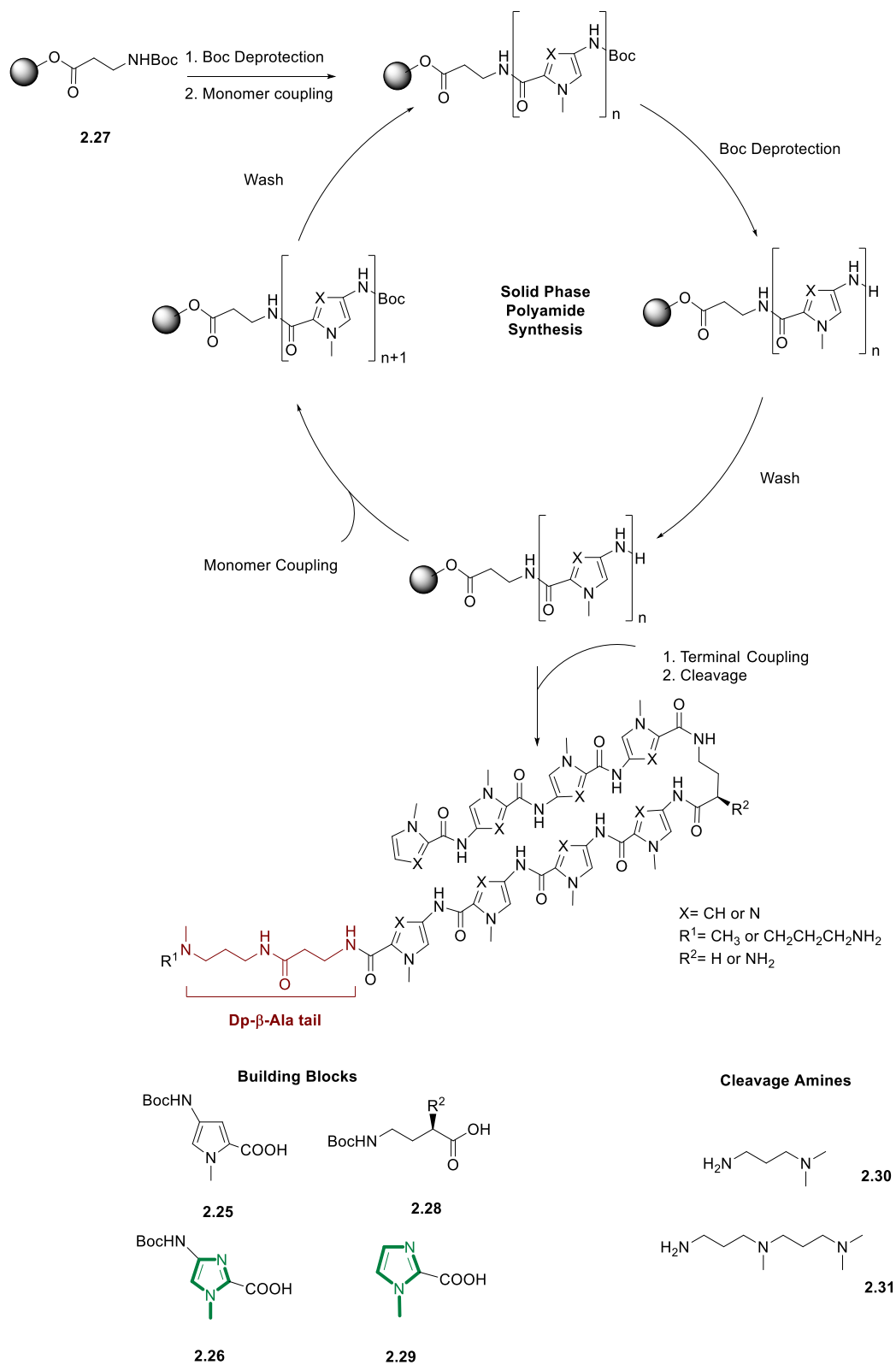


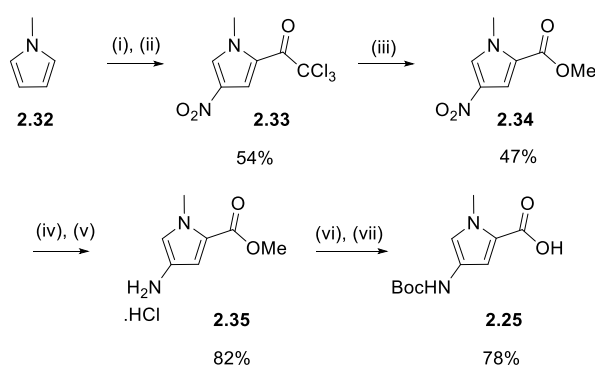
Figure 2.2 Structures of Boc-Py-OH and Boc-Im-OH building blocks.



Scheme 2.3 Schematic representation of a solid phase synthesis cycle using Boc chemistry and Boc-β-Ala-PAM resin.

2.2.1.1 Synthesis of Boc-Py-OH monomer

The synthesis of **2.25** was optimised by Baird *et al.* from the commercially available methylpyrrole **2.32** (Scheme 2.4).¹⁷⁸ Following trichloroacetylation, nitration using concentrated nitric acid in acetic anhydride provided the desired regioisomer **2.33** in 47% yield. Esterification and Pd-catalysed hydrogenation afforded **2.35** as a chloride salt after bubbling of HCl gas. Boc protection and saponification afforded **2.25** over 7 steps in 14% yield without any chromatographic purification.

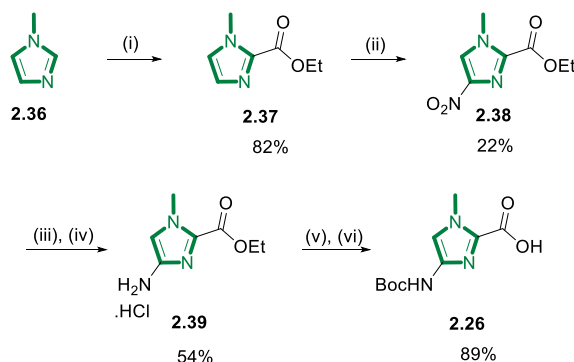


Scheme 2.4 Synthesis of Boc-Py-OH monomer (**2.25**) developed by Baird *et al.*¹⁷⁸ *Reagents and conditions:* (i) Trichloroacetyl chloride, rt, 3 h. (ii) 90 % nitric acid, Ac₂O, -40 °C, 4 h. (iii) NaOMe/MeOH, rt, 2 h. (iv) H₂ (1 atm), 10 % Pd/C, rt, 48 h. (v) HCl. (vi) 10 % Na₂CO₃ (aq), Boc₂O, rt, 3 h. (vii) 2 M NaOH (aq.), 60 °C, 6 h.

2.2.1.2 Synthesis of Boc-Im-OH

As for the Boc-Py-OH, Boc-Im-OH **2.26** was generated from the commercially available N-methylimidazole **2.36** (Scheme 2.5). In this case, the ethyl ester **2.37** was obtained by treatment with ethyl chloroformate and triethylamine.¹⁷⁸ Selective nitration at position 4 was achieved by maintaining the temperature at 0 °C and quenching the reaction after 50 minutes. Analogous to the pyrrole monomer (**2.25**), the ammonium salt **2.39** was obtained after Pd-catalysed hydrogenation followed by bubbling of HCl gas. Final Boc protection and saponification

afforded **2.26** in an overall yield of 8%. The limiting step is the nitration, which returns the desired isomer in 22% yield.



Scheme 2.5 Synthesis of Boc-Im-OH monomer (**2.26**) utilised by Dervan *et. al.*¹⁷⁸ *Reagents and conditions:* (i) Ethyl chloroformate, TEA, -20 °C, 36 h. (ii) H₂SO₄ (conc.), HNO₃, 0 °C, 50 min. (iii) H₂ (1 atm) 10 % Pd/C, rt, 48 h. (iv) HCl. (v) Boc₂O, DIPEA, 60 °C, 18 h. (vi) 1 M NaOH (aq.), 60 °C, 3 h.

2.2.2 Choice of the support in solid phase synthesis of polyamides

Since solid phase synthesis was first reported, a wide range of solid supports, reagents and strategies have been described for enhancing the efficiency of the method. For Py-Im polyamides, the choice of the solid support is also crucial for the design of the ligand. In the first approach proposed by Baird *et al.*,¹⁷⁸ the base labile resin Boc- β -Ala-PAM was chosen due to its thermal stability and resistance to strong acid that allows facile Boc deprotection of the amine of the monomer units. After completion of the desired monomer couplings, polyamides are cleaved from the β -Ala-PAM resin through treatment with *N,N*-dimethylpropylamine (**2.30**, Dp) (Scheme 2.3). This provides polyamides containing a Dp- β -Ala tail on the C-terminus, which was originally thought to have minimal influence on the binding selectivity.¹⁷⁸ However, it was later demonstrated that the Dp- β -Ala tail preferentially binds to A/T over G/C regions of dsDNA, thus complicating the design of polyamides

targeting G/C rich region of DNA.^{181,182} Furthermore, cellular internalisation studies reported the presence of the β -alanine moiety as detrimental for cellular uptake of the polyamides.¹⁴¹

Alternative solid supports were proposed for the purpose of addressing these limitations. For example, Belitsky *et al.* introduced the use of the Kaiser oxime resin, which allowed access to a wide range of compounds lacking the β -alanine moiety (Figure 2.3).¹⁸³ However, the Kaiser oxime resin is more sensitive to acidic conditions, therefore requires more diluted deprotecting “cocktails”, which in turn results in longer deprotection times. Another limitation observed was the increased thermal sensitivity of the resin, which renders it susceptible to the harsher conditions required for difficult couplings.

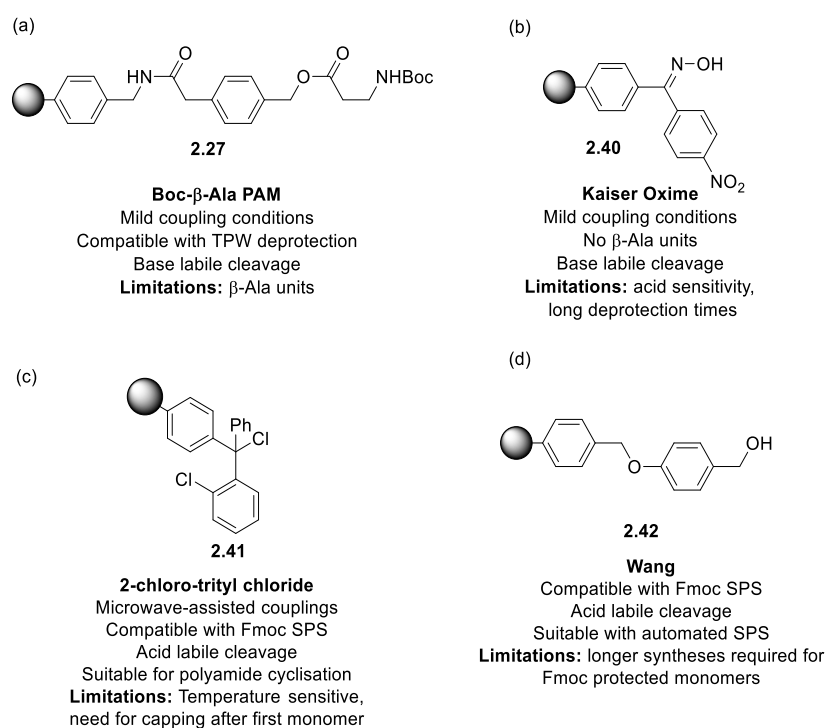
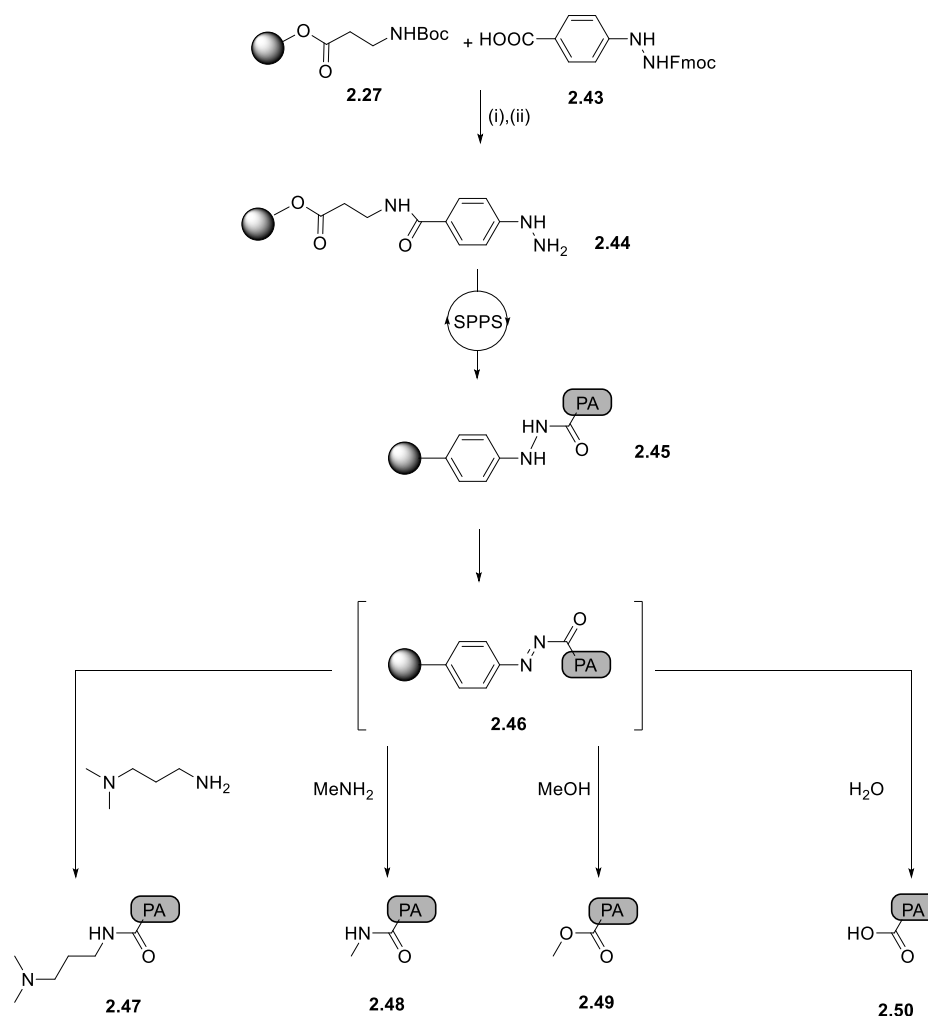


Figure 2.3 Structures of solid supports commonly used for the synthesis of Py-Im polyamides. (a) Boc- β -Ala-PAM.¹⁷⁸ (b) Kaiser oxime.¹⁸³ (c) 2-chloro-trityl chloride resin.¹⁰³ (d) Wang resin.¹⁷²

Previous work in our laboratory also investigated the use of “safety-catch” linkers for polyamides synthesis (Figure 2.3).¹⁸⁴ Hydrazine linkers have found regular use in peptide synthesis as they combine the stability of certain solid supports with the flexibility of modifying the C-terminus upon cleavage.^{185,186} For this method, an aromatic hydrazine is

incorporated between Boc- β -ala-PAM and the first ring of the polyamide scaffold. Before the cleavage, the linker is normally activated through an oxidation step to convert the hydrazine to an active diazene species (**2.46**), which then become susceptible to attack by a wide range of nucleophiles (Scheme 2.6). Although it allows for diverse functionalisation at the C-terminus, this strategy is mainly limited by large batch-to-batch variability in final product purity and coupling efficiency.

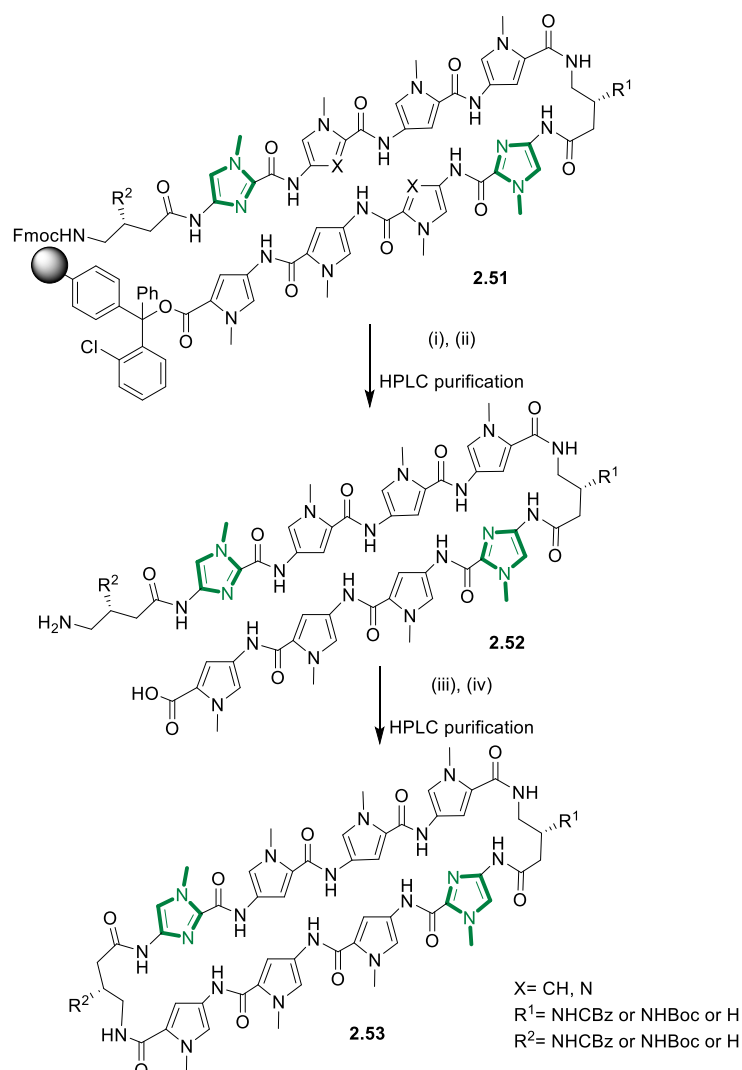
The use of the acid labile 2-chlorotrityl chloride (**2.41**) and Wang (**2.42**) resins has also been explored, allowing the automation of polyamide synthesis using standard peptide synthesisers and microwave-assisted couplings (Figure 2.3).^{103,172} Compatible Fmoc protected monomers are obtained in good yields and on multigram scales. Although there was relevant improvement associated with the automation of the process, this procedure was not considered optimal for the screening of novel coupling conditions. For these reasons, the highly-stable Boc- β -Ala-PAM remains the most frequently used resin for the development of innovative synthetic strategies and incorporations of novel building blocks.



Scheme 2.6 Schematic representation of functionalisation via “safety catch” hydrazine linkers. *Reagents and conditions:* (i) TPW, then HATU, DIPEA, rt, 2 h. (ii) 20% piperidine in DMF, rt, 20 min.¹⁸⁴

2.2.3 Limitations of solid phase synthesis methods

While solid phase methods proved ideal for the synthesis of libraries of hairpin and linear polyamides, these techniques are more laborious for H-pin and cyclic architectures.^{96,103} In the case of cyclic polyamides, a microwave-assisted method using 2-chlorotrityl chloride resin was developed (Scheme 2.7). Although these conditions reduced the coupling times to 30 minutes, the synthesis was complicated by the need for double HPLC purification, which was necessary both after cleavage from the resin and after DPPA-mediated cyclisation.



Scheme 2.7 Schematic representation of the microwaved-assisted synthesis of cyclic polyamides libraries proposed by Li *et al.*¹⁰³ *Reagents and conditions:* (i) 30% HFIP, DCM, rt, 1 h. (ii) 20% piperidine in DMF, rt, 30 min. (iii) DPPA, DIPEA, rt, 20 h. (iv) 10% TFMSA in TFA or TFA, rt, 15 min.

In the case of H-pin polyamides, solid phase synthesis was possible through incorporation of building blocks containing N-alkene functionalities (**2.54**) (Scheme 2.8). In this method, a Grubbs-catalysed metathesis step produced an inter-strand linkage on the solid support.⁹⁶ Following completion and cleavage of **2.57**, reduction, catalysed by ammonium formate and Pd on carbon, afforded alkylic derivative **2.58** that was isolated by HPLC purification.⁹⁶ This synthetic strategy was limited by the possibility of obtaining exclusively symmetric compounds, therefore restricting the sequences of DNA that could be targeted.

Other limitations of solid phase synthesis include poor scalability (normally reactions are limited to milligrams scale), the need for HPLC purifications, and poor atom efficiency. For this reason, the Dervan lab demonstrated the possibility of synthesising 8-ring polyamides in gram scale using solution phase synthesis, with only one chromatographic step being required.¹⁷⁵ In this method, the scaffold of a polyamide targeting the sequence 5'WGWWCW (where W=A/T) present in the ARE, was retrosynthetically divided into 4 main fragments (Figure 2.4).

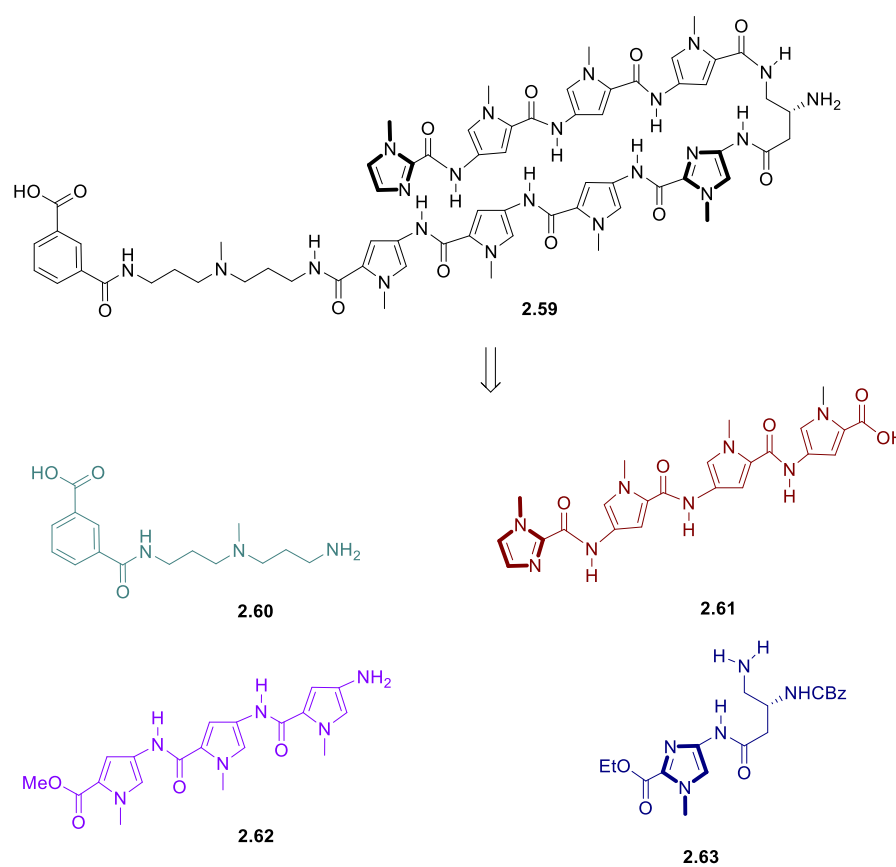
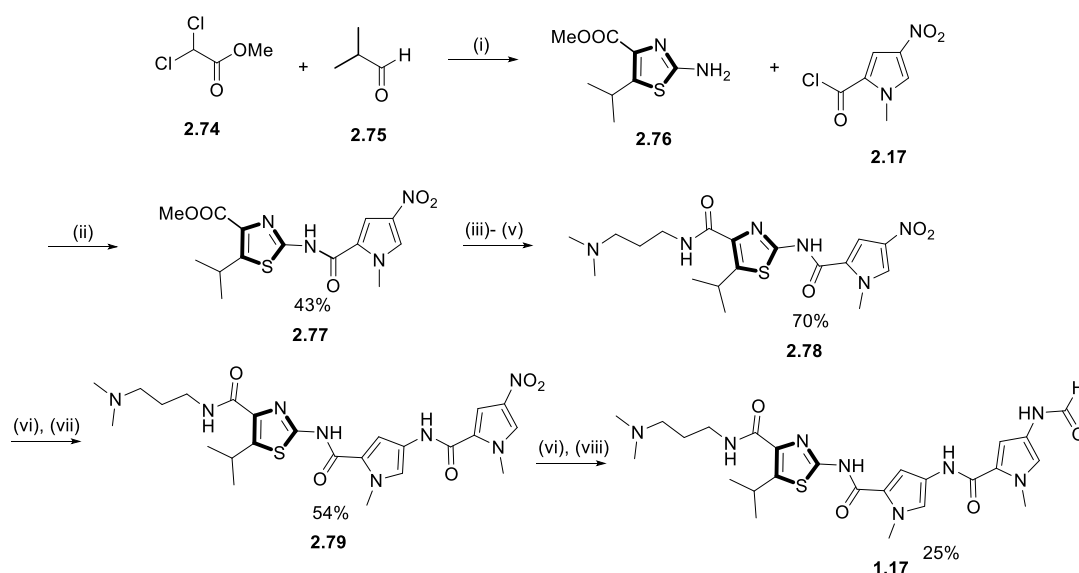


Figure 2.4 Retrosynthetic fragmentation for solution synthesis of **2.59** targeting the ARE dsDNA sequence.¹⁷⁵

The desired fragments were obtained through a series of high yielding deprotection and coupling steps and isolated by precipitation. This allowed the preparation of **2.59** with a purity above 90% and overall 34% yield, with only one chromatographic step following removal of the Cbz group (Scheme 2.9).

2.3 Incorporation of Nt in MGBs

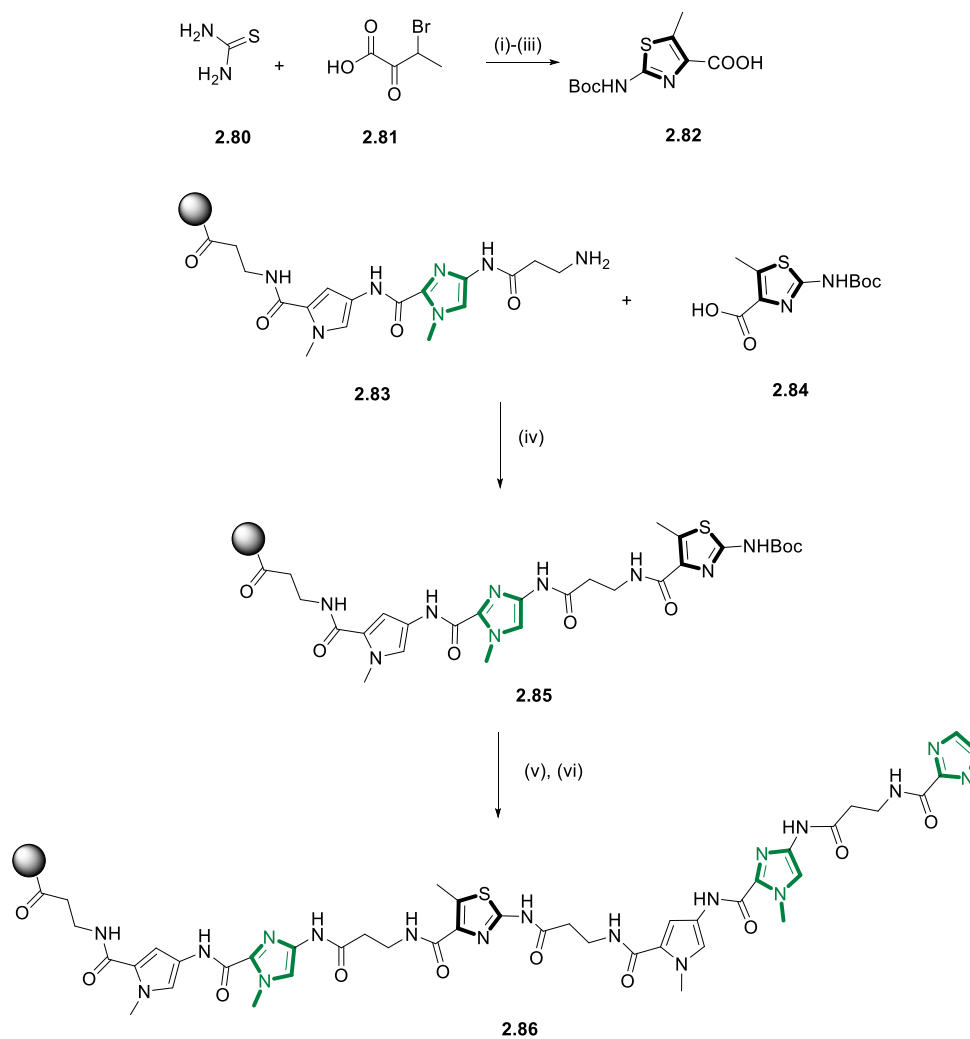
Incorporation of Nt building blocks into small untethered MGBs has been explored by the Suckling group for the synthesis of thiazotropsin A (Scheme 2.10).¹⁶² The isopropyl-substituted building block **2.76** was synthesised using a Darzens condensation.¹⁸⁷ The resulting amine was coupled to acid chloride **2.17** in 43% yield. Ester hydrolysis, followed by coupling with **2.30** provided **2.78**. Reduction of the nitro group of **2.78** followed by coupling with **2.33** provided **2.79**, which was reduced to the amine form and directly formylated to obtain thiazotropsin A in 25% yield.



Scheme 2.10 Synthetic approach used by Khalaf *et al.* for the synthesis of thiazotropsin A (**1.17**)¹⁶²
Reagents and conditions: (i) MeONa, 0 °C, 1 h. then thiourea, Δ, 4 h. (ii) N-methylmorpholine, rt. (iii) 0.25 M KOH (aq.), Δ, 4 h. (iv) SOCl₂, Δ, 3 h. (v) **2.30**, N-methylmorpholine, rt. (vi) H₂ atm., Pd on C. (vii) **2.33**, rt. (viii) Ethylformate, Δ, 12 h.

5-methyl-substituted Nt has also been incorporated into 1:1 linear motifs using Boc-β-Ala-PAM by the Dervan group (Scheme 2.11).¹⁴⁷ In this case, the monomeric unit **2.82** was prepared after bromination of 2-oxobutanoic acid, followed by treatment with thiourea and Boc protection. This was used directly for solid phase synthesis. HBTU-mediated coupling of

the acid was achieved after 20 hours at 37 °C. After deprotection, the HBTU-activated β -alanine unit was incorporated by coupling over 48 hours at 37 °C. The synthesis of **2.86** was completed using standard solid phase conditions.¹⁴⁷



Scheme 2.11 Synthetic approach used by Marques *et al.* for the synthesis of **2.86**.¹⁴⁷ *Reagents and conditions:* (i) thiourea, neat, rt, 16 h. (ii) (Boc)₂O, DIPEA, DMAP, 60 °C, 12 h. (iii) 1 M NaOH (aq.), 35 °C, 1 h. (iv) HBTU, 37 °C, 20 h. (v) TPW, then HBTU, 37 °C, 48 h. (vi) standard SPS conditions.

2.4 Aims and objectives of this chapter

Nt building blocks provide an attractive alternative to C-terminus and γ -turn modifications for tuning the pharmacokinetic properties of Py-Im polyamides (Section 1.11). Furthermore, alkyl substitutions at position 5 of an Nt building block represent a potential modification for modulation of the overall hydrophobicity of the polyamide.¹⁶² The objective of this chapter is to establish synthetic routes for the preparation of Nt-modified polyamides in defined positions. These were identified as position 4, 6 and 8 of an 8-ring hairpin polyamide, which will be used to probe the influence of an Nt unit on the polyamide architecture when incorporated between Py and γ -turn (position 4), two Py rings (position 6) and at the N-terminus (position 8) (Figure 2.5).

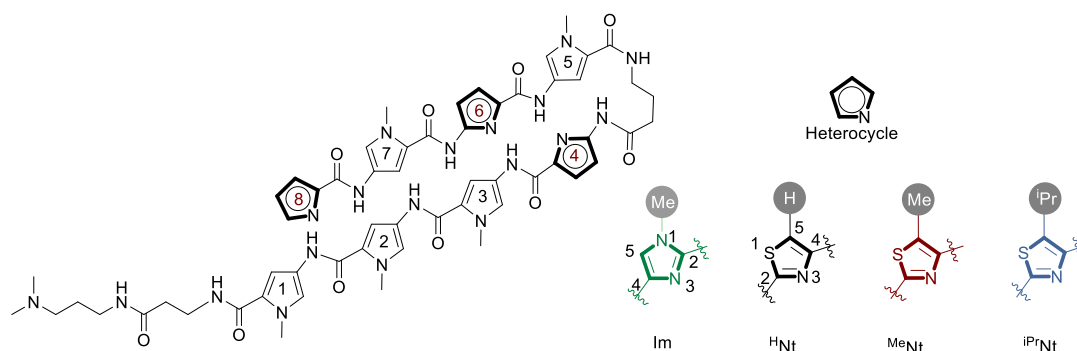


Figure 2.5 Positions of the polyamide scaffold proposed for incorporation of Nt building blocks and heterocycles analysed in this work.

The specific aims of this chapter are as follows:

- (i) Synthesis of Boc-Py-OH (**2.25**) and Boc-Im-OH (**2.26**) building blocks.
- (ii) Optimisation of previous conditions for the solid phase synthesis of Py-Im polyamides.
- (iii) Synthesis of Nt building blocks with 3 different substitutions at position 5 (**2.87-91**) to be used in the solid phase synthesis of Nt-containing polyamides (Figure 2.6).

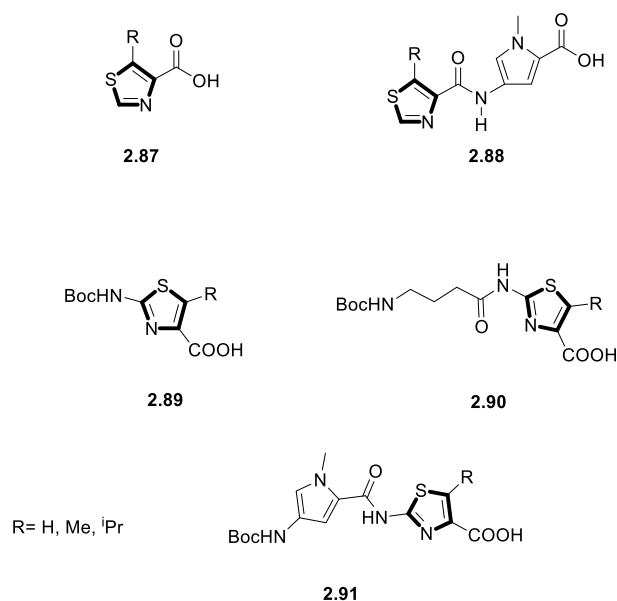
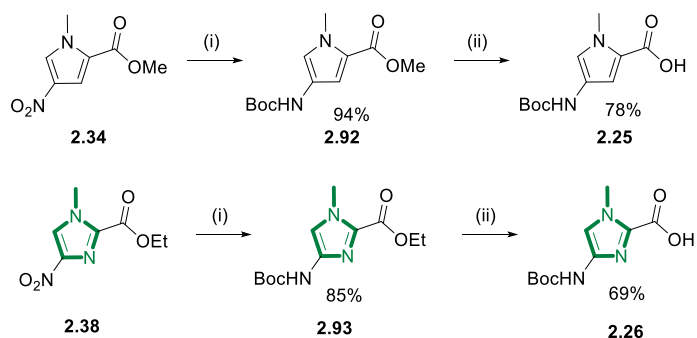


Figure 2.6 Structures of Nt building blocks for SPS.

2.5 Results and discussion

2.5.1 Synthesis of Boc-protected Py and Im amino acid monomers

Boc-Py-OH (**2.25**) and Boc-Im-OH (**2.26**) were synthesised starting from the commercially available nitro-derivatives **2.34** and **2.38** (Scheme 2.12).¹⁷⁸



Scheme 2.12 Reagents and conditions: (i) Boc₂O (1.25 eq.), Pd/C 10% (cat.), H₂ atm., rt, 3 h. (ii) 1 M NaOH (aq.), 70 °C, 4 h.

One-pot reduction and Boc protection of **2.34** and **2.38** afforded the esters **2.92-93** in 94% and 85% yield, respectively. In both cases, sufficiently pure products were isolated after recrystallisation from DCM using hexane. The direct protection of the corresponding amine was used to bypass HCl-mediated salt formation.¹⁷⁸ This resulted in increased yields and simplified synthetic protocols. Saponifications of **2.92** and **2.93** were carried out in a 1:1 mixture of methanol: 1 M aqueous NaOH at 70°C. The Boc-Py-OH building block **2.25** was then isolated following acidification (pH 3) and extraction with ethyl acetate. As reported by Dervan *et al.*, acidification of **2.26** resulted in formation of a slushy white solid, insoluble in organic solvents.¹⁷⁸ Isolation was therefore achieved by lyophilisation after vacuum filtration and several water washes to remove residual salts.

2.5.2 *Optimisation of standard solid phase synthesis protocols*

Initial attempts of polyamide synthesis using manual solid phase highlighted the necessity for further optimisation of existing protocols.^{176,178,188} This was mainly dictated by the large batch-to-batch variability observed during the synthesis of standard Py-Im polyamides. Factors influencing coupling efficacy and yields include resin batch lot, swelling procedure, aromatic composition of the polyamides, coupling agents, reaction times, cleavage conditions and purifications strategies. Most of these complications are not reported in the general literature, which typically focused on the preparation of 6-ring hairpin polyamides.^{178,183,188} The following optimisation aimed to provide reproducible conditions to avoid frustrating coupling failures, which often require complicated HPLC purification of the desired product in the presence of truncated sequences.

2.5.2.1 *Preliminary procedure and choice of the coupling reagents*

Several optimisations of the first reported protocol for solid phase synthesis of Py-Im polyamides by Baird *et al.* have been previously attempted by our group. Initially, preliminary

washing rounds of the resin with DCM and DMF (normally 4×1 minutes for each solvent) were performed prior to the start of the synthesis, with the aim of removing dust and breaking resin aggregates formed during storage. This was generally followed by a swelling cycle (2×30 minutes) using DCM, which guarantees homogeneous accessibility of the amine moieties on the bead surface, resulting in enhanced crude purity. A mixture of TFA, phenol and water (92.5%, 5% and 2.5% respectively, named TPW) was used for deprotection of the Boc group in place of thiol-containing cleavage cocktails previously reported.¹⁷⁸ As first reported by Krutzik *et al.*, phenol represents an optimal alternative to thiophenol as a “scavenger” of tertbutyl cations to prevent capping of the amines on the support.¹⁸⁸

A wide range of deprotection times (from 15 seconds to 30 minutes) have been previously reported for removal of the Boc group from Boc-Py-OH and Boc-Im-OH monomers depending on the resin, the cleavage cocktail and building blocks.¹⁸⁸ In our lab, deprotection times were optimised for TPW to 7 minutes (divided in 3 rounds; 1×1 minutes and 2×3 minutes) for Boc-Py-OH and Boc- γ -turn-OH building blocks, and 21 minutes (divided in 3 rounds 1×1minutes, 2×10 minutes) for Boc-Im-OH, which is notoriously more difficult to deprotect.^{176,188,189} The high stability of Boc- β -Ala-PAM also permits Fmoc deprotection, which allows for the use of the Fmoc-protected commercially available hydrazine “safety-catch” linker (**2.43**, Scheme 2.6). This is normally carried out using a 20% solution of piperidine in DMF for 20 min (divided in 2×10 minutes rounds) at room temperature. The number of base deprotection steps was limited to prevent unwanted resin cleavage.

After deprotection, the resin was normally washed with DMF (4×1 minutes) and conditioned with the appropriate anhydrous solvent used in the coupling. The choice of activating agent is also essential to achieve coupling completion. In general, 7-aza-benzotriazole esters (OAt, **2.94-2.98**) have been shown to be more efficient compared to benzotriazole esters (OBt), initially reported by Baird *et al.* (Figure 2.7).^{178,188} The list of activating agents to be used in combination with carboxylic acid and amine is outlined in Table 2.1.

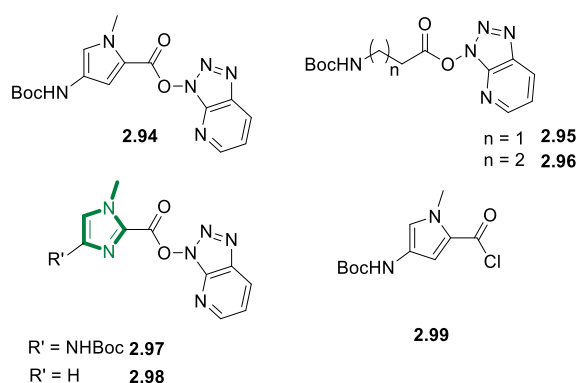
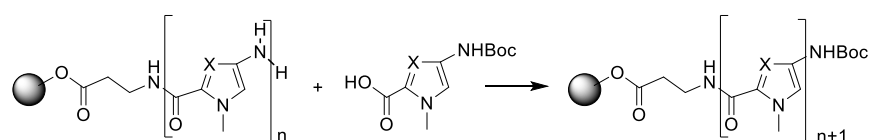


Figure 2.7 Activated species of the monomers used in solid phase synthesis of Py-Im polyamides.

Table 2.1 List of coupling reagents and conditions used in solid phase synthesis of Py-Im polyamides.

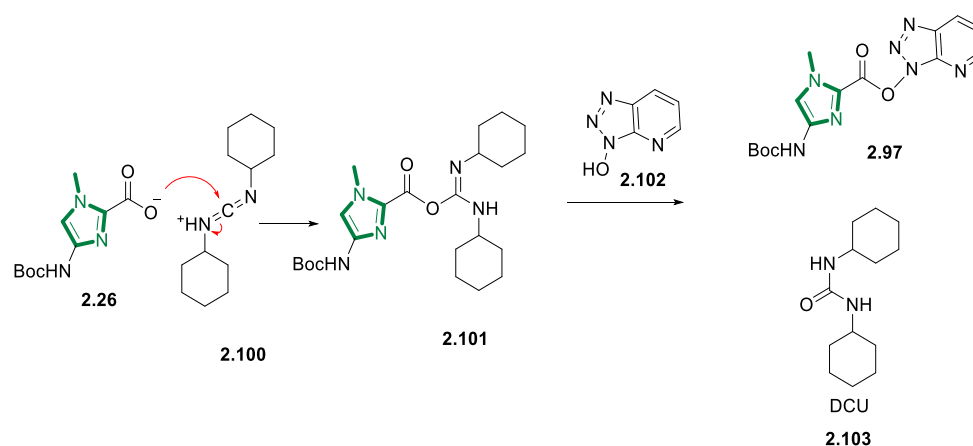


Activated acid	Amine on support	Reagents used	Active species	Reaction time
Boc-Py-OH	NH ₂ -Py	HATU	2.94	2h
Boc-β-Ala-OH	NH ₂ -Py	HATU	2.95	2h
Boc-γ-turn-OH	NH ₂ -Py	HATU	2.96	2h
Boc-Im-OH	NH ₂ -Py	DCC/HOAt	2.97	6h
Boc-Im-OH	NH ₂ -Im	DCC/HOAt	2.97	6h
Boc-β-Ala-OH	NH ₂ -Im	DCC/HOAt	2.95	12h
Boc-γ-Turn	NH ₂ -Im	DCC/HOAt	2.96	12h
Boc-Py-OH	NH ₂ -Im	BTC	2.99	3h
Im-COOH	NH ₂ -Py	DCC/HOAt	2.98	6h

The uronium salt HATU was used for couplings involving only Boc-Py-OH, Boc-β-Ala-OH and Boc-γ-turn-OH building blocks. An excess of carboxylic acid compared to the activating agents was normally used to prevent possible capping of the immobilised amine on the support. Owing to the brief activation time (20 minutes) and high coupling efficiency, HATU represented the first choice among activating agents. Coupling times as low as 15 minutes

have been previously reported.¹⁸⁸ However, large variability was often observed, in particular for couplings in the second half of 8-ring polyamides. For this reason, a coupling time of 2 hours was normally used to guarantee completion of the condensation.

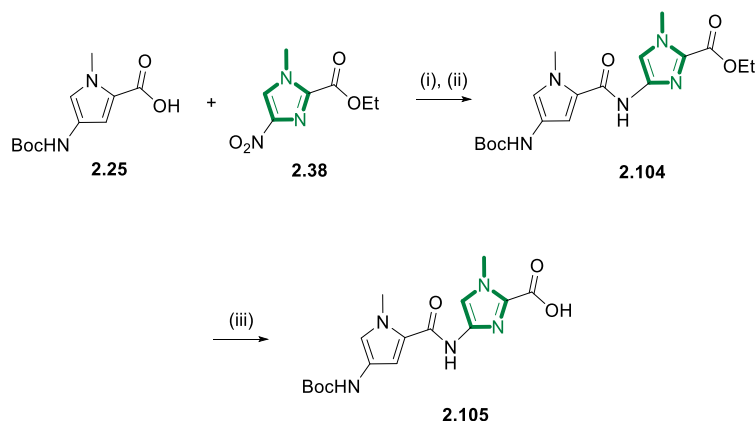
In contrast to Boc-Py-OH and Boc- γ -turn-OH building blocks, HATU-mediated activation has been reported to be unsuccessful for Boc-Im-OH, which is normally activated using a combination of DCC and HOAt.¹⁸⁸ This mixture has been proven to be more effective in the generation of the azabenzotriazole esters **2.97** and **2.98**, although, longer activation times are required (2 hours). Formation of the activated species is followed by precipitation of the side product dicyclohexylurea (DCU, **2.103**), which is removed by filtration prior to coupling (Scheme 2.13). More challenging were the couplings involving the deactivated amine of Boc-Im-OH. In fact, condensations between an immobilised Im-NH₂ and Boc- γ -turn-OH were normally achieved using DCC/HOAt for 12 hours to ensure completion of the coupling. The most challenging was the condensation between an immobilised Im-NH₂ and Boc-Py-OH (**2.25**), which was incomplete even after double coupling using DCC/HOAt for 12 hours. For this reason, the dimeric building block **2.105** is generally incorporated onto the solid support using standard DCC/HOAt conditions (Scheme 2.14).¹⁷⁸



Scheme 2.13 Mechanism of Boc-Im-OH activation by DCC/HOAt.

Alternatively, couplings between an immobilised Im-NH₂ and Boc-Py-OH were achieved in good yield using triphosgene as the activating agent.^{174,190,191} In the putative mechanism, the

carboxylic acid is activated to the acyl chloride intermediate **2.99**, which more rapidly reacts with the immobilised Im-NH₂. As previously demonstrated, triphosgene can efficiently be used for all the condensations of the polyamide, increasing the overall yield and purity of the synthesis.¹⁹¹ However, because of the need for dry solvent, as well as some variability observed between different batches of resin, triphosgene-mediated couplings were limited to Im-NH₂ to Boc-Py-OH condensations.



Scheme 2.14 Synthetic approach designed by Baird *et al.*¹⁷⁸ *Reagents and conditions:* (i) **2.38**, Pd/C 10% (cat.), H₂ atm., rt, 2 h. (ii) **2.25**, DCC/HOBt, DIPEA, 60 °C, 8 h. (iii) 1 M NaOH (aq.), 60 °C, 6 h.

2.5.2.2 Cleavage of the resin and purification of polyamides

The progress of the polyamide synthesis was monitored after difficult couplings and/or at key points of the polyamide scaffold. The inherently low reactivity of the immobilised aromatic amines Py-NH₂ and Im-NH₂ used for the synthesis of polyamides prevents utilisation of the ninhydrin test, which is normally used for peptide sequencing. For this reason, cleavage tests were normally used to check if couplings were successful. In this procedure, 1-2 mg of the resin were cleaved using the amine Dp (**2.30**) at 90 °C for 1 hour. After the resin was removed by filtration, the crude material was analysed by LC-MS. Upon completion of the polyamides, the entire batch of the resin was cleaved with the desired cleaving amine for 16 hours at 70 °C. The resin was removed by filtration and rinsed with DCM. The polyamides were collected following precipitation with cold diethyl ether and centrifugation, then purified by RP-HPLC.

Alternatively, the filtered resin was washed with methanol, and the crude mixture was directly used for HPLC purification.

2.5.3 Preparation of polyamides incorporating Nt at the N-terminus (position 8)

The polyamide scaffold targeting the sequence 5'WGWWCW (W= A/T) found in the ARE sequence was first considered in this study. Incorporation of Nt building blocks in the terminal position of 8-ring hairpin polyamides was first evaluated (Figure 2.8). If the hypothesis that Nt building blocks retain the binding selectivity for G nucleotides was proven correct, then these polyamides would target 5'WGWWCW found in the ARE dsDNA sequence. Monomers **2.87a-c** and dimers **2.88a-c** were prepared in order to evaluate different coupling conditions (Figure 2.8). While direct coupling of monomers would not require any further solution chemistry, the use of dimers would reduce the number of coupling steps required on the resin, which could improve final compound purity.^{180,184}

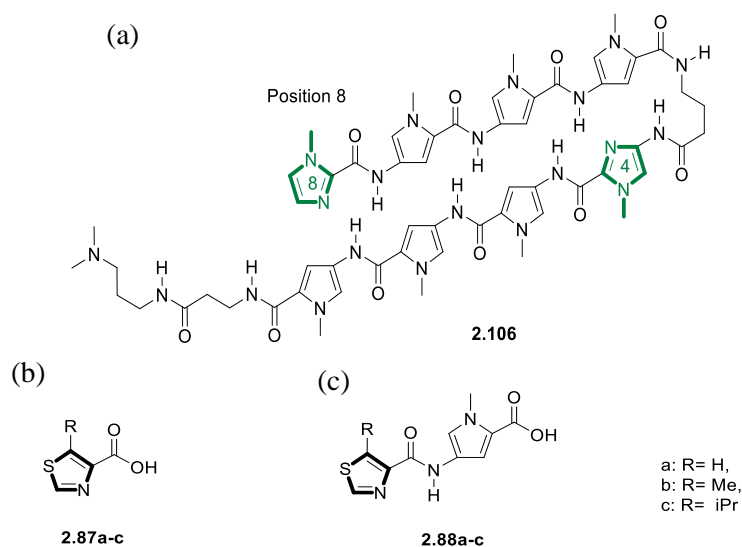
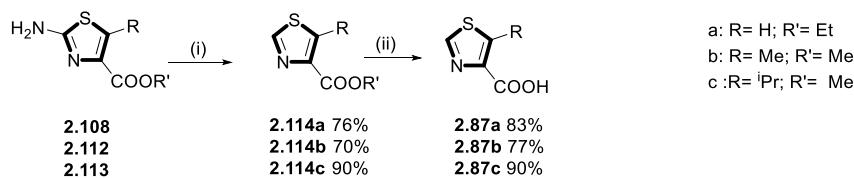


Figure 2.8 Structures of (a) the 8-ring hairpin polyamide targeting the ARE. (b) Nt terminal monomers and (c) dimers.

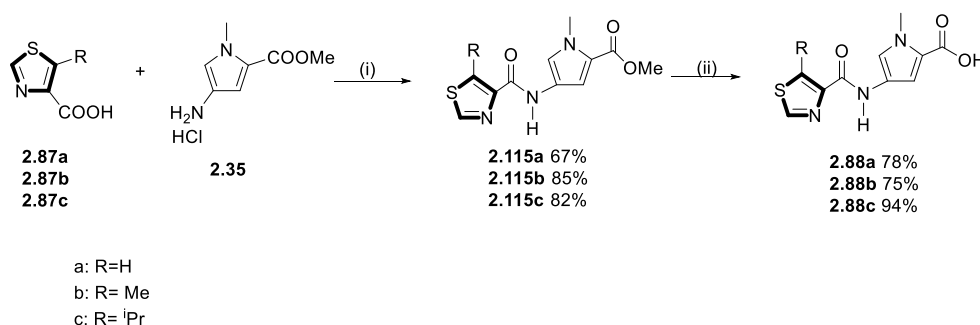
nitrite in the presence of hypophosphorous acid, followed by ester hydrolysis using NaOH 1M at room temperature, provided the desired monomers **2.87a-c** (Scheme 2.17).¹⁹³



Scheme 2.17 Reagents and conditions: (i) NaNO₂ (2 eq.), H₃PO₂ 50% in water (24 eq.), 0 °C/rt, 1 h/2 h. (ii) 1 M NaOH (aq.), rt, 2 h.

2.5.3.2 Preparation of terminal Nt dimers

Preparation of dimers **2.88a-c** was achieved by coupling **2.87a-c** with **2.35** (Scheme 2.18). The resulting esters **2.115a-c** were hydrolysed to provide the desired acids **2.88a-c** in 75-94 % yields.

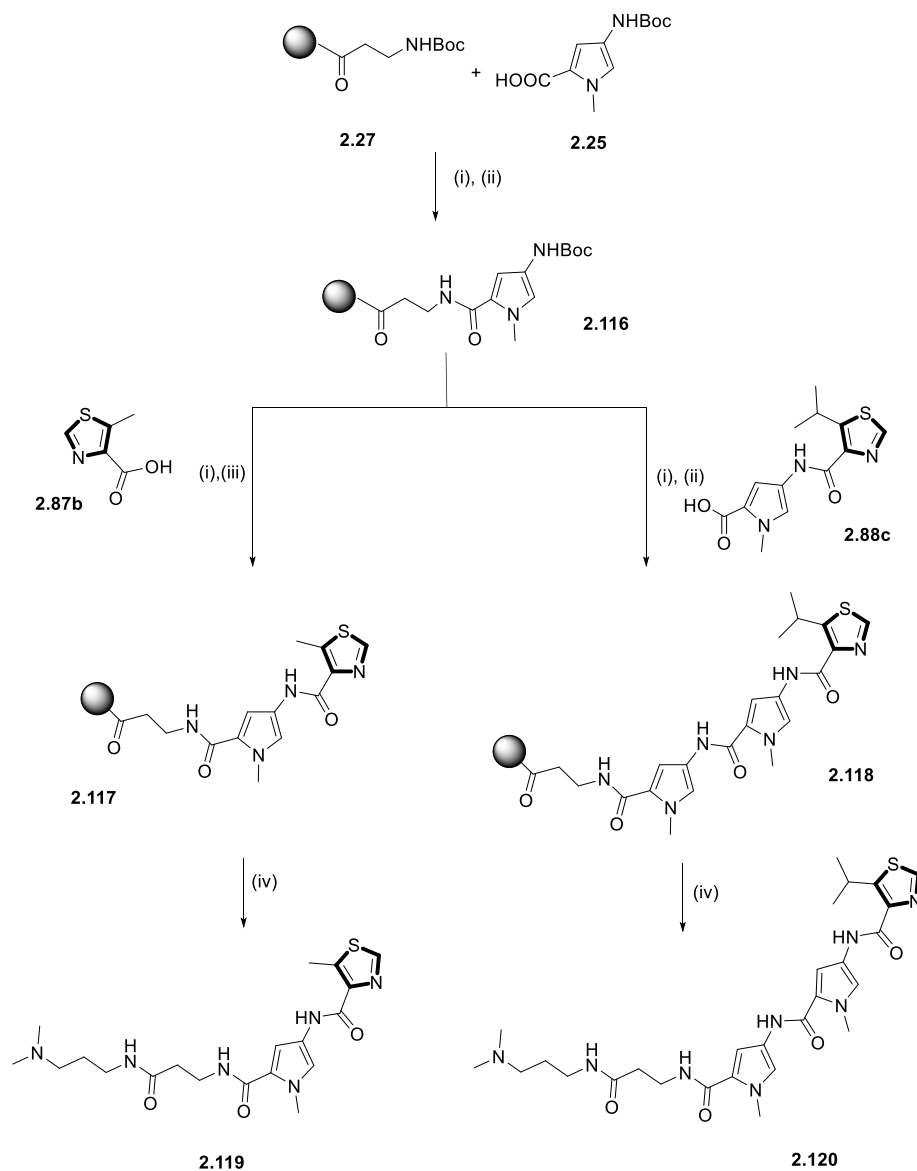


Scheme 2.18 Reagents and conditions: (i) HATU (1 eq.), DIPEA (9 eq.), rt, 12 h. (ii) 1 M KOH (aq.), rt, 12 h.

2.5.3.3 Preparation of Nt polyamides by solid phase synthesis

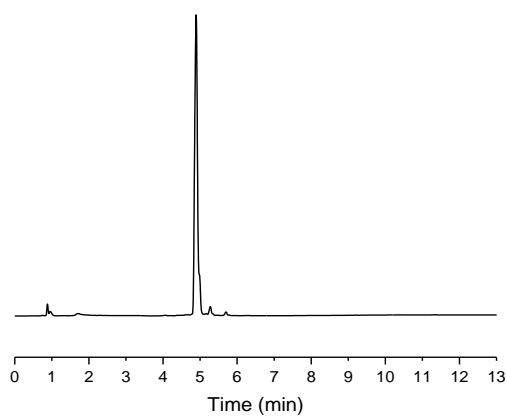
Before attempting the solid phase synthesis of 8-ring polyamides, shorter polymers **2.119-120** were prepared to optimise the SPS conditions (Scheme 2.19). Coupling of the Boc-Py-OH monomer (**2.25**) to Boc-β-Ala-PAM solid support using HATU afforded **2.116**, which was then coupled with monomer **2.87b** or dimer **2.88c** to yield **2.117** and **2.118** in good purity, as suggested by LC-MS analyses following cleavage with **2.30** (Figure 2.9). A 6 hour

HOAt/DCC coupling time was required to form **2.117**, while a standard 2 hour HATU coupling was used to obtain **2.118**.



Scheme 2.19 Reagents and conditions: (i) TPW (1× 1 min, 2× 3 min). (ii) **2.25**, HATU (3.8 eq.), DIPEA (9 eq.), 2 h. (iii) HOAt (3.8 eq.), DCC (3.6 eq.), DIPEA (9 eq.), 6 h. (iv) **2.30**, 70 °C, 12 h.

(a)



(b)

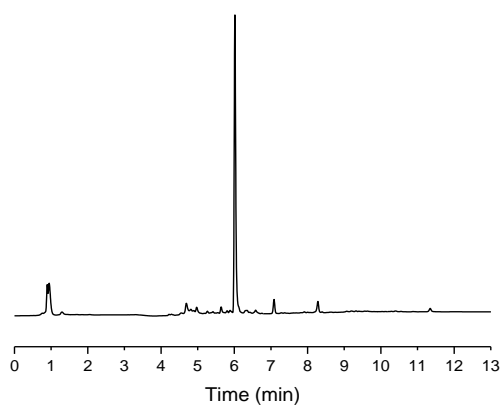
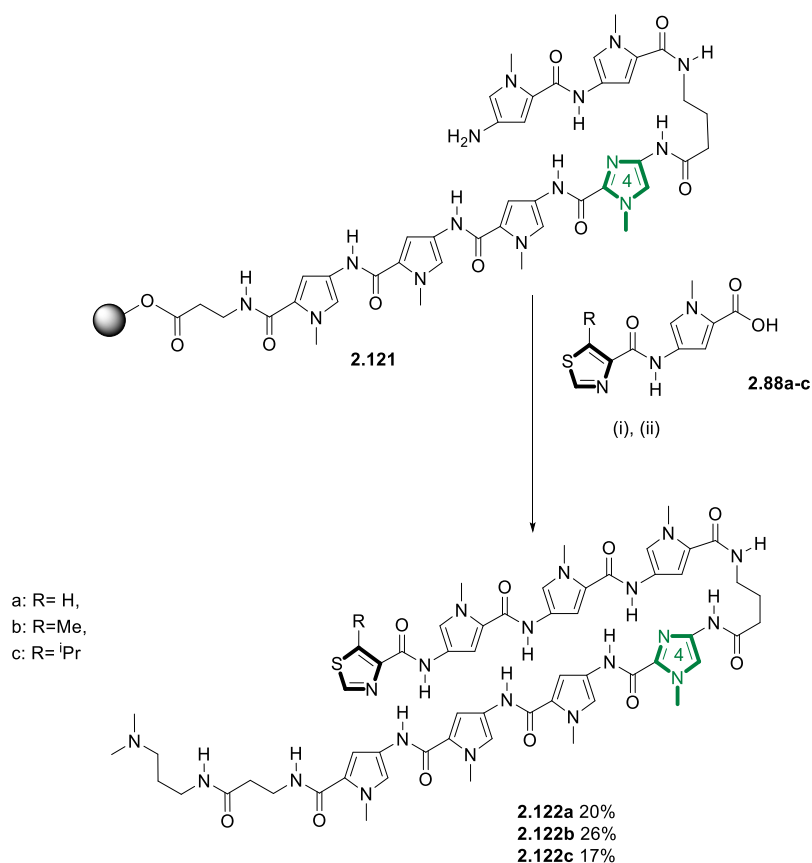


Figure 2.9 LC-MS chromatogram of crude (a) **2.119** and (b) **2.120**.

Due to the shorter activation time required to couple **2.88c** to immobilised, deprotected **2.116**, the use of dimers was preferred for the synthesis of 8-ring hairpin polyamides (**2.122a-c**) (Scheme 2.20). The hexamer (**2.121**) was prepared using standard solid phase synthesis conditions and coupled with dimers (**2.88a-c**) to provide the desired compounds in 17-26% overall yield after RP-HPLC purification.



Scheme 2.20 Reagents and conditions: (i) HATU (3.8 eq.), DIPEA (9 eq.), rt, 2 h. (ii) **2.30**, 70 °C, 12 h.

2.5.4 Attempts to prepare polyamides containing Nt at an internal position (position 4)

Position 4 of **2.106** was investigated for the incorporation of Nt building blocks on the 8-ring hairpin polyamide scaffold targeting the ARE sequence 5'WGWWCW (W= A/T). As for position 8, monomer **2.89** and dimer **2.90** building blocks were synthesised to screen for the optimal SPS conditions (Figure 2.10). In this case, both solution and solid phase conditions were problematic and prevented the preparation of polyamides incorporating Nt in position 4. For these reasons, the synthetic investigation was limited to installing ⁱPrNt building blocks (**2.89-2.90**).

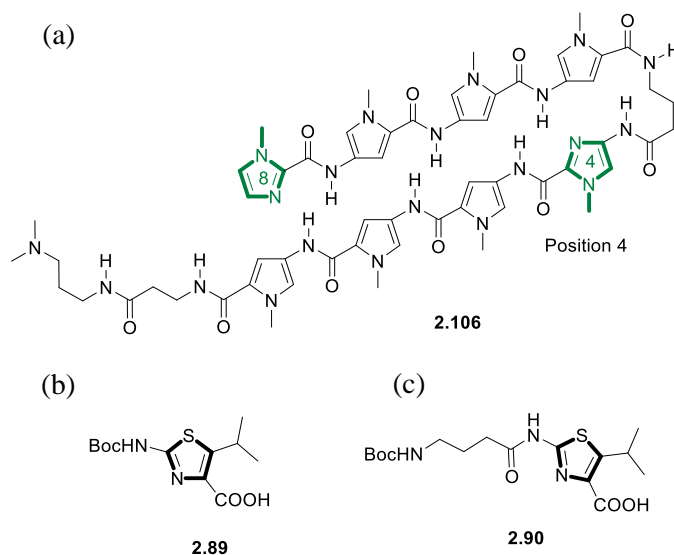
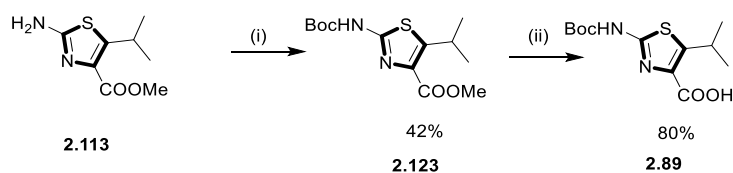


Figure 2.10 Structures of (a) the 8-ring hairpin polyamide targeting the ARE sequence. (b) Nt monomer and (c) dimer for incorporation in position 4 of the polyamide scaffold.

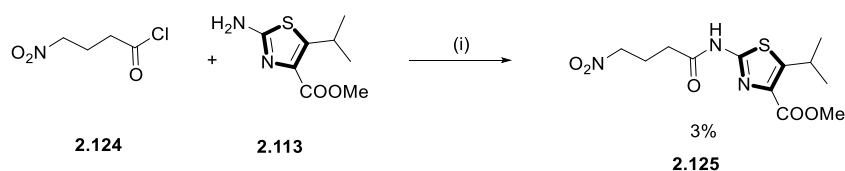
Incorporation of ^{15}N building blocks in position 4 of the polyamide scaffold through solid phase required the protection of the amine moiety of the Nt ring to afford the Boc protected **2.123**, followed by ester hydrolysis to obtain the monomer **2.89** as outlined in Scheme 2.21.



Scheme 2.21 Reagents and conditions: (i) Boc₂O (2 eq.), DIPEA (2 eq.) 60 °C, 3 d. (ii) 1M NaOH (aq.), 70 °C, 4 h.

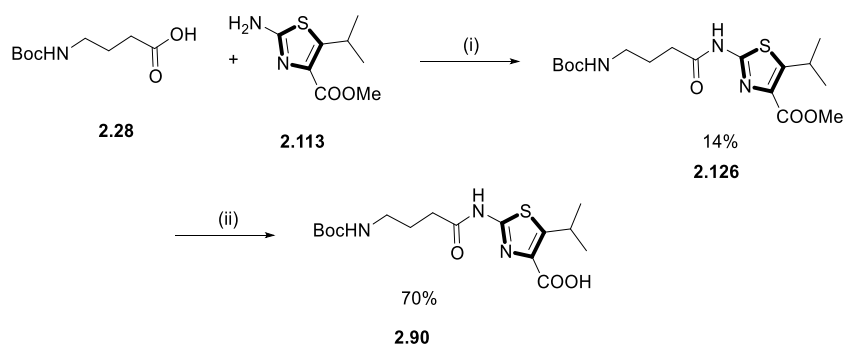
Low reactivity of the amine of **2.113** had been previously reported for 2-amino-4-carboxylate thiazole by Marques *et al.*¹⁴⁷ For this reason, Boc protection of **2.113** was achieved in 42% yield using an excess of di-*tert*-butyl dicarbonate, at 60 °C in DMF over 3 days. Furthermore, silica gel chromatography was necessary to separate **2.123** from the unreacted free amine **2.113**. Saponification using NaOH aqueous solution (1 M) provided the desired monomer **2.89** in 80% yield.

Synthesis of dimer **2.90** was initially attempted by coupling **2.124** with **2.113** (Scheme 2.22). The resulting nitro dimer **2.125** was obtained in 3% yield after silica gel chromatography. With three steps remaining to synthesise the desired building block, this method was deemed unsuitable for preparation of building blocks for solid phase synthesis.



Scheme 2.22 Reagents and conditions: (i) TEA (3 eq.), rt, 12 h.

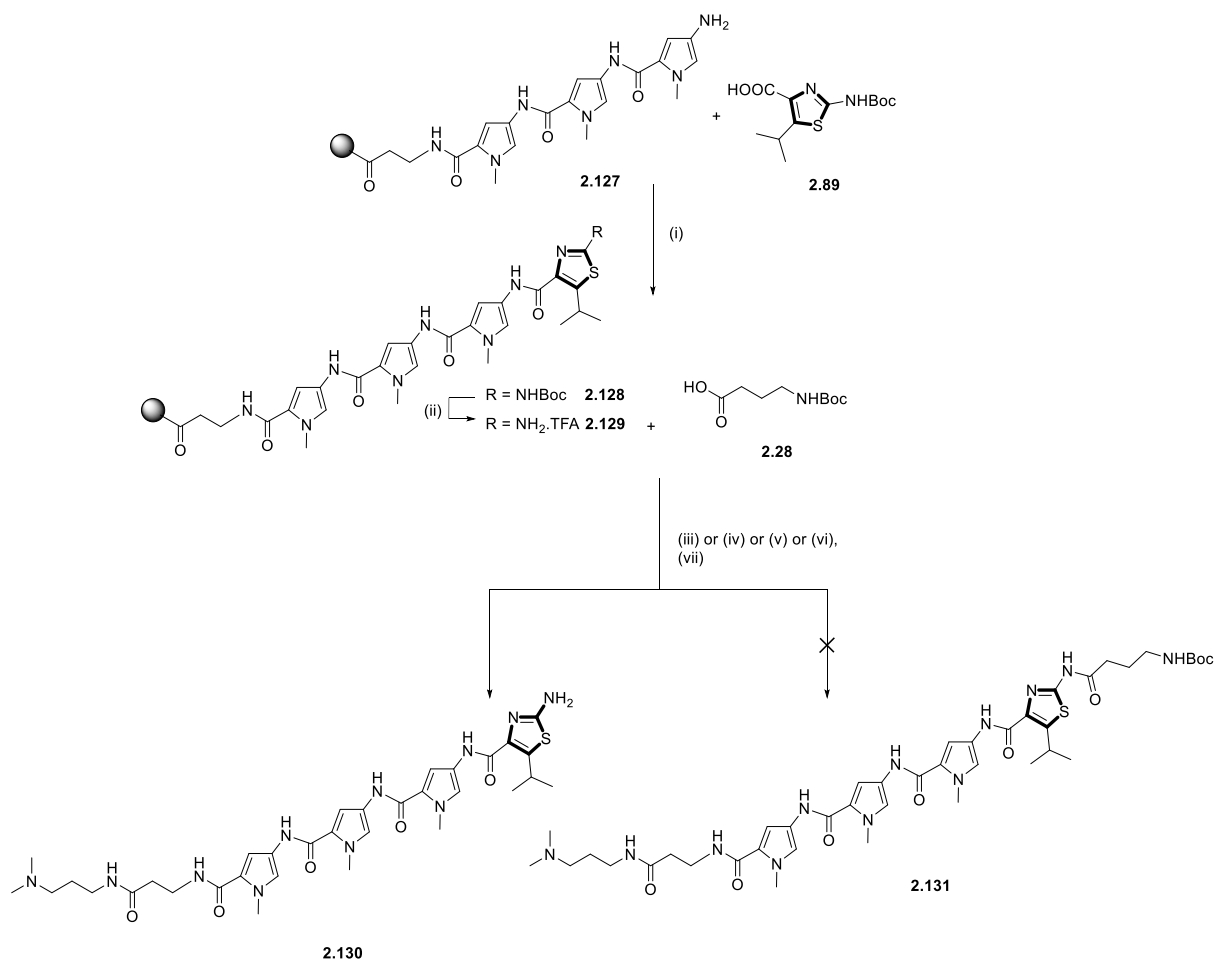
Improvement of the yield was achieved using previously reported conditions for Nt inactivated amines. Direct coupling of the Boc- γ -turn-OH (**2.28**) to **2.113**, using an excess of EDC for 8 days, afforded **2.126** in 14% yield, which was isolated through column chromatography (Scheme 2.23). Ester hydrolysis provided **2.90** in 70% yield.¹⁹⁴



Scheme 2.23 Reagents and conditions: (i) EDC (3 eq.), DMAP (3 eq.), rt, 8 d. (ii) 1M NaOH (aq.), rt, 12 h.

As a first attempt, monomer **2.89** was used to incorporate Nt in position 4 of the polyamide scaffold via solid phase synthesis (Scheme 2.24). The Boc protected monomer **2.89** was condensed with trimer **2.127**. Incorporation of **2.89** was achieved after 12 hours of HATU-mediated coupling. The Boc group of **2.128** was removed using TPW (1 \times 1 minute, 2 \times 3

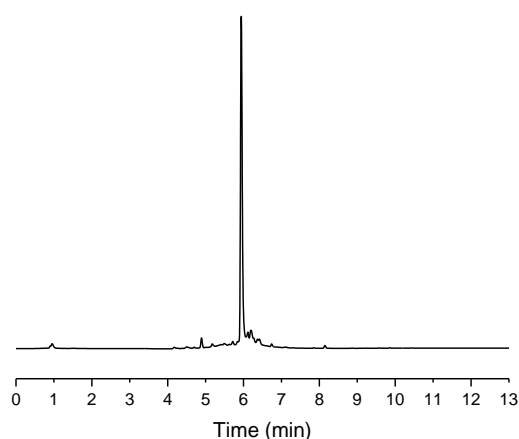
minutes) to obtain **2.129**. Coupling of Boc- γ -turn-OH (**2.28**) to **2.129** did not produce the desired **2.131**. For this reason, harsher conditions for the coupling were explored (Table 2.2). Unfortunately, none of these conditions resulted in the formation of **2.131**, and a singular peak corresponding to the amine **2.130** was observed on LC-MS chromatogram in all cases (Figure 2.11). It was therefore concluded that internal monomer **2.89** was not suitable for solid phase synthesis.



Scheme 2.24 Reagents and conditions: (i) HATU (3.8 eq.), DIPEA (9 eq.), 12 h. (ii) TPW (1 \times 1 min, 2 \times 3 min). (iii-vi) table below. (vii) **2.30**, 75 °C, 1 h.

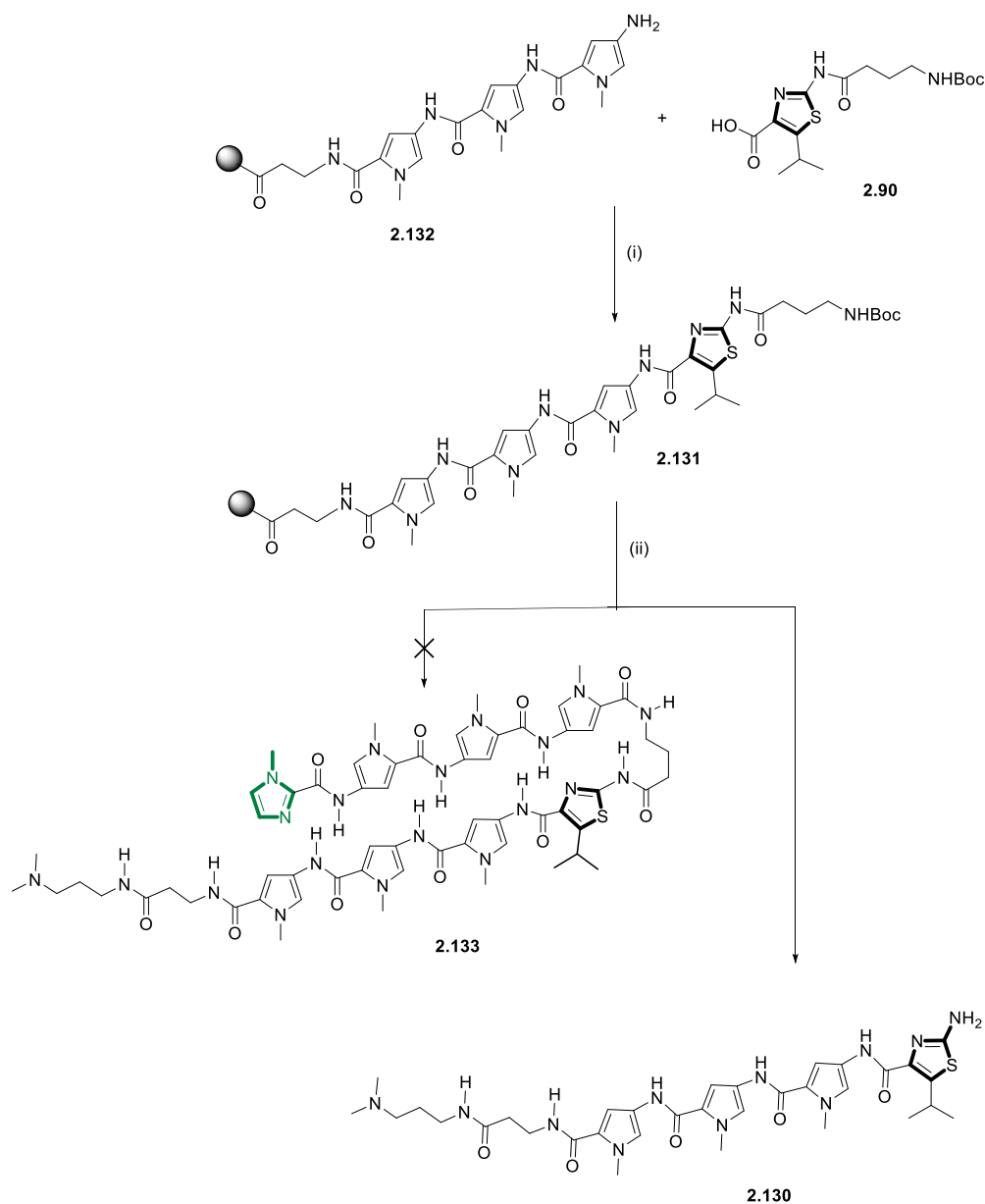
Table 2.2 Coupling conditions attempted for the synthesis of **2.131**.

Conditions	Activated acid	Amine on support	Reagents used	Temperature	Reaction time	Yield
iii	Boc- γ -turn-OH	NH ₂ -Nt	HATU	40°C	72h	0%
iv	Boc- γ -turn-OH	NH ₂ -Nt	DCC/HOAt	40°C	72h	0%
v	Boc- γ -turn-OH	NH ₂ -Nt	BTC	40°C	72h	0%
vi	Boc- γ -turn-OH	NH ₂ -Nt	EDC/DMAP	40°C	72h	0%

**Figure 2.11** LC-MS chromatogram corresponding to **2.130** observed after coupling conditions reported in Table 2.2.

Dimer **2.90** was used in an attempt to circumvent this difficult coupling. As for monomer **2.89**, incorporation of the dimer to **2.90** was achieved using HATU as the condensing agent (Scheme 2.25). However, the LC-MS chromatogram obtained following a cleavage test highlighted the presence of various impurities that could not be attributed to any component of the coupling (Figure 2.12a). In order to synthesise polyamide **2.133**, elongation of monomers was continued to completion without further cleavage tests. Upon cleavage from the resin, the final crude mixture was analysed using LC-MS. Surprisingly, mass spectrometric analysis suggested the presence of amine **2.130**, instead of the complete polyamide **2.133**. Moreover, the crude material from this attempt displayed the same retention time found for **2.130** in the attempt to

couple the monomer **2.89**. This indicates a high sensitivity of the amide bond between Nt and the γ -turn, either to the strongly acidic conditions used for Boc deprotection or the long exposure to basic conditions, normally used at the end of the synthesis to cleave the polyamide from the support. Therefore, the synthesis of polyamides incorporating Nt in position 4 was abandoned.



Scheme 2.25 (i) HATU (3.8 eq.), DIPEA (9 eq.), 12 h. (ii) Standard deprotection/coupling procedure to complete **2.133**.

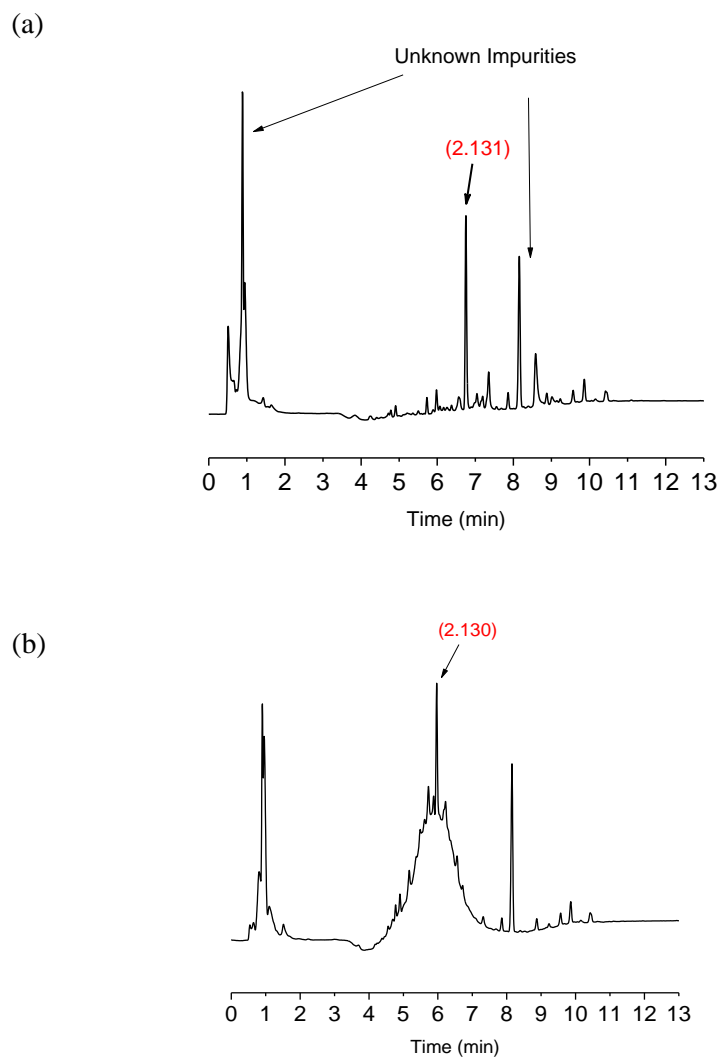


Figure 2.12 (a) LC-MS chromatogram observed after incorporation of **2.90**. (b) LC-MS chromatogram observed at the end of the synthesis of **2.133**.

2.5.5 Preparation of polyamides incorporating Nt at an internal position (position

6)

The synthetic study was completed by investigating position 6 of 8-ring hairpin polyamides. The archetypal polyamide containing Im in this position selects the DNA sequence 5'WWGWGC according to the established binding rules.³ Although this sequence deviates from the ARE target sequence (5'WGWWCW) of interest, incorporation of Nt building blocks at this position would provide insight on the impact of dsDNA binding affinity and selectivity

when Nt is incorporated in an internal position. Due to the low reactivity of Nt amines observed during the investigation of position 4, dimers **2.91a-c** were directly used for solid phase synthesis.

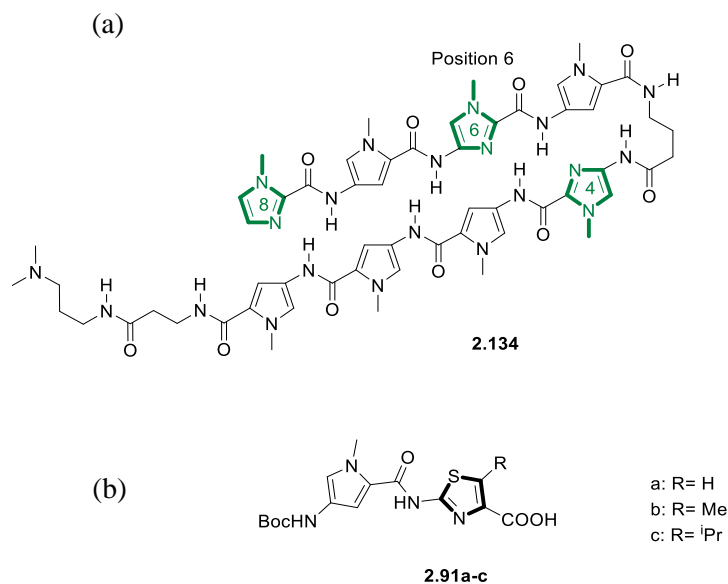
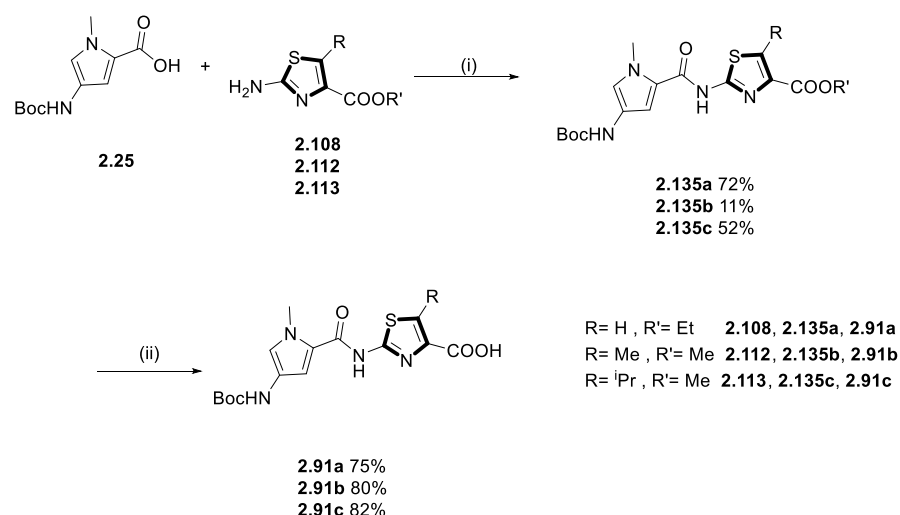


Figure 2.13 Structures of (a) the 8-ring hairpin polyamide targeting the DNA sequence 5'WGWCW. (b) Nt dimers used for incorporation within position 6 of the polyamide scaffold.

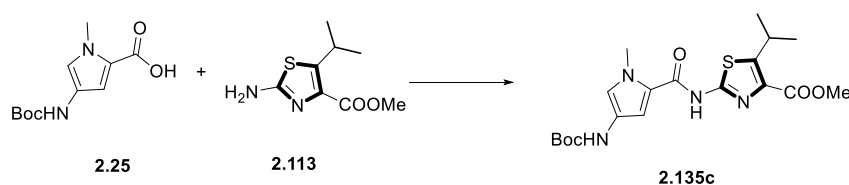
2.5.5.1 Preparation of Nt-Py dimers

The low reactivity of Nt amines was again confirmed for coupling with aromatic carboxylic acids. A set of standard conditions was unsuccessfully attempted for the formation of the amide bonds (Table 2.3). The desired compounds **2.91a-c** were isolated using conditions previously reported for unreactive amines (Scheme 2.26).¹⁹⁴ The unreacted amine and carboxylic acid were removed through silica gel chromatography to obtain dimer **2.135a-c** in adequate purity to proceed with the synthesis. After ester hydrolysis, **2.91a-c** were directly used for the preparation of polyamides by solid phase synthesis.



Scheme 2.26 Reagents and conditions: (i) EDC (4 eq.), DMAP (3 eq.), rt, 5-10 d. (ii) 1 M NaOH (aq.), rt, 12 h.

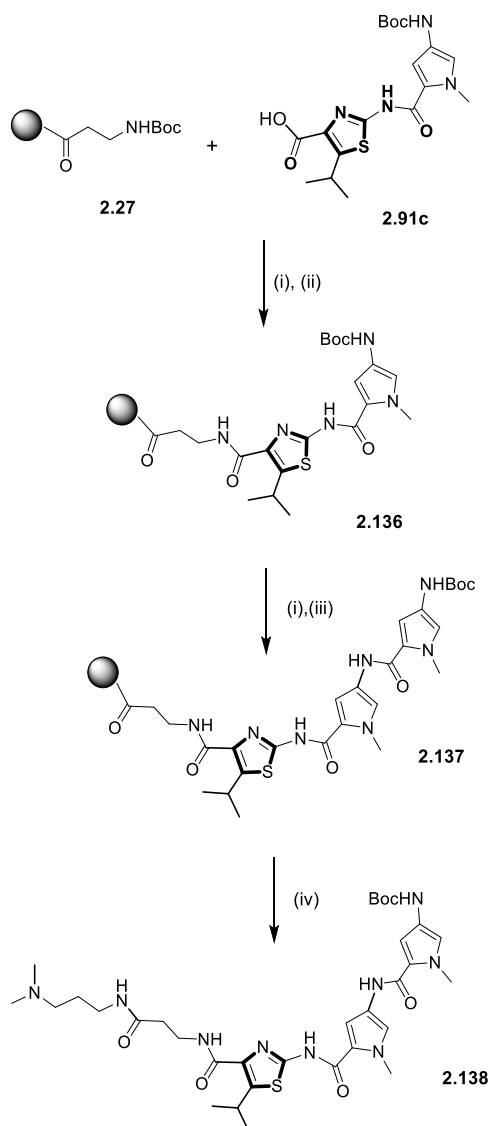
Table 2.3 Coupling conditions attempted for the synthesis of **2.135c**.



Entry	Method	Conditions	Product (%)
1	CDI, TEA	60 °C, 3 d	traces
2	POCl ₃ , Pyridine	rt, 3 d	traces
3	Triphosgene, collidine, DIPEA	60 °C, 16 h	0
4	HOAt/DCC, DIPEA	rt, 3 d	traces
5	DPPA, TEA	16 h	0
6	HATU, DIPEA	12 h	0
7	EDC, DMAP	5-10 d	11-72%

2.5.5.2 Preparation of Nt polyamides by solid phase synthesis

A preliminary test was carried out to investigate the stability of the dimer to solid phase synthesis conditions (Scheme 2.27). For this reason, dimer **2.91c** was incorporated onto the solid support using standard HATU conditions. Boc-Py-OH monomer **2.25** was coupled to **2.136** to obtain **2.138** in good purity as indicated by LC-MS after cleavage (Figure 2.14). This suggested the greater stability of the amide when located between two aromatic rings.



Scheme 2.27 Reagents and conditions: (i) TPW (1 × 1 min, 2 × 3 min). (ii) HATU (3.8 eq.), DIPEA (9 eq.), 12 h. (iii) **2.25**, HATU (3.8 eq.), DIPEA (9 eq.), 2 h. (iv) **2.30**, 70 °C, 12 h.

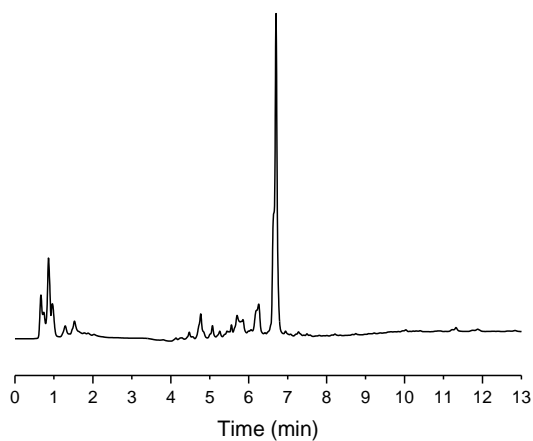
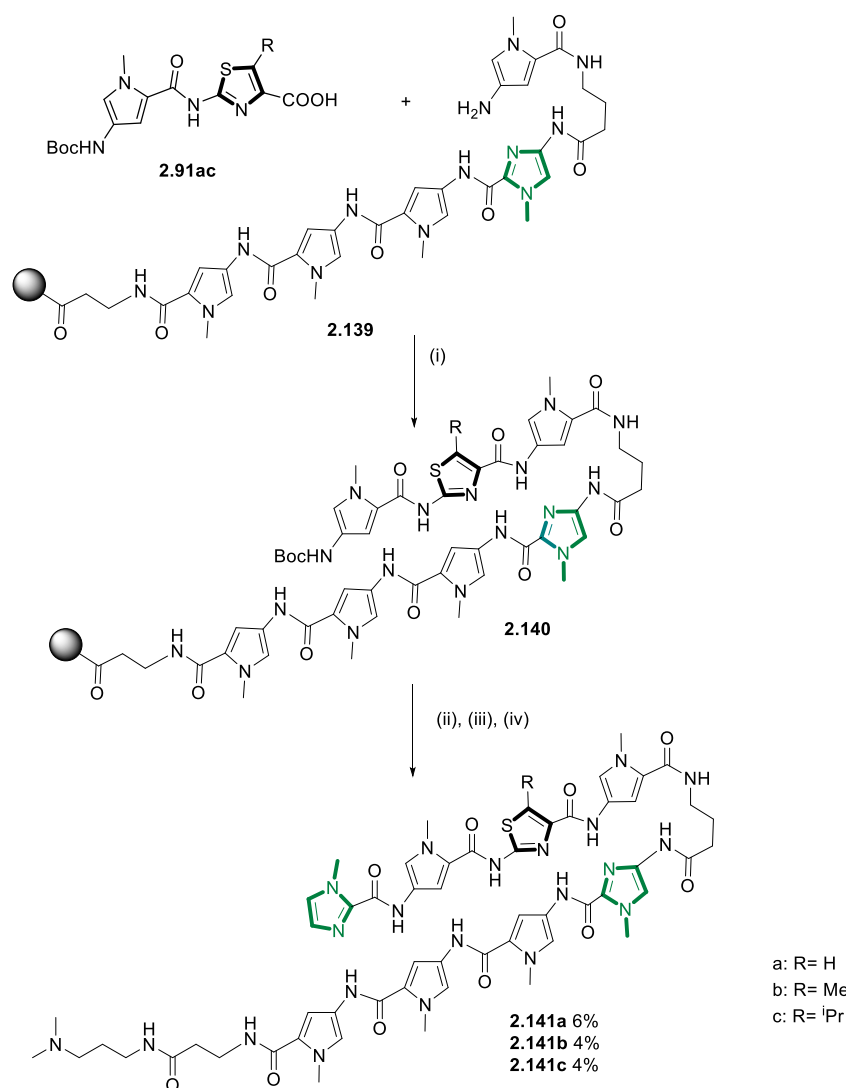


Figure 2.14 LC-MS chromatogram of crude **2.138**.

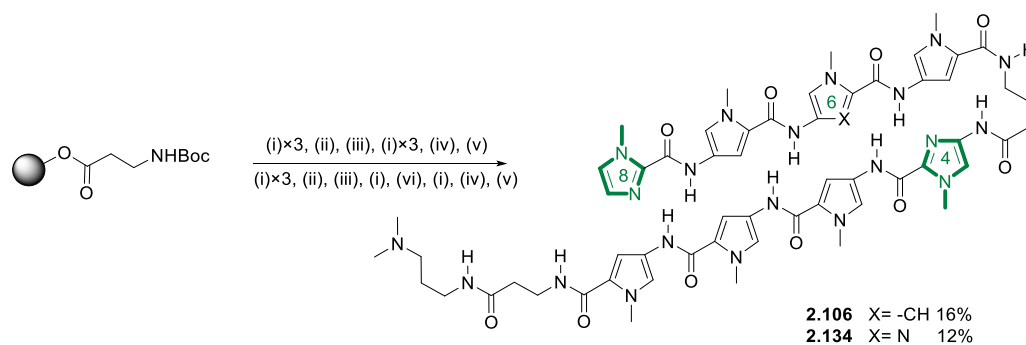
Compound **2.140** was prepared by incorporating dimers **2.91a-c** using solid phase synthesis (Scheme 2.28). After Boc deprotection and coupling of the terminal Im-OH unit **2.29** using DCC/HOAt, the resin was cleaved. Although lower crude purity was observed compared to polyamides incorporating Nt in the terminal position, polyamides **2.141a-c** were isolated by RP-HPLC for preliminary determination of DNA binding (Chapter 3).



Scheme 2.28 Reagents and conditions: (i) HATU (3.8 eq.), DIPEA (9 eq.), 12 h. (ii) TPW (1 \times 1min, 2 \times 3 min). (iii) **2.29**, HOAt (3.8 eq.), DCC (3.6 eq.), DIPEA (9 eq.), rt, 6 h. (iv) **2.30**, 70 $^{\circ}\text{C}$, 12 h.

2.5.6 Preparation of Im-containing polyamides

Polyamides **2.102** and **2.129** containing Im units in position 4, 6 and 8 were synthesised according to standard procedures (Scheme 2.29).



Scheme 2.29 *Reagents and conditions:* (i) TPW (1× 1 min, 2× 3 min); **2.26** HATU (3.8 eq.), DIPEA (9 eq.), 2 h. (ii) TPW (1× 1 min, 2× 3 min); **2.27**, HOAt (3.8 eq.), DCC (3.6 eq), DIPEA (9 eq.), 6 h. (iii) TPW (1× 1 min, 2× 10 min); **2.28**, HOAt (3.8 eq.), DCC (3.6 eq.), DIPEA (9 eq.), 6 h. (iv) TPW (1× 1 min, 2× 3 min); **2.29**, HOAt (3.8 eq.), DCC (3.6 eq), DIPEA (9 eq.), 6 h. (v) **2.30**, 70 °C, 16 h. (vi) TPW (1× 1 min, 2× 3 min); **2.26**, BTC (0.33 eq.), collidine (3 eq.), DIPEA (9 eq.), 3h.

2.6 Summary

This chapter described the preparation of 8-ring hairpin polyamides containing Nt building blocks in positions 6 and 8 (Figure 2.15). Fundamental to optimising polyamide synthesis was the exploration of solid phase conditions. Monomeric and dimeric Nt building blocks (**2.87-91**) modified at position 5 of the aromatic ring were prepared using solution phase synthesis. Procedures for solid phase incorporation of Nt were designed and employed for the preparation of 8-ring hairpin polyamides containing Nt in position 6 and 8. Attempts to prepare polyamides containing Nt in position 4 of the polyamide scaffold were unsuccessful owing to the low reactivity of the Nt amine combined with an enhanced sensitivity towards acidic and/or basic conditions of the amide bond between Nt and the γ -turn. Taken collectively, a suite of 8 polyamides was successfully prepared in suitable quantities for evaluation of the dsDNA binding behaviour.

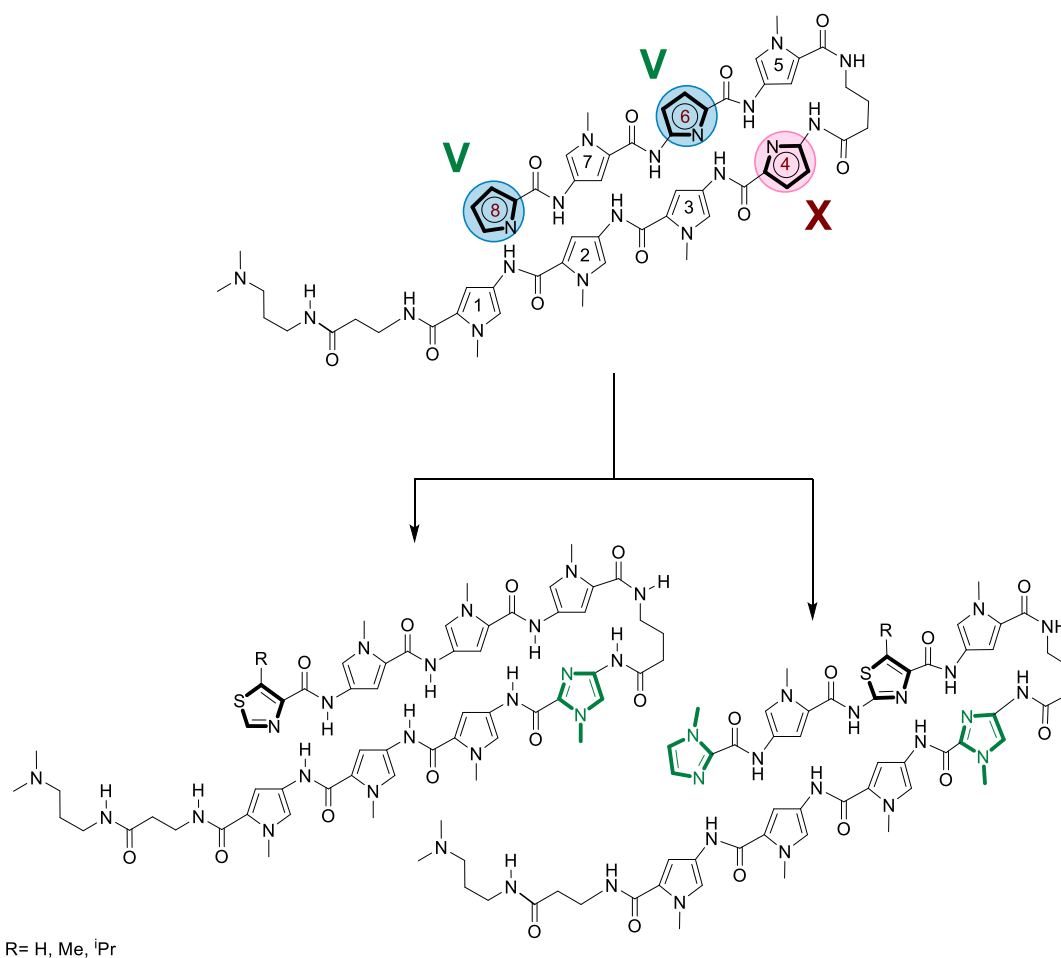


Figure 2.15 Schematic representation of the positions in the polyamide scaffold evaluated for Nt incorporation and structures of the synthesised Nt-containing polyamides (V and X refers to successful and unsuccessful modified positions, respectively).

2.7 Experimental

2.7.1 General procedure

All reagents and solvents were obtained from commercial sources and used without further purification unless otherwise stated.

2.7.2 NMR spectroscopy

NMR spectroscopy was carried out using a Bruker AV3 400 MHz and AV 500 MHz spectrometers. All chemical shifts (δ) were referenced to the deuterium lock and are reported

in parts per million (ppm). Coupling constants are quoted in hertz (Hz). Abbreviations for splitting patterns are s (singlet), d (doublet), t (triplet), q (quartet), sep (septet), br (broad) and m (multiplet). All NMR data were processed using iNMR.

2.7.3 HPLC experiments

Semi-preparative HPLC purification was carried out on a Dionex Ultimate 3000 series HPLC using a 150×21.2mm Kinetex 5 μ m C18 column. Analytical HPLC traces were recorded on a Shimadzu Prominence HPLC or Dionex Ultimate 3000 series HPLC using a 250×4.6 mm Aeris 3.6 μ m Widedpore XB-C18. In all cases, 0.1% v/v TFA in water/acetonitrile was used as buffer. Details of HPLC gradients are reported in the Appendix.

2.7.4 IR data and melting points

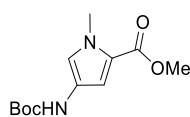
IR data were collected on an Agilent spectrometer and the data were processed using Spectrum One software. Only major absorbances are reported. Melting points of solids were obtained using a Griffin melting point apparatus.

2.7.5 Mass spectrometry

LC-MS was carried out on an Agilent HPLC in conjunction with an Agilent quadrupole mass spectrometer. The mass of the polyamides was measured on a MALDI-TOF Kratos Axima CFR using α -cyano-4-hydroxycinnamic acid matrix. HRMS spectra of small molecules were measured on a Bruker microTOFq High Resolution Mass Spectrometer.

2.7.6 Synthetic procedures for small molecules

*Methyl 4-((tert-butoxycarbonyl)amino)-1-methyl-1H-pyrrole-2-carboxylate (2.92)*¹⁹⁵



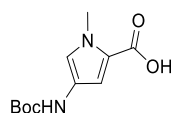
A solution of **2.34** (15 g, 0.08 mol), *diterbutyl* dicarbonate (21.32 g, 0.098 mol) and Pd/C 10% wt (catalytic amount) in EtOAc/MeOH (1:1, 300 mL) was shaken for 6 hours at room temperature in a Parr hydrogenator apparatus at 55 psi. The mixture was filtered through celite and the solution was concentrated *in vacuo*. The residue was then recrystallized from DCM/Hexane and dried to give **2.92** as a white powder (19.1 g, 94%).

$^1\text{H-NMR}$ (d_6 -DMSO, 500 MHz) δ : 1.44 (s, 9H, Boc- CH_3), 3.71 (s, 3H, O- CH_3), 3.79 (s, 3H, Py- CH_3), 6.62 (s, 1H, Py- H), 7.10 (s, 1H, Py- H), 9.10 (s, 1H, CONH).

$^{13}\text{C-NMR}$ (d_6 -DMSO, 125 MHz) δ : 28.1, 36.0, 50.8, 78.5, 107.3, 118.6, 119.2, 123.1, 152.7, 160.7.

LC-MS (+ve mode): 255.1 [M+H] $^+$

4-((*tert*-butoxycarbonyl)amino)-1-methyl-1H-pyrrole-2-carboxylic acid (**2.25**)¹⁹⁵

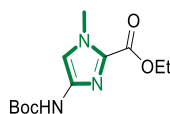


A solution of **2.92** (2.5 g, 0.01 mol) in a 1:1 solution of MeOH and 1 M LiOH (aq.) (22 mL) was heated to 70 °C for 4 hours. The residue was dried *in vacuo* and dissolved in water. The solution was acidified to pH 3 with 1 M HCl (aq.) to form a precipitate. The mixture was then extracted with EtOAc (5 \times 15 mL) and the organic phases were combined, dried over Na₂SO₄ and concentrated *in vacuo*. The residue was then recrystallized from DCM/hexanes to give the pure compound **2.25** as a white powder (1.87 g, 78%).

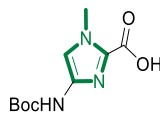
$^1\text{H-NMR}$ (d_6 -DMSO, 500 MHz) δ : 1.44 (s, 9H, Boc- CH_3), 3.77 (s, 3H, Py- CH_3), 6.57 (s, 1H, Py- H), 7.04 (s, 1H, Py- H), 9.04 (s, 1H, CONH), 12.07 (br, 1H, COOH).

$^{13}\text{C-NMR}$ (d_6 -DMSO, 101 MHz) δ : 28.2, 36.0, 78.4, 107.4, 118.7, 119.7, 122.8, 152.7, 161.9.

LC-MS (-ve mode): 239.0 [M-H] $^-$

*Ethyl 4-((tert-butoxycarbonyl)amino)-1-methyl-1H-imidazole-2-carboxylate (2.93)*¹⁹⁶

To a solution of **2.38** (1.90 g, 9.5 mmol) and di-*tert*-butyldicarbonate (2.50 g, 11.4 mmol) in MeOH (60 mL) was added Pd/C 10% wt (catalytic amount). The mixture was stirred at room temperature under a hydrogen atmosphere for 6 hours. The mixture was filtered through celite and the solvent removed *in vacuo* to yield the crude product as an orange solid. Pure product **2.93** was isolated after recrystallisation from DCM/hexane as a yellow powder (2.19 g, 85 %). ¹H-NMR (CDCl₃, 500 MHz) δ: 1.41 (t, 3H, *J* = 7 Hz, O-CH₂CH₃), 1.49 (s, 9H, Boc-CH₃), 3.96 (s, 3H, Im-CH₃), 4.38 (q, 2H, *J* = 7 Hz, O-CH₂CH₃), 6.89 (s, 1H, CONH), 7.21 (s, 1H, Im-H). ¹³C-NMR (CDCl₃, 125 MHz) δ: 14.3, 28.3, 35.9, 61.4, 80.8, 113.2, 131.4, 137.9, 152.7, 158.8. LC-MS (+ve mode): 214.2 [M-*tert*Butyl+H]⁺

*4-((tert-butoxycarbonyl)amino)-1-methyl-1H-imidazole-2-carboxylic acid (2.26)*¹⁹⁶

2.93 (2.80 g, 10.4 mmol) was dissolved in a solution of MeOH:1 M LiOH (aq.) (60 mL, 60 mmol). The solution was stirred for 2 hours at room temperature and reaction monitored by TLC. MeOH was removed under reduced pressure and the aqueous solution was acidified to pH 3 using 1 M HCl (aq.), causing the precipitation of a slushy white solid. The solid was washed by repeated additions of acidified water (pH 3), centrifugation and removal of the supernatant. The purified product **2.26** was then lyophilised, leaving an off-white fine solid (1.71 g, 69 %).

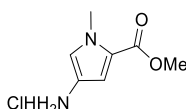
¹H-NMR (d₆-DMSO, 500 MHz) δ: 1.44 (s, 9H, Boc-CH₃), 3.87 (s, 3H, Im-CH₃), 7.25 (s, 1H,

Im-*H*), 9.59 (s, 1H, CONH).

^{13}C -NMR (d_6 -DMSO, 125 MHz) δ : 28.0, 35.4, 79.0, 113.4, 131.8, 137.6, 152.6, 159.8.

LC-MS (-ve mode): 240.2 [M-H] $^-$

*Methyl 4-amino-1-methyl-1H-pyrrole-2-carboxylate hydrochloride (2.39)*¹⁹⁷



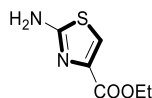
2.25 (1 g, 3.93 mmol) was dissolved in 4 M HCl (13 mL) in dioxane and the resulting mixture was stirred for 3 hours at room temperature. Hexanes (10 mL) was added to form a precipitate, that was collected and freeze-dried to give **2.39** as a pale yellow solid (0.741 g, 99%).

^1H -NMR (d_6 -DMSO, 500 MHz) δ : 3.75 (s, 3H, O- CH_3), 3.86 (s, 3H, Py- CH_3), 6.82 (s, 1H, Py-*H*), 7.26 (s, 1H, Py-*H*), 10.21 (s, 3H, $\text{NH}_2\cdot\text{HCl}$).

^{13}C -NMR (d_6 -DMSO, 125 MHz) δ : 36.5, 51.3, 111.5, 114.0, 120.8, 123.7, 160.2.

LC-MS (+ve mode): 155.0 [M+H] $^+$

*Ethyl 2-aminothiazole-4-carboxylate (2.108)*¹⁹⁸



Ethyl bromopyruvate (**2.107**) (2 g, 0.01 mol) was slowly added to a solution of thiourea (0.94 g, 0.012 mol, **2.81**) in dry MeOH (20 mL) and heated to reflux for 2 hours. The solvent was removed *in vacuo* and water (50 mL) was added to the crude residue. The pH was adjusted to 9 using solid K_2CO_3 , which resulted in the formation of a precipitate. The precipitate was collected, washed several times with water and dried to give **2.108** as pale-yellow powder (1.27 g, 74%).

^1H -NMR (CDCl_3 , 400 MHz) δ : 1.37 (t, 3H, $J = 7.1$ Hz, O- CH_2CH_3), 4.36 (q, 2H, $J = 7.1$ Hz, O- CH_2CH_3), 5.27 (br, 2H, Nt- NH_2), 7.44 (s, 1H, Nt-*H*).

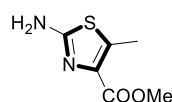
^{13}C -NMR (CDCl_3 , 101 MHz) δ : 14.5, 61.3, 117.8, 143.2, 161.5, 167.8.

LC-MS (+ve mode): 173.0 $[\text{M}+\text{H}]^+$

General procedure for the synthesis of 2.112-2.113

Sodium metal (1 eq.) was added in two portions to dry MeOH (50 mL) at 0 °C under a nitrogen atmosphere and with vigorous stirring. This solution was then added over one hour to a solution of **2.74** (1 eq.) and the respective aldehyde (**2.75**, **2.109**) (1.1 eq.) in dry Et₂O (50 mL) at 0 °C. The reaction mixture was stirred for one hour at 0 °C, after which Et₂O (150 mL) and brine (50 mL) were added. The organic phase was isolated, dried over Na₂SO₄ and concentrated *in vacuo* to give a yellow-orange oil. This oil was then added to a solution of thiourea (0.85 eq., **2.81**) in dry MeOH (60 mL). The solution was heated to reflux for 4 hours and the solvent was removed *in vacuo*. Water (100 mL) was added and the pH was adjusted to 9 using NH₄OH (18 M). The aqueous phase was then extracted several times with DCM (5 × 30 mL). The organic phases were combined, dried over Na₂SO₄, and concentrated *in vacuo* to afford the pure desired compounds.

Methyl 2-amino-5-methylthiazole-4-carboxylate (**2.112**)¹⁹³



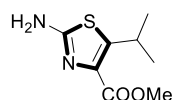
Yellow solid

Yield: 58%

^1H -NMR (CDCl_3 , 400 MHz) δ : 2.59 (s, 3H, Nt-CH₃), 3.85 (s, 3H, O-CH₃), 5.39 (br, 2H, Nt-NH₂).

^{13}C -NMR (CDCl_3 , 101 MHz) δ : 12.8, 51.9, 135.6, 136.4, 162.9, 163.4.

LC-MS (+ve mode): 173.0 $[\text{M}+\text{H}]^+$

*Methyl 2-amino-5-isopropylthiazole-4-carboxylate (2.113)*¹⁹³

Pale yellow solid

Yield: 72%

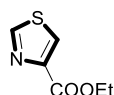
¹H-NMR (CDCl₃, 400 MHz) δ : 1.24 (d, 6H, $J = 6.8$ Hz, Nt-CH₃), 3.85 (s, 3H, O-CH₃), 4.02 (sep, 1H, $J = 6.8$ Hz, Nt-CH), 5.30 (br, 2H, Nt-NH₂).

¹³C-NMR (CDCl₃, 101 MHz) δ : 25.0, 27.7, 51.9, 134.7, 150.2, 162.8, 163.5.

ESI-MS (+ve mode): 223.1 [M+Na]⁺

General procedure for the synthesis of 2.114a-c

A solution of NaNO₂ (2 eq.) in water (6 mL) was slowly added to a solution of the corresponding aminothiazole (**2.108**, **2.112**, **2.113**) (1 eq.) in H₃PO₂ 50% water (24 eq.) at -5 °C. The resulting solution was stirred for 1 hour at 0 °C and for 2 hours at room temperature. The solution was cooled down to 0 °C and neutralized using 2 M NaOH (aq.). The reaction mixture was extracted with Et₂O (3 × 50 mL) and the organic layers were combined, dried over Na₂SO₄ and concentrated *in vacuo* to give an orange oil. The crude product was purified by silica gel chromatography (1:1 EtOAc/hexanes) to afford the desired product.

*Ethyl thiazole-4-carboxylate (2.114a)*¹⁹³

Pale yellow solid

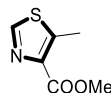
Yield: 76%

¹H-NMR (CDCl₃, 400 MHz) δ : 1.40 (t, 3H, $J = 7.1$ Hz, O-CH₂CH₃), 4.42 (q, 2H, $J = 7.1$ Hz, O-CH₂CH₃), 8.23 (d, 1H, $J = 2.1$ Hz, Nt-H), 8.84 (d, 1H, $J = 2.1$ Hz Nt-H).

^{13}C -NMR (CDCl_3 , 101 MHz) δ : 14.5, 61.7, 127.3, 153.5, 161.4 (one carbon not observed).

LC-MS (+ve mode): 158.1 $[\text{M}+\text{H}]^+$

*Methyl 5-methylthiazole-4-carboxylate (2.114b)*¹⁹³



Pale yellow solid

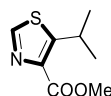
Yield: 70%

^1H -NMR (CDCl_3 , 400 MHz) δ : 2.79 (s, 3H, Nt- CH_3), 3.93 (s, 3H, O- CH_3), 8.56 (s, 1H, Nt- H).

^{13}C -NMR (CDCl_3 , 101 MHz) δ : 13.0, 52.7, 141.9, 145.0, 149.2, 162.9.

LC-MS (+ve mode): 158.1 $[\text{M}+\text{H}]^+$

*Methyl 5-isopropylthiazole-4-carboxylate (2.114c)*¹⁹³



Pale yellow solid

Yield: 90%

^1H -NMR (CDCl_3 , 400 MHz) δ : 1.32 (d, 6H, $J = 6.8$ Hz, Nt- CH_3), 3.93 (s, 3H, O- CH_3), 4.18 (sep, 1H, $J = 6.8$ Hz, Nt- CH), 8.60 (s, 1H, Nt- H).

^{13}C -NMR (CDCl_3 , 101 MHz) δ : 25.3, 27.8, 52.2, 140.1, 149.2, 159.2, 162.8.

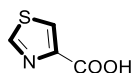
LC-MS (+ve mode): 186.0 $[\text{M}+\text{H}]^+$

General procedure for compounds 2.87a-c

1 M NaOH (aq.) (5-10 eq.) was added to a solution of the corresponding thiazole **2.114a-c** (1 eq.) in THF (20 mL). The resulting solution was stirred for 2 hours at room temperature. The solvent was removed *in vacuo* and the pH was adjusted to 2 through the addition of 1 M HCl

(aq.). The water phase was extracted with EtOAc (3× 30 mL) and the organic layers were combined, dried over Na₂SO₄ and concentrated *in vacuo* to give the desired compounds.

*Thiazole-4-carboxylic acid (2.87a)*¹⁹³



White solid

Yield: 83%

¹H-NMR (d₆-DMSO, 400 MHz) δ: 8.49 (d, 1H, *J* = 2.0 Hz, Nt-*H*), 9.14 (d, 1H, *J* = 2.0 Hz, Nt-*H*), 13.00 (br, 1H, COOH).

¹³C-NMR (d₆-DMSO, 101 MHz) δ: 128.6, 148.0, 155.0, 162.1.

LC-MS (-ve mode): 128.1 [M-H]⁻

*5-methylthiazole-4-carboxylic acid (2.87b)*¹⁹³



Pink solid

Yield: 77%

¹H-NMR (CDCl₃, 400 MHz) δ: 2.87 (s, 3H, Nt-*CH*₃), 8.61 (s, 1H, Nt-*H*).

¹³C-NMR (CDCl₃, 101 MHz) δ: 13.0, 141.1, 145.6, 149.3, 162.7.

LC-MS (-ve mode): 141.9 [M-H]⁻

*5-isopropylthiazole-4-carboxylic acid (2.87c)*¹⁹³



Yellow solid

Yield: 90%

$^1\text{H-NMR}$ (d_6 -DMSO, 400 MHz) δ : 1.26 (d, 6H, $J = 6.8$ Hz, Nt- CH_3), 4.06 (sep, 1H, $J = 6.8$ Hz, Nt- CH), 8.92 (s, 1H, Nt- H), 12.91 (broad, 1H, COOH).

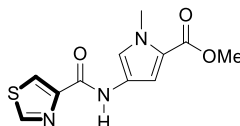
$^{13}\text{C-NMR}$ (d_6 -DMSO, 101 MHz) δ : 25.0, 27.1, 141.0, 150.3, 157.0, 163.3.

LC-MS (+ve mode): 172.1 $[\text{M}+\text{H}]^+$

General procedure for 2.115a-c

A solution of **2.87a-c** (1 eq.), HATU (1 eq.) and DIPEA (6 eq.) in dry DMF (10 mL) was stirred at room temperature for 30 minutes and then added to a solution of **2.35** (1 eq.) and DIPEA (3 eq.) in dry DMF (5 mL). The mixture was stirred overnight at room temperature and pH 3 water was added to form a precipitate that was collected after centrifugation and removal of the supernatant. The resulting solid was then lyophilised to give the desired compound.

Methyl 1-methyl-4-(thiazole-4-carboxamido)-1H-pyrrole-2-carboxylate (2.115a)



Pale yellow solid

Yield: 67%

m.p. = 125-130 °C

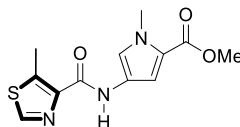
$^1\text{H-NMR}$ (CDCl_3 , 500 MHz) δ : 3.80 (s, 3H, O- CH_3), 3.92 (s, 3H, Py- CH_3), 6.83 (d, 1H, $J = 1.8$ Hz, Py- H), 7.52 (d, 1H, $J = 1.8$ Hz, Py- H), 8.21 (d, 1H, $J = 2$ Hz, Nt- H), 8.77 (d, 1H, $J = 2$ Hz, Nt- H), 9.08 (s, 1H, CONH).

$^{13}\text{C-NMR}$ (CDCl_3 , 101 MHz) δ : 37.0, 51.3, 108.3, 120.1, 121.1, 121.3, 123.5, 150.9, 152.9, 158.0, 161.6.

IR (neat) ν_{max} : 3353, 3116, 1690, 1653, 1500, 1256.

ESI (+ve mode) HRMS calculated for $[M+Na]^+$ 288.0415, found 288.0413.

Methyl 1-methyl-4-(5-methylthiazole-4-carboxamido)-1H-pyrrole-2-carboxylate (2.115b)



Pale yellow solid

Yield: 85%

m.p. = 127-132 °C

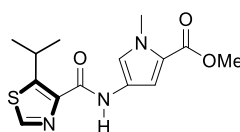
$^1\text{H-NMR}$ (CDCl_3 , 500 MHz) δ : 2.90 (s, 3H, Nt- CH_3), 3.81 (s, 3H, O- CH_3), 3.92 (s, 3H, Py- CH_3), 6.82 (d, 1H, $J = 2$ Hz, Py- H), 7.52 (d, 1H, $J = 2$ Hz, Py- H), 8.51 (s, 1H, Nt- H), 9.16 (s, 1H, CONH).

$^{13}\text{C-NMR}$ (CDCl_3 , 101 MHz) δ : 12.8, 36.9, 51.2, 108.3, 120.0, 120.9, 121.7, 141.3, 143.4, 148.4, 159.8, 161.7.

IR (neat) ν_{max} : 3347, 1698, 1658, 1558.

ESI (+ve mode) HRMS calculated for $[M+Na]^+$ 302.0570, found 302.0563.

Methyl 4-(5-isopropylthiazole-4-carboxamido)-1-methyl-1H-pyrrole-2-carboxylate (2.115c)



Pale yellow solid

Yield 82%

m.p.=100-102 °C

$^1\text{H-NMR}$ (CDCl_3 , 500 MHz) δ : 1.36 (d, 6H, $J = 7$ Hz, Nt- CH_3), 3.81 (s, 3H, O- CH_3), 3.91 (s, 3H, Py- CH_3), 4.49 (sep, 1H, $J = 7$ Hz, Nt- CH), 6.82 (d, 1H, $J = 1.7$ Hz, Py- H), 7.53 (d, 1H, $J = 1.7$ Hz, Py- H), 8.54 (s, 1H, Nt- H), 9.21 (s, 1H, CONH).

^{13}C -NMR (CDCl_3 , 101 MHz) δ : 25.5, 27.5, 36.9, 51.2, 108.3, 120.0, 120.9, 121.7, 142.0, 148.5, 155.4, 159.6, 161.7.

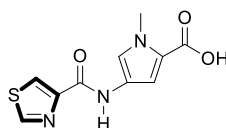
IR (neat) ν_{max} : 3352, 2956, 1693, 1656, 1545, 1253.

ESI (+ve mode) HRMS calculated for $[\text{M}+\text{H}]^+$ 330.0883, found 330.0869.

General procedure for the synthesis of 2.88a-c

1 M KOH (aq.) (5-10 eq.) was added to a solution of **2.115a-c** (1 eq.) in THF (10 mL) and stirred at room temperature for 4-12 hours. The THF was removed *in vacuo* and the aqueous phase was acidified to pH 2 with 1 M HCl (aq.) and extracted with EtOAc (3×10 mL). The organic layers were combined, washed with brine, dried over Na_2SO_4 and concentrated *in vacuo* to give the desired compound.

1-methyl-4-(thiazole-4-carboxamido)-1H-pyrrole-2-carboxylic acid (2.88a)



Brown solid

Yield 78%

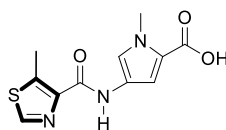
m.p. = 210-219 °C

^1H -NMR (d_6 -DMSO, 500 MHz) δ : 3.83 (s, 3H, Py- CH_3), 7.02 (d, 1H, $J = 1.8$ Hz, Py- H), 7.50 (d, 1H, $J = 1.8$ Hz, Py- H), 8.39 (d, 1H, $J = 2$ Hz, Nt- H), 9.23 (d, 1H, $J = 2$ Hz, Nt- H), 10.48 (s, 1H, CONH), 12.18 (br, 1H, COOH).

^{13}C NMR (d_6 -DMSO, 101 MHz) δ : 36.2, 109.1, 119.8, 120.3, 120.6, 122.0, 124.5, 150.7, 154.9, 157.7, 161.9.

IR (neat) ν_{max} : 3385, 3086, 2990 (br), 1683, 1627, 1584.

ESI (-ve mode) HRMS calculated for $[\text{M}+\text{Na}]^+$ 250.0292, found 250.0281.

1-methyl-4-(5-methylthiazole-4-carboxamido)-1H-pyrrole-2-carboxylic acid (2.88b)

Dark yellow solid

Yield 75%

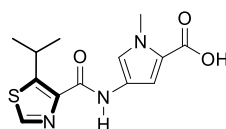
m.p. = 190-195 °C

$^1\text{H-NMR}$ (d_6 -DMSO, 400 MHz) δ : 2.80 (s, 3H, Nt- CH_3), 3.84 (s, 3H, Py- CH_3), 7.01 (d, 1H, J = 1.7 Hz, Py- H), 7.52 (d, 1H, J = 1.7 Hz, Py- H), 8.97 (s, 1H, Nt- H), 10.33 (s, 1H, CONH), 12.14 (br, 1H, COOH).

$^{13}\text{C-NMR}$ (d_6 -DMSO, 151 MHz) δ : 12.2, 36.2, 108.49, 119.6, 120.5, 122.1, 140.2, 143.5, 150.3, 159.3, 161.9.

IR (neat) ν_{max} : 3395, 3079, 2924 (br), 1660, 1653.

ESI (+ve mode) HRMS calculated for $[\text{M}+\text{H}]^+$ 288.0413 found 288.0412.

4-(5-isopropylthiazole-4-carboxamido)-1-methyl-1H-pyrrole-2-carboxylic acid (2.88c)

Dark yellow solid

Yield 94%

m.p.=167-170 °C

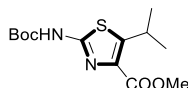
$^1\text{H-NMR}$ (d_6 -DMSO, 400 MHz) δ : 1.29 (d, 6H, J = 7 Hz, Nt- CH_3), 3.83 (s, 3H, Py- CH_3), 4.32 (sep, 1H, J = 7 Hz, Nt- CH), 7.00 (d, 1H, J = 1.7 Hz, Py- H), 7.53 (d, 1H, J = 1.7 Hz, Py- H), 9.00 (s, 1H, Nt- H), 10.32 (s, 1H, CONH), 12.14 (br, 1H, COOH).

^{13}C -NMR (d_6 -DMSO, 101 MHz) δ : 25.0, 26.6, 36.1, 108.9, 119.7, 120.6, 122.0, 142.3, 150.2, 153.7, 159.0, 161.9.

IR (neat) ν_{max} : 3399, 3077, 2958 (br), 1655, 1550.

ESI (+ve mode) HRMS calculated for $[\text{M}+\text{Na}]^+$ 316.0726, found 316.0722.

Methyl 2-((tert-butoxycarbonyl)amino)-5-isopropylthiazole-4-carboxylate (2.123)



Ditertbutyldicarbonate (12 mL, 0.05 mol) and DIPEA (8.7 mL, 0.05 mol) were added to a solution of **2.113** (5 g, 0.022 mol) in peptide grade DMF (50 mL) and the solution was heated to 60 °C for 3 days. Water (200 mL) was added and the mixtures extracted with Et₂O (3 × 50 mL). The organic phases were combined, dried over Na₂SO₄ and concentrated *in vacuo*. The residue was purified through silica gel chromatography (EtOAc/hexanes 3:7) to afford the pure desired compound **2.123** as a white solid (2.77 g, 42 %).

m.p. = 125-130 °C

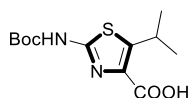
^1H -NMR (CDCl₃, 500 MHz) δ : 1.32 (d, 6H, J = 6.9 Hz, Nt-CH₃), 1.51 (s, 9H, Boc-CH₃), 3.88 (s, 3H, O-CH₃), 4.10 (sep, 1H, J = 6.9 Hz, Nt-CH), 8.69 (broad, 1H, CONH).

^{13}C -NMR (CDCl₃, 125 MHz) δ : 24.9, 27.7, 28.3, 52.0, 83.0, 133.8, 152.6, 152.8, 155.7, 162.8.

IR (neat) ν_{max} : 2964, 1720, 1703, 1556, 1151.

ESI (+ve mode) HRMS calculated for $[\text{M}+\text{H}]^+$ 301.1217, found 301.1220.

2-((tert-butoxycarbonyl)amino)-5-isopropylthiazole-4-carboxylic acid (2.89)



1 M NaOH (aq.) (5 mL) was added to a solution of **2.123** (0.145 g, 0.48 mmol) in EtOH (15 mL) and heated to 70 °C for 4 hours. The mixture was cooled to room temperature and dried

in vacuo. The residue was dissolved in water and acidified to pH 3 through the addition of 1M HCl (aq.). The mixture was extracted with EtOAc (3× 10 mL), dried over Na₂SO₄ and concentrated *in vacuo* to afford the desired compound **2.89** as a white solid (0.11 g, 80%).

m.p. = 255-260 °C

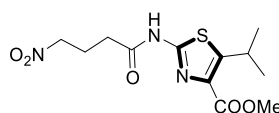
¹H-NMR (d₆-DMSO, 500 MHz) δ: 1.25 (d, 6H, *J* = 6.9 Hz, Nt-CH₃), 1.46 (s, 9H, Boc-CH₃), 3.99 (sep, 1H, *J* = 6.9 Hz, Nt-CH).

¹³C-NMR (d₆-DMSO, 125 MHz) δ: 24.7, 26.8, 27.8, 81.2, 135.0, 149.4, 153, 155.0, 163.5.

IR (neat) ν_{\max} : 2979 (br), 2959, 1720, 1672, 1573, 1160.

ESI (-ve mode) HRMS calculated for [M-H]⁻ 285.0915, found 285.0912.

Methyl 5-isopropyl-2-(4-nitrobutanamido)thiazole-4-carboxylate (2.125)



2.124 was freshly prepared by suspending 4-nitrobutanoic acid (0.5 g, 3.76 mmol) in thionyl chloride (1.2 mL, 17 mmol) and heating to reflux for 4 hours. The thionyl chloride was removed under reduced pressure and the resulting oil was washed with hexanes (4× 10 mL) and used in the next step without any further purification. TEA (1.1 mL, 7.7 mmol) was added to a solution of **2.113** (0.568 g, 2.55 mmol) in dry DCM (10 mL) and the mixtures cooled to 0 °C. **2.124** was dissolved in dry DCM (5 mL) and added to the first solution. The mixture was stirred at room temperature for 12 hours. DCM (10 mL) was added to the reaction and the mixture was washed with water (3× 10 mL) and pH 3 water (3× 10 mL). The organic phases were combined, washed with brine, dried over Na₂SO₄ and adsorbed onto silica. The compound was purified by silica gel column chromatography (1:1 EtOAc/hexanes) to give the desired compound as a white solid (0.025 g, 3 %).

m.p. = 145-150 °C

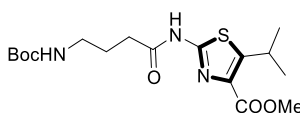
$^1\text{H-NMR}$ (CDCl_3 , 500 MHz) δ : 1.35 (d, 6H, $J = 6.8$ Hz, Nt- CH_3), 2.39-2.42 (m, 2H, CH_2), 2.60-2.62 (m, 2H, CH_2), 3.89 (s, 1H, O- CH_3), 4.08 (sep, 1H, $J = 6.8$ Hz, Nt- CH), 4.52-4.54 (m, 2H, CH_2), 10.26 (s, 1H, CONH).

$^{13}\text{C-NMR}$ (CDCl_3 , 125 MHz) δ : 22.3, 25.0, 27.9, 32.0, 52.2, 74.4, 133.5, 153.3, 154.3, 162.8, 169.6.

IR (neat) ν_{max} : 3228, 1716, 1692, 1543, 1348, 1205, 1168.

ESI (+ve mode) HRMS calculated for $[\text{M}+\text{H}]^+$ 316.0962, found 316.0964.

Methyl 2-(4-((tert-butoxycarbonyl)amino)butanamido)-5-isopropylthiazole-4-carboxylate
(**2.126**)



A solution of **2.28** (1.59 g, 7.8 mmol), EDC (4.88 g, 26 mmol) and DMAP (2.9 g, 23.7 mmol) in dry DMF (10 mL) was stirred for 2 hours at room temperature. **2.113** (1.57 g, 6.7 mmol) was added and the mixture was stirred for 8 days. Water (20 mL) was added and the mixture extracted with EtOAc (3x 10 mL). The organics were combined and washed with 10% aqueous citric acid (6x 15 mL) and NaHCO_3 sat. (4x 15 mL), and brine and dried over Na_2SO_4 . The residue was adsorbed onto silica and purified by silica gel column chromatography (1:1 EtOAc/hexanes) give the desired **2.126** as a white solid (0.36 g, 14%).

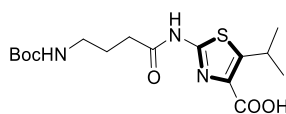
m.p. = 125-130 °C

$^1\text{H-NMR}$ (CDCl_3 , 500 MHz) δ : 1.33 (d, 6H, $J = 6.9$ Hz, Nt- CH_3), 1.42 (s, 9H, Boc- CH_3), 1.89-1.91 (m, 2H, CH_2), 2.47-2.49 (m, 2H, CH_2), 3.18-3.22 (m, 2H, CH_2), 3.88 (s, 3H, O- CH_3), 4.07 (sep, 1H, $J = 6.9$ Hz, Nt- CH), 4.73 (br, CONH), 10.35 (br, 1H, CONH).

$^{13}\text{C-NMR}$ (CDCl_3 125 MHz) δ : 24.9, 26.0, 27.8, 28.5, 33.3, 39.7, 52.1, 79.7, 133.5, 153.0, 154.3, 156.5, 162.9, 171.2.

IR (neat) ν_{max} : 3381, 3262, 2970, 1724, 1687, 1538, 1201, 1162.

ESI (+ve mode) HRMS calculated for $[\text{M}+\text{H}]^+$ 386.1744, found 386.1745.

2-(4-((tert-butoxycarbonyl)amino)butanamido)-5-isopropylthiazole-4-carboxylic acid (2.90)

1 M NaOH (aq.) (6 ml) was added to a solution of **2.126** (0.33 g, 0.9 mmol) in THF (10 mL) and the mixture was stirred for 12 hours at room temperature. The THF was removed under reduced pressure and the aqueous phase was washed with EtOAc (2× 10 mL). The water was acidified to pH 3 using 1 M HCl and extracted with EtOAc (3× 10 mL). The organic phases were combined, washed with brine and dried over Na₂SO₄. The residue was concentrated *in vacuo* to give the desired compound as a white powder (0.234 g, 70%).

m.p. = 165-170 °C

¹H-NMR (d₆-DMSO, 500 MHz) δ: 1.26 (d, 6H, *J* = 6.8 Hz, Nt-CH₃), 1.36 (s, 9H, Boc-CH₃), 1.67-1.71 (m, 2H, CH₂), 2.37-2.40 (m, 2H, CH₂), 2.91-2.95 (m, 2H, CH₂), 4.0 (sep, 1H, *J* = 6.8 Hz, Nt-CH), 6.79 (br, 1H, CONH), 12.22 (br, 1H, CONH), 12.67 (br, COOH).

¹³C-NMR (CDCl₃, 125 MHz) δ: 24.7, 25.1, 26.8, 28.2, 32.3, 77.5, 134.7, 149.5, 153.3, 155.6, 163.6, 171.3.

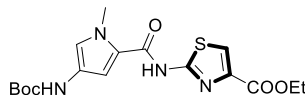
IR (neat) ν_{max}: 2990 (br), 1681, 1562, 1519, 1164.

ESI (+ve mode) HRMS calculated for [M+H]⁺ 372.1588, found 372.1588.

General procedure for 2.135a-c

A solution of **2.25** (1 eq.), EDC (4 eq.) and DMAP (3 eq.) in dry DMF (10 mL) was stirred at room temperature for 2 hours. The amine (**2.108**, **2.112**, **2.113**) (1.5 eq.) was added and the mixture was stirred for 5-10 days. Water was added (50 mL) and the mixture was extracted with EtOAc (3× 15 mL). The organic layers were combined, washed with brine and dried over Na₂SO₄, adsorbed onto silica and purified through silica gel column chromatography (3:7 EtOAc/hexanes), to give the desired compound.

*Methyl 2-(4-((tert-butoxycarbonyl)amino)-1-methyl-1H-pyrrole-2-carboxamido)thiazole-4-carboxylate (2.135a)*¹⁹⁴



White solid

Yield: 72%

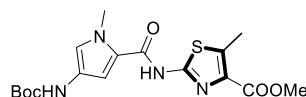
¹H-NMR (CDCl₃, 500 MHz) δ: 1.41 (t, 3H, *J* = 7.1 Hz, OCH₂-CH₃), 1.51 (s, 9H, Boc-CH₃), 3.96 (s, 3H, Py-CH₃), 4.41 (q, 2H, *J* = 7.1 Hz, O-CH₂CH₃), 6.37 (s, 1H, Py-*H*), 6.83 (s, 1H, Py-*H*), 7.29 (s, 1H, CONH), 7.80 (s, 1H, Nt-*H*).

¹³C-NMR (CDCl₃, 101 MHz) δ: 14.4, 28.1, 28.5, 37.2, 61.5, 105.0, 120.5, 121.2, 122.0, 122.9, 141.7, 153.3, 158.4, 161.6 (one carbon not observed).

IR (neat) ν_{\max} : 3287, 1716, 1686, 1660, 1232, 1161.

ESI (+ve mode): 417.1 [M+Na]

Methyl 2-(4-((tert-butoxycarbonyl)amino)-1-methyl-1H-pyrrole-2-carboxamido)-5-methylthiazole-4-carboxylate (2.135b)



Pale yellow solid

Yield: 11%

m.p.: 100-102°C

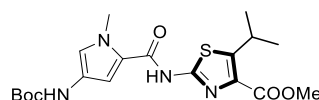
¹H-NMR (CDCl₃, 500 MHz) δ: 1.51 (s, 9H, Boc-CH₃), 2.69 (s, 3H, Nt-CH₃), 3.94 (2×s, 6H, N-CH₃ and O-CH₃), 6.44 (s, 1H, Py-*H*), 6.90 (s, 1H, CONH), 7.30 (s, 1H, Py-*H*).

¹³C-NMR (CDCl₃, 101 MHz) δ: 12.6, 28.5, 37.2, 52.2, 80.7, 104.6, 121.2, 122.9, 134.9, 139.1, 153.9, 158.0, 162.8 (2 carbons not observed).

IR (neat) ν_{\max} : 3346, 2967, 1705, 1548, 1162.

ESI (+ve mode) HRMS calculated for $[M+Na]^+$ 417.1203, found 417.1183.

Methyl 2-(4-((*tert*-butoxycarbonyl)amino)-1-methyl-1H-pyrrole-2-carboxamido)-5-isopropylthiazole-4-Carboxylate (**2.135c**)



Pale yellow solid

Yield: 52%

m.p.: 95-98°C

$^1\text{H-NMR}$ (CDCl_3 , 500 MHz) δ : 1.35 (d, 6H, $J = 6.5$ Hz, Nt- CH_3), 1.49 (s, 9H, Boc- CH_3), 3.89 (s, 3H, O- CH_3), 3.94 (s, 3H, Py- CH_3), 4.10 (sep, 1H, $J = 6.5$ Hz, Nt- CH), 6.41 (s, 1H, Py- H), 6.67 (s, 1H, CONH), 7.21 (s, 1H, Py- H).

$^{13}\text{C-NMR}$ (CDCl_3 , 101 MHz) δ : 24.9, 27.67, 37.2, 52.1, 80.6, 104.7, 120.5, 121.2, 122.8, 133.5, 152.9, 154.2, 158.3, 162.8.

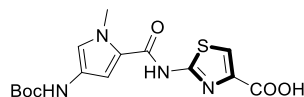
IR (neat) ν_{\max} : 2922, 1714, 1658, 1537, 1158.

ESI (+ve mode) HRMS calculated for $[M+Na]^+$ 445.1516, found 445.1495.

General procedure for 2.91a-c

1 M KOH (aq.) (5-10 eq.) was slowly added to a solution of ester **2.135a-c** (1 eq.) in THF (10 mL) and stirred overnight at room temperature. THF was removed under reduced pressure, and the aqueous phase was acidified to pH 3 with 1 M HCl (aq.) and extracted with EtOAc (3 \times 5 mL). The organic layers were combined, washed with brine and dried over Na_2SO_4 and taken to dryness to give the desired product.

2-(4-((*tert*-butoxycarbonyl)amino)-1-methyl-1*H*-pyrrole-2-carboxamido)thiazole-4-carboxylic acid (**2.91a**)¹⁹⁴



Yellow solid

Yield: 75%

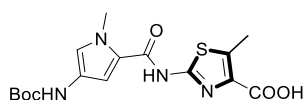
¹H-NMR (d₆-DMSO, 400 MHz) δ: 1.46 (s, 9H, Boc-CH₃), 3.86 (s, 3H, Py-CH₃), 7.14 (s, 1H, Py-H), 7.17 (s, 1H, Py-H), 7.88 (s, 1H, Nt-H), 9.17 (s, 1H, CONH).

¹³C-NMR (d₆-DMSO, 125 MHz) δ: 28.2, 36.5, 78.4, 106.6, 120.1, 120.2, 121.2, 123.9, 142.9, 152.8, 158.3, 159.1, 162.7.

IR (neat) ν_{max}: 2930 (br), 1684, 1670, 1240, 1165.

ESI (-ve mode): 365.0 [M-H]⁻

2-(4-((*tert*-butoxycarbonyl)amino)-1-methyl-1*H*-pyrrole-2-carboxamido)-5-methylthiazole-4-carboxylic acid (**2.91b**)



Pale yellow solid

Yield: 80%

m.p.: decomposition

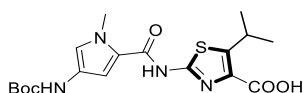
¹H-NMR (d₆-DMSO, 400 MHz) δ: 1.45 (s, 9H, Boc-CH₃), 2.60 (s, 3H, Nt-CH₃), 3.84 (s, 3H, Py-CH₃), 7.13 (s, 1H, Py-H), 7.15 (s, 1H, Py-H), 9.17 (s, 1H, CONH), 12.30 (s, 1H, CONH); 12.67 (br, 1H, COOH).

¹³C NMR (d₆-DMSO, 101 MHz) δ: 12.1, 28.2, 36.5, 78.4, 106.6, 120.1, 123.0, 136.3, 152.9, 153.8, 159.0, 163.7 (2 carbons not observed).

IR (neat) ν_{\max} : 2960 (br), 1664, 1654, 1542, 1166.

ESI (+ve mode) HRMS calculated for $[M+Na]^+$ 403.1047, found 403.1028.

2-(4-((*tert*-butoxycarbonyl)amino)-1-methyl-1*H*-pyrrole-2-carboxamido)-5-isopropylthiazole-4-carboxylic acid (**2.91c**)



Pale yellow solid

Yield: 82%

m.p.: decomposition (> 200 °C).

$^1\text{H-NMR}$ (d_6 -DMSO, 400 MHz) δ : 1.28 (d, 6H, $J = 6.8$ Hz, Nt- CH_3), 1.45 (s, 9H, Boc- CH_3), 3.84 (s, 3H, Py- CH_3), 4.03 (sep, 1H, $J = 6.8$ Hz, Nt- CH), 7.12 (s, 1H, Py- H), 7.16 (s, 1H, Py- H), 9.14 (s, 1H, CONH), 12.30 (br, 1H, CONH), 12.71 (br, 1H, COOH).

$^{13}\text{C-NMR}$ (d_6 -DMSO 101, MHz) δ : 24.8, 26.9, 28.2, 36.5, 78.4, 106.6, 120.1, 122.9, 134.7, 149.4, 152.9, 153.9, 159.0, 163.7 (one carbon not observed).

IR (neat) ν_{\max} : 2954 (br), 1723, 1666, 1158.

ESI (+ve mode) HRMS calculated for $[M+Na]^+$ 431.1360, found 431.1339.

2.7.7 General procedure for synthesis of polyamides

All polyamides were prepared using Boc- β -Ala-PAM resin (0.5 mmol/g loading)(Sigma-Aldrich). After swelling the resin for 1 hour in DCM, the Boc group was removed using TPW (TFA 92.5%, Phenol 5%, water 2.5%) solution (1 \times 1 minute, 2 \times 3 minutes for Boc-Py-OH and Boc- γ -turn-OH or 1 \times 1 minute, 2 \times 10 minutes for Boc-Im-OH).

Py-NH₂/Py-OH, turn-NH₂/Py-OH coupling

Following deprotection the resin was washed with DMF (4 \times 1 minute), conditioned with anhydrous DMF and flushed with nitrogen. The monomer (4 eq.) was activated with HATU

(3.6 eq.) and DIPEA (9 eq.) for 20 minutes in DMF (1.5 mL). The monomer solution was then added to the resin and shaken for 2 hours at room temperature.

Py- NH₂/Im-OH and Im-NH₂/turn-OH coupling

After deprotection, the resin was washed with DMF (4× 1 minute), conditioned with anhydrous DMF and flushed with nitrogen. The monomer (4 eq.) was activated with DCC (3.6 eq.) and HOAt (3.8 eq.) in DMF (1.5 mL) for 2 hours. The activated monomer was filtered to remove the precipitated DCU and the solution was added to the resin. Following the addition of DIPEA (9 eq.), the reaction was shaken for 6-12 hours at room temperature.

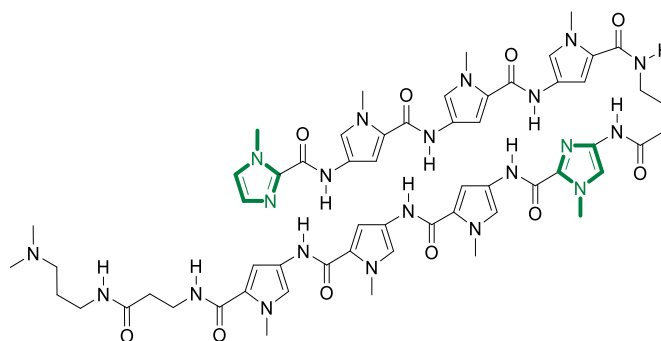
Py-NH₂/Nt-OH coupling

Following deprotection, the resin was washed with DMF (4× 1 minute), conditioned with anhydrous DMF and flushed with nitrogen. The solution with the monomer/dimer (4 eq.) was activated with HATU (3.6 eq.) and DIPEA (9 eq.) for 30 minutes in DMF (1.5 mL). The solution was then added to the resin and shaken for 12 hours at room temperature.

General cleavage procedure

Polyamides were cleaved from the resin using dimethylaminopropylamine (**2.30**, 1.5 mL) and heating at 70 °C for 16 hours. The resin was removed by filtration, washed with MeOH (3× 1 mL). The crude polyamide was then purified through RP-HPLC.

Compound **2.106**



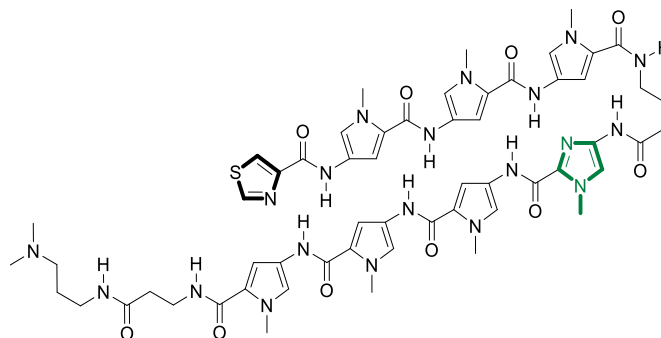
Yield: 16% (19.5 mg)

$^1\text{H-NMR}$ (d_6 -DMSO, 500 MHz) δ : 1.71-1.83 (m, 4H), 2.34-2.38 (m, 4H), 2.63-2.64 (m, 1H), 2.74 (s, 3H), 2.75 (s, 3H), 2.99-3.03 (m, 2H), 3.10-3.14 (m, 2H), 3.19-3.23 (m, 2H), 3.37-3.41 (m, 2H), 3.80 (s, 6H), 3.85-3.86 (m, 12H), 3.96 (s, 3H), 4.00 (s, 3H), 6.88 (d, 1H, $J = 1.7$ Hz), 6.90 (d, 1H, $J = 1.7$ Hz), 7.06-7.07 (m, 2H), 7.16-7.17 (m, 4H), 7.22 (d, 1H, $J = 1.7$ Hz), 7.23 (d, 1H, $J = 1.7$ Hz), 7.27 (d, 1H, $J = 1.7$ Hz), 7.29 (d, 1H, $J = 1.7$ Hz), 7.41 (s, 1H), 7.46 (s, 1H), 8.02-8.03 (m, 2H), 9.21 (s, 1H), 9.88 (s, 1H), 9.90 (s, 1H), 9.93 (s, 1H), 9.95 (s, 1H), 9.96 (s, 1H) 10.24 (s, 1H), 10.45 (s, 1H).

HPLC purity (310 nm): 99%; Retention time: 20.15 min.

MALDI-TOF-MS: $[\text{M}+\text{H}]^+$ calculated for $\text{C}_{58}\text{H}_{72}\text{N}_{21}\text{O}_{10}^+$ 1222.58, found 1222.30.

Compound **2.122a**

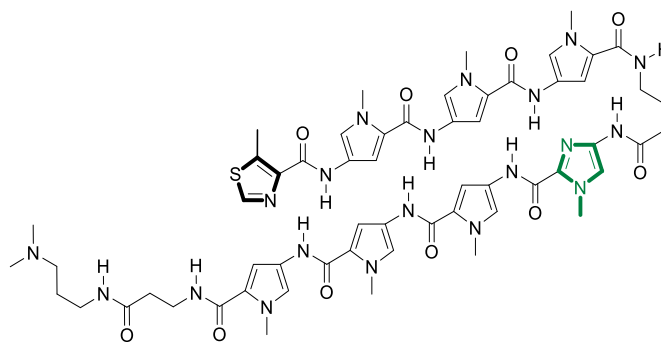


Yield: 20% (24.5 mg)

$^1\text{H-NMR}$ (d_6 -DMSO, 500 MHz) δ : 1.72-1.83 (m, 4H), 2.34-2.38 (m, 4H), 2.63-2.64 (m, 1H), 2.74 (s, 3H), 2.75 (s, 3H), 2.99-3.03 (m, 2H), 3.10-3.14 (m, 2H), 3.19-3.23 (m, 2H), 3.37-3.41 (m, 2H), 3.81 (s, 6H), 3.85-3.87 (m, 12H), 3.96 (s, 3H), 6.88 (d, 1H, $J = 1.7$ Hz), 6.90 (d, 1H, $J = 1.7$ Hz), 7.06-7.07 (m, 2H), 7.16-7.18 (m, 4H), 7.23 (d, 1H, $J = 1.7$ Hz), 7.24 (d, 1H, $J = 1.7$ Hz), 7.27 (d, 1H, $J = 1.7$ Hz), 7.33 (d, 1H, $J = 1.7$ Hz), 7.46 (s, 1H), 8.02-8.03 (m, 2H), 8.40 (d, 1H, $J = 2$ Hz), 9.24 (d, 1H, $J = 2$ Hz) 9.34 (s, 1H), 9.88 (s, 1H), 9.90 (s, 1H), 9.93 (s, 1H), 9.96 (s, 1H), 9.97 (s, 1H), 10.24 (s, 1H), 10.44 (s, 1H).

HPLC purity (310 nm): 98.77%; Retention time: 20.65 min.

MALDI-TOF-MS: $[\text{M}+\text{H}]^+$ calculated for $\text{C}_{57}\text{H}_{69}\text{N}_{20}\text{O}_{10}\text{S}^+$ 1225.52 found 1225.25.

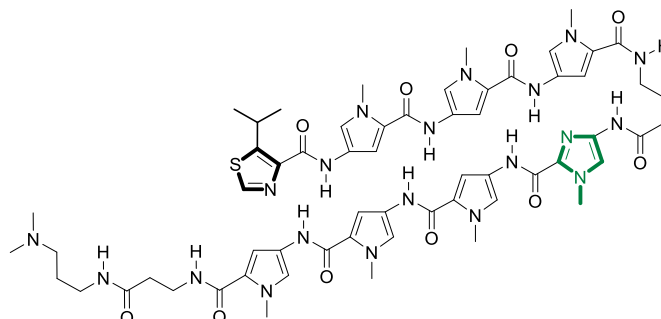
Compound **2.122b**

Yield: 26% (32.2 mg)

$^1\text{H-NMR}$ (d_6 -DMSO, 400 MHz) δ : 1.71-1.83 (m, 4H), 2.32-2.39 (m, 4H), 2.66-2.68 (m, 1H), 2.74 (s, 3H), 2.75 (s, 3H), 2.80 (s, 3H), 2.97-3.03 (m, 2H), 3.09-3.14 (m, 2H), 3.18-3.23 (m, 2H), 3.36-3.41 (m, 2H, obstructed by water peak), 3.81 (s, 6H), 3.85-3.86 (m, 12H), 3.96 (s, 3H), 6.88 (d, 1H, $J = 1.7$ Hz), 6.90 (d, 1H, $J = 1.7$ Hz), 7.06-7.07 (m, 2H), 7.16-7.17 (m, 2H), 7.17 (d, 1H, $J = 1.7$ Hz), 7.20 (d, 1H, $J = 1.7$ Hz), 7.22 (d, 1H, $J = 1.7$ Hz), 7.23 (d, 1H, $J = 1.7$ Hz), 7.27 (d, 1H, $J = 1.7$ Hz), 7.32 (d, 1H, $J = 1.7$ Hz), 7.46 (d, 1H, $J = 1.7$ Hz), 8.02-8.05 (m, 2H), 8.97 (s, 1H), 9.21 (s, 1H), 9.88 (s, 2H), 9.94 (s, 2H), 9.96 (s, 1H), 10.24 (s, 1H), 10.29 (s, 1H).

HPLC purity (310 nm): 99%; Retention time: 21.28 min.

MALDI-TOF-MS: $[\text{M}+\text{H}]^+$ calculated for 1239.54 $\text{C}_{58}\text{H}_{71}\text{N}_{20}\text{O}_{10}\text{S}^+$, found 1239.26.

Compound **2.122c**

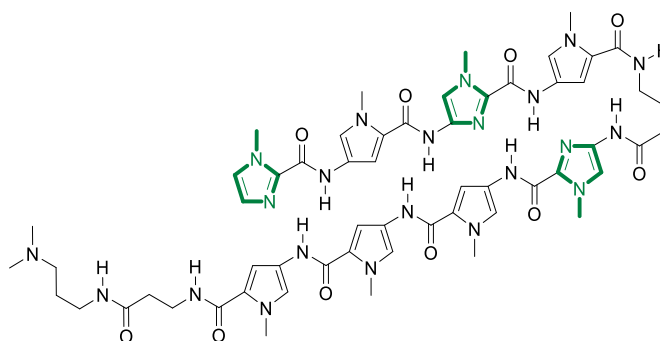
Yield: 17% (21.5 mg)

$^1\text{H-NMR}$ (d_6 -DMSO, 500 MHz) δ : 1.31 (d, 6H, $J = 6.8$ Hz), 1.73-1.83 (m, 4H), 2.34-2.38 (m, 4H), 2.63-2.64 (m, 1H), 2.73 (s, 3H), 2.74 (s, 3H), 2.98-3.02 (m, 2H), 3.10-3.14 (m, 2H), 3.19-3.23 (m, 2H), 3.37-3.41 (m, 2H), 3.81 (s, 6H), 3.85-3.86 (m, 12H), 3.96 (s, 3H), 4.33 (sep, 1H, $J = 6.8$ Hz), 6.88 (d, 1H, $J = 1.7$ Hz), 6.90 (d, 1H, $J = 1.7$ Hz), 7.06-7.07 (m, 2H), 7.16-7.18 (m, 3H), 7.21 (d, 1H, $J = 1.7$ Hz), 7.23 (d, 1H, $J = 1.7$ Hz), 7.24 (d, 1H, $J = 1.7$ Hz), 7.27 (d, 1H, $J = 1.7$ Hz), 7.33 (d, 1H, $J = 1.7$ Hz), 7.46 (s, 1H), 8.01-8.06 (m, 2H), 9.02 (s, 1H), 9.48 (s, 1H), 9.88 (s, 1H), 9.89 (s, 1H), 9.94 (s, 2H), 9.96 (s, 1H), 10.25 (s, 1H), 10.29 (s, 1H).

HPLC purity (310 nm): 99%; Retention time: 22.79 min.

MALDI-TOF-MS: $[\text{M}+\text{H}]^+$ calculated for $\text{C}_{60}\text{H}_{75}\text{N}_{20}\text{O}_{10}\text{S}^+$ 1267.57, found 1267.24.

Compound **2.134**

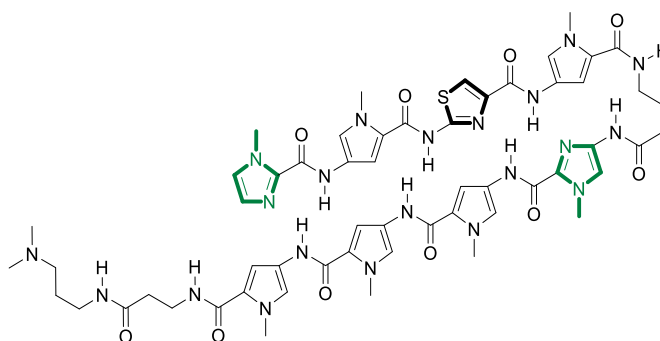


Yield: 11.6% (2.8 mg)

HPLC purity (300 nm): 92%; Retention time: 17.05 min.

MALDI-TOF-MS: $[\text{M}+\text{H}]^+$ calculated for $\text{C}_{57}\text{H}_{71}\text{N}_{22}\text{O}_{10}^+$ 1223.57, found 1223.19.

Compound **2.141a**

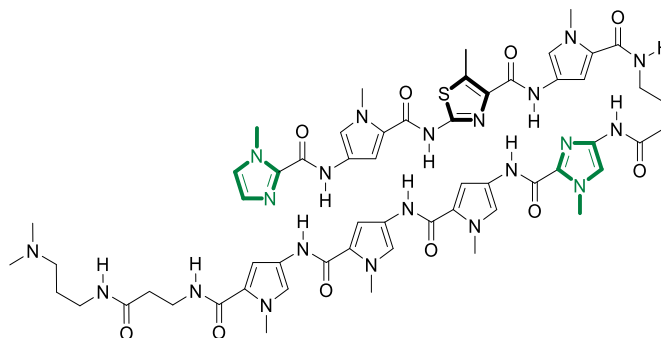


Yield: 5.7% (1.4 mg)

HPLC purity (300 nm): 96.7%; Retention time: 14.78 min.

MALDI-TOF-MS: $[M+Na]^+$ calculated for $C_{56}H_{67}N_{21}NaO_{10}S^+$ 1248.50, found 1248.28.

Compound **2.141b**

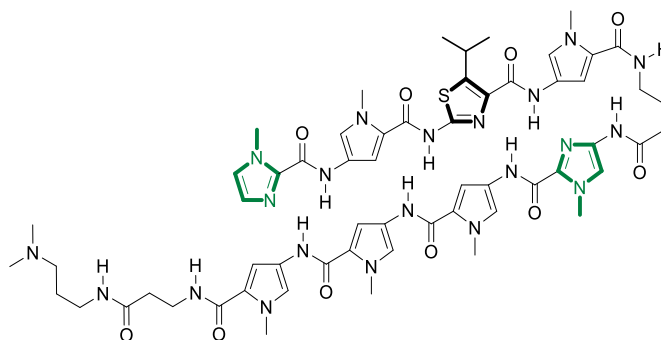


Yield: 3.9% (2.4 mg)

HPLC purity (300 nm): 95.4%; Retention time: 15.36 min.

MALDI-TOF-MS: $[M+H]^+$ calculated for $C_{57}H_{70}N_{21}O_{10}S^+$ 1240.53, found 1240.02.

Compound **2.141c**



Yield: 4.4% (2.2 mg)

HPLC purity (300 nm): 98.5%; Retention time: 16.71 min.

MALDI-TOF-MS: $[M+H]^+$ calculated for $C_{59}H_{74}N_{21}O_{10}S^+$ 1268.56, found 1268.23.

Chapter 3

Assessing the DNA binding profile of Nt- containing polyamides

3.1 Determination of MGBs interaction with dsDNA

The interaction between an MGB and its dsDNA sequence is governed by the free energy of binding ΔG described by Equation 3.1.^{199,200}

$$\Delta G = \Delta H - T\Delta S \quad (\text{Equation 3.1})$$

The ΔG value is determined by the variations in enthalpy (ΔH) and entropy (ΔS), which can contribute positively or negatively to the interaction. In a spontaneous process, the total contribution of enthalpy and entropy results in negative ΔG values. The enthalpy (ΔH) reflects the introduction or removal of the forces responsible for the formation of a stable complex (*i.e.*, hydrogen bonds and van der Waals).²⁰¹ An enthalpically driven binding is characterised by a negative ΔH value (favoured), indicating that the interaction is exothermic, with heat released upon complex formation.²⁰¹ On the other hand, positive ΔH values (disfavoured) are typical of an endothermic process, requiring an intake of heat for the interaction to occur. The entropy (ΔS) is related to the overall disorder of the system. Upon interaction, the limited degree of motion of the ligand/dsDNA complex as compared to unbound molecules results in negative (disfavoured) entropic contributions. However, changes to the surrounding environment also contribute to the free energy of molecular interactions. When an MGBs binds to dsDNA, ordered water molecules associated with the DNA are displaced, resulting in an increase in the entropy of the environment. This entropic increase can lead to an overall positive ΔS value (favoured) for the interaction. The overall strength of the binding is normally expressed using the equilibrium dissociation constant (K_D), which is related to the free energy of binding by Equation 3.2.

$$\Delta G = RT \ln K_D \quad (\text{Equation 3.2})$$

Solving for the equilibrium dissociation constant K_D provides kinetic information, since it is also defined as the ratio between the rate constants k_{on} and k_{off} , as shown (Equation 3.3).²⁰²

$$K_D = \frac{k_{off}}{k_{on}} \quad (\text{Equation 3.3})$$

The determination of the thermodynamic and kinetic parameters of the MGB:dsDNA interaction is essential for establishing the binding affinity and selectivity, and directing the synthetic design of novel analogues. However, this analysis can be complicated by the presence of multiple species contributing to the overall equilibrium, with each component characterised by different kinetics. In the case of hairpin polyamides, the self-aggregation of the ligands has been notoriously problematic for the definition of kinetic and thermodynamic parameters (Figure 3.1).^{163,164,197,203,204}

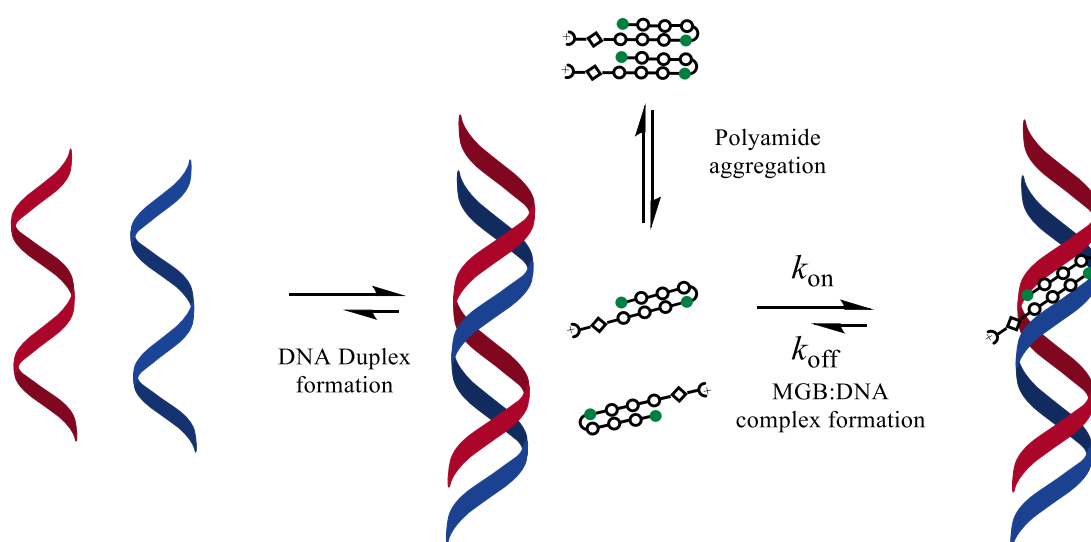


Figure 3.1 Schematic representation of the species present at equilibrium during the formation of a polyamide·dsDNA complex.

3.2 Techniques used to study polyamide·dsDNA complexes

3.2.1 UV melting analysis

UV melting can be used to quickly assess the binding behaviour of libraries of compounds.^{205,206} Due to the aromatic nature of the nucleobases, DNA absorbs ultraviolet radiation at a wavelength of 260 nm. UV melting analysis exploits the hypochromic effect, which is the reduced UV absorption of DNA assuming a double-helix conformation. This

effect results from the electronic interactions between the nucleobases, which are paired and stacked in the double-helix. In a melting experiment, the dsDNA is denatured by heating, which disrupts the hydrogen bonds between the bases and causes the transition to ssDNA.²⁰⁶ This transition is observable as a sigmoidal trend in the UV absorbance at 260 nm (Figure 3.2). This sigmoidal shape is a hallmark of dsDNA melting and is derived from the sequential increase in mobility due to the dissociation of Watson-Crick base pairs, which results in a gradient increase of the absorbance/temperature plot.²⁰⁶ The temperature of melting (T_m) is defined as the temperature when half of the DNA has melted and is influenced by the length, sequence composition and salt concentration.²⁰⁷ For example, the greater number of hydrogen bonds in long GC-rich dsDNA sequences results in a higher T_m . The same trend is observed when the concentration of positive ions is increased, which reduces the electrostatic repulsion between the negative phosphates.²⁰⁷ The melting curve can be used to calculate the enthalpy (ΔH) for the transition of a non-self-complementary duplex to its single-stranded form using Equation 3.4:

$$\Delta H = -6RT_m^2 \left[\frac{\delta \alpha}{\delta T} \right]_{T=T_m} \quad (\text{Equation 3.4})$$

where R is the gas constant and α is the fraction of DNA in the single-stranded form (0.5 at the T_m).²⁰⁵

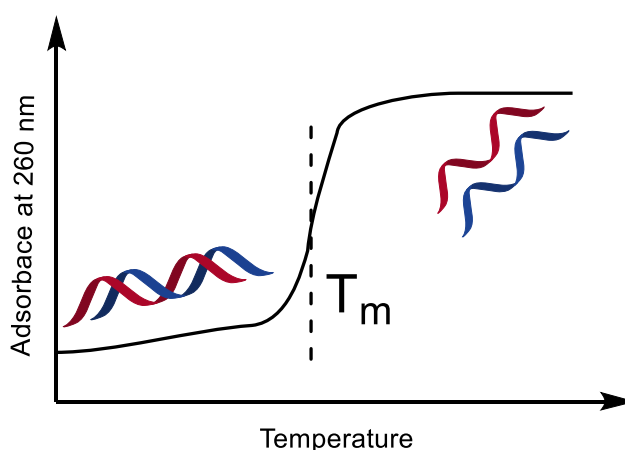


Figure 3.2 Typical sigmoidal curve obtained from a dsDNA melting experiment.

In an MGB·dsDNA complex, the ligand stabilises the duplex, thus increasing the T_m required to promote ligand dissociation and subsequent dsDNA melting. For this reason, DNA melting experiments provides quick, qualitative information on binding affinity and selectivity of dsDNA binding ligands. The increase in T_m can be correlated to the binding affinity using Equation 3.5:²⁰⁶

$$\frac{1}{T_m^\circ} - \frac{1}{T_m} = \frac{R}{\Delta H^\circ} [N_h \ln(1 + K_h a) - N_{ss} \ln(1 + K_{ss} a)] \quad (\text{Equation 3.5})$$

where T_m° and T_m are the melting temperatures of dsDNA in the absence and presence of the ligand respectively; ΔH° is the enthalpy change for the nucleic acid alone, K_h and K_{ss} are the binding constants of the ligand to dsDNA and ssDNA, respectively; N_h and N_{ss} are the number of binding sites of the ligand in dsDNA and ssDNA, respectively; and a is the free ligand concentration at the melting temperature.

While the term accounting for the interactions with single-stranded DNA can potentially be approximated to zero, the presence of a variable in the equation associated with the nature of the ligand (*i.e.*, activity) complicates the direct comparison of different compounds. For this reason, only T_m values are commonly reported for melting experiments for the determination of dsDNA·ligand binding affinity.^{108,197,208} The main advantages of this technique are the relatively low cost and the requirement of small amounts of nucleic acid and ligands. However, melting experiments do not provide any structural information or kinetic data associated with the binding. Another important limitation is that the binding interaction is derived at the melting temperature and not at physiological conditions.

3.2.2 DNase I footprinting

DNase I footprinting is a biochemical method that is often used to complement UV melting for determination of binding affinity and selectivity of MGBs for dsDNA.²⁰⁹ In this experiment, an oligonucleotide fragment containing several potential target sequences is treated with an endonuclease (*e.g.*, DNase I), which cleaves dsDNA. The ligand protects its

bound sequence and thus produces a “footprint”, which can be visualised by an electrophoretic mobility assay (Figure 3.3). Normally, one end of the DNA is radiolabelled with ^{32}P , allowing quick and facile examination of the gel.²⁰⁹ In contrast to UV melting, this method provides information about the sequence-selectivity of a given ligand in the presence of a range of potential target sequences under physiological conditions. Binding constants can be estimated by titrating the ligand at different concentrations and measuring the intensity of the resulting bands. Binding kinetics can also be established by determining the rate of appearance and disappearance of the footprint.²⁰⁹

The primary disadvantage of DNase I footprinting is that the calculation of the apparent dissociation constant for high-affinity binders (*i.e.*, low nanomolar binding affinity) is obfuscated by the lack of enough unbound dsDNA, required for a meaningful analysis. Other limitations are related to the mechanism of cleavage of DNase I. For example, the digestion is biased against the rigid GC regions of DNA, which as a result are particularly challenging substrates to evaluate. Furthermore, the bending of the phosphodiester backbone induced by DNase I upon cleavage causes a footprint that is staggered about 2-3 bases in the 3' direction relative to the ligand binding site.²⁰⁹ Regardless of its limitations, DNase I footprinting is currently the most widely used technique for the screening of sequence-selective ligands.

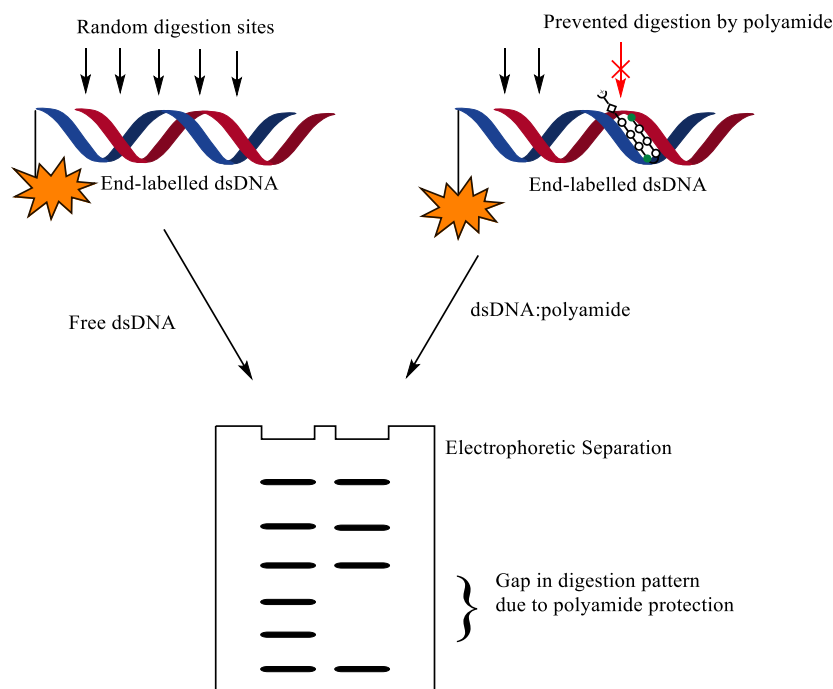


Figure 3.3 Schematic representation of DNase I footprinting experiment.

Single nucleotide resolution can be achieved using alternative footprinting techniques such as methidiumpropyl-EDTA-Fe(II) (MPE-Fe(II)).^{209,210} This experiment exploits the intercalating action of MPE, which, in the presence of an Fe(II) complex and molecular oxygen, cleaves dsDNA in a targeted manner. To further increase the resolution, the EDTA-Fe(II) complex can also be incorporated in the tail of the polyamides to direct the cleavage in the proximity of the binding site (Figure 3.4, **3.1**).

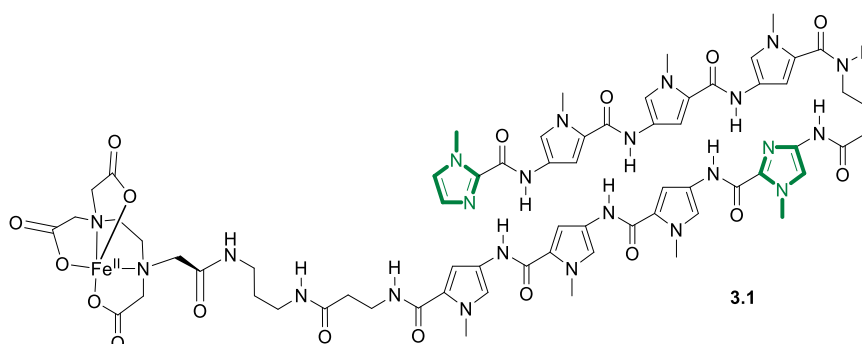


Figure 3.4 Structure of an EDTA(Fe(II)) derivatised MGBs.²¹⁰

3.2.3 Isothermal titration calorimetry (ITC)

The most common technique used for the calculation of ligand·substrate thermodynamics is isothermal titration calorimetry (ITC), which measures the amount of heat dissipated to the surrounding environment upon molecular interactions.^{78,204,211–213} The instrument consists of two separated cells: a reference cell containing buffer and an analytical cell containing the substrate (Figure 3.5). The change in enthalpy is measured by titrating small aliquots of the ligand into the substrate cell. The heat produced by the binding event following each titration point is calculated as the instrument applies thermal energy to maintain a constant temperature. When saturation of the binding site is reached, the heat variation is solely caused by dilution effects.²¹¹

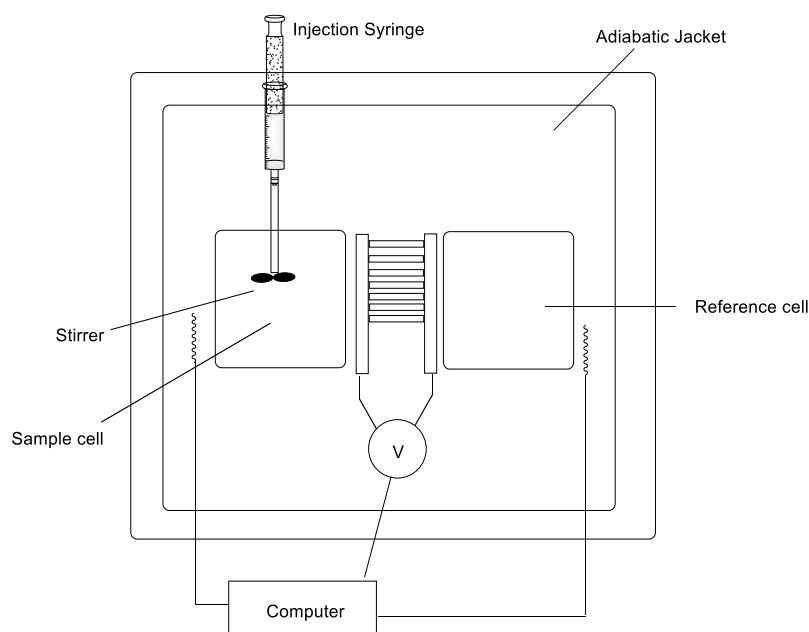


Figure 3.5 Schematic representation of the ITC instrument.²¹²

This directly provides the enthalpy value of the interaction from which the affinity constant and free energy can be derived using mathematical models, chosen according to available information on the binding.²¹² ITC has been widely used for measuring the variation of ΔH induced by the non-covalent forces, such as hydrogen bonding and van der Waals interactions,

essential to MGB-dsDNA interaction.^{78,157,164,204,208,214} However, ITC studies on MGBs are restricted to relatively small-sized molecules. This limitation is related to the chemical structure of most MGBs, which are composed of aromatic rings separated by amide bonds, in which inter-ligand aggregation arises from π - π interactions and hydrogen bonds between amide N-H and oxygen (Figure 3.6). Owing to the aggregation interference, the assumption that the equilibrium is reversible and has been reached is not always valid. As reported by the Wilson group, the contribution of aggregates is negligible for small MGBs at the concentrations required for ITC.¹⁹⁷ However, aggregation becomes significant for polyamides composed of 8 or more aromatic rings, which complicates the number of species involved in the equilibrium. In order to address this problem, the Wilson group reported 8-ring hairpin polyamides containing a doubly-charged tail aimed to reduce the entity of aggregation by electrostatic repulsion (Figure 3.6, **3.2**).¹⁹⁷ While this allows the calculation of binding enthalpies for 8-ring hairpin polyamides without significantly affecting the binding selectivity, the propensity of Py-Im polyamides to aggregate has restricted the use of this technique.

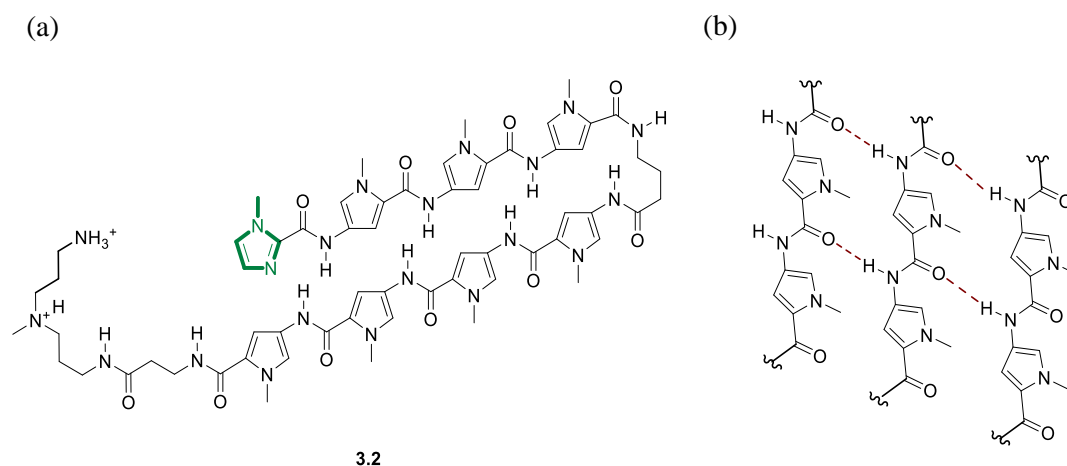


Figure 3.6 (a) Structure of 8-ring polyamide **3.2** used for ITC experiments. (b) Schematic representation of the putative hydrogen bond pattern responsible for polyamide aggregation.

3.2.4 Determining the aggregation properties of MGBs in aqueous solution

Several studies have sought to explore the influence of MGB aggregation on biological activity and pharmacokinetic properties. For example, Parkinson *et al.* analysed the aggregation of small untethered MGBs, known to bind dsDNA with a 2:1 stoichiometry.^{163,164,203} Using a combination of NMR, ITC and molecular dynamics studies, it was established that these compounds pre-assemble in solution in the same dimeric head-to-tail conformation assumed when bound to dsDNA (see section 1.7.2.1 and 1.10.1). Direct evidence of this was found using the thiazotropsin analogue **3.3** through intermolecular nuclear Overhauser effect (nOe) cross-correlations, which could only be explained by the presence of a pre-organised assembly of the ligand (Figure 3.7).¹⁶⁴ Although this pre-organisation contributes favourably to the entropy of the binding to dsDNA, an inverse correlation was observed between the extent of aggregation and antibacterial activity, presumably due to the decreased cell permeability of larger aggregates compared to monomers.²⁰³

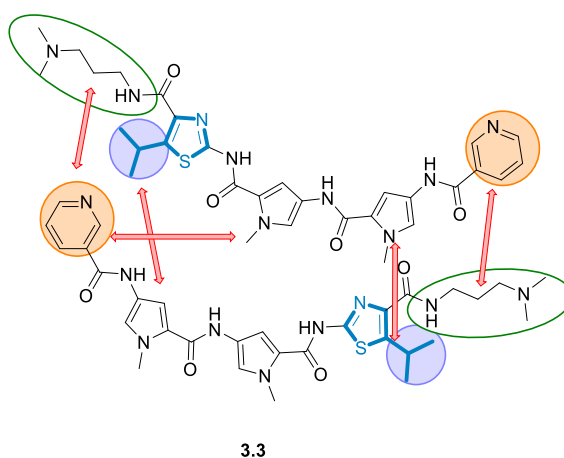


Figure 3.7 Schematic representation of the thiazotropsin analogue **3.3** used for aggregation analysis. Double-arrows indicate the observed nOe correlations that suggest self-assembly.¹⁶⁴

The Dervan group also explored the aggregation effects of MGBs, using larger 8-ring polyamides.¹⁴⁶ In this study, the assembly of a variety of polyamides with different architectures (*i.e.*, hairpin and cyclic) and substituents at the C-terminus and γ -turn was

assessed using dynamic light scattering (DLS). All the compounds were found to form particles with radii ranging from 50 to 500 nm at 1-4 μM concentrations. *In vivo* co-administration of polyamides with the solubilising agent 2-hydroxypropyl- β -cyclodextrin (**3.4**) resulted in increased solubility of the polyamides, with no effect on their circulating levels or cellular internalisation (Figure 3.8).¹⁴⁶

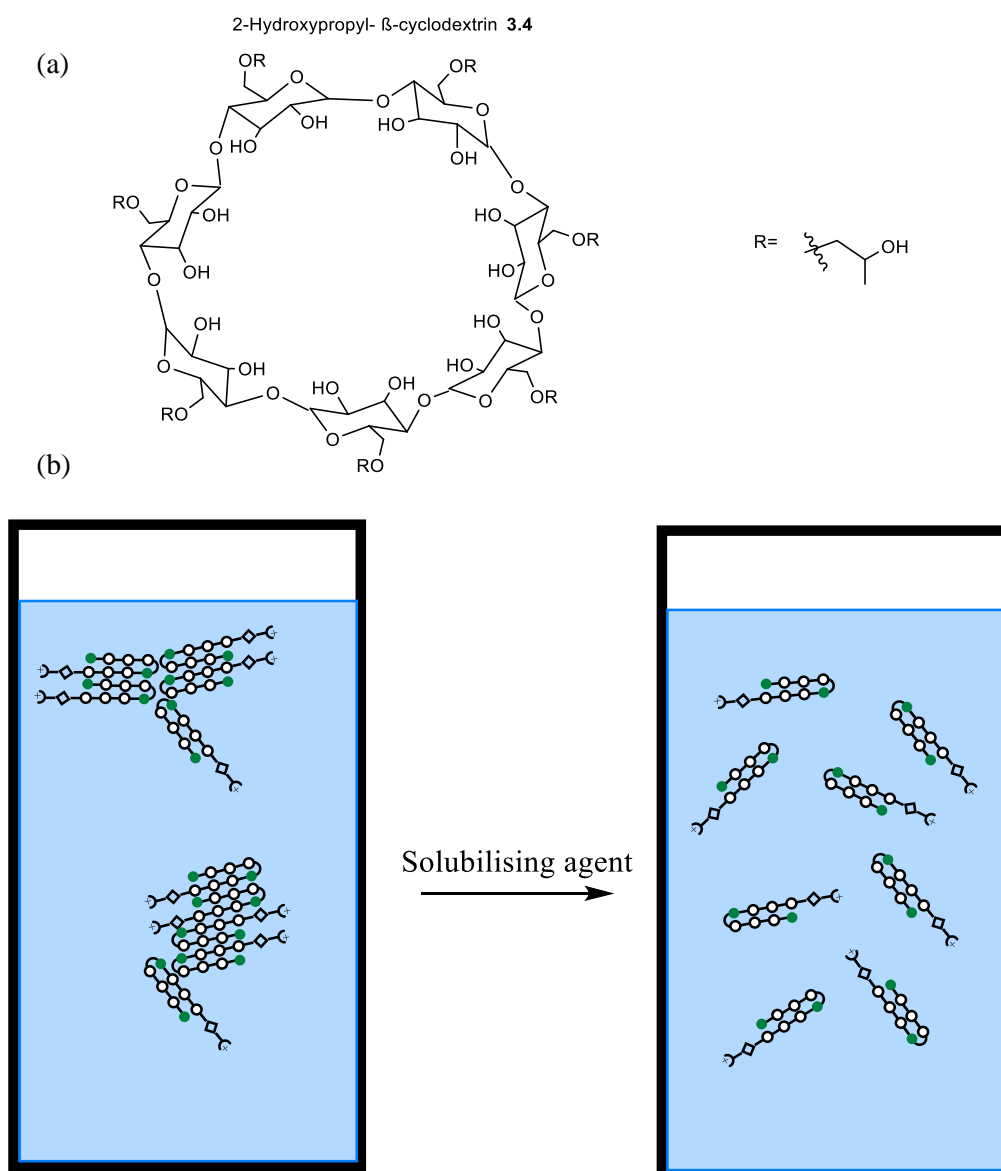


Figure 3.8 (a) Structure of the solubilising agent 2-hydroxypropyl- β -cyclodextrin (**3.4**) used to reduce polyamides aggregation. (b) Schematic representation of the putative effect of 2-hydroxypropyl- β -cyclodextrin on a solution of polyamide.¹⁴⁶

3.2.5 *Surface plasmon resonance*

Surface plasmon resonance (SPR) can be used to determine the kinetic parameters of MGB·dsDNA complexes, complementing the thermodynamic signature obtained by ITC.^{208,214,215} This technique is particularly advantageous as it measures the variation of the refractive index generated upon binding of an analyte to its substrate, which is immobilised on a metal surface, and therefore does not require any molecular labelling. A typical SPR system consists of a gold-coated glass surface functionalised with a target substrate. For the analysis of MGBs, this is normally a hairpin dsDNA.^{197,208} When the surface is irradiated with polarised light of a certain angle of incidence, the beam is partially absorbed, generating a horizontal plasmon wave, while the remaining photons are reflected off (Figure 3.9). The refractive angle of the absorbed light depends on the mass of the surface. For this reason, the change in mass resulting from a binding event can be detected as deviation of this angle. To determine association and dissociation constants, a microfluidic system introduces first ligand, and then buffer, onto the surface and the variation of the reflective angle is measured (Figure 3.9). The reduced influence of the aggregation, derived from the continuous flow of the ligand solution, has allowed the investigation of large Py-Im polyamides.²⁰⁸

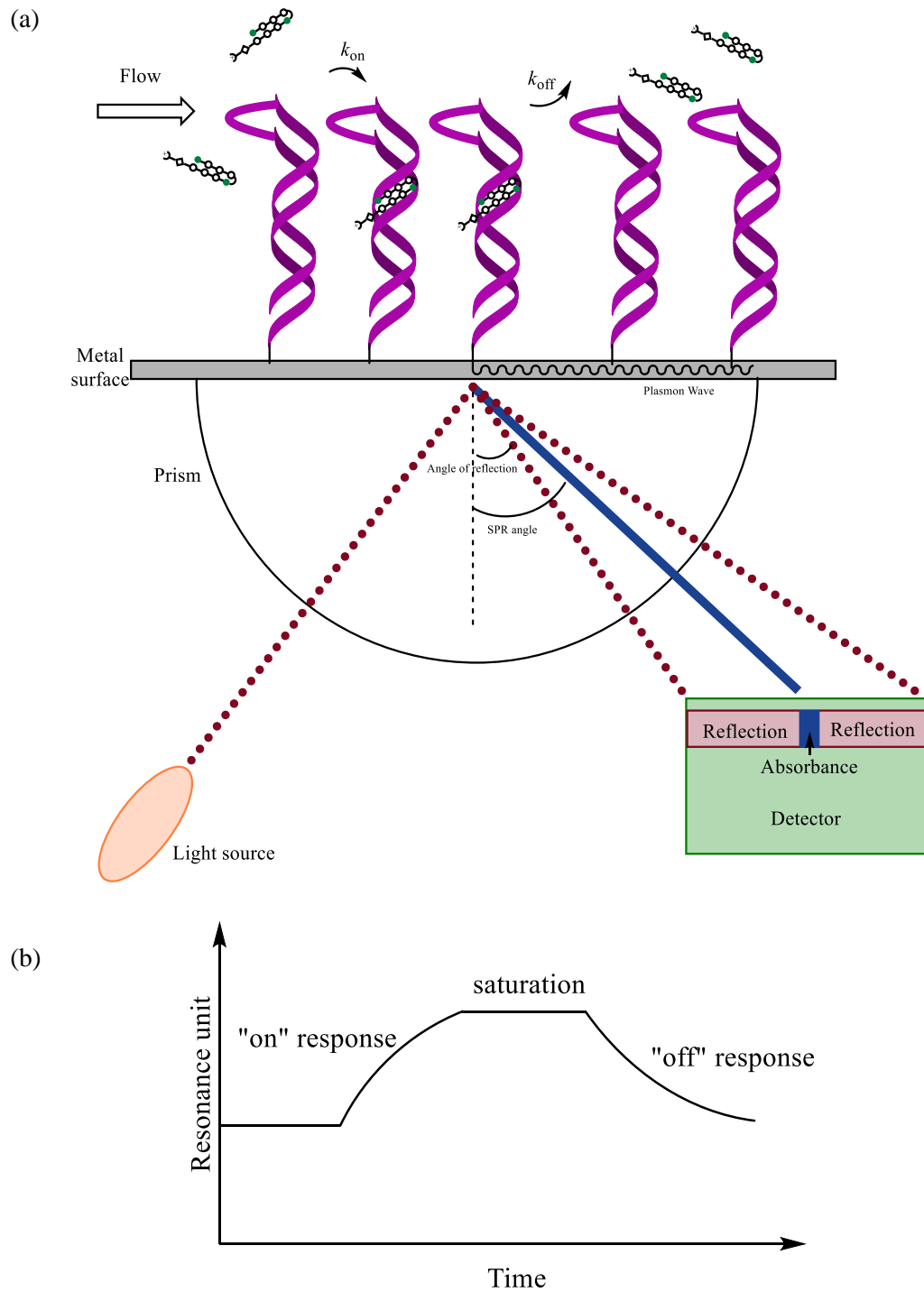


Figure 3.9 (a) Schematic representation of the SPR experiment. (b) SPR sensogram.

3.2.6 SwitchSENSE[®] method to determine kinetic parameters of protein·DNA interactions

SwitchSENSE[®] technology is an emerging technique, representing an alternative to SPR for evaluation of binding affinity and kinetics.^{216,217} The principle at the base of this technology is the controlled movement of DNA levers immobilised on a gold surface. By exploiting the negative charges of the phosphate backbone of DNA, it is possible to alternatively switch the DNA near to or far from the surface by applying alternating electric potential.²¹⁶ The movement is monitored through fluorescence variation generated by a fluorophore attached to the distal portion of the DNA. The levers can be functionalised with the substrate of interest by chemical ligation or exploiting Watson and Crick base pairing. Upon ligand binding, the increased mass of the levers alters their hydrodynamic friction, thus slowing the movement (Figure 3.10). This deceleration can be used to determine affinity constants and on- and off-rates of analytes for their biomolecular target. This new technology has recently found wide applications including investigation of protein·RNA, protein·DNA and protein·protein interactions.^{216,218} Section 3.4.1.4 of this chapter describes the first use of SwitchSENSE[®] for measuring interactions between MGBs and dsDNA.

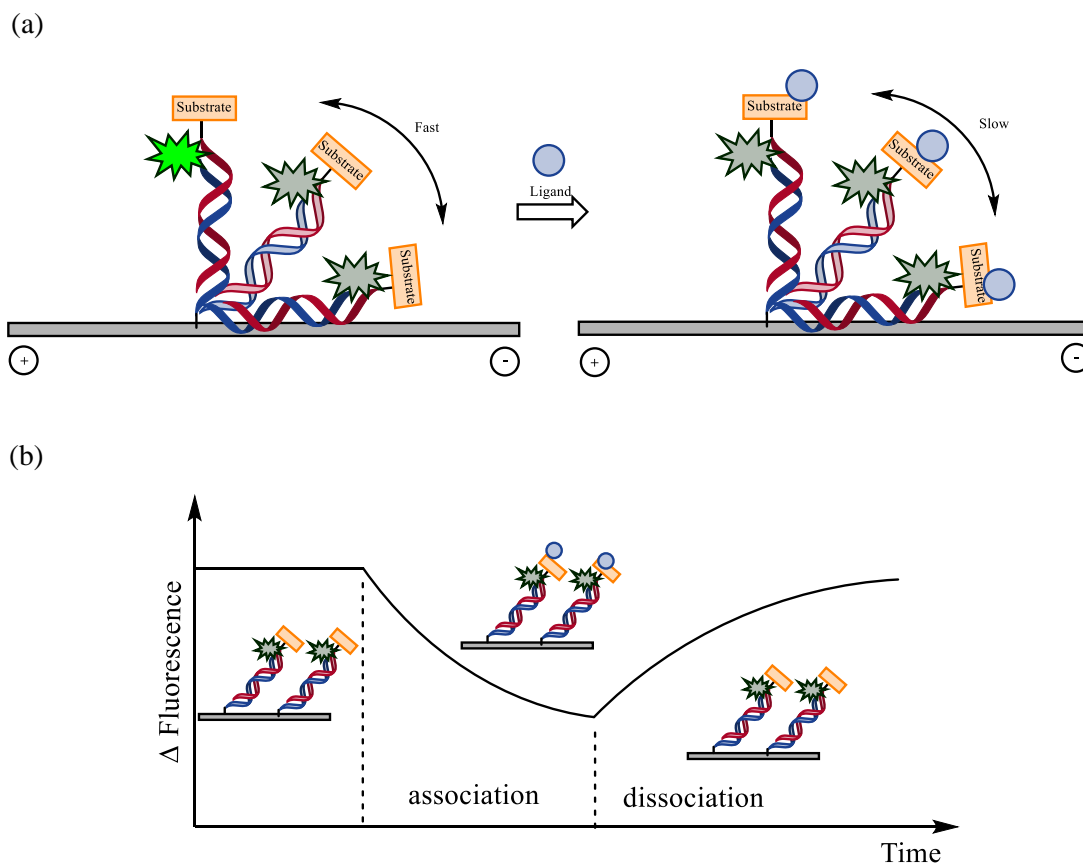


Figure 3.10 (a) Schematic representation of the SwitchSENSE[®] experiment. (b) Typical fluorescence response obtained in the SwitchSENSE[®] experiment.

3.2.7 Circular Dichroism

Another important technique used to assess general MGBs·dsDNA interactions is circular dichroism (CD).^{208,219–222} In CD experiments, circularly polarised light is applied to optically active compounds, which absorb right- and left-handed polarised light unequally, due to the coupling between non-orthogonal magnetic and electric transition moments. This is measured in ellipticity (ψ) (Figure 3.11). Most MGBs are not chiral and therefore are CD silent. However, upon interactions with dsDNA, they become part of an optically active entity, resulting in induced CD signals (ICD).²¹⁹ This change in intensity upon ligand titration at wavelengths that are not influenced by dsDNA signal can be plotted and used for binding affinity determination.²¹⁹ Most importantly, CD offers structural information about the binding

mechanism. The intensity and sign of the ICD are dependent on an intricate combination of electronic couplings between ligand and dsDNA electric and magnetic transition moments.²²³ While understanding the signal arising from intercalating agents requires further knowledge of the directionality of transition moments of the ligand in relation to the dsDNA pseudo dyad axis, a common trend is normally observed for MGBs that share complementary shape with the minor groove of dsDNA. For example, ICD arising from DNA ligands that direct their transition moments along the groove (*i.e.*, with an inclination of ca 45° to the bases) has high positive intensity (Figure 3.11). Furthermore, the shape of the signal can also be used to predict dimeric binding and higher-ordered structures. For example, a bis-signate signal has been observed for compounds binding in a 2:1 MGB·dsDNA ratio or hairpin conformation.²¹⁹ This is a result of the close proximity of transition moments with equal or similar energy. Despite the lower resolution compared to other techniques used for structural determination (such as NMR and X-ray), CD is a relatively fast method for preliminary evaluation of the binding mode of novel compounds, and also provides information on the binding affinity.

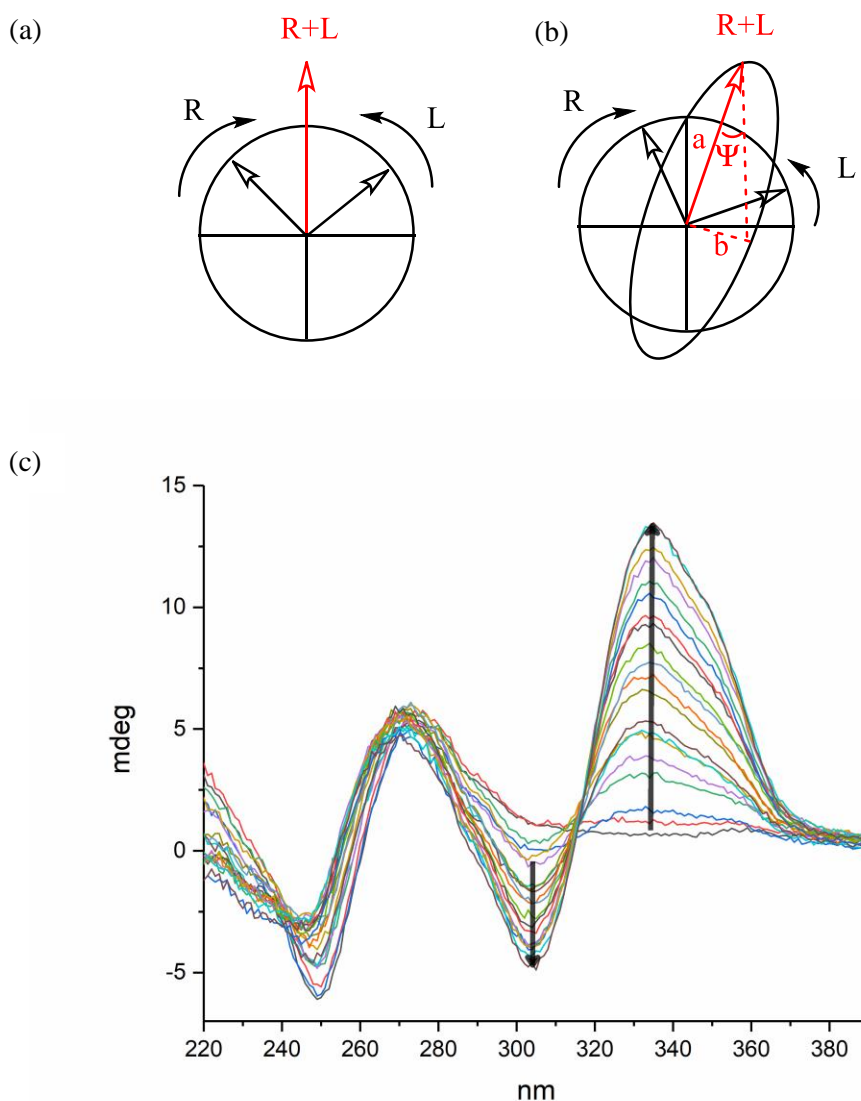


Figure 3.11 (a) Schematic representation of the plane of polarised light characterised by left- and right-handed polarised lights of equal intensity. (b) Schematic representation of ellipticity (ψ) resulting from a difference in absorption of the left and right circularly polarised light. (c) Example of the bis-signate CD spectra obtained upon titration of a hairpin MGB to the target dsDNA sequence.²¹⁹

3.3 Aims and objectives of this chapter

The objective of this chapter is to determine the binding affinities and selectivities of 8-ring hairpin polyamides containing Nt units in positions 6 and 8 and to compare them to their

cognate Im-containing counterparts. To ease the description of the data, polyamides synthesised in Chapter 2 are renamed as outlined in Figure 3.12.

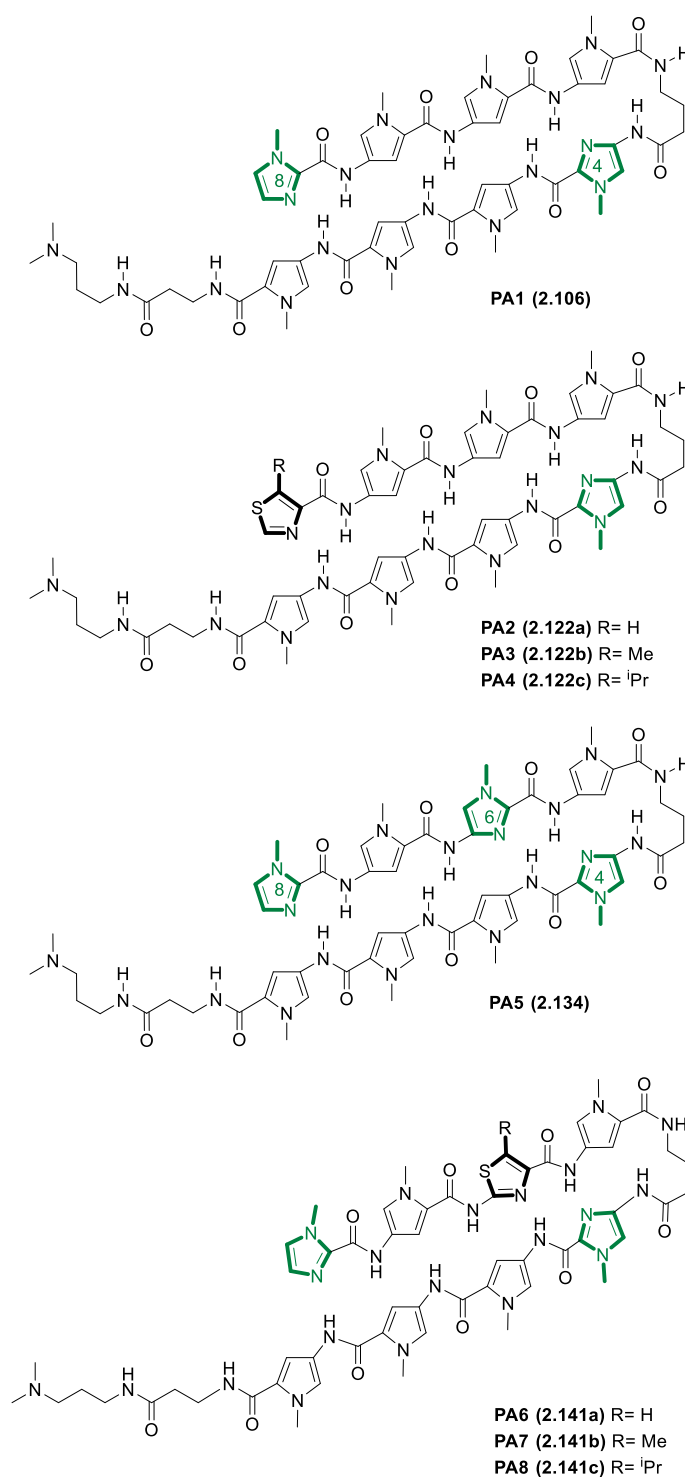


Figure 3.12 Structures of polyamides **PA1-8** as referred to in this chapter.

The specific aims are as follows:

- (i) Qualitative evaluation of the dsDNA binding affinity and selectivity of selected polyamides using DNase I footprinting and UV melting stabilisation.
- (ii) Defining the thermodynamic and kinetic parameters of dsDNA binding of selected polyamides using SwitchSENSE[®].
- (iii) Preliminary structural characterisation of selected polyamides using CD.

3.4 Results and Discussion

3.4.1 Binding characteristics of polyamides containing Nt in position 8

3.4.1.1 UV melting stabilisation

The melting stabilisations of **PA1-4** were determined for four dsDNA fragments, each containing a putative target sequence where the novel building block is tested against the four possible base pairs (5'-ATXTACT, where **X** = G, C, A or T; Figure 3.13). A 19.8 °C stabilisation was observed for the Im-containing **PA1** in the presence of the target sequence 5'-ATGTACT, with a clear binding preference for G at this position (Figure 3.13-14, Table 3.1). Enhanced thermal stabilisation for all three Nt-containing polyamides (**PA2-4**) relative to **PA1** was observed for 5'-ATGTACT, with **PA2** exhibiting the highest degree of thermal stabilisation of the polyamide series (Figure 3.13-14, Table 3.1). However, the Nt-containing polyamides are compromised in their specificity for G nucleotides. In general, the 5'-ATTTACT sequence was well tolerated by **PA2-4**. The overall binding selectivity was estimated using $\Delta\Delta T_m$ values that were calculated for each polyamide relative to the thermal stabilisation observed for 5'-ATGTACT (Table 3.2). **PA2** induced similar stabilization for the target sequences 5'-ATGTACT and 5'-ATTTACT with a $\Delta\Delta T_m$ of 0.8 °C. Surprisingly, **PA3** generated higher stabilisation for 5'-ATTTACT compared to 5'-ATGTACT ($\Delta\Delta T_m$ -0.5 °C). The sequence discrimination of **PA4** was higher than **PA2-3**, with a $\Delta\Delta T_m$ of 1.5 °C

stabilisation for 5'-ATGTACT relative to 5'-ATTTACT, suggesting that the larger isopropyl substitution influences the binding selectivity.

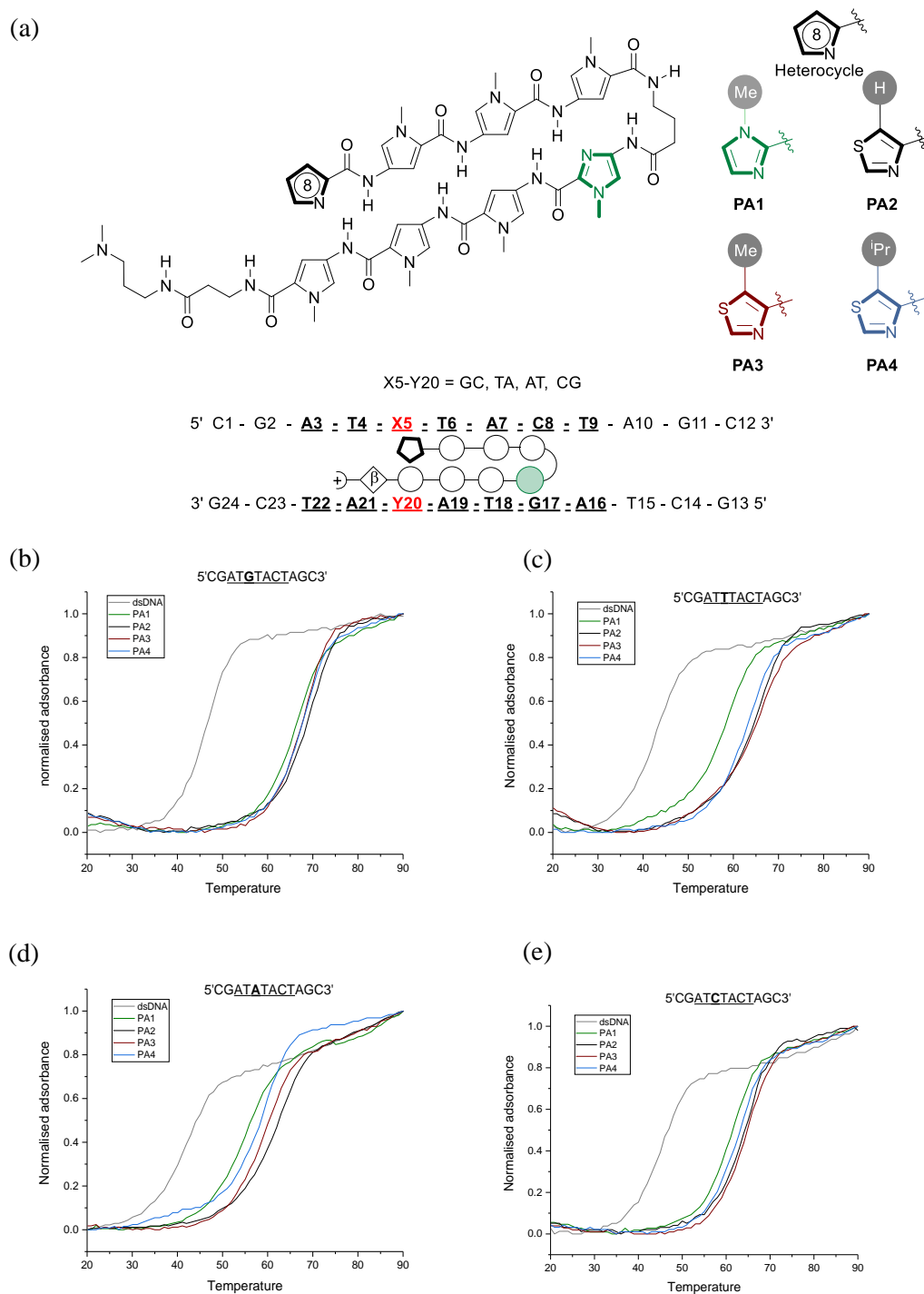


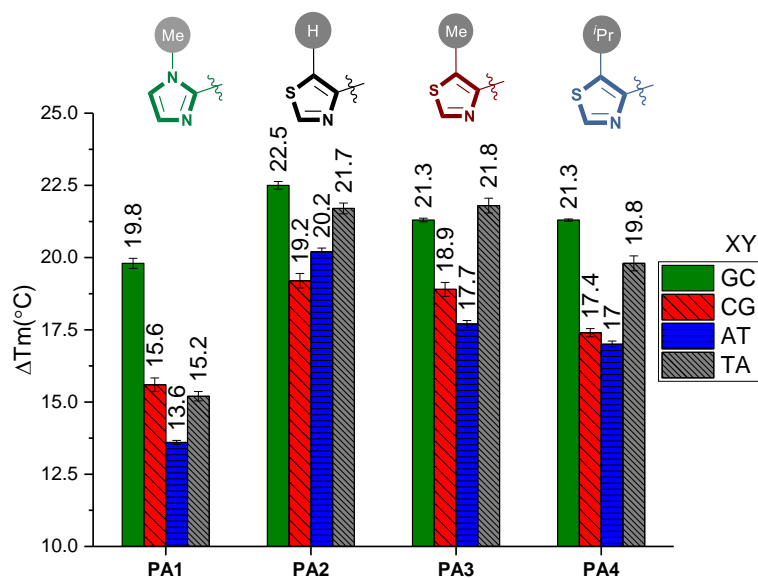
Figure 3.13 (a) Schematic representation of the polyamides interacting with the four alternative dsDNA sequences. Melting curves of **PA1-4** binding to (b) 5'-AT**G**TACT, (c) 5'-AT**T**TACT, (d) 5'-AT**A**TACT and (e) 5'-AT**C**TACT.

Table 3.1 T_m analyses of PA1-4 in complex with target dsDNA sequences.

	5'CGAT G TACTAGC3'		5'CGAT T TACTAGC3'	
	T_m (°C)	ΔT_m	T_m (°C)	ΔT_m
Free DNA	46.1±0.1	-	42.7±0.1	-
PA1	66.0±0.1	19.8±0.2	57.9±0.1	15.2±0.2
PA2	68.6±0.1	22.5±0.1	64.4±0.1	21.7±0.2
PA3	67.4±0.1	21.3±0.1	64.4±0.2	21.8±0.3
PA4	67.4±0.1	21.3±0.1	62.4±0.1	19.8±0.3

	5'CGAT A TACTAGC3'		5'CGAT C TACTAGC3'	
	T_m (°C)	ΔT_m	T_m (°C)	ΔT_m
Free DNA	41.7±0.1	-	45.1±0.1	-
PA1	55.2±0.1	13.6±0.1	60.7±0.1	15.6±0.2
PA2	61.9±0.1	20.2±0.1	64.3±0.1	19.2±0.2
PA3	59.4±0.1	17.7±0.1	64.0±0.2	18.9±0.3
PA4	58.7±0.1	17.0±0.1	62.5±0.1	17.4±0.1

Average values ± standard errors obtained from three experiments

**Figure 3.14** UV-melting stabilisation of PA1-4.**Table 3.2** $\Delta\Delta T_m$ values of PA1-4.

	5'-AT G TACT	5'-AT T TACT	5'-AT A TACT	5'-AT C TACT
PA1	-	4.6	6.2	4.2
PA2	-	0.8	2.3	3.3
PA3	-	-0.5	3.6	2.4
PA4	-	1.5	4.3	3.9

$$\Delta\Delta T_m = \Delta T_m(5\text{-ATG}TACT) - \Delta T_m(5\text{-ATX}TACT)$$

3.4.1.2 Influence of polyamide aggregation on the sample preparation for UV melting analysis

The sample solutions used in the melting experiments were prepared with an excess of the polyamide compared to dsDNA (1.75:1 ratio). This was necessary to ensure binding saturation and avoid double sigmoidal profiles in the melting plot (Figure 3.15). This was in contrast to previously reported conditions, where an exact 1:1 stoichiometry was used to produce a singular melting transition.¹⁰⁸

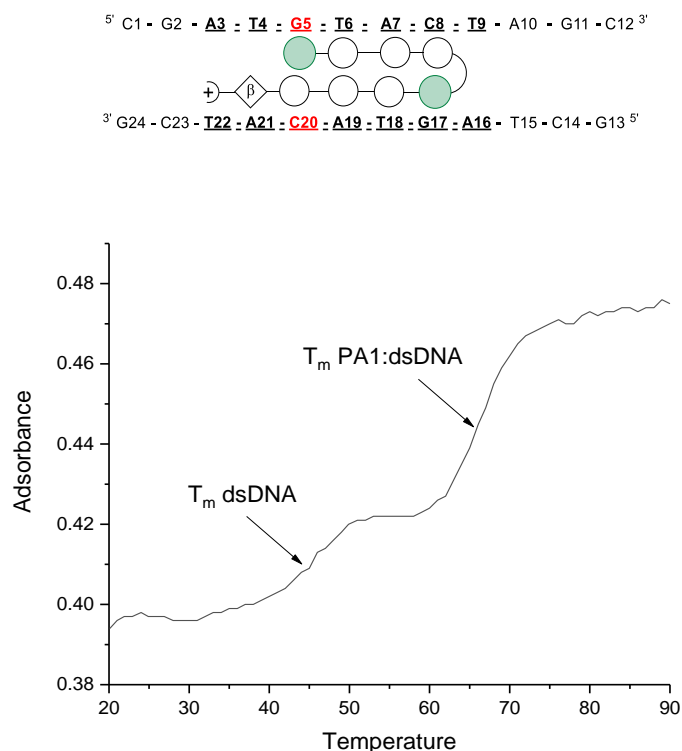


Figure 3.15 Typical double-sigmoidal trend observed when a 1:1 **PA1**:dsDNA ratio is used.

One possible explanation for this difference is due to the method used for the quantification of the polyamide stock solutions. Several works determined the polyamide concentration using an empirical extinction coefficient calculated by adding $8690 \text{ M}^{-1} \text{ cm}^{-1}$ for each aromatic ring present in the polyamide scaffold.^{108,139,146,210} This protocol is common as a limited amount of polyamide is often obtained from challenging synthetic procedures, which makes it difficult

to determine extinction coefficients experimentally. However, a wide range of extinction coefficient values has been observed when they were experimentally calculated in a variety of solvent mixtures and buffers, indicating the variability of this measurement.¹⁷⁵ Since the syntheses of **PA1-4** are high-yielding, we possessed enough material to prepare the stock solution by weighing the compounds and diluting with water to a final concentration of 100 μM . The inaccuracy of the empirical extinction coefficient was further confirmed when it was used to back-calculate the concentration of the stock solution, which in each case resulted in a 1-2-fold more dilute solution as compared to the experimentally-determined concentration. This explained the difference in stoichiometry between our experiment and the literature. Therefore, assuming our method of quantification to be more precise, it was hypothesised that the aggregation of the polyamides would explain the need for an excess of ligands to reach binding saturation. Evidence for aggregation in water was found by comparing 1D $^1\text{H-NMR}$ spectra of **PA1** and **PA3** in DMSO-d_6 and D_2O (Figure 3.16).

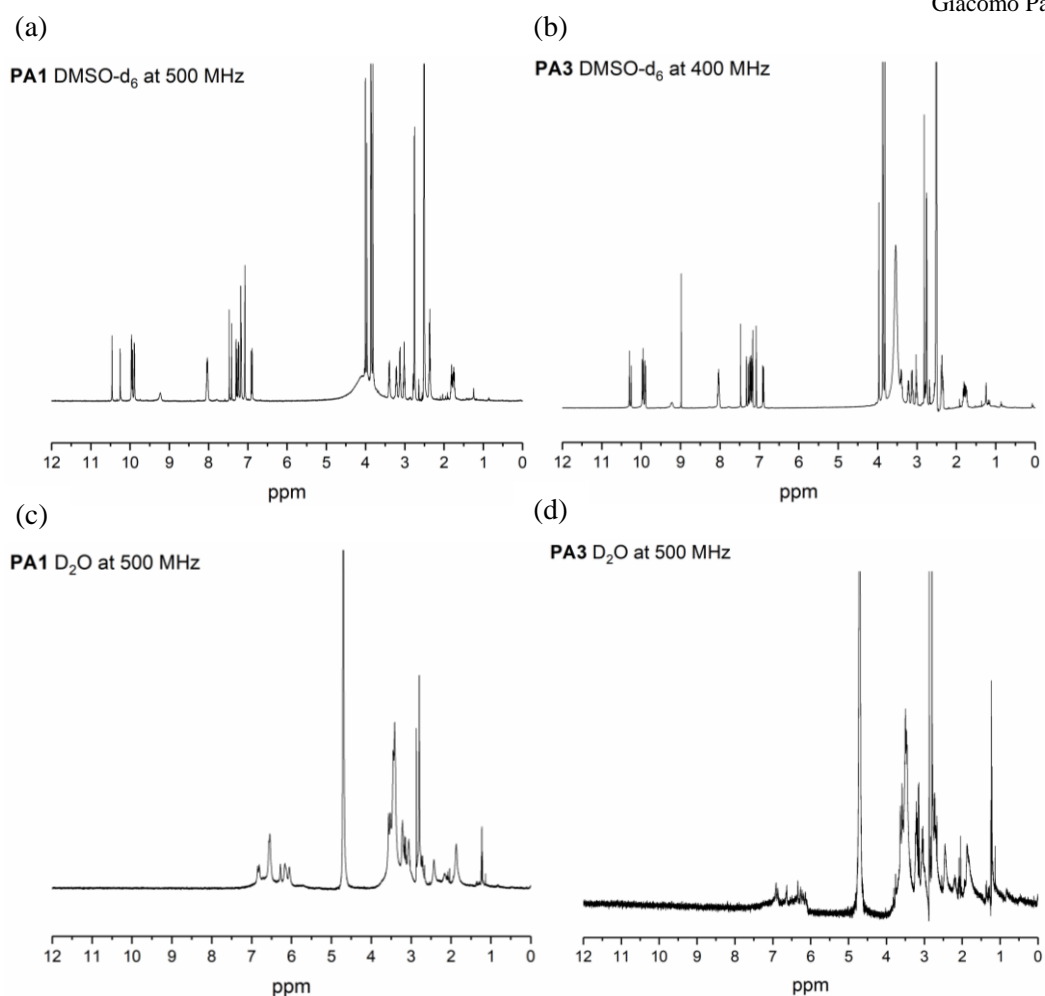


Figure 3.16 1D $^1\text{H-NMR}$ of (a) **PA1** in DMSO-d_6 . (b) **PA3** in DMSO-d_6 . (c) **PA1** in D_2O (2 mM). (d) **PA3** in D_2O (2 mM).

The broadening of the signals observed for D_2O solutions compared to the DMSO-d_6 spectra suggested that both the polyamides aggregate in D_2O . The linewidths were somewhat less defined for **PA3** compared to **PA1** at the same concentration in D_2O , suggesting that the presence of Nt increases the tendency of the polyamide to aggregate. Evidence for aggregation was also found when reproducing the melting experiment. In fact, an increased amount of ligand (up to 4 equivalents compared to dsDNA) from the same stock solution was necessary to saturate the binding site and reproduce the same melting plot after 24-72 hours. This is likely a result of the dynamic nature of the aggregation process. This was confirmed by the enhanced line broadening that was observed when repeating the NMR experiment after

leaving the sample in D₂O for seven days (Figure 3.17). Taken collectively, these observations suggest that both Nt and Im-containing polyamides slowly aggregate in water and, therefore a higher concentration of polyamide is necessary to saturate the dsDNA binding site. This finding is consistent with previous observations used to justify the impossibility of ITC measurements for 8-ring hairpin polyamides, which were therefore not attempted for **PA1-8**.¹⁹⁷

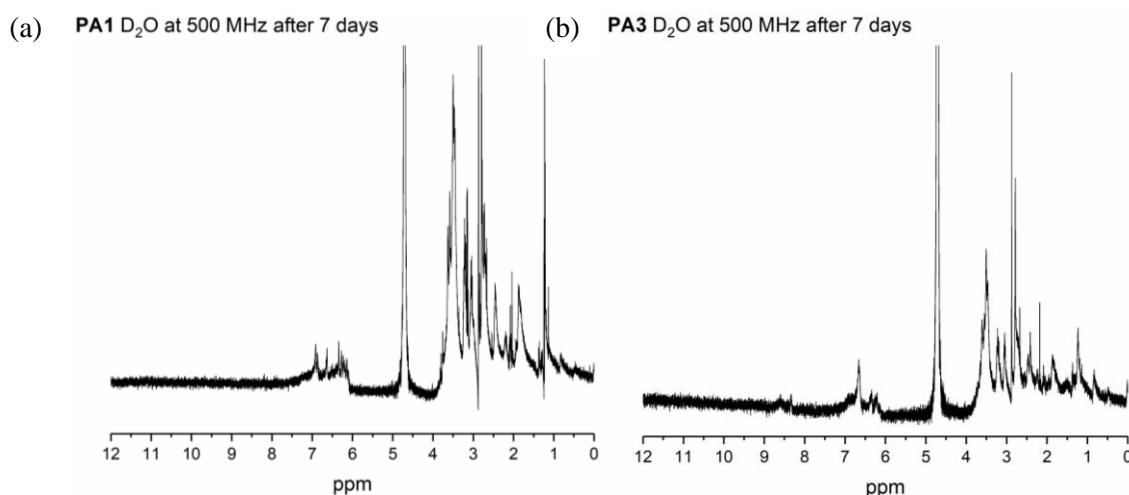


Figure 3.17 1D ¹H-NMR of (a) **PA1** (b) **PA3** in D₂O after 7 days.

3.4.1.3 DNase I footprinting

Further insight into binding selectivity was obtained from DNase I footprinting performed by Professor Keith Fox at the University of Southampton. In this experiment, **PA1-4** were titrated against radiolabelled DNA fragments containing a 75 base-pair insert with variants of the putative binding sites (Figure 3.18). As expected from UV melting studies, **PA1** revealed a bias towards binding to 5'-ATGTACT (Figure 3.18). Similarly, **PA4** exhibited a clear preference for 5'-ATGTACT. In contrast, **PA2** and **PA3** manifested more promiscuous DNA binding profiles, characterised by footprints at 5'-ATTTACT, 5'-ATATACT and 5'-ATGTACT. These results are consistent with UV melting studies and confirm the role of the isopropyl group in defining the binding affinity.

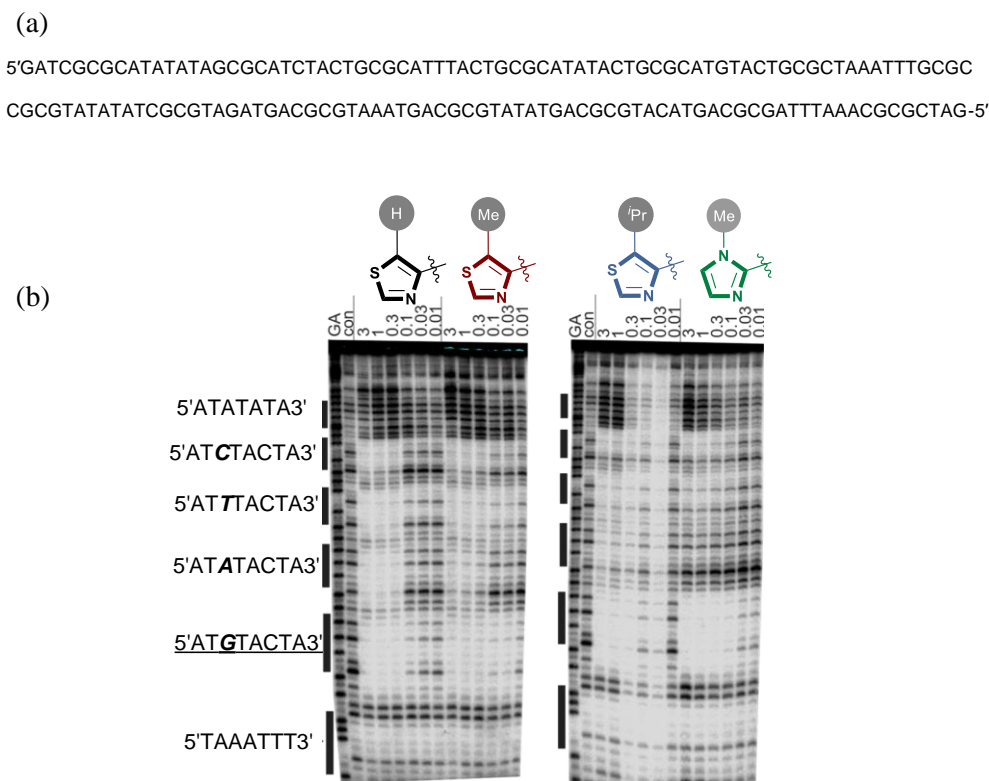


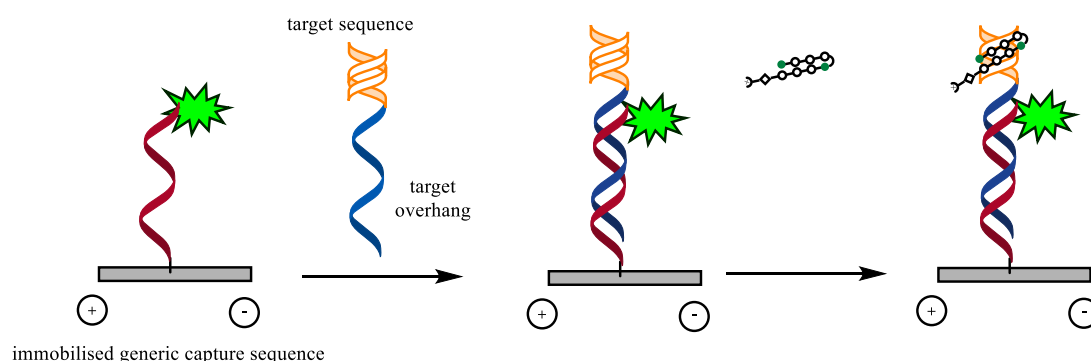
Figure 3.18 (a) Sequence cloned into the BamHI site of pUC19 to generate the footprinting fragments. (b) DNase I footprint titration using an engineered DNA 75 bp fragment in the presence of **PA1-4** at concentrations [3-0.01 μ M].

3.4.1.4 *SwitchSENSE[®]* for measurement of binding kinetics and thermodynamics

In order to gain further knowledge of the binding behaviours of Nt-containing polyamides (i.e., kinetics and thermodynamics), an experiment based on SwitchSENSE[®] technology was designed in collaboration with Dr Thomas Welte (DynamicBiosensors). This study focused primarily on the comparison between the G-selective **PA1** and **PA4**, therefore only **PA3** among the more promiscuous binders **PA2-3** was examined as a control.

The basic design of the experiment uses an immobilised, generic fluorophore-labelled ssDNA lever to capture a dsDNA fragment containing the polyamide target sequence via annealing to its complementary 3' overhang (Figure 3.19).

(a)



(b)

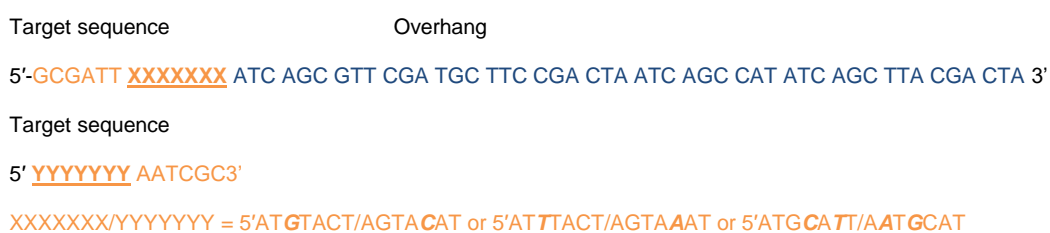


Figure 3.19 (a) Schematic representation of the SwitchSENSE[®] experimental design. (b) Sequences of the target complementary and overhang sequences used in the experiment.

At first, various concentrations of the polyamide (450 pM-25.5 nM) were injected into a continuous flow system (1000 μ l/min) for 10 minutes to be analysed. The change of the hydrodynamic friction of the DNA duplex induced upon binding of the polyamide was measured by the fluorescence variation relative to the unbound DNA, providing the association rate constant (k_{on}). The injected solution was then replaced by buffer and the fluorescence was monitored until the initial level was achieved to provide the dissociation constant (k_{off}). Monitoring the melting temperatures of bound and unbound sequences via fluorescence provided a sigmoidal trend similar to those observed in the UV melting experiments. The analyses were carried out on the match sequence 5'ATGTACT, the single mismatch 5'ATTTACT - which is tolerated by PA3 - and the double mismatch 5'ATGCATT, which should not be recognised by any of the polyamides.

The calculated binding dissociation constants (K_D) of the Im-containing PA1 for the three dsDNA sequences ranked as expected (Figure 3.20). No relevant fluorescence response was

measured for the double mismatch 5'ATGCATT, suggesting a lack of interaction of **PA1** with this sequence. A slower association and faster dissociation of **PA1** were observed for the single mismatch 5'ATTTACT ($k_{\text{on}} 5.37 \cdot 10^6 \text{ M}^{-1}\text{s}^{-1}$, $k_{\text{off}} 7.07 \cdot 10^{-3} \text{ s}^{-1}$) compared to the match 5'ATGTACT ($k_{\text{on}} 7.49 \cdot 10^6 \text{ M}^{-1}\text{s}^{-1}$, $k_{\text{off}} 1.91 \cdot 10^{-3} \text{ s}^{-1}$), resulting in an overall 5.2- fold higher K_D than the match sequence, 5'ATGTACT, confirming the sequence selectivity of the archetypal Im-containing **PA1**.

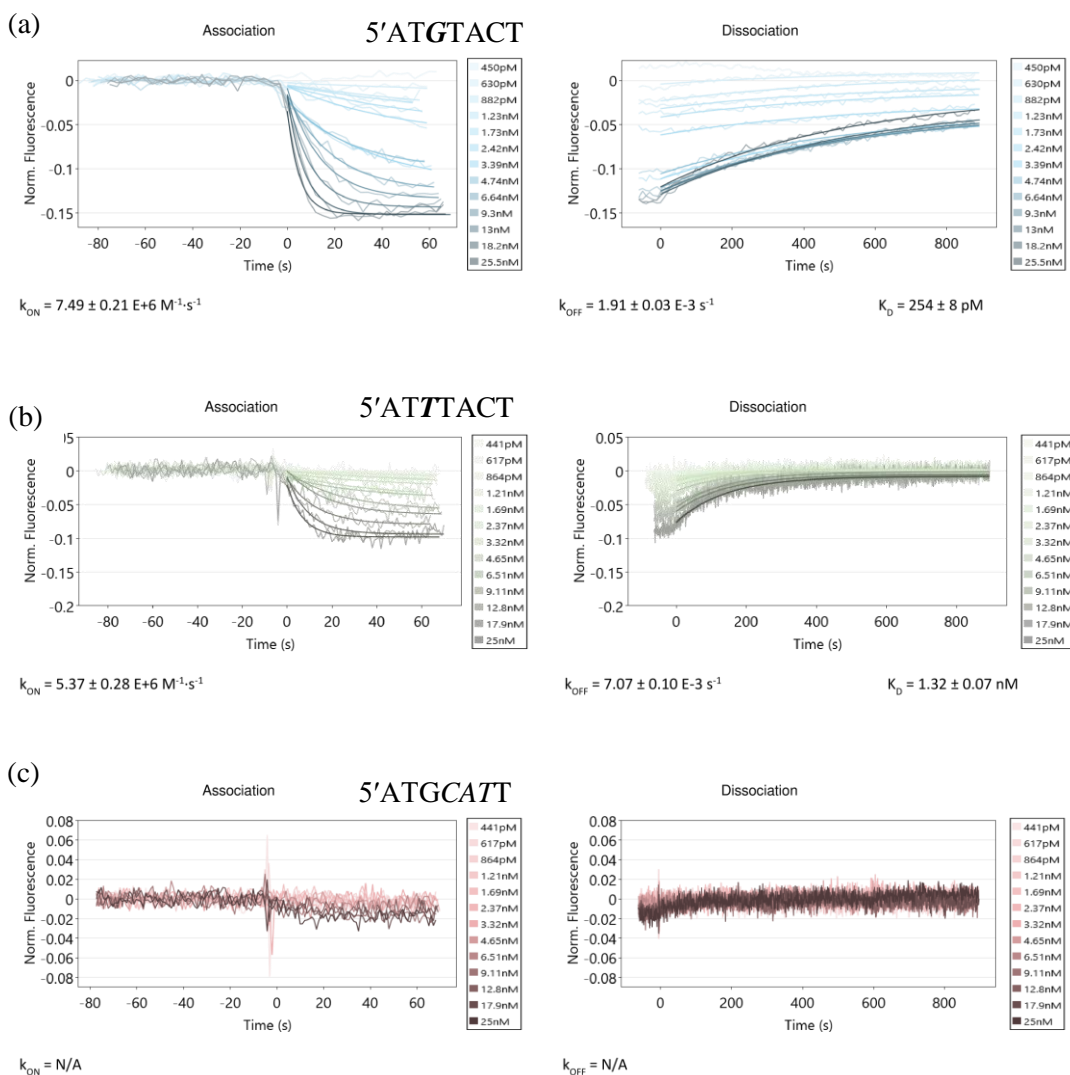


Figure 3.20 Binding curves of **PA1** obtained for (a) the match 5'ATGTACT, (b) the single mismatch 5'ATTTACT and (c) the double mismatch 5'ATGCATT.

As for the UV melting experiments, **PA1** generated higher ΔT_m for the match 5'ATGTACT (5.1 °C) compared to the single mismatch 5'ATTTACT (2.0 °C) (Figure 3.21a-b). No thermal stabilisation was observed for the double mismatch 5'ATGCATT (Figure 3.21c).

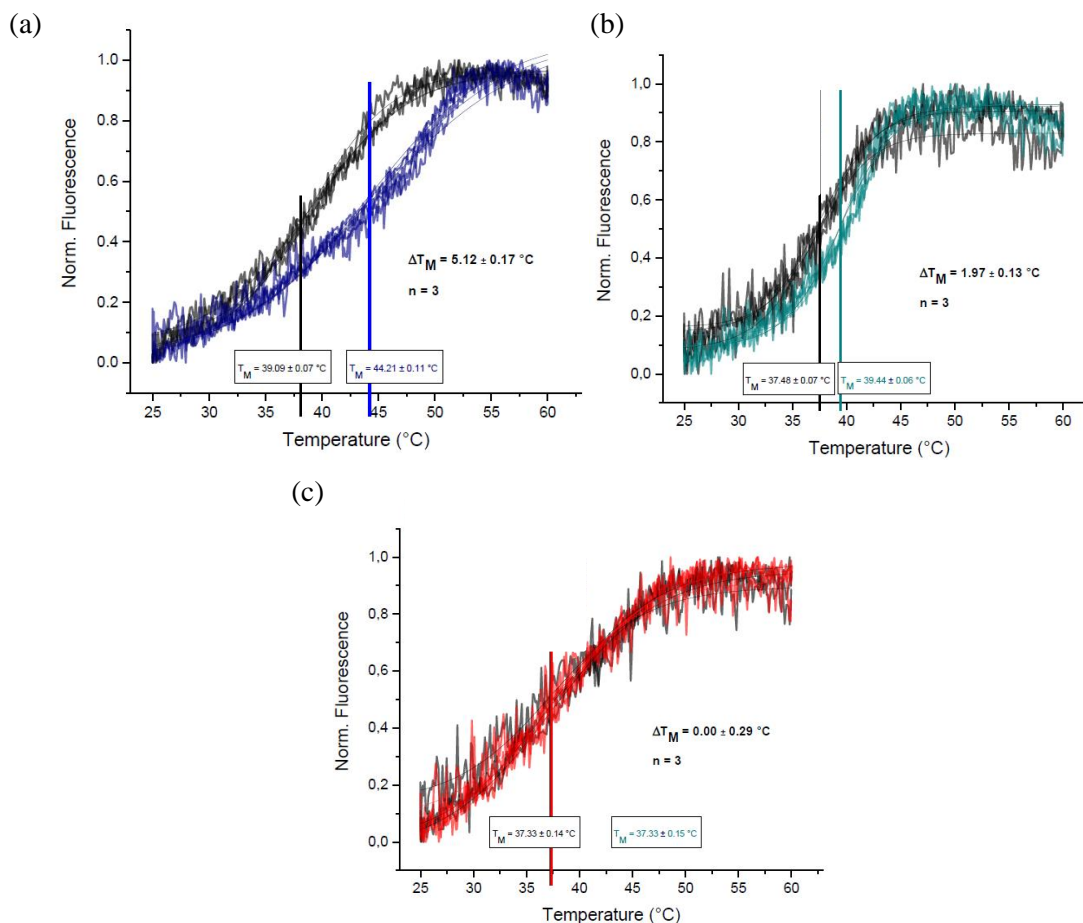


Figure 3.21 Fluorescence melting curves of **PA1** obtained for (a) the match 5'ATGTACT, (b) the single mismatch 5'ATTTACT and (c) the double mismatch 5'ATGCATT.

These analyses were repeated for **PA3** and **PA4** (Figure 3.22, Figure 3.24). In agreement with DNase I and UV melting analysis, **PA3** exhibited similar binding affinities for the single mismatch 5'ATTTACT ($K_D = 1.25$ nM) and the match 5'ATGTACT ($K_D = 1.17$ nM) (Figure 3.22a-b). This promiscuous binding, introduced by the presence of the Nt building block, was also confirmed by the observation of some weaker interactions with the double mismatch sequence 5'ATGCATT (Figure 3.22c).

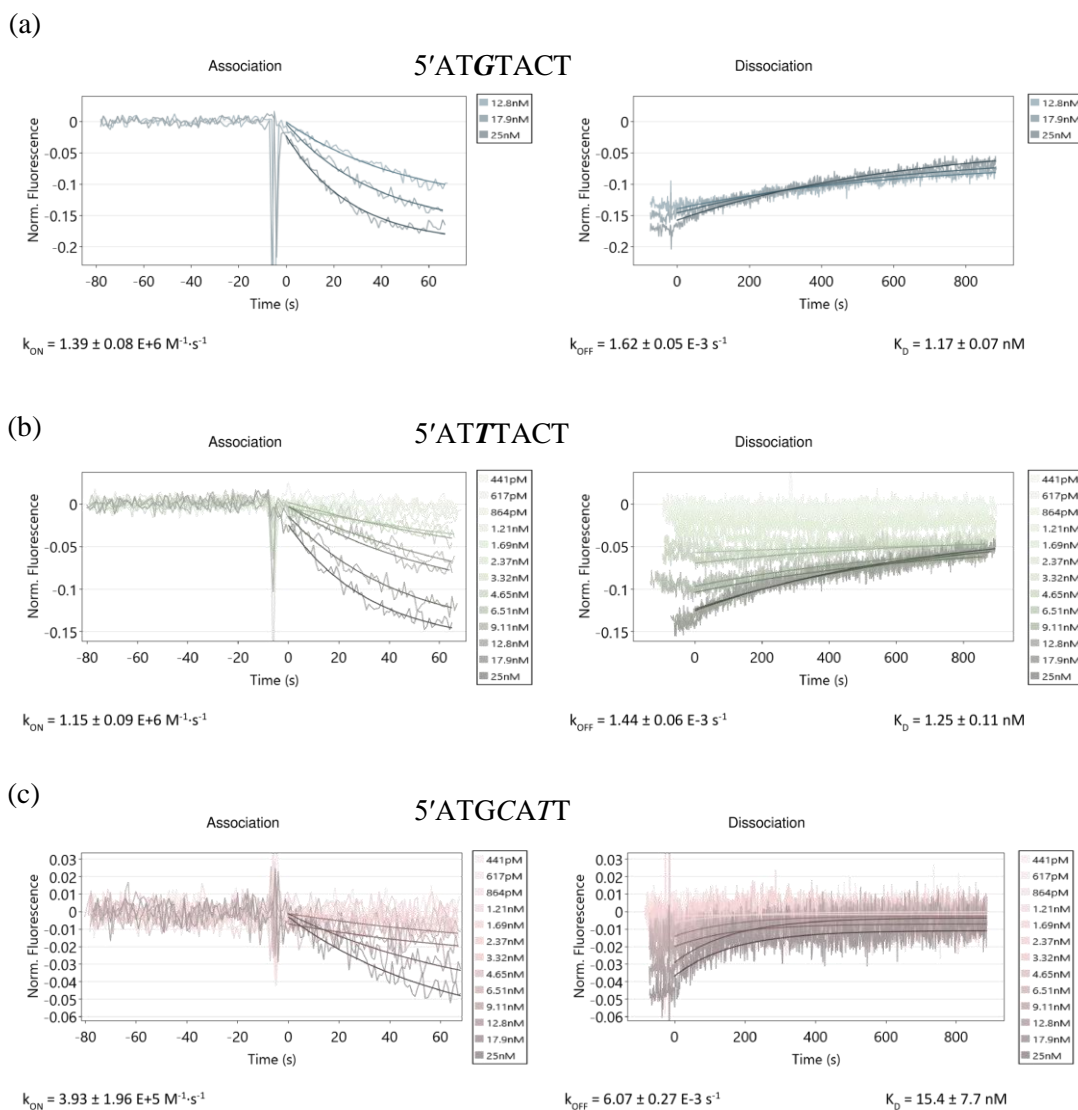


Figure 3.22 Binding curves of **PA3** obtained for (a) the match 5'ATGTACT, (b) the single mismatch 5'ATTACT and (c) the double mismatch 5'ATGCATT.

Similar melting stabilisations by **PA3** were observed for the match 5'ATGTACT and mismatch 5'ATTACT sequences; however, in this case, the higher stabilisation was measured for 5'ATGTACT (6.6 °C) compared to 5'ATTACT (5.9 °C) (Figure 3.23a-b). As expected lower stabilisation was measured for the double mismatch 5'ATGCATT (2.9 °C) (Figure 3.23c).

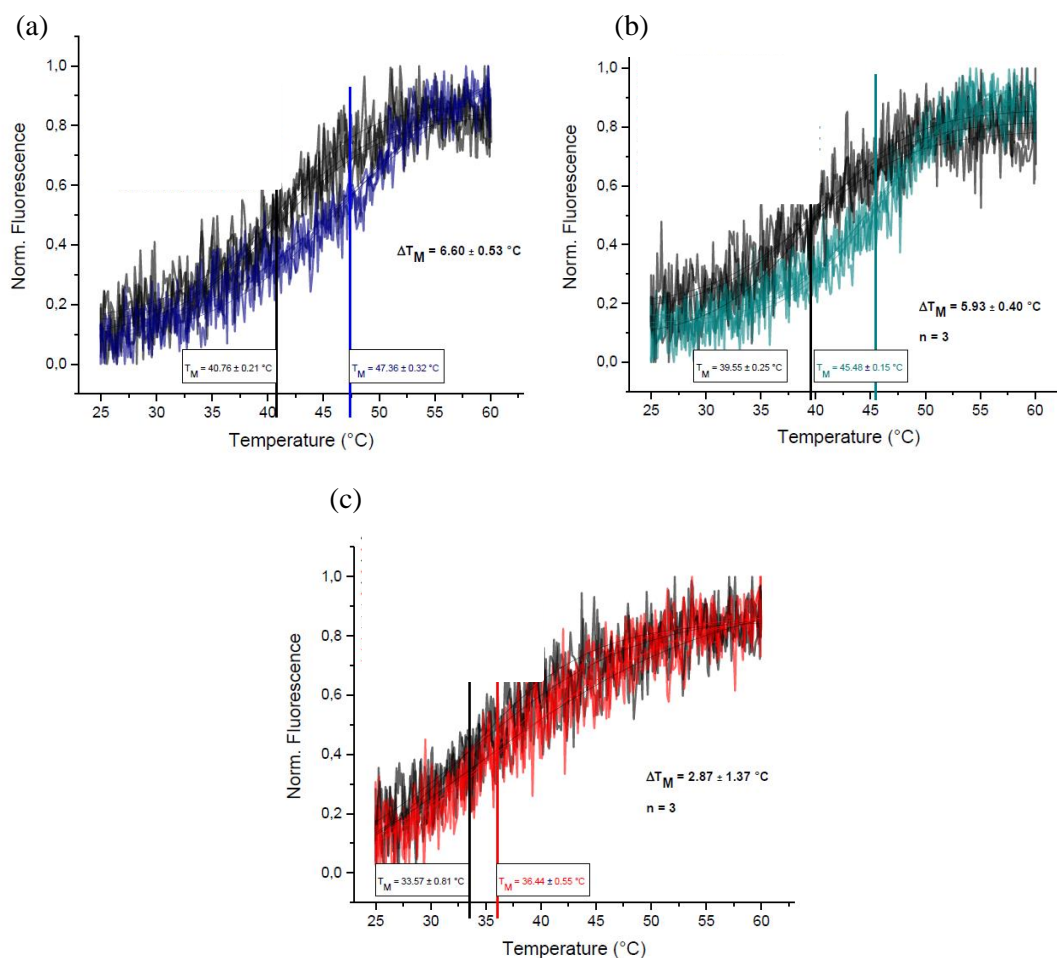


Figure 3.23 Fluorescence melting curves of **PA3** obtained for (a) the match 5'ATGTACT, (b) the single mismatch 5'ATTTACT and (c) the double mismatch 5'ATGCATT.

On the other hand, **PA4** only produced a measurable fluorescence response for the target sequence 5'ATGTACT ($K_D = 1.97$ nM) (Figure 3.24a). Unexpectedly, only a weak signal was observed for 5'ATTTACT (Figure 3.24b). This is in contrast with fluorescence and UV melting experiments, which suggested a relatively strong interaction for this sequence (5.8 °C and 19.8 °C, respectively) (Figure 3.14 and Figure 3.25b).

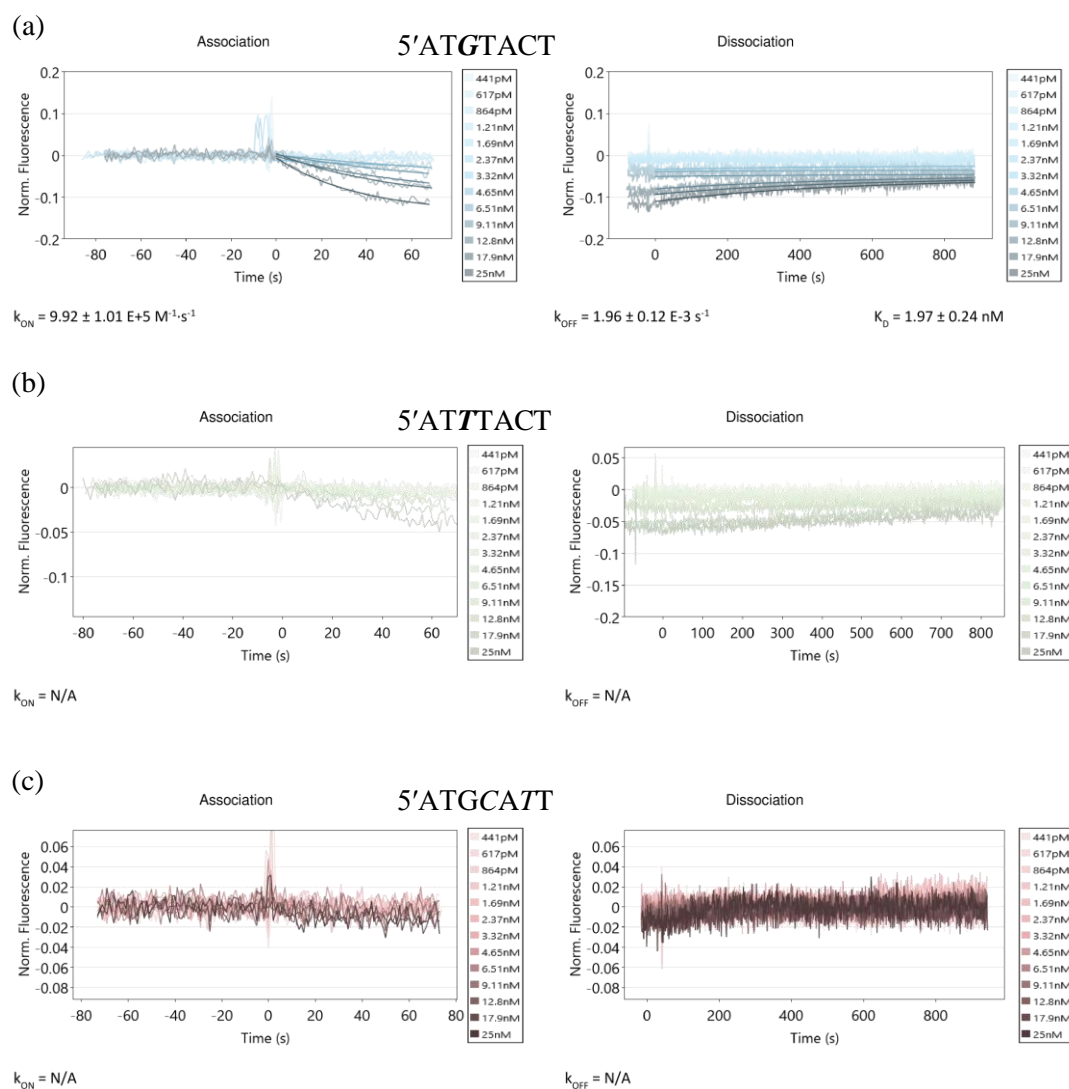


Figure 3.24 Binding curves of **PA4** obtained for (a) the match 5'ATGTACT, (b) the single mismatch 5'ATTACT and (c) the double mismatch 5'ATGCATT.

Consistent with the UV melting experiments, the highest melting stabilisation was observed for the match sequence 5'ATGTACT (6.4 °C), with no measurable stabilisation for the double mismatch 5'ATGCATT, confirming the G-selectivity of **PA4**.

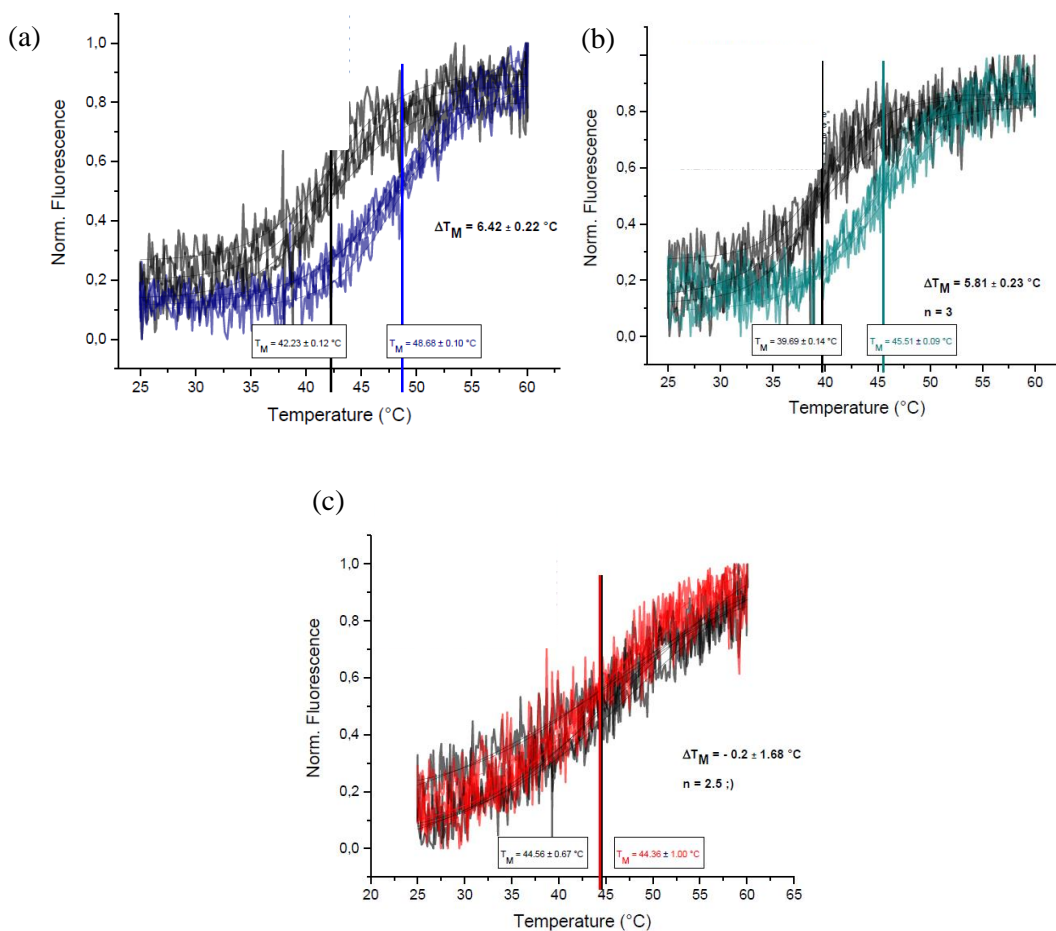


Figure 3.25 Fluorescence melting curves of **PA4** obtained for (a) the match 5'ATGTACT, (b) the single mismatch 5'ATTTACT and (c) the double mismatch 5'ATGCATT.

Taken collectively, these data confirm the binding selectivity observed by UV melting and DNase I footprinting. Furthermore, the direct comparison of the polyamide dissociation constants suggests that each polyamide possessed a unique binding signature. For instance, the lowest K_D measured for 5'ATGTACT was observed for **PA1** ($K_D = 0.254$ nM) (Table 3.3), suggesting the higher binding affinity of this polyamide for the target sequence. However, this is in contrast to what has been observed from UV and fluorescence thermal stabilisation experiments that indicated higher stabilisation for the Nt-containing **PA3** and **PA4**. This can possibly be explained by comparing the association rates. **PA1** has the highest k_{on} ($7.49 \pm 0.21 \times 10^6$ M⁻¹s⁻¹) (Table 3.3), indicating the fast formation of the complex. Increasing the

dimension of the ring (*i.e.*, **PA3**) and steric bulk of the substituent (*i.e.*, **PA4**) leads to slower association rate constants of $1.39 \pm 0.08 \times 10^6 \text{ M}^{-1}\text{S}^{-1}$ (Table 3.4) and $9.92 \pm 1.01 \times 10^5 \text{ M}^{-1}\text{S}^{-1}$ (Table 3.5), respectively. This inverse correlation between k_{on} and the dimensions might suggest that the larger ligands, such as **PA3-4**, use a different, slower mechanism to adapt to the shape of dsDNA. On the other hand, the higher melting stabilisation suggests that once the complex is formed, the binding is somewhat tighter for **PA3** and **PA4** compared to **PA1**. However, very similar dissociation rate constants were observed for the three polyamides binding to target sequence 5'ATGTACT. This effect is particularly relevant for the more encumbered **PA4** targeting the mismatch 5'ATTTACT, which despite producing only a very weak interaction that did not allow the calculation of the association and dissociation rate constants, resulted in a significant melting stabilisation (5.8 °C) (Table 3.5). This difference can also be explained by the propensity of the polyamides to aggregate. Because of the lower tendency of **PA1** to aggregate, more molecules would be available for interaction with dsDNA. As a consequence of the enhanced aggregation resulting from Nt incorporation, the association rate would be lower. Further experiments to quantify the aggregation rates of **PA1**, **PA3** and **PA4** will be necessary to verify these hypotheses.

Table 3.3 Overview of k_{on} , k_{off} , K_{D} and ΔT_{m} of **PA1**.

PA1	$k_{\text{on}} (\text{M}^{-1}\text{s}^{-1})$	$k_{\text{off}} (\text{s}^{-1})$	$K_{\text{D}} (\text{nM})$	ΔT_{m} at 20 °C
5'ATGTACT	$7.49 \pm 0.21 \times 10^6$	$1.91 \pm 0.003 \times 10^{-3}$	0.254 ± 8	5.12 ± 0.17
5'ATGCATT	No interaction	No interaction	No interaction	0.00 ± 0.29
5'ATTTACT	$5.37 \pm 0.28 \times 10^6$	$7.07 \pm 0.01 \times 10^{-3}$	1.320 ± 70	1.97 ± 0.13

Table 3.4 Overview of k_{on} , k_{off} , K_{D} and ΔT_{m} of **PA3**.

PA3	$k_{\text{on}} (\text{M}^{-1}\text{s}^{-1})$	$k_{\text{off}} (\text{s}^{-1})$	$K_{\text{D}} (\text{nM})$	ΔT_{m} at 20 °C
5'ATGTACT	$1.39 \pm 0.08 \times 10^6$	$1.62 \pm 0.05 \times 10^{-3}$	1.17 ± 0.07	6.60 ± 0.53
5'ATGCATT	$3.93 \pm 1.96 \times 10^5$	$6.07 \pm 0.27 \times 10^{-3}$	15.4 ± 7.7	2.87 ± 1.37
5'ATTTACT	$1.15 \pm 0.09 \times 10^6$	$1.44 \pm 0.06 \times 10^{-3}$	1.25 ± 0.11	5.93 ± 0.4

Table 3.5 Overview of k_{on} , k_{off} , K_{D} and ΔT_{m} of **PA4**.

PA4	$k_{\text{on}} (\text{M}^{-1}\text{s}^{-1})$	$k_{\text{off}} (\text{s}^{-1})$	$K_{\text{D}} (\text{nM})$	ΔT_{m} at 20 °C
5'ATGTACT	$9.92 \pm 1.01 \times 10^5$	$1.96 \pm 0.12 \times 10^{-3}$	1.97 ± 0.24	6.42 ± 0.22
5'ATGCATT	No interaction	No interaction	No interaction	-0.2 ± 1.68
5'ATTTACT	Weak interaction	Weak interaction	Weak interaction	5.81 ± 0.23

3.4.1.5 Circular Dichroism

Circular dichroism was used to gain insight into the structural features of **PA1-4** bound to dsDNA. The calculation of the binding affinity constant by CD was prevented by the aggregation of the polyamides, which were required in large excess compared to dsDNA, in order to reach saturation. For this reason, CD was used exclusively as a qualitative technique for the determination of the binding mode. The matched dsDNA sequence 5'ATGATCT was added with two different concentrations of **PA1**, **PA3** and **PA4** (1.5:1 and 5:1 ratio polyamide-dsDNA). In all cases, similar behaviours, consistent with minor groove binding (*i.e.*, positive peak at ca. 330 nm) were observed for all polyamides (Figure 3.26).²¹⁹ The negative peak at 303 nm was similar to previously observed MGBs, and is attributed to the hairpin conformation assumed by the polyamides in the minor groove.

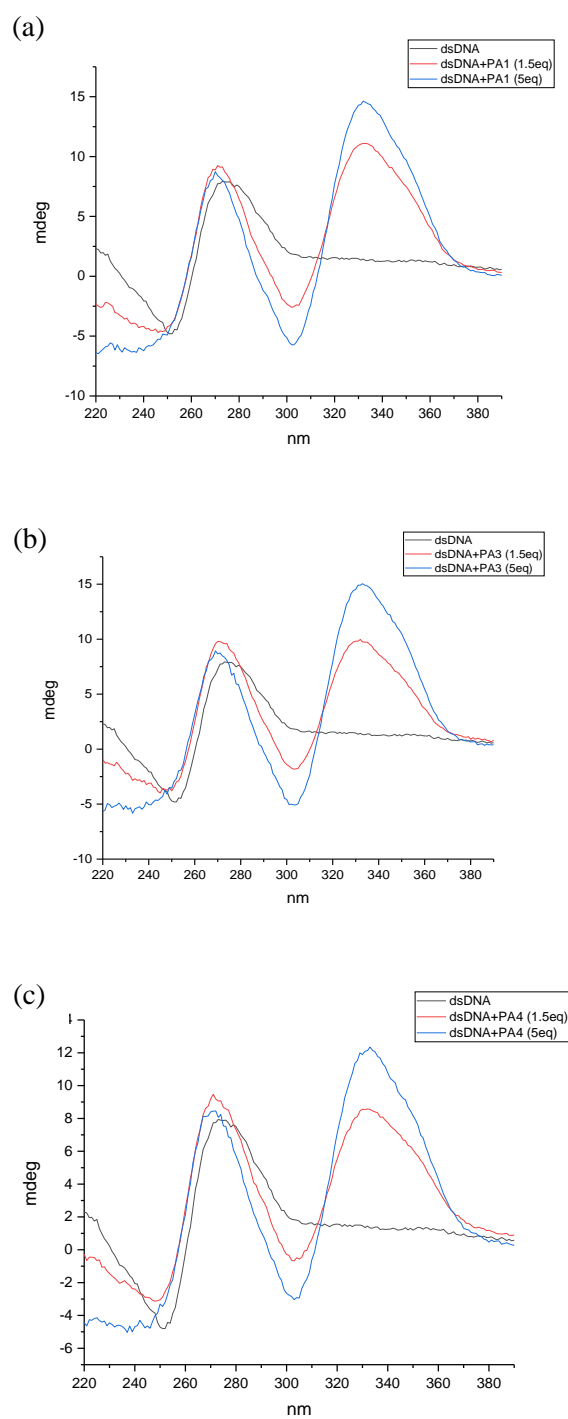


Figure 3.26 CD spectra of (a) **PA1**, (b) **PA3** and (c) **PA4** in complex with d(CGATGATCATGC)-(CGATGATCATCG).

Taken collectively, these data suggest that Nt-containing polyamides (**PA3-4**) bind to dsDNA via a similar binding mode to Im-containing polyamides.

3.4.2 Binding characteristics of polyamides containing Nt in position 6

3.4.2.1 UV melting stabilisation

The UV melting stabilisation induced by **PA5-8** was measured on four dsDNA sequences containing the potential target sites 5'ATGAXCT (where X=G/C/A/T) (Figure 3.27). Due to the lower quantity afforded by their synthesis, **PA5-8** were quantified using the empirically calculated extinction coefficient of $69200 \text{ M}^{-1}\text{cm}^{-1}$ at approximately λ_{max} of 310 nm. Saturation of the dsDNA binding site was observed in all cases upon addition of 1.25-1.5 equivalents of a freshly prepared solution of polyamide. In this series, the Im-containing **PA5** induced the highest melting stabilisation for 5'ATGAGCT, suggesting higher binding affinity compared to Nt polyamides (**PA6-8**) (Figure 3.28 and Table 3.6). **PA6**, which contained the least hindered $^{\text{H}}\text{Nt}$, produced the highest melting stabilisation for 5'ATGAGCT amongst the Nt polyamides, suggesting that the larger steric bulk in position 5 of Nt in **PA7-8** decreases the binding affinity towards dsDNA (Figure 3.28 and Table 3.6). In contrast to the previous set of polyamides, a clear discrimination for the target 5'ATGAGCT was observed for all the polyamides (*i.e.*, **PA5-8**, Table 3.7). Once again, the higher degree of selectivity was observed for the Im-containing **PA5**. Good selectivity was also shown by **PA6**. Despite generally having low binding affinities, G-selective binding was also observed for **PA7-8**.

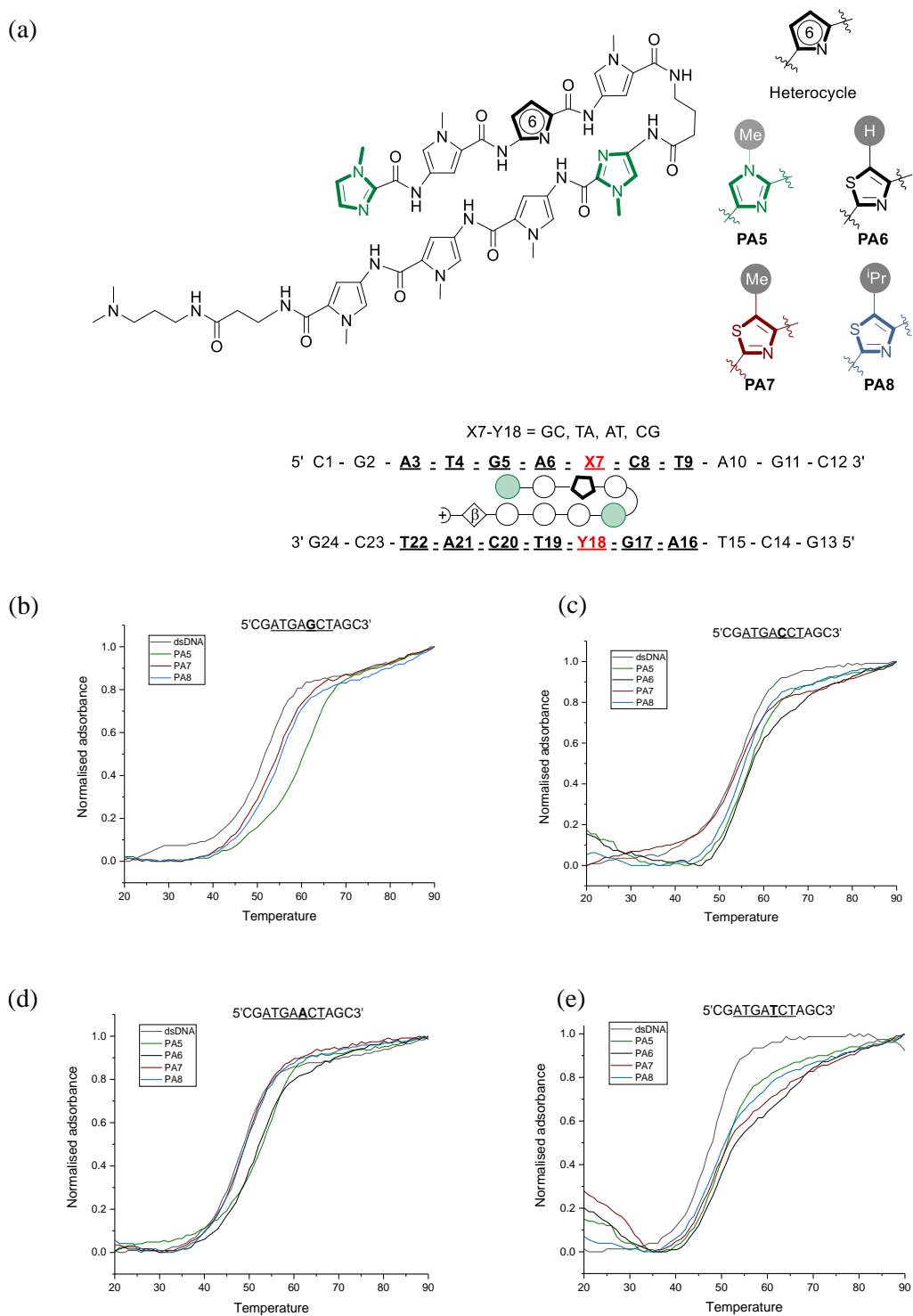
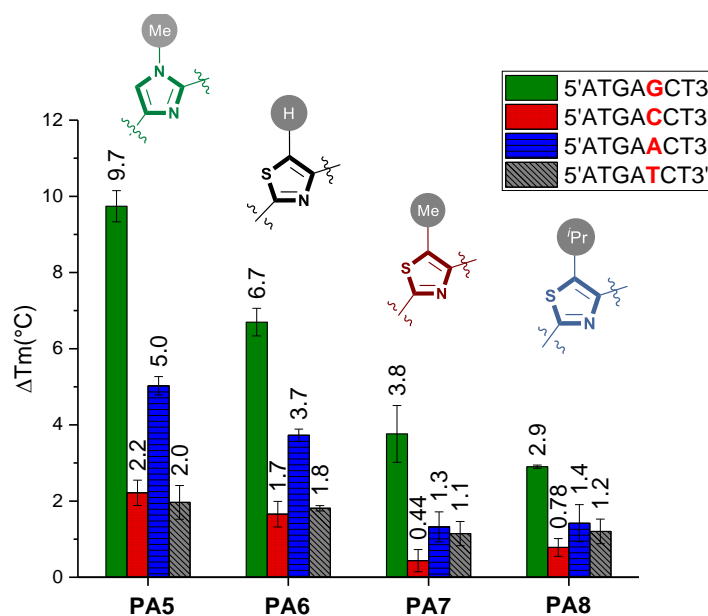


Figure 3.27 (a) Schematic representation of the polyamides interacting with the four alternative dsDNA sequences. Melting curves of **PA5-8** binding to (b) 5'ATGAGCT, (c) 5'ATGACCT, (d) 5'ATGAACT and (e) 5'ATGATCT.

Table 3.6 T_m analyses of PA5-8 in complex with target dsDNA sequences.

	5'CGATGAGCTAGC3'		5'CGATGACCTAGC3'	
	T_m (°C)	ΔT_m	T_m (°C)	ΔT_m
Free DNA	51.2±0.1	-	53.1±1.1	-
PA5	60.9±0.3	9.7±0.4	55.3±1.3	2.2±0.2
PA6	57.9±0.7	6.7±0.7	54.7±0.9	1.7±0.3
PA7	55.0±0.1	3.8±0.04	53.5±0.9	0.4±0.3
PA8	54.1±0.2	2.9±0.4	53.8±1.3	0.8±0.3

	5'CGATGA A CTAGC3'		5'CGATGA T CTAGC3'	
	T_m (°C)	ΔT_m	T_m (°C)	ΔT_m
Free DNA	47.7±0.2	-	47.4±0.3	-
PA5	52.7±0.3	5.0±0.2	49.3±0.7	2.0±0.4
PA6	51.4±0.2	3.7±0.2	49.2±0.3	1.8±0.1
PA7	49.0±0.2	1.3±0.4	48.5±0.6	1.1±0.3
PA8	49.1±0.3	1.4±0.5	48.6±0.6	1.2±0.3

Average values \pm x standard errors**Figure 3.28** UV-melting stabilisation of PA5-8.**Table 3.7** $\Delta\Delta T_m$ values of PA5-8.

	5'-ATGAGCT	5'-ATGACCT	5'-ATGA A CT	5'-ATGA T CT
PA5	-	7.5	4.7	7.7
PA6	-	5.0	3.0	4.9
PA7	-	3.4	2.5	2.7
PA8	-	2.1	1.5	1.7

$$\Delta\Delta T_m = \Delta T_m(5\text{'-ATG}T\text{ACT}) - \Delta T_m(5\text{'-AT}X\text{TACT})$$

UV thermal stabilisation analysis of **PA5-8** confirmed our initial hypothesis that Nt rings can function as G-selective building blocks, similarly to Im. However, the overall lower melting stabilisation observed for Nt-containing polyamides **PA6-8** suggests that

the incorporation of an Nt building block in an internal position of the polyamide affects the capability of the ligand to adapt to the shape of the minor groove. This was further shown by the very low stabilisation observed for the more sterically hindered **PA7** and **PA8**, which suggested that bulky substituents at position 5 of Nt heterocycles have a negative influence on dsDNA binding (Figure 3.29).

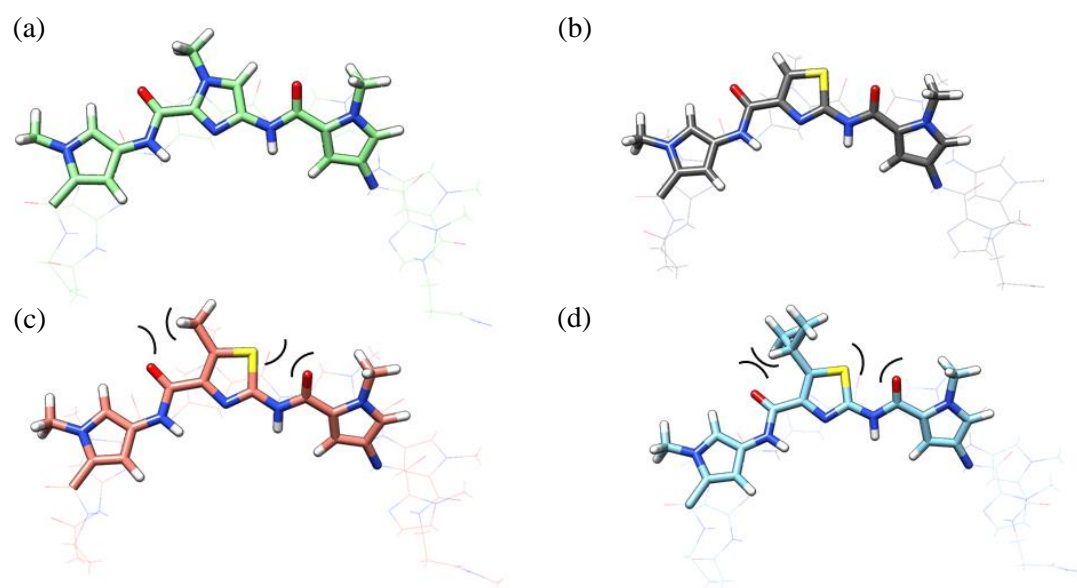


Figure 3.29 Depiction of energy-minimised structures of (a) **PA5**, (b) **PA6**, (c) **PA7** and (d) **PA8** using the GAFF force field (putative steric clashes affecting the DNA binding are highlighted for **PA7-8**).

Because of the limited binding affinity and the more challenging synthetic preparation, the analysis of **PA5-8** was restricted to these data.

3.5 Summary

In this chapter, the binding affinities and selectivities for dsDNA were established for two series of Nt-containing polyamides, in which the novel building block was incorporated in one of the two different positions within the scaffold. Two distinct binding behaviours were observed within these series (Figure 3.30). UV and fluorescence melting experiments of **PA2-4**, which incorporated Nt into position 8, indicated a higher affinity for dsDNA compared to

the Im analogue **PA1**, although this was at the expense of the sequence selectivity. In particular, the mismatch sequence 5'ATTTACT produced equal or higher melting stabilisation for **PA2-3** compared to the expected target 5'ATGTACT. Interestingly, the bulkier isopropyl substitution of **PA4** increases selectivity towards G with respect to **PA2-3**, suggesting the importance of the substitution in position 5 of Nt when determining the selectivity.

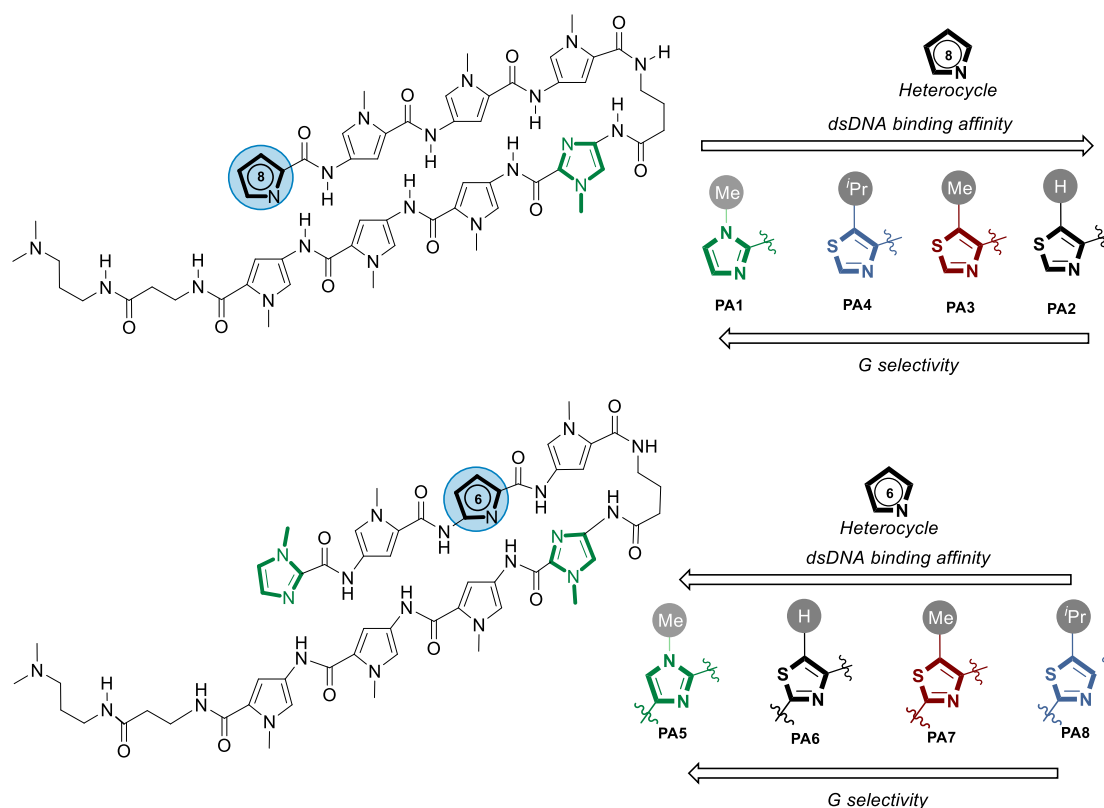


Figure 3.30 Schematic representation of the binding affinity and selectivity of **PA1-8**.

In contrast, weaker binding affinities were observed for **PA6-8** containing the Nt in position 6 compared to the Im-cognate **PA5**. In this case, the large steric bulk at the position 5 of Nt showed a negative influence on the binding, presumably because this prevents the optimal folding to match the curvature of the minor groove. However, in contrast to **PA2-4**, **PA6-8** retained G-selectivity.

This might suggest that the amide substitutions at either side of Nt more strictly direct Nt(N3) towards the groove, therefore assuming a more suitable conformation for hydrogen bonding with G(N2). On the other hand, it is possible that the increased binding affinity and lower selectivity of **PA2-4** is related to the enhanced flexibility of the terminal Nt, which can more easily adapt to the shape of the minor groove in different sequences of dsDNA (Figure 3.31).

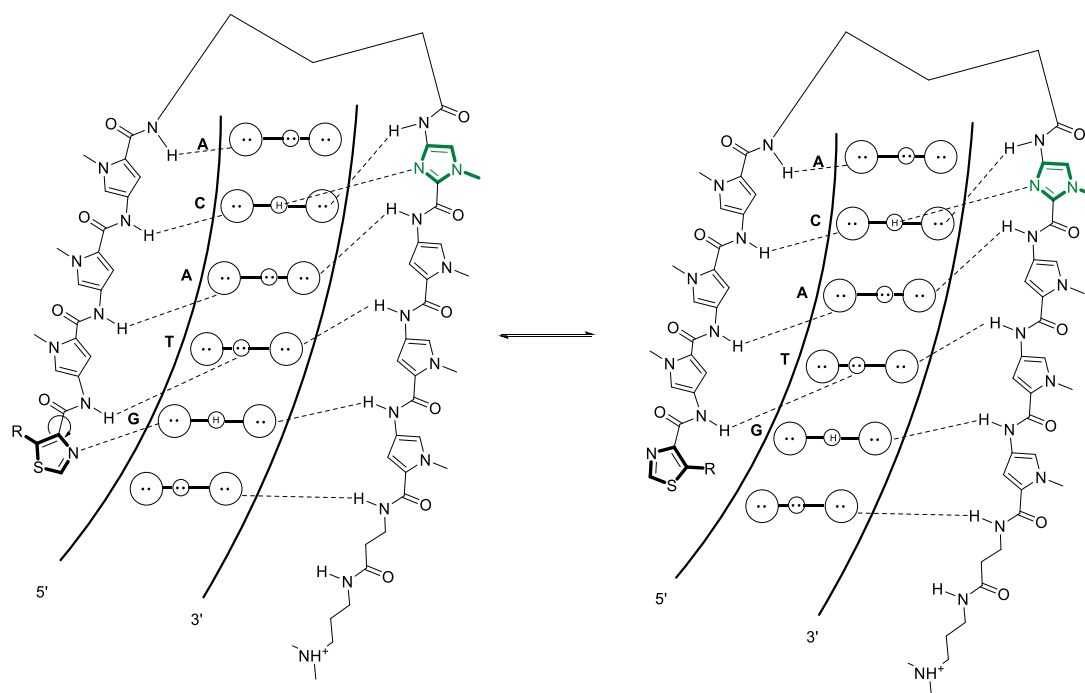


Figure 3.31 Schematic representation of the two possible conformations assumed by Nt polyamides.

These observed binding differences prompted the structural characterisation of selected Im- and Nt-containing polyamides (Chapter 4). Since the G-selective binding of **PA4** was encouraging for the development of novel polyamides with enhanced hydrophobicity targeting the ARE sequence, the structural evaluation was focused on establishing the binding differences of Nt and Im at the terminal position of the polyamide scaffold.

3.6 Experimental

3.6.1 *UV melting analysis*

Melting temperature analyses were performed on a Shimadzu UV spectrometer UV 1800 equipped with a thermo-controlled cell holder possessing a cell path length of 1 cm. An aqueous solution consisting of 10 mM monobasic phosphate, 10 mM dibasic phosphate, 100 mM NaCl, at pH 7.0 was used as the analytical buffer. Duplex DNA and the polyamides were mixed at a stoichiometry of 1:1.75 to give a final dsDNA concentration of 1 μ M for each experiment. Denaturing profiles were recorded at $\lambda = 260$ nm from 20 to 90 $^{\circ}$ C, with a heating rate of 1 $^{\circ}$ C/min. The reported melting temperatures were calculated from the averages of the maxima of the first derivative plots of three independent denaturing profiles.

3.6.2 *DNase I Footprinting*

DNA fragments for footprinting were prepared by ligating synthetic oligonucleotides, which had been designed to contain six heptanucleotide potential ligand binding sites, each separated by GCGC, into the BamHI site of pUC19. Successful clones were sequenced by Eurofins MWG Operon, Ebersberg, Germany.

Radiolabelled DNA fragments were obtained by cutting the plasmid with HindIII and SacI and labelling the 3'-end of the HindII site with α - 32 P[dATP] using *exo*-Klenow fragment of DNA polymerase I (New England Biolabs). The labelled DNA was separated from the remainder of the plasmid on a 6% polyacrylamide gel, eluted and dissolved in 10 mM Tris-HCl pH 7.5, containing 0.1 mM EDTA at a concentration of at least 10 c.p.s./ μ L as determined on a hand-held Geiger counter (approximately 10 nM DNA).

DNase I footprinting was performed by mixing 1.5 μ L radiolabelled DNA with 1.5 μ L ligand (diluted in 10 mM Tris-HCl containing 10 mM NaCl, from a 10 mM stock solution in DMSO). Samples were equilibrated for at least 30 minutes before adding 2 μ L DNase I (typically 0.01

units/ml) which had been diluted in 20 mM NaCl, 2 mM MgCl₂, 2 mM MnCl₂. The reaction was stopped after 1 minute by adding 4.5 μL 80% formamide containing 10 mM EDTA, 1 mM NaOH and bromophenol blue. Samples were denatured by heating at 100 °C for 3 minutes before cooling on ice and loading onto a 40 cm, 8% denaturing polyacrylamide gel containing 8 M urea, which was run at 1500 V for about 90 minutes. The gel was then fixed in 10% (v/v) acetic acid, dried onto 3MM filter paper and exposed overnight to a phosphor screen. The screen was scanned using a Typhoon phosphorimager.

3.6.3 *Circular Dichroism*

All CD measurements were performed on a Chirascan spectropolarimeter using a 1 cm cell. All spectra were recorded between $\lambda = 220$ and 390 nm and corrected for background contributions (step size 1 nm, count time per point 0.5 s). A 1 μM solution of the desired dsDNA in pH 7 PBS buffer was added with the ligand. After each addition, the solution was allowed to incubate for 3 minutes prior to measurement.

Chapter 4

Structural basis of Nt-containing polyamides binding to their target dsDNA

4.1 NMR structural determination of MGB·dsDNA complexes

Since the first X-ray structure of netropsin bound to dsDNA containing the target sequence 5'AATT,²²⁴ numerous MGBs have been structurally characterised to provide insight into their mechanism of action.^{80,83,85,132,133,225} In the case of Py-Im polyamides, the evident compression of the major groove induced in the central portion of the target sequence has been attributed to the disruption of TFs binding to their associated response element.^{132,133} Despite its high resolution, X-ray crystallography is reliant on finding suitable conditions for crystallisation. Moreover, artefacts resulting from crystal packing forces cannot be excluded during structural analysis.²²⁶ For these reasons, NMR spectroscopy is widely considered to be a powerful alternative structural tool to characterise biomacromolecules. The main advantage of this technique over X-ray crystallography is that it provides structural information in solution, which is more representative of biological conditions. However, a key restriction of this technique is the limited size of the macromolecules that can be analysed.²²⁶ In the case of nucleic acids, oligonucleotides larger than 20 base pairs result in complicated analyses associated with a large degree of signal overlap.¹³ Despite these limitations, the structural elucidation of linear and 6-ring hairpin and cyclic polyamides has been possible through a careful design of the sequence being analysed, which has facilitated the assignment.^{12,106,227} This has provided essential information on the binding mode of these compounds and has dictated the direction for the development of improved MGBs.

4.1.1 Principles of Nuclear Magnetic Resonance

Nuclear magnetic resonance exploits the magnetic properties of nuclei with spin (I) equal $\frac{1}{2}$ (such as ^1H , ^{13}C and ^{31}P), which are subjected to an external magnetic field (B_0) to assume two opposite spin orientations aligned with or opposite to B_0 (Figure 4.1).²²⁸ As a consequence of this, two energy levels are generated, each associated with a magnetic quantum number (m).

As dictated by the Boltzmann equation (Equation 4.1), the two levels are characterised by a small difference in population, with a small excess in favour of the level with lower energy.²²⁸

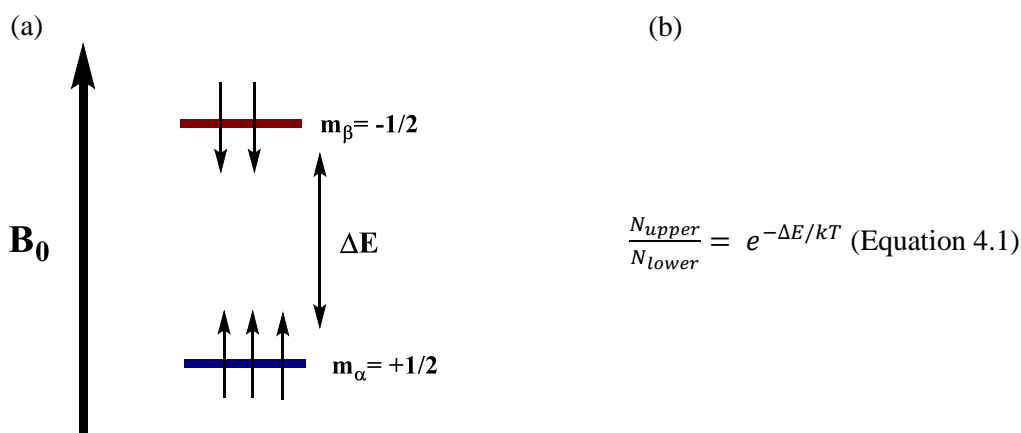


Figure 4.1 (a) Diagram of spin orientations of nuclei with $I = 1/2$ when B_0 is applied. (b) Boltzmann equation. N_{upper} and N_{lower} indicate the population of upper and lower energy states, k is the Boltzmann constant and T is temperature (K).

In a standard one-dimensional (1D) experiment, the energy necessary to flip the spin is provided by a radiofrequency pulse (RF), which allows the transition of nuclei from the low energy level to the higher, and thus equalises the populations of the two states (saturation). This energy is unique for each nucleus and is dependent on the magnetic field B_0 as in Equation 4.2:

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

where h is the Planck constant and γ is the gyromagnetic ratio, which is specific for each nucleus.

The radiofrequency is applied only for a few milliseconds, after which the system eventually returns to the ground state. This generates a free induction decay (FID), which is measured as voltage and is subsequently converted to the frequency domain via a Fourier transformation. Each nucleus of the molecules in the analysis experiences a different local magnetic field and resonates at a slightly different frequency, generating the chemical shift.²²⁸ Furthermore, the transitions between multiple energy levels, originated from close active spins in the molecule,

generate scalar (through-bond J coupling) and dipolar couplings (nuclear Overhauser effect, nOe) (Figure 4.2).²²⁸

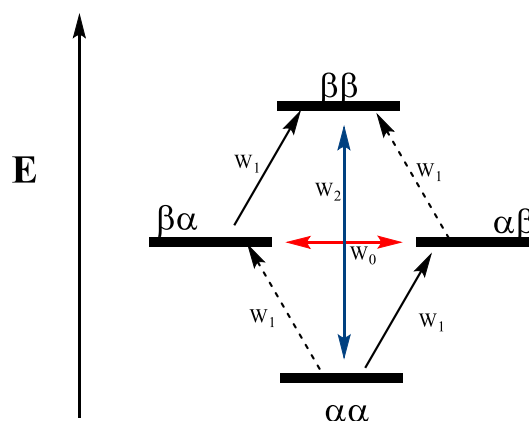


Figure 4.2 Schematic representation of the energy levels in a two-spin (α - β) system (solid and dashed black arrow W_1 = single quantum transition, red arrow W_0 = zero quantum transition, blue arrow W_2 = double quantum transition).²²⁸

Owing to the large number of NMR-active nuclei in dsDNA, 1D ^1H spectra normally appear strongly overlapped. However, because of the oligomeric nature of dsDNA, a ^1H spectrum can be divided into distinct regions (Figure 4.3).¹³ The imino protons which are involved in the base pairing are normally found in the most downfield region of the spectrum (12-14 ppm). The aromatic protons of the bases resonate between 8.5 and 7 ppm with the exception of H5 of cytosine, which are found between 5.5 to 6 ppm. Sugar protons cover a wide area of the spectra, normally between 6 and 2 ppm. The H1' are found more downfield because they experience deshielding from both the nitrogen of the base and the oxygen of the sugar. H3', H4' and H5'/H5'' protons are normally found in a crowded region between 5 to 3.5 ppm. On the other hand, H2' and H2'' protons are found between 3 and 1.7 ppm. Farther upfield is the characteristic methyl signal of thymine, which normally resonates at 1.5 ppm. The most common solvent mixture utilised for the analysis of dsDNA is 90% H_2O /10% D_2O , which provides information on the exchangeable protons (*i.e.*, imino and amino) involved in the base pairing and, thus, on the formation of an assembled DNA duplex.^{12,13,82,229} However, 99% D_2O is also used in DNA analysis for simplifying regions of the spectrum where exchangeable

protons appear. Furthermore, 99% D₂O provides better signal resolution because it does not require solvent suppression, which could potentially hide some resonances in the vicinity of the HOD signal (4.74 ppm) or generate artefacts in 2D experiments.¹³

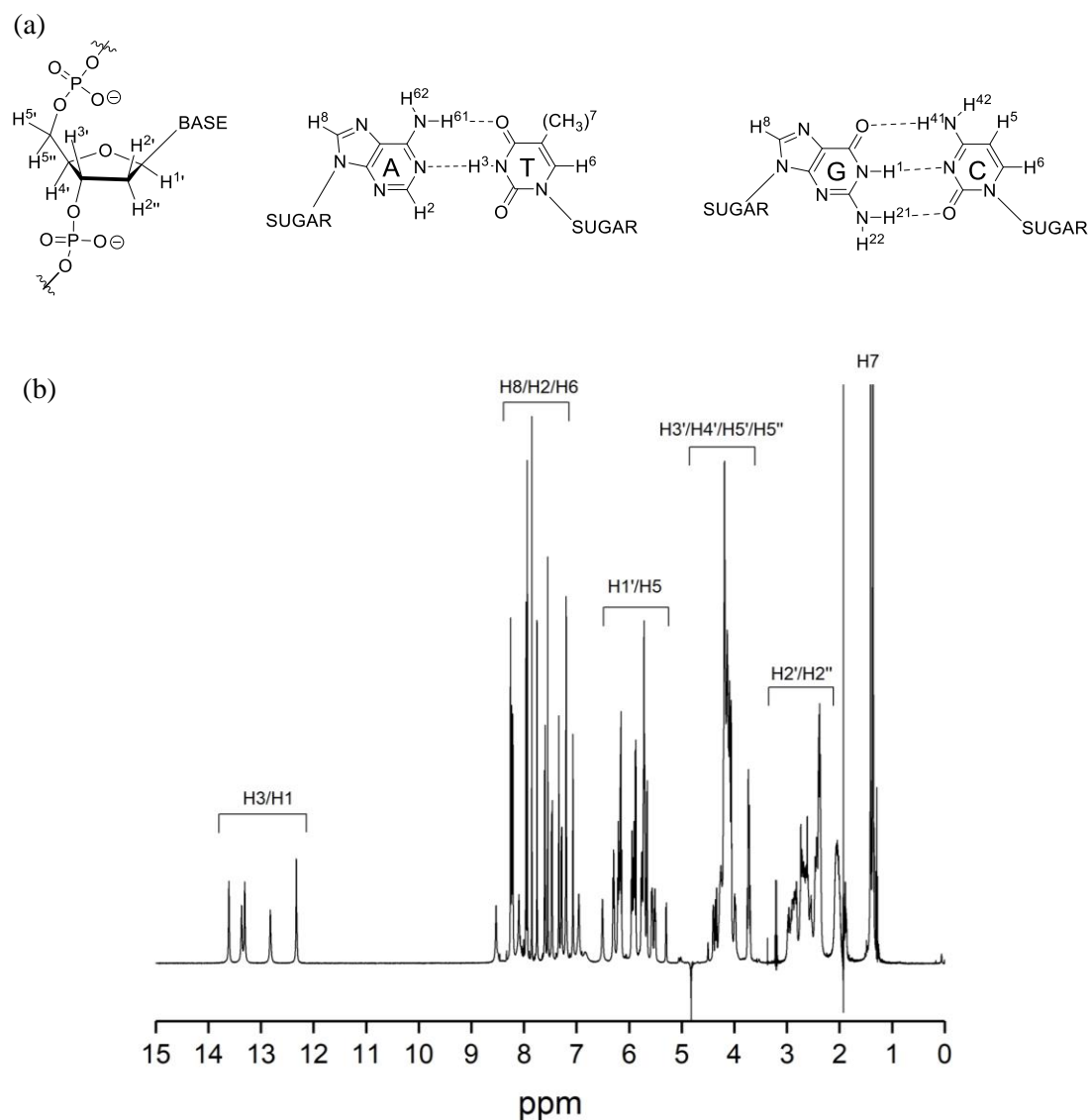


Figure 4.3 (a) Nomenclature of sugar and base protons of DNA. (b) ¹H-NMR of d(CGATGTACATCG)₂ in 90% H₂O/10% D₂O.

4.1.2 1D-NMR studies of ligand-dsDNA interactions

1D NMR experiments represent the first approach in the study of ligand·DNA interactions.^{13,82,230} The imino proton region (12-14 ppm) is the common starting point of this

analysis, which is based on the observation of the change in shape and chemical shifts of proton resonances, as a reflection of the binding interactions. The observed responses are correlated with exchange rate of the assembly/disassembly of the complex in relation with the NMR timescale and provide information on the strength of the binding.¹³

As a result of the strong binding affinity of MGBs to their target dsDNA sequence, the equilibrium of complex formation is likely to have a slow exchange rate (Figure 4.4). If this is considerably slower than the NMR timescale (>100 ms), then the spectrum will be characterised by a well-defined peak shape. For this reason, when titrating high-affinity ligands it is possible to clearly detect bound and unbound species involved in the equilibrium, which eventually convert into a singular species upon saturation of the binding site (see section 4.3.2). In the case of weaker binders (*i.e.*, mismatch target sequence or non-specific interaction), if the equilibrium is biased towards the unbound state (Figure 4.4), the exchange rate will be fast on the NMR timescale. This is normally detected as increased broadening of the signals with no substantial change in the chemical shift of the peaks.¹³ More complicated is the evaluation of binding events with exchange at rates similar to the NMR timescale. In these particular cases, line broadening and coalescence are observed, which strongly affect the quality of the spectrum.¹³

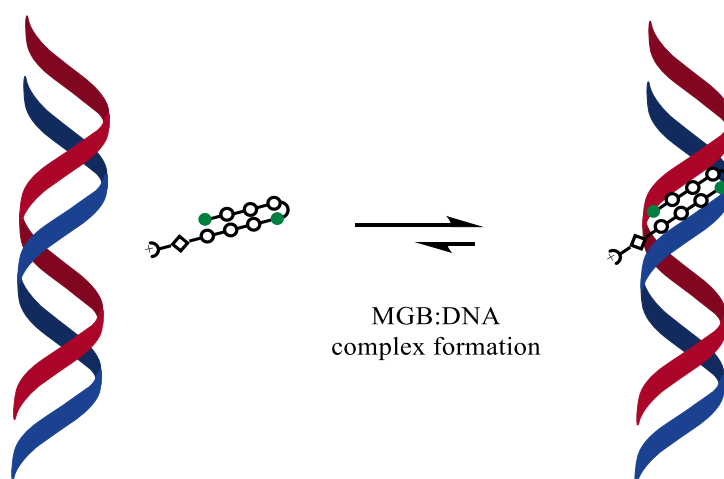


Figure 4.4 Schematic representation of the equilibrium involved in the formation of MGB-DNA complex.

Some preliminary structural characterisation of the binding can be also obtained from 1D ^{31}P -NMR, which gives information on the perturbation of the phosphodiester backbone. In a typical B-DNA conformation, the phosphodiester backbone is in an equilibrium between a B_I and B_{II} conformation, classified by the difference value (ε - ζ) between ε ($\text{C4}'\text{-C3}'\text{-O3}'\text{-P}$) and ζ ($\text{C3}'\text{-O3}'\text{-P-O5}'$) angles of -90° and $+90^\circ$, respectively (Figure 4.5).^{231,232} This element of flexibility in the backbone is directly correlated to the dimension of the grooves and is therefore fundamental for the shape readout of dsDNA by proteins.²³² The chemical shift of the ^{31}P resonances is an average representation of this equilibrium. With a few exceptions, the B_I contribution normally prevails and results in a chemical shift at around -1 ppm.^{13,82,231} In contrast, positive chemical shift is observed for B_{II} conformation.¹³ Clear deviations from the standard chemical shift of the unbound DNA that are seen upon addition of the ligand reflect the conformational changes of the phosphodiester backbone, providing quick and qualitative evidence of the interaction (see section 4.3.4).

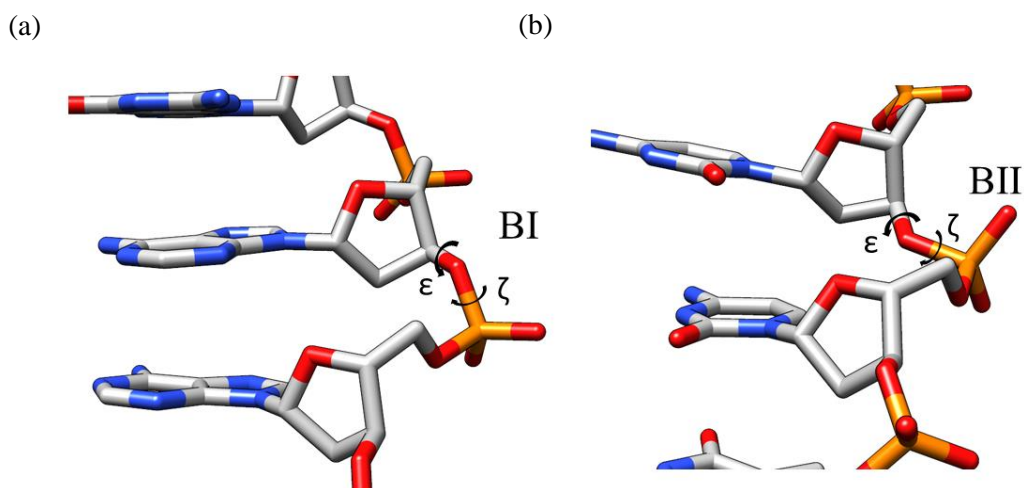


Figure 4.5 X-ray structure of (a) B_I (PDB 1EHV)²³³ and (b) B_{II} phosphate linkages (PDB 3GGI).

4.1.3 2D-NMR and structural assignment of dsDNA

The assignment of nucleic acids is commonly achieved using a combination of 2D experiments. The majority of the information is normally obtained by 2D [^1H , ^1H] NOESY experiments. This methodology exploits the nuclear Overhauser effect (nOe), which arises from through-space magnetisation transfer between nuclei located $<6 \text{ \AA}$ apart from each other. In a standard NOESY experiment, this correlation is obtained through three sequential 90° pulses (Figure 4.6).²²⁸

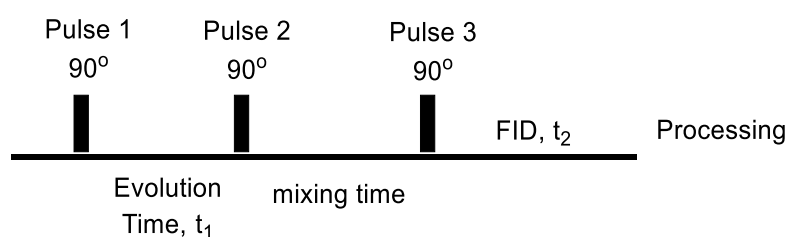


Figure 4.6 Schematic of the pulse sequence used in a NOESY experiment.

After a first 90° pulse, transverse magnetization evolves during time t_1 , then a second 90° pulse generates the longitudinal magnetization that is transferred during nOe mixing times through cross-relaxation between nuclei via zero (W_0)- or second (W_2)-order quantum transitions.²²⁸ The final 90° pulse creates the transverse magnetization that is detected. The tumbling rate and viscosity of the system play important roles in the nOe effect.²²⁸ For the analysis of macromolecules, zero-order transitions are prevalent and result in negative nOe signals. On the other hand, double-order quantum transitions are predominant for small molecules, producing positive nOes.²²⁸ The difference in the nOe sign provides a first indication of the binding event, as a transition from positive to negative peaks of the small molecule should be observed upon interaction.^{13,203} Strong intra- and inter-nucleotide nOe correlations are exploited for sequential assignment of the DNA strands. Key regions of the DNA NOESY spectrum are located between 6-8 ppm and 5-6.5 ppm, which include H1' and H6/H8 resonances. In a B-DNA conformation, the sequential walk along the DNA strand is permitted by a strong correlation between H1' and H6/H8 of the adjacent base in the 3'

direction (Figure 4.7).¹³ For samples in 90% H₂O/10% D₂O, the imino and amino protons nOe correlations allow the user to walk across the base pairs to complete the assignment of the DNA skeleton.¹³

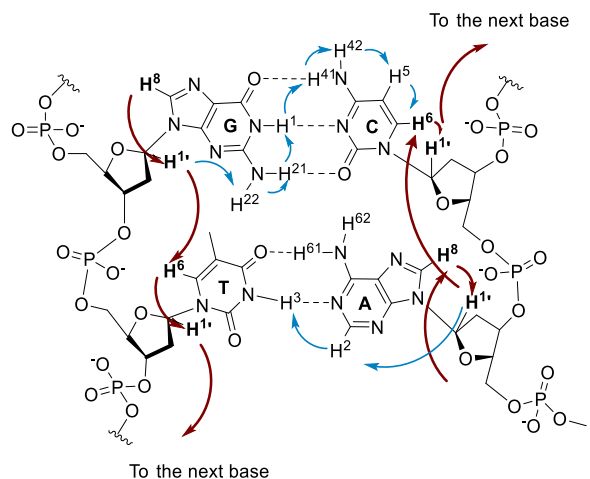


Figure 4.7 nOe contacts used for the assignment of B-DNA (red arrows = strand contacts; blue arrows = base-pair contacts).

C and T nucleobases can be identified by correlation spectroscopy (COSY) and total correlation spectroscopy (TOCSY) experiments using the C(H5)-C(H6) and T(H6)-T(H7) cross-peaks. The COSY experiment consists of two 90° pulses separated by the evolution time t_1 , producing cross-peaks between protons directly linked to each other and separated by no more than three bonds (Figure 4.8a). For analysis of biomacromolecules, the double-quantum filter (DQF) variant of COSY is employed as it increases resolution by introducing a third 90° pulse that filters out the single quantum transitions (W_1) of uncoupled signals (Figure 4.8b).²²⁸ This reduces the intensity and sharpens the line-shape of the diagonal peaks, allowing analysis of cross-peaks in its vicinity. The TOCSY experiment provides longer scalar coupling correlations (up to 5 bonds) by replacing the second pulse of the COSY with a spin-lock pulse (closely spaced 180° pulses, Figure 4.8c). This suppresses chemical shift evolution and results in enhanced J coupling strength, which can be observed as cross-peaks.²²⁸ COSY and TOCSY experiments are also fundamental for the assignment of the sugar protons, which cannot be

unambiguously distinguished by the sole NOESY experiment because of the high degree of peak overlap. In a standard B-DNA conformation, assignment of H2', H2'', H3' and H4' is achieved using a pattern of cross-peak correlations outlined in Figure 4.8d. H5' and H5'' represent the most challenging assignments as they are strongly overlapped in the proximity of the diagonal.

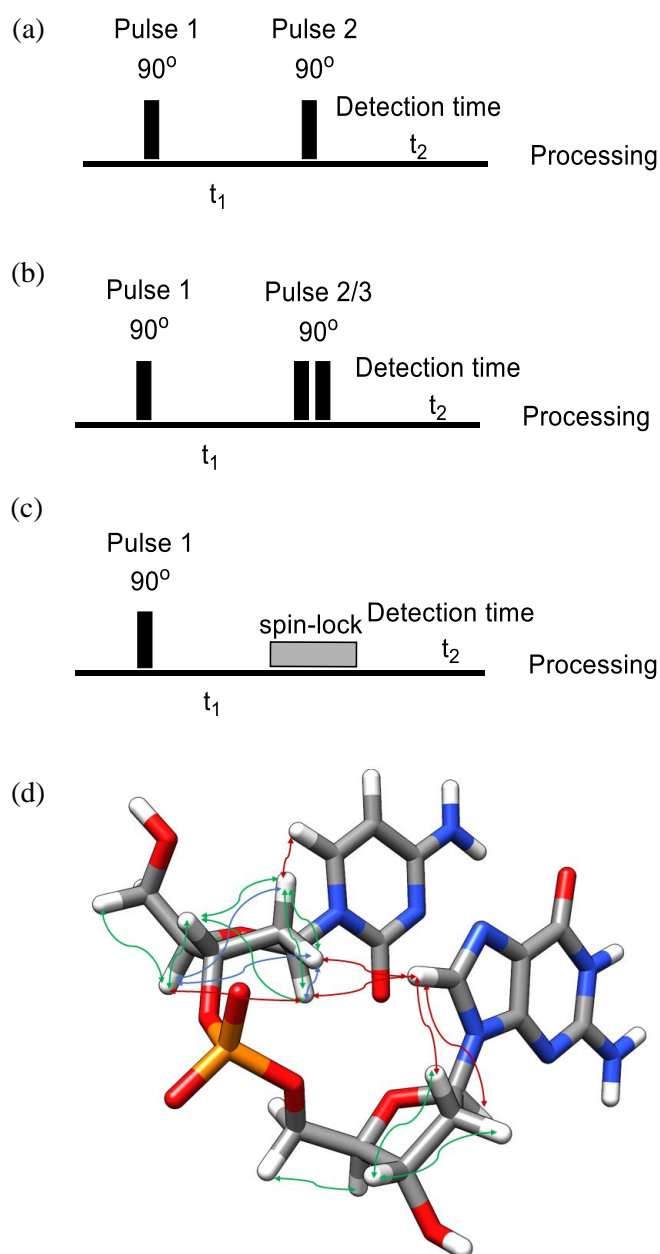


Figure 4.8 Schematic of the pulse sequence used in (a) COSY, (b) DQF-COSY (c) TOCSY experiments. (d) Representation of the correlations between sugar protons observed in B-DNA (red arrow = nOe, green arrow = nOe and COSY blue arrow = TOCSY and nOe).

Further simplification of the assignment of the sugar protons is provided by heteronuclear experiments such as [^1H - ^{13}C] HSQC, which detect correlations between directly bonded ^1H and ^{13}C nuclei. The more complex pulse sequence of this experiment can be divided into three main blocks (Figure 4.9a).²²⁸ In block A, an initial 90° pulse generates transverse magnetization of ^1H . Dipolar heteronuclear coupling evolution is allowed for a fixed time by applying a simultaneous 180° pulse on both ^1H and ^{13}C in the middle of this period to refocus chemical shift evolution. Another 90° pulse is then introduced, allowing ^{13}C to evolve in chemical shift for a period t_1 , during which ^1H chemical shift is refocused by subsequent 180° pulse (block B). Concomitant 90° and 180° pulses are then applied to reverse the magnetization (block C) followed by the detection.²²⁸ The HSQC of B-DNA appears divided in distinct clusters, each belonging to a certain type of C-H correlation (Figure 4.9b). As ^{13}C chemical shift perturbations are normally less pronounced upon ligand binding, HSQC is a useful tool for identification of those protons that are predominantly affected by the interaction, resulting in resonances that appear outside the expected cluster (see section 4.3.4).¹³

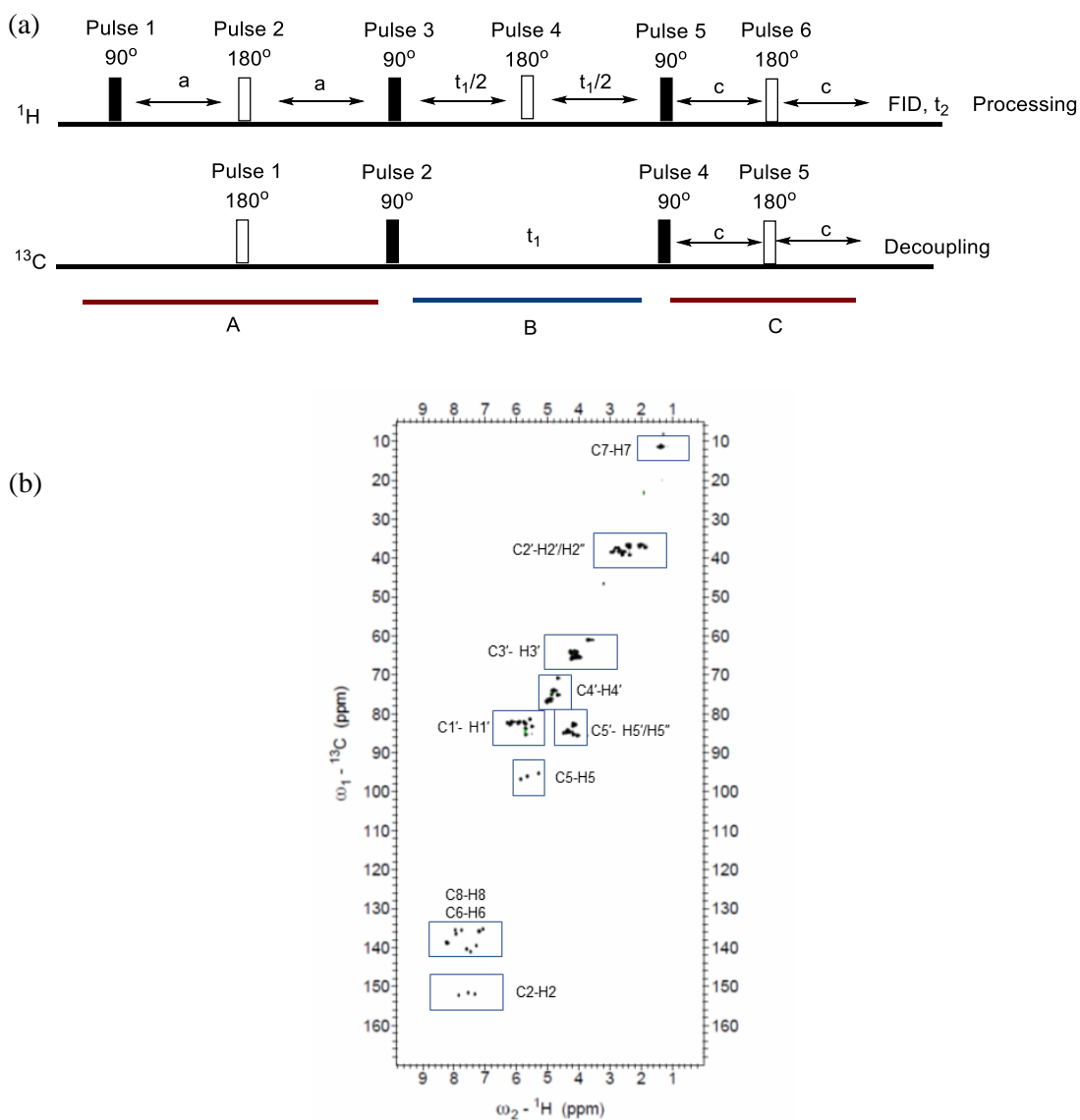


Figure 4.9 (a) Schematic of the pulse sequence used in HSQC experiment. (b) HSQC of d(CGATGTACATCG)₂.

Another commonly used heteronuclear experiment for DNA analysis is [^1H - ^{31}P] COSY, which provides the ability to distinguish ^{31}P resonances through scalar correlation with sugar protons H3', H5'/H5'' and H4' of two consecutive nucleotides (Figure 4.10a).²³⁴ In this experiment, ^1H is saturated by alternating 120° pulses at the beginning of the sequence (Figure 4.10b). A 90° pulse on the ^{31}P generates transverse magnetization, which evolves only under the effect of

scalar coupling by a simultaneous ^1H and ^{31}P 180° pulse. The final 90° pulses provides the observable magnetization that is recorded.²³⁴

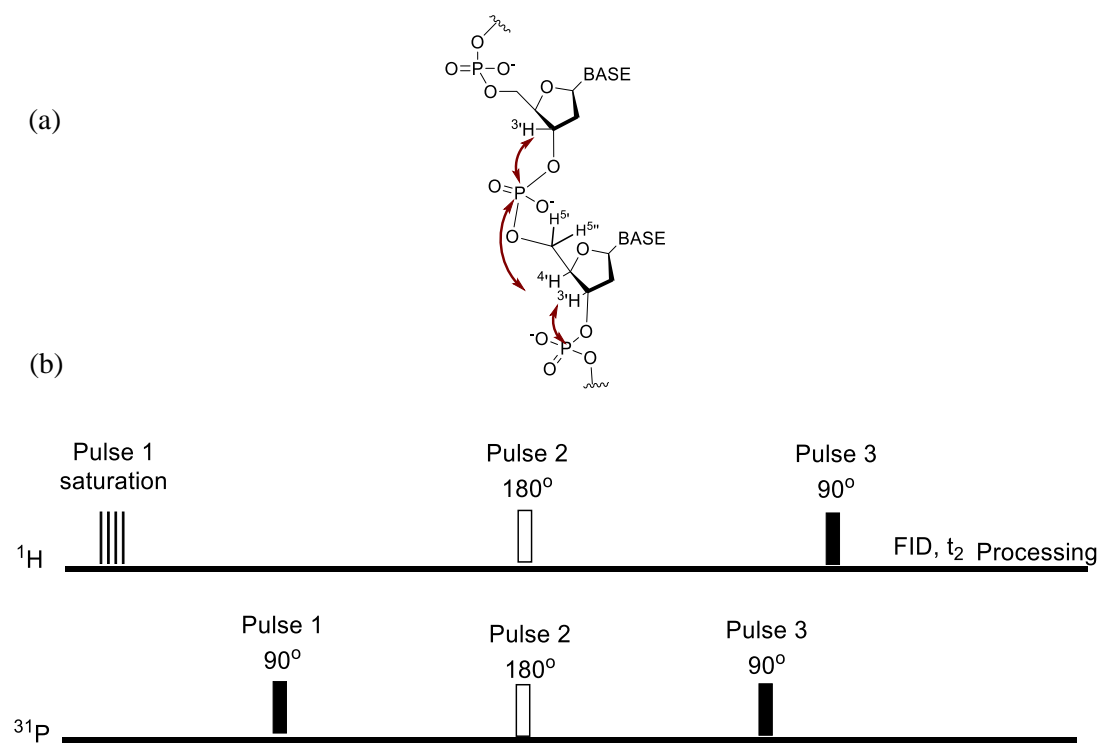


Figure 4.10 (a) Cross peaks observed in a $[^1\text{H}\text{-}^{31}\text{P}]$ COSY. (b) Schematic representation of the pulse sequence of $[^1\text{H}\text{-}^{31}\text{P}]$ COSY experiment.²³⁴

4.1.4 Principles of molecular dynamics and NMR-structural determination

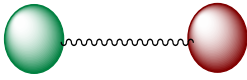
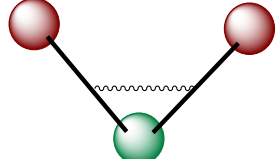
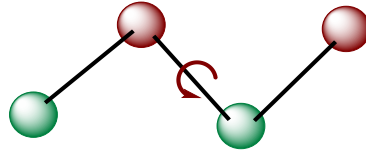
The information obtained from the NMR analyses can be combined with a molecular dynamic approach to obtain an experimentally-derived 3D structure.^{12,82,163} Molecular dynamics are computer simulations that use a molecular mechanics force field to determine interactions and motion between atoms and molecules solving the Newtonian classical equations of motion (Equation 4.3):

$$f_i(t) = m_i a_i(t) = - \frac{\delta V(x(t))}{\delta x_i(t)} \quad (\text{Equation 4.3})$$

where $f_i(t)$ is the net force acting on atom i at a given time t , a is the acceleration of the corresponding atom and m is the mass.²³⁵ Through this approach, it is possible to obtain the

configuration of the system at time t by the vector $x(t)$ and an empirical potential energy function δV , which is used to approximate interaction between particles. This is known as the force field, which defines the potential energy as the ensemble of forces acting on a system of molecular coordinates.²³⁵ The Amber force field is one of the most-used for proteins and nucleic acids.²³⁶ This force field describes the element type, hybridization and bonding characteristics of each atom of the molecule and is comprised of the sum of bonding (*i.e.*, bond stretching, torsions of angle and dihedrals between connected atoms) and non-bonding interactions (*i.e.*, van der Waals and electrostatic interaction), which are defined through a series of mathematical approaches (Table 4.1).

Table 4.1 The functional form of the Amber force field parameters. The parameters listed are: K_r , K_n and K_ϕ = force constants; r = current bond length and r_0 = bond length at equilibrium; θ = current bond angle; θ_0 = bond angle at the equilibrium; ϕ = current dihedral angle, ϕ_0 = dihedral angle at the equilibrium; ϵ_{ij} = depth of the energy well, R_{ij} = current internuclear distance between atom ij ; $R_{0,ij}$ = equilibrium internuclear distance between ij , q = point charge, ϵ = permittivity.²³⁵

Potential Energy	
$\sum_{bonds} \frac{K_r}{2} (r - r_0)^2$	
$\sum_{angles} \frac{K_\theta}{2} (\theta - \theta_0)^2$	
$\sum_{dihedrals} \frac{K_\phi}{2} [1 + \cos(n\phi - \phi_0)]$	
$\sum_{i < j} \epsilon_{ij} \left[\left(\frac{R_{0,ij}}{R_{ij}} \right)^{12} - 2 \left(\frac{R_{0,ij}}{R_{ij}} \right)^6 \right]$	Van der Waals (Lennard-Jones)
$\sum_{i < j} \frac{q_i q_j}{\epsilon R_{ij}}$	Electrostatic

In MD simulations, the position of each atom at time t is solved by splitting the equation of motions in small time steps (δt) of 1-2 femtoseconds, knowing the initial velocities and acceleration and approximating the force to be constants.²³⁵ After each step, acceleration and velocities are back-calculated from the forces of the new position, before advancing to the next time interval.

Using this approach, MD simulations consist of a series of snapshots which provide a statistical ensemble of conformations of a microscopic system which does not exchange matter or energy with the surroundings. To mimic a more natural macroscopic behaviour, functions such as the thermostat (which simulates fluctuations of kinetic energy) or the barostat (which controls the pressure scaling the volume of the system) are commonly used.²³⁵ Periodic boundary conditions are also employed to mimic the bulk of a macroscopic system. In this approach, coordinates and velocities are stored in a unit cell, which is replicated in all directions to form an infinite lattice.²³⁵ In this method, the computational costs of the simulation are greatly increased by accounting for the non-bonding interactions, which need to be considered for all the atoms in the cell and all the periodic images,²³⁷ for this reason, a cut-off distance of 9-11 Å is normally set for van der Waals interactions. Due to their intrinsically long-range nature, this same approach cannot be used for electrostatic interactions, which are normally treated with the particle-mesh Ewald algorithm and thus this approach accounts differently for short and long-range interactions and reduces the computational cost.²³⁷

In NMR-restrained molecular dynamics, the MD empirical approach is complemented by the experimental information derived from the nOe cross-peaks observed in the NMR, which are correlated to internuclear distances according to:

$$\sigma \propto \frac{1}{r^6}$$

where σ is nOe intensity and r is the distance between the nuclei.²²⁸ In the Isolated Spin Pair Approximation (ISPA), the distances between the nuclei of the DNA are calculated by using nOe intensities of known distances, such as C(H5)-C(H6), as a reference.²³⁸ However, this approximation is only valid assuming an isotropic motion between two isolated protons and does not take into account perturbation from neighbouring protons or the spin diffusion effect that makes the intensities unreliable.

The Matrix Analysis of Rates for Discerning GeometRy in Aqueous Solution (MARDIGRAS) is an extremely useful approach to reduce the errors induced by these effects.²³⁸ The MARDIGRAS algorithm requires an original model of the structure to calculate a matrix of mixing coefficients. Experimental intensities are then used to calculate a new rate matrix. This procedure is iterated to reduce the error between calculated and observed intensities, thus providing a set of distances.²³⁸ The main disadvantage of this approach is the requirement for a starting model for the initial matrix calculation. This is normally generated from previously reported crystal structures or using computational tools.

4.2 Aims and objectives of this chapter

The objective of this chapter is to provide structural insight on the binding mode of Nt-containing polyamides. Data arising from UV thermal melting analysis, DNase I footprinting and switchSENSE[®] technology revealed a distinct binding mode for Nt-containing polyamides compared to the Im-containing **PA1**. Of particular interest is the influence of substituents in the 5-position of Nt, which appear to have a role in polyamide selectivity. For this reason, NMR studies were undertaken to understand the structural differences between Nt-containing polyamide **PA3-4** and the Im polyamide (**PA1**) when bound to the target sequence 5'ATGTACA.

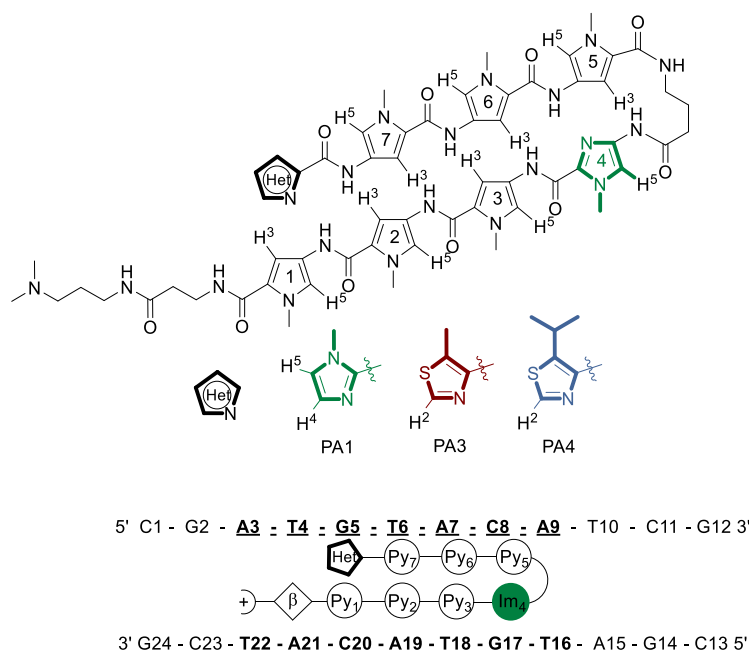


Figure 4.11 Nomenclature and schematic representations of **PA1**, **PA3** and **PA4** in complex with **ODN1**.

The specific aims of this chapter are as follows:

- (i) NMR assignment of **PA1** and **PA3-4** in complex with the self-complementary dodecamer oligodeoxyribonucleotide **ODN1** sequence d(CGATGTACATCG)₂ containing the target sequence 5'ATGTACA.
- (ii) Comparison of the binding characteristics of **PA1** and **PA3-4** by chemical shift perturbation and nOe analysis.
- (iii) Structural determination of the polyamides·dsDNA complexes using NMR-restrained MD.

4.3 Results and Discussion

4.3.1 NMR assignment of ODN1

The assignment of the unbound **ODN1** (d(CGATGTACATCG)₂) was achieved using a suite of 2D-NMR techniques in two different solvent mixtures: DQF-COSY, TOCSY and

heteronuclear experiments [^1H , ^{13}C] HSQC and [^1H , ^{31}P] COSY were acquired in 99% D_2O , and a set of 2D-NOESY experiments with different mixing times was recorded in 99% D_2O and 90% $\text{H}_2\text{O}/10\%$ D_2O . The cross-peak between C1(H6) and C1(H5'/H5'') was first identified from the distinct upfield chemical shift (3.72 ppm) of C1(H5'/H5''), which is due to the lack of the phosphate at the 5' terminus in synthetic DNA.¹³ This served as the starting point for the sequential walk along the **ODN1** (Figure 4.12).¹³ NOESY spectra in 90% $\text{H}_2\text{O}/10\%$ D_2O were used to identify the imino and amino protons and to complete the assignment of the bases. Imino proton resonances corresponding to the Watson-Crick base pairs at the edges of the DNA fragment were not observed. The assignment of the sugar protons was obtained from a combination of NOESY, DQF-COSY and TOCSY correlations arising from the identified H1'.

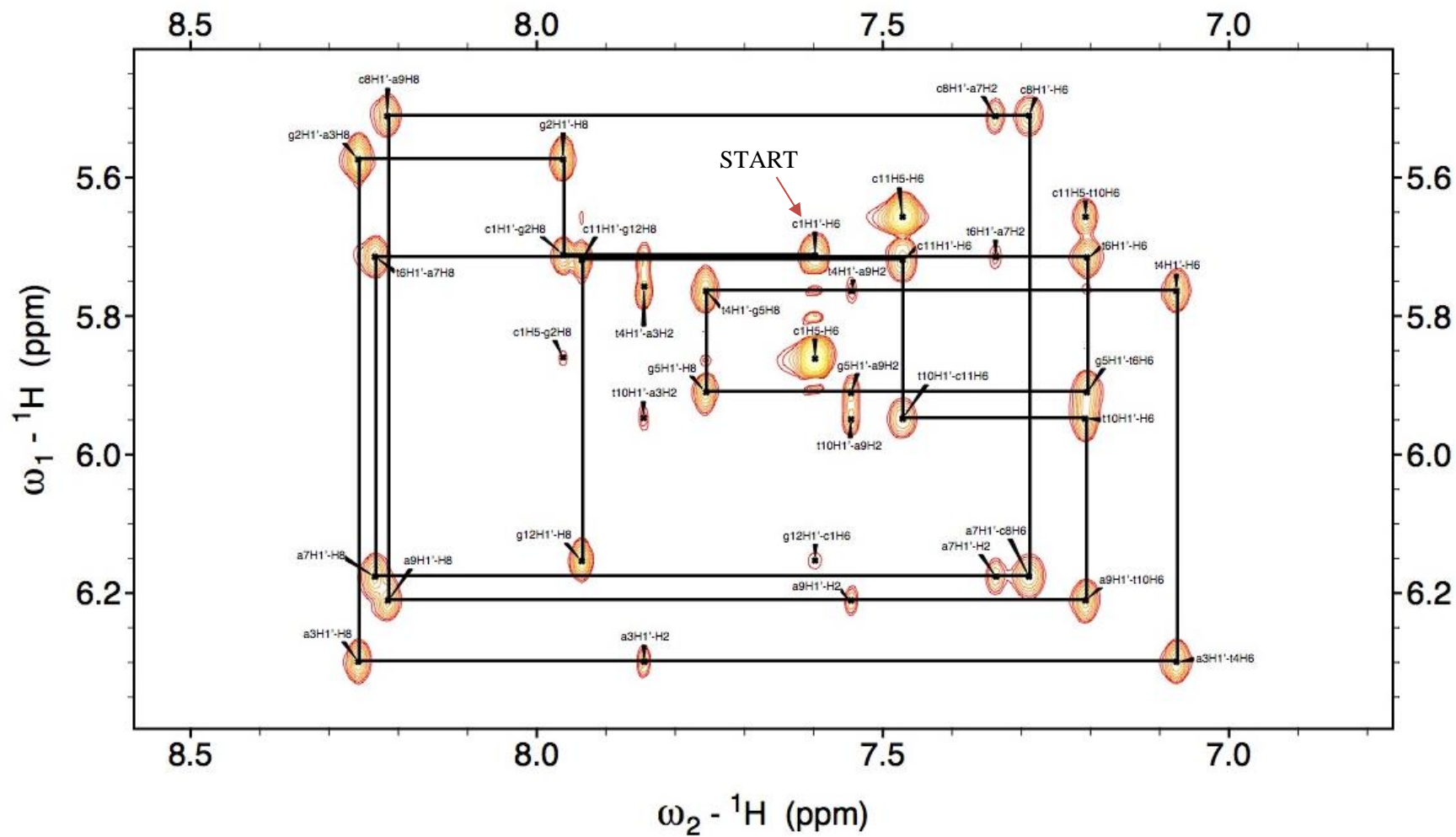


Figure 4.12 Depiction of the nOe walk of **ODN1** obtained from the NOESY spectrum at 250 ms mixing time in 99% D₂O.

Non-quaternary carbons were easily identified using 2D [^1H , ^{13}C] HSQC (Figure 4.13). Lastly, ^{31}P chemical shifts were unambiguously determined via COSY cross-correlation with H3' and/or H4' of the two adjacent nucleotides, with the exception of the terminal G12(P) that did not produce any unambiguous correlation (Figure 4.14).

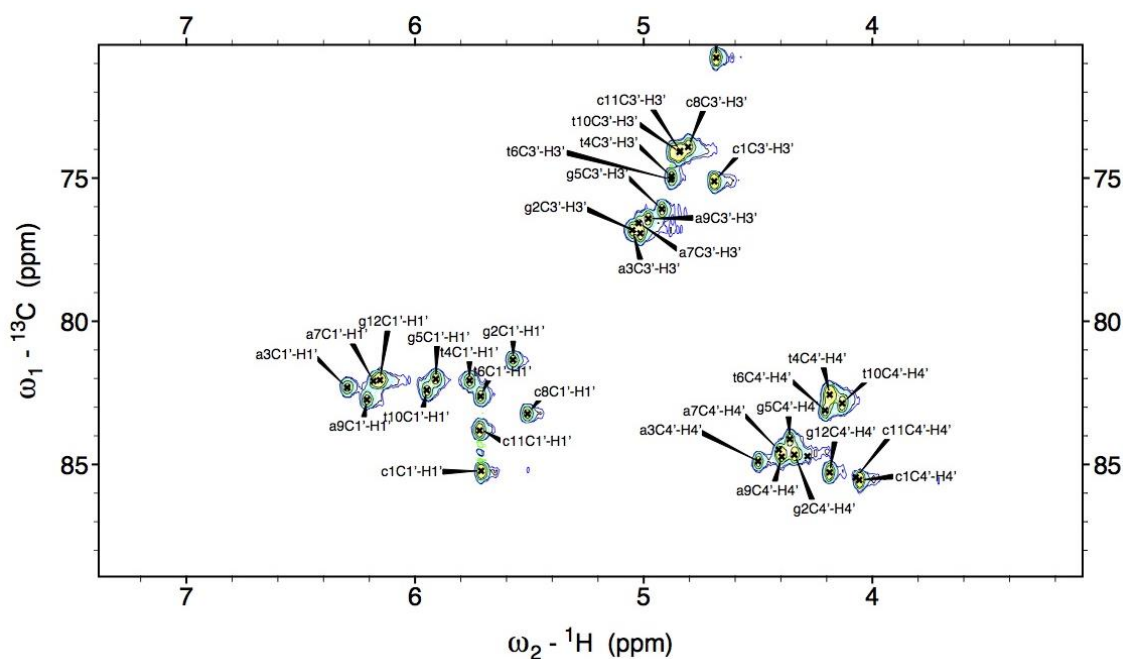


Figure 4.13 Selected region of the [^1H , ^{13}C] HSQC of ODN1.

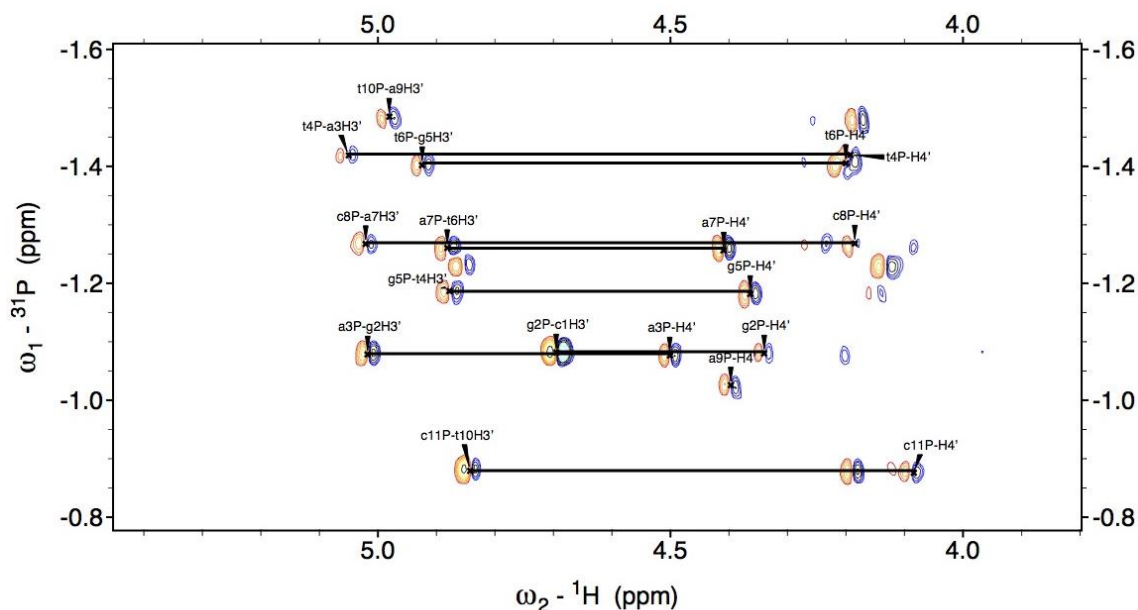


Figure 4.14 [^1H , ^{31}P] COSY of **ODN1** (black lines indicate the dinucleotide step determined through the ^{31}P correlation).

4.3.2 Titration of polyamides into a dsDNA solution

Three aliquots of **ODN1** were titrated with each of the analysed polyamides (*i.e.*, **PA1**, **PA3** and **PA4**). The samples were gently warmed after each addition to reduce the effect of polyamide aggregation and facilitate the formation of the complex. The defined line-shape of the new peaks corresponding to the bound species was suggestive of a strong binding event for all the polyamides. Saturation of the binding site was determined by disappearance of the peaks of unbound **ODN1**, which were replaced by the signal of a singular complex species in each case. A 1:1 stoichiometric complex for each polyamide was observed by 1D ^1H NMR, characterised by a doubling of the number of DNA imino proton signals and a single set of polyamide resonances (Figure 4.15). This was further confirmed by the presence of a single proton for the aromatic Nt(H2) of **PA3-4**.

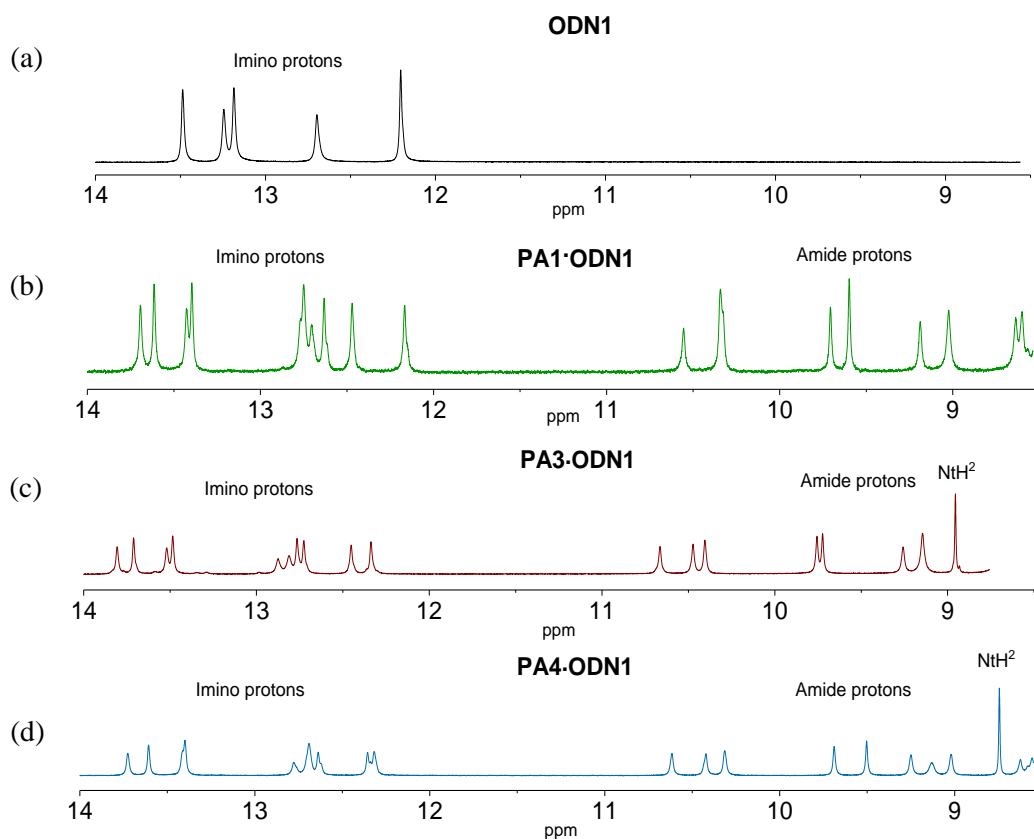


Figure 4.15 Selected region of 1D $^1\text{H-NMR}$ of (a) ODN1 and complexes (b) PA1·ODN1, (c) PA3·ODN1 and (d) PA4·ODN1 at 1:1 stoichiometry.

4.3.3 Assignment of PA·ODN1 complexes

Assignment of the DNA within the polyamide complexes PA1·ODN1, PA3·ODN1 and PA4·ODN1 was achieved in a similar fashion, despite the complication arising from the asymmetry of the complexes. Therefore, a sequential walk was necessary for each strand of the dsDNA (Figure 4.16-17, see Appendix 132-135 for PA1·ODN1 and PA3·ODN1)

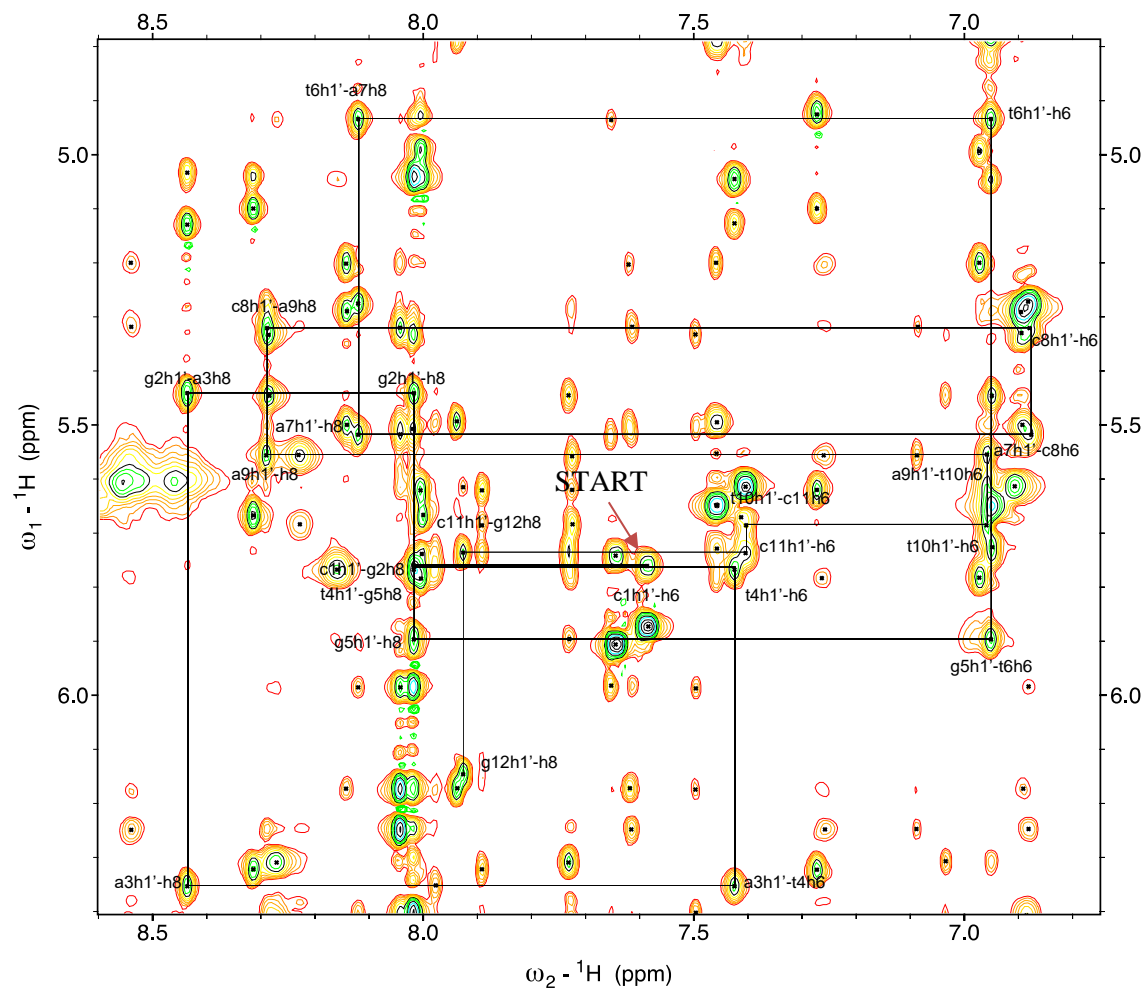


Figure 4.16 Depiction of the nOe walk of the strand C1-G12 PA4-ODN1 obtained from NOESY spectrum at 250 ms in 90% H₂O/10% D₂O.

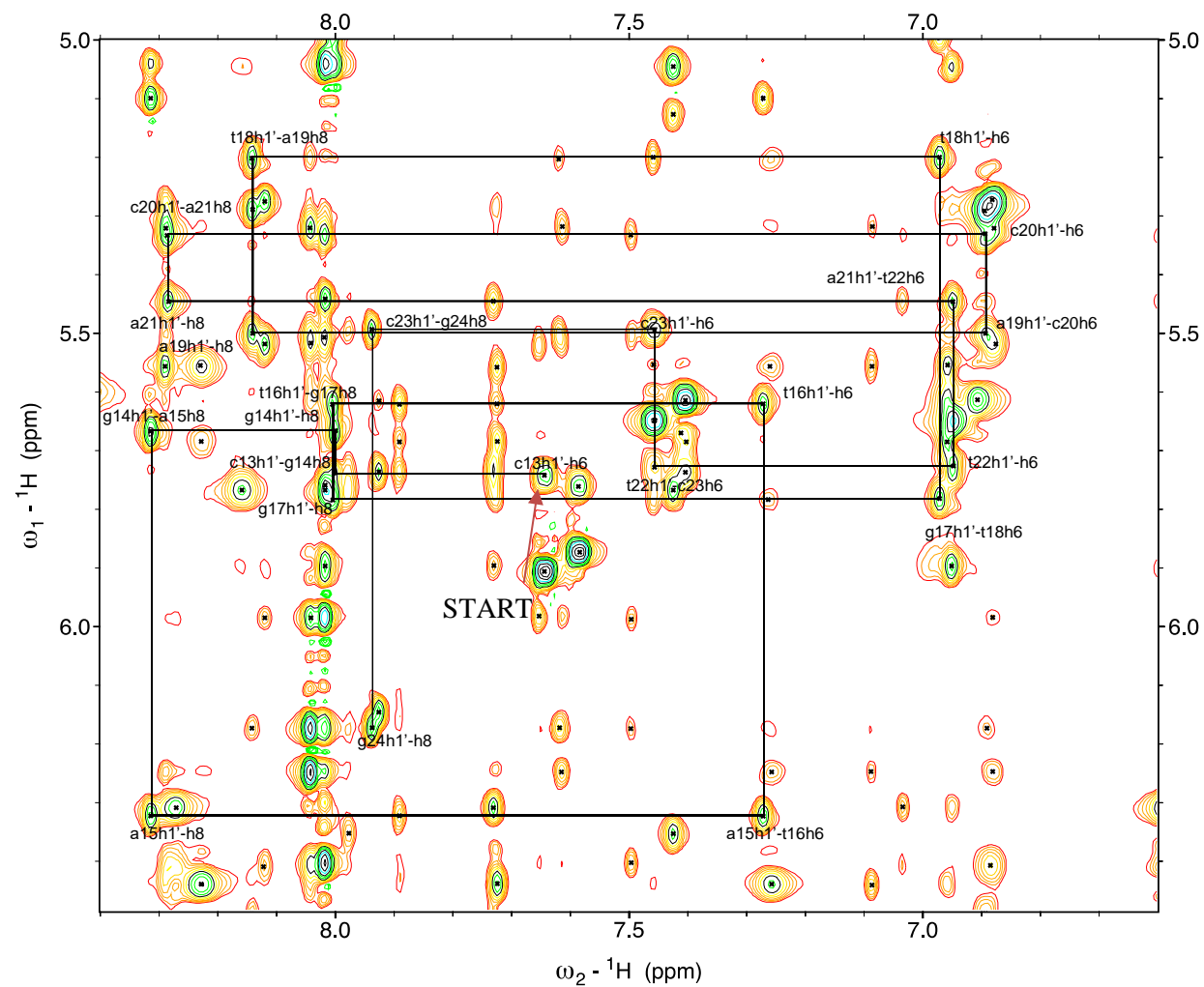


Figure 4.17 Depiction of the nOe walk of the strand C13-G24 of **PA4:ODN1** obtained from NOESY spectrum at 250 ms in 90% H₂O/10% D₂O.

The identification of the remaining sugar and base protons, carbons and phosphorus atoms of the DNA in each complex was then completed as previously described for unbound **ODN1**.

Assignment of the polyamides was achieved by a sequential nOe walk across aromatic and amide protons (Figure 4.19). For **PA3ODN1** and **PA4ODN1**, the NtH2 proton was diagnostic (8.97 ppm for **PA3**; 8.87 ppm for **PA4**) and was used as a starting point for the structural assignment of these polyamides. Strong nOe cross-correlations were observed from Nt(H2) to the adjacent Py7(NH) and Py7(H3), which were used in conjunction to identify Py7(H5) via NOESY and TOCSY correlations. nOe correlations between Py7(H5)/Py7(H3) and (Me)Py7 completed the assignment of Py7 resonances. This same pattern was observed for all the aromatic rings of the polyamides.

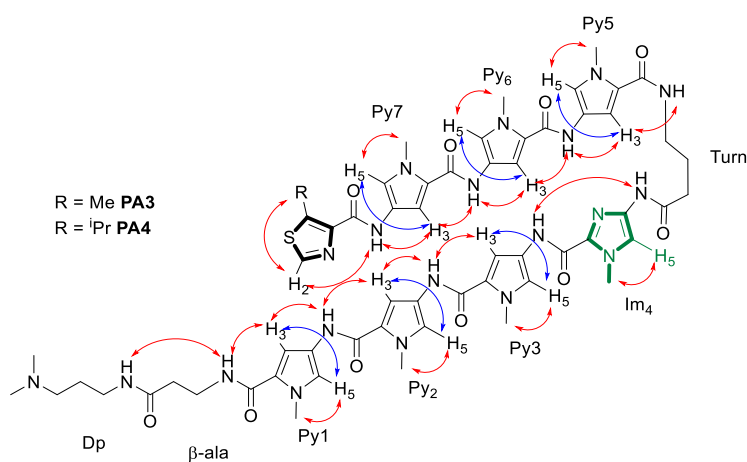


Figure 4.18 (a) Schematic depiction of the main cross-correlations used for **PA3-4** assignment (red arrows = nOe cross-correlations, blue arrows = TOCSY and NOE cross-correlations).

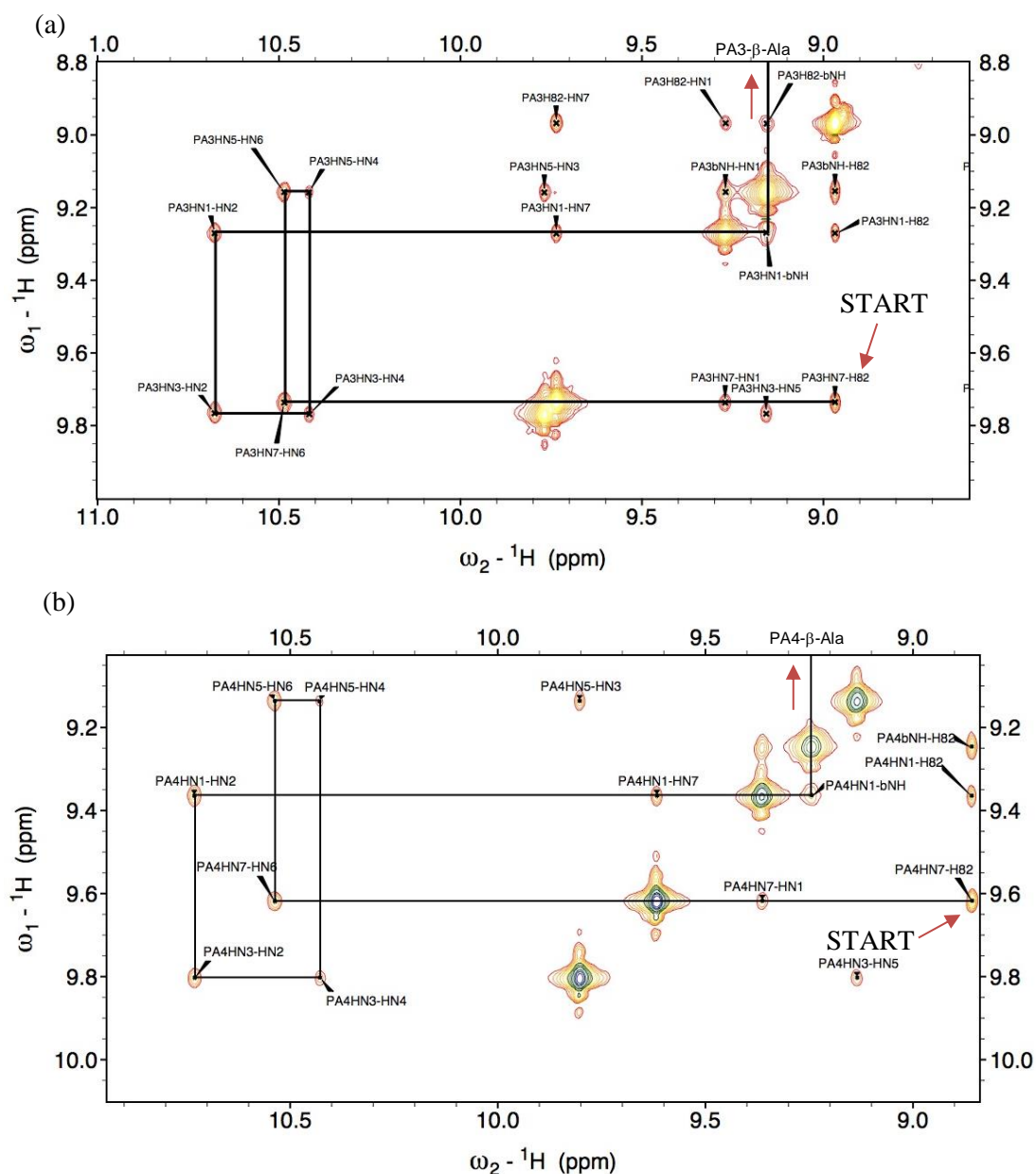


Figure 4.19 nOe walk of the amide protons of (a) **PA3** and (b) **PA4** in complex with **ODN1**.

In the case of **PA1**, a similar sequential assignment was carried out after identification of resonances Im8(H4-H5) (Figure 4.20).

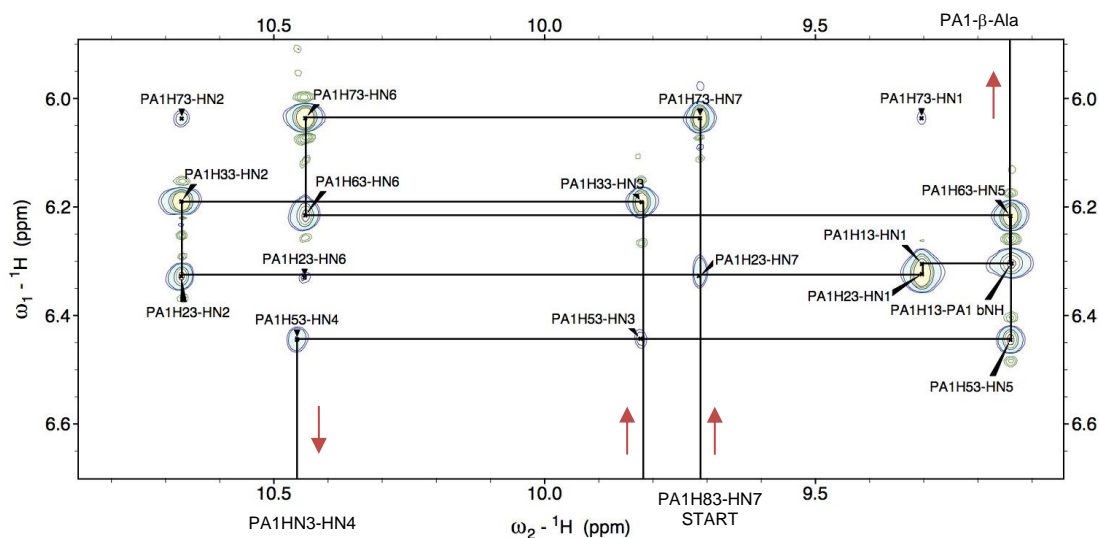


Figure 4.20 nOe walk of **PA1** in complex with **ODN1**.

Finally, TOCSY and COSY NMR data were used to assign the alkyl protons of the γ -turn, β -alanine and Dp moieties of each polyamide. Each of the methylene groups appeared as diastereotopic protons, with a distinct chemical shift originating from the different surroundings of each proton when in complex with dsDNA.¹² Due to considerable overlap the methyl groups of Dp could not be unambiguously identified.

4.3.4 Chemical shift analysis of binding mode of PA·ODN1 complexes

At first, the binding modes of **PA1** and **PA3-4** were examined by looking at the perturbations of the chemical shifts of **ODN1** resonances, which provide information about which protons are located in proximity to the ligand. Given the planar and aromatic nature of the polyamides, the larger perturbations originate from the ring currents characteristic of aromatic systems, which induce either an upfield or downfield shift, dependent upon whether the proton in analysis is coplanar with the ring (deshielded) or sits above or below the ring (shielded) (Figure 4.21).^{12,239}

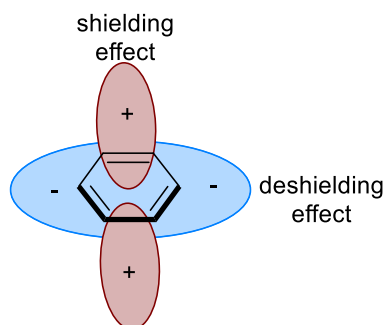


Figure 4.21 Schematic representation of ring current effects originated by the aromatic rings of the polyamides.

For all the complexes, perturbation of the H4' proton could be easily depicted from [^1H , ^{13}C] HSQC (Figure 4.23) and [^1H , ^{31}P] COSY (Figure 4.23, Appendix 136-138). Strong upfield perturbation was observed for H4' of nucleotides located in the central 5'-ATGTACA sequence, which is a diagnostic marker for minor groove binding.^{12,14,106}

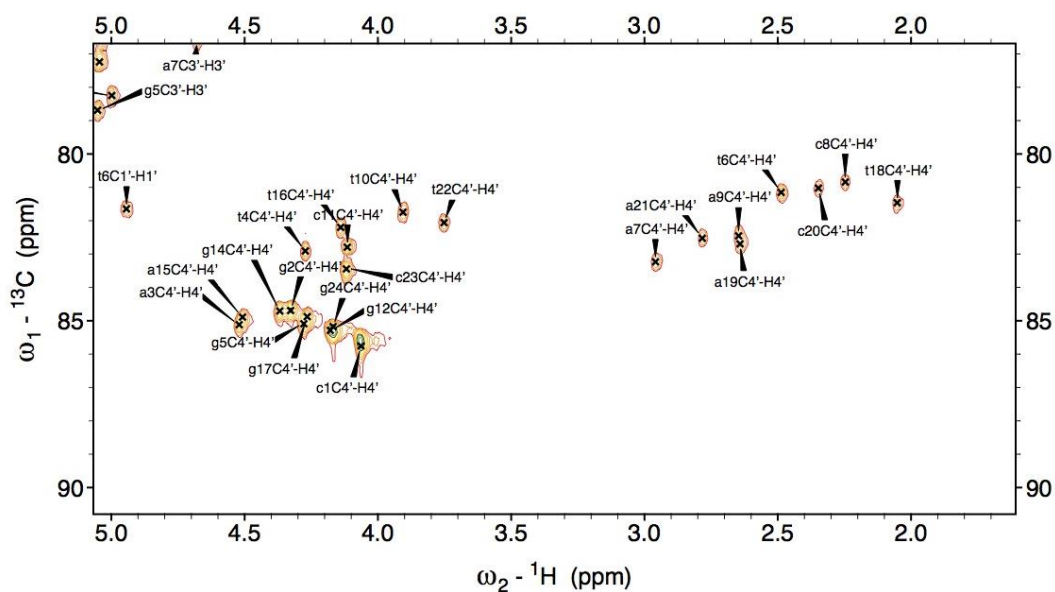


Figure 4.22 Selected region of the [^1H , ^{13}C] HSQC of PA4-ODN1.

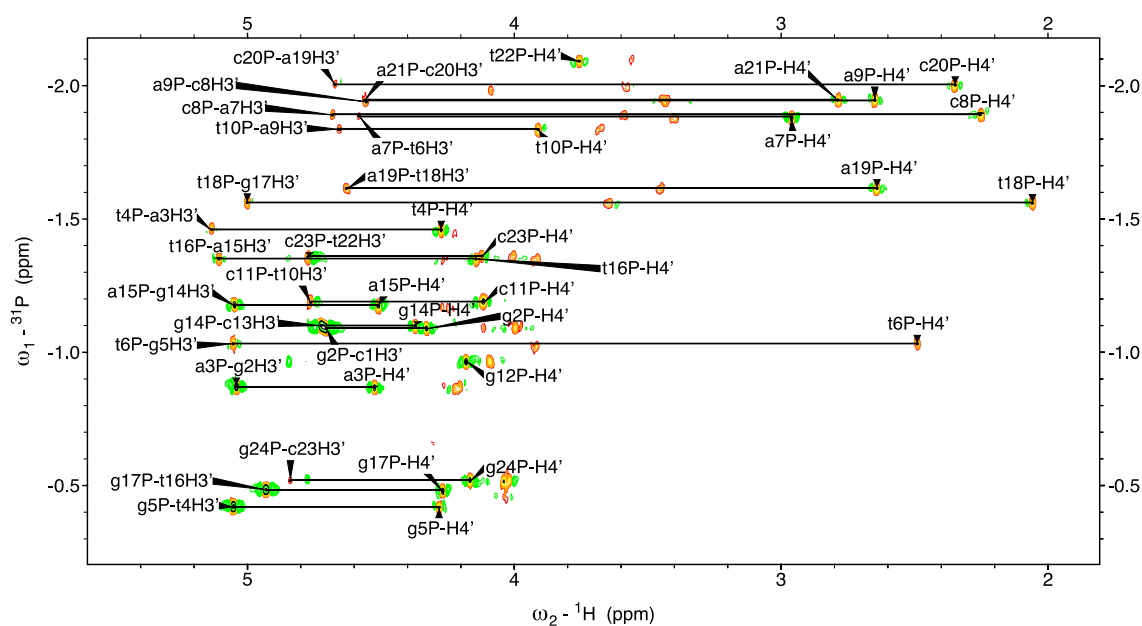


Figure 4.23 [^1H , ^{31}P] COSY spectrum of the **PA4-ODN1** complex.

Upfield chemical shifts were also observed for the H1' protons of the 5'-ATGTACA sequence and the imino protons of the central T6 and T18, which provided further indication of minor groove binding by all the polyamides. In contrast, the downfield chemical shifts observed for A7(H2) and A19(H2) relative to the unbound **ODN1** suggested that the position of this atom is in alignment with one of the aromatic rings of the polyamide (Appendix 113-131). ^{31}P NMR spectra of the polyamide·dsDNA complexes revealed broadening of the ^{31}P resonances with downfield shifts of G5(P) and G17(P), and upfield shifts of A9(P) and A21(P); these are noteworthy features relative to free **ODN1** and suggest relevant distortion of the phosphodiester backbone at these positions (Figure 4.24).

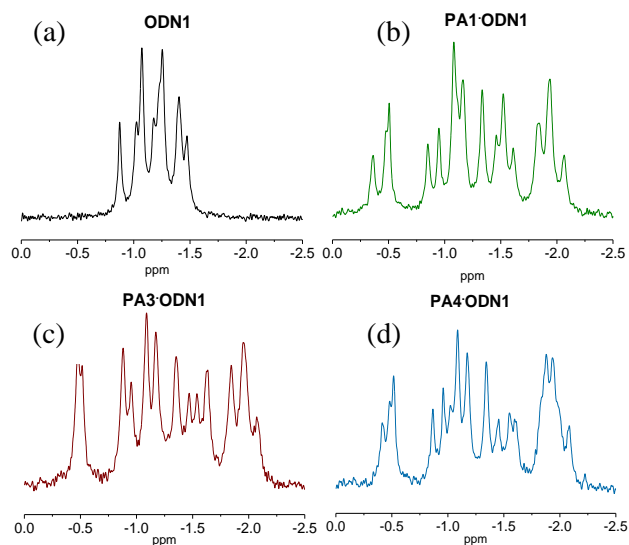


Figure 4.24 ^{31}P - NMR spectra of (a) free **ODN1**, (b) **PA1·ODN1**, (c) **PA3·ODN1** and (d) **PA4·ODN1** complexes.

The chemical shift perturbations of **PA3·ODN1** and **PA4·ODN1** complexes reveal subtle, but distinct differences in DNA duplex distortion relative to the **PA1·ODN1** complex. For example, a significant upfield shift of H4' (T6 and A7) and imino G5(H1) protons in the **PA1·ODN1** complex is observed relative to **PA3·ODN1** and **PA4·ODN1** (Figure 4.25a). In the case of T6(H3) and T18(H3) imino protons, a more significant upfield shift is evident for the Nt-containing polyamides **PA3·ODN1** and **PA4·ODN1** relative to **PA1·ODN1** (Figure 4.25c). Small differences in chemical shifts were also observed between G5(H1') and T6 (H1'), which were found more downfield for **PA1·ODN1** compared to **PA3·ODN1** and **PA4·ODN1** (Figure 4.25b). Further distinct chemical shift changes for G5(P), T6(P) and A7(P) were observed across all three complexes, which suggested different levels of phosphodiester backbone distortion imparted by each polyamide (Figure 4.25d).

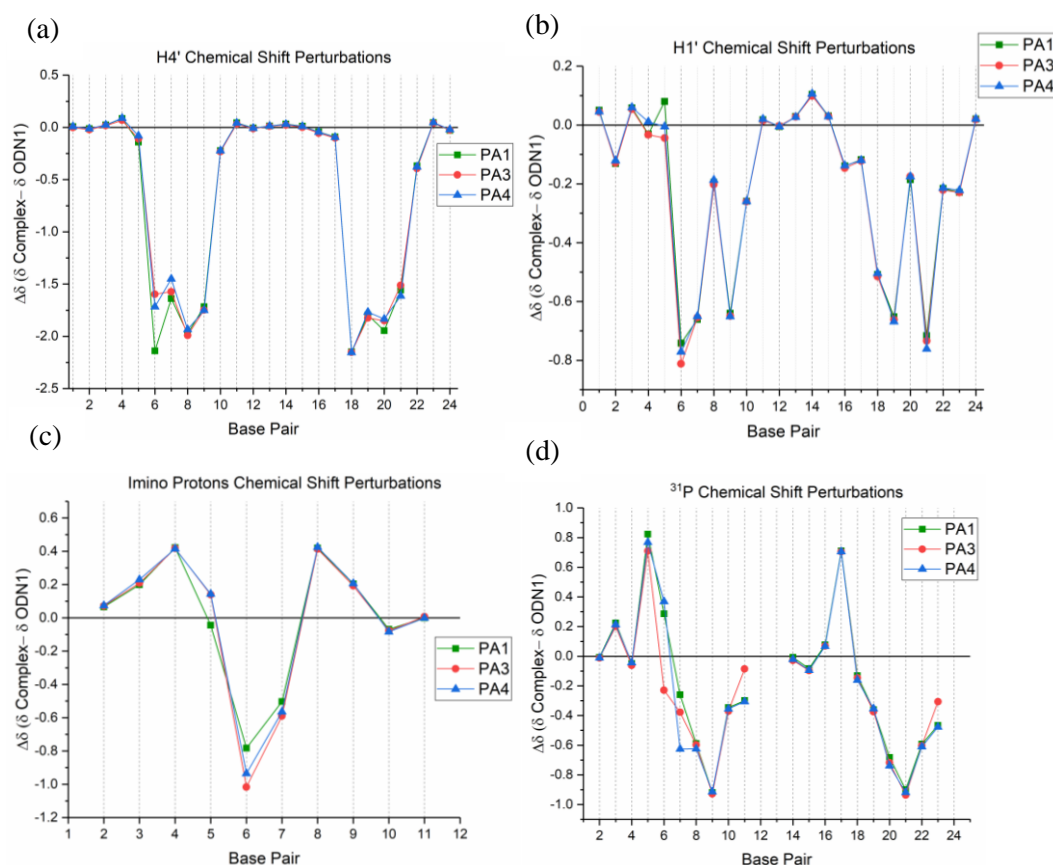


Figure 4.25 Chemical shift perturbation plots of (a) H4', (b) H1', (c) imino and (d) ^{31}P of ODN1.

4.3.5 Comparative nOe analysis of the polyamides binding mode when in complex with ODN1

Further information on the binding modes of the polyamides was obtained from the nOe correlations observed between the ligands and ODN1. Strong nOe correlations between amide NH, Py(H3) and H1' sugar protons of the adjacent base were used to confirm the position of the polyamide relative to the ODN1 sequence. nOe cross-peaks from Nt(H2) of both Nt building blocks to the β -alanine and Py1 protons in PA3·ODN1 and PA4·ODN1 suggest that both Nt-containing polyamides bind to their target sequence in a hairpin conformation (Figure 4.26).

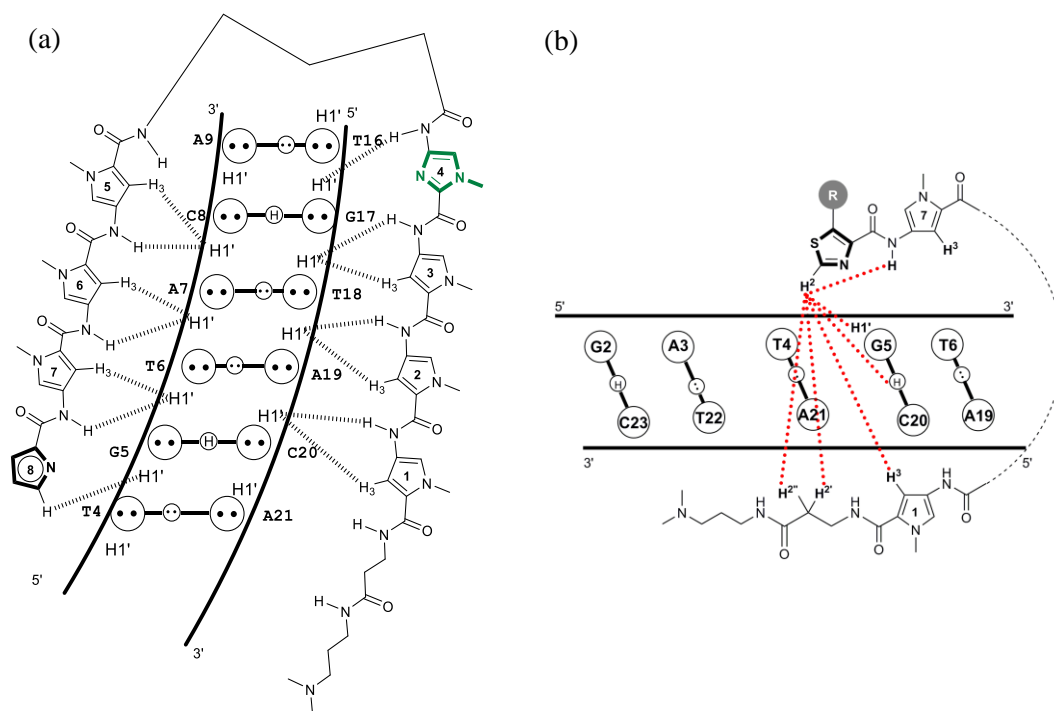


Figure 4.26 (a) Relevant inter-residue nOe cross-correlations observed for **PA1-ODN1**, **PA3-ODN1** and **PA4-ODN1** complexes. (b) Schematic representation of strong inter and intra molecular nOe cross-correlations of Nt(H2) (**PA3** and **PA4**) in complex with **ODN1**.

The nOe data also provided insight into the nature of Nt-binding to G5 in the minor groove. nOe correlations from Nt(H2) of **PA3** and **PA4** to G5(H1') and the G5(N2) exocyclic amine suggest that both Nt building blocks have the endocyclic N3 directed towards the floor of the minor groove. The presence of a hydrogen bond between Nt(N3) and the exocyclic G5(N2) amine is implied from the slower exchange rate of the exocyclic N-H of G5(N2) not involved in Watson-Crick base-pairing (G5(H22)) and nOe cross-correlations from G5(H22) to G5(H1) and T6(H3), all of which are absent in the NMR data for the free **ODN1** duplex (Figure 4.27). Relevant differences in the chemical shift of G5(H22) in **PA1-ODN1** (7.87 ppm) and Nt polyamides (**PA3-ODN1** 6.76 ppm, and **PA4-ODN1** 6.59 ppm) indicate that the presence of the endocyclic sulphur in Nt building blocks influences Nt-G5(N2) pairing.

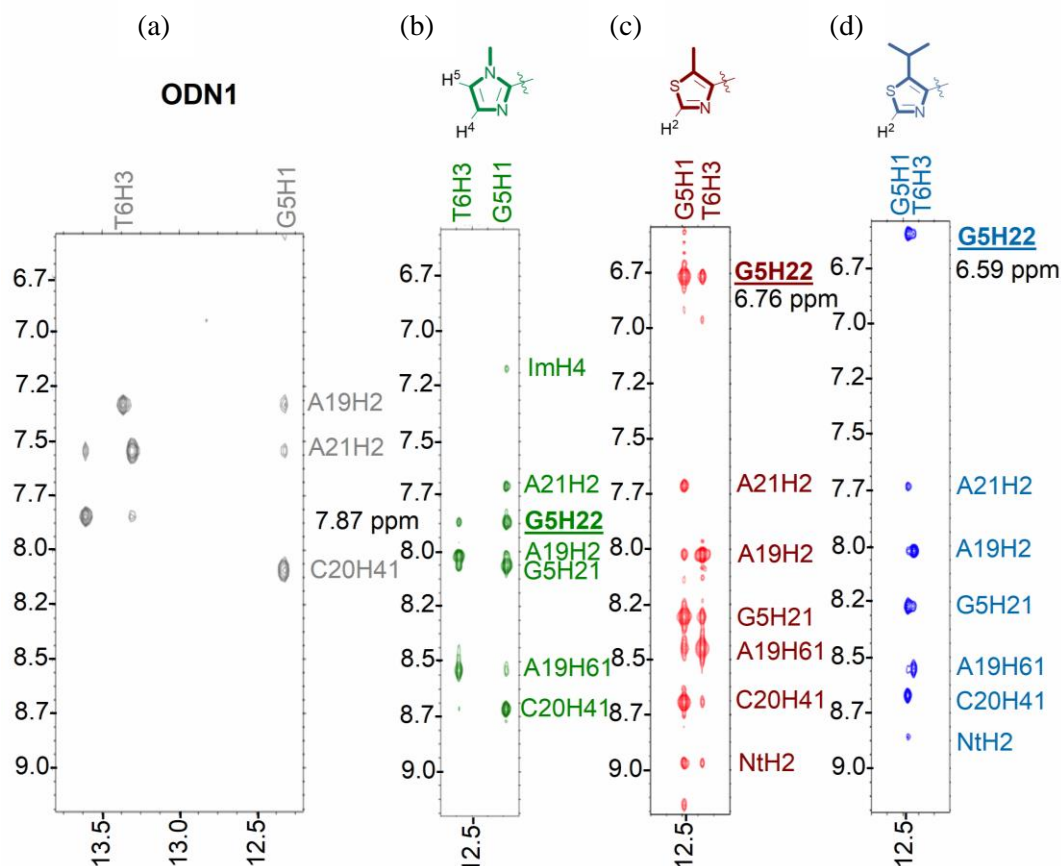


Figure 4.27 Strip plots from 2D [^1H , ^1H] NOESY NMR data of (a) free **ODN1**, (b) **PA1·ODN1**, (c) **PA3·ODN1** and (d) **PA4·ODN1**.

Taken collectively, the NMR data of each polyamide-dsDNA complex correlates with minor groove binding in a hairpin conformation with the implied formation of a hydrogen bond between the exocyclic amine of G5(N2) (*i.e.*, G5(H22)) and the endocyclic N3 of Im8 and both Nt building blocks.

4.3.6 NMR-restrained molecular dynamics of PA·ODN1 complexes

NMR-derived 3D structures of free **ODN1** and **PA1-3-4** complexes were obtained by converting the nOe intensities of not overlapping peaks into distance restraints using MARDIGRAS.^{106,157,238} The resulting distances were initially applied with a force constant of 10 kcal/mol Å^2 to a starting model obtained by manual docking of the ligand into the dsDNA

according to the experimentally observed nOe contacts. Initial simulations were performed using an implicit solvent and the resulting structures were used for a second round of distance calculations. This process was repeated using explicit solvent conditions. The ten most representative geometric conformational states of each model obtained from molecular dynamics simulations were then minimised and chosen as representative ensembles (Figure 4.28, Table 4.2).

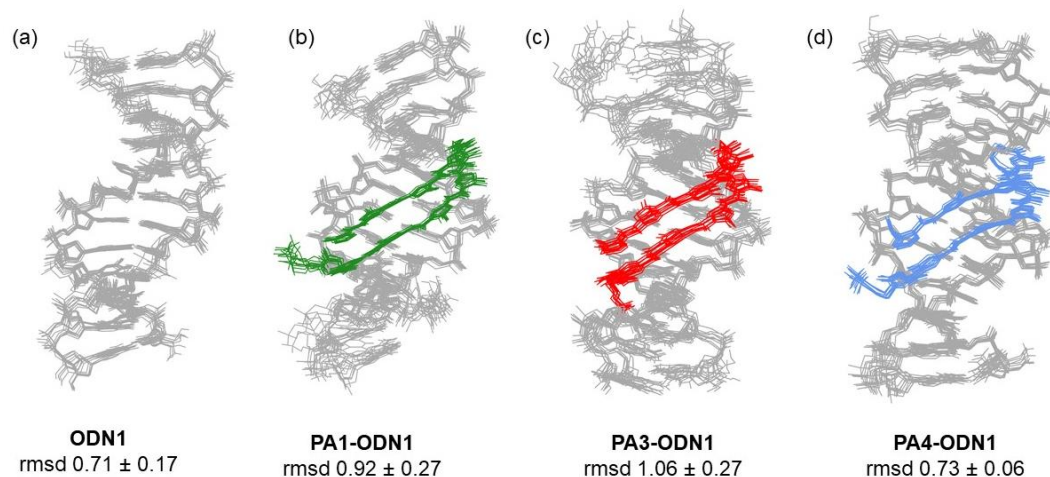


Figure 4.28 Minimised MD ensembles of (a) **ODN1** and complexes (b) **PA1-ODN1**, (c) **PA3-ODN1**, (d) **PA4-ODN1** (rmsd values were calculated considering all atoms towards the most representative conformer).

Table 4.2 Statistics and outputs of NMR restraints used for MD model determinations.

NMR	Total restraints ^a	DNA	Polyamide	Inter-residue	Rmsd (bond) ^b	Rmsd (angle) ^b	Penalty (kcal) ^b
ODN1	355	355	-	-	0.0069	4.847	774.759
PA1-ODN1	301	137	48	116	0.0065	4.569	586.407
PA3-ODN1	327	146	62	119	0.0072	4.520	476.145
PA4-ODN1	356	141	83	132	0.0073	4.584	678.406

^a includes distance and hydrogen bonds restraints

^b output values that refers to the last frame of 500 ps NMR restrained MD

The structural analyses were carried out on average minimised structures obtained from the selected ensemble. All three polyamides retain a ring-over-ring conformation in each complex, which has previously been observed for hairpin and cyclic polyamides.^{12,106,133} However, the Nt-containing polyamides (**PA3-4**) adopt a distinct dsDNA binding geometry compared to **PA1**. In the case of **PA1-ODN1**, the smaller Im8 building block induces a more pronounced

curvature within the minor groove relative to **PA3-ODN1**, which results in shorter distances between Im8(N3) and Py6/Py7(NH) and their **ODN1** hydrogen bonding partner (*i.e.*, G5H22, T6(O2), A7(N3); Figure 4.29a-b). In the **PA4-ODN1** complex, the steric clash between the isopropyl group and (Me)Py1 prevents close contact between Nt(N3) and the exocyclic G5(N2) amine (Figure 4.29c).

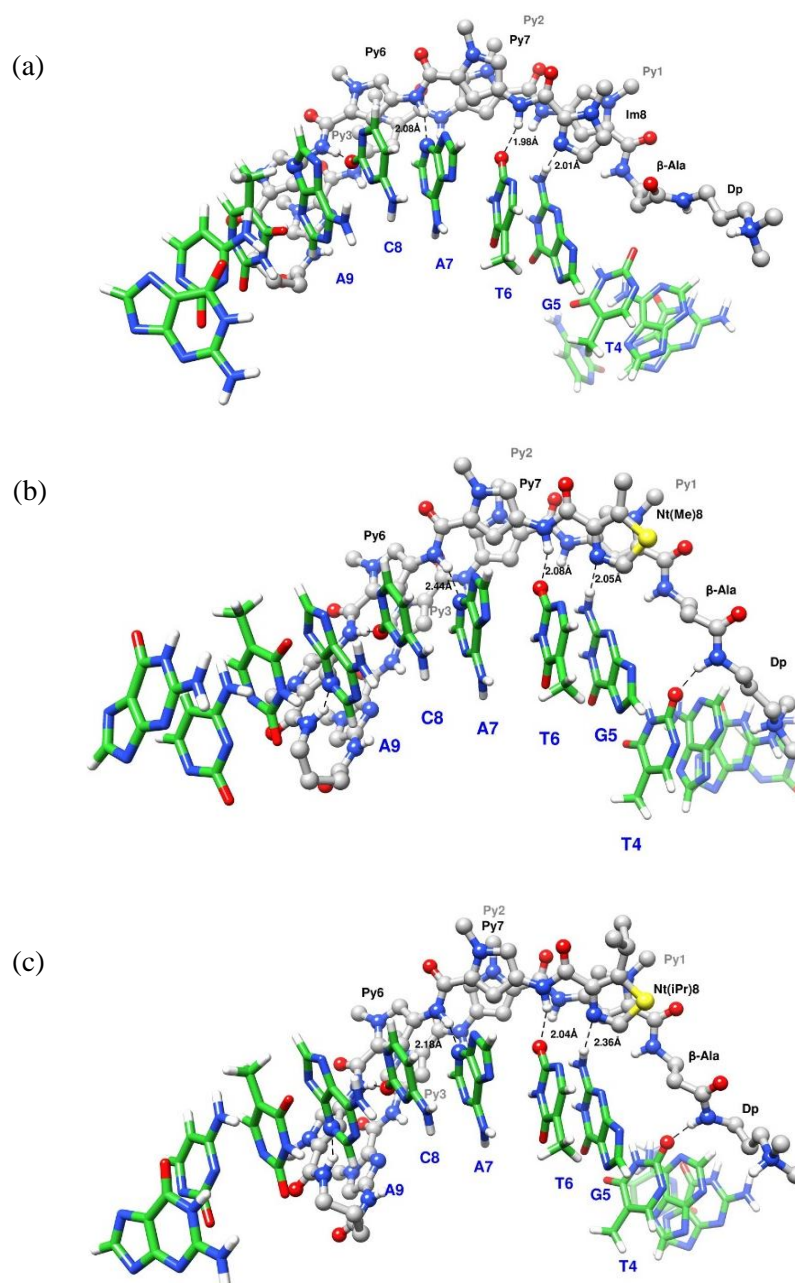


Figure 4.29 View of the curvature and anticipated hydrogen-bonding pattern between (a) **PA1-ODN1**, (b) **PA3-ODN1** and (c) **PA4-ODN1**.

As a result of these differences in binding mode, an altered inclination is observed for the base pairs targeted by the polyamides. This effect is more pronounced in the **PA1·ODN1** complex compared with **PA3·ODN1** and **PA4·ODN1**, which induces evident inclination of the DNA axis.

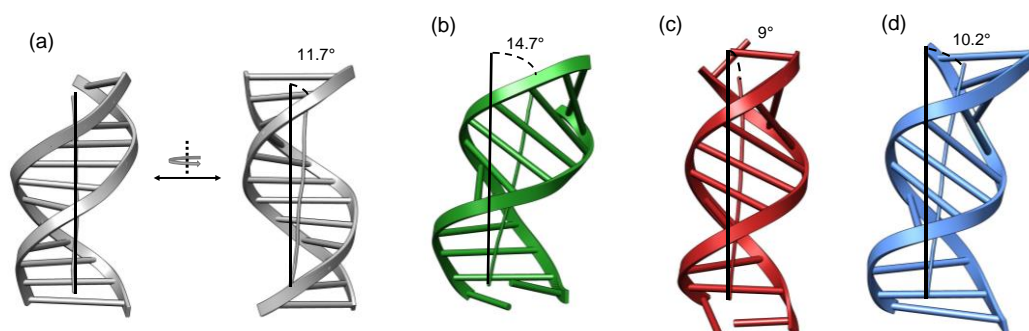


Figure 4.30 Details of DNA axis inclination of (a) **ODN1**, (b) **PA1·ODN1**, (c) **PA3·ODN1** and (d) **PA4·ODN1** (base pairs at the edges of the sequence were excluded from the calculation).

The minor and major groove geometries of each complex reveal a number of distinct structural features. **PA1·ODN1** and **PA3·ODN1** complexes are characterised by an enlarged minor groove relative to **PA4·ODN1**. This is most likely due to deeper minor groove penetration of the less hindered Im8 and ^{Me}Nt8 (Figure 4.31). An unexpected observation is the difference in the major groove width between **PA1·ODN1** and **PA4·ODN1**. A more pronounced narrowing of the major groove is observed at the site of ^{iPr}Nt binding to G (*i.e.*, Nt(N3)·G5(N2)), changing from 22.4 Å in the free duplex (*i.e.*, **ODN1**) to 12.2 Å in **PA4·ODN1** (Figure 4.31). In contrast, only a slight narrowing (21.0 Å) of the major groove is observed for the **PA1·ODN1** complex, which is consistent with crystal structural studies of a cyclic polyamide-dsDNA complex.^{132,133} For the **PA3·ODN1** complex, an intermediate change (16 Å) in the major groove width relative to **PA4·ODN1** suggests that the increased steric bulk of the isopropyl substituent in **PA4·ODN1** plays a significant role in inducing major groove compression of the DNA duplex.

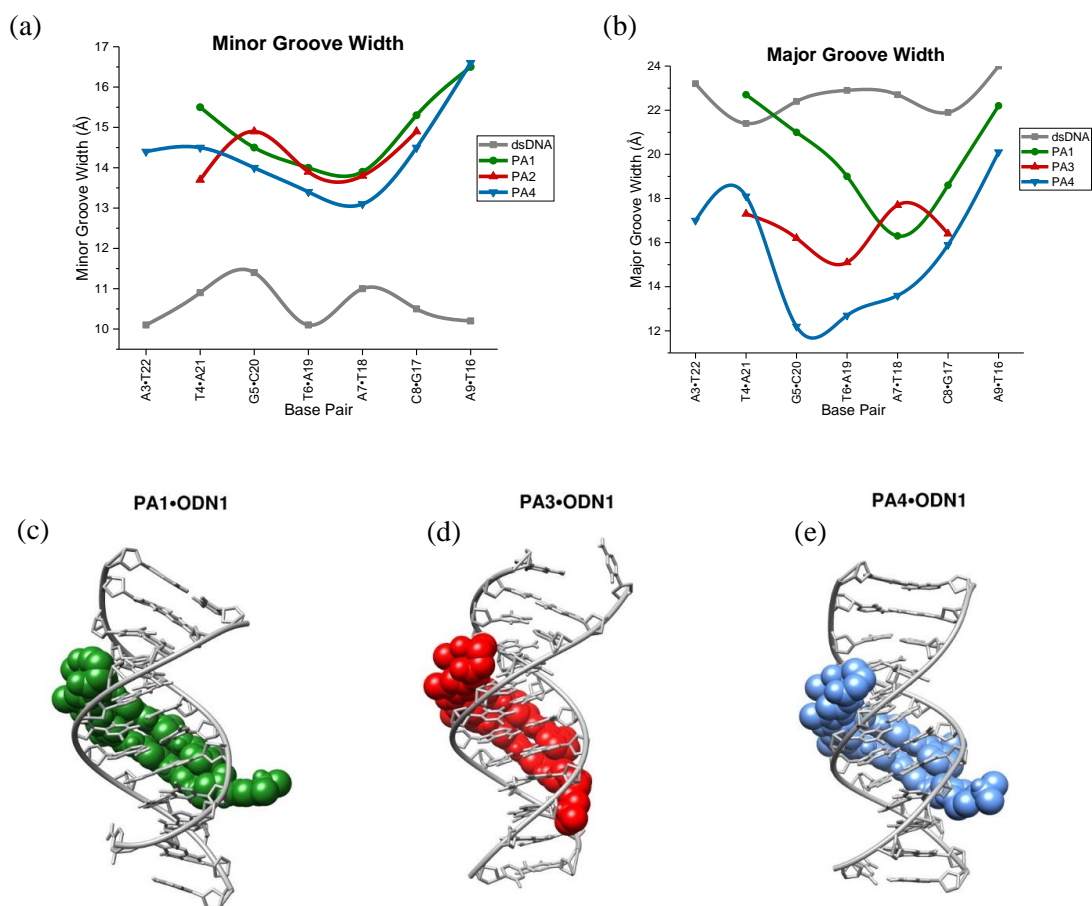


Figure 4.31 (a) Minor Groove and (b) Major Groove width of ODN1 and PA1-ODN1, PA3-ODN1, PA4-ODN1. Major groove view of (c) PA1-ODN1 (d) PA3-ODN1 (e) PA4-ODN1.

Structural changes to the Watson-Crick pairing profile in each polyamide·DNA complex also differ. In **PA1-ODN1**, the larger propeller twist in T4·A21 and increased buckle of T6·A19 contribute to inducing a widening of the major groove, which is typified by the larger shift changes of the imino and H4' protons (see Appendix 140-156). In the case of the **PA4-ODN1** complex, the more sterically encumbered ^{iPr}Nt induces significant changes in the χ torsional angle (O4'-C1'-N9-C4) of the G5 (-87.9°) nucleosidic bond relative to **PA3-ODN1** (-103.2°) and **PA1-ODN1** (-117.1°) (Figure 4.32). Finally, the significant deviations in the ³¹P resonances of G5(P), T6(P) and A7(P) in **PA4-ODN1** relative to **PA1-ODN1** are indicative of significant changes in the phosphodiester bond angles ζ (C3'-O3'-P-O5') and ϵ (C4'-C3'-O3'-P) located in close proximity to the site of Nt-binding (Figure 4.32).

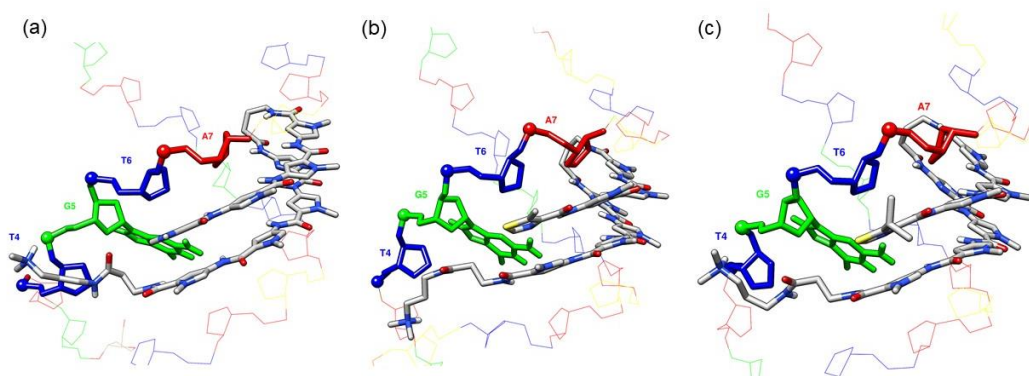


Figure 4.32 Details of the phosphodiester backbone perturbation induced over T4-G5-T6-A7 by (a) **PA1**, (b) **PA3**, (c) **PA4** (P atoms are highlighted as spheres. For clarity, all base rings are hidden with the exception of G5).

4.3.7 Rationalisation of the differences in duplex distortion of Nt-containing polyamides bound to the target ODN1

These experiments were designed to probe how structural changes to the *N*-terminal heterocycle of a hairpin polyamide affects the dsDNA binding mode. Despite the slower association rate, UV and fluorescence thermal stabilisation suggest that the replacement of an *N*-terminal Im unit for an Nt heterocycle enhances overall dsDNA binding, but at the expense of sequence selectivity for the target G nucleotide. Our working hypothesis is that the overall differences in hydrophobicity, hydrogen bond strength and steric bulk between Nt and Im heterocycles are contributing factors. Underpinning all of these is the replacement of the nitrogen N1 of Im with a sulfur (S1) atom, which in Nt is also oriented differently, occupying the same position as the smaller carbon (C5) atom in the Im unit. This key structural change appears to contribute to a general loss of G-selectivity for **PA2-3**, in which there is minimal steric bulk in the 5-position relative to the archetypal hairpin **PA1**.

This has implications in, for example, weaker hydrogen bonding between the exocyclic N-H of G5(N2) (*i.e.*, G5(H22)) and the N3 of Nt compared to the more basic ImN3, implied by a downfield shift of G5H22 for the Im-containing polyamide (**PA1-ODN1**, 7.87 ppm) compared to Nt polyamide (**PA3-ODN1** 6.76 ppm, and **PA4-ODN1** 6.59 ppm). The higher dsDNA

melting stabilisation observed for Nt polyamides (**PA2-4**) compared to the archetypical Im polyamide (**PA1**) is attributed to a stronger overall hydrophobic interaction for the Nt-containing **PA2-4**, which appears to compensate for the weaker hydrogen bonding interaction with G5(H22), and is most likely responsible for an increase in non-specific dsDNA binding for **PA2** and **PA3**. However, the installation of an isopropyl group (*i.e.*, **PA4**) increases G selectivity with respect to **PA2-3**. Whilst these changes in G selectivity cannot be rationalised by this study alone, the steric effect of the isopropyl group in **PA4** appears to bias the binding towards wider minor groove GC regions.

A structural rationalisation for the observed dsDNA binding differences of **PA3-4** relative to **PA1** would likely involve an interplay between the geometrical differences of the Nt and Im heterocycles, and their inherent differences in their respective non-covalent interactions with dsDNA in the minor groove. For example, the stronger hydrogen-bonding as well as the smaller overall dimension of the Im8 unit in **PA1** results in deeper penetration into the dsDNA minor groove. From a topological perspective, this induces an overall increase in dsDNA bending in the **PA1·ODN1** complex (14.7°), which concomitantly compresses the major groove around the central T6·A7·C8 region of the DNA duplex. In the case of **PA3·ODN1**, the weaker hydrogen-bonding interaction, as well as the steric effect of the sulfur in ^{Me}Nt reduces the overall bending (9°), which in turn compresses the major groove over a larger DNA region (T4·G5·T6·A7·C8). The 5-position substituent of the ^{iPr}Nt building block of **PA4·ODN1** only partially affects the overall DNA bending (10.2°). A unique characteristic of the ^{iPr}Nt unit is that the larger steric bulk induces a more extensive compression of the major groove (in particular over G5), which is characterised by an increased hydrogen bond distance between the N3 of ^{iPr}Nt and G5(H22) arising from reduced minor groove penetration. Taken collectively, the steric effect of the isopropyl Nt substituent in **PA4** enhances the selective recognition of Nt-containing hairpin polyamides by compensating for the weaker hydrogen

bonding interaction with the exocyclic N2 of G with an increase in the overall hydrophobicity of this *N*-terminal building block.

4.4 Summary

This chapter described the NMR structural characterisation of **ODN1** in complex with **PA1**, **PA3** and **PA4**. The chemical shift perturbations of **ODN1** and nOe contacts originated from the binding suggest that all the polyamides bind to the target sequence 5'ATGTACA in the minor groove in a hairpin conformation. Similar to Im(N3), Nt(N3) is facing towards the groove, hydrogen bonding with G(N2). However, the combination of lower hydrogen bonding strength, larger dimension and increased hydrophobicity results in evident differences in the structural distortion imparted on the dsDNA structure by Nt-containing polyamides. In particular, a larger compression of the major groove is observed when Nt is functionalized with an isopropyl group at position 5 (*i.e.*, **PA4**). While further structural analysis using mismatch sequences is necessary to gain further information on the role of the isopropyl group in sequence selectivity, this study provides evidence of a unique binding mode for an 8-ring polyamide incorporating ⁱPrNt at the N-terminus, which has possible implications for its biological activity.

4.5 Experimental

Two separated batches of HPLC purified self-complementary d(CGATGTACATCG)₂ (**ODN1**) were purchased from Eurogentec. Each sample (1.56 μmol) was dissolved in 1.2 mL of 100 mM phosphate-buffered water (pH = 7.4) containing 0.1 mM deuterated trimethylsilylpropionic acid sodium salt-*d*₄, divided in two aliquots (600 μL each) and lyophilised to dryness. These samples were re-dissolved in 600 μL of 90% H₂O with 10% D₂O or 99% D₂O. The solutions were transferred into two separate 5 mm high-precision NMR

tubes for **ODN1** data acquisition. Following data acquisition, the two 99% D₂O samples were lyophilised and re-dissolved in 10%D₂O/90%H₂O. Concentrated solutions of the **PA1/3/4** (16 mM in no more than 10% d₆-DMSO in water) were added in small aliquots to their respective samples. After each addition of polyamide to the buffered solution containing **ODN1**, 1D ¹H NMR spectra were acquired until the end point of the titration was reached as determined by inspection of the imino proton resonance region of the spectra. Following acquisition of the full set of data in 10%D₂O/90%H₂O, the samples were freeze-dried and re-dissolved in D₂O to complete the acquisition of all data sets.

NMR Data Acquisition

NMR data were typically acquired on a Bruker AVANCE-II⁺ 600 NMR spectrometer operating at a magnetic field strength of 14.1 Tesla (600.13 MHz for proton resonance) under 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 probehead [TBI-z] equipped with an actively shielded z-gradient coil for delivery of pulsed field gradients was manually tuned for a [¹H, ¹³C, ³¹P] frequency configuration on each sample 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) ¹H NMR data were acquired using either presaturation (pulse program zgpgpr) or double pulsed field gradient spin-echo (pulse program zgesgp) to eliminate the residual solvent resonance. Data were typically acquired for 90% H₂O samples with between 64 and 256 transients into 32 K data points (acquisition time: 1.09 s) over a frequency width equivalent to 20.0276 ppm centred 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 μs across the samples studied. Data acquired for D₂O-dissolved samples were acquired over a frequency width equivalent to 12.0166 ppm (acquisition time: 2.27 s).

One-dimensional (1D) ^{31}P - $\{^1\text{H}\}$ NMR spectra were acquired over a frequency width of 3 ppm into 512 data points (acquisition time: 0.35 s) centred at $\delta^{31}\text{P} = -1.185$ ppm with the decoupler offset set to $\delta^1\text{H} = 4.00$ ppm. Between 512 and 4096 transients were acquired with a relaxation delay of 0.5 s.

Proton inversion recovery profiles were measured on one sample to gauge the requirement for relaxation delays during the acquisition of 2D NOESY data sets used in structure calculations. The largest T_1 values were measured as 5.65 s (for adenine H2 protons). Relaxation delays were typically set to 2.0 s for most data acquisitions with the exception of NOESY data sets whose relaxation delays were typically set to 7 s to allow at least one T_1 period of magnetization recovery without excessively extending 2D NOESY data acquisitions.

Two-dimensional (2D) NMR data sets were acquired as follows for samples dissolved in D_2O . Proton-only data were acquired over frequency widths equivalent to 12 ppm (7194 Hz) in both F2 and F1 centred 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 t_1 increments for acquisition times (aq) of 285 ms (ω_2) and $\text{aq}_{\text{max}} = 71$ ms (ω_1) and a total data accumulation time of 10 hrs. 2D TOCSY data (pulse program: dipsi2phpr) were acquired into 2048 data points (ω_2 acquisition time = 142 ms) with 8 transients for each of 512 States-TPPI t_1 increments ($\text{aq}_{\text{max}} \omega_1 = 35.4$ 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. 2D [^1H , ^{31}P] correlation data were acquired for assignment purposes only according to the constant-time procedure (pulse program: nahpctco).²³⁴ Data were acquired over frequency widths equivalent to 7 ppm (ω_2) and 2 ppm (ω_1) centred at $\delta^1\text{H} = 4.701$ ppm and $\delta^{31}\text{P} = -1.185$ ppm respectively with 128 transients for each of 128 State-TPPI t_1 increments and a relaxation delay of 1.5 s to give a total data accumulation period of 2.75 h. 2D [^1H , ^{13}C] correlation data were acquired using echo/anti-echo-TPPI data acquisition schemes with sensitivity improvement, with (pulse program: hsqcedetgpcsp.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 t_1 increments into 2048 data points for frequency widths equivalent to 10 ppm (F2) and 170 ppm (F1) with 128 transients per t_1 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 t_1 increment for a total data accumulation time of 20 h. For samples dissolved in 90% H₂O, 2D NOESY data were acquired with excitation sculpting for solvent suppression (pulse program: noesyegpph) using State-TPPI for 1024 t_1 increments and 4096 data points over F2 and F1 frequency widths equivalent to 25 ppm centred at the solvent frequency resonance. Data were acquired with 16 transients per t_1 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 for data assignment and reduction in preparation for molecular structure calculations.

Preliminary Molecular Dynamics

Preliminary molecular dynamics for distance restraint calculations of free **ODN1** and the complexes were run in explicit solvent using the SANDER module of AmberTools16.²⁴⁰ The starting model of **ODN1** was generated using the UCSF CHIMERA program.²⁴¹ The structures of each polyamide•**ODN1** complex were generated using an X-ray-derived structure of a cyclic polyamide•dsDNA complex as the starting model (PDB ID 3OMJ).¹³² Input files (prmtop and inpcrd) were generated using the nucleic acid force field (LEAP module) ff99bsc1 and GAFF for **ODN1** and ligands respectively. 22 and 21 Na⁺ ions for the free **ODN1** and the complexes were added to neutralize the negative charge of the resultant complexes. 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 interaction were applied with the particle mesh Ewald (PME) method to account for long range electrostatic interactions. Initial minimisations of 1000 steps

(500 steps of steepest descent and 500 steps of conjugate gradient minimization) were performed on each complex, keeping the solute fixed with a 500 kcal/mol·Å² force constant. The entire system was then minimised 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 resulting 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 evaluated for general consistency with the nOe data and used for distance calculations

NMR Distance Restraints

In all cases, the distance restraints were obtained from 2D [¹H-¹H] NOESY (250 ms). NMR-FAM SPARKY was used for the assignment process.²⁴² Non-overlapping peaks were integrated using the Gaussian fit to generate the intensity file for MARDIGRAS.²³⁸ The RST files obtained from the preliminary molecular dynamics were converted to pdb file format for use in MARDIGRAS through CORMA.IN employing an isotropic model and relative error of 10%. The correlation times used were roughly established in each case by checking MARDIGRAS calculation outputs from 1 ns to 8 ns. Absolute un-normalised noise was defined as a fraction of the smallest peak that could be integrated in the spectrum. All models were calculated using the Methyl jump 3 model. MARDIGRAS was run for 50 cycles. The resulting .dst files were associated with RAND-RESTR to generate an average distance file. These coordinates were converted to restraints to be used in AMBER employing M2AHOMO or MARDI2AMBER and filtering out selected out-of-range distances. The force constants applied were 10 kcal/mol·Å 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.

NMR-restrained Molecular Dynamics

Initial molecular dynamics simulations were performed using the generalised Born implicit-solvent model. After initial minimisation (500 total steps, 250 steepest descent and 250 conjugated gradient), 100 ps MD were calculated under isothermic conditions. NMR restraints were gradually increased over the first 20 ps and kept constant over the remaining 80 ps. After error checking, the RST file was used for a second cycle of MARDIGRAS distance calculations. The second round of molecular dynamics was carried out in explicit solvent using the same condition employed in the preliminary simulations. Newly generated NMR restraints were gradually applied over the first 20 ps of the simulation and kept constant for the remaining 80 ps. The resulting RST file was then used for the final cycle of MARDIGRAS calculations and the obtained restraints used for a third MD simulation. Production runs of 500 ps were obtained during this stage. The production runs were analysed using the CPPTRAJ module of AMBERTools16. Clustering of the last 400 ps was carried out using UCSF Chimera to produce ten representative structures, which were minimised and deposited in the ProteinDataBank (PDB accession codes 5OCZ, 5OE1, 5ODF and 5ODM). An average minimised structure obtained from this ensemble was obtained using the CPPTRAJ module of AMBERTools16 and used to calculate the DNA parameters taking a complementary approach using Curves+ and 3DNA software.^{243,244}

Chapter 5

Future directions towards the development of polyamides as gene regulatory agents

This thesis describes the preparation and binding evaluation of 8-ring hairpin polyamides containing the novel Nt building block, which had been identified as a possible alternative to C-terminus and γ -turn modifications for improving the pharmacokinetics properties of Py-Im polyamides. The original objective was to develop analogues for the recognition of the ARE sequence 5'WGWWCW (where W = A/T), which has previously been targeted by Py-Im polyamides for the treatment of prostate cancer.^{131,134}

The initially hypothesised advantages resulting from the introduction of Nt into an 8-ring hairpin polyamide were:

- (i) Decreased toxicity associated with the Im building block.^{159,162,245}
- (ii) Enhanced polyamide lipophilicity without a significant increase in molecular weight, thus improving the cellular uptake and pharmacokinetics properties.
- (iii) Facile substitutions at position 5 of the Nt building block to modulate polyamide properties.

The general outcome of this work is that the synthesis, binding affinity and selectivity of polyamides containing Nt is contingent on the position of the Nt building block in the polyamide scaffold and its substitution at position 5.

The preparation of hairpin polyamides containing Nt in an internal position was complicated by the low reactivity of the Nt amine, which was unsuitable for solid phase synthesis. The preparation of polyamides containing Nt in an internal position was successful only when the Nt amine was coupled to an aromatic ring (*i.e.*, position 6). Preliminary dsDNA binding evaluation of this series of polyamides suggested a decreased binding affinity compared to the Im-containing parent compound. While G-selectivity was observed for all the polyamides, only the polyamide containing ^HNt (**PA6**), retained sufficient affinity. This suggests that substitutions at position 5 of Nt have a detrimental effect on the binding and thereby prevent the modulation of the hydrophobic character of the polyamide.

A different binding behaviour was observed for polyamides incorporating Nt at the N-terminus (position 8) of the polyamide scaffold which were prepared in moderate yields and as larger scales (20-30 mg) by solid phase synthesis. Through a combination of biophysical and biochemical methods, it was demonstrated that Nt-containing polyamides retain high binding affinity for dsDNA, however, with lower selectivity for G as compared to Im. Interestingly, the analogue containing the more encumbered isopropyl substitution at position 5 (**PA4**) showed the greatest G-selectivity among the Nt-containing polyamides. Structural studies showed that the ^{iPr}Nt-containing polyamide induced greater compression of the major groove compared to the parent Im building blocks, which may help to account for its selectivity. For these reasons, the ^{iPr}Nt-containing **PA4** has been identified as a potential candidate for targeting the ARE sequence 5'WGWWCW in prostate cancer cells.

Further developments of this work would include:

- (i) Exploring how hydrophobicity impacts cellular uptake and pharmacokinetic properties of linear and hairpin analogues containing ^{iPr}Nt in the terminal position of the polyamide scaffold.
- (ii) Understanding how the larger compression of the major groove impacts on the biological activity of Nt-polyamides in prostate cancer cell lines. As one of the main mechanisms of action of Py-Im polyamides has been associated with the allosteric disruption of AR·ARE recognition caused by the distortion of dsDNA induced by the binding of the polyamide, the potential of this larger perturbation could translate to enhanced cytotoxicity in prostate cancer cells.^{132,133}

Furthermore, the different binding signatures observed between ^{Me}Nt and ^{iPr}Nt suggest that distinct binding features are produced even by minor modification of the polyamide scaffold such as small aliphatic modifications. It is therefore possible to tune the pharmacokinetic and activity profiles, which are associated with structural perturbation of dsDNA, through defined modifications of the polyamide scaffold. As a consequence of this, further investigations could

focus on understanding the influence of large isopropyl substitution on N1 of the imidazole ring. A working hypothesis is that this building block would combine the enhanced G-selectivity of the Im ring with the larger distortion of the major groove induced by the isopropyl (Figure 5.1)

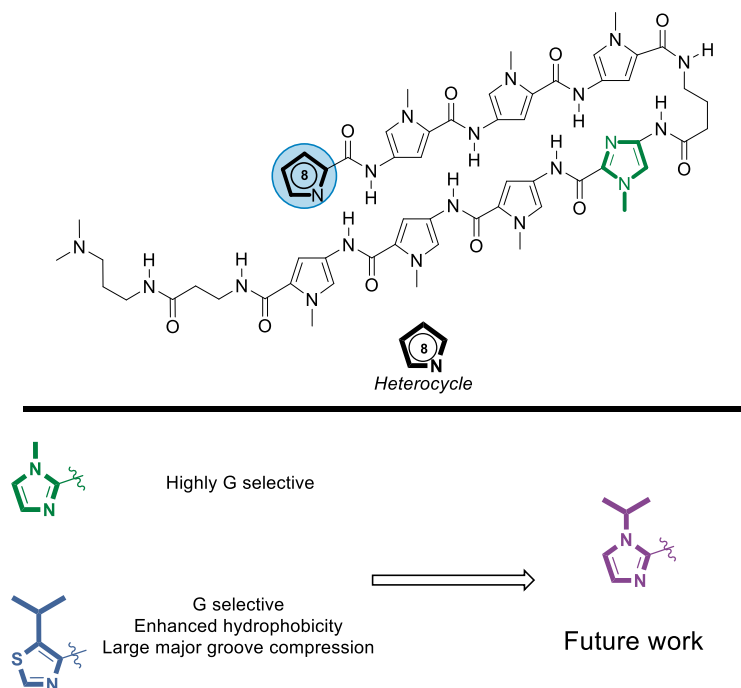


Figure 5.1 Schematic representation of the future directions of this project.

References

- (1) Rohs, R.; Jin, X.; West, S. M.; Joshi, R.; Honig, B.; Mann, R. S. Origins of Specificity in Protein-DNA Recognition. *Annu. Rev. Biochem.* **2010**, *79* (1), 233–269.
- (2) Dervan, P. B.; Edelson, B. S. Recognition of the DNA Minor Groove by Pyrrole-Imidazole Polyamides. *Curr. Opin. Struct. Biol.* **2003**, *13* (3), 284–299.
- (3) Dervan, P. Molecular Recognition of DNA by Small Molecules. *Bioorg. Med. Chem.* **2001**, *9*, 2215–2235.
- (4) Bewley, C.; Gronenborn, M.; Clore, G. M. Minor Groove-Binding Architectural Proteins: Structure, Function, and DNA Recognition. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 105–131.
- (5) Wei, D.; Wilson, W. D.; Neidle, S. Small-Molecule Binding to the DNA Minor Groove Is Mediated by a Conserved Water Cluster. *J. Am. Chem. Soc.* **2013**, *135* (4), 1369–1377.
- (6) Finlay, A. C.; Hochstein, F. A.; Sobin, B. A.; Murphy, F. X. Netropsin, a New Antibiotic Produced by a Streptomyces. *J. Am. Chem. Soc.* **1951**, *73*, 341–343.
- (7) Arcamone, F.; Penco, S.; Orezzi, P.; Nicoletta, V.; Pirelli, A. Structure and Synthesis of Distamycin A. *Nature* **1964**, *203*, 1064.
- (8) Baraldi, P. G.; Tabrizi, M.; Preti, D.; Fruttarolo, F.; Avitabile, B.; Bovero, A.; Pavani, G.; del Carretero, M. C. N.; Romagnoli, R. DNA Minor-Groove Binders. Design, Synthesis and Biological Evaluation of Ligands Structurally Related to CC-1065, Distamycin, and Anthramycin. *Pure Appl. Chem.* **2003**, *75* (2–3), 187–194.
- (9) Chen, X.; Ramakrishnan, B.; Sundaralingam, M. Crystal Structures of the Side-by-Side Binding of Distamycin to AT-Containing DNA Octamers d(ICITACIC) and d(ICATATIC). *J. Mol. Biol.* **1997**, *267* (5), 1157–1170.
- (10) Fede, A.; Billeter, M.; Leupin, W.; Wuthrich, K. Determination of the NMR Solution Structure of the Hoechst 33258-d(GTGGAATTCCAC)₂ Complex and Comparison with the X-Ray Crystal Structure. *Structure* **1993**, *1* (3), 177–186.
- (11) Davies, D. B.; Eaton, R. J.; Baranovsky, S. F.; Veselkov, A. N. NMR Investigation of the Complexation of Daunomycin with Deoxytetranucleotides of Different Base Sequence in Aqueous Solution. *J. Biomol. Struct. Dyn.* **2000**, *17* (5), 887–901.
- (12) Lamamie De Clairac, R. P.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. NMR Characterization of Hairpin Polyamide Complexes with the Minor Groove of DNA. *J. Am. Chem. Soc.* **1997**, *119* (34), 7909–7916.
- (13) Parkinson, J. A.; Brkljaca, R.; Urban, S.; Hollingsworth, K.; Turnbull, B. Selected

- Applications of NMR Spectroscopy. In *Modern NMR Techniques for Synthetic Chemistry*; Fisher, J., Ed.; New Directions in Organic & Biological Chemistry; CRC Press, 2014; pp 263–328.
- (14) Pelton, J. G.; Wemmer, D. E. Binding Modes of Distamycin A with d(CGCAAATTTGCG)₂ Determined by Two-Dimensional NMR. *J. Am. Chem. Soc.* **1990**, *112*, 1393–1399.
- (15) Bai, X.; McMullan, G.; Scheres, S. H. W. How Cryo-EM Is Revolutionizing Structural Biology. *Trends Biochem. Sci.* **2018**, *40* (1), 49–57.
- (16) Withers, J. M.; Padroni, G.; Pauff, S. M.; Clark, A. W.; Mackay, S. P.; Burley, G. A. *Comprehensive Supramolecular Chemistry II - DNA Minor Groove Binders as Therapeutic Agents*, 2nd ed.; Elsevier Inc., 2016.
- (17) Watson J.D.; Crick, F. H. C. A Structure for Deoxyribose Nucluc Acid. *Nature* **1953**, *171*, 737–738.
- (18) Fonseca Guerra, C.; Bickelhaupt, F. M.; Snijders, J. G.; Baerends, E. J. The Nature of the Hydrogen Bond in DNA Base Pairs: The Role of Charge Transfer and Resonance Assistance. *Chem. A Eur. J.* **1999**, *5*, 3581–3594.
- (19) Yakovchuk, P.; Protozanova, E.; Frank-Kamenetskii, M. D. Base-Stacking and Base-Pairing Contributions into Thermal Stability of the DNA Double Helix. *Nucleic Acids Res.* **2006**, *34* (2), 564–574.
- (20) Wing, R.; Drew, H.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E. Crystal Structure Analysis of a Complete Turn of B-DNA. *Nature* **1980**, *287*, 755–758.
- (21) Basham, B.; Schroth, G. P.; Ho, P. S. An A-DNA Triplet Code: Thermodynamic Rules for Predicting A- and B-DNA. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92* (14), 6464–6468.
- (22) Auffinger, P.; Westhof, E. Water and Ion Binding around r(UpA)₁₂ and d(TpA)₁₂ Oligomers - Comparison with RNA and DNA (CpG)₁₂ duplexes. *J. Mol. Biol.* **2001**, *305* (5), 1057–1072.
- (23) Xiong, Y.; Sundaralingam, M. Crystal Structure and Conformation of a DNA–RNA Hybrid Duplex with a Polypurine RNA Strand: d(TTCTTBr5CTTC)–r(GAAGAAGAA). *Structure* **1998**, *6* (12), 1493–1501.
- (24) Pallan, P. S.; Lubini, P.; Bolli, M.; Egli, M. Backbone-Base Inclination as a Fundamental Determinant of Nucleic Acid Self- and Cross-Pairing. *Nucleic Acids Res.* **2007**, *35* (19), 6611–6624.
- (25) Rich, A.; Nordheim, A.; Wang, A. H.-J. The Chemistry and Biology of Left-Handed Z-DNA. *Annu. Rev. Biochem.* **1984**, *53* (1), 791–846.

- (26) Ha, S. C.; Lowenhaupt, K.; Rich, A.; Kim, Y.-G.; Kim, K. K. Crystal Structure of a Junction between B-DNA and Z-DNA Reveals Two Extruded Bases. *Nature* **2005**, *437*, 1183-1186.
- (27) Lercher, L.; McDonough, M. A.; El-Sagheer, A. H.; Thalhammer, A.; Kriaucionis, S.; Brown, T.; Schofield, C. J. Structural Insights into How 5-Hydroxymethylation Influences Transcription Factor Binding. *Chem. Commun.* **2014**, *50* (15), 1794–1796.
- (28) Chatake, T.; Tanaka, I.; Umino, H.; Arai, S.; Niimura, N. The Hydration Structure of a Z-DNA Hexameric Duplex Determined by a Neutron Diffraction Technique. *Acta Crystallogr. Sect. D* **2005**, *61* (8), 1088–1098.
- (29) Brown, T. *Genomes 2nd Edition*; Oxford: Wiley-Liss, 2002.
- (30) Yang, L.; Orenstein, Y.; Jolma, A.; Yin, Y.; Taipale, J.; Shamir, R.; Rohs, R. Transcription Factor Family-specific DNA Shape Readout Revealed by Quantitative Specificity Models. *Mol. Syst. Biol.* **2017**, *13* (2), 910.
- (31) Gorin, A. A.; Zhurkin, V. B.; Wilma K. B-DNA Twisting Correlates with Base-Pair Morphology. *J. Mol. Biol.* **1995**, *247* (1), 34–48.
- (32) Mack, D. R.; Chiu, T. K.; Dickerson, R. E. Intrinsic Bending and Deformability at the T-A Step of CCTTTAAAGG: A Comparative Analysis of T-A and A-T Steps within A-Tracts. *J. Mol. Biol.* **2001**, *312* (5), 1037–1049.
- (33) Olson, W. K.; Gorin, A. A.; Lu, X.-J.; Hock, L. M.; Zhurkin, V. B. DNA Sequence-Dependent Deformability Deduced from Protein-DNA Crystal Complexes. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95* (19), 11163–11168.
- (34) Dey, B.; Thukral, S.; Krishnan, S.; Chakrobarty, M.; Gupta, S.; Manghani, C.; Rani, V. DNA-Protein Interactions: Methods for Detection and Analysis. *Mol. Cell. Biochem.* **2012**, *365* (1–2), 279–299.
- (35) Klug, A. The Discovery of Zinc Fingers and Their Development for Practical Applications in Gene Regulation and Genome Manipulation. *Q. Rev. Biophys.* **2010**, *43* (1), 1–21.
- (36) Pavletich, N. P.; Pabo, C. O. Zinc Finger-DNA Recognition: Crystal Structure of a Zif268-DNA Complex at 2.1 Å. *Science* **1991**, *252* (5007), 809–817.
- (37) Foppoli, C.; Coccia, R.; Perluigi, M. Chapter 6 - Role of Oxidative Stress in Human Papillomavirus-Driven Cervical Carcinogenesis A2 Cancer. In *Cancer*; Academic Press: San Diego, 2014; pp 51–61.
- (38) Ellenberger, T. E.; Brandl, C. J.; Struhl, K.; Harrison, S. C. The GCN4 Basic Region Leucine Zipper Binds DNA as a Dimer of Uninterrupted α Helices: Crystal Structure

- of the Protein-DNA Complex. *Cell* **1992**, *71* (7), 1223–1237.
- (39) Lodish, H.; Arnauld, B.; Zipursky, S. L.; Matsudaira, P.; Baltimore, D.; Darnell, J. *Molecular Cell Biology*; 4th ed., New York: W.H. Freeman, 2000.
- (40) Muthurajan, U. M.; Park, Y. J.; Edayathumangalam, R. S.; Suto, R. K.; Chakravarthy, S.; Dyer, P. N.; Luger, K. Structure and Dynamics of Nucleosomal DNA. *Biopolymers* **2003**, *68* (4), 547–556.
- (41) Kornberg, R. D. Chromatin Structure: A Repeating Unit of Histones and DNA. *Science* **1974**, *184* (4139), 868–871.
- (42) Van Helden, P. D. Histone H10: A Maintainer of the Differentiated Cell State? *Int. J. Biochem.* **1985**, *17* (3), 381–385.
- (43) Luger, K.; Mäder, A. W.; Richmond, R. K.; Sargent, D. F.; Richmond, T. J. Crystal Structure of the Nucleosome Core Particle at 2.8 Å Resolution. *Nature* **1997**, *389*, 251–260.
- (44) Ebright, R. H. RNA Polymerase: Structural Similarities between Bacterial RNA Polymerase and Eukaryotic RNA Polymerase II. *J. Mol. Biol.* **2000**, *304* (5), 687–698.
- (45) Liu, X.; Bushnell, D. A.; Kornberg, R. D. RNA Polymerase II Transcription: Structure and Mechanism. *Biochim. Biophys. Acta - Gene Regul. Mech.* **2013**, *1829* (1), 2–8.
- (46) Nogales, E.; Louder, R. K.; He, Y. Structural Insights into the Eukaryotic Transcription Initiation Machinery. *Annu. Rev. Biophys.* **2017**, *46* (1), 59–83.
- (47) Gnatt, A. L.; Cramer, P.; Fu, J.; Bushnell, D. A.; Kornberg, R. D. Structural Basis of Transcription: An RNA Polymerase II Elongation Complex at 3.3 Å Resolution. *Science* **2001**, *292* (5523), 1876–1882.
- (48) Collingwood, T. N.; Urnov, F. D.; Wolffe, a. P. Nuclear Receptors: Coactivators, Corepressors and Chromatin Remodeling in the Control of Transcription. *J. Mol. Endocrinol.* **1999**, *23* (3), 255–275.
- (49) Kosa, P. F.; Ghosh, G.; DeDecker, B. S.; Sigler, P. B. The 2.1 Å Crystal Structure of an Archaeal Preinitiation Complex: TATA-Box-Binding Protein/transcription Factor (II)B core/TATA-Box. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94* (12), 6042–6047.
- (50) Cramer, P.; Armache, K.-J.; Baumli, S.; Benkert, S.; Brueckner, F.; Buchen, C.; Damsma, G. E.; Dengl, S.; Geiger, S. R.; Jasiak, A. J.; et al. Structure of Eukaryotic RNA Polymerases. *Annu. Rev. Biophys.* **2008**, *37* (1), 337–352.
- (51) Steitz, T. A. DNA Polymerases: Structural Diversity and Common Mechanisms. *J. Biol. Chem.* **1999**, *274* (25), 17395–17398.

- (52) Steltz, T. A. A Mechanism for All Polymerases. *Nature* **1998**, *391* (6664), 231–232.
- (53) McKenna, N. J.; O'Malley, B. W. Combinatorial Control of Gene Expression by Nuclear Receptors and Coregulators. *Cell* **2002**, *108* (4), 465–474.
- (54) Aranda, A.; Pascual, A. Nuclear Hormone Receptors and Gene Expression. *Physiol. Rev.* **2001**, *81* (3), 1269–1304.
- (55) Huang, P.; Chandra, V.; Rastinejad, F. Structural Overview of the Nuclear Receptor Superfamily: Insights into Physiology and Therapeutics. *Annu. Rev. Physiol.* **2010**, *72* (1), 247–272.
- (56) Vanden, J. P. Nuclear Receptors : A Brief Overview. *Nucl. Recept. Resour.* **2009**, 1–5.
- (57) Shiau, A. K.; Barstad, D.; Loria, P. M.; Cheng, L.; Kushner, P. J.; Agard, D. A.; Greene, G. L. The Structural Basis of Estrogen Receptor/Coactivator Recognition and the Antagonism of This Interaction by Tamoxifen. *Cell* **2018**, *95* (7), 927–937.
- (58) Knauert, M. P.; Glazer, P. M. Triplex Forming Oligonucleotides: Sequence-Specific Tools for Gene Targeting. *Hum. Mol. Genet.* **2001**, *10* (20), 2243–2251.
- (59) Kurreck, J. Antisense Technologies. *Eur. J. Biochem.* **2003**, *270* (8), 1628–1644.
- (60) Koeller, K. J.; Harris, G. D.; Aston, K.; He, G.; Castaneda, C. H.; Melissa, A.; Edwards, T. G.; Wang, S.; Nanjunda, R.; Wilson, W. D.; et al. Recognition in Their Use as Gene-Specific and Antiviral Agents. *Med Chem (Los Angeles)* **2014**, 338–344.
- (61) Zhu, W.; Wang, Y.; Li, K.; Gao, J.; Huang, C.-H.; Chen, C.-C.; Ko, T.-P.; Zhang, Y.; Guo, R.-T.; Oldfield, E. Antibacterial Drug Leads: DNA and Enzyme Multitargeting. *J. Med. Chem.* **2015**, *58*, 1215–1227.
- (62) Cai, X.; Gray, P. J.; Von Hoff, D. D. DNA Minor Groove Binders: Back in the Groove. *Cancer Treat. Rev.* **2009**, *35* (5), 437–450.
- (63) Sharma, V. K.; Sharma, R. K.; Singh, S. K. Antisense Oligonucleotides: Modifications and Clinical Trials. *Med. Chem. Commun.* **2014**, *5* (10), 1454–1471.
- (64) Asensio, J. L.; Brown, T.; Lane, A. N. Solution Conformation of a Parallel DNA Triple Helix with 5' and 3' Triplex–duplex Junctions. *Structure* **1999**, *7* (1), 1–11.
- (65) Aklank, J.; Wang, G.; Vasquez, K. M. DNA Triple Helices: Biological Consequences and Therapeutic Potential. *Biochimie* **2008**, *90* (8), 1117–1130.
- (66) Gupta, A.; Mishra, A.; Puri, N. Peptide Nucleic Acids: Advanced Tools for Biomedical Applications. *J. Biotechnol.* **2017**, *259*, 148–159.

- (67) Betts, L.; Josey, J. A.; Veal, J. M.; Jordan, S. R. A Nucleic Acid Triple Helix Formed by a Peptide Nucleic Acid-DNA Complex. *Science* **1995**, *270* (5243), 1838–1841.
- (68) Rasmussen, F. W.; Bendifallah, N.; Zachar, V.; Shiraishi, T.; Fink, T.; Ebbesen, P.; Nielsen, P. E.; Koppelhus, D. U. Evaluation of Transfection Protocols for Unmodified and Modified Peptide Nucleic Acid (PNA) Oligomers. *Oligonucleotides* **2006**, *16* (1), 43–57.
- (69) Wittung, P.; Kajanus, J.; Edwards, K.; Haaima, G.; Nielsen, P. E.; Norden, B.; Malmström, B. G. Phospholipid Membrane Permeability of Peptide Nucleic Acid. *FEBS Lett.* **1995**, *375* (3), 27–29.
- (70) Ivanova, G. D.; Arzumanov, A.; Abes, R.; Yin, H.; Wood, M. J. A.; Lebleu, B.; Gait, M. J. Improved Cell-Penetrating peptide-PNA Conjugates for Splicing Redirection in HeLa Cells and Exon Skipping in Mdx Mouse Muscle. *Nucleic Acids Res.* **2008**, *36* (20), 6418–6428.
- (71) Reddy, B. S. P.; Sondhi, S. M.; Lown, J. W. Synthetic DNA Minor Groove-Binding Drugs. *Pharmacol. Ther.* **1999**, *84* (1), 1–111.
- (72) Ramos, J. P.; Le, V. H.; Lewis, E. a. Role of Water in Netropsin Binding to an A2T2 Hairpin DNA Site: Osmotic Stress Experiments. *J. Phys. Chem. B* **2013**, *117*, 15958–15965.
- (73) Marky, L.; Breslauer, K. J. Origins of Netropsin Binding Affinity and Specificity: Correlations of Thermodynamic and Structural Data. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84* (13), 4359–4363.
- (74) Rajagopalan, M.; Ayyer, J.; Sasisekharan, V. Interaction of Non-Intercalative Drugs with DNA: Distamycin Analogues. *J. Biosci.* **1985**, *7* (1), 27–32.
- (75) Asagi, M.; Toyama, A.; Takeuchi, H. Binding Affinity and Mode of Distamycin A with A/T Stretches in Double-Stranded DNA: Importance of the Terminal A/T Residues. *Biophys. Chem.* **2010**, *149* (1–2), 34–39.
- (76) Baliga, R.; Crothers, D. M. On the Kinetics of Distamycin Binding to Its Target Sites on Duplex DNA. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97* (14), 7814–7818.
- (77) Chen, F. M.; Sha, F. Circular Dichroic and Kinetic Differentiation of DNA Binding Modes of Distamycin. *Biochemistry* **1998**, *37* (32), 11143–11151.
- (78) Lewis, E. A.; Munde, M.; Wang, S.; Rettig, M.; Le, V.; MacHha, V.; Wilson, W. D. Complexity in the Binding of Minor Groove Agents: Netropsin Has Two Thermodynamically Different DNA Binding Modes at a Single Site. *Nucleic Acids Res.* **2011**, *39* (22), 9649–9658.

- (79) Van Hecke, K.; Nam, P. C.; Nguyen, M. T.; Van Meervelt, L. Netropsin Interactions in the Minor Groove of d(GGCCAATTGG) Studied by a Combination of Resolution Enhancement and Ab Initio Calculations. *FEBS J.* **2005**, *272* (14), 3531–3541.
- (80) Uytterhoeven, K.; Sponer, J.; Van Meervelt, L. Two 1 : 1 Binding Modes for Distamycin in the Minor Groove of d(GGCCAATTGG). *Eur. J. Biochem.* **2002**, *269* (12), 2868–2877.
- (81) Chen, X.; Ramakrishnan, B.; Rao, S. T.; Sundaralingam, M. Binding of Two Distamycin A Molecules in the Minor Groove of an Alternating B-DNA Duplex. *Nat. Struct. Biol.* **1994**, *1* (3), 169–175.
- (82) Anthony, N. G.; Johnston, B. F.; Khalaf, A. I.; MacKay, S. P.; Parkinson, J. A.; Suckling, C. J.; Waigh, R. D. Short Lexitropsin That Recognizes the DNA Minor Groove at 5'-ACTAGT-3' Understanding the Role of Isopropyl-Thiazole. *J. Am. Chem. Soc.* **2004**, *126* (2), 11338–11349.
- (83) Kopka, M. L.; Goodsell, D. S.; Han, G. W.; Chiu, T. K.; Lown, J. W.; Dickerson, R. E. Defining GC-Specificity in the Minor Groove: Side-by-Side Binding of the Di-Imidazole Lexitropsin to C-A-T-G-G-C-C-A-T-G. *Structure* **1997**, *5* (8), 1033–1046.
- (84) Kissinger, K.; Krowicki, K.; Dabrowiak, J. C.; Lawn, J. W. Imidazole Lexitropsins That Display Enhanced GC Sequence Dependent DNA. *Biochemistry* **1987**, *26* (18), 5590–5595.
- (85) Goodsell, D. S.; Ho Leung Ng; Kopka, M. L.; Lown, J. W.; Dickerson, R. E. Structure of a Dicationic Monoimidazole Lexitropsin Bound to DNA. *Biochemistry* **1995**, *34* (51), 16654–16661.
- (86) Kielkopf, C. L.; Baird, E. E.; Dervan, P. B.; Rees, D. C. Structural Basis for G•C Recognition in the DNA Minor Groove. *Nat. Struct. Biol.* **1998**, *5*, 104.
- (87) Urbach, A. R.; Szewczyk, J. W.; White, S.; Turner, J. M.; Baird, E. E.; Dervan, P. B. Sequence Selectivity of 3-Hydroxypyrrole/pyrrole Ring Pairings in the DNA Minor Groove. *J. Am. Chem. Soc.* **1999**, *121* (50), 11621–11629.
- (88) Kielkopf, C. L.; Bremer, R. E.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. Structural Effects of DNA Sequence on T A Recognition by Hydroxypyrrole/Pyrrole Pairs in the Minor Groove. *J. Mol. Biol.* **2000**, *295*, 557–567.
- (89) Melander, C.; Herman, D. M.; Dervan, P. B. Discrimination of A/T Sequences in the Minor Groove of DNA within a Cyclic Polyamide Motif. *Chem. - A Eur. J.* **2000**, *6* (24), 4487–4497.
- (90) Kielkopf, C. L.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. A Structural Basis for Recognition of A.T and T.A Base Pairs in the Minor Groove of B-DNA. *Science* **1998**, *282* (5386), 111–115.

- (91) Urbach, A. R.; Dervan, P. B. Toward Rules for 1 : 1 Polyamide : DNA Recognition. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (8), 4343–4348.
- (92) Erwin, G. S.; Grieshop, M. P.; Ali, A.; Qi, J.; Lawlor, M.; Kumar, D.; Ahmad, I.; McNally, A.; Teider, N.; Worringer, K.; et al. Synthetic Transcription Elongation Factors License Transcription across Repressive Chromatin. *Science* **2017**, *358*, 1217–1622.
- (93) Burnett, R.; Melander, C.; Puckett, J. W.; Son, L. S.; Wells, R. D.; Dervan, P. B.; Gottesfeld, J. M. DNA Sequence-Specific Polyamides Alleviate Transcription Inhibition Associated with Long GAA·TTC Repeats in Friedreich's Ataxia. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103* (31), 11497–11502.
- (94) Swalley, S. E.; Baird, E. E.; Dervan, P. B. Effects of Gamma-Turn and Beta-Tail Amino Acids on Sequence-Specific Recognition of DNA by Hairpin Polyamides. *J. Am. Chem. Soc.* **1999**, *121* (6), 1113–1120.
- (95) Urbach, A. R.; Love, J. J.; Ross, S. A.; Dervan, P. B. Structure of a β -Alanine-Linked Polyamide Bound to a Full Helical Turn of Purine Tract DNA in the 1:1 Motif. *J. Mol. Biol.* **2002**, *320* (1), 55–71.
- (96) Olenyuk, B.; Jitianu, C.; Dervan, P. B. Parallel Synthesis of H-Pin Polyamides by Alkene Metathesis on Solid Phase. *J. Am. Chem. Soc.* **2003**, *125* (16), 4741–4751.
- (97) Sharma, S. K.; Tandon, M.; Lown, J. W. Design and Synthesis of Novel Thiazole-Containing Cross-Linked Polyamides Related to the Antiviral Antibiotic Distamycin. *J. Org. Chem* **2000**, No. 18, 1102–1107.
- (98) Hare, C. C. O.; Mack, D.; Tandon, M.; Sharma, S. K.; Lown, J. W.; Kopka, M. L.; Dickerson, R. E.; Hartley, J. A. DNA Sequence Recognition in the Minor Groove by Crosslinked Polyamides : The Effect of N-Terminal Head Group and Linker Length on Binding Affinity and Specificity. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *99*, 72-77.
- (99) Chen, Y.-H.; Lown, J. W. A New DNA Minor Groove Binding Motif: Cross-Linked Lexitropsins. *J. Am. Chem. Soc.* **1994**, *116* (16), 6995–7005.
- (100) Herman, D. M.; Baird, E. E.; Dervan, P. B. Tandem Hairpin Motif for Recognition in the Minor Groove of DNA by Pyrrole - Imidazole Polyamides. *Chem. Eur. J.* **1999**, *5* (3), 975–983.
- (101) Mrksich, M.; Parks, M. E.; Dervan, P. B. Hairpin Peptide Motif. A New Class of Oligopeptides for Sequence-Specific Recognition in the Minor Groove of Double-Helical DNA. *J. Am. Chem. Soc.* **1994**, *116* (18), 7983–7988.
- (102) Greenberg, W. a.; Baird, E. E.; Dervan, P. B. A Comparison of H-Pin and Hairpin Polyamide Motifs for the Recognition of the Minor Groove of DNA. *Chem. - A Eur. J.* **1998**, *4* (5), 796–805.

- (103) Li, B. C.; Montgomery, D. C.; Puckett, J. W.; Dervan, P. B. Synthesis of Cyclic Py-Im Polyamide Libraries. *J. Org. Chem.* **2013**, *78* (1), 124–133.
- (104) Yamamoto, M.; Bando, T.; Morinaga, H.; Kawamoto, Y.; Hashiya, K.; Sugiyama, H. Sequence-Specific DNA Recognition by Cyclic Pyrrole-Imidazole Cysteine-Derived Polyamide Dimers. *Chem. - A Eur. J.* **2014**, *20* (3), 752–759.
- (105) Cho, J.; Parks, M. E.; Dervan, P. B. Cyclic Polyamides for Recognition in the Minor Groove of DNA. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92* (22), 10389–10392.
- (106) Zhang, Q.; Dwyer, T. J.; Tsui, V.; Case, D. A.; Cho, J.; Dervan, P. B.; Wemmer, D. E. NMR Structure of a Cyclic Polyamide-DNA Complex. *J. Am. Chem. Soc.* **2004**, *126* (25), 7958–7966.
- (107) Herman, D. M.; Baird, E. E.; Dervan, P. B. Stereochemical Control of the DNA Binding Affinity, Sequence Specificity, and Orientation Preference of Chiral Hairpin Polyamides in the Minor Groove. *J. Am. Chem. Soc.* **1998**, *120* (7), 1382–1391.
- (108) Dose, C.; Farkas, M. E.; Chenoweth, D. M.; Dervan, P. B. Next Generation Hairpin Polyamides with (R) -3, 4-Diaminobutyric Acid Turn Unit. *J. Am. Chem. Soc.* **2008**, *130* (11), 6859–6866.
- (109) Kers, I.; Dervan, P. B. Search for the Optimal Linker in Tandem Hairpin Polyamides. *Bioorganic Med. Chem.* **2002**, *10* (10), 3339–3349.
- (110) Kawamoto, Y.; Sasaki, A.; Chandran, A.; Hashiya, K.; Ide, S.; Bando, T.; Maeshima, K.; Sugiyama, H. Targeting 24 Bp within Telomere Repeat Sequences with Tandem Tetramer Pyrrole-Imidazole Polyamide Probes. *J. Am. Chem. Soc.* **2016**, *138* (42), 14100–14107.
- (111) Kawamoto, Y.; Sasaki, A.; Hashiya, K.; Ide, S.; Bando, T.; Maeshima, K.; Sugiyama, H. Tandem Trimer Pyrrole-imidazole Polyamide Probes Targeting 18 Base Pairs in Human Telomere Sequences. *Chem. Sci.* **2015**, *6*, 2307–2312.
- (112) Maeshima, K.; Janssen, S.; Laemmli, U. K. Specific Targeting of Insect and Vertebrate Telomeres with Pyrrole and Imidazole Polyamides. *EMBO J.* **2001**, *20* (12), 3218–3228.
- (113) Swalley, S. E.; Baird, E. E.; Dervan, P. B. A Pyrrole-Imidazole Polyamide Motif for Recognition of Eleven Base Pair Sequences in the Minor Groove of DNA. *Chem. Eur. J.* **1997**, *3* (10), 1600–1607.
- (114) Wang, X.; Nagase, H.; Watanabe, T.; Nobusue, H.; Suzuki, T.; Asami, Y.; Shinojima, Y.; Kawashima, H.; Takagi, K.; Mishra, R.; et al. Inhibition of MMP-9 Transcription and Suppression of Tumor Metastasis by Pyrrole-Imidazole Polyamide. *Cancer Sci.* **2010**, *101* (3), 759–766.

- (115) Trauger, J. W.; Baird, E. E.; Dervan, P. B. Cooperative Hairpin Dimers for Recognition of DNA by Pyrrole-Imidazole Polyamides. *Angew. Chemie - Int. Ed.* **1998**, *37* (10), 1421–1423.
- (116) Nickols, N. G.; Szablowski, J. O.; Hargrove, A. E.; Li, B. C.; Raskatov, J. A.; Dervan, P. B. Activity of a Py-Im Polyamide Targeted to the Estrogen Response Element. *Mol. Cancer Ther.* **2013**, *12* (5), 675–684.
- (117) Nickols, N. G.; Dervan, P. B. Suppression of Androgen Receptor-Mediated Gene Expression by a Sequence-Specific DNA-Binding Polyamide. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 10418–10423.
- (118) Muzikar, K.; Nickols, N. G.; Dervan, P. B. Repression of DNA-Binding Dependent Glucocorticoid Receptor-Mediated Gene Expression. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (39), 16598–16603.
- (119) Davey, R. A.; Grossmann, M. Androgen Receptor Structure, Function and Biology: From Bench to Bedside. *Clin. Biochem. Rev.* **2016**, *37* (1), 3–15.
- (120) Tan, M. E.; Li, J.; Xu, H. E.; Melcher, K.; Yong, E.-L. *Androgen Receptor: Structure, Role in Prostate Cancer and Drug Discovery.*; Nature Publishing Group, 2015; Vol. 36, pp 3–23.
- (121) Matsumoto, T.; Sakari, M.; Okada, M.; Yokoyama, A.; Takahashi, S.; Kouzmenko, A.; Kato, S. The Androgen Receptor in Health and Disease. *Annu. Rev. Physiol.* **2013**, *75*, 201–224.
- (122) Horie-Inoue, K.; Bono, H.; Okazaki, Y.; Inoue, S. Identification and Functional Analysis of Consensus Androgen Response Elements in Human Prostate Cancer Cells. *Biochem. Biophys. Res. Commun.* **2004**, *325* (4), 1312–1317.
- (123) Tomlins, S. A.; Laxman, B.; Varambally, S.; Cao, X.; Yu, J.; Helgeson, B. E.; Cao, Q.; Prensner, J. R.; Rubin, M. A.; Shah, R. B.; et al. Role of the TMPRSS2-ERG Gene Fusion in Prostate Cancer. *Neoplasia* **2008**, *10* (2), 177–188.
- (124) Wang, Z., Wang, Y., Zhang, J., Hu, Q., Zhi, F., Zhang, S., Mao, D., Zhang, Y., Liang, H. Significance of the TMPRSS2:ERG Gene Fusion in Prostate Cancer. *Mol. Med. Rep.* **2017**, No. 16, 5450–5458.
- (125) Shaffer, P. L.; Jivan, A.; Dollins, D. E.; Claessens, F.; Gewirth, D. T. Structural Basis of Androgen Receptor Binding to Selective Androgen Response Elements. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (14), 4758–4763.
- (126) Nelson, P. S.; Clegg, N.; Arnold, H.; Ferguson, C.; Bonham, M.; White, J.; Hood, L.; Lin, B. The Program of Androgen-Responsive Genes in Neoplastic Prostate Epithelium. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (18), 11890–11895.

- (127) Karayi, M. K.; Markham, A. F. Molecular Biology of Prostate Cancer. *Prostate Cancer Prostatic Dis.* **2004**, *7* (1), 6–20.
- (128) Krause, W. C.; Shafi, A. A.; Nakka, M.; Weigel, N. L. Androgen Receptor and Its Splice Variant, AR-V7, Differentially Regulate FOXA1 Sensitive Genes in LNCaP Prostate Cancer Cells. *Int. J. Biochem. Cell Biol.* **2014**, *54*, 49–59.
- (129) Eisermann, K.; Wang, D.; Jing, Y.; Pascal, L. E.; Wang, Z. Androgen Receptor Gene Mutation, Rearrangement, Polymorphism. *Transl. Androl. Urol.* **2013**, *2* (3), 137–147.
- (130) Jacobs, C. S.; Dervan, P. B. Modifications at the C-Terminus to Improve Pyrrole-Imidazole Polyamide Activity in Cell Culture. *J. Med. Chem.* **2009**, *52* (23), 7380–7388.
- (131) Hargrove, A. E.; Martinez, T. F.; Hare, A. A.; Kurmis, A. A.; Phillips, J. W.; Sud, S.; Pienta, K. J.; Dervan, P. B. Tumor Repression of VCaP Xenografts by a Pyrrole-Imidazole Polyamide. *PLoS One* **2015**, *10* (11), e0143161.
- (132) Chenoweth, D. M.; Dervan, P. B. Structural Basis for Cyclic Py-Im Polyamide Allosteric Inhibition of Nuclear Receptor Binding. *J. Am. Chem. Soc.* **2010**, *132* (41), 14521–14529.
- (133) Chenoweth, D. M.; Dervan, P. B. Allosteric Modulation of DNA by Small Molecules. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (32), 13175–13179.
- (134) Yang, F.; Nickols, N. G.; Li, B. C.; Marinov, G. K.; Said, J. W.; Dervan, P. B. Antitumor Activity of a Pyrrole-Imidazole Polyamide. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (5), 1863–1868.
- (135) Martinez, T. F.; Phillips, J. W.; Karanja, K. K.; Polaczek, P.; Wang, C.-M. C. M.; Li, B. C.; Campbell, J. L.; Dervan, P. B.; Martínez, T. F.; Phillips, J. W.; et al. Replication Stress by Py-Im Polyamides Induces a Non-Canonical ATR-Dependent Checkpoint Response. *Nucleic Acids Res.* **2014**, *42* (18), 11546–11559.
- (136) Xu, L.; Wang, W.; Gotte, D.; Yang, F.; Hare, A. A.; Welch, T. R.; Li, B. C.; Shin, J. H.; Chong, J.; Strathern, J. N.; et al. RNA Polymerase II Senses Obstruction in the DNA Minor Groove via a Conserved Sensor Motif. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113* (44), 12426–12431.
- (137) Kielkopf, C. L.; Ding, S.; Kuhn, P.; Rees, D. C. Conformational Flexibility of B-DNA at 0.74 Å Resolution: d(CCAGTACTGG)₂. *J. Mol. Biol.* **2000**, *296* (3), 787–801.
- (138) Yang, F.; Nickols, N. G.; Li, B. C.; Szablowski, J. O.; Hamilton, S. R.; Meier, J. L.; Wang, C. M.; Dervan, P. B. Animal Toxicity of Hairpin Pyrrole-Imidazole Polyamides Varies with the Turn Unit. *J. Med. Chem.* **2013**, *56* (18), 7449–7457.

- (139) Raskatov, J. A.; Hargrove, A. E.; So, A. Y.; Dervan, P. B. Pharmacokinetics of Py-Im Polyamides Depend on Architecture: Cyclic versus Linear. *J. Am. Chem. Soc.* **2012**, *134* (18), 7995–7999.
- (140) Belitsky, J. M.; Leslie, S. J.; Arora, P. S.; Beerman, T. a.; Dervan, P. B. Cellular Uptake of N-methylpyrrole/N-Methylimidazole Polyamide-Dye Conjugates. *Bioorganic Med. Chem.* **2002**, *10* (10), 3313–3318.
- (141) Best, T. P.; Edelson, B. S.; Nickols, N. G.; Dervan, P. B. Nuclear Localization of Pyrrole-Imidazole Polyamide-Fluorescein Conjugates in Cell Culture. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (21), 12063–12068.
- (142) Meier, J. L.; Montgomery, D. C.; Dervan, P. B. Enhancing the Cellular Uptake of Py-Im Polyamides through next-Generation Aryl Turns. *Nucleic Acids Res.* **2012**, *40* (5), 2345–2356.
- (143) Nishijima, S.; Shinohara, K. ichi; Bando, T.; Minoshima, M.; Kashiwazaki, G.; Sugiyama, H. Cell Permeability of Py-Im-Polyamide-Fluorescein Conjugates: Influence of Molecular Size and Py/Im Content. *Bioorganic Med. Chem.* **2010**, *18* (2), 978–983.
- (144) Han, Y.-W.; Sugiyama, H.; Harada, Y. The Application of Fluorescence-Conjugated Pyrrole/imidazole Polyamides in the Characterization of protein–DNA Complex Formation. *Biomater. Sci.* **2016**, *4* (3), 391–399.
- (145) Raskatov, J.; Puckett, J. W.; Dervan, P. B. A C-14 Labeled Py-Im Polyamide Localizes to a Subcutaneous Prostate Cancer Tumor. *Bioorg. Med. Chem.* **2014**, *22* (16), 10–17.
- (146) Hargrove, A. E.; Raskatov, J. A.; Meier, J. L.; Montgomery, D. C.; Dervan, P. B. Characterization and Solubilization of Pyrrole-Imidazole Polyamide Aggregates. *J. Med. Chem.* **2012**, *55* (11), 5425–5432.
- (147) Marques, M. A.; Doss, R. M. M.; Urbach, A. R. R.; Dervan, P. B. B. Toward an Understanding of the Chemical Etiology for DNA Minor-Groove Recognition by Polyamides. *Helv. Chim. Acta* **2002**, *85* (12), 4485–4517.
- (148) Doss, R. M.; Marques, M.; Foister, S.; Dervan, P. B. DNA Minor-Groove Recognition by 3-Methylthiophene/pyrrole Pair. *Chem. Biodivers.* **2004**, *1* (6), 886–899.
- (149) Foister, S.; Marques, M. A.; Doss, R. M.; Dervan, P. B. Shape Selective Recognition of T·A Base Pairs by Hairpin Polyamides Containing N-Terminal 3-Methoxy (and 3-Chloro) Thiophene Residues. *Bioorg. Med. Chem.* **2003**, *11* (20), 4333–4340.
- (150) Plouvier, B.; Bailly, C.; Houssin, R.; Rao, K. E.; Lown, W. J.; Henichart, J.; Waring, M. J. DNA-Sequence Specific Recognition by a Thiazole Analogue of Netropsin : A Comparative Footprinting Study. *Nucleic Acids Res.* **1991**, *19* (21), 5821–5829.

- (151) Beker, W.; Szarek, P.; Komorowski, L.; Lipiński, J. Reactivity Patterns of Imidazole, Oxazole, and Thiazole as Reflected by the Polarization Justified Fukui Functions. *J. Phys. Chem. A* **2013**, *117* (7), 1596–1600.
- (152) Rao, K Ekambareswara, Shea R.G., Yadagiri B., L. J. W. Molecular Recognition between Oligopeptides and Nucleic Acids: DNA Sequence Specificity and Binding Properties of Thiazole-Lexitropsins. *Anticancer. Drug Des.* **1990**, *5*, 3–20.
- (153) Kumar, R.; Lown, J. W. Synthesis and Antitumor Cytotoxicity Evaluation of Novel Thiazole-Containing Glycosylated Polyamides. *European J. Org. Chem.* **2003**, No. 24, 4842–4851.
- (154) Sharma, S. K.; Tandon, M.; Lown, J. W. Design and Synthesis of New Thiazolated Cross-Linked DNA Binding Polyamides for Altered Sequence Recognition. *Tetrahedron* **2002**, *58* (17), 3417–3421.
- (155) Nguyen, D. H.; Szewczyk, J. W.; Baird, E. E.; Dervan, P. B. Alternative Heterocycles for DNA Recognition : An N -Methylpyrazole / N -Methylpyrrole Pair Specifies for A T / T A Base Pairs. *Bioorg. Med. Chem.* **2001**, *9*, 7–17.
- (156) James, P. L.; Merkina, E. E.; Khalaf, A. I.; Suckling, C. J.; Waigh, R. D.; Brown, T.; Fox, K. R. DNA Sequence Recognition by an Isopropyl Substituted Thiazole Polyamide. *Nucleic Acids Res.* **2004**, *32* (11), 3410–3417.
- (157) Alniss, H. Y.; Salvia, M. V.; Sadikov, M.; Golovchenko, I.; Anthony, N. G.; Khalaf, A. I.; MacKay, S. P.; Suckling, C. J.; Parkinson, J. A. Recognition of the DNA Minor Groove by Thiazotropsin Analogues. *Chembiochem* **2014**, *15* (13), 1978–1990.
- (158) Anthony, N. G.; Breen, D.; Clarke, J.; Donoghue, G.; Drummond, A. J.; Ellis, E. M.; Gemmell, C. G.; Helesbeux, J. J.; Hunter, I. S.; Khalaf, A. I.; et al. Antimicrobial Lexitropsins Containing Amide, Amidine, and Alkene Linking Groups. *J. Med. Chem.* **2007**, *50* (24), 6116–6125.
- (159) Suckling, C. J. The Antibacterial Drug MGB-BP3: From Discovery to Clinical Trial. *Chem. Biol. INTERFACE* **2015**, *5*, 166–174.
- (160) Barrett, M. P.; Gemmell, C. G.; Suckling, C. J. Minor Groove Binders as Anti-Infective Agents. *Pharmacol. Ther.* **2013**, *139* (1), 12–23.
- (161) Khalaf, A. I.; Anthony, N.; Breen, D.; Donoghue, G.; Mackay, S. P.; Scott, F. J.; Suckling, C. J. Amide Isosteres in Structure-Activity Studies of Antibacterial Minor Groove Binders. *Eur. J. Med. Chem.* **2011**, *46* (11), 5343–5355.
- (162) Khalaf, A. I.; Waigh, R. D.; Drummond, A. J.; Pringle, B.; McGroarty, I.; Skellern, G. G.; Suckling, C. J. Distamycin Analogues with Enhanced Lipophilicity: Synthesis and Antimicrobial Activity. *J. Med. Chem.* **2004**, *47* (8), 2133–2156.

- (163) Alniss, H. Y.; Anthony, N. G.; Khalaf, A. I.; Mackay, S. P.; Suckling, C. J.; Waigh, R. D.; Wheate, N. J.; Parkinson, J. a. Rationalising Sequence Selection by Ligand Assemblies in the DNA Minor Groove: The Case for Thiazotropsin A. *Chem. Sci.* **2012**, *3* (3), 711-722.
- (164) Salvia, M. V.; Addison, F.; Alniss, H. Y.; Buurma, N. J.; Khalaf, A. I.; Mackay, S. P.; Anthony, N. G.; Suckling, C. J.; Evstigneev, M. P.; Santiago, A. H.; et al. Thiazotropsin Aggregation and Its Relationship to Molecular Recognition in the DNA Minor Groove. *Biophys. Chem.* **2013**, *179*, 1–11.
- (165) Wemmer, D. E.; Lown, J. W.; Bathini, Y.; Geierstanger, B. H.; Dwyer, T. J. Design and Binding of a Distamycin A Analog to D(CGCAAGTTGGC)-D(GCCAACTTGCG): Synthesis, NMR Studies, and Implications for the Design of Sequence-Specific Minor Groove Binding Oligopeptides. *J. Am. Chem. Soc.* **1992**, *114* (15), 5911–5919.
- (166) Danuta, D. New Solid Phase Synthesis of Distamycin Analogues. *Molecules* **2011**, *16* (4), 3066–3076.
- (167) D’Alessio, R.; Geroni, C.; Biasoli, G.; Pesenti, E.; Grandi, M.; Mongelli, N. Structure-Activity Relationship of Novel Distamycin a Derivatives : Synthesis and Antitumor Activity. *Bioorg. Med. Chem. Lett.* **1994**, *4* (12), 1467–1472.
- (168) Juguet, M.; Lautru, S.; Francou, F. X.; Nezbedová, Š.; Leblond, P.; Gondry, M.; Pernodet, J. L. An Iterative Nonribosomal Peptide Synthetase Assembles the Pyrrole-Amide Antibiotic Congocidine in *Streptomyces Ambofaciens*. *Chem. Biol.* **2009**, *16* (4), 421–431.
- (169) Hao, C.; Huang, S.; Deng, Z.; Zhao, C.; Yu, Y. Mining of the Pyrrolamide Antibiotics Analogs in *Streptomyces Netropsis* Reveals the Amidohydrolase-Dependent “Iterative Strategy” Underlying the Pyrrole Polymerization. *PLoS One* **2014**, *9* (6), e99077.
- (170) Vingadassalon, A.; Lorieux, F.; Juguet, M.; Le Goff, G.; Gerbaud, C.; Pernodet, J.-L.; Lautru, S. Natural Combinatorial Biosynthesis Involving Two Clusters for the Synthesis of Three Pyrrolamides in *Streptomyces Netropsis*. *ACS Chem. Biol.* **2015**, *10* (2), 601–610.
- (171) Bialer, M.; Yagen, B.; Mechoulam, R. A Total Synthesis of Distamycin A, an Antiviral Antibiotic. *Tetrahedron* **1978**, *34* (15), 2389–2391.
- (172) Wurtz, N. R.; Turner, J. M.; Baird, E. E.; Dervan, P. B. Fmoc Solid Phase Synthesis of Polyamides Containing Pyrrole and Imidazole Amino Acids. *Org. Lett.* **2001**, *3* (8), 1201–1203.
- (173) Ohtsuki, A.; Kimura, M. T.; Minoshima, M.; Suzuki, T.; Ikeda, M.; Bando, T.; Nagase, H.; Shinohara, K.; Sugiyama, H. Synthesis and Properties of PI Polyamide-SAHA Conjugate. *Tetrahedron Lett.* **2009**, *50* (52), 7288–7292.

- (174) Fang, L.; Yao, G.; Pan, Z.; Wu, C.; Wang, H.; Burley, G. A.; Su, W. Fully Automated Synthesis of DNA-Binding Py-Im Polyamides Using a Triphosgene Coupling Strategy. *Org. Lett.* **2015**, *17*, 158–161.
- (175) Chenoweth, D. M.; Harki, D. A.; Dervan, P. B. Solution-Phase Synthesis of Pyrrole-Imidazole Polyamides. *J. Am. Chem. Soc.* **2009**, *131* (20), 7175–7181.
- (176) Pauff, S. M.; Fallows, A. J.; Mackay, S. P.; Su, W.; Cullis, P. M.; Burley, G. A. Pyrrole-Imidazole Polyamides: Manual Solid-Phase Synthesis. *Curr. Protoc. nucleic acid Chem.* **2015**, *63*, 8.10.1-8.10.41.
- (177) Marques, M.; Doss, R. M.; Foister, S.; Dervan, P. B. Expanding the Repertoire of Heterocycle Ring Pairs for Programmable Minor Groove DNA Recognition. *J. Am. Chem. Soc.* **2004**, *126* (33), 10339–10349.
- (178) Baird, E. E.; Dervan, P. B. Solid Phase Synthesis of Polyamides Containing Imidazole and Pyrrole Amino Acids. *J. Am. Chem. Soc.* **1996**, *118* (9), 6141–6146.
- (179) Merrifield, R. B. Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.* **1963**, *85* (14), 2149-2154.
- (180) Wetzler, M.; Wemmer, D. E. Facile Dimer Synthesis for DNA-Binding Polyamide Ligands. *Org. Lett.* **2010**, *12* (15), 3488–3490.
- (181) Kelly, J. J. Binding Site Size Limit of the 2:1 Pyrrole-Imidazole Polyamide-DNA Motif. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93* (14), 6981–6985.
- (182) Parks, M. E.; Baird, E. E.; Dervan, P. B. Optimization of the Hairpin Polyamide Design for Recognition of the Minor Groove of DNA. *J. Am. Chem. Soc.* **1996**, *118* (26), 6147–6152.
- (183) Belitsky, J. M.; Nguyen, D. H.; Wurtz, N. R.; Dervan, P. B. Solid-Phase Synthesis of DNA Binding Polyamides on Oxime Resin. *Bioorg. Med. Chem.* **2002**, *10* (8), 2767–2774.
- (184) Fallows, A. J.; Singh, I.; Dondi, R.; Cullis, P. M.; Burley, G. A. Highly Efficient Synthesis of DNA-Binding Polyamides Using a Convergent Fragment-Based Approach. *Org. Lett.* **2014**, 31–34.
- (185) Millington, C. R.; Quarrell, R.; Lowe, G. Aryl Hydrazides as Linkers for Solid Phase Synthesis Which Are Cleavable under Mild Oxidative Conditions. *Tetrahedron Lett.* **1998**, *39* (39), 7201–7204.
- (186) Ni, S.; Zhang, H.; Huang, W.; Zhou, J.; Qian, H.; Chen, W. The Application of an Aryl Hydrazine Linker Prevents Beta-Elimination Side Products in the SPPS of C-Terminal Cysteine Peptides. *J. Pept. Sci.* **2010**, *16* (6), 309–313.

- (187) Barton, A.; Breukelman, S. P.; Kaye, P. T.; Meakins, G. D. The Preparation of Thiazole-4- and -5-Carboxylates, and an Infrared Study of Their Rotational Isomers. *J. Chem. Soc., Perkin Trans I* **1982**, No. 0, 159–164.
- (188) Krutzik, P. O.; Chamberlin, A. R. Rapid Solid-Phase Synthesis of DNA-Binding Pyrrole-Imidazole Polyamides. *Bioorg. Med. Chem. Lett.* **2002**, *12* (16), 2129–2132.
- (189) Krutzik, P. O.; Chamberlin, A. R. Synthesis of DNA-Binding Polyamides. Robust Solid-Phase Methods for Coupling Heterocyclic Aromatic Amino Acids. *Methods Mol. Biol.* **2002**, *201*, 77–92.
- (190) Fang, L.; Wu, C.; Yu, Z.; Shang, P.; Cheng, Y.; Peng, Y.; Su, W. Triphosgene-Mediated Couplings in the Solid Phase: Total Synthesis of Brachyestemin A. *European J. Org. Chem.* **2014**, No. 34, 7572–7576.
- (191) Su, W.; Gray, S. J.; Dondi, R.; Burley, G. A. Highly Efficient Synthesis of DNA-Binding Hairpin Polyamides via the Use of a New Triphosgene Coupling Strategy. *Org. Lett.* **2009**, No. 13, 15–18.
- (192) El-Subbagh, H. I.; Abadi, A. H.; Lehmann, J. Synthesis and Antitumor Activity of Ethyl 2-Substituted-Aminothiazole-4-Carboxylate Analogs. *Arch. Pharm. (Weinheim)*. **1999**, *332* (4), 137–142.
- (193) Cui, Y. M.; Huang, Q. Q.; Xu, J.; Chen, L. L.; Li, J. Y.; Ye, Q. Z.; Li, J.; Nan, F. J. Identification of Potent Type I MetAP Inhibitors by Simple Bioisosteric Replacement. Part 1: Synthesis and Preliminary SAR Studies of Thiazole-4-Carboxylic Acid Thiazol-2-Ylamide Derivatives. *Bioorg. Med. Chem. Lett.* **2005**, *15* (16), 3732–3736.
- (194) Brucoli, F.; Hawkins, R. M.; James, C. H.; Jackson, P. J. M.; Wells, G.; Jenkins, T. C.; Ellis, T.; Kotecha, M.; Hochhauser, D.; Hartley, J. A.; et al. An Extended Pyrrolobenzodiazepine-Polyamide Conjugate with Selectivity for a DNA Sequence Containing the ICB2 Transcription Factor Binding Site. *J. Med. Chem.* **2013**, *56* (16), 6339–6351.
- (195) Iyer, P.; Srinivasan, A.; Singh, S. K.; Mascara, G. P.; Zayitova, S.; Sidone, B.; Fouquerel, E.; Svilar, D.; Sobol, R. W.; Bobola, M. S.; et al. Synthesis and Characterization of DNA Minor Groove Binding Alkylating Agents. *Chem. Res. Toxicol.* **2013**, *26* (1), 156–168.
- (196) Jaramillo, D.; Liu, Q.; Aldrich-Wright, J.; Tor, Y. Synthesis of N-Methylpyrrole and N-Methylimidazole Amino Acids Suitable for Solid-Phase Synthesis. *J. Org. Chem.* **2004**, *69* (23), 8151–8153.
- (197) Wang, S.; Aston, K.; Koeller, K. J.; Harris, G. D.; Rath, N. P.; Bashkin, J. K.; Wilson, W. D. Modulation of DNA–polyamide Interaction by β -Alanine Substitutions: A Study of Positional Effects on Binding Affinity, Kinetics and Thermodynamics. *Org. Biomol. Chem.* **2014**, *12* (38), 7523–7536.

- (198) Ries, O.; Granitzka, M.; Stalke, D.; Ducho, C. Concise Synthesis and X-Ray Crystal Structure of N¹-Benzyl-2-(Pyrimidin-4'-Ylamino)-Thiazole-4-Carboxamide (Thiazovivin), a Small-Molecule Tool for Stem Cell Research. *Synth. Commun.* **2013**, *43* (21), 2876–2882.
- (199) Haq, I. Thermodynamics of Drug-DNA Interactions. *Arch. Biochem. Biophys.* **2002**, *403* (1), 1–15.
- (200) Chaires, J. B. Drug – DNA Interactions. *Curr. Opin. Struct. Biol.* **1998**, No. 8, 314–320.
- (201) Kuriyan, J.; Konforti, B.; Wemmer, D. *The Molecules of Life*; 1st ed., Garland Science Taylor and Francis group, 2012.
- (202) Ritzefeld, M.; Sewald, N. Real-Time Analysis of Specific Protein-DNA Interactions with Surface Plasmon Resonance. *J. Amino Acids* **2012**, *2012*, 1–19.
- (203) Parkinson, J. A.; Scott, F. J.; Suckling, C. J.; Wilson, G. Exceptionally Strong Intermolecular Association in Hydrophobic DNA Minor Groove Binders and Their Potential Therapeutic Consequences. *Medchemcomm* **2013**, *4* (7), 1105–1108.
- (204) Buurma, N. J.; Haq, I. Calorimetric and Spectroscopic Studies of Hoechst 33258: Self-Association and Binding to Non-Cognate DNA. *J. Mol. Biol.* **2008**, *381* (3), 607–621.
- (205) Marky, L. A.; Breslauer, K. J. Calculating Thermodynamic Data for Transitions of Any Molecularity from Equilibrium Melting Curves. *Biopolymers*. 1987, pp 1601–1620.
- (206) Wilson, W. D.; Fariol, T. A.; Maria, F.-S.; Ted, R. *Drug Nucleic Acid Interactions Protocols -Evaluation of Drug Nucleic Acid Interactions by Thermal Melting Curves*; 1997.
- (207) Schildkraut, C.; Lifson, S. Dependence of the Melting Temperature of DNA on Salt Concentration. *Biopolymers* **1965**, *3* (2), 195–208.
- (208) Wang, S.; Nanjunda, R.; Aston, K.; Bashkin, J. K.; Wilson, W. D. Correlation of Local Effects of DNA Sequence and Position of Beta-Alanine Inserts with Polyamide-DNA Complex Binding Affinities and Kinetics. *Biochemistry* **2012**, *51* (49), 9796–9806.
- (209) Hampshire, A. J.; Rusling, D. A.; Broughton-Head, V. J.; Fox, K. R. Footprinting: A Method for Determining the Sequence Selectivity, Affinity and Kinetics of DNA-Binding Ligands. *Methods* **2007**, *42* (2), 128–140.
- (210) Trauger, J. W.; Dervan, P. B. Footprinting Methods for Analysis of Pyrrole-Imidazole polyamide/DNA Complexes. *Methods Enzymol.* **2001**, *340* (1997), 450–466.
- (211) Buurma, N. J.; Haq, I. Advances in the Analysis of Isothermal Titration Calorimetry

Data for Ligand-DNA Interactions. *Methods* **2007**, *42* (2), 162–172.

- (212) Kabiri, M.; Unsworth, L. D. Application of Isothermal Titration Calorimetry for Characterizing Thermodynamic Parameters of Biomolecular Interactions: Peptide Self-Assembly and Protein Adsorption Case Studies. *Biomacromolecules* **2014**, *15* (10), 3463–3473.
- (213) Falconer, R. J. Applications of Isothermal Titration Calorimetry - the Research and Technical Developments from 2011 to 2015. *J. Mol. Recognit.* **2016**, *29* (10), 504–515.
- (214) Nanjunda, R.; Wilson, W. D. Binding to the DNA Minor Groove by Heterocyclic Dications: From AT-Specific Monomers to GC Recognition with Dimers. *Curr. Protoc. nucleic acid Chem.* **2012**, Unit8.8, doi: 10.1002/0471142700.nc0808s51.
- (215) Chavda, S.; Liu, Y.; Babu, B.; Davis, R.; Sielaff, A.; Ruprich, J.; Westrate, L.; Tronrud, C.; Ferguson, A.; Franks, A.; et al. Hx, a Novel Fluorescent, Minor Groove and Sequence Specific Recognition Element: Design, Synthesis, and DNA Binding Properties of P-Anisylbenzimidazole Imidazole/pyrrole-Containing Polyamides. *Biochemistry* **2011**, 3127–3136.
- (216) Rant, U. Sensing with Electro-Switchable Biosurfaces. *Bioanal. Rev.* **2012**, *4* (2–4), 97–114.
- (217) Cléry, A.; Sohier, T. J. M.; Welte, T.; Langer, A.; Allain, F. H. T. switchSENSE: A New Technology to Study Protein-RNA Interactions. *Methods* **2017**, *118–119*, 137–145.
- (218) Langer, A.; Hampel, P. A.; Kaiser, W.; Knezevic, J.; Welte, T.; Villa, V.; Maruyama, M.; Svejda, M.; Jähner, S.; Fischer, F.; et al. Protein Analysis by Time-Resolved Measurements with an Electro-Switchable DNA Chip. *Nat. Commun.* **2013**, *4*, 2099.
- (219) Garbett, N. C.; Ragazzon, P. A.; Chaires, J. B. Circular Dichroism to Determine Binding Mode and Affinity of Ligand-DNA Interactions. *Nat. Protoc.* **2007**, *2* (12), 3166–3172.
- (220) Chang, Y.-M.; Chen, C. K.-M.; Hou, M.-H. Conformational Changes in DNA upon Ligand Binding Monitored by Circular Dichroism. *Int. J. Mol. Sci.* **2012**, *13* (3), 3394–3413.
- (221) Caesar, C. E. B.; Johnsson, R.; Ellervik, U.; Fox, K. R.; Lincoln, P.; Nordén, B. A Polarized-Light Spectroscopy Study of Interactions of a Hairpin Polyamide with DNA. *Biophys. J.* **2006**, *91* (3), 904–911.
- (222) Gondeau, C.; Maurizot, J. C.; Durand, M. Circular Dichroism and UV Melting Studies on Formation of an Intramolecular Triplex Containing Parallel T*A:T and G*G:C Triplets: Netropsin Complexation with the Triplex. *Nucleic Acids Res.* **1998**, *26* (21), 4996–5003.

- (223) Holmgaard List, N.; Knoops, J.; Rubio-Magnieto, J.; Idé, J.; Beljonne, D.; Norman, P.; Surin, M.; Linares, M. Origin of DNA-Induced Circular Dichroism in a Minor-Groove Binder. *J. Am. Chem. Soc.* **2017**, *139* (42), 14947–14953.
- (224) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. The Molecular Origin of DNA-Drug Specificity in Netropsin and Distamycin. *Proc. Natl. Acad. Sci. U. S. A.* **1985**, *82* (5), 1376–1380.
- (225) Gottesfeld, J. M.; Melander, C.; Suto, R. K.; Raviol, H.; Luger, K.; Dervan, P. B. Sequence-Specific Recognition of DNA in the Nucleosome by Pyrrole-Imidazole Polyamides. *J. Mol. Biol.* **2001**, *309* (3), 615–629.
- (226) Krishnan, V. V.; Rupp, B. Macromolecular Structure Determination: Comparison of X-Ray Crystallography and NMR Spectroscopy. In *eLS*; John Wiley & Sons, Ltd, 2001.
- (227) Hawkins, C. A.; Peláez de Clairac, R.; Dominey, R. N.; Baird, E. E.; White, S.; Dervan, P. B.; Wemmer, D. E. Controlling Binding Orientation in Hairpin Polyamide DNA Complexes. *J. Am. Chem. Soc.* **2000**, *122* (22), 5235–5243.
- (228) Keeler, J. *Understanding NMR Spectroscopy*; Wiley, 2010.
- (229) Bunkenborg, J.; Behrens, C.; Jacobsen, J. P. NMR Characterization of the DNA Binding Properties of a Novel Hoechst 33258 Analogue Peptide Building Block. *Bioconjug. Chem.* **2002**, *13* (5), 927–936.
- (230) Adrian, M.; Heddi, B.; Phan, A. T. NMR Spectroscopy of G-Quadruplexes. *Methods* **2012**, *57* (1), 11–24.
- (231) Heddi, B.; Foloppe, N.; Bouchemal, N.; Hantz, E.; Hartmann, B. Quantification of DNA BI/BII Backbone States in Solution. Implications for DNA Overall Structure and Recognition. *J. Am. Chem. Soc.* **2006**, *128* (28), 9170–9177.
- (232) Oguey, C.; Foloppe, N.; Hartmann, B. Understanding the Sequence-Dependence of DNA Groove Dimensions: Implications for DNA Interactions. *PLoS One* **2010**, *5* (12), 1–8.
- (233) Johansson, E.; Parkinson, G.; Neidle, S. A New Crystal Form for the Dodecamer C-G-C-G-A-A-T-T-C-G-C-G: Symmetry Effects on Sequence-Dependent DNA Structure. *J. Mol. Biol.* **2000**, *300* (3), 551–561.
- (234) Carlomagno, T.; Hennig, M.; Williamson, J. R. A Novel PH-CT-COSY Methodology for Measuring JPHcoupling Constants in Unlabeled Nucleic Acids. Application to HIV-2 TAR RNA. *J. Biomol. NMR* **2002**, *22* (1), 65–81.
- (235) De Vivo, M.; Masetti, M.; Bottegoni, G.; Cavalli, A. Role of Molecular Dynamics and Related Methods in Drug Discovery. *J. Med. Chem.* **2016**, *59* (9), 4035–4061.

- (236) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25* (9), 1157–1174.
- (237) Vařeková, R. S.; Koča, J.; Zhan, C. G. Complexity and Convergence of Electrostatic and van Der Waals Energies within PME and Cutoff Methods. *Int. J. Mol. Sci.* **2004**, *5*, 154–173.
- (238) Borgias, B. A.; James, T. L. MARDIGRAS-A Procedure for Matrix Analysis of Relaxation for Discerning Geometry of an Aqueous Structure. *J. Magn. Reson.* **1990**, *87* (3), 475–487.
- (239) Embrey, K. J.; Searle, M. S.; Craik, D. J. Interaction of Hoechst 33258 with the Minor Groove of the A + T-Rich DNA Duplex d(GGTAATTACC)₂ Studied in Solution by NMR Spectroscopy. *Eur. J. Biochem.* **1993**, *211* (3), 437–447.
- (240) D.A. Case, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, D. Greene, N. Homeyer, S. Izadi, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. Mermelstein, K.M. Merz, G. Monard, H., D. M. Y. and P. A. K. Amber17. 2017.
- (241) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605–1612.
- (242) Lee, W.; Tonelli, M.; Markley, J. L. NMRFAM-SPARKY: Enhanced Software for Biomolecular NMR Spectroscopy. *Bioinformatics* **2015**, *31* (8), 1325-1327.
- (243) Blanchet, C.; Pasi, M.; Zakrzewska, K.; Lavery, R. CURVES+ Web Server for Analyzing and Visualizing the Helical, Backbone and Groove Parameters of Nucleic Acid Structures. *Nucleic Acids Res.* **2011**, *39* (suppl_2), W68–W73.
- (244) Lu, X. J.; Olson, W. K. 3DNA: A Software Package for the Analysis, Rebuilding and Visualization of Three-Dimensional Nucleic Acid Structures. *Nucleic Acids Res.* **2003**, *31* (17), 5108–5121.
- (245) Nickols, N. G.; Jacobs, C. S.; Farkas, M. E.; Dervan, P. B. Improved Nuclear Localization of DNA-Binding Polyamides. *Nucleic Acids Res.* **2007**, *35* (2), 363–370.

Appendix

Contents

1.1	HPLC and LC-MS method parameters	252
1.1.1	Analytical HPLC generic methods	252
1.1.2	Semiprep HPLC generic method	253
1.1.3	LC-MS	253
2	Spectroscopic characterization of compounds described in Chapter 2	255
3	NMR chemical shifts of PA·ODN1 complexes	310
4	Analysis of NMR-derived structures	328

1.1 HPLC and LC-MS method parameters

1.1.1 Analytical HPLC generic methods

Column specification: Aeris 3.6 μm WIDEPORÉ XB-C18, 250 x 4.6 mm.

Column temperature 25 °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

1.1.2 Semiprep HPLC generic method

Column specification: Kinetex 5 μ m C18 100A, 150x 21.2mm

Column temperature 25 °C

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
41	90	10

1.1.3 LC-MS

LC Conditions: The HPLC analysis was conducted on a Zorbax 45mm x 150mm C18 column at 40 °C.

The solvents employed were:

A = 5 mM ammonium acetate in water.

B = 5 mM ammonium acetate in acetonitrile.

The gradient employed was:

Time (min)	Flow Rate (ml/min)	% A	% B
0	1	95	5
1.48	1	95	5
8.5	1	0	100
13.5	1	0	100
16.5	1	95	5
18	1	95	5

The UV detection signal was recorded at 254 nm.

MS Conditions

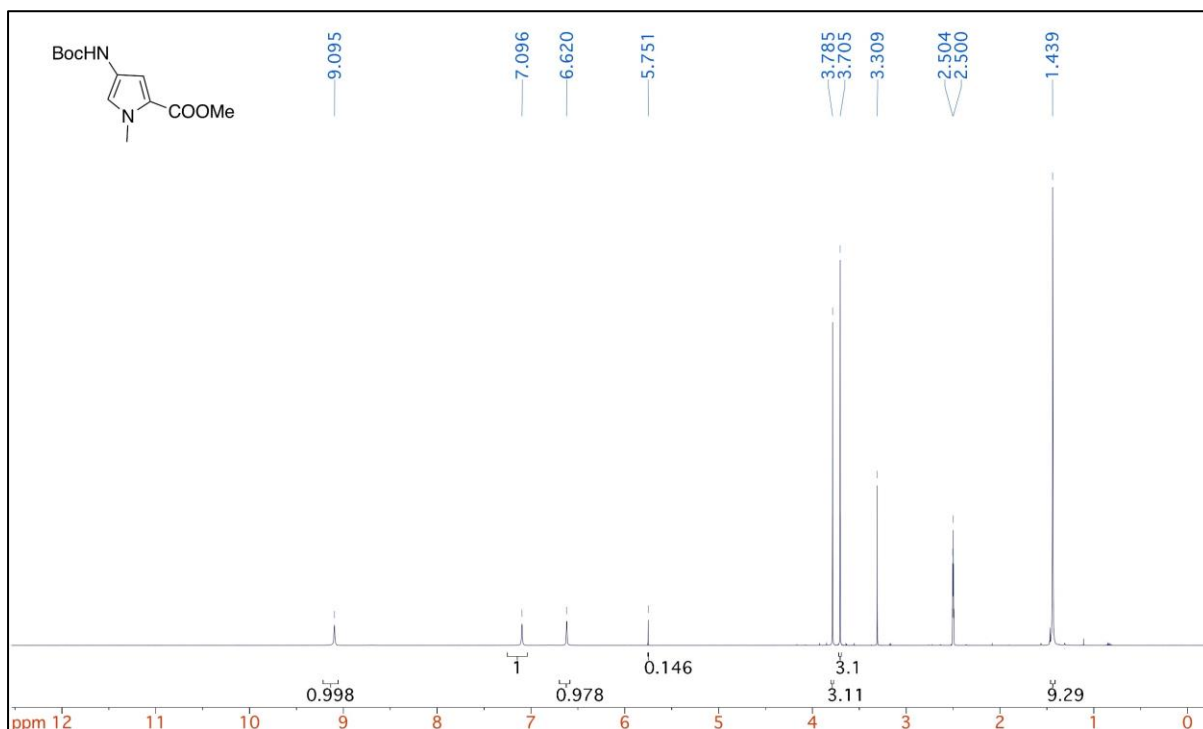
MS: Agilent Quadrupole

Ionisation mode: Positive and/or negative electrospray.

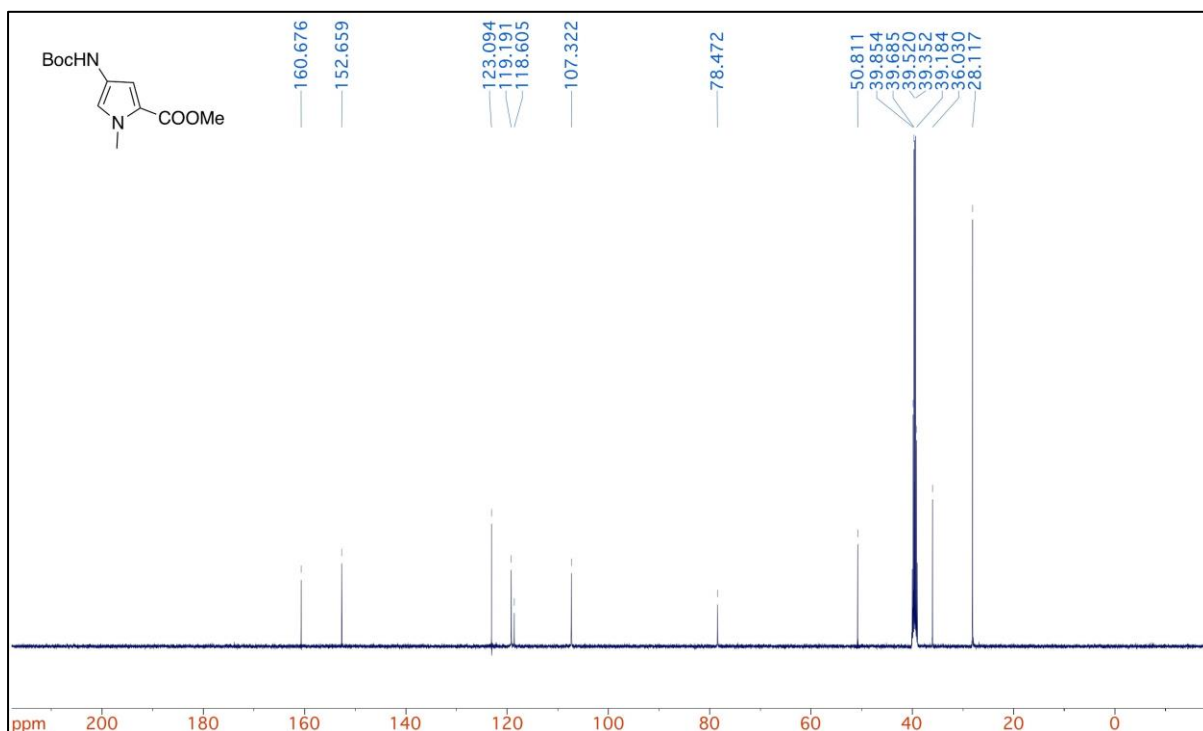
Scan Range: 100 to 1000 AMU positive, 120-1000 AMU negative.

2 Spectroscopic characterization of compounds described in Chapter 2

Methyl 4-((tert-butoxycarbonyl)amino)-1-methyl-1H-pyrrole-2-carboxylate (**2.92**)

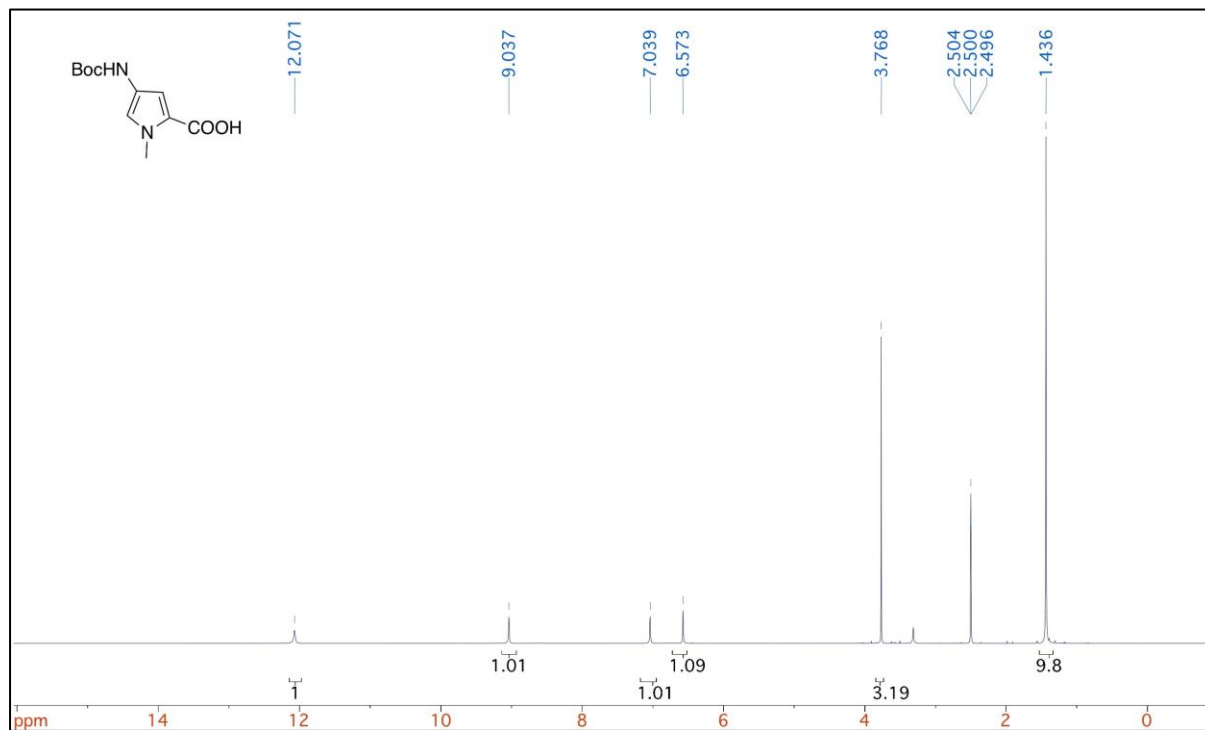


Appendix 1 $^1\text{H-NMR}$ of **2.92**.

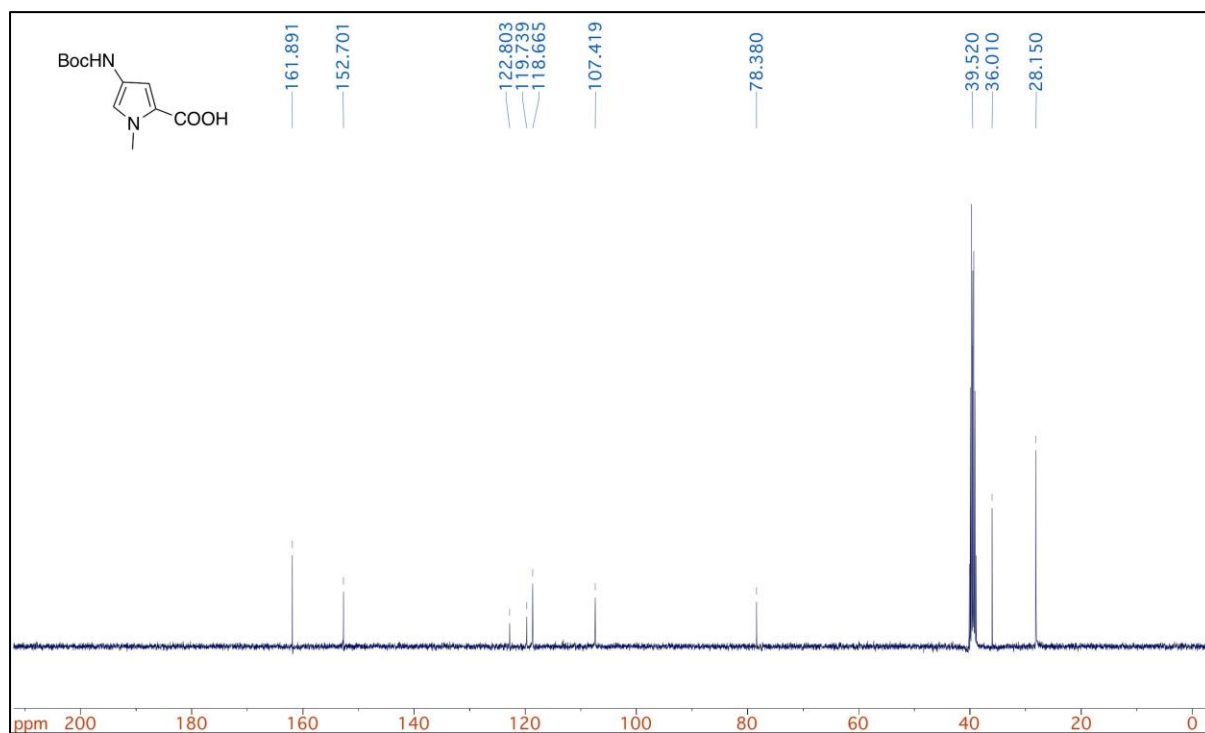


Appendix 2 $^{13}\text{C-NMR}$ of **2.92**.

4-((*tert*-butoxycarbonyl)amino)-1-methyl-1*H*-pyrrole-2-carboxylic acid (**2.25**)

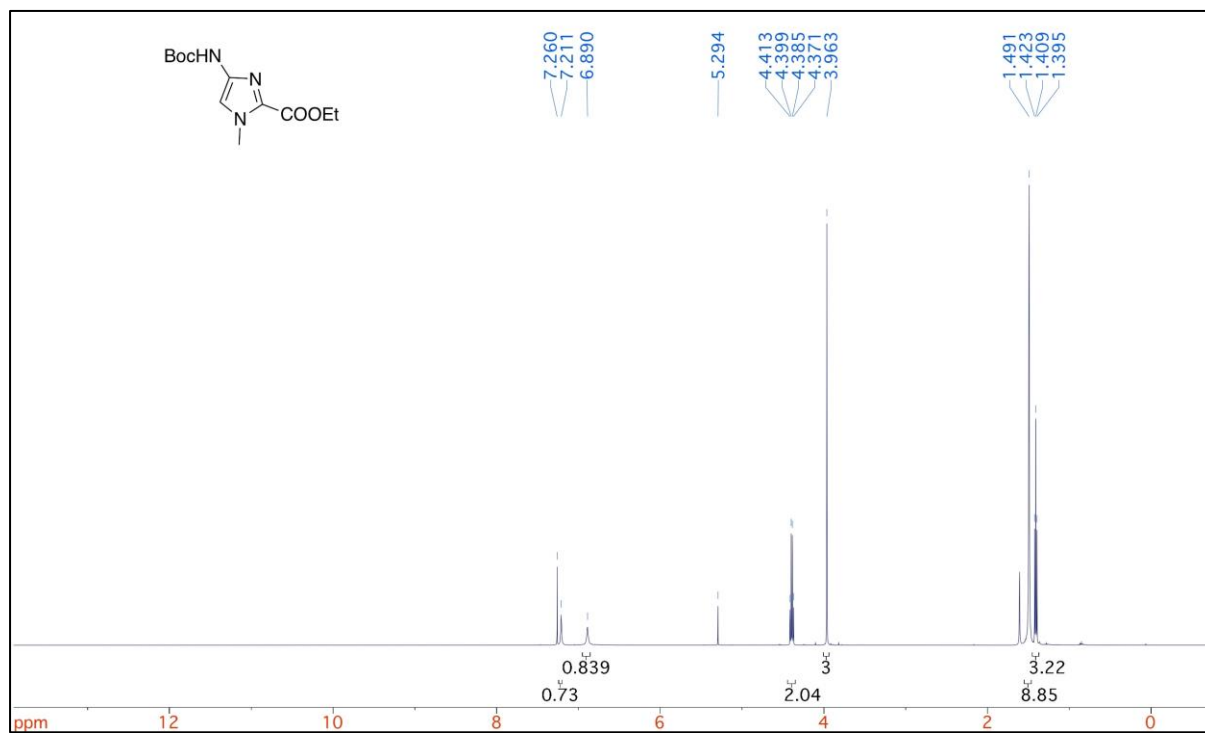


Appendix 3 ¹H-NMR of **2.25**.

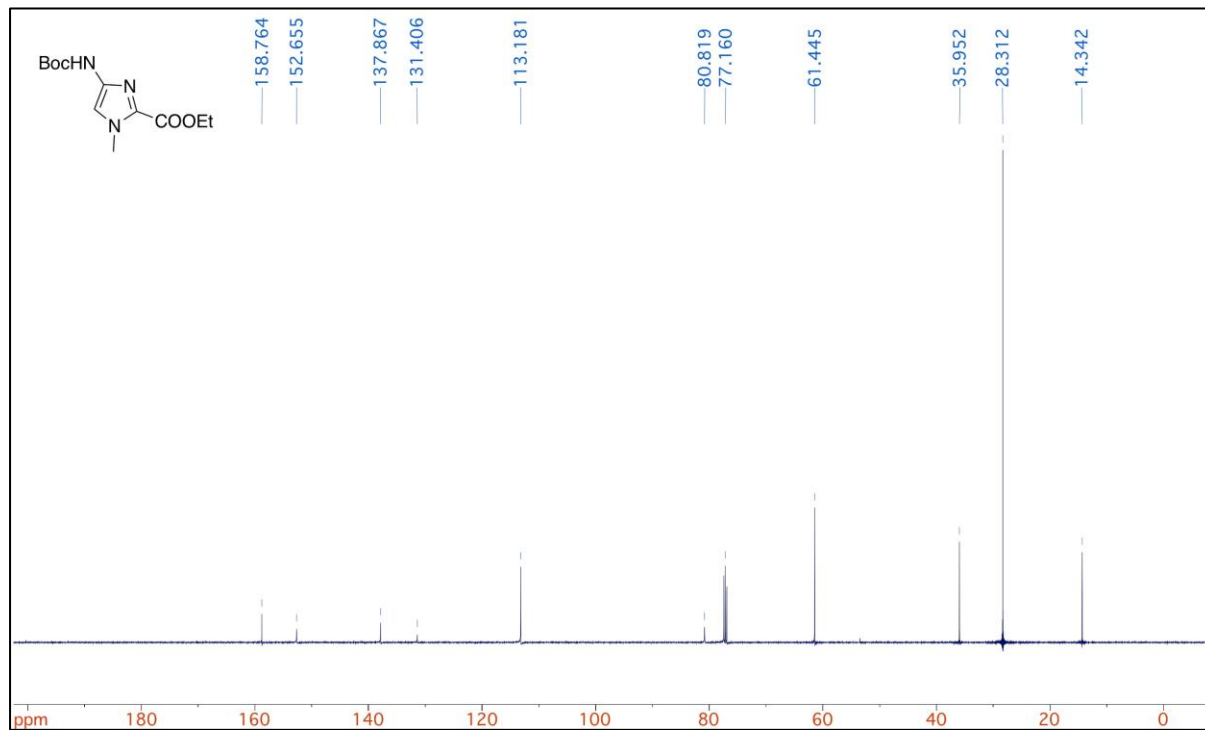


Appendix 4 ¹³C-NMR of **2.25**.

Ethyl 4-((tert-butoxycarbonyl)amino)-1-methyl-1H-imidazole-2-carboxylate (**2.93**)

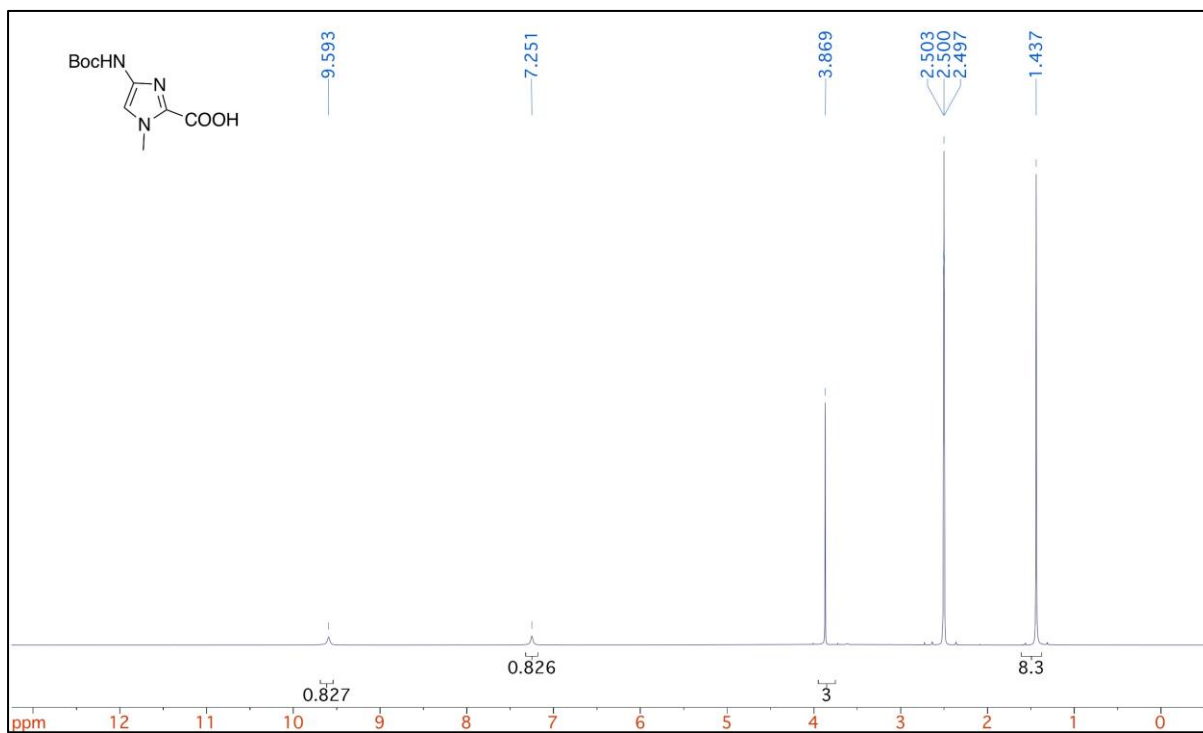


Appendix 5 $^1\text{H-NMR}$ of **2.93**.

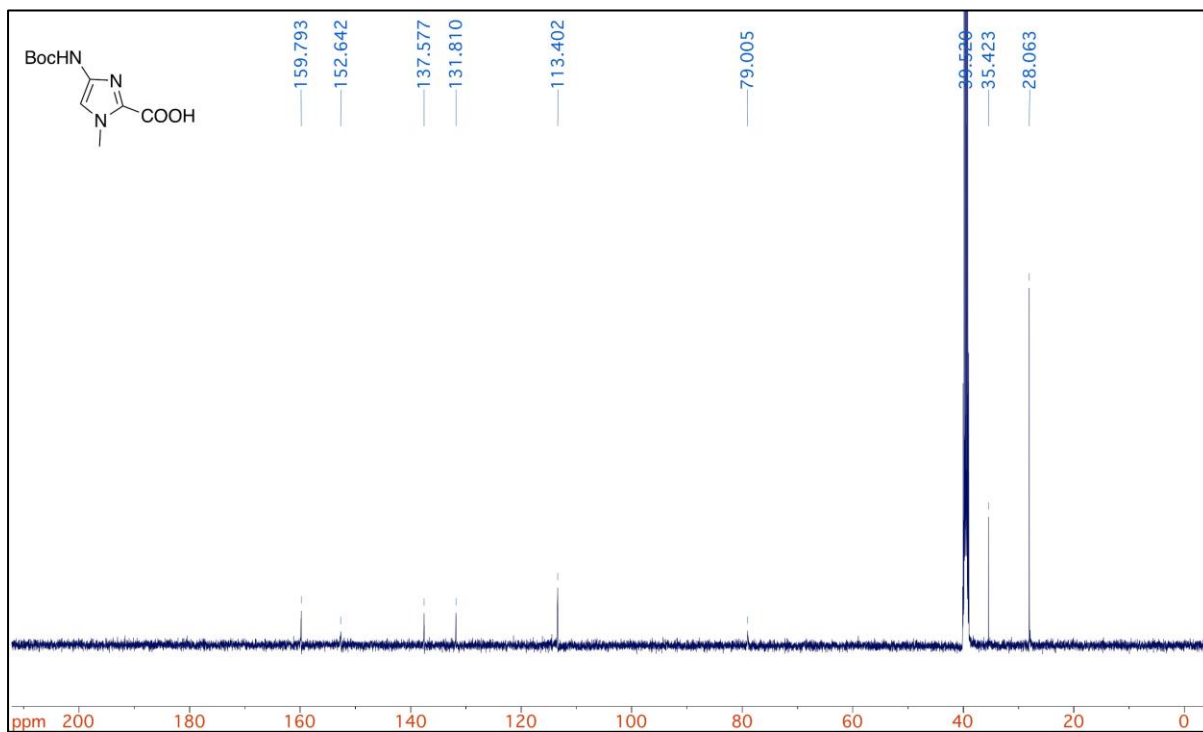


Appendix 6 $^{13}\text{C-NMR}$ of **2.93**.

4-((tert-butoxycarbonyl)amino)-1-methyl-1H-imidazole-2-carboxylic acid (**2.26**)

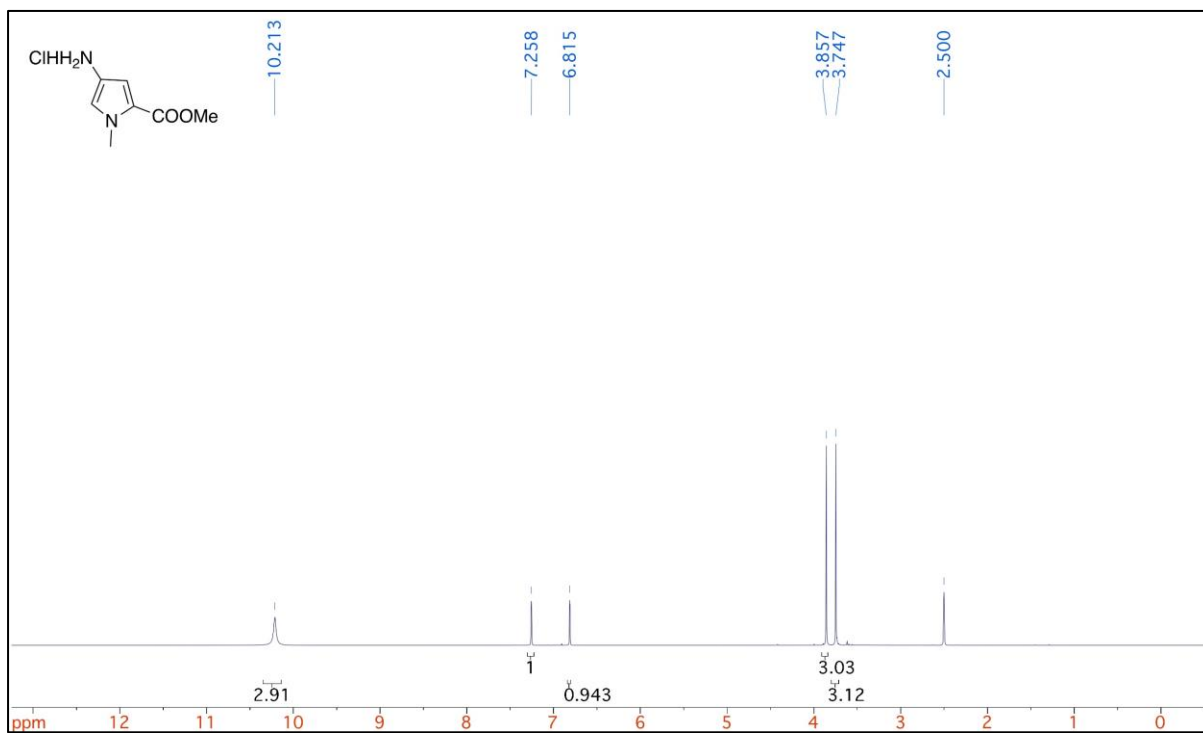


Appendix 7 ¹H-NMR of **2.26**.

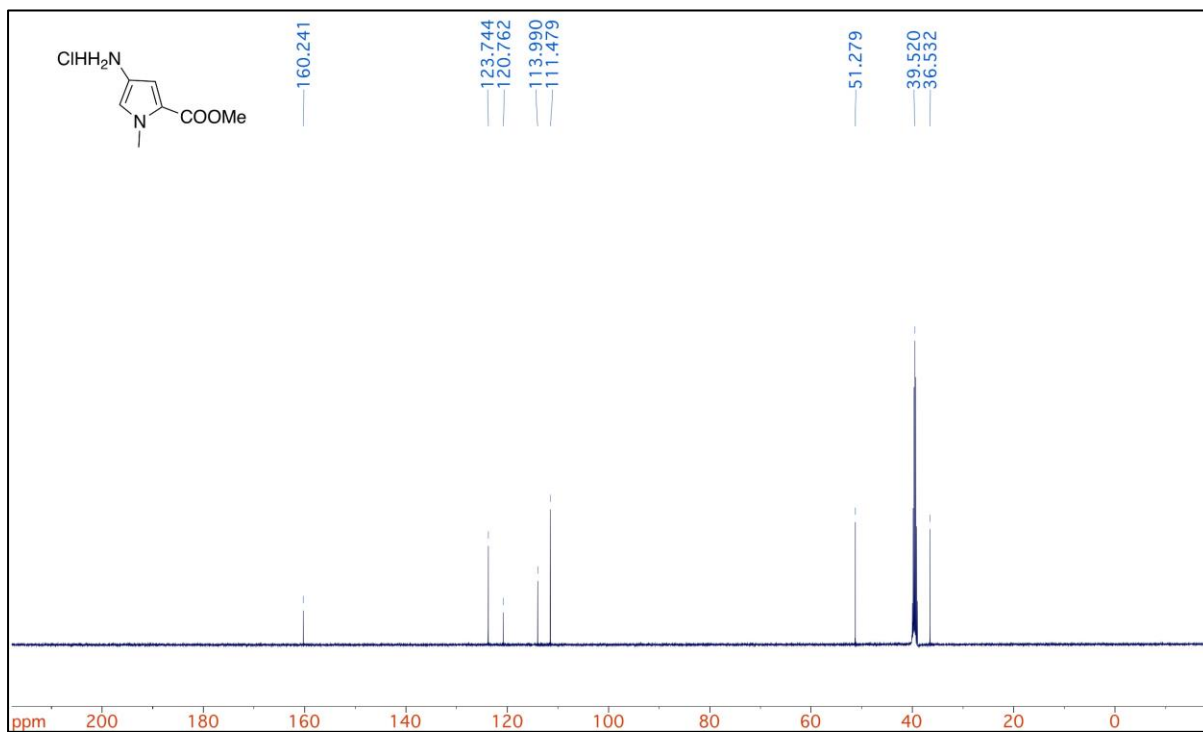


Appendix 8 ¹³C-NMR of **2.26**.

Methyl 4-amino-1-methyl-1H-pyrrole-2-carboxylate hydrochloride (**2.39**)

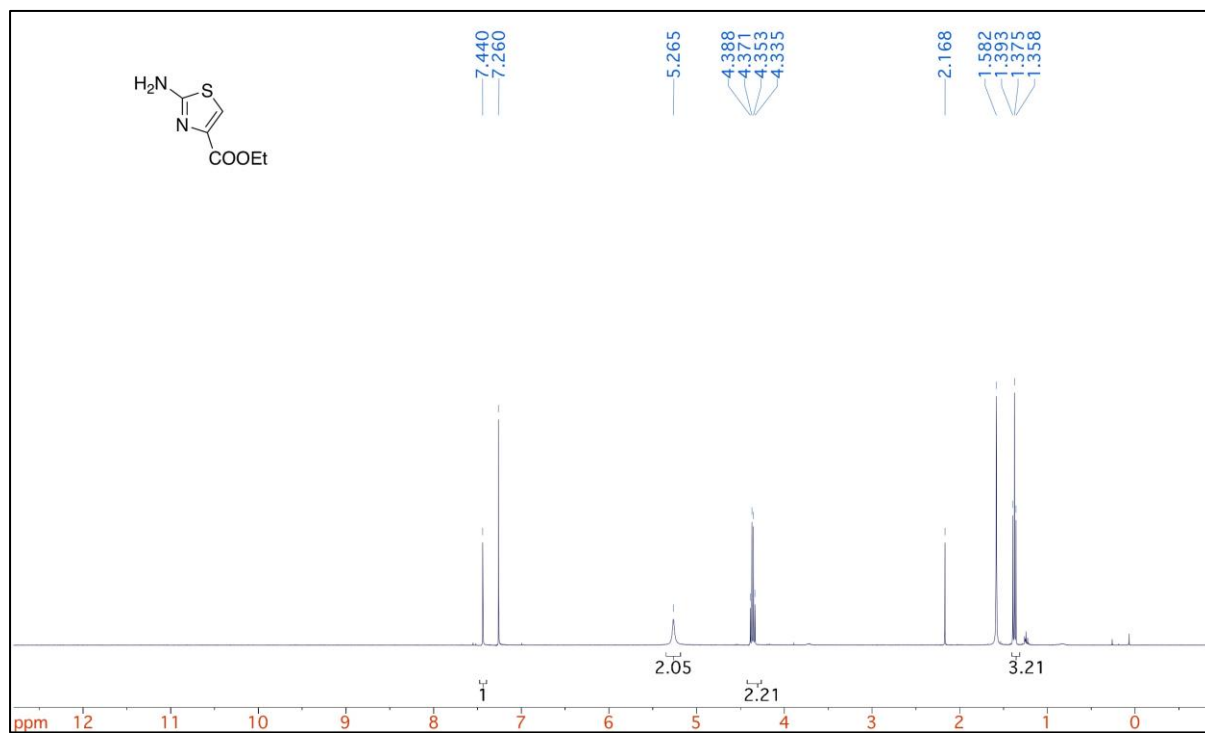


Appendix 9 $^1\text{H-NMR}$ of **2.39**.

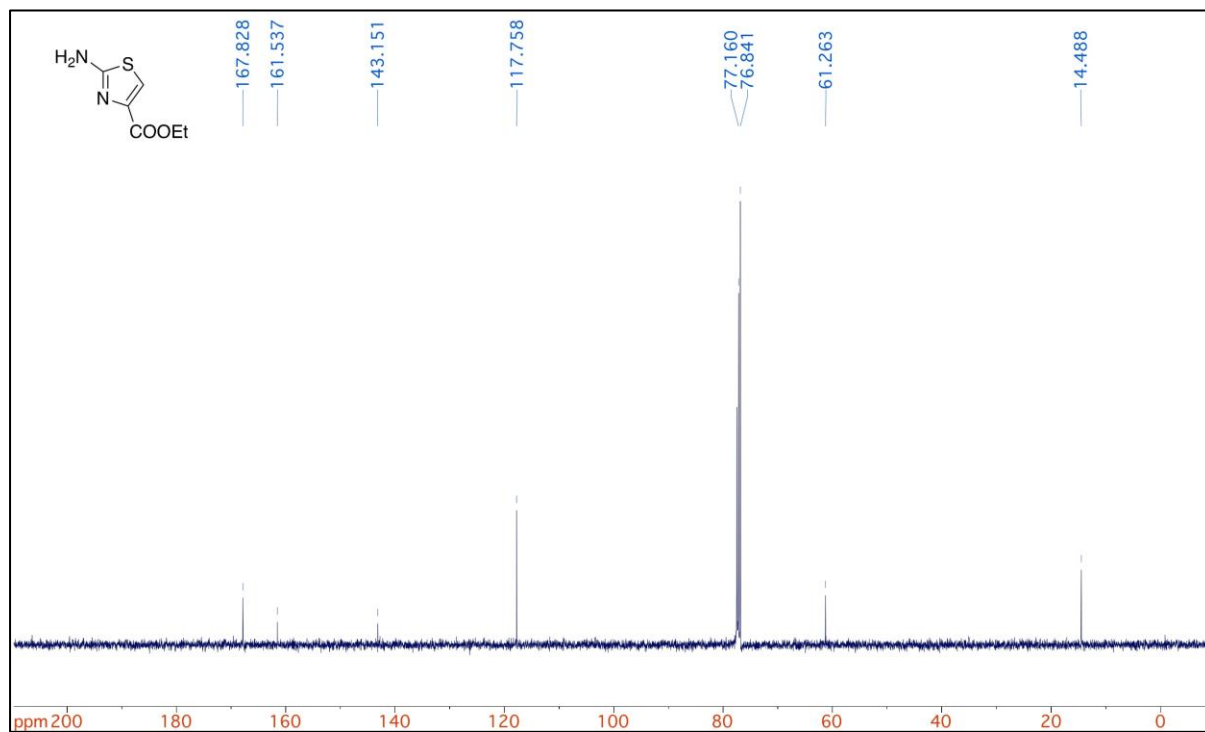


Appendix 10 $^{13}\text{C-NMR}$ of **2.39**.

Ethyl 2-aminothiazole-4-carboxylate (**2.108**)

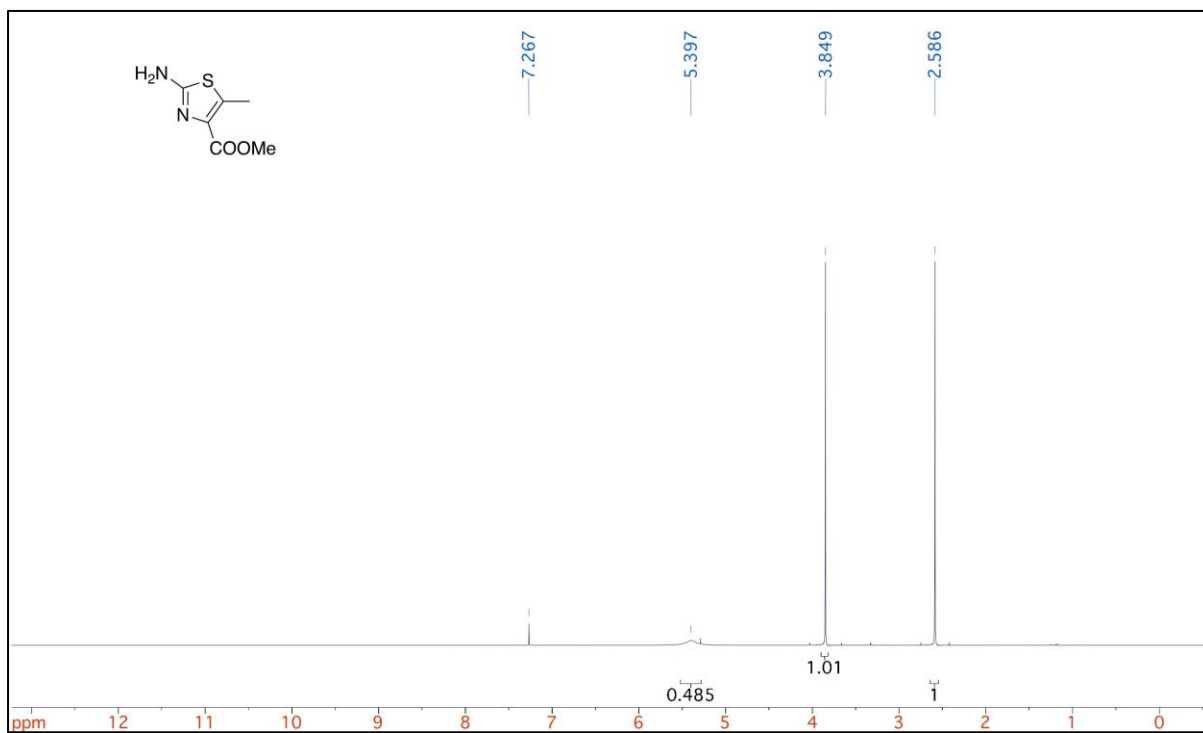


Appendix 11 ¹H-NMR of **2.108**.

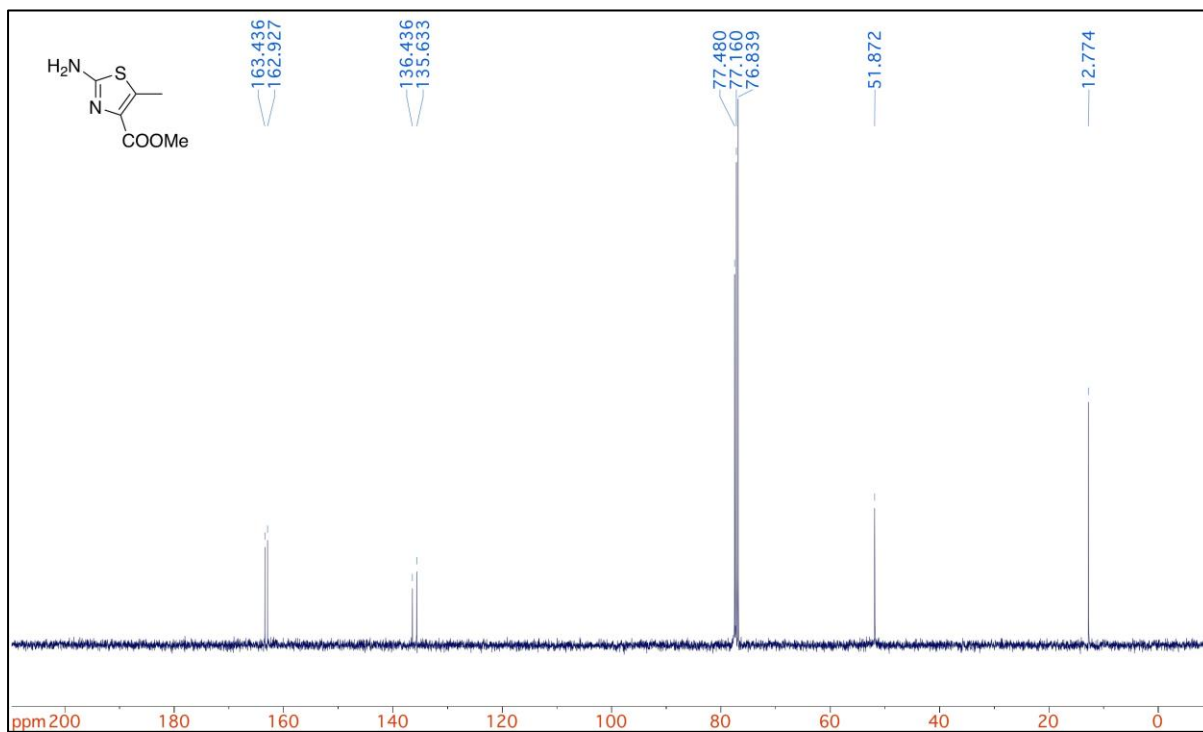


Appendix 12 ¹³C-NMR of **2.108**.

Methyl 2-amino-5-methylthiazole-4-carboxylate (**2.112**)

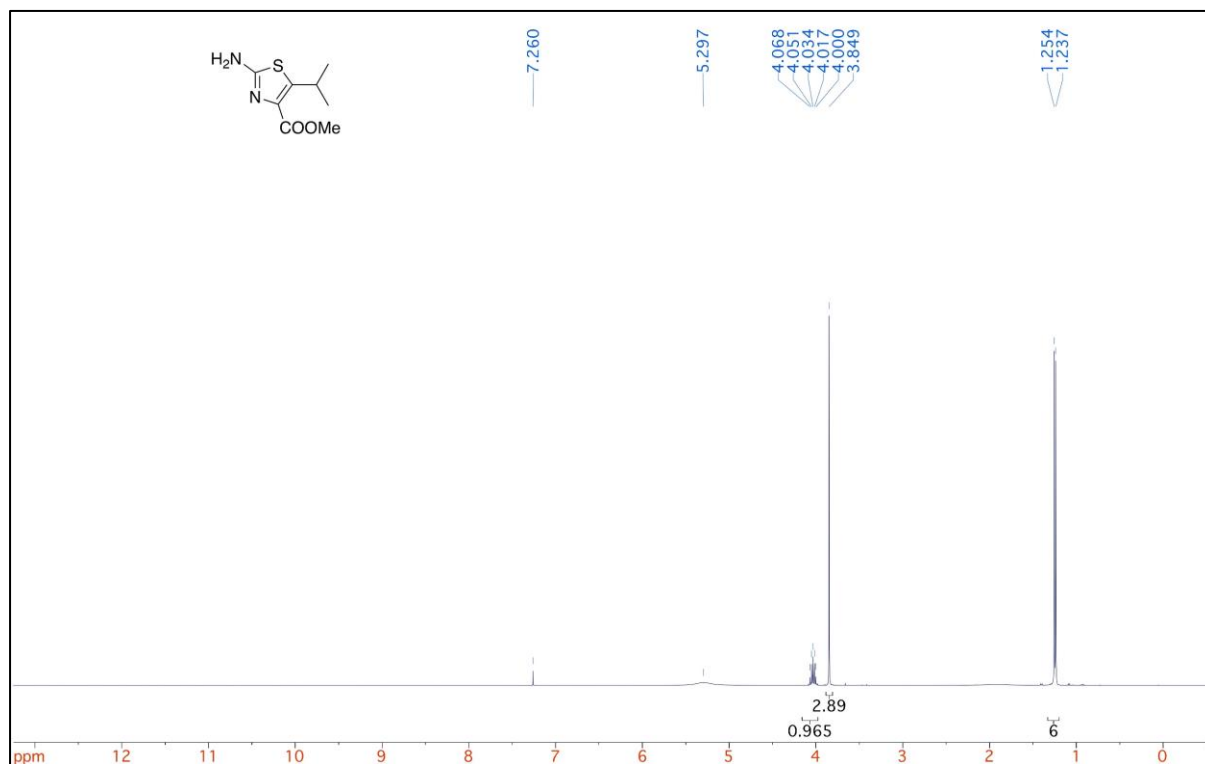


Appendix 13 ¹H-NMR of **2.112**.

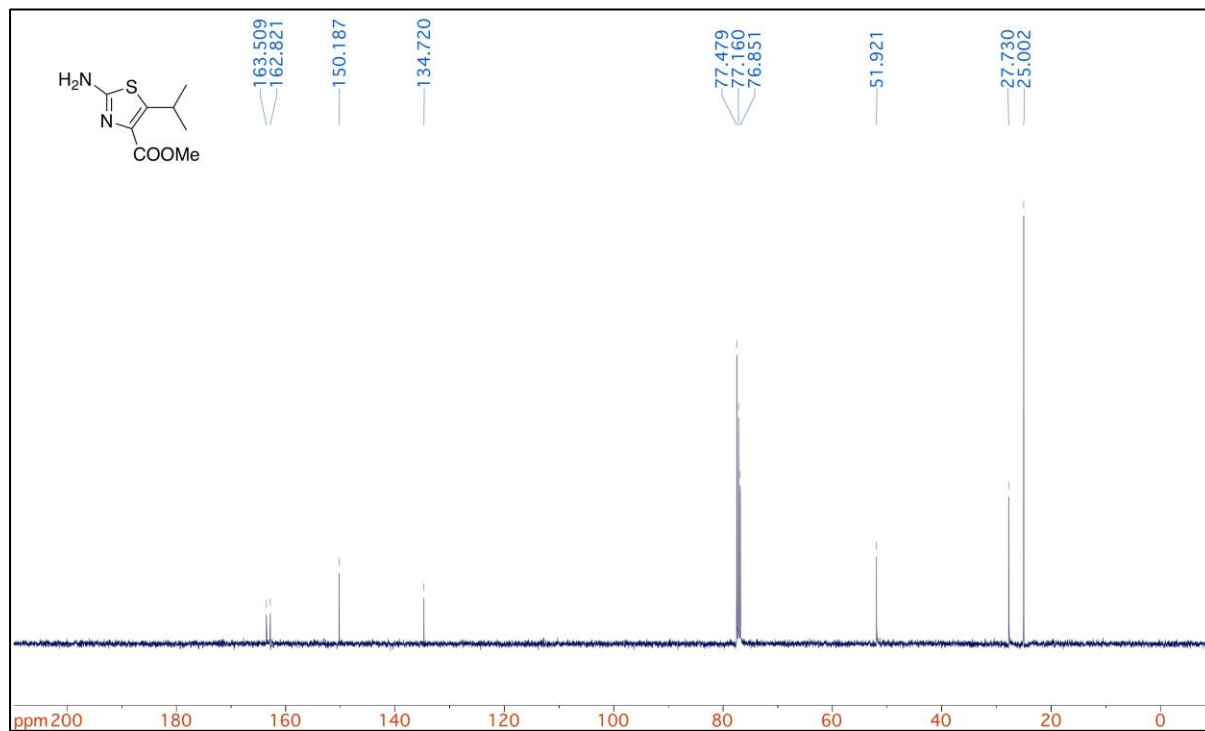


Appendix 14 ¹³C-NMR of **2.112**.

Methyl 2-amino-5-isopropylthiazole-4-carboxylate (**2.113**)

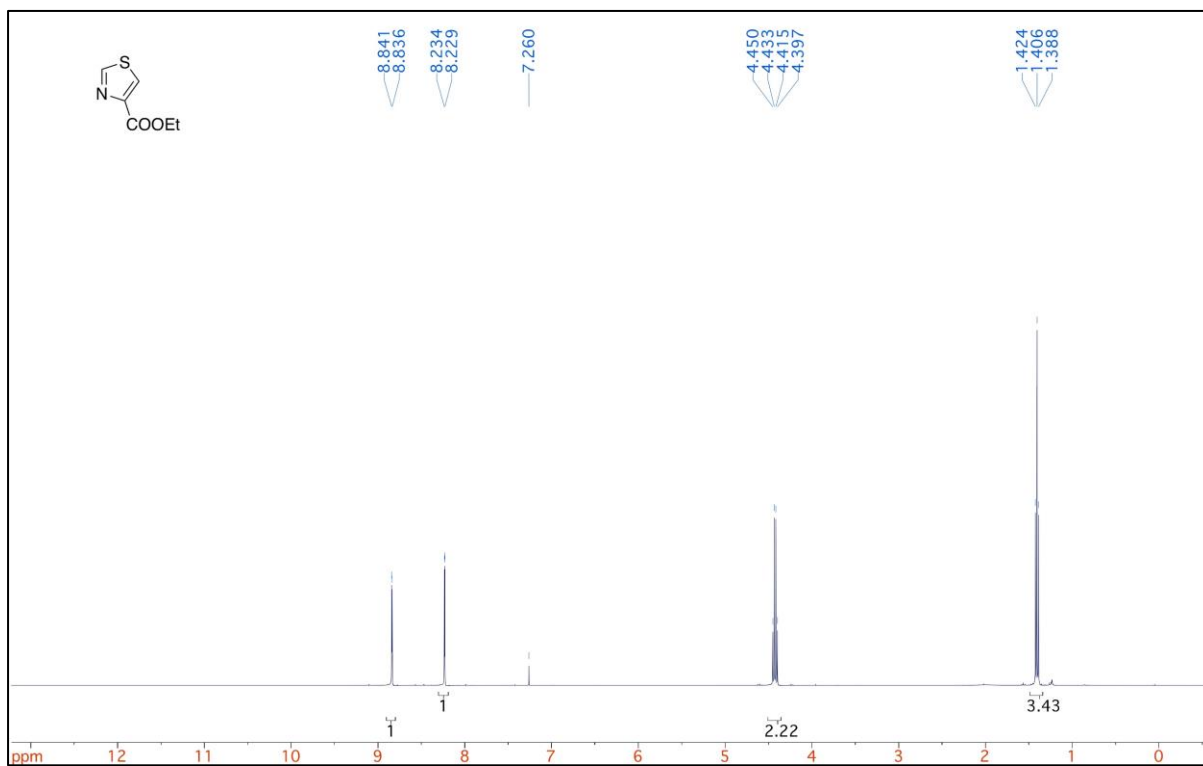


Appendix 15 $^1\text{H-NMR}$ of **2.113**.

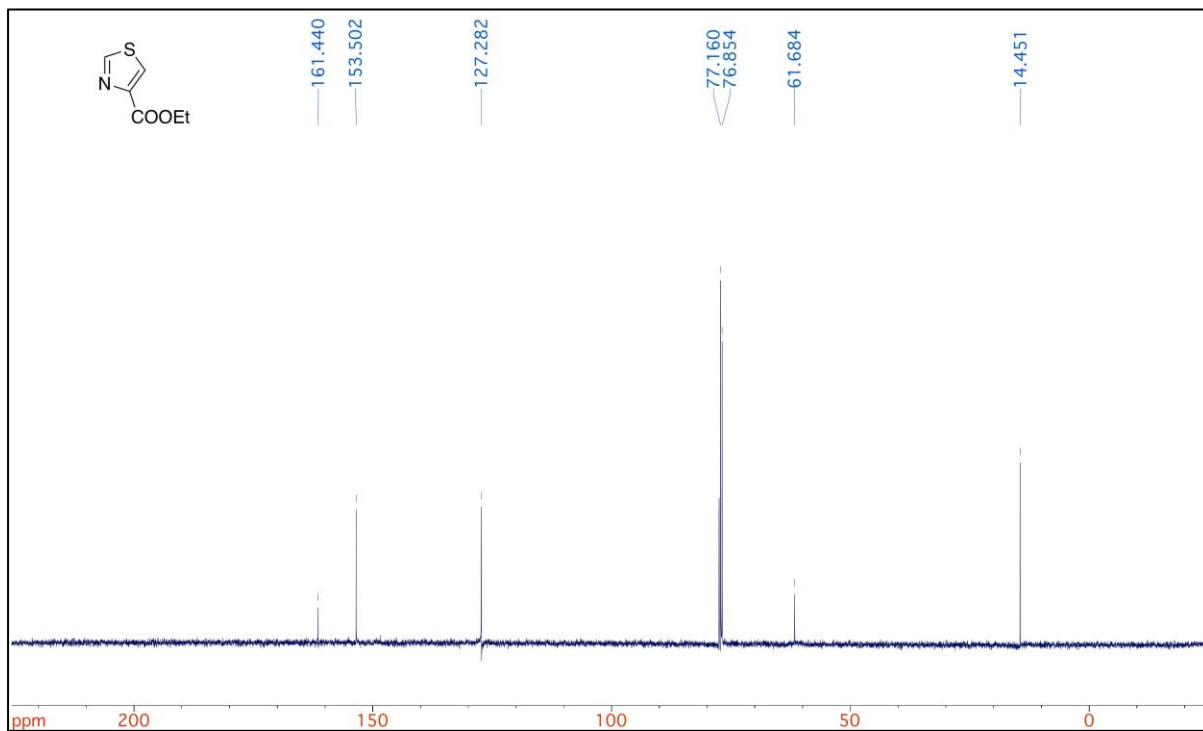


Appendix 16 $^{13}\text{C-NMR}$ of **2.113**.

Ethyl thiazole-4-carboxylate (**2.114a**)

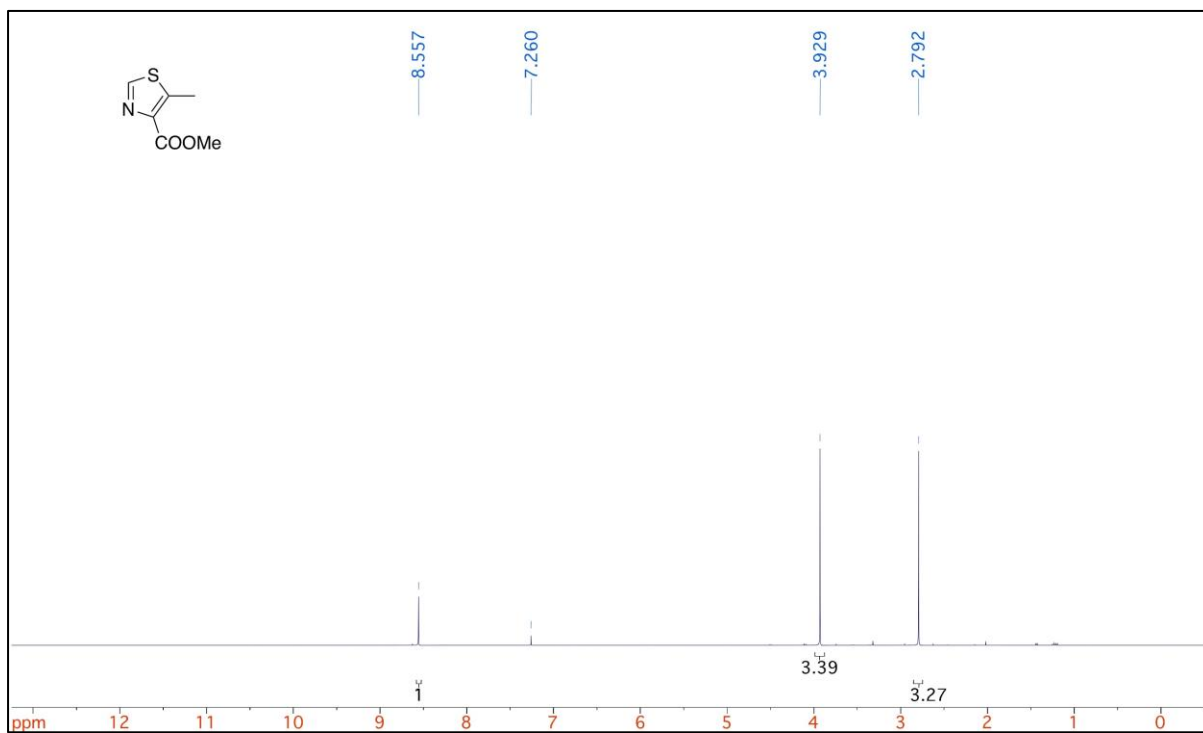


Appendix 17 ¹H-NMR of **2.114a**.

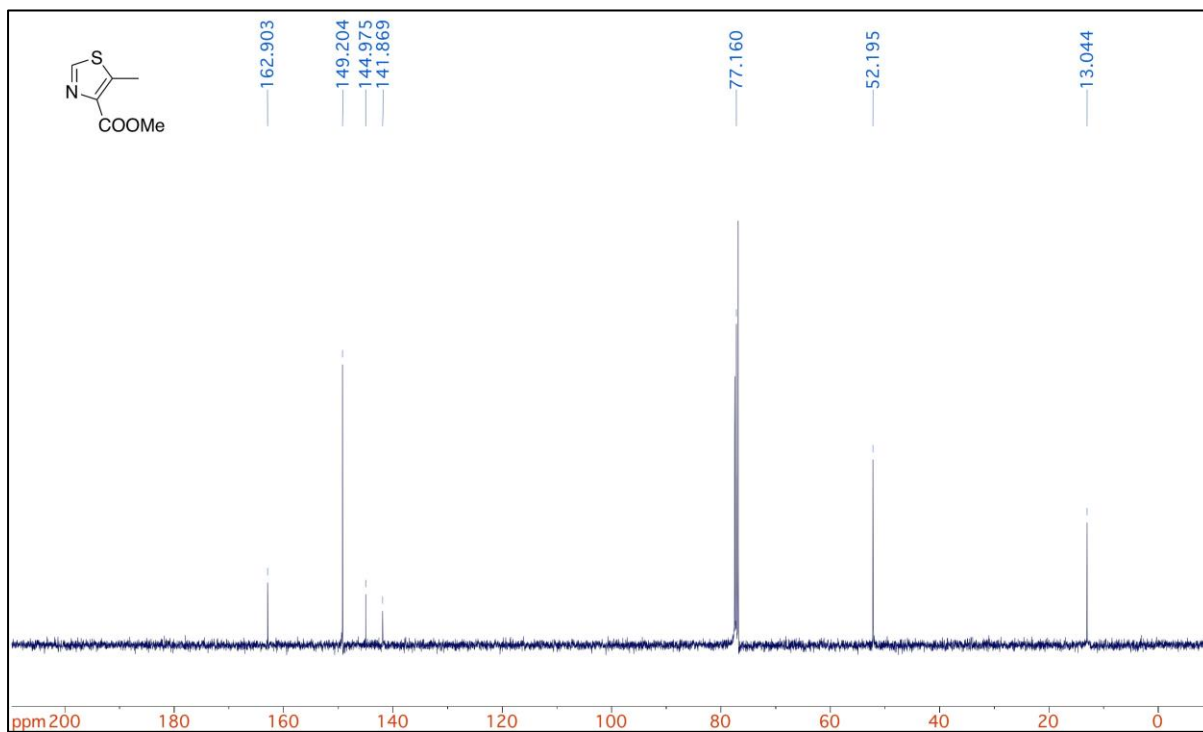


Appendix 18 ¹³C-NMR of **2.114a**.

Methyl 5-methylthiazole-4-carboxylate (**2.114b**)

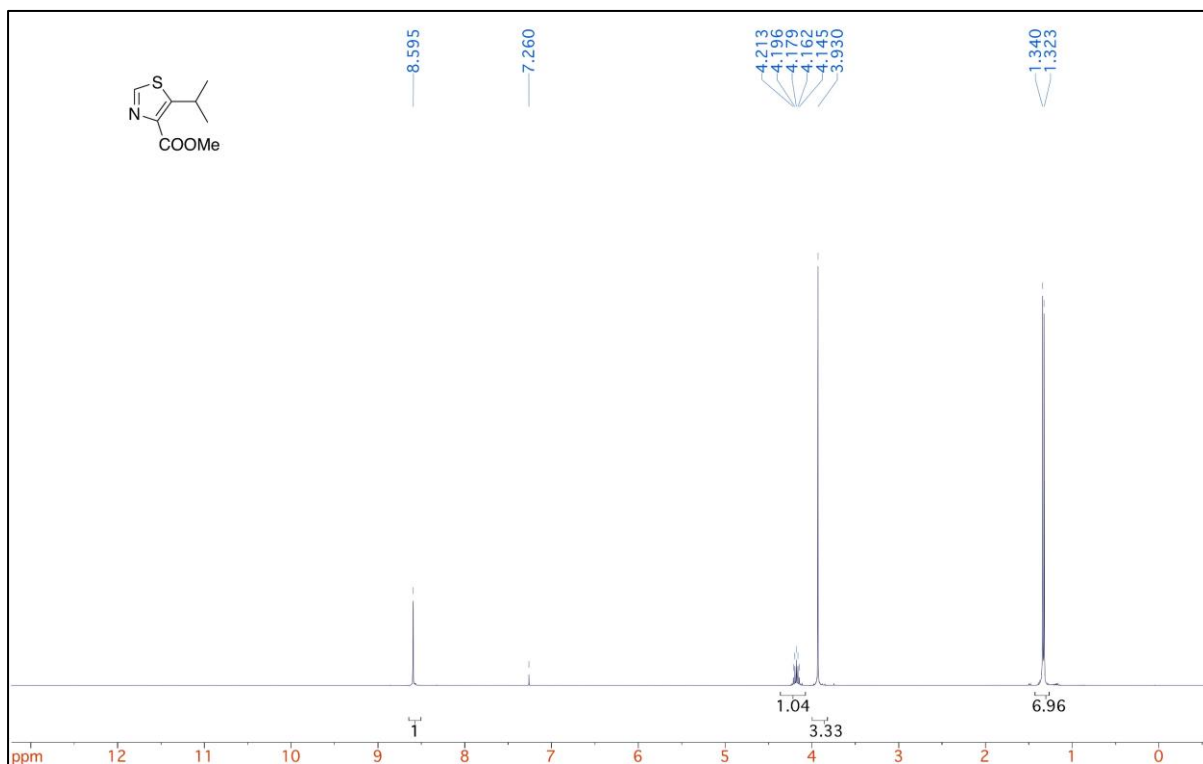


Appendix 19 ¹H-NMR of **2.114b**.

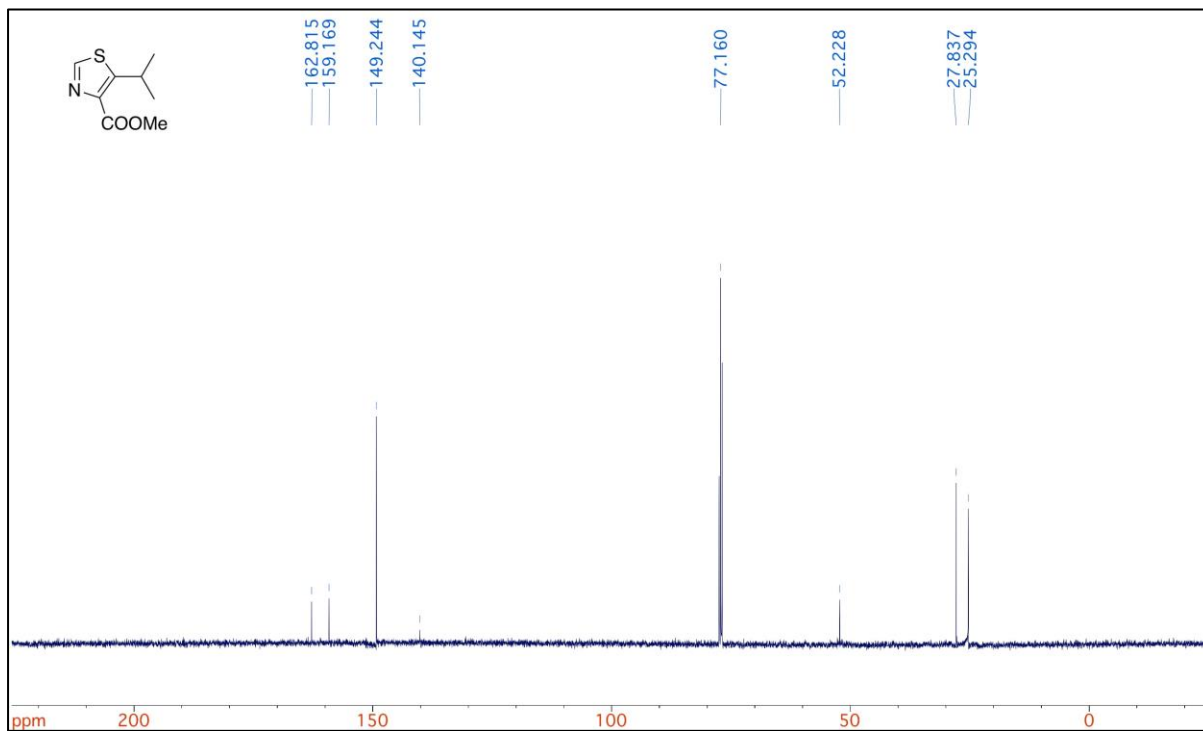


Appendix 20 ¹³C-NMR of **2.114b**.

Methyl 5-isopropylthiazole-4-carboxylate (**2.114c**)

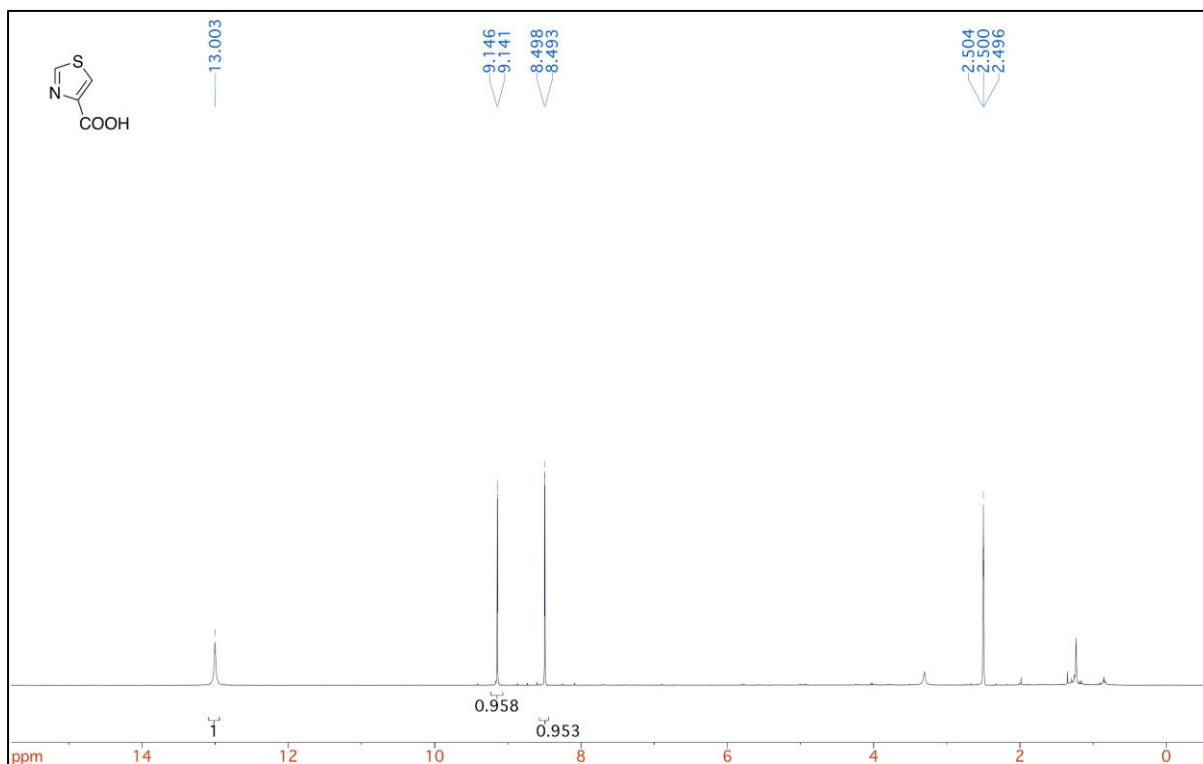


Appendix 21 ¹H-NMR of **2.114c**.

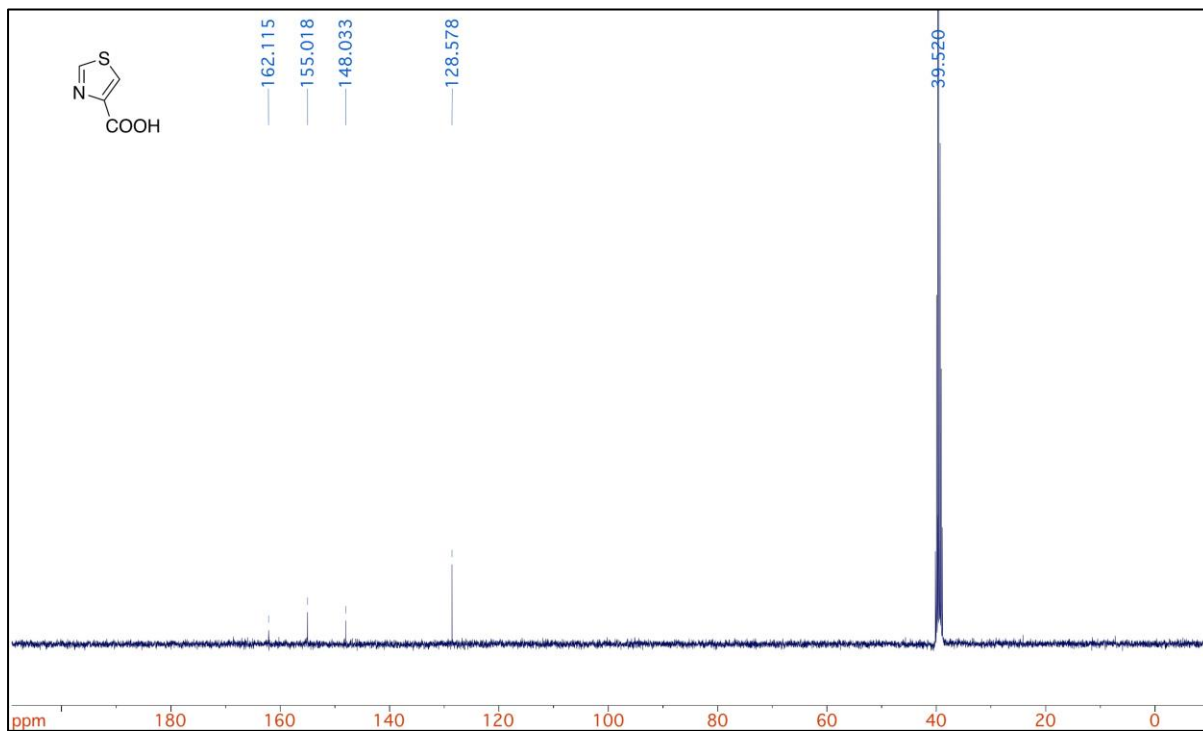


Appendix 22 ¹³C-NMR of **2.114c**.

Thiazole-4-carboxylic acid (**2.87a**)

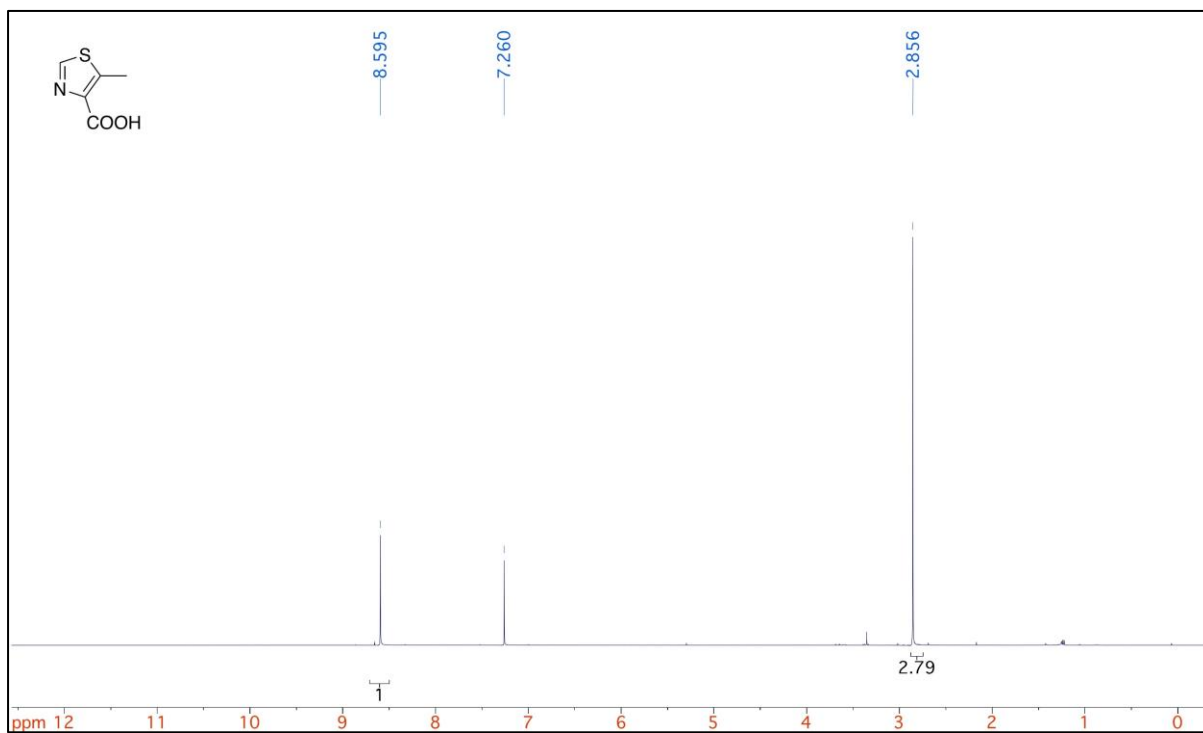


Appendix 23 ¹H-NMR of **2.87a**.

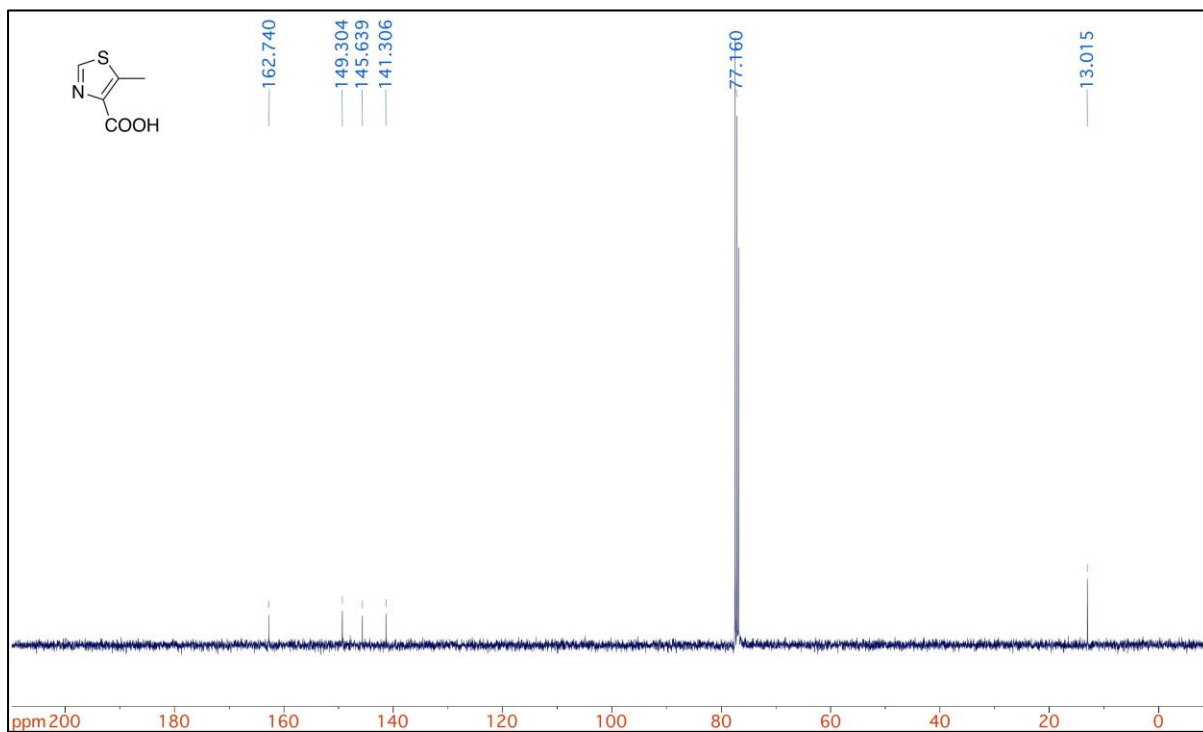


Appendix 24 ¹³C-NMR of **2.87a**.

5-methylthiazole-4-carboxylic acid (**2.87b**)

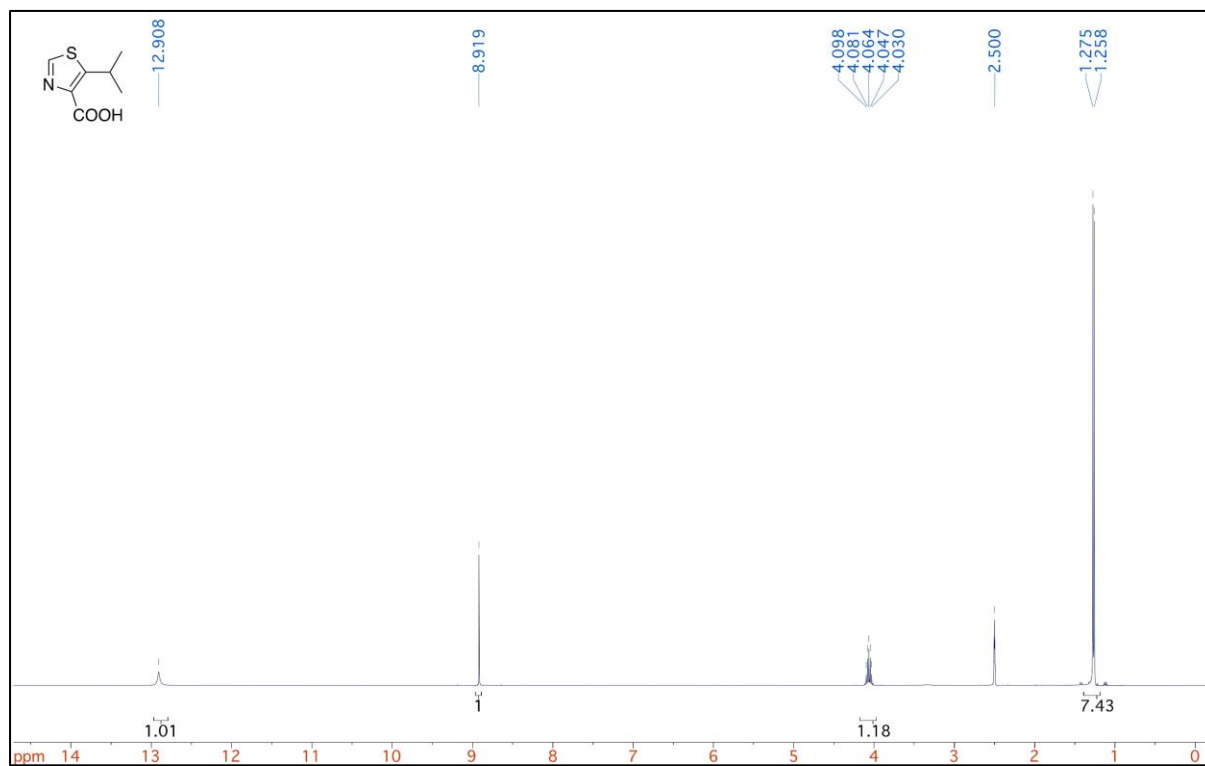


Appendix 25 $^1\text{H-NMR}$ of **2.87b**.

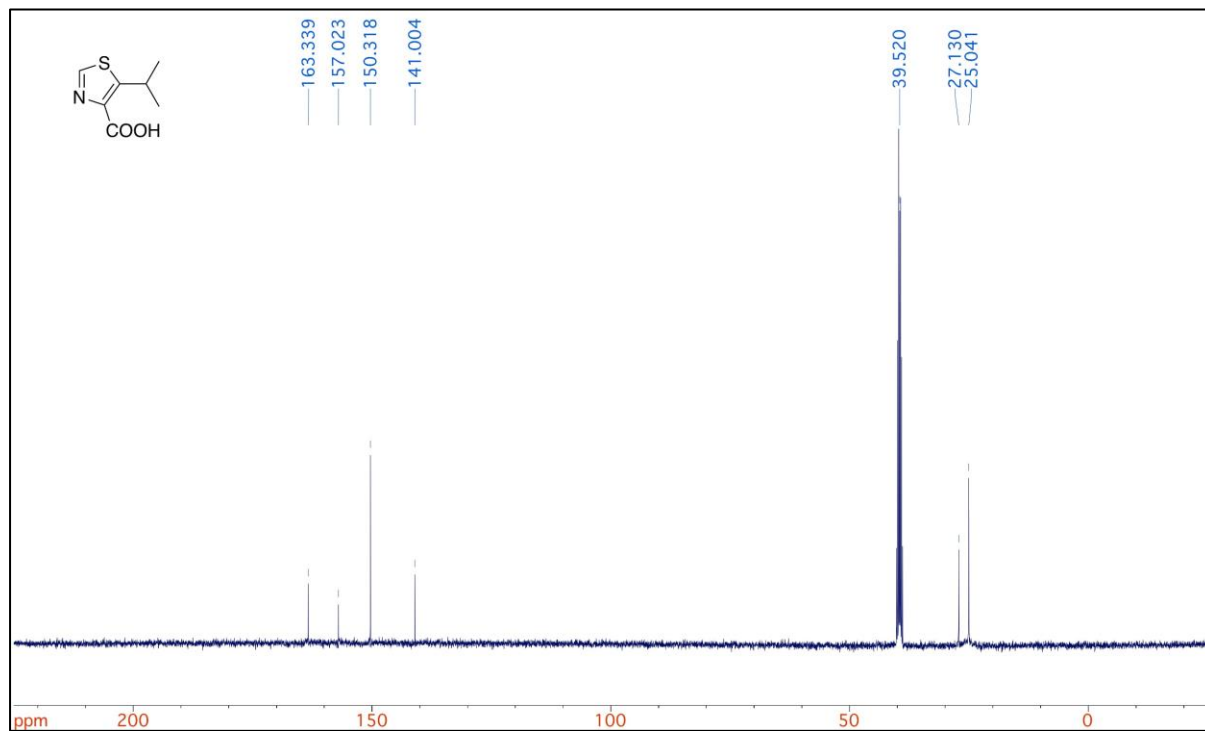


Appendix 26 $^{13}\text{C-NMR}$ of **2.87b**.

5-isopropylthiazole-4-carboxylic acid (**2.87c**)

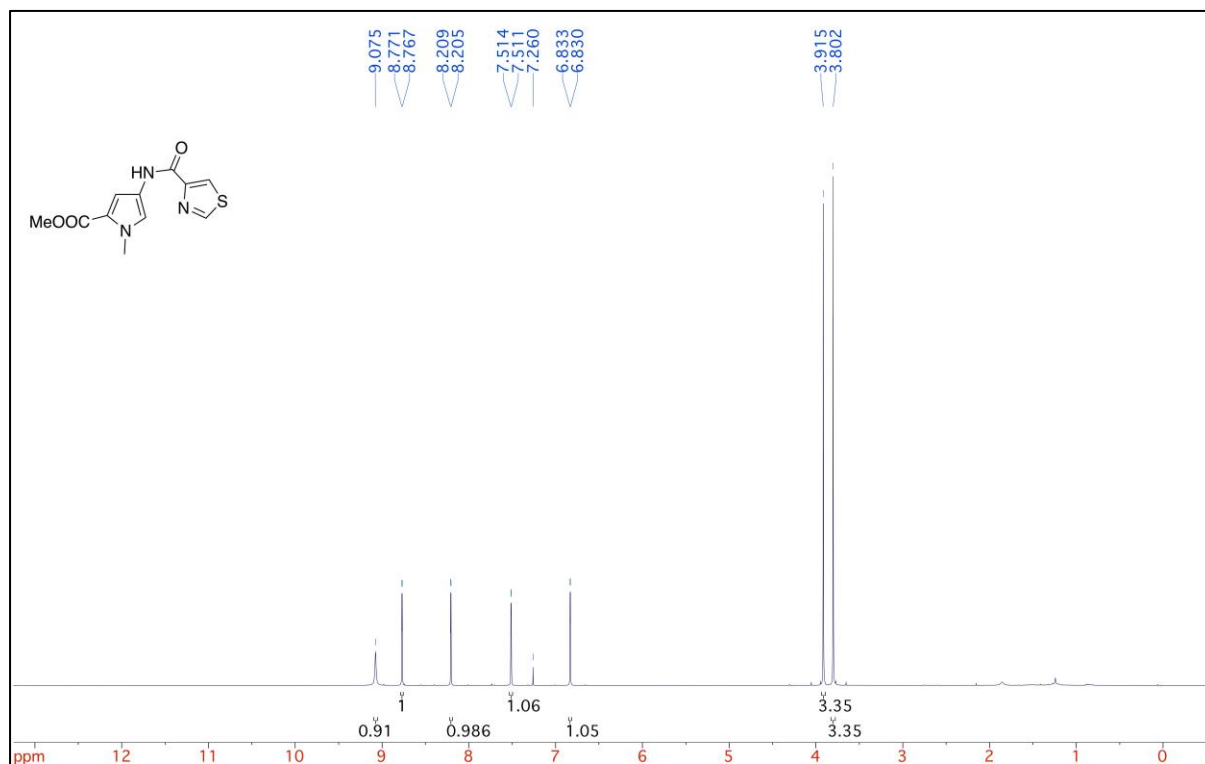


Appendix 27 ¹H-NMR of **2.87c**.

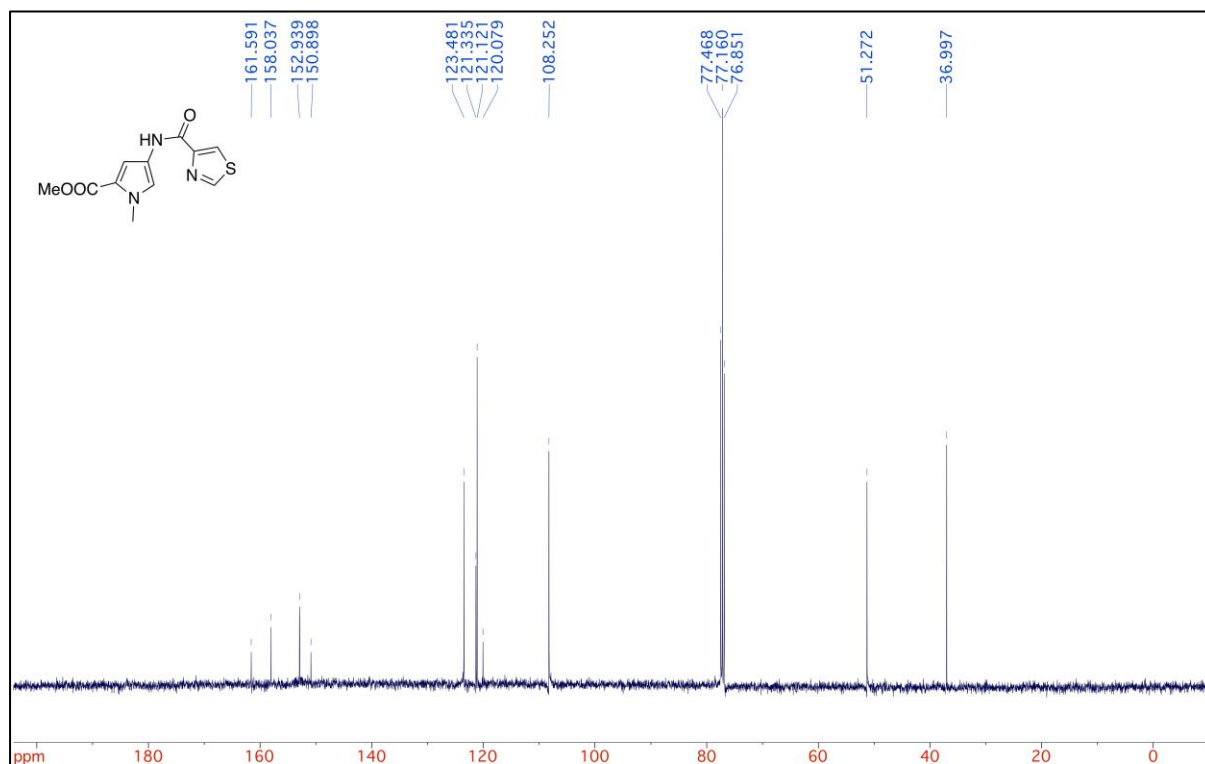


Appendix 28 ¹³C-NMR of **2.87c**.

Methyl 1-methyl-4-(thiazole-4-carboxamido)-1H-pyrrole-2-carboxylate (**2.115a**)

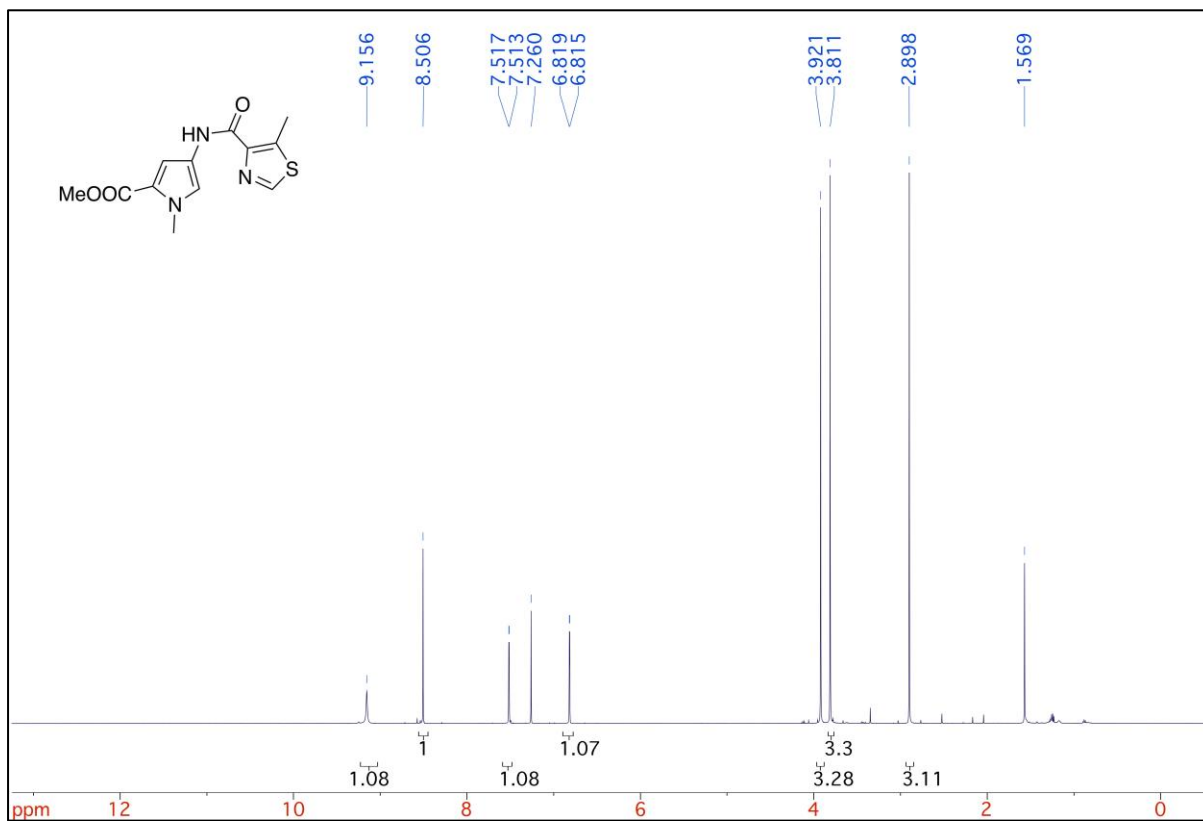


Appendix 29 ¹H-NMR of **2.115a**.

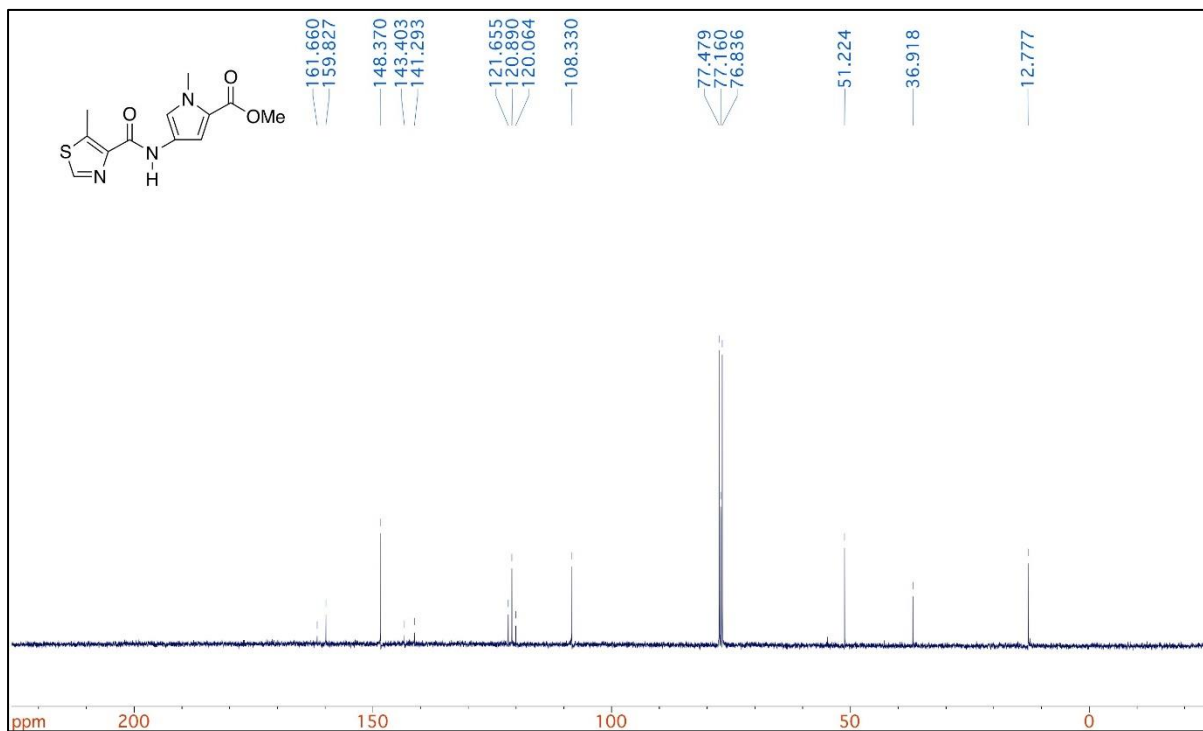


Appendix 30 ¹³C-NMR of **2.115a**.

Methyl 1-methyl-4-(5-methylthiazole-4-carboxamido)-1H-pyrrole-2-carboxylate (**2.115b**)



Appendix 33 ¹H-NMR of **2.115b**.

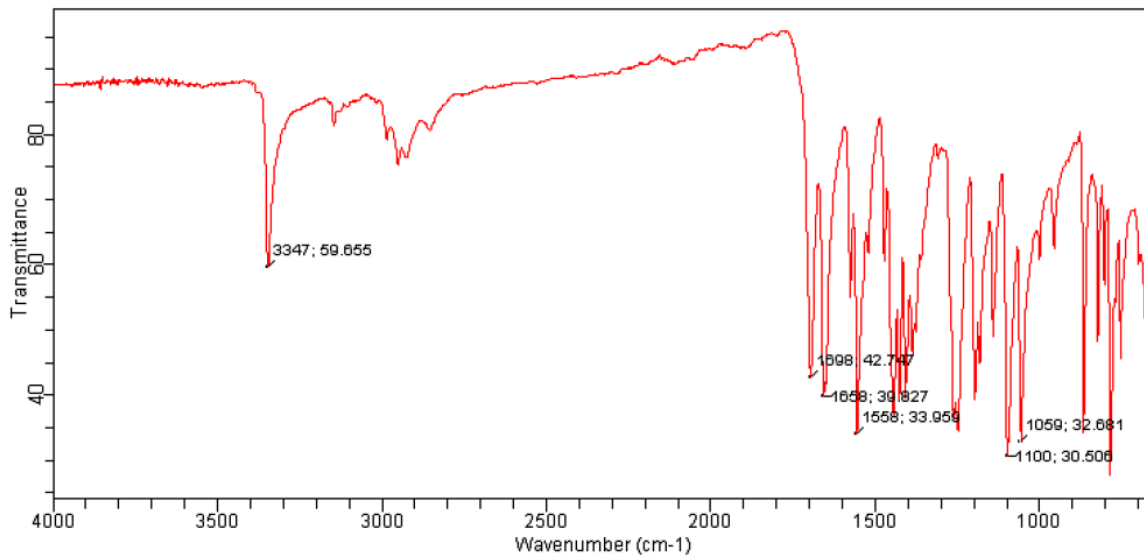


Appendix 34 ¹³C-NMR of **2.115b**.

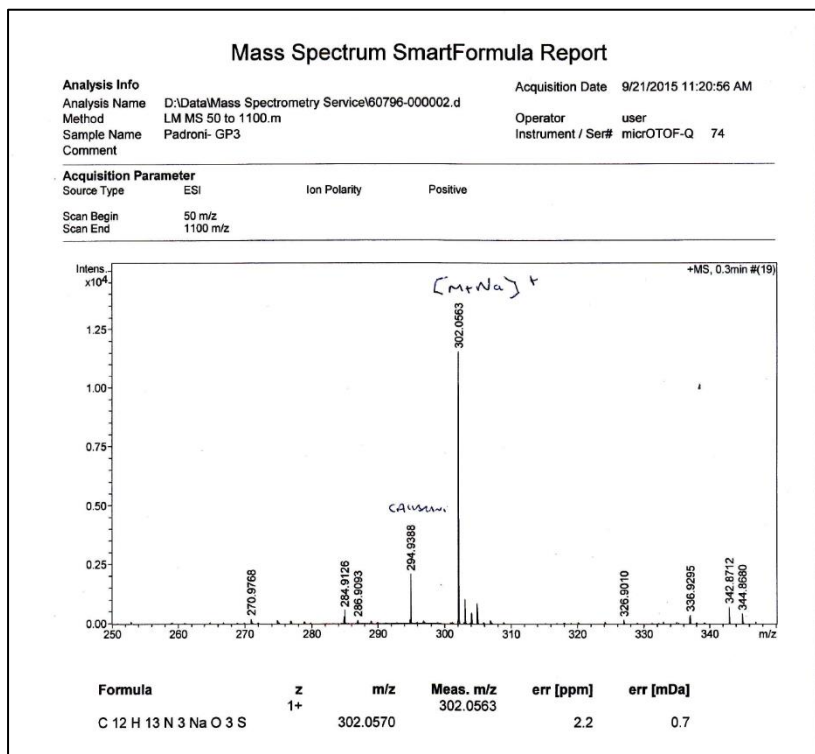


Agilent Technologies

Sample ID: ThCH3PyCOOMe Method Name: STUDENT ATR 32 4cm
Sample Scans: 32 User: STUDENT
Background Scans: 32 Date/Time: 27/08/2015 09:06:08
Resolution: 4 cm-1 Range: 4,000.00 - 650.00
System Status: Good Apodization: Happ-Genzel
File Location: C:\Program Files\Agilent\MicroLab PC\Results\STUDENT ATR 32
4cm\ThCH3PyCOOMe_2015-08-27T09-07-29.a2r

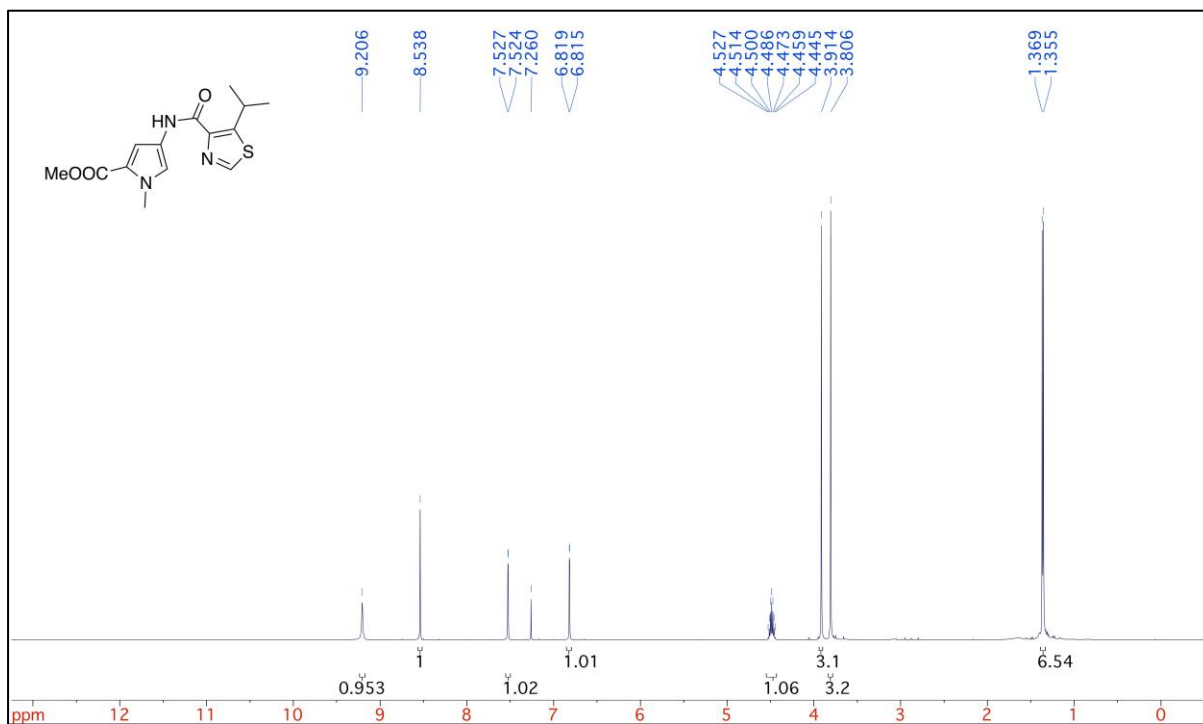


Appendix 35 IR of 2.115b.

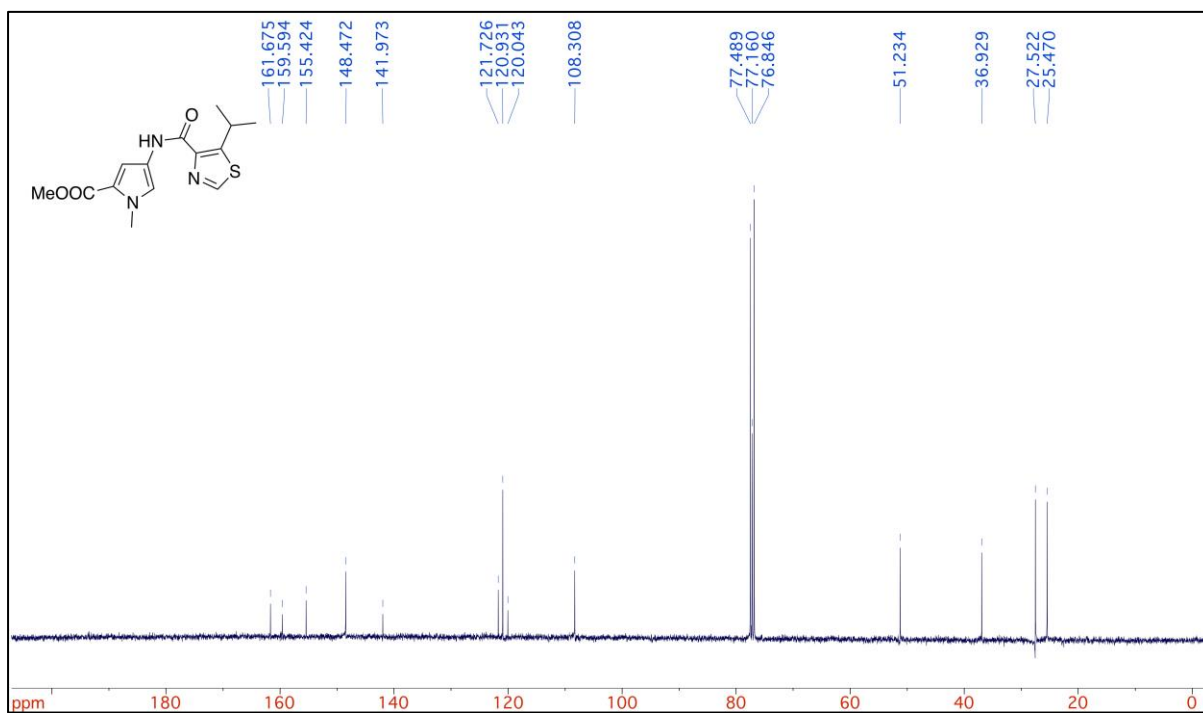


Appendix 36 HRMS of 2.115b.

Methyl 4-(5-isopropylthiazole-4-carboxamido)-1-methyl-1H-pyrrole-2-carboxylate (**2.115c**)



Appendix 37 ¹H-NMR of **2.115c**.

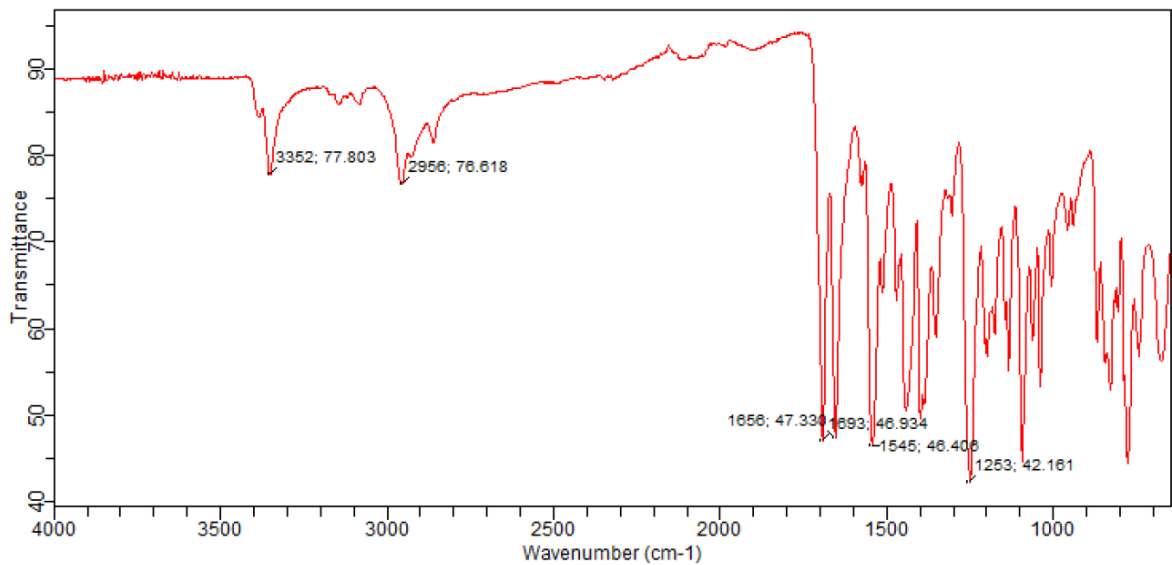


Appendix 38 ¹³C-NMR of **2.115c**.

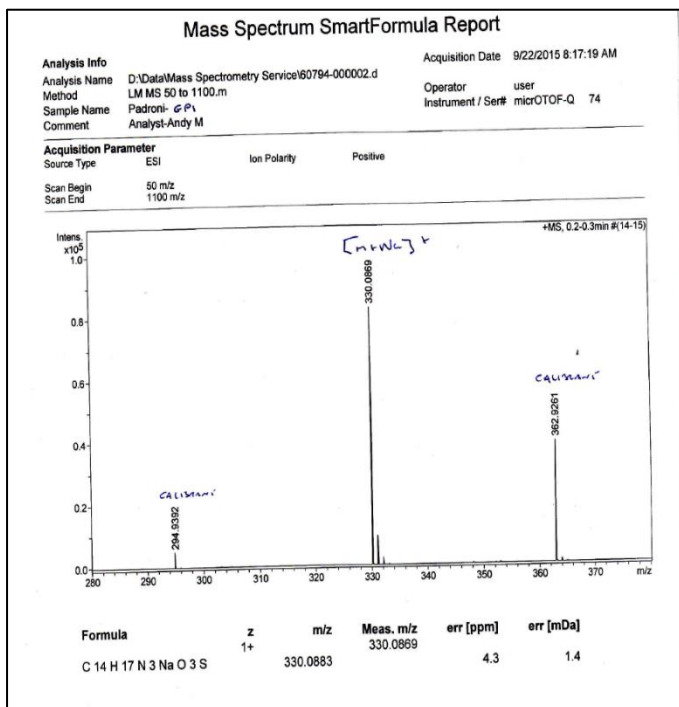


Agilent Technologies

Sample ID: ThiPrPyCOOMe Method Name: STUDENT ATR 32 4cm
Sample Scans: 32 User: STUDENT
Background Scans: 32 Date/Time: 10/09/2015 16:32:42
Resolution: 4 cm⁻¹ Range: 4,000.00 - 650.00
System Status: Good Apodization: Happ-Genzel
File Location: C:\Program Files (x86)\Agilent\MicroLab PC\Results\STUDENT ATR 32
4cm\ThiPrPyCOOMe_2015-09-10T16-34-23.a2r

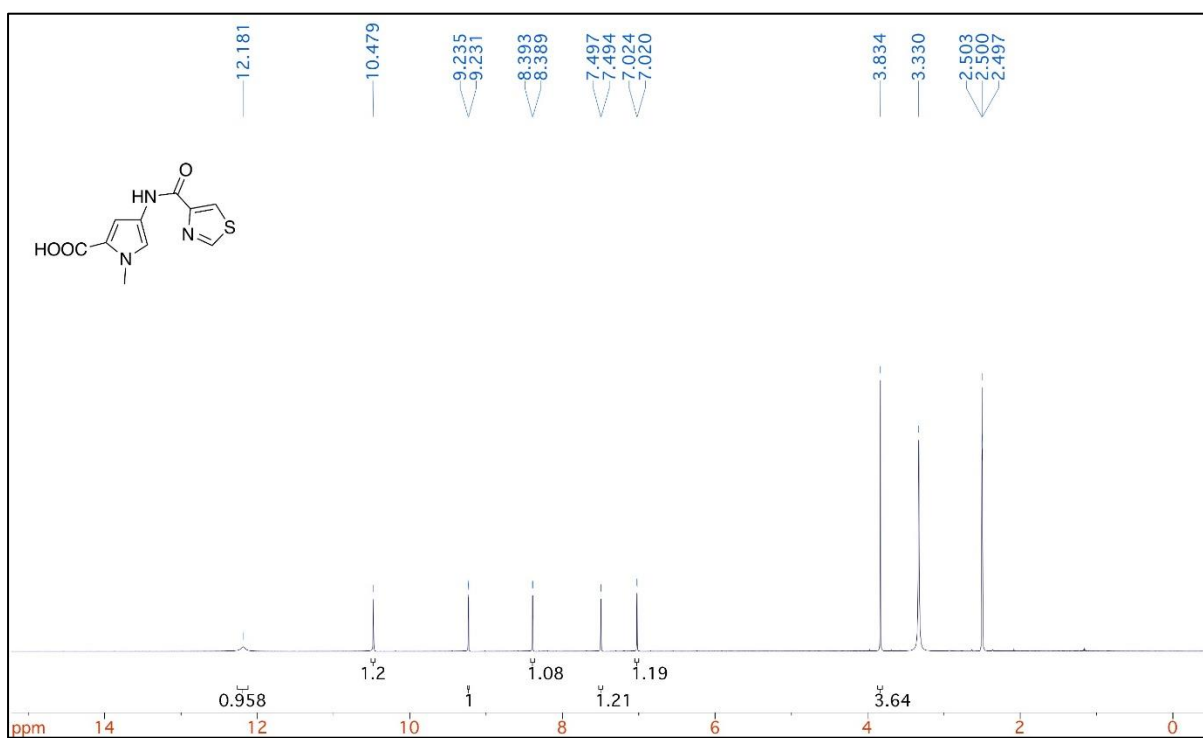


Appendix 39 IR of 2.115c.

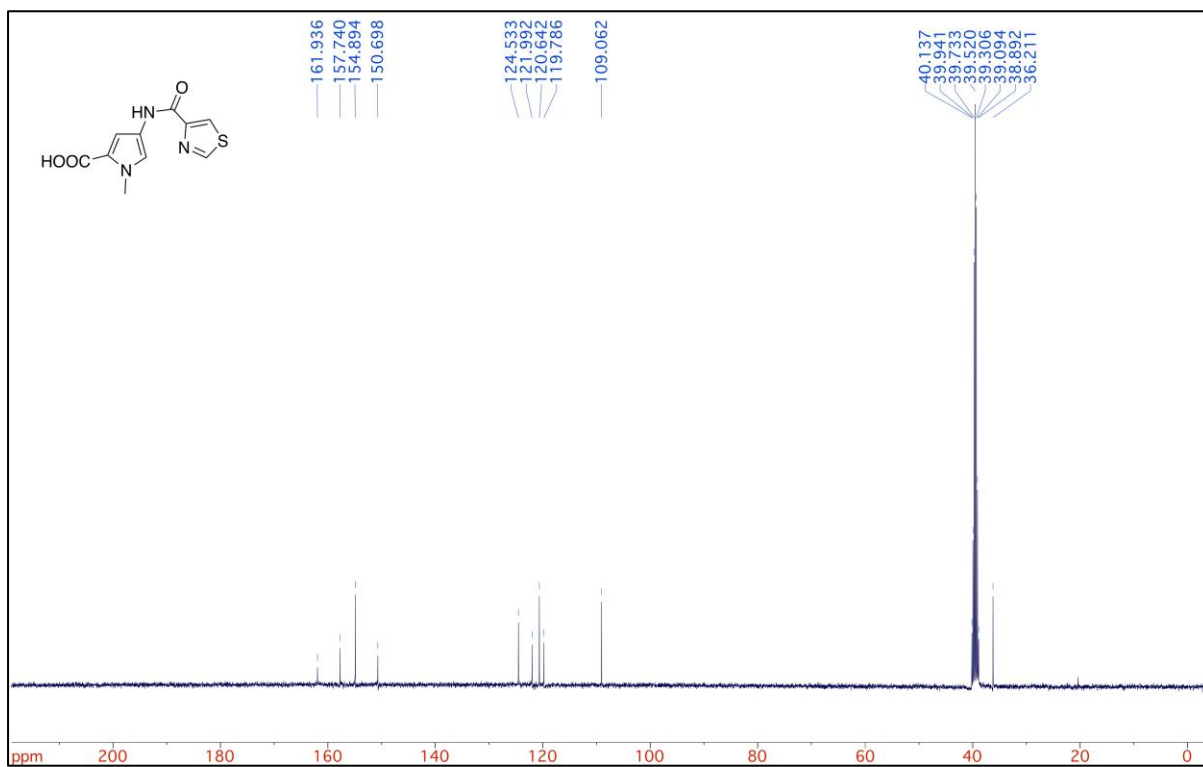


Appendix 40 HRMS of 2.115c.

1-methyl-4-(thiazole-4-carboxamido)-1H-pyrrole-2-carboxylic acid (2.88a)



Appendix 41 ¹H-NMR of 2.88a.



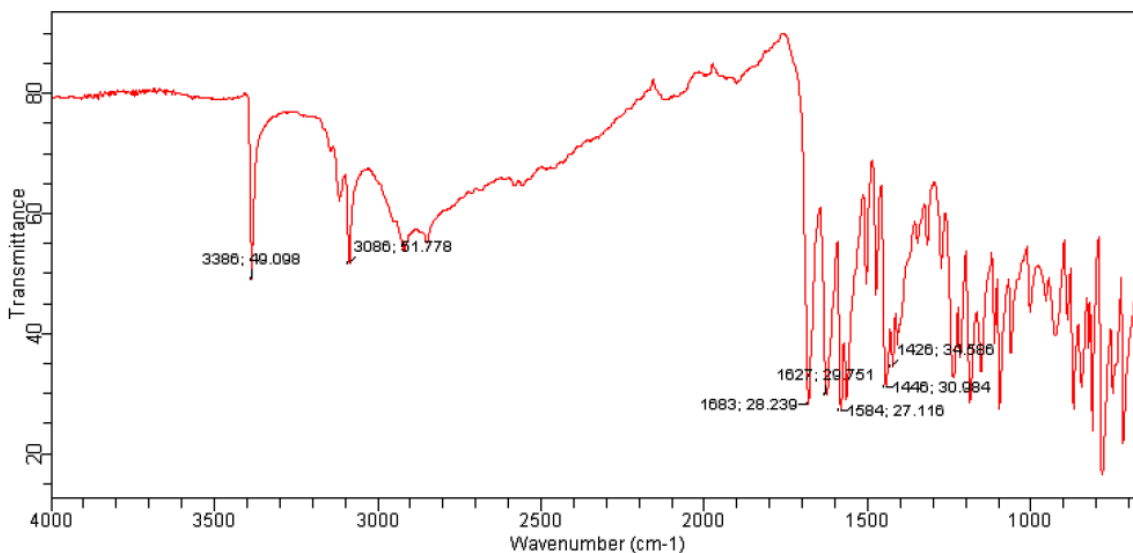
Appendix 42 ¹³C-NMR of 2.88a.



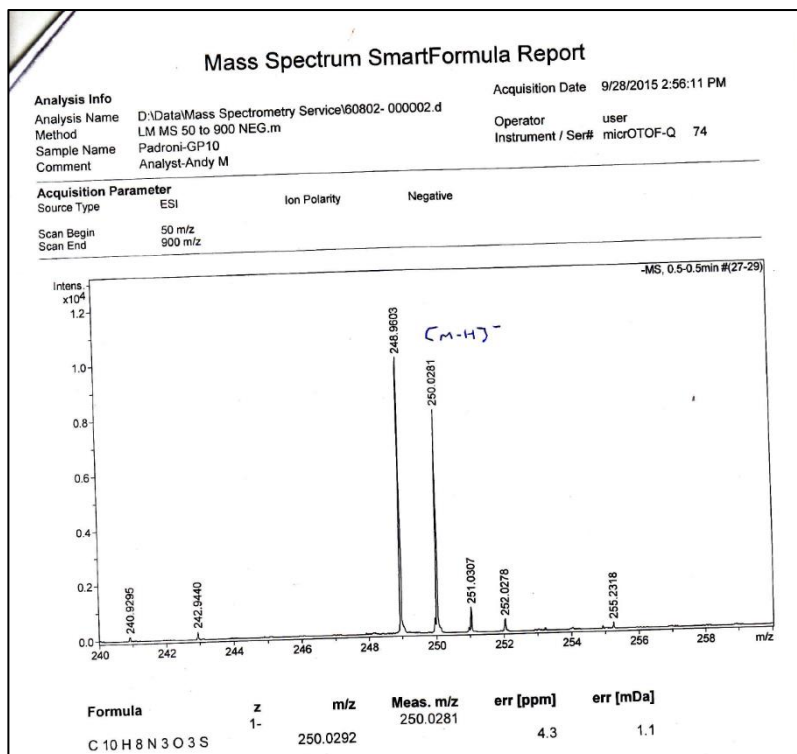
Agilent Technologies

Sample ID: ThHPyCOOH
Sample Scans: 32
Background Scans: 32
Resolution: 4 cm⁻¹
System Status: Good
File Location: C:\Documents and Settings\lan\Desktop\Giacomo\ThHPyCOOH_2015-08-27T08-56-43.a2r

Method Name: STUDENT ATR 32 4cm
User: STUDENT
Date/Time: 27/08/2015 08:55:04
Range: 4,000.00 - 650.00
Apodization: Happ-Genzel

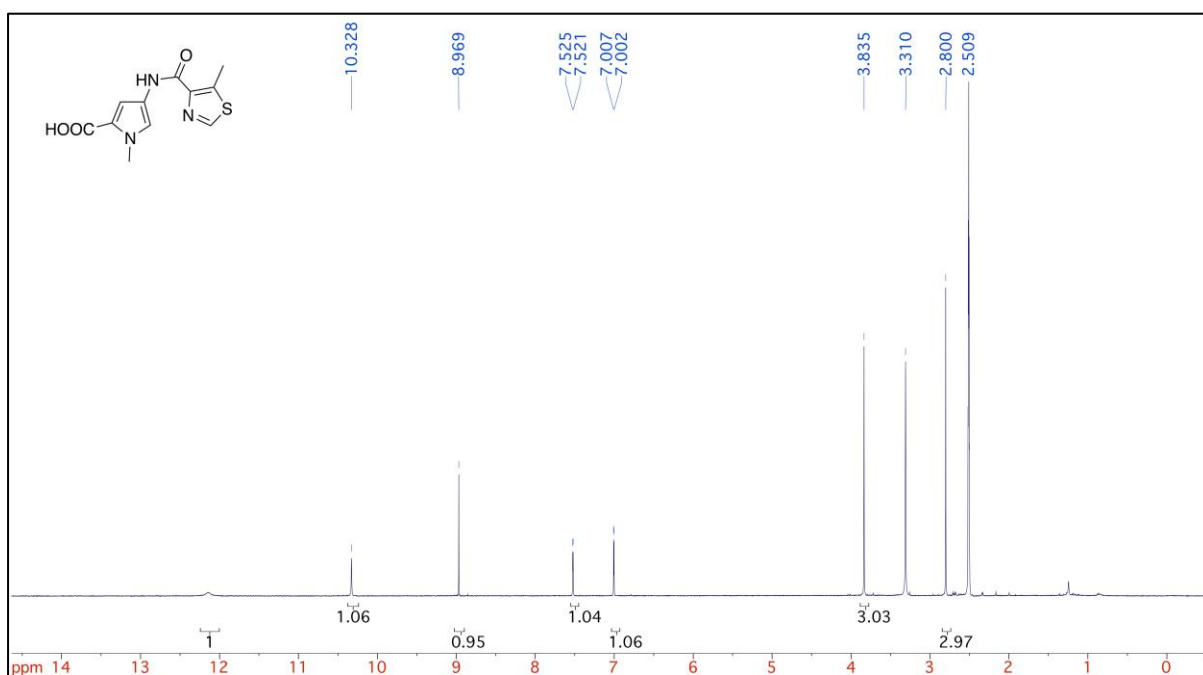


Appendix 43 IR of 2.88a.

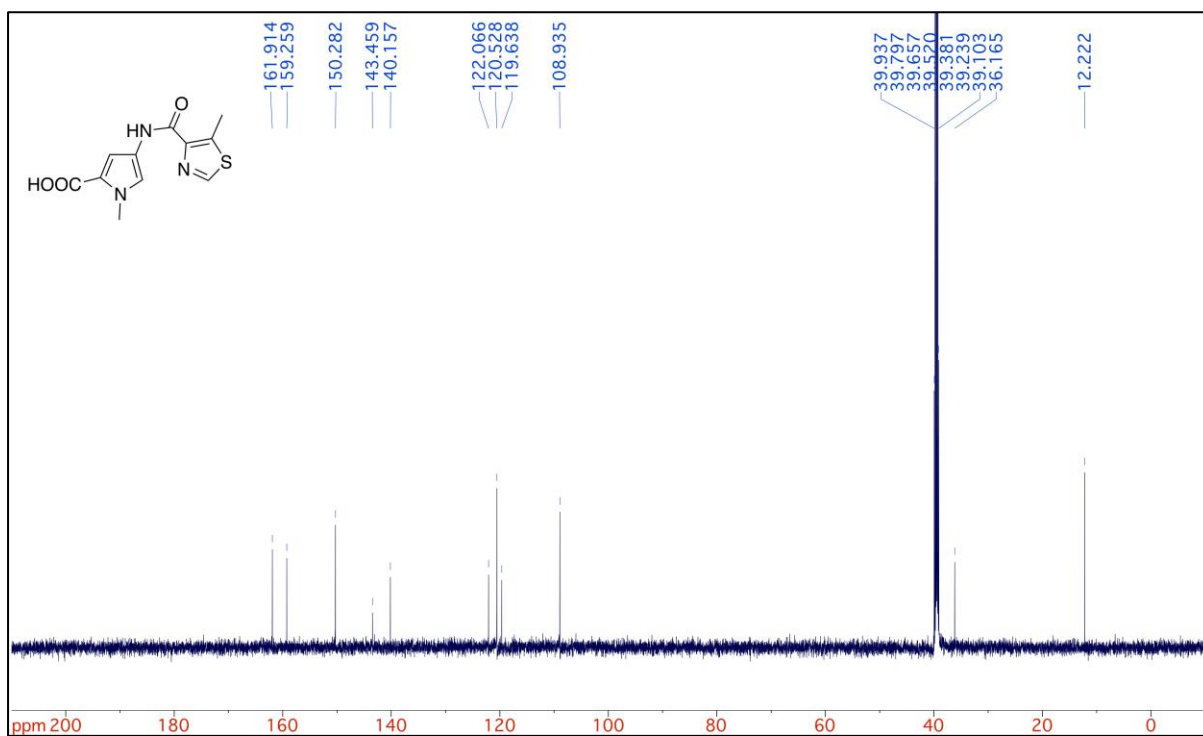


Appendix 44 HRMS of 2.88a.

1-methyl-4-(5-methylthiazole-4-carboxamido)-1H-pyrrole-2-carboxylic acid (2.88b)



Appendix 45 ¹H-NMR of 2.88b.

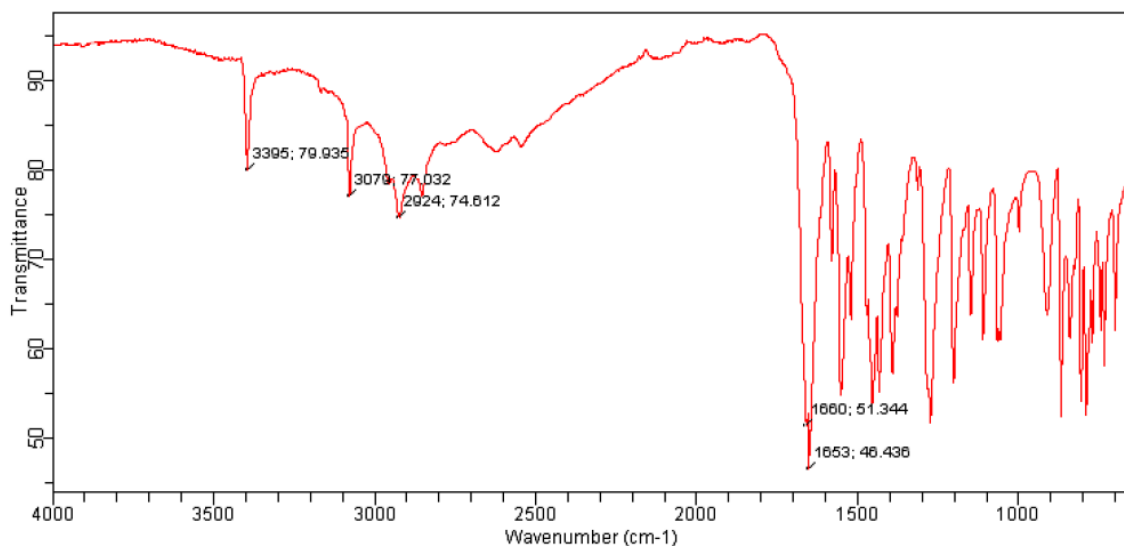


Appendix 46 ¹³C-NMR of 2.88b

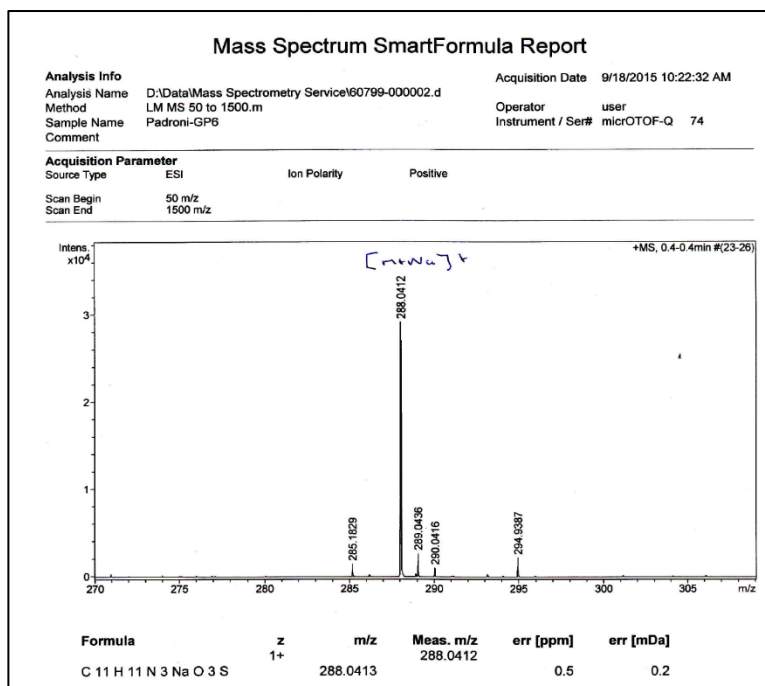


Agilent Technologies

Sample ID: ThCH3PyCOOH Method Name: STUDENT ATR 32 4cm
Sample Scans: 32 User: STUDENT
Background Scans: 32 Date/Time: 27/08/2015 09:27:28
Resolution: 4 cm⁻¹ Range: 4,000.00 - 650.00
System Status: Good Apodization: Happ-Genzel
File Location: C:\Program Files\Agilent\MicroLab PC\Results\STUDENT ATR 32
4cm\ThCH3PyCOOH_2015-08-27T09-28-31.a2r

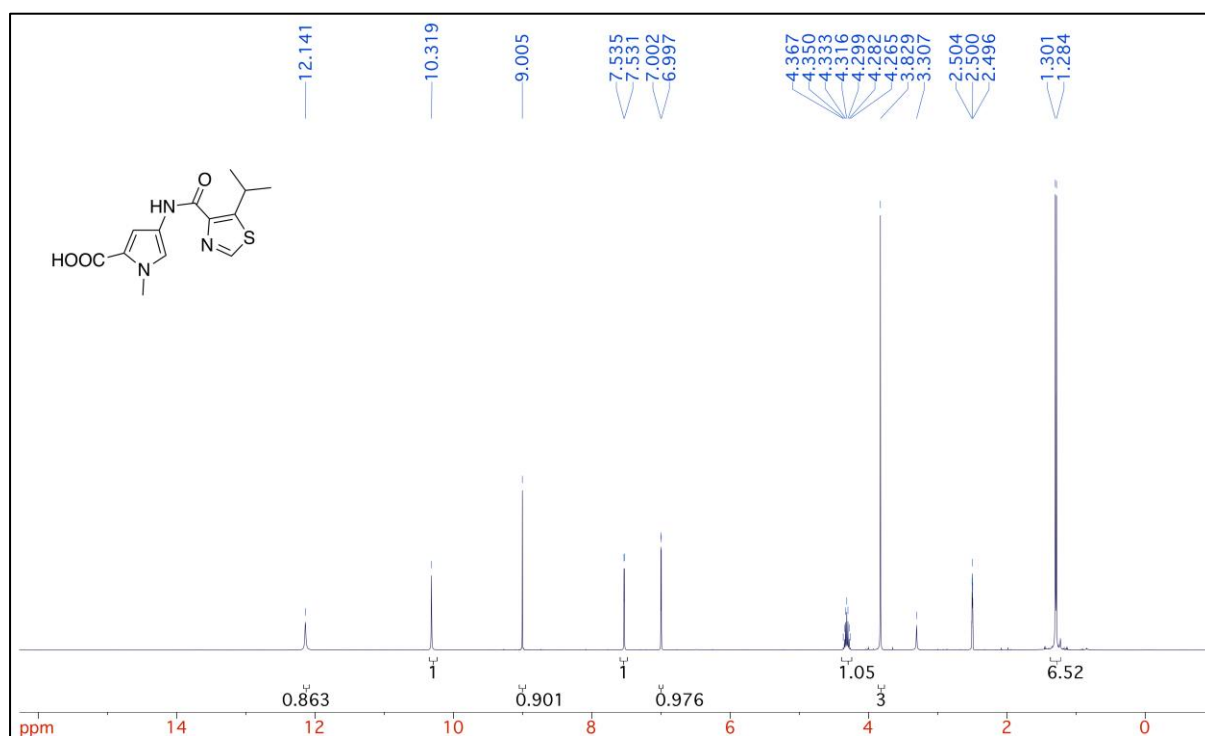


Appendix 47 IR of 2.88b.

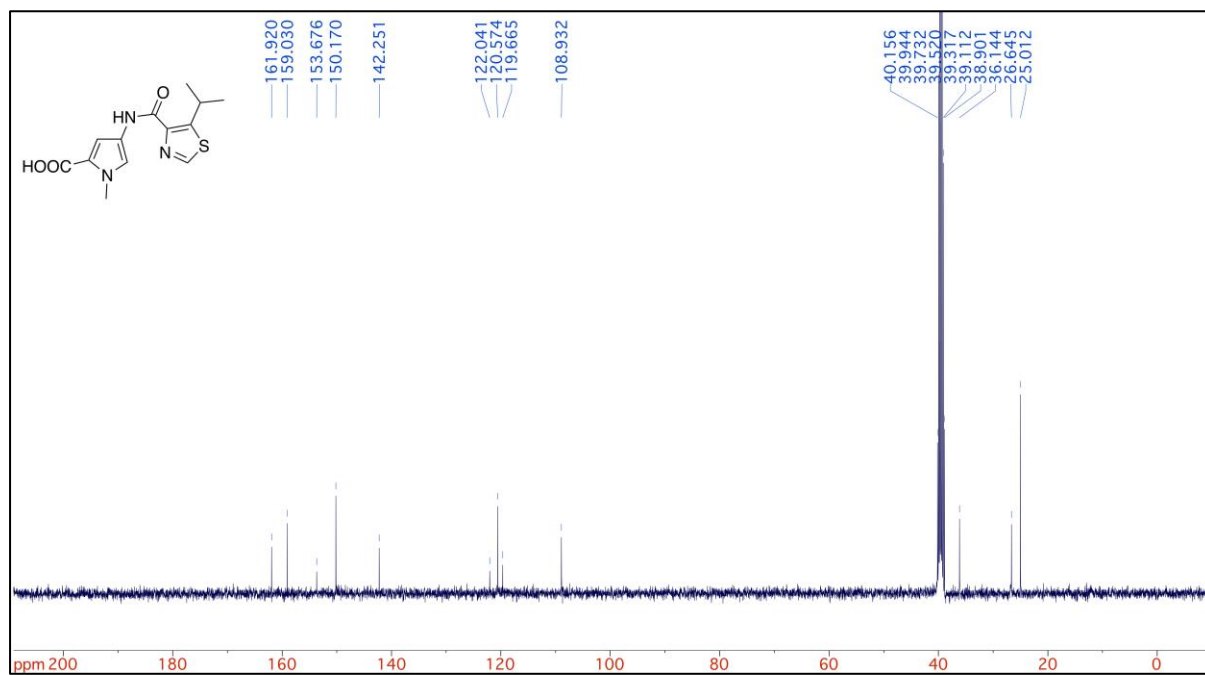


Appendix 48 HRMS of 2.88b.

4-(5-isopropylthiazole-4-carboxamido)-1-methyl-1H-pyrrole-2-carboxylic acid (**2.88c**)



Appendix 49 ¹H-NMR of **2.88c**.

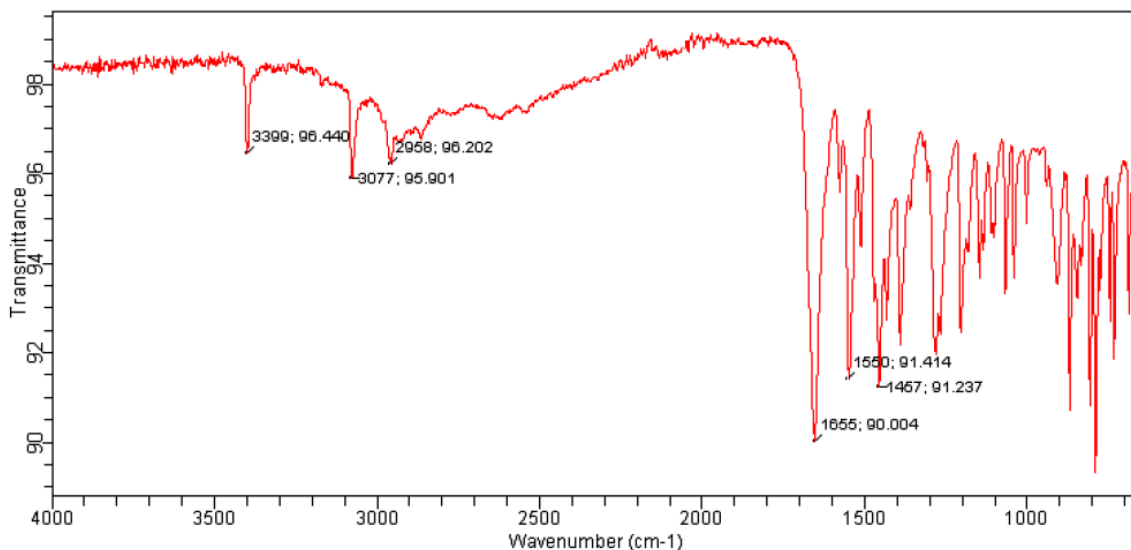


Appendix 50 ¹³C-NMR of **2.88c**.

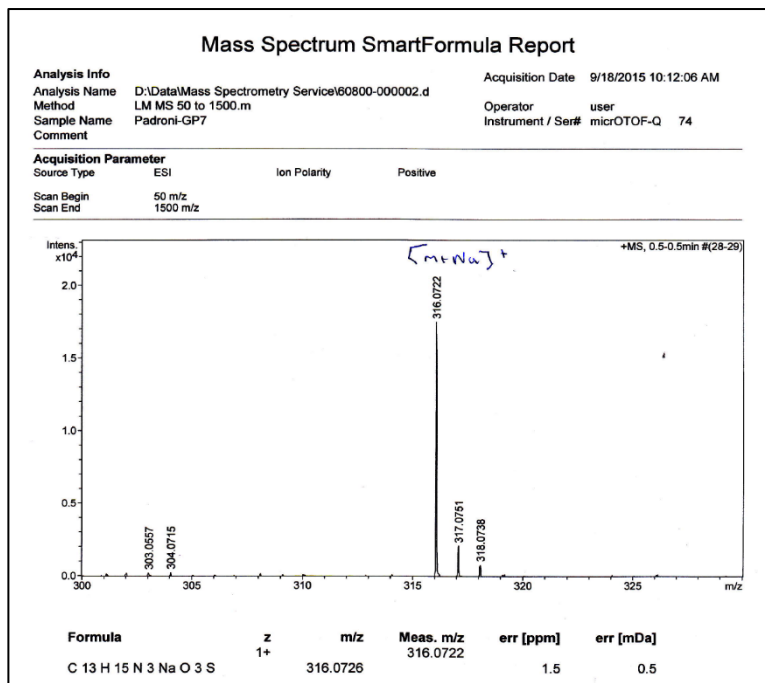


Agilent Technologies

Sample ID: ThiPrPyCOOH Method Name: STUDENT ATR 32 4cm
Sample Scans: 32 User: STUDENT
Background Scans: 32 Date/Time: 27/08/2015 09:29:53
Resolution: 4 cm⁻¹ Range: 4,000.00 - 650.00
System Status: Good Apodization: Happ-Genzel
File Location: C:\Program Files\Agilent\MicroLab PC\Results\STUDENT ATR 32
4cm\ThiPrPyCOOH_2015-08-27T09-30-38.a2r

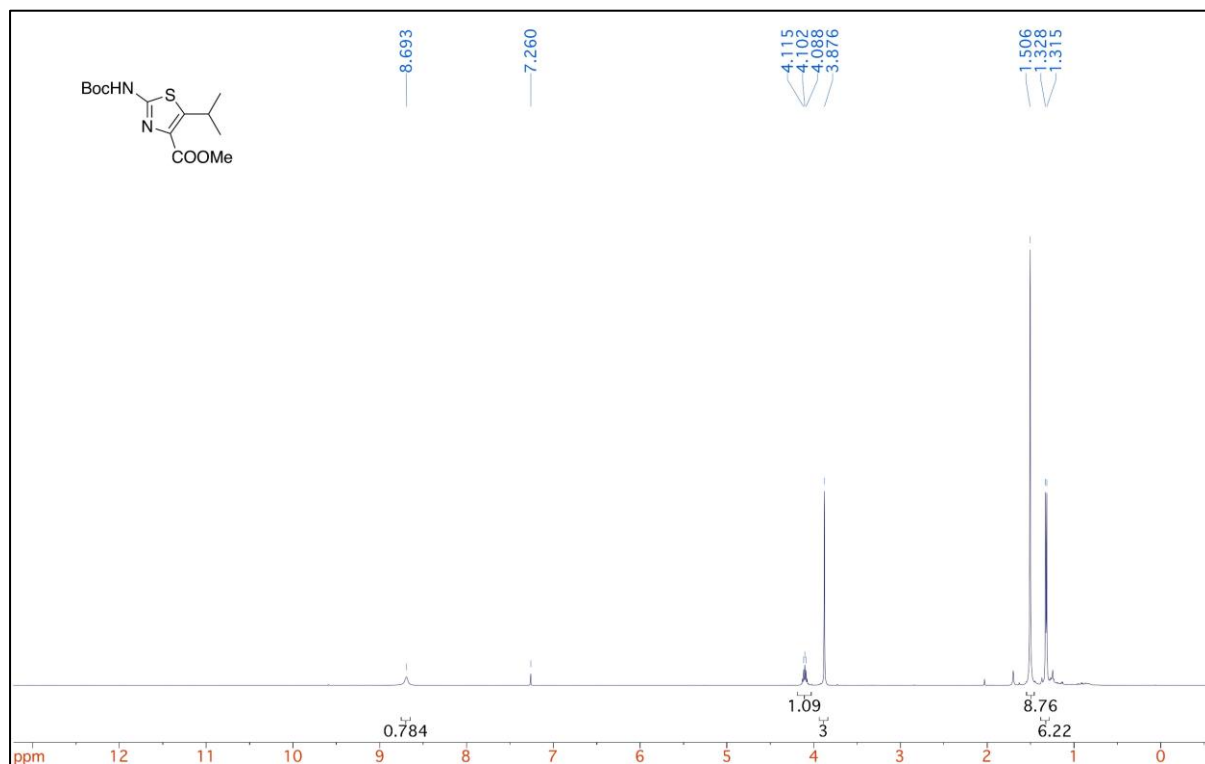


Appendix 51 IR of 2.88c.

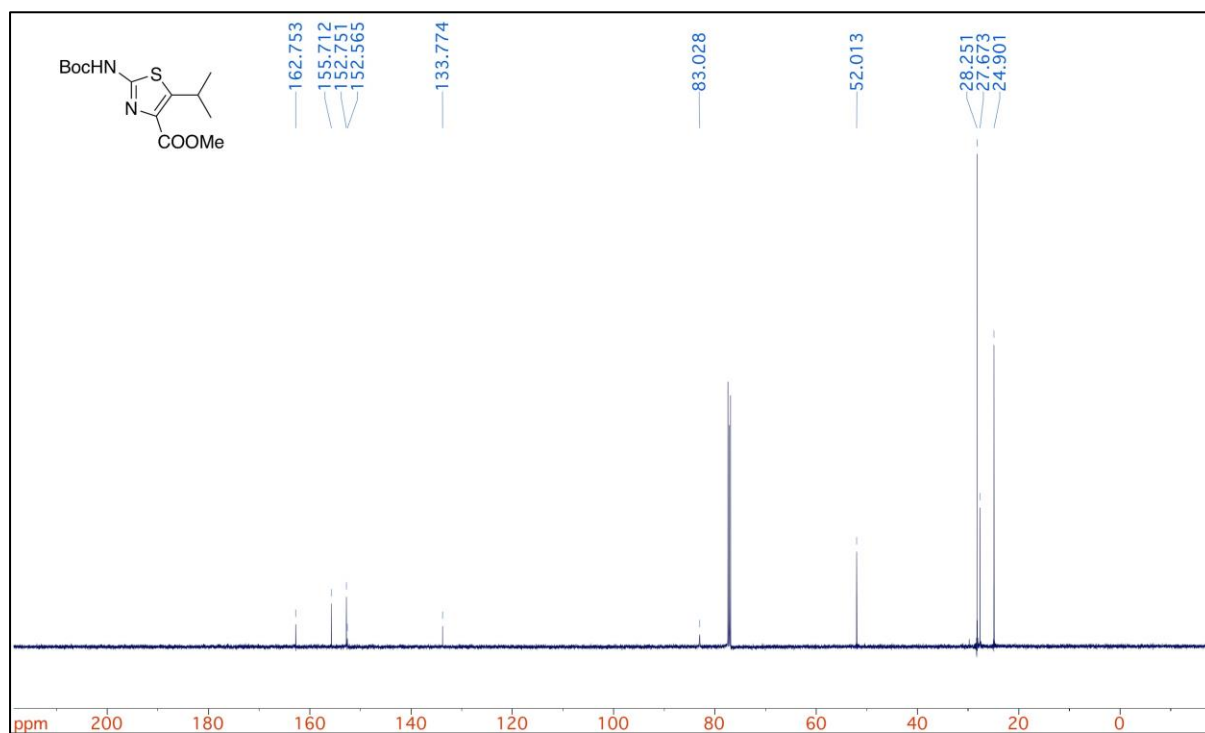


Appendix 52 HRMS of 2.88c.

Methyl 2-((tert-butoxycarbonyl)amino)-5-isopropylthiazole-4-carboxylate (**2.123**)



Appendix 53 ¹H-NMR of **2.123**.



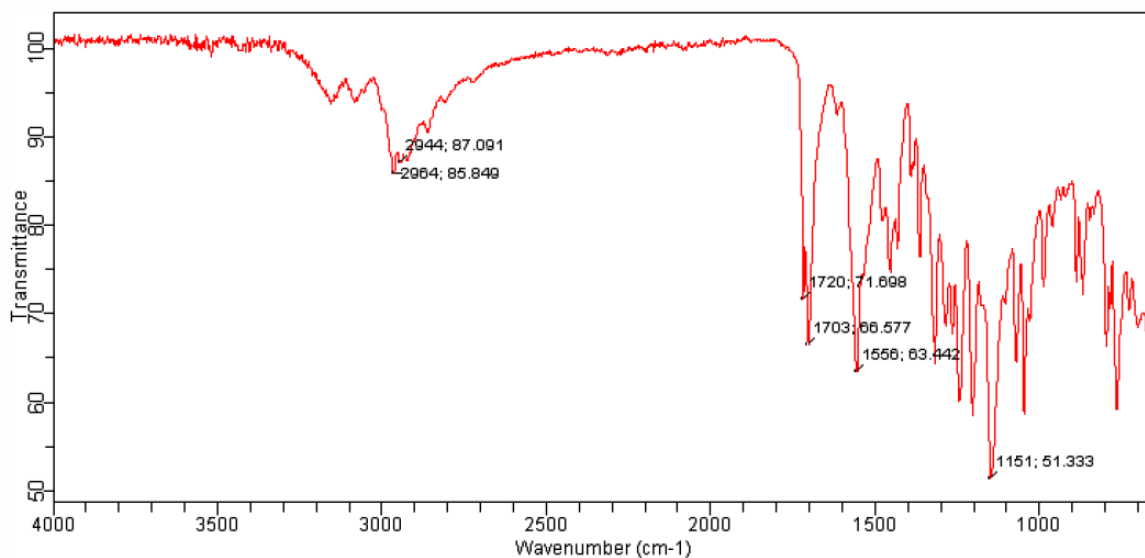
Appendix 54 ¹³C-NMR of **2.123**.



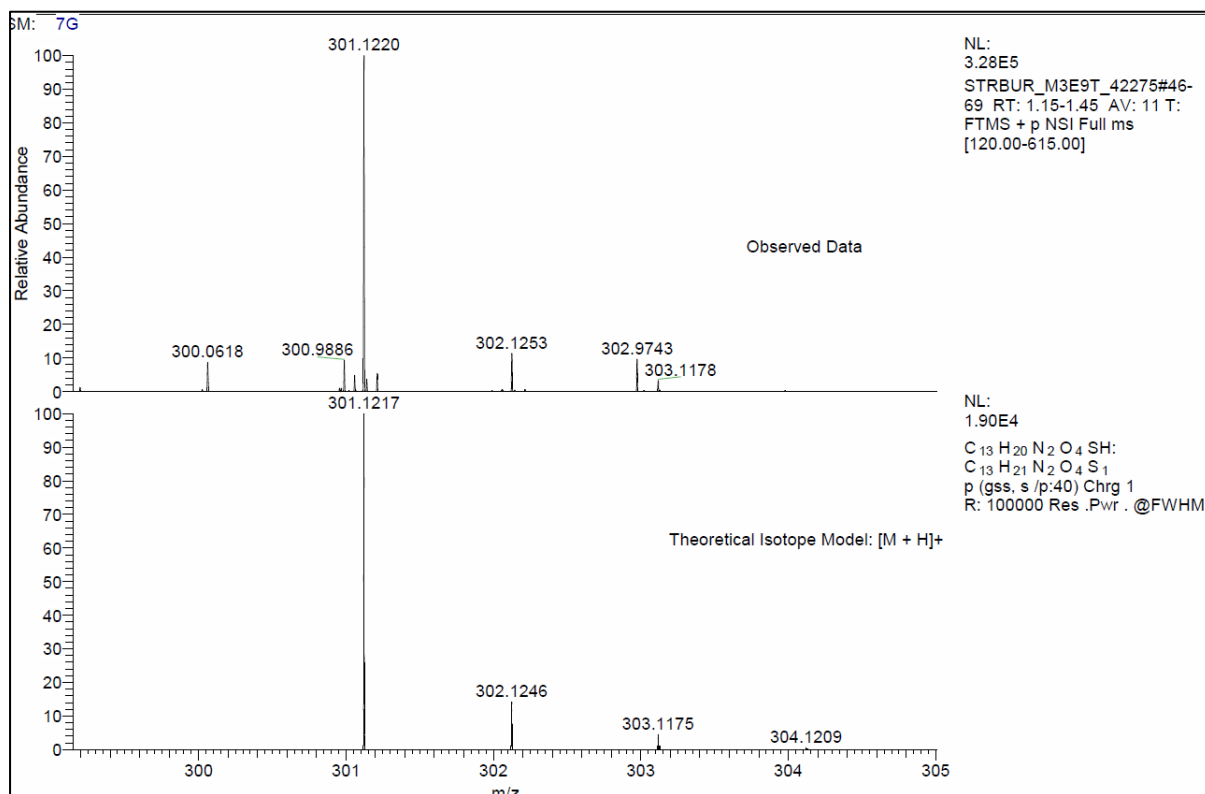
Agilent Technologies

Sample ID: BocNtiPrCOOMe
Sample Scans: 32
Background Scans: 32
Resolution: 4 cm⁻¹
System Status: Good
File Location: C:\Program Files\Agilent\MicroLab PC\Results\STUDENT ATR 32 4cm\BocNtiPrCOOMe_2018-02-15T10-47-39.a2r

Method Name: STUDENT ATR 32 4cm
User: STUDENT
Date/Time: 15/02/2018 10:46:34
Range: 4,000.00 - 650.00
Apodization: Happ-Genzel

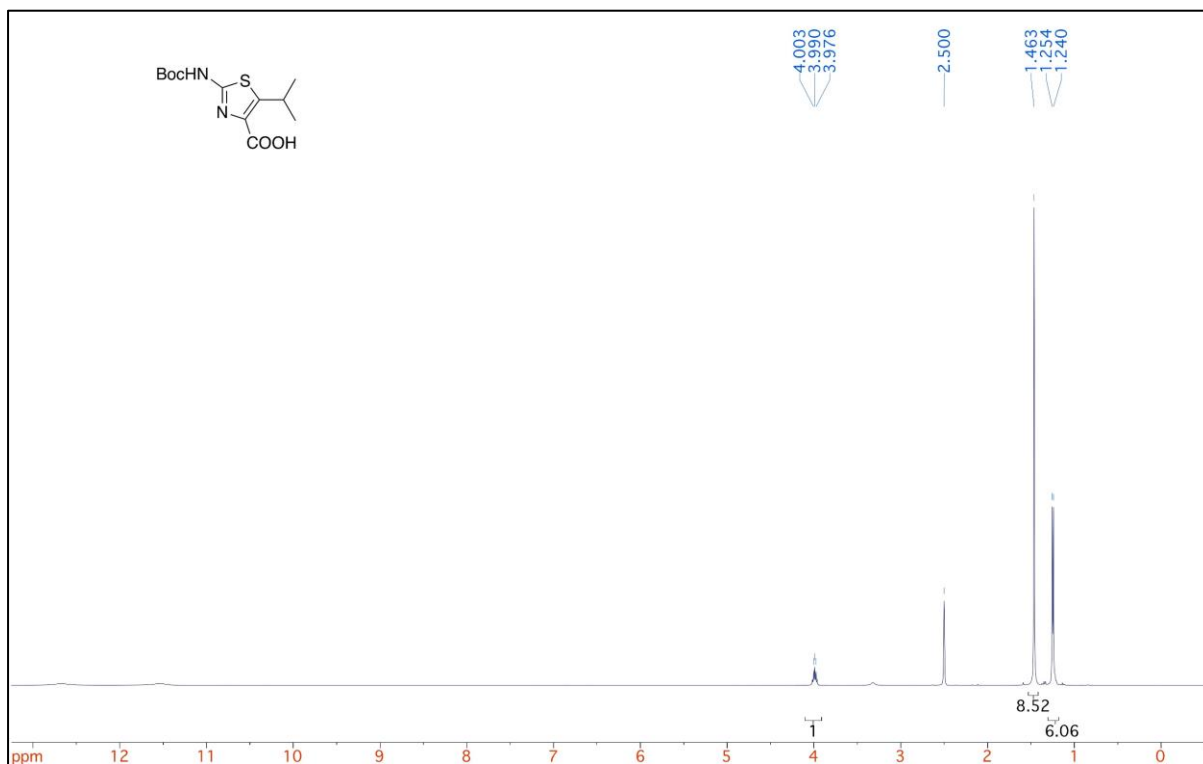


Appendix 55 IR of 2.123.

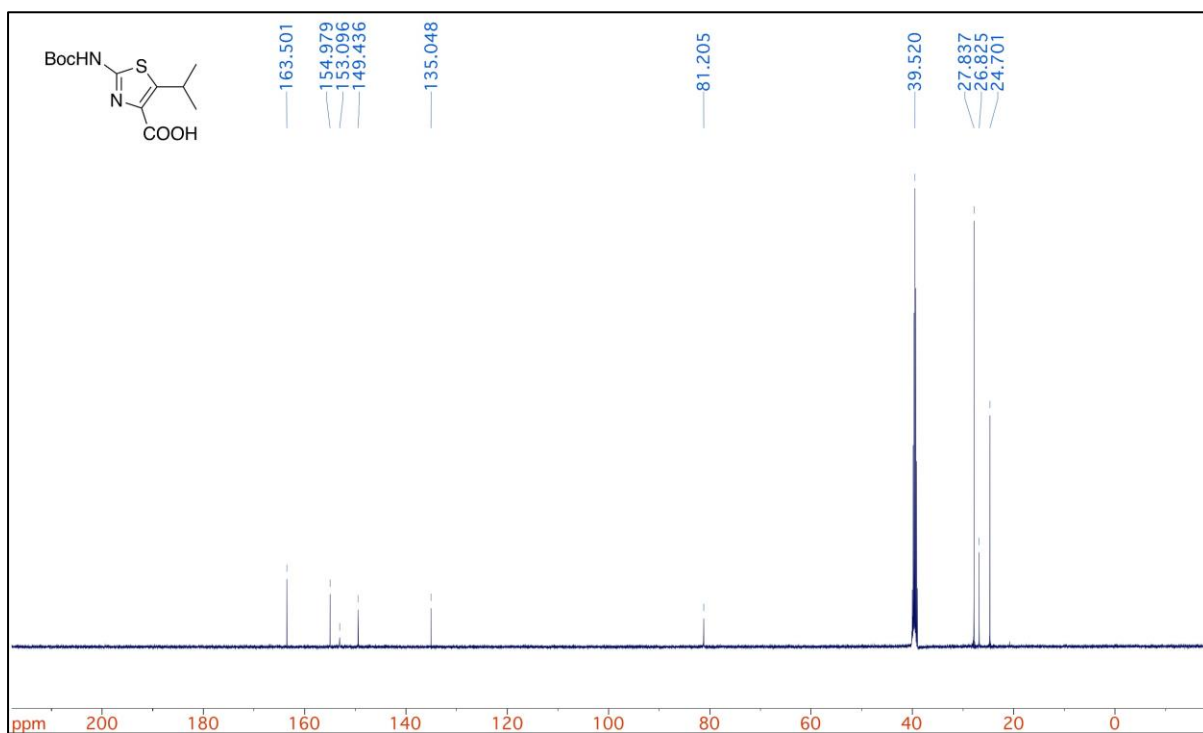


Appendix 56 HRMS of 2.123.

2-((tert-butoxycarbonyl)amino)-5-isopropylthiazole-4-carboxylic acid (**2.89**)



Appendix 57 ¹H-NMR of **2.89**.



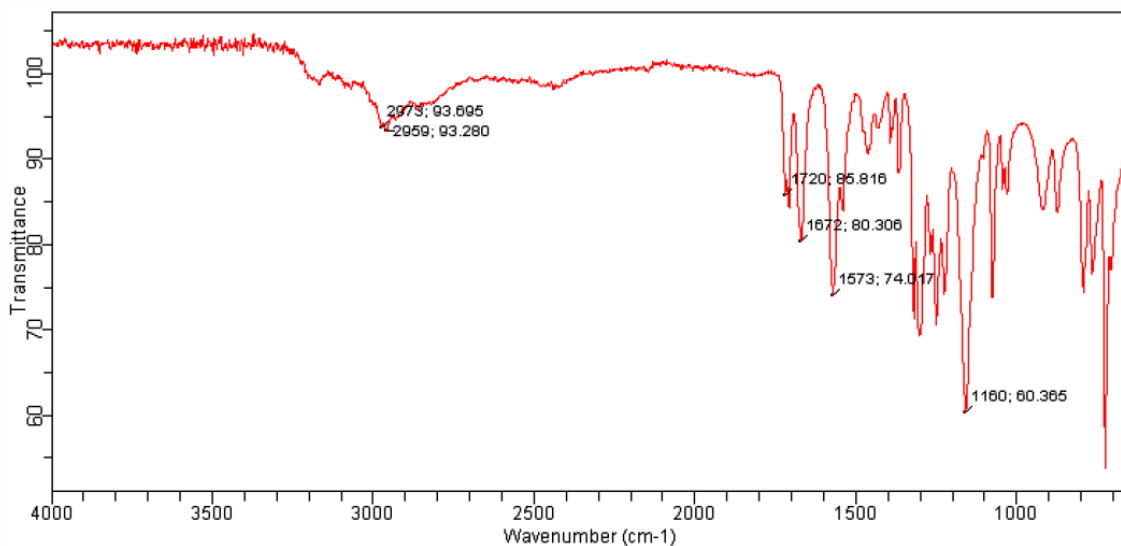
Appendix 58 ¹³C-NMR of **2.89**.



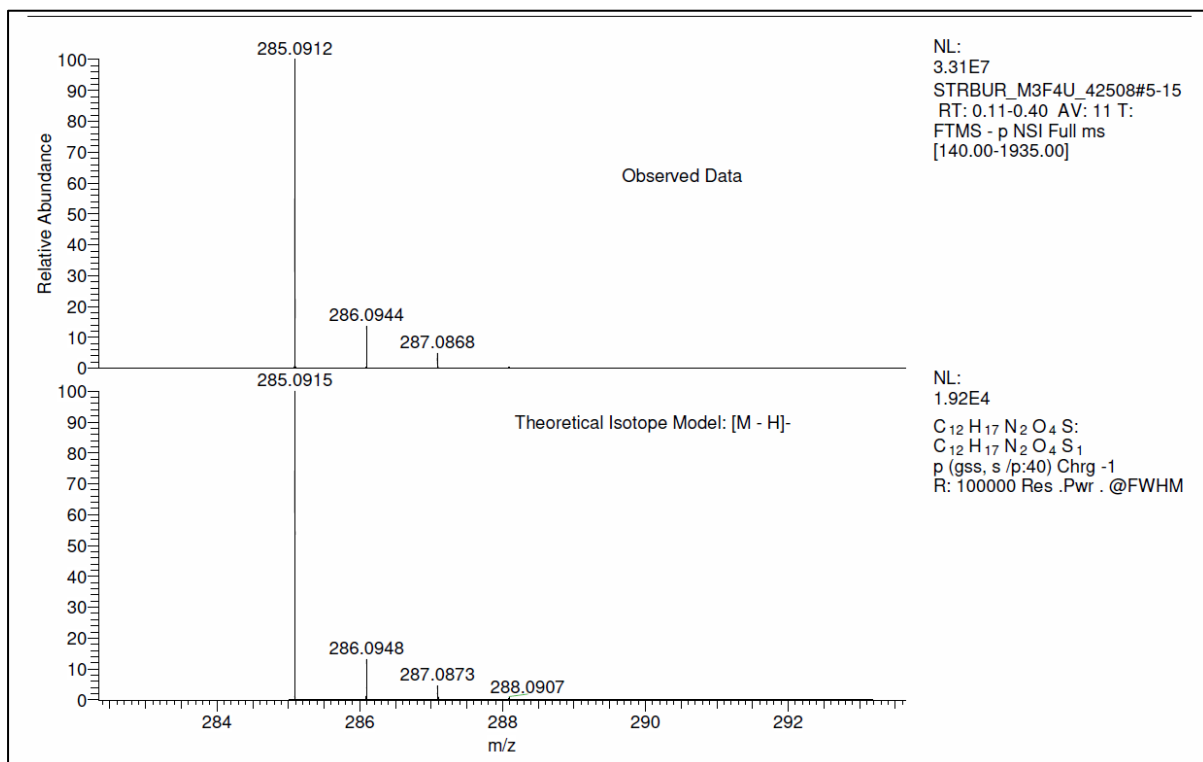
Agilent Technologies

Sample ID: BocNtiPrCOOH
Sample Scans: 32
Background Scans: 32
Resolution: 4 cm⁻¹
System Status: Good
File Location: C:\Program Files\Agilent\MicroLab PC\Results\STUDENT ATR 32 4cm\BocNtiPrCOOH_2018-02-15T10-50-20.a2r

Method Name: STUDENT ATR 32 4cm
User: STUDENT
Date/Time: 15/02/2018 10:49:25
Range: 4,000.00 - 650.00
Apodization: Happ-Genzel

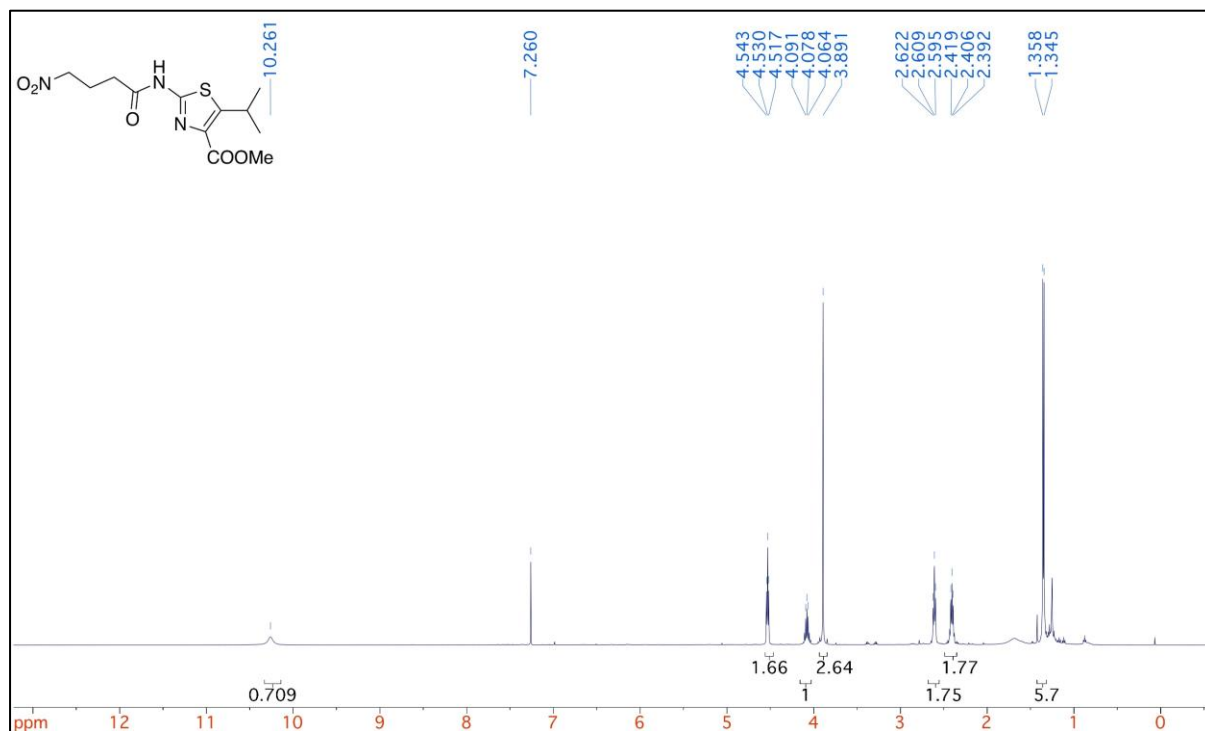


Appendix 59 IR of 2.89.

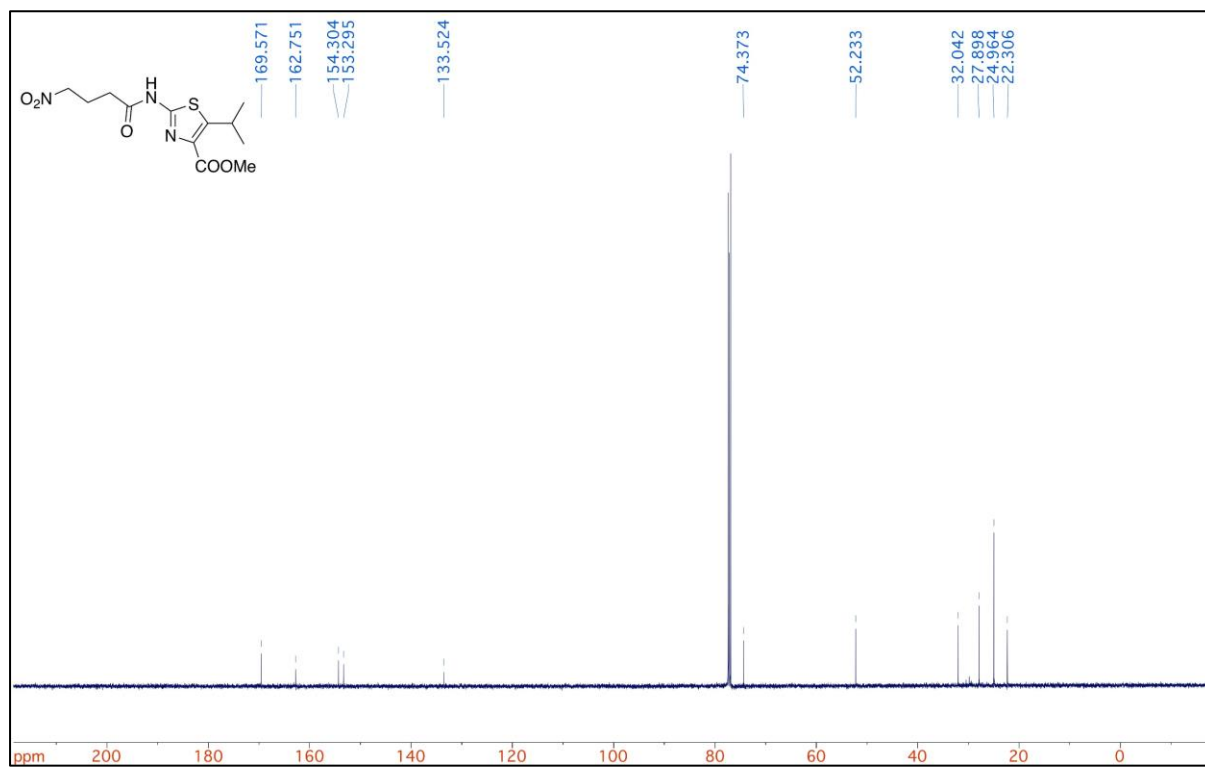


Appendix 60 HRMS of 2.89.

Methyl 5-isopropyl-2-(4-nitrobutanamido)thiazole-4-carboxylate (**2.125**)



Appendix 61 ¹H-NMR of **2.125**.

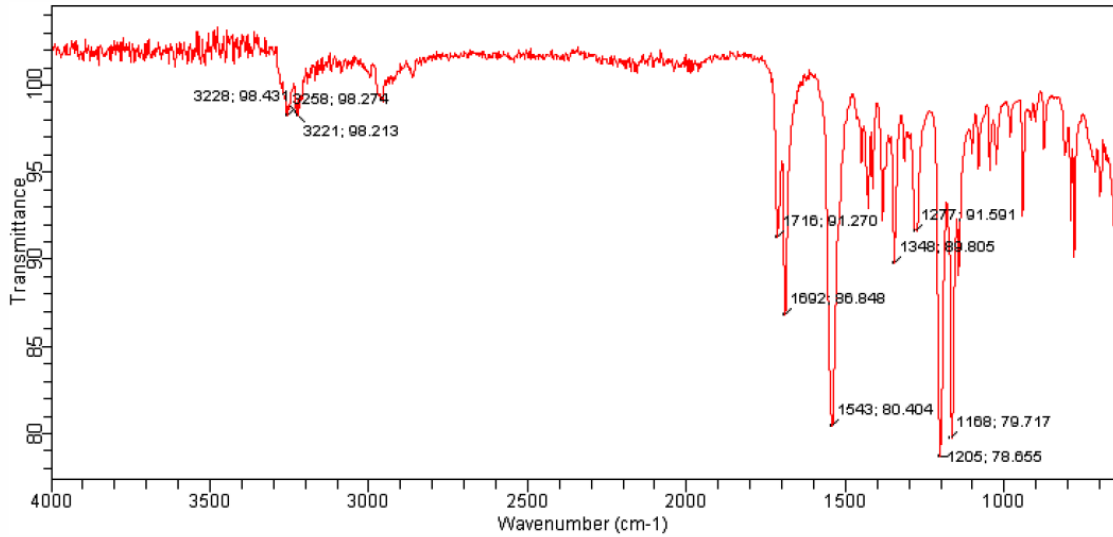


Appendix 62 ¹³C-NMR of **2.125**.

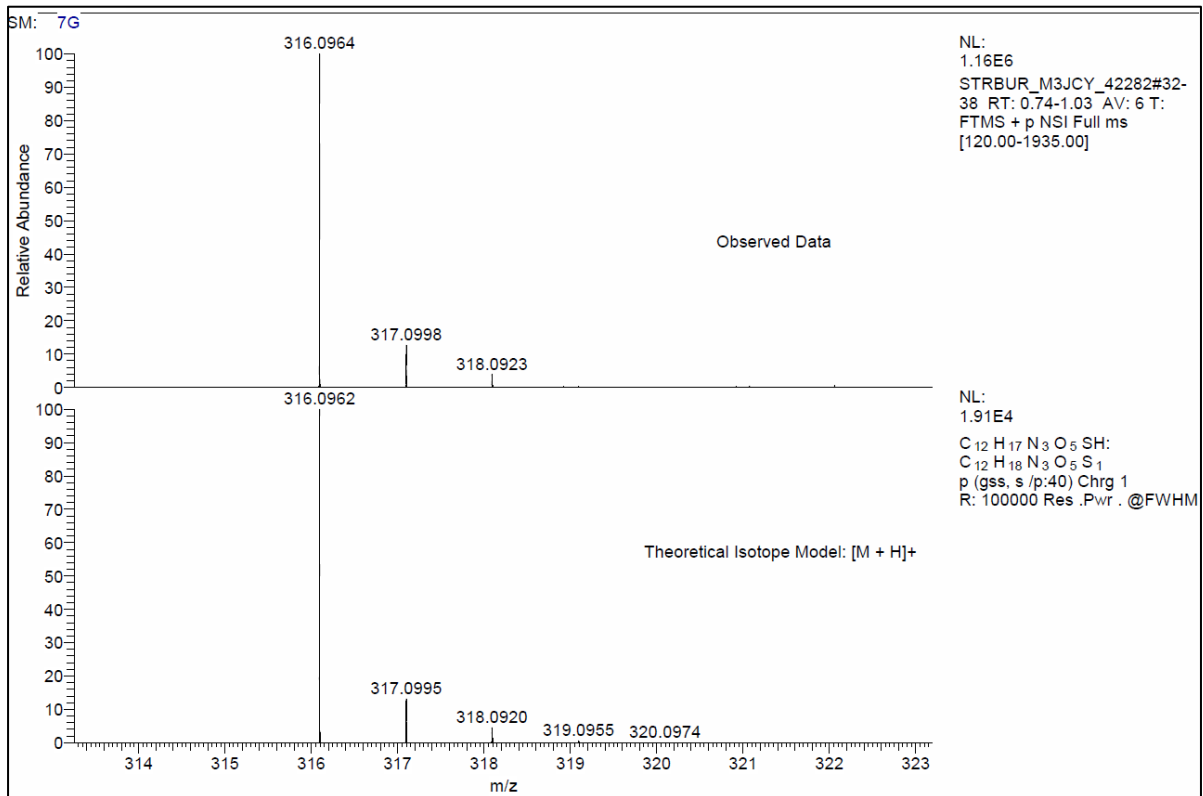


Agilent Technologies

Sample ID: NO2GABANTiPrCOOMe Method Name: STUDENT ATR 32 4cm
Sample Scans: 32 User: STUDENT
Background Scans: 32 Date/Time: 15/02/2018 10:51:47
Resolution: 4 cm-1 Range: 4,000.00 - 650.00
System Status: Good Apodization: Happ-Genzel
File Location: C:\Program Files\Agilent\MicroLab PC\Results\STUDENT ATR 32
4cm\NO2GABANTiPrCOOMe_2018-02-15T10-52-58.a2r

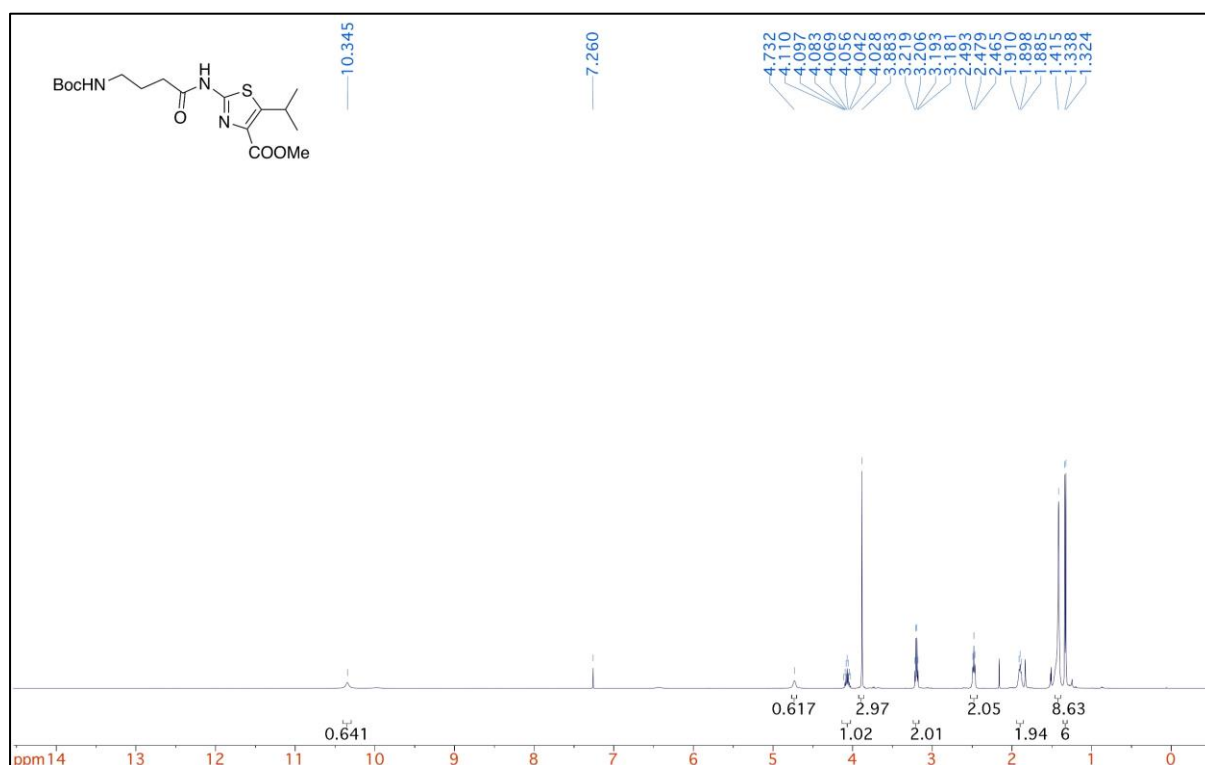


Appendix 63 IR of 2.125.

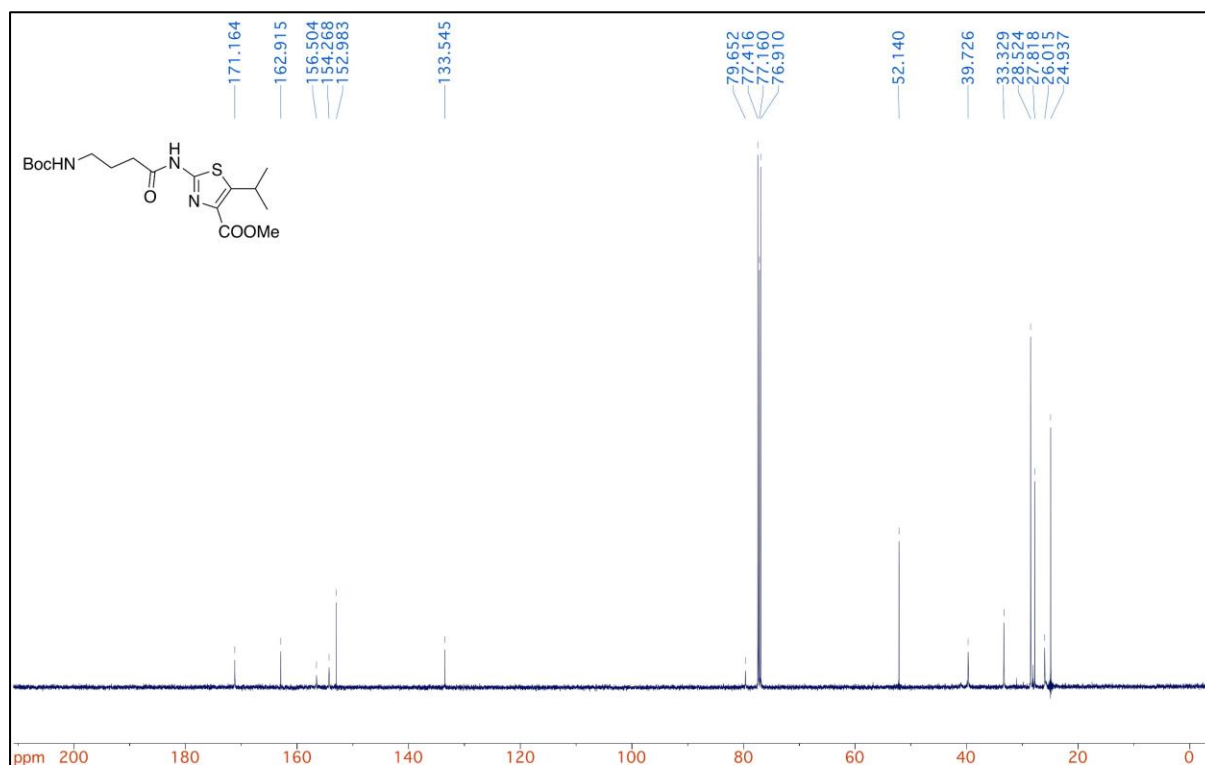


Appendix 64 HRMS of 2.125.

Methyl 2-(4-((tert-butoxycarbonyl)amino)butanamido)-5-isopropylthiazole-4-carboxylate (**2.126**)



Appendix 65 $^1\text{H-NMR}$ of **2.126**.

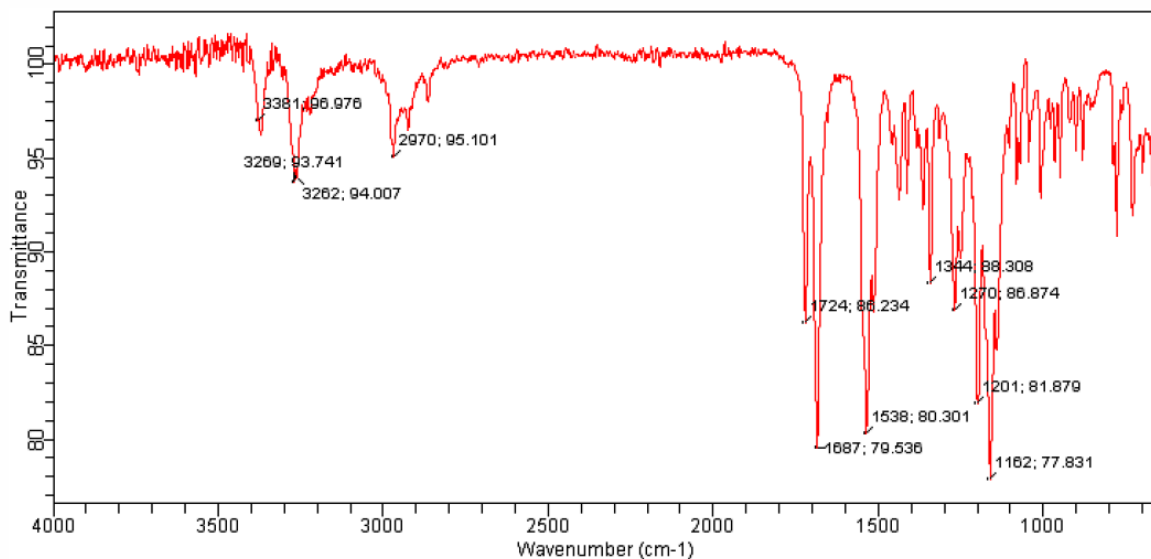


Appendix 66 $^{13}\text{C-NMR}$ of **2.126**.

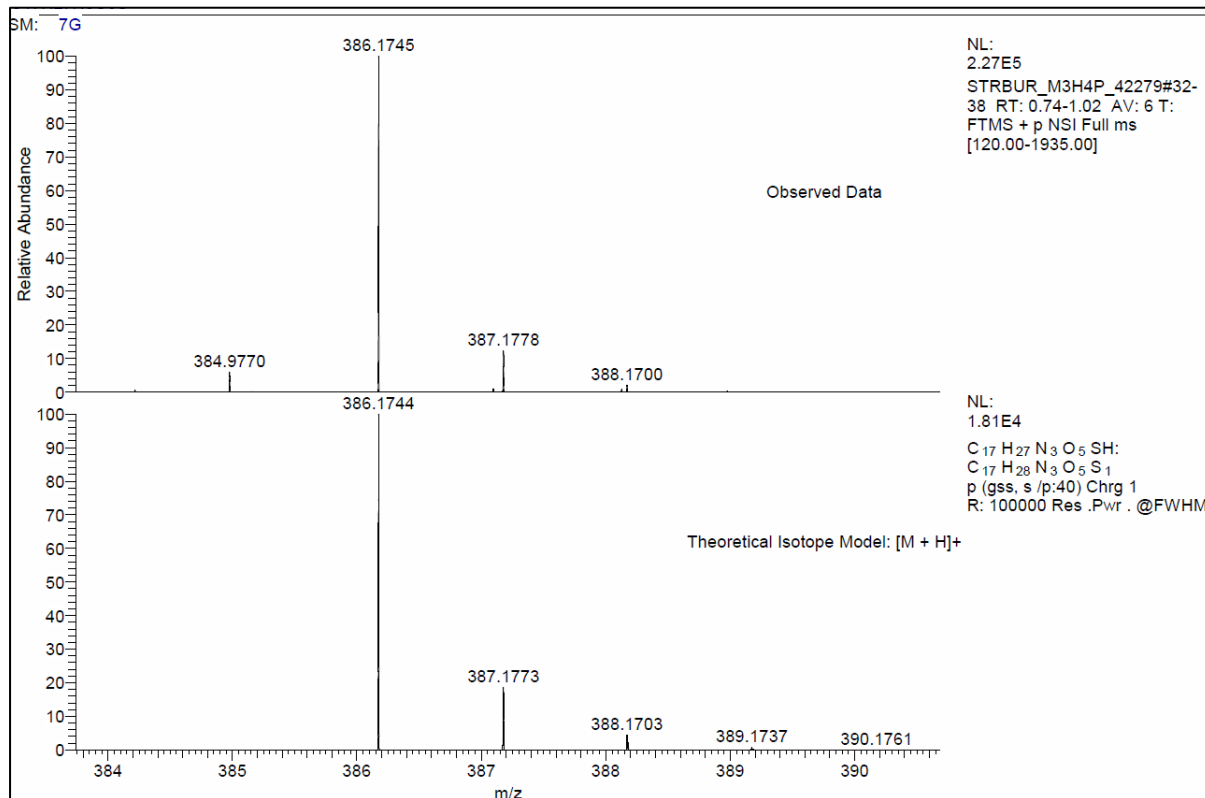


Agilent Technologies

Sample ID: BocNHGABANTiPrCOOMe Method Name: STUDENT ATR 32 4cm
Sample Scans: 32 User: STUDENT
Background Scans: 32 Date/Time: 15/02/2018 10:54:15
Resolution: 4 cm⁻¹ Range: 4,000.00 - 650.00
System Status: Good Apodization: Happ-Genzel
File Location: C:\Program Files\Agilent\MicroLab PC\Results\STUDENT ATR 32
4cm\BocNHGABANTiPrCOOMe_2018-02-15T10-55-41.a2r

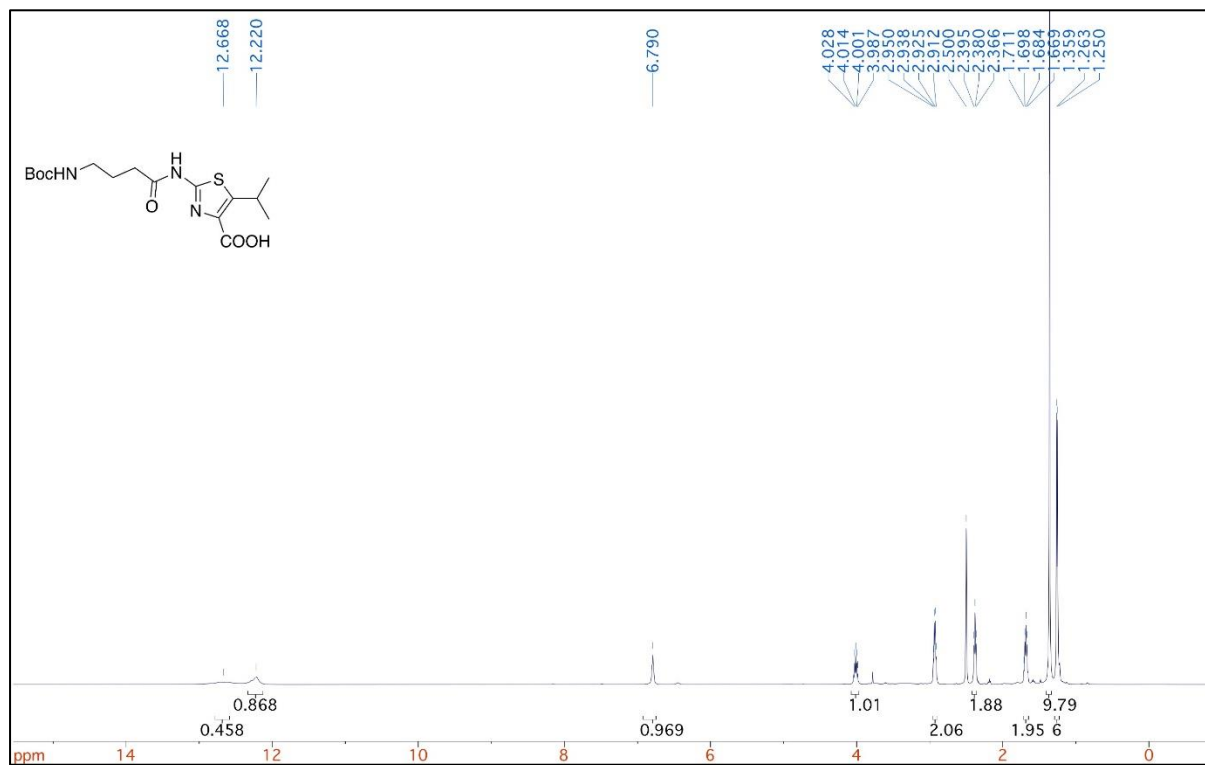


Appendix 67 IR of 2.126.

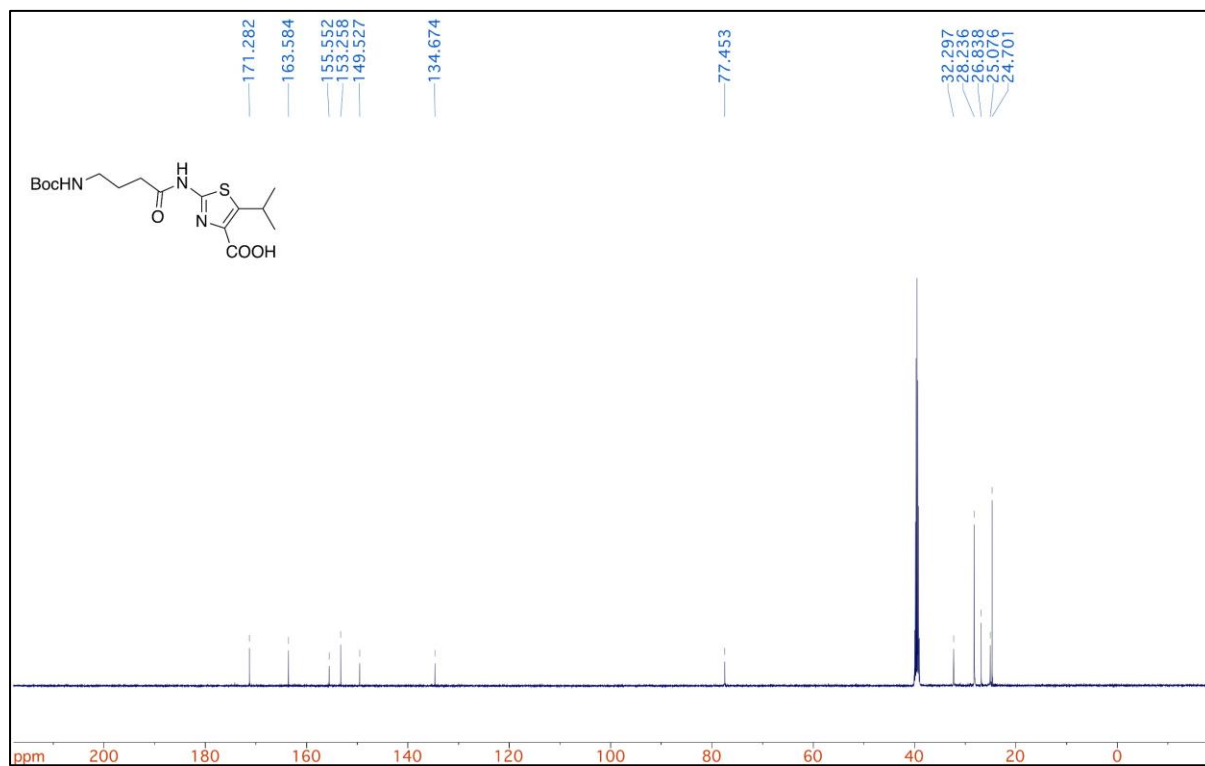


Appendix 68 HRMS of 2.126.

2-(4-((*tert*-butoxycarbonyl)amino)butanamido)-5-isopropylthiazole-4-carboxylic acid (**2.90**)



Appendix 69 ¹H-NMR of **2.90**.

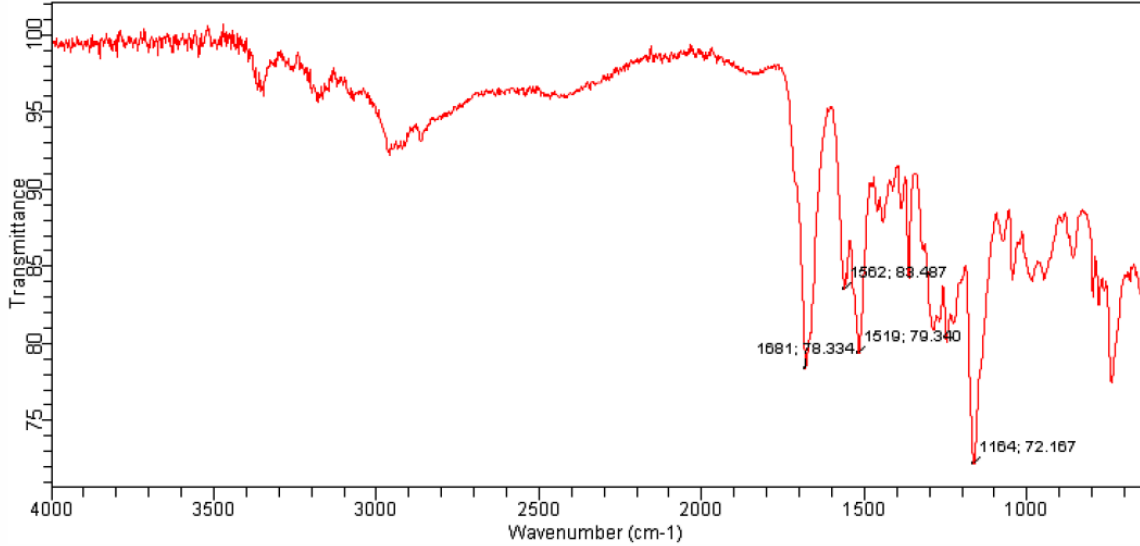


Appendix 70 ¹³C-NMR of **2.90**.

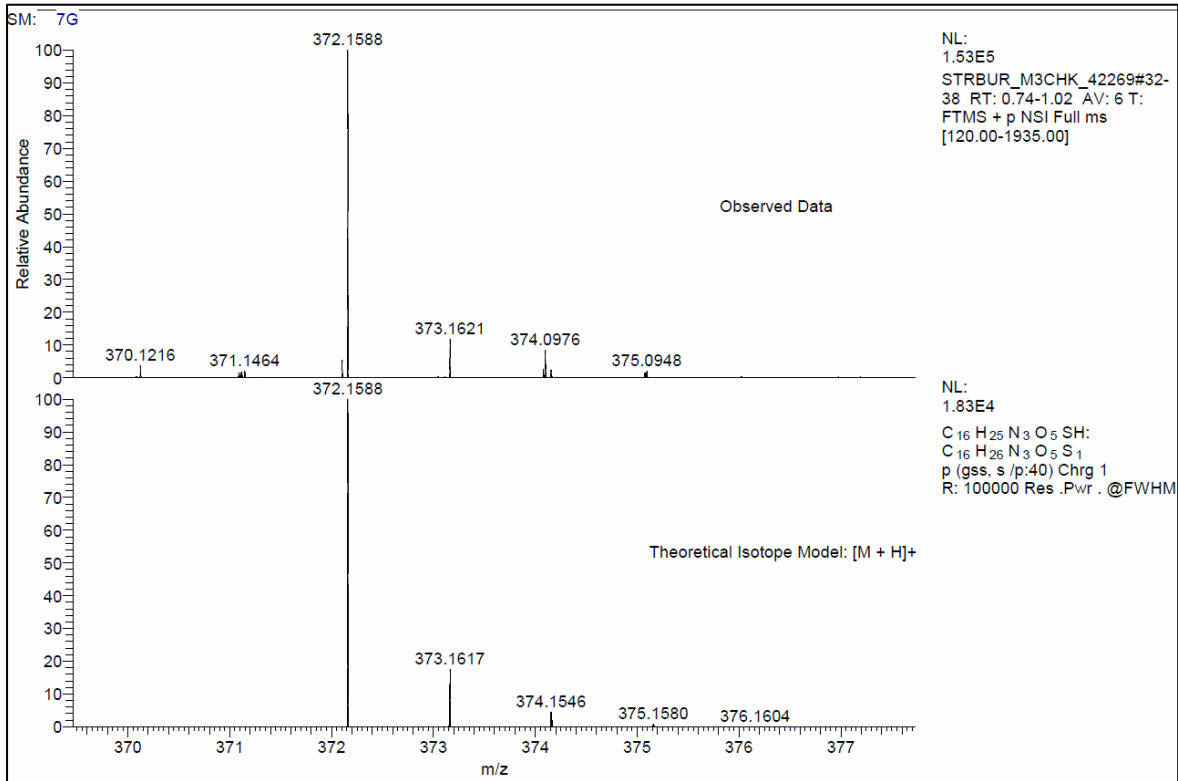


Agilent Technologies

Sample ID: BocNHGABANTiPrCOOH Method Name: STUDENT ATR 32 4cm
Sample Scans: 32 User: STUDENT
Background Scans: 32 Date/Time: 15/02/2018 10:56:50
Resolution: 4 cm⁻¹ Range: 4,000.00 - 650.00
System Status: Good Apodization: Happ-Genzel
File Location: C:\Program Files\Agilent\MicroLab PC\Results\STUDENT ATR 32
4cm\BocNHGABANTiPrCOOH_2018-02-15T10-58-08.a2r

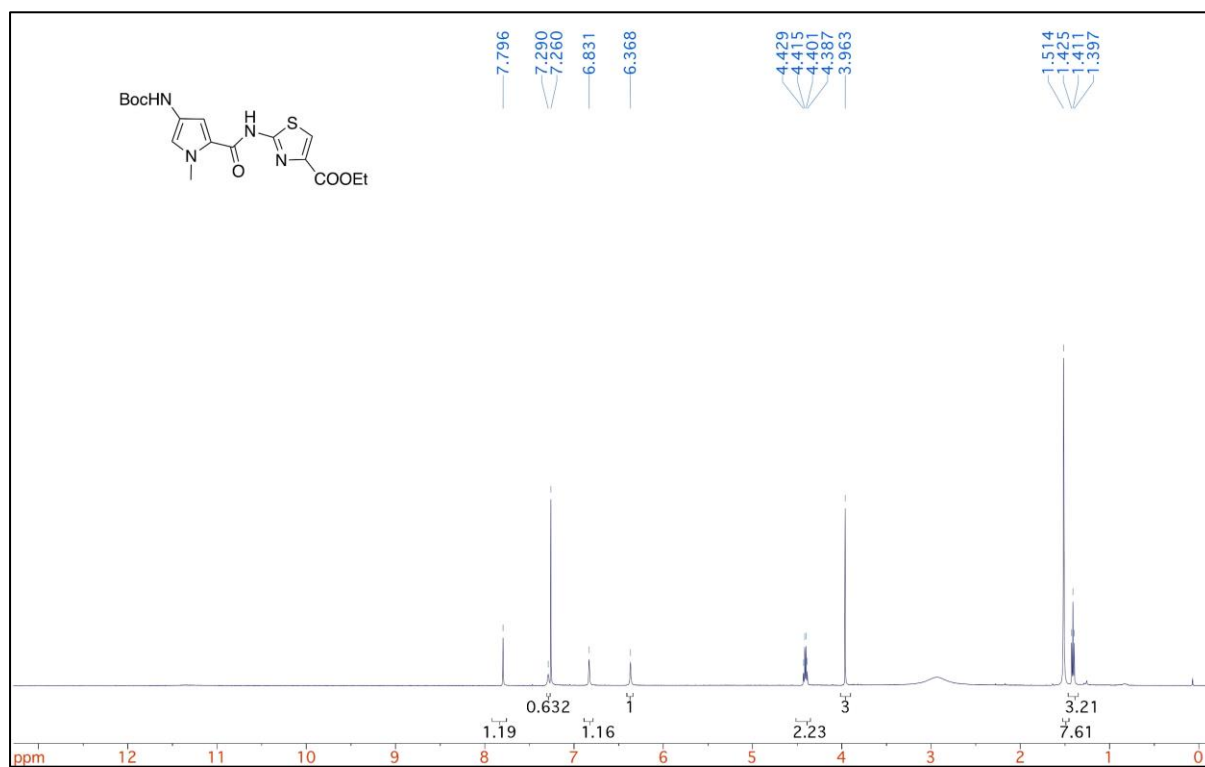


Appendix 71 IR of 2.90.

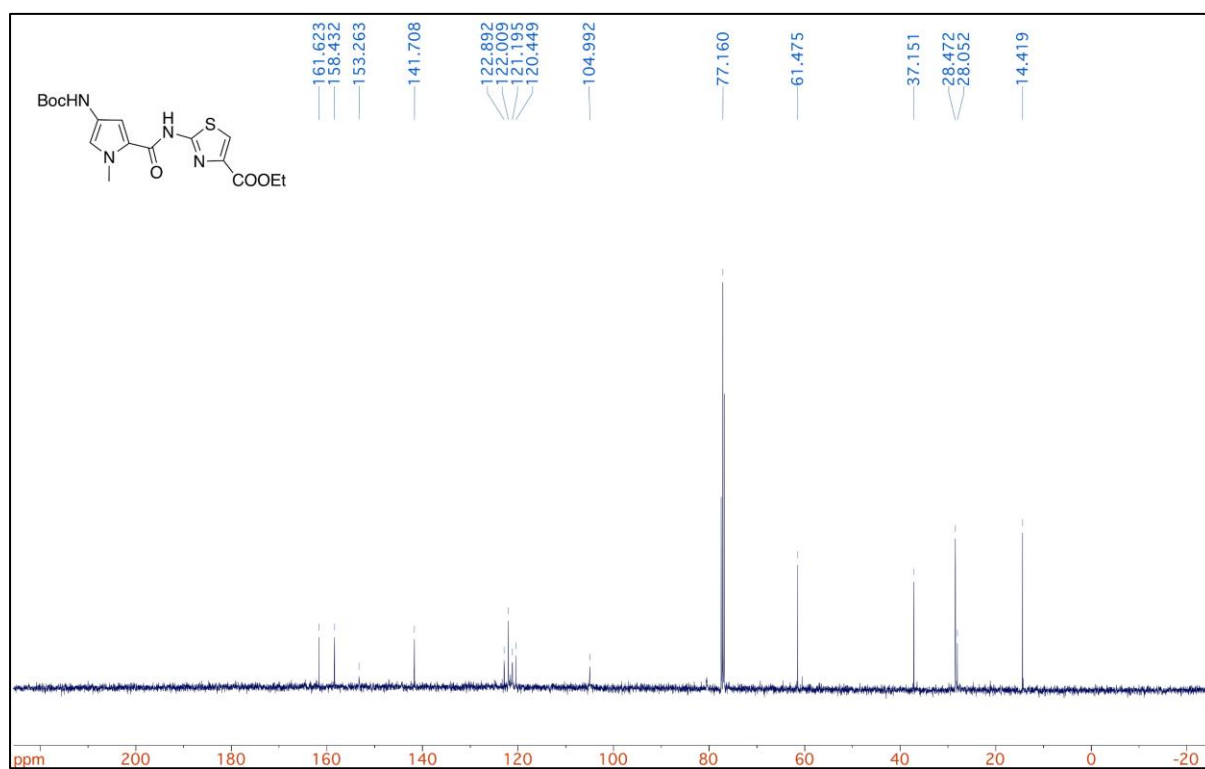


Appendix 72 HRMS of 2.90.

Methyl 2-(4-((tert-butoxycarbonyl)amino)-1-methyl-1H-pyrrole-2-carboxamido)thiazole-4-carboxylate (**2.135a**)

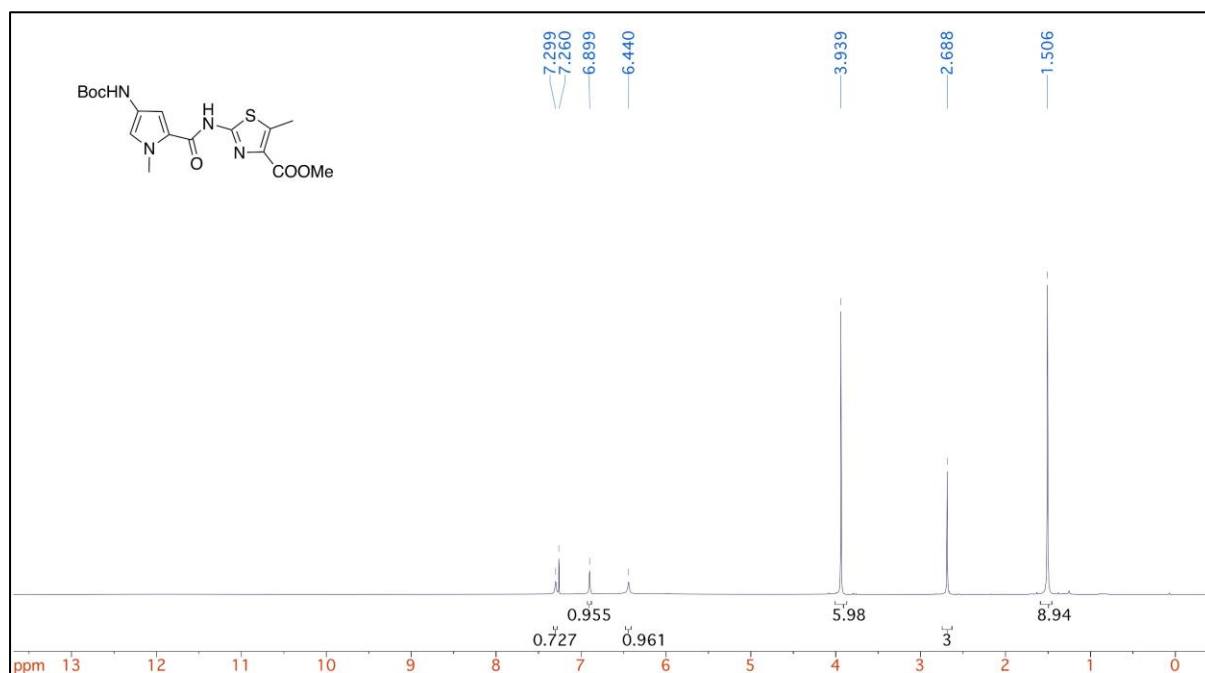


Appendix 73 ¹H-NMR of **2.135a**.

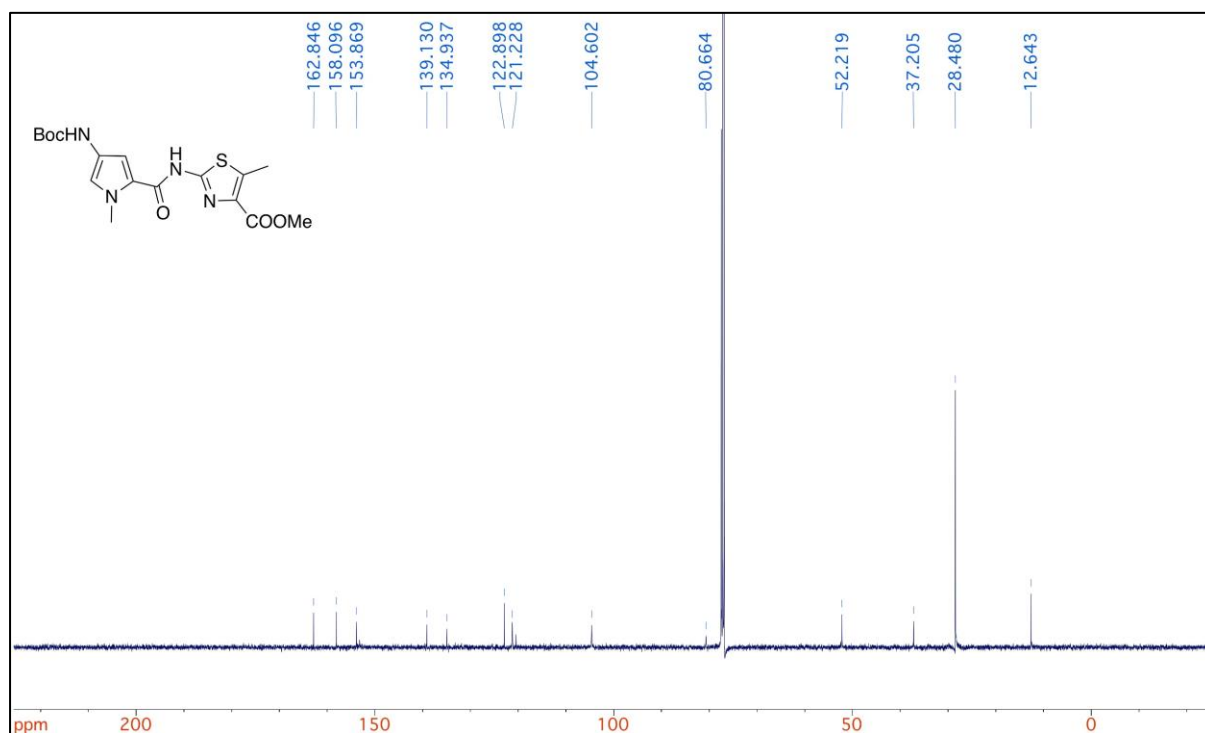


Appendix 74 ¹³C-NMR of **2.135a**.

Methyl 2-(4-((tert-butoxycarbonyl)amino)-1-methyl-1H-pyrrole-2-carboxamido)-5-methylthiazole-4-carboxylate (**2.135b**)



Appendix 75 ¹H-NMR of **2.135b**.

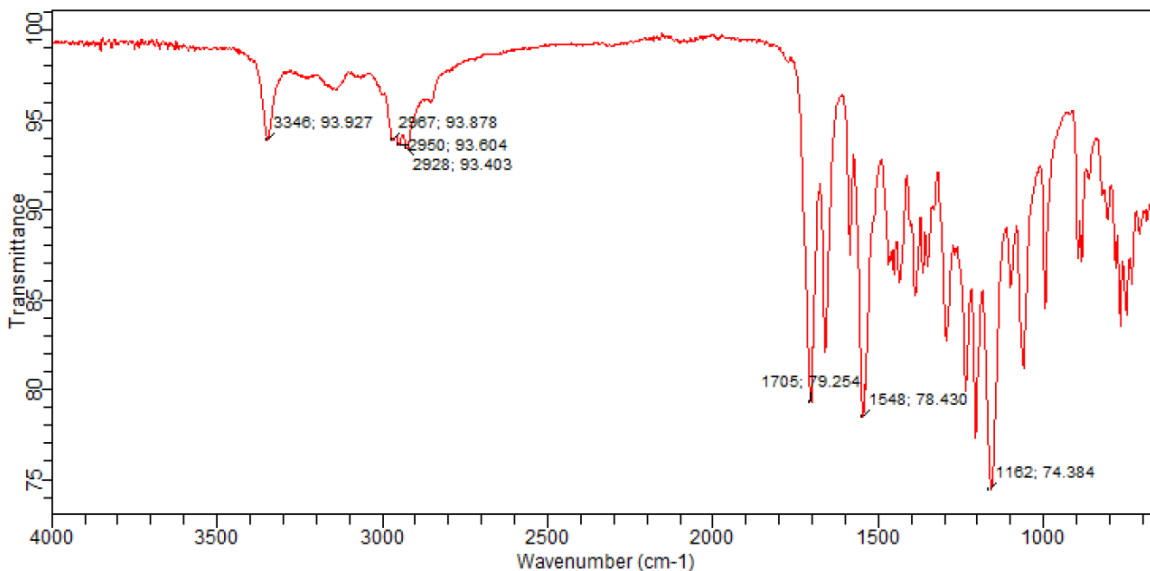


Appendix 76 ¹³C-NMR of **2.135b**.

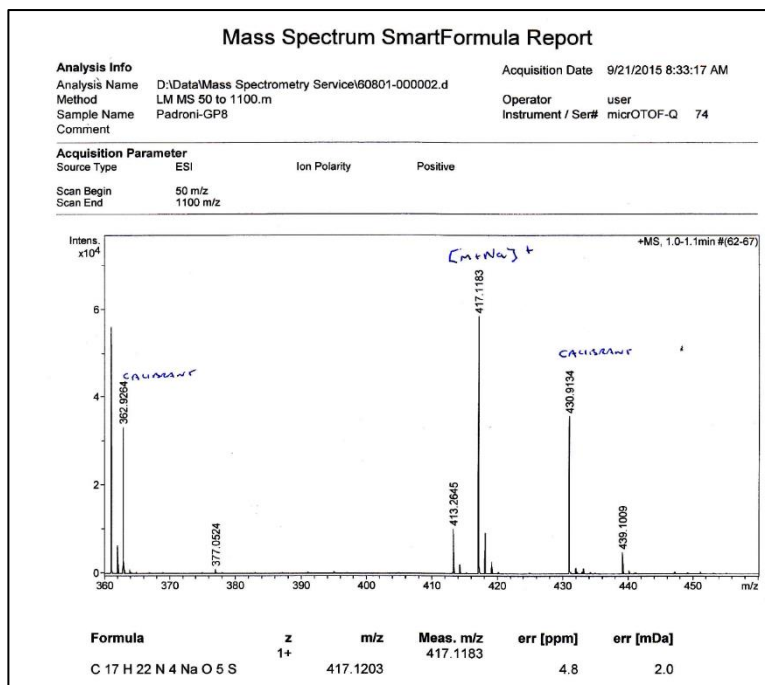


Agilent Technologies

Sample ID: BocNHPyThCH3COOMe Method Name: STUDENT ATR 32 4cm
Sample Scans: 32 User: STUDENT
Background Scans: 32 Date/Time: 10/09/2015 16:30:09
Resolution: 4 cm⁻¹ Range: 4,000.00 - 650.00
System Status: Good Apodization: Happ-Genzel
File Location: C:\Program Files (x86)\Agilent\MicroLab PC\Results\STUDENT ATR 32
4cmBocNHPyThCH3COOMe_2015-09-10T16-31-05.a2r

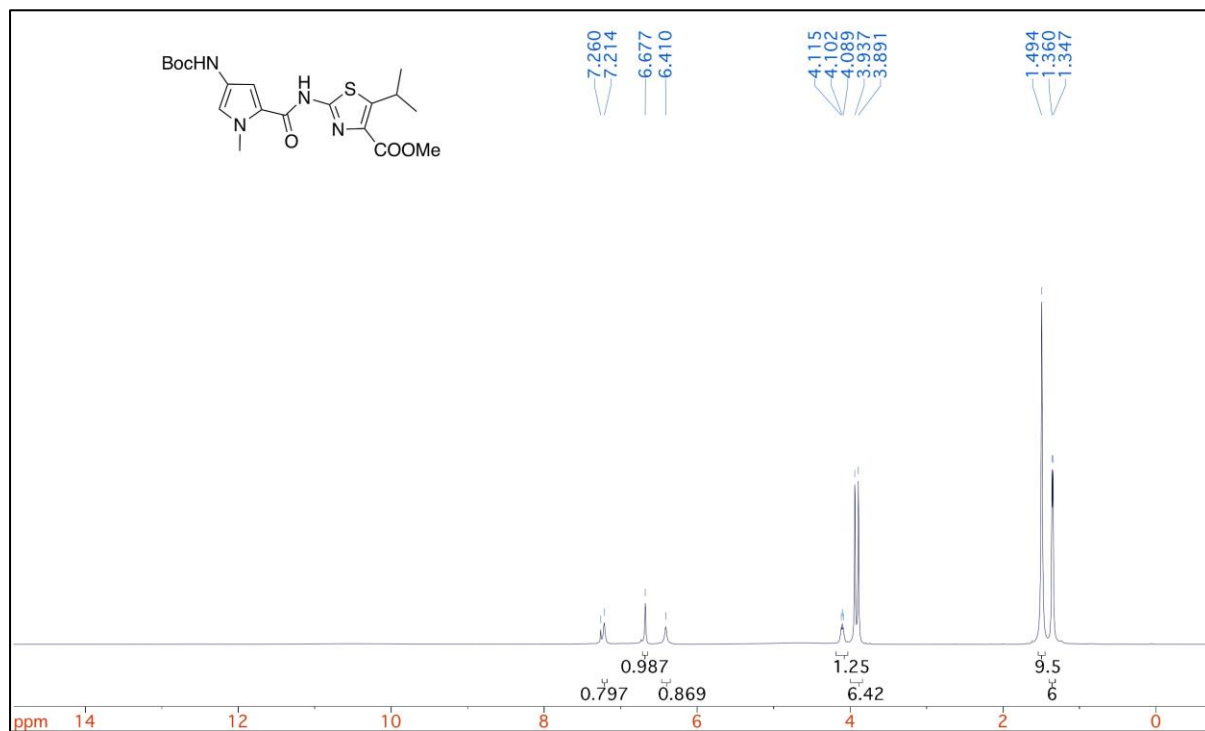


Appendix 77 IR of 2.135b.

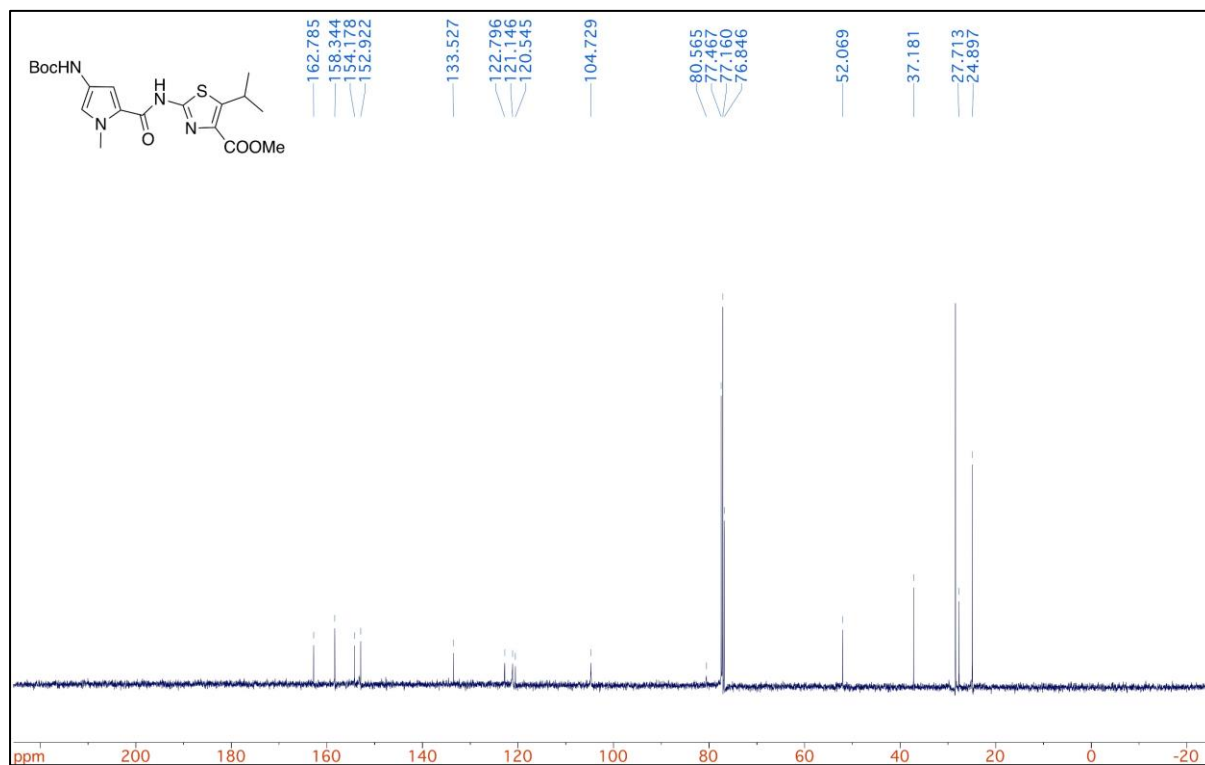


Appendix 78 HRMS of 2.135b.

Methyl 2-(4-((*tert*-butoxycarbonyl)amino)-1-methyl-1*H*-pyrrole-2-carboxamido)-5-isopropylthiazole-4-Carboxylate (**2.135c**)



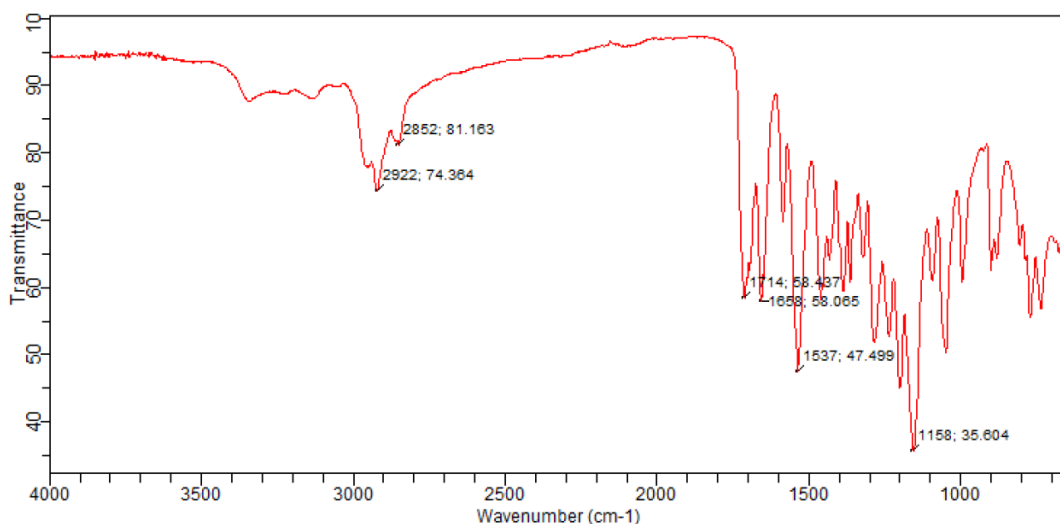
Appendix 79 ¹H-NMR of 2.135c.



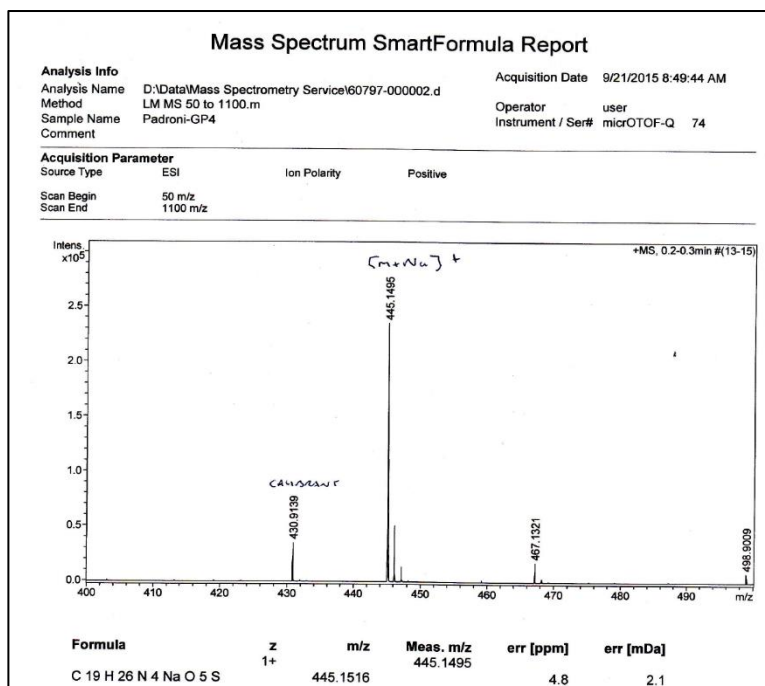
Appendix 80 ¹³C-NMR of 2.135c.



Sample ID: BocNHPyThiPrCOOMe Method Name: STUDENT ATR 32 4cm
 Sample Scans: 32 User: STUDENT
 Background Scans: 32 Date/Time: 10/09/2015 16:27:48
 Resolution: 4 cm-1 Range: 4,000.00 - 650.00
 System Status: Good Apodization: Happ-Genzel
 File Location: C:\Program Files (x86)\Agilent\MicroLab PC\Results\STUDENT ATR 32
 4cmBocNHPyThiPrCOOMe_2015-09-10T16-28-55.a2r

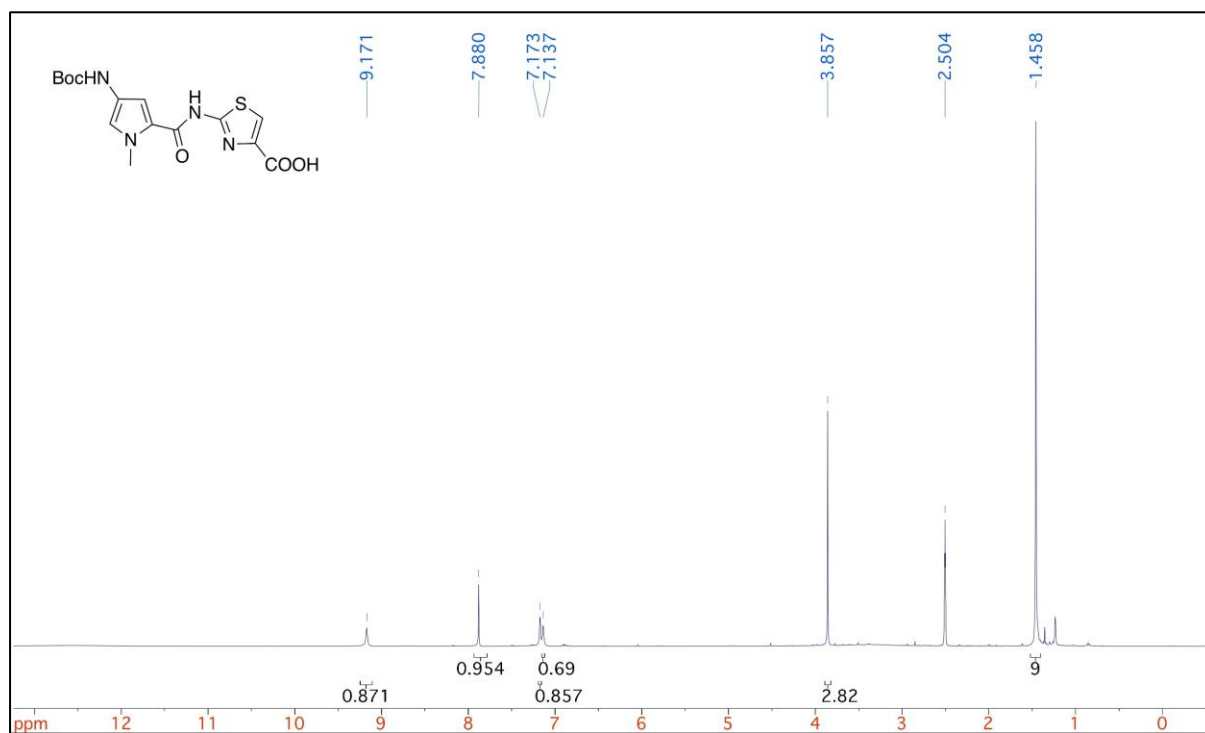


Appendix 81 IR of 2.135c.

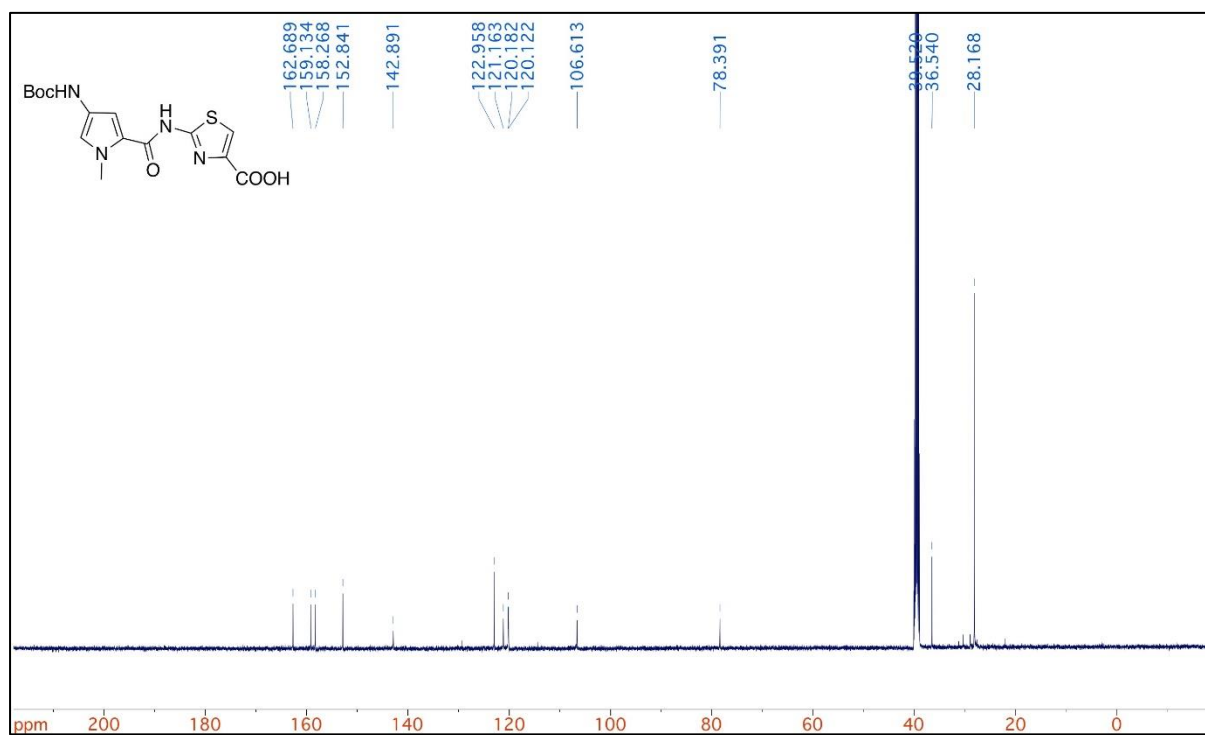


Appendix 82 HRMS of 2.135c.

2-(4-((*tert*-butoxycarbonyl)amino)-1-methyl-1*H*-pyrrole-2-carboxamido)thiazole-4-carboxylic acid (2.91a)⁶

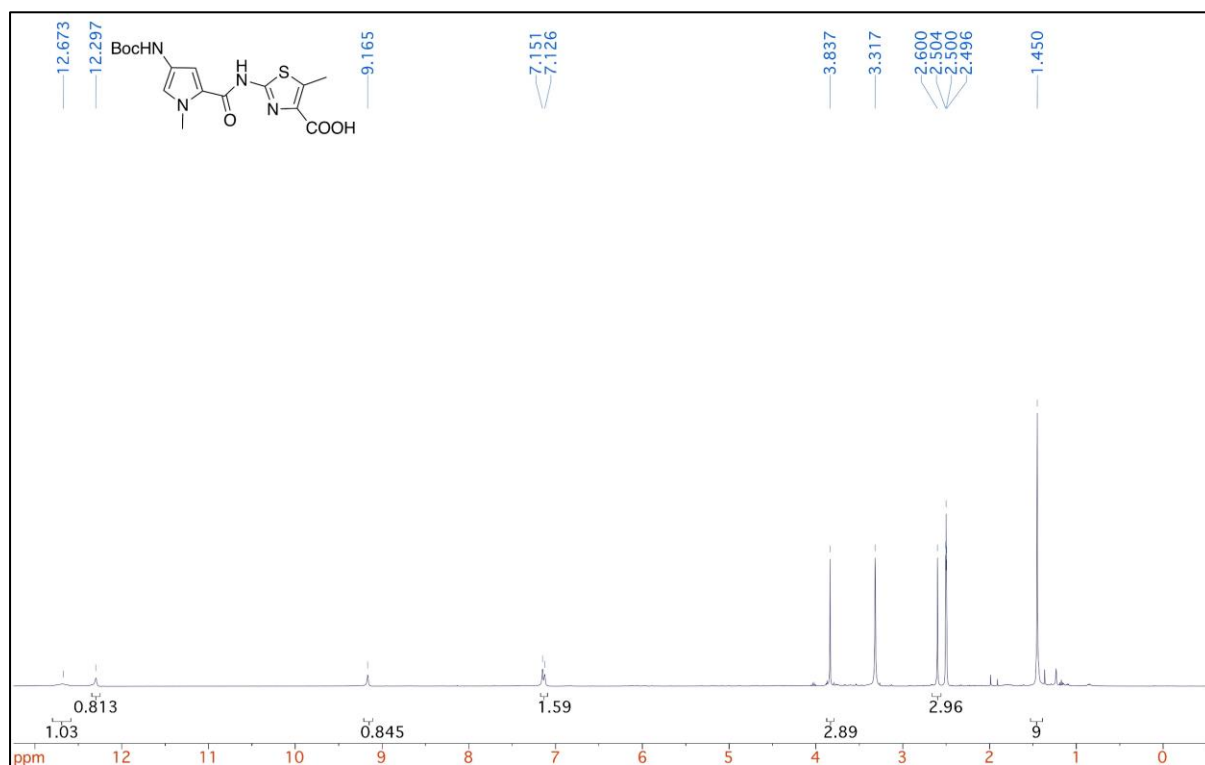


Appendix 83 ¹H-NMR of 2.91a

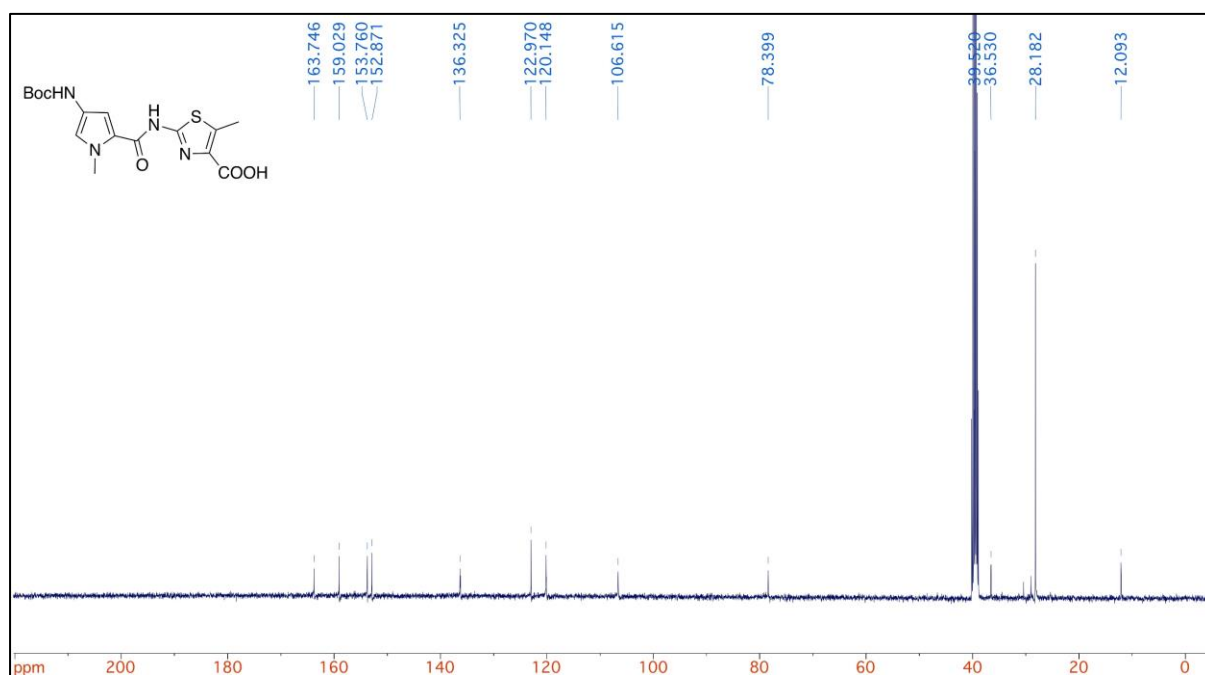


Appendix 84 ¹³C-NMR of 2.91a.

2-(4-((*tert*-butoxycarbonyl)amino)-1-methyl-1*H*-pyrrole-2-carboxamido)-5-methylthiazole-4-carboxylic acid (**2.91b**)



Appendix 85 ¹H-NMR of **2.91b**



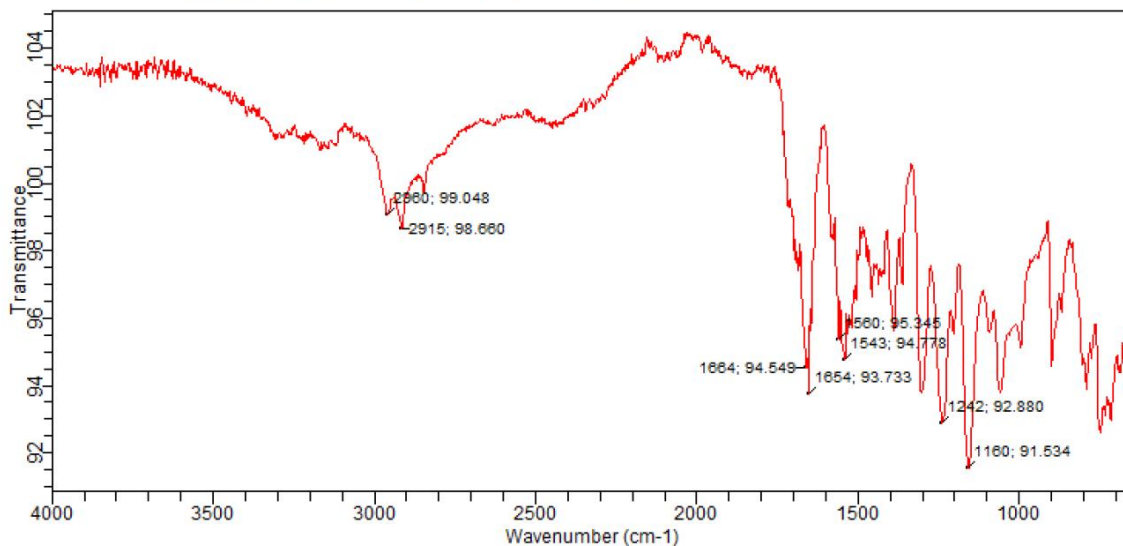
Appendix 86 ¹³C-NMR of **2.91b**



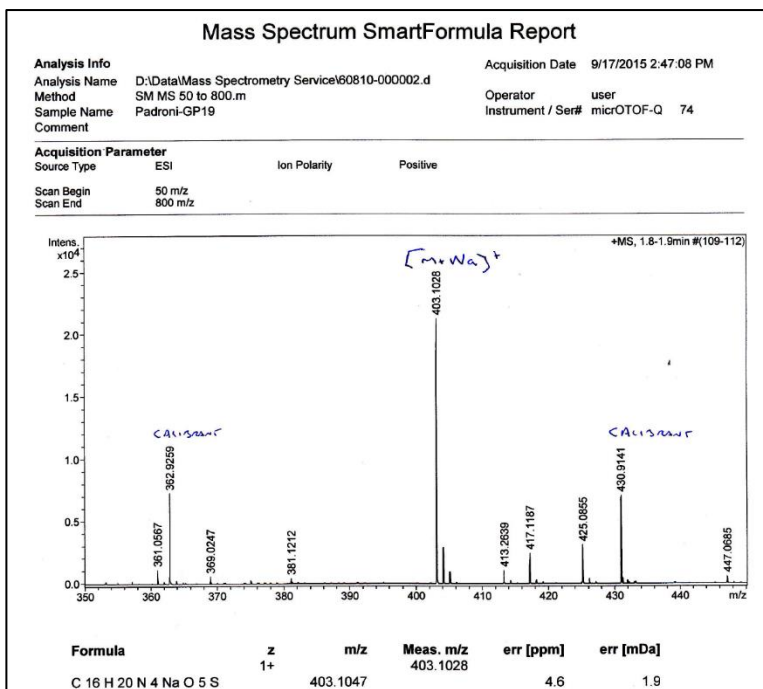
Agilent Technologies

Sample ID: BocNHPyThCH3COOH
Sample Scans: 32
Background Scans: 32
Resolution: 4 cm-1
System Status: Good
File Location: C:\Program Files (x86)\Agilent\MicroLab PC\Results\STUDENT ATR 32 4cm\BocNHPyThCH3COOH_2015-09-22T10-36-35.a2r

Method Name: STUDENT ATR 32 4cm
User: STUDENT
Date/Time: 22/09/2015 10:35:05
Range: 4,000.00 - 650.00
Apodization: Happ-Genzel

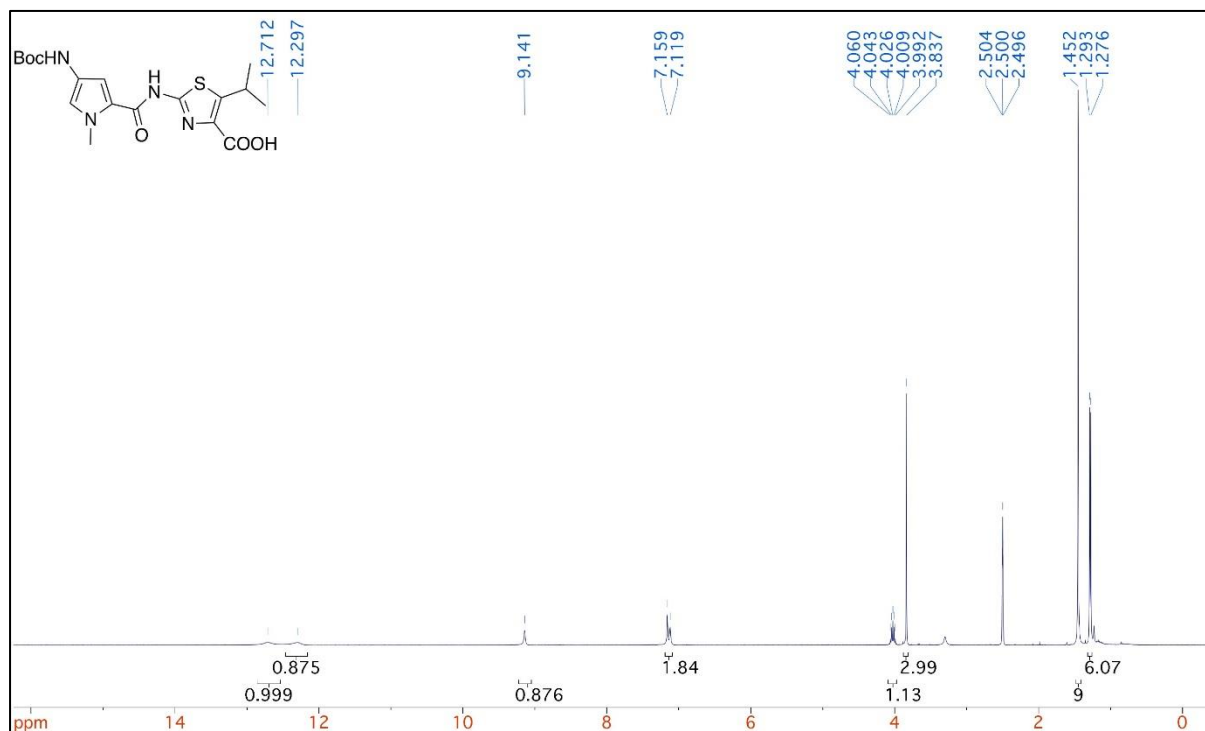


Appendix 87 IR of 2.91b.

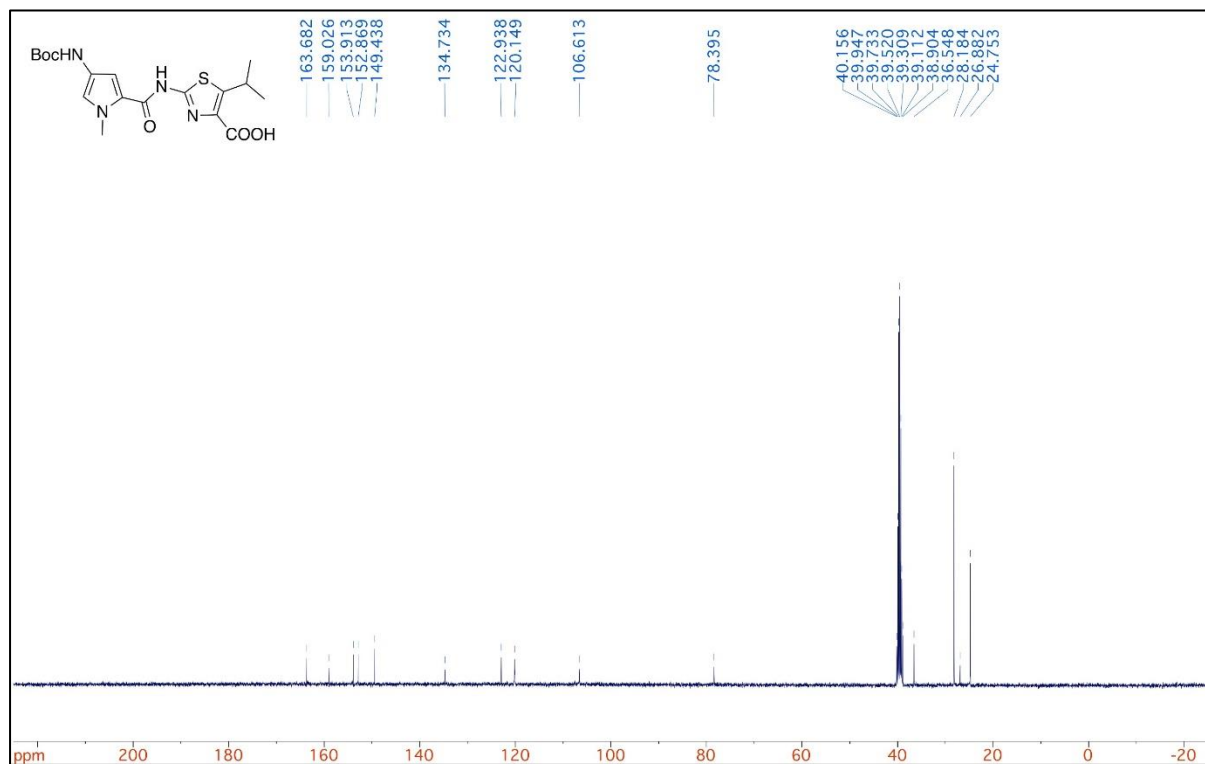


Appendix 88 HRMS of 2.91b.

2-(4-((*tert*-butoxycarbonyl)amino)-1-methyl-1*H*-pyrrole-2-carboxamido)-5-isopropylthiazole-4-carboxylic acid (**2.91c**)



Appendix 89 ¹H-NMR of **2.91c**.

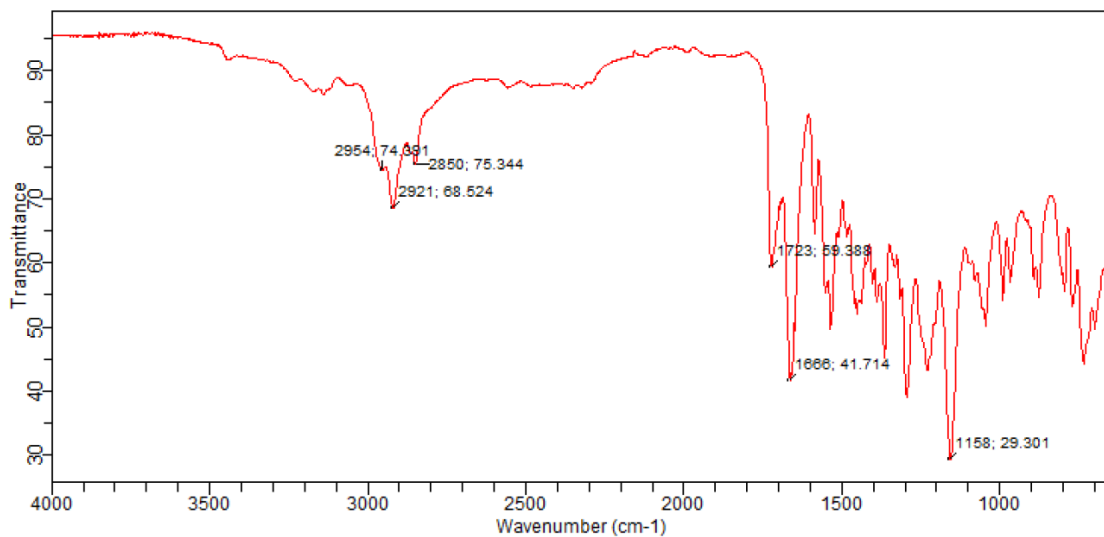


Appendix 90 ¹³C-NMR of **2.91c**.

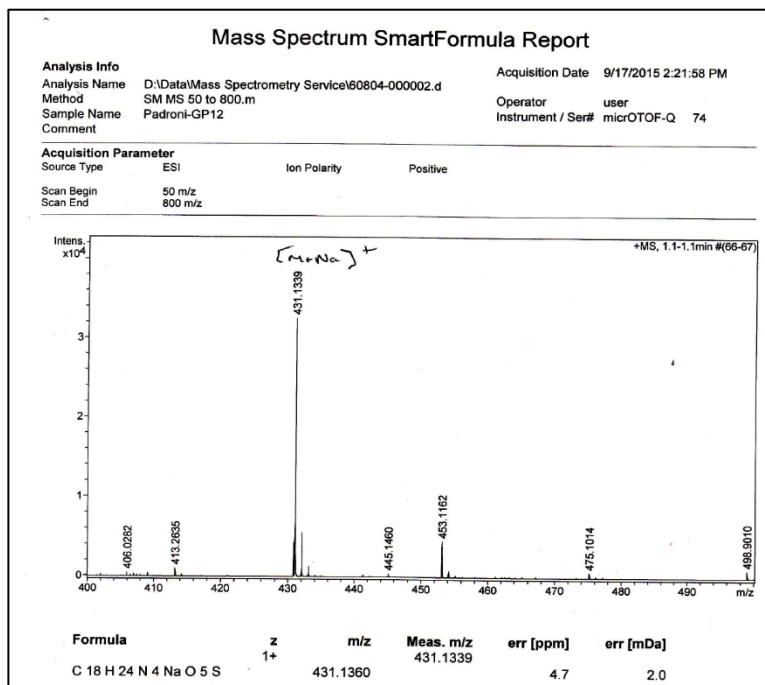


Agilent Technologies

Sample ID: BocNHPyThiPrCOOH Method Name: STUDENT ATR 32 4cm
Sample Scans: 32 User: STUDENT
Background Scans: 32 Date/Time: 10/09/2015 16:25:00
Resolution: 4 cm-1 Range: 4,000.00 - 650.00
System Status: Good Apodization: Happ-Genzel
File Location: C:\Program Files (x86)\Agilent\MicroLab PC\Results\STUDENT ATR 32
4cm\BocNHPyThiPrCOOH_2015-09-10T16-26-23.a2r

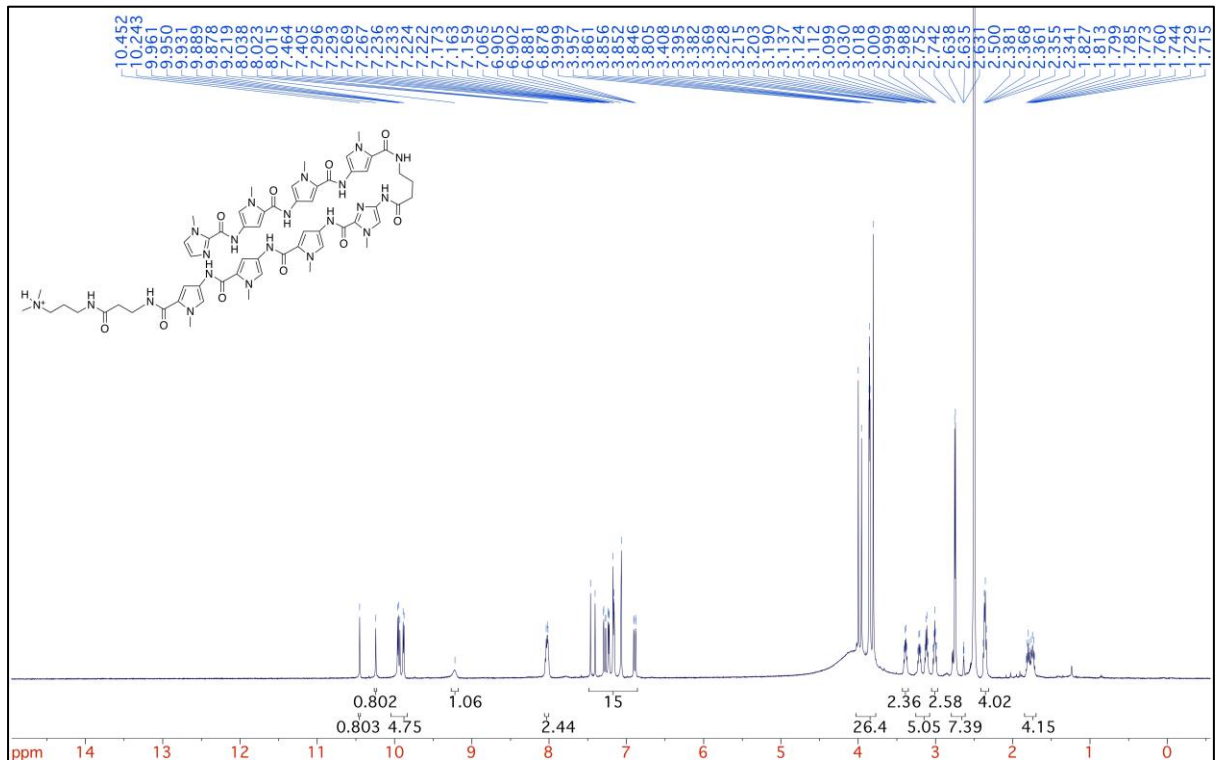


Appendix 91 IR of 2.91c.

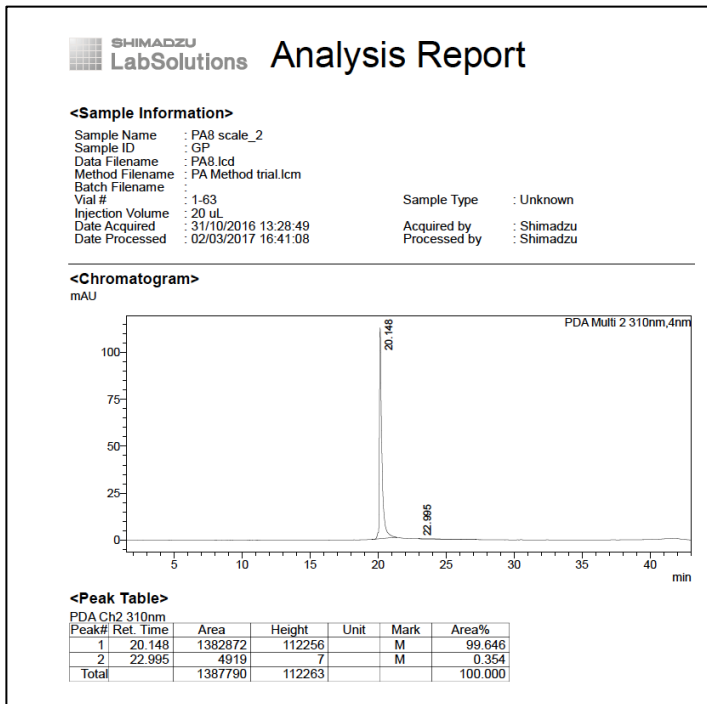


Appendix 92 HRMS of 2.91c.

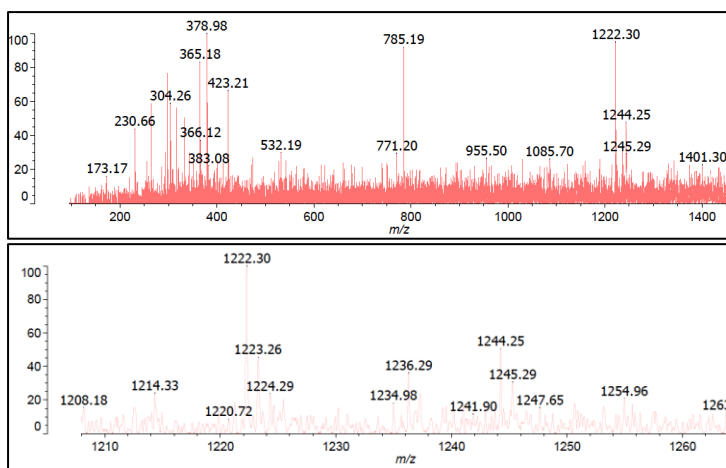
PA1 (2.106)



Appendix 93 ¹H-NMR of PA1.

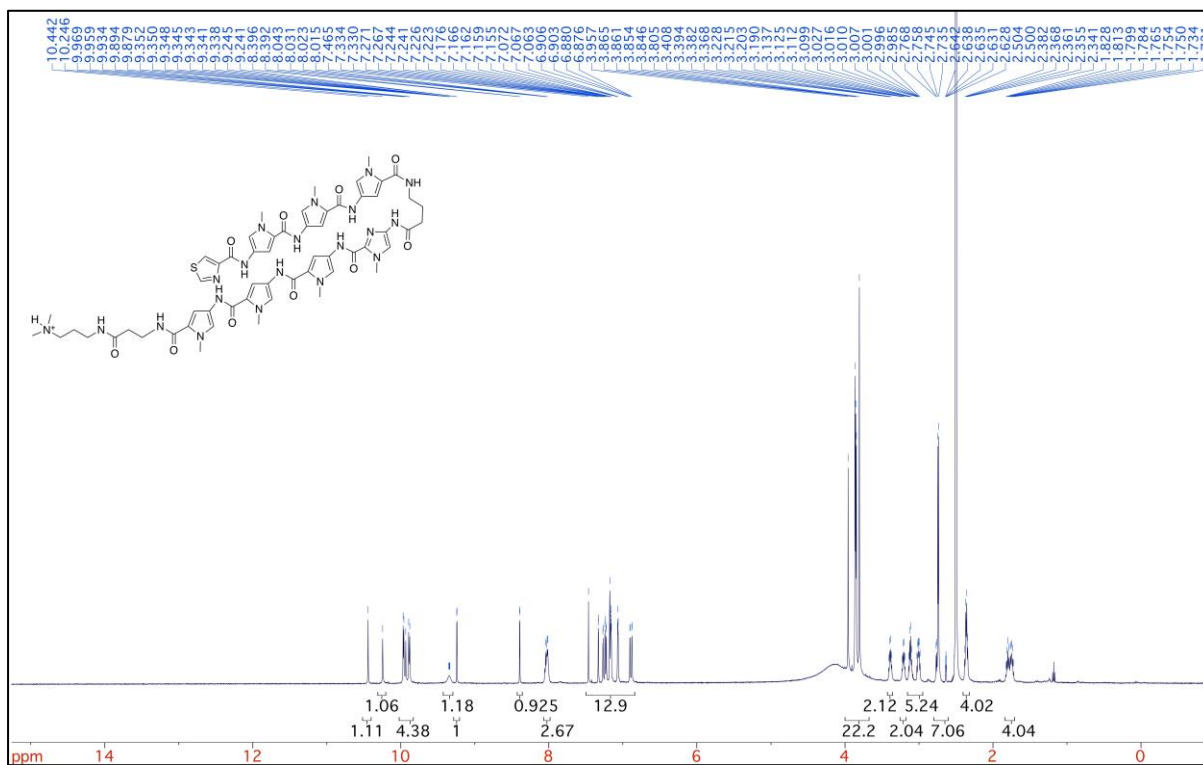


Appendix 94 HPLC of PA1.



Appendix 95 MALDI-TOF of PA1.

PA2 (2.122a)



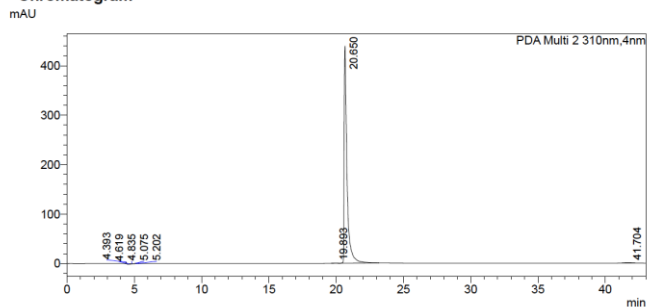
Appendix 96 ¹H-NMR of PA2.

SHIMADZU LabSolutions Analysis Report

<Sample Information>

Sample Name : PA5
 Sample ID : Gp
 Data Filename : GpPA5-7_11042017_002.lcd
 Method Filename : PA Method trial.lcm
 Batch Filename : GpPA5-7.lcb
 Vial # : 1-2
 Injection Volume : 10 uL
 Date Acquired : 11/04/2017 10:44:50
 Date Processed : 11/04/2017 11:27:54
 Sample Type : Unknown
 Acquired by : Shimadzu
 Processed by : Shimadzu

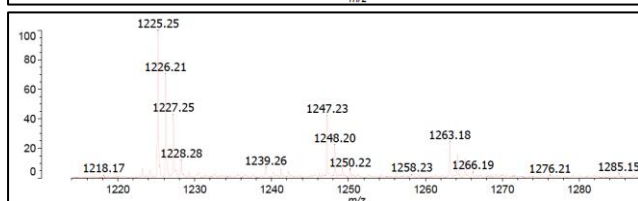
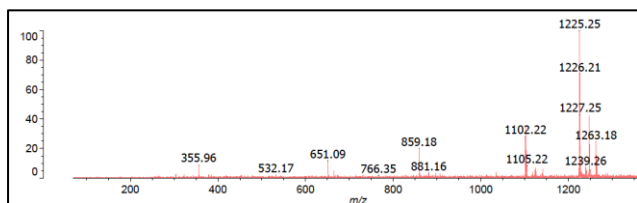
<Chromatogram>



<Peak Table>

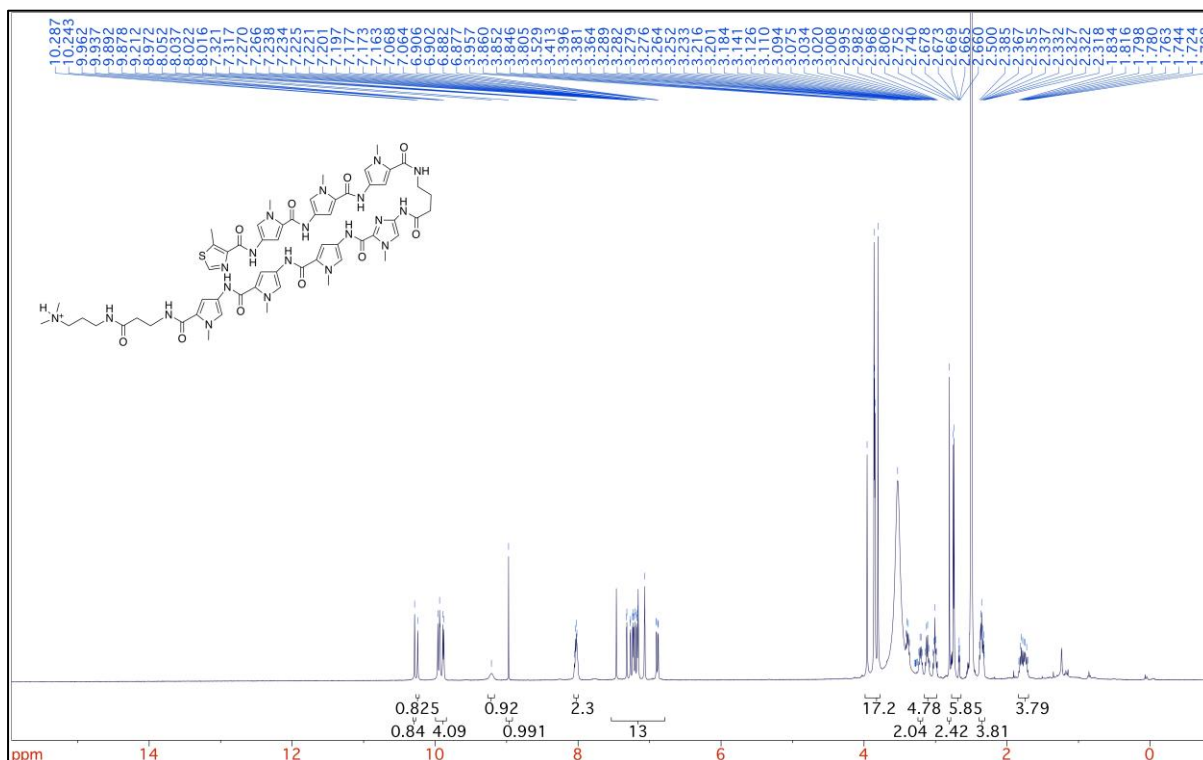
Peak#	Ret. Time	Area	Height	Area%
1	4.393	14260	3250	0.219
2	4.619	3278	478	0.050
3	4.835	8373	638	0.129
4	5.075	5584	684	0.086
5	5.202	3131	443	0.048
6	19.893	3754	204	0.058
7	20.650	6417977	439346	98.766
8	41.704	41817	819	0.644
Total		6498175	445862	100.000

Appendix 97 HPLC of PA2.

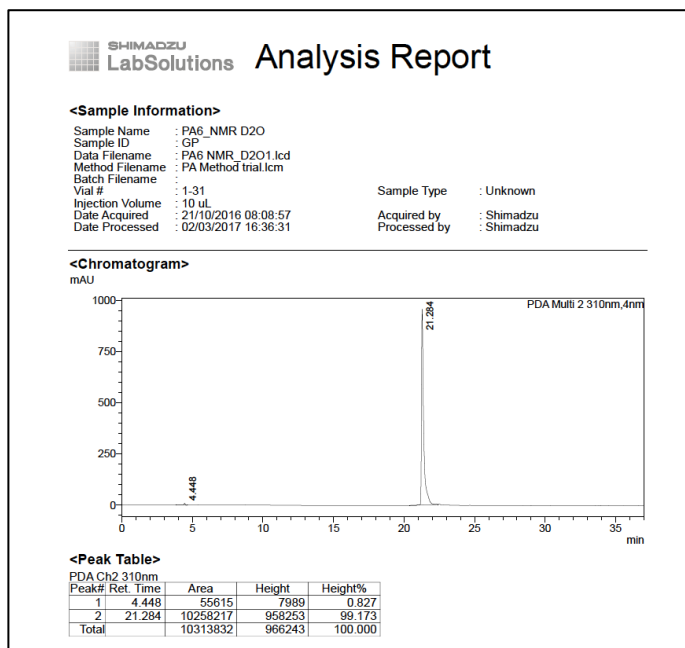


Appendix 98 MALDI-TOF of PA2.

PA3 (2.122b)



Appendix 99 ¹H-NMR of PA3



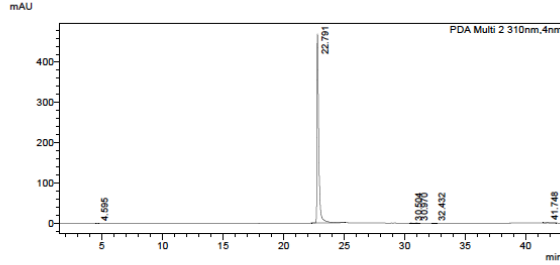
Appendix 100 HPLC of PA3.

SHIMADZU LabSolutions Analysis Report

<Sample Information>

Sample Name : PA7
 Sample ID : GP
 Data Filename : PA7.lcd
 Method Filename : PA Method trial.lcm
 Batch Filename :
 Vial # : 1-62
 Injection Volume : 20 uL
 Date Acquired : 31/10/2016 12:23:51
 Date Processed : 31/10/2016 17:45:41
 Sample Type : Unknown
 Acquired by : Shimadzu
 Processed by : Shimadzu

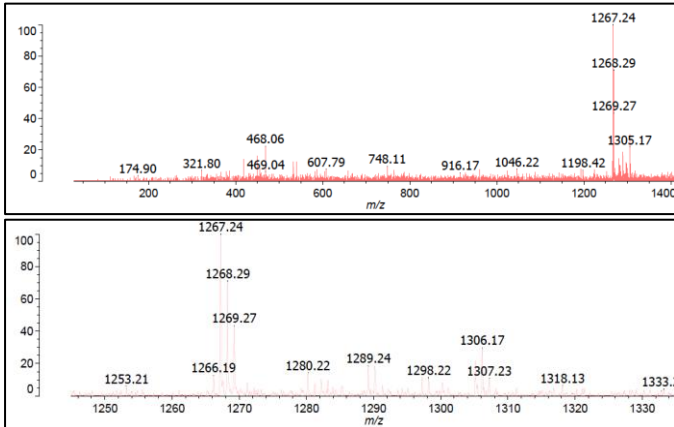
<Chromatogram>



<Peak Table>

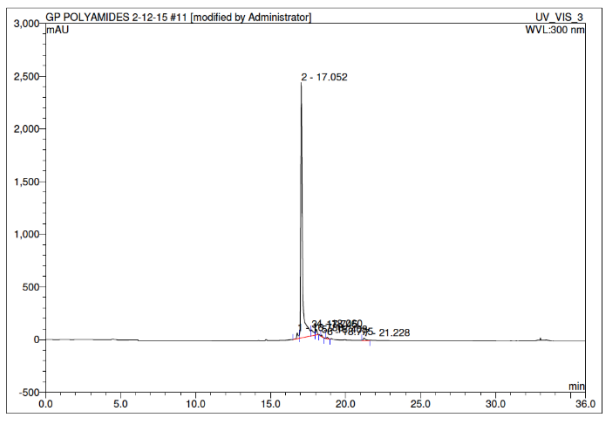
Peak#	Ret. Time	Area	Height	Conc.	Unit	Mark	Name
1	4.595	1625	304	0.033			
2	22.791	4918119	469676	99.638			
3	30.504	1241	131	0.025			
4	30.970	2154	183	0.044			
5	32.432	1402	156	0.028			
6	41.748	11467	323	0.232			
Total		4936008	470776				

Appendix 103 HPLC of PA4.



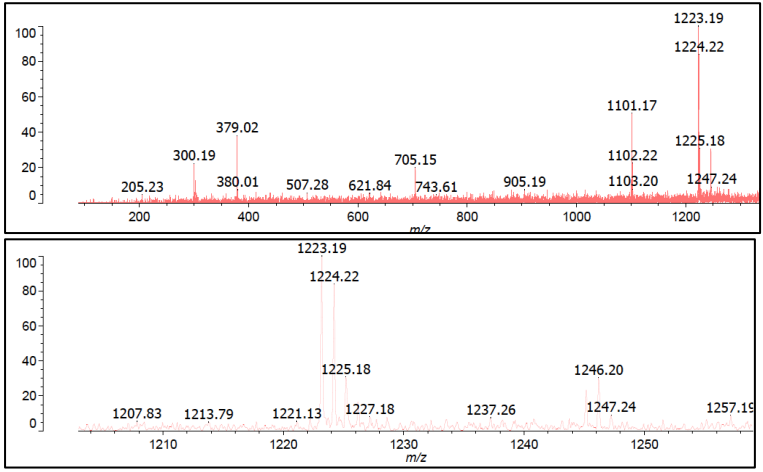
Appendix 104 MALDI-TOF of PA4.

PA5 (2.134)



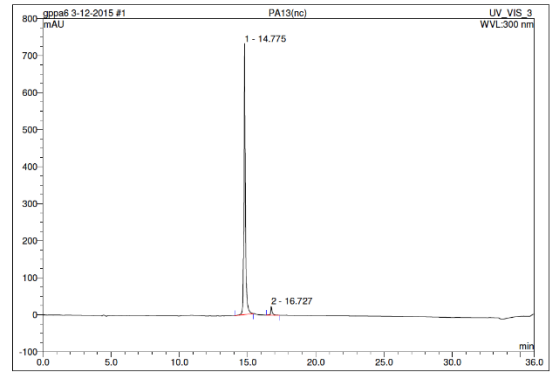
No.	Ret.Time min	Peak Name	Height mAU	Area mAU*min	Rel.Area %	Amount	Type
1	16.79	n.a.	55,079	6,382	1.88	n.a.	BM
2	17.05	n.a.	2422.449	311.542	91.56	n.a.	M
3	17.75	n.a.	55,126	11,145	3.28	n.a.	M
4	18.06	n.a.	54,519	5,524	1.62	n.a.	MB
5	18.41	n.a.	16,252	1,317	0.39	n.a.	BMB
6	18.78	n.a.	12,937	1,410	0.41	n.a.	BMB
7	21.23	n.a.	19,723	2,943	0.86	n.a.	BMB
Total:			2636.085	340.263	100.00	0.000	

Appendix 105 HPLC of PA5.



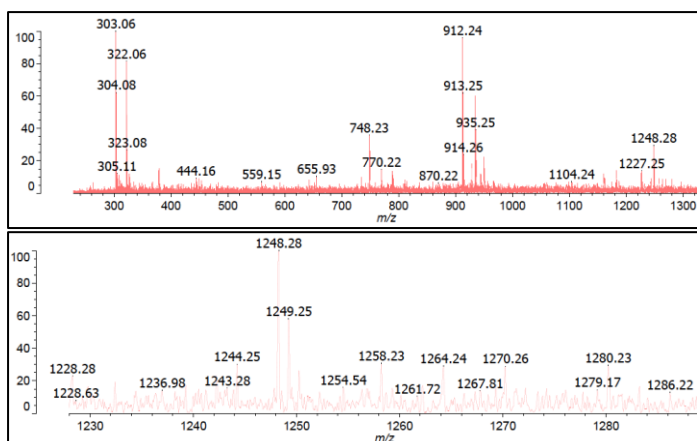
Appendix 106 MALDI-TOF of PA5.

PA6 (2.141a)



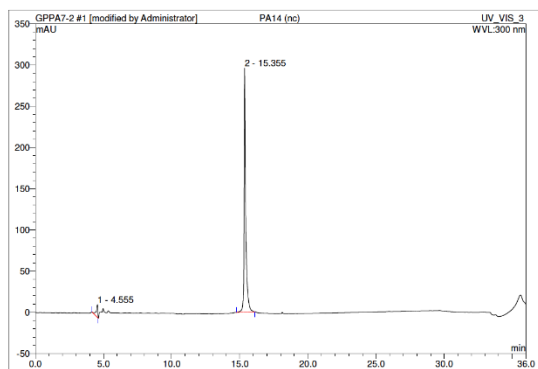
No.	Ret.Time min	Peak Name	Height mAU	Area mAU*min	Rel.Area %	Amount	Type
1	14.78	n.a.	731.705	91.653	96.74	n.a.	BMB
2	16.73	n.a.	23.669	3.088	3.26	n.a.	BMB
Total:			755.374	94.741	100.00	0.000	

Appendix 107 HPLC of PA6.



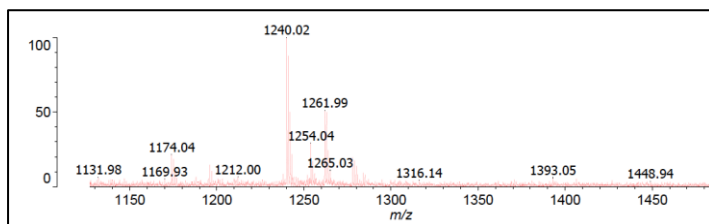
Appendix 108 MALDI-TOF of PA6.

PA7 (2.141b)



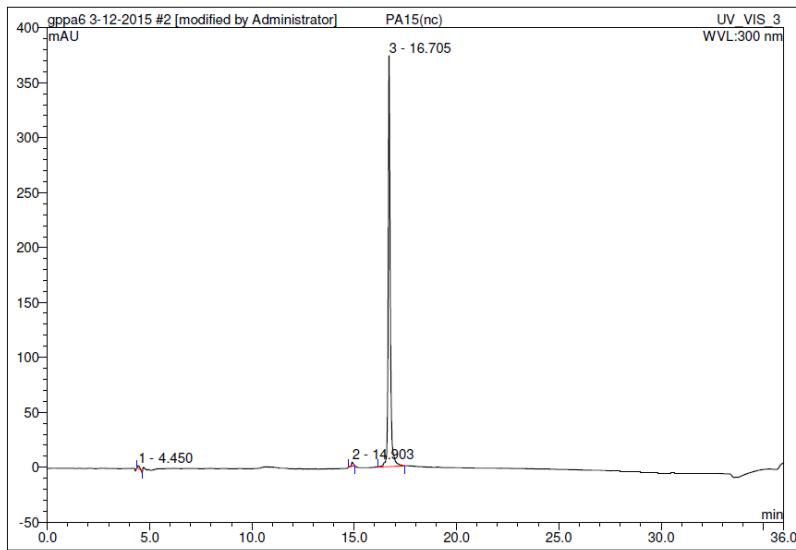
No.	Ret.Time min	Peak Name	Height mAU	Area mAU*min	Rel.Area %	Amount	Type
1	4.56	n.a.	15.359	2.003	4.65	n.a.	BMB
2	15.36	n.a.	295.949	41.062	95.35	n.a.	BMB
Total:			311.307	43.065	100.00	0.000	

Appendix 109 HPLC of PA7.



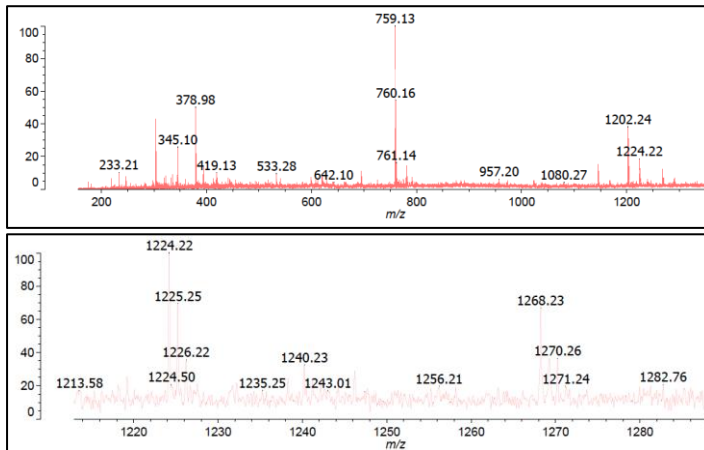
Appendix 110 MALDI-TOF of PA7.

PA8 (2.141c)



No.	Ret.Time min	Peak Name	Height mAU	Area mAU*min	Rel.Area %	Amount	Type
1	4.45	n.a.	2.465	0.259	0.60	n.a.	BMB*
2	14.90	n.a.	3.616	0.375	0.87	n.a.	BMB*
3	16.71	n.a.	373.890	42.640	98.54	n.a.	BMB
Total:			379.971	43.273	100.00	0.000	

Appendix 111 HPLC of PA8.



Appendix 112 MALDI-TOF of PA8.

3 NMR chemical shifts of PA·ODN1 complexes

Appendix 113 ¹H sugar chemical shifts of free ODN1 (H5' and H5'' were not stereochemically assigned. “'” and “''” refer to the most upfield and downfield signal, respectively).

ODN1	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 114 ¹H base chemical shifts of free ODN1.

ODN1	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 115 ¹³C base chemical shifts of free ODN1.

ODN1	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
T6	82.61	-	36.73	75.04	83.1	-	n.d	135.8	11.38	-
A7	82.09	151.9	38.04	76.57	84.49	-	n.d	-	-	138.9
C8	83.23	-	37.23	73.9	n.d.	95.27	n.d	139.5	-	-
A9	82.73	151.6	38.52	76.41	84.72	-	n.d	-	-	138.8
T10	82.41	-	36.52	74.09	82.87	-	n.d	135.8	11.39	-
C11	83.82	-	36.59	74.05	85.45	96.03	n.d	141.1	-	-
G12	82.05	-	39.15	70.79	85.28	-	n.d	-	-	136.5

Appendix 116 ¹H sugar chemical shifts of **PA1-ODN1** (H5' and H5'' were not stereochemically assigned. “'” and “''” refer to the most upfield and downfield signal, respectively).

ODN1	H1'	H2'	H2''	H3'	H4'	H5'	H5''
C1	5.764	1.857	2.368	4.709	4.064	3.709	3.709
G2	5.43	2.769	2.857	5.033	4.32	3.979	4.095
A3	6.35	2.83	3.006	5.129	4.52	4.205	4.205
T4	5.726	2.505	2.596	5.036	4.261	n.d	n.d
G5	5.982	2.611	2.611	5.026	4.213	3.618	3.752
T6	4.963	1.696	2.17	4.57	2.052	3.392	3.612
A7	5.507	2.175	2.685	4.674	2.763	n.d	3.576
C8	5.311	1.447	2.297	4.542	2.191	n.d	n.d
A9	5.567	2.183	2.558	4.655	2.671	3.427	3.562
T10	5.684	1.864	2.289	4.766	3.903	n.d	n.d
C11	5.734	1.965	2.357	4.767	4.113	n.d	n.d
G12	6.146	2.359	2.587	4.674	4.175	3.622	3.75
C13	5.743	1.969	2.43	4.726	4.071	3.741	3.741
G14	5.665	2.771	2.889	5.042	4.365	4.008	4.112
A15	6.323	2.701	2.978	5.101	4.505	4.262	4.262
T16	5.618	2.247	2.373	4.923	4.137	n.d	n.d
G17	5.784	2.638	2.638	4.993	4.262	3.624	3.753
T18	5.197	1.701	2.234	4.622	2.047	3.433	3.56
A19	5.516	2.174	2.702	4.666	2.607	3.434	3.566
C20	5.32	1.482	2.28	4.545	2.219	n.d	n.d
A21	5.491	2.181	2.482	4.638	2.833	3.446	3.571
T22	5.727	1.83	2.094	4.777	3.752	n.d	n.d
C23	5.487	2.098	2.306	4.789	4.11	n.d	n.d
G24	6.171	2.395	2.616	4.683	4.16	3.621	3.752

Appendix 117 ¹H base chemical shifts of **PA1-ODN1**.

	H1	H2	H21	H22	H3	H41	H42	H5	H6	H61	H62	H7	H8
C1	-	-	-	-	-	n.d.	n.d.	5.877	7.584	-	-	-	-

G2	12.88	-	n.d.	n.d.	-	-	-	-	-	-	-	-	8.019
A3	-	7.963	-	-	-	-	-	-	-	n.d.	n.d.	-	8.436
T4	-	-	-	-	13.73	-	-	-	7.42	-	-	1.546	-
G5	12.28	-	8.064	7.867	-	-	-	-	-	-	-	-	7.966
T6	-	-	-	-	12.58	-	-	-	6.96	-	-	1.307	-
A7	-	8.049	-	-	-	-	-	-	-	8.451	5.58	-	8.125
C8	-	-	-	-	-	8.752	6.402	5.27	6.875	-	-	-	-
A9	-	7.733	-	-	-	-	-	-	-	n.d.	n.d.	-	8.297
T10	-	-	-	-	13.54	-	-	-	6.964	-	-	1.441	-
C11	-	-	-	-	-	8.559	6.91	5.614	7.408	-	-	-	-
G12	n.d.	-	n.d.	n.d.	-	-	-	-	-	-	-	-	7.925
C13	-	-	-	-	-	n.d.	n.d.	5.911	7.645	-	-	-	-
G14	12.82	-	n.d.	n.d.	-	-	-	-	-	-	-	-	8.003
A15	-	7.896	-	-	-	-	-	-	-	n.d.	n.d.	-	8.318
T16	-	-	-	-	13.51	-	-	-	7.274	-	-	1.521	-
G17	12.75	-	8.521	7.285	-	-	-	-	-	-	-	-	8.006
T18	-	-	-	-	12.86	-	-	-	6.973	-	-	1.449	-
A19	-	8.019	-	-	-	-	-	-	-	8.541	5.574	-	8.145
C20	-	-	-	-	-	8.716	6.423	5.273	6.869	-	-	-	-
A21	-	7.705	-	-	-	-	-	-	-	n.d.	n.d.	-	8.288
T22	-	-	-	-	13.81	-	-	-	6.952	-	-	1.368	-
C23	-	-	-	-	-	8.63	6.954	5.655	7.46	-	-	-	-
G24	n.d.	-	n.d.	n.d.	-	-	-	-	-	-	-	-	7.937

Appendix 118 ¹³C ODN1 assigned chemical shifts of PA1-ODN1.

ODN1	C1'	C2	C3'	C4'	C5	C6	C7	C8
C1	85.3	-	75.72	85.78	96.95	140.5	-	-
G2	81.38	-	76.54	84.69	-	-	-	135.5
A3	82.05	152	77.35	85.17	-	-	-	138.7
T4	83.34	-	75.52	82.90	-	136.3	11.36	-
G5	82.56	-	78.75	84.91	-	-	-	135.8
T6	81.83	-	73.68	80.83	-	136.5	11.54	-
A7	80.74	152	76.15	82.99	-	-	-	139.6
C8	81.58	-	72.78	80.81	96.14	139.8	-	-
A9	80.52	151.5	74.22	82.41	-	-	-	139.5
T10	81.4	-	73.04	81.75	-	135.9	11.27	-
C11	83.61	-	75.11	83.44	95.96	141.1	-	-
G12	82.04	-	70.71	85.36	-	-	-	136.2
C13	85.3	-	75.09	85.54	96.85	140.6	-	-
G14	n.d.	-	77.18	84.71	-	-	-	136.8

A15	82.2	152.3	77.06	84.91	-	-	-	138.6
T16	82.4	-	73.22	82.19	-	136.2	11.36	-
G17	82.66	-	78.28	84.91	-	-	-	135.4
T18	81.9	-	73.88	81.5	-	136.4	12.16	-
A19	81.28	152.4	75.54	82.68	-	-	-	139.7
C20	81.56	-	72.89	80.81	96.12	139.6	-	-
A21	80.56	150.9	74.44	82.48	-	-	-	139.3
T22	81.42	-	73.54	82.08	-	135.9	11.3	-
C23	84.07	-	73.92	82.79	96.15	141.2	-	-
G24	82.26	-	70.9	85.36	-	-	-	136.5

Appendix 119 ¹H PA1 chemical shifts of PA1-ODN1 (H2'/H2'', H3'/H3'' and H4'/H4'' were not stereochemically assigned. “''' and “''” refer to the most upfield and downfield signal, respectively).

PA1	H2'	H2''	H3	H3'	H3''	H4	H4'	H4''	H5	CH₃	NH
Dp	3.151	3.17	-	1.938	2.126	-	2.952	3.619	-	n.d.	8.094
β-Ala	1.664	2.155	-	2.886	3.791	-	-	-	-	-	9.138
Py1	-	-	6.305	-	-	-	-	-	7.105	3.617	9.303
Py2	-	-	6.328	-	-	-	-	-	7.527	3.711	10.67
Py3	-	-	6.19	-	-	-	-	-	7.607	3.86	9.821
Im4	-	-	-	-	-	-	-	-	7.464	3.944	10.46
Turn	1.583	2.5	-	2.163	2.292	-	2.764	3.444	-	-	8.244
Py5	-	-	6.444	-	-	-	-	-	7.094	3.566	9.14
Py6	-	-	6.216	-	-	-	-	-	7.608	3.784	10.44
Py7	-	-	6.036	-	-	-	-	-	7.56	3.901	9.713
Im8	-	-	-	-	-	7.17	-	-	7.057	3.962	-

Appendix 120 ¹³C PA1 chemical shifts of PA1-ODN1.

PA1	C3	C4	C5	CM
Py1	106.6	-	119.3	36.2
Py2	102.5	-	121.7	37.08
Py3	104.3	-	-	37.84
Im4	-	113.9	-	36.37
Py5	105.3	-	119	36.25
Py6	102.7	-	122	37.14
Py7	104.3	-	120.7	37.56
Im8	-	130.2	124.4	35.55

Appendix 121 ¹H sugar chemical shifts of PA3-ODN1 (H5' and H5'' were not stereochemically assigned. “''' and “''” refer to the most upfield and downfield signal, respectively).

ODN1	H1'	H2'	H2''	H3'	H4'	H5'	H5''
C1	5.759	1.851	2.372	4.699	4.058	3.707	3.707

G2	5.432	2.761	2.852	5.03	4.316	3.973	4.098
A3	6.346	2.828	3.007	5.122	4.515	4.203	4.261
T4	5.724	2.492	2.575	5.03	4.254	n.d.	n.d.
G5	5.858	2.678	2.587	5.029	4.251	n.d.	n.d.
T6	4.893	1.721	2.211	4.564	2.609	3.384	3.851
A7	5.514	2.182	2.684	4.671	2.831	3.579	n.d.
C8	5.304	1.439	2.304	4.539	2.173	3.421	3.782
A9	5.558	2.174	2.561	4.653	2.645	3.55	3.663
T10	5.683	1.86	2.288	4.763	3.897	3.443	3.659
C11	5.731	1.959	2.352	4.816	4.101	n.d.	n.d.
G12	6.148	2.364	2.592	4.674	4.174	4.084	n.d.
C13	5.742	1.965	2.431	4.72	4.065	3.735	3.735
G14	5.66	2.764	2.884	5.037	4.36	4.002	4.115
A15	6.321	2.696	2.971	5.094	4.5	4.256	4.255
T16	5.611	2.243	2.369	4.921	4.132	4.178	4.248
G17	5.78	2.633	2.634	4.99	4.258	3.644	3.837
T18	5.19	1.697	2.244	4.62	2.046	n.d.	n.d.
A19	5.506	2.16	2.699	4.664	2.604	3.576	3.78
C20	5.331	1.511	2.296	4.549	2.316	3.445	3.777
A21	5.473	2.178	2.475	4.642	2.886	3.537	n.d.
T22	5.722	1.835	2.097	4.765	3.741	3.531	3.736
C23	5.487	2.095	2.308	4.791	4.11	n.d.	n.d.
G24	6.17	2.396	2.613	4.676	4.162	4.048	n.d.

Appendix 122 ¹H base chemical shifts of PA3-ODN1.

ODN1	H1	H2	H21	H22	H3	H41	H42	H5	H6	H61	H62	M7	H8
C1	-	-	-	-	-	n.d.	n.d.	5.879	7.591	-	-	-	-
G2	12.89	-	n.d.	n.d.	-	-	-	-	-	-	-	-	8.015
A3	-	7.964	-	-	-	-	-	-	-	n.d.	n.d.	-	8.429
T4	-	-	-	-	13.72	-	-	-	7.415	-	-	1.554	-
G5	12.46	-	8.303	6.76	-	-	-	-	-	-	-	-	7.993
T6	-	-	-	-	12.35	-	-	-	6.955	-	-	1.346	-
A7	-	8.041	-	-	-	-	-	-	-	8.454	n.d.	-	8.119
C8	-	-	-	-	-	8.608	6.397	5.273	6.876	-	-	-	-
A9	-	7.731	-	-	-	-	-	-	-	n.d.	n.d.	-	8.297
T10	-	-	-	-	13.53	-	-	-	6.962	-	-	1.437	-
C11	-	-	-	-	-	8.395	6.907	5.614	7.403	-	-	-	-
G12	n.d.	-	n.d.	n.d.	-	-	-	-	-	-	-	-	7.933
C13	-	-	-	-	-	n.d.	n.d.	5.912	7.648	-	-	-	-
G14	12.82	-	n.d.	n.d.	-	-	-	-	-	-	-	-	8.001

A15	-	7.891	-	-	-	-	-	-	-	n.d.	n.d.	-	8.312
T16	-	-	-	-	13.5	-	-	-	7.27	-	-	1.517	-
G17	12.74	-	8.494	7.239	-	-	-	-	-	-	-	-	8.002
T18	-	-	-	-	12.78	-	-	-	6.969	-	-	1.438	-
A19	-	8.022	-	-	-	-	-	-	-	8.446	5.602	-	8.141
C20	-	-	-	-	-	8.691	6.501	5.315	6.914	-	-	-	-
A21	-	7.711	-	-	-	-	-	-	-	n.d.	n.d.	-	8.299
T22	-	-	-	-	13.82	-	-	-	6.95	-	-	1.355	-
C23	-	-	-	-	-	8.63	6.951	5.649	7.457	-	-	-	-
G24	n.d.	-	n.d.	n.d.	-	-	-	-	-	-	-	-	7.941

Appendix 123 ¹³C ODN1 assigned chemical shifts of PA3-ODN1.

ODN1	C1'	C2	C3'	C4'	C5	C6	C7	C8
C1	85.29	-	75.63	85.7	96.83	140.4	-	-
G2	81.44	-	76.59	84.72	-	-	-	135.7
A3	82.04	152.1	77.39	85.14	-	-	-	138.9
T4	83.11	-	75.63	84.94	-	136.3	11.42	-
G5	82.6	-	78.54	n.d.	-	-	-	135.7
T6	82.32	-	74.19	82.66	-	-	11.84	-
A7	81.47	152.1	n.d.	83.11	-	-	-	140
C8	81.68	-	73.02	80.77	96.28	139.9	-	-
A9	80.55	151.5	74.22	82.42	-	-	-	139.7
T10	81.44	-	73.71	81.74	-	135.9	11.36	-
C11	83.56	-	73.97	82.79	96.02	141.1	-	-
G12	82.07	-	70.82	85.33	-	-	-	136.6
C13	85.3	-	75.08	85.48	96.92	140.5	-	-
G14	85.33	-	77.24	84.71	-	-	-	136.2
A15	82.25	152.4	77.12	84.88	-	-	-	138.7
T16	82.44	-	73.2	82.23	-	136.2	11.47	-
G17	82.64	-	78.27	82.89	-	-	-	137.1
T18	81.98	-	74.01	81.47	-	136.4	12.22	-
A19	81.06	152.5	n.d.	82.3	-	-	-	139.8
C20	81.62	-	72.99	80.96	96.35	139.8	-	-
A21	80.67	150.9	74.57	82.59	-	-	-	139.5
T22	81.48	-	73.11	82.07	-	-	11.32	-
C23	84.11	-	75.14	83.45	96.17	141.2	-	-
G24	82.25	-	70.82	85.33	-	-	-	136.5

Appendix 124 ¹H PA3 assigned chemical shifts of PA3-ODN1 (H2'/H2'', H3'/H3'' and H4'/H4'' were not stereochemically assigned. "'' and '''' refer to the most upfield and downfield signal, respectively).

PA3	H2	H2'	H2''	H3	H3'	H3''	H4	H4'	H4''	H5	CH₃	NH
------------	-----------	------------	-------------	-----------	------------	-------------	-----------	------------	-------------	-----------	-----------------------	-----------

Dp	-	3.144	3.17	-	1.943	2.137	-	2.975	3.612	-	n.d.	8.135
β-Ala	-	1.669	2.078	-	2.892	3.779	-	-	-	-	-	9.156
Py1	-	-	-	6.307	-	-	-	-	-	7.099	3.543	9.27
Py2	-	-	-	6.363	-	-	-	-	-	7.523	3.775	10.68
Py3	-	-	-	6.17	-	-	-	-	-	7.623	3.871	9.769
Im4	-	-	-	-	-	-	-	-	-	7.46	3.947	10.42
Turn	-	2.173	2.48	-	1.555	2.272	-	2.769	3.448	-	-	8.23
Py5	-	-	-	6.428	-	-	-	-	-	7.084	3.542	9.158
Py6	-	-	-	6.231	-	-	-	-	-	7.611	3.782	10.49
Py7	-	-	-	5.99	-	-	-	-	-	7.628	3.921	9.737
Nt8	8.968	-	-	-	-	-	-	-	-	-	2.773	-

Appendix 125 ¹³C **PA3** assigned chemical shifts of **PA3-ODN1**.

PA3	C2	C3	C4	C5	CM
Py1	-	106.4	-	120.2	36.29
Py2	-	102.7	-	121.9	37.25
Py3	-	104.4	-	122.1	37.83
Im4	-	-	-	114.1	36.42
Py5	-	105.3	-	119.1	36.17
Py6	-	102.8	-	121.8	n.d.
Py7	-	104.3	-	121.1	37.64
NtMe8	152.5	-	-	-	12.44

Appendix 126 ¹H sugar chemical shifts of **PA4-ODN1** (H5' and H5'' were not stereochemically assigned. “'” and “''” refer to the most upfield and downfield signal, respectively).

ODN1	H1'	H2'	H2''	H3'	H4'	H5'	H5''
C1	5.76	1.855	2.372	4.708	4.062	3.706	3.706
G2	5.44	2.768	2.856	5.033	4.321	3.977	4.102
A3	6.352	2.827	3.009	5.128	4.521	4.209	4.266
T4	5.768	2.509	2.606	5.044	4.267	n.d.	n.d.
G5	5.897	2.639	2.639	5.046	4.275	n.d.	n.d.
T6	4.934	1.724	2.234	4.574	2.474	n.d.	n.d.
A7	5.518	2.222	2.684	4.675	2.952	3.39	n.d.
C8	5.319	1.453	2.292	4.546	2.247	n.d.	n.d.
A9	5.556	2.175	2.564	4.649	2.647	n.d.	n.d.
T10	5.684	1.86	2.292	4.763	3.903	n.d.	n.d.
C11	5.736	1.962	2.354	n.d.	4.115	4.068	n.d.
G12	6.146	2.362	2.594	4.680	4.179	4.074	n.d.
C13	5.741	1.966	2.431	4.725	4.069	3.742	3.742
G14	5.667	2.763	2.888	5.041	4.366	4.004	4.119
A15	6.321	2.699	2.976	5.1	4.505	4.238	4.266

T16	5.619	2.249	2.378	4.924	4.138	4.191	4.251
G17	5.782	2.639	2.639	4.993	4.265	4.038	4.104
T18	5.201	1.702	2.25	4.623	2.044	3.650	3.849
A19	5.5	2.171	2.708	4.668	2.636	n.d.	n.d.
C20	5.331	1.51	2.320	4.557	2.337	n.d.	n.d.
A21	5.446	2.176	2.473	4.634	2.78	3.566	n.d.
T22	5.729	1.838	2.101	4.774	3.747	n.d.	n.d.
C23	5.494	2.093	2.306	4.829	4.120	n.d.	n.d.
G24	6.172	2.398	2.621	4.675	4.167	4.023	n.d.

Appendix 127¹H base chemical shifts of PA4•ODN1.

ODN1	H1	H2	H21	H22	H3	H41	H42	H5	H6	H61	H62	H7	H8
C1	-	-	-	-	-	n.d.	n.d.	5.872	7.586	-	-	-	-
G2	12.89	-	n.d.	n.d.	-	-	-	-	-	-	-	-	8.017
A3	-	7.977	-	-	-	-	-	-	-	n.d.	n.d.	-	8.437
T4	-	-	-	-	13.72	-	-	-	7.425	-	-	1.556	-
G5	12.47	-	8.273	6.593	-	-	-	-	-	-	-	-	8.018
T6	-	-	-	-	12.43	-	-	-	6.952	-	-	1.31	-
A7	-	8.042	-	-	-	-	-	-	-	8.461	n.d.	-	8.12
C8	-	-	-	-	-	8.739	6.407	5.275	6.881	-	-	-	-
A9	-	7.725	-	-	-	-	-	-	-	8.698	n.d.	-	8.289
T10	-	-	-	-	13.53	-	-	-	6.959	-	-	1.44	-
C11	-	-	-	-	-	8.558	6.906	5.613	7.405	-	-	-	-
G12	n.d.-	-	-	-	-	-	-	-	-	-	-	-	7.926
C13	-	-	-	-	-	n.d.	n.d.	5.904	7.645	-	-	-	-
G14	12.82	-	n.d.	n.d.	-	-	-	-	-	-	-	-	8.002
A15	-	7.891	-	-	-	-	-	-	-	n.d.	n.d.	-	8.314
T16	-	-	-	-	13.51	-	-	-	7.272	-	-	1.517	-
G17	12.75	-	8.54	7.259	-	-	-	-	-	-	-	-	8.005
T18	-	-	-	-	12.8	-	-	-	6.972	-	-	1.439	-
A19	-	8.021	-	-	-	-	-	-	-	8.551	n.d.	-	8.142
C20	-	-	-	-	-	8.672	6.489	5.288	6.893	-	-	-	-
A21	-	7.732	-	-	-	-	-	-	-	8.636	n.d.	-	8.285
T22	-	-	-	-	13.84	-	-	-	6.949	-	-	1.349	-
C23	-	-	-	-	-	8.629	6.947	5.648	7.458	-	-	-	-
G24	n.d.	-	n.d.	n.d.	-	-	-	-	-	-	-	-	7.938

Appendix 128 ¹³C ODN1 assigned chemical shifts of PA4•ODN1.

ODN1	C1'	C2	C3'	C4'	C5	C6	C7	C8
C1	85.37	-	75.65	85.76	96.91	140.4	-	-

G2	81.37	-	76.57	84.69	-	-	-	-
A3	81.95	152	77.36	85.12	-	-	-	138.8
T4	83.36	-	75.6	82.91	-	136.3	11.35	-
G5	82.66	-	78.69	85.09	-	-	-	-
T6	81.65	-	73.43	81.15	-	135.8	11.55	-
A7	81.53	152.4	76.6	83.23	-	-	-	139.8
C8	81.57	-	72.91	80.84	96.28	139.9	-	-
A9	80.48	151.5	74.23	82.45	-	-	-	139.4
T10	81.45	-	73.02	81.74	-	n.d.	11.31	-
C11	83.69	-	n.d.	82.8	95.99	141.1	-	-
G12	82.05	-	70.93	85.28	-	-	-	136.3
C13	85.21	-	75.06	n.d.	96.81	140.5	-	-
G14	81.44	-	77.23	84.71	-	-	-	-
A15	82.21	152.3	77.08	84.89	-	-	-	138.5
T16	82.39	-	73.21	82.2	-	136.1	11.38	-
G17	82.65	-	78.25	84.88	-	-	-	-
T18	81.94	-	73.88	81.46	-	136.4	12.16	-
A19	80.86	151.9	75.49	82.7	-	-	-	139.6
C20	81.57	-	73.06	81.03	96.23	139.7	-	-
A21	80.57	150.9	74.06	82.52	-	-	-	-
T22	81.43	-	73.57	82.05	-	-	11.22	-
C23	84.06	-	73.94	83.45	96.17	141.2	-	-
G24	82.21	-	70.67	85.17	-	-	-	136.5

Appendix 129 ¹H PA4 assigned chemical shifts of PA4-ODN1(H2'/H2'', H3'/H3'' and H4'/H4'') were not stereochemically assigned. “'''” and “''''” refer to the most upfield and downfield signal, respectively).

PA4	H2	H2'	H2''	H3	H3'	H3''	H4	H4'	H4''	H5	H6	CH₃	NH
Dp	-	3.134	3.19	-	1.949	2.139	-	2.98	3.626	-	-	n.d.	8.16
βAla	-	1.617	2.055	-	2.889	3.794	-	-	-	-	-	-	9.246
Py1	-	-	-	6.308	-	-	-	-	-	7.035	-	3.56	9.364
Py2	-	-	-	6.403	-	-	-	-	-	7.497	-	3.697	10.73
Py3	-	-	-	6.172	-	-	-	-	-	7.619	-	3.874	9.803
Im4	-	-	-	-	-	-	-	-	-	7.459	-	3.949	10.43
Turn	-	2.102	2.509	-	1.592	2.3	-	2.754	3.445	-	-	-	9.135
Py5	-	-	-	6.438	-	-	-	-	-	7.088	-	3.574	10.54
Py6	-	-	-	6.248	-	-	-	-	-	7.615	-	3.782	9.618
Py7	-	-	-	5.984	-	-	-	-	-	7.654	-	3.923	8.229
Nt8	8.859	-	-	-	-	-	-	-	-	-	4.138	1.287(H7) 1.531(H8)	-

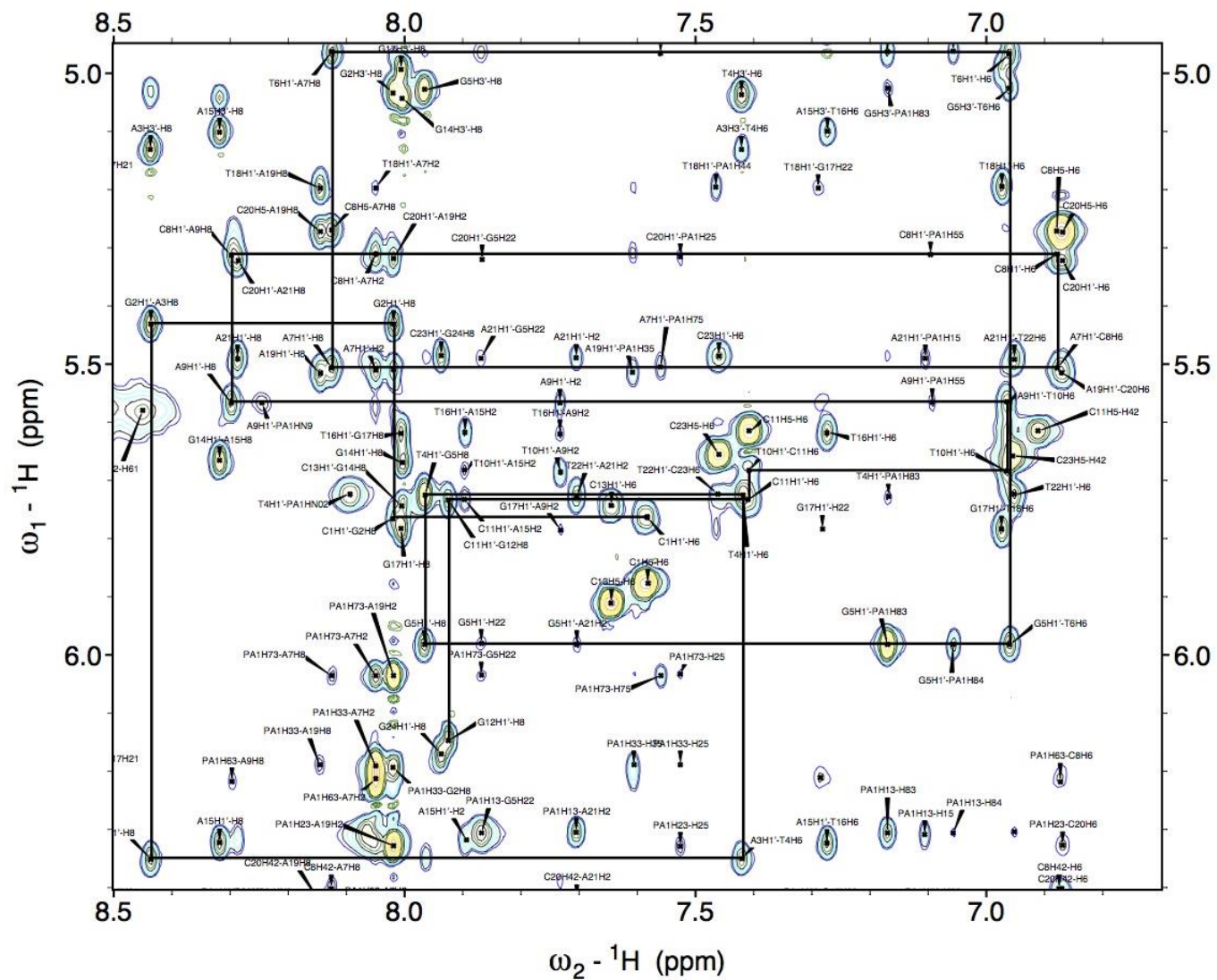
Appendix 130 ¹³C PA4 assigned chemical shifts of PA4-ODN1.

PA4	C2	C3	C4	C5	C6	C7	C8	CM
------------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------

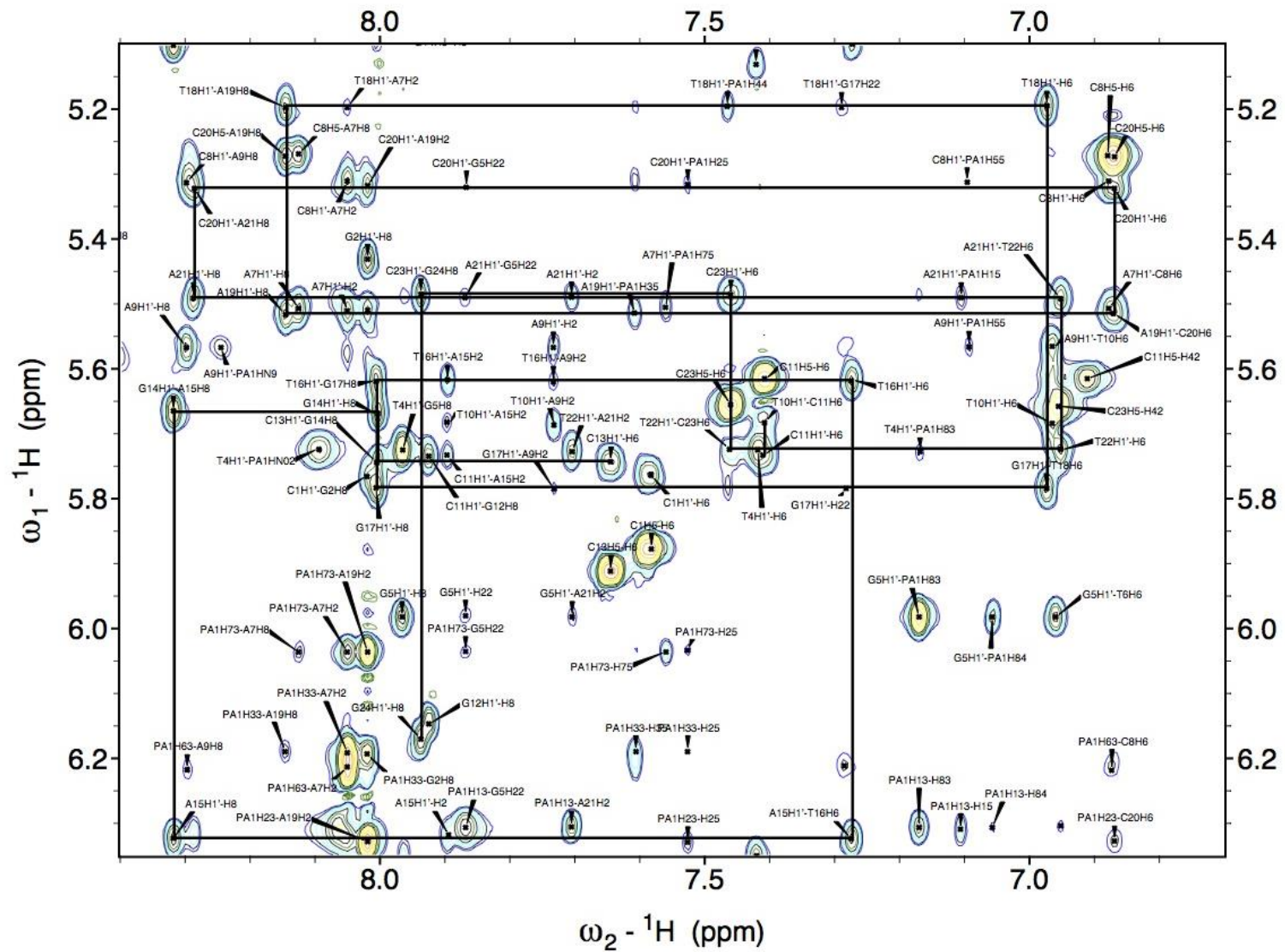
Py1	-	106.6	-	119.9	-	-	-	36.57
Py2	-	102.5	-	121.9	-	-	-	37.06
Py3	-	104.3	-	121.9	-	-	-	37.73
Im4	-	-	-	113.9	-	-	-	36.4
Py5	-	105.3	-	119	-	-	-	36.27
Py6	-	102.6	-	121.8	-	-	-	37.25
Py7	-	104	-	121	-	-	-	37.63
Nt8	151.9	-	-	-	27.6	25.15	24.5	-

Appendix 131 ³¹P assigned chemical shifts **ODN1, PA1·ODN1, PA3·ODN1 and PA4·ODN1.**

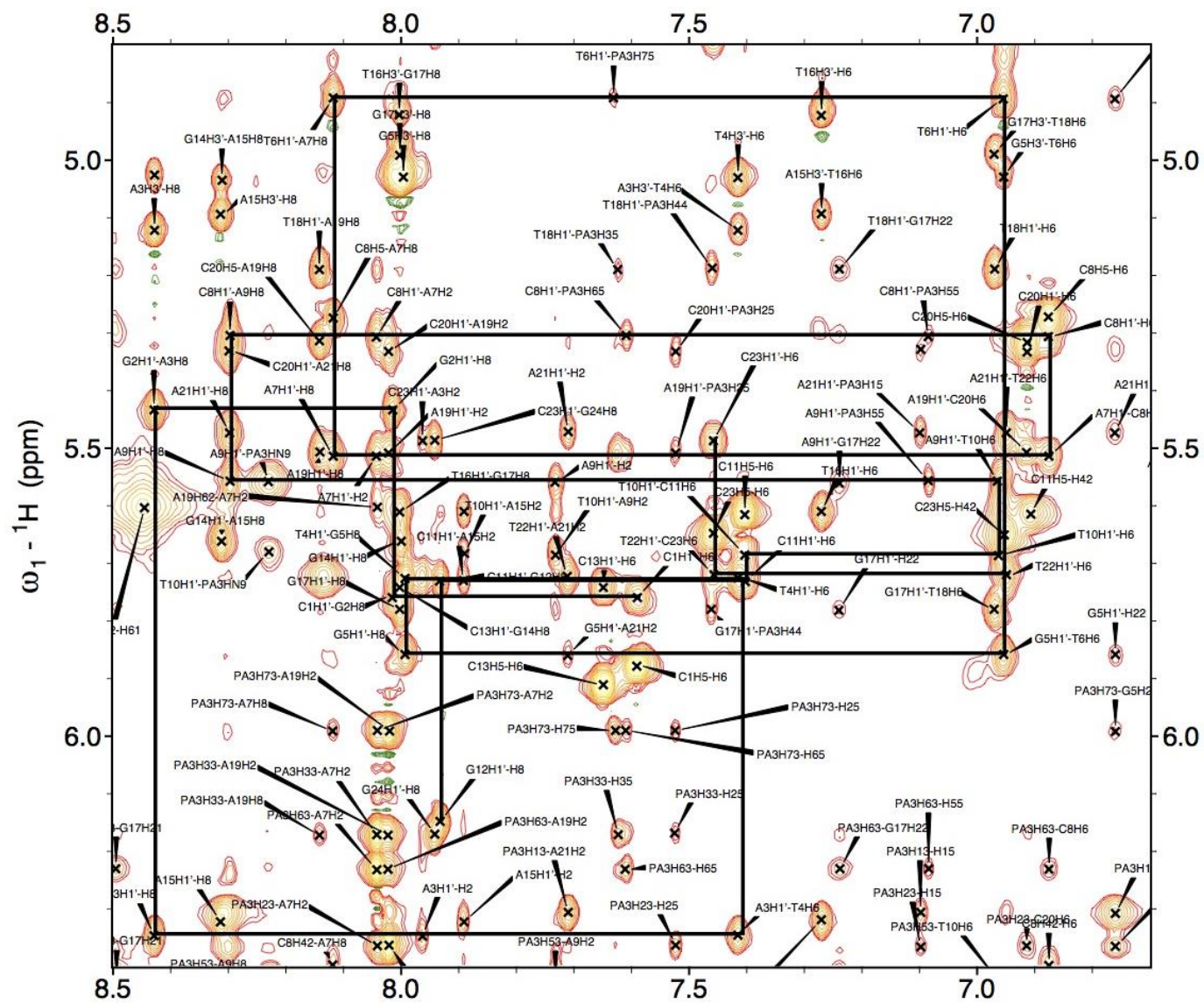
³¹P	ODN1	PA1·ODN1	PA3·ODN1	PA4·ODN1
C1	-	-	-	-
G2	-1.079	-1.087	-1.091	-1.09
A3	-1.082	-0.858	-0.881	-0.869
T4	-1.417	-1.466	-1.478	-1.46
G5	-1.187	-0.364	-0.477	-0.418
T6	-1.401	-1.115	-1.63	-1.032
A7	-1.262	-1.522	-1.64	-1.887
C8	-1.267	-1.856	-1.864	-1.891
A9	-1.027	-1.946	-1.955	-1.94
T10	-1.482	-1.828	-1.853	-1.837
C11	-0.883	-1.182	-0.968	-1.189
G12	n.d.	-0.951	-0.961	-0.964
C13	-	-	-	-
G14	-1.079	-1.087	-1.107	-1.099
A15	-1.082	-1.165	-1.179	-1.176
T16	-1.417	-1.34	-1.348	-1.35
G17	-1.187	-0.477	-0.482	-0.481
T18	-1.401	-1.532	-1.547	-1.562
A19	-1.262	-1.622	-1.636	-1.615
C20	-1.267	-1.95	-1.985	-2.006
A21	-1.027	-1.927	-1.963	-1.945
T22	-1.482	-2.074	-2.085	-2.092
C23	-0.883	-1.349	-1.189	-1.36
G24	n.d.	n.d.	n.d.	n.d.



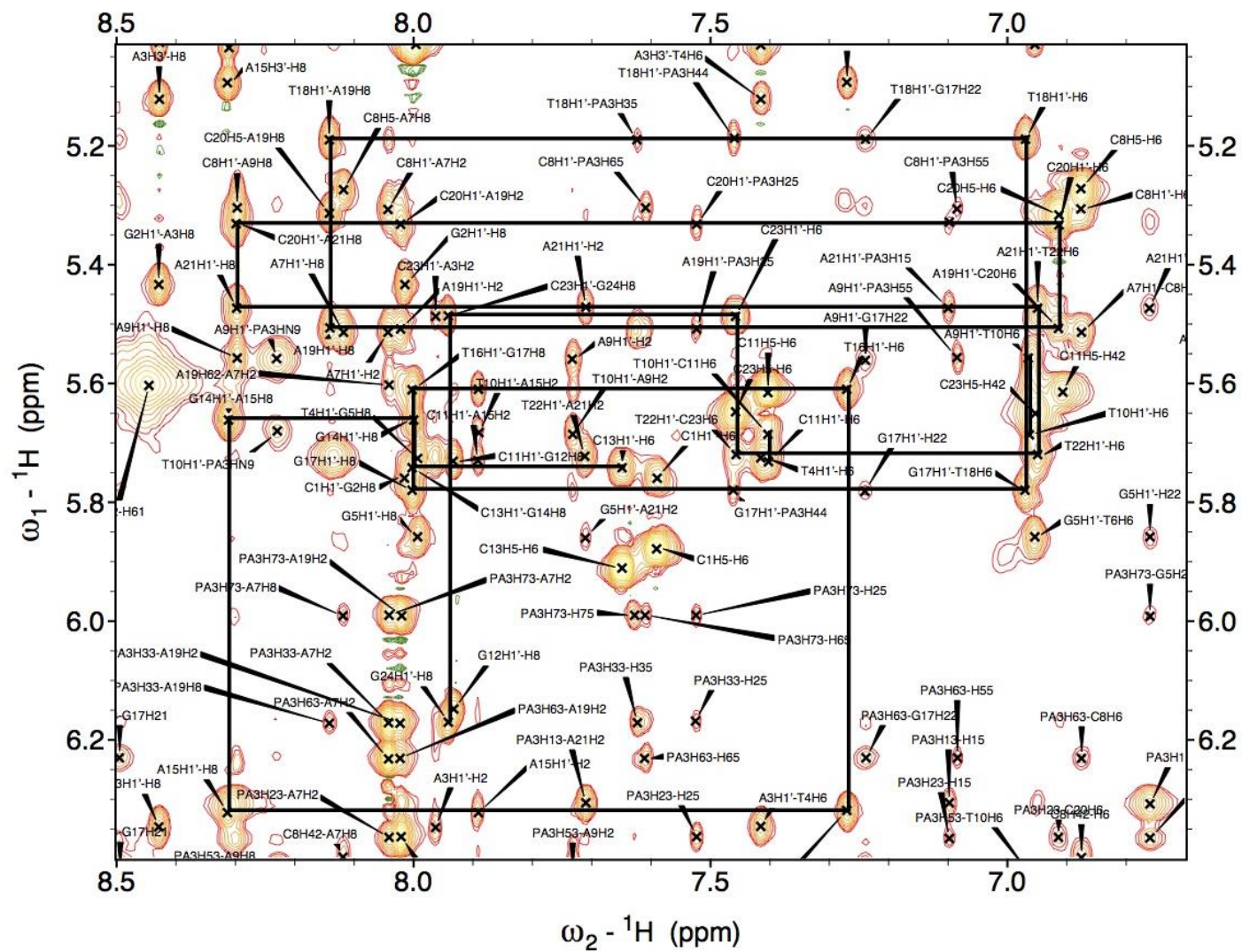
Appendix 132 Depiction of the nOe walk of the strand C1-G12 PA1-ODN1 obtained from NOESY spectrum at 250 ms in 90% H₂O/10% D₂O



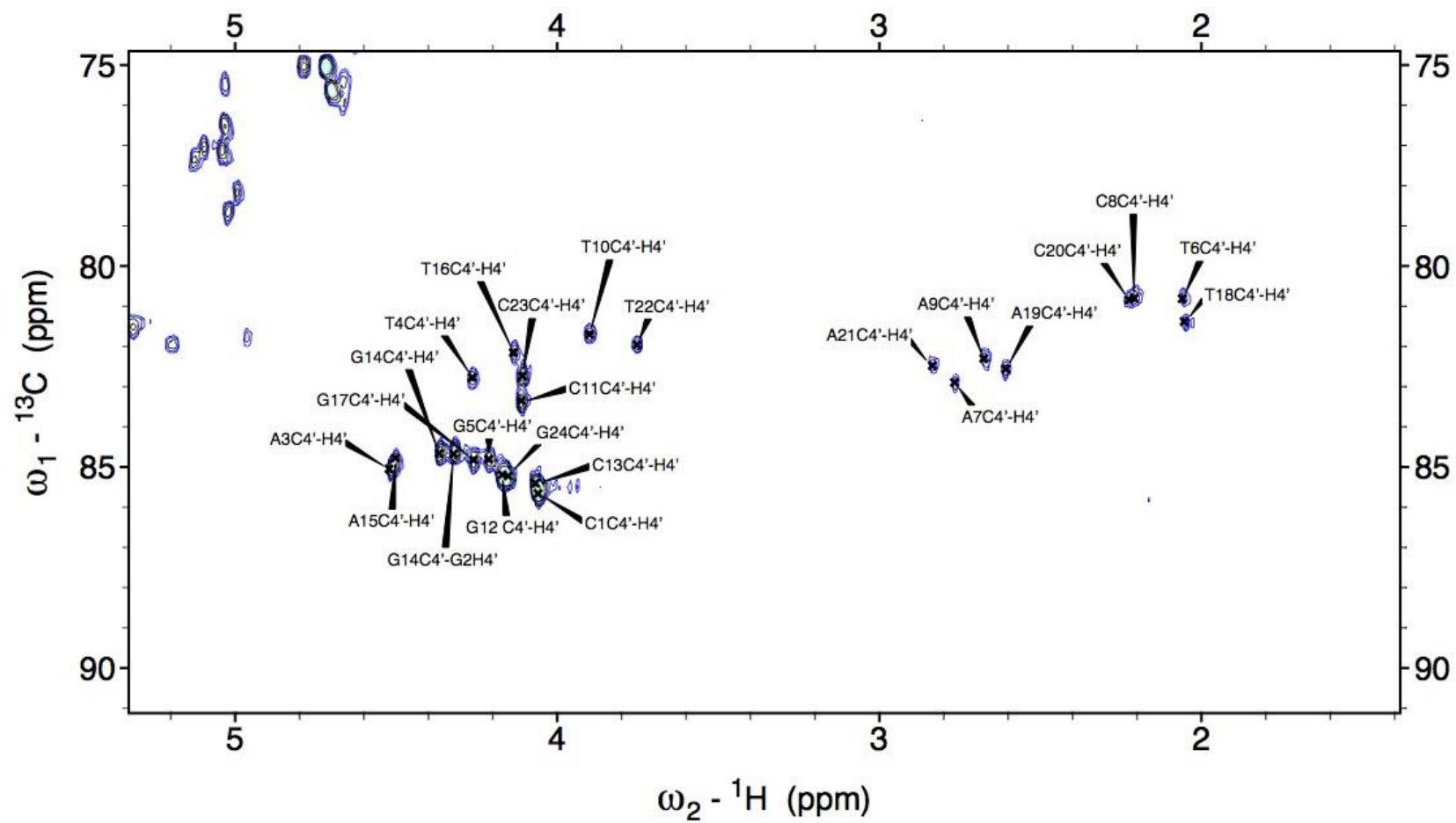
Appendix 133 Depiction of the nOe walk of the strand C13-G24 PA1-ODN1 obtained from NOESY spectrum at 250 ms in 90% H₂O/10% D₂O



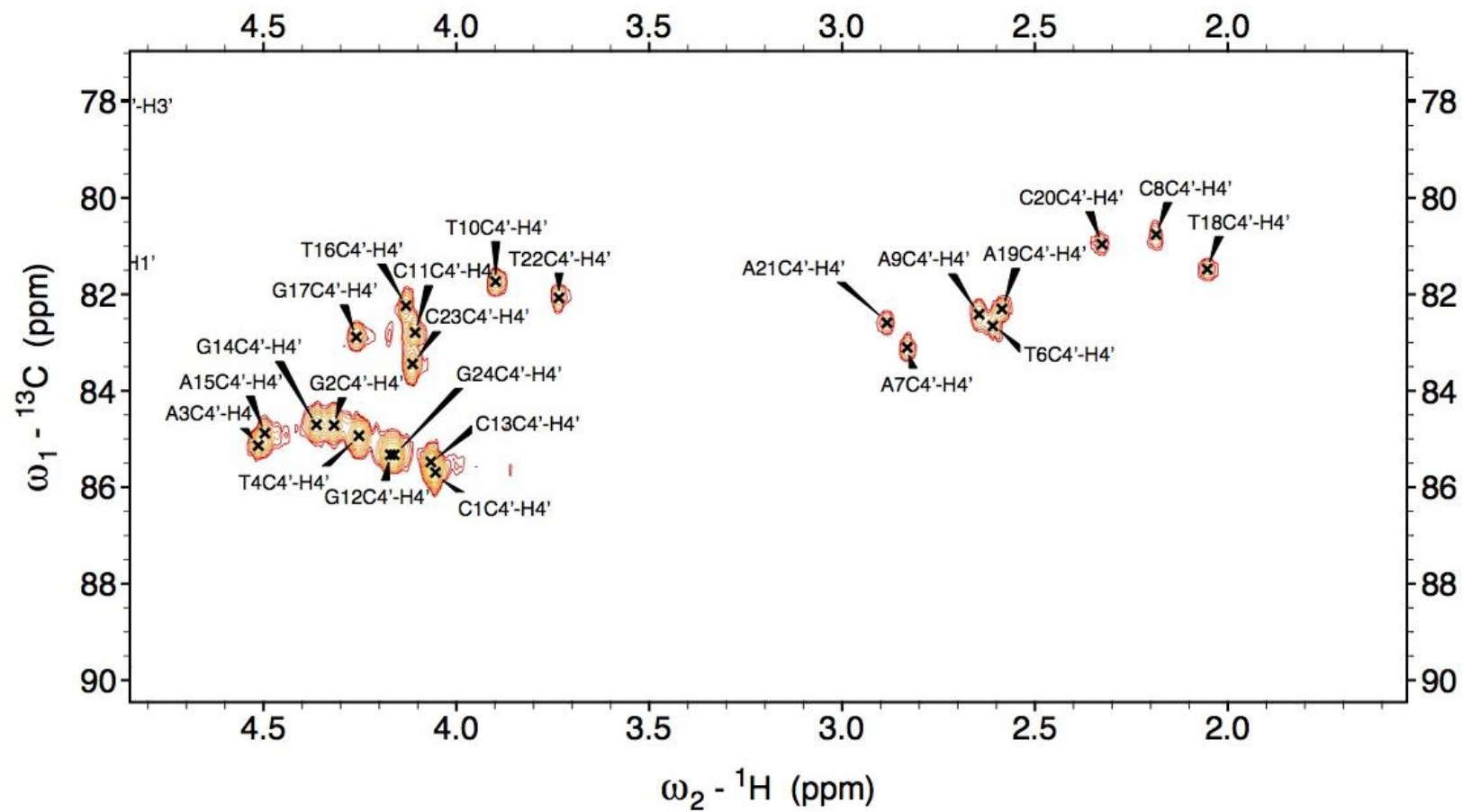
Appendix 134 Depiction of the nOe walk of the strand C1-G12 PA3-ODN1 obtained from NOESY spectrum at 250 ms in 90% H₂O/10% D₂O.



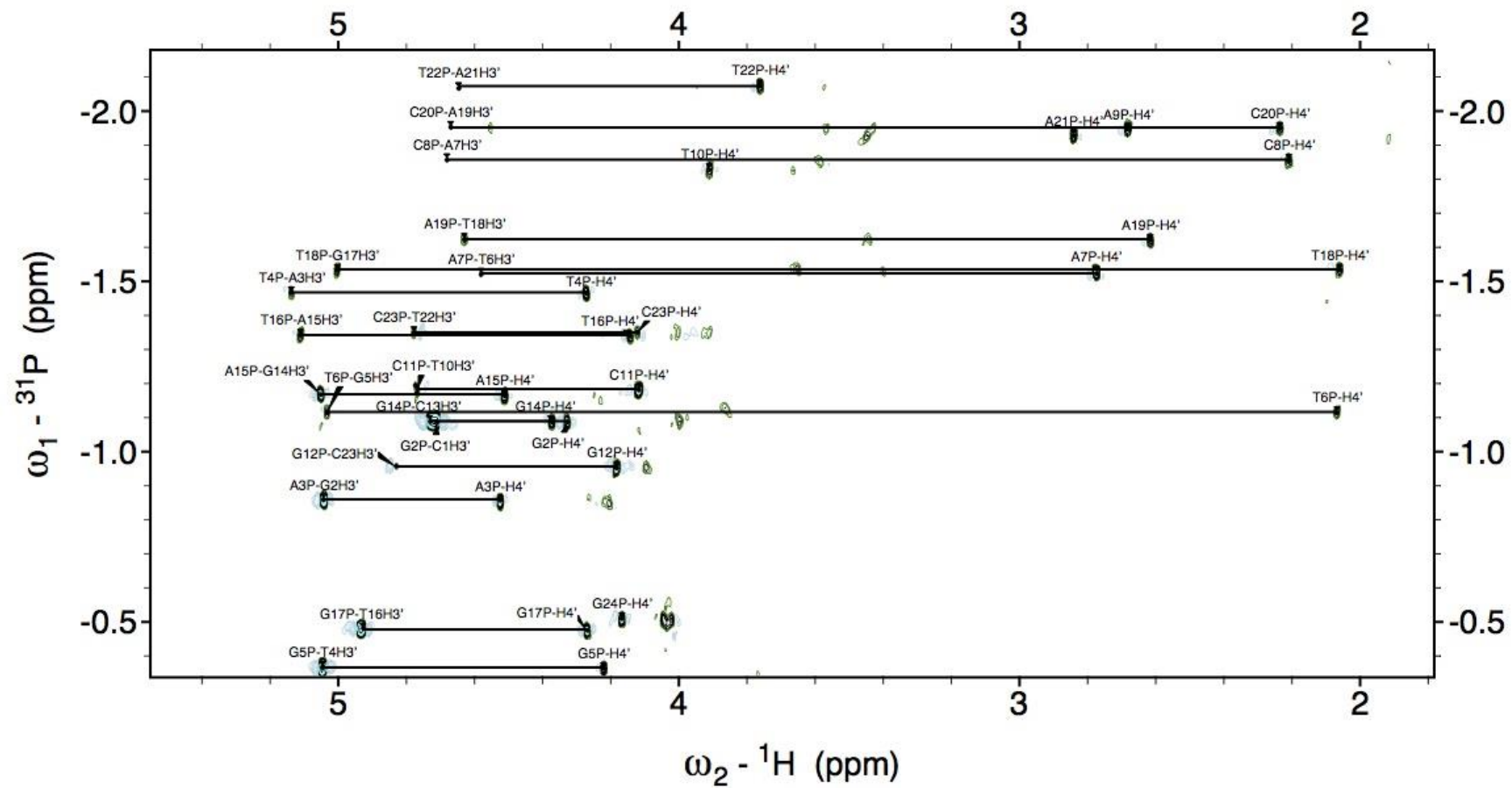
Appendix 135 Depiction of the nOe walk of the strand C13-G24 PA3-ODN1 obtained from NOESY spectrum at 250 ms in 90% H₂O/10% D₂O



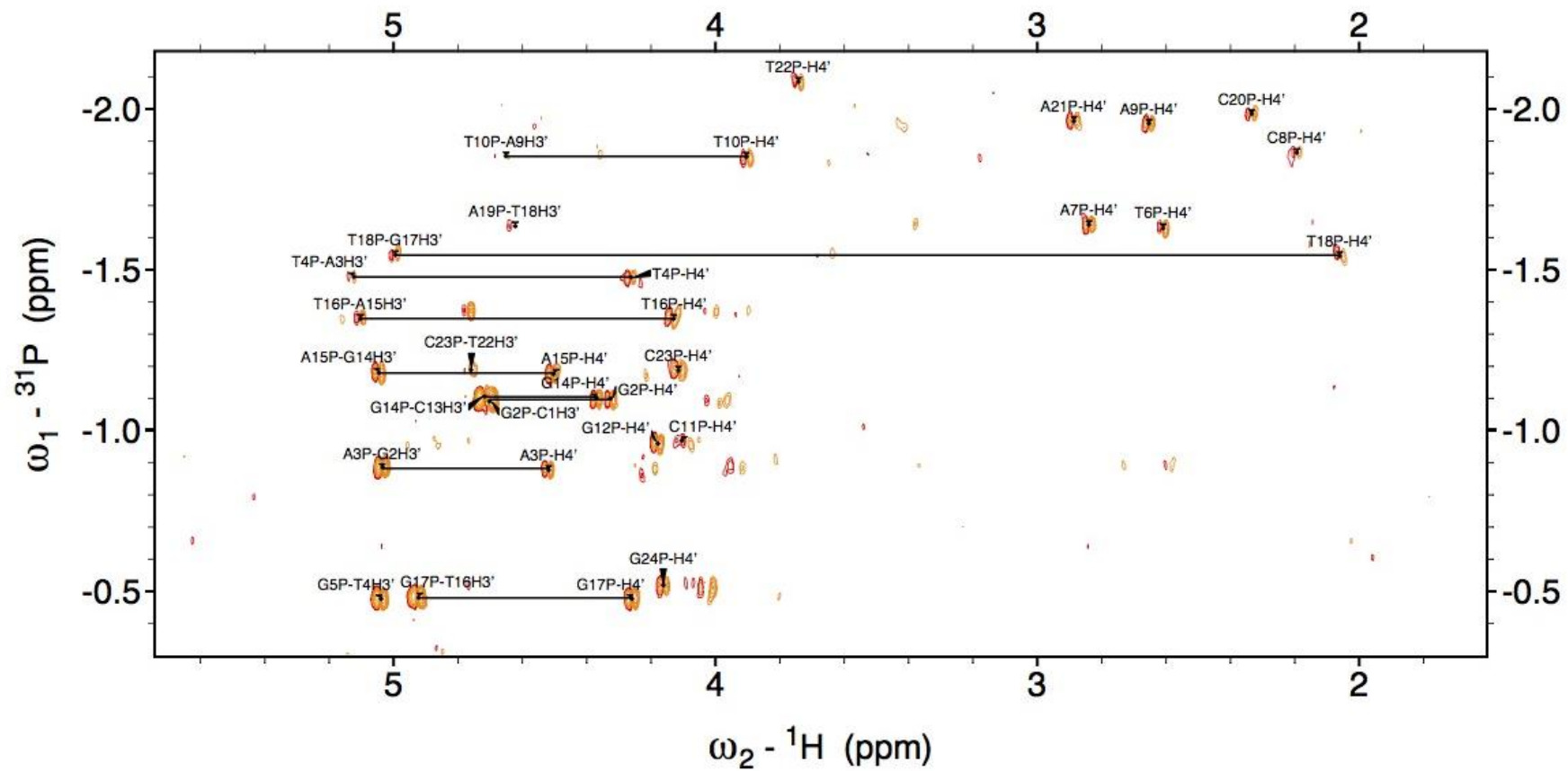
Appendix 136 Selected region of the [^1H , ^{13}C] HSQC of PA1-ODN1.



Appendix 137 Selected region of the [${}^1\text{H}$, ${}^{13}\text{C}$] HSQC of PA3-ODN1.



Appendix 138 [^1H , ^{31}P] COSY spectrum of the PA1-ODN1 complex.



Appendix 139 [^1H , ^{31}P] COSY spectrum of the PA3-ODN1 complex.

4 Analysis of NMR-derived structures

Appendix 140 Minor and Major Grooves width (P-P) calculated from averaged minimised structures using 3DNA.

	Minor Groove				Major Groove			
	ODN1	PA1	PA3	PA4	ODN1	PA1	PA3	PA4
1 CG/CG	---			---	---			---
2 GA/TC	---	---	---	---	---	---	---	---
3 AT/AT	10.1	---	---	14.4	23.2	---	---	17
4 TG/CA	10.9	15.5	13.7	14.5	21.4	22.7	17.3	18.1
5 GT/AC	11.4	14.5	14.9	14	22.4	21	16.2	12.2
6 TA/TA	10.1	14	13.9	13.4	22.9	19	15.1	12.7
7 AC/GT	11	13.9	13.8	13.1	22.7	16.3	17.7	13.6
8 CA/TG	10.5	15.3	14.9	14.5	21.9	18.6	16.4	15.9
9 AT/AT	10.2	16.5	---	16.6	24	22.2	---	20.1
10 TC/GA	---	---	---	---	---	---	---	---
11 CG/CG	---	---		---	---	---		---

Appendix 141 Axis parameters of ODN1 (calculated using Curves+).

BP-Axis	Xdisp	Ydisp	Inclin	Tip	Ax-bend
1) G 2-C 23	-2.52	0.06	5.9	-1.1	---
2) A 3-T 22	-1.7	0.65	10.7	0.6	1.9
3) T 4-A 21	-1.06	-0.01	5.8	-9.8	1.7
4) G 5-C 20	-1.78	-0.65	-2.9	6.6	1.4
5) T 6-A 19	-1.42	-0.82	5	8.1	0.9
6) A 7-T 18	-1.43	0.72	5.6	-8.3	1.4
7) C 8-G 17	-1.96	0.6	-4.1	-7	1.4
8) A 9-T 16	-1.16	-0.03	5.5	9.5	2
9) T 10-A 15	-1.8	-0.62	9.7	-1	1.9
10) C 11-G 14	-2.59	-0.05	6.1	2.2	1.5
Average:	-1.74	-0.02	4.7	0	Total bend 10.8

Appendix 142 Local base-pair parameters of ODN1 (calculated using 3DNA).

bp	Shear	Stretch	Stagger	Buckle	Propeller	Opening
1 C-G	0.09	-0.04	0.09	-0.16	4.63	-2.97
2 G-C	-0.25	-0.01	0.28	21.24	-2.62	-1.11
3 A-T	0.14	-0.03	-0.41	6.85	-24.76	1.1
4 T-A	-0.24	0.01	-0.29	4.62	-22.1	1.48
5 G-C	-0.26	-0.07	0.21	-10.67	-19.55	-0.63
6 T-A	-0.3	0.1	0.22	-12.72	-6.6	-5.27

7 A-T	0.34	0.06	0.08	8.51	-7.13	-3.17
8 C-G	0.24	-0.07	0.21	8.96	-19.19	-1.04
9 A-T	0.28	0.02	-0.36	-7	-21.93	1.72
10 T-A	-0.14	-0.01	-0.37	-7.48	-23.42	1.14
11 C-G	0.24	-0.02	0.27	-20.1	-1.32	-1.56
12 G-C	-0.11	-0.05	0.02	-0.08	5.49	-2.69
	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
<b>ave.</b>	0	-0.01	0	-0.67	-11.54	-1.08
<b>s.d.</b>	0.24	0.05	0.27	11.49	11.43	2.19

**Appendix 143** Local base-pair step parameters of **ODN1** (calculated using 3DNA).

<b>step</b>	<b>Shift</b>	<b>Slide</b>	<b>Rise</b>	<b>Tilt</b>	<b>Roll</b>	<b>Twist</b>
<b>1 CG/CG</b>	-0.26	-0.45	2.77	-3.06	9.88	15.3
<b>2 GA/TC</b>	0.17	-0.55	3.66	5.53	7.44	31.23
<b>3 AT/AT</b>	0.43	-0.67	3.16	-1.19	-2.58	32.08
<b>4 TG/CA</b>	-0.36	-1.15	3.52	-6.82	14.76	24.62
<b>5 GT/AC</b>	0.18	-0.71	3.29	3.47	0.51	28.93
<b>6 TA/TA</b>	0.1	1.34	2.9	1.48	-12.1	38.49
<b>7 AC/GT</b>	-0.25	-0.79	3.23	-4.82	-0.16	29.53
<b>8 CA/TG</b>	0.49	-1.27	3.56	8.12	13.98	24.34
<b>9 AT/AT</b>	-0.46	-0.7	3.14	0.23	-3.34	31.84
<b>10TC/GA</b>	-0.21	-0.54	3.63	-5.15	7.9	29.74
<b>11CG/CG</b>	0.38	-0.44	2.8	3.99	9.07	13.79
	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
ave	0.0	2 -0.54	3.24	0.16	4.12	27.26
s.d	0.3	4 0.68	0.33	4.83	8.3	7.35

Appendix 144 Phosphate backbone angles values of **ODN1** (calculated using 3DNA).

Strand I							
base	alpha	beta	gamma	delta	epsilon	zeta	chi
1 C	---	---	54.6	132.3	-165.6	-86.8	-106
2 G	-69.7	165.8	49.3	134.5	-76.9	150.1	-94.7
3 A	-29.4	94.7	7.8	77.8	41	50.9	-106.7
4 T	-133.9	-169.1	54.5	107.9	-177.7	-87.2	-112.6
5 G	-59	168.7	53.4	108.5	-162.8	-84.7	-133.8
6 T	-72.6	175.4	41.6	126.4	-89.4	155	-100.7
7 A	-59.7	114.3	47.4	139.8	-80.7	164.4	-98.5
8 C	-58	115.8	35.8	106.1	-167.6	-83	-118.1
9 A	-69.3	168.8	42.6	76.6	39.5	52.9	-123.3
10 T	-119.5	-172.7	53.9	131.3	-169.9	-90.2	-112.1
11 C	-80.3	-179.7	45.8	118.9	-160.4	-83.1	-110.6
12 G	-87	162.3	47.6	79	---	---	-107.6

Strand II							
base	alpha	beta	gamma	delta	epsilon	zeta	chi
1 G	-91.3	163	46.8	78.1	---	---	-107.4
2 C	-78.9	-179.7	47.3	120.5	-159	-81.9	-113.6
3 T	-118.2	-174.4	55	131.3	-171	-91	-112.8
4 A	-72	166.7	43.6	77	41.4	51.5	-122.7
5 C	-56.2	113.5	36.6	104.3	-164.9	-83.1	-119
6 A	-58.8	112.6	45.6	138.1	-80.9	161	-96
7 T	-72.5	-179.2	40.2	127.6	-90	157.9	-97.1
8 G	-59.1	168.8	52.1	110.8	-165.1	-86.1	-134.2
9 T	-126.6	-174.2	54.2	106.8	-176.5	-86.2	-115.5
10 A	-33.7	97.1	8	75.7	40.3	51.4	-107.2
11 G	-71.9	164.3	47.8	131.4	-76.9	152.2	-93.3
12 C	---	---	-69.3	147.1	-167.1	-85.4	-97.8

Appendix 145 Axis parameters of **PA1-ODN1**. (calculated using Curves+).

BP-Axis	Xdisp	Ydisp	Inclin	Tip	Ax-bend
1) G 2-C 23	-3.42	-0.12	23.8	-2.8	---
2) A 3-T 22	-3.43	1.11	24.1	-1.2	1.4
3) T 4-A 21	-1.06	0.37	20.5	-6.8	1.4
4) G 5-C 20	-1.72	-0.35	14.5	7.5	1.5
5) T 6-A 19	-2.55	-0.67	17.3	-0.4	1.6
6) A 7-T 18	-2.61	0.46	17.4	3.9	2.4
7) C 8-G 17	-1.7	0.31	18.6	-9.5	2.1
8) A 9-T 16	-0.57	-0.54	21.7	8.6	2.1
9) T 10-A 15	-2.43	-1.62	24.4	0.8	2.5
10) C 11-G 14	-2.93	-1.72	27.9	0.3	2.7
Average:	-2.24	-0.28	21	0.1	Total bend 14.2

Appendix 146 Local base-pair parameters of **PA1-ODN1** (calculated using 3DNA).

Bp	Shear	Stretch	Stagger	Buckle	Propeller	Opening
1 G-C	-0.35	-0.07	-0.02	21.75	12.94	-2.54
2 A-T	0.25	-0.04	-0.65	16.63	-24.07	2.63
3 T-A	0.35	0.3	-0.39	1.57	-30.57	29.08
4 G-C	0.11	-0.01	0.3	2.42	3.26	-0.31
5 T-A	-0.24	0.05	-0.1	12.47	-18.88	-7.75
6 A-T	0.2	0.06	0.03	-1.87	-24.96	-8.56
7 C-G	-0.11	-0.04	0.03	-3.34	3.9	-2.76

8 A-T	-0.42	0.24	-0.17	-9.43	-32.43	22.36
9 T-A	-0.15	0.01	-0.11	-16.17	-26.24	3.7
10 C-G	0.39	-0.06	0.38	-19.52	4.52	-2.62
11 G-C	-0.3	-0.08	-0.05	0.12	-11.21	-0.8
~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
<b>ave.</b>	-0.02	0.03	-0.07	0.42	-13.07	2.95
<b>s.d.</b>	0.29	0.13	0.28	12.88	16.42	11.94

**Appendix 147** Local base-pair step parameters of **PA1-ODN1** (calculated using 3DNA).

<b>step</b>	<b>Shift</b>	<b>Slide</b>	<b>Rise</b>	<b>Tilt</b>	<b>Roll</b>	<b>Twist</b>
<b>1 GA/TC</b>	-0.13	0.08	3.79	1.2	14.62	28.64
<b>2 AT/AT</b>	1.81	-1.15	3.62	1.47	9.02	30.77
<b>3 TG/CA</b>	-0.92	-0.48	3.28	-6.67	19.58	17.46
<b>4 GT/AC</b>	-1.17	-0.33	3.04	2.73	0.33	28.73
<b>5 TA/TA</b>	-0.3	0.6	3.5	-2.53	11.93	37.14
<b>6 AC/GT</b>	0.91	0.24	3.27	0.35	-2.67	33.21
<b>7 CA/TG</b>	1.46	-0.07	3.35	2.84	24.97	21.52
<b>8 AT/AT</b>	-0.74	-0.92	3.38	-2.84	10.6	33.78
<b>9 TC/GA</b>	0.73	-0.63	3.61	2.53	12.4	24
<b>10 CG/CG</b>	-0.06	-0.28	2.93	5.39	8.09	19.35
~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
ave.	0.16	-0.29	3.38	0.45	10.89	27.46
s.d.	1.02	0.53	0.27	3.52	8.15	6.61

Appendix 148 Phosphate backbone angle values of **PA1-ODN1**(calculated using 3DNA).

Strand I							
base	alpha	beta	gamma	delta	epsilon	zeta	chi
1 G	-172.9	179.3	59.7	142.4	-89.9	162.1	-69.4
2 A	-66.3	113.6	51.9	97.3	60.1	51.3	-114.9
3 T	179.9	-159.8	56.3	76.3	-160.5	-69.2	-118.5
4 G	-74.5	173	48.8	74.4	35.5	53.3	-117.1
5 T	-120.4	178.9	62.4	110.4	-166.2	-89.5	-117.1
6 A	-68.9	165.8	57.7	120.3	-175	-90.7	-109.3
7 C	-64.4	168	57.5	123	-166.5	-86.5	-112.8
8 A	-71.3	158.7	57.8	93.8	-175.9	-85.2	-129.5
9 T	-71.2	171	49.8	78.2	-154.3	-62.2	-145.2
10 C	-79.2	178.4	49.7	72.6	48.1	59	-133.5
11 G	-113.1	173.6	48.1	139.8	---	---	-103.6
Strand II							
base	alpha	beta	gamma	delta	epsilon	zeta	chi

1 C	-67	-178.9	61	133.6	-171.7	-90.6	-120.2
2 T	-72.9	172.4	52.8	78.1	-161	-72.7	-146.7
3 A	-66.8	161.5	57.8	85.8	-175.1	-83	-131.8
4 C	-63.2	168.8	55.6	108.1	-170.7	-89	-121
5 A	-72.4	162.8	56.9	113.9	-178.7	-88.6	-108.9
6 T	-62.8	118.3	50.9	109.7	-157.9	-88.3	-121.9
7 G	-73	179.8	49	140.8	-86.1	151.7	-92.8
8 T	-66.3	169.9	55.6	82.6	-139.7	-73.8	-137.4
9 A	-65.3	107.9	48.8	83.9	-174.2	-79.6	-137.6
10 G	-94.9	162.9	41.2	130.1	-89.9	160.7	-80.3
11 C	---	---	54.1	142.9	-140.2	-66.7	-132.9

Appendix 149 Axis parameters of **PA3-ODN1** (calculated using Curves+).

(A) BP-Axis	Xdisp	Ydisp	Inclin	Tip	Ax-bend
1) G 2-C 23	-0.56	0.89	4.7	4.8	---
2) A 3-T 22	-0.6	1.23	10.5	6.9	1.7
3) T 4-A 21	0.39	1.26	8.7	0.1	1.2
4) G 5-C 20	-0.07	0.58	7.5	7.6	1
5) T 6-A 19	-0.51	-0.27	10.7	2.7	1.3
6) A 7-T 18	-0.82	-0.65	14.1	1.7	1.6
7) C 8-G 17	-1.31	-1.11	17	-7.1	1
8) A 9-T 16	-0.75	-0.8	19.3	4.2	0.8
9) T 10-A 15	-0.94	-1.38	18.4	-0.7	1.5
10) C 11-G 14	-2.84	-0.49	20.3	-6.9	1.7
Average:	-0.8	-0.07	13.1	1.3	Total bend 9

Appendix 150 Local base-pair parameters of **PA3-ODN1** (calculated using 3DNA).

bp	Shear	Stretch	Stagger	Buckle	Propeller	Opening
1 G-C	-1.25	-0.03	0.32	7.15	-9.63	9.48
2 A-T	0.17	0.02	-0.22	-2.52	-13.65	6.58
3 T-A	-0.24	-0.01	0.29	-3.82	-11.86	-3.52
4 G-C	0.13	-0.01	0.06	-1.55	-4.96	-0.47
5 T-A	-0.21	0.08	0.04	0.44	-18.69	-9.17
6 A-T	0.33	0.05	0.11	-4.68	-21.39	-4.99
7 C-G	-0.22	0.02	-0.03	1.94	-1.61	-2.13
8 A-T	-0.12	0.22	0.08	-5.57	-32.48	18.72
9 T-A	0.35	0.81	0.6	-10.66	-16.28	28.43

10 C-G	0.76	-0.21	0.14	-4.85	1.29	0.18
~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
<b>ave.</b>	-0.03	0.1	0.14	-2.41	-12.92	4.31
<b>s.d.</b>	0.54	0.27	0.22	4.84	10.03	11.7

**Appendix 151** Local base-pair step parameters of **PA3-ODN1** (calculated using 3DNA).

<b>step</b>	<b>Shift</b>	<b>Slide</b>	<b>Rise</b>	<b>Tilt</b>	<b>Roll</b>	<b>Twist</b>
<b>1 GA/TC</b>	-1.23	0.03	3.69	0.09	8.16	43.54
<b>2 AT/AT</b>	-0.14	0.48	3.48	-6.82	-0.99	33.54
<b>3 TG/CA</b>	-1.31	-0.07	3.13	-3.17	11.41	31.12
<b>4 GT/AC</b>	-0.85	-0.02	3.12	-1.61	1.03	31.32
<b>5 TA/TA</b>	0.41	0.39	3.29	1.05	7.76	36.58
<b>6 AC/GT</b>	0.77	-0.02	3.03	3.63	1.26	31.17
<b>7 CA/TG</b>	1.74	0.51	3.38	3.55	21.14	29.21
<b>8 AT/AT</b>	0.99	-0.44	3.22	-0.66	8.67	30.16
<b>9 TC/GA</b>	-0.76	0.08	3.23	3.36	7.65	33.75
~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
ave.	-0.04	0.1	3.29	-0.06	7.34	33.3
s.d.	1.08	0.31	0.2	3.5	6.67	4.4

Appendix 152 Phosphate backbone angle values of **PA3-ODN1** (calculated using 3DNA).

Strand I							
base	alpha	beta	gamma	delta	epsilon	zeta	chi
1 G	-63.6	169.9	70.5	146.1	175.1	-94.5	-114.8
2 A	-65.7	-146	44.3	143.7	-175.8	-92.3	-89.5
3 T	-68.2	167.5	56.4	113.6	171.6	-98.8	-112.4
4 G	-59.4	-165.2	48.8	136.8	-167.6	-98.7	-103.2
5 T	-71.5	162.7	56.5	95.8	-177.3	-81.3	-126.1
6 A	-67.7	161	61.9	90.8	-176.1	-68.5	-135.4
7 C	-64.6	167.9	68.6	130.9	-160	-83.4	-114.4
8 A	-71	160.1	54.3	108.3	-175.6	-88.9	-114.3
9 T	-64.3	165.2	55.5	79	-171.2	-77.8	-141.8
10 C	-65	173.6	61.2	142.2	-173.9	-82.8	-105
Strand II							
base	alpha	beta	gamma	delta	epsilon	zeta	chi
1 C	-68.1	175.1	53.8	142.5	-173.2	-81.1	-101.3
2 T	-62.4	130.3	59.9	135.8	-172.5	-89.1	-113.7
3 A	-72.6	166.1	47.4	133.1	-121	171.1	-94.6
4 C	-61.6	168.6	57.1	127.6	-165.2	-84.2	-109.5
5 A	-73.9	158.1	62.3	109.8	-178.8	-88.6	-115.3

6 T	-60.7	169.5	64.6	124.8	-156	-93.2	-116.6
7 G	-67.5	-177.5	53.8	126	-174.6	102.5	-102
8 T	-74.7	166.1	51.5	83.3	-166.4	-83.6	-123.6
9 A	-61.9	-170.6	54.2	138.5	-169.5	-91.3	-100.4
10 G	-54.5	-161.2	66.8	146	169.8	-94.1	-101

Appendix 153 Axis parameters of **PA4-ODN1** (calculated using Curves+).

BP-Axis	Xdisp	Ydisp	Inclin	Tip	Ax-bend
1) G 2-C 23	-2.86	1.05	19	1	---
2) A 3-T 22	-2.15	1.3	16.8	3.2	0.6
3) T 4-A 21	-0.52	0.81	16.8	-4.3	0.9
4) G 5-C 20	-0.65	1.02	14.2	3.7	0.9
5) T 6-A 19	-0.58	0.3	11.2	-3.6	1
6) A 7-T 18	-0.62	-0.39	12.7	4.6	2
7) C 8-G 17	-0.74	-1.32	14.7	-5.6	1.4
8) A 9-T 16	0.09	-0.42	19.1	10.5	1.3
9) T 10-A 15	-0.32	-1.63	16.3	1.1	1.2
10) C 11-G 14	-1.29	-1.23	22.1	3.7	1.3
Average:	-0.96	-0.05	16.3	1.4	Total bend 10.2

Appendix 154 Local base-pair parameters of **PA4-ODN1** (calculated using 3DNA).

bp	Shear	Stretch	Stagger	Buckle	Propeller	Opening
1 C-G	0.89	0.02	-0.02	20.13	-38.03	11.29
2 G-C	-0.43	-0.14	-0.21	0.01	-6.54	-1.44
3 A-T	0.07	-0.04	-0.21	0.22	-19.18	1.32
4 T-A	-0.1	0.02	-0.07	2.9	-24.64	7.13
5 G-C	0.12	-0.05	-0.06	3.62	-6.73	-1.3
6 T-A	-0.11	0.07	-0.03	0.94	-23.79	-10.54
7 A-T	0.23	0.15	-0.13	-7.45	-22.97	-11.88
8 C-G	-0.14	-0.01	0	-7.55	-4.56	-2.47
9 A-T	-0.33	0.17	-0.35	-16.85	-32.4	19.13
10 T-A	-0.03	0.25	0.62	-16.53	-11.77	15.87
11 C-G	0.54	-0.16	0.36	-10.41	14.28	-2.36
12 G-C	-0.46	-0.08	0.23	12.01	-10.45	1.45
	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
<b>ave.</b>	0.02	0.02	0.01	-1.58	-15.57	2.18
<b>s.d.</b>	0.39	0.12	0.27	10.96	14.2	9.58

**Appendix 155** Local base-pair step parameters of **PA4-ODN1** (calculated using 3DNA).

step	Shift	Slide	Rise	Tilt	Roll	Twist
<b>1 CG/CG</b>	-1.44	-1.16	3.59	-2.74	23.28	30.8
<b>2 GA/TC</b>	-0.23	-0.68	3.34	-3.01	9.55	26.24
<b>3 AT/AT</b>	0.42	-0.77	3.11	0.94	4.04	31.87
<b>4 TG/CA</b>	-1.43	0.6	3.31	-1.1	18.26	35.26
<b>5 GT/AC</b>	-1.02	0.34	3.15	-3.12	0.97	34.82
<b>6 TA/TA</b>	0.01	0.68	3.34	1.99	12.07	35.7
<b>7 AC/GT</b>	0.99	0.18	3.12	2.39	-1.32	34.88
<b>8 CA/TG</b>	1.83	1.64	3.56	3.11	24.32	33.89
<b>9 AT/AT</b>	0.52	-0.58	3.26	-7.48	4.06	33.27
<b>10 TC/GA</b>	-0.12	0.61	3.39	1.8	10.19	23.85
<b>11 CG/CG</b>	-0.42	0.13	2.88	0.92	0.39	25.52
~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
<b>ave.</b>	-0.08	0.09	3.28	-0.57	9.62	31.4
<b>s.d.</b>	1.01	0.82	0.21	3.23	9.09	4.3

**Appendix 156** Phosphate backbone angle values of **PA4-ODN1** (calculated using 3DNA).

<b>Strand I</b>							
base	alpha	beta	gamma	delta	epsilon	zeta	chi
<b>1 C</b>	---	---	59.5	81.4	-168.5	-80.2	-142.6
<b>2 G</b>	-68.3	-163.9	55.2	141.4	-161.5	-83.7	-116.1
<b>3 A</b>	-80.3	160.1	52.9	85.2	179.7	-78.5	-137.4
<b>4 T</b>	-68.9	164.1	60.4	70	75.8	75	-150.5
<b>5 G</b>	-141.8	-173.6	60.3	151	-146.2	-132.5	-87.9
<b>6 T</b>	-64.9	142.6	56.9	93.6	-174	-84.1	-122.2
<b>7 A</b>	-63.2	169	58.6	126.8	178.5	-92.4	-109.1
<b>8 C</b>	-60.3	173.6	58.4	122.2	-165	-87.5	-114.1
<b>9 A</b>	-67.3	160.4	60.3	129.7	-179.7	-90.4	-105.1
<b>10 T</b>	-70.9	176.3	47.3	74.8	-153.1	-75	-139.8
<b>11 C</b>	-70	169.2	58.8	141.6	-117.5	165.8	-100.5
<b>12 G</b>	-62.7	124.5	55.2	141.4	---	---	-119.9
<b>Strand II</b>							
base	alpha	beta	gamma	delta	epsilon	zeta	chi
<b>1 G</b>	-64	177.4	56.2	137.1	---	---	-127.4
<b>2 C</b>	-68	169.4	64.6	125.4	179.8	-93.1	-120.3
<b>3 T</b>	-74.9	165.8	57.6	84.9	-168.1	-70	-138.8
<b>4 A</b>	-73.5	159.3	61.6	106.9	-176.1	-95.8	-114.7
<b>5 C</b>	-61.1	167.6	61.5	128.4	-162.3	-87	-111.5
<b>6 A</b>	-76.6	161.2	60.8	114.1	178.4	-90.9	-110.3

---

<b>7 T</b>	-60.3	175.3	62	131.5	-159.6	-89.2	-110
<b>8 G</b>	-65.1	-175.5	54.6	140.7	163.1	-95.4	-90
<b>9 T</b>	-63.3	166.4	59.7	85.3	-140.1	-92.2	-132.1
<b>10 A</b>	-64.9	106.9	54.8	84.3	176.1	-85.7	-129.7
<b>11 G</b>	-106.9	167.7	42.5	131	-93.2	159.4	-73.6
<b>12 C</b>	---	---	58.9	147	-139.7	-63	-144.9

---