To Affinity & Beyond: Exploring Sequence Selectivity of Small Molecules for the DNA Minor Groove

Sarah Laughlin-Toth
Georgia State University

Follow this and additional works at: https://scholarworks.gsu.edu/chemistry_diss

Recommended Citation
Laughlin-Toth, Sarah, "To Affinity & Beyond: Exploring Sequence Selectivity of Small Molecules for the DNA Minor Groove." Dissertation, Georgia State University, 2016.
https://scholarworks.gsu.edu/chemistry_diss/127

This Dissertation is brought to you for free and open access by the Department of Chemistry at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Chemistry Dissertations by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.
A common approach to treating many illnesses is through targeting a specific protein. Gene-related illnesses, however, are particularly difficult to treat and targeting the gene with small molecules for drug development is an alternative approach. This is an attractive area because it provides new and potential methods for regulating the expression of diseased-genes. Dicationic diamidines are a class of small molecules which bind in the DNA minor groove where most gene control proteins, such as transcription factors, do not interact. These small molecules can modulate various processes through allosteric interactions and provide therapeutic potential. Typically, these compounds have an inherent selectivity towards sites rich in A·T base pairs (bps). Many synthetic efforts are used to design diamidines specific to target mixed-site sequences (G·C-containing) but
these efforts remain challenging. To overcome this obstacle, our recognition repertoire must be extended to include G·C bps and distinguishability among A·T bp sites.

Many biophysical methods can be used to investigate small molecule-DNA interactions and the development of competition mass spectrometry, in particular, has helped do this. Competition electrospray ionization mass spectrometry (ESI-MS) is a powerful, novel screening technique used to identify biomolecular interactions. This method identifies important information such as stoichiometry, relative binding affinity, and cooperativity and can be used for a number of analyses. It is particularly useful when applied as a competition assay to quickly and accurately pinpoint the preferred target site of a compound among many DNA sequences. Our competition ESI-MS methodology has been used to investigate the sequence specific interaction of diamidines with DNA and discover binding sites for synthetic compounds. Combining ESI-MS with other biophysical techniques has successfully identified patterns of recognition and the selectivity of DNA minor groove binding compounds. With this information, we have developed a detailed understanding of the variations in sequences and their effects on compound recognition. This will ultimately lead to more sequence specific, rationally designed compounds with fewer off-target effects.

INDEX WORDS: DNA, Electrospray ionization mass spectrometry (ESI-MS), Microstructure, Minor groove binder, Minor groove recognition, Molecular dynamics (MD), Sequence specificity, Surface plasmon resonance (SPR)
TO AFFINITY & BEYOND: EXPLORING SEQUENCE SELECTIVITY OF SMALL MOLECULES FOR THE DNA MINOR GROOVE

by

SARAH LAUGHLIN-TOTH

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2016
TO AFFINITY & BEYOND: EXPLORING SEQUENCE SELECTIVITY OF SMALL 
MOLECULES FOR THE DNA MINOR GROOVE

by

SARAH LAUGHLIN-TOTH

Committee Chair: W. David Wilson

Committee: David W. Boykin
Markus W. Germann

Electronic Version Approved:

Office of Graduate Studies
College of Arts and Sciences
Georgia State University
December 2016
DEDICATION

To my husband, Chris:

Thank you for being my biggest supporter. Through many late nights, early mornings, lost weekends and holidays, and through moments of up and down. I would not have been able to achieve this without unconditional love and support from you.

To my mother and stepfather: Tonhja and Layton

To my sister and family: Jess, Ryan, Jacob, Nathan and Mallory

To my other parents: Glenn and Jean

&

To Dad
ACKNOWLEDGEMENTS

To say that I am appreciative for the guidance I have received under David Wilson’s advisement would be an understatement. Patience, words of encouragement, the freedom to think and do for myself, the ability to be taught and to teach others – these are only a few qualities I learned or improved on as a student, but ones I will carry with me wherever I go. I cannot imagine a better group to have joined and to have spent the last five years with. David and Carol Wilson have helped me to become both best best scientist (and person) I can be, and for this I will always be grateful.

This dissertation would not have been possible without Dr. Siming Wang. Her support, patience, kindness, and willingness to teach me was the life force of this project.

Over the years, I have received helpful advice and constructive criticism from my committee members, Drs. David Boykin and Markus Germann, which made each project the absolute best it could be.

The work in Chapter 5 would not have been possible without Dr. Ivaylo Ivanov and E. Kathleen Carter. Their computational insight and enthusiasm revolutionized the way I think about DNA shape and drug design.

Many thangs to: My Chinese sister, Dr. Shuo Wang, and to Dr. Ananya Paul for their years of guidance and friendship, as well as all of the Wilson group members (past and present) from
whom I continue to learn; Discussions about graduate life and how to improve it with Drs. Giovanni Gadda and Donald Hamelberg; Preparation for life as a Ph.D. student comes from Dr. Binghe Wang; Many wonderful collaborations with Dr. Moses Lee of Murdock Trust, Dr. Bruce Armitage of Carnegie Mellon University, Drs. Stephen Neidle and John Hartley of University College London, and Drs. Gregory Poon and David Boykin of Georgia State University; And lastly to Susan Whittington for her focused approach to navigating life.

And finally to my friends, family, and colleagues whom I met along the way because knowing each person has positively affected me and for this I am thankful.
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS ........................................................................................................ v

LIST OF TABLES .................................................................................................................. xv

LIST OF FIGURES ............................................................................................................... xvi

1 INTRODUCTION ............................................................................................................... 1

1.1 DNA and the minor groove .......................................................................................... 1

1.2 Classical minor groove binding agents ...................................................................... 1

1.3 Electrospray ionization mass spectrometry and DNA .............................................. 4

  1.3.1 Background ........................................................................................................... 4

  1.3.2 Small molecule studies by ESI-MS ...................................................................... 5

  1.3.3 Application to other systems .............................................................................. 6

1.4 Protein complexes with nucleic acids by ESI-MS ..................................................... 8

1.5 Purpose of dissertation .............................................................................................. 9

1.6 References ................................................................................................................... 9

2 A NOVEL APPROACH USING ELECTROSPRAY IONIZATION MASS SPECTROMETRY TO STUDY COMPETITIVE BINDING OF SMALL MOLECULES WITH MIXED DNA SEQUENCES ........................................................................................................ 15

  2.1 Abstract ................................................................................................................... 15

  2.2 Key words ................................................................................................................ 16

  2.3 Abbreviations .......................................................................................................... 16
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.4</td>
<td>Introduction</td>
<td>16</td>
</tr>
<tr>
<td>2.5</td>
<td>Results</td>
<td>17</td>
</tr>
<tr>
<td>2.5.1</td>
<td>Testing the experimental conditions and ESI-MS method with known minor groove binders</td>
<td>17</td>
</tr>
<tr>
<td>2.5.2</td>
<td>Using DB293 to test monomer versus dimer binding</td>
<td>19</td>
</tr>
<tr>
<td>2.5.3</td>
<td>DNA and DNA-complexes are stable</td>
<td>20</td>
</tr>
<tr>
<td>2.5.4</td>
<td>Sensitivity of DNA and DNA-complexes</td>
<td>21</td>
</tr>
<tr>
<td>2.6</td>
<td>Discussion</td>
<td>23</td>
</tr>
<tr>
<td>2.7</td>
<td>Conclusion</td>
<td>24</td>
</tr>
<tr>
<td>2.8</td>
<td>Experimental protocol</td>
<td>25</td>
</tr>
<tr>
<td>2.8.1</td>
<td>DNA and compounds</td>
<td>25</td>
</tr>
<tr>
<td>2.8.2</td>
<td>Electrospray ionization mass spectrometry</td>
<td>25</td>
</tr>
<tr>
<td>2.9</td>
<td>References</td>
<td>27</td>
</tr>
<tr>
<td>3</td>
<td>RESOLUTION OF MIXED SITE DNA COMPLEXES WITH DIMER-FORMING MINOR GROOVE BINDERS BY USING ELECTROSPRAY IONIZATION MASS SPECTROMETRY: COMPOUND STRUCTURE AND DNA SEQUENCE EFFECTS</td>
<td>28</td>
</tr>
<tr>
<td>3.1</td>
<td>Abstract</td>
<td>28</td>
</tr>
<tr>
<td>3.2</td>
<td>Key words</td>
<td>29</td>
</tr>
<tr>
<td>3.3</td>
<td>Abbreviations</td>
<td>29</td>
</tr>
<tr>
<td>3.4</td>
<td>Introduction</td>
<td>30</td>
</tr>
</tbody>
</table>
3.5 Compound design ............................................................................................................. 33
3.6 Results .................................................................................................................................. 34

3.6.1 Structural effects on selective recognition using analogues of 1.................. 34

3.6.2 Molecular modeling of the compounds illustrates the effect of structural
conformation on minor groove recognition................................................................. 43

3.6.3 DNA sequence and directionality influence selective recognition............ 45

3.6.4 Additional evidence for DNA complex formation with 1 and analogues by
thermal melting and circular dichroism ................................................................. 47

3.7 Discussion .......................................................................................................................... 50

3.8 Conclusion ......................................................................................................................... 57

3.9 Experimental procedures ............................................................................................... 57

3.9.1 Materials......................................................................................................................... 57

3.9.2 Electrospray ionization mass spectrometry ............................................................. 59

3.9.3 Thermal melting ............................................................................................................ 59

3.9.4 Circular dichroism ....................................................................................................... 59

3.9.5 Molecular modeling .................................................................................................... 60

3.10 References .......................................................................................................................... 60

4 MIXED UP MINOR GROOVE BINDERS: CONVINCING A·T SPECIFIC
COMPOUNDS TO RECOGNIZE A G·C BASE PAIR......................................................... 64

4.1 Abstract ............................................................................................................................. 64
DNA MICROSTRUCTURE INFLUENCES BINDING OF SMALL
MOLECULES DESIGNED TO TARGET MIXED-SITE SEQUENCES

5.1 Abstract ................................................................. 118

5.2 Key words ........................................................................ 119

5.3 Abbreviations ..................................................................... 119

5.4 Introduction ....................................................................... 119

5.5 Results and discussion ..................................................... 122

5.5.1 Competition electrospray ionization mass spectrometry identifies new
interactions ......................................................................... 122

5.5.2 Surface plasmon resonance confirms sequence specific behavior identified
by ESI-MS ........................................................................ 97

5.5.3 Molecular dynamics identifies microstructural differences in the
experimental DNAs ............................................................ 101

5.6 Conclusion ....................................................................... 118

5.7 Materials and methods ...................................................... 118

5.7.1 Compound and dsDNA ................................................... 118

5.7.2 Electrospray ionization mass spectrometry ....................... 119

5.7.3 Molecular dynamics simulations ................................... 120

5.7.4 Biosensor surface plasmon resonance ........................... 124

5.7.5 Competition biosensor-surface plasmon resonance ........... 125

5.8 References ...................................................................... 127
APPENDIX A. MODULATION OF TOPOISOMERASE IIα EXPRESSION AND CHEMOSENSITIVITY THROUGH TARGETED INHIBITION OF NF-κ: DNA BY A DIAMINO p-ANISYL-BENZIMIDAZOLE (Hx) POLYAMIDE

A.1 Abstract .................................................................................................................. 138
A.1.1 Background .......................................................................................................... 138
A.1.2 Methods ............................................................................................................... 138
A.1.3 Results ................................................................................................................. 138
A.1.4 Conclusions ......................................................................................................... 139
A.1.5 General significance ............................................................................................ 139
A.2 Key words ............................................................................................................... 139
A.3 Introduction ............................................................................................................. 139
A.4 Methods and materials .......................................................................................... 143
A.4.1 Synthesis of polyamides ..................................................................................... 143
A.4.2 Thermal denaturation studies ............................................................................. 143
A.4.3 Circular dichroism ............................................................................................... 144
A.4.4 Surface plasmon resonance ............................................................................... 144
A.4.5 DNase I footprinting .......................................................................................... 146
A.4.6 Cell lines and culture conditions ......................................................................... 146
A.4.7 Electrophoretic mobility shift assays ................................................................. 146
A.4.8  Immunofluorescence staining and confocal microscopy.......................... 147
A.4.9  Quantitative real time PCR................................................................. 147
A.4.10 Immunoblotting ................................................................................. 148
A.4.11 Cell viability....................................................................................... 148
A.4.12 Statistical analysis............................................................................... 149
A.5  Results...................................................................................................... 149
A.5.1  DNA binding affinity and sequence selectivity of diamino Hx-polyamides 2
& 3 ............................................................................................................. 149
A.5.2  Binding of diamino HxI*P and HxIP* on the topoIIα promoter.......... 151
A.5.3  Disruption of the NF-γ:ICB2 interaction by diamino Hx-polyamides..... 153
A.5.4  Nuclear uptake of diamino Hx-polyamides ........................................ 153
A.5.5  Biological activity of diamino Hx-polyamides in A549 cells................. 158
A.5.6  HxIP*-mediated chemosensitisation to topo II poisons..................... 158
A.6  Discussion .............................................................................................. 162
A.7  Conclusion ............................................................................................. 167
A.8  Acknowledgements ................................................................................. 168
A.9  References............................................................................................... 168

APPENDIX B. THE SEARCH FOR SELECTIVITY OF RNA G-QUADRUPLEXES BY
SYNTHETIC COMPOUNDS............................................................................. 173
B.1  Abstract .................................................................................................. 173
B.2 Key words .............................................................................................................. 174
B.3 Abbreviations ........................................................................................................ 174
B.4 Introduction .......................................................................................................... 174
B.5 Materials and methods ......................................................................................... 176
  B.5.1 DNA and compounds ....................................................................................... 176
  B.5.2 Electrospray ionization mass spectrometry ....................................................... 178
  B.5.3 Fluorescence polarization spectroscopy ............................................................. 179
B.6 Results and discussion ......................................................................................... 180
  B.6.1 Sequence specific investigations of small molecule interaction with ALS-related quadruplexes ...................................................................................... 180
  B.6.2 Sequence specific investigations of small molecule interaction with human telomere (HTel21) quadruplexes ................................................................. 186
B.7 Conclusions .......................................................................................................... 194
B.8 References .......................................................................................................... 195
LIST OF TABLES

Table 3.1 ΔTm values of dimer-forming compounds with select DNA sequences. .......... 48

Table 3.2 ΔTm values of 1 with ATGA and mutant sequences. .................................. 48

Table 4.1 Thermal melting studies (ΔTm[^a], °C) of the designed heterocyclic amidine compounds with mixed DNA sequences. .................................................. 96

Table 5.1 Comparison of kinetics-fitted and steady-state equilibrium binding constants (K_D x 10^{-9} M). ........................................................................................................ 99

Table 5.2 Mean (M), standard deviation (SD), and 99.9% confidence interval (CI) of the propeller twist over 50,000 frames for the 13 base pair sequences. ................. 112

Table 5.3 Mean (M), standard deviation (SD), and 99.9% confidence interval (CI) of the propeller twist over 50000 frames for the 14 base pair sequences. ......................... 113
LIST OF FIGURES

Figure 1.1 DNA and the minor groove................................................................. 2

Figure 1.2 Small molecule inhibition of protein-DNA interactions.......................... 3

Figure 1.3 Classical minor groove binding compounds........................................... 4

Figure 1.4 X-ray crystal structure of DB75 with 5'-AATT-3'..................................... 4

Figure 1.5 Cartoon of electrospray ionization of DNA.............................................. 6

Figure 1.6 Adjustment of DNAs to achieve different molecular weights for ESI-MS........ 7

Figure 2.1 Spectra of mixed DNA sequences forming complexes with DB75 and Netropsin. ........................................................................................................................................ 18

Figure 2.2 Using DB293 to test monomer versus dimer binding.............................. 20

Figure 2.3 Peak comparison of R1, ATATAT, AAATTT, and Net-DNA complexes........ 22

Figure 2.4 Raw, convoluted data to observe competitive binding of Net with R1 and AAATTT............................................................................................................................... 23

Figure 3.1 Structures of dicationic diamidine minor groove binding compounds used to investigate dimer formation in mixed-sequence sites.................................................. 31

Figure 3.2 Mixed hairpin DNA sequences used to screen interactions for monomer and dimer-forming complex interactions with multiple sequences................................. 36

Figure 3.3 Example ESI-MS spectra of compound 1 titrated with multiple DNA sequences. ........................................................................................................................................ 37

Figure 3.4 Spectra of DNA sequences titrated with analogues 3 - 8............................. 39

Figure 3.5 Mixed DNA sequence results with 3......................................................... 41

Figure 3.6 Comparison of compound 1 and analogues 3, 4, 5, and 6......................... 43

Figure 3.7 Models of compound 3 recognizing the mixed sequences ATGA as a dimer..... 44
Figure 3.8 ATGA cognate and sequence variants used to examine the sequence specificity of 1. ................................................................. 46
Figure 3.9 Spectra of ATGA cognate and ATGA mutant sequences with compound 1. ...... 46
Figure 3.10 Comparison of relative abundances and thermal melting................................. 49
Figure 3.11 Comparison of ATGA and ATGA-mutant sequence CD spectra......................... 51
Figure 3.12 Spectrum of DNA and 1 run using 2.5 μM concentration of DNA...................... 56
Figure 4.1 Representative minor groove binding compounds and test sequences................. 94
Figure 4.2 UV melting profiles of DNA with minor groove binding compounds. ................. 96
Figure 4.3 Surface plasmon resonance sensorgrams for DB2120. ..................................... 98
Figure 4.4 Competition mass spectrometry........................................................................ 100
Figure 4.5 Circular dichroism titration spectra of DB2120 with A4GT4.............................. 101
Figure 4.6 DNase I footprinting densitometry analysis....................................................... 102
Figure 4.7 DNase I footprinting experiments...................................................................... 103
Figure 4.8 Molecular docking model of DB2120 with the A4GT4 sequence......................... 105
Figure 5.1 Test compound and sequences for ESI-MS....................................................... 121
Figure 5.2 Representative ESI-MS spectra of DB2277 binding with DNA......................... 94
Figure 5.3 Comparison of deconvoluted peak intensities from titrated DB2277 with DNA. 95
Figure 5.4 SPR sensorgram and binding curves of DB2277 and DNA............................... 98
Figure 5.5 2D contour histograms for minor groove widths of unbound DNA. ................. 102
Figure 5.6 2D histograms of most probable minor groove widths for unbound DNA. ....... 104
Figure 5.7 2D histograms of DNA minor groove depths for unbound DNA. ..................... 106
Figure 5.8 2D histograms of minor groove width for AAAAGTTTT................................. 107
Figure 5.9 2D histograms of minor groove width for AATTGCAATT.............................. 108
Figure 5.10 2D histograms of major groove widths for DNA sequences. .......................... 109
Figure 5.11 2D histograms of major groove depths for DNA sequences. .......................... 110
Figure 5.12 Most probable propeller twist per base pair. .................................................. 111
Figure 5.13 Average helical base pair step parameters for one and two G·C base pair sequences............................................................................................................. 115
Figure 5.14 2D contour histograms of minor groove width for 1:1, DB2277-DNA complexes............................................................................................................................ 117
Figure 5.15 Cartoon representations of groove widths. ....................................................... 122
1 INTRODUCTION

1.1 DNA and the minor groove

The central dogma of molecular biology is the simplest explanation of how genetic information is passed on. Simply put, DNA is transcribed to RNA which is then translated to protein. Within this process, however, are many intricacies. DNA – short for deoxyribonucleic acid – is the carrier of genetic information but is itself a highly complex system. For example, Figure 1.1 illustrates a canonical B-form DNA double-helix strand. The DNA is composed of two complementary strands of nucleotides joined together by hydrogen bond interactions. Nucleotide sequences are read from the O5’ to O3’ direction and read as 5’ to 3’ (Figure 1.1C). Due to the specific configuration of the helical duplex, two particular formations occur which are referred to as the major and minor grooves (Figures 1.1B & C). The major groove is often found as having a deep and wide groove which is where many proteins bind, or interact, with the DNA. On the other hand, the minor groove is often identified as having a deep and narrow groove. Figure 1.1C shows two examples of thymine-adenine (T·A) and cytosine-guanine (C·G) complementary base pairs and the locality of the major and minor grooves per base pair. For G·C base pairs, three hydrogen bond interactions occur. The amino group of C and keto group of G extend into the major groove while the keto group of C and amino groups of G point into the minor groove. For A·T base pairs, only two hydrogen bond interactions are formed. Here, the methyl group of T and keto group of A protrude into the major groove while a second keto group of T sits in the minor groove.

1.2 Classical minor groove binding agents

The minor groove of DNA has become an increasingly interesting area of focus in the drug design community. Many DNA-binding proteins target the major groove leaving the minor groove exposed. An exposed minor groove opens the prospect to using small molecules to target the minor
groove and modulate genetic processes. Several cartoon examples are shown in Figure 1.2. The top example, labeled as 1a and 1b illustrates the uses of a small molecule acting as an inhibitor. First, the small molecule binds in the minor groove (1a) and in doing so causes some type of allosteric or conformational change in the DNA shape or structure (1b). A second possible inhibition pathway follows the same DNA sequence except here, the protein is able to bind first with the DNA (2a). In this scenario, the protein interacts with the major groove of the DNA leaving the backside of minor groove accessible to a small molecule. Here, (2b) the small molecule binds with DNA and causes the protein to dissociate (2c) via allosteric interactions. A third possible approach, not illustrated in Figure 1.2, also involves the enhancement or upregulation of DNA-protein interactions through small molecule binding.

A classical minor groove binding compound has several key characteristics and three examples are shown in Figure 1.3. Both Netropsin and Distamycin are naturally occurring polyamides derived from bacteria while DB75 [1] is a synthetic dicationic diamidine made by Dr.
David W. Boykin and co-workers at Georgia State University. The shape, length, and charge of a minor groove binding molecule are critical for binding interactions. As shown with the three examples below, a crescent shape, end positive charges, and hydrogen bond donor/acceptor groups are all crucial for high affinity binding with DNA. Positive charges on the compound are also important as it allows for electrostatic interactions with the negative phosphate backbone of the DNA which helps stabilize binding. Likewise, hydrogen bond donor and acceptor groups on the compound with the bases increase binding affinity and specificity. Evidence of minor groove binding is shown in Figure 1.4 with a crystal structure of DB75 in a 5’-AATT-3’ binding site.[2] This example clearly demonstrates not only hydrogen bond and electrostatic interactions, but also a shape complementary to the minor groove. The length of the small molecule is equally important. For instance, if the compound is too short it may lose specificity; too long or if bulky substituents are present, it can also reduce specificity through increasing insolubility and/or obstruction into the minor groove.
1.3 Electrospray ionization mass spectrometry and DNA

1.3.1 Background

The continued development of electrospray ionization (ESI) methods and the extensive improvements in commercial mass spectrometry (MS) instruments over the last two decades have brought ESI-MS experiments into the forefront of analysis of bio-macromolecules and their complexes. The majority of ESI-MS reports on biological systems to date have involved proteins.
but nucleic acids are attracting increasing attention. Early studies on double-helical DNA involved establishing conditions for preserving the duplex in the gas phase and evaluating different volatile solution buffers/salts for optimum ESI-MS conditions.[3] There are numerous reports on duplex stability [3-5], dissociation to single strands [6-8], effects of ESI conditions [9, 10], solution composition [11-14] and nucleic acid sequence [7,15]. These key studies and more have provided a strong foundation for ESI experiments on nucleic acid-small molecule complexes.

1.3.2 Small molecule studies by ESI-MS

The early studies on ESI-MS requirements for stable duplex [3, 16] quickly evolved into important studies on DNA complexes with metal ions [7, 17], organic compounds [18, 19] and proteins [20, 21]. Studies of organic systems complement extensive solution biophysical studies that have two important goals: (i) develop a better understanding of the fundamental features of nucleic acid interactions and (ii) design nucleic acid-targeting agents for biotechnology and therapeutics. Compounds that bind in the DNA minor groove have a variety of structural features that affect their affinity, stoichiometry and sequence specificity. All of these features, as well as cooperativity for compounds with a stoichiometry greater than one, can be investigated by ESI-MS methods. The earliest reports of DNA complexes by ESI-MS were with the polyamide minor groove binder, Distamycin A.[16, 22] The complex was intact in the gas phase and gave a 1:1 binding stoichiometry for a 12 base pair duplex that had an AAATTT base sequence binding site at low ratios of Distamycin to DNA complex. As the ratio of Distamycin to duplex was increased, a 2:1 bound species was observed and these results are in agreement with solution experimental findings.[23-25] The excellent agreement between species present and their ratio dependencies was a significant example that nucleic acid complex stoichiometry, cooperativity and relative affinities could be determined by ESI-MS.
Figure 1.5 Cartoon of electrospray ionization of DNA.
Droplets of DNA in solution are ejected from a capillary at high voltage towards a counter electrode. The electric potential causes the droplets to disperse into smaller particles until the solvent evaporates. The particles will then be carried toward a mass analyzer where they will be separated by a mass-over-charge ratio (m/z).

The successful, initial studies of small molecule-DNA complexes moved into ESI-MS experiments with a well-known variety of agents and a wide variety of minor groove binders of quite different structure.[26, 27] All of the results with minor groove binders gave excellent agreement with the solution studies. It is now quite clear that ESI-MS experiments, when properly conducted, will provide complementary and very useful results for analysis of DNA-small molecule complexes. The method is quite versatile and experiments can be rapidly conducted so that it is an attractive addition to other powerful, biophysical approaches for DNA complex analysis. The ability to quickly analyze non-covalent complexes, with good sensitivity and low sample consumption, as well as the variety of information provided, makes ESI-MS a very valuable tool.

1.3.3 Application to other systems

ESI-MS is often used for characterizing small, organic molecules but has become a powerful tool for large biomolecular systems. For small molecule-DNA investigations, excellent consistency using ESI-MS and other biophysical methods has been found. Competition ESI-MS is particularly appealing since multiple interactions between ligand and DNA can be analyzed.
Figure 1.6 Adjustment of DNAs to achieve different molecular weights for ESI-MS. Target base pair sites in the stem of the DNA are preserved. (A) DNA with unadjusted molecular weights; (B) complex and unbound DNA peaks are not distinguishable upon addition of ligand due to overlapping peaks for free DNA and DNA-ligand complexes (e.g., black and red); (C) modifications in the hairpin loop by incorporation of various bases allows the DNA stem to be preserved while creating distinguishable molecular weights; (D) complexes and free DNA become easily identifiable. $T =$ thymidine, $C =$ cytidine, $U =$ deoxyuridine.

Simultaneously. Analyzing interactions between a single DNA and a single ligand is not efficient for screening a library of compounds. With a competition method, a large number of interactions can be studied in much less time than with the conventional approach of one DNA and one ligand. Valuable information is gathered quickly with regards to preferential binding of a ligand to DNA. Although several of the studies described later will focus primarily on the binding of small molecule ligands to DNA, this technique is not limited to these specific interactions and can be applied to other biomolecules of interest, including proteins [4, 20, 28-31], carbohydrates [32, 33] and other types of nucleic acids such as RNA [34-44] and peptide nucleic acids [45-47].
1.4 Protein complexes with nucleic acids by ESI-MS

The ETS family of transcription factors (TF) comprises a major class of transcriptional regulators across many species, including humans.[48] Humans also express various oncogenic mutations of the ETS TFs that are associated, for example, with bone, breast, and prostate tumors.[49, 50] All ETS TFs have similar DNA binding domains that are highly conserved in structure with a 5’-GGAA/T-3’ consensus central binding site.[51] Genomic analyses have identified the ETS member PU.1 as a pioneering transcription factor [52] that can overcome chromatin packaging to bind chromosomal DNA. Because of the special properties of PU.1, it is important to understand how DNA recognition by PU.1 is differentiated from other ETS proteins. There is, thus, an essential need for a broad range of methods and studies, including ESI-MS, to address the physical mechanisms of sequence recognition by the PU.1 TF.

Targeting TF-DNA complexes, either for inhibition or enhancement, is very attractive for the treatment of a number of different diseases. This could be done by targeting the TF or the DNA binding domain. Unfortunately, it has proved very difficult to target TFs and they are frequently referred to as “undruggable”.[53, 54] An alternative is to target the DNA binding domain of the TF with designed small molecules and this approach is gaining increasing attention. ESI-MS has developed into a very attractive method to evaluate both small molecule binding to specific DNA sequences as well as for their effects on TF-DNA complexes. The wealth of important obtainable information illustrates the power of ESI-MS in both the characterization of TF-DNA complexes and in the discovery and development of TF inhibitors. Although ESI-MS studies of protein-DNA complexes are relatively rare, a number of studies have appeared in recent years and it is likely that many more are in progress.
1.5 Purpose of dissertation

Much of the work described in this dissertation focuses on extending the minor groove recognition repertoire through sequences with additional G·C bps and distinguishability among variable A·T bp regions. With several minor groove binder and DNA interaction model systems, a novel screening method using ESI-MS was developed to identify new DNA-small molecule interactions. Beginning with well-characterized minor groove binders with known DNAs, ESI-MS conditions were varied to find optimal conditions for detecting these systems. Once optimal conditions were chosen, sequences were strategically designed to not only interact with minor groove binding compounds, but also to be detectable by ESI-MS (details describing the design will be discussed in the following chapter). This ESI-MS screening method was then applied to the discovery of new target binding sites for newly synthesized compounds and/or in expanding of target binding sites for existing compounds. A combination of other biophysical techniques were used to validate and/or complement results obtained by ESI-MS including thermal melting ($T_m$), circular dichroism (CD), molecular dynamics (MD), and surface plasmon resonance (SPR). Our ESI-MS method has been applied to the study of DNA-protein complexes, RNA-small molecule interactions, quadruplex-DNA complexes and DNA-aminoglycoside interactions; however, only a portion of our findings will be discussed in the remaining chapters.

1.6 References


2 A NOVEL APPROACH USING ELECTROSPRAY IONIZATION MASS SPECTROMETRY TO STUDY COMPETITIVE BINDING OF SMALL MOLECULES WITH MIXED DNA SEQUENCES


Copyright © 2014 Analytical & Bioanalytical Chemistry

My contribution to this chapter was sample preparation, spectral analyses, and writing.

2.1 Abstract

Minor groove binding compounds have been shown to induce changes in global DNA conformation, allosterically inhibiting DNA-protein interactions necessary for transcriptional processes. Many minor groove binders are specific for A·T base pairs but have little preference over alternating A/T or A-tract sequences. Few compounds, other than polyamides, show selectivity for mixed sequences with A·T and G·C base pairs. Electrospray ionization mass spectrometry (ESI-MS) can provide insight on the stoichiometry and relative affinities in minor
groove recognition of different DNA sequences with a library of minor groove binders. A goal in our current research is to develop new compounds that recognize mixed sequences of DNA. In an effort to optimize screening for compounds that target mixed A·T and G·C base pair sequences of DNA, ESI-MS was used to study the competitive binding of compounds with a mixed set of DNA sequences. The method identified preferred binding sites, relative affinities, and concentration-dependent binding stoichiometry for the minor groove binding compounds Netropsin and DB75 with A·T-rich sequences, and DB293 with ATGA and AT sites.

2.2 Key words
Electrospray ionization mass spectrometry; Minor groove binders; Minor groove recognition; Mixed DNA sequences, Selectivity

2.3 Abbreviations
Base pairs (bp); Deoxyribonucleic acid (DNA); Dissociation equilibrium constant (K_D); Electrospray ionization mass spectrometry (ESI-MS); Mass-over-charge ratio (m/z); Minor groove (MG); Netropsin (Net); Surface plasmon resonance (SPR)

2.4 Introduction
Non-polyamide minor groove (MG) binders target A·T-rich sites with variable distinction among A·T sequences.[1] Netropsin (Net) and DB75 (Figure 2.1) are MG binding compounds which bind with 1:1 stoichiometry to A·T-rich sites.[2, 3] Polyamides can selectively target G·C base pairs (bps) in a DNA sequence, but few non-polyamide MG binders target mixed sequence sites. DB293 (Figure 2.2) is the first dicationic diamidine to strongly recognize a sequence with mixed bp.[4] It binds in the MG of ATGA sequences as an antiparallel stacked dimer with positive cooperativity. Using the known interactions of Net, DB75, and DB293 as reference points, we are developing a mixed sequence method to screen DNA-MG binders using electrospray ionization
spectrometry (ESI-MS). Because of its improved versatility, ESI-MS has widely increased its utility for studying biomacromolecules and is ideal for characterizing systems with non-covalent interactions. We report a method that provides rapid screening of MG compound libraries for interactions, which can be easily characterized based on differences in structure and DNA sequence. To the best of our knowledge, this is the first example using ESI-MS to screen the competitive interaction of ligands with multiple DNA sequences in a single sample. Using this approach, we can begin to define a basis for specificity of drug binding at a target sequence. With enough DNAs, this method can screen target sites and complexes of interest. Information including relative binding affinity, stoichiometry, and binding cooperativity can be determined.

2.5 Results

2.5.1 Testing the experimental conditions and ESI-MS method with known minor groove binders

Net is a well-understood MG binder which forms high-affinity 1:1 complexes with the A·T-rich sequences selected for our experiments.[2] Figure 2.1B is a deconvoluted spectrum of AAATTT and R1 sequences at a molar ratio of [0:1] Net:DNA, and both DNA sequences have single peaks with similar intensity and minimal background noise. Adding Net to form a molar ratio of [1:1] Net to AAATTT led to a large decrease in peak intensity for the oligonucleotide compared to R1 and the appearance of a peak corresponding to a 1:1 Net:DNA complex, which indicates strong binding between Net and AAATTT. A [2:1] sample showed an additional decrease in parent peak intensity, indicating that Net had completely saturated the binding site of the AAATTT. Even at a [2:1] ratio, no 2:1 species was observed. This indicated that binding is 1:1, as expected. Similar to Net, DB75 is also a well-characterized MG binder that binds with high affinity as a monomer to A·T-rich regions. Experiments were performed using both MG binders with
Figure 2.1 Spectra of mixed DNA sequences forming complexes with DB75 and Netropsin.

(A) DNA sequences used to simultaneously test binding of multiple sequences: ATATAT, AAATTT, ATGA, and R1. Concentration of DNA is 5 μM. (B-D) Spectra of DNA and Net binding with increasing ligand concentrations. Ligand-to-DNA ratios are [0:1], [1:1], and [2:1], respectively. (E-G) Spectra show the competitive binding of Net and DB75 with multiple DNA sequences. Respective ratios are shown at [0.5:1], [1:1], and [2:1]. Free DNA is indicated by a sequence name above the corresponding peak (AAATTT, m/z 7,921.5) and ligand-DNA complex as name + n (AAATTT + Net, m/z 8,352.5).

ATATAT, AAATTT, and R1 to test competitive binding. Equal concentrations of DB75 and Net were added at a mole to mole ratio of [0.5:1] for a single ligand to DNA. At this ratio, the tallest peak corresponded to 1:1 AAATTT with Net. Smaller peaks show ATATAT + Net and ATATAT + DB75 complexes, but no binding is detected for AAATTT + DB75 (Figure 2.1E). An increase in ligand concentration to [1:1] showed a different pattern in complex formation where AAATTT
+ Net was the most abundant. Peak intensities increased to roughly 50% for ATATAT + DB75 and ATATAT + Net, while a new peak for AAATTT + DB75 was detected (Figure 2.1F). Upon further increasing ligand concentrations to [2:1], a new series of peaks appeared and free ATATAT or AAATTT decreased (Figure 2.1G). The tallest peak became AAATTT + Net, and the intensity of ATATAT + Net surpassed ATATAT + DB75. The peaks corresponding to the parent nucleotides disappeared, indicating complete complexation. This increase in intensity for the ATATAT + Net complex likely occurred once all free AAATTT was consumed due to the high affinity of Net for AAATTT. Residual Net could then saturate the ATATAT binding site. The high binding affinity of Net for AT sequences is apparent, as is the competitive binding of Net over DB75 for AT sequences. Surprisingly, as the concentration was increased, new peaks were detected for DB75 + R1. DB75, known to bind strongly in the MG of AT sequences, has also been reported to intercalate weakly with G·C sites, which is likely the case for R1 + DB75 at increased concentrations. The intercalated binding of DB75 occurs because the stronger binding Net saturates the A·T-rich MG site. As the free concentration of DB75 reaches a high level, intercalation at the available G·C sites then ensues. All of the results with Net and DB75 are in good agreement with published results but also provide new insight into the interactions of these compounds with different DNA sequences.

2.5.2 Using DB293 to test monomer versus dimer binding

Previous work with DB293 indicates it forms a cooperatively stacked dimer with ATGA and a 1:1 complex with AATT. Thus, the binding of DB293 provides a more complex ESI-MS test system.[4] Solutions of DB293 were mixed with R1, AAATTT, and ATATAT, as well as ATGA (Figure 2.2). Initially, only monomer complexes with ATATAT and AAATTT are formed at a [1:1] ratio with little preference for either sequence. At a [2:1] ratio, a 1:1 species between
Figure 2.2 Using DB293 to test monomer versus dimer binding.
Spectra of mixed DNA sequences titrated with DB293. Unbound DNA is indicated by a sequence name above the corresponding peak (ATGA, m/z 6,686.5) and ligand-DNA complex as name+n (ATGA+2, m/z 7,375.0). DNA concentration is 5 μM for each sequence. Molar ratios of [0:1], [1:1], [2:1], and [4:1] correspond to (A-D), respectively.

Both A-T sequences and DB293 is again observed. For ATGA and DB293, no 1:1 complex is detected but a 2:1 species is observed. The lack of a 1:1 complex between DB293 and ATGA is indicative of strong cooperative dimer binding and is in agreement with biosensor-SPR results.[4] When the concentration of DB293 was doubled to a [4:1] ratio, the pattern remained the same. R1 showed no interaction with DB293 at any of the ligand concentrations. The peak at 7,375.5 m/z not only indicates a 2:1 complex for DB293 with ATGA, but having the highest intensity suggests a greater affinity for the cooperative dimer with ATGA over monomer formation with ATATAT and AAATTTT sequences.

2.5.3 DNA and DNA-complexes are stable

We adjusted each parameter individually to monitor for changes in ligand binding and signal intensity for the DNA-MG binder systems (Figure 2.3). There were no changes in signal
intensity which indicates the ESI-MS conditions used in this work are appropriate. We have found that relatively weak complexes (\(i.e. K_D > 0.1 \text{mM}\)), such as external electrostatic complexes, are easily dissociated during injection of the sample and into the gas phase, which filters the results and emphasizes more specific and relative strong complexes over nonspecific binding. The raw, deconvoluted spectra show multiple charged species (Figure 2.4) for free DNA and DNA-complexes with the most abundant charge states having net charges of -4, -5, and -6. For these systems, the peaks observed correspond to DNA with partially neutralized backbones. Neutralization occurs during the electrospray process through interaction of the ammonium ions with the phosphate backbone. The degree of neutralization and net charge depends on ammonium acetate concentration, size of the DNA, and the conditions used which were chosen based on earlier reports.[5, 6]

2.5.4 Sensitivity of DNA and DNA-complexes

Differences in the complex conformation following ionization was an initial concern, meaning the complex detected in the gas phase may not reflect the complex found in solution. A conformational change could, therefore, have an effect on the response and relative binding affinity. For example, after the first DB293 compound binds with ATGA, it is believed that the MG widens to accommodate the second, more favorable molecule. This is a conformational change in the microstructure of the ATGA minor groove and yet, response factors for systems that undergo microstructural changes are similar.[7, 8] This assumption for similar response allows an easy comparison of the peaks. The relative affinities of free DNA and DNA-complexes are, thus, approximated by monitoring changes in relative intensities before and after addition of the compound. Additionally, molecular dynamics simulations of DNA-MG binder complexes \textit{in vacuo} have been reported using conditions to mimic the environment in a mass spectrometer.[9]
Figure 2.3 Peak comparison of R1, ATATAT, AAATTT, and Net-DNA complexes. DNA sequences and Net concentrations each 5 µM. Top row (A) capillary voltage, (B) collision energy, and (C) cone voltage conditions. Bottom row (D) desolvation temperature, (E) extraction voltage, and (F) source temperature conditions. Peaks in order of least to greatest m/z: R1, ATATAT, ATATAT + Net, AAATTT, and AAATTT + Net.

The simulations suggested little conformational difference between solution and gaseous environments as long as the DNA charge is low in the gas phase (bound ions). This provides additional support that the relative affinities of the DNA-complexes observed using ESI-MS are comparable to those found by other methods. The actual process of solution to gas phase can pose some difficulty for solutions requiring an aqueous environment. A more volatile solvent, such as methanol, is often times added to samples to increase the transition, and as a result the signal/noise ratio (S/N). With our conditions, addition of MeOH (up to 20% v/v) led to some reduction in the amount of complex observed with a concomitant increase in free DNA. The addition of MeOH did, however, result in some general improvement in S/N. Due to the lower boiling point of methanol, only a small percentage is required to facilitate transition into the gas phase. The use of
Figure 2.4 Raw, convoluted data to observe competitive binding of Net with R1 and AAATTT. Concentrations of AAATTT, R1, and Net are 5 μM each. Titration ratio is expressed as [1:1] where molar ratio is Net to AAATTT. The most abundant species are the -4, -5, and -6 charge states.

methanol raised concerns regarding the appropriate amount to add and possible effects the solvent could have on the DNA structure. The binding mechanism for MG binders in contingent on the shape and width of the MG and the selectivity of these compounds is also dependent on solvent effects. Based on our findings, solutions containing variable methanol concentrations did not show significant conformational effects based on the amount of complex observed.

2.6 Discussion

Using a mixed set of DNA sequences, we can quickly and accurately evaluate relative affinities, stoichiometry, and cooperativity. These tests show that ESI-MS not only detects binding using control DNAs and MG binders, but also enables competition among ligands and DNAs to be observed. ESI-MS can screen multiple sequences simultaneously, and alternative binding modes for MG binding compounds can be detected. Our intention for developing this method is to complement information obtained by other techniques such as footprinting or next-generation assays. The goal is to determine differences in binding for compounds with a closely related set of sequences having systematic variations. The DNA has a known binding site, in addition to sequence variants (e.g., ATAT, AATT, and AAAA), so we can focus on sequence specificity and
compound selectivity. The strongest binding ligands will interact with available binding sites first, which makes this a true competitive assay.

Analyzing interactions between ligand and a single DNA is not efficient. One can study any number of interactions in the same time it takes to study one. The key is in ligands and sequences which form complexes with distinguishable molecular weights. By using sequences with like compositions, the sensitivities of DNA and DNA complexes are also similar; thus, limitations such as response factors are overlooked.[6, 8] This feature allows direct comparison of unbound DNA and complexes to determine the relative binding affinity. It further demonstrates that our method to investigate the competitive and selective binding between multiple, mixed DNA-ligand interactions in a single sample not only is more efficient but also provides important new information with DNA interactions.

2.7 Conclusion

We have developed a novel approach to observe competitive binding and to screen for DNA-ligand interactions. To the best of our knowledge, this is the first example using ESI-MS to simultaneously examine ligand binding with multiple DNA sequences. Our findings show excellent consistency between ESI-MS and other biophysical methods. This technique has many favorable features: it is rapid and convenient and requires small amounts of sample. It allows direct comparison of relative binding affinities in addition to stoichiometry and cooperativity. Using this innovative method, one could theoretically screen dozens of sequences and obtain a large amount of information from a single sample, reducing reagents used and time spent cleaning between sample runs. More importantly, this method is not limited to DNA and small molecules and can be applied to other biomacromolecular interactions including proteins, RNA, and carbohydrates.
2.8 Experimental protocol

2.8.1 DNA and compounds

Ligand stock solutions were prepared in doubly distilled water at a concentration of 1 mM. DNA sequences were purchased from Integrated DNA Technologies (IDT, Coralville, IA). Based on the predicted amount of DNA provided by IDT, DNAs were dissolved to concentrations near 1 mM in doubly distilled water. All sequences were converted to ammonium acetate salts by three times dialysis in 150 mM NH₄OAc vacuum-filtered buffer (0.22 μM Millipore filter, pH 6.7) using a 1000 Da cut-off membrane (Spectrum Laboratories, Rancho Dominguez, CA). Following dialysis, concentrations of DNA were spectroscopically determined at 260 nm with extinction coefficients calculated using the nearest-neighbor method.[7] Sequences were denatured at 95 °C and immediately quenched on ice to initiate hairpin formation. Ligand and dialyzed DNA solutions were stored at 4 °C. Titration experiments were performed with a mixed set of DNA in a solution with a total volume of 100 μL. DNAs were diluted to 5 μM each in 150 mM NH₄OAc (pH 6.7) and stored at 4 °C. Titration ratios are expressed as compound-to-single-DNA, instead of compound-to-total-DNA concentration. For example, 20 μM DB293 to 5 μM ATGA is expressed as a [4:1] ratio. R1 was used as a reference because it contained no known target sequence.

2.8.2 Electrospray ionization mass spectrometry

ESI-MS experiments were performed using a Waters (Milford, MA) Micromass Q-TOF in negative ion mode using the MassLynx 4.1 software. Conditions were chosen based on published methods [5] and optimized as shown in Figure 2.3. Capillary voltage, 2200 V; sample cone voltage, 30 V; extraction cone voltage, 3 V; source block temperature, 70 °C; desolvation temperature, 100 °C, and sample injection flow rate, 5 μL/min. A volatile solvent, such as MeOH,
is often added to facilitate solution-to-gas phase transition, however, the addition of a solvent was not necessary using our conditions.

The actual process of solution to gas phase can pose some difficulty for solutions requiring an aqueous environment. A more volatile solvent, such as methanol, is often times added to samples to increase the transition, and as a result the signal/noise ratio (S/N). With our conditions, addition of MeOH (up to 20% v/v) led to some reduction in the amount of complex observed with a concomitant increase in free DNA. The addition of MeOH did, however, result in some general improvement in S/N. Due to the lower boiling point of methanol, only a small percentage is required to facilitate transition into the gas phase. The use of methanol raised concerns regarding the appropriate amount to add and possible effectes the solvent could have on the DNA structure. The binding mechanism for MG binders in contingent on the shape and width of the MG and the selectivity of these compounds is also dependent on solvent effects. Based on our findings, solutions containing variable methanol concentrations did not show significant conformational effects based on the amount of complex observed.

The instrument was flushed with 150 mM NH₄OAc prior to sample injection. Scans were collected every 1.0 sec for 10 min with the final 2 min averaged. Raw spectra of the free DNA and DNA-complexes showed multiply charged species ranging 300 – 3000 m/z and the most abundant peaks belonged to -4, -5, and -6 charge states (Figure 2.4). Spectra were deconvoluted using the Maximum Entropy 1 function which calculates the molecular ion (M) based on the equations below, where $H$ is the proton molecular weight, $m'$ and $m''$ are specific m/z values, and $z'$ is the charge state for $m'$.

$$z' = \frac{(m'' - H)}{(m' - m'')}$$  \hspace{0.5cm} (1)

$$M = z' (m' + H)$$  \hspace{0.5cm} (2)


2.9 References


3 RESOLUTION OF MIXED SITE DNA COMPLEXES WITH DIMER-FORMING MINOR GROOVE BINDERS BY USING ELECTROSPRAY IONIZATION MASS SPECTROMETRY: COMPOUND STRUCTURE AND DNA SEQUENCE EFFECTS


Copyright © 2015 Chemistry – A European Journal

My contribution to this chapter was in sample preparation and spectra analyses for mass spectrometry, thermal melting studies, circular dichroism studies, molecular modeling studies, and writing.

3.1 Abstract

Small molecule targeting of the DNA minor groove is a promising approach to modulate genomic processes necessary for normal cellular function. For instance, dicationic diaminodines, a
well-known class of minor groove binding compounds, have been shown to inhibit interactions of transcription factors binding to genomic DNA. The applications of these compounds could be significantly expanded if we understand sequence specific recognition of DNA better and could use the information to design more sequence specific compounds. Aside from polyamides, minor groove binders typically recognize DNA at A-tract or alternating A·T base pair sites. Targeting sites with G·C base pairs, referred to here as mixed base pair sequences, is much more difficult than those rich in A·T base pairs. Compound 1 is the first dicationic diamidine reported to recognize a mixed base pair site. It binds in the minor groove of ATGA sequences as a dimer with positive cooperativity. Due to the well-characterized behavior of 1 with ATGA and A·T-rich sequences, it provides a paradigm for understanding the elements that are key for recognition of mixed sequence sites. Electrospray ionization mass spectrometry (ESI-MS) is a powerful method to screen DNA complexes formed by analogues of 1 for specific recognition. We also report a novel approach to determine patterns of recognition by 1 for cognate ATGA and ATGA-mutant sequences. We found that functional group modifications and mutating the DNA target site significantly affect binding and stacking, respectively. Both compound conformation and DNA sequence directionality are crucial for recognition.

3.2 Key words
Dimerization; DNA recognition; Mass spectrometry; Minor groove binder; Mixed DNA sequence

3.3 Abbreviations
Deoxyribonucleic acid (DNA); Electrospray ionization mass spectrometry (ESI-MS); Association constant ($K_A$); Minor groove binder (MGB); Mass-over-charge ($m/z$); Surface plasmon resonance (SPR)
3.4 Introduction

Genetic processes are regulated by transcription factors (TF) that target specific DNA sequences. Typically, conformational changes or other processes, such as hydration, that yield strong interactions with bases in the recognition site are involved in binding.[1-3] A major goal of fields from chemical biology to therapeutic development is control of gene expression through TF modulation by small molecules that target DNA.[4, 5] Instead of targeting the major groove, like most TFs, a more effective approach involves using small molecules to form a complex in the minor groove of DNA and allosterically modulate transcription factor binding.[6-9] Both inhibition and enhancement of TF complexes are possible with this approach. Typical minor groove binding compounds are relatively planar, crescent-shaped structures which match the geometry of the minor groove. Reversible binders typically have positively charged groups and form non-covalent interactions with DNA sites through electrostatic, hydrogen bonds, and van der Waals contacts. Although most minor groove binding structural types have a high affinity for Atract and multiple A·T sequences, they do not distinguish well between various A·T base pair sequences.[10, 11]

Polyamides (PA), the paradigm minor groove binding compounds which can recognize mixed or A·T and G·C-containing sequences pose difficulties with solubility, aggregation, and synthetic costs.[12, 13] Dicationic diamidines are a class of minor groove binding drugs which have overcome many of the issues encountered by PAs, but lack the sequence specific targeting characteristics of PAs. A breakthrough compound for dicationic diamidines is compound 1 (Figure 3.1) since it recognizes a target site with a G·C base pair in addition to A·T.[14-16] It is exceptional since it dimerizes in the minor groove of ATGA sequences with positive cooperativity, in spite of being a dication. Earlier reports revealed two binding constants for the dimer, the second
Figure 3.1 Structures of dicationic diamidine minor groove binding compounds used to investigate dimer formation in mixed-sequence sites.

Compound 1 is a reference compound known to dimerize in the mixed sequence site ATGA. Compound 2 is a classical minor groove binding compound known to recognize A·T-rich sites. Compounds 3 to 8 are analogues of 1. Molecular weights are listed below the respective structures.

$K_A$ value considerably higher (> 20-fold) than the first, demonstrating positive cooperativity in binding of 1.[15] The first diamidine molecule is believed to insert itself in the minor groove followed by slight widening in the groove width to accommodate the second, energetically more favorable molecule. The second 1 molecule inserts itself into the groove and participates in $\pi-\pi$ stacking with the first molecule in an antiparallel fashion. Surprisingly, the four positive charges, which would be expected to repel each other due to their close proximities, do not inhibit the cooperative binding.

A similar and symmetric ligand, compound 2, is a well characterized dicationic diphenyl diamidine which behaves as a classical minor groove binding compound by recognizing A·T-rich
sequences with 1:1 stoichiometry.[17] A single, asymmetric modification from benzimidazole to phenyl is the only difference between 1 and 2. Both compounds recognize A·T sequences but 1 has a higher affinity for dimer formation with ATGA. Alternatively, variation in the flanking sequence of the target site has also been demonstrated to affect the binding affinity of small molecules in the minor groove.[15] For instance, the binding affinity of 1 for ATGA can be affected by sequences flanking the target binding site. These examples provide important insights into structural and sequence-dependent effects of minor groove recognition.

A current research goal is to identify how variations in compound structure can affect the relative affinity for specific DNA sequences and how different sequences will be recognized by a single compound. Due to the large number of potential drug candidates and DNA sequences, a robust method to screen DNA and small molecule interactions is essential. Electrospray ionization mass spectrometry (ESI-MS) is a powerful method to investigate minor groove binder-DNA complexes.[18-21] It has been demonstrated that ESI-MS can be used for studying biological macromolecular systems such as DNA complexes because the soft ionization conditions used allow the non-covalent interactions that occur to remain essentially intact.[22-26] Necessary information such as stoichiometry and relative binding affinities can be determined directly, rapidly, and with little material. It is especially useful when examining interactions between DNA and small molecules. We recently reported a high-throughput method using ESI-MS to simultaneously screen multiple DNA-minor groove binder interactions.[27] This technique is advantageous over other screening methods because ESI-MS is gentle enough to detect complexes yet powerful enough to sort out similar complexes. The complexes detected are of minor groove binding compounds having relatively high binding affinities so that they can be detected at low concentrations. Previously, we reported the versatility of our ESI-MS method by demonstrating
the cooperative dimer-forming nature of 1 with ATGA as well as monomer binding for A-T-rich sites.

Our goal in this report is to identify the features of 1 that make it ideal for dimerization with an ATGA sequence. Interactions of structurally similar compounds are compared with a mixed set of multiple DNA sequences since small modifications can affect minor groove recognition. The motifs chosen have the potential for dimer formation based on their similarities to the parent compound 1. Next, DNA-minor groove binder complexes are studied using the parent compound, 1, and mutated target sequences. This is a novel approach to gain insight into how 1 forms a sequence specific dimer in the minor groove. It allows a better understanding of how sequence composition and directionality can affect the selectivity of 1 using cognate and variant ATGA sequences and the method can identify other compounds and sequences for mixed-site dimers.

3.5 Compound design

Visually, the structural and conformational characteristics of 1 are typical for heterocyclic cations that bind specifically in the minor groove at A-T sequence sites. With 1, however, two molecules can form a unique, antiparallel stacked tetracationic dimer in the minor groove and recognize an ATGA sequence with positive cooperativity.[14] The phenyl-furan-benzimidazole system clearly has features that optimize stacking in sequences having a wider minor groove and altering these functional groups can modulate dimerization.[28-29] It is not clear that 1 is the optimum structure for this recognition mode or what other sequences could be recognized in a similar complex. To address these questions and better understand the molecular features that are required for the cooperative dimer complex, a number of analogues of 1 were prepared. The effects
of structural and chemical changes on minor groove recognition of the ATGA target site and related sequences were then investigated with ESI-MS.

Modifications of the furan group give 3, 4, and 5 (Figure 3.1). Analogue 4 contains an imidazole where one nitrogen is adjacent to the benzimidazole, while 3 is an isomer of 4 in which the nitrogen is positioned away from the benzimidazole and the third analogue, 5, is a pyrazole substituted system. For this class of compounds, modifying the furan group from a single hydrogen bond acceptor to a system containing both a donor and an acceptor should better define the stacking effects and hydrogen bonding found in ATGA recognition. The benzimidazole-amidine of 1 provides a strong minor groove recognition module. The indole analogue 6 provides structural similarity to both 1 and DAPI [30] and the indole can preserve strong binding to the minor groove in A·T sequences, but its effects on dimer formation are unknown. In functional groups with multiple nitrogens, this modification may affect the stacking and/or hydrogen bonding required for dimerization.

Lastly, two compounds have additions at the phenyl groups which lengthen the structure by including a second benzimidazole between the phenyl and amidine. Analogues 7 and 8 are isomers and differ in the benzimidazole-phenyl connectivity at the para- and meta-phenyl positions, respectively. This modification was chosen to determine how length, hydrogen bonding capability, and curvature of the moiety could affect DNA-ATGA interactions, and specifically dimer formation.

3.6 Results

3.6.1 Structural effects on selective recognition using analogues of 1

To begin our investigation of the effects of different functional groups and substitutions on minor groove recognition, a test was conducted using 1 as a reference with a mixed set of DNA
sequences including ATATAT, AAATTT, and ATGA (Figure 3.2) since the binding affinities and modes of these sequences have been extensively studied with 1.

A titration assay was performed with increasing concentrations of 1 to DNA and the spectra are shown in Figure 3.3. For each titration, the concentrations are expressed as a mole to mole ratio of 1 to a single DNA sequence to evaluate competition among the DNAs for ligand binding. This procedure allows lower concentrations of ligand to be used and enhances the preferred binding for one sequence over another. The spectrum in Figure 3.3A shows only DNA, where no 1 was included. Peaks are labeled as the “sequence name” over the corresponding molecular weight (m/z). In the following titrations (Figures 3.3B and 3.3C) where 1 is added, peaks begin to show for complexes formed between DNA and ligand. For example, a complex formed between 1 and AAATTT is shown at m/z 8,266 and labeled as AAATTT + (1) 1, where the integer in parentheses is the stoichiometric value for one molecule of 1 bound to AAATTT. In Figure 3.3B, the binding of two 1 compounds to ATGA is observed which indicates dimerization of the ligand with ATGA as expected. It is interesting to note that no 1:1 binding of compound 1 with ATGA is detected, which is reasonable since the affinity of the second molecule is more than 10-fold greater than binding of the first molecule.[15] The monomeric binding of 1 binding with AAATTT and ATATAT sequences and dimeric binding to ATGA is in agreement with the literature. The most distinctive characteristic of 1 is its ability to selectively bind as a cooperative dimer with ATGA while forming only monomer complexes with A·T-rich sequences. These results are clearly observed in Figure 3.3, further illustrating the positive cooperativity of 1 with ATGA by ESI-MS.

To expand our understanding of how structural modifications of the 1 motif affects recognition in the DNA minor groove, the analogues of 1 were screened with the same mixed set of DNAs previously tested with 1. The structural conformation of each analogue can vary
considerably depending on the modification made, and the intrinsic groove width for each DNA depends on the base pair sequence. The groove width of A-tract sequences are the most narrow of sequences compared to alternating A·T sites, followed by mixed sequence sites, and G·C-rich sequences having the widest groove. Therefore, interaction between DNA and ligand is contingent on the inherent minor groove width matching the conformational space of the ligand(s). The A·T-rich sequences would be expected to bind the ligand as a monomer while the mixed sequence site, ATGA, should be able to recognize two ligands. The minor groove of the R1 sequence is too wide and has the steric effects of the G·C hydrogen bond in the minor groove making it difficult to recognize small molecules, such as our dicationic diamidines, by the R1 reference sequence. As aforementioned, complexes formed are labeled as “sequence name” + (n) ligand, where (n) is the stoichiometric value for one ligand molecule bound to DNA. Screening of the analogues first began with the compounds extended in length, 7 and 8. With 7 at a [1:1] ratio, only small peaks
Figure 3.3 Example ESI-MS spectra of compound 1 titrated with multiple DNA sequences. Free DNA sequences are apparent by the sequence “name” above the corresponding peak (e.g., AAATTT m/z 7921.5) and ligand–DNA complex as “name + (n ligands bound) ligand name” (e.g., ATGA+(2) 1, m/z 7375). Concentrations of 1 are expressed as a mole-to-mole ratio for 1-to-DNA and range [0:1] to [2:1]. Note that the positive cooperative nature of 1 binding to ATGA is indicated by increasing peak for the dimer species and no detectable 1:1 species. (A) [0:1], (B) [1:1] and (C) [2:1].

For AAATTT + (1) 7 and ATATAT + (1) 7 were detected (data not shown). A two-fold increase in ligand concentration showed higher peak intensities for AAATTT and ATATAT. In Figure 3.4A, a [4:1] titration of 7 with mixed sequences showed large peak intensities for 7 with the A·T sequences, but no ATGA complex. Alternatively, no complexes with ATGA, ATATAT, or AAATTT sequences were detected with 8, an isomer of 7. Formation of complexes is contingent on the compound having a complementary shape to fit in the minor groove. For instance, based on
Figure 3.4 Spectra of DNA sequences titrated with analogues 3 - 8.
Unbound DNAs are indicated by the sequence “name” above the respective peak (e.g. AAATTT, m/z 7921.5) and ligand–DNA complex as “name + (n ligands bound) ligand name” (e.g., AAATTT+(1) 3, m/z 8265.5). Molar ratios are expressed as [4:1] where ligand is to DNA. (A) 7, (B) 8, (C) 4, (D) 5, (E) 6 and (F) 3.

the intrinsic helical nature of the DNA minor groove, and due to the extreme curvature of 8, one would expect to find a weakened interaction with any DNA (Figure 3.4B), as observed.

In our investigation of compound conformational space affecting recognition of ATGA, derivatives with the furan replaced by other five atom heterocycles were evaluated. A titration with the imidazole, 4, at [1:1] detected no complex interactions. On increasing concentrations of 4 to [2:1], 1:1 binding for AAATTT and ATATAT and 2:1 stoichiometry with ATGA were observed with similar intensities for both A·T complexes (data not shown). After further increasing the
concentration of 4 to [4:1], dimerization was enhanced and the peak intensity for ATGA increased relative to AAATTT + (1) 4 and ATATAT + (1) 4. In this case, the intensity for the A·T sequences were comparable to one another (Figure 3.4C). A higher peak intensity of 4 with ATGA was observed with nearly equal intensities to A·T complexes. It is important to note that upon increasing the concentration of 4, specifically from [1:1] to [2:1], only dimer complexes between 4 and ATGA were observed. The lack of any detectable 1:1 species illustrates the positive and cooperative binding behavior of 4 with ATGA.

Analogue 5, which also has a central nitrogen heterocycle, was examined with the mixed DNA set. At lower concentrations of 5, a small peak for a complex formed with ATATAT was detected along with cooperative dimer binding with ATGA. Unlike its isomer 4, 5 showed a small peak at a molar ratio of [1:1] corresponding to ATATAT + (1) 5. At a [2:1] ratio, a complex with AAATTT was identified, but with less affinity than with ATGA and ATATAT. A 2:1 complex with 5 and ATGA was detectable with a higher intensity than AAATTT and ATATAT + (1) 5. By again doubling the concentration of 5 to [4:1], the intensity of ATGA + (2) 5 increased to more than double that of free ATGA, and increases in both ATATAT and AAATTT complexes were observed (Figure 3.4D). Based on the spectra shown in Figures 3.4C & 3.4D, ATGA recognition as a dimer appears stronger with the pyrazole system found in 5 over the imidazole arrangement of 4. However, neither of these two systems is preferred over the furan found in 1 based on results obtained using ESI-MS with mixed sequences.[27]

Due to its structural similarity to 1, the indole-substituted analogue, 6, would be expected to recognize ATGA as a dimer. At a mole:mole ratio of [1:1], a small complex peak for ATATAT + (1) 6 was detected. As the concentration was increased to [2:1], monomers with both A·T
Figure 3.5 Mixed DNA sequence results with 3.
Spectrum expanded between the range m/z 7,250 to 7,950 to highlight the unexpected dimerization of two molecules of 3 bound to R1. The molar ratio shown is [4:1].

sequences were observed with comparable intensities, but no complexes formed with ATGA. Finally, after again doubling the concentration of 6, a peak corresponding to ATGA + (2) 6 was detected with the intensity of nearly half that of free ATGA but with no 1:1 ATGA peak (Figure 3.4E). Analogue 6 exhibits 1:1 stoichiometry with ATATAT and AAATTT; however, the highest complex peak corresponded to ATATAT + (1) 6 instead of the ATGA complex. Unlike 1 at lower concentrations, 6 formed monomeric complexes with near equal proportions from A·T complexes. Increasing concentrations of 6 showed cooperative dimerization with ATGA, but the preference for A·T sequences was greater than ATGA. This set of DNAs with 6 indicates that substituting the benzimidazole with an indole negatively affects the relative affinity for ATGA recognition as a stacked dimer versus monomer A·T binding by a surprisingly large amount.

To investigate the relationship between ATGA recognition and the arrangement of hydrogen bond donors/acceptors, 3, an isomer of 4, was screened with the DNA set. The titration at [1:1] displayed a pattern similar to that found with 5 with only a small peak for ATATAT + (1) 3. Peak intensities from a [2:1] titration showed cooperative binding between ATGA and two molecules of 3 with a higher relative intensity than ATATAT + (1) 3. Doubling the concentration for 3 to [4:1] showed a dramatic increase in dimerization with ATGA (Figure 3.4F), an increase
in ATATAT + (1) 3, and a new peak corresponding to AAATTT + (1) 3. The absence of any detectable 1:1 species again highlights the positive and cooperative behavior of 3 with ATGA. However, at [4:1] there was an additional – and relatively high – peak at m/z 7,680.5. Unexpectedly for this minor groove binding series, the new peak matches a 2:1 stoichiometry for a 3 complex with the reference DNA, R1 (Figure 3.5). This reference sequence has none of the usual diamidine minor groove binding sites and showed no interaction with the other compounds found in Figure 3.1.

As expected with 3 and DNA, monomer binding with the A·T sequences was observed at lower concentrations. At higher concentrations, cooperative dimerization with ATGA was observed with a higher abundance relative to the A·T sequences. The peak corresponding to two 3 molecules and R1, with higher relative intensity than complexes with the A·T sequences was, however, unexpected and was not observed with the isomer 4. The dimeric binding of 3 to R1 is likely attributed to a common TGA, found in both ATGA and the CTGA in R1. The cooperativity of binding from 3 is comparable to 1 with ATGA. Structural similarity would suggest similar interactions with the mixed set of DNAs since 4 and 3 are isomers of each other with only the inner imidazole reversed. Surprisingly, however, 3 showed a higher specificity for ATGA than 4 and 1, in addition to dimerization with R1. This rather significant difference in complex formation with 3 and 4 was certainly unexpected and illustrates the power of the ESI-MS mixed DNA sequences approach to discover new binding modes and sequences.
Figure 3.6 Comparison of compound 1 and analogues 3, 4, 5, and 6.
Left illustrates the electrostatic potential map for the compounds with increasing electronegativity (navy to red). The right column shows a side view of the twists experienced in the overall structures. Molecules were minimized and electrostatic potential maps calculated using Spartan.

3.6.2 Molecular modeling of the compounds illustrates the effect of structural conformation on minor groove recognition

In an effort to explain the compound differences in recognition of ATGA, \textit{ab initio} calculations and molecular modeling were performed to better understand the conformation of 1 and its analogues and understand how slight differences in composition can affect overall conformation (Figure 3.6). The conformation and curvature of 3 are very similar to 1 and yet, interestingly, the behavior of 3 is different from its 4 isomer. The planarity of 3 matches that of 1 as does the electrostatic potential map. These qualities of 3 innately enhance its ability to stack as
Figure 3.7 Models of compound 3 recognizing the mixed sequences ATGA as a dimer. 
(A) The spaced-filled model illustrates the stacked dimer formation of 3 in the minor groove of ATGA. (B) Side view of the stacked compounds. The curvature of the bottom molecule (orange) turns in towards the floor of the minor groove whereas the top molecule (green) faces out toward the solvent. (C) The stacked 3 dimer interactions with the base pairs 5’-ATGA-3’ and 3’-TCAT-5’. H-bond interactions between the base pairs are shown as dashed lines with distances in Å.

a dimer with mixed sequences including sequences with multiple G.C base pairs (i.e. CTGA of R1). Structural information of 4 and 5 were also compared to 1. Surprisingly, the electrostatic potential map for 5 is very similar to 1 but the overall structures do not match. A model of 5 shows a twist in the dihedral angle between the phenyl and pyrazole ring systems. This twist likely arises from the hydrogens of the phenyl and pyrazole groups in close enough proximity to clash which is relieved by a 20° rotation. A twist of the same degree is also observed in 4 between its phenyl and imidazole groups. As with 5, 4 likely experiences clashing between the phenyl and imidazole hydrogens. It appears that a mostly planar conformation, such as that found in 1 and 3, is necessary for strong dimerization in the minor groove of ATGA. Compromising this planarity appears to hinder the ability of 5 and 4 to recognize ATGA presumably due to the conformational changes required to fit the minor groove, particularly as a stacked system. For instance, modifying the core imidazole system in 3 to 4 results in a decreased curvature for 4 compared to 3 which has a more crescent shape.
A model for two molecules of 3 bound in the minor groove of ATGA is illustrated in Figure 3.7 and is based on our current understanding of the interactions between 1 and ATGA.[31] Figure 3.7A is a model portraying the stacked recognition for two molecules of 3 in the ATGA minor groove. The 5′-ATGA-3′ is represented in cyan while the complementary 5′-TCAT-3′ is purple. A side view (Figure 3.7B) of the two stacked molecules illustrates the antiparallel, stacked nature of the compounds. The bottom ligand, in orange, binds in an orientation in which the benzimidazole-amidine motif is at the 3′ end of 5′-ATGA-3′. The N-H group of the central imidazole is solvent accessible while the nitrogen faces the floor of the minor groove to act as a hydrogen bond acceptor with the amino group of G (Figure 3.7C). The N-H of the benzimidazole faces the floor of the minor groove while the attached amidine can hydrogen bond with the keto oxygen of T (adjacent to C) on the complementary strand. The top ligand, shown as green, is orientated with the phenyl-amidine group at the 5′ end of 5′-TCAT-3′ with the curvature facing away from the ATGA minor groove. This arrangement moves the amidine groups apart and helps prevent electrostatic repulsion. The adjacent amidine is also capable of forming a hydrogen bond with the carbonyl oxygen of T of the 3′ end of 3′-TACT-5′. This indicates that rearrangement of nitrogens in the central ring system clearly has an overall effect on binding with ATGA. Reversing the central imidazole ring in 3 so that the two nitrogens are facing the benzimidazole-amidine system increases its curvature to more closely match the contour of the minor groove and improves its affinity for mixed DNA sequences.

3.6.3 DNA sequence and directionality influence selective recognition

An alternative approach to investigate minor groove binding is with the modification of a known target sequence to understand sequence specificity of a single compound and was inspired by the surprising interaction of 3 with R1 (CTGA). Figure 3.8 shows a scheme of the cognate
Figure 3.8 ATGA cognate and sequence variants used to examine the sequence specificity of 1. Base pairs flanking the target sites were maintained to allow similar response. Loops were modified for distinguishability using ESI-MS.

ATGA and ATGA-mutant sequences studied simultaneously with 1. Strong and separated peak intensities were observed for complexes of 1 with ATGA and ATGA-mutant sequences at lower concentrations. Complex peaks had lower intensities compared with peaks of free DNA, but dimerization was observed with ATGA, TTGA, and ATAA. Additional peaks were present for ATAA + (1) 1 and ATGT + (1) 1. Peak intensities for ATAA + (1) 1 and TTGA + (2) 1 were comparable to ATGA + (2) 1. At a [4:1] ratio, the peak for ATGA + (2) 1 showed the highest abundance of the dimer complexes (Figure 3.9). It was followed next, in decreasing order, by
TTGA, ATAA, and ATGT dimer complexes. Monomer complexes were also detected for ATAA, ATGT, and TTGA; however, peak intensities for ATGT + (1) 1 and TTGA + (1) 1 were difficult to distinguish from background noise. The cooperative binding of ATGA is evident by 2:1 complexes and no 1:1 interactions detected. Of the DNA sequences which have both monomer and dimer complexes, the intensity for ATAA + (1) 1 was greater than ATAA + (2) 1. This differs from TTGA where the peak for 2:1 was greater than 1:1. Intensities for ATGT + (1) 1 and ATGT + (2) 1 were nearly equal to each other and no complex between AGTA and 1 was observed.

The strong, cooperative binding of 1 with ATGA indicates a preference over all other similar sequence variants and is clearly optimized for dimer formation. As evident in Figure 3.9, smaller peaks occurred between 1 and ATAA in 1:1 and 2:1 stoichiometries, with a slight preference for the monomer complex, indicating low cooperativity for dimer formation. Dimerization was shown with TTGA at a higher relative abundance. A 1:1 complex with TTGA is also detected, but with lower intensity and only at higher concentrations. Peaks were visible for both monomer and dimer-ATGT complexes at m/z 7,322 and 7,668.5, respectively. However, the intensities were low and signals nearly merged with the background. The strongest dimer-forming complexes were ATGA and TTGA in which peak intensities continued to increase as the concentration of 1 increased.

### 3.6.4 Additional evidence for DNA complex formation with 1 and analogues by thermal melting and circular dichroism

Thermal melting is a robust method to qualitatively measure the stability of DNA and DNA complexes and offers valuable insight on single complexes for comparison with our ESI-MS competition experiments. The $\Delta T_m$ values (Table 3.1) suggest ATGA has a higher affinity for 6 over 4 and 5. The preference of 3 for ATGA over R1 (i.e. CTGA) is in agreement with the results
Table 3.1 ΔTm values of dimer-forming compounds with select DNA sequences.
Changes in thermal stability are color-coordinated based on increase in the melting temperature (white-orange). Thermal melting studies were performed in triplicate ± 0.5 °C.

<table>
<thead>
<tr>
<th></th>
<th>[1:1]</th>
<th>[2:1]</th>
<th>[4:1]</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 with ATGA</td>
<td>2.7</td>
<td>7.3</td>
<td>12.6</td>
</tr>
<tr>
<td>4 with ATGA</td>
<td>2.0</td>
<td>4.5</td>
<td>7.7</td>
</tr>
<tr>
<td>5 with ATGA</td>
<td>2.3</td>
<td>4.6</td>
<td>6.5</td>
</tr>
<tr>
<td>3 with ATGA</td>
<td>4.0</td>
<td>7.3</td>
<td>11.0</td>
</tr>
<tr>
<td>3 with CTGA (R1)</td>
<td>2.3</td>
<td>4.0</td>
<td>6.8</td>
</tr>
<tr>
<td>1 with AAATTT</td>
<td>8.9</td>
<td>11.1</td>
<td>12.4</td>
</tr>
<tr>
<td>1 with ATATAT</td>
<td>7.0</td>
<td>10.7</td>
<td>13.0</td>
</tr>
<tr>
<td>6 with AAATTT</td>
<td>8.3</td>
<td>12.6</td>
<td>14.5</td>
</tr>
<tr>
<td>6 with ATATAT</td>
<td>6.3</td>
<td>10.6</td>
<td>14.4</td>
</tr>
</tbody>
</table>

Table 3.2 ΔTm values of 1 with ATGA and mutant sequences.
ΔTm values of 1 with ATGA and mutant sequences. Values shown were performed at a molar ratio of [4:1] of 1 to DNA. Changes in thermal stability are color-coordinated based on increase in the melting temperature (white-orange). Thermal melting studies were performed in triplicate ± 0.5 °C.

<table>
<thead>
<tr>
<th></th>
<th>AGTA</th>
<th>ATGT</th>
<th>ATGA</th>
<th>TTGA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 with DNA</td>
<td>6.0</td>
<td>8.5</td>
<td>11.0</td>
<td>11.6</td>
</tr>
</tbody>
</table>

shown in Figures 3.4F & 3.5. This is expected since the hydrogen bonding pattern necessary for recognition will be disrupted when substituting adenosine to cytidine (ATGA → CTGA). The peak intensities and ΔTm values for 1 and its analogues with initial mixed sequences are compared in Figure 3.10A with a superimposed model of the compounds (Figure 3.10B) to illustrate differences in structural conformations. The values for 1 and ATGA were taken from previously published results using ESI-MS.[27] A comparison of ΔTm values for 1 with ATGA (Table 3.2) shows that 1 prefers TTGA and ATGA over ATGT and AGTA. This is consistent with the results
Figure 3.10 Comparison of relative abundances and thermal melting.

(A) Comparison of the relative peak intensities (3%) for complexes and ΔTm values (0.5 °C) for mixed DNA sequences with 1 and the dimer-forming analogues 3-6. ΔTm values (secondary y-axis) are for dimer complexes formed between ligands and ATGA at a [4:1] molar ratio. (B) Structural variability and spatial arrangement for dimer-forming compounds are shown by superimposing the molecules over their mutual phenyl-amidines; 1 (tan), 3 (blue), 4 (green), 5 (orange), and 6 (pink).

obtained using ESI-MS, as shown in Figure 3.9, which indicates that the choice and arrangement of base pairs in the target site plays a key role in forming complexes between the parent compound 1 and DNA.

It is interesting to see that the DNA complexes formed are well-defined. The binding behavior established by 1 translates well for some of its analogues with ATGA recognition.
Specifically, cooperative binding is observed with increasing concentrations of ligand where the complexes formed between ATGA and 1 or its analogues show only 2:1 dimer ligand to DNA complexation. Any 1:1 complexes formed are at very low concentrations and, therefore, not detected. This illustrates that as concentrations of ligand are increased, so does the amount of complexes detected. The results correlate well with the thermal melting studies which allows the relative affinities of DNA and DNA-ligand complexes to be directly compared.

The overall structure of DNA and DNA complexes can be evaluated using circular dichroism spectroscopy (CD). CD studies are useful in examining the global conformation of DNA and other biomolecules. Studies were performed to compare the conformation of ATGA to its mutated sequences (Figure 3.11). Curves were normalized and superimposed to facilitate the comparison. No change in CD signal was detected when comparing ATGT or AGTA to our cognate sequence, and only minute differences in the normalized signals of ATAA and TTGA were observed when compared to ATGA. The spectra showed that ATGA and its mutant sequences were very similar in their overall DNA conformations. Based on the structural similarities of ATGT and AGTA when compared to ATGA, it is becomes more clear that it is sequence which plays a direct role for recognition by 1. The slight variation in the CD curves are likely attributed to the individual base pair stacking which will affect the DNA microstructure but not the overall conformation. This may offer an explanation as to why interaction of 1 with ATAA can occur by both monomerization and dimerization since the microstructures between ATAA and mixed sites will be different.

3.7 Discussion

This report indicates that several features of the DNA complexes of 1 and analogues with DNA are critical in whether the compounds bind as a cooperative dimer or monomer. For DNA,
the groove width and sequences of hydrogen bond donors and acceptors on the base pair edges at the floor of the minor groove are crucial. For the compounds, the stacking ability in the correct conformation to place hydrogen bond donors and acceptors to match the groups on DNA are necessary. These features are illustrated in Figure 3.7 and are based on the known binding mode of 1 with ATGA.\cite{14, 15} Because of the complexity of these factors, experimental methods are required to determine whether the compounds bind as cooperative dimers, monomers, or if binding occurs at all. To do this in a competitive format for evaluating a large number of DNA sequences and compounds as rapidly as possible while using little sample, we have established the competitive ESI-MS method. Our method provides detailed insight into the features necessary to form a stacked cooperative dimer with DNA. We have discovered an entirely new and unexpected binding mode for 3. This is the first report of recognition by a synthetic, non-polyamide compound.
for a multiple G·C sequence such as R1 with positive cooperativity. The analogue, 3, forms not only a strong cooperative dimer complex with ATGA, but also forms a strong cooperative dimer with CTGA in R1. The core site, CTGA, has none of the traditional sites found in sequences recognized by 1 or similarly designed compounds. Steric hindrance from the hydrogen bond between the carbonyl group of cytidine and the amino group of guanidine in the minor groove likely affects the stacking ability of 3 and yet the compound is still capable of forming a strong positive cooperative dimer complex with CTGA. It is also interesting to see that the isomer of 3, 4, is unable to bind CTGA and binds weakly with ATGA. A seemingly subtle reorientation of the central imidazole places it in a way so that the compound is unable to form hydrogen bonds with the bases in the minor groove. The strategic placement of groups that act as hydrogen bond donors or acceptors is key for stacked binding in the minor groove and these complexes are detected using a competitive ESI-MS method. This important discovery provides a new paradigm for rationally designed, synthetic compounds to recognize mixed and/or G·C-rich sequences.

Parent compound 1 binds in the minor groove of ATGA as a dimer and recognizes A·T-rich sequences as a monomer. In contrast to this generalization, detection of ATAA as both monomer and dimer by 1 expands our understanding of earlier evidence of dimerization of 1 at an ATAA site.[15] For A·T-rich sequences, the minor groove of A-tracts is distinguished by a narrower groove width while alternating A·T sequences, including those with the TpA step, are wider.[10, 33] Monomer and dimer binding should be possible for an ATAA sequence due to the TpA step and wider minor groove. Based on the dual recognition of ATAA by 1 as a monomer and dimer and with nearly equal intensities, the ATAA minor groove is more closely related to alternating A·T sequences than A-tracts. Binding of 1 to ATAA as a monomer can slightly narrow the groove width while binding as a dimer can slightly widen the groove.
For our mixed sequence mutants, the melting temperature of the free DNAs fall within ± 1.0 °C of ATGA and so the particular arrangement of the base pairs in these sequences does not have a large effect on the thermal stability of free DNA. The sequence arrangement, however, has an effect on binding of 1. For instance, mutation of the cognate sequence to read GT in the 5′ to 3′ direction within AGTA or ATGT, results in a significant decrease in binding of 1 compared to ATGA. Footprinting studies with 1 and a single mutation from ATGA to AGTA have shown similar results with no AGTA recognition.[16] Additionally, as evident with TTGA at the 5′ end when A is replaced by T, cooperative binding of 1 is present, but decreased. These results suggest that because the base pair composition is very well maintained (G·C and A·T content), it is the stacking of the base pairs AA·TT vs. AT·AT that influences changes in minor groove microstructure and affects the affinity and binding mode of 1. Further investigations are necessary to identify minor groove microstructures for sequences with similar structures to ATGA.

Based on the ESI-MS studies of 1 with several DNAs, we can now see that it binds as a highly cooperative dimer to ATGA-like sequences but as a monomer to A-tract sequences. Based on the structural similarity of benzimidazole and indole groups, we expected the indole analogue of 1, 6, to bind as a similar cooperative dimer. With a few exceptions, however, dimerization among minor groove binders containing an indole system is rare.[29, 32] Most indole-containing minor groove binders recognize A·T sequences strictly as a monomer. For instance, DAPI, the most thoroughly studied indole-containing compound, binds A·T sequences as a monomer only.[30] More interestingly, however, is the higher affinity of 6 over 1 for ATGA which is unexpected since the curvature and conformation of the benzimidazole and indole systems are essentially the same (Figure 3.6). Biosensor-SPR studies (not published) have shown that 6 binds as a strong dimer to ATGA with a higher affinity over 1 which is in agreement with the thermal
melting studies, however, the results from ESI-MS are not completely consistent. In the mass spectra, the 6-ATGA relative peak abundances are not as high as one would anticipate based on the results with 1. At this time, it is not completely understood why the 6-DNA peaks, which includes 6 with ATGA and both A·T sequences, are less than expected. This is especially surprising since there has been excellent correlation between ESI-MS and thermal melting with 1 and the other analogues. One possible explanation may be technique-related in which the compound interacts with the injection tubing so that the total concentration of 6 in the sample solution decreases below the expected amount. A lower concentration of 6 would then result in less 6 complex formed and lower abundances of 6-DNA complexes detected.

To examine competition for DNA sites by 1 and analogues using ESI-MS, proper care must be taken to ensure that the molecular weights of the small molecules and their complexes, and all possible stoichiometries, are distinguishable. On the other hand, another approach is to examine the binding of a single compound with an array of target sequences and their mutations. Different DNA sequences can be examined simultaneously in this way as long as the molecular weights of the DNAs and complexes are distinguishable. A combination of an ATGA cognate sequence, ATGA-mutant sequences, and a reference DNA (R2) were screened with 1. To obtain different molecular weights for the variants, such as ATGA and AGTA which have the same stem molecular weights, the hairpin loops of the DNAs were altered with different numbers of thymidine and cytidine or by incorporation of a deoxyuridine so that the flanking base pairs were preserved.

In the spectra shown, peaks of the systems correspond well to their expected molecular weights (i.e. m/z) for free DNA and DNA-ligand complexes. The ionization process of ESI-MS results in multiply charged species and for the raw data, every system shows multiple, charge states. Due to the nature of the analyte and negative mode analysis, the most abundant charge states
range between -3 and -6. These lower net charges indicate the DNA backbone becomes partially neutralized during the electrospray process during which ammonium ions transfer a proton to the phosphate backbone and the ammonia molecules evaporate. The amount of neutralization occurred depends on the size of the DNA, concentration of ammonium ions, and instrument parameters used.[18, 37] Positively charged dicationic diamidines help in neutralizing the backbone, however, the presence of ligand does not affect the overall charge after forming a complex. For instance, peaks remain the most abundant in -4 and -5 charge states for both free AAATTT and AAATTT + (1) 1 complexes. The spectral peaks are transformed via deconvolution – the ability to transform multiple charge peaks into the single peak, zero charge molecular ion species. Deconvolution greatly simplifies the spectra for optimum visualization and is achieved by multiplying the charge of the species by its respective m/z.

Lower DNA concentrations such as 2.5 μM have been tested and not surprisingly, there is little difference in the peak intensities when comparing 2.5 μM of DNA versus 5 μM of DNA. The level of cooperativity is still observed, and is in agreement with earlier reports from our group demonstrating the cooperative binding of 1 to ATGA by ESI-MS using 5 μM of DNA.[27] For our systems, there is a general preference for using 5 μM of DNA since it results in a larger signal for the DNA and/or complexes over using 2.5 μM. A spectrum using 2.5 μM concentrations of DNA with compound 1 is shown in Figure 3.12. Due to the nature of compounds 1 – 8 and other dicationic diamidines, an unknown amount of ligand is often lost during the injection process. At times, the ligand will presumably become stuck and remain fixed to the inside of the injection tubing, therefore reducing the total ligand concentration. This phenomenon has been experienced on multiple occasions and requires thorough cleanings of the instrument between different samples. Samples containing DNA only (no compound) are routinely injected before beginning
any new analysis to check for and remove residual ligand through binding of free DNA. Results can be successfully quantified using ESI-MS, as long as the specific response sensitivity and the concentrations are accurately known. It is possible to determine an equilibrium binding constant for DNA and small molecule systems and there are examples in literature demonstrating this.[36, 38] The ability to determine binding constants for dicationic diamidines is primarily limited to the loss of ligand during injection and response factors for the DNA and complexes, and these limitations influence our preference to use ESI-MS for qualitative purposes only.

Other methods can also be used, with or without ESI-MS, to efficiently screen for DNA binding compounds. For instance, thermal melting studies are commonly used to screen for binding of ligand to DNA. Additional techniques can include fluorescence assays, competitive dialysis experiments,[34] and separation techniques such as gel electrophoresis [35]. While these methods can provide important information, they can often demand more time and sample than ESI-MS. The ESI-MS technique reported here is rapid and convenient, requires little sample, and can provide quantitative information.[36] However, the most important feature from this method is that it can offer quick insight into the preferential binding of ligands based on compound structure and/or DNA sequence. With this, one can determine the stoichiometry, relative affinity,
the binding mode (cooperative vs. non-specific) and it can even be used to determine heterodimeric binding.

3.8 Conclusion

Mixed DNA sequence investigations using ESI-MS has allowed the discovery of important features of 1 and analogues with ATGA and mutant sequences. For specificity and cooperative binding affinity to ATGA, these results show that at this point, 1 is the optimum compound. The results also show that 3 binds very well to ATGA but has many other strong interactions. A surprising result is that 3 binds quite well as a 2:1 dimer species to the G·C sequence, R1, which was selected because heterocyclic dicationic diamidines have not been observed to bind to such G·C-rich sequences. Analogue 4, the imidazole isomer of 3, does not bind as well to ATGA and does not bind at all to R1. The surprising binding of 3 needs additional investigation.

For sequence specificity, the sequence ATAA is capable of binding 1 as both a monomer and dimer, despite containing no G or C bases in the target site. In sequences containing a G·C base pair, the order of base pairs played a strong role in recognition by 1 such that the GpT and TpG steps had surprisingly different binding modes. Sites with TG have a preference for 1, whereas, GT sites tend to avoid forming complexes. Overall, ATGA remains the preferred site for cooperative 2:1 binding of 1 and these results further illustrate that sequence is crucial for minor groove recognition.

3.9 Experimental procedures

3.9.1 Materials

Compounds 1,[28] 2,[39] and 3 [40] were synthesized using previously reported methods, and syntheses for the new analogues 4 – 8 are available in the Electronic Supporting Information of Laughlin S, et al. Chemistry, 2015, 24 (14): 5528-5539. All compound stock solutions were
prepared in doubly distilled water at a concentration of 1 mM. DNA sequences were purchased from Integrated DNA Technologies (IDT, Coralville, IA). Based on the predicted amount of DNA provided by IDT, DNAs were dissolved in doubly distilled water (1 mM). All sequences were converted to ammonium acetate salts by three steps of dialysis in 0.15 M ammonium acetate vacuum-filtered buffer (0.22 μM Millipore filter, pH 6.7) using a 1000 Da cut-off membrane (Spectrum Laboratories, Rancho Dominguez, CA). Following dialysis, concentrations of DNA were spectroscopically determined at 260 nm with extinction coefficients calculated using the nearest-neighbor method.[41] Sequences were denatured at 95 °C and immediately quenched on ice to initiate hairpin formation. Ligand stock solutions and dialyzed DNAs were stored at 4 °C.

Titration experiments were performed with a mixed set of DNAs in a single Eppendorf tube (100 μL, total volume). Ammonium acetate buffer was used due to its volatility under mass spectrometric conditions.[18, 23] DNAs were diluted (5 μM, 0.15 M ammonium acetate buffer, pH 6.7) with the appropriate concentration of ligand, vortexed, and stored at 4 °C until injection. Ratios with no surrounding punctuation refer to stoichiometry (i.e. 1:1 is ligand:DNA) whereas titration ratios are enveloped by brackets. For example, titration ratios are written as \([n:m]\), where \(n\) and \(m\) are empirical concentrations of ligand and DNA, respectively. Titration ratios were prepared as compound-to-single-DNA. The ligand-to-single-DNA approach is more desirable for competitive binding analyses using multiple DNA sequences and avoids higher ratios being prepared. Two distinct hairpin DNA sequences were used to compare the formation and relative abundances of free DNA and DNA-complexes to a reference peak. For experiments with \(\mathbf{1}\) and its analogues, \(\mathbf{R1}\) was used because it contained no known target sequence. \(\mathbf{R2}\) was later used as a reference to compare \(\mathbf{1}\) with mutant DNA sequences due to the reference base pair composition which consisted of GC base pairs only in the DNA stem.
3.9.2 Electrospay ionization mass spectrometry

ESI-MS experiments were performed using a Waters Micromass Q-TOF (Waters, Milford, MA) in negative ion mode and MassLynx 4.1 software. Capillary voltage was set to 2500 V, sample cone voltage to 30 V, and extraction cone voltage at 3 V. Source block temperature was set to 70 °C and desolvation temperature at 100 °C. Prior to injection, the instrument was flushed with ammonium acetate buffer (0.15 M). Samples were injected at a rate of 5 μL·min⁻¹ and run for several minutes until the MassLynx chromatogram reached stabilization. Scanned peaks ranged m/z 300-3000 and the most abundant peaks observed belonged to -3 to -6 charge states. Scans were averaged over the last 2 min of analysis. Spectra were deconvoluted for comparative purposes. Deconvolution was achieved through multiplying peak intensities (m/z) by the charge (z) using the Maximum Entropy 1 Function (MassLynx 4.1).

3.9.3 Thermal melting

Thermal melting studies were performed in cacodylate buffer (0.01 M cacodylic acid, 1 mM EDTA, 0.1 M NaCl, pH 7.1) using a Cary 300 UV-Vis spectrophotometer (Varian, Walnut Creek, CA) and a 1 cm quartz cuvette. Compound concentrations were chosen to give the desired ratio of compounds to hairpin DNA (3 μM). Scans were run from 25 °C to 95 °C at a rate of 0.5 °C·min⁻¹.

3.9.4 Circular dichroism

Circular dichroism studies were performed using DNA prepared in cacodylate buffer (5 μM, 0.01 M cacodylic acid, 1 mM EDTA, 0.1 M NaCl, pH 7.1) using a Jasco J-810 Spectropolarimeter (Jasco Analytical Instruments Inc., Easton, MD) and a 1 cm quartz cuvette. Scans were performed at a rate of 50 nm·min⁻¹ from 320 nm to 220 nm, acquired in triplicate, and averaged.
3.9.5 **Molecular modeling**

*Ab initio* calculations were performed in Spartan 10. Structures were minimized in the equilibrium geometry setting using a Hartree-Fock wavefunction and 6-31G* basis set. Molecules were set to dications in a vacuum environment. Canonical B-form doubled stranded DNA was built using the Sybyl software and coordinates saved as PDB file. Hydrogen atoms were added to DNA using xLeap, solvated within a 10.0 Å TIP3PBOX waterbox, and neutralized by sodium ions. DNA minimization was achieved using AMBER99 force fields. DNA was visualized in VMD and coordinates were saved. The DNA sequence was then visualized and modeled with compound 3 using Chimera 1.8.1.

3.10 References


40. Batista-Parra, A. *PhD thesis.* Georgia State University (USA) **2003**.

4 MIXED UP MINOR GROOVE BINDERS: CONVINCING A·T SPECIFIC COMPOUNDS TO RECOGNIZE A G·C BASE PAIR

Ananya Paul, Rupesh Nanjunda, Arvind Kumar, Sarah Laughlin, Raja Nhili, Sabine Depauw, Shelby Sheldon Deuser, Yun Chai, Arpana S. Chaudhary, Marie-Helene David-Cordonnier, David W. Boykin, W. David Wilson


Copyright © 2015 Bioorganic Medicinal Chemistry Letters

My contribution to this chapter was in sample preparation, performance, and analysis of the results by electrospray ionization mass spectrometry, and portions of writing.

4.1 Abstract

DNA minor-groove-binding compounds have limited biological applications, in part due to problems with sequence specificity that cause off-target effects. A model to enhance specificity has been developed with the goal of preparing compounds that bind to two A·T sites separated by
G·C base pairs. Compounds of interest were probed using thermal melting, circular dichroism, mass spectrometry, biosensor-SPR, and molecular modeling methods. A new minor groove binder that can strongly and specifically recognize a single G·C base pair with flanking A·T sequences has been prepared. This multi-site DNA recognition mode offers novel design principles to recognize entirely new DNA motifs.

4.2 Key words

Biophysical analysis; Heterocyclic diamidines; Mixed DNA recognition; Rational design; Sequence specificity

4.3 Introduction

The systematic design and preparation of compounds that can recognize mixed base pair (bp) nucleic acid sequences is a very important goal in DNA small-molecule applications. Such specific recognition would allow new diagnostic applications in vitro while in vivo they could provide new therapeutics and gene-specific probes. This goal has been difficult to reach, however, and to date only polyamides have had significant success in design of compounds for mixed-sequence DNA recognition.[1-4] Unfortunately, polyamides can be limited by synthetic costs, aggregation and cell-uptake problems.[5, 6] In an interesting contrast, relatively simple A·T sequence selective minor-groove binders have had good success in therapeutic targeting of DNA in cells ranging from various types of cancers [7-13] to bacterial [4-16] and parasitic microorganisms [17-22]. DNA minor-groove binding heterocyclic diamidines, such as pentamidine and berenil, are examples with a long history of therapeutic use.[20, 23-25] DAPI is another heterocyclic diamidine that readily enters cells and is frequently used to stain nuclear DNA in imaging.[26] Limited selectivity and off-target effects, however, have restricted the therapeutic
applications of these types of compounds, and methods to increase their binding selectivity and affinity are essential for development of improved agents.

The starting points for our development of new sequence specific agents are synthetic compounds known to recognize A·T sequences as well as the extensive literature on A·T specific minor-groove binders. The heterocyclic diamidines described above are all A·T site specific. For recognizing sequences containing G·C bp, new functional groups must be incorporated into the A·T specific agents. Such compounds would be helpful in a number of areas such as targeting the kinetoplast DNA (kDNA) of parasitic microorganisms, where a large number of A·T bp sites are frequently separated by a G·C bp [27-29], as well as transcription factor (TF) inhibition in gene control applications [30]. Targeting transcription factors has been recognized as a promising but difficult direction for therapeutic development.[31, 32] In our initial efforts in this area, compounds with alkyl or alkyl-aromatic linked amidine-benzimidazole-phenyl (ABP) motifs were designed, synthesized and evaluated for binding to A·T sites separated by one or two G·C base pairs.[33] The goal of these studies was to determine how linker length and rigidity affect the ability of ABP modules to recognize two A·T sites.

In the progress to our next generation sequence specific minor groove binders, we have prepared a series of potential mixed-sequence DNA binding molecules from the existing pure A·T sequence binders. In the design path to G·C bp recognition by rationally designed small molecules, we have incorporated groups that can act as H-bond accepting units in the minor groove of DNA (Figure 4.1A). Some very encouraging initial success in targeting mixed bp DNA sequences was found with a relatively small, dissymmetric, diamidine compound, DB2277 (Figure 4.1A), which recognizes a single G in sequences such as AAGTT through an aza-benzimidazole group.[34, 35]
In this report we describe a more general approach by connecting two A·T specific binding modules with H-bond accepting linking units. The results show critical effects of molecular flexibility, position of the linker and heteroatom for mixed sequence specific DNA recognition.

Detailed biophysical studies of the compounds in Figure 4.1 showed very strong and selective recognition of a G·C bp separating two A·T sites with much weaker recognition of pure A·T sites by a pyridyl-linked compound, DB2120. The rational design and preparation of a non-polyamide compound that can selectively recognize long, mixed bp sequences is a significant step forward in the design of minor-groove binders for potential diagnostic and therapeutic use. While
there is much to learn, the new results offer clear directions and leads for the development of a DNA minor groove-binding language.

**Compound design and preparation:** The design platform for the compounds in Figure 4.1 starts with the classical A·T-specific minor-groove binding compound, DB2119.[33] The ABP A·T recognition unit was held constant for binding to the flanking A·T sites while other molecular units of DB2119 were varied and the derivatives were evaluated for G·C bp recognition. DB2370 is similar to DB2119 but without the -O-CH₂- linker groups and with the central phenyl replaced with a pyridine. It is a direct derivative of a classical triphenyl A·T-specific minor-groove binder. Combinations of these two gives DB2120 with a central pyridine. The DB2120 and DB2370 pair tell us if the pyridine-N can serve as an H-bond acceptor for the minor groove G-NH and whether the interaction is better with or without the flexible -O-CH₂-. DB2319 is the same as DB2120 but with the amidine cationic groups changed to methyl piperazinyl which is in Hoechst nuclear DNA staining compounds (Figure 4.1). This cationic group change will indicate the contribution of the cations to binding and in cell studies may affect the types of cells that can be targeted with this molecular platform.[30] The new bis-benzimidazoles DB2120, DB2319, and DB2370 were synthesized employing standard methodology for coupling the proper substituted phenylenediamine with the appropriate bis-aldehyde; Schemes and experimental details are provided in the Supplementary Material section of Paul A, et al. *Bioorg. Med. Chem. Lett.* 2015, 25 (21): 4927-4932.

4.4 Results

4.4.1 Thermal melting

The compounds in Figure 4.1 were selected based on $T_m$ screening studies of a large number of synthetic compounds and they illustrate how to incorporate G·C bp recognition into an
UV melting profiles at 260nm of (A) A4GT4, (B) A4T4, and (C) A4GCT4, hairpin DNA sequences in the absence and presence of the heterocyclic cations. The experiments were conducted in Tris-HCl buffer. The listed values are for 2:1 [ligand]/[DNA] ratio and an average of two independent experiments with a reproducibility of ± 0.5 ºC.

**Table 4.1 Thermal melting studies (ΔTm [a], ºC) of the designed heterocyclic amidine compounds with mixed DNA sequences.**

<table>
<thead>
<tr>
<th>Compound</th>
<th>A4T4</th>
<th>A4GT4</th>
<th>A4GCT4</th>
</tr>
</thead>
<tbody>
<tr>
<td>DB2119</td>
<td>21</td>
<td>17</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>DB2120</td>
<td>15</td>
<td>22</td>
<td>8</td>
</tr>
<tr>
<td>DB2319</td>
<td>14</td>
<td>19</td>
<td>6</td>
</tr>
<tr>
<td>DB2370</td>
<td>24</td>
<td>19</td>
<td>5</td>
</tr>
</tbody>
</table>

[a]ΔTm = Tm (the complex) - Tm (the free DNA). The listed values are for 2:1 [ligand]/[DNA] ratio and an average of two independent experiments with a reproducibility of ± 0.5 ºC.

A·T binding compound. The crucial breakthrough in developing strong and selective G·C bp recognition in this series was achieved with the pyridyl derivative of DB2119, compound DB2120. The phenyl to pyridyl change resulted in an increase in the thermal stability of the single G·C-containing A4GT4 motif (ΔTm = 22 ºC; **Figure 4.2, Table 4.1**) but a decrease in A·T binding
The piperazinyl, DB2319, had some selectivity for G·C bp but weaker binding while DB2370 did not show significant selectivity (ΔT_m = 19 °C and ΔT_m = 24 °C for A4GT4 and A4T4, respectively). These results clearly indicate that the pyridine nitrogen in DB2120 is serving as an H-bond acceptor for the exocyclic amino group of guanine to give a considerable increase in thermal stability of the complex. All compounds showed limited enhancement in thermal stabilities for two G·C-containing sequences such as A4GCT4 (Table 4.1).

### 4.4.2 Biosensor-surface plasmon resonance

Biosensor-surface plasmon resonance (SPR) is an effective technique to quantitatively monitor biomolecular interactions in real-time to obtain binding affinity, kinetics, and stoichiometry.[33-38] Biosensor experiments were conducted with DB2120 based on the T_m screening results and Figure 4.3 shows sensorgrams of DB2120 with A4GT4, A4IT4 (I = inosine, 2-amino of guanine replaced with H), and A4T4 motifs. It is apparent from these results that the interactions of DB2120 with DNAs with and without a G·C bp are quite distinct. The most striking result is the interaction with the single G·C bp-containing A4GT4, and particularly noteworthy is the very slow dissociation of DB2120 from the DNA complex (Figure 4.3A). Even with a quite long experimental dissociation time (1 h), very little dissociation of DB2120 from the complex is observed under these conditions. Global kinetics fitting yielded a single binding site and an approximate $K_A = 6.6 \times 10^{10} \text{ M}^{-1}$ ($K_D = 1.5 \times 10^{-11} \text{ M}$) for DB2120 at 0.1 M Na^+. This is an impressive 2000-fold increase compared to the parent compound, DB2119. DB2120 binds to A4T4 in a monomer complex with a 200-fold lower affinity compared to A4GT4, indicating high specificity for the single GC bp sequence. The sensorgrams of A4T4 show an off-rate that is much faster, and complete dissociation from the complex occurs within the first few minutes of the dissociation phase (Figure 4.2C). With the A4GCT4 sequence, no binding is observed for DB2120.
Figure 4.3 Surface plasmon resonance sensorgrams for DB2120. (A) A4GT4, (B) A4IT4, and (C) A4T4 sequences, respectively. The injected concentrations of DB2120 with A4GT4 are 2, 3, 5, 7, 10 and 15 nM whereas with A4IT4 and A4T4 sequences the injected concentrations are higher; 3, 5, 7, 10, 15, 20, 30 and 5, 7, 10, 15, 20, 30, and 50 nM, respectively. The black lines for all three sequences represent global kinetic fitting using a 1:1 interaction model.

Under this experimental condition, in agreement with $T_m$ results. When G (A4GT4) is substituted by I (A4IT4) at the minor groove of the mixed-DNA sequence, although the groove width remains similar [39], the binding affinity of the A4IT4–DB2120 complex decreases by 80-fold ($K_D = 1.2 \times 10^{-9}$ M). This phenomenon is expected by the interruption of H-bonding between the pyridyl-N and G-NH$_2$ (Figure 4.3B) in the A4IT4 complex and confirms the DB2120 to G-NH H-bond.

In summary, SPR and $T_m$ results converge to indicate very strong and selective monomer binding of DB2120 to A·T binding sites separated by a single G·C bp. No other compound from Figure 4.1 showed significant DNA sequence specific recognition. The SPR results clearly highlight the success and the difficulties (compare DB2120 and DB2370) in the development of G·C specific recognition.

4.4.3 Competition electrospray ionization mass spectrometry

Competition electrospray ionization mass spectrometry (ESI-MS) is a method that we have developed to provide comparative information on the stoichiometry, cooperativity, and relative
affinity for DNA complexes.[40, 41] The use of multiple DNA sequences simultaneously mixed with a small molecule creates a competitive binding environment for comparison of DNA–ligand interactions. Since the free ligand concentration for all of the DNAs is the same, peak intensities of the free, unbound DNA and the DNA-small molecule complexes offer insight into the binding mode and preferred binding site(s). Each titration series contains a pure G·C sequence used as an internal standard that does not bind ligand. This offers a simple comparison for peak intensities, since the DNA molecular weights are adjusted to be different. To help adjust molecular weights among this set of sequences, substitution of guanine by inosine in the bp(s) nearest the hairpin loop were incorporated (Figure 4.4A).

The top spectrum in Figure 4.4B is an example of free DNA prior to titration. The bottom spectrum (Figure 4.4C) illustrates DNA interactions with DB2120 at a [4:1] added ratio. The free A4GT4 sequence is no longer visible and shows that the DNA has formed a 1:1, DB2120–A4GT4 complex, as indicated by the tall peak at m/z 10,324. It is striking that there are only very weak complexes for A4GCT4 and A4T4 DNAs at this ratio and is in direct agreement with Tm and SPR experiments. The relative peak intensities for the titrations are compared for DB2120 and DB2370 in Figure 4.4D and shown as compound ratio versus relative abundance. DB2119 showed relatively weak binding to A4T4 and a small signal for A4GT4 at the [4:1] ratio, in agreement with Tm and SPR results. For DB2370 at a [4:1] ratio, the majority of A4GT4 and A4T4 DNAs are bound with nominal amounts of free DNA, whereas most of the A4GCT4 is left unbound. B2370 thus has no significant selectivity for A4GT4 over A4T4. All compounds show very poor binding to the two GC bps sequence (A4GCT4).
Figure 4.4 Competition mass spectrometry.
(A) Sequences used for competition ESI mass spectrometry. Example spectra of (B) free DNA and (C) DNA and DB2120-DNA complexes. (D) Comparison of the peak intensities for DNA and ligand–DNA complexes for a series of titrations of [1:1], [2:1], and [4:1] and are expressed as a mole to mole ratio of ligand to DNA single sequence (e.g., [2:1] = 10 μM DB2120 to 5 μM A4T4).
Figure 4.5 Circular dichroism titration spectra of DB2120 with A4GT4. Ligand was titrated until no further increase in the induced CD signals were observed. The insets show the ligand:DNA ratio. The strong induced CD signal indicates a very strong complex formation in the DNA minor groove.

4.4.4 Circular dichroism

Circular dichroism (CD) is a powerful technique to detect conformational changes of biomolecules, as well as small molecule binding modes using pattern recognition.[42, 43] Figure 4.5 shows the titration CD spectra of DB2120 with the A4GT4 sequence. The compound exhibits strong, positive induced CD signals between 300 and 400 nm that have been found to be indicative of minor-groove binding, as expected from the structure. Small and consistent changes in the CD spectral region of DNA (230-290 nm) are observed with incremental titration of ligands, indicating only minor conformational changes in DNA upon complex formation.

4.4.5 DNase I footprinting

Do these results on sequence specificity by DB2120 hold up with a much longer DNA? To answer this question, binding to a long DNA fragment was also assessed using DNase I footprinting experiments. A DNA fragment containing A4CT4A4GT4, A4GCT4A4GCT4, A4T4A4T4, AATTGCAATTAATTGCAATT, and AATTCAATTAATTG AATT sequences
Figure 4.6 DNase I footprinting densitometry analysis.
Binding of the phenyl (DB2119) and pyridyl (DB2120) compounds to the 3'-end radio-labeled 208 bp DNA fragment (Experimental details in Supplementary Material) containing the five different DNA sequences (A4CT4-A4GT4, A4T4-A4T4, A4GCT4-A4GCT4, AATTCAATT-AATTGAATT, and AATTGCAATT-AATTGCAATT, localized as black lines) was quantified relative to unbound DNA treated with DNase I in similar manner. The primary bound portions of DNA are localized as red boxes. The single G-C bp recognition by the pyridyl compound is visually apparent.

(black lines, Figure 4.6) was prepared (Experimental Section, Supplementary Material) and incubated with increasing concentrations of DB2120 and DB2119 prior to mild DNase I treatment (Figure 4.6, experimental gel electrophoresis results shown in Figure 4.7). The corresponding densitometric analyses for DB2120 and DB2119 (Figure 4.6) show very different results for the two compounds that structurally differ in only by replacing a CH by N. Strong and selective binding of DB2120 to the AATTCAATTAATTGAATT site contrasts with the absence of binding to AATTGCAATTAATTGCAATT. Binding is also seen for DB2120 with A4CT4A4GT4 but no significant binding to a similar A·T sequence without a G·C base pair (A4T4A4T4). In contrast, DB2119 shows weaker binding to all of the binding sites under the gel conditions, in agreement with the biophysical methods discussed above.
Figure 4.7 DNase I footprinting experiments.

The 208bp DNA fragment containing the 5 sequences of interest (AATTGCAATT•AATTGCAATT, AATTCAATT•AATTGAATT, A4GCT4•A4GCT4, A4CT4•A4GT4 and A4T4•A4T4) was incubated with increasing concentrations of DB2119 and DB2120, as indicated on the top of the lanes (µM) prior to being subjected to DNase I mild digestion. The digested products were separated on an 8% polyacrylamide gel containing 8 M urea. The sequence was numbered and identified relative to classic DMS-piperidine treated DNA samples identifying guanines as strong, cleaved bands and adenines as much weaker bands (G-track lanes).
4.4.6 Molecular modeling

For molecular insights into the structural parameters dictating G·C bp recognition by DB2120, molecular modeling studies were conducted with the A4GT4 sequence using experimentally derived information. Docking analysis reveals that DB2120 complements the DNA minor-groove shape and makes excellent van der Waals interactions with the walls of the minor groove (Figure 4.8A). The two terminal ABP modules are oriented parallel to the groove walls, while the central pyridine ring is deeply embedded in the minor groove. As expected, the inner-facing pyridine nitrogen is involved in H-bonding to the central G-NH that extends into the minor-groove (Figure 4.8B, middle). Additional H-bonds are formed between the benzimidazole and amidine groups with AT bp (Figure 4.8B, left and right). This orientation provides excellent minor-groove recognition units for both G·C and A·T bp. In addition, the linker flexibility allows molecular twist along the groove that orients the two ABP units in such a way that the individual units on either side of the pyridine ring form strong interactions with both strands of the duplex. The specific interactions observed between DB2120 and the minor groove of A4GT4 dictate the sequence selectivity observed for this compound and provide a rational explanation for the observed mixed-sequence DNA recognition.

4.5 Discussion

In summary, new ideas for the design of a variety of cell-permeable compounds for sequence specific recognition of DNA are needed to take advantage of the wealth of new and expanding genomic information. The variety is required because of the different uptake characteristic of different cells and the different pharmacokinetics or ADME properties of compounds of different structure.[30, 44] In this project, A·T bp recognition groups were linked by test modules for potential G·C bp recognition with the long-term objective of developing a new
Figure 4.8 Molecular docking model of DB2120 with the A4GT4 sequence.

(A) The model represents the minor (left) and major (right) groove views of DB2120 with A4GT4 sequence. The A-T DNA bases are represented as stick models in red and the central G-C bp is green. The DNA backbone is represented as a tube in gray color. DB2120 is represented in a space-filling model with the carbons light blue, nitrogen dark blue, hydrogens orange and oxygen red. (B) Important interactions between different sections of DB2120 and DNA are illustrated. The central G-C bp is used as the reference for base numbering of the leading and complementary strands. The benzimidazole ring NH and the amidine unit at the top of DB2120 in this orientation forms strong interactions with the carbonyl groups of T+2 and T+3 bases in the minor groove of the leading strand (left). The pyridine nitrogen makes strong interactions with the G-NH2 (ball and stick) in the minor groove (middle). The benzimidazole NH at the bottom end forms bifurcated interactions with A-2 N3 on the 5'-AAACTTT-3’ strand and T+3 carbonyl groups of the 5'-AAAGTTT-3’ strand, respectively. The amidine unit forms strong interactions with the carbonyl group of T+4 of the AAACCTT strand (right). All the interactions contribute to the overall stability of complex formation with the pyridine ring conferring G-C recognition as it was designed to do.

A paradigm for sequence specific DNA minor groove recognition using biologically-compatible heterocyclic cations.

The pyridyl analogue of DB2119, DB2120, is a significant advance in the design of compounds that are highly selective for single G-C bp recognition with flanking A-T sequences. As noted above, such sequences appear often in the kDNA of parasitic microorganisms as well as
TF promoter sites. SPR results clearly show that the slow dissociation rate of DB2120 from the A4GT4 sequence is the primary factor in the strong and selective binding observed. The compound binds rapidly and forms numerous van der Waals and H-bond interactions (Figure 4.8) that must be disrupted to dissociate the complex. The absence of the intervening G-C bp in the A4T4 and inosine motifs clearly eliminates the H-bonding interaction between the G-NH$_2$ group and DB2120 and gives faster dissociation and weaker binding. The DNA interactions of DB2120 offer insights that can be integrated into future compound design schemes. The significant loss in binding selectivity by DB2370, due to the absence of the -O-CH$_2$- linker, also indicates that not only the molecular geometry but also the flexibility of DB2120 play very critical roles for the selective recognition of the G-NH$_2$ group. The sequence composition, which primarily governs the local microstructure of the minor groove, also strongly dictates the strength and stoichiometry of binding. While the narrower minor groove of pure A-tract motifs with enhanced negative electrostatic potential offer a more conducive environment for a highly favorable monomer complex formation, the somewhat wider grooves of other A·T or more G·C bp-containing sequences [33] may promote strong dimeric complex formation.

The SPR, mass spectrometry, and DNase I footprinting results strongly support mixed bp DNA recognition by DB2120 but not the phenyl derivative, DB2119. This is particularly important since the DNase I cleavage experiments are done with long DNA sequences that are quite different from the shorter segments used in our other biophysical studies. An example densitometry trace of a DNase I footprinting gel that compares DB2119 to results with DB2120 (Figure 4.6) shows that AAAACTTTTAAAAGTTTT and AATTCAATTAATTGAATT sequences are the strongest binding sites for DB2120 with much weaker binding to two GC and pure AT sequences.
The molecular model in Figure 4.8 provides information to explain the strong and specific binding of DB2120 to the A4GT4 site. The pyridyl-N forms a strong and direct H-bond to the G-amino-NH that projects into the minor groove. The two -O-CH$_2$- flexible linking groups of DB2120 allow rotations of the terminal phenyl-benzimidazole–amidine units such that they are able to fit snugly into the minor groove in the flanking A-tract sequences. Without these groups, DB2370 has poor GC selectivity. Both benzimidazole-NH and amidine-NH groups form strong, direct H-bonds with C=O groups of thymines in the minor groove (Figure 4.8). The amidine cationic charges are well-placed to form ionic interactions with the anionic DNA backbone. The entire DB2120 molecule is involved in optimizing the complex with the DNA minor groove. An electron density calculation (Figure 4.1) shows the positive electron density on the amidines and benzimidazole-NH groups, and the negative charge density on the pyridyl-N help account for strong binding between DB2120 and the DNA minor groove.

4.6 Conclusion

The combined biophysical results reveal that effective recognition of A·T-rich sequences with intervening G·C bps can be achieved with heterocyclic cations that have appropriately placed H-bond acceptors flanked by A·T bp recognition modules and appropriate flexibility. This information should assist significantly in the rational design of further potent heterocyclic cations to target larger mixed DNA sequences with more G·C base pairs.

4.7 Materials and methods

4.7.1 DNA oligonucleotides

For the thermal melting and circular dichroism spectroscopy experiments, hairpin DNA oligomers used were A4GT4 [5’-CCAAAAAGTTTTGCTCTCAAAAACCTTTTGG-3’], A4IT4 [5’-CCAAAAITTTTGCTCTCAAAAACCTTTTGG-3’], A4T4 [5’-
CCAAAATTGTGCTCTCAAAAATTTTG-3’], and with A4GCT4 [5’-CCAAAAGCTTTTGCTCTCAAAAAGCTTTTG-3’], with the hairpin loop underlined (Figure 4.1). Lyophilized DNA oligomers were purchased from Integrated DNA Technologies, Inc. (IDT, Coralville, IA) via HPLC purification. Doubly distilled water was added to the solid DNAs to bring the concentration to approximately 1.0 mM, based on the reported amount of DNA from IDT. The molar concentrations of these hairpin DNAs were then determined using a Cary 300 Bio UV–Vis spectrophotometer (Varian, Walnut Creek, CA) at 260 nm based on the molar extinction coefficients ($\varepsilon = 260$ nm) calculated by the nearest-neighbor method.[45]

**4.7.2 Thermal melting (Tm)**

Thermal melting experiments were performed on a Cary 300 Bio UV-Vis spectrophotometer (Varian). The concentration of each hairpin DNA sequence was 3 μM and experiments were in buffer (50 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA, pH 7.4) in 1 cm quartz cuvettes at 2:1 ligand-DNA ratio. The DNA solutions were annealed prior to being tested. The spectrophotometer was set at 260 nm, 0.5 °C/min increase beginning at 25 °C, well below the DNA melting temperature and ending well above it or at 95 °C. The absorbance of the buffer was subtracted, and a graph of normalized absorbance vs. temperature was created using KaleidaGraph software. The $\Delta T_m$ values were calculated using a combination of the derivative function and estimation from the normalized graphs.

**4.7.3 Biosensor surface plasmon resonance (SPR)**

SPR measurements were performed with a four-channel BIAcore T200 optical biosensor system (GE Healthcare, Inc., Piscataway, NJ). A streptavidin-derivatized (SA) sensor chip was prepared for use by conditioning with a series of 60 s injections of 1 M NaCl in 50 mM NaOH (activation buffer) followed by extensive washing with HBS buffer (10 mM HEPES, 150 mM NaCl).
NaCl, 3 mM EDTA, and 0.05% P20, pH 7.4). Biotinylated-DNA samples (A4T4, A4GT4, A4IT4, and A4GCT4 hairpins) of 25-30 nM were prepared in HBS buffer and immobilized on the flow cell surface by noncovalent capture as previously described.[36, 37] Flow cell 1 was left blank as a reference, while flow cells 2-4 were immobilized with DNA by manual injection of DNA stock solutions (flow rate of 1 μL/min) until the desired amount of DNA response units (RU) was obtained (320 - 330 RU). Ligand solutions were prepared with degassed and filtered 50 mM Tris-HCl buffer (pH 7.42) by serial dilutions from a concentrated stock solution. Typically, a series of different ligand concentrations (1 nM to 1 μM) were injected over the DNA sensor chip at a flow rate of 100 μL/min until a constant steady-state response was obtained (3 min), followed by buffer flow for ligand dissociation (10 - 60 min). After each cycle, the sensor chip surface was regenerated with a 10 mM glycine solution (pH 2.5) for 30 s followed by multiple buffer injections to yield a stable baseline for the following cycles. RU_{obs} was plotted as a function of free ligand concentration (C_{free}), and the equilibrium binding constants (K_A) were determined either with a one-site binding model (K_2 = 0) or with a two-site model, where r represents the moles of bound compound/mol of DNA hairpin duplex and K_1 and K_2 are macroscopic binding constants.

\[ r = \frac{[(K_1 \cdot C_f) + (2K_1 \cdot K_2 \cdot C_f^2)]}{1 + (K_1 \cdot C_f) + (K_1 \cdot K_2 \cdot C_f^2)} \]  

(1)

RU_{max} in the equation was used as a fitting parameter, and the obtained value was compared to the predicted maximal response per bound ligand to independently evaluate the stoichiometry.[37] Kinetic analyses were performed by globally fitting the binding results for the entire concentration series using a standard 1:1 kinetic model with integrated mass transport-limited binding parameters as described previously.[37]
4.7.4 Mass spectrometry (ESI-MS)

Electrospray ionization mass spectrometry experiments were performed using a Waters Micromass Q-ToF (Milford, MA) in negative ion mode and MassLynx 4.1 software for data analysis. Hairpin DNA sequences (Integrated DNA Technologies) were designed to maintain flanking base pairs in the duplex stem while altering only the target binding site with A4T4: [5'-GCAGAGATTTTTCGAGTTTTCCGAGAAAATTTTCG-3'], A4GT4: [5'-CGAAGGTTTTTCGGGGGCGAAAAGCTTTTCG-3'], A4GCT4: [5'-CGAAGGCTTTTCCGAAAAAGCTTTTCG-3'], and GC internal standard: [5'-GCAGAGGCTTTTTCGCGCGCAGGCG-3'] with the hairpin loop underlined. Guanine base pairs not in the target site nor adjacent to the target site were substituted with an inosine to create sequences with distinguishable molecular weights. Buffer solution of 0.15 M ammonium acetate buffer (pH 6.7) was prepared in doubly distilled water and filtered using a 0.22 µm hydrophilic cellulose filter (Millipore, Darmstadt, Germany). DNAs were dissolved in 0.15 M ammonium acetate buffer to a concentration near 1.0 mM and dialyzed in a 1000 Da MWCO dialysis membrane tubing (Spectrum Laboratories, Rancho Dominguez, CA) at 4 ºC with 3x buffer exchange. Following dialysis, DNA concentrations were spectroscopically determined using extinction coefficients provided by IDT based on the nearest neighbor method.[45] DNA sequences were heated to 95 ºC and quenched on ice to initiate hairpin formation. DNAs were prepared in 0.15 M ammonium acetate buffer to concentrations of 5 µM. Titrated amounts of compound were added based on a mole to mole ratio of compound to a single DNA sequence (e.g. 10 µM compound to 5 µM A4GT4 to create a [2:1] concentration ratio) and 5% methanol (v/v) to increase sample volatility. Capillary, sample cone, and extraction voltages were set to 2100 V, 25 V, and 1.5 V, respectively. Source block and desolvation temperatures were respectively set to 80 ºC and 110 ºC. Collision energy
was set to 2 V. Cone gas and desolvation gas flows were at 50 L/h and 350 L/h. Samples were injected at a flow rate of 5 µL/min and run for ≈ 10 min until the MassLynx chromatogram reached stabilization. Scanned peaks were analyzed over $m/z$ 300 – 3,000 range and the final 2 min averaged. The most abundant peak intensities observed belonged to charge states of -3 to -6. Spectra were deconvoluted using the ‘Maximum Entropy 1’ function (MassLynx) which multiplies each peak ($m/z$) by its respective charge ($z$) to produce a single-charge peak for each DNA and/or complex and allow for a simplified comparison. Between each set of titrations, the instrument was thoroughly cleaned using HPLC-grade methanol, acetonitrile, methanol/acetonitrile mixture (1:1) and finally flushed with 0.15 M ammonium acetate buffer. Estimated peak intensities for the titration comparisons are within ± 3% of peak values.

4.7.5 Circular dichroism spectroscopy (CD)

Circular dichroism experiments were performed on a Jasco J-810 CD spectrometer in 1 cm quartz cell at 25 ºC. The hairpin DNA (5 µM) sequence in buffer (50 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA, pH 7.4) was added to the cell prior to the experiment and then the compound was added to the hairpin DNA solution and incubated for 10 min to achieve equilibrium binding for the DNA complex. For each titration point, four spectra were averaged from 500 to 220 nm with scan speed 50 nm/min, and a response time of 1 s. Buffer subtracted graphs were created using the KaleidaGraph software.

4.7.6 DNase I footprinting experiments

[5’-gatcggtgcaAATTTTgcgtcAAAAGTTTTtggtgcgAAAAGCTTTTccggggaATTGAAATTg gcgccAATTGCAATTcggcc3’] and [5’- gatcgccgAATTGCAATTggccgccAATTCAATTgccgg ggAAAAAGCTTTTgcaccAAAACTTTTggacgcAAAATTTTgcac-3’] oligonucleotides
(Eurogentec, Belgium) were annealed and cloned in BamH I site of the pBSIISK vector (Stratagene, La Jolla, CA). The DNA fragment encompassing this sub-cloned sequence was obtained from ClaI-NotI, double digestion and 3’-end-labeled using γ-[32P]-dCTP (Perkin-Elmer, France) and Klenow enzyme. The resulting 208-bp radio-labeled DNA fragment was then purified and the DNase I footprinting experiments performed as previously described.[48]

4.7.7 Structural model calculation methods

Molecular modeling studies were initiated with conformational analysis of the tested compounds with a molecular mechanics MMFF approximation level with the Spartan’10 software package (Wavefunction Inc. Irvine, CA). The Spartan’10 software package was employed to optimize the final geometry by using ab initio calculations with density functional theory (DFT), B3LYP at the 6-31G* approximation level. The alignment of compounds was done using the ‘Align Database’ option of the QSAR module in SYBYL-X1.2 software package on a Windows workstation.[47]

Molecular docking and visualization studies were performed with the SYBYL-X1.2 software. The initial DNA duplex [5’- d(GCAAAAGTTTTCC)-3’· 5’-d(GGAAAACCTTTTG)-3’] was constructed in the ‘Biopolymer-Build DNA Double Helix’ module employing regular B-DNA parameters. The DNA was next energy minimized for a maximum of 100 iterations using the conjugate gradient algorithm and Tripos force field, with a termination gradient of 0.1 kcal/mol Å. The three-dimensional structure of DB2120 was built, assigned Gasteiger-Hückel charges and minimized using the Tripos force field until a terminating conjugate gradient of 0.01 kcal/mol Å or the maximum 1000 iterations was reached.[47]

During the docking process, DB2120 and the DNA were in separate molecular areas within the SYBYL graphical user interface. DB2120 was manually inserted into the DNA minor groove
and the ‘Flexidock’ module was then employed. Ten different random numbers were designed and employed by the genetic algorithm, one at a time, for a total of 10 docking trials. 456,000 generations were calculated and assigned for docking DB2120 into the DNA minor groove.[47-49] A large amount of generations ensured that the lowest-energy conformations were obtained. Both the ligand and the bound DNA were permitted torsional flexibility in the docking process. Atomic charges were computed using Kollman all-atom for DNA and Gasteiger-Hückel for the ligand. All possible H-bond sites were selected for the DNA-ligand complex. From each docking, the 20 lowest-energy structures were selected.

4.8 References and notes


47. SYBYL Molecular Modeling Software, Version X1.2, Tripos Inc.: St. Louis, MO, **2010**.


5 DNA MICROSTRUCTURE INFLUENCES BINDING OF SMALL MOLECULES DESIGNED TO TARGET MIXED-SITE SEQUENCES

Sarah Laughlin-Toth, E. Kathleen Carter, Ivaylo Ivanov, W. David Wilson


Copyright © 2016 Nucleic Acids Research

My contribution to this chapter was in sample preparation, performance and analysis of the samples using ESI mass spectrometry and surface plasmon resonance, and writing.

5.1 Abstract

Specific targeting of protein-nucleic acid interactions is an area of current interest, for example, in the regulation of gene-expression. Most transcription factor proteins bind in the DNA major groove; however, we are interested in an approach using small molecules to target the minor groove to control expression by an allosteric mechanism. In an effort to broaden sequence recognition of DNA-targeted-small-molecules to include both A·T and G·C base pairs, we recently
discovered that the heterocyclic diamidine, DB2277, forms a strong monomer complex with a DNA sequence containing 5’-AAAGTTT-3’. Competition mass spectrometry and surface plasmon resonance identified new monomer complexes, as well as unexpected binding of two DB2277 with certain sequences. Inherent microstructural differences within the experimental DNAs were identified through computational analyses to understand the molecular basis for recognition. These findings emphasize the critical nature of the DNA minor groove microstructure for sequence-specific recognition and offer new avenues to design synthetic small molecules for effective regulation of gene-expression.

5.2 Key words

Competition ESI mass spectrometry; Competition SPR; DNA shape; Microstructure; Mixed site; Molecular dynamics; Sequence specificity

5.3 Abbreviations

Base pair (bp); Deoxyribonucleic acid (DNA); Electrospray ionization mass spectrometry (ESI-MS); Molecular dynamics (MD); Surface plasmon resonance (SPR); Transcription factor (TF)

5.4 Introduction

Regulation of the binding affinity in protein-nucleic acid complexes is an attractive concept for development of novel therapeutics and agents for control of gene expression.[1-4]. Several innovative approaches have used small molecules to target disease-associated DNA binding transcription factors or TFs.[5-15] Most TFs of interest bind in the major groove [16] and an alternative approach to control expression is to use small molecules to modulate TF activities by interacting directly with the minor groove of DNA where most of these agents bind.[17-19] There are two possible mechanisms whereby a minor groove binding compound could disrupt protein-
nucleic acid interactions in the major groove to modulate TF association. First, when bound to the minor groove, the small molecule could distort DNA so that the structure of the TF no longer complements its target recognition site, such as an allosteric inhibition mechanism.[20, 21] Alternatively, direct competition is another possible mechanism which may be significant for TFs that position side chains into or near the DNA minor groove. By knowing how small molecule inhibitors recognize DNA, it is possible to preemptively block TF binding to DNA. Our main goal is to understand, in detail, the minor groove binding variations of synthetic small molecules with different DNA sequences and how they vary with sequence-dependent DNA structure.

Small molecules that bind in the minor groove of DNA have been validated for this approach from studies using synthetic polyamides.[22-24] However, polyamides have limitations such as aggregation and cell uptake and a wider variety of agents is needed for diverse biological systems.[25, 26] We are approaching this problem with a class of sequence-specific, DNA-targeted minor groove binders based on a heterocyclic cation design since these compounds have shown good cell uptake and biological properties through human clinical studies.[27, 28] Few non-polyamide minor groove agents, including heterocyclic diamidines, have been identified to selectively recognize mixed, A·T and G·C base pair-containing DNA sequences.[29, 30] This constitutes a significant barrier to progress in the area of designed synthetic agents for the disruption of TF-DNA complexes. To interact with the edges of A·T base pairs in the minor groove, compounds must have hydrogen bond donor groups for the thymidine carbonyl and an N3 of adenine acceptor. To recognize a G·C base pair, the compound must have an acceptor to hydrogen bond to the guanine NH$_2$ group. It is also critical that a successful small molecule have the appropriate shape and charge to complement the DNA minor groove.[31, 32]
A synthetic effort has led to cationic diamidines that strongly and selectively recognize the minor groove in mixed-site DNA sequences.\cite{33} The lead compound in this development is DB2277, which contains a nitrogen hydrogen bond acceptor in an aza-benzimidazole (Figure 5.1). Strong binding of DB2277 requires the 2-amino group of guanine and suggests an aza-N-G-NH$_2$ hydrogen bond. These observations show that DB2277 binds best to mixed-site sequences with a single G-C base pair flanked by A-T base pair sites.\cite{34} Key questions in the design effort for new mixed-sequence minor groove compounds that recognize G-C base pairs with flanking A-T base pairs: In addition to monomer binding to recognize a single G-C base pair, can the compound form dimers to recognize two G-C base pair sequences? What is the effect of the flanking A-T base pairs? How could this influence binding affinity?

**Figure 5.1 Test compound and sequences for ESI-MS.**

(A, B) Structures of DB2277 and (C) DNA sequences used to screen for binding with the DB2277 using ESI-MS.
To address these questions, a systematic set of DNAs were tested with DB2277 and their interactions, affinities and stoichiometries were investigated. The composition of A·T base pairs was maintained \( (i.e. \) number of A·T base pairs per binding site) to see how interactions vary due to A·T base pair order with one and two, central G·C base pairs. Electrospray ionization mass spectrometry (ESI-MS) and surface plasmon resonance (SPR) were used to examine stoichiometry and binding behavior for the DNA-DB2277 complexes. Significant variations in affinity and stoichiometry for binding of DB2277 to the different, closely related sequences were observed. To help understand these sequence-dependent variations, extensive molecular dynamics (MD) simulations were conducted to provide specific details regarding the structural properties intrinsic to each DNA sequence that govern small molecule recognition. Large differences in the local DNA structure were observed with these closely related sequences and the differences correlate with observed differences in DB2277 binding affinity and stoichiometry. The results described here provide new and fundamental information in design research for DNA sequence-specific recognition and structural complementarity between a small molecule and its target site.

### 5.5 Results and discussion

#### 5.5.1 Competition electrospray ionization mass spectrometry identifies new interactions

Competition electrospray ionization mass spectrometry (ESI-MS) can simultaneously identify affinity, stoichiometry and cooperativity in multiple DNA-small molecule interactions.[35] The molecular weights of all possible DNA species are controlled through sequence modifications, making each sequence distinguishable (Figure 5.1C). The target binding sites of the tested DNA sequences are listed as DNA 1-9 in Figure 5.2. DNA 2, which contains the AAAAGTTTT target site, was used as a reference to compare binding due to the extensive data available for DB2277
Figure 5.2 Representative ESI-MS spectra of DB2277 binding with DNA.

(A) DNA sequences in the absence of DB2277 with m/z 1,580-1,780 signifying -6 charged species. Molar concentration ratio of [DB2277] to [DNA] expressed as: (B) [0.5 to 1], (C) [1 to 1], (D) [1.5 to 1], (E) [2 to 1], (F) [2.5 to 1], and (G) [3 to 1]. Concentrations of DNA were fixed at 5 µM. Unbound DNA, 1 to 1, and 2 to 1 complexes labelled above respective boxes.
Figure 5.3 Comparison of deconvoluted peak intensities from titrated DB2277 with DNA. Comparison of relative peak intensities for AAAATTTT, AAAAGTTTT, AAAACTTTT, AAAAGCTTTT, AAAAGCAATT, AATTGCAATT, AATTCGAATT, AATTCGAATT, ATATGCATAT, and their DB2277 complexes. Molar ratios are expressed as [DB2277] to [DNA]. Concentrations of DB2277 increased from [0.5 to 1] to [3 to 1] with a fixed DNA concentration of 5 µM. Complexes are expressed as unbound DNA (empty), 1:1 (checkered), and 2:1 (solid) complexes.

with that and similar sequences. Target binding sites were designed to test the interactions of the compound with a range of closely related DNA sequences. Each sequence has two set of A·T base
pairs as AAAA, AATT or ATAT. Only DNA 1 lacks a central G·C base pair. Two categories of mixed sequences are grouped with either one or two G·C base pairs.

**Figure 5.2** illustrates the changes in relative peak intensity for DB2277-DNA complexes. Changes in the relative intensities are based on the binding of DB2277 to DNA and compared to a reference sequence present in the sample. In the presence of DB2277, peak intensities for free DNA disappear while intensities for DB2277-DNA complexes emerge. For instance, in Figure 5.2B, half of the unbound AAAAGTTTT (DNA 2) is present at a concentration molar ratio of [0.5 to 1] as well as ≈ 50% of 1:1 complex. As the concentration of DB2277 is increased, unbound AAAAGTTTT decreases with a concurrent increase in 1:1 binding. Based on these results, sequences which contain A-tracts (DNA 1-3, 9) prefer 1:1 binding. The -6 charged species were used for illustrative purposes in Figure 2 since they were the most abundant of the multiply charged species. Relative binding affinities were measured using deconvoluted spectra which takes into account all multiply charged species (**Figure 5.3**).

Sequences in the AATT subcategory (DNA 5-8) allow us to examine the transition from A-tracts to sites with an ApT base pair step. Surprisingly, for this closely related sequence, the ESI-MS results show that 2:1 binding is strongly preferred over 1:1 complex formation for AATT DNAs. Finding such preference was especially interesting since many minor groove binding compounds cannot differentiate among A·T base pair sites.[36, 37] Based on the results from AATT sequences, one might expect to find strong 2:1 complexes formed between DB2277 and ATAT sequences since alternating A·T sites have wide minor grooves similar to AATT. In another surprise, the compound preferentially formed a 1:1 complex with ATATGCATAT (**Figure 5.2**). DB2277 can bind tightly to sites with a single G base flanked by A-tract sites [31]; however, little is known regarding sequences with two G·C base pairs. In summary, the ESI-MS results show that
monomer complexes are the preferred systems for AAAA and ATAT base pair sites that flank a core G·C base pair whereas 2:1 complexes are preferred for sequences containing AATT sites.

5.5.2 **Surface plasmon resonance confirms sequence specific behavior identified by ESI-MS**

Surface plasmon resonance (SPR) is a powerful method to define the thermodynamic and kinetic properties of biomolecular interactions.[38, 39]. In our experiments, increasing concentrations of DB2277 were injected over a set of immobilized DNA sequences. Binding curves for DB2277 are shown in **Figure 5.4**. Binding affinities for steady-state equilibrium and kinetics-fitted analyses are compared in **Table 5.1** as well as the binding on and off-rates. The unexpected 2:1 binding of AATTGCAATT was of considerable interest, especially since it contains two central G·C base pairs. To directly compare flanking base pair sequence and its role in small molecule recognition, binding of DB2277 was measured with AATTGCAATT, as well as with AAAAGCTTTT and ATATGCATAT since all three sequences contain the same G·C base pair core. A representative sensorgram of DB2277 binding to AATTGCAATT is shown in **Figure 5.4A**. Results of AATTGCAATT binding from SPR are in direct agreement with those obtained from ESI-MS with two binding sites, $K_{D1}$ and $K_{D2}$, near 15 nM and 30 nM, respectively. Also in agreement, a strong 1:1 complex for DB2277 and AAAAGCTTTT was observed ($K_D \approx 50$ nM). With the ATATGCATAT sequence, a weaker 1:1 complex was formed ($K_D \approx 100$ nM) and a second, much weaker 2:1 complex was detected at high compound concentrations, which were well above the first $K_D$. These results are in agreement with those obtained by ESI-MS for preferred 1:1 binding.

Competition SPR [40] was used to measure the binding of three single G·C base pair sequences against the original immobilized DNA. Similar to direct-binding SPR experiments, the
Figure 5.4 SPR sensorgram and binding curves of DB2277 and DNA. 
(A) SPR sensorgram of AATTGCAATT binding DB2277. Injected concentrations of DB2277 shown are 5, 10, 30, 70 and 200 nM. (B) Steady-state fits for binding with AAAAGCTTTT, AATTGCAATT and ATATGCATAT fit using a two site binding model. (C) Competition SPR steady-state fits of competitor DNA sequences AAAAGTTTT, AATTGAATT and ATATGATAT fit using a one-site binding model.

DB2277 was added to the sample solution and the observed response at steady-state is plotted to determine the binding constant. In the competition SPR experiments, the compound was held at a fixed concentration while the competing DNA was added to the sample solution. The observed response, however, decreased as concentrations of competing DNA were increased, which resulted in less available free compound in solution. Calculated dissociation constants of AAAAGTTTT, AATTGAATT and ATATGATAT were determined from Figure 5.4C and are listed in Table 5.1.
Table 5.1 Comparison of kinetics-fitted and steady-state equilibrium binding constants (KD x 10^-9 M).
Kinetic rates and fits were determined using direct-binding SPR. Steady-state fits were compared using both direct-binding and competition binding SPR. Sequences with multiple binding constants are listed as KD1 and KD2.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Kinetic Rates</th>
<th>Dissociation Constants</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ka (10^6·M^-1·s^-1)</td>
<td>kd (10^-1·s^-1)</td>
</tr>
<tr>
<td>AAAAGCTTTT</td>
<td>5.3 ± 0.7</td>
<td>2.4 ± 0.5</td>
</tr>
<tr>
<td>ATATGCATAT</td>
<td>1.6 ± 0.2</td>
<td>1.4 ± 0.2</td>
</tr>
<tr>
<td>AATTGCAATT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KD1</td>
<td>1.0 ± 0.2</td>
<td>0.1 ± 0.01</td>
</tr>
<tr>
<td>KD2</td>
<td>0.5 ± 0.08</td>
<td>0.2 ± 0.02</td>
</tr>
<tr>
<td>AAAAGTTTTT</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>ATATGATAT</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>AATTGAATT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KD12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KD1</td>
<td>4.4 ± 1.0</td>
<td>1.3 ± 0.4</td>
</tr>
<tr>
<td>KD2</td>
<td>3.1 ± 1.2</td>
<td>1.6 ± 0.7</td>
</tr>
<tr>
<td>b KD12</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ND Not determined
a Determined using competition SPR
b KD12 value determined by √KD1 · KD2 with KD1 and KD2 values obtained through direct-binding SPR
c Value determined by direct-binding SPR and fit with one-site binding model
d Value determined by direct-binding SPR and fit with two-site binding model
The observed response (RU_{obs}) was plotted as a function of competitor DNA concentration [38, 39] and fit to a 1:1 binding model using Equation 5.2.

Using competition SPR, the strongest 1:1 complex formed within this DNA series was AAAAGTTTTT (K_D \approx 4 \text{ nM}) as expected from ESI-MS and literature.[33, 34] A 10-fold weaker complex was formed with the compound and AAAAGCTTTT. Results by competition SPR for AATTGAATT show it formed a strong 1:1 complex with a binding constant of 40 nM. Since its two G-C counterpart (i.e. AATTGCAATT) forms both 1:1 and 2:1 complexes, a direct-binding SPR approach was also used to compare the binding affinities and determine if multiple binding modes occur. In this experiment, AATTGAATT forms both 1:1 and 2:1 complexes with K_D values near 25 nM and 60 nM, respectively. Competition SPR results were fit using a one-site binding model; however, when more than one compound binds (e.g. AATTGAATT) the calculated value for two binding constants is K_{D12} or \sqrt{K_{D1} \cdot K_{D2}}. Various forms of analyzing DB2277 with AATTGAATT were compared, such as direct-binding SPR, competition SPR, and using one and two-site binding models. Likewise, a comparison of kinetics-fitted and steady-state binding constants, determined by the two SPR methods, are in excellent agreement for AATTGAATT (Table 5.1).

Interestingly, association rate constants for 1:1 binding of DB2277 to DNA are similar for all sequences (k_a \approx 10^6 \text{ M}^{-1}\text{s}^{-1}) whereas association rate constants are comparatively slower for the second DB2277 molecule binding with AATTGCAATT or AATTGAATT. On the other hand, the second off-rate for DB2277 is faster than the first off-rate. The calculated binding constants of AATTGAATT are similar to those for AATTGCAATT (Table 5.1) and further suggest a binding mechanism for the compound unique to the AATT sequences. Results obtained by ESI-MS and SPR are in excellent agreement and indicate binding of DB2277 differs when the order of flanking...
A·T base pairs is varied. Clearly, the exact order of the flanking bases influences binding of the test compound since both global and local structure of the DNA are contingent on base pair sequence. In order to probe the basis of these binding differences in molecular detail, we turned to molecular dynamics simulations.

5.5.3 Molecular dynamics identifies microstructural differences in the experimental DNAs

Honig, Rohs and others have shown that the local structure within the DNA minor groove can depend on base pair sequence.[41-44] Such microstructural variations may explain why binding of DB2277 varies greatly even though base pair composition is maintained. To elucidate how DNA microstructure may influence small molecule binding, extensive molecular dynamics (MD) simulations of a systematic set of closely related experimental DNA sequences with central G·C base pairs flanked by A·T base pairs of different sequence were carried out. Variations of the resulting structures in the MD trajectories were analyzed with Curves+ [45] to predict their roles in DB2277-DNA recognition. We measured helical parameters with specific emphasis on minor groove width and depth (Figures 5.5-5.7). Additionally, simulations for select sequences were repeated and analyzed over 100 ns intervals to validate structure and flexibility convergences (Figures 5.8, 5.9). This is the first highly detailed structural analysis on the effects of systematic changes in DNA sequence focused on one and two core G·C base pairs flanked by varying sequences of A·T-rich sites over long trajectories for 200 ns.[46, 47]

Varying the A·T flanking sequence around the core G·C produced a surprisingly large deviation in minor groove widths and depths. Comparison of the flanking sequences reveals unique groove width variations. For example, the 2D contour histogram for the minor groove width of AAAAGTTTT (Figure 5.5C) indicates a high probability of adopting a narrow (4.5 Å) and deep
Figure 5.5 2D contour histograms for minor groove widths of unbound DNA. Representative manta plots for (A) 13 and (B) 14 base pair sequences. Number of points along the spline of the backbone was increased from 10 to 50 to produce 601 points for 13 base pair sequences and 651 points for 14 base pair sequences. Using these two dimensional distance matrices, a smooth line was interpolated along the global minimum (red arrow) to construct contoured histograms of the minor groove width. Contour 2D histograms of minor groove widths vs. base pair sequence. (C) AAAAGTTTT, (D) AAAAGCTTTT, (E) AATTGAATT, (F) ATTGCAATT, (G) AATTGGAATT, (H) ATATGATAT, and (I) ATATGCATAT. The color gradient indicates increasing probability (navy to red) of observing a minor groove width (in Å).
(5.0 Å) minor groove with little variation along the target binding. This sequence has the highest binding affinity of all the DNA sequences investigated and is explained by the inherently narrow and deep groove pre-formed for energetically favorable binding of the compound. DB2277 and similar compounds thus bind and fit well into A-tract flanking sequences in strong 1:1 complexes.

In contrast, AATTGAATT has a strong preference toward maintaining a narrow and deep groove at the terminal AA·TT regions. The groove width increases to 8.0 Å, becoming much wider than AAAAGTTTTT, at the central G·C base pair of the sequence (Figure 5.5E). The depth of the sequence at the central G·C is also less than the AAAAGTTTTT sequence. Due to the change in groove width and depth, the central G·C therefore provides a less favorable binding site for DB2277 monomer binding. The competition SPR results have, thus, revealed sequence-specific variations in both stoichiometry and affinity. We speculate that a single DB2277 binds first with AATTGAATT and because of the wider groove and variable depth, it is possible that two molecules of DB2277 can fit the optimum groove structure by staggered stacking at the central G·C base pair.

Finally, within the single G·C base pair series, the widest and most shallow measurable groove occurs in the ATATGATAT sequence throughout the course of the trajectory. Unlike the previous two sequences, it is energetically unfavorable for ATATGATAT to exist in a deep and narrow groove conformation. Instead, there is a strong preference for the groove to remain wide (8.0 Å) and shallow (4.0 Å). The MD described combination of an intrinsically wide groove and shallow depth would be expected to bind two DB2277 molecules. Our findings, however, suggest the DNA is ill-suited for binding a curved, planar small molecule such as DB2277. Instead, the wide and shallow groove must undergo an induced fit to bind DB2277 with a high deformation energy penalty.
Figure 5.6 2D histograms of most probable minor groove widths for unbound DNA.
Most probable groove widths per base pair for (A) one G·C and (B) two G·C base pair sequences. 2D histograms of minor groove widths vs. base pair sequence. (C) AAAAGTTTT, (D) AAAAGCTTTT, (E) AATTGAATT, (F) AATTGCAATT, (G) AATTGGAATT, (H) ATATGATAT, and (I) ATATGCATAT. The color gradient indicates increasing probability (blue to red) of observing a minor groove width (in Å).
Sequences with two G·C base pairs were next simulated to better understand the structural (dis)similarities among sequences with one or two central G·C base pairs in the target binding site. Altering the core from G to GC increases the overall probability of adopting a wide and shallow groove. For instance, a contour 2D histogram of AAAAGCTTTT in Figure 5.5D indicates a higher probability of the groove width expanding to 12.0 Å compared to AAAAGTTTT. It is also less probable for AAAAGCTTTT to maintain a deep groove, but rather becomes shallowest with the additional G·C base pair in its core. This observed decrease in probability of a narrow and deep minor groove may help to explain the weaker binding found by DB2277 with AAAAGCTTTT compared to binding with AAAAGTTTT ($K_D = 49.6 \pm 1.1$ nM and $K_D = 4.4 \pm 0.7$ nM, respectively).

A wide groove also exists at the GC region in AATTGCAATT (Figures 5.5F, 5.6F), and is more probable than its AATTGAATT counterpart. In addition to the widened groove, a correlated decrease in groove depth also occurs at this region (Figure 5.7D). Comparing both histograms, it is evident that the shallowest region occurs at the GpC step and is much more pronounced in the AATTGCAATT sequence over AATTGAATT. Interestingly, altering the central G·C base pairs in the sequence to create AATTGGAATT reveals little change in width or depth of the minor groove (Figures 5.5-5.7). Little change in groove width and depth between AATTGCAATT and AATTGGAATT explains why little detectable difference occurred for binding of DB2277 to these two sequences. The MD results therefore provide a rationale for why the AATT sequences are favorable for a 2:1 complex in both ESI-MS and SPR.

Like AATTGCAATT, the ATATGCATAT sequence also has a higher probability of existing in a wider state than its counterpart ATATGATAT (Figure 5.6B). Changing G to GC in the core of the alternating A·T base pair flanking sequences stabilizes a wider minor groove. This
2D histograms of DNA minor groove depths for unbound DNA. 2D histograms of minor groove depths vs. base pair sequence. (A) AAAAGTTTT, (B) AAAAGCTTTT, (C) AATTGAATT, (D) AATTGCAATT, (E) AATTGGAATT, (F) ATATGATAT, and (G) ATATGCATAT. The color gradient indicates increasing probability (blue to red) of observing a minor groove depth (in Å).

Figure 5.7 2D histograms of DNA minor groove depths for unbound DNA.

stability is even more evident when looking at the minor groove depth histograms (Figure 5.7). For the sequence ATATGCATAT, there is a clear preference for shallow groove depth throughout the entire sequence and unlike ATATGATAT, does not break at its core G·C region. This would, therefore, indicate that a wide and stable minor groove within the two GC sequence is consistent for ATATGCATAT in having the lowest binding affinity for DB2277. Upon binding a single DB2277 molecule, the flexibility of the ATAT sequence allows it to favorably constrict to a narrow
Figure 5.8 2D histograms of minor groove width for AAAAGTTTT. (A) 100 ns original simulation (B) 200 ns original simulation (C) 100 ns for repeated simulation (D) 200 ns repeated simulation. The color gradient indicates increasing probability (navy to red) of distance in Angstroms (in Å) for each base pair.

groove, rather than binding two molecules in an unfavorable wider groove conformation. These MD simulations complement ESI-MS and SPR studies and indicate that DB277 binding should be more favorable where the minor groove is intrinsically narrow and deep and is related to the pre-organized groove width prior to binding the compound. These local structural differences also influence how and where the molecule will bind in the minor groove. It is somewhat surprising that the large diversity of microstructural characteristics, such as groove width, observed for the minor groove are not found in the major groove, which has a much more constant structure (Figures 5.10, 5.11).
A comparison of the sequences with matched flanking sites (AAAAGTTTT vs. AAAAGCTTTT) shows little variation in local DNA structure (Figure 5.5). On the other hand, a comparison of unmatched flanking sequences, for example AAAA to AATT, indicates a larger variation in microstructure which in turn governs binding stoichiometry. For sequences with AAAA sites, there are similar distributions of a well-maintained narrow groove (4.5-5.0 Å) along the target site in the region of AAAAGTTTT and AAAAGCTTTT. Likewise, there is a consistently wide groove for both ATATGATAT and ATATGCATAT. With a range of 7.0-8.5 Å, the narrowest regions occur along the ends of the target site while the widest portions are at the T to G/C transitions (i.e. TpG and TpC of the complementary strand). Alternatively, large groove
Figure 5.10 2D histograms of major groove widths for DNA sequences.
2D histograms of major groove widths. A) AAAAGTTTT B) AAAAGCTTTT; C) AATTGAATT D) AATTGCAATT E) AATTGGAATT F) ATATGATAT G) ATATGCATAT. The color gradient indicates increasing probability (blue to red) of observing a minor groove width (in Å).

width variations within the target site occur for both AATTGAATT and AATTGCAATT. Specifically in AATT regions at the ApT base steps, there exist large differences in minor groove width ($\Delta_{\text{width}} \approx 3.5 \text{ Å}$) compared to variations of $\Delta_{\text{width}} < 2 \text{ Å}$ among AAAA and ATAT flanks width. This type of intra-target-site variation is significant to the AATT sequences resulting in two bound DB2277 molecules. Current observations of a monomeric DB2277-AATT system would suggest that a wide minor groove at the core G or G·C base pair region, adjacent to a narrow AATT
Figure 5.11 2D histograms of major groove depths for DNA sequences.
2D histograms of major groove depths vs. base pair sequence. (A) AAAAGTTTT, (B) AAAAGCTTTT, (C) AATTGAATT, (D) AATTGCAATT, (E) AATTGGAATT, (F) ATATGATAT and (G) ATATGCATAT. The color gradient indicates increasing probability (blue to red) of observing a minor groove depth (Å).

site, may support staggered stacking of two DB2277 molecules at the G/C core with the unstacked ends of DB2277 in the AATT sites.

The sequence-dependency of other helical parameters for the DNA was also compared. Differences in propeller twist were very informative and the averages for each sequence are shown in Figure 5.12 (see also Tables 5.2, 5.3). For all sequences there is a characteristic “W” shape to each curves but the range varies. For instance, the degree of propeller twist for AAAAGTTTT is
Figure 5.12 Most probable propeller twist per base pair. (A) One G·C and (B) two G·C base pair sequences. When the DNA sequences experience a high propeller twist the groove width is narrow. When the propeller twist is relieved, the groove is broadened. The terminal base pair caps were removed.

large and quite constant along the target site. On the other hand, AATTGAATT has a much wider range of propeller twist along the target binding site. Interestingly, the AATTGAATT sequence is closely related to the AATTGCAATT sequence, which may partially explain the similar binding measured for both sequences by SPR (Table 5.1). In general, sequences with consecutive A or T bases (e.g. AAAAA) have steric clash due to CH₃ groups of thymidine,[48] Propeller twists in A·T base pairs form bifurcated hydrogen bonds between the NH₂ of adenosine to O4 of the adjacent thymidine on the complementary strand (Figure 5.12A) which reduces the amount of fluctuation. Likewise, sequences with AATT have two consecutive A·T base pairs and are also likely to form bifurcated hydrogen bonds; however, two fewer possible hydrogen bonds likely increases the
Table 5.2 Mean (M), standard deviation (SD), and 99.9% confidence interval (CI) of the propeller twist over 50,000 frames for the 13 base pair sequences. Central G·C base pair is shaded in gray.

<table>
<thead>
<tr>
<th>Base Pair</th>
<th>CGAAAAGTTTTCG</th>
<th>99.9% CI</th>
<th>CGAATTTGAATTCG</th>
<th>99.9% CI</th>
<th>CGATATGATATCG</th>
<th>99.9% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD)</td>
<td></td>
<td>M (SD)</td>
<td></td>
<td>M (SD)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-6.80 (28.31)</td>
<td>[-7.22, -6.38]</td>
<td>-6.71 (24.85)</td>
<td>[-7.08, -6.35]</td>
<td>-10.28 (19.28)</td>
<td>[-10.57, -10.00]</td>
</tr>
<tr>
<td>2</td>
<td>-4.72 (10.13)</td>
<td>[-4.87, -4.58]</td>
<td>-7.57 (10.07)</td>
<td>[-7.72, -7.42]</td>
<td>-10.21 (9.71)</td>
<td>[-10.36, -10.07]</td>
</tr>
<tr>
<td>3</td>
<td>-14.95 (8.18)</td>
<td>[-15.08, -14.83]</td>
<td>-18.01 (8.12)</td>
<td>[-18.13, -17.89]</td>
<td>-16.08 (7.97)</td>
<td>[-16.20, -15.96]</td>
</tr>
<tr>
<td>4</td>
<td>-17.84 (8.08)</td>
<td>[-17.96, -17.72]</td>
<td>-22.19 (7.38)</td>
<td>[-22.30, -22.08]</td>
<td>-15.78 (8.15)</td>
<td>[-15.90, -15.66]</td>
</tr>
<tr>
<td>8</td>
<td>-17.60 (7.41)</td>
<td>[-17.70, -17.49]</td>
<td>-18.00 (7.65)</td>
<td>[-18.11, -17.89]</td>
<td>-14.94 (7.92)</td>
<td>[-15.05, -14.82]</td>
</tr>
<tr>
<td>12</td>
<td>-3.61 (10.36)</td>
<td>[-3.46, -3.76]</td>
<td>-5.87 (10.99)</td>
<td>[-6.04, -5.72]</td>
<td>-10.13 (10.82)</td>
<td>[-10.29, -9.97]</td>
</tr>
<tr>
<td>13</td>
<td>-6.76 (23.21)</td>
<td>[-6.42, -7.10]</td>
<td>-7.43 (27.72)</td>
<td>[-7.84, -7.03]</td>
<td>-20.33 (45.85)</td>
<td>[-21.01, -19.66]</td>
</tr>
</tbody>
</table>
Table 5.3 Mean (M), standard deviation (SD), and 99.9% confidence interval (CI) of the propeller twist over 50000 frames for the 14 base pair sequences.
Central G·C base pairs are shaded in gray.

<table>
<thead>
<tr>
<th>Base Pair</th>
<th>CGAAAAAGCTTTTCG</th>
<th>CGAATTGCAATTTCG</th>
<th>CGATATGCATATCG</th>
<th>CGAATTGGAAATTTCG</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD)</td>
<td>99.9% CI</td>
<td>M (SD)</td>
<td>99.9% CI</td>
</tr>
<tr>
<td>1</td>
<td>-16.58 (39.12)</td>
<td>[-17.15, -16.00]</td>
<td>-34.27 (55.20)</td>
<td>[-35.08, -33.46]</td>
</tr>
<tr>
<td>8</td>
<td>-12.90 (8.03)</td>
<td>[-13.02, -12.79]</td>
<td>-4.50 (9.19)</td>
<td>[-4.63, -4.36]</td>
</tr>
<tr>
<td>9</td>
<td>-17.50 (7.88)</td>
<td>[-17.62, -17.39]</td>
<td>-12.61 (8.69)</td>
<td>[-12.73, -12.48]</td>
</tr>
<tr>
<td>14</td>
<td>-12.55 (34.08)</td>
<td>[-13.05, -12.05]</td>
<td>-7.93 (17.12)</td>
<td>[-8.18, -7.67]</td>
</tr>
</tbody>
</table>
structural flexibility as shown by the break at the G or GC core in the minor groove width comparisons. Alternatively, sequences with alternating A·T base pairs do not experience steric clash of the CH₃ groups and are therefore more flexible with lower propeller twist.

In addition to propeller twist of our sequences, other base step parameters were evaluated for influence on DNA structure. Comparing the single G-containing sequences, only nominal differences occur in all of the parameters except for roll (Figure 5.13). Roll increases after thymidine-purine steps in AATTGAATT and ATATGATAT. On the other hand, comparison of the two central GC sequences shows interesting results within helical parameters. Significant increases are seen at TpA steps for slide, rise and twist while other sequences are constant. With shift, both up and down changes are seen in ATATGCAAT (Figure 5.13) while no consistent patterns occur within the set of sequences. Increases also occur in slide at TpA steps for ATATGCAAT. Additionally, tilt and roll decrease as AATTGAATT transitions to AATTGCAATT at the CpA step. It is interesting to note how the addition of the second core G·C base pair causes ATATGCAAT to become an outlier compared to the other sequences. Specifically, the marked deviations that occur at the pyrimidine-purine steps exhibit the most dramatic changes in helical parameters. Propeller twists, rolls, and tilts likely compensate for each other in AAAAA·TTTT and AATT sites due to the bifurcated hydrogen bonding networks.

The observable changes in ATATGCAAT for every measurable parameter is likely from an inherent flexibility due to the alternating 5’ to 3’ purine-pyrimidine steps that can perturb canonical B-DNA conformations.[49] The ATATGCAAT sequence is the only consistently alternating purine-pyrimidine sequence within this series and is the sequence with the lowest binding affinity for DB2277. There is a high degree of dynamic helical bending in ATATGCAAT compared to AAAAA·TTTT. Early reports by Charney and co-workers demonstrated alternating
Figure 5.13 Average helical base pair step parameters for one and two G·C base pair sequences.

(A, G) Shift, (B, H) slide, and (C, I) rise (i.e. translational parameters) are reported in Å. (D, J) Tilt, (E, K) roll, (F, L) twist (i.e. rotational parameters) are reported in degrees. The terminal base pair caps were removed.
poly (dA·dT) sequences are nearly twice as flexible as “random” DNA.[50] Therefore, the apparent increased flexibility and dynamic bending of our alternating purine-pyrimidine sequences can explain the relatively poor binding of DB2277 with ATATGCATAT and ATATGATAT. These findings suggest that for ATATGCATAT, no single base pair parameter contributes substantially to minor groove width or depth. Instead, minor groove characteristics are a collective contribution of intra and inter-base pair parameters.

Simulations of the 1:1, monomeric complexes were next performed for DB2277 binding with single G·C sequences. Because the DNA sequences are asymmetric about the DB2277 binding site (i.e. 5’-AAAAGTTTT-3’ vs. 5’-AAAACTTTT-3’) and because of asymmetry in the small molecule, DB2277 was oriented in both the 5’ to 3’ and the 3’ to 5’ directions, totalling six simulations. For all orientations (six total), two-dimensional contour histograms of the simulated complexes are shown in Figure 5.14. To our surprise, the minor groove width distributions for the 1:1 complexes changed markedly and were nearly identical for all the simulated orientations. Upon binding DB2277, the preferred sequence, AAAAGTTTT, undergoes very small change in minor groove width. In both AATTGAATT and ATATGATAT simulations, the minor groove becomes constricted at the central G·C base pair, indicative of an induced fit recognition mechanism. This phenomenon is especially prevalent in ATATGATAT, yielding a $\Delta \text{width} \approx 3.5 \text{ Å}$. The comparison of effects for AAAAGTTTT and ATATGATAT is interesting since the intrinsic minor groove structure of AAAAGTTTT did not change much upon binding DB2277, in contrast to ATATGATAT. This phenomenon is worth noting since AAAAGTTTT showed the highest affinity for DB2277 while ATATGATAT had the lowest affinity within this series. Regardless of sequence, in the presence of DB2277, minor groove width conforms to the same pattern at its target binding site. Therefore, sequences that the start (free) and end (bound) most similarly have the
**Figure 5.14** 2D contour histograms of minor groove width for 1:1, DB2277-DNA complexes. (A) AAAAGTTTT, (B) AATTGAATT and (C) ATATGATAT. The color gradient indicates increasing probability (navy to red) of distance in Angstroms (in Å) for each base pair.

most favorable binding as a result of lower deformation energy of the DNA. This implies that the sequence with the highest binding affinity for our test compound already has a shape
complementary to the small molecule and further suggests that inherent microstructure of the DNA strongly influences binding affinity.

5.6 Conclusion

In this study, the molecular basis for sequence-specific binding by a synthetic minor groove binder is explained by inherent differences in the local DNA structure for an investigated set of sequences. This is the first reported use of competition mass spectrometry to identify unique DNA-ligand interactions that are explained by highly detailed, long time-scale molecular dynamics simulations. Our current understanding suggests that planar, synthetic small molecules, such as our test compound, bind best to sequences with a narrow and deep minor groove. Increased flexibility in specific sequences contributes to a wide and shallow groove that is unfavorable for strong 1:1 binding, while unexpected 2:1 binding of the compound for certain sequences further illustrates the sequence-dependent, microstructural variations within DNA. These findings emphasize the need for structural complementarity between the shape of a designed small molecule binder and the local structure of the DNA minor groove, and is therefore critical for understanding small molecule, sequence-specific recognition of DNA. Such site-specific recognition would prove useful for selectively targeting and modulating transcription factor activity and can become a powerful therapeutic tool for treating genetic-related diseases.

5.7 Materials and methods

5.7.1 Compound and dsDNA

Synthesis of compound DB2277 can be found in Chai et al.[33] Stock solution of 1.5 mM DB2277 was prepared by dissolving solid in doubly distilled water and storing at 4 °C until use. DNA sequences were bought from Integrated DNA Technologies (Coralville, IA, USA). Sequences were dissolved in the appropriate experimental buffer and spectroscopically determined
with extinction coefficients at 260 nm provided by IDT calculated using the nearest-neighbor method.[51] All buffers were filtered and degassed using a 0.22 µm hydrophilic cellulose filter (Millipore, Darmstadt, Germany).

5.7.2 Electrospray ionization mass spectrometry

Every sequence was designed to have a unique molecular weight so that all DNAs could be combined into a single sample vial for mass spectrometry experiments. Distinguishability was possible by substitution of guanine for inosine (/ideoxyI/) which removes NH₂ from the N3 position of G, addition of a phosphate group (p) at the 3’ terminus, or modification of the hairpin loop with pyrimidine bases (i.e. thymidine and/or cytidine). Stock solutions of 1 mM DNA were prepared by dissolving DNA in 150 mM NH₄OAc (pH 6.8) and dialyzed using a 1000 Da MWCO (Spectrum Labs, Rancho Dominguez, CA, USA) in 150 mM NH₄OAc buffer with 3x buffer exchange. Reference DNA: 5’-CGCGCGCGCTTTTGCGCGCGCG-3’; DNA 1: 5’-GCGAAAATTTTTCGCGAAAAATTTTTCG-3’; DNA 2: 5’-CGAAAAAGTTTTTCGCGAAAAAATTTTTTCG-3’; DNA 3: 5’-CGAAAAAGCTTTTTCGCGAAAAAGCTTTTTCG-3’; DNA 4: 5’-CGATATGCATATCGCCCGATATGCATATCG-3’; DNA 5: 5’-CGAATTTGAATTTCGTCCCCGGAATTTCCATTCG-3’; DNA 6: 5’-CGAATTGCAATTCGTCTCCCGGAATTGCAATTCG-3’; DNA 7: 5’-CGAATTGCAATTCGTCTCCCGGAATTGCAATTCG-3’; DNA 8: 5’-CGAAAGCAATTCGTTTTTCGAATTGCATTTCCGpG-3’; DNA 9: 5’-CGAAAACCTTTTGCGCCCCCGAAAAGTTTTTGCpC-3’.

Samples were prepared by combining 10 µM of each DNA sequence into a single vial. A G·C base pair sequence (i.e. no AT base pairs present) of equimolar concentration was used as a
reference sequence for which complex peak intensities were compared. In the same vial, DB2277 was added so that the total concentration of compound was equivalent to the desired molar concentration ratio of the total DNA concentration. For example, a [3 to 1] concentration ratio of DB2277 to DNA = [DNA 1] 10 µM + [DNA 2] 10 µM + … + [DNA 9] 10 µM = 90 µM total concentration DNA in sample (9 x 10 µM = 90 µM) and, therefore, 270 µM of DB2277 in the sample gives [270 to 90] or [3 to 1]. An additional 5% MeOH (v/v) was added to the sample prior to injection to help facilitate gas-phase transition and give a total sample volume of 75 µL.

Samples for ESI-MS were using a Waters Micromass ESI-Q-ToF spectrometer (Waters Corporation, Milford, MA, USA) and analyzed with MassLynx 4.1 software. Samples were scanned from \( m/z \) 500-3,000 in negative ion mode at a flow rate of 5 µL∙min\(^{-1}\) with the final two minutes of the chromatogram (≈ 200 scans) averaged. Spectral peaks were deconvoluted. Capillary 2000 V, sample cone 20 V, extraction cone 1.8 V, source 70 °C, desolvation 100 °C, cone gas flow 30 L∙hr\(^{-1}\), desolvation gas flow 450 L∙hr\(^{-1}\), ion energy 2.5 V, collision energy 2 V, RF1 lens -125 V, RF2 lens -103 V, and acceleration lens 88 V. A comparison of deconvoluted peak intensities is shown in Figure 5.2. Some variations exist in peak intensity for free DNA since the ionization efficiencies and response sensitivities are different among the sequences and so samples of free DNA are analyzed as a reference. Upon complex formation, however, response sensitivities even out, as indicated by the excellent agreement in stoichiometry and relative binding affinities for ESI-MS with other biophysical and bioanalytical methods.

5.7.3 Molecular dynamics simulations

The sequence dependent major and minor groove widths and depths were analyzed using molecular dynamics trajectories.[52-55] Seven double-stranded sequences used in the computational simulations mirrored the hairpin sequences used in the competition mass
spectrometry studies: 5’-CGAAAGTTTTTCG-3’, 5’-CGAATTGAATTCG-3’, 5’-CGATATGATATCG-3’, 5’-CGAAAAGCTTTTCG-3’, 5’-CGAATTGCAATTCG-3’, 5’-CGATATGCATATCG-3’, including a comparative sequence with two GG bases in the core as 5’-CGAATTGGAATTCG-3’. Nucleic Acid Builder (NAB) was used to construct canonical B-form dsDNA models for the above sequences.[56] Optimization and electrostatic potential calculations for DB2277 were performed using the DFT/B3LYP [57-60] method with the 6-31+G* basis set [61-63] in Gaussian09 (Gaussian, Inc., 2009, Wallingford, CT). Atom charges were calculated using the Merz-Singh-Kollman (MK) Scheme.[64, 65] Force field parameter and AMBER preparation files were generated using ANTECHAMBER [66] with the GAFF force field.[67] The angle and dihedral parameters of DB2277 that were not defined by the GAFF force field were modified in the force field modification (frcmod) file to fit ab initio calculations from potential energy scans performed in Guassian09. Three of the seven sequences, 5’-CGAAAGTTTTTCG-3’, 5’-CGAATTGAATTCG-3’, 5’-CGATATGATATCG-3’, were simulated with DB2277 bound in the minor groove. Because these sequences are non-palindromic about the DB2277 binding site (i.e. 5’-AAAAGTTTT-3’ vs. 5’-AAAACTTTT-3’) and because of the asymmetry in DB2277, it was oriented in both 5’ to 3’ and 3’ to 5’ directions. The small molecule was manually docked at the central G of the Watson strand of the canonical B-form dsDNA models produced by NAB using the visualization software, VMD.[68] All systems were solvated with TIP3P water [69] in a truncated octahedral water box extending 10.0 Å from the dsDNA in each dimension and neutralized with Na+ ions in TLeap (AMBER 2015, University of California, San Francisco). Additional Na+ and Cl- ions were added to reach a salt concentration of 150 mM for each dsDNA system.
Figure 5.15 Cartoon representations of groove widths.
Cartoon representation of minor groove width (cyan) and depth (dashed line) and major groove width (green). Groove depths are defined as the distance between the center of the width vector (cyan or green) and the mid-point of the vector constructed using the C6 of pyrimidines and the C8 of purines (orange), which defines the corresponding base pair.

All systems were relaxed over 5,000 steps of steepest-descent minimization with positional restraints imposed on the nucleic acid residues. Heating 0 K to 310 K was carried out over 10 ps in the canonical ensemble with 5 kcalÅ²mol⁻¹ harmonic restraints enforced on the heavy atoms of the nucleic acid residues. For the DB2277-DNA complex systems, 5 kcalÅ²mol⁻¹ distance restraints were applied between hydrogen bond donor and acceptor pairs in the capping base pairs, and between the nitrogen of the DB2277 aza-benzimidazole ring and the guanine NH₂ group. The harmonic restraints were released over eight stages of equilibration in the isothermal-isobaric ensemble (T = 310 K, P = 1 atm) with a 2 fs time step for a total of 500 ps using the Berendsen algorithm [70] to control temperature and Monte Carlo barostat as implemented in AMBER14. SHAKE was imposed on bonds involving hydrogen atoms [71] and electrostatic interactions were treated using the smooth Particle Mesh Ewald method [72] with a 10.0 Å cutoff. Distance restraints
were monotonically released to 1 kcal·Å²·mol⁻¹ for the capping base pairs and 0 kcal·Å²·mol⁻¹ for the aza-N·G-NH₂ hydrogen bond over 50 ns after positional restraints were fully released. The 1 kcal·Å²·mol⁻¹ distance restraints on the capping base pairs were maintained during the production simulations to prevent fraying which could skew the dynamics of the DB2277-DNA complex. Production level simulations were extended to 200 ns and trajectory snapshots were saved every 1 ps. All simulations were performed using the PMEMD CUDA [73] module of AMBER14 with the parm99 force field and the parmbsc0 + ε/ζOL1+χOL4 force field modifications for DNA [74-76].

Trajectories were preprocessed using CPPTRAJ module of AMBERTOOLS 15 to produce 50,000 snapshots for analysis and visualization in VMD. Major and minor groove width and depth, and propeller twist for each base pair were calculated as well as base pair step translational and rotational helical parameters using Curves+ and Canal programs for the unbound complexes.[77] Minor groove width of the bound complexes was also calculated using the Curves+ and Canal programs. Contoured histograms were produced by increasing the number of points per base pair that define the backbone spline from 10 to 50 and collecting the resultant minor groove widths from the two-dimensional distance matrix produced by Curves+ (Figure S2). A smooth line was interpolated along the entire groove for each frame analyzed. Width and depth data generated from Curves+ was then binned into 100 evenly spaced bins and represented as histograms using MATLAB (The MathWorks, 2016, Natick, MA, USA). To establish the convergence of our simulations, repeats of the 200 ns simulations for two sequences with randomized initial velocities. For both Simulation 1 and Simulation 2, the minor groove width was analyzed in 100 ns sections. It can be seen clearly that for both sequences, the minor groove width is nearly identical in all
cases (Figures 5.8, 5.9). Additionally, PDB coordinates of DB2277 bound to AAAAGTTTTT can be found as Supplementary Data.

5.7.4 Biosensor surface plasmon resonance

SPR experiments were performed using a four-channel Biacore T200 (GE Healthcare, Inc., Piscataway, NJ, USA) and the Biacore T200 Evaluation Software. A streptavidin-coated sensor chip was prepared by series injections of 1 M NaCl in 50 mM NaOH at 60 s intervals. Washing of the chip surface followed using HBS buffer (10 mM HEPES, 150 mM NaCl, 3 mM EDTA, 0.05% P20, pH 7.4). Following chip washing, biotin-5’-end labelled hairpin DNA sequences (5’-/5Biosg/CTGCTCTCAAAAGCTTTTGG-3’, 5’-/5Biosg/GGCAATTGCAATTCC-3’, 5’-/5Biosg/GGCAATTGCAATTCC-3’) were dissolved to 25 nM concentrations in HBS buffer and immobilized to the flow cell surface via non-covalent capture using previously described methods.[78, 79] The first flow cell (FC1) was intentionally left blank as a reference cell while FC2, FC3 and FC4 were immobilized with 50 nM stock solutions of AAAAGCTTTT, AATTGCAATT and ATATGCATAT sequences, respectively. Immobilization was achieved by manual injection of the DNAs at a flow rate of 1 µL·min⁻¹ until response units (RU) of 310-330 were reached. Compound solution concentrations of DB2277 were prepared in 50 mM Tris-HCl buffer (pH 7.4) ranging 10 nM – 1 µM and were injected over the sensor chip at a flow rate of 100 µL·min⁻¹ until steady-state responses were reached. Buffer was then flowed over the chip surface to dissociate bound DB2277 from DNA. Following each cycle, the surface of the sensor chip was regenerated with 10 mM glycine (pH 2.5) for 30 s and rinsed with three injections of experimental buffer to produce a stable baseline for the following cycles. The observed response
(RU_{obs}) was plotted as a function of free compound concentration (C_f) and dissociation binding constants (K_D) determined using the following equations:

\begin{align*}
(5.1a) \quad RU_{obs} &= \frac{\text{MW of ligand}}{\text{MW of DNA}} \times \frac{\text{RU of ligand}}{\text{RU of DNA}} \\
(5.1b) \quad RU_{obs} &= \frac{[(K_1 \cdot C_f) + (2K_1 \cdot K_2 \cdot C_f^2)] \times RU_{max}}{1 + (K_1 \cdot C_f) + (K_1 \cdot K_2 \cdot C_f^2)} \\
(5.1c) \quad RU_{obs} &= \frac{[(K_1 \cdot K_2 \cdot C_f^3) + (3K_1 \cdot K_2 \cdot C_f^3) + (3K_1 \cdot K_2 \cdot K_3 \cdot C_f^3)] \times RU_{max}}{1 + (K_1 \cdot C_f)^2 + (K_1 \cdot K_2 \cdot C_f^2) + (3K_1 \cdot K_2 \cdot K_3 \cdot C_f^3)}
\end{align*}

Binding plots were fit using a binding model of “n+1”, where n equals the number of binding sites determined for a sequence from ESI-MS. An “n+1” model was used to confirm the binding constants were not averaged (i.e. K_{12}) thus providing an additional, low affinity binding constant well above the measured concentration range (i.e. K_D > 10^{-6} \text{ M}). Data were fit using Equations 5.1b and 5.1c in KalediaGraph (Synergy Software, Reading, PA, USA). RU_{max} represents the maximum measured response for bound DB2277 to DNA from the equation and K_1, K_2 and K_3 are macroscopic association constants in units M^{-1}. Equilibrium binding constants were converted to dissociation constants (K_D) and reported in units of concentration, M. The microscopic association and dissociation rates, k_a and k_d, (i.e. on and off-rates) were calculated at low concentrations of DB2277 using Biacore T200 Evaluation Software and kinetics fit binding constants were compared to values obtained by steady-state models.

5.7.5 Competition biosensor-surface plasmon resonance

The same streptavidin-coated sensor chip previously described was also used for competition SPR analyses. Samples containing unlabeled hairpin DNA sequences (5’-CGAAAAGTTTTTCGGCTCTCCGAAAAACTTTTCG-3’, 5’-CGAAAAAGTTTTTCGGCTCTCCGAAAAACTTTTCG-3’)
CGAATTGAATTCCGTCTCCGAATTCAATTCG-3’, and 5’-
CGATATGATATCGGCTCTCCGATATCATATCG-3’) were added to a constant concentration of DB2277 and flowed over the chip surface. The added DNA sequences in solution compete for binding to DB2277, which results in the decrease of RU

Observe. Solutions were prepared in 50 mM Tris-HCl buffer with constant 100 nM DB2277 compound. Competing DNA concentrations ranging 0 nM – 2.5 µM were injected over the sensor chip at a flow rate of 100 µL·min⁻¹ until steady-state responses were reached. Experimental buffer (i.e. no DB2277) was flowed over the chip surface to dissociate bound DB2277 from DNA and competing DNA. The sensor chip surface was regenerated with 0.5 M NaCl for 30 s and rinsed with three injections of experimental buffer to produce a stable baseline for the next cycle. For chip regeneration, no detectable differences were observed in the baseline stabilization among 10 mM glycine, 1 M NaCl, and 0.5 M NaCl solutions. A lower salt concentration of 0.5 M NaCl was used for competition SPR analyses to ensure a constant RU of immobilized 5’-end labeled biotinylated DNA was maintained, meaning none of the immobilized DNA dissociated during regeneration.

To determine the solution dissociation constant (K_s) of the competing DNA, a one-site binding model was used to determine K_s of the competing DNA with DB2277 in solution by the following equation:

\[
(5.2) \text{RU}_{obs} = \frac{\left( [L]_T - 0.5(K_s + [L]_T + [D]_T) - \sqrt{[K_s + [L]_T + [D]_T]^2 - 4([L]_T[D]_T)} \right) \times RU_{max}}{\left( [L]_T - 0.5(K_s + [L]_T + [D]_T) - \sqrt{[K_s + [L]_T + [D]_T]^2 - 4([L]_T[D]_T)} \right) + K_D}
\]

The RU

Observe was plotted against total competing DNA concentration to determine dissociation constants of the competitor DNA (K_s) in solution. Equation 5.2 was derived specifically for competition SPR analyses since the concentration of DNA-ligand complex in solution cannot be...
determined. It substitutes a conventional one-site binding model with a quadratic formula [40] which includes total concentrations of DB2277 ligand and competing DNA in solution as $[L]_T$ and $[D]_T$, respectively. The $K_D$ in Equation 5.2 is an averaged equilibrium dissociation constant of $K_1$ and $K_2$ values (i.e. $K_{12}$) for DB2277 binding with AAAAGCTTTT, AATTGCAATT or ATATGCATAT previously determined by Equations 5.1b or 5.1c. Here, $[L]_T$, $[D]_T$, $RU_{max}$ and $K_D$ are used as fitting parameters to determine $K_S$ using KalediaGraph. Dissociation constants for AATTGAATT using competition SPR are compared to values obtained direct-binding SPR and are reported in Table 5.1.

5.8 References


6 CONCLUSIONS

In this dissertation, recognition of the DNA minor groove by small molecules was investigated. The use of electrospray ionization mass spectrometry was the primary tool used to identify ligand-DNA interactions. In addition to the findings by electrospray ionization mass spectrometry (ESI-MS), other biophysical methods were used to validate our findings.

We first introduced the development of a novel technique using ESI-MS to observe competitive binding of well-characterized compounds with various DNA sequences. We highlighted the many advantages competition ESI-MS has to offer such as the rapid and convenient sample analysis, as well as the small amounts of sample required. More importantly, competition ESI-MS allowed easy identification of stoichiometry, cooperativity, and a direct comparison of relative binding affinities. In theory, any number of sequences and ligands can be simultaneously analyzed as long as the molecular weights of each species and their potential complexes are discernable. With one sample, lots of information can be gathered all the while reducing reagents used and time spent cleaning between sample runs. More importantly, it is not limited to DNA and can be applied to other biomacromolecular interactions.

Our competition ESI-MS method was next applied to investigations of mixed DNA sequences with small molecules. Many important features were discovered using systematic variations of a test compound and its target binding site. Specificity and cooperativity of the test compound against several analogues determined the parent compound was optimum for binding with the target sequence. One such analogues showed unexpected binding with both target and mixed-site reference sequences. Several mutant sequences displayed unusual binding patterns strictly based on sequence such that a simple reversal of two bases would result in complete loss
of binding of the parent compound. Overall, the consensus binding site remained the preferred sequence and further illustrated the importance of base pair sequence in minor groove recognition.

In the next project, several diamidines were tested for selectivity among a set of mixed-site sequences. Recognition groups within the compounds, such as hydrogen bond acceptors and donors, and flexibility of the compound were strategically modified to determine the effects on minor groove binding. Competition ESI-MS, in combination with other biophysical methods, confirmed the strong and selective recognition of a diamidine for a single G-containing sequence. The information gathered provided additional information for the rational design of more specific compounds to target longer mixed-site sequences.

Finally, ESI-MS was used to identify additional binding interactions of a well-characterized compound with new mixed-site sequences. Competition ESI-MS initially provided information such as stoichiometry and cooperativity. More detailed studies using SPR showed direct correlation for binding with results obtained by ESI-MS. MD simulations unveiled a distinct pattern in the DNA microstructure for several sequences which later explained the intrinsic binding behavior of our test compound. Complementary evidence from both experimental and computational methods provided a rationale for the sequence-dependent behavior of the compound binding in the minor groove.

The common denominator in each of these projects is the application of competition ESI-MS to identify interactions of minor groove binding compounds with DNA. This method is versatile in its ability to provide quick and accurate information regarding stoichiometry, cooperativity, and relative binding affinity. ESI-MS, in conjunction with other analytical and biophysical techniques, is capable of identifying patterns in minor groove recognition which are based on structural features found (1) in the compound and (2) within the minor groove itself. This
wealth of information is useful in developing a new paradigm for sequence specific recognition of the DNA minor groove by synthetic small molecules which will ultimately lead to desired regulation of genetic expression.
APPENDIX A. MODULATION OF TOPOISOMERASE IIα EXPRESSION AND CHEMOSENSITIVITY THROUGH TARGETED INHIBITION OF NF-γ: DNA BY A DIAMINO p-ANISYL-BENZIMIDAZOLE (Hx) POLYAMIDE


Copyright © 2016 Biochimica et Biophysica Acta – Gene Regulatory Mechanisms

My contribution to this project was the SPR sample preparation, performance, and analysis of results.
A.1 Abstract

A.1.1 Background

Sequence specific polyamide HxIP 1, targeted to the inverted CCAAT Box 2 (ICB2) on the topoisomerase IIα (topo IIα) promoter can inhibit NF-Y binding, re-induce gene expression and increase sensitivity to etoposide. To enhance biological activity, diamino-containing derivatives (HxI*P 2 and HxIP* 3) were synthesised incorporating an alkyl amino group at the N1-heterocyclic position of the imidazole/pyrrole.

A.1.2 Methods

DNase I footprinting was used to evaluate DNA binding of the diamino Hx-polyamides, and their ability to disrupt the NF-Y:ICB2 interaction assessed using EMSAs. Topo IIα mRNA (RT-PCR) and protein (Immunoblotting) levels were measured following 18 h polyamide treatment of confluent A549 cells. γH2AX was used as a marker for etoposide-induced DNA damage after pre-treatment with HxIP* 3 and cell viability was measured using Cell-Titer Glo®.

A.1.3 Results

Introduction of the N1-alkyl amino group reduced selectivity for the target sequence 5′-TACGAT-3′ on the topo IIα promoter, but increased DNA binding affinity. Confocal microscopy revealed both fluorescent diamino polyamides localised in the nucleus, yet HxI*P 2 was unable to disrupt the NF-Y:ICB2 interaction and showed no effect against the downregulation of topo IIα. In contrast, inhibition of NF-Y binding by HxIP* 3 stimulated dose-dependent (0.1–2 μM) re-induction of topo IIα and potentiated cytotoxicity of topo II poisons by enhancing DNA damage.
A.1.4 Conclusions

Polyamide functionalisation at the N1-position offers a design strategy to improve drug-like properties. Dicationic HxIP* 3 increased topo IIα expression and chemosensitivity to topo II-targeting agents.

A.1.5 General significance

Pharmacological modulation of topo IIα expression has the potential to enhance cellular sensitivity to clinically-used anticancer therapeutics. This article is part of a Special Issue entitled: Nuclear Factor Y in Development and Disease, edited by Prof. Roberto Mantovani.

A.2 Key words

DNA-binding polyamides; Sequence selectivity; Transcription factor-DNA interactions; Gene modulation; NF-Y; Topoisomerase IIα (Topo IIα); Chemosensitisation

A.3 Introduction

Gene expression is precisely regulated by the binding of the transcription machinery to specific DNA sequences. Dysregulation of transcriptional activity leading to aberrant gene expression is a fundamental driver of a diverse array of human diseases. Sequence selective (P) pyrrole-(I) imidazole polyamides are able to modulate gene expression through binding non-covalently to specific DNA sequences and disrupting the DNA interactions of transcription factors. These reversible DNA minor groove binders can arrange in a stacked, antiparallel 2:1 (ligand:DNA) orientation and afford programmable sequence recognition, governed by the side-by-side heterocyclic ring pairing rules.[1-4] A/P pairing degenerately targets A·T or T·A, whereas P/I recognises C·G and I/P preferentially binds to G·C. The reported anti-cancer biological activity of these cell permeable small molecules in both cellular and in vivo studies has
highlighted the potential of therapeutic strategies that directly target transcription factor-DNA interfaces known to be implicated in certain malignant phenotypes.[5-8]

Nuclear Factor Y (NF-Y) is a heterotrimeric CCAAT-binding transcription factor involved in cell differentiation, proliferation and implicated in cancer progression.[9-13] NF-Y has been shown to bind to the promoter of the essential DNA processing enzyme, topoisomerase IIα (topo IIα) and regulates its transcription through interactions with the inverted CCAAT box (ICB) sequences located within the promoter.[14-16] Topo IIα plays a critical role in DNA metabolism, maintaining genomic stability [17], and is the target of clinically-used chemotherapeutic agents etoposide and doxorubicin [18], with low levels of topo IIα conferring cellular resistance to these anticancer agents.[19-22] NF-Y acts as both an activator and repressor of topo IIα transcription with increased association of NF-Y to the promoter exerting a negative effect at confluence.[15, 16] The ICB2 has been identified as the crucial DNA regulatory element and its interaction with

**Figure A.1. Schematic representation of topoIIα.**
Schematic representation of the confluence-induced downregulation of topo IIα, mediated by the repressive binding of NF-Y to the ICB2. The ICB sequence ATTTGG is highlighted in blue and the 5'-flanking sequence of the ICB2 is outlined with a dashed box.
NF-Y (Figure A.1) mediates the confluence-induced downregulation of topo IIα and reduced chemosensitivity to topo II targeting therapeutics.

Chemical approaches that re-induce topo IIα expression have the potential to increase cellular sensitivity to topo II poisons and to this end, our group has used various DNA binding small molecules and polyamides to inhibit the repressive activity of NF-Y on the topo IIα promoter.[16, 23-27] Most recently, we reported the synthesis and biological activity of a novel polyamide incorporating the p-anisylbenzimidazole (Hx) DNA recognition element.[28, 29] Designed to enhance polyamide-DNA binding and cellular uptake, the Hx moiety exhibits intrinsic fluorescence upon binding DNA enabling the direct visualisation of polyamide nuclear localisation. Hx-polyamide HxIP (1) (Figure A.2A) designed to target the 5′-flanking sequence 5′-TACGAT-3′ of the ICB2 (Figure A.1), binds to DNA with high affinity and sequence selectivity, and disrupts the NF-Y:ICB2 interaction resulting in the upregulation of topo IIα expression at confluence. HxIP pre-treatment enhanced etoposide-induced DNA damage, providing further evidence that sequence specific polyamides can re-sensitise confluence-arrested cancer cells to topo IIα poisons.[28]

The development of HxIP provides a new framework for the design of fluorescent sequence selective DNA binding molecules, distinct in configuration from the prototypical hairpin polyamide, yet capable of efficient nuclear localisation and in vitro gene regulation. In parallel, we have continued to explore an alternative strategy for further optimisation of polyamide physicochemical and DNA binding properties, through the introduction of an additional alkyl amino group at the N1 position of the heterocyclic rings.[30-33] DNA binding studies by our groups revealed the small diamino polyamide containing an orthogonal positioned propyl amino group, f-IP*I (*denotes modified heterocycle) to have greater binding affinity than its monoamino
Figure A.2. Design of the diamino Hx-polyamides 2 and 3.
(A) Structures of the monoamino Hx-polyamide, HxIP 1 and the orthogonally positioned diamino-containing derivatives, HxI*P 2 and HxIP* 3. The asterisk (*) denotes position of the N-alkyl amino group. (B) Schematic models of polyamides 2 and 3 binding to the 5'-TACGAT-3' sequence on the 5'-flank of the ICB2 in an antiparallel 2:1 fashion as an overlapped stacked dimer.

counterpart and analogous sequence selectivity. Importantly, the inclusion of an extra amino group, which is cationic at physiological pH, also increases water solubility and may afford greater polyamide nuclear uptake. We aim to exploit this potential combination of improved DNA binding and solubility properties to engineer a potentially more potent generation of dicationic polyamides.

This study presents the DNA binding and biological activities of the diamino Hx-polyamides HxI*P 2 and HxIP* 3, which incorporate the N1-alkyl amino group modification
(Figure A.2A). Polyamides 2 and 3 are functionalised derivatives of HxIP 1 and like their monoamino predecessor are designed to target the ICB2 5’-flanking sequence 5’-TACGAT-3’ on the topo IIα promoter (Figure A.2B) and disrupt NF-Y binding, inducing topo IIα expression at confluence. Comparison with the monoamino HxIP 1 will reveal the effect the inclusion of an N1-alkyl amino group in the Hx-framework has on DNA binding and the feasibility of using functionalisation of the N1 position to enhance polyamide cellular uptake and biological activity. Additionally, we shall assess the potential chemosensitising effects of the diamino generation of polyamide inhibitors of NF-Y:DNA binding and the effectiveness of pharmacological modulation of topo IIα expression as a strategy for overcoming the drug resistance exhibited by confluence-arrested cells.

A.4 Methods and materials

A.4.1 Synthesis of polyamides

Details of the synthesis and characterisation of diamino Hx-polyamides 2 and 3 are provided in the Supplemental Materials and Methods (Scheme S1).

A.4.2 Thermal denaturation studies

Thermal denaturation (ΔT_M) studies were performed using a Cary Bio 100 spectrophotometer UV–Vis instrument (Palo Alto, CA) as previously described by Chavda et al.[29] DNA oligomers were purchased from Operon and the sequences are provided in the Supplemental materials. Experiments for diamino Hx-polyamides 2 and 3 were performed at a concentration of 3 μM ligand and 1 μM DNA. All melts were performed in 10-mm path length quartz cells. T_M values were determined as the maximum of the first derivative, and ΔT_M values are the difference between the melting temperatures of ligand bound DNA and free DNA.
A.4.3 Circular dichroism

Circular dichroism (CD) studies were carried out as previously reported,[30] using an OLIS (Bogart, GA, USA) DSM20 spectropolarimeter. Experiments were conducted at ambient temperature in a 1-mm path length quartz cell using phosphate A PO₄ buffer (10 mM sodium phosphate, 1 mM EDTA, pH 6.2). Buffer and stock DNA were added to the cuvette to give a final DNA concentration of 9 μM. Each diamino Hx-polyamide (in 500 μL in ddH₂O) was titrated in increments of 1 M equivalent into the relevant DNA (160 μL of 9 μM DNA). Each run was performed over 400–220 nm. The CD response at the $\lambda_{\text{max}}$ of the induced peak was plotted against the mole ratio of ligand to DNA.

A.4.4 Surface plasmon resonance

Surface plasmon resonance (SPR) measurements were performed with a four-channel BIAcore T100 optical biosensor system (Biacore, GE Healthcare Inc.). A streptavidin-coated sensor chip was prepared by a series injections of 1 M NaCl in 50 mM NaOH for periods of 60 s followed by washing of the chip surface with HBS buffer (10 mM HEPES, 150 mM NaCl, 3 mM EDTA, 0.05% P20, pH 7.4). Biotin-5′-end labelled DNA hairpin oligomers dissolved to 25 nM concentrations in HBS buffer were immobilized to the flow cell surface via non-covalent capture using previously described methods.[34, 35] The sequences of the three hairpin oligomers are detailed in the Supplemental Materials. Three flow cells were used to immobilise the DNA, while a fourth was left blank as a control. Immobilization was achieved by manual injection of the DNAs at a flow rate of 1 μL/min until response units (RU) of 350–400 were reached. Diamino polyamide 3 was dissolved in 500 μL MeOH and 0.9 mol eq. of HCl was added to form a salt. Methanol was removed using N₂ at 25 °C leaving a yellow salt. The compound was then re-dissolved in ddH₂O to a concentration near 1 mM and spectroscopically determined by UV–Vis
using the extinction coefficient ($\epsilon_{322\ nm} = 29,129\ \text{L/(mol\cdot cm)}$) and stored at 4°C until use (within 2 weeks). Compound solution concentrations of 3 were prepared in 10 mM cacodylic acid (CCL), 1 mM EDTA, and 200 mM NaCl buffer (pH 6.5) ranging 2 nM to 1 μM and were injected over the sensor chip at a flow rate of 100 μL/min for 180 s. Buffer was then flowed over the chip surface for 600 s to dissociate bound HxIP* 3 from DNA. Following each cycle, the surface of the sensor chip was regenerated with 1 M NaCl for 30 s and rinsed with three injections of experimental buffer to produce a stable baseline for the following cycles. The relative response units (RU) were plotted as a function of free compound concentration ($C_f$) and equilibrium binding constants ($K_{eq}$) were determined using a two-site cooperative binding model with Equation 1.

\[
\begin{align*}
  r &= \frac{RU}{RU_{\text{max}}} = \frac{[(K_1 \cdot C_f) + (2K_1 \cdot K_2 \cdot C_f^2)] \times RU_{\text{max}}}{1 + (K_1 \cdot C_f) + (K_1 \cdot K_2 \cdot C_f^2)} \\
  \text{Equation 1}
\end{align*}
\]

Here, $K_1$ and $K_2$ are macroscopic binding constants in units per concentration (M$^{-1}$). The maximum obtainable response (RU$_{\text{max}}$) in Equation 1 was calculated from the product of immobilized DNA response units, HxIP* molecular weight, refractive index of HxIP*, and the inverse DNA molecular weight. The obtained maximum value was compared to predict RU$_{\text{max}}$ of bound HxIP* to determine the stoichiometry. Equilibrium binding constants ($K_{eq}$) were determined using both global kinetics fits and/or steady-state binding models. Two equilibrium binding constants, $K_1$ and $K_2$, were determined, multiplied by each other, and the square root of the resulting calculated value gave an averaged equilibrium binding constant ($K_{eq}$) to compare with previous results.[29]

\[
K_{eq} = \sqrt{K_{D1} \cdot K_{D2}}
\]

(2)
A.4.5  **DNase I footprinting**

The 5′ radiolabelled DNA substrate corresponding to the *topo IIα* promoter was prepared by PCR amplification and then isolated and purified as previously reported.[16] The DNase I footprinting reactions were performed as described by Kiakos et al. [28] and resolved on a 10% denaturing polyacrylamide-urea gel (National Diagnostics) by electrophoresis for 3 h at 1650 V (55 °C) in 1 × TBE buffer. The gel was then transferred onto Whatman 3MM paper, dried and exposed overnight to Fuji medical X-ray film to visualise the radioactive signal.

A.4.6  **Cell lines and culture conditions**

NIH3T3 mouse fibroblast cells were obtained from CR-UK London Research Institute, and the A549 cell line was purchased from the European Collection of Cell Cultures. All cell lines were maintained in DMEM (Sigma-Aldrich) supplemented with 2 mM l-glutamine (Sigma-Aldrich) and 10% foetal bovine serum (Gibco, Life Technologies).

A.4.7  **Electrophoretic mobility shift assays**

Electrophoretic mobility shift assay (EMSA) experiments were conducted as outlined by Kiakos et al. [28] using nuclear protein extracts from NIH3T3 fibroblast cells and the ICB2-containing oligonucleotide; 5′-GGCAAGCTACGATTGGTTCTTCTGACG-3′ (sense); 5′-CGTCCAGAAGAACCAATCGTAGCTTGCC-3′ (antisense). Oligonucleotides containing a mutated ICB2 were used as specific competitors, with the wild-type ICB motif replaced by AAACC and GGTTT in sense and antisense oligonucleotides, respectively. For supershift experiments the nuclear extract and radiolabelled oligonucleotide were incubated with anti-NF-YA antibody (Abcam) for 1 h.
**A.4.8 Immunofluorescence staining and confocal microscopy**

To assess polyamide nuclear uptake, A549 and NIH3T3 cells grown on 13-mm glass cover slips were treated with different concentrations of polyamides 2 and 3 for various incubation times. Cells were fixed (2% PFA), washed with PBS and then permeabilised to allow nuclear DNA staining with PI (2 μg/mL). Alternatively, MitoTracker® Red was used to stain the mitochondria, without fixation or permabilisation. The blue fluorescence of the Hx-polyamides was excited with a UV-laser (364 nm) and detected using the DAPI filter. PI/MitoTracker® Red was excited with an argon-ion laser (488 nm) and detected at 642 nm. The fluorescence of the polyamide was detected and imaged by z-stack acquisition using the Zeiss 510 UV-VIS microscope and the LSM510 software. Immunofluorescence was used to assay for H2AX foci induction after polyamide-etoposide combination treatment. A549 cells were washed with PBS, fixed with 2% PFA and permeabilised with PBS containing 0.5% Triton X-100. Cells were subsequently blocked in PBS 5% BSA for 1 h and incubated with anti-phospho-histone H2A. X mouse monoclonal antibody (1:100; Millipore) diluted in PBS 1% BSA for 1 h. After three washes, cells were incubated with the goat anti-mouse secondary antibody Alexa Fluor 488 (1:100; Life Technologies) for another hour. Following further three washes in PBS 0.1% Triton X-100, nuclei were stained using a PI solution (2 μg/mL; Sigma-Aldrich). γH2AX levels were then visualised using confocal microscopy and the number of foci per nuclei was quantified using the CellProfiler software.[36, 37]

**A.4.9 Quantitative real time PCR**

RNA was isolated using the RNeasy Plus Mini Kit (Qiagen) and cDNA was generated with the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems). Real time PCR was carried out using the ABI PRISM 7000 Sequence Detection System (Applied Biosystems) and
Taqman assay probes (Life Technologies) for the target gene, *topo IIα* (Hs01032137_m1) and the internal control, *GAPDH* (Hs03929097_g1). The relative quantification of topo IIα mRNA levels from untreated and polyamide treated confluent A549 cells was determined using the $2^{\Delta\Delta Ct}$ method and normalised to the internal control.

**A.4.10 Immunoblotting**

Nuclear extracts from A549 cell lines were prepared using a Nuclear Extraction Kit (Active Motif) following the manufacturer's protocol. Protein concentration was quantified using the Bio-Rad DC protein assay. Nuclear proteins were separated on a NuPAGE 7% Tris-acetate gel (Life Technologies), transferred to PVDF membranes and blocked (5% w/v non-fat dry milk in 1 × TBS, 0.1% Tween-20) for 1 h at ambient temperature. Topo IIα was identified by overnight incubation with topo IIα rabbit polyclonal antibody (1:5000, kindly provided by Dr. I.D. Hickson, Weatherall Institute of Molecular Medicine, Oxford, UK) and subsequent incubation for 1 h with anti-rabbit secondary antibody. Chemiluminescent visualisation was performed using ECL reagent (Amersham) and exposure onto autoradiography film (Kodak-X-Omat). Protein levels of γH2AX were also assessed after polyamide-etoposide combination treatment. For the detection of phosphorylated histone H2AX (Ser139), proteins were separated using a NuPAGE 4–12% Bis-tris gel and probed using the anti-phospho-histone H2AX monoclonal antibody (1:500, Millipore). Lamin (Cell Signalling) was used as a loading control for the nuclear protein extracts.

**A.4.11 Cell viability**

The CellTiter-Glo luminescent assay (Promega) was used as per manufacturer's instructions to assess viability of A549 cancer cells after combination treatment with polyamide 3 and etoposide (400 μM) or doxorubicin (75 μM).
**A.4.12 Statistical analysis**

For statistical analysis of significance, data were analysed using either the unpaired, two-tailed Student's *t*-test or one-way analysis of variance (ANOVA) as appropriate. Results were considered statistically significant at a *p* value < 0.05 and were calculated using GraphPad Prism 6.

**A.5 Results**

**A.5.1 DNA binding affinity and sequence selectivity of diamino Hx-polyamides 2 & 3**

The DNA binding properties of the diamino Hx-polyamides 2 and 3 were investigated to assess the effect of introducing an alkyl amino group at the heterocyclic N1 position. Thermal denaturation analysis probed the binding affinity and selectivity of polyamides 2 and 3 by measuring their ability to stabilise duplex DNA. The Hx recognition element behaves similarly to two consecutive pyrrole units and as a result, polyamides 1, 2 and 3 can degenerately bind sequences 5′-ATCGAT-3′ and 5′-TACGAT-3′, with the latter found on the 5′-flank of the ICB2. The Δ*T*<sub>M</sub> data presented in Table A.1 indicate that polyamides 2 (Δ*T*<sub>M</sub> = 30 °C)

<table>
<thead>
<tr>
<th></th>
<th>ATCGAT</th>
<th>ACGCGT</th>
<th>ATGCAT</th>
<th>AAATTT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ΔT&lt;sub&gt;M&lt;/sub&gt; (°C); [K&lt;sub&gt;eq&lt;/sub&gt; M&lt;sup&gt;-1&lt;/sup&gt;]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HxIP (1)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>15; [3 × 10&lt;sup&gt;6&lt;/sup&gt;]</td>
<td>0; [&lt; 10&lt;sup&gt;5&lt;/sup&gt;]</td>
<td>-</td>
<td>5; [4 × 10&lt;sup&gt;6&lt;/sup&gt;]</td>
</tr>
<tr>
<td>HxI*P (2)</td>
<td>30</td>
<td>13</td>
<td>-</td>
<td>15</td>
</tr>
<tr>
<td>HxIP* (3)</td>
<td>32; [3 × 10&lt;sup&gt;6&lt;/sup&gt;]</td>
<td>13; [&lt; 10&lt;sup&gt;5&lt;/sup&gt;]</td>
<td>[2 × 10&lt;sup&gt;6&lt;/sup&gt;]</td>
<td>18</td>
</tr>
</tbody>
</table>

<sup>a</sup> Values presented for HxIP (1) were previously reported.[29]
and 3 \((\Delta T_M = 32 \, ^\circ C)\) bind to the target sequence \((5'-\text{ATCGAT}-3')\) with greater affinity than their monoamino counterpart 1 \((\Delta T_M = 15 \, ^\circ C)\), displaying a 2-fold increase in \(\Delta T_M\) relative to HxIP 1.[29] These results show that the additional positively charged alkyl amino group contributes favourably to DNA stabilisation. However, polyamides 2 and 3 also showed increased \(\Delta T_M\) values for the non-cognate sequences \(5'-\text{ACGCGT}-3'\) and \(5'-\text{AAATTT}-3'\), suggesting an overall decrease in sequence selectivity relative to 1, which has previously been shown to have little effect on the stabilisation of non-cognate sequences.[29] The increased stabilisation of the oligonucleotides by dicationic Hx-polyamides 2 and 3 demonstrated by thermal denaturation is consistent with previously published of comparable N1-derivatized dicationic polyamides.[32, 33]

Circular dichroism (CD) studies confirmed that both diamino Hx-polyamides 2 and 3 bind effectively to the \(5'-\text{ATCGAT}-3'\) target sequence in the minor groove (Figure A.3A), as shown by the emergence of a strong DNA-induced ligand band at \(\approx 330 \, \text{nm}\). The observation of a clear isodichroic point at \(305 \, \text{nm}\) suggests that 2 and 3 bind through a single mechanism, in a presumed side-by-side, antiparallel, stacked orientation. However, HxIP* 3 induced a strong CD band and clear isodichroic point when titrated to the non-cognate sequences, demonstrating that DNA sequence selectivity is reduced by the presence of the second alkyl amino group and corroborating the findings from the DNA denaturation experiments. HxI*P 2 also bound to the non-cognate sequences, but overlaid CD spectra showed that the CD bands were weaker and the isodichroic points less distinct than those observed for 3, confirming the overall lower binding affinity of polyamide 2 and highlighting that the position of the N1 modification affects DNA binding affinity.
To further investigate the effect of the N1-alkyl amino group, the binding constants ($K_{eq}$) of the diamino Hx-polyamide 3 were determined using the surface plasmon resonance (SPR) biosensor assay. The sensorgrams and steady-state plots shown in Figures A.3B & A.3C, revealed that HxIP* 3 binds to the consensus 5’-ATCGAT-3’ site with high affinity ($K_{eq} \approx 3 \times 10^6 \text{ M}^{-1}$ or a $K_D$ of approximately 0.3 μM). Interestingly, diamino Hx-polyamide 3 also binds to the non-cognate sequence 5’-ATGCAT-3’ ($K_{eq} \approx 2 \times 10^6 \text{ M}^{-1}$) almost as strongly as to the consensus sequence, but showed little affinity for 5’-ACGCGT-3’ ($K_{eq} < 10^5 \text{ M}^{-1}$). The favourable binding of Hx-polyamide 3 to the non-cognate sequence 5’-ATGCAT-3’ could be due to strong electrostatic interactions between the protonated alkyl amino groups and the negatively charged phosphodiester backbone of DNA, which could compromise the weaker forces, such as hydrogen bonding and hydrophobic interactions that govern sequence selectivity. Alternatively, we cannot rule out the possibility that a stacked, anti-parallel dimer of Hx-polyamide 3 could interact with 5’-ATGCAT-3’ in a reversed alignment of 3’–to-5’, rather than the typical 5’-to-3’ manner.

A.5.2 Binding of diamino HxI*P and HxIP* on the topoIIα promoter

DNase I footprinting studies on the biologically relevant topo IIα promoter (Figure A.4) reveal that HxI*P 2 binds to the 5’-TACGAT-3’ target sequence of the 5’-flank of the ICB2 with weaker affinity (3 μM) than the monoamino HxIP 1 (1 μM, [28]) despite showing greater affinity in the biophysical studies. HxIP* 3 displayed a 2-fold enhancement of binding affinity relative to polyamide 1, generating a footprint evident at 0.5 μM, which is in strong agreement with the SPR results. Incorporating an additional alkyl amino group had a detrimental effect on the sequence selectivity, with both polyamides 2 and 3 showing off-target binding at a sequence (5’-TTGGTT-3’) overlapping the ICB3.
Figure A.3. Biophysical comparison of the diamino Hx-polyamide DNA binding characteristics. (A) CD spectra for HxIP* 3 (top panel) and HxI*P 2 (bottom) binding with the target sequence 5'-ATCGAT-3' and non-consensus sequences 5'-AAATTT-3' and 5'-ACGCGT-3'. (B) SPR sensorgram of HxIP* 3 with the target 5'-ATCGAT-3' sequence at 100, 200 and 300 nM concentrations of compound. Thick grey lines represent true association and dissociation and thin black lines represent global kinetics fits. (C) Steady-state analyses of HxIP* 3 with 5'-ATGCAT-3' (●, solid line), 5'-ATCGAT-3' (▲, dashed line), and 5'-ACGCGT-3' (◆, dotted line) sequences. SPR experiments were run in 10 mM CCL, 1 mM EDTA, 200 mM NaCl, 0.05% P20 (pH 6.5) at 25 °C.
A.5.3 Disruption of the NF-γ:ICB2 interaction by diamino Hx-polyamides

The diamino polyamides inhibit the NF-Y binding to the ICB2 in a cell-free system with contrasting effectiveness as shown by electrophoretic mobility shift assay (EMSA) experiments (Figure A.5). A radiolabelled oligonucleotide containing the ICB2 and target flanking sequence 5′-TACGAT-3′ was incubated with nuclear protein extract from NIH3T3 cells after pre-treatment with HxI*P 2 or HxIP* 3. Supershift studies using an antibody specific for NF-YA confirmed the presence of NF-Y within the ICB2-bound protein complex (Figure A.5C). HxI*P 2 only affected NF-Y binding at higher concentrations of 10–20 μM, whereas HxIP* 3 displayed a dose-dependent inhibition of the NF-Y:ICB2 interaction, evident at doses ≥ 3 μM. In addition, diamino HxIP* 3 was able to displace NF-Y already bound to the ICB2 (Figure A.5B), when the radiolabelled oligonucleotide was incubated with the nuclear extract prior to the addition of the polyamide. The comparative EMSA study of polyamides 2 and 3 emphasises the importance of optimising the DNA binding affinity and sequence selectivity to deliver polyamide-directed interference of specific protein-DNA interactions.

A.5.4 Nuclear uptake of diamino Hx-polyamides

Emission studies confirmed that the diamino-containing HxI*P 2 and HxIP* 3 retain the characteristic fluorescent output exhibited by Hx-polyamides upon binding DNA and UV excitation (322 nm), with emission bands detected at 370 and 375 nm for compounds 2 and 3, respectively (data not shown). Previously reported cellular uptake studies of HxIP 1 utilised the intrinsic fluorescence of the Hx fluorophore to visualise the rapid polyamide localisation in the nucleus.[28] Confocal microscopy and flow cytometry experiments were also used in this study to assess the nuclear uptake of polyamides 2 and 3. A549 and NIH3T3 cells were treated with
Figure A.4. Binding of diamino Hx-polyamides 2 and 3 to the topo IIα promoter. 
Autoradiograms of DNase I footprinting gels, HxI*P 2 (left) and HxIP* 3 (right). The concentrations (μM) used are shown at the top of the gel. G + A represents a formic acid-piperidine marker specific for purines. The positions of the ICB1, ICB2, ICB3 and the target sequence are indicated.
Figure A.5. Inhibition of NF-Y binding to ICB2 by the diamino Hx-polyamides.
Electrophoretic mobility shift assays (EMSAs) using a radiolabelled oligonucleotide containing the ICB2 and target sequence 5′-TACGAT-3′ were pre-incubated with increasing concentrations of polyamide 2 or 3 for 1 h at room temperature prior to addition of the NIH3T3 nuclear extract. 0, control reaction containing the oligonucleotide and nuclear extract without polyamide; C, reaction in the presence of an excess of unlabelled competitor oligonucleotide of the same sequence as the control reaction; M, reaction in the presence of an excess of unlabelled oligonucleotide with a mutation to the ICB2 motif. (A) Comparative analysis of the inhibitory effects of HxI*P 2 and HxI*P 3 on the binding of NF-Y containing protein complexes to the ICB2. (B) EMSA showing HxI*P 3 is able to displace NF-Y already bound to the ICB2 when the radiolabelled oligonucleotide is pre-incubated with the nuclear extract prior to the addition of the polyamide. (C) Supershift analysis using an anti-NF-YA antibody confirmed the presence of NF-Y in the protein complex bound to the radiolabelled oligonucleotide.
increasing concentrations of polyamides for 24 h and representative confocal microscopy images are shown in Figure A.6. Nuclear uptake of 2 and 3 was confirmed by co-localisation of the polyamide fluorescence signal (blue) and propidium iodide (PI) signal (red). HxI*P 2 and HxIP* 3 nuclear staining increased in a dose-dependent manner, most prominent after exposure to 20 μM of each polyamide. Flow cytometry analysis (Figure S2, Supplementary Material) confirmed the confocal microscopy findings, with a 24 h exposure to each diamino polyamide resulting in a concentration-dependent increase in the median fluorescence intensity of the distribution of the treated cell populations when compared to the DMSO-treated controls. A 41-fold and a 26-fold increase were observed after treatment with HxI*P 2 and HxIP* 3, respectively. It is not clear whether this difference in fluorescent output is due to the enhanced nuclear uptake of 2 or because of lower sequence selectivity, leading to promiscuous nuclear DNA binding and therefore greater fluorescence signal.

To examine polyamide uptake in live, unfixed NIH3T3 cells, the MitoTracker® Red dye, which stains the mitochondria, was used to define the cytoplasm without cell fixation or permeabilisation. The preferential accumulation of the polyamides in the nucleus was observed in NIH3T3 cells, with no overlap between the diamino polyamides and the MitoTracker fluorescence signals (Figure S3A, Supplementary Material). A time course experiment revealed the rapid nuclear uptake and sustained localisation of both diamino analogues following exposure to 20 μM, with blue fluorescence signal visible after just 1 h and still evident after 48 h (Figures S3B & C). These results highlight the advantage of incorporating the fluorogenic Hx moiety to monitor cellular uptake.
Figure A.6. Visualisation of HxI*P 2 and HxIP* 3 nuclear localisation.
A549 (top panels) and NIH3T3 (bottom panels) cells were treated with the indicated concentrations of HxI*P (left) or HxIP* (right) for 24 h, washed with PBS, and fixed with 2% paraformaldehyde (PFA). They were subsequently permeabilised and the nuclei were stained with propidium iodide (PI) before confocal microscopy imaging. The composite image presents the superimposed overlay of diamino Hx-polyamide fluorescence and the PI fluorescence. No polyamide fluorescent signal was detected in the control, untreated cells under the same observation settings.
A.5.5 Biological activity of diamino Hx-polyamides in A549 cells

To explore whether the introduction of a second alkyl amino group enhanced polyamide biological activity, topo IIα expression in confluent A549 cancer cells was evaluated by measuring topo IIα mRNA and protein levels after polyamide treatment (Figure A.7). HxIP* 3 induced a dose-dependent increase in topo IIα mRNA levels as shown by Quantitative RT-PCR, with expression upregulated by 2.1-fold after 18 h treatment with 2 μM polyamide. In contrast, HxI*P 2 had no effect on topo IIα mRNA expression at the same concentration. Immunoblot analysis confirmed the differing biological activities of the diamino HxIP analogues. Treatment with HxIP* 3 mediated a concentration-dependent (0.1–2 μM) enhancement of topo IIα protein levels relative to the untreated cells, with a stimulatory effect exerted at 0.1 μM, whereas, HxI*P 2 had no effect on the nuclear topo IIα content (Figure A.7B). With both diamino polyamides confirmed to localise in the nucleus via confocal microscopy, these findings indicate that the strikingly different in vitro biological activities displayed by 2 and 3 are dictated by the position of the additional alkyl amino group and its effect on DNA binding. Importantly, topo IIα expression in A549 cells is upregulated by the dicationic polyamide 3 at a lower concentration than reported for the monocationic polyamide 1.[28] The enhanced polyamide efficacy may be a consequence of increased DNA binding affinity and water solubility arising from the inclusion of a second positively charged amino group.

A.5.6 HxIP*-mediated chemosensitisation to topo II poisons

The NF-Y-induced transcriptional downregulation of topo IIα at confluence results in cellular resistance to anti-cancer therapeutics targeting the activity of the enzyme. Previously, we reported that the re-induction of topo IIα expression after treatment with HxIP 1 re-sensitised confluent A549 cells to the DNA damaging effects of etoposide, as shown by the increased levels
**Figure A.7. Effect of the diamino HxIP derivatives on topo IIα expression in confluent A549 cells.**

(A) Quantification of topo IIα mRNA levels via RT-PCR analysis. Cells were maintained at confluency before 18 h treatment with increasing concentrations of HxI*P 2 or HxIP* 3. mRNA levels are compared relative to untreated confluent A549 cells (CON). Error bars represent the SEM from three biological replicates. Statistical significance was calculated using one-way ANOVA analysis (*p < 0.05). (B) Immunoblot analysis of confluent A549 nuclear extracts probed with topo IIα antibody following 18 h treatment with HxI*P 2 or HxIP* 3. Lamin is shown as a loading control.

of DNA damage marker, γH2AX.[28] Here, we show the effects of combining topo II poisons with the more biologically active polyamide, HxIP* 3 (**Figure A.8**). Following pre-treatment of confluent A549 cells with increasing concentrations of HxIP* for 24 h, cells were exposed to 50 μM etoposide for 2 h and allowed to recover in drug-free medium for a further 24 h before visualisation of γH2AX. Representative immunofluorescence confocal microscopy images shown in **Figure A.8A**. A revealed that HxIP*-etoposide combination treatment regimes initiated a polyamide dose-dependent increase in the number of γH2AX foci, relative to the etoposide treatment alone. The mean number of foci per nuclei was quantified (**Figure A.8B**) using
Compared to etoposide alone (~0.46 γH2AX foci), combination treatments with 1 μM (~1.6 γH2AX foci) and 2 μM HxIP* (~2.5 γH2AX foci) induced statistically significant 3.5-fold and 5.4-fold increases in γH2AX foci, respectively (Figure A.8B). HxIP* treatment resulted in negligible levels of γH2AX as it does not induce DNA damage due to the non-covalent binding nature of polyamides. Immunoblot analysis (Figure A.8C) corroborated these findings and after pre-treatment with HxIP* there was a concentration-dependent increase in the phosphorylation of H2AX relative to etoposide alone, with the most significant increase observed after pre-treatment with 5 μM HxIP*. A sustained, dose-dependent upregulation of nuclear topo IIα expression by HxIP* was observed in parallel and correlated with the enhanced levels of γH2AX.

HxIP* can mediate the chemosensitisation of confluent A549 cells to the cytotoxic effects of etoposide and doxorubicin. Cell viability was assessed following single agent HxIP* or etoposide treatment, and after HxIP*-etoposide combinations, where 6 h etoposide exposure followed 24 h pre-treatment with HxIP*. Combination of etoposide with increasing concentrations of HxIP* caused a synergistic dose-dependent decrease in cell viability relative to the untreated control. The combination of etoposide (400 μM) with 5 μM HxIP* reduced cell viability by an additional 50% compared to cells treated with etoposide alone (Figure A.8D). HxIP* alone had little effect on viability and displayed no cytotoxic effects below 100 μM in exponential or confluent A549 cells following 24 h treatment (Figure S4A, Supplementary Materials). Whereas, etoposide induced-cytotoxicity following 6 h exposure was attenuated in confluent-arrested cells (Figure S4B, Supplementary Materials), further demonstrating the confluence-mediated resistance to topo II poisons. Finally, polyamide induced cellular sensitisation was also observed in combination with
Figure A.8. Potentiation of topo II poison induced-cytotoxicity by HxIP*.
A549 cells were pre-treated with the indicated concentrations of HxIP* for 24 h and exposed to etoposide (50 μM) for 2 h. After 24 h in drug free medium, they were analysed for γH2AX levels by confocal microscopy and immunoblotting. (A) Representative images of A549 cells. (B) Number of γH2AX foci per nuclei for the indicated treatment combinations, as quantified by CellProfiler Software. (C) Immunoblot analysis of the nuclear levels of γH2AX and topo IIα, 24 h after pre-treatment with HxIP* for 24 h, treatment with etoposide for 2 h and their combination. (D) Confluent A549 cells were treated with HxIP* (2 and 5 μM) or etoposide (400 μM) and HxIP*-etoposide combination, where 6 h etoposide exposure followed 24 h pre-treatment with HxIP*. Cell viability (%) was measured using the CellTiter-Glo® assay. Luminescence values are normalised to those of the untreated control. (E) Viability was also assessed after 6 h treatment with doxorubicin (75 μM) and HxIP*-doxorubicin combinations, where 6 h exposure followed 24 h HxIP* pre-treatment. All data are represented as mean ± SEM (n = 3). Statistical significance was calculated using one-way ANOVA analysis (*p < 0.05, **p < 0.01, ***p < 0.001).

doxorubicin (Figure A.8E). Pre-treatment with 5 μM HxIP* enhanced doxorubicin induced-cytotoxicity causing an additional 22% decrease in viability relative to cells treated with doxorubicin (75 μM) alone. These results confirm the potential application of NF-Y modulating polyamides as chemosensitising agents to increase the cytotoxic potency of topo II poisons and reverse the resistance of confluence-arrested cancer cells.

A.6 Discussion

DNA binding polyamides with programmed sequence recognition are able to chemically control transcription and their gene regulatory activities have been confirmed in various biological contexts, targeting a range of transcription factors such as nuclear hormone receptors [6, 8], hypoxia-inducible factor 1 (HIF-1) [38], nuclear factor κβ (NF-κβ)[7] and c-Myc [39]. Nevertheless, design strategies that further enhance polyamide activity are still required in order to realise the therapeutic potential of this class of small molecules. Recent advances within the field have centred on improving the often-modest nuclear uptake properties, which is essential for polyamide activity as gene control agents.[40-43] We reported that the incorporation of the p-
anisylbenzimidazole (Hx) DNA recognition element into a simple triamide structure increases nuclear uptake and the subsequent biological activity of the polyamide HxIP 1, targeting the repressive DNA binding of transcription factor NF-Y to the ICB2 of the topo IIα promoter.[28] Here, we present the continued evolution of NF-Y targeting polyamides, and explore the modification of the Hx-polyamide framework through the introduction of an alkyl amino group at the N1 position of either the imidazole (HxI*P 2) or pyrrole (HxIP* 3).

This approach to polyamide functionalization was driven by the higher binding affinity and more effective inhibition of transcription factor-DNA interactions by polyamides integrating pyrrole N1-alkyl spermine/spermidine groups.[44, 45] The results of ΔTm and CD studies confirm that the presence of the N1-alkyl amino group increases binding affinity of the diamino Hx-polyamides 2 and 3 to the cognate sequence, because of electrostatic interactions between the additional positively charged amino group and the negatively charged phosphodiester groups of the DNA backbone. However, the position of the N1 modification also affected the binding affinity of isomers 2 and 3. The inclusion of the propyl amino group at the C-terminal heterocycle of the HxIP design (HxIP* 3) caused a greater enhancement of binding affinity than derivatisation of the central heterocycle (HxI*P 2). The difference in affinities between the two diamino analogues became more apparent following footprinting studies to assess their binding to the biologically relevant 5′-TACGAT-3′ sequence on the topo IIα promoter. Diamino HxIP* 3 displayed greater binding affinity (0.5 μM) than the monoamino HxIP 1 (1 μM),[28] whereas HxI*P 2 binds to the target sequence with reduced affinity (3 μM). The weaker binding affinity of polyamide 2 relative to 3 may be due to the closer proximity of the cationic N1-alkyl amino groups of the stacked HxI*P 2 dimers when bound in a 2:1 configuration. Electrostatic repulsions between the positively charged groups and steric effects between the alkyl chains could cause non-optimal polyamide-
DNA binding orientations and compromise DNA binding affinity. N1 modification to C-terminal pyrrole on the other hand, limits the electrostatic and steric clash as the propyl amino groups of the stacked HxIP* 3 dimers are positioned at the maximum possible distance apart.

The importance of the position of the orthogonal propyl amino group in the polyamide design was previously revealed when the binding affinities of 5′-ACGCGT-3′ targeting diamino triamides were shown to be dependent upon the position of the modified heterocycle, without significantly affecting DNA sequence selectivity.[30-33] In contrast, the diamino Hx-polyamides 2 and 3 both display reduced DNA sequence selectivity relative to monoamino 1, with the integration of the N1-alkyl amino group at the imidazole (HxI*P 2) further compromising selectivity compared to attachment to the pyrrole ring (HxIP* 3). This reduced selectivity is attributed to the increased positive electrostatic potential of the diamino polyamides, due to the presence of a second cationic moiety, resulting in greater attraction to the negative potential in the minor groove of A·T-rich sequences. These findings demonstrate how modifications can dramatically influence polyamide-DNA binding characteristics, and underline the importance of functionalities which can enhance DNA binding affinity without affecting sequence selectivity to deliver optimised binding properties.

Engineered to derepress the NF-Y-mediated downregulation of topo IIα, the diamino Hx-polyamides displayed markedly different in vitro biological activities, seemingly dictated by the position of the N1-alkyl amino group and its effect on the polyamide's ability to disrupt the repressive NF-Y:ICB2 interaction. HxI*P 2 had no effect on topo IIα expression at mRNA or protein levels in A549 cancer cells, whereas HxIP* 3 induced upregulation in a dose-dependent manner. Previously, polyamides that displayed no biological effect were assumed to be incapable of penetrating into the nucleus.[27] Here, confocal microscopy and flow cytometry studies
exploiting the inherent fluorescence of the diamino Hx-polyamides confirmed the rapid and sustained nuclear accumulation of both analogues. Therefore, the inability of HxI*P 2 to control topo IIα transcription can be attributed to the polyamide's attenuated DNA binding properties and the resulting moderate inhibition of the NF-Y:DNA interaction. In contrast, diamino polyamide 3 not only exhibited a greater biological effect than 2, but also stimulated upregulation of topo IIα at a lower concentration than the monoamino HxIP 1.[28] EMSA studies revealed that HxIP* 3 and HxIP 1 inhibited the NF-Y:ICB2 interaction with comparable efficiency despite footprinting studies showing diamino polyamide 3 to bind to the target sequence 5′-TACGAT-3′ with higher affinity.[28] This suggests that the greater induction of topo IIα expression by HxIP* 3 is not solely a result of superior DNA binding. The presence of an additional cationic functionality also increases the aqueous solubility of diamino Hx-polyamides. Taken together, these results indicate that the improved biological effects of diamino HxIP* 3 derive from improved physicochemical properties including enhanced water solubility arising from the inclusion of a second positively charged alkyl amino group.

The introduction of a N1-alkyl amino group and its position influenced the DNA binding and biological activities of diamino Hx-polyamides 2 and 3, and highlights why polyamide functionalisation has been widely explored as a strategy to advance their therapeutic potential (reviewed in [46]). Dervan and co-workers have reported on the effect of altering the composition of the C-terminal tail group of non-fluorescent polyamides.[40, 41, 47] The introduction of an isophthalic acid (IPA) group at the C-terminus preserved DNA binding affinity and selectivity, and increased potency in cell culture.[41] The attached IPA group enhanced nuclear localisation and enabled non-conjugated polyamides to replicate the efficient uptake of fluorophore-conjugated derivatives to deliver greater biological effects. An additional adjustment to the C-terminus group
via the substitution of the amide linkage with an oxime linkage between the aliphatic linker and the aromatic tail group further enhanced the potency of polyamides targeting the androgen response element.[42] In a different approach, Sugiyama's group tested the introduction of a hydrophilic methoxypolyethylene glycol (PEG) 750 group as a strategy to address the poor aqueous solubility of their hairpin polyamide seco-CBI conjugates.[48] Conjugates modified by PEGylation at the hydroxyl group of the seco-CBI moiety showed moderately higher solubility and caused greater cytotoxic effects in A549 and DU145 cell lines due to improved cell permeability. Furthermore, work in Dervan's group showed that variation of the hairpin polyamide γ-aminobutyric acid turn (γ-turn) significantly affected the biological efficacy, pharmacokinetics and toxicity.[43, 49, 50] The integration of an aryl group at the β-position of the γ-turn enhanced biological activity against nuclear receptor mediated transcription by two orders of magnitude.[43] However, further studies confirmed that modifications to the α- and β-positions of the γ-turn dramatically altered systemic toxicity of polyamides in mice.[49, 50]

The inclusion of a second alkyl amino group did not change the inherently non-cytotoxic properties of the Hx-polyamides and no evidence of polyamide-induced DNA damage was detected after exposure to diamino polyamide 3. However, when used in combination with etoposide and doxorubicin, HxIP* 3 stimulated the chemosensitisation of confluence-arrested cancer cells to the DNA damaging effects of these topo II targeting agents. Enhancing cellular sensitivity to the cytotoxic effects of topo II poisons through abrogation of the repressive NF-Y binding to the ICB2 was first demonstrated using bis-benzimidazole minor groove binder Hoechst 33342.[16] Upregulation of topo IIα expression in confluent mouse fibroblast NIH3T3 cells significantly lowered the IC50 value of etoposide in combination treatments and inspired the development of sequence specific DNA interacting agents with greater selectivity for the critical
ICB2 sequence. Hairpin polyamide JH-37 bound to the 5′-TTGGT-3’ sequence overlapping the ICB2 and ICB3, and increased formation of DNA strand breaks when NIH3T3 cells were exposed to polyamide-etoposide treatments, resulting in a synergistic reduction in cell viability.[25] Recently, HxIP 1 was the first ICB2-targeting polyamide to sensitisce confluent cancer cells to etoposide.[28] Here, diamino HxIP* 3, our most potent NF-Y inhibiting polyamide to date, re-activated topo IIα expression and the subsequent increase of etoposide-generated DNA damage significantly reduced cell viability in combination treatments. The enhancement of etoposide and doxorubicin induced-cytotoxicity by diamino Hx-polyamide 3 reaffirmed the feasibility of direct modulation of transcription factor NF-Y activity as an approach to chemosensitisation within the context of cellular confluency.

A.7 Conclusion

Polyamide functionalisation via the N1 position of the pyrrole and imidazole rings in the Hx-framework presents a promising approach to improving the drug-like properties of these small molecules. The introduction of an additional cationic alkyl amino functionality enhances water solubility and facilitates the nuclear uptake of the diamino polyamides 2 and 3. However, the N1-alkyl amino group had a detrimental effect on DNA sequence selectivity, and binding affinity for the target sequence 5′-TACGAT-3’ on the topo IIα promoter was dependent upon the position of the modified heterocycle. The diamino polyamides displayed strikingly different in vitro biological activities. HxI*P 2 showed no effect against the downregulation of topo IIα at confluence, while HxIP* 3 stimulated a dose-dependent upregulation of topo IIα expression. This discrepancy is seemingly dictated by their contrasting abilities to disrupt the NF-Y:ICB2 interface. Overall, the correct positioning of a second alkyl amino group in the HxIP structure generated a more potent dicationic polyamide, benefiting from superior solubility relative to its monocationic counterpart.
Targeted inhibition of NF-Y binding by HxIP* 3 and the subsequent re-induction of topo IIα levels potentiated the cytotoxicity of topo II poisons. These results further highlight the need for in vivo investigation of the efficacy of drug combination strategies that employ polyamides to overcome resistance to clinically-used anticancer therapeutics.

A.8 Acknowledgements

This work was supported by a program grant from Cancer Research UK (C2259/A16569 to J.A.H and D.H), a Medical Research Council Doctoral Training Grant awarded to UCL (L.P) and a grant from the National Science Foundation (CHE 0809162 to M.L and W.D.W). The biosensor-SPR work was supported by the National Institutes of Health for W.D.W (GM 111749).

A.9 References


APPENDIX B. THE SEARCH FOR SELECTIVITY OF RNA G-QUADRUPLEXES

BY SYNTHETIC COMPOUNDS

Tam Vo, Sarah Laughlin-Toth, Stephen Neidle, David W. Boykin, W. David Wilson

My contribution to this project was ESI mass spectrometry sample preparation and analysis, and writing.

B.1 Abstract

Telomeres are the ends of chromosomes which help protect the chromosome from degradation during the cell life cycle. The ends are made up of guanine-rich repeats with areas that form structures known as G-quadruplexes. These structures are important in cellular processes and have been identified specifically in relation to certain cancers and neurodegenerative diseases. G-quadruplex topologies are contingent on solution conditions and can occur for both DNA and RNA nucleotide forms. Several RNA quadruplexes have been identified in individuals having dementia with correlations of large quadruplex repeats of 5’-GGGGCC-3’. Little is known regarding RNA G-quadruplexes so their structure and function are of great interest. We have investigated the interactions of four quadruplex-binding compounds for selectivity of RNA quadruplexes over DNA quadruplexes. Additionally, the compounds were also tested with both DNA and RNA forms of the human telomeric sequence 5’-TTAGGG-3’ and 5’-UUAGGG-3’, respectively. Here, we
found morpholine-containing compound, CM03, showed the highest affinity of the compounds tested. Furthermore, CM03 showed very good selectivity for RNA quadruplexes with multiple binding stoichiometries. These results are an exciting step towards the targeting of selectivity for specific quadruplex structures, particularly RNA quadruplexes, with the potential to treat neurodegenerative diseases.

**B.2 Key words**

Anisotropy; G-quadruplex; Mass spectrometry; Neurodegenerative; RNA; Telomere

**B.3 Abbreviations**

Deoxyribonucleic acid (DNA); Electrospray ionization mass spectrometry (ESI-MS); Fluorescence polarization (FP); Guanine (G); Human telomere DNA 21mer (HTel21); Ribonucleic acid (RNA); Human telomere RNA 21mer (rHTel21); telomeric RNA (TERRA)

**B.4 Introduction**

Telomeres are the end regions of chromosomes and contain repeats of nucleotide sequences.[1] These end caps protect the termini of the chromosome from degradation and/or merging with nearby chromosomes. During replication, DNA polymerases are restricted from copying the DNA towards the end cap resulting in a shortened chromosome. Telomeres can, therefore, be considered barriers to protect the chromosome and gene during cell division. In addition to acting as a protector, telomeres also act as regulators for the cell life cycle and signal senescence or cell death. In humans, it is common to find tandem repeats of the nucleotide sequence, 5’-TTAGGG-3’. Often times, the guanine or G-rich regions of these repeats form a unique planar quartet structure called G-quadruplexes (Figure B.1A). These planar regions, or tetrads, are coordinated with four guanine bases that are hydrogen bonded through Hoogsteen base interactions with a core cation.[2]
Figure B.1. Cartoon representation of G-quadruplex formation.
(A) G-quadruplex tetrad formed by Hoogsteen hydrogen bonded guanines coordinated by a core monovalent metal cation. (B) Cartoon representation of stacked tetrads which form the G-quadruplex. (C) Parallel and (D) antiparallel topologies of G-quadruplexes, indicated by 5' to 3' arrows as representative nucleotide backbone.

The stability and topology of a G-quadruplex is often influenced by the internal cation bound within the core. In general, monovalent cations including as K⁺ and Na⁺ are the dominant cations.[3] G-quadruplexes can take on several specific topologies in which the sequence can be intramolecular or intermolecular.[4] The intramolecular topology consists of a single strand whereas an intermolecular topology can be made up of two or four strands. Within these topologies, the backbone can run parallel or antiparallel (Figures B.1C & B.1D). Little is known
regarding the topology of RNA quadruplexes except that they are usually limited to forming the parallel topology due to steric interactions by the hydroxyl groups of the RNA ribose sugar. On the other hand, much more is known DNA G-quadruplexes which can adopt any type of structure.

RNA G-quadruplexes have been identified as markers for neurodegenerative diseases, such as amyotrophic lateral sclerosis (ALS) and other types of dementia, associated with the C9orf72 gene. Healthy individuals were identified to contain an average 2-8 repeats of 5’-rGrGrGrCrC-3’.[5, 6] Unfortunately, individuals with a mutation were found to have repeats ranging from several hundred to thousands. Alternatively, the human RNA telomeric sequence, or TERRA, has also been identified with 5’-rUrUrArGrGrG-3’ repeats in vertebrates. Less is known regarding its function compared to the DNA form but is involved in telomerase regulation processes.[7-9] ESI-MS[10] studies by Collie and co-workers revealed a dimer system of two TERRA while later NMR solution studies showing binding of an acridine ligand bound to two parallel TERRA strands associated at their 5’ ends.[11] It therefore comes as no surprise that targeting G-quadruplexes with small molecules has become a hot topic in the potential treatment of certain cancers and other genetically-related diseases due to the inherent relevance of quadruplexes. If a stacked dimer system can be selectively targeted, RNA quadruplexes may be able to offer an attractive target for potential drug candidates.

### B.5 Materials and methods

**B.5.1 DNA and compounds**

DNA and RNA quadruplexes were purchased from Integrated DNA Technologies (IDT, Coralville, IA, USA). DB1246 and DB1247 were synthesized by Mohamed A. Ismail in Prof. David W. Boykin’s group at GSU. Two sets of sequences were studied by ESI-MS. The first series, 5-GGG GCC GGG GCC GGG GCC GGG G-3 (G4-ALS) and 5-rGrGrG rGrCrC rGrGrG rGrCrC
Figure B.2 Morpholine-containing compounds. (A) MM41 and (B) CM03 tested with DNA and RNA quadruplexes.

rGrGrG rGrCrG rGrGrG rG-3 (rG4-ALS) was designed to mimic ALS-related sequences. The second set of sequences were designed based on the human telomere, 5’-GGG TTA GGG TTA GGG TTA GGG-3’ (HTel21) and 5’-rGrGrG rUrUrA rGrGrG rUrUrArGrGrG rUrUrA rGrGrG-3’ (rHTel21). Stock solutions of 0.5 mM of the ALS-related sequences were dissolved in 150 mM NH$_4$OAc (pH 6.8) prepared in nuclease-free water. Stock solution of 1 mM of the HTel21-related sequences were dissolved in 150 mM NH$_4$OAc (pH 6.8) prepared in nuclease-free water. All samples were aliquot into 25 µL volumes into 200 µL PCR tubes and heated to 98 °C for 10 min in a thermocycler. The heating block was powered off and the samples were slowly cooled for additional 15 min while remaining within the thermocycler. Samples were transferred to a plastic, autoclaved rack, slowly cooled an additional 30 min at 25 °C, and stored at 4 °C until use (within 2 days). Concentrations were spectroscopically determined using extinction coefficients provided by IDT using the nearest-neighbor method.[12]
B.5.2 Electrospray ionization mass spectrometry

Samples were prepared to quadruplex concentrations of 20 µM using 150 mM NH₄OAc (pH 6.8) prepared in nuclease-free water. In separate vials, molar concentrations of compound to quadruplex were prepared to create ratios of [1:1], [2:1], and [4:1]. Compound was added so that the total concentration of compound was equivalent to the desired molar concentration ratio of the total DNA concentration. For example, to prepare a [4:1] concentration ratio of CM03 and MM41 to DNA would be the equivalent of \([\text{DNA}] = 20 \, \mu M + [\text{CM03}] = 40 \, \mu M + [\text{MM41}] = 40 \, \mu M\). Total concentration of DNA in the sample is 20 µM with a total compound concentration of 80 µM (2 x 40 µM = 80 µM) and, therefore, 80 µM of total compound in the sample gives [80:20] or [4:1]. Likewise, a [4:1] ratio of a single DNA to compound would be the equivalent of DNA at 20 µM with CM03 only at 80 µM. An additional 5% MeOH (v/v) was added to the sample seconds prior to injection to help facilitate gas-phase transition to give a total sample volume of 100 µL.

Samples for ESI-MS were run using a Waters Micromass ESI-Q-ToF spectrometer (Waters Corp., Milford, MA, USA) and analyzed with MassLynx 4.1 software. They were scanned from
$m/z$ 500-3000 in negative ion mode for DNA. Negative ion mode was attempted for RNA but resulted in poor signal therefore positive ion mode was used. Flow rates for all samples were of 5 µL/min and the final two minutes of the chromatogram ($\approx$ 200 scans) was averaged. Spectral peaks were deconvoluted using the MaxEnt 1 function in MassLynx 4.1. A G4T4G4 dimer quadruplex was used as a standard to calibrate the instrument parameters. ESI source: capillary 2200 V, sample cone -130 V, extraction cone -115 V, source 40 °C, desolvation 60 °C, cone gas 30 L/hr, desolvation gas flow 450 L/hr. Quadrupole: ion energy, 2.5 V; collision energy, 2 V; RF1 lens, -125; RF2 lens, -103 V. Time-of-flight: acceleration lens 88 V.

**B.5.3 Fluorescence polarization spectroscopy**

Fluorescence polarization spectroscopic studies were performed using 1 mM stock solutions of quadruplex sequences in 150 mM NH$_4$OAc (pH 6.8) in nuclease-free water using a Varian Cary Eclipse Fluorescence Spectrophotometer (Agilent Technologies, Santa Clara, CA) using a 3 mL volume, 4-sided quartz cuvette. Compounds were dissolved to 1 mM stock solutions in nuclease-free water and stored in the dark at 4 °C. Samples were prepared by first diluting compound in the cuvette with 10 mM Tris-HCl (pH 7.4), 50 mM KCl in nuclease-free water, except for CM03 which was also tested using 200 mM KCl. Quadruplex was then titrated in with concentrations of both compound and quadruplex appropriately adjusted in calculations. Anisotropy titrations were used to measure the binding affinity of MM41 to RNA and DNA quadruplex sequences. Direct fluorescence titrations were used to measure the affinity of CM03 to RNA and DNA quadruplexes. Data was fitted using a one-site binding model (1 to 1) to obtain the equilibrium dissociation binding constant ($K_D$).
B.6 Results and discussion

B.6.1 Sequence specific investigations of small molecule interaction with ALS-related quadruplexes

There are many biophysical techniques which can be used to measure the binding of these compounds with DNA and RNA quadruplex systems; however, there is some difficulty in investigating RNA due to instability and/or decomposition of the nucleic acids by naturally-occurring RNase enzymes which can easily break down the nucleic acid. One method which is useful for solution measurements of such systems is fluorescence polarization (FP) spectroscopy. Additionally, because of its native solution properties, low concentrations (picomolar to nanomolar) of sample can be used along with rapid analysis of the sample. Fortunately, both series of compounds, including DB1246 and DB1247 and CM03 and MM41, have inherent fluorescence properties. Figure B.4 shows the anisotropy of DB1246, DB1247 and MM41 with G4-ALS. Beginning with compound at 50 nM concentrations, G4-ALS was titrated in, resulting in increasing anisotropy values ($r$), indicative of ligand-G4 complex formation. Based on the FP results of G4-ALS, the quadruplex sequence showed weak binding with DB1246 ($K_D \approx 240$ nM). On the other hand, both DB1247 and MM41 showed 5-fold stronger binding for the same sequence. Values are listed in Table B.1. Interestingly, interaction between CM03 and the G4-ALS quadruplex sequence was difficult to observe due to significant decreases in fluorescence intensity upon complex formation, therefore, in an effort to measure binding of CM03, direct fluorescence quenching was used instead. Figure B.5 illustrates quenching of 25 nM CM03 upon titrating in DNA. Based on the fluorescence quenching experiments, binding of CM03 with G4-ALS showed the strongest interaction of within this series of compounds.
Figure B.4 Fluorescence anisotropy measurements for G4-ALS quadruplex DNA. (A) DB1246, (B) DB1247, and (C) MM41. Compound concentrations were held near 50 nM in 10 mM Tris-HCl, 50 mM KCl in nuclease-free water (pH 7.4) with increasing concentrations of G4-ALS.

Figure B.5 Fluorescence titration assays of quadruplex DNA and CM03. Fluorescence intensity vs. DNA concentration (left) and fluorescence scans of CM03 upon increasing concentrations of DNA (right). Fluorescence intensities decrease with increasing concentrations of quadruplex species. [CM03] is 25 nM in 10 mM Tris-HCl (pH 7.4) and 50 mM KCl. Maximum λ_{em} = 570 nm for CM03. Data fit using a one-site binding model.
Table B.1 Equilibrium binding constants ($K_D$) of G-4 binding compounds with DNA and RNA quadruplexes.

<table>
<thead>
<tr>
<th></th>
<th>DB1246</th>
<th>DB1247</th>
<th>CM03</th>
<th>MM41</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quadruplex DNA</td>
<td>243 ± 33</td>
<td>40.3 ± 9</td>
<td>0.49 ± 0.8</td>
<td>49.2 ± 6.9</td>
</tr>
<tr>
<td>Quadruplex RNA</td>
<td>410 ± 79</td>
<td>258 ± 60</td>
<td>5.3 ± 1.4</td>
<td>6 ± 1.4</td>
</tr>
</tbody>
</table>

To confirm the one-site binding model used FP analyses, compound and G4-ALS DNA quadruplex complexes were evaluated using electrospray ionization mass spectrometry (ESI-MS). Samples of quadruplex DNA were prepared with and without various ratios of compound. Figure B.6 shows an ESI-MS mass spectrum of G4-ALS with two internal NH$_4^+$ cations within the tetrad core, in the absence of compound. A -5 species was chosen since it had the highest intensity of all observed charge states. Following addition of compound, the -5 remained the dominant charge state even after complex was formed. Binding of CM03 to DNA shows a preferred binding more of 1 to 1 stoichiometry, but some 2 to 1 binding is observed at high concentrations as illustrated in a [4:1] molar concentration ratio (Figure B.7A). Alternatively, DB1246 in Figure B.7B shows similar preferential binding to DNA at high compound concentrations with an additional peak corresponding to 2 to 1 binding stoichiometry. Results by FP indicate very strong binding of CM03 (subnanomolar) to quadruplex DNA; however, the peak intensities in Figure B.7A suggest otherwise, meaning peak intensities for the CM03-G4 complex should be more intense. Likewise, peak intensities for DB1246-G4 complexes should be less in comparison to CM03 complexes since binding is nearly 500-fold weaker. It is believed that that much of the compound is lost during the injection due to the increased number of charges and size of the ligand for MM41 and DB1247, respectively. Regardless, results by ESI-MS agreement well with FP for a one-to-one binding mode.
**Figure B.6 ESI-MS spectrum of DNA quadruplex.**

MW of DNA is 6,940.5. The -5 charge peak was the most abundant species and is shown here with m/z of 1,394. Tallest peak corresponds to quadruplex DNA with two internal NH$_4^+$ ions. Quadruplex shown is in the absence of any quadruplex-binding compounds. DNA concentration is 20 µM and analyzed in negative ion mode.

**Figure B.7 ESI-MS spectra of quadruplex DNA (20 µM) in the presence of ligand.**

(A) CM03 and (B) CM03 each at [4 to 1] molar concentration ratios of [compound to DNA]. Spectra m/z ranges 1,250 – 1,700. Left peaks correspond to unbound DNA. Green labeled peaks shows CM03-DNA and DB1246-DNA complexes of 1 to 1 stoichiometry. The purple label in (A) is of a 2 to 1 stoichiometry of 2 CM03-DNA complex.
A second sequence was investigated with the same set of fluorescent small molecules using the ALS-related RNA quadruplex, rG4-ALS. Using the aforementioned polarization experiments for DB1246, DB1247 and MM41, Figure B.8 shows a similar increase in anisotropy values upon binding with the RNA quadruplex. Similarly, CM03 was again investigated using fluorescence quenching experiments. Unlike the DNA quadruplex sequence which showed a clear preference for CM03 over the other compounds, the RNA quadruplex showed weak binding of DB1246 and DB1247 compounds over the morphonolino series. Table B.1 also lists the equilibrium dissociation binding constants.

Analyses of rG4-ALS with DB1246, DB1247, CM03 and MM41 were expected to produce similar results using ESI-MS as were obtained by FP. The compound-RNA interactions were unfortunately not detectable in negative ion mode due to specific properties of the RNA quadruplex which cannot be explained at this time. While optimizing conditions for ESI-MS, it was found that RNA quadruplex was surprisingly better-detected using ESI positive ion mode.

That being said, results by fluorescence polarization of DB1246 showed the weakest binding for
Figure B.9 Molecular mechanics energy minimized structures. (A) MM41 and (B) CM03, tested with DNA and RNA quadruplexes.

Figure B.10 Fluorescence titration assays of quadruplex RNA with CM03. Fluorescence intensity vs. RNA concentrations (left) and fluorescence scans of CM03 upon addition of RNA (right). Fluorescence intensities decrease with increasing concentrations of quadruplex species. [CM03] is 25 nM in 10 mM Tris-HCl (pH 7.4) and 50 mM KCl. Maximum $\lambda_{em} = 570$ nm for CM03. Data fit using a one-site binding model.
both DNA and RNA quadruplex sequences while CM03 was the strongest binding compound. DB1246 and DB1247, dicationic diamidines, do not have the typical planar, conjugated structure found with many end-stacking compounds. Binding of these compounds with quadruplex can occur through end-stacking, groove-binding, or intercalation and may explain why binding is weak. On the other hand, both MM41 (Figure B.9A) and CM03 (Figure B.9B) share a similar planar structure found within many end-stacking compounds and are similar in RNA binding strength. The RNA quadruplex structure is not selective to either of these compounds unlike the DNA quadruplex which preferentially binds CM03 over MM41 based on the apparent binding constants.

B.6.2 Sequence specific investigations of small molecule interaction with human telomere (HTel21) quadruplexes

To better understand sequence selectivity of the CM03 and MM41 compounds, these compounds were investigated with additional biologically important quadruplex sequences. The human telomere sequences studied contain 21 nucleotide bases with both DNA (HTel21) and RNA (rHTel21) forms. Here, fluorescence polarization and mass spectrometry were again used to study the binding interactions of the DNA and RNA telomeric sequences with CM03 and MM41. Figure B.11 compares the binding of MM41 to DNA and RNA telomeric sequences. Based on the larger increase in anisotropy values (r) at the same quadruplex concentrations, these results indicate that rHTel21 binds MM41 nearly 15 times stronger than the DNA quadruplex. Additionally, the anisotropy value for rHTel21 at saturation is nearly twice as high as the DNA. This anomaly may be due simply to the structure of RNA compared to DNA and/or may suggest more than 1 to 1 binding for MM41 with RNA.
Figure B.11 Fluorescence anisotropy titrations of DNA and RNA with MM41.
Fluorescence anisotropy measurements of MM41 with (A) HTel21 DNA quadruplex (B) TERRA or rHTel21 RNA quadruplex. Compound concentrations were held near 50 nM in 10 mM Tris-HCl, 50 mM KCl in nuclease-free water (pH 7.4) with increasing concentrations of quadruplex.

Table B.2 Equilibrium binding constants (KD) of G4-binding compounds with TERRA and HTel21 quadruplexes.

<table>
<thead>
<tr>
<th>Sequences</th>
<th>K_D (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MM41</td>
</tr>
<tr>
<td>Q-DNA</td>
<td>20 ± 2</td>
</tr>
<tr>
<td>Q-RNA</td>
<td>50 mM KCl</td>
</tr>
<tr>
<td></td>
<td>200 mM KCl</td>
</tr>
</tbody>
</table>

Due to the decreased fluorescence experienced by CM03 when quadruplexes are titrated into the sample solution, fluorescence quenching titrations were again performed to measure the binding affinity of CM03 to both DNA and RNA telomeric sequences. The insets in Figures B.12A & B.12B are the fluorescence scans from which the relative intensity values at λ_em = 570 nm were plotted as a function of quadruplex concentration. The results in Figure B.12 show the
Fluorescence titrations of DNA and RNA with CM03 in 50 mM KCl.

Fluorescence quenching measurements of CM03 with (A) HTel21 DNA quadruplex (B) TERRA or rHTel21 RNA quadruplex. Insets are of the direct fluorescence scan with a maximum $\lambda_{em} = 570$ nm. CM03 concentrations were held near 50 nM in 10 mM Tris-HCl, 50 mM KCl in nuclease-free water (pH 7.4) with increasing concentrations of quadruplex.

Fluorescence titrations of DNA and RNA with CM03 in 200 mM KCl.

Fluorescence quenching measurements of CM03 with (A) HTel21 DNA quadruplex (B) TERRA or rHTel21 RNA quadruplex. Insets are of the direct fluorescence scan with a maximum $\lambda_{em} = 570$ nm. CM03 concentrations were held near 50 nM in 10 mM Tris-HCl, 200 mM KCl in nuclease-free water (pH 7.4) with increasing concentrations of quadruplex.
decrease in fluorescence intensity for CM03 with HTel21 and rHTel21. Binding constants are listed in Table B.2. These values indicate CM03 has a 60-fold stronger preference for RNA over DNA with a $K_D < 1$ nM. To determine a more accurate a binding constant for the RNA sequence, not in the subnanomolar range, salt concentrations were increased to 200 mM KCl with DNA and RNA and CM03. Fluorescence quenching results for CM03 with DNA and RNA quadruplexes are shown in Figures B.13 and binding constants listed in Table B.2. Based on the marked decrease in fluorescence intensity for CM03 with RNA compared to DNA, results using 200 mM KCl agree with those obtained in buffer with 50 mM KCl that CM03 interactions with RNA are stronger than with DNA (90-fold difference).

Based on the calculated binding constants obtained by FP, both MM41 and CM03 show high affinity binding to DNA and RNA telomeric sequences. Interestingly, anisotropy values for MM41 with DNA and RNA quadruplexes indicate a binding mode different for RNA since the r-value was double that of DNA at saturation (i.e. 0.3 to 0.15). At 50 mM KCl, both MM41 and CM03 are selective for the RNA quadruplex with nanomolar and subnanomolar binding constants, respectively. At 50 mM KCl, the DNA quadruplex was more selective for MM41 over CM03 with a binding constant. Higher salt concentrations were then used to obtain a more reasonable binding constant for CM03 with RNA and showed an even more marked preference for RNA over DNA. At this point, we can only speculate the observable differences in binding are due to differences in quadruplex structures between the DNA and RNA. An example of the NMR solution structures of both DNA and RNA are shown in Figure B.14. These structures suggest two very different conformations for the quadruplexes with an expected monomer system for DNA in green, (Figure B.14B) and a dimer system for RNA shown in blue. Superimposition of the two quadruplexes in Figure B.14C further illustrates the distinct conformational characteristics of RNA and DNA.
Figure B.14 Comparison of human DNA and RNA G-quadruplex structures. (A) NMR solution structures of DNA human G-quadruplex and TERRA RNA G-quadruplex, respectively. (B) Single structures of DNA quadruplex and TERRA. (C) Superimposition of DNA (green) and RNA (blue) quadruplexes to compare monomer and dimer quadruplexes. DNA PDB ID: 2kf8, TERRA PDB ID: 2m18.

Figure B.15 ESI-MS spectrum of HTel21 DNA quadruplex. MW of HTel21 DNA is 6,653.3. Tallest peak corresponds to quadruplex DNA with two internal NH$_4^+$ ions. HTel21 shown in the absence of quadruplex-binding compound. HTel21 DNA concentration is 20 µM and analyzed in negative ion mode.

In an effort to determine if the anisotropy values for RNA and DNA were, in fact, accurate in predicting more than 1 to 1 stoichiometry, competition mass spectrometry was used. Both CM03 and MM41 have different molecular weights and can, therefore, be combined in a single sample.
**Figure B.16 ESI-MS spectra of ligand-DNA complexes.**

ESI-MS spectra of HTelo21 DNA (20 µM) in the presence of MM41 (40 µM) and CM03 (40 µM) at a [4 to 1] molar concentration ratio of [compound to DNA]. Top spectrum m/z ranges 6500 – 9000. Blue inset shows free, unbound DNA. Red inset shows CM03-HTel21 and MM41-HTel21 complexes of 1 to 1 stoichiometry. Green inset shows three systems with 2 to 1 stoichiometry of 2 CM03-HTel21, a heterodimer of CM03-MM41-HTel21 (green arrow in center), and 2 MM41-HTel21 complexes.

The compounds compete for binding of the same quadruplex structure which allows determination of the preferred compound and stoichiometry. DNA HTel21 was first investigated in the absence and presence of the compounds. **Figure B.15** of HTel21 is shown and analyzed in negative ion mode. The tallest peak corresponds to the quadruplex with two internal NH₄⁺ cations. Varying concentrations of compounds were then added to the DNA quadruplex, up to [4 to 1] molar concentration ratios, to study the competitive binding. At a [4 to 1] ratio, the tallest peak intensities belong to a 1 to 1 binding stoichiometry for DNA-CM03 followed by 1 to 1 of DNA-MM41 (**Figure B.16**). This was surprising to see since based on FP results, the DNA quadruplex showed stronger binding to MM41 over CM03. However, there was a significant difference in binding
Figure B.17 ESI-MS spectra of RNA in the absence of ligand.
(A) ESI-MS spectrum of HTelo21 RNA (rHTel21) with (B) no detectable dimer system present. MW of rHTel21 is 6,905.2. Tallest peak corresponds to quadruplex RNA with one internal NH$_4^+$ ion. rHTel21 shown here in the absence of quadruplex-binding compound at a concentration of 20 µM and analyzed in positive ion mode.

when salt concentrations were varied in studying the RNA. We can only postulate that since MS experiments require salt concentrations higher than 50 mM KCl (i.e. 150 mM NH$_4$OAc) that perhaps the difference in ionic strength and salt may explain these differences. To our surprise, at these concentration 2 to 1 binding also occurs for 2 CM03 + DNA and 2 MM41 + DNA. More interesting is the heterodimer, 2 to 1 stoichiometry which occurs for 1 CM03 + 1 MM41 + DNA and the higher peak intensity over the homodimer systems.

The interaction of CM03 and MM41 were next evaluated with the RNA telomeric sequence by competition mass spectrometry. Using the same conditions as described with the HTel21, various concentration ratios were examined up to a molar concentration ratio up to [4 to 1]. The first question regarding RNA quadruplex asks what stoichiometry the RNA is present in in solution since Phan and co-workers recently published an NMR solution structure of the TERRA sequence as a dimer complex.[13] To answer this, rHTel21 was first analyzed in the absence of compound. The m/z ranges in which both monomer and dimer conformations would be present revealed little
Figure B.18 ESI-MS spectra of ligand-RNA complexes.
ESI-MS spectra of rHTelo21 RNA (20 µM) in the presence of MM41 (40 µM) and CM03 (40 µM) at a [4 to 1] molar concentration ratio of [compound to DNA]. Top spectrum of (A) m/z ranges 6,500 – 9,000. Blue inset shows where free, unbound RNA would be. Red inset of (A) shows CM03-rHTel21 and MM41-rHTel21 complexes of 1 to 1 stoichiometry with CM03-rHTel21 highly preferred over MM41 binding. (B) ESI-MS spectrum shows 2 to 1 stoichiometric binding of 2 CM03-rHTel21.

Figure B.19 ESI-MS spectra of ligand-RNA complexes.
ESI-MS spectra of rHTelo21 RNA (20 µM) in the presence of MM41 (40 µM) and CM03 (40 µM) at a [4 to 1] molar concentration ratio of [compound to DNA]. Top spectrum m/z ranges 13,800 – 15,600. Blue inset shows dimer RNA. Green inset shows three systems with 2 to 1 stoichiometry of 2 CM03-rHTel21, a heterodimer of CM03-MM41-rHTel21, and 2 MM41-rHTel21 complexes. Heterodimer binding of one CM03 and one MM41 to rHTel21 is preferred of the typical 2 to 1 binding of ligand to RNA.
to no RNA dimer (Figure B.17A & B.17B). Upon addition of both CM03 and MM41 at a molar concentration ratio of [4 to 1], strong peak intensities were observed for RNA-CM03 complexes for both 1 to 1 and 2 to 1 stoichiometry in Figures B.18A & B.18B, respectively. When a 1 to 1 and 1:2 to 1 complexes are formed between RNA and a ligand, there is a very clear and distinct preference for RNA to bind with CM03 as little to no complex is detectable for RNA with MM41. These results agree with those obtained using anisotropy. On the other hand, the m/z range from 13,850 – 15,600 reveals an interesting RNA dimer complex. Not only was a small amount of ligand-free-RNA-dimer detected but 2 to 2 stoichiometries were also present (i.e. 2 ligands + 2 RNA) as shown in Figure B.19. In agreement with FP, CM03 binds much stronger to RNA than MM41 based on the relative peak intensities. More interesting, however, was the heterodimer complex formation for CM03 + MM41 with 2 RNA quadruplexes which had a higher peak intensity than the homodimer, 2:2 complexes. This phenomenon, as observed with DNA, indicates a binding mechanism unique to the combination of CM03 and MM41 with quadruplexes.

B.7 Conclusions

We investigated the binding of several small molecules with DNA and RNA quadruplexes. Fluorescence polarization experiments revealed the binding interactions of diamidine and morpholine-containing compounds with GC-rich quadruplexes with preferential binding of the quadruplexes to the compound CM03. The diamidine compounds DB1246 and DB1247, on the other hand, were generally weaker binding than the morpholine compounds. A second set of DNA and RNA quadruplexes, based on the human telomeric sequence (HTel21) were investigated with the morpholine compounds CM03 and MM41. The RNA quadruplex bound very strong with CM03. Fluorescence and competition mass spectrometry studies confirmed the existence of an
induced dimer for RNA. The unexpected selectivity of CM03 for RNA shows evidence of differences among quadruplexes and may provide significant therapeutic potential.

B.8 References


