The Thesis Committee for Maria I. Lambousis
Certifies that this is the approved version of the following thesis:

Pendant NDI Bisintercalator Derivatives

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor:

Brent L. Iverson

Eric V. Anslyn

Pendant NDI Bisintercalator Derivatives

by

Maria I. Lambousis

Thesis

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Arts

The University of Texas at Austin

August 2017

Dedication

To my Family –

Parents, Brother, and + Γιαγιά Χριστίνα +

Acknowledgements

I would like to thank Professor Brent Iverson for his strong support throughout my time in graduate school. While I expected to gain a great deal of scientific knowledge, I came away having learned far more about myself. Without his continued encouragement and insights, this road would have been a whole lot bumpier.

My fellow Iverson lab mates! After long hours working away together, bonds were made… But seriously, the checkup texts, encouragement, and friendship has really helped and made the daily work pleasant. Who can forget the hours of Wicked and Hamilton, Liz? Recall the Easter egg hunts and hallway shootouts. Better yet, remember the delicious treats of doughnuts, Diet Coke, mochi cupcakes, bubble tea (my 1st intro thanks to Brian), vanilla-hazelnut Einstein Bros. coffee (not regular brew, Chris), Fricano's and of course Taco Joint!

Beyond the immediate crew, Dr. Shoulders, Steve Sorey, and Angela Spangenberg have been invaluable. Thank you for helping acquire 2D NMR spectra of dilute samples and working with me in analyzing information hard to untangle. Even in the last days, thank you to Angela for going back through years of CDs to find a lost file that beautifully displayed the definite answer to the diastereomer question. I also want to thank everyone who guided me in learning new molecular biology skills, starting with Amy and continuing with Costas, the Joes, Rasha, and the ever instructional and informational Carl. Thank you also to Dr. Brodbelt, Dr. Anslyn, and Ms. Betsy Hamblen for their understanding and support.

To Dad, Mom, and lil' Bro, I can't express how thankful and proud I am to claim you as my family. You are my solid foundation that I build upon to the best of my abilities.

Abstract

Pendant NDI Bisintercalator Derivatives

Maria I. Lambousis, MA

The University of Texas at Austin, 2017

Supervisor: Brent Iverson

Sequence specific binding of DNA by small molecules potentially offers the ability to control gene expression. In the past, our laboratory has developed sequence specific threading polyintercalators based on naphthalene diimide (NDI) groups linked head-to-tail by peptides. This design yielded a tetraintercalator with the ability to bind its specific 14 base pair site with a 57 day dissociation half-life, a record for a synthetic DNA-binding molecule (Rhoden Smith and Iverson, 2013). A longer hexaintercalator binds specifically to a 22 base pair site, the longest for a synthetic non-nucleic acid based-DNA binding molecule (Rhoden Smith and Iverson, 2013).

The following work attempts to expand our family of DNA binding molecules by focusing on an alternative bisintercalator design where peptide backbones vary while incorporating NDI units at the ε-amino group of lysine residues, as initially proposed by Dr. Chelsea Martinez (Martinez, 2011). Several pendant or comb-like polyintercalator derivatives analyzed by DNase I footprinting show a preference for GC-rich sequences.

Table of Contents

List of Tables

List of Figures

List of Schemes

CHAPTER 1

DNA-Binding Molecules and Brief Overview of Expanding the Genetic Code

1.1 DNA DRUG TARGET

The template used by organisms to live, replicate, and carry out functions from the simplest to the most complex is known as deoxyribonucleic acid (DNA). In the central dogma of molecular biology, DNA is recognized as the genetic material from which RNA is transcribed and subsequently translated into proteins. The protein products then control cellular functions (Niedle, 2010). Most common drugs today target these downstream protein products, which exist as multiple copies per cell and are regenerated according to cellular need. While effective, each of those individual drugs has undergone several rounds of screening and optimization to target one site on one protein. A more efficient method in treating disease would be to target DNA, as each cell contains only one copy and functions as the control center. Ideally, a modular set of binding molecules with great tunability could target very specific DNA sites to control gene expression. In practice, such molecules would need to bind DNA very specifically to avoid interfering with processes beside the intended. From a clinical perspective, a DNA-binding therapeutic that could find its way to the target DNA site, bind specifically and stay bound for an extended period, would be a relief to patients who now take multiple pills several times a day.

1.1.1 Structure of DNA

Understanding the basic structure of DNA is important in designing molecules intended to bind DNA. While there are three well-known conformations of DNA: B, A, and Z, the most relevant under normal physiological conditions is B form. The following structural information assumes an ideal B form, free of dynamic character that allows for

its transcription, small molecule binding, and similar events. DNA is formed by two polymeric strands of deoxyribonucleotides, where electron rich nitrogenous bases are connected to deoxyribose sugars at the 1' sugar position (Niedle, 2010). Each sugar is connected to the next sugar via a phosphodiester linkage at the 3' and 5' sugar positions. The overall structure then, has a nonpolar hydrophobic interior and a negatively-charged backbone. The two strands are held together through Watson-Crick hydrogen bonds between bases (Vazquez et al., 2003). There are four total DNA bases of two subcategories, pyrimidine and purine. Pyrimidines, **C**ytosine and **T**hymine, are aromatic heterocyclic nitrogenous molecules. Purines, **A**denine and **G**uanine, are composed of a pyrimidine ring fused to an imidazole ring. Traditional pairing rules couple adenine with thymine using 2 hydrogen bonds and guanine with cytosine through 3 hydrogen bonds, making GC bases harder to separate, and increasing melting temperatures for GC-rich sections (Figure 1.1). The chiral centers of the sugars are responsible for B form DNA's characteristic right-handed helical twist (Cheng and Pettitt, 1992). The 36° rotation per base pair results in approximately 10 base pairs for every turn of the helix. The off-set twisting of the phosphodiester backbone produces two grooves down the length of the DNA (Figure 1.2). The major groove is wider and deeper than the shallower but longer minor groove (Niedle, 2010).

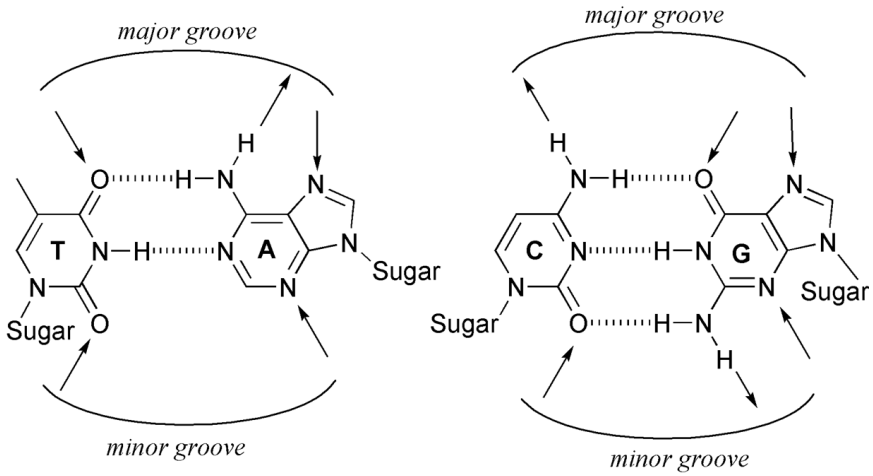


Figure 1.1: Traditional Watson-Crick hydrogen bonding base pairs. Arrows indicate where exposed functional groups in major and minor grooves may act as H bond donors or acceptors (Vazquez et al., 2003).

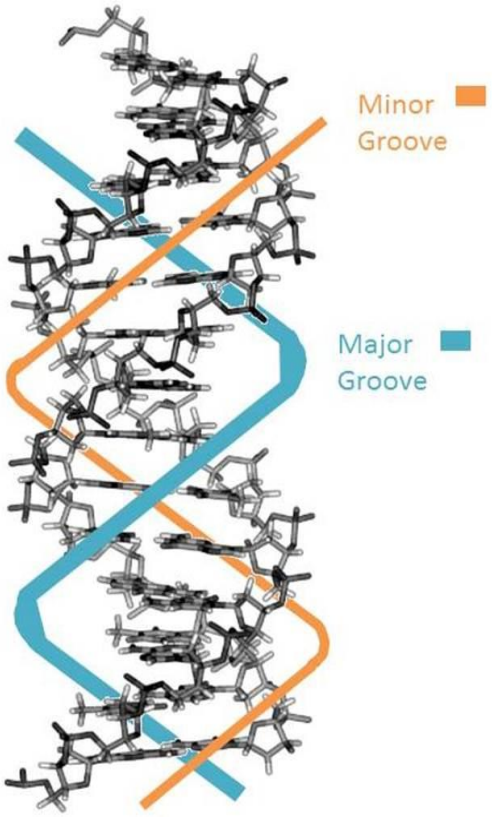


Figure 1.2: Minor groove of DNA double helix outlined in orange and major groove outlined in blue.

1.2 VARIOUS MODES OF BINDING TO DNA

As described above, DNA is a potentially great target for therapeutics. While not all of the following molecules may proceed to become pharmaceutical treatments, an understanding of triplex-forming oligonucleotides (TFOs), peptide nucleic acids (PNAs), minor groove-binding polyamides, and intercalators have made great strides in synthetically based DNA-binding recognition. An overview of these various DNA binding motifs is discussed in section 1.2, followed by a more detailed view of intercalation presented in section 1.3.

1.2.1 Triple-Helix Forming Oligonucleotides

As previously described, Watson-Crick base pairing involves H-bonds between complementary purine-pyrimidine base pairs, giving the traditional B-form DNA structure. When a third strand of DNA, usually polypurine, recognizes and binds a duplex in the major groove, another mode of base pairing is observed (Moser and Dervan, 1987). Hoogsteen base pairs utilize a different geometry than Watson-Crick base pairs, while still employing hydrogen bonding to interact with the other strands (Figure 1.3). In Hoogsteen pairs, the angle between the two glycosidic bonds is greater, the distance between pairs is greater, and the H-bonds differ in location (Cheng and Pettitt, 1992). The resulting three strand nucleotide structures are fittingly termed triple-helix (triplex) forming oligonucleotides (TFO), where the third strand has the ability to bind in the same 5' to 3' direction (parallel) as the polypurine strand or in the anti-parallel direction creating reverse Hoogsteen bonds. Sequence specificity resembles traditional binding in that G-C base pairs will be bound by either a G or a C and A-T base pairs will be bound by an A or T, but there are differences. In polypurine sequences, TFOs best occur at low pH as a protonated cytosine is necessary to bind guanine (Duca et al., 2008). In this case, the polypyrimidine third strand will bind parallel along the major groove as T **:** A-T and

C^+ **:** G-C. When a polypurine third strand binds in an anti-parallel direction with reverse-Hoogsteen hydrogen bonds, the specificity follows A **:** A-T and G **:** G-C. A mix of Hoogsteen or reverse-Hoogsteen binding occurs in parallel or anti-parallel triplexes respectively when the third strand is a mix of purines and pyrimidines with T **:** A-T and G **:** G-C binding (Figure 1.3) (Duca et al., 2008).

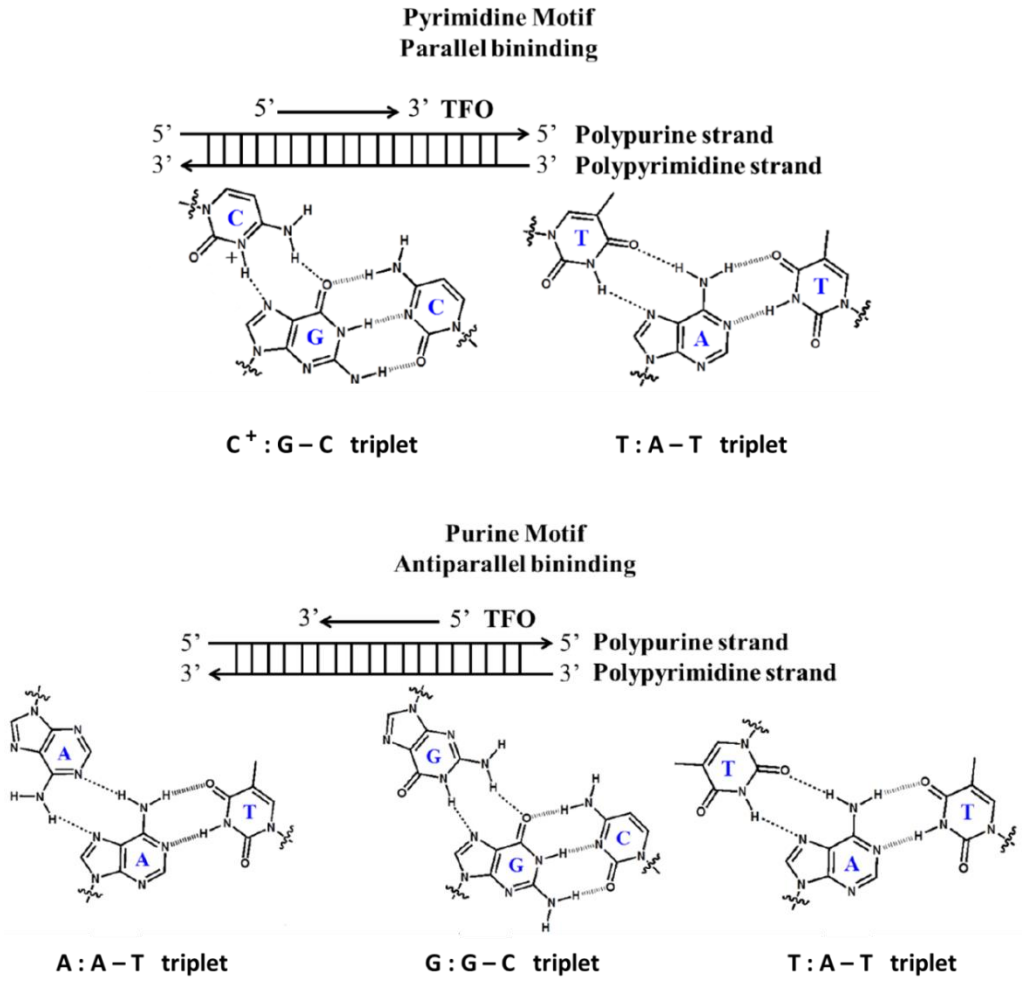


Figure 1.3: Various TFO motifs, with **top** parallel binding and **bottom** anti-parallel binding. Dotted lines indicate (**:**) Hoogsteen H-bonds and dashed lines indicate (–) traditional Watson-Crick hydrogen bonds (Holman, 2011).

While all these triplex variants are possible, the TFO approach to sequence-specific DNA recognition is mostly limited to purine sequences. As previously mentioned, these polypurine sequences are also pH dependent. In addition, charge repulsions must be overcome from the third anionic backbone, which also requires divalent cations to stabilize the triplex, leading to non-physiological salt conditions.

Some methods have been developed to overcome some of these difficulties. Triplex formation at physiological pH became possible through the use of the modified nucleobase 5-methylcytosine (Lee et al., 1984). Furthermore, anionic charge repulsions can be alleviated by replacing the third strand's sugar-phosphate backbone with a phosphoramidate backbone (Giovannangeli et al., 1996). Many have also incorporated multiple DNA binding motifs into one recognition molecule in attempts to overcome the polypurine limitation and enhance thermal stability (Kukreti et al., 1997; Moriguchi et al., 2011; Pedersen et al., 2008).

1.2.2 Peptide Nucleic Acids

Peptide nucleic acids (PNAs) are another class of DNA binding molecules that can recognize complementary DNA sequences. As the name suggests, these structures are a hybrid composition of polypeptides and nucleobases. More specifically, the sugar phosphate backbone of DNA is replaced by a synthetic peptide backbone made of *N*-(2-aminoethyl)-glycine units, with a methylene carbonyl linker connecting the nucleotide bases to this peptide backbone (Figure 1.4). Synthesis is generally easy as PNAs can be made by standard solid-phase peptide synthesis protocols. The resulting structure is an uncharged, achiral, DNA mimic. Peptide nucleic acids are not readily degradable in cells as they are resistant to hydrolytic cleavage and are chemically stable (Ray and Dorden,

2000). Due to their nucleobases, PNAs are capable of sequence-specific recognition of both DNA and RNA. Without a negatively-charged phosphate backbone, PNAs lack much of the electrostatic repulsion found in double-stranded (ds) DNA. As a result, PNA complexes show great thermal stability. In cases where a single PNA strand interacts with a single DNA strand, traditional Watson-Crick hydrogen bonding holds the two strands together. While these complexes are more thermally stable than ds DNA, even a single base mismatch can strongly affect PNA-DNA hybridization (Ray and Norden, 2000).

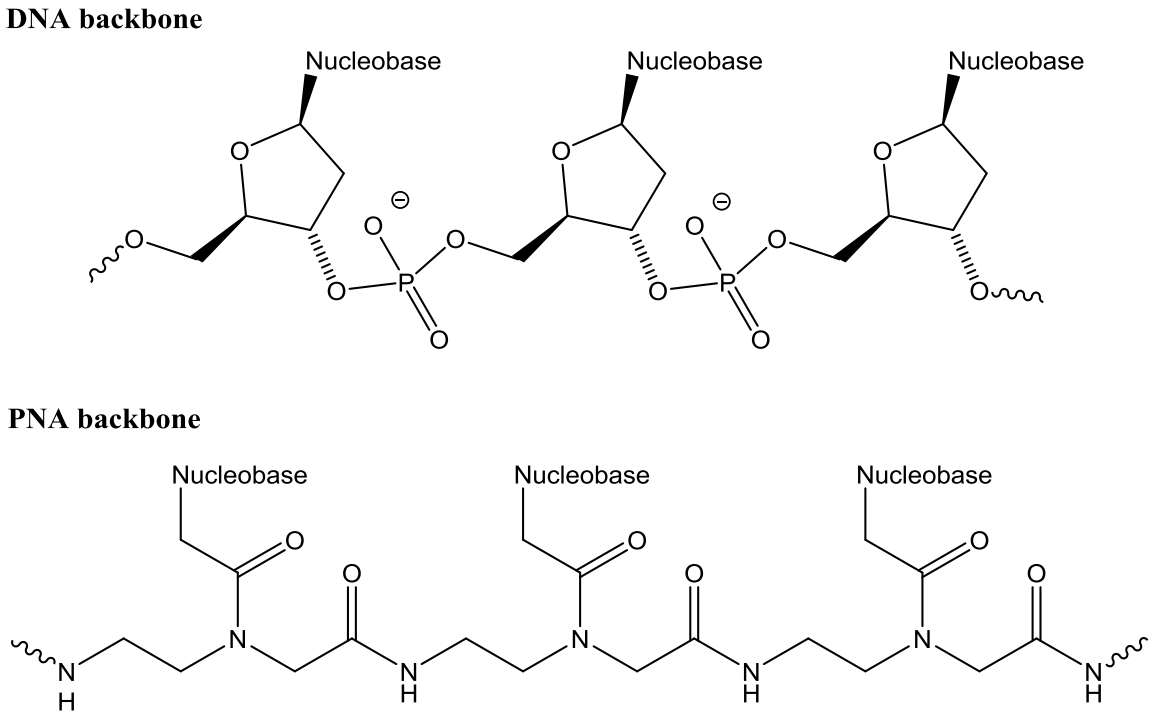


Figure 1.4: Comparison of general structures of DNA backbone and PNA backbone.

Multi-strand complexes may also form. In a 2:1 PNA:DNA complex, both Watson-Crick and Hoogsteen base pairing occurs. Complexes of PNA with duplex DNA require the DNA to have a homopurine sequence for PNA recognition that proceeds by strand invasion. In this case, a stable triplex is formed where the PNA doubles up to form a PNA-DNA-PNA complex with the other DNA strand looped out. A homopyrimidine PNA strand recognizes and Watson-Crick base pairs with the homopurine DNA in an antiparallel orientation. Simultaneously, a second PNA strand will Hoogsteen hydrogen bond with the same DNA strand in a parallel fashion. Multiple models of PNA-dsDNA complexes are possible as is exemplified in Figure 1.5 below. With the ability to prevent transcription or *sometimes* terminate transcription elongation (Nielsen, Egholm and Buchardt, 1994; Lohse et al., 1999) or even promote transcription (Nielsen, 2001) and engage in targeted gene repair (Nielsen, 2010), the various binding modes of PNAs have produced interesting results.

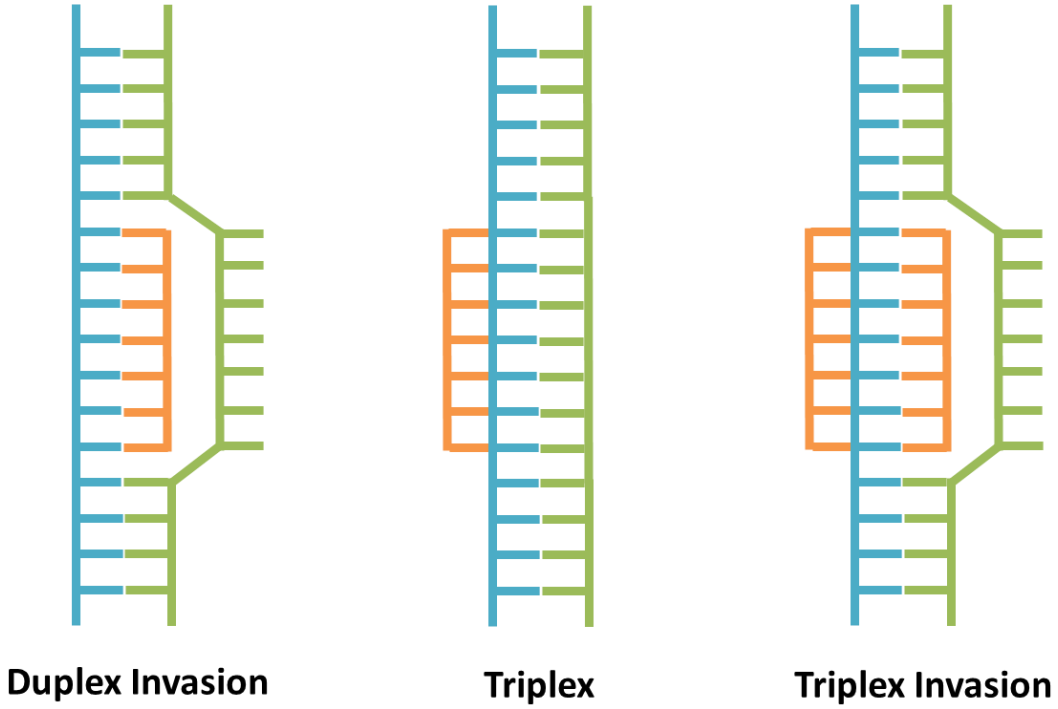


Figure 1.5: Depictions of different PNA binding modes with double stranded DNA. PNA strands are shown in orange and complementary DNA strands are colored green and blue.

1.2.3 Polyamides

Polyamides have proven to be some of the most successful DNA targeting molecules. Using the naturally-occurring antibiotic distamycin as a template, Peter Dervan and coworkers pioneered work with programmable DNA binding polyamides (Figures 1.6 and 1.7) (Dervan, 2001; Melander, 2004)). The three *N*-methylpyrrole amino acids (Py) of distamycin bind AT-rich sequences from the minor groove of DNA (Acramone, 1964). Similarly, synthetic "hairpin" polyamides comprised of Py analogues bind in the DNA minor groove as an antiparallel 2:1 complex. Pairing rules were developed by including imidazole (Im) and *N*-methyl-3-hydroxypyrrole (Hp) moieties

(Figure 1.6) (Dervan, 2001). The Im groups make contact with G,C base pairs by the exocyclic amino group, and Hp moieties distinguish between T,A and A,T base pairs, binding the latter. Furthermore, introduction of β-alanine as a sequence specific flexible spacer allows for adjustment of the polyamides to fit curvature differences between the polyamide and DNA, while recognizing AT-rich sites. Paired antiparallel combinations of Py, Im, and Hp recognize specific DNA base pairs, and follow a clear set of pairing rules (Table 1.1) (Melander, 2004). Additionally, γ-aminobutyric acid (GABA) has been used as a turn unit in hairpin configurations, keeping rings properly paired when folded (Figure 1.7) (Meier et al., 2012). With these pieces, dimeric polyamides were able to target in a cooperative manner a 16 base pair binding site (Trauger et al., 1998).

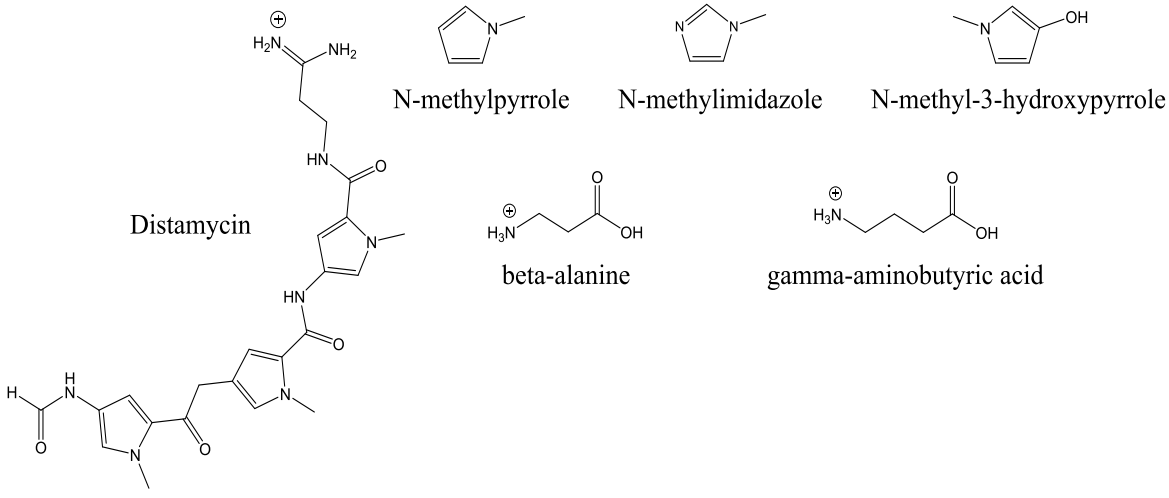


Figure 1.6: Structures of distamycin and polyamide building blocks Py, Im, Hp, β-alanine, and γ-aminobutyric acid.

Pair	GC	CG	TA	AT
Im/Py	+	–	–	–
Py/Im	–	+	–	–
Hp/Py	–	–	+	–
Py/Hp	–	–	–	+

Table 1.1: Pairing rules for polyamides binding in DNA minor groove with a 2:1 stoichiometry. Im = *N*-methylimidazole, Py = *N*-methylpyrrole, Hp = *N*-methyl-3-hydroxypyrrole (Dervan, 2001).

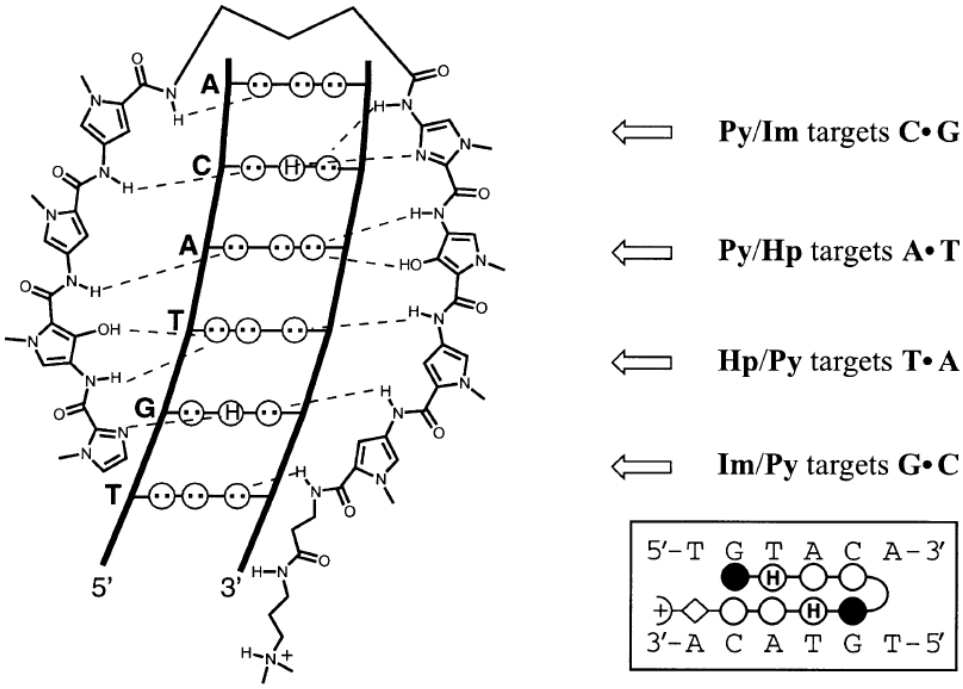


Figure 1.7: Model of antiparallel hairpin ImHpPyPy-γ-ImHpPyPy-β-Dp polyamide bound to 5′-TGTACA-3′ along minor groove. Within DNA model, circles with H signify guanine's 2-amino group, circles with dots denote lone pairs, and dashed lines indicate hydrogen bonds. Smaller representation (bottom right) shows Py as blank white circles, Im as black circles, and Hp as white circles with H (Dervan, 2001).

With their sequence specificity and tunability, polyamides have a large range of applications. Polyamides were developed with the capability of *in vivo* sequence-specific binding of nucleosome bound DNA and transcription prevention by allosterically inhibiting transcription factors (Gottesfeld et al., 2002). Some polyamides were even found to inhibit HIV virus replication (Dickinson et al., 1998). More recently, cyclic polyamide derivatives have been synthesized and show similar affinities for their predicted sequences as their hairpin polyamide counterparts (Li et al., 2013). Some traditional PyIm polyamides have also been found to inhibit major groove CpG methylation in a sequence-specific manner that could lead to the ability to desilence certain genes (Kang, Meier and Dervan, 2014). While arguably the most successful synthetic sequence-specific DNA binding molecules, improvements can still be made to target longer and more difficult sequences. The development of high-throughput screening methods of polyamides will aid researchers in future progress (Meier et al., 2012).

1.2.4 Intercalation

Another mode of binding to double-stranded DNA is through intercalation, generally described as a noncovalent interaction in which a planar aromatic moiety inserts between DNA base pairs. First discovered by Lerman in the early sixties, but hypothesized even earlier by Oster (1951) and Heilweil and Winkle (1955), intercalation causes many changes to take place in the DNA structure (Lerman, 1961; Mukherjee and Sasikala, 2013). Such intuitively destructive effects as unwinding, lengthening, and stiffening of the DNA are energetically offset by the stabilizing effect of removing the intercalator from a polar aqueous environment (Wheate, 2007). The base pairs separate by 3.4 Å as the intercalator slides into place and the reduction of rotation depends upon

the specific intercalator (Strekowski and Wilson, 2007). In order for the DNA to accommodate such rearrangement, only one intercalator can bind for every two base pairs, as described by the nearest neighbor exclusion principle (Bond et al., 1975; Crothers, 1968). Therefore, once one intercalator has bound, a second intercalator is inhibited from binding in the adjacent space.

There are a variety of classes of intercalators, from anthracenes to ellipticines to naphthalimides (NI) among others (Figure 1.8). These intercalators maintain a completely planar polycyclic aromatic moiety within the nucleotide sandwich. However, atypical intercalators that do not exhibit the traditional planar fused aromatic ring structure exist (Fekry et al., 2011).

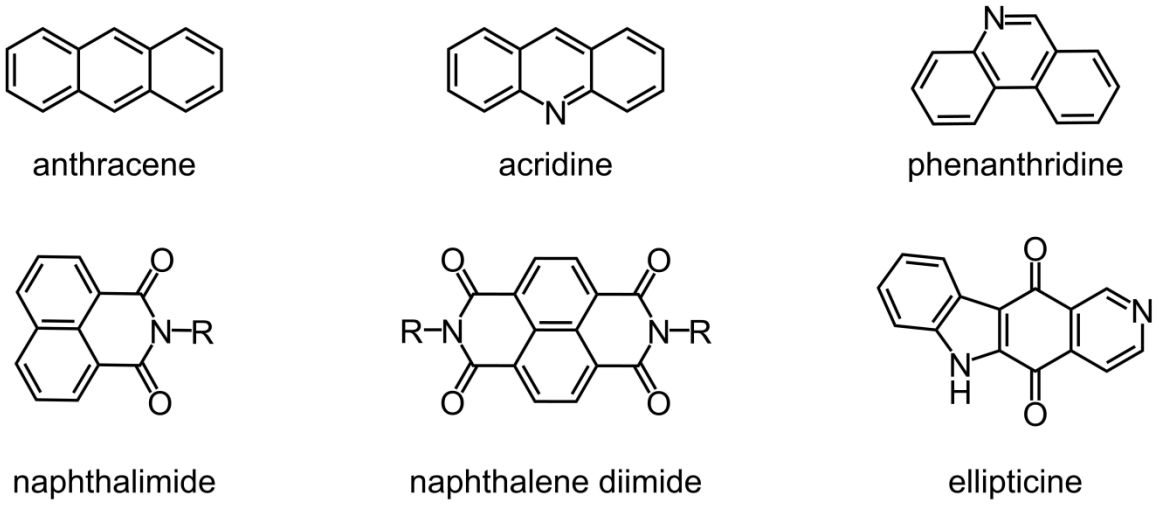


Figure 1.8: Examples of intercalating moieties.

In the interest of selectively targeting specific areas of DNA for genetic manipulation, intercalators would seem an unconventional choice as they generally do not have a preference for particular DNA sequences. However, in accompaniment with

another DNA recognizing motif such as a polyamide, sugar, or peptide among others, intercalator-based molecules can exhibit DNA recognition specificity.

1.3 INTERCALATORS

1.3.1 Monointercalators

The classical intercalator is one that simply inserts between DNA base pairs, parallel to the faces of the nucleotide bases, without extending into the DNA grooves. Proflavin is a typical example of such a classical intercalator (Figure 1.9). Monointercalators, as the name indicates, contain only a single group inserting between base pairs. This core intercalating unit may be substituted, which allows for further interaction between the intercalator molecule and the DNA.

Many monointercalators are found in the biological world and typically serve as dyes or medicines like ethidium bromide or actinomycin D. Until the discovery of penicillin, proflavine and other acridine dye derivatives served as antibacterial agents and wound antiseptics. Their discontinued use was not necessarily due to lack of effectiveness (Mukherjee and Sasikala, 2013). Like many intercalators, these derivatives are cytotoxic and mutagenic but they are also detrimental because they lack selectivity in binding, finding their way into both host and foreign DNA. In the 1940s, the first antibiotic with anti-cancer properties was discovered in the *Streptomyces* species of bacteria and called actinomycin D (Mukherjee and Sasikala, 2013). Proflavin and actinomycin D function as DNA and RNA polymerase inhibitors. The bacterial *Streptomyces* species provided another discovery about ten years later, when the first anthracycline antibiotic, daunorubicin (daunomycin) was isolated and led to the emergence of its own new class of antibiotics (Mukherjee and Sasikala, 2013).

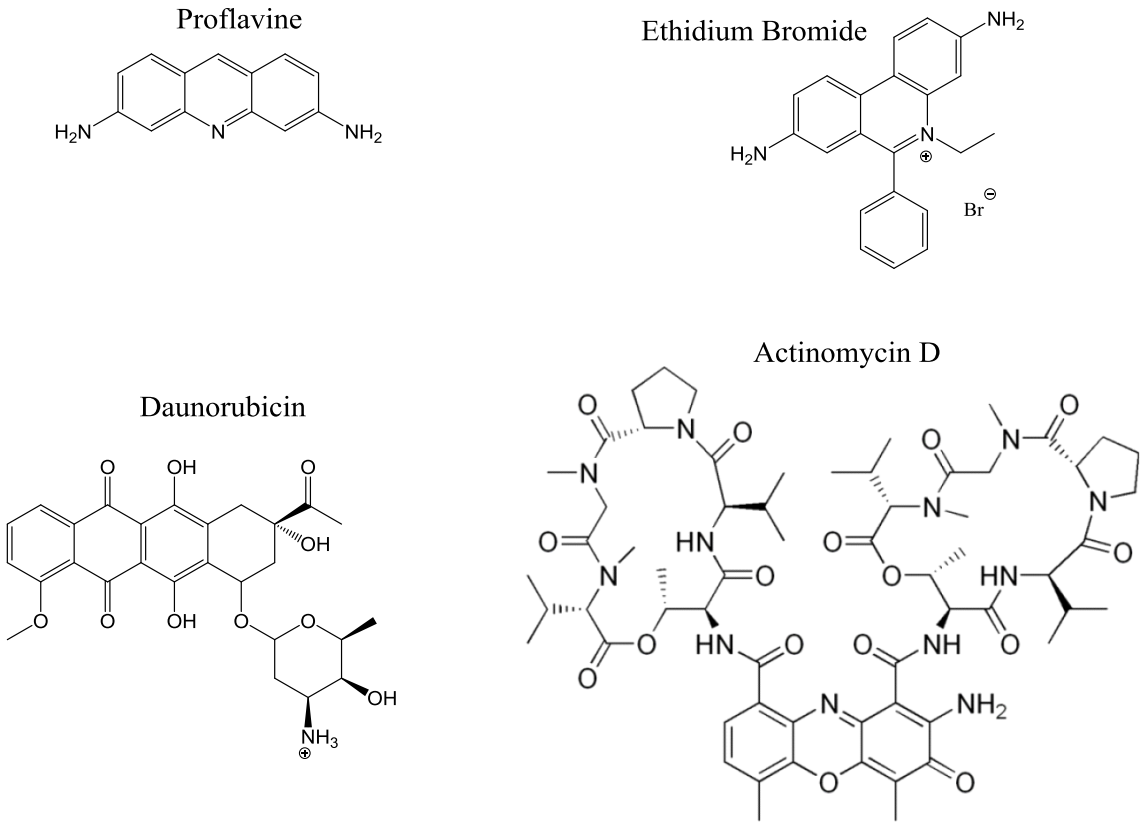


Figure 1.9: Examples of monointercalators.

Recently, a set of 1,8-naphthalimides with varied substituents at position 3 were studied for their binding DNA. Common derivatives included in the study are mitonafide and amonafide. A distinct pattern regarding sequence specificity did not emerge. The naphthalimides were found to not intercalate in AT-rich segments. While a G-C pair is necessary, it is not enough for binding to occur. Additionally, a single GpT step is enough for binding to an AT-rich sequence, but not sufficient to bind to GC-rich DNA (Johnson

et al., 2015). This intercalation study exemplifies the complicated nature of sequence specificity in DNA intercalation, and the difficulties in designing such molecules.

A general understanding of intercalator function is known. Intercalators generally find their way to the DNA and favorable desolvation prompts the intercalators to sandwich within the hydrophobic nucleobase centers (Sasikala and Mukherjee, 2013; Martinez and Iverson, 2012). This is accompanied by an increase in distance between negatively charged phosphate groups that release shielding cations into the aqueous solution (Strekowski and Wilson, 2007). Yet, the exact process of the intercalation is still unclear. A minimum two-step process has been proposed and it may even be that those two or more steps might differ depending on the nature of the particular intercalator (Li and Crothers, 1969; Chaires et al., 1985). Newly developed models of the intercalation process are in close agreement with experimental kinetic data gathered over the decades. Using molecular dynamics (MD) analysis, Sasikala and Muherjee (2013) have proposed that for the monointercalator proflavine, a two-step intercalation process is plausible. The first step is supposed to be a fast initial binding to the minor groove, before a second slow step involves a rearrangement to intercalate through the major groove. The reversal, deintercalation through the major groove is purported to be slow due to the large activation energy barrier required to rearrange back into a groove bound state (Sasikala and Mukherjee, 2013).

The ability of intercalators to find specific sequences of DNA is another area lacking clarity. Most intercalators do not exhibit sequence specificity but some show a preference for GC or AT base pair steps. Both monointercalators proflavine and ethidium show a small preference for CpG steps (Mukherjee and Sasikala, 2013). Müller and Crothers noted while examining a series of intercalators that as the absorbance maximum of the individual intercalator appeared at longer wavelengths, there was an increase in an

intercalator's preference for GC base pairs. They suggested the more polar GC base pair was increasing the polarization of the intercalator (Müller and Crothers, 1975). Sharples and Brown confirmed this conjecture, finding a strong relationship between an intercalator's charge transfer affinity and its preference for GC base pair binding (Sharples and Brown, 1976).

Beyond a broad preference for GC or AT binding sites, some specificity can be imparted. Often only a stronger affinity is observed as the intercalator-DNA association is made stronger with the addition of positively charged (amino) groups present on a protruding substituent group. While affinity is important, there must be a balance between attraction to the DNA in general and the binding to a particular location on the DNA over other sites. Intercalator substituents may lie in either the minor groove or the major groove, or both grooves. Molecules like actinomycin D and duanorubicin (Figure 1.9) are considered threading intercalators. A threading design, explained further in section **1.3.3**, can be more stable as the substituents can interact with the DNA grooves.

1.3.2 Bisintercalators

Connecting two monointercalating moieties with a substituent linker produces a bisintercalator. Likewise, linking more intercalating units together produces longer polyintercalators. While a single intercalator unit can have measureable effects, these longer molecules provide a way to tune the location and duration of binding.

The first discovered bisintercalator natural product was echinomycin (Waring and Wakelin, 1974). It binds to CpG steps surrounded by AT bases, and its peptide linker lies in the minor groove. Like many of the first discovered monointercalators, echinomycin acts as an antibiotic.

Another antibiotic, and derivative of anthracycline, WP631 is a bisintercalator of duanorubucin linked by p-xylene at the amino sugars (Robinson et al., 1997; and Chaires et al., 1997). This bisintercalator binds DNA with its linker in the minor groove, between CpG steps and unlike many other bisintercalators, covers 4 base pairs between the intercalator units. Commonly bisintercalators will span 2 base pairs. WP631 is also the first clear example of an intercalator inhibiting transcription by blocking transcription factor binding (Martin et al., 1999; Portugal et al., 2001).

More unique recognition sites are also possible. Holliday junctions are areas of DNA where a dsDNA molecule will intertwine with a second dsDNA molecule to produce a junction with four separate single strands of DNA, usually occurring during genetic recombination. Such a site is recognized by a bisacridine intercalator (Brogden et al., 2007). The linker allows for the correct distance to place the intercalator units at their two unique sites. Primarily consisting of methylenes, the linker does not seem to drive site recognition. Another acridine intercalator, PT-BIS(ACRAMTU), binds AT DNA and has the ability to block a restriction enzyme from DNA cleavage. It is a bit more unique in that its structure utilizes platinum and so is a metal complex (Choudhury and Bierbach, 2005).

Metal-based molecules that interact with DNA generally consist of coordinating ligands with d6 octahedral or d8 square planar transition metals (Zeglis et al., 2007). The field originated with the discovery of cisplatin as an anticancer drug covalently attaching to DNA. Metallo-insertion is another mode of binding to DNA. In this case, the coordinating ligands are very large and aromatic. When the molecule inserts, one of the nucleobases flips out of the helix. Insertion sites tend to be thermodynamically weaker areas of DNA like mismatch sites, single base bulges, or abasic sites (Zeglis et al., 2007). Metallo-intercalators commonly have rhodium or ruthenium centers and function like

traditional intercalators with the coordination ligand as the intercalating unit. Metallo-interclators are not only capable of recognizing specific sequences, many have the ability to interact with different chiral environments differently depending on the different metallo-intercalator enantiomers (Onfelt et al., 2001; Zeglis et al., 2007). Generally, the design of metallo-intercalators with more than one intercalating unit connects two metal centers with a shared coordinating ligand in a pendant type of arrangement.

1.3.3 Polyintercalation Modes

Two similar but distinct polyintercalator designs are possible. In a pendant design, the intercalator units appear to hang off of a long chain of linked substituents like pendants hanging into the space between base pairs from a necklace that lies along a single DNA groove. In a second intercalation mode the substituents are connected to the intercalator units in a head-to-tail manner, producing a linear design mimicking a thread being stitched in and out of fabric, switching grooves as it passes through the base pairs of DNA. In order for full intercalation to take place in the threading head-to-tail design, one substituent group must pass entirely through the DNA. This makes the process of intercalation more difficult as the DNA must accommodate greater disruption. However, the resulting complex is expected to be more stable and may have greater specificity as the substituents in both grooves can interact with the DNA, and it is intuitively harder to remove such a molecule once it has intercalated (Figure 1.10).

Figure 1.10: Cartoon of threading head-to-tail (left) and pendant (right) intercalating motifs. The linkers are depicted in orange and the intercalators in blue diamonds.

Beside the metallo-intercalators mentioned above, porphyrins are also compounds capable of intercalation in a pendant manner. In one case, a tris-intercalator was designed with an intercalating central porphyrin ring (non-metallated) and two intercalating acridines on either side connected by flexible linkers and agrinyl side groups. Experiments with this molecule maintained intercalation of all three units (Far et al., 2004). Another pendant style tris-intercalator used only acridine moieties to bind DNA using a similar flexible aminoalkyl liner. It was found to prefer poly-AT DNA (Laugaa et al., 1985). Yet, longer polyintercalator derivatives were not able to bind with all four or more acridine units, indicating the linker length needed optimization (Wirth et al., 1988).

Naphthalimides are another class of intercalators and whose bisintercalators are capable of antitumor activity. One such bisintercalator, elinafide (LU 79553) has been investigated by the Brana lab (Bousquet et al, 1995). It has also been shown to preferentially bind to TpG steps, with its alkylamine linker in the major groove (Gallego and Reid, 1999). Although elinafide has not passed clinical trials, Gallego and others

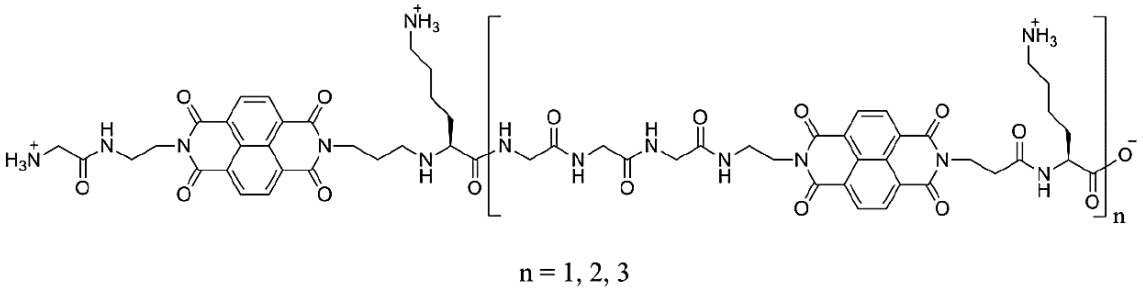
have continued to research similar bisintercalators with modified linkers and substituted naphthalimide cores (Gonzales-Bulnes and Gallego, 2012).

Cousins to the naphthalimides, naphthalene diimides (NDI) are well known intercalators with a propensity to bind to GC-rich DNA. One recent NDI intercalator utilizes imidazolium groups as groove binders, thus combining two ways of binding DNA (Suseela et al., 2016). A set of molecules were made, varying the charge on the imidazolium groups and length between those groups and NDI. The imidazolium groups showed a slight preference for binding to the major groove and absence of a positive charge reduced the propensity for groove binding, which also affected intercalation. The linker length between NDI and imidazole indicated a shorter linker offered better binding. The intercalator with the shortest linker and a positively charged imidazole on both sides proved to be cytotoxic against HeLa cells and inhibited topoisomerase I by binding to supercoiled DNA.

Many of the longer polyintercalators have come from the Takenaka and Iverson groups primarily using NDI as the intercalating unit. The foundations of NDI as a threading intercalator, however, were laid by the Wilson group. The first such description included linkers attached to the imide group, lying within both major and minor grooves, and showed higher binding affinities than the naphthalimide cousins that did not have a threading design. Importantly, it was also shown that bulky substituents could pass through the DNA to accomplish the threading motif (Tanious et al., 1991; Yen et al., 1982). This information coupled with the results of later experiments (Marchetti et al., 2015; McKnight et al., 2011; Rhoden Smith et al., 2012; Tumiatti et al., 2009) related to intercalator linker character give rise to the core design of linking NDI units with peptides usually carrying at least one positive charge in the form of an amino group. The

effects of changing the linker lengths have played an important role in designing longer sequence specific polyintercalators.

The Iverson group began designing intercalators in the 1990s with NDI as the core intercalating moiety joined in a head-to-tail threading manner utilizing a peptide linker. Lokey et al. reported the first fully bound tetraintercalator (Lokey et al., 1997). The study of this tetraintercalator with a strong preference for poly (dGdC) over poly (dAdT) sequences did not include structural support for its mode of binding. This soon led to the first octaintercalator, binding 16 base pairs but without specificity beyond a preference for GC-rich DNA (Murr et al., 2001).



n = 1, 2, 3

Figure 1.11: Initial Iverson Lab threading NDI polyintercalators.

Concurrent with the above studies, a combinatorial library of 360 NDI bisintercalators with varied linkers was manually synthesized and screened by DNase I footprinting on a 231 base pair DNA fragment (Guelev et al., 2000). Manually synthesizing the library was manageable due to the ease of converting naphthalene dianhydride with a primary amine into NDI derivatives, and then by split-pool simple Fmoc solid phase synthesis into the bisintercalators (Guelev et al., 2001b). The goal was to learn more about how different peptide linkers might affect sequence specificity,

among other binding properties. Two bisintercalators came out of this study, those with the linkers Glycine$_3$Lysine (G3K) and β-Alanine$_3$Lysine (B3K). They each recognize a distinct 6 base pair binding site and surprisingly span four base pairs between NDI units, rather than the more common two base pairs. Additionally, NMR structural studies determined that both bind by threading bisintercalation. NMR showed the B3K linker location in the minor groove, and binding to 5' – CG|ATAA|GC – 3', intercalating at GpA and ApG steps (Guelev et al., 2002). The G3K bisintercalator linker lays unexpectedly in the major groove, and binds to 5' – CG|GTAC|CG – 3' with NDIs at GpG and CpC steps (Guelev et al., 2001a). The important result was that these two threading bisintercalators exhibited sequence specific binding to DNA, with recognition through opposite DNA grooves.

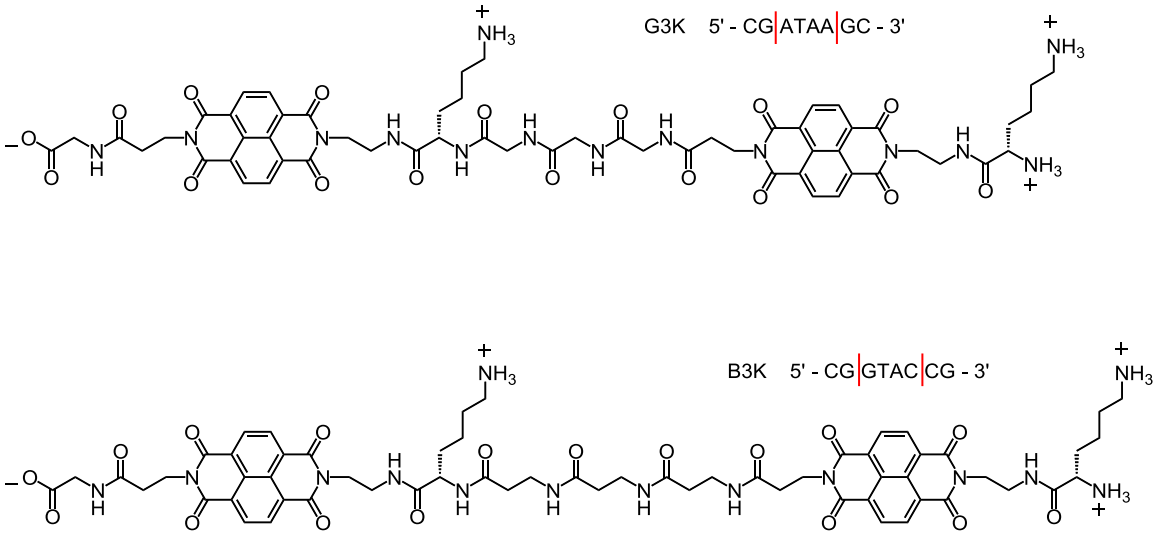


Figure 1.12: Threading bisintercalators G3K and B3K with respective binding sites.

Another tetraintercalator was then designed utilizing the B3K and G3K intercalator information to thread through the minor, then major, then minor grooves and bind to a specific sequence comprised of the original bisintercalator sites. However, rather than use the G3K linker, an adipic acid linker was substituted, mimicking the length and hydrophobicity of the original linker but providing C2 symmetry to the tetraintercalator. The symmetry then allowed for simplified NMR structural analysis of the tetraintercalator in its palindromic 14 base pair DNA binding site 5' – G|ATAA|GTAC|TATT|C – 3' (Lee et al., 2004). Kinetic studies followed that indicated a dissociation half-life of 16 days, setting a record at the time for a synthetic molecule binding DNA (Holman et al., 2011). This record was then surpassed by adding one methylene unit (using pimelic acid rather than adipic acid) to the major groove binding section of the tetraintercalator, now displaying a dissociation half-life of 57 days from its preferred 14 base pair binding site (Rhoden Smith and Iverson, 2013). A hexaintercalator was also synthesized and analyzed, but while binding a record 22 base pair binding site, it displayed a considerably faster dissociation rate compared to the tetraintercalator (Rhoden Smith and Iverson, 2013).

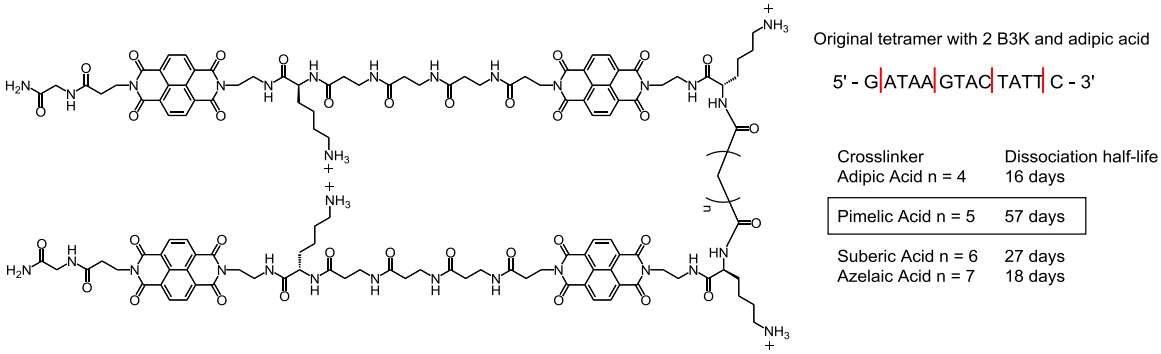


Figure 1.13: Tetraintercalators with binding site and dissociation half-life.

Clearly, optimizing the linkers joining intercalator units together is of great significance to both binding specificity and stability. With a strong foundation of head-to-tail threading intercalators already laid, questions arise: *Can new and possibly longer binding sites be accessed by utilizing a pendant threading design? Can an entirely new library be built using engineered bacteria with a modified NDI amino acid as the intercalator and pendant linker?* Mimicking the manual combinatorial library of Guelev, a high-throughput combinatorial library of pendant bisintercalators, if successful, could very quickly lead to many new sequence-specific polyintercalators with access to new binding sites.

1.4 General Principles of Genetic Code Expansion

As stated previously, DNA is transcribed to RNA and RNA is translated into proteins. In translation, a group of three successive nucleotides, known as codons, specify an individual amino acid. Traditionally there are 64 codons, but only 20 amino acids. This uneven pairing means the code for turning nucleic acid into protein is degenerate. Transfer RNAs (tRNAs) serve as an adapter to bridge nucleic acids and peptide sequences. Aminoacyl tRNA synthetases (aaRSs) are enzymes that covalently bind tRNAs to their respective amino acid. (Figure 1.14). The ribosome then catalyzes a reaction between the tRNA and the template messenger RNA (mRNA), transferring the amino acid onto a growing peptide chain. When an aminoacyl tRNA synthetase goes to esterify a tRNA with the proper amino acid, the aaRS does not necessarily make the pairing by recognizing the tRNA anticodon. If there is a mutation in the anticodon but the aaRS-tRNA recognition site is unchanged, the amino acid has now changed to be encoded by a new codon. It is this flexible relationship between the aaRS, tRNA, and

amino acid that in recent years has been exploited to incorporate non-canonical or "unnatural" amino acids within ribosomally produced polypeptides, thereby expanding the genetic code (Hendrickson et al., 2004). Molecular biology now uses this "flaw" as a tool to study protein structure and function, evolve new organisms with varied properties, and possibly even run high-throughput syntheses of modified peptides among the range of applications.

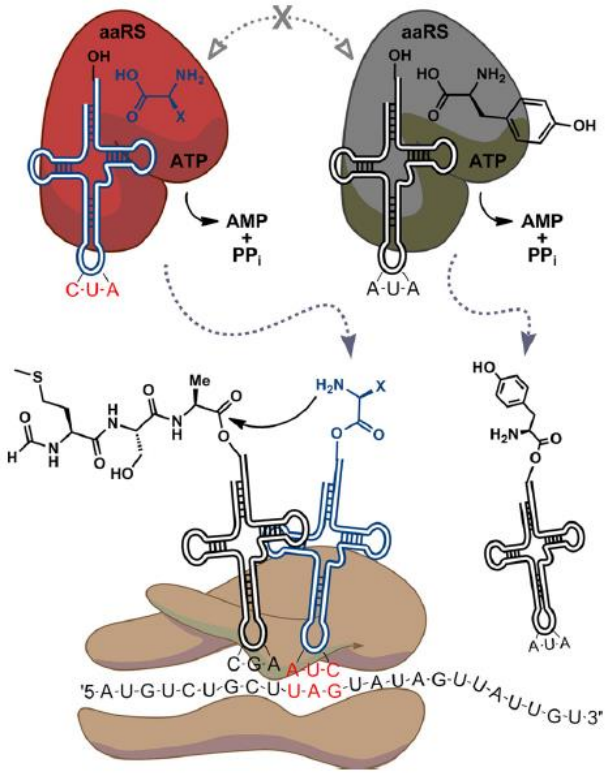


Figure 1.14: Translation machinery where two orthogonal aaRS/tRNA systems are displayed. The red aaRS aminoacylates its blue tRNA (red CUA anticodon) with a blue unnatural amino acid. The native aaRS/tRNA pair in green and black (AUA anticodon) does not interfere with the 'unnnatural' pair nor vice versa. The brown ribosome accepts the red/blue acylated $tRNA_{CUA}$ matching the UAG codon. The unnatural amino acid is thus incorporated into the growing peptide (Young and Schultz, 2010).

The Schultz group has made great strides in creating organisms that genetically encode many more than the 20 canonical amino acids (Young and Schultz, 2010). Multiple methods can be used to accomplish this task, but there are criteria to be taken into account. The new unnatural amino acid must be metabolically stable and not already be a substrate for any aaRSs native to that organism. The new tRNA must be the only tRNA within that system to recognize the unique codon while also being specific to accept only that new amino acid. And the new aaRS/tRNA pair must be orthogonal to that organism's naturally occurring aaRS/tRNA pairs while being functional in that environment (Liu and Schultz, 2010).

1.4.1 Incorporation of Large Modified Amino Acids

Currently, more than 150 unnatural amino acids have been genetically encoded into various organisms (Dumas et al., 2015). While methods are consistently improving to incorporate a wider variety of molecules, adding large polyaromatic moieties has either been less successful or not as necessary to those working in the field. Aromatic groups like derivatives of naphthalene, anthracene, and pyrene have been tethered to amino acids (Figure 1.15) (Hohsaka et al., 1999; Speight et al., 2013).

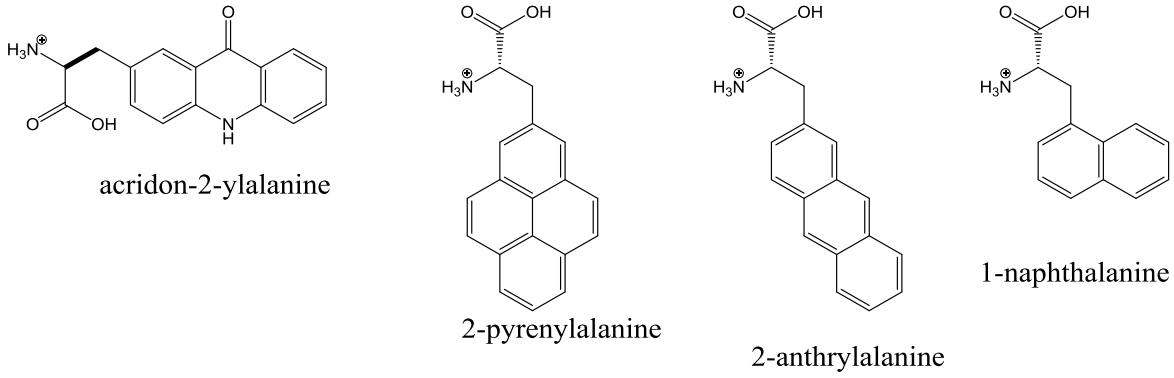


Figure 1.15: Examples of genetically encoded large aromatic unnatural amino acids.

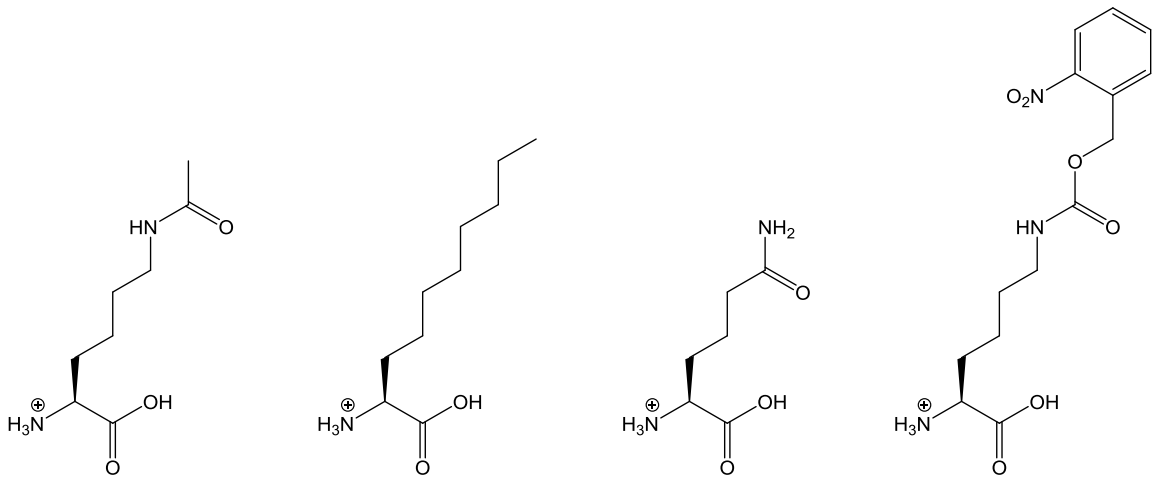


Figure 1.16: Examples of genetically encoded long chain unnatural amino acids (Liu and Schultz, 2010).

More unnatural amino acids are now utilizing lysine, having a longer side chain, but usually carry smaller groups at the *N*-ε position (Yanagisawa et al., 2008; Mukai et al., 2008; Li et al., 2009; Fekner et al., 2009). Finding a synthetase and ribosome able to accommodate both a long side chain and large, bulky group seem to be limiting factors (Armen et al., 2010; Hohsaka et al., 1999; Lacey et al., 2013). Something approaching the length and size of an NDI modified lysine is an encoded coumarin lysine analogue (Luo et al., 2014). Providing multiple functions, these analogues can be used as fluorescent cellular probes and can trigger protein function by optical activation (Luo et al., 2014).

CHAPTER 2

Optimization of Pendant NDI Bisintercalators

2.1 CHAPTER SUMMARY

2.1.1 Goals

The overarching goal is to create a vast library of pendant intercalators by high-throughput synthesis through ribosomally produced unnatural peptides using an expanded genetic code. Such a library could then be screened for the most sequence specific binding molecules, expanding our access to unique DNA sequences and the ability to modulate gene expression. The immediate goal is to determine the optimal linker length between each NDI intercalating unit to provide a basic template for interchanging of linker amino acids.

2.1.2 Approach

Previously, six pendant bisintercalators of two lengths and varying charge were synthesized and preliminarily evaluated by DNase I footprinting, possibly indicating the same multiple binding sites across all derivatives (Martinez, 2011). Two new pendant bisintercalator derivatives have been synthesized with one less amino acid in the backbone while continuing to monitor the effect of charge by incorporating lysine in one derivative where the other has glycine. These were analyzed by DNase I footprinting for binding to a DNA sequence containing all possible six base pair palindromic binding sites (Hampshire and Fox, 2008) and subsequently to a sequence designed to include sites of interest based on the palindromic sequence footprinting results.

2.1.3 Results

The two new pendant bisintercalator derivatives display binding to the same GC-rich DNA sites as the longer derivatives. Changing from glycine to lysine did not change the binding locations. With increasing charge the binding affinity appears to increase.

2.2 BACKGROUND

2.2.1 Threading to Pendant Design

Sequence specific binding of DNA by small molecules could possibly impart the ability to control gene expression. One manner of binding to double-stranded DNA is through intercalation, generally described as a planar aromatic moiety inserting between DNA base pairs. Most intercalators do not have a preference for particular DNA sequences unless in accompaniment with some other DNA recognizing motif such as peptide nucleic acids, triplex forming oligonucleotides, and polyamides. Yet, even when intercalators incorporate these features, there is no guarantee of great sequence specificity. An octakisintercalator, built with eight intercalating units connected by peptide linkers, only bound to regions of GC-rich DNA with no higher affinity towards any one particular sequence (Murr et al., 2001).

Our group has developed a set of modular polyintercalating molecules based on the electron deficient intercalator 1,4,5,8-naphthalenetetracarboxylic diimide unit (NDI) connected to other NDIs in a head-to-tail manner via flexible peptide linkers (Guelev et al., 2001, 2002). Most notable of our molecules are a hexaintercalator that binds to a 22 base pair site, the longest for a synthetic non-nucleic acid based DNA binding molecule, and a tetraintercalator that binds its 14 base pair site with a record setting dissociation half-life of 57 days for a synthetic DNA binding molecule (Figure 2.1) (Rhoden Smith and Iverson, 2013; Holman et al., 2011).

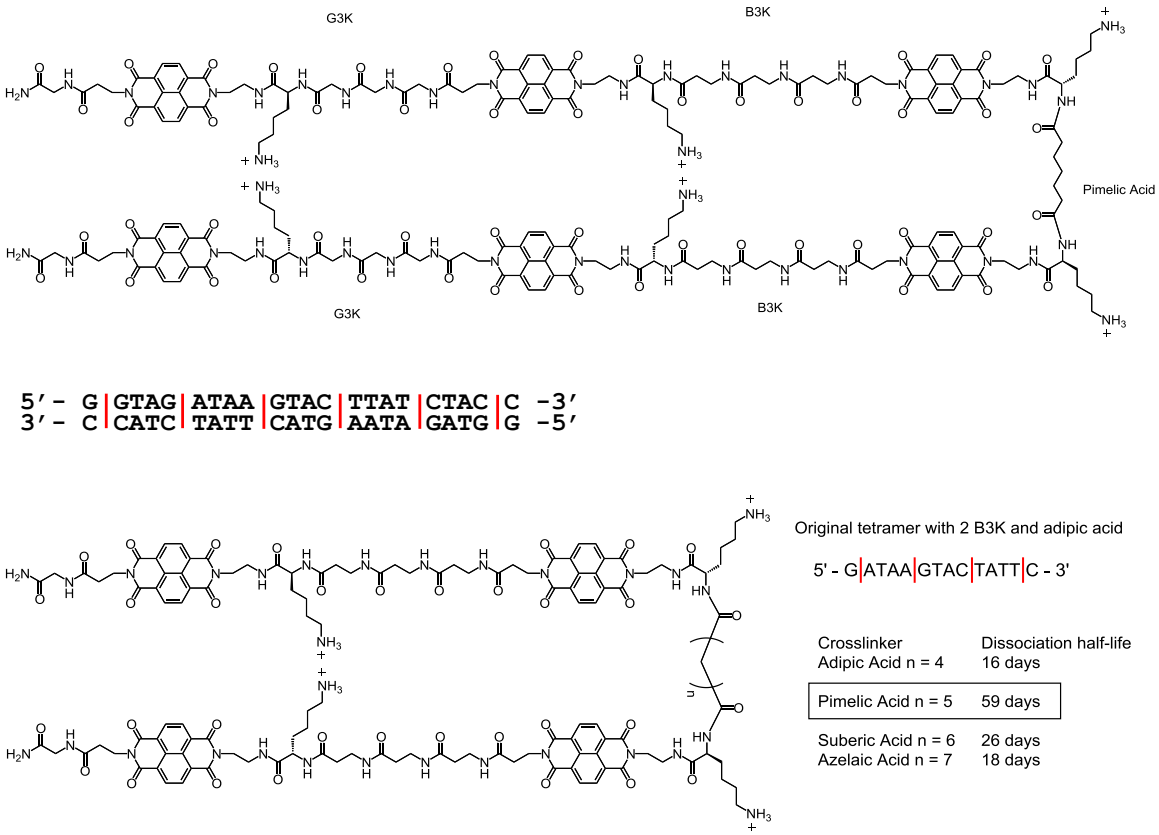


Figure 2.1: Top: Hexaintercalator with its 22 base pair binding site. Bottom: Tetraintercalator and dissociation half-life.

Two NDI threading bisintercalators, with linkers Glycine$_3$Lysine (G3K) and β-Alanine$_3$Lysine (B3K), have been quite successful as the basis for constructing our sequence specific polyintercalators. The discovery of these bisintercalators from DNase I footprinting by screening of a manually synthesized library was followed by NMR determination of each molecule's binding site (Guelev et al., 2000, 2001, 2002). The G3K bound to the palindromic 5' – CG|GTAC|CG – 3' 6 base pair binding site in the major groove of the DNA, while the B3K bound in the minor groove to another 6 base

pair binding site 5’ – CG|ATAA|GC – 3’ (Figure 2.2). A tetraintercalator made by combining two B3K and a G3K with a binding site comprised of the respective bisintercalator sites was studied by Amy Rhoden Smith. More unique results arose from a modified tetraintercalator where two B3K were joined with adipic acid rather than G3K, to impart C2 symmetry and allow for easier NMR structural analysis. This tetraintercalator bound a 14 base pair site 5’ – G|ATAA|GTAC|TATT|C – 3’ in a threading manner along the minor groove, through the major groove, and again into the minor groove (Lee et al., 2004). The dissociation half-life of the molecule from its binding site was 16 days, a record at the time (Holman et al., 2011). Binding was improved to a 57 day half-life by adding one methylene unit to the major groove binding portion, using pimelic acid (Figure 2.1) (Rhoden Smith and Iverson, 2013).

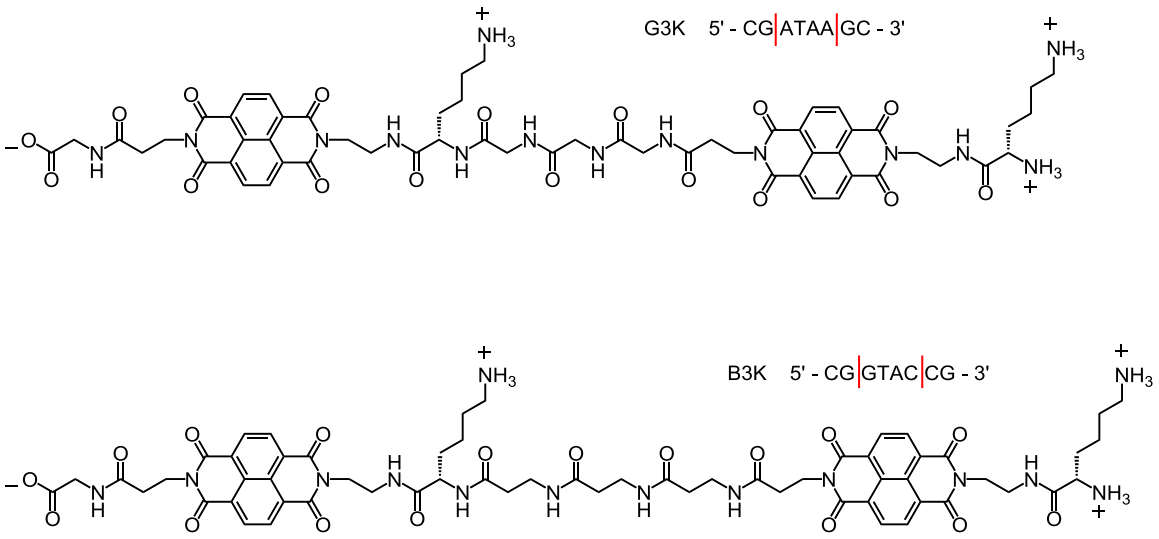


Figure 2.2: Threading bisintercalators G3K and B3K with respective binding sites.

Many substantial findings have been made using the original G3K and B3K bisintercalators. Yet, we are limited because currently we have only one unique sequence for each DNA groove. While there are more molecules to be synthesized, the G3K binds with less affinity than the B3K, so analyzing a tetraintercalator using two G3K and one B3K for example would likely bind with lower affinity than our prior tetraintercalators. Part of the strong binding exhibited by our intercalators is due to their threading design, but the design can also be a hindrance. Longer polyintercalators, like our hexaintercalator, may also bind with lower affinity considering how finely tuned each linker section must be to bind longer sequences. In addition, longer threading polyintercalators are expectedly more difficult to completely intercalate all units considering the degree of rearrangement and unwinding the DNA must experience.

Rather than continue making variations on the same theme, Chelsea Martinez decided it may be more fruitful to expand our group of intercalators by utilizing a different binding topology. She designed a set of pendant, or comb-like, bisintercalators that incorporate the intercalating moiety off of an amino acid side chain, removing NDI from the backbone. The result creates a branching molecule rather than a threading linear molecule (Figure 2.3). More specifically, lysine residues are modified at the *N*ε position with an NDI unit similar to intercalators synthesized by the Takenaka group (Martinez, 2011; Nojima et al., 2003). Ethylenediamine modifies the second imide of the NDI, technically also making these threading intercalators, but the overall design is that of a pendant. The ability to modify this second NDI imide position offers the option of later combining the pendant design with our well-defined threading bisintercalators to access even more DNA sequences through a variety of groove binding topologies.

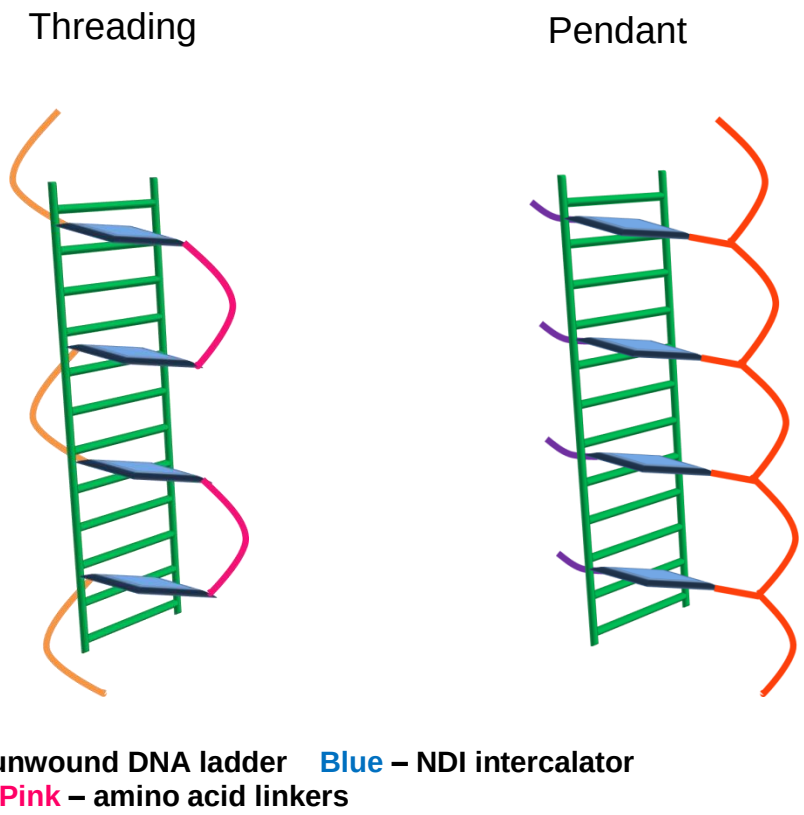


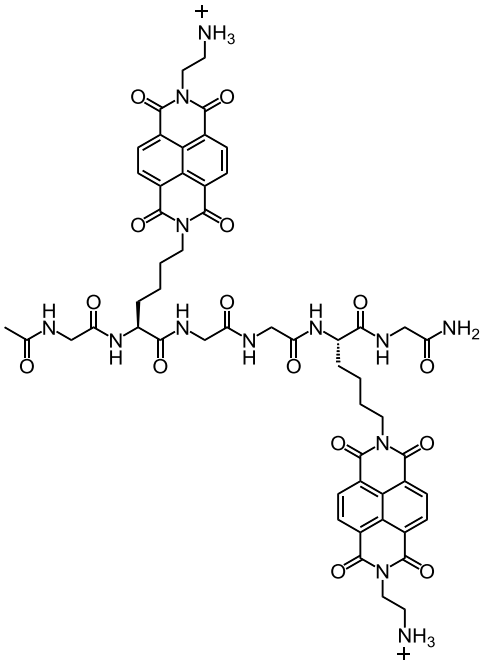
Figure 2.3: Cartoon of threading vs. pendant modes of intercalation shown on unwound DNA ladder.

The synthesis of our intercalators follows solid phase peptide synthesis (SPPS), even the pendant derivatives, as the pendant modification takes place during production of the NDI monomers. While relatively straight forward, this modification may pose a new risk of racemization. Any new piece introduced to a known system can yield unexpected results.

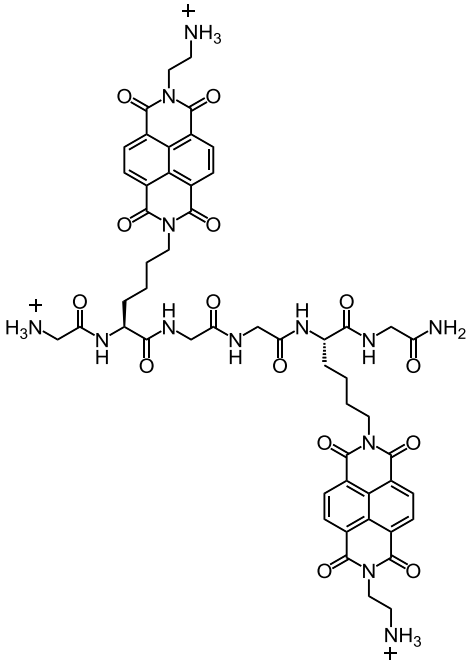
The original set of six pendant bisintercalators **1-6** (Figures 2.4 and 2.5) were designed to resemble the G3K and B3K bisintercalators primarily in length. Differences in charge were expected to affect affinity, though binding specificity could also be

different. Flexibility and hydrogen bonding differ from the threading intercalators as the NDI moves to a branched position, allowing more freedom of movement and reducing H-bonding closest to the NDI due to the multiple lysine methylenes. In attempting to target the major groove, the length between NDI imide nitrogens of **1**-**3** was kept close to that of G3K, equating to 18 atoms. Similarly, in targeting the minor groove the linker length between each NDI of **4**-**6** and that of B3K was kept as consistent as possible at 21 atoms while considering how an extra amino acid would be needed for the pendant linker to account for the extra methylene of the β amino acid. Taking into account the number of amide bonds and amino acids, the pendant intercalators are slightly longer than their threading counterparts. This was expected to be advantageous as longer linkers seem to allow for better binding (Rhoden Smith and Iverson, 2013). Each of the six pendant bisintercalators was modeled *in silico* with the respective NMR determined six base pair DNA binding site of the threading bisintercalators to ensure similar length and stable geometry (Martinez, 2011). Preliminary DNase I footprinting results with DNA sequences containing all possible six base pair palindromic sites indicated all six pendant derivatives bound to the same multiple GC-rich sites, possibly four base pairs, not the anticipated six base pairs as the model threading intercalators. Differences in charge showed some increased affinity as charge increased. Considering the longer intercalators bound the same sites as the shorter derivatives, an optimal linker length had not been found.

1



2 / C



3

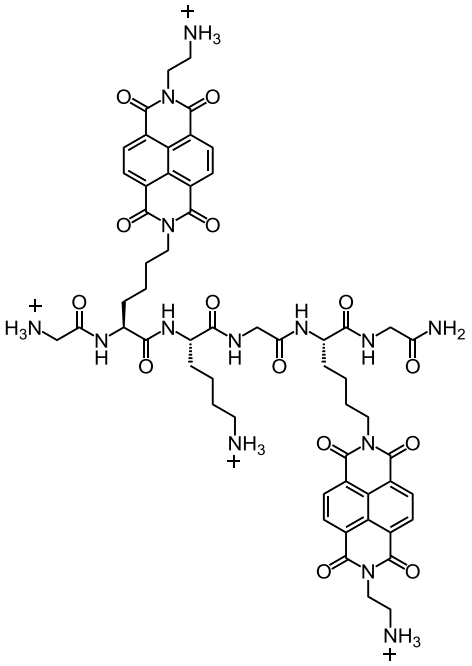
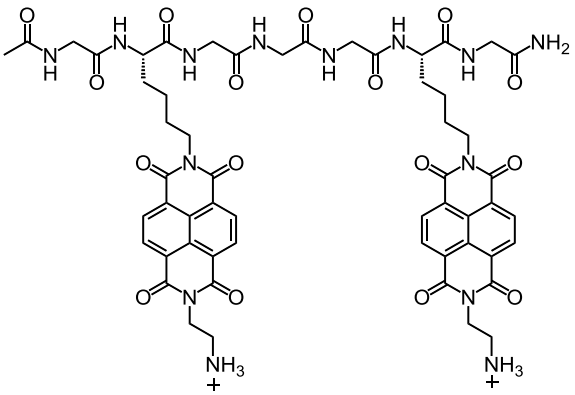
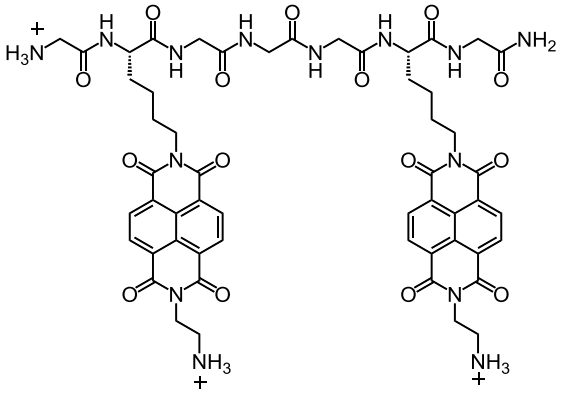


Figure 2.4: 18 atom Pendant bisintercalators **1**-**3**.

4



5 / D



6

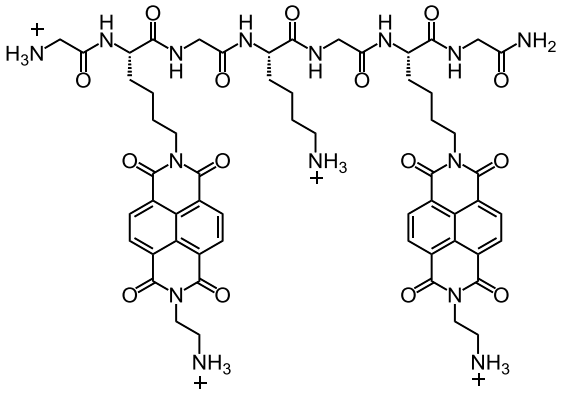


Figure 2.5: 21 atom Pendant bisintercalators **4**-**6**.

Herein is described the design, synthesis, and footprinting of two new shorter pendant NDI bisintercalator derivatives (Figure 2.6 **A** and **B**) and a comparison to two previously described longer pendant NDI bisintercalators (Figure 2.7 **C** and **D**). In striving to optimize the pendant design, the new shorter derivatives have 15 atoms between DNI imide nitrogens, one amino acid less than the molecules with 18 atom linkers. This new length will be a determining factor of future pendant designs. Additionally, intercalator **A** with a 15 atom linker contains a lysine residue to ensure selection of a linker with the appropriate number of charges. No clear footprint would indicate at least one NDI unit is not fully intercalated and the new linker is likely too short. Based on the nearest neighbor exclusion principle, having two intercalating moieties between adjacent base pairs is unlikely to occur due to lack of space, as the extent of rearrangement needed to allow one unit to intercalate encroaches on the area separating neighboring bases (Williams et al., 1992; Bond et al., 1975; Crothers, 1968). More specific binding would indicate the longer, previously designed derivatives are too long. Longer linkers may lack interaction with the DNA, making them unable to impart required specificity.

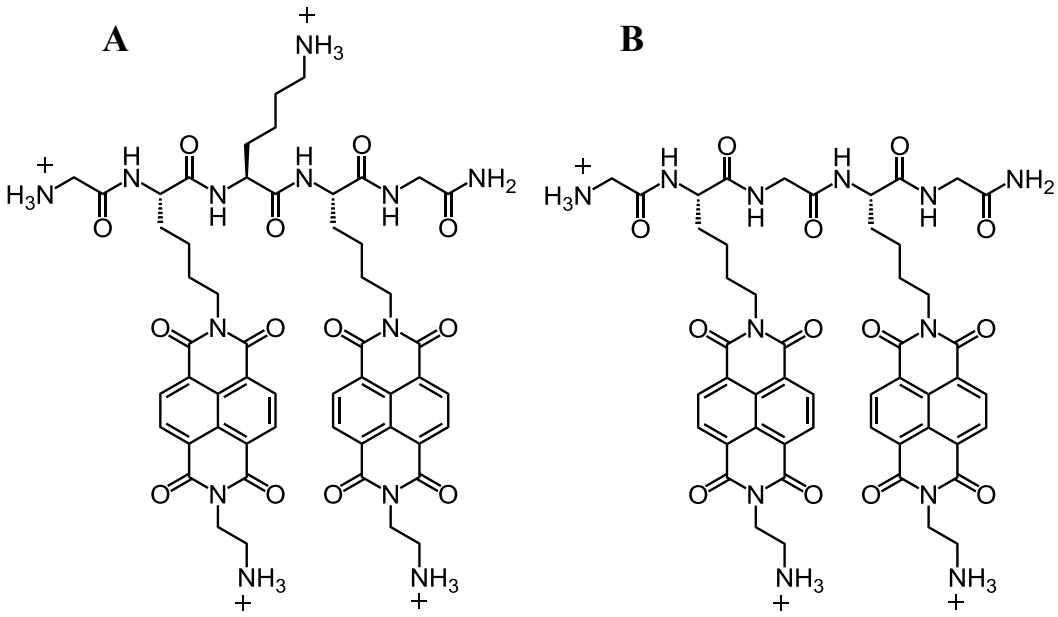
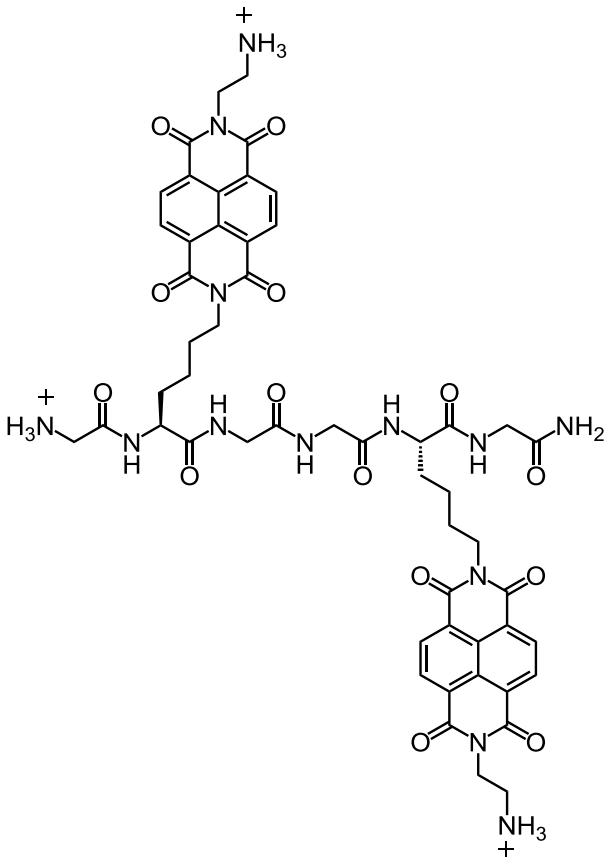
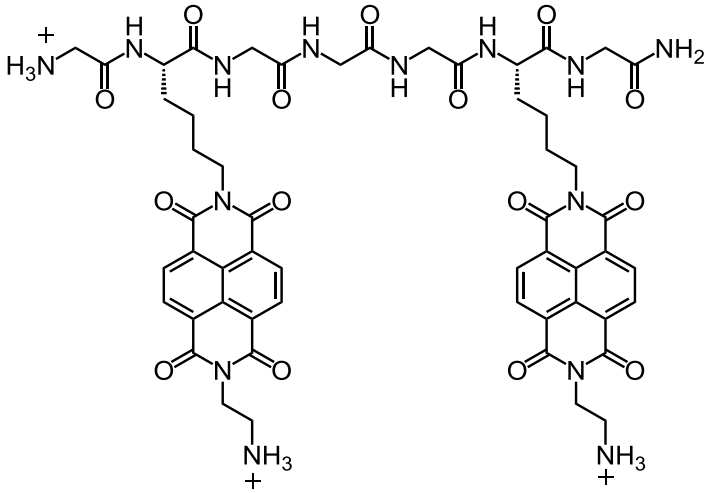


Figure 2.6: Pendant bisintercalators **A** and **B**.



C



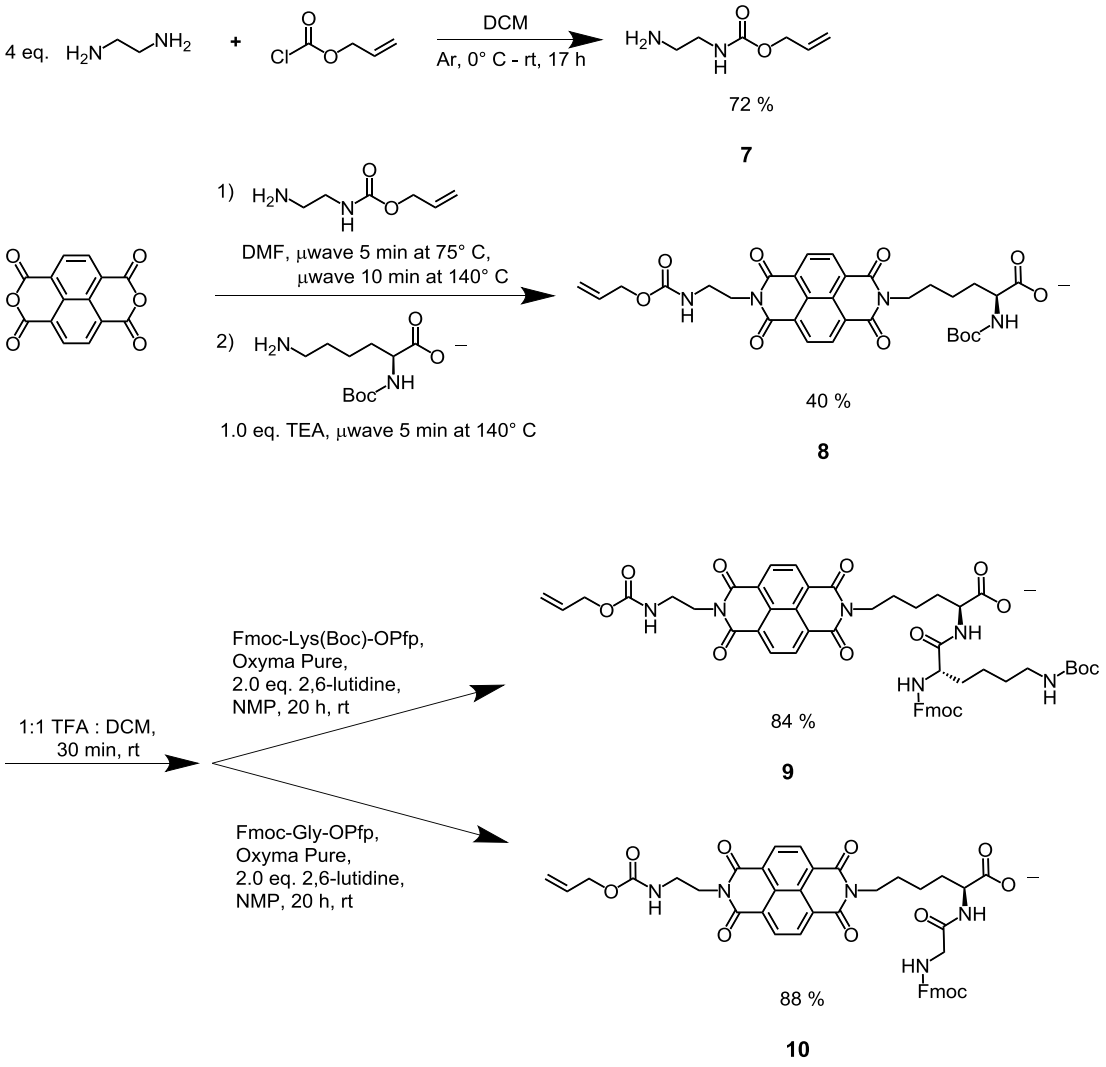
D

Figure 2.7: Pendant bisintercalators **C** and **D**.

2.3 RESULTS

2.3.1 Synthesis

Synthesis of pendant bisintercalators **A**, **B**, **C**, and **D** (Scheme 2.1) began with an NDI-lysine monomer that was subsequently coupled to another amino acid in solution, followed by solid phase peptide synthesis (SPPS) similar to the protocol outlined by Guelev (Guelev et al., 2001b; Martinez, 2011; Tambara et al., 2011). Standard Fmoc SPPS employed orthogonal Alloc protection for the ethylenediamine NDI side chain and orthogonal *t*-Boc protection for the lysine side chains. Two different monomers were synthesized Fmoc-Lys(Boc)-Lys(NDI-Alloc)-OH (**9**) and Fmoc-Gly-Lys(NDI-Alloc)-OH (**10**) for later analysis of charge effects.

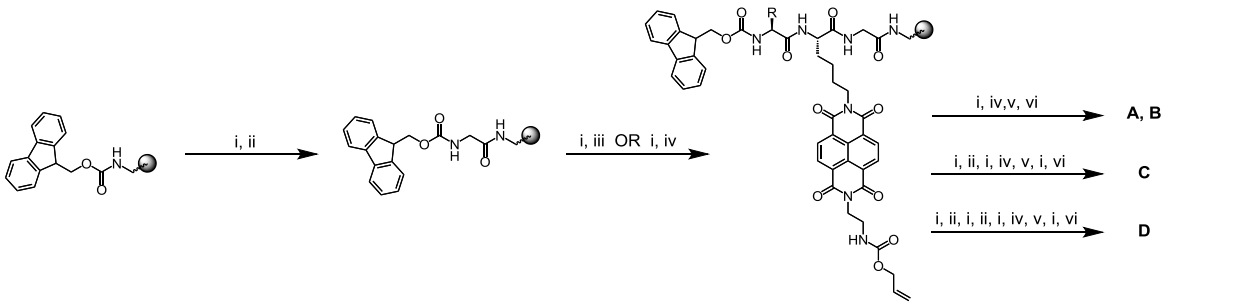


Scheme 2.1: Solution synthesis of NDI-lysine monomers for SPPS.

Previous pendant intercalators **3** and **6** also incorporated lysine in the linker; the lysine was placed at least one amino acid residue removed from the previously coupled modified NDI-lysine monomer (Martinez, 2011). By shortening to the 15 atom linker, no individual amino acids were added by SPPS between monomers, necessitating a new Fmoc-Lys(Boc)-Lys(NDI-Alloc)-OH (**9**) monomer. This was easily made by using

Fmoc-Lys(Boc)-OPfp rather than Fmoc-Gly-OPfp during the solution phase peptide coupling reaction.

Changes were also made in the SPPS procedure (Scheme 2.2). The capping steps have been eliminated as the Fmoc protecting group appears too labile under coupling conditions; capping results in predominantly truncated molecules with an acetamide terminus. Similar Fmoc instability was observed in developing previous synthetic protocols (Guelev et al., 2001b). Coupling conditions now utilize 2,4,6-trimethylpyridine (TMP/collidine) for the base due to its ability to reduce epimerization (Han et al., 1997; Carpino et al., 1996; Subiros-Funosas et al., 2010). Also, the Alloc deprotection is done in a 1:1, THF : DMSO solvent system, increasing yield (Brase et al., 2002, 2003; Rigby et al., 1986; Albericio et al., 2000).



Scheme 2.2: Solid phase peptide synthesis for bisintercalators **A**-**D**.

Prior to the changes made in the solid phase protocol, bisintercalator **A** could not be isolated. Multiple products giving identical masses led to the idea of isomers through epimerization, though it was unclear which chiral centers were changing configuration and at what step of the synthesis (Figure 2.8). In an attempt to test racemization of the

microwave reaction, the Boc protected NDI monomer (**8**) was measured for its specific rotation, synthesized with and without base, as well as the Fmoc protected G-K-NDI monomer (**10**). The measured optical rotation values are close to zero. Purchased Boc-Lys-OH was used as a standard and also has an optical rotation close to zero, but matches the manufacturer's specific rotation. The measured values are too close to the instrument threshold for confident analysis. Evidence of racemization was not conclusively found in monomer **8** nor **10**.

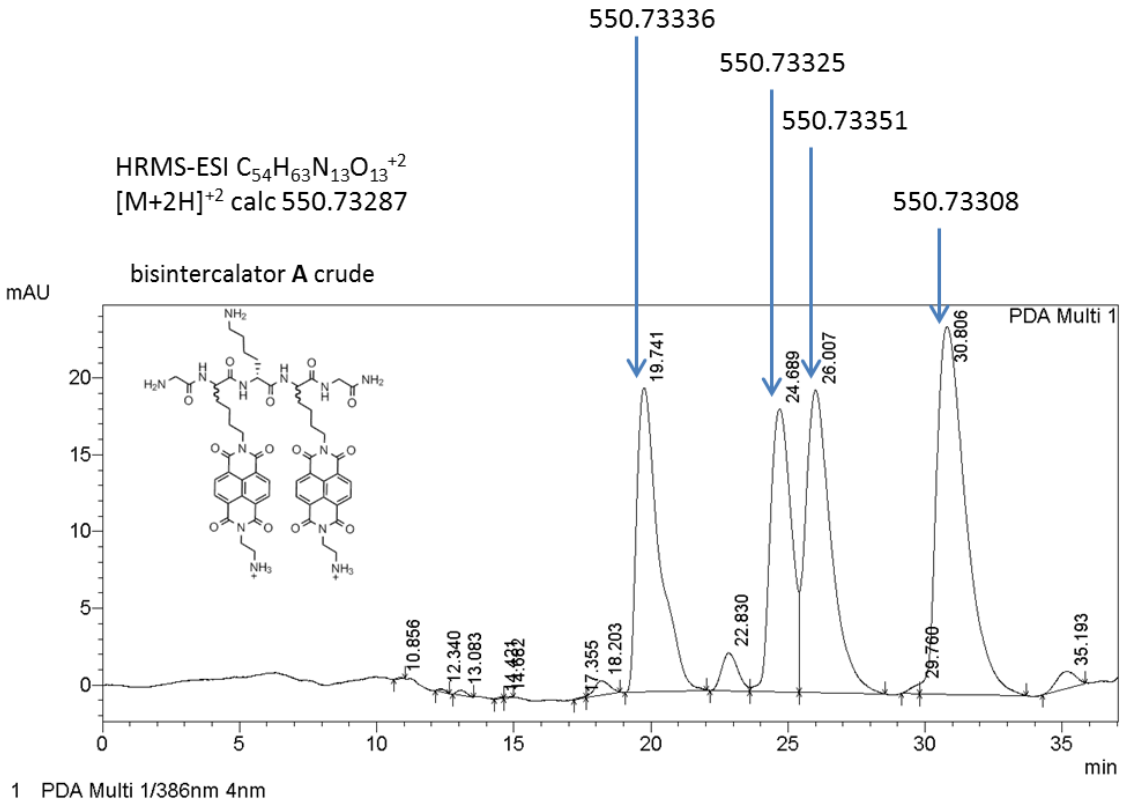
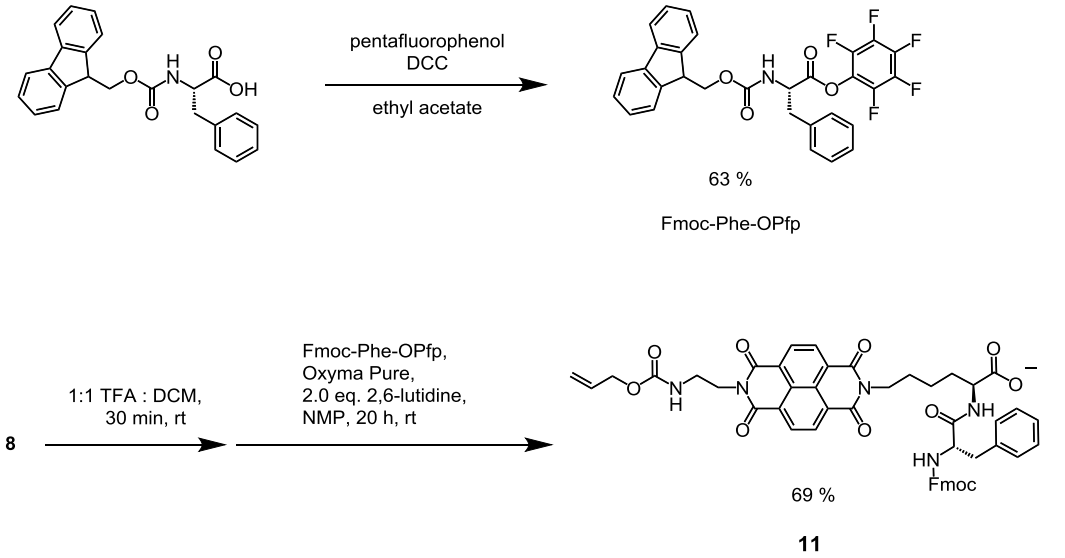


Figure 2.8: Crude solution of bisintercalator **A** run on Shimadzu LCsolution HPLC after synthesis using old SPPS protocol. Indicated peaks were isolated and observed HRMS-ESI values reported above respective peaks.

Sample	Concentration (g/mL)	α Observed optical rotation (measured)	[α] specific rotation (calculated)
Boc-K-OH Bottle 1	0.02 in ddH_2O	+ 00.04 ° + 00.05 ° + 00.04 °	+ 4 ° + 5 ° + 4 °
Boc-K-OH Bottle 2	0.02 in ddH_2O	+ 00.05 ° + 00.05 ° + 00.05 °	+ 5 ° + 5 ° + 5 °
Boc-K-OH Bottle 3	0.02 in ddH_2O	+ 00.04 ° + 00.05 ° + 00.04 °	+ 4 ° + 5 ° + 4 °
(**8**) Boc-K(NDI-Alloc)OH Batch 1 – used Et_3N	0.01 in DMF	– 00.01 ° – 00.01 ° – 00.01 °	– 2 ° – 2 ° – 2 °
(**8**) Boc-K(NDI-Alloc)OH Batch 2 – used Et_3N	0.01 in DMF	– 00.01 ° 00.00 ° – 00.02 ° – 00.02 ° – 00.02 °	– 2 ° 0 ° – 4 ° – 4 ° – 4 °
(**8**) Boc-K(NDI-Alloc)OH Batch 3 – used Et_3N	0.01 in DMF	– 00.01 ° – 00.01 ° – 00.01 ° – 00.01 °	– 2 ° – 2 ° – 2 ° – 2 °
(**8**) Boc-K(NDI-Alloc)OH **No** Et_3N used Measurements taken of series dilution	0.04 0.02 0.01 0.04 in DMF	+ 00.01 ° 00.00 ° – 00.01 ° + 00.01 °	+ 0.5 ° 0 ° – 2 ° + 0.5 °
(**10**) Fmoc-G-K(NDI-Alloc)OH Batch 1 – used base in multiple steps	0.01 in DMF	+ 00.01 ° + 00.01 ° + 00.01 °	+ 2 ° + 2 ° + 2 °
(**10**) Fmoc-G-K(NDI-Alloc)OH Batch 2 – used base in multiple steps	0.01 in DMF	+ 00.02 ° + 00.02 ° + 00.02 °	+ 4 ° + 4 ° + 4 °

Table 2.1: Optical rotation study where [α] is calculated by $[\alpha]^{T}_{\lambda} = \alpha / (l \cdot c)$ and Temperature (T) is 25 ° C, λ is sodium D line wavelength, and the pathlength (l) is 0.5 dm. Rudolph Autopol II Automatic Polarimeter.

The solid phase synthesis was then examined for proof of a scrambled chiral center. Another variation of the NDI monomer was made using phenylalanine (Scheme 2.3). The aromatic phenyl group provides significant structural difference allowing for faster and clearer monitoring of epimerization during SPPS by automated LCMS. In monitoring each SPPS step, LCMS spectra indicate epimerization taking place during coupling steps involving the Fmoc-F-Lys(NDI-Alloc)-OH monomer (**11**) (Figure 2.9). A clean 15 atom bisintercalator using phenylalanine was synthesized, to also monitor alloc deprotection, according to Scheme 2.2 by replacing the first monomer coupling with (**11**) and the second with (**9**), producing K-K(NDI)-F-K(NDI)-G (Figure 2.10). The epimerization hypothesis is supported by NMR of Fmoc-Lys-Lys(NDI-Alloc)-Gly (its synthesis utilizes monomer **9**) showing splitting between 52 ppm and 56 ppm in ^{13}C NMR (Figure 2.11). In the case of a single species, the chiral carbon of the NDI modified lysine would be expected to present as a single peak. Similarly, a single product peak would be expected on an HPLC trace of non-epimerized intercalator (Figure 2.12).



Scheme 2.3: Phenylalanine monomer synthesis.

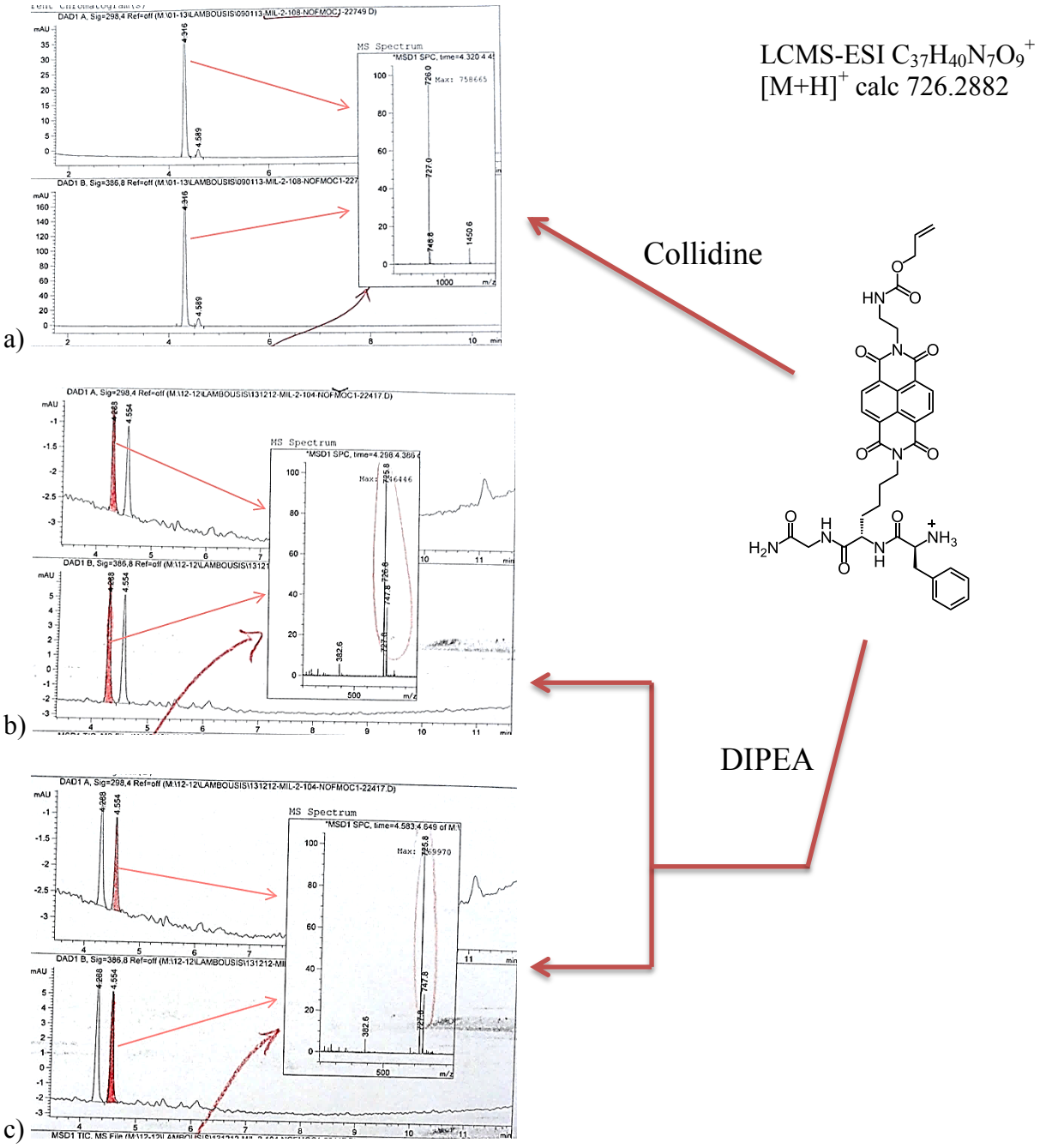


Figure 2.9: LCMS traces of Phe-Lys(NDI-Alloc)-Gly are shown with absorbance at 298.4 nm (top) and 386.8 nm (bottom) and m/z (right). Solid phase synthesis using the base collidine in coupling reactions produces predominantly (95 %) the S, S isomer (a). Use of diisopropylethylamine during coupling reactions yields a 1:1 mixture of isomers with one scrambled chiral center (b and c).

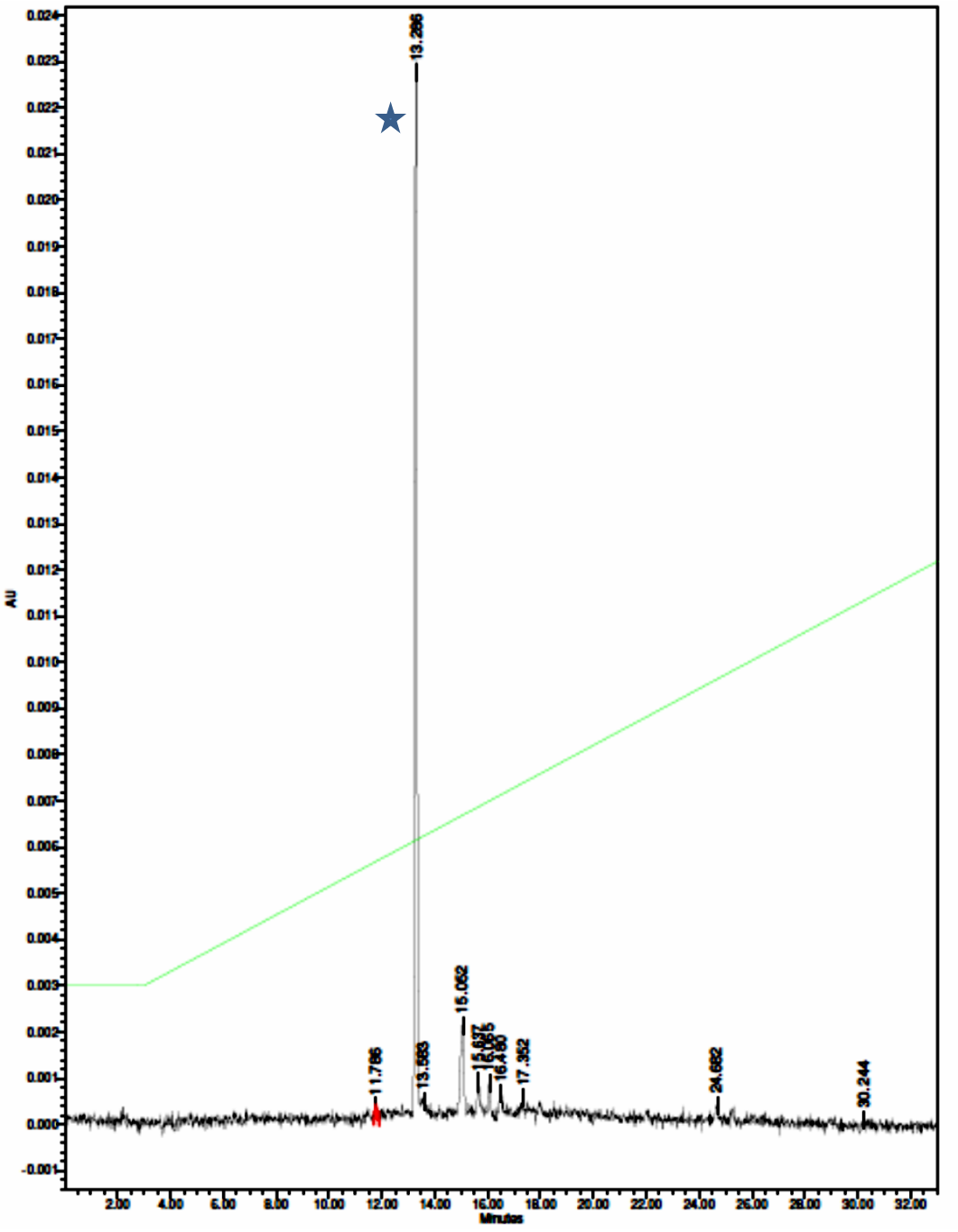


Figure 2.10: Analytical HPLC trace of crude K-K(NDI)-F-K(NDI)-G single peak eluting at 13.286 minutes, marked by star.

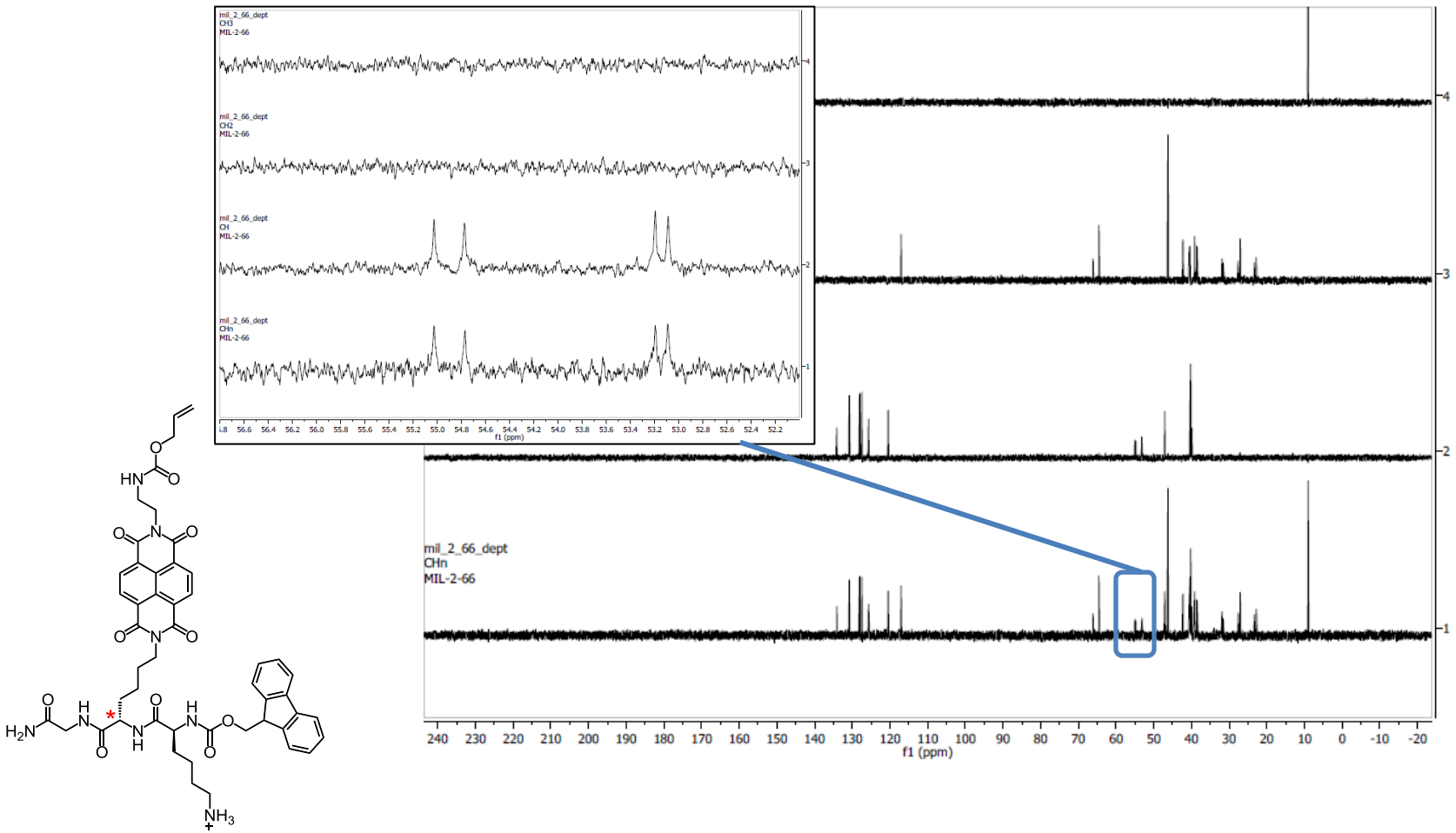


Figure 2.11: NMR DEPT (DMSO-d_6) of Fmoc-Lys-Lys(NDI-Alloc)-Gly with magnified insert of 56–52 ppm region showing peak splitting at the NDI modified lysine chiral center, indicating presence of diastereomers.

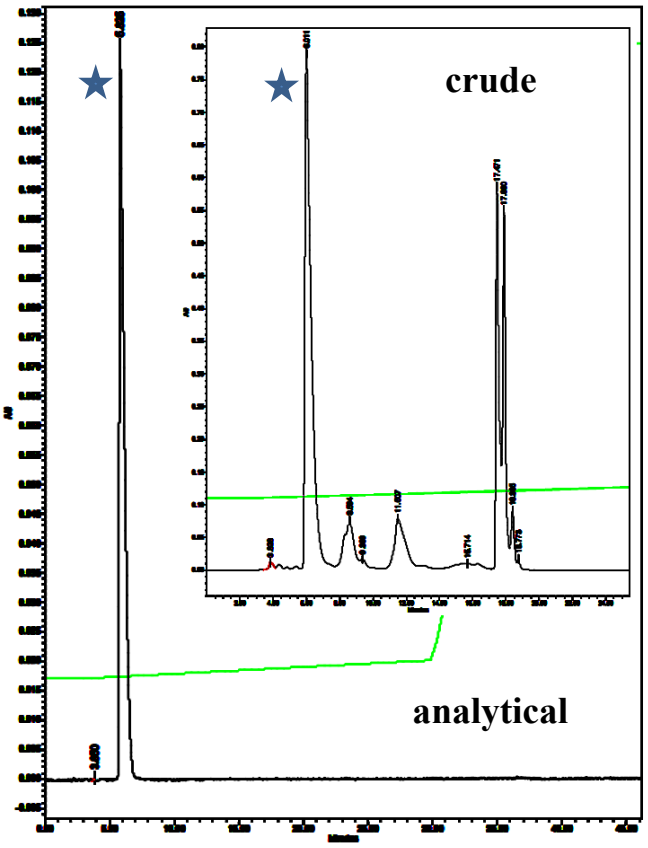


Figure 2.12: Crude HPLC trace of bisintercalator **A** at 6.011minutes within subsequent analytical trace, eluting at 5.825 minutes and marked by star. Synthesis of **A** followed Scheme 2.2.

2.3.2 DNase I Footprinting Using Sequences Hex A and Hex B

The preliminary DNase I footprinting experiments were all run with DNA sequences Hex A and Hex B generously provided by the Fox group (Hampshire and Fox, 2008). These two sequences contain every six base pair palindromic binding site. The reverse of each Hex sequence was also used to visualize possible intercalator binding sites. The four intercalators each bind to the same GC-rich sites on the DNA regardless of length or charge. Molecule **A**, +5 charged, did have greater affinity to the negatively

charged DNA than molecule **B**, +4 charged, as expected. Several possible binding sequences were plucked from these results for further study (Figure 2.18).

```
HexAfor
5′-
GGATCCCGGGATATCGATATATGGCGCCAAATTTAGCTATAGATCTAGAATTCCGGACCGCGGTT
TAAACGTTAACCGGTACCTAGGCCTGCAGCTGCGCATGCTAGCGCTTAAGTACTAGTGCACGTGG
CCATGGATCC-3′

HexArev
5′-GGATCCATGGCCACGTGCACTAGTACTTAAGCGCTAGCATGCGCAGCTGCAGGCCTAGGTAC
CGGTTAACGTTTAAACCGCGGTCCGGAATTCTAGATCTATAGCTAAATTTGGCGCCATATATCGA
TATCCCGGGATCC-3′

HexBfor
5′-GGATCCGGCCGATCGCGAGCTCGAGGGCCCTAATTAGCCGGCAATTGCAAGCTTATAAGCGC
GCTACGTATACGCGTACGCGCGTATATACATATGTACATGTCGACGTCATGATCAATATTCGAAT
TAATGCATGGATCC-3′

HexBrev
5′-GGATCCATGCATTAATTCGAATATTGATCATGACGTCGACATGTACATATGTATATACGCGC
GTACGCGTATACGTAGCGCGCTTATAAGCTTGCAATTGCCGGCTAATTAGGGCCCTCGAGCTCGC
GATCGGCCGGATCC-3′
```

Figure 2.13: All four Hex A and Hex B DNA footprinting sequences are written in the traditional 5′ to 3′ and the top strands designated as for (forward). The reverse sequences (rev) are the reverse complement of the forward sequences but incorporated within the pUC19 vector (Appendix) in the 5′ to 3′ direction as template strands.

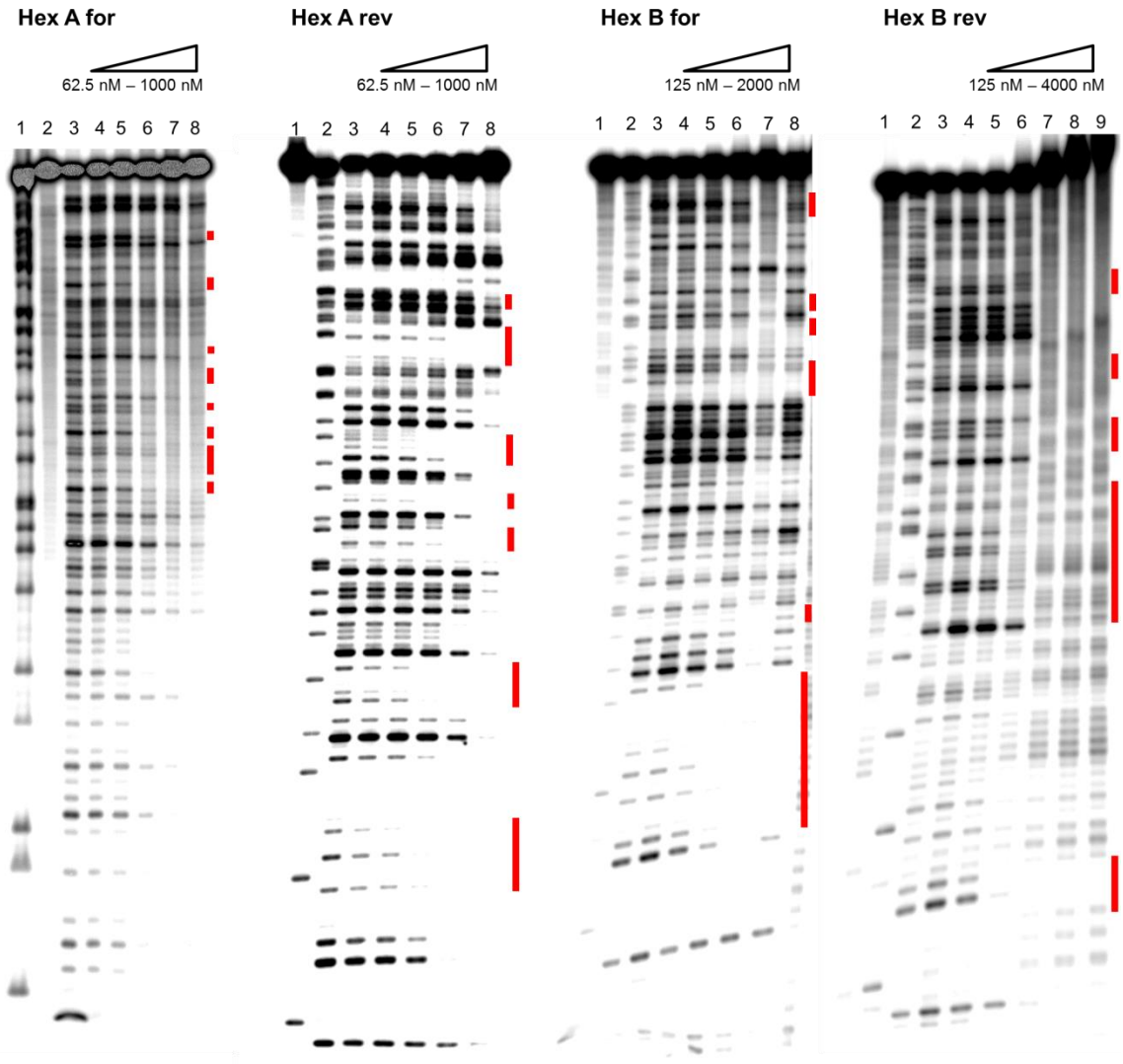


Figure 2.14: Footprints of intercalator **A** incubated with Hex sequences. For HexAfor, lane 1 has an adenine-specific cleavage reaction (Iverson and Dervan, 1987), lane 2 has undigested DNA, lane 3 has digested DNA with no intercalator, lanes 4-8 have 62.5, 125, 250, 500, and 1000 nM intercalator. HexArev is the same as HexAfor except lane 1 has undigested DNA and lane 2 has the adenine-specific reaction. For HexBfor lane 1 contains undigested DNA, lane 2 contains an adenine-specific cleavage reaction, lane 3 contains digested DNA with no intercalator, lanes 4-8 contain 125, 250, 500, 1000, and 2000 nM intercalator. HexBrev is the same as HexBfor with an additional lane 9 containing 4000 nM intercalator **A**. The binding sites are indicated in red.

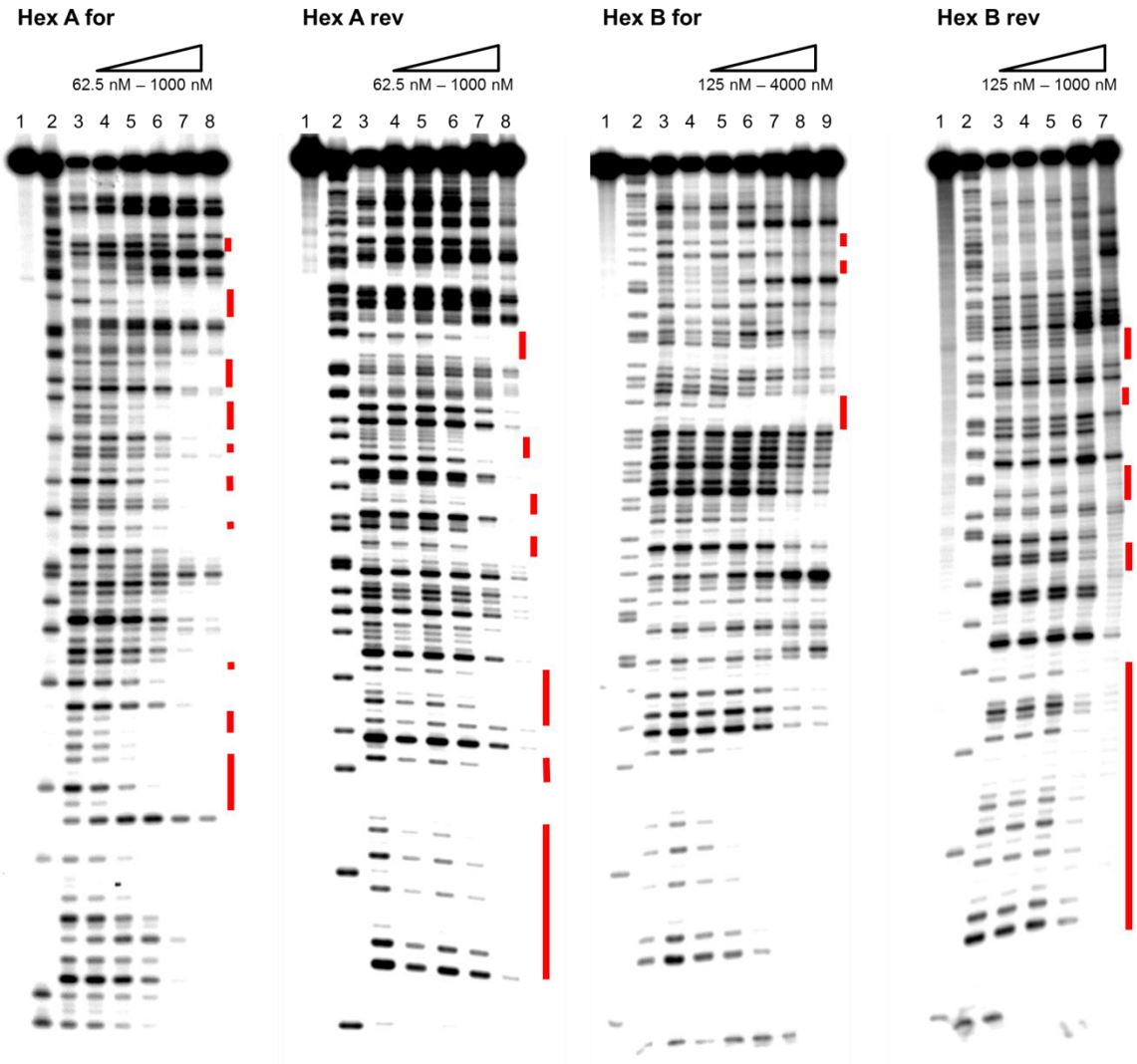


Figure 2.15: Footprints of intercalator **B** incubated with Hex sequences. For HexAfor and HexArev, lane 1 has undigested DNA, lane 2 has an adenine-specific cleavage reaction (Iverson and Dervan, 1987), lane 3 has digested DNA with no intercalator, lanes 4-8 have 62.5, 125, 250, 500, and 1000 nM intercalator. For HexBfor lane 1 contains undigested DNA, lane 2 contains an adenine-specific cleavage reaction, lane 3 contains digested DNA with no intercalator, lanes 4-9 contain 125, 250, 500, 1000, 2000, and 4000 nM intercalator HexBrev is the same as HexBfor but stops after 1000 nM intercalator **B**. The binding sites are indicated in red.

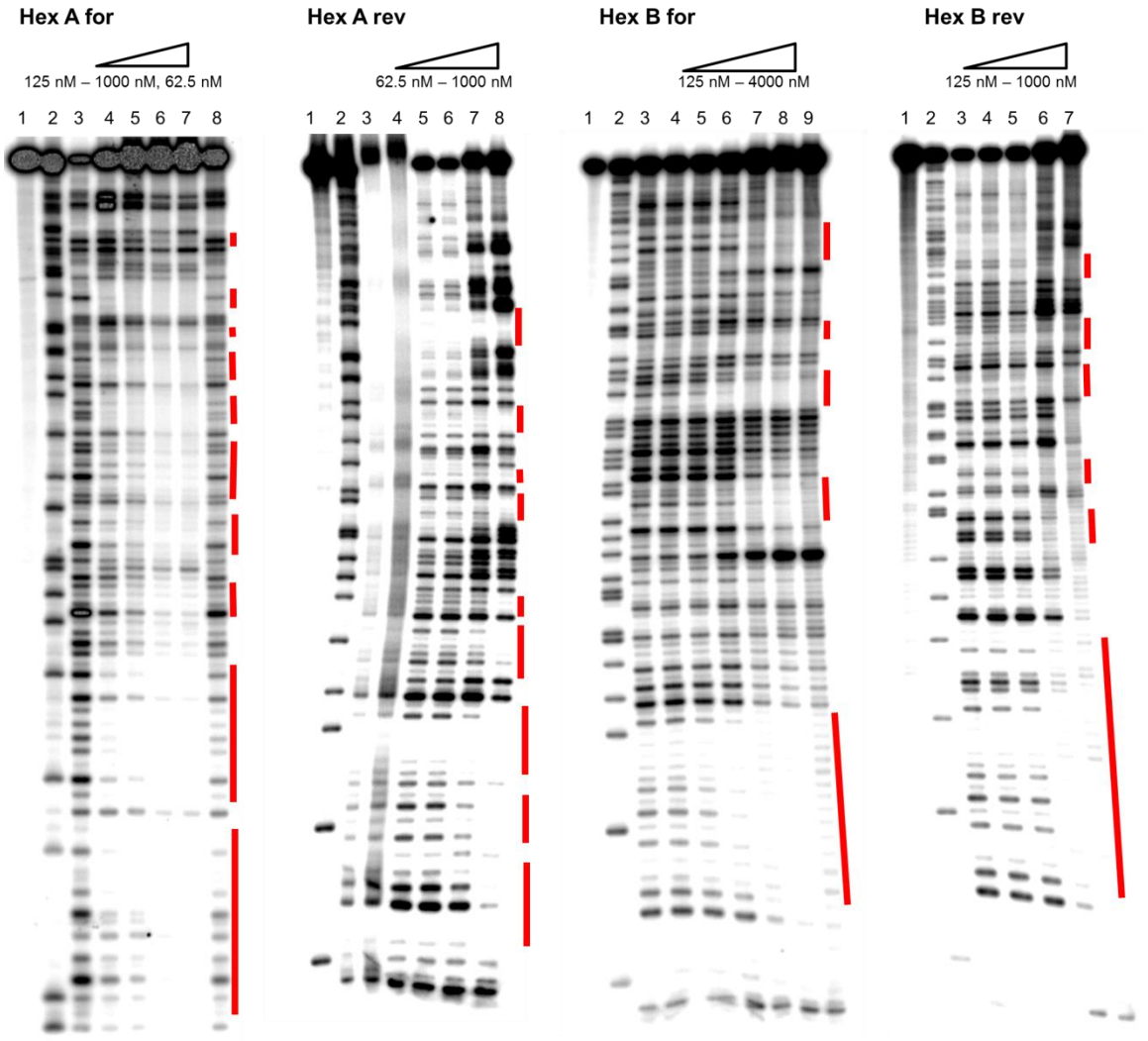


Figure 2.16: Footprints of bisintercalators **C** incubated with Hex sequences. For HexAfor and HexArev, lane 1 has undigested DNA, lane 2 has an adenine-specific cleavage reaction (Iverson and Dervan, 1987), lane 3 has digested DNA with no intercalator, lanes 4-8 have 62.5, 125, 250, 500, and 1000 nM intercalator. For HexBfor lane 1 contains undigested DNA, lane 2 contains an adenine-specific cleavage reaction, lane 3 contains digested DNA with no intercalator, lanes 4-9 contain 125, 250, 500, 1000, 2000, and 4000 nM intercalator HexBrev is the same as HexBfor but stops after 1000 nM intercalator **C**. The binding sites are indicated in red.

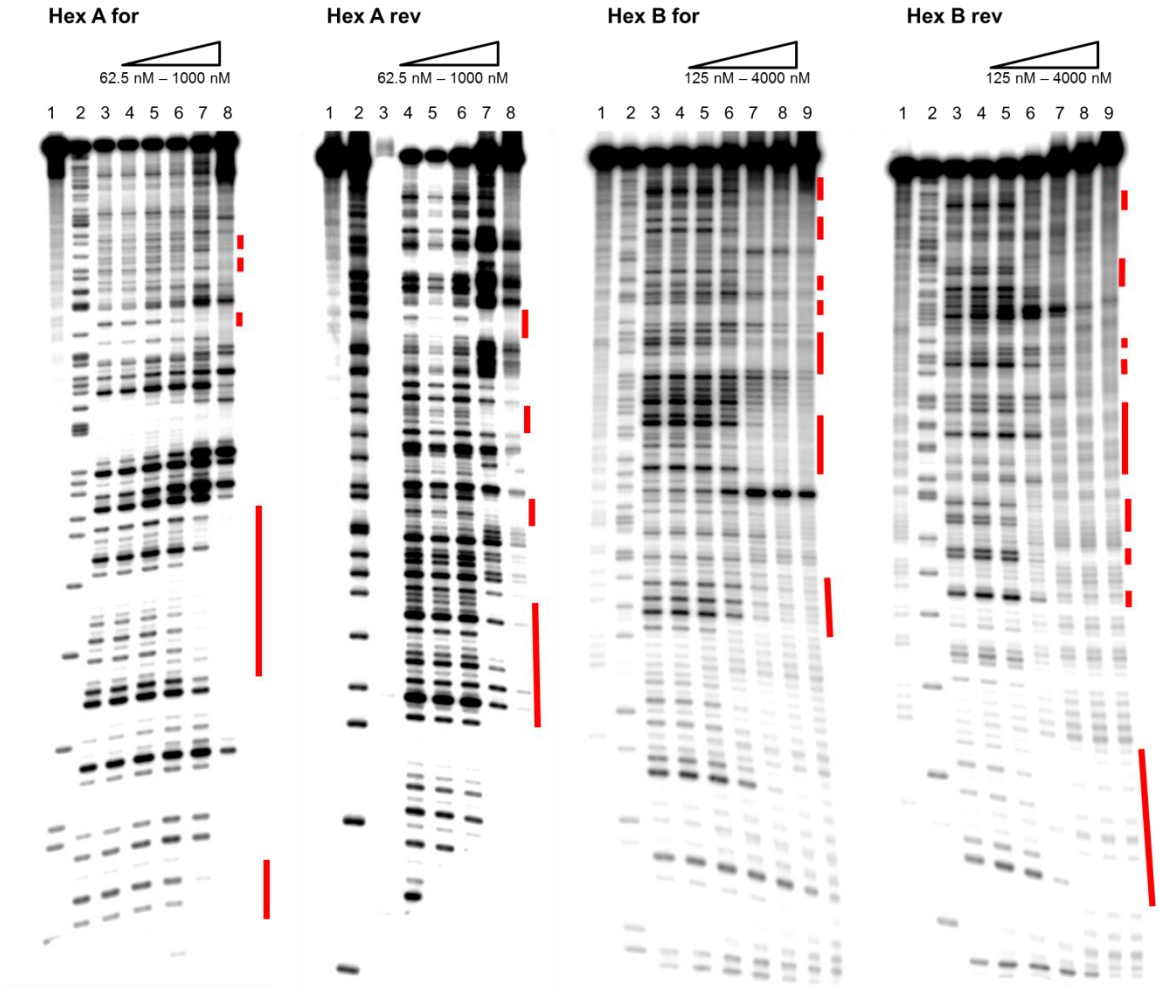


Figure 2.17: Footprints of bisintercalators **D** incubated with Hex sequences. For HexAfor and HexArev, lane 1 has undigested DNA, lane 2 has an adenine-specific cleavage reaction (Iverson and Dervan, 1987), lane 3 has digested DNA with no intercalator, lanes 4-8 have 62.5, 125, 250, 500, and 1000 nM intercalator. For HexBfor and HexBrev lane 1 contains undigested DNA, lane 2 contains an adenine-specific cleavage reaction, lane 3 contains digested DNA with no intercalator, lanes 4-9 contain 125, 250, 500, 1000, 2000, and 4000 nM intercalator **D**. The binding sites are indicated in red.

2.3.3 Binding Site Sequence Design

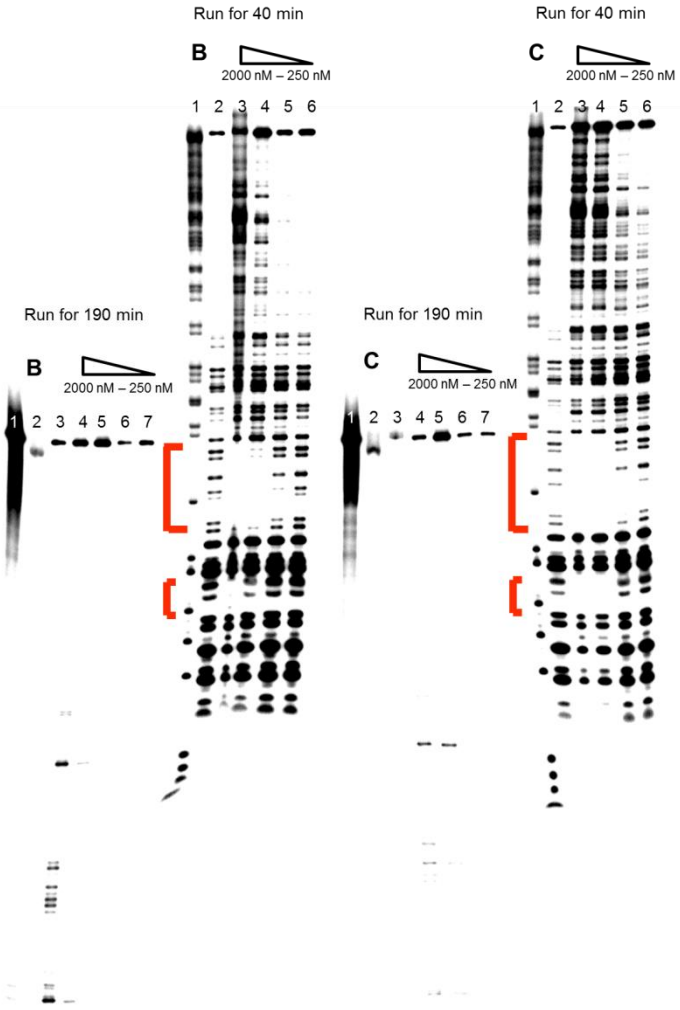
The new binding site candidates chosen from sequences Hex A and B were built into a new DNA sequence, "Pendant Intercalation Sites" (PIS), separated by ten bases to allow for one full helical twist of the DNA between bound molecules. The PIS sequence was also designed to include SfiI restriction enzyme sites for possible insertion into a vector. Additionally, the sequence extends beyond the SfiI sites to allow for vector insertion by overlap extension polymerase chain reaction (OLE PCR) and Gibson cloning (Bryksin, 2010; Gibson, 2009). Incorporation into a vector and subsequent transformation provides a continuous supply of the DNA sequence and a shorter route to possible future experiments with the pendant intercalator system *in vivo*. The 381 base pair sequence was ordered as a double-stranded fragment with blunt ends. OLE PCR was not successful. Gibson cloning of PIS with the pMoPac16 vector (Appendix Figures A3 and A4) was successful, but subsequent isolation of the sequence from the vector for footprinting was not high yielding due to a secondary amplification product. Instead, PIS stock for footprinting experiments was created by an initial extra round of PCR on the small amount of ordered gene fragment, generating a 316 base pair sequence.

5'*-GGCTTTACACTTTATGCTTCCGGCTCGTATGTTGGCAACAAAGGTGATGAGGAGC
TACATCGGCCCAGCCGGCCTCGTAGAGTCGAAGTGCACATACGTCTTCCTGACGATGCG
TTGCGCAGCTGATACTGGCGTAGCAGCTGTTGGCCACTAGTAGCAGCTAGGTTCTACTA
TTAAGTAGCAGCTTCCGGGTACTAGTAGAGGGAAGTTGGTAACTGCCAGCTGGTACGCC
TATAGTAGCAGCCGCGTATCTCACTAGTGCGCGCAGCAGGTGCTCCGCGGACTAGAAGC
ACTTCATGAGGCCTCGGGGGCCCATCATCCTCCACCATGTCAGCATCAGTAGTCATACG
GACGCTGAGCAAAGCAGACTACGAGAAACAC-3'

Figure 2.18: PIS gBlock sequence containing possible pendant binding sites (underlined) derived from the Hex footprinting experiments. The red portion is eliminated after initial PCR and is not present in the footprinting experiments.

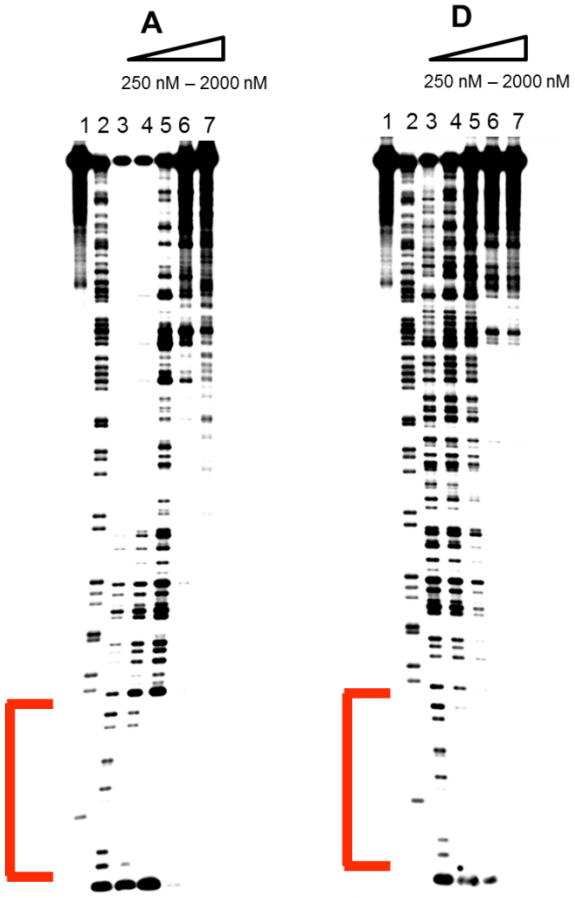
2.3.4 DNase I Footprinting Using PIS

Evaluation of intercalators **A-D** binding to PIS was accomplished by concentration dependent DNase I footprinting. All four bisintercalator derivatives show a preference for GC-rich sequences, in particular 5' – GGCC – 3'. The SfiI site, being a long GC-rich sequence, fortuitously provides a clear binding site. Non-linear attenuation of bands results in abrupt footprints at 1000 nM of each intercalator.



5′*- … ATGAGGAGCTACATCGGCCCAGCCGGCCTCGTAGA … - 3′

Figure 2.19: DNase I footprints of PIS with bisintercalators **B** and **C**, with dominant binding sites indicated in red. A single 10 % denaturing PAGE gel was continuously run with two loading times, and the second loading taking place 2.5 hours after the fist. For the 190 min. runs: lane 1 contains undigested PIS DNA, lane 2 contains an adenine-specific cleavage reaction (Iverson and Dervan, 1987), lane 3 contains digested PIS with no intercalator, lanes 4-7 contain 2000, 1000, 500, and 250 nM intercalator. The same batch of samples was used in the 40 min. runs as in the 190 min. For the 40 min. runs: lane 1 contains an adenine-specific cleavage reaction, lane 2 contains digested PIS with no intercalator, lanes 3-6 contain decreasing concentrations of intercalator.



5′*- … ATCGGCCCAGCCGGCCTCGTAGA … - 3′

Figure 2.20: DNase I footprints of PIS with bisintercalators **A** and **D**, with dominant binding site indicated in red on an 8 % denaturing PAGE gel over 75 min. Lane 1 contains undigested PIS DNA, lane 2 contains an adenine-specific cleavage reaction (Iverson and Dervan, 1987), lane 3 contains digested PIS with no intercalator, lanes 4-7 contain 250, 500, 1000, and 2000 nM intercalator.

2.4 DISCUSSION

Ensuring diastereomerically pure intercalators are being analyzed is of paramount importance as future ribosomally produced NDI pendant bisintercalators would consist of all L amino acids. Originally, the longer pendant derivatives utilized only the Fmoc-Gly-Lys(NDI-Alloc)-OH monomer (**10**) in SPPS coupling steps with PyBop and the base diisopropylethylamine. Once the new Fmoc-Lys(Boc)-Lys(NDI-Alloc)-OH monomer (**9**) was introduced, purification by HPLC and characterization by mass spectrometry indicated four different compounds of identical mass. While a similar result may have occurred with the original glycine monomer, stereoisomers likely were too similar for routine detection. The 2^n rule gives the maximum number of stereoisomers possible, where n= number of chiral centers. Four stereoisomers appeared to have been synthesized from a structure containing three chiral centers. If only two of those chiral centers racemized, at the modified NDI-lysines, the rule predicts $2^2 = 4$ isomers. This realization of having synthesized diastereomers called into question the entire synthesis. Any reaction involving base could racemize an amino acid. It would seem especially plausible for this to happen at elevated temperatures particularly during the initial microwave reaction. Thus, optical rotation measurements were taken of (**8**) Boc-(Lys-NDI-Alloc)-OH batches made with varying amounts of base. The measurements were nearly zero and close to the detection limit of the polarimeter, making it difficult to draw any definite conclusions. NMR spectra of the Fmoc-Lys(Boc)-Lys(NDI-Alloc)-OH monomer (**9**) did not confirm nor refute the presence of diastereomers.

Since a diastereomer was not readily detectable, the epimerization could be taking place at another point in the synthesis. Base is also present in SPPS coupling steps. Diisopropylethylamine was initially used in the solid phase coupling steps. While PyBop is used as an additive to both activate the carboxy end of the peptide and reduce

racemization, it was not enough to hinder abstraction of the proton from the chiral center. NDI is known to stabilize anions and may have done so in this case (Guha and Saha, 2010). Folding of the flexible lysine side chain may position NDI above the negatively charged carbon, stabilizing it long enough for the proton to be replaced from either side. Switching to collidine (2,4,6-trimethylpyridine) reduced the epimerization, likely a result of collidine's size and steric bulkiness (Han et al., 1997; Carpino et al., 1996; Subiros-Funosas et al., 2010).

DNase I footprinting of the four pendant bisintercalators with the Hex sequences showed binding to the same locations across all intercalators, indicating both the 18 and 21 atom derivatives were too long. A commonly bound sequence 5' – GGCC – 3' among the GC-rich sequences implicates a four base pair binding site. As the shortest 15 atom intercalators are too short to bind comfortably to a 6 base pair binding site, and nearest neighbor exclusion principle eliminates binding in adjacent spaces, a four base pair site is most likely.

All four intercalators were carried over to the next round of experiments to verify the prior results and narrow down possible specific binding sites. As before in footprinting with the Hex sequences, DNase I footprinting with PIS showed molecules **A, B, C, and D** all bound the same sites. Inclusion of the SfiI sites, in particular 5'* – GGCCCAGCCGGCC – 3' that appears at the bottom of the gel, proved to be serendipitously useful in determining both preferences for GC-rich sites as well as possible cooperative binding demonstrated by the abrupt, non-linear footprints. As NDI monomers are known to prefer GC DNA sites, the linkers in all of our pendant derivatives are unlikely to be directing binding location.

2.5 CONCLUSION

As designed, the pendant bisintercalators prefer a four base pair, GC-rich binding site, and the shortest derivatives seem to be of optimal length. Lack of hydrogen-bonding sites on the lysine chain near the NDI does not provide the necessary interactions likely needed to impart greater sequence specificity by the linker. Similar binding seen by all three lengths of derivatives indicates that the linkers are not driving the location of binding, rather the NDIs, which are known to bind GC-rich sites, are providing the little specificity seen by DNase I footprinting. In order to proceed to *in vivo* studies, the pendant design must show the ability to bind specifically based on linker authority to provide modular tunability. It may be possible to improve linker specificity by changing the amino acid to which the NDIs are attached. Choosing an amino acid with fewer methylene units replaced by hydrogen-bonding capable functionalities might make a more sequence specific pendant intercalator design. The most useful information in selecting the optimal amino acid may come from a closer NMR structural study of the DNA bound G3K and B3K bisintercalators and tetraintercalator to identify specific hydrogen bonds close to the intercalation sites.

2.6 MATERIALS AND METHODS

2.6.1 General Methods

Solvents and starting materials were used without further purification unless otherwise noted. Organic chemicals and solvents were obtained from Sigma-Aldrich or Fisher, resins and amino acids were obtained from Novabiochem, oligonucleotides were obtained from Integrated DNA Technologies (IDT), and enzymes were obtained from New England Biolabs or Fermentas. Analytical and semi-preparative HPLC were performed on a Waters Delta 600 system with a photodiode array detector using Vydac C-18 columns, unless

otherwise noted. An Agilent 8453 Spectrophotometer was used to perform UV-Vis analyses. Most NMR spectra were obtained with a Varian DirectDrive 400 MHz and 2D spectra were obtained with a Varian DirectDrive 600 MHz. High resolution mass spectra were recorded on a 9.4T IonSpec HiResESI FT-ICR.

2.6.2 Synthesis

Allyl-2-aminoethyl carbamate (Alloc Ethylenediamine; 7): Ethylenediamine (50 mL, 0.75 mol) was dissolved in 100 mL DCM in a 500 mL oven-dried and argon purged round bottom flask with stir bar and cooled to 0 °C on ice. In a separate dried 250 mL round bottom flask were combined allyl chloroformate (20 mL, 0.19 mol) and 175 mL DCM and sonicated under argon for 10 minutes. The chloroformate solution was then poured into a dry addition funnel and added dropwise to the ethylenediamine solution while stirring over 17 hours and solution allowed to warn to room temperature. The slurry was then concentrated *in vacuo* at room temperature to produce a thick white opaque liquid with some white solid. To this was added 200 mL of chilled water and vigorously shaken to produce white solid fluff that was removed by vacuum filtration. The filtrate was extracted using DCM (1 x 600 mL), (2 x 500 mL), (3 x 250 mL), and (1 x 200 mL). Then concentrated *in vacuo* to remove all solvent and produce a thin clear oil (19.73 g, 72 %). ^{1}H NMR (400 MHz, $CDCl_3$) δ 5.79 (br m, 1H), 5.63 (br s, 1H), 5.18 (d, 1H), 5.13 (d, 1H), 4.47 (d, 2H), 3.15 (q, 2H), 2.72 (t, 2H), 1.38 (br s, 2H) ppm; ^{13}C NMR (400 MHz, $CDCl_3$) δ 156.56, 132.91, 117.53, 65.44, 43.73, 41.67; HRMS-ESI $C_6H_{12}N_2O_2Na^+$ $[MNa]^+$ calcd 167.0899, found 167.0791.

Boc-(Lys-NDI-Alloc)-OH (8): 1,4,5,8-Naphthalenetetracarboxylic dianhydride (5 g, 18.6 mmol) and Alloc ethylenediamine (2.74 g, 19.0 mmol) were suspended in 200 mL of dry DMF, argon purged, and stirred or sonicated until homogeneous. The mixture

was then heated in a microwave reactor (CEM Microwave Accelerated Reaction System, model MARS®) for 5 minutes at 75 °C then for 10 minutes at 140 °C (P = 600 S). The solution was then cooled to 65 °C, at which point Boc-lysine-OH (4.6g, 18.7 mmol) and triethylamine (2.65 mL, 19.0 mmol) were added. The solution was then heated for 5 minutes at 140 °C. It was then allowed to cool to 50 °C before removing the solvent *in vacuo*. The product mixture was the purified by column chromatography using 5 % TEA / 5 % MeOH in DCM. Solvent was removed *in vacuo* from the collected product fractions. The dark tan solid was then dissolved in a minimal amount of 10 % MeOH in DCM (12.5 mL) and acidified with AcOH (2 mL) to pH 4 – 5, resulting in partial precipitation. Further product was precipitated upon slow addition of hexanes (30 mL). The solid was then triturated using water and again hexanes prior to drying *in vacuo*. (Yield 40%). ^{13}C NMR (400 MHz, DMSO-d_6) δ= 175.0, 163.9, 161.7, 156.0, 134.1, 131.0, 125.2, 78.4, 69.2, 63.0, 54.1, 45.7, 36.2, 31.2, 30.0, 28.6, 23.0; ^{1}H NMR (400 MHz, DMSO-d_6) δ= 8.58 (s, 4H), 7.29 (t, 1H), 6.68 (d, 1H), 5.72-5.81 (m, 1H), 5.16 (dq, 1H), 5.06 (dq, 1H), 4.34 (d, 2H), 4.13 (t, 1H), 4.00 (t, 2H), 3.73-3.79 (m, 2H), 3.31-3.6 (m, 2H), 1.53-1.71 (m, 4H), 1.32 (s with broad shoulder, 9H); HRMS-ESI calc'd for $C_{31}H_{34}N_4O_{10}$ Na^+ $[MNa]^+$ calcd 645.2275, found 645.2165.

Boc deprotection: Boc-(Lys-NDI-Alloc)-OH (1.00 g, 1.61 mmol) was dissolved in 4.64 mL of DCM and purged with argon for several minutes while stirring. An equivalent volume, 4.64 mL of TFA was added slowly to the rapidly stirring mixture. The reaction was left to stir for 30 minutes at room temperature. Nitrogen gas was used to blow off about half of the solvent before adding 8 – 10 mL of MeOH to azeotropically remove TFA. Once the yellowish tan clay-like solid began to crack, it was placed under vacuum for several hours to ensure complete removal of acid and solvent. Material was immediately used in solution phase coupling without further purification.

Fmoc-Phe-OPfp: (Green and Berman, 1990; Kisfaludy and Schon 1983; Ikeda, Nakamura and Saito, 2002). In an oven-dried and argon purged round bottom flask with stir bar were added *N*-α-Fmoc-L-phenylalanine (3.022 g, 7.800 mmol) and pentafluorophenol (1.428 g, 7.758 mmol) and were dissolved in 23 mL of dry ethyl acetate with 2 mL of dry DMF. The stirring reaction solution was cooled in an ice bath and purged under argon before addition of N,N'-Dicyclohexylcarbodiimide (1.610 g, 7.803 mmol) and kept stirring cold for two hours then at room temperature overnight. The reaction mixture was filtered to remove the white precipitated dicyclohexylurea and the liquid concentrated *in vacuo*. The resulting solid was triturated with hexanes and subsequently recrystallized using ethyl acetate to give a white solid (Yield 63%). ^{13}C NMR (400 MHz, DMSO-d_6) δ= 173.8, 169.0, 156.4, 144.2, 144.1, 141.2, 137.0, 129.7, 129.5, 128.8, 128.6, 128.1, 127.5, 127.2, 125.5, 120.5, 66.2, 55.7, 47.0, 36.3; ^{1}H NMR (400 MHz, DMSO-d_6) δ= 8.27 (d, 2H), 7.87 (d, 2H), 7.64 (q, 2H), 7.39 (d, 2H), 7.34 (m, 2H), 7.31 (dd, 2H), 7.24 (dd, 1H), 4.71 (m, 1H), 4.34 (m, 2H), 4.20 (t, 1H), 3.23-3.13 (m, 2H); HRMS-ESI calc'd for $C_{30}H_{20}F_5NO_4$ Na^+ $[MNa]^+$ calcd 576.12050, found 576.12060.

General Soln Phase Coupling: Fmoc-Lys(Boc)-Lys(NDI-Alloc)-OH (9); Fmoc-Gly-Lys(NDI-Alloc)-OH (10); Fmoc-Phe-Lys(NDI-Alloc)-OH (11): To the monomer NDI boc-deprotected solid was added 13.4 mL NMP and argon purged while stirring for several minutes until entirely dissolved. To the solution were added ethyl 2-cyano-2-(hydroxyimino)acetate (0.228 g, 1.61 mmol), Fmoc-*amino acid*-OPfp (**9**: *Lys(Boc)*; **10**: *Gly*; **11**: *Phe*) (**9,10**: 1.61 mmol; **11**: 0.80 mmol (all amounts for **11** were half of those for **9,10**)), and slowly added 2,6-lutidine (0.372 mL, 3.21 mmol) and left to stir at room temperature under argon for 20 hours. Once the reaction was complete, the solution was added to 200 mL of DCM and partitioned with 200 mL of 0.2 M citric acid

buffer pH 4.5. The product was extracted from the aqueous buffer using another 200 mL portion of DCM. The two organic fractions were combined and washed using citric acid buffer (1x200 mL), water (1 x 200 mL), and brine (5 x 200 mL). The organic layer was dried using sodium sulfate and filtered before concentrating *in vacuo*. Having left a small amount of solvent, 150 mL of DCM was added to the concentration and transferred to a large beaker. This was vigorously stirred to completely dissolve any solid. While stirring, 900 mL of Hexanes was slowly added to elicit precipitation of product which was subsequently filtered. Lastly, the yellow solid was placed under vacuum. (**9**: 1.3183 g, 84 %; **10**: 1.1342g, 88 %; **11**: 0.4942g, 69 %) **9:** ^{13}C NMR (400 MHz, DMSO-d_6) δ = 174.0, 172.6, 163.0, 162.8, 156.6, 156.4, 156.0, 144.3 144.1, 141.0, 134.1, 130.7, 128.0, 127.5, 126.5, 126.42, 126.36, 126.2, 125.8, 125.7, 120.4, 117.1, 77.8, 66.1, 64.6, 54.9, 52.3, 48.9, 47.1, 38.5, 32.1, 31.2, 30.6, 29.7, 29.5, 28.7, 27.5, 23.4, 23.3, 17.7; ^{1}H NMR (400 MHz, DMSO-d_6) δ= 8.60 (s, 4H), 8.08 (d, 1H), 7.82 (d, 2H), 7.70 (t, 2H), 7.45 (d, 1H), 7.37 (t, 2H), 7.29 (t, 3H), 6.77 (t, 1H), 5.79 (m, 1H), 5.18 (dq, 1 H), 5.07 (dq, 1H), 4.37 (d, 2H), 4.27-4.14 (m, 6H), 4.01 (t, 3H), 3.34 (s, 2H), 2.89 (m br, 2H), 1.83-1.74 (m br, 1H), 1.72-1.57 (m br, 4H), 1.55-1.49 (m br, 1H), 1.47-1.32 (m br, 6H), 1.36 (s, 9H); **10:** ^{13}C NMR (400 MHz, DMSO-d_6) δ = 174.0, 169.5, 162.9, 156.9, 156.6, 144.23, 144.20, 141.0, 134.1, 130.7, 128.0, 127.5, 126.3, 125.7, 120.5, 117.1, 66.2, 64.6, 52.4, 49.0, 47.0, 43.6, 38.5, 31.4, 30.6, 29.5, 28.1, 27.5, 23.4, 17.7; ^{1}H NMR (400 MHz, DMSO-d_6) δ= 8.61 (s, 4H), 8.10 (d, 1H), 7.84 (d, 2H), 7.69 (d, 2H), 7.52 (t, 1H), 7.39 (t, 2H), 7.31 (t, 3H), 5.9 (m, 1H), 5.18 (dq, 1 H), 5.07 (dq, 1H), 4.37 (d, 2H), 4.27-4.11 (m, 6H), 4.02 (t, 2H), 3.66 (d, 2H), 3.34 (s, 4H), 1.83-1.73 (m br, 1H), 1.72-1.60 (m br, 3H), 1.46-1.35 (m br, 2H); **11:** ^{13}C NMR (600 MHz, DMSO-d_6) δ = 174.2, 173.9, 172.1, 163.0, 162.9, 156.6, 156.1, 144.1, 144.0, 141.0, 138.5, 134.1, 130.7, 130.6, 129.6, 128.4, 127.9, 127.4, 126.60, 126.57, 126.43, 126.40, 126.3, 126.2, 125.7, 125.6, 120.4, 117.0, 66.0, 64.6, 56.5,

52.2, 48.9, 46.9, 40.5, 38.5, 37.8, 31.2, 30.5, 29.4, 27.5, 23.4, 17.6; ^{1}H NMR (600 MHz, DMSO-d_6) δ= 12.63 (s br, 1H), 8.54 (m, 4H), 8.20 (d, 1H), 7.77 (d, 2H), 7.60-7.53 (m, 2H), 7.36-7.32 (t, 2H), 7.31-7.18 (m, 7H), 7.18-7.14 (m, 1H), 5.79 (m, 1H), 5.19 (dq, 1H), 5.08 (dq, 1H), 4.37 (d, 2H), 4.29-4.22 (m, 2H), 3.35 (d br, 4H), 2.91 (dd, 1H), 2.67 (m, 1H), 1.86-1.81 (m, 1H), 1.75-164 (m, 3H), 1.48-1.41 (m, 2H); HRMS-ESI calc'd for (**9**) $C_{52}H_{56}N_6O_{13}Na^+$ $[MNa]^+$ 995.3905, found 995.3802; calc'd for (**10**) $C_{43}H_{39}N_5O_{11}Na^+$ $[MNa]^+$ 824.25, found 824.2501; calc'd for (**11**) $C_{50}H_{45}N_5O_{11}Na^+$ $[MNa]^+$ 914.30078, found 914.29986.

Solid Phase Peptide Synthesis: All six pendant-NDI bisintercalators were synthesized using standard Fmoc solid phase synthesis methods beginning with Rink Amide MBHA resin (loading 0.64 mmol/g), commercially available Fmoc-Gly-OH and Fmoc-Lys-OH amino acids, and the Fmoc-Lys-Lys(NDI-Alloc)-OH Fmoc-Gly-Lys(NDI-Alloc)-OH monomers **9** and **10**. The resin was added to 5-mL fritted syringes (Torviq) and swelled in DMA for 45 minutes. Coupling steps proceded by rotary shaking the resin in a DMF solution containing 3 eq protected amino acid, 3 eq PyBOP, and 6 eq Collidine. Deprotection steps were carried out by shaking the resin in a DMF solution containing 20% piperidine/DMF. No capping steps were included. Between steps, the resin was rinsed with (3 x 5) mL DMF, (3 x 5) mL DCM, and (3 x 5) mL DMF. Prior to removing the Alloc protecting groups from the NDI sidechains, the resin was vacuum dried and thrice purged with argon, then re-swollen in THF for an hour. The Alloc groups were then removed by shaking the resin for 15 hours in a solution composed of 6 eq. thiosalicylic acid and 6 eq. tetrakis(triphenylphosphine)palladium(0) in a 50% THF / 50% DMSO solvent system. The peptides were then rinsed with (3 x 5) mL 0.5% (v/v) collidine in DMF, (3 x 5) ml 0.5% (w/v) sodium dimethyldithiocarbamate in DMF, and 0.5% (v/v) collidine in DMF and then rinsed with (3 x 5) mL DMF, (3 x 5) mL DCM,

and (3 x 5) mL DMF. This was followed by a final Fmoc deprotection using 20 % piperidine/DMF and washed with (3 x 5) mL DMF, (3 x 5) mL DCM, (3 x 5) mL DMF and (8 x 5) mL DCM. The resin was immediately subjected to cleavage with (3 x 5) mL 95% TFA / 5% H_2O cleavage solution with each round of cleavage shaken for 1 hour. The resulting solutions were concentrated by blowing nitrogen gas over the top of scintillation vials containing the solutions. The resulting sticky red-orange residues were dissolved in 1.5 mL H_2O and 0.1 mL MeOH and transferred to 2 mL centrifuge tubes with a pinhole at the top for lyophilization. Once lyophilized the burnt orange colored airy solids were purified by HPLC.

HPLC purification: Solutions of **A-D** were clarified using a 0.2-μm RC syringe filter and subjected to reverse-phase HPLC purification in a water/acetonitrile solvent system with 0.1% TFA. The first three minutes are isocratic 14 % organic, followed by a gradient over the next 27 minutes increasing to 16.5 % organic, and once again held at 16.5 % organic for another 3 minutes. Each of the four bisintercalators **A-D** eluted between 6 minutes and 22 minutes.

Pendant Intercalator A. HRMS-ESI calc'd for $C_{54}H_{61}N_{13}O_{13}$ $[M+H]^+$ calcd 1100.4585, found 1100.4611.

Pendant Intercalator B. HRMS-ESI calc'd for $C_{50}H_{52}N_{12}O_{13}$ Na^+ $[MNa]^+$ calcd 1051.3669, found 1051.3668.

Pendant Intercalator C. HRMS-ESI calc'd for $C_{52}H_{55}N_{13}O_{14}$ Na^+ $[MNa]^+$ calcd 1108.3884, found 1108.3885.

Pendant Intercalator D. HRMS-ESI calc'd for $C_{54}H_{58}N_{14}O_{15}$ Na^+ $[MNa]^+$ calcd 1165.4098, found 1165.4097.

2.6.3 Cloning

Cloning of PIS into pMoPac16 Vector: Footprinting analysis of the bisintercalators with the Hex sequences yielded 14 possible pendant intercalator binding sites to be further analyzed. A new sequence, PIS, was designed to include these binding sites. In order to preserve the sequence, it would be cloned into a vector, thus the inclusion of a SfiI restriction site on both sides of the sequence of interest. This new DNA fragment also was designed to contain an area for primer overlap beyond the SfiI sites not only to ensure the complete sequence for analysis could be amplified, but also to provide for various methods of cloning the fragment into a vector. The designed DNA fragment named PIS for pendant intercalation sites, was purchased from IDT as a gBlock® Gene Fragment.

Overlap extension PCR attempting to incorporate the PIS fragment into a pMoPac16 vector (Bryksin, 2010, Appendix Figures A3 and A4) was unsuccessful. The vector overlapping ends of the PIS gBlock acted as primers for the 20 μL overlap extension PCR. Phusion DNA polymerase was used for the reaction of 1 μL of 3 ng/μL pMoPac16 and 2.3 μL of 20 ng/μL PIS. The reaction was heated at 94 °C for 2 minutes, followed by 25 cycles of 30 seconds at 98 °C, 40 seconds at 58 °C, and 7 minutes at 72 °C, and ending with 10 minutes at 72 °C. The annealing temperature was derived from the priming portion of the PIS insert. A control was also made without the PIS insert. DpnI was then added to digest the template over one hour at 37 °C. Following desalting using nitrocellulose filters and electroporation, the cells were plated onto Luria media with ampicillin, and incubated at 37 °C overnight. Both the control and PIS plates had very few, small colonies.

The PIS DNA fragment was successfully cloned into the pMoPac16 plasmid using the Gibson assembly method (Gibson et al., 2009). Primers MIL-PI+V-for and MIL-

PI+V-rev from IDT were used with Phusion DNA polymerase to linearize the pMoPac16 vector (20 ng) (Table 2.2). This PCR reaction was placed in the thermocycler and heated at 94 °C for 2 minutes, followed by 25 cycles of 30 seconds at 98 °C, 40 seconds at 53 °C, and 7 minutes at 72 °C, and ending with 10 minutes at 72 °C, resulting in the amplified linear vector with sticky ends to later overlap the PIS fragment. The original methylated DNA from the bacterial vector was eliminated by digesting the PCR product with DpnI for one hour at 37°C. The resulting mixture was cleaned using the last few steps of a miniprep procedure. Qiagen Buffer PB was added using 5x the volume of the reaction mixture and transferred to a miniprep (or miniElute) column and centrifuged. The eluate was discarded and 750 µL Buffer PE was added to the column, centrifuged and discarded. The column was spin dried for 1 minute. The linear DNA was eluted from the column using 25 µL of distilled deionized water. The Gibson assembly was prepared on ice with the linearized vector (100 ng), PIS gBlock® fragment insert (20 ng) and Gibson Master Mix and diluted with water up to 20 µL total reaction volume. A control without the PIS fragment insert was also prepared. These reactions were incubated in a thermocycler for 60 minutes at 50 °C. Once the reaction was completed, the product was desalted, electroporated, plated with ampicillin, and incubated at 37 °C overnight. The control plate had few, if any colonies, while the Gibson plate had plenty of growth. Growths were made using Luria broth and ampicillin and incubated at 37 °C overnight. The DNA was isolated and purified using a Qiagen Miniprep Kit and subsequently submitted for sequencing. A stock was made using a bacterial growth of the pMoPac16 + PIS and glycerol and stored at -80 °C.

Primer Name	Primer Sequence
MIL-PI+V-for	CCTCCCCATGTCAGCATCAGTAGTCATACGGACGCTGAGCAAAGCAGACTAC
MIL-PI+V-rev	GCTCCTCATCACCTTTGTTGCCAACATACGAGCCGGAAGCAT
MIL-PI-for	GCAACAAAGGTGATGAGGAGC
MIL-PI-rev	CGTATGACTACTGATGCTGACATGG

Table 2.2: Primers utilized in PIS experiments.

2.6.4 DNase I Footprinting

Sequence Isolation, ^{32}P Labeling, and DNase I Footprinting: The Hex sequences within individual pUC19 vectors (Appendix Figures A1 and A2) were isolated by double restriction enzyme digests. Hex A forward and Hex A reverse were isolated from each vector by double digestion using HindIII-HF and SacI-HF (NEB) at 37 °C for 4 hours, while Hex B forward and Hex B reverse were digested using EcoRI-HF and PstI-HF (NEB) under the same conditions (Hampshire and Fox, 2008; Martinez, 2011). Each of the 3′-^{32}P-end labeled Hex sequences was then produced by incubation at 25 °C for 25 minutes with [α-^{32}P]-ATP (12.5 μL, 125 μCi) (PerkinElmer, EasyTide) and Klenow Fragment (3′ → 5′ exo-) (NEB) which lacks exonuclease activity in either direction. The labeled DNA was purified by 5 % native PAGE. The band with the desired length DNA was identified by exposing the plastic wrapped gel to a phosphor screen for up to a minute then slicing out a segment of the gel and rewrapping the gel before exposing again. The gel excision and exposure was repeated twice more before imaging the screen using ImageQuant TL v2005 and identifying the desired gel slice. The gel band was then excised from the gel, and divided and eluted into several 700 μL aliquots of 1x Tris-EDTA (TE) buffer at 40 °C with shaking at 400 rpm overnight. The subsequent ethanol precipitation is described below.

An alternative to using restriction enzymes for sequence isolation and labeling is to prepare the desired fragment by PCR. The PCR method alters the position of the radiolabel, placing it on the 5′ end of the DNA. All of the Hex sequences and the PIS sequence have been radiolabeled using the PCR method. Primers were designed and ordered from Integrated DNA Technologies (IDT) to amplify each desired sequence either from a vector or from the PIS sequence from gBlock® Gene Fragment (Table 2.2 and Table 2.3). The ordered PIS fragment was only 200 ng, restricting its number of uses. For use in this PCR method, part of the PIS fragment was amplified using primers MIL-PI-for and MIL-PI-rev (Table 2.2). While error can be introduced and propagated by multiple PCR reactions, having enough template predominated. This first amplification of PIS was then used as the stock template for the following PCR with radiolabel.

Primer Name	Primer Sequence
MIL-HexAf-for	GGTTTTCCCAGTCACGACG
MIL-HexAf-rev	CTATGACCATGATTACGCCAAGC

Table 2.3: Primers utilized in Hex radiolabeling for footprinting.

Each forward primer was 5′-^{32}P-end labeled using T4 polynucleotide kinase (Fermentas) and [γ-^{32}P]-ATP (12.5 μL, 125 μCi) (PerkinElmer, EasyTide) according to the manufacturers protocol. Once complete, the 100 μL labeling reaction was ended by addition of EDTA (5 μL, 0.5 M, pH 8.0). The labeled DNA was extracted four times with a phenol-chloroform-isoamyl alcohol mixture (25:24:1, 100 μL) with the DNA remaining in the top yellow aqueous layer. An illustra NICK column (GE Healthcare) was used to remove unincorporated radiolabel from the labeled DNA primer by loading the 100 μL DNA solution and eluting with water following the manufacturer's protocol. The 400 μL

aliquot containing the labeled forward primer was ethanol precipitated. A reverse primer solution was made using 3 μL of the reverse primer (20 μM) and diluting up to 50 μL with water. The forward primer pellet was resolubilized with the reverse primer solution. DNA containing the desired binding site from the appropriate vector or gBlock® was PCR amplified using this radiolabeled forward primer / reverse primer solution and Taq polymerase (NEB) by standard protocols. After preheating the thermocycler to 95 °C, the PCR reaction was placed in the thermocycler and heated at 95 °C for 2 minutes, followed by 30 cycles of 30 seconds at 95 °C, 30 seconds at 56 °C, and 1 minute at 72 °C, and ending with 10 minutes at 72 °C, resulting in a 5′-radiolabeled DNA oligonucleotide. The Hex PCR product was purified by 5 % native PAGE and the PIS PCR product by 3.5 % native PAGE. The band containing the desired length DNA was hot enough to be identified by TLC shadowing. The gel band was then excised from the gel, and divided and eluted into several 700 μL aliquots of 1x Tris-EDTA (TE) buffer at 40 °C with shaking at 400 rpm overnight.

DNase I footprinting was performed according to a previously reported procedure (Trauger and Dervan, 2001). After eluting the desired DNA overnight, the elution buffer was divided into 400 μL aliquots, and 150 μL was set aside for the adenine-specific cleavage reaction (A reaction). Each aliquot was ethanol precipitated, including that for the A reaction. One aliquot was resuspended with 400 μL 1x TE buffer by vortexing and transferred to another aliquot. This was repeated until all the radiolabeled DNA was recombined in one 400 μL aliquot, excluding the A reaction aliquot. The recombined aliquot was ethanol precipitated, resuspended in sodium phosphate buffer (20 mM, pH = 7.5) with $MgCl_2$ (2 mM) until a 20 μL aliquot gave a reading of 80-100 CPM, filtered, and stored at -20 °C. The A reaction DNA was resuspended in 160 μL deionized water and stored at -20 °C. The bisintercalators were incubated with radiolabeled DNA at 37 °C

overnight. Incubations were digested with 3 U/mL DNase I for 10 min. The adenine-specific cleavage reaction was carried out according to published procedure (Iverson and Dervan, 1987). Hex DNA fragments were separated on a 12 % denaturing polyacrylamide gel and the PIS DNA fragment on an 8 % denaturing polyacrylamide gel. The gels were exposed to a phosphor screen overnight and were imaged and analyzed using Quantity One 4.6.3 (Bio-Rad) and ImageQuant TL v2005.

Appendix

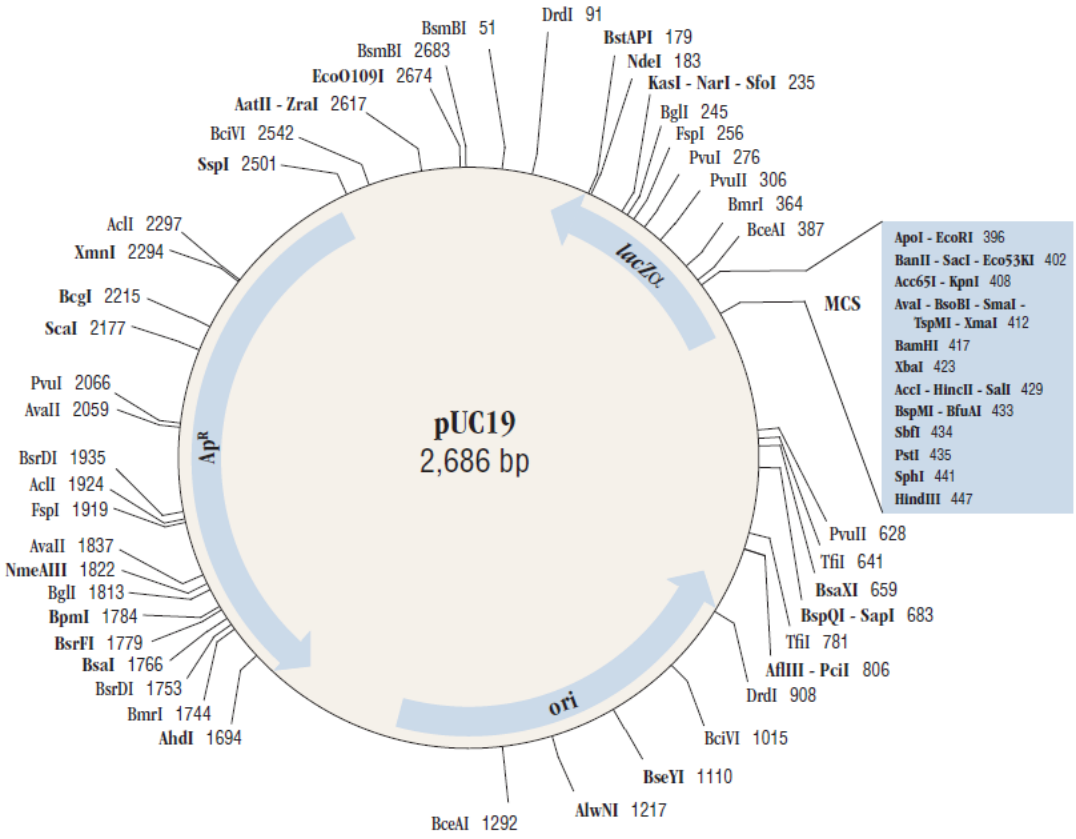


Figure A1: Vector map of pUC19 (NEB; Yanisch-Perron, 1985).

Figure A2: pUC19 vector sequence without Hex sequences.

TCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTG
TAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCT
TAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACAGATGCG
TAAGGAGAAAATACCGCATCAGGCGCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTG
CGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCC
AGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTCGAGCTCGGTACCCGGGGATCCTCTA
GAGTCGACCTGCAGGCATGCAAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCC
GCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCT
AACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAA
TGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTC
GCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAG
AATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCC
GCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAG
GTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTG
TTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGC
TCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGT
TCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGC
CACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAG
TGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTT
CGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCA
AGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCT
CAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCT
TTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAAT
GCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTC
GTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACG
CTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAA
CTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGT
TTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAG
CTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCG
GTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAAT
TCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGA
ATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAA
CTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGA
TCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGG
GTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCA
TACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAA
TGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGA
AACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTC

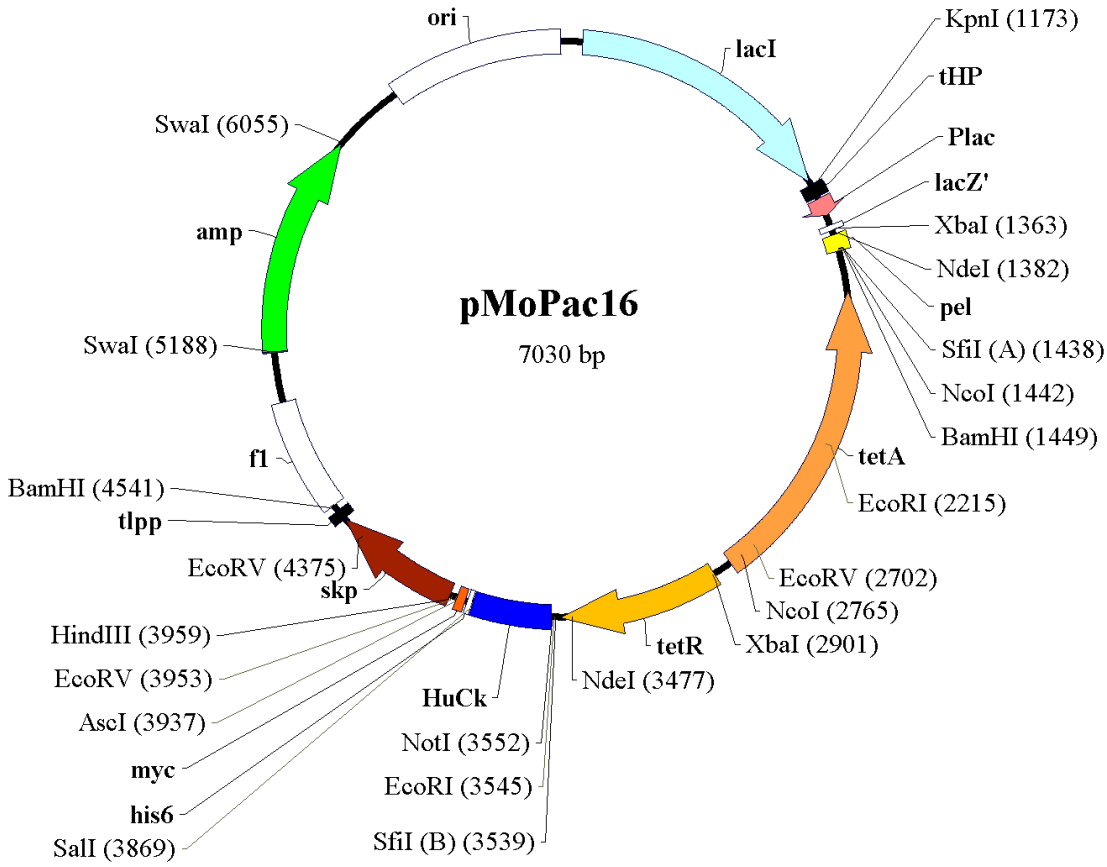


Figure A3: Vector map of pMoPac16 (Hayhurst et al., 2003).

Figure A4: pMoPac16 vector sequence without PIS sequence.

ACCCGCCACCATCGAATGGCGCAAAACCTTTCGCGGTATGGCATGATAGCGCCCGGAAGAGAGTCAATTCA
GGGTGGTGAATGTGAAACCAGTAACGTTATACGATGTCGCAGAGTATGCCGGTGTCTCTTATCAGACCGTT
TCCCGCGTGGTGAACCAGGCCAGCCACGTTTCTGCGAAAACGCGGGAAAAAGTGGAAGCGGCGATGGCGGA
GCTGAATTACATTCCCAACCGCGTGGCACAACAACTGGCGGGCAAACAGTCGTTGCTGATTGGCGTTGCCA
CCTCCAGTCTGGCCCTGCACGCGCCGTCGCAAATTGTCGCGGCGATTAAATCTCGCGCCGATCAACTGGGT
GCCAGCGTGGTGGTGTCGATGGTAGAACGAAGCGGCGTCGAAGCCTGTAAAGCGGCGGTGCACAATCTTCT
CGCGCAACGCGTCAGTGGGCTGATCATTAACTATCCGCTGGATGACCAGGATGCCATTGCTGTGGAAGCTG
CCTGCACTAATGTTCCGGCGTTATTTCTTGATGTCTCTGACCAGACACCCATCAACAGTATTATTTTCTCC
CATGAAGACGGTACGCGACTGGGCGTGGAGCATCTGGTCGCATTGGGTCACCAGCAAATCGCGCTGTTAGC
GGGCCCATTAAGTTCTGTCTCGGCGCGTCTGCGTCTGGCTGGCTGGCATAAATATCTCACTCGCAATCAAA
TTCAGCCGATAGCGGAACGGGAAGGCGACTGGAGTGCCATGTCCGGTTTTCAACAAACCATGCAAATGCTG
AATGAGGGCATCGTTCCCACTGCGATGCTGGTTGCCAACGATCAGATGGCGCTGGGCGCAATGCGCGCCAT
TACCGAGTCCGGGCTGCGCGTTGGTGCGGACATCTCGGTAGTGGGATACGACGATACCGAAGACAGCTCAT
GTTATATCCCGCCGTTAACCACCATCAAACAGGATTTTCGCCTGCTGGGGCAAACCAGCGTGGACCGCTTG
CTGCAACTCTCTCAGGGCCAGGCGGTGAAGGGCAATCAGCTGTTGCCCGTCTCACTGGTGAAAAGAAAAAC
CACCCTGGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGAC
AGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGGTACCCGATAAAAGCGGCTTCCTGACAGGAGGCCGTTTT
GTTTTGCAGCCCACCTCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTT
ATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGACCA
TGATTACGAATTTCTAGAGAAGGAGATATACATATGAAATACCTATTGCCTACGGCAGCCGCTGGATTGTT
ATTACTCGCGGCCCAGCCGGCGGATCCTTGCTATTTACCGCGGCTTTTTATTGAGCTTGAAAGATAAATAA
AATAGATAGGTTTTATTTGAAACTAAATCTTCTTTATCGTAAAAAATGCCCTCTTGGGTTATCAAGAGGGT
CATTATATTTCGCGGAATAACATCATTTGGTGACGAAATAACTAAGCACTTGTCTCCTGTTTACTCCCCTG
AGCTTGAGGGGTTAACATGAAGGTCATCGATAGCAGGATAATAATACAGTAAAACGCTAAACCAATAATCC
AAATCCAGCCATCCCAAATTGGTAGTGAATGATTATAAATAACAGTAAACAGTAATGGGCCAATAACACCG
GTTGCATTGGTAAGGCTCACCAATAATCCCTGTAAAGCACCTTGCTCATGACTCTTTGTTTGGATAGACAT
CACTCCCTGTAATGCAGGTAAAGCGATCCCACCACCAGCCAATAAAATTAAAACAGGGAAATCTAACCAAC
CTTCAGATATAAACGCTAAAAAGGCAAATGCACTACTATCTGCAATAAATTCGAGCAGTACTGCCGTTTTT
TCGCCCCATTTAGTGGCTATTCTTCCTGCCACAAAGGCTTGGAATACTGAGTGTAAAAGACCAAGACCCGC
TAATGAAAAGCCAACCATCATGCTATTCCATCCAAAACGATTTTCGGTAAATAGCACCCACACCGTTGCGG
GAATTTGGCCTATCAATTGCGCTGAAAAATAAATAATCAACAAAATGGGCATCGTTTTAAATAAAGTGATG
TATACCGAATTCGATTGCGTCTCAACCCCTACTTCGGTATCTGTATTATCACGTGTATTTTTGGTTTCACG
GAACCAAAACATAACCACAAGGAAAGTGACAATATTTAGCAACGCAGCGATAAAAAAGGGACTATGCGGTG
AAATCTCTCCTGCAAAACCACCAATAATAGGCCCCGCTATTAAACCAAGCCCAAAACTTGCCCCTAACCAA
CCGAACCACTTCACGCGTTGAGAAGCTGAGGTGGTATCGGCAATGACCGATGCCGCGACAGCCCCAGTAGC
TCCTGTGATCCCTGAAAGCAAACGGCCTAAATACAGCATCCAAAGCGCACTTGAAAAAGCCAGCAATAAGT
AATCCAGCGATGCGCCTATTAATGACAACAACAGCACTGGGCGCCGACCAAATCGGTCAGACATTTTTCCA
AGCCAAGGAGCAAAGATAACCTGCATTAACGCATAAAGTGCAAGCAATACGCCAAAGTGGTTAGCGATATC
TTCCGAAGCAATAAATTCACGTAATAACGTTGGCAAGACTGGCATGATAAGGCCAATCCCATCGAGTAACG
TAATTACCAATGCGATCTTTGTCGAACTATTCATTTCACTTTTCTCTATCACTGATAGGGAGTGGTAAAAT
AACTCTATCAATGATAGAGTGTCAACAAAAATTAGGAATTAATGATGTCTAGATTAGATAAAAGTAAAGTG
ATTAACAGCGCATTAGAGCTGCTTAATGAGGTCGGAATCGAAGGTTTAACAACCCGTAAACTCGCCCAGAA
GCTAGGTGTAGAGCAGCCTACATTGTATTGGCATGTAAAAAATAAGCGGGCTTTGCTCGACGCCTTAGCCA
TTGAGATGTTAGATAGGCACCATACTCACTTTTGCCCTTTAGAAGGGGAAAGCTGGCAAGATTTTTTACGT
AATAACGCTAAAAGTTTTAGATGTGCTTTACTAAGTCATCGCGATGGAGCAAAAGTACATTTAGGTACACG
GCCTACAGAAAAACAGTATGAAACTCTCGAAAATCAATTAGCCTTTTTATGCCAACAAGGTTTTTCACTAG
AGAATGCATTATATGCACTCAGCGCTGTGGGGCATTTTACTTTAGGTTGCGTATTGGAAGATCAAGAGCAT
CAAGTCGCTAAAGAAGAAAGGGAAACACCTACTACTGATAGTATGCCGCCATTATTACGACAAGCTATCGA
ATTATTTGATCACCAAGGTGCAGAGCCAGCCTTCTTATTCGGCCTTGAATTGATCATATGCGGATTAGAAA

AACAACTTAAATGTGAAAGTGGGTCTTAAAAGCCCCATCGGCCTCGGGGGCCGAATTCGCGGCCGCTGCAC
CATCTGTCTTCATCTTCCCGCCATCTGATGAGCAGTTGAAATCTGGAACTGCCTCTGTTGTGTGCCTGCTG
AATAACTTCTATCCCAGAGAGGCCAAAGTACAGTGGAAGGTGGATAACGCCCTCCAATCGGGTAACTCCCA
GGAGAGTGTCACAGAACAGGACAGCAAGGACAGCACCTACAGCCTCAGCAGCACCCTGACGCTGAGCAAAG
CAGACTACGAGAAACACAAAGTCTACGCCTGCGAAGTCACCCATCAGGGCCTGAGTTCGCCCGTCACAAAG
AGCTTCAACCGCGGAGAGTCAGTCGACCATCATCATCACCATCACGGGGCCGCAGAACAAAAACTCATCTC
AGAAGAGGATCTGAATGGGCGCGCCGCATAGTGATATCGCAAGCTTTAAGGAGATATATATATGAAAAAGT
GGTTATTAGCTGCAGGTCTCGGTTTAGCACTGGCAACTTCTGCTCAGGCGGCTGACAAAATTGCAATCGTC
AACATGGGCAGCCTGTTCCAGCAGGTAGCGCAGAAAACCGGTGTTTCTAACACGCTGGAAAATGAGTTCAA
AGGCCGTGCCAGCGAACTGCAGCGTATGGAAACCGATCTGCAGGCTAAAATGAAAAAGCTGCAGTCCATGA
AAGCGGGCAGCGATCGCACTAAGCTGGAAAAAGACGTGATGGCTCAGCGCCAGACTTTTGCTCAGAAAGCG
CAGGCTTTTGAGCAGGATCGCGCACGTCGTTCCAACGAAGAACGCGGCAAACTGGTTACTCGTATCCAGAC
TGCTGTGAAATCCGTTGCCAACAGCCAGGATATCGATCTGGTTGTTGATGCAAACGCCGTTGCTTACAACA
GCAGCGATGTAAAAGACATCACTGNCGACGTACTGAAACAGGTTAAATAATAAGACCTGTGAAGTGAAAAA
TGGCGCACATTGTGCGACATTTTTTTTGTCTGCCGTTTACCGCTACTGCGTCACGGATCCCCACGCGCCCT
GTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTA
GCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAA
TCGGGGCATCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTG
ATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTT
AATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGG
GATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACA
AAATATTAACGTTTACAATTTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTT
TTCTAAATACATTCAAATATGTATCCGCTCATGTCGAGACGTTGGGTGAGGTTCCAACTTTCACCATAATG
AAATAAGATCACTACCGGGCGTATTTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAATTTAAATGAGT
ATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGA
AACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCA
ACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTG
CTATGTGGCGCGGTATTATCCCGTGTTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCA
GAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTAT
GCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAG
GAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAA
TGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGCAGCAATGGCAACAACGTTGCGCAAACTAT
TAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCA
GGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGG
GTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGG
GGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGG
TAAATTTAAATAGGCAGTTATTGGTGCCCTTAAACGCCTGGTGCTACGCCTGAATAAGTGATAATAAGCGG
ATGAATGGCAGAAATTCGAAAGCAAATTCGACCCGGTCGTCGGTTCAGGGCAGGGTCGTTAAATAGCCGCT
TATGTCTATTGCTGGTTTACCGGTTTATTGACTACCGGAAGCAGTGTGACCGTGTGCTTCTCAAATGCCTG
AGGCCAGTTTGCTCAGGCTCTCCCCGTGGAGGTAATAATTGCTCGACATGACCAAAATCCCTTAACGTGAG
TTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCG
CGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTAC
CAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCG
TAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGT
GGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGC
AGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGA
TACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAG
CGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTG
TCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAA
AACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATG

References

Acramone, F.; Penco, S.; Orezzi, P.; Nicolella, V.;Pirelli, A. *Nature.* **1694,** 203, 1064 – 1065.

Albericio, F. *Biopolymers (Pepetide Science).* **2000,** 55, 123 – 139.

Armen, R.S.; Schiller, S.M.; Brooks III, C.L. *Proteins.* **2010,** 78, 1926 – 1938.

Bond, P.J.; Langridge, R.; Jennette, K.W.; Lippard, S.J. *Proc. Natl. Acad. Sci. U.S.A.* **1975,** *72*, 4825 – 4829.

Bousquet, P.F.; Brana, M.F.; Conlon, D.; Fitzgerald, K.M., Perron, D., Cocchiaro, C., Miller, R., Moran, M., George, J., Qian, X.D., et al. *Cancer Res.* **1995,** 55, 1176 – 1180.

Brase, S.; Kirchhoff, J.H.; Kobberling, J. *Tetrahedron.* **2003,** 59, 885 – 939.

Brase, S.; Kobberling, J.; Griebenow, N. Organopalladium Reactions in Combinatorial Chemistry. In *Handbook of Organopalladium Chemistry for Organic Synthesis, Vol 2*; Negishi, E., Ed.; John Wiley & Sons, Inc., New York, **2002**; pp 3102 – 3126.

Brogden, A.L.; Hopcroft, N.H.; Searcey, M.; Cardin, C.J. *Angew. Chem. Int. Ed. Engl.* **2007,** 46, 3850 – 3854.

Bryksin, A.V.; Matsumura, I. *BioTechniques.* **2010,** 48, 463 – 465.

Carpino, L.A.; Ionescu, D.; El-Faham, A. *J. Org. Chem.* **1996,** 61, 2460 – 2465.

Chaires, J.B.; Dattagupta, N.; Crothers, D.M. *Biochemistry.* **1985,** 24, 260 – 267.

Chaires, J.B.; Leng, F.; Przewloka, T.; Fokt,I.; Ling, Y-H.; Perez-Soler, R.; Priebe, W. *J. Med. Chem.* **1997,** 40, 3, 261 – 266.

Cheng, Y-K.; Pettett, M. *Prog. Biophys. Molec. Biol.* **1992,** 58, 225 – 257.

Choudhury, J.R.; Bierbach, U. *Nucleic Acids Res.* **2005,** 33, 5622 – 5632.

Crothers, D.M.; *Biopolymers.* **1968,** 6, 575 – 584.

Dervan, P.B. *Bioorg. Med. Chem.* **2001,** 9, 2215 – 2235.

Dickinson, L.A.; Gulizia, R.J.; Trauger, J.W.; Baird, E.E.; Mosier, D.E.; Gottesfeld, J.M.; Dervan, P.B. *Proc. Natl. Acad. Sci. U.S.A.* **1998,** 95, 22, 12890 – 12895.

Duca, M.; Vekhoff, P.; Oussedik, K.; Halby, L.; Arimondo, P.B. *Nucleic Acids Res.* **2008,** 36, 16, 5123 – 5138.

Dumas, A.; Lercher, L.; Spicer, C. D.; Davis, B. G. *Chem. Sci.* **2015,** 6, 50 – 69.

Far, S.; Kossanyi, A.; Verchere-Beaur, C.; Gresh, N.; Taillandier, E.; Perree-Fauvet, M. *Eur. J. Org. Chem.* **2004,** 1781 – 1797.

Fekner, T.; Li, X.; Lee, M. M.; Chan, M. K. *Angew. Chem. Int. Ed. Engl*. **2009**, 48, 1633 – 1635.

Fekry, M.I.; Szekely, J.; Dutta, S.; Breydo, L.; Zang, H.; Gates, K.S. *J. Am. Chem. Soc.* **2011,** 133, 17641 – 17651.

Gallego, J.; Reid, B.R. *Biochemistry.* **1999,** 38, 15104 – 15115.

Gibson, D.G.; Young, L.; Chuang, R.; Venter, J.C.; Hutchison III, C.A.; Smith, H.O. *Nat. Methods.* **2009,** 6, 343 – 345.

Giovannangeli, C.; Perrouault, L.; Escude, C.; Gryaznov, S.; Helene, C. *J. Mol. Biol.* **1996,** 261, 386 – 398.

Gottesfeld, J.M.; Belitsky, J.M.; Melander, C.; Dervan, P.B.; Luger, K.*J. Mol. Biol.* **2002,** 321, 249 – 263.

Green, M.; Berman, J. *Tetrahedron Lett.* **1990,** 31, 41, 5851 – 5852.

Guelev, V.M.; Harting, M.T.; Lokey, R.S.; Iverson, B.L. *Chem. Biol.* **2000,** 7, 1 – 8.

Guelev, V.M.; Sorey, S.; Hoffman, D.W.; Iverson, B.L. *J. Am. Chem. Soc.* **2002,** 124, 2864 – 2865.

Guelev, V.M.; Lee, J.; Ward, J.; Sorey, S.; Hoffman, D.W.; Iverson, B.L. *Chem. Biol.* **2001a,** 8, 415 – 425.

Guelev, V.M.; Cubberley, M.S.; Murr, M.M.; Lokey, R.S.; Iverson, B.L. *Method Enzymol.* **2001b,** 340, 556 – 570.

Guha, S.; Saha, S. *J. Am. Chem. Soc.* **2010,** 132, 17674 – 17677.

Hampshire, A.J.; Fox, K.R. *Anal. Biochem.* **2008,** 374, 298 – 303.

Han, Y.; Albericio, F.; Barany, G. *J. Org. Chem.* **1997,** 62, 4307 – 4312.

Hendrickson, T. L.; de Crécy-Lagard, V.; Schimmel, P. *Annu. Rev. Biochem.* **2004,** 73, 147 – 176.

Hohsaka, T.; Kajihara, D.; Ashizuka, Y.; Murakami, H.; Sisido, M. *J. Am. Chem. Soc.* **1999,** 121, 34 – 40.

Holman, G.G. PhD. Dissertation, The University of Texas at Austin, 2011.

Holman, G.G.; Zewail-Foote, M.; Rhoden Smith, A.; Johnson, K.A.; Iverson, B.L. *Nat. Chem*. **2011,** 3, 875 – 881.

Ikeda, H.; Nakamura, Y.; Saito, I. *Tetrahedron Lett.* **2002,** 43, 5525 - 5528

Iverson, B.L.; Dervan, P.B. *Nucleic Acids Res.* **1987,** 15, 7823 – 7830.

Johnson, C.A.; Hudson, G.A.; Hardebeck, L.K.E.; Jolley, E.A.; Ren, Y.; Lewis, M.; Znosko, B.M. *Bioorg. Med. Chem.* **2015,** 23, 3586 – 3591.

Kang, J.S.; Meier, J.L.; Dervan, P.B. *J. Am. Chem. Soc.* **2014,** 136, 3687 – 3694.

Kisfaludy, L.; Schon, I. *Synthesis.* **1983,** 4, 325 – 327.

Kukreti, S.; Sun, J-S.; Garestier, T.; Helene, C. *Nucleic Acids Res.* **1997,** 25, 21, 4264 – 4270.

Lacey, V.K.; Louie, G.V.; Noel, J.P.; Wang, L. *ChemBioChem*. **2013,** 14, 2100 – 2105.

Laugaa, P.; Markovits, J.; Delbarre, A.; Le Pecq, J.B.; Roques, B.P. *Biochemistry*. **1985,** 24, 5567 – 5575.

Lee, J.; Guelev, V.M.; Sorey, S.; Hoffman, D.W.; Iverson, B.L. *J. Am. Chem. Soc.* **2004,** 126, 14036 – 14042.

Lee, J.S.; Woodsworth, M.L.; Latimer, L.J.; Morgan, A.R. *Nucleic Acids Res.* **1984,** 12, 6603 – 6614.

Lerman, L.S. *J. Mol. Biol*. **1961,** 3, 18-30, IN13 – IN14.

Li, B.C.; Montgomery, D.C.; Puckett, J.W.; Dervan, P.B. *J. Org. Chem.* **2013,** 78, 124 – 133.

Li, H.J.; Crothers, D.M. *J. Mol. Biol.* **1969,** 39, 3, 461 – 477.

Li, W-T.; Mahapatra, A.; Longstaff, D. G.; Bechtel, J.; Zhao, G.; Kang, P. T.; Chan, M. K.; Krzycki. J. A.; *J. Mol. Biol*. **2009,** 385, 1156 – 1164.

Liu, C.C.; Schultz, P. G. *Annu. Rev. Biochem*. **2010,** 79, 413 – 444.

Lohse, J.; Dahl, O.; Nielsen, P.E. *Proc. Natl. Acad. Sci. U.S.A*. **1999,** 96, 21, 11804 – 11808.

Lokey, R.S.; Kwok, Y.; Guelev, V.; Pursell, C.J.; Hurley, L.H.; Iverson, B.L. *J. Am. Chem. Soc.* **1997,** 119, 7202 – 7210.

Luo, J.; Uprety, R.; Naro, Y.; Chou, C.; Nguyen, D. P.; Chin, J. W.; Deiters, A. *J. Am. Chem. Soc.* **2014,** 136, 15551 – 15558.

Marchetti, C.; Minarini, A.; Tumiatti, V.; Moraca, F.; Parrotta, L.; Alcaro, S.; Rigo, R.; Sissi, C.; Gunaratnam, M.; Ohnmacht, S.A.; Neidle, S. et al. *Bioorg. Med. Chem.* **2015,** 23, 13, 3819 – 3830.

Martin, B.; Vaquero, A.; Priebe, W.; Jose, P. *Nucleic Acids Res.* **1999,** 27, 3402 – 3409.

Martinez, C. R. PhD. Dissertation, The University of Texas at Austin, 2011.

Martinez, C.R.; Iverson, B.L. *Chem. Sci.* **2012,** 3, 2191 – 2201.

McKnight, R.E.; Reisenauer, E.; Pintado, M.V.; Polasani, S.R.; Dixon, D.W. *Bioorg. Med. Chem. Lett.* **2011,** 21, 4288 – 4291.

Meier, J.L.; Yu, A.S.; Korf, I.; Segal, D.J.; Dervan, P.B. *J. Am. Chem. Soc.* **2012,** 134, 17814 – 17822.

Melander, C.; Burnett, R.; Gottesfeld, J.M. *J. Biotech.* **2004,** 112, 195 – 220.

Moriguchi, T.; Zafrul Azam, A.T.M.; Shinozuka, K. *Bioconjugate Chem.* **2011,** 22, 1039 – 1045.

Moser, H.E.; Dervan, P.B. *Science.* **1987,** 238, 645 – 650.

Mukai, T.; Kobayashi, T.; Hino, N.; Yanagisawa, T.; Sakamoto, K.; Yokoyama, S. *Biochem. Biophys. Res. Commun.* **2008,** 371, 818 – 822.

Mukherjee, A.; Sasikala, W.D. *Adv. Protein Chem. Struct. Biol.* **2013,** 92, 1-62.

Muller, W.; Crothers, D.M. *Eur. J. Biochem.* **1975,** 54, 267 – 277.

Murr, M.M.; Harting, M.T.; Guelev, V.; Ren, J.; Chaires, J.B.; Iverson, B.L. *Bioorg. Med. Chem.* **2001,** *9*, 1141 – 1148.

Niedle, S. *Principles of Nucleic Acid Structure*; Elsevier Science, 2010.

Nielsen, P.E. *Curr. Med. Chem.* **2001,** 8, 545 – 550.

Nielsen, P.E. *ChemBioChem.* **2010,** 11, 2073 – 2076.

Nielsen, P.E.; Egholm, M.; Buchardt, O. *Gene.* **1994,** 149, 139 – 145.

Nojima, T.; Ohtsuka, K.; Nagamatsu, T.; Takenaka, S. *Nucleic Acids Symp. Ser.* **2003,** 3, 123 – 124.

Onfelt, B.; Lincoln, P.; Norden, B. *J. Am. Chem. Soc.* **2001,** 123, 16, 3630 – 3637.

Pedersen, E.B.; Osman, A.M.A.; Globisch, D.; Paramasivam, M.; Cogoi, S.; Bomholt, N.; Jorgensen, P.T.; Xodo, L.E.; Filichev, V.V. *Nucleic Acids Symp. Ser.* **2008,** 52, 37 – 38.

Portugal, J.; Martin, B.; Vaquero, A.; Ferrer, N.; Villamarin, S.; Priebe, W. *Curr. Med. Chem.* **2001,** 8, 1 – 8.

Ray, A.; Norden, B. *FASEB J.* **2000,** 14, 9, 1041 – 1060.

Rhoden Smith, A.; Iverson, B.L. *J. Am. Chem. Soc.* **2013,** 135, 12783 – 12789.

Rhoden Smith, A.; Ikkanda, B.A.; Holman, G.G.; Iverson, B.L. *Biochemistry.* **2012,** 51, 4445 – 4452.

Rigby, H.H.; Moore, T.L.; Rege, S. *J. Org. Chem.* **1986,** 51, 2402 – 2404.

Robinson, H., Priebe, W., Chaires, J.B., Wang, A.H.J. *Biochemistry*. **1997,** 36, 8663 – 8670.

Sasikala, W.D.; Mukherjee, A. *Phys. Chem. Chem. Phys.* **2013,** 15, 6446 – 6455.

Sharples, D.; Brown, J.R. *FEBS Letters.* **1976,** 69, 37 – 40.

Speight, L. C.; Muthusamy, A. K.; Goldberg, J. M.; Warner, J. B.; Wissner, R. F.; Willi, T. S.; Woodman, B. F.; Mehl, R. A.; Petersson, E. J. *J. Am. Chem. Soc.* **2013,** 135, 18806 – 18814.

Strekowski, L.; Wilson, B. *Mutation Research.* **2007,** 623, 3 – 13.

Subiros-Funosas, R.; El-Faham, A.; Albericio, F. *Org. Biomol. Chem.* **2010,** 8, 3665 – 3673.

Suseela, Y.V.; Das, S.; Pati, S.K.; Govindaraju, T. *ChemBioChem.* **2016,** 17, 2162 – 2171.

Tambara, K.; Ponnuswamy, N.; Hennrich, G.; Pantoş, G.D *J. Org. Chem.* **2011,** 76, 3338 – 3347.

Tanious, F.A.; Yen, S.F.; Wilson, W.D. *Biochemistry.* **1991,** 30, 1813 – 1819.

Trauger, J.W.; Dervan, P.B. *Method Enzymol.* **2001,** 340, 450 – 466.

Trauger, J.W.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1998,** 120, 3534 – 3535.

Tumiatti, V.; Milelli, A.; Minarini, A.; Micco, M.; Gasperi Campani, A.; Roncuzzi, L.; Baiocchi, D.; Marinello, J.; Capranico, G.; Zini, M. et al. *J. Med. Chem.* **2009,** 52, 7873 – 7877.

Vazquez, M.E.; Caamano, A.M.; Mascarenas, J.L. *Chem. Soc. Rev.* **2003,** 32, 338 – 349.

Waring, M.J.; Wakelin, L.P.G. *Nature.* **1974,** 252, 653 – 657.

Williams, L.D.; Elgi, M.; Gao, Q.; Rich, A. (**1992**). DNA Intercalation: Helix Unwinding and Neighbor-Exclusion. In Struct. Funct., Proc. Conversation Discip. Biomol. Stereodyn., 7th, R.H. Sarma, and M.H. Sarma, eds. (Schenectady, NY: Adenine Press), pp. 107–125.

Wirth, M.; Buchardt, O.; Koch, T.; Nielsen, P.E.; Norden, B. *J. Am. Chem. Soc.* **1988,** 110, 932 – 939.

Yanagisawa, T.; Ishii, R.; Fukunaga, R.; Kobayashi, T.; Sakamoto, K.; Yokoyama, S. *Chem. Biol.* **2008,** 15, 1187 – 1197.

Yen, S.F.; Gabbay, E.J.; Wilson, W.D. *Biochemistry.* **1982,** 21, 2070 – 2076.

Young, T.S.; Schultz, P.G. *J. Biol. Chem.* **2010,** 285, 15, 11039 – 11044.

Zeglis, B.M.; Pierre, V.C.; Barton, J.K. *Chem. Commun.(Camb).* **2007,** 44, 4565 – 4579.

Vita

Maria I. Lambousis was born and raised in Cheyenne, Wyoming. She obtained her B.S. in Chemistry and completed the Honor's Program at the University of Wyoming. While at UW, she was introduced to research by Dr. Mark Mehn before conducting nucleic acid chirality studies and spectroscopic detection of Z-DNA under Dr. Milan Balaz. Following graduation in 2010, she entered the graduate program at the University of Texas at Austin and joined the lab of Dr. Brent Iverson.

Permanent email: mlambousis@utexas.edu

This thesis was typed by the author.