The Dissertation Committee for Garen Gilman Holman Certifies that this is the approved version of the following dissertation:

Binding Studies of a Sequence Specific Threading NDI Tetraintercalator

Committee:

Brent L. Iverson, Supervisor

Eric V. Anslyn

David W. Hoffman

Sean M. Kerwin

Jonathan L. Sessler

Binding Studies of a Sequence Specific Threading NDI Tetraintercalator

by

Garen Gilman Holman, B.S.

Dissertation

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

Doctor of Philosophy

The University of Texas at Austin

August 2011

Dedication

To my wonderful wife and children.

Acknowledgements

I would like to thank my advisor, Brent Iverson, for the patience and motivation that he is able to provide. He has provided more than scientific advice, but constant reminding of how successful we can become by just being nice to those who we work with and teach.

It goes without saying, that the dynamic in our lab is one of friendship and trust, much to be said for a group of scientist trying to make their way through graduate school. As a group, we were able to make it through all challenges, both scientific and personal.

Finally, I would like to thank my wife and family for the constant joy and perspective that they bring to my life. Without them, I would be nothing. The happiness and joy that I have found in my family provide me the desire and will to succeed. My new sons have given me a whole new outlook on life and relationships.

Binding Studies of a Sequence Specific Threading NDI Tetraintercalator

Garen Gilman Holman, Ph.D.

The University of Texas at Austin, 2011

Supervisor: Brent L. Iverson

A series of studies from our lab have investigated the threading polyintercalator approach to sequence specific DNA binding using a 1,4,5,8-naphthalene tetracarboxylic diimide (NDI) intercalating unit connected by flexible peptide linkers. Herein is a report of the sequence specificity, as well as a detailed kinetic analysis, of a threading NDI tetraintercalator. DNase I footprinting using two ~500 base pair DNA fragments containing one designed binding site for the tetraintercalator confirmed highly sequence specific binding. Kinetic analyses include ^{1}H NMR, gel mobility-shift assays, and stopped-flow UV measurements to reveal a polyintercalation binding mode that demonstrates significant similarities between association rate profiles and rate constants for the tetraintercalator binding to its preferred versus a random oligonucleotide sequence. Sequence specificity was found to derive almost entirely from large differences in dissociation rates from the preferred versus random oligonucleotide sequences. Interestingly, the dissociation rate constant of the tetraintercalator complex dissociating from its preferred binding site was extremely slow, corresponding to a 16 day half-life at a benchmark 100 mM [Na^{+}]. This dissociation result for the tetraintercalator is one of the

longest bound half-lives yet measured, and to the best of our knowledge, the longest for a DNA binding small molecule. Such a long-lived complex raises the possibility of using threading polyintercalators to disrupt biological processes for extended periods.

Current focus is given to deciphering a mechanism for the molecular recognition of the tetraintercalator preferred binding site within a long sequence of DNA. Initial DNase I footprinting results on an approximate 500mer DNA sequence containing three sequential preferred binding sites reveal that the tetraintercalator likely locates its designed binding site by a macro- or microscopic dissociation/re-association type of mechanism. Cooperativity is a possible ally to binding, leaving future studies to distinguish the mechanism for molecular recognition in a manner that is capable of circumventing cooperative binding. Taken together, the threading polyintercalation binding mode presents an interesting topology to sequence specific DNA binding. Extraordinarily long dissociation rates from preferred binding sites offers many future possibilities to disrupt biological processes *in vivo*.

Table of Contents

List of Tables

List of Figures

List of Schemes

Chapter 1

Overview of Small Molecules Binding to DNA

1.1 Small Molecules as Molecular Biology Tools for Targeting DNA

Research focused on the discovery and development of new therapeutic drugs has been a topic of interest for several decades.[1] So much has new drug discovery been desired, that technology and innovation have brought about the advent of new high throughput screening methods[2,3,4] and both rational[5,6] and structure based[7,8,9] drug design in searching for potentially interesting therapeutics.

The central dogma of molecular biology[10] states that genetic material is stored in DNA, which is transcribed into single stranded mRNA. The mRNA carries the code into the cytoplasm from the nucleus, where it is then translated via the ribosome and individual amino acid building blocks into proteins for some ultimate function (**Figure 1.1**). The resultant proteins control cellular function through their roles as enzymes, receptors, antibodies, structural functions, and a variety of other functions. There is no shortage of cellular targets for specific diseases, with small molecules that are capable of targeting all aspects of the central dogma. For instance, targeting DNA is to target the information storage of a cell, possibly to disrupt either cellular replication totally,[11,12,13] or just to regulate individual gene expression.[14,15] Targeting RNA directly is an approach to disrupt the message before it can be put into action to translate proteins or to reverse-transcribe DNA.[16,17,18] Aside from targeting directly information storage (DNA) or the process of delivering the information (RNA), specific cellular functions can be stopped by targeting the individual proteins responsible for crucial functions.[19,20,21]

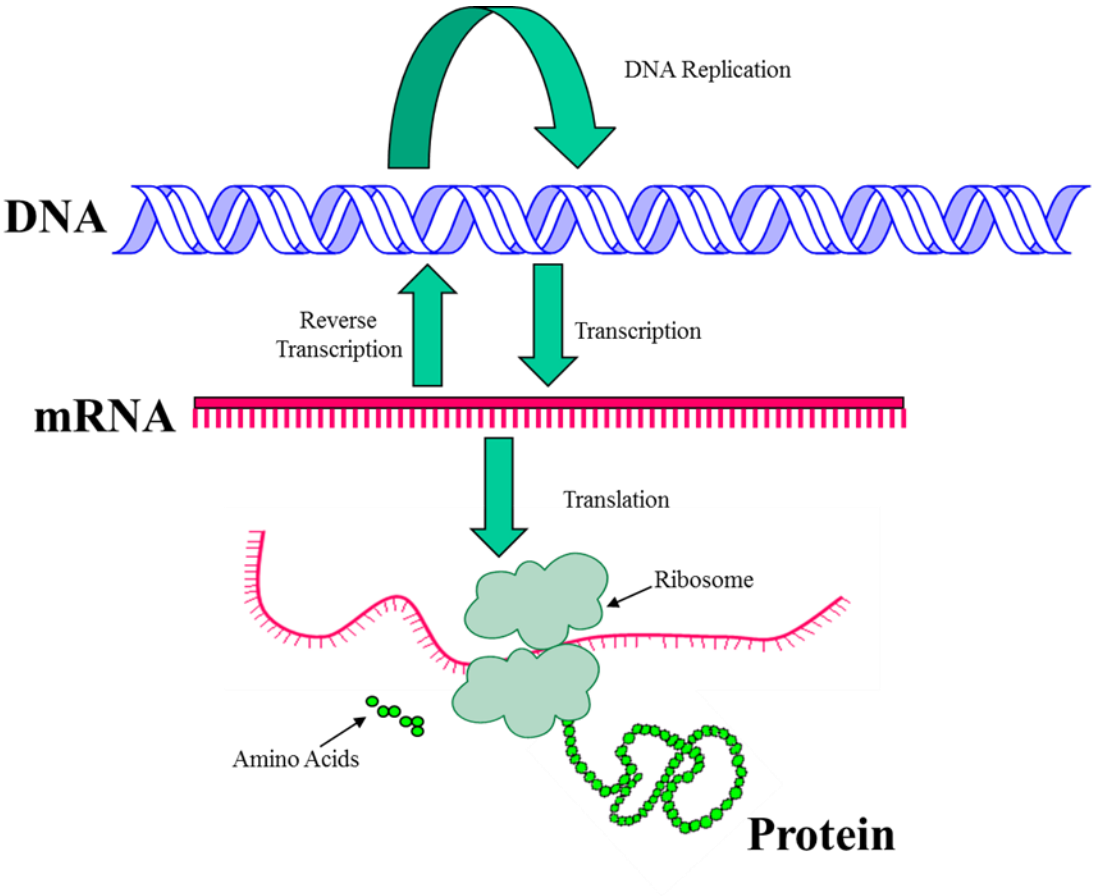


Figure 1.1 Central Dogma of Molecular Biology as proposed by F. Crick.[10]

While mRNA is capable of reverse transcription to produce DNA, one key concept of the central dogma is that ultimately, DNA is responsible for the replication and survival of the cell. Put another way, DNA is the single "master key" of information inside of a cell that is used to produce multiple copies of RNA, which in turn produce multiple copies of proteins. Knowing this, DNA is a highly desirable target, provided that it is mostly localized in the nucleus of each cell. Targeting DNA provides the highest level of control, in that blocking the information storage can disrupt transcription via sequence-specific DNA binding, from which the downfield production of mRNA and proteins are affected. It would also appear to be more efficient to target DNA since there is only one genomic copy inside of a non-dividing cell (not including mitochondrial DNA) as opposed to the continued reproduction of transcribed mRNA and translated proteins over the life of the cell.

Molecular recognition of DNA is important for binding specific sequences of DNA, especially in the case of trying to regulate a specific gene.[22,23] In order to target DNA of some length N, the potential sequences can be composed of four possible DNA nucleotides (A, G, C, T) producing up to 4^N possibilities of N base pair (bp) sequences.[24] For instance, several small-molecules target DNA sequences of between 4-6 bp,[25,26] of which there would be $2.5 \times 10^2 - 4 \times 10^3$ possible sequences, respectively. Increasing the binding site size to recognizing 16 base pairs allows for $\sim 4 \times 10^9$ possible sequences. These different sequence possibilities allows for the determination of the uniqueness of a targeted sequence within a particular genome. Provided that the human genome contains approximately 3×10^9 bp, the likelihood that a binding site sequence of *n* bp would occur (utilizing only A, G, C, or T), would be $3 \times 10^9 \times 4^{-n}$.[24] Using this, recognizing a binding site approaching 16 bp should occur only once, or not at all, within the human genome.[5,24]

1.2 STRUCTURAL PROPERTIES OF DNA

DNA is a biopolymer that is built using four nucleobase subunits. The individual nucleobases are adenine, guanine, cytosine, and thymine. The individual bases are connected to the 1' position of a 2'-deoxyribose sugar, and so named as the appropriate deoxynucleosides (deoxyadenosine, deoxyguanosine, thymidine, and deoxycytidine). The deoxynucleoside bases are covalently attached to each other via a phosphodiester backbone through both the 3' and 5' positions on the deoxyribose sugar to produce a given single stranded DNA (ssDNA) sequence (**Figure 1.2**).[27] DNA in solution typically exists as a double-stranded helix, which is the combination of two antiparallel strands of ssDNA connected via Watson-Crick base pairing. Adenine base pairs to thymine and guanine base pairs to cytosine through two and three hydrogen bonds each, respectively (**Figure 1.3**). The phosphate backbone of the DNA is negatively charged and thereby hydrophilic, while the nucleobases in the interior of the helix are aromatic and hydrophobic surfaces that are capable of stacking. While hydrogen-bonds are largely responsible for the specificity of base-pairing interactions, it is the favorable hydrophobic interactions and aromatic base stacking of the interior bases and electrostatic forces of the phosphate backbone that are responsible for the overall stability of the DNA double-helix.[27]

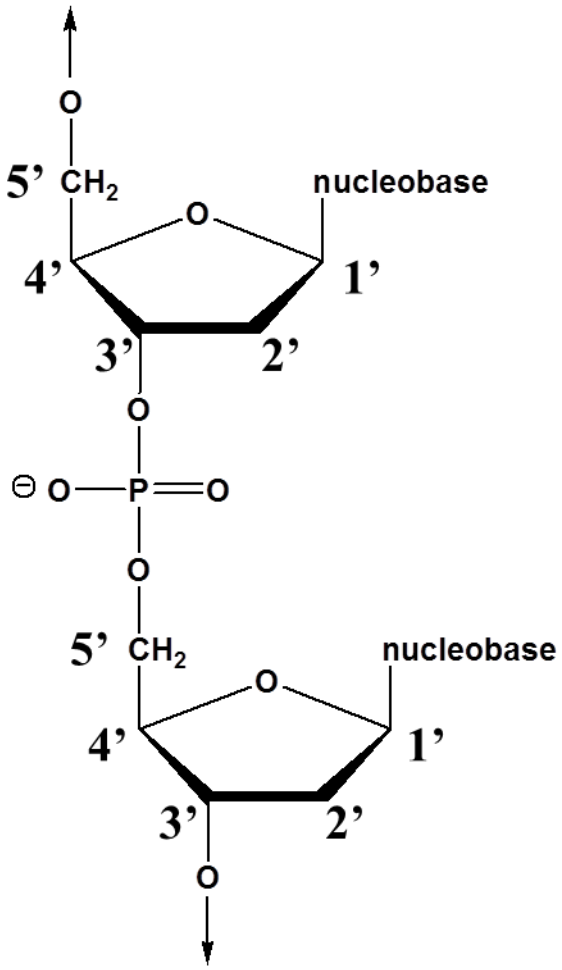


Figure 1.2 Chemical structure of the phosphodiester backbone of DNA with the labeled carbons as named around the deoxyribose ring.

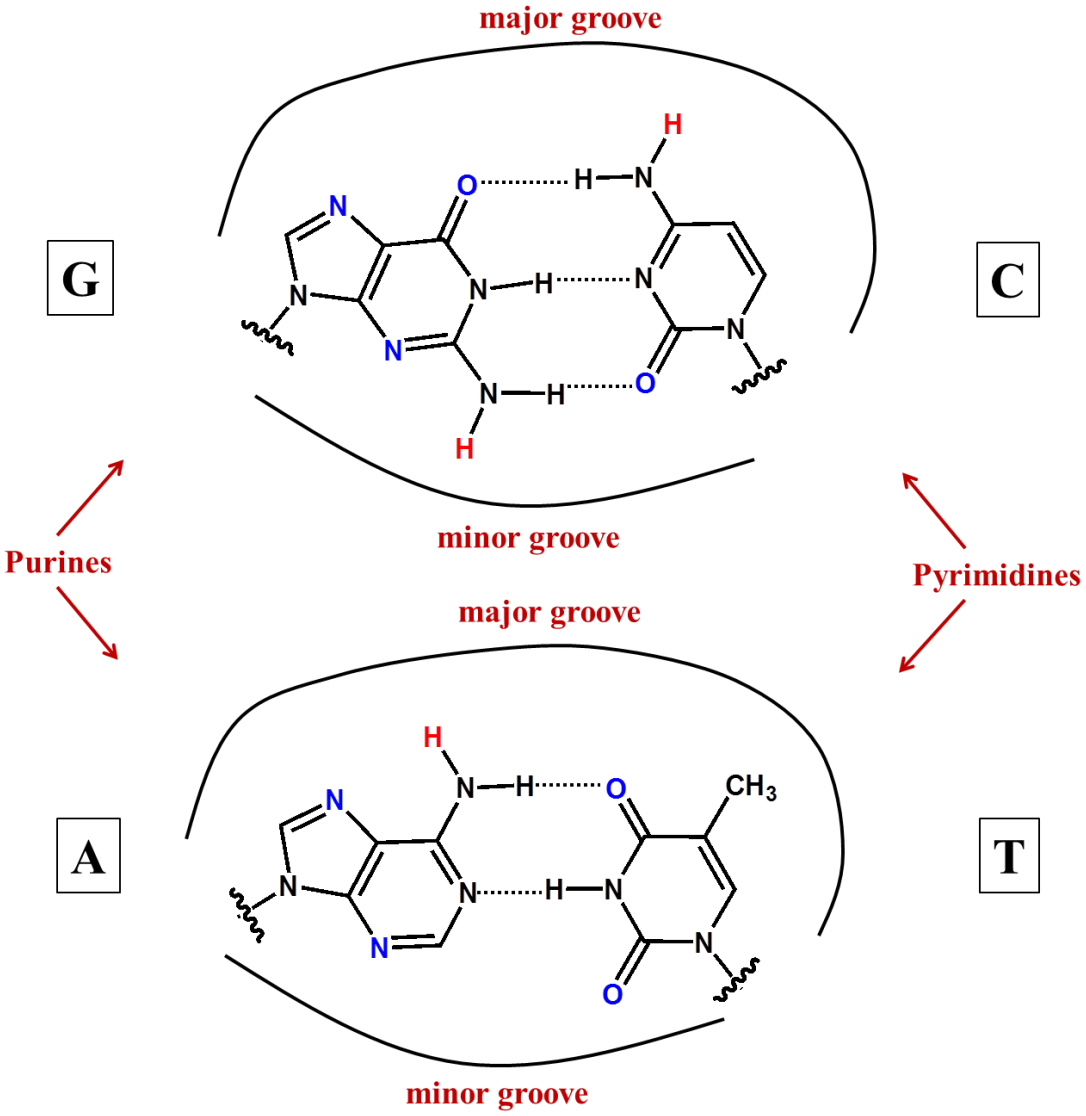


Figure 1.3 Watson-Crick base pairing of Guanine-Cytosine and Adenine-Thymine with hydrogen bonds indicated as hashed lines, potential hydrogen-bond donors shown in red, and potential hydrogen-bond acceptors shown in blue. Major and minor grooves are as labeled. Purine bases G & A, as well as pyrimidine bases C & T are indicated with arrows.

There are several configurations of DNA, with A-, B-, and Z-form conformations (A- & B-form DNA can be seen in **Figure 1.4**). The most common structure under physiological conditions is B-form DNA, which has a right-handed helical twist with 10 stacked base pairs per 360° rotation of the helix. The stacked nucleobases are nearly perpendicular to the helical axis, with a 3.4 Å separation between each stacked base. The deoxyribose ring conformation adopts a C2'-*endo* sugar pucker. Typical B-form DNA has a shallow and wide major groove, and a relatively deep and narrow minor groove. One characteristic of sequence-specific recognition of DNA is that many small molecules and proteins can differentiate between the localized topologies of the DNA grooves, which will vary depending on the specific nucleobase composition of the sequence.

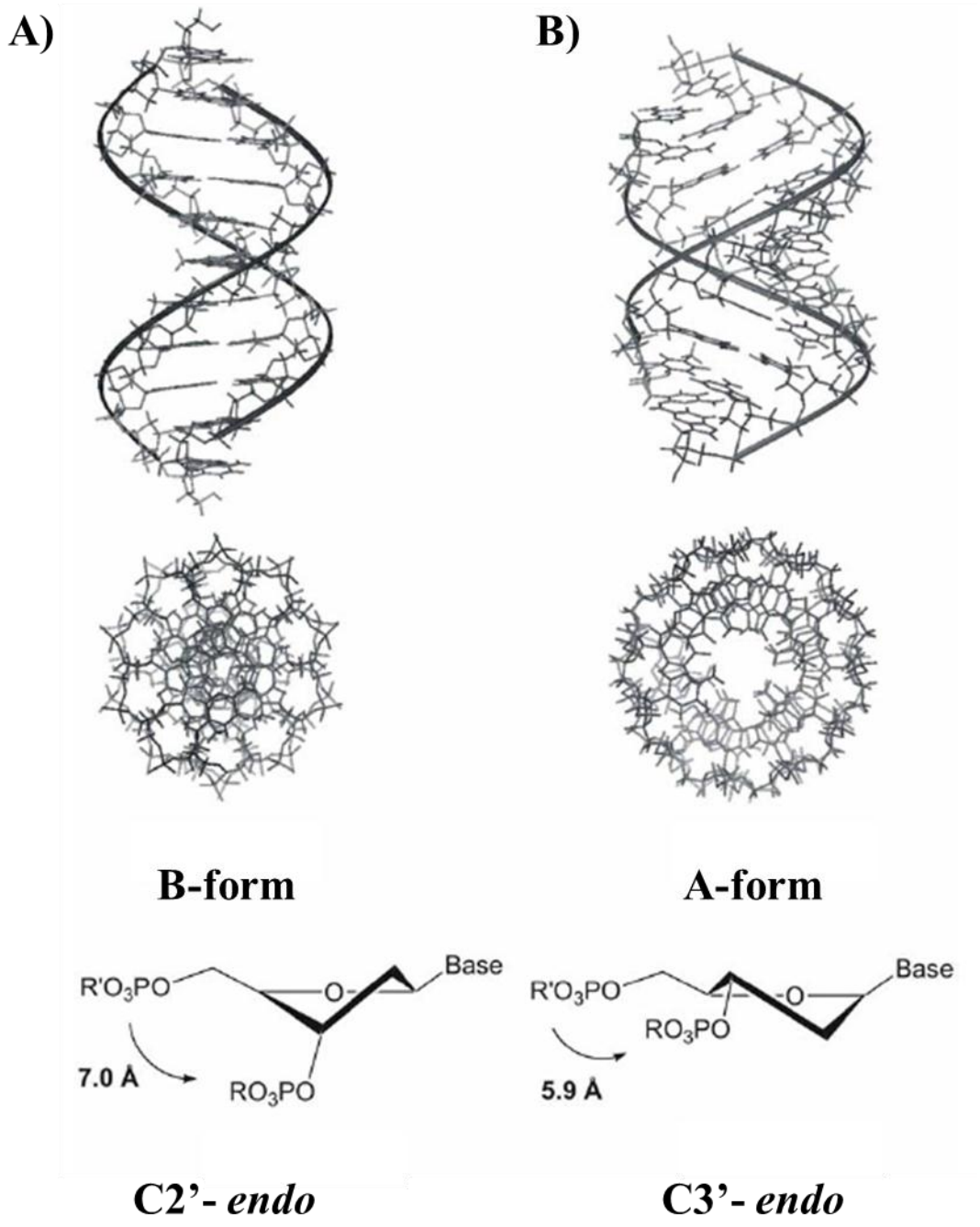


Figure 1.4 A) side view, top view, and C2'-*endo* sugar pucker of B-form DNA. **B**) side view, top view, and C3'-*endo* sugar pucker of A-form DNA.[28]

Although B-form DNA is the most common physiological form, A or Z form DNA might also be present based upon a range of conditions that include the solvent, salt conditions, nucleobase sequence, as well as topological environments resulting from bound ligands.[29,30] A-DNA is more compact than B-form, with 11 bp per turn and a 2.54 Å helical rise, although it is also right-handed in the helical twist.[27] Compared to B-DNA, A-form DNA contains a fairly deep and narrow major groove, with a wider and shallower minor groove and an overall C3'-*endo* sugar pucker. Z-DNA is different altogether, with a left-handed helical twist and 12 bp per turn.[27]

DNA methylation is another aspect of DNA structure that can affect recognition. DNA methylation occurs at the 5-position of cytosine, located in the major groove, in a consecutive CG step of eukaryotic cells via methyltransferase enzymes (**Figure 1.5**). DNA methylation is thought to contribute to the epigenetic regulation of genes.[31] This aspect of structural modification should be considered when designing molecules with likely interactions in the major groove of DNA.

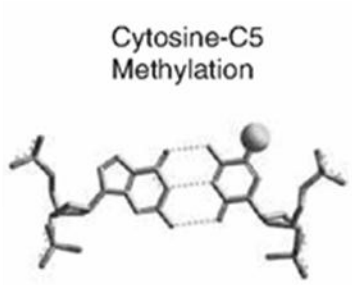


Figure 1.5 Location of DNA methylation shown as space-filled methyl at the 5-position on cytosine.[31]

One further consideration for designing small molecules that bind DNA is the actual DNA presentation within a cell. As opposed to free duplex DNA in solution, cellular DNA often resides in a highly compacted chromatin system. First, approximately 140 bp of DNA are wrapped around a core histone octamer protein assembly, creating the basic nucleosome complex that is disc-like in shape (**Figure 1.6**). With some DNA excluded from the nucleosome acting as spacer DNA, nucleosomes will then pack together to create a filament of approximately 100 Å in diameter. This filament is then coiled on itself into a spring shaped superstructure called a fiber, which is about 300 Å in diameter.[27] A cartoon representation of higher-order packing of DNA to create the fiber can be seen in **Figure 1.7**. It is then possible for even higher-order arrangements by attaching these fibers to other protein cores depending on the particular phase of the cell cycle. Ultimately, these highly compacted structures of cellular DNA should be considered when designing small molecules that bind DNA, due to the difference in accessibility versus standard duplex DNA in solution. Because cellular DNA does not necessarily favor a standard solution-phase equilibrium with small molecules (presumably due to limited accessibility of the packing structures), an extremely long off-rate would be a highly desirable characteristic of such a ligand, provided that once bound to the DNA, it would ideally stay in a bound state long enough to block transcription at the particular site.

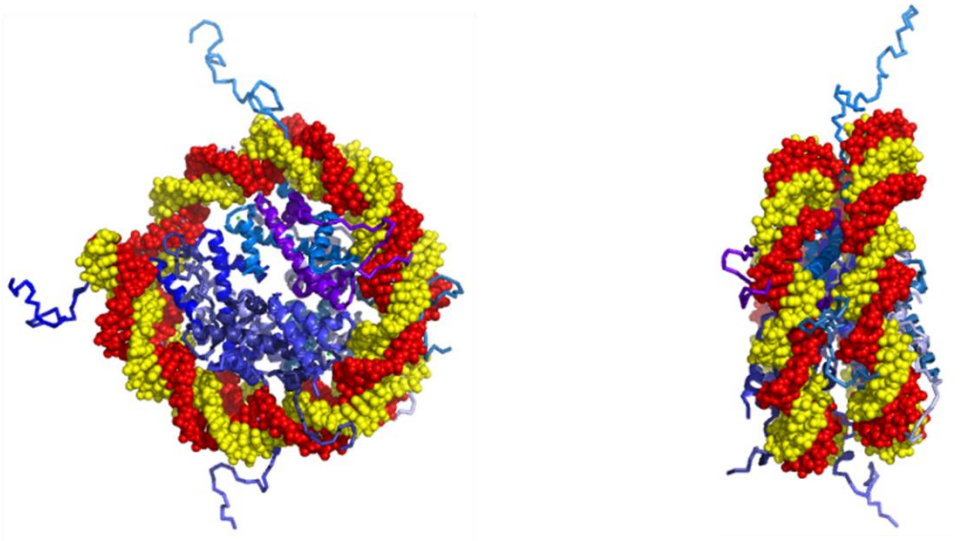


Figure 1.6 Two different views of an X-ray crystal structure for the nucleosome core particle NCP147[32] (PDB ID: PD0287) with the histone proteins shown as different shades of blue, and the DNA with red and yellow strands. The two views were produced by The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.

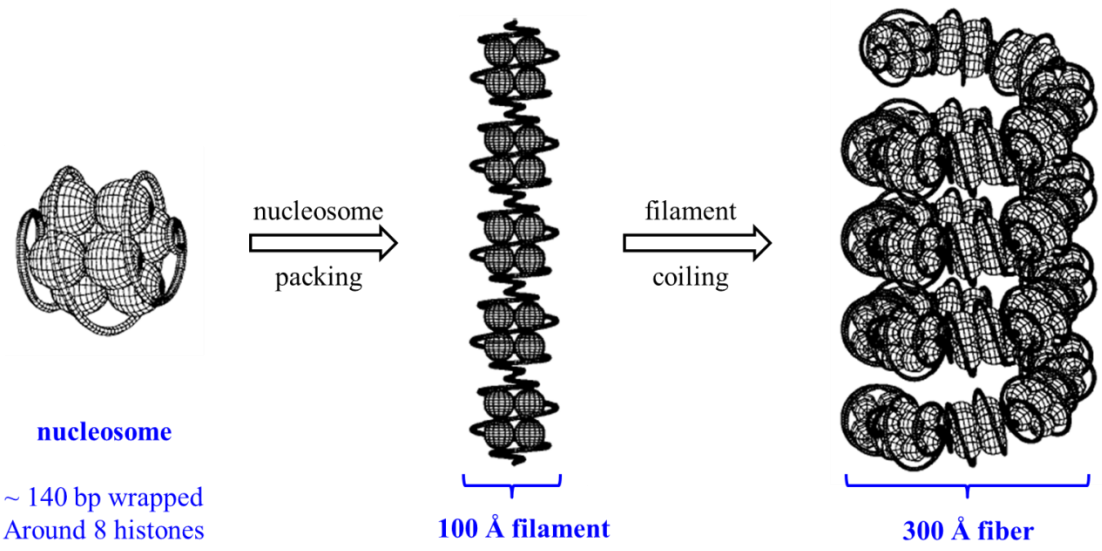


Figure 1.7 Cartoon of higher-order packing of DNA.

1.3 Binding Interactions in DNA Recognition

DNA is a biopolymer that experiences a wide range of interactions both in solution as well as *in vivo*. The phosphate backbone of the DNA is a polyanion, with each phosphate being negatively charged. Too many negative charges lined up is an unfavorable interaction. However, cations are able to ion-pair with the phosphate backbone, thereby reducing the repulsive forces experienced between phosphate groups. Manning's polyelectrolyte theory was one of the first attempts to describe the ion-pairing, or condensation, of cations to polyanions such as DNA.[33] A few of the most common electrostatic interactions between the DNA backbone and cations occur with Na^+ and Mg^{2+}. Record *et al.* have observed that approximately 0.88 Na^+ ions are condensed per phosphate, and that there are approximately 0.37 – 0.50 Mg^{2+} ions condensed per phosphate along the DNA backbone.[34] These electrostatic interactions serve to stabilize the DNA polyanion in solution, so that the overall charge density is close to unity for solubility purposes.[33]

Proteins or small molecules which bind DNA are often charged species, and are initially attracted to the DNA in solution via electrostatic forces. While these initial electrostatic interactions are not necessarily specific due to the even distribution of charges on the phosphate backbone, they do provide a platform by which the ligand can search the localized DNA topology for more favorable binding conformations. Upon the ligand binding DNA, the ion-paired cations are then released into solution, which provides a favorable entropic component to the binding event.[34]

Hydrogen bonding between the nucleobases, as well as potential hydrogen bond interactions with the edges of the DNA bases, is critical to both DNA structure and recognition. The typical Watson-Crick base pairs result from a favorable geometry between nucleobases on opposite strands of the DNA. One mismatch in the base pairing

can result in bulges of the double helix. If the mismatch is not repaired, mutagenesis of the particular gene can occur, which ultimately could result in diseases such as colorectal cancer.[35,36]

The edges of the nucleobases within the DNA grooves contain many potential opportunities for hydrogen bonding. **Figure 1.3** above shows the atoms that can act as hydrogen bond donors in red, with atoms that can accept a hydrogen bond shown in blue. The importance of this characteristic is that the patterns of donor/acceptor ability of DNA containing G•C or C•G base pairs can produce a sequence specific readout in the major groove, which would be further distinguished in the event of cytosine methylation. While the donor/acceptor readout of the minor groove is symmetrical, work from the Dervan lab has pioneered a chemical key-code that uses polyamides to distinguish between any combination of base pairs within the minor groove based solely on hydrogen bonding and geometry configurations.[37]

In addition to electrostatic and hydrogen bonding interactions, hydrophobic interactions also play a crucial role in DNA chemistry. It is largely the favorable hydrophobic stacking interactions of the aromatic nucleobases that stabilize the formation of the double helix.[27] These interactions arise from the mixing of aromatic orbitals and nonpolar surfaces of molecules rearranging to maximize their interactions with each other in such a way as to minimize the interactions with polar surfaces such as water.[38] Many binding events to DNA involve a hydrophobic component, or more specifically, maintain a desolvation driving force as to reduce the interactions with water. Interactions such as DNA intercalation play strongly to hydrophobic base-stacking, while groove binding interactions can exhibit many favorable Van der Waals interactions between the wall of the minor groove and small molecules containing aliphatic linkers as well as entropic contributions from the release of bound water molecules.

With all of these different forces affecting the DNA double helix in solution, DNA is not static in solution. Quite to the contrary, DNA is undergoing constant dynamic fluctuations, some of which are necessary to its biological role. For instance, several laboratories have studied the DNA dynamics, or "breathing", in detail.[39,40,41,42] It has been observed that for B-form DNA, A•T or T•A base pairs open at a rate of 1 ± 10 ms, while G•C and C•G base pairs breathe at a rate of 5 ± 50 ms.[41] The slower rate for GC/CG base pairing is an expected outcome with one extra hydrogen bond between the nucleobases. For both A•T/T•A and G•C/C•G base pairs, the actual lifetime of the open base is on the order of nanoseconds, presumably due to the unfavorable disruption of the hydrophobic base stacking interactions. The importance of the dynamic nature of DNA in solution has been implicated due to the recognition of DNA breathing by several enzymes. For instance, Cao *et al.* have demonstrated that some DNA base flipping is crucial to the ability of uracil DNA glycosylase to replace uracil nucleobases inserted into DNA.[43] It has also been observed that upon protein binding, some DNA bases can flip out to promote specific sequence recognition, and ultimately facilitate the melting process leading to DNA replication and transcription.[44,45,46] Recent work has focused on trying to study the rapid breathing interaction, and perhaps indicate the ability of dynamic DNA breathing at junctions of ssDNA and dsDNA that are potentially capable of recruiting enzymes to these localities.[47]

Taken together, electrostatic, hydrogen bonding, and hydrophobic interactions all play a role in the overall favorability of ligands binding to DNA. So while some interactions might contribute more to overall specificity than others, it is no doubt a combination of the many possible interactions, coupled with the dynamic nature of DNA, that lead to the end result of specific molecular recognition.

1.4 SMALL MOLECULES THAT BIND DNA

As discussed above, there are many potential biological targets for molecules which could act as useful therapeutics. Limiting the field to molecules that bind DNA still offers a near limitless array of possible individual candidates. Even distinguishing between molecules that may be defined as "small" can be a vague task. This section will consider the potential for molecules to bind DNA which are significantly smaller than most proteins, which eliminates a vast category of candidates, including proteins that are well known to bind DNA, often through specific contacts in the major groove.[48,49] There are still numerous examples of small molecule binding motifs towards DNA, for instance, systems where small-molecule sapphyrin derivatives can bind DNA through the chelation of the phosphate groups of the DNA backbone, and may be used to transport interesting functional groups to the double helix.[50,51,52,53] Some porphyrins are also capable of intercalating between base pairs of the DNA.[54] Other types of DNA binding small molecules such as cisplatin or enediynes focus on covalently altering or oxidatively damaging the DNA, thereby disrupting transcription which can ultimately lead to cell death.[55,56,57] While there are many more potential systems of small-molecule-DNA interactions, this section will focus on examples of systems involving groove binding and intercalation, two binding modes closely associated with the work contained in this dissertation.

1.4.1 Minor Groove Binders

The minor groove binding mode is among the most studied, and well understood, topologies of small molecules that bind to DNA. One common aspect of such small molecules is there characteristic crescent shape, which closely matches the curvature of the DNA minor groove. Minor groove binders are often positively charged, and contain

nitrogen heterocycles connected by amide bonds, both of which typically present multiple hydrogen bond donating sites. The positive charges often provide for the initial electrostatic attraction to the DNA. The shape and hydrogen bond donating capacity for minor groove binders allows for a strong desolvation driving force, releasing water molecules from the groove, as well as close Van der Waals contacts with the groove walls and optimal geometry for hydrogen bonding to the nitrogenous bases.[58] Several minor groove binding small molecules can be seen in **Figure 1.8**.

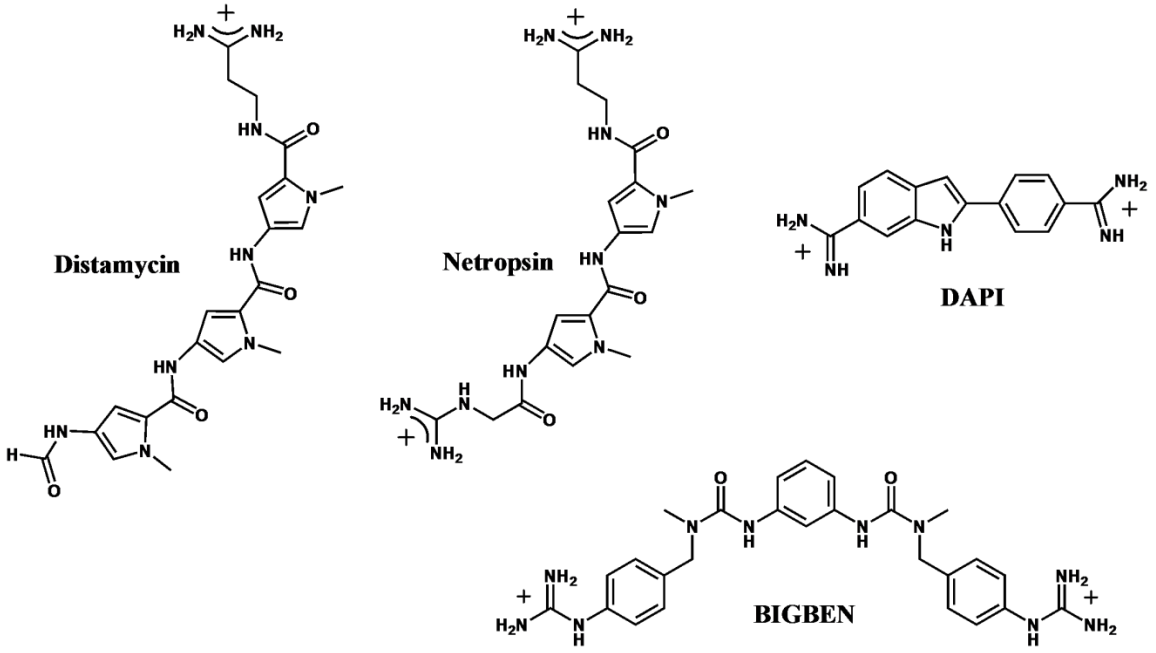


Figure 1.8 Chemical structures of a variety of minor groove binding ligands.

Recall that from Figure 1.3 above, the minor groove edge of the nitrogenous bases contain four potential hydrogen bond accepting sites, with only one hydrogen bond

donating site with a G•C or C•G base pair. While most minor groove binders have traditionally preferred A,T rich DNA sequences, the hydrogen bonding characteristics in the minor groove have become crucial to the eventual design of a "key code" to read a diverse set of unique sequences containing A•T/T•A or G•C/C•G base pairs as will be discussed in more detail below.[37] The ability for small molecules to bind DNA has demonstrated some potential therapeutic usefulness by possibly blocking transcription or inhibiting the DNA topoisomerase enzymes.

For instance, the natural products netropsin and distamycin from *Streptomyces netropsis*[59] and *distallicus*[60], respectively, are very well characterized systems of minor groove binding molecules that have shown interesting therapeutic properties.[61,62] Netropsin and distamycin are composed of two and three N-methylpyrrole (Py) units, respectively, and both bind A,T rich DNA with similar association constants on the order of ~10^6 M^{-1}.

There are many synthetic minor groove binding analogues such as DAPI,[63] BIGBEN,[6] Hoechst 33258,[64] and several other lexitropsin derivatives.[62,65] Perhaps the most successful example of synthetic minor groove binders are the polyamides developed by the Dervan Lab. The first crystal structure of a 1:1, netropsin- A,T rich DNA complex, detailed the significance of the hydrogen bonding contacts in the minor groove, and it was proposed by Dickerson and Lown, that exchanging a Py unit for an N-methylimidazole unit (Im) in netropsin might facilitate recognition of a G•C or C•G pair.[58,66] This hypothesis, coupled with a discovery of antiparallel binding of distamycin to a sequence of A,T DNA in a 2:1 complex by the Wemmer group,[67] has enabled the Dervan lab to develop a string of sequence specific polyamides with comparable DNA affinities to those of proteins (K_A ~ 10^8 – 10^9 M^{-1}).[68,69,70,71,72] At the time, they had discovered they could pair Py and Im groups to distinguish between three of the four

possible base pair sequences as shown in **Table 1.1**. Antiparallel combinations of Im/Py and Py/Im could recognize G•C and C•G base pairs, respectively, while Py/Py was selective for both A•T and T•A base pairs.

Antiparallel Pair	G•C	C•G	A•T	T•A
Im/Py	+	-	-	-
Py/Im	-	+	-	-
Py/Py	-	-	+	+

Table 1.1 Minor groove base pairing specificity for antiparallel polyamides. A (+) indicates specific recognition and (–) indicates no specific recognition.

It was at this point, that structure based design through modeling led to the design, synthesis, and confirmation that an N-methylhydroxypyrrole (Hp) unit could be used in conjunction with an antiparallel Py unit to differentiate between A•T and T•A base pairs.[37] The results led to the first synthetic "key code" for designing small molecules that could potentially recognize any designed sequence with relatively high affinity. The new base pairing code can be seen in **Table 1.2**, and the structure of the $(ImHpPyPy)_2$•5'-CCAGTACTGG-3' complex that confirmed the new recognition is shown in **Figure 1.9**.[73] Using the four base pair key code to specific binding, many new developments have come from minor groove binding polyamides, including the ability to inhibit gene expression,[15,22] and even a few studies indicating that polyamides may be capable of up-

regulating gene expression as well.[74,75] Ultimately, the Dervan lab postulated that because of the sequence dependent microstructure of DNA, they could potentially bind ~ 50% of the DNA sites on any promoter, thereby targeting most of the relevant transcription factors of therapeutic interest.[76] Also important in the context of this work is the discovery that some of their synthetic derivatives are capable of binding to nucleosomal DNA. This is accomplished where the minor groove is facing away from the histone octamer, and also noted was that the nucleosome remained in a folded state upon polyamide binding.[77]

Antiparallel Pair	G•C	C•G	A•T	T•A
Im/Py	+	-	-	-
Py/Im	-	+	-	-
Py/Hp	-	-	+	-
Hp/Py	-	-	-	+

Table 1.2 New base pairing "key code" utilizing antiparallel pairing of Hp with Py to discriminate between A•T and T•A base pairs. A (+) indicates specific recognition and (–) indicates no specific recognition.

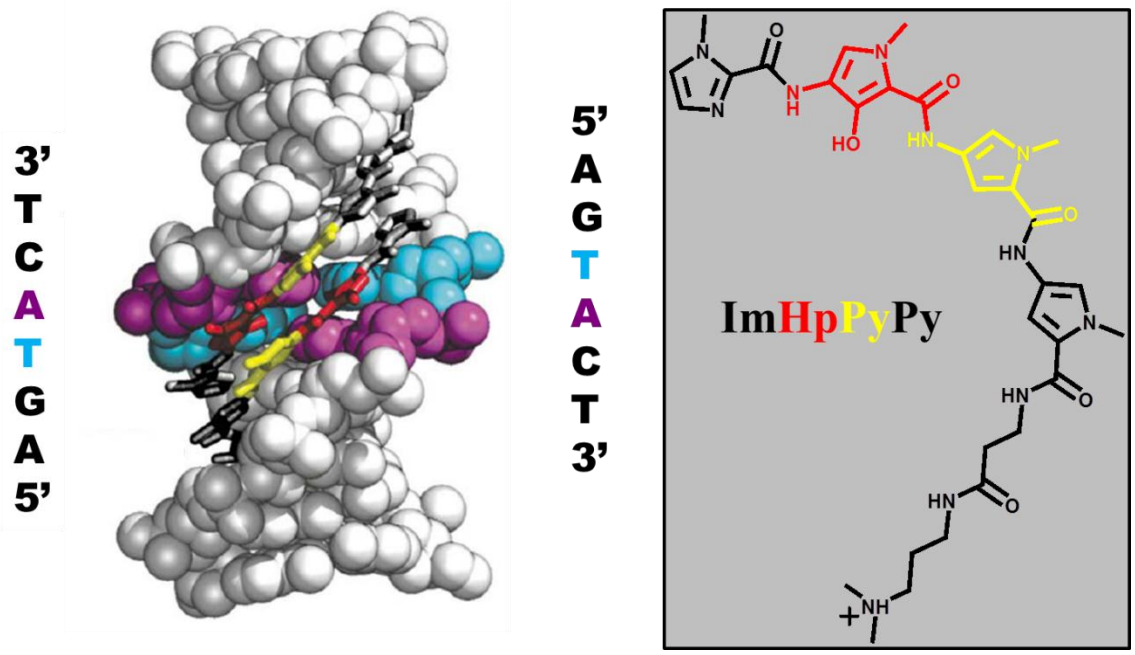


Figure 1.9 Left: Space-filled model showing the 2:1 complex of $(ImHpPyPy)_2$•5’-CCAGTACTGG-3’. The adenosine nucleobase is shown in purple, and the thymidine shown in blue. The Hp and Py pair are shown in red and yellow, respectively.[73] Right: chemical structure of ImHpPyPy color coded identical to the NMR model structure.

1.4.2 Triplex Forming Oligonucleotides

Triple helix (triplex) forming oligonucleotides (TFO) are composed of a third strand of DNA that binds the major groove of dsDNA via Hoogsteen and reverse Hoogsteen base pairing. Triplex oligonucleotides should quite possibly be disqualified from being labeled as "small molecules", however, their ability for sequence specific recognition at relatively high affinities ($K_D \sim 10^{-7} - 10^{-9}$ M) through the major groove and potential uses for gene regulation merits an appropriate discussion.[78]

The first observation for triplex formation was by Felsenfeld *et al*. in 1957, although, the first example of sequence specific targeting of DNA occurred much later by Moser and Dervan.[79,80] There are two basic forms of triplex molecules, those of a pyrimidine binding motif and those of a purine binding motif, both of which recognize stretches of purine rich dsDNA through the major groove. The pyrimidine motif binds to the purine-rich DNA strand in a parallel fashion via Hoogsteen base pairs at less than physiological pH levels. The sequence specificity is for the TFO thymine (T) to bind the adenine (A) of an A•T pair, while a protonated cytosine (C^+) binds to the guanine (G) of a G•C pair. The purine motif specificity is shown through the third strand binding of A to an A•T pair, G to a G•C pair, and T to an A•T pair. This purine TFO exhibits an antiparallel binding to the purine-rich strand of DNA, and accomplishes this through reverse Hoogsteen base pairing ("reverse" due to the antiparallel recognition).[81] Representations of the triplex binding motifs and directionalities can be seen in **Figure 1.10**.

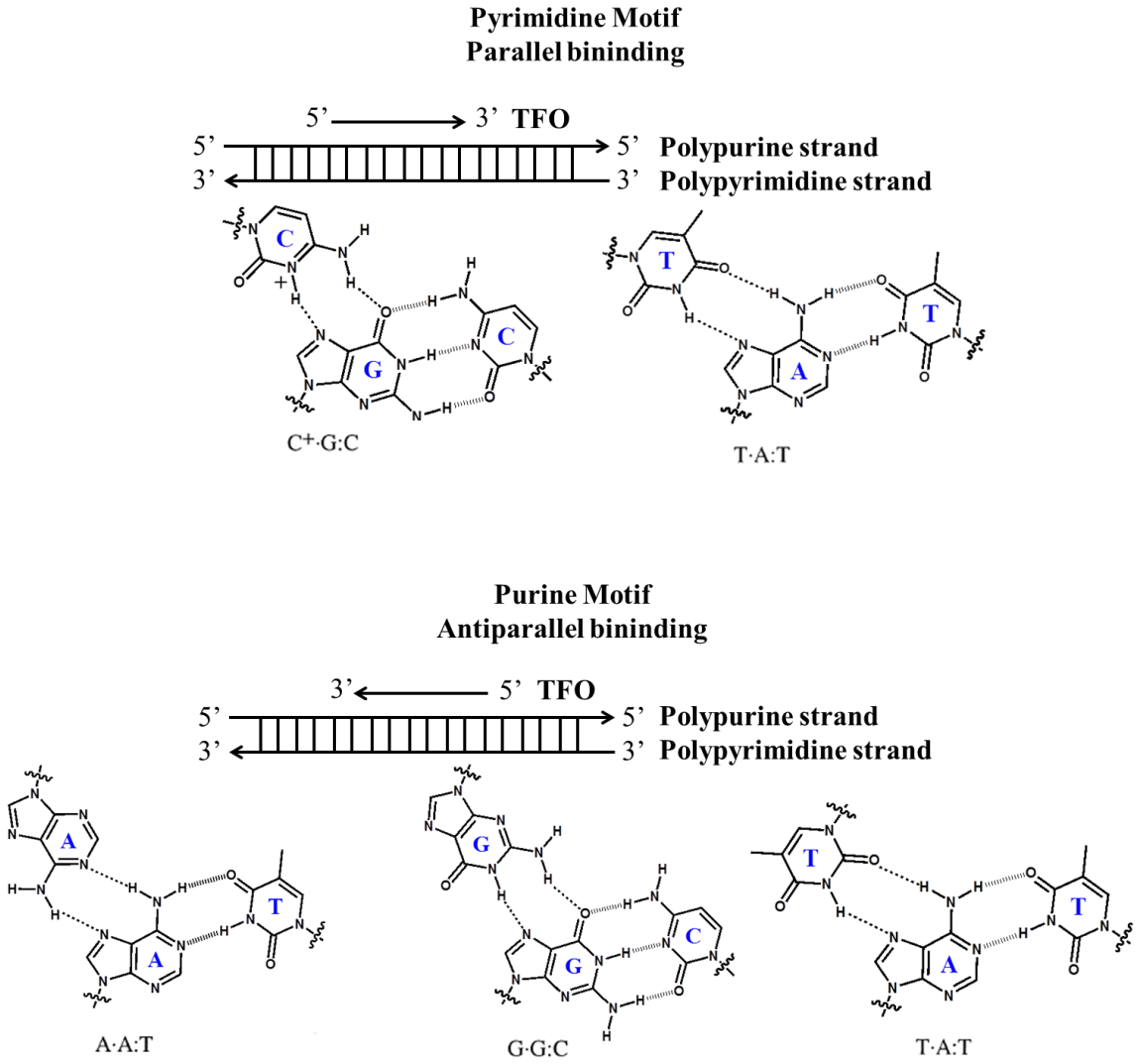


Figure 1.10 Pyrimidine and purine binding motifs showing (| | |) Watson-Crick, (···) Hoogsteen, and (···) reverse Hoogsteen hydrogen bonding of nucleobases. The Directionality of triplex binding oligonucleotide (TFO) is also indicated for each motif as labeled.

Triplex binding motifs can offer several potential therapeutic applications. Recall that their binding is through the major groove of DNA, which could potentially inhibit many proteins because of their tendency towards major groove recognition. As mentioned above, one of the first examples of triplex targeting of DNA actually attached an Fe·EDTA cleaving moiety to the TFO, and demonstrated sequence specific cleavage of the Watson-Crick dsDNA.[80] Used in such a way, this approach demonstrated a concept that TFOs might act as "molecular scissors", to specifically cleave desired sequences of DNA. To date, the therapeutic benefits of TFOs have largely centered on site-directed mutagenesis, TFO induced recombination, and inhibition of transcription by binding to gene promoter regions of DNA.[81]

Outside of the typical cellular uptake and delivery problems experienced by many DNA binding molecules, there are still many limitations to triplex oligonucleotides. One problem is that a third anionic backbone in such close proximity to the other two Watson-Crick backbones can be destabilizing, which is why divalent cations like Mg^{2+} serve to stabilize triplex formation.[81] Other key issues can range from requiring non-physiological salt or low pH conditions to unstable complexes resulting from a mismatch in the purine-rich strand of DNA.[82] There have been many efforts to circumvent these issues, such as the development of synthetic analogue bases like 2-aminopyridine, which can reduce the need for lower pH systems.[83] New approaches to triplex formation focus on the use of aminoglycosides, such as neomycin, to further stabilize triplex formation.[84]

1.4.3 Peptide Nucleic Acids

Peptide nucleic acids (PNA) are a group of sequence specific molecules that are essentially DNA analogs. However, the sugar-phosphate backbone has been replaced

with a neutral peptidic backbone (**Figure 1.11**). The first observation of PNA binding to dsDNA revealed that the neutral backbone alleviates anionic repulsion so much that it had increased affinity over that of the complement DNA strand.[85] A particular benefit to the original design is that the peptide backbone is biologically stable and oligomers of PNAs are easily accessible.[31] PNAs contain the same nucleobases as DNA and thereby offer the ability for targeting specific DNA sequences, through triplex formation or strand displacement, with the potential to regulate gene expression.

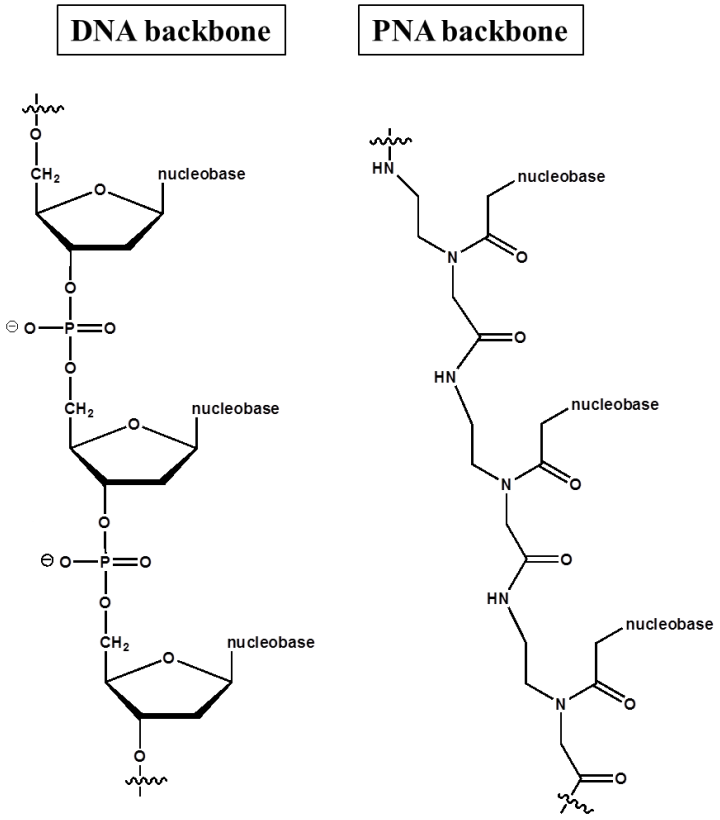


Figure 1.11 Chemical structure comparison of DNA and PNA backbones as labeled.

There are several observed models of PNA-dsDNA complex formation, such as duplex invasion,[86] triplex formation,[85,87] triplex invasion,[88] and double duplex invasion[89] (**Figure 1.12**). The PNA triplex invasion model (PNA_2-DNA) is the most common complex and offers higher stability than that of typical dsDNA, with observed thermal melting stabilities of approximately 70°C. The work associated with this finding concludes that the rate of PNA formation, therefore, is a kinetically controlled process.[90] Not surprisingly, complexed PNA rate of formation is inversely related to the concentration of cations such as Na^+, and even more so for Mg^{2+}, because of the stabilizing effects that the cations provide to the anionic double helix.

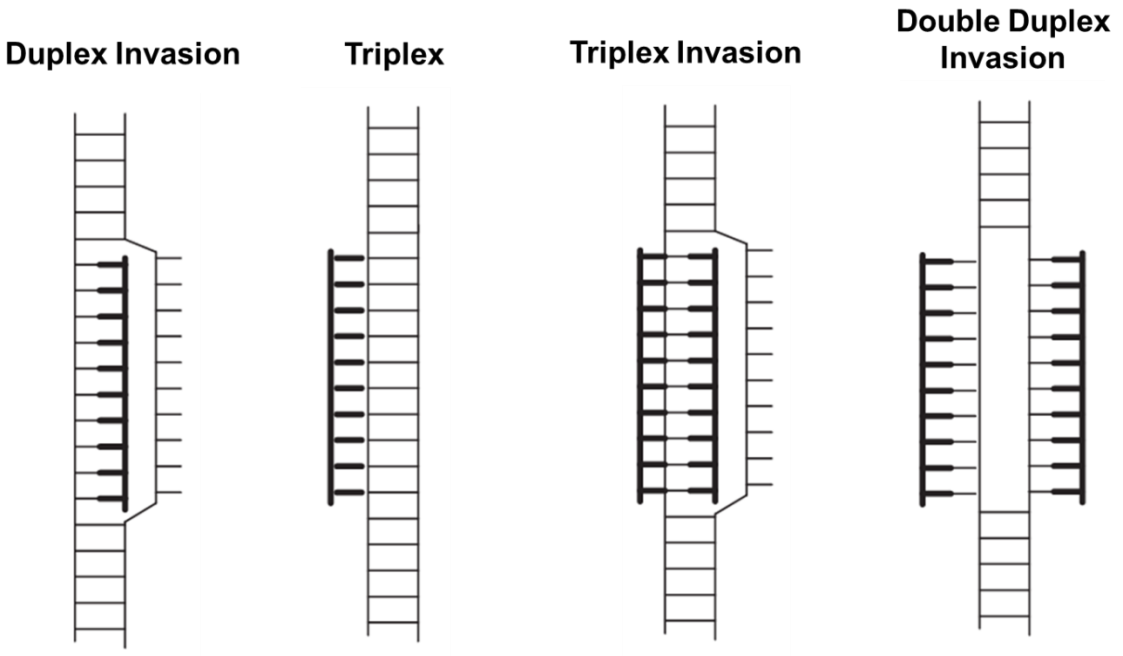


Figure 1.12 Cartoon models of different PNA-dsDNA complexes as labeled. PNA strands indicated in bold.

Not only do PNAs typically show binding constants of $K_D \sim 10^{-9}$ M, but they also have produced interesting therapeutic results in uses such as gene targeting, transcription interference, targeted gene repair, antibacterial capability, and perhaps the most interesting potential for RNA interference.[31,91] As expected, these promising uses do not come without limitations. One of the largest drawbacks of PNAs is that their neutral backbone, which eliminates repulsion, also serves to decrease solubility.[92] This is in addition to poor cellular uptake and very poor physiological pharmacokinetics due to the rapid excretion by the kidneys.[93] Notwithstanding, PNAs still offer relative simplicity of design, providing for a continued interest in the field. This same simplicity of a peptidic backbone that can encode genetic information, has also led to experiments testing the possibility of PNAs as a possible developmental accomplice in the origin of life.[94]

1.4.4 Intercalators

Intercalation is a well-studied binding mode for small-molecules to DNA, first observed by Lerman in 1961.[95] Intercalation is described by a planar aromatic (often heteroaromatic) ring that is inserted between adjacent base pairs of the double helix. The process of intercalation causes the nucleobases above and below to vertically separate, an accommodation made possible by the partial lengthening and unwinding of the helix. This results in a slight increase in the intrinsic viscosity of the helix, causing an overall stiffening of the DNA.[96] Crothers observed that the number of typical monointercalators bound to DNA per base pair was approximately 0.5, introducing his nearest "neighbor-exclusion principle".[97] This principle states that the two sites directly adjacent on either side of an intercalation site will remain unoccupied as a result of the induced helical conformations to allow for intercalation (a C3'-*endo* sugar pucker can be adopted in the

pyrimidine deoxyribose ring). Intercalators have been well characterized since their discovery, and have had many clinically relevant applications.[98,99] For these reasons, intercalators continue to be part of an intriguing research field with the potential at discovering new intercalation systems that might show potential therapeutic properties.

There are a variety of compounds that can be classified as intercalators. The first class is composed of monointercalators, or molecules with one intercalating moiety. This class includes many examples, a few of which are proflavine, ethidium, daunomycin (daunorubicin), Adriamycin (doxorubicin), and actinomycin D (**Figure 1.13**). Monointercalators are less sequence specific and more prone to search the DNA for suitable nucleobase steps. Müller and Crothers observed that the preference of many intercalators for G•C base pairs increases as the absorbance maximum of the particular chromophore (intercalator) shifts to longer wavelengths. They proposed that this was presumably due to an increased polarization of the intercalator by the more polar G•C base pair.[100] Sharples and Brown investigated these results, and using a multi-parameter regression, confirmed the previous observation in that there was a strong correlation between charge transfer affinity of an intercalator and the respective preference for G•C base pairs in DNA.[101] These results suggest that electron deficient intercalating units of the proper size should produce some specificity for GC rich DNA, and indeed this has been the case for many intercalators.[102]

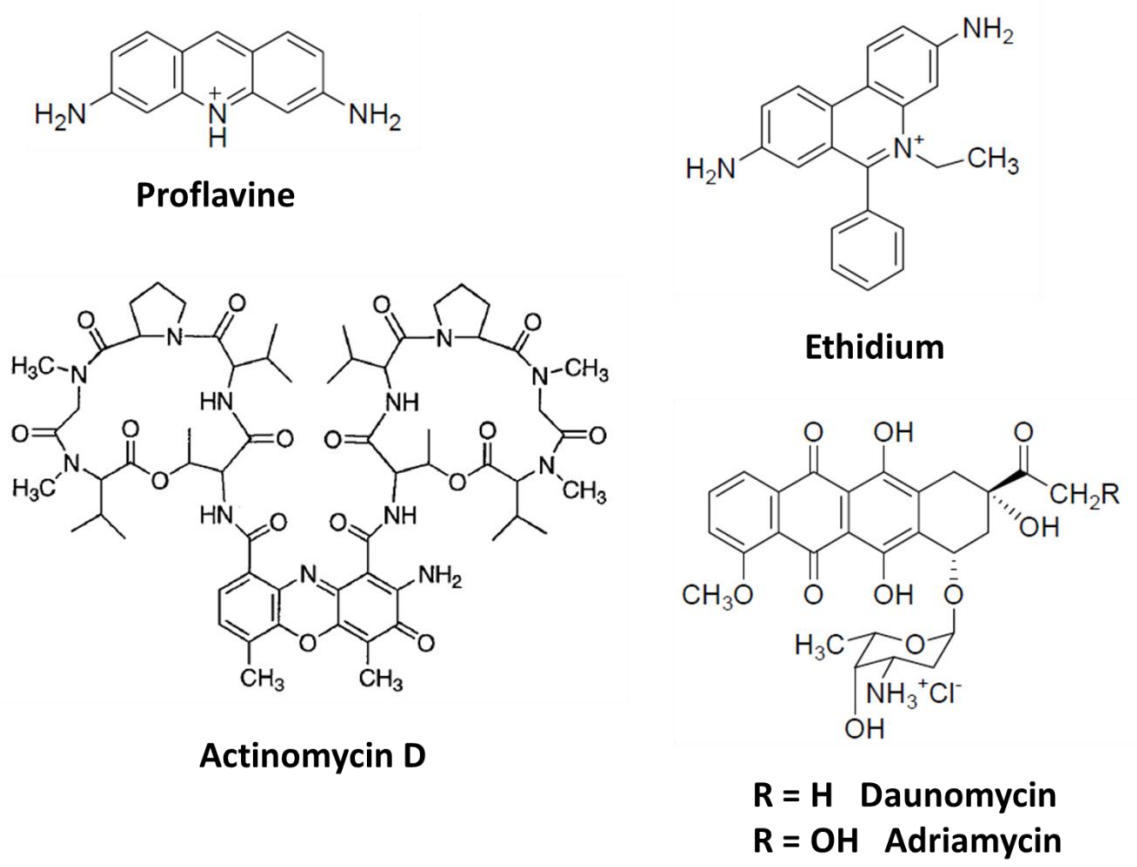


Figure 1.13 Chemical structures of some monointercalators.

For instance, ethidium and proflavine show a slight preference for pyrimidine-(3'-5')-purine steps, more specifically, CpG steps.[103] Ethidium has found its main use as a fluorescent stain of DNA, while proflavine has had some past clinical use as a topical antiseptic agent. Ethidium and proflavine are relative simple monointercalators, both offering association constants of $K_A \sim 10^4$ M^{-1}.[104,105,106] Monointercalators such as daunomycin and actinomycin D are natural products isolated from *Streptomyces coeruleorubidus* and *Streptomyces antibioticus*, respectively. They are structurally more

complex, which allows for more specific recognition elements upon binding DNA. This increased recognition from the extra functional groups results in an increase in the overall association constant, as opposed to simple monointercalators, which are on the order of $K_A \sim 10^6 - 10^7$ M^{-1}.[11,107]

Daunomycin prefers (A/T)pGpC or (A/T)pCpG sites, and contains a daunosamine ring that resides in the minor groove of DNA. A crystal structure shows that the N-3' of the amino sugar ring forms a hydrogen bond with the O2 of the adjacent thymine. It has been suggested that this electrostatic interaction helps position the sugar in a way that significant Van der Waals contacts occur between the intercalator and the DNA.[108] These extra interactions with DNA above the favorable intercalation interactions allow daunomycin and its analogs to be used as some of the most valuable drugs currently approved by the FDA, seeing application in the treatments of a variety of cancers ranging from ovarian and breast cancers to acute leukemias.[99] However, their effectiveness does not come without restrictions, as they have been shown to be cardio-toxic, likely the result of their interference of the topoisomerase II enzyme.[98]

Like daunomycin, actinomycin has been extensively studied and has been in clinical use since about 1943. The preferred recognition sequence is a GpC step, with the two depsipeptide (so named for the replacement of an amide bond by an ester) rings making a near symmetrical complex that provides for many favorable Van der Waals contacts in the minor groove of the DNA.[109] One particular feature likely responsible for the biological activity of the actinomycin-DNA complex is a characteristically slow dissociation rate, with a $k_d \sim 0.01$ sec^{-1}. Müller and Crothers postulated that the basis for the slow dissociation is structural, with a slow reversal of the conformational change of the depsipeptide rings.[11] Taken together, the extra functionality contained in daunomycin and actinomycin D provide for more specific contacts and higher binding affinities.

These characteristics, as well as slower off-rates, seem to provide for more practical clinical uses for the treatment of a variety of diseases.[99]

Metallo-intercalators, discovered by Lippard in the early 70's,[110] are another class of intercalating moiety composed of a kinetically inert d^6 octahedral or d^8 square planar transition metal with coordinated ligands attached.[111] It is actually the coordinated ligands that are responsible for the intercalation process, with the most common being the phi and dppz ligands coordinated to either ruthenium or rhodium metal centers (**Figure 1.14**). One interesting characteristic of metallo-intercalators is their ability to distinguish the chiral environment through different modes of binding based upon the individual enantiomer of the metallo-intercalator being examined. For example, the Barton lab has discovered the ability for a Δ enantiomer of $[Ru(phen)_3]^{2+}$ to preferentially intercalate via the major groove of DNA, while the Λ-enantiomer prefers to bind DNA through a minor groove binding mode.[112] It is the photochemical and photophysical properties of metallo-intercalators that have proven the most interesting. Ruthenium metallo-intercalators have found their primary purpose as molecular "light switches", which are capable of detecting small amounts of DNA in solution. Rhodium metallo-intercalators however, have been used as UV activated probes, which can cleave the DNA backbone upon excitation.[113] It has been observed that the metallo-intercalators show a slight preference for A,T-rich DNA, and have binding constants on the order of $K_A \sim 10^4 - 10^5$ M^{-1}.[111,114]

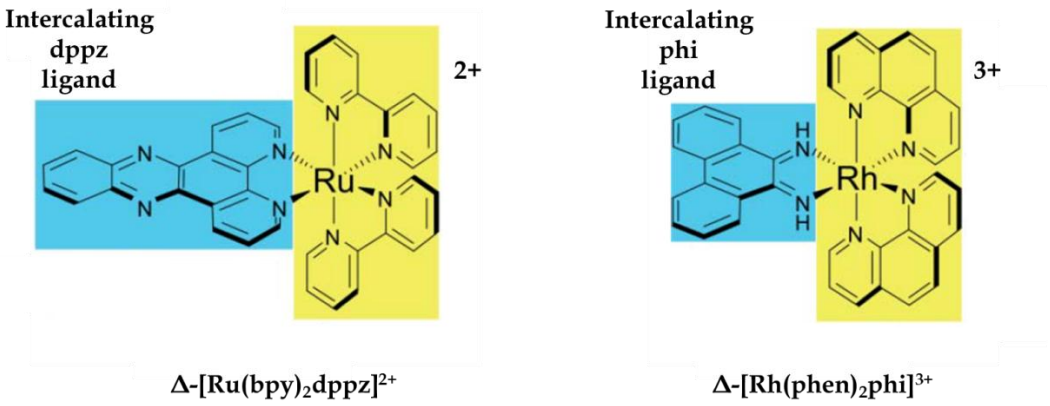
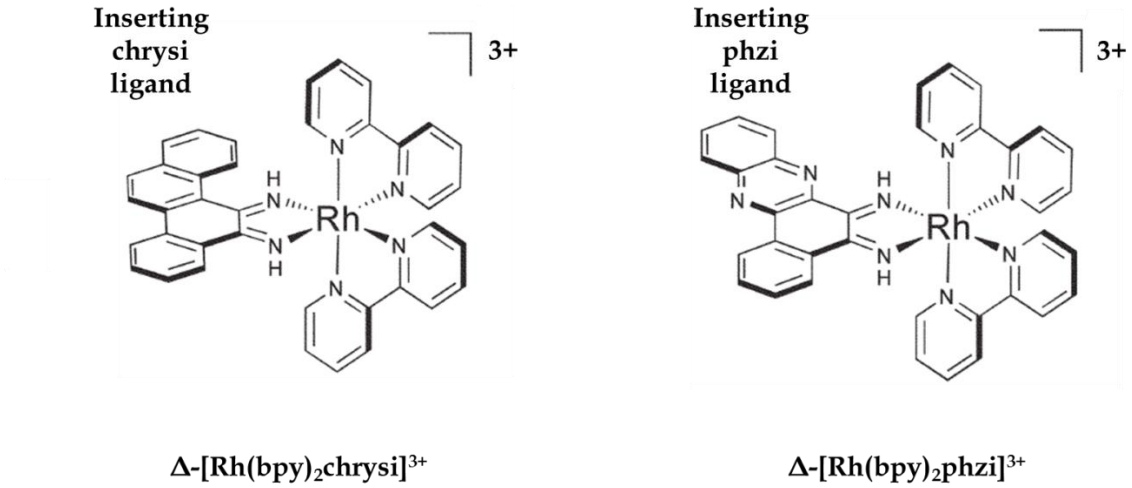


Figure 1.14 Δ-Ruthenium and Δ-rhodium metallo-intercalators with the dppz and phi intercalating ligands labeled and in blue, respectively. Ancillary ligands are shown highlighted in yellow.[111]

Slightly different than the intercalators, are metallo-insertors, which have shown promise as diagnostic probes capable of detecting mismatched bases (**Figure 1.15**). These molecules have a coordinating ligand that is too large for typical intercalation, thereby causing one of the nucleobases to flip out upon insertion. This feature makes metallo-insertors extremely capable at detecting mismatched base pairs in DNA, due to the displaced nucleobase from non-standard Watson-Crick base pairing. While not ultimately a tool for treating cancer, these diagnostic properties are very useful for the detection of potentially mutagenic genes due to their cleavage capabilities localized at the mismatched site.[115]



Δ-[Rh(bpy)$_2$chrysi]$^{3+}$ **Δ-[Rh(bpy)$_2$phzi]$^{3+}$**

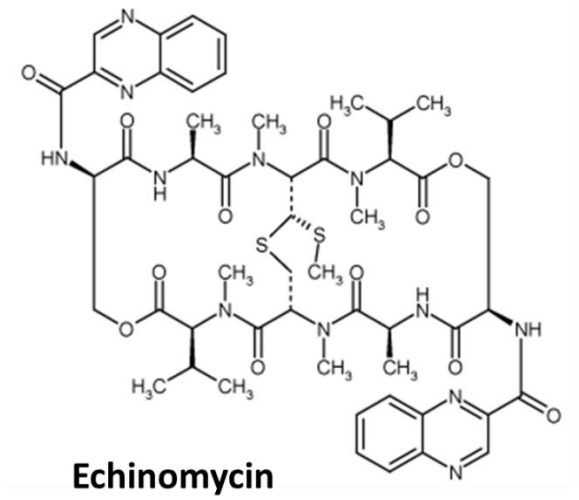
Figure 1.15 Chemical structures of two Δ-rhodium metallo-insertors.[111]

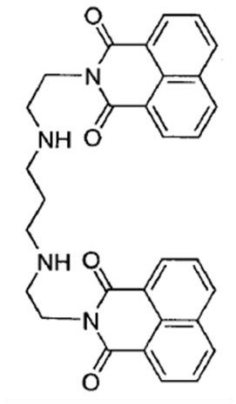
1.4.5 Polyintercalators

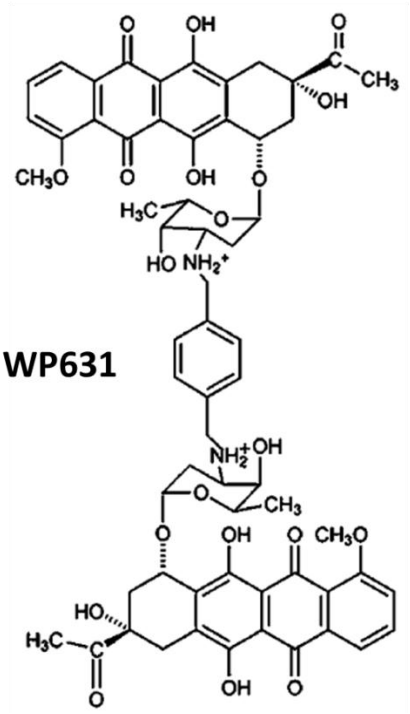
As discussed above, a 16 bp binding site is thought to be the correct length to specify a unique single sequence within the human haploid genome. Monointercalation only involves the recognition of between 2 – 4 bp per intercalator. One particular strategy to recognize longer sequences of DNA is by connecting two or more of the individual monointercalation units, creating bisintercalators and polyintercalators, respectively. In addition to recognizing longer sequences, connecting intercalating units should allow for the binding constants to theoretically be squared without any consideration of affinity between the linker and DNA since the binding free energies of monointercalators should be additive ($\Delta G_{bis} = \Delta G_{mono} + \Delta G_{mono}$).[7,116] Recall from above, that binding constants ranging from 10^4 – 10^6 M^{-1} should then be approximately 10^8 – 10^{12} M^{-1} upon dimerization. Ultimately, this increase in affinity should ideally produce a decrease in dissociation rate from the DNA, and thereby more significant biological effects. The

concept of building polyintercalators and their improved DNA binding properties has been extensively reviewed by Wakelin.[117]

Several bisintercalators can be seen in **Figure 1.16**, including the natural product echinomycin, isolated from *Streptomyces echinatus*.[99] The two intercalating units are quinoxaline rings connected by a cyclic octadepsipeptide. It has been shown to intercalate CpG steps from the minor groove of DNA, with the linker spanning two base pairs (4 bp binding site).[26,118] The two alanine residues in the linker have been shown to make important hydrogen bonding contacts to the adjacent interior bases, and coupled with a large number of Van der Waals interactions, maintains an approximate affinity of $K_A \sim 10^6$ M^{-1}.[119] Like many other intercalators, echinomycin produces enhanced biological activity such as antitumor, antiviral, and antibiotic activity which made it a promising for a candidate of phase II clinical trials. However, it was ultimately disregarded due to an unacceptable amount of increased toxicity.[99]







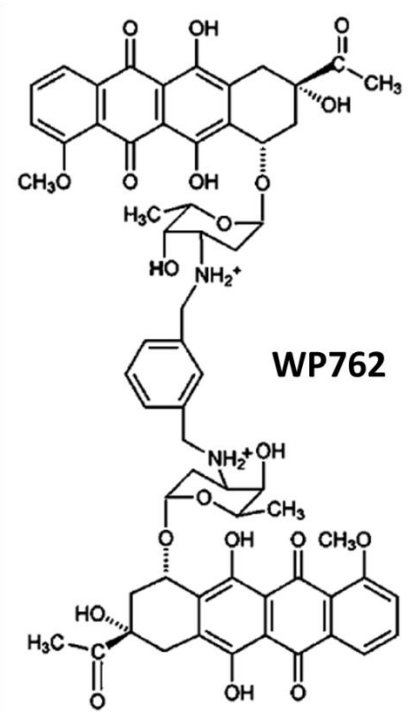


Figure 1.16 Chemical structures of a few representative bisintercalators.

LU 79553 (**Figure 1.16**) is another bisintercalator composed of two naphthalimide moieties connected by an aminoalkyl linker. Bailly *et al*. has shown LU 79553 to preferentially bind TpG and GpT steps from within a restriction digest of plasmid DNA. Their results suggested this preference was accomplished through the major groove of DNA, with an overall association constant of $K_A \sim 2 \times 10^6$ M^{-1}.[120] It was not until Gallego and Reid had produced an NMR structural study that the drug was confirmed to bind the oligonucleotide d(AT|GC|AT) through a mode of bisintercalation at the TpG and CpA steps, with the linker spanning 2 bp in the major groove of DNA.[121] This approach to DNA intercalation is not common, provided that most intercalation is through the minor groove.[120] Gallego proposed that a hydrogen bond between the amino group of the linker and the O6 of the guanine residue was partially responsible for the specificity. Interestingly, they observed a dynamic ring flipping of the intercalated mono-naphthalimide ring, which they rationalized by an electrostatic repulsion between the electron rich guanine (N7 and O6 on the major groove edge) and the oxygens on the imide intercalating unit. The observed exchange rate (flipping) was on the time scale of approximately 20 - 1800 sec^{-1} at 2 - 36 °C, respectively. Limited motion was observed in the linker, to which they proposed that the many electrostatic contacts between the aminoalkyl linker and the major groove appear to be the main factor governing the unique groove recognition.[121] As a topoisomerase II inhibitor, LU 79553 has shown remarkable activity against a wide range of human tumor xenografts cultured in nude mice which include lung, ovarian, colon, and breast cancers, respectively. These same experiments in mice also indicated that tumor regression has occurred on top of the tumor growth inhibition. However, phase I clinical studies indicated that neuro-muscular toxicity effects were too great to proceed with cancer treatments in humans.[122,123]

The Chaires lab has been in pursuit of rationally designing higher affinity bisintercalators with increased specificity towards DNA. Their studies have focused on linking the monointercalator daunorubicin (daunomycin, **Figure 1.13**). The first successful attempt at this rational design proceeded through studying the high-resolution structure of the daunorubicin-DNA complex, and connecting two daunorubicin moieties with a *p*-xylene linker. The result was the new bisintercalating anthracycline antibiotic, WP631 (**Figure 1.16**). Their proposed goal was to achieve the above mentioned multiplicative properties to the binding constants of the new bisintercalator, relative to that of the monointercalating daunorubicin. High-resolution NMR of the WP631-$d(ACGTACGT)_2$ complex confirmed bisintercalation, with the daunorubicins inserted between the CpG steps and the linker spanning four bp in the minor groove of the duplex.[124] A G,C-rich preference was confirmed by DNA melting studies on a homogeneous, 214 bp DNA fragment. It was determined that WP631 produced a binding constant of $K_A \sim 3 \times 10^{11}$ M^{-1} at 20 °C with herring sperm DNA.[7] While the overall affinity is extremely impressive, rivaling most other protein-DNA interactions, it was not the theoretical square (10^{14} M^{-1}) of the parent daunorubicin (10^{7} M^{-1}). The thermodynamic studies suggested that the lower than expected binding constants (and ultimately the binding free energy) was the result of a larger disfavored entropic contribution for WP631, as opposed to that for daunorubicin itself. Not to be discounted, the Chaires group had achieved a rationally designed bisintercalator that is appropriately classified as an ultra-tight affinity, and had also shown it to overcome a multidrug resistance in cultured cell lines, of which daunorubicin alone was not capable of producing.[7]

Improving upon their own design, the Chaires lab altered the linker between daunorubicin units from a *para* to a *meta*-xylene conformation to produce the

bisintercalator WP762 (**Figure 1.16**). The observed binding constant to herring sperm DNA at 20 °C was 7 x 10^{12} M^{-1}.[125] The difference in affinity between WP762 and the aforementioned WP631 was explained through molecular modeling, which indicated the *m*-xylene linker was more closely aligned to the wall of the minor groove. This conformation likely produces more favorable Van der Waals contacts than that of the *p*-xylene linker of WP631. The cytotoxicity of WP762 compares favorably to that of other anthracyclines, with nanomolar concentrations able to inhibit cell growth better than the previous WP631 in a breast carcinoma cell line. While altering the linker produced a higher affinity than that of WP631, the bisintercalator WP762 is still short of the theoretical 10^{14} M^{-1} value for the squared monomeric daunorubicin affinity. Again, they concluded that this lower than expected value was due to the unfavorable entropic penalty of connecting two daunorubicins molecules together.[125] From their studies, it appears that connecting two intercalating moieties does indeed increase affinity, however, the entropic penalties associated with the process are unlikely to allow a full multiplicative effect in binding constants (additive in binding free energy), and quite possibly might be amplified when connecting more than two intercalators together.

The amplified decrease in the multiplicative binding constants as polyintercalators add units is partially supported by the aminoacridine class of intercalators, which have been studied with respect to the creation of both dimer and trimer derivatives connected by flexible aminoalkyl linkers (**Figure 1.17**). Capelle *et al*. have studied a group of diacridine and monoacridine polyintercalators with respect to their kinetics towards poly(dAdT) dsDNA. They had observed that the association rate constant, k_a, was approximately 3 x 10^7 M^{-1} $sec^{-1}$for both the mono and diacridine intercalators. The difference, however, was that the dissociation rate constant, k_d, for the bisintercalating diacridines was estimated at 0.015 sec^{-1}, which resulted in an overall affinity of 10^9 – 10^{10}

M^{-1} at pH = 5.0 for the diacridines, compared with 10^5 – 10^6 M^{-1} for the monoacridines.[126] This result was convincing in that increasing the intercalating units resulted in increased affinities, however, Atwell *et al*. observed that increasing to triacridine derivatives resulted in rate constants and affinities of approximately the same order of magnitude as the diacridine analogues.[127] It should be noted that the kinetic incubations were maintained at a pH = 5.0 to prevent self-stacking, as well as potential issues measuring the true dissociation due to an observed rapid exchange between different strands of DNA. There has also been debate on the ability of triacridine derivatives to bind DNA through true trisintercalation, or if bisintercalation is the apparent binding mode with the interior acridine unit bound in a groove, especially provided that the diacridines appear to be just as cytotoxic as the triacridines against a L1210 leukemia cell culture.[117,127]

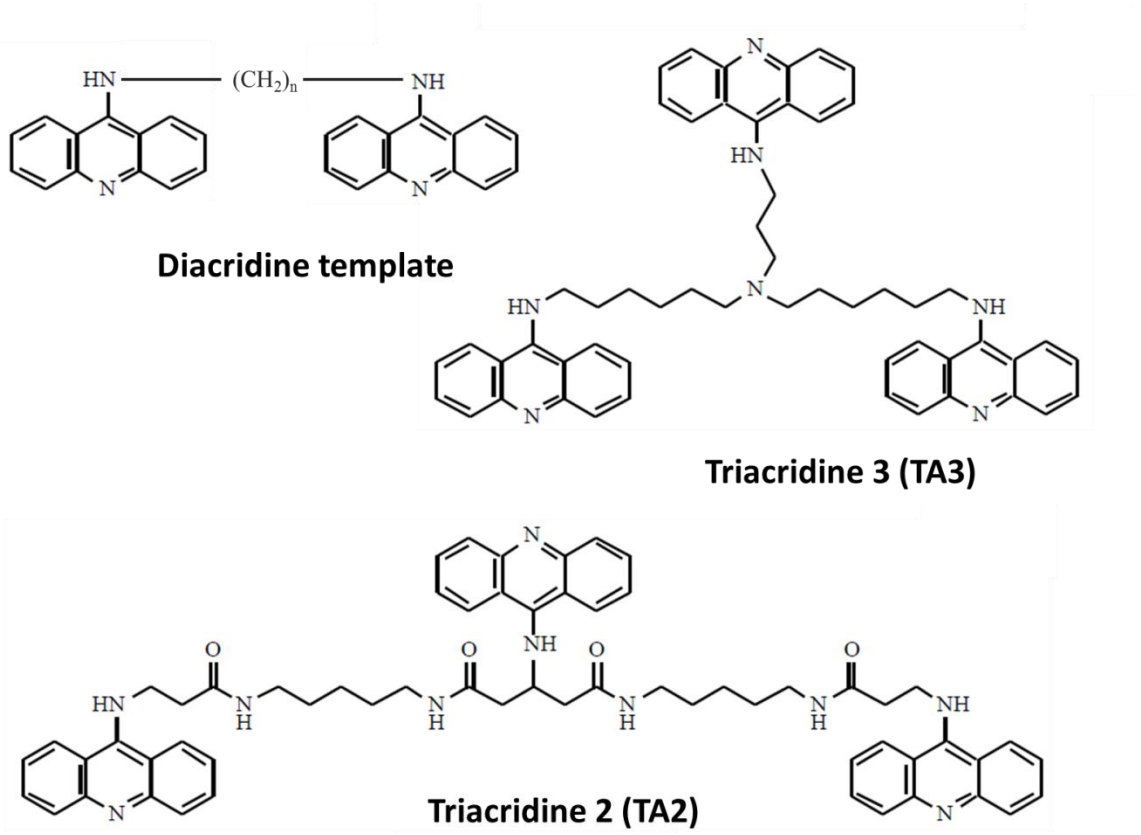


Figure 1.17 Chemical structures of diacridine template and triacridine polyintercalators TA2 and TA3.[117]

1.4.6 Threading Intercalators

Threading intercalation is different from standard intercalation, in that the aromatic molecule is still inserted between the nucleobases, but one end protrudes into the major groove, while the opposite end resides in the minor groove simultaneously. The topic of threading intercalation has been extensively reviewed by Takenaka and Takagi,[128] and several representative threading intercalators can be seen in **Figure 1.18**. Threading intercalation was first described regarding the natural product nogalamycin

(isolated from *Streptomyces nogalator*) binding to dsDNA. Similar to daunorubicin (daunomycin) and doxorubicin (adriamycin), nogalamycin is a structurally complex molecule containing two bulky sugar substituents on opposite sides of the anthracycline core, which requires that one of the sugar groups must thread through the stacked nucleobases of the double helix to bind via threading intercalation. A neutral nogalose sugar and a methyl ester reside at one end of the molecule, while a positively charged bicyclic amino sugar inhabits the opposite end, giving the molecule a characteristic dumbbell-like shape. A crystal structure of two nogalamycin molecules bound to the self-complementary, 5' cytosine methylated duplex, $d(CGTACG)_2$, has confirmed the threading intercalation binding mode between each CpG step.[129] They observed that the positively charged bicyclo amino sugar resides in the major groove, stabilized by hydrogen bonds to the C•G base pair. The neutral nogalose sugar is located in the minor groove, and appeared to be devoid of any significantly stabilizing interactions. The threading intercalative binding mode has been correlated with the extremely slow off-rates observed for nogalamycin, with a $k_d \sim 1 \times 10^{-3}$ sec^{-1} and an overall affinity of approximately $10^5 - 10^6$ M^{-1}.[25,130] Slower dissociation off-rate constants have been linked to the ability of nogalamycin to inhibit DNA-directed RNA synthesis as well as its toxicity towards both gram-positive bacteria and some tumors.[131,132]

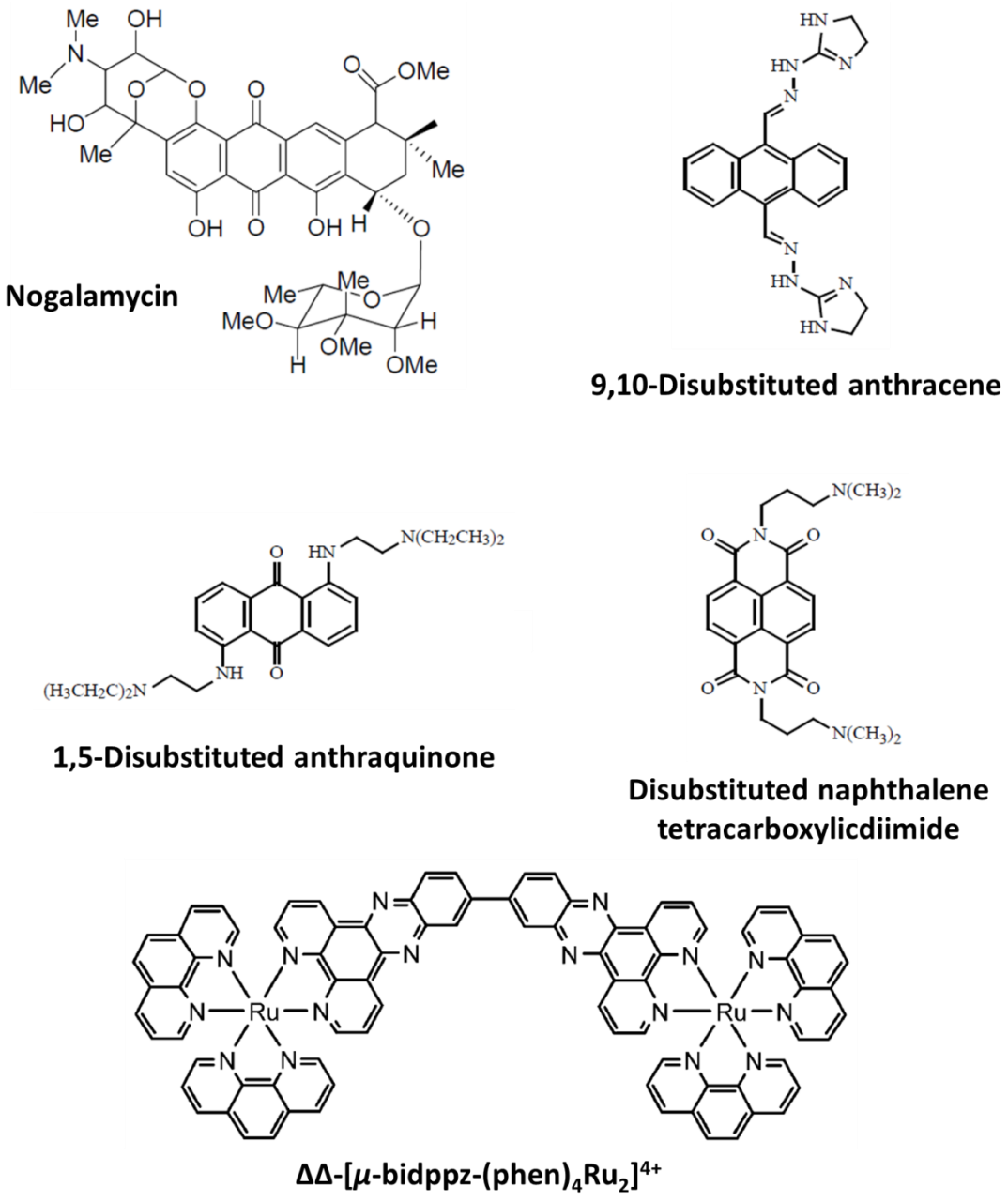


Figure 1.18 Chemical structures of representative threading intercalators.

As stated above, threading intercalation is linked to larger binding constants when compared with standard intercalators, likely due to increased favorable interactions in both grooves simultaneously. It has been suggested that the slower dissociation off-rates characteristic of threading intercalators could produce significantly increased biological activity, as well as potential uses for probing the dynamic nature of DNA in solution.[99,128] Naphthalene tetracarboxylic diimide (NDI) derivatives have become increasingly popular as threading intercalators because of the synthetically easy derivatization of the parent NDI structure. The Wilson lab was the first to characterize the interactions of an NDI threading intercalator (**Figure 1.18**).[133] Kinetic studies indicated threading intercalation as the binding mode, with interesting effects observed in the salt dependence of dissociation. Typical dicationic intercalators, such as propidium, show a quantitative and linear dependence on the salt concentration. For instance, propidium was observed in a log (K_D) versus –log([Na^+]) to have a linear slope with an approximate value of 2. This value suggests the displacement of two ion pairs upon binding to dsDNA.[105] Contrary to propidium, the observed NDI compound was also dicationic, however produces a linear slope of approximately 1. The resultant conclusion was that the NDI places one charged linker in each groove simultaneously, and that the rate-limiting step for dissociation was dependent upon breaking only one of the electrostatic interactions between the intercalator linker and the DNA backbone. Measured under comparable conditions, they observed that NDI associates and dissociates from DNA more slowly than propidium, as well as other typical monointercalators. The apparent on-rate was $k_a \sim 1.4 \times 10^5$ M^{-1} sec^{-1}, and the apparent off-rate was $k_d \sim 0.20$ sec^{-1}, providing an overall affinity of $K_A \sim 7 \times 10^5$ M^{-1}.[134] The reported rate constants strongly support that only modestly sized side chains can have dramatic effects on DNA interaction kinetics compared with standard intercalation.

For more than a decade, the Iverson lab has systematically designed various NDI-based bisintercalators and polyintercalators to target DNA in a sequence specific fashion. Initially, Lokey *et al*. synthesized a novel class of NDI intercalators, in which a range of one to four NDI intercalating units were connected by flexible peptide linkers in a head-to-tail fashion on solid phase support.[135] This work represented the first known tetraintercalating species that bound to dsDNA; however, without structural evidence a threading intercalation binding mode could not be confirmed. The polyintercalators were studied kinetically using UV/Vis spectroscopy, and it was apparent that all of the NDI derivatives exhibited a dramatic preference for poly (dGdC) over poly (dAdT), as was also evident through the GC specificity of DNase I footprinting studies. Murr *et al*. has extended the NDI derivatives to a record breaking eight NDI units, to synthesize the first known oktakis-intercalator.[136] In the same study, proposed dissociation constants for a mono-, bis-, tetra-, and octaintercalators were estimated from the composite data to poly (dGdC) DNA at approximately $K_D \sim 5\times10^{-6}$, 2.5×10^{-7}, 1.25×10^{-7}, and 1.56×10^{-8} M, respectively.

Parallel to the efforts of developing and studying longer polyintercalating molecules, our lab was also pursuing a combinatorial approach to examine varying linker compositions for NDI bisintercalators in order to determine how different peptides might alter the binding properties, especially specificity, towards dsDNA. Guelev *et al*. observed via DNase I footprinting, that by changing the composition of the peptide linker, the sequence specificity was also changed.[137] Two particular bisintercalators were observed to recognize distinct sequences on a 231 bp DNA restriction fragment of pBR322. These results, combined with later high-resolution NMR structural studies, determined that a dicationic bisintercalator with a –Gly-Gly-Gly-Lys linker (**G_3K**) specifically bound the duplex d(CG|GTAC|CG)$_2$ via threading bisintercalation with ~ 10^6

M^{-1} affinity.[138] This bisintercalator (compound **1**, **Figure 1.19 A**), intercalated between the GpG and CpC steps, with the linker spanning four base pairs in the major groove of the duplex. Recall from earlier discussion, small molecules binding to DNA through any mode, rarely do so through the major groove. This characteristic allows for compound **1** to possibly serve a useful function in the process of creating longer polyintercalators. The second dicationic bisintercalator (compound **2**, **Figure 1.19 B**) of interest contained a –βAla-βAla-βAla-Lys- linker (**βA_3K**), and was also examined through high-resolution NMR bound to its specific sequence of d(CG|ATAA|GC)·(GC|TTAT|CG) with approximately the same affinity as compound **1**.[139] This specificity was achieved through bisintercalation at the GpA and ApG steps with the linker spanning four base pairs, although located in the minor groove of the DNA. Comparing the structural studies of compounds **1** and **2** suggests that the driving force for the specific sequence recognition between intercalation sites is influenced not only by hydrogen bonding and Van der Waals contacts, but also by the different linker lengths of the bisintercalators. The longer **βA_3K** linker follows the deep minor groove, while the shorter **G_3K** linker cuts across the shallow major groove. The real benefit to our laboratory research was that we now had two bisintercalating ligands that demonstrate sequence specific binding through opposite grooves on the DNA, a useful tool for threading intercalators.

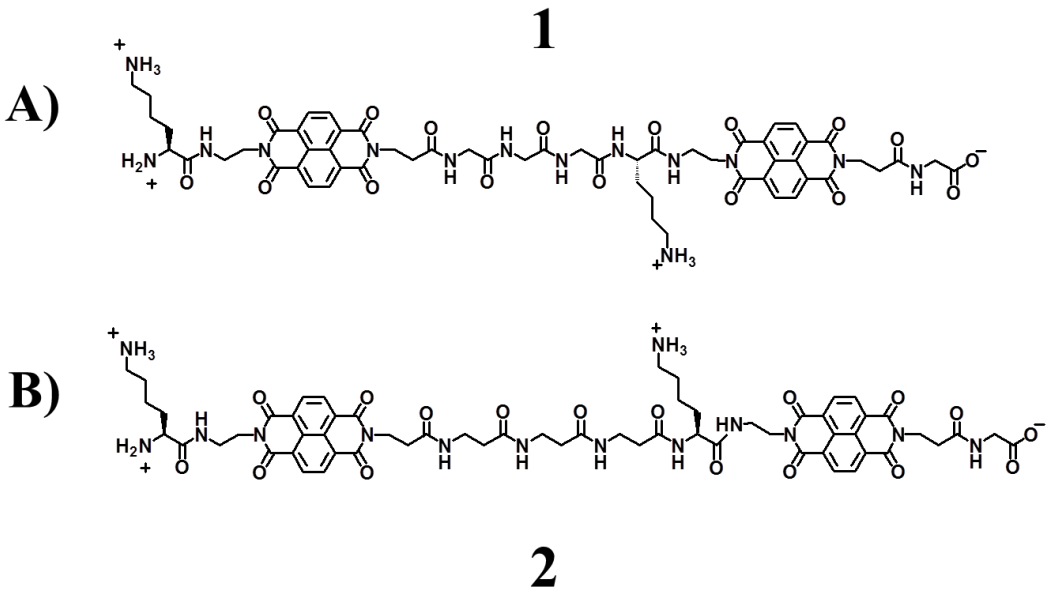
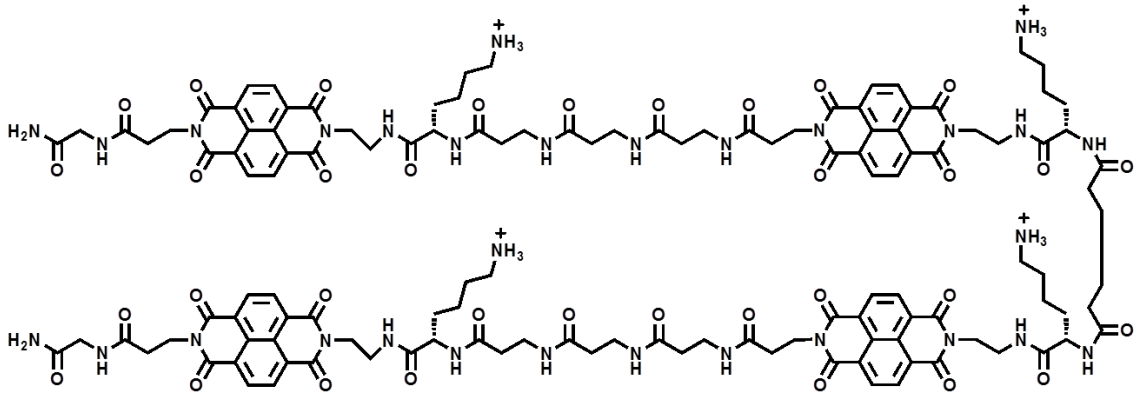


Figure 1.19 A) Chemical structure of bisintercalator **1**, containing a –$(Gly)_3$Lys- linker. B) Chemical structure of bisintercalator **2**, containing a –$(\beta Ala)_3$Lys- linker.

The next step was then to rationally design a modular system, by connecting multiple NDI units with the alternating minor groove and major groove specific linkers. The product of this design was a novel threading tetraintercalator (compound **3**, **Figure 1.20**).[8] This design was attempted by connecting two of the minor groove bisintercalator **2** using a symmetrical adipic acid for the major groove linker. Modeling results confirmed the likely conservation of a few key hydrogen bonding interactions similar to the **G_3K** linker of compound **1** with the interior $d(GTAC)_2$ sequence , and the use of this residue would impart C_2 symmetry to the overall molecule. The result was a C_2 symmetric tetraintercalator rationally designed to bind the 14 bp sequence

d(G|ATAA|GTAC|TTAT|C)$_2$ through a threading polyintercalation mode where the linkers would alternate in the fashion of minor groove - major groove – minor groove. High resolution NMR studies verified the first ever threading tetraintercalation binding topology with a 1:1 sequence specificity and the anticipated minor-major-minor groove linker orientation with the NDI units intercalating at the GpA, ApG, CpT, and the TpC steps.[8] These results verified that the prediction of sequence specific threading polyintercalators could be achieved through a modulated system connecting bisintercalators.



3

Figure 1.20 Chemical structure of tetraintercalator **3**.

A similar approach to sequence specific recognition of both grooves simultaneously was also completed in our lab. Chu *et al*. was able to take the two groove specific linkers mentioned above, and create a macrocyclic bisintercalator.[140] This design

paired the major groove **G_3K** linker with the minor groove **βA_3K** linker. The resulting specificity was dominated by the major groove recognition element, with this macrocyclic bisintercalator preferring the sequence d(CG|GTAC|CG)$_2$. 2D NMR structural analysis confirmed a 1:1 stoichiometry and the expected placement of **G_3K** in the major groove and **βA_3K** in the minor groove, thereby forming a pseudocatenane-like complex with the DNA. These examples demonstrate the importance of understanding sequence specific recognition of the linker elements, and how they can be used to modulate the rational design of other threading intercalators.

Current work in our lab is focused on extending our understanding of linker-directed sequence specificity. We have developed a tetraintercalator using the same platform as described above, however the use of **G_3K** was used for the major groove recognition element. This removes the original C_2 symmetric design, which actually serves to impart a true N→C directionality to the molecule. While there remains to be a structural characterization of the complex, altering the major groove binding element has been shown through sequence specific DNase I footprinting to change the sequence recognition in a subtle fashion. The resulting sequence preference is for d(G|ATAA|GTAC|CTAA|G)·(C|TTAG|GTAC|TTAT|C).[(Smith et. al, manuscript in preparation.)]

1.5 Molecules That Bind DNA Through High Affinity Non-covalent Interactions

One might imagine two opposite extremes to therapeutic applications of molecules that bind to DNA. If it was desired to destroy a cell, a particular sequence specificity is ultimately not crucial if a molecule has the ability to bind in such a way that many molecules are bound all over the DNA so that they interfere with any possible replication or transcription processes. On the other hand, if cell death was not necessary, but rather only a known genetic process was undesirable, then targeting that specific gene

for regulation would require a ligand to be extremely sequence specific and that it would stay bound long enough as to interfere with only the transcription of that individual gene. The former approach would most likely require high drug concentrations, and result in more cytotoxic side effects while relying on the cancerous cells to uptake the drug more readily. The latter approach, rather, would conceivably result in fewer side effects since theoretically, only one molecule would be required per cell resulting in the inhibition of only one gene, all of this assuming that it was even capable of locating the DNA *in vivo* and that no other desirable gene transcription would overlap with the targeted binding site. It could also be possible to kill the cell with such a sequence specific approach if a vital cellular function was shut down. For this latter approach, the most desirable drug characteristics would be the specificity of a unique sequence and an extraordinarily slow dissociation rate-constant from the binding site in order to interfere with transcription of the gene when it occurs *in vivo*. The following section will focus on the characteristics important to the latter approach to DNA binding.

High affinity binding to DNA requires a series of stabilizing interactions, and the more stabilizing interactions can induce a sequence specific recognition, which would thereby be closely associated with strong interactions towards DNA. The combination of sequence specificity and high affinity binding to DNA are of significant importance with regard to the ability of a ligand to interfere, or alter biological processes.[5]

The lactose (*lac*) repressor protein is considered a high affinity system that has served as a benchmark system for studying gene regulation.[141] The *lac* repressor protein is capable of binding the major groove of DNA, more specifically in the operator region of the operon (**Figure 1.21**).[142,143] In this complex, RNA polymerase is inhibited from the transcription of the operon, and in this way the *lac* repressor is effective at regulating gene expression. The *lac* repressor/operator complex is classified as a high affinity

system, with an estimated affinity of 10^{13} M^{-1}.[144] The rate constants have also been experimentally determined, with approximate values for the on-rate at $k_a \sim 10^9$ M^{-1} sec^{-1}, and approximate off-rate of $k_d \sim 10^{-4}$ sec^{-1}.[145]

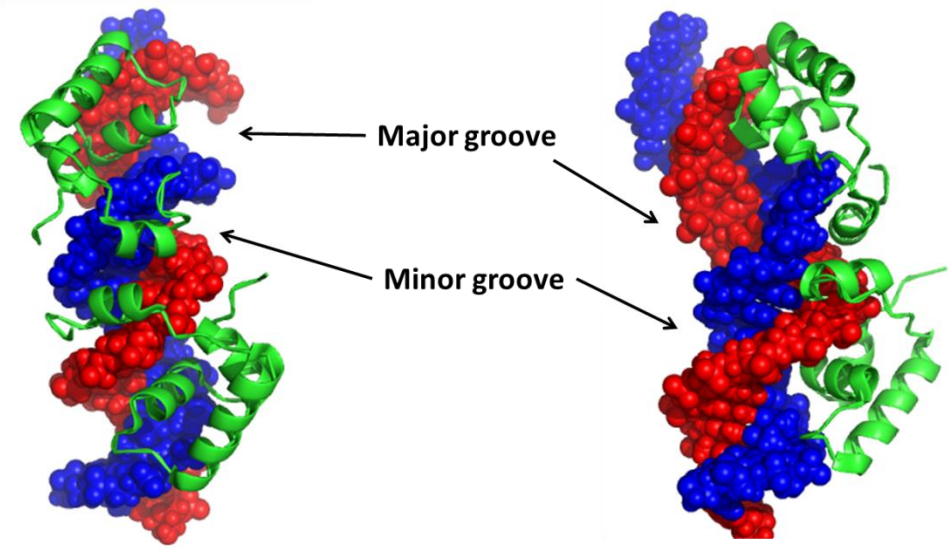


Figure 1.21 Two different views of an X-ray crystal structure for the lactose repressor DNA binding domain (green) complexed with the O1 operator DNA (red & blue strands), PDB ID: 2KEI.[143] The two views were produced by The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.

The RNA polymerase holoenzyme is another protein-DNA interaction characterized as high affinity. The core enzyme of RNA polymerase contains four subunits (alpha, beta, beta′, and omega) that can act to transcribe RNA from the DNA template, although this core enzyme is not efficient at recognizing promoter regions of the DNA. The holoenzyme however, contains another subunit (sigma) that is capable of increasing the affinity towards the promoter region of the DNA. The overall affinity is

approximately 10^{11} M^{-1}, with a dissociation rate-constant of approximately $k_d \sim 3 \times 10^{-6}$ sec^{-1} at 37°C.[146] Interestingly, Hinkle *et al*. have observed a unique temperature dependence on dissociation of the RNA polymerase holoenzyme. At 25°C, the dissociation rate-constant is approximately, $k_d \sim 1.3 \times 10^{-5}$ sec^{-1}, a faster dissociation time. They proposed that this is most likely due to the need for the DNA double helix to melt, so that the RNA polymerase holoenzyme can transcribe the ssDNA, therefore rationalizing the higher affinity at physiological temperatures.[146]

As previously discussed, many small molecules that recognize DNA in a sequence specific fashion also exhibit quite large affinity constants. For instance, the sequence specific polyamides developed in the Dervan lab offer high affinities that range from $K_A \sim 10^8 - 10^{11}$ M^{-1}, with the best (slowest) dissociation rate constants at approximately $k_d \sim 2 \times 10^{-3}$ sec^{-1}.[147] Their tandem hairpin polyamides are able to produce affinity constants in the range of approximately $K_A \sim 10^{12}$ M^{-1}, however no dissociation rate constant was reported.[148]

Triplex forming oligonucleotides and peptide nucleic acids also bind dsDNA with relatively high affinities. Maher III and co-workers have done a thorough kinetic analysis on a representative TFO, to which they observed an approximate affinity of $K_A \sim 10^8$ M^{-1}, with the respective rate constants $k_a \sim 2.2 \times 10^3$ M^{-1} sec^{-1} and $k_d \sim 2.2 \times 10^{-5}$ sec^{-1}.[149] Demidov *et al*. performed kinetic characterization on 10mer PNAs, and observed a dissociation rate constant of $k_d \sim 2.5 \times 10^{-4}$ sec^{-1}.[90] Some modified PNA derivatives are capable of achieving overall sub-nanomolar affinity constants of $K_A \sim 10^9 - 10^{10}$ M^{-1}.[150]

With regard to intercalators there are three particular examples of high affinity binding. Recall WP762 (**Figure 1.16**) developed in the Chaires lab, that was capable of achieving an affinity with herring sperm DNA of $K_A = 7.3 \times 10^{12}$ M^{-1}.[125] Laugâa and co-workers have studied di- and triacridine derivatives in an effort to produce higher affinity

analogues. In doing so, they discovered a triacridine derivative, AcTri2 (TA2, **Figure 1.17**), that showed rate constants to poly (dAdT) DNA of $k_a \sim 2.6 \times 10^7$ M^{-1} sec^{-1}, $k_d \sim 1.2 \times 10^{-4}$ sec^{-1}, and an overall affinity constant of $K_A \sim 2.2 \times 10^{11}$ M^{-1}.[151] Perhaps the most astonishing dissociation rate-constant yet seen for DNA binding small molecules, is that of a ruthenium bisintercalator designed in the Nordén and Lincoln laboratories. The metallo-bisintercalator, ΔΔ-[μ-bidppz-$(phen)_4Ru_2]^{4+}$, was shown to display a dissociation rate constant from poly (dAdT) DNA of 1.2×10^{-4} sec^{-1} at 37°C as obtained from an Arrhenius plot estimation, which upon extrapolation to 25°C is approximately $k_d \sim 5 \times 10^{-6}$ sec^{-1} (**Figure 1.18**).[152]

1.6 Other High Affinity Systems

It is interesting to compare the dissociation rate constants provided above for DNA binding interactions to those for other reported long-lived complexes. A comprehensive table including experimentally determined, and some extrapolated values, for the kinetic parameters surrounding a diverse set of DNA binding molecules, as well as those presented in this section are shown in **Table 1.3**.

Among the list of extremely slow dissociating systems are a handful of antibody-antigen interactions. The Lerner group has developed a combinatorial methodology for improving the affinity of antibodies, and in the process, developed a single-chain variable fragment (scFv) anti-tumor necrosis factor-alpha (TNF-α) mutant, cb2-6, which displayed a dissociation constant of $k_d \sim 4.5 \times 10^{-7}$ sec^{-1}, and an overall affinity for TNF-α of $K_A \sim 1 \times 10^{12}$ M^{-1}.[153] The Wittrup laboratory has produced two highly engineered antibodies, a scFv 4-4-20 antibody mutant 4 M5.3 that binds a fluorescein-biotin hapten, and an MFE-23 scFv mutant sm3E that binds a carcinoembryonic antigen. These evolved antibodies

display dissociation rate constants of $k_d \sim 1.4 \times 10^{-6}$ and 3×10^{-7} sec^{-1} at 25°C against their antigens, respectively, and an overall affinity for the 4 M5.3 antibody of $K_A \sim 2.1 \times 10^{13}$ M^{-1}.[154,155]

The "benchmark" by which all non-covalent interactions are measured, has for some time been the avidin-biotin system.[156] This particular system consists of four avidin protein subunits, each able to bind a biotin molecule with an apparent dissociation rate constant $k_d \sim 7.5 \times 10^{-8}$ sec^{-1} with an overall for affinity for biotin of $K_A \sim 10^{15}$ M^{-1}.[157] Impressive as femtomolar affinity is, the Whitesides group has developed a multivalent vancomycin system that tops even the affinity of avidin-biotin. This system relies upon a trivalent vancomycin carboxamide that binds to a trivalent ligand derived from the D-Ala-D-Ala peptide with a dissociation rate constant of approximately $k_d \sim 4 \times 10^{-8}$ sec^{-1} and an impressively high affinity of $K_A \sim 2.5 \times 10^{16}$ M^{-1}.[158]

Guest	Host	Binding Mode	k_{on} (M^{-1} s^{-1})	k_{off} (s^{-1})	K_A (M^{-1})	Reference
Proflavine	5'-Py-Pu-3' steps	M / Int	$1.4x10^7$	$4.2x10^2$	$\sim 10^4$	(106)
Ethidium	5'-Py-Pu-3' steps	M / Int	$8.8x10^5$	29	$1.5x10^4$	(104,105)
Propidium	5'-Py-Pu-3' steps	M / Int	$1.1x10^6$	9.5	$\sim 10^5$	(105)
Daunomycin	(A/T)GC or (A/T)CG	M / Int	$\sim 10^6$	~0.06	$1.6x10^7$	(7)
Actinomycin D	GpC steps	M / Int	$\sim 10^4$	$\sim 10^{-2}$	$\sim 10^6$	(11)
Acridine monomer	poly(dAdT)	M / Int	$3x10^7$	56	$5.3x10^5$	(126)
Nogalamycin	CpG steps	Th/ M/ Int	~440	$1x10^{-3}$	$4.4x10^5$	(25,130)
NDI monomer	poly(dGdC)	Th/ M/ Int	$1.4x10^5$	0.2	$7x10^5$	(134)
Porphyrin	poly(dGdC)	M / Int	-	-	$8x10^5$	(54)
Echinomycin	Internal CpG steps	B / Int	$1.1x10^4$	$1.3x10^{-3}$	$\sim 8x10^6$	(26,118,119)
LU-79553	$d(ATGCAT)_2$	B / Int	-	-	$2x10^6$	(120,121)
Acridine dimer	poly(dAdT)	B / Int	$3x10^7$	$\sim 1.5 \times 10^{-2}$	$\sim 10^9$-10^{10}	(126)
WP631	d(C|G(A/T)(A/T)C|G)	B / Int	-	-	$2.7x10^{11}$	(7)
WP762	$d(GC\|GTAC(GC)_2$	B / Int	-	-	$7.3x10^{12}$	(125)
ΔΔ-[μ-bidppz-$(phen)_4Ru_2]^{4+}$	poly(dAdT)	Th / B / Int	-	$5x10^{-6}$	-	(152)
AcTri2	poly(dAdT)	T / Int	$2.6x10^7$	$1.2x10^{-4}$	$2.2x10^{11}$	(151)
Polyamides	poly(dAdT) / universal	Minor Groove	$\sim 10^6$ - 10^7	$2x10^{-3}$	10^9 - 10^{12}	(147,148)
Triplex oligonucleotides	purine rich DNA	Major groove	$2.2x10^3$	$2.2x10^{-5}$	$\sim 10^8$ - 10^9	(149)
Peptide Nucleic Acids	purine rich / universal	Major groove (others)	$\sim 10^5$ - 10^6	$2.5x10^{-4}$	$\sim 10^9$ - 10^{10}	(90,150)
RNA pol holoenzyme	promoter DNA	dsDNA/ssDNA	$3x10^5$	$3x10^{-6}$	$\sim 10^{11}$	(146)
lac repressor	lac operon of DNA	Major groove	$\sim 10^9$	$\sim 10^{-4}$	$\sim 10^{13}$	(144,145)
cb2-6 (anti-TNF-α Ab)	TNF-α	Ab-antigen	-	4.54×10^{-7}	-	(153)
4M5.3 (4-4-20 Ab)	fluoroscein-biotin	Ab-antigen	-	1.6×10^{-6}	2.1×10^{13}	(154)
sm3E (MFE-23 Ab)	carcinoembryonic antigen	Ab-antigen	-	3×10^{-7}	-	(155)
Avidin	biotin	ligand-protein	-	7.5×10^{-8}	$\sim 10^{15}$	(157)
Tris(vancomycin) (RV_3)	tris(D-Ala-D-Ala)	receptor-ligand	$\sim 10^9$	$4x10^{-8}$	$\sim 2.5 \times 10^{16}$	(158)

Table 1.3 Kinetic parameters of variety of DNA binding molecules and high-affinity systems. Approximated values (~) are based on other experimentally determined values provided, or on a range of observed values. M = mono, B = bis, T = tris, Int = intercalation, Th = threading, Ab = antibody.

Chapter 2

Specificity Studies of a Threading Tetraintercalator using DNase I Footprinting

2.1 Chapter Summary

2.1.1 Goals

Determine if a threading NDI tetraintercalator is capable of selectively locating, and specifically binding a preferred designed binding site that is inserted within a large fragment of double-stranded DNA (dsDNA).

2.1.2 Approach

Synthesize an NDI tetraintercalator using a solid phase peptide synthesis protocol established by Guelev *et al.*[139] and expanded by Lee and co-workers.[8] A plasmid construct can then be prepared through ligation of an insert containing the tetraintercalator binding site. Specific primers can then be constructed for use in a PCR reaction to produce a desired long sequence of dsDNA, which by design, also contains the tetraintercalator binding site. Incubation of the tetraintercalator with long sequences of DNA can then be analyzed via DNase I footprinting for the selectivity and specificity of a preferred DNA binding site.

2.1.3 Results

Synthesis and purification of the NDI-based tetraintercalator was achieved according to Lee *et al.*[8] It was shown that the tetraintercalator binds to its preferred binding site d(GATAAGTACTTATC)$_2$ with high specificity. DNase I footprinting confirms the ability of the tetraintercalator to selectively locate its designed binding site

when inserted into either a 388mer or 467mer sequence of plasmid DNA, and bind with relatively high affinity.

2.2 INTRODUCTION

Research focused on the pursuit of small molecules that bind specifically to DNA has been an important pursuit for several decades. Many studies have provided insights into DNA recognition and have even shown the potential to modulate gene expression *in vivo*.[15,14] Polyamides are one class of small molecules that have shown true promise towards binding specific sequences of DNA with high affinity. The Dervan lab has developed a scaffold using these polyamides that is capable of recognizing specific sequences of DNA in a programmable way via parallel binding interactions in the minor groove.[76,5] Also useful for recognizing specific sequences of DNA is the formation of triplex DNA, which is able to accomplish this recognition based upon complementary interactions with the existing DNA bases through Hoogsteen base pairing in the major groove.[159,84,82] An interesting feature of both polyamides and triplex DNA, is their ability to recognize long sequences of DNA in both a programmable and modular fashion. While triplex DNA constructs can be regarded as programmable within the limitations of either purine rich or pyrimidine rich stretches of DNA, the polyamides developed in the Dervan lab represent the best accomplishment towards true sequence programmability with synthetic molecules.[76]

With regard to potential clinical interest, the ability to target longer sequences of DNA is important for imparting specificity, especially when one considers all the possible choices present to a small molecule interacting with a genome of DNA either *in vitro* or *in vivo*. Dervan and co-workers have calculated that a sequence spanning approximately 16 bp would be required to ensure the recognition of only one unique

sequence within the human genome.[5] The Iverson lab has focused on the development of scaffolds that utilize threading polyintercalation as a method for constructing small molecules, in a modular fashion, in an effort to specifically recognize and bind longer sequences of DNA.

Our group has developed a novel class of threading polyintercalators based upon the electron-deficient 1,4,5,8-naphthalenetetracarboxylic diimide unit (NDI), a known threading intercalator.[133,134,135] As described earlier, Guelev *et al*. analyzed a library of NDI-based small molecules with varying peptide linker sequences, and found a bisintercalator, compound **1** (**Figure 1.19 A),** with the internal **G_3K** linker that specifically bound to the sequence d(CG|GTAC|CG)$_2$,[137] A 2D NMR structural analysis later confirmed that **1** bound to the DNA sequence in a threading manner, with the **G_3K** linker spanning four base pairs in the major groove, while NDI units intercalate between the GpG and CpC steps of the (+) strand of the DNA as shown in **Figure 2.1 A**.[138] Interestingly, by changing to the **βA_3K** linker, the bisintercalator compound **2** (**Figure 1.19 B**) specifically binds the sequence d(CG|ATAA|GC)·d(GC|TTAT|CG). 2D NMR structural analysis again confirmed threading bisintercalation and that **βA_3K** spans 4 base pairs between intercalation sites. However, the longer and more flexible **βA_3K** linker is able to follow the narrower minor groove (**Figure 2.1 B**) as opposed to cutting across the wider major groove, as does the shorter **G_3K** linker.[139] The NDI units of bisintercalator **2** intercalate between the GpA and ApG steps on the (+) strand of the DNA. Stated again, these structural studies have shown that by simply varying the peptide linker between NDI units, we can access both the major and minor grooves while specifically targeting preferred sequences of DNA.

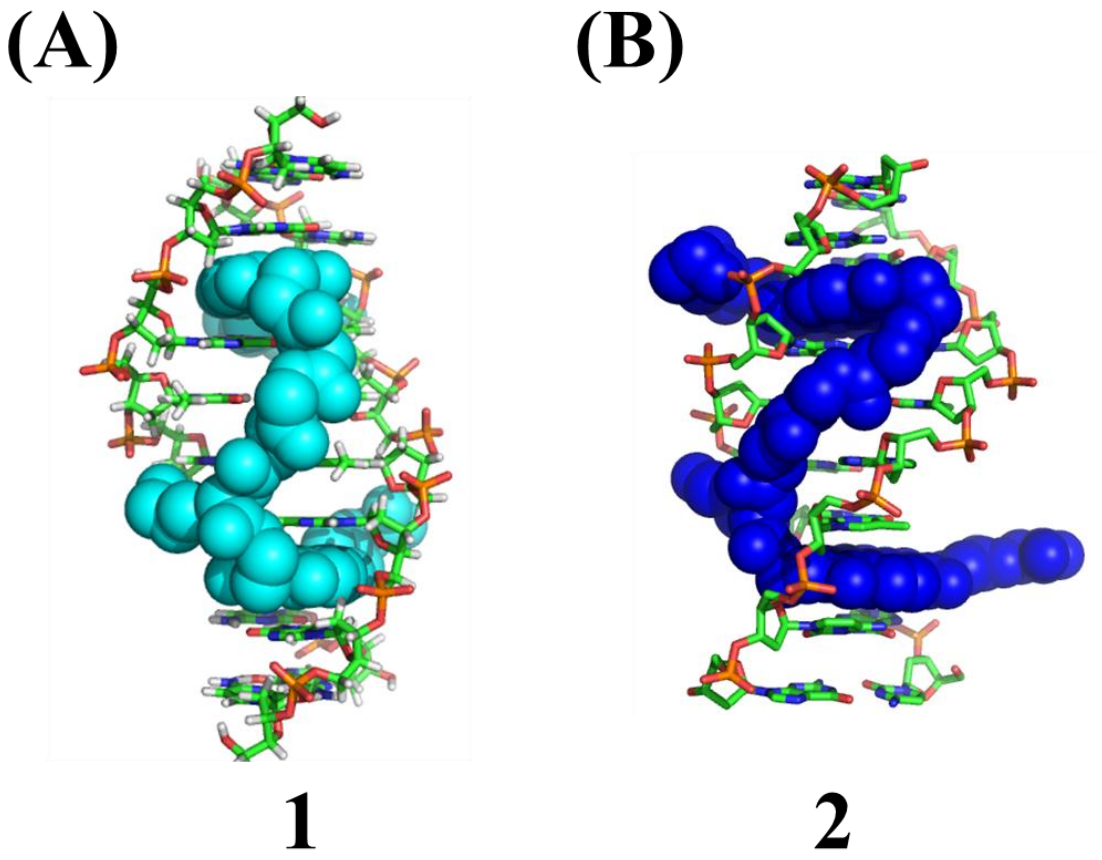


Figure 2.1 Models of the A) bisintercalator **1** with a –$(Gly)_3$Lys- linker bound to $d(CGGTACCG)_2$[138] and B) bisintercalator **2** with a –$(\beta Ala)_3$Lys- linker bound to d(CGATAAGC)·(GCTTATCG).[139]

With this information, the next step was to attempt to target both grooves of DNA simultaneously, all while still recognizing specific sequences of DNA. The ability to combine bisintercalators by simply crosslinking two of them together would allow for a modular approach to developing threading polyintercalators which would be capable of recognizing longer sequences of DNA. A cartoon representation of the concept used by Lee and co-workers to develop a C_2-symmetric tetraintercalator is shown in **Figure 2.2**.

The design would allow access to both grooves in an alternating fashion by crosslinking two bisintercalator **2** molecules with an adipic acid linker on solid phase resin (**Figure 2.3**).[8] The proposed binding topology would thereby alternate the placement of the linkers in the order of minor groove - major groove - minor groove. The reasons for choosing adipic acid as a linker were three-fold; it had C_2 symmetry, it would reduce the number of steps by half to synthesize the tetraintercalator in a head-to-tail fashion, and it was shown through molecular modeling calculations to maintain some of the key hydrogen bonding contacts to 5'-GTAC-3' in the major grove analogous to the –$(Gly)_3$Lys- linker of bisintercalator **1**.[28] Tetraintercalator **3** was thereby designed to bind the 14bp sequence d(G|ATAA|GTAC|TTAT|C)$_2$, a hybrid of the sequences bound by bisintercalators **1** and **2**. The structure and naming convention of **3**, as well as its designed binding site sequence, can be seen in **Figure 2.4 (A & B)**. 2D NMR structural data confirmed binding to d(G|ATAA|GTAC|TTAT|C)$_2$ with a 1:1 stoichiometry, as well as the pattern of groove binding that alternates the linker placement in the order minor groove, major groove, and minor groove accessed simultaneously as shown in **Figure 2.5**.[8] It was also observed that upon binding d(G|ATAA|GTAC|TTAT|C)$_2$, tetraintercalator **3** induces a slight distortion in the axis of the DNA towards the major groove. This is not totally unexpected, since many intercalators promote helix unwinding and other slight DNA distortions upon complete intercalation between DNA base pairs.[160,107]

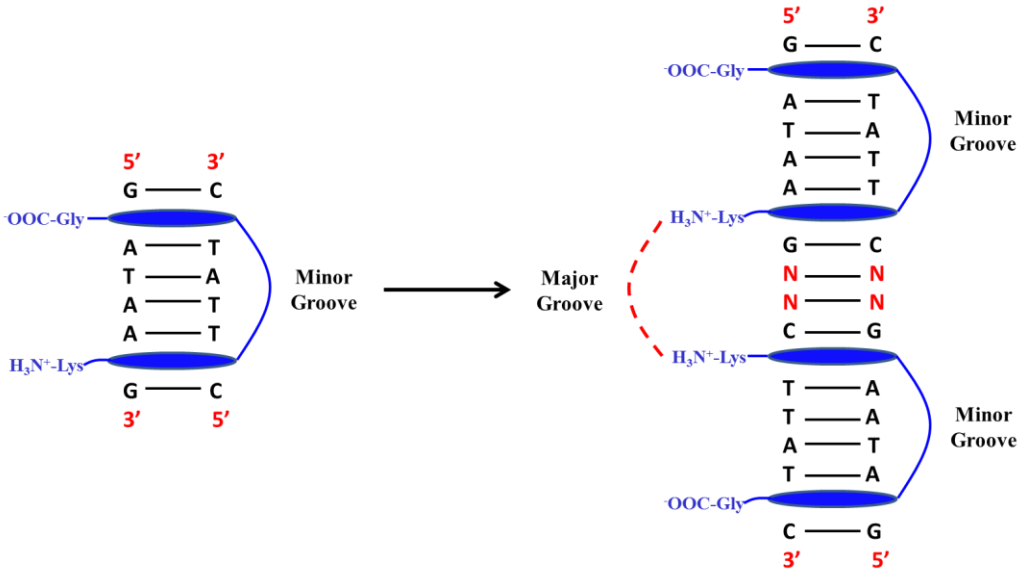


Figure 2.2 Cartoon representation of the rational design of connecting two bisintercalator **2** molecules bound to dsDNA. DNA grooves labeled with respect to the appropriate placement of the linkers connecting NDI units upon binding dsDNA.

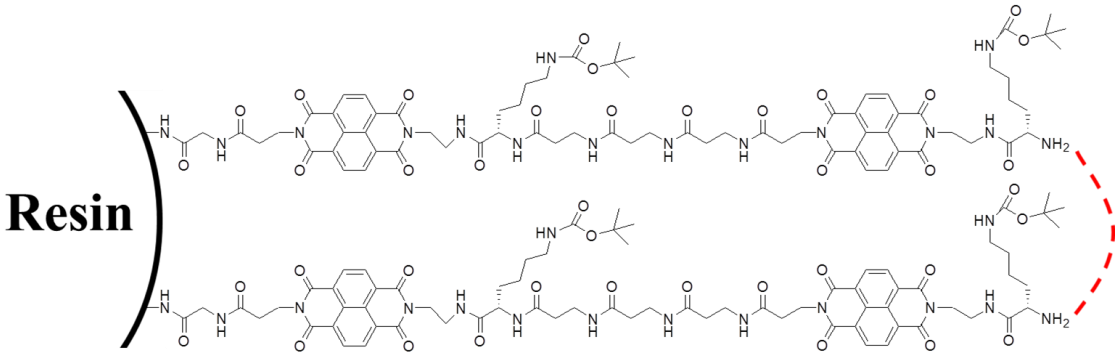
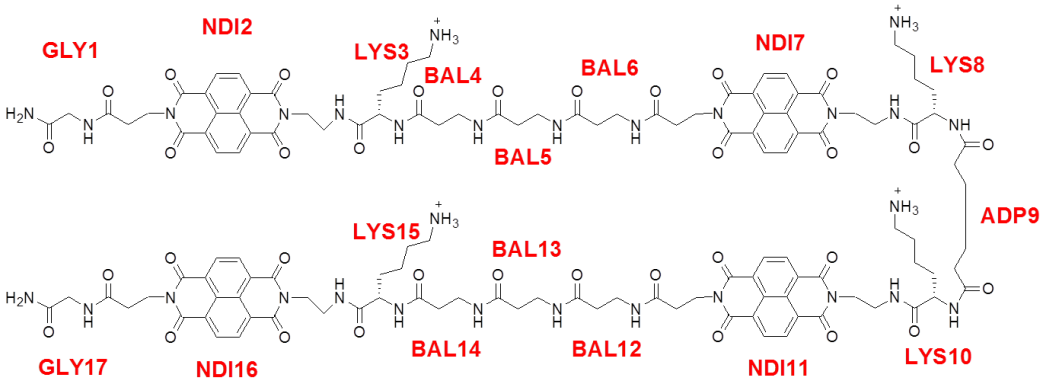


Figure 2.3 Cartoon showing two compound **2** bisintercalators attached on resin. Proposed crosslinking using adipic acid is indicated by the dashed red line.

A)



Tetraintercalator **3**

B)

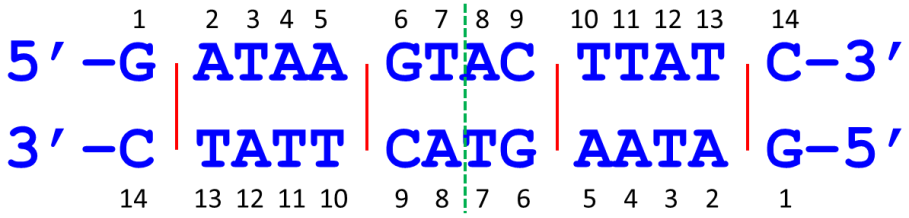


Figure 2.4 A) Structure and naming convention of the tetraintercalator **3**; B) The palindromic DNA sequence used in the NMR study with intercalation sites indicated by red lines, and C_2 symmetry indicated by a green hash.[8]

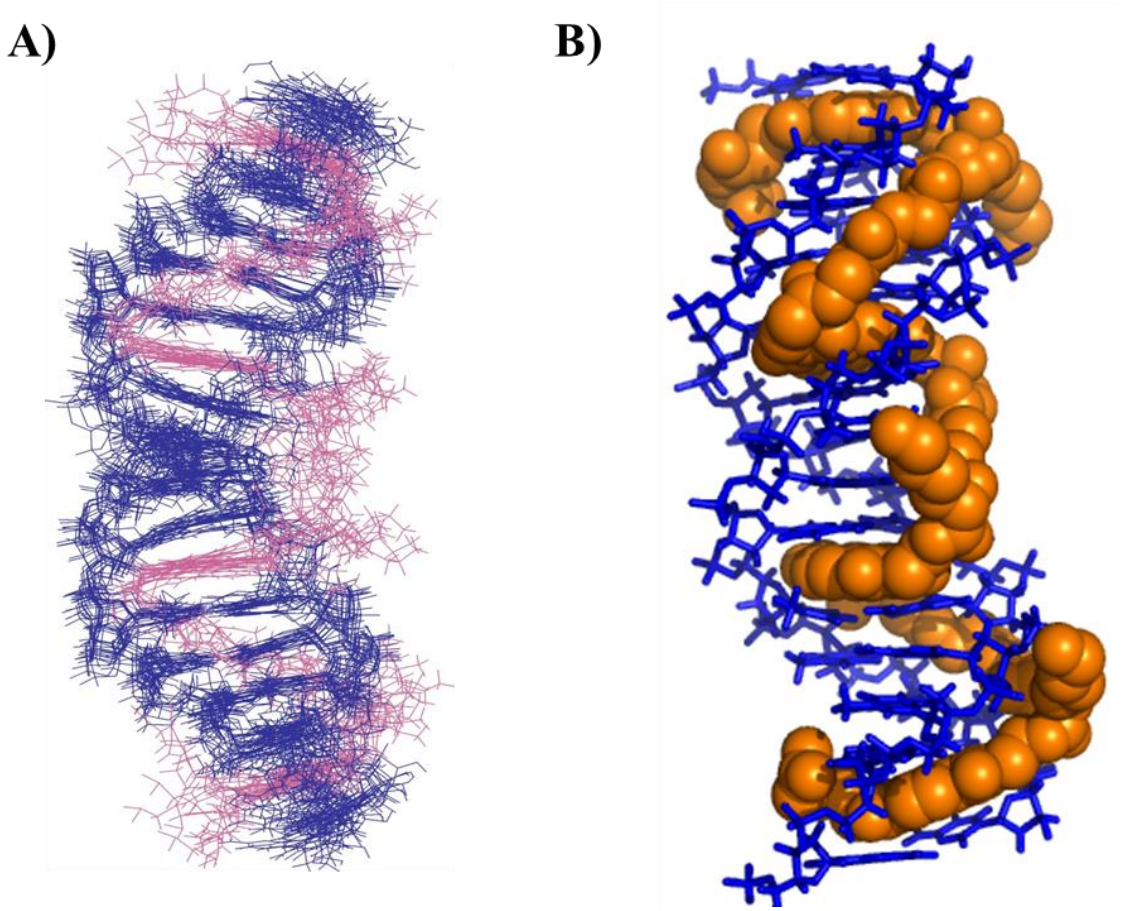


Figure 2.5 A) 12 superposed lowest energy conformations of **3** bound to the sequence d(GATAAGTACTTATC)$_2$ using distance constraints from NOE data of the 2D NMR. B) Space-filling model of **3**-d(GATAAGTACTTATC)$_2$ complex showcasing a threading polyintercalation mode of binding where the linkers alternate grooves in the order minor-major-minor. Models and cartoon illustrations were produced by The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.

Most DNA intercalators are capable of specifically recognizing a few bases, usually G,C-rich, in a DNA sequence.[107,161] However, small stretches of G,C-rich DNA within a relatively long sequence are frequent, thereby reducing the overall specificity for a unique sequence of DNA as stated above. This can be problematic, especially for intercalators that have practical purposes such as Adriamycin, which is used clinically and will target all available DNA with 5'-GpC-3' sequences.[162] These types of intercalators have practical applications because cancerous cells will be more likely to take in intercalators, and as a result, transcription of the cancerous DNA is shut down with greater frequency. If we can develop threading polyintercalators that are capable of selectively and specifically targeting longer sequences of DNA with relatively high affinity, we can possibly circumvent the likelihood of significant binding to just any random composition of DNA that is presented to the small molecules. This would be one step closer to learning how to reduce undesired side-effects produced by the lack of specificity offered by other, clinically relevant, DNA intercalators.[163,164] Research focus in this area might be able to produce likely scaffolds for the future development of new drugs designed to inhibit gene expression.

In light of the importance of designing small molecules able to bind longer sequences of DNA specifically, it is our task to see if tetraintercalator **3** designed by Lee *et al*. is specific in its binding to dsDNA.[8] We would like to determine if **3** can specifically locate its designed 14 bp binding site when presented with a long sequence of DNA that is composed of random nucleobases, a situation more similar to what would be encountered in a typical cellular environment. When considering the amount of molecular rearrangement that must occur for complete polyintercalation as shown in the NMR structure of **3** bound to its 14 bp sequence in **Figure 2.5 B**, it is not obvious that tetraintercalator **3** would be able to locate its designed binding site from within a longer

sequence of DNA. The molecular rearrangement necessary for dissociation from other non-specific binding sites could potentially be too arduous a process, thereby interfering with the ability of **3** to locate and fully associate with its designed binding site. Producing this answer would provide valuable insights into further development of threading polyintercalators in a modular fashion, as well as possibly offering itself as a potential scaffold to other labs that synthesize small molecules capable of specifically binding DNA.

2.3 RESULTS

2.3.1 Design

As previously discussed, Lee *et al*. described the development of a threading, tetraintercalator synthesized in a modular fashion, using adipic acid as an analogue to the –$(Gly)_3$Lys- linker of **1** in order to crosslink two molecules of compound **2**. The presumption was that by using the linkers from two previously described bisintercalators[138,139], tetraintercalator **3** would bind to a hybrid 14 bp sequence, d$(GATAAGTACTTATC)_2$, with a 1:1 stoichiometry.[8] The NMR structural analysis confirmed 1:1 binding to the designed DNA sequence, however, it remained unknown whether the tetraintercalator was indeed only specific for, and could locate that binding site preferentially when presented with a longer random sequence of DNA. Again, considering the molecular rearrangement necessary for complete association and/or dissociation from DNA, the ability of **3** to selectively "scan" the available DNA bases and specifically locate its designed binding site is not apparent.

In order to determine the specificity for **3**, two complementary 70mer oligonucleotide duplexes containing the designed binding site were ordered for ligation

into plasmid DNA as a binding site insert (**BS**) (**Figure 2.6**). Two plasmid vectors were chosen, pAK400[165] as shown in **Figure 2.7** which conveys chloramphenicol resistance, as well as pMoPac16[166], which is a variant of pAK400 that carries an ampicillin resistance gene and can be seen in **Figure 2.8**. Either plasmid represents a random composition of dsDNA nucleobases, similar to what any small molecule might encounter when binding dsDNA *in vivo*. Forward and reverse primers can be designed such that a PCR reaction provides a desired sequence of dsDNA of a particular length that also contains the binding site after ligation into the plasmid.

While tetraintercalator **3** was never optimized for work *in vivo*, it is important to verify that the molecule has the ability to distinguish its designed binding site from other random sequences of dsDNA in the event that future *in vivo* work is carried out, or if **3** shows potential as a scaffold for the development of future DNA binding molecules. We designed forward and reverse primers that would produce a sequence of dsDNA of approximately 400 bp, such that **3** can be presented with random dsDNA when locating its preferred binding site, yet still allowing for attainable DNase I footprinting of **3** bound to DNA.

Also included in the **BS** insert was the sequence $d(GGATCC)_2$, a known binding site for the bisintercalator **1.**[135,138] This serves the two purposes of acting as a calibration for concentration dependence and as an internal control during the DNase I footprinting experiments. Because it is part of the hybrid tetraintercalator sequence, we will also be able to see if there is any reasonable affinity for that particular sequence of dsDNA in addition to the affinity for the designed binding site.

(+) Strand

5' - CGG CCA CAT GCG GTA CCG ACA CGT AGT CGA CTA CAG ACG TAG TGA TAA GTA CTT ATC ACA TGG GCC TCG G - 3'

(-) Strand

3' - TCG GCC GGT GTA CGC CAT GGC TGT GCA TCA GCT GAT GTC TGC ATC ACT ATT CAT GAA TAG TGT ACC CGG A - 5'

Figure 2.6 Sequence of the 70mer DNA used as the **BS** insert into pAK400 and pMoPac16 vectors.[165,166] The sequences for the binding sites of compounds **1** and **3** are shown in red and the sticky ends are highlighted in green.

Figure 2.7 Vector map for pAK400.[165]

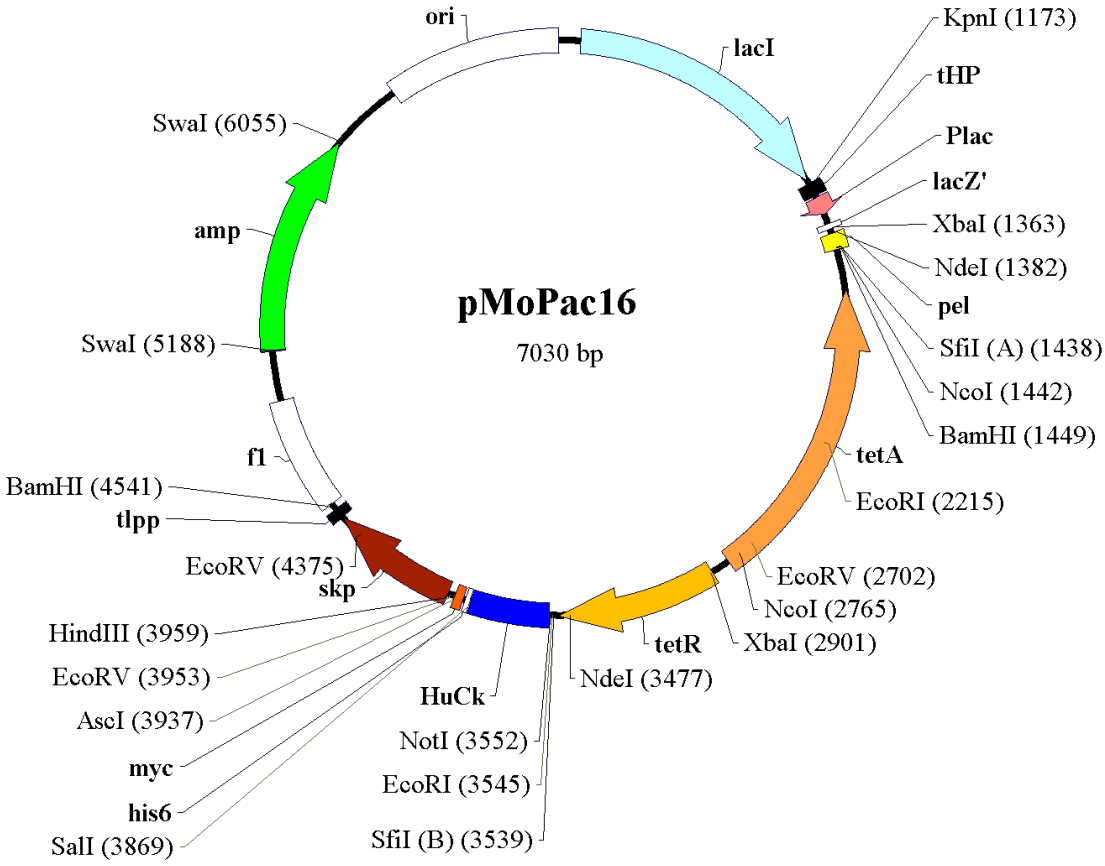
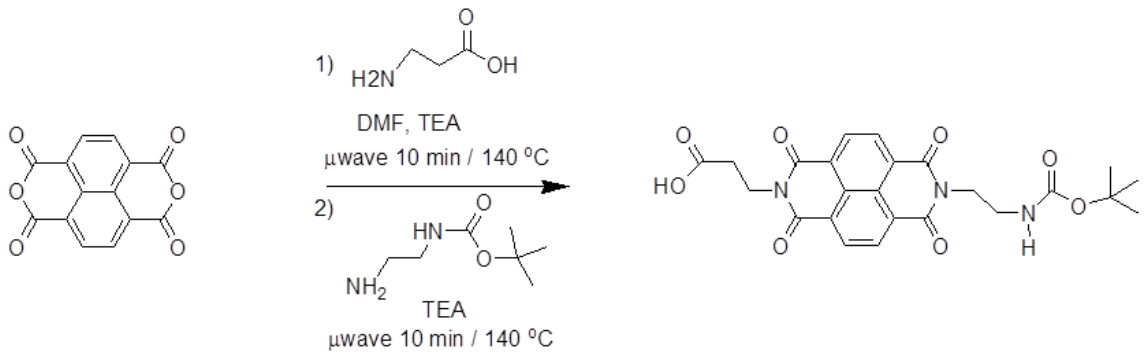


Figure 2.8 Vector map of pMoPac16.[166]

2.3.2 Synthesis

The synthesis of **3** began with a similar synthesis of the NDI monomer to the one outlined by Guelev *et al*.[167], however, a microwave reactor was used in place of heating at reflux as shown in Scheme 2.1. Following the reaction in the microwave reactor, synthesis continued on solid phase using Rink Amide resin (NovaBiochem) and an FMOC-based peptide synthesis protocol to produce the NH_2-Lys-NDI-$(\beta Ala)_3$-Lys-NDI-Gly-Resin dimer as described by Guelev and co-workers.[139] A prolonged incubation with limiting amounts of adipic acid, followed by cleavage from the resin using trifluoroacetic acid, produced the desired tetraintercalator **3**.[8] Purification of **3** was then achieved using preparative HPLC, resulting in an analytically pure sample.



Scheme 2.1: Microwave reactor synthesis of the NDI Monomer.

The vector pAK400 (**Figure 2.7**) was chosen for insertion of the binding site insert, **BS**, because it contains two *Sfi*I restriction sites. The *Sfi*I restriction enzyme cleaves dsDNA and produces a 3 base-pair "sticky end". The importance of the *Sfi*I restriction site is that it recognizes two successive sequences, 5'-GGCC-3', separated by five generic DNA bases, cleaving after the fourth generic base as read 5'→3' on both of the complement strands (**Figure 2.9**). The two *Sfi*I restriction sites in the pAK400 plasmid have an altered generic sequence between the two 5'-GGCC-3' recognition sites, which allows for directionality to be implemented into the **BS** insert, so that ligation of the insert into the plasmid after a *Sfi*I restriction digest can only occur in one direction because of the non-complementary sticky ends of the **BS** insert. The pAK400 vector was digested using the *Sfi*I enzyme followed by purification and extraction from an Agarose gel.

Both the forward and reverse strands of the **BS** insert were ordered. Annealing of the **BS** insert was then followed by ligation into the *Sfi*I digested pAK400 vector as shown in **Figure 2.10** to prepare the desired **pAK400-BS** vector containing the tetraintercalator **3** binding site.

Figure 2.9 Recognition and cleavage for *Sfi*I restriction enzyme.

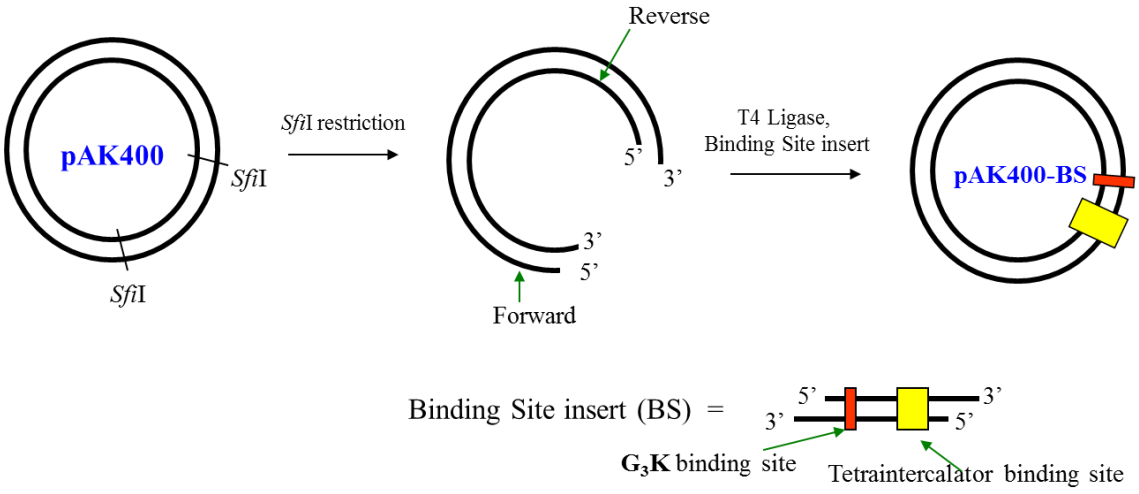


Figure 2.10 Cartoon showing **BS** insertion into *Sfi*I digested pAK400 to prepare **pAK400-BS**.

The same **BS** insert as mentioned above was also ligated into the vector pMoPac16 as shown in **Figure 2.11**. The pMoPac16 vector is a variant of the pAK400 vector containing similar restriction sites, however, it carries a resistance gene to the antibiotic ampicillin. Insertion of **BS** into pMoPac16 was done according to the same procedure as for the preparation of **pAK400-BS**, but use of the **pMoPac16-BS** vector offers a slightly longer and more varied sequence of DNA bases that can be presented to compound **3** during DNase I footprinting experiments.

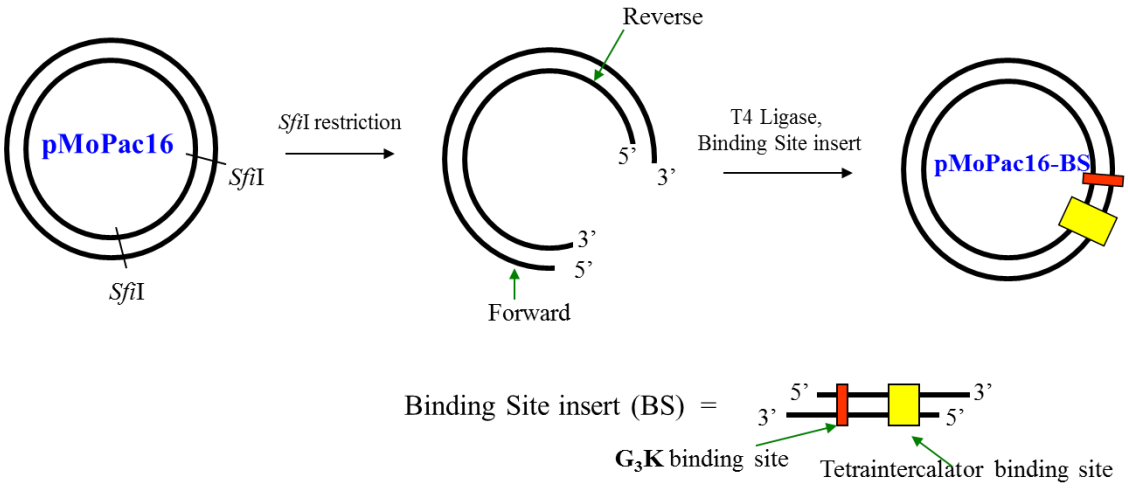


Figure 2.11 Cartoon showing **BS** insertion into *Sfi*I digested pMoPac16 to prepare **pMoPac16-BS**.

2.3.3 DNase I footprinting with the pAK400-BS construct

PCR primers for **pAK400-BS** (**Figure 2.12**) were designed such that the forward and reverse primers will produce a sequence of dsDNA which places the tetraintercalator **3** binding site approximately 75 – 100 bp from the 5' end of the forward strand. This allows for accurate tracking during the sequencing in an 8% denaturing PAGE gel via the xylene cyanol dye. The forward and reverse primer design also incorporate similar melting temperatures approximately 5°C above the temperature for the elongation phase of the PCR reaction, typically 55°C. The forward primer can then be 5'-^{32}P-phosphorylated using [γ-^{32}P]-ATP and T4 polynucleotide kinase according to the enzyme manufacturer's protocol. Purification of the radiolabeled forward primer is then accomplished by phenol-chloroform-isoamyl alcohol extraction and elution from a NICK™ column.

A PCR reaction (**Figure 2.13**) consisting of the **pAK400-BS** construct, ^{32}P end-labeled forward primer, reverse primer, dNTPs, and *Taq* DNA polymerase produces a 388mer PCR fragment of 5' end-labeled dsDNA containing the tetraintercalator **3** binding site. The DNA is then purified by a 5% native PAGE gel and extracted via the crush/soak method of elution. The (+) strand of the 388mer sequence is shown in **Figure 2.14**.

Forward Primer

5'- GGAAACAGCTATGACCATGATTACGAATTTCTAGAGAAG **– 3'**

Reverse Primer

5'- GCGAAAGGAGCGGGCG - **3'**

Figure 2.12 Sequences of the forward and reverse **pAK400-BS** primers.

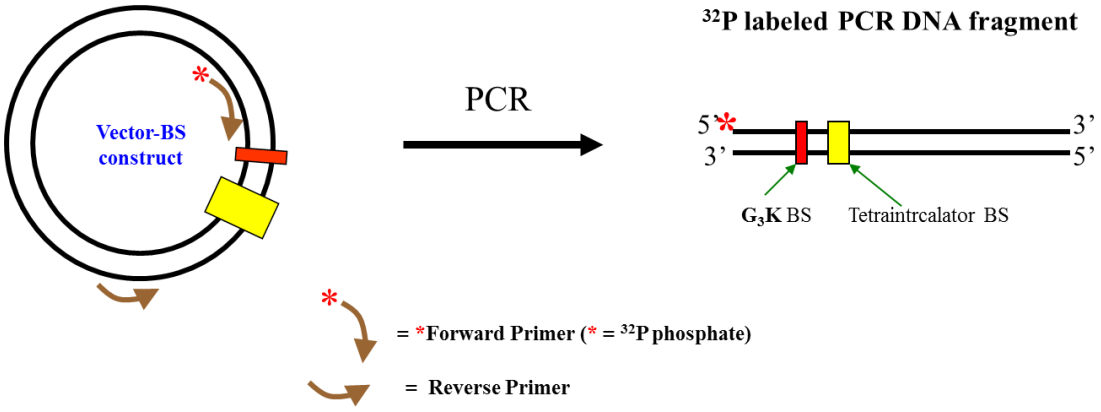


Figure 2.13 Cartoon of PCR amplification of 5', ^{32}P end labeled DNA containing both compound **1** and **3** binding sites as applied to either the **pAK400-BS** or **pMoPac16-BS** constructs.

pAK400-BS Sequence

5' – GGAAACAGCTATGACCATGATTACGAATTTCTAGAGAAGGAGATATACATATGAA

ATACCTATTGCCTACGGCAGCCGCTGGATTGTTATTACTCGCGGCCCAGCCGGCCAC

ATGCGGTACCGACACGTAGTCGACTACAGACGTAGTGATAAGTACTTATCACATGGG

CCTCGGGGGCCGATCACCATCATCACCATCATTAGTAAGCTTGACCTGTGAAGTGAA

AAATGGCGCACATTGTGCGACATTTTTTTTGTCTGCCGTTTACCGCTACTGCGTCAC

GGATCCCCACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGC

GCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGC - 3'

Figure 2.14 388mer DNA sequence of the (+) strand following PCR on **pAK400-BS** construct. Compound **1** and **3** binding sites are highlighted in red.

The DNase I footprinting data with the 388mer sequence is shown in **Figure 2.15**. As expected, bisintercalator **1** binds the sequence d(GGTACC)$_2$ at a concentration as low as 62.5 nM. Unexpectedly, a footprint also appears at the 12 bp GC rich sequence d(CGGCCCAGCCGG)·(CCGGCTGGGCCG), presumably as a pair of dimers. There is no observable footprint of **1** with the tetraintercalator **3** binding site.

The footprint of tetraintercalator **3** shows specificity only for its designed binding site d(GATAAGTACTTATC)$_2$ within the presented 388mer dsDNA of the **pAK400-BS** PCR fragment at room temperature. Specific binding is observed down to 62.5 nM, indicating a relatively high affinity for the preferred binding site. It is assumed that the same threading topology occurs upon binding the preferred site within the 388mer as

observed in the previous 2D NMR structural characterization towards the 14 bp binding site.[8] The same incubation was also run at identical concentrations, but at a temperature of 37°C, to see if there was an effective difference in specificity and/or affinity between the room temperature and physiological temperature incubations. The resulting footprint shows no observable difference in affinity for the designed binding site between the two temperatures, and specificity is still apparent at 37°C.

The 500 nM lane at room temperature shows potential problems with the digestion of the DNA. This result is unexpected from previous DNase I footprinting experiments with other threading intercalators in our lab, however as discussed in a later section, is occasionally an issue with the concentration dependent DNase I footprinting experiments at higher concentrations of **3**.

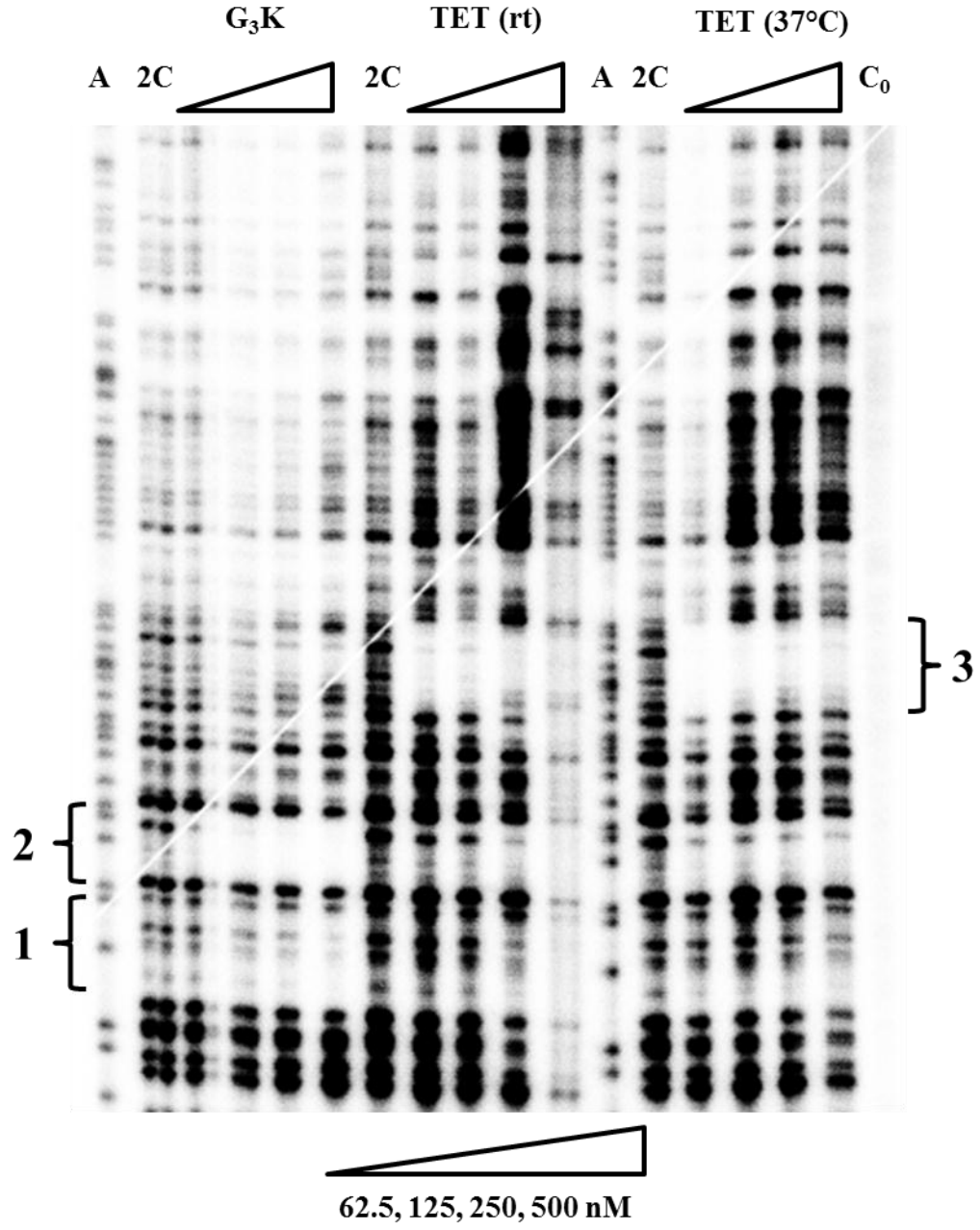


Figure 2.15 DNase I footprint of 388mer PCR fragment from **pAK400-BS** of compound **1** at room temperature and with **3** at both room temperature and 37°C respectively, both with incubation times of 24 hrs. Lane A represents an adenine-specific sequencing reaction. Lane 2C contains DNA digested with DNase I and no **3**. Lane C_0 contains DNA with no DNase I digestion or **3**. Sequences at 1: 5'-CGGCCCAGCCGG-3'; 2: 5'-GGTACC-3'; 3: 5'-GATAAGTACTTATC-3'.

2.3.4 DNase I footprinting with the pMoPac16-BS construct

PCR primers for the **pMoPac16-BS** vector were designed (**Figure 2.16**) using the same requirements as mentioned above for the **pAK400-BS** vector. In an identical PCR reaction to those discussed previously, the **pMoPac16-BS** vector can be used to produce a 5', ^{32}P end-labeled 467mer to be purified and eluted from a 5% non-denaturing PAGE gel. The (+) strand sequence of the 467mer can be seen in **Figure 2.17**.

Forward Primer

5'- GGAGATATACATATGAAATACCTATTGCCTACGGCAG -3'

Reverse Primer

5'- CTGACTCTCCGCGGTTGAAGC -3'

Figure 2.16 Sequences of the forward and reverse **pMoPac16-BS** primers.

pMoPac16-BS Sequence

5' – GGAGATATACATATGAAATACCTATTGCCTACGGCAGCCGCTGGATTGTTATTACT

CGCGGCCCAGCCGGCCACATGCGGTACCGACACGTAGTCGACTACAGACGTAGTGAT

AAGTACTTATCACATGGGCCTCGGGGGCCGAATTCGCGGCCGCTGCACCATCTGTCT

TCATCTTCCCGCCATCTGATGAGCAGTTGAAATCTGGAACTGCCTCTGTTGTGTGCCT

GCTGAATAACTTCTATCCCAGAGAGGCCAAAGTACAGTGGAAGGTGGATAACGCCCT

CCAATCGGGTAACTCCCAGGAGAGTGTCACAGAACAGGACAGCAAGGACAGCACCT

ACAGCCTCAGCAGCACCCTGACGCTGAGCAAAGCAGACTACGAGAAACACAAAGTCT

ACGCCTGCGAAGTCACCCATCAGGGCCTGAGTTCGCCCGTCACAAAGAGCTTCAACC

GCGGAGAGTCAG - 3'

Figure 2.17 DNA sequence of the 467mer (+) strand following PCR on **pMoPac16-BS** construct. Compounds **1** and **3** binding sites are highlighted in red.

The results of a concentration dependent DNase I footprinting gel on the 467mer **pMoPac16-BS** fragment with a concentration range of 31.25 nM – 500 nM over 48 hours can be seen in **Figure 2.18**. As observed for the footprinting experiment with the 388mer **pAK400-BS** fragment in **Figure 2.15**, tetraintercalator **3** shows specificity only for the preferred sequence d(GATAAGTACTTATC)$_2$, with a similar affinity between a concentration range of 62.5 nM – 500 nM. Again, we assume that this process occurs via the same threading polyintercalation binding topology as observed by Lee and co-workers.[8] The best complete footprints appear in the concentration range of 125 nM to 250 nM after an incubation time of 48 hours. Upon closer inspection, there appears to be inconsistency in the 500 nM lane, with a definite footprint at the designed binding site, but slightly less distinct than at 250 nM. This effect will be discussed further in a later section.

To analyze a time dependence of **3** binding its specific 14 bp binding site, two separate gels were run. 62.5 nM can be seen in **Figure 2.19**, where specific binding occurs between 36 and 48 hr. **Figure 2.20** shows the results of a time dependence incubation at 125 nM and 250 nM. The results show that at 62.5 nM, tetraintercalator **3** produces a footprint at the designed binding site, d(GATAAGTACTTATC)$_2$, between 36 and 48 hours. At 125 nM, the tetraintercalator **3** footprint at the designed binding site begins to appear between incubation times of 8 to 16 hours, and becomes distinct between 16 and 24 hours. The lanes for 250 nM show that a footprint begins to appear around 4 hours, and becomes a distinct footprint after 8 hours. All of the lanes indicate that the **3** continues to show specificity only for its designed binding site amongst the 467mer dsDNA from the **pMoPac16-BS** vector.

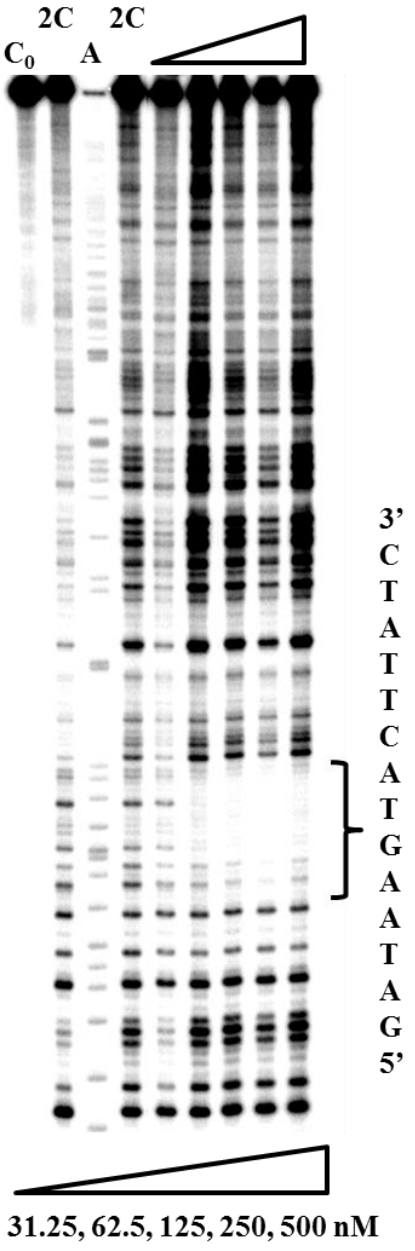


Figure 2.18 DNase I footprint of a concentration dependent experiment with a 467mer PCR fragment from **pMoPac16-BS** incubated with **3** for 48 hr, and digested for 4 min. Lane A represents adenine-specific sequencing reaction. Lane 2C contains DNA with DNase I but no added **3**, with the first 2C lane (from left to right) being digested 7 min, and the second 2C lane digested 4 min. Lane C_0 contains DNA with no DNase I or **3**.

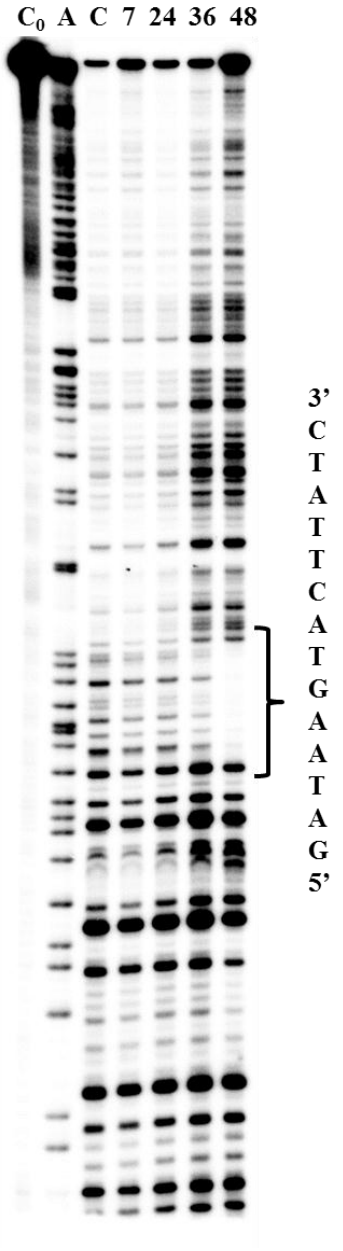


Figure 2.19 DNase I footprint of 467mer PCR fragment from **pMoPac16-BS** incubated with 62.5 nM of **3**. The time-trial is in hours. Lane A represents an adenine-specific sequencing reaction. Lane C contains DNA with DNase I digestion but no **3**. Lane C_0 contains DNA with no DNase I digestion and no **3**. All other lanes represent incubation times in hours.

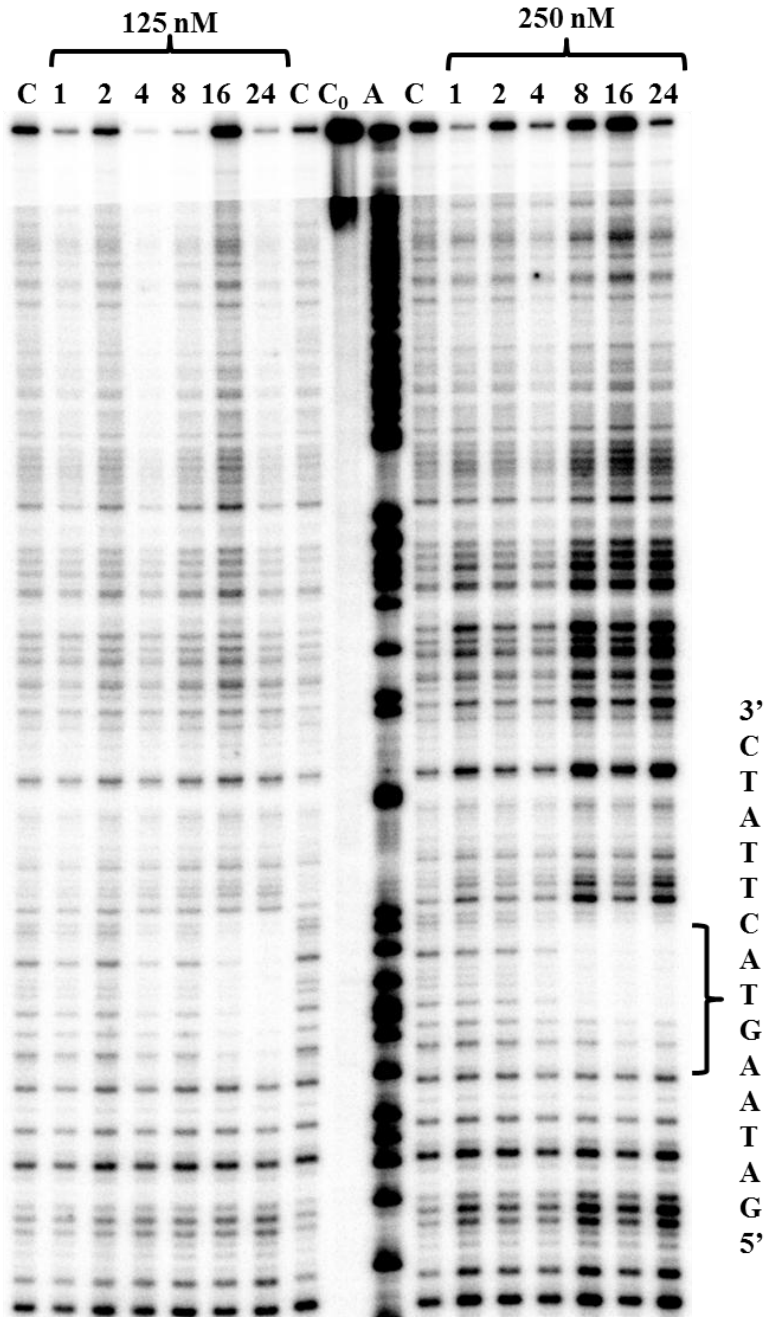


Figure 2.20 DNase I footprint of 467mer PCR fragment from **pMoPac16-BS** incubated with **3**. Time-trial in hours at both 125 nM and 250 nM of **3**. Lane A represents an adenine-specific sequencing reaction. Lane C contains DNA with DNase I digestion but no **3**. Lane C_0 contains DNA with no DNase I digestion and no **3**. All other lanes represent incubation times in hours.

2.4 DISCUSSION

2.4.1 Description of DNase I footprinting with 388mer from pAK400-BS construct

As stated previously, the bisintercalator **1** binding site was included in the binding insert primarily for use as a concentration calibration and internal control for the DNase I footprinting experiments. The results as shown in **Figure 2.15** show that bisintercalator **1** does indeed bind to the sequence $d(GGTACC)_2$ with high affinity, but also showed a slight affinity for the 12 base pair sequence 5'-CGGCCCAGCCGG-3'. While this result was not expected, it is not surprising given previous work in our lab showing the affinity of **1** for GC rich DNA.[135,137] Although this sequence had not previously been tested, it contains similar elements to previous DNase I footprinting experiments for which bisintercalator **1** showed a slight affinity. It is likely that, as proposed by Lokey *et al*.[135], some degree of cooperativity in binding exists, due to the nature of the compound **1** footprint spanning l2 bases of a GC rich DNA sequence. Interestingly, **1** shows no particular affinity for the tetraintercalator **3** binding site $d(GATAAGTACTTATC)_2$, even though the binding site contains the sequence 5'-GTAC-3', which the bisintercalator **1** – $(Gly)_3$Lys- linker is known to recognize via its interactions in the major groove of DNA.

Figure 2.15 shows that the footprinting data for tetraintercalator **3** to its designed 14 base pair binding site, $d(GATAAGTACTTATC)_2$, is highly specific. Binding on the gel occurs within a dynamic concentration range of 62.5 nM to 500 nM. This result shows the ability of **3** to scan available DNA bases and locate a preferred binding site within a longer sequence of dsDNA. The molecular rearrangement necessary for association and/or dissociation when complete polyintercalation occurs appears to be a minimal factor with respect to tetraintercalator **3** locating its preferred binding site. This could indicate that complete polyintercalation of the tetraintercalator does not fully occur with non-specific sequences of DNA. Another interpretation could infer that the

dissociation rates from the specific interactions at the preferred binding site are much slower than those dissociation rates corresponding to non-specific binding. A detailed kinetic analysis is presented in a later chapter.

We also ran the incubations for tetraintercalator **3** at both room temperature and at 37°C for 24 hours to see if there were any temperature-related differences in affinity or specificity. The observed results indicate that there is no difference in the binding affinity for the sequence 5'-GATAAGTACTTATC-3'. Also, specificity is maintained for the tetraintercalator towards its designed binding site.

2.4.2 Description of DNase I footprinting of 467mer from pMoPac16-BS construct

The footprint of tetraintercalator **3** on the 467mer from **pMoPac16-BS** shown in **Figure 2.18** is consistent with the previous 388mer footprint for a high specificity towards the designed 5'-GATAAGTACTTATC-3' binding site. The 467mer originates from PCR of the **pMoPac16-BS** construct, which presents different possible sequences of DNA bases to compound **3**. These data, taken together, indicate that **3** is capable of locating its preferred binding site and specifically distinguishing it from other possible sequences of DNA upon incubation. Also inferred from these results, is that the molecular rearrangement necessary for dissociation from non-specific sequences does not apparently interfere with the ability of the tetraintercalator to specifically locate and associate with its designed binding site. As stated previously, this might indicate that complete polyintercalation does not occur with respect to interactions with other sequences of DNA. Also possible, and more likely the case, is that the dissociation rates from the specific binding site are presumably much slower than the dissociation rates from non-specific interactions.

Most often, complete footprints become distinct around 125 nM for our standard 24 – 48 hr incubation times. The results of the time dependent study in **Figures 2.19** & **2.20** show that beginning at 62.5 nM, the tetraintercalator **3** footprint begins to appear between 36 and 48 hours, with the eventual distinction of a complete footprint at 48 hours. The 125 nM time trial results show an initial footprint beginning between 8 and 16 hours, with distinction between 16 and 24 hours after the incubation begins. The time dependent footprint was also completed at a 250 nM concentration of **3**. A footprint is visible in a shorter time, between 2 to 4 hours after incubation at 250 nM, as would be expected at a higher concentration of tetraintercalator **3**. Also occurring more quickly is a complete footprint at the designed binding site after only 8 hours of incubating at 250 nM. These results allow us to visualize the complete intercalation taking place on a time scale of hours when locating a preferred binding site out of nearly 500 DNA base pairs. Recall that for the 48 hr concentration dependence study mentioned above, a distinct footprint was observed over the effective range of **3** between 62.5 and 500 nM. Knowing that the footprint becomes distinct after 36, 16, and 8 hours for 62.5 nM, 125 nM, and 250 nM, respectively, it is logical to conclude that the dissociation from the preferred binding site is extremely slow provided that the full footprint remains even after a 48 hr incubation.

Again, a detailed kinetic analysis detailing the association and dissociation rates from a preferred binding site oligonucleotide will be discussed in a following chapter. However, this should not discount the data observed for **3** to specifically locate its preferred binding site from among nearly 500 DNA bases. It is logical to assume that specific binding would take longer when discriminating increasing lengths of DNA base pairs. It is also not unreasonable to observe that the association rates with the designed binding site are relatively slow, provided previous works monitoring a variety of

intercalators slow association rates with DNA.[134,168,169,152] Given the amount of molecular rearrangement that must occur for complete threading intercalation, we would anticipate presumably slow association rates as well as slow dissociation rates from the DNA.

Sometimes observed in concentration dependent studies of the 388mer and 467mer (**Figures 2.15** & **2.18**), when compound **3** approaches 500 nM, some lanes and footprints appear inconsistent when compared to other lower concentrations. Other, non-published, work in our lab led us to propose that this is most likely due to having reached too high a concentration of DNA intercalator with too many charges associated with a particular molecule. We have hypothesized that because of the initial non-specific electrostatic interactions occurring between the positively charged lysine side chains of our intercalators and the negatively charged DNA phosphate backbone, our molecules can coat the DNA, which would interfere with the DNase I digestion. The predicted result would be lanes that are inconsistent, usually appearing to have too high radio-label loading due to a lack of digestion, leading to higher concentrations of longer sequences of DNA that are undigested. While footprints are still present at the preferred binding sites, the inconsistent nature of the 500 nM lanes leads to an interesting question concerning the amount of positive charges, as well as where they are located within our molecules. More positive charges per polyintercalator unit might increase the non-specific electrostatic interactions with random sequences of DNA, thereby limiting specificity and hindering DNase I digestion. Also, as a result of salt dependent dissociation kinetic studies discussed in a later chapter, we believe that charge placement might alter dissociation from specific binding sites as a more complex correlated mechanism to that proposed for dicationic NDI intercalators by the Wilson lab.[134]

2.5 CONCLUSIONS

This work shows that tetraintercalator **3** is highly specific for its designed 14 bp binding site d(GATAAGTACTTATC)$_2$. The ability to discriminate between available DNA sequences in preference for its designed binding site has been shown via DNase I footprinting experiments with both a 388mer and 467mer sequence of dsDNA. The observed time for binding site recognition from approximately 500 base pairs is approximately 4-8 hours at 125 nM and 2-4 hours at 250 nM. The distinct footprint remaining after 48 hours would presume an extremely slow dissociation process from its preferred binding site. This observation is not surprising considering the molecular rearrangement necessary for complete dissociation from the 14 bp designed binding site.

This work further supports the results of Lee *et al.*[8], which has detailed the ability of **3** to bind its designed 14 bp binding site with a 1:1 stoichiometry via a threading polyintercalation binding mode. It is assumed that the same threading binding topology is maintained during binding of the preferred site within the longer sequence of DNA. The modularity of this system proves a potentially useful method for the development of future polyintercalators as small molecules able to search out and specifically bind dsDNA.

Work is currently underway in our laboratory to obtain specificity data as it applies to a specific binding site located within a linearly cut plasmid, as well as the ability of **3** to block transcription at a particular restriction site. Preliminary results obtained by a fellow group member are promising with respect to the ability of **3** to locate, and block transcription at its designed binding site.

We are also in the process of developing even larger threading polyintercalators based off of combining tetraintercalator scaffolds to bind even longer sequences of dsDNA. This is being done in concert with examining the quantity and location of

positive charges contained within our threading polyintercalators, and the resulting specificity and kinetic effects on the respective ability to bind dsDNA.

2.6 MATERIALS AND METHODS

2.6.1 General

Materials. Rink Amide resin (0.72 mmol/g), PyBOP, and all necessary Fmoc protected amino acids were purchased from NovaBiochem. All enzymes were purchased from New England Biolabs (Beverly, MA) and DNA manipulations were carried out according the manufacturer's provided protocols unless otherwise noted. QIAprep® Spin plasmid purification, QIAquick® Gel extraction, and QIAquick® PCR purification kits were purchased from QIAGEN (Valencia, CA). All growth media were purchased from Fisher. All other chemicals were purchased from Aldrich, Acros, or Fisher unless otherwise noted. All [γ-^{32}P]-ATP was purchased as 0.250 millicure from Perkin-Elmer Life & Analytical (Waltham, MA). All DNA oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA) purified by desalting or polyacrylamide gel electrophoresis. The plasmid vectors pAK400 and pMoPac16 were graciously provided by the Iverson and Georgiou labs (Institute for Cellular and Molecular Biology, University of Texas at Austin). The Jude1 (DH10B) strain of *E. coli* was provided by the Iverson and Georgiou labs (University of Texas at Austin).

General Procedures. ^{1}H NMR spectra were recorded on Varian 300 or 500 MHz spectrometer. Silica gel 60 F254 glass-based plates (Merck) were used for TLC. Flash chromatography was carried out using SiliCycle SiliaFlash P60 40-63 μm 60Å silica gel. Autoradiography was performed on a GE Typhoon Trio^{+} phosphor imager. Di-deoxy sequencing was performed using a capillary based AB 3730XL DNA analyzer and

sequenced using the BRHO6 primer (5'-GCGGATAACAATTTCAGACAG-3'). UV-Vis spectroscopy was performed on either a Perkin-Elmer Lambda 35 spectrophotometer, an Agilent 8453 spectrophotometer, or an ND1000 Nanodrop spectrophotometer for synthetic or plasmid oligonucleotides. Analytical and semi-preparative HPLC was performed using a Waters HPLC system equipped with a Waters 2996 photodiode array detector and Vydac protein and peptide C18 preparative column were solvent A was 0.1% TFA in ddi H_2O and solvent B was 0.1% TFA in HPLC grade acetonitrile.

2.6.2 Synthesis

N-(2-*tert*-Butoxycarbonylaminoethyl)-N'-(2-carboxyethyl)-1,4,5,8-naphthalenetetracarboxylic diimide (NDI monomer). The NDI monomeric unit was synthesized as described by Guelev *et al*.[167] excepting the use of a microwave reactor to replace heating at reflux (Scheme 2.1). β-Alanine (0.094 g, 1.0385 mmol) was sonicated in 16 ml DMF to effect dissolution. 1,4,5,8-naphthalenetetracarboxylic dianhydride (0.2786 g, 1.0385 mmol) was added to the solution, followed by the addition of TEA (0.15 ml, 1.0385 mmol), the mixture was transferred to a 3-neck 100 ml round bottomed flask. The mixture was heated in a microwave reactor (Mars) at power = 600W, max = 100%, and spin = 3. The procedure followed a 2 min ramp to 140ºC constant temperature heating for 10 min, then cooling to 65ºC. A mixture of tert-butyl 2-aminoethylcarbamate (0.1604g, 1.0385 mmol) and TEA (0.15 ml, 1.0385 mmol) was then added to the 3-neck round bottom flask. The flask was placed back into the microwave reactor and the same settings and procedure were used as previously stated. The flask was cooled to room temperature and the solvent removed in vacuo. The crude product was dissolved in CH_2Cl_2/TEA (99:1) and purified by column chromatography with 0-10% MeOH gradient

in 10% TEA in CH_2Cl_2. The product fraction was concentrated and 10 ml of HOAc was added to the 10 ml re-dissolved solution of 10% MeOH/ CH_2Cl_2 in order to neutralize the product. 200 ml of hexanes were added to facilitate precipitation, then the solution was filtered and washed with hexanes, followed by drying in vacuo to yield the product as a tan, flaky solid (0.0851g, 17%). ^{1}H NMR (500 MHz, DMSO-d6): δ 12.40 (br s, 1H, COO*H*), 8.64 (s, 4H, aromatic), 6.92 (t, 1H, N*H*CO, J = 6.3 Hz), 4.27 (t, 2H, CH_2, J = 7.5 Hz), 4.14 (t, 2H, CH_2, J = 5.7 Hz), 3.27 (br q, 2H, CH_2-(NH)), 2.62 (t, 2H, CH_2(COOH), J = 7.5 Hz), 1.20 (s, 9H, 3CH_3); HRMS-CI predicted for $C_{24}H_{24}N_3O_8$[M+H]+: 482.1558, found 482.1577.

Tetraintercalator 3 Synthesis. The Fmoc-Lys-NDI-(β-Ala)$_3$-Lys-Ndi-Gly-resin was prepared as previously described by Guelev *et al*.[139] excepting the use of Rink Amide resin, and thereafter followed the procedure of Lee *et al*. for the complete synthesis of the tetraintercalator[8] (**Figure 2.4**). The resulting product was purified using HPLC at 14 ml/min over the following gradient: 0-2 min at 100% A; 2-9 min to 82% A; 9-70 min to 77% A. The product eluted at 53 minutes and lyophilization of pure fractions (**Figure 2.21 (A)** & **(B)**) gave the tetraintercalator as a white powder. ^{1}H NMR spectra were in agreement with the previously published data[8] and HRMS-ESI predicted for $C_{128}H_{150}N_{30}O_{34}[M+2H]^{+2}$: 1325.5466, found 1325.8, predicted for $C_{128}H_{151}N_{30}O_{34}[M+3H]^{+3}$: 884.0336, found 884.5, and predicted for $C_{128}H_{152}N_{30}O_{34}[M+4H]^{+4}$: 663.681, found 663.6.

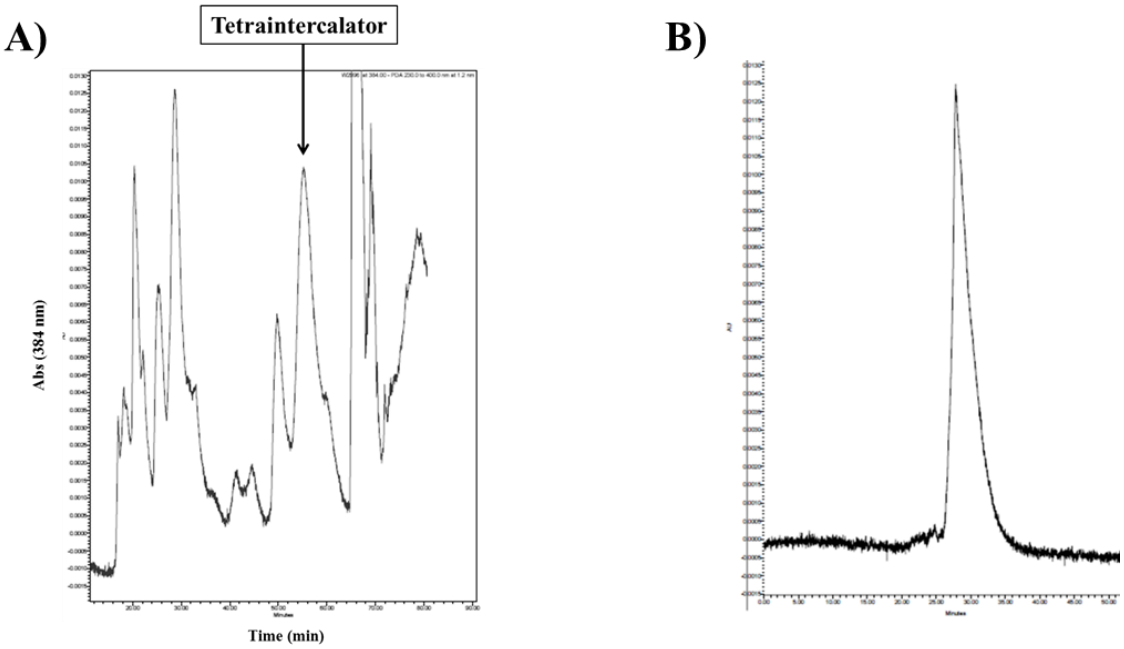


Figure 2.21 A) Crude and B) Analytically pure HPLC traces of tetraintercalator **3**.

2.6.3 pAK400-BS single binding site construct

For the construction of **pAK400-BS** (**Figure 2.10**), 0.5 ml of the Jude1 (DH10B) strain of *E. coli* (graciously given by the Georgiou & Iverson labs, University of Texas at Austin) contains the pAK400 plasmid[165] was added to 20 ml of Terrific Broth (TB) media containing 2% glucose and chloramphenicol (30 μg/ml). The liquid culture was incubated at 37ºC overnight on a shaker, removed, and pAK400 was isolated using a QIAprep® spin plasmid miniprep kit. The concentration was determined using a nanodrop spectrophotometer. pAK400 (50 ng/μL) was digested with SfiI restriction enzyme in a 50 μL reaction volume for 4 hr at 50ºC, according to the protocol provided with the enzyme. The restriction fragments were then purified by a 1% agarose gel at a constant 110V, and visualized using UV light. The pAK400 fragment was extracted using a

QIAquick® gel extraction kit, and resuspended in 150 μL ddi H_2O. The concentration was determined by nanodrop as mentioned above. The forward and reverse strands of the designed *Sfi*I compatible binding insert (**Figure 2.6**) containing bisintercalator **1** and tetraintercalator **3** binding sites (**BS**) were ordered from Integrated DNA Technologies as PAGE purified and 5' phosphorylated oligonucleotides. Both strands were re-suspended to a stock concentration of 100 μM. The two strands were annealed at a concentration of 12.5 μM of double stranded DNA, in a thermal-cycler (Applied Biosystems 2720 Thermal-Cycler) at 95ºC for 5 min, and allowed to cool at a rate of 1ºC per min to room temperature. Ligation of the binding insert into *Sfi*I digested pAK400 was done at a 3:1 ratio, respectively, with 100 ng of vector per reaction in a total volume of 20 μL. 1.0 μL (400 cohesive units) of T4 DNA Ligase enzyme was added, and ligation took place at room temperature for 4 hr, followed by a 20 min incubation at 65ºC to heat-kill the ligase according to the protocol provided with the enzyme. The ligated vector (**pAK400-BS**) was then desalted using a desalting membrane (Millipore) in ddi H_2O. 1 μL of the desalted vector was electroporated (1.8 KV) into 35 μL Jude1 (DH10B) *E. coli* using a 1.0 mm wide electroporation cuvette (VWR), and recovered into 1.0 ml SOC media at 37ºC for 1 hr. SOC/chloramphenicol (30 μg/ml)/Agar plates were poured and spotted with the recovered Jude1 *E. coli* cells, then placed upside down in a 30ºC oven and grown overnight. A single colony was transferred to 2.0 ml TB with 2% glucose and chloramphenicol (30 μg/ml) and incubated overnight at 37ºC. **pAK400-BS** was isolated using a QIAprep® spin plasmid miniprep kit and the concentration was determined by nanodrop. Successful ligation of the binding site insert into pAK400 was confirmed by di-deoxy sequencing.

2.6.4 pMoPac16-BS single binding site construct

For the construction of **pMoPac16-BS** (**Figure 2.11**), pMoPac16[166] was electroporated into a Jude1 (DH10B) *E. coli* strain as described above for **pAK400-BS**, but ampicillin (100 μg/ml) was used for selection. A 3.0 ml TB/2% glucose liquid culture of Jude1 cells containing pMoPac16 was incubated at 37ºC overnight. pMoPac16 was isolated using a QIAprep® spin plasmid miniprep kit and the concentration determined by nanodrop. pMoPac16 was digested by a *Sfi*I enzyme and purified as explained for **pAK400-BS**. The forward and reverse strands of the binding site insert **BS** used for ligation into pAK400 were annealed and cloned into the vector pMoPac16 as described for **BS** ligation into pAK400. The concentrations were verified by nanodrop. Successful ligation was confirmed for **pMoPac16-BS** by di-deoxy sequencing.

2.6.5 Preparation of 5' – ^{32}P end labeled DNA fragments

Forward and reverse primers for either **pAK400-BS** or **pMoPac16-BS** were ordered and re-suspended to 100 μM stock solutions in ddH_2O. 20 μM working solutions in ddH_2O were then prepared. 5'-^{32}P end-labeling proceeded as a slightly modified protocol from Dervan *et al*.[170] for 1 hour at 37°C in a 100 μL reaction volume as follows: 3 μL (20 μM) forward primer; 10 μL 10X T4PNK Buffer; 56 μL ddH_2O; 25 μL (41.75 pmol, 250 μCi) [γ-^{32}P]-ATP; 6 μL (20 units) T4 polynucleotide kinase. After end-labeling, 5 μL of 0.5 M EDTA was added, the end-labeled primer was then extracted four times with 100 μL (25:24:1) (v/v/v) of phenol-chloroform-isoamyl alcohol. The aqueous layer was then purified on a NICK™ column according to the manufacturer's protocol and collected as a 400 μL aliquot. To the aliquot was added 24 μL of 4M NaCl, 1 μL glycogen (20 mg/ml), and 860 μL ice cold EtOH. The sample was vortexed and placed in dry ice 15 min, followed by centrifugation at 13.2K rpm for 25 min. The pellet was washed with 100 μL 70% EtOH in ddH_2O, followed by centrifugation at 13.2K rpm for

2 min, and the pellet was allowed to air dry for 10 min. To the pellet was added 3 µL (20 µM) reverse primer and 57 µL ddiH_2O.

PCR was then used to amplify a 388mer or 467mer fragment from **pAK400-BS** or **pMoPac16-BS** respectively, at a reaction volume of 100 µL as follows: 50 µL of forward/reverse primer solution; 32 µL ddiH_2O; 10 µL Thermopol buffer; 5 µL (4 mM) dNTPs; 1 µL 100X BSA; 1 µL (12 ng/µL) plasmid. Heat the sample at 70°C 5 min, 1 µL (5 units) *Taq* DNA polymerase was added, place in thermal-cycler according to the following reaction conditions for PCR: 30 cycles of [94°C 60 s, 55°C 60 s, 72°C 90 s]; 72°C 10 min; 4°C hold. The fragment was then purified using 5% native PAGE at 75 W and ≤ 250 V. The 5', ^{32}P end labeled fragments were then extracted from the gel by exposing to a phosphor screen for 3 min, and the bands were removed by cutting the gel with a razor. (Note: some instances allow for the placement of the gel on a TLC plate, and cutting the bands from the gel using a UV light for visualization of the PCR DNA product) Autoradiography was used to verify that the labeled fragments were removed completely, followed by using a crush/soak method in 800 µL 1X TE buffer in a shaker at 37°C overnight. Ethanol precipitation with $1/10^{th}$ volume of 3M sodium acetate and 3x volume of 100% EtOH was placed on dry ice for 15 min, followed by centrifugation at 13.2K rpm for 25 min. The pellet was washed with 300 µL of 70% EtOH in ddiH_2O, spun at 13.2K rpm for 7 min, the pellet was allowed to air dry for 15 min. The pellet was re-suspended in 400 µL of 20 mM Na_2PO_4 with 2 mM $MgCl_2$. The radiolabel was quantified using a Geiger counter.

2.6.6 DNase I Footprinting

The incubations for DNase I footprinting were performed at a total volume of 24 µL as follows: 4 µL of tetraintercalator to provide desired final concentration; dilute

stock 5' end-labeled DNA until 20 μL reads at ~ 20,000 counts per minute (cpm). Incubate at room temperature for desired times and/or concentrations. A DNase I solution[170] was then prepared to a total volume of 87 μL as follows: 17.5 μL ddi H_2O; 2.5 μL DNase I buffer; 5 μL DNase I (4 units per μL); 62 μL $NaPO_4$(20 mM)/$MgCl_2$(2 mM) buffer. A DNase I Stop solution[170] was then prepared to a total volume of 1400 μL as follows: 40 μL glycogen (20 mg/ml); 40 μL sonicated calf thymus DNA (1 mM bp); 107 μL ddi H_2O; 788 μL NaCl (4 M); 425 μL EDTA (0.5 M, pH 8.0). The incubations were then digested by adding 2 μL of DNase I solution for 4 min. 50 μL of DNase I stop solution was then added, followed by 314 μL of ddi H_2O. 975 μL of room temperature EtOH was added to the solution, followed by centrifugation at 13.2K rpm for 25 min. The solution removed and the pellet was washed with 300 μL of 70% EtOH in ddi H_2O, followed by centrifugation at 13.2K rpm for 3 min. The resulting pellet was then re-suspended in 15 μL ddi H_2O, frozen in liquid nitrogen, and lyophilized overnight to facilitate re-suspension in denaturing formamide buffer. The resulting digested DNA was separated using 8% denaturing PAGE at 75 W and ≤ 2000 V. The sequencing gels were extracted from the plates using 3 mm Whatman chromatography paper (Fisher) and dried for 45 min on a Model 583 gel dryer from Bio-Rad (Hercules, CA). After drying, the gels were exposed on a phosphor screen (18 - 24 hr), and imaged using either Quantity One 4.5 software from Bio-Rad (Hercules, CA) or ImageQuant TL software from GE Healthcare (Piscataway, NJ).

Chapter 3

Binding Kinetics for a Threading Tetraintercalator

3.1 Chapter Summary

3.1.1 Goals

Determine the apparent k_{on} and k_{off} of tetraintercalator **3** for both its designed 14 bp binding site and a random sequence of dsDNA. Use these data to explain the observed sequence specificity and determine a salt dependence for dissociation from the designed 14 bp preferred site. Taken together, the determined values for k_{on} and k_{off} can be utilized to propose an overall apparent dissociation constant, K_D.

3.1.2 Approach

A multi-stage approach was used to determine apparent values for the association of tetraintercalator **3** to its preferred binding site. 1D ^{1}H NMR was used for a structural verification while monitoring association of **3** to its preferred binding site. Gel mobility-shift assays were then used to monitor association, providing a higher resolution study than NMR. While structural verification is absent from gel mobility-shifts, the non-covalent interactions must still be strong enough to produce a gel shift. Stopped-flow UV-vis spectroscopy was also used to monitor association of **3** with both its preferred binding site and a non-target sequence of DNA.

Gel mobility-shift assays were also used to monitor dissociation of **3** from its preferred binding site under a variety of [Na^+] concentrations. Stopped-flow UV-vis spectroscopy was used for monitoring dissociation of **3** from a non-target sequence of DNA using surfactant sequestering techniques to monitor complete dissociation.

3.1.3 Results

Association of **3** with its preferred binding site was successfully monitored using 1D ^{1}H NMR, gel mobility-shift assays, and stopped-flow UV-vis spectroscopy. The results from the NMR studies produced structural verification of association which was complete after the first scan, thus providing a lower limit to association between **3** and a preferred binding site equivalent to k_{on} = 9.4 x 10^{2} M^{-1} s^{-1}. Gel mobility-shift assays for association using a concentration range of **3** and preferred binding site of 250 nM to 2500 nM, produced an estimated k_{on} = 2.7 ± 0.5 x 10^{5} M^{-1} s^{-1}. This value is about an order of magnitude slower than that obtained by stopped-flow UV-vis, with an apparent association rate between **3** and its preferred binding site of k_{on} = 3.4 ± 0.1 x 10^{6} M^{-1} s^{-1} for the faster step of a four step model. This value was slightly faster than association between **3** and a non-target sequence obtained by stopped-flow equal to k_{on} = 2.5 ± 0.1 x 10^{6} M^{-1} s^{-1} for the faster step of a three step model.

Dissociation of **3** with its preferred binding site was monitored using gel mobility-shift assays where **3** was competed off using excess unlabeled preferred binding site. The resultant dissociation rates at varying [Na^{+}] from 25 mM to 200 mM ranged from k_{d} = 2.3 ± 0.3 x 10^{-7} sec^{-1} to k_{d} = 9 ± 1.5 x 10^{-7} sec^{-1}, respectively. Dissociation from non-target DNA was monitored by stopped-flow UV-vis and found to be equal to k_{d} = 0.019 s^{-1} for the slower of a two-step dissociation process.

Under these conditions, an apparent overall dissociation constant using a benchmark 100 mM NaCl concentration and stopped-flow association rates provide K_{D} = 7.6 x 10^{-9} M (7.6 nanomolar) for **3** with a non-target sequence and an overall apparent K_{D} = 1.5 x 10^{-13} (150 femtomolar) for **3** with its preferred binding site.

3.2 INTRODUCTION

The kinetics of association and dissociation for a variety of DNA binding molecules have been investigated and can serve as a useful benchmark for understanding the kinetics surrounding threading polyintercalation (**Table 1.3**, Ch. 1). The significance of experimentally determining the binding kinetics for drug-DNA interactions revolves around the ability to propose possible reaction mechanisms that can elucidate the molecular details of binding.[107] For example, Chaires *et al*. had shown that the kinetics of interaction between daunomycin and calf thymus DNA raised the possibility of a multi-step association process for eventual intercalation at preferred sites on the DNA. In this process, three steps were proposed that might correspond to the rapid formation of an initial "outside" complex, which would then be followed by drug intercalation between the DNA base pairs, followed by either the conformational adjustment of the drug-DNA complex or the possible redistribution of the bound drug to preferred sites via an internal transfer mechanism. The resulting data provided an overall binding constant for daunomycin ranging from $10^6 – 10^7$ M^{-1}.[107]

Many intercalators are positively charged molecules. So while intercalation has its own desolvation driving force, there is an electrostatic component of binding which can be better understood by studying the salt dependence of binding to DNA. Several studies from other laboratories are detailed below, and demonstrate some of the extensive work performed with respect to the salt dependence on DNA binding. At this time, we do not understand enough about the individual steps involved in the association of compound **3** to its preferred binding site, so the salt dependence effects on association are not of particular value to our current work. However, future elucidation of the mechanism of association might make useful the data, and provide us the ability to partition an

electrostatic component of binding which justifies the coverage of salt dependence on DNA binding molecules.

Manning's polyelectrolyte theory proposed that monovalent salts, or counterions, were capable of condensing onto polyanions until the charge density is below some critical value for the solution, typically with a charge density close to unity.[33] Record *et al.* then applied this theory to obtain a general thermodynamic analysis of the monovalent salt effects on the observed association constants, K_{obs}, surrounding ligand-DNA interactions. Their findings indicated that the condensation of approximately 0.88 Na^+ ions occurs per phosphate on the native DNA polyanion, however, the full extent of Na^+ binding is determined only by the axial charge density of the DNA and not by the $[Na^+]$.[34] By analyzing a variety of data, Record *et al.* found that the formation of ion pairs from ligand-DNA interactions releases the condensed counterions, and that the entropic effect of this counterion release provides a major contribution to the overall free energy of formation between the ligand-DNA complexes.[34] Moreover, it was shown that there is a correlation in the log K_{obs} versus –log $[Na^+]$ plots, where the slope of the linear fit is equivalent to the number of counterions released, thereby indicating the number of electrostatic interactions taking place during binding.[34] Lohman *et al.* have also been instrumental in the analysis of understanding the salt effects on the kinetics of proteins that bind to DNA.[171] The basis for all of these findings has led to many studies focusing on understanding the salt dependence on the kinetics of binding between a variety of ligand-DNA interactions.

Wilson and co-workers have produced extensive results detailing the monovalent salt dependence of interactions between small molecules that bind to DNA. They have found that the monocationic intercalator ethidium produces a slope of approximately 1 in a log K_{obs} versus –log $[Na^+]$ plot, thus indicating formation of one ion pair between the

DNA and intercalator. In the same study, they observed that propidium, a similar dicationic intercalator, produced a slope from the above mentioned plot equal to 2, corresponding to the formation of two ion pairs as was expected for the molecule.[105] A proposed model for the system suggests the formation of a loosely bound electrostatic ligand-DNA complex, followed by the formation of an intercalation pocket in the DNA and subsequently followed by ligand intercalation which can be used to estimate the salt dependent dissociation rates. A linear relationship was observed for log k_d versus –log [Na^+] for dissociation, with slopes equal to 0.31 and 0.85 for ethidium and propidium respectively. While the slopes for the dissociation kinetic plots do not directly equal the number of counterions released (or ion pairs formed), it was observed that the values for these plots do correlate to the ion pairs formed based upon the charge of the molecule under observation.[105] Taken together, the experimental values for the slopes of the salt dependence studies can be powerful indicators of the ion pairs formed, and even subtle differences in the actual binding mode as discussed below.

Results from the Chaires lab have also outlined salt dependent relationships for the high affinity dicationic bisintercalator, WP631. They have found that the slope resultant from the salt dependence relationship on the binding constant to be equal to 1.63.[7] Recall that for Na^+, 0.88 ions are released per electrostatic interaction, resulting in a predicted charge of +1.85 (1.63 ÷ 0.88) for WP631, in close agreement with the actual 2+ charge on the molecule. WP631 is essentially two daunorubicin monointercalators attached via a *p*-xylene linker, and similar salt dependence studies with the 1+ daunorubicin resulted in a slope value of 0.85.[7] This produces a predicted charge of +0.96, again, in close agreement with the actual +1 charge of the intercalator. Later experiments slightly modified the linker to an *m*-xylene moiety, producing WP762, a similar dicationic bisintercalator again connecting two daunorubicin monointercalators.

The effects of this slight alteration resulted in a slope of 1.54 for the salt dependence on binding. [125] This slope value predicts a charge of +1.75 for WP762, also in close agreement with the actual 2+ charge on the molecule.

Interestingly, threading intercalation produces depressed slope values for dissociation plots of log k_d versus –log [Na^+] compared to standard intercalation. This was the case reported for a dicationic NDI threading intercalator with a slope equal to 0.3, compared to that of 0.8 for a dicationic propidium intercalator.[134] The Wilson lab has also outlined the effects of altering substituents, and their positions on a core anthracene-9-10-dione intercalator. They discovered that by altering the positions of the substituents around the ring, the salt dependence studies were able to distinguish between normal intercalation and threading intercalation for the family of derivatives with values that were in agreement with previously studied normal intercalating and threading intercalating compounds.[172] It is important to note, however, that other studies have shown that threading intercalators do not produce salt dependent dissociation slopes that correlate directly to the ionic charge and binding mode for the ligand. Threading ruthenium intercalators from the Lincoln and Nordén labs yield salt dependent dissociation slopes that range from 0.09 to 2.4 for dicationic and tetracationic ligands, respectively.[173] While a full understanding of the salt dependence on dissociation kinetics remains unclear, it is useful as a benchmark for similar systems of ligand-DNA interactions.

For instance, salt dependent kinetic studies have been used to distinguish separate binding modes to DNA. 4',6-Diamidino-2-phenylindole (DAPI) is a dicationic diamidine similar to the minor groove binders such as berenil. Upon studying the salt effects on the kinetics of binding for DAPI with both poly(AT) and poly(GC) sequences of DNA, Wilson *et al.* found that both slopes of the equilibrium constant, K_{obs}, produced values of

approximately 2.3, in agreement with the 2+ charge for the formation of two ion pairs.[174] Interestingly, the log-log plots for the dissociation rates varied, with slopes equal to 0.8 and 1.8 for GC rich and AT rich sequences, respectively.[174] Coupled with other experimental observations, they were able to conclude that the dissociation slope values of 1.8 for AT rich DNA correspond to groove binding, while the slope value of 0.8, in close agreement with a dicationic propidium intercalator, for GC rich DNA corresponds to intercalation.[174] The values of the salt dependence on the dissociation rates is capable of predicting different binding topologies, as well as distinguishing subtle differences in intercalative binding as mentioned above.

Outside of salt dependence, it is interesting to consider the dissociation rate constants for other reported long-lived complexes. An overall long-term goal of developing small molecules that bind DNA is to discover new and interesting systems that might have some practical application for regulating gene expression or interrupting transcription. One facet of these sorts of practical application is that the small molecules binding to DNA have characteristically slow dissociation rates, so that biological processes might be disrupted over the course of the life of a cell. With this desirable slow dissociation property of small molecule-DNA interactions in mind, it is important to consider other systems that display particularly high affinities, and more specifically, very slow dissociation rates. One such system, perhaps even the "yard-stick" by which high affinity non-covalent systems are measured, is that of avidin-biotin. The avidin-biotin system is a protein-ligand system with a $K_D = 10^{-15}$ M^{-1}, and reported off-rates ranging from 9×10^{-8} s^{-1} through 7.5×10^{-8} s^{-1} ($t_{1/2} \approx 89 - 107$ days, respectively).[156,157] A synthetic multivalent *tris*-vancomycin-*tris*-D-Ala-D-Ala complex has been developed in the Whitesides group with an even higher affinity than the avidin-biotin system, in which they reported a $K_D \approx 10^{-17}$ M^{-1}, with a corresponding dissociation off-rate of $k_d \approx 4 \times 10^{-8}$

s^{-1} ($t_{1/2} \approx 200$ days).[158] A few highly engineered antibody-antigen systems have also been developed with dissociation off-rates ranging from 1.4×10^{-6} s^{-1} through 3×10^{-7} s^{-1} ($t_{1/2} \approx$ 5 – 27 days).[154,155,153] More closely related to our own system, threading ruthenium intercalators have been shown to produce extremely long dissociation times with an off-rate of 5.02×10^{-6} s^{-1} ($t_{1/2} \approx 1.6$ days) at 37°C.[152] Extrapolation via the Arrhenius equation to 25°C results in an off-rate of 1.24×10^{-6} s^{-1}, or a half-life of 6.5 days.

An anticipated consequence of the threading polyintercalator design of **3** is an extremely slow dissociation from its preferred DNA sequence. This expectation is based on the observation that for one of the two internal NDI units to dissociate, the adjacent terminal NDI must dissociate first. For the entire molecule to dissociate, the other two NDI units would have to also dissociate in a coordinated fashion before rebinding occurs. The unlikely nature of this highly concerted set of motions should reduce the dissociation rate significantly. One might expect the association rate to be proportionately slow as well due to the complicated nature of the bound topology.

3.3 RESULTS

3.3.1 Design

Many separate methods were used to study the kinetics between **3** and three oligonucleotide sequences, both with and without the preferred binding site, as shown in **Figure 3.1**. However, conventional techniques were problematic with regard to examining the binding constants to DNA and will be discussed, as appropriate, below. Because initial UV-visible spectroscopic studies were inadequate for fast initial data collection, a three-step approach was used to determine on rates.

oligo 1 with binding site

5'-CATTTGATAAGTACTTATCGGCTC-3'

3'-GTAAACTATTCATGAATAGCCGAG-5'

oligo 2 without binding site

5'-CATTTAACAACATGTTGTTGGCTC-3'

3'-GTAAATTGTTGTACAACAACCGAG-5'

oligo 3 NMR binding site

5'-GATAAGTACTTATC-3'

3'-CTATTCATGAATAG-5'

Figure 3.1 Oligonucleotide duplex sequences used in the UV, ^{1}H NMR, gel mobility-shift, and Stopped-flow studies. Abbreviated as **oligo 1-3** as referenced above.

1D ^{1}H NMR was used to obtain a structural verification of **3** association with the preferred binding site, **oligo 3,** as used in the previous NMR structural studies which had confirmed 1:1 stoichiometric binding.[8] A simplified one step reaction process can be expressed as seen in **scheme 3.1**.

Scheme 3.1 Simplified reaction scheme of **3**-DNA complex formation.

Scheme **3.1** is of the reaction type A + B → P. However, studying association reactions where $[A]_0 = [B]_0$ is problematic since division by zero occurs in the integrated rate equation for the **3**-DNA complex formation (**equation (1**)), where $[\mathbf{3}]_0 = [DNA]_0$.

$$\left(\frac{1}{[B]_0 - [A]_0}\right) \ln\left(\frac{[A]_0[B]}{[B]_0[A]}\right) = kt \qquad \textbf{(1)}$$

However, because **3**-DNA complex formation is stoichiometric for tetraintercalator **3** and the preferred DNA binding site ([**oligo 1**]), at any given point in time along the reaction coordinate [**3**] = [**oligo 1**] and the reaction can be simplified to A + A → P. The integrated rate equation for stoichiometric binding then simplifies to that shown by **equation (2)**, where a simple plot of 1/[A] vs. *t* should yield a straight line with the observed association constant equal to the slope and y-intercept equal to $1/[A]_0$.[38]

$$\frac{\mathbf{1}}{[\boldsymbol{A}]} = \boldsymbol{kt} + \frac{\mathbf{1}}{[\boldsymbol{A}]_\mathbf{0}} \qquad \textbf{(2)}$$

Next, gel mobility-shift assays can be analyzed for association between compound **3** and the respective oligonucleotide sequences. A cartoon gel shift representation is shown in **Figure 3.2**. Duplex DNA separation on a native PAGE gel is the result of a change in the overall mass-to-charge ratio, as well as the changes in the

rigidity of the duplex. Ultimately, with a polyanion such as DNA, the size and shape of the fragment will determine the separation. The four positive charges on the lysine side chains of **3** should effectively cancel a net of four negative charges from the DNA phosphodiester backbone. Coupled with a likely more rigid polyintercalation complex,[96] if the interaction between **3** and the respective DNA oligo is strong enough, this reduction in net charge should produce a reasonable gel shift during the non-denaturing PAGE separation. A gel shift will thereby provide information regarding the non-covalent interactions that, while not structurally definitive, are strong enough to elicit retardation in the DNA duplex during electrophoresis. By quantifying the ratios between bound and unbound 5'-^{32}P end labeled oligos using densitometry, these values can be plotted and fit using the same stoichiometric binding equation (**equation (2)** above) in order to extract estimated values for k_{on}.

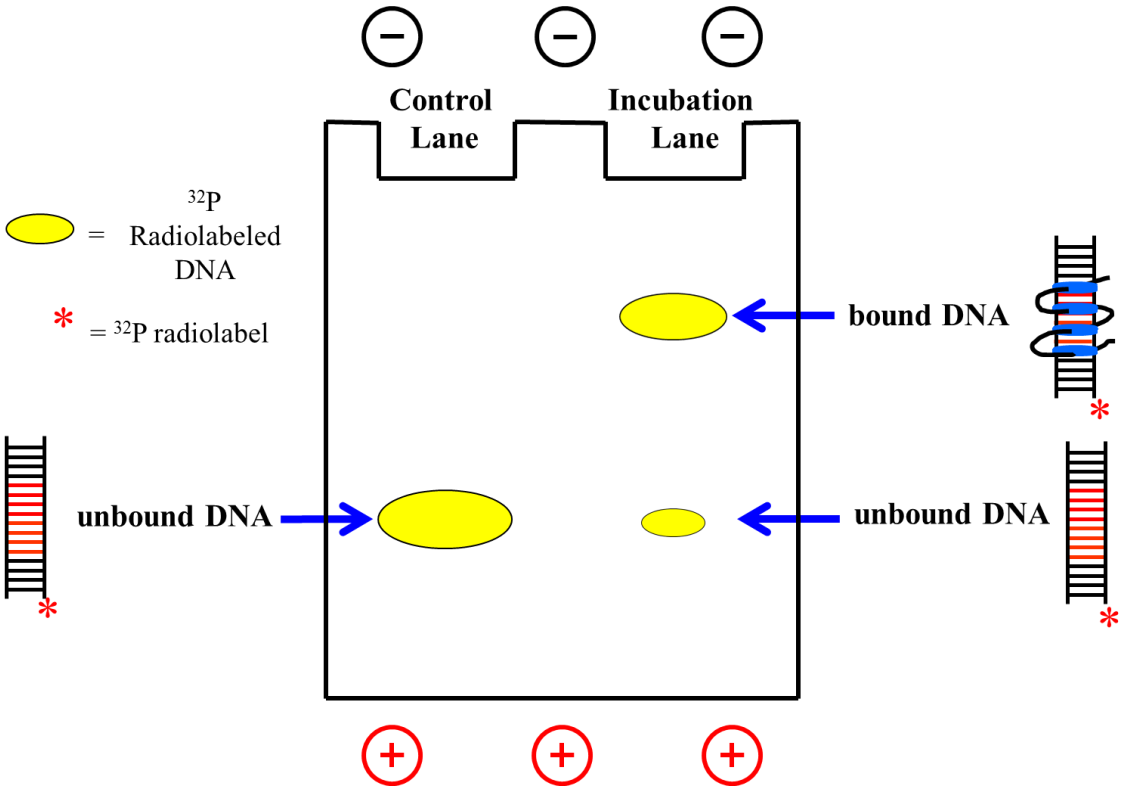


Figure 3.2 Cartoon representation of a gel mobility-shift assay. Control lane represents labeled oligonucleotide with no ligand. Incubation lane represents labeled oligonucleotide with ligand. Incubation lane indicates partial binding where the labeled bound DNA has retarded migration during electrophoresis.

Finally, stopped-flow UV-visible spectroscopy was used as a method to examine association. The fast initial interactions between **3** and respective oligonucleotide sequences will provide data that describes the early binding process, however, it cannot deduce any structural information surrounding complete polyintercalation. The values obtained for k_{on} in this way can then be compared with those obtained via lower resolution, albeit more structural, experiments to arrive at a final apparent value for k_{on} through spectroscopic methods.

Threading intercalators often produce both slower association and dissociation rates.[134] To quantify a value for k_{off}, a gel mobility-shift dissociation assay was utilized where **3** bound to labeled **oligo 1** could be competed off using an excess of unlabeled **oligo 1**, and thereafter monitoring the change between bound and unbound DNA. The ratios between the bound and unbound 5'-end labeled **oligo 1** can then be quantified and plotted for multiple incubations at varying NaCl concentrations. After fitting the data to a monoexponential curve, a value for k_{off} can be determined. Plotting the log k_d versus the $-\log [Na^+]$ should then reveal the salt dependence of the dissociation rate on $[Na^+]$.

The off-rate from a non-specific sequence, **oligo 2**, was monitored using stopped-flow UV-vis and compared with that of the specific k_{off}. These experiments were also performed at multiple concentrations of $[Na^+]$ to determine the relationship to the salt dependence of the interaction by plotting the log k_d versus $-\log [Na^+]$ and fitting the data.

3.3.2 Association Kinetic Studies

3.3.2.1 UV-Visible Spectroscopy

During intercalation, the NDI chromophore is stacked between the DNA bases, resulting in a hypochromic shift in the absorbance maximum at 386 nm.[135] This shift allows UV-Visible spectroscopy to be used to provide data regarding the association of **3** to **oligo 1**. Solutions were prepared in 10 mM PIPES buffer (1mM EDTA, 100 mM NaCl, pH = 7.0) such that stock concentrations of **3** were approximately 3 μM so that the absorbance values were measureable in the linear range with respect to Beer's law. A stoichiometric amount of **oligo 1** in the same PIPES buffer as above was added and quickly mixed and placed in the spectrophotometer for data collection. **Figure 3.3** shows a basic association experiment monitored out to 600s and 1600s respectively. After 100 s, the spectra is virtually unchanged over time varying by only 0.001 absorbance unit, corresponding to only 0.2 nM concentration change. From the point of initial mixing to the first acquired absorbance value, a portion of the initial absorbance change cannot be measured, presumably due to the fast electrostatic interactions during association. Due to this inadequate acquisition time, stopped-flow UV spectroscopy was thereafter pursued for collecting absorbance data at much faster time points.

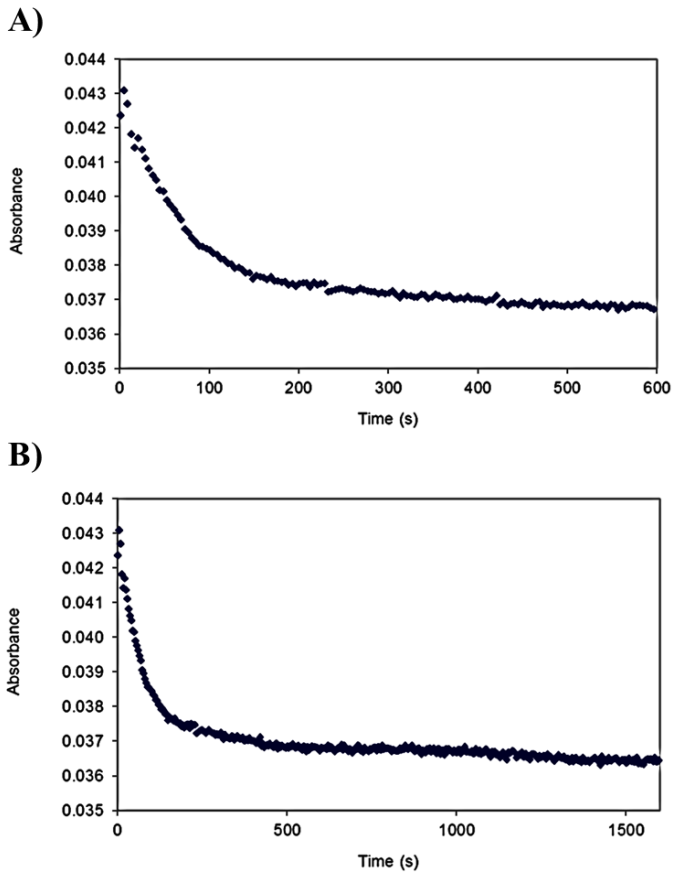


Figure 3.3 UV-vis association studies of **3** binding to **oligo 1** shown out to A) 600 s and B) 1600 s.

3.3.2.2 1D ^{1}H NMR Kinetics

1D ^{1}H NMR studies were performed with tetraintercalator **3** bound to the same 14-mer binding site oligonucleotide (**oligo 3, Figure 3.1**) used in the previous NMR structural analysis.[8] Due to solubility restrictions for compound **3**, stock solutions of both **oligo 3** and tetraintercalator **3** were made at 120 μM in phosphate buffer. 1D ^{1}H NMR spectra were then recorded for **oligo 3** at a final concentration of 60 μM using a standard

jump-return solvent suppression method. Following the acquisition of **oligo 3**, tetraintercalator **3** was mixed directly in an NMR tube with a stoichiometric amount of **oligo 3** at a final concentration of 60 μM. The shifts of the imino proton peaks were then compared with those previously assigned by Lee *et al*. for the same DNA duplex as seen in **Figure 3.4 (A & B).**[8] After a total of 5 minutes 38 seconds to mix and scan the sample, analysis of the spectra reveals that all signals are of identical agreement to the fully equilibrated sample described by Lee and co-workers.[8] In other words, **3** is fully intercalated in the **oligo 3** binding site with the linker in the expected minor groove - major groove - minor groove topology after 5.6 minutes.

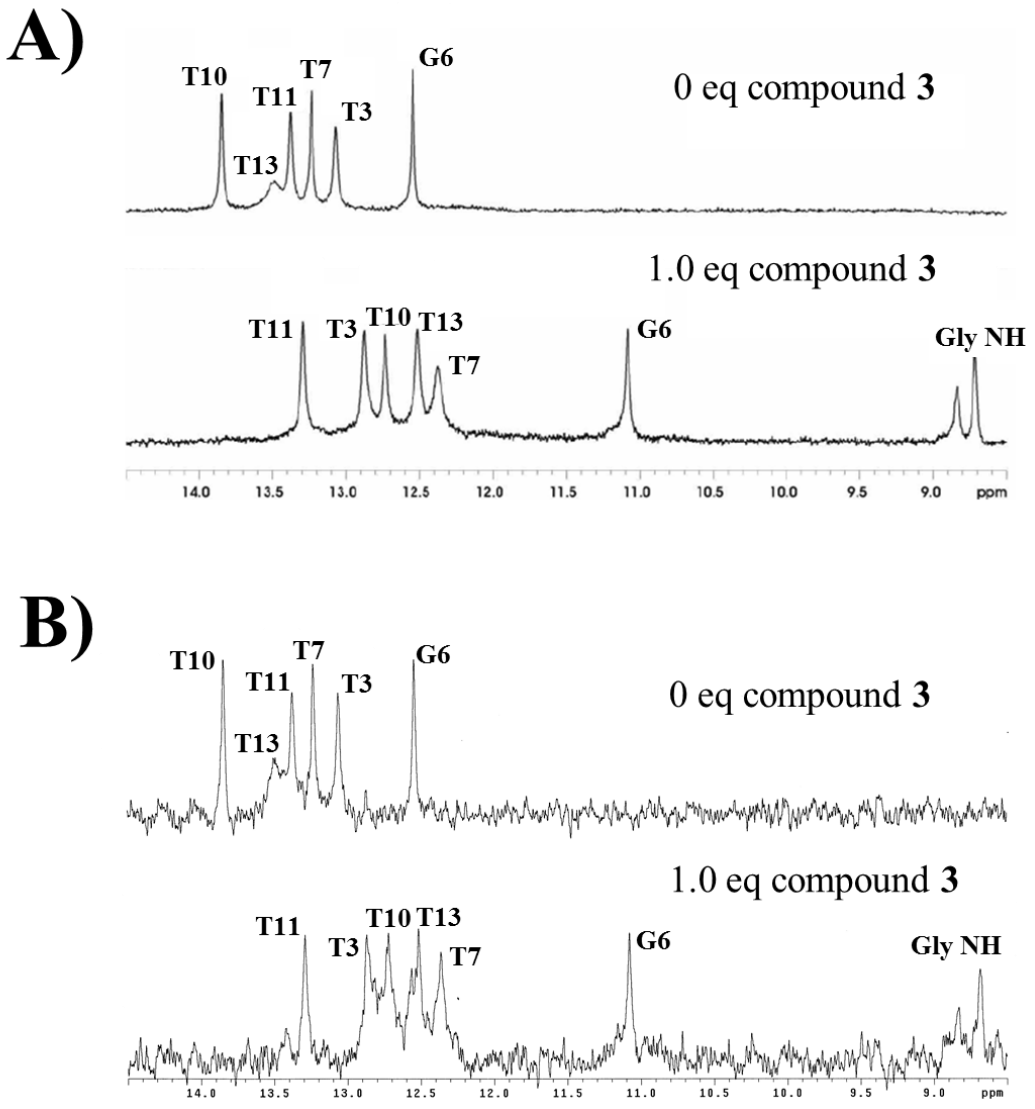


Figure 3.4 **A)** 1D ^{1}H NMR spectra of compound **3** titration into d(GATAAGTACTTATC)$_2$ completed by Lee *et. al*[8] **B)** 1D ^{1}H NMR spectra of 1:1 association kinetic run taken at 5 min 38 sec at a final concentration of 60 μM **3**:**oligo 3** in 30 mM phosphate buffer at pH = 7.0 at room temperature. Imino protons as observed in spectral region of 8.5 ppm to 14.5 ppm.

To determine if the complex would rearrange over time, the experiment was repeated as above, and the **3**-**oligo 3** NMR sample was monitored at different time points with no observable change over 24 hours as shown in **Figure 3.5**. Because no spectral change occurred over 24 hours, if it is assumed that binding is then at least 95% complete (based upon the detection limits of ^{1}H NMR) and the fastest scan time of 338 sec is used as a lower limit, we can determine an association constant by using **equation** (**2**) from above. By plugging in the single data point where [A] = unbound **oligo 3**, and coupled with a point for 100% unbound (60 μM) at t = 0 sec, we can generate a line. While this line only includes a starting point and a finishing point, and is therefore uninformative towards a true association rate, it can still provide an absolute lower limit to association, with slope = k_{obs}, which produced a value of approximately 937 M^{-1} s^{-1} for the lower association limit.

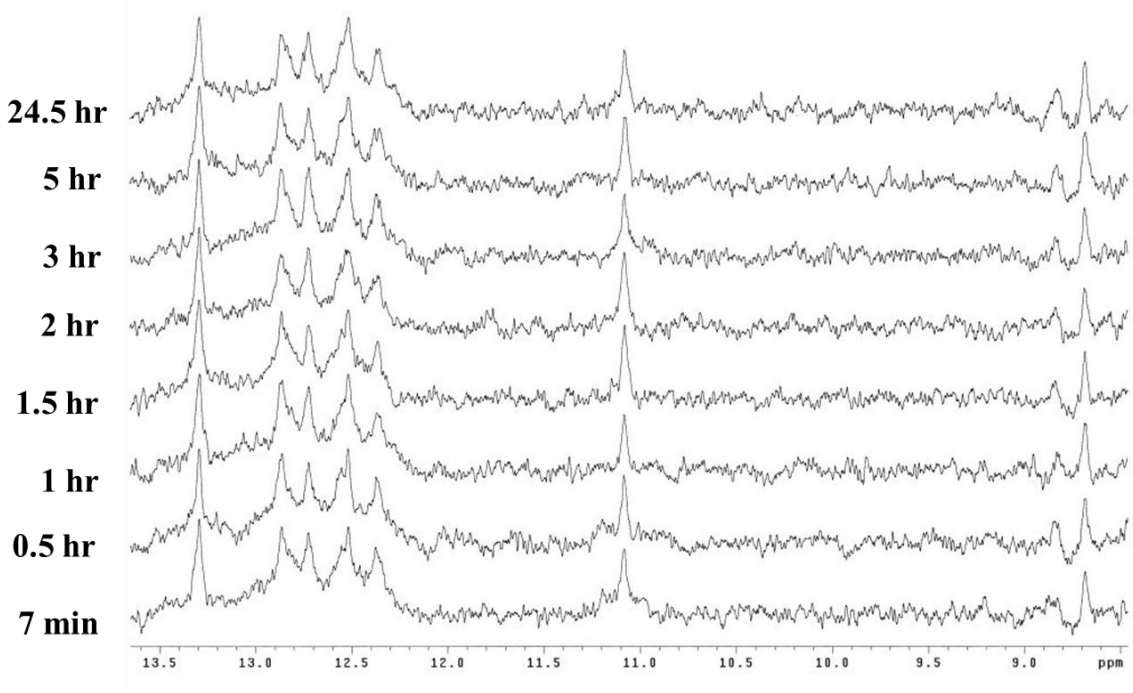


Figure 3.5 1D ^{1}H NMR stacked spectra of 1:1 association kinetic run taken at different time points and at a final concentration of 60 μM **3**:**oligo 3** in 30 mM phosphate buffer at pH = 7.0 at room temperature.

3.3.2.3 Gel Mobility-Shift Association Assays

As a qualitative approach to monitoring association, the gel mobility-shift assay was used. Before association rates could be measured, it was determined if **3** was capable of producing a gel shift upon incubation with respective specific and non-specific oligonucleotides (**oligo 1** & **oligo 2**, **Figure 3.1**). This was tested by incubating tetraintercalator **3** with a stoichiometric amount of respective ^{32}P radiolabeled oligonucleotide in 10 mM PIPES buffer (1mM EDTA, 100 mM NaCl, pH = 7.0) for 12 hours. An aliquot of the incubation was then mixed briefly with non-denaturing loading buffer and separated using native PAGE for 1.5 hours. As seen in **Figure 3.6**, a clear gel shift is present for specific binding between **3** and **oligo 1** as indicated by lane 2. Interestingly, no gel shift is observed for non-specific binding between **3** and **oligo 2** in lane 4 of **Figure 3.6**, indicating that non-specific interactions are not sufficiently strong to produce a gel shift.

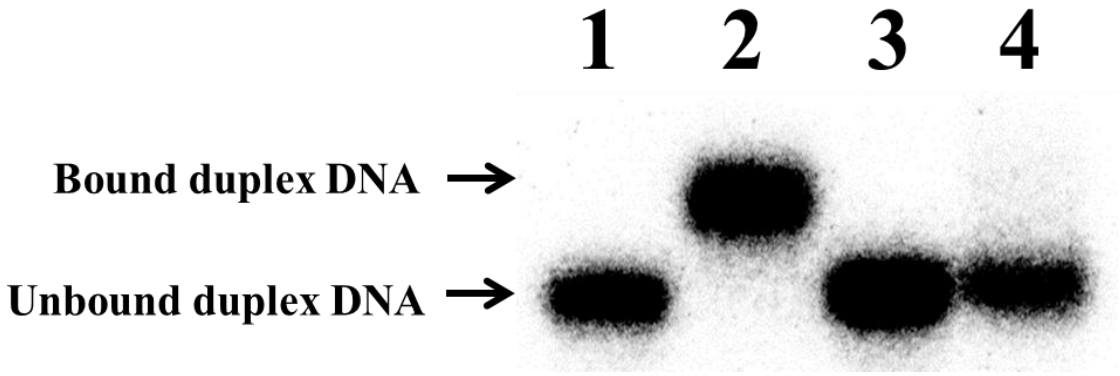


Figure 3.6 Representative association gel shifts. Lane 1 represents a control and contains **oligo 1** only. Lane 2 represents specific binding and contains a stoichiometric amount of **3** and **oligo 1**. Lane 3 represents a control and contains only **oligo 2**. Lane 4 represents non-specific binding and contains a stoichiometric amount of **3** and **oligo 2**.

Further, preliminary results are consistent with the need to adjust the concentration of **oligo 1** and tetraintercalator **3** lower, to the point that incomplete binding could be detected as a percent unbound by gel electrophoresis. At concentrations above 2.5 μM, complete association between **3** and **oligo 1** was observed even at the first data point around 2 minutes. Due to uncertainties regarding exactly how long the components were interacting before entering the gel, these experiments were used to gain only a rough estimate of an association rate constant. This estimate was used to be compared with, and inform, the more accurate stopped-flow spectroscopic measurements to be described below. One particular advantage of this approach is that non-specific association does not produce a gel shift, so the estimated association rate constant measured in this way is expected to represent sequence specific binding.

Association rate incubations were run in the same PIPES buffer as mentioned above. The concentrations of **oligo 1** and tetraintercalator **3** were lowered to 250 nM, and

after approximately 2 minutes of incubation before loading, triplicate measurements indicated 11.7 % ± 0.4 % of unbound **oligo 1** remained. After approximately 5 minutes of incubation before loading, triplicate measurements of the 250 nM incubation indicated a remaining 4.7 % ± 0.3 % of unbound **oligo 1**. Due to the observation that concentrations below 250 nM contained too small a quantity of radiolabel to accurately quantify the data, this same analysis was used to monitor concentrations ranging from 250 nM to 2,500 nM using equation (**2**) discussed above. A representative association gel after a 2 minute incubation time can be seen in **Figure 3.7**. Using autoradiography and densitometry to quantify the percent unbound **oligo 1**, the molar concentration can then be calculated. Plotting 1/[A] vs. *t* yields a straight line with the slope equal to k_{on} (M^{-1} s^{-1}) for association of **3** to **oligo 1** where $[A]_0$ is the initial molar concentration of unbound **oligo 1**, [A] is the molar concentration of unbound **oligo 1** at time, *t*, and *k* is the estimated association constant for **3-oligo 1** complex formation. A plot of 1/[A] versus *t* for gel shift association assays can be seen in **Figure 3.8**. The values obtained for the slope in this manner can be seen in **Table 3.1**. Averaging across the above mentioned concentration range provides an estimate to the association rate constant with a value for the apparent $k_{on} = 2.7 \pm 0.5 \times 10^5$ M^{-1} s^{-1}. (**Note:** Amy Smith of the Iverson lab has since produced more precise methods for measuring the association rate by gel shift. These studies indicate a more appropriate $k_{on} \sim 1 \times 10^4$ M^{-1} s^{-1}.)

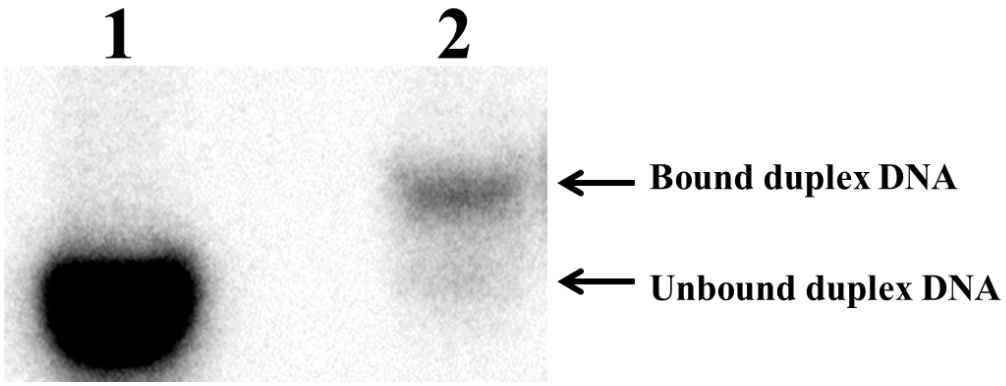


Figure 3.7 Representative association gel for an approximate incubation time of 2 min. Lane 1 represents a control at 1000 nM **oligo 1** only. Lane 2 represents association at 250 nM of both **3** and **oligo 1**.

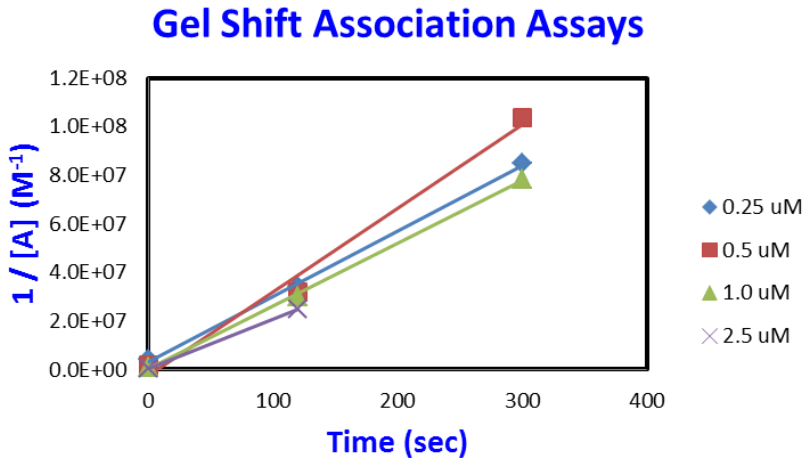


Figure 3.8 Plot of 1/[A] vs. *t* for gel shift association assays for a range of concentrations with a stoichiometric amount of **3** and **oligo 1** for 0.25 through 2.5 μM as indicated in the legend. Association linear fits are shown for the time points at $t = 0$, 120, 300 sec respectively.

[oligo 1] (μM)	k_a x 10^5 (M^{-1} s^{-1})
0.25	2.69
0.50	3.42
1.00	2.58
2.50	2.03

Table 3.1 Table containing the observed values for k_{on} (M^{-1} s^{-1}) as provided by the slope for the linear fits of **Figure 3.9**, with a standard deviation value of 0.5 x 10^5 (M^{-1} s^{-1}).

3.3.2.4 Stopped-Flow UV-Visible Spectroscopy

Stopped flow association kinetics for the reaction of the tetraintercalator 3 with DNA. The rates of association between tetraintercalator **3** and the DNA oligonucleotides containing either the target (**oligo 1**) or non-target binding sequence (**oligo 2**) were analyzed using stopped flow methods. As mentioned above, hypochromism is known to accompany intercalation, so the time-dependent decrease at the tetraintercalator visible absorbance maximum (386 nm) was monitored following the mixing of equal volumes of solutions containing **3** and the respective specific or non-specific DNA oligonucleotide. Absorbance traces were collected for fixed concentrations of **3** mixed with increasing DNA oligonucleotide concentrations such that the final ratios after mixing encompassed a range from 1:1 to 1:8 for **3**:oligonucleotide (**Figure 3.9**). An initial rapid decrease in absorbance is observed, reflecting a fast association of **3** with the DNA oligonucleotides. The absorbance change levels off by 10 s and no significant further change in absorbance is observed up to 600 s.

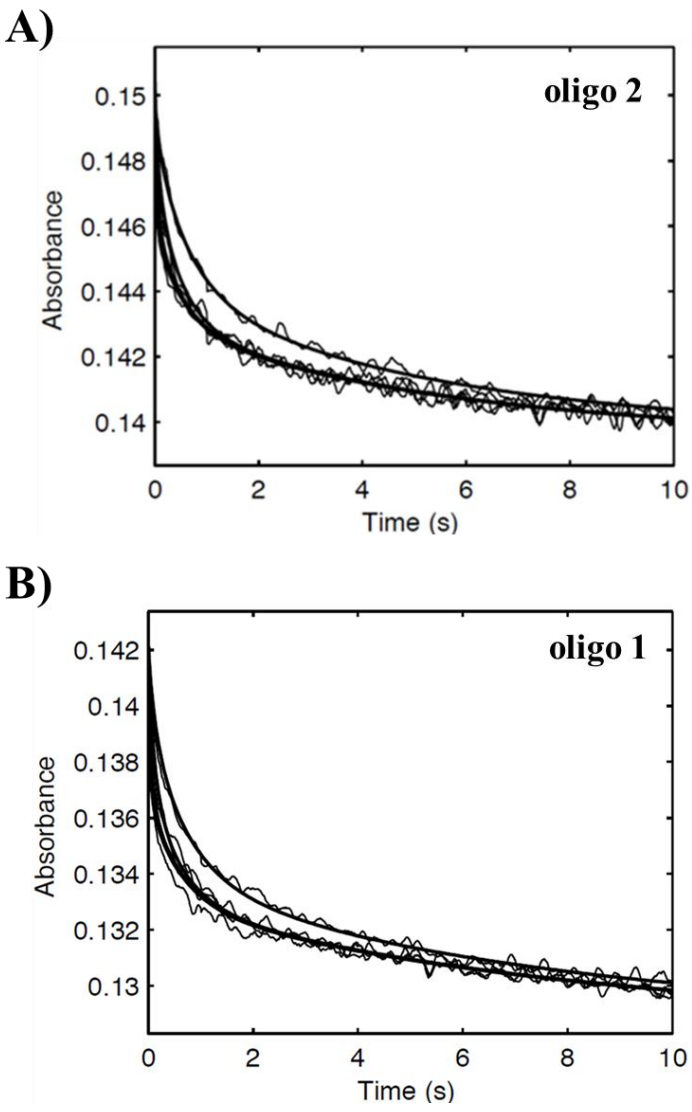
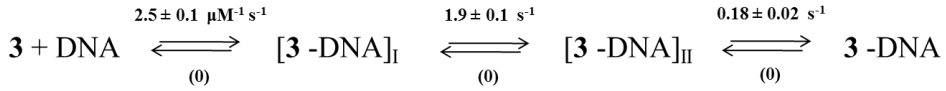


Figure 3.9 Representative stopped flow absorbance traces of the association of **3** (2 μM) with **A**) **oligo 2** and **B**) **oligo 1** at four different DNA concentrations (2, 4, 8, 16 μM) in 10 mM PIPES buffer (pH 7.0), 1 mM EDTA, 100 mM NaCl, and monitored at 386 nm at room temperature. The solid lines represent the best global fit to a three-step consecutive mechanism according to scheme 1a using KinTek Explorer Software.

Single or two-step models did not fit the data, therefore the kinetic traces obtained at each DNA concentration within the first 10 seconds were fit globally[175] to a minimal three-step irreversible mechanism (**scheme 3.2A**). The model assumes an initial fast interaction followed by slower consecutive rearrangement steps, where [**3**-DNA]$_I$ and [**3**-DNA]$_{II}$ represent two different intermediate complexes.[176] Although this model describes the kinetic behavior for both DNA sequences, the fit is best for the association of tetraintercalator **3** with **oligo 2**. In the case of **3** binding to **oligo 1**, an additional step was required in order to better account for the fast phase, particularly at lower DNA concentration, which is a key component of the data (**scheme 3.2B**). The additional step fits the kinetic data but the parameters are not sufficiently well constrained to uniquely define the rate constants for the two isomerization steps. Nonetheless, globally fitting the data according to both mechanisms yielded well-defined second order association rate constants for the two DNA sequences, which were $3.4 \pm 0.1 \times 10^{-6}$ $M^{-1}s^{-1}$ and $2.5 \pm 0.1 \times 10^{-6}$ $M^{-1}s^{-1}$ (k_1) for **3** binding to **oligo 1** and **oligo 2**, respectively. The largely irreversible rearrangement steps are also consistent with the remarkably slow off-rate of **3** from the target sequence.

A)



B)

Scheme 3.2 Rate constants were derived from global fitting the concentration dependence of the data series to a three step model. Both models assume irreversible steps toward complex formation. (A and B) Model that best fit the association of **3** with **oligo 2** and **oligo 1**, respectively.

3.3.3 Dissociation Kinetic Studies

3.3.3.1 UV-Visible Spectroscopy

Standard UV-visible spectroscopy dissociation experiments were performed using a final 2% sodium dodecyl sulfate (SDS) detergent in 10 mM PIPES buffer (pH = 7.0) to sequester **3** upon dissociation from either specific (**oligo 1**) or non-specific DNA (**oligo 2**).[1,135] While NDI intercalators typically have sufficiently slow dissociation rates to allow monitoring by UV-vis,[134,177] compound **3** dissociation profiles demonstrated unique characteristics with extraordinarily long dissociation times.

Figure 3.10A shows a UV dissociation profile of **3** from the non-specific **oligo 2**. When plotted with a corresponding UV dissociation profile of **3** from **oligo 1** under identical concentrations and conditions (**Figure 3.10B**), it is clear that a stopped-flow apparatus is necessary to collect the initial dissociation data from non-specific DNA, as will be described in a later section below.

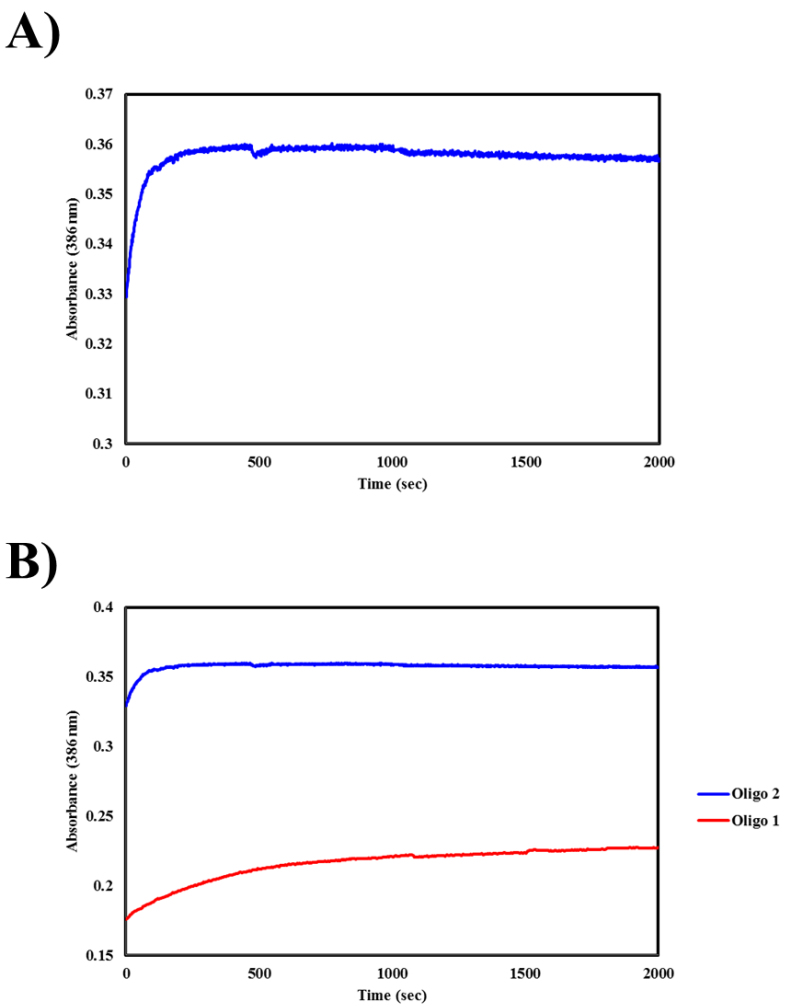


Figure 3.10 UV-vis dissociation profiles of **3** from A) **oligo 2** only and B) an overlay of **oligo 1** (red) and **oligo 2** (blue) dissociations at identical concentrations. All dissociations were induced by 2% SDS at room temperature in 10 mM PIPES buffer (pH = 7.0).

It is also apparent from **Figure 3.10 (B)** that dissociation of **3** from the specific sequence of **oligo 1** is a much slower process than dissociation from non-specific DNA. Further investigation of this process was pursued by monitoring the dissociation of **3** from **oligo 1** out to 100 hours by UV-visible spectroscopy as shown in **Figure 3.11**. Not only is dissociation of **3** from **oligo 1** still incomplete after 100 hours, but the ability to accurately collect data is compromised due to the lack of consistent measurements from the extended use of both the UV and visible lamps over a time course of days.

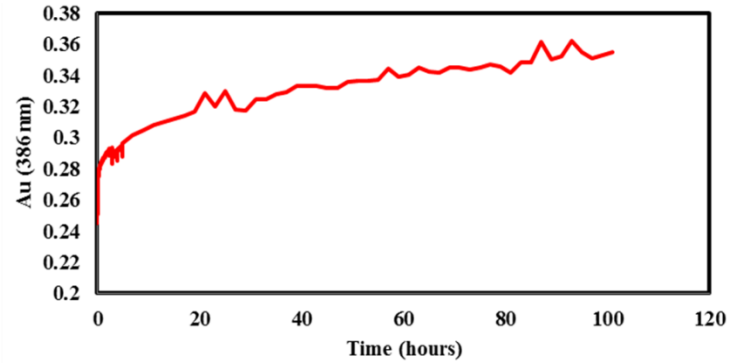


Figure 3.11 UV-visible dissociation profile of **3** from **oligo 1** over the course of 100 hours as induced by 2% SDS in 10 mM PIPES buffer (pH = 7.0) at room temperature.

One assumption of an SDS-induced dissociation reaction is that sequestered ligands are not able to re-associate with the DNA after dissociation. To test this, re-association controls were run to determine if a solution of **3** mixed in 2% SDS prior to addition of an **oligo 1**/2% SDS solution would indicate any association of **3** to **oligo 1** in the presence of the detergent. As shown in **Figure 3.12**, 2% SDS is not sufficient to

prevent re-association of **3** with the specific **oligo 1**, and further increasing the SDS percent composition can result in compromising the integrity of the DNA double helix itself.[178] While acceptable for dissociation of **3** from non-specific **oligo 2**, standard SDS-induced dissociation monitored by UV-vis spectroscopy in determining the off-rates of **3** from its preferred binding site in **oligo 1** is not an appropriate methodology.

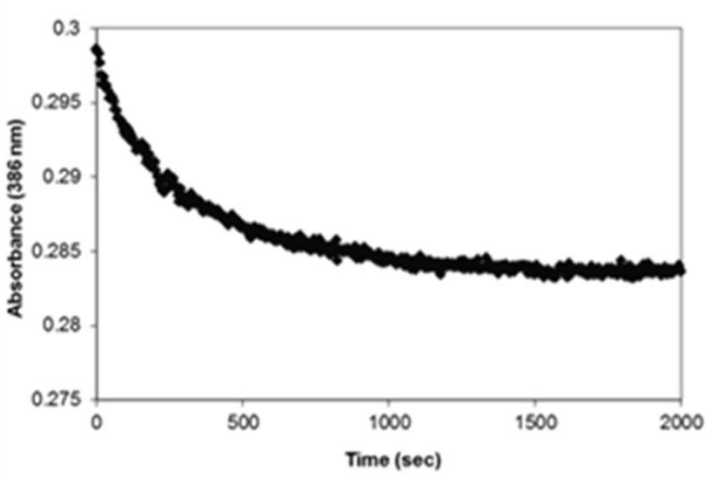


Figure 3.12 UV-visible re-association profile of a solution of **3** in 2% SDS added to a solution of **oligo 1** in 2% SDS at room temperature. All solutions in 10 mM PIPES buffer (pH = 7.0).

3.3.3.2 Gel Mobility-Shift Dissociation Assays

To quantify a value for k_{off}, a gel mobility-shift assay was designed such that tetraintercalator **3** was able to fully associate with a stoichiometric amount of 5'-P^{32}-labeled **oligo 1** duplex (see association **Figure 3.6** above). The incubation mixture was then separated by native polyacrylamide gel electrophoresis (PAGE). Bound

tetraintercalator **3** retarded migration of **oligo 1** on the gel as shown above for the association experiments. A control oligonucleotide (**oligo 2**) that did not contain the preferred binding site was also incubated with **3**, and as expected, the non-target sequence did not produce an observable gel-shift.

Gel mobility-shift assays were then utilized to determine the rate constant for **3** dissociating from its preferred binding site. To measure the dissociation rate constants, a 100-fold excess of unlabeled oligonucleotide was added to a solution containing the ^{32}P-labeled **oligo 1** pre-equilibrated with a stoichiometric amount of tetraintercalator **3**. The 100-fold excess of unlabeled **oligo 1** insured that any **3** to dissociate from the ^{32}P-labeled **oligo 1** would be captured by unlabeled duplex. In this way, the rate constant of **3** dissociating from labeled **oligo 1** was determined by quantifying the PAGE separated bound to unbound ratio versus time using autoradiography.

Dissociation experiments using **3** bound to **oligo 1** were carried out in 10 mM PIPES buffer (1 mM EDTA, pH = 7.0) using a range of NaCl concentrations. Gel-shifts could be accurately analyzed up to about 34 days before loss of radioactivity reduced the signal-to-noise ratio to below useful levels. Representative data of gel shifts are shown in **Figure 3.13**.

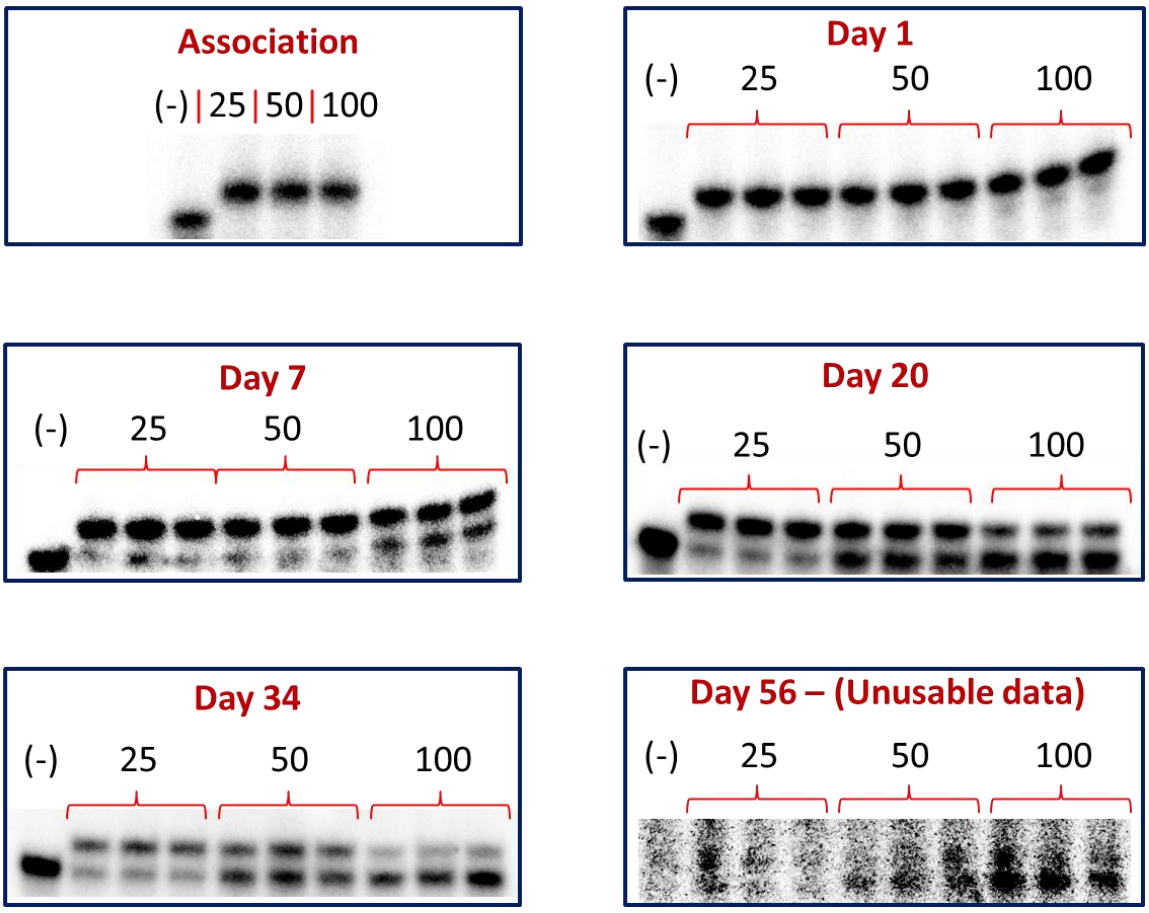


Figure 3.13 Representative Gel-shift dissociation assays in triplicate at the respective labeled time points and at a range of [Na^+] from 25 mM to 100 mM as labeled above the lanes. All incubations were performed at 1.1 µM of **3** at room temperature in 10 mM PIPES (1 mM EDTA, pH = 7.0) with 100-fold excess of unlabeled **oligo 1**.

An autoradiography analysis was performed on the bands corresponding to the bound and unbound oligonucleotides which were quantified using densitometry. These values represent a ratio, which can be plotted as either percent bound or percent unbound versus time. The quantified data was fit to a standard monoexponential decay equation (equation (**3**)[178]):

$$A = A_0 e^{-kt} + b_0 \qquad (\mathbf{3})$$

The dissociation data was plotted by the percent bound **oligo 1** by **3**, which can be seen in **Figure 3.14** with their respective curve fits. The dissociation off-rates and their respective dissociation half-lives are shown in **Table 3.2**. The dissociation rate constant, $k_d \approx k_{off}$, increased from k_{off} = 2.3 ± 0.3 x 10^{-7} sec^{-1} to k_{off} = 9 ± 1.5 x 10^{-7} sec^{-1} as the concentration of NaCl was increased from 25 mM to 200 mM. At 100 mM NaCl, the benchmark dissociation rate constant was found to be k_{off} = 5 ± 0.5 x 10^{-7} sec^{-1}, corresponding to a dissociation half-life of $t_{1/2}$ = 16 days. A dependence on NaCl is not surprising given the four lysine residues within **3** that presumably make electrostatic contacts with the phosphodiester backbone of DNA. A plot of log k_d as a function of –log [Na^+] was linear with a slope of 0.59 (**Figure 3.15)**, which while not directly comparable due to a more complex binding topology, has some correlations with previous salt dependent dissociation studies of threading NDI intercalators.[134]

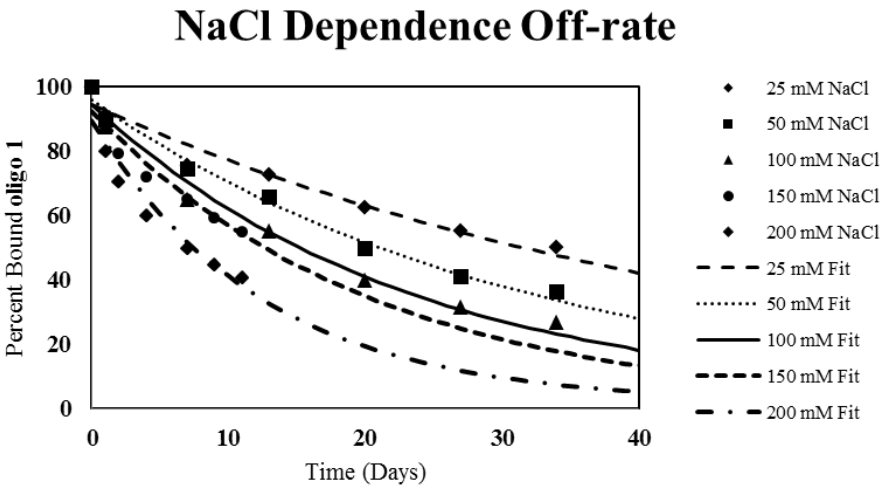


Figure 3.14 Dissociation plot of the percent of bound **oligo 1** by **3** versus time along with monoexponential curve fits at varying concentrations of [Na^+].

[NaCl] (mM)	k_d x 10^{-7} (s^{-1})	$t_{1/2}$ (days)
25	2.3	32
50	3.3	23
100	5.0	16
150	5.7	13
200	9.3	8

Table 3.2 Dissociation rate constants of **3** from **oligo 1** calculated from the monoexponential fits to the gel-shifts generated at a range of [Na^+] from 25 mM to 200 mM. Error levels estimated at ± 17%. Half-life values derived from the monoexponential fit curves are shown in days.

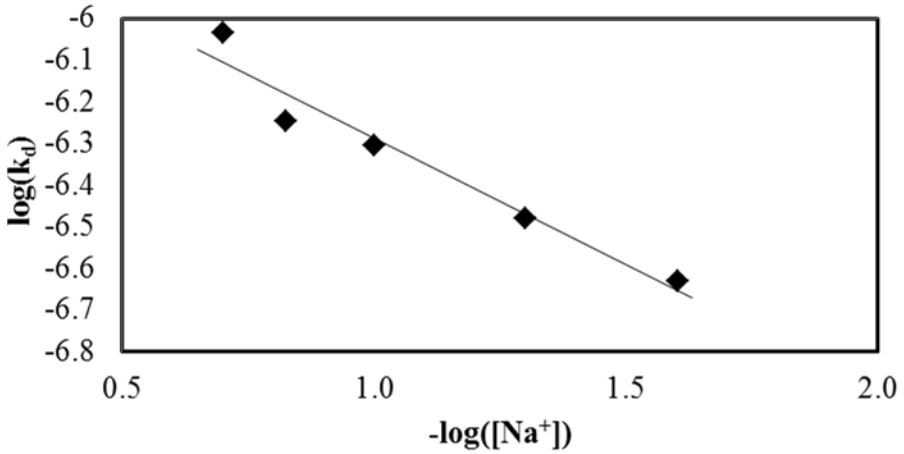


Figure 3.15 Plot of log(k_d) vs – log([Na^+]) for dissociation of **3** from **oligo 1** with slope of 0.59. Experiments were conducted at room temperature in 10 mM PIPES buffer (pH = 7.0) with 1 mM EDTA at varying [Na^+].

3.3.3.3 Stopped-Flow UV-Visible Spectroscopy

Dissociation kinetics of 3 from a randomly chosen DNA sequence. In order to quantify the sequence specific behavior of **3**, the dissociation rate constant from a randomly chosen sequence, **oligo 2**, was also determined. As mentioned previously, gel electrophoresis could not be used to monitor the dissociation because a gel shift could not be detected for the **3-oligo 2** complex, presumably due to rapid dissociation. Instead, the dissociation rate of **3** from **oligo 2** was determined using stopped-flow methods in the presence of 2% sodium dodecyl sulfate (SDS). In these experiments, the **3-oligo 2** complex was mixed with equal volumes of buffer containing 2% SDS and the absorbance at 386 nm was followed for 200 s (**Figure 3.16**) It is assumed that upon dissociation, tetraintercalator **3** is precluded from rebinding due to the detergent action of the SDS.

This assumption was tested by carrying out controls in which the same concentration of SDS was added to the non-specific **oligo 2** solution prior to adding **3** and no re-association binding was observed. The absorbance trace was fit to a double exponential function, suggesting a two-step mechanism, with the first-order dissociation rate constants of the fast and slow steps, k_1 and k_2, determined to be 0.29 s^{-1} and 0.019 s^{-1}, respectively, at a benchmark 100 mM NaCl. The relative amplitudes of the faster and slower phases were 34% and 66%, respectively. Although detergent-induced dissociation is a well-established method to determine dissociation rates of DNA-intercalator complexes[1,107], SDS has been shown to enhance the dissociation rate somewhat.[179] Thus, the experimental values reported here should be considered upper-limit estimates. Using the slower dissociation step for comparison, these results demonstrate that the dissociation of **3** from **oligo 1** containing the preferred binding site is roughly 4 x 10^4 times slower compared to **oligo 2**, verifying that the differences in kinetic off-rates is largely, if not exclusively, responsible for sequence selectivity. Salt dependence of the dissociation rate was also examined for the **3-oligo 2** dissociation, (**Table 3.3**) and a plot of log k_d as a function of –log [Na^+] was linear with a slope of 0.54 (**Figure 3.17**), similar to the value of 0.59 for the same salt dependence dissociation plot with the specific **oligo 1** sequence.

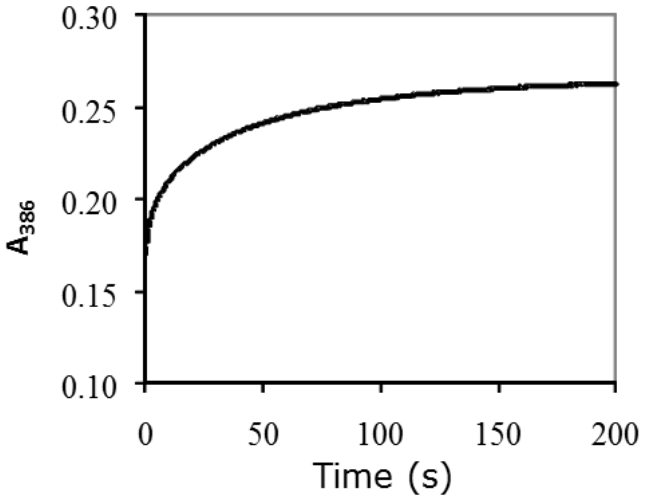


Figure 3.16 Representative stopped flow dissociation trace of **3**-**oligo 2** complex. The equilibrated complex was mixed at room temperature with equal volumes of 4% SDS in 10 mM PIPES (pH 7.0), 1 mM EDTA, and 100 mM NaCl. The data was fit to a double-exponential.

$[Na^+]$ (mM)	k_1 (s^{-1})	A_1 (%)	k_2 (s^{-1})	A_2 (%)	k_d (s^{-1})
50	**0.27**	**30**	**0.012**	**70**	**0.088**
100	**0.29**	**34**	**0.019**	**66**	**0.11**
150	**0.38**	**35**	**0.031**	**65**	**0.15**
200	**0.39**	**40**	**0.038**	**60**	**0.18**
300	**0.48**	**39**	**0.055**	**61**	**0.22**

Table 3.3 Dissociation rate constants of **3** from **oligo 2** at varying concentrations of NaCl. A_1 and A_2 are the relative amplitudes for the double exponential according to $k_{d,app} = A_1k_1+A_2k_2$. In order to obtain the faster rates and corresponding amplitudes, the fits to the traces obtained at 10 s intervals were coerced to the rates and amplitudes obtained from the 200 s trace.

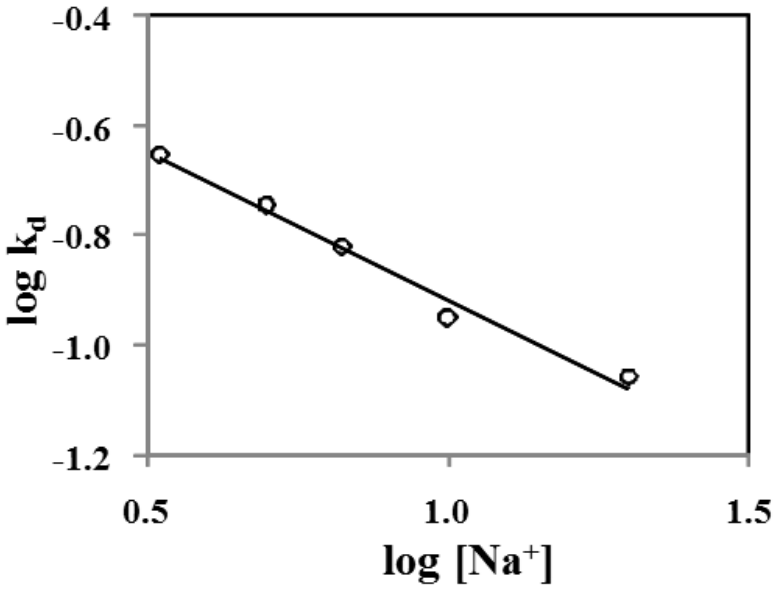


Figure 3.17 Plot of log k_d versus – log [Na^+] for dissociation of **3** from **oligo 2** with a slope of 0.54.

3.3.4 Further Binding Studies

Isothermal titration calorimetry (ITC) is a common technique to determine the binding energetics of multiple host-guest systems. As such, it is a technique capable of measuring thermodynamic data, more specifically the enthalpy (ΔH), surrounding the interactions between a ligand and DNA, and ultimately producing a binding curve as a result of the experiment. ITC can be advantageous in that it measures the heat released or absorbed from a series of injections while able to reach equilibrium in solution, versus other techniques like surface plasmon resonance (SPR) in which one of the substrates is typically immobilized on a chip. However, one drawback is that ITC can require higher concentrations of samples,[180] typically no lower than 500 μM. While the solubility of **3** is

approximately 0.4 mg / ml (~150 μM) in H_2O, ITC was approached as a potential method for obtaining binding energetics of the **3-oligo 1** complex. The results can be seen in **Figure 3.18**. It was presumed that the heat flow changes measured in the experiment were due largely to the electrostatic interactions as the result of the release of 4 Na^+ ions upon binding of the tetracationic tetraintercalator. From the given data, no appropriate binding curve could be produced and no further attempts at ITC experiments were pursued. It was later discovered that the dissociation of **3** from **oligo 1** displays a half-life of 16 days at 100 mM NaCl. Provided that ITC measures the heat evolved or absorbed as the system moves to equilibrium, it becomes apparent now that ITC is not particularly appropriate for our system due to the extended times to reach equilibrium.

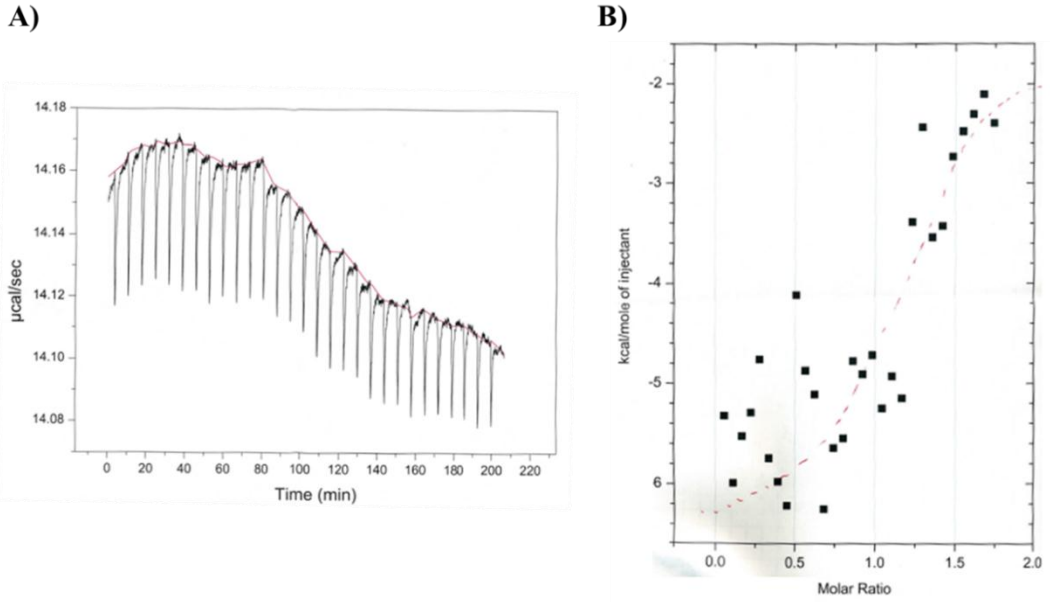


Figure 3.18 A) ITC injection data from titration of **3** into a solution of **oligo 1** in 10 mM PIPES buffer (pH = 7.0) with 1 mM EDTA and 100 mM NaCl. B) Resultant binding curve plot.

Differential scanning calorimetry (DSC) is perhaps the most desirable method for obtaining thermodynamic data related to our system of **3** bound to its preferred binding site. Micro-DSC experiments have become widely used in the field of small molecule-DNA interactions for probing insights into the thermodynamic properties of such systems.[7,181] While ITC measures the energy change surrounding a system where the reaction is kept at a constant temperature, DSC measures the energy required to maintain a zero temperature differential between a reference and sample cell as a function of temperature over time. DSC coupled with thermal melting curve studies can produce a binding curve, from which the binding affinities may be derived with the simultaneous thermodynamic data.

We approached both thermal melting studies and DSC for a potential method to determine a melting temperature for both the **3-oligo 1** and **3-oligo 2** complexes, and obtain thermodynamic data simultaneously. [182] Thermal melting experiments were performed and can be seen in **Figure 3.19** monitored at 260 nm. However, the melting curve was too broad for accurate interpretation and was not a reversible process (**Figure 3.19 (C)**). The derivative of the melting curves suggests a two-phase process for melting as shown in **Figure 3.19 (B)**.

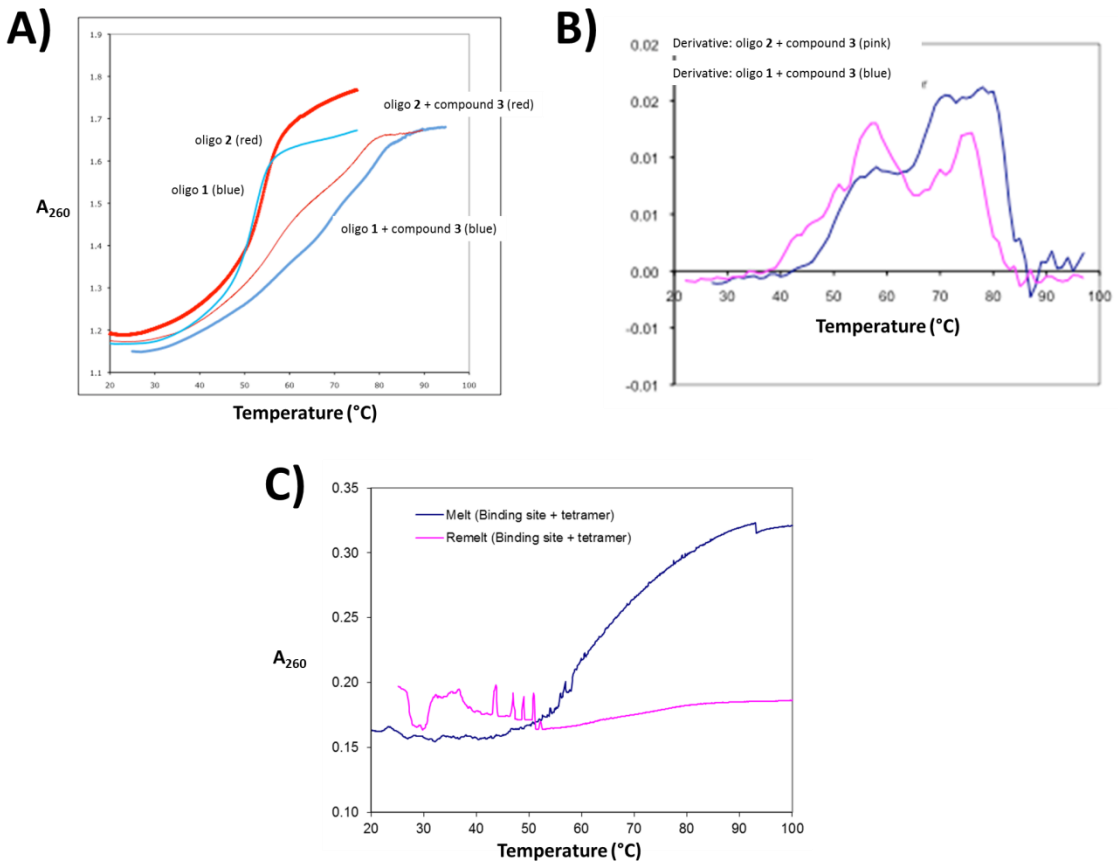


Figure 3.19 (A) Thermal melting studies for **3** with both **oligo 1** and **oligo 2** at 260 nm. (B) First derivative for melting of the **3**-**oligo 1** and **3**-**oligo 2** complexes. (C) Thermal melting of the **3**-**oligo 1** complex, followed by a re-melt of the complex.

The first attempts of the DSC experiment were to observe the signals produced for the individual strands melting and re-annealing, and after several trial runs, the signals were slightly apparent only after using the equivalent of 10 mg of the **oligo 2** duplex in 0.1 ml H_2O (~ 6.8 mM) (**Figure 3.20**). It was determined that at this time, DSC is not a practical methodology due to the requirements and expense of using so much

oligonucleotide per run coupled with the solubility restrictions of tetraintercalator **3**. Typical experiments examining ligand-DNA interactions have been performed using cheaper, bulk carrier DNA, such as from calf-thymus or herring sperm, but the use of such DNA doesn't detail sequence specific interactions.[7] Our approach using oligonucleotides to monitor sequence specific interactions may be useful in the future, with the potential collaboration with research groups using a Micro-cal DSC, which would require less material and present a potential solution to obtaining thermodynamic data pertinent to our system.

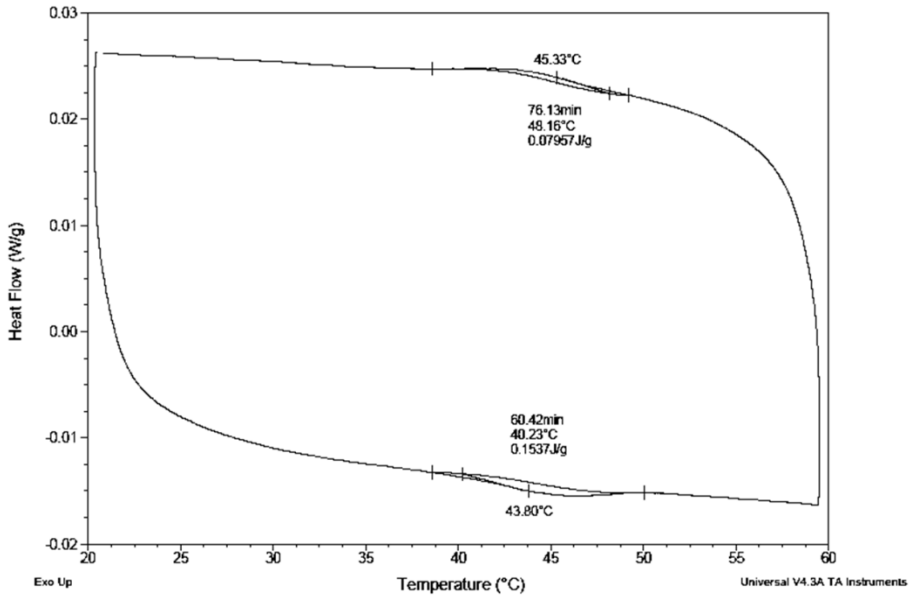


Figure 3.20 DSC run with **oligo 2** in 10 mM PIPES buffer (pH = 7.0), 1 mM EDTA, 100 mM NaCl.

Circular dichroism spectroscopy can be used as a low-resolution technique, compared to two-dimensional NMR and X-ray crystallography, to determine the binding mode and affinity for small molecules bound to DNA. This is done by monitoring an induced CD (ICD) signal as a result of the asymmetry of the entire complexed system. By extension, the ICD signals can produce a binding curve when analyzing a titration experiment, the results of which can be fit and used to calculate an apparent binding constant. While intercalators are known to produce a small ICD signal, minor-groove binders are capable of producing ICD signals orders of magnitude larger.[183] While our system is characterized as threading polyintercalation, it thereby relies upon a combination of intercalation and groove-binding in both the major and minor grooves of DNA. Because of this mixed binding topology, CD spectroscopy was used to monitor if tetraintercalator **3** could produce an ICD signal when mixed in a solution of **oligo 1** at multiple time points. The results of this experiment can be seen in **Figure 3.21** and show that no ICD signal is produced upon **3** binding to **oligo 1**, even after an incubation time of 96 hours. Further, the previous ^{1}H NMR data as seen above, coupled with the association gel shifts provide an effective concentration range of 250 nM – 60 μM of stoichiometric **3** and **oligo 1**, both of which indicating binding after a matter of minutes. Taken together, CD spectroscopy of the **3-oligo 1** complex does not produce a measurable ICD signal at reasonable concentration ranges, and at this time is not a useful method to experimentally determine a binding constant via a titration experiment.

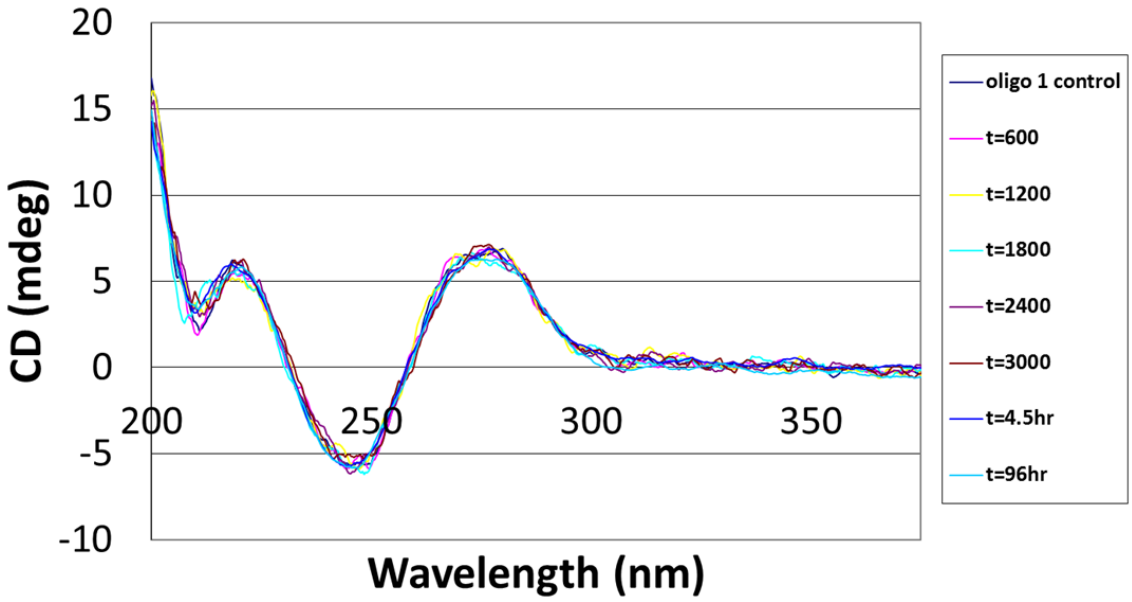


Figure 3.21 CD spectra taken at various time points in minutes (hour time points are indicated) of a mixture of **3** into a solution of **oligo 1**. **Oligo 1** control is a run with no added **3** as a baseline control. After mixing **3** with **oligo 1**, no ICD signal is observed out to 96 hours.

Surface plasmon resonance (SPR) is a technique where a host (ligand) is immobilized either covalently or non-covalently (avidin-biotin reactions) to a chip. The guest (analyte) is dissolved in a suitable buffer and passed over the chip. When interactions between the ligand and analyte take place, this changes the refractive index of the chip, and as such can monitor the association and dissociation of an analyte to a ligand. SPR has been used on multiple occasions to measure the binding parameters of small molecules binding to DNA.[184,185] Due to the current uncertainty of the mechanism

for molecular recognition of **3** to its preferred binding site within dsDNA, one primary concern is the suggested use of hairpin oligonucleotides for immobilization on a sensor SPR chip. The interactions of some small molecules with hairpin oligonucleotides has been addressed,[186] but the complex topology of the system between **3** and its preferred binding site is a cause for experimental concern. Also, there are other particular drawbacks to this type of SPR experiment, mostly that extremely slow dissociation rates (resultant from an overall K_D < 10 nM) are not adequately measured due to the extremely long incubation times to reach equilibrium. An approximate "rule of thumb" is that the experiments require $4.6/k_{off}$ seconds to reach an acceptable equilibrium before measuring dissociation.[187] Provided that using our previously determined value of $k_{off} = 4.94 \times 10^{-7}$ sec^{-1}, that would require a time equivalent to 107 days for the SPR equilibrium to be reached, making this experimental method un-economical at the current time.

3.4 DISCUSSION

One important attribute of other biologically relevant anthracycline intercalators, such as daunomycin, is their overall affinity for the targeted DNA sequence.[188] While the affinity for a particular sequence is important in determining if a small-molecule is capable of showing biological activity, the dissociation rate from the DNA is also critical when observing the significance of interactions with DNA on a biological time-scale. Stopped-flow kinetic measurements were suitable for measuring the dissociation of **3** from the random sequence **oligo 2**. While standard sodium dodecyl sulfate facilitated spectrophotometric methods are known to enhance cationic intercalator dissociation from DNA[179,152] our attempts shown above for sequence specific binding have indicated that

dissociation was still too slow for accurate spectrophotometric data collection. Therefore, we monitored the dissociation rates of **3** from its designed 14 bp binding site within **oligo 1** by a gel mobility-shift assay as seen in **Figure 3.13**.

In order to study association rates with tetraintercalator **3**, a multi-staged approach was pursued. First, a lower limit to the association rate constant was established using a rapidly scanned 1D ^{1}H NMR approach that confirmed the threading polyintercalation topology of the bound structure. Second, the gel mobility-shift assay used for measuring dissociation was also used for providing a rough estimate of the association rate constant. Third, stopped-flow measurements were used to provide an accurate spectroscopic measurement of tetraintercalator **3** binding to the preferred **oligo 1** sequence.

The observed sequence specificity could be the consequence of differences in association rates, dissociation rates or a combination of the two. Based on a 4 x 10^{4}-fold slower dissociation rate constant observed for **3** binding to its preferred sequence compared to a randomly chosen sequence, it is safe to say that specificity is based largely on differences in the dissociation rates. This conclusion was reinforced by the stopped-flow association data that revealed very similar association rate constants for the preferred versus a random sequence.

The threading polyintercalation topology is intuitively expected to manifest an extremely slow dissociation due to the seemingly unlikely molecular rearrangements required for full dissociation. In addition, threading intercalator monomers display relatively slow dissociation compared to non-threading monointercalators. Indeed, using 100 mM NaCl as benchmark conditions, the observed k_d = 5 ± 0.5 x 10^{-7} sec^{-1} for **3** dissociating from its preferred binding sequence, **oligo 1**, corresponds to a dissociation half-life of $t_{1/2}$ = 16 days. We envision that multiple individual steps will be required for full dissociation, so these values are best thought of as corresponding to being rate-

limiting out of several individual rate constants. It is important to note that to the extent these conditions resemble biologically relevant parameters, a dissociation half-life of this time-frame should be capable of disrupting biological interactions perhaps for multiple generations of microorganisms.

The complex topology of a bound threading polyintercalator also raises questions about the detailed mechanism of association. Intuitively, one would predict a rapid and nonspecific association between the positively-charged **3** and the negatively-charged DNA driven largely by electrostatic attraction. Following that, one could envision a large number of distinct states as the molecule searches various topologies and sequences before settling into the final threading polyintercalation binding mode. The stopped-flow measurements confirmed at least a three-step process, consistent with expectation. Perhaps surprisingly, there were significant similarities between the association rate profiles and rate constants for **3** binding to its preferred versus a random sequence, although detailed curve fitting indicated there may be some subtle mechanistic differences between the two.

Unfortunately, no detailed mechanistic interpretation is possible based on the stopped-flow data alone because spectrophotometric monitoring cannot be used to gain an unambiguous structural understanding of the process. For example, it is possible that the observed decreases in NDI absorbance are the result of a step in which the NDI units are simply buried in a DNA groove and not yet intercalated. In addition, some mechanistic features could be very hard to detect as distinct steps because one could imagine several different bound topologies with just two NDI units intercalated, all of which would have very similar overall absorbance values.

The lack of structural information in the stopped-flow experiment was addressed by using 1D ^{1}H NMR studies and qualitative gel-shift measurements. A detailed analysis

of the entire spectrum and especially the imino proton signals verified that at a stoichiometric concentration of 60 μM and after the approximately 5.6 minute mixing time, all of **3** was bound to its preferred site in the expected threading polyintercalator topology as previously assigned by Lee and co-workers (**Figure 3.4**).[8] Unfortunately, because the ^{1}H NMR experiment requires higher concentrations than gel mobility-shifts for detection, the complex was already fully formed before the first scan could be acquired, and no detailed rate information could be obtained.

On the other hand, the gel mobility-shift association experiments were used to provide a qualitative estimate of the association rate constant for **3** binding to its preferred sequence in a manner that survives PAGE. As encountered by other labs working with high affinity DNA-binding small molecules, concentration restrictions and lack of accurate timing between initial mixing and gel loading can limit the accuracy of this approach.[7] Recall that the gel-shift was only observed for **3** binding its preferred sequence, not a randomly chosen sequence, so this measurement is assumed to verify binding to the oligonucleotide in a sequence-specific fashion. The approximate association rate constant obtained in this way was $0.3 \pm 0.05\ \mu M^{-1}\ sec^{-1}$, a value that is consistent with full binding under the conditions of the NMR experiment, but is still about an order of magnitude slower than the slower of the two rate constants determined by stopped-flow kinetic experiments. At this time it is not possible to reconcile whether this difference is due to experimental error in the gel-shift measurement (uncertain incubation time, etc.), spectroscopically silent steps in the full association mechanism, or a combination of the two. (**Note:** Refer to Amy Smith dissertation for more appropriate measurements of the association gel shift experiments.)

Currently, it is not clear what an equilibrium association constant, K_a, actually means in the case of **3** binding its preferred binding site due to the complex mechanism of

association and presumably dissociation as well. Nevertheless, for **3** binding its preferred site at 100 mM NaCl, we can derive a benchmark value of 6 x 10^{11} M^{-1} for the apparent association constant, K_a, using the slowest values determined via gel-shift measurements, and thus presumably rate-determining steps. A similar analysis assigns an apparent K_a = 1.6 x 10^7 M^{-1} for **3** binding a randomly chosen sequence. Despite the quantitative caveats already mentioned, these values do justify designation of this interaction as being qualitatively both highly sequence selective and relatively high affinity.

The salt dependence on dissociation of **3** from both **oligo 1** and **oligo 2** were also monitored. The log-log plots of the dissociation rate constant versus [Na^+] both gave a linear fit with a slope of 0.59 and 0.54 for dissociation of **3** from **oligo 1** and **oligo 2**, respectively (**Figures 3.15** & **3.17** above). The thermodynamic basis for specificity can be analyzed via log-log plots of K_{obs} versus [Na^+]. If the resultant slope is the same for both specific and non-specific interactions, the conclusion is then that both interactions involve the same number of electrostatic interactions, and sequence specificity would then be the result of only the non-electrostatic component of binding.[34] While log-log plots of k_d versus [Na^+] are not directly comparable to those of the overall rate constant, K_{obs}, recall that the on-rates for association of **3** to both **oligo 1** and **oligo 2** are similar with both being described by at least three mechanistic steps. Presumably then, the similarities of slopes for the log k_d versus -log [Na^+] would indicate that specificity is determined by some non-electrostatic component of binding. This result is mostly likely from the difference in the thermodynamically more favorable threading polyintercalation topology of the **3-oligo 1** complex, noting that Lee and co-workers have previously established by ^{1}H NMR spectra, that **3** did not bind a random DNA duplex via a complete threading polyintercalation topology.[8]

These slope values from the log k_d versus -log [Na^+] plots can also be used to indicate the binding mode, as molecules known to demonstrate different DNA binding modes give different and characteristic slopes. For instance, dicationic threading NDI intercalators have a slope around 0.3 while groove binders produce a slope between 1.6-1.8.[134,174] The slope observed for the tetracationic compound **3**, 0.59, is nearly twice that previously reported by the Wilson group for a threading dicationic NDI derivative, consistent with the threading mode of intercalation by **3** and all four positive charges being involved in binding. It is worth noting, that work out of the Nordén lab concerning charged threading ruthenium intercalators[173] has shown some inconsistencies with establishing salt dependency with dissociation rates using the model[134] proposed by Tanious *et al*.

3.5 CONCLUSIONS

As expected, the association mechanism was found to be complex, and at least three mechanistic steps were detected spectroscopically. Future work will expand upon these studies by probing the mechanism by which threading polyintercalators can ”scan” large DNA segments of random sequence to finally reside in the preferred binding site, some of which is detailed in a following chapter.

Sequence specific binding was examined for the threading tetraintercalator **3** binding to DNA through a detailed kinetic analysis. In particular, a roughly 4 x 10^4-fold preference for a specific 14 base-pair sequence was determined relative to a randomly chosen sequence. The specificity was entirely due to differences in dissociation rate constants, which is also supported by the similar slopes for the dissociation log k_d – log

[Na^+] plots of **3** from both **oligo 1** and **oligo 2**. Importantly, the slow dissociation rate constant seen with **3** dissociating from its preferred binding site on **oligo 1** corresponds to a complex half-life of around 16 days at 100 mM NaCl, sufficient to disrupt biological processes for perhaps generations of microorganisms. In addition, noting that adding between 4 and 5 kcal/mol of binding energy to the dissociation rate constant reaches a bound half-life of 90 years, we will be exploiting the modular nature of our threading polyintercalators to make longer molecules with bound half-lives relevant to mammalian lifetimes.

3.6 FUTURE DIRECTIONS

At the current time, an insufficient knowledge of the mechanism of association has made salt dependence studies and an estimation of an association on-rate, k_{on}, inappropriate. Further investigation might illuminate a few likely possibilities, thereby providing reason to examine the Na^+ effects on association, eventually leading to an overall estimation of K_{obs} values at a range of Na^+ concentrations. A log-log plot of the K_{obs} versus [Na^+] would then allow for the estimation of the non-electrostatic contribution to binding by the extrapolation of the plots to a standard state of 1 M [Na^+].[34] Coupled with possible DSC studies using a Micro-cal DSC, a full thermodynamic analysis of the sequence specific binding of **3** to its preferred binding site might illuminate interesting characteristics surrounding threading polyintercalation. Perhaps more importantly, we may be able to distinguish the thermodynamic data for sequence-specific binding within a longer sequence of DNA, as opposed to the more traditional thermodynamic

measurements on poly(dAdT), poly(dGdC), or carrier DNA (calf thymus, herring sperm, etc.).[7]

3.7 Materials and Methods

3.7.1 General

Materials. Rink Amide resin (0.72 mmol/g), benzotriazol-1-yl-oxytrispyrrolidinophosphonium hexafluorophosphate (PyBOP), and all necessary Fmoc protected amino acids were purchased from Novabiochem. All enzymes were purchased from New England Biolabs unless otherwise noted. All other chemicals were purchased from Aldrich, Acros, or Fisher unless otherwise noted. All DNA oligonucleotides were purchased from IDT DNA (Coralville, IA) purified by polyacrylamide gel electrophoresis. γ-P^{32} ATP was purchased from Perkin Elmer Life & Analytical.

All of the stopped-flow kinetic experiments were performed by Dr. Maha Zewail Foote of Southwestern University in Georgetown, Texas.

3.7.2 UV-vis Kinetics

In a typical association experiment, 500 μL of 0.5 mM of DNA stock solution in 10 mM Na-PIPES buffer (containing 50 mM NaCl, pH = 7.0) was carefully mixed at room temperature with 500 μL of 0.3 μM **3** stock solution into a masked reduced volume quartz cuvette (Type 9B, 10 mm path length) (Starna Cells, Atascadero, CA). A Teflon cap was placed on the cuvette and wrapped with parafilm. The cuvette was placed in a cell attached to a water re-circulating bath set at 24°C, and the UV absorbance monitored at 386 nm for 36,000 s, with scans every 4 s.

For a typical dissociation experiment, 0.5 mL of the resulting association incubation of DNA/compound **3** complex solution was mixed with 0.5 mL of 4% SDS (10 mM PIPES, 50 mM EDTA, pH = 7.0) solution into a masked reduced volume quartz cuvette as described above. A Teflon cap was placed on top and wrapped with parafilm. The cuvette was then placed in a cell attached to a water re-circulating bath set at 24°C, and the UV absorbance monitored at 386 nm for 36,000 s, with scans every 4 s, and thereafter up to 100 hr with scans every 4 hr. (Model 8453 UV-Vis spectrophotometer, Agilent Technologies). The error estimate was determined to be ± 8% for this experiment.

3.7.3 Isothermal Titration Calorimetry

ITC experiments were performed on a VP-ITC (Microcal Inc., Northhampton, MA). Solutions of **3** and **oligo 1** were prepared in 10 mM PIPES buffer (pH = 7.0) with 1 mM EDTA, and 100 mM NaCl. (Concentrations determined by UV vis, λ_{max} = 385 nm, ε = 51,300 M^{-1} cm^{-1})The buffer solutions were sonicated and filtered prior to addition of either compound **3**, or **oligo 1**. Before adding to the ITC cells, all solutions were de-gassed on the ITC pump for 30 minutes. To the reference cell was added 1.4 ml of buffer solution (10 mM PIPES, 1 mM EDTA, 100 mM NaCl). To the sample cell was added 1.11 ml of the solution of **oligo 1** at a final concentration of 7.0 μM. 290 μL of a solution of **3** (52.45 μM) was injected as 29 individual 10 μL injections with 7 min between injections at 25°C. The data was analyzed and fitted using Origin 7.0 Software (Northhampton, MA).

3.7.4 Circular Dichroism

Solutions of **3**, **oligo 1**, and **oligo 2** were prepared in 10 mM PIPES buffer (pH = 7.0) with 1 mM EDTA and 100 mM NaCl to a final concentration of 2.8 μM. Each experiment was performed on a Jasco J-810 CD spectrophotometer and monitored under the following conditions: wavelength from 200 – 450 nm, a scan rate of 200 nm min^{-1}, a bandwidth of 1 nm, averaging time of 0.125 s, and the number of scans equal to 4 at 24°C. Control spectra were taken for each oligonucleotide duplex diluted with the same buffer at a final concentration of 1.4 μM. The same self-masking reduced volume quartz cuvette (Type 9B, 10 mm path length) (Starna Cells, Atascadero, CA) was used for each experimental run, cleaned and prepared as described.[183] To determine if an ICD signal was present, a scan was taken for **oligo 1** as prepared above at 1.4 μM as a control. Next an equal amount of a prepared solution of **3** (2.8 μM) was mixed with a 2.8 μM solution of **oligo 1**, to effect a stoichiometric amount of **3** and **oligo 1**, and complex formation was monitored over the course of 96 hours.

3.7.5 1D ^{1}H NMR Sample Preparation and Spectroscopy

Solutions were prepared in 90% H_2O / 10% D_2O containing 30 mM phosphate buffer (pH = 7.0) and 100 mM NaCl. To one solution was added **oligo 3** to a final concentration of 120 μM verified by nanodrop spectroscopy (ε = 143,400 M^{-1} cm^{-1} ssMW = 4,261.8 g mol^{-1}). To the other solution was added compound **3** to a final concentration of 120 μM as verified by UV-vis spectroscopy (ε = 51,300 M^{-1} cm^{-1}).[135] Control spectra were obtained for **oligo 3** at 60 μM final concentration. The kinetic association was then monitored by mixing a stoichiometric amount of compound **3** as prepared above, directly into an NMR tube containing **oligo 3** as prepared above, to a final 1:1 concentration of 60 μM.

Experiments were performed using a Varian DirectDrive 600 MHz spectrometer. 1D NMR spectra in 90% H_2O / 10% D_2O were taken at 27°C using a jump-return solvent suppression method. All spectra were processed using VNMRJ (Varian, Inc.).

3.7.6 Gel Mobility-Shift Assays for Dissociation

The synthetic oligonucleotides (**oligo 1, oligo2**) (page purified) forward strand was 5' end labeled and purified as previously described for DNase I footprinting.[170] After annealing with the complement strand, the DNA was purified on an 8% non-denaturing polyacrylamide gel and extracted as described above. After EtOH precipitation, the DNA was resuspended in 200 µl of ddiH_2O (concentration determined by nanodrop spectrophotometer). Incubations were prepared at 24°C in 10mM PIPES buffer (pH = 7.0), 1mM EDTA, with NaCl concentrations ranging from 25 – 200 mM. Compound **3** was added in 0.1 µM excess to drive association. Incubations were separated on an 8% non-denaturing PAGE run at 75 W and 250 V. Gels were extracted and dried as described for sequencing gels above. The gel was then exposed on a phosphor screen for 15 minutes, and analyzed using Image Quant 4.1 imaging software to verify complete complex association. To effect dissociation, non-labeled **oligo 1** was added in 100-fold excess to the associated **3-oligo 1** complex. Aliquots were then taken at representative times and separated on an 8% non-denaturing polyacrylamide gel. After extraction and drying as described previously, gels were exposed to a phosphor screen for 15 min – 48 h, and analyzed using ImageQuant 4.1 imaging software to quantify amount of bound and unbound radiolabeled **oligo 1**.

3.7.7 Gel Mobility-Shift Assays for Association

The synthetic oligonucleotide (**oligo 1**) (page purified) was labeled and purified as described for dissociation above. Concentrations were quantified using a nanodrop spectrophotometer. Incubation were prepared in 10 mM PIPES buffer (pH = 7.0), 1 mM EDTA, and 100 mM NaCl at 24°C. To the buffer was added 5 nM of radiolabeled **oligo 1**, unlabeled **oligo 1** to arrive at a final concentration of **oligo 1** (labeled/unlabeled) ranging from 250 to 2500 nM. A stoichiometric amount of compound **3** was added and mixed to a final volume of 50 μL to begin association. Aliquots where then taken from respective incubations and mixed briefly with non-denaturing loading buffer containing xylene cyanol and bromophenol blue for tracking prior to loading on an 8% non-denaturing PAGE for separation at 75 W and 250 V. Gels were extracted and dried as mentioned above, then exposed to a phosphor screen for 4 hr. The unbound and bound **oligo 1** was then quantified using ImageQuant 4.1 imaging software.

3.7.8 Stopped-Flow Kinetics

Kinetic data was acquired using a KinTek stopped-flow instrument equipped with 2.5-cm path length optical cell. The change in absorbance at 386 nm was monitored for the association and dissociation of **3** with DNA duplexes containing the target binding site or a random sequence at 24°C. Kinetic experiments of DNA association were conducted at constant tetramer concentration and varying DNA concentration. DNA duplexes (ranging from 4 to 32 μM) in buffer containing 10 mM Pipes (pH 7.0), 1 mM EDTA , and 100 mM NaCl were mixed with equal volumes of 4 μM **3** in the same buffer. No absorbance change occurred when the tetramer was mixed with buffer alone. For the dissociation experiments, equilibrated tetramer-oligonucleotide complexes in 10 mM Pipes (pH 7.0), 1 mM EDTA, and the designated sodium chloride concentration (50-300

mM) were rapidly mixed with a 4% SDS solution in the same buffer. Data collected at each timescale consisted of one thousand data points. Global data fitting of the averaged kinetic traces was done using the KinTek Global Kinetic Explorer software (KinTek Corp, Austin, TX).

3.7.9 Isothermal Melting Studies and DSC

Equal concentrations of either **oligo 1** or **oligo 2** were incubated with a stoichiometric amount of compound **3** (2 μM) for 3 days at room temperature. Thermal melting was accomplished by monitoring the absorbance at 260 nm from 20°C to 100°C increasing by 1°C per minute.

DSC trials were run only with a 6.8 mM solution of **oligo 2** in H_2O, using a volume of 200 μL on a DSC Q100 by TA Instruments.

Chapter 4

Investigating a Mechanism for Molecular Recognition and Divalent Salt Dependence on Dissociation

4.1 Chapter Summary

4.1.1 Goals

There are two parts to this chapter, with the first goal to examine the mechanism for molecular recognition for a threading NDI tetraintercalator towards a specific 14 bp sequence of dsDNA inserted within a longer sequence of dsDNA.

They second goal is to observe the salt dependent effects of divalent Mg^{2+} cations on the dissociation rate constant, k_{off}, of tetraintercalator **3** for its designed 14 bp binding site in a mixed monovalent/divalent solution when the incubations are held at a constant ionic strength of 100 mM.

4.1.2 Approach

Synthesize NDI tetraintercalator **3** as discussed in a previous chapter. A plasmid construct can then be prepared through ligation of an insert containing three sequential tetraintercalator binding sites, separated by 10 bp each (one full turn of the double helix). Primers can then be designed and ordered for use in a PCR reaction to create a desired sequence of dsDNA, which also contains the three sequential tetraintercalator binding sites at different distances from each terminal end of the duplex. This positioning of the three binding sites should allow us to distinguish between two extreme mechanisms for molecular recognition, either a macroscopic association/dissociation, or a zipper-like sliding mechanism. Incubation of tetraintercalator **3** with the DNA containing the three

binding sites (**3BS**) can then be analyzed for specificity and a populated preference at each site via DNase I footprinting.

To study the salt dependence for dissociation of **3** with its binding site (**oligo 1**, **Figure 3.1** in Ch. 3), gel mobility-shift assays were used to monitor the dissociation of **3** from **oligo 1** under a variety of mixed monovalent [Na^+] and divalent [Mg^{2+}] conditions held at a constant 100 mM ionic strength. Autoradiographic densitometry can then be used to quantify the percent of bound and unbound **oligo 1** for each incubation, which can then be plotted versus time. Curve fitting the resultant dissociation plots to a monoexponential decay function will allow for the observed k_{off} to be extracted. Comparisons with the off-rate data from the previous chapter will then provide a more complete picture of salt effects on the rate of dissociation with respect to a threading polyintercalation mode of binding.

4.1.3 Results

Synthesis and purification of the NDI-based tetraintercalator was achieved as described earlier. It was shown that tetraintercalator **3** binds to each of its preferred binding sites, $d(GATAAGTACTTATC)_2$, with high specificity and with approximately equal populations at each of three sequential binding sites after approximately 12 hours. These results suggest a mechanism for molecular recognition being similar to that of a pure macroscopic association/dissociation mechanism. However, cooperativity of binding cannot be ruled out at this time. Further experiments will be necessary to discriminate between favorable distortions in the double helix that lead to cooperative binding, and a true association/dissociation type of mechanism.

Dissociation of **3** with its preferred binding site, **oligo 1**, was accomplished using gel mobility-shift assays where **3** was competed off using excess unlabeled preferred

binding site in 100 mM ionic strength monovalent/divalent salt incubations. The resultant dissociation rates at varying $[Na^+]/[Mg^{2+}]$ from 100/0 mM to 0/33 mM ranged from an apparent k_d = 5.0 ± 0.5 x 10^{-7} sec^{-1} to k_d = 6.8 ± 0.7 x 10^{-6} sec^{-1}, respectively. This represents a change in dissociation rate-constant of approximately one order of magnitude between a 100 mM ionic strength solution of only monovalent $[Na^+]$ to only divalent $[Mg^{2+}]$. These results are as expected, due to the increased affinity of Mg^{2+} for the DNA phosphate backbone over Na^+.

4.2 INTRODUCTION

The topics discussed in this chapter are independent studies with respect to the mechanism for molecular recognition, and the effects of the divalent Mg^{2+} cation on the dissociation of **3** from its preferred binding site on **oligo 1**. However, the two topics have been combined due to the relationship between the mechanisms for specific binding and the respective dependence upon the salt concentrations.

4.2.1 On the Mechanism for Molecular Recognition

Biological processes that occur *in vivo* no doubt require the regulatory proteins involved in DNA regulation to translocate from one site to another on the particular DNA strand(s) involved in order to locate the specific site of interest. This process is incredibly complex when imagining that a specific DNA sequence of interest would likely be surrounded by millions of potential DNA binding sites within a genome, all of which would also convey many of the same structural aspects of the specific binding site of interest. Simple three-dimensional diffusion in solution would noticeably be a slow

process, provided the number of opportunities for non-specific interactions during a typical macroscopic association/dissociation process, meaning that a DNA binding ligand must fully dissociate from the DNA before re-associating to a different sequence repeatedly until it eventually locates a preferred site.

In terms of outlining possible mechanisms for the molecular recognition of specific sequences of DNA, the von Hippel lab has derived models and theory in an attempt to describe four potential mechanistic pathways for an apparent association process of the *E. coli lac* repressor to operator DNA that is faster than that of a typical diffusion-limited process ($k_a \sim 10^{10}$ M^{-1} sec^{-1} as opposed to a diffusion limit of $\sim 10^7 - 10^8$ M^{-1} sec^{-1}).[189,190] This finding led to the presumption that there must be at least two-steps involved in the association process. The first step likely being some non-specific association controlled by three-dimensional diffusion, followed by at least a second step that facilitates the transfer of the binding ligand to the specific site with fewer dimensional freedoms to diffusion, since the target-containing species has already been located in solution.

The first possible model is that for a macroscopic dissociation/re-association process within the domain of the DNA molecule, or an intra-molecular process. In this process, a DNA binding ligand would non-specifically associate with a molecule of DNA, followed by a series of complete dissociation and re-association processes to the same molecule of DNA (hence within the same DNA domain), until eventually locating the specific binding site.

Similar to the above model, is one in which a microscopic dissociation/re-association process can take place. Because all of the DNA base pairs are connected, a correlated search for the specific binding site is possible, where the DNA binding ligand would dissociate, however, it is still in close enough proximity to re-associate to a nearby

site. This type of mechanistic process is commonly referred to as "hopping", in that the ligand can sample several nearby sites through the hopping, or microscopic dissociation/re-association process taking place. Most likely, the previous two types of facilitated ligand transfer are in concert with each other while the ligand searches for the target DNA sequence.

Intersegment transfer is another model, which could be thought of as a ligand being associated to the DNA, when another segment of the same DNA comes within a close enough proximity to allow for a loosely bound ligand to each of the DNA segments during the dynamic process of DNA fluctuation in solution. This process is thought only to occur when a ligand has two separate DNA recognition elements.[191] Upon the segments separating, the bound ligand could potentially leave with either the original or the new segment, thereby possibly facilitating a relocation of the ligand on the same molecule of DNA.

A sliding mechanistic process is the last pathway discussed by Berg *et al*.[189] Sliding involves a non-specific association with DNA, followed by the movement of the bound ligand in a one-dimensional projection along the DNA helix. The ligand can thereby sample many correlated DNA base pairs until it locates its specific binding site or dissociates completely from the DNA. Further studies from the von Hippel group have determined to the most likely transfer process for the association of the *lac* repressor to operator DNA follows a two-step sliding mechanism.[192,193] A cartoon representation summarizing each of the models with respect to the *lac* repressor/operator system can be seen in **Figure 4.1**.

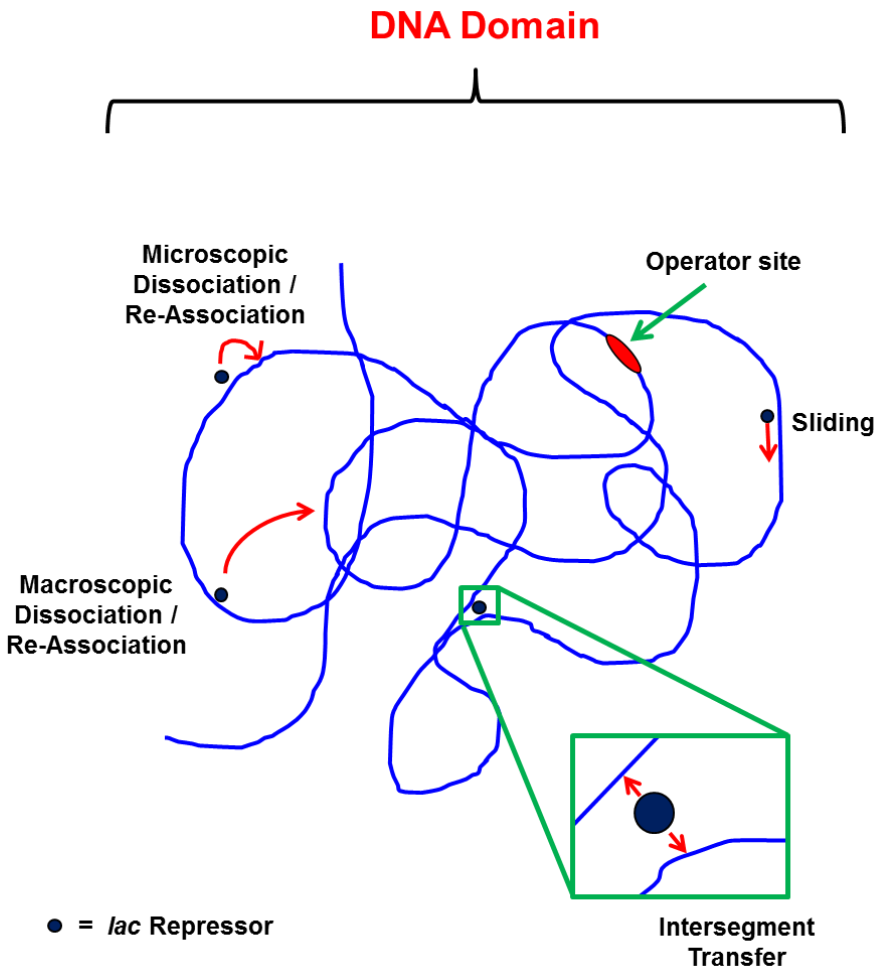


Figure 4.1 Cartoon representation of four facilitated diffusion models as indicated in the figure for a theoretical *lac* repressor/operator system. The DNA molecule is indicated by a blue line, with red arrows showing the respectful paths of diffusion.

Essentially, facilitated diffusion of a DNA binding ligand to a specific site can be summarized in a way where the non-specific associations with a DNA strand actually reduce the amount of time to locate a specific binding site. This is accomplished by increasing the initial target size, the simultaneous reducing of the dimensionality of diffusion of the ligand through close contacts of the DNA, or perhaps by the intra-

molecular re-associations that principally reduce the volume of the solution for which the ligand must search for the target site. This reduced volume is because the ligand is constrained to search only one DNA molecule for a target sequence as opposed to a solution of DNA molecules for the same sequence. Current research outlines examples of proteins that "slide" (one dimensional),[193,194] as well as those which "hop" (three dimensional) to locate specific binding sites.[195] Not surprisingly, several examples of DNA binding proteins express a hybrid of facilitated diffusion pathways to locate their specific binding sites.[196,197]

4.2.2 Divalent Cation Effects on Ligand-DNA Interactions

As discussed in Chapter 3, DNA binding molecules are often positively charged, especially the intercalating class of compounds. These charged species rely, at least partially, on some electrostatic component of binding that can be better studied with respect to the salt dependence in solution. The previous chapter discussed the importance of a dependence upon monovalent salt species such as Na^+, although physiological systems contain many electrolyte species, including divalent ions such as Mg^{2+}. The importance of the studies contained in this chapter, with respect to the divalent Mg^{2+} cation, are not currently known. It is the purpose to cover an understanding of how divalent species and solutions containing mixed monovalent/divalent salts generally affect DNA, and how the presence of divalent species in solution can alter the overall binding interactions to DNA. With this, we might report our results in the event that future salt dependent studies with compound **3** might illuminate our current understanding to our particular threading polyintercalation system.

First, the effects of ions on only DNA has been theoretically examined through Manning's polyelectrolyte theory, which has been extended to divalent counterions as well as to mixed monovalent/divalent systems on the helix-coil transition of DNA. Like monovalent ions, divalent salts such as Mg^{2+} also condense onto polyanions, with Mg^{2+} ions condensing to DNA with a higher preference over monovalent counterions. This preference is shown through the correlation of the linear charge density of the polyanion to the dimensionless parameter, ξ,[198,199] with a value of 4.2 for standard double helical B-form DNA in solution. In solutions containing divalent cations, the condensation of the ions to the DNA reduces ξ to ½, as opposed to a value of approximately 1 for strictly monovalent systems.[200] This is a mathematical manipulation performed by Manning, and suggests that when present in solution, divalent counterions will condense onto the polyanion (DNA) until ξ reduces to ½, or until the population of the divalent species has been depleted. Below $\xi = 1$, no further monovalent ion condensation occurs. This means that when $\xi \leq 1$, no monovalent ions should be condensed, but rather they would have been released into solution in mixed monovalent/divalent systems by the ion-pairing of the divalent species, demonstrating mathematically, a preference for Mg^{2+} condensation to DNA over that of Na^{+}.[200] More current Metropolis Monte Carlo and Poisson-Boltzmann calculations have been performed, and have shown good agreement to Manning's counterion condensation theory described both above and in Chapter 3.[201]

Thomas, as well as Dove and Davidson, have experimentally measured the melting transitions of DNA in the presence of ions such as Na^{+} and Mg^{2+}, and have shown that the double helix is stabilized more so by divalent ions than monovalent ions and that the affinities for DNA are stronger at lower ionic strengths (less than 500 mM).[202,203] The results from Dove and Davidson also suggested that increased stabilization from the Mg^{2+} was likely also a result of the appropriate size of the ion,

perhaps being more capable of achieving favorable stabilizing effects due to sterics, leading to increased affinities for the DNA phosphates as well.[203] Rowatt and Williams have determined the approximate affinities for several cations to DNA, with the overall dissociation constant for Mg^{2+} to be $K_D \sim 6.5 \times 10^{-7}$ M.[204]

Record *et al.* have also studied the ability of Mg^{2+} to compete with proteins for binding sites on DNA, with a reported site range for Mg^{2+} between 2.0 – 2.7 phosphates per bound Mg^{2+}.[34] Stated differently, that is 0.37 – 0.50 Mg^{2+} ions per phosphate, as compared with the previously discussed 0.88 Na^+ ions per phosphate. This competition for the DNA effectively reduces the overall binding constant, K_{obs}, for the protein-DNA interaction.[34] Not only is the binding constant of the interaction affected, but they also found that experimental results confirm Manning's counterion condensation theory with respect to Mg^{2+} ions reducing the Na^+ ion condensation on the DNA.[205] This in turn decreases the quantity of Na^+ ions released upon the protein binding the DNA, and ultimately reduces the Na^+ dependence on the overall K_{obs} for the protein-DNA interaction.[34] In these mixed systems, K_{obs} is often not dependent upon the ionic strength of the solution, but usually to the direct binding of both Na^+ and Mg^{2+} on the DNA.[171] It is noted then, that a plot of the log (k_d) versus the -log([Na^+]) for the dissociation of a protein or ligand at different concentrations of Mg^{2+} is expected to produce a non-linear dependence to the [Na^+].

4.3 RESULTS

4.3.1 Design

4.3.1.1 Distinguishing Two Extreme Mechanisms for Molecular Recognition

There are two possible extreme mechanisms that could take place while **3** searches for its 14 bp binding site within a long sequence of DNA. The first extreme would be a macroscopic association/dissociation mechanism, where **3** fully intercalates the DNA followed by a complete dissociation from the DNA, repeated until the preferred binding site is located. In this case, specificity could be controlled by either a fast on-rate to the specific binding site and/or a very slow off-rate from the preferred site. The next extreme would be a zipper-like sliding mechanism, where **3** would enter the double-helix from one of the two ends, and move up and down the helix with respect to the base pairs opening and closing as it searches for the preferred binding site. This case for specific binding would be determined by a slow off-rate from the binding site. The reality most likely falls somewhere in between these two extremes, however, if we can obtain experimental evidence supporting one extreme over the other, we will be able to develop a more complete picture for the dynamic mechanism of molecular recognition.

To differentiate between the two extremes discussed above, we can design a long sequence of DNA containing three sequential preferred binding sites (hereafter referred to as **3BS**). Each binding site can then be separated by 10 bp of DNA to ensure a full turn in the DNA between binding sites in an effort to avoid any helical distortions brought about upon binding of the DNA by compound **3**. In the event of a zipper-like sliding mechanism, the positioning of the three binding sites with respect to the proximity of the terminal ends of the duplex should facilitate discrimination between the two outside binding sites if they are located closer to one end of the overall sequence. Using a DNase

I footprinting analysis, we should be able to observe the relative population of each binding site with respect to the other two. Specificity was previously demonstrated for the binding of **3** to its preferred binding site in Ch. 2. These studies attempt to extract mechanistic information, so the most appropriate type of experiment would be a time-dependence study at a concentration that yields good resolution, or 250 nM of compound **3** as demonstrated in Ch. 2. **Figure 4.2** shows a cartoon representation of the design for a **3BS** sequence appropriate for such a time-dependent study.

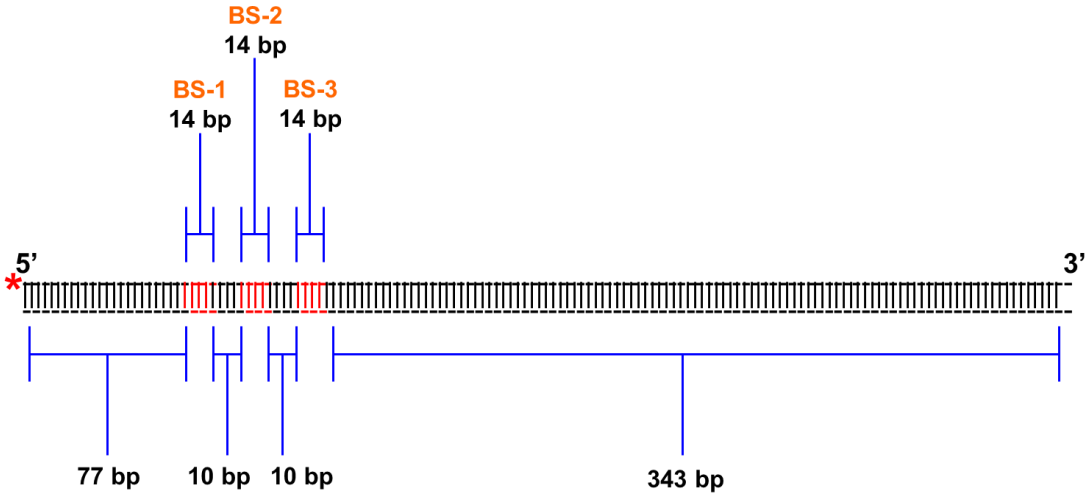


Figure 4.2 Cartoon representation of a 482mer, **3BS** construct, with each of the three respective binding sites, d(GATAAGTACTTATC)$_2$, shown in red and named as indicated. Base pair lengths and separations are shown with brackets. For DNase I footprinting purposes, the 5' end radiolabel placement is indicated by a red asterisk.

Leaving one flanking end much longer than the other, if compound **3** enters through an end of the DNA helix, this design should create a "race" towards the preferred site. We would anticipate, that on average, the binding site adjacent the shorter flanking sequence (**BS-1, Figure 4.2**) should be populated more-so than that of the binding site adjacent the longer flanking sequence (**BS-3, Figure 4.2**), due to a fewer amount of base pairs to traverse. We would expect the middle binding site (**BS-2, Figure 4.2**) to be relatively unoccupied, due to the trapping of **3** in one of the outside binding sites first and thereby blocking any further molecules of **3** from passing through to the middle.

In the event of a macroscopic association/dissociation for an extreme mechanism, we would anticipate that all three of the binding sites produce a similar occupation by **3**. This would be due to the likelihood that any particular sequence is just as capably of being bound by **3**. Also, no blocking of the interior binding site by the two bound exterior sites should be encountered due to the need for **3** to continuously dissociate completely from the DNA and then re-associate until locating the preferred site.

In order to distinguish the two mechanistic extremes, two complementary 85mer oligonucleotides containing the three successive binding sites separated by 10 bp were ordered, and annealed as the three binding site insert (**3BS**) (**Figure 4.3**). The plasmid vector pMoPac16,[166] as shown in **Figure 2.8** of Ch. 2, was chosen for ligation of the **3BS** insert. Forward and reverse primers were then designed and ordered such that a PCR reaction with the **pMoPac16-3BS** construct produced a 482mer sequence of dsDNA with the binding site region 77 bp from the 5' end, and 343 bp from the 3' end as shown above in **Figure 4.2**. Previous DNase I footprinting studies have verified the specificity of **3** for its designed binding site. This work will be similar, in that compound **3** will be presented with approximately 500 bp of random plasmid DNA, in search for three preferred binding sites.

(+) Strand

5' - CGGCCACATGGATAAGTACTTATCGACTACTGCTGATAAGTACTTATC
GACTACTGCTGATAAGTACTTATCACATGGGCCTCGG - 3'

(-) Strand

3' - TCGGCCGGTGTACCTATTCATGAATAGCTGATGACGACTATTCATGAATAG
CTGATGACGACTATTCATGAATAGTGTACCCGGA - 5'

Figure 4.3 Sequence of the 85mer DNA used as the **3BS** insert into pMoPac16. The sequences for the three binding sites of compound **3** are shown in red and the sticky ends are highlighted in green.

4.3.1.2 Divalent Mg^{2+} Dissociation Studies

The gel mobility-shift studies in Ch. 3 indicate the ability of **3** to illicit a retardation of bound **oligo 1** upon separation using an 8% native PAGE gel. We can use an identical experimental setup, with the addition of added Mg^{2+}, to quantify the percent bound and unbound **oligo 1**, and ultimately extract a value for k_{off}, in the same fashion as put forth in Ch. 3.

4.3.2 Synthesis

The vector pMoPac16 (**Figure 2.8**, Ch. 2) was chosen for insertion of the **3BS** insert, again**,** because it contains two *Sfi*I restriction sites with asymmetric sticky ends that allow for a controlled directional insertion. Both the forward and reverse strands of

the **3BS** insert were ordered. After annealing the complementary **3BS** strands, the insert was ligated into the *Sfi*I digested pMoPac16 vector to prepare the desired **pMoPac-3BS** vector containing three sequential tetraintercalator **3** binding sites.

4.3.3 DNase I Footprinting with the pMoPac16-3BS Construct

The same PCR primers for **pMoPac16-BS** (**Figure 2.16**, Ch. 2) were also used with **pMoPac16-3BS**, with a design such that the forward and reverse primers will produce a sequence of dsDNA which places the first tetraintercalator **3** binding site (**BS-1, Figure 4.1**) 77 bp from the 5' end of the forward strand and the last binding site (**BS-3, Figure 4.1**) 343 bp from the 3' end of the forward strand. Similar melting properties were incorporated into both primers, as discussed previously in Ch. 2. An identical protocol was followed for the labeling, PCR amplification, and extraction of the final 482mer **3BS** DNA sequence. The (+) strand of the 482mer sequence is shown in **Figure 4.4**.

pMoPac16-3BS Sequence

5' – GGAGATATACATATGAAATACCTATTGCCTACGGCAGCCGCTGGATTGTTATTACT

CGCGGCCCAGCCGGCACATGGATAAGTACTTATCGACTACTGCTGATAAGTACTTATC

GACTACTGCTGATAAGTACTTATCACATGGGCCTCGGGGGCCGAATTCGCGGCCGCT

GCACCATCTGTCTTCATCTTCCCGCCATCTGATGAGCAGTTGAAATCTGGAACTGCCT

CTGTTGTGTGCCTGCTGAATAACTTCTATCCCAGAGAGGCCAAAGTACAGTGGAAGG

TGGATAACGCCCTCCAATCGGGTAACTCCCAGGAGAGTGTCACAGAACAGGACAGCA

AGGACAGCACCTACAGCCTCAGCAGCACCCTGACGCTGAGCAAAGCAGACTACGAG

AAACACAAAGTCTACGCCTGCGAAGTCACCCATCAGGGCCTGAGTTCGCCCGTCACA

AAGAGCTTCAACCGCGGAGAGTCAG - 3'

Figure 4.4 DNA sequence of the 482mer (+) strand following PCR on **pMoPac16-3BS** construct. The three compound **3** binding sites are highlighted in red.

The DNase I footprinting data of 250 nM of compound **3** incubated with the 482mer **3BS** sequence in a time-dependence study over a 36 hour time frame is shown in **Figure 4.5**. The footprint of **3** continues to show specificity only for its designed binding site $d(GATAAGTACTTATC)_2$ within the presented 482mer sequence of dsDNA. Density trace analysis, as well as the obvious band enhancements between binding sites, indicates that binding has occurred at each of the three sites after only 4 hours of incubation time. However, it is clear from the data, that a distinct footprint does not become apparent until between 12 and 16 hours of incubation, and does so with an approximately equal occupation of all three sites. This is similar to the incubation

concentration of 250 nM of **3** becoming apparent after 4 hours and distinct after 8 hours of incubation in the single binding site of **pMoPac16-BS** shown in **Figure 2.20** of Ch. 2.

Density trace analysis reveals that each of the three binding sites share a relatively equal population at each site. However, a simple visual inspection seems to reveal more occupation at **BS-1** compared to either **BS-2** or **BS-3** after 12 hours. To address this, the same incubations were loaded on another 8% denaturing PAGE in a stacked manner, such that one loading was separated from an identical loading on the same gel delayed by 50 minutes. The results of the DNase I footprint for the stacked gel can be seen in **Figure 4.6**. It is clear that upon allowing the gel to run longer to obtain increased resolution at **BS-3**, the equal occupation of all three binding sites is more apparent visually, in agreement with the approximated occupations from the respective density traces.

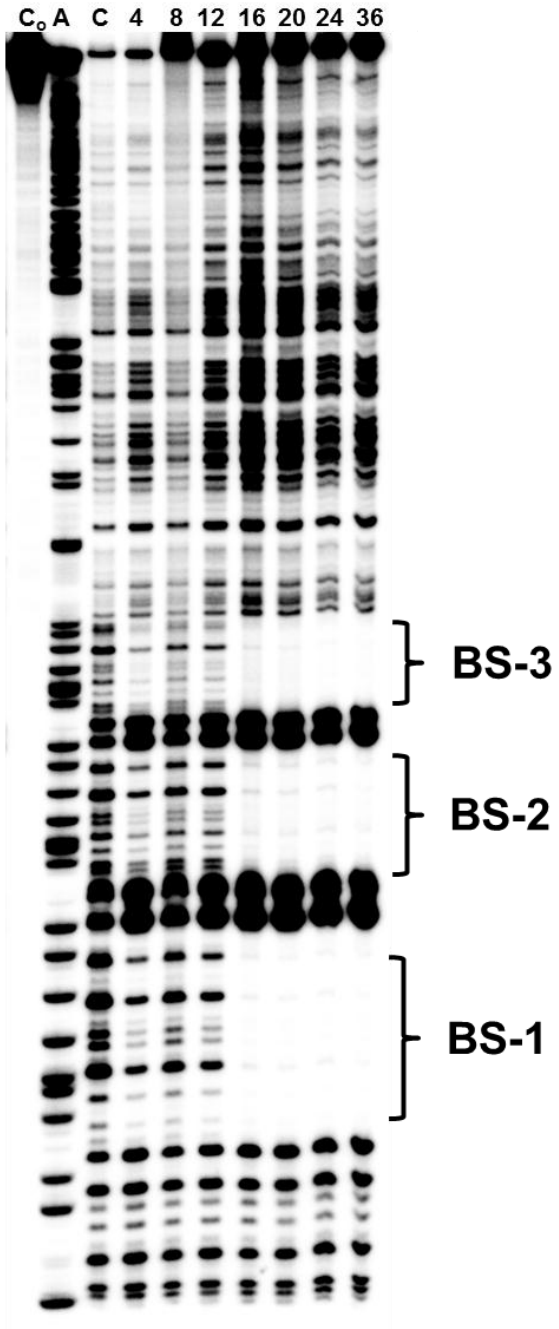


Figure 4.5 DNase I footprint of 482mer PCR fragment from **pMoPac16-3BS** incubated with 250 nM of **3**. The respective binding sites are labeled as in **Figure 4.2**. Lane A represents an adenine-specific sequencing reaction. Lane C contains DNA with DNase I digestion but no **3**. Lane C_0 contains DNA with no DNase I digestion and no **3**. All other lanes represent incubation times in hours.

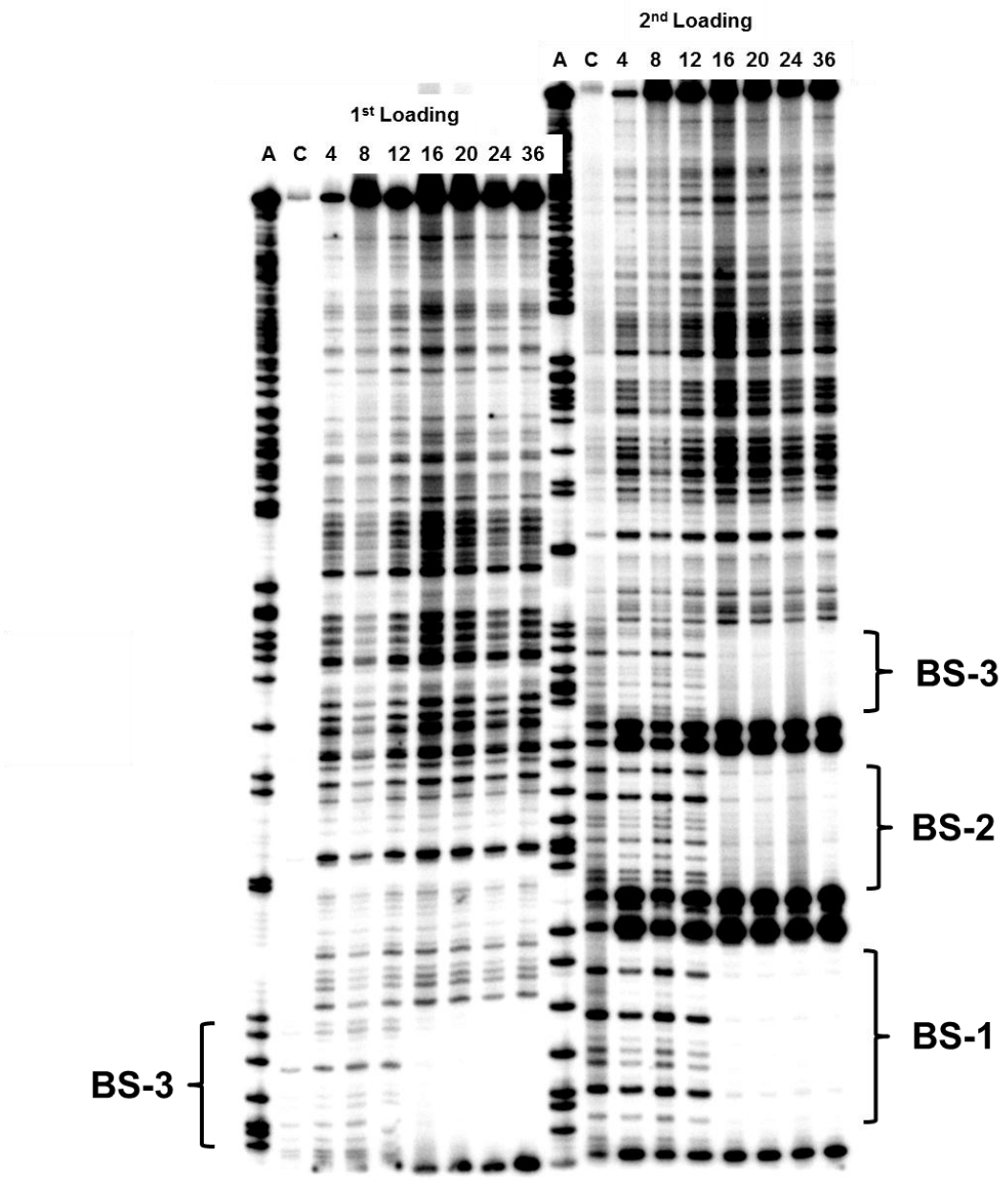


Figure 4.6 DNase I footprint from the previous incubation seen in **Figure 4.5**. However, the loading was repeated twice, with the 2nd loading delayed by 50 minutes from the 1st loading. The respective binding sites are labeled as in **Figure 4.2**. Lane A represents an adenine-specific sequencing reaction. Lane C contains DNA with DNase I digestion but no **3**. Lane C_0 contains DNA with no DNase I digestion and no **3**. All other lanes represent incubation times in hours.

4.3.4 Gel Mobility-Shift Dissociation Assays in Mixed Na^+/Mg^{2+} Solutions

To quantify a value for k_{off}, a gel mobility-shift assay was used identical to that performed in Ch. 3 for dissociation of **3** from **oligo 1**. However, incubations were designed with varying levels of $[Na^+]$ and $[Mg^{2+}]$, such that the overall ionic strength remained at a constant 100 mM. As discussed previously, quantification of the percent bound and unbound **oligo 1** by autoradiography allows for the determination of the respective dissociation off-rate. Representative data of gel-shifts are shown in **Figure 4.7**, and the respective values were plotted as the percent of unbound **oligo 1** versus time. The quantified data was then fit to a standard monoexponential decay equation (equation (**3**), Ch. 3) and can be seen in **Figure 4.8**.

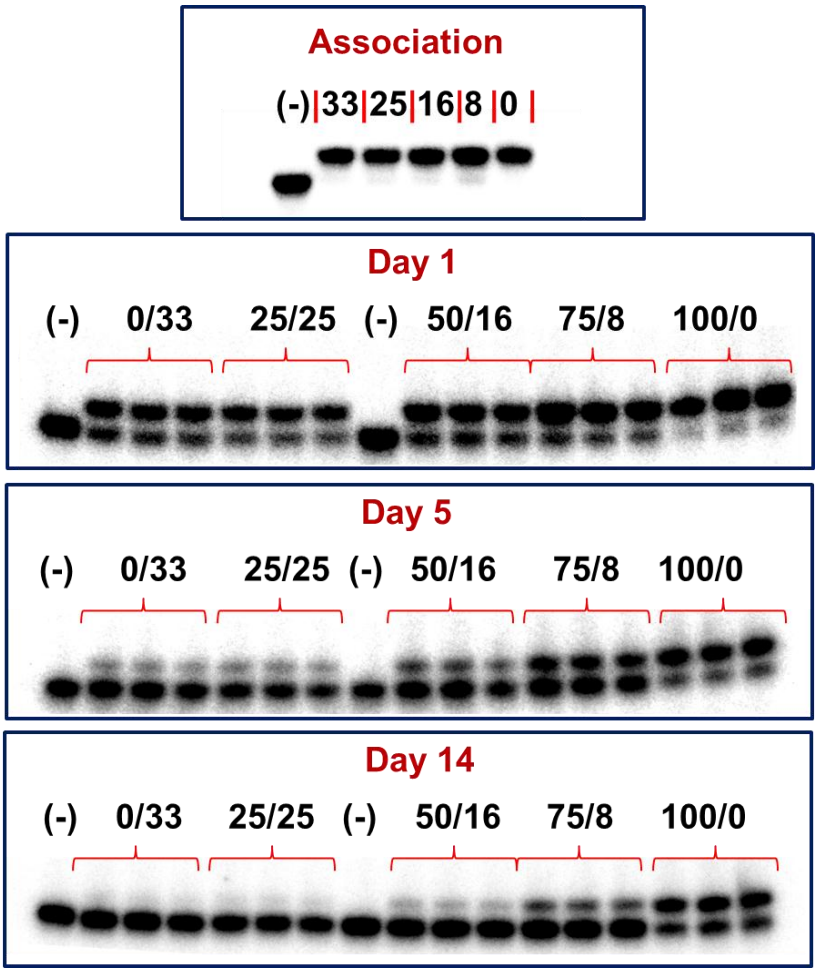


Figure 4.7 Representative Gel-shift dissociation assays in triplicate at the respective labeled time points and at a constant ionic strength of 100 mM with mixed $[Na^{+}]/[Mg^{2+}]$ (mM) ratios as labeled above the lanes. All incubations were performed at 1.0 µM of **3** at 24°C in 10 mM PIPES (1 mM EDTA, pH = 7.0) with 100-fold excess of unlabeled **oligo 1**.

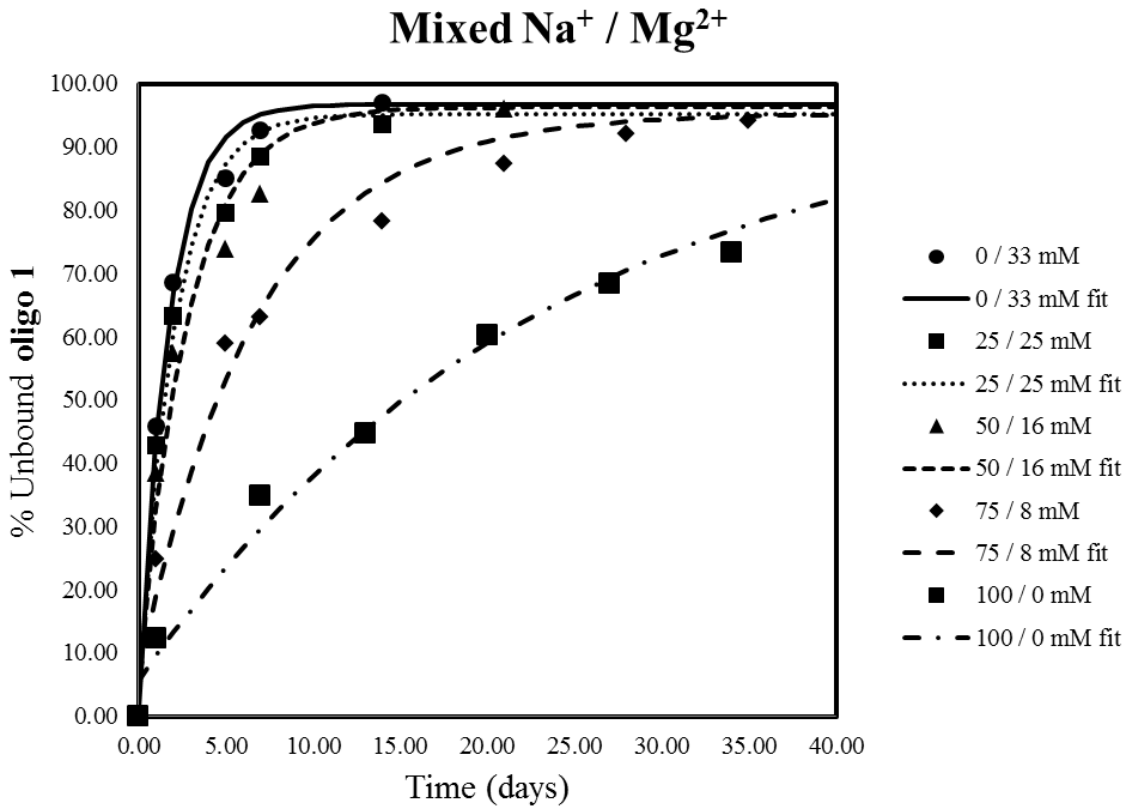


Figure 4.8 Dissociation plot of the percent of unbound **oligo 1** by **3** versus time along with monoexponential curve fits at varying concentration mixtures of $[Na^+]/[Mg^{2+}]$ at a constant 100 mM ionic strength.

The dissociation rate constants were measured for five separate ratios of $[Na^+]/[Mg^{2+}]$ maintained at a constant ionic strength of 100 mM. The following off-rates for the respective ratio of $[Na^+]$ / $[Mg^{2+}]$ were found to increase 10-fold as follows: 100/0 mM ,k_{off} = 5.0 ± 0.5 x 10^{-7} sec^{-1}; 75/8 mM, k_{off} = 1.7 ± 0.3 x 10^{-6} sec^{-1}; 50/16 mM, k_{off} = 4.2 ± 0.7 x 10^{-6} sec^{-1}; 25/25 mM, k_{off} = 5.8 ± 0.9 x 10^{-6} sec^{-1}; 0/33 mM, k_{off} = 6.8 ± 0.7 x 10^{-6} sec^{-1}. The corresponding dissociation off-rate constants and their respective

dissociation half-lives are combined in **Table 4.1**. The respective half-lives range from $t_{1/2}$ ~ 28 – 390 hours for the ratios of 0/33 mM to 100/0 mM of $[Na^+]/[Mg^{2+}]$, respectively. While there is a dependence upon NaCl, a plot of log k_d as a function of – log $[Na^+]$ did not (**Figure 4.9**) reveal a linear relationship due to the presence of the divalent Mg^{2+} cations in solution.

$[Na^+]$ (mM)	$[Mg^{2+}]$ (mM)	k_d x 10^{-6} (s^{-1})	$t_{1/2}$ (hr)
100	0	0.5	390
75	8	1.7	111
50	16	4.1	46
25	25	5.8	33
0	33	6.8	28

Table 4.1 Dissociation rates of **3** from **oligo 1** calculated from the monoexponential fits to the gel-shifts generated at a fixed ionic strength of 100 mM, varying the mixture of [Na+] and [Mg2+]. Error levels for k_d from curve fits calculated at below ± 17%. Half-lives are derived from the off-rates and are shown in hours.

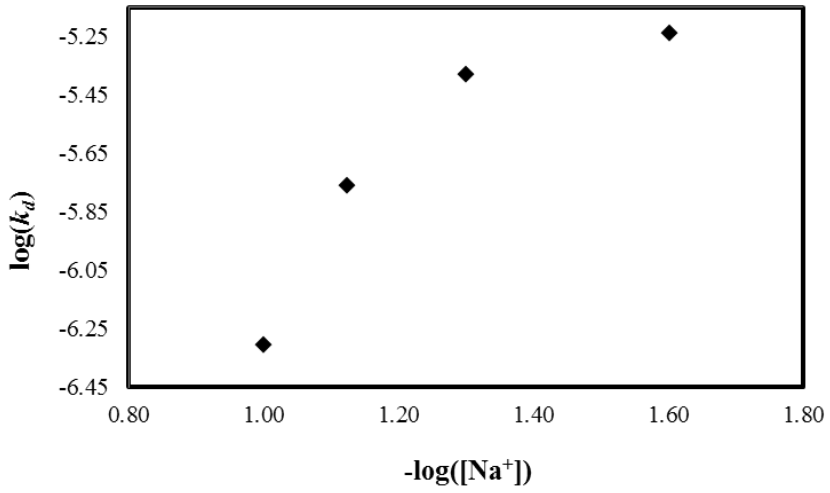


Figure 4.9 Plot of log(k_d) vs –log[Na^+] at varying ranges of [Mg^{2+}] and [Na^+] with a constant ionic strength of 100 mM.

4.4 DISCUSSION

4.4.1 Discussion of DNase I footprinting of 3 with a 3BS DNA sequence

The results of the DNase I footprint in a time dependent study are shown in **Figure 4.5**, and indicate an approximately equal population of all three 14 bp tetraintercalator binding sites using 250 mM of compound **3**. The binding at each site begins after only 4 hours, evidenced through the enhanced bands between binding sites from increased DNase I cleavage, which are most likely due to favorable distortions for the enzyme caused by the binding of **3** to the double helix. However, the footprints do not become truly distinct until between a 12 and 16 hour incubation. The population of each binding site is again, approximately equivalent by densitometry, and the stacked gel shown in **Figure 4.6** shows that at higher resolution of **BS-3**, the occupation is

comparable to that of **BS-1** and **BS-2**. In agreement with the previous time-dependent DNase I footprinting results from Ch. 2, each binding site remains occupied throughout the 36 hour duration of the incubation, which presumably indicates a consistency in the observed remarkably slow dissociation rates from a short 20 bp oligonucleotide discussed in Ch. 3.

Recall from previous discussions in Ch. 2, that in order for compound **3** to fully associate and/or dissociate, an appreciable amount of molecular rearrangement in the DNA helix must take place. This probable rearrangement makes it intuitive that compound **3** would demonstrate slow association on-rates to the specific site, but perhaps more importantly, show remarkably slow dissociation off-rates from the preferred binding site as well. Indeed, this seems to be the case, in that the time dependent analysis reveals distinct footprints for nanomolar concentrations of **3** at the preferred binding site on the order of hours. The slower association to the specific binding site as compared with the stopped flow or ^{1}H NMR studies of Ch. 3 are expected, in that the kinetic analysis were only to a short oligonucleotide, circumventing the need to actually "search" the DNA for its preferred site. One important aspect of the stopped-flow kinetics to the short oligonucleotide, however, was that it confirms our hypothesis of a multi-step association to the DNA binding site.

The equal population of the three binding sites likely indicate that a mechanism for the molecular recognition of the preferred binding site follow some form of a macroscopic dissociation/re-association pathway. While not necessarily considered a "facilitated" search due to on-rates at below diffusion limited values, it is still important to consider how our molecule locates its preferred site. It is highly probably, that the pathway includes the use of both macro- and microscopic dissociation/re-association events, however, there is one particular caveat. While densitometry reveals some binding

to each site after only 4 hours, it is curious that rather than the binding sites appearing more completely occupied by **3** over time (a crescendo appearance to the occupation), that instead they reveal very sudden distinctly bound states between 12 and 16 hours. This leads us to believe that there may be some element of cooperative binding that is occurring. Early DNase I footprinting studies with threading polyintercalators from our lab indicated a likely cooperativity in binding with respect to a bisintercalating compound.[135] This might be due to the event of one threading intercalator binding, thereby creating some favorable helix distortion which might enhance the opportunity for binding by other threading intercalators in solution. Cooperativity in binding to DNA has also been observed for the bisintercalator echinomycin.[206]

4.4.2 Discussion of Divalent Cation Effects on 3-oligo 1 Dissociation

Dissociation of **3** from **oligo 1** was monitored for incubations containing the mixed monovalent and divalent cations of $[Na^+]$ and $[Mg^{2+}]$ maintained at a constant ionic strength of 100 mM. Previous experiments in our lab indicated enhanced dissociation rates in the presence of $[Mg^{2+}]$, so we established an experiment to test this interaction. We had anticipated that the divalent Mg^{2+} would facilitate dissociation due to its ability to disrupt, potentially two ion-pairs of compound **3** with **oligo 1**. The measured dissociation rates (**Table 4.1**) were obtained by fitting the percent of unbound **oligo 1** versus time to a standard monoexponential decay equation (equation **(3)**, Ch. 3).[178] Dissociation rates showed about a 10-fold range from 100/0 mM of $[Na^+]/[Mg^{2+}]$ with $k_d = 5.0 \times 10^{-7}\ s^{-1}$ to a solution of 0/33 mM of $[Na^+]/[Mg^{2+}]$ with $k_d = 6.8 \times 10^{-6}\ s^{-1}$. Plotting log k_d versus –log $[Na^+]$ indicated a dependence on $[Na^+]$ that is not linear. Also, the

increased dissociation rate is expected, provided that Mg^{2+} not only acts as a counter ion in solution, but also as a competitor to the binding of **3** to **oligo 1**. As a result of the direct binding of cations to DNA, Lohman has shown that the effects of Mg^{2+} on protein binding to DNA in mixed valence ion systems are much larger than expected based solely on the ionic strength of the solution.[171] It is not surprising then, that in a mixed Na^+/Mg^{2+} system held at a constant ionic strength, the rate constants observed for **3** dissociating from **oligo 1** are faster at higher concentrations of Mg^{2+}. It is also the case that the competition between Na^+ and Mg^{2+} to re-condense to **oligo 1** upon dissociation of **3** would result in a non-linear dependence upon $[Na^+]$ in a log (k_d) vs –log ($[Na^+]$) plot.[171] Even with the competition from Mg^{2+}, **3** still shows a remarkably slow dissociation rate from **oligo 1** with a half-life equal to 28 hours under the fastest of dissociation conditions (0/33 mM $[Na^+]/[Mg^{2+}]$).

As described elsewhere,[171] a plot of log (k_d) vs –log ($[Na^+]$) should not be linear, which was also the result of our plot seen in **Figure 4.9**. As previously discussed above, our results indicate that the rate of dissociation is not dependent upon ionic strength alone, but that the direct binding of the mixed $[Na^+]$ and $[Mg^{2+}]$ change the amount of condensed cations on the DNA. Also, the non-linear slope for the dissociation rate log (k_d) vs –log ($[Na^+]$) plot support that, in our system, the Mg^{2+} affinity for DNA is acting in competition not only with the Na^+, but also compound **3** in binding to **oligo 1**. Provided these results, it is still remarkable that at 0/33 mM $[Na^+]/[Mg^{2+}]$, direct competition between $[Mg^{2+}]$ and compound **3** still reveals a dissociation half-life of $t_{1/2}$ = 28 hours.

4.5 CONCLUSIONS

4.5.1 On the Mechanism for Molecular Recognition

It was our observation that from DNase I footprinting studies on a 482mer strand of dsDNA containing three sequential tetraintercalator binding sites, that all three revealed approximately equal population between 12 and 16 hours with completely distinct footprints. While this result infers that there is not particularly any facilitated diffusion in the sense of speeding up association rates past diffusion limits, it is a likely indication of some form of macro- or microscopic dissociation/re-association pathway in locating the preferred binding sites. However, it is noted that previous reports of very similar threading bisintercalators demonstrate cooperativity in binding to GC-rich DNA.[135] This observation prevents us from claiming one mechanistic pathway over another due to the high possibility of cooperative binding in our experimental design. Future work as discussed below will outline some potential experiments in which more information might be extracted in an attempt to describe the system of tetraintercalator **3** binding to its preferred site within a long sequence of dsDNA.

4.5.2 Salt Dependence of Mixed Na^+/Mg^{2+} Solutions

It was observed that the dissociation of **3** from its preferred binding site in **oligo 1** was substantially increased through the addition of divalent Mg^{2+} to the incubations. These results are not unexpected, as Mg^{2+} is capable of directly binding the DNA in an effort to compete with compound **3**. While these experiments were performed under a constant ionic strength environment, it quickly became obvious that the dependence on Na^+ in a mixed valence environment is not a linear function (**Figure 4.9**), indicating that ionic strength is not nearly as important for enhanced dissociation as is the concentration

of divalent Mg^{2+} species. In order to draw proper conclusions for the salt dependence of k_d and k_a, an accurate value for K_{obs} must be attainable.[171] At this time, it is not apparent how we might achieve a truly accurate representation for the association rate constant, however, a deeper understanding of the mechanism for molecular recognition might illuminate this task in future experiments.

4.6 FUTURE DIRECTIONS

There are many possibilities for experimental designs that might elucidate the mechanism for molecular recognition of proteins to specific sequences of DNA. However, careful selection must be shown in order to perform the proper experiments for our system which relies upon a small molecule interaction with the DNA. For instance, one of the most common ways for performing extremely fast association assays with proteins that demonstrate faster than diffusion on-rates has been to use quantitative filter binding techniques.[207] This technique utilizes radioactive DNA, which can be detected at very dilute concentrations, in an effort to express protein binding. The proteins adhere to the filter, and unbound DNA can be washed away, leaving only the bound radiolabeled DNA to be quantified. The drawbacks to this approach are that it relies upon the protein to adhere to the filter where our system uses a small molecule, as well as the fact that "bound protein" does not confer any type of sequence specificity. This type of technique is not necessary at this time, provided that our system does not display association on-rates that rival diffusion limits.

There are many possibilities that alter the length of the DNA molecules available in solution, as well as the location of the target binding sites within each molecule. Such

techniques allow for a partitioning between three-dimensional and one-dimensional facilitated transfer mechanisms, and might be of potential use in our system.[208] The one drawback to several of these designs is that the enzymes are capable of cleaving the DNA so that different fragments can be quantitatively analyzed, which is in a direct conflict with a system such as ours where binding occurs without any type of restriction. It might be possible to covalently attach some forms of Fe(II)·EDTA moieties to our molecule, which are capable of cleaving the DNA through oxidation,[209] however it is not currently understood how this might affect the overall association and dissociation of compound **3** to its preferred site.

Covalent crosslinking of the DNA base pairs would prevent any type of zipper-like sliding mechanism discussed above, however, so should the bound **3**-DNA complex. This leads to the conclusion that a zipper-like sliding mechanism can be ruled out at this point. However, we cannot yet rule out a classical sliding mechanism, in which the electrostatic contacts hold the ligand within the ionic radius of the DNA, and thereby allow it to scan the DNA in a one dimensional fashion. The bound **3**-DNA complex does not significantly enhance the radius of the DNA, as well as not necessarily adequately blocking out the phosphate backbones either. This might still allow for sliding to occur. Therefore, it might be more appropriate to "block" the tetraintercalator binding sites using a much larger molecule, such as the *lac* repressor protein by placing adjacent operator binding sites to the 14 bp preferred site of compound **3**.

I have currently designed and synthesized a **pMoPac16-2BS** vector (the **2BS** insert can be seen in **Figure 4.10**), where two sequential tetraintercalator binding sites are separated by 34 base pairs (~ 3.5 full turns of the helix) so that we might be able to rule out the potential cooperativity in binding. The design allows for PCR amplification where one binding site is only 20 bp in from the 5' side, with 34 bp separating the next binding

site, and an overall sequence length of 500 base pairs (**Figure 4.10**). With 88% of the overall sequence being to the 3' side of the 2nd binding site, we might expect compound **3** to first associate with that portion of the sequence, allowing for a likely binding site discrimination in the case of one dimensional sliding. The spacing between binding sites should also preclude any cooperativity in binding. It is likely that to truly use this experimental design, a few more "two binding site" sequences should be designed, such that the binding sites are placed in different locations and with different separations within the overall sequence, similar to those types of experiments discussed for proteins above.[208] Since our system does not currently allow for DNA cleavage, some care would have to be shown in determining the proper incubation times for DNase I footprinting analysis. It may be possible to take advantage of a quench-flow DNase I footprinting technique which couples the ability of fast data collection in a stopped-flow type apparatus with quantitative DNase I footprinting techniques, which might be the best approach for illuminating the potential mechanistic pathway.[210] This rapid footprinting technique has been useful in providing further mechanistic detail for the *lac* repressor/operator system indicating that both sliding and intersegment transfer might play important roles in the target site location.[211]

(+) Strand

5'- CGGCCACATGGATAAGTACTTATCGACTACTGCTGACTACTGCTGACTGACTACTGCTGA
TAAGTACTTATCACATGGGCCTCGG-3'

(-) Strand

3'- TCGGCCGGTGTACCTATTCATGAATAGCTGATGACGACTGATGACGACTGACTGATGAC
GACTATTCATGAATAGTGTACCCGGA -5'

pMoPac16-2BS Sequence

5'- GCGGCCCAGCCGGCCACATGGATAAGTACTTATCGACTACTGCTGACTACTGCTGACTG

ACTACTGCTGATAAGTACTTATCACATGGGCCTCGGGGGCCGAATTCGCGGCCGCTGCACC

ATCTGTCTTCATCTTCCCGCCATCTGATGAGCAGTTGAAATCTGGAACTGCCTCTGTTGTGT

GCCTGCTGAATAACTTCTATCCCAGAGAGGCCAAAGTACAGTGGAAGGTGGATAACGCCCT

CCAATCGGGTAACTCCCAGGAGAGTGTCACAGAACAGGACAGCAAGGACAGCACCTACAG

CCTCAGCAGCACCCTGACGCTGAGCAAAGCAGACTACGAGAAACACAAAGTCTACGCCTG

CGAAGTCACCCATCAGGGCCTGAGTTCGCCCGTCACAAAGAGCTTCAACCGCGGAGAGTC

AGTCGACCATCATCATCACCATCACGGGGCCGCAGAACAAAAACTCATCTCAGAAGAGGAT

CTGAATGGGCGCGCCG -3'

Forward pMoPac16-2BS Primer:

5'- GCGGCCCAGCCGGCC -3'

Reverse pMoPac16-2BS Primer:

5'- CGGCGCGCCCATTCAG -3'

Figure 4.10 Both (-) and (+) strands for the designed **2BS** insert, (+) strand of the 500mer sequence, and primers for **pMoPac16-2BS** construct. Binding sites are highlighted in red, with *SfiI* sticky ends highlighted in green.

In light of learning more details regarding the mechanism for molecular recognition, it would also be interesting to observe the Na^+ dependence upon the association of **3** to **oligo 1**. Most likely, a dependence upon the direct binding of the cations rather than the ionic strength would be observed in mixed monovalent/divalent association reactions, which would again infer a multi-step association reaction mechanism, in support of the stopped-flow kinetic data presented in Chapter 3.[171]

The door is undoubtedly open with respect to this project. There are still many details to be worked through in order to further understand a system of threading polyintercalation. While focus is currently being given to developing other polyintercalators capable of binding altered DNA binding sites, or perhaps longer sequences of DNA, the full capabilities of our molecules are currently unknown. What we do know, is that there is a truly unique and useful quality of our threading polyintercalators in their extraordinarily long dissociation rates from preferred binding sites. Compound **3** represents the most likely candidate for understanding, in more detail, the full range of characteristics involved in threading polyintercalation as a useful binding mode of DNA.

4.7 MATERIALS AND METHODS

4.7.1 General

General chemicals and synthesis of compound **3** follow the same details and procedures as outlined in Chapters 2 and 3.

4.7.2 pMoPac16-3BS Three Binding Site Construct

The vector pMoPac16-**3BS** was prepared exactly as described in Ch. 2 for the **pAK400-BS** construct. Successful ligation of the **3BS** insert was confirmed for **pMoPac16-3BS** by di-deoxy sequencing.

4.7.3 Preparation of 5' – ^{32}P end labeled DNA fragments

Forward and reverse primers for **pMoPac16-3BS** were ordered and re-suspended to 100 μM stock solutions in ddiH_2O. The procedure for end-labeling is detailed in Ch. 2. The 482mer **3BS** dsDNA sequence was produced by PCR amplification from **pMoPac16-3BS**, followed by purification and extraction from a native 5% PAGE gel as described in Ch. 2.

4.7.4 DNase I Footprinting

The incubations for DNase I footprinting were performed as discussed in Ch. 2. The resulting digested DNA fragments were then separated using an 8% denaturing

PAGE at 75 W and ≤ 2000 V. The sequencing gels were extracted from the plates using 3 mm Whatman chromatography paper (Fisher) and dried for 45 min on a Model 583 gel dryer from Bio-Rad (Hercules, CA). After drying, the gels were exposed on a phosphor screen (18 - 24 hr), and imaged using either Quantity One 4.5 software from Bio-Rad (Hercules, CA) or ImageQuant TL software from GE Healthcare (Piscataway, NJ).

4.7.5 Gel Mobility-Shift for Dissociation including Mg^{2+}

The synthetic oligonucleotide, **oligo 1**, was prepared as described in Ch. 3 (concentration determined by nanodrop spectrophotometer). Incubations were prepared at 24°C in 10mM PIPES buffer (pH = 7.0), 1mM EDTA, with mixed NaCl/$MgCl_2$ concentrations of 100/0 mM, 75/8.33 mM, 50/16.66 mM, 25/25 mM, and 0/33.33 mM, so that the final ionic strength was equal to 100 mM. Compound **3** was added in 0.1 μM excess to drive association, although, at concentrations of $MgCl_2$ ≥ 16 mM, it was necessary to increase excess **3** up to ~ 0.7 μM excess to drive complete association. Incubations were separated on an 8% non-denaturing PAGE run at 75 W and 250 V. Gels were extracted and dried as described for sequencing gels above. The gel was then exposed on a phosphor screen for 15 minutes, and analyzed using Image Quant 4.1 imaging software to verify complete complex association. To effect dissociation, non-labeled **oligo 1** was added in 100-fold excess to the associated **3-oligo 1** complex. Aliquots were then taken at representative times and separated on an 8% non-denaturing polyacrylamide gel. After extraction and drying as described previously, gels were exposed to a phosphor screen for 15 min – 48 h, and analyzed using ImageQuant 4.1 imaging software to quantify amount of bound and unbound radiolabeled **oligo 1**.

References

(1) Müller, W.; Crothers, D. M. *Journal of Molecular Biology* **1968**, *35*, 251–290.
(2) Zhang, J.-H. Chung, T. D. Y.; Oldenburg, K. R. *Journal of Biomolecular Screening* **1999**, *4*, 67-73.
(3) Hallikas, O.; Taipale, J. *Nat Protoc* **2006**, *1*, 215-222.
(4) Rishi, V. Potter, T. Laudeman, J. Reinhart, R. Silvers, T. Selby, M. Stevenson, T. Krosky, P. Andrew Stephen; Acharya, A. Moll, J. Jun Oh, W. Scudiero, D. Shoemaker, R., H.; Vinson, C. *Analytical Biochemistry* **2005**, *340*, 259-271.
(5) Dervan, P. B. *Science* **1986**, *232*, 464-471.
(6) Watts, C. R. Kerwin, S. M. Kenyon, G. L. Kuntz, I. D.; Kallick, D. A. *Journal of the American Chemical Society* **1995**, *117*, 9941–9950.
(7) Leng, F. Priebe, W.; Chaires, J. B. *Biochemistry* **1998**, *37*, 1743–1753.
(8) Lee, J. Guelev, V. Sorey, S. Hoffman, D. W.; Iverson, B. L. *J. Am. Chem. Soc* **2004**, *126*, 14036–14042.
(9) Chu, Y. Sorey, S. Hoffman, D. W.; Iverson, B. L. *Journal of the American Chemical Society* **2007**, *129*, 1304-1311.
(10) Crick, F. *Nature* **1970**, *227*, 561-563.
(11) Müller, W.; Crothers, D. M. *Journal of Molecular Biology* **1968**, *35*, 251–290.
(12) Portugal, J.; Waring, M. J. *Nucleic acids research* **1986**, *14*, 8735.
(13) Wang, A. H. J. Ughetto, G. Quigley, G. J.; Rich, A. *Biochemistry* **1987**, *26*, 1152–1163.
(14) Nickols, N. G.; Dervan, P. B. *Proceedings of the National Academy of Sciences* **2007**, *104*, 10418.
(15) Muzikar, K. A. Nickols, N. G.; Dervan, P. B. *Proceedings of the National Academy of Sciences* **2009**, *106*, 16598.
(16) Fire, A. Xu, S. Montgomery, M. K. Kostas, S. A. Driver, S. E.; Mello, C. C. *Nature* **1998**, *391*, 806-811.
(17) Elbashir, S. M. Harborth, J. Lendeckel, W. Yalcin, A. Weber, K.; Tuschl, T. *Nature* **2001**, *411*, 494-498.
(18) Drygin, D. Lin, A. Bliesath, J. Ho, C. B. O'Brien, S. E. Proffitt, C. Omori, M. Haddach, M. Schwaebe, M. K. Siddiqui-Jain, A. Streiner, N. Quin, J. E. Sanij, E. Bywater, M. J. Hannan, R. D. Ryckman, D. Anderes, K.; Rice, W. G. *Cancer Research* **2010**, *71*, 1418-1430.
(19) Wei, W.-H. Fountain, M. Magda, D. Wang, Z. Lecane, P. Mesfin, M. Miles, D.; Sessler, J. L. *Org. Biomol. Chem.* **2005**, *3*, 3290.
(20) Bonacci, T. M. Mathews, J. L. Yuan, C. Lehmann, D. M. Malik, S. Wu, D. Font, J. L. Bidlack, J. M.; Smrcka, A. V. *Science* **2006**, *312*, 443 -446.
(21) Ong, C. C. Jubb, A. M. Haverty, P. M. Zhou, W. Tran, V. Truong, T. Turley, H. O'Brien, T. Vucic, D. Harris, A. L. Belvin, M. Friedman, L. S. Blackwood, E. M. Koeppen, H.; Hoeflich, K. P. *Proceedings of the National Academy of Sciences* **2011**, *108*, 7177 -7182.
(22) Gottesfeld, J. M. Neely, L. Trauger, J. W. Baird, E. E.; Dervan, P. B. *Nature* **1997**, *387*, 202-205.

(23) Gottesfeld, J. M. Neely, L. Trauger, J. W. Baird, E. E.; Dervan, P. B. *Nature* **1997**, *387*, 202-205.
(24) Cantor, C.; Smith, C. *Genomics: The Science and Technology Behind the Human Genome Project*; Wiley, 1999.
(25) Egli, M. Williams, L. D. Frederick, C. A.; Rich, A. *Biochemistry* **1991**, *30*, 1364–1372.
(26) Van Dyke, M.; Dervan, P. *Science* **1984**, *225*, 1122 -1127.
(27) Niedle, S. *Principles of Nucleic Acid Structure*; Elsevier, 2007.
(28) Lee, J. Toward Threading Polyintercalators with Programmed Sequence Specificity, University of Texas at Austin, 2003.
(29) Ivanov, V. Grzeskowiak, K.; Zocchi, G. *J. Phys. Chem. B* **2003**, *107*, 12847–12850.
(30) Mazur, A. K. *J. Am. Chem. Soc* **2003**, *125*, 7849–7859.
(31) Gunter, M. *The Chemical Biology of Nucleic Acids*; Wiley, 2010.
(32) Davey, C. A. Sargent, D. F. Luger, K. Maeder, A. W.; Richmond, T. J. *Journal of Molecular Biology* **2002**, *319*, 1097-1113.
(33) Manning, G. S. *J. Chem. Phys.* **1969**, *51*, 924.
(34) Record, J. Lohman, T. M.; Haseth, P. de *Journal of Molecular Biology* **1976**, *107*, 145-158.
(35) Esteller, M. *European Journal of Cancer* **2000**, *36*, 2294–2300.
(36) Jardim, M. J. Wang, Q. Furumai, R. Wakeman, T. Goodman, B. K.; Wang, X.-F. *Mol. Biol. Cell* **2009**, *20*, 3801-3809.
(37) White, S. Szewczyk, J. W. Turner, J. M. Baird, E. E.; Dervan, P. B. *Nature* **1998**, *391*, 468-471.
(38) Anslyn, E. V.; Dougherty, D. A. *Modern Physical Organic Chemistry*; University Science Books, 2006.
(39) Early, T. A. Kearns, D. R. Hillen, W.; Wells, R. D. *Biochemistry* **1981**, *20*, 3756–3764.
(40) Leroy, J. L. Gao, X. Gueron, M.; Patel, D. J. *Biochemistry* **1991**, *30*, 5653–5661.
(41) Leroy, J. L. Gao, X. Misra, V. Gueron, M.; Patel, D. J. *Biochemistry* **1992**, *31*, 1407–1415.
(42) Folta-Stogniew, E.; Russu, I. M. *Biochemistry* **1994**, *33*, 11016–11024.
(43) Cao, C. Jiang, Y. L. Stivers, J. T.; Song, F. *Nature Structural & Molecular Biology* **2004**, *11*, 1230–1236.
(44) Lyakhov, I. G. Hengen, P. N. Rubens, D.; Schneider, T. D. *Nucleic Acids Research* **2001**, *29*, 4892 -4900.
(45) Schneider, T. D. *Nucleic Acids Research* **2001**, *29*, 4881 -4891.
(46) Alexandrov, B. S. Gelev, V. Yoo, S. W. Alexandrov, L. B. Fukuyo, Y. Bishop, A. R. Rasmussen, K. Ø.; Usheva, A. *Nucleic Acids Research* **2010**, *38*, 1790 -1795.
(47) Jose, D. Datta, K. Johnson, N. P.; von Hippel, P. H. *Proceedings of the National Academy of Sciences* **2009**, *106*, 4231 -4236.
(48) Glover, J. N. M.; Harrison, S. C. *Nature* **1995**, *373*, 257-261.
(49) Jamieson, A. C. Miller, J. C.; Pabo, C. O. *Nat Rev Drug Discov* **2003**, *2*, 361-368.
(50) Iverson, B. L. Shreder, K. Kral, V.; Sessler, J. L. *Journal of the American Chemical Society* **1993**, *115*, 11022-11023.

(51) Iverson, B. L. Shreder, K. Král, V. Sansom, P. Lynch, V.; Sessler, J. L. *Journal of the American Chemical Society* **1996**, *118*, 1608-1616.
(52) Sessler, J. L. Sansom, P. I. Král, V. O'Connor, D.; Iverson, B. L. *Journal of the American Chemical Society* **1996**, *118*, 12322-12330.
(53) Arakawa, H. Ahmad, R. Naoui, M.; Tajmir-Riahi, H.-A. *Journal of Biological Chemistry* **2000**, *275*, 10150 -10153.
(54) Pasternack, R. F. Gibbs, E. J.; Villafranca, J. J. *Biochemistry* **1983**, *22*, 2406-2414.
(55) ROSENBERG, B. VANCAMP, L. TROSKO, J. E.; MANSOUR, V. H. *Nature* **1969**, *222*, 385-386.
(56) Kalben, A. Pal, S. Andreotti, A. H. Walker, S. Gange, D. Biswas, K.; Kahne, D. *Journal of the American Chemical Society* **2000**, *122*, 8403-8412.
(57) Sherman, C. L. Pierce, S. E. Brodbelt, J. S. Tuesuwan, B.; Kerwin, S. M. *J Am Soc Mass Spectrom* **2006**, *17*, 1342-1352.
(58) Kopka, M. L. Yoon, C. Goodsell, D. Pjura, P.; Dickerson, R. E. *Journal of Molecular Biology* **1985**, *183*, 553-563.
(59) Finlay, A. C. Hochstein, F. A. Sobin, B. A.; Murphy, F. X. *Journal of the American Chemical Society* **1951**, *73*, 341-343.
(60) Chandra, P. Zimmer, C.; Thrum, H. *FEBS Letters* **1970**, *7*, 90-94.
(61) Wegner, M.; Grummt, F. *Biochemical and Biophysical Research Communications* **1990**, *166*, 1110-1117.
(62) Reddy, B. S. Sondhi, S. M.; Lown, J. W. *Pharmacology & therapeutics* **1999**, *84*, 1–111.
(63) Wilson, W. D. Tanious, F. A. Barton, H. J. Jones, R. L. Strekowski, L.; Boykin, D. W. *Journal of the American Chemical Society* **1989**, *111*, 5008-5010.
(64) Guan, L. L. Zhao, R.; Lown, J. W. *Biochemical and Biophysical Research Communications* **1997**, *231*, 94-98.
(65) O'Hare, C. C. Mack, D. Tandon, M. Sharma, S. K. Lown, J. W. Kopka, M. L. Dickerson, R. E.; Hartley, J. A. *Proceedings of the National Academy of Sciences* **2002**, *99*, 72 -77.
(66) Lown, J. W. Krowicki, K. Bhat, U. G. Skorobogaty, A. Ward, B.; Dabrowiak, J. C. *Biochemistry* **1986**, *25*, 7408-7416.
(67) Pelton, J. G.; Wemmer, D. E. *Proceedings of the National Academy of Sciences of the United States of America* **1989**, *86*, 5723.
(68) Mrksich, M. Wade, W. S. Dwyer, T. J. Geierstanger, B. H. Wemmer, D. E.; Dervan, P. B. *Proceedings of the National Academy of Sciences of the United States of America* **1992**, *89*, 7586.
(69) Szewczyk, J. W. Baird, E. E.; Dervan, P. B. *Journal of the American Chemical Society* **1996**, *118*, 6778-6779.
(70) Herman, D. M. Baird, E. E.; Dervan, P. B. *Journal of the American Chemical Society* **1998**, *120*, 1382-1391.
(71) Turner, J. M. Swalley, S. E. Baird, E. E.; Dervan, P. B. *Journal of the American Chemical Society* **1998**, *120*, 6219-6226.
(72) Herman, D. M. Turner, J. M. Baird, E. E.; Dervan, P. B. *Journal of the American Chemical Society* **1999**, *121*, 1121-1129.

(73) Kielkopf, C. L. White, S. Szewczyk, J. W. Turner, J. M. Baird, E. E. Dervan, P. B.; Rees, D. C. *Science* **1998**, *282*, 111 -115.
(74) Mapp, A. K. Ansari, A. Z. Ptashne, M.; Dervan, P. B. *Proceedings of the National Academy of Sciences* **2000**, *97*, 3930 -3935.
(75) Ansari, A. Z. Mapp, A. K. Nguyen, D. H. Dervan, P. B.; Ptashne, M. *Chemistry & Biology* **2001**, *8*, 583-592.
(76) Dervan, P. B. *Bioorganic and Medicinal Chemistry* **2001**, *9*, 2215–2236.
(77) Gottesfeld, J. Melander, C. Suto, R. Raviol, H. Luger, K.; Dervan, P. *Journal of Molecular Biology* **2001**, *309*, 615-629.
(78) Vasquez, K. M.; Wilson, J. H. *Trends in Biochemical Sciences* **1998**, *23*, 4-9.
(79) Felsenfeld, G. Davies, D. R.; Rich, A. *Journal of the American Chemical Society* **1957**, *79*, 2023-2024.
(80) Moser, H.; Dervan, P. *Science* **1987**, *238*, 645 -650.
(81) Chan, P. P.; Glazer, P. M. *Journal of molecular medicine* **1997**, *75*, 267–282.
(82) Fox, K. R. *Current Medicinal Chemistry* **2000**, *7*, 17–37.
(83) Cassidy, S. A. Slickers, P. Trent, J. O. Capaldi, D. C. Roselt, P. D. Reese, C. B. Neidle, S.; Fox, K. R. *Nucleic Acids Research* **1997**, *25*, 4891 -4898.
(84) Arya, D. P. *Accounts of Chemical Research* **2010**, *44*, 134-146.
(85) Nielsen, P. Egholm, M. Berg, R.; Buchardt, O. *Science* **1991**, *254*, 1497 -1500.
(86) Nielsen, P. E.; Christensen, L. *Journal of the American Chemical Society* **1996**, *118*, 2287-2288.
(87) Egholm, M. Buchardt, O. Nielsen, P. E.; Berg, R. H. *Journal of the American Chemical Society* **1992**, *114*, 1895–1897.
(88) Nielsen, P. E. Egholm, M.; Buchardt, O. *Journal of Molecular Recognition* **1994**, *7*, 165-170.
(89) Lohse, J. Dahl, O.; Nielsen, P. E. *Proceedings of the National Academy of Sciences* **1999**, *96*, 11804 -11808.
(90) Demidov, V. V. Yavnilovich, M. V. Belotserkovskii, B. P. Frank-Kamenetskii, M. D.; Nielsen, P. E. *Proceedings of the National Academy of Sciences* **1995**, *92*, 2637 -2641.
(91) Nielsen, P. E. *Chem. Eur. J. of Chem. Bio.* **2010**, *11*, 2073-2076.
(92) Zhilina, Z. V. Ziemba, A. J.; Ebbinghaus, S. W. *Current Topics in Medicinal Chemistry* **2005**, *5*, 1119–1131.
(93) Hamzavi, R. Dolle, F. Tavitian, B. Dahl, O.; Nielsen, P. E. *Bioconjugate Chemistry* **2003**, *14*, 941-954.
(94) Bohler, C. Nielsen, P. E.; Orgel, L. E. *Nature* **1995**, *376*, 578-581.
(95) Lerman, L. S. *Journal of Molecular Biology* **1961**, *3*, 18-30, IN13-IN14.
(96) Gago, F. *Methods* **1998**, *14*, 277-292.
(97) Crothers, D. M. *Biopolymers* **1968**, *6*, 575-584.
(98) Wheate, N. J. Brodie, C. R. Collins, J. G. Kemp, S.; Aldrich-Wright, J. R. *Mini Reviews in Medicinal Chemistry* **2007**, *7*, 627–648.
(99) Brana, M. F. Cacho, M. Gradillas, A. Pascual-Teresa, B.; Ramos, A. *Current pharmaceutical design* **2001**, *7*, 1745–1780.
(100) Müller, W.; Crothers, D. M. *European Journal of Biochemistry* **1975**, *54*, 267-277.

(101) Sharples, D.; Brown, J. R. *FEBS Letters* **1976**, *69*, 37-40.
(102) Strekowski, L.; Wilson, B. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **2007**, *623*, 3–13.
(103) Patel, D. J.; Canuel, L. L. *Proc Natl Acad Sci U S A* **1977**, *74*, 2624-2628.
(104) Bresloff, J. L.; Crothers, D. M. *Journal of Molecular Biology* **1975**, *95*, 103-110.
(105) Wilson, W. D. Krishnamoorthy, C. R. Wang, Y. H.; Smith, J. C. *Biopolymers* **1985**, *24*, 1941–1961.
(106) Li, H. J.; Crothers, D. M. *Journal of Molecular Biology* **1969**, *39*, 461-477.
(107) Chaires, J. B. Dattagupta, N.; Crothers, D. M. *Biochemistry* **1985**, *24*, 260–267.
(108) Moore, M. H. Hunter, W. N. d' Estaintot, B. L.; Kennard, O. *Journal of molecular biology* **1989**, *206*, 693–705.
(109) Mirau, P. A.; Shafer, R. H. *Biochemistry* **1982**, *21*, 2626-2631.
(110) Jennette, K. W. Lippard, S. J. Vassiliades, G. A.; Bauer, W. R. *Proceedings of the National Academy of Sciences* **1974**, *71*, 3839 -3843.
(111) Zeglis, B. M. Pierre, V. C.; Barton, J. K. *Chem. Commun.* **2007**, 4565.
(112) Barton, J. K. *Science* **1986**, *233*, 727-734.
(113) Sitlani, A. Long, E. C. Pyle, A. M.; Barton, J. K. *Journal of the American Chemical Society* **1992**, *114*, 2303-2312.
(114) Howe-Grant, M.; Lippard, S. J. *Biochemistry* **1979**, *18*, 5762-5769.
(115) Jackson, B. A.; Barton, J. K. *Journal of the American Chemical Society* **1997**, *119*, 12986-12987.
(116) Jencks, W. P. *Proceedings of the National Academy of Sciences* **1981**, *78*, 4046 -4050.
(117) Wakelin, L. P. G. *Medicinal Research Reviews* **1986**, *6*, 275-340.
(118) Waring, M. J.; Wakelin, L. P. G. *Nature* **1974**, *252*, 653-657.
(119) Fox, K. R. Wakelin, L. P. G.; Waring, M. J. *Biochemistry* **1981**, *20*, 5768-5779.
(120) Bailly, C. Braña, M.; Waring, M. J. *European Journal of Biochemistry* **1996**, *240*, 195-208.
(121) Gallego, J.; Reid, B. R. *Biochemistry* **1999**, *38*, 15104-15115.
(122) Bousquet, P. F. Bra\ na, M. F. Conlon, D. Fitzgerald, K. M. Perron, D. Cocchiaro, C. Miller, R. Moran, M. George, J. Qian, X. D.; others *Cancer research* **1995**, *55*, 1176.
(123) Awada, A. Thödtmann, R. Piccart, M. J. Wanders, J. Schrijvers, A. H. G. J. Von Broen, I.-M.; Hanauske, A. R. *European Journal of Cancer* **2003**, *39*, 742-747.
(124) Robinson, H. Priebe, W. Chaires, J. B.; Wang, A. H.-J. *Biochemistry* **1997**, *36*, 8663-8670.
(125) Portugal, J. Cashman, D. J. Trent, J. O. Ferrer-Miralles, N. Przewloka, T. Fokt, I. Priebe, W.; Chaires, J. B. *J. Med. Chem* **2005**, *48*, 8209–8219.
(126) Capelle, N. Barbet, J. Dessen, P. Blanquet, S. Roques, B. P.; Le Pecq, J. B. *Biochemistry* **1979**, *18*, 3354-3362.
(127) Atwell, G. J. Baguley, B. C. Wilmanska, D.; Denny, W. A. *Journal of Medicinal Chemistry* **1986**, *29*, 69–74.
(128) Takenaka, S.; Takagi, M. *Bull. Chem. Soc. Jpn.* **1999**, *72*, 327-337.
(129) Williams, L. D. Egli, M. Qi, G. Bash, P. van der Marel, G. A. van Boom, J. H. Rich, A.; Frederick, C. A. *Proceedings of the National Academy of Sciences* **1990**, *87*, 2225 -2229.

(130) Fox, K. R. Brassett, C.; Waring, M. J. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1985**, *840*, 383-392.
(131) Bhuyan, B. K.; Reusser, F. *Cancer Research* **1970**, *30*, 984 -989.
(132) Li, L. H. Kuentzel, S. L. Murch, L. L. Pschigoda, L. M.; Krueger, W. C. *Cancer Research* **1979**, *39*, 4816 -4822.
(133) Yen, S. F. Gabbay, E. J.; Wilson, W. D. *Biochemistry* **1982**, *21*, 2070–2076.
(134) F. A Tanious; Yen, S. F.; Wilson, W. D. *Biochemistry* **1991**, *30*, 1813-1819.
(135) Lokey, R. S. Kwok, Y. Guelev, V. Pursell, C. J. Hurley, L. H.; Iverson, B. L. *J. Am. Chem. Soc* **1997**, *119*, 7202–7210.
(136) Murr, M. M. Harting, M. T. Guelev, V. Ren, J. Chaires, J. B.; Iverson, B. L. *Bioorganic & Medicinal Chemistry* **2001**, *9*, 1141-1148.
(137) Guelev, V. M. Harting, M. T. Lokey, R. S.; Iverson, B. L. *Chemistry & Biology* **2000**, *7*, 1-8.
(138) Guelev, V. Lee, J. Ward, J. Sorey, S. Hoffman, D. W.; Iverson, B. L. *Chemistry & Biology* **2001**, *8*, 415-425.
(139) Guelev, V. Sorey, S. Hoffman, D. W.; Iverson, B. L. *J. Am. Chem. Soc* **2002**, *124*, 2864–2865.
(140) Chu, Y. Hoffman, D. W.; Iverson, B. L. *Journal of the American Chemical Society* **2009**, *131*, 3499–3508.
(141) Wilson, C. J. Zhan, H. Swint-Kruse, L.; Matthews, K. S. *Biophysical Chemistry* **2007**, *126*, 94-105.
(142) Gilbert, W.; Müller-Hill, B. *Proceedings of the National Academy of Sciences* **1966**, *56*, 1891 -1898.
(143) Romanuka, J. Folkers, G. E. Biris, N. Tishchenko, E. Wienk, H. Bonvin, A. M. J. J. Kaptein, R.; Boelens, R. *Journal of Molecular Biology* **2009**, *390*, 478-489.
(144) Riggs, A. D. Suzuki, H.; Bourgeois, S. *Journal of Molecular Biology* **1970**, *48*, 67-83.
(145) Pfahl, M. *J. Bacteriol.* **1979**, *137*, 137-145.
(146) Hinkle, D. C.; Chamberlin, M. J. *Journal of Molecular Biology* **1972**, *70*, 157-185.
(147) Baliga, R. Baird, E. E. Herman, D. M. Melander, C. Dervan, P. B.; Crothers, D. M. *Biochemistry* **2001**, *40*, 3-8.
(148) Herman, D. M. Baird, E. E.; Dervan, P. B. *Chemistry - A European Journal* **1999**, *5*, 975-983.
(149) Maher, L. J. Dervan, P. B.; Wold, B. J. *Biochemistry* **1990**, *29*, 8820-8826.
(150) Hansen, M. E. Bentin, T.; Nielsen, P. E. *Nucleic Acids Research* **2009**, *37*, 4498 -4507.
(151) Laugaa, P. Markovits, J. Delbarre, A. Le Pecq, J. B.; Roques, B. P. *Biochemistry* **1985**, *24*, 5567-5575.
(152) Westerlund, F. Nordell, P. Nordén, B.; Lincoln, P. *J. Phys. Chem. B* **2007**, *111*, 9132–9137.
(153) Rajpal, A. Beyaz, N. Haber, L. Cappuccilli, G. Yee, H. Bhatt, R. R. Takeuchi, T. Lerner, R. A.; Crea, R. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102*, 8466.
(154) Boder, E. T. Midelfort, K. S.; Wittrup, K. D. *Proc Natl Acad Sci U S A* **2000**, *97*, 10701-10705.

(155) Graff, C. P. Chester, K. Begent, R.; Wittrup, K. D. *Protein Engineering Design and Selection* **2004**, *17*, 293 -304.
(156) Green, N. M. *Biochemical Journal* **1963**, *89*, 585.
(157) Piran, U.; Riordan, W. J. *Journal of immunological methods* **1990**, *133*, 141–143.
(158) Rao, J. Lahiri, J. Isaacs, L. Weis, R. M.; Whitesides, G. M. *Science* **1998**, *280*, 708-711.
(159) Roberts, R. W.; Crothers, D. M. *Proceedings of the National Academy of Sciences of the United States of America* **1991**, *88*, 9397.
(160) Persil, \Ö; Hud, N. V. *Trends in biotechnology* **2007**, *25*, 433–436.
(161) Lane, M. J. Dabrowiak, J. C.; Vournakis, J. N. *Proceedings of the National Academy of Sciences of the United States of America* **1983**, *80*, 3260.
(162) Waring, M. J. *Sequence-specific DNA Binding Agents*; 1st ed. Royal Society of Chemistry, 2006.
(163) Arcamone, F. *Cancer research* **1985**, *45*, 5995.
(164) Ashley, N.; Poulton, J. *Oncogene* **2009**, *28*, 3880-3891.
(165) Krebber, A. Bornhauser, S. Burmester, J. Honegger, A. Willuda, J. Bosshard, H. R.; Plückthun, A. *Journal of immunological methods* **1997**, *201*, 35–55.
(166) Hayhurst, A. Happe, S. Mabry, R. Koch, Z. Iverson, B. L.; Georgiou, G. *Journal of immunological methods* **2003**, *276*, 185–196.
(167) Guelev, V. M. Cubberley, M. S. Meredith M Murr; Lokey, R. S.; Iverson, B. L. In *Methods in Enzymology*; Academic Press, 2001; Vol. Volume 340, pp. 556-570.
(168) Wakelin, L. P. G.; Waring, M. J. *Journal of Molecular Biology* **1980**, *144*, 183–214.
(169) Lokey, R. S. Kwok, Y. Guelev, V. Pursell, C. J. Hurley, L. H.; Iverson, B. L. *J. Am. Chem. Soc* **1997**, *119*, 7202–7210.
(170) Trauger, J. W.; Dervan, P. B. In *Methods in Enzymology*; Academic Press, 2001; Vol. 340, pp. 450-466.
(171) Lohman, T. M.; von Hippel, P. H. *Critical Reviews in Biochemistry and Molecular Biology* **1986**, *19*, 191–245.
(172) Tanious, F. A. Jenkins, T. C. Neidle, S.; Wilson, W. D. *Biochemistry* **1992**, *31*, 11632-11640.
(173) Önfelt, B. Lincoln, P.; Nordén, B. *J. Am. Chem. Soc.* **2001**, *123*, 3630-3637.
(174) Wilson, W. D. Tanious, F. A. Barton, H. J. Jones, R. L. Fox, K. Wydra, R. L.; Strekowski, L. *Biochemistry* **1990**, *29*, 8452–8461.
(175) Johnson, K. A. Simpson, Z. B.; Blom, T. *Analytical Biochemistry* **2009**, *387*, 20-29.
(176) Bevilacqua, P. C. Kierzek, R. Johnson, K. A.; Turner, D. H. *Science* **1992**, *258*, 1355-1358.
(177) Chu, Y. DNA threading intercalation: Building sequence-specific linear rigidified and cyclic bisintercalators, THE UNIVERSITY OF TEXAS AT AUSTIN, 2008.
(178) Fox, K. R. *Drug-DNA Interaction Protocols*; Methods in Molecular Biology; Humana Press, 1997; Vol. 90.
(179) Westerlund, F. Wilhelmsson, L. M. Nordén, B.; Lincoln, P. *J. Am. Chem. Soc* **2003**, *125*, 3773–3779.
(180) Wiseman, T. Williston, S. Brandts, J. F.; Lin, L. N. *Analytical biochemistry* **1989**, *179*, 131–137.

(181) Chaires, J. B. *Biophysics* **2008**, *37*.
(182) McGhee, J. D. *Biopolymers* **1976**, *15*, 1345–1375.
(183) Garbett, N. C. Ragazzon, P. A.; Chaires, J. B. *Nat Protoc* **2007**, *2*, 3166-3172.
(184) Davis, T. M.; Wilson, W. D. *Analytical Biochemistry* **2000**, *284*, 348–353.
(185) Miao, Y. Cui, T. Leng, F.; Wilson, W. D. *Analytical biochemistry* **2008**, *374*, 7–15.
(186) Nguyen, B.; Wilson, W. D. *The Journal of Physical Chemistry B* **2009**, *113*, 14329–14335.
(187) Van Der Merwe, P. A. *Protein-ligand interactions: hydrodynamics and calorimetry: a practical approach* **2001**, 137.
(188) Valentini, L. Nicolella, V. Vannini, E. Menozzi, M. Penco, S.; Arcamone, F. *Farmaco Sci* **1985**, *40*, 377-390.
(189) Berg, O. G. Winter, R. B.; Von Hippel, P. H. *Biochemistry* **1981**, *20*, 6929–6948.
(190) Riggs, A. D. Bourgeois, S.; Cohn, M. *Journal of Molecular Biology* **1970**, *53*, 401-417.
(191) Halford, S. E. *Nucleic Acids Research* **2004**, *32*, 3040-3052.
(192) Winter, R. B.; Von Hippel, P. H. *Biochemistry* **1981**, *20*, 6948–6960.
(193) Winter, R. B. Berg, O. G.; Von Hippel, P. H. *Biochemistry* **1981**, *20*, 6961–6977.
(194) Givaty, O.; Levy, Y. *Journal of molecular biology* **2009**, *385*, 1087–1097.
(195) Hedglin, M.; O'Brien, P. J. *ACS chemical biology* **2010**, *5*, 427–436.
(196) Gowers, D. M. Wilson, G. G.; Halford, S. E. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102*, 15883.
(197) Halford, S. E. *Biochem. Soc. Trans* **2009**, *37*, 343.
(198) Manning, G. S. *J. Chem. Phys.* **1969**, *51*, 924.
(199) Manning, G. S. *Biopolymers* **1972**, *11*, 937-949.
(200) Manning, G. S. *Biopolymers* **1972**, *11*, 951-955.
(201) Pack, G. R.; Lamm, G. *International Journal of Quantum Chemistry* **1993**, *48*, 213-230.
(202) Thomas, R. *Biochimica et Biophysica Acta* **1954**, *14*, 231-240.
(203) Dove, W. F.; Davidson, N. *Journal of Molecular Biology* **1962**, *5*, 467–478.
(204) Rowatt, E.; Williams, R. J. *J. Inorg. Biochem* **1992**, *46*, 87-97.
(205) De Marky, N.; Manning, G. S. *Biopolymers* **1975**, *14*, 1407-1422.
(206) Bailly, C. Hamy, F. Waring, M. J.; others *Biochemistry* **1996**, *35*, 1150–1161.
(207) von Hippel, P. H.; Berg, O. G. *Journal of Biological Chemistry* **1989**, *264*, 675.
(208) Halford, S.; Szczelkun, M. *European Biophysics Journal* **2002**, *31*, 257-267.
(209) Moser, H.; Dervan, P. *Science* **1987**, *238*, 645 -650.
(210) Hsieh, M.; Brenowitz, M. In *RNA Polymerase and Associated Factors, Part B*; Academic Press, 1996; Vol. Volume 274, pp. 478-492.
(211) Hsieh, M.; Brenowitz, M. *Journal of Biological Chemistry* **1997**, *272*, 22092 -22096.

Vita

Garen Gilman Holman was raised in central Arkansas, where he graduated from the University of Central Arkansas with an A.C.S. certified B.S. degree in Chemistry, and a minor in mathematics. He began his undergraduate research at UCA in the lab of Dr. R. Tarkka, where he created a parallel library of peptides on solid phase in search for a mimic to the rat μ-opiod receptor. He later joined the research group of Dr. L. Isom, where he modeled water-pi interactions using a variety of PDB crystal structures of short duplex DNA. Following graduation in 2006, he was admitted to the University of Texas at Austin, where he joined the lab of Dr. B. Iverson in pursuit of a Ph.D. in organic chemistry as well as an introduction to the world of teaching.

Permanent email: garen_holman@hotmail.com

This dissertation was typed by the author.