The Dissertation Committee for Yongjun Chu certifies that this is the approved version of the following dissertation:

DNA Threading Intercalation: Building Sequence-Specific Linear Rigidified and Cyclic Bisintercalators

Committee:

Brent L. Iverson, Supervisor

Eric V. Anslyn

David W. Hoffman

Jonathan L. Sessler

Christian P. Whitman

DNA Threading Intercalation: Building Sequence-Specific Linear Rigidified and Cyclic Bisintercalators

by

Yongjun Chu, B.S.; M.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 2007

Dedication

To my family

Acknowledgements

I would like to thank my advisor, Professor Brent L. Iverson, for all his advising and help. His understanding, encouragement and support have helped me to enjoy my graduate school experience. I cannot imagine having worked for anyone else. I also would like to thank Professor David W. Hoffman for his tremendous help on NMR data collection, and for being inspirational during the discussion of NMR data and structural simulation.

All of my fellow group members, past or current, have provided me kind help that I will never forget. They have been the best group of individuals I could have been around in the laboratory. I am thankful for their friendship. I also want to express my gratitude to Steve Sorey for helping me collect some 1D NMR data.

Sincere love and appreciation go to my wife, Sandy Liu, for being a wonderful support system during stressful times, as well as to my family for their love and encouragement. I would not be where I am today without all of them.

Finally, I would like to thank UT-Austin for accepting me into this fine program and NIH for providing the financial support for all the work I have done.

DNA Threading Intercalation: Building Sequence-Specific Linear Rigidified and Cyclic Bisintercalators

Publication No. _____

Yongjun Chu, Ph.D.

The University of Texas at Austin, 2007

Supervisor: Brent L. Iverson

The threading polyintercalators are based on a 1,4,5,8-naphthalenetetracarboxylic diimide (NDI) unit that binds DNA via threading intercalation with two imide groups situated in different DNA grooves. In order to extend this strategy to bind DNA with higher affinity and more programmable specificity, it was envisioned that the linkers connecting the NDI units had to be rigidified. Thus a rigid, spiro-cyclic linker was designed and synthesized in the context of a bisintercalator, **C1**. The new linker has several sites for possible addition of functional groups as recognition elements. DNAse I footprinting results showed that **C1** has a higher binding affinity ($K_D \sim 10^{-7}$ M) toward 5'-GGTACC-3' sequence than the previously developed dimer **G₃K**, which has a flexible

linker. NMR structural analysis of the **C1**-d(CGGTACCG)$_2$ complex has revealed the versatility of threading polyintercalation based on NDI moieties by verifying the binding of the **C1** linker in the minor groove with two NDI units intercalating between GpG steps. The observed binding specificity of **C1** is the result of interplay of different factors, such as overall linker length, electrostatic and hydrophobic complementarity to their preferred grooves.

The fact that **C1** and **G$_3$K** can target the same DNA sequence via different grooves, but with different linker structures, prompted us to explore the possibility of creating cyclic bisintercalating molecules. The first example, **CBI-1**, has been efficiently synthesized through a solid phase synthesis strategy, in which Gly$_3$Lys, a major groove recognition element for d(GGTACC)$_2$, was linked at one side and (β-Ala)$_3$Lys, a perfect match for a 4-base pair span in the minor groove, was connected on the other side. A dissociation kinetics study on poly(dGdC) indicated a slow dissociation process for **CBI-1**. Data from DNAse I footprinting and NMR structural studies confirmed that **CBI-1** forms a tightly bound complex with DNA d(CG│GTAC│CG)$_2$, in which two NDI units intercalate between GC pairs, with linkers interacting with the major and minor grooves simultaneously. **CBI-1** also exhibits improved sequence specificity compared to the linear dimer **G$_3$K** by only binding 5'-GGTACC-3' sequence. All the results demonstrate that cyclic threading intercalation is a new and effective approach to specifically target DNA sequences.

Table of Contents

List of Tables

List of Figures

xiii

xv

Chapter 1   Overview of DNA Recognition by Small Molecules

1.1 Developing sequence-specific DNA binding ligands as potential therapeutic agents and biological tools

Targeting specific gene sequences has been of great interest to both the pharmaceutical community as well as basic scientists. According to the "central dogma of molecular biology", the genetic information encoded in the double-stranded DNA structure is first transferred to single-stranded messenger RNA structure via the transcription process (Figure 1.1). Following transport to the cytoplasm, the mRNA is translated into polypeptides or protein products on the ribosome with the aid of tRNA. Since many diseases are directly related to malfunctioning proteins (such as enzymes and transcription factors), it is reasonable to propose that controlling gene expression might be able to prevent the appearance of those malfunctioning proteins. Historically, there have been two ways to achieve this goal, namely, targeting DNA or RNA structure. In this dissertation, the focus will be on targeting DNA structure.

Since the discovery that the short double-stranded RNAs regulate gene expression, targeting mRNA to knock-out unwanted gene expression has been an extremely active research area (Fire, 1998; Elbashir, 2001). However, compared to targeting mRNA, targeting DNA structure is, in some sense, a more efficient approach. This is because inside of the non-dividing cells there is only one copy of the DNA sequence, but many copies of transcribed RNA molecules that are replenished with time.
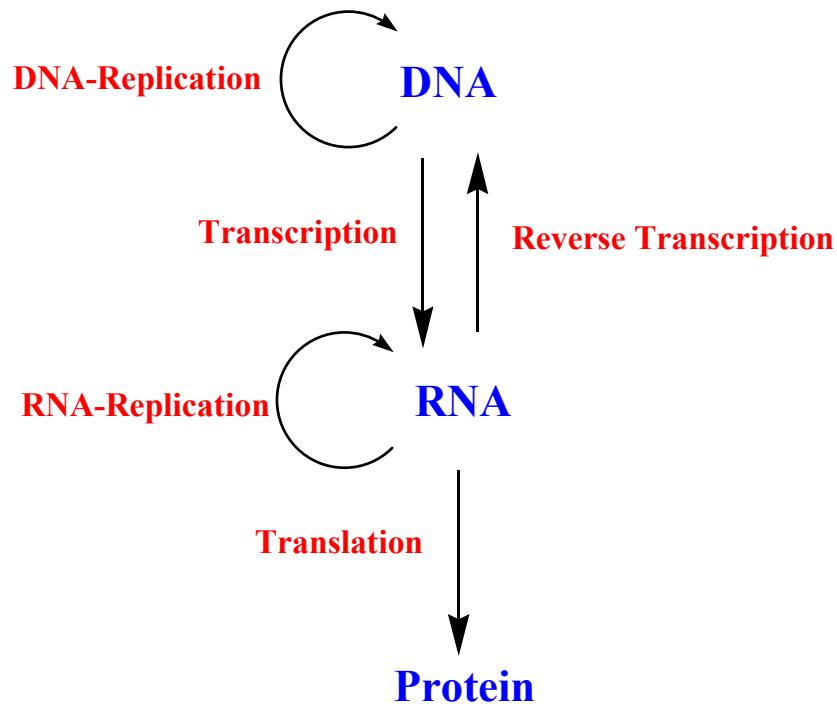
Figure 1.1 The central dogma of molecular biology.

The number of possible strings of a length of N (made out of four possible letters A, G, T, C) is 4^N. Assuming all four bases being distributed equally in a genome, one can calculate the size of a string that constitutes a unique "address". By using the 4^N rule, there exist $\sim 10^6$ possible 10-bp sequences, $\sim 10^9$ 15-bp sequences, and $\sim 4 \times 10^9$ 16-bp sequences. The human haploid genome contains approximately 3.2×10^9 base pairs (Cantor and Smith, 1999), coding for around 30, 000 genes. Interestingly, this only corresponds to about 2 % of the genome base pairs. In theory, a sequence of 16-18 base pairs would only occur once or not at all in the human genome (Cantor and Smith, 1999; Dervan, 1986). Depending on the content of the sequence of interest, however, the size of

site to be recognized in practice could be substantially lower or higher than 16-18 base range.

DNA methylation and chromatin packaging (in eukaryotes) should be considered when designing a ligand for DNA readout. DNA methylation in mammalian genomes is predominantly at the 5-position of cytosine in CG steps. As a result, when targeting the minor groove by small molecules, methylation may not be a factor, but it should affect major groove recognition.

DNA in chromatin is highly compressed through a complex scheme of folding and compaction. First, DNA is wrapped around a disk-shaped protein assembly of eight histone molecules to form the nucleosome core particle. Nucleosomes are later subject to higher order levels of organization. Nucleosomes are the primary substrates for most crucial biological activities in the genome. The structure and dynamics of nucleosomal DNA, along with spatial position of histone octamers, directly determine if the DNA-binding proteins or small organic ligands can interact with specific sequences.

Studies have shown that with concentrations in the micro-molar range (e.g. 5-20 μM), some minor groove binders (e.g. distamycin) tend to shift the DNA positioning along the histone octamer, while mono-intercalators (e.g. actinomycin) bind to DNA sequences with little effect on the histone-DNA complex folding (Portugal and Waring, 1986; Portugal and Waring, 1987; Fitzgerald and Anderson 1999). It has been reported that minor groove binders can access much of the sequences in the nucleosome except at sequences where sites are completely blocked by the interactions between DNA and the histone octamer (Gottesfeld, 2001). Studying how the chromatin packing regulates gene

expression has been a very active research area (Farkas, 2000; Wolffe and Guschin, 2000; Francis, 2004).

Despite the challenges to achieve *in vivo* DNA recognition, cellular DNA remains an attractive drug target due to the modular structure of DNA and the easy handling of DNA synthesis, cloning, and sequencing. In general, the coding regions and regulatory sequences are considered to be the best sites for drug targeting. However, because promoters and operator sequences are shared between genes, it would be difficult to make a specific gene target based on binding to promoter regions alone. On the other hand, it has been reported that the sequences flanking promoters may be appropriate candidates for targeting (Dickinson, 1998; Praseuth, 1999; Lavie, 2004).

1.2 Double-stranded DNA structure

Under physiological condition, the duplex DNA structure is predominately in the classic B form as shown in Figure 1.2. In B-DNA, the right-handed double helix has 10 base pair per complete turn, with the two polynucleotide chains wound anti-parallel to each other and linked by Watson-Crick A-T and G-C pairs. The base pair separation is the same as the helical rise, 3.4 Å. It has a wide, shallow major groove and a narrow, deep minor groove and a C2'-endo sugar pucker pattern. Two other forms, A-DNA and Z-DNA have also been observed in fiber and crystal structures. A-DNA is produced under conditions of low humidity and in solutions when water activity is reduced by the addition of various alcohols. A-DNA, observed in RNA-DNA and RNA-RNA helices, is a more compact structure with 11 base pairs per turn. As shown in Figure 1.2, A-DNA has a helical rise of 2.54 Å and is wider than B-DNA in diameter. A DNA also features a

4

relatively deep, narrow major groove as well as a wide, shallow minor groove. The A-DNA has a C3'-endo sugar conformation. Z-DNA is a left-handed helix with a zig-zag appearance of the sugar-phosphate backbone. It exists in the alternating sequence poly(dCdG).poly(dGdC) and is thought to be the structure formed by this sequence in solution in high salt conditions.

DNA structure is determined by a variety of factors: base sequence, solvent, counter-ion environment, ligand binding, topological strain, and crystal packing forces. The three major polymorphs observed are A, B and Z duplexes and they can be inter-convertible under certain environmental conditions (Ivanov, 2003; Mazur, 2003). This indicates that the three forms should really be thought of as low-energy points along a continuum of structures.
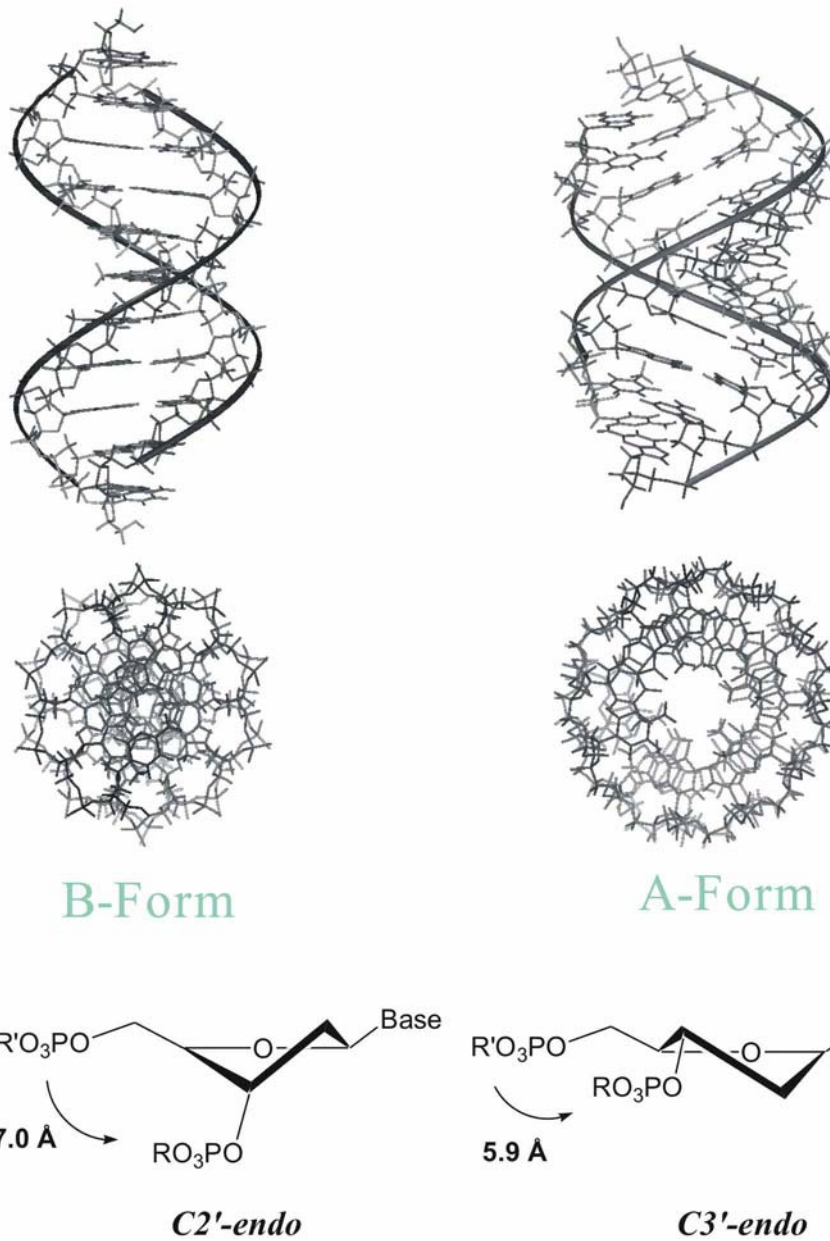
Figure 1.2 Models of various DNA structures: B- and A-form DNA and their deoxyribose sugar conformations (Lee, 2004(b)).

1.3 DNA dynamics

Under normal physiological conditions, DNA is undergoing a wide variety of thermally induced fluctuations, occurring from picosecond to millisecond timescales. These fluctuations play crucial roles in maintaining the normal biological functions and activities of DNA. For example, the fast fluctuations in local helical conformation of DNA (picosecond to nanosecond timescale) enable DNA-binding proteins to probe the base sequence indirectly via local changes in mechanical and dynamic behavior of DNA.

Base pair disruption is more of a dramatic change for duplex DNA structure, in contrast to local helical conformation changes. Functionally, this is a necessary step in making reactive sites on the bases accessible for chemical attack, such as an enzyme-catalyzed DNA methylation reaction (Klimasauskas, 1994) or an enzymatic repair for damaged bases (Cao, 2004). For DNA replication and transcription, base pair disruption has to be the first step in a larger scale disruption of the double helix (Schneider, 2001; Lyakhov, 2001). Because the bases are held within the duplex both by Watson-Crick hydrogen bonding and by base stacking, base opening is a step requiring higher activation energies, typically on the order of 10-20 kcal/mol. As a result, base opening occurs with much longer timescales, on the order of tens of milliseconds (Guéron and Leroy, 1992). Using NMR spectroscopy, base pair opening has been studied indirectly by following the exchange of labile base protons with the surrounding solvent (Englander and Kallenbach, 1984; Leroy, 1988; Folta-Stogniew and Russu, 1994; Guéron and Leroy, 1995). Within B-DNA, A-T and G-C pairs are typically associated with opening times of 1±10 and 5±50 ms, respectively, while open bases have lifetimes of the order of a few nanoseconds

(Guéron and Leroy, 1992). Certain sequences, such as A tracts (A_nT_m, where n ≥3, m ≥ 0), can lead to an order of magnitude slower exchange rates for thymine (Leroy, 1988), while runs of GC pairs appear to accelerate guanine imino exchange rate (Dornberger, 1999).

The conformational changes involved in opening and/or the exact nature of an open state, however, are usually inaccessible to experimental methods due to their higher energies and resulting short lifetimes. It appears that calculations from molecular mechanics (Ramstein and Lavery, 1990; Bernet, 1997; Chen, 1998), molecular dynamics (Keepers, 1984) and Brownian dynamics (Briki, 1991) may provide a realistic picture of dynamic DNA in solution (reviewed in Auffinger and Westhof, 1998; Beveridge and McConnell, 2000; Giudice and Lavery 2002; Priyakumar and MacKerell, 2006(a)). This is mainly thanks to increasing computational power which allows longer simulation times with inclusion of explicit solvent and counterions, improved treatment of long-range electrostatic interactions (Cheatham, 1995) and improvements in force fields and simulation techniques (Cheatham, 1999; Foloppe and MacKerell, 2000; Banavali and MacKerell, 2002; Giudice, 2003; Giudice and Lavery 2003; Priyakumar and MacKerell, 2006(b); Bouvier and Grubmüller, 2007).

It is noteworthy that recently some new experimental techniques based on measuring the time-resolved fluorescence emission of a probe chromophore interacting with extrahelical bases may provide complementary information about the base pair opening in the near future (Neely, 2005; O'Neil and Wiest; 2005).

In order to understand the thermodynamics of DNA-ligand binding with high affinity, a thorough understanding of dynamic behavior of the complex may be necessary.

1.4 The thermodynamics of DNA binding

The energetics of DNA-drug (Chaires, 1998(a); Haq and Ladbury, 2000) and DNA-transcription factor (Patikoglou and Burley, 1997) interactions have been thoroughly reviewed. This section merely serves to cover some of the fundamental aspects encountered in the analysis of DNA-ligand binding thermodynamics.

The binding free energy can reasonably be parsed into five terms:

$$\Delta G_{obs} = \Delta G_{conf} + \Delta G_{t+r} + \Delta G_{hyd} + \Delta G_{pe} + \Delta G_{mol}$$

ΔG_{obs} is the experimentally observed binding free energy, estimated from association constant K by the standard Gibbs relationship, $\Delta G_{obs} = -RT \ln K$.

- ΔG_{conf} is the free energy cost resulting from the loss of conformational degrees of freedom upon binding. Estimates of the cost of immobilizing a single rotatable bond vary between 0.2-1 kcal/mol, but 0.5-0.6 kcal/mol is commonly used for making estimates (Mammen, 1998; Chaires, 1998(a)). Estimates for the free energy of formation of an intercalation cavity in DNA are 4-10 kcal/mol, presumably dependent on the sequence.

- ΔG_{t+r} quantifies the loss of translational/rotational degrees of freedom on bimolecular association. It is commonly estimated at 15 kcal/mol (Finkelstein and Janin, 1989; Spolar and Record, 1994). It depends logarithmically on concentration, molecular mass, and temperature (Mammen, 1998).

- ΔG_{hyd} is the free energy coming from the hydrophobic transfer of ligand from solution into a binding site. There are two main approaches to evaluating ΔG_{hyd} (Spolar and Record, 1994; Chaires, 1998(a)). An experimental approach involves the measurement of heat capacity change (ΔC_p) of binding. ΔC_p is a measure of the temperature dependence of ΔH and ΔS, and a large negative ΔC_p of binding is considered a sign of relatively large hydrophobic effect. ΔG_{hyd} [kcal/mol] has been estimated as 0.080 K x ΔC_p. If a value for ΔC_p is unavailable, a theoretical estimate may be made based on the change in non-polar solvent accessible surface area (SASA). One estimate is $\Delta G_{hyd} \sim -0.022$ cal/mol•A^2 x $\Delta SASA$ (Spolar and Record, 1994).

- ΔG_{pe} is the free energy contribution coming from the release of counterions from the DNA upon ligand binding, commonly referred to as "polyelectrolyte effect" (Manning, 1978; Record, 1978). This energy contribution is dependent on the ionic strength of the solution. The quantity is calculated as

  $$\Delta G_{pe} = \mathbf{Z}\, \varphi\, RT\, \ln[M^+]$$

  where $\mathbf{Z}$ is the charge on the ligand, $[M^+]$ the monovalent cation concentration and φ is the fraction of monovalent cation associated per DNA phosphate ($\varphi = 0.88$ for B-DNA (Record , 1978)). Compared with the free energy contribution from hydrophobic transfer process, ΔG_{pe} is much smaller in magnitude, although a favorable effect.

- ΔG_{mol} refers to the contributions from the non-covalent interactions formed between ligand and DNA, such as hydrogen bond formation, van der Waals

10

interactions, specific electrostatic bonding formation and dipole-dipole interaction.

1.5 DNA recognition

Before we look at the examples of DNA-ligand or DNA-protein complexes, an analysis of structural features of A•T and G•C pairs may help understand the DNA recognition phenomena better. As shown in Figure 1.3, the edge of the bases in duplex DNA is available for recognition by ligands with H-bonding complementarities for direct readout. In the major groove, a C•G pair has a pattern for H-bonding (donor, acceptor, acceptor) which is different from a G•C pair (acceptor, acceptor, donor). For A•T and T•A base pairs, the donor/acceptor patterns are the same, but the presence of the methyl group on T make it possible to effectively discriminate the AT pair from TA pair. In the minor groove, however, the patterns of H-bonding are symmetrical for each base pair, which means a clear discrimination based only on this is impossible.
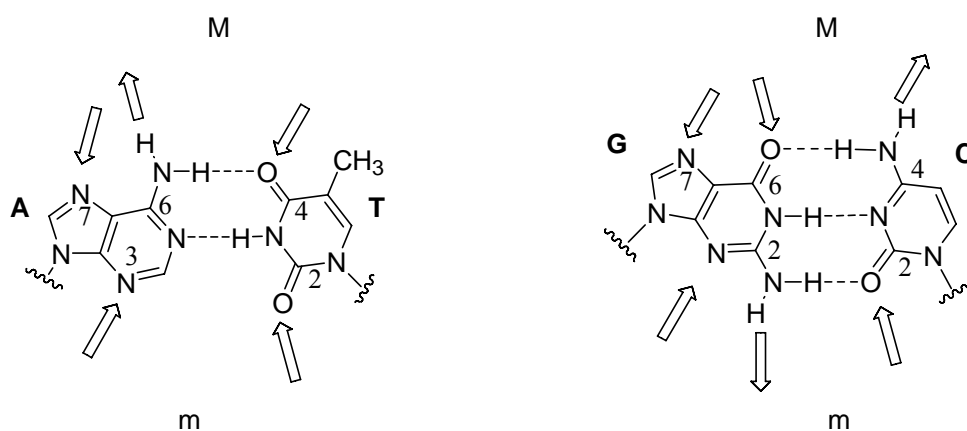
Figure 1.3 The display of H-bond donors and acceptors in A•T and G•C pairs. The arrows indicate the directionality of H-bonds. M and m represent the major and minor grooves respectively.

11

Indirect sequence readout may be enhanced by the hydrogen bonding and/or electrostatic interactions of phosphate groups with basic amino acid side chains. In addition, the "walls" of the minor groove of B-DNA have apolar character. Carbon and hydrogen atoms of sugar residues, such as C1', H1', C4', H4', C5' and H5' form a large part of the van der Waals and solvent-accessible surface of the minor groove "walls" (Neidle, 2002). The minor groove depth in G/C regions is generally shallower due to the presence of the exocyclic amino group at the C2 position. In terms of electronic character, the GC pair is more electron-rich than the AT pair. Thus, electron-deficient planar molecules tend to stack to GC pairs preferentially (Neidle, 2002). The methyl group on thymine could become essential in determining the major groove readout. It can make favorable van der Waals contact with hydrophobic groups or side chains from leucine, isoleucine and alanine alike. On the other hand, the bulky methyl group can sterically clash with the groups from the ligand, forcing the ligand to bind the G/C regions. The exocyclic amino group of G in the minor group can also drive the minor-groove binders to the A/T regions.

1.6 Zinc finger binding motif

Zn finger motif is one of a few commonly observed DNA-protein binding motifs. Due to its potential to be a universal readout of DNA sequences, the Zn finger motif has been studied extensively. There are several different families of proteins containing Zn fingers. A typical example is Zif268, the mouse transcription factor, which contains three zinc fingers. Each zinc finger, made up of ~30 amino acids, consists of two anti-parallel β

sheets followed by one α-helix (Figure 1.4). These are held in place by the zinc ion. It is the helix that is involved in binding to the recognition sequence in the major groove of DNA. The protein recognizes a 9 base DNA sequence, 5'-GCG TGG GCG-3', with a K_D value between ~ 0.01 nM and 10 nM (Elrod-Erickson and Pabo, 1999).

The crystal structure of Zif268 bound to DNA revealed that the amino acids at position −1, 3 and 6 of the α-helix make contacts with bases in the primary DNA strand, while the amino acid at the 2 position interacts with the complementary strand of the DNA (Pavletich and Pabo, 1991). Multiple H-bonding formations, such as the ones between the arginine residues in the protein and the guanine bases in the DNA, and hydrophobic interactions were responsible for the observed binding sequence specificity and high affinity (Figure 1.4).
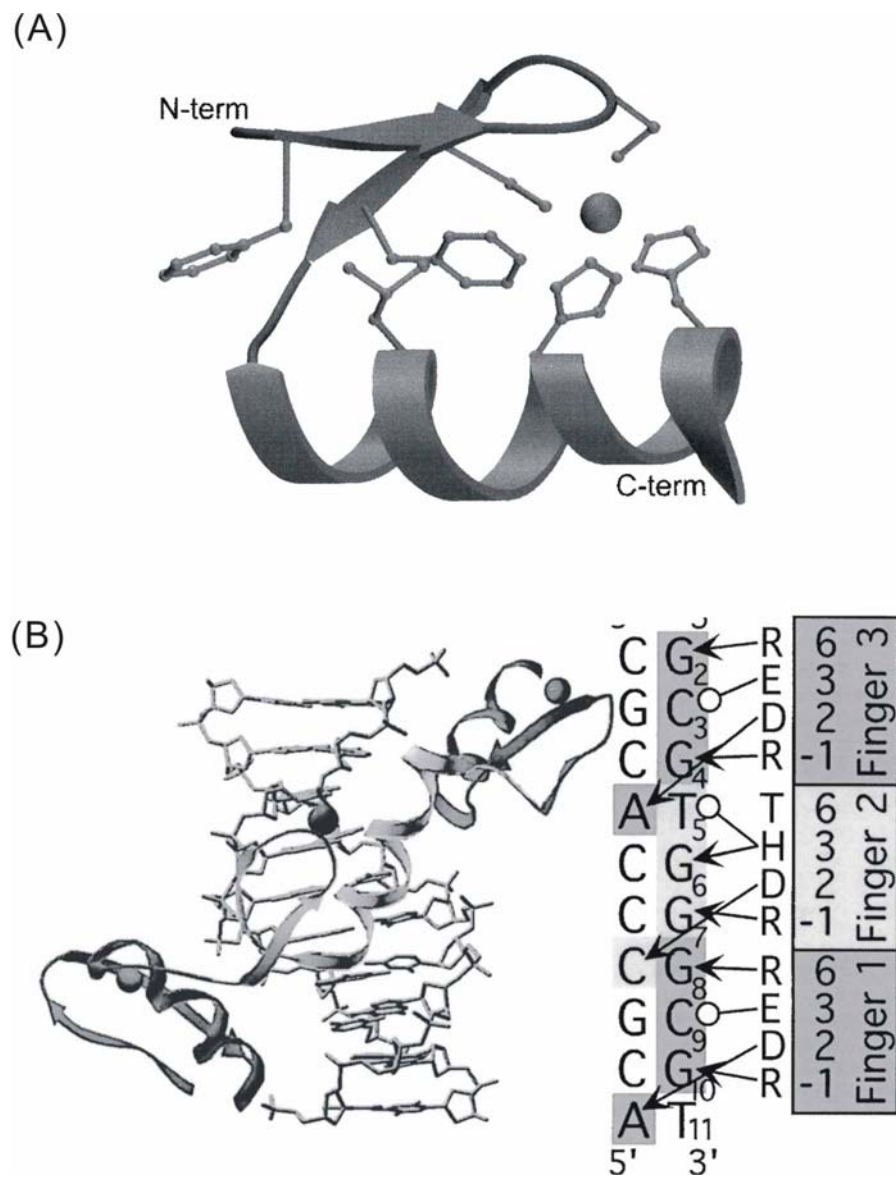
Figure 1.4 (A) The $\beta\beta\alpha$ motif from finger 2 of Zif268. The side chains of cysteines and histidines are involved in zinc coordination (Wolfe, 2000). (B) Structures of the three fingers of Zif268 bound to DNA. Base contacts made from positions −1, 2, 3, and 6 of each α-helix are indicated (Wolfe, 2000). Reprinted, with permission, from the Annual Review of Biophysics and Biomolecular Structure, Volume 29 (c) 2000 by Annual Reviews (www.annualreviews.org).

14

The sequence recognition pattern of Zif268 demonstrated its potential to be engineered to recognize any possible DNA sequences. Tremendous effort has been devoted to attaining this goal, however, identification of a recognition code that would correlate specific amino acid residues with specific bases has not yet been successful. A library enrichment scheme reported recently by Isalan, Klug and Choo (Choo and Isalan, 2000) appears to be promising in making any DNA sequence accessible (Jamieson, 2003)

1.7 Triplex Forming Oligonucleotides

The first reported triplex was a 1:1:1 three stranded polynucleotide complex, poly(U•AU). Other triplexes were later discovered, such as poly(T•AT) and poly(C^+•GC) (Moser and Dervan, 1987). NMR and crystallographic structural studies suggested a model in which the third oligonucleotide strand in the triplex occupies the major groove. Hoogsteen H-bonding to the purine strand was seen parallel to this strand (Figure 1.5). Anti-parallel triplexes have also been developed, such as the G•GC triplex, with the third purine strand being anti-parallel to the purine strand of the duplex. Equilibrium dissociation constants (K_D) were measured on the order of 10^{-7} – 10^{-9} M for 17-34 nt oligos containing G/A and G/T (Vasquez and Wilson, 1998). A single base mismatch could result in the destabilization of the complex at ~ 3.2 - 4 kcal/mol, a value comparable to that for duplex DNA (Roberts and Crothers, 1991).
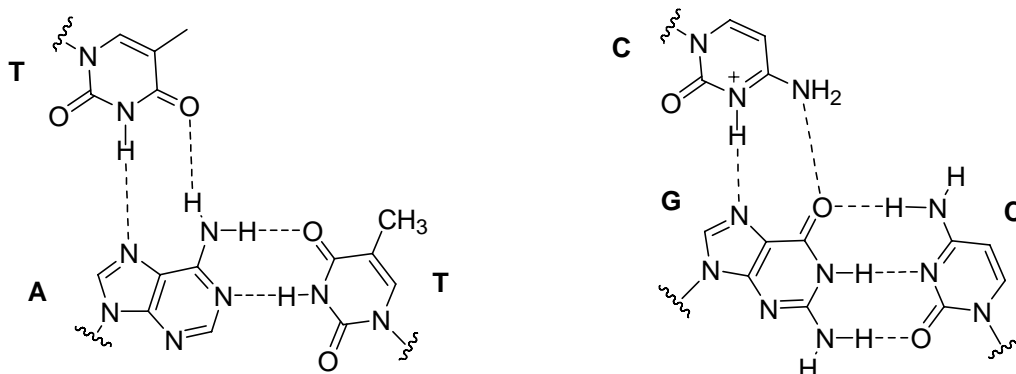
Figure 1.5 The T•AT and C+•GC triplex hydrogen-bonding arrangements.

Targeting of mixed purine/pyrimidine sequences (Fox, 2000) via G•TA, T•CG or other triplets recognition is generally possible, but at much reduced affinity. This is mainly due to two factors. The first one is the limitation of the possible available H-bonds from T or C to the third strand (Figure 1.6). The second one, probably the major factor, is that the insertion of purines into the third strand DNA will often result in unfavorable distortions in its structure, hence hindering its binding to the duplex DNA.
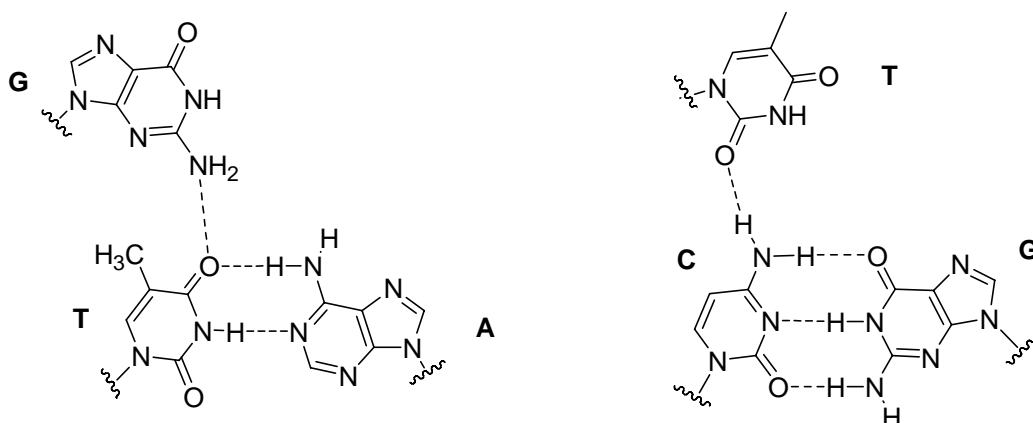
Figure 1.6 The G•TA and T•CG base triplets.

In order to recognize all four Watson-Crick base pairs in a duplex (GC, CG, AT and TA), numerous attempts have been directed to develop non-standard bases. For example, 5-(1-propargylamino)-2'-deoxyuridine, after being incorporated into a third strand of olignucleotide increases the binding affinity by 100 times due to favorable electrostatic interaction with the adjacent phosphate groups (Bijapur, 1999). However, a general strategy has not yet emerged from such studies.

The real interest in triplex phenomenon comes from its potential of being used as an antigene agent to specifically inhibit transcription of a particular gene. A variety of *in vitro* and *in vivo* experiments have been carried out to test this triplex approach. It has been demonstrated that N3'-P phosphoramidite-modified oligonucleotides inhibited transcription elongation of a HIV polypurine tract in cell culture (Faria, 2000).

Overall, the triplex approach is still under development and some of the problems associated with its usage in cell culture, including poor cell permeability, triplex instability, and nuclease degradation, are still waiting to be solved.

1.8 DNA binding small molecules

DNA-interactive small molecules can be categorized into groove binders and intercalators based on their modes of interaction. However, some binding ligands may exhibit both modes of interaction with DNA.

1.8.1 Minor groove binding molecules

Many minor groove binders, including Hoechst 33258, berenil, DAPI, netropsin, and distamycin show biological activities. They can function as simple blockers of

transcription, or as inhibitors of DNA topoisomerase enzymes. Structurally, they are flat, crescent-shaped molecules bearing positive charges (Figure 1.7). Without disturbing the DNA structure, they bind to the minor groove of A/T rich sequences with typical binding affinities on the order of $K_D \sim 10^{-6}$ M. The dominant driving forces for ligand-DNA binding probably come from van der Waals and hydrophobic interactions with the groove walls. Hydrogen-bonding between a ligand and a thymine O2 or an adenine N3 atom on DNA determine specificity. With the entering of a ligand, water molecules located in the minor groove of A/T "spine of hydration" regions are displaced. The non-polar interactions between the ligand and groove walls provide the driving force for this to happen.
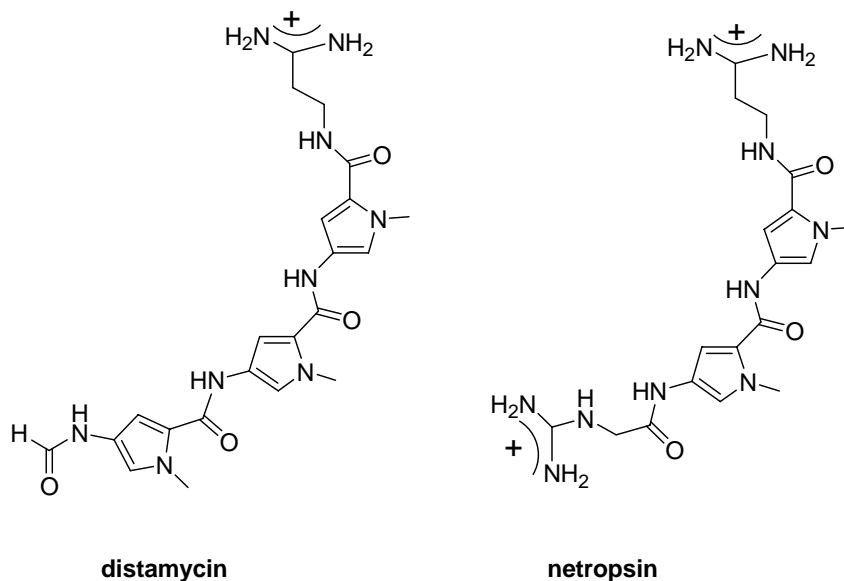
Figure 1.7 The structures of distamycin and netropsin.

Distamycin and netropsin are two naturally occurring antibiotics and are well studied. Both contain multiple *N*-methyl pyrrole and amide units, with netropsin having

18

two cationic charges and distamycin one charge. They bind to the minor groove of A/T –
containing oligonucleotides as do other simple minor groove binders. Each NH group in
the amide unit participates in hydrogen bonding with a thymine O2 and/or an adenine N3
(Kopka, 1986).

A significant discovery, initially from NMR studies, demonstrated that distamycin
can form a 2:1 complex with duplex DNA d(CGCAAATTTGCG)$_2$ (Pelton and Wemmer,
1989). This, along with crystallographic analysis of other 2:1 distamysin/DNA sequences
has greatly influenced the subsequent sequence specific design. These studies confirmed
that in the 2:1 distamycin/d(CGCAAATTTGCG)$_2$ complex, each distamycin molecule
interacts with just one DNA strand in an antiparallel fashion. The narrow minor groove
has to be widened in order to accommodate both distamycin molecules. In the 1:1
distamycin/d(CGCAAATTTGCG)$_2$ complex, each amide group forms a three-center
hydrogen bond with any two adenine N3 and/or thymine O2 atoms.

1.8.2 Sequence-specific synthetic polyamides

Following the discovery of 2:1 distamycin/d(CGCAAATTTGCG)$_2$ complex, a
significant amount of studies lead by the Dervan group have been carried out, aiming at
the goal of sequence recognition at the gene level (Dervan, 2001). First, the *N*-methyl
pyrrole (Py) group was changed to an imidazole (Im) group, so that the recognition
pattern can be switched from an A-T pair to a G-C pair. Then, molecules were
constructed by connecting two antiparallel peptides via γ-aminobutyric acid linkers,
which bind DNA by forming a hairpin in the minor groove. In this new code, an Im-Py
would recognize a G-C base pair but not a C-G pair, while a Py-Im pair binds to a C-G

19

only. Py-Py pairs bind to both A-T and T-A base pairs (White, 1997). Finally, a replacement of pyrrole with a hydroxypyrrole (Hp) made the recognition of an A-T pair over a T-A pair possible by creating a steric clash between the edge of the adenine base and the hydroxyl group on Hp (Figure 1.8). One extra hydrogen bond was also picked up by this new OH group involving the carbonyl O atom on thymine. Thus, a Hp-Py pair would preferentially bind A-T, while Py-Hp would recognize T-A, and a "four base pair code" was accomplished for the first time, although the tails of these molecules have to located at A-T sites (White, 1998). These hairpin polyamides have extremely high selectivity and site affinities, with typical binding constants (K_D) in the 10^{-9} M range.
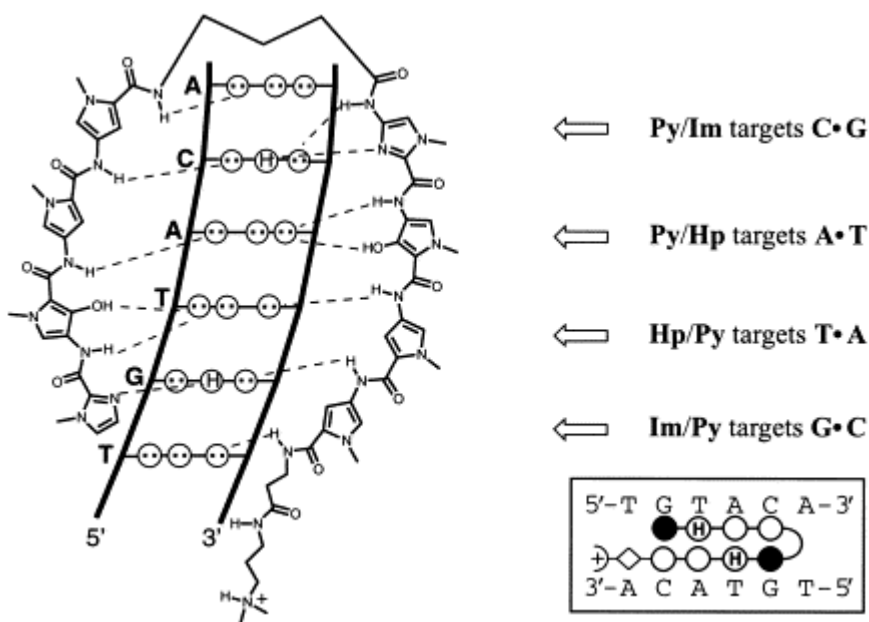
Figure 1.8 Binding model for the complex formed between ImHpPyPy-γ-ImHpPyPy-β-Dp and a 5′-TGTACA-3′ sequence (Dervan, 2001).

Coverage of extended sequences by polyamides was initially achieved by linking of additional Py/Im units. Further addition of Py or Im units resulted in reduced affinities, however, due to mismatch in the curvatures of the DNA minor groove and the ligands. A flexible β-alanine residue ("β") was then introduced to accommodate the DNA curvature, leading to an increase in affinities. A β-β pair binds preferentially to A-T/T-A pairs (~ 20-fold over G/C pairs). As shown in Figure 1.9, a dimer composed of pyrrole, imidazole and β-alanine recognition units can recognize a 16 bp sequence 5'-ATAAGCAGCTGCTTTT-3' with a dissociation constant of $K_D \sim 30$ pM. The dissociation constants for molecules with single base mismatches were 10-35-fold less (Trauger, 1998).
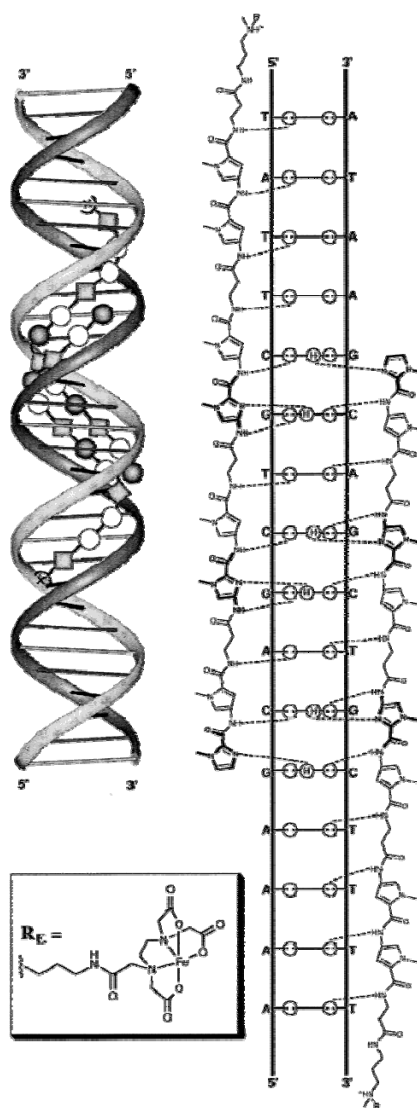
Figure 1.9 A polyamide dimer that recognizes 16 bp 5'-ATAAGCAGCTGCTTTT-3' in the minor groove (Trauger, 1998). Reprinted with permission, from Journal of the American Chemical Society, Copyright 1998, American Chemical Society.

Biological activity tests have shown that polyamides can effectively compete with regulatory proteins at transcription sites (Gottesfeld, 1997; Dickinson, 1998; Mapp, 2000; Bremer, 2001). An impressive example was the repression of the 5S RNA gene by a polyamide that binds to a d(AGTACT)$_2$ sequence within the TFIIIA transcription factor

binding domain. Because the polyamide binds to this d(AGTACT)$_2$ sequence much tighter than TFIIIA, successful inhibition of transcription has been achieved both *in vitro* and *in vivo*.

Although the polyamide approach by targeting programmable polyamides to the desired sequences in minor groove has been very successful so far, there are still some obstacles left to be conquered. For example, the polyamides do not target all the long sequences (larger than 9 bp) equally well as the short ones. Apparently, the local structures of DNA are sequence dependent. Also the cell permeability for polyamides is not as good as other minor-groove drugs such as Hoechst 33258.

1.8.3 Intercalative binding agents

The main interest in intercalating drugs lies in their potential for anti-tumor activity. For example, the anthracycline antibiotic adriamycin and its synthetic mimetic mitoxantrone are important, anti-cancer drugs in widespread clinical use. It is generally accepted that these drugs act as cellular topoisomerase I and II poisons (Foye, 1995; Pommier, 1998; Burden and Osheroff, 1998). The DNA binding properties of intercalating agents usually correlate well with their biological activities.

The intercalating unit, sometimes called the chromophore, is an extended planar heteroaromatic ring system. Electron-deficiency is important, and being positively charged is common. Intercalation leads to the extension of duplex structure by ~3.4 Å per bound drug molecule and the unwinding of duplex in the proximity of the intercalation site. This greatly decreases the chance for another chromophore to intercalate within the nearby base pairs (the neighbor-exclusion principle; Crothers, 1968).

Simple intercalators, such as proflavine and ethidium bromide have a small preference for a pyrimidine-3',5'-purine sequence step (Figure 1.10).
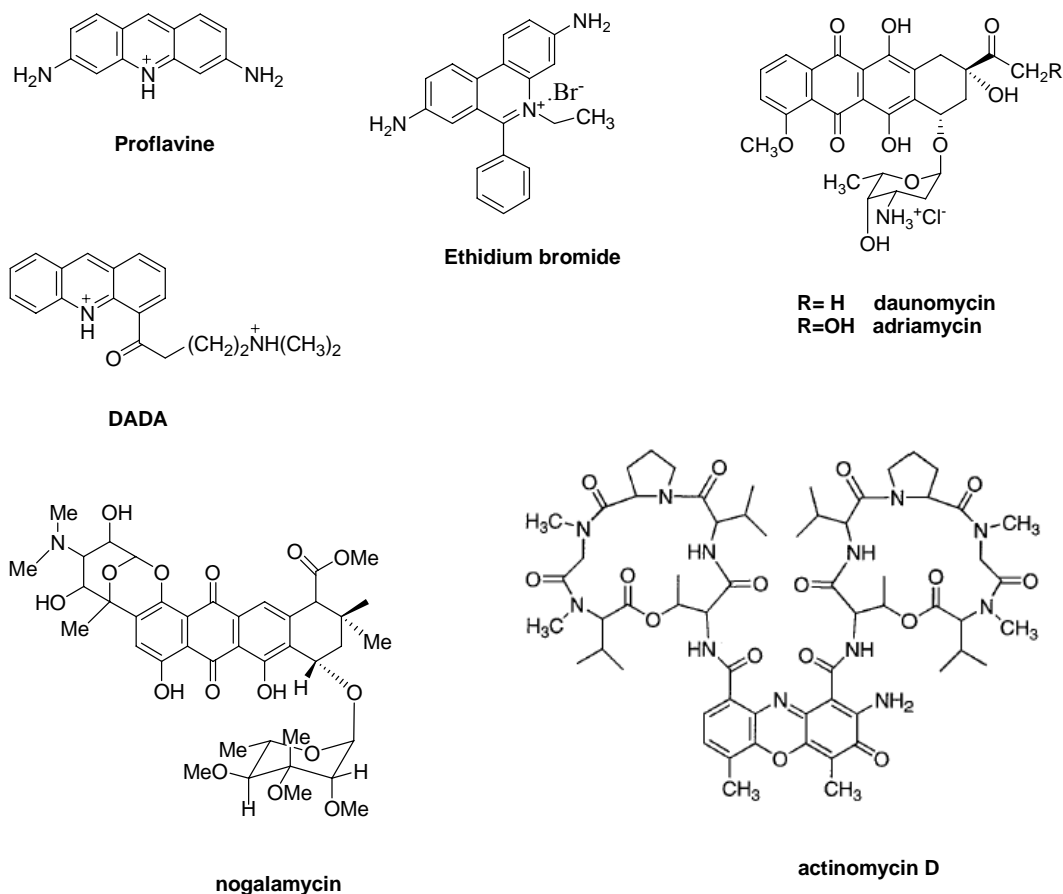
Figure 1.10 Chemical structures of some typical intercalators.

Complex intercalators including actinomycin D and daunomycin have side chains, such as peptide units attached (Figure 1.10). These attached groups reside in a groove, stabilized by van der Waals interactions and/or hydrogen bonding to the nearby bases, thus giving extra sequence specificity to the intercalating chromophore. For example, the threonine residue in the side chain of actinomycin forms hydrogen-bonds to the N2 and

N3 atoms of guanine, giving a binding site preference for the sequence of GpX (X : any one of the bases). Structural studies on a d(CGTACG)$_2$-nogalamycin complex revealed an interesting result (Egli, 1991; Williams and Searle, 1999), in that the aglycone side group is located in the minor groove and the amino sugar group in the major groove. Hydrogen-bonding to the N2 and O6 atoms of guanines was also seen. Importantly, for the intercalating chromophore to get into the intercalating site, the two bulky side groups have to pass through the intercalation site. Relatively slow dissociation and association processes would be expected for this system. Results from a stopped-flow kinetic study confirmed this claim. A very small dissociation rate constant (k_d = 0.001 s^{-1}) was determined for nogalamycin in the complexation with calf thymus DNA, which was in sharp contrast with classic intercalators, such as ethidium bromide which has a rate constant k_d = 13.4 s^{-1} (Fox, 1985; Wilson, 1985). The small dissociation rate constant of nogalamycin has been linked to its biological activity (Denny and Wakelin, 1986). A few intercalators have been found to have their side groups interacting with the major groove, not the minor groove like the majority of intercalators. For example, the dimethylaminoethyl side group of DADA, a 9-amino derivative of the acridine carboxamide parent structure, was found located in the major groove of duplex d(CGTACG)$_2$ due to hydrogen-bond formation between the protonated N atom and the edge atoms of a guanine base in the major groove.

1.8.4 Bisintercalators

In order to improve the binding affinity of single intercalators, as well as allow the recognition of extended sequences of DNA, two (or more) intercalating units can be covalently connected through one or more linkers. In theory, even with no favorable interactions between the DNA and the linker(s), bisintercalation should result in an increase in binding affinity. In the best scenario, the binding constant of a bisintercalator (K_{bis}) should be larger than (rather than equal to) the product of binding constants from each of the corresponding monomers (K_{mon}), that is $K_{bis} > K_{mon} \cdot K_{mon}$. This additional energy could theoretically derive from the entropy gain achieved by combining two molecules into one (Jencks, 1981).

In reality, this is never the case. For example, the binding affinities of bisintercalators, built from a series of 9-aminoacridine units and connected by a polymethylene chain were found to be only 10-15 times higher than that of the 9-aminoacridine monomer ($K_D \sim 10^{-4}$ M) (Figure 1.11, reviewed in Wakelin, 1986). Generally, the binding affinity of a bisintercalator is closely related to the linker length and structure, and how much the overall DNA structure has been changed.

Ditercalinium is a bisintercalator made of two 7H-pyrido[4,3-c]carbazole chromophores and a rigid diethylbipiperidine chain as the linker. It binds DNA with K_D >10^{-7} M but without much specificity (Figure 1.11, reviewed in Gago, 1998). Crystal structural studies of a ditercalinium-d(CGCG)$_2$ complex showed that bisintercalation occurs via the major groove, spanning two base pairs, with electrostatic contacts between the charged nitrogen atoms of the drug and guanine residues in the floor of the groove.
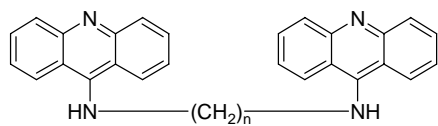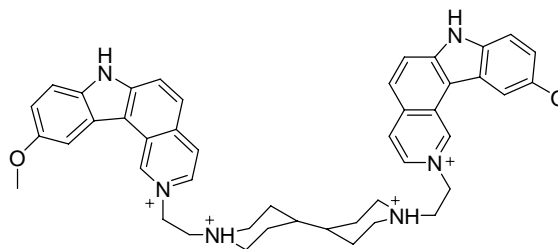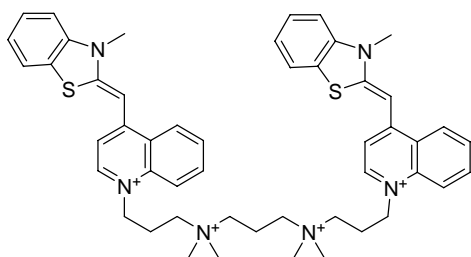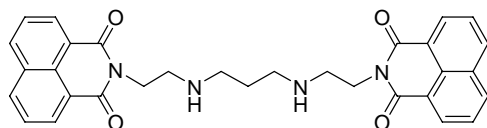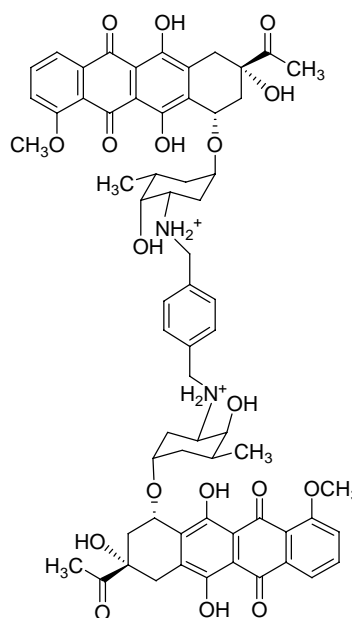
Figure 1.11 The chemical structures of some of the bisintercalators.

Two examples of sequence specific bisintercalators are TOTO and LU79553 (Figure 1.11). The NMR structure of TOTO-d(CGC|TA|GCG)$_2$ revealed that the polyamine linker resides in the minor groove. Sequence specificity was ascribed to thiazole orange preferential intercalation in the PyPy (PuPu) steps. A sequence preference

for CpT over CpC steps was due to favorable non-polar interactions between the thiazole ring, which protrudes toward the major groove, and the methyl groups of thymine (Spielmann, 1998). LU 79553 binds to TpG and GpT steps, with a $K_D \sim 5 \times 10^{-7}$ M (Bailly 1996). NMR structural studies have the drug bisintercalating in the oligonucleotide d(AT|GC|AT), with the linker residing in the major groove and spanning two base pairs. It was suggested that a hydrogen bond was formed between the linker amino group and O6 of the guanine residues (Gallego and Reid, 1999). A $K_D \sim 10^{-4}$ M has been estimated for monoimide; analogous to the monoimide unit of LU 79553A described above, monoimide has (A/T)G binding specificity. The electrostatic interactions between the carbonyl oxygens of the imide and the N2-amino group of guanine in the minor groove were thought to be the origin of specificity (Liu, 1996). Interestingly, in the NMR structure of the LU79553-d(AT│GC│AT)$_2$, a dynamic ring flipping of the intercalated rings was observed and was ascribed to the electrostatic repulsion in the major groove, thought to be present between the imide oxygens and the N7 / O6 electron-rich edge of guanine (Gallego and Reid, 1999).

A novel series of synthetic bisintercalating analogues of daunomycin have been developed by a structure-based drug design approach. Starting from the hexanucleotide-daunomycin crystal structures (with two drug molecules bound to d(C|GATC|G)$_2$), the two minor-groove located sugar groups were connected by a *p*-xylene linker group through the exocyclic amino groups on each sugar ring, thus forming the bisintercalator WP631 (Wang, 1987; Hu, 1997). WP631 exhibits the binding affinity and binding mode predicted for such a bisintercalator, although its anti-tumor activity is not vastly superior

28

to that of a single daunomycin molecule. The predominant binding mode of WP631 to heterogeneous DNA was found to be through bisintercalation, with a $K_D \sim 3 \times 10^{-12}$ M (0.2 M salt) (K_D for daunomycin is $\sim 10^{-7}$ M). The linker of WP631 resides in the minor groove as expected (Leng, 1998). WP631 was successfully predicted to bind G|C(T/A)(T/A)G|C and C|G(T/A)(T/A)C|G sequences, but not sequences containing G/C pairs in the middle due to steric clash with the exocyclic amino groups of guanine in the minor groove. It is noteworthy that the successful design of WP631 may indicate that optimal bisintercalation requires four base pairs between intercalated units.

1.9 Threading intercalators

Nogalamycin is a naturally occurring threading intercalator (as described above). Upon binding to DNA, the anthraquinone chromophore intercalates between two DNA base pairs, with one of the two sugar substituents interacting with the major groove and the other with the minor groove. A variety of synthetic threading intercalators have also been developed, including 1,5-disubstituted anthraquinones, 9,10-disubstituted anthracenes and naphthalene diimide (NDI) derivatives (Figure 1.12). The threading intercalation binding mode may have several advantages pertaining to biological activity, such as simultaneous blockade of both DNA grooves which may help effectively regulate gene expression, improved linker-DNA interactions, as well as slower dissociation rates from cellular DNA (Lamberson, 1991).
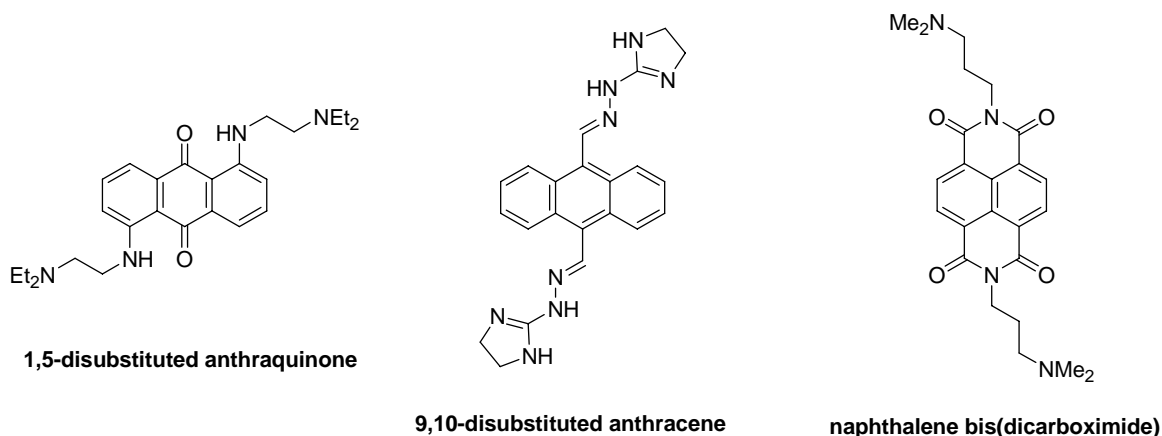
Figure 1.12 Structures of some synthetic threading intercalators.

NDI derivatives have become major players in this field due to synthetically easy derivatization of the parent naphthalene diimide structure. Wilson *et al.* investigated kinetic behaviors of symmetric NDI monomers with alkyl amino substituents attached to both sides (Tanious, 1991; Yen, 1982). Stopped-flow kinetic experiment results revealed that both association and dissociation rates of the diimide are significantly slower than those of classical intercalators with similar binding constants, a result that was ascribed to a threading intercalation mode of binding. For the dissociation process, it was proposed that the rate determining step involves only one ion pair to be broken; then the free side chain could slide between base pairs to have both side chains in the same groove, followed by fast dissociation of the NDI unit.

For the past ten years or so, the Iverson group has been involved in systematically developing various NDI-based bisintercalators and polyintercalators to target duplex DNA structures. At the early stage, Lokey et al. synthesized a novel class of molecules, called "serpentercalators", in which various numbers of NDI intercalating units were

connected together by flexible peptide linkers in a head-to-tail fashion on solid phase support (Figure 1.13; Lokey, 1997). These molecules showed full intercalation on double-stranded DNA based on viscometric titration and the unwinding studies. Kinetic studies using UV/Vis revealed that these molecules have a dramatic preference for poly(dGdC) over poly(dAdT). DNAse I footprinting results were also consistent with this discovery. The number of NDI intercalating units has been extended up to eight (Murr, 2001). The dissociation constants (K_D) for mono-, bis-, tetra-, and octa-intercalators were estimated from footprinting data (based on poly(dGdC)) to be 5×10^{-6}, 2.5×10^{-7}, 1.25×10^{-7}, and 1.56×10^{-8} M, respectively.

Later, a combinatorial approach was applied to examine how the peptide linker compositions would modulate the binding properties of NDI-based bisintercalators. It turned out that change of the linker peptide composition resulted in altered sequence specificity, based on DNAse I footprinting on a 231 bp restriction fragment of pBR 322 plasmid (Guelev, 2000). Two bisintercalators were identified that recognized distinct sequences. Combined with NMR structural studies, it was determined that compound **G$_3$K**, with a tris-glycine-lysine linker, exhibited a dissociation constant in the sub-micromolar range and good sequence specificity. Compound **G$_3$K** prefers d(CG│GTAC│CG)$_2$ site at 1:1 ratio (Guelev, 2001(a); Figure 1.14(A)). The NDI units are intercalated between the GpG steps with the connecting peptide linker spanning four base pairs in major groove (Figure 1.14 (A)). It was interesting that the linker resides in the major groove. Compound **(β-A)$_3$K**, having a tris-β-alanine-lysine linker, was shown to bind tightly to the sequence d(CGATAAGC)·d(GCTTATCG). The derived NMR

model has NDI units intercalating at the ApG and GpA steps, with the linker spanning four base pairs in the minor groove (Guelev, 2002; Figure 1.14 (B)).

Specific recognition is thought to be achieved largely through favorable solvophobic interactions and steric complementarity between ligand and DNA. In the major groove of d(CGGTACCG)$_2$, the shorter Gly-Gly-Gly-Lys linker spans diagonally across a channel formed by the two opposing thymine methyl groups of the TA·TA step, covering the shortest path along the DNA major groove. The longer, more hydrophobic β-Ala-β-Ala-β-Ala-Lys linker takes a longer path along the hydrophobic minor groove of the A/T rich sequence. In addition, possible hydrogen bonds were proposed to explain the observed binding specificities for **G$_3$K** and **(β-A)$_3$K**, such as the one between the amide proton of the linker of **G$_3$K** and O6 of guanine in the major groove of d(CGGTACCG)$_2$, and the one between the linker amide proton of **(β-A)$_3$K** and N3 of an adenine in the minor groove of d(CGATAAGC)$_2$. Structural modeling studies indicated that in both complexes the lysine side chains make non-specific contacts with the DNA and have indirect effect on specificity, via peptide conformation. It appears that among different factors which contribute to the observed sequence recognition, the length of linkers plays a key role in directing the groove selection for bisintercalators.

Structural studies on these NDI bisintercalators verified that both the major and the minor grooves of DNA are accessible through the threading polyintercalation strategy. Next, a polyintercalator was designed to recognize both grooves in an alternating fashion with programmable sequence specificity. The resulting symmetric tetramer (**5.1**), comprising two identical bisintercalator units ((**β-A)$_3$K**) and a flexible

aliphatic chain linker was synthesized (Lee, 2004(a); Figure 1.15). NMR structural studies revealed that this compound binds to a 14 bp DNA sequence d(GATAAGTACTTATC)$_2$ with 1:1 stoichiometry. A threading polyintercalation binding pattern was confirmed, in which the tris-β-alanine-lysine linker resides in the minor groove as predicted, while an adipic acid linker interacts with the wall of major groove (Figure 1.16). This is the first structural analysis of any DNA intercalating molecules larger than a bisintercalator. It verified that the sequence-specific binding of longer threading polyintercalators can be predicted from bisintercalator modules. However, it has to be pointed out that, upon binding to the tetramer, the DNA duplex structure is distorted. The linker structures, particularly the one within the major groove, clearly needed further optimization since it was significantly disordered in the NMR constrained dynamics calculation.

Figure 1.13 (A) Polyintercalators containing 1,4,5,8-naphthalenecarboxylic diimide chromophore show preference for GC rich region in DNase I Footprinting (Lokey, 1997). (B) The combinatorial library work identified different binding sites for two different molecules, **G₃K** and **(β-A)₃K** (Guelev, 2000).
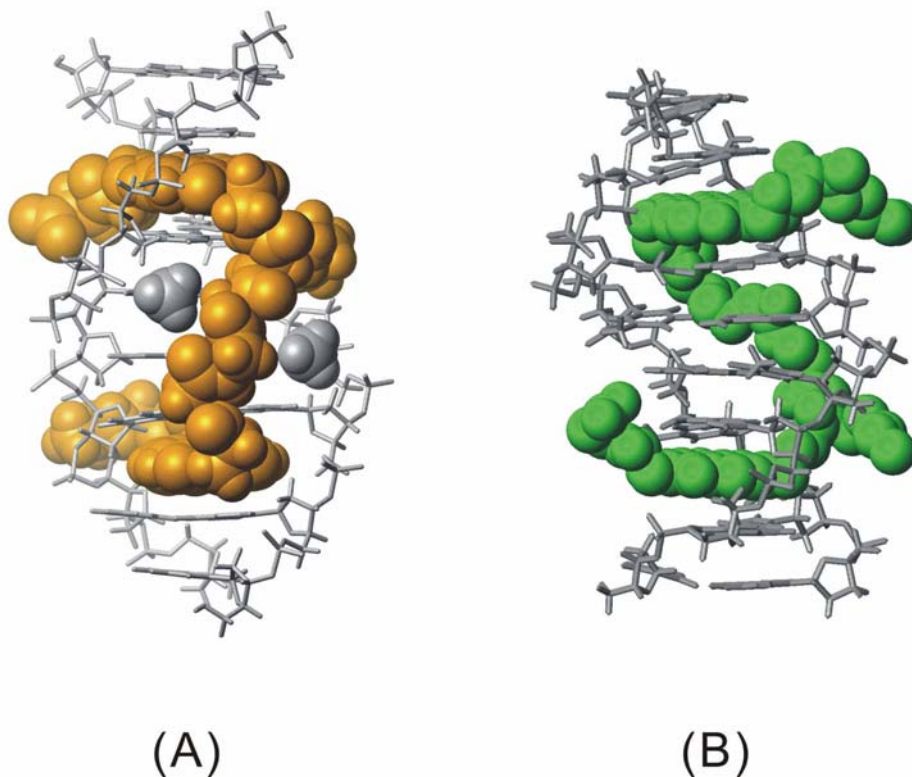
Figure 1.14 Spacefill models of the ligand-DNA complex based on the NMR data. (A)* compound **G₃K**-d(CGGTACCG)₂ complex (Guelev, 2001(a))　(B)** compound **(β-A)₃K**-d(CGATAAGC)·d(GCTTATCG) complex (Guelev, 2002).

*Reprinted from Chemistry & Biology, 8, Peptide bisintercalator binds DNA via threading mode with sequence specific contacts in the major groove, Guelev, V., *et al.*, 415-425, Copyright 2001, with permission from Elsevier.

** Reprinted with permission, from Journal of the American Chemical Society, Copyright 2002, American Chemical Society.
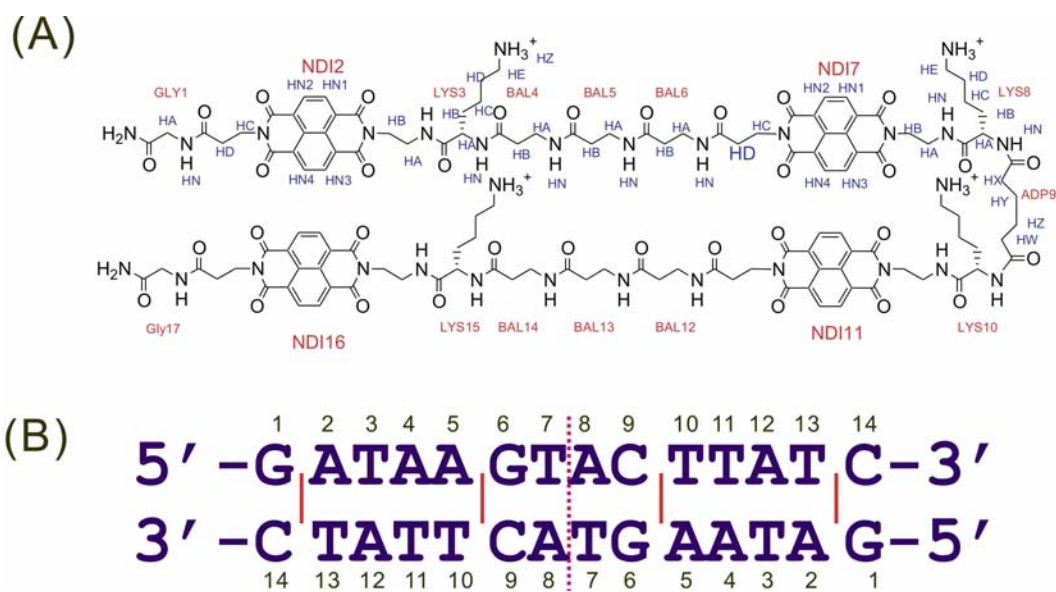
Figure 1.15 (A) Structure and naming convention of compound **5.1**; (B) the DNA sequence used in the NMR study (Lee, 2004(a)).
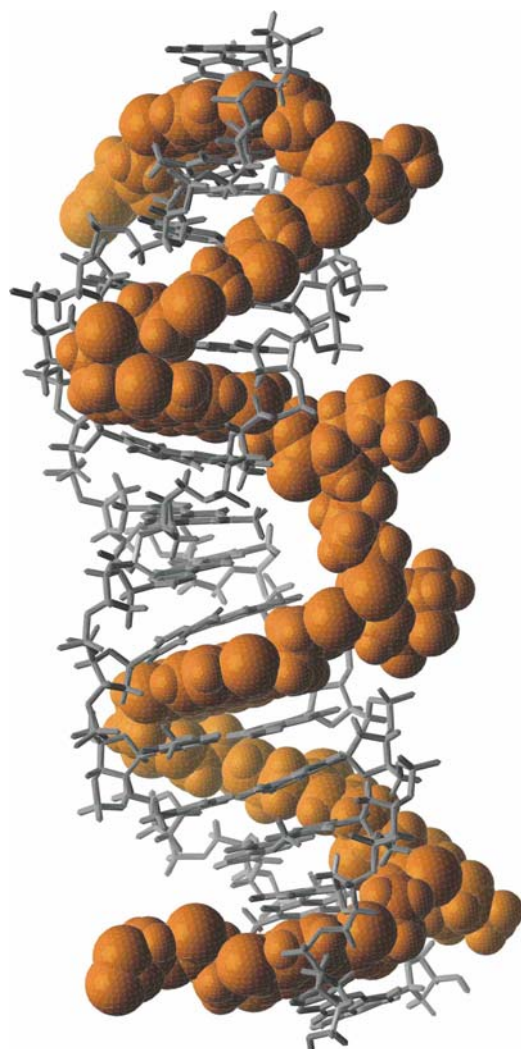
Figure 1.16 Space filling model of the tetramer-DNA complex that represents unique threading polyintercalation mode of binding, with linkers alternating in the order of minor groove, major groove, minor groove (Lee, 2004(a)). Reprinted with permission, from Journal of the American Chemical Society, Copyright 2004, American Chemical Society.

Chapter 2   Synthesis and DNA Binding Studies of Bisintercalators with a Novel Spiro-cyclic Linker

2.1 Chapter summary

2.1.1 Goals

To synthesize second generation, rigid linkers intended to enhance affinity and specificity in the context of threading polyintercalators.

2.1.2 Approach

Two spiro-cyclic rigid linkers were designed to fit within the groove of double-stranded DNA and serve as a scaffold for sequence recognition. The linkers were synthesized via different highly stereo-selective synthetic routes. The corresponding bisintercalators were made by solution phase synthetic strategy. Their binding properties were analyzed using UV-vis dissociation studies of poly(dAdT), poly(dGdC) and calf thymus (CT) DNA, followed by DNAse I footprinting on a 232 base pair fragment plasmid and a synthetic, rationally designed 92mer DNA.

2.1.3 Results

The *cis* and *trans* oriented spiro-cyclic linkers were synthesized and used to create corresponding bisintercalators. Analysis of dissociation kinetics indicated that the *cis* oriented bisintercalator has significantly slower dissociation from poly(dGdC) and calf thymus (CT) DNA, but reproducibly faster dissociation from poly(dAdT), compared to

the *trans* oriented dimer and a flexible dimer control. DNAse I footprinting on 232 DNA plasmid indicated that the *cis* oriented bisintercalator prefers to bind to GGNNCC sequences. Further experiments on a synthetic 92mer DNA showed that this *cis* dimer binds to GGTACC with highest affinity, followed by GGGCCC, GGATCC, and GGCGCC. These data thus indicate that the *cis*-oriented linker confers both greater affinity and specificity, consistent with the original design goals.

2.2 Introduction

Bisintercalation represents an effective strategy for targeting specific DNA sequences for purposes of cancer therapy and antibiotic development (Chaires, 1998(b)). This approach was first investigated in the 1970's following the discovery and elucidation of the biological activities of echinomycin, a natural bisintercalator consisting of two quinoxaline moieties linked by a bicyclic peptide (Waring and Wakelin, 1974). Over the past 30 years, a variety of DNA-binding bisintercalating molecules have been developed. These bisintercalators generally consist of two aromatic units connected by a linker chosen to modify binding affinity or specificity (Le Pecq, 1975). Some of the bisintercalators have exhibited promising antitumor activities, such as an acridine-based agent DACA (Bridewell, 2001), bisanthracycline derivative WP631 (Chaires, 1997; Leng, 1998; Hu, 1997), bisimidazoacridone compound WMC26 (Cholody, 1995), and bisnaththalimide related compounds such as LU79553 (Bousquet, 1995).

Threading polyintercalators have attracted attention recently due to their ability to occupy and interact strongly with both the minor and major grooves of DNA simultaneously. As a result, the threading polyintercalator design promises high DNA

binding affinity and specificity, a slow rate of dissociation, and an enhanced ability to block DNA-protein interactions (Takenaka and Takagi, 1999). The 1,4,5,8-naphthalenetetracarboxlic diimide (NDI) unit is a well-studied threading intercalator, in which functional groups attached to the two imide nitrogen atoms will reside in the different grooves of DNA (Yen, 1982; Tanious, 1991).

Using a combinatorial approach, two sequence-specific threading bisintercalators, **G₃K** and **(β-A)₃K**, linked by natural or non-natural peptides were identified, with the first one binding to the major groove of d(CGGTACCG)₂ (Guelev, 2001(a)), and the second one binding to the minor groove of d(CGATAAGC)·(GCTTATCG) (Guelev, 2002). Modeling has indicated that linker length, electrostatic complementarity and sterics are all important for determining groove location and sequence specificity. The NDI ring itself exhibits a preference for stacking in an NpG step where N is any one of the four bases. This specificity is thought to be due in part to the favorable electrostatic interactions between the C=O of NDI and exocyclic NH_2 group of G in the minor groove (Liu, 1996). Building on this strategy, a modular symmetric tetramer was synthesized, composed of two minor groove binding bisintercalators connected by an adipic acid linker intended to bind in the major groove. NMR structural studies verified that this tetramer binds to a specific 14 base pair DNA sequence in a threading polyintercalation mode, with linkers alternating in the expected minor groove, major groove, minor groove pattern (Lee, 2004(a)). However, in this same analysis, the major groove binding linker exhibited significant disorder, leading to the conclusion that a properly designed, rigidified linker could greatly improve binding specificity and affinity of the

bisintercalators. Ideally, rigidifying the linker will decrease the entropy loss upon binding that occurs with the flexible linkers, and in addition prevent unwanted modes of interaction with alternative DNA sequences. Chaires and coworkers have achieved remarkable increases in both the affinity and specificity of binding by choosing appropriate linkers to connect two anthracycline units (Chaires, 1997; Leng, 1998; Hu, 1997).

This chapter describes the design and synthesis of rigidified linkers intended to be a scaffold for future linker designs. DNA binding studies are reported in the context of threading bisintercalation.

2.3 Results and discussion

2.3.1 Linker design

The extreme flexibility of our first generation linkers likely limits overall specificity as well as possible programmable recognition of different sequences of DNA. Specific recognition based on shape and/or hydrogen bonding complementarity may be compromised by the flexibility inherent in the unconstrained peptide chains. A conformationally restricted linker scaffold, of the proper size and geometry, should be able to improve not only specificity of DNA recognition, but overall binding affinity as well (Chaires, 1997). The important caveat here is that a rigidified scaffold should be readily synthesizeable and have the potential to be derivatized with various function groups capable of interacting with specific sites on the DNA bases.

Shown in Figure 2.1 are two novel rigid scaffolds. The two heterocycles are intended to 1) lock the central cyclohexane ring into a chair conformation and 2) form hydrogen bonds with bases on the floor of a DNA groove. Next generation designs would involve derivatization of the central six-membered ring to achieve alternative specificities. Preliminary computer modeling indicates these rigid scaffolds could be compatible with binding in either the minor or major grooves of six DNA base pair sequences, increasing the potential applications of the design (Figure 2.2). The same modeling indicated that the *cis* isomer should be more compatible with the curvature of DNA compared to the *trans* derivative. Scheme 2.1 shows the retrosynthetic analysis of **C1** and **D1**. In order to make these two intercalators, two structurally different bis-amino acids are needed.
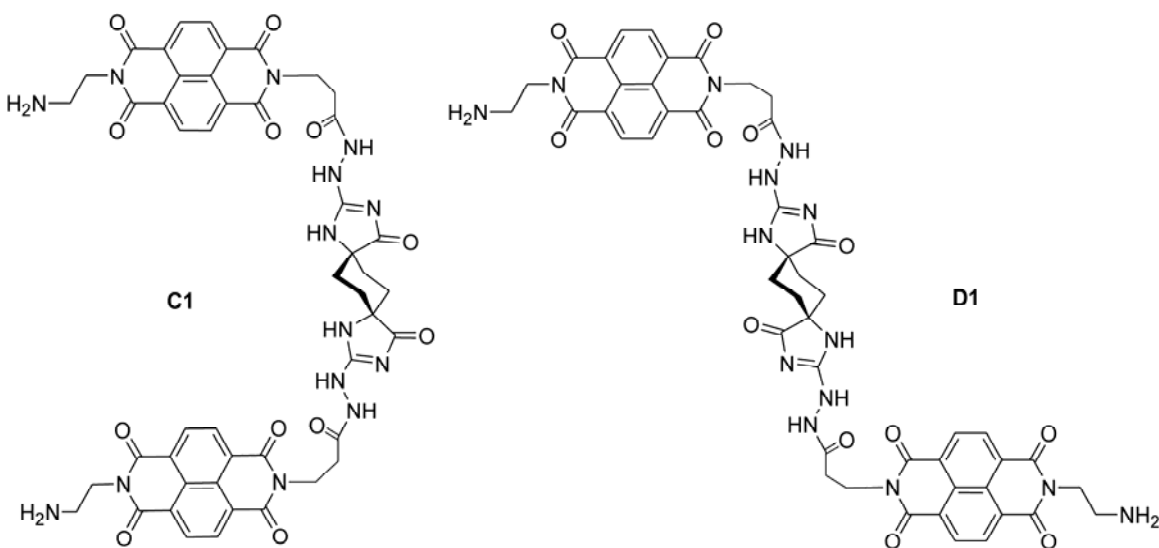
Figure 2.1 The structures of two dimers, **C1** and **D1**, containing trans and cis spiro-cyclic linkers, respectively.

(a)   (b)

(c)   (d)
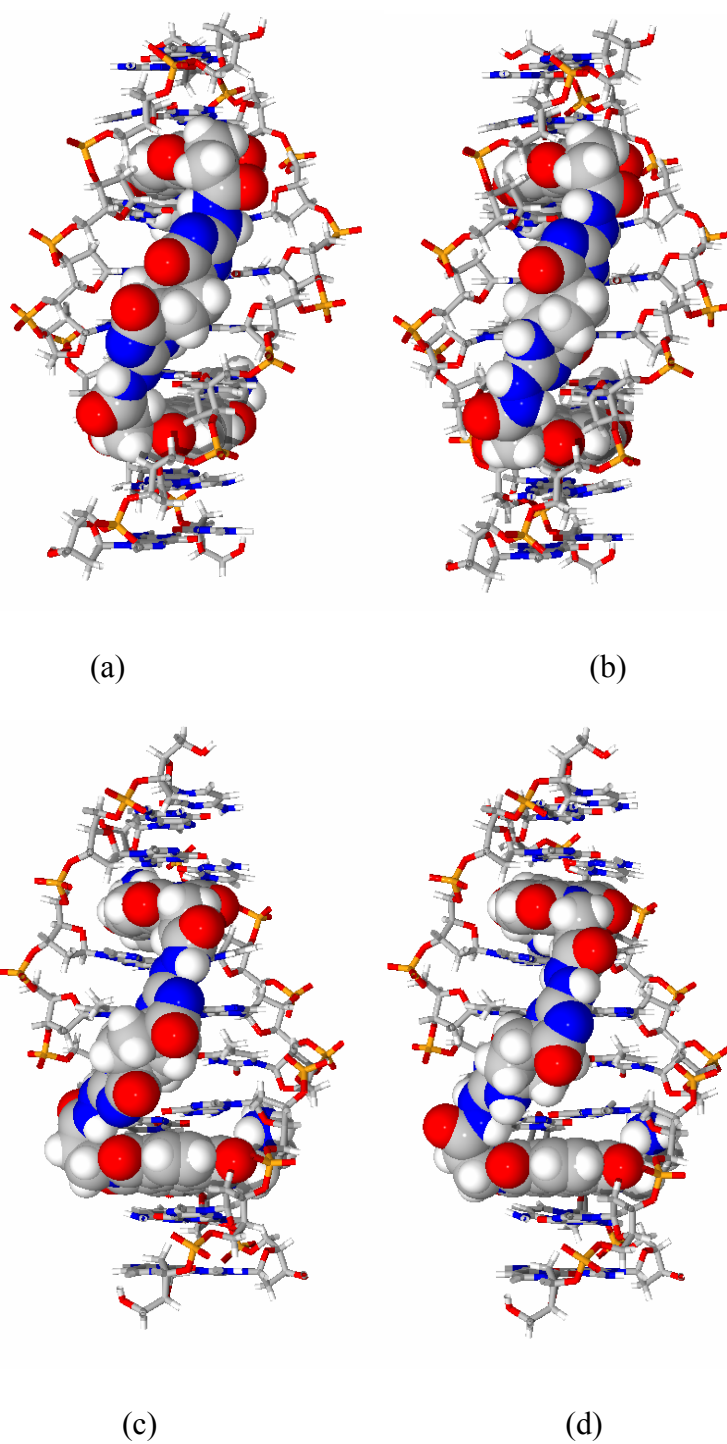
Figure 2.2 View from the minor groove (a and b) and major groove (c and d) of the **C1** (left) and **D1** (right) bisintercalator/d(CGGTACCG)$_2$ complexes.

Scheme 2.1 Retrosynthetic analysis of bisintercalators **C1** and **D1**.
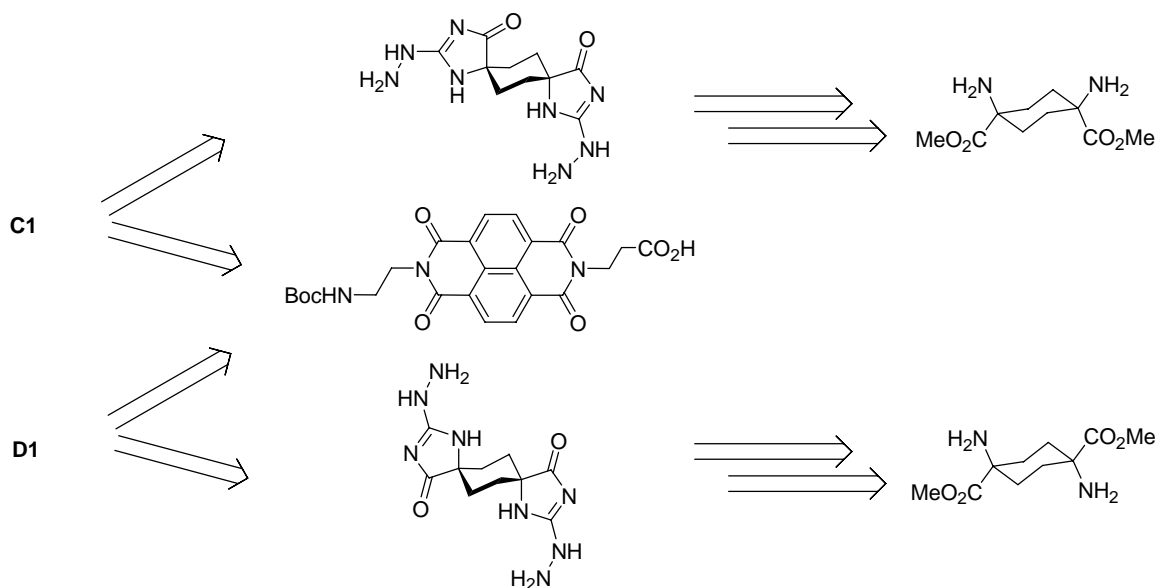
2.3.2 Synthesis

2.3.2.1 *trans*-Spiro linker

A Bucherer-Bergs reaction gave the two bis-spirohydantoins shown (**1** and **2**) in excellent yields (Haroutounian, 1989; Gadwood, 1993) (Scheme 2.2). Both structures were found to be in the *trans* geometry based on NOESY spectra (not shown). The kinetic and/or thermodynamic reasons for the exclusive formation of the *trans* geometry (i.e. no *cis* products were observed) are currently unknown. The *trans*-1,4-diamino-1,4-dicarboxylic acid dimethyl ester (**3**) was synthesized in good yield. The structure of intermediate **5** was confirmed by X-ray crystallography (Figure 2.3).

Scheme 2.2
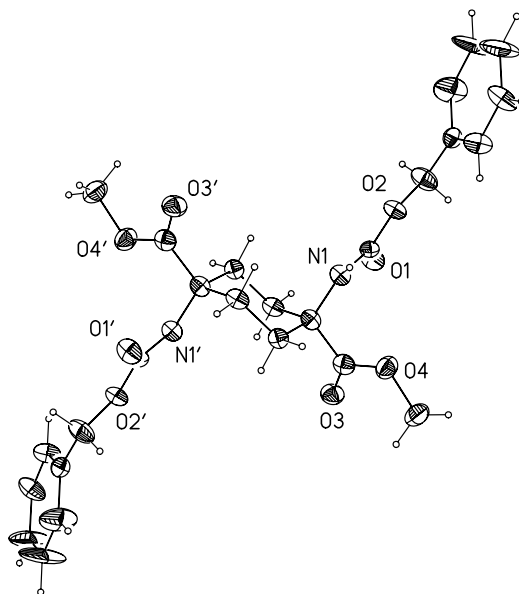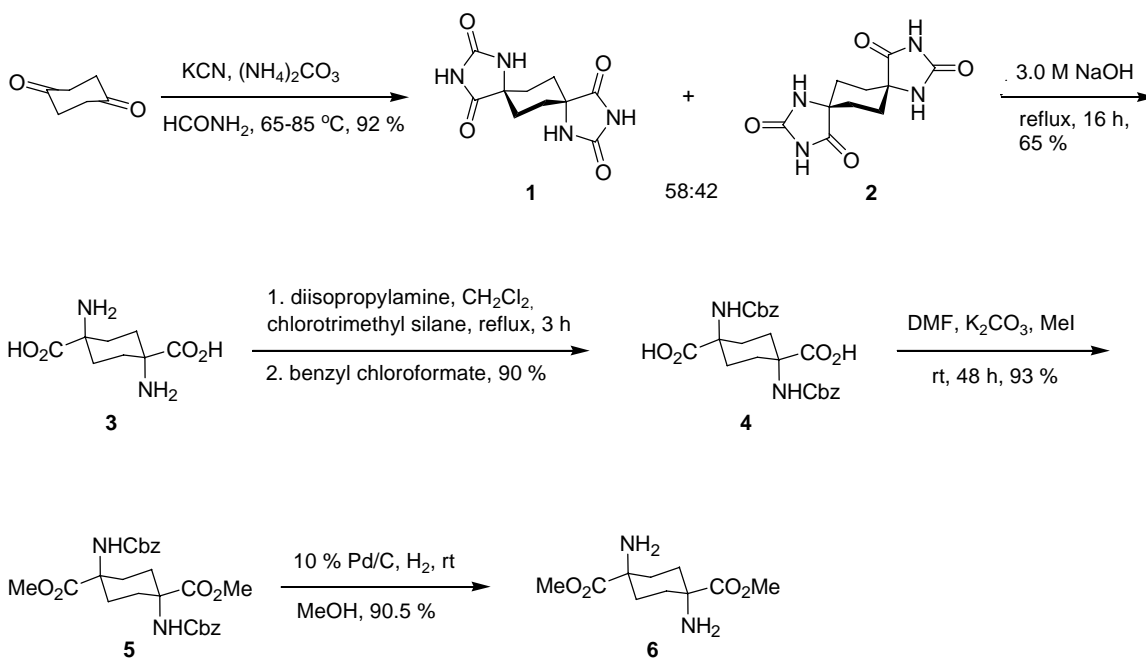
Figure 2.3 Displacement ellipsoid diagram for **5**.

45
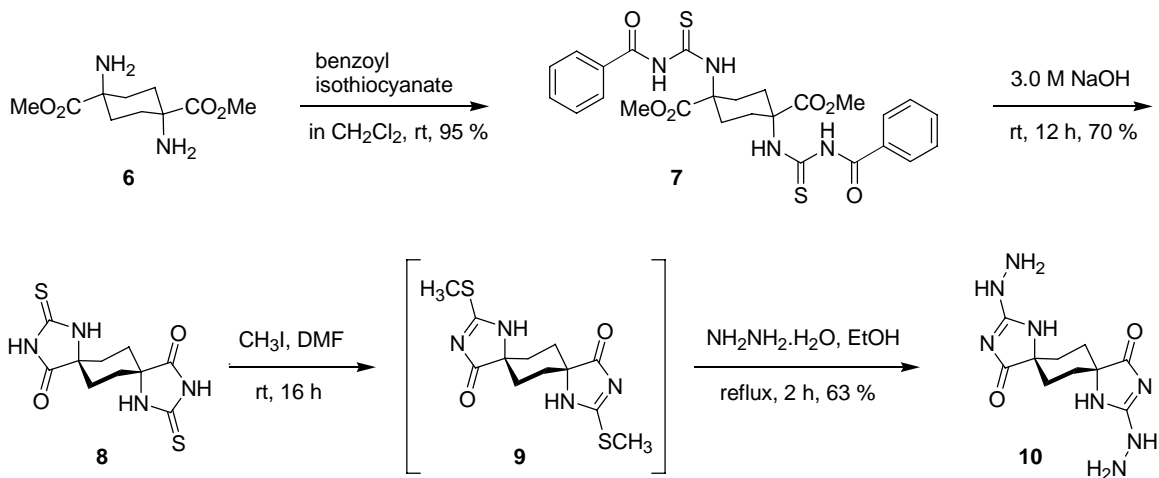
Scheme 2.3

Scheme 2.3 describes the conversion of diamino dimethyl carboxylate **6** into the desired *trans* linker **10**. The easy removal of benzoyl group makes benzoyl isothiocyanate as an ideal reagent to form a precursor to the *trans*, bis-spirothiohydantoin **8**. Using mild base, intermediate **8** was derived from cyclization in 75% yield. The structure of **8** was confirmed by X-ray crystallography (Figure 2.4). Subsequent methylation and hydrazine substitution gave the desired *trans* linker **10** in good overall yield.
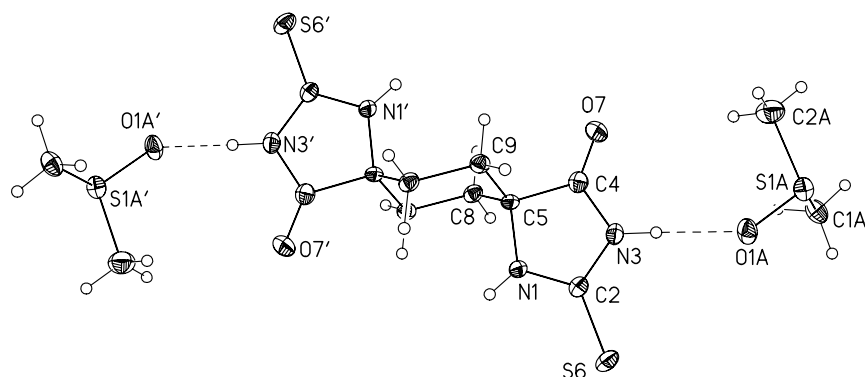
Figure 2.4 Displacement ellipsoid diagram for **8**:DMSO.

2.3.2.2 *cis*-Spiro linker for C1

Since the *cis* oriented bishydantoins could not be made via Bucherer-Bergs reaction when the two carbonyl groups of 1,4-cyclohexane were cyclized simultaneously, the two cyclization reactions were carried out separately. For this purpose, mono-protected 1,4-cyclohexanedione was used as shown in Scheme 2.4.

Scheme 2.4
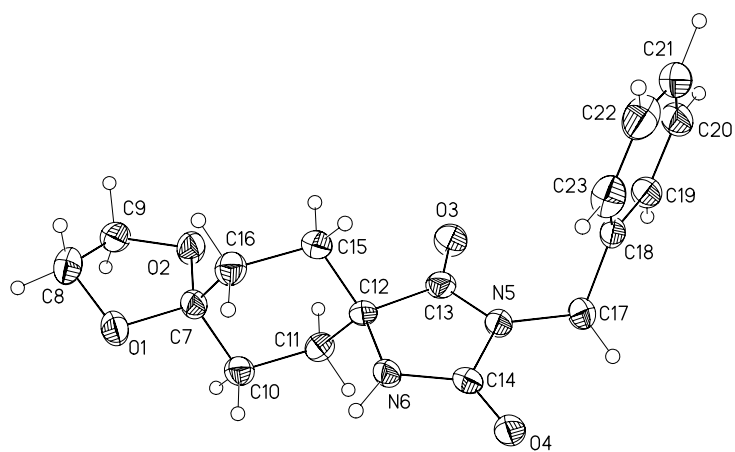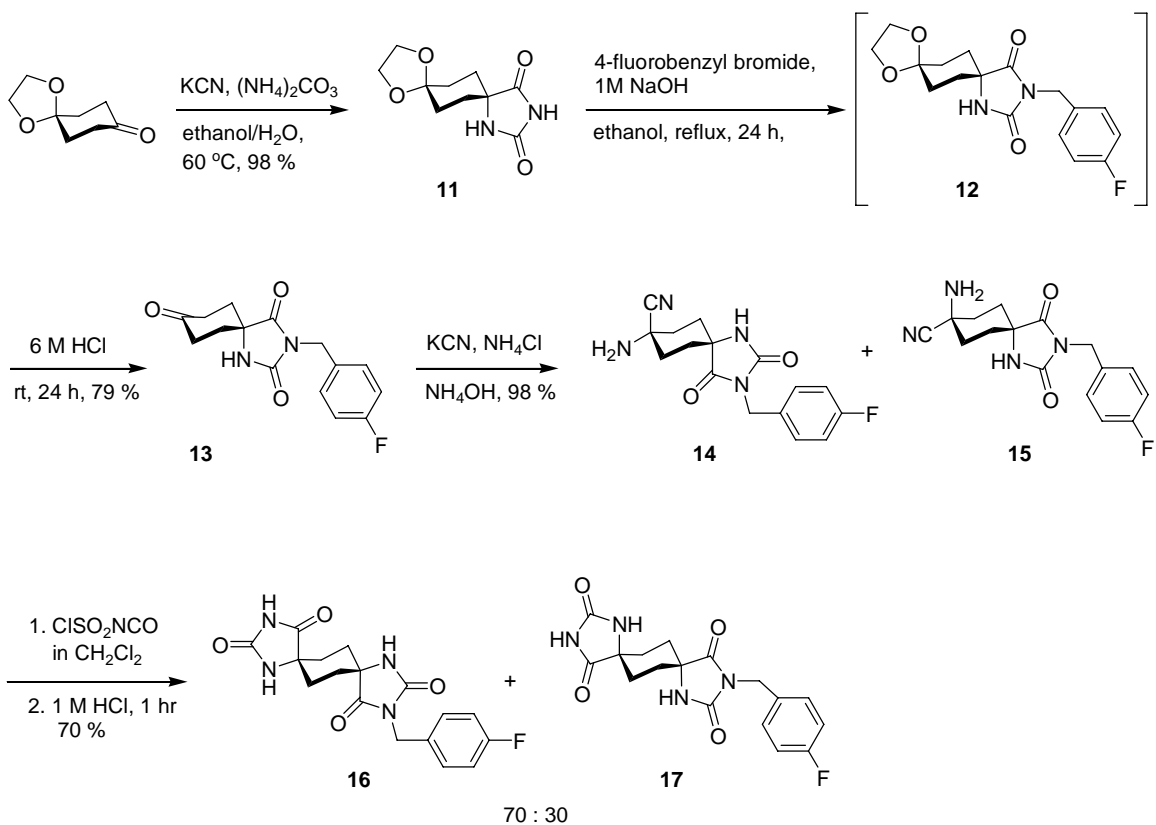
Figure 2.5 Displacement ellipsoid diagram for **12a** (an analogue of **12**).

The Bucherer-Bergs reaction was carried on the mono-protected 1,4-cyclohexanedione to give mono-spirohydantoin **11**. NMR analysis confirmed that **11** contained the NH and CO groups in axial and equatorial orientations, respectively. Alkylation gave compound **12**, whose geometry was confirmed by X-ray crystallography (Figure 2.5). Acidic deprotection gave compound **13** in high yield. Based on the difference between A-values of -NH$_2$ and -CN, we anticipated that a Strecker reaction would give a *cis, trans*, mixture of amino monohydantoins (Crooks, 1986). However, following a Strecker reaction on intermediate **13**, cyclization using chlorosulfonyl isocyanate and 1 M HCl gave two bis-spirohydantoins **16** and **17**. Their NOESY spectra (not shown) suggested both were of the *trans* geometry.

The fact that no *cis* isomer was found under the above reaction conditions might indicate that the reaction occurring on one carbonyl group directs the reaction pattern on the other. In addition, the unfavorable interaction between NH and NH$_2$ may contribute as well. Interestly, in the case of an analogous five-member ring system, a *cis* and *trans* mixture was formed (Levins and Schafmeister, 2003).

Scheme 2.5
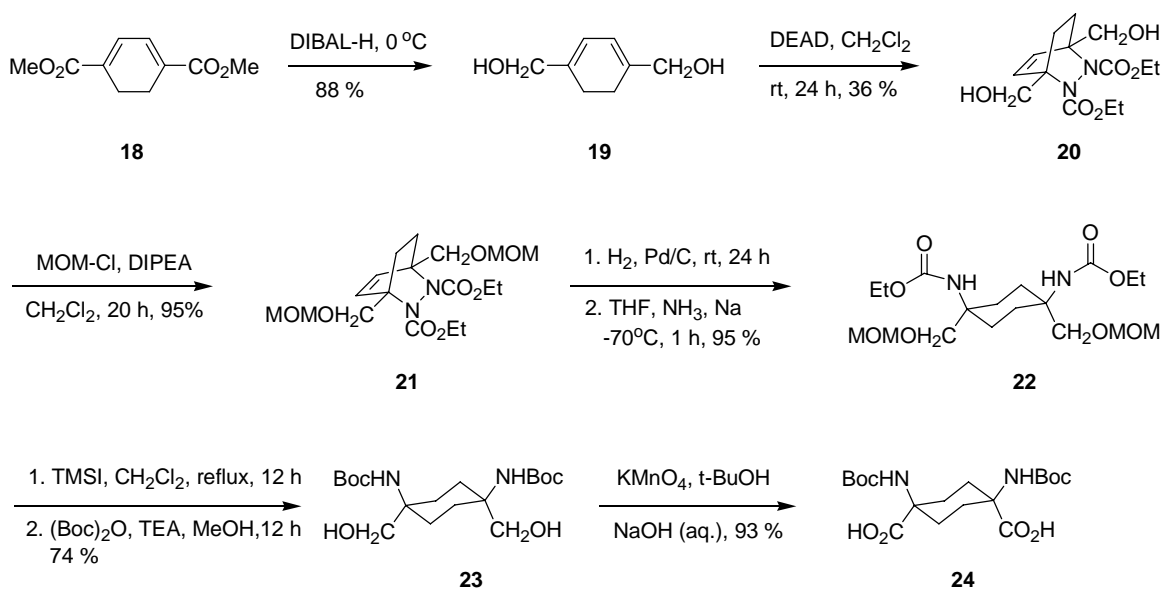
A second attempt was made to obtain the *cis* structure based on a Diels-Alder strategy (Scheme 2.5) via key intermediate **24**. Starting from **18**, which can be made in 3 steps (Chapman, 1970), DIBAl-H reduction at 0 °C and Diels-Alder reaction with DEAD gave compound **20** in a modest yield due to competing side reactions such as aromatization of **19** and the ene reaction (Franzus and Surridge, 1962; Shah and George, 1971; Jenner and Salem, 1990). Enough material could be generated to continue the synthesis, but this step needs to be optimized. Following protection of the hydroxyl groups, reductive cleavage using Na in ammonia gave compound **22** in high yield. The two hydroxyl groups were successfully oxidized using basic KMnO$_4$.

Scheme 2.6
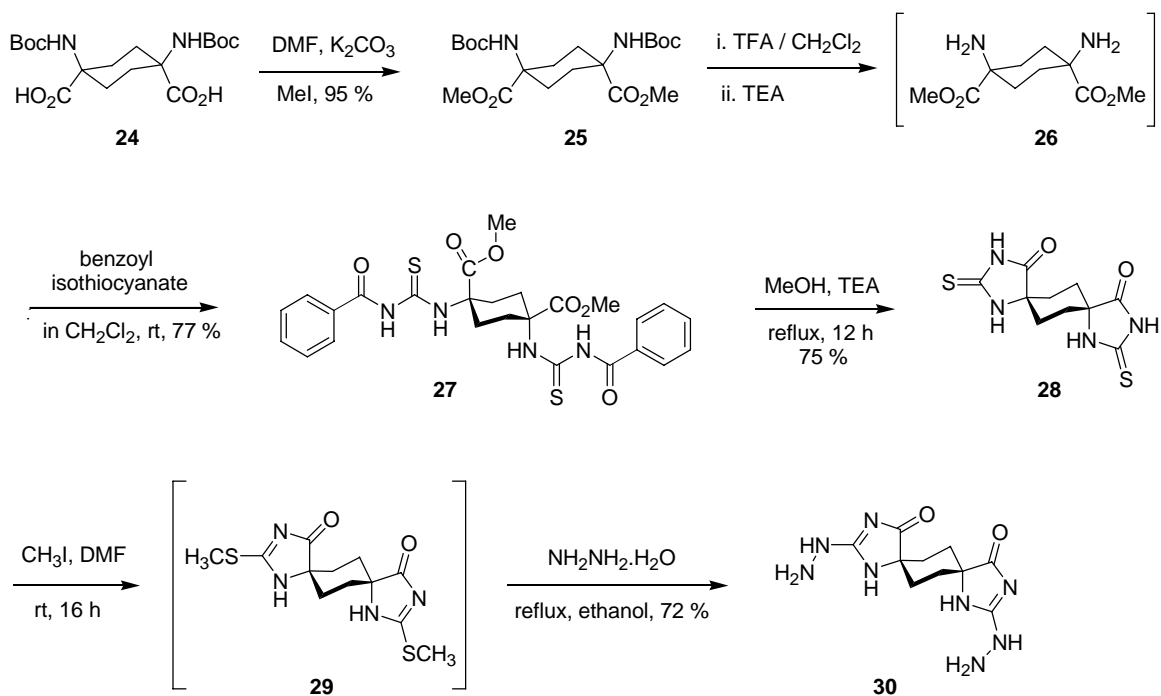
Shown in Scheme 6 are the final steps to make the *cis* linker **30** analogous to the procedures used in the *trans* linker synthesis. The *cis* geometry of bis-spirothiohydantoin intermediate **28** was confirmed by X-ray crystallography (Figure 2.6).

The corresponding *trans* and *cis* NDI dimers **D1** and **C1** were synthesized in high yields using PyBOP, HOBT as the coupling reagents (Scheme 2.7) to attach previously reported NDI derivatives to the *trans* and *cis* linkers (Lokey, 1997).
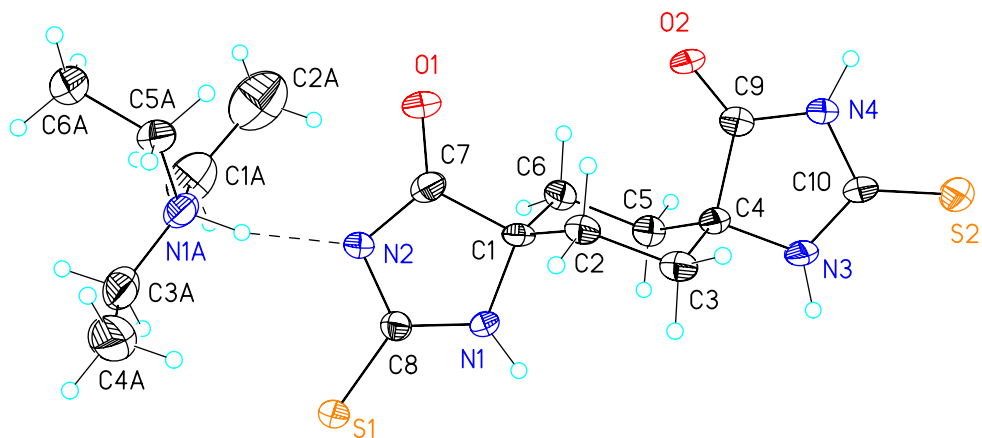
Figure 2.6 Displacement ellipsoid diagram for **28**:triethyl amine.

Scheme 2.7 Solution phase dimer synthesis.

| 31 | 30 | 10 |

1. PyBOP, HOBT, DIPEA in DMF | 2. TFA in CH$_2$Cl$_2$, 60-70 %

C1   or   D1
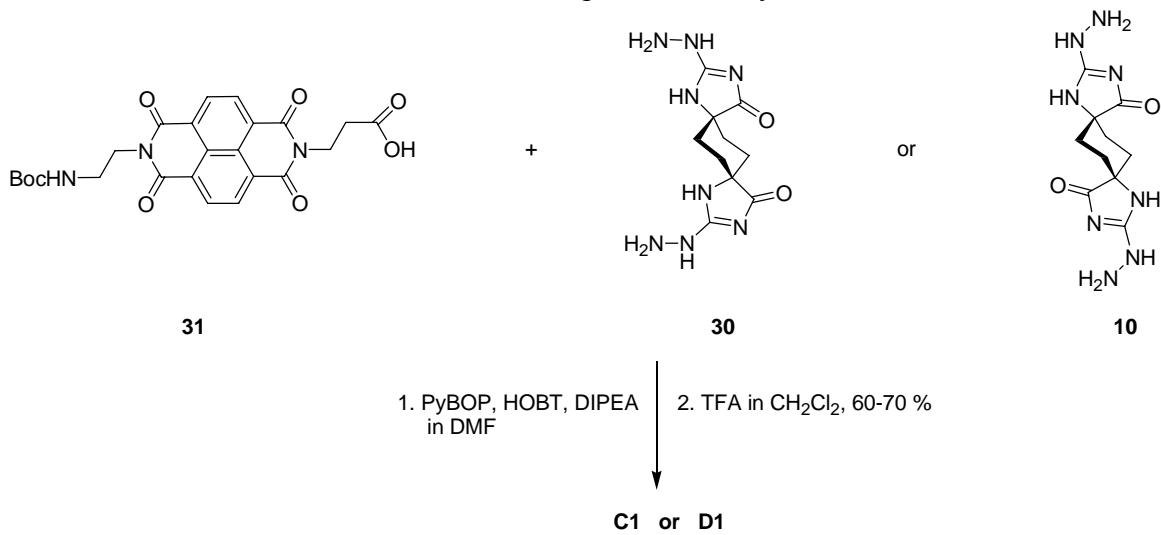
2.3.3 Kinetics studies

As a preliminary investigation into the interaction of **D1** and **C1** with DNA, dissociation rates were determined for CT DNA, poly(dGdC) and poly(dAdT). These values were compared to similar data obtained with a previously reported dimer containing a flexible peptide linker (referred to as **G₃K**) (Guelev, 2001(a)). For NDI bisintercalators, dissociation rates are generally sufficiently slow to allow monitoring using a spectrophotometer without the need for a stopped-flow apparatus. Dissociation rate measurements were carried out with 2% SDS in the buffer to sequester the dissociated intercalators following dissociation from the DNA (Lokey, 1997). Control studies indicated that the dimers did not reassociate with the DNA under these conditions. All dissociation profiles were adequately described by one phase exponential decay, $A = a_0 \, \mathrm{Exp} \, (-kt) + b_0$ (Chen, 1997).

The dissociation rate data reveal that all dimers exhibit a dramatic preference for poly(dGdC) over poly(dAdT) (Table 2.1 and Figures 2.7, 2.8 and 2.9), a trend seen with previously reported NDI based intercalation (Lokey, 1997). The dissociation half-lives measured using poly(dGdC) for all of the dimers extend over 10 minutes, with **C1** exhibiting a 50 minute half-life. In contrast, the dissociations of the three dimers from poly(dAdT) were almost complete within 1 minute, with the *cis* dimer **C1** exhibiting the shortest half-life. Taken together, our data indicate that the *cis* dimer **C1** binds with greatest overall affinity to poly(dGdC), and with the greatest overall ability to discriminate between poly(dGdC) and poly(dAdT).

Table 2.1 Dissociation rates and half-live times

DNA type	D1		C1		G$_3$K	
	k (min^{-1})	t (min)	K (min^{-1})	t (min)	k (min^{-1})	t (min)
Poly(dAdT)	1.02	0.682	2.86	0.243	1.10	0.629
Poly(dGdC)	0.0528	13.1	0.0138	49.9	0.0471	14.7
CT DNA	0.51	1.4	0.13	5.3	0.37	1.9
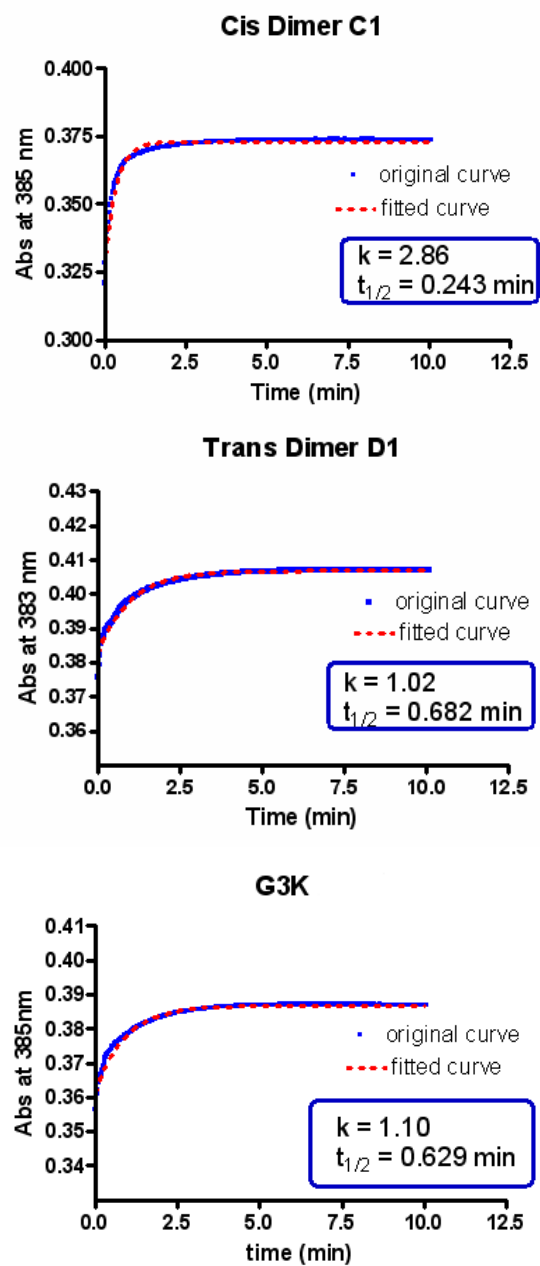
Figure 2.7 UV dissociation profile of compounds **C1**, **D1** and **G₃K** from poly(dAdT) in 2% SDS and the fitted curves.
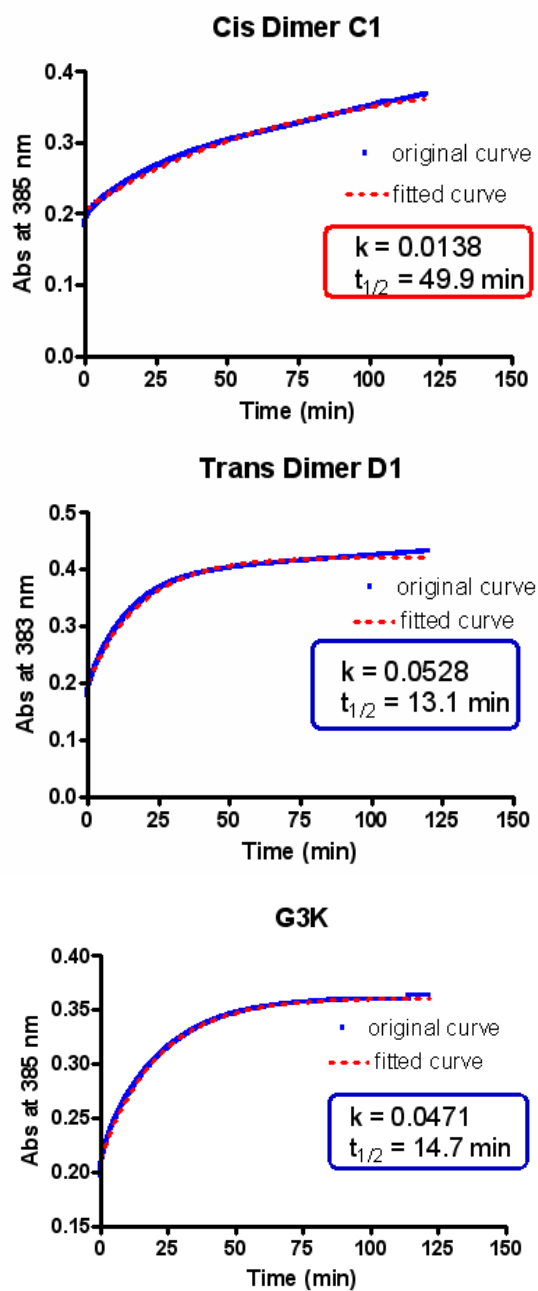
Figure 2.8 UV dissociation profile of compounds **C1**, **D1** and **G₃K** from poly(dGdC) in 2% SDS and the fitted curves.
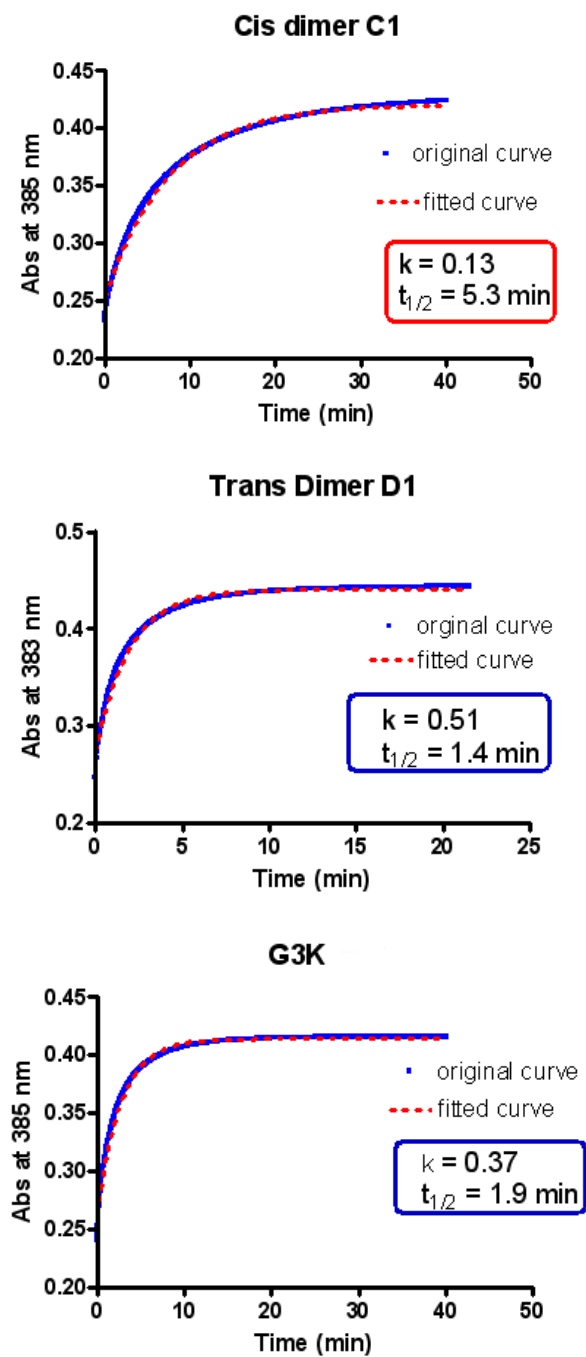
Figure 2.9 UV dissociation profile of compounds **C1**, **D1** and **G₃K** from CT DNA in 2% SDS and the fitted curves.

2.3.4 DNAse I footprinting

The distinct dissociation kinetics profile of **C1** suggested that **C1** binds strongly to GC rich sequences. In order to further examine its binding affinity and sequence specificity, DNAse I footprinting of **C1** on a naturally occurring plasmid fragment (232 bp) was carried out. This DNA fragment was chosen because it contains multiple palindromic sequences and has been successfully used to screen the binding specificity of NDI ligands previously (Guelev, 2000). As shown in Figure 2.10, **C1** binds to multiple 6 bp sequences. Detailed sequence analysis suggested that **C1** prefers to bind to GGNNCC sequences, consistent with the previous observation that NDI-based intercalating units possess a preference for GpG steps (Guelev, 2001(a)). Based on this information, a synthetic 92 bp DNA fragment was constructed, containing five 5'-GGNNCC-3' sites (Figure 2.11). The DNAse I footprinting result of **C1** on this DNA was shown in Figure 2.12. As reported previously, **G₃K** has a distinct binding site at 5'-GGTACC-3' with a K_D ~100 nM (Guelev, 2001(a)). **D1** showed some non-specific binding behavior, from targeting less than six-base pairs to only binding at high concentration (> 0.25 μM). **C1**, however, shows very different binding characteristics from the others. It not only binds to GGTACC with an even higher affinity than **G₃K**, but it also binds to other sequences with good affinity, such as GGGCCC, GGCGCC, and GGATCC. The variety of binding sites with good binding affinity (K_D ~100 nM) demonstrates the potential of this spiro tricyclic scaffold being a platform for future polyintercalator devlopment.
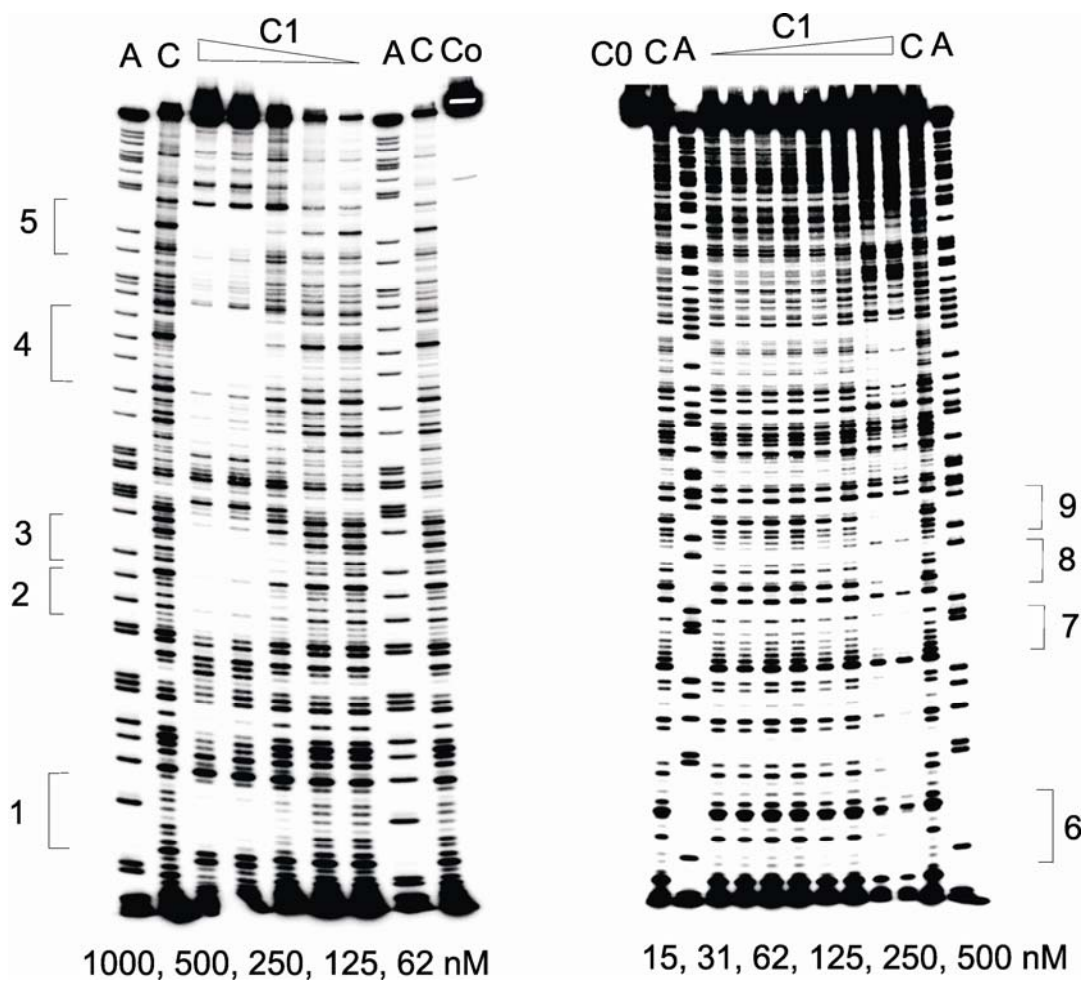
Figure 2.10 The DNAse I footprinting of **C1** on a 232 bp plasmid fragment: (+) 1. GGTAGT; 2. ACTCAG; 3. GGCACC; 4. CCTCGG; 5. GGTACT. (-) 6. GGCGAT; 7. GCGGGA; 8. GGCCCG; 9. GGTACT.

(+) strand:
5'GGTAATTTATTAT**GGTACC**CATATTAAA**GGATCC**AATAAATTA**GGGCCC**CAT TATTA**GGCGCC**TATTATAAT**GGGGCC**TTATAATTATTATTGG-3'

(-) strand:
3'CCATTAAATAATA**CCATGG**TATAATTT**CCTAGG**TTATTTAAT**CCCGGG**TA ATAAT**CCGCGG**ATAATATTA**CCCCGG**AATATTAATAATAACC-5'

Figure 2.11 The DNA sequence of a 92mer DNA.
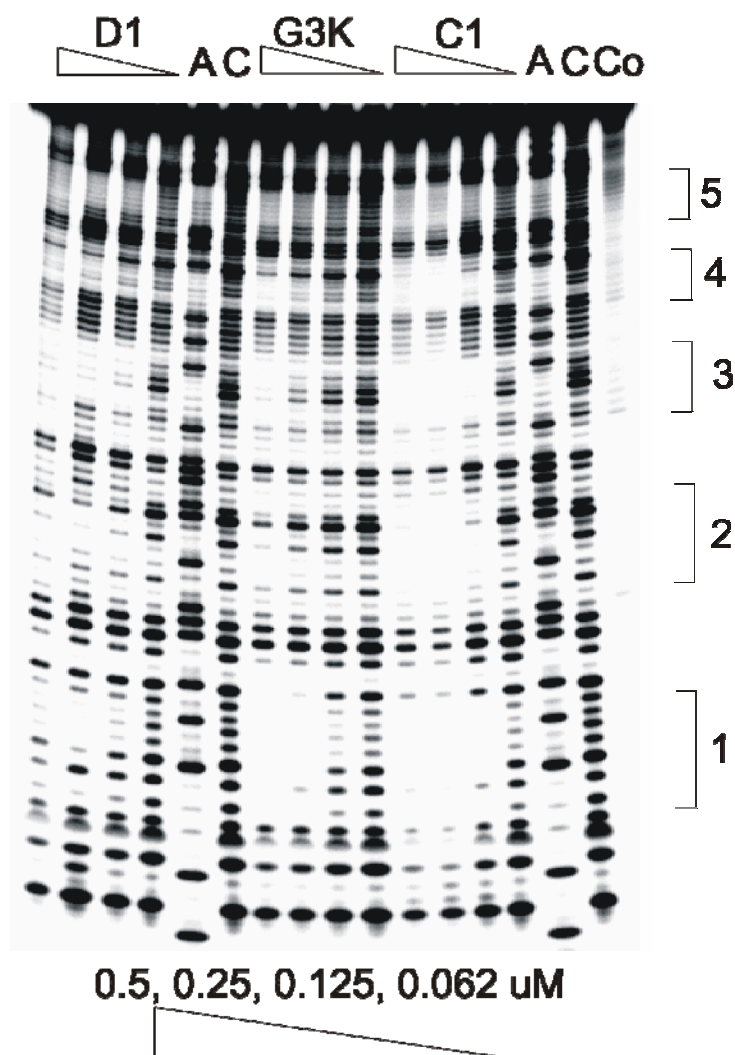
Figure 2.12 The DNAse I footprinting of **C1**, **D1** and **G₃K**: 1. GGTACC; 2. GGATCC; 3. GGGCCC; 4. GGCGCC; 5. GGGGCC. Lane A represents Adenine-specific sequencing reaction. Lane Co contains DNA without DNAse I. Lane C contains DNA with DNAse but no ligands.

2.4 Conclusion

Two novel threading bisintercalators containing *cis* and *trans* oriented spiro-cyclic linkers were successfully synthesized, although preparation of the *cis* derivative proved to be a considerable challenge. Data from dissociation kinetics suggested that the

DNA binding behavior of the *cis* oriented dimer **C1** is different from that of the *trans* oriented dimer **D1** and linear dimer **G₃K** based on studies with poly(dGdC), poly(dAdT), and CT DNA. The *cis* dimer **C1** exhibits the slowest off-rate towards poly(dGdC) and the fastest off-rate with poly(dAdT). Further DNAse I footprinting experiments demonstrated that **C1** binds to GGGCCC, GGTACC and GGCGCC sequences with good affinity, but with highest affinity for GGTACC sequence, indicating the strong possibility of using **C1** as the scaffold for further ligand development. The most straightforward interpretation of these preliminary results is that by constraining the linker in the appropriate *cis* geometry, the binding affinity has been enhanced compared to our flexible linker designs. Future functionization of the central ring of **C1** should bestow the new ligands with desired sequence specificity.

2.5 Experimental

2.5.1 General

All chemicals were purchased from Aldrich, Acros, or Novabiochem. Dry solvents were either purchased or dried using common laboratory techniques. All chemical reactions were carried out in oven-dried glassware under an argon atmosphere. Silica gel 60 F_{254} glass-based plates (Merck) were used for TLC. Flash chromatography was carried out using ICN SiliTech 32-63D 60 Å silica gel. All NMR spectra were recorded on Varian 300 or 500 MHz instruments. CDCl₃ (δ_H = 7.24 and δ_C = 77.0 ppm) or DMSO-d_6 (δ_H = 2.49 and δ_C = 39.5) or D₂O (δ_H = 4.67 ppm) were used as solvents. Assignment of ¹³C signals are based on ¹H, ¹³C-correlated 2D NMR spectra.

1,3,9,11-Tetraaza-dispiro[4.2.4.2]tetradecane-2,4,10,12-tetraone (1, *α*) and (2, *β*). To a 40 mL Ace pressure tube were added 1,4-cyclohexanedione (1.12 g, 0.01 mol), potassium cyanide (1.43 g, 0.022 mol) and ammonium carbonate (4.8 g, 0.05 mol). Formamide (35 mL) was then added and the pressure tube was tightly closed. The tube was heated at 60 °C for 24 h, then 85 °C for another 24 h with vigorous stirring. Vacuum filtration at room temperature gave a white solid after rinsing with cold water and ethanol. The filtrate was treated with enough mixed ion-exchange resin (AG® 501-X8 (D) Resin, 20-50 mesh, BIO-RAD) to extract remaining cyanide anions. The crude solid was washed with CH_2Cl_2 (3×20 mL), giving a pure solid product (2.32 g, 92%): ^{1}H NMR (500 MHz, DMSO-d_6): δ 10.57 (br, s, 2H), 8.53 (s, 1H, H1' of **1**), 8.18 (s, 1H, H1' of **2**), 2.10-2.05 (m, 4H, equatorial of **2**), 1.99-1.91 (m, 4H, equatorial of **1**), 1.68-1.63 (m, 4H, axial of **2**), 1.62-1.57 (m, 4H, axial of **1**); ^{13}C NMR (125.7 MHz, DMSO-d_6): δ 177.4 (C4' of **2**), 177.3 (C4' of **1**), 155.8 (C2' of **1**), 155.7 (C2' of 2), 60.3 (C1 and C4 of **1**), 59.4 (C1 and C4 of **2**), 28.8 (C2, C3, C5 and C6 of **2**), 28.4 (C2, C3, C5 and C6 of **1**) ; HRMS-CI calcd for $C_{10}H_{13}N_4O_4[M+ H]^+$: 253.0937, found 253.0926.

trans-**1,4-Diamino-cyclohexane-1,4-dicarboxylic acid (3).** The dihydantoins **1** and **2** (1.16 g, 0.0046 mol) were suspended in 12 mL 3 M NaOH aqueous solution and the mixture was heated at reflux (120 °C) while stirring for 16 h. The reaction mixture then was cooled to room temperature and filtered through a fritted glass filter. The filtrate was brought to pH 2 with 3 M HCl at 0 °C and a white precipitate formed. After 10 minutes at

0 °C, the mixture was filtered through a fine fritted glass filter. The solid was washed with copious cold deionized water and dried at 50 °C under vacuum oven over 12 h, yielding the desired product **3** (0.59 g, 65%): ^{1}H NMR (300 MHz, D_2O + NaOD): δ 1.75 (d, 4H, equatorial, J = 9 Hz), 1.21 (d, 4H, axial, J = 9 Hz); ^{13}C NMR (75.5 MHz, D_2O + NaOD, uncorrected): δ 184.8 (COOH), 57.5 (C1 and C4), 31.1 (C2, C3, C5 and C6). No useful mass data were available due to insolubility of this compound in common solvents.

***trans*-1,4-Bis-(benzyloxycarbonylamino)-1,4-cyclohexanedicarboxylic acid (4).** To a 50 mL flask equipped with a magnetic stir bar and rubber septum were added diamino acids **3** (0.6 g, 0.003 mol). The solid was suspended in 20 mL of CH_2Cl_2. Diisopropylamine (1.72 mL, 0.0099 mol) was added with stirring. Chlorotrimethylsilane (1.71 mL, 0.0135 mol) was then added slowly, the rubber septum replaced with a reflux condenser, and the solution was heated at reflux under nitrogen for 4 h. The condenser was removed and replaced with another rubber septum. The flask was cooled in an ice bath, and benzyl chloroformate (Cbz-Cl) (0.81 mL, 0.006 mol) was added in one portion via syringe. The stirred solution was allowed to warm and stirred at RT overnight. The mixture was concentrated to a pink syrup, dissolved in 100 mL 1.0 M aqueous NaOH and transferred to a separatory funnel. The aqueous solution was washed with ether (2×100 mL). The ether washes were combined and back extracted with water (2×200 mL). The aqueous layers were combined and acidified to pH 2 with 3 M HCl at 0 °C. White solid formed and was filtered through a fine fritted glass filter. The solid was washed with cold water then air-dried to give the desired product **4** (1.27 g, 90%): ^{1}H NMR (300 MHz,

DMSO-d_6): δ 12.40 (br s, 2H, COOH), 7.56 (s, 2H, 2NH), 7.36-7.31 (m, 10H, aromatic), 5.00 (s, 4H, 2CH_2Ph), 1.92-1.82 (m, 8H, 4CH$_2$ of central ring); ^{13}C NMR (75 MHz, DMSO-d_6): δ 176.0 (COOH), 155.5 (CONH), 137.0, 128.3, 127.8 and 127.7 (aromatic), 65.1 (CH$_2$Ph), 57.4 (C1 and C4), 26.4 (C2, C3, C5 and C6); HRMS-CI calcd for C$_{24}$H$_{27}$N$_2$O$_8$[M+ H]$^+$: 471.1767, found 471.1774.

trans-1,4-Bis-(benzyloxycarbonylamino)-1,4-dimethoxycarbonyl-cyclohexane   (5).
Dicarboxylic acid **4** (1.955 g, 4.1 mmol) was dissolved in dry DMF (33 mL) in a 100 mL round bottom flask, and then anhydrous K$_2$CO$_3$ (1.72 g, 12.4 mmol) was added with vigorous stirring. MeI (0.65 mL, 10 mmol) was added slowly by syringe and the reaction was stirred at room temperature for 48 h. 120 mL of H$_2$O was added at the end of the reaction and the product precipitated. Filtration in a fine fritted glass filter followed by washing with cold water gave the desired product (1.9 g, 93%, R_f = 0.25 in 1/49 CH$_3$OH/CH$_2$Cl$_2$): ^{1}H NMR (300 MHz, DMSO-d_6): δ 7.76 s, 2H, 2CONH), 7.37-7.30 (m, 10 H, aromatic), 5.01 (s, 4H, 2CH_2Ph), 3.56 (s, 6H, 2CH$_3$), 1.92-1.82 (m, 8H, 4CH$_2$ of central ring); ^{13}C NMR (75.5 MHz, DMSO-d_6): δ 174.7 (CO$_2$CH$_3$), 155.5 (CONH), 137.0, 128.4, 127.8 and 127.7 (aromatic), 65.2 (CH$_2$Ph), 57.6 (C1 and C4), 52.0 (CH$_3$), 26.4 (C2, C3, C5 and C6); HRMS-CI calcd for C$_{26}$H$_{31}$N$_2$O$_8$[M+ H]$^+$: 499.2080, found 499.2075.

trans-1,4-Diamino-1,4-dimethoxycarbonyl-cyclohexane (6). Ester **5** (0.5 g, 1 mmol) was suspended in 20 mL methanol in a 50 mL three-neck round bottom flask which was

pre-saturated with dry Ar. 10% Pd/C (0.15 g) was added slowly. The solution was degassed, placed under hydrogen atmosphere (10 psi) and stirred for 1.5 h. The reaction mixture was filtered through celite on a fine fritted glass filter. The filtrate was evaporated at room temperature then dried *in vacuo*. The dried solid was used directly for the next step without further purification (0.21 g, 90.5%, R_f = 0.33 in 2/3 CH_3OH/CH_2Cl_2): ^{1}H NMR (300 MHz, pyridine-d_5): δ 3.59 (s, 6H, 2CH_3), 2.51-2.45 (m, 4H, equatorial), 2.03 (br s, 4H, 2NH_2), 1.63-1.57 (m, 4H, axial); ^{13}C NMR (75.5 MHz, pyridine-d_5): δ 178.6 (CO_2CH_3), 56.9 (C1 and C4), 51.8 (CH_3), 30.5 (C2, C3, C5 and C6); HRMS-CI calcd for $C_{10}H_{19}N_2O_4[M+ H]^+$: 231.1345, found 231.1347.

***trans*-1,4-Bis-((N-benzoyl(thiocarbamoyl)amino)-1,4-dimethoxycarbonyl-cyclohexane (7)**. Diamine **6** (0.115 g, 0.5 mmol) was dissolved in dry CH_2Cl_2 (10 mL) in a 50 mL round bottom flask. Benzoyl isothiocyanate (0.15 mL, 1.1 mmol) was added by syringe with stirring. The solution became cloudy after stirring for a few hours. The reaction was terminated after 24 h by adding another 10 mL of CH_2Cl_2 to dilute the mixture and the solid was isolated by filtration using a fine fritted glass filter (0.26 g, 95%, R_f = 0.35 in 1/1 CH_3OH/CH_2Cl_2): ^{1}H NMR (300 MHz, DMSO-d_6): δ 11.56 (s, 2H, 2(C=O)NH(C=S)), 11.43 (s, 2H, 2NH(S=C)), 7.99-7.93 (m, 4H, aromatic), 7.68-7.64 (m, 2H, aromatic), 7.58-7.54 (m, 4H, aromatic), 3.61 (s, 6H, 2CH_3), 2.57-2.52 (m, 4H, equatorial), 1.97-1.93 (m, 4H, axial); ^{13}C NMR (75.5 MHz, DMSO-d_6): δ 179.8 (C=S), 172.0 (CO_2Me), 169.0 (CONH), 133.3, 131.9, 128.8 and 128.5 (aromatic), 70.1 (CH_2Ph),

60.0 (C1 and C4), 52.3 (CH$_3$), 26.4 (C2, C3 C5 and C6); HRMS-CI calcd for C$_{26}$H$_{29}$N$_4$O$_6$S$_2$[M+ H]$^+$: 557.1529, found 557.1525.

***trans*-Cyclohexane-1, 4-bis-spiro-5',5''-(2', 2''-thiohydantoin) (8).** Isothiocyanate **7** (0.51 g, 0.9 mmol) was dissolved in 10 mL 3 M NaOH aqueous solution. After stirring for 5 minutes, white precipitate formed within the flask. Stirring was continued for 16 h. The reaction was terminated by adding 1.0 M HCl to bring the solution pH to neutral. After being stirred for additional one hour the mixture was filtered and rinsed with cold water to give the crude product. The product was purified by crystallization in DMSO (0.18 g, 70%): ^{1}H NMR (300 MHz, DMSO-d_6): δ 11.65 (br s, 2H, 2(C=O)NH(C=S)), 10.70 (br s, 2H, 2(C=S)NH), 2.02 (d, 4H, equatorial, J = 9.3 Hz), 1.69 (d, 4H, axial, J = 9.3 Hz); ^{13}C NMR (75.5 MHz, DMSO-d_6): δ 181.4 (C=S), 178.0 (C=O), 63.8 (C1 and C4), 27.7 (C2, C3, C5 and C6); HRMS-CI calcd for C$_{10}$H$_{13}$N$_4$O$_2$S$_2$[M+ H]$^+$: 285.0480, found 285.0488.

***trans*-2,10-Dihydrazino-1,3,9,11-tetraaza-dispiro[4.2.4.2]tetradeca-1,9-diene-4,12-dione (10).** Thiohydantoin **8** (1.0 g, 3.5 mmol) was dissolved in anhydrous DMF (10 mL) with gentle heating in 25 mL round bottom flask. Methyl iodide (0.88 mL, 14 mmol) was then added via syringe. The flask was clear and light yellowish for the first few hours, and then some pale yellowish precipitate formed inside. After stirring the mixture for 12 h at room temperature, the solvent was removed at room temperature *in vacuo*. The dry solid was used directly for the next step due to its presumed moisture sensitivity. Solid

methylsulfanyl hydantoin **9** was dissolved in 30 mL of ethanol in a round bottom flask.

Hydrazine monohydrate (0.85 mL, 17.6 mmol) was added dropwise to the mixture. The

clear solution was heated at reflux for 2 hr, forming a white precipitate. The flask was

cooled down to room temperature and vacuum filtration gave a white powder that was

further dried *in vacuo* (0.62 g, 63%): ^{1}H NMR (300 MHz, DMSO-d_6): δ 8.90 (br s, 2H,

2NH), 8.28 (s, 2H, 2N*H*NH$_2$), 4.49 (s, 4H, 2NH$_2$), 1.99 (d, 4H, equatorial, J = 8.6 Hz),

1.34 (d, 4H, axial, J = 8.6 Hz); ^{13}C NMR (75.5 MHz, DMSO-d_6 + TFA): δ 178.0 (C=O),

158.9 (NH-C=N), 64.0 (C1 and C4), 29.1 (C2, C3, C5 and C6); HRMS-CI calcd for

$C_{10}H_{17}N_8O_2[M+H]^+$: 281.1474, found 281.1465.

9,12-Dioxa-1,3-diaza-dispiro[4.2.4.2]tetradecane-2,4-dione (11). To a 40 mL pressure

tube, 1,4-cyclohexanedione monoethylene acetal (1.56g, 0.01 mol), potassium cyanide

(0.65 g, 0.01 mol) and ammonium carbonate (2.4 g, 0.025 mol) were added followed by

35 mL ethanol/water (v/v = 1:1). The reaction tube was tightly capped and heated at 65

°C for 24 h with stirring. The reaction mixture was cooled down to 0 °C using an ice

bath. Approximately half of the solvent was removed by vacuum evaporation and the

solid was isolated by vacuum filtration. The filtrate was treated with mixed ion exchange

resin (AG$^®$ 501-X8 (D) Resin, 20-50 mesh, BIO-RAD) to extract the remaining KCN.

The solid product was rinsed with dichloromethane and dried *in vacuo* (2.21 g, 98%): [1]H

NMR (300 MHz, DMSO-d_6): δ 10.55 (bs, 1H, (C=O)NH(C=O)), 8.44 (s, 1H, C=ONH-

C), 3.86 (s, 4H, O(CH$_2$)$_2$O), 1.87-1.54 (m, 8H, central ring); ^{13}C NMR (75.5 MHz,

DMSO-d_6): δ 178.2 (C4'), 156.3 (C2'), 106.8 (C1), 63.7 and 63.6 (OCH$_2$CH$_2$O), 61.0

(C4), 31.3 (C2 and C6), 29.8 (C3 and C5); HRMS-CI calcd for $C_{10}H_{15}N_2O_4[M+ H]^+$: 227.1032, found 227.1028.

3-(4-Fluoro-benzyl)-1,3-diaza-spiro[4.5]decane-2,4,8-trione (13). Hydantoin **11** (1.13 g, 0.005 mol) was suspended in 1.0 M NaOH solution (5 mL) and ethanol (5 mL) and the mixture was heated at reflux for 15 min. To this solution was added 4-fluorobenzyl bromide (0.59 mL, 0.0048 mol) dropwise through the top of a reflux condenser. The mixture was heated at reflux for 24 h and allowed to cool in ice bath. The resulting precipitate was washed with water and recrystallized from ethanol to give pure product **12. 12** was then dissolved in 3 M HCl (10 mL) and ethanol (10 mL). The clear solution was stirred at room temperature for 16 h. The reaction was terminated by adding 3 M NaOH to adjust the pH of the solution to 7.5. CH_2Cl_2 (4×40 mL) was used to extract the product from the mixture. The CH_2Cl_2 layer was dried (Na_2SO_4), followed by vacuum evaporation to yield the crude product. Flash silica gel chromatography (1/19 CH_3OH/CH_2Cl_2) gave the product **13** (1.14 g, 79%, R_f = 0.33 in 1/19 CH_3OH/CH_2Cl_2): ^{1}H NMR (300 MHz, DMSO-d_6): δ 9.07 (s, 1H, NH), 7.31-7.26 (m, 2H, aromatic), 7.18-7.12 (m, 2H, aromatic), 4.54 (s, 2H, CH_2Ph), 2.55-2.48 (m, 2H, axial on C2 and C6), 2.38-2.32 (m, 2H, equatorial on C2 and C6), 2.17-2.08 (m, 2H, axial on C3 and C5), 1.95-1.89 (m, 2H, equatorial on C3 and C5); ^{13}C NMR (75.5 MHz, DMSO-d_6): δ 208.0 (CO), 175.8 (C1'), 161.0 (d, aromatic C-F, $^1J_{CF}$ = 242 Hz), 155.5 (C3'), 132.9 (aromatic), 129.3 (d, aromatic C-(C-C-F), $^3J_{CF}$ = 8.8 Hz), 115.3 (d, aromatic C-(C-F), $^2J_{CF}$ = 21.5

Hz), 59.8 (C4), 40.4 (CH_2Ph), 36.3 (C2 and C6), 32.9 (C3 and C5); HRMS-CI calcd for $C_{15}H_{16}N_2O_3F[M+ H]^+$: 291.1145, found 291.1141.

8-Amino-3-(4-fluoro-benzyl)-2,4-dioxo-1,3-diaza-spiro[4.5]decane-8-carbonitriles (14 and 15). To a 100 mL round bottom flask was added 4-fluorobenzyl hydantoin cyclohexanone **13** (0.6 g, 2.06 mmol), potassium cyanide (0.28 g, 4.3 mmol), ammonium chloride (0.24 g, 4.4 mmol), ammonium hydroxide (15 mL) and ethanol (30 mL). The flask was stirred at room temperature for 16 h. The clear reaction solution was then transferred to a separatory funnel and extracted with CH_2Cl_2 (4×50 mL). The aqueous layer was treated with mixed ion exchange resin ($AG^{®}$ 501-X8 (D) Resin, 20-50 mesh, BIO-RAD) to absorb KCN. The organic layer was dried over Na_2SO_4, followed by vacuum evaporation to give the desired product (0.64 g, 98%, **14**-77%, **15**-23%, R_f (**14**) = 0.16, R_f (**15**) = 0.22 in 1/19 CH_3OH/CH_2Cl_2); ^{1}H NMR (300 MHz, DMSO-d_6): δ 8.86 (s, 1H, NH, **14**), 8.78 (s, 1H, NH, **15**), 7.32-7.10 (m, 8H, aromatic, **14** and **15**), 4.51 (s, 2H, CH_2Ph, **14**), 4.48 (s, 2H, CH_2Ph, **15**), 2.68 (br s, 2H, NH_2, **14**), 2.54 (br s, 2H, NH_2, **15**), 2.01-1.68 (m, 16H, central ring, **14** and **15**); ^{13}C NMR (75.5 MHz, DMSO-d_6): δ 176.1 (C1'), 163.0 and 159.8 (d, C-F, J = 243 Hz), 155.4 (C3' of **14**), 155.3 (C3' of **15**), 133.0 and 132.9 (aromatic), 129.3 and 129.2 (aromatic of **14**), 128.4 and 128.3 (aromatic of **15**), 125.2 (CN of **15**), 124.0 (CN of **14**), 115.9 and 115.2 (aromatic of **14**, J = 21.0 Hz), 114.9 and 114.6 (aromatic of **15**, J = 20.9 Hz), 62.1 (C1 of **15**), 59.7 (C1 of **14**), 50.2 (C4 of **14**), 47.9 (C4 of **15**), 40.3 (CH_2Ph), 32.2 (C2 and C6 of **14**), 31.0 (C2 and C6 of **15**), 30.1

(C3 and C5 of **14**), 27.8 (C3 and C5 of **15**); HRMS-CI calcd for $C_{16}H_{18}N_4O_2F[M+ H]^+$: 317.1414, found 317.1415.

trans-3-(4-Fluoro-benzyl)-1,3,9,11-tetraaza-dispiro[4.2.4.2]tetradecane-2,4,10,12-tetraone (16 and 17). To a mixture of **14** and **15** (100 mg, 0.316 mmol) in 4 mL of CH_2Cl_2 was added chlorosulfonyl isocyanate (48 mg, 0.34 mmol). After being stirred for 10 min at room temperature, the clear solution was concentrated *in vacuo* into a pale yellow form. Following addition of 3 mL of 1 N HCl the suspension was stirred for 10 minute at room temperature, then heated in an oil bath at 100 °C for 1 h. After approximately 15 min of heating, the reaction mixture became homogeneous, and then a precipitate formed. The reaction mixture was cooled to room temperature; the solid was filtered, washed with water, and dried to give 0.08 g of a mixture (70%) of **16** and **17** (7:3). ^{1}H NMR (300 MHz, DMSO-d_6): δ 10.63 (s, 1H, NH, **17**), 10.60 (s, 1H, NH, **16**), 8,98 (s, 1H, NH, **17**), 8.66 (s, 1H, NH, **16**), 8.61 (s, 1H, NH, **17**), 8.25 (s, 1H, NH, **16**), 7.28-7.24 (m, aromatic, **16** and **17**), 7.17-7.13 (m, aromatic, **16** and **17**), 4.50 (s, 2H, C*H₂*Ph, **17**), 4.49 (s, 2H, C*H₂*Ph, **16**), 2.12-2.07 (m, 4H, central ring, **16**), 2.00-1.97 (m, 4H, central ring, **17**), 1.70-1.60 (m, 8H, central ring, **16** and **17**); ^{13}C NMR (75.5 MHz, DMSO-d_6): δ 178.0 (C1' of **16**), 177.8 (C1' of **17**), 176.1 (C1" of **16**), 176.0 (C1" of **17**), 161.0 (d, C-F, J_{CF} = 239 Hz, **16** and **17**), 156.4 (C3' of **16** and **17**), 155.6 (C3" of **16**), 155.5 (C3" of **17**), 133.0 (aromatic, **16** and **17**), 129.2 (d, *C*-(C-C-F), $^3J_{CF}$ = 8 Hz, **16**), 129.1 (d, *C*-(C-C-F), $^3J_{CF}$ = 8 Hz, **17**), 115.3 (d, *C*-(C-F), $^2J_{CF}$ = 21.9 Hz, **16** and **17**), 60.8 (C1 of central ring of **17**), 60.0 (C4 of central ring of **17**), 59.6 (C1 of central ring of **16**),

58.9 (C4 of central ring of **16**), 40.2 (CH_2Ph of **17**), 40.1 (CH_2Ph of **16**), 29.1 (C2 and C6 of central ring of **16**), 29.0 (C3 and C5 of central ring of **16**), 28.6 (C2 and C6 of central ring of **17**), 28.5 (C3 and C5 of central ring of **17**); HRMS-CI calcd for $C_{17}H_{18}N_4O_4F[M+H]^+$: 361.1312, found 361.1297.

(4-Hydroxymethyl-cyclohexa-1,3-dienyl)-methanol (19). Starting material **18** (1.8 g, 9.25 mmol) was dissolved in anhydrous THF (20 mL) at 0 °C, which was then subject to ultra-pure Ar flushing for 10 min. 37 mL of DIBAL-H in THF(1.0 M, 37 mmol) was added to the above solution slowly via a double-ended needle. The clear solution was stirred rigorously at 0 °C for 30 min then quenched with MeOH (3 mL), followed by adding powdered $Na_2SO_4 \cdot 10H_2O$ (25 g). The mixture was stirred at room temperature for 1 hr then filtered through celite, and the filter cake was washed by copious EtOAc. The solution was evaporated *in vacuo* yielding a light orange oil. The oil was dissolved in EtOAc and applied to a flash column using a CH_2Cl_2/EtOAc gradient to elute the product (1.14 g, 88%, R_f = 0.24 in 3/2 CH_2Cl_2/EtOAc). Due to the instability of the product, it was used immediately for the next step: ^{1}H NMR (300 MHz, CDCl$_3$): δ_H 5.89 (s, 2H, CH-CH), 4.15 (d, 4H, 2CH_2OH, J = 5.7 Hz), 2.22 (s, 4H, 2CH$_2$), 1.93 (t, 2H, 2OH, J = 5.7 Hz); ^{13}C NMR (75.5 MHz, CDCl$_3$): δ_C 137.6 (C1 and C4), 119.0 (C2 and C3), 66.0 (CH$_2$OH), 23.9 (C5 and C6); HRMS-CI calcd for $C_8H_{12}O_2$ [M$^+$]: 140.0837, found 140.0839.

1,4-Bis-hydroxymethyl-2,3-diaza-bicyclo[2.2.2]oct-5-ene-2,3-dicarboxylic acid diethyl ester (20). Diethyl azodicarboxylate (DEAD) (5.0 mL, 11 .0 mmol) was added to a CH_2Cl_2 (40 mL) solution of **19** (1.40 g, 10 mmol). The solution was stirred at room temperature for 24 hr. The adduct **20** was isolated by column chromatography using a CH_2Cl_2/EtOAc gradient to give a viscous, colorless oil (1.13 g, 36%, R_f = 0.32 in 1/1 CH_2Cl_2/EtOAc): ^{1}H NMR (300 MHz, $CDCl_3$): δ 6.86 (d, 1H, C*H*=CH, J = 8.1 Hz), 6.26 (d, 1H, CH=C*H*, J = 8.1 Hz), 4.27-4.04 (m, 8H, 2C*H_2*OH and 2OC*H_2*CH_3), 2.39-2.32 (m, 1H, C*H*HCH_2), 1.98-1.91 (m, 1H, CH*H*CH_2), 1.45-1.37 (m, 1H, CH_2CH*H*), 1.29-1.19 (m, 7H, CH_2C*H*H and 2CH_3); ^{13}C NMR (75.5 MHz, $CDCl_3$): δ 159.3(CO_2Et), 158.1(CO_2Et), 136.5(C=C), 135.5(C=C), 66.4, 65.9, 65.2, 63.2, 62.8, 62.5, 27.5(CH_2CH_2), 26.7(CH_2CH_2), 14.5(CH_3), 14.3(CH_3); HRMS-CI calcd for $C_{14}H_{23}N_2O_6$ [M+H]$^+$ 315.1556, found 315.1543.

1,4-Bis-methoxymethoxymethyl-2,3-diaza-bicyclo[2.2.2]oct-5-ene-2,3-dicarboxylic acid diethyl ester (21). A solution of **20** (1.0 g, 3.1 mmol), MOM-Cl (0.5 mL, 6.25 mmol), DIPEA (1.40 mL, 7.75 mmol) in dry CH_2Cl_2 (40 mL) under N_2 was stirred at room temperature for 24 h, then quenched with 40 mL saturated $NaHCO_3$ solution. The layers were separated, and the aqueous layer was extracted with CH_2Cl_2. The combined organic layers were washed with water, brine, dried (Na_2SO_4), filtered and concentrated *in vacuo*. Flash chromatography on silica gel (CH_2Cl_2/EtOAc, 3:2) gave **21** as a colorless oil (1.18 g, 95%); ^{1}H NMR (400 MHz, $CDCl_3$): δ 6.65 (d, 1H, C*H*=CH, J = 8.0 Hz), 6.51 (d, 1H, CH=C*H*, J = 8.0 Hz), 4.74-4.72 (m, 4H, 2CH_2), 4.54-4.49 (m, 2H, CH_2), 4.26-

4.03 (m, 6H, 3CH$_2$), 3.39 (s, 6H, 2CH$_3$), 2.34-3.30 (m, 1H, C*H*HCH$_2$), 1.88-1.81 (m, 2H,

C*H$_2$*CH$_2$), 1.35-1.28 (m, 1H, CH*H*CH$_2$), 1.26 (t, 3H, CH$_3$, *J* = 6.8 Hz), 1.19 (t, 3H, CH$_3$,

J = 7.2 Hz); ^{13}C NMR (100 MHz, CDCl$_3$): δ 158.9 (*C*O$_2$Et), 157.7 (*C*O$_2$Et), 137.7(*C*=C)),

134.8 (C=*C*), 96.6(O*C*H$_2$O), 96.4(O*C*H$_2$O), 69.8, 69.3, 62.8, 61.9, 61.6, 61.2, 55.2, 55.1,

28.1(*C*H$_2$CH$_2$), 25.0(*C*H$_2$CH$_2$), 14.3(*C*H$_3$CH$_2$), 14.2(*C*H$_3$CH$_2$); HRMS-CI calcd for

C$_{18}$H$_{31}$N$_2$O$_8$ [M+H]$^+$ 403.2080, found 403.2082.

cis-(4-Ethoxycarbonylamino-1,4-bis-methoxymethoxymethyl-cyclohexyl)-carbamic

acid ethyl ester (22). Under dry Ar, 10 % Pd/C (0.15 g) catalyst was added slowly into a

solution of **21** (0.55 g, 1.37 mmol) in CH$_2$Cl$_2$ (30 mL). The solution was degassed, placed

under hydrogen atmosphere (10 psi) and stirred for 24 h. The reaction mixture was

filtered through celite on a fine fritted glass filter and the activated carbon was washed

with additional EtOAc. The filtrate was evaporated *in vacuo* to dryness then redissolved

in THF (40 mL). The THF solution was then transferred to a three neck flask into which

NH$_3$ was bubbled at -70 °C until approximately 150 mL were condensed. Excess Na

metal was added and the solution turned dark blue. The resulting mixture was stirred at -

70 °C for 1 h. The reaction was quenched by carefully adding solid NH$_4$Cl, and NH$_3$ was

allowed to evaporate slowly. The residue was diluted with EtOAc (150 mL) and filtered,

and the solvents were evaporated *in vacuo*. The resulting crude product was purified by

flash chromatography over silica gel (CH$_2$Cl$_2$/EtOAc 5:1) to give **22** as a colorless paste

(0.52 g, 95%); ^{1}H NMR (300 MHz, CDCl$_3$): δ 4.70 (s, 2H, 2NHCO), 4.58 (s, 4H, 2CH$_2$),

4.03 (q, 4H, 2CH$_2$, *J* = 7.2 Hz), 3.66 (s, 4H, 2CH$_2$), 3.32 (s, 6H, 2CH$_3$), 2.02-1.94 (m,

4H, equatorial), 1.70-1.64 (m, 4H, axial), 1.20 (t, 6H, 2CH$_3$, J = 7.2 Hz); ^{13}C NMR (75.5 MHz, CDCl$_3$): δ 155.1 ($CO_2C_2H_5$), 96.6 (OCH_2O), 69.4(OCH_2C), 60.2(OCH_2CH$_3$), 55.2 (OCH_3), 54.3 (C1 and C4 of the ring), 27.3(C2, C3, C5 and C6 of the ring), 14.5 (CH_3CH$_2$); HRMS-CI calcd for C$_{18}$H$_{35}$N$_2$O$_8$ [M+H]$^+$ 407.2393, found 407.2388.

cis-(4-*tert*-Butoxycarbonylamino-1,4-bis-hydroxymethyl-cyclohexyl)-carbamic acid *tert*-butyl ester (23). Iodotrimethyl silane (TMSI, 0.58 mL, 4.3 mmol) was added dropwise into a solution of **22** (174 mg, 0.43 mmol) in CH$_2$Cl$_2$ (20 mL). The reaction mixture was heated under reflux for 18 h and quenched with MeOH (1.0 mL), leading to the formation of a fine precipitates. Vacuum filtration gave a yellowish solid (presumably an iodide salt of the fully deprotected dihydroxyl diamine). This highly hygroscopic solid was used directly to the next step without further purification. MeOH (10 mL), di-*tert*-butyl-dicarbonate (0.25 g, 1.14 mmol) and triethyl amine (0.24 mL, 1.71 mmol) were added to the above solid and the reaction mixture was stirred at room temperature overnight, then quenched with 40 mL saturated NaHCO$_3$ solution. The layers were separated and the aqueous layer was extracted with EtOAc (2×40 mL). The combined organic layers were washed with water, brine, dried (Na$_2$SO$_4$), filtered and concentrated *in vacuo*. Flash chromatograph on silica gel (CH$_2$Cl$_2$/MeOH 9:1) gave **23** as a colorless solid (119 mg, 74%); ^{1}H NMR (300 MHz, CDCl$_3$): δ 4.61(s, 2H, 2NHCO), 3.70 (s, 4H, 2CH_2OH), 1.72 (s, 8H, equatorial and axial), 1.44 (s, 18H, 2C(CH$_3$)$_3$); ^{13}C NMR (75.5 MHz, CDCl$_3$): δ 156.2(CO_2C(CH$_3$)$_3$), 80.2 (CH$_2$OH), 68.0 (C(CH$_3$)$_3$), 55.6 (C1 and C4),

28.3 (CH$_3$), 27.8 (C2, C3, C5 and C6); HRMS-CI calcd for C$_{18}$H$_{35}$N$_2$O$_6$ [M+H]$^+$ 375.2495, found 375.2498.

cis-1,4-Bis-*tert*-butoxycarbonylamino-cyclohexane-1,4-dicarboxylic acid (24). A solution of bis-alcohol **23** in *tert*-butyl alcohol (1mL/0.1 mmol of alcohol) was treated at room temperature with sodium hydroxide solution (4 equiv., 0.5 M in water) and KMnO4 solution (4 equivalents, 10% water) and the resulting mixture was stirred overnight. After quenching the reaction with an excess of aqueous sodium thiosulfate (5%), the mixture was washed with diethyl ether and the aqueous solution was acidified to a pH of 1-2 with 1M HCl at 0 °C. The mixture was extracted with ethyl acetate three times and the combined organic layers were dried (Na$_2$SO$_4$), filtered and concentrated *in vacuo* to yield product **24** as a colorless, sticky solid, that was used for the next step without further purification (93%); ^{1}H NMR (300 MHz, CD$_3$OD): δ 2.19-2.15 (m, 4H, equatorial), 1.89-1.84 (m, 4H, axial), 1.42 (s, 18 H, 2C(CH$_3$)$_3$); ^{13}C NMR (75.5 MHz, CDCl$_3$): δ 177.6 (CO$_2$H), 157.3 (CONH), 80.1 (*C*(CH$_3$)$_3$), 58.9 (C1 and C4), 30.1 (C2, C3, C5 and C6), 28.7 (CH$_3$); HRMS-CI calcd for C$_{18}$H$_{31}$N$_2$O$_8$ [M+H]$^+$ 403.2080, found 403.2075.

cis-1,4-Bis-*tert*-butoxycarbonylamino-cyclohexane-1,4-dicarboxylic acid dimethyl ester (25). Ester **25** was prepared from **24** by following the procedure described for the synthesis of **5** in 95% yield; ^{1}H NMR (300 MHz, CDCl$_3$): δ 4.83 (s, 2H, 2NH), 3.73 (s, 6H, 2CH$_3$), 2.26-2.21 (m, 4H, equatorial), 1.88-1.82 (m, 4H, axial), 1.43 (s, 18H, 2C(CH$_3$)$_3$); ^{13}C NMR (75.5 MHz, CDCl$_3$): δ 173.9 (*C*O$_2$CH$_3$), 154.8 (CONH), 80.1

($C(CH_3)_3$), 57.7 (C1 and C4), 52.3 (OCH_3), 31.4 (C2, C3, C5 and C6), 28.2 (CH_3); HRMS-CI calcd for $C_{20}H_{35}N_2O_8$ $[M+H]^+$ 431.2393, found 431.2378.

cis-1,4-Bis-((N-benzoyl(thiocarbamoyl)amino)-1,4-dimethoxycarbonyl-cyclohexane (27). The ester **25** (0.7 g, 1.63 mmol) was treated with 35 mL CH_2Cl_2 / TFA (3/7) at 0 °C. The reaction mixture was stirred at room temperature for 4h. The volatiles were removed under vacuum and the residue was kept in a dry ice/acetone cold bath, to which a diluted TEA (5 mL) in CH_2Cl_2 (10 mL) solution was added slowly (**caution**: fumes formed). After 10 min of stirring, the cold bath was removed and the reaction mixture was warmed to room temperature. The volatiles were all removed *in vacuo*, giving **26** as a colorless oil, which was redissolved in dry CH_2Cl_2 (15 mL). Benzoyl isothiocyanate (0.48 mL, 3.58 mmol) was added by syringe. The clear solution was stirred for 12 h then concentrated *in vacuo*. Flash chromatograph on silica gel (9/1 CH_2Cl_2/EtOAc) gave **27** as a yellowish oil (0.70 g, 77%); ^{1}H NMR (300 MHz, $CDCl_3$): δ 11.15 (s, 2H, 2NHCO), 8.93 (s, 2H, 2NHC=S), 7.84-7.82 (m, 4H, aromatic), 7.63-7.61 (m, 2H, aromatic), 7.54-7.50 (m, 4H, aromatic), 3.79 (s, 6H, 2CH_3), 2.70-2.65 (m, 4H, equatorial), 2.27-2.22 (m, 4H, axial); ^{13}C NMR (75.5 MHz, $CDCl_3$): δ 180.8 (C=S), 173.0 (CO_2Me), 168.0 (CONH), 134.3, 132.9, 128.8 and 128.6 (aromatic), 70.2 ($CH_2C_6H_5$), 60.2 (C1 and C4), 52.5 (CH_3), 26.1 (C2, C3 C5 and C6); HRMS-CI calcd for $C_{26}H_{29}N_4O_6S_2$ $[M+H]^+$ 557.1529, found 557.1544.

cis-2,10-Dithioxo-1,3,9,11-tetraaza-dispiro[4.2.4.2]tetradecane-4,12-dione (28). A mixture of **27** (0.55 g, 0.99 mmol) and triethylamine (1.5 mL) in methanol (10 mL) was heated at reflux for 18 h. The solvent was evaporated; ethanol (10 mL) was added to the residue and the solution was adjusted to a pH of 5.0 with a 5% aqueous HCl. The precipitate was collected by filtration to give **28** as white solid (0.21 g, 75 %); ^{1}H NMR (300 MHz, DMSO-d_6): δ 11.7 (bs, 2H, 2NH(C=O)(C=S)), 10.4 (s, 2H, 2NHC=S), 2.09-2.05 (m, 4H, equatorial), 1.79-1.74 (m, 4H, axial); ^{13}C NMR (75.5 MHz, DMSO-d_6): δ 182.1 (C=S), 178.2 (C=O), 63.5 (C1 and C4), 27.9 (C2, C3, C5 and C6); HRMS-CI calcd for $C_{10}H_{13}N_4O_2S_2$[M+ H]$^+$: 285.0480, found 285.0478.

cis-2,10-Dihydrazino-1,3,9,11-tetraaza-dispiro[4.2.4.2]tetradeca-1,9-diene-4,12-dione (30). Compound **30** was prepared from **28** by following the procedure described for the synthesis of **10** in 72% overall yield in two steps; ^{1}H NMR (300 MHz, DMSO-d_6): δ 9.03 (bs, 2H, 2NH), 7.29 (s, 2H, 2NH(NH$_2$), 4.46 (s, 4H, 2NH$_2$), 1.97-1.93 (m, 4H, equatorial), 1.62-1.57 (m, 4H, axial); ^{13}C NMR (75.5 MHz, DMSO-d_6 + TFA): δ 177.8 (C=O), 158.6 (NH$_2$-C=N), 64.3 (C1 and C4), 29.0 (C2, C3, C5 and C6); HRMS-CI calcd for $C_{10}H_{17}N_8O_2$[M+ H]$^+$: 281.1474, found 281.1482.

Bisintercalater D1. The linker **10** (0.05 mg, 0.18 mmol) was added into a DMF solution (15 mL) containing naphthalenetetracarboxylic diimide **31** (0.26 g, 0.54 mmol)(Lokey, 1997), PyBOP (0.279 g, 0.54 mmol), HOBT (0.082 g, 0.54 mmol) and DIPEA (0.19 mL, 1.1 mmol). The reaction mixture was stirred at room temperature for 20 h then filtered,

and the crude solid product was rinsed with DMF (2×10 mL) then ethanol (2×10 mL). The solid was dried *in vacuo*, then redissolved in 1:1 TFA/CH_2Cl_2 (20 mL). The solvent was removed under vacuum after 15 min of stirring. The resulting pink solid was dissolved in 0.1% TFA/H_2O and purified by reverse-phase preparative HPLC (Amersham, AKTA Purifier, Hi-Pore RP-318 250×10 mm column) using 0.1% TFA/water as solvent A and 0.1% TFA/ACN as solvent B. The product fractions were combined and lyophilized to yield the product **D1** (0.12 g, 70%) as a light orange solid; ^{1}H NMR (500 MHz, D_2O): δ 8.60-8.56 (m, 8H), 4.68 (bs, 4H), 4.41 (t, 4H, J = 6.5 Hz), 3.35 (bs, 4H), 2.75 (t, 4H, J = 6.5 Hz), 1.52 (bs, 8H); UV-vis: λ_{max} ($\log_{10}\varepsilon$) 383 (4.56), 363 nm (4.46); MALDI (Applied Biosystems 4700 Proteomics Analyzer, Foster City, CA) calcd for $C_{48}H_{43}N_{14}O_{12}$ [M + H][+]: 1007, found 1007.

Bisintercalater C1. The linker **30** (20.5 mg, 0.073 mmol) was added into a DMF solution containing naphthalenetetracarboxylic diimide **31** (106 mg, 0.22 mmol) (Lokey, 1997), PyBOP (114 mg, 0.22 mmol), HOBT (34 mg, 0.22 mmol) and DIPEA (0.08 mL, 0.44 mmol). The reaction mixture was stirred for 20 h at room temperature. Diethyl ether (20 mL) was added to the above clear solution at - 20 °C and a precipitate formed. Vacuum filtration gave a light orange solid that was then dissolved in 1:1 TFA/CH_2Cl_2 (20 mL). The solvents were removed under vacuum after 15 min of stirring at room temperature. The resulting solid was dissolved in 0.1 % aqueous TFA and purified following the same procedure as for **D1**. The product fractions were combined and lyophilized to yield the product **C1** (44 mg, 60%) as a light orange solid; ^{1}H NMR (500

MHz, D$_2$O): δ 8.41 (d, 4H, aromatic Hs, $J = 6.0$ Hz), 8.30 (d, 4H, aromatic Hs, $J = 6.0$ Hz), 4.68 (bs, 4H), 4.43 (t, 4H, $J = 6.1$ Hz), 3.35 (t, 4H, $J = 6.1$ Hz), 2.65 (bs, 4H) , 2.29 (m, 4H), 1.97(m, 4H); UV-vis: λ_{max} (log$_{10}\varepsilon$) 384 (4.42), 364 nm (4.43); MALDI (Applied Biosystems 4700 Proteomics Analyzer, Foster City, CA) calcd for C$_{48}$H$_{43}$N$_{14}$O$_{12}$ [M + H]$^+$: 1007, found 1007.

2.5.2 Dissociation kinetics

In a typical 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 first incubated with 500 µl of 0.04 mM compound stock solution for 2 hr. Next, 0.5 mL of the resulting DNA/compound complex solution was mixed with 0.5 mL of 4% SDS solution in a quartz cuvette with 1 cm path length. The UV absorbance was monitored at 383-385 nm (depending on the linker structure of the dimer) after 20 seconds of mixing (Hewlet-Packard 8553 Multitransport UV-vis spectrophotometer). The error estimate was determined to be ± 8% for this experiment.

2.5.3 DNAse I Footprinting

Plasmid pBR322 (New England BioLabs, Ipswich, MA) was digested with *Nhe*I, dephosphorylated with CIAP, 5'-^{32}P-end labeled with [γ-^{32}P]-ATP and T4 kinase, digested with *Eco*RI (all enzymes were purchased from New England Biolabs) and purified by native polyacrylamide gel electrophoresis (PAGE), following standard protocols (Sambrook and Russell, 2001). The 92 bp synthetic fragment (PAGE grade) was purchased from Midland Certified (Midland, TX) and labeled (^{32}P) similarly. The

DNAse I (Amersham Biosciences, Piscataway, NJ) footprinting was carried out according to the procedure described previously (Lee, 2004(b)). The DNA fragments were separated on an 8% (231 bp) or 12% (92 bp) denaturing polyacrylamide gel. The gels were exposed on a phosphor screen and analyzed with Quantity One 4.5 software from Bio-Rad (Hercules, CA).

2.5.4 Computer simulation

Molecular dynamics and geometry optimization computations were performed in HyperChem 7.0 (Hypercube Inc., 1115 NW 4th Street, Gainesville, FL 32601) using the Amber force field. PDB coordinates from the **G₃K**-d(CGGTACC)₂ were used and the linker segment was modified as needed (Guelev, 2001(a)). During the calculation, DNA coordinates were fixed and molecular dynamics were performed on the dimer to anneal the system to obtain a lower energy minimum. Initial structures were subjected to 15 picoseconds of molecular dynamics at 1000 K to allow high degree randomization of the initial model, and then the temperature was slowly lowered to 300 K over 10 picoseconds. After annealing a final geometry optimization was performed using the Fletcher-Reeves conjugate gradient algorithm, with a convergence cutoff value of 0.01 kcal·mol^{-1}.

2.5.5 X-ray single crystallographic analysis

Crystallographic summary for **5** ($C_{26}H_{30}N_2O_8$). Colorless prismatic crystals, monoclinic, $P2_1/c$ (No. 14), Z = 2 in a cell of dimensions: a = 12.7888(9), b = 10.5763(7), c = 9.7425(5)Å, β = 105.064(4)°, V = 1272.47(14)Å^3, ρ_{calc} = 1.30 g·cm^{-3}, μ = 0.097 mm^{-1},

F(000) = 528. A total of 4024 reflections were measured, 2200 unique (R_{int} = 0.056), on a Nonus Kappa CCD using graphite monochromatized Mo Kα radiation (λ = 0.71073Å) at -120 ˚C. The structure was refined on F^2 to an R_w = 0.254, with a conventional R = 0.100 (1451 reflections with F_o > 4[$\sigma(F_o)$]), and a goodness of fit = 1.489 for 198 refined parameters.

Crystallographic summary for **8** ($C_{10}H_{12}N_4O_2S_2$ – $2C_2H_6SO$). Colorless prismatic crystals, monoclinic, P2$_1$/n (No. 14), Z = 2 in a cell of dimensions: a = 8.0735(1), b = 9.2155(2), c = 13.6410(3)Å, β = 99.819(1)°, V = 1000.04(3)Å^3, ρ_{calc} = 1.46 g-cm^{-3}, μ = 0.502 mm^{-1}, F(000) = 464. A total of 4213 reflections were measured, 2275 unique (R_{int} = 0.028), on a Nonus Kappa CCD using graphite monochromatized Mo Kα radiation (λ = 0.71073Å) at -120 ˚C. The structure was refined on F^2 to an R_w = 0.0769, with a conventional R = 0.0349 (1765 reflections with F_o > 4[$\sigma(F_o)$]), and a goodness of fit = 1.010 for 167 refined parameters.

Crystallographic summary for **12a** (an analogue of **12**) ($C_{17}H_{20}N_2O_4$). Colorless needles, triclinic, P-1 (No. 2), Z = 2 in a cell of dimensions: a = 5.6474(2), b = 11.2341(4), c = 12.4607(4)Å, α = 99.586(2), β = 97.391(2), γ = 91.882(2)°, V = 771.84(5)Å^3, ρ_{calc} = 1.36 g-cm^{-3}, μ = 0.098 mm^{-1}, F(000) = 336. A total of 5249 reflections were measured, 3497 unique (R_{int} = 0.026), on a Nonus Kappa CCD using graphite monochromatized Mo Kα radiation (λ = 0.71073Å) at -120 ˚C. The structure was refined on F^2 to an R_w =

0.108, with a conventional R = 0.0431 (2188 reflections with $F_o > 4[\sigma(F_o)]$), and a goodness of fit = 1.006 for 289 refined parameters.

Crystallographic summary for **28** $((C_{10}H_{11}N_4S_2O_2)^{1-} (C_6H_{16}N)^{1+})$. Colorless lathes, monoclinic, P2$_1$/c (No. 14), Z = 8 in a cell of dimensions: a = 12.2470(3), b = 11.9535(3), c = 27.0891(7)Å, β = 96.674(1)°, V = 3938.82(17)Å^3, ρ_{calc} = 1.30 g-cm^{-3}, μ = 0.290 mm^{-1}, F(000) = 1648. A total of 11100 reflections were measured, 6658 unique (R_{int} = 0.070), on a Nonus Kappa CCD using graphite monochromatized Mo Kα radiation (λ = 0.71073Å) at -120 °C. The structure was refined on F^2 to an R_w = 0.140, with a conventional R = 0.0617 (3580 reflections with $F_o > 4[\sigma(F_o)]$), and a goodness of fit = 1.168 for 460 refined parameters.

Chapter 3   Structural Characterization of a Rigidified Threading Bisintercalator

3.1 Chapter summary

3.1.1 Goals

To explore the sequence-specific interaction of DNA with a new threading bisintercalator (**C1**) consisting of two intercalating 1,4,5,8-naphthalenetetracarboxylic diimide (NDI) units connected by a rigid, tricyclic spiro linker.

3.1.2 Approach

NMR spectroscopy was used to develop a structural model of **C1** complexed to d(CGGTACCG)$_2$, using distance constraints. Residual dipolar coupling (RDC) techniques were also used to confirm the structural model derived primarily from NOESY.

3.1.3 Results

According to the derived model, the central cyclohexane ring of the linker connecting the two NDI units of **C1** lie flat in the minor groove of DNA. Linker length, hydrogen bonding, steric, and hydrophobic factors all appear to contribute to the observed sequence specificity of binding. These results serve to illustrate the versatility of threading polyintercalation given that in a previous study, a ligand consisting of two NDI units joined by a flexible peptide linker was shown to bind sequence specifically

within the major groove of this same sequence of DNA. As such, this is the first example of a single approach to target the same DNA sequence in both the minor and major grooves.

3.2 Introduction

The bisintercalator **C1** with a novel, rigid linker has been synthesized and found to bind poly(dGdC) with high binding affinity (Figure 3.1) (Chu, 2006). DNAse I footprinting suggested that **C1** binds GGNNCC sequences (see Chapter 2). Considering that **C1** has 2-fold symmetry, a non-natural 92 base pair DNA fragment composed of different combinations of GGNNCC sequences was subjected to footprinting, with results suggesting that **C1** prefers the sequence GGTACC ($K_D \sim 10^{-7}$ M), although interactions with GGGCCC, GGATCC, and GGCGCC were also seen. There was no apparent binding to other sequences of the DNA used in the footprinting experiments (Mazzitelli, 2007).

In this chapter, we report the results of experiments directed toward the structural analysis of the interactions of **C1** with d(CGGTACCG)$_2$ and present a model of the **C1**-d(CGGTACCG)$_2$ complex.
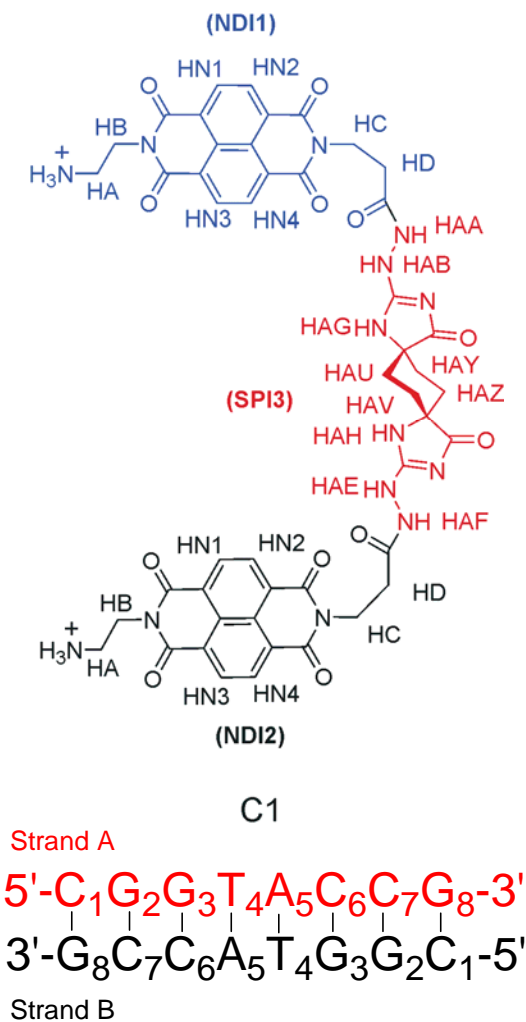
Figure 3.1 Chemical structure and naming scheme of ligand **C1** and the oligonucleotide sequence used in the NMR studies.

3.3 Results

Titration of d(CGGTACCG)$_2$ with bisintercalator C1. A titration of d(CGGTACCG)$_2$ with **C1** was monitored using 1D ^{1}H NMR spectroscopy (Figure 3.2). When 0.4 equivalents of **C1** was added to the free d(CGGTACCG)$_2$ in solution, many new peaks appeared. The peaks of the free d(CGGTACCG)$_2$ remained unchanged in both

position and shape, but intensity decreased. These observations indicate a relatively tight and specific binding between **C1** and d(CGGTACCG)$_2$, with disassociation process being slow on the NMR time scale. After adding 1.0 equivalent of ligand **C1**, the resonances of the free d(CGGTACCG)$_2$ disappeared entirely, and a new set of resonances emerged from the upfield shifted signals of the complex.
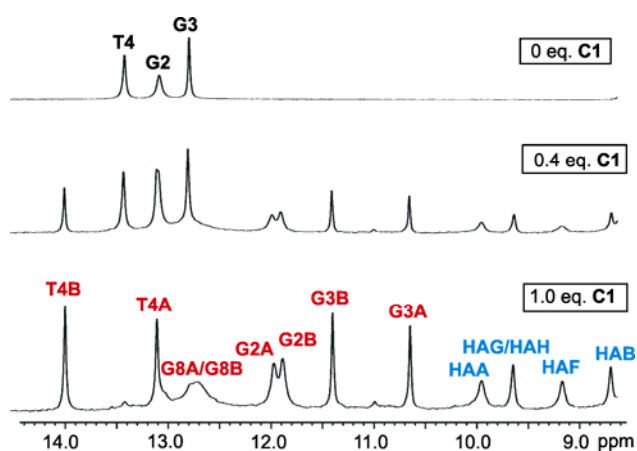
Figure 3.2 1D ^{1}H NMR spectra of **C1** titration into d(CGGTACCG)$_2$ in H$_2$O/D$_2$O (9:1) with 30 mM Na phosphate buffer at 25 °C.

Interestingly, the number of resonances in the **C1**-d(CGGTACCG)$_2$ complex significantly increased compared to that of the free d(CGGTACCG)$_2$, suggesting that the 2-fold symmetry of the d(CGGTACCG)$_2$ is not retained in the complex. The most notable chemical shift differences between free DNA and ligand-DNA complex are observed for the central TA base pairs, as well as for the G3-C6 base pair. The imino proton region of the 1D spectrum suggests binding between **C1** and d(CGGTACCG)$_2$ with a 1:1 stoichiometry. Upfield shifting of the imino resonances of base pairs G2-C7

and G3-C6 are consistent with intercalation at these sites (Feigon, 1984). Resonances in the 8.5-10.0 ppm range were also observed from the exchangeable protons of **C1** upon binding (9.9, 9.6, 9.2 and 8.6 ppm, respectively).

The loss of 2-fold symmetry for **C1**-d(CGGTACCG)$_2$ complex is unexpected, since both **C1** and d(CGGTACCG)$_2$ have 2-fold symmetry prior to mixing together. It may indicate that **C1** does not have 2-fold symmetry when in a conformation that would bind to DNA. However, this phenomenon is not unprecedented, being observed in a study of a WP652-d(TGTACA)$_2$ complex, where both the ligand and d(TGTACA)$_2$ have individual 2-fold symmetry yet the complex does not (Robinson, 1997).

Incubation of d(GGATCC)$_2$, d(GGGCCC)$_2$, and d(GGCGCC)$_2$ with bisintercalator C1. Consistent with the footprinting results showing lower levels of binding affinity to these sequences (Mazzitelli, 2007), incubation of d(GGATCC)$_2$, d(GGCGCC)$_2$, and d(GGGCCC)$_2$ with 1.0 equivalent of **C1** revealed no single distinct complex in any case. There were too many peaks present to represent a single bound species. For example, in the titration using the d(GGGCCC)$_2$ (Figure 3.3), it appeared that there were two different complexes at a ratio of 3:7 formed, preventing any reliable peak assignments. These complexes were not investigated further.
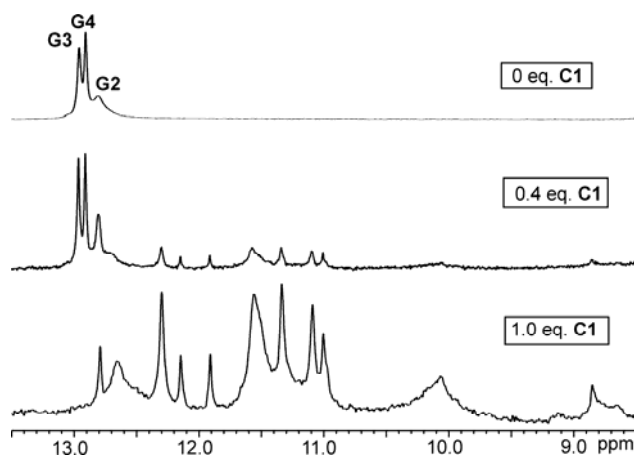
Figure 3.3 1D ^{1}H NMR spectra of **C1** titration into d(AGGGCCCT)$_2$ in H$_2$O/D$_2$O (9:1) with 30 mM Na phosphate buffer at 25 °C. No single distinct complex could be identified.

Assignment of the C1-d(CGGTACCG)$_2$ complex. The signals of the **C1**-d(CGGTACCG)$_2$ complex were assigned by following the H6/8-H1' and H6/8-H2'/H2" NOE connectivities (Figure 3.4 and Table 3.1) (Hare, 1983; Wüthrich, 1986). The inter-nucleotide NOE connectivities for **C1**-d(CGGTACCG)$_2$ complex are weak or interrupted at the steps G2-G3 and C6-C7 (Figures 3.5 and 3.6). A TOCSY spectrum showed four strong cross peaks in the aromatic-to-aromatic region, arising from the aromatic protons of two NDI units.
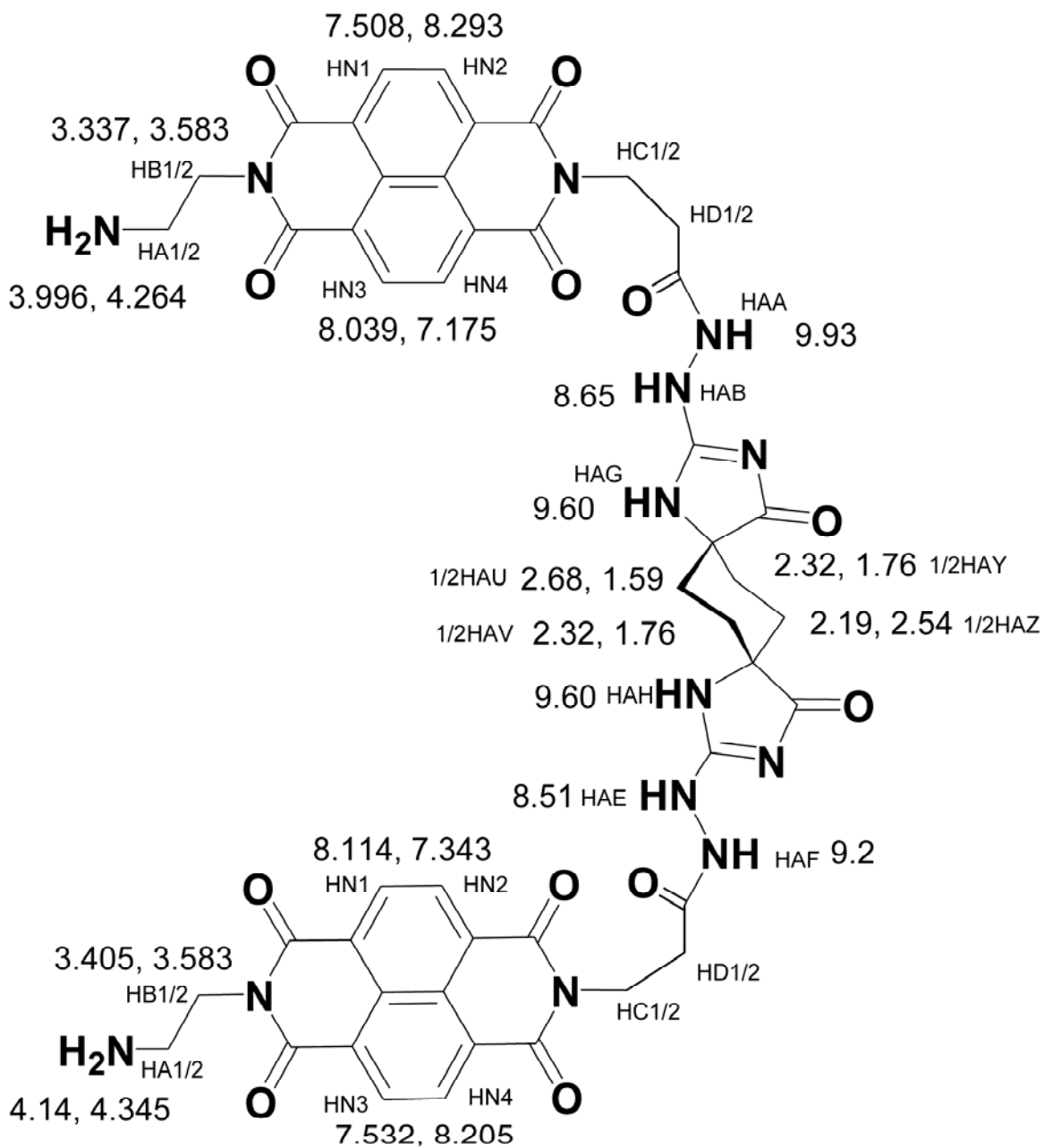
Figure 3.4 Chemical shifts of ligand **C1** protons which have NOE contacts with DNA bases.

Table 3.1 Chemical shifts (ppm) of the DNA protons in the complex **C1**-d(CGGTACCG)$_2$, 30 mM Na phosphate, pH 7.5, 25 °C (15 °C for imino and amino protons). The spectra were referenced to the solvent HOD peak.

DNA base	C1A	G2A	G3A	T4A	A5A	C6A	C7A	G8A
H1'	5.55	5.62	5.45	5.31	5.94	5.88	5.59	6.14
H2'	1.42	2.63	2.65	2.13	2.66	2.39	1.92	2.63
H2''	2.07	2.73	2.53	2.36	2.70	2.39	2.28	2.36
H3'	4.52	4.99	4.88	4.84	5.00	4.94	4.75	4.67
H4'	3.91	4.23	4.42	4.18	4.27	4.06	4.15	4.17
H6/H8	7.36	7.81	7.92	7.28	8.21	7.39	7.69	7.93
H5/H2/HM	5.71			1.32	7.27	5.38	5.96	
NH$_2$	6.94/8.16	5.42/n.a.	5.00/7.20		n.a.	6.53/6.90	7.19/7.75	n.a.
HN		11.97	10.62	13.16				n.a.

DNA base	C1B	G2B	G3B	T4B	A5B	C6B	C7B	G8B
H1'	5.55	5.51	5.78	4.97	6.24	5.84	5.50	6.14
H2'	1.42	2.63	2.55	2.08	2.73	2.43	2.00	2.63
H2''	2.07	2.84	2.79	2.18	3.00	2.69	2.28	2.36
H3'	4.52	5.00	4.86	4.74	5.04	5.00	4.74	4.67
H4'	3.91	4.24	4.52	4.06	4.38	4.10	4.25	4.17
H6/H8	7.36	7.82	7.81	7.13	8.16	7.65	7.76	7.93
H5/H2/HM	5.71			1.32	7.51	5.77	5.96	
NH$_2$	6.94/8.16	5.42/8.08	5.31/8.51		6.17/7.29	7.18/7.50	7.13/7.87	n.a.
HN		11.86	11.38	14.04				n.a.
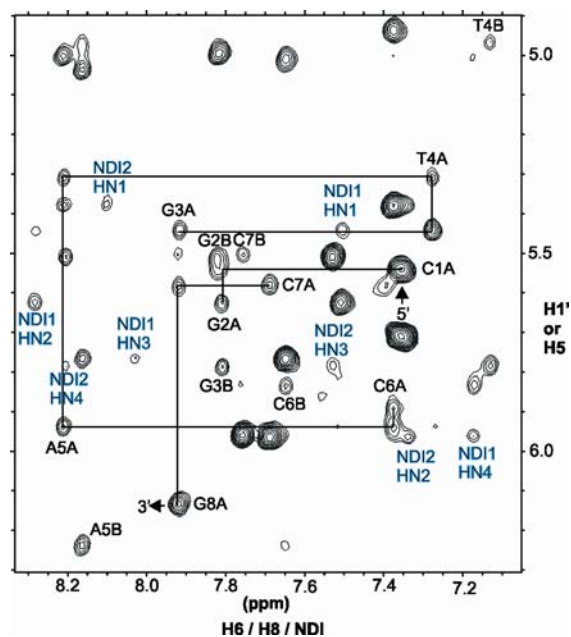
(a)
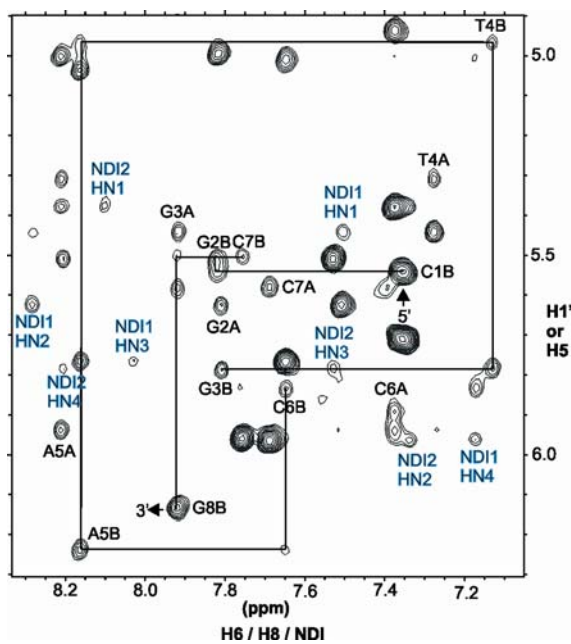
(b)

Figure 3.5 Contour plot of the 2D NOESY spectrum (D$_2$O, 60 ms mixing time, 25 °C) of the **C1**-d(CGGTACCG)$_2$ complex showing the DNA H6/H8 to H1' connectivities. (a) strand A; (b) strand B. Note that the connectivities are interrupted at the G2/G3 and C6/C7 steps.

91
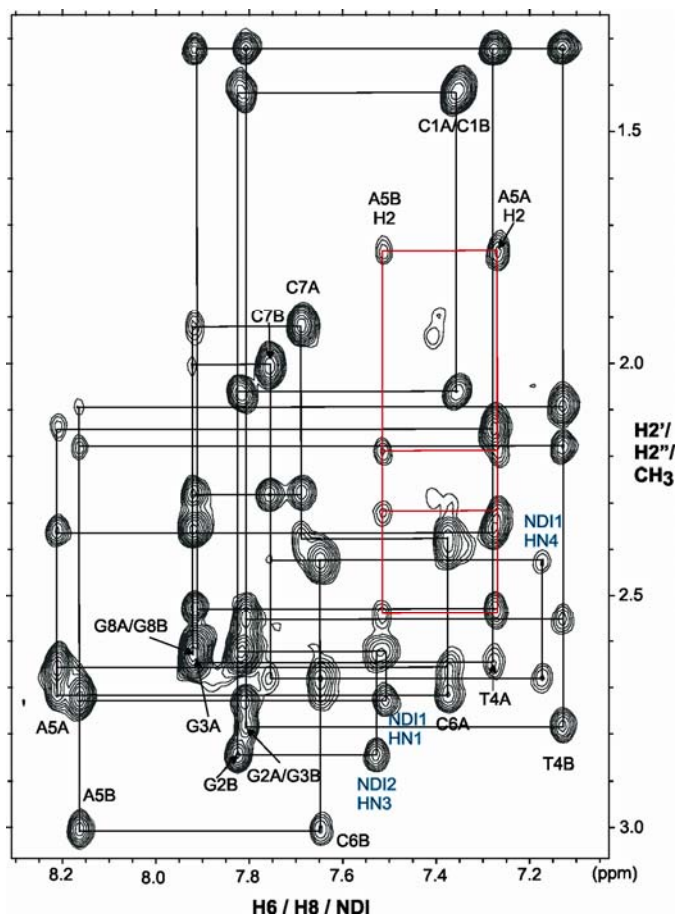
Figure 3.6 Contour plot of the NOESY spectrum (D$_2$O, 60 ms mixing time, 25 °C) of the **C1**-d(CGGTACCG)$_2$ complex showing the DNA H6/H8 to H2'/H2"/CH$_3$ connectivities (black lines) and intermolecular NOEs between the ligand linker and H2 of the adenine residues (red lines).

These assignments were further supported by the NOE cross peaks observed between protons from NDI and d(CGGTACCG)$_2$ (Figure 3.7). The four aromatic protons on NDI1 had NOEs with G2A, G3A, C6B and C7B, while four aromatic protons on NDI2 were observed to have NOEs at C6A, C7A, G2B and G3B. Cross peaks in the NOE spectrum were observed between the methylene protons on the cyclohexane ring and H2 of A5A and A5B, as well as H1' and H4' of A5A and A5B (Figures 3.6 and 3.8).

92

NOEs were also seen between the exchangeable protons on the linker of **C1** and H1' and H2 of A5A and A5B, H1' of C6B, and weak NOE cross peaks were observed between the methylene protons at the two termini of **C1** and H8 of G2 for both strands A and B (Figure 3.8).
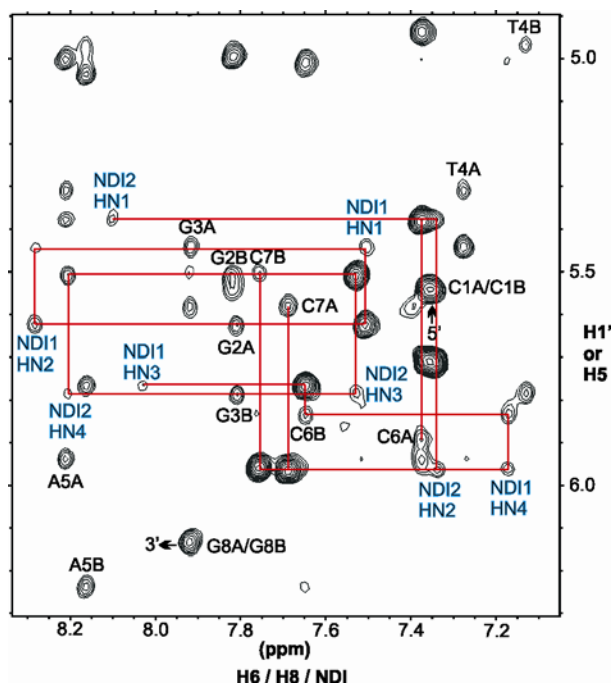
Figure 3.7 Intermolecular NOEs between protons from the ligand (2 NDI units) and d(CGGTACCG)$_2$ (H1' and H5) at the intercalation sites (D$_2$O, 60 ms mixing time, 25 °C).
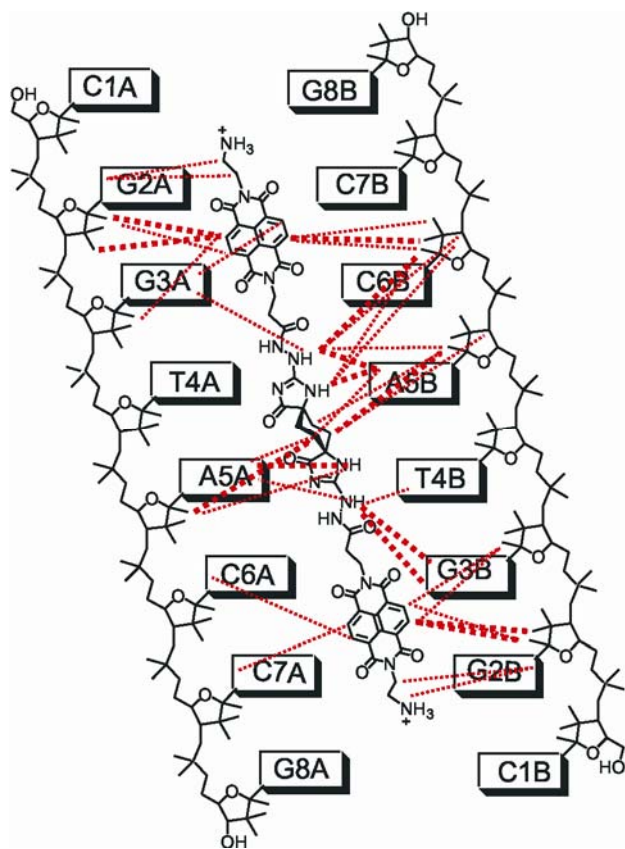
Figure 3.8 Diagram of the observed intermolecular NOEs in the **C1**-d(CGGTACCG)$_2$ complex. The thick dashed lines indicate 'strong' and 'medium' NOEs; thin dotted lines represent weak NOEs.

The NOE data are consistent with a model in which the NDI units are intercalated into the GpG step in a threading manner, with the tricyclic spiro linker located in the minor groove and both termini protruding into the major groove. Having a linker residing in the minor groove turns out to be opposite to that reported previously with a different threading bisintercalator, **G$_3$K**, which binds in the major groove of the same d(CGGTACCG)$_2$ sequence (Guelev, 2001(a)).

The conformations of the ribose rings were determined from a qualitative analysis of the intensities of intra-nucleotide cross peaks in the NOE spectrum (Table 3.2)

(Wijmenga, 1993). Data indicate that all nucleotides except C1A, C1B, G2A and C6A are in an S-type (C2'-endo sugar pucker) conformation, as is typical for B-form DNA, and consistent with the x-ray crystal structures of other intercalators complexed with d(CGTACG)$_2$ (Wang, 1987; Adams, 1999;). The NMR data support nucleotides C1A, C1B, G2A and C6A having an N-type (C3'-endo sugar pucker) conformation.

Table 3.2 Qualitative analysis of DNA sugar conformation from NOE intensities observed in NOESY spectrum in D$_2$O (mixing time 60 ms, pH 7.5, 25 °C, S = strong, M = medium, W = weak).

DNA bases	H3'-H6/8	H2"-H6/8(n+1)	H2'-H6/8(n+1)	H2"-H4'	H1'-H4	Sugar Pucker
C1A	**M**	**M**	**M**	**S**	**S**	**North**
G2A	**Overlap**	**n.a.**	**n.a.**	**M**	**M**	**North**
G3A	W	M	W	W	M	South
T4A	W	M	W	W	W	South
A5A	W	M	M	Overlap	W	South
C6A	**M**	**W**	**W**	**M**	**M**	**North**
C7A	W	M	W	W	W	South
G8A	W			M	M	South
C1B	**M**	**M**	**M**	**S**	**S**	**North**
G2B	W	n.a.	n.a.	M	M	South
G3B	W	M	W	W	W	South
T4B	n.a.	M	W	W	n.a.	South
A5B	W	M	W	W	W	South
C6B	W	W	W	M	M	South
C7B	n.a.	M	W	W	n.a.	South
G8B	W			M	M	South

Ring flipping. It has been recently reported that in a structurally related naphthalimide bisintercalator LU 79553 ligand-DNA complex, the imide chromophores exhibit rotational ring flipping within the intercalation site (TpG step) on the millisecond time scale (Gallego and Reid, 1999). The ring flipping was ascribed to the electrostatic

repulsion in the major groove, thought to be present between the imide oxygens and the N7 / O6 electron-rich edge of guanine (Gallego and Reid, 1999). A possible relationship between the ring flipping rate and the antitumor activity of a given intercalator was proposed (Gallego and Reid, 1999).
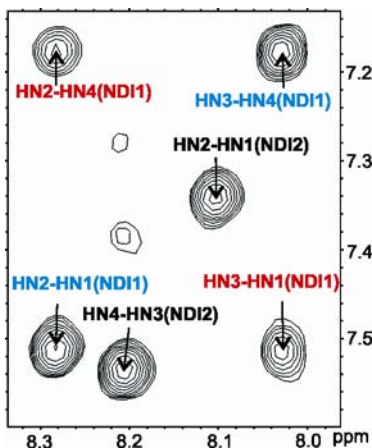
Figure 3.9 Aromatic region of the 2D NOESY spectrum of the complex **C1**-d(CGGTACCG)$_2$ (25 °C, 60 ms mixing time).

As shown in the aromatic-aromatic region of the NOESY spectrum (Figure 3.9), NOE cross peaks were observed between HN1 and HN3, as well as HN2 and HN4 in NDI1. No such cross peaks were observed for aromatic protons in NDI2. The distance between NH1 and HN3 (and HN2 and HN4) is more than 6.9 Å (Lokey and Iverson, 1995), ruling out the possibility that these cross peaks were caused by dipolar relaxation. Similar unusual cross peaks have been observed for both NDI rings in a minor groove binding bisintercalator **(β-A)$_3$K** (Guelev, 2002). In this latter case, a combination of NOESY and ROESY data were used to support a model in which the aromatic NDI rings within GpA steps flipped on a millisecond time scale (Guelev, 2002). In the case of the

C1-d(CGGTACCG)$_2$ complex, it thus appears that the NDI1 ring is experiencing ring flipping on the NMR time scale, while the NDI2 ring is not.

Structure modeling. The structural analysis of the **C1**-d(CGGTACCG)$_2$ complex was carried out using the simulated annealing protocols within the CNS program suite (Tables 3.2 and 3.3) (Brünger, 1998). Twenty different starting structures were used for the simulated annealing, to ensure that a full range of structures consistent with the NMR data were identified. The position determined for the ligand was found to be similar in all of the derived structures. An ensemble of eight lowest-energy structures is shown in Figure 3.10(a), with the median structure shown separately in Figure 3.10(b). The average rmsd for this family is 1.10 Å for the ligand and d(CGGTACCG)$_2$, and 0.57 Å for the ligand and bases between two intercalators (G3 to C6). In all eight calculated structures, NDI1 stacks in the G2-G3 step, while NDI2 is located in the C6-C7 step. The distances between adjacent purine bases have been almost doubled in order to accommodate the NDI rings, however, the H-bonding network for the G-C base pairs remains intact. The tricyclic spiro linker appears to fit snugly in the minor groove. Interestingly, the central cyclohexane ring lies flat in the minor groove, occupying the TpA step and it appears locked in a single chair conformation.

Table 3.2 Intermolecular ligand-DNA NOEs (60 ms mixing time, 25 °C) used in the modeling.

	Ligand Residues	Proton	DNA Residue	Proton	Intensity
1	NDI1	HN1	G2A	H1'	M
2	NDI1	HN1	G2A	H2''	M
3	NDI1	HN1	G3A	H1'	W
4	NDI1	HN1	G3A	H8	VW*
!5	NDI1	HN1	G3A	H4'	VW*
6	NDI1	HN2	G2A	H1'	W
7	NDI1	HN2	G3A	H1'	VW
!8	NDI1	HN2	G3A	H4'	VW*
9	NDI1	HN3	C6B	H5	VW
10	NDI1	HN4	C7B	H5	VW
11	NDI1	HN4	C6B	H2'	W
12	NDI1	HN4	C6B	H2''	M
13	NDI1	HN4	C6B	H1'	W
14	NDI1	HN4	C6B	H6	VW
15	NDI2	HN1	C6A	H5	W
16	NDI2	HN2	C7A	H5	W
17	NDI2	HN3	G3B	H1'	W
18	NDI2	HN3	G3B	H8	W
19	NDI2	HN3	G2B	H2''	M
20	NDI2	HN3	G2B	H1'	M
21	NDI2	HN4	G3B	H1'	W
22	NDI2	HN4	G2B	H1'	W
23	NDI1	HA1/2	G2A	H8	W
24	NDI1	HB1/2	G2A	H8	W
25	NDI2	HA1/2	G2B	H8	W
26	NDI2	HB1/2	G2B	H8	VW
!27	SPI3	1/2HAU	A5B	H4'	W*
28	SPI3	1/2HAU	A5B	H1'	W
!29	SPI3	1/2HAV	A5B	H4'	W
30	SPI3	1/2HAV	A5B	H1'	M
31	SPI3	1/2HAZ	A5A	H1'	M
32	SPI3	1/2HAZ	A5A	H2	W
33	SPI3	1/2HAZ	A5B	H2	M
34	SPI3	HAA	C6B	H1'	VW
35	SPI3	HAA	G3A	H21/2	VW
36	SPI3	HAB	G3A	H21/2	W
37	SPI3	HAB	C6B	H2''	W
38	SPI3	HAB	C6B	H4'	W
39	SPI3	HAB	C6B	H1'	S
40	SPI3	HAB	A5B	H1'	W
41	SPI3	HAB	A5B	H2	M
42	SPI3	HAB	G3A	H1	W
43	SPI3	HAG	C6B	H4'	W
44	SPI3	HAG	A5B	H2''	VW
45	SPI3	HAG	C6B	H1'	W
46	SPI3	HAG	A5B	H1'	M
47	SPI3	HAH	A5A	H2	M
48	SPI3	HAG	A5B	H2	M
49	NDI1	HN3	G2A	H1	VW
50	SPI3	HAB	G3A	H21/2	W
51	NDI1	HN3	G3A	H1	W
52	NDI2	HN1	G3B	H1	W

* NOEs was observed from 200 ms mixing time. ! NOEs were not counted in the simulation.

Table 3.3 Statistics for an ensemble of eight lowest energy **C1**-d(CGGTACCG)$_2$ structures.

Restraints for the structural calculation	
total number of distance restraints	227
DNA-DNA NOEs	145
DNA-ligand NOEs	48
hydrogen bondings[a]	34
Statistics for the structural calculation	
number of NOE violations > 0.2 Å	1.0
number of NOE violations > 0.5 Å	0
rmsd to the mean structure (whole complex)	1.10 Å
rmsd to the mean structure (bases 3-6)	0.57 Å
rmsd for covalent bonds	0.0029 Å
rmsd for covalent angles	0.81°
rmsd for improper angles	0.41°

a. These restraints were obtained from the dna-rna_restraints.def file in the CNS program.

(a)   (b)   (c)
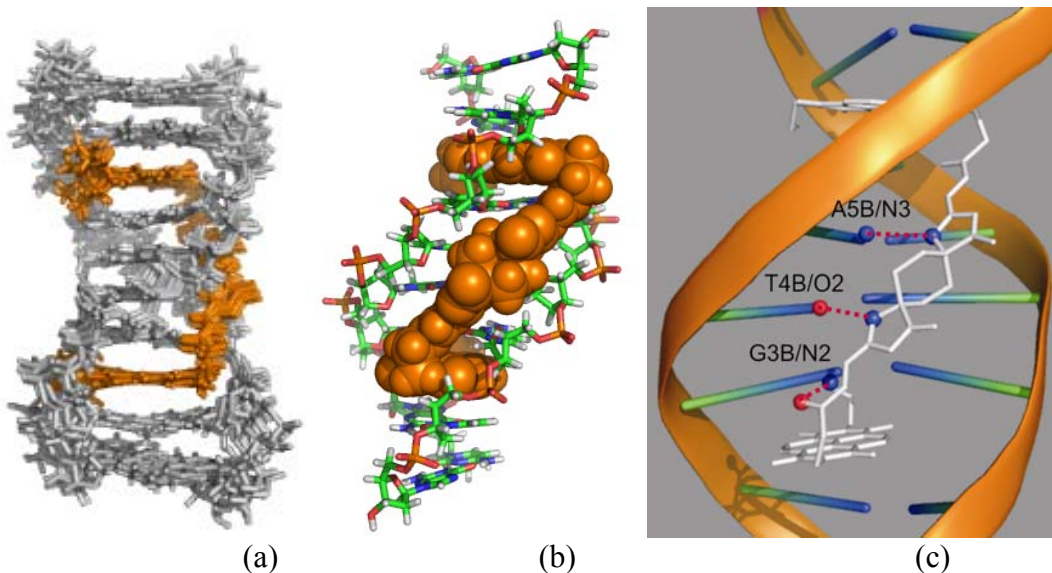
Figure 3.10 (a) Superposed structures of **C1**-d(CGGTACCG)$_2$ complex of the final 8 ensembles with the lowest energy; (b) space-filling model showing the linker in the minor groove; (c) proposed hydrogen bonds, top: N-HAG—N3/A5B; middle: N-HAH—O2/T4B; bottom: N2-H2(G3B)—O=C(NDI2). Hydrogen bonds were proposed when heavy atoms were between 2.5 and 3.4 Å. The figures were created in PyMol (DeLano, 2002).

There appear to be a few hydrogen bonds formed between the groups from the ligand and the edges of DNA bases in the ensemble of eight calculated lowest-energy structures (Figure 3.10c). The carbonyl group close to the NDI2 was found to form a strong hydrogen bond to the amino group of G3B (bond distance 2.8 Å, bond angle 164°). In addition, the orientation of amino N-HAG near the NDI1 unit and its large downfield shift of resonance upon binding suggest a hydrogen bond between N-HAG and N3 of A5B. However, this bond could be a weak one, as indicated by relatively long bond length (N-H···N close to 3.35 Å), although the N-H···N angle is favorable (close to 162°). In a related DNA-ligand complex, a similar NH-O6 hydrogen bond was observed for the corresponding ammonium group positioned two carbon atoms away from the imide ring (Gallego and Reid, 1999). Another weak hydrogen bond could be proposed between N-HAH and O2 of T4B (bond distance 3.4 Å, bond angle 161°).

Residual dipolar couplings in the C1-d(CGGTACCG)$_2$ complex. The analysis of residual dipolar couplings (RDCs) has recently emerged as an experimental NMR technique that can yield angular information that complements the inter-H atom distance information provided by the nuclear Overhauser effect. This promising new tool has been used as an aid in determining the global structures of helical nucleic acids (Lipsitz, 2004) and assessing the bend of DNA structures (MacDonald and Lu, 2002). In analyzing nucleic acid structures in solution, RDCs have been derived from ^{1}H-^{1}H, ^{15}N-^{1}H, ^{13}C-^{1}H and ^{15}N-^{13}C couplings (Sibille, 2001; Jaroniec, 2005; Boisbouvier, 2003). The

observation of RDCs requires that the tumbling of the molecule of interest be made slightly anisotropic, which can be achieved by adding an appropriate amount of an alignable liquid crystalline material such as Pf1 filamentous phage (Hansen, 1998). Pf1-phage has been described as inert to nucleic acid, merely serving as the alignment medium to create an anisotropic magnetic environment (Hansen, 1998). Residual dipolar couplings are sensitive to the angle between a vector connecting the coupled nuclei and a vector defining the average orientation of the molecule in the magnetic field. In the case of the **C1**-d(CGGTACCG)$_2$ complex, the rings of the two NDI intercalators are flat, so that vectors between pairs of protons HN1-HN2 and HN3-HN4 are restricted to be in the same plane. If the rings of NDI1 and NDI2 were parallel, and not rotated relative to each other, the HN1-HN2 and HN3-HN4 protons of each ring would be expected to exhibit the same dipolar couplings.

Addition of Pf1-phage to a sample of **C1**-d(CGGTACCG)$_2$ complex results in a change in the ^{1}H-^{1}H three bond coupling constants for the four pairs of aromatic H atoms on the two planar NDI units (Figure 3.11), verifying the presence of residual dipolar coupling. The observed RDCs for NDI1 are 4.4 Hz for each of HN1-HN2 and HN3-HN4, consistent with the vectors defined by HN1-HN2 and HN3-HN4 being parallel to each other. Smaller RDCs of 2.3 Hz were observed of the corresponding pairs of H atoms in NDI2. The observation of different RDCs for the two NDI units is consistent with the structural model derived from NOE data, where the aromatic rings of the two NDI groups are parallel but not fully overlapping with each other (Figure 3.12), however, it should be noted that this interpretation assumes that the rates of ring flipping do not differently

101

affect the observed RDCs in the two NDI units. Addition of phage does not result in a significant change in the chemical shifts recorded in 1D and 2D ^{1}H NMR spectra of the **C1**-d(CGGTACCG)$_2$ complex, showing that the structure is not significantly changed by the phage (compare Figures 3.2 to 3.13).
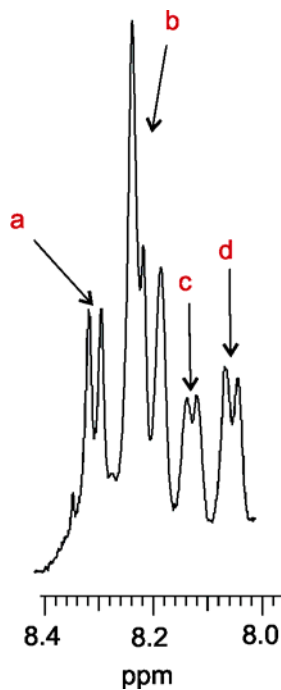
Figure 3.11 RDCs for 4 pairs of aromatic H atoms on two NDI units. Ordinary scalar coupling and RDC each contribute to the observed splitting of the peaks. The derived RDCs for the proton pairs are: a. HN1-HN2 (NDI1)--- 4.5 ± 0.3 Hz; b. HN3-HN4 (NDI2)---2.3 ± 0.3 Hz; c. HN1-HN2 (NDI2)---2.3 ± 0.3 Hz; d. HN3-HN4 (NDI1)---4.3 ± 0.3 Hz.
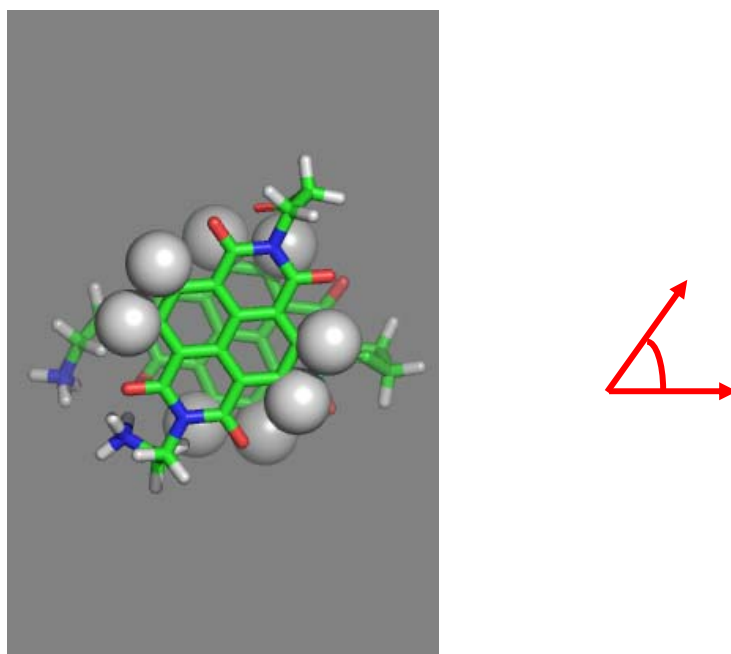
Figure 3.12 Skeletal drawing of two NDI units in the model indicating an angle of ~ 60° on overlap, viewed along the long axis of the complex **C1**-d(CGGTACCG)$_2$ (all other atoms were not shown). The figure was created in PyMol (DeLano, 2002).
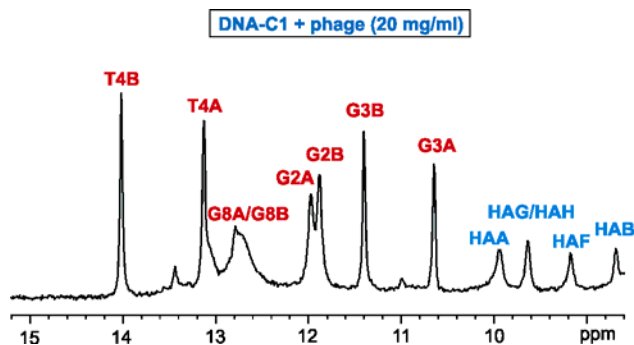
Figure 3.13 1D ^{1}H NMR spectrum of **C1**-d(CGGTACCG)$_2$ complex after adding Pf1 Phage (final concentration: ~ 20 mg/ml) in H$_2$O/D$_2$O (9:1), 30 mM Na phosphate buffer.

Through-space dipolar couplings were also observed between G2B/H1' and NDI2/HN3, and G2A/H1' and NDI1/HN1, which were detected as new peaks appearing in the 2QF-COSY spectrum after the addition of phage (Figure 3.14). These inter-nucleotide dipolar couplings are entirely consistent with the structural model of the **C1**-

103

d(CGGTACCG)$_2$ complex, where one NDI is located between G2 and G3, and the other NDI is between C6 and C7. A few other inter-nucleotide dipolar couplings were also identified, including C1A/H2"-G2A/H8 and G3B/H2"-T4B/H6 (not shown), also consistent with the structural model.
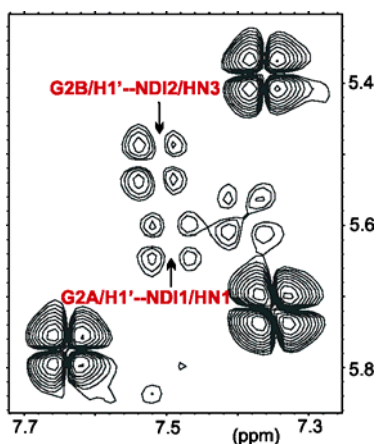
Figure 3.14 2QF-COSY spectrum of **C1**-d(CGGTACCG)$_2$ using Pf1-Phage as the co-solvent showing RDCs between the intercalators and G2.

3.4 Discussion

The rational design of a rigid, bisintercalator linker for the purpose of increasing binding affinity and specificity toward DNA presents a number of significant challenges. Generally speaking, rigid linkers should decrease an unfavorable entropic contribution to binding. In practice, it can be difficult to achieve simultaneously the goals of precise control of functional group orientation and idealize the rigidity of their connections.

The detailed structure of complex **C1**-d(CGGTACCG)$_2$, derived from NMR data and shown here, may shed some light on these issues. **C1**, essentially composed of two

NDI intercalators and a reasonably rigid, non-charged, tricyclic spiro linker associates firmly with the GGTACC sequence through combined minor groove binding and bisintercalation. The novel linker aligns perfectly with the minor groove's curvature, as well as providing the correct spacing between the two NDI aromatic units.

Previous quantitative DNAse I footprinting results suggested that **C1** binds best to 5-GGTACC-3' sequence, with an affinity of $K_D \sim 10^{-7}$ M (Mazzitelli, 2007). The NMR titration experiments reported here clearly suggested a specific binding mode for **C1** interacting with d(CGGTACCG)$_2$ at exactly a 1:1 molar ratio. No such single species were observed for complexes of **C1** with d(GGATCC)$_2$, d(GGCGCC)$_2$ or d(GGGCCC)$_2$.

As shown in the NMR modeling structure, the tricyclic spiro structure binds in the minor groove of the GTAC sequence. The close proximity of the ligand linker to minor groove is not only indicated by the numerous NOEs observed between the central cyclohexane ring and the edge of DNA bases (mostly to A5A and A5B), but also the loss of C2 symmetry of the linker. In the unbound state, the ligand is symmetrical as indicated by the ^{1}H NMR spectrum: only three sets of exchangeable protons along the linker and two sets of non-exchangeable protons on the cyclohexane ring as expected for equatorial and axial protons, respectively (Chu, 2006). Upon binding to d(CGGTACCG)$_2$, most resonances for DNA and **C1** lost their degeneracy, as the complex is no longer 2-fold symmetric. In other words, **C1** does not have 2-fold symmetry when in the bound conformation. The number of cyclohexane resonances increased from two in the free state to six when bound, consistent with the central cyclohexane ring being locked into a single chair conformation in the minor groove. It is thus reasonable to propose that the

locked chair conformation structure of cyclohexane ring upon binding gives rise to the loss of the C2 symmetry of the entire **C1**-d(CGGTACCG)$_2$ complex.

The previously developed peptide linked bisintercalator **G$_3$K** was demonstrated to bind to the same sequence as **C1**, but in the major groove and with 2-fold lower overall binding affinity (Mazzitelli, 2007). In that case, structural analysis led to the conclusions that the glycine$_3$-lysine linker was the proper length and possessed complementary electrostatics to the floor of the major groove in the GGTACC sequence.

The most intriguing aspect of the present work is that the rigid new linker of **C1** prefers the same DNA sequence as the **G$_3$K** molecule, but binds in the minor groove. This unprecedented change in binding modality is most likely due to a combination of several factors, including linker length, sterics, electrostactic interactions, and hydrophobic contacts.

For example, the linker length for **C1** (from the top C=O to the bottom C=O) is 13.9 Å, almost identical to the effective length of the linker of another minor groove binding bisintercalator **(β-A)$_3$K** (Guelev, 2002), while for **G$_3$K** the length is 11.8 Å (from top C=O to the bottom N-H) (Guelev, 2001(a)). The major groove can require a slightly shorter linker to span the same number of bases compared to the minor groove. This is because the major groove is wider, allowing for a linker to adopt more of a "diagonal" geometry, requiring less length. The narrower minor groove requires a linker to traverse the entire length between bases (Saenger, 1984). Consistent with this notion, our two bisintercalators with slightly longer lengths (**C1** and **(β-A)$_3$K**) preferred the minor groove, while the shorter linker of **G$_3$K** preferred the major groove.

Steric interactions are likely another factor important for determining groove binding selectivity. Modeling indicates that in the major groove of 5'-GTAC-3', the exocyclic amino group from A and methyl group from T make it too crowded to accommodate the bulky cyclohexane ring of the **C1** linker. There is no steric crowding in the minor groove of the same sequence. In fact, the modeling indicated that the relatively flat, hydrophobic 5'-GTAC-3' minor groove could easily accommodate the tricyclic spiro linker in an orientation that provides for van der Waals contacts between the central cyclohexane CH groups and the hydrophobic walls of the minor groove.

The NMR structural models also provide a hydrogen bonding explanation for **C1** favoring the binding of 5'-GTAC-3' over 5'-GATC-3'. Two such hydrogen bonds that likely assist the specific binding are the N–HAH—O2 of T4B and N–HAG—N3 of A5B as indicated in Figure 3.10c. Modeling also indicated there is a strong hydrogen bond formed between the carbonyl group close to the NDI2 and NH_2 of G3B. Although it is not entirely clear why the two NDI rings have different flipping rates, the NDI2 carbonyl to G3B hydrogen bond may play an important role in this dynamic behavior.

3.5 Conclusions

NMR analysis of the bisintercalator **C1**-d(CGGTACCG)$_2$ complex has revealed the versatility of threading polyintercalation based on NDI moieties by verifying binding of the **C1** linker in the minor groove. This same DNA sequence was recognized by another NDI based bisintercalator, **G$_3$K**, that preferred the major groove. The structures of these two linkers are very different and appear to achieve recognition of the same

sequence by having overall length, shape, electrostatic and hydrophobic complementarity to their preferred grooves.

There are several positions on the central cyclohexane ring of **C1** that are in close proximity to the DNA bases when bound, and could thus be used as potential sites for derivatization in order to modify sequence specificity in a rational way. In addition, it has not escaped our attention that it should be possible to create a cyclic bisintercalating molecule utilizing the tricyclic spiro linker presented here as a minor groove recognition element, combined with a **G₃K** major groove recognition element for the d(CGGTACCG)₂ sequence. A cyclic bisintercalator that recognizes simultaneously both grooves of DNA should provide unique opportunities to explore new DNA binding modes, possibly at the limit of what is accessible from a molecular dynamics point of view.

3.6 Materials and methods

Sample preparation. The synthesis for compound **C1** has been described in Chapter 2 (Chu, 2006). The DNA (gel filtration grade, Midland Certified, Midland, TX) was initially dissolved in 0.7 ml 30 mM Na phosphate buffer, pH 7.5. Prior to addition of **C1**, the d(CGGTACCG)₂ sample was diluted to 10 ml in chilled water to which ligand was added and the sample was lyophilized. Samples used for NMR spectra in D_2O solvent were lyophilized twice from D_2O and finally dissolved in 0.7 ml 99.9% D_2O (Cambridge Isotope, Cambridge, MA, USA). The final concentration of the complex used in obtaining NMR spectra was approximately 1 mM.

NMR spectroscopy. Spectra were obtained using a 500 MHz Varian Unity-Inova spectrometer. 2D NOESY (60 and 200 ms mixing time), TOCSY (50 ms mixing time) and COSY spectra were acquired in D_2O solvent at 25 °C using a sweep width of 6000 Hz, with 2048 complex points (t2) and 512 (t1) complex points being acquired in each dimension; presaturation was used to remove the residual HOD signal. To observe solvent-exchangeable protons, NOESY (200 ms mixing time) and TOCSY spectra were acquired at 15 and 25 °C in 9:1 H_2O:D_2O solvent, using a sweep width of 12200 Hz; NOESY spectra were acquired using the jump-return method for solvent suppression, so as not to saturate the signals of protons that exchange rapidly with the solvent. [1]H spectra were referenced to the H_2O resonance at 4.75 ppm at 25 °C. A ^{13}C-^{1}H HSQC obtained at natural ^{13}C abundance was useful in confirming the DNA and ligand resonance assignments. The 1D spectra were processed using VNMR software (Varian). 2D spectra were processed using NmrPipe (Delaglio, 1995) and displayed using Sparky (Goddard and Kneller, 2004).

For RDC experiments, the sample was prepared by adding Pf1-phage solution (Profos AG, Germany; protease free) to the **C1**-d(CGGTACCG)$_2$ solution in sodium phosphate buffer, followed by gently mixing and then storage at 4 °C before use. The final sample contained 9:1 H_2O:D_2O as well as 1 mM **C1**-d(CGGTACCG)$_2$ complex, 20 mg/mL Pf1-phage, 26 mM Na phosphate buffer pH 7.5, 4 mM potassium phosphate, 0.7 mM $MgCl_2$, and 0.007% NaN_3. All spectra for RDC experiments were taken at 25 °C.

Modeling. Structure calculation of the **C1**-d(CGGTACCG)$_2$ complex was carried out using the simulated annealing protocol within the CNS program (Brünger, 1998). Distance restraints for non-exchangeable protons were derived from the observed cross peak intensities in the D$_2$O NOESY spectrum (60 ms mixing time). NOEs were grouped as strong (0-3.5 Å), medium (0-4.0 Å), weak (2.5-5.0 Å) and very weak (2.5-5.5 Å). The distance restraints involved methylene protons were determined by adding an extra 0.7 Å to distances measured from the center of methylene group. The cross peak intensities in 200 ms NOESY spectrum using 9:1 H$_2$O:D$_2$O as solvent were used to produce the distance restraints for exchangeable protons. The dna-rna_restraints.def file within CNS program was used to define hydrogen bonds for those identified Watson-Crick base pairs. The base pair planarity ($\pm 10^\circ$) was confined by applying torsion angle restraints. The DNA backbone dihedral angles were fixed to be within a 40° range of the values typical for B-form DNA ($\alpha = -46^\circ \pm 40^\circ$, $\beta = -147^\circ \pm 40^\circ$, $\gamma = 36^\circ \pm 40^\circ$, $\delta = 157^\circ \pm 40^\circ$, $\varepsilon = 155^\circ \pm 40^\circ$, $\zeta = -96^\circ \pm 40^\circ$) for the base pairs at intercalation sites, and a $\pm 25^\circ$ range for the rest base pairs (Lee, 2004(b)). Based on qualitative analysis of intra-ribose NOE intensities (supporting information), each ribose ring puckering pattern was assigned as C2' or C3' endo accordingly. All torsion angles of ribose rings were restrained to within 25° of the value typical of each conformation.

The parallhdg.dna parameter file in CNS was used to produce force field parameters for the DNA. Parameter and topology files for **C1** were created with the assistance of Dundee PRODRG2 Server (Schuettelkopf and van Aalten, 2004) and

Gerard Kleywegt's XPLO2D server (version 050802/3.3.2) (Kleywegt and Jones, 1997).

The atomic charges and bond lengths of **C1** were either adopted from the crystal structure

of an NDI derivative (Lokey and Iverson, 1995) or calculated using *ab initio* method in

HyperChem 7.0 (Hypercube Inc., 1115 NW 4th Street, Gainesville, FL 32601).

A variety of starting structures were used for the simulated annealing, including

those where two strands of DNA were not base paired, and the ligand was away from

DNA by 30 Å. In the first stage a 60 ps of torsion-angle molecular dynamics was applied

to those starting structures at 20,000 K to allow a high degree of randomization of the

initial models. In the second stage, the temperature was slowly decreased to 2,000 K over

a period of 60 ps, then to 300 K within 15 ps using Cartesian molecular dynamics.

Finally, the structures were subjected to 2000 steps of conjugate-gradient minimization.

The annealing process was repeated until the total energy was near a consistent minimum

value (Lee, 2004(b)).

Chapter 4   Cyclic Bisintercalators: Synthesis and DNA Binding Studies

4.1 Chapter summary

4.1.1 Goals

Establish a general approach to synthesize cyclic NDI-based structures. Study the binding behavior of cyclic bisintercalator toward DNA.

4.1.2 Approach

Solid phase organic synthesis strategy was applied to build the cyclic structures. The DNA binding ability of cyclic bisintercalator was characterized through dissociation kinetics studies, DNAse I footprinting and NMR spectroscopy.

4.1.3 Results

A general and efficient solid phase synthetic strategy was developed to synthesize NDI-based cyclic compounds. The number of NDI units and the sequence of amino acids within the cyclic structure can be manipulated as desired. For the first time, a sequence-specific cyclic DNA bisintercalator was identified based on data from DNAse I footprinting and NMR structural studies. It was shown that this cyclic bisintercalator interacts with $d(CGGTACCG)_2$ specifically through threading intercalation by spanning 4 base pairs, with one linker interacting with the major groove and the other one residing in the minor groove.

4.2 Introduction

Among the different DNA binding motifs studied so far, intercalation has gone from simple monointercalation, to bisintercalation. and now to threading polyintercalation, which has attracted considerable interest due to slow dissociation kinetics from the DNA (Tanious, 1991; Yen, 1982), while offering the potential to be a versatile DNA binding motif (Guelev, 2001(a); Guelev, 2002; Lee, 2004(a); Chu, 2007). For the threading polyintercalators studied so far, the groove selectivity and sequence specificity are derived from the linker structure. As described earlier, **G$_3$K**, a bisinteralatior having a linker composition of Gly$_3$Lys, prefers to bind to d(GGTACC)$_2$ with the linker in the major groove (Guelev, 2001(a)), while (β-Ala)$_3$Lys, from the bisintercalator **(β-A)$_3$K**, prefers to interact with d(GATAAG)·(CTTATC) in the minor groove (Guelev, 2002). As discussed in previous chapters, **C1**, a NDI-based bisintercalator with a rigid, tricyclic spiro linker was found to bind d(GGTACC)$_2$ sequence tightly and specifically, the same sequence as **G$_3$K** binds, but with the linker residing in the minor groove (Chu, 2007). The fact that both bisintercalators bind to the same DNA sequence specifically, but with linkers residing in different grooves, prompts us to explore the possibility of designing a sequence-specific cyclic bisintercalator as shown in the cartoon in Figure 4.1. The motivation for doing so is that an extremely slow dissociation kinetic process would be achieved for this type of binding and the sequence specificity would also be greatly enhanced compared to a simple bisintercalator with only one linker, since both grooves would be accessed simultaneously. Over and above any practical considerations, a cyclic bisintercalator would open up new questions concerning

113

association kinetics and DNA base stacking dynamics, since substantial DNA distortion would have to occur upon binding.
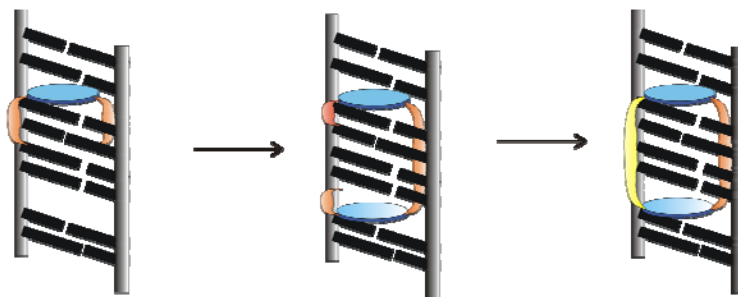
Figure 4.1 The cartoon showing the evolution of intercalation binding approach: from simple intercalation, to threading bisintercalation, to cyclic bisintercalation.

Cyclic DNA bisintercalators have been studied before, with mixed success and little in the way of definite structural characterization. For example, a macrocyclic bisacridine (Figure 4.2, bisacridine A) has been synthesized and its binding behavior toward double-stranded (ds) DNA were characterized (Zimmerman, 1989; Veal, 1990). No conclusive evidence supported any one binding mode. The cyclic bisacridine was seen to either interact with four base pairs $d(CGCG)_2$ by spanning two base pairs or it could bind two base pairs by spanning a single base pair, in violation of the neighbor-exclusion principle. This lack of a unique binding mode was most likely due to the lack of specificity of the cyclic structure linkers with either DNA grooves. Another cyclic bisacridine with two polyammonium connecting chains has also been studied for its ability to recognize an abasic site within duplex DNA (Figure 4.2, bisacridine B) (Jourdan, 1999). A major complex between DNA (11mer) and the cyclic ligand was identified in which one acridine unit stacked within the abasic site. However, the binding

detail for the minor complex was not determined due to a lack of suitable NMR structural information.

A   **B**
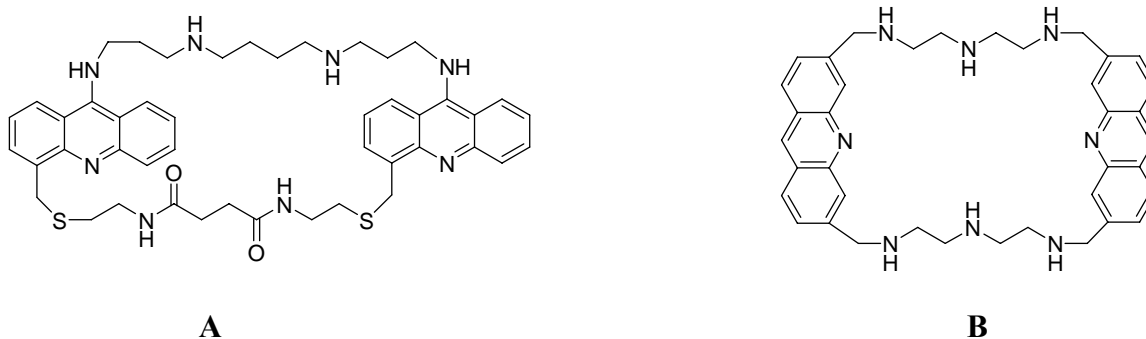
Figure 4.2 Previous reported two cyclic bisintercalators, bisacridine A and B.

Double-stranded DNA structure is known to be very dynamic in solution, with base pairs opening and reforming rapidly at room temperature. For example, imino proton half-lives are typically in the 1-50 ms range (Guéron and Leroy, 1992). A cyclic bisintercalator may be able to explore this millisecond "breathing" behavior of duplex DNA by sliding through DNA base pairs to form a catenane-like complex structure.

In this chapter, a general and efficient synthetic method to make NDI based bisintercalators is presented. One such structure, **CBI-1**, was first characterized by studying its dissociation kinetic behavior to poly(dGdC), a polymer that NDI-based intercalators bind tightly (Lokey, 1997). Sequence specificity was further analyzed using DNAse I footprinting. Finally, detailed NMR studies were carried out to explore the binding details of the cyclic bisintercalator toward a specific DNA sequence identified in the footprinting results.

4.3 Results

Design of the cyclic bisintercalator.   As discussed in previous chapters, the linker structure connecting the intercalators plays a key role in directing polyintercaltors to bind DNA either through the major or minor grooves selectively. The first generation cyclic bisintercalator has one linker from **G₃K** and the other from **(β-A)₃K** as shown in Figure 4.3. Thus, the cyclic bisintercalator has the recognition element for the d(GGTACC)₂ sequence in the major groove and yet should recognize d(GATAAG)·(CAAATC) sequence in the minor groove (but in the opposite direction). Note that the linker length of (β-Ala)₃Lys is expected to fit the minor groove of d(GGTACC)₂ by spanning four base pairs according to our modeling. Similarly, the linker length of Gly₃Lys would perfectly match the major groove of d(GATAAG)·(CTTATC). It will be interesting to determine which sequence preference dominates.
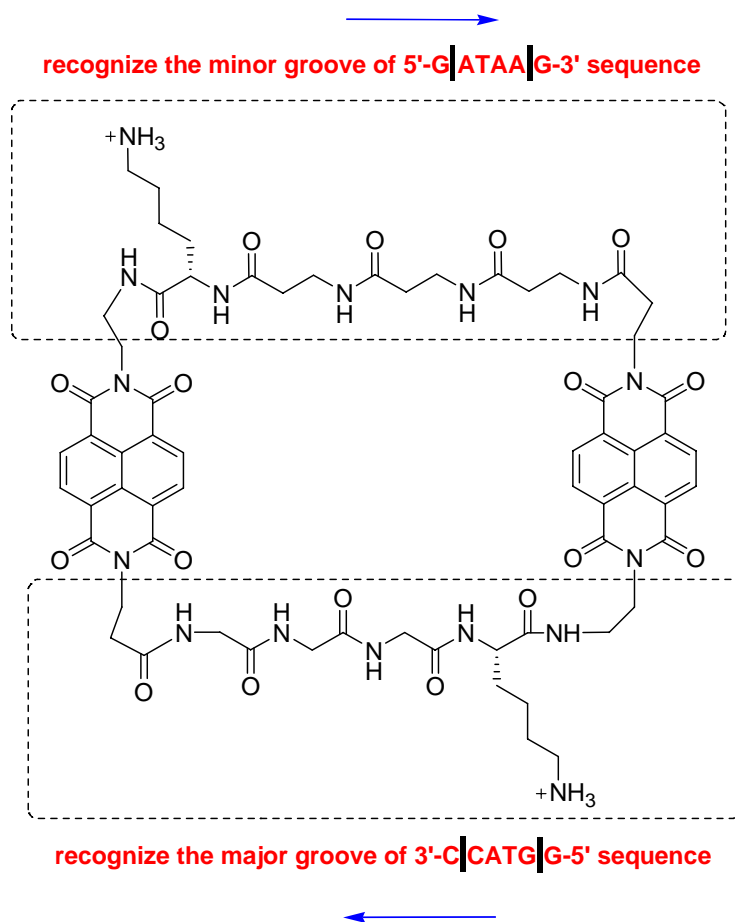
Figure 4.3 The chemical structure of cyclic bisintercalator **CBI-1**.

Synthesis of cyclic bisinteralator CBI-1. The presence of the lysine residue on **CBI-1** makes it possible to carry out the solid phase synthesis of **CBI-1** using a side-chain to resin link. Choosing the right type of resin is the first key step. PEG-based resins have excellent swelling properties over a wide range of solvents, from water to toluene. Several low loading PEG-based resins, such as NovaSyn TGA resin, are commercially available. It has been reported that higher loading resins gave the poorest results (Albericio, 1993). Also, the tentacle nature of the resin helps minimize inter-chain

oligomerization. The lysine side chain can be attached to the resin through a carbamate linkage as shown in Scheme 4.1. This carbamate linkage should be TFA-labile, since the linker contains a 4-methoxy phenol component. Thus, an active carbonate resin was first formed using DSC and DMAP (catalyst). FmocHN-Lys-OAll, which was synthesized in high yield (Alsina, 1998), was attached to the resin. The three glycine amino residues were attached to the lysine-modified resin using standard solid phase peptide synthesis protocol. The carboxylate group of lysine on the resin was fully deprotected using $Pd(PPh_3)_4$ as catalyst within 2.5 hr (Kates, 1994), followed by a coupling of NDI monomer **4**, NH_2-$(CH_2)_2$-NDI-$(CH_2)_2$-CO_2All. This unit was produced by treating the BocNH-$(CH_2)_2$-NDI-$(CH_2)_2$-CO_2All with 1:1 TFA / DCM (Scheme 4.2). On the *N*-terminal side, another NDI monomer, Fmoc-Lys(Boc)-$(CH_2)_2$-NDI-$(CH_2)_2$-CO_2H was attached (Guelev, 2001(b)), followed by attachment of three β-alanine residues (Scheme 4.3). The allyl protecting group on the *C*-terminal side was removed as described above. Using PyBOP / HOAt under weakly basic conditions, the final cyclization step was completed within 18 hours (Scheme 4.4). Other activating reagent combinations such as PyBOP / HOBT, DIPICI / HOBT didn't result in clean cyclization. Upon cyclization, the desired product was cleaved off the resin in 1.5 hr by using TFA/TIPS/H_2O (95/2.5/2.5) (Scheme 4.4). The identity of **CBI-1** was confirmed by ESI-MS and NMR results.

The analytical HPLC profiles for the crude mixture before and after the cyclization are displayed in Figure 4.4. Only one major species appeared in both mixtures, and the cyclization step went with nearly 100% conversion, resulting in relatively easy purification. The low loading NovaSyn TGA resin appears to have been

the right choice to for solid phase synthesis of cyclic NDI bisintercalators. Interestingly, the synthetic effort was also attempted on 2-chlorotrityl resin, a popular resin for non-natural peptide synthesis. Although the attachment of FmocHN-Lys-OAll to this resin at the beginning stage was fairly straightforward with good yield (~ 50%), the deprotection of the allyl group was problematic, even when using excess Pd(PPH$_3$)$_4$/TIPS in CH$_2$Cl$_2$ (Thieriet, 2000). In addition, too many by-products were formed, making final purification difficult. Although **LBI-1**, the linear precursor of **CBI-1**, was formed on 2-chlorotrityl resin among many other by-products, the desired compound **CBI-1** couldn't be identified based on data from analytical HPLC and ESI-MS. Possible reasons for the failure on 2-chlorotrityl resin might involve the high initial loading (1.5 mmol/g) and/or relatively narrow solvent swelling properties for this polystyrene-based resin.

(a)   (b)
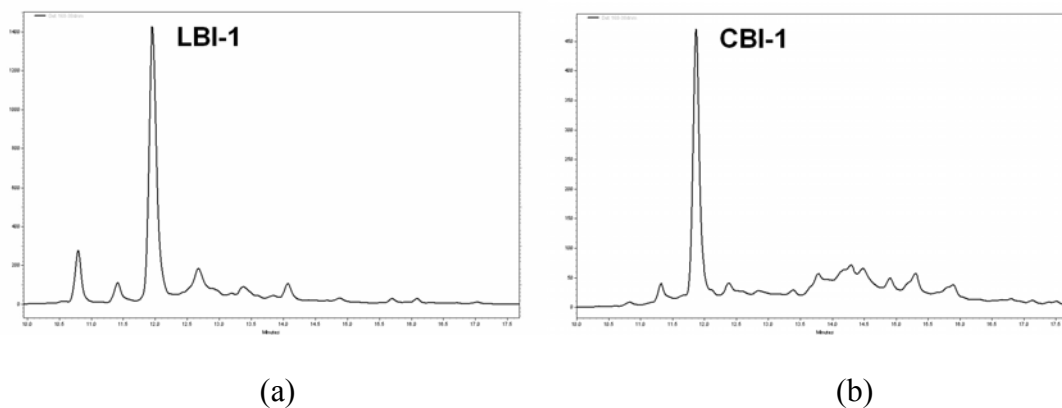
Figure 4.4 HPLC profiles of the crude mixtures before (a) and after (b) cyclization on NovaSyn TGA resin.

Scheme 4.1

NovaSyn TGA resin

DSC, DMAP

1

Fmoc-Lys-OAll

DIPEA

2

Fmoc SPPS

3 couplings

2

1. Pd(PPh$_3$)$_4$

2. NH$_2$-(CH$_2$)$_2$-NDI-(CH$_2$)$_2$-CO$_2$All

4

3

CO$_2$All

5

Scheme 4.2

4: NH$_2$-(CH$_2$)$_2$-NDI-(CH$_2$)$_2$-CO$_2$All

Scheme 4.3

7

1. 20 % piperidine / DMF

5

2. Fmoc-Lys(BOC)-(CH$_2$)$_2$-NDI-(CH$_2$)$_2$-CO$_2$H
6

Fmoc SPPS   1. Pd(PPh$_3$)$_4$

3 couplings   2. 20 % piperidine / DMF

8
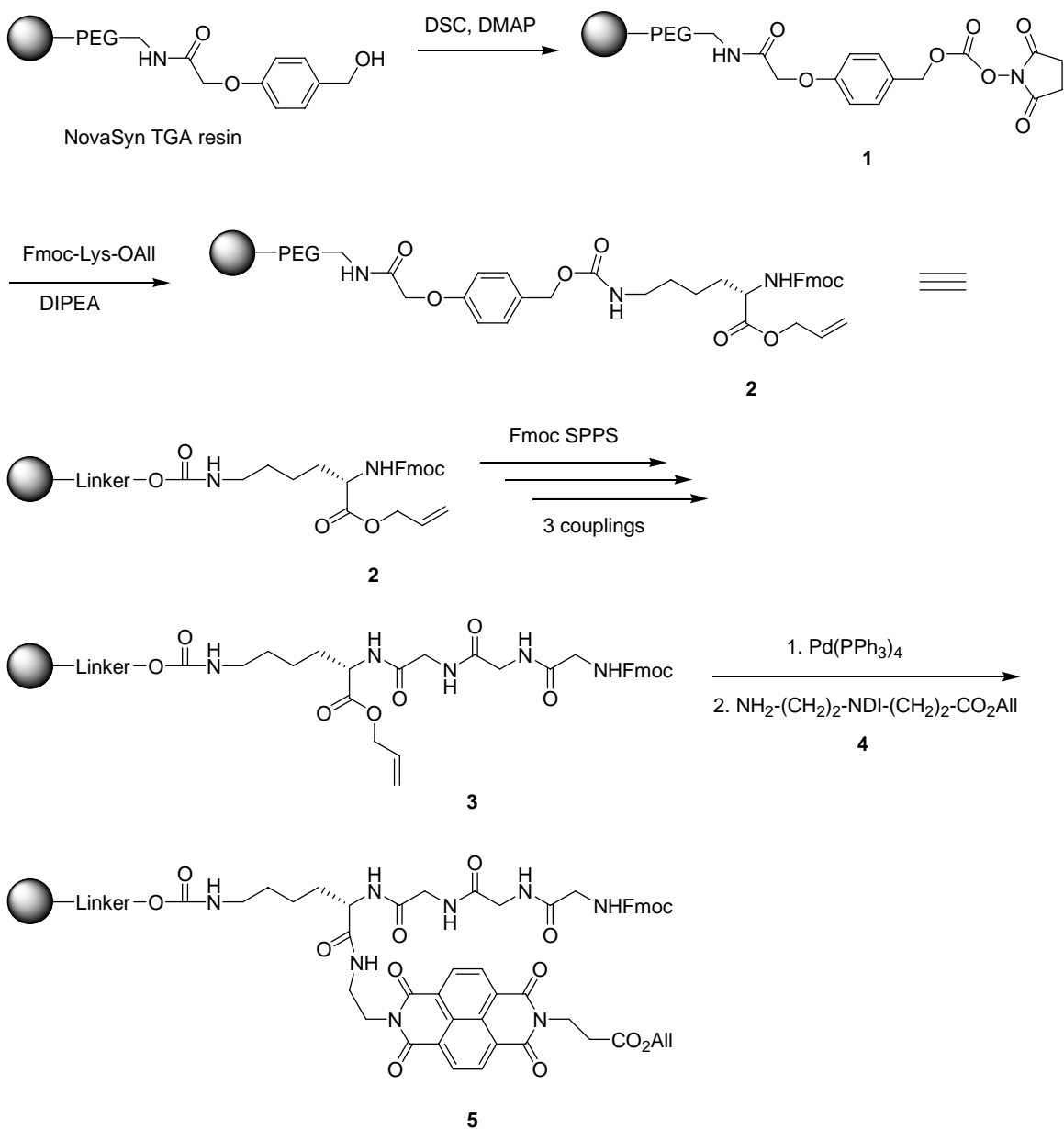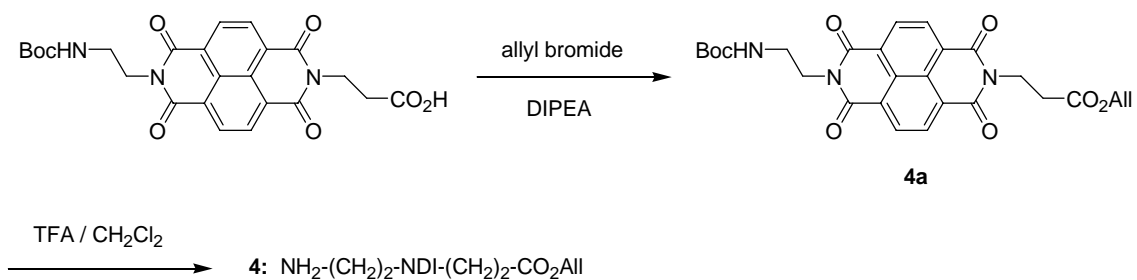
Scheme 4.4
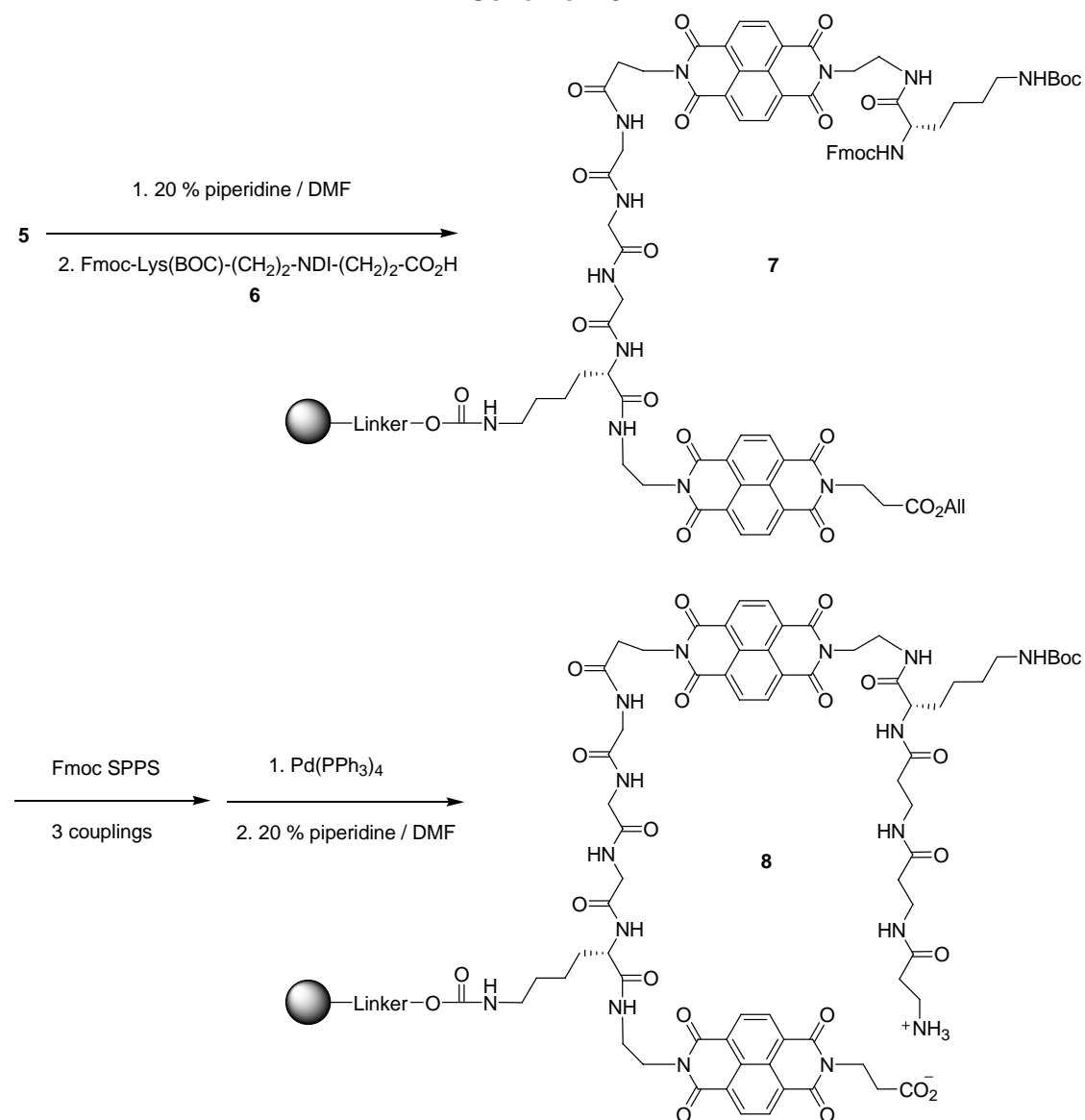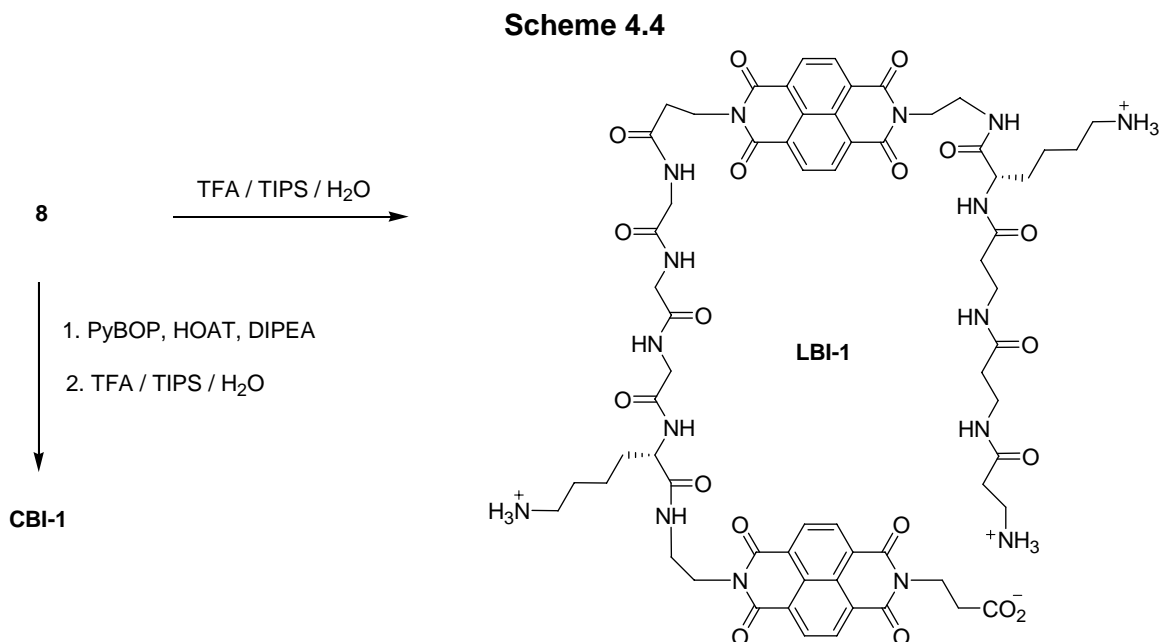
Synthesis of cyclic bisinteralator CBI-2. The chemical structure of **CBI-2** was shown in Figure 4.5. This cyclic structure was derived directly from the linear bisintercalators **C1** (Chu, 2006) and **G₃K**, with the former binding to the minor groove of d(GGTACC)₂ specifically while the latter binding to the major groove of the same sequence (Figure 4.6). It is expected that this cyclic structure would bind to d(GGTACC)₂ with high affinity and specificity and extremely slow dissociation kinetics. The rigid linker structure from **C1** is expected to prevent the self-association of two NDI units, thus greatly increasing the chance for **CBI-2** to target 6 base pairs directly, rather than some possible intermediate binding states, such as 5, 4 or 3 bp.
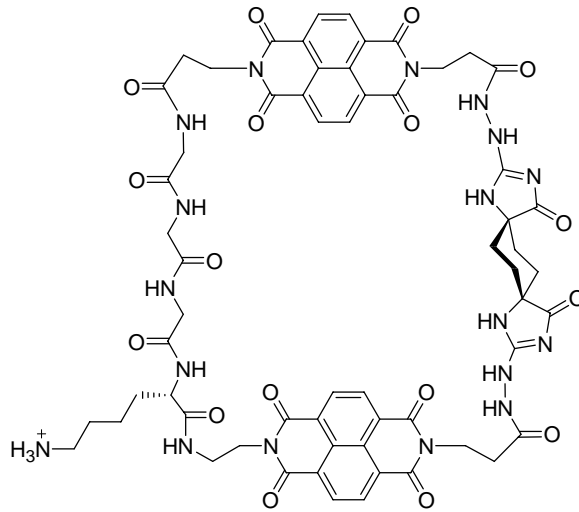
Figure 4.5 The chemical structure of **CBI-2**.
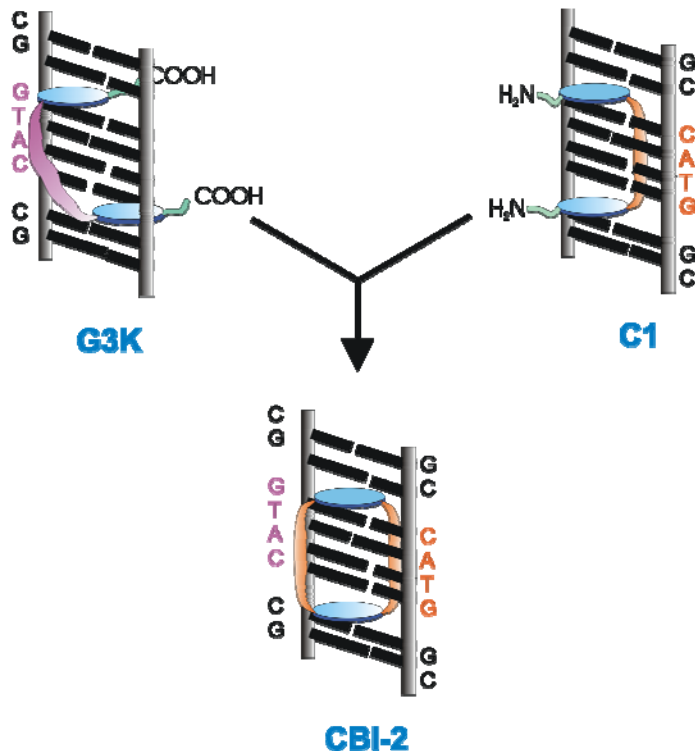
Figure 4.6 The cartoon showing how the **CBI-2** is derived.

As shown in Scheme 4.5, the synthesis began with orthogonally protected Lys-Gly-Gly-Gly on the resin, followed by the attachment of mono-protected NDI biscarboxylic acid **9**, HO_2C-$(CH_2)_2$-NDI-$(CH_2)_2$-CO_2Fm, which was made through DIPCDI-activated coupling reaction (Scheme 4.6) (Kessler and Siegmerier, 1983). Since the Fm and allyl protecting groups are orthogonal to each other, it is expected that the spiro linker **11** should be attached to only one of the two carboxylate groups upon selective deprotection (Scheme 4.7). If the spiro linker is kept in excess (e.g. 5 equiv relative to the resin loading), the chances of getting both amino groups coupled to the resin (cross coupling) should be extremely low. Another Fmoc protected NDI monomer **13**, HO_2C-$(CH_2)_2$-NDI-$(CH_2)_2$-NHFmoc (Scheme 4.8), can thus be attached to the resin. This Fmoc-protected, NDI-based non-natural amino acid **13** may be valuable in the synthesis of other NDI-related peptides as well. Upon removal of the allyl and Fmoc groups successively, the desired cyclic, rigid bisintercalator could be made through the on-resin cyclization. Unfortunately, the coupling step between the spiro linker and the carboxylate group on the resin did not go as expected, possibly due to poor solubility of **11** in DMF or NMP. Different solvent mixtures, such as DMF/DMSO, NMP/DMSO, were examined, but no desired product could be identified. Microwave-assisted coupling method (heating up to 75 °C for coupling step) may be tried to enhance the coupling efficiency in the future.

Scheme 4.5

3

1. 20 % piperidine / DMF

2. HO$_2$C-(CH$_2$)$_2$-NDI-(CH$_2$)$_2$-CO$_2$Fm

9

10

Scheme 4.6

Fm-OH

DMAP, DIPCDI

9a

TFA / triethyl silane / CH$_2$Cl$_2$

9: HO$_2$C-(CH$_2$)$_2$-NDI-(CH$_2$)$_2$-CO$_2$Fm

Scheme 4.7

1. 20% piperidine / DMF

10 ✕

2.

11

12

●—Linker—O—C(=O)—NH—...—CO₂All

1. HO₂C-(CH₂)₂-NDI-(CH₂)₂-NHFmoc

13

2. Pd(PPh₃)₄

3. 20 % piperidine / DMF

4. PyBOP, HOAT, DIPEA

14

●—Linker—O—C(=O)—NH—...

TFA / TIPS / H₂O

CBI-2

Scheme 4.8

13: HO₂C-(CH₂)₂-NDI-(CH₂)₂-NHFmoc
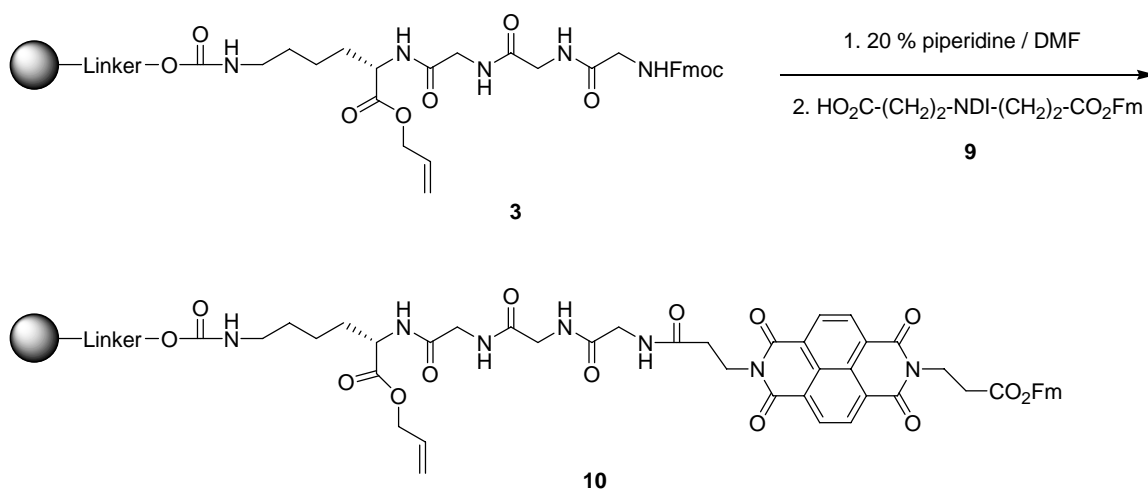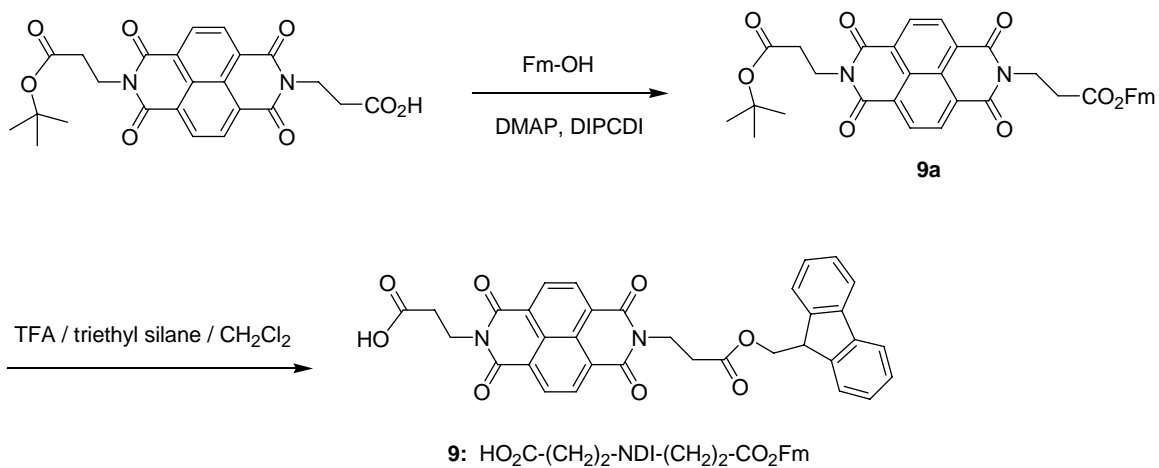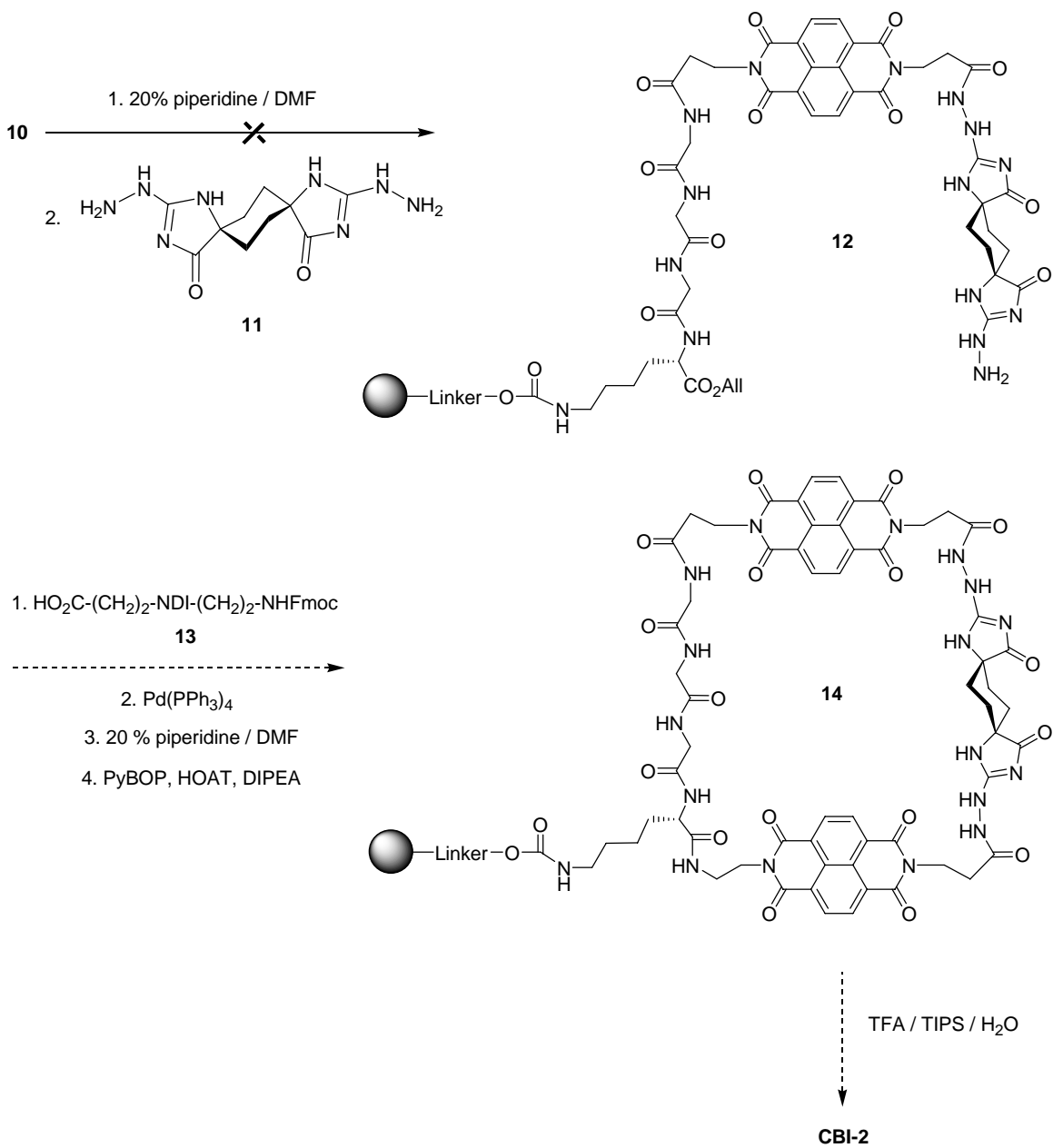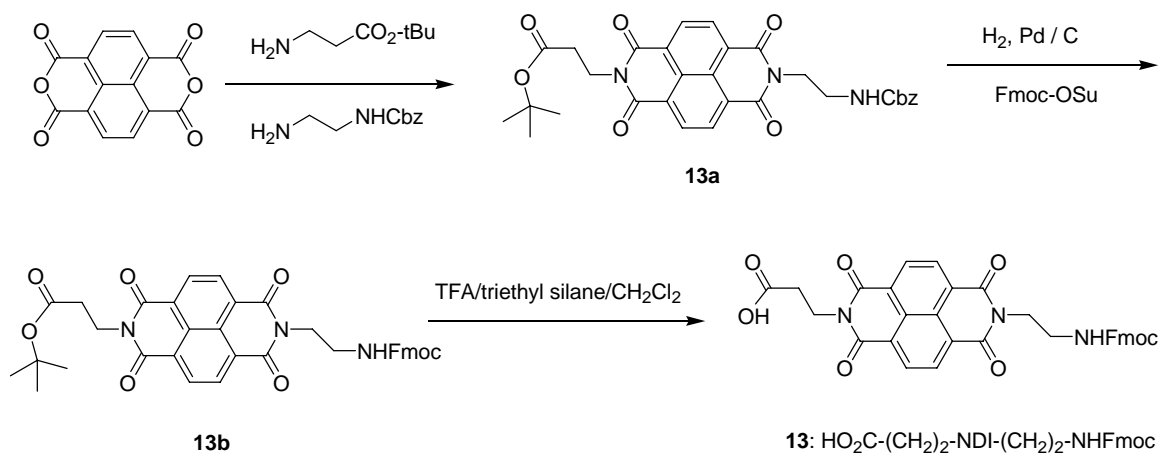
Dissociation kinetics of CBI-1 on poly(dGdC). It has been established that previous NDI-based intercalators bind poly(dGdC) DNA more tightly than poly(dAdT) (as shown in Chapter 2). Therefore, as an initial binding test, the dissociation kinetic behavior of **CBI-1** on poly(dGdC) was carried out. The results are shown in Figure 4.7. Prior to binding to DNA, the UV spectrum of **CBI-1** was no different than a linear bisintercalator, with the peak at 386 nm being lower than that at 363 nm, indicating a self-stacking in aqueous solution between two NDI units which is based on a comparison to previous work. The molar extinction coefficient for **CBI-1** is 25,000 L·mol^{-1}·cm^{-1} (at 386 nm), slightly smaller than the values determined for linear bis-NDI molecules such as **LBI-1** (26,000-27,000 L·mol^{-1}·cm^{-1}) (Guelev, 2001(b)). Upon adding poly(dGdC), the spectrum of the **CBI-1** changed. Apparently, the self-stacking of two NDI units was broken and the observed hypochromicity (particularly at 363 nm) suggested that both NDI units are likely intercalated between GC pairs.

The dissociation kinetics profile is shown in Figure 4.8 and is adequately described as one phase exponential decay, $A = a_0 \text{ Exp } (-kt) + b_0$ (k: dissociation rate constant, $t_{1/2}$: half life) (Chen, 1997). The half-life for **CBI-1**-poly(dGdC) complex was determined to be 41 minutes, significantly longer than the half life of a typical linear dimer, which is around 15 minutes (11.5 min for **LBI-1**; Chu, 2006). Therefore, it is reasonable to assume that the cyclic bisintercalator **CBI-1** does indeed intercalate through GC pairs with one linkage in the major groove and the other in the minor groove of poly(dGdC). If true, an interesting question concerns how many GC the **CBI-1** spans in the **CBI-1**-poly(dGdC) complex.

(a)

(b)

(c)
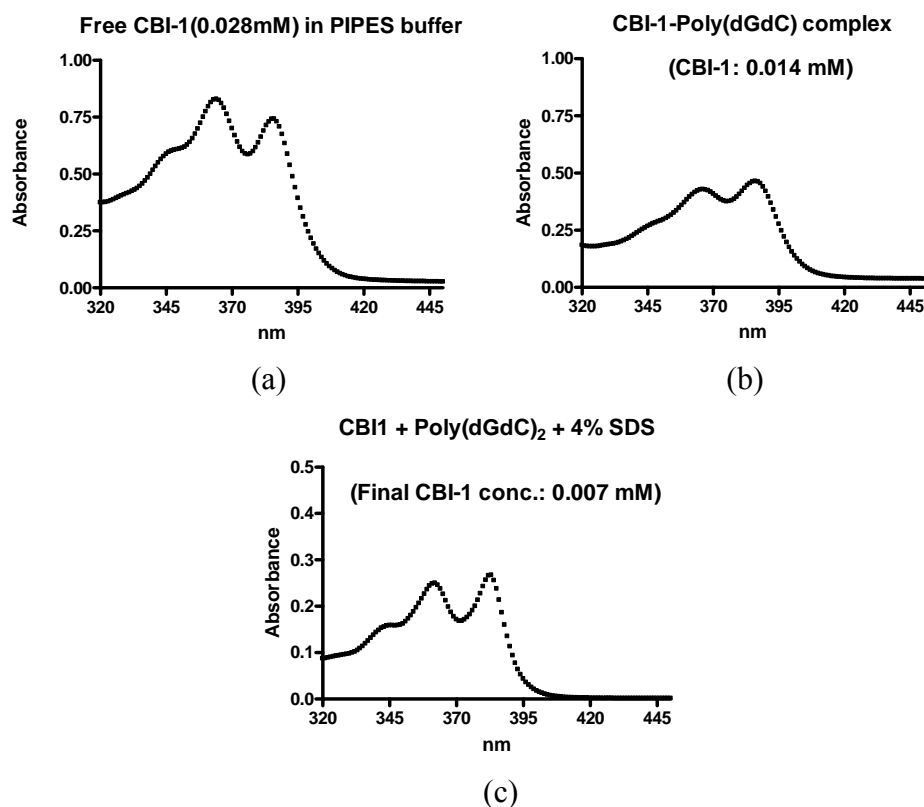
Figure 4.7 The UV-vis profiles of (a) **CBI-1** only; (b) **CBI-1**-Poly(dGdC) complex; (c) **CBI-1**-Poly(dGdC) upon the addition of 4% SDS.
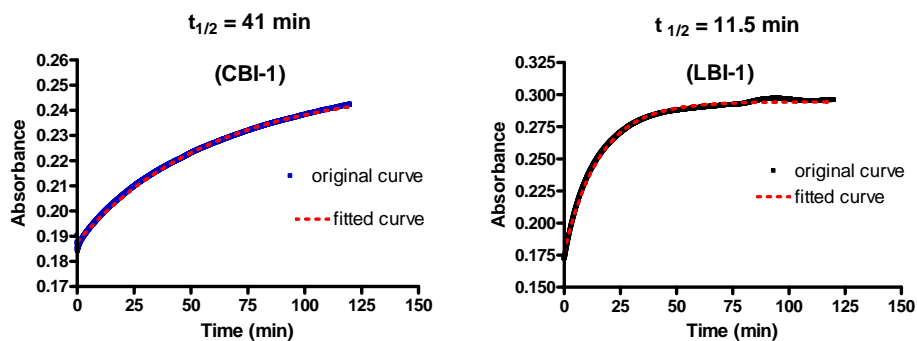
Figure 4.8 The dissociation kinetics profiles of **CBI-1** and **LBI-1** from Poly(dGdC) upon the addition of 4% SDS.

DNAse I footprinting of CBI-1 on a 69mer DNA. The preferred binding sequences could include GGTACC and GATAAG, since **CBI-1** has a major groove recognition component Gly₃Lys and a minor groove recognition element (β-Ala)₃Lys for these sequences, respectively. As shown in Figure 4.9, the 69mer DNA contains multiple possible binding sites for **CBI-1**, such as 5'-GGTACC-3', 5'-GATAAG-3', etc.

(+) strand:

5' GGT AAT TTA TTA T**GG TAC C**AT ATT AAT T**GA TAA G**TA **CTT ATC** ATT AAT TAT **GAT AAG** TTA TAA TTA ACC 3'

(-) strand:

3' CCA TTA AAT AAT A**CC ATG G**TA TAA TTA A**CT ATT C**AT **GAA TAG** TAA TTA ATA **CTA TTC** AAT ATT AAT TGG 5'

Figure 4.9 The sequence of a 69mer DNA used in DNAse I footprinting experiment.

129
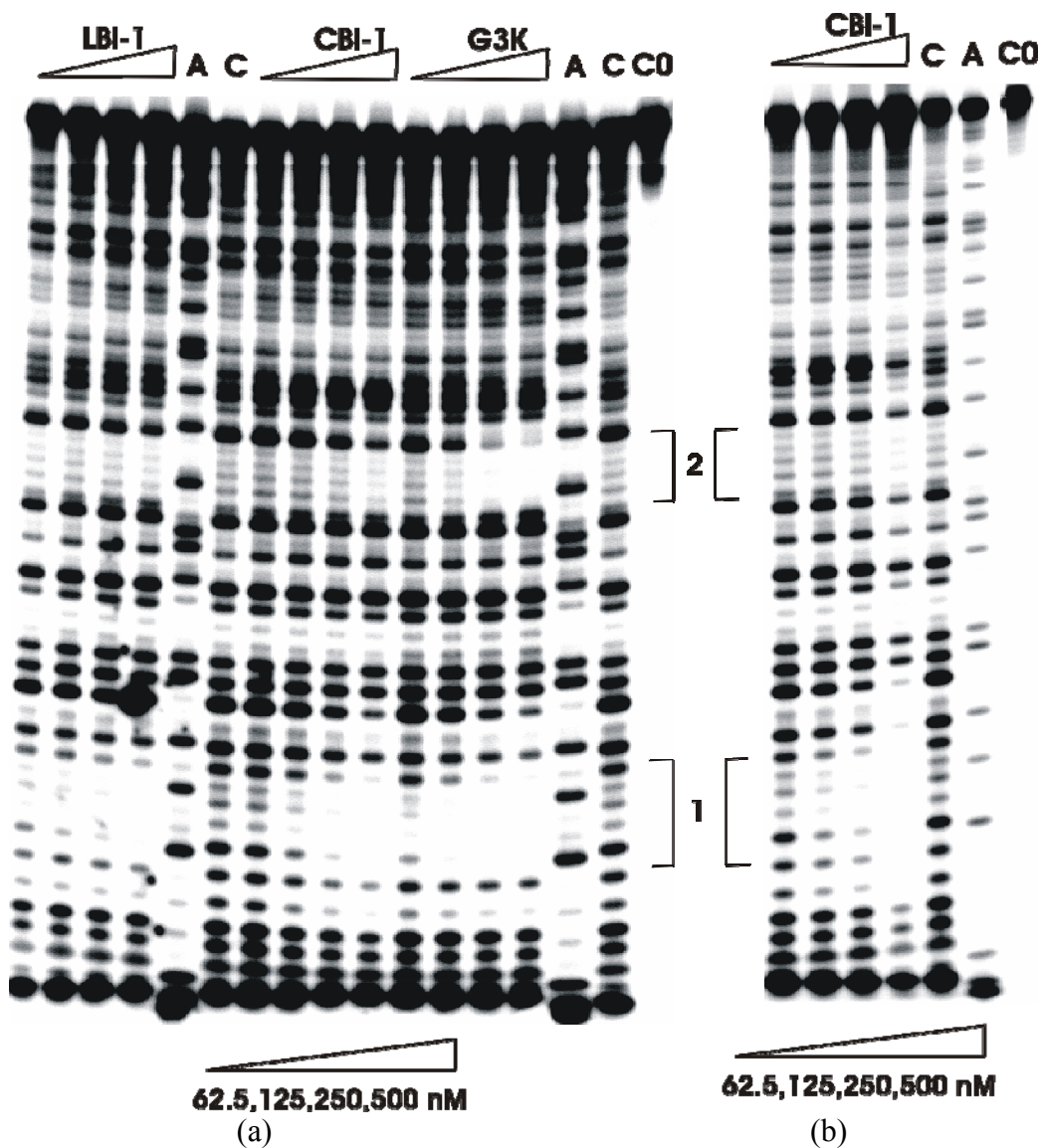
Figure 4.10 The DNAse I footprinting of different dimers on a 69mer DNA: (a) dimers of **LBI-1**, **CBI-1** and **G₃K** after a incubation time of 1.5 hours; (b) **CBI-1** after incubation time of 38 hours. Lane A represents Adenine-specific sequencing reaction. Lane Co contains DNA without DNAse I. Lane C contains DNA with DNAse I but no ligand. Sequences at 1: 5'-GGTACC-3'; 2: 5'-AGTACT-3'.

The DNAse I footprinting data is shown in Figure 4.10. As expected based on previous work, the linear dimer **G₃K** binds to the d(GGTACC)₂ sequence with high

130

affinity. Unexpectedly, **G₃K** also binds to another palindromic sequence, d(AGTACT)$_2$, a specificity that has not been seen before. For the linear dimer **LBI-1** (linear structure right before the cyclization) and cyclic dimer **CBI-1**, both were found to bind only d(GGTACC)$_2$ sequence ($K_D \sim 10^{-7}$ M), although with slightly lower binding affinity than dimer **G₃K**. Interestingly, at **CBI-1** lanes no footprint was found at sequence d(GATAAG)·(CTTATC), a sequence that linear dimer **(β-A)₃K** binds strongly and specifically (Guelev, 2002). This clearly indicated that it is the Gly₃Lys linkage that dominates the binding specificity of the cyclic structure **CBI-1**. Although **LBI-1** and **CBI-1** shared similar binding affinity and specificity, **CBI-1** has a clear cut footprint at the GGTACC site, while **LBI-1** doesn't. This may indicate that the flexible (β-Ala)₃Lys side chain of **LBI-1** may exist in multiple conformations with regard to interacting with DNA groove(s). The footprinting results (Figure 4.4 (b)) from incubating the ligand **CBI-1** with DNA for up to 38 hours did not show any significant change in specificity compared to an incubation time of 1.5 hours, although the longer incubation time did appear to give a little more bound cyclic bisintercalators.

Titration of d(CGGTACCG)₂ with cyclic bisintercalator CBI-1. The naming convention of **CBI-1** and DNA are shown in Figure 4.11. A titration of d(CGGTACCG)$_2$ with **CBI-1** was monitored by 1D ^{1}H NMR spectroscopy (Figure 4.12). Upon adding 0.5 equivalents of **CBI-1** to the free DNA, a few new sharp peaks appeared while the peaks of the free DNA remained unchanged in both position and shape, although the intensity decreased significantly. These observations indicate a relatively tight and specific binding

131

between **CBI-1** and d(CGGTACCG)$_2$, with dissociation kinetics being slow on NMR the time scale. After adding 1.0 equivalent of ligand **CBI-1**, the peaks from the free DNA disappeared, and a new set of resonances emerged, suggesting specific binding between **CBI-1** and d(CGGTACCG)$_2$ with a 1:1 stoichiometry. Assignments for all the imino and amide protons in the complex were made based on the analysis of 2D NMR data (in 9:1 H$_2$O/D$_2$O, at 10 or 25 °C; see below).
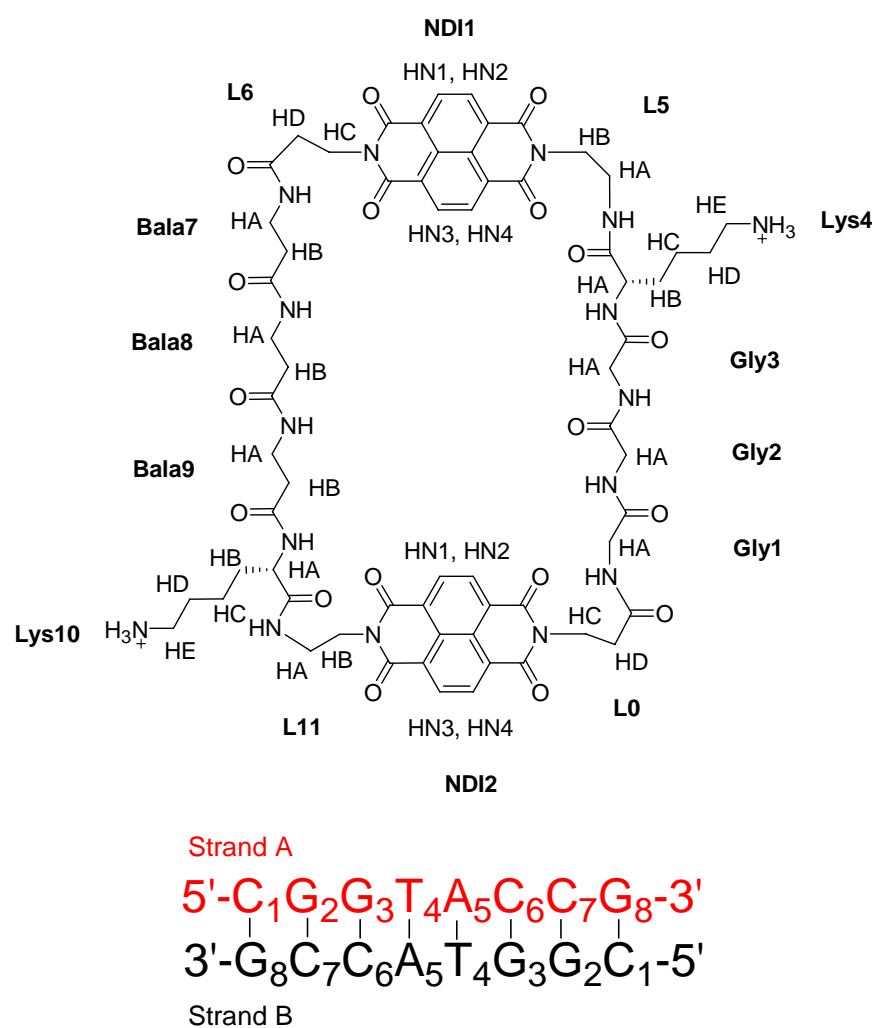
Figure 4.11 The naming convention of compound **CBI-1** and DNA sequence used in the NMR study.
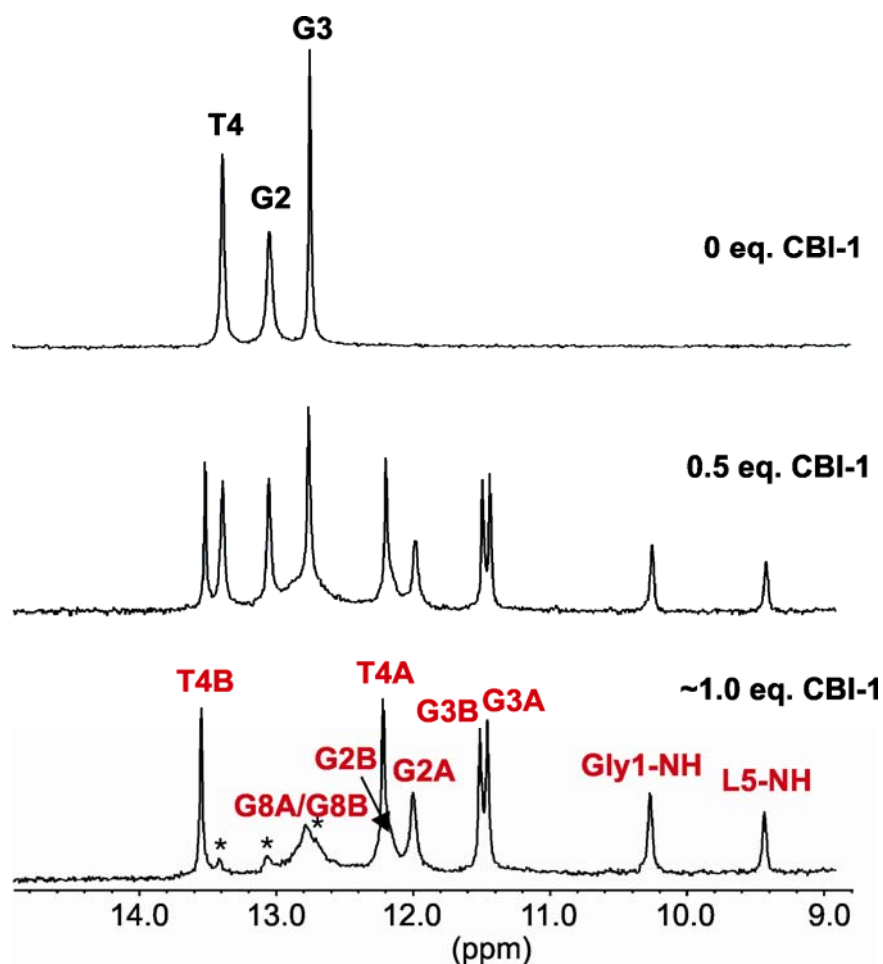
Figure 4.12 1D ^{1}H NMR spectra of **CBI-1** titration into d(CGGTACCG)$_2$ in H$_2$O/D$_2$O (9:1) with 30 mM Na phosphate buffer at 25 °C (* a trace amount of free DNA left in solution).

The titration data clearly indicated that **CBI-1**-d(CGGTACCG)$_2$ complex doesn't retain the 2-fold symmetry of the free d(CGGTACCG)$_2$. This is expected since the ligand **CBI-1** is not symmetrical. The most notable chemical shift differences between free DNA and ligand-DNA complex are observed for the central TA base pairs, as well as for the G3-C6 base pair. Upfield shifting of the imino resonances of base pairs G2-C7 and G3-C6 are consistent with intercalation at these sites (Feigon, 1984). Resonances in the

9.0-10.4 ppm range were also observed from the exchangeable protons of **CBI-1** upon binding (10.3 and 9.4 ppm, respectively).

Assignment of the CBI-1-d(CGGTACCG)$_2$ complex. The signals from the intercalators in the ligand-DNA complex were first identified. In the aromatic-aromatic region of TOCSY (in D$_2$O) spectrum, only four cross peaks appeared; they were assigned to 4 pairs of aromatic protons on two NDI rings as in the past (Guelev, 2002; Lee, 2004(a); Chu, 2007). The assignments for the **CBI-1**-d(CGGTACCG)$_2$ complex were assigned by following the H6/8-H1' and H6/8-H2'/H2'' NOE connectivities as shown in Figures 4.13, 4.14 and 4.16 (Wüthrich, 1986). The internucleotide NOE connectivities for DNA d(CGGTACCG)$_2$ were weak or interrupted at the steps G2-G3 and C6-C7 (Figures 4.13 and 4.14). New connectivities appeared at these steps (Figure 4.15). Two protons on NDI1 have NOEs with H1's of G2A, G3A and C7B. Three protons on NDI2 were observed having through-space interactions with H1's of G2B and G3B, and H5 of C6A. NOE cross peaks between H2'/H2''s of G2A and NDI1, G2B and NDI2 were also identified (Figure 4.16). The NMR data suggests that **CBI-1** interacts with d(CGGTACCG)$_2$ with the two NDI rings intercalating between GpG steps in a threading manner. These conclusions suggest that one linker must be located in the major groove and the other in the minor groove, as expected based on our original design strategy.
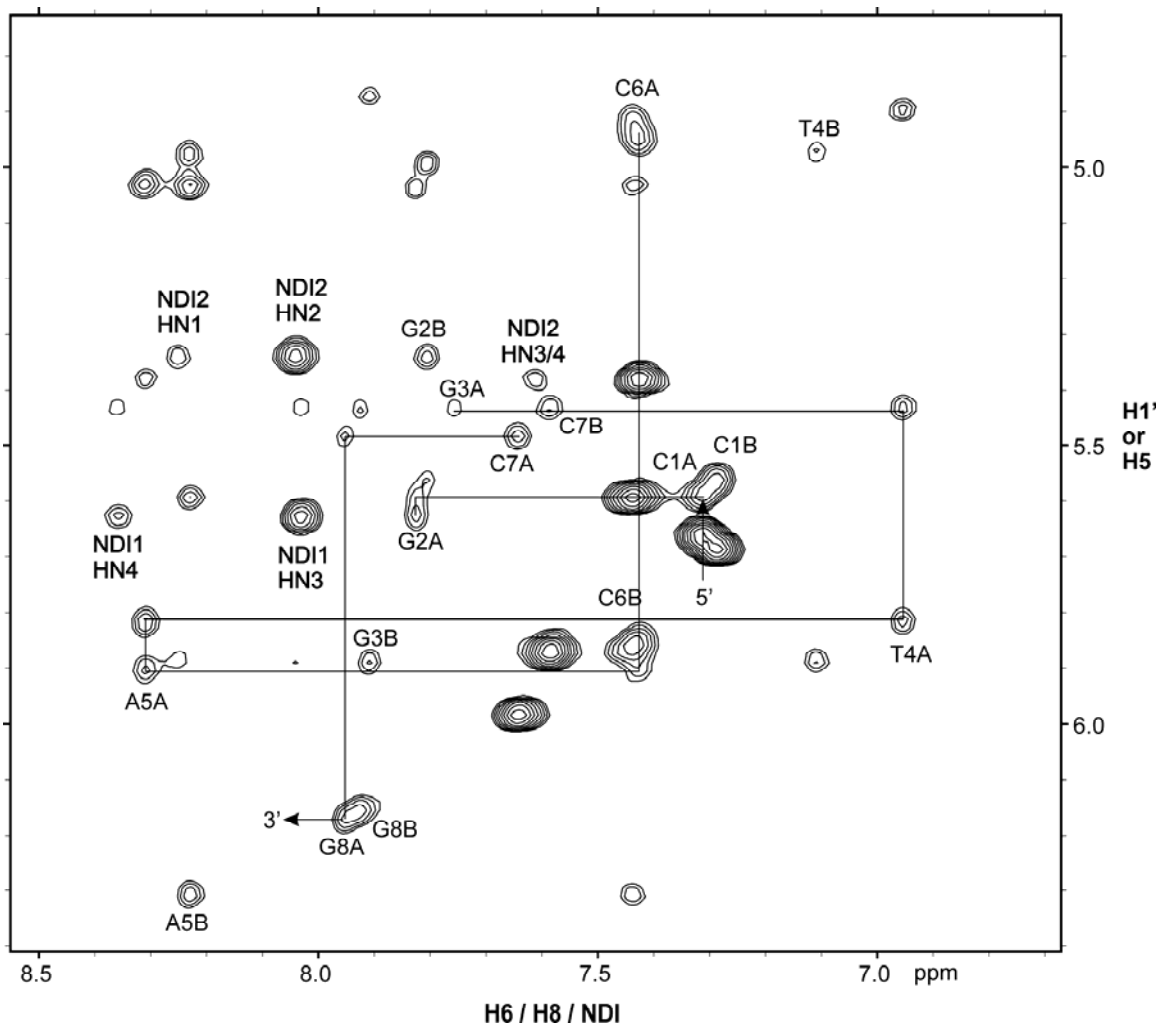
Figure 4.13 Contour plot of the 2D NOSEY spectrum (D$_2$O, 80 ms mixing time, 25 °C) of the **CBI-1**-d(CGGTACCG)$_2$ complex showing the DNA H6/8 to H1' connectivities for strand A. Note that the connectivities are interrupted at the G2/G3 and C6/C7 steps.
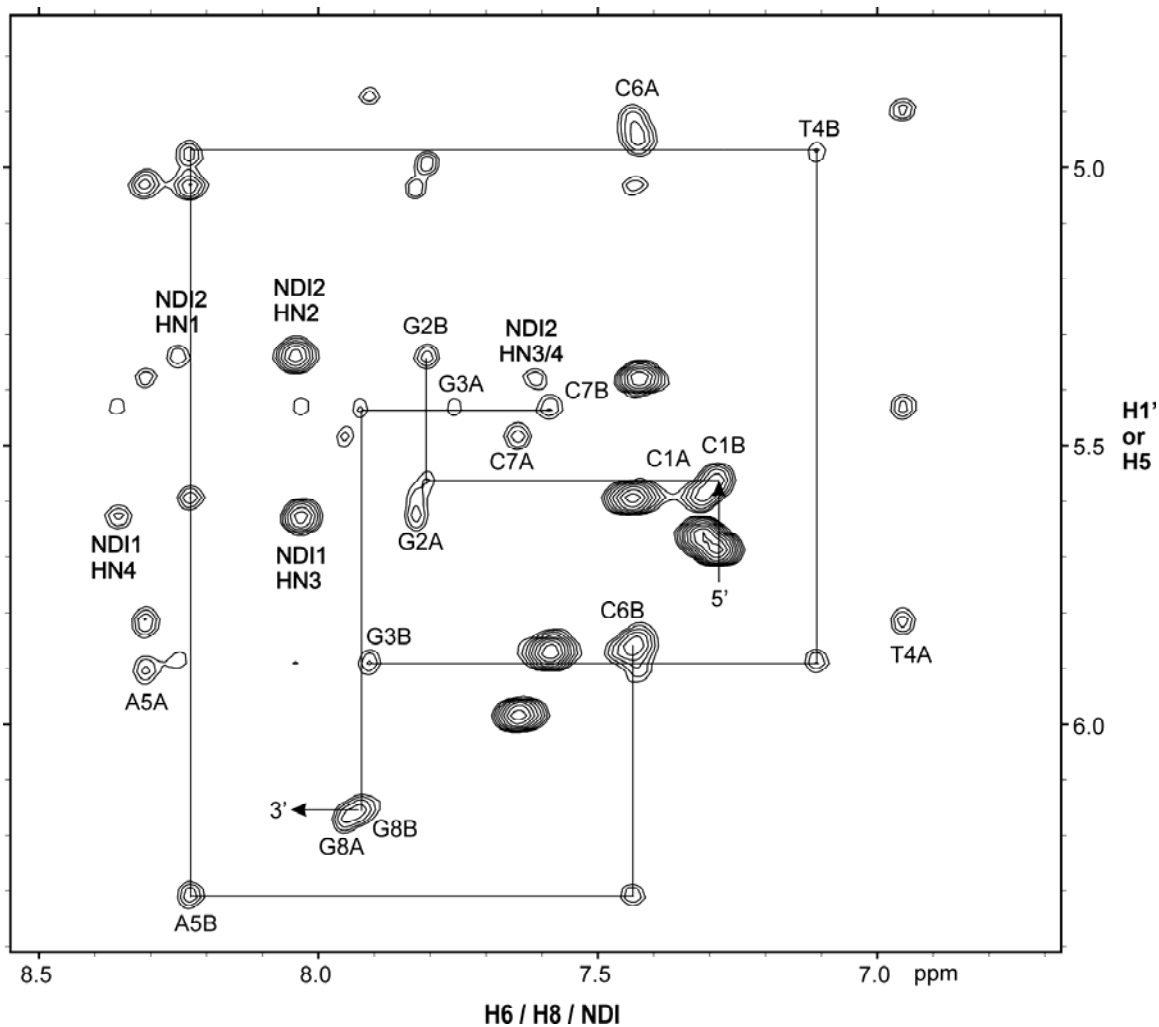
Figure 4.14 Contour plot of the 2D NOSEY spectrum (D$_2$O, 80 ms mixing time, 25 °C) of the **CBI-1**-d(CGGTACCG)$_2$ complex showing the DNA H6/8 to H1' connectivities for strand B. Note that the connectivities are interrupted at the G2/G3 and C6/C7 steps.
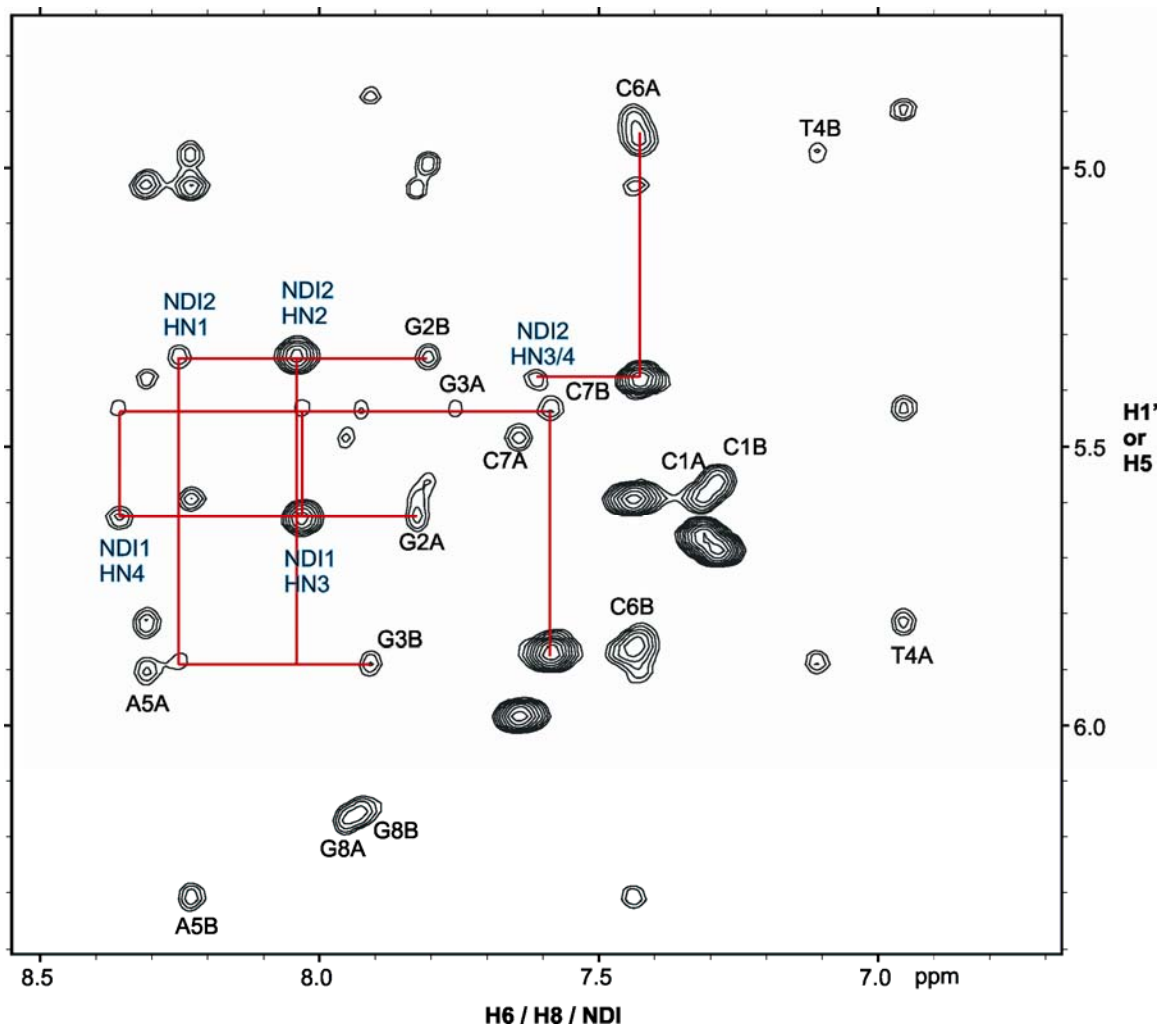
Figure 4.15 Intermolecular NOEs between protons from 2 NDI rings and DNA d(CGGTACCG)$_2$ (H1' and H5) at the intercalation sites (D$_2$O, 80 ms mixing time, 25 °C).
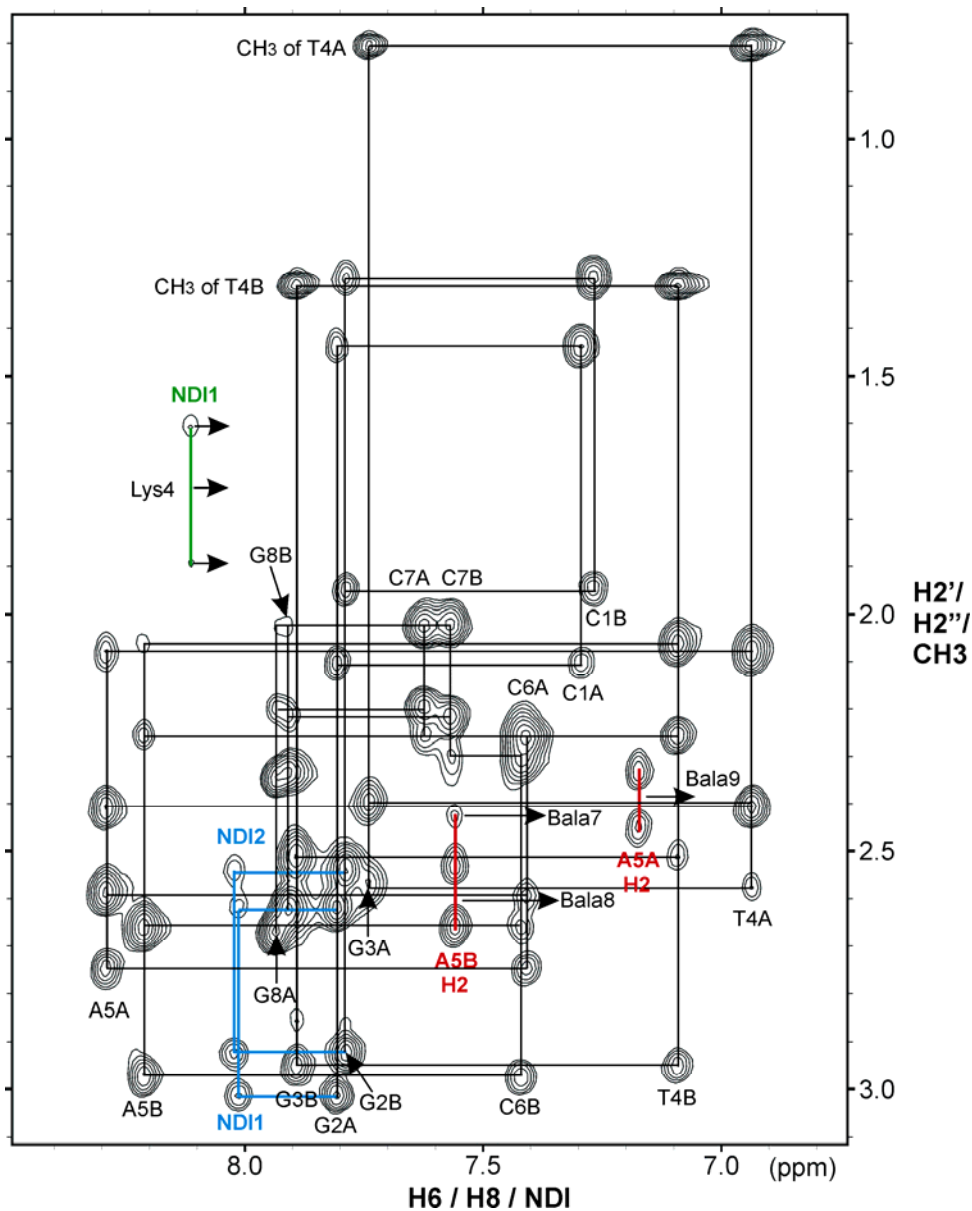
Figure 4.16 Contour plot of the NOESY spectrum (D$_2$O, 80 ms mixing time, 25 °C) of the **CBI-1**-d(CGGTACCG)$_2$ complex showing the DNA H6/H8 to H2'/H2''/CH$_3$ connectivities (black lines). Also shown are intramolecular NOEs between the NDI1 aromatic proton and Lys4 side chain protons (green line) and intermolecular NOEs between NDI aromatic protons and H2'/H2'' of G2 (blue lines), the ligand linker and H2 of the adenine residues (red lines).

As shown in the Figure 4.16, H2 of A5B shows NOE cross peaks with methylene protons of Bala8 and Bala7; H2 of A5A shows NOE cross peaks with methylene protons of Bala9. NOEs were also observed between H2 of A5A, A5B and NH of Bala9; H1' of A5a, A5B and NH of Bala9 (Figure 4.17 (a)). No NOE cross peaks were observed between the methyl group on T4 and any of the protons from the (β-Ala)$_3$Lys linker. Taken together, these interactions confirm that the (β-Ala)$_3$Lys linker resides in the minor groove.

Extensive NOEs were seen between CH$_3$ of T4A or T4B and the amide protons on the Gly$_3$Lys linker as shown in Figure 4.17 (b). For example, strong NOE cross peaks were evident for pairs: Gly1/NH—T4B/Me, Gly2/NH—T4B/Me, Gly2/NH—T4A/Me, Gly3/NH—T4B/Me and Gly3/NH—T4A/Me. Thus NMR structural information has confirmed that the Gly$_3$Lys linker is located in the major groove. Additional evidence includes the observed NOEs between NH$_2$ of C6A and the amide proton (10.26 ppm) of Gly1 on the Gly$_3$Lys linker as well as the NH$_2$ of C6B and amide proton of Lys4 (8.81 ppm) (data not shown).

Weak NOE cross peaks were seen between aromatic protons on NDI1 ring and protons from Lys4 side chain as shown in Figure 4.16. This should put Lys4 close to G3A-C6B base pair, and Lys10 close to C6A-G3B pair. This information further supports our assignment of the spatial positions of eight amino acids relative to the DNA bases in the complex (Figure 4.18).

Overall, the NOE data are consistent with a model in which the NDI units are intercalated into the GG step in a threading manner, with the (β-Ala)$_3$Lys linker residing

in the minor groove and the Gly$_3$Lys linker in the major groove. A structure that matches the model is shown in Figure 4.19, obtained through energy minimization (the DNA coordinates were derived from complex d(CGGTACCG)$_2$-**G$_3$K** NMR structure). It is clear from Figure 4.19 that Gly$_3$Lys fits snugly into the major groove, with methylene groups of the lysine side chain interacting closely with A5B in the major groove. In the minor groove, the β-alanine backbone fits well with the narrow minor groove but the lysine side chain has to point away from the minor groove, due to the chiral character of lysine residue and the orientation of the whole molecule. More rigorous molecular modeling, based on CNS simulation program package is currently underway. It will be interesting to see whether the overall B-form of the DNA is still retained in the complex.
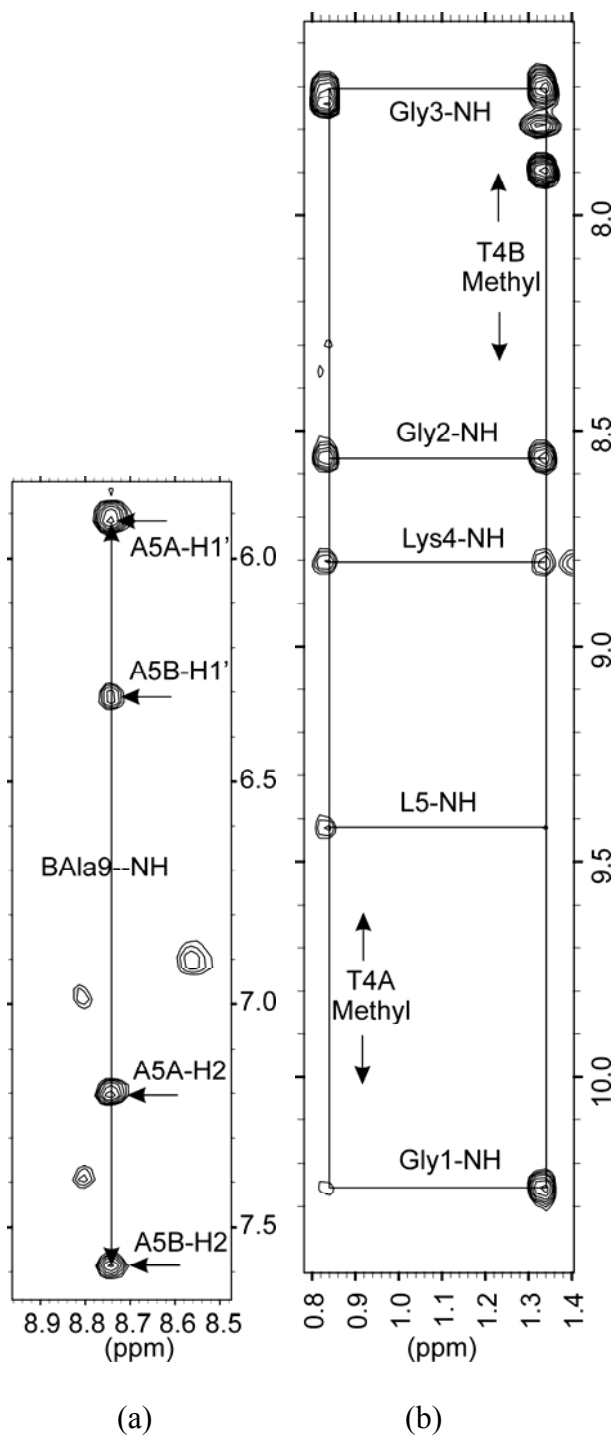
Figure 4.17 Expansion of NOESY spectrum of **CBI-1**-DNA complex. (a) Adenine H2 protons have NOEs with Bala9 residue of the ligand; (b) methyl protons of T4 have NOEs with amide protons of the ligand.
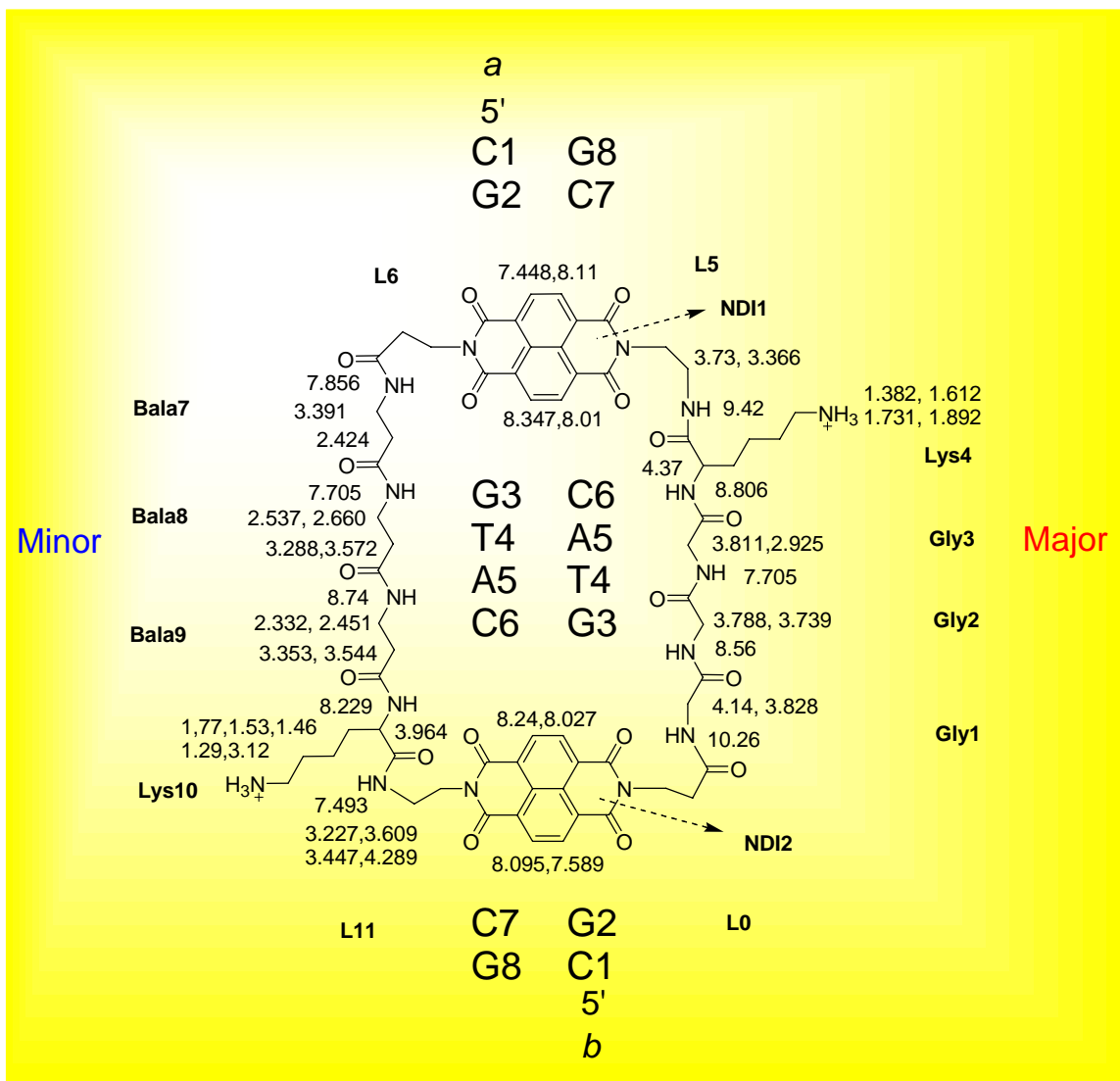
Figure 4.18 Chemical shifts of ligand **CBI-1** protons which have NOE contacts with DNA bases.

<div align="center">(a)   (b)   (c)</div>
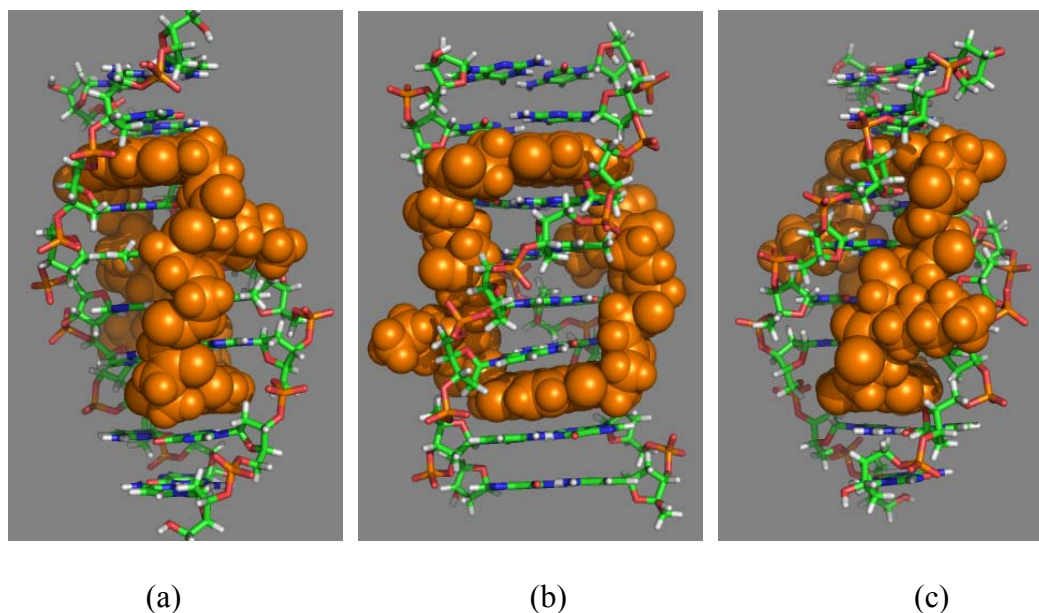
Figure 4.19 A energy-minimized structure of the complex **CBI-1**-d(CGGTACCG)$_2$, based on the model derived from NMR structural information. (a) View toward to the major groove; (b) view from the front; (c) view toward the minor groove. The models were created in PyMol (Delano, 2002).

4.4 Discussion

Selecting NovaSyn TGA resin as the resin to synthesize cyclic compounds proved to be good choice. The on-resin cyclization strategy presented here is an efficient and fast way to make cyclic threading bisintercalators. The purification step, which is usually a time-consuming task, was greatly simplified by the efficient peptide chain elongation and final cyclization. Hopefully, this solid phase synthetic strategy will turn out to be a relatively general approach. We have successfully synthesized a cyclic mono-NDI compound using this strategy (Chu and Iverson, unpublished data). Unfortunately,

despite repeated attempts we were unable to attach the spiro tri-cyclic linker **11** to the resin. Hopefully, conditions can be found in the future that overcome such limitations.

The hypochroimicity observed in the UV-Vis dissociation kinetics study with poly(dGdC) indicated an occurrence of intercalation of the cyclic structure, **CBI-1** between GC pairs. The half-life for the dissociation process after adding SDS detergent was more than 40 minutes, much longer than the time (~ 15min) needed for a typical linear NDI-based dimer to dissociate from the DNA. Considering the inherent flexibility for both linkers, (β-Ala)$_3$Lys and Gly$_3$Lys of **CIB-1**, and the known strong self-stacking tendency of two NDI rings, two unfavorable factors for ligand binding, the observed 41 minutes of half-life time is impressively long, implying a threading intercalation binding mode for this cyclic bisintercalator.

The footprinting data on a 69mer of synthetic DNA clearly showed that the binding between **CBI-1** and 5'-GGTACC-3' sequence is highly specific. The observed dissociation constant is in the sub-micromolar range (K_D ~ 10^{-7} M). Interestingly, **LBI-1** also binds to this sequence only. For bisintercalator **G$_3$K**, it binds not only to the d(GGTACC)$_2$, but also to d(AGTACT)$_2$ with reasonable affinity. Clearly, the hydrophobic, minor-groove preferring (β-Ala)$_3$Lys linking component (Guelev, 2002) altered the binding specificity of both **LBI-1** and **CBI-1**, exhibiting a different specificity from that of Gly$_3$Lys alone. Nevertheless, the footprinting studies confirmed that the Gly$_3$Lys linker dominated the specificity of **CBI-1**, since no binding was seen at the (β-Ala)$_3$Lys preferred 5'-GATAAG-3' sequence.

The specific and tight binding between d(CGGTACCG)$_2$ and **CBI-1** was confirmed by a detailed NMR structural study. For the first time, well-resolved 1D and 2D spectra were obtained for a cyclic bisintercalator binding to DNA duplex. Definite evidence was found to support the model in which **CBI-1** has two NDI rings intercalating between GpG steps, with the linker Gly$_3$Lys interacting with the major groove wall and the linker (β-Ala)$_3$Lys residing in the minor groove, consistent with the initial design.

The most intriguing aspect about this binding is that in order for **CBI-1** to bisintercalate fully into d(CGGTACCG)$_2$, disruption of at least four Watson-Crick base pairs must occur. A mechanistic study concerning the ligand association/dissociation process must await future work. Fortunately, much progress has recently been made in the theoretical study of DNA dynamics (reviewed in Giudice and Lavery 2002; Priyakumar and MacKerell, 2006(a)).

Nogalamycin, a naturally occurring antibiotic having two bulky groups, serves as a good example of a threading intercalator. Structural studies on a d(CGTACG)$_2$-nogalamycin complex revealed that the aglycone side group is located in the minor groove and the amino sugar group is in the major groove, with hydrogen bonds to the N2 and O6 atoms of guanines (Egli, 1991; Williams and Searle, 1999). For the intercalating chromophore of nogalamycin to get into the intercalating site, the two bulky side groups have to effectively pass through and/or open up the intercalation site substantially as does a cyclic bisintercaltor such as **CBI-1**. Within B-DNA, A-T and G-C pairs are typically associated with opening times of 1±10 and 5±50 ms, respectively, while open bases have lifetimes on the order of a few nanoseconds (Guéron and Leroy, 1992). Certain

sequences, such as A tracts (AnTm, where n ≥ 3, m ≥ 0), can lead to an order of magnitude slower exchange rate for thymine (Leroy, 1988), while runs of GC pairs appear to accelerate guanine imino exchange rate (Dornberger, 1999). For **CBI-1** to achieve the final interaction mode with DNA, the four base pairs G-C, T-A, A-T and C-G must be broken. It is unlikely that all four base pair disruptions occur simultaneously. It probably occurs in a step-wise fashion. New hydrogen bonds formed between the DNA grooves and linkers may stabilize the **CBI-1**-DNA complex. In particular, the downfield shifts of amide protons from Gly1 and L5 (10.26 and 9.42 ppm) may indicate two hydrogen bonds formed in the major groove. Similar hydrogen bonds have been routinely observed in other NDI based threading intercalators (Guelev, 2001(a)).

The footprinting data surprisingly showed that **CBI-1** does not have significant affinity for the 5'-GATAAG-3' sequence. Because the (β-Ala)$_3$Lys linker is known to interact strongly with the minor groove of this sequence, it must be that the Gly$_3$Lys linker cannot be accommodated within the major groove of 5'-GATAAG-3'. The lysine side chain of Gly$_3$Lys would be pointing away from the major groove due to the steric hindrance provided by the lysine residue (Figure 4.20(a)) (Guelev, 2002). Modeling has indicated that in order to avoid direct contact with the hydrophilic/polar solvent surface, the hydrophobic lysine side chain might be forced to interact with the pocket formed by three methyl groups through van der Waals/hydrophobic interactions. This pulls the lower part of the ligand toward the top, resulting in the severe deformation of the NDI ring, a highly unfavorable movement energetically. There could also be other factors involved in making this binding orientation unfavorable.

146

(a)   (b)
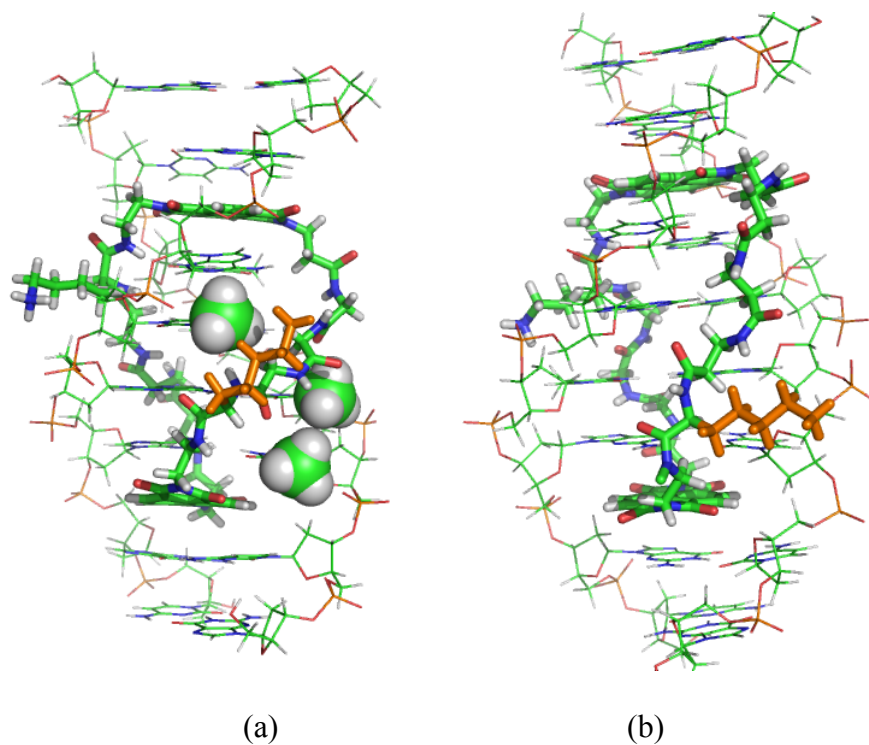
Figure 4.20 (a) Possible binding model between **CBI-1** and d(CGATAAGC)·(GCTTATCG); view toward the major groove. (b) Binding model between **CBI-1** and d(CGGTACCG)$_2$; view toward the minor groove. Lysine side chains were in orange color; methyl groups from thymines were in spacefill model. The figures were created in PyMol (Delano, 2002).

The (β-Ala)$_3$Lys linker is relatively hydrophobic, containing 10 methylene groups between the two intercalating units. It is probably well suited for binding in the narrow and hydrophobic minor groove of an AT rich region, but the lysine side chain has to be pointing out, away from the DNA d(CGGTACCG)$_2$ (Guelev, 2002). Nevertheless, as shown in Figure 4.20 (b), it seems that the lysine side chain can manage to fit in the hydrophobic pocket formed by H1', C1', H4', C4', H5' and C5' on the sugar (Neidle, 2002), without significantly distorting the **CBI-1** structure.

147

In summary, it appears that either the d(GGTACC)$_2$ or d(GATAAG)·(CTTATC) sequences do not match perfectly with **CBI-1** for both the major and minor grooves. More advanced modeling studies based on the NMR data will likely provide more insights regarding the sequence specificity of **CBI-1**. The information gained from the structural study may help to design new cyclic bisintercalators.

4.5 Conclusion

The structural study of a cyclic bisintercalator **CBI-1** binding to d(CGGTACCG)$_2$ sequence has verified that the cyclic NDI-based bisintercalator has the ability to place linkers in the major and minor grooves of the same DNA sequence in a modular fashion. The solid phase synthesis for such compounds turned out to be highly efficient. Current work is directed toward optimizing the linker compositions, so that cyclic bisintercalators can bind targeted sequences with higher affinity and specificity.

4.6 Experimental

4.6.1 General

All chemicals were purchased from Aldrich, Acros, or Novabiochem. Dry solvents were either purchased or dried using common laboratory techniques. Silica gel 60 F$_{254}$ glass-based plates (Merck) were used for TLC. Flash chromatography was carried out using ICN SiliTech 32-63D 60 Å silica gel. All NMR spectra were recorded on Varian 300 or 500 MHz instruments. CDCl$_3$ (δ_H = 7.24 and δ_C = 77.0 ppm) or D$_2$O (δ_H = 4.67 ppm) were used as solvents.

3-[7-(2-tert-Butoxycarbonylamino-ethyl)-1,3,6,8-tetraoxo-3,6,7,8-tetrahydro-1H-benzo[lmn][3,8]phenanthrolin-2-yl]-propionic acid allyl ester (4a, BocNH-(CH$_2$)$_2$-NDI-(CH$_2$)$_2$-CO$_2$All). To a 50 ml round bottom flask was added *N*-(2-tert-butoxycarbonylaminoethyl)-*N'*-(2-carboxyethyl)-1,4,5,8-naphthalenetetracarboylic diimide (Guelev, 2001(b)) (or BocNH-(CH$_2$)$_2$-NDI-(CH$_2$)$_2$-CO$_2$H) (2 g, 4.16 mmol), allyl bromide (20 ml, 110 mmol), and DIPEA (1.5 ml, 8.3 mmol). The flask was heated at 75-80 °C for 4 hours with stirring under Ar atmosphere. The flask was then cooled down to room temperature, followed by addition of 100 ml CHCl$_3$. The organic layer was washed by saturated NaHCO$_3$ (3 × 100 ml), saturated NaCl (3 × 100 ml), dried (anhydrous Na$_2$SO$_4$) and concentrated *in vacuo*. Column chromatography using CH$_2$Cl$_2$/EtOAc (3:1) as the eluting solvents gave a white solid (1.7 g, 79 %, R_f = 0.47 in 3/1 CH$_2$Cl$_2$/EtOAc): ^{1}H NMR (400 MHz, CDCl$_3$): δ 8.71 (s, 4H, aromatic), 5.91-5.81 (m, 1H, allylic), 5.26 (dq, 1H, allylic, J_1 = 1.6 Hz, J_2 = 17.2 Hz), 5.16 (dq, 1H, allylic, J_1 = 1.6 Hz, J_2 = 10.0 Hz), 4.56 (dt, 2H, C*H*$_2$, J_1 = 1.6 Hz, J_2 = 5.6 Hz), 4.49 (t, 2H, C*H*$_2$, J = 7.2 Hz), 4.34 (t, 2H C*H*$_2$, J = 5.6 Hz), 3.55-3.50 (m, 2H, C*H*$_2$), 2.79 (t, 2H, C*H*$_2$, J = 7.2 Hz), 1.19 (s, 9H, C(C*H*$_3$)$_3$); ^{13}C NMR (100 MHz, CDCl$_3$): δ 170.6, 163.0, 162.6, 156.1, 131.9, 131.0, 126.7, 126.6, 126.3, 118.5, 79.2, 65.5, 40.7, 39.1, 38.0, 37.9, 36.5, 32.5, 28.1; HRMS-CI calcd for C$_{27}$H$_{28}$N$_3$O$_8$ [M+H]$^+$ 522.1876, found 522.1876.

3-{7-[2-(9H-Fluoren-9-ylmethoxycarbonyl)-ethyl]-1,3,6,8-tetraoxo-3,6,7,8-tetrahydro-1H-benzo[lmn][3,8]phenanthrolin-2-yl}-propionic acid tert-butyl ester (9a, t-BuO$_2$C-(CH$_2$)$_2$-NDI-(CH$_2$)$_2$-CO$_2$Fm). To a solution of 3-[7-(2-tert-

149

Butoxycarbonyl-ethyl)-1,3,6,8-tetraoxo-3,6,7,8-tetrahydro-1H-

benzo[lmn][3,8]phenanthrolin-2-yl]-propionic acid (or t-BuO$_2$C-(CH$_2$)$_2$-NDI-(CH$_2$)$_2$-

CO$_2$H) (1.7 g, 3.65 mmol) in CH$_2$Cl$_2$ (35 ml) was added Fm-OH (0.65 g, 3.31 mmol),

DIPCDI (0.565 ml, 3.65 mmol) and DMAP (44.5 mg, 0.365 mmol). The cloudy solution

was stirred at room temperature for 4 hours. Vacuum filtration gave a clear solution,

which was concentrated. Flash column chromatography using CH$_2$Cl$_2$/EtOAc (4:1) as the

eluting solvents gave a yellowish solid (1.88 g, 80 %, R_f = 0.71 in 4/1 CH$_2$Cl$_2$/EtOAc):

^{1}H NMR (400 MHz, CDCl$_3$): δ 8.73 (s, 4H, aromatic), 7.70 (d, 2H, aromatic, J = 7.6

Hz), 7.54 (d, 2H, aromatic, J = 7.6 Hz), 7.34 (t, 2H, aromatic, J = 7.6 Hz), 7.22 (t, 2H,

aromatic, J = 7.6 Hz), 4.54 (t, 2H, CH_2, J = 7.2 Hz), 4.46 (t, 2H, CH_2, J = 7.2 Hz), 4.37

(d, 2H, CH_2, J = 7.6 Hz), 4.16 (t, 1H, CH, J = 7.2 Hz), 2.86 (t, 2H, CH_2, J = 7.2 Hz), 2.68

(t, 2H, CH_2, J = 7.6 Hz), 1.41 (s, 9H, C(CH_3)$_3$); ^{13}C NMR (100 MHz, CDCl$_3$): δ 170.9,

170.1, 162.6, 162.5, 143.6, 141.2, 131.1, 131.0, 127.7, 127.0, 126.7, 126.6, 126.4, 125.0,

120.0, 81.0, 66.8, 46.6, 36.7, 36.6, 33.7, 32.6, 28.0; HRMS-CI calcd for C$_{38}$H$_{33}$N$_2$O$_8$

[M+H]$^+$ 645.2237, found 645.2242.

3-{7-[2-(9H-Fluoren-9-ylmethoxycarbonylamino)-ethyl]-1,3,6,8-tetraoxo-3,6,7,8-

tetrahydro-1H-benzo[lmn][3,8]phenanthrolin-2-yl}-propionic acid tert-butyl ester

(13b, t-BuO$_2$C-(CH$_2$)$_2$-NDI-(CH$_2$)$_2$-NHFmoc). To a solution of 1,4,5,8-

naphthalentetracarboxylic diimide (4.0 g, 14.8 mmol) in DMF (200 ml) was added β-

alanine t-butyl ester hydrochloride (2.7g, 14.8 mol) and triethyl amine (4.18 ml, 29.6

mmol). In a microwave machine (MARS, CEM Corporation), the mixture was heated to

140 °C within 2 min, and stayed at 140 °C for 10 more min (P = 300 W, max 100%, S = 3). Upon cooling down the mixture to around 50 °C, *N*-Z-ethylenediamine hydrochloride (3.43 g, 14.8 mmol) and triethyl amine (4.18 ml, 29.6 mmol) were added to the mixture. The mixture was again heated to 140 °C within 2 min, and stayed at 140 °C for another 10 min (P = 600 W, max 100%, S = 3). The mixture was cooled down to room temperature. Solvents were removed by *in vacuo*. Flash column chromatography (3:1 CH_2Cl_2/EtOAc) gave a pink solid of **13a** (3.4 g, 40%, R_f = 0.35 in 3:1 CH_2Cl_2/EtOAc). Then **13a** was dissolved in 1:1 CH_2Cl_2/MeOH (150 ml) and purged with argon. Pd/C (0.8 g, 10 %) was added and the mixture was charged with H_2. After 2 hours of stirring, the reaction mixture was filtered through Celite; the filtrate was concentrated to dryness to give a deep-green solid which was used for next step directly. The solid was dissolved in CH_2Cl_2 (50 ml), followed by the addition of 9-fluorenylmethyl-succinimidyl carbonate (Fmoc-OSu, 2.0 g, 5.9 mmol). Some precipitates formed quickly. After 15 min of stirring of the mixture at room temperature, the reaction mixture was filtered to give a clear filtrate. The filtrate was concentrated and flash column chromatography (3:1 CH_2Cl_2 / EtOAc) gave a pure yellowish solid of **13b** (1.55 g, 40 % in two steps, R_f = 0.52 in 3:1 CH_2Cl_2/EtOAc): ^{1}H NMR (300 MHz, $CDCl_3$): δ 8.64 (d, 2H, aromatic, J = 7.5 Hz), 8.58 (d, 2H, aromatic, J = 7.5 Hz), 7.67 (d, 2H, aromatic, J = 7.2 Hz), 7.41-7.32 (m, 4H, aromatic), 7.22 (t, 2H, aromatic, J = 7.2 Hz), 5.16 (bs, 1H, N*H*CO), 4.44-4.39 (m, 4H, 2C*H*₂), 4.12-4.09 (m, 2H, C*H*₂), 3.70-3.66 (m, 3H, C*H*₂ and C*H*), 2.65 (t, 2H, CH₂, J = 7.8 Hz), 1.40 (s, 9H, C(C*H*₃)₃); HRMS-CI calcd for $C_{38}H_{34}N_3O_8$ [M+H]$^+$ 660.2346, found 660.2349.

4.6.2 Solid phase synthesis

Resin loading. NovaSyn TGA resin (1.0g, loading 0.28 mmol/g) was swollen with DMF for 30 min in a 3-way glass reactor. DSC (717 mg, 10 equiv) and DMAP (34 mg, 1.0 equiv) in DMF (7 ml) were added to the resin and the reaction went on for 2 hr at room temperature with continuous mechanical shaking. The resin was washed with DMF (6X), followed by addition of a solution of Fmoc-Lys(Boc)-OAll (1.2 g, 10 equiv), DIPEA (0.975 ml, 20 equiv) in DMF (6ml). After 6 hr of mechanical shaking the reaction was stopped by draining, then washing with DMF (6X), DCM (4X), DMF (4X). Fmoc protecting group was cleaved off by 20% piperidine/DMF solution (3×6 ml $\times 10$ min). The resin loading can be estimated by determining the concentration of Fmoc cleavage product in the solution. Using an extinction coefficient of 7800 at wavelength of 300 nm, through Beer's Law, the resin loading was estimated to be 0.17 mmol/g.

Linear and cyclic bisintercalators synthesis. For amino acid coupling the conditions were 5 equiv of Fmoc-protected amino acid (relative to the determined resin loading), 4.9 equiv of HATU and 10 equiv of DIPEA in 6 ml of DMF for 50 min at room temperature. The unreacted amino termini were capped with 5% (v/v) acetic anhydride-5% (v/v) 2,6-lutidine in NMP (30 min). 20% Piperidine/DMF solution (3×6 ml $\times 10$ min) was used to cleave off the Fmoc group. Between every two steps, the resin was washed with DMF (6X), DCM (4X), and DMF (4X) successively. For the coupling of Fmoc-Lys(Boc)-NDI abduct to the resin, NDI abduct (3 equiv), PyBOP (3 equiv) and HOBT (3 equiv) were dissolved in NMP (6ml) completely. This solution and N-methylmorpholine (NMM, 6

equiv) were added to the resin and the coupling reaction proceeded for 50 min. After the solvent was drained, the resin was washed with NMP (6X) and the coupling was repeated. The resin was washed again with NMP (6X) after the second coupling, followed by the DMF-DCM-DMF washes. De-allylation was carried out by treating the resin with Pd(PPh$_3$)$_4$ (3 equiv) in CHCl$_3$/AcOH/NMM (v/v, 37:2:1) (15 ml /g of resin) under argon for 2.5 hr at room temperature. The resin was then washed with 0.5% (v/v) DIPEA/DMF, 0.5% (w/w) sodium diethyldithiocarbamate / DMF and again 0.5% (v/v) DIPEA/DMF, followed by the usual DMF-DCM-DMF washes. The final cyclization was initiated by adding a solution of HOAT (5 equiv), PyBOP (5 equiv) and DIPEA (10 equiv) in 6 ml of NMP to the resin. After 18 hr, the resin was negative to the Kaiser ninhydrin test. Final cleavage of the linear or cyclic products were performed with 95/2.5/2.5 TFA/triisopropyl silane (TIPS)/H$_2$O (v/v) at 25 °C for 1.5 hr. The TFA solution is concentrated to dryness, followed by addition of 10 ml of H$_2$O. The aqueous solution was washed three times with diethyl ether and lyophilized overnight. The crude linear or cyclic bisintercalators were dissolved in 0.1% TFA/H$_2$O and purified by reverse-phase preparative HPLC (Waters, C18 VYDAC HPLC column, Cat.# 218TP1022) using 0.1% TFA/water (v/v) as solvent A and 0.1% TFA/ACN (v/v) as solvent B. In order to get extremely pure compounds, two runs of purifications were carried out. For the first run, a gradient of 5% of B to 95% of B over 40 min was applied. For the second run, a typical gradient is composed of three stages: 5% B to 12% B over 1 min; 12% B to 25% B over 43 min; 25% B to 95% B over 3 min. The product fractions were combined and lyophilized to yield the linear or cyclic bisintercalators (~1.5% overall) as white powder.
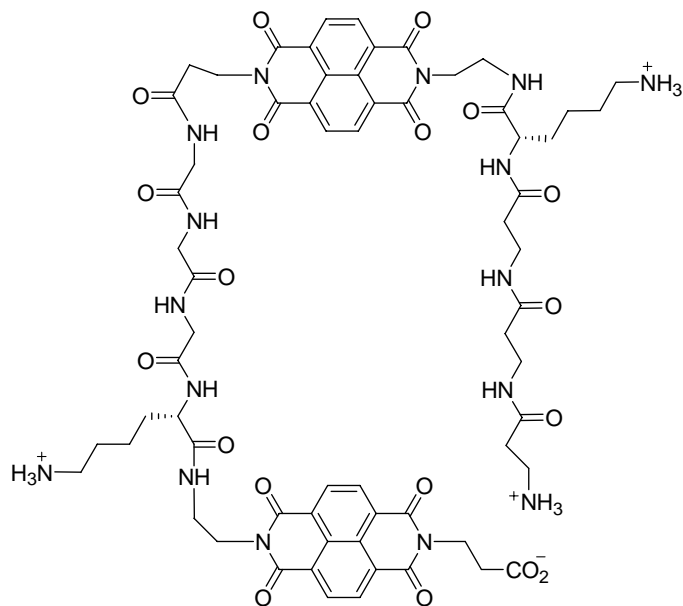
LBI-1

^{1}H NMR (500 MHz, D$_2$O): δ 8.40 (d, 2H, J = 7.5 Hz), 8.33 (d, 2H, J = 7.5 Hz), 8.24 (d, 2H, J = 7.5 Hz), 8.19 (d, 2H, J = 7.5 Hz), 4.40-4.32 (m, 4H), 4.10-3.97(m, 6H), 3.90 (d, 2H, J = 4.5 Hz), 3.88-3.85 (m, 2H), 3.50-3.38 (m, 4H), 3.29-3.26 (m, 2H), 3.16 (t, 2H, J = 6.5 Hz), 3.05-3.00 (m, 1H), 2.96-2.91(m, 1H), 2.85 (t, 4H, J = 7.5 Hz), 2.81-2.76 (m, 6H), 2.55 (t, 2H, J = 6.5 Hz), 2.25 (t, 2H, J = 6.5 Hz), 2.21-2.10 (m, 2H), 1.65-1.11 (m, 12H); LCMS (ESI) m/z found 1385.3 (MH$^+$), 693.5 (M2H^{+2}), predicted 1385.5(MH$^+$);

154
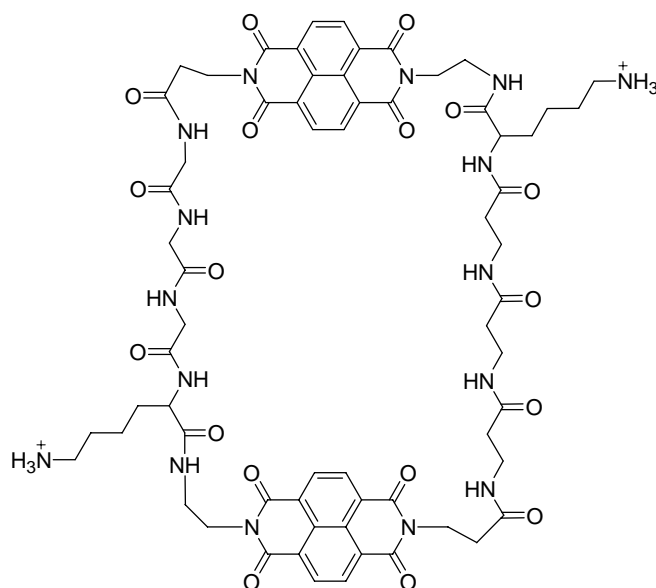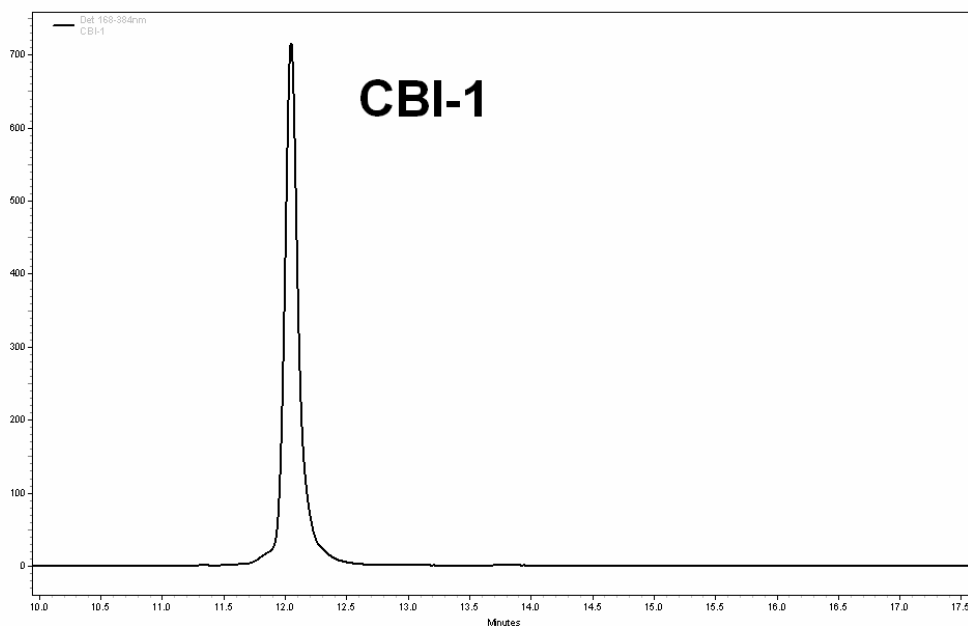
CBI-1

HPLC profile of the purified CBI-1

^{1}H NMR (500 MHz, D$_2$O): δ 8.38 (d, 2H, J = 7.5 Hz), 8.36 (d, 2H, J = 7.5 Hz), 8.24 (d, 2H, J = 7.5 Hz), 8.17 (d, 2H, J = 7.5 Hz), 4.36 (t, 2H, J = 7.0 Hz), 4.28 (t, 2H, J = 7.0 Hz), 4.10-3.94 (m, 6H), 3.92-3.86 (m, 4H), 3.80-3.77 (m, 2H), 3.50-3.25 (m, 10H), 2.91-

155

2.77 (m, 6H), 2.62 (t, 2H, J = 7.0 Hz), 2.45-2.42 (m, 2H), 2.39 (t, 2H, J = 6.5 Hz), 2.33-2.29 (m, 2H), 1.66-1.18 (m, 12H); LCMS (ESI) m/z found 1367.4 (MH$^+$), 684.5 (M2H^{+2}), predicted 1367.5(MH$^+$).

4.6.3 Dissociation kinetics

In a typical 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 first incubated with 500 µl of 0.028 mM compound stock solution for 2 hr. Next, 0.5 mL of the resulting DNA/compound complex solution was mixed with 0.5 mL of 4% SDS solution in a quartz cuvette with 1 cm path length. The UV absorbance was monitored at 385 nm after 20 seconds of mixing (Hewlet-Packard 8553 Multitransport UV-vis spectrophotometer). The error estimate was determined to be ± 8% for this experiment.

4.6.4 DNAse I footprinting

The 69 bp synthetic fragment (PAGE grade) was purchased from Midland Certified (Midland, TX) and labeled (^{32}P) as described in Chapter 2. The DNAse I (Amersham Biosciences, Piscataway, NJ) footprinting was carried out according to the procedure described previously (Lee, 2004(b)). The DNA fragments were separated on a 12 % denaturing polyacrylamide gel. The gels were exposed on a phosphor screen for 24 - 48 hours and the image was analyzed with Quantity One 4.5 software from Bio-Rad (Hercules, CA).

4.6.5 NMR sample preparation and NMR spectroscopy

The DNA (gel filtration grade, Midland Certified, Midland, TX) was initially dissolved in 0.7 ml 30 mM Na phosphate buffer, pH 7.5. Prior to addition of **CBI-1**, the d(CGGTACCG)$_2$ sample was diluted to 10 ml in chilled water to which ligand was added and the sample was lyophilized. Samples used for NMR spectra in D$_2$O solvent were lyophilized twice from D$_2$O and finally dissolved in 0.7 ml 99.9% D$_2$O (Cambridge Isotope, Cambridge, MA, USA). The final concentration of the complex used in obtaining NMR spectra was approximately 1 mM.

Spectra were obtained using a 500 MHz Varian Unity-Inova spectrometer. 2D NOESY (80 and 250 ms mixing time), TOCSY (50 ms mixing time) and COSY spectra were acquired in D$_2$O solvent at 25 °C using a sweep width of 8000 Hz, with 2048 complex points (t2) and 512 (t1) complex points being acquired in each dimension; presaturation was used to remove the residual HOD signal. To observe solvent-exchangeable protons, NOSEY (10 °C, 125 ms mixing time), NOESY (25 °C, 80 and 250 ms mixing time) and TOCSY (25 °C) spectra were acquired in 9:1 H$_2$O:D$_2$O solvent, using a sweep width of 12210 Hz; NOESY spectra were acquired using the jump-return method for solvent suppression, so as not to saturate the signals of protons that exchange rapidly with the solvent. NOSEY spectra at ^{1}H spectra were referenced to the H$_2$O resonance at 4.75 ppm at 25 °C. The 1D spectra were processed using VNMR software (Varian). 2D spectra were processed using NmrPipe (Delaglio, 1995) and displayed using Sparky (Goddard and Kneller, 2004).

4.6.6 Computer simulation

Molecular geometry optimization computations were performed in HyperChem 7.0 (Hypercube Inc., 1115 NW 4th Street, Gainesville, FL 32601) using the Amber force field. DNA PDB coordinates which were derived directly from NMR structures of **G₃K**-d(CGGTACCG)₂ or **(β–A)₃K**-d(CGATAAGC)·(GCTTATCG) were used in the minimization and the linker segment was manually modified as described in the text (Guelev, 2001(a); Guelev, 2002). During the calculation, DNA coordinates were fixed and the geometry optimization was performed only on the ligand by using the Fletcher-Reeves conjugate gradient algorithm, with a convergence cutoff value of 0.01 kcal·mol^{-1}.

Chapter 5   Future Directions

Future studies may be focused on two areas. The first one would be involved in expanding the linker rigidification research project. The second area would be to systematically study how the changes made on the chemical structures of the linkers of cyclic bisintecalators will affect the DNA binding specificity and affinity.

The rigidified bis-intercalator **C1** with a spiro, tricyclic and unfunctionalized linker structure was shown to bind all GGTACC, GGGCCC, GGATCC, and GGCGCC sequences with good affinities, although highest for GGTACC sequence. It is possible to attach a few functional groups to the central cyclohexyl ring to make it selective for one sequence over the others. For example, preliminary modeling studies indicated that by putting two CH_2OH groups at opposite positions on this cyclohexane ring (see the structures below), the resulting bisintercalator can fit well within the minor groove of GGCGCC sequence, but not the others mentioned above. The synthesis and purification (of diastereomers) could be a challenge, but the resulting linker and even dimer will be easier to handle because of its enhanced solubility in organic or aqueous solvents than the unfunctionalized linker. Thus, the yield for the final coupling step (to the NDI units) will be significantly increased. 5'-GGCGCC-3' is an important DNA sequence which is part of the primary binding site of the long terminal repeat of HIV-1 retrovirus (Wain-Hobson, 1985; Ratner, 1985; Mohammadi, 1998).
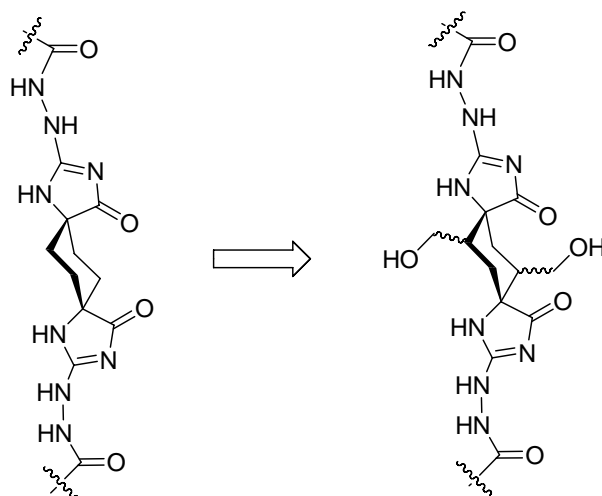
The discovery of a specific cyclic bisintercalator **CBI-1** for d(CGGTACCG)$_2$ demonstrates a new approach to target DNA. The linker compositions for both linkers of CBI-1 may be altered rationally. Currently, it appears that changing composition of minor groove recognition linker (β-Ala)$_3$Lys looks very promising. A quick project would involve changing the L-lysine residue within the (β-Ala)$_3$Lys linker to glysine or D-lysine to see if the binding affinity can be increased for d(GGTACC)$_2$. This sequence (occurring only once) is part of gene sequence of a key bifunctional enzyme which regulates the cellular level of fructose-2,6-bisphosphate, a powerful activator of glycolysis. Over-expression of this enzyme has been related to the breast and colon malignant tumor growth (Minchenko, 2005).

In addition, the solid phase synthesis strategy developed for the synthesis of **CBI-1** could be easily adapted to synthesize all kinds of DAN-NDI (donor-acceptor) mixed cyclic structures.

Appendix A: Topology file for dimer C1 used in CNS simulation

```
!TOPOlogy
set echo=false end
MASS CT   12.01100
MASS HA   1.00800
MASS NX1  14.00700
MASS CX2  12.01100
MASS OX3  15.99900
MASS CB   12.01100
MASS CA   12.01100
MASS C    12.01100
MASS O    15.99900
mass HC   1.008
mass NH3  14.007

{ Note: edit masses if necessary }
 MASS NX7    14.00700 ! assuming N -> 14.00700 + 1.008 * 0 (Hs)
 MASS HX8     1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS NX9    14.00700 ! assuming N -> 14.00700 + 1.008 * 0 (Hs)
 MASS HX10    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS CX11   12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS NX12   14.00700 ! assuming N -> 14.00700 + 1.008 * 0 (Hs)
 MASS HX13    1.00800 ! assuming H -> 1.00800 + 1.008 * 0 (Hs)
 MASS NX14   14.00700 ! assuming N -> 14.00700 + 1.008 * 1 (Hs)
 MASS CX15   12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS OX16   15.99900 ! assuming O -> 15.99900 + 1.008 * 0 (Hs)
 MASS CX17   12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS CX18   12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS CX21   12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS CX24   12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS CX27   12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)
 MASS CX30   12.01100 ! assuming C -> 12.01100 + 1.008 * 0 (Hs)

AUTOgenerate ANGLes=TRUE END

RESIdue NDI
GROUp
 ATOM NZ    TYPE NH3        CHARge -0.043 END
 ATOM HZ1   TYPE HC         CHARge  0.257 END
 ATOM HZ2   TYPE HC         CHARge  0.257 END
 ATOM HZ3   TYPE HC         CHARge  0.257 END
 ATOM CA    TYPE CT         CHARge -0.155 END
 ATOM HA1   TYPE HA         CHARGE  0.161 END
 ATOM HA2   TYPE HA         CHARGE  0.161 END
 ATOM CB    TYPE CT         CHARge  0.037 END ! based on THY charges
 ATOM HB1   TYPE HA         CHARGE  0.155 END ! based on THY charges
 ATOM HB2   TYPE HA         CHARGE  0.155 END ! based on THY charges
 ATOM NI1   TYPE NX1        CHARge -0.373 END ! based on THY charges
 ATOM CI1   TYPE CX2        CHARge  0.361 END ! same
 ATOM CI2   TYPE CX2        CHARge  0.361 END ! same
 ATOM OI1   TYPE OX3        CHARge -0.331 END ! same
```

```
ATOM OI2   TYPE OX3        CHARge -0.331 END ! same
ATOM CN1   TYPE CB         CHARge  0.016 END !MIDDLE OF RING
ATOM CN2   TYPE CB         CHARge -0.015 END !  ''
ATOM CN3   TYPE CB         CHARge -0.132 END !NEXT TO CARBONYL
ATOM CN4   TYPE CB         CHARge -0.058 END !   "
ATOM CN5   TYPE CB         CHARge -0.132 END !   "
ATOM CN6   TYPE CB         CHARge -0.058 END !   "
ATOM CN7   TYPE CA         CHARge -0.015 END ! OUTSIDE
ATOM CN8   TYPE CA         CHARge -0.058 END ! "
ATOM CN9   TYPE CA         CHARge -0.015 END ! "
ATOM CN0   TYPE CA         CHARge -0.058 END ! "
ATOM HN1   TYPE HA         CHARGE  0.157 END
ATOM HN2   TYPE HA         CHARGE  0.166 END
ATOM HN3   TYPE HA         CHARGE  0.157 END
ATOM HN4   TYPE HA         CHARGE  0.166 END
ATOM NI2   TYPE NX1        CHARge -0.332 END
ATOM CI3   TYPE CX2        CHARge  0.342 END
ATOM CI4   TYPE CX2        CHARge  0.342 END
ATOM OI3   TYPE OX3        CHARge -0.285 END
ATOM OI4   TYPE OX3        CHARge -0.285 END
ATOM CC    TYPE CT         CHARge  0.0   END
ATOM HC1   TYPE HA         CHARge  0.132 END
ATOM HC2   TYPE HA         CHARge  0.132 END
ATOM CD    TYPE CT         CHARge -0.194  END
ATOM HD1   TYPE HA         CHARge  0.117  END
ATOM HD2   TYPE HA         CHARge  0.117  END
ATOM C     TYPE C          CHARge  0.288 END
ATOM O     TYPE O          CHARge -0.371 END

BOND NZ HZ1     BOND NZ HZ2      BOND NZ HZ3       BOND NZ   CA
BOND CA CB      BOND CB NI1      BOND CA HA1       BOND CA HA2
BOND CB HB1     BOND CB HB2      BOND NI1 CI1
BOND NI1 CI2    BOND CI1 OI1     BOND CI2 OI2      BOND CI1 CN3
BOND CI2 CN5    BOND CN3 CN7     BOND CN3 CN1      BOND CN1 CN2
BOND CN5 CN1    BOND CN5 CN9     BOND CN9 CN0      BOND CN7 CN8
BOND CN8 CN4    BOND CN2 CN4     BOND CN2 CN6      BOND CN0 CN6
BOND CN6 CI4    BOND CI4 NI2     BOND CN4 CI3      BOND CI3 NI2
BOND CN7 HN1    BOND CN8 HN2     BOND CN9 HN3      BOND CN0 HN4
BOND CI3 OI3    BOND CI4 OI4     BOND NI2 CC       BOND CC CD
BOND CC HC1     BOND CC HC2      BOND CD HD1       BOND CD HD2
BOND CD C       BOND C O


DIHEDRAL CN5 CN1 CN2 CN4     !180.0
DIHEDRAL CN3 CN1 CN2 CN6     !180.0

DIHEDRAL OI1 CI1 NI1 CB      !  0.0
DIHEDRAL OI2 CI2 NI1 CB      !  0.0
DIHEDRAL OI3 CI3 NI2 CC      !  0.0
DIHEDRAL OI4 CI4 NI2 CC      !  0.0

DIHEDRAL OI1 CI1 CN3 CN1   !180.0
DIHEDRAL OI2 CI2 CN5 CN1   !180.0
```

```
DIHEDRAL OI3 CI3 CN4 CN2    !180.0
DIHEDRAL OI4 CI4 CN6 CN2    !180.0

DIHEDRAL HN1 CN7 CN8 HN2     !0.0
DIHEDRAL HN3 CN9 CN0 HN4     !0.0

DIHEDRAL CI1 CN3 CN1 CN2      !180
DIHEDRAL CI2 CN5 CN1 CN2      !180
DIHEDRAL CI3 CN4 CN2 CN1      !180
DIHEDRAL CI4 CN6 CN2 CN1      !180

DIHEDRAL CN3 CN7 CN8 CN4     !0
DIHEDRAL CN5 CN9 CN0 CN6     !0

DIHEDRAL CN1 CN2 CN4 CN8     !0
DIHEDRAL CN1 CN2 CN6 CN0     !0
DIHEDRAL CN2 CN1 CN3 CN7     !0
DIHEDRAL CN2 CN1 CN5 CN9     !0

DIHEDRAL HN1 CN7 CN3 CN1     !180
DIHEDRAL HN2 CN8 CN4 CN2     !180
DIHEDRAL HN3 CN9 CN5 CN1     !180
DIHEDRAL HN4 CN0 CN6 CN2     !180

DIHEDRAL CI1 NI1 CB CA    !0.0
DIHEDRAL CI3 NI2 CC CD    !0.0

DIHEDRAL NI2 CC CD C      !0.0
DIHEDRAL CC CD C O        !0.0

DIHEDRAL NI1 CB CA NZ      !0.00   {sd = 0.031}   3   0.0000

IMPRoper  HN3  HN4  HN2  HN1 ! Needs to be zero
IMPRoper  OI1  OI2  OI4  OI3 ! Needs to be zero
IMPRoper  CN3 CN5 CN6 CN4 ! Needs to be zero
IMPRoper  NI1 CN1 CN9 CN0 !NEEDS TO BE 180
IMPRoper  NI2 CN2 CN8 CN7 !NEEDS TO BE 180
IMPRoper  CI1 CI2 CI4 CI3 !NEEDS TO BE ZERO

IMPRoper  HZ1 HZ2 CA HZ3 !methyl NZ  500.00   0   -66.4262


 ACCEptor  O    C
 ACCEptor  OI1    CN1
 ACCEptor  OI2    CN2
 ACCEptor  OI3    CN3
 ACCEptor  OI4    CN4
END { NDI }

!spiro-linker
RESIdue SPI

 { Note: electrostatics should normally not be used in }
```

```
{ crystallographic refinement since it can produce }
{ artefacts. For this reason, all charges are set to }
{ zero by default. Edit them if necessary }

GROUp
 ATOM  NAK   TYPE NX7    CHARge  -0.244   END ! Nr of Hs =   0
 ATOM  HAA   TYPE HX8    CHARge   0.299   END ! Nr of Hs =   0
 ATOM  NAL   TYPE NX9    CHARge  -0.182   END ! Nr of Hs =   0
 ATOM  HAB   TYPE HX10   CHARge   0.285   END ! Nr of Hs =   0
 ATOM  CAM   TYPE CX11   CHARge   0.263   END ! Nr of Hs =   0
 ATOM  NAS   TYPE NX12   CHARge  -0.365   END ! Nr of Hs =   0
 ATOM  HAG   TYPE HX13   CHARge   0.233   END ! Nr of Hs =   0
 ATOM  NAD   TYPE NX14   CHARge  -0.358   END ! Nr of Hs =   1
 ATOM  CAE   TYPE CX15   CHARge   0.238   END ! Nr of Hs =   0
 ATOM  OAA   TYPE OX16   CHARge  -0.329   END ! Nr of Hs =   0
 ATOM  CAT   TYPE CX17   CHARge   0.039   END ! Nr of Hs =   0
 ATOM  CAY   TYPE CX18   CHARge  -0.145   END ! Nr of Hs =   0
 ATOM 1HAY   TYPE HA     CHARge   0.108   END ! Nr of Hs =   0
 ATOM 2HAY   TYPE HA     CHARge   0.108   END ! Nr of Hs =   0
 ATOM  CAZ   TYPE CX21   CHARge  -0.145   END ! Nr of Hs =   0
 ATOM 1HAZ   TYPE HA     CHARge   0.108   END ! Nr of Hs =   0
 ATOM 2HAZ   TYPE HA     CHARge   0.108   END ! Nr of Hs =   0
 ATOM  CAU   TYPE CX24   CHARge  -0.145   END ! Nr of Hs =   0
 ATOM 1HAU   TYPE HA     CHARge   0.108   END ! Nr of Hs =   0
 ATOM 2HAU   TYPE HA     CHARge   0.108   END ! Nr of Hs =   0
 ATOM  CAV   TYPE CX27   CHARge  -0.145   END ! Nr of Hs =   0
 ATOM 1HAV   TYPE HA     CHARge   0.108   END ! Nr of Hs =   0
 ATOM 2HAV   TYPE HA     CHARge   0.108   END ! Nr of Hs =   0
 ATOM  CAW   TYPE CX30   CHARge   0.039   END ! Nr of Hs =   0
 ATOM  NAX   TYPE NX12   CHARge  -0.365   END ! Nr of Hs =   0
 ATOM  HAH   TYPE HX13   CHARge   0.233   END ! Nr of Hs =   0
 ATOM  CAF   TYPE CX15   CHARge   0.238   END ! Nr of Hs =   0
 ATOM  OAB   TYPE OX16   CHARge  -0.329   END ! Nr of Hs =   0
 ATOM  NAG   TYPE NX14   CHARge  -0.358   END ! Nr of Hs =   1
 ATOM  CAN   TYPE CX11   CHARge   0.263   END ! Nr of Hs =   0
 ATOM  NAO   TYPE NX9    CHARge  -0.182   END ! Nr of Hs =   0
 ATOM  HAE   TYPE HX10   CHARge   0.285   END ! Nr of Hs =   0
 ATOM  NAP   TYPE NX7    CHARge  -0.244   END ! Nr of Hs =   0
 ATOM  HAF   TYPE HX8    CHARge   0.299   END ! Nr of Hs =   0

BOND  NAK   HAA        BOND  NAK   NAL        BOND  NAL   HAB
BOND  NAL   CAM        BOND  CAM   NAS        BOND  CAM   NAD
BOND  NAS   HAG        BOND  NAS   CAT        BOND  NAD   CAE
BOND  CAE   OAA        BOND  CAE   CAT        BOND  CAT   CAY
BOND  CAT   CAU        BOND  CAY  1HAY        BOND  CAY  2HAY
BOND  CAY   CAZ        BOND  CAZ  1HAZ        BOND  CAZ  2HAZ
BOND  CAZ   CAW        BOND  CAU  1HAU        BOND  CAU  2HAU
BOND  CAU   CAV        BOND  CAV  1HAV        BOND  CAV  2HAV
BOND  CAV   CAW        BOND  CAW   NAX        BOND  CAW   CAF
BOND  NAX   HAH        BOND  NAX   CAN        BOND  CAF   OAB
BOND  CAF   NAG        BOND  NAG   CAN        BOND  CAN   NAO
BOND  NAO   HAE        BOND  NAO   NAP        BOND  NAP   HAF
```

```
{ Note: edit these DIHEdrals if necessary }

!DIHEdral  NAL  CAM  NAS  HAG ! flat ? (0 degrees = cis) 1.47
!DIHEdral  NAL  CAM  NAS  CAT ! flat ? (180 degrees = trans)181.49
!DIHEdral  NAL  CAM  NAD  CAE ! flat ? (180 degrees = trans)179.70
!DIHEdral  HAG  NAS  CAT  CAE ! flat ? (180 degrees = trans)176.22
!DIHEdral  CAM  NAD  CAE  OAA ! flat ? (180 degrees = trans)180.44
!DIHEdral  OAA  CAE  CAT  NAS ! flat ? (180 degrees = trans)182.12


!DIHEdral  CAF  CAW  NAX  HAH ! flat ? (180 degrees = trans)181.36
!DIHEdral  NAX  CAW  CAF  OAB ! flat ? (180 degrees = trans)177.72
!DIHEdral  CAW  NAX  CAN  NAO ! flat ? (180 degrees = trans)178.69
!DIHEdral  HAH  NAX  CAN  NAO ! flat ? (0 degrees = cis)    -1.34
!DIHEdral  OAB  CAF  NAG  CAN ! flat ? (180 degrees = trans)182.04
!DIHEdral  CAF  NAG  CAN  NAO ! flat ? (180 degrees = trans)180.09


!DIHEdral  NAX  CAN  NAO  NAP ! flat ? (180 degrees = trans)180.00
!DIHEdral  NAG  CAN  NAO  NAP ! flat ? (0 degrees = cis)   0.00
!DIHEdral  HAE  NAO  NAP  HAF ! flat?(180 degrees = trans)180.00
!DIHEdral  CAN  NAO  NAP  HAF ! flat ? (0 degrees = cis) 0.00




{ Note: edit these IMPRopers if necessary }
IMPRoper  NAL  NAK  HAB  CAM ! chirality or flatness 0.02
IMPRoper  CAM  NAL  NAS  NAD ! chirality or flatness -2.07
IMPRoper  NAS  CAM  HAG  CAT ! chirality or flatness 0.01
IMPRoper  CAE  NAD  OAA  CAT ! chirality or flatness 0.15

IMPRoper  HAA  NAK  NAL  CAM ! flat ? (0 degrees = CIS)      0.00
IMPRoper  NAK  NAL  CAM  NAS ! flat ? (180 degrees = trans) 180.00


!IMPRoper  NAL  CAM  NAD  CAE ! flat ? (180 degrees = trans) 180.00
!IMPRoper  NAL  CAM  NAS  CAT ! flat ? (180 degrees = trans) 180.00


IMPRoper  CAM  NAD  CAE  CAT ! flat ? (0 degrees = CIS)      0.00
IMPRoper  NAD  CAE  CAT  NAS ! flat ? (0 degrees = CIS)      0.00
IMPRoper  CAE  CAT  NAS  CAM ! flat ? (0 degrees = CIS)      0.00
IMPRoper  CAT  NAS  CAM  NAD ! flat ? (0 degrees = CIS)      0.00
IMPRoper  NAS  CAM  NAD  CAE ! flat ? (0 degrees = CIS)      0.00




IMPRoper  CAU   NAS  CAE  CAY ! chirality or flatness   69.0091
IMPRoper  CAV   NAX  CAF  CAZ ! chirality or flatness  -69.2039
IMPRoper  1HAY  2HAY  CAT  CAZ ! chirality or flatness  -67.9409
IMPRoper  1HAZ  2HAZ  CAW  CAY ! chirality or flatness   67.8271
IMPRoper  2HAU  1HAU  CAT  CAV ! chirality or flatness   67.9811
IMPRoper  2HAV  1HAV  CAW  CAU ! chirality or flatness  -67.7693
```

```
   IMPRoper   NAO   NAP   HAE   CAN ! chirality or flatness   0.00
   IMPRoper   CAN   NAO   NAX   NAG ! chirality or flatness  -0.32
   IMPRoper   NAX   CAW   HAH   CAN ! chirality or flatness  -0.02
   IMPRoper   CAF   CAW   OAB   NAG ! chirality or flatness   0.42

   IMPRoper   HAF   NAP   NAO   CAN ! flat ? (0 degrees = CIS)        0.00
   IMPRoper   NAP   NAO   CAN   NAX ! flat ? (180 degrees = trans) 180.00

  !IMPRoper   NAO   CAN   NAG   CAF ! flat ? (180 degrees = trans) 180.00
  !IMPRoper   NAO   CAN   NAX   CAW ! flat ? (180 degrees = trans) 180.00

   IMPRoper   CAN   NAG   CAF   CAW ! flat ? (0 degrees = CIS)        0.00
   IMPRoper   NAG   CAF   CAW   NAX ! flat ? (0 degrees = CIS)        0.00
   IMPRoper   CAF   CAW   NAX   CAN ! flat ? (0 degrees = CIS)        0.00
   IMPRoper   CAW   NAX   CAN   NAG ! flat ? (0 degrees = CIS)        0.00
   IMPRoper   NAX   CAN   NAG   CAF ! flat ? (0 degrees = CIS)        0.00


   { Note: edit any DONOrs and ACCEptors if necessary }
DONOr HAA   NAK
DONOr HAB   NAL
DONOr HAE   NAO
DONOr HAF   NAP
DONOr HAG   NAS
DONOr HAH   NAX
ACCEptor   OAA   CAE
ACCEptor   OAB   CAF

END { RESIdue SPI }


presidue LNK1      ! PEPTide bond link, for NDI_SPI
  add bond 1C 2NAK

  add angle 1CD 1C 2NAK
  add angle 1O 1C 2NAK
  add angle 1C  2NAK 2NAL
  add angle 1C  2NAK 2HAA

  add improper 1O 1C  2NAK  2NAL          ! planar 1C
  add improper 2HAA 2NAK  1C  1CD         ! planar 2N
  add improper 1CD 1C  2NAK  2NAL         ! planar peptide
end

!patch LNK1
! reference=1=( resid 1 ) reference=2=( resid 3 )
! end


presidue LNK2      !
  add bond 1C 2NAP

  add angle 1CD 1C 2NAP
```

```
   add angle 1O 1C 2NAP
   add angle 1C  2NAP 2NAO
   add angle 1C  2NAP 2HAF

   add improper 1O 1C  2NAP  2NAO          ! planar 1C
   add improper 2HAF 2NAP  1C   1CD        ! planar 2N
   add improper 1CD 1C  2NAP  2NAO          ! planar peptide
end

!patch LNK2
 !reference=1=( resid 2 ) reference=2=( resid 3 )
 !end

set echo=true end
```

Appendix B: Parameter file for dimer C1 used in CNS simulation

```
set echo off message off end
{ Note: edit if necessary }
BOND NX7   HX8    1000.0  1.000 ! Nobs =    1
BOND NX7   NX9    1000.0  1.246 ! Nobs =    1
BOND NX9   HX10   1000.0  1.000 ! Nobs =    1
BOND NX9   CX11   1000.0  1.346 ! Nobs =    1

BOND CX11 NX12    1000.0  1.335 ! Nobs =    1
BOND CX11 NX14    1000.0  1.326 ! Nobs =    1
BOND NX12 HX13    1000.0  0.999 ! Nobs =    1
BOND NX12 CX17    1000.0  1.470 ! Nobs =    1
BOND NX14 CX15    1000.0  1.327 ! Nobs =    1
BOND CX15 OX16    1000.0  1.228 ! Nobs =    1
BOND CX15 CX17    1000.0  1.518 ! Nobs =    1

BOND CX17 CX18    1000.0  1.552 ! Nobs =    1
BOND CX18 HA      1000.0  1.000 ! Nobs =    1
BOND CX18 CX21    1000.0  1.542 ! Nobs =    1

BOND CX17 CX24    1000.0  1.552 ! Nobs =    1
BOND CX24 HA      1000.0  1.000 ! Nobs =    1
BOND CX24 CX27    1000.0  1.542 ! Nobs =    1

BOND CX21 HA      1000.0  1.000 ! Nobs =    1
BOND CX21 CX30    1000.0  1.541 ! Nobs =    1

BOND CX27 HA      1000.0  1.000 ! Nobs =    1
BOND CX27 CX30    1000.0  1.541 ! Nobs =    1

BOND CX30 CX15    1000.0  1.518 ! Nobs =    1
BOND CX30 NX12    1000.0  1.470 ! Nobs =    1


{ Note: edit if necessary }
ANGLe HX8   NX7   NX9     500.0   118.44 ! Nobs =    1
ANGLe NX7   NX9   HX10    500.0   120.62 ! Nobs =    1
ANGLe NX7   NX9   CX11    500.0   120.94 ! Nobs =    1
ANGLe HX10  NX9   CX11    500.0   118.39 ! Nobs =    1
ANGLe NX9   CX11 NX12     500.0   126.29 ! Nobs =    1
ANGLe NX9   CX11 NX14     500.0   121.25 ! Nobs =    1
ANGLe NX12 CX11 NX14      500.0   112.38 ! Nobs =    1
ANGLe CX11 NX12 HX13      500.0   125.09 ! Nobs =    1
ANGLe CX11 NX12 CX17      500.0   109.83 ! Nobs =    1
ANGLe HX13 NX12 CX17      500.0   125.08 ! Nobs =    1
ANGLe CX11 NX14 CX15      500.0   108.57 ! Nobs =    1
ANGLe NX14 CX15 OX16      500.0   123.70 ! Nobs =    1
ANGLe NX14 CX15 CX17      500.0   110.89 ! Nobs =    1
ANGLe OX16 CX15 CX17      500.0   125.41 ! Nobs =    1

ANGLe NX12 CX17 CX15      500.0    98.14 ! Nobs =    1
```

```
ANGLe NX12 CX17 CX18      500.0   109.07 ! Nobs =    1
ANGLe NX12 CX17 CX24      500.0   109.07 ! Nobs =    1
ANGLe CX15 CX17 CX18      500.0   111.95 ! Nobs =    1
ANGLe CX15 CX17 CX24      500.0   111.95 ! Nobs =    1
ANGLe CX18 CX17 CX24      500.0   111.84 ! Nobs =    1

ANGLe CX17 CX18 HA        500.0   108.43 ! Nobs =    1
ANGLe CX17 CX18 CX21      500.0   113.60 ! Nobs =    1
ANGLe HA   CX18 HA        500.0   109.47 ! Nobs =    1
ANGLe HA   CX18 CX21      500.0   108.41 ! Nobs =    1

ANGLe CX17 CX24 HA        500.0   108.43 ! Nobs =    1
ANGLe CX17 CX24 CX27      500.0   113.60 ! Nobs =    1
ANGLe HA   CX24 HA        500.0   109.47 ! Nobs =    1
ANGLe HA   CX24 CX27      500.0   108.41 ! Nobs =    1

ANGLe CX18 CX21 HA        500.0   108.67 ! Nobs =    1
ANGLe CX18 CX21 CX30      500.0   112.59 ! Nobs =    1
ANGLe HA   CX21 HA        500.0   109.51 ! Nobs =    1
ANGLe HA   CX21 CX30      500.0   108.65 ! Nobs =    1

ANGLe CX24 CX27 HA        500.0   108.67 ! Nobs =    1
ANGLe CX24 CX27 CX30      500.0   112.59 ! Nobs =    1
ANGLe HA   CX27 HA        500.0   109.51 ! Nobs =    1
ANGLe HA   CX27 CX30      500.0   108.65 ! Nobs =    1

ANGLe CX21 CX30 CX27      500.0   112.08 ! Nobs =    1
ANGLe CX21 CX30 NX12      500.0   109.87 ! Nobs =    1
ANGLe CX21 CX30 CX15      500.0   114.07 ! Nobs =    1
ANGLe NX12 CX30 CX15      500.0    98.14 ! Nobs =    1
ANGLe NX12 CX30 CX27      500.0   109.87 ! Nobs =    1
ANGLe CX15 CX30 CX27      500.0   114.07 ! Nobs =    1

ANGLe CX30 NX12 HX13      500.0   125.08 ! Nobs =    1
ANGLe CX30 NX12 CX11      500.0   109.83 ! Nobs =    1
ANGLe CX30 CX15 OX16      500.0   125.41 ! Nobs =    1
ANGLe CX30 CX15 NX14      500.0   110.89 ! Nobs =    1

{ Note: edit if necessary }
!DIHEdral HX8  NX7  NX9  HX10   500.0   0   180.00 ! Nobs =    1
!DIHEdral HX8  NX7  NX9  CX11   500.0   0   0.00 ! Nobs =    1
!DIHEdral NX7  NX9  CX11  NX12  500.0   0   180.00 ! Nobs =    1
!DIHEdral NX7  NX9  CX11  NX14  500.0   0   0.00 ! Nobs =    1

!DIHEdral NX9  CX11  NX12  HX13  500.0   0   0.00 ! Nobs =    1
!DIHEdral NX9  CX11  NX12  CX17  500.0   0   180.00 ! Nobs =  1
!DIHEdral NX9  CX11  NX14  CX15  500.0   0   180.00 ! Nobs =  1
!DIHEdral HX13  NX12  CX17  CX15  500.0   0   180.00 ! Nobs =  1
!DIHEdral CX11  NX14  CX15  OX16  500.0   0   180.00 ! Nobs =    1
!DIHEdral OX16  CX15  CX17  NX12  500.0   0   180.00 ! Nobs =    1

!DIHEdral CX15  CX30  NX12  HX13  500.0   0   180.00 ! Nobs =    1
!DIHEdral NX12  CX30  CX15  OX16  500.0   0   180.00 ! Nobs =    1
```

```
!DIHEdral CX30  NX12  CX11  NX9   500.0   0   180.00 ! Nobs =   1


{ Note: edit these IMPRopers if necessary }
IMPRoper  NX9   NX7   HX10  CX11   500.0   0    0.00! flatness
IMPRoper  CX11  NX9   NX12  NX14   500.0   0    0.00! flatness
IMPRoper  NX12  CX11  HX13  CX17   500.0   0    0.00! flatness
IMPRoper  CX15  NX14  OX16  CX17   500.0   0    0.00! flatness

IMPRoper  HX8   NX7   NX9   CX11   500.0   0    0.00! flatness
IMPRoper  NX7   NX9   CX11  NX12   500.0   0   180.00! flatness

!IMPRoper  NX9   CX11  NX14  CX15   500.0   0   180.00! flatness
!IMPRoper  NX9   CX11  NX12  CX17   500.0   0   180.00! flatness

IMPRoper  CX11  NX14  CX15  CX17   500.0   0    0.00! flatness
IMPRoper  NX14  CX15  CX17  NX12   500.0   0    0.00! flatness
IMPRoper  CX15  CX17  NX12  CX11   500.0   0    0.00! flatness
IMPRoper  CX17  NX12  CX11  NX14   500.0   0    0.00! flatness
IMPRoper  NX12  CX11  NX14  CX15   500.0   0    0.00! flatness


IMPRoper  CX24  NX12  CX15  CX18   500.0   0    69.01! flatness
IMPRoper  CX27  NX12  CX15  CX21   500.0   0   -69.20! flatness
IMPRoper  HA    HA    CX17  CX21   500.0   0   -67.94! flatness
IMPRoper  HA    HA    CX30  CX18   500.0   0    67.83! flatness
IMPRoper  HA    HA    CX17  CX27   500.0   0    67.98! flatness
IMPRoper  HA    HA    CX30  CX24   500.0   0   -67.77! flatness


IMPRoper  NX12  CX30  HX13  CX11   500.0   0    0.00! flatness
IMPRoper  CX15  CX30  OX16  NX14   500.0   0    0.00! flatness


!IMPRoper  NX9   CX11  NX12  CX30   500.0   0   180.00! flatness

IMPRoper  CX11  NX14  CX15  CX30   500.0   0    0.00! flatness
IMPRoper  NX14  CX15  CX30  NX12   500.0   0    0.00! flatness
IMPRoper  CX15  CX30  NX12  CX11   500.0   0    0.00! flatness
IMPRoper  CX30  NX12  CX11  NX14   500.0   0    0.00! flatness

{ Note: edit if necessary }
NONBonded NX7  0.2384  2.8509    0.2384  2.8509
NONBonded HX8  0.0498  1.4254    0.0498  1.4254
NONBonded NX9  0.2384  2.8509    0.2384  2.8509
NONBonded HX10 0.0498  1.4254    0.0498  1.4254
NONBonded CX11 0.1200  3.7418    0.1000  3.3854
NONBonded NX12 0.2384  2.8509    0.2384  2.8509
NONBonded HX13 0.0498  1.4254    0.0498  1.4254
NONBonded NX14 0.2384  2.8509    0.2384  2.8509
NONBonded CX15 0.1200  3.7418    0.1000  3.3854
NONBonded OX16 0.1591  2.8509    0.1591  2.8509
NONBonded CX17 0.1200  3.7418    0.1000  3.3854
```

170

```
NONBonded CX18 0.1200  3.7418     0.1000  3.3854
NONBonded CX21 0.1200  3.7418     0.1000  3.3854
NONBonded CX24 0.1200  3.7418     0.1000  3.3854
NONBonded CX27 0.1200  3.7418     0.1000  3.3854
NONBonded CX30 0.1200  3.7418     0.1000  3.3854


!DIIMIDE BONDS
 BOND     NH3  HC              1000.0         1.04! from protein file
 BOND     NH3  CT              1000.0         1.489! from protein file

 BOND     CT   NX1             1000.0         1.47!
 BOND     NX1  CX2             1000.0         1.40!
 BOND     CX2  OX3             1000.0         1.23!
 BOND     CX2  CB              1000.0         1.40!
 BOND     CB   CB              1000.0         1.40

 BOND     CT   CT              1000.0         1.565!from sybyl
 BOND     CT   HA              1000.0         1.106!from sybyl
 BOND     CB   CA              1000.0         1.404!from sybyl
 BOND     CA   CA              1000.0         1.397!from sybyl
 BOND     CA   HA              1000.0         1.092!from sybyl
 BOND     C    CT              1000.0         1.502!from sybyl
 BOND     C    O               1000.0         1.231! from protein file

! DIIMIDE ANGLES
 ANGLEs    CB CB CB            500.0      120.0
 ANGLES    OX3 CX2 NX1         500.0      120.0
 ANGLES    CT NX1 CX2          500.0      120.0
 ANGLES    OX3 CX2 CB          500.0      120.0
 ANGLES    CX2 CB CB           500.0      120.0
 ANGLES    NX1 CT CT           500.0      109.5
 ANGLES    NX1  CT   HA        500.0      109.5
 ANGLES    CX2  NX1  CX2       500.0      120.0
 ANGLES    NX1  CX2  CB        500.0      120.0
 ANGLES    CX2  CB   CA        500.0      120.0
 ANGLES    CA   CB   CB        500.0      120.0
 ANGLe     HA CT   HA          500.0      109.5  ! Nobs =     1
 ANGLe     HA CT   CT          500.0      109.80 ! Nobs =     1
 ANGLe     CB  CA   HA         500.0      120.67 ! Nobs =     1
 ANGLe     CA  CA HA           500.0      118.96 ! Nobs =     1
 ANGLe     CT CT C             500.0      110.65 ! Nobs =     1
 ANGLe     HA CT C             500.0      110.74 ! Nobs =     1
 ANGLe     CT C O              500.0      122.56 ! Nobs =     1
 ANGLe     CB CA CA            500.0      120.71 ! Nobs =     1
 ANGLe     CT   CT NH3         500.0      111.8939 ! Nobs =     1
 ANGLe     HA CT NH3           500.0      108.9390  ! Nobs =    1
 ANGLe     CT NH3 HC           500.0      119.4693  ! Nobs =    1
 ANGLe     HC NH3 HC           500.0 {SD = 0.031} 108.1992!from Lys

! DIIMIDE IMPROPERS
 IMPRoper  HA    HA    HA    HA    750.0    0      0.0
 IMPRoper  OX3   OX3   OX3   OX3   750.0    0      0.0
 IMPRoper  NX1   CB    CA    CA    750.0    0    180.0


                              171
```

```
 IMPRoper  CX2   CX2   CX2   CX2      750.0    0      0.0
 IMPROper  CB    CB    CB    CB       750.0    0      0.0

 IMPRoper  HC    HC   CT    HC        500.00   0    -66.4262 ! from protein
file


! DIIMIDE DIHEDRALS
 DIHEDRAL  CB    CB    CB    CB       750.0    0    180.0 !THESE are 4
 DIHEDRAL  OX3   CX2   NX1   CT       750.0    0      0.0 !planarity:
 DIHEDRAL  OX3   CX2   CB    CB       750.0    0    180.0 !USING 750.0
 DIHEDRAL  HA    CA    CA    HA       750.0    0      0.0 !(500)
 DIHEDRAL  CX2   CB    CB    CB       750.0    0    180.0 !not $kdih (0)
 DIHEDRAL  CB    CA    CA    CB       750.0    0      0.0 !
 DIHEDRAL  CB    CB    CB    CA       750.0    0      0.0 !
 DIHEDRAL  HA    CA    CB    CB       750.0    0    180.0 !
 DIHEDRAL  CX2   NX1   CT    CT         0.0    6      0.0 !0826
 DIHEDRAL  NX1   CT    CT    NH3        0.0    3      0.0 !0826
 DIHEDRAL  NX1   CT    CT    C          0.0    3      0.0 !0826
 DIHEDRAL  CT    CT    C     O          0.0    6      0.0 !0826

NONBonded CT   0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
NONBonded NX1  0.2384  2.8509    0.2384  2.8509 ! assuming Nitrogen
NONBonded CX2  0.1200  3.7418    0.1000  3.3854 ! assuming Carbon
NONBonded OX3  0.1591  2.8509    0.1591  2.8509 ! assuming Oxygen
NONBonded CA   0.1200  3.7418    0.1000  3.3854 ! assuming Carbon


!from protein parameter file

NONBonded  CB      0.1450  3.3409      0.1450  3.3409
NONBonded  C       0.0903  3.3409      0.0903  3.3409
NONBonded  HA      0.0045  2.2272      0.0045  2.2272
NONBonded  HC      0.0498  2.2272      0.0498  2.2272
NONBonded  NH3     0.1592  3.0068      0.1592  3.0068
NONBonded  O       0.2342  2.7755      0.2342  2.7755


!parameters for amide bond linkage (from protein parameter file)

BOND  C    NX7     1000.000 {sd=     0.001}     1.329

ANGLe  CT   C     NX7     500.00 {sd=  0.031}    116.1998
ANGLe  O    C     NX7     500.00 {sd=  0.031}    122.9907
ANGLe  C    NX7   NX9     500.00 {sd=  0.031}    121.6541
ANGLe  C    NX7   HX8     500.00 {sd=  0.031}    119.2489


IMPRoper  O    C    NX7    NX9    500.00 {sd=  0.031}    0    -0.0057
IMPRoper  HX8   NX7   C    CT     500.00 {sd=  0.031}    0     0.0023
IMPRoper  CT   C    NX7    NX9    500.00 {sd=  0.031}    0    180.0


set echo on message on end
```

Bibliography

Adams, A.; Guss, J. M.; Collyer, C. A.; Denny, W. A.; Wakelin, L. P. Crystal structure of the topoisomerase II poison 9-amino-[N-(2-dimethylamino)ethyl]acridine-4-carboxamide bound to the DNA hexanucleotide d(CGTACG)$_2$. *Biochemistry* **1999**, 38, 9221-9233.

Albericio, F.; Barany, G.; Fields, G. B.; Hudson, D.; Kates, S. A.; Lyttle, M. H.; Sole, N. A. Allyl-based orthogonal solid phase peptide synthesis. In "*Peptides 1992, Proc. 22nd European Peptide Symposium*"; Schneider, C. H. and Eberle, A. N. (eds.) ESCOM, Leiden, 1993, pp. 191-3.

Alsina, J.; Rabanal, F.; Chiva, C.; Giralt, E.; Albericio, F. Active carbonate resins: application to the solid-phase synthesis of alcohol, carbamate and cyclic peptides. *Tetrahedron* **1998**, 54, 10125-10152.

Auffinger, P.; Westhof, E. Simulations of the molecular dynamics of nucleic acids. *Curr. Opin. Struct. Biol.* **1998**, 8, 227-236.

Bailly, C.; Braña, M; Waring, M. J. Sequence-selective intercalation of antitumour bisnaphthalimides into DNA. Evidence for an approach via the major groove. *Eur. J. Biochemistry* **1996**, 240, 195-208.

Banavali, N. K.; MacKerell, A. D. Jr. Free energy and structural pathways of base flipping in a DNA GCGC containing sequence. *J. Mol. Biol.* **2002**, 319, 141-160.

Bernet, J.; Zakrzewska, K.; Lavery, R. Modelling base pair opening: the role of helical twist. *J. Mol. Struct. (Theochem)* **1997**, 398-399, 473-482.

Beveridge, D. L.; McConnell, K. J., Nucleic acids: theory and computer simulation, Y2K. *Curr. Opinion Struct. Biol.* **2000**, 10, 182-196.

Bijapur, J.; Keppler, M. D.; Bergqvist, S.; Brown, T.; Fox, K. R. 5-(1-Propargylamino)-2'-deoxyuridine (U^p): a novel thymidine analogue for generating DNA triplexes with increased stability. *Nucleic Acids Res.* **1999**, 27, 1802-1809.

Boisbouvier, J.; Delaglio, F.; Bax, A. Direct observation of dipolar couplings between distant protons in weakly aligned nucleic acids. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, 100, 11333-11338.

Bousquet, P. F.; Braña, M. F.; Conlon, D.; Fitzgerald, K. M.; Perron, D.; Cocchiaro, C.; Miller, R.; Moran, M.; George, J.; Qian, X.-D.; Keilhauer, G.; Romerdahl, C. A. Preclinical evaluation of LU 79553: a novel bis-naphthalimide with potent antitumor activity. *Cancer Res.* **1995**, 55, 1176-1180.

Bouvier, B.; Grubmüller, H. A Molecular dynamics study of slow base flipping in DNA using conformational flooding. *Biophys. J.* **2007**, 93, 770-786.

Bremer, R. E.; Wurtz, N. R.; Szewczyk, J. W.; Dervan, P. B.. Inhibition of major groove DNA binding bZIP proteins by positive patch polyamides. *Bioorg. & Med. Chem.* **2001**, 9, 2093-2103.

Bridewell, D. J.; Finlay, G. J.; Baguley, B. C. Topoisomerase I/II selectivity among derivatives of N-[2-(dimethylamino)ethyl]acridine-4-carboxamide (DACA). *Anti-Cancer Drug Des.* **2001**, *16*, 317-324.

Briki, F.; Ramstein, J.; Lavery, R.; Genest, D. Evidence for the stochastic nature of base pair opening in DNA: a Brownian dynamics simulation. *J. Am. Chem. Soc.* **1991**, 113, 2490-2493.

Brünger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J.-S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr.* **1998**, D54, 905-921.

Burden, D. A.; Osheroff, N. Mechanism of action of eukaryotic topoisomerase II and drugs targeted to the enzyme. *Biochim. Biophys. Acta* **1998**, 1400, 139-154.

Cantor, C.; Smith, C. L. Genomics. The science and technology behind the human genome project. Wiley and Sons, New York, NY, 1999.

Cao, C; Jiang, Y.; Stivers, J.; Song, F.. Dynamic opening of DNA during the enzymatic search for a damaged site base. *Nat. Struct. Mol. Biol.* **2004**, 11, 1230-1236.

Chaires, J. B.; Leng, F.; Przewloka, T.; Fokt, I.; Ling, Y. H.; Perez-Soler, R.; Priebe, W. Structure-based design of a new bisintercalating anthracycline antibiotic. *J. Med. Chem.* **1997**, 40, 261-266.

Chaires, J. B. Energetics of drug-DNA interactions. *Biopolymers* **1998(a)**, 44, 201-215.

Chaires, J. B. Drug-DNA interactions. *Curr. Opin. Struct. Biol.* **1998 (b)**, 8, 314-320.

Chapman, N. B.; Sotheeswaran, S.; Toyne, K. J. Preparation of 4-substituted 1-methoxycarbonylbicyclo[2.2.2]octanes, 4-substituted 1-phenylbicyclo[2.2.2]octanes, 4-substituted 1-p-nitrophenylbicyclo[2.2.2]octanes and 1,4-disubstituted bicyclo[2.2.2]octanes. *J. Org. Chem.* **1970**, 35, 917-923.

Cheatham, T. E., III; Miller, J. L.; Fox, T.; Darden, T. A.; Kollman, P. A. Molecular dynamics simulation on solvated biomolecular systems: the particle mesh Ewald method

leads to stable trajectories of DNA, RNA and proteins. *J. Am. Chem. Soc.* **1995**, 117, 4193-4194.

Cheatham, T. E., III; Cieplak, P.; Kollman, P. A. A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dyn.* **1999**, 16, 845-862.

Chen, F. M. *Methods for the Studies of Drug Dissociation from DNA*; Humana Press: Totowa, NJ, 1997.

Chen, Y. Z.; Mohan, V.; Griffey, R. H. The opening of a single base without perturbations of neighboring nucleotides: a study on crystal B-DNA duplex d(CGCGAATTCGCG)$_2$. *J. Biomol. Struct. Dyn.* **1998**, 15, 765–777.

Cholody, W. M.; Hernandez, L.; Hassner, L.; Scudiero, D. A.; Djurickovic, D. B.; Michejda, C. J. Bisimidazoacridones and related compounds: new antineoplastic agents with high selectivity against colon tumors. *J. Med. Chem.* **1995**, 38, 3043-3052.

Choo, Y.; Isalan, M. Advances in zinc finger engineering. *Curr. Opin. Struct. Biol.* **2000**, 10, 411-416.

Chu, Y.; Lynch, V.; Iverson, B. L. Synthesis and DNA binding studies of bisintercalators with a novel spiro-cyclic linker. *Tetrahedron* **2006**, 62, 5536-5548.

Chu, Y.; Sorey, S.; Hoffman, D. W.; Iverson, B. L. Structural characterization of a rigidified threading bisintercalator. *J. Am. Chem. Soc.* **2007**, 129, 1304-1311.

Crooks, S. L.; Robinson, M. B.; Koerner, J. F.; Johnson, R. L. Cyclic analogs of 2-amino-4-phosphonobutanoic acid (APB) and their inhibition of hippocampal excitatory transmission and displacement of [3H]APB binding. *J. Med. Chem.* **1986**, 29, 1988-1995.

Crothers, D. M. Calculation of binding isotherms for heteroheneous polymers. *Biopolymers* **1968**, 6, 575-584.

Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **1995**, 6, 277-293.

DeLano, W. L. The PyMol Molecular Graphics System; DeLano Scientific, San Carlos, CA, USA, 2002.

Denny, W. A.; Wakelin, L. P. G. Kinetic and equilibrium studies of the interaction of amsacrine and anilino ring-substituted analogues with DNA. *Cancer Res.* **1986**, 46, 1717-1721.

Dervan, P. B. Design of sequence-specific DNA-binding molecules. *Science* **1986**, 232, 464-471.

Dervan, P. B. Molecular recognition of DNA by small molecules. *Bioorg. & Med. Chem*. **2001**, 9, 2215-2235.

Dickinson, L. A.; Gulizia, R. J.; Trauger, J. W.; Baird, E. E.; Mosier, D.E.; Gottesfeld, J. M.; Dervan, P. B. Inhibition of RNA polymerase II transcription in human cells by synthetic DNA-binding ligands. *Proc. Natl. Acad. Sci. USA* **1998**, 95, 12890-12895.

Dornberger, U.; Leijon, M.; Fritzsche, H. High base pair opening rates in tracts of GC base pairs. *J. Biol. Chem.* **1999**, 274, 6957–6962.

Egli, M.; Williams, L. D.; Frederick, C. A.; Rich, A. DNA-nogalamycin interactions. *Biochemistry* **1991**, 30, 1364-1372.

Elbashir, S. M.; Harborth, J.; Lendeckel, W.; Yalcin, A.; Weber, K.; Tuschl, T. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **2001**, 411, 494-498.

Elrod-Erickson, M.; Pabo, C. O. Binding studies with mutants of Zif268: contribution of individual side chains to binding affinity and specificity in the Zif268 zinc finger-DNA complex. *J. Biol. Chem.* **1999**, 274, 19281-5.

Englander, S. W.; Kallenbach, N. R. Hydrogen exchange and structural dynamics of proteins and nucleic acids. *Q. Rev. Biophys.* **1984**, 16, 521–655.

Faria, M.; Wood, C. D.; Perrouault, L.; Nelson, J. S.; Winter, A.; White, M. R. H.; Hlélèn, C.; Giovannangeli, C. Targeted inhibition of transcription elongation in cells mediated by triplex forming oligonucleotides. *Proc. Natl. Acad. Sci. USA* **2000**, 97, 3862-3867.

Farkas, G., Leibovitch, B. A. and Elgin, S. C. Chromatin organization and transcriptional control of gene expression in Drosophila. *Gene* **2000**, 253, 117-136.

Feigon, J.; Denny, W. A.; Leupin, W.; Kearns, D. R. Interactions of antitumor drugs with natural DNA: ^{1}H NMR study of binding mode and kinetics. *J. Med. Chem.* **1984**, 27, 450-465.

Finkelstein, A. V.; Janin, J. The price of lost freedom: entropy of bimolecular complex formation. *Prot. Eng.* **1989**, 3, 1-3.

Fire, A.; Xu, S.; Montgomery, M. K.; Kostas, S. A.; Driver, S. A.; Mello, C. C. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **1998**, 391, 806-811.

Fitzgerald, D. J.; Anderson, J. N. Selective nucleosome disruption by drugs that bind in the minor groove of DNA. *J. Biol. Chem.* **1999**, 274, 27128-27138 and references therein.

Foloppe, N. ; MacKerell, A. D., Jr. All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comput. Chem.* **2000**, 21, 86–104.

Folta-Stogniew, E.; Russu, I. M. Sequence dependence of base-pair opening in a DNA dodecamer containing the CACA/GTGT sequence motif. *Biochemistry* **1994**, 33, 11016-11024.

Fox, K. R.; Brassett, C.; Waring, M. J. Kinetics of dissociation of nogalamycin from DNA: comparison with other anthracycline antibiotics. *Biochim. Biophys. Acta*, General Subjects, **1985**, 840, 383-392.

Fox, K. R. Targeting DNA with triplexes. *Curr. Med. Chem.* **2000**, 7, 17-37.

Foye, W. O. (ed.) Cancer Chemotherapeutic Agents. The American Chemical Society, Washington, D.C., 1995.

Francis, N. J.; Kingston, R. E.; Woodcock, C. L. Chromatin compaction by a polycomb group protein complex. *Science* **2004**, 306, 1574-1577.

Franzus, B.; Surridge, J. H. The reaction of ethyl azodicarboxylate with 1,3- and 1,4-cyclohexadienes. *J. Org. Chem.* **1962**, 27, 1951-1957.

Gadwood, R. C.; Kamdar, B. V.; Cipkus Dubray, L. A.; Wolfe, M. L.; Smith, M. P.; Watt, W.; Mizsak, S. A.; Groppi, V. E. Synthesis and biological activity of spirocyclic benzopyran imidazolone potassium channel openers. *J. Med. Chem.* **1993**, 36, 1480-1487.

Gago, F. Stacking interactions and intercalative binding. *Methods* **1998**, 14, 277-292.

Gallego, J.; Reid, B. R. Solution structure of a complex between DNA and the antitumour bisnaphthalimide LU-79553: Intercalated ring flipping on the millisecond time scale. *Biochemistry* **1999**, 38, 15104-15115.

Giudice, E.; Lavery, R. Simulations of nucleic acids and their complexes. *Acc. Chem. Res.* **2002**, 35, 350-357.

Giudice, E.; VaÂrnai, P.; Lavery, R. Base pair opening within B-DNA: free energy pathways for GC and AT pairs from umbrella sampling simulations. *Nucleic Acids Res.* **2003**, 31, 1434-1443.

Giudice, E.; Lavery, R. Nucleic acid base pair dynamics: the impact of sequence and structure using free-energy calculations. *J. Am. Chem. Soc*. **2003**, 125, 4998-4999.

Goddard, T. D.; Kneller, D. G. SPARKY 3; University of California, San Francisco, 2004.

Gottesfeld, J. M.; Neely, L.; Trauger, J. W.; Baird, E. E.; Dervan, P. B. Regulation of gene expression by small molecules. *Nature* **1997**, 387, 202-205.

Gottesfeld, J. M.; Melander, C.; Suto, R. K.; Raviol, H.; Luger, K.; Dervan, P. B. Sequence-specific recognition of DNA in the nucleosome by pyrrole-Imidazole polyamides. *J. Mol. Biol.* **2001**, 309, 615-629.

Guelev, V. M.; Harting, M. T.; Lokey, R. S.; Iverson, B. L. Altered sequence specificity identified from a library of DNA-binding small molecules. *Chemistry & Biology* **2000**, 7, 1-8.

Guelev, V.; Lee, J.; Ward, J.; Sorey, S.; Hoffman, D. W.; Iverson, B. L. Peptide bisintercalator binds DNA via threading mode with sequence specific contacts in the major groove. *Chemistry & Biology* **2001(a)**, 8, 415-425.

Guelev, V. M.; Cubberley, M. S.; Murr, M. M.; Lokey, R. S.; Iverson, B. L. Design, synthesis, and characterization of polyintercalating ligands. *Methods Enzymol.* **2001(b)**, 340, 556-570.

Guelev, V.; Sorey, S.; Hoffman, D. W.; Iverson, B. L. Changing DNA grooves - a 1,4,5,8-naphthalene tetracarboxylic diimide bis-intercalator with the Linker (β-Ala)$_3$Lys in the minor groove. *J. Am. Chem. Soc.* **2002**, 124, 2864-2865.

Guéron, M.; Leroy, J. L. Base-pair opening in double-stranded nucleic acids. In *Nucleic Acids and Molecular Biology*. Eckstein, F. and Lilley, D. M. J. (eds), Springer-Verlag, New York, 1992, Vol. 6, pp. 1-22.

Guéron, M.; Leroy, J. L. Studies of base pair kinetics by NMR measurement of porton exchange. *Methods Enzymol.* **1995**, 261, 383-413.

Hansen, M. R.; Mueller, L.; Pardi, A. Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat. Struct. Boil.* **1998**, 5, 1065-1074.

Haq, I.; Ladbury, J. Drug–DNA recognition: energetics and implications for design. *J. Mol. Recognit.* **2000**, 13, 188-197

Hare, D. R.; Wemmer, D. E.; Chou, S.-H.; Drobny, G.; Reid, B. R. Assignment of the non-exchangeable proton resonances of d(C-G-C-G-A-A-T-T-C-G-C-G) using two-dimensional nuclear magnetic resonance methods. *J. Mol. Biol.* **1983**, 171, 319-336.

Haroutounian, S.; Georgiadia, M. P.; Polissiou, M. G. Stereochemistry of the Bucherer-Bergs reaction and a modified Strecker reaction on tetrahydro-2H-pyran-3-ones. *J. Heterocycl. Chem.* **1989**, 26, 1283-1287.

Hu, G. C.; Shui, X.; Leng, F.; Priebe, W.; Chaires, J. B.; Williams, L. D. Structure of a DNA-bisdaunomycin complex. *Biochemistry* **1997**, 36, 5940-5946.

Ivanov, V.; Grzeskowiak, K.; Zocchi, G. Evidence for an intermediate state in the B-to-Z transition of DNA. *J. Phys. Chem. B*. **2003**, 107, 12847-12850.

Jamieson, A. C.; Miller, J. C.; Pabo, C. O. Drug discovery with engineered zinc-finger proteins. *Nature Rev. Drug Discov.* **2003**, 2, 361−368.

Jaroniec, C. P.; Boisbouvier, J.; Tworowska, I.; Nikonowicz, E. P.; Bax, A. Accurate measurement of 15N-13C residual dipolar couplings in nucleic acids. *J. Biomol. NMR* **2005**, 31, 231-241.

Jencks, W. P. On the attribution and additivity of binding energies. *Proc. Natl. Acad. Sci. USA* **1981**, 78, 4046-4050.

Jenner, G.; Salem, R. B. Anatomy of ene and Diels-Alder reactions between cyclohexadienes and azodicarboxylates. *J. Chem. Soc., Perkin Trans. 2*, **1990**, 1961-1964.

Jourdan, M.; Garcia, J.; Lhomme, J.; Teulade-Fichou, M.-P.; Vigneron, J.-P.; Lehn, J.-M. Threading bis-intercalation of a macrocyclic bisacridine at abasic sites in DNA: nuclear magnetic resonance and molecular modeling study. *Biochemistry* **1999**, 38, 14205-14213.

Kates, S. A.; Daniels, S. B.; Sole, N. A.; Barany, G.; Albericio, F. Automated allyl chemistry for solid-phase peptide synthesis: applications to cyclic and branched peptides. In "*Peptids, Chemistry, Structure & Biology, Proc. 13th American Peptide Symposium*", Hodges, R. S. and Smith, J. A. (Eds.), ESCOM, Leiden, 1994, pp. 113-115.

Keepers, J.; Kollman, P. A.; James, T. L. Molecular mechanical studies of base-pair opening in d(CGCGC):d(GCGCGC), dG5·dC5, d(TATAT):d(ATATA) and dA5·dT5 in the B and Z forms of DNA. *Biopolymers* **1984**, 23, 2499–2511.

Kessler, H.; Siegmeier, R. 9-Fluorenylmethyl esters as carboxyl protecting group. *Tetrahedron Lett.* **1983**, 24, 281-282.

Kleywegt, G. J.; Jones, T. A. Model building and refinement practice. *Methods Enzymol.* **1997**, 277, 208-230.

Klimasauskas, S.; Kumar, S.; Roberts, R. J.; Cheng, X. HhaI methyltransferase flips its target base out of the DNA helix. *Cell* **1994**, 76, 357–369.

Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. Binding of an antitumor drug to DNA: netropsin and C-G-C-G-A-A-T-T-^{Br}C-G-C-G. *J. Mol. Biol.* **1986**, 183, 553-563.

Lamberson, C. Topologically constrained DNA intercalators: synthesis and DNA binding studies. University of Illinois, Ph.D. thesis, 1991, pp. 23-26.

Lavie, L.; Maldener, E.; Brouha, B.; Meese, E. U.; Mayer, J. The human L1 promoter: variable transcription initiation sites and a major impact of upstream flanking sequence on promoter activity. *Genome Res.* **2004**, 14, 2253-2260

Lee, J.; Guelev, V.; Sorey, S.; Hoffman, D. W.; Iverson, B. L. NMR structural analysis of a modular threading tetraintercalator bound to DNA. *J. Am. Chem. Soc.* **2004(a)**, 126, 14036-14042.

Lee, J. Toward threading polyintercalators with programmed sequence specificity. University of Texas at Austin, Ph.D. thesis, **2004(b)**.

Leng, F.; Priebe, W.; Chaires, J. B. Ultratight DNA binding of a new bisintercalating anthracycline antibiotic. *Biochemistry* **1998**, 37, 1743-1753.

Le Pecq, J. B.; Le Bret, M.; Barbet, J.; Roques, B. DNA polyintercalating drugs: DNA binding of diacridine derivatives. *Proc. Natl Acad. Sci. USA* **1975**, 72, 2915-2919.

Leroy, J. L.; Charretier, E.; Kochoyan, M.; Guéron, M. Evidence from base-pair kinetics for two types of adenine tract structures in solution: their relation to DNA curvature. *Biochemistry* **1988**, 27, 8894-8898.

Levins, C. G.; Schafmeister, C. E. The synthesis of functionalized nanoscale molecular rods of defined length. *J. Am. Chem. Soc.* **2003**, 125, 4702-4703.

Lipsitz, R. S.; Tjandra, N. Residual dipolar couplings in NMR structure analysis. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, 33, 387-413.

Liu, Z. –R.; Hecker, K. H.; Rill, R. L. Selective DNA binding of (N-alkylamine)-substituted naphthalene imides and diimides to G+C-rich DNA. *J. Biomol. Struct. Dynam.* **1996**, 14, 331-339.

Lokey, R. S.; Iverson, B. L. Synthetic molecules that fold into a pleated secondary structure in solution. *Nature (London)* **1995**, 375, 303-305.

Lokey, R. S.; Kwok, Y.; Guelev, V.; Pursell, C.; Hurley, L.; Iverson, B. L. A new class of polyintercalating molecules. *J. Am. Chem. Soc.* **1997**, 19, 7202-7210.

Lyakhov, I. G.; Hengen, P. N.; Rubens, D.; Schneider, T. D. The P1 phase replication protein RepA contacts an otherwise inaccessible thymine N3 protein the DNA distortion or base flipping. *Nucl. Acids Res.* **2001**, 29, 4892-4900.

MacDonald, D.; Lu, P. Residual dipolar couplings in nucleic acid structure determination. *Curr. Opin. Struct. Biol.* **2002**, 12, 337-343.

Mammen, M.; Choi, S.-K.; Whitesides, G. M. Polyvalent interactions in biological systems: implications for design and use of multivalent ligands and inhibitors. *Angew. Chem. Int. Ed.* **1998**, 37, 2754-2794.

Manning, G. S. The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Q. Rev. Biophys.* **1978,** 2, 179-246 .

Mapp, A. K.; Ansari, A. Z.; Ptashne, M.; Dervan, P. B. Activation of gene expression by small molecule transcription factors. *Proc. Natl. Acad. Sci. USA* **2000**, 97, 3930-3935.

Mazur, A. K. Titration *in silico* of reversible B $\leftrightarrow$ A transitions in DNA. *J. Am. Chem. Soc.* **2003**, 125, 7849-7859.

Mazzitelli, C. L.; Chu, Y.; Reczek, J. J.; Iverson, B. L.; Brodbelt, J. S. Screening of threading bisIntercalators binding to duplex DNA by Electrospray ionization Tandem Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2007**, 18, 311-321.

Minchenko, O. H.; Ochiai, A.; Opentanova, T. O.; Minchenko, D. O.; Caro, J.; Komisarenko, S. V.; Esumi, H. Overexpression of 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase-4 in the human breast and colon malignant tumors. *Biochemie* **2005**, 87, 1005-1010.

Mohammadi, S.; Perree-Fauvet, M.; Gresh, N.; Hillairet, X. K.; Taillandier, E. Joint molecular modeling and spectroscopic studies of DNA complexes of a bis(arginyl) conjugate of a tricationic porphyrin designed to target the major groove. *Biochemistry* **1998**, 37, 6165-6178.

Moser, H.; Dervan, P. B. Sequence-specific cleavage of double helical DNA by triple helix formation. *Science* **1987**, 238, 645-650.

Murr, M.; Harting, M.; Guelev, V.; Ren, J.; Chaires, J. B.; Iverson, B. L. An oktakis-intercalating molecule. *Bioorg. Med. Chem.* **2001**, 9, 1141-1148

Neely, R. K.; Daujotyte, D.; Grazulis, S.; Magennis, S. W.; Dryden, D. T. F.; Klimasauskas, S.; Jones, A. C. Time-resolved fluorescence of 2-aminopurine as a probe of base flipping in M.HhaI-DNA complexes. *Nucleic Acids Res.* **2005**, 33, 6953–6960.

Neidle, S. *Nucleic acid structure and recognition*. Oxford University Press Inc., New York, 2002.

O'Neil, L. L.; Wiest, O. A selective, noncovalent assay for base flipping in DNA. *J. Am. Chem. Soc.* **2005**, 127, 16800-16801.

Patikoglou, G.; Burley, S. K. Eukaryotic transcription factor-DNA complexes. *Ann. Rev. Biophys. Biomol. Struct.* **1997**, 26, 289-325.

Pavletich, N. P.; Pabo, C. O. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* **1991**, 252, 809-17.

Pelton, J. G.; Wemmer, D. E. Structural characterization of a 2:1 distamycin A·d(CGCAAATTGGC) complex by two dimensional NMR. *Proc. Natl. Acad. Sci. USA* **1989**, 86, 5723-5727.

Pommier, Y.; Pourquier, P.; Fan, Y.; Strumberg, D. Mechanism of action of eukaryotic DNA topoisomerase I and drugs targeted to the enzyme. *Biochim. Biophys. Acta* **1998**, 1400, 83-106.

Portugal, J.; Waring, M. J. Antibiotics which can alter the rotational orientation of nucleosome core DNA. *Nucleic Acids Res.* **1986**, 14, 8735-8754.

Portugal, J.; Waring, M. J. Interaction of nucleosome core particles with distamycin and echinomycin: analysis of the effect of DNA sequences. *Nucleic Acids Res.* **1987**, 15, 885-903.

Praseuth, D.; Guieysse, A. L.; Helene, C. Triple helix formation and the antigene strategy for sequence-specific control of gene expression. *Biochim. Biophys. Acta* **1999**, 1489, 181-206.

Priyakumar, U. D.; MacKerell, A. D. Computational approaches for investigating base flipping in oligonucleotides. *Chem. Rev.* **2006(a)**, 106, 489-505.

Priyakumar, U. D.; MacKerell, A. D. NMR imino proton exchange experiments on duplex DNA primarily monitor the opening of purine bases. *J. Am. Chem. Soc.* **2006(b)**, 128, 678-679.

Ramstein, J.; Lavery, R. Base pair opening pathways in B-DNA. *J. Biomol. Struct. Dyn.* **1990**, 7, 915–933.

Ratner, L.; Haseltine, W.; Patarca, R.; Livak, K. J.; Starcich, B.; Josephs, S. F.; Doran, E. R.; Rafalski, J. A.; Whitehorn, E. A.; Baumeister, K. Complete nucleotide sequence of the AIDS virus, HTLV-III. *Nature* **1985**, 313, 277-284.

Record, M. T.; Anderson, C. F.; Lohman, T. Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: the role of ion association and release, screening and ion effects on water activity. *Q. Rev. Biophys.* **1978**, 11, 103–178.

Roberts, R. W.; Crothers, D. M. Specificity and stringency in DNA triplex formation. *Proc. Natl. Acad. Sci. USA* **1991**, 88, 9397-401.

Robinson, H.; Priebe, W.; Chaires, J. B.; Wang, A. H.-J. Binding of two novel bisdaunorubicins to DNA studied by NMR spectroscopy. *Biochemistry* **1997**, 36, 8663-8670.

Saenger, W. Principles of Nucleic Acid Structure. Springer-Verlag, New York, N.Y., 1984.

Sambrook, J.; Russell, D. W. *Molecular cloning : a laboratory manual*; 3rd ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y., 2001, 5.4-5.86.

Schneider, T. D. Strong minor groove base conservation in sequence logos implies DNA distortion or base flipping during replication and transcription initiation. *Nucl. Acids Res.* **2001**, 29, 4881-4891.

Schuettelkopf, A. W.; van Aalten, D. M. F. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr.* **2004**, D60, 1355-1363.

Shah, A.; George, M. V. Thermal and photochemical additions of azo esters to unsaturated systems—II: additions to olefins and dienes. *Tetrahedron* **1970**, 27, 1291-1301.

Sibille, N.; Pardi, A.; Simorre, J. P.; Blackledge, M. Refinement of local and long-range structural order in theophylline-binding RNA using 13C-1H residual dipolar couplings and restrained molecular dynamics. *J. Am. Chem. Soc.* **2001**, 123, 12135-12146.

Spielmann, H. P. Dynamics of a bis-intercalator DNA complex by 1H-detected natural abundance 13C NMR spectroscopy. *Biochemistry* **1998**, 37, 16863-76

Spolar, R.; Record, M. T. Coupling of local folding to site-specific binding of proteins to DNA. *Science* **1994**, 263, 777-784.

Takenaka, S.; Takagi, M. Threading intercalators as a new DNA structural probe. *Bull. Chem. Soc. Jpn.* **1999**, 72, 327-337.

Tanious, F.; Yen, S.; Wilson, W. D. Kinetic and equilibrium analysis of a threading intercalation mode: DNA sequence and ion effects. *Biochemistry* **1991**, 30, 1813-1819.

Thieriet, N.; Guibe, F.; Albericio, F. Solid-phase peptide synthesis in the reverse (N→C) direction. *Org. Lett.* **2000**, 2, 1815-1817.

Trauger, J. W.; Baird, E. E.; Dervan, P. B. Recognition of 16 base pairs in the minor groove of DNA by a pyrrole-imidazole polyamide dimer. *J. Am. Chem. Soc.* **1998**, 120, 3534-3535.

Vasquez, K. M.; Wilson, J. H. Triplex-directed modification of genes and gene activity. *Trends Biochem. Sci.* **1998**, 23, 4-9.

Veal, J. M.; Li, Y.; Zimmerman, S. C.; Lamberson, C. R.; Cory, M.; Zon, G.; Wilson, W. D. Interaction of a macrocyclic bisacridine with DNA. *Biochemistry* **1990**, 29, 10918-27.

Wakelin, L. Polyfunctional DNA intercalating agents. *Med. Res. Reviews* **1986**, 6, 275-340 and references therein.

Wain-Hobson, S.; Sonigo, P.; Danos, O.; Cole, S.; Alizon, M. Nucleotide sequence of the AIDS virus, LAV. *Cell* **1985**, 40, 9-17.

Wang, A. H. -J.; Ughetto, G.; Quigley, G. J.; Rich, A. Interactions between an anthracycline antibiotic and DNA: molecular structure of daunomycin complexed to d(CpGpTpApCpG) at 1.2-Å resolution. *Biochemistry* **1987**, 26, 1152-1163.

Waring, M. J.; Wakelin, L. P. G. Echinomycin: a bifunctional intercalating antibiotic. *Nature* **1974**, 252, 653-657.

White, S.; Baird, E. E.; Dervan , P. B. On the pairing rules for recognition in the minor groove of DNA by pyrrole-imidazole polyamides. *Chem. Biol.* **1997**, 4, 569-578.

White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B. Recognition of the four Watson-Crick base pairs in the DNA minor groove by synthetic ligands. *Nature* **1998**, 391, 468-471.

Wijmenga, S. S.; Mooren, M. W.; Hilbers, C. W. NMR of nucleic acids; from spectrum to structure. In *NMR of macromolecules. A practical approach.* (Roberts, G. C. K., ed.), IRL Press, New York, N. Y., 1993.

Williams, H. E. L.; Searle, M. S. Structure, dynamics and hydration of the nogalamycin-d(ATGCAT)2 complex determined by NMR and molecular dynamics simulations in solution, *J. Mol. Biol.* **1999**, 290, 699-716.

Wilson, W. D.; Krishnamoorthy, C. R.; Wang, Y. -H.; Smith, J. C. Mechanism of intercalation: ion effects on the equilibrium and kinetic constants for the interaction of propidium and ethidium with DNA. *Biopolymers* **1985**, 24, 1941-1961.

Wolfe, S. A.; Nekludova, L.; Pabo, C. O. DNA recognition by Cys2His2 Zinc Finger proteins. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, 29, 183–212.

Wolffe, A. P.; Guschin, D. Review: chromatin structural features and targets that regulate transcription. *J. Struct. Biol.* **2000**, 129, 102-122.

Wüthrich, K. NMR of proteins and nucleic acids. John Wiley and Sons, New York, N. Y. 1986.

Yen, S.-F.; Gabbay, E. J.; Wilson, W. D. Interaction of aromatic imides with deoxyribonucleic acid. Spectrophotometric and viscometric studies. *Biochemistry* **1982**, 21, 2070-2076.

Zimmerman, S. C.; Lamberson, C. R.; Cory, M.; Fairley, T. A. Topologically constrained bifunctional intercalators: DNA intercalation by a macrocyclic bisacridine. *J. Am. Chem. Soc.* **1989**, 111, 6805-6809.

Vita

Yongjun Chu was born and raised in Anhui, China. After graduation from Yuantan High School in Anhui, China in 1991, he was enrolled in the Department of Chemical Engineering of Harbin Engineering University in Harbin, China, following his love for chemistry and engineering. He received a Bachelor degree with honors in Chemical Engineering in 1995. Then he entered the graduate school of Dalian Institute of Chemical Physics, Chinese Academy of Sciences in Dalian, China, aiming to understand why the heterogeneous catalysts work so efficiently in industry. In 1998, he was awarded the Master of Science in Physical Chemistry (Catalysis) under the supervision of Professor Xin Qin. From 1998 to 1999, he worked as a research associate in Xin's group. During those four years of stay at Dalian Institute of Chemical Physics, he learned that heterogeneous catalysis was too complicated to be explained and predicted by any theories.

In the fall of 1999, he came to the United States, starting as a graduate student at the Chemistry Department of the University of Alabama, Tuscaloosa, AL. He was excited about the development of the Bioorganic Chemistry field. He then joined Dr. Woski's group, where he began working DNA chemistry. In December 2001, he received another M.S., but in the field of Organic Chemistry. He joined Dr. Iverson's group as a Ph.D. candidate at The University of Texas at Austin in Spring of 2002, continuing to work in the Bioorganic Chemistry field.

Permanent Address: State Key Laboratory of Catalysis, Dalian Institute of Chemical Physics, 457 Zhongshan Road, Dalian, China, 116023.

This dissertation was typed by the author.