J|A|C|S
ARTICLES
Published on Web 01/17/2007

Structural Characterization of a Rigidified Threading Bisintercalator

Yongjun Chu, Steven Sorey, David W. Hoffman, and Brent L. Iverson*

Contribution from the Department of Chemistry and Biochemistry, The University of Texas at Austin, Austin, Texas 78712

Received September 15, 2006; E-mail: biverson@mail.utexas.edu

Abstract: NMR spectroscopy was used 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. A structural model of **C1** complexed to $d(CGGTACCG)_2$ was calculated using distance constraints obtained from solution NMR data. The model was also supported by the results from residual dipolar coupling (RDC) measurements obtained using Pf1-phage as a cosolvent. According to the model, the central cyclohexane ring of the linker connecting the two NDI units lies 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 consisiting 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.

Introduction

Bisintercalation represents an effective strategy for targeting specific DNA sequences for purposes of cancer therapy and antibiotic development.[1] This approach was first investigated in the 1970s following the discovery and elucidation of the biological activities of echinomycin, a natural bisintercalator consisting of two quinoxaline moieties linked by a bicyclic peptide.[2] 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.[3] Some of the bisintercalators have exhibited promising antitumor activities, such as an acridine-based agent, DACA,[4] bisanthracycline derivative WP631,[5] bisimidazoacridone compound WMC26,[6] and bisnaththalimide-related compounds such as LU79553.[7]

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.[8] The 1,4,5,8-naphthalenetetracarboxylic 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.[9] Using a combinatorial approach, two sequence-specific threading bisintercalators, $\mathbf{G_3K}$ and $\mathbf{(\beta\text{-}Ala)_3K}$, linked by two related peptides were identified, with the first one binding to the *major* groove of $d(GGTACC)_2$[10] and the second one binding to the *minor* groove of $d(AGATGA)_2$.[11] Modeling has indicated that the 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 the exocyclic NH_2 group of G in the minor groove.[12] 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

(1) Chaires, J. B. *Curr. Opin. Struct. Biol.* **1998**, *8*, 314−320.
(2) (a) Waring, M. J.; Wakelin, L. P. G. *Nature (London)* **1974**, *252*, 653−657. (b) Wakelin, L. P. G.; Waring, M. J. *Biochem. J.* **1976**, *157*, 721−740. (c) Shafer, R. H.; Waring, M. J. *Biopolymers* **1980**, *19*, 431−443. (d) Van Dyke, M. M.; Dervan, P. B. *Science* **1984**, *225*, 1122−1127.
(3) Le Pecq, J. B.; Le Bret, M.; Barbet, J.; Roques, B. *Proc. Natl Acad. Sci. U.S.A.* **1975**, *72*, 2915−2919.
(4) Bridewell, D. J.; Finlay, G. J.; Baguley, B. C. *Anti-Cancer Drug Des.* **2001**, *16*, 317−324.
(5) (a) Chaires, J. B.; Leng, F.; Przewloka, T.; Fokt, I.; Ling, Y. H.; Perez-Soler, R.; Priebe, W. *J. Med. Chem.* **1997**, *40*, 261−266. (b) Leng, F.; Priebe, W.; Chaires, J. B. *Biochemistry* **1998**, *37*, 1743−1753. (c) Hu, G. G.; Shui, X.; Leng, F.; Priebe, W.; Chaires, J. B.; Williams, L. D. *Biochemistry* **1997**, *36*, 5940−5946. (d) Portugal, J.; Martin, B.; Vaquero, A.; Ferrer, N.; Villamarin, S.; Priebe, W. *Curr. Med. Chem.* **2001**, *8*, 1−8.
(6) Cholody, W. M.; Hernandez, L.; Hassner, L.; Scudiero, D. A.; Djurickovic, D. B.; Michejda, C. J. *J. Med. Chem.* **1995**, *38*, 3043−3052.
(7) 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. *Cancer Res.* **1995**, *55*, 1176−1180.
(8) Takenaka, S.; Takagi, M. *Bull. Chem. Soc. Jpn.* **1999**, *72*, 327−337.
(9) (a) Yen, S.-F.; Gabby, E. J.; Wilson, D. *Biochemistry* **1982**, *21*, 2070−2076. (b) Tanious, F. A.; Yen, S.-F.; Wilson, D. *Biochemistry* **1991**, *30*, 1813−1819. (c) Lokey, R. S.; Kwok, Y.; Guelev, V.; Pursell, C. J.; Hurley, L. H.; Iverson, B. L. *J. Am. Chem. Soc.* **1997**, *119*, 7202−7210.
(10) Guelev, V.; Lee, J.; Ward, J.; Sorey, S.; Hoffman, D. W.; Iverson, B. L. *Chem. Biol.* **2001**, *8*, 415−425.
(11) Guelev, V.; Sorey, S.; Hoffman, D. W.; Iverson, B. L. *J. Am. Chem. Soc.* **2002**, *124*, 2864−2865.
(12) Liu, Z.-R.; Hecker, K. H.; Rill, R. L. *J. Biomol. Struct. Dyn.* **1996**, *14*, 331−339.

10.1021/ja066480x CCC: $37.00

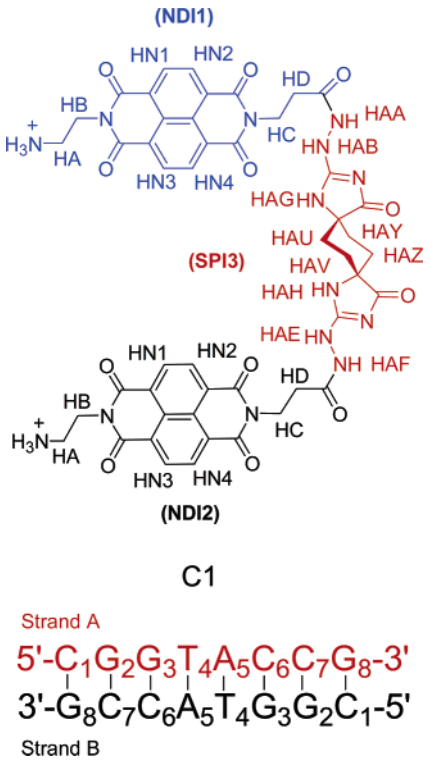

Figure 1. Chemical structure and naming scheme of ligand **C1** and the d(CGGTACCG)$_2$ oligonucleotide sequence used in the NMR studies.

pair (bp) DNA sequence in a threading polyintercalation mode, with linkers alternating in the expected minor groove, major groove, minor groove pattern.[13]

In an attempt to synthesize threading polyintercalators with enhanced affinity and altered specificity, we have designed a rigid, functionalizable linker to replace the flexible, linear peptide-based linkers we have used in our previous NDI-based threading polyintercalators. 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 co-workers have achieved remarkable increases in both the affinity and specificity of binding by choosing appropriate linkers to connect two anthracycline units.[5a−c]

The bisintercalator **C1** with a novel, rigid linker has been synthesized and found to bind poly(dGdC) with high binding affinity (Figure 1).[14] DNAse I footprinting suggested that **C1** binds six bp GGNNCC sequences (Chu and Iverson, unpublished results). Considering that **C1** has 2-fold symmetry, a nonnatural 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 ($K_a \approx 10^7$ M^{-1}), 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.[15]

In the present study, 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.

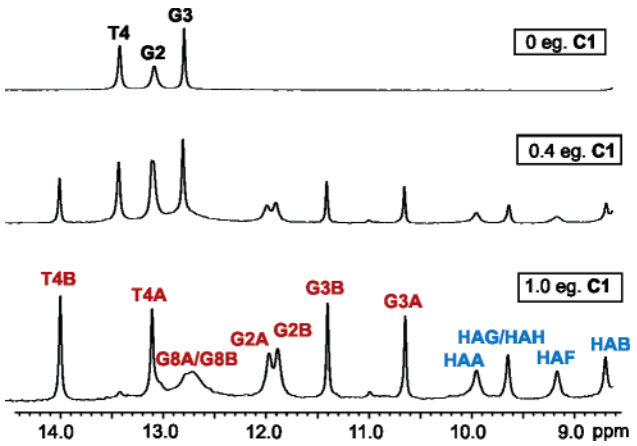

Figure 2. 1D ^{1}H NMR spectra of **C1** titration into d(CGGTACCG)$_2$ in H_2O/D_2O (9:1) with 30 mM sodium phosphate buffer at 25 °C.

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 2). When 0.4 equiv 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 the intensity decreased. These observations indicate a relatively tight and specific binding between **C1** and d(CGGTACCG)$_2$, with the association and disassociation processes being slow on the NMR time scale. After addition of 1.0 equiv 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.

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 d(CGGTACCG)$_2$ is not retained in the complex. The most notable chemical shift differences between free DNA and the 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.[16] 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).

The loss of 2-fold symmetry for the **C1**−d(CGGTACCG)$_2$ complex is unexpected, since both **C1** and d(CGGTACCG)$_2$ have 2-fold symmetry prior to mixing together. This 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.[17]

(13) Lee, J.; Guelev, V.; Sorey, S.; Hoffman, D. W.; Iverson, B. L. *J. Am. Chem. Soc.* **2004**, *126*, 14036−14042.
(14) Chu, Y.; Lynch, V.; Iverson, B. L. *Tetrahedron* **2006**, *62*, 5536−5548.
(15) Mazzitelli, C. L.; Chu, Y.; Reczek, J. J.; Iverson, B. L.; Brodbelt, J. S. *J. Am. Soc. Mass Spectrom.*, in press.
(16) Feigon, J.; Denny, W. A.; Leupin, W.; Kearns, D. R. *J. Med. Chem.* **1984**, *27*, 450−465.
(17) Robinson, H.; Priebe, W.; Chaires, J. B.; Wang, A. H.-J. *Biochemistry* **1997**, *36*, 8663−8670.

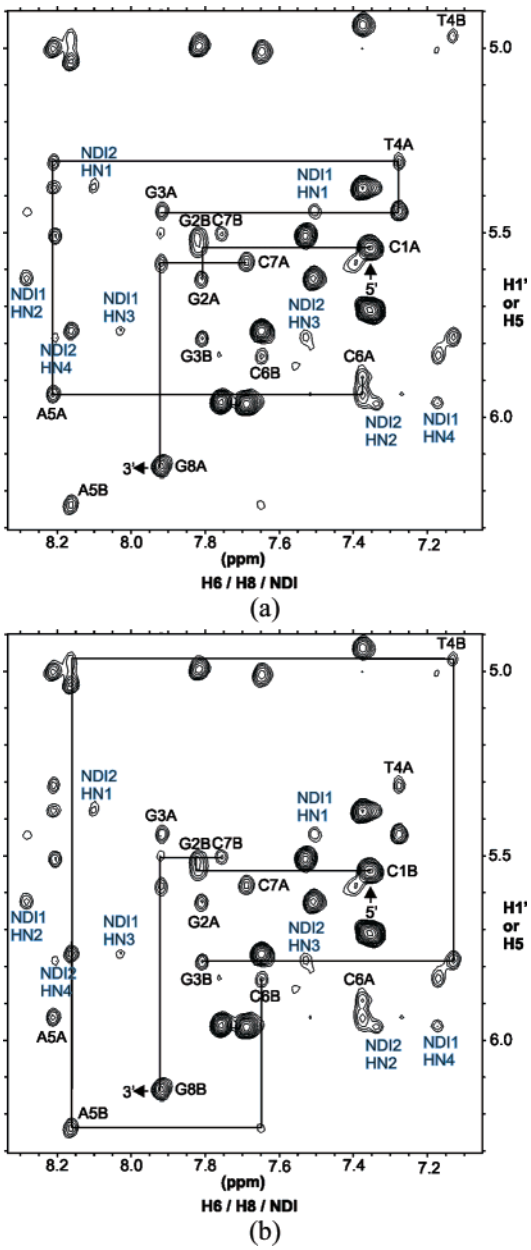

Figure 3. Contour plot of the 2D NOESY spectrum (D_2O, 60 ms mixing time, 25 °C) of the **C1**−d(CGGTACCG)$_2$ complex showing the DNA H6/H8 to H1′ connectivities: (a) strand A; (b) strand B. Note that the connectivities are interrupted at the G2/G3 and C6/C7 steps.

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,[15] incubation of d(GGATCC)$_2$, d(GGCGCC)$_2$, and d(GGGCCC)$_2$ with 1.0 equiv 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 d(GGGCCC)$_2$ (Figure S1, Supporting Information), 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.

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.[18,19] The internucleotide NOE connectivities for the **C1**−d(CGGTACCG)$_2$ complex are weak or interrupted at steps G2−G3 and C6−C7 (Figures 3 and S2, Supporting Information). A TOCSY spectrum showed four strong cross-peaks in the aromatic-to-aromatic region, arising from the aromatic protons of two NDI units.

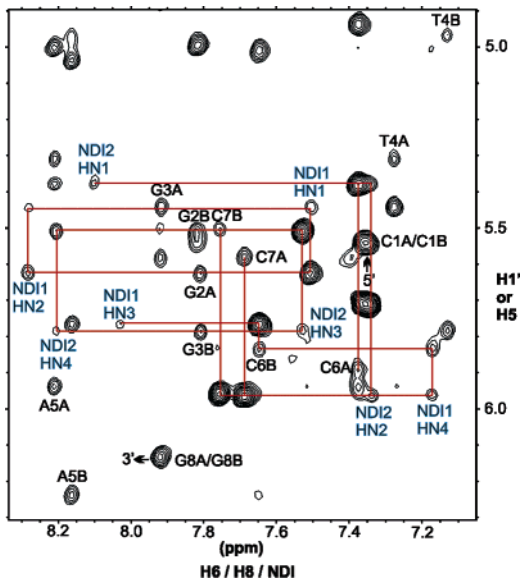

Figure 4. Intermolecular NOEs between protons from the ligand (two NDI units) and DNA d(CGGTACCG)$_2$ (H1′ and H5) at the intercalation sites (D_2O, 60 ms mixing time, 25 °C).

These assignments were further supported by the NOE cross-peaks observed between protons from NDI and d(CGGTACCG)$_2$ (Figure 4). 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. NOEs were also seen between the exchangeable protons on the linker of **C1** and H1′ and H2 of A5A and A5B and H1′ of C6B (Figure 5), 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.

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 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 that reported previously with a different threading bisintercalator, **G$_3$K**, which binds in the major groove of the same d(CGGTACCG)$_2$ sequence.[10]

The conformations of the ribose rings were determined from a qualitative analysis of the intensities of intranucleotide cross-peaks in the NOE spectrum.[20] 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

(18) Hare, D. R.; Wemmer, D. E.; Chou, S.-H.; Drobny, G.; Reid, B. R. *J. Mol. Biol.* **1983**, *171*, 319−336.
(19) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; J. Wiley & Sons: New York, 1986.
(20) Wijmenga, S. S.; Mooren, M. 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, 1993.

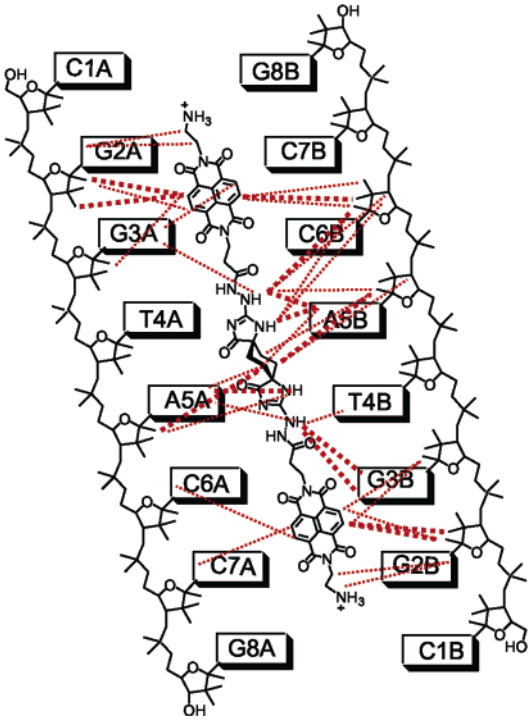

Figure 5. Diagram of the observed intermolecular NOEs in the **C1**−d(CGGTACCG)$_2$ complex. The thick dashed lines indicate "strong" and "medium" NOEs; thin dotted lines represent weak NOEs.

complexed with d(CGTACG)$_2$.[21] The NMR data support nucleotides C1A, C1B, G2A, and C6A having an N-type (C3′-endo sugar pucker) conformation.

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.[22] A possible relationship between the ring flipping rate and the antitumor activity of a given intercalator was proposed.[22]

As shown in the aromatic−aromatic region of the NOESY spectrum (Figure 6), 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,[23] ruling out the possibility of these cross-peaks being caused by dipolar relaxation. Similar unusual cross-peaks have been observed for both NDI rings in a minor groove binding bisintercalator, **(*β*-Ala)$_3$K**.[11] 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.[11] 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

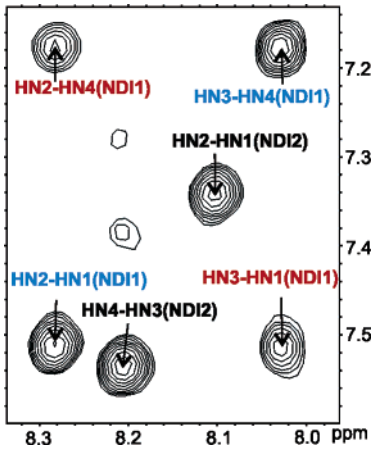

Figure 6. Aromatic region of the 2D NOESY spectrum of the complex **C1**−d(CGGTACCG)$_2$ (25 °C, 60 ms mixing time).

Table 1. Statistics for an Ensemble of the Eight Lowest Energy **C1**−d(CGGTACCG)$_2$ Structures

Restraints for the Structural Calculation	
total number of distance restraints	227
number of DNA−DNA NOEs	145
number of DNA−ligand NOEs	48
number of 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.

annealing protocols within the CNS program suite (Table 1).[24] 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 7a, with the median structure shown separately in Figure 7b. 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 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.

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 7c). 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

(21) (a) Wang, A. H.; Ughetto, G.; Quigley, G. J.; Rich, A. *Biochemistry* **1987**, *26*, 1152−1163. (b) Adams, A.; Guss, J. M.; Collyer, C. A.; Denny, W. A.; Wakelin, L. P. *Biochemistry* **1999**, *38*, 9221−9233. (c) Wang, A. H.; Ughetto, G.; Quigley, G. J.; Rich, A. *J. Biomol. Struct. Dyn.* **1986**, *4*, 319−342.
(22) Gallego, J.; Reid, B. R. *Biochemistry* **1999**, *38*, 15104−15115.
(23) Lokey, R. S.; Iverson, B. L. *Nature (London)* **1995**, *375*, 303−305.
(24) 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. *Acta Crystallogr.* **1998**, *D54*, 905−921.

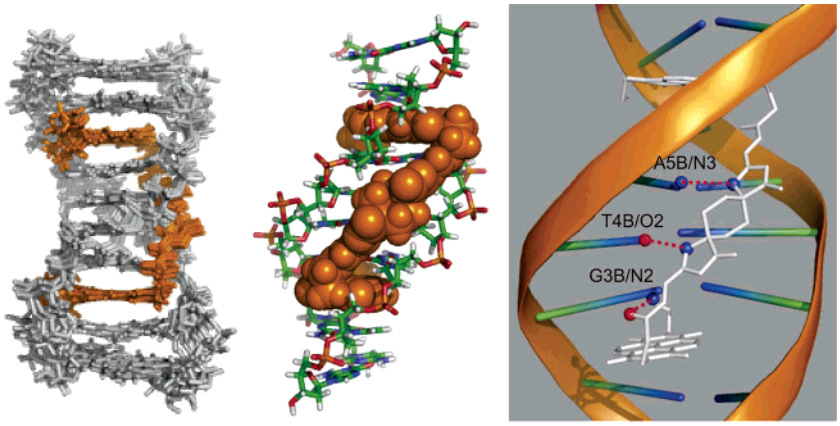

Figure 7. (a) Superposed structures of the **C1**−d(CGGTACCG)$_2$ complex of the final eight ensembles with the lowest energy. (b) Space-filling model showing the linker in the minor groove. (c) Proposed hydrogen bonds: top, N-HAG−N3/A5B; middle, N-HAH−O2/T4B; bottom, N2-H2 (G3B)−O═C (NDI2). Hydrogen bonds were proposed when heavy atoms were between 2.5 and 3.4 Å. The figures were created in PyMOL.[25]

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 the 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.[22] 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[26] and assessing the bend of DNA structures.[27] 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.[28−30] 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.[31] Pf1-phage has been described as inert to nucleic acid, merely serving as the alignment medium to create some anisotropic magnetic environment.[31] 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.

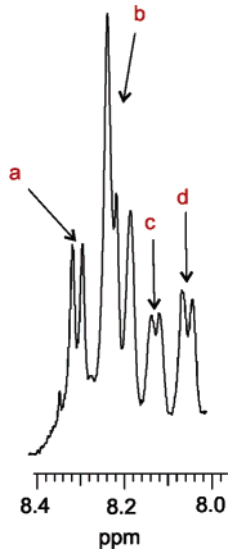

Figure 8. RDCs for four pairs of aromatic H atoms on two NDI units. Ordinary scalar coupling and RDC each contribute to the observed splitting of the peaks. The derived RDCs for the proton pairs are (a) HN1−HN2 (NDI1), 4.5 ± 0.3 Hz, (b) HN3−HN4 (NDI2), 2.3 ± 0.3 Hz, (c) HN1−HN2 (NDI2), 2.3 ± 0.3 Hz, and (d) HN3−HN4 (NDI1), 4.3 ± 0.3 Hz.

Addition of Pf1-phage to a sample of the **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 8), consistent with the appearance 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 for 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 planes of the aromatic rings of the two NDI groups are not parallel (Figure 9); however, it should be noted that this interpretation assumes that the rates of ring flipping do not differently affect the observed RDCs in the two NDI units.

(25) DeLano, W. L. *The PyMol Molecular Graphics System*; DeLano Scientific: San Carlos, CA, 2002.
(26) Lipsitz, R. S.; Tjandra, N. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 387−413.
(27) MacDonald, D.; Lu, P. *Curr. Opin. Struct. Biol.* **2002**, *12*, 337−343.
(28) Sibille, N.; Pardi, A.; Simorre, J. P.; Blackledge, M. *J. Am. Chem. Soc.* **2001**, *123*, 12135−12146.
(29) Jaroniec, C. P.; Boisbouvier, J.; Tworowska, I.; Nikonowicz, E. P.; Bax, A. *J. Biomol. NMR* **2005**, *31*, 231−241.
(30) Boisbouvier, J.; Delaglio, F.; Bax, A. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 11333−11338.
(31) Hansen, M. R.; Mueller, L.; Pardi, A. *Nat. Struct. Biol.* **1998**, *5*, 1065−1074.

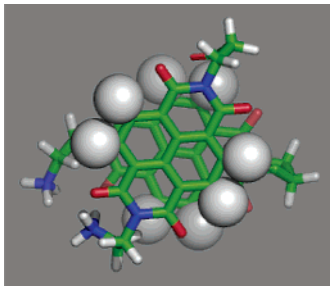

Figure 9. Skeletal drawing of two NDI units in the model indicating an angle of ~60° on overlap, viewed along the long axis of the complex **C1**−d(CGGTACCG)$_2$ (all other atoms are not shown). The figure was created in PyMOL.[25]

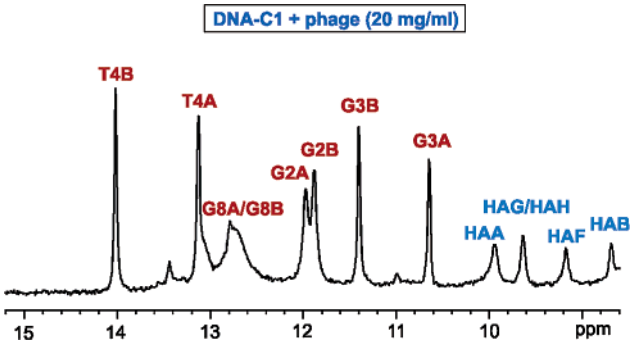

Figure 10. 1D ^{1}H NMR spectrum of the **C1**−d(CGGTACCG)$_2$ complex after addition of Pf1-phage (final concentration ~20 mg/mL) in H_2O/D_2O (9:1) and 30 mM sodium phosphate buffer.

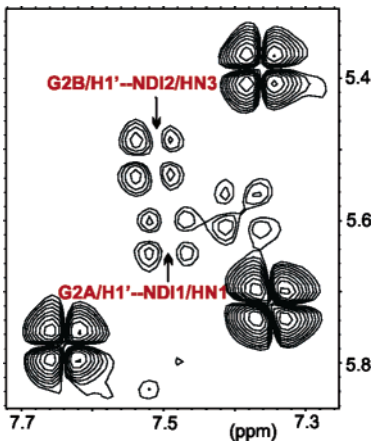

Figure 11. 2QF-COSY spectrum of **C1**−d(CGGTACCG)$_2$ using Pf1-phage as the cosolvent showing RDCs between the intercalators and G2.

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 Figure 2 to Figure 10).

Through-space dipolar couplings were also observed between G2B/H1′ and NDI2/HN3 and between G2A/H1′ and NDI1/HN1, which were detected as new peaks appearing in the 2QF-COSY spectrum after the addition of phage (Figure 11). These internucleotide dipolar couplings are entirely consistent with the structural model of the **C1**−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 internucleotide dipolar couplings were also identified, including C1A/H2″−G2A/H8 and G3B/H2″−T4B/H6 (not shown), also consistent with the structural model.

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 idealization of the rigidity of their connections.

The detailed structure of the 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, noncharged, 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 provides the correct spacing between the two NDI aromatic units.

Previous quantitative DNAse I footprinting results suggested that **C1** binds best to the 5′-GGTACC-3′ sequence, with an affinity of $K_a \approx 10^7$ M^{-1}.[15] 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 the minor groove is indicated by not only the numerous NOEs observed between the central cyclohexane ring and the edge of the DNA bases (mostly to A5A and A5B), but also the loss of *C*2 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 nonexchangeable protons on the cyclohexane ring as expected for equatorial and axial protons, respectively.[14] 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 2 in the free state to 6 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 the cyclohexane ring upon binding gives rise to the loss of the *C*2 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.[15] In that case, structural analysis led to the conclusions that the glycine$_3$−lysine linker was the proper length and possessed electrostatics complementary to that of 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-Ala)$_3$Lys**,[11] while for **G$_3$K** the length is 11.8 Å (from the top C═O to the bottom N−H).[10] 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.[32] Consistent with this notion, our two bisintercalators with slightly longer lengths (**C1** and **(β-Ala)$_3$Lys**) 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 7c. 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.

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 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 **G$_3$K** 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. These and many other studies are currently under way in our laboratory.

(32) Saenger, W. In *Principles of Nucleic Acid Structure*; Cantor, C. R., Ed.; Springer-Verlag: New York, 1984.

Materials and Methods

Sample Preparation. Compound **C1** synthesis has been described in detail elsewhere.[14] The DNA (gel filtration grade, Midland Certified, Midland, TX) was initially dissolved in 0.7 mL of 30 mM sodium 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 of 99.9% D_2O (Cambridge Isotope, Cambridge, MA). 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 times), 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 (t_2) and 512 (t_1) 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[33] and displayed using Sparky.[34]

For RDC experiments, the sample was prepared by adding a certain volume of 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 storing 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 sodium 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.[24] Distance restraints for nonexchangeable protons were derived from the observed cross-peak intensities in the D_2O 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 involving methylene protons were determined by adding an extra 0.7 Å to the distances measured from the center of the methylene group. The cross-peak intensities in the 200 ms NOESY spectrum using 9:1 H_2O/D_2O as the solvent were used to produce the distance restraints for exchangeable protons. The dna-rna_restraints.def file within the CNS program was used to define hydrogen bonds for those identified Watson−Crick base pairs. The base pair planarity (±10°) 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 ($\alpha = -46° \pm 40°$, $\beta = -147° \pm 40°$, $\gamma = 36° \pm 40°$, $\delta = 157° \pm 40°$, $\epsilon = 155° \pm 40°$, $\zeta = -96° \pm 40°$) for the base pairs at intercalation sites, and a ±25° range for the rest of the base pairs.[13] On the basis of 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**

(33) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277−293.
(34) Goddard, T. D.; Kneller, D. G. *SPARKY 3*; University of California: San Francisco, 2004.

were created with the assistance of Dundee PRODRG2 Server[35] and Gerard Kleywegt's XPLO2D Server (version 050802/3.3.2).[36] The atomic charges and bond lengths of **C1** were either adopted from the crystal structure of an NDI derivative[23] or calculated using the ab initio method in HyperChem 7.0 (Hypercube Inc., 1115 NW 4th St., 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 60 ps of torsion-angle molecular dynamics was applied to those starting structures at 20000 K to allow a high degree of randomization of the initial models. In the second stage the temperature was slowly decreased to 2000 K over a period of 60 ps and 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.[13]

Acknowledgment. This paper is dedicated to Professor Peter Dervan on the occasion of his 60th birthday. This work was supported by the National Institutes of Health (Grant GM-069647) and a grant from the Welch Foundation (F-1353). We thank Dr. Vladimir Guelev for helpful discussions on RDC experiments.

Supporting Information Available: Tables of DNA sugar conformation analysis and chemical shifts of DNA protons and figures of 1D NMR spectra of **C1** titration into $d(AGGGCCCT)_2$ and DNA H6/8-H2′/H2″ 2D connectivities of the $d(CGGTACCG)_2$−**C1** complex. This material is available free of charge via the Internet at http://pubs.acs.org.

JA066480X

(35) Schuettelkopf, A. W.; van Aalten, D. M. F. *Acta Crystallogr.* **2004**, *D60*, 1355−1363.
(36) Kleywegt, G. J.; Jones, T. A. *Methods Enzymol.* **1997**, *277*, 208−230.