J|A|C|S ARTICLES

Published on Web 02/23/2009

A Pseudocatenane Structure Formed between DNA and A Cyclic Bisintercalator

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

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

Received July 21, 2008; E-mail: biverson@mail.utexas.edu

Abstract: Targeting double-stranded DNA with small molecules remains an active area of basic research. Herein is described a cyclic DNA bisintercalator that is based on two naphthalene diimide (NDI) intercalating units tethered by one linking element specific for binding in the minor groove and the other linking element specific for binding in the major groove. DNase I footprinting revealed a strong preference for binding the sequence 5′-GGTACC-3′. NMR structural studies of the complex with $d(CGGTACCG)_2$ verified a pseudocatenane structure in which the NDI units reside four base pairs apart, with one linker segment located in the minor groove and the other in the major groove consistent with the linker designs. To the best of our knowledge, this is the first structurally well-characterized pseudocatenane complex between a sequence specific cyclic bisintercalator and intact DNA.

Introduction

The DNA molecule remains an important focus of research geared toward developing new approaches to cancer and antimicrobial therapies. An advantage of targeting DNA with small molecules is the potentially general nature of the approach in that once sequence selectivity and delivery are solved for one case, using similar approaches to treating a wide variety of other diseases should be enabled. Over the past few decades, several approaches for DNA recognition have been explored. A series of polyamide oligomers developed by Dervan and co-workers that target the minor groove of DNA have proven to be a particularly successful and general approach.[1] Follow-up studies showed that these polyamides can function as gene regulators in vivo. Another well-studied approach to sequence specific binding of DNA targets the major groove using a third oligodeoxynucleotide strand.[2] In a related approach, so-called PNAs, peptide nucleic acids, can be used to target both single and double-stranded DNA sequences.[3−6]

We have been pursuing an alternative approach that we have termed threading polyintercalation. By tethering a threading intercalating unit, naphthalene diimide (NDI), with appropriately chosen linkers, molecules have been produced and characterized that bind selectively to specific sequences of DNA with all the NDI units fully intercalated while the linkers reside in the minor and major grooves in an alternating fashion.[7−10] Threading polyintercalators have attracted attention because they interact strongly with both DNA grooves, and the threading intercalation interaction is generally characterized by a relatively slow dissociation rate from DNA.[11,12]

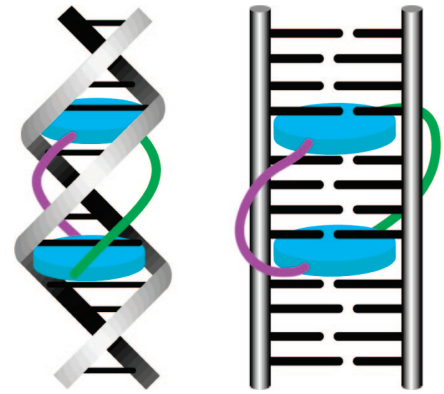


Figure 1. Schematic representations of a cyclic bisintercalator bound to DNA to form a pseudocatenane complex.

Cyclic bisintercalation is shown conceptually in Figure 1. The double-stranded DNA structure is known to be highly dynamic in solution, with base pairs opening and reforming rapidly at room temperature. For example, the imino protons, which are intrinsic to the Watson−Crick hydrogen bonds between base pairs, exhibit exchange half-lives that are typically in the 1−50 ms range.[13] A cyclic bisintercalator may be able to exploit this millisecond "breathing" behavior of duplex DNA by sliding through the temporarily disrupted DNA base pairs to form a pseudocatenane-like complex structure.

(1) Dervan, P. B. *Bioorg. Med. Chem.* **2001**, *9*, 2215–2235.
(2) Moser, H.; Dervan, P. B. *Science* **1987**, *238*, 645–650.
(3) Nielsen, P. E.; Egholm, M.; Berg, R. H.; Buchardt, O. *Science* **1991**, *254*, 1497–1500.
(4) Kaihatsu, K.; Huffman, K, E.; Corey, D. R. *Biochemistry* **2004**, *43*, 14340–14347.
(5) Hu, J.; Corey, D. R. *Biochemistry* **2007**, *46*, 7581–7589.
(6) Englund, E. A.; Xu, Q.; Witschi, M. A.; Appella, D. H. *J. Am. Chem. Soc.* **2006**, *128*, 16456–16457.
(7) Guelev, V.; Lee, J.; Ward, J.; Sorey, S.; Hoffman, D. W.; Iverson, B. L. *Chem. Biol.* **2001**, *8*, 415–425.
(8) Guelev, V.; Sorey, S.; Hoffman, D. W.; Iverson, B. L. *J. Am. Chem. Soc.* **2002**, *124*, 2864–2865.
(9) Lee, J.; Guelev, V.; Sorey, S.; Hoffman, D. W.; Iverson, B. L. *J. Am. Chem. Soc.* **2004**, *126*, 14036–14042.
(10) Chu, Y.; Sorey, S.; Hoffman, D. W.; Iverson, B. L. *J. Am. Chem. Soc.* **2007**, *129*, 1304–1311.
(11) Tanious, F.; Yen, S.; Wilson, W. D. *Biochemistry* **1991**, *30*, 1813–1819.
(12) Yen, S.-F.; Gabbay, E. J.; Wilson, W. D. *Biochemistry* **1982**, *21*, 2070–2076.

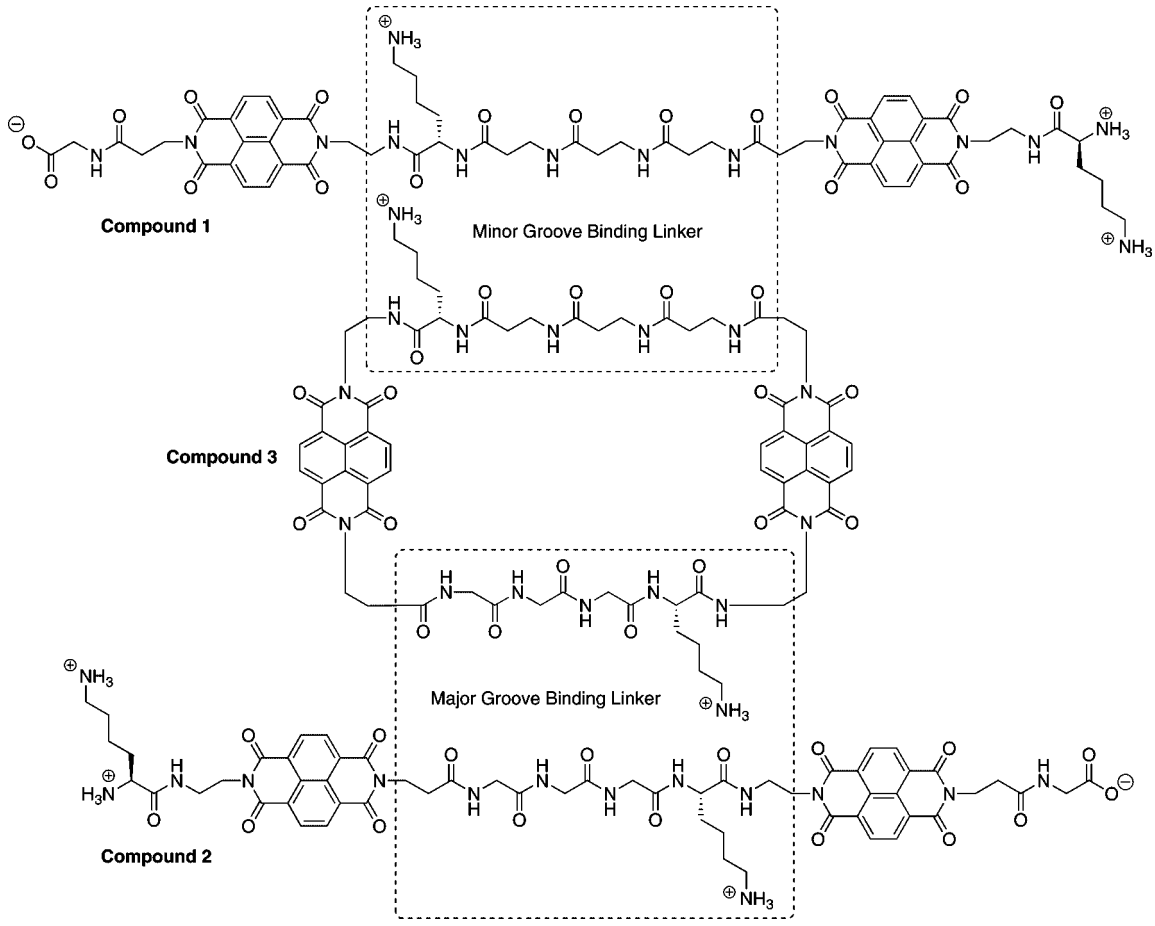


Figure 2. Structures of previously reported linear bisintercalators **1** and **2** along with cyclic bisintercalator **3** highlighting the minor and major groove binding linker elements.

A cyclic binsintercalator, and the pseudocatenane structure formed upon its complexation with DNA, could prove interesting for a number of reasons. For example, characterization of cyclic bisintercalator association kinetics could be used to probe DNA breathing dynamics. In addition, it is reasonable to anticipate extremely slow dissociation kinetics because significant distortion of the DNA double helix would be required for release of a bound bisintercalator. Finally, a cyclic bisintercalator would have the unprecedented ability to interact with DNA base pairs in both grooves simultaneously, providing unique opportunities for powerful sequence recognition.

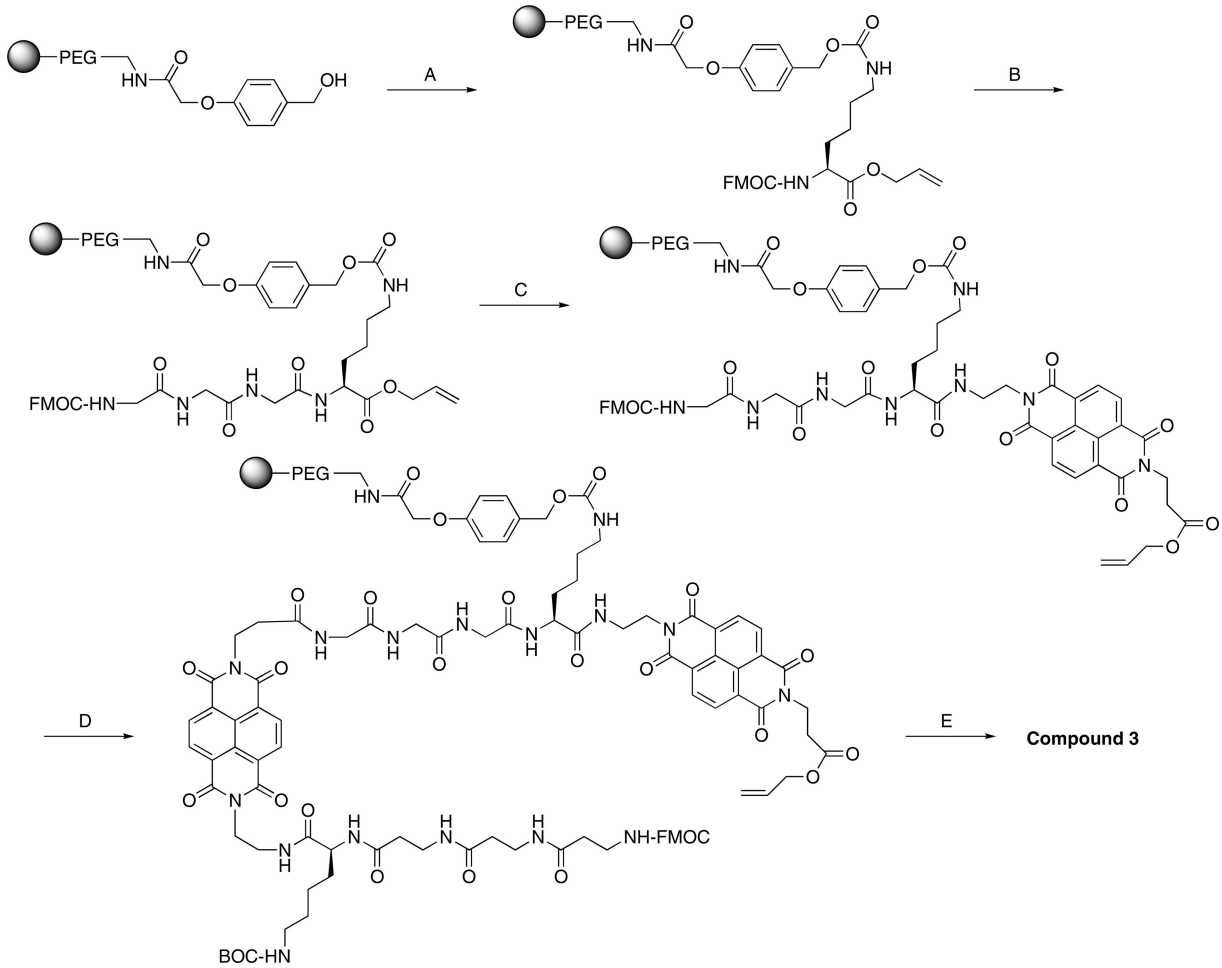
Molecules intended to act as cyclic DNA bisintercalators have been studied before, but to the best of our knowledge, there has never been definitive structural characterization of these molecules when bound sequence selectively to intact double-stranded DNA. For example, a macrocyclic bisacridine derivative was reported and its binding behavior toward duplex DNA characterized.[14,15] Evidence for more than one binding mode was obtained, involving complexes that spanned either one or two base pairs. Another cyclic bisacridine with two polyammonium connecting chains has also been studied for its ability to recognize abasic sites within duplex DNA.[16] A major complex between DNA (11mer) and the cyclic ligand was identified in which one acridine unit stacked within the abasic site as expected. However, a minor complex was also seen but not structurally characterized.

Previous work in our laboratory has demonstrated that groove selectivity and sequence specificity of threading polyintercalators is modulated in dramatic fashion by the linking units. For example, the linear bisintercalator **1**, having a linker composed of $(\beta\text{-Ala})_3$Lys (Figure 2), specifically interacts with d(GATAAG)·d(CTTATC) and has the linker bound in the minor groove.[8] On the other hand, bisintercalator **2**, with the linker Gly_3Lys, prefers to bind $d(GGTACC)_2$, and in this case the linker is found in the major groove.[7] Based on these structures, a tetraintercalator (not shown) was synthesized, and NMR analysis verified that when bound to DNA, it exhibited a unique topology with linkers residing in alternating grooves (minor groove−major groove−minor groove) analogous to how similar linkers behaved in the linear binsintercalators **1** and **2**.[9]

Herein is described the synthesis and characterization of a first generation NDI-based cyclic bisintercalator, **3**, which combines the linkers from linear bisintercalators **1** and **2**. These linkers were chosen because they reside in the minor and major grooves, respectively. In particular, we have previously proposed that the three additional methylene units of the $(\beta\text{-Ala})_3$Lys linker of **1** provide more length and hydrophobic surface area that combine to favor binding in the minor groove, while the

(13) Guéron, M.; Leroy, J. L. In *Nucleic Acids and Molecular Biology*; Eckstein, F., Lilley, D. M. J., Eds.; Springer-Verlag: New York, 1992; Vol. 6, pp 1−22.
(14) Zimmerman, S. C.; Lamberson, C. R.; Cory, M.; Fairley, T. A. *J. Am. Chem. Soc.* **1989**, *111*, 6805–6809.
(15) Veal, J. M.; Li, Y.; Zimmerman, S. C.; Lamberson, C. R.; Cory, M.; Zon, G.; Wilson, W. D. *Biochemistry* **1990**, *29*, 10918–27.
(16) Jourdan, M.; Garcia, J.; Lhomme, J.; Teulade-Fichou, M.-P.; Vigneron, J.-P.; Lehn, J.-M. *Biochemistry* **1999**, *38*, 14205–14213.

Scheme 1. Solid-Phase Synthesis of Cyclic Bisintercalator **3**[a]



[a] A. (1) *N,N′*-disuccinimidyl carbonate (DSC); (2) Fmoc-Lys-OAll; B. Fmoc SPPS, three couplings; C. (1) $Pd(PPh_3)_4$; (2) NH_2-$(CH_2)_2$-NDI-$(CH_2)_2$-CO_2All; D. (1) 20% piperidine/DMF; (2) Fmoc-Lys(Boc)-$(CH_2)_2$-NDI-$(CH_2)_2$-CO_2H; (3) Fmoc SPPS, three couplings; E. (1) $Pd(PPh_3)_4$; (2) 20% piperidine/DMF; (3) PyBOP, HOAt; (4) TFA/triisopropylsilane (TIPS)/H_2O.

shorter Gly_3Lys linker of **2** is too short for minor groove binding but can be accommodated in the wider major groove where the distance required to connect adjacent intercalated NDI units is shorter because the linker can "cut across" the groove in a more diagonal fashion.

Interactions of **3** binding to DNA were studied using DNase I footprinting to establish sequence specificity. Detailed NMR studies confirmed the expected pseudocatenane structure of **3** bound to its preferred sequence of double-stranded DNA, representing the first example of a structurally characterized cyclic bisintercalator complex with intact DNA.

Results

Synthesis of Cyclic Bisinteralator 3. The presence of the lysine residue on **3** makes it possible to carry out a solid-phase synthesis in which the lysine side chain is attached to the resin, enabling synthesis on either the N-terminal or C-terminal ends during construction of the cyclized structure. The low loading PEG-based NovaSyn TGA resin was used as the solid support and an acid-labile carbamate linkage was used to attach the ε-amino of lysine[17] (Scheme 1). Three glycine residues were attached to the lysine N-terminus using standard solid-phase peptide synthesis procedures, then the allyl function on the carboxyl group of lysine was removed by treatment with $Pd(PPh_3)_4$.[18] The NDI monomer, NH_2-$(CH_2)_2$-NDI-$(CH_2)_2$-CO_2All (Scheme S1, Supporting Information), was then attached to the C-terminus. Next, on the N-terminal side, another NDI monomer, Fmoc-Lys(Boc)-$(CH_2)_2$-NDI-$(CH_2)_2$-CO_2H was attached,[19] followed by attachment of three β-alanine residues. The C-terminal allyl protecting group on the C-terminal side was removed. The final cyclization step was completed in high yield after 18 h using PyBOP/1-hydroxy-7-azabenzotriazole (HOAt) under weakly basic conditions (Scheme 1). Interestingly, other activating reagent combinations such as PyBOP/1-hydroxybenzotriazole (HOBt) or *N,N′*-diisopropylcarbodiimide (DIPICI)/HOBt proved to be significantly less effective for cyclization.

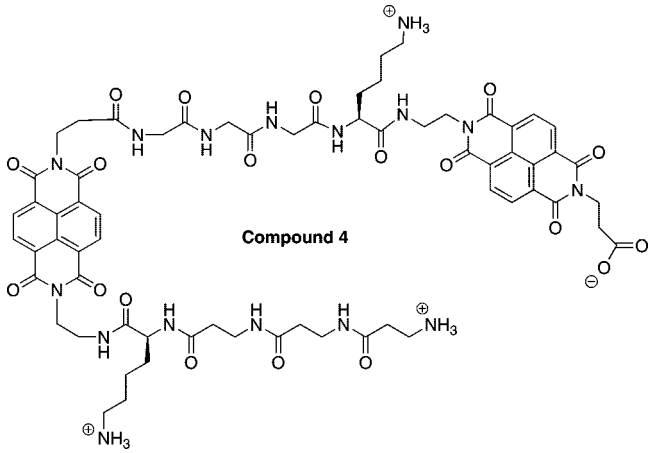
The analytical HPLC profiles for the crude material cleaved from the resin before and after the cyclization displayed only one major species each, indicating that the cyclization step proceeded with extremely high efficiency (Figure S1).

(17) Alsina, J.; Rabanal, F.; Chiva, C.; Giralt, E.; Albericio, F. *Tetrahedron* **1998**, *54*, 10125–10152.

(18) Kates, S. A.; Daniels, S. B.; Sole, N. A.; Barany, G.; Albericio, F. In *Peptides, Chemistry, Structure & Biology, Proceedings of the 13th American Peptide Symposium*; Hodges, R. S., Smith, J. A., Eds.; ESCOM: Leiden, 1994; pp 113−115.

(19) Guelev, V. M.; Cubberley, M. S.; Murr, M. M.; Lokey, R. S.; Iverson, B. L. *Methods Enzymol.* **2001**, *340*, 556–570.

To serve as a noncyclized control, compound **4** was prepared as the linear version of **3**, isolated prior to cyclization.



DNase I Footprinting of 2, 3, and 4. DNase I footprinting studies were carried out using a synthetic double-stranded sequence (see below) designed to probe whether **3** prefers to bind 5′-GGTACC-3′, which is favored by the Gly_3Lys linker (compound **2**), or 5′-GATAAG-3′ preferred by the $(\beta\text{-Ala})_3Lys$ linker (compound **1**). Note that there are actually three 5′-GATAAG-3′ sites to serve as an extra check for specificity (vide infra).

(+) strand:

5' GGT AAT TTA TTA TGG TAC CAT ATT AAT TGA TAA GTA CTT ATC ATT AAT TAT GAT AAG TTA TAA TTA ACC 3'

(-) strand:

5' GGT TAA T TA TAA CTT ATC ATA ATT AAT GAT AAG TAC TTA TCA ATT AAT ATG GTA CCA TAA TAA ATT ACC 3'

The DNase I footprinting data is shown in Figure 3. On the basis of previous work, it was expected that the linear dimer compound **2** would bind with high affinity to the single 5′-GGTACC-3′ sequence, designated as sequence 1 in Figure 3. Unexpectedly, **2** also binds to another palindromic sequence, 5′-AGTACT-3′ (designated as sequence 2 in Figure 3), a specificity that has not been observed previously for this molecule. Both the cyclic compound **3** and the linear analogue compound **4** bind preferentially to the 5′-GGTACC-3′ sequence ($K_D \sim 10^{-7}$ M), although compound **4** shows a less distinct footprint, possibly a result of its more flexible nature. For compound **3**, no footprints were found at any of the 5′-GATAAG-3′ sequences. A reasonable conclusion to draw from these data is that the Gly_3Lys linkage dominates the binding specificity of **3**. In addition, incubating **3** with DNA for 38 h (data not shown) did not show any significant change in specificity compared to an incubation time of 1.5 h, although the longer incubation time did produce a more clear footprint at 125 nM.

Titration of d(CGGTACCG)$_2$ with Compound 3. Based on the preference of **3** for binding 5′-GGTACC-3′, an NMR structural analysis was carried out using the self-complementary oligonucleotide 5′-CGGTACCG-3′. The naming convention of **3** and the d(CGGTACCG)$_2$ duplex is shown in Figure 4. A titration of d(CGGTACCG)$_2$ with **3** was monitored by 1D ^{1}H NMR spectroscopy (Figure 5). Upon adding 0.5 equiv of **3** to the DNA duplex, several new sharp peaks appeared while the peaks of the free DNA remained unchanged in both position

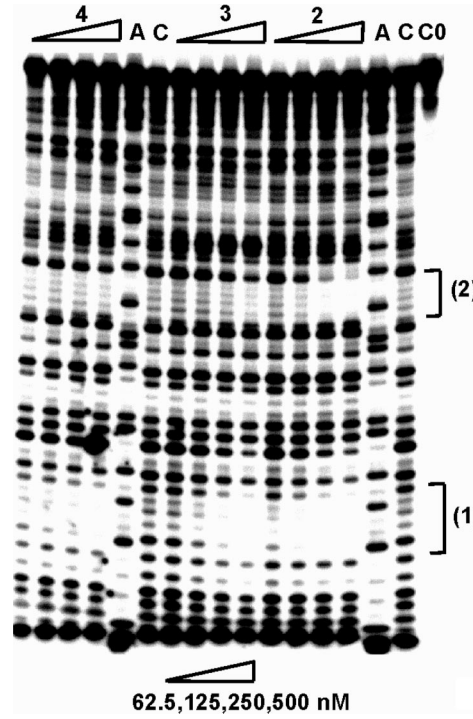


Figure 3. DNase I footprinting of **2**, **3**, and **4** after 1.5 h incubation. Lane A is an adenine-specific sequencing reaction. Lane C0 contains DNA without DNase I. Lane C contains DNA with DNase I but no ligand. The sequences 5′-GGTACC-3′ (1) and 5′-AGTACT-3′ (2) are highlighted with brackets.

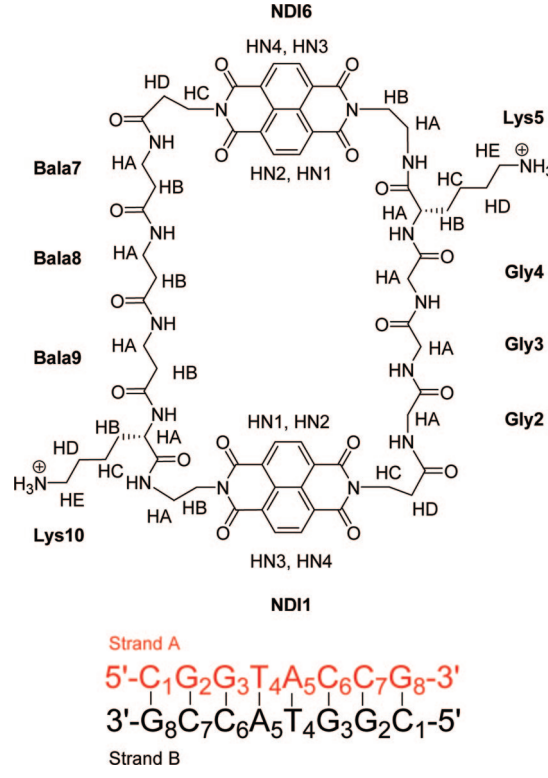


Figure 4. The naming convention of compound **3** and the d(CGGTACCG)$_2$ duplex sequence used in the NMR study.

and shape, although their intensity decreased significantly. These observations indicate a relatively tight and specific binding between **3** and d(CGGTACCG)$_2$, with dissociation kinetics being slow; no effects on line widths indicative of chemical

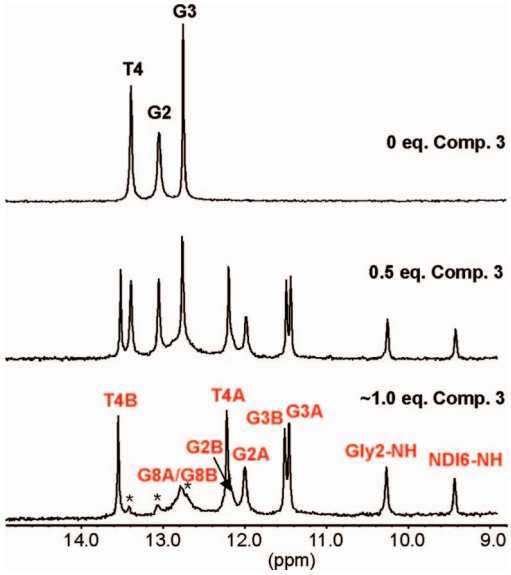


Figure 5. 1D ^{1}H NMR spectra of **3** titrated into d(CGGTACCG)$_2$ in H_2O/D_2O (9:1) with 30 mM Na phosphate buffer at 25 °C (* trace amount of free DNA left in solution).

exchange on the NMR time scale are apparent. After adding 1.0 equiv of **3**, the peaks from the free DNA disappeared, being replaced by the same peaks that had appeared upon addition of 0.5 equiv of **3**. These data suggest specific binding as a single species between **3** 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_2O/D_2O, at 10 or 25 °C; see below).

The **3**−d(CGGTACCG)$_2$ complex does not retain the 2-fold symmetry of the palindromic d(CGGTACCG)$_2$ duplex. This is expected because the ligand **3** does not have 2-fold symmetry. 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. Upfield shifting of the imino resonances of base pairs G2-C7 and G3-C6 are consistent with NDI intercalation at these sites.[20] Similarly, resonances in the 9.0−10.4 ppm range were also observed from the exchangeable protons of **3** upon binding (10.3 and 9.4 ppm, respectively).

Assignment of the 3-d(CGGTACCG)$_2$ Complex. The signals from the NDI units in the **3**−d(CGGTACCG)$_2$ complex were identified in the aromatic−aromatic region of the TOCSY (in D_2O) spectrum. Only four cross-peaks appeared, and they were assigned to four pairs of aromatic protons on the two NDI rings as in previous NDI-DNA structures.[8−10] The assignments for the rest of the **3**−d(CGGTACCG)$_2$ complex were made by following the H6/8-H1′ and H6/8-H2′/H2″ NOE connectivities as shown in Figures 6 and 7.[21] The internucleotide NOE connectivities for the d(CGGTACCG)$_2$ duplex were weak or interrupted at the steps G2−G3 and C6−C7 (Figure 6). New connectivities appeared at these steps (Figure 8). Two aromatic protons on NDI6 have NOEs with H1's of G2A, G3A, and C7B. Three aromatic protons on NDI1 were observed having through-space interactions with H1's of G2B and G3B, and H5 of C6A (Figure 8). NOE cross-peaks between H2′/H2's of G2A and NDI6, G2B and NDI1, were also identified (Figure 7). These data confirm that **3** interacts with d(CGGTACCG)$_2$ through intercalation of the two NDI rings between GpG steps.

As shown in Figure 7, H2 of A5B shows NOE cross-peaks with the methylene protons of Bala8 and Bala7, and 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 (Figures S2a and 9, Supporting Information). 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 9 (see also Figure S2b, Supporting Information). For example, strong NOE cross-peaks were evident for the following pairs: Gly1/NH−T4B/Me, Gly2/NH−T4B/Me, Gly2/NH−T4A/Me, Gly3/NH−T4B/Me, and Gly3/NH−T4A/Me. In addition, there are observed NOEs between NH_2 of C6B and the amide proton of Lys5 (8.81 ppm) (Figure 9). Thus, taken together, the NMR data has confirmed that the Gly$_3$Lys linker is located in the major groove.

Weak NOE cross-peaks were seen between aromatic protons on the NDI6 ring and protons from the Lys5 side chain as shown in Figure 7. This indicates that Lys5 is close to the G3A-C6B base pair and. Similarly, that Lys10 is close to the C6A-G3B pair. This information further supports our assignment of the spatial positions of the eight linker amino acids relative to the DNA bases in the complex.

Taken together, the NOE data are consistent with a model in which the NDI units are intercalated into the GG steps 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.

DNA Conformational Analysis. The conformations of the ribose rings were determined from a qualitative analysis of the intensities of intranucleotide cross-peaks in the COSY and NOESY spectra (Table S1).[22] The data indicate that all nucleotides except T4A, A5A, C6A, A5B, and C6B 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)$_2$.[23,24] The nucleotides of T4A, A5A, C6A, A5B, and C6B appeared to be mixtures of S- and N-type (C3′-endo sugar pucker) puckers and were constrained as C3′-endo conformation in the simulation.

In order to determine the DNA backbone conformation 1D ^{31}P and 2D ^{1}H−^{31}P correlated NMR spectra were acquired (Figures S3 and S4, Supporting Information). ^{31}P chemical shifts were assigned using the observed H3′-P and H4′-P correlations in a 2D ^{1}H−^{31}P COSY spectrum. With the exception of the phosphate group between the second and third nucleotides on each DNA strand (specifically, G3A and G3B), ^{31}P chemical shifts were found to be within the range typical of phosphate groups in A-form or B-form DNA. The ^{31}P chemical shifts and couplings were used to limit the possible conformations of the backbone dihedrals during simulations.

(20) Feigon, J.; Denny, W. A.; Leupin, W.; Kearns, D. R. *J. Med. Chem.* **1984**, *27*, 450–465.
(21) Wüthrich, K. *NMR of proteins and nucleic acids*; John Wiley and Sons: New York, 1986.
(22) 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, 1993.
(23) Wang, A. H. -J.; Ughetto, G.; Quigley, G. J.; Rich, A. *Biochemistry* **1987**, *26*, 1152–1163.
(24) Adams, A.; Guss, J. M.; Collyer, C. A.; Denny, W. A.; Wakelin, L. P. *Biochemistry* **1999**, *38*, 9221–9233.

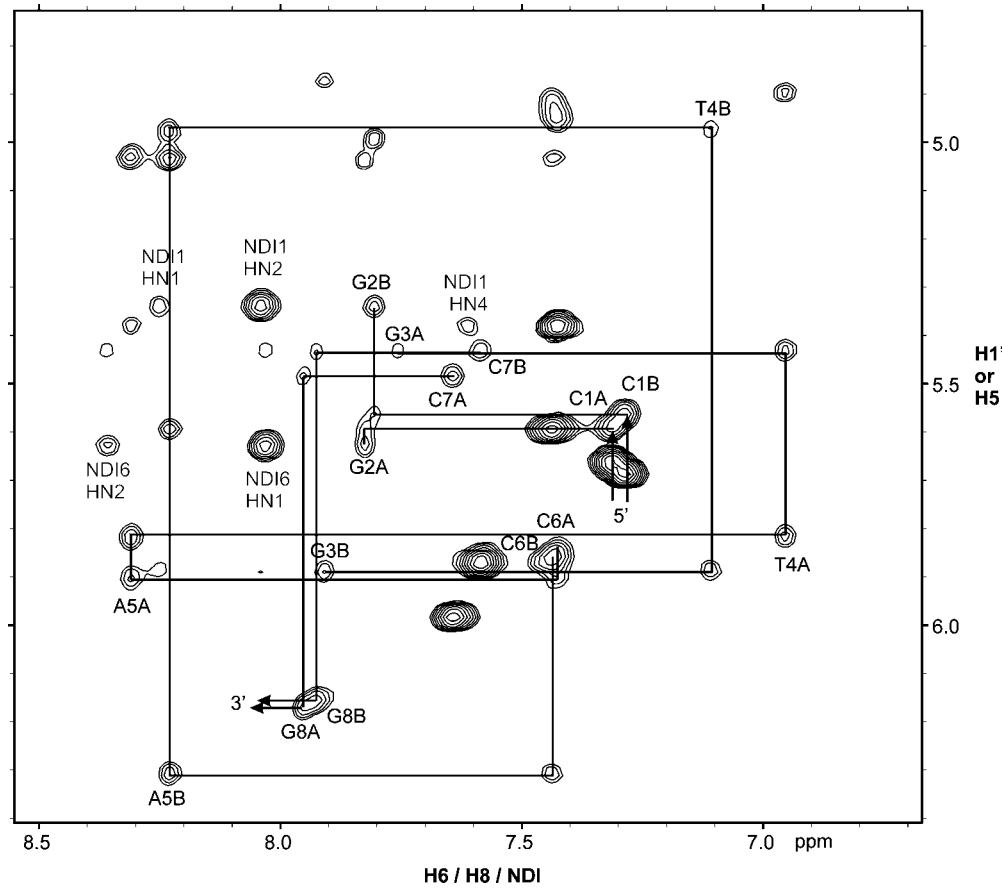


Figure 6. Contour plot of the 2D NOESY spectrum (D_2O, 250 ms mixing time, 25 °C) of the **3**−d(CGGTACCG)$_2$ complex showing the DNA H6/8 to H1′ connectivities for strands A and B. Note that the connectivities are interrupted at the G2/G3 and C6/C7 steps.

Structure Modeling. The structural analysis of the **3**−d(CGGTACCG)$_2$ complex was carried out using the simulated annealing protocols within the CNS program suite (Table 1).[25] A set of 60 starting structures with a wide variety of initial conformations was used for simulated annealing to ensure that at the end of the simulated annealing process, the full range of structures consistent with the NMR-derived structural restraints were identified. The location and orientation of the cyclic bisintercalator was found to be the same in all of the derived structures. An ensemble of eleven lowest-energy structures is shown in Figure 10a−c, with one representative structure shown separately in Figure 10d,e. Close inspection of the ensemble reveals that the DNA bends toward to the major groove slightly upon the binding of ligand **3**. The average rmsd for this family is 2.17 Å for the ligand and d(CGGTACCG)$_2$ (Table 1). In all 11 calculated structures, NDI1 stacks in the G2−G3 step, while NDI6 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 hydrogen bonding between the G-C base pairs remains intact. The Gly$_3$-Lys linker appears to fit snugly in the major groove, with the amide bond formed between Gly3 and Gly4 interacting closely with the pocket produced by two methyl groups from T4A and T4B. The Lys5 side chain is extended across the major groove toward the phosphate backbone of strand B, consistent with earlier findings using compound **2** bound to DNA.[7] On the other hand, the modeling indicated that the (β-Ala)$_3$-Lys linker is interacting more loosely with the narrow minor groove of d(CGGTACCG)$_2$, as there are relatively fewer consistent element overlaps seen with the ligand in the minor groove among the eleven low energy structures. The side chain of Lys10 appears to point away from the minor groove toward bulk solvent. However, in all 11 structures methylene protons HB1/2 of NDI1 are in close contact with the strand A, consistent with the extensive NOEs seen.

Discussion

The on-resin cyclization strategy presented here proved to be a convenient method for the production of **3**. Final purification, which is usually a time-consuming task for such large molecules, was greatly simplified by the efficiencies of the peptide chain elongation and final cyclization steps. This solid-phase synthetic strategy could 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).

To the best of our knowledge, the **3**−d(CGGTACCG)$_2$ complex has produced the first well-resolved 1D and 2D spectra for a cyclic bisintercalator binding to an intact DNA duplex. Strong NOE evidence confirmed a pseudocatenane structure in which **3** has two NDI rings intercalating between GpG steps, with the linker Gly$_3$Lys interacting with the walls of the major

(25) 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., Sect. D: Biol. Crystallogr.* **1998**, *D54*, 905–921.

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

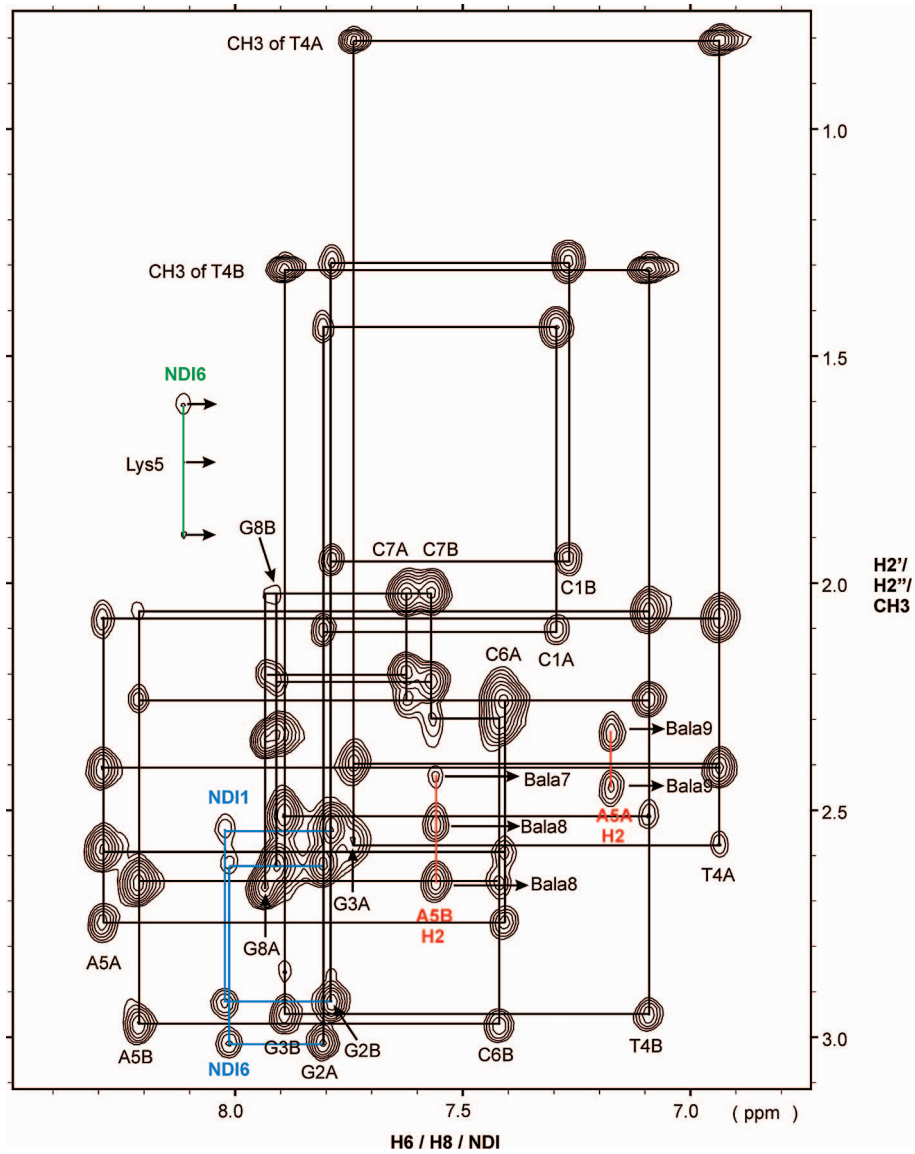


Figure 7. Contour plot of the NOESY spectrum (D_2O, 250 ms mixing time, 25 °C) of the **3**−d(CGGTACCG)$_2$ complex showing the DNA H6/H8 to H2′/H2′/CH$_3$ connectivities (black lines). Also shown are intramolecular NOEs between the NDI6 aromatic proton and Lys5 side chain protons (green line) and intermolecular NOEs between NDI aromatic protons and H2′/H2′ of G2 (blue lines), the ligand linker and H2 of the adenine residues (red lines).

groove and the linker (β-Ala)$_3$Lys residing in the minor groove, consistent with the initial design. Although not confirmed by any experiment directly, we believe it is reasonable to assume that binding of **3** to a 5′-GGTACC-3′ sequence in the context of a long DNA duplex strand (Figure 3) involves formation of the same pseudocatenane structure identified in the **3**−d(CGGTACCG)$_2$ complex.

The DNase I footprinting data for **3** revealed a strong preference for binding the 5′-GGTACC-3′ sequence with an apparent dissociation constant in the micromolar to submicromolar range ($K_D \sim 10^{-7}$ M). No binding was observed at any of the three 5′-GATAAG-3′ sites. The most reasonable explanation for this specificity is that Gly$_3$Lys is the dominant linker, perhaps due to the favorable contacts between the linker and the DNA major groove predicted by the computer simulations. The (β-Ala)$_3$Lys linker can accommodate binding in the minor groove of 5′-GGTACC-3′, although significant disorder was seen in the computer simulations. At this point it is not clear whether the observed preference is due to the overriding strength of the interaction between Gly$_3$Lys and 5′-GGTACC-3′ in the major groove, or the fact that this linker cannot be accommodated in the major groove of 5′-GGTACC-3′ (preferred by (β-Ala)$_3$Lys), or a combination of the two effects.

The detailed mechanism for even relatively simple monointercalators involves several distinct steps[27,28] and remains an area of active interest.[29] An intriguing aspect of the pseudocatenane structure described here is that in order for **3** to bisintercalate fully into a 5′-GGTACC-3′ sequence, it is possible that the disruption of at least the four internal Watson−Crick base pairs must occur. Because of this added complexity, there are undoubtedly numerous steps involved in the association/dissociation of **3** with DNA beyond those seen with monointercalators. In particular, one can imagine that the rate-limiting steps involved with pseudocatenane formation with **3** are tied to multiple base-pair breathing motions of the DNA duplex.

(27) Chaires, J. B.; Dattagupta, N.; Crothers, D. M. *Biochemistry* **1985**, *24*, 260–267.
(28) Rizzo, V.; Sacchi, N.; Menoozi, M. *Biochemistry* **1989**, *28*, 274–282.
(29) Mukherjee, A.; Lavery, R.; Bagchi, B.; Hynes, J. T. *J. Am. Chem. Soc.* **2008**, *130*, 9747–9755.

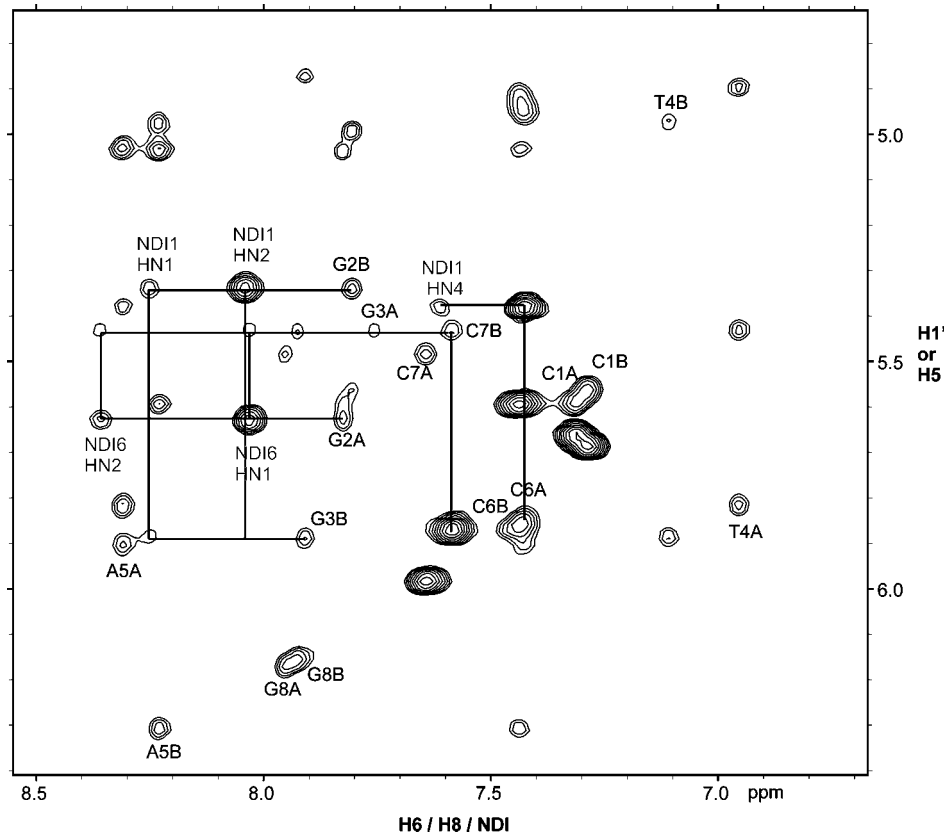


Figure 8. Intermolecular NOEs between protons from the two NDI rings and DNA d(CGGTACCG)$_2$ (H1′ and H5) at the intercalation sites (D_2O, 250 ms mixing time, 25 °C).

A related question concerns the mechanism by which **3** ends up at its preferred binding site in the context of long stretches of other sequences (Figure 3). Several potential pathways can be envisioned. For example, some type of initial "outside" complex can be formed that facilitates movement along the DNA duplex. In this scenario, **3** might associate loosely with one of the DNA grooves, presumably guided by electrostatic attraction, followed by relatively rapid movement along the groove. An analogous step has been proposed for monointercalator binding.[27]

At some point, pseudocatenane formation at a preferred sequence must occur. One can imagine at least two limiting scenarios, the first of which is a conceptually simple association−dissociation searching mechanism that is composed of a repeated series of pseudocatenane formation−dissociation events along the length of the DNA. Preferred sequences might be distinguished by either faster overall pseudocatenane formation rates (likely related to enhanced base pair breathing motions), or slower dissociation rates, or both. By virtue of repeated association−dissociation events, different regions of the DNA would be "sampled" with bound molecules in a stochastic fashion, eventually equilibrating to a preponderance of molecules at the highest affinity site(s) due to a longer overall residence time. Alternatively, one can imagine a scenario in which initial formation of a pseudocatenane occurs somewhere along the double helix, followed by the molecule "sliding", i.e. a series of base-pair disruptions as the molecule remains fully intercalated while moving up and down the sequence until a preferred binding site is reached. This latter pathway is conceptually analogous to the proposed "redistribution to preferred sites by internal transfer" proposed by Chaires and Crothers for monointercalators such as daunomycin.[27]

Importantly, the scenarios for DNA sequence scanning by **3** mentioned above, as well as others that can be envisioned, are not mutually exclusive and the true situation may turn out to be best described as an exceedingly complex hybrid of several different processes. Nevertheless, a detailed mechanistic study examining the **3** association−dissociation process is currently underway in our laboratory.

Materials and Methods

DNase I Footprinting. The 69 bp synthetic fragment (PAGE grade) was purchased from Midland Certified (Midland, TX) and labeled (^{32}P) as described.[30] The DNase I (Amersham Biosciences, Piscataway, NJ) footprinting was carried out according to the procedure described previously.[30] The DNA fragments were separated on a 12% denaturing polyacrylamide gel. The gels were exposed on a phosphor screen for 24−48 h, and the image was analyzed with Quantity One 4.5 software from Bio-Rad (Hercules, CA).

NMR Sample Preparation and NMR Spectroscopy. The DNA (gel filtration grade, Midland Certified, Midland, TX) was initially dissolved in 0.7 mL of 30 mM Na phosphate buffer, pH 7.5. Prior to addition of **3**, the d(CGGTACCG)$_2$ sample was diluted to 10 mL in chilled water to which the 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.

Spectra were obtained using a 500 MHz Varian Unity-Inova spectrometer. 2D NOESY (50, 80, and 250 ms mixing time), TOCSY (50 ms mixing time), and COSY spectra were acquired in

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

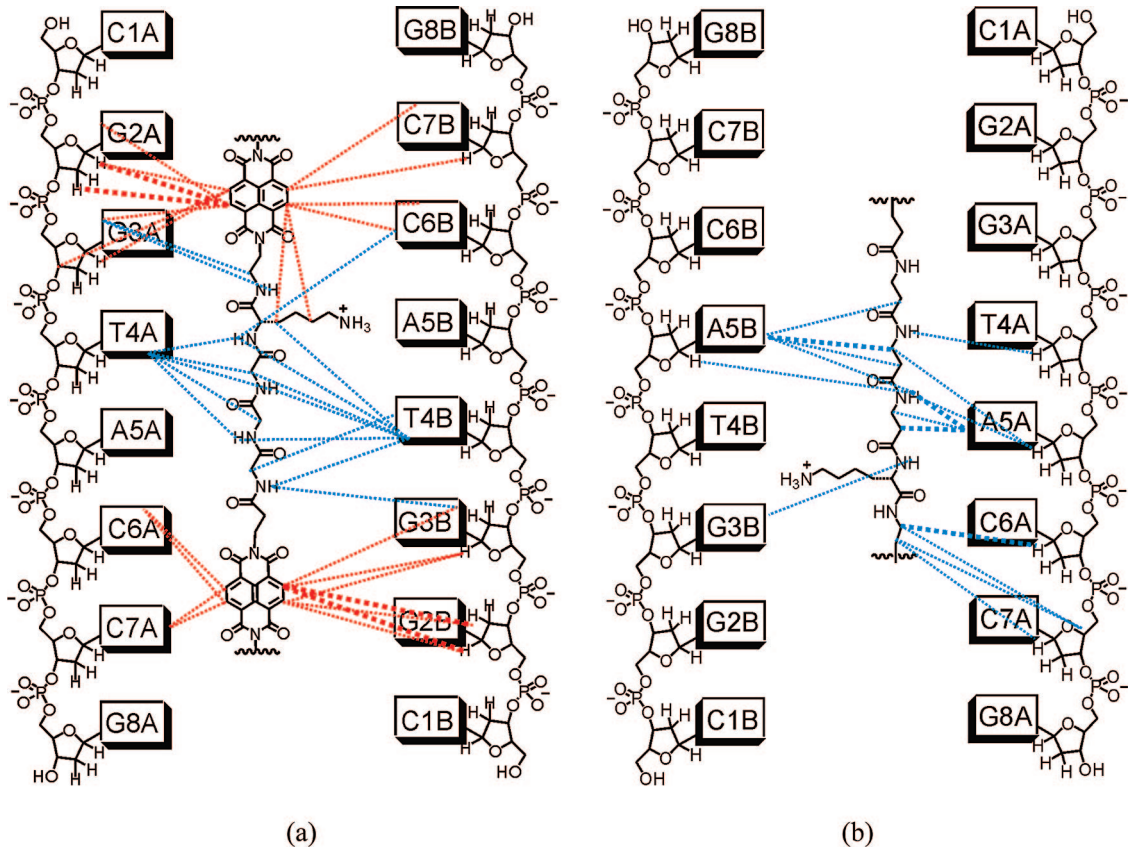


Figure 9. Diagram of the observed intermolecular NOEs in the **3**−d(CGGTACCG)$_2$ complex. The thick dashed lines indicate 'strong' and 'medium' NOEs; thin dotted lines represent weak NOEs. Blue lines indicate NOEs involving linker protons, and red lines indicate NOEs involving NDI protons. (a) NOEs observed in the intercalation sites and the major groove; (b) NOEs observed in the minor groove.

Table 1. Statistics for an Ensemble of Eleven Lowest Energy **3**−d(CGGTACCG)$_2$ Structures

Restraints for the Structural Calculation	
total number of distance restraints	268
DNA-DNA NOEs	131
DNA-ligand NOEs	99
hydrogen bondings[a]	38
Statistics for the Structural Calculation[b]	
number of NOE violations >0.5 Å	0.4
number of NOE violations >1.0 Å	0.0
rmsd to the mean structure (whole complex)	2.17 Å
rmsd for covalent bonds	0.0028 Å
rmsd for covalent angles	0.95°
rmsd for improper angles	0.37°

[a] These restraints were obtained from the dna-rna_restraints.def file in the CNS program. [b] Structural statistics are based on a superposition of 11 structures with nearly equal values of energy and representative of the full range of structures that are consistent with the NMR-derived distance and angle constraints.

D_2O 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, NOESY (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_2O:D_2O 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. NOESY spectra at ^{1}H spectra were referenced to internal 2,2-dimethylsilapentane-5-sulfonic acid (DSS) as recommended by Wishart et al.[31] A proton-detected ^{1}H−^{31}P correlated 2D spectrum[32] was acquired using sweep widths of 8000 and 5000 Hz and 2048 and 512 complex points for the ^{1}H and ^{31}P dimensions, respectively, and a total acquisition time of 36 h. A ^{31}P 1D NMR spectrum was acquired at phosphorus frequency of 202 MHz and a total acquisition time of 12 h. ^{31}P chemical shifts are relative to external aqueous trimethyl phosphate. The 1D spectra were processed using VNMR software (Varian). 2D spectra were processed using NmrPipe[33] and displayed using Sparky.[34]

Structural Modeling. Structure calculation of the **3**−d(CGGTACCG)$_2$ complex was carried out using the simulated annealing protocol within CNS program.[25] Distance restraints for nonexchangeable protons were derived from the observed cross-peak intensities in the D_2O NOESY spectrum (80 ms mixing time). NOEs were grouped as strong (0−3.5 Å), medium (0−4.0 Å), weak (0−5.0 Å), and very weak (0−5.5 Å). The distance restraints involving methylene protons were determined by adding an extra 0.7 Å to distances measured from the center of the methylene group. The cross-peak intensities in 80 ms NOESY spectrum using 9:1 H_2O:D_2O 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 the identified Watson−Crick base pairs. The DNA backbone dihedral angles were loosely restricted, depending on whether the nucleotides were near the intercalation sites and whether the regular NOE patterns typical of B-form DNA were observed. Dihedral angles for nucleotides 1,

(31) Wishart, D. S.; Bigam, C. G.; Yao, J.; Abildgaard, F.; Dyson, H. J.; Oldfield, E.; Markley, J. L.; Sykes, B. D. *J. Biomol. NMR* **1995**, *6*, 135–140.
(32) Sklenar, V.; Migoshiro, H.; Zon, G.; Bax, A. *FEBS Lett.* **1986**, *208*, 94–98.
(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.

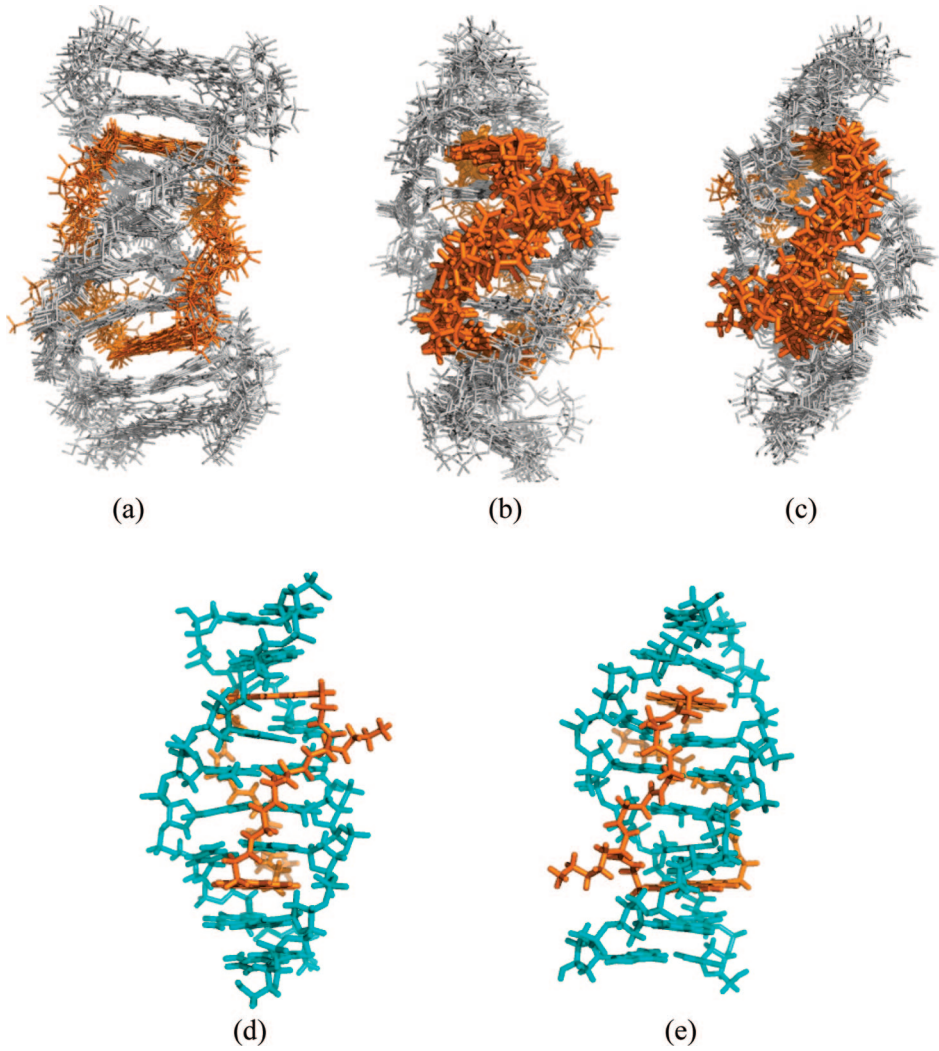


Figure 10. (a−c) Superposed structures of **3**−d(CGGTACCG)$_2$ complex of the final 11 ensembles with the lowest energy, with (a) showing the overall view, (b) view toward to the major groove, and (c) view toward the minor groove. (d and e) A representative calculated structure showing the linkers in the major groove and minor groove, respectively. The figures were created in PyMOL.[26]

4, 5, 8 A and B were restricted within a ±25° range of values typical of B-DNA while other nucleotides were restricted within a wider range ($\alpha = -46° \pm 40°$, $\beta = -147° \pm 40°$, $\gamma = 36° \pm 40°$, $\delta = 157° \pm 40°$, $\varepsilon = 155° \pm 40°$, $\zeta = -96° \pm 40°$). α, β, and ζ torsion angle restraints between nucleotides G2A and G3A and G2B and G3B were entirely removed because of the proximity to the intercalating ligand and the upfield-shifted ^{31}P resonances of G3A and G3B. The base pair planarity (±10°) was confined using torsion angle restraints. On the basis of qualitative analysis of deoxyribose conformations (Table S1, Supporting Information), each ribose ring puckering pattern was assigned as C2′- or C3′-endo accordingly. All torsion angles of the 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 **3** were created manually. The atomic charges and bond lengths of the NDI unit were derived from ab initio calculation. The bond linkage of NDI unit was based on the crystal structure of an NDI derivative.[35]

A very wide variety of starting structures were used for the simulated annealing. In the first stage a 60 ps of torsion-angle molecular dynamics was applied to those starting structures at 20 000 K to allow 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, 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 is near a consistent minimum value.[9,10]

Acknowledgment. This work was supported by the National Institutes of Health (GM-069647) and grants from the Welch Foundation (F-1188 and F-1353). We thank Dr. Steven Sorey for technical help during NMR spectra acquisition.

Supporting Information Available: Solid-phase synthesis detail and characterization data of **3** and **4**, NOESY, 1D ^{31}P spectrum and 2D ^{1}H−^{31}P correlation spectrum of **3**−d(CGGTACCG)$_2$ complex, ^{1}H chemical shifts of DNA duplex in the complex and DNA sugar conformation analysis data. This material is available free of charge via the Internet at http://pubs.acs.org.

(35) Lokey, R. S.; Iverson, B. L. *Nature (London)* **1995**, *375*, 303–305.

JA805676W