J|A|C|S
ARTICLES
Published on Web 10/13/2004

NMR Structural Analysis of a Modular Threading Tetraintercalator Bound to DNA

Jeeyeon Lee, Vladimir Guelev, Steven Sorey, David W. Hoffman, and Brent L. Iverson*

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

Received June 21, 2004; E-mail: biverson@mail.utexas.edu

Abstract: The synthesis and NMR structural studies are reported for a modular threading tetraintercalator bound to DNA. The tetraintercalator design is based on 1,4,5,8-tetracarboxylic naphthalene diimide units connected through flexible peptide linkers. Aided by an overall C_2 symmetry, NMR analysis verified a threading polyintercalation mode of binding, with linkers alternating in the order minor groove, major groove, minor groove, analogous to how a snake might climb a ladder. This study represents the first NMR analysis of a threading tetraintercalator and, as such, structurally characterizes a new topology for molecules that bind to relatively long DNA sequences with extensive access to both DNA grooves.

Introduction

The study of synthetic molecules that specifically bind to DNA is a challenging area of molecular recognition research. Such molecules have the potential to elucidate new mechanisms underlying DNA recognition and ultimately modulate gene expression in vitro and in vivo. An important challenge is to develop new molecular scaffolds with which to explore sequence-specific DNA recognition. Particularly successful and well-studied examples include the polyamide minor groove-binding derivatives of netropsin and distamycin derivatives[1−4] and the various nucleic acid-based approaches[5,6] to triple-strand formation. For example, a modular approach to long sequence recognition was reported for a polyamide that demonstrated the ability to recognize, specifically, 16 base pairs of DNA at subnanomolar concentrations.[7] For the polyamides, sequence-specific recognition of relatively long sequences has been achieved in a programmable fashion by establishing a direct readout between DNA recognition elements and complementary functional groups.

Intercalation is a well-studied mode of small molecule binding to DNA; however, thus far, polyintercalation has generally not proven to be an important strategy for binding long stretches of DNA with sequence specificity and high affinity. A notable exception is the report of a relatively rigid carbohydrate linker that was designed to fit between two intercalating units and generated a bisintercalator that bound to the expected sequence with very high affinity.[8]

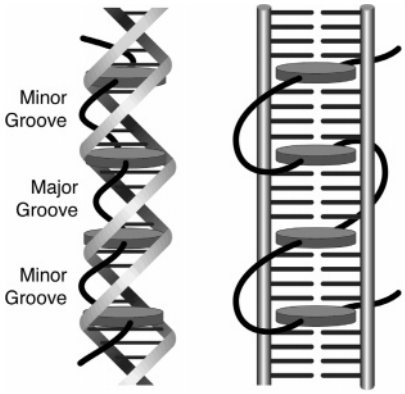

Figure 1. Threading tetraintercalator that binds a 14 base pair sequence of DNA.

We have developed a novel class of DNA-binding polyintercalators in which 1,4,5,8-tetracarboxylic naphthalene diimide (NDI) units are connected with flexible peptide linkers.[9] NDI is classified as a threading intercalator, in that the attachment of functional groups to both diimide nitrogen atoms creates a situation in which one functional group will reside in the major groove and the other in the minor groove[9−11] upon intercalation. A reasonable assumption is that DNA breathing must occur to allow for threading intercalation, implying characteristically slow association and dissociation rates[9] for threading intercalators.

A threading polyinteracalation design (Figure 1) has a number of potential advantages, including (1) interactions with functional groups in both grooves, providing greater opportunity for sequence recognition; (2) possible disruption of protein−DNA

(1) Dervan, P. B. *Science* **1986**, *232*, 464−471.
(2) White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *Nature (London)* **1998**, *391*, 468−471.
(3) Singh, M. P.; Lown, J. W. *Prog. Med. Chem.* **1996**, *1*, 149−171.
(4) Reddy, P. M.; Bruice, T. C. *J. Am. Chem. Soc.* **2004**, *126*, 3736−3747.
(5) Moser, H. E.; Dervan, P. B. *Science* **1987**, *238*, 645−650.
(6) Gowers, D. M.; Fox, K. R. *Nucleic Acids Res.* **1999**, *27*, 1569−1577.
(7) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 3534−3535.
(8) Chaires, J. B.; Leng, F.; Przewloka, T.; Fokt, I.; Ling, Y. H.; Pérez-Soler, R.; Priebe, W. *J. Med. Chem.* **1997**, *40*, 261−266.
(9) 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) Yen, S.-F.; Gabby, E. J.; Wilson, D. *Biochemistry* **1982**, *21*, 2070−2076.
(11) Tanious, F. A.; Yen, S.-F.; Wilson, D. *Biochemistry* **1991**, *30*, 1813−1819.

10.1021/ja046335o CCC: $27.50

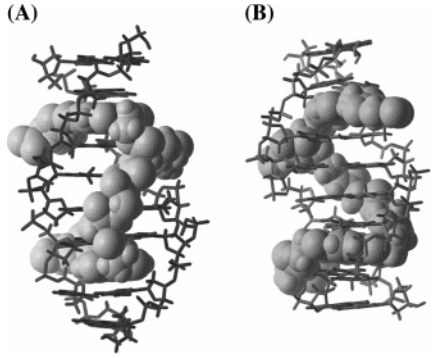

Figure 2. Views from the major groove of (A) the **1**−d(CGGTACCG)$_2$ complex[15] and (B) the **2**−d(CGATAAGC)·d(GCTTATCG) complex.[16]

interactions in both grooves; (3) potentially very slow off rates, considering the large molecular rearrangements required for dissociation; and (4) recognition of relatively long DNA sequences with a relatively low molecular weight molecule (assuming an extended conformation when bound). Takenaka et al.[12] and Lamberson[13] have also reported threading polyintercalator designs, although no structural studies have appeared for these systems.

Library studies, using peptide-linked NDI dimers, identified linkers with different DNA sequence preferences.[14] Two dimers bound to their preferred sequences[15,16] were analyzed by NMR (Figure 2). These studies generated models of bound complexes that were consistent with the expected threading mode of intercalation. Importantly, the dimer with a -Gly$_3$-Lys- linker (compound **1**, Figure 3) was found to have the linker in the major groove,[15] while the dimer with a -βAla$_3$-Lys- linker (compound **2**, Figure 3) was found to have the linker in the minor groove.[16] In both cases, the linkers were observed to reside in an entirely extended conformation, spanning four base pairs between intercalation sites. Herein, the synthesis and characterization by NMR of a modular threading tetraintercalator **3**, designed to have linkers that alternate grooves based on similarities to compounds **1** and **2**, are reported.

Design. For the design of tetramer **3**, the -βAla$_3$-Lys- linker of **2** was used as a minor groove-binding element, and a novel major groove-binding linker, namely, adipic acid between Lys residues, was chosen based on the key elements of the -Gly$_3$-Lys- linker from compound **1** studied previously.[15] The -βAla$_3$-Lys- linker of **2** places a total of 10 methylene units in the minor groove and is the proper length to span four base pairs when fully extended (Figure 2).[16] Thus, it can be thought of as a classical, relatively hydrophobic minor groove-binding element,[17] analogous to the role proposed by Dervan and coworkers for hairpin polyamides containing a β-Ala unit.[18] On the basis of footprinting and the DNA alkylation pattern of an *N*-bromoacetyl derivative, **2** was shown to have a preference for binding d(CGATAAGC)·d(GCTTATCG). Specific hydrogen bonds were proposed between the linker amide NH groups closest to both NDI rings and the T residue O2 atoms of the DNA.[16]

Adipic acid between Lys residues was chosen as the major groove binding linker in analogy to the -Gly$_3$-Lys- linker from compound **1** that was studied previously.[15] Footprinting analysis for **1** revealed a preference for binding d(GGTACC)$_2$. Note that this linker is three atoms shorter than the minor groove binding -βAla$_3$-Lys- linker, consistent with -Gly$_3$-Lys- spanning four base pairs through a shorter, diagonal orientation in the wider major groove. The minor groove is too narrow to allow a similar diagonal linker orientation. For the -Gly$_3$-Lys- linker, specific hydrogen bonds were proposed between the amide NH groups closest to each NDI ring and two G residue O6 atoms.[15] Steric complementarity between the linker and the floor of the major groove in the central TA region of d(GGTACC)$_2$ was thought to contribute to specificity, as well. Computer modeling indicated that the proposed hydrogen bonds to G residue O6 atoms and overall linker length would be retained with the adipic acid linker, along with a close steric fit along the bottom of the major groove.

The adipic acid linker greatly facilitated synthesis of the desired tetramer on the solid-phase resin (vida infra). In addition, an adipic acid-based linker has C_2 symmetry, like the preferred d(GGTACC)$_2$ DNA binding site. Because of the adipic acid linker, the entire tetramer **3** has C_2 symmetry, simplifying, by one-half, the number of signals that must be assigned in an NMR analysis of tetramer **3** bound to a similarly C_2 symmetric DNA binding site.

The predicted DNA binding site for tetramer **3** is the self-complementary, 14 base pair sequence 5′-G|ATAA|GTAC|TTAT|C-3′, duplex **(4)$_2$** (Figure 4), a hybrid of the d(CGATAAGC)·d(GCTTATCG) and d(GGTACC)$_2$ sites, which are preferred by **2** and **1**, respectively. The linkers of tetramer **3** are expected to be bound in the order minor groove, major groove, minor groove. Tetramer **3** and the DNA binding site share many common elements with the previous NMR dimer studies,[15,16] facilitating initial peak assignments.

Synthesis. Synthesis of tetramer **3** was carried out on solid phase (Figure 5). Using an Fmoc-based strategy and Rink amide resin (Novabiochem), an NH$_2$-Lys-NDI-βAla$_3$-Lys-NDI-Gly-NH-resin dimer was synthesized as described previously.[19] Prolonged incubation with limited adipic acid under peptide-coupling conditions resulted in the desired tetramer structure. Because the synthesis was based on assembly of two dimers, purification of tetramer **3** was relatively straightforward using preparative HPLC, resulting in analytically pure material in 17% overall yield based on resin loading.

Titration of Duplex (4)$_2$ with Tetramer 3. Titration of the d(GATAAGTACTTATC)$_2$ DNA duplex, **(4)$_2$**, with tetramer **3** was monitored using 1-D ^{1}H NMR spectroscopy (Figure 6) in 30 mM sodium phosphate buffer (pH 7.5). After adding 1 equiv of ligand, the imino proton resonances of the free DNA disappeared, and new resonances from the complex appeared as a separate set of upfield-shifted signals. The number of observed resonances in the spectrum of the complex did not change compared to those of the uncomplexed **(4)$_2$**, which is

(12) Takenaka, S.; Nishira, S.; Tahara, K.; Kondo, H.; Takagi, M. *Supramol. Chem.* **1993**, *2*, 41−46.
(13) Lamberson, C. Ph.D. Dissertation, University of Illinois, 1991.
(14) Guelev, V. M.; Harting, M. T.; Lokey, R. S.; Iverson, B. L. *Chem. Biol.* **2000**, *7*, 1−8.
(15) Guelev, V.; Lee, J.; Ward, J.; Sorey, S.; Hoffman, D. W.; Iverson, B. L. *Chem. Biol.* **2001**, *8*, 415−425.
(16) Guelev, V.; Sorey, S.; Hoffman, D. W.; Iverson, B. L. *J. Am. Chem. Soc.* **2002**, *124*, 2864−2865.
(17) Bailly, C.; Chaires, J. B. *Bioconjugate Chem.* **1998**, *9*, 513−538.
(18) Wemmer, D. E. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 439−461.
(19) Guelev, V. M.; Cubberley, M. S.; Murr, M. M.; Lokey, R. S.; Iverson, B. L. *Methods Enzymol.* **2001**, *340*, 556−570.

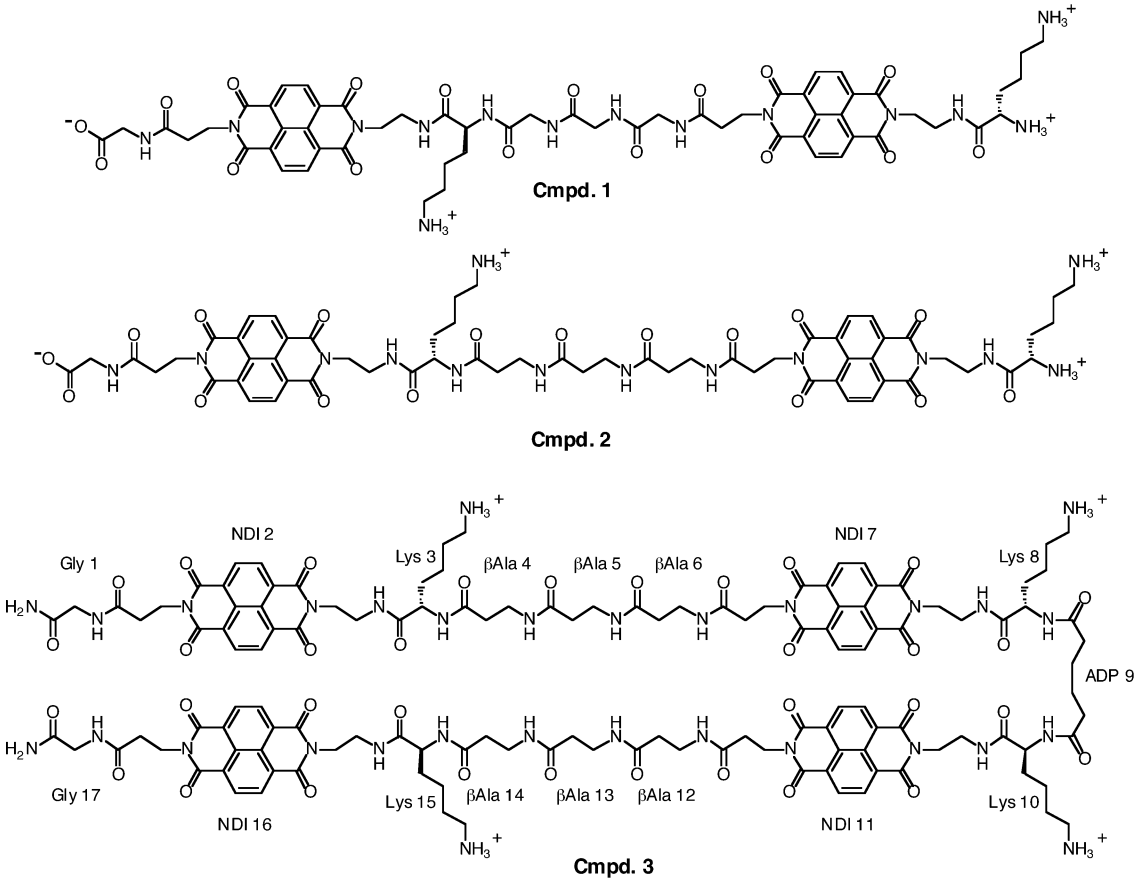

Figure 3. Structures of compounds **1**−**3**. Abbreviations used for the residues of **3** are shown.

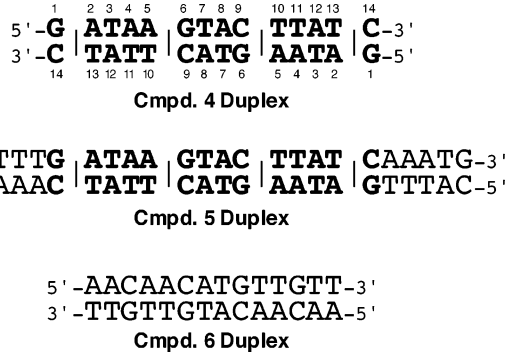

Figure 4. Sequences of the self-complementary oligonucleotides **4**−**6** used in these studies. Abbreviations used for the oligonucleotide $(\mathbf{4})_2$ duplex are shown.

consistent with the expectation that C_2 symmetry is retained upon formation of the $\mathbf{3}-(\mathbf{4})_2$ complex.

Specificity of the interaction was indicated by two control experiments. In the first, the 14-mer binding sequence was imbedded in a longer 24-mer oligonucleotide (compound **5**) with the sequence d(CATTT**GATAAGTACTTATC**AAATG)$_2$. A similar 1-D titration showed peak shifts of both tetramer **3** and DNA peaks that were identical to those seen with the $\mathbf{3}-(\mathbf{4})_2$ complex, verifying binding at the same site. Fortuitous chemical shift overlap within the flanking sequences precluded an unambiguous assignment of the entire complex; so, it is not presented here. In a second control, a presumed nonbinding 14-mer oligonucleotide (compound **6**) with the sequence d(AACAACATGTTGTT)$_2$ was used. As expected, no discrete complex was seen with $(\mathbf{6})_2$ when 1 equiv of **3** was added, as evidenced by only peak broadening, with the emergence of many broad peaks too numerous to represent a single bound species.

NMR Assignments and Sites of Intercalation. Each signal of the free DNA and complex with **3** was assigned using standard 2-D homonuclear NMR methods.[20] The aromatic protons in the NDI groups of **3** were well-resolved. The NOE connectivities between sequential DNA nucleotides should be interrupted or weak for all steps where the NDI units are intercalated, and such interruptions are seen between G1/C14: A2/T13 and A5/T10:G6/C9. The sites of NDI intercalation were further indicated by a loss of fast exchange of the amino protons on A2, A5, and G6 and by characteristic upfield shifts of 1.45, 1.1, and 1.01 ppm for the imino protons of G6, T10, and T13, respectively, upon binding **3**. The imino proton of G1 could not be detected with or without bound **3**, presumably due to its rapid exchange with solvent.

The expected NOE cross-peaks between the protons of the NDI groups and those of the DNA were observed at the intercalating sites (Figure 7). Most of the NDI DNA intermolecular NOEs involved the central bases (A5, G6, C9, and T10). Fewer intermolecular NOE cross-peaks were observed for nucleotides A2, T13, and C14, presumably due to fraying of the terminal bases. No NOEs could be confirmed between NDI and G1, likely for the same reason.

Strong NOE cross-peaks were observed between the methylene protons of each β-Ala residue of **3** and the H2 protons of

(20) Hare, D. R.; Wemmer, D. E.; Chou, S.; Drobny, G. *J. Mol. Biol.* **1983**, *171*, 319−336.

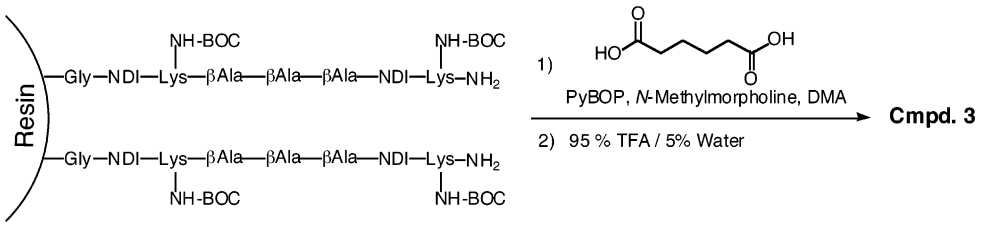

Figure 5. Solid-phase synthetic scheme used to construct compound **3**. The NDI dimer molecule was constructed first and then linked into a tetramer by reaction with limiting adipic acid.

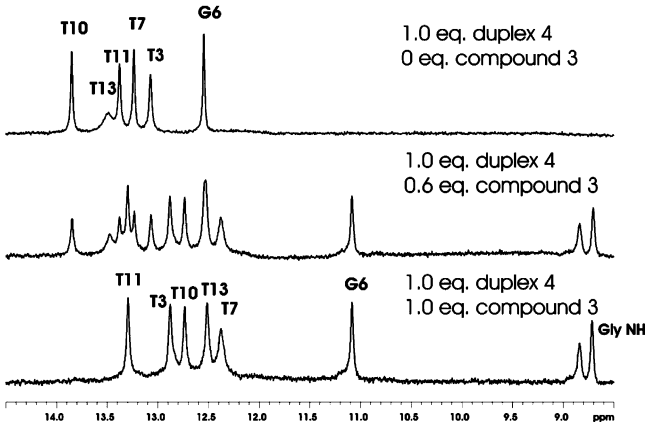

Figure 6. One-dimensional ^{1}H NMR spectra of tetramer **3** titration into d(GATAAGTACTTATC)$_2$ (**4**) in H_2O with 30 mM sodium phosphate buffer (pH 7.5, 27 °C).

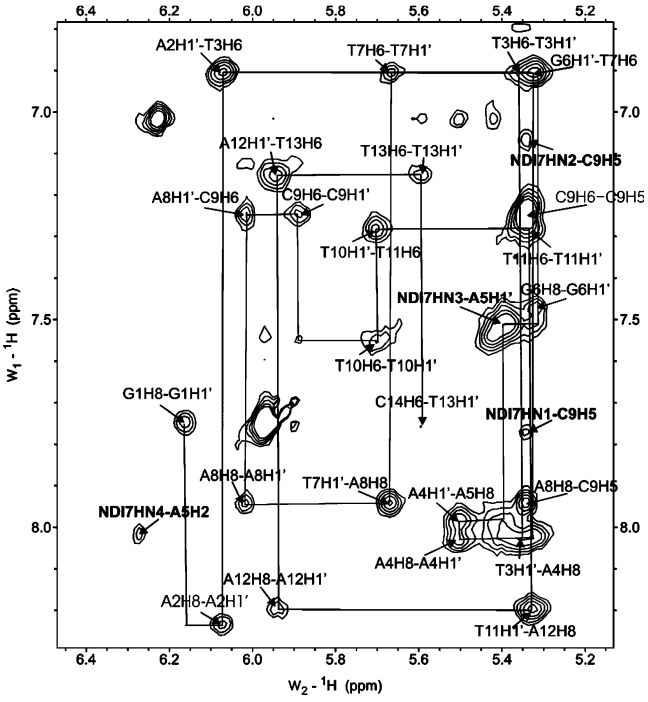

Figure 7. Expansion of the NOESY spectrum (mixing time 60 ms, pH 7.5, 27 °C) of the **3**−**(4)$_2$** complex. Sequential aromatics to H1′ connectivities are shown as lines.

nucleotides A4, A5, and A12, as well as between the imino protons of T3 and T11 (Figure 8). These NOE cross-peaks are only consistent with the β-Ala residues being located in the minor groove of the DNA duplex. In contrast, the methylene protons of the adipic acid group of **3** exhibited NOE cross-peaks with the methyl protons of T7, consistent with being located in the major groove.

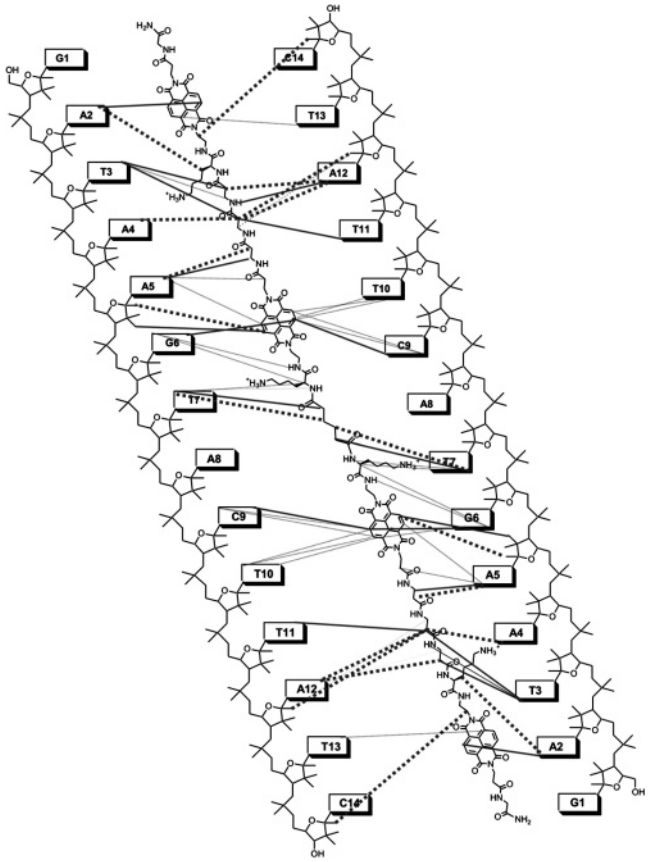

Figure 8. Diagram of the observed NOEs. The strong NOEs are shown as thick dotted lines; the medium NOEs are shown as solid lines, and the weak NOEs are shown as thin dotted lines.

DNA Sugar Conformation Analysis. NOE data for the **3**−**(4)$_2$** complex revealed that the majority of nucleotides are in an S-type (C2′-endo sugar pucker) conformation.[21] However, nucleotides on the 5′ side of the intercalating sites (G1, A5, C9, and T13) are in an N-type (C3′-endo) conformation, presumably induced by NDI intercalation. This conclusion was also supported by the strong intranucleotide H2″-H4′ NOE cross-peaks between H6/H8 and H2′/H2″ inside these same four nucleotides, as well as those to the residue on the 5′ side.[22] C3′-endo sugar pucker in the nucleotides on the 5′ side of an intercalation site has been observed in other DNA−intercalator complexes.[23] The preponderance of S-type C2′-endo pucker of ribose rings points toward an overall B-form DNA helix, except

(21) 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.
(22) Wang, A. H.; Robinson, H. In *Nucleic Acid Targeted Drug Design*; Propst, C. L., Perun, T. J., Eds.; Marcel Dekker: New York, 1992; pp 17−64.
(23) Saenger, W. In *Principles of Nucleic Acid Structure*; Cantor, C. R., Ed.; Springer-Verlag: New York, 1984.

Table 1. Structural Statistics for the Final 12 Ensembles of the **3**−**(4)$_2$** Complex

Restraints for Structural Calculation	
total number of distance restraints	341
DNA−DNA NOEs	192
DNA−ligand NOEs	76
hydrogen bonding[a]	58
long-range distances[b]	15
Statistics for Structural Calculation	
number of NOE violations >0.2 Å	3.58
number of NOE violations >0.5 Å	0
number of torsion angle violations >5°	1.08
number of torsion angle violations >6.5°	0
averaged rmsd to mean structure for backbone	1.79 Å
averaged rmsd to mean structure for non-H atoms	1.62 Å
rmsd to the averaged structure (all)	1.80 Å
rmsd to the averaged structure (bases 2−5 and 10−13)	1.06 Å
rmsd for covalent bonds	0.003 Å
rmsd for covalent angles	2.19°
rmsd for NOE violations	0.45 Å

[a] To clearly define Watson−Crick base pairs, hydrogen-bonding restraints based on the distances of two atoms involved in base pairing were used. These restraints were obtained from the dna-rna_restraints.def file in the CNS program. [b] To define the overall twist of the DNA helix, these long-range distances were determined by averaging the analogous distances in several PDB files of mono/bisintercalators−DNA complexes.[25−27] The values of the CNS energy function, averaged for a set of 12 low-energy structures, are $E_{total} = 899$, $E_{bond} = 16.7$, $E_{angle} = 1751$, $E_{improper} = 19.4$, $E_{vdw} = -434$, $E_{dih} = 1.1$, $E_{elect} = -536$, $E_{noe} = 35.0$, and $E_{cdih} = 18.8$.

for those nucleotides on the 5′ side of the intercalation sites (G1, A5, C9, and T13).

Description of the 3−(4)$_2$ Complex. The **3**−**(4)$_2$** complex exhibits well-resolved NMR signals of the appropriate number, most likely indicating one predominant species that is not exchanging with other species slower than, or on, the NMR time scale. Although faster exchange between multiple species cannot be rigorously ruled out, the unique threading mode of polyintercalation would require substantial molecular rearrangements of both **(4)$_2$** and **3** to allow conversion to alternative binding modes, an unlikely situation given the NMR time scale boundary. Additionally, the similarity in the upfield-shifted signals in the NMR spectra observed between the **3**−**(4)$_2$** and **3**−**(5)$_2$** complexes, combined with the clear lack of specific binding in the case of **3** and **(6)$_2$**, supports the notion that a predominant bound species is being observed.

Structure calculations were performed using the simulated annealing protocol within the CNS program suite[24] (Table 1). Overall, the NMR structural analysis of the **3**−**(4)$_2$** structure is most consistent with the predicted threading mode of polyintercalation (Figure 9). As exhibited by all 12 of the calculated low-energy structures, overwhelming evidence indicates NDI intercalation at G1/C14:A2/T13 and A5/T10:G6/C9, meaning that the linkers span four base pairs between intercalation sites. The DNA is in an overall right-handed B-form helix conformation (C2′-endo sugar pucker), except that the bases on the 5′ side of the intercalation site give signals that are consistent with that of the C3′-endo sugar pucker. NOEs indicating close contacts were identified between the -β-Ala- linker methylene groups and adenosine H2 protons deep in the minor groove, consistent with the NMR studies of **2**,[16] as well as with the previous work, indicating β-Ala as a minor groove binding element.[17,18] NOEs are also seen between the adipic acid linker and the methyl groups of T7 in the major groove. Taken together, these key NOEs, as well as many others, confirm threading polyintercalation, with linkers alternating in the order minor groove, major groove, minor groove, consistent with the original modular design.

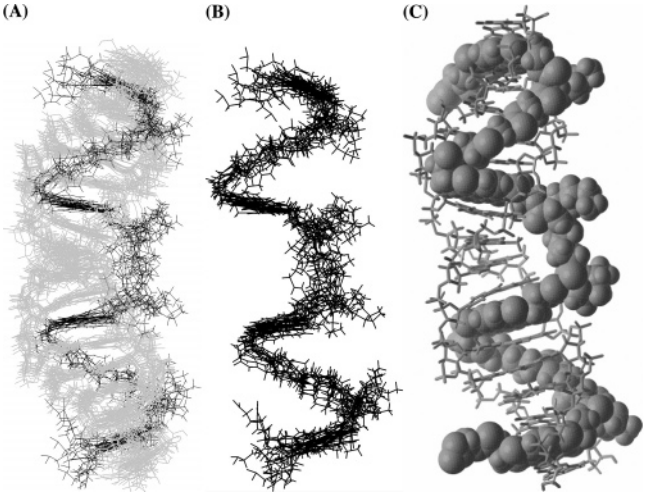

Figure 9. Superposed structures of the final 12 ensembles with the lowest energy (A and B) and space-filling model (C) of the tetramer **3**−DNA **(4)$_2$** complex.

Various amounts of DNA bending are observed in the central T7-A8 region of the calculated low-energy structures (Figure 9). Energies involved in modest amounts of DNA bending such as this are relatively small.[25] NMR studies using NOE distances have difficulty gauging global features of elongated DNA (such as bending) due to the lack of long-range distance information; so, at this point, the relative importance of possible DNA bending remains uncertain. Interestingly, there is an unusually high upfield shift of 0.85 ppm for the imino proton chemical shift of residue T7 upon binding to **3**, perhaps indicating the presence of altered base stacking in this region, as might occur in the presence of some helix bending.

Hydrogen Bonding. Analysis of possible specific hydrogen bonds in the ensemble of 12 calculated lowest-energy structures suggests some consensus hydrogen bonding between **3** and the DNA, as well as some hydrogen bonds present in only a subset of structures. For this analysis, hydrogen bonds were assumed when heavy atom distances were between 2.5 and 3.0 Å. The amide NH protons between the terminal NDI groups (2/16) and Lys 3/15 have a consensus hydrogen bond with the O2 atom of T13, and amide NH protons on the amide bond between β-Ala 6/12 and the internal NDI groups (7/11) have a consensus hydrogen bond with the O2 atom of T10 (Figure 10). These proposed hydrogen bonds are consistent with those postulated in the case of the compound **2** DNA NMR analysis,[16] and it is likely that they play a significant role in the specific binding of **3** to the DNA.

Several other hydrogen bonds between NH protons of the -βAla$_3$-Lys- linker and the DNA appear in at least three of the calculated structures, although they never appear simultaneously in the same structure. In particular, hydrogen bonds appear between the NH of β-Ala 5/13 and the O2 position of T11, the NH of β-Ala 4/14 and the N3 position of A4, and the NH of

(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.

(25) Hartmann, B.; Lavery, R. *Q. Rev. Biophys.* **1996**, *29*, 309−368.

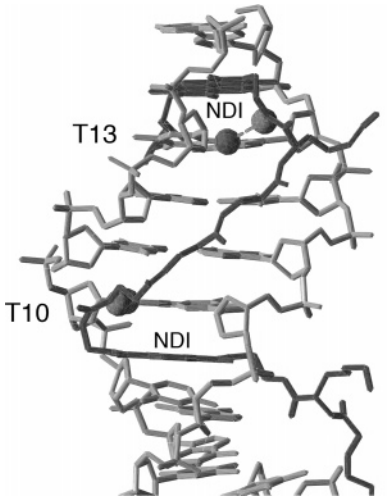

Figure 10. Representative structure emphasizing the consensus hydrogren bonds seen in the minor groove of the **3**−**(4)$_2$** complex. Locations of the hydrogen-bonded atoms are indicated by the large spheres. The NDI residues and relevant bases (T10 and T13, respectively) are labeled.

Lys 3/15 and the O4 position of T13. Surprisingly, the amide groups adjacent to the NDI units in the major groove do not appear to be taking part in significant hydrogen bonding according to the modeling based on NMR constraints, even though such hydrogen bonds were assumed in the original design.

Materials and Methods

Materials. Rink amide resin (0.72 mmol/g), benzotriazol-1-yl-oxytrispyrrolidinophosphonium hexafluorophosphate (PyBOP), and all of the Fmoc-protected amino acids were purchased from Novabiochem. Adipic acid, *N*,*N*-dimethylacetamide (DMA), 2-propanol, and *N*-methylmorpholine (NMP) were purchased from Aldrich. All DNA oligonucleotides (5′-GATAAGTACTTATC-3′, 5′-CATTTGATAAG-TACTTATCTTTATCAAATG-3′, and 5′-AACAACATGTTGTT-3′) were purchased from Midland Certified (gel filtration grade) and used for NMR without further purification.

Synthesis. Fmoc-lysine-NDI-(β-alanine)$_3$-lysine-NDI-glycine-resin was prepared as previously described.[19] After the mixture was treated with 20% piperidine and washed, an activated ester solution containing 0.3 equiv of adipic acid, 0.6 equiv of PyBOP, and 0.6 equiv of NMP dissolved in DMA was added to the resin, and the mixture was shaken for 1.5 h. One more portion of the same activated ester solution was then added to the resin, and the mixture was shaken for an additional 4 h to complete the cross-linking reaction. After a Kaiser test[26] showed a negative result for free amine, the resin was washed with methanol, followed by dichloromethane, and then dried in vacuo. TFA (95%) and water (5%) were added to the resin, and the mixture was shaken for 2 h. The mixture was then filtered, and the filtrate was precipitated with ether. The precipitate was washed with ether three times and dried in vacuo. The crude product was dissolved in 0.1% aqueous TFA and purified using reverse-phase preparative HPLC in 0.1% TFA in a water/0.1% TFA in an acetonitrile system. The collected pure fractions of the major peak were lyophilized to yield a pale yellow powder. Overall isolated yield of **3** was 17% based on resin loading data provided by the manufacturer: ^{1}H NMR (500 MHz, D_2O) δ 8.13−8.30 (m, 16H), 4.22 (t, 4H, J = 7 Hz), 4.14−4.09 (m, 8H), 4.07−4.04 (m, 4H), 4.03−3.99 (m, 2H), 3.90 (dd, 2H, J = 9.0, 5.0 Hz), 3.80 (s, 4H), 3.51−3.42 (m, 8H), 3.40−3.30 (m, 8H), 3.25−3.14 (m, 4H), 2.87 (t, 4H, J = 7.5 Hz), 2.83 (t, 4H, J = 7.5 Hz), 2.64 (t, 4H, J = 7.0 Hz), 2.52 (t, 4H, J = 7.0 Hz), 2.36−2.33 (m, 8H), 2.27 (t, 4H, J = 7.0 Hz), 2.07−1.98 (m, 4H), 1.63−1.39 (m, 16H), 1.32−1.16 (m, 12H); ESI-MS found, 1325.9 (MH^{2+}/2); predicted, 1325.5.

NMR Sample Preparation and NMR Spectroscopy. The DNA sample was prepared for NMR spectroscopy by dissolving the sample in 30 mM sodium phosphate (pH 7.5). The DNA concentration in a 2-D ^{1}H NMR experiment was typically 1 mM, as determined by UV absorption at 260 nm (ϵ = 143 400 M^{-1} cm^{-1}). For **3**-DNA titration experiments, the DNA was diluted with 5 mL of water, and then a ligand stock solution (quantified by UV absorption, ϵ = 51 300 M^{-1} cm^{-1})[9] in water was added slowly to the DNA. The **3**−**(4)$_2$** complex was lyophilized from H_2O and dissolved in 650 μL of a 90% H_2O/10% D_2O solution. This titration and lyophilization process was repeated to obtain various ratios of **3**:**(4)$_2$**. For spectra in D_2O, the samples were lyophilized twice from D_2O and finally dissolved in 650 μL of 99.99% D_2O.

NMR experiments were performed using a Varian INOVA 500 MHz spectrometer. One-dimensional spectra in 90% H_2O/10% D_2O solvent were obtained at 27 °C using a jump-return solvent-suppression method. Two-dimensional NOESY, TOCSY, and DQF-COSY spectra were obtained at 5 or 27 °C in 90% H_2O/10% D_2O and 100% D_2O using presaturation for solvent suppression. NOESY spectra were acquired with a mixing time of 60 and 200 ms, and TOCSY spectra were determined with a mixing time of 50 ms. All spectra were recorded with 512 (t1) and 2048 (t2) complex points. The spectra were processed using VNMR (Varian, Inc.).

Structure Determination. Structure calculations were performed using the simulated annealing protocol within the CNS program suite,[24] with the aim of determining the full range of possible structures for the DNA−ligand complex that are consistent with the NMR-derived restraints, while having reasonable molecular geometries. Distance restraints were derived from the observed cross-peak intensities in the NOESY spectrum obtained in the D_2O solvent with a mixing time of 60 ms. Distance restraints were grouped as 1.8−3.5 Å, 1.8−4.0 Å, 2.5−5.0 Å, and 2.5−5.5 Å for strong, medium, weak, and very weak NOEs, respectively. For NOEs involving methyl protons, distances were measured from the center of the methyl group, and 1 Å was added to the interproton distance restraint. For NOEs involving methylene protons, distances were measured from the center of the methylene group, and 0.7 Å was added to the interproton distance restraint. Distance restraints for exchangeable protons were obtained from cross-peaks observed in a 200 ms NOESY spectrum in a 90% H_2O/10% D_2O solvent. Hydrogen bonds for the clearly identified Watson−Crick base pairs were defined using the dna-rna_restraints.def file in the CNS program. Torsion angle restraints were used to fix the base pairs within ±10° of planarity. Dihedral angles along the DNA backbone were restrained to be within a 40° range of the values typical for B-form DNA (a = −46 ± 40°, b = −147 ± 40°, g = 36 ± 40°, d = 157 ± 40°, e = 155 ± 40°, z = −96 ± 40°) for the base pairs adjacent to the intercalation sites and a ±25° range for other base pairs. Ribose rings were assigned an S- or N-type conformation based on a qualitative analysis of intraribose NOE intensities, and torsion angles were restrained to within a ±25° range of the value typical of each conformation. Since the short-range NMR-derived distances were insufficient for defining the overall twist of the DNA helix, 15 additional long-range restraints were included to define the diagonal distances between the terminal residues of opposite strands; these long-range distances were determined by averaging the analogous distances in several PDB files of mono/bisintercalators−DNA complexes.[27−30] These

(26) Stewart, J. M.; Young, J. D. *Solid-Phase Peptide Synthesis*; Pierce Chemical Company: Rockford, Illinois, 1984.

(27) Robinson, H.; Priebe, W.; Chaires, J. B.; Wang, A. H. *Biochemistry* **1997**, *36*, 8663−8670.
(28) Lisgarten, J. N.; Coll, M.; Portugal, J.; Wright, C. W.; Aymami, J. *Nat. Struct. Biol.* **2002**, *9*, 57−60.
(29) Adams, A.; Guss, J. M.; Collyer, C. A.; Denny, W.; Wakelin, L. P. G. *Biochemistry* **1999**, *38*, 9221−9233.
(30) Wang, A. H.; Robinson, H. In *Nucleic Acid Targeted Drug Design*; Propst, C. L., Perun, T. J., Eds.; Marcel Dekker: New York, 1992.

15 long-range restraints are in the form of distances; for example, the phosphate of nucleotide 2 on DNA strand 1 was restrained to be 54.0 ± 10.0 Å from the phosphate of nucleotide 2 on strand 2 of the DNA.

The simulated annealing protocol within CNS was used to find the structures of the **3**−DNA complex that were consistent with the NOE-derived restraints, while having reasonable molecular geometries. Force-field parameters for the DNA are described by the CNS parallhdg.dna parameter file. Parameter and topology files for the heteroresidues within **3** (NDI, ADP, and β-Ala) that are not in the CNS protein library were created with the assistance of Gerard Kleywegt's XPLO2D server (version 011023/2.9.3) and then were manually corrected with values based on those used in CNS for protein residues.

Starting structures included models in which the DNA was unduplexed; the ligand was separated by 20 Å from the DNA, and **3** was partially threaded or fully threaded in the DNA. A square-well function with a soft asymptote was used to describe the NOE-derived distance restraints. In the first stage of the simulated annealing, the structures were subjected to 60 ps of torsion-angle molecular dynamics at 20 000 K to allow a high degree of randomization of the initial model; then, the temperature was slowly reduced to 2000 K over a period of 60 ps and was followed by the second cooling stage in which the temperature was reduced to 300 K for 15 ps using Cartesian molecular dynamics. The structures were finally subjected to 2000 steps of conjugate-gradient minimization. The simulated annealing was repeated until the total energy was at or very near a consistent minimum value.

Conclusions

An NMR analysis has verified for the first time threading polyintercalation for a tetrameric intercalator bound to duplex DNA. As such, this work represents the first structural characterization of a new binding topology for molecules that bind to relatively long DNA sequences, a topology analogous to the way a snake might try to climb a ladder. The threading polyintercalator scaffold will allow unprecedented access to functional groups in both DNA grooves, with which one can derive extended sequence specificity in a relatively low molecular weight species. It is reasonable to anticipate that the relative lack of functionality combined with flexibility of the current linker designs will limit access to highly sequence-selective binding molecules. We are therefore currently investigating rigidified linkers with additional hydrogen bonding sites in preparation for the construction of next generation threading polyintercalators.

Acknowledgment. This work was supported by the American Cancer Society (RPG-97-085-01-CDD), the Welch Foundation (F-1188), and the National Institutes of Health (RO1 GM55646).

JA046335O