ELSEVIER

Chemistry & Biology 8 (2001) 415–425

Chemistry & Biology

www.elsevier.com/locate/chembiol

Research Paper

Peptide *bis*-intercalator binds DNA via threading mode with sequence specific contacts in the major groove

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

Department of Chemistry and Biochemistry, The University of Texas at Austin, Austin, TX 78722, USA

Received 4 December 2000; revisions requested 30 January 2001; revisions received 8 February 2001; accepted 15 February 2001
First published online 9 March 2001

Abstract

Background: We previously described a general class of DNA polyintercalators in which 1,4,5,8-naphthalenetetracarboxylic diimide (NDI) intercalating units are connected via peptide linkers, resulting in the first known *tetrakis*- and *octakis*-intercalators. We showed further that changes in the composition of the peptide tether result in novel DNA binding site specificities. We now examine in detail the DNA binding mode and sequence specific recognition of Compound 1, an NDI *bis*-intercalator containing the peptide linker gly-gly-gly-lys.

Results: ^{1}H-NMR structural studies of Compound 1 bound to d(CGGTACCG)$_2$ confirmed a threading mode of intercalation, with four base pairs between the diimide units. The NMR data, combined with DNAse I footprinting of several analogs, suggest that specificity depends on a combination of steric and electrostatic contacts by the peptide linker in the floor of the major groove.

Conclusions: In view of the modular nature and facile synthesis of our NDI-based polyintercalators, such structural knowledge can be used to improve or alter the specificity of the compounds and design longer polyintercalators that recognize correspondingly longer DNA sequences with alternating access to both DNA grooves.

Keywords: DNA recognition; Major groove binding; 1,4,5,8-Naphthalenetetracarboxylic diimide; NMR; Threading intercalation

1. Introduction

DNA intercalators represent a diverse source of novel DNA ligands of theoretical and medicinal interest. 1,4,5,8-Naphthalenetetracarboxylic diimide (NDI) derivatives have been previously shown to intercalate into DNA in a threading manner [1], and we previously described a modular NDI–peptide system [2] created to allow for threading polyintercalation with alternating blockade of both DNA grooves [3–7]. We recently showed that changes in the peptide linker lead to novel specificity [8] and also extended the number of NDI units from four [2] to eight [9], without loss of full intercalation (based on viscometry and UV measurements). However, the precise binding mode of such large linear molecules appears sensitive to DNA sequence and the composition of the tether connecting the intercalating units [10,11]. Despite a wealth of structural information on *mono*- and *bis*-intercalators, there is a conspicuous lack of data for the few known synthetic *tris*- and *tetra*-intercalators.

As a starting point for the design of diimide-based polyintercalators with programmed DNA binding specificity, we now have examined the DNA binding mode and sequence specific recognition of Compound 1 (Fig. 1), an NDI *bis*-intercalator containing the peptide linker gly-gly-gly-lys. The structures of several threading *mono*- and *bis*-intercalators [4,6,7,12,13], and a few non-threading intercalators which bind via the major groove [14–18] have been reported, including the *bis*-imide elinafide (LU 79553) [15] which is structurally related to NDI. However, prior to the present work, no structural data were available for an NDI–DNA complex. Our results confirm a threading intercalation of the NDI units and indicate placement of the peptide linker in the major groove, potentially opening the way to threading polyintercalation.

* Correspondence: Brent L. Iverson;
E-mail: biverson@utxvms.cc.utexas.edu

1074-5521/01/$ – see front matter
PII: S1074-5521(01)00013-8

2. Results and discussion

2.1. NMR studies of Compound 1 in complex with d(CGGTACCG)$_2$

Previous footprinting studies indicated potential Compound 1 binding sites [8]. Out of several sequences, (5′-CGGTACCG-3′)$_2$ was chosen for detailed study because preliminary experiments with the complex displayed well-resolved spectra. The imino proton region of the one-dimensional (1D) spectrum for the Compound 1–(5′-CGGTACCG-3′)$_2$ complex suggests binding with a 1:1 stoichiometry. Upfield shifts of the imino resonances of base pairs G2–C7 and G3–C6, as well as T4A–A5B are consistent with intercalation [19]. The C2 symmetry of the complexed DNA is distorted upon binding (Table 1), the differences between corresponding residues being most notable for the central TA step, while base pairs C1–G8 and G2–C7 virtually retain their symmetry in the bound species. In the exchangeable 1D spectrum, there are also significant downfield shifts of Compound 1 amide protons from Gly-3 and Lys-6 (10.1 and 9.2 ppm, respectively).

Four cross peaks were observed in the aromatic-to-aromatic region of the COSY spectrum, corresponding to eight distinct aromatic NDI protons (Fig. 2A). Those can be grouped by chemical shift into two similar sets, consistent with the two NDI moieties intercalated in a similar fashion. The resulting proton assignment was further supported by NOE cross peaks with the DNA (Fig. 2B). However, the data were not sufficient to unequivocally assign diimide protons D1 vs. D5 and D2 vs. D6. Therefore, two separate sets of NOE restraints were used in the restrained molecular dynamics (rMD) calculations and the resulting models were compared (see below).

The NOE data are consistent with a model in which the NDI units are similarly intercalated into the GG steps in a threading manner, with the connecting peptide linker spanning four base pairs diagonally across the major groove, and both termini protruding into the minor groove. DNA protons were assigned following the H6(8)–H2′/H2″ and H6(8)–H1′ NOE connectivities [20]. The connectivities are weak or interrupted for the steps G2–G3 and C6–C7 (Fig. 2B). NOE cross peaks between the H2′/H2″ and H1′ protons of the DNA and NDI are observed in place of several of the missing DNA connectivities, locating the intercalation sites between the two G residues at the (GG)–(CC) steps. Several cross peaks are observed between the diimide protons and the C6 and C7 amino protons. Additional intermolecular NOE cross peaks (Fig. 3) were observed in the major groove between the intercalator side chains L2 and L3 and H8 of G2 and G3, and in the minor groove between the methylene groups of Lys-1, L1 and L4 of Compound 1, and H1′ and H4′ of DNA bases C6 and C7. Linker residues Gly-3, Gly-4, and Gly-5 all have NOE contacts with the major groove methyl groups of T4A and T4B. In the exchangeable NOESY spectrum of the Compound 1–DNA complex, the base pairing is intact, as judged by the observation of all internal imino protons, although the G2A and B, and T4A imino protons only become visible in the NOESY spectrum below 10°C.

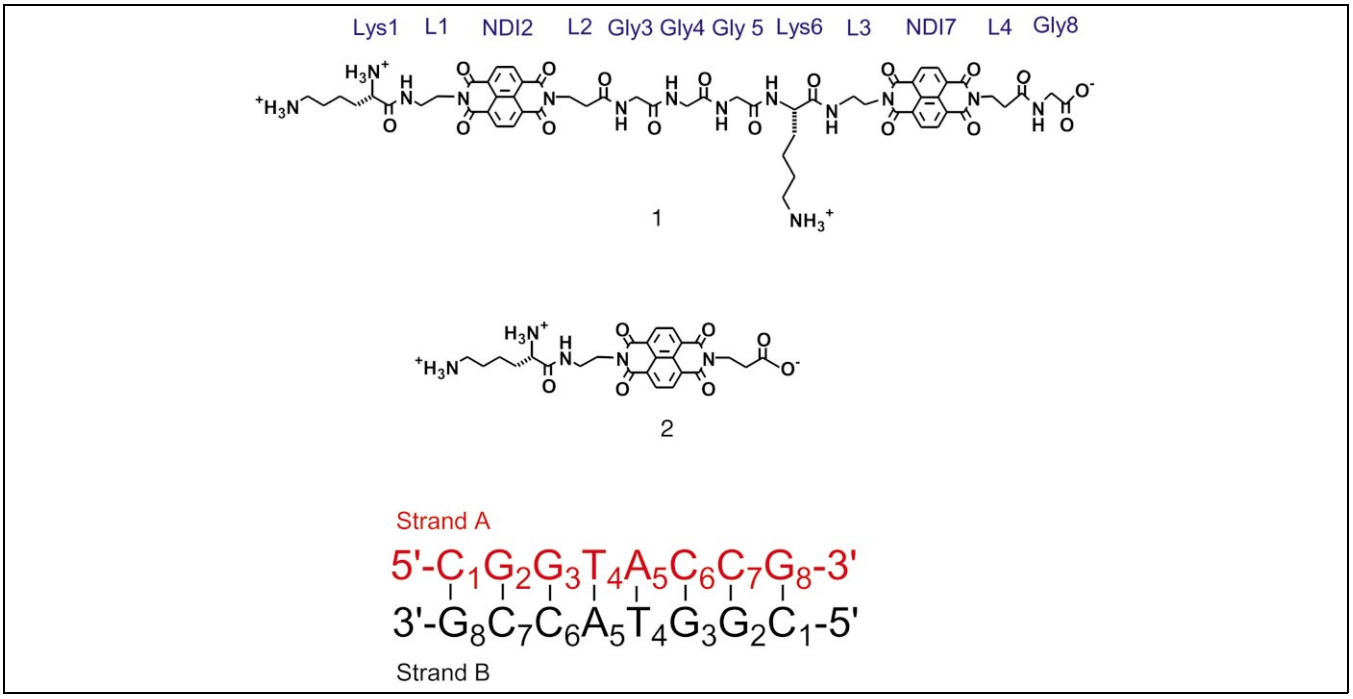


Fig. 1. Structure and naming convention of Compound 1, the corresponding mono-NDI derivative 2 and the DNA oligonucleotide used in the NMR study.

Table 1
DNA chemical shifts

Base	H6/8	H2/H5/M	H1′	H2′	H2″	H3′	H4′	NH2*	NH2*	NH*
C1A	*7.26* (*7.58*)	*5.63* (*5.90*)	*5.55* (*5.77*)	*1.41* (*1.88*)	*2.04* (*2.36*)	4.50	3.88	6.88 (n.a.)	8.04 (n.a.)	–
G2A	7.78 (7.92)	–	5.48 (5.58)	2.56 (2.75)	2.90 (2.75)	4.96	4.21	n.a.	n.a.	*12.03* (*13.08*)
G3A	7.75 (7.76)	–	*5.76* (*6.00*)	2.41 (2.56)	**2.85** (**2.78**)	4.80	4.51	*5.27* (*6.94*)	7.45 (n.a.)	*11.45* (*12.80*)
T4A	*6.98* (*7.22*)	***1.22*** (***1.48***)	***4.97*** (***5.74***)	1.97 (2.08)	***2.14*** (***2.47***)	4.71	3.94	–	–	*12.05* (*13.42*)
A5A	8.16 (8.26)	**7.54** (**7.49**)	**6.27** (**6.22**)	2.61 (2.70)	**2.92** (**2.85**)	4.96	4.33	(6.98)	–	–
C6A	7.37 (7.30)	*5.52* (*5.28*)	**5.94** (**5.84**)	*2.26* (*1.97*)	**2.26** (**2.36**)	4.87	3.90	*6.92* (*6.46*)	8.04 (8.08)	–
C7A	7.55 (7.43)	*5.86* (*5.62*)	*5.35* (*5.68*)	1.90 (1.97)	2.13 (2.32)	4.64	4.07	*7.17* (*6.90*)	*7.66* (*8.59*)	–
G8A	7.86 (7.94)	–	6.07 (6.17)	2.57 (2.62)	*2.26* (*2.60*)	4.59	4.10	–	–	–
C1B	*7.26* (*7.58*)	*5.63* (*5.90*)	*5.56* (*5.77*)	*1.43* (*1.88*)	*2.06* (*2.36*)	4.50	3.88	6.88 (n.a.)	8.04 (n.a.)	–
G2B	7.79 (7.92)	–	5.51 (5.58)	2.56 (2.75)	*2.95* (*2.75*)	4.96	4.21	8.10 (n.a.)	n.a. (n.a.)	*12.02* (*13.08*)
G3B	7.65 (7.76)	–	*5.39* (*6.00*)	2.37 (2.56)	***2.52*** (***2.78***)	4.72	4.42	*5.65* (*6.94*)	7.96 (n.a.)	*11.46* (*12.80*)
T4B	*6.88* (*7.22*)	***0.76*** (***1.48***)	*5.82* (*5.74*)	2.00 (2.08)	*2.42* (*2.47*)	4.83	4.17	–	–	***13.68*** (***13.42***)
A5B	8.20 (8.26)	***7.07*** (***7.49***)	***5.93*** (***6.22***)	*2.48* (*2.70*)	**2.66** (**2.85**)	4.96	4.18	n.a. (6.98)	–	–
C6B	7.43 (7.30)	5.36 (5.28)	***6.14*** (***5.84***)	*2.33* (*1.97*)	**2.49** (**2.36**)	4.88	4.05	*6.82* (*6.46*)	*7.75* (*8.08*)	–
C7B	7.53 (7.43)	*5.83* (*5.62*)	5.51 (5.68)	*1.76* (*1.97*)	2.11 (2.32)	4.58	4.07	*7.18* (*6.90*)	*7.76* (*8.59*)	–
G8B	7.85 (7.94)	–	6.06 (6.17)	2.58 (2.62)	*2.27* (*2.60*)	4.59	4.10	–	–	–

Chemical shifts (ppm) of the complexed and free DNA at 27°C (*10°C) in 30 mM Na phosphate buffer (pH 7.5). Numbers are shown in italics if the difference in shift in the free and complexed DNA is greater than 0.2 ppm and in bold if the difference between the corresponding residues in the two strands in the complex differ by more than 0.2 ppm.

2.2. *Modeling*

A variety of starting structures were used in constructing a model of the ligand-bound DNA, so that the full range of structures that are consistent with the NMR data could be identified. The position of the ligand was found to be similar in all derived structures. Qualitative analysis of the H1′–H2′/H2″ and H3′–H2′ coupling constants and

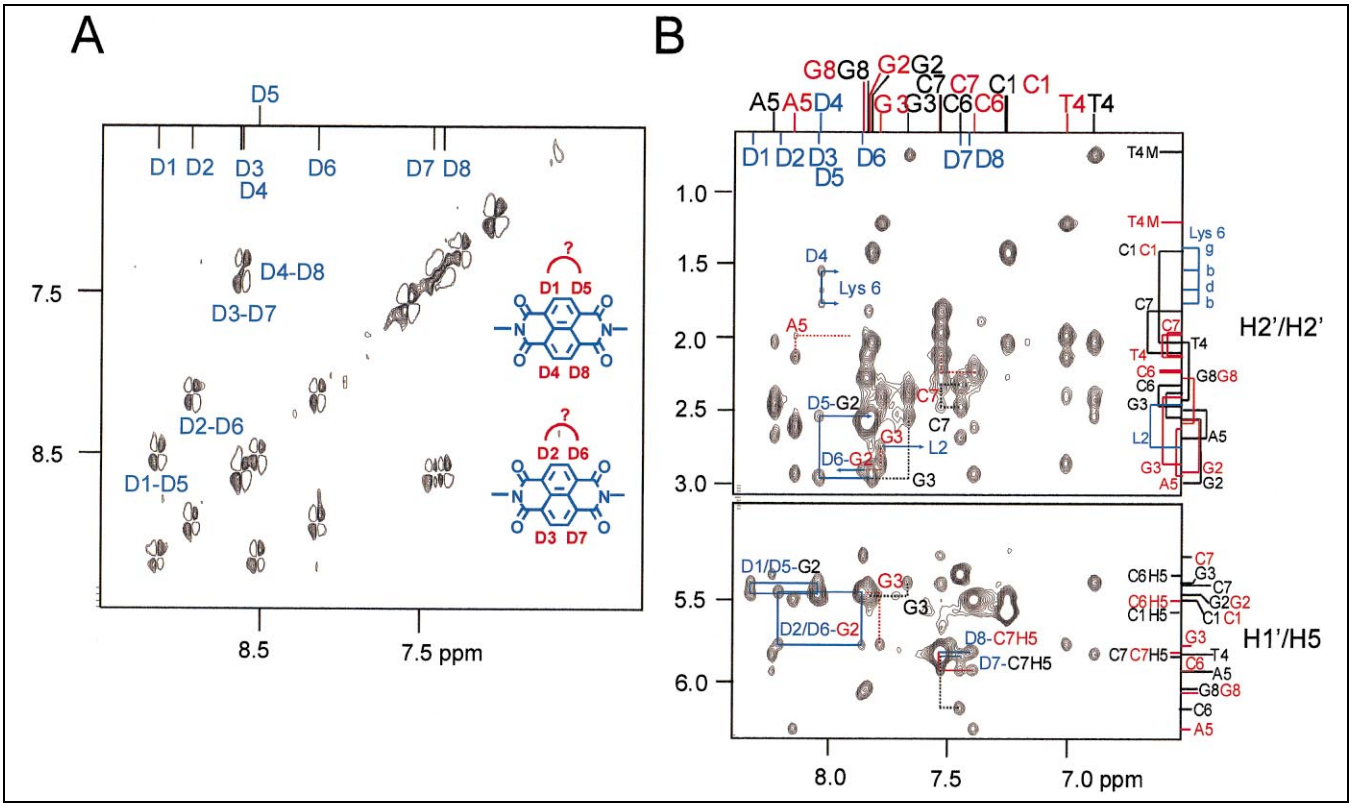
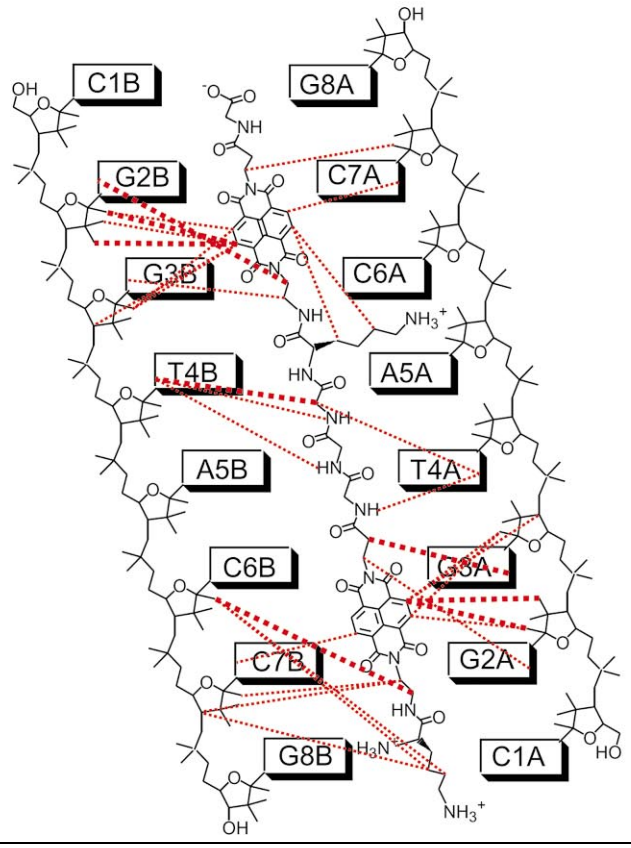


Fig. 2. A: Aromatic region of the DQF-COSY spectrum (D_2O, 27°C) of the Compound 1–DNA complex, showing the eight distinct NDI proton resonances. The assignments of the NDI rings are based on the COSY cross peaks shown, chemical shift and NOEs with the DNA and Lys-6 (shown in B). Protons D1–D5 and D2–D6 could not be unambiguously assigned. B: Sequential DNA H6/H8 to H2′/H2″ and H6/H8 to H1′ connectivities in the NOESY spectrum (D_2O, 60 ms mixing time, 10°C) of the Compound 1–DNA complex. Red lines, strand A; black lines, strand B. Weak or missing NOE connectivities between G2B–G3B, G2A–G3A, C6B–C7B and C6A–C7A are 'replaced' by NOEs between the intercalator protons and DNA (blue lines): G2B–D5, D5–G3B, G2A–D6, D6–G3A, D7–C7A and D8–C7B. Also note the Lys-6–D4 NOEs.

A



B

NOE	LSR Residue	Proton	DNA Residue	Proton	Intensity
1	NDI 7	D5	G2B	H2'	weak
2	NDI 7	D5	G2B	H2''	medium
3	NDI 7	D5	G2B	H1'	strong
4	NDI 7	D5	G3B	H1'	weak
5	NDI 7	D5	G3B	H4'	weak
6	NDI 7	D1	G2B	H1'	weak
7	NDI 7	D1	G3B	H1'	weak
8	NDI 7	D8	C7A	H5	weak
9	NDI 2	D6	G2A	H2''	medium
10	NDI 2	D6	G2A	H1'	strong
11	NDI 2	D6	G3A	H1'	weak
12	NDI 2	D6	G3A	H4'	weak
13	NDI 2	D2	G2A	H1'	weak
14	NDI 2	D2	G3A	H1'	weak
15	NDI 2	D7	C7B	H5	weak
16	L1	Hb*	C6B	H1'	medium
17	L1	Ha*	C7B	H1'	weak
18	L1	Ha*	C7B	H4'	weak
19	L2	Hg*	G2A	H8	weak
20	L2	Hb*	G3A	H8	medium
21	L3	Ha*	G2B	H8	medium
22	L3	Hb*	G3B	H8	weak
23	L4	Ha*	C7A	H1'	weak
24	Lys 1	Hb*,Hg*	C6B	H1'	weak
25	Lys 1	Hg*	C7B	H4'	weak
26	Gly 3	HN	T4A	M	weak
27	Gly4	HN	T4B	M	weak
28	Gly 5	Ha*	T4B	M	strong
29	Gly 5	Ha*	T4A	M	weak
30	Gly 5	HN	T4B	M	weak

Fig. 3. A: Diagram of the observed intermolecular NOEs (60 ms mixing time, 10°C). The thick dashed lines indicate 'strong' and 'medium' intensity cross peaks; 'weak' NOEs are shown as thin, dotted lines. B: DNA–ligand NOEs used in the rMD calculations. Asterisks indicate that peaks were observed for both diastereotopic protons and the NOE restraints employed pseudoatoms.

the relative intensities of the H6(8)–H2′/H2″/H3′ NOEs (Table 2) suggest that the DNA is in a B-like form [20], with the furanose rings of the internal DNA residues except C6A, C6B and A5B, in the C2′-endo (S) conformation. The latter three residues appear to be mixtures of S- and N-type puckers. In the rMD simulations all residues except C6A, C6B, and A5B were constrained as C2′-endo, while those residues were kept in the C3′-endo conformation to produce the DNA model.

B-like DNA in the complex is consistent with the crystal structures of several intercalators complexed to $(C\underline{GTAC}G)_2$ (Fig. 4). The chemical shift differences between T4A and T4B (Table 1) suggest that binding of the asymmetric ligand causes a distortion of the central (AT).(AT) region. Our model lacks sufficient 'resolution' to determine the exact nature of this deformation. Perhaps of some significance is a displacement of the T4A residue outward into the major groove, which occurs in the model, irrespective of the A5B pucker and linker–DNA NOE restraints used. Such distortion is consistent with the interrupted T4AH2′–A5AH8 NOE connectivity, and with the unusual upfield chemical shifts of T4A protons HN3 and H1′, which are perhaps more exposed to the aromatic ring currents of G3A and A5A, respectively [21].

The observed NOEs place the NDI groups between the G2 and G3 of the two GG–CC base pairs, diagonally extending toward C6–C7 in the minor groove. As evident from Fig. 5, the ambiguity in the assignment of the diimide protons D1–D5 and D2–D6 affects mostly the depth of intercalation. Assuming that an upfield shift roughly corresponds to the extent of overlap with the DNA bases [21], the deeper intercalation geometry shown is more consistent with the observed chemical shifts.

In the recently reported complex of the structurally related naphthalimide *bis*-intercalator LU 79553 [15], the two [mono]imide chromophores are intercalated via the major groove with a similar geometry to what is seen in our complex. Interestingly, the imide chromophores of LU exhibit ring flipping within the intercalation site on the millisecond time scale. Such dynamic behavior was ascribed to unfavorable electrostatic repulsions between the intercalating aromatic units and the DNA bases. Judging from the NOESY and ROESY data (not shown), ring flipping was not evident for the intercalating diimide units

Table 2
Qualitative analysis of DNA sugar pseudorotation angles (P)

Base	$(H6(8)–H2'/2'')_i$ NOESY	$(H6(8)–H2'/2'')_S$ NOESY	H1′–H2′/2″ NOESY	H1′–H2′/2″ COSY	H3′–H2′/2″ COSY	H1′–H4′ NOESY	H3′–H4′ COSY	H2′/2″–H4′ NOESY	H6(8)–H3′ NOESY	P
C1A	S[a]/M[b]	M/S	W[c]/S	S/M	**S/W**	M	**M**	M/M	**M**	Mix[e]
G2A	S/M	–/–	W/M	S/W	W/W	M	W	W/W	W	South[f]
G3A	S/M	W/M	W/M	S/W	W/W	M	W	W/W	W	South
T4A	S/M	W/M	W/M	S/W	n.a.	M	n.a.	**W/M**	W	South
A5A	S/M	W/M	W/S	S/W	W/W	W	W	W/W	W	South
C6A	overlap	overlap	overlap	overlap	overlap	M	**S**	overlap	**M**	Mix
C7A	S/M	W/S	W/S	S/W	n.a.	M	n.a.	W/W	W	South
G8A	S[d]/M	–	M/S	S/M	**S/W**	M	**M**	M/M	**M**	Mix
C1B	S/M	M/S	W/S	S/M	**S/W**	M	**M**	M/M	**M**	Mix
G2B	S/M	–/–	W/M	S/W	W/W	M	W	W/W	W	South
G3B	S/M	W/M	W/M	S/W	n.a.	M	n.a.	W/W	W	South
T4B	S/M	W/M	W/M	S/W	W/W	M	W	W/W	W	South
A5B	S/M	W/M	W/M	**M/M**	W/W	M	**M**	W/W	W	Mix
C6B	S/M	M/W	M/S	**M/M**	**M/W**	M	**S**	overlap	**M**	Mix
C7B	S/M	W/S	W/S	S/W	**M/W**	M	W	M/W	W	South
G8B	S[d]/M	–	M/S	S/M	**S/W**	M	**M**	M/M	**M**	Mix

Qualitative analysis of DNA sugar pseudorotation angles (P) from interproton distances and coupling constants affected by deoxyribose conformation. Boldface entries suggest C3′-endo ('North') contribution to the sugar conformation.
[a]S, strong.
[b]M, medium.
[c]W, weak cross peak intensity.
[d]H2′/H″ assignments based on shift and relative H6(8)–H2′/2″ NOE intensity. In all cases except for the 3′-terminal G8 residues and C6A, H2′ is upfield of H2″.
[e]Mix, mixed C2′-endo/C3′-endo deoxyribose conformation.
[f]South, predominantly C2′-endo deoxyribose conformation.

of Compound 1. This may be due to increased steric strain provided by the additional diimide side chains facing the minor groove, or the different intercalation steps (TG vs. GG) in the two complexes.

The peptide linker of Compound 1 extends diagonally across the DNA major groove, with the central Gly residue placed between the two methyl groups of T4A and T4B. The DQF-COSY data for side chains L2 and L3 are consistent with the gauche conformation. Several NOE cross peaks between diimide proton D4 and Lys-6 Hβ and Hγ (Fig. 2B) indicate that the Lys-6 side chain is extended across the major groove toward the phosphate backbone of strand A. TOCSY cross peaks between the side chain Hε, Hδ, Hγ and the exchangeable amino protons suggest hydrogen bonding with the DNA for both lysine residues. According to our model, for the internal Lys-6 this probably occurs at the phosphate group of nucleotide A5A, and/or N7 of the same adenine (Fig. 6).

There appear to be specific hydrogen bonds to the DNA at each end of the four amino acid linker. The orientation of the amide proton of L3 and its large downfield shift suggest a hydrogen bond between L3 NH and O6 of G3B. Importantly, an analogous NH–O6 hydrogen bond was observed for a corresponding ammonium group positioned two carbon atoms away from the imide ring of LU 70553 [15]. The amide proton of Gly-3, adjacent to the other NDI ring, but three carbon atoms away, resonates even further downfield than L3 suggesting that it may be hydrogen bonded to O6 of G3A.

Fig. 6 shows several structures obtained by rMD with and without the inclusion of the suspected hydrogen bonds (shown in Fig. 6C), introduced as distance restraints between Compound 1 and the DNA. The major change in the model from these additional restraints occurs in the lower portion of the linker, which is now tucked more deeply along the bottom of the DNA groove.

The threaded termini of the ligand are not well-defined by the NMR data and may possess greater conformational flexibility. The COSY spectrum indicates *trans* torsion angles for linkers L1 and L4. Weak NOEs between Lys-1 and L1 and the C6 and C7 residues in the minor groove (Fig. 3), the lack of other such NOEs upstream or downstream from the GG step, and the size of the DNAse I footprints of Compound 1 (Fig. 7) are consistent with a model in which the positively charged Lys-1 residue is folded sideways toward the phosphate backbone, instead of laid out inside the minor groove. Only one weak intermolecular contact was identified for the negatively charged C-terminus of the ligand, between the linker L4 and the ribose H1′ of C7A (Fig. 3). This is consistent with the proposed orientation of the adjacent NDI 7 ring, and places Gly-8 in the minor groove.

2.3. Footprinting of Compound 1 and analogs

In order to investigate the fine specificity of binding, DNAse I footprinting studies with Compound 1, several derivatives, and a monomer control 2 (Fig. 1) were carried

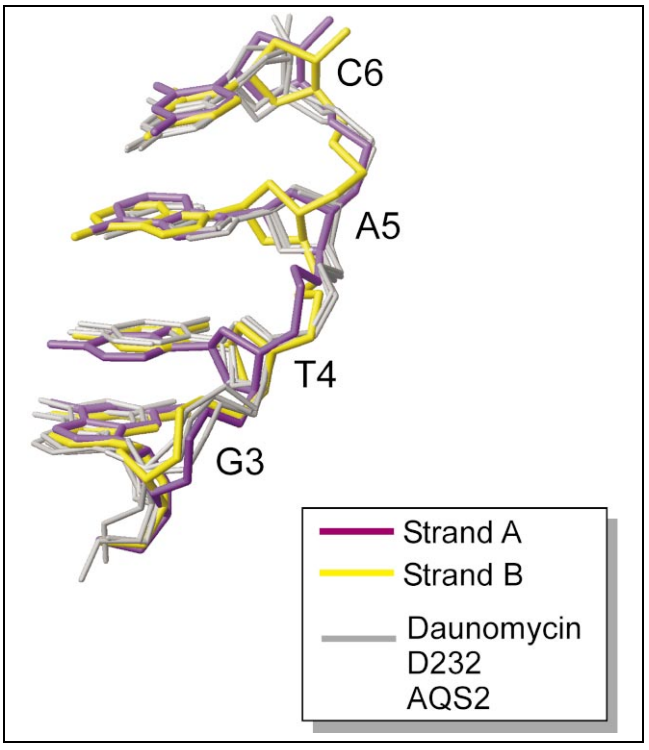


Fig. 4. A: Comparison of the crystal structures of three intercalators: daunomycin (PDB ID 1D11, [32]), D232 (PDB ID 1C9Z, [16]) and AQS2 (PDB ID 386D, [18]) complexed to CGTACG, and the two GGTACC strands of the DNA–Compound 1 model, obtained from a B-DNA starting structure, with the mixed C2′-endo/C3′-endo sugars (C6A, C6B and A5B) restrained as C3′-endo.

out using a synthetic 77 bp DNA fragment containing several 5′-GGNNCC-3′ sites (Fig. 7A). Compound 1 binds specifically to the sequence 5′-GGPyPuCC-3′ ($K_d \sim 100$ nM, estimated from the footprints) with a preference for 5′-GGTACC-3′ over 5′-GGCGCC-3′. Binding is significantly less for the sequences 5′-GGATCC-3′ and 5′-GGGCCC-3′. The monomer control, H_2N-Lys-NDI-Gly-COOH, binds roughly equally well to any of the 5′-GGN-3′ sites, with binding affinity ~1000-fold lower than the dimer. It is known that *mono*-intercalators can bind preferentially to certain helical structures (e.g. ethidium bromide prefers the A-form polyA.polydT over the 'B′-form' polydA.polydT [22], and B-form polydGdC over the Z-form [23]). It has also been observed that a GGT triad has some unusual conformational properties [24,25], and that a TA step (but not an AT step) between adjacent intercalation sites possibly 'transmits' a conformational change required for intercalation [26,27]. However, the NDI monomer 1 does not appear to discriminate between the four different GGN combinations. This suggests that the specificity of the dimer 1 for the central TA step of GGTACC is at least in part due to the peptide linker, and not the conformational properties of the DNA sequence alone.

In order to investigate the influence of the positively charged Lys side chains, we conducted footprinting studies on several analogs that differed in the position of Lys residues (Fig. 7B). From the footprinting data, it appears that a Lys residue in position 3 impedes binding to 5′-GGTACC-3′, while positions 6 and 1 are 'neutral'. Only placing a Lys at position 8 significantly improves binding. This observed enhancement of binding due to a positively charged residue in the C-terminal position (position 8) can, in retrospect, be anticipated. Compound 1 possesses an overall negatively charged glycine C-terminus in close proximity to the phosphate backbone of the minor groove that might have a destabilizing effect on binding. On the other hand, a Lys residue in position 8 will result in an overall neutral C-terminus. Note that moving the Lys residues around the molecule does not significantly alter sequence preference, indicating that the Lys side chains of Compound 1 make primarily electrostatic contributions to overall binding as opposed to specificity determining contacts with the DNA bases.

2.4. Molecular basis for observed sequence specificity

2.4.1. NDI intercalation

Rill and co-workers have shown that NDI rings possess

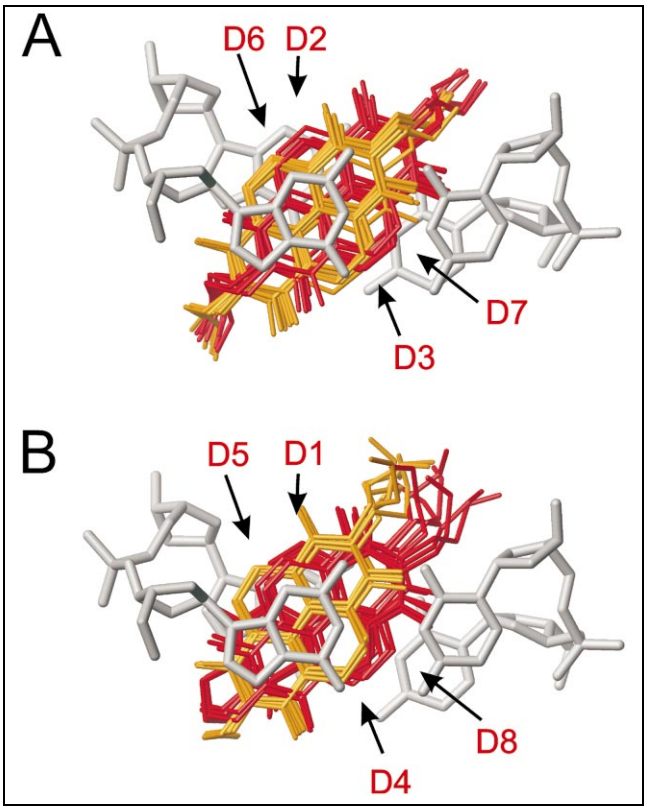


Fig. 5. Intercalation geometry of the aromatic rings NDI 2 (A) and NDI 7 (B), obtained from rMD calculations of the ligand on a rigid DNA with two alternative assignments for protons D1/D5 and D2/D6 (red and orange). The upfield δ shifts of the NDI protons (due to DNA base ring currents) qualitatively match the geometry shown in red: D1, least overlap (most downfield shift); D8, most overlap (most upfield shift).

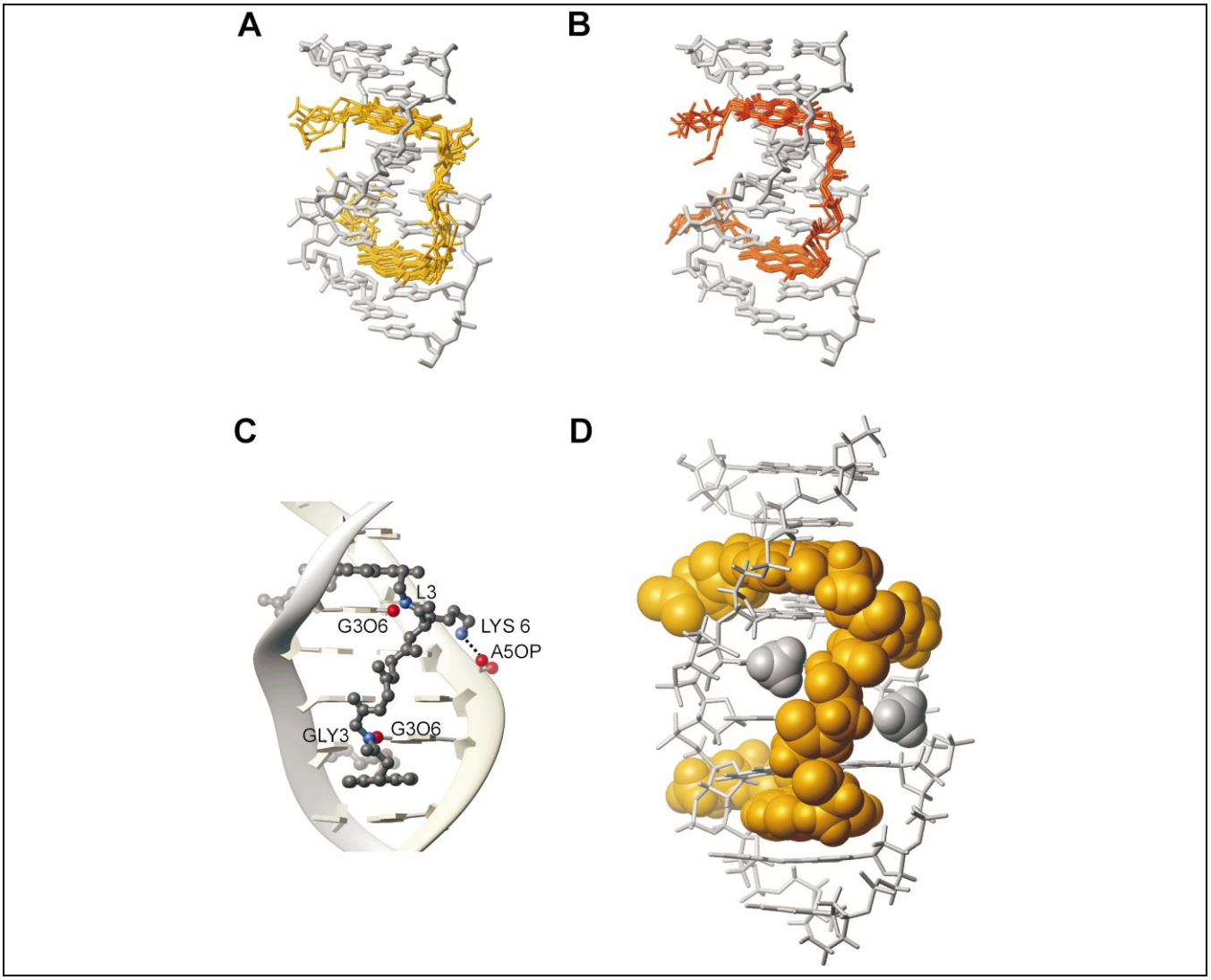


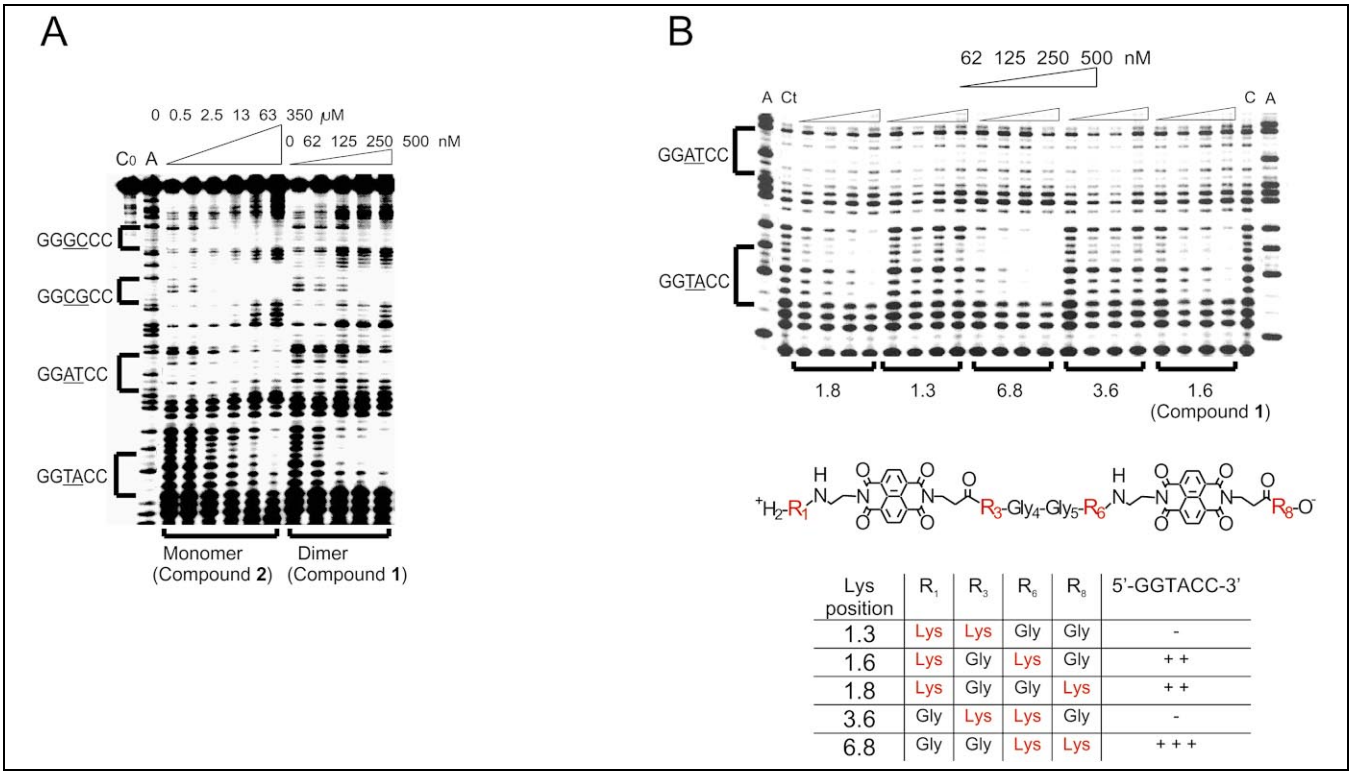
Fig. 6. Model of the ligand–CGGTACC complex based on the NMR data. Seven structures obtained without (A) and with (B) the postulated intermolecular H-bonds shown in C; D: average minimized model from B showing the thymine methyl groups in the major groove. The figures were created in MOLMOL [33].

5′-NG-3′ specificity and suggested that this was due to favorable electrostatic interactions between the carbonyl groups of NDI and the exocyclic N^2 amino group of guanine in the minor groove [28]. Based on our footprinting results for the monomeric derivative, the NDI amino acid units possess a preference for binding G–G steps. Such preference may be enhanced by specific hydrogen bonds between the amide N–H groups at each end of the linker. Importantly, we previously identified a Compound 1 analog with strong binding to a DNA site which contained no G–G steps [8]. This is consistent with linker-dependent sequence specificity, although, as discussed below, it is not entirely clear whether the linker directly contributes to the free energy of specific binding. Alternatively, a linker could act by destabilizing binding to disfavored sequences, as has been proposed for the *bis*-intercalator TOTO [6].

2.4.2. Sterics

According to the NOE data, residues Gly-4 and Gly-5 are positioned between the methyl groups of the thymidine residues of 5′-GGTACC-3′, in what could be considered a relatively narrow, hydrophobic cleft in the major groove of this sequence (Fig. 8A) [29]. The methylene group of Gly-5 is particularly buried in the cleft, perhaps providing a desolvation driving force for sequence specific binding. Inverted positioning of the thymidine methyl groups, as occurs in the sequence 5′-GGATCC-3′, changes the topology of the major groove, eliminating the cleft and possibly providing a steric barrier to linker binding along the bottom of the groove. Such a steric bias is consistent with the poor binding to 5′-GGATCC-3′ relative to 5′-GGTACC-3′ observed in the footprinting studies.

The exocyclic amino groups of cytosine residues also



Lys position	R_1	R_3	R_6	R_8	5'-GGTACC-3'
1.3	Lys	Lys	Gly	Gly	-
1.6	Lys	Gly	Lys	Gly	+ +
1.8	Lys	Gly	Gly	Lys	+ +
3.6	Gly	Lys	Lys	Gly	-
6.8	Gly	Gly	Lys	Lys	+ + +

Fig. 7. DNAse I footprinting of Compound 1 and several analogs. A: Comparison of 1 and the monomer analog 2 (Fig. 1) on a 77 bp nucleotide containing four different 5′-GGNNCC-3′ sites (located within the brackets). B: Effect of Lys position on GGTACC binding. Lanes C0 contain no DNAse I, lanes Ct contain DNAse I but no compound, lanes A contain adenine chemical cleavage reaction [34].

put steric bulk into the bottom of major groove. Similar to the situation with the thymidine methyl groups, the sequence 5′-GGCGCC-3′ allows for unhindered linker binding along a cleft in the bottom of the groove, while the inverted sequence 5′-GGGCCC-3′ does not possess such a cleft and may, in fact, have a topology that interferes with linker association. These steric parameters are also consistent with the sequence preferences of 5′-GGCGCC-3′ over 5′-GGGCCC-3′ seen in the footprinting experiments.

2.4.3. Electrostatic complementarity

While the steric landscapes of 5′-GGTACC-3′ and 5′-GGCGCC-3′ are both amenable to linker binding in a cleft along the bottom of the major groove, the electrostatic patterns of these two sequences are markedly different (Fig. 8B). Close inspection of the calculated electrostatic potential surfaces for our model (Fig. 8C) reveals that the amide H atoms along the entire linker backbone are adjacent to atoms of complementary high electron density (red color), perhaps providing an electrostatic component to the driving force for recognition of this sequence. The other sequences have quite different electrostatic patterns at the bottom of their major grooves, expected to provide differing levels of electrostatic complementarity, even when linker flexibility is considered.

3. Significance

What emerges from our structural analysis is the following model of sequence specificity. The NDI intercalating units possess a preference for binding G–G steps, that is either enhanced by, or perhaps the result of, specific hydrogen bonds between the amide N–H groups at each end of the linker.

The strong preference of Compound 1 for the internal Py–Pu sequence appears to be the result of steric and electrostatic complementarity between the amino acid linker and floor of the major groove, likely assisted by desolvation (i.e. the hydrophobic effect) of the Gly-5 methylene unit buried among the methyl groups in the case of 5′-GGTACC-3′. At the present time, it is not possible to determine with any greater detail the relative contributions of the listed steric, electrostatic, and hydrophobic factors responsible for the observed sequence specificity. It is also not possible to entirely rule out contributions from other influences such as unique conformational or dynamic characteristics of the different DNA sequences. Nevertheless, the structural studies reported here have provided considerable insight into the mode of binding of the NDI-based polyintercalators, and as such provide a structural foundation for the construction of next-generation molecules with altered sequence specificity programmed by design. To further assist in this process, we are cur-

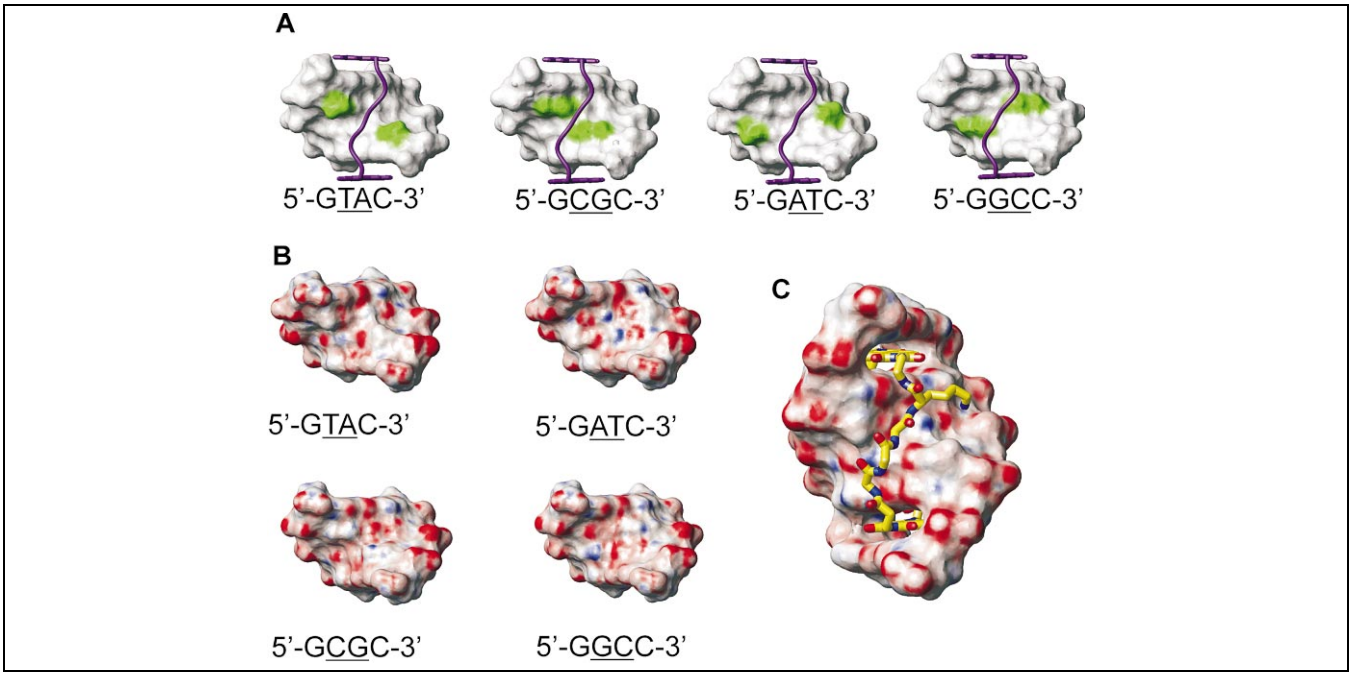


Fig. 8. A: Contact surface of the major groove of four GNNC nucleotides in the canonical B-form. The exocyclic thymine methyl groups, cytosine amino groups and cytosine H5 hydrogens are colored green. Superimposed in purple is a ribbon representation of the peptide backbone of Compound 1 from the NMR model. B and C: Contact surface colored by electrostatic potential (blue, positive; red, negative) of the four GNNC nucleotides and the ligand–GGTACC complex (NMR model). The figures were created in MOLMOL [33].

rently investigating by NMR the details of DNA binding of a second derivative with a different peptide linker and entirely different DNA specificity [8].

A remarkable feature of Compound 1 is that DNA sequence specificity is observed for such a flexible molecule. Coupled to the modular nature and facile synthesis of our NDI-based polyintercalators, it should prove possible to use structural knowledge to design longer polyintercalators that recognize correspondingly longer DNA sequences. The immediate challenge will be to find specific linkers that favor binding in the minor groove, so that these may be introduced into a hybrid molecule that recognizes both grooves in an alternating fashion. If possible, such an approach would set the stage for threading polyintercalators that achieve the promise of programmable sequence specificity of very long stretches of DNA.

4. Materials and methods

4.1. Sample preparation

Compound synthesis has been described in detail elsewhere [11]. The DNA (gel filtration grade, Midland Certified, Midland, TX, USA) was initially dissolved in 0.5 ml 30 mM Na phosphate buffer, pH 7.5. Prior to addition of compound, the DNA samples were diluted to ~20 ml chilled water to which Compound 1 (~2 μM) was added and the samples were lyophilized. For spectra in D_2O the samples were lyophilized twice from increasing purity D_2O and finally suspended in 0.5 ml 99.96% D_2O (Cambridge Isotope, Cambridge, MA, USA). The final concentration of the complex was approximately 2 mM; a precipitate formed at higher concentrations.

4.2. NMR spectroscopy

Spectra were obtained on a 500 MHz Varian Inova spectrometer. NOESY (60 and 200 ms mixing times), DQF-COSY, TOCSY (100 ms mixing time) and ROESY spectra in D_2O were acquired with States phase cycling, a spectral width of 6000 Hz in 2048 complex points (*t2*) and 512 (*t1*) experiments, using presaturation to remove the HOD signal. The spectra were processed with VNMR (Varian) using a Gaussian window function and referenced to H_2O. To observe solvent-exchangeable protons, NOESY (150 ms mixing time) and TOCSY spectra were acquired in 9:1 H_2O:D_2O, with a spectral width of 13 154 in 4096 complex points, using the 1–1 jump-return water suppression sequence [30] and presaturation, respectively.

4.3. Modeling

Restrained molecular dynamics (rMD) simulations were performed in XPLOR 3.1 [31], with additional parameters for the NDI residues. A total of 119 DNA–DNA and 28 DNA–ligand NOE restraints were grouped as 'strong' (0–2.5 A), 'medium' (0–3.5 A) and 'weak' (0–5 A). The initial structure was generated by manually docking the compound to B-form DNA, followed by geometry optimization. To generate the final DNA model, multi-

ple rMD runs were performed using the repel option for the non-bonded energy term. NOE restraints were constant throughout the calculation and initially excluded amino acid linker to DNA contacts. All deoxyribose rings were restrained to the C2′- or C3′-endo conformation (see text). Included in the calculation were NOE restraints representing Watson–Crick hydrogen bonding, planarity restraints for all DNA bases and torsion angle restraints to keep the end-bases in the B-conformation. The resulting structures were averaged and geometry optimized using a full VDW energy term. The DNA was fixed and Compound 1 was subjected to rMD with gradual increase of the NOE restraints (with and without a set of H-bond restraints).

4.4. DNAse I footprinting

The 77 bp synthetic fragment was purchased from Midland Certified. Footprinting was performed as previously described [8,35]. A binding constant for Compound 1 was visually estimated as the concentration of ligand at which protection from DNAse cleavage is roughly one half the maximum.

Acknowledgements

This work was supported by the American Cancer Society (RPG-97-085-01-CDD), the Welch Foundation (F-118) and the National Institutes of Health (RO1 GM55646).

References

[1] F. Tanious, S. Yen, W.D. Wilson, Kinetic and equilibrium analysis of a threading intercalation mode: DNA sequence and ion effects, Biochemistry 30 (1991) 1813.
[2] R.S. Lokey, Y. Kwok, V. Guelev, C. Pursell, L. Hurley, B.L. Iverson, A new class of polyintercalating molecules, J. Am. Chem. Soc. 119 (1997) 7202.
[3] S. Takenaka, S. Nishira, K. Tahara, H. Kondo, M. Takagi, Synthesis and characterization of novel tris-intercalators having potentially two different DNA binding modes, Supramol. Chem. 2 (1993) 41.
[4] J.M. Veal, Y. Li, S.C. Zimmerman, C.R. Lamberson, M. Cory, G. Zon, W.D. Wilson, Interaction of a macrocyclic bisacridine with DNA, Biochemistry 29 (1990) 10918–10927.
[5] C. Lamberson, Topologically constrained DNA intercalators: Synthesis and DNA binding studies, Ph.D. thesis, University of Illinois, IL, 1991, pp. 23–26.
[6] M. Petersen, A.A. Hamed, E. Petersen, J.P. Jacobsen, Bis-intercalation of homodimeric thiazole orange dye derivatives in DNA, Bioconjug. Chem. 10 (1999) 66–74.
[7] M. Jourdan, J. Garcia, J. Lhomme, M.P. Teulade-Fichou, J.P. Vigneron, J.M. Lehn, Threading bis-intercalation of a macrocyclic bisacridine at abasic sites in DNA: nuclear magnetic resonance and molecular modeling study, Biochemistry 38 (1999) 14205–14213.
[8] V. Guelev, M. Harting, R.S. Lokey, B.L. Iverson, Altered sequence specificity identified from a library of DNA-binding small molecules, Chem. Biol. 7 (2000) 1.
[9] M. Murr, M. Harting, V. Guelev, J. Ren, J.B. Chaires, B.L. Iverson, An oktakis-intercalating molecule, Bioorg. Med. Chem., in press.
[10] L. Wakelin, Polyfunctional DNA intercalating agents, Med. Res. Rev. 6 (1986) 275.
[11] V. Guelev, M. Cubberley, M. Murr, R.S. Lokey, B.L. Iverson, Design, synthesis and characterization of DNA polyintercalators, Methods Enzymol., in press.
[12] Y. Liaw, Y. Gao, H. Robinson, G. Van der Marel, J. Van Boom, A. Wang, Antitumour drug nogalamycin binds DNA in both grooves simultaneously: molecular structure of nogalamycin-DNA complex, Biochemistry 28 (1989) 9913–9918.
[13] H. Robinson, W. Priebe, J.B. Chaires, A.H. Wang, Binding of two novel bisdaunorubicins to DNA studied by NMR spectroscopy, Biochemistry 36 (1997) 8663–8670.
[14] Q. Gao, L.D. Williams, M. Egli, D. Rabinovich, S.-L. Chen, G.J. Quigley, A. Rich, Drug-induced DNA repair: X-ray structure of a DNA-ditercalinium complex, Proc. Natl. Acad. Sci. USA 88 (1991) 2422–2426.
[15] J. Gallego, B. Reid, Solution structure of a complex between DNA and the antitumour bisnaphthalimide LU-79553: Intercalated ring flipping on the millisecond time scale, Biochemistry 38 (1999) 15104.
[16] X. Shui, M.E. Peek, L.A. Lipscomb, M. Gao, C. Ogata, B.P. Roques, C. Garbay-Jaureguiberry, A.P. Wilkinson, L.D. Williams, Effects of cationic charge on three-dimensional structures of intercalative complexes: structure of a bis-intercalated DNA complex solved by MAD phasing, Curr. Med. Chem. 7 (2000) 59–71.
[17] C.L. Kielkopf, K.E. Erkkila, B.P. Hudson, J.K. Barton, D.C. Rees, Structure of a photoactive rhodium complex intercalated into DNA, Nat. Struct. Biol. 7 (2000) 117–121.
[18] S.M. Gasper, B. Armitage, X. Shui, G.G. Hu, C. Yu, G.B. Schuster, L.D. Williams, Three-dimensional structure and reactivity of a photochemical cleavage agent bound to DNA, J. Am. Chem. Soc. 120 (1998) 12402–12409.
[19] J. Feigon, W.A. Denny, W. Leupin, D.R. Kearns, Interactions of antitumor drugs with natural DNA: 1H NMR study of binding mode and kinetics, J. Med. Chem. 27 (1984) 450–465.
[20] K. Wüthrich, NMR of Proteins and Nucleic Acids, Wiley, New York, 1986.
[21] S.S. Wijmenga, M. Kruithof, C.W. Hilbers, Analysis of 1H-chemical shifts in DNA: assessment of the reliability of chemical shift calculations for use in structure refinement, J. Biomol. NMR 10 (1997) 337–350.
[22] J. Ren, J.B. Chaires, Sequence and structural selectivity of nucleic acid binding ligands, Biochemistry 3 (1999) 16067–16075.
[23] F. Pohl, T.M. Jovin, W.R. Baehr, J.J. Holbrook, Ethidium bromide as a cooperative effector of a DNA structure, Proc. Natl. Acad. Sci. USA 69 (1972) 3805–3809.
[24] H. Rozenberg, D. Rabinovich, F. Frolow, R.S. Hegde, Z. Shakked, Structural code for DNA recognition revealed in crystal structures of papillomavirus E2-DNA targets, Proc. Natl. Acad. Sci. USA 95 (1998) 15194–15199.
[25] X.-J. Lu, Z. Shakked, W.K. Olson, A-form conformational motifs in ligand-bound DNA structures, J. Mol. Biol. 300 (2000) 819–840.
[26] C. Bailly, F. Hamy, M.J. Waring, Cooperativity in the binding of echinomycin to DNA fragments containing closely spaced CpG sites, Biochemistry 35 (1996) 1150–1161.
[27] F. Gago, Stacking interactions and intercalative binding, Methods 14 (1998) 277–292.
[28] Z.-R. Liu, K.H. Hecker, R.L. Rill, Selective DNA binding of (N-alkylamine)-substituted naphthalene imides and diimides to G+C-rich DNA, J. Biomol. Struct. Dynam. 14 (1996) 331–339.
[29] C. Bailly, S. Crow, A. Minnock, M.J. Waring, Demethylation of thymine residues affects DNA cleavage by endonucleases but not sequence recognition by drugs, J. Mol. Biol. 291 (1999) 561–573.
[30] V. Sklenar, A. Bax, Spin echo water suppression for the generation of pure-phase two dimensional NMR spectra, J. Magn. Reson. 74 (1987) 469.

[31] A. Brünger, XPLOR 3.1, Yale University Press, New Haven, CT, 1993.
[32] A.H. Wang, G. Ughetto, G.J. Quigley, A. Rich, Interactions between an anthracycline antibiotic and DNA: molecular structure of daunomycin complexed to d(CpGpTpApCpG) at 1.2-A resolution, Biochemistry 26 (1987) 1152–1163.
[33] R. Koradi, M. Billeter, K. Wüthrich, MOLMOL: a program for display and analysis of macromolecular structures, J. Mol. Graph. 14 (1996) 51–55.
[34] B.L. Iverson, P.B. Dervan, Adenine-specific chemical sequencing reaction, Methods Enzymol. 218 (1993) 222–227.
[35] K.R. Fox, DNAse I footprinting, in: K.R. Fox (Ed.), Drug–DNA Interaction Protocols, Humana Press, Totowa, NJ, 1997, pp. 1–22.