VOLUME 119, NUMBER 34
AUGUST 27, 1997

JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

NMR Characterization of Hairpin Polyamide Complexes with the Minor Groove of DNA

Rafael Peláez Lamamie de Clairac,[†] Bernhard H. Geierstanger,[†] Milan Mrksich,[‡] Peter B. Dervan,[‡] and David E. Wemmer*,[†]

Contribution from the Department of Chemistry, University of California, Berkeley, California 94720, and Arnold and Mabel Beckman Laboratories of Chemical Synthesis, California Institute of Technology, Pasadena, California 91125

Received December 16, 1996[⊗]

Abstract: Polyamides containing *N*-methylimidazole (Im) and *N*-methylpyrrole (Py) amino acids can be combined in antiparallel side-by-side dimeric complexes for sequence-specific recognition in the minor groove of DNA. Covalently linking polyamide subunits has led to designed ligands with both increased affinity and specificity. Simple aliphatic amino acid linkers serve as internal guide residues for turn vs extended binding in a head-to-tail-linked polyamide motif. Polyamides of sequence composition ImPyPy-X-PyPyPy containing linkers of incremental length (X = 3-aminopropionic acid (β), 4-aminobutyric acid (γ), or 5-aminovaleric acid (δ)) in complex with an undecamer DNA duplex containing a 5′-(A,T)G(A,T)$_3$-3′ target site were structurally characterized using NMR spectroscopy. Previous quantitative DNase I footprinting studies identified γ as the highest affinity of these "turn" linkers. NMR titrations and 2D NOESY data combined with restrained molecular modeling reveal that polyamides with β, γ, and δ linkers all may adopt a hairpin structure. Modeling supports the idea that the linkers in the β and δ complexes adopt an energetically less favorable turn geometry than the γ linker and confirms that the three-carbon γ linker is sufficient and optimal for the hairpin conformation.

Introduction

A dimeric arrangement of pyrrole−imidazole polyamides in the minor groove has developed into a general model for the sequence-specific recognition of DNA.[1−5] Two ligands containing *N*-methylpyrrole (Py) and *N*-methylimidazole (Im) amino acids simultaneously occupy the DNA minor groove (Figure 1). Polyamides are bound antiparallel to one another, each making contact to its adjacent DNA strand. The DNA-binding sequence specificity of these small molecules depends on the sequence of side-by-side amino acid pairings. A pairing of imidazole opposite pyrrole recognizes a G·C base pair, while a pyrrole/imidazole combination targets C·G.[2−6] Pyrrole/pyrrole

[†] University of California, Berkeley.
[‡] California Institute of Technology.
[⊗] Abstract published in *Advance ACS Abstracts*, August 15, 1997.

(1) (a) Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5723−5727. (b) Pelton, J. G.; Wemmer, D. E. *J. Am. Chem. Soc.* **1990**, *112*, 1393−1399. (c) Fagan, P. A.; Wemmer, D. E. *J. Am. Chem. Soc.* **1992**, *114*, 1080−1081.

(2) (a) Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 8783−8794. (b) Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 7586−7590. (c) Wade, W. S.; Mrksich, M.; Dervan, P. B. *Biochemistry* **1993**, *32*, 11385−11389.

(3) (a) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 2572−2576. (b) Geierstanger, B. H.; Jacobsen, J. P.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Biochemistry* **1994**, *33*, 3055−3062.

(4) Geierstanger, B. H.; Dwyer, T. J.; Bathini, Y.; Lown, J. W.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, *115*, 4474−4482.

(5) (a) Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Science* **1994**, *266*, 646−650. (b) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1995**, *117*, 3325−3332.

(6) (a) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 9892−9899. (b) Dwyer, T. J.; Geierstanger, B. H.; Mrksich, M.; Derrvan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, *115*, 9900−9906. (c) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 3663−3664.

S0002-7863(96)04314-4 CCC: $14.00

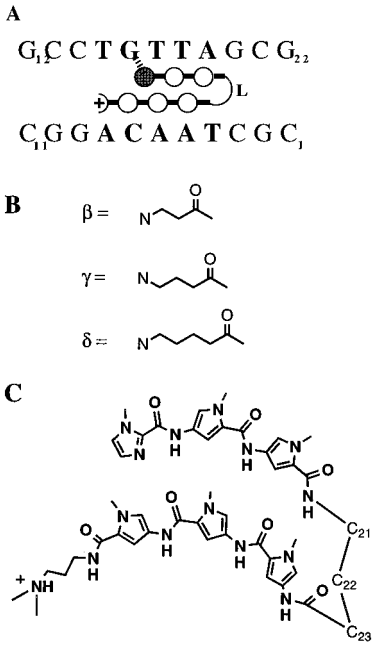

Figure 1. (A) Schematic representation of the antiparallel side-by-side DNA binding motif. Open circles represent *N*-methylpyrrole units, while shaded circles indicate *N*-methylimidazole rings. (B) Structures of the linkers studied, abbreviations to the left. (C) Structure of the head-to-tail "hairpin" polyamides.

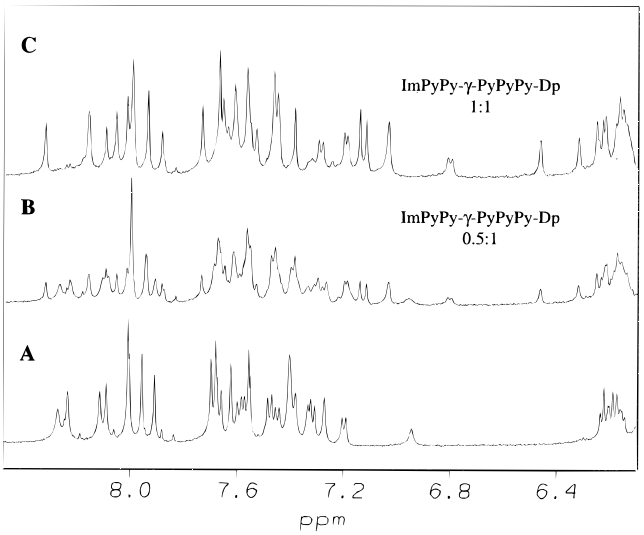

Figure 2. Aromatic region of ^{1}H NMR spectra of d(CGCTAACAGGC)·d(GCCTGTTAGCG) in complex with ImPyPy-γ-PyPyPy-Dp (B and C). Free DNA is shown in panel A. The molar ligand to DNA duplex ratios are indicated for each spectrum.

pairs bind to either A·T or T·A rendering the recognition code partially degenerate.[1−4]

Different polyamides can be combined in heterodimeric complexes allowing the recognition of sequences not accessible to homodimers.[3,4] Accordingly, the polyamide ImPyPy was combined with the polyamide PyPyPy (distamycin) and shown to recognize 5′-(A,T)G(A,T)$_3$-3′ sequences.[3] Linking the central pyrrole rings via methylene chains increased the binding specificity and affinity of the heterodimer.[6] A more general synthetic methodology has recently been employed to link the N-terminus of PyPyPy and the C-terminus of ImPyPy.[7] The amino acids 3-aminoproprionic acid [β-alanine (β)], 4-aminobutyric acid (γ) and 5-aminovaleric acid (δ) were investigated as potential turn-forming linkers (Figure 1). The arrangement of the ring systems in the complex should be a conserved feature when the PyPyPy and the ImPyPy moieties fold back onto themselves to form a "hairpin" structure. The resulting side-by-side combination of the imidazole group of the ImPyPy and a pyrrole ring of the PyPyPy moiety allows for specific interactions with the G·C base pair in the 5′-(A,T)G(A,T)$_3$-3′ binding sites (Figure 1). A (dimethylamino)propyl tail (-Dp) is added to approximate the propylamidine tail of distamycin.

Quantitative DNase I footprinting experiments revealed that ImPyPy-γ-PyPyPy-Dp binds a 5′-TGTTA-3′ target site with an equilibrium association constant of 7.6×10^7 M^{-1}, an affinity enhancement of at least 2 orders of magnitude relative to the unlinked monomers ImPyPy and PyPyPy.[7] ImPyPy-γ-PyPyPy-Dp binds the designated matched site, 5′-TGTTA-3′, with an association constant 24- and 100-fold greater than the single base pair mismatched sites 5′-TGACA-3′ and 5′-TTTTT-3′, respectively.[7] Affinity cleaving studies on the γ-linked polyamide are consistent with monomeric binding in a single

Table 1. Ligand−DNA and Intraligand NOE Contacts for the ImPyPy-γ-PyPyPy-Dp Complex with d(CGCTAACAGGC)·d(GCCTGTTAGCG)[a]

PyPyPy moiety	DNA	ImPyPy moiety
	DNA to Ligand	
$N(CH_3)_2$	T15 C1′	
1H3-3 1NH-4 C18H1[b] C18H2[b] C19H1[b] C19H2[b]	A8 C2H	
1H3-3 1NH-4	A8 C1′H	
1H3-3 1H5-3	A8 C4′H	
	G16 C1′H	2H4-1
$N(CH_3)_2$	G16 C4′H	
$N(CH_3)_2$	G16 C5′H[b]	
1NH-3 1H3-3 1NH-4	G16 NH2 H21/22[b]	2NH-1
1H3-2 1NH-3	C7 C1′H	
1H3-2 1H5-2	C7 C4′H	
	T17 C1′H	2NH-1
	T17 C4′H	2H4-1 2H5-1 2N1-CH_3
1H3-1 1NH-2 1H3-2 1NH-3	A6 C2H	2NH-1 2H3-2 2NH-2
1H3-1 1NH-2	A6 C1′H	
1H3-1 1H5-1 1N1-CH_3	A6 C4′H	
	T18 C1′H	2H3-2 2NH-2
	T18 C4′H	2H3-2 2H5-2
1NH-1 1H3-1	A5 C2H	2NH-2 2H3-3 2NH-3 C22H1
1NH-1	A5 C1′H	
	A19 C2H	2H3-3 2NH-3 C21H1 C21H2 C22H1 C22H2 C23H1 C23H2
	A19 C1′H	2NH-2 2H3-3 2NH-3
	A19 C4′H	2H3-3 2H5-3
	PyPyPy to ImPyPy	
C19H1[b] C19H2[b] C20H1[b] $N(CH_3)_2$ 1H3-3		2H4-1
C18H1 1N3-CH_3		2H5-1
1N3-CH_3 1H5-3		2N1-CH_3
1H3-2		2H3-2
1H5-2		2N2-CH_3
1N2-CH_3		2N2-CH_3 2H5-2
1H3-1		2H3-3
1N1-CH_3		2H5-3 2N3-CH_3
1H5-1		2N3-CH_3
1NH-1		C22H1 C22H2 C23H1 C23H2

[a] Identified in the H_2O NOESY acquired at 100 ms mixing time. [b] Protons not stereo-specifically assigned.

(7) Mrksich, M.; Parks, M. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 7983−7988.

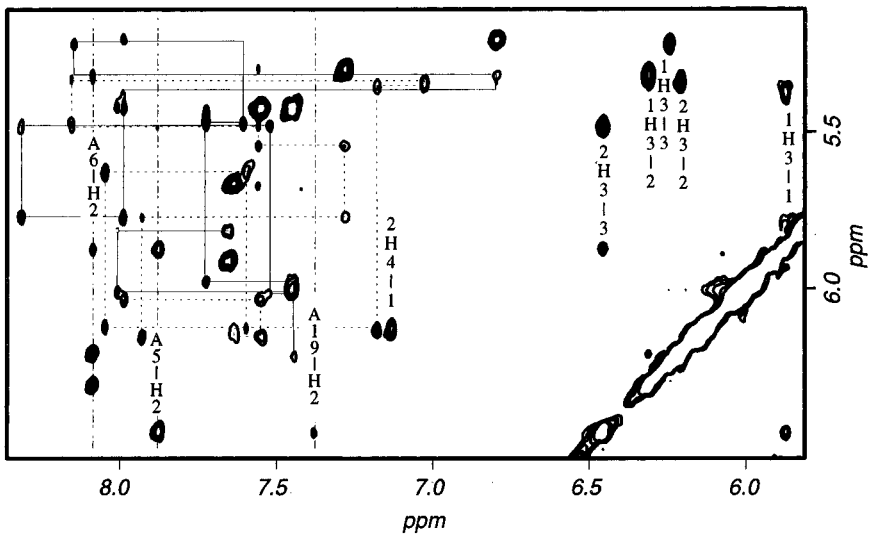

Figure 3. Expansion of the aromatic and H1′ region of NOESY spectra (in D_2O, 500 MHz, 25 °C, τ_{mix} = 200 ms) of d(CGCTAACAGGC)·d(GCCTGTTAGCG) in complex with ImPyPy-γ-PyPyPy-Dp Sequential aromatic to C1′H connectivities for the 5′-TAACA-3′ strand are shown as solid lines; those for the 5′-TGTTA-3′ strand are shown as dashed lines. Labels below or above a cross peak denote the chemical shift along the ω_2−(horizontal) axis while labels to the left or right of a peak indicate the chemical shift along the ω_1−(vertical) axis. Ligand protons are labeled according to Figure 5.

orientation as predicted by the "hairpin" model.[8] The β- and δ-linked polyamides, ImPyPy-β-PyPyPy-Dp and ImPyPy-δ-PyPyPy-Dp, bind all three sites with association constants of less than 2×10^6 M^{-1}, leaving some question as to whether they bind in the same mode.[7] Here we employ nuclear magnetic resonance and molecular modeling to directly characterize of the polyamides ImPyPy-X-PyPyPy-Dp complexed with a 5′-TGTTA-3′ target site.

Materials and Methods

Sample Preparation. NMR samples contained 10 mM sodium phosphate buffer (pH 7.0) in 0.5 mL 99.96% D_2O (Cambridge Isotope Laboratories) or a 90% H_2O/10% D_2O mixture. Ligand stock solutions were prepared in 99.96% D_2O from the solid HCl salts and stored at −70 °C. Their concentrations were 10 mM as determined by UV absorbance at 306 nm ($\epsilon \approx 7 \times 10^4$ M^{-1} cm^{-1}). DNA samples were 1 mM duplex as determined by UV absorbance at 80 °C [$\epsilon_{260} = 1.04 \times 10^5$ M^{-1} cm^{-1} for d(CGCTAACAGGC) and 1.01×10^5 M^{-1} cm^{-1} for d(GCCTGTTAGCG)].[9]

NMR Experiments and Signal Assignments. NMR experiments were performed at 600 MHz on a Bruker AMX-600 or at 500 MHz on a General Electric GN-Omega spectrometer. Ligands were titrated into the NMR sample containing duplex DNA in approximately 0.2 mol equiv per addition. 1D spectra in D_2O (averaged 128 scans) were acquired with 4096 complex points over a spectral width of 5000 Hz (500 MHz). NOESY spectra in D_2O (200 ms mixing time) were collected with 1024 complex points in t_2 using a spectral width of 5000 Hz (500 MHz). A total of 424−505 t_1 experiments with 48−64 scans were recorded and zero-filled to 1 K. For experiments in D_2O presaturation pulses were applied during the recycle delay (2 s) and the mixing period to suppress the residual HDO resonance. For the ImPyPy-γ-PyPyPy-Dp complex DQF-COSY (512 t_1 experiments, 48 scans) and TOCSY (470−502 t_1 experiments, 64 scans) spectra with mixing times of 40, 70, and 90 ms were similarly collected in D_2O. A TOCSY spectrum for the ImPyPy-β-PyPyPy-Dp complex (60 ms mixing time) was acquired with 498 t_1 experiments (64 scans). NOESY spectra in water were acquired at 25 °C, replacing the last 90° pulse by a 1−1 jump and return sequence for solvent suppression as described previously.[2b] The spectra were collected into 2048 complex points in t_2 using a spectral width of 13514 Hz at 600 MHz. A total of 485−509 t_1 experiments (48−64 scans) were recorded for NOESY spectra of ImPyPy-γ-PyPyPy-Dp (100 and 200 ms mixing time), theImPyPy-β-PyPyPy-Dp, and ImPyPy-δ-PyPyPy-Dp complexes (only 200 ms mixing time) and zero-filled to 2 K. All 2D spectra were acquired using TPPI. The data were processed with FELIX (version 2.30β, Biosym, San Diego) on Silicon Graphics workstations. Skewed sine bell functions were used for apodization of the free induction decays. DNA and ligand resonances were assigned using standard sequential methods,[10] and as previously described.[1] NOE contacts between C4′H DNA protons and H3, H5, and *N*-methyl pyrrole protons were assigned in analogy to the contacts observed in complexes previously characterized.[5a]

Molecular Modeling Using Restrained Energy Minimization. A molecular model of the ImPyPy-γ-PyPyPy-Dp complex with d(CGCTAACAGGC)·d(GCCTGTTAGCG) was generated using the Biosym molecular modeling package InsightII (Biosym, San Diego) running on Silicon Graphics work stations. The DNA model was constructed using the Biopolymer module of InsightII from standard B-form DNA, which is consistent with the NMR data for the three complexes. Coordinates for the ImPyPy-γ-PyPyPy-Dp molecule were derived from the model of the heterodimeric complex of ImPyPy and PyPyPy with d(GCCTAACAAGG)·d(CCTTGTTAGGC).[3b] The ligand molecules were linked and modified using the Builder module. Energy minimizations using Discover (with the AMBER forcefield) were performed with the ring systems of the ImPyPy-γ-PyPyPy-Dp in an antiparallel hairpin structure. Partial atomic charges were then calculated using MOPAC (AM1). The preformed hairpin polyamide was roughly oriented in the 5′-TGTTA-3′·5′-TAACA-3′ binding site. Restrained energy minimizations (Discover) were performed on the complex, as described previously.[3b] A total of 53 intermolecular ligand−DNA and 21 intramolecular ligand restraints were derived from NOESY data at 100 ms mixing time. NOE cross peaks were classified semiquantitatively into three categories: strong (1.8−2.5 Å), medium (2.5−3.7 Å) or weak (3.7−5.0 Å) relative to the volume integrals of cytosine H5-H6 cross peaks. (Listings of the intermolecular ligand-DNA and intramolecular ligand restraints are available as Supporting Information). The model fulfills all restraints to within 0.1 Å. The complexes of ImPyPy-β-PyPyPy-Dp, and ImPyPy-δ-PyPyPy-Dp with d(CGCTAACAGGC)·d(GCCTGTTAGCG) were modeled similarly with the use of restraints derived for the ImPyPy-γ-PyPyPy-Dp complex but excluding restraints involving the linker regions. Energy minimization was combined with molecular dynamics at temperatures of up to 600 K to relax the linker conformation and ligand arrangement.

(8) White, S.; Baird, E. E.; Dervan, P. B. *Biochemistry* **1996**, *35*, 12532−12537.

(9) Warshaw, M.; Cantor, C. *Biopolymers* **1970**, *9*, 1079−1103.

(10) (a) Hare, D.; Wemmer, D. E.; Chou, S.-H.; Drobny, G.; Reid, B. R. *J. Mol. Biol.* **1983**, *171*, 319−336. (b) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; John Wiley & Sons, New York, 1986.

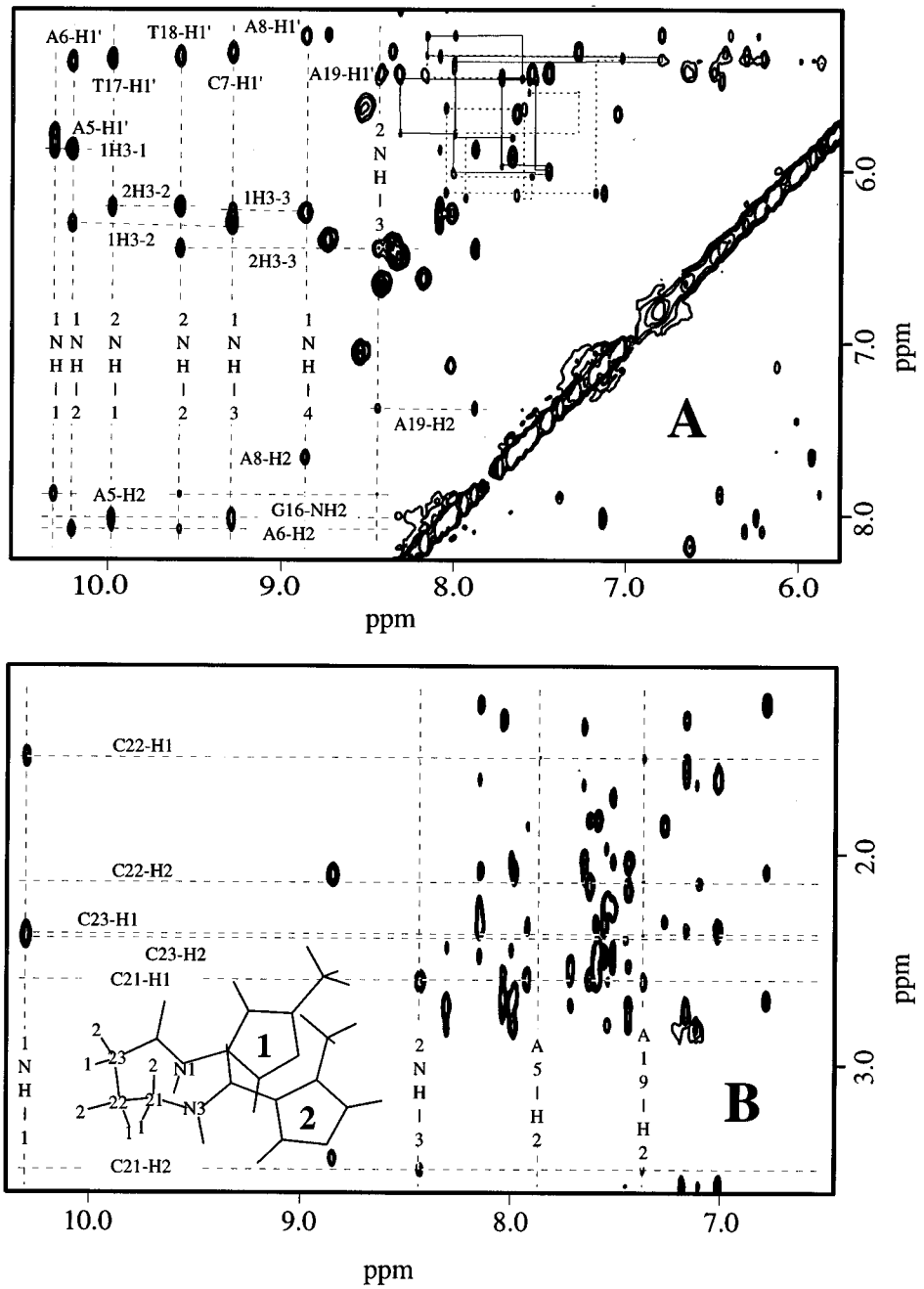

Figure 4. Expansions of NOESY spectra of the ImPyPy-γ-PyPyPy-Dp complex with d(CGCTAACAGGC)·d(GCCTGTTAGCG) (in 90% H_2O/10% D_2O, 25 °C). (A) Contacts between ligand amides and pyrrole protons with DNA protons in the minor groove (τ_{mix} = 200 ms): Sequential aromatic to C1′H connectivities for the 5′-TAACA-3′ strand are shown as solid lines; those for the 5′-TGTTA-3′ strand are shown as dashed lines. (B) Contacts involving linker protons (τ_{mix} = 100 ms): Labels below or above a cross peaks identify the proton along the ω_2−(horizontal) axis while labels to the left or right of a peak identify the proton along the ω_1−(vertical) axis. The protons C21H1, C22H1, and C23H2 are the proR protons, and C21H2, C22H2, and C23H1 are correspondingly the proS. Ligand protons are labeled according to Figure 5.

Results

NMR Titrations. One-dimensional NMR spectra were recorded at 25 °C during titrations of d(CGCTAACAGGC)·d(GCCTGTTAGCG) with ImPyPy-β-PyPyPy-Dp, ImPyPy-γ-PyPyPy-Dp, and ImPyPy-δ-PyPyPy-Dp (Figure 2 and Supporting Information). Each polyamide forms a well-defined complex with a 1:1 ligand/duplex stoichiometry. All three polyamides dissociate slowly on the NMR time scale from the complex as indicated by two sets of DNA resonance lines for complexed and free at substoichiometric amounts of ligand (Figure 2B). A second minor complex was detected in the binding of ImPyPy-β-PyPyPy-Dp to d(CGCTAACAGGC)·d(GCCTGTTAGCG) (Supporting Information).

Characterization of the ImPyPy-γ-PyPyPy-Dp Complex. NOESY data acquired in D_2O and H_2O (Figures 3 and 4, Table 1) confirm the formation of a "hairpin" structure for the ImPyPy-γ-PyPyPy-Dp polyamide when bound to d(CGCTAACAGGC)·d(GCCTGTTAGCG). NOE contacts between ligand protons and DNA sugar and adenine C2 protons indicate the polyamide is bound in the minor groove (Table 1 and Figure 5). The ImPyPy moiety lies against residues 5′-G16-T17-T18-A19-3′ and the PyPyPy moiety contacts the opposite strand spanning 5′-AACA-3′. The "hairpin" structure of ImPyPy-γ-PyPyPy-Dp is further confirmed by intramolecular NOEs between the H4-1 and H5-1 protons on the imidazole ring of ImPyPy and the methylene protons of the positively charged tail group (Dp). Weak NOE

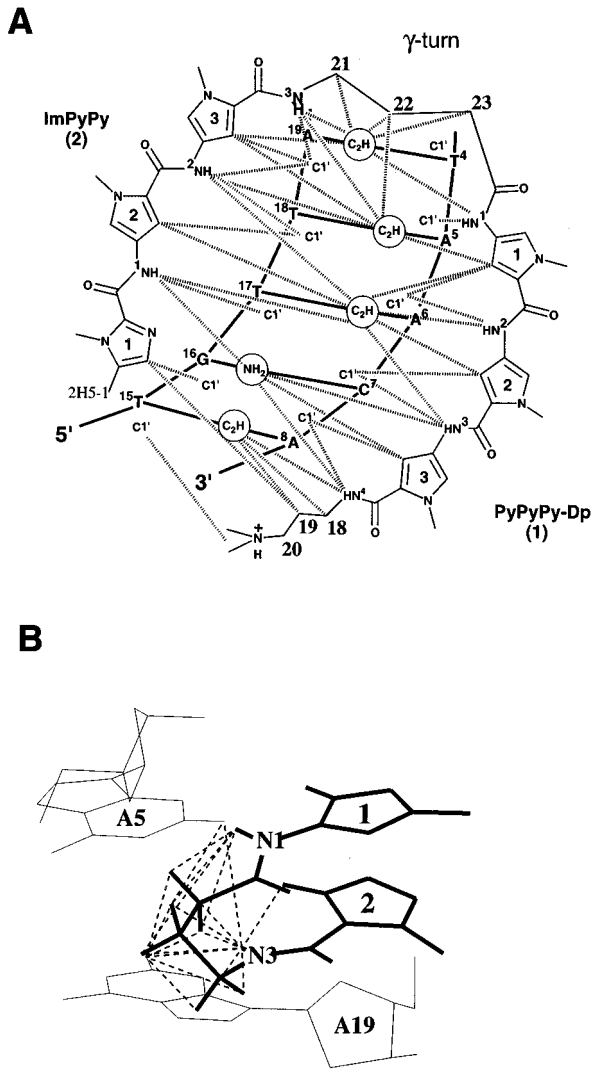

Figure 5. Schematic representation of the NOE data obtained for the ImPyPy-γ-PyPyPy-Dp complex with d(CGCTAACAGGC)·d-(GCCTGTTAGCG). (A) Selected intermolecular ligand−DNA NOEs: The PyPyPy and ImPyPy moieties are referred to as 1 and 2 respectively. The numbering system adopted is indicated by the superscripts of the amide nitrogens and the numbers in the rings. The nomenclature used includes the moiety number, the type of atoms, and their numbers and the corresponding number of the ring or amide (2H5-1 is indicated). (B) NOEs that define the linker conformation of ImPyPy-γ-PyPyPy-Dp in the complex.

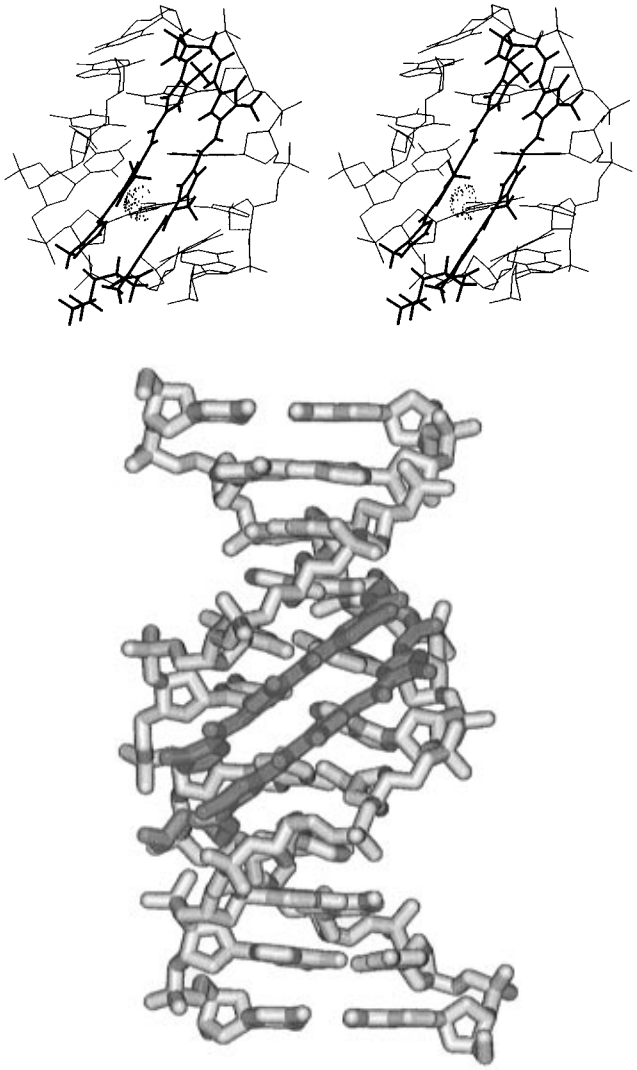

Figure 6. (A, top) Ligand binding region of the molecular model of ImPyPy-γ-PyPyPy-Dp in complex with d(CGCTAACAGGC)·d-(GCCTGTTAGCG) obtained by energy minimization using semiquantitative distance restraints derived from NOESY data. Stereorepresentation of the complex: For clarity, hydrogen atoms are omitted for the DNA model but not for the ligand molecules. Only the five central base pairs are shown. The ligand is shown in thick lines. The guanine amino group recognized by the imidazole nitrogen of ImPyPy is highlighted as van der Waals surface. (B, bottom) Stick drawing of the full complex shown expanded in A.

cross peaks are observed between the protons 1H3-2 and 2H3-2 of the two central pyrrole rings as well as between 2H3-3 and 1H3-1. Contacts between the H5 and methyl groups of the pyrrole and imidazole rings to the corresponding protons on the stacked ring further confirm the side-by-side arrangement.

Consistent with the pairing rules, the imidazole ring of ImPyPy-γ-PyPyPy-Dp recognizes the guanine amino group in the 5′-(A,T)G(A,T)$_3$-3′ target site. The formation of a *specific hydrogen bond* between the imidazole nitrogen of the ImPyPy moiety and the guanine amino group of G16 is indicated by NOE cross peaks to these amino protons and by the downfield shift of the proton not involved in Watson−Crick hydrogen bonding (Figure 4A). In this particular complex, the chemical shift values of the two amino protons are accidentally degenerate. The remaining NOE contacts of the ImPyPy-γ-PyPyPy-Dp complex are analogous to those observed for the ImPyPy-Dp·Distamycin heterocomplex with 5′-(A,T)G(A,T)$_3$-3′ sequences, indicating that the antiparallel polyamide arrangement is conserved.[3]

On the basis of sequential intramolecular NOEs to neighboring pyrrole H3 protons and on NOE contacts to sugar C1′ and adenine C2 protons, the ligand amide protons (including those on the linker 1NH-1 and 2NH-3) must point into the DNA minor groove (Figure 4A). Resonances of the methylene protons of the linker were stereospecifically assigned on the basis of the cross peak intensities in the 100 ms H_2O NOESY (Figures 4B and 5B). The C21 proton pair was identified on the basis of the characteristic chemical shift values due to proximity of the amide nitrogen. 2NH-3 has a very strong cross peak to C21H1 that resonates at 2.6 ppm while a weaker cross peak is observed to the C21H2 proton at 3.5 ppm. Neither proton shows a NOE cross peak to the amide 1NH-1 at the other end of the linker. A very strong cross peak to 1NH-1 identifies the proton at 2.5 ppm as C23H1, overlapping with the geminally paired C23H2 proton. The remaining methylene proton pair (C22) has resonances at 1.6 and 2.2 ppm. The proton C22H1 at 1.6 ppm is close in space to 1NH-1 while also showing a weak NOE cross peak to 2NH-3. C22H2 exhibits a weak NOE cross peak

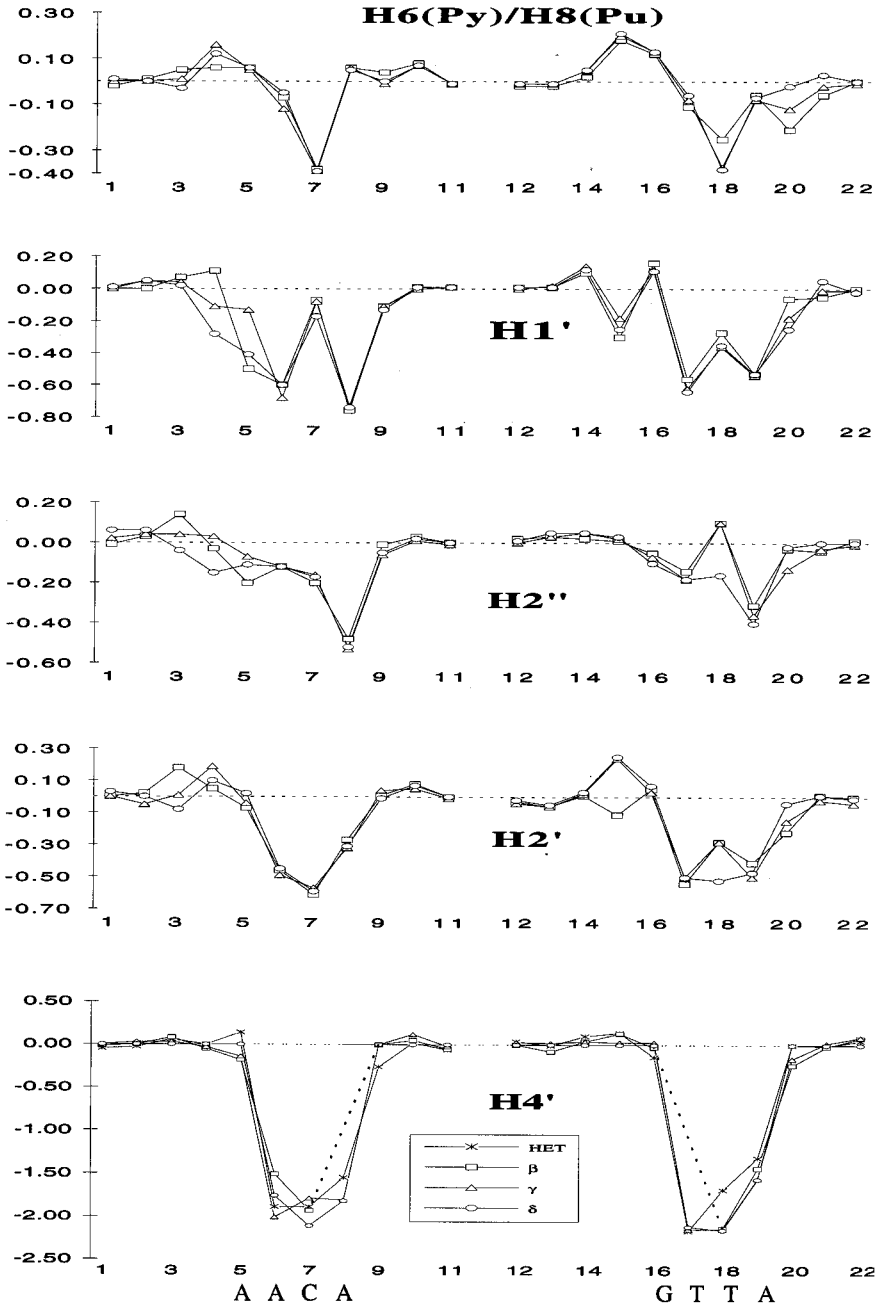

Figure 7. Chemical shift changes of representative DNA protons induced by complex formation with the three hairpin polyamides ($\beta = \square$, $\gamma = \triangle$, $\delta = \bigcirc$) and with the unlinked Distamycin:ImPyPy-Dp heterodimer (*, only for the H4′ protons) shown as difference in ppm between complexed and free DNA versus the number of the base they belong to (numbering according to Figure 1). The H2′/H2″ were not stereospecifically assigned, and the most upfield proton was always considered to be the H2′. Dashed lines are used to connect two nonsequential points if the corresponding values for the intermediate one(s) is(are) not known.

to 1NH-1 but none to 2NH-3. The linker conformation is further defined by intermolecular NOE contacts to the DNA. C21H1 contacts A19 C2H and G20 C1′H (identified in 200 ms D_2O NOESY) while C21H2 shows only a weak NOE to A19 C2H. C22H1 has a cross peak to A19 C2H and a weaker one to A5 C2H, while C22H2 exhibits only a very weak one to A19 C2H. Only a weak NOE peak between A19 C2H and the C23 proton pair is observed. The proximity of C23 to nucleotide A5 is indicated by an NOE peak to A5 C1′H identified in the 200 ms D_2O NOESY (NOE could not be identified in the H_2O NOESY spectra due to lower resolution and reduced intensities because of the excitation profile of the spectra acquired in H_2O). Restrained molecular modeling of the ImPyPy-γ-PyPyPy-Dp complex yielded the linker conformation shown in Figures 5B and 6.

Molecular Modeling of the "Hairpin" Complexes. A molecular model of the ImPyPy-γ-PyPyPy-Dp complex with d(CGCTAACAGGC)·d(GCCTGTTAGCG) was generated by restrained energy minimization using semiquantitative distance restraints derived from NOESY data acquired at 100 ms mixing time (Figure 6). The "hairpin" complex of ImPyPy-γ-PyPyPy-Dp is similar to previously characterized dimeric polyamide complexes. Intermolecular hydrogen bonds were assigned by the InsightII program for each amide proton on the ligand to the N3 of adenines or the O2 of pyrimidine bases which face into the minor groove. The imidazole nitrogen is also well positioned for a hydrogen bond to the non-Watson−Crick-paired amino proton of guanine G16 (Figure 6). The γ linker sits deep in the minor groove at the 3′ end of the 5′-TGTTA-3′ binding site. The amide bonds remain planar with the amino proton

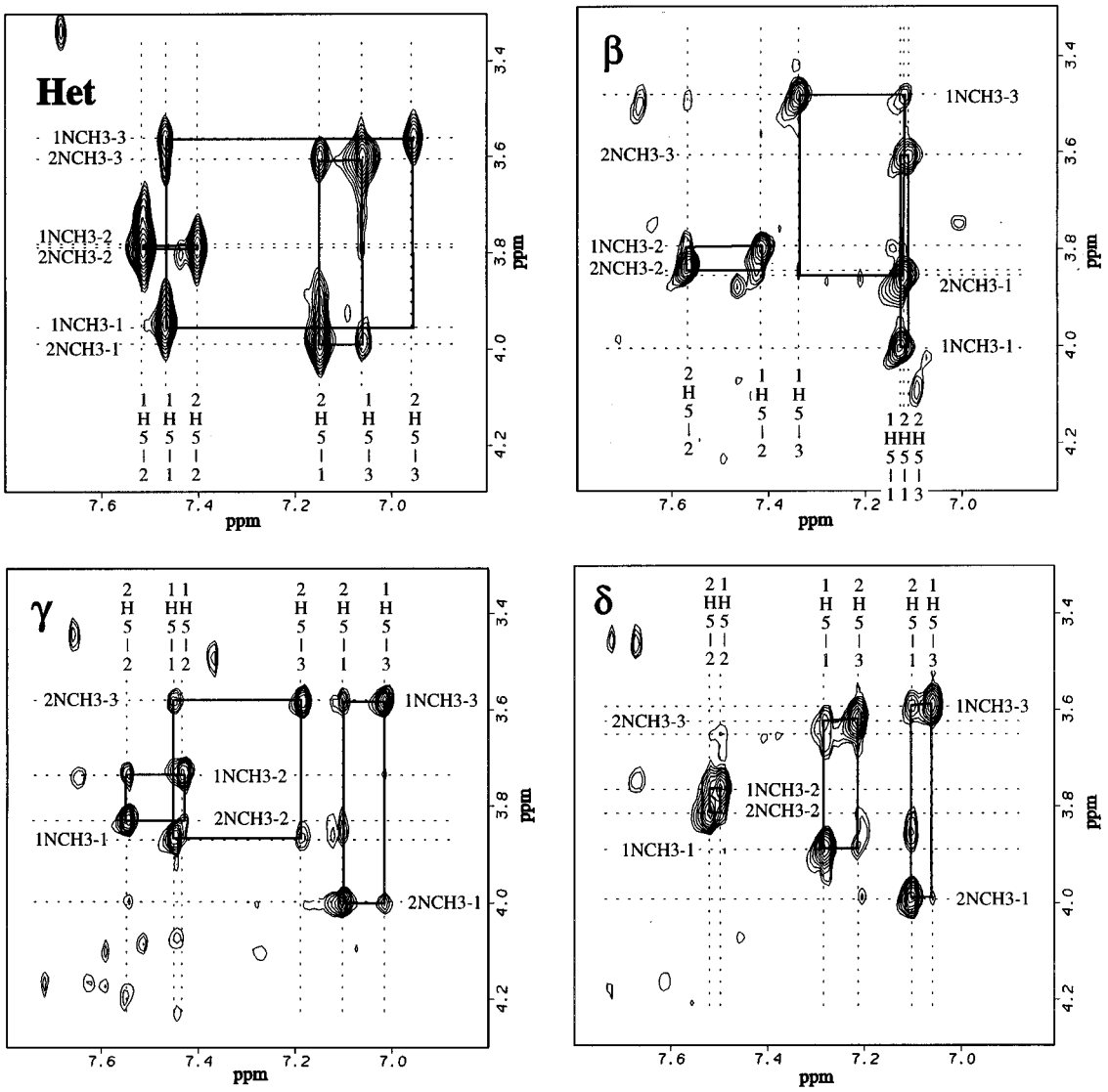

Figure 8. Pyrrole-H5 to *N*-methyl regions of the D_2O NOESYS at $\tau_m = 200$ ms for the unlinked Distamycin:ImPyPy-Dp heterodimeric complex (Het), the β, γ, and δ complexes. Labels below or above a cross peaks identify the proton along the ω_2-(horizontal) axis while labels to the left or right of a peak identify the proton along the ω_1-(vertical) axis. Ligand protons are labeled according to Figure 5.

pointing into the minor groove suggesting that *the turn is formed without significant conformational strain*.

Identical intermolecular NOE contacts (Table 1 and Supporting Information) and very similar chemical shift values (Figure 7) indicate similar structures for the ImPyPy-γ-PyPyPy-Dp, the ImPyPy-β-PyPyPy-Dp, and the ImPyPy-δ-PyPyPy-Dp complexes. The pyrrole H5 proton to *N*-methyl region of the NOESY spectrum of the three complexes and of the "unstrained" distamycin:ImPyPy-Dp complex[3] also show very similar NOE patterns, indicating a conserved arrangement of the ring systems (Figure 8). We therefore used the restraints derived for ImPyPy-γ-PyPyPy-Dp to model the other two complexes. In both cases the linkers were not restrained. A comparison of the complexed ligands indicates that all three linkers can adopt a turn conformation, without disturbing the stacking of the rings, that is characteristic for the antiparallel side-by-side motif. In order to accommodate the turn the methylenes of the linkers can no longer adopt a staggered conformation. Geometrically the β-linker is too short to accommodate the turn without bending the flanking amide bonds out of the pyrrole ring plane. In contrast to the γ linker, the δ linker is too long to avoid steric clashes without disrupting the stacking of the rings. The experimental data show that the strain is accommodated in the linker and its interaction with DNA rather than through shifting the pyrrole/imidazole ring positions. The footprinting results, with the NMR data indicating similar binding modes for the three "hairpin" polyamides, indicate that the *energy of turn formation is minimal for the* γ linker.

Discussion

Consistent with previous quantitative footprint titration results,[7] NMR titrations verify that the polyamide ImPyPy-γ-PyPyPy-Dp binds with a 1:1 stoichiometry to the minor groove of d(CGCTAACAGGC)·d(GCCTGTTAGCG). Confirmation of the "hairpin" structure in the complex is clearly indicated by inter- and intramolecular NOEs. The ImPyPy moiety contacts the 5′-GTTA-3′ strand in the minor groove while the PyPyPy portion of the polyamide contacts the opposite minor groove strand, 5′-AACA-3′. For a related system, it has been proposed that linkers shorter than four methylene units cannot accommodate the turn and thus necessarily form intermolecular dimers.[11] In the hairpin series extensive intra- and intermolecular NOE contacts unambiguously place the linkers deep in

the minor groove and forming a tight turn, even for the short, β linker. The observed complex is consistent with the previously determined characteristics of the side-by-side antiparallel motif of these minor groove binders, namely (a) extensive hydrogen bonding between ligand amide protons and acceptor groups on the DNA bases, (b) good van der Waals contacts between the stacked ligands and the wall of the minor groove, and (c) positively charged tail groups that are positioned deep in the minor groove and when placed at the carboxy terminus point toward the 3′ end of the contacting DNA strand (Figure 6). There is some local distortion of the DNA evident in the model, including buckle and tilt of base pairs in the binding site. However the analysis of the NMR data was only semiquantitative, and extensive sampling of structures was thus not warranted. While discussion of details of the distortions induced by binding await further NMR data collection and analysis, Figure 6b shows that gross distortion of the structure does not occur.

The DNA sequence selectivity of this class of ligands relies on the specific recognition of G by the imidazole ring. NOESY data and molecular modeling indicate formation of a specific hydrogen bond between the imidazole nitrogen and the G16 amino proton which is not involved in base pairing. Also, the linker moiety sits deep in the groove, making van der Waals contacts with the walls and the bottom of the groove. The A,T preference of the linker shown in the footprinting studies[7] (γ binds TGACA with a $K_{app} \sim 3.2 \times 10^6$ but does not protect significantly a TGACG site) suggests avoidance of the protruding amino group of guanine, reminiscent of the A,T preference of distamycin and consistent with the observed position of the linker in the groove. The head-to-tail linkage also enforces bidentate binding to DNA, reducing monodentate binding of the PyPyPy segment to narrow TTTTT sites. Previous studies on dimers formed by linkage of the central pyrrole rings have shown that the monodentate binding was not greatly disfavored by this linkage, the sequence specificity (<30-fold) arising from enhanced binding in the bidentate motif.[6a] For the hairpin polyamides, the selectivity (~100-fold) seems to be further increased by reducing the affinity of the monodentate binding[7] (however, binding in alternative motifs is still possible[12]).

Molecular modeling using semiquantitative restraints derived from NOESY data indicates that the characteristic side-by-side ligand arrangement is preserved in the ImPyPy-γ-PyPyPy-Dp, ImPyPy-β-PyPyPy-Dp, and ImPyPy-δ-PyPyPy-Dp complexes. Modeling supports previous footprinting results which indicate that the linker in the δ and the β complexes adopts an energetically less favorable geometry than the γ linker. β is too short to make the turn easily, thus requiring energetically unfavorable twisting of the amides connecting the linker to the ring system. With δ, on the other hand, it appears that unfavorable contacts with the DNA cause the decrease in binding affinity. Footprinting data[7] and NMR restrained modeling suggest that the three-carbon linker of the ImPyPy-γ-PyPyPy-Dp polyamide is sufficient and optimal for the "hairpin" design.

Acknowledgment. We are grateful to National Institutes of Health (GM-43129 D.E.W. and GM-27681 P.B.D.) and the National Foundation for Cancer Research for research support, to the Ralph M. Parson Foundation for a graduate fellowship to M.M., and to the U.S. Department of Energy, National Science Foundation and Keck Foundation for instrumentation grants. R.P.L.deC. gratefully acknowledges a postdoctoral fellowship from the Fundación Ramón Areces. R.P.L.deC., B.H.G., and D.E.W. also thank T. D. Cole for helpful discussions and assistance in the early stages of the project.

Supporting Information Available: Listings of the intermolecular ligand−DNA and ligand−ligand restraints and the achieved distances, and chemical shift assignments for the DNA and the ligands in the ImPyPy-γ-PyPyPy-Dp, ImPyPy-δ-PyPyPy-Dp and ImPyPy-β-PyPyPy-Dp complexes (12 pages). See any current masthead page for ordering and Internet access instructions. The coordinates for the model are available from D.E.W. upon request.

JA964314R

(11) (a) Chen, Y.-H.; Lown, J. W. *J. Am. Chem. Soc.*, **1994**, *116*, 6995−7005. (b) Chen, Y.-H.; Lown, J. W. *Biophys. J.* **1995**, *68*, 2041−2048.

(12) Trauger, J. W.; Baird, E. E.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *116*, 6160−6166.