NMR Characterization of the Aliphatic β/β Pairing for Recognition of A·T/T·A Base Pairs in the Minor Groove of DNA

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

Contribution from the Department of Chemistry, University of California, Berkeley, California 94720-1460, and Arnold and Mabel Beckman Laboratories of Chemical Synthesis, California Institute of Technology, Pasadena, California 91125

Received October 19, 1998. Revised Manuscript Received February 1, 1999

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. Because the curvature of four or five contiguous Im−Py rings does not perfectly match the canonical B-helix, β-alanine (β) residues have been inserted to reset the register. Complexes of three pyrrole−imidazole polyamides of sequence composition ImPyPy-X-PyPyPy-Dp, where X = Py, β, or glycine (G), bound to a 13 base pair DNA duplex containing a 9 base pair 5′-TGTATATCA-3′ match site were characterized by NMR. NMR titrations and NOESY data combined with restrained molecular modeling show that each polyamide adopts an extended antiparallel dimeric conformation with the ligands fully overlapped around a central Py/Py, G/G, or β/β pair. Conformational exchange is seen near the linker for the G-linked complex, but not with the β or Py linkers. In addition to providing the first direct structural evidence for formation of the aliphatic β/β pairing in the minor groove, models support the idea that the β linker of ImPyPy-β-PyPyPy-Dp provides an optimal combination of size, flexibility, and alignment of the polyamide-paired aromatic subunits in extended, dimeric 2:1 complexes.

Introduction

Polyamides containing *N*-methylpyrrole (Py), *N*-methylimidazole (Im), and *N*-methyl-3-hydroxypyrrole (Hp) amino acids are synthetic ligands that have an affinity and a specificity for DNA comparable to those of naturally occurring DNA-binding proteins.[1] DNA recognition depends on side-by-side ring pairings in the minor groove.[2−5] A pairing of Im opposite Py targets a G·C base pair, while a Py/Im pair targets C·G.[2] A Py/Py combination is degenerate, targeting both T·A and A·T base pairs.[2,3] An Hp/Py pair discriminates T·A from A·T and both from G·C/C·G.[5] The generality of these pairing rules has been demonstrated by their success in designing ligands that recognize a variety of sequences[1−11] and is supported directly

[†] University of California.

[‡] California Institute of Technology.

(1) For subnanomolar binding, see: (a) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1996**, *382*, 559−561. (b) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 6953−6961. (c) Turner, J. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 7636−7644. (d) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Angew. Chem., Int. Ed. Engl.* **1998**, *37*, 1421−1423. (e) Turner, J. M.; Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 6219−6226.

(2) For specificity of Im/Py pairings, see: (a) Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 8783−8794. (b) Wade, W. S.; Mrksich, M.; Dervan, P. B. *Biochemistry* **1993**, *32*, 11385−11389. (c) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 2572−2576. (d) White, S.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 8756−8765.

(3) For Py/Py pairing, see: (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. Biomol. Struct. Dyn.* **1990**, *8*, 81−97. (c) Pelton, J. G.; Wemmer, D. E. *J. Am. Chem. Soc.* **1990**, *112*, 1393−1399. (d) Wemmer, D. E.; Fagan, P. A.; Pelton, J. G. In *Molecular Basis of Specificity in Nucleic Acid-Drug Interactions*; Pullman, B., Jortner, J., Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1990; pp 95−101. (e) Fagan, P. A.; Wemmer, D. E. *J. Am. Chem. Soc.* **1992**, *114*, 1080−1081. (f) White, S.; Baird, E. E.; Dervan, P. B. *Biochemistry* **1996**, *35*, 12532−12537.

(4) For Im/Im pairing, see: (a) Dwyer, T. J.; Geierstanger, B. H.; Bathini, Y.; Lown, J. W.; Wemmer, D. E. *J. Am. Chem. Soc.* **1992**, *114*, 5911−5919. (b) Singh, S. B.; Ajay; Wemmer, D. E.; Kollman, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 7673−7677. (c) White, S.; Baird, E. E.; Dervan, P. B. *Chem. Biol.* **1997**, *4*, 569−578.

(5) For Hp/Py pairing, see: (a) White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *Nature* **1998**, *391*, 468−471. (b) Kielkopf, C. L.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Science* **1998**, *288*, 111−115.

(6) For extended dimers, see: (a) Kelly, J. J.; Baird, E. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 6981−6985. (b) Trauger, J. W.; Baird, E. E.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6160−6166. (c) Geierstanger, B. H.; Mrksich M.; Dervan, P. B.; Wemmer, D. E. *Nature, Struct. Biol.* **1996**, *3*, 321−324. (d) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *Chem.—Eur. J.* **1997**, *3*, 1600−1607. (e) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 3534−3535.

(7) For central-ring-bridged dimers, see: (a) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 3663. (b) Dwyer, T. J.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, *115*, 9900. (c) Chen, Y. H.; Lown, J. W. *J. Am. Chem. Soc.* **1994**, *116*, 6995. (d) Alsaid, N. H.; Lown, J. W. *Tetrahedron Lett.* **1994**, *35*, 7577. (e) Alsaid, N. H.; Lown, J. W. *Synth. Comm.* **1995**, *25*, 1059 (f) Chen, Y. H.; Yang, Y. W.; Lown, J. W. *J. Biomol. Struct. Dyn.* **1996**, *14*, 341. (g) Singh, M. P.; Wylie, W. A.; Lown, J. W. *Magn. Reson. Chem.* **1996**, *34*, S55. (h) Greenberg, W. A.; Baird, E. E.; Dervan, P. B. *Chem.—Eur. J.* **1998**, *4*, 796−805.

(8) For hairpin motif s, see: (a) Mrksich, M.; Parks, M. E.; Dervan, P. B. *J. Am. Chem. Soc.*, **1994**, *116*, 7983−7988. (b) Parks, M. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6147−6152. (c) Parks, M. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6153−6159. (d) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 8198−8206. (e) Pilch, D. S.; Poklar, N. A.; Gelfand, C. A.; Law, S. M.; Breslauer, K. J.; Baird, E. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 8306−8311.

(9) For chiral hairpins, see: Herman, D. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 1382.

(10) For cyclic polyamides, see: Cho, J. Y.; Parks, M. P.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 10389.

(11) For solid-phase polyamide syntheses, see: Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6141.

10.1021/ja983671u CCC: $18.00
Published on Web 03/19/1999

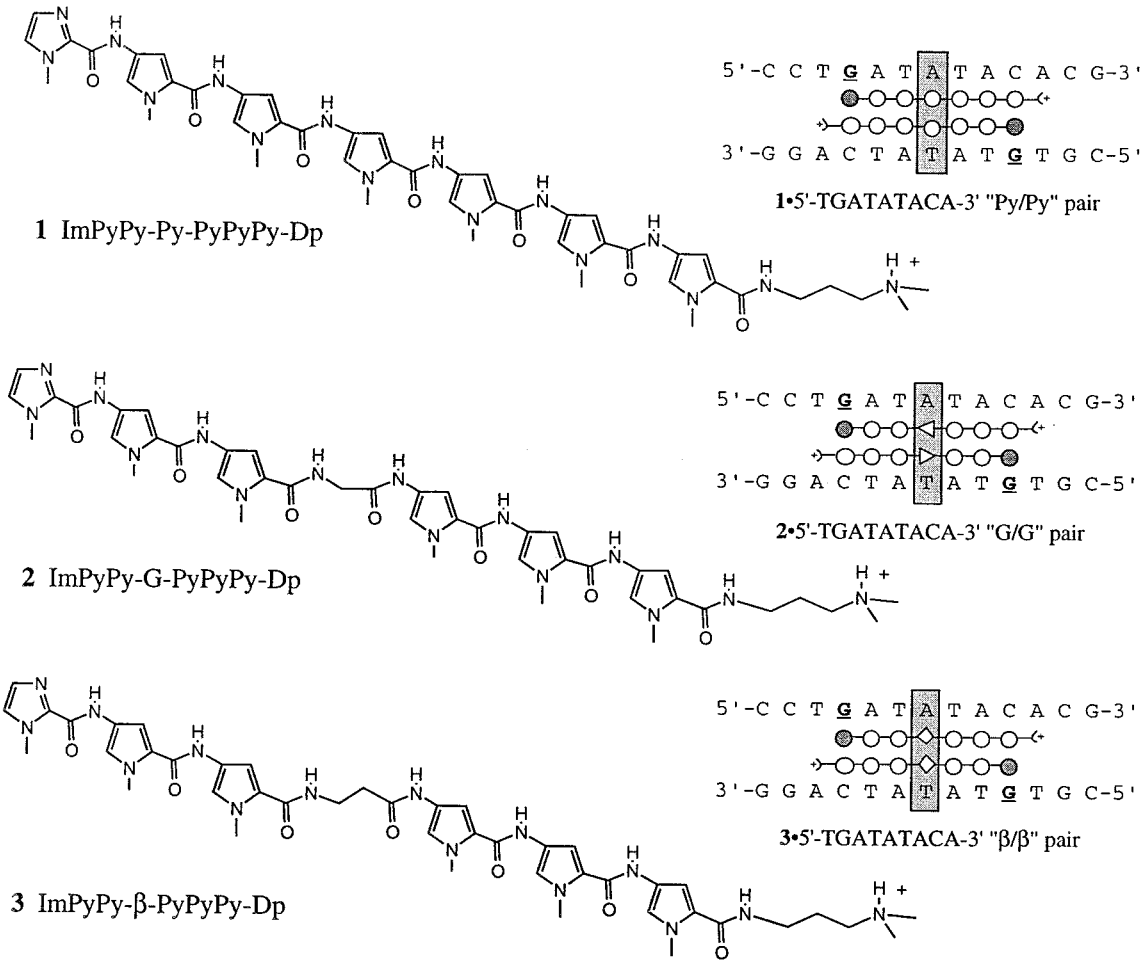

Figure 1. Chemical structures of head-to-tail linked polyamides together with schematic representations of the complexes they form. In the schematics, open circles represent *N*-methylpyrrole rings, while shaded circles indicate *N*-methylimidazole rings, triangles represent glycine, and diamonds represent β-alanine.

by NMR[12] and X-ray[5b,13] structure studies. Eight-ring pyrrole−imidazole polyamides have been shown to be cell permeable and to inhibit the transcription of designated genes in cell culture.[14] This provides impetus to understand the structural basis of several useful motifs which recognize a broad DNA-sequence and site size repertoire.

Binding Site Size Limitations. In the fully overlapped 2:1 binding mode, polyamide dimers containing solely pyrrole and imidazole amino acids can recognize a maximum binding site size of seven base pairs (five contiguous rings) before ligand affinity and specificity decrease.[6a] High-resolution X-ray studies indicate that the polyamide rise per aromatic amino acid residue matches the DNA rise per base pair; however, the curvature of the crescent-shaped polyamide dimer was found to be overwound relative to the pitch of the DNA helix.[13c] Polyamides of sequence composition ImPyPy-X-PyPyPy-Dp, where X is Py, glycine (G), or β-alanine (β), form overlapped 2:1 complexes at a 5′-TGATATACA-3′ site (Figure 1). A central Py/Py, G/G, or β/β pair is expected to form for each respective polyamide. Footprinting reveals that both binding affinity and sequence specificity increase for X = Py < G < β, indicating that the flexible aliphatic amino acid "springs" effectively reset the polyamide curvature with the minor groove for recognition of larger binding site sizes.[6] For example, the polyamide ImPyPy-β-PyPyPy-Dp (β/β pair) binds to the nine base pair 5′-TGTTAAACA-3′ target site with an equilibrium association constant (K_a) of $K_a = 8 \times 10^8$ M^{-1}, a 10-fold higher affinity than the "pyrrole-linked" polyamide ImPyPy-Py-PyPyPy-Dp.[6b] We have employed a combination of nuclear magnetic resonance and restrained molecular modeling to directly characterize binding of the extended polyamides ImPyPy-β-PyPyPy-Dp and ImPyPy-G-PyPyPy-Dp to a 5′-TGTATATCA-3′ target site. As a control, the parent ligand containing exclusively aromatic amino acids, ImPyPy-Py-PyPyPy-Dp, was also characterized. We report the effects of the central β/β, G/G, and Py/Py pairs on local structure and dynamics.

(12) For NMR studies, see: (a) 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. (b) Geierstanger, B. H.; Dwyer, T. J.; Bathini, Y.; Lown, J. W.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, *115*, 4474−4482. (c) Geierstanger, B. H.; Jacobsen, J. P.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Biochemistry* **1994**, *33*, 3055−3062. (d) Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Science* **1994**, *266*, 646−650. (e) de Clairac, R. P. L.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1997**, *119*, 7909−7916. (f) Wemmer, D. E.; Geierstanger, B. H.; Fagan, P. A.; Dwyer, T. J.; Jacobsen, J. P.; Pelton, J. G.; Ball, G. E.; Leheny, A. R.; Chang, W. H.; Bathini, Y.; Lown, J. W.; Rentzeperis, D.; Marky, L.; Singh, S. B.; Ajay; Kollman, P. A. In *Structural Biology: The State of the Art. Proceedings of the Eighth Conversation on Biomolecular Stereodynamics*; Sarma, S. M., Ed.; Adenine Press: Guilderland, NY, 1994; Vol. 2, pp 301−323.

(13) For X-ray studies, see: (a) Chen, X.; Ramakrishnan, B.; Rao, S. T.; Sundaralingham, M. *Nature, Struct. Biol.* **1994**, *1*, 169. (b) Chen, X.; Ramakrishnan, B.; Sundaralingham, M. *J. Mol. Biol.* **1997**, *267*, 1157. (c) Kielkopf, C. L.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Nature, Struct. Biol.* **1998**, *5*, 104−109.

(14) For cell culture, see: (a) Gottesfield, J. M.; Nealy, L.; Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1997**, *387*, 202−205. (b) Dickenson, L. A.; Guzilia, P.; Trauger, J. W.; Baird, E. E.; Mosier, D. M.; Gottesfeld, J. M.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 12890−12895.

Materials and Methods

Synthesis of Ligands and Oligonucleotides. ImPyPy-G-PyPyPy-Dp, ImPyPy-β-PyPyPy-Dp, and ImPyPy-Py-PyPyPy-Dp were synthesized and purified as described previously.[6a,8a] All DNA oligonucleotides were prepared on an automated synthesizer on a 1 μmol scale and purified using HPLC as described previously.[12]

Sample Preparation. NMR samples contained 10 mM sodium phosphate buffer in 0.5 mL of 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 duplexes as determined by UV absorbance at 80 °C using extinction coefficients calculated from the standard values for random coil monomers.[15]

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. For competition experiments, the concentrations were typically 0.1 mM in the initial complexes. The competing species were added in approximately 0.25−0.5 mol equiv per addition. 1D spectra in D_2O (an average of 128−2048 scans) were acquired with 4096 or 8192 complex points over a spectral width of 5000 Hz (500 MHz). NOESY spectra in D_2O (200 ms mixing time) were collected with 1024 or 2048 complex points in t_2 using a spectral width of 5000 Hz (500 MHz); 478−870 t_1 experiments with 32−64 scans were recorded and zero-filled to 1024 or 2048 points. TOCSY spectra (τ_{mix} = 40 and 70 ms) in D_2O (490−504 t_1 experiments, 64 scans) were collected for the β_2 complex and for the Py_2 complex (τ_{mix}=70 ms, 490 t_1 experiments, 64 scans). For experiments in D_2O, presaturation pulses were applied during the recycle delay (2 s) and the mixing period to suppress the solvent resonance. NOESY spectra in water were acquired at 45 °C for the G/G complex (τ_{mix} = 200 ms) and 25 °C for the β/β complex (150 and 200 ms mixing times) and the Py/Py complex (150 and 200 ms mixing times), replacing the last 90° pulse by a 1−1 jump and return sequence for solvent suppression as described previously.[12a,b] The spectra were collected into 2048 complex points in t_2 using a spectral width of 13 514 Hz at 600 MHz; 438−475 t_1 experiments (64 scans) were recorded and zero-filled to 2K. All 2D spectra were acquired using TPPI. The data were processed with FELIX (versions 2.30β and 95.0β, Biosym, San Diego, CA) 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[16,17] and as previously described.[12] NOE contacts between C4′H DNA protons and H3, H5, and *N*-methyl pyrrole protons of the ligands were assigned in analogy to the contacts observed in complexes previously characterized.[12a]

Molecular Modeling Using Restrained Energy Minimization. Molecular models of the complexes of d(CGTGTATATCAGG)·d(CCTGATATACACG) (henceforth just one strand of the underlined binding site will be shown for simplicity) with ImPyPy-β-PyPyPy-Dp and ImPyPy-Py-PyPyPy-Dp were obtained using the Biosym molecular modeling package InsightII installed on Silicon Graphics workstations. The initial DNA model was constructed using the Biopolymer module of InsightII as the standard B-form, since the NMR data indicated that the DNA remains basically as its B-form in the complex. Coordinates for the polyamides were derived from the model of the monomeric complex of ImPyPy-β-PyPyPy-Dp.[12e] The linkers were modified as needed using the Builder module. Energy minimizations using Discover (with the AMBER force field) were performed with the ring systems in an antiparallel, dimeric arrangement. Partial atomic charges were then calculated using MOPAC (AM1). The polyamides were roughly oriented within the 5′-TGTATATCA-3′ binding site by manual docking. Restrained energy minimizations (Discover) were then performed on the complex, as described previously.[12e] Forty-seven and 41 intermolecular ligand−DNA and 22 intraligand restraints were derived from NOESY data at 150 and 100 ms mixing times, for the β_2 and Py_2 complexes, respectively. NOE cross-peaks were classified semiquantitatively into three categories: strong (1.8−2.5 Å), medium (2.5−3.7 Å), and 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). Energy minimizations after restrained molecular dynamics with temperatures of up to 600 K were used to sample conformational space more extensively.

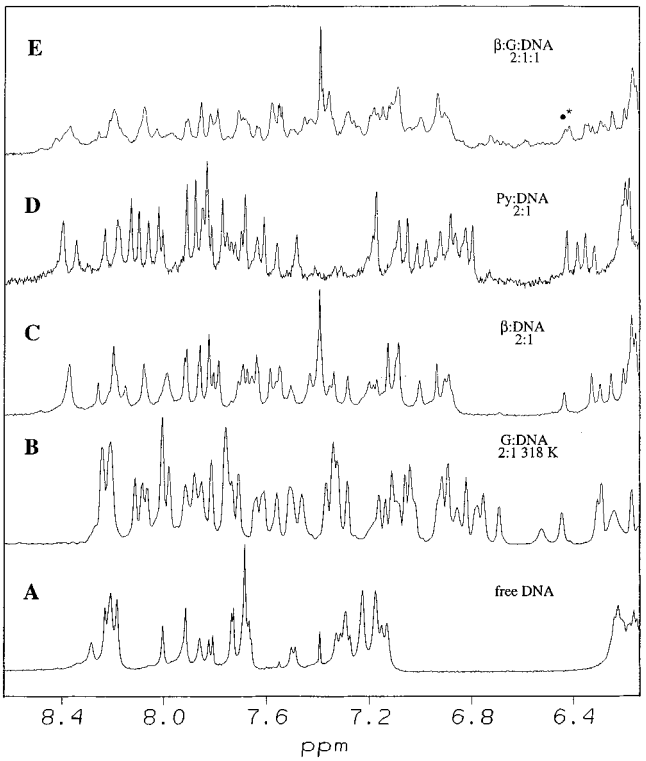

Figure 2. Aromatic region of the ^1H NMR (at 500 MHz) spectra in D_2O of (A) free d(CGTGTATATCAGG)·d(CCTGATATACACG) [DNA], (B) the 2:1 "extended" complex of DNA with ImPyPy-G-PyPyPy-Dp [G] at 45 °C, (C) the β_2 complex with DNA at 25 °C, (D) the 2:1 Py_2 complex, and (E) an approximately 1:1 mixture of the G−β−DNA (and/or β−G−DNA) hetero complex(es) (∗) and the β_2−DNA homodimeric complex (●) formed at 2:1:1 β:G:DNA stoichiometry. The molar ratios are indicated for each spectrum.

Results

NMR Titrations. One-dimensional spectra were recorded at 25 °C at points in independent titrations of TGTATATCA with ImPyPy-G-PyPyPy-Dp, ImPyPy-β-PyPyPy-Dp, and ImPyPy-Py-PyPyPy-Dp (Figure 1). For the G- and Py-linked ligands, a single form of complex was seen at all stoichiometries, the resonances of the complex in slow exchange with those from free DNA. At the final 2:1 ligand:DNA ratio, only resonances from the complex were seen (Figure 2B,D). For the G_2 complex, a few resonances were distinctly broadened even at 2:1 stoichiometry. These resonances sharpened markedly with increasing temperature, with only a few resonances still broadened at 45 °C (Figure 2B). Since the broadening is independent of stoichiometry and occurs only for a subset of resonances, this must arise from conformational exchange from within the complex rather than from dissociation of the complex. There is no evidence for such broadening in the other complexes studied.

(15) Warshaw, M.; Cantor, C. *Biopolymers* **1970**, *9*, 1079−1103.
(16) Hare, D. R.; Wemmer, D. E.; Chou, S.-H.; Drobny, G.; Reid, B. R. *J. Mol. Biol.* **1983**, *171*, 319−336.
(17) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; J. Wiley & Sons: New York, 1986.

For titration of TGTATATCA with ImPyPy-β-PyPyPy-Dp, complexes of 1:1 stoichiometry are formed at low ligand concentration, as determined by the ratio of free duplex to complex. When the total ligand:DNA ratio reaches 1:1, a new species forms, increasing in amount until it is the only form of complex present at a final stoichiometry of 2:1 (Figure 2D). Exchange between both forms of the complex and the free DNA is slow. Full characterization of the 2:1 complex, described in further detail below, shows that it is the extended, side-by-side complex anticipated. The qualitative similarity of these complexes can be seen from the induced chemical shift changes, Figure 3.

To determine if a β/G pair is structurally compatible with the DNA minor groove, the β_2 and G_2 DNA complexes were mixed. Combining equal amounts of the G_2 and β_2 complexes with TGTATATCA leads to new resonances (not from either homo complex) (Figure 2F), indicating formation of a new complex containing one β and one G ligand. Further addition of ImPyPy-β-PyPyPy-Dp up to 2:1:1 β:G:TGTATATCA stoichiometry yields a roughly 1:1 ratio of the homodimeric to the heterodimeric complex. These data suggest that the heteropairs of β/G and G/β can form stable complexes with DNA.

Characterization of the β_2 Complex. Numerous intermolecular ligand−DNA and ligand−ligand contacts are observed in the NOESY data from the 2:1 complex of ImPyPy-β-PyPyPy-Dp with TGTATATCA, confirming the simultaneous binding of the two ligands in the minor groove (Table 1, Figures 4B and 5, and Supporting Information). One polyamide again lies along 5′-GTATATCA-3′, and the other along 5′-GATATACA-3′. Hydrogen bond formation between the imidazole nitrogens of both ligands and the N2 amino protons of G4 and G17 is observed. This indicates that the β linkers are able to span one base pair when opposite one another in the minor groove. The details in the linker region are poorly characterized, as the eight methylene protons of the two linkers only show detectable NOE contacts to the C2 proton of A20 (Table 1). Sequential NOEs to the pyrrole H3 protons and intermolecular contacts to sugar H1's and adenine C2 protons place the amide protons of both ligands deep in the minor groove. The close proximity of the ligand tails to the DNA is also indicated by cross-peaks between the methyl protons and the H1′ protons of the thymines T3 and T16 that flank the binding site.

Characterization of the G_2 Complex. NOESY data acquired in D_2O at 25 °C confirm the 2:1 fully overlapped conformation for ImPyPy-G-PyPyPy-Dp when bound to TGTATATCA (Figure 5 and Supporting Information). NOE contacts between the ligand protons and the sugar and adenine C2 protons of the DNA indicate that two ligands are bound in the minor groove at the expected locations. One of the ligands contacts the residues 5′-GTATATCA-3′ and the other spans 5′-GATATACA-3′ (Figure 5). Each imidazole ring within the antiparallel, side-by-side ligands recognizes one G on its corresponding strand (G4 and G17, respectively). This arrangement is confirmed by intermolecular NOEs from the imidazole H4-1 and H5-1 protons of one ligand to the protons of the positively charged tail of the other ligand at both ends of the binding site. In addition, the terminal pyrrole and/or imidazole H5 of one ligand has cross-peaks to the *N*-methyl protons of the corresponding stacked ring of the other ligand, and the H3 pyrrole protons of rings 2 and 5 of the ligands also show cross-peaks (Supporting Information). Contacts between the H3 of the pyrroles and the adenine C2 protons also support this ligand arrangement (Figure 5). Line broadening precludes complete characterization of the complex. The resonances corresponding to the protons of the G linkers show no NOEs, nor do the H3 protons of the pyrrole rings immediately preceding the linker, 1H3-3 and 2H3-3. The resonances of the protons of the rings following the linker are somewhat sharper and show some weak cross-peaks. The resonances corresponding to the DNA protons in this region are also broadened and show no contacts or only very weak ones. This indicates a dynamic process involving the ligand linkers and the central region of the DNA-binding site.

Table 1. Ligand−DNA and Intraligand NOE Contacts for the β_2 Complex with d(CGTGTATATCAGG)·d(CCTGATATACACG)[a,b]

DNA to Lig.

Ligand 1	DNA Strand 1	DNA Strand 2	Ligand 2
	T3 H1'		$2N(CH_3)_2$
	T3 H4'		
		A24 H1'	2H3-6 2NH-7
		A24 H2	2C18-H1/2 2C19-H1/2 2C20-H1/2
		A24 H4'	2H3-6 2H5-6
1H4-1	G4 H1'		
1NH-1		G4 NH_2	2NH-6
		C23 H1'	2H3-5 2NH-6
		C23 H4'	2H3-5
1NH-1	T5 H1'		
1H5-1	T5 H4'		
		A22 H1'	2H3-4 2NH-5
1H3-2		A22 H2	2H3-5 2NH-5 2H3-6 2NH-6
1H3-2 1NH-2	A6 H1'		
1H3-3 1NH-2		A6 H2	2H3-4 2H3-5 2NH-5
1H5-2	A6 H4'		
		T21-H1'	2NH-4
1H3-3 1NH-3	T7 H1'		
1H3-3 1H5-3	T7 H4'		
		A20 H1'	2H3-3 2NH-3
1C21-H1/2 1C22-H1/2		A20 H2	2NH-3 2C21-H1/2 2C22-H1/2
		A20 H4'	2H3-3 2H5-5
1NH-4	A8 H1'		
1H3-4 1NH-4 1NH-5		A8 H2	2H3-3
		T19 H1'	2H3-2 2NH-2
		T19 H4'	2H3-2
1H3-4 1NH-5	T9 H1'		
1H5-4	T9 H4'		
		A18 H1'	2NH-1
1H3-5 1NH-6		A18 H2	2H3-2 2NH-2
1H3-5 1NH-6	C10 H1'		
1H3-5 1H5-5	C10 H4'		
		G17 H1'	2H4-1
1NH-6		G17 NH_2	2NH-1
1H3-6 1NH-7	A11 H1'		
1NH-2 1C18-H1/2 1C19-H1/2 1C20-H1/2		A11 H2	
1H3-6 1H5-6	A11 H4'		
$1N(CH_3)_2$		T16 H1'	
		T16 H4'	

Interligand

Ligand 1	Ligand 2
1H4-1	2C18-H1/2 2C19-H1/2 2C20-H1/2 $1N(CH_3)_2$
1H5-2	2H5-6
1H5-3	2H5-5
1H5-5	2H5-3
1H5-6	2H5-2
1H5-1	2C20-H1/2 $2N(CH_3)_2$ 2N6-CH_3
1H3-2	2H3-5
1C19-H1/2 1C20-H1/2 $1N(CH_3)_2$	2H4-1 2H5-1

[a] Identified in the H_2O NOESY spectrum acquired at 100 ms mixing time. [b] C18,19,20 protons are not stereospecifically assigned.

The spectra acquired at 45 °C (Figure 4A) display the same basic features as the ones at 25 °C (Supporting Information). The resonances corresponding to the pyrrole rings flanking the linker are somewhat sharper and show broad NOE cross-peaks to the DNA. However, the linker protons, the C2 proton of A20, and the ligand amide protons 1NH-4 and 2NH-4 can still not be identified. A NOESY spectrum acquired in 90% H_2O/10% D_2O solvent was used to perform the sequential intramolecular assignment of all H3 and amide protons, except the NH-4 proton of each ligand. The imidazole N of one ligand forms a specific hydrogen bond with the guanine amino proton of G4, consistent with the rules for recognition of G·C base pairs. This is indicated by cross-peaks between the amino protons and ligand protons (Supporting Information), as previously described.[12] Such NOE cross-peaks are not observed for the broader G17 amino resonances, but their chemical shifts suggest interaction with the nitrogen of the adjacent imidazole ring. These observa-

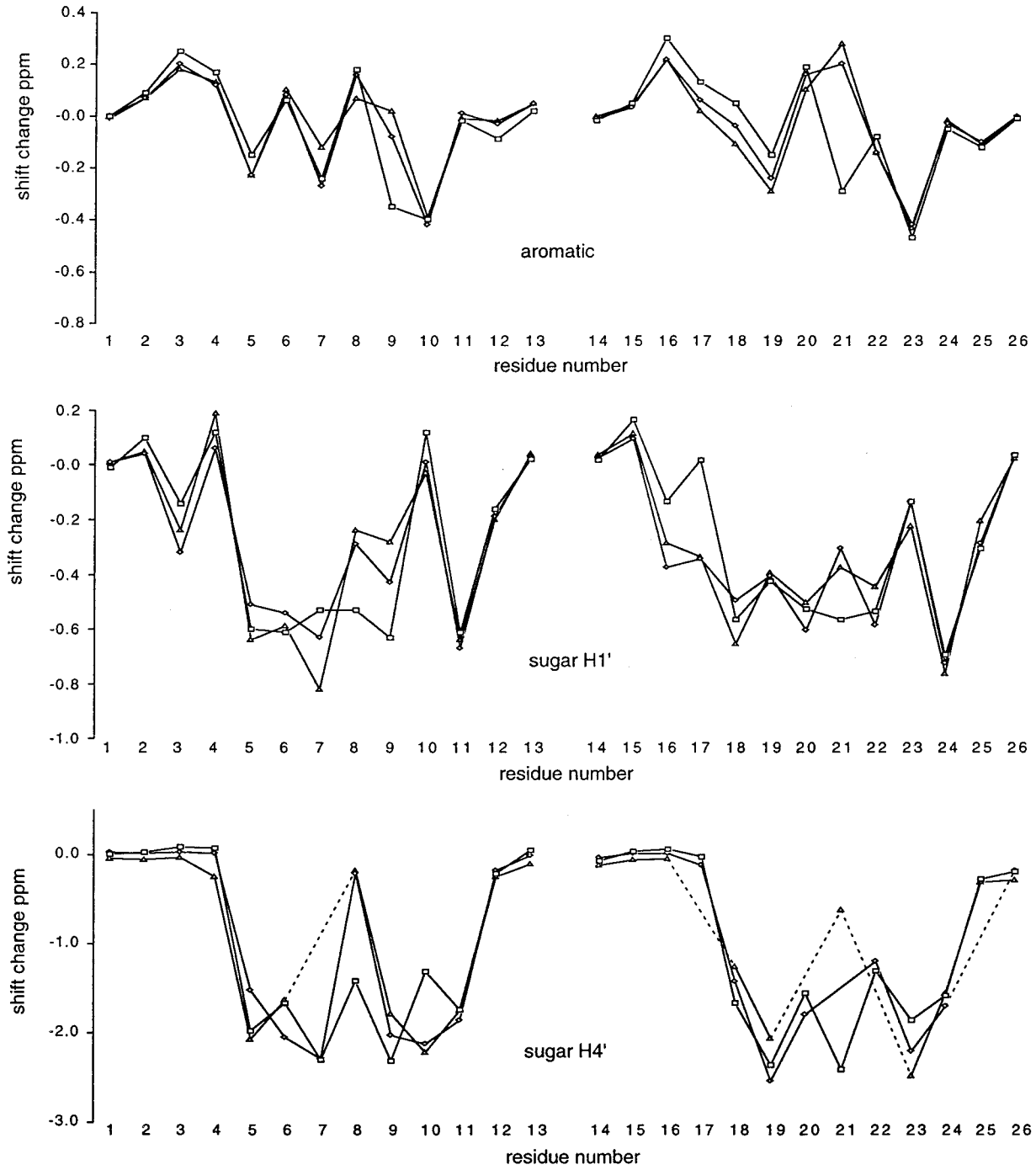

Figure 3. Chemical shift changes of representative DNA protons induced by complex formation with the three polyamides: $\beta = \diamond$, G = $\triangle$, Py = $\square$. Differences in ppm between complexed and free DNA are plotted versus the number of the residue to which they belong. Dashed lines are used to connect two nonsequential points if the corresponding value(s) for the intermediate one(s) is (are) not known.

tions suggest that the G linkers are not able to span the central base pair when opposite one another in the minor groove, resulting in a nonoptimal arrangement of the ligands in the complex.

Characterization of the Py_2 Complexes. Intermolecular NOESY patterns similar to those of the complexes described above confirm the formation of a 2:1 complex of ImPyPy-Py-PyPyPy-Dp with TGTATATCA (Table 2; Figures 2, 3, 4C, and 5). One of the polyamides lies along 5′-GTATATCA-3′, and the other contacts lie along 5′-GATATACA-3′. The presence of the additional Py ring in the linker region shifts nearby sugar resonances upfield with respect to those of the other two complexes (Figure 3). Hydrogen bond formation between the imidazole nitrogen of one of the ligands and the amino protons of G4 is observed. For G17, the rotation about the C−N bond is slowed (as indicated by sharper lines than seen in the spectrum of free DNA) but there is a smaller downfield shift of the non-Watson−Crick hydrogen-bonded amino proton, suggesting a somewhat different geometry and possibly a weaker hydrogen bond. Sequential amide−H3−amide NOEs within the polyamides and intermolecular contacts of these protons to sugar H1's and several adenine C2 protons verify that the amide protons of both ligands reside deep in the minor groove for the full length of each ligand (Table 2). The close proximity of the ligand tails to the DNA is also indicated by cross-peaks between the methyl and methylene protons of the Dp tail and the H1′ protons of T3, G4, T16, and G17 that flank the binding site.

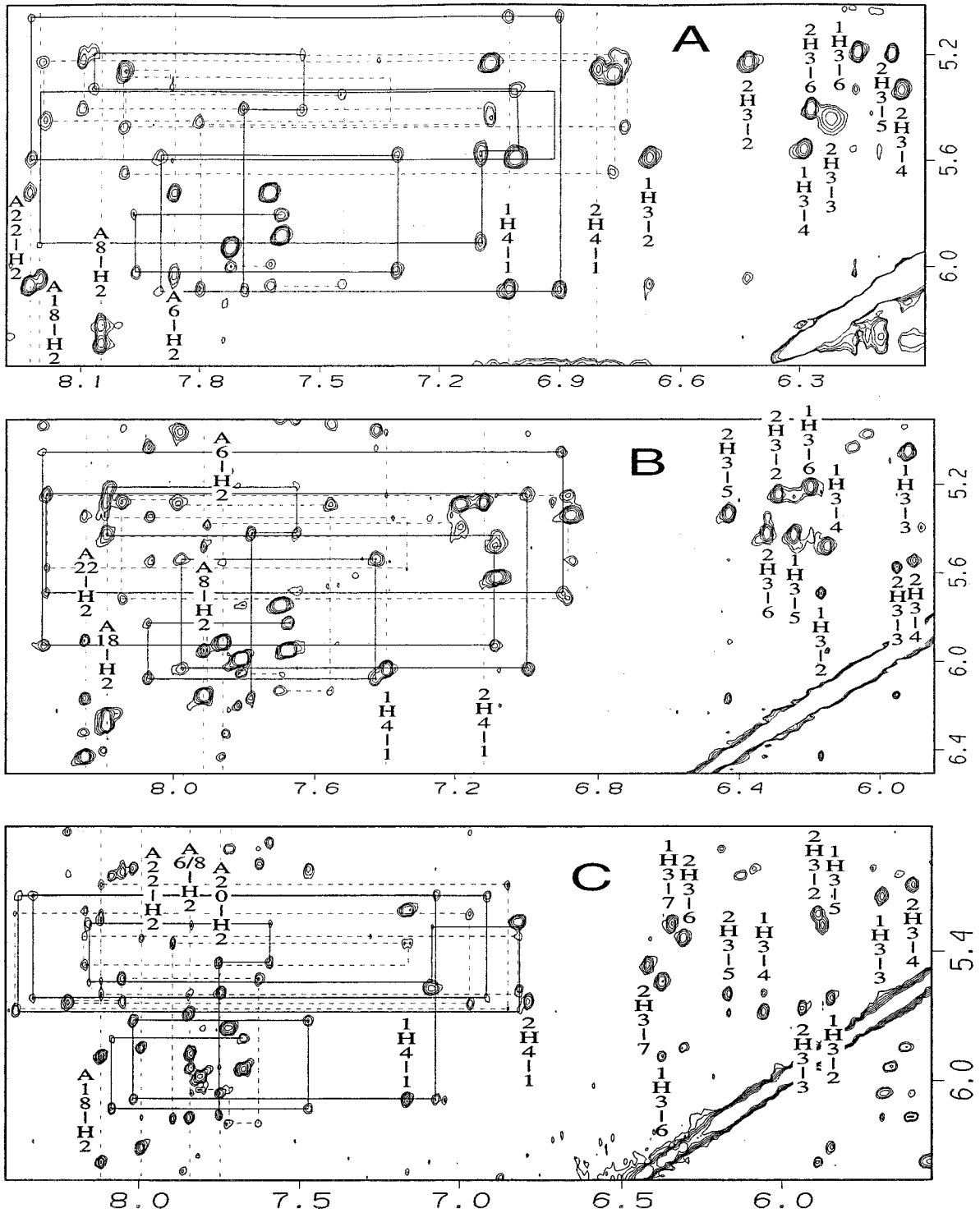

Figure 4. Expansion of the aromatic and H1′ region of NOESY spectra (in D_2O, 500 MHz, τ_{mix} = 200 ms) of d(CGTGTATATCAGG)·d(CCTGATATACACG) in complexes with ImPyPy-G-PyPyPy-Dp at 45 °C (A), ImPyPy-*β*-PyPyPy-Dp at 25 °C (B), and ImPyPy-Py-PyPyPy-Dp (C). Sequential aromatic to C1′H connectivities for the 5′-TGTATATCA-3′ strand are shown as solid lines; those for the 5′-TGATATACA-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. The ligands making contacts to $T_3G_4T_5A_6T_7A_8T_9C_{10}A_{11}$ and $T_{16}G_{17}A_{18}T_{19}A_{20}T_{21}A_{22}C_{23}A_{24}$ 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 ligand number, the type of atom and its number, and the corresponding number of the ring or amide.

Molecular Models of the β_2 and Py_2 Complexes. Restrained molecular dynamics using semiquantitative distances derived from NOESY data allow generation of molecular models for the 2:1 ImPyPy-*β*-PyPyPy-Dp:TGTATATCA and the 2:1 ImPyPy-Py-PyPyPy-Dp:TGTATATCA complexes (Figure 6). In the *β*/*β* complex, the linkers adopt an extended conformation in which the methylene protons of the two ligands avoid one another. Since there are few restraints from the *β*-alanine protons, more than one conformation is seen in the structures. The local conformation of the linkers is correlated on the two ligands, probably just from steric interactions. The data do not indicate clearly whether these multiple forms might be present

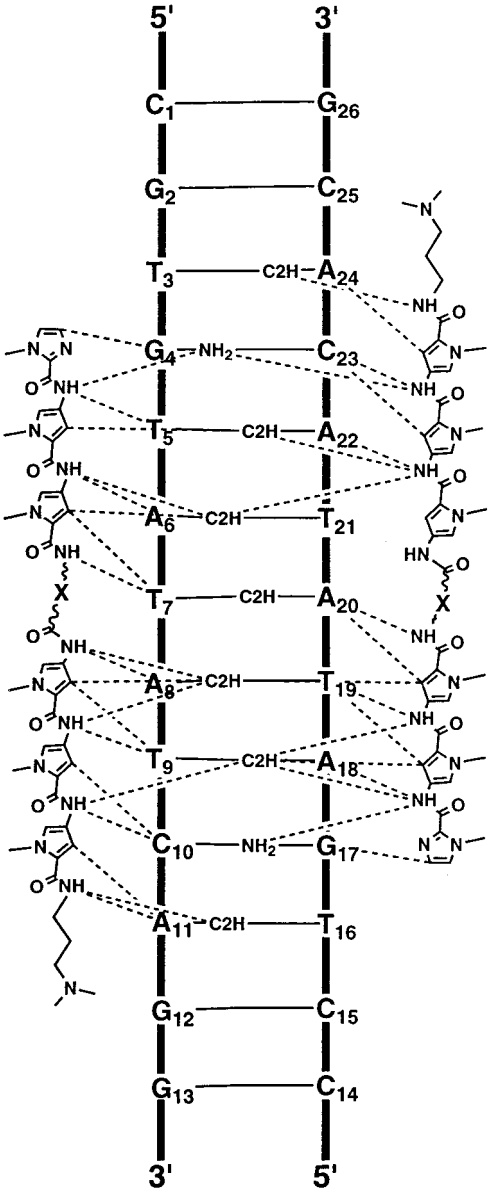

Figure 5. Schematic representation of the X_2−DNA complexes. Observed intermolecular ligand−DNA NOE-derived contacts are indicated for the case of the β_2 complex.

in rapid equilibrium or if, in fact, just one is populated but the lack of stereospecific assignments of the β-alanine protons leaves this region undefined. Fitting the two relatively bulky β-alanine segments opposite to each other also requires the ring systems to slightly modify their relative alignments, also indicated by direct contacts between the Py H5's of rings diagonally opposed across the groove (Table 1), not seen in side-by-side complexes with unlinked, three-ring ligands. Considerable twisting of the ligands is observed within the β-alanine moieties, which can occur due to the higher flexibility of this region compared to the pyrrole ring system. For the more rigid Py ligand, the twisting observed for the polyamide is accommodated relatively uniformly over the full length of the ligand. Hydrogen bonds between the amide protons of the ligand and the N3 of adenines or the O2 of pyrimidine bases facing into the minor groove were assigned by the InsightII program for both complexes. The Im nitrogens are also within hydrogen-bonding distances to the G4 and G17 amino protons that point out into the groove, as seen in other Im-containing-ligand complexes.[12,13] The G/G complex was not modeled because of the lack of constraints in the G linker region due to line broadening.

Table 2. Ligand−DNA and Intraligand NOE Contacts for the Py_2 Complex with d(CGTGTATATCAGG)·d(CCTGATATACACG)[a,b]

DNA to Lig.

Ligand 1	DNA Strand 1	DNA Strand 2	Ligand 2
	T3 H1'		2N(CH3)2
	T3 H4'		
		A24 H1'	2H3-7 2NH-7
		A24 H2	2C18-H1/2 2C19-H1/2 2C20-H1/2
		A24 H4'	2H3-6 2H5-6
1H4-1	G4 H1'		
1NH-1 1H4-1		G4 NH2	2NH-6
	G4 H4'		1N(CH3)2
		C23 H1'	2H3-6 2NH-6
		C23 H4'	2H3-6
1NH-1	T5 H1'		
	T5 H4'		
		A22 H1'	2H3-5 2NH-5
1H3-2		A22 H2	2H3-6 2NH-5 2H3-6 2NH-6
		A22 H4'	2H3-5
1H3-2 1NH-2	A6 H1'		
1H3-3 1NH-2		A6 H2	2H3-4 2H3-5 2NH-5
1H3-2	A6 H4'		
		T21-H1'	2H3-4 2NH-4
		T21-H4'	2H3-4
1H3-3 1NH-3	T7 H1'		
1H3-3	T7 H4'		
		A20 H1'	2H3-3 2NH-3
1H3-4 1NH-4		A20 H2	2NH-3 2H3-4 2NH-4
		A20 H4'	2H3-3 2H5-5
1NH-4 1H3-4	A8 H1'		
1H3-4 1NH-4 1H3-5		A8 H2	2H3-3 2NH-4
1H3-4	A8 H4'		
		T19 H1'	2H3-2 2NH-2
		T19 H4'	2H3-2
1H3-5 1NH-5	T9 H1'		
1H3-5	T9 H4'		
		A18 H1'	2NH-1
1NH-5 1H3-6 1NH-6		A18 H2	2NH-1 2H3-2 2NH-2 2NH-1
1H3-6 1NH-6	C10 H1'		
1H3-6	C10 H4'		
		G17 H1'	2H4-1
1N(CH3)2		G17-H4'	
1NH-6		G17 NH2	2NH-1 2H4-1
1H3-7 1NH-7	A11 H1'		
1NH-7 1C18-H1/2 1C19-H1/2 1C20-H1/2 1H3-7		A11 H2	
1H3-7	A11 H4'		
1N(CH3)2		T16 H1' T16 H4'	

Interligand

Ligand 1	Ligand 2
1H4-1	2H3-7 2C18-H1/2 2C19-H1/2 2C20-H1/2
1H3-2	2H3-6
1H3-3	2H3-5
1H3-4	2H3-4
1H3-6	2H3-2
1C19-H1/2 1C20-H1/2 1N(CH3)2	2H4-1

[a] Identified in the H_2O NOESY spectrum acquired at 200 ms mixing time. [b] C18,19,20 protons are not stereospecifically assigned.

Discussion

The studies described here indicate binding of ImPyPy-X-PyPyPy-Dp polyamides in the extended, overlapped 2:1 ligand arrangement suggested for (T,A)G(T,A)$_5$C(T,A) sequences. The NMR data confirm this binding mode for all three linkers studied, X = G, β, and Py on the sequence AGTATATCT. Quantitative footprinting studies have shown that on a related sequence, TGTTAAACA, the affinities of G/G, β/β, and Py/Py complexes are $K_a = 1.4 \times 10^8$, 7.8×10^8, and 9.7×10^7 M^{-1}, respectively.[6b] In addition, sequence specificity is enhanced for the β/β complex. Despite the different sizes and flexibilities of the linkers, complexes with similar overall structures are formed. However, there are small differences in local structure which are likely the basis for the differences in affinity and specificity.

Considerable line broadening of resonances near the linker (both from ligand and DNA) is seen for the G_2 complex at low temperatures, with marked sharpening as the temperature is raised. This indicates a conformational exchange process on the

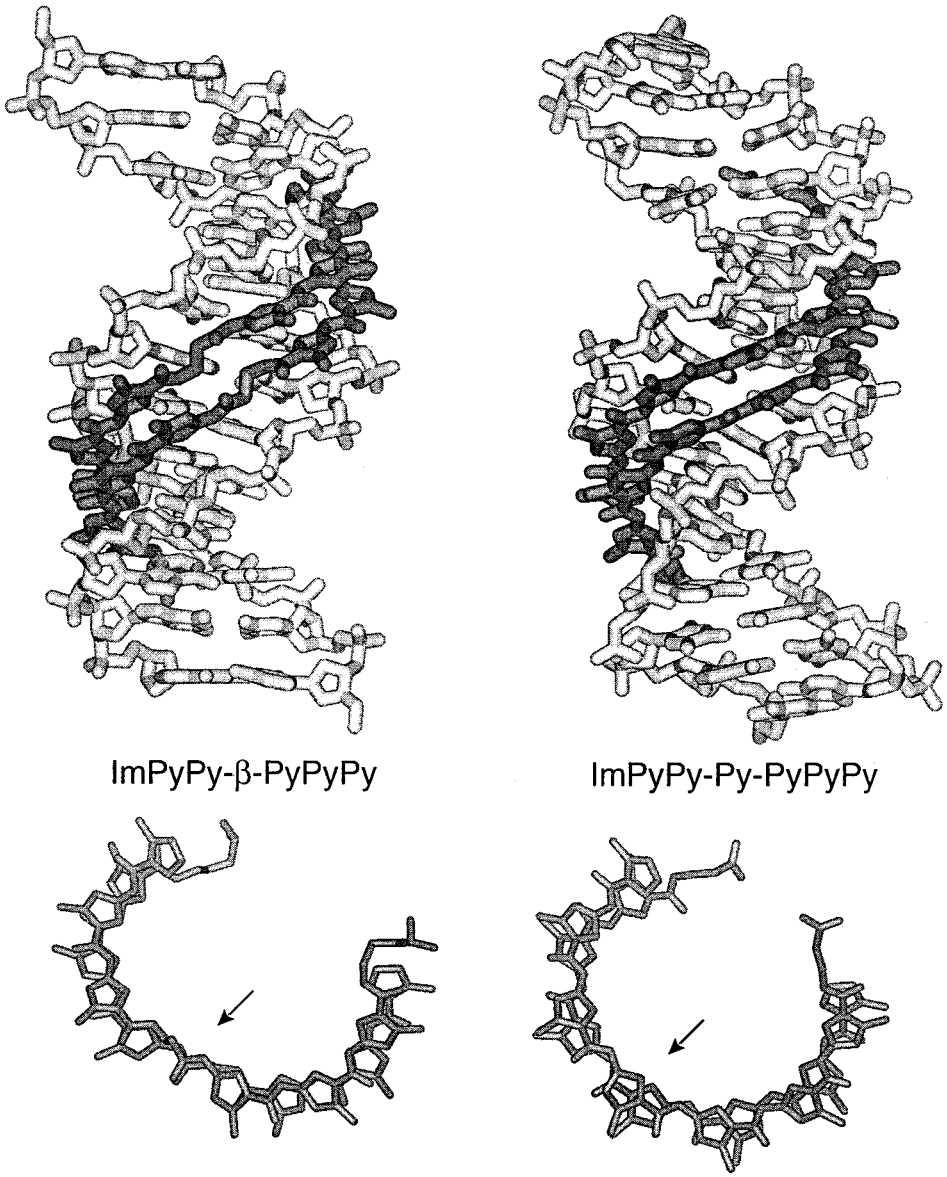

Figure 6. Top: Molecular models of the ImPyPy-*β*-PyPyPy [left] and ImPyPy-Py-PyPyPy [right] complexes. These were obtained by restrained molecular dynamics using semiquantitative distances derived from NOESY data. The ligands are drawn dark for clarity. Bottom: Ligands from the complexes without the DNA, rotated 90°. The site of the "linker" residue is indicated by an arrow.

millisecond time scale involving the linker. Since the complex is pseudosymmetric, with the center of symmetry at the linker, a probable explanation is that the methylene groups of the two glycines do not fit well side-by-side and hence stagger to avoid the steric clash. If the glycines are not equivalent in this staggered geometry, then the two exchanging identities would lead to line broadening such as that observed. Since the broadening occurs only near the linker, the structural differences between the two forms of complex (in both ligand and DNA) can be accommodated locally, not perturbing residues further away. These observations suggest that the G linker does not allow for a good fit of both ligands opposite one another in the minor groove. It is also significant that one of the imidazole rings seems to interact less well with the adjacent guanosine amino group (indicated by a greater rate of rotation of the amino than is typical in such complexes, suggesting a weakened hydrogen bond). The glycine linker introduces greater curvature and is shorter than the other linkers studied; both features may contribute to the less than optimal fit.

Titrations of ImPyPy-*β*-PyPyPy-Dp into 5′-TGTATATCA-3′ followed by NMR showed different complexes at low and high ligand:DNA ratios. Below 1:1 ligand:DNA ratios, a complex consistent with a 1:1 stoichiometry is formed, in which the ligand is assumed to form a hairpin. Studies with binding sites analogous to half of the target site in this work have shown that thermodynamically disfavored *β*-alanine hairpin complexes can form under the conditions of the NMR experiments.[12e] Upon further ligand addition, a new complex appears in slow exchange with the DNA and the first complex. This complex has a 2:1 stoichiometry and shows all of the features characteristic of the previously studied side-by-side, antiparallel motif of this class of minor-groove ligands, including contacts of one ligand with each strand of DNA, both ligands contacting adenosine H2's in the middle of the groove, and distinctive shifts of both ligand and DNA resonances. The hydrogen bonds between the imidazole nitrogens and the guanine amino protons of G4 and G17 are indicated by the slowing of the amino group rotation and by the substantial downfield shift of the non-Watson−Crick amino proton resonance. The behavior appears to be equivalent at the two ends of the complex, in contrast to the behavior of the complex with the glycine-linked ligand. The molecular models obtained from restrained energy minimization are

consistent with these observations: *the 3−β−3 dimer curvature is lower than that of 3-Py-3 and fits better with the shape of the DNA helix* (Figure 6).

In the titration of 5′-TGTATATCA-3′ with ImPyPy-Py-PyPyPy-Dp, a single complex forms at all stoichiometries. This complex again shows all of the hallmarks of the side-by-side family of complexes. The NMR data show clearly that each ligand contacts one DNA strand continuously along the full length of each ligand and resides deep in the minor groove, very much like the ligands in previously described structures of side-by-side complexes. The differences from the *β*/*β* complex are fairly subtle, a more distributed twist driven by the lack of a flexible linker and a shift of relative ring positions along the groove. The observed drop in affinity for long polypyrrole ligands could be due to mismatch of length or curvature. The differences seen in ring positions in the *β*/*β* vs Py/Py complexes are consistent with this idea. Chemical shifts of the amino protons of contacted G residues indicate an inequivalence of H-bonds at the two ends of the complex, suggesting an energetic difference between them. This would be consistent with a length mismatch, i.e. inability of the ligands to "reach" both G residues in an optimum geometry.

The NMR data from the complexes discussed indicate that there is steric conflict between paired glycine linkers, which is apparently reduced for *β* linkers. In addition, the geometry in the G-linked ligand may weaken the interaction between the imidazole and the contacted guanosine residue. The two-carbon *β* linker of the ImPyPy-*β*-PyPyPy-Dp polyamide reduces linker conflict and has sufficient flexibility to allow better contact to the DNA in the "overlapped" extended mode, consistent with the fact that the tightest binding was observed for this ligand.[6] The rigid Py linker has poorer specificity due to its inability to optimize contacts to the DNA without compromising ligand−ligand interactions. The use of the flexible *β* linker expands the binding site size targetable with side-by-side Py-Im polyamides and will facilitate the design of new polyamide motifs targeted for other sequences.

Acknowledgment. We are grateful to the National Institutes of Health (Grant GM 43129 to D.E.W. and Grant GM 27681 to 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., to the Howard Hughes Medical Institute for a predoctoral fellowship to E.E.B, and to the U.S. Department of Energy and the National Science Foundation for instrumentation grants. R.P.L.d.C. and C.J.S. gratefully acknowledge postdoctoral fellowships from the Fundación Ramón Areces and the Deutsche Forschungsgemeinschaft, respectively. R.P.L.d.C., B.H.G., and D.E.W. also thank T. D. Cole and P. A. Fagan for helpful discussions.

Supporting Information Available: Listings of chemical shifts for the ligand protons in the complexes, intermolecular contacts in the G_2 complex at 25 and 45 °C, chemical shifts for the DNA protons of the G/G (at 25 and 45 °C), *β*/*β*, and Py/Py complexes, and intermolecular ligand−DNA and ligand−ligand restraints and achieved distances for the ImPyPy-*β*-PyPyPy-Dp complex (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA983671U