J|A|C|S
ARTICLES
Published on Web 10/09/2002

Analysis of Hairpin Polyamide Complexes Having DNA Binding Sites in Close Proximity

Cheryl A. Hawkins,[†] Eldon E. Baird,[‡] Peter B. Dervan,[‡] and David E. Wemmer*,[†]

Contribution from the Department of Chemistry, MC-1460, University of California, Berkeley, California 94720, and Division of Chemistry and Chemical Engineering, 164-30, California Institute of Technology, Pasadena, California 91125

Received March 4, 2002

Abstract: The binding of two hairpin polyamide ligands at adjacent sites on DNA has been studied using NMR spectroscopy. The ligands ImPyPy-γ-PyPyPy-Gly-Dp and Ac-ImPyPy-γ-PyPyPy-Gly-Dp were studied binding to oligomers containing one or two matched binding sites: 5′-XGTTA-3′ and 5′-TAACX$_N$GTTA-3′, where X is G, C, or A and $N = 0$, 1 or 2. At these sites the C-terminal ring shows an equilibrium between normal and inverted conformations. Better binding was observed with the ligand running 5′ to 3′ along the contacted strand than in the opposite direction. Complexes of DNAs with two binding sites indicated that at least one spacing base pair was required, and that the identity of this base pair was not critical. Binding with 5′ to 3′ contact is again preferred. Demonstrated binding at adjacent sites indicates that it may be possible to engineer cooperative binding for enhanced specificity or affinity.

Introduction

Over the past decade it has been shown that polyamides containing *N*-methylimidazole (Im), *N*-methylpyrrole (Py), and *N*-methylhydroxypyrrole (Hp) rings can be designed to recognize DNA sequence-specifically through minor groove contacts.[1−7] The binding affinities rival those of DNA binding proteins; hence, these ligands are capable of competing with transcription factors and thereby inhibiting gene expression *in vivo*.[8] The sequence specificity of these ligands is determined by the sequence of side-by-side pairings of the aromatic rings: *N*-methylimidazole is specific for guanine, through hydrogen bonding between the amino of G and the imidazole; *N*-methylhydroxypyrrole is specific for thymine through a hydrogen bond from the hydroxy and the thymine carbonyl; *N*-methylpyrrole binds at adenine, cytosine, or thymine.[1−3,7] Thus Im/Py, Py/Im, Hp/Py, and Py/Hp pairs are specific for G•C, C•G, T•A, and A•T base pairs respectively, and Py/Py binds at either A•T or T•A base pairs.

Side-by-side pairings comprising this recognition motif may be formed between separate molecules (a noncovalent dimer) or between two linked halves of the same molecule (a covalent dimer).[4−6] The most studied ligands are "hairpins" consisting of two sets of rings linked by γ-aminobutyric acid group (henceforth γ-linker).[6] This linker allows the molecule to fold back upon itself with the two planar, aromatic ring systems stacked side-by-side within the DNA minor groove. These ligands, prepared by solid-phase synthesis, also typically contain a β-alanine-dimethylaminopropyl (β-Dp) "tail".[9] Both the γ-linker and the tail contact the minor groove and contribute to affinity and sequence specificity.[10,11] These aliphatic groups have been shown to have strong preference for sites with A,T base pairs.

* Corresponding author. E-mail address: dewemmer@lbl.gov.
† University of California, Berkeley.
‡ California Institute of Technology.

(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) Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Science* **1994**, *266*, 646−650. (a) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1995**, *117*, 3325−3332. (c) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 8198−8206. (d) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 6953−6961. (e) Kielkopf, C. L.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Nat. Struct. Biol.* **1998**, *5*, 104−109.
(4) (a) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 5, 2572−2576. (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.
(5) (a) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 9892−9899. (b) Dwyer, T. J.; Geierstanger, B. H.; Mrksich, M.; Dervan, 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. (d) Chen, Y. H.; Lown, J. W. *J. Am. Chem. Soc.* **1994**, *116*, 6995−7005.
(6) (a) Mrksich, M.; Parks, M. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 7983−7988. (b) Pilch D. S.; Poklar, N.; 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. (c) de Clairac, R. P. L.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1997**, *119*, 7909−7916.
(7) (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**, *282*, 111−115. (c) White, S.; Turner, J. M.; Szewczyk, J. W.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1999**, *121*, 260−261.
(8) (a) Gottesfeld, J. M.; Neely, L.; Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1997**, *387*, 202−205. (b) Dickinson, L. A.; Gulizia, R. J.; Trauger, J. W.; Baird, E. E.; Mosier, D. E.; Gottesfeld, J. M.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 12890−12895.
(9) Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6141−6146.
(10) Parks, M. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6147−6152.
(11) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1999**, *121*, 1113−1120.

10.1021/ja020335b CCC: $22.00 © 2002 American Chemical Society

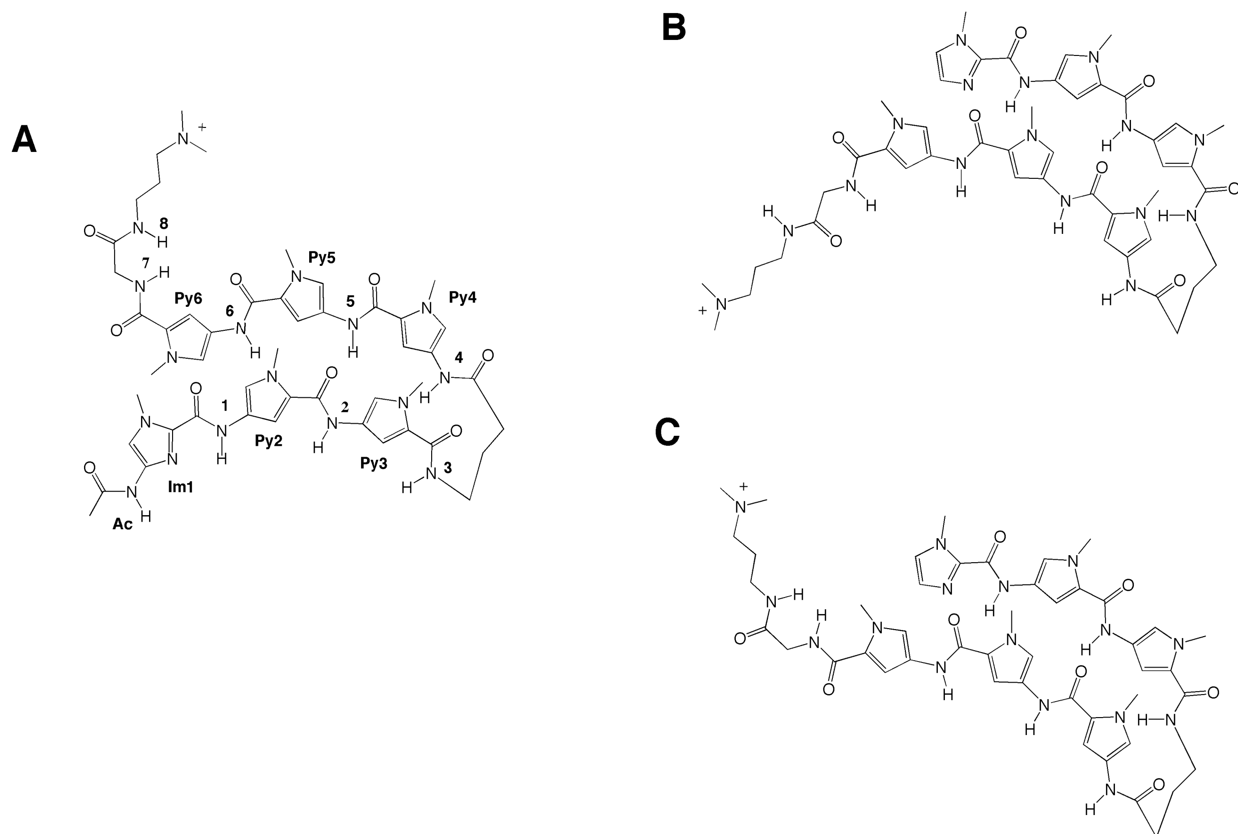

Figure 1. (A) Structure of Ac-ImPyPy-γ-PyPyPy-Gly-Dp with the C-terminal pyrrole in the inverted conformation. The numbering of aromatic rings and amide groups is shown. (B) Structure of ImPyPy-γ-PyPyPy-Gly-Dp with the C-terminal pyrrole in the normal conformation. The -Gly-Dp tail of the ligand is shown in an extended conformation that would potentially allow the tail to contact the groove, which has not been observed with these ligands. Note that in this conformation the second carbonyl oxygen of the tail is directed toward the floor of the groove which is probably unfavorable. (C) The -Gly-Dp tail of the ligand is shown in a "kinked" conformation that would allow the C-terminal pyrrole rings of two ligands in close proximity to spend time in the normal conformation.

In analogy to peptides, the polyamides have N- and C-termini, with the tail normally at the C-terminus. It has been noted that polyamide ligands bind in a strongly preferred orientation, with the N-terminus of the ligand at the 5′ end of the run of bases contacted on one strand, referred to as 5′ directionality.[1] It has been shown that ligands with an N-terminal acetyl (Ac) group can bind in the reverse orientation, 3′ directionality.[12,13] The ability to bind with 3′ directionality was found to be completely correlated with inversion of the C-terminal ring of the ligand, Figure 1, which removes the tail from contact with the groove.[13] The presence of an N-terminal Ac group promotes the inversion of the C-terminal ring through steric interactions of the acetyl and tail. A similar effect is observed when glycine rather than β is used as a spacer in the tail. The deletion of the single methylene group changes the shape of the tail. With the glycine-containing tails the contacts with the floor of the groove are lost, and this leads to an equilibrium between normal and inverted conformations for the last ligand ring.

In all binding modes the affinity of the ligands increases with length up to five ring pairs.[14] Beyond five pairs a mismatch between the shape of the ligands and DNA groove reduces both affinity and specificity. On the basis of recent crystallographic work it appears that the ligands have a curvature that is somewhat too high for optimal contact with the groove.[15] It has been shown that β-alanine incorporated in place of a pyrrole ring can reset the register of the ligand by relaxing the curvature.[16] These β residues have the same sequence specificity as a pyrrole ring when paired with other moieties: β/β pairs and β/Py pairs recognize A•T or T•A, and Im/β and β/Im are specific for G•C and C•G, respectively.[17] The use of β in this way has allowed sites of up to 16 base pairs to be recognized.[18] While the longer ligands have higher affinity and target a small number of sites within a genome, the longer ligands may have reduced cell permeability. Therefore, developing a system of smaller ligands that can cooperatively bind to a long target site may be of benefit for future in vivo uses of these polyamides. Previous work along these lines used a hairpin molecule, which incorporates a β between two sets of rings and a modified tail, to cooperatively bind as a dimer of hairpins.[19]

(12) White, S.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 8756−8765.
(13) Hawkins, C. A.; de Clairac, R. P. L.; Dominey, R. N.; Baird, E. E.; White, S.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **2000**, *122*, 5235−5243.
(14) (a) Kelly, J. J.; Baird, E. E.; Mrksich, M.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 6981−6985. (b) Turner, J. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 7636−7644.
(15) Kielkopf, C. L.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Nat. Struct. Biol.* **1998**, *5*, 104−109.
(16) (a) Trauger, J. W.; Baird, E. E.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6160−6166. (b) de Clairac, R. P. L.; Seel, C. J.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1999**, *121*, 2956−2964.
(17) Turner, J. M.; Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 6219−6226.
(18) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 3534−3535.
(19) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Angew. Chem., Int. Ed.* **1998**, *37*, 1421−1423.

Figure 2. Schematics of the complexes studied. (A) ImPyPy-γ-PyPyPy-Gly-Dp/GGTTT. (B) ImPyPy-γ-PyPyPy-Gly-Dp/CGTTT. (C) ImPyPy-γ-PyPyPy-Gly-Dp/CCAAA. (D) 2:1 ImPyPy-γ-PyPyPy-Gly-Dp/TAACGTTA. (E) 2:1 ImPyPy-γ-PyPyPy-Gly-Dp/TAACTGTTA and 2:1 Ac-ImPyPy-γ-PyPyPy-Gly-Dp/TAACTGTTA. (F) ImPyPy-γ-PyPyPy-Gly-Dp/TAACGGTTA. (G) 2:1 Ac-ImPyPy-γ-PyPyPy-Gly-Dp/ATTGTCAAT. (H) 2:1 Ac-ImPyPy-γ-PyPyPy-Gly-Dp/ATTGTACAAT.

In the present work we examine complexes of ImPyPy-γ-PyPyPy-Gly-Dp with DNA sequences of the CSGTTT type (S = G or C) to determine whether binding with the terminal ring inverted completely removes the specificity of the tail for T,A base pairs, Figure 2A–C. Examination of these complexes also has provided further insight into the interactions of the tails with DNA. In addition, to explore the possibility of expanding the size of sequences that can be recognized on DNA through cooperative binding of hairpins, ImPyPy-γ-PyPyPy-Gly-Dp and AcImPyPy-γ-PyPyPy-Gly-Dp were studied in 2:1 ligand:DNA complexes in which the ligands are close together, Figure 2D–H. Binding in this geometry should allow additional interactions to be engineered that would lead to cooperative binding. The number of base pairs between the ligands was varied to determine the spacing which allows for optimal binding. Complexes were characterized structurally using two-dimensional NMR.

Experimental Section

The ligands studied were synthesized as described previously.[9] The ligand concentrations were determined by UV absorbance, the extinction coefficient used for ImPyPy-γ-PyPyPy-Gly-Dp was 51 300 M^{-1} cm^{-1} at 306 nm and for Ac-ImPyPy-γ-PyPyPy-Gly-Dp was 48 800 M^{-1} cm^{-1} at 312 nm.[10] Synthetic oligonucleotides were purified by reverse-phase HPLC and desalted on Waters' Sep Pak Cartridges. Concentrations of the single strands were measured by UV absorbance, using extinction coefficients calculated from the base sequences of the strands. The single strands were combined in 1:1 molar ratios to create duplex DNA, which was confirmed by NMR. Duplex DNA concentrations used in the various experiments ranged from 1.0 to 2.4 mM, with a total volume of 0.2 mL in Shigemi NMR tubes. The solutions were also 25 mM K_2HPO_4 (pH 7.0) and 0.25 mM EDTA, with 99.96% D_2O (Cambridge Isotopes) or 90% H_2O/10% D_2O as solvent.

Spectra were recorded on a Bruker DRX-500 spectrometer. The duplex DNA samples were titrated to form 1:1 or 2:1 ligand-to-DNA complexes based on OD, and confirmed by 1-D proton NMR experiments. The 1-D spectra were 8192 complex points and an average of 64 scans. The NOESY and TOCSY spectra in D_2O were collected with 64 scans of 1024 complex points in t_2, and 512 t_1 experiments were recorded. The experiments used spectral widths of 5000 Hz in each dimension. Presaturation pulses were used for water suppression on D_2O samples. NOESY experiments in 90% H_2O/10% D_2O were collected with 64 scans and 2048 complex points in t_2, and 256 complex points in t_1. Spectral widths of 10 000 Hz were used. A combination of a 1–1 jump and return sequence and a gradient pulse of 2.5 G/cm for 1 ms during the mixing time were used for water suppression. NOESY experiments were acquired with a mixing time of 200 ms, and the TOCSY experiments used mixing times of 100 ms. Experiments in D_2O were acquired with TPPI, 90% H_2O/10% D_2O experiments were acquired with States-TPPI, and all were taken at 25 °C unless specified otherwise. The data were processed using Felix 97.0 from Biosym, followed by Facelift 2.1 for baseline correction. Molecular modeling of complexes was performed with InsightII by Biosym/MSI. Energy minimization was completed using Discover with an AMBER force field. Restraints were included with a harmonic force constant of 25 kcal/mol/Å^2.

Results

X-ImPyPy-γ-PyPyPy-Gly-Dp with SGTTT Sites. Previous studies of hairpin ligands had always used DNA sequences which contained A,T base pairs in the region contacted by the ligand tails, based on the known preference of -Dp and β-Dp tails for such sequences.[11] Since ligands which contain glycine

in the tails seemed not to contact the groove under any circumstances, we believed that such ligands might have no sequence preference at the base pairs normally contacted by the tails.[13] To initially explore this issue, *in silico* models were generated for complexes of the ligands ImPyPy-γ-PyPyPy-Gly-Dp and AcImPyPy-γ-PyPyPy-Gly-Dp with DNA sequences that placed a G•C base pair at the primary tail contact site. The sequence d(CCT**GGTTT**GG) •d(CC**AAACC**AGG) was used in the two models, and ligands were docked at the highlighted central sites with the C-terminal rings in the inverted conformation. The models suggested that the acetyl group in AcImPyPy-γ-PyPyPy-Gly-Dp would have steric conflicts with the amino group of the terminal guanine of the binding site. For this reason, ImPyPy-γ-PyPyPy-Gly-Dp was chosen to make three complexes of this type with the oligonucleotides: d(C**GGTTT**AA)• d(TT**AAACC**G), henceforth termed [GGTTT]; d(C**CGTTT**AA)• d(TT**AAACG**G), termed [CGTTT]; and d(G**CCAAA**TT)• d(AA**TTTGG**C), termed [CCAAA], Figure 2A−C. The oligonucleotides were designed without G•C base pairs at one end to reduce the possibility that the ligand would bind there as well as at the desired target sites.

Titrations of ImPyPy-γ-PyPyPy-Gly-Dp into a solution of the [GGTTT] oligomer indicated formation of a well-defined complex, in slow exchange with the free DNA in solution. From NOESY data collected for a 1:1 ligand:DNA complex it was apparent that the ligand bound the expected binding site, with ImPyPy contacting 5′-$G_3T_4T_5$-3′ and PyPyPy contacting 5′-$A_{12}A_{13}C_{14}$-3′. The NOEs and line widths of resonances from the C-terminal pyrrole ring of the ligand indicate that this ring undergoes exchange between the normal and inverted conformations, as has been seen in all the other 5′-directional complexes with the glycine- or acetyl-containing ligands. The ImPyPy-γ-PyPyPy-Gly-Dp:[GGTTT] complex is very similar to the previously described complex of ImPyPy-γ-PyPyPy-Gly-Dp with d(CCT**TGTTT**GG)•d(CC**AAACA**AGG), termed [TGTTT], although there appear to be minor differences in the behavior of the C-terminal pyrrole, Py6.[11] In the ImPyPy-γ-PyPyPy-Gly-Dp:[TGTTT] complex, the H3 and H5 of Py6 show NOEs to ligand *N*-methyls, and DNA H1's and H4's, which helped define the conformational equilibrium between normal and inverted geometries for Py6, Figure 3A. For ImPyPy-γ-PyPyPy-Gly-Dp:[GGTTT] there are no clear NOEs from the Py6 protons to the DNA, and the cross-peaks to the *N*-methyls are obviously broader, Figure 3B. We established previously that the chemical shift of the Py6 H3 resonance is upfield of that from H5 in the normal conformation, but the positions are reversed in the inverted conformation.[13] In complexes in which Py6 is undergoing fast exchange an average position is seen, weighted by the relative populations of the normal and inverted conformations. For the ImPyPy-γ-PyPyPy-Gly-Dp:[TGTTT] complex the H3 resonance is slightly upfield of that from H5. For ImPyPy-γ-PyPyPy-Gly-Dp:[GGTTT], the H3 resonance is just downfield of that from H5, but the shift difference is very small. The line widths and shifts indicate that Py6 resonances in the ImPyPy-γ-PyPyPy-Gly-Dp:[TGTTT] complex are near the fast-exchange limit, while in the ImPyPy-γ-PyPyPy-Gly-Dp:[GGTTT] complex the exchange rate is slower, placing resonances more toward intermediate exchange. For the ImPyPy-γ-PyPyPy-Gly-Dp:[TGTTT] complex the similarities in the intensities of the NOEs between the H3 and H5 of Py6 to neighboring amide shows that there are nearly equal populations of the two conformations of Py6 Figure S1A (Supporting Information). Because the H3 and H5 chemical shifts are closer in ImPyPy-γ-PyPyPy-Gly-Dp:[GGTTT] and the lines are broader, their NOE cross-peaks to NH6 overlap, precluding an analysis of their intensities, Figure S1B (Supporting Information). However, based on chemical shifts the population of the inverted conformation is probably somewhat higher in the [GGTTT] complex than in that with [TGTTT].

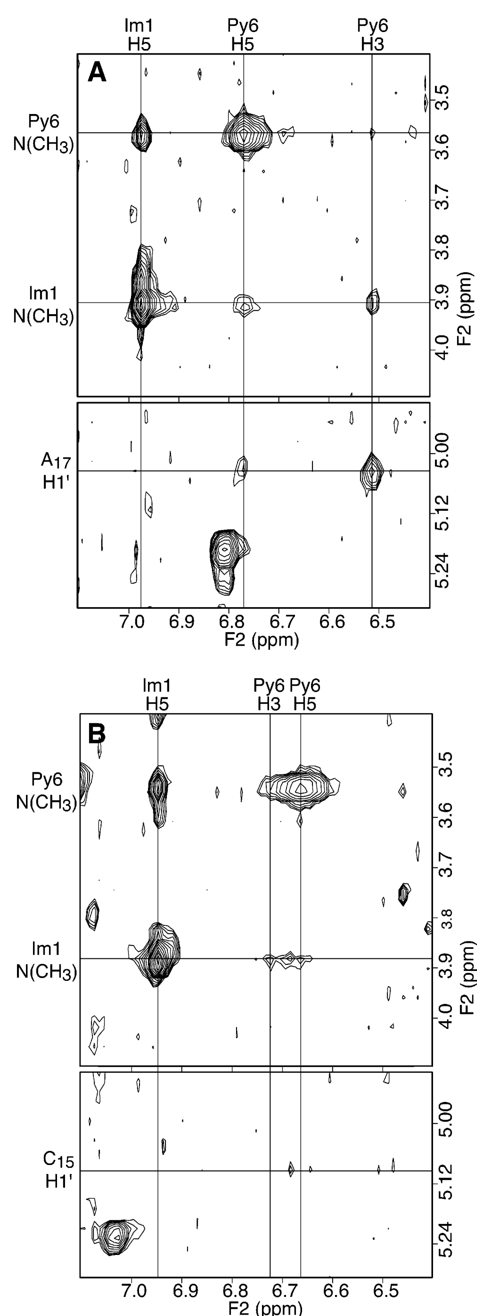

Figure 3. NOESY regions (D_2O, 500 MHz, 25 °C, t_{mix} = 200 ms) showing the Py6 H3 and H5 cross-peaks to ligand *N*-methyls and DNA H1's. (A) ImPyPy-γ-PyPyPy-Gly-Dp/TGTTT. (B) ImPyPy-γ-PyPyPy-Gly-Dp/GGTTT. Note that the H3 and H5 protons of Py6 do not show NOEs to C_{15} H1′.

When ImPyPy-γ-PyPyPy-Gly-Dp was titrated into a solution with [CGTTT], the lines in the 1-D NMR spectra were somewhat broader than in the [GGTTT] complex, the increased broadening probably arising from faster dissociation (weaker

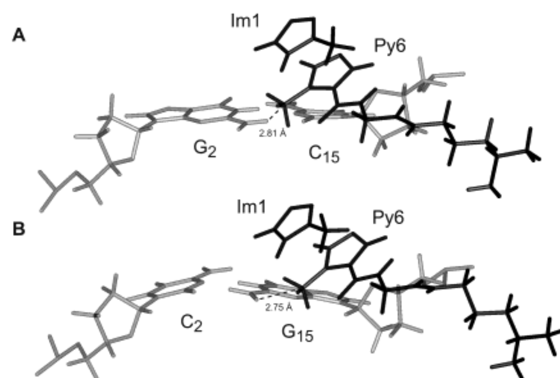

Figure 4. Part of the models of (A) ImPyPy-γ-PyPyPy-Gly-Dp/GGTTT and (B) ImPyPy-γ-PyPyPy-Gly-Dp/CGTTT, showing the Py6's in the inverted conformation near the G_2−C_{15} base pair of (A) and C_2-G_{15} base pair of (B). The distances between the carbon of the Py6 *N*-methyl groups to the closest proton of the amino groups of the guanines of the G_2-C_{15} base pair of (A) and C_2-G_{15} base pair of (B) are shown. The distances are similar in the two models.

binding). The 2-D spectra could be assigned when data were collected at 15 °C to reduce the broadening. At 15 °C the [CGTTT] complex is very similar to [GGTTT]. Py6 exchanges between normal and inverted conformations, and the H3 and H5 resonances are very close in chemical shift, but with the H3 resonance very slightly upfield from that from the H5, and again no distinct NOEs to the DNA. *In silico* models of ImPyPy-γ-PyPyPy-Gly-Dp with [GGTTT] and [CGTTT] were built and examined to try to identify the basis for the affinity difference, Figure 4; however, there were no apparent steric conflicts, suggesting that the difference arises from more subtle conformational differences in the DNA or complex.

In titrations of the [CCAAA] oligomer with ImPyPy-γ-PyPyPy-Gly-Dp the DNA resonances broadened, but did not sharpen after stoichiometric addition of ligand. This indicates that a well-defined complex with CCAAA did not form, either due to low affinity or exchange among multiple binding sites. The target binding site on the oligonucleotide was such that the ligand should have bound in the 3′ orientation. These observations are consistent with previous work indicating that such complexes are less favorable than those bound with 5′ directionality.[20]

Complexes with Two ImPyPy-γ-PyPyPy-Gly-Dp Ligands. An *in silico* model was generated of d(CCTAACGTTAGG)$_2$, termed [TAACGTTA] with two ImPyPy-γ-PyPyPy-Gly-Dp ligands symmetrically bound with 5′ directionality and the C-terminal pyrroles in the inverted conformation. The ligands, placed in standard hairpin conformations, had the N-terminus of one ligand in close proximity to the opposite ligand's C-terminus, Figure 2D. No obvious steric conflicts were present, suggesting that the ligands should be able to bind to immediately adjacent sites. A sample of the [TAACGTTA] oligomer was prepared, and titrated with the ImPyPy-γ-PyPyPy-Gly-Dp ligand. During the course of the titration, the DNA resonances broadened and remained broad even at a 2:1 stoichiometry, Figure S2A (Supporting Information). The initial broadening could be explained by ligand exchange between the two available binding sites on each oligonucleotide at an intermediate rate; however, were this the sole source of broadening, the spectra should have sharpened as all sites became occupied. The large line widths made it impossible to assign the resonances of the complex and, hence, to determine its structure.

(20) Hawkins, C. A. Ph.D. Dissertation, University of California, Berkeley, 2001.

To determine whether spacing between the ligands was a critical issue a 2:1 ImPyPy-γ-PyPyPy-Gly-Dp complex with d(CC**TAACTGTTA**GG)•d(CC**TAACAGTTA**GG), termed [TAACTGTTA], was created, Figure 2E. This sequence incorporates a T•A base pair spacer between the two ligand binding sites on the oligonucleotide. In a titration of DNA with ligand the spectra broadened and then sharpened again near a 2:1 ratio, Figure S2B (Supporting Information). The 2:1 complex was well-behaved, and 2D spectra were collected, Figure 5A. Since the complex is symmetric except for the central T•A base pair, several of the protons are degenerate in chemical shift with the equivalent protons in the other half of the complex; nonetheless, the structure of the complex is well-defined by the data.

An unexpected finding in this complex is that the C-terminal pyrrole rings of both ligands populate both normal and inverted conformations (based both on NOE contacts and chemical shifts as discussed above). Binding with the pyrroles in the normal state and the tails in an extended conformation, Figure 1B, is sterically impossible in this complex since the extended conformation would lead to complete overlap of the tail with the imidazole ring of the second ligand. Hence, both of the tails must be in a "kinked" conformation, Figure 1C. The lack of NOE contacts between the protons of the Gly-Dp segment of the tail and the minor groove (NOEs which are seen for β-Ala-Dp ligands) in all complexes indicates that such kinked conformations are the norm for Gly-containing tails.

An additional 2:1 complex of ImPyPy-γ-PyPyPy-Gly-Dp was made using d(CC**TAACGGTTA**GG)•d(CC**TAACCGTTA**GG), termed [TAACGGTTA], Figure 2F. This oligonucleotide again contained a single base pair separating the two hairpins, but in this case a G•C base pair. NOEs between the first three rings of the ligand and the minor groove clearly indicated formation of the intended complex. However, the protons of the PyPyPy segments of the ligands were broad, as were the H1′ protons of their contacted DNA strand. The amide-to-H1′ region exhibits this best in that NH1, NH2, NH3 are sharp with strong cross-peaks to ligand H3's and DNA H1′ protons but the NH4, NH5, and NH6 are broad, and the cross-peaks are ambiguous, Figure 5B. The localization of the broadening shows that it arises from a slow conformational exchange process within the complex. The most likely candidate is the exchange of Py6 between normal and inverted geometries, with the rate slowed relative to the complexes described above to move into the intermediate exchange regime.

AcImPyPy-γ-PyPyPy-Gly-Dp Double-Hairpin Complexes. Since previous work had indicated that the introduction of an N-terminal acetyl group affected ligand orientation and bound conformation, we explored double hairpin complexes of this ligand as well.[11,12] The ligand was modeled in complexes with [TAACTGTTA], Figure 2E, and d(CC**ATTGTCAAT**GG)•d(CC**ATTGACAAT**GG), termed [ATTGTCAAT], Figure 2G. The latter has 3′-directional binding sites, and the C-terminal pyrrole rings were modeled solely in the inverted conformation, consistent with previous experimental observations.[13] The modeling initially indicated that a single base pair separating the sites should allow space for the acetyl groups in double-hairpin complexes.

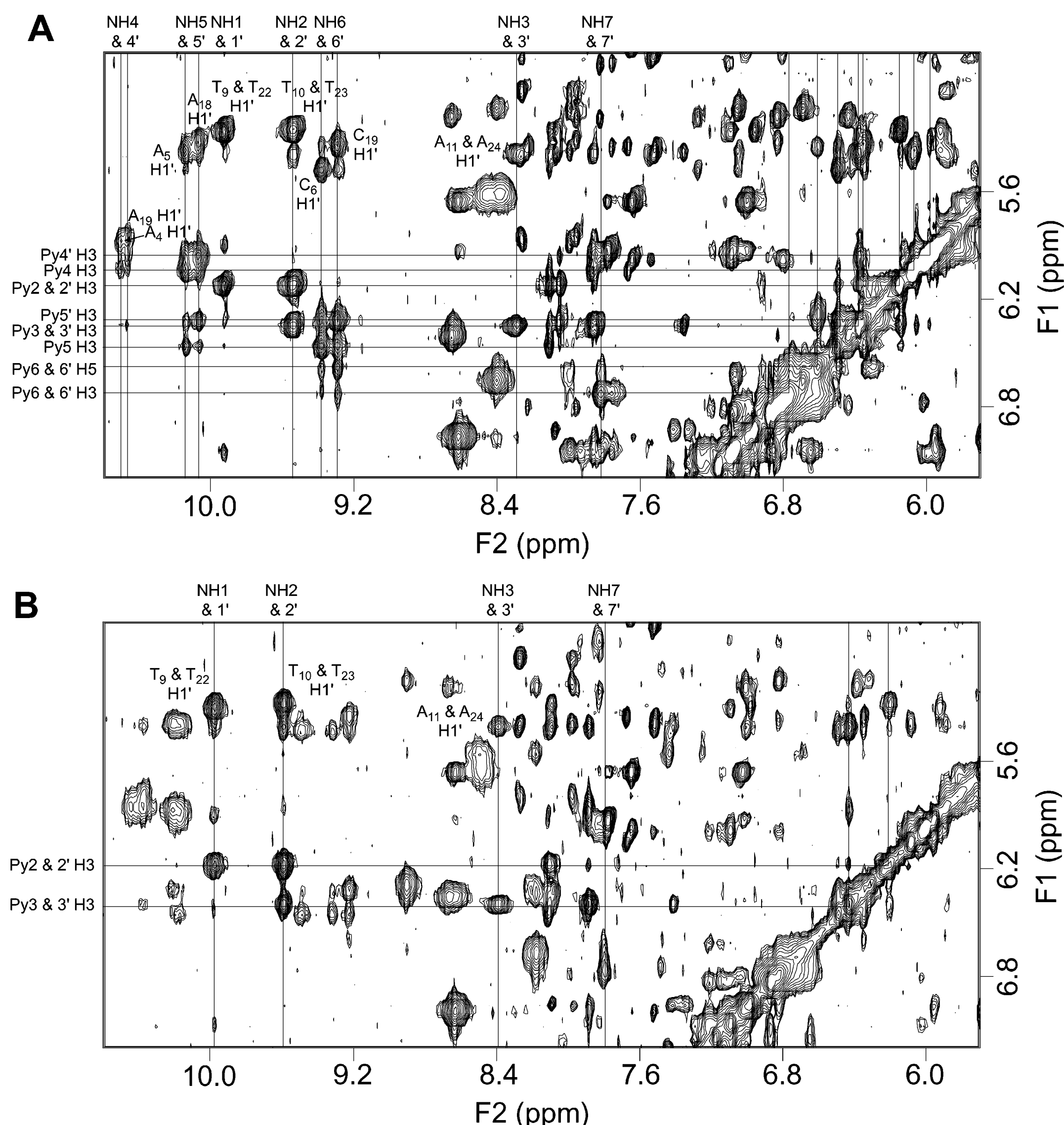

Figure 5. 1-D NOESY regions (90% H_2O/10% D_2O, 500 MHz, 25 °C, t_{mix} = 200 ms) of the ligand amides to pyrrole H3's and DNA H1's. (A) 2:1 ImPyPy-γ-PyPyPy-Gly-Dp/TAACTGTTA complex. (B) 2:1 ImPyPy-γ-PyPyPy-Gly-Dp/TAACGGTTA complex. Note that all of the cross-peaks are sharp in (A) but in (B) the resonances from the PyPyPy half of the ligand are broad and unassignable.

The 2:1 AcImPyPy-γ-PyPyPy-Gly-Dp:[TAACTGTTA] complex was well-behaved. As in other 5′-directional complexes of this ligand the C-terminal pyrrole rings exchange between the normal and the inverted conformations.[13] Each of the Py6 H3 and H5 protons (with nearly equivalent shifts for the two molecules in the almost symmetric complex) has a single chemical shift, which indicates that both C-terminal rings are in fast exchange between the two conformations. The H3 protons of the C-terminal pyrrole rings are downfield of the H5 protons by approximately 0.7 ppm. This indicates that there is a greater population of the rings in the inverted conformation than the normal conformation. Unfortunately, this cannot be further evaluated by the relative intensities of the NOEs between the amide protons N-terminal to the C-terminal rings and the H3 and H5 protons of the C-terminal ring because the H5 protons are broad, and thus have weak cross-peaks even to the nearby *N*-methyls. This may arise from the fact that the chemical shift difference between the inverted and normal conformation is greater for the H5 protons than for the H3 protons. For the particular flipping rate of the Py6 ring the H3 protons may be near the fast-exchange limit, while the H5 protons remain broadened by the exchange process. Alternatively there could be an as-yet uncharacterized dynamic process in one of the conformational states.

The 2:1 AcImPyPy-γ-PyPyPy-Gly-Dp:[ATTGTCAAT] complex was less well-behaved. The NOESY spectra exhibited chemical exchange cross-peaks, indicating that the ligand was binding at more than just the two expected hairpin sites. In the amide proton region almost every resonance has exchange cross-peaks to two other resonances indicating slow exchange to other sites. The cross-peak pattern is extremely complicated due to the combination of direct NOEs, exchange cross-peaks, and exchange-transferred NOEs making it impossible to completely assign the spectra.

In the models it was apparent that the acetyl groups in the 5′-orientation complex extended somewhat toward the walls of the minor groove, while in the 3′-orientation complex they both extended toward the center of the groove, Figure 6. Although the acetyl groups did not overlap in the model, it seemed possible that an unfavorable interaction of these might be

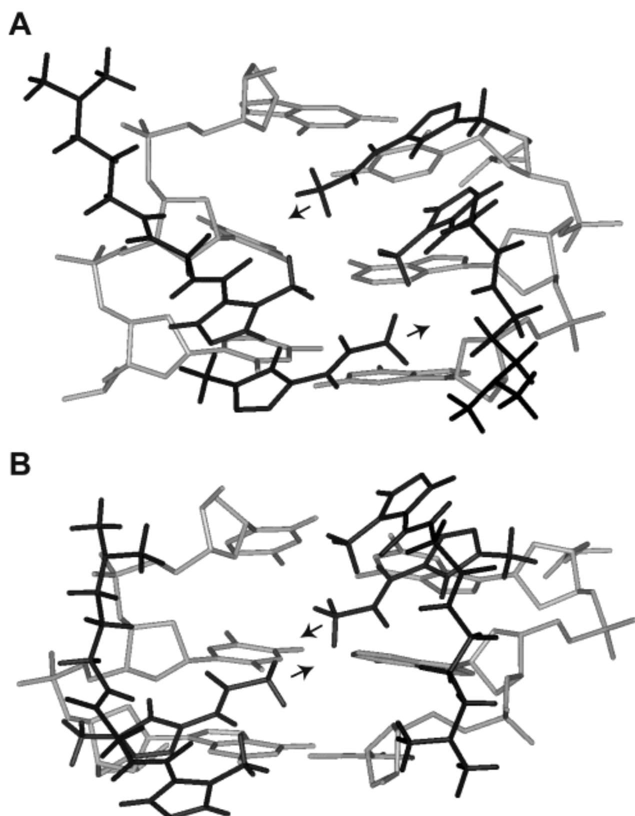

Figure 6. Part of the models of (A) 2:1 Ac-ImPyPy-γ-PyPyPy-Gly-Dp/TAACTGTTA complex and (B) 2:1 Ac-ImPyPy-γ-PyPyPy-Gly-Dp/AT-TGTCAAT complex, showing the area of the acetyl groups of the two ligands. Note that the two acetyls, as noted by the arrows, are turned away from one another in (A) but turn in toward each other in (B). The hydrogens on the DNA have been removed for clarity.

responsible for the observed exchange behavior. To test this possibility, a further 2:1 complex of AcImPyPy-γ-PyPyPy-Gly-Dp with d(CC**ATTGTACAAT**GG)$_2$, termed [ATTGTACAAT] was prepared which has an additional base pair between the two ligand binding sites, Figure 2H. This complex was well-behaved and completely symmetric as expected. NOEs indicate that the Py6 rings are only in the inverted conformation, consistent with previously studied 1:1 3′-directional complexes of this ligand.[13]

Discussion

To apply solid-phase synthesis methods to polyamide ligands it was convenient to begin with a single amino acid attached to resin, then to couple activated building blocks (pyrrole, imidazole, or alkyl amino acids), and at the last step to cleave from the resin with an alkylamine.[9] This process incorporates the starting amino acid into the ligand tail, extending it relative to the natural products distamycin and netropsin from which the polyamides are derived.[1] Footprinting and affinity cleavage studies done to characterize ligands with the extended tails indicated that a glycine-Dp tail decreased affinity and specificity; however, β-alanine-Dp increased affinity while maintaining specificity relative to the shorter tail without any amino acid.[10] Modification of the N-terminus by addition of an acetyl or -Dp group lead to indications of binding with the orientation reversed relative to all complexes previously characterized.[12,13] The NMR studies demonstrated that reversed binding was associated with an inversion of the C-terminal ring in the groove, which directs the tail out away from the DNA, indicating that contacts between the tails and DNA are critical features determining binding orientation. The tendency for the ligands to bind in the reversed orientation, with the last ring inverted, was enhanced in ligands which contain a glycine linker in the tail which, together with the lower affinity of reversed binding, can be attributed to the loss of favorable interactions between the tail and the bottom of the groove in glycine-linked compounds.

Quantitative footprinting studies defined the specificity of the β-Ala-Dp tails for A,T base pairs.[11] In the present work complexes of ImPyPy-γ-PyPyPy-Gly-Dp with [***G***GTTT] and [***C***GTTT] were prepared for comparison with the previously described [***T***GTTT] complex to determine whether there is specificity of Gly-Dp at the base pair normally contacted by the ligand tail. The features of all three of these complexes are very similar, with the expected 5′ directionality of binding and an equilibrium between normal and inverted conformations of the Py6 ring. In no case is there any evidence for contacts between the tail and the bottom of the groove. Ligands with β-Dp tails fail to bind with an A,T-to-G,C substitution at the same site where the Gly-Dp tail is indifferent to sequence, indicating a drop in binding constant of over 200-fold.[11] The small apparent difference in affinity between [***G***GTTT] and [***C***GTTT] (evidenced by the difference in broadening during titrations) probably results from some coupling of DNA sequence, conformation, and binding affinity rather than any direct contacts between the altered base pair and ligand. Such differences are analogous to small differences in affinity, resulting from A•T to T•A changes at base pairs contacted by Py/Py ring pairs of the ligands.[21] It has not yet been possible to clarify the structural basis for such differences. Although the lower affinity of the Gly-Dp tail ligands, due to lost contacts between the tail and DNA, may be viewed as a negative feature, removing the requirement for two A,T base pairs for the tails to interact with in target sequences may be an advantage for recognition of very G•C rich sequences.

Favorable interactions between polyamides have been exploited to generate cooperative binding of ligands as a mechanism for extending the recognized binding site. When this was done with polyamide hairpins, the tails had to be short to allow the ligands to fit closely together.[19] In the single report on this type of binding mode the hydroxyethyl amide tail was paired with an internal β-ala residue and contacted an A,T base pair. A possible alternative basis for interaction of ligands would be through derivatized tails, which could point away from the DNA, allowing a wide variety of mechanisms to be used for favorable interactions between ligands. To understand the constraints imposed by steric interactions of the aromatic portions of the ligands on ligand binding at proximal sites we have examined a number of complexes with two potential ligand binding sites at different spacings. Computer modeling suggested that it would be possible to bind ImPyPy-γ-PyPyPy-Gly-Dp ligands at immediately adjacent sites, with the Py6 ring in an inverted conformation. However titrations of this ligand into a sample of a DNA oligomer containing the double site [TAACGTTA] yielded broad lines even at a 2:1 stoichiometry. This indicates that both sites cannot be occupied with the same affinity as an individual site [CGTTA]. However, when the spacing was

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

increased to a single base pair, [TAAC**T**GTTA] or [TTAC**G**GTTA], then well-defined complexes did form. Surprisingly, the observed NOEs within the ligand indicate that both normal and inverted conformations of the Py6 ring are sampled, with relative populations very similar to what was observed for an individual site. If bound in an extended conformation, the tail would interfere with binding of neighboring ligand; hence, the Gly-Dp tail must be kinked when the Py6 ring is in the normal conformation. A very similar complex forms on the [TAAC**T**GTTA] oligomer with the N-terminally acetylated ligand, although intramolecular interactions lead to a greater population of the inverted state of Py6 in this case. When the sites are reversed, giving the oligomer [ATTGTCAAT] which requires 3′ directionality of ligand binding, then multiple exchanging complexes are formed. Insertion of an additional spacing pair to give [ATTG**TA**CAAT] does yield a well-defined, symmetric complex with AcImPyPyPy-γ-PyPyPy-Gly-Dp that has both ligands bound at the expected reversed site. These data indicate that the introduction of the acetyl group does affect the "in-out" equilibrium of the last ring, as in previously described single-ligand complexes, but also increases the spacing requirement when binding with 3′ directionality.

In all of the double-hairpin complexes the ligand tails are sufficiently close that they could be modified to allow favorable interactions, for example attachment of a metal chelating group on each. By combining an increased number of rings in the ligands (8- and 10-ring hairpins have been characterized previously), together with additions to stimulate cooperative binding, it should be possible to target specific sites of 10 or more base pairs.

Acknowledgment. We thank the National Institutes of Health for grant support (GM43129 to D.E.W. and GM27681 to P.B.D.) and acknowledge instrumentation grants from DOE (DE FG05-86ER75281) and NSF (DMB 86-09305 and BBS 87-20134).

Supporting Information Available: NOEs between the H3 and H5 of Py6 to neighboring amides, Figure S1A, and 1-D titrations, Figure S1B (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA020335B