VOLUME 122, NUMBER 22
JUNE 7, 2000

JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Controlling Binding Orientation in Hairpin Polyamide DNA Complexes

Cheryl A. Hawkins, Rafael Peláez de Clairac, Raymond N. Dominey, Eldon E. Baird, Sarah White, Peter B. Dervan, and David E. Wemmer*

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

Received January 10, 2000

Abstract: The effects of N-terminal acetylation and C-terminal tail structure on the orientation of binding of imidazole/pyrrole polyamide DNA ligands has been investigated. We find that N-terminal acetylation leads to an intramolecular steric clash for hairpin ligands bound in the minor groove, promoting a rotation of the spatially close C-terminal pyrrole ring. This in turn leads to loss of contacts between the tail and the groove, removing the preference for 5′-to-3′ orientational binding typical of this class of ligand. Similarly, introduction of a glycine linker into the tail leads to a direct steric clash with the groove, again promoting rotation of the attached ligand ring. The effects of acetylation and a glycine in the tail are additive. The implications for the design of sequence-specific ligands are discussed.

Introduction

Over the past several years a design has emerged for sequence specific DNA ligands which target the minor groove.[1−7] The ligands are related to the natural product distamycin, Figure 1, and comprise pyrrole (Py), hydroxypyrrole (Hp), or imidazole (Im) rings, linked together by amide bonds or short alkyl chains.[2−7] In the DNA bound complexes two sets of rings (which can be from separate molecules, or parts of a single one) stack side-by-side across the minor groove in an antiparallel arrangement.[3,5,6] Base-specific recognition occurs through hydrogen bonding from the imidazole ring nitrogens to the amino groups of guanosines or from the OHs of the hydroxypyrroles to the carbonyls of thymidines, in either case with a pyrrole against the opposite strand.[4−7] It has been found that pyrrole/pyrrole pairs are degenerate for A-T or T-A combinations.[8] To achieve optimum affinity and specificity it has been shown that stacked ring systems should be linked head to tail by a propyl chain (γ-amino butyric acid, abbreviated γ), causing the ligand to fold back and form a "hairpin" structure.[6] In addition,

(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. A. Chem. Soc.* **1990**, *112*, 1393−1399. (c) Fagan, P. A.; Wemmer, D. E. *J. A. Chem. Soc.* **1992**, *114*, 1080−1081.

(2) (a) Dwyer, T. J.; Geierstanger, B. H.; Bathini, Y.; Lown, J. W.; Wemmer, D. E. *J. A. Chem. Soc.* **1992**, *114*, 5911−5919. (b) Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. A. Chem. Soc.* **1992**, *114*, 8783−8794. (c) 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. (d) Wade, W. S.; Mrksich, M.; Dervan, P. B. *Biochemistry* **1993**, *32*, 11385−11389.

(3) (a) Mrksich, M.; Dervan, P. B. *J. A. Chem. Soc.* **1993**, *115*, 2572−2576. (b) Geierstanger, B. H.; Dwyer, T. J.; Bathini, Y.; Lown, J. W.; Wemmer, D. E. *J. A. 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.

(4) (a) Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Science* **1994**, *266*, 646−650. (a) Mrksich, M.; Dervan, P. B. *J. A. Chem. Soc.* **1995**, *117*, 3325−3332. (c) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. A. Chem. Soc.* **1996**, *118*, 8198−8206. (d) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. A. 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.

10.1021/ja0001198 CCC: $19.00
Published on Web 05/18/2000

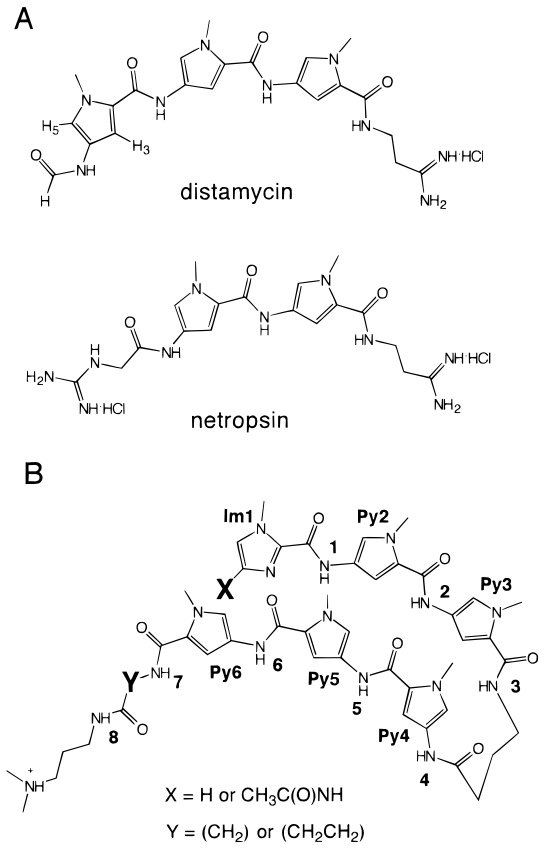

Figure 1. (A) Structures of distamycin and netropsin, the ring proton numbering is shown on the first ring of distamycin. (B) Ligands used in this work, showing the numbering of the aromatic rings and amide protons.

paralleling the natural product and also for synthetic convenience, the C-terminal end of the molecule has a charged tail. This is now typically a β-alanine (β) linked by an amide to a dimethylaminopropyl (Dp) group.[9] Such tails contribute to affinity, and have a substantial preference for contacting A-T base pairs in the DNA.[10] The affinity of ligands increases with the number of pyrrole or imidazole rings up to five rings; further rings decrease both affinity and specificity due to a mismatch between curvature of the ligand and DNA groove.[11] Ligands with four or five rings per module (eight or 10 total) reach affinities of 10^9, have good sequence discrimination, and have been shown to compete with proteins for binding sites on DNA.[12,13] These features have generated substantial interest in such compounds for both biochemical and, potentially, therapeutic applications. In the present work we describe experiments which further our understanding of the specificity of these ligands and which suggest new options for recognizing some sequences.

Early in the study of complexes of distamycin it was found that there is a strong orientational specificity in binding.[1] In particular, when distamycin binds to a sequence with three or four adenosines on one strand, the N-terminal formyl end of the ligand is always directed to the 5′ end of the run of adenosines (referred to henceforth as 5′ directional binding). In binding to a DNA oligomer containing 5′-AAAA-3′, the orientational preference was estimated to be ≥20:1. Netropsin, Figure 1, was also studied in complex with the 5′-AAAA-3′ sequence, but the orientation ratio was then only about 3:2.[14] The most obvious difference between distamycin and netropsin is the fact that netropsin has two alkyl tails, while distamycin has just one. When side-by-side 2:1 complexes of distamycin were discovered the orientational binding phenomonon remained. Distamycin was always bound with the formyl group at the 5′ end of any repeated adenosine sequence.[1] Modeling studies did not generate an obvious basis for this effect. The synthetic ligand ImPyPy-Dp was shown to bind to sites of the type WGWCW (where W is A-T or T-A).[15] However there was no indication of binding to WCWGW, which would require just a reversal of orientation, again indicating a strong energetic preference for one specific orientation in binding.

In the course of developing solid-phase synthesis methods for this class of ligands, it became clear that starting with an amino acid linked to resin would be convenient and that the final release could then be done with an organic base (dimethylaminopropyldiamine = Dp) to give a final amino acid-Dp tail.[9] In early studies both glycine and β-alanine were used as the resin linker. When the ligands generated were characterized by quantitative footprinting, it was noted that ligands with the glycine tail had lower affinity and somewhat reduced specificity.[16] Another seeming unrelated observation from early studies was that introduction of an acetyl (Ac) group to the ligand ImPyPy-Dp (in analogy to the formyl of distamycin, but synthetically more convenient to introduce and more chemically stable) to give Ac-ImPyPy-Dp lead to a dramatic drop in affinity.[17] When acetylation and the gly spacer in the tail were combined in a hairpin ligand, and complexes were studied by NMR and affinity cleavage, it became apparent that binding was occurring "backward" or 3′ directional, the ligand running 3′-to-5′ along the contacted strand of DNA.[18] In the present work we describe detailed NMR studies of a series of ligands with the same ring system, ImPyPy-γ-PyPyPy, but different combinations of H- or Ac- at the head, and gly-Dp or β-ala-Dp in the tail, Figure 1. These studies show that a combination of

(5) (a) Mrksich, M.; Dervan, P. B. *J. A. Chem. Soc.* **1993**, *115*, 9892−9899. (b) Dwyer, T. J.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. A. Chem. Soc.* **1993**, *115*, 9900−9906. (c) Mrksich, M.; Dervan, P. B. *J. A. Chem. Soc.* **1994**, *116*, 3663−3664. (d) Chen, Y. H.; Lown, J. W. *J. A. Chem. Soc.* **1994**, *116*, 6995−7005.

(6) (a) Mrksich, M.; Parks, M. E.; Dervan, P. B. *J. A. 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. A. 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. A. Chem. Soc.* **1999**, *121*, 260−261.

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

(9) Baird, E. E.; Dervan, P. B. *J. A. Chem. Soc.* **1996**, *118*, 6141−6146.

(10) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. A. Chem. Soc.* **1999**, *121*, 1113−1120.

(11) Kelly, J. J.; Baird, E. E.; Mrksich, M.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 6981−6985.

(12) Trauger, J. W.; Baird, E. E.; Mrksich, M.; Dervan, P. B. *J. A. Chem. Soc.* **1996**, *118*, 6160−6166.

(13) (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.

(14) Fagan, P. A. Ph.D. Thesis, University of California Berkeley, 1996.

(15) 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.; Renzeperis, D.; Marky, L. A.; Singh, S.; Kollman, P. *Structural Biology: The State of the Art*; Adenine Press: New York, NY, 1993; pp 301−323.

(16) Parks, M. E.; Baird, E. E.; Dervan, P. B. *J. A. Chem. Soc.* **1996**, *118*, 6147−6152.

(17) Wade, W. S. Ph.D. Thesis, California Institute of Technology, 1989.

(18) White, S.; Baird, E. E.; Dervan, P. B. *J. A. Chem. Soc.* **1997**, *119*, 8756−8765.

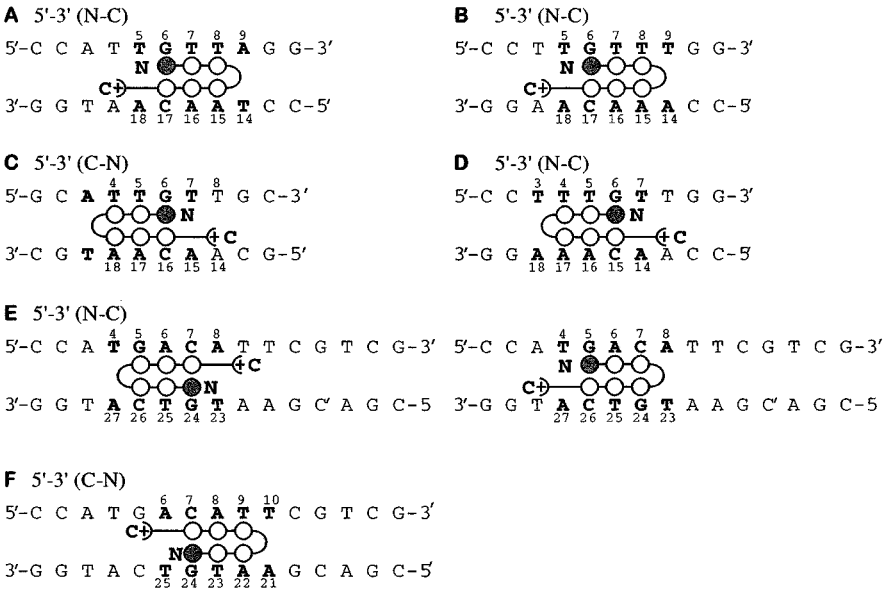

Figure 2. Schematic representation of binding modes. Shaded circles represent imidazoles, and open circles, pyrroles. (A) Site bound by ImPyPy-γ-PyPyPy-β-Dp and Ac-ImPyPy-γ-PyPyPy-β-Dp on *TGTTA*. (B) *TGTTT* site bound by Ac-ImPyPy-γ-PyPyPy-β-Dp, Ac-ImPyPy-γ-PyPyPy-gly-Dp, and ImPyPy-γ-PyPyPy-gly-Dp. (C) Ac-ImPyPy-γ-PyPyPy-β-Dp in complex with *ATTGT*. (D) Site recognized by Ac-ImPyPy-γ-PyPyPy-gly-Dp on *TTTGT*. (E) First figure is the single mismatch site where ImPyPy-γ-PyPyPy-β-Dp binds to *TGACATT*. The second figure shows an equivalent site which not recognized by the ligand. (F) *TGACATT* site bound by both Ac-ImPyPy-γ-PyPyPy-β-Dp and Ac-ImPyPy-γ-PyPyPy-gly-Dp.

ligand−ligand and ligand−DNA interactions control the affinity and orientation of binding.

Experimental Section

The ligands used in this work were synthesized and purified as described previously.[9] Ligand concentrations were determined by UV absorbance assuming an extinction coefficient of 52 000 M^{-1} cm^{-1} at 304 nm. Oligonucleotides prepared on an automated synthesizer, purified by HPLC, and then desalted using a SepPak (Waters). Concentrations of the single strands were measured by UV absorbance, applying extinction coefficients caculated from the base composition. Single strands were mixed to make duplex based on these concentrations. Complete formation of duplex was confirmed by NMR. Duplex DNA concentrations used for titrations and structural studies by NMR ranged from 0.8 to 3.9 mM, using a volume of 0.2 mL. in Shigemi NMR tubes. Solutions were 25 mM K_2HPO_4 (pH 7.0), 0.25 mM EDTA in "100%" D_2O (Cambridge isotopes) or 90% H_2O/10% D_2O.

NMR spectra were collected on either a G.E. GN-Omega 500 or a Bruker AMX 600 spectrometer. Titrations were followed by running a 1D spectrum after each addition, with 2D spectra collected at the final 1:1 stoichiometry. NOESY and TOCSY spectra run in D_2O had 64 or 80 scans averaged for 480 to 512 t_2 values, collecting 1024 complex points in t_2 over a 10 ppm spectral width. The residual solvent signal was saturated by low power irradiation during the relaxation delay and mixing time. NOESY experiments in H_2O were collected for 480−512 t_1 values, with 2048 complex points in t_2 and a spectral width of 20 ppm. A 1−1 jump and return sequence was used to suppress the solvent signal. The mixing time was 200 ms in NOESY experiments, and 40−100 ms in TOCSY experiments. Natural abundance $^1H-^{13}C$ HMQC spectra were recorded on the 600 MHz spectrometer with 190 t_1 values, with 1024 complex t_2 points, averaging 192 scans per value. All experiments were acquired with TPPI to generate quadrature in the indirect dimension. The temperature was 25 °C unless otherwise specified. NMR data processing was done with Felix 95.0, and modeling with InsightII, both from Biosym/MSI.

Results

ImPyPy-γ-PyPyPy-β-Dp Complex. As the ligand ImPyPy-γ-PyPyPy-β-Dp was titrated into a solution of the DNA oligomer d(CCATTGTTAGG)•d(CCTAACAATGG), abreviated *TGTTA* henceforth, a single set of new resonances appeared in slow exchange with free DNA. Conversion from free DNA to complex was complete at a 1:1 ligand:DNA ratio. NOESY data were collected in both H_2O and D_2O solution for characterization. In accordance with previous studies of hairpin ligands,[6c] binding occurs in the minor groove at the site indicated schematically in Figure 2a. NOEs are observed from the ligand aromatic H3s and amides to DNA H1′s, H4′s, and H2′s. Similarly ligand H5′s and *N*-methyls contact DNA H4′s. The hairpin geometry of the ligand is further confirmed by NOEs between H3 protons of rings stacked across the groove, as well as by analogous contacts between H5s and *N*-methyls. In this complex the imidazole ring is positioned over the amino group of guanosine, forming a hydrogen bond. The β-ala-Dp tail lies in the minor groove, making contact with A-T base pairs, Figure 3A. The β-ala residue contacts the A18 H2, while the Dp propyl contacts A19 H2, and the methyls contact A3 H2. Further contacts are seen between the tail and the H1's of the residues in this segment. All features of this complex are exactly analogous to the previously described hairpin complex of ImPyPy-γ-PyPyPy-Dp, except that the contacts of the tail are extended by one base pair as might be expected from insertion of the β-ala linker.[16] Each half of the ligand runs 5′-to-3′ along the contacted DNA strand, although an equivalent sequence in the opposite orientation is available.

Ac-ImPyPy-γ-PyPyPy-β-Dp Complexes. Titration of Ac-ImPyPy-γ-PyPyPy-β-Dp into the *TGTTA* oligomer indicated complex formation, but more resonances were present than expected for a single form of complex. NOESY data showed that the major form of complex corresponded to binding at the same site as the nonacetylated version of the ligand, Figure 2a. Although the second form of complex present could not be clearly identified from NOEs, the resonances which appeared to be affected included the A3-T4-T5 segment, and examination of the sequence suggested this site might have the orientation of the ligand reversed. To examine the complexes being formed in more detail, two new oligomers were made such that part of one of the potential binding sites was eliminated. The oligomers studied were d(CCTTGTTTGG)•d(CCAAACAAGG), desig-

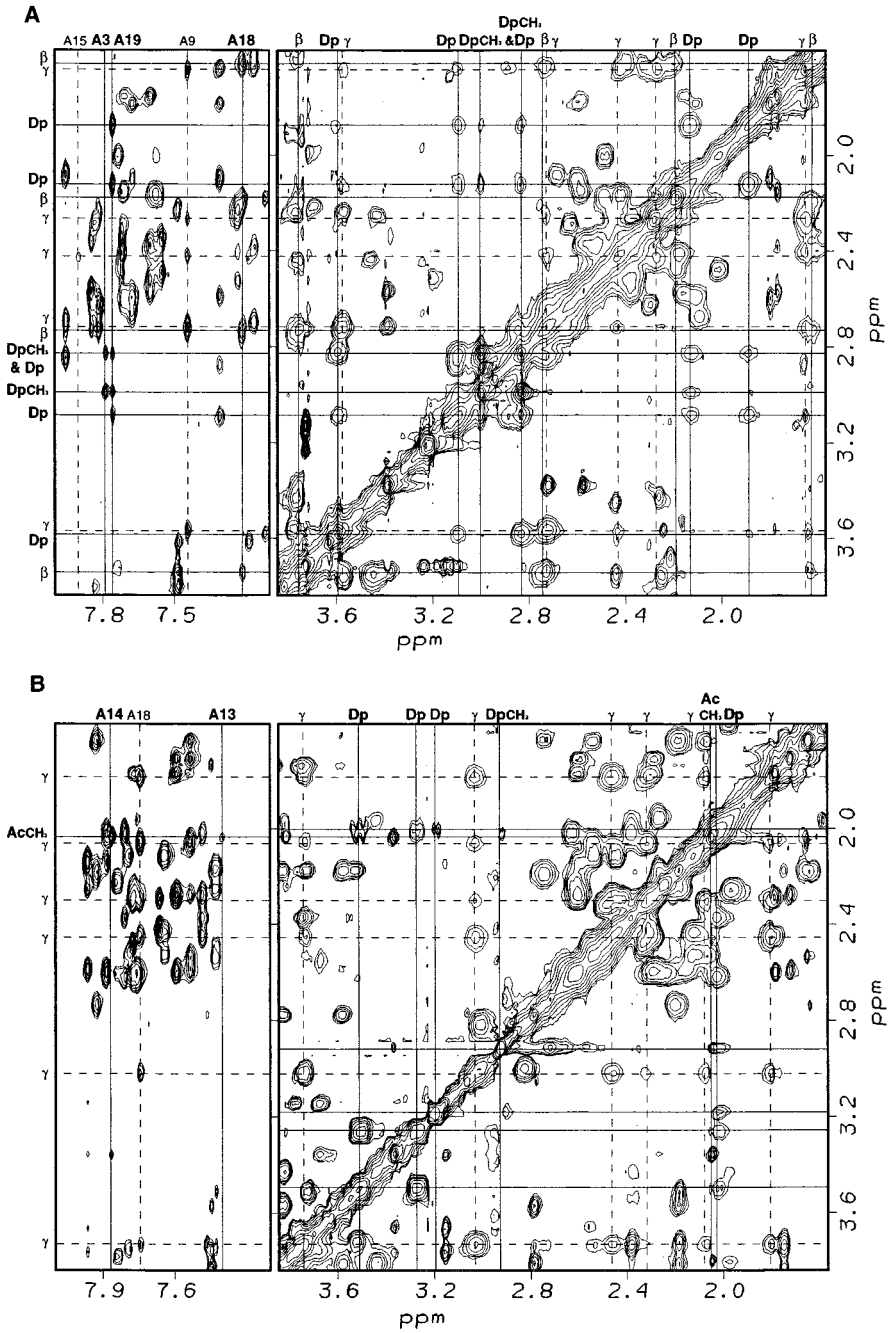

Figure 3. NOESY regions of ligand showing cross-peaks among methylene protons of the ligand, and between ligand methylene and adenine H2′s (D_2O, 500 MHz, 25 °C, τ_{mix} = 200 ms). (A) ImPyPy-γ-PyPyPy-β-Dp complex with *TGTTA*. NOEs between adenine H2′s and γ-linker, and β-ala and Dp of the tail methylenes are present. (B) Ac-ImPyPy-γ-PyPyPy-gly-Dp complex with *TTTGT*. γ-linker methylenes have NOEs with adenine H2′s, but the glycine and Dp methylenes of the tail do not.

nated *TGTTT*, and d(GCATTGTTGC)•d(GCAACAATGC) designated *ATTGT*. In titrations of both oligomers it was apparent that one of the binding sites had been eliminated, resulting in simpler spectra.

The complex of *TGTTT* with Ac-ImPyPy-γ-PyPyPy-β-Dp showed intermolecular contacts indicating a binding site analogous to one of the forms observed with *TGTTA*, as shown schematically in Figure 2b. Most resonances of the complex were sharp, and NOEs between ligand and DNA followed typical patterns, verifying unambiguously the site of ligand binding. However, unlike most complexes of this class studied, the resonances of the C-terminal pyrrole (Py_6) and those of the amide and methyl protons of the acetyl, and the β-linker were substantially broadened. Evidence presented below indicates that this broadening arises from exchange between the two possible orientations of this ring with respect to the rest of the ligand and DNA, with significant populations of both orientations and exchange on a millisecond time scale. In one orientation the tail extends along the groove, while in the other it is pointed up, away from the groove.

Ac-ImPyPy-γ-PyPyPy-β-Dp also formed a well-defined complex with the *ATTGT* oligomer. In this case NOE contacts showed clearly that the bound orientation was reversed from previous complexes—the N-terminus of the ligand at the 3′ end of the set of contacted bases, Figure 2c. The basic pattern of NOEs was similar to the *TGTTA* complex, the pyrrole H3 protons contacting sugar H1′ and H4's, and the γ linker contacts the A3 H2. The intraligand NOEs typical for 3′ directional binding vs 5′ directional binding are summarized in Figure 4. However, in stark contrast to the nonacetylated ligand complex

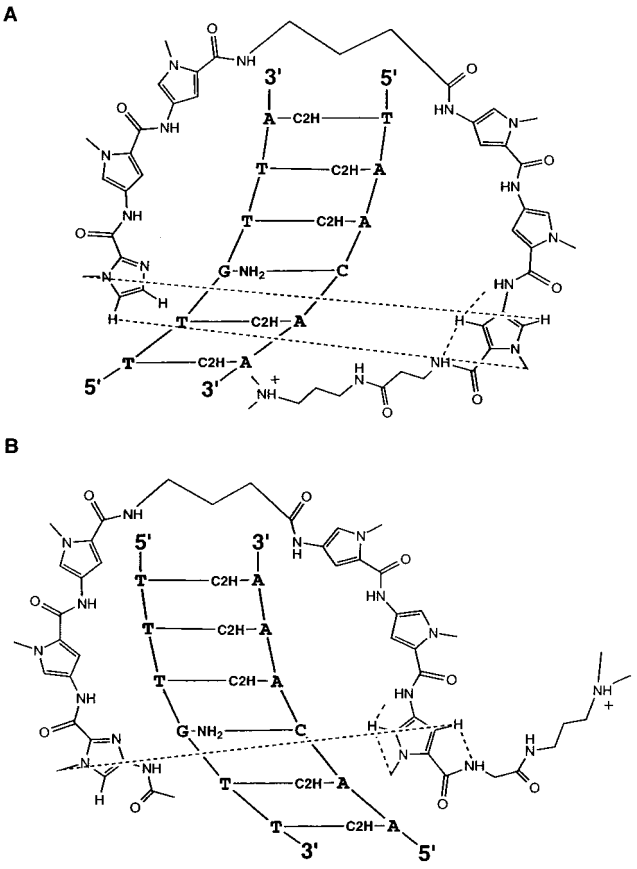

Figure 4. Schematics showing differences in intraligand NOEs between a typical 5′ directional complex and those of a 3′ directional complex with an inverted C-terminal pyrrole and an acetyl group. (A) ImPyPy-γ-PyPyPy-β-Dp complex with *TGTTA*. (B) Ac-ImPyPy-γ-PyPyPy-gly-Dp complex with *TTTGT*.

with *TGTTA*, **no** NOEs are seen between the β-Dp tail and the DNA, Figure 3b. To unambiguously assign the resonances of pyrrole ring 6 (attached to the tail) natural abundance $^{13}C-^{1}H$ correlation experiments were done. The chemical shifts for the ^{13}C resonances of the C3 and C5 of the pyrrole are distinct and hence allow clear identification of the corresponding attached proton resonances. The chemical shifts of the protons are reversed relative to typical complexes, the H5 occurring downfield of the H3. In addition the H5 and *N*-methyl show NOEs to protons at the bottom of the minor groove, while the H3 proton has an NOE to the *N*-methyl of the ring 1 imidazole, Figure 5. In the amide region, the H5 proton has an NOE to NH6, which faces into the groove, while the H3 proton has an NOE to NH7, which points out of the groove with the rest of the tail, Figure 6. Together these observations demonstrate clearly that the orientation of the last pyrrole ring of the ligand has flipped 180° relative to previously characterized complexes. In this orientation the ligand tail points out, away from the groove of the DNA, shown schematically in Figure 7. This binding orientation appears to be driven both by a clash between the tail and the acetyl group across the groove and also by a clash of the tail with the walls of the DNA groove. These possibilities are discussed in more detail in subsequent sections. The observation of the rotated conformer in this complex supports the argument that it also occurs in the *TGTTT* complex, explaining the exchange broadening observed and the atypical chemical shifts observed for ligand resonances in that complex.

ImPyPy-γ-PyPyPy-gly-Dp Complex. As noted above, footprinting studies indicated somewhat reduced specificity and affinity when gly rather than β-ala was used to link the last

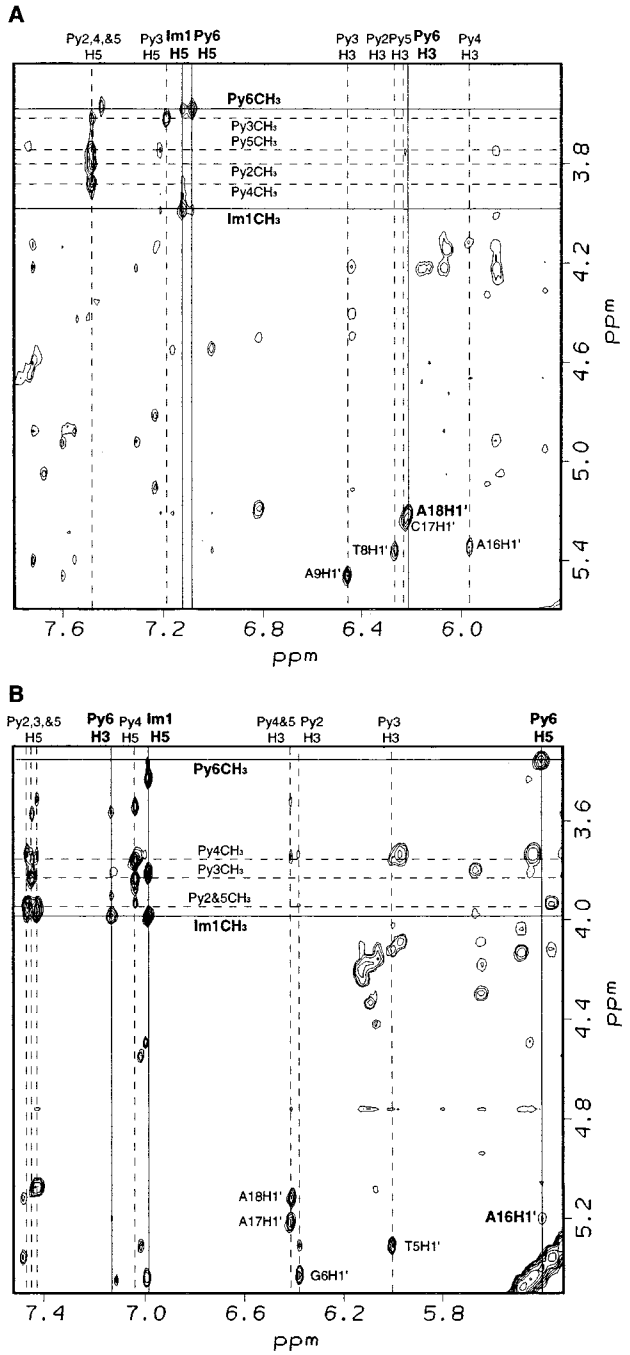

Figure 5. NOESY regions showing cross-peaks between ligand H3's and H5's to ligand *N*-methyls and DNA H1′s (D_2O, 500 MHz, 25 °C, τ_{mix} = 200 ms). (A) ImPyPy-γ-PyPyPy-β-Dp complex with *TGTTA*. Shown are the ligand H3-to-DNA H1′ contacts which indicate that the H3 protons of the ligand face toward the floor of the groove. Also, ligand H5 to the *N*-methyl of the same aromatic ring and the *N*-methyl of the cross groove aromatic ring, evidence that the ligand forms a hairpin conformation. (B) Ac-ImPyPy-γ-PyPyPy-gly-Dp complex with *TTTGT*. The significant difference from (A) is that the C-terminal pyrrole (Py6) H5 has a H1′ contact because it faces into the groove, while the Py6 H3 has a NOE to the Im1 *N*-methyl.

ring and the Dp part of the tail.[16] Structural studies were undertaken to try to understand the basis for this behavior. A well-defined complex was obtained in a titration of the ImPyPy-γ-PyPyPy-gly-Dp ligand with the *TGTTT* oligomer. NOESY data on this complex confirm that binding occurs with 5′-to-3′ directionality and that the ligand rings contact the GTT and AAC bases quite analogously to the equivalent ligand which has a

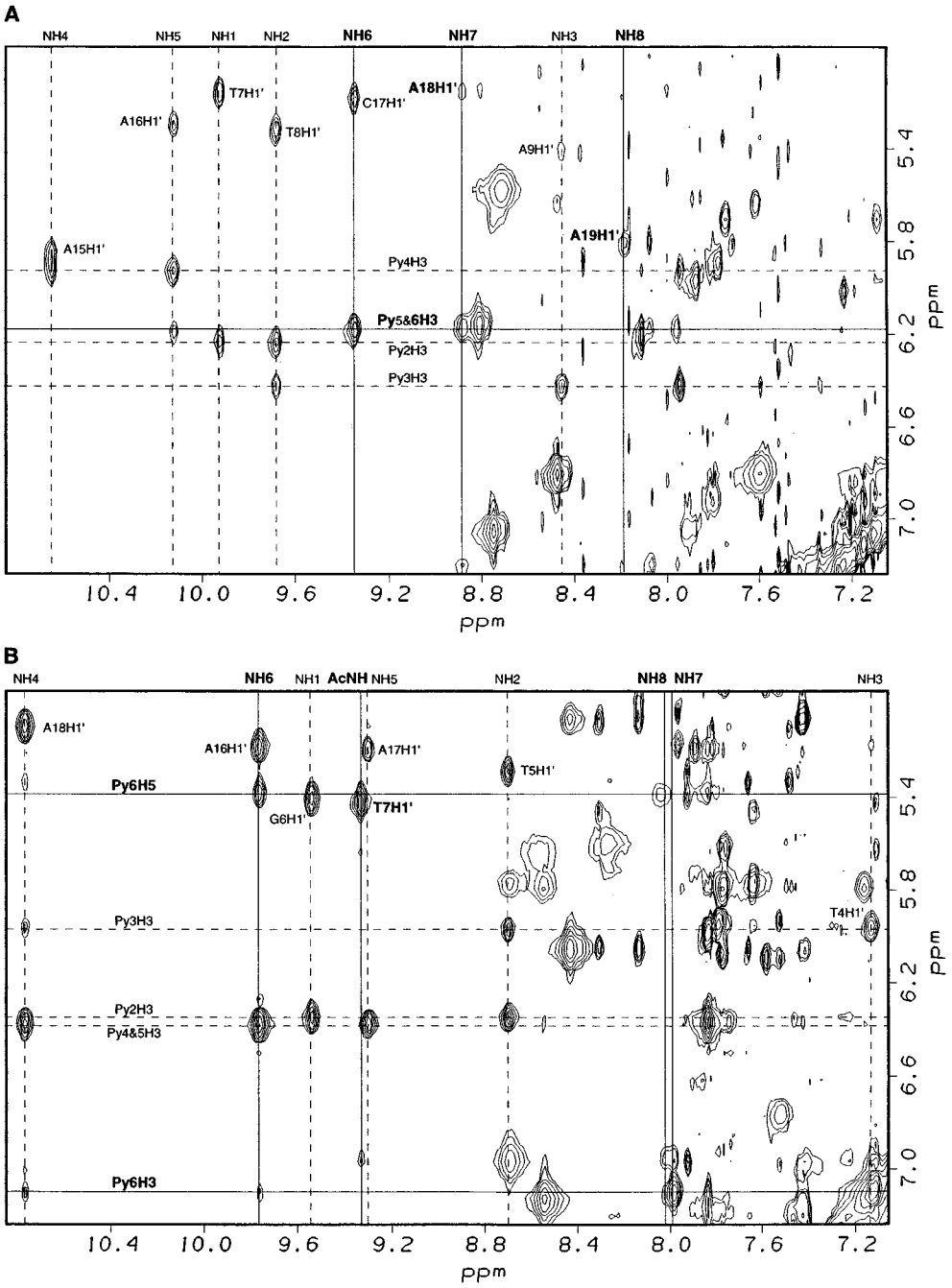

Figure 6. Expansion of the amide and aromatic to H1′ cross-peaks in NOESY spectra (90% H_2O/10% D_2O, 500 MHz, 25 °C, $\tau_{mix} = 200$ ms). (A) ImPyPy-γ-PyPyPy-β-Dp complex with *TGTTA*. Both amide protons of the tail make contacts to H1′s, indicating that the tail contacts the groove. Both NH6 and NH7 have NOEs to Py6 H3, which means that the C-terminal pyrrole is in the normal conformation. (B) Ac-ImPyPy-γ-PyPyPy-gly-Dp complex with *TTTGT*. Amide proton of the acetyl group (AcNH) has a NOE to a H1′, indicating that the amide proton is in the groove. The inverted C-terminal pyrrole (Py6) is evident in this region because NH6, which points into the groove apparent from a H1′ contact, has an NOE to Py6 H5, whereas NH7, which does not face into the groove, has an NOE to Py6 H3.

-Dp tail alone. The NOE data do indicate, however, that the last pyrrole ring is exchanging rapidly between normal and inverted states. NOEs are observed from both the H3 and H5 to the connecting amide proton. The relative intensities of these NOEs suggest nearly equal populations of the two conformers. The lack of linebroadening shows that the exchange between these conformers is rapid. There are no contacts observed between the tail and protons of the DNA, indicating that the tail is out of the groove even when the last pyrrole is bound in the normal geometry.

Ac-ImPyPy-γ-PyPyPy-Gly-Dp Complexes. Titrations of the *TGTTT* oligomer with the N-terminally acetylated, gly-linked tail ligand indicate that one predominant complex formed. NOEs again indicate clearly that the ligand is bound to the GTT•AAC site, running in the conventional 5′-to-3′ direction. The NOEs of the last pyrrole ring indicate that both the normal and inverted orientations of this ring are substantially populated, with the inverted somewhat preferred. The exchange between these conformers is rapid since only a single resonance, without substantial broadening, is observed for each proton. In the NOESY spectra there are exchange cross-peaks connecting the major form of complex to a minor one, behavior not seen for the β-containing ligand. The identity of the minor binding site is not apparent.

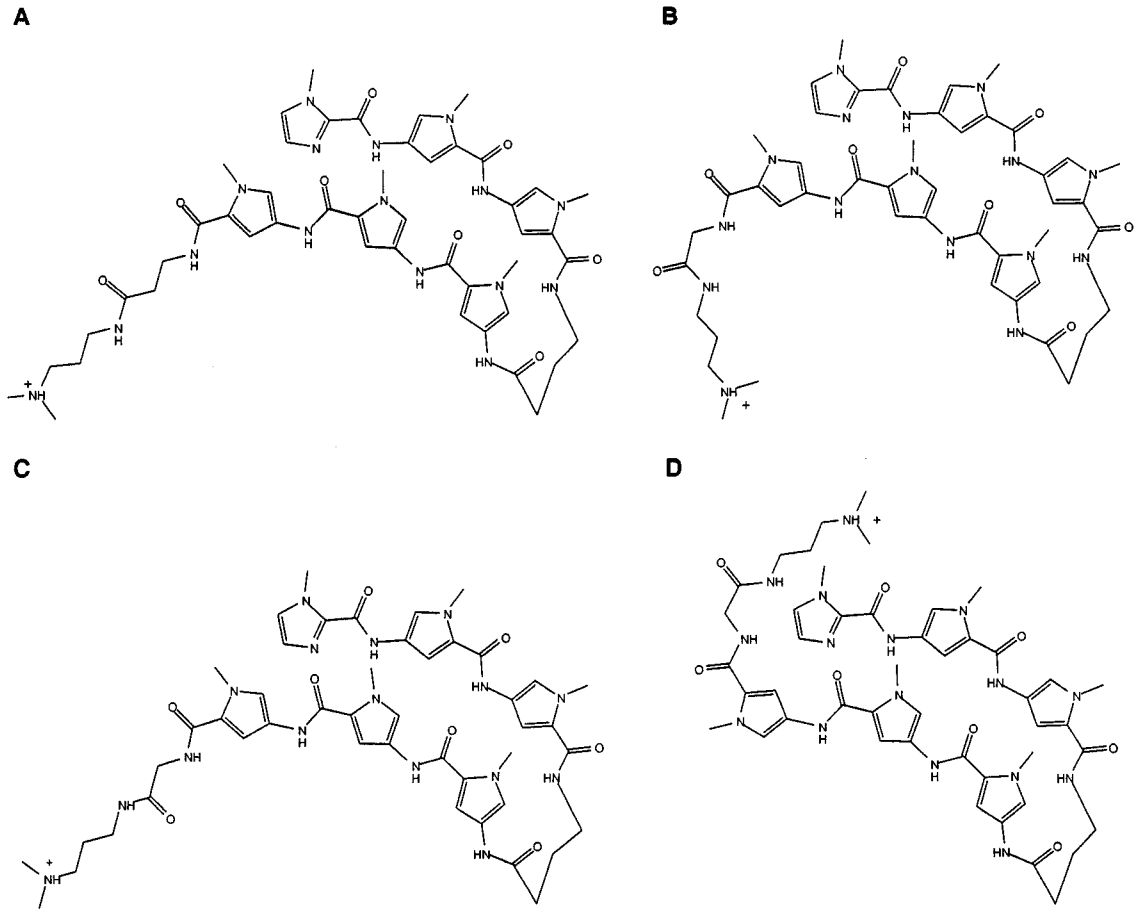

Figure 7. (A) ImPyPy-γ-PyPyPy-β-Dp shown with the C-terminal pyrrole, Py6, in the normal conformation. The β-ala allows the tail, with the amide protons facing down into the groove, to follow the curve of the DNA helix with the rest of the ligand. (B) ImPyPy-γ-PyPyPy-gly-Dp shown with the Py6 in the normal conformation with both amide carbonyl in the tail pointing up such that they would not contact the groove. It is obvious that in this conformation the tail of the ligand would sterically clash with the floor of the groove. (C) ImPyPy-γ-PyPyPy-gly-Dp with Py6 in the normal conformation with one amide carbonyl pointing down. With this geometry, the tail does not sterically clash with the groove, but the amide carbonyl pointing down is unfavorable. (D) ImPyPy-γ-PyPyPy-gly-Dp with Py6 in the inverted conformation. The tail cannot make contact with the groove.

To examine possible binding with 3′-to-5′ directionality, a titrations was done with the oligomer d(CCTTTGTTGG)• d(CCAACAAAGG), termed *TTTGT*. In this case a well-defined complex formed, and NOE contacts indicate clearly that binding occurs in the 3′-to-5′ orientation. In the complex pyrrole ring 6 shows NOEs corresponding to the inverted binding mode only, and no contacts between the tail and DNA are observed.

Matched Site vs Mismatched Binding. In footprinting studies of this family of ligands there were sites protected in which the actual binding site was unclear. We carried out NMR studies of an oligomer containing one such site, d(CCATGACATTCGTCG)• d(CGACGAATGTCATGG), termed *TGACATT*, in complex with ligands. This oligomer does not contain the expected optimum binding site of 5′-WGWWW-3′. With the nonacetylated, β-linked tail ligand a single well-defined complex formed. NOEs indicate that binding occurred in the normal 5′ to 3′ direction, at with the first three rings aligned with GTC, and the second three with GAC. This positions pyrrole rings 3 and 4 over a G-C base pair, which lowers the affinity relative to a fully matched site. There are normal contacts observed between the β-Dp tail and the bottom of the minor groove, specifically the adenosine H2 protons of A8, A21, and A22. Just considering the target sequence, it seemed possible that the ligand could also have bound the same mismatch sequence but with the imidazole contacting G5 rather than G24, Figure 2e. In this case the tail would be pointing toward the end of the oligomer rather than the center, and perhaps it is less favorable electrostatic interaction of the charged tail with the phosphates that leads to the observed binding site. Alternatively the fact that there are three A-T base pairs in this orientation, rather than two in the opposite, may affect binding affinity. With the gly-linked tail ligand the resonances broadened during the titration with this oligomer. Even after addition of a stoichiometric amount of ligand resonances remained broad, apparently due to chemical exchange between different forms of complex present.

When the complex was made with the acetylated version of either the β-ala- or gly-linked tail ligand, a well-defined complex was again formed. However, in this case binding occurred in the 3′-to-5′ orientation, the first three rings contacting G24, T23, A22 and the second set of three contacting T9, A8, C7, Figure 2f. There is no evidence for any contact between either the β-Dp or gly-Dp tail and the groove, and NOEs indicate that the pyrrole to which the tail is attached is always inverted. There is no evidence for binding at the mismatched site occupied by the nonacetylated, β-linked tail ligand. For the *TGACATT*/AcImPyPy-γ-PyPyPy-gly-Dp complex the chemical shifts are very similar to those in the inverted β-containing tail complex on the same oligomer, which is not surprising since the only difference in the ligands is in the tail which is not in contact with the DNA.

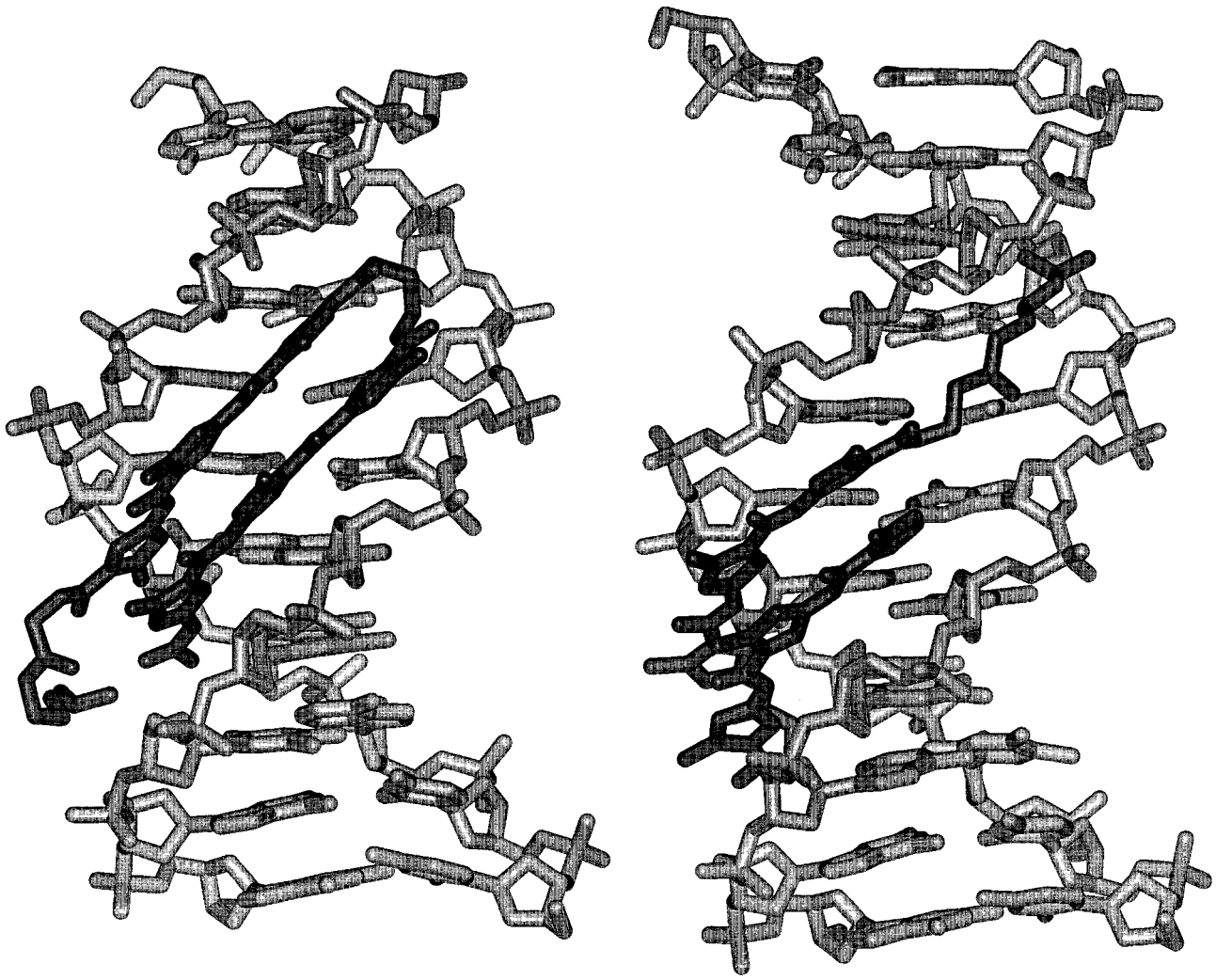

Figure 8. Molecular models obtained by simulated annealing and energy minimization using semiquantitative distance restraints derived from NOESY data. The hydrogens have been removed for clarity. (A) ImPyPy-γ-PyPyPy-β-Dp in complex with *TGTTA*. The tail contacts the groove with the rest of the ligand. (B) Complex of Ac-ImPyPy-γ-PyPyPy-gly-Dp and *TTTGT*. Note that the C-terminal pyrrole is in the inverted conformation and the tail is out of the groove.

Discussion

The fact that distamycin prefers to bind to "A-tract" DNAs in one orientation, in both the 1:1 and 2:1 modes, has been known for some time[1], but addressing the basis for this behavior has been difficult. By comparing the effects of relatively small modifications of covalent structure in a series of hairpin ligands, we have established that interactions of the alkyl tail with DNA are the dominant source of the orientational preference.

Geometry of the Linkers. When amino acid linkers were first added to the ligand tails to facilitate solid-phase synthesis, it was found that the β-alanine linker increased the affinity of the ligand relative to the dimethyldiaminopropyl tails used previously.[9,16] This was reasonable since the hydrophobic surface area of the tail available to contact the adenosine H2 protons in the minor groove was increased, while the geometry of the tail was essentially the same. As one would anticipate the longer tail contacts one additional A-T base pair. The reduction by one methylene unit in the gly-linked tail leads to somewhat different geometric constraints, Figure 7. For the tail to be maintained in an extended conformation along the groove the amide bond must be rotated by 180° relative to the β-ala-linked tail, pointing the amide carbonyl rather than the amide hydrogen into the groove. The carbonyl must be desolvated to fit into the groove, but there are no hydrogen bond donors available on the DNA to compensate at A-T sequences making binding less favorable than for the β-ala linker. If, on the other hand, the amide remained oriented with the N−H into the groove, then the remainder of the tail would clash with the groove unless there is also a rotation about the N−C bond of the glycine. In this alternate conformation for the N−C bond the tail ends up pointing away from the groove, leading to a loss of interaction between the Dp part of the tail and the DNA. Thus there appears to be no way for the gly-linked ligand to maintain optimum contact between the tail and the groove, which is supported by the experimental observation that contacts between a gly-linked tail and the groove have not been seen in any complex studied by NMR.

Steric Interactions of the Ligand N- and C-termini. In complexes of both distamycin and netropsin there are hydrogen bonds from each ligand amide group to an acceptor on the edge of a base pair in the minor groove, which are believed to contribute to the affinity of the ligands. Homologues of distamycin with additional pyrrole rings were synthesized and footprinted many years ago, verifying that they had increased affinity and required longer sequences of A-T base pairs for tight binding. These ligands were synthesized with either an N-terminal acetyl group, or an acetyl-EDTA conjugate for affinity cleavage studies. There were no apparent anomalies in

the binding behavior of these ligands. However, when one of the first G-C specific ligands was made, ImPyPy-Dp, it was somewhat surprising that addition of an N-terminal acetyl group to the imidazole ring lead to a large drop in affinity, whereas the nonacetylated form bound tightly and specifically to 5′-WGWCW-3′ sites as a 2:1 complex.[17] In the series of hairpin ligands discussed here the addition of the γ-linker between the sets of pyrrole/imidazole rings increases affinity, making it possible to study the effects of acetylation directly. The behavior of the acetylated -β-Dp and -gly-Dp tail ligands indicates the presence of a steric clash between the acetyl group and the tail. In each case the presence of the acetyl group leads to at least partial reversal of the orientation of pyrrole ring 6, to which the tail is attached. In the 5′ directional complex with the β-linked ligand with *TGTTT* this is manifested by intermediate exchange between conformers with Py_6 in the normal orientation and the 180° rotated form. In the rotated form favorable contacts between the tail and the groove are lost, indicating that the energetic cost of the clash is similar to the favorable interactions of the tail with the groove.

Tails Clash with the Groove When Binding with 3′-to-5′ Directionality. When the DNA sequence allows 3′-to-5′ binding the sixth pyrrole ring is always rotated to position the tail out of the groove, with no evidence for the equilibrium between the two rotational states as occurs with 5′-to-3′ binding sites. Examination of the ligands in complexes with DNA shows that there is a consistent twist of the ligand (leading to the induced CD signal seen for these complexes) which allows it to better match the curvature of the DNA groove.[6b] When the ligand is positioned such that the N-terminus is at the 5′ end of the contacted DNA strand then the twist naturally extends the tail along the groove. However in the reversed binding mode, with the ligand running 3′-to-5′ the twist positions the tail so that it runs into the wall of the minor groove. This enhances the tendency of the last pyrrole ring to rotate, which forces the tail completely out of the groove for the acetylated ligands in all cases studied.

Interactions of the Tail Dominate the Orientation of Binding. For the ligands described in the present work the interactions of the tail with the groove appear to be responsible for the 5′-to-3′ orientational preference which has normally been seen for this family of ligands. With the nonacetylated, β-linked ligand there is substantial interaction of the tail with the groove when bound in the 5′-to-3′ orientation, Figure 8a. That this interaction is substantial is shown clearly by the binding behavior with the *TGACAAT* oligomer, a mismatched site is occupied (typically leading to $\geq$10-fold reduction in affinity) rather than a matched 3′-to-5′ oriented site. However when the tail has a glycine or the N-terminus is acetylated, either of which compromises the contacts between the tail and the groove, Figure 8b, then 3′-to-5′ orientational binding occurs. When binding sites are available with both orientations in the same oligomer then there is an equilibrium between them, leading to exchange broadening which complicates analysis. However, this exchange behavior indicates that the binding in the two orientations must have very similar affinity. Thus, tails generate orientational preference by a combination of favorable interactions when binding 5′-to-3′ and unfavorable interactions when binding 3′-to-5′.

Implications for Ligand Design. Optimized hairpin ligands contain a γ linker and a β-ala tail. Both of these have a significant preference for binding at A-T base pairs, the γ contacting one A-T pair, and the β-ala contacting at least two, providing restrictions on the sequences which can be targeted by this class of ligand. In the present work we have shown that using an acetylated, gly-linked ligand causes rotation of the pyrrole ring with tail attached, leading to complete loss of interactions between the tail and DNA. This removes the requirement for two of the A-T pairs in the target sequence relative to a nonacetylated, β-linked ligand. In addition ligands can then be designed to target sequences binding in the 3′-to-5′ orientation. Another feature of the "tail out" binding mode is that ligands can be designed to bind with ring systems of two different ligands overlapped, impossible if the tail is in the groove. The effects of the steric interactions found in this work suggest that other covalent modifications are possible which could help control the binding mode.

JA0001198