J|A|C|S
ARTICLES
Published on Web 06/02/2004

NMR Structure of a Cyclic Polyamide−DNA Complex

Qing Zhang,[†,‡] Tammy J. Dwyer,[§,||] Vickie Tsui,[||] David A. Case,[||] Junhyeong Cho,[⊥] Peter B. Dervan,[⊥] and David E. Wemmer*[,†,‡]

Contribution from the Department of Chemistry, University of California, Berkeley, California 94720-1460, Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, Department of Chemistry, University of San Diego, 5998 Alcala Park, San Diego, California 92110, Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, and Division of Chemistry & Chemical Engineering, California Institute of Technology, Pasadena, California 91125

Received July 17, 2003; E-mail: dewemmer@lbl.gov

Abstract: The solution structure of a cyclic polyamide ligand complexed to a DNA oligomer, derived from NMR restrained molecular mechanics, is presented. The polyamide, *cyclo*-γ-ImPyPy−γ-PyPyPy*−, binds to target DNA with a nanomolar dissociation constant as characterized by quantitative footprinting previously reported. 2D ^{1}H NMR data were used to generate distance restraints defining the structure of this cyclic polyamide with the DNA duplex d(5′-GCCTGTTAGCG-3′):d(5′-CGCTAACAGGC-3′). Data interpretation used complete relaxation matrix analysis of the NOESY cross-peak intensities with the program MARDIGRAS. The NMR-based distance restraints (276 total) were applied in restrained molecular dynamics calculations using a solvent model, yielding structures with an rmsd for the ligand and binding site of ∼1 Å. The resulting structures indicate some distortion of the DNA in the binding site. The constraints from cyclization lead to altered stacking of the rings in the halves of the cyclic ligand relative to unlinked complexes. Despite this, the interactions with DNA are very similar to what has been found in unlinked complexes. Measurements of ligand amide and DNA imino proton exchange rates indicate very slow dissociation of the ligand and show that the DNA can undergo opening fluctuations while the ligand is bound although the presence of the ligand decreases their frequency relative to the free DNA.

Introduction

A design for ligands capable of targeting many different specific DNA sequences evolved from the pyrrole-polyamide natural products distamycin and netropsin, both of which bind to the minor groove with strong preference for A,T-rich DNA sequences. A variety of structural studies established that the binding preference of such crescent-shaped, aromatic ligands arises from the particularly narrow minor groove of DNA at A,T-rich segments.[1] It was subsequently shown that distamycin can also bind in a 2:1 complex, with two molecules stacked side-by-side in the groove, removing the requirement for a narrow groove.[2] Exploiting this 2:1 binding mode, it has been possible to make new ligands with specific hydrogen bond acceptors (imidazole [Im] replacing pyrrole [Py]) or donors (2-hydroxypyrrole [Hp] replacing pyrrole) that interact with strand-specific complementary groups on the DNA (the amino of guanosine for imidazole and the thymine carbonyl for hydroxypyrrole) to increase sequence specificity and broaden the range of sequences that can be targeted (for history and recent reviews, see refs 3 and 4). Quantitative footprinting, affinity cleavage, NMR, and X-ray diffraction have all been used to define the sequence preferences and structures of polyamide complexes with DNA.[5−8]

Individual ligands containing three pyrrole or imidazole rings bind with moderate affinity. For example, two molecules of the three-ring polyamide ImPyPy-Dp bind cooperatively to a five base-pair site, TGACT, with an apparent first-order binding constant $K = 2 \times 10^5$ M^{-1}.[9] This is significantly lower than the affinity of distamycin for most four base pair A,T sequences, although addition of a formamido group has been shown to raise the affinity to near that of distamycin.[10] In addition, recognition of many DNA sequences using the 2:1 binding motif requires

[†] University of California.
[‡] Lawrence Berkeley National Laboratory.
[§] University of San Diego.
[||] The Scripps Research Institute.
[⊥] California Institute of Technology.

(1) Kopka, M. L.; Goodsell, D. S.; Han, G. W.; Chiu, T. K.; Lown, J. W.; Dickerson, R. E. *Structure* **1997**, *5*, 1033−1046.
(2) Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5723−5727.
(3) Wemmer, D. E. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *29*, 439−461.
(4) Dervan, P. B. *Bioorg. Med. Chem.* **2001**, *9*, 1723−1733.
(5) Chen, X.; Ramakrishnan, B.; Rao, S. T.; Sundaralingam, M. *Nat. Struct. Biol.* **1994**, *1*, 169−175.
(6) Geierstanger, B. H.; Dwyer, T. J.; Bathini, Y.; Lown, J. W.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, *115*, 4474−4482.
(7) Geierstanger, B. H.; Jacobsen, J. P.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Biochemistry* **1994**, *33*, 3055−3062.
(8) Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 7586−7590.
(9) Wade, W. S.; Mrksich, M.; Dervan, P. B. *Biochemistry* **1993**, *32*, 11385−11389.
(10) Lacy, E. R.; Le, N. M.; Price, C. A.; Lee, M.; Wilson, W. D. *J. Am. Chem. Soc.* **2001**, *124*, 5153−2163.

10.1021/ja0373622 CCC: $27.50

two different ligands, one to sequence specifically contact each strand while binding cooperatively at the desired target sequence. However, in such a situation, it is possible for the ligands to bind in other combinations at different sequences, which must be considered in evaluating the overall sequence specificity of binding. To improve both the affinity and the specificity, covalently linked versions of polyamide ligands have been made. Two different types of bridges between the pyrrole/imidazole ring systems have been prepared. In the first, the ligands are connected by extending the *N*-methyl groups on the Im/Py rings to *N*-alkyl-*N*.[11−13] These linkers point away from the DNA, not interacting with it at all. Different linker lengths have been tested, three carbons being the shortest that can bridge and give the properly stacked arrangement. Longer alkyl groups allow the pyrrole/imidazole ring systems to take on different staggered positions. In both cases, the affinity of binding is increased by linkage. The second type of bridge links the polyamide units head to tail, giving structures which have generally been referred to as hairpins.[14−16] Again, different length alkyl groups were tested: a linker with two methylenes (β-amino proprionic acid) allowed folding back to make the hairpin but only with significant strain, reducing the affinity; a three-carbon linker (γ-amino butyric acid) seemed optimal; and four (δ-amino pentanoic acid) leads to steric clashes with the walls of the groove, again reducing affinity.[14,17] The attachment points for the linker lie deep in the groove, and there are contacts between the linker and the groove that lead to a favorable contribution to the binding affinity, as well as specificity for A,T base pairs at the site of the linker.[18] Quantitative footprinting[11] showed that the hairpin-linked ligands bind with a higher affinity than unlinked dimers, and also enhanced specificity as the covalently linked dimers did not footprint at A,T sites which were bound by the unlinked version. The hairpin ImPyPy-γ-PyPyPy-Dp was shown to have an equilibrium association constant of 8×10^7 M^{-1} for the "matched" DNA target site 5′-TGTTA-3′/5′-TAACA-3′, an increase of about 400-fold over unlinked dimers.[19]

It is clear that part of the enhancement in binding affinity upon linking the ligands comes through the reduction in entropy loss upon binding that occurs with the individual ligands. This idea has been carried further by closing the hairpins to form a macrocycle.[20] The first such "cyclic polyamide" ligand used a second -γ- linker, and a derivatized pyrrole to maintain solubility and retain a positive charge favorable for DNA interaction. This cyclic polyamide ligand, Figure 1, is a direct extension from the first hairpin, binding the same DNA targets. Quantitative footprinting indicated the binding constant for the site 5′-TGTTA-3′/5′-TAACA-3′ to be $K \geq 2.9 \times 10^9$ M^{-1} (pH 7.0, 22 °C), which is an enhancement of about 30-fold relative to the hairpin version. However, footprinting on other related DNA sequences indicated that the decrease in affinity at single base mismatched sites was less than that for the hairpin, indicating a decrease in overall specificity. Subsequent work with other cyclic ligands containing a total of eight pyrrole/imidazole rings, and an amino-derivatized linker at one end to improve solubility instead of the modified pyrrole, showed that both high specificity and affinity can be obtained with this design.[21]

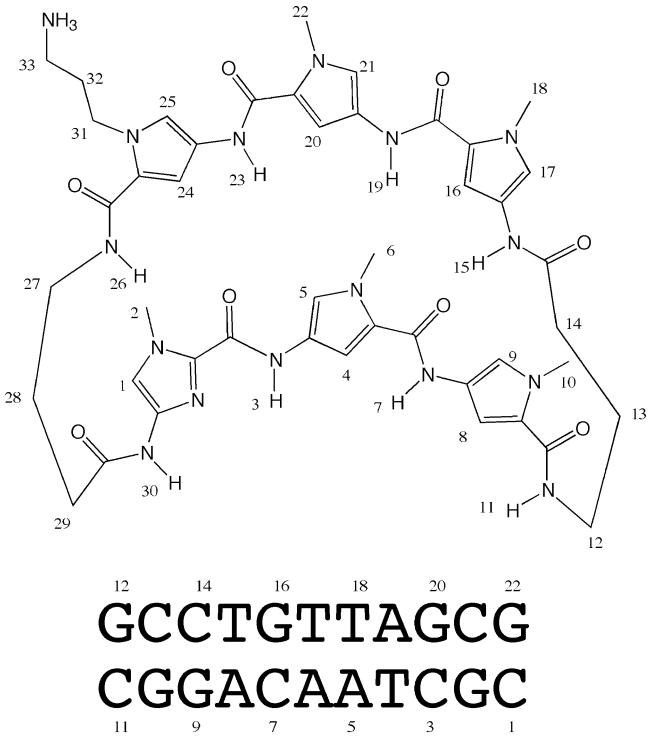


Figure 1. (A) The structure of the cyclic polyamide used in the studies described is shown with the DNA sequence and numbering schemes used for each. (B) The structures of distamycin and the hairpin equivalent of the cyclic polyamide are shown for reference.

Here, we describe structural results from 2D NMR investigations of the six ring cyclic polyamide *cyclo*-γ-ImPyPy−γ-PyPyPy*− bound to the duplex DNA oligomer d(GCCTGTTAGCG)−d(CGCTAACAGGC). In addition, measurements of hydrogen exchange rates of both the DNA imino protons and the amides of the ligand were performed to investigate the dissociation rate of the ligands and to probe conformational fluctuations in the complex.

Materials and Methods

Synthesis of the Cyclic Polyamide and Oligonucleotides. The cyclic polyamide was synthesized as reported previously.[20] The oligomers d(GCCTGTTAGCG) and d(CGCTAACAGGC) were synthesized using the phosphite triester method on an Applied Biosystems 381A synthesizer and were purified by HPLC as described previously.[2] Oligomer concentrations were determined from UV absorbance at 260 nm. The oligomers were mixed in a 1:1 stoichiometry, annealed at 85 °C, and then dialyzed against 2 L of 0.2 M NaCl, 2 L of 0.02 M NaCl, and 2 L of deionized water subsequently in a flow dialysis apparatus

(11) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 9892−9899.
(12) Dwyer, T. J.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, *115*, 9900−9906.
(13) Chen, Y.-H.; Yanwu, Y.; Lown, J. W. *J. Biomol. Struct. Dyn.* **1996**, *14*, 341−355.
(14) Mrksich, M.; Parks, M. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 7983−7988.
(15) Parks, M. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6147−6152.
(16) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 8198−8206.
(17) de Clairac, R. P. L.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1997**, *119*, 7909−7916.
(18) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1999**, *121*, 1113−1120.
(19) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 3663−3664.
(20) Cho, J.; Parks, M. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 10389−10392.
(21) Herman, D. M.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1999**, *121*, 1121−1129.

using a membrane with a 1 kD molecular weight cutoff. The purity of the duplex DNA sample was greater than 95% based on NMR. This duplex DNA is henceforth referred to as [TGTTA].

NMR samples of the complex were prepared by dissolving the [TGTTA] oligomer duplex in 0.5 mL of 10 mM sodium phosphate buffer with 0.1 mM EDTA (pH 7.0) and then lyophilizing and redissolving in a 90% H_2O/10% D_2O mixture or 99.96% D_2O (Cambridge Isotope Laboratories) to a final volume of 0.5 mL. A 22 mM stock solution (based on UV absorbance) of cyclic polyamide was prepared in DMSO, due to its low solubility in H_2O. The sample concentration used for NOE and HSQC data was ~4 mM in the complex.

Formation of the [TGTTA]:cyclic polyamide complex was followed by NMR. 1D spectra of the sample were acquired on a General Electric GN-500 spectrometer after each addition of ca. 0.25 mol equivalent of the cyclic polyamide in DMSO solution. Spectra were acquired at 25 °C using 8192 complex points, 64 scans, and a spectral width of 12 048 Hz, with a 1:1 jump and return sequence[22] applied to suppress the residual HDO resonance.

T_1 values were measured for nonexchangeable protons in the DNA/cyclic polyamide complex in D_2O and were used to correct the integrals for cross-peaks in the 2D spectra, using the following relation:

$$F = \frac{1}{(1 - \exp(-t_{rd}/T_1))}$$

where F is the integral correction factor, t_{rd} is the total recycle delay in the experiment, and T_1 is the measured longitudinal relaxation time.

NOESY, COSY, and TOCSY spectra in D_2O were acquired using TPPI[23] on a General Electric GN-Omega 500 MHz spectrometer. Data were collected at 15, 25, and 35 °C to resolve some cross-peak overlaps in the spectra. NOESY spectra were collected with 1024 complex points in t_2 with a spectral width of 4504 Hz and a mixing time of 200 ms. 512 t_1 increments were recorded, and then the data were zero-filled to 1024. For each t_1 value, 64 scans were averaged using a recycle delay of 2 s. Presaturation was applied during the recycle delay and the mixing time to suppress the residual HDO resonance. COSY spectra were collected with 2048 complex points in t_2 with a spectral width of 4504 Hz. 512 t_1 experiments were recorded and then zero-filled to 1024 points. For each t_1 value, 48−80 scans were collected using a recycle delay of 2 s with presaturation of the HDO resonance. TOCSY spectra were collected with 1024 complex points in t_2 with a spectral width of 4504 Hz and mixing times of 50−75 ms. 512 t_1 experiments were recorded and then zero-filled to 1024 points in Felix 2.3.0. For each t_1 value, 64 scans were collected using a recycle delay of 2 s with presaturation of the HDO resonance. Natural abundance $^1H-^{13}C$ HSQC data to aid assignment were acquired at 25 °C using States-TPPI on a Bruker AMX 600 MHz spectrometer. Spectra were collected with 2048 complex points in t_2 with a spectral width of 6667 Hz. The ^{13}C carrier was set to 85 ppm. 256 complex points were recorded in t_1 (^{13}C dimension) with a spectral width of 7496 Hz. For each t_1 value, 64 scans were collected using a recycle delay of 2 s with presaturation of the HDO resonance. NOESY spectra in H_2O were acquired on a Bruker AMX 600 MHz spectrometer. Water suppression was achieved with a 1:1 jump and return sequence.[22] A spectrum at 25 °C was acquired for the complex, and one at 15 °C was acquired for the free DNA. The pulse sequence "delay-90°$_x$-t_1-90°$_{-x}$-τ_{mix}-90°$_x$-Δ^{11}-90°$_\phi$-t_2" was used with quadrature detection using TPPI. 2048 complex points were acquired in t_2 with a spectral width of 13 514 Hz and a mixing time of 200 ms. 512 t_1 experiments with 64 scans were recorded and zero-filled to 1024 in Felix 2.3. The delay period Δ^{11} (45 μs) was calibrated to optimize the excitation by creating a maximum in the imino proton region (11−14 ppm) and a null at the water resonance. All NOESY spectra in D_2O for calculating the distance restraints were acquired at 25 °C using TPPI on a Bruker AMX 600 MHz spectrometer. The spectra were collected with 2048 complex points in t_2, a spectral width of 6024 Hz, for mixing times of 50, 100, 150, and 200 ms. 512 t_1 experiments were recorded and then zero-filled to 1024 in Felix 2.3. For each t_1 value, 48 scans were signal averaged using a recycle delay of 2.1 s with presaturation of the HDO resonance. NOE cross-peaks were then integrated manually and corrected with calculated integral correction factors.

All data were processed with Felix 2.3 (Biosym.) on a SGI workstation. The 2D data were apodized with a skewed sine bell function in both dimensions (800 or 1600 points, phase 60°, skew 0.5−0.7 in t_2; 128 or 512 points, phase 60°, skew 0.5−0.7 in t_1). The first row of the data matrix was multiplied by a factor in the range of 0.3−0.7 prior to Fourier transformation in t_1 to suppress t_1 ridges.

Exchange rates for the ligand amide protons were determined from the intensities of peaks in 1D spectra taken as a function of time after dissolving a deuterated sample of complex in 90% H_2O/10% D_2O. Intensities were fit to a simple exponential curve to obtain the exchange rate. Measurements were done at pH 7.0 and 9.1, 25 °C. Imino proton exchange rates in both the free DNA duplex and the complex were determined at pH 8.8 as a function of ammonia concentration by measuring the line width contribution due to exchange (the dominant contribution at high pH and added ammonia).[24] The rates were extrapolated to infinite ammonia concentration to determine the base pair opening rates.

The program MARDIGRAS[25−27] was used to derive distance estimates from NOESY cross-peak intensities. To evaluate uncertainties in the distance estimates, the RANDMARDI (Random Error MARDIGRAS)[28] procedure was used. This approach simulates the effects of experimental noise by redoing the distance estimates with different random additions to each input NOE intensity, scaled appropriately for the experimental noise in the data.

NOESY spectra were collected with mixing times of 50, 100, 150, and 200 ms and were analyzed to generate distance restraints. The upper and lower diagonal NOE cross-peaks of each of the four assigned NOESY spectra were integrated using

(22) Sklenar, V.; Brooks, B. R.; Zon, G.; Bax, A. *FEBS Lett.* **1987**, *216*, 249−252.
(23) Drobny, G.; Pines, A.; Sinton, S.; Weitekamp, D. P.; Wemmer, D. E. *Faraday Div. Chem. Soc. Symp.* **1979**, *13*, 49−55.
(24) Gueron, M.; Leroy, J. L. *Methods Enzymol.* **1995**, *261*, 383−413.
(25) Borgias, B. A.; Gochin, M.; Kerwood, D. J.; James, T. L. *Prog. Nucl. Magn. Reson. Spectrosc.* **1990**, *22*, 83−100.
(26) Borgias, B. A.; James, T. L. *J. Magn. Reson.* **1988**, *79*, 493−512.
(27) Borgias, B. A.; James, T. L. *J. Magn. Reson.* **1990**, *87*, 475−487.
(28) Liu, H.; Spielmann, H. P.; Ulyanov, N. B.; Wemmer, D. E.; James, T. L. *J. Biomol. NMR* **1995**, *6*, 390−402.

Felix 2.3, creating a total of four peak intensity sets. Peak overlap of some NOESY cross-peaks reduced the accuracy of the integration, and hence these cross-peaks were not included in the RANDMARDI calculations. For the calculations, a noise level of 1−5 times the integrated intensity of the smallest NOE cross-peak was used as the "noise" range. Thirty different random additions were done for each experimental data set, and MARDIGRAS was applied to each. Resulting distances were then averaged to form one bounds file from each of the four data sets. The distance estimates from the four mixing times were then combined to give the final distance restraint file. Upper and lower bounds in the distance restraint file were average interproton distances ± 1 standard deviation calculated from all of the MARDIGRAS runs. The width of the restraints (difference between lower and upper bounds) varies from 0.12 to 2.8 Å. All of the intermolecular restraints used are given in BioMagResBank deposition 5927. The accuracy of the RANDMARDI procedure was tested by comparison of the calculated distances and distances derived from covalent bonds in the cyclic polyamide. Distances within the cyclic polyamide were calculated, and good agreement was found between the experimentally determined distances and those calculated from known covalent bond lengths and angles, indicating that the RANDMARDI procedure functioned properly. The final set of 276 restraints included 29 Watson−Crick H-bond restraints, 187 intraDNA, 25 intraligand, and 35 intermolecular distance restraints.

The approach for computing structures used in this study was patterned after that of Smith, et al.[29] All energy minimization and restrained molecular dynamics (rMD) calculations were performed with the SANDER module of AMBER 6[30] on a Silicon Graphics Origin cluster. The NAB molecular manipulation language[31] was used to create a series of 40 DNA structures. These structures differed in the four helical parameters inclination, rise, twist, and *x*-displacement to give a diverse set of A-form and B-form DNA oligomers. Each member of this ensemble was subjected to 1000 steps of steepest descent energy minimization followed by heating to 700 K for 20 ps of rMD SA. A total of 216 DNA−DNA constraints were used including Watson−Crick hydrogen bonding constraints. The force constants for all DNA−DNA constraints were increased linearly from 3.2 to 32 kcal over 3 ps, and remained at 32 kcal for the final 17 ps of the simulation. The resulting DNA structures had a mean pairwise rmsd of 4.79 Å.

A second ensemble of 40 starting conformations of the ligand was generated by taking snapshots from a 20 ps MD simulation of 1 at 300 K. Each of the 40 ligand structures was placed at a distance of 10 Å from a randomly chosen member of the DNA ensemble using NAB. In each pair, the ligand was positioned in the correct binding orientation with respect to the minor groove and docked into the minor groove of its DNA partner via the following rMD procedure. All constraints used in the DNA equilibration plus 34 ligand−DNA constraints and 25 ligand−ligand constraints were utilized. Force constants for all intramolecular constraints were kept constant at 32 kcal during the simulation, while those for intermolecular constraints were ramped linearly from 0 to 32 kcal over 20 ps. Each complex was heated to 700 K over 3 ps and held at that temperature for 3 ps. The system was then cooled to 0 K over 14 ps.

The refinement was completed with two cycles of rMD simulated annealing using all 276 constraints. In each cycle, the complex was heated to 1000 K over 2 ps and held at this temperature for 3 ps of dynamics. The system was then cooled to 0 K over 15 ps. Force constants for all constraints were ramped linearly from 3.2 to 32 kcal over the first 3 ps of the simulation and held at 32 kcal for the remaining 17 ps.

To establish the "correct" number of structures to represent the ensemble, the statistical analysis suggested by Smith, et al.[29] was used. The 40 structures were placed in random order, and the mean all-atom pairwise rmsd was taken for the first two structures in the group, then the first three structures in the group, etc. This procedure was repeated 500 times, each time with a different random ordering of the 40 structures. The analysis predicts the minimum number of structures sufficient to represent the conformational space consistent with the experimental data, in the limit of the change in the standard deviation of rmsd with the addition of a structure to the family. It was determined that an ensemble of 20 structures was adequate to describe the complex.

The 20 structures in the final ensemble were chosen to minimize both the molecular mechanics energy and the constraint violation energy, coordinates were deposited in the RCSB database as entry 1PQQ. A summary of the energy and rmsd characteristics of the 20 structures is given in Table 1.

Table 1. Summary of Molecular Energies, rmsd's, and Violations of the 20 Analyzed Structures

Molecular Mechanics Energy (kcal)	
EAmber	−6224.6 ± 6.8
Eviol	17.4 ± 1.8
Average Pairwise rmsd (Å)	
DNA	1.99
ligand + DNA	1.92
ligand + DNA binding site	0.98

The chemical shifts were calculated using the SHIFTS program that can be downloaded at http://www.scripps.edu/case. Following the procedure of Case,[32] a set of "dummy" points were placed at various positions in the vicinity of the ring at a fixed distance of 2.0 Å from the van der Waals surface. The chemical shifts of the points were computed, and ring-current parameters were calculated. Values for ring susceptibility anisotropies for imidazole (−99.0 (ppm Å^3)/molecule) and pyrrole (−96.4 (ppm Å^3)/molecule) rings were determined. With the ring-current information and the structure of the ligand:DNA complex (in Brookhaven PDB format) as input, SHIFTS uses simple empirical models to compute the proton chemical shifts for the specified pyrrole protons of the ligand.[33,34]

(29) Smith, J. A.; Gomez-Paloma, L.; Case, D. A.; Chazin, W. J. *Magn. Reson. Chem.* **1996**, *34*, S147−S155.
(30) Case, D. A.; Pearlman, D. A.; Caldwell, J. C.; Cheatham, T. E., III; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. *AMBER* **1999**, *6*, University of California, San Francisco.
(31) Macke, T. Ph.D. Dissertation, Scripps Research Institute, La Jolla, CA, 1996.
(32) Case, D. A. *J. Biomol. NMR* **1995**, *6*, 341−346.
(33) Osapay, K.; Case, D. A. *J. Am. Chem. Soc.* **1991**, *113*, 9436−9444.
(34) Dejaegere, A. P.; Bryce, R. A.; Case, D. A. In *Modeling NMR Chemical Shifts*; Facelli, J. C., de Dios, A. C., Eds.; American Chemical Society: Washington, DC, 1999; pp 194−206.

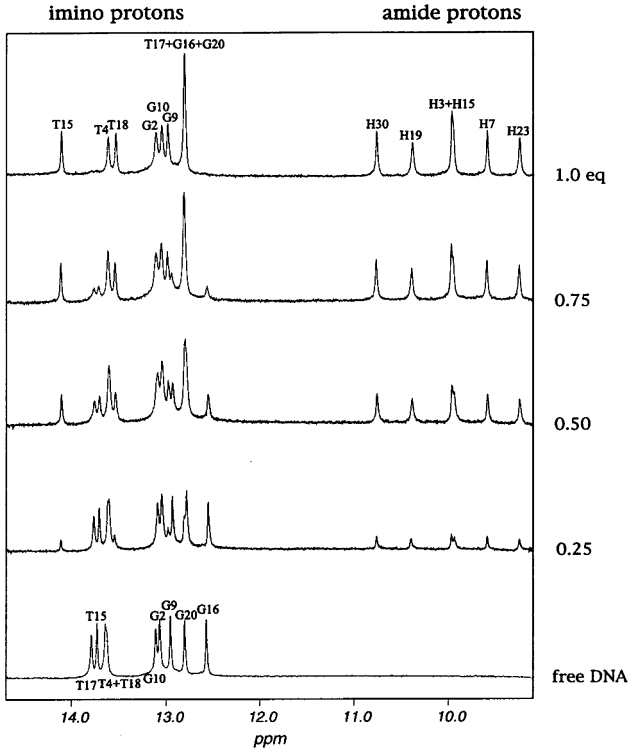


Figure 2. The downfield portion of the proton NMR spectra of DNA at different points during a titration with the cyclic polyamide. The ratio of ligand to DNA is shown at the right of each spectrum. The identities of the imino and amide resonances are shown above the peaks.

Results

Figure 2 shows sections of NMR spectra taken at points in the titration of DNA with ligand. Upon addition of the ligand resonances from the ligand appear, free DNA resonances decrease intensity and new resonances from DNA in the complex grow in, indicating slow exchange between free DNA and complex, as expected for this high affinity ligand. Substantial shifting of the imino protons of A−T base pairs occurs as expected because they constitute most of the ligand binding site. Resonances in the 9−11 ppm range were also observed from the ligand amide protons, as has been described in other complexes.

The numbering schemes for DNA and the cyclic polyamide are shown in Figure 1. Resonance assignment followed conventional NOESY-based methods.[35] The assignments for DNA aromatic base protons and sugar H1's are shown on a section of the NOESY spectrum in Figure 3. Assignment of sugar 3′,4′ and 5′,5″ protons was facilitated by using natural abundance $^{1}H-^{13}C$ HSQC correlations, in which the sugar positions can be unambiguously identified by the carbon chemical shift, Figure 4. Imino and amino proton resonances were assigned from NOESY data in H_2O solution.[36] The cyclic polyamide ligand protons were assigned through a combination of intra- and intermolecular NOEs in both D_2O and H_2O NOESY data and were confirmed by the $^{1}H-^{13}C$ HSQC data (Figure 4). The chemical shifts of the DNA and ligand are given in BioMagResBank deposition 5927. The NOESY cross-peaks between ligand and DNA are consistent with the expected binding site for this ligand, the pyrrole and imidazole rings contacting G T T on strand 1, and A A C on strand 2.

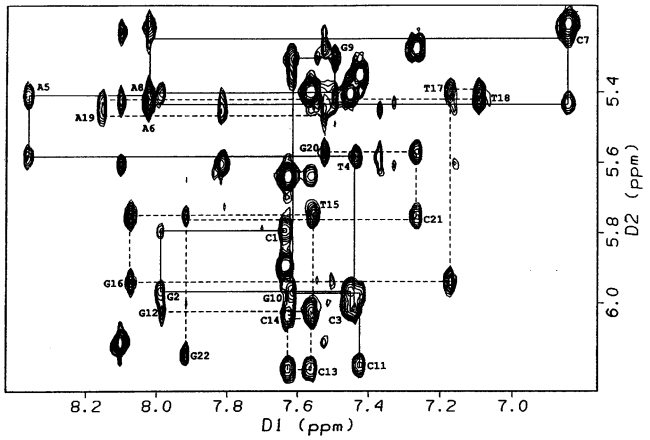


Figure 3. A portion of a NOESY spectrum of the 1:1 polyamide:DNA complex taken in D_2O solution is shown. Sequential connectivities are indicated for the two DNA strands by solid and dashed lines. Intraresidue peaks are labeled with the residue.

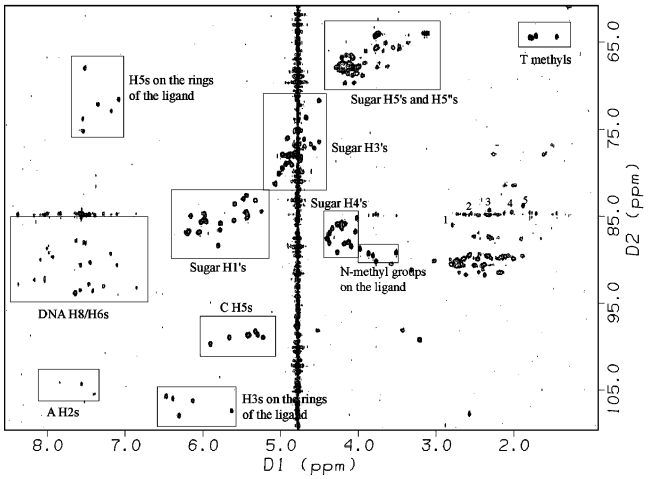


Figure 4. The natural abundance, proton-detected ^{13}C HSQC of the 1:1 cyclic polyamide ligand complex is shown. The spectrum was collected with folding in the ω_1 dimension; reported chemicals shifts are corrected for folding; folded peaks are shown in gray rather than black.

The ensemble of 20 structures used to represent the polyamide:DNA complex is shown in Figure 5A, with the average structure shown separately in Figure 5b. The average all-atom pairwise rmsd for this family is 1.92 Å for ligand and full DNA, but 0.98 Å for the ligand and DNA in the immediate binding site (T4 to A8 and T15 to A19). The ligand fits snugly into the minor groove, with the individual imidazole/pyrrole amide units twisted to follow the shape of the groove in both curvature and pitch. Both -γ- linkers are positioned near the bottom of the groove making hydrophobic contacts, as do the single -γ- linkers in hairpin complexes. The ligand rings are stacked about 3.2 Å apart. The propylammonium of pyrrole 6 in the ligand points away from the groove, but can still make favorable electrostatic contacts with the phosphates of the DNA backbone. The data do not define its position except for the connection to the pyrrole ring.

On the inside of the ligand, hydrogen bonds are formed between the amides of the ligand and acceptors on the edges of

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

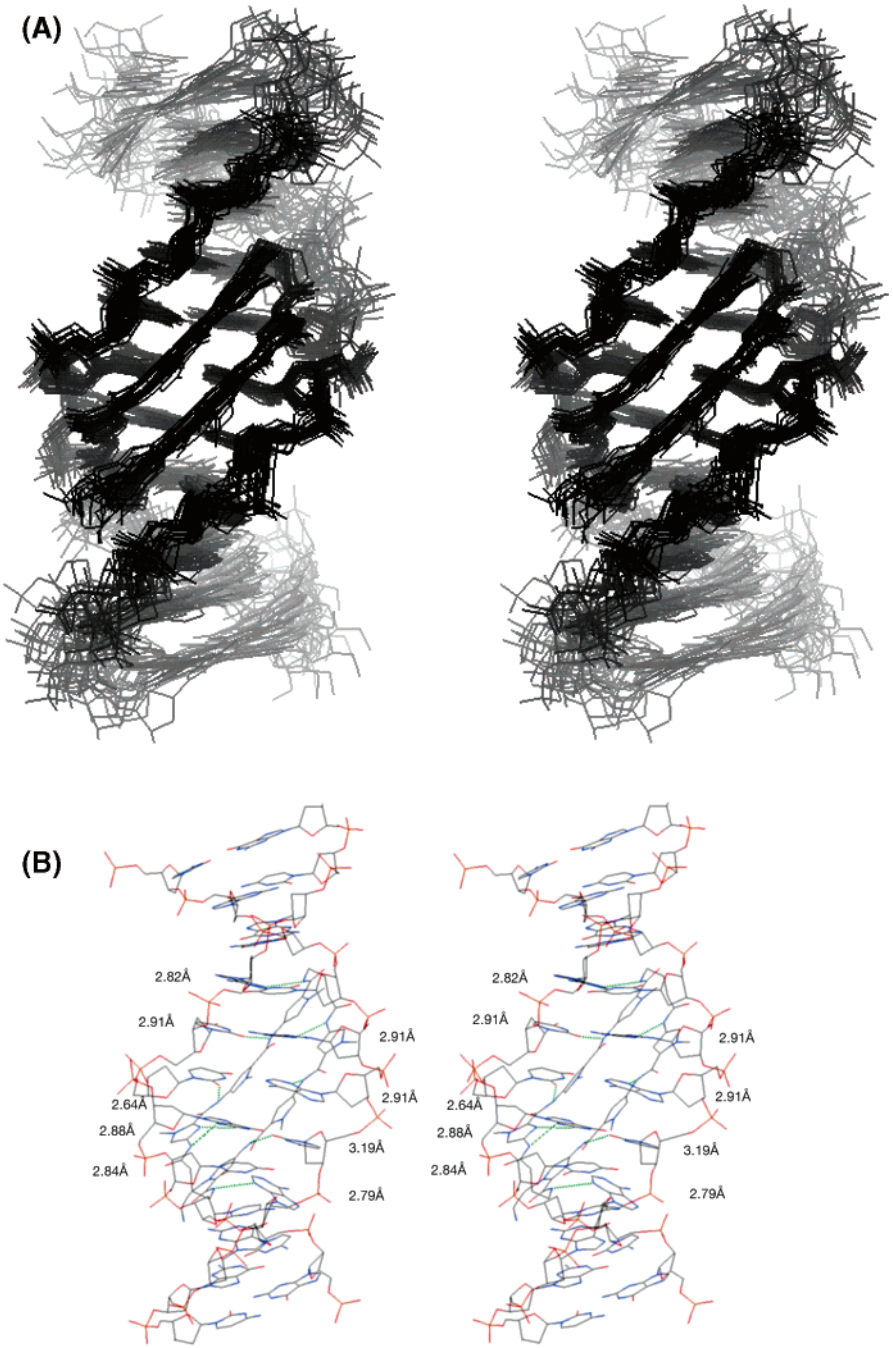


Figure 5. (A) The 20 computed structures used to represent the cyclic ligand complex structure are shown. (B) The structure of the minimized averaged structure from the family in (A) is shown. Hydrogen bonds between the ligand and DNA are labeled with the donor acceptor heavy atom distance.

the DNA bases, as well as between the G5 amino group and the ligand imidazole, Figure 5b. All of the ligand amides are hydrogen bonded to the DNA, with hydrogen bond lengths (2.6−3.2 Å, average 2.9 Å) that are quite similar to those found in the crystal structure of other polyamide ligand complexes (ranging from 2.8 to 3.4 Å).

Helical parameters for DNA structures in the complex were derived using Curves 4.[37] Figure 6 shows plots of the values of rise, twist, *x*-displacement, inclination, and sugar pucker. A majority of the parameters are near the value of standard B-form DNA, although the twist is somewhat reduced, falling between the typical B- and A-form DNA values. The average *x*-displacement is −1.0 Å, placing base pairs near the helical axis, as in B-DNA for which it is −0.8 Å. Overall, the sugar puckers are near C2′ as endo in typical B-form DNA, although C7 and A19 are close to the C4′-exo or O4′-endo range. The width of the minor groove is widened somewhat to accommodate the stacked aromatic rings at the center of the ligand. The pattern of structural variations suggests that contact with specific parts of the ligand leads to changes in the local helical parameters.

When samples of ligand were added to D_2O solutions of DNA, resonances from amide protons of the ligand were observed which persisted for many hours. Such behavior had

(37) Lavery, R.; Sklenar, H. *J. Biomol. Struct. Dyn.* **1988**, *6*, 63−91.

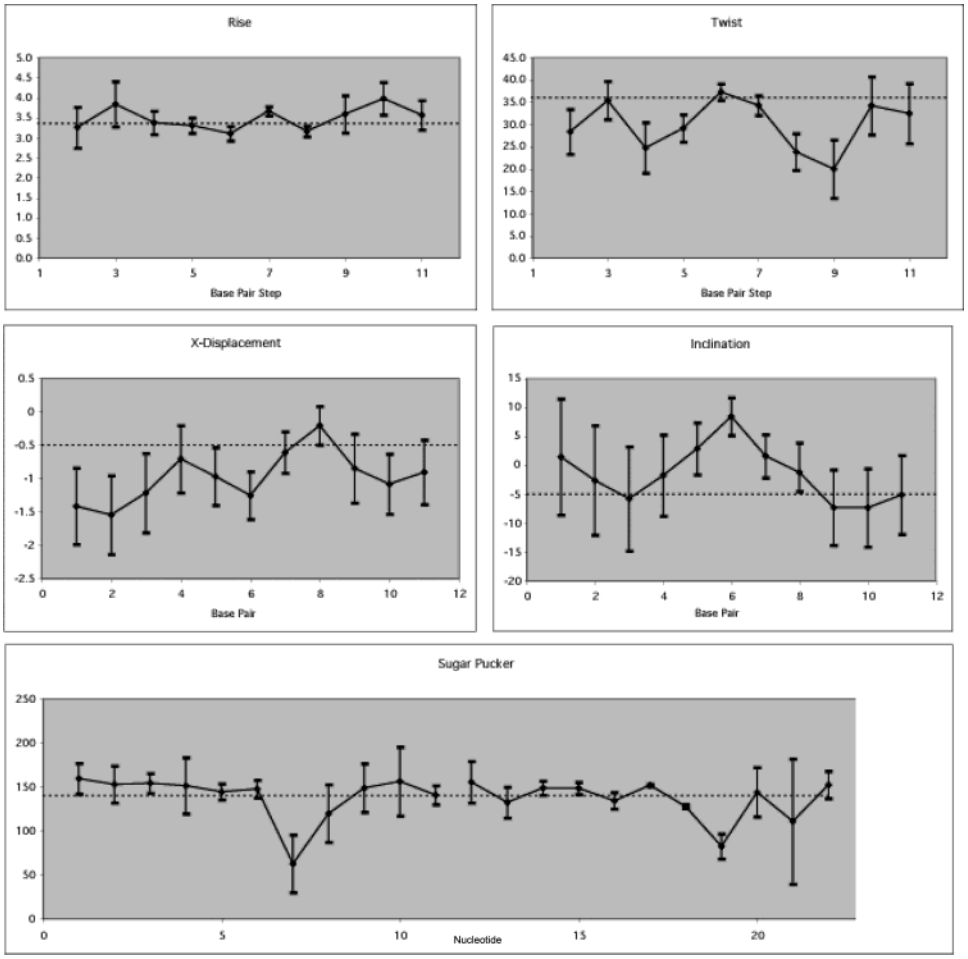


Figure 6. Helical parameters for the cyclic ligand complex are presented. The dashed lines represent the expected value for B-form DNA. The solid line connects the average value of the parameter for each base, base pair, or base pair step, with the range seen in the family of 20 structures shown by the vertical bar at each point.

not been observed with unlinked complexes and indicated slow dissociation of ligand from the complex. Measurements were repeated at high pH (9.1), except starting with deuterated ligand and exchanging in protons. High pH accelerates the rate of exchange of free ligand, so that each dissociation event should lead to exchange. The most slowly exchanging amide on the ligand still had a lifetime of about 8 min, which is hence a reasonable estimate of the lifetime of the complex. The opening rates of the base pairs, determined from imino proton exchange rates, indicated a stabilization of the base pairs in the complex relative to the free DNA.

Discussion

The solution structure of the DNA/cyclic polyamide complex is consistent with the binding site identified in previous footprinting studies.[20] The basic features are characteristic for this class of minor groove ligands.[2,38] There is good van der Waals contact between the stacked ligands and the wall of the DNA minor groove and extensive hydrogen bonding between ligand amide protons and acceptor groups on the DNA bases. A hydrogen bond between the *N*-methylimidazole and the guanosine amino group anchors the cyclic polyamide, defining its orientation on the DNA. The structure indicates that the positively charged extension of the derivatized pyrrole ring of the ligand points away from the DNA minor groove.

The average pairwise rmsd for the central region of the 20 final NMR structures is about 1 Å, giving a fairly good definition of features in the complex. It is of interest to compare the NMR-derived structure for this cyclic ligand complex with those of crystal structures, as well as previous lower resolution NMR structures, of unlinked side-by-side polyamide complexes. In the first 2:1 complex crystallized, distamycin with $d(ICICICIC)_2$,[5] the two distamycin molecules are well stacked in the minor groove of the DNA, with the pyrrole rings of one ligand stacked upon the amide groups of the other. This stacking arrangement was defined in the first NMR studies, and has since been found in crystal structures of other side-by-side complexes,[1,39] indicating a minimum configuration in the stacking energy. The two crescent-shaped distamycins twist and wrap around the helical surface of the minor groove and extend to nearly the full length of the octamer DNA duplex. In the $(ICICICIC)_2$ complex, the two distamycins overlap for two of the three pyrrole rings, although NMR data indicate that in such sequences probably there is an equilibrium between the complex with all three rings overlapped, and the partially overlapped mode found in the crystals.[38] Very similar stacking was also found in the crystal

(38) Pelton, J. G.; Wemmer, D. E. *J. Am. Chem. Soc.* **1990**, *112*, 1393−1399.

(39) Chen, X.; Ramakrishnan, B.; Sundaralingam, M. *J. Mol. Biol.* **1997**, *267*, 1157−1170.

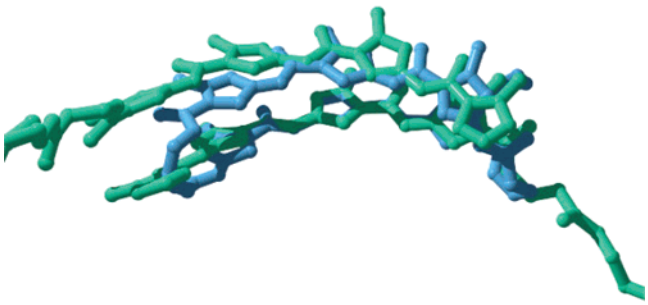


Figure 7. An overlay is shown of the cyclic polyamide from the complex (blue) and the unlinked ligands ImPyPyPy from an X-ray crystal structure (Kielkopf, 1998b). The central ring of the lower ligand and the central ring of the cyclic ligand were overlapped to clearly see the difference in relative position of the upper portion of the ligand (or second ligand).

structure of ImPyPyPy bound in a 2:1 complex with d(CCAG-TACTGG)$_2$.[40] This supports the notion that the staggered conformation, in which the rings of one ligand stack on the amide groups of the other, is the most energetically favorable geometry. The cyclic polyamide ligand in complex with DNA forms a very similar crescent shape to conform to the shape of the DNA minor groove. The most obvious difference between the distamycin dimer and the cyclic polyamide is that, in the cyclic polyamide, the rings of one side of the analogue stack immediately above on the rings of the other (Figure 7). This geometry is enforced by the relatively short linkers at each end of the ligand. The different stacking arrangement changes the positions of hydrogen bond donors on the ligand, which point into the groove.

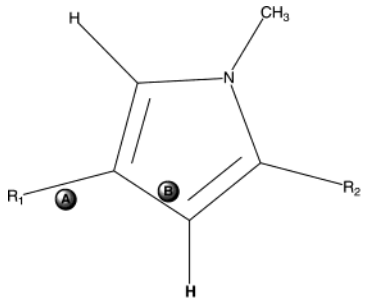
The DNA base pairs in the binding site are somewhat less regular in the complex than in free DNA or in the unlinked complexes that have been reported. We do not have a structure of the unliganded DNA of this specific sequence to make a detailed comparison. Helix parameters are shown in Figure 6 as a function of position along the DNA. In the structures, there is a slight displacement of base pairs away from the helix axis, to a larger extent than seen in the crystal structure of distamycin with (ICICICIC)$_2$. There is also an inclination of the base pairs, maximizing at the center of the binding site. The T4−A19 base pair has a high propeller twist and is somewhat buckled, apparently due to contact with the -γ- linker that abuts it. The distortion of this base pair seem somewhat larger than that at A8−T15, which is contacted by the other -γ- linker. It may be that proximity to the imidazole, which makes a hydrogen bond to G5, leaves it in a somewhat different position. It is also possible that the flanking propylamine-derivatized pyrrole affects the local geometry in trying to optimize electrostatic interactions and that this could be related to the reduced sequence specificity of this particular ligand. The longer cyclic pyrroles, which yield good specificity and have the amine moved onto one of the linkers, might have less structural distortion at this point. One sugar on each strand is displaced from the normal C2′ endo for B-DNA. On the strand contacted by the Im residue it is G5 (also having an unusual phosphate position), and on the opposite strand it is T15. The groove width assessed by the C1′ to C1′ distance is about the same as that of regular B-DNA (most in the 10.7−11 Å range, B-DNA 10.9 Å), but the phosphate−phosphate distance (measured P to P) is somewhat expanded (13−15 Å, B-DNA 11.7 Å), as has been seen in other 2:1 complexes, to accommodate the two ligands. Because the phosphate positions are not determined directly by the NMR data collected, this parameter is not very well determined by them. However, in the structures, the phosphate between G5 and T6 is consistently different in local geometry from the others, seemingly pushed out by contact with the Im ring of the ligand. Of course, it is not possible to separate and assess the energetics of local distortions, but it is likely that some of these reflect contacts that could affect affinity and specificity.

Although some DNA base pairs are distorted in the structure of the complex, none of the base pair hydrogen bonds is broken upon binding to the ligand, and all expected imino resonances were observed in spectra taken in H_2O solutions. From the structure of the complex, nine intermolecular hydrogen bonds could be identified between the ligand and DNA. Base pair A8T15 has one hydrogen bond to the ligand, C7G16 has three, A6T17 has two, A5T18 has two, and T4A19 has one (Figure 5b). Overall, this array of hydrogen bonds between ligand and DNA is similar to what has been seen in unlinked complexes. The fact that the relative positions of donors in the ligand are different, as discussed above, suggests that the DNA structure rearranges slightly to enable hydrogen bonding. The imino exchange data show that the opening rates of the central, most contacted base pairs C7G16, A6T17, and A5T18 are decreased with respect to the free DNA duplex by about 2 orders of magnitude, while those for base pairs next to the -γ- linkers, A8T15 and T4A19, are only decreased slightly, if at all. The central base pairs with more hydrogen bonds to the ligand show increased stability, indicating that these are favorable interactions despite local DNA distortions. In the two A−T pairs at the edges of the binding site, the single hydrogen bond in each case is to the adenosine. The lack of stabilization of these base pairs by the presence of the ligand suggests that the thymine can shift to the open state without breaking the hydrogen bond to the partner A. It should be possible to test this idea in the future by comparing the exchange behavior in the complex for which the A−T pairs are swapped across the strands, in which case the hydrogen bond would probably be to the T carbonyl.

Unusual chemical shifts for deoxyribose H4′ resonances were observed for some residues in the complex. NOESY data indicated resonances at 2.78 ppm (A6 H4′), 1.86 ppm (C7 H4′), 2.00 ppm (T17 H4′), 2.29 ppm (T18 H4′), and 2.53 ppm (A19 H4′), which were confirmed with the natural abundance HSQC data (Figure 4). The structure determined using complete matrix analysis shows that these protons are immediately above the pyrrole/imidazole rings of the polyamide (Figure 5b). The structure of the complex indicates the upfield shifted chemical shifts are caused by ring currents, with the magnitude of shift correlated with the distance between the shifted proton and the center of the aromatic ring. This provides independent validation of the orientation of the ligand in the minor groove.

A similar trend is observed for the calculated chemical shifts of the polyamide ligand ring protons. Figure 8 shows ring-current calculations for the pyrrole ring protons that point into the minor groove, H4, H8, H16, H20, and H24. These chemical shifts are primarily influenced by the location of the closest ring on the "other half" of the cyclic ligand. We examined both the "ring-over-ring" conformation as found in the current NMR refinement and the "ring-over-amide" conformation that re-

(40) Kielkopf, C. L.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Nat. Struct. Biol.* **1998**, *5*, 104−109.



Proton	Chemical Shift A (ppm)	Chemical Shift B (ppm)
H4	7.40	6.94
H8	6.82	6.53
H16	5.54	6.14
H20	6.90	5.82
H24	6.66	6.44

Figure 8. A typical pyrrole ring in the cyclic ligand showing the relative stacking interaction and calculated chemical shifts of ligand protons (numbering scheme given in Figure 1A). The solid circles A and B correspond to the projection of the center of the ring opposite it in the ligand. Circle A represents ring-over-amide stacking (generated artificially to mimic the conformations seen in 2:1 noncyclic complexes), whereas circle B represents the ring-over-ring stacking found in the current NMR refinement. The chemical shifts were calculated using the SHIFTS program at http://www.scripps.edu/case. Values for ring susceptibility anisotropies were extracted from density function calculations following the procedure of Case (1995): −99.0 (ppm Å^3)/molecule for imidazole and −96.4 (ppm Å^3)/molecule for pyrrole.

sembles the geometry seen in 2:1 complexes of noncyclic ligands.[40] As the stacking of the side-by-side ligands approaches ring-over-ring, the proton resonance typically shifts upfield, relative to that of the same proton in a structure whose stacking is ring-over-amide. As Figure 8 shows, the orientation resulting from "ring-over-ring" fits these shifts better than that from "ring-over-amide", again providing an independent validation of our model for the ligand conformation.

The measured exchange rates of the ligand amide protons indicate that dissociation of the cyclic polyamide from the complex is slow. Rather similar results based on fluorescence measurements have recently been reported for another cyclic ligand containing a total of eight Py/Im rings.[41] The association rates are high for both the hairpin and the cyclic ligands, with the factor of 2 higher rate for the cyclic version corresponding to the increase in affinity. The amide exchange behavior of the cyclic complex described here is consistent with the same pattern for six ring hairpins and cyclic ligands. Further work comparing the behavior of hairpin versions of ligands may help to understand more about the mechanism of dissociation.

The relaxation matrix analysis of NMR data, together with modern molecular mechanics calculations, yields solution structures of ligand−DNA complexes that can help in understanding the details of the interactions that lead to affinity and specificity. Hairpin and cyclic polyamides are synthetic ligands that have affinity and specificity for DNA comparable to that of many naturally occurring DNA-binding proteins. As such, these small molecule ligands can control gene expression, for example, an eight-ring polyamide targeted to a specific region of the transcription factor TFIIIA binding site interfering with 5S RNA gene expression in Xenopus kidney cells,[42] and two polyamides binding DNA sequences immediately adjacent to binding sites for the transcription factors Ets-1, lymphoid-enhancer binding factor 1, and TATA-box binding protein, which inhibit HIV type 1 transcription in cell-free assays.[43] However, despite these successes, it is clear that the understanding of the combinations of affinity and specificity that are required for good in vivo activity is incomplete. Furthermore, for applications, in vivo cellular uptake, clearance, and metabolism are also issues. The linked polyamide ligands are of higher molecular weight than most drugs, and so minimizing size while maximizing binding affinity is important. The lower solubility of the cyclic ligand relative to hairpins could reflect higher lipophilicity and hence indicate better penetrability into cells, which is also desirable. However, it is important to understand the binding modes of the ligands, and hence the rules for targeting them. The structures reported here show that the cyclic ligand has slightly altered stacking of the Im/Py segments, but despite this binds to the DNA and interacts with it in a manner very similar to that of the more studied hairpin versions of these polyamide ligands.

Acknowledgment. This work was supported by NIH grants GM45811 (to D.A.C.) and GM43129 (to D.E.W.), and GM 27681 (to P.B.D.), and F33 GM63420 (to T.J.D.).

Supporting Information Available: Tables of electrostatic charges and AMBER parameters developed for polyamide ligands; ring-current information and the structure of the ligand: DNA complex in PDF format. This material is available free of charge via the Internet at http://pubs.acs.org.

JA0373622

(41) Baliga, R.; Baird, E. E.; Herman, D. M.; Melander, C.; Dervan, P. B.; Crothers, D. M. *Biochemistry* **2001**, *40*, 3−8.

(42) Gottesfeld, J. M.; Neely, L.; Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1997**, *387*, 202−205.

(43) 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.