

Structure of a β -Alanine-linked Polyamide Bound to a Full Helical Turn of Purine Tract DNA in the 1:1 Motif

Adam R. Urbach¹, John J. Love², Scott A. Ross^{1,3} and Peter B. Dervan^{1*}

¹*Division of Chemistry and Chemical Engineering
California Institute of Technology, Pasadena, CA 91125, USA*

²*Department of Chemistry, San Diego State, San Diego, CA 92182, USA*

³*Howard Hughes Medical Institute, Pasadena, CA 91125 USA*

Polyamides composed of *N*-methylpyrrole (Py), *N*-methylimidazole (Im) and *N*-methylhydroxypyrrole (Hp) amino acids linked by β -alanine (β) bind the minor groove of DNA in 1:1 and 2:1 ligand to DNA stoichiometries. Although the energetics and structure of the 2:1 complex has been explored extensively, there is remarkably less understood about 1:1 recognition beyond the initial studies on netropsin and distamycin. We present here the 1:1 solution structure of ImPy- β -Im- β -ImPy- β -Dp bound in a single orientation to its match site within the DNA duplex 5'-CCAAAGAGAAGCG-3'/5'-CGCTTCTCTTTGG-3' (match site in bold), as determined by 2D ¹H NMR methods. The representative ensemble of 12 conformers has no distance constraint violations greater than 0.13 Å and a pairwise RMSD over the binding site of 0.80 Å. Intermolecular NOEs place the polyamide deep inside the minor groove, and oriented N–C with the 3'–5' direction of the purine-rich strand. Analysis of the high-resolution structure reveals the ligand bound 1:1 completely within the minor groove for a full turn of the DNA helix. The DNA is *B*-form (average rise = 3.3 Å, twist = 38°) with a narrow minor groove closing down to 3.0–4.5 Å in the binding site. The ligand and DNA are aligned in register, with each polyamide NH group forming bifurcated hydrogen bonds of similar length to purine N3 and pyrimidine O2 atoms on the floor of the minor groove. Each imidazole group is hydrogen bonded *via* its N3 atom to its proximal guanine's exocyclic amino group. The important roles of β -alanine and imidazole for 1:1 binding are discussed.

© 2002 Elsevier Science Ltd. All rights reserved

Keywords: NMR structure; β -alanine; polyamides; purine tract; DNA recognition

*Corresponding author

Introduction

High-resolution structural studies of small-molecule-DNA interactions improve our understanding of molecular recognition processes and aid in the design of sequence-specific DNA-binding ligands. Two global motifs are well-recognized for targeting the minor groove of DNA based on analogues of the *N*-methylpyrrole-carboxamide

ring (Py) of the natural products netropsin and distamycin. There exist 1:1 and 2:1 ligand-DNA stoichiometries with quite different rules for recognition and, hence, different sequence specificities.^{1–3} This ambiguity of sequence targeting by a single ligand, depending on stoichiometry, raises an important design issue for the field.⁴ The structure and energetics of the 2:1 motif have been characterized extensively;^{3,5} however, less is understood about 1:1 recognition beyond the initial studies on netropsin and distamycin. Netropsin binds 1:1 in the minor groove of *B*-form DNA in A/T regions.⁶ The minor groove of A/T regions is both narrow due to propeller twisting of the base-pairs and relatively deep due to lack of the protruding 2-amino group on guanine.⁷ NMR and X-ray structures of netropsin and distamycin bound to DNA reveal that the two (and three) Py rings fit snugly in the minor groove in a 1:1 complex.^{1,8,9}

Abbreviations used: COSY, correlated spectroscopy; Dp, dimethylaminopropylamine; DQF, double quantum filtered; Hp, *N*-methylhydroxypyrrole; Im, *N*-methylimidazole; MPE, methidiniumpropyl-EDTA; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; Py, *N*-methylpyrrole; rMD, restrained molecular dynamics; SA, simulated annealing; β , β -alanine.

E-mail address of the corresponding author: dervan@caltech.edu

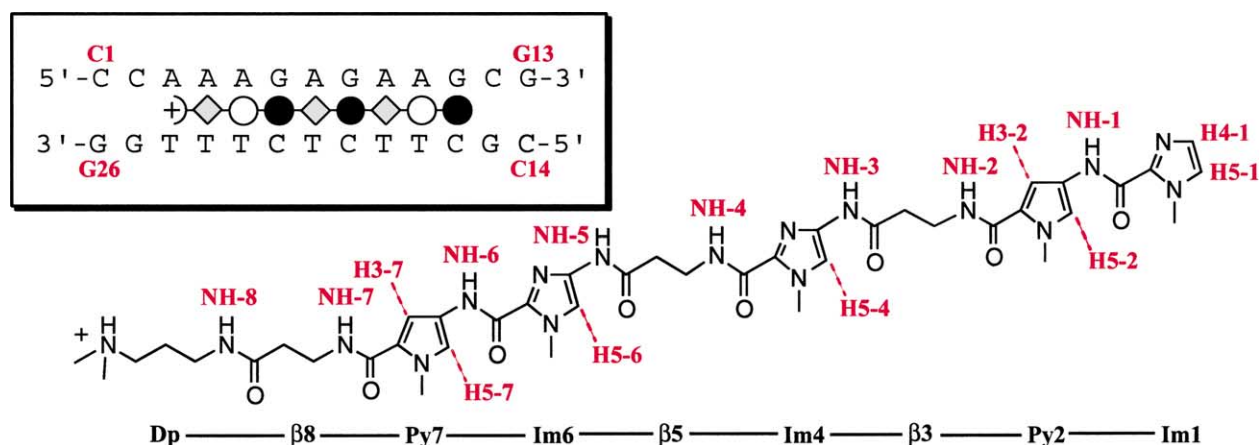


Figure 1. Nomenclature for the 1:1 complex studied by NMR. The chemical structure of polyamide 1 is shown in black with relevant protons labeled in red. Residue names are shown below the structure, linked by line segments. Inset: DNA sequence with a dot model of the bound polyamide. DNA residue numbers are indicated in red, along with a dot model of the bound ligand. Filled circles represent imidazole; open circles represent pyrrole; diamonds are for β -alanine; and the half-circle is the dimethylaminopropylamide tail.

Understanding the sequence-dependent microstructure of DNA is a key issue for the field of ligand-DNA interactions. Analysis of numerous B-DNA single-crystal X-ray structures reveals that certain base-base steps are more deformable than others.¹⁰ In particular, purine-purine steps such as A-A and G-A are inclined to be rigid structures with a narrow minor groove and large negative propeller twist. It has been suggested that optimal base stacking is the primary factor governing this feature.¹¹ The single spine of hydration observed in the minor groove of A-tract-like structures is also thought to provide stabilization of this geometry.¹⁹ Polyamides displace the spine of hydration in a multidentate fashion by forming hydrogen bonds from each amide NH to proximal purine N3 and pyrimidine O2 atoms,^{1,12,13} which provides a large entropic driving force for binding.¹⁴ The narrow minor groove of A-tracts accommodates netropsin and distamycin in a 1:1 complex well, but the plasticity of T-A steps is required in order to expand the minor groove for accommodation of two ligands.^{2,10}

The X-ray structure of netropsin bound as a 1:1 complex to DNA inspired the lexitropsin model, where it was predicted that replacing one or both Py residues in netropsin with *N*-methylimidazole-carboxamide (Im) would confer G-C recognition by simultaneously alleviating a steric interaction with the C3-H of Py and forming a hydrogen bond from Im-N3 to G-NH2.¹ Subsequent footprinting experiments revealed that Im-Py oligoamides tolerated G-C base-pairs, but were not very sequence-specific, presumably due to Im tolerance of A-T base-pairs as well.¹⁵ Remarkably, the structural basis for the lexitropsin model, as envisioned in the 1:1 motif, was never verified structurally.

In 1989, Wemmer and co-workers made the important observation that distamycin A can bind A-T tracts of DNA in an antiparallel 2:1 fashion, as well as 1:1, even at low ligand/DNA stoichiometries (i.e. Py/Py pairs, as well as Py, prefer A/

T over G/C).² It was discovered subsequently that the unsymmetrical ring pair Im/Py could distinguish G-C from C-G and both from A-T and T-A.¹⁶ Further invention of the new ring pair Hp/Py completed a recognition code to target all four Watson-Crick base-pairs of DNA.^{3,5,13,17} However, certain sequence contexts, most notably those containing 5'-GA-3' and 5'-GNG-3' steps, have emerged as difficult targets for 2:1 binders. The aliphatic amino acid β -alanine (β), previously used to relax the over-curvature of long contiguous aromatic subunits,^{18–20} was incorporated between imidazole groups as Im- β -Im to target 5'-GNG-3' successfully.²¹

In a breakthrough report, the 1:1 motif was revived by Laemmli and co-workers and used to effect phenotypic changes in *Drosophila melanogaster* by binding to 5'-GAGAA-3' repeats in satellite regions of the fly genome.²² In the Laemmli design, β -alanine residues were used to add flexibility between ring subunits, allowing the Im residues to register optimally with the DNA helix. The resulting molecules bound 1:1 with subnanomolar affinity in a strongly oriented fashion. In order to understand how to better control 1:1 binding, efforts have been made recently to determine the sequence requirements for 1:1 recognition.²³ It was found from footprinting studies that Im residues tolerate G/C base-pairs, whereas other residues, including Py and β , are A/T-specific presumably for steric reasons.

In the course of these experiments, several observations were made that would require structural data for explanation. In particular we were interested in elucidating the role of β -alanine^{4,18} as well as the sequence-dependent orientation of the ligand.^{22,23} In order to understand this unique mode of recognition, we set out to solve the high-resolution solution structure of an Im and β -rich polyamide bound to its purine-rich cognate match site in a 1:1 complex. We present here the structure

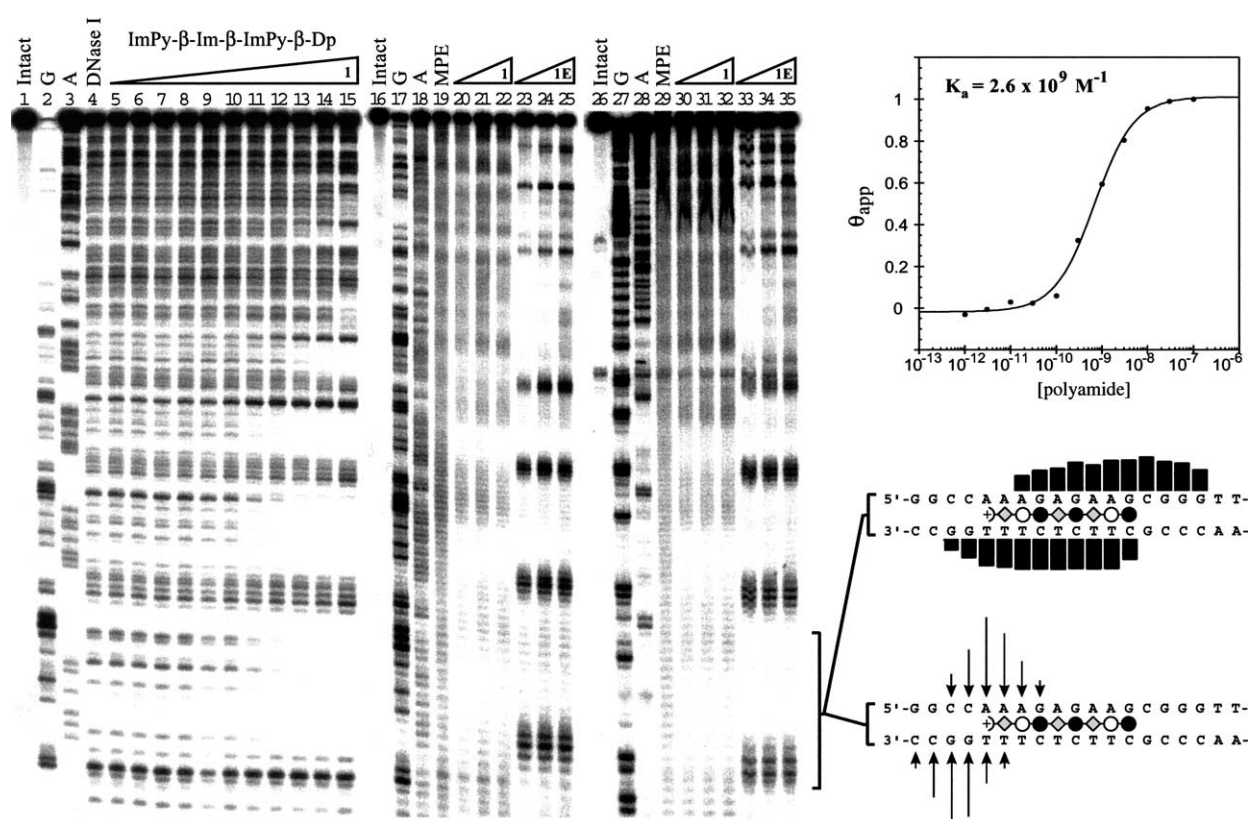


Figure 2. Biophysical characterization. Lanes 1–15, quantitative DNase I footprint titration experiment for polyamide 1 on the 289 bp 5'-end-labeled PCR product from pAU20: lane 1, intact DNA; lane 2, G reaction; lane 3, A reaction; lane 4, DNase I standard; lanes 5–15, 1, 3, 10, 30, 100, 300 pM, 1, 3, 10, 30, 100 nM 1, respectively. Lanes 16–25, MPE and affinity cleavage experiments with 1 and 1E, respectively, on the 289 bp 5'-end-labeled PCR product from pAU20: Lane 16, intact DNA; lane 17, G reaction; lane 18, A reaction; lane 19, MPE standard; lanes 20–22, 10, 30, 100 nM 1, respectively; lanes 23–25, 10, 30, 100 nM 1E, respectively. Lanes 26–35, MPE and affinity cleavage experiments with 1 and 1E, respectively, on the 289 bp 3'-end-labeled restriction fragment of pAU20: lane 26, intact DNA; lane 27, G reaction, lane 28, A reaction, lane 29, MPE standard; lanes 30–32, 10, 30, 100 nM 1, respectively; lanes 33–35, 10, 30, 100 nM 1E, respectively. Right: Analysis of the polyamide-d(CCAAAGAGAAGCG)-d(CGCTTCTCTTTGG) binding site at the bottom of the gels: top right, binding isotherm for the DNase titration with the binding constant (K_a), as determined from a non-linear least-squares fit, shown in bold type; (bottom right) schematics illustrating observed protection (middle) and cleavage (bottom) patterns derived from the MPE and affinity cleavage experiments, respectively. The relative heights of the bars and arrows indicate relative intensities of protection and cleavage, respectively. Polyamides are drawn as oriented, 1:1 complexes, as observed in these experiments.

of the polyamide ImPy- β -Im- β -ImPy- β -Dp (Dp, dimethylaminopropylamine) bound to a full turn of the 13-mer DNA duplex 5'-CCAAAGAGAGCG-3'/5'-CGCTTCTCTTTGG-3' (match site in bold) (Figure 1). From the high-resolution structure we are able to observe directly the hydrogen bonds between Im-N3 and G-NH2 in the 1:1 motif.^{1,15} The role of β -alanine for 1:1 binding in this purine-rich tract is discussed. In addition, the structure reveals a possible explanation for the sequence-dependent orientation of polyamides in the 1:1 motif.

Results

Footprinting and affinity cleavage

To simplify the NMR spectral assignments, we screened several truncated polyamides to identify

the minimal complex size that retains the established characteristics of 1:1 polyamide/DNA recognition.²³ In the last round of this iterative process of design, synthesis, and characterization, three DNA-binding sites of sequence type 5'-ccAAAGAGAAGA_ncg-3' (flanking G-C clamp-sare in lower-case) were examined as candidates for NMR studies, such that the only variable among the sites was the number of A-T base-pairs (A_n , $n = 0, 1$, and 2) beyond the N terminus of the polyamide-binding site. To determine binding affinity, site size, and orientation, the DNA sequences were cloned into plasmid pAU20 and characterized in complex with polyamide 1 (Im-Py- β -Im- β -Im-Py- β -Dp) by DNase I and methidiumpropyl-EDTA (MPE) footprinting, as well as affinity cleavage analysis (Figure 2). Quantitative DNase I footprinting of compound 1 on pAU20 revealed essentially equivalent binding affinities for the three sites at $K_a \sim 2$ nM, indicating

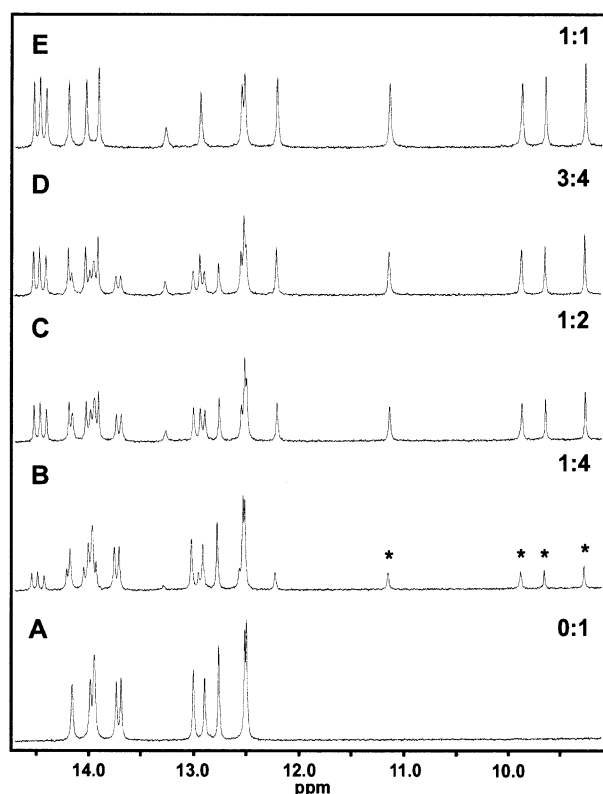


Figure 3. Far downfield region (9.0–14.5 ppm) of the ^1H NMR (at 600 MHz, 25 °C) spectra in 9:1 $^2\text{H}_2\text{O}:\text{H}_2\text{O}$ of d(CCAAAGAGAAGCG)·d(CGCTTCTCTTTGG) with (a) no ligand added; (b) 1:4 ligand/DNA; (c) 1:2 ligand/DNA; (d) 3:4 ligand/DNA; and (e) 1:1 ligand/DNA stoichiometry. Molar ratios are indicated at the right. Ligand aromatic amide NH protons are indicated by asterisks. Peaks at the left are from thymine and guanine imino protons.

that the additional A·T base-pairs (A_n) were not necessary to stabilize the complex. MPE and affinity cleavage experiments of compounds **1** and **1E**, respectively, show the polyamide bound to its match sequence, 5'-AAAGAGAAG-3', oriented N–C with the 3'–5' direction of the purine-rich strand. Both experiments show a 3'-shift between upper and lower strands, characteristic of minor groove binding.²⁴ The smallest sequence, 5'-ccAAAGAGAAGcg-3', was chosen for NMR studies because it retains 1:1 polyamide/DNA binding properties and high affinity.

Titration to 1:1 polyamide to DNA stoichiometry

The titration of polyamide **1** to the NMR sample of d(CCAAAGAGAAGCG)·d(CGCTTCTCTTTGG) is shown in Figure 3. Chemical shift perturbation in the uncrowded DNA imino region upon addition of polyamide was used to monitor the degree of complex formation. The number of imino peaks doubles upon addition of sub-stoichiometric amounts of polyamide. A single set of imino peaks is restored at a 1:1 polyamide to DNA stoichiometry. Only one set of polyamide

signals is observed, as indicated with asterisks in Figure 3, which further confirms a 1:1 ligand to DNA stoichiometry. Terminal base-pair imino resonances were not observed.

Spectral assignments

Methods for assigning protons in polyamide·DNA complexes have been well-established.^{2,25–29} The intramolecular DNA cross-peaks were assigned as described:³⁰ 16 pairs of H5'/H5'' protons were assigned non-stereospecifically. All other non-exchangeable protons were assigned, and all exchangeable protons were assigned except for those on the terminal C1·G26 and C13·G14 base-pairs. Sequential assignments for guanine and thymine imino protons were determined from the NOESY spectrum in protiated solvent. These facilitated the assignment of A-H2 and C-amino protons. The C-amino protons correlate strongly to each other and to the vicinal C-H5 proton, allowing identification of C-H5–C-H6 crosspeaks. It is worth pointing out that the process of assigning imino → amino → A-2H/CH5/CH6 was straightforward, and proved essential for the unambiguous assignment of the aromatic-H1' region. NOEs were observed between each A-H2 proton and the H1' protons of its 3'-neighbor and its base-paired thymidine's 3'-neighbor. This pattern is indicative of a narrow minor groove.³¹

The nomenclature for polyamide protons is shown in Figure 1. All polyamide protons were assigned. Amide NH, Py-H3, and Im-H4 protons were assigned on the basis of intra- and intermolecular connectivities, as observed in the NOESY spectrum in H_2O . Methylene protons for each β -alanine residue were identified by a unique set of strong intermolecular contacts with each A-H2 proton. Geminal pairs were assigned non-stereospecifically on the basis of the patterns observed in the DQF-COSY and NOESY experiments. Ring H5 protons were identified by contacts to H4' ribose protons as well as strong contacts to the respective proximal *N*-methyl group. The intermolecular connectivity pattern observed for these protons defines the orientation of each Im and Py residue such that the *N*-methyl group points out of the groove. Polyamide chemical shift values are similar to those reported by Wemmer and coworkers for 1:1 and 2:1 polyamide complexes.^{20,26,29}

The NOESY spectrum of the 1:1 complex in H_2O is shown in Figure 4. Sequential ligand assignments were based on a clear pattern of intermolecular NOEs from ligand protons to assigned DNA protons. Strong intermolecular NOEs between Py-H3 and A-H2 protons, as well as contacts between amide NH and H1' protons in the binding site, orient the polyamide N–C with respect to the 3'–5' direction of the purine-rich strand. Each polyamide residue contacts H1', H4', and/or base protons on both strands of DNA. This is further evidence for 1:1 complex formation in the minor groove with a single binding orientation.

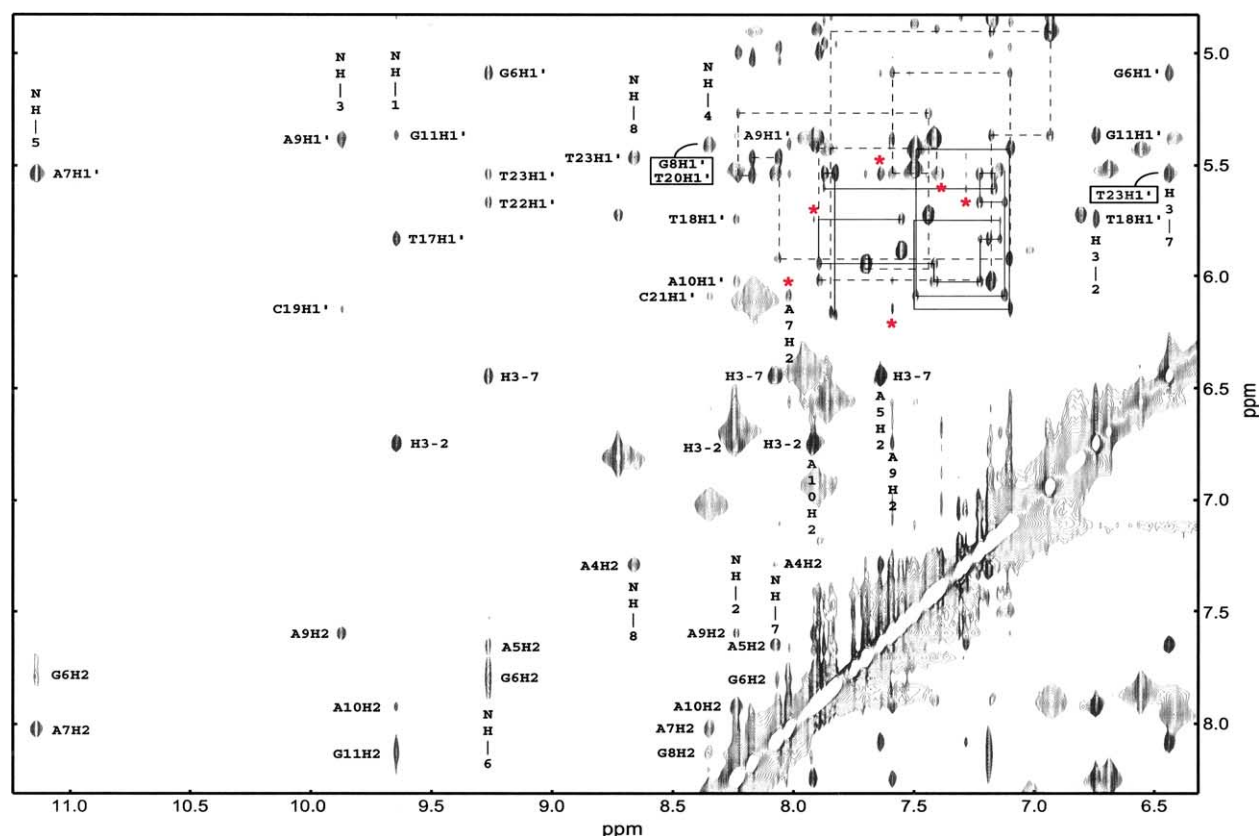


Figure 4. Expansion of the aromatic and amide region of the NOESY spectrum of the 1:1 polyamide-d(CCAAAGAG-AAGCG)-d(GCGTTCTCTTTGG) complex (9:1 H₂O: ²H₂O, 10 mM sodium phosphate (pH 7.0), 25 °C; 75 ms mixing). Sequential aromatic to H1' connectivities for the purine-rich strand are shown as dotted lines; those for the pyrimidine-rich strand as continuous lines. Crosspeaks are labeled according to their chemical shifts along ω_1 (vertical axis, label beside the peak) and along ω_2 (horizontal axis, label above or below the peak). Labeling conventions for the DNA are the residue name, the residue number, the proton name (e.g. A7H2 = adenine 7, H2 proton); the ligand is named as the proton name, the residue number (e.g. NH-7 = amide NH of pyrrole 7). Red asterisks above or below a peak indicate a cross-strand A-2H to H1' NOE (type *a* distance³¹).

Selected intermolecular contacts are illustrated in Figure 5, with continuous and broken lines indicating contacts to the floor and walls of the minor groove, respectively. A homogeneous distribution of contacts is observed, which defines the position of the ligand within the minor groove unambiguously.

Distance constraints

Polyamide binding induces a high dispersion of NOESY crosspeaks. This enabled the identification of a large number of discrete NOEs, which allowed the use of 508 distance constraints for structure calculations. The distribution of experimental constraints is relatively homogeneous throughout the binding region, with fewer constraints for the terminal base-pairs. The majority of constraints were derived from the NOESY spectrum in ²H₂O, with an additional 115 distance constraints from the NOESY spectrum in H₂O. Methods for converting NOE intensities to upper bound distance constraints are detailed in Experimental. Forty Watson-Crick hydrogen bond constraints were applied, based on the observed cytosine amino,

and guanosine and thymidine imino chemical shifts in the spectral region indicative of cross-strand hydrogen bonding. The final list of 548 experimental constraints is available from the Brookhaven Data Bank under accession code 1LEJ.

Structure calculations

Following the approach described by Chazin and co-workers, a significant effort was made during the structure calculations to sample and represent conformational space consistent with the input data.^{32,33} The starting ensemble of 40 structures, which differ considerably in helical geometry (RMSD 3.46 Å), were generated using the Nucleic Acids Builder program NAB†. These structures were positioned with the polyamide and docked using a restrained molecular dynamics (rMD) simulated annealing (SA) protocol in the AMBER 6.0 software package.³⁴ The docked structures converged to an RMSD of 1.37 Å, and subsequent rMD SA did not improve the total

† Macke & Case: <http://www.scripps.edu/case>

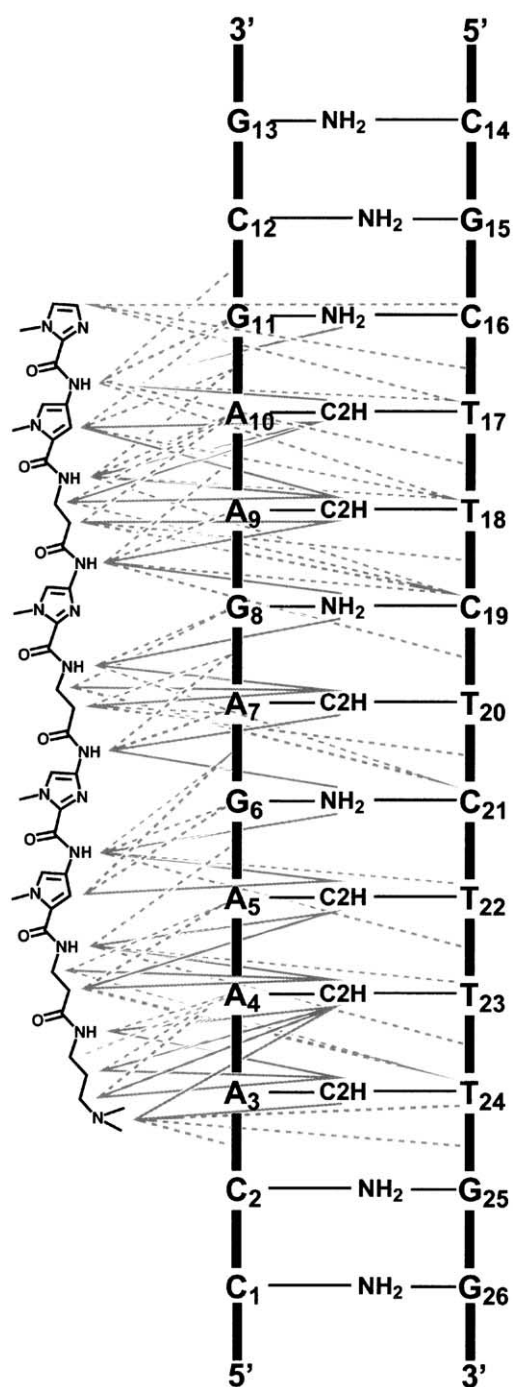


Figure 5. A representation of the 1:1 polyamide/DNA complex. Observed intermolecular NOEs from the polyamide to the floor and walls of the minor groove are indicated by continuous and broken gray lines, respectively. The DNA is shown as a ladder with residue numbers indicated at each rung in bold type. Connectivity lines to a residue number indicate a contact to H1'; those ending between residue numbers indicate H4' and/or H5'/H5'' protons.

energy or RMSD of the ensemble. The 40 docked structures were sorted by increasing residual constraint violation energy, and the 12 structures of lowest violation energy were chosen as the final ensemble (RMSD 1.12 Å). A summary of relevant

Table 1. Structural statistics for the final structural ensemble of the polyamide-DNA complex

E_{total}	$-1407.4 (\pm 17.0)^a$
$E_{\text{Lennard-Jones}}$	$-583.7 (\pm 14.7)^a$
$E_{\text{violation}}$	$0.9 (\pm 0.2)^a$
NOE violations ≥ 0.13 Å	0 ^b
NOE violations ≥ 0.10 Å	0.42 ^b
<i>Progression of the NMR refinement (RMSD, Å)^c</i>	
The 40 starting DNA structures	3.46
The 40 docked complexes	1.37
The 12 lowest constraint energy structures	1.12 (0.76)
Core eight base pair binding site ^d	0.80 (0.54)

^a Average AMBER energies in kcal/mol.

^b Average number of violations per structure.

^c Mean pairwise RMSD between the structures.

^d The core binding site includes all heavy atoms in DNA residues A4–G11, C16–T23, and polyamide residues Im1–Py7. The RMSD from the mean is given in parentheses.

statistics for the structure calculations is given in Table 1. These statistics demonstrate excellent agreement with the input data (low violation energy) as well as non-bonding stabilization of the structure due to ligand binding (large negative Lennard-Jones energy).

Discussion

Confirmation of oriented 1:1 binding

Several independent measurements confirm formation of the 1:1 ligand:DNA complex presented here, in which the polyamide is oriented N–C with respect to the 3'–5' direction of the purine-rich strand. The binding isotherm derived from quantitative DNase I footprinting (Figure 2) fits an $n = 1$ Hill equation.³⁵ The 1D NMR titration (Figure 3) shows the appearance of a single set of polyamide resonances with concomitant shifting of the DNA peaks at 1:1 stoichiometry. By contrast, polyamides that form 2:1 complexes with high cooperativity are known to do so even at low (0.25:1) ligand/DNA ratios, resulting in the existence of two sets of DNA resonances at 1:1 stoichiometry.^{2,20,26} Additionally, the NOESY spectrum reveals a large number of intermolecular contacts from ligand protons to both strands of DNA with similar intensity. This is in sharp contrast to reported 2:1 complexes, where ligands contact only the proximal strand.^{2,20,26}

The affinity cleavage and MPE footprinting experiments reveal a characteristic 3'-shift of cleavage and protection intensities, respectively, between the upper and lower DNA strands, which is indicative of minor groove binding.²⁴ The existence of cleavage on only one side of the binding site, even when the concentration of the polyamide is much greater than that of the DNA, is evidence for a single binding orientation. The pattern of intermolecular contacts observed in the NOESY spectra (Figures 4 and 5) confirms the

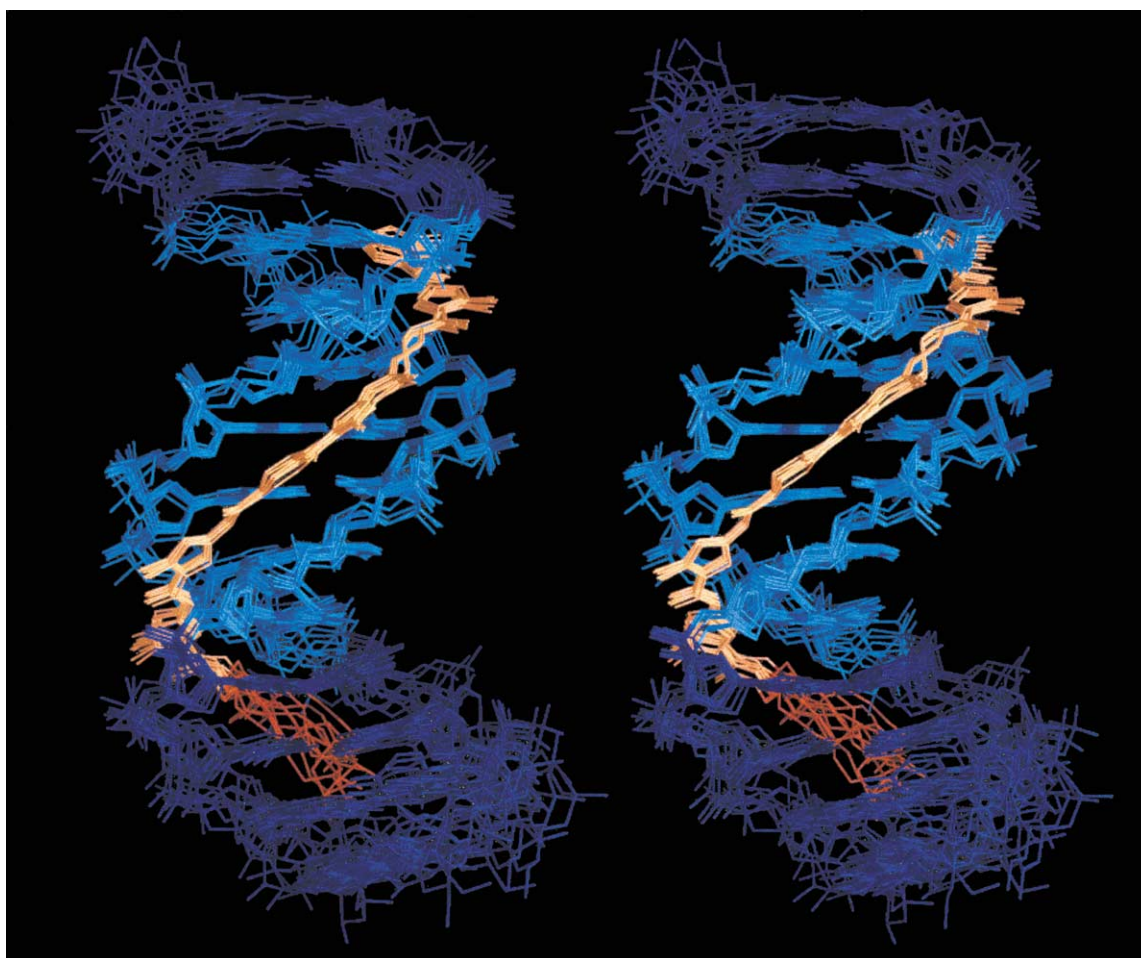


Figure 6. Stereo view of the final ensemble of 12 structures superimposed over the core binding site. The DNA is shown in blue, the ligand in gold. The binding site is indicated by a lighter color in the center of the complex.

DNA binding site and orientation, in addition to the ligand-DNA recognition code of Im to G-C, and Py and β to A/T base-pairs.

Characterization of the complex

Overall features

The final ensemble of 12 structures is shown in Figure 6 with the core binding site highlighted in the center. The binding site is well defined, and it consists of the five aromatic residues linked by two flexible β residues (ImPy- β -Im- β -ImPy). The RMSD for the ensemble is 1.12 Å, converging to 0.80 Å for all heavy atoms in the binding site (Table 1). The ends of the DNA as well as the C-terminal β -Dp tail of the polyamide sample a larger region of conformational space, and consequently are defined more poorly. The DNA has average B-form values for rise per residue (3.3 Å) and inter-base-pair twist (38°). Sugar pucker values span the range from C2'-*endo* to C1'-*exo*, as observed for A-tract structures,³⁶ with an average phase angle value of 148°.

Minor groove width and propeller twist

Figure 7 shows plots of minor groove width and propeller twist for the complex, which were calculated using the CURVES program, as described by Lavery and co-workers.^{37,38} Overall, the complex displays a narrow minor groove and a large negative propeller twist, which are features typically associated with A-tract structures.³⁹ The narrow minor groove is confirmed by the observation of medium-intensity cross-strand NOEs (type *a* distance³¹) between each adenine H2 and H1' of its base-paired thymidine's 3'-neighbor (red stars in Figure 4). The minor groove in the structural ensemble is much narrower than observed for ligand-free duplex DNA containing similar sequences-5'-AAAGAA-3' by NMR⁴⁰ as well as 5'-AAAGAAAA-3' and short A-tracts by X-ray crystallography.^{41,42} However, it has been proposed that ligand binding in a 1:1 mode can induce the walls of the minor groove to close down to maximize van der Waals contacts.^{10,43} Large negative propeller twist is commonly associated with a narrow minor groove. Lu and co-workers observe a significant decrease in propeller twist upon

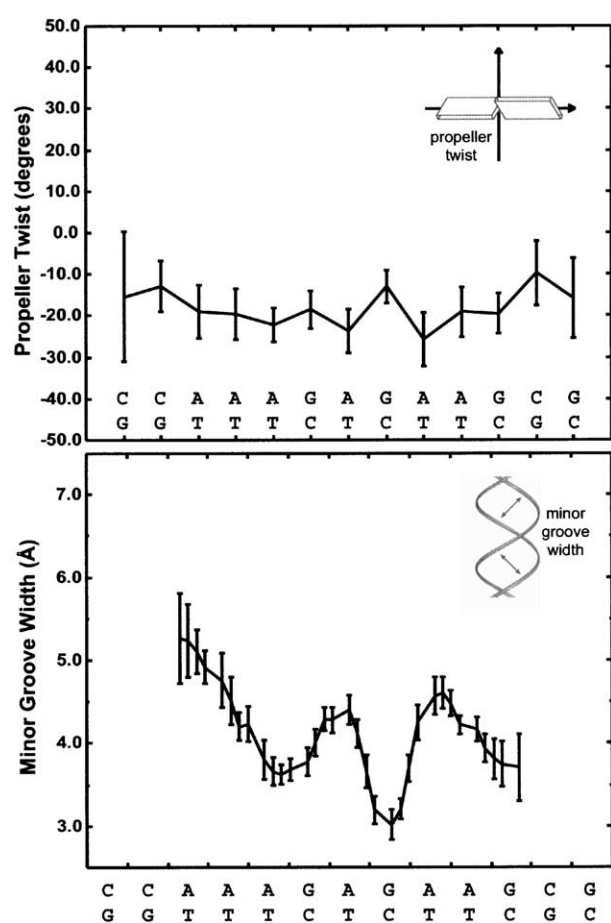


Figure 7. Plots of average propeller twist (top) and minor groove width (bottom) at each DNA residue, determined using the CURVES program as described by Lavery and co-workers.^{37,38} Average values are connected by a continuous line, and the *y*-axis error bars indicate one standard deviation from the average over the ensemble of 12 structures. Schematics illustrating the helical parameters are given in the upper right of each plot.

interruption of an A-tract with guanine, and they attribute this to a disruption in the spine of hydration by the G-NH2 group.⁴⁰ We observe a consistently large degree of propeller twist throughout the complex, even though there are

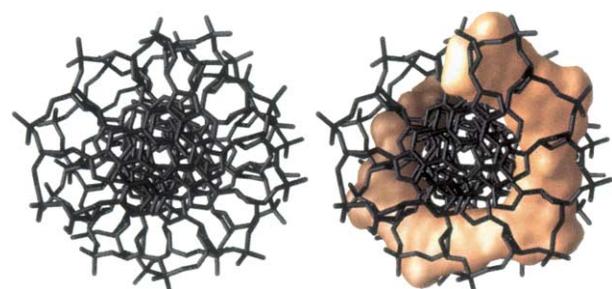


Figure 8. Top view of the polyamide-DNA mean structure shown (left) with and (right) without the ligand. The DNA is shown as a stick model in black, and the ligand is rendered with a molecular surface in gold.

multiple guanine bases interrupting the short A-tracts. Therefore, this effect is likely to be stabilized by the polyamide as discussed below.

Ligand structure

Figure 8 displays an axial view of the DNA helix with and without the polyamide shown. It is clear from this perspective that the DNA has a characteristic B-form geometry. The helical axis is relatively straight, and the polyamide wraps completely around the DNA, binding to a full turn of 9.5 base-pairs. The ligand is bound in the minor groove to its cognate nine base-pair match site, as determined by footprinting studies, with Im across from G-C, and Py and β across from A-T base-pairs. The orientation of the polyamide is N-C with the 3'-5' direction of the purine-rich strand.

The core binding site contains β residues in positions 3 and 5 (ImPy- β 3-Im- β 5-ImPy; see Figure 1). The structural ensemble reveals a single orientation for β 5. However, equal populations of two binding modes are observed for β 3, which adopts a straight conformation and a bent conformation with virtually identical AMBER energies. We were interested chiefly in assessing the impact of these two conformations on DNA recognition. Therefore, we averaged the coordinates for the five bent and the seven straight conformers within the final ensemble, minimized the mean structures, and then superimposed the flanking Py2 and Im4 rings of the two mean structures. The superposition reveals an interesting result: the two β -alanine conformations do not affect the positioning or relative orientation of the flanking aromatic rings. Specifically, the dihedral angles between the aromatic ring planes of Py2 and Im4 are virtually identical (30° and 33°) in spite of the two β -alanine conformations. This result could be attributed to a higher definition of position for β 5 from intermolecular NOEs, or it could be that the role of β -alanine is more to provide the flexibility needed for polyamide residues to properly align with DNA base-pairs and less for specific base recognition.

Amide-DNA interactions

In the absence of a ligand, minor groove hydration is thought to stabilize the propeller twisted base-pairs *via* hydrogen bonding to the N3 and O2 groups on the proximal purine and pyrimidine residues, respectively.⁴⁴ Netropsin¹ and distamycin⁹ have been observed to replace the spine of hydration by forming bifurcated bonds from each amide NH to the same N3 and O2 atoms in the 1:1 motif. Indeed, we observe this phenomenon. Figure 9 illustrates the interactions observed between ligand NH and purine-N3/pyrimidine-O2 groups, showing the core binding site of the NMR structure at left with the average (\pm standard deviation) distances given in the schema at the right. The amide NH groups to the

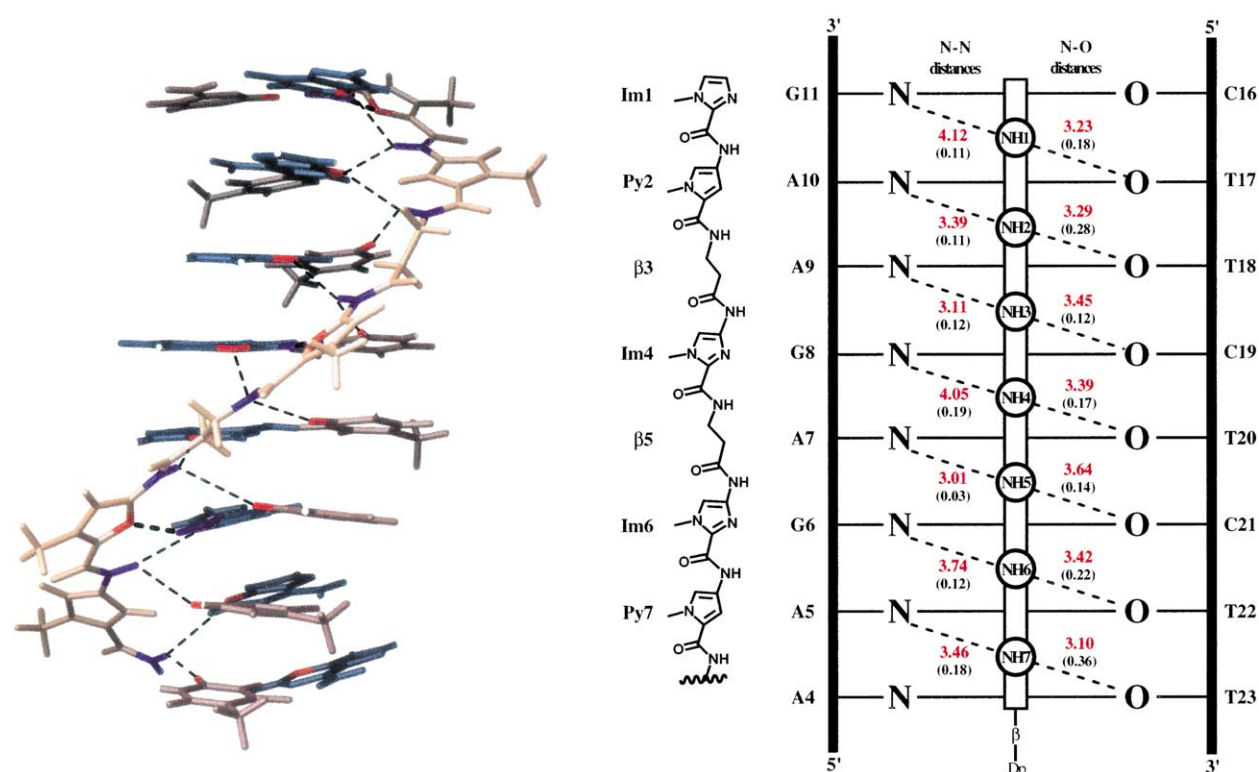


Figure 9. Schematic of polyamide NH to purine N3 and pyrimidine O2 contacts for DNA residues A4–G11–C16–T23 and polyamide residues Im1–Py7. The Figure has three parts, which are all in vertical register. Left: Mean structure of the polyamide-DNA complex showing the base-pairs as a stack down the center with purines in dark gray, pyrimidines in light gray, and ligand residues in yellow. Hydrogen bond donor and acceptor pairs are in blue and red, respectively. Amide NH to purine N3 and pyrimidine O2 interactions are indicated by broken lines; imidazole-N3 to guanine NH2 hydrogen bonds are shown as thick dotted lines. Center: Polyamide chemical structure. Right: Diagram of amide NH to purine N3 and pyrimidine O2 interactions. The DNA is shown as a ladder with each rung containing the residue number as well as purine N3 and pyrimidine O2 atoms indicated by bold N and O, respectively. The ligand is illustrated as a long bar with embedded circles containing the amide NH number, as defined in Figure 1. Broken lines connect NH to N3 and O2 atoms. Average distances over the final 12 structures for these interactions are given in red next to the broken lines, with standard deviations in parentheses.

C-terminal side of each Im residue have longer NH to N3 distances than would be ideal for hydrogen bonding. This small deviation could be due to competition from the Im-N3 to properly hydrogen bond to G-NH2. The distance values are roughly equal from the left to the right side of the minor groove, which shows the relatively central location of the ligand between the two DNA strands on the floor of the minor groove. The distances are also similar along the length of the DNA (top to bottom in Figure 9), attesting to an optimal ligand-DNA register along the length of the minor groove between polyamide and DNA residues for a complete turn of the helix. This is evidence for excellent shape complementarity between polyamide and DNA, and it suggests that the driving force for polyamide-DNA register is the proper alignment of amide NH groups with respect to the DNA base pairs.

The lexitropsin model

Dickerson and Lown proposed that substitution of one or more Py units in netropsin with Im

should confer G/C specificity in a 1:1 complex.^{1,15} Attempts have been made to verify this specific interaction in X-ray cocrystal structure analysis of imidazole-containing netropsin analogs with DNA sequences containing G/C base-pairs.^{45,46} However, Im does not bind across from G in these structures, and therefore the key Im-N3 to G-NH2 hydrogen bond in 1:1 binding has yet to be verified by a high-resolution structure. Considerable X-ray and NMR structural data exist to support the formation of Im-N3 to G-NH2 hydrogen bonds for the Im/Py pair in the 2:1 motif.^{12,26,27,47} In the 1:1 complex, we find a specific oriented hydrogen bond from Im-N3 to G-NH2. In fact, the ImPy sections of the polyamide presented here are structurally similar to the compounds originally synthesized by Lown and co-workers.^{15,45} Figure 10 shows a portion of the binding site with polyamide residues Im6–β5–Im4 binding across from DNA purine residues G6–A7–G8, respectively. The Im-N3 to G-NH2 hydrogen bonds (Im6–G6 lower left, Im4–G8 upper right) are shown by dotted lines.

Hydrogen bonding frequency was tested using the CARNAL module of the AMBER 6.0 software

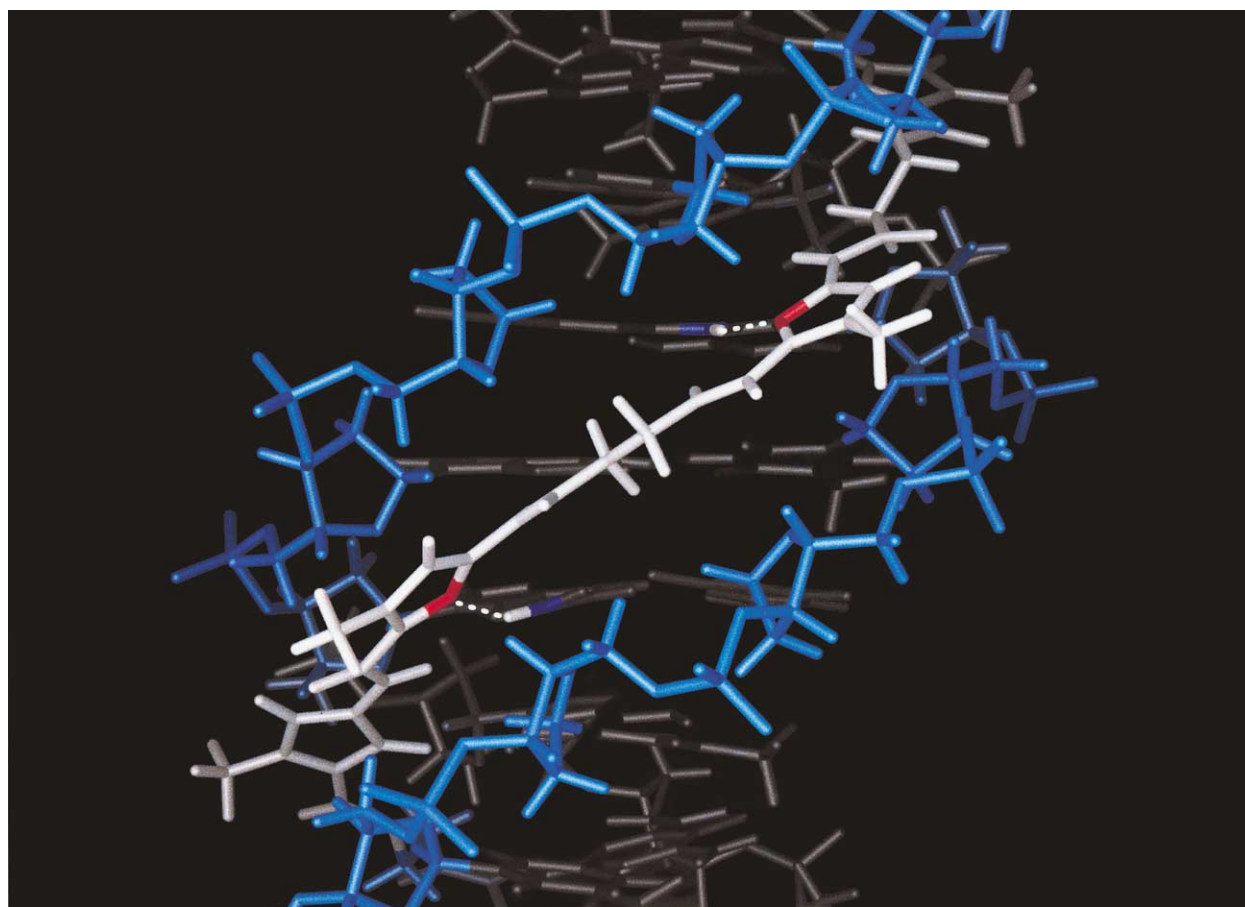


Figure 10. View into the minor groove of the mean structure, with polyamide residues Im6– β 5–Im4 proximal to DNA residues G6–A7–G8, respectively. The DNA bases are in black; the backbone is in blue; the polyamide is in white. Hydrogen bonds between imidazole N3 and guanine NH2 are shown by dotted lines connecting the donor (blue) and acceptor (red) pairs.

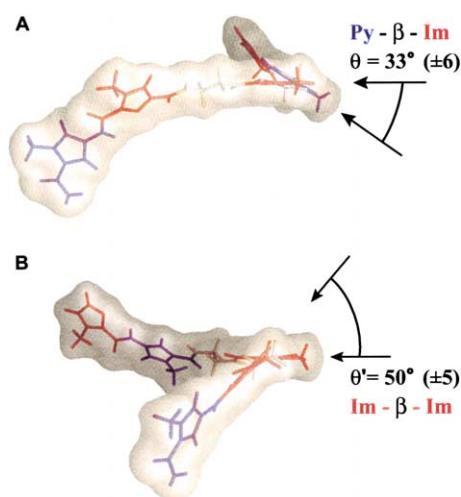


Figure 11. Polyamide ring- β ring dihedrals. The polyamide from the mean structure is shown as a stick model with a transparent molecular surface. Imidazole groups are in red, with β -alanine residues in white and pyrroles in blue. (a) View down the dihedral axis between Py2 and Im4 (Py- β -Im). (b) View down the dihedral axis between Im4 and Im6 (Im- β -Im). Arrows point along the plane of each ring, perpendicular to the view. θ and θ' prime denote the average ring to ring dihedral angles (\pm standard deviation) over the 12 final structures.

package.³⁴ The final mean structure was subjected to 15 ps of restrained molecular dynamics at 25 °C, and snapshots of the trajectory were analyzed for hydrogen bond formation within 3.3 Å N-N distance and 30° N-H...N angle. It was found for 30 snapshots that Im1 and Im6 residues, which are locked into two contiguous ring subunits (i.e. Im1-Py2 and Im6-Py7), form hydrogen bonds with greater frequency (93%) than the central Im4 residue (80%). This result suggests that the central Im4, which is flanked by two β residues, is sampling more conformational space and is therefore more conformationally flexible at room temperature than Im1 and Im6.

The importance of β -alanine

We have reported a study in which the Im- β -Im and Im- β -Py subunits of a 1:1 binding polyamide were replaced with Im-Py-Im and Im-Py-Py, respectively.⁴ We found that the β to Py substitution was tolerated for Im- β -Py with a small energetic penalty. However the Im- β -Im to Im-Py-Im mutation eliminated specific binding completely. In accordance with β /Py mutation studies in the 2:1 motif, it is likely that the β is

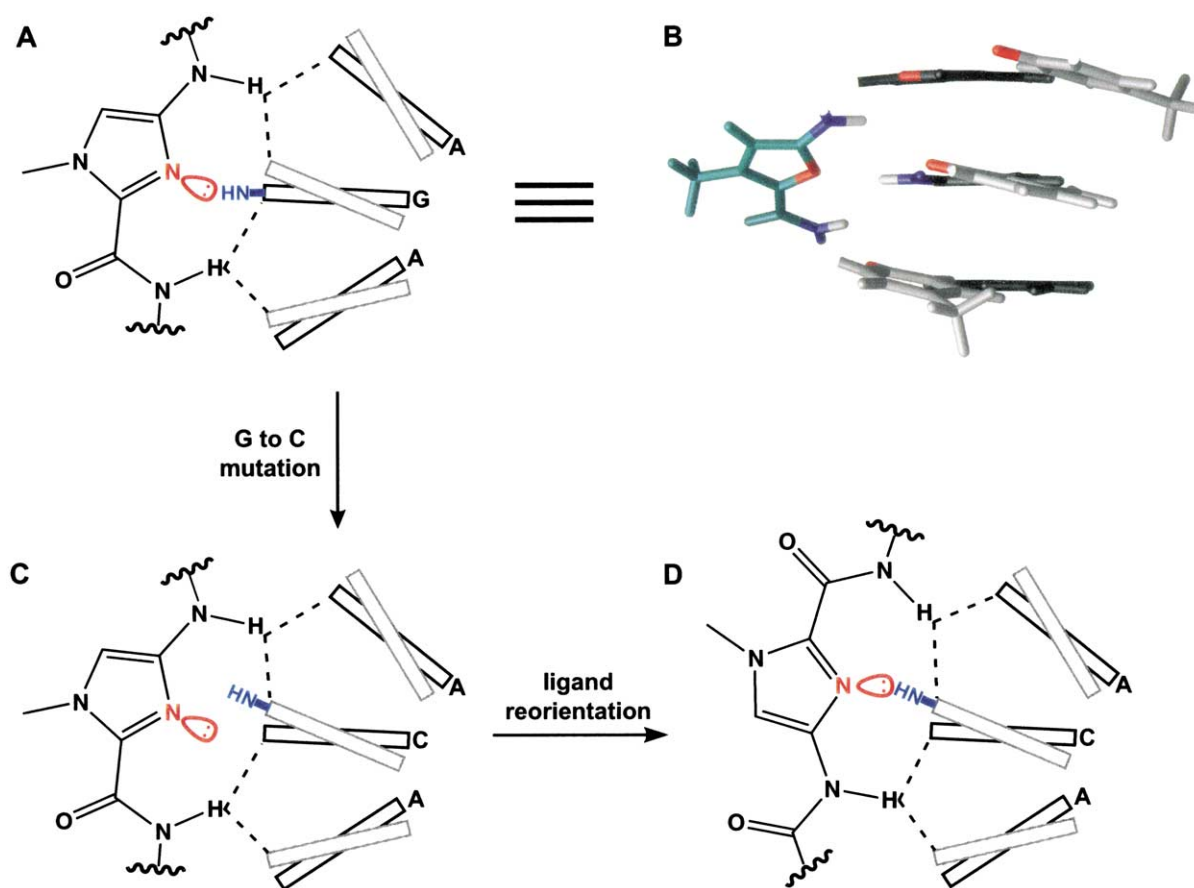


Figure 12. G/C-dependence of polyamide orientation. (a) Schematic illustrating the optimal alignment of polyamide NH groups with the DNA bases and showing that the inherent geometry of the Im residue coupled with the propeller twist of the G/C base-pair allows overlap between Im-N3 and G-NH2 groups. Propeller twisted base-pairs are shown as gray boxes (pyrimidines) crossing over black boxes (purines), with purine labels at the right. The Im-N3 sp^2 orbital is shown as its lone pair in red. Guanine's NH2 group is drawn as HN- in blue. The bases are rotated away from the center in order to project the natural 38° twist onto two dimensions. (b) Analogous section of the mean NMR structure showing Im4 next to DNA residues A7-G8-A9-T20-C21-T22. Purine bases are shown in black; pyrimidines in gray. Hydrogen bond donors and acceptors are shown in blue and red, respectively. (c) Upon mutation of G to C, the model suggests that the sp^2 orbital of Im-N3 would not overlap with G-NH2. (d) Therefore, the ligand binds in the opposite orientation in order to restore this interaction.

needed to reset the register for the following (C-terminal) residue.²¹ It appears that the β residue is necessary for the two flanking Im residues in Im- β -Im to orient properly in order to form hydrogen bonds. The Im- β -Py subunit forms only one Im-G hydrogen bond, and therefore the flexibility is not as critical. This is further supported by the ring-ring dihedral angles between the plane of Im4 and the planes of Py2 (Im- β -Py) and Im6 (Im- β -Im). Figure 11 shows axial views of these dihedral angles with average values (\pm standard deviation) for the 12 structures in the final ensemble. The Im/Py dihedral for Im- β -Py of 33° is similar to the Im-Py-Py dihedral angles (measured for bolded residues) observed in 2:1 crystal structures.^{12,13} The Im/Im dihedral angle for Im- β -Im subunit is a significant 17° larger. Although a high-resolution structure has yet to be determined for an Im-Py-Im-containing compound, we suggest that the increase in dihedral for Im- β -Im is due to the need for proper orien-

tation in order to maximize Im-N3 to G-NH2 hydrogen bonding. It is interesting that only one β -alanine conformation is observed in Im4- β 5-Im6 versus two degenerate conformers for Py2- β 3-Im4. If indeed the two β conformers are real, it is conceivable that β 5 is conformationally constrained due to the propensity of the flanking Im residues to orient properly for hydrogen bond formation and thus force a larger ring-ring dihedral angle.

Footprinting data have provided a partial recognition code for the 1:1 motif,²³ whereby Im, Py, and β will bind across from A/T base-pairs, but only Im is able to tolerate G/C. The structure presented here reveals the specific interactions underpinning the Im-G specificity and provides the first high-resolution structural model of a β residue in a 1:1 complex with DNA. Strong NOEs are observed from methylene protons in each β residue to the proximal A-H2 proton on the floor of the minor groove, as observed in the 2:1 motif.²⁰

The structural ensemble shows these protons to be in contact directly, with distances in the range of 2.3 to 2.5 Å. This would suggest a strong steric violation when β is placed across from a G-C base-pair, which helps to explain the 20 to 50-fold loss in affinity for this mismatch.²³

The Sequence-dependence of ligand orientation

Laemmli and co-workers made the remarkable observation that the β -linked polyamides in a 1:1 complex prefer a single orientation, N-C with respect to the 3'-5' direction of 5'-AAGAGA-AGAG-3', but in the opposite C-N orientation with respect to this strand when the G residues are mutated to C. In the 1:1 complex, the polyamide does not distinguish G-C and C-G base-pairs.^{22,23} On the basis of the structure, we believe that polyamide orientation is governed by a combination of the inherent geometry in the amide-Im-amide unit in combination with the negative propeller twisting of base-pairs. We assume that the uniform alignment of amide NH groups observed here is the driving force for the register between polyamide and DNA. The sp^2 lone-pair orbital on the Im-N3 atom has an inherently different orientation with respect to the flanking NH groups based on covalent bonding geometry. Due to the propeller twisting of base-pairs, the G-NH2 group is oriented more favorably for interaction with the sp^2 orbital of Im-N3 when the polyamide is oriented N-C with respect to the 3'-5' direction of the G-containing strand. Figure 12 details this model, showing a section of the NMR structure in (b) with a schematic of this orientation shown in (apart A). When the G residues are mutated to C, the lone pair of each Im-N3 will not overlap with G-NH2 (Figure 12(c)) unless the ligand orients in the opposite direction (Figure 12(d)). However, this model does not explain the observation that the polyamide will orient N-C with the 3'-5' direction of the poly(dA) tract.

Conclusions

Solution structure determination by NMR is an essential tool for studying molecular recognition phenomena. DNA structure determined by NMR is often poorly defined in the absence of a bound ligand, and in the present study it was indeed the chemical shift dispersion induced by the ligand that allowed for the virtually complete assignment of all NMR spectra, which resulted in the use of a large number of NOE-derived distance constraints to enforce the geometry of the complex. This enabled the determination at high resolution of the solution structure of a 1:1 polyamide:DNA complex comprising one full turn of the DNA helix.

The structure presented here offers a close look at a polyamide containing imidazole and β -alanine residues bound to its DNA match site in a 1:1 motif. The complex reveals B-form DNA with a

narrow minor groove and a large degree of negative propeller twist, which has been demonstrated to be stabilized by bifurcated hydrogen bonds donated from each polyamide NH group to proximal purine N3 and pyrimidine O2 atoms. Stabilization of the negative propeller twist by this interaction in addition to the inherently rigid and narrow minor groove is thought to be the reason that polyamides would bind 1:1 in A-tract-like sequences, but would have difficulty binding as 2:1.

The observed homogeneous register of amide NH groups with respect to the DNA is thought to be the driving force for optimal ligand-DNA alignment. If this is so, the previously established G/C-dependent orientation preference of the polyamide could be explained by an inherent asymmetry in the projected angle of the Im-N3 lone pair sp^2 orbital with respect to the amide NH groups. Therefore, overlap of this orbital with the propeller-twisted guanine's NH2 group is optimal when the polyamide is oriented N-C with respect to the 3'-5' direction of the guanine-containing strand.

The final structural ensemble reveals specific hydrogen bonds between Im-N3 and G-NH2. The Im to Im dihedral within the Im- β -Im subunit requires a large dihedral angle so that both rings orient properly to form hydrogen bonds. Additionally, we are now able to understand from a structural perspective the observed A/T specificity of β -alanine within the 1:1 motif, based on its close contacts with the floor of the minor groove. These results set the stage for a more critical design of next-generation polyamides for the discrimination of 1:1 *versus* 2:1 binding modes.

Experimental

Materials

Boc- β -PAM resin was purchased from Peptides International. *N,N*-Dimethylaminopropylamine, 3,3'-diamino-*N*-methyldipropylamine, ethylenediamine-tetraacetic dianhydride, triethylamine, and acetic acid were purchased from Aldrich. Acetonitrile was Omnisolv grade from EM. Dimethylsulfoxide/*N*-methylpyrrolidone (DMSO/NMP) (1:1, v/v) and diisopropylethylamine (DIEA) were from Applied Biosystems. Potassium chloride was from Fluka. RNase-free water was from US Biochemicals. *Bam* HI, *Hind* III, and phage T4 DNA ligase enzymes were from Boehringer Mannheim. pUC19 was from New England Biolabs. Centricon YM-3 dialysis filters were from Amicon. 2H_2O ("100%") was from Cambridge Isotope Laboratories. NMR tubes were type 535 from Wilmad.

Oligonucleotides were synthesized at the Biopolymer Synthesis Center at the California Institute of Technology (Caltech). Plasmid sequencing was carried out at the Sequence/Structure Analysis

Facility at Caltech. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) was performed at the Protein and Peptide Microanalytical Facility at Caltech. UV-visible spectra were obtained on a Beckman DU-7400 spectrophotometer. HPLC analysis was performed on a Beckman Gold system using a Rainin C₁₈ silica, Microsorb MV, 5 μ m, 300 mm \times 4.6 mm column. Preparatory reversed-phase HPLC was carried out on a Beckman Gold system using a Waters DeltaPak 25 mm \times 100 mm, 100 μ m C₁₈ silica column equipped with a guard. FPLC was performed on a Pharmacia LKB system. All buffers were 0.2 μ m filtered.

Polyamide synthesis

Im-Py- β -Im- β -Im-Py- β -Pam resin was prepared in ten steps from Boc- β -PAM resin as described.⁴⁸ Internal imidazole residues were incorporated as the Boc- β -Im-COOH dimer. The polyamide was cleaved from the solid support by aminolysis of the resin ester linkage using dimethylaminopropylamine (Dp) or 3,3'-diamino-N-methyldipropylamine (Dp-NH₂) to provide Im-Py- β -Im- β -Im-Py- β -Dp (**1**) or Im-Py- β -Im- β -Im-Py- β -Dp-NH₂ (**1-NH₂**), respectively, upon purification by reversed-phase preparatory HPLC in 0.1% (w/v) aqueous trifluoroacetic acid (TFA_(aq)) using a 0.33% acetonitrile/minute gradient. A sample of **1-NH₂** was treated with an excess of EDTA dianhydride (1:1:1 DMSO/NMP/DIEA, 55 $^{\circ}$ C, 15 minutes), and the remaining anhydride was hydrolysed (0.1 M NaOH, 55 $^{\circ}$ C, 10 min). The EDTA conjugate, Im-Py- β -Im- β -Im-Py- β -Dp-NH-EDTA (**1-E**), was then purified by reversed-phase preparatory HPLC in 0.1% TFA_(aq) with a 0.33% acetonitrile/minute gradient. MALDI-TOF-MS (monoisotopic): (**1**) 914.5 (914.5 calculated for C₄₁H₅₆N₁₇O₈⁺); (**1-NH₂**) 957.5 (957.5 calculated for C₄₃H₆₁N₁₈O₈⁺); (**1-E**) 1231.3 (1231.6 calculated for C₅₃H₇₅N₂₀O₁₅⁺).

Construction of plasmid DNA

The plasmid pAU20 was constructed by inserting the hybridized oligonucleotide, 5'-GATCC-GGCCAAAGAGAAGCGGGTTGGCCAAAGAG-AAGACGGGTTGGCCAAAGAGAAGAACGGGT-TGGCCAAAGAGAAGAACGGGGA-3'/5'-AGCTTCCCCTTCTTCTTGGCCAAACCCGTTCTTCTT-TTGGCCAAACCCGTTCTTCTTGGCCAAACCCG-CTTCTCTTTGGCCG-3', into the *Bam*HI/*Hind*III polycloning site in pUC19, with subsequent transformation into JM109 subcompetent cells (Promega), according to standard protocols.⁴⁹ Plasmid DNA was isolated using WizardPlus Midi Preps from Promega.

DNA radiolabeling and footprinting methods

The 5' and 3' end-labeling of plasmid pAU20 as well as the Affinity Cleavage, DNase I, and MPE

footprinting experiments were performed exactly in accordance with published protocols.³⁵ The PCR method was chosen for 5' end-labeling, employing two primer oligonucleotides, ³²P-5'-AATTCGAGCTCGGTACCCGG-3' (forward) and 5'-CTGGCACGACAGGTTTCCCG-3' (reverse), to complement the pUC19 *Eco*RI and *Pvu*II sites, respectively, such that amplification by PCR generates the 289 bp, 3'-filled *Eco*RI/*Pvu*II restriction fragment.

NMR sample preparation

DNA purification

The dimethoxytrityl-protected complementary oligonucleotides 5'-CCAAAGAGAAGCG-3' (purine strand) and 5'-CGCTTCTCTTGG-3' (pyrimidine strand) were synthesized at the Biopolymer Synthesis Center at Caltech. Each strand was purified separately by reversed phase preparatory HPLC in 100 mM triethylammonium acetate (TEAA), pH 6.5, with a gradient of 1% acetonitrile/minute. The appropriate fractions were combined, lyophilized to dryness, and resuspended in RNase-free water. The pH was lowered to 5.0 with acetic acid to remove the DMT group, and complete deprotection was observed by analytical HPLC after 12 hours. The deprotected single-stranded oligonucleotides were quantified by measuring their UV absorbance at 260 nm using the calculated extinction coefficients, $\epsilon = 137,300 \text{ M}^{-1} \text{ cm}^{-1}$ (purine strand) and $\epsilon = 107,200 \text{ M}^{-1} \text{ cm}^{-1}$ (pyrimidine strand): 7.0 μ mol of each strand was combined, and the single strands were hybridized by heating the 4.9 ml sample at 90 $^{\circ}$ C for ten minutes, with slow cooling to room temperature over six hours. Duplex DNA was separated from a slight excess of single-stranded DNA by FPLC purification using a HiTrap Q strong anion-exchange column (Pharmacia) in 200 mM KCl with a 1% per minute gradient of 1.5 M KCl. The appropriate fractions were combined and concentrated using a Centricon YM-3, 3000 Da cutoff filter from Amicon. The filter was used to dialyze the DNA sample, reducing the KCl concentration to approximately 1 μ M. The sample was lyophilized to dryness and then dissolved in 10 mM sodium phosphate (pH 7.0), quantified by measuring UV absorbance, distributed into four 2.83 μ mol aliquots, and then lyophilized to dryness. A 4.35 mM NMR sample was prepared by dissolving a 2.83 μ mol aliquot of purified duplex DNA in 650 μ l of 9:1 H₂O/²H₂O.

The 1:1 polyamide: DNA complex preparation

Polyamide concentration was estimated using the molar extinction coefficient, $\epsilon = 31,000$. A polyamide stock solution was prepared by dissolving purified polyamide **1** in 10 mM sodium phosphate (pH 7.0) in 9:1 H₂O/²H₂O. The polyamide stock was titrated into the 4.35 mM DNA

sample in very small increments in order to determine empirically a 1:1 stoichiometry by observing complete disappearance of free DNA peaks. The final complex concentration was 3.67 mM. For experiments carried out in $^2\text{H}_2\text{O}$, the complex was lyophilized twice from $^2\text{H}_2\text{O}$.

NMR experiments and data processing

NMR experiments were performed at 25 °C on a Varian INOVA 600 MHz spectrometer. The 1D spectra and 2D NOESY spectra in 9:1 $\text{H}_2\text{O}/^2\text{H}_2\text{O}$ were acquired using WATERGATE⁵⁰ water suppression containing a W5 pulse element.⁵¹ Pre-saturation of the residual solvent signal was used in the acquisition of 1D spectra and 2D double-quantum-filter (DQF)-COSY, TOCSY, and NOESY for the sample in $^2\text{H}_2\text{O}$. The TOCSY spectra were acquired with mixing times of 40 and 100 ms, and the NOESY spectra with mixing times of 50, 75, 100, and 150 ms. All the 2D spectra were acquired on the 1:1 polyamide–DNA complex. Spectral widths were 12,500 Hz for the sample in protiated solvent and 6000 Hz for the sample in deuterated solvent. All spectra were recorded with $512 t_1 \times 2048 t_2$ complex points. Data were processed on a SUN Ultra 10 workstation using VNMR (Varian, Inc.) or NMRPipe.⁵² Resonance assignment was performed using NMRVIEW 4.1.2.⁵³ Chemical shifts were referenced relative to TSP *via* the residual solvent resonance at 4.7718.

Distance constraints

In the initial round of structure calculations, resolved and unambiguously assigned cross-peak volumes from the 75 ms $^2\text{H}_2\text{O}$ NOESY spectrum were converted to upper-bound distance constraints using the isolated spin-pair approximation, plus 20% as an error approximation.⁵⁴ Once the first set of reasonable structures was obtained, a set of representative structures was used as the starting point for two rounds of hybrid relaxation matrix calculations using the MARDIGRAS program.⁵⁵ NOE cross-peak intensities from the 150 ms $^2\text{H}_2\text{O}$ NOESY spectrum were input to the MARDIGRAS program. The uncertainty in these intensities was accounted for by adjusting the upper bound distances in proportion to their magnitude as follows: <2.0 , +0.35 Å; 2–3, +0.7 Å; 3–4, +1.1 Å; 4–5, +1.5 Å; >5.0 , +1.9 Å.³² The constraints were supplemented with additional distance constraints from labile protons identified in the 100 ms NOESY spectrum in protiated solvent. All distance constraints assigned from NOEs were given a lower bound of 1.8 Å as the sum of van der Waals radii. Overlapped cross-peaks from the $^2\text{H}_2\text{O}$ and H_2O NOESY spectra were assigned upper bound distance constraints of 5.5 and 6.0 Å, respectively, on the basis of calculated distances for the smallest observable cross-peak intensities. Methylene and methyl groups were restrained as Q and M pseudoatoms, respect-

ively, as defined by MARDIGRAS and AMBER 6.0 software.^{34,45}

Watson–Crick hydrogen bonding constraints were included on the basis of characteristic cross-peaks observed at 12–14 ppm between labile protons in the 100 ms H_2O NOESY spectrum. Upper and lower bound constraints were assigned for each guanosine imino, thymidine imino, and cytosine amino group in all but the terminal base-pairs. $\text{N}\cdots\text{O}$ or $\text{N}\cdots\text{N}$ distances,⁵⁶ as well as $\text{H}\cdots\text{N}$ and $\text{H}\cdots\text{O}$ distances were enforced as follows: G-O6 to C-N4, 2.81–3.01 Å; G-N1 to C-N3, 2.85–3.05 Å; G-N1H to C-N3, 1.80–2.20 Å; G-O6 to C-N4H1, 1.76–2.16 Å; A-N1 to T-N3, 2.72–2.92 Å; A-N1 to T-N3H, 1.67–2.07 Å.

Structure calculations

Calculation strategy was based on previously reported protocols.^{32,57} Restrained molecular dynamics (rMD) calculations were performed on a Dell Dimension 8100 workstation using the AMBER 6.0 software package.³⁴ All calculations were carried out *in vacuo* with a distance-dependent dielectric and a cut-off distance of 15.0 Å for non-bonded interactions. A force constant of $20 \text{ kcal mol}^{-1} \text{ Å}^{-2}$ (1 cal = 4.184 J) was applied to interproton distances exceeding the upper and lower constraints, in a smooth parabolic fashion for 0.5 Å and then linearly.

In all, 40 starting DNA structures with variable x -displacement, incline, rise, and twist values were constructed using Nucleic Acids Builder (NAB) software.⁵⁸ While holding base-pair geometry constant, the DNA backbone was relaxed using 1000 steps of steepest-descent minimization (average pairwise all-atom RMSD 3.46 Å). The polyamide was constructed using the LEaP module of the AMBER 6.0 software package,³⁴ with partial charges assigned on the basis of an AM1 calculation using the AMPAC module of InsightII (Molecular Simulations, Inc.).

The polyamide was aligned with the DNA minor groove and positioned approximately 12 Å from the helix using NAB. Docking was performed using a 15 ps rMD simulated annealing (SA) calculation consisting of 4 ps of linear heating to 600 K, 5 ps of high-temperature dynamics, and 6 ps of linear cooling to 0 K. Further rMD SA calculations did not improve the total energy or RMSD. The family of 40 docked structures had an all-atom pairwise RMSD of 1.37 Å. The final structures were ranked by increasing residual constraint violation energy, and a calculation was performed using FINDFAM, which indicated that a minimum of 12 structures would be adequate to represent the input data accurately.⁵⁷ Accordingly, the 12 structures of lowest violation energy were chosen as the final structural ensemble (Figure 6). The core binding site of the final ensemble, which is defined as polyamide residues 1–7 (ImPy- β -Im- β -ImPy) and DNA purine tract residues A4–G11 and C16–T23, has an all-atom RMSD of 0.80 Å

and RMSD from the mean of 0.54 Å. The ensemble coordinates are available from the Brookhaven Data Bank under accession code 1LEJ.

Distances and angles in the final structural ensemble were examined using Insight II software. DNA helical parameters and groove dimensions were calculated using CURVES 5.3.³⁷ The molecular structures in Figures 6, and 8–12 were rendered in GRASP.⁵⁹

Acknowledgments

The authors thank Professor Tammy Dwyer for generously providing the ligand force-field parameters. Professor Dwyer and Professor David Case provided many helpful suggestions that led to significant improvements in the structure calculations. We are grateful to Jason Schnell for assembling the ligand in LEaP and for helpful discussions. A.R.U. thanks Professor Tom James for help with MARDIGRAS and Professor Richard Lavery for help with CURVES. This work was supported by the National Institutes of Health (GM 27681).

References

- Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P. & Dickerson, R. E. (1985). The molecular origin of DNA–drug specificity in netropsin and distamycin. *Proc. Natl Acad. Sci. USA*, **82**, 1376–1380.
- Pelton, J. G. & Wemmer, D. E. (1989). Binding modes of distamycin A with d(C G C A A A T T T G C G)₂ determined by two-dimensional NMR. *J. Am. Chem. Soc.* **112**, 1393–1399.
- Dervan, P. B. (2001). Molecular recognition of DNA by small molecules. *Bioorg. Med. Chem.* **9**, 2215–2235.
- Dervan, P. B. & Urbach, A. R. (2001). The importance of β-alanine for recognition of the minor groove of DNA. In *Essays in Contemporary Chemistry, from Molecular Structure Toward Biology* (Quinkert, G. & Kisakürek, M. V., eds), pp. 327–339, Verlag Helvetica Chimica Acta, Zurich.
- Kielkopf, C. L., Bremer, R. E., White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E. *et al.* (2000). Structural effects of DNA sequence on T·A recognition by hydroxypyrrole/pyrrole pairs in the minor groove. *J. Mol. Biol.* **295**, 557–567.
- Wartell, R. M., Larson, J. E. & Wells, R. D. (1974). Netropsin: a specific probe for A–T regions of duplex deoxyribonucleic acid. *J. Biol. Chem.* **249**, 6719–6731.
- Fratini, A. V., Kopka, M. L., Drew, H. R. & Dickerson, R. E. (1982). Reversible bending and helix geometry in a B-DNA dodecamer-CGCGAATTBrCGCG. *J. Biol. Chem.* **257**, 14686–14707.
- Patel, D. J. (1982). Antibiotic-DNA interactions: intermolecular nuclear Overhauser effects in the netropsin-d(C-G-C-G-A-A-T-T-C-T-C-G) complex in solution. *Proc. Natl Acad. Sci. USA*, **79**, 6424–6428.
- Coll, M., Frederick, C. A., Wang, A. H.-J. & Rich, A. (1987). A bifurcated hydrogen-bonded conformation in the d(A(T) base pairs of the DNA dodecamer d(CGCAAATTTGCG) and its complex with distamycin. *Proc. Natl Acad. Sci. USA*, **84**, 8385–8389.
- Dickerson, R. E. (2001). Helix structure and molecular recognition by B-DNA. In *Oxford Handbook of Nucleic Acid Structure* (Neidle, S., ed.), pp. 145–197, Oxford University Press, New York, USA.
- Hunter, C. A. (1993). Sequence-dependent DNA structure: the role of base stacking interactions. *J. Mol. Biol.* **230**, 1025–1054.
- Kielkopf, C. L., Baird, E. E., Dervan, P. B. & Rees, D. C. (1998). Structural basis for G·C recognition in the DNA minor groove. *Nature Struct. Biol.* **5**, 104–109.
- Kielkopf, C. L., White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., Dervan, P. B. & Rees, D. C. (1998). A structural basis for recognition of A·T and T·A base pairs in the minor groove of B-DNA. *Science*, **282**, 111–115.
- Chalikian, T. V., Plum, G. E., Sarvazyan, A. P. & Breslauer, K. J. (1994). Influence of drug-binding on DNA hydration-acoustic and densimetric characterization of netropsin binding to the poly(dAdT)·poly(dAdT) and poly(dA)·poly(dT) duplexes and the poly(dT)·poly(dT)·poly(dT) triplex at 25 °C. *Biochemistry*, **33**, 8629–8640.
- Lown, J. W., Krowicki, K., Bhat, U. G., Skorobogaty, A., Ward, B. & Dabroziak, J. C. (1986). Molecular recognition between oligopeptides and nucleic acids—novel imidazole-containing oligopeptides related to netropsin that exhibit altered DNA-sequence specificity. *Biochemistry*, **25**, 7408–7416.
- Wade, W. S., Mrksich, M. & Dervan, P. B. (1992). Design of peptides that bind in the minor groove of DNA at 5'-(A,T)G(A,T)C(A,T)-3' sequences by a dimeric side-by-side motif. *J. Am. Chem. Soc.* **114**, 8783–8794.
- White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E. & Dervan, P. B. (1998). Recognition of the four Watson–Crick base pairs in the DNA minor groove by synthetic ligands. *Nature*, **391**, 468–471.
- Youngquist, R. S. & Dervan, P. B. (1987). A synthetic peptide binds 16 base pairs of A,T double helical DNA. *J. Am. Chem. Soc.* **109**, 7564–7566.
- Trauger, J. W., Baird, E. E., Mrksich, M. & Dervan, P. B. (1996). Extension of sequence-specific recognition on the minor groove of DNA by pyrrole-imidazole polyamides to 9–13 base pairs. *J. Am. Chem. Soc.* **118**, 6160–6166.
- de Clairac, R. P. L., Seel, C. L., Geierstanger, B. H., Mrksich, M., Baird, E. E., Dervan, P. B. & Wemmer, D. E. (1999). NMR characterization of the aliphatic β/β pairing for recognition of A·T/T·A base pairs in the minor groove of DNA. *J. Am. Chem. Soc.* **121**, 2956–2964.
- Turner, J. M., Swalley, S. E., Baird, E. E. & Dervan, P. B. (1998). Aliphatic/aromatic amino acid pairings for polyamide recognition in the minor groove of DNA. *J. Am. Chem. Soc.* **120**, 6219–6226.
- Janssen, S., Durussel, T. & Laemmli, U. K. (2000). Chromatin opening of DNA satellites by targeted sequence-specific drugs. *Mol. Cell*, **6**, 999–1011.
- Urbach, A. R. & Dervan, P. B. (2001). Toward rules for 1:1 polyamide:DNA recognition. *Proc. Natl Acad. Sci. USA*, **98**, 4343–4348.
- Taylor, J. S., Schultz, P. G. & Dervan, P. B. (1984). DNA affinity cleaving-sequence specific cleavage of DNA by distamycin-EDTA-Fe(II) and EDTA-distamycin-Fe(II). *Tetrahedron*, **40**, 457–465.
- Pelton, J. G. & Wemmer, D. E. (1988). Structural modeling of the distamycin A-d(CGCGAATTCGCG)₂ complex using 2D NMR and molecular mechanics. *Biochemistry*, **27**, 8088–8096.

26. Mrksich, M., Wade, W. S., Dwyer, T. J., Geierstanger, B. H., Wemmer, D. E. & Dervan, P. B. (1992). Anti-parallel side-by-side dimeric motif for sequence-specific recognition of the minor groove of DNA by the designed peptide 1-methylimidazole-2-carboxamide netropsin. *Proc. Natl Acad. Sci. USA*, **89**, 7586–7590.
27. Dwyer, T. J., Geierstanger, B. H., Bathini, Y., Lown, J. W. & Wemmer, D. E. (1992). Design and binding of a distamycin A analog to d(C G C A A G T T G G C)-d(G C C A A C T T G C G): synthesis, NMR studies, and implications for the design of sequence-specific minor groove binding oligopeptides. *J. Am. Chem. Soc.* **114**, 5911–5919.
28. Dwyer, T. J., Geierstanger, B. H., Mrksich, M., Dervan, P. B. & Wemmer, D. E. (1993). Structural analysis of covalent peptide dimers, bis(pyridine-2-carboxamidonetropsin)(CH₂)_{3–6}, in complex with 5'-TGACT-3' sites by two-dimensional NMR. *J. Am. Chem. Soc.* **115**, 9900–9906.
29. Geierstanger, B. H., Mrksich, M., Dervan, P. B. & Wemmer, D. E. (1996). Extending the recognition site of designed minor groove binding molecules. *Nat. Struct. Biol.* **3**, 321–324.
30. Wütrich, K. (1986). *NMR of proteins and nucleic acids*, Wiley, New York, USA.
31. Nadeau, J. G. & Crothers, D. M. (1989). Structural basis for DNA bending. *Proc. Natl Acad. Sci. USA*, **86**, 2622–2626.
32. Eis, P. S., Smith, J. A., Rydzewski, J. M., Case, D. A. & Boger, D. L. (1997). High resolution solution structure of a DNA duplex alkylated by the antitumor agent duocarmycin SA. *J. Mol. Biol.* **272**, 237–252.
33. Schnell, J. R., Ketchum, R. R., Boger, D. L. & Chazin, W. J. (1999). Binding-induced activation of DNA alkylation by duocarmycin SA: insight from the structure of an indole derivative-DNA adduct. *J. Am. Chem. Soc.* **121**, 5645–5652.
34. Kollman, P. A., Case, D. A., Merz, K., Cheatham, T. E. III, Simmerling, C., Darden, T., *et al.* (2000). *AMBER 6.0*, University of California, San Francisco, CA.
35. Trauger, J. W. & Dervan, P. B. (2001). Footprinting methods for analysis of pyrrole-imidazole polyamide/DNA complexes. *Methods Enzymol.* **340**, 450–466.
36. Celda, B., Widmer, H., Leupin, W., Chazin, W. J., Denny, W. A. & Wütrich, K. (1989). Conformational studies of d(AAAAATTTT)₂ using constraints from nuclear Overhauser effects and from quantitative analysis of the cross-peak fine structures in two-dimensional ¹H nuclear magnetic resonance spectra. *Biochemistry*, **28**, 1462–1471.
37. Lavery, R. & Sklenar, H. (1988). The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dyn.* **6**, 63–91.
38. Stofer, E. & Lavery, R. (1994). Measuring the geometry of DNA grooves. *Biopolymers*, **34**, 337–346.
39. Crothers, D. M. & Shakked, Z. (2001). DNA bending by adenine–thymine tracts. In *Oxford Handbook of Nucleic Acid Structure* (Neidle, S., ed.), pp. 455–470, Oxford University Press, New York.
40. MacDonald, D., Herbert, K., Zhang, X., Polgruto, T. & Lu, P. (2001). Solution structure of an A-Tract DNA bend. *J. Mol. Biol.* **306**, 1081–1098.
41. Han, G. W., Kopka, M. L., Cascio, D., Grzeskowiak, K. & Dickerson, R. E. (1997). Structure of a DNA analog of the primer for HIV-1 RT second strand synthesis. *J. Mol. Biol.* **269**, 811–826.
42. Shatzky-Schwartz, M., Arbuckle, N. D., Eisenstein, M., Rabinovich, D., Bareket-Samish, A., Haran, T. E. *et al.* (1997). X-ray and solution studies of DNA oligomers and implications for the structural basis of A-tract-dependent curvature. *J. Mol. Biol.* **267**, 595–623.
43. Bostock-Smith, C. E., Harris, S. A., Laughton, C. A. & Searle, M. S. (2001). Induced fit DNA recognition by a minor groove binding analogue of Hoechst 33258: fluctuations in DNA A tract structure investigated by NMR and molecular dynamics simulations. *Nucl. Acids. Res.* **29**, 693–702.
44. Dickerson, R. E. & Drew, H. R. (1981). Structure of a B-DNA dodecamer II. Influence of base sequence on helix structure. *J. Mol. Biol.* **149**, 761–786.
45. Goodsell, D. S., Ng, H. L., Kopka, M. L., Lown, J. W. & Dickerson, R. E. (1995). Structure of a dicationic monoimidazole lexitropsin bound to DNA. *Biochemistry*, **34**, 16654–16661.
46. Kopka, M. L., Goodsell, D. S., Han, G. W., Chiu, T. K., Lown, J. W. & Dickerson, R. E. (1997). Defining GC-specificity in the minor groove: side-by-side binding of the di-imidazole lexitropsin to C-A-T-G-G-C-C-A-T-G. *Structure*, **5**, 1033–1046.
47. Geierstanger, B. H., Mrksich, M., Dervan, P. B. & Wemmer, D. E. (1994). Design of a G-C-specific DNA minor groove-binding peptide. *Science*, **266**, 646–650.
48. Baird, E. E. & Dervan, P. B. (1996). Solid phase synthesis of polyamides containing imidazole and pyrrole amino acids. *J. Am. Chem. Soc.* **118**, 6141–6146.
49. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
50. Piotto, M., Saudek, V. & Sklenar, V. (1992). Gradient-tailored excitation for single-quantum NMR-spectroscopy of aqueous-solutions. *J. Biomol. NMR*, **2**, 661–665.
51. Liu, M. L., Mao, X. A., Ye, C. H., Huang, H., Nicholson, J. K. & Lindon, J. C. (1998). Improved WATERGATE pulse sequences for solvent suppression in NMR spectroscopy. *J. Magn. Reson.* **132**, 125–129.
52. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPIPE—a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.
53. Johnson, B. A. & Blevins, R. A. (1994). NMR VIEW—a computer-program for the visualization and analysis of NMR data. *J. Biomol. NMR*, **4**, 603–614.
54. Barsukov, I. L. & Lian, L.-Y. (1993). Structure determination from NMR data I. Analysis of NMR data. In *NMR of Macromolecules* (Roberts, G. C. K., ed.), pp. 315–357.
55. Borgias, B. A. & James, T. L. (1990). MARDIGRAS—a procedure for matrix analysis of relaxation for discerning geometry of an aqueous structure. *J. Magn. Reson.* **87**, 475–487.
56. Schmitz, U., Sethson, I., Egan, W. M. & James, T. L. (1997). Solution structure of a DNA octamer containing the Pribnow box *via* restrained molecular dynamics simulation with distance and torsion angle constraints derived from two-dimensional nuclear magnetic resonance spectral fitting. *J. Mol. Biol.* **227**, 510–531.
57. Smith, J. A., Gomez-Paloma, L. G., Case, D. A. & Chazin, W. J. (1996). Molecular dynamics docking driven by NMR-derived restraints to determine the

- structure of the calicheamicin gamma (I) (I) oligosaccharide domain complexed to duplex DNA. *Magn. Reson. Chem.* **34S**, 147–155.
58. Macke, T. (1996). NAB: A language for molecular manipulation. PhD Thesis, The Scripps Research Institute
59. Nicholls, A., Sharp, K. & Honig, B. (1991). GRASP—graphical representation and analysis of structural properties. *Proteins: Struct. Funct. Genet.* **11**, 281–296.

Edited by P. Wright

(Received 7 February 2002; received in revised form 26 April 2002; accepted 29 April 2002)



<http://www.academicpress.com/jmb>

Supplementary Material comprising two Tables of chemical shift data for the 1:1 complex is available on IDEAL.