1:1 MOTIF FOR DNA RECOGNITION BY β-ALANINE-LINKED POLYAMIDES

Thesis by

Adam Robert Urbach

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2002

(Submitted May 17, 2002)

To my Father

Acknowledgements

I would like to thank my Ph. D. advisor, Professor Peter Dervan, for giving me just the right mixture of freedom and guidance in order to pursue my interests and yet to turn my energy into useful science. The training I've received has far exceeded anything I could have imagined just a few years ago. I am particularly grateful to Professor Dervan for supporting my decision to leave Caltech during my second year and then taking me back into his group with enthusiasm and compassion. My thesis committee has taught me, among other things, the important lesson of how to better approach the writing and defending of research proposals. I am grateful to Professors Robert H. Grubbs, John D. Roberts, Richard W. Roberts, and Stephen L. Mayo for these lessons as well as helpful conversations over the years. Special thanks goes to Professor John D. Roberts for the privilege of his discerning attention during this past year.

The work of Professor Uli Laemmli provided much inspiration for the theses presented here. I look to Professor Laemmli with great admiration and respect as a scientist, and I am grateful to him for advocating my work during his visit to Caltech. Much appreciation goes to Professor John Love and Dr Scott Ross for teaching me macromolecular NMR and being supportive throughout our fruitful collaboration. Professors David Case and Tammy Dwyer provided many helpful suggestions that led to significant improvements in the structure calculations, and their generous input is most appreciated. During my last year at Caltech, I had the pleasure of collaborating with Michael "Meaty" Marques and Ray Doss on several interesting projects. I am very grateful to Michael and Ray for helping me finish research while I was writing props.

I wish to acknowledge an extraordinary group of colleagues in the Dervan lab who have managed somehow to put up with me over the years. David Herman is the embodiment of positive vibe, infecting all who know him. David's friendship has helped to keep me sane during many long nights of James Brown in the computer room–- adventures to the Goodland and afternoons with Abe will never be forgotten, not to mention learning to surf in overhead conditions. Thanks to John Trauger for showing me that ego is not a necessary component of ambition or success. I am grateful to John for his friendship and for generously sharing his apartment when I needed a place to live. Much appreciation goes to Ben Edelson for many stimulating discussions and for proofreading numerous works of literature. Ben's integrity and humility are an exceptional example to follow. Thanks to Will Greenberg and Dave Liberles for being excellent labmates and for many good times with Gamesa. I thank Ken Brameld for getting me interested in Himalayan trekking. Thanks to Clay Wang for great dinner parties and to Meredith Howard for lively conversations. Victor Rucker has made life in Church 308 memorable, and I thank him for being himself. I thank Shane Foister and Ramez Elgammal for much hospitality, sometimes southern. Roland Bürli and Ralf Jäger provided some much needed help with my candidacy props. I thank Aileen Chang, Jason Belitsky, Adam Kerstien, Amanda Cashin, Eric Fechter, Philipp Weyermann, Christoph Briehn, Leonard Prins, Anna Mapp, Paul Floreancig, Tom Minehan, Christian Melander, Bogdan Olenyuk, and (last but certainly not least) Bobby Arora for making the Dervan group a more pleasant place to live.

The staff at Caltech is very accommodating, and I would especially like to thank Dian Buchness, Chris Smith, Margot Hoyt, Lynne Martinez, Lindy Alo, Tom Dunn, and Steve Gould for their exceptional help over the years. I thank Darryl Willick for saving me from computer disaster on several occasions.

Along the path that led to Caltech, I thank David Arnold for taking in a problem child; Professor Jonathan Sessler for introducing me to organic chemistry and chemical research; and Faiz Kayyem and Cindy Bamdad for helping me find the path back to grad school.

Life is meaningless without friends and family. The enormous support provided by this special group has made my life complete. Thanks to Mimi, Papa, Mom, Jim, Rene, Joe, Lucy, Siggy, and Susan for their support and understanding. I thank Bryn, Allison, Dian, Mark, Kutty, Denise, Ashley, and the Herman family for making me part of their family. I am grateful to Dr Art Herman for solid advice when I needed it. I have had the privilege of getting to know Professor Michael Waring during my time at Caltech. Michael's support during my candidacy made a world of difference, and I will always be grateful. I thank Professor David Laude for his friendship and lessons outside the box. I am grateful to Sifu Ken Edwards for teaching me Tai Chi.

There are three people whom I would especially like to acknowledge for their endless support and friendship. Rudy Emmelot has been an outstanding friend to me these past few years. Rudy's generosity is sometimes overwhelming, and I cannot in these short words express to him the full extent of my gratitude for his companionship both at home and traveling across the globe. I could only be half the person I am without the love and support of my wonderful wife, Dana, whose support and devotion mean everything to me. I look forward most to our lives together. To my father, to whom this thesis is dedicated, goes my deepest respect and appreciation for his friendship, support, and countless self-sacrifices from the very beginning.

Abstract

Polyamides composed of N-methylpyrrole (Py), N-methylimidazole (Im), and 3-hydroxypyrrole (Hp) amino acids linked by beta-alanine (β) bind in the minor groove of DNA in 1:1 and 2:1 ligand:DNA complexes. Although the energetics and structure of the 2:1 motif have been explored extensively, there is remarkably less understood about 1:1 recognition beyond the initial studies on netropsin and distamycin. Laemmli and coworkers used β-linked polyamides, which bind in a 1:1 motif, to effect phenotypic changes in *Drosophila melanogaster*. The thesis work described here investigates Laemmli's 1:1 motif in order to further understand and exploit this novel mode of DNA recognition.

By selectively replacing Py residues with β it was found that the Im-β-Im subunit is important for high-affinity binding in 1:1 and 2:1 modes. This study also demonstrates that a single ligand can target very different DNA sequences based on 1:1 or 2:1 binding. This ambiguity of sequence targeting based on stoichiometry was addressed. It was discovered that hairpin and 1:1 binding modes, which are dependent on ligand conformation, are controlled by changing the linker between polyamide subunits.

The possibility of developing a 1:1 recognition code was explored by selectively mutating polyamide residues and DNA base pairs and comparing the association constants for the resulting complexes. It was found that Im residues tolerate all four Watson-Crick base pairs; Py and β residues are specific for A•T and T•A base pairs; and Hp specifies a single base pair, A•T, in the sequence context 5'-AAAGAGAAGAG-3'. Attempts to improve upon this recognition code using novel heterocyclic amino acids, such as furan, thiophene, thiazole,

and hydroxythiophene, are presented. The sequence-dependence of ligand orientation and the effect of ligand size on binding affinity were also explored.

The NMR structure of a 1:1 polyamide:DNA complex was determined. It reveals B-form DNA with a narrow minor groove and large negative propeller twist, which are shown to be stabilized by bifurcated hydrogen bonds between polyamide NH groups and purine N3 and pyrimidine O2 atoms. The first direct evidence is provided for hydrogen bond formation between Im-N3 and guanine NH2 in the 1:1 motif, thus confirming the original lexitropsin model.

Table of Contents

page

List of Figures and Tables

INTRODUCTION page

RESULTS AND DISCUSSION

Introduction

Biochemical instructions for life on Earth are stored in each organism's deoxyribonucleic acid (DNA). The massive international effort to sequence the human genome underscores the importance of the information contained in the roughly 30,000 genes, each of which encodes different protein or ribonucleic acid (RNA) products (Venter et al., 2001). Proper cellular function depends on specific protein-DNA interactions necessary for regulating gene expression, and it is the misregulation of gene expression that is responsible for many disease states, including certain cancers. Synthetic molecules that bind to predetermined DNA sequences and regulate gene expression would therefore offer great benefit to human medicine (Gottesfeld et al., 1997; Dickinson et al., 1998; Mapp et al., 2000).

DNA Structure. Double-helical DNA is composed of two polydeoxy-ribonucleotide strands aligned in an antiparallel fashion and associated through specific hydrogen bonds between the heterocyclic bases, adenine (A), thymine (T), guanine (G), and cytosine (C), such that A pairs with T and G with C (Figure 1) (Saenger, 1984; Neidle, 1999). Common B-form DNA is characterized by a wide and shallow major groove and a narrow and deep minor

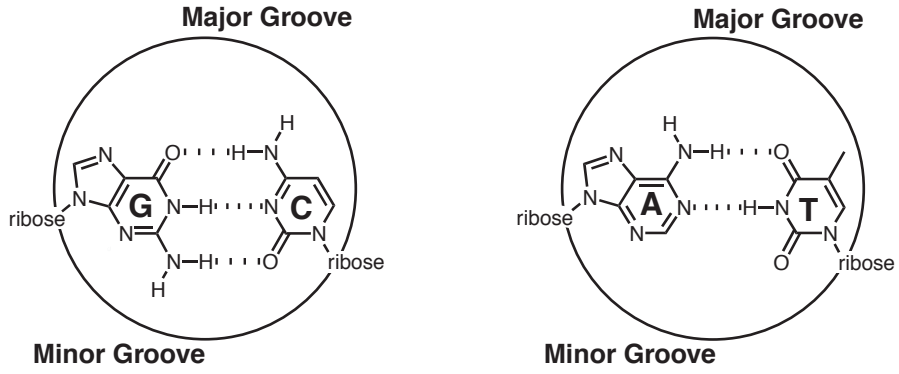


Figure 1 Chemical structure of G•C (left) and A•T (right) base pairs. The major and minor grooves are indicated as the regions spanning between the ribose backbones.

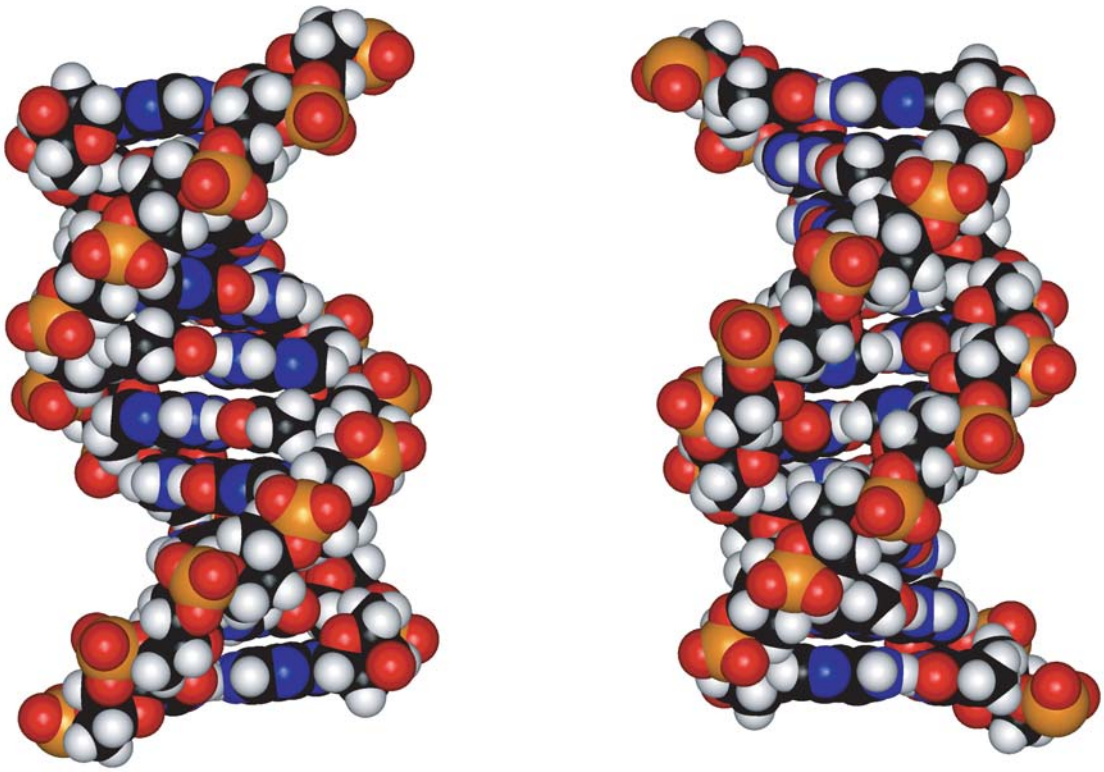


Figure 2 Space-filling model of a ten base-pair B-form DNA duplex viewed into (left) the major groove and (right) the minor groove. Carbon atoms are black, oxygens are red, nitrogens are blue, hydrogens are white, and phosphorus atoms are orange.

groove (Figure 2). DNA sequences can be distinguished by the pattern of functional groups, e.g., hydrogen bond donors and acceptors, displayed on the edges of the base pairs. However, the sequence-dependent variation in conformation and counterion organization that distinguishes local DNA microstructure (Saenger, 1984) makes it difficult to design molecules with optimal shape and electrostatic complementarity to a particular DNA sequence.

Native DNA Recognition. Nature has selected for numerous DNA-binding proteins capable of specific sequence recognition based on ensembles of electrostatic and shape-selective interactions. Conversely, the enormous diversity in protein structure makes *de novo* protein design for the recognition of specific DNA sequences a challenging prospect. Although numerous structural motifs have been identified for protein-DNA recognition (Figure 3) (Love et al., 1995; Kim et al., 1993; Pavletich and Pabo, 1991; Ellenberger et al., 1992), a general recognition code correlating target DNA sequence with amino acid sequence composition has yet to be identified. On the other hand, small molecules are typically more limited in conformational flexibility than proteins, offering the chemist a more controllable platform for fine-tuning the shape complementarity necessary for DNA sequence discrimination. Nature has also provided a number of structurally diverse small molecules that recognize DNA by binding in the minor groove, intercalating between base pairs, or both (Figure 4) (Gao et al., 1992; Kamitori and Takusagawa, 1992; Paloma et al., 1994; Coll et al., 1987). The N-methylpyrrole carboxamide (Py) backbone of the antitumor antibiotics netropsin and distamycin A presents an attractive context for the design of polyamide ligands with altered DNA sequence selectivity.

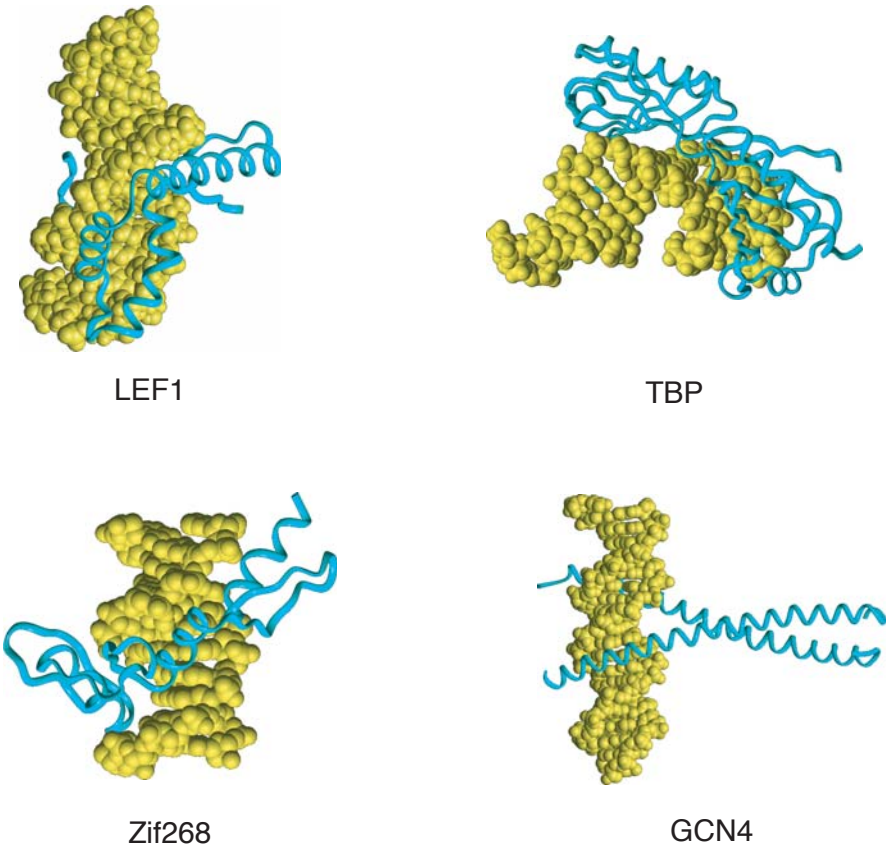


Figure 3 X-ray crystal structures of four protein-DNA complexes showing the diversity of structural motifs for protein-DNA recognition.

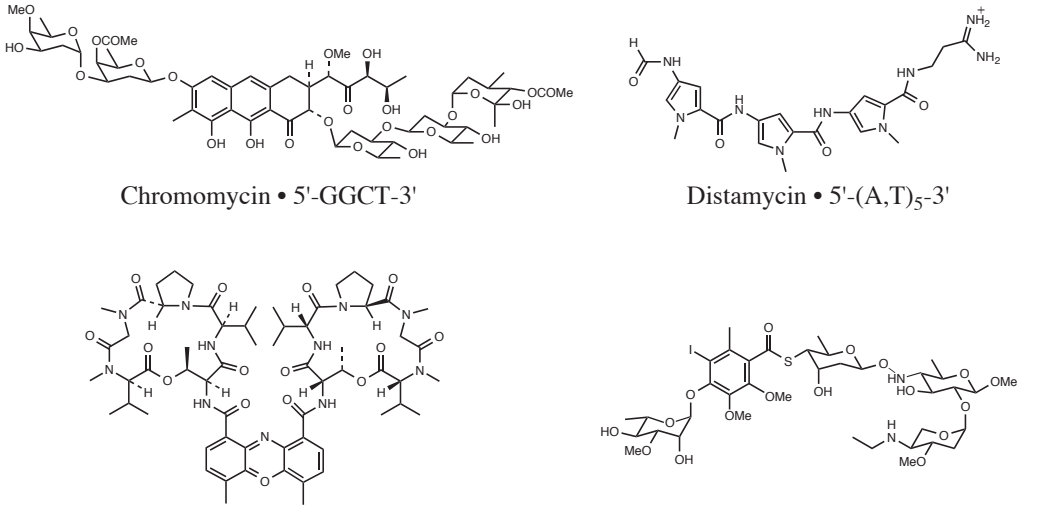


Figure 4 Chemical structures of several naturally occurring, DNA-binding small molecules with their optimal target DNA sequences indicated.

Minor Groove Recognition by Designed Ligands. The minor groove of DNA can be characterized by a somewhat uniform display of chemical functionality. A•T base pairs present relatively symmetric hydrogen bond acceptors, N3 of A and O2 of T. G•C base pairs present similar groups, N3 of G and O2 of C, in addition to the hydrogen bond-donating 2-amino group of guanine (G-NH_2) (Figure 5). The minor groove of A,T-tracts is both narrow due to propeller twisting of the base pairs and relatively deep due to lack of the protruding G-NH_2 (Fratini et al., 1982).

Understanding the sequence-dependent microstructure of DNA is of key importance for the study 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 (Dickerson, 2001). In particular, purine-purine steps such as A-A and G-A are inclined to be more 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 (Hunter, 1993).

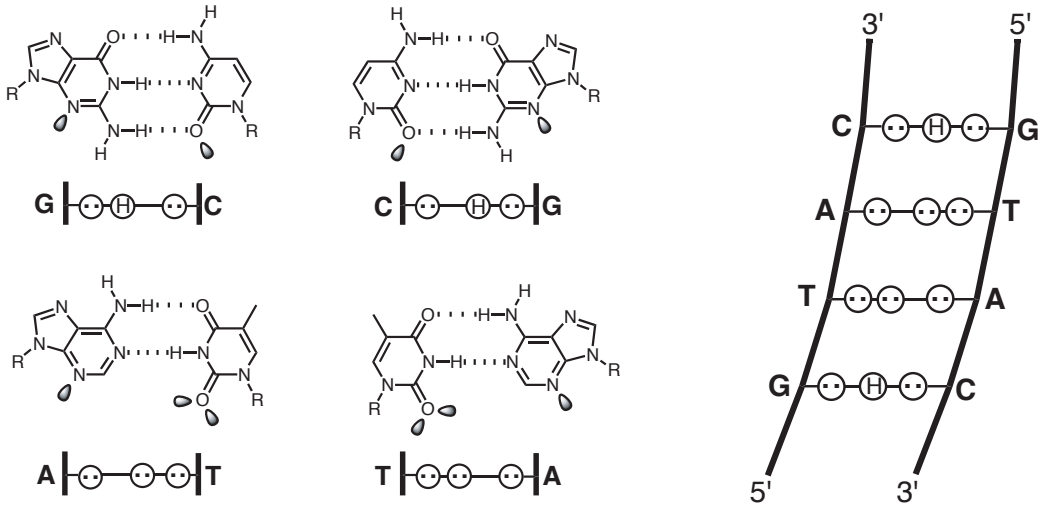


Figure 5 Minor groove hydrogen bonding patterns of the four Watson-Crick base pairs. Circles with dots represent lone pairs on purine N3 or pyrimidine O2, and circles with an H represent the 2-amino group of guanine (G-NH_2).

Distamycin and netropsin bind in the minor groove of A,T-tract DNA (Zimmer and Wahnert, 1986). Upon binding, these ligands displace the spine of hydration in a multidentate fashion, forming hydrogen bonds to proximal purine N3 and pyrimidine O2 atoms (Kopka et al., 1985; Coll et al., 1987), which provides a large entropic driving force for binding (Chalikian et al., 1994). Moreover, the bound ligands fit snugly, making extensive van der Waals contacts to the walls of the minor groove. The x-ray structure of netropsin bound as a 1:1 complex to DNA (Figure 6, left) 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 G-NH_2 to Im-N3 (Kopka et al., 1985). Subsequent footprinting experiments revealed that Im-Py polyamides tolerate G•C base pairs but show little sequence-specificity (Lown et al., 1986). Remarkably, the structural basis for the lexitropsin model, as envisioned in a 1:1 complex, was never verified structurally.

Wemmer and coworkers made the unanticipated observation that distamycin can bind A•T tracts of DNA in an antiparallel 2:1 fashion (Figure 6, right), even at low ligand:DNA stoichiometries (Pelton and Wemmer, 1989). Therefore, Py/Py pairs, as well as Py, prefer A,T over G,C. It was subsequently discovered by Dervan and coworkers at Caltech that the unsymmetrical ring pair Im/Py can distinguish G•C from C•G and both from A•T and T•A base pairs (Dervan, 2001). Further invention of the new ring pair Hp/Py (Hp = 3-hydroxypyrrole, Hp/Py specifies T•A) completed a recognition code to target all four Watson-Crick base pairs in the minor groove of DNA (Figure 7).

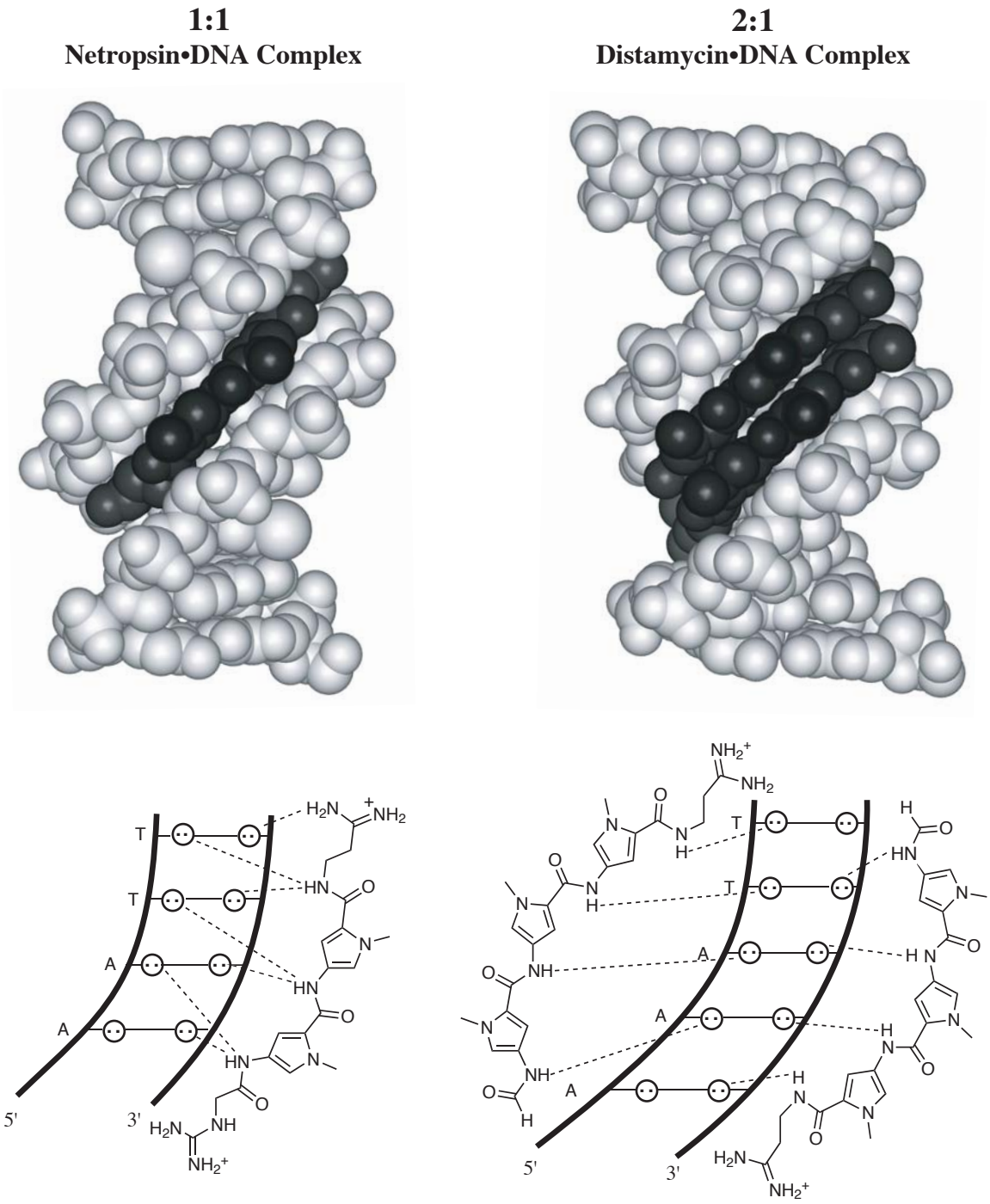


Figure 6 (left) X-ray crystal structure of the 1:1 netropsin•DNA complex; (right) NMR structure of the 2:1 distamycin•DNA complex. Binding models of each complex are shown below with hydrogen bonds indicated as dashed lines.

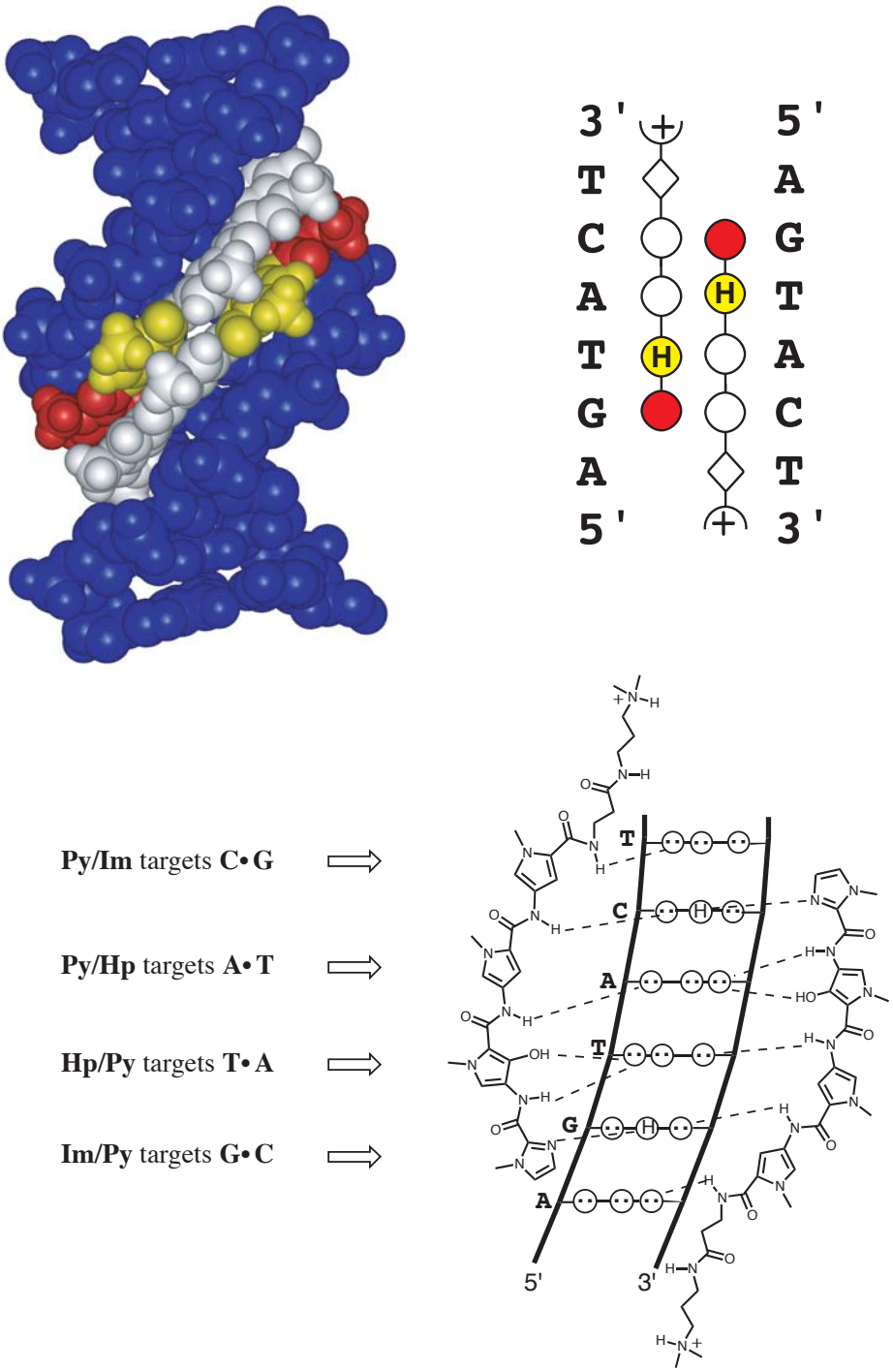


Figure 7 Pairing Rules. (top left) x-ray crystal structure of ImHpPyPy-β-Dp (Dp = dimethylaminopropylamide) bound in a 2:1 complex with its target DNA site, 5'-AGTACT-3' (Kielkopf et al., 1998b). A schematic is shown to the right, which represents Im residues as red circles, Hp as yellow circles containing an H, Py as white circles, and β as white diamonds. The hydrogen bonding schematic at the bottom shows the pairing rules for targeting all four Watson-Crick base pairs of DNA.

Antiparallel polyamide subunits stacked in a 2:1 complex can be covalently linked, head-to-tail, to form hairpin polyamides with substantially increased binding affinity and sequence specificity (Mrksich et al., 1994). Polyamides composed of multiple contiguous Py and Im residues are overcurved with respect to the DNA helix (Kelly et al., 1996). However, Py residues can be substituted with the flexible, A,T-specific beta-alanine residue (β), in order to relax ligand curvature and restore ligand-DNA alignment, thereby restoring binding affinity (Trauger et al., 1996a; Turner et al., 1998; de Clairac et al., 1999) (Figure 8). An important example of affinity restoration via β substitution is the hairpin polyamide ImPyImPy-γ-ImPyImPy-β-Dp (Dp = dimethylaminopropylamide), which targets the core sequence 5′-GCGC-3′, according to the pairing rules, but with low affinity (Swalley et al., 1997). By judicious replacement of Py with β-alanine, i.e., Im-Py-Im → Im-β-Im, a second generation hairpin polyamide, Im-β-ImPy-γ-Im-β-ImPy-β-Dp, restores the dissociation constant (K_D) to subnanomolar (Figure 8) (Turner et al., 1998).

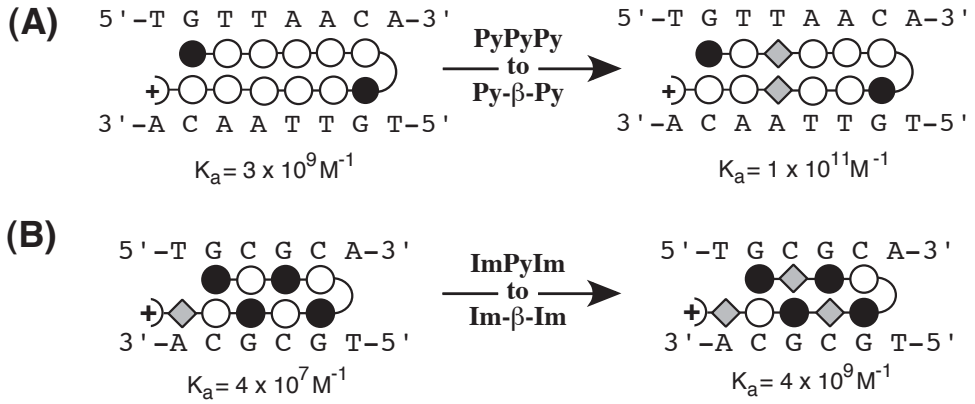


Figure 8 Restoration of binding affinity upon replacement of Py with β. Im and Py residues are illustrated as shaded and nonshaded circles, respectively; gray diamonds indicate β; and semicircles represent the γ-turn residue.

Limitations. Sequence specificity in a genomic context will require high fidelity targeting of long DNA sequences (16-20 base pairs). Employing the current state-of-the-art hairpin polyamides may be problematic because hairpin molecular weight increases dramatically with binding site size. This increase can be deleterious to cellular and nuclear uptake, which are essential for therapeutic applications. Therefore, the search for alternative binding modes remains an important direction. The "pairing rules" have proven useful for recognition of hundreds of predetermined DNA sequences by designed polyamides. However, sequence-dependent structural variation is thought to reduce binding affinity at numerous DNA sequences, most notably those containing 5'-GA-3' and 5'-GNG-3' steps (Swalley et al., 1997; Herman, 2001). The development of higher fidelity recognition rules, which will require a better understanding of sequence-dependent structural effects, is an important goal**.**

The 1:1 Motif. Although the structure and energetics of the 2:1 motif have been explored extensively (Dervan, 2001; Kielkopf et al., 2000), relatively little was known (prior to the thesis presented here) about the 1:1 motif beyond the initial work on netropsin and distamycin. In a recent breakthrough, Laemmli and coworkers reported the use of Py-Im polyamides to effect phenotypic changes in *Drosophila melanogaster* (Janssen et al., 2000a; 2000b). In their report, 5'-GAGAA-3' repeat sequences were targeted at very high affinity using β-rich polyamides that bind in a 1:1 fashion and in a sequence-dependent, single orientation (Figure 9). This new mode of 1:1 recognition offers the opportunity to target longer DNA sequences with considerably smaller molecules, due to the requirement for only a single subunit per binding site, which may further confer the benefit of increased cellular uptake. Additionally, the 5'-GAGAA-3' sequence

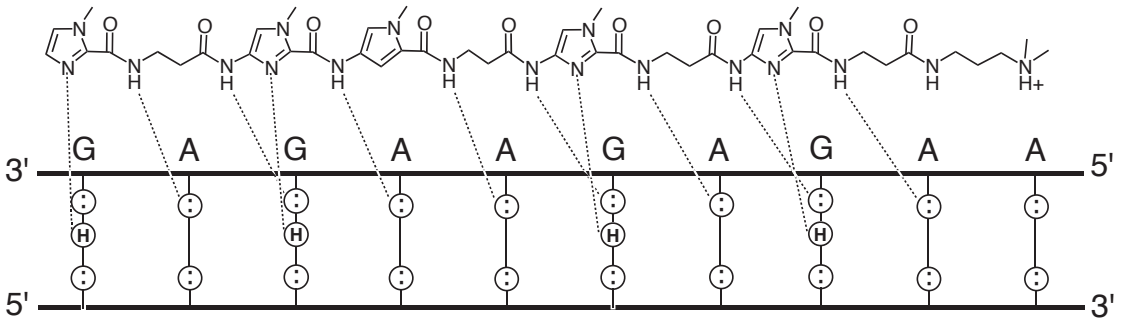


Figure 9 Laemmli model for 1:1 polyamide-DNA recognition using a β-linked Py-Im polyamide. Putative hydrogen bonds are indicated by dashed lines.

is almost entirely composed of 5'-GA-3' units, which are problematic for hairpin polyamides. Therefore, this class of molecules holds the potential to address numerous problems in DNA recognition and, hence, merits further exploration.

Description of this Work. The thesis presented here investigates Laemmli's 1:1 motif in order to further understand and exploit this novel mode of DNA recognition. First, the importance of β-alanine for 1:1 recognition was examined. In this study, Py residues within Im**Py**Py and Im**Py**Im contexts were replaced with β, and it was found using quantitative DNase I footprinting that the Im-β-Im subunit is important for high-affinity binding. This study also demonstrates the capability of a single ligand to bind very different DNA sequences depending on the stoichiometry of complexation, i.e., 1:1 or 2:1. Since the 1:1 and 2:1 motifs have inherently different rules for recognition, this finding poses the design problem of how to control the binding mode and therefore the DNA sequence target (Dervan and Urbach, 2001). Next, the possibility of developing a 1:1 recognition code was explored. By selectively mutating

polyamide residues and DNA base pairs, and comparing the association constants for the resulting complexes, it was found that Im residues tolerate all four Watson-Crick base pairs; Py and β residues are specific for A•T and T•A base pairs; and Hp specifies a single base pair, A•T, in the sequence context 5'-AAAGAGAAGAG-3' (Urbach and Dervan, 2001). Efforts to improve upon this recognition code using novel heterocyclic amino acids, such as furan, thiophene, thiazole, and hydroxythiophene, are presented (Marques et al., in preparation). Additionally, the sequence-dependence of ligand orientation and the effect of ligand size on binding affinity were explored. The intellectual core of this thesis is supported by the determination and analysis of the solution structure of a 1:1 polyamide-DNA complex using two-dimensional NMR methods. The high-resolution structure reveals B-form DNA with a narrow minor groove and large negative propeller twist, which are shown to be stabilized by bifurcated hydrogen bonds between polyamide NH groups and purine N3 and pyrimidine O2 atoms at each base step. The first direct evidence is provided for hydrogen bond formation between Im-N3 and G-NH_2 in the 1:1 motif, thus confirming the original lexitropsin model (Urbach et al., 2002). Finally, the ambiguity of sequence targeting depending on stoichiometry was addressed. It was discovered that hairpin and 1:1 binding modes, which are dependent on ligand conformation, can be effectively controlled by changing the linker between polyamide subunits. In the system examined, a β linkage specifies the 1:1 mode by >150-fold. Replacement of β with α-(*R*)-acetamido-γ-aminobutyric acid enforces an 82,000-fold reversal of specificity in preference for the hairpin motif (Urbach et al., *in preparation*).

Results and Discussion

The Importance of β-Alanine for Recognition of the Minor Groove of DNA[1]

Purpose. The recent finding by Laemmli and coworkers that the β-linked polyamide Im-β-ImPy-β-Im-β-Im-β-Dp binds GAGAA tracks in a 1:1 stoichiometry and in a single orientation raises some interesting questions regarding the nature of this binding mode (Janssen et al., 2000). The favorable Im-β-Im composition (Turner et al., 1998) is present in their polyamide, leading us to ponder its importance and generality. On one hand, the 1:1 complex is an important observation by Laemmli, which could expand the sequence repertoire for DNA targeting. On the other hand, the fact that β-linked Py/Im polyamides can bind *both* 1:1 and 2:1 in the minor groove raises the issue that a single polyamide molecule may bind very different sequences depending on stoichiometry (Figure 10).

Approach. To explore the effects of β linkage at different polyamide positions, we synthesized three polyamides: Im**Py**ImPy-β-Im**Py**ImPy-β-Dp (**1**) and Im-**β**-ImPy-β-Im-**β**-ImPy-β-Dp (**2**), which vary in their Im-**X**-Im composition (X= Py or β); and Im-β-ImPy**Py**Im-β-ImPy-β-Dp (**3**), wherein an internal β residue is changed to Py (Figure 11). To ask if a single ligand can target different sequence contexts depending on stoichiometry, footprinting and affinity cleavage experiments were performed for **1**-**3** at designed 2:1 and 1:1 sites, 5′-AGCGCAGCGCT-3′ and 5′-AAGAGAAGAG-3′, respectively.

---

[1] The text of this section is taken from Dervan and Urbach, 2001.

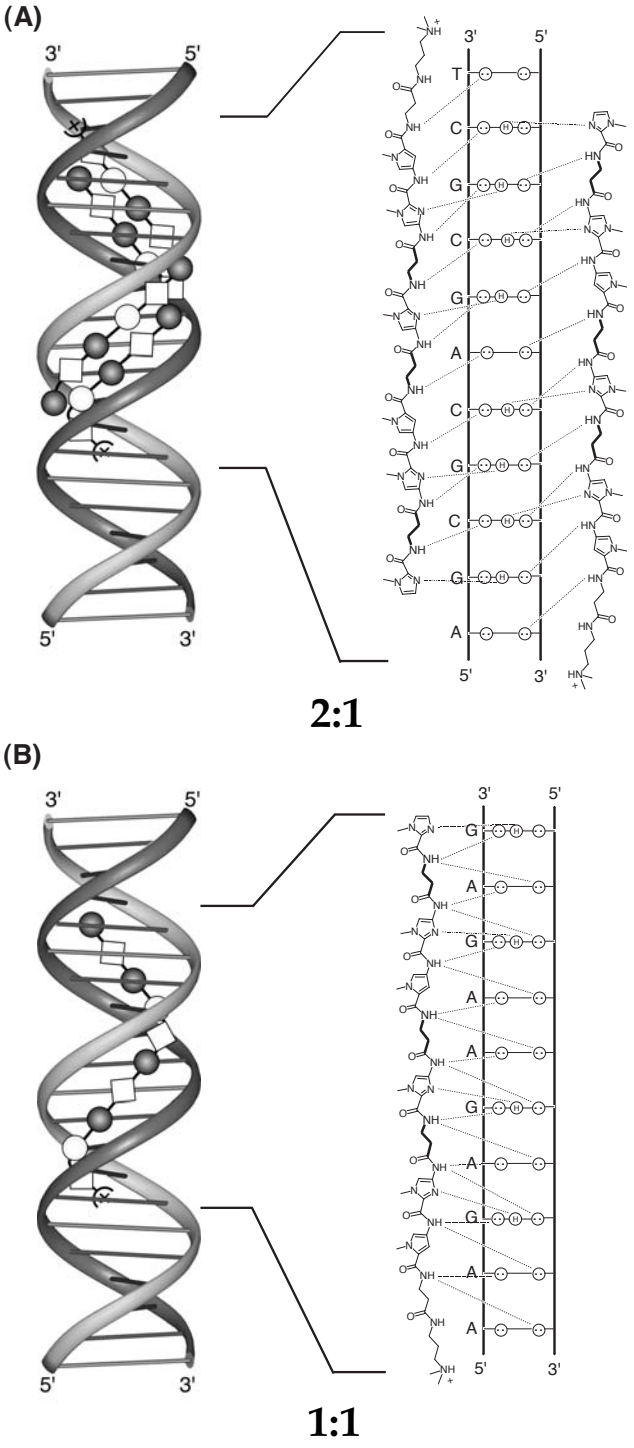


Figure 10 Stoichiometry-dependent sequence targeting. Polyamide **2**, shown here, can bind in 2:1 (A) and 1:1 (B) ligand:DNA complexes, specifying the very different sequences, 5'-AGCGCAGCGCT-3 and 5'-AAGAGAAGAG-3', respectively.

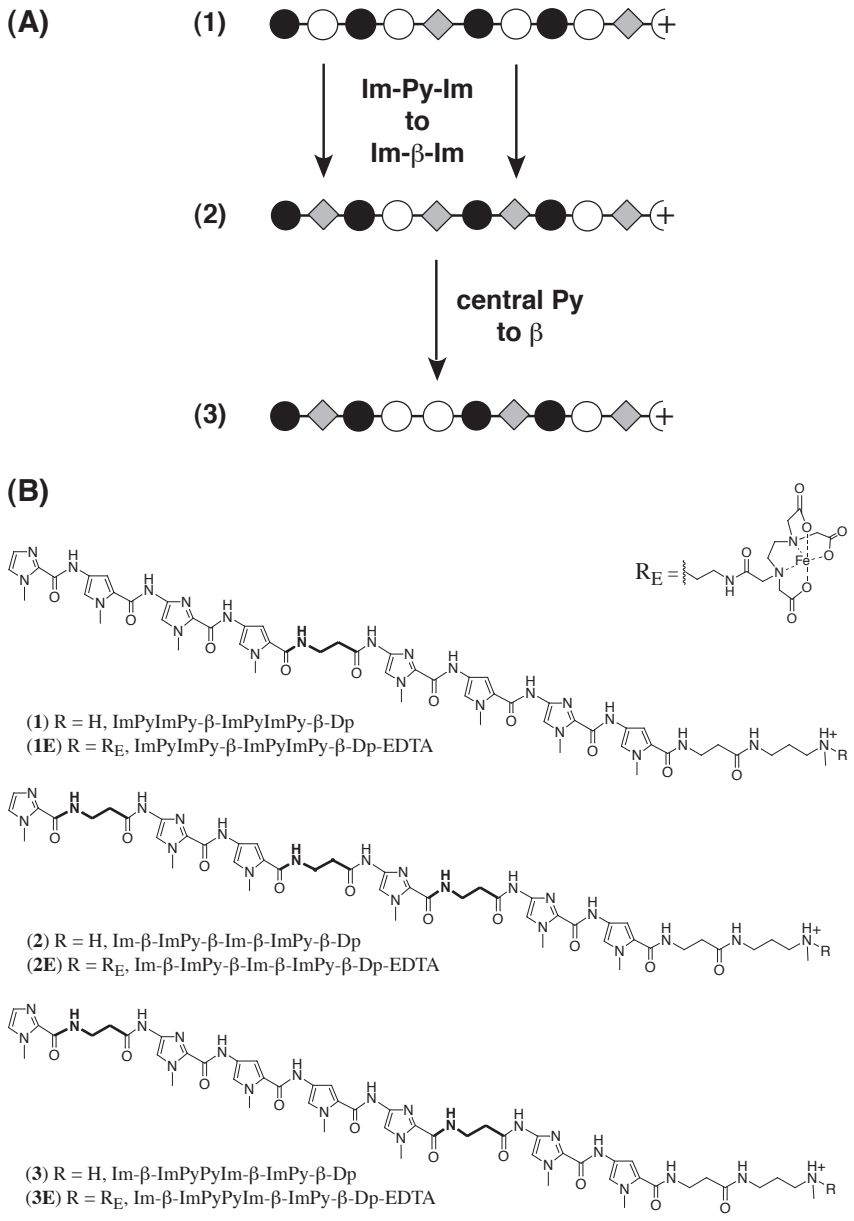


Figure 11 (A) Selective Py → β substitutions in polyamides **1**, **2**, and **3** examined in this study. (B) Chemical structures of these compounds and their EDTA conjugates, **1E**, **2E**, and **3E**. Polyamide sequences are indicated beneath each structure.

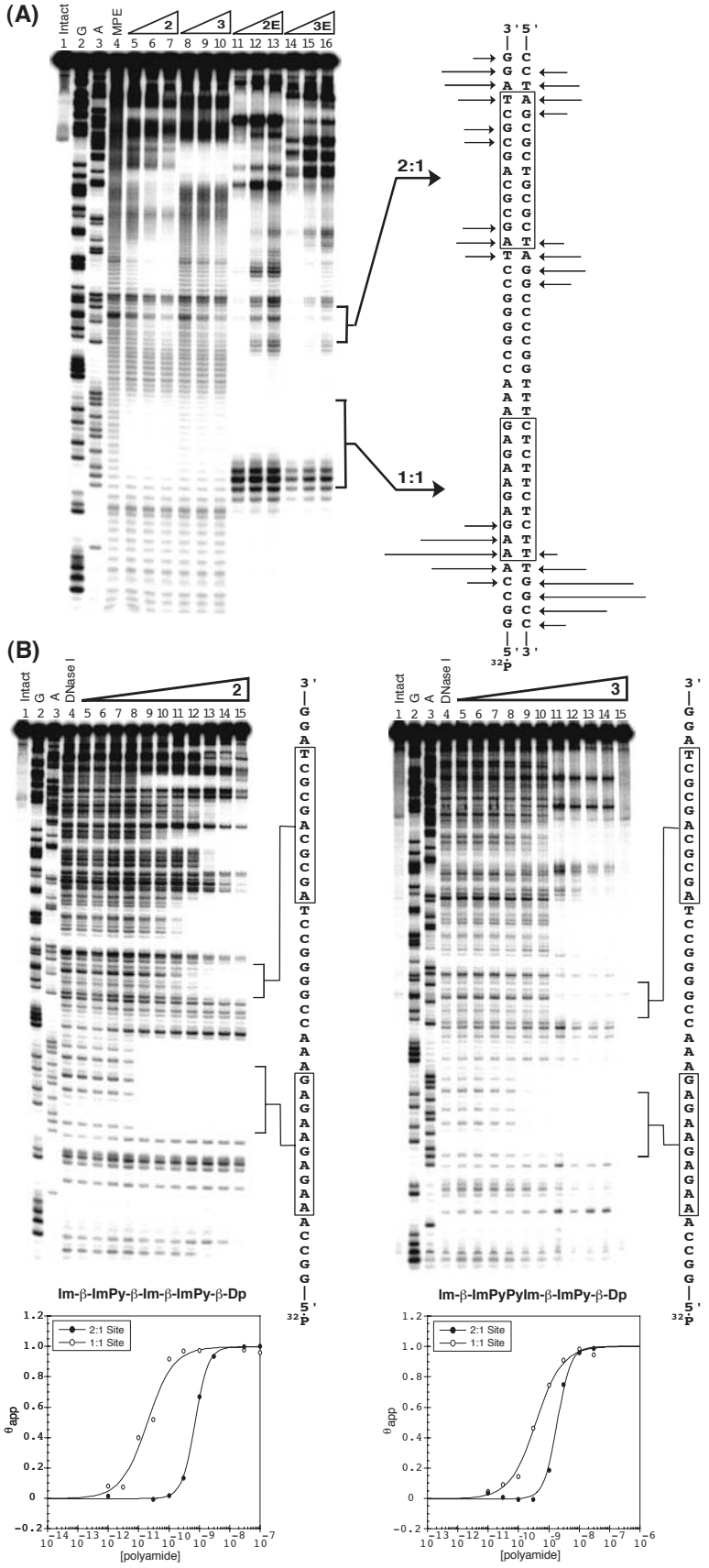
Synthesis. Polyamides **1-3** were synthesized in 12 steps from Boc-β-Pam-resin (1.25 g resin, 0.26 mmol/g substitution) using previously described solid-phase methods (Baird and Dervan, 1996). Nonterminal imidazole residues were introduced as dimers, Boc-Py-Im-COOH and Boc-β-Im-COOH. The polyamide

was cleaved from the solid support by aminolysis of the resin ester linkage using dimethylaminopropylamine (Dp) or 3,3'-diamino-N-methyldipropylamine (Dp-NH_2). Products were purified by reversed-phase preparatory HPLC to provide ImPyImPy-β-ImPyImPy-β-Dp (**1**), ImPyImPy-β-ImPyImPy-β-Dp-NH_2 (**1**-NH_2), Im-β-ImPy-β-Im-β-ImPy-β-Dp (**2**), Im-β-ImPy-β-Im-β-ImPy-β-Dp-NH_2 (**2**-NH_2), Im-β-ImPyPyIm-β-ImPy-β-Dp (**3**), and Im-β-ImPyPyIm-β-ImPy-β-Dp-NH_2 (**3**-NH_2). The purified polyamides with primary amines at the C terminus were treated with an excess of EDTA dianhydride (DMSO/NMP, DIEA, 55 °C, 15 min) and the remaining anhydride was hydrolyzed (0.1 N NaOH, 55 °C, 10 min). The EDTA conjugates were then purified by reversed-phase preparatory HPLC to yield **1E**, **2E**, and **3E**.

MPE•Fe(II) Footprinting and Affinity Cleaving. The plasmid pAU9 was constructed, which contains two binding sites, 5'-AGCGCAGCGCT-3' and 5'-AAGAGAAGAG-3', in order to examine 2:1 and 1:1 polyamide:DNA binding, respectively. MPE•Fe(II) footprinting (Van Dyke and Dervan, 1983) on the 3'- and 5'-^{32}P end-labelled 253 base pair *EcoRI*/*PvuII* restriction fragment (3') or PCR product (5') from plasmid pAU9 (28.6 mM HEPES, 285.7 mM NaCl, pH 7.0, 22°C) reveals that polyamides **2** and **3**, each at 100 nM concentration, bind the designed sites 5'-AGCGCAGCGCT-3' and 5'-AAGAGAAGAG-3' (Figure 12). No footprinting is observed for polyamide **1** at concentrations ≤1 μM.

Affinity cleaving experiments (Schultz and Dervan, 1984) on the same DNA fragments (28.6 mM HEPES, 285.7 mM NaCl, pH 7.0, 22 °C) with the EDTA•Fe(II) analogs **1E**, **2E**, and **3E** reveals 3'-shifted cleavage patterns indicating minor groove binding for polyamides **2E** (at 10 nM) and **3E** (at 100 nM) (Figure 12-A). Polyamides **2E** and **3E** cleave both sides of site,

Figure 12 (A) MPE footprinting (polyamides **2** and **3**) and affinity cleaving experiments (**2E** and **3E**) on the 5'-end-labelled 253 bp PCR product from pAU9: lane 1, intact DNA; lane 2, G reaction; lane 3, A reaction; lane 4, MPE standard; lanes 5-7, 10 nM, 30 nM, and 100 nM **2**, respectively; lanes 8-10, 10 nM, 30 nM, and 100 nM **3**, respectively; lanes 11-13, 1 nM, 3 nM, and 10 nM **2E**, respectively; lanes 14-16, 10 nM, 30 nM, and 100 nM **3E**, respectively. Relative cleavage intensities and MPE footprints for each binding site are illustrated at right by arrows and boxes, respectively. (B, top) Quantitative DNase I footprint titration experiments on the 5'-end-labelled 253 bp PCR product from pAU9 (left) with Im-β-ImPy-β-Im-β-ImPy-β-Dp, **2**: lane 1, intact DNA; lane 2, G reaction; lane 3, A reaction; lane 4, DNase I standard; lanes 5-15, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, and 100 nM **2**, respectively; and (right) with Im-β-ImPyPyIm-β-ImPy-β-Dp, **3**: lane 1, intact DNA; lane 2, G reaction; lane 3, A reaction; lane 4, DNase I standard; lanes 5-15, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, 100 nM, 300 nM, and 1 μM **3**, respectively. 2:1 and 1:1 binding sites are shown boxed at the right of each gel. (bottom) Binding isotherms for each DNase I footprint titration experiment directly above. The data points for the 5'-AGCGCAGCGCT-3' (2:1) site are indicated by filled circles and the 5'-GAGAAGAGAA-3' (1:1) site by open circles. The solid lines are best fit Langmuir binding isotherms obtained by a nonlinear least-squares algorithm, as described in the Experimental.


(A)
Intact
G
A
MPE
2
3
2E
3E
2:1
1:1
(B)
DNase I
Im-β-ImPy-β-Im-β-ImPy-β-Dp
Im-β-ImPyPyIm-β-ImPy-β-Dp
2:1 Site
1:1 Site
θapp
[polyamide]

5'-AGCGCAGCGCT-3', consistent with binding as an antiparallel dimer. However, **2E** and **3E** reveal cleavage only at the 5'-end of the site, 5'-AAGAGAAGAG-3', suggesting one orientation and a 1:1 polyamide:DNA binding stoichiometry, similar to the finding of Laemmli (Janssen et al., 2000). No cleavage is observed for polyamide **1E** at concentrations ≤ 1 μM.

Quantitative DNase I Footprint Titrations. Quantitative DNase I footprint titration experiments (Brenowitz et al., 1986; Trauger and Dervan, 2001) were performed to determine the apparent equilibrium dissociation constants (K_D = concentrations of polyamide bound at half-saturation) of **1-3** at each of the target sites (Figure 12-B). The 5'-AGCGCAGCGCT-3' site was bound by **2** and **3** with similar affinities, K_D = 0.63 nM and 1.9 nM, respectively, displaying cooperative binding isotherms (n=2) for both molecules, consistent with 2:1 binding. However, the 5'-AAGAGAAGAG-3' site was bound 16-fold more tightly by polyamide **2** (K_D = 0.02 nM) than polyamide **3** (K_D = 0.32 nM), displaying noncooperative binding isotherms (n=1) for both molecules, consistent with 1:1 stoichiometry. Polyamide **1** bound with poor affinity and at concentrations ≥100 nM in a nonspecific manner (Table 1).

Table 1 Apparent Equilibration Dissociation Constants (K_D)

Polyamide 1	Polyamide 2	Polyamide 3
2:1		
5'-A G C G C A G C G C T-3' 3'-T C G C G T C G C G A-5' **≥ 100 nM**	5'-A G C G C A G C G C T-3' 3'-T C G C G T C G C G A-5' **0.63 nM (±0.08)**	5'-A G C G C A G C G C T-3' 3'-T C G C G T C G C G A-5' **1.9 nM (±0.1)**
1:1		
5'-A A G A G A A G A G-3' 3'-T T C T C T T C T C-5' **≥ 100 nM**	5'-A A G A G A A G A G-3' 3'-T T C T C T T C T C-5' **0.020 nM (±0.002)**	5'-A A G A G A A G A G-3' 3'-T T C T C T T C T C-5' **0.32 nM (±0.04)**

Discussion. Polyamides **2** and **3**, which preserve the Im-β-Im unit, bind DNA sites of similar size (10 and 11-base pairs), but remarkably different sequence compositions, which is related to the stoichiometry of complexation in the minor groove. In the original "lexitropsin" model, based on 1:1 binding of polyamides to DNA, Im was proposed to specify G•C/C•G > A•T/T•A (Kopka et al., 1985). We now know this not to be the case. In the study detailed in the next section, Im, in the sequence context of polyamide **2**, is shown to bind all four base pairs (within a factor of 2) (Urbach and Dervan, 2001). From the crystal structure of the 1:1 netropsin:DNA complex, we understand the molecular mechanism by which Py specifies A•T/T•A > G•C/C•G (Kopka et al., 1985). However, this 1:1 recognition code of Py selecting A•T/T•A > G•C/C•G must now be modified to include the judicious placement of β for A•T/T•A recognition to reset the curvature in the 1:1 motif. This is substantiated by our observation that Im-β-Im binds 5′-GAG -3′ with higher affinity than Im-Py-Im (Table 1).

There are several implications for the addition of a 1:1 motif to current 2:1 targeting of mixed G,C/A,T sequences of DNA in the minor groove. Assuming cell permeability is related to size, one could argue that a 1:1 binding molecule will occupy larger binding site sizes for the same molecular weight as a hairpin polyamide. In fact, the difference of one methylene unit (β vs. γ) can direct whether a polyamide binds extended as 1:1 complex or folds as a hairpin (Mrksich et al., 1994; Trauger et al., 1996a). We face the added complexity of the same molecule binding very different DNA sites depending on stoichiometry. If specificity is the goal, this raises the challenge for chemists to design next-generation polyamides, which enforce 2:1 vs. 1:1 binding in the minor groove.

Toward Rules for 1:1 Polyamide:DNA Recognition[2]

Purpose. At the forefront of the endeavor to control gene expression by small molecules is the elucidation of chemical principles for direct read-out of predetermined sequences of double-stranded DNA. Although there exists a significant body of literature on the 1:1 mode of binding (Lown et al., 1986), much of this was carried out before quantitative footprinting methods were introduced to the field. In an effort to characterize more rigorously the 1:1 motif, we address several questions quantitatively: 1) Can a 1:1 recognition code be established, which uses individual ligand residues (e.g. Py, Im, Hp or β) to specify individual Watson•Crick base pairs? 2) How is orientation related to overall DNA sequence type? and 3) What is the effect of ligand size on binding affinity in the 1:1 motif?

Specificity of Py, Im, Hp and β

Approach. The polyamide Im-β-ImPy-β-Im-β-ImPy-β-Dp (**2**) was chosen as the template to examine the specificity at Im, Py and β residues in an oriented 1:1 complex with DNA. Because Hp/Py pairs were shown to discriminate T•A from A•T in the 2:1 motif, a second ligand Im-β-ImHp-β-Im-β-ImPy-β-Dp (**4**) was prepared to explore any possible specificity the Hp residue may have in a 1:1 complex. Specificity at a single and unique carboxamide position was determined by varying a single base pair within the parent sequence context, 5'-AAAGAGAAGAG-3', to all four Watson•Crick base pairs and comparing the relative affinities for the four possible complexes. To meet this end, three

[2] The text of this section is taken from Urbach and Dervan, 2001.

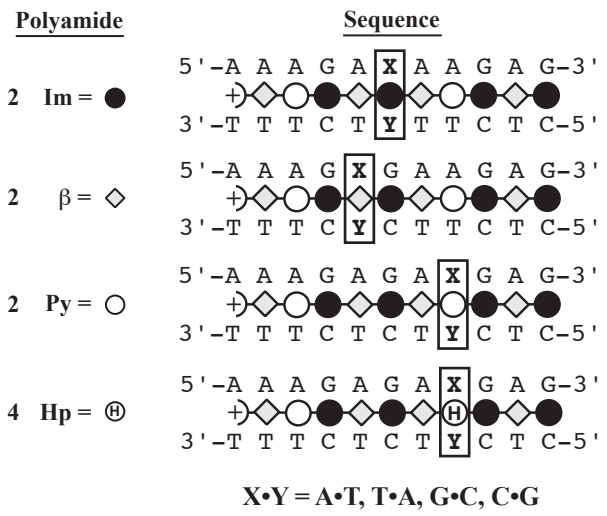


Figure 13 Examination of sequence selectivity at a single imidazole (Im), beta alanine (β), pyrrole (Py), or hydroxypyrrole (Hp) position within the parent context, 5'-AAAGAGAAGAG-3'. Imidazole and pyrrole rings are represented as shaded and nonshaded circles, respectively; β-alanines are shown as gray diamonds; and hydroxypyrrole is indicated by a circle containing the letter H.

plasmids were cloned, each containing four binding sites: pAU8 (for Im), 5'-AAAGA**X**AAGAG-3'; pAU15 (for β), 5'-AAAG**X**GAAGAG-3'; and pAU12 (for Py and Hp) 5'-AAAGAGA**X**GAG-3', where **X** = A, T, G, and C (Figure 13). The Hp-containing polyamide (**4**) was synthesized by solid phase methods, as described previously (Urbach et al., 1999).

DNA Binding Affinity and Sequence Specificity. Quantitative DNase I footprinting was carried out for polyamides **2** and **4** on PCR products of pAU8, pAU15, and pAU12 (Figure 14). The variable base pair position was chosen opposite the amino acid residue in question. Specificity of Im: polyamide **2** binds the four DNA sites 5'-AAAGA**X**AAGAG-3' (**X** =A, T, G, C), with similar high affinities, K_a = 2.6 – 1.1 x 10^{10} M^{-1}, revealing that Im tolerates all four base pairs.

Figure 14 Quantitative DNase I footprint titration experiments for polyamide **2** on the 298 bp, 5'-end-labelled PCR product of plasmids pAU8 (A), pAU15 (B), and pAU12 (C), as well as polyamide **4** on the PCR product of pAU12 (D): (A) and (C), lane 1, intact DNA; lane 2, G reaction; lane 3, A reaction; lane 4, DNase I standard; lanes 5 – 14, 300 fM, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM **2**, respectively; (B) lane 1, intact DNA; lane 2, G reaction; lane 3, A reaction; lane 4, DNase I standard; lanes 5 – 15, 100 fM, 300 fM, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM 1, respectively; (D) lane 1, intact DNA; lane 2, G reaction; lane 3, A reaction; lane 4, DNase I standard; lanes 5 – 14, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, 100 nM, 300 nM **4**, respectively. Each footprinting gel is accompanied by the following: (above) binding schematic with the mutated position boxed; (left, top) chemical structure of the monomer of interest; and (left, bottom) Langmuir binding isotherms for the four designed sites. θ_{norm} values were obtained using a nonlinear, least-squares algorithm.

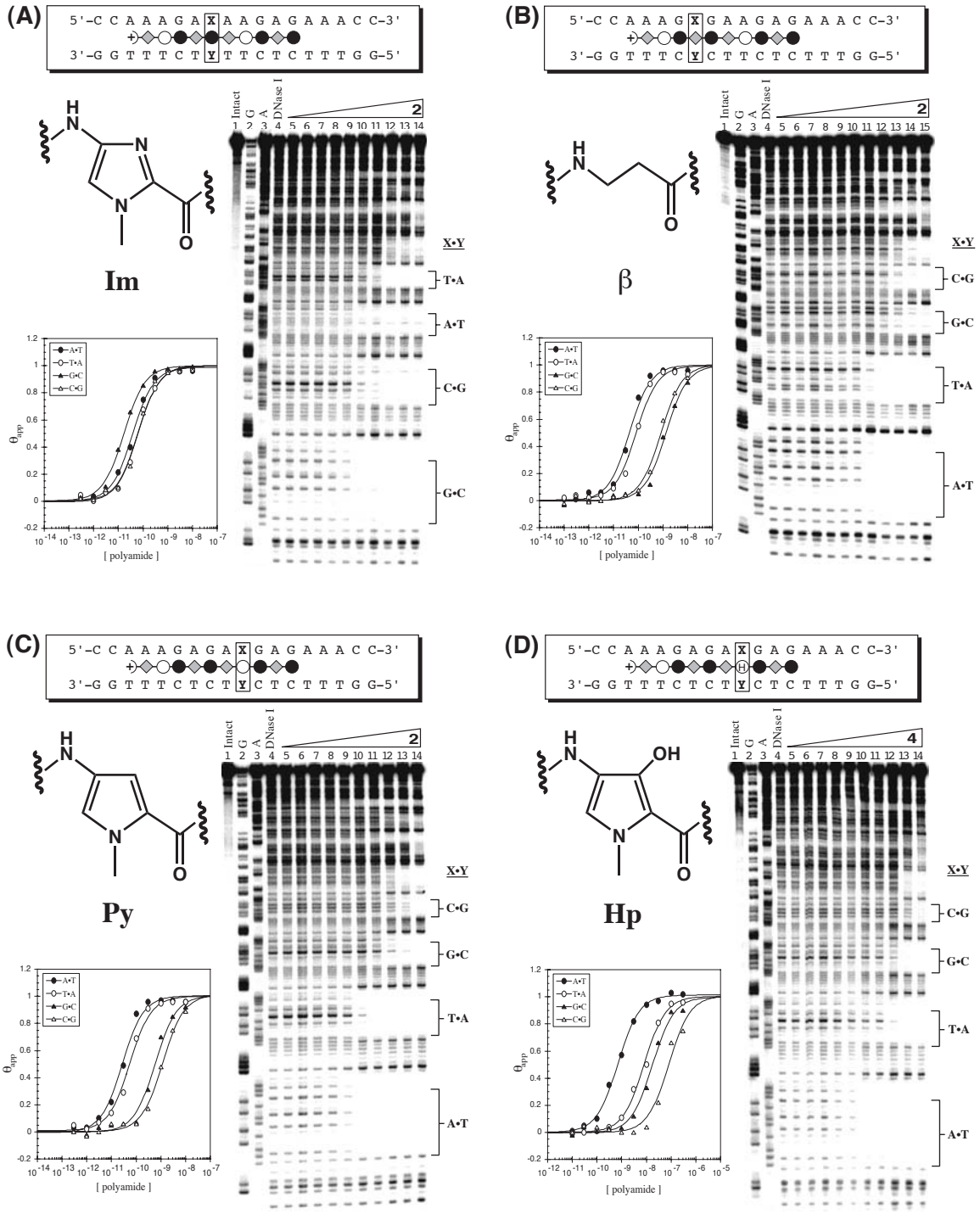

(A)
5'-C C A A A G A X A A G A G A A A C C-3'
3'-G G T T T C T Y T T C T C T T T G G-5'
Im
Intact
G
A
DNase I
X•Y
T•A
A•T
C•G
G•C
A•T
T•A
G•C
C•G
θapp
[polyamide]
(B)
5'-C C A A A G X G A A G A G A A A C C-3'
3'-G G T T T C Y C T T C T C T T T G G-5'
β
(C)
5'-C C A A A G A G A X G A G A A A C C-3'
3'-G G T T T C T C T Y C T C T T T G G-5'
Py
(D)
5'-C C A A A G A G A X G A G A A A C C-3'
3'-G G T T T C T C T Y C T C T T T G G-5'
OH
Hp

Table 2 Equilibrium Association Constants, K_a (M^{-1})[a,b]

Residue	A•T	T•A	G•C	C•G
Im (**2**)[c]	2.5 (± 0.2) x 10^{10}	1.1 (± 0.1) x 10^{10}	2.6 (± 0.4) x 10^{10}	1.3 (± 0.3) x 10^{10}
β (**2**)	2.4 (± 0.1) x 10^{10}	1.3 (± 0.1) x 10^{10}	4.3 (± 1.4) x 10^{8}	7.8 (± 1.9) x 10^{8}
Py (**2**)	3.4 (± 0.3) x 10^{10}	1.5 (± 0.2) x 10^{10}	1.8 (± 0.2) x 10^{9}	8.6 (± 1.5) x 10^{8}
Hp (**4**)	1.6 (± 0.2) x 10^{9}	1.3 (± 0.2) x 10^{8}	4.9 (± 0.8) x 10^{7}	1.0 (± 0.3) x 10^{7}

[a] Values reported are the mean values from at least three DNase I footprint titration experiments, with the standard deviation given in parentheses. [b] Assays were performed at 22 ^{o}C in a buffer of 10 mM Tris•HCl, 10 mM KCl, 10 mM, $MgCl_2$, and 5 mM $CaCl_2$ at pH 7.0. [c] The number in parentheses indicates the compound containing the unique residue.

Specificity of Py and β: polyamide **2** binds the target sites, 5'-AAAGAGA**X**GAG-3' and 5'–AAAG**X**GAAGAG-3', respectively (**X** = A, T, G, C), with high affinity and in both cases displays a preference for A•T and T•A > G•C and C•G by at least a factor of 10. Remarkably, substituting one Py residue with Hp afforded the most specific polyamide (**4**), which binds the sequences 5'-AAAGAGA**X**GAG-3' with a modest loss in affinity, characteristic of the Hp residue (White et al., 1998; Kielkopf et al., 1998b), and with a tenfold single site preference for **X** = A > T > G/C (Table 2). Langmuir binding isotherms for each complex fit well to an n = 1 Hill equation (see Experimental for equation), which is consistent with a 1:1 ligand:DNA stoichiometry (Figure 14). To establish that the specificity of Im and β is not position dependent, controls were performed on polyamide **2** at different Im (pAU16, 5'-AAAGAGAA**X**AG-3') and β positions (pAU13, 5'-AAAGAG**X**AGAG-3'). The observed complex affinities and sequence specificities were similar to that described above for pAU8 and pAU15, respectively. The more significant effect of position on the sequence specificity of Hp is presented in the next section.

Discussion. This quantitative study helps to elucidate the current state-of-the-art for 1:1 polyamide:DNA complexes and creates a baseline for future specificity studies. An Im residue binds each of the four Watson•Crick base pairs with high affinity, whereas a β or Py residue prefers A,T over G,C base pairs but does not discriminate A from T. Steric inhibition between the exocyclic amino group of guanine and the Py and β residues may explain their A,T preference, as previously suggested by Dickerson from x-ray structural analysis of 1:1 complexes. Based on the study of netropsin bound in a 1:1 complex with DNA, the promiscuous nature of the Im residue accepting G,C as well as A,T was anticipated (Kopka et al., 1985). The unanticipated result of this study is the observation that an Hp residue in this polyamide sequence context can distinguish one of the four Watson•Crick base pairs. Whether the hydroxyl moiety lies asymmetrically in the cleft between A and T and makes a specific hydrogen bond to the O2 of T, as observed for Hp/Py recognition of T•A (Kielkopf et al., 1998b), is unclear. However, the fact that a single aromatic carboxamide residue can select one of the four Watson•Crick base pairs within the 1:1 motif is an encouraging step toward a set of rules for DNA recognition similar to the 2:1 motif. Whether new aromatic residues can be invented to complete a 1:1 recognition code is addressed in the next section.

Specificity of Novel Heterocyclic Amino Acids[3]

Approach. In an effort to improve upon the 1:1 recognition code established in the previous section, we address the issues of whether new heterocycles can be developed to discriminate between the four Watson-Crick base pairs in the 1:1 motif, and whether we can understand the relationship between overall heterocycle structure and DNA sequence specificity. This specificity has been attributed largely to the unique functionality presented by each heterocycle to the floor of the minor groove. However, little has been done to assess the ramifications of functional groups pointing away from the DNA. Figure 15 shows a family of five-membered aromatic, heterocyclic residues grouped into columns by the type of functionality directed toward the DNA minor groove. Py and 1H-pyrrole (Nh) project a hydrogen with positive

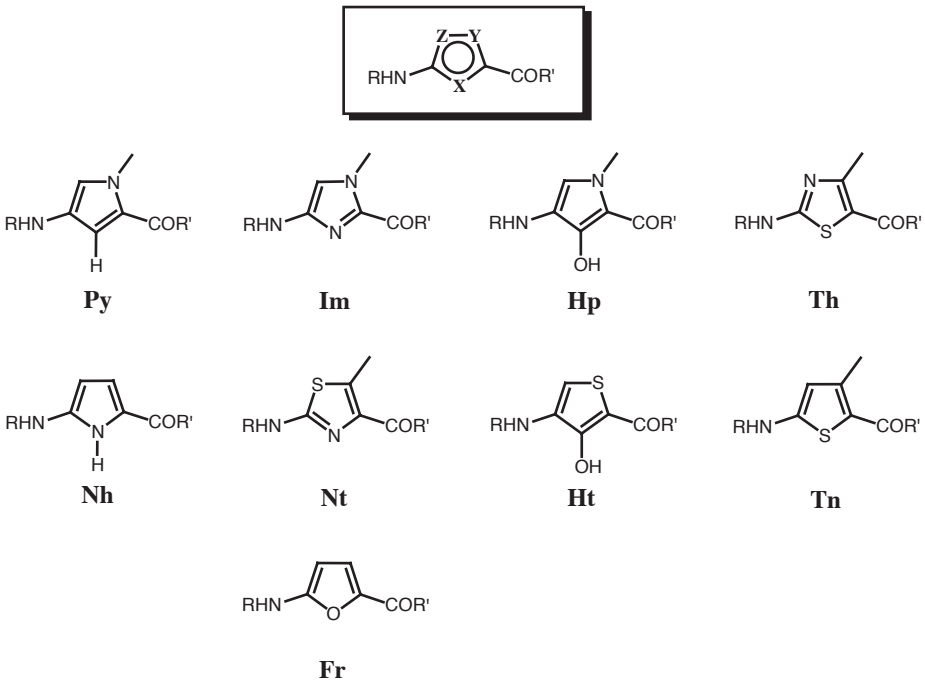
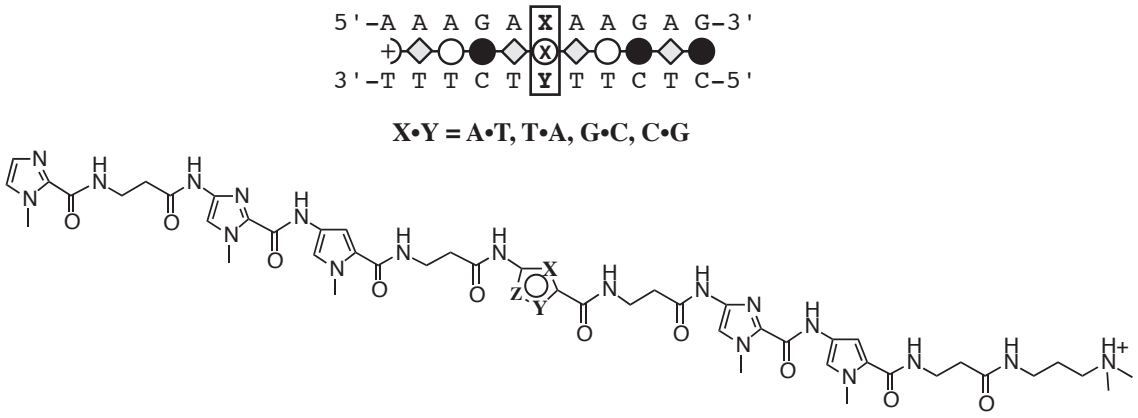


Figure 15 Family of five-membered aromatic heterocycles.

[3] The text of this section is taken from Marques et al., in preparation.

potential toward the DNA; Im, 5-methylthiazole (Nt), and furan (Fr) project an sp^2 lone pair from nitrogen or oxygen; Hp and 3-hydroxythiophene (Ht) project a hydroxyl group; and 4-methylthiazole (Th) and 4-methylthiophene (Tn) project an sp^2 lone pair from sulfur. Comparative analysis of new residues within this five-membered heterocyclic framework should enable us to retain overall ligand morphology and to observe the effects of small structural changes, such as single atom substitution, on DNA base pair specificity. Specificity at the unique carboxamide position was determined by varying a single base pair (**X**) within the sequence context 5′-AAAGA**X**AAGAG-3′ to all four Watson•Crick base pairs and comparing the relative affinities for the four possible complexes. *Ab initio* computational modeling of the heterocyclic amino acids was implemented to derive their inherent geometric and electronic parameters. The combination of these techniques has provided an interesting perspective on the origin of DNA sequence discrimination by polyamides.

Synthesis. Synthesis of Boc-protected Nh, Fu, Ht, Nt, and Tn amino acids and the corresponding polyamides Im-β-ImPy-β-**X**-β-ImPy-β-Dp (**X** = unique heterocycle) required new solution and solid-phase synthetic methodologies, which were developed by Michael Marques and Raymond Doss and will be reported elsewhere (Marques et al., in preparation). The structures of polyamides Im-β-Im-Py-β-**Py**-β-Im-Py-β-Dp (**5**), Im-β-Im-Py-β-**Hp**-β-Im-Py-β-Dp (**6**), Im-β-Im-Py-β-**Nh**-β-Im-Py-β-Dp (**7**), Im-β-Im-Py-β-**Ht**-β-Im-Py-β-Dp (**8**), Im-β-Im-Py-β-**Fr**-β-Im-Py-β-Dp (**9**), Im-β-Im-Py-β-**Nt**-β-Im-Py-β-Dp (**10**), Im-β-Im-Py-β-**Tn**-β-Im-Py-β-Dp (**11**), and Im-β-Im-Py-β-**Th**-β-Im-Py-β-Dp (**12**), with the parent Im polyamide (**2**) are shown in Figure 16.



(**2**) Im-β-ImPy-β-**Im**-β-ImPy-β-Dp, X = N, Y=N-Me, Z = C-H
(**5**) Im-β-ImPy-β-**Py**-β-ImPy-β-Dp, X = C-H, Y=N-Me, Z = C-H
(**6**) Im-β-ImPy-β-**Hp**-β-ImPy-β-Dp, X = C-OH, Y=N-Me, Z = C-H
(**7**) Im-β-ImPy-β-**Nh**-β-ImPy-β-Dp, X = NH, Y=C-H, Z = C-H
(**8**) Im-β-ImPy-β-**Ht**-β-ImPy-β-Dp, X = S-OH, Y=S, Z = C-H
(**9**) Im-β-ImPy-β-**Fr**-β-ImPy-β-Dp, X = O, Y=C-H, Z = C-H
(**10**) Im-β-ImPy-β-**Nt**-β-ImPy-β-Dp, X = N, Y=C-Me, Z = S
(**11**) Im-β-ImPy-β-**Tn**-β-ImPy-β-Dp, X = S, Y=C-Me, Z = C-H
(**12**) Im-β-ImPy-β-**Th**-β-ImPy-β-Dp, X = S, Y=C-Me, Z = N

Figure 16 Chemical structures for 1:1 polyamides containing novel heterocyclic residues, with the variable positions in the central ring indicated by the letters X, Y, and Z. A binding model is shown at top with the variable polyamide position indicated by a circle containing the letter X proximal to the variable base pair, X•Y.

DNA Binding Affinity and Sequence Specificity. Quantitative DNase I footprinting was carried out for polyamides **5** – **12** on the 298 bp PCR product of pAU8 (Figures 17 and 18). The variable base pair position was installed opposite the amino acid residue in question. Equilibrium association constants (K_a) for 1:1 polyamides containing Im, Py, and Hp residues tested against the four Watson-Crick base pairs have been discussed. However, in that study only the Im specificity experiment was performed at the more flexible central residue, as with the new polyamides reported here. Therefore, new polyamides containing Py and Hp residues at the central position have been included in this study for a

Figure 17 (A-D) Quantitative DNase I footprint titration experiments for polyamides **5**-**8**, respectively, on the 298 bp, 5'-end-labelled PCR product of plasmid pAU8: (A and B) lane 1, intact DNA; lane 2, G reaction; lane 3, A reaction; lane 4, DNase I standard; lanes 5-15, 100 fM, 300 fM, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM polyamide, respectively. (C) lane 1, intact DNA; lane 2, G reaction; lane 3 A reaction; lane 4, DNase I standard; lanes 5-15, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, 100 nM polyamide, respectively. (D) lane 1, intact DNA; lane 2, G reaction; lane 3 A reaction; lane 4, DNase I standard; lanes 5-15, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, 100 nM, 300 nM polyamide respectively. Each footprinting gel is accompanied by the following: (left, top) chemical structure of the residue of interest; and (left bottom) Langmuir binding isotherm for the four designed sites. θ_{norm} values were obtained using a nonlinear least-squares fit.

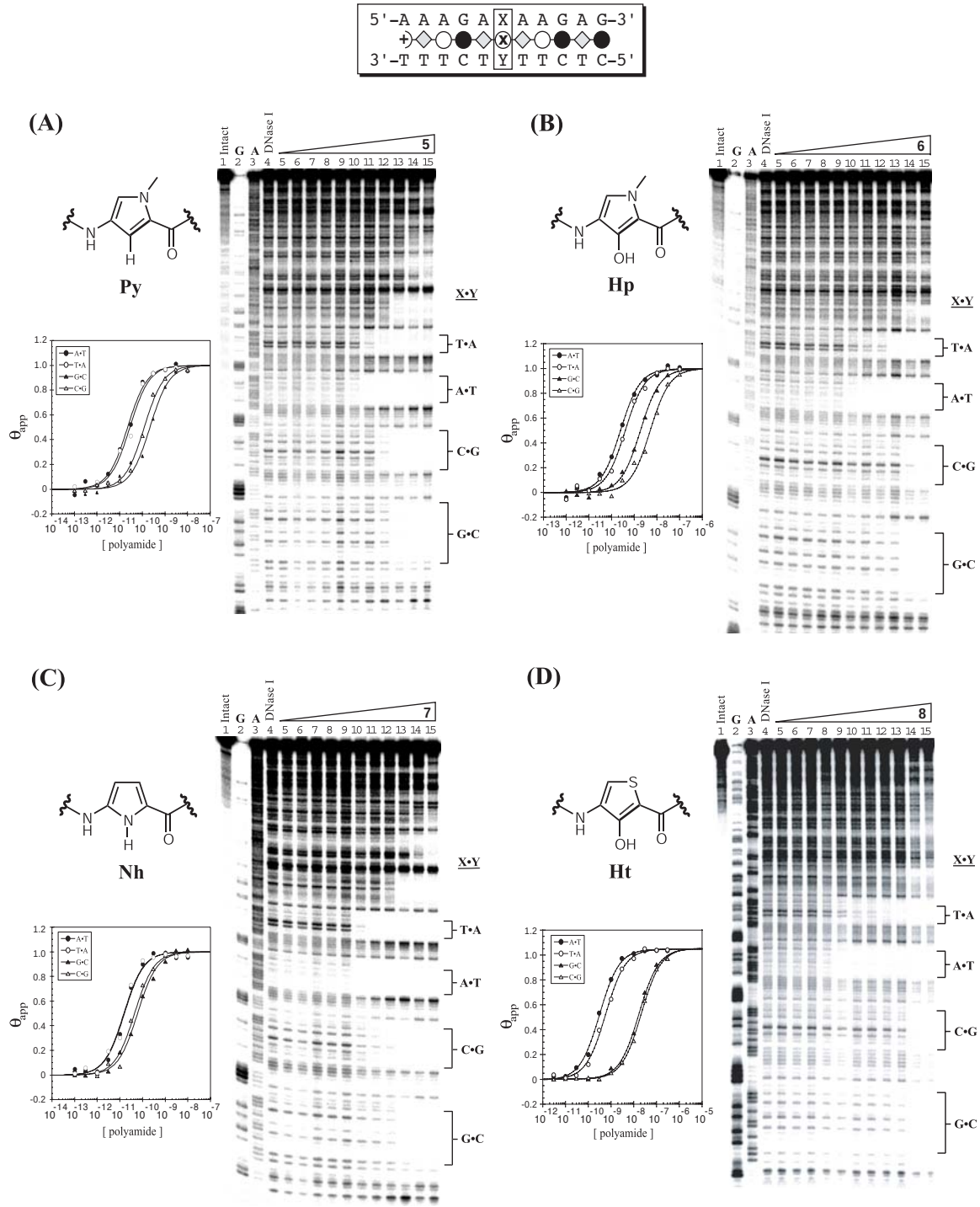
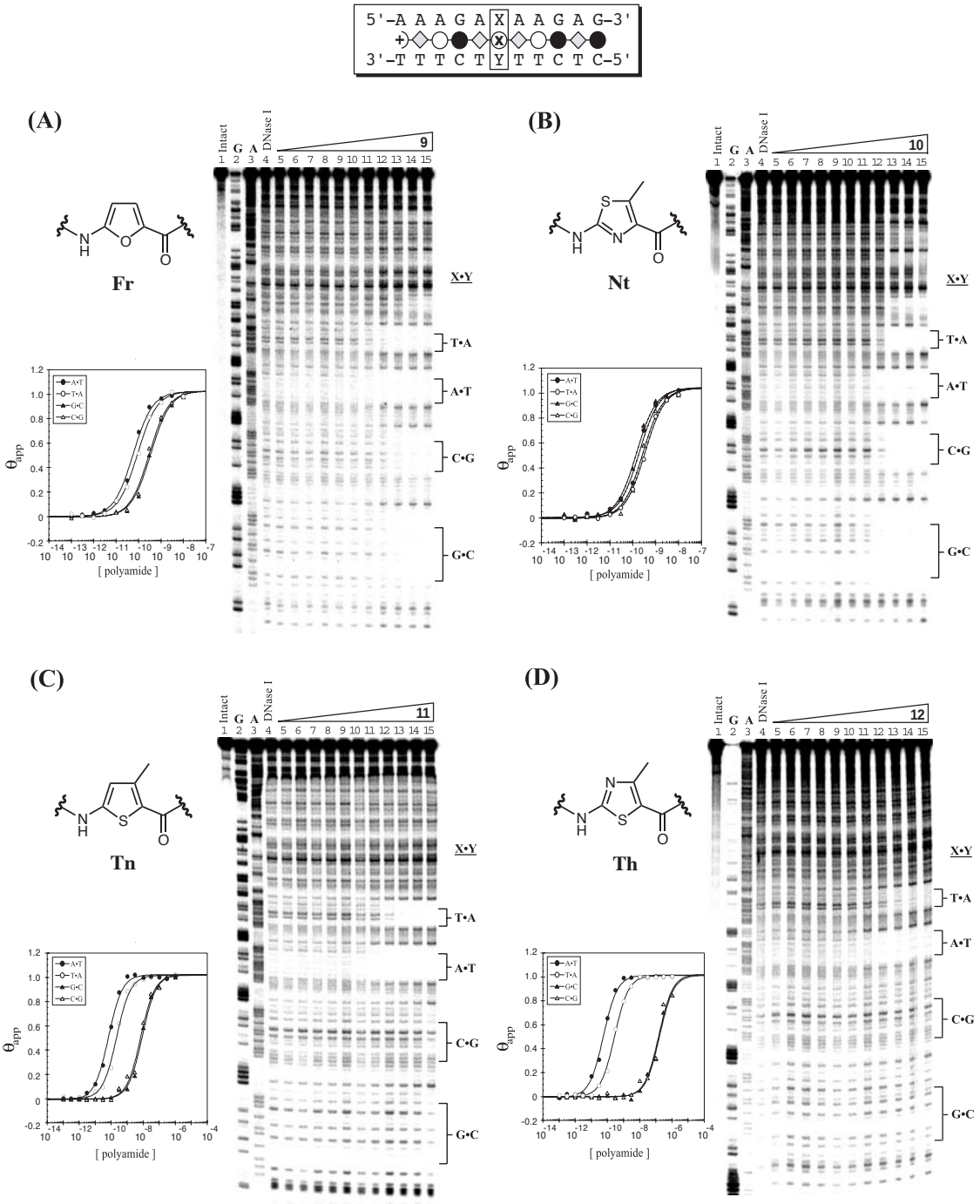

5'–A A A G A X A A G A G–3'
3'–T T T C T Y T T C T C–5'
(A)
Py
(B)
Hp
(C)
Nh
(D)
Ht
Intact
G A
DNase I
X•Y
T•A
A•T
C•G
G•C
θapp
[polyamide]
A•T
T•A
G•C
C•G

Figure 18 (A-D) Quantitative DNase I footprinting experiments for polyamides **9**-**12**, respectively, on the 298 bp, 5'-end-labelled PCR product of plasmid pAU8: lane 1, intact DNA; lane 2, G reaction; lane 3, A reaction; lane 4, DNase I standard; lanes 5-15, 100 fM, 300 fM, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, respectively. Each footprinting gel is accompanied by the following: (left, top) Chemical structure of the residue of interest; and (left bottom) Langmuir binding isotherm for the four designed sites. θ_{norm} values obtained using a nonlinear least-squares fit. Isotherms for C and D were generated from gels run out to a final concentration of 1 uM (not shown).


5'-A A A G A X A A G A G-3'
3'-T T T C T Y T T C T C-5'
(A)
Fr
(B)
Nt
(C)
Tn
(D)
Th
Intact
G A
DNase I
X•Y
T•A
A•T
C•G
G•C
θapp
[polyamide]
A•T
T•A
G•C
C•G

more controlled comparison. Polyamide **5** (Py) binds with very high affinity (K_a ~ 6 x 10^{10} M^{-1}) at the **X** = A, T sites (5'-AAAGA**X**AAGAG-3') with a 5- to 10-fold preference over **X** = G, C (Table 3). Polyamide **6** (Hp) binds with lower affinity (K_a ~ 3 x 10^9 M^{-1}) but with similar specificity to **5**, preferring **X** = A, T > G, C by 5- to 10-fold. The Nh-containing polyamide (**7**) bound with very high affinity to the **X** = A, T sites (K_a = 7.5 x 10^{10} M^{-1}) but with a mere 3- to 5-fold selectivity over the high-affinity **X** = G, C sites. Compound **8** (Ht) bound with subnanomolar affinities to the **X** = A, T sites, similar to **6** but with ≥ 40-fold specificity for **X** = A, T > G, C. Polyamide **9** (Fr) showed high affinity for the **X** = A, T sites (K_a ~ 10^{10} M^{-1}) with a small 2- to 4-fold preference over **X** = **G**, **C**. The 5-methylthiazole-containing polyamide (**10**, Nt), which places the thiazole ring *nitrogen* into the floor of the minor groove, bound all four sites with similar high affinities (K_a ~ 5 x 10^9 M^{-1}). Thiophene-containing polyamide (**11,** Tn) showed modest single-site specificity, binding the **X** = **A** site at K_a = 3.0 x 10^{10} M^{-1} with 5-fold preference over

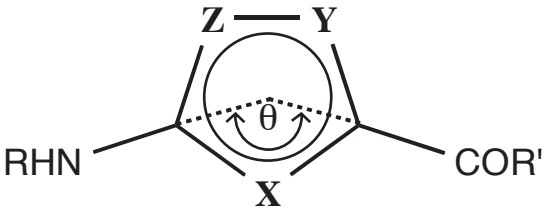
Table 3 Equilibrium Association Constants, K_a (M^{-1})[a,b] for Polyamides Containing Novel Heterocycles (X) within Im-β-ImPy-β-X-β-ImPy-β-Dp

Ring	A•T	T•A	G•C	C•G
Im (**2**)[c]	2.5 (± 0.2) x 10^{10}	1.1 (± 0.1) x 10^{10}	2.6 (± 0.4) x 10^{10}	1.3 (± 0.3) x 10^{10}
Py (**5**)	7.2 (± 0.3) x 10^{10}	5.3 (± 0.1) x 10^{10}	3.2 (± 0.4) x 10^9	9.4 (± 0.2) x 10^9
Hp (**6**)	3.9 (± 0.1) x 10^9	2.5 (± 0.3) x 10^9	5.3 (± 0.5) x 10^8	1.9 (± 0.5) x 10^8
Nh (**7**)	7.5 (± 0.2) x 10^{10}	7.4 (± 0.1) x 10^{10}	1.6 (± 0.2) x 10^{10}	2.3 (± 0.1) x 10^{10}
Ht (**8**)	2.8 (± 0.5) x 10^9	1.6 (± 0.6) x 10^9	3.8 (± 1.3) x 10^7	3.7 (± 0.7) x 10^7
Fr (**9**)	2.2 (± 0.5) x 10^{10}	1.0 (± 1.3) x 10^{10}	4.4 (± 0.5) x 10^9	5.0 (± 0.5) x 10^9
Nt (**10**)	5.4 (± 0.9) x 10^9	2.9 (± 0.6) x 10^9	8.0 (± 1.3) x 10^9	4.2 (± 0.6) x 10^9
Tn (**11**)	3.0 (± 0.2) x 10^{10}	5.7 (± 0.4) x 10^9	8.1 (± 0.4) x 10^7	8.3 (± 0.2) x 10^7
Th (**12**)	1.5 (± 0.2) x 10^{10}	3.0 (± 0.7) x 10^9	6.5 (± 0.5) x 10^6	7.4 (± 0.5) x 10^6

[a] Values reported are the mean values from at least three DNase I footprint titration experiments, with the standard deviation given in parentheses. [b] Assays were performed at 22 ^oC in a buffer of 10 mM Tris•HCl, 10 mM KCl, 10 mM, $MgCl_2$, and 5 mM $CaCl_2$ at pH 7.0. [c] The number in parentheses indicates the compound containing the unique residue.

X = **T** and ~ 70-fold preference over **X** = **G**, **C**. 3-methylthiophene-containing polyamide (**12**, Th), which places the thiazole ring *sulfur* into the floor of the minor groove, bound with similar **X** = A, T affinity as **11** (Tn) but with > 400-fold preference over **X** = G, C. In all cases, binding isotherms fit well to an n = 1 Hill equation, which is consistent with a 1:1 polyamide:DNA stoichiometry (Figures 17 and 18).

Calculations. Molecular modeling calculations were preformed by Michael Marques using *Spartan Essential* software package (Wavefunction Inc.). Each ring was first minimized using an AM1 model, followed by *Ab initio* calculations using the Hartree-Fock model and a 6-31G* polarization basis set. Each heterocycle exhibited a unique geometric and electronic profile (Figure 19). Bonding geometry for imidazole, pyrrole, and 3-hydroxypyrrole were in excellent agreement with coordinates derived from x-ray structures of polyamides containing these heterocycles (Kielkopf et al., 1998a; 1998b). The overall curvature of each monomer was calculated to be the sweep angle (θ)



Ring	X	Y	Z	θ (*degrees*)	Charge on X (*e*)
Fr	O	C-H	C-H	*126*	*-0.31*
Nt	N	C-Me	S	*127*	*-0.60*
Ht	O-H	S	C-H	*133*	*+0.40*
Nh	N-H	C-H	C-H	*136*	*+0.34*
Im	N	N-Me	C-H	*137*	*-0.71*
Py	C-H	N-Me	C-H	*146*	*+0.21*
Hp	O-H	N-Me	C-H	*148*	*+0.50*
Tn	S	C-Me	C-H	*149*	*-0.21*
Th	S	C-Me	N	*153*	*-0.25*

Figure 19 Geometric and electrostatic profiles for nine heterocyclic amino acids, derived from *ab initio* molecular modeling calculations using *Spartan Essential* software (Wavefunction, Inc.). (Top) Schematic illustrating the amide-ring-amide angle of curvature, θ. X, Y, and Z denote variable functionality at the different ring positions for each heterocycle. (Bottom) Table listing the functional groups at X, Y, and Z, along with the angle θ, and the electrostatic partial charge on X. For Ht, Nh, Py, and Hp, the positive charge on X is listed for the H atom.

created by the theoretical intersection of the two ring-to-amide bonds in each ring. The structures were ranked by increasing θ as follows: Fr > Nt > Ht > Nh > Im > Py > Hp > Tn > Th. The ring atom in closest proximity to the floor of the DNA minor groove was examined for partial charge. The structures were ranked by decreasing partial charge on this atom as follows: Hp > Ht > Nh > Py > Tn > Th > Fr > Im.

Discussion. The single-subunit•DNA complexes of the 1:1 motif provide a relatively flexible system for the exploration of novel recognition elements. Due to the conformational freedom imparted by the β residues, changes in heterocycle geometry do not have as much of an impact on DNA sequence recognition as in the hairpin motif (Marques et al., in preparation). In fact, all 1:1 compounds in this study bind with high affinity to the **X** = A, T sites but with varying degrees of **X** = A, T > G, C specificity. The high-resolution solution structure presented later in this Thesis reveals an important register of amide NH groups with the purine N3 and pyrimidine O2 groups on the floor of the DNA minor groove (Urbach et al., 2002). Given this alignment as a driving force for DNA recognition in the 1:1 motif, one may view the subtle differences in heterocycle curvature as merely placing the central ring atom (**X** in Figure 19) closer to or farther from the DNA. In this view, increasing the ring curvature decreases the polyamide-DNA intimacy, thereby diminishing DNA specificity. The results presented here fit well within this ideology.

Polyamides **5** (Py) and **7** (Nh) present a hydrogen with a positive potential to the minor groove floor. Both compounds exhibit a modest 3- to 5-fold selectivity for **X** = A, T > G, C, but **7** binds with higher affinity to all sites. The selectivity is probably due to the negative steric X-H to G-NH_2 interaction (X =

C3 for Py and N1 for Nh), which was predicted by Dickerson and coworkers for netropsin and supported by NMR studies discussed later in this Thesis (Kopka et al., 1985; Urbach et al., 2002). The higher affinity for **7** may be attributed to a combination of greater positive charge on N1-H and higher ring curvature, both of which should reduce specificity.

Polyamides **2** (Im), **9** (Fr), and **10** (Nt) present a small atom with an sp^2 lone pair directed toward the minor groove floor. Polyamide **2** was discussed in the previous section, binding all sites with high affinity and displaying virtually no discrimination between sites. **9** and **10** behave quite similarly. It is likely that the small atom (N for Im and Nt or O for Fr) presented to the DNA provides no steric clash with G-NH_2, and therefore all sites are bound with similarly high affinities.

Polyamides **6** (Hp) and **8** (Ht) present a hydroxyl group to the DNA minor groove. Previously in both hairpin and 1:1 systems, hydroxypyrrole successfully discriminated between A•T and T•A base pairs (White et al., 1998; Urbach and Dervan, 2001). Yet when flanked on both sides by β-alanine residues, as with the Hp compound presented here, single base-pair specificity is lost. This loss may be attributed to a larger degree of conformational freedom afforded to the Hp ring by the *two* flanking aliphatic linkers, (Urbach et al., 2002). Nonetheless, both **6** and **8** exhibit significant **X** = A, T > G, C specificity, as expected from a negative 3C-OH to G-NH_2 steric clash.

Polyamides **11** (Tn) and **12** (Th) present a sulfur atom with an sp^2 lone pair to the DNA minor groove. These compounds exhibit substantial **X** = A, T > G, C specificity ranging from ≥ 70 to ≥ 2300-fold. This remarkable selectivity may be attributed to the decreased curvature of thiazole and thiophene rings, which

forces a more intimate interaction of the large sulfur atom and the minor groove floor. In the case of **X** = G and C, this interaction is very negative, resulting in complete loss in measurable binding affinity.

Sequence Dependence of Polyamide Orientation

Approach. Because Im binds all four base pairs at a single position, it would be interesting to ask what happens if one varies simultaneously the base pairs proximal to all four Im residues. To meet this end, the plasmid pAU18 was prepared, which contains the four binding sites 5'-AAA**X**A**X**AA**X**A**X**AAA-3' (**X** = G, C, A, and T). Equilibrium association constants were determined for the complexes, and the orientation at each site was determined by affinity cleaving experiments.

DNA Binding Affinity and Sequence Specificity. Quantitative DNase I footprint titrations were performed to determine the equilibrium association constants for polyamide **2** at the designed sites on pAU18 (Figure 20). A slight preference for the **X** = G site is revealed, although all measured sites are bound with high affinity. Affinity cleavage analysis with polyamide **2** confirms a single orientation when **X** = G (Figure 20), consistent with 1:1 binding, which reverses when **X** = C, similar to observations made by Laemmli and coworkers (Janssen et al., 2000a). Although the **X** = A site lacks a DNase I footprint because of a characteristic lack of cleavage at A-tracts by DNase, the cleavage pattern of **2** is visible by affinity cleavage and is oriented to the 5' side of this binding site, similar to **X** = G. The cleavage pattern is broader than that observed at the other sites, likely because of an ensemble of slipped complexes. The **X** = T site reveals cleavage, however, on both sides of the binding site and to different extents.

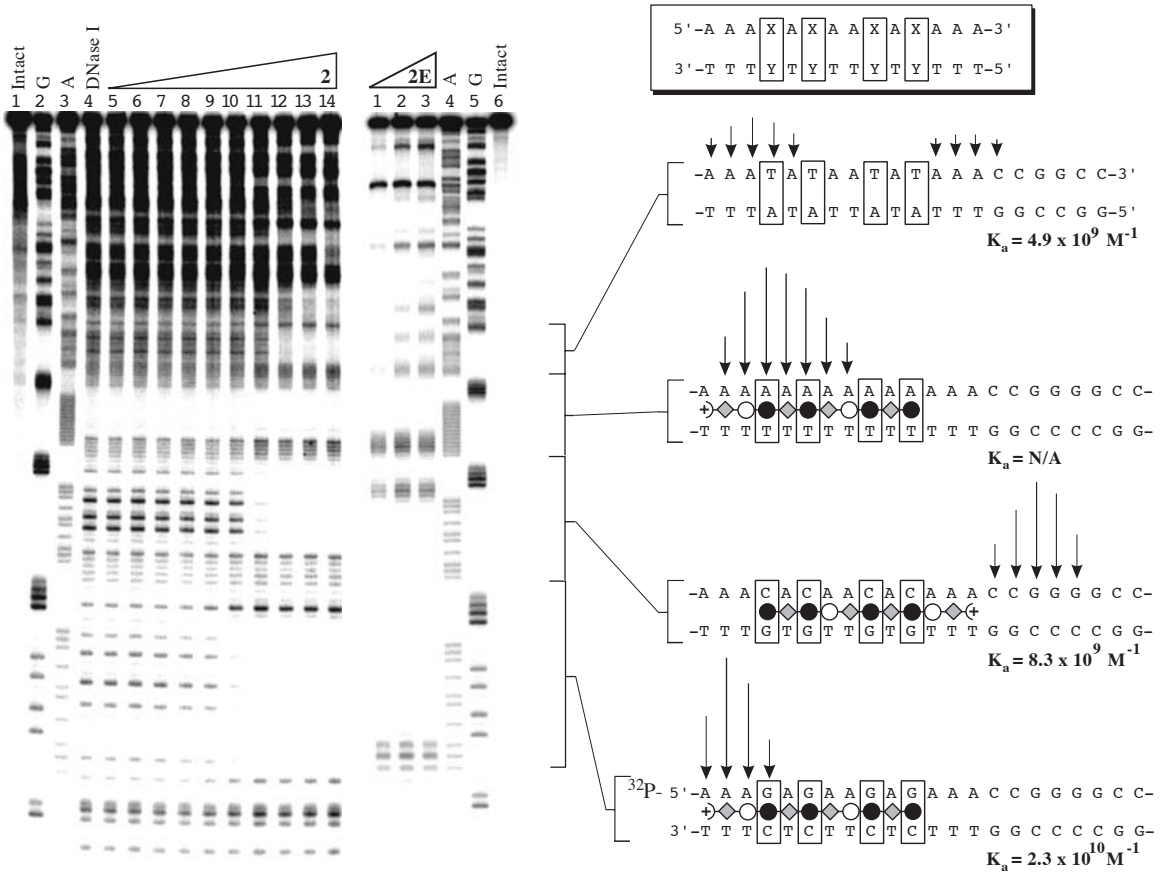


Figure 20 (Left) Quantitative DNase I footprint titration experiment for polyamide **2** on the 298 bp, 5'-end-labelled PCR product from plasmid pAU18: lane 1, intact DNA; lane 2, G reaction; lane 3, A reaction; lane 4, DNase I standard; lanes 5 – 14, 300 fM, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM **2**, respectively. (Middle) Affinity cleavage experiment with polyamide **2E** on the same PCR product of plasmid pAU18: lanes 1-3, 1 nM, 3 nM, and 10 nM **2E**, respectively; lane 4, A reaction; lane 5, G reaction; lane 6, intact DNA. (Right) Schematic illustrating the observed affinity cleavage patterns with arrows representing relative cleavage intensities. Polyamides are drawn as oriented, 1:1 complexes, as observed by affinity cleavage. Equilibrium association constants, K_a, are listed below each binding site.

Because of the similarity in DNA sequence at this site when read 5' → 3' or 3' → 5', it is possible that polyamide **2** is binding 1:1 at this site but with a slight preference for one orientation. However, due to the plasticity of 5'-TA-3' steps (Dickerson, 2001), this site likely has the capacity to accommodate two ligands, and therefore the binding mode is unclear.

Discussion. It is interesting that although a single Im residue displays no significant base preference, the multiple G to C base mutation experiment described here demonstrates that the four Im residues taken together profoundly influence the orientation preference of the polyamide with respect to G•C vs. C•G base pairs. The similarity of orientation for X = G and A suggests that the purine rich strand dominates the orientation preference, i.e., polyamide N – C aligns in the minor groove 3' – 5' with the purine-rich strand. This phenomenon is addressed more extensively in the discussion of the high-resolution NMR structure. A hypothesis is offered there to explain the G/C-dependence of polyamide orientation based on the inherent bonding geometry of an Im residue combined with the large negative propeller twisting of base pairs in the 1:1 complex.

Ligand Size Limitations in the 1:1 Motif

Approach. Determination of macromolecular structure using NMR methods requires extensive deconvolution of two-dimensional spectra. To simplify this process, the size of the molecule or complex can be reduced. However, spectral quality, and therefore the quality of the final structure, depends on a high-affinity complex. For this reason, we were interested in elucidating the effects of ligand size on binding affinity in the 1:1 motif. Using the high-affinity, eleven-residue polyamide Im-β-ImPy-β-Im-β-ImPy-β-Dp (**2**) as the parent for this study, a series of truncated polyamides, designed as analogues of **2** containing 9, 8, and 6 residues and retaining the overall Im-β-ImPy-β-type sequence (Figure 21), were prepared and examined for binding affinity in complex with the parent sequence 5'-AAAGAGAAGAG-3'.

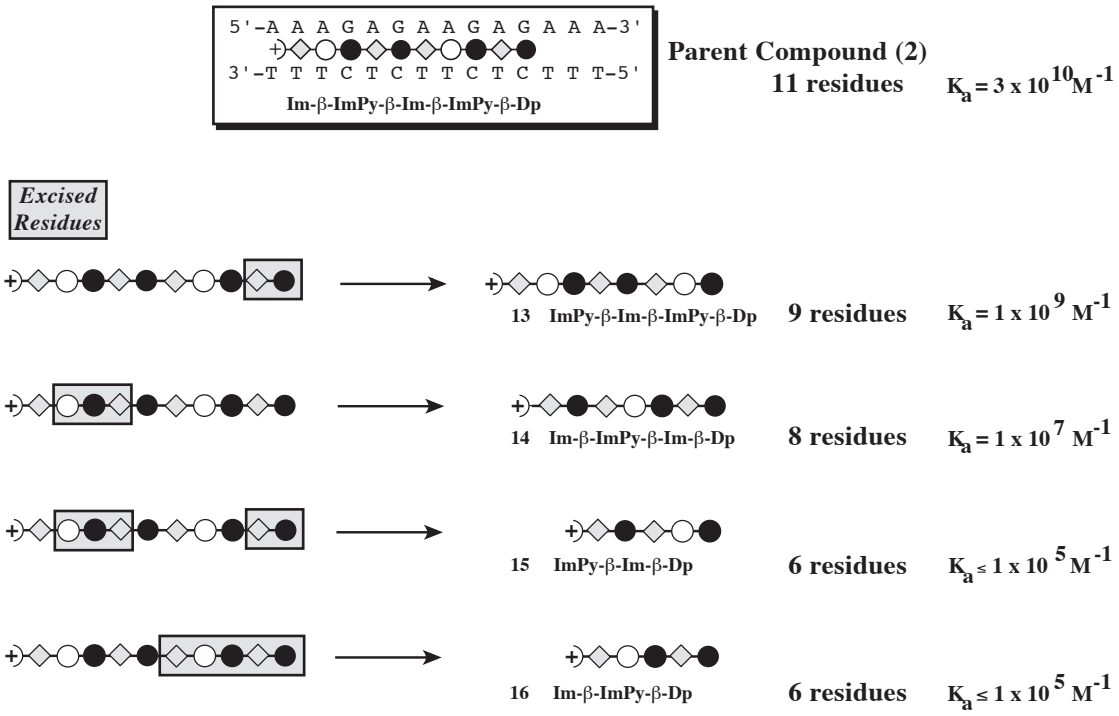


Figure 21 The effect of ligand size on binding affinity in the 1:1 motif. The 11 residue parent polyamide (**2**) is shown at the top. At left, **2** is shown with gray boxes drawn to indicate the residues that were removed in the design of polyamides **13** – **16**. The number of residues and equilibrium binding constants are given at right.

DNA Binding Affinity and Sequence Specificity. Polyamides ImPy-β-Im-β-ImPy-β-Dp (**13**), Im-β-ImPy-β-Im-β-Dp (**14**), ImPy-β-Im-β-Dp (**15**), and Im-β-ImPy-β-Dp (**16**) were prepared by standard solid phase protocols (Baird and Dervan, 1996). The equilibrium association constants for polyamides **13** – **16** against the sequence 5'-AAAGAGAAGAG-3' were determined by quantitative DNase I footprinting on the 298 base pair PCR product of pAU18 (Figure 21). It was found that the 9-residue polyamide (**13**) binds with nanomolar affinity (K_a = 10^9 M^{-1}). However, the 8-residue polyamide (**14**) binds with 100-fold reduced affinity ($K_a = 10^7 M^{-1}$). Binding affinities for the 6-residue polyamides (**15** and **16**) were too low to be determined by this method.

Discussion. These results clearly demonstrate the precipitous decrease in binding affinity as a function of ligand size. The 100-fold decrease in binding affinity from 9 to 8 residues demonstrates an important limitation of the 1:1 motif. However, since longer binding sites are required for greater specificity in a genomic context, this limitation is not so unfavorable. For the purposes of determining a high-resolution structure by NMR, the higher affinity polyamides **2** and **13** are desirable because their conformational mobility on the NMR time scale is likely to be considerably reduced as a function of their affinity. Because **13** is two residues smaller than **2**, while still retaining high affinity, **13** is the preferred target for NMR studies. The question of how far the DNA binding site can be reduced while retaining a high-affinity complex will be addressed in the next section.

NMR Structure of a 1:1 Polyamide-DNA Complex[4]

Purpose. 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. In the course of these thesis studies, several observations were made that would require structural data for explanation. In particular, we were interested in elucidating the role of β-alanine, as well as the sequence-dependent orientation of the ligand (Janssen et al., 2000a; Dervan and Urbach, 2001; Urbach and Dervan, 2001).

Approach. In order to understand the unique mode of 1:1 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. The solution structure of the polyamide ImPy-β-Im-β-ImPy-β-Dp bound to a full turn of the 13-mer DNA duplex 5'-CC**AAAGAGAAG**CG-3'•5'-CG**CTTCTCTTT**GG-3' (match site bolded) is presented here (Figure 22). Footprinting and affinity

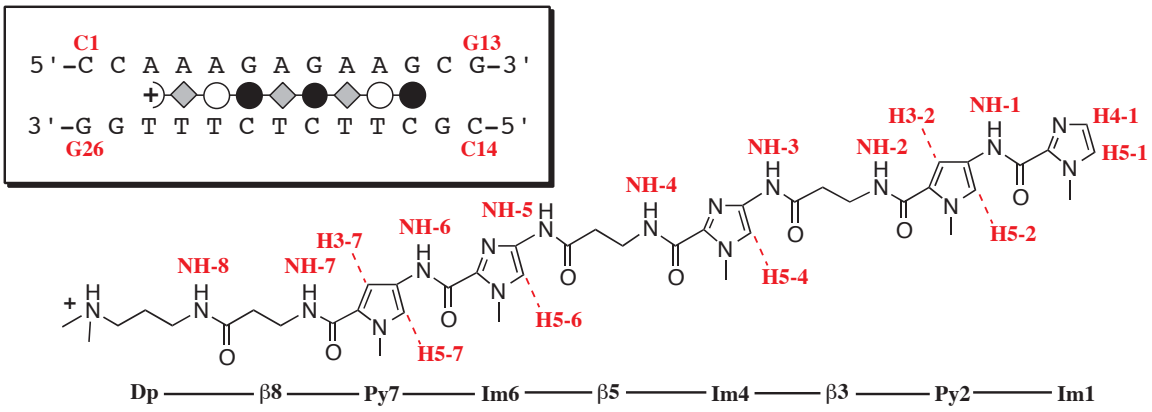


Figure 22 Nomenclature for the 1:1 complex studied by NMR. The chemical structure of polyamide **13** is shown in black with relevant protons labeled in red. Residue names are shown below the structure, linked by line segments. (inset) Polyamide binding model: DNA residue numbers are indicated in red, along with a dot model of the bound ligand.

[4] The text of this section is taken from Urbach et al., 2002.

cleavage experiments were employed to determine the minimal complex size that retains the established properties of 1:1 polyamide-DNA recognition (Janssen et al., 2000a; Urbach and Dervan, 2001). The DNA and polyamide were synthesized and purified in large quantity, and a sample of the 1:1 complex was prepared at 3.67 mM concentration (10 mM sodium phosphate, pH 7.0). Two-dimensional TOCSY, NOESY, and DQF-COSY experiments were performed, and the spectra were fully assigned. Distance constraints were derived from the NOESY spectra using hybrid matrix relaxation calculations in the MARDIGRAS program. The structural ensemble was generated and refined using AMBER 6.0 software.

DNA Binding Affinity and Ligand Orientation. To simplify the NMR spectral assignments, several truncated polyamides were screened in order to identify the minimal high-affinity ligand size. The ligand size study detailed in the previous section was the first round of this iterative process of design, synthesis, and characterization, wherein it was found that the polyamide ImPy-β-Im-β-ImPy-β-Dp (**13**) is the smallest ligand of its class that retains high affinity. In the second iteration, we were interested in determining the minimal DNA binding site for polyamide **13** that retains high affinity. To this end, three DNA binding sites of sequence type 5'-ccAAAGAGAAG$\mathbf{A}_n$cg-3' (flanking G•C clamps 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 ($\mathbf{A}_n$, **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 **13** by DNase I and methidiumpropyl-EDTA (MPE) footprinting. Affinity cleavage analysis was

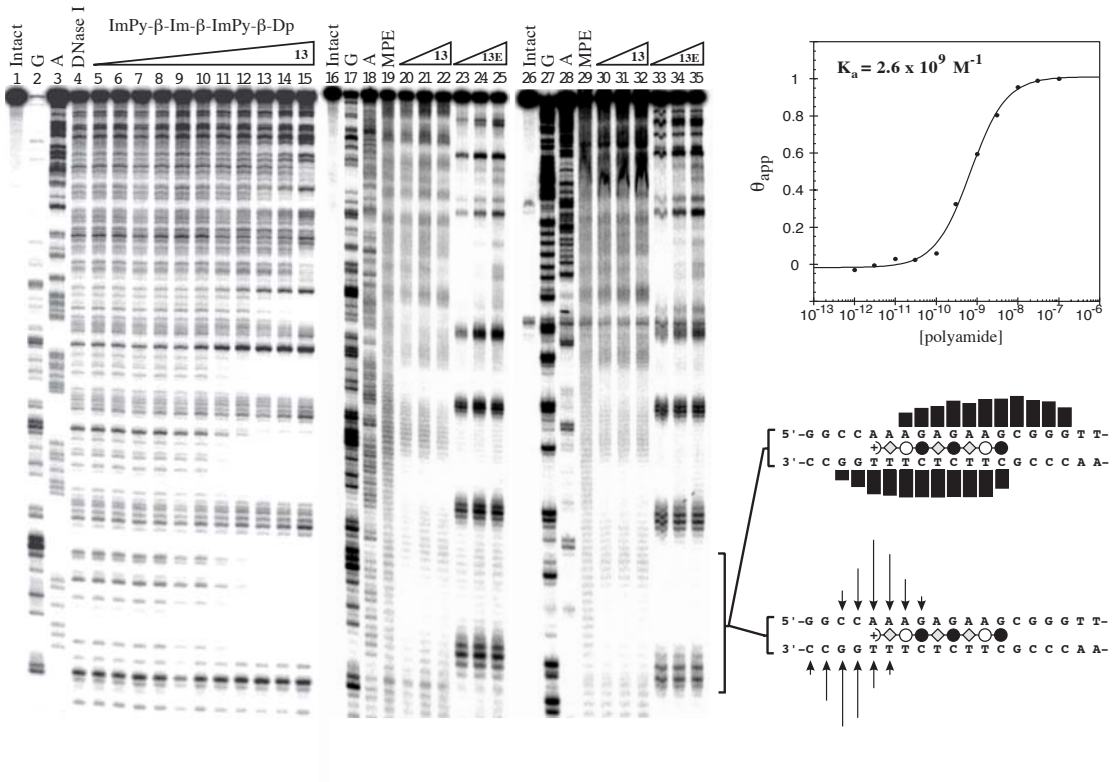


Figure 23 Biophysical characterization. (lanes 1-15) Quantitative DNase I footprint titration experiment for polyamide **13** on the 289-bp 5'-end-labeled PCR product from pAU20: lane 1, intact DNA; lane 2, G reaction; lane 3, A reaction; lane 4, DNase I standard; lanes 5-15, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, 100 nM **13**, respectively. (lanes 16-25) MPE and affinity cleavage experiments with **13** and **13E**, respectively, on the 289-bp 5'-end-labelled PCR product from pAU20: lane 16, intact DNA; lane 17, G reaction; lane 18, A reaction; lane 19, MPE standard; lanes 20-22, 10 nM, 30 nM, 100 nM **13**, respectively; lanes 23-25, 10 nM, 30 nM, 100 nM **13E**, respectively. (lanes 26-35) MPE and affinity cleavage experiments with **13** and **13E**, respectively, on the 289-bp 3'-end-labelled restriction fragment of pAU20: lane 26, intact DNA; lane 27, G reaction, lane 28, A reaction, lane 29, MPE standard; lanes 30-32, 10 nM, 30 nM, 100 nM **13**, respectively; lanes 33-35, 10 nM, 30 nM, 100 nM **13E**, respectively. (right) Analysis of the polyamide•d(CCAAAGAGAAGCG)•d(CGCTTCTCTTTGG) binding site at the bottom of the gels: (right top) Langmuir binding isotherm for the DNAse titration with the binding constant (K_a), as determined from a nonlinear least squares fit, shown in bold type; (right bottom) 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.

performed on the C-terminal EDTA conjugate (**13E**) (Figure 23). Quantitative DNase I footprinting of compound **13** on pAU20 revealed essentially equivalent binding affinities for the three sites at $K_a \sim 2 \times 10^9$ M^{-1}, indicating that the additional A•T base pairs flanking the binding site ($\mathbf{A_n}$) were not necessary to stabilize the complex. MPE and affinity cleavage experiments of compounds **13**

and **13E**, 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, which is characteristic of minor groove binding (Taylor et al., 1984). 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:DNA Stoichiometry. The titration of polyamide **13** to the NMR sample of d(CCAAAGAGAAGCG) • d(CGCTTCTCTTTGG) is shown in Figure 24. Chemical shift perturbation in the uncrowded DNA imino region upon polyamide addition 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-DNA stoichiometry. Only one set of

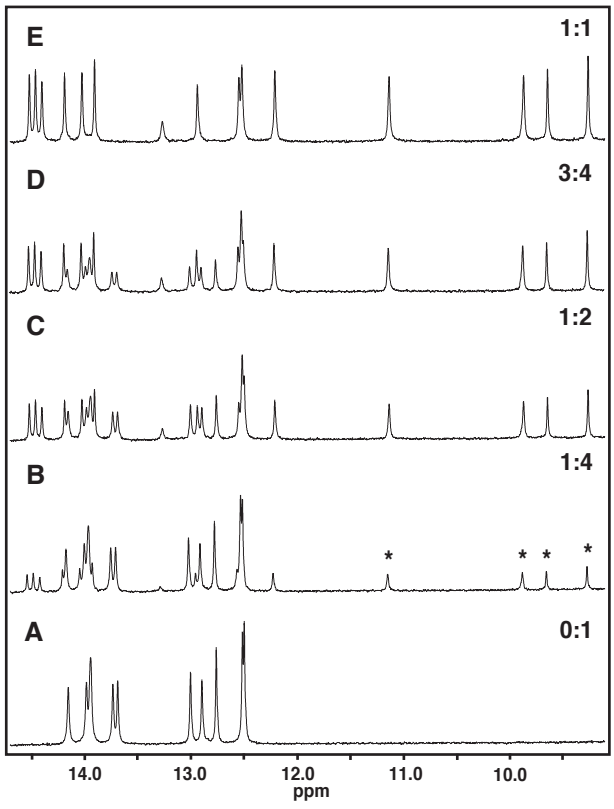


Figure 24 Far downfield region (9.0 – 14.5 ppm) of the ^{1}H NMR (at 600 MHz, 25 °C) spectra in 9:1 D_2O:H_2O of d(C C A A A G A G A A G C G) • d(C G C T T C T C T T T G G) with (A) no ligand added; (B) 1:4 ligand:DNA; (C) 1:2 ligand:DNA; (D) 3:4 ligand:DNA; and (E) 1:1 ligand:DNA stoichiometry. Molar ratios are indicated at right. Ligand aromatic amide NH protons are indicated by asterisks. Peaks at left are from thymine and guanine imino protons.

polyamide signals is observed, as indicated with asterisks in Figure 24, which further confirms a 1:1 ligand:DNA stoichiometry. Terminal base pair imino resonances were not observed.

Spectral Assignments. Methods for assigning protons in polyamide•DNA complexes have been well established (Pelton and Wemmer, 1988; Pelton and Wemmer, 1989; Mrksich et al., 1992; Dwyer et al., 1992; Dwyer et al., 1993; Geierstanger et al., 1996). DNA chemical shifts for the 1:1 complex are listed in Table 4. The intramolecular DNA crosspeaks were assigned as previously described (Wütrich, 1986). 16 pairs of H5'/H5" protons were assigned nonstereospecifically. All other nonexchangeable protons were

Table 4. DNA Chemical Shift Assignments for the 1:1 Polyamide-DNA Complex.*

	H6/H8	CH5/A2H	H1'	H2'	H2"	H3'	H4'	GH1/TH3	GH2	CH41	CH42
Strand 1											
C1	7.72	5.95	5.98	2.03	2.48	4.66	4.09	-----	-----	6.93	7.91
C2	7.47	5.74	5.28	1.83	2.13	4.78	4.01	-----	-----	6.81	8.73
A3	8.25	7.41	5.56	2.76	2.76	5.01	4.32	-----	-----	-----	-----
A4	8.19	7.31	5.48	2.76	2.71	5.05	4.35	-----	-----	-----	-----
A5	8.08	7.66	5.94	2.63	2.71	4.98	4.15	-----	-----	-----	-----
G6	7.13	-----	5.11	1.93	2.54	4.72	3.28	12.54	7.76	-----	-----
A7	7.62	8.04	5.56	2.06	2.63	4.74	3.04	-----	-----	-----	-----
G8	7.43	-----	5.42	2.41	2.48	4.90	4.03	12.23	6.58	-----	-----
A9	7.93	7.61	5.40	2.19	2.70	4.91	3.82	-----	-----	-----	-----
A10	7.91	7.94	6.04	2.56	2.73	5.02	4.27	-----	-----	-----	-----
G11	7.20	-----	5.39	2.03	2.15	4.63	3.52	12.57	8.12	-----	-----
C12	6.96	4.91	4.85	1.41	1.83	4.62	3.28	-----	-----	6.11	8.16
G13	7.87	-----	6.18	2.36	2.64	4.69	4.20	not ob.	not ob.	-----	-----
Strand 2											
C14	7.58	5.89	5.76	1.70	2.33	4.67	4.04	-----	-----	7.02	8.35
G15	7.92	-----	5.96	2.68	2.79	4.99	4.38	13.28	7.23	-----	-----
C16	7.44	5.39	6.04	2.06	2.61	4.67	4.35	-----	-----	6.42	7.96
T17	7.26	1.52	5.85	1.78	2.50	4.65	3.45	14.53	-----	-----	-----
T18	7.17	1.57	5.76	1.91	2.49	4.67	3.40	14.47	-----	-----	-----
C19	7.53	5.53	6.17	2.15	2.43	4.88	3.99	-----	-----	6.69	8.25
T20	7.13	1.52	5.44	1.91	2.37	4.70	3.41	14.41	-----	-----	-----
C21	7.52	5.44	6.10	2.15	2.45	4.71	4.06	-----	-----	6.56	7.86
T22	7.16	1.48	5.68	1.93	2.44	4.68	3.04	14.19	-----	-----	-----
T23	7.26	1.70	5.56	1.78	2.29	4.62	3.08	13.93	-----	-----	-----
T24	7.20	1.69	5.63	1.88	2.13	4.86	3.94	14.03	-----	-----	-----
G25	7.90	-----	5.55	2.70	2.70	4.97	4.33	12.96	6.81	-----	-----
G26	7.85	-----	6.19	2.39	2.54	4.65	4.20	not ob.	not ob.	-----	-----

* Nonlabile proton chemical shifts are from the NOESY spectrum in D20 (75 ms mixing). Labile proton chemical shifts are from the NOESY spectrum in H20 (100 ms mixing).

Table 5. Polyamide Chemical Shift Assignments.*

Residue:	**Im-1**	**Py-2**	**Beta-3**	**Im-4**	**Beta-5**	**Im-6**	**Py-7**	**Beta-8**		**Dp-9**
H4-1	7.21									
H3	-----	6.77	-----	-----	-----	-----	6.46	-----		
H5	7.33	7.77	-----	7.62	-----	7.73	7.55	-----		
NCH-3	4.12	4.03	-----	4.12	-----	4.14	3.94	-----		
		C12-H1	2.91	C20-H1	2.56		C34-H1	2.42	C37-H1	3.12
methylene		C12-H2	2.94	C20-H2	3.04		C34-H2	2.65	C37-H2	3.38
protons		C13-H1	3.70	C21-H1	3.61		C35-H1	3.43	C38-H1	1.94
		C13-H2	4.01	C21-H2	4.13		C35-H2	3.53	C38-H2	1.94
									C39-H1	3.12
									C39-H2	3.18
									Dimethyl	2.90
Amides	NH-1	NH-2	NH-3	NH-4	NH-5	NH-6	NH-7	NH-8		
	9.65	8.24	9.87	8.35	11.14	9.26	8.08	8.66		

* Nonlabile proton chemical shifts are from the NOESY spectrum in D20 (75 ms mixing). Labile proton chemical shifts are from the NOESY spectrum in H20 (100 ms mixing).

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 noteworthy to point 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 (Nadeau and Crothers, 1989).

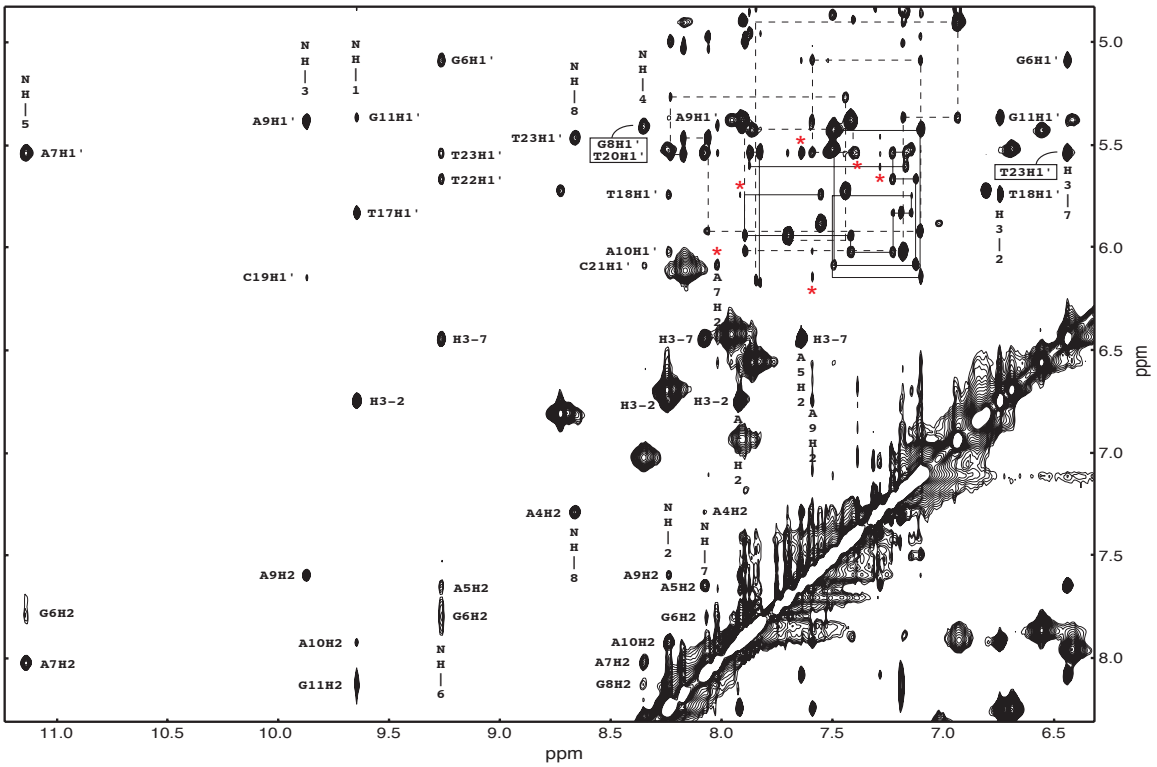


Figure 25 Expansion of the aromatic and amide region of the NOESY spectrum of the 1:1 polyamide • d(CCAAAGAGAAGCG) • d(GCGTTCTCTTTGG) complex (9:1 H_2O:D_2O, 10 mM sodium phosphate, pH = 7.0; 25 °C; 75 ms mixing). Sequential aromatic to H1' connectivities for the purine-rich strand are shown as dotted lines; those for the pyrimidine-rich strand as solid 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 residue name, residue number, proton name (e.g. A7H2 = Adenine 7, H2 proton); ligand is named as proton name, 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* distance, Nadeau and Crothers, 1989).

Nomenclature for polyamide protons is shown in Figure 22. All polyamide protons were assigned, and their chemical shifts are listed in Table 5. Amide NH, Py-H3, and Im-H4 protons were assigned based on intra- and inter-molecular 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 nonstereospecifically based on the patterns observed in the DQF-COSY

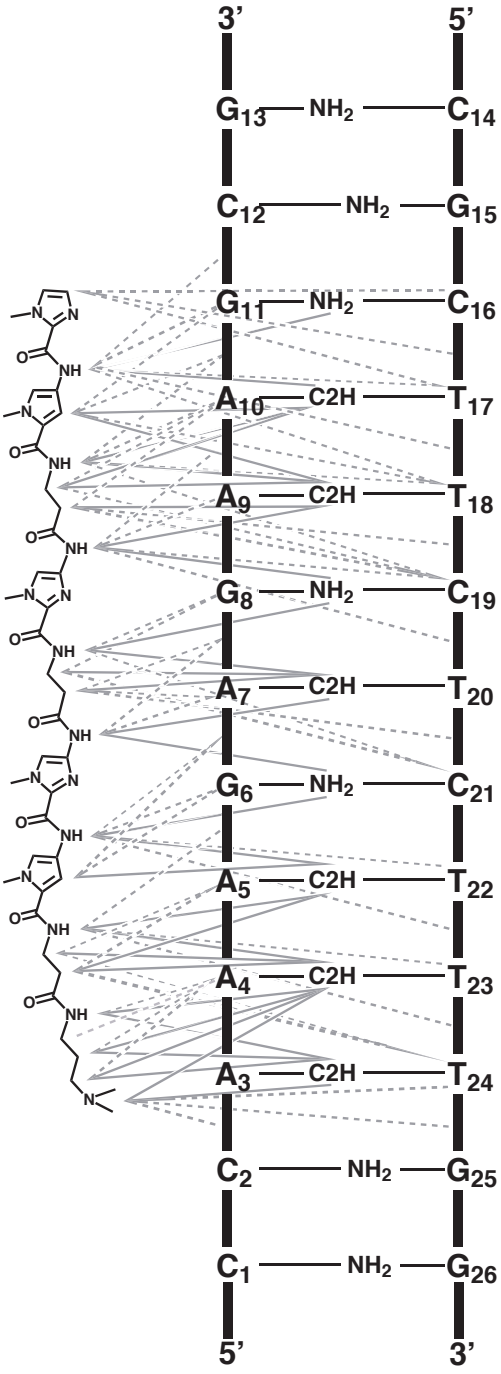


Figure 26 Model of the 1:1 complex showing observed intermolecular NOEs from the polyamide to the floor and walls of the minor groove as solid and dashed gray lines, respectively. The DNA is shown as a ladder with residue numbers at each rung. Lines to a residue number show a contact to H1'; lines ending between residue numbers indicate H4' and/or H5'/H5" protons.

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. 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 shifts are similar to those previously reported by Wemmer and coworkers for 1:1 and 2:1 complexes (Mrksich et al., 1992; Geierstanger et al., 1996; de Clairac et al., 1999).

The NOESY spectrum of the 1:1 complex in H_2O is shown in Figure 25. 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. Selected intermolecular contacts are illustrated in Figure 26, with solid and dashed lines indicating contacts to the floor and walls of the minor groove, respectively. A homogeneous distribution of contacts is observed, which unambiguously defines the position of the ligand within the minor groove. The NOESY and spectra are provided in Appendix A, with each region expanded and annotated.

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 D_2O, with an additional 115 distance constraints from the NOESY spectrum in H_2O. Methods for converting NOE intensities to upper bound distance constraints are detailed in the Experimental section. Forty Watson-Crick hydrogen bond constraints were applied on the basis of 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 provided in Appendix B and is available from the Brookhaven Data Bank under accession code 1LEJ.

Structure Calculations. Following the approach of Chazin and coworkers, a significant effort was made during the structure calculations to

sample and represent conformational space consistent with the input data (Eis et al., 1997; Schnell et al., 1999). 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; Macke and Case, http://www.scripps.edu/case). 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 (Kollman et al., 2000). The docked structures converged to an RMSD of 1.37 Å, and subsequent rMD SA did not improve the total 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 statistics for the structure calculations is given in Table 6. These statistics demonstrate excellent agreement with the input data (low violation energy) as well as nonbonding stabilization of the structure due to ligand binding (large negative Leonard Jones energy).

Table 6. Statistics for the Final Structural Ensemble of the Polyamide•DNA Complex

E_{total}	-1407.4 (±17.0)[a]
$E_{Leonard\ Jones}$	-583.7 (±14.7)[a]
$E_{violation}$	0.9 (±0.2)[a]
NOE Violations ≥ 0.13 Å	0[b]
NOE Violations ≥ 0.10 Å	0.42[b]
Progression of the NMR Refinement (RMSD, Å)[c]	
40 Starting DNA Structures	3.46
40 Docked Complexes	1.37
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.

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 23) fits an n=1 Hill equation. The 1D NMR titration (Figure 24) 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 (Pelton and Wemmer, 1989; Mrksich et al., 1992; de Clairac et al., 1999). 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 (Pelton and Wemmer, 1989; Mrksich et al., 1992; de Clairac et al., 1999).

The affinity cleavage and MPE footprinting experiments (Figure 23) 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 (Taylor et al., 1984). The existence of cleavage on only one side of the binding site, even when the [polyamide] >> [DNA], is evidence for a single binding orientation. The pattern of intermolecular contacts observed in the NOESY spectra (Figures 25 and 26) confirms the 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.

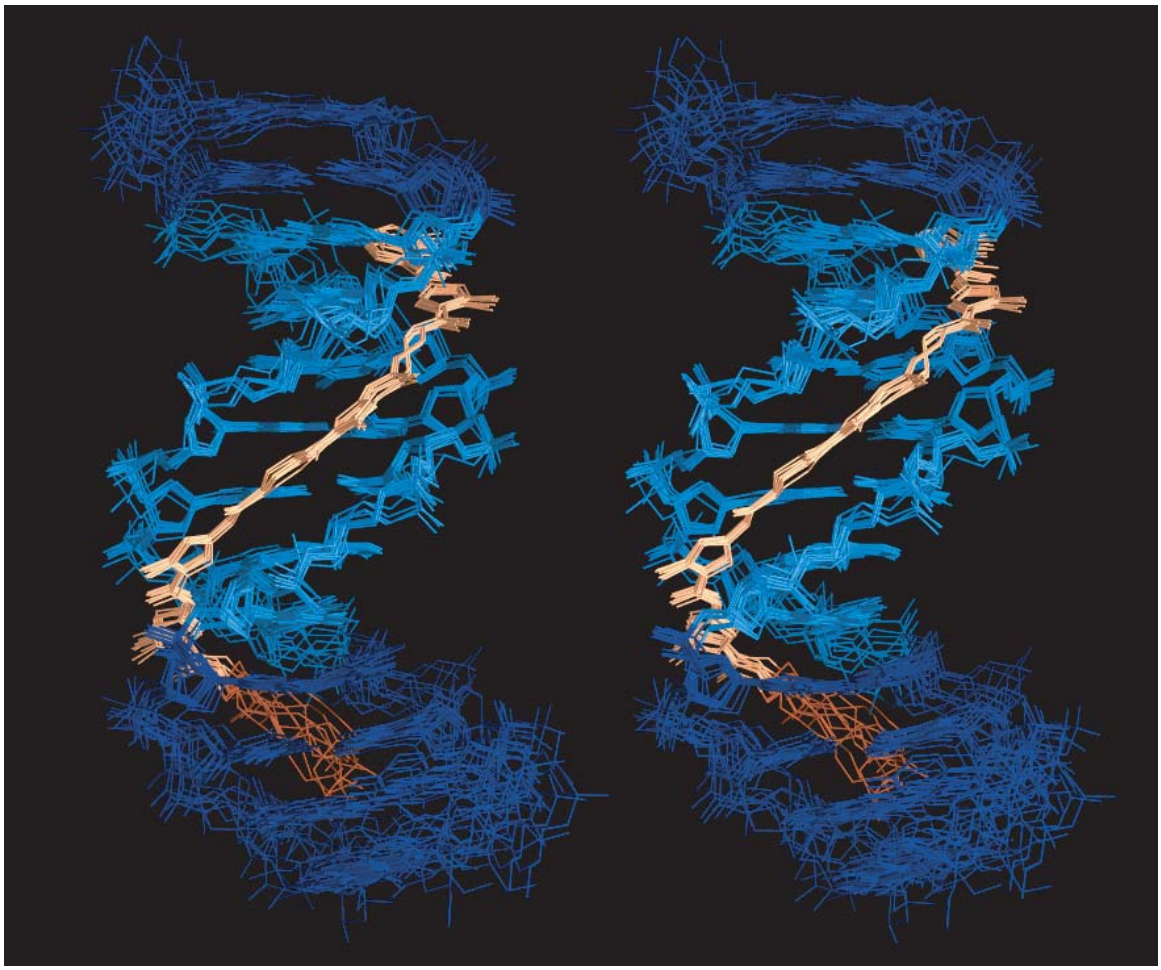


Figure 27 Stereo view of the final ensemble of 12 structures superimposed over the binding site. DNA is shown in blue, ligand in gold. The binding site is shown as a lighter color in the center of the complex.

Characterization of the Complex. The final ensemble of 12 structures is shown in Figure 27 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 6). 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 more poorly defined. 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 previously observed for A-tract structures (Celda et al., 1989), with an average phase angle value of 148°. Helical parameters are listed in Appendix C.

Minor Groove Width and Propeller Twist. Figure 28 shows plots of minor groove width and propeller twist for the complex, which were calculated using the CURVES program, as described by Lavery and coworkers (Lavery and Sklenar, 1988; Stofer and Lavery, 1994). Overall, the complex displays a narrow minor groove and a large negative propeller twist, which are features typically associated with A-tract structures (Crothers and Shakked, 2001). The narrow minor groove is confirmed by the observation of medium intensity cross-strand NOEs (type *a* distance, Nadeau and Crothers, 1989) between each adenine H2 and the H1' of its base-paired thymidine's 3'-neighbor (red stars in Figure 25). The minor groove in the structural ensemble is much narrower than observed for ligand-free duplex DNA containing similar sequences – 5'-AAAGAA-3' by NMR (MacDonald et al., 2001) as well as 5'-AAAGAAAA-3' and short A-tracts by x-ray crystallography (Han et al., 1997; Shatzky-Schwartz et al., 1997). However, it has

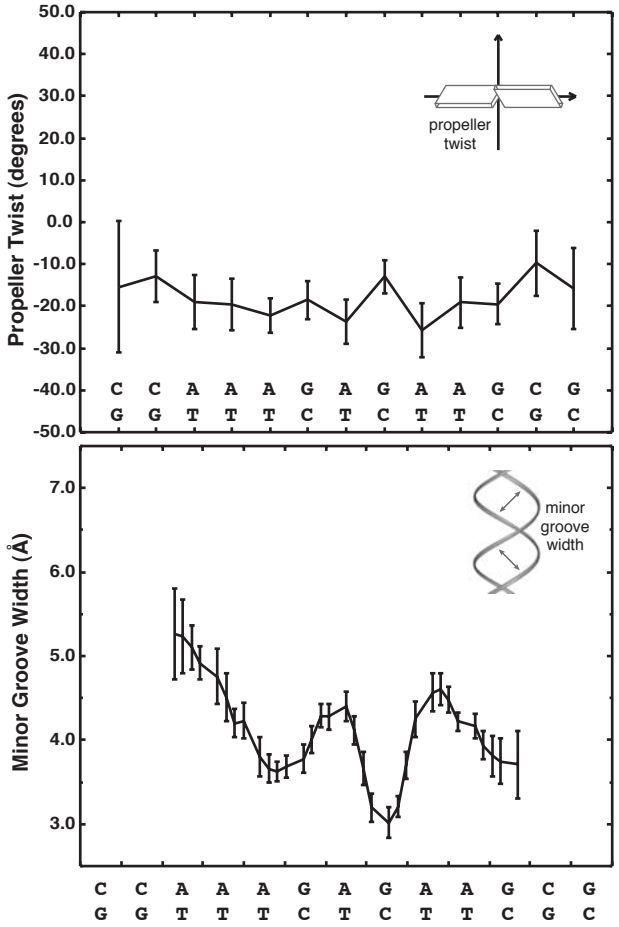


Figure 28 Plots of average propeller twist (top) and minor groove width (bottom) at each DNA residue, determined using the CURVES program as described by Lavery and coworkers (Lavery and Sklenar, 1988; Stofer and Lavery, 1994). Average values are connected by a solid line, and the y-axis error bars indicate 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.

been proposed that ligand binding in a 1:1 mode can induce the walls of the minor groove to close down in order to maximize van der Waals contacts (Dickerson, 2001; Bostock-Smith et al., 2001). Large negative propeller twist is commonly associated with a narrow minor groove. Lu and coworkers observe a significant decrease in propeller twist upon interruption of an A-tract with guanine, and they attribute this to a disruption in the spine of hydration by the G-NH_2 group (MacDonald et al., 2001). We observe a consistently large degree of propeller twist throughout the complex, even though there are multiple guanines interrupting the short A-tracts. Therefore, this effect is likely to be stabilized by the ligand, as discussed below. Helical parameters are provided in Appendix C.

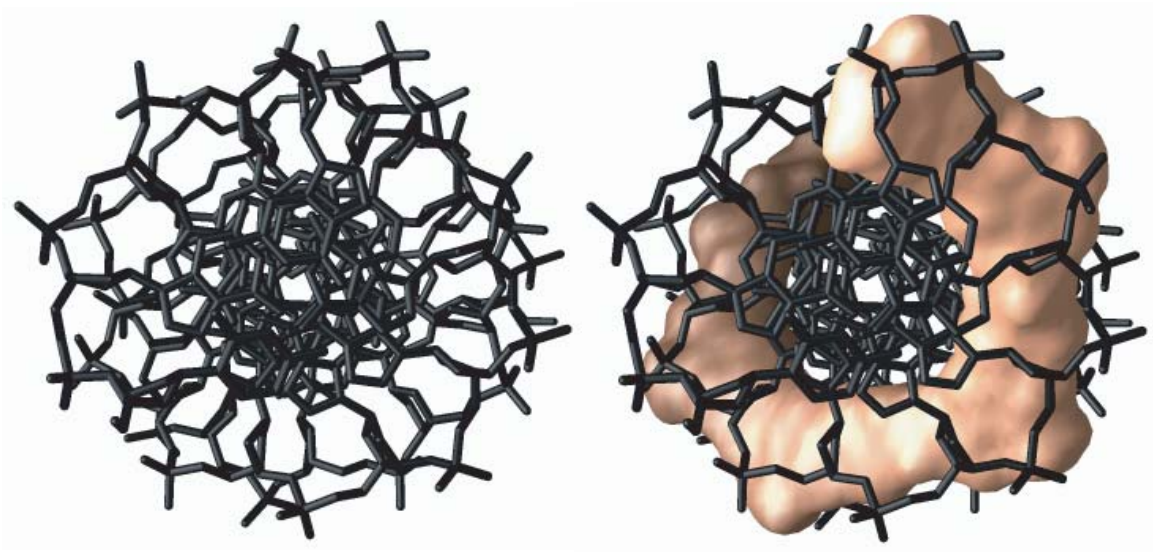


Figure 29 Top view of the polyamide•DNA mean structure with (left) and without (right) the ligand. The DNA is shown as a stick model in black, and the ligand is rendered with a molecular surface in gold.

Ligand Structure. Figure 29 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-**β3**-Im-**β5**-ImPy – see Figure 22). The structural ensemble reveals a single orientation for β5. However, equal populations of two binding modes are observed for β3, which adopts a straight conformation and a bent conformation with virtually identical AMBER energies. We were chiefly interested in assessing the impact of

these two conformations on DNA recognition. Therefore, we averaged the coordinates for the 5 bent and 7 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 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 β5 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 (Dickerson and Drew, 1981). Netropsin (Kopka et al., 1985) and distamycin (Coll et al., 1987) have been observed to displace 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 30 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 (± standard deviation) distances given in the schematic at right. Amide NH groups to the 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

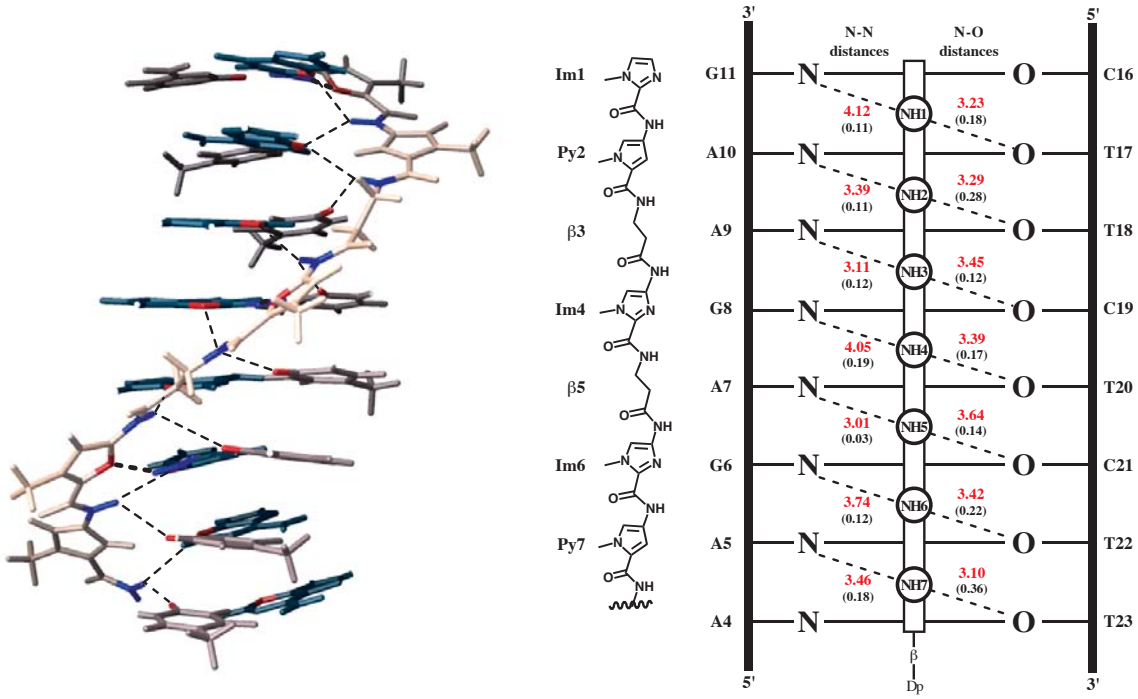


Figure 30 Schematic of polyamide NH to purine N3 and pyrimidine O2 contacts for DNA residues A4-G11•C16-T23 and polyamide residues Im1-Py7. The figure has three parts, which are all in vertical register. (left) Mean structure of the polyamide•DNA complex showing the base pairs as a stack down the center 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 dashed 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 imbedded circles containing the amide NH number, as defined in Figure 22. Dashed 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 dashed lines, with standard deviations in parentheses.

bond to G-NH_2. The distance values are roughly equal from the left to 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 30), 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 part of 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 (Kopka et al., 1985; Lown et al., 1986). 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 (Goodsell et al., 1995; Kopka et al., 1997). However, Im does not bind across from G in these structures, and therefore the key Im-N3 to G-NH_2 hydrogen bond in 1:1 binding was not verified by a high-resolution structure. Considerable x-ray and NMR structural data exists to support the formation of Im-N3 to G-NH_2 hydrogen bonds for the Im/Py pair in the 2:1 motif (Kielkopf et al., 1998a; Mrksich et al., 1992; Dwyer et al., 1992; Geierstanger et al., 1994). In the 1:1 complex presented here, we find a specific oriented hydrogen bond from Im-N3 to G-NH_2. In fact, the ImPy sections of the polyamide presented here are structurally similar to the compounds originally synthesized by Lown and coworkers (Lown et al., 1986; Goodsell et al., 1995). Figure 31 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-NH_2 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 package (Kollman et al., 2000). 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

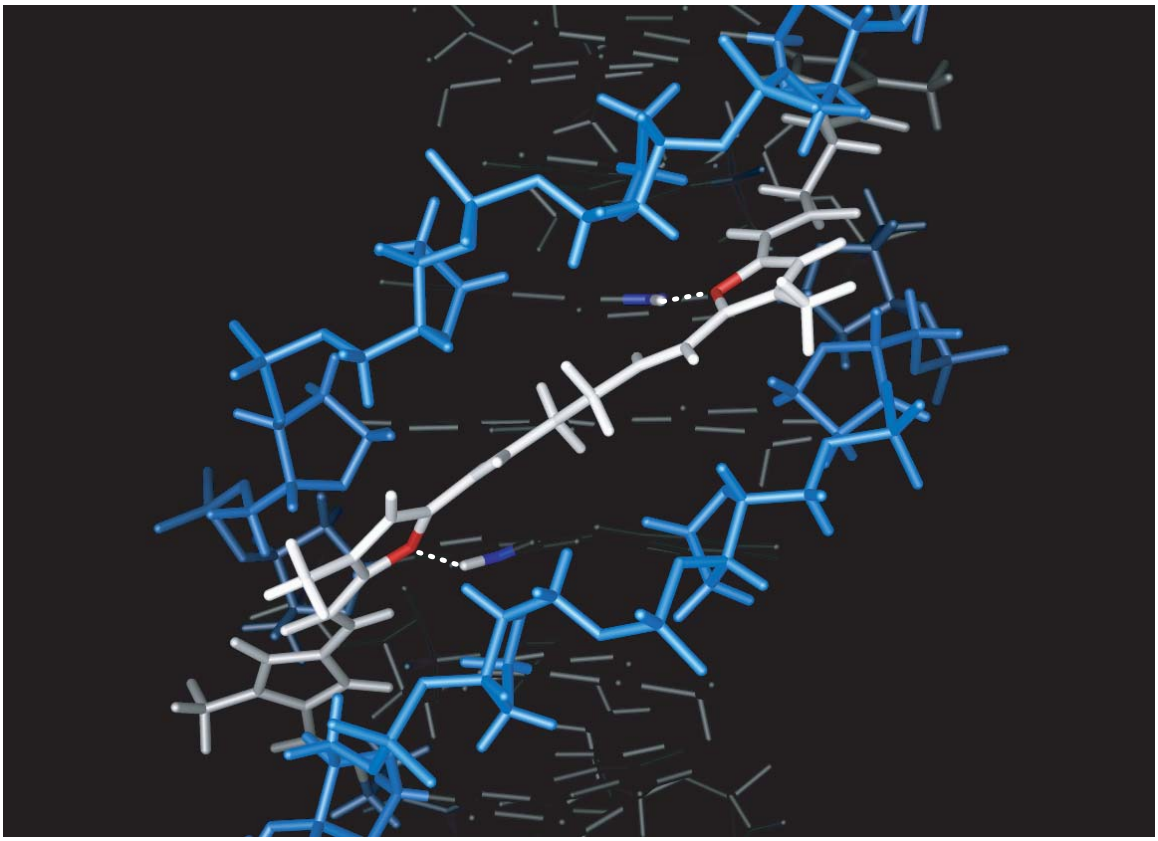


Figure 31 View into the minor groove of the mean structure, with polyamide residues Im6–β5–Im4 proximal to DNA residues G6–A7–G8, respectively. The DNA bases are in black; the backbone is in blue; the polyamide is in white. Hydrogen bonds between imidazole N3 and guanine NH2 are shown by dotted lines connecting the donor (blue) and acceptor (red) pairs.

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 previously 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 (Dervan and Urbach, 2001). 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 completely eliminated specific binding. In accordance with β/Py mutation studies in the 2:1 motif, it is likely that the β is needed to reset the register for the following (C-terminal) residue (Turner et al., 1998). It appears that β 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 dihedrals between the plane of Im4 and the planes of Py2 (Im–β–Py) and Im6 (Im–β–Im). Figure 32 shows axial views of these dihedrals with average values (± 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** dihedrals (measured for bolded residues) observed in 2:1 crystal structures (Kielkopf et al., 1998a; 1998b). The Im/Im dihedral for Im–β–Im subunit is a significant 17°

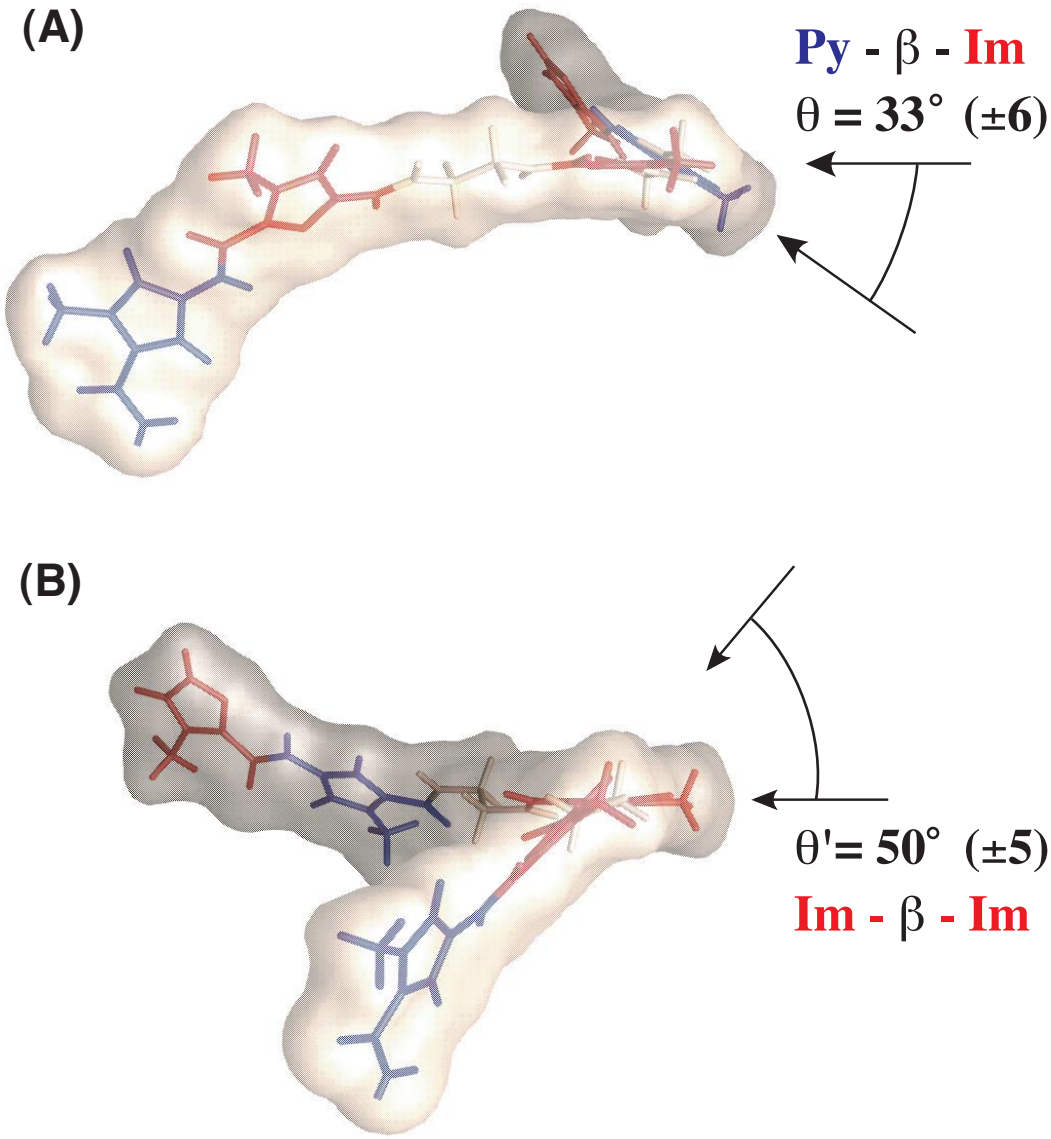


Figure 32 Polyamide ring–β–ring dihedrals. The polyamide from the mean structure is shown as a stick model with a transparent molecular surface. Imidazoles are in red, with β-alanines in white and pyrroles in blue. (A) View down the dihedral axis between Py2 and Im4 (Py–β–Im). (B) View down the dihedral 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 (± standard deviation) over the 12 final structures.

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 orientation in order to maximize Im-N3 to G-NH_2 hydrogen bonding. It is interesting to point out that only one β-alanine conformation is observed in Im4–β5–Im6 vs. two degenerate conformers for Py2–β3–Im4. If indeed the two β3 conformers are real, it is conceivable that β5 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.

Footprinting data provided earlier in this thesis has established a partial recognition code for the 1:1 motif, 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 (de Clairac et al., 1999). The structural ensemble shows these protons to be directly in contact, with internuclear 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.

The Sequence-Dependence of Ligand Orientation. Laemmli and coworkers 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

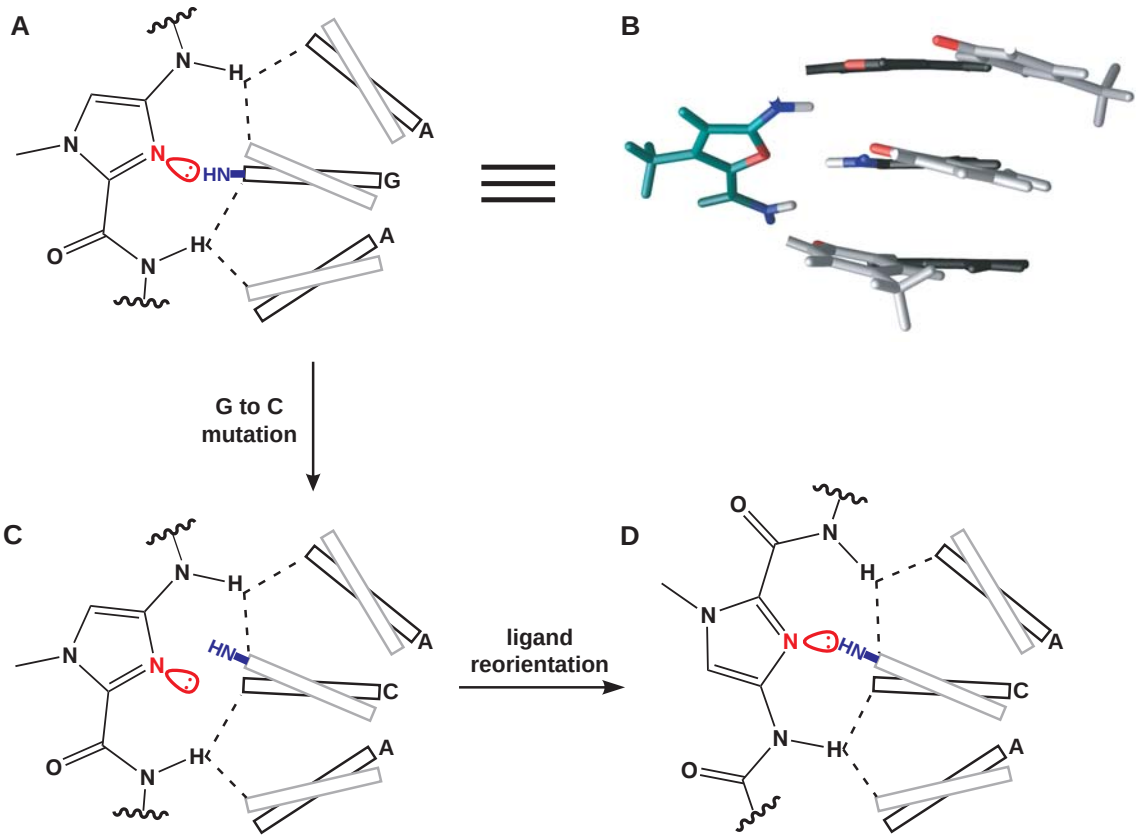


Figure 33 G/C-Dependence of polyamide orientation. (A) Schematic illustrating the optimal alignment of polyamide NH groups with the DNA bases and showing that the inherent geometry of the Im residue coupled with the propeller twist of the G/C base pair allows overlap between Im-N3 and G-NH_2 groups. Propeller twisted base pairs are shown as gray boxes (pyrimidines) crossing over black boxes (purines), with purine labels at right. The Im-N3 sp^2 orbital is shown as its lone pair in red. Guanine's NH_2 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-NH_2. Therefore, the ligand binds in the opposite orientation (D) in order to restore this interaction.

5' – AAGAGAAGAG – 3', but in the opposite C – N oritation 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 (Janssen et al., 2000; Urbach and Dervan, 2001). Based on the structure presented here, 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-NH_2 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 33 details this model, showing a section of the NMR structure (part B) with a schematic of this orientation (part A). When the G residues are mutated to C, the lone pair of each Im-N3 will not overlap with G-NH_2 (part C) unless the ligand orients in the opposite direction (part D). This model could apply to the orientation preference at poly A (Urbach and Dervan, 2001) by considering the partial positive potential presented by A-H2, similar to G-NH_2 but weaker, which could interact favorably with Im-N3.

Summary. 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 though 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 NH_2 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-NH_2. The Im – β – Im subunit requires a large dihedral 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 polyamides that can discriminate between 1:1 and 2:1 binding modes.

Linker-Dependent Conformational Control of Polyamide-DNA Binding Modes[5]

Purpose. It has been demonstrated that polyamides of type Im-β-ImPy can bind in 1:1 and 2:1 modes depending on rules for recognition inherent to each motif (Dervan and Urbach, 2001). This ambiguity of sequence targeting depending on stoichiometry has posed a serious design problem for the DNA recognition field. If a polyamide is to be effectively sequence selective in a genomic context, it will be necessary to control its binding mode. For example, the polyamide of sequence Im-β-ImPy-**β**-Im-β-ImPy-β-Dp (**2**) (Dp = dimethylaminopropylamide) targets 5'-AAAGAGAAGAG-3' as a 1:1 complex and 5'-TAGCGCAGCGCTA-3' as a 2:1 complex. The γ-linked analogue, Im-β-ImPy-**γ**-Im-β-ImPy-β-Dp (**17**), binds as a hairpin with high affinity to its target sequence, 5'-TAGCGCT-3' (Turner et al., 1998). Given the structural similarity between **2** and **17**, i.e., one methylene unit difference, it was a simple prediction that **17** would bind in a 1:1 mode. We examined this possibility and found that, indeed, **17** binds in both hairpin and 1:1 conformations (Figure 34). Therefore, an accurate prediction of the target DNA sequence would require the ability to control the ligand conformation. Here we ask the question whether polyamide conformation, and therefore the polyamide•DNA binding mode, can be controlled by tailoring the linkage between subunits.

Approach. The concept of linker-controlled binding modes is a familiar one. Trauger et al. demonstrated that by increasing the aliphatic linker length from β to γ, the hairpin binding mode is favored over the extended 2:1 motif

[5] The text of this section is taken from Urbach et al., in preparation.

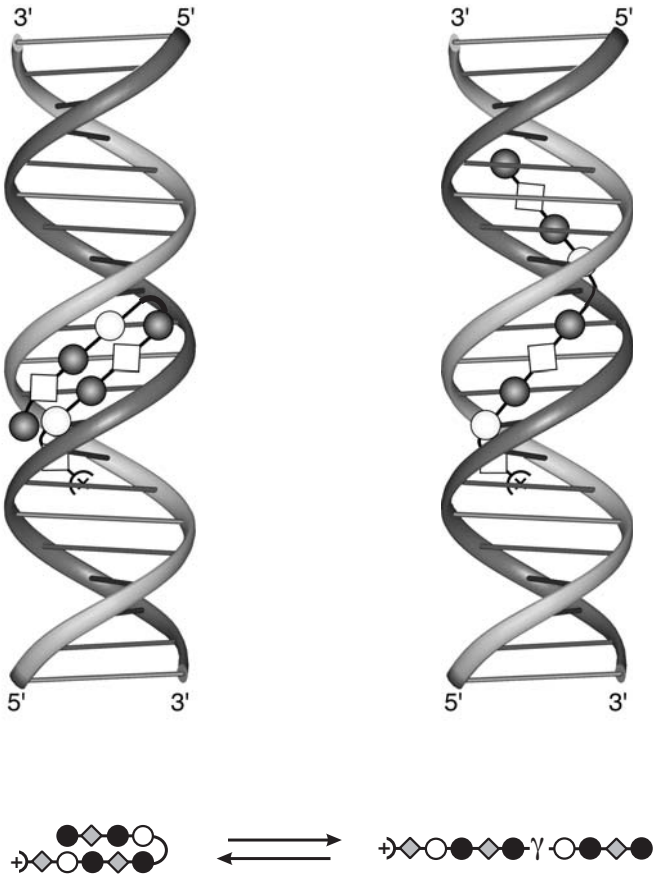


Figure 34 Equilibrium between hairpin (left) and extended (right) conformational binding modes for compound **17**. The cartoons at top represent DNA with polyamide bound in the respective conformation. DNA is shown as a ladder. Polyamides are shown as dot models, with shaded and nonshaded circles representing imidazole and pyrrole, respectively, and gray diamonds indicating beta-alanine. The γ-turn residue is shown both as a semicircle connecting the two subunits and as the symbol, γ.

(Trauger et al., 1996b). Wemmer and coworkers showed by NMR that β-linked polyamides can adopt a hairpin conformation, although the hairpin conformation is more accessible by the γ linker (de Clairac et at., 1997). During the course of this study, Boger and coworkers reported that the turn conformation of a β linker can be reinforced by substituting the α-(R)-proton of β with -OCH_3 (Woods et al., 2002). Herman et al. demonstrated the use of α-(R)-amino-substituted γ (the Herman turn, $^{H_2N}\gamma$) to increase the binding affinity of hairpin polyamides (Herman et al., 1998). We hypothesized that the Herman turn could be used to disfavor the extended 1:1 binding mode, while favoring the

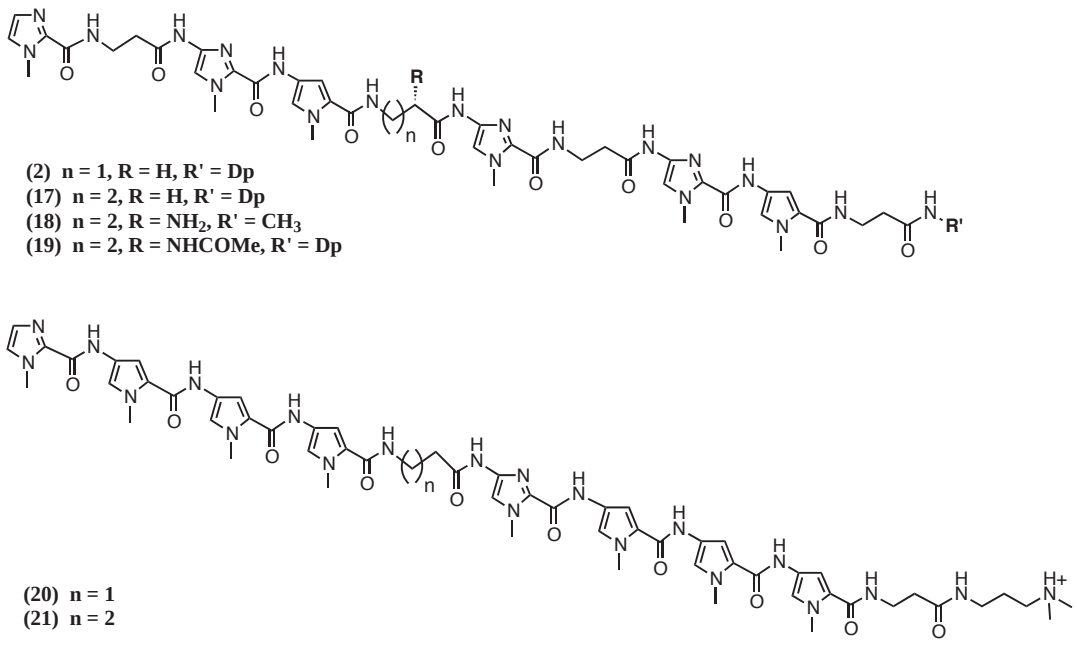


Figure 35 Chemical structures of polyamides **2**, **17**, **18**, **19**, **20**, and **21**.

hairpin mode. Accordingly, the polyamide Im-β-ImPy-^{H_2N}**γ**-Im-β-ImPy-β-Dp (**18**) and its acetamide analogue, Im-β-ImPy-^{Ac}**γ**-Im-β-ImPy-β-Dp (**19**), were prepared (Figure 35). In addition, we were interested in studying the conformational effects of linkage on eight-ring polyamides, such as ImPyPyPy-**β**-ImPyPyPy-β-Dp (**20**) and ImPyPyPy-**γ**-ImPyPyPy-β-Dp (**21**). As described above, the β-rich series (**2**, **17**, **18**, and **19**) targets hairpin and 1:1 binding sites, 5'-TAGCGCT-3' and 5'-AAAGAGAAGAG-3', respectively. Polyamide **21** is known to bind as a hairpin with subnanomolar affinity to the sequence 5'-TAGTACT-3' (Trauger et al., 1996b). The sequence 5'-AAAAAGAAAAG-3' was designed for extended 1:1 binding of polyamides **20** and **21** based on rules for 1:1 recognition (Urbach and Dervan, 2001). For simplicity, a plasmid containing all four binding sites was

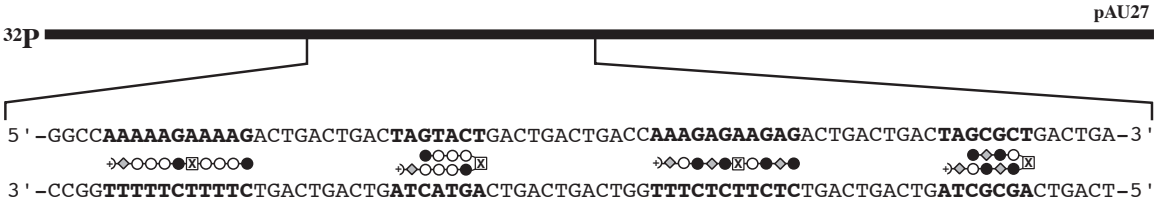


Figure 36 The designed insert cloned into plasmid pAU27. The targeted recognition sites are shown in bold type. Two sets of polyamides are shown as dot models. The two at left represent compounds **20** and **21** bound in the predicted 1:1 and hairpin conformations. The two at right represent compounds **2**, **17**, **18**, and **18** bound in the predicted 1:1 and hairpin conformations. The variable linker position is shown as a square containing the letter X.

constructed (Figure 36). Equilibrium association constants and binding site sizes for these compounds at their target sites were determined in order to derive the relative preference of each compound for hairpin versus 1:1 binding modes.

DNA Binding Affinity and Sequence Specificity. The synthesis and characterization of **2**, **17**, and **21** have been reported previously (Trauger et al., 1996b; Turner et al., 1998; Urbach and Dervan, 2001). Polyamides **18**, **19**, and **20** were prepared according to standard solid phase protocols (Baird and Dervan, 1996; Herman et al., 1998). Quantitative DNase I footprint titrations (Trauger and Dervan, 2001) were performed for polyamides **2** and **17**-**21** on the 288 bp PCR product of pAU27 in order to compare the equilibrium association constants for the resulting complexes. Polyamides **2**, **17**, **18**, and **19** were designed to bind in hairpin and 1:1 modes at DNA sites 5'-TAGCGCT-3' and 5'-AAAGAGAAGAG-3', respectively (Figure 37). Compound **2** (β linker) binds to the 1:1 site with very high affinity (K_a = 1.5 x 10^{10} M^{-1}), displaying more than 150-fold specificity versus the hairpin site (K_a = 9.7 x 10^7 M^{-1}) (Table 7). Polyamide **17** (γ linker) binds with similar affinities to the 1:1 (K_a = 1.3 x 10^8 M^{-1}) and 2:1 (K_a = 7.6 x 10^8 M^{-1}) sites, showing a modest 5-fold preference for the hairpin site. The ^{H_2N}γ-linked

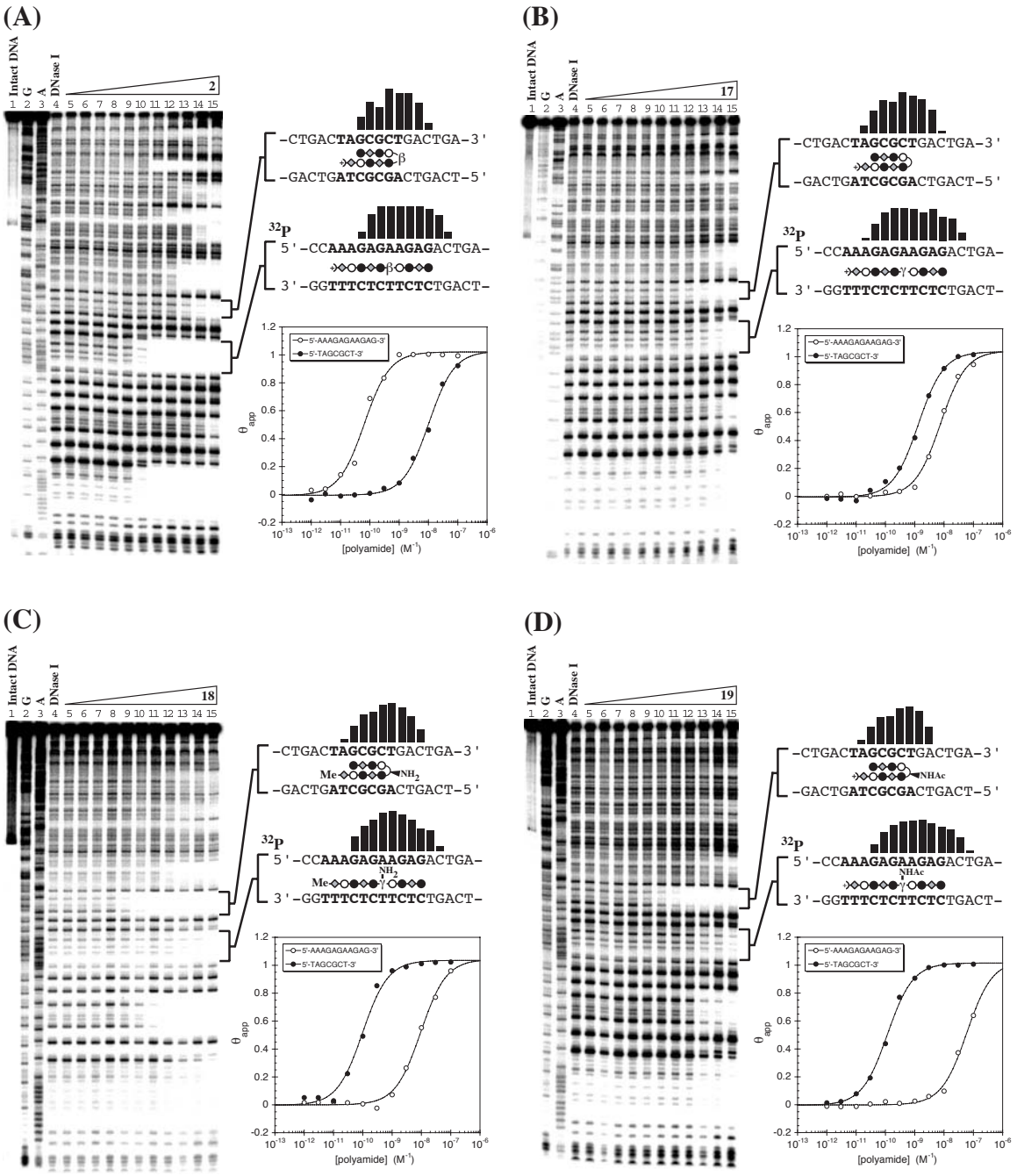


Figure 37 (A – D) Quantitative DNase I footprinting experiments for polyamides **2**, **17**, **18**, and **19**, respectively, on the 288 bp, 5'-end-labeled PCR product of plasmid pAU27: lane 1, intact DNA; lane 2, G reaction; lane 3, A reaction; lane 4, DNase I standard; lanes 5-15, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, 100 nM polyamide, respectively. Each footprinting gel is accompanied by the following: (right top) Schematic illustrating the observed protection pattern derived from the MPE footprinting experiment, with the polyamides shown in the observed conformation; (right bottom) Langmuir binding isotherms derived from the DNase I footprinting experiment for the two designed sites, 5'-AAAGAGAAGAG-3' and 5'-TAGCGCT-3', as determined from a nonlinear least squares fit.

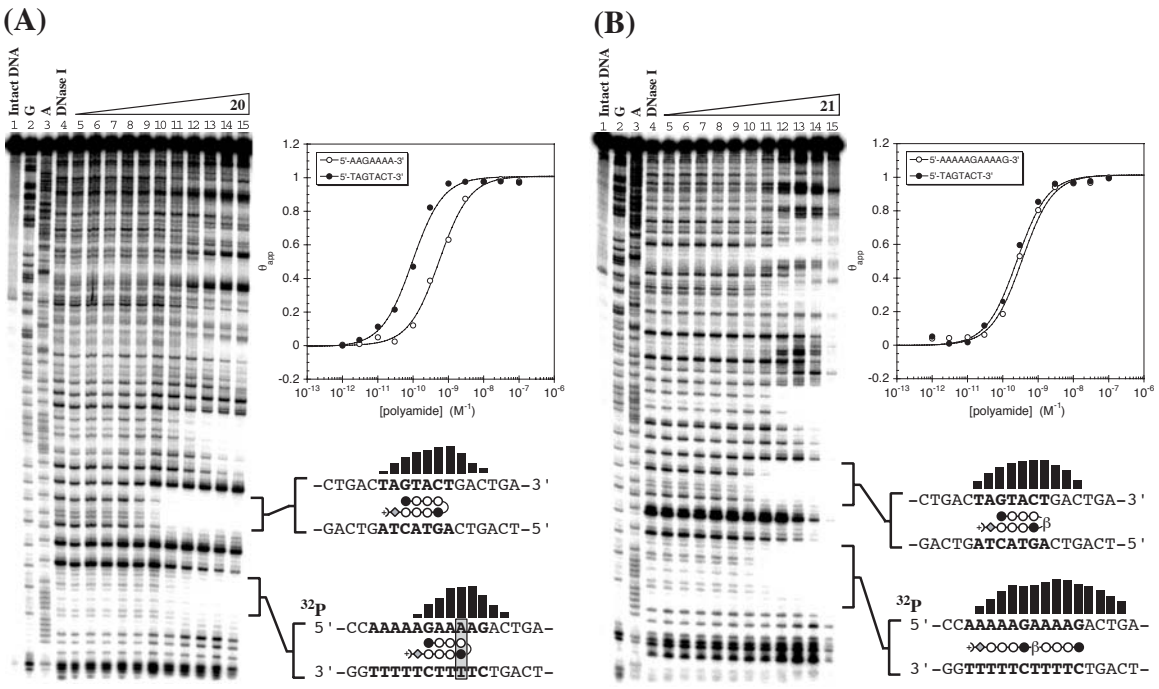


Figure 38 (A and B) Quantitative DNase I footprinting experiments for polyamides **20** and **21**, respectively, on the 288 bp, 5'-end-labeled PCR product of plasmid pAU27: lane 1, intact DNA; lane 2, G reaction; lane 3, A reaction; lane 4, DNase I standard; lanes 5-15, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, 100 nM polyamide, respectively. Each footprinting gel is accompanied by the following: (right top) Langmuir binding isotherms derived from the DNase I footprinting experiment for the two designed sites, 5'-AAAAAGAAAAG-3' and 5'-TAGTACT-3', as determined from a nonlinear least squares fit; (right bottom) schematic illustrating the observed protection pattern derived from the MPE footprinting experiment, with the polyamides shown in the observed conformation.

polyamide **18** binds with very high affinity to the hairpin site (K_a = 1.2 x 10^{10} M^{-1}), exhibiting more than 150-fold specificity versus the 1:1 site (K_a = 7.6 x 10^7 M^{-1}), which is an almost exact reversal of specificity compared to compound **2**. Polyamide **19** (^{Ac}γ-linker) binds at high affinity to the hairpin site (K_a = 8.5 x 10^9 M^{-1}), with greater than 500-fold specificity versus the 1:1 site (K_a = 1.6 x 10^7 M^{-1}).

Polyamides **20** and **21** were designed to bind in hairpin and 1:1 modes at DNA sites 5'-TAGTACT-3' and 5'-AAAAAGAAAAG-3', respectively (Figure 38). Compound **20** (β linker) binds with similar affinity ($K_a \approx$ 3 x 10^9 M^{-1}) to both

Table 7. Equilibrium Association Constants, K_a (M^{-1})*

Polyamide	Linker	Hairpin	1:1	Specificity (Hairpin/1-1)
Im-β-ImPy-		5'-TAGCGCT-3'	5'-AAAGAGAAGAG-3'	
(2)	β	9.7 (± 0.8) x 10^7	1.5 (± 0.5) x 10^{10}	0.0065
(17)	γ	7.6 (± 0.6) x 10^8	1.3 (± 0.5) x 10^8	5.8
(18)	H_2Nγ	1.2 (± 0.2) x 10^{10}	7.6 (± 0.2) x 10^7	160
(19)	Ac γ	8.5 (± 0.5) x 10^9	1.6 (± 0.4) x 10^7	530
ImPyPyPy-		5'-TAGTACT-3'	a b 5'-AAAAAGAAAAG-3'**	
(20)	β	3.9 (± 0.7) x 10^9	2.7 (± 0.7) x 10^9 *a*	1.4
(21)	γ	1.0 (± 0.2) x 10^{10}	1.8 (± 0.2) x 10^9 *b*	5.6

* Values reported are the mean values from at least three DNase I footprint titration experiments, with the standard deviation given in parentheses. Assays were performed at 22 °C in a buffer of 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$ at pH 7.0
** Observed binding sites are indicated as boxes **a** and **b**.

designed sites (Table 7). Polyamide **21** (γ-linker) binds with very high affinity to the hairpin site (K_a = 1.0 x 10^{10} M^{-1}), exhibiting a modest 5-fold selectivity over its single base-pair mismatch hairpin site, 5'-AAGAA**A**A-3' (mismatch base bolded), within the 1:1 binding site (K_a = 1.8 x 10^9 M^{-1}). All binding isotherms fit well to an n = 1 Hill equation, which is consistent with 1:1 polyamide:DNA stoichiometry (Figures 37 and 38).

Binding Site Size. Evidence for hairpin and 1:1 binding modes is provided by methidium propyl EDTA (MPE) footprinting, wherein the binding site size can be determined at high resolution (Trauger and Dervan, 2001). Binding modes were deduced from MPE footprint sizes on the basis of previously characterized hairpin and 1:1 complexes for polyamides **2**, **17**, and **21** at identical sites (Trauger et al., 1996b; Turner et al., 1998; Urbach and Dervan, 2001). For example, the eight-ring hairpin, **21**, occupies a seven base-pair binding site, whereas the prototypical 1:1 binder, **2**, occupies an eleven base-pair binding site. Figures 37 and 38 display occupation histograms derived from the

MPE gels (not shown), illustrated to the right of the respective DNase gel. Polyamides **2**, **17**, **18**, and **19** were designed to bind in hairpin and 1:1 modes at DNA sites 5'-TAGCGCT-3' and 5'-AAAGAGAAGAG-3', respectively (Figure 37); and polyamides **20** and **21** were designed to bind in hairpin and 1:1 modes at DNA sites 5'-TAGTACT-3' and 5'-AAAAAGAAAAG-3', respectively (Figure 38). In all cases, polyamides occupy their target hairpin sites as hairpins. Polyamides **2** and **17-20** bind to their target 1:1 sites in an extended 1:1 binding mode. However, polyamide **21** occupies a smaller site within its designed 1:1 site. Close inspection of the occupied site reveals the binding mode of **21** as a hairpin bound to a single base-pair mismatch site (Figure 38 at left). Inspection of the relative footprint sizes in the DNase gels further supports the MPE results.

Discussion. Linker-dependent conformational control of binding modes is established here for two distinct classes of hairpin polyamides–those containing contiguous rings, and those with beta alanine-endowed conformational flexibility. In the flexible series, **2**, **17**, **18** and **19**, compounds of type Im-β-ImPy-**X**-Im-β-ImPy-β-R (**X** = β, γ, ^{H_2N}γ, and ^{Ac}γ; R = Me or Dp) bind in hairpin and extended 1:1 modes to the designed sites, 5'-TAGCGCT-3' and 5'-AAAGAGAAGAG-3', respectively (Figure 37). Previous studies in this thesis on the β-linked polyamide **2** demonstrated its high affinity for the 1:1 site. The >150-fold specificity of **2** for 1:1 versus hairpin binding sites, presented here, may be attributed to the steric destabilization involved in the β residue adopting a hairpin conformation (de Clairac et al., 1997). Moreover, structural studies in 1:1 and 2:1 modes support the notion that β confers optimal polyamide-DNA alignment in an extended conformation (de Clairac et al., 1999; Urbach et al.,

2002). Inspection of the DNase gel reveals that **2** bound with very high affinity to the 5'-AAAAAGAAAAG-3' site. Although this site was not intended for **2**, the rules for 1:1 recognition would suggest this as a match site (Urbach and Dervan, 2001). The γ-linked polyamide **17** contains one additional methylene unit in the linker, which provides the flexibility needed to fold into a hairpin, as established previously (de Clairac et al., 1997; Turner et al., 1998). Interestingly, **17** discriminates least between the hairpin and 1:1 sites. This result poses an important problem because hairpin polyamides, which most commonly employ the γ linker, are the most sequence-specific class of DNA-binding polyamides. Therefore, a linker that favors hairpin formation while disfavoring 1:1 and other alternative binding modes would be of great value.

It was predicted from molecular modeling of the 1:1 NMR complex that adding functionality to an aliphatic linker in the 1:1 motif would cause steric destabilization of this binding mode. Furthermore, we postulated that incorporation of the Herman turn ($^{H_2N}\gamma$) would favor hairpin binding while disfavoring the 1:1 mode. Indeed, this result was observed for the $^{H_2N}\gamma$-linked polyamide **18**, which shows >150-fold preference for the hairpin binding site versus the 1:1 site. Therefore, polyamides **2** and **18** exhibit a reversal of preference for the hairpin and 1:1 binding sites. It should be noted that **18** contains a truncated tail in order to reduce bias in this series of compounds by maintaining a single positive charge. We have also tested the doubly charged Dp-tail analogue of **18**, finding that its recognition properties are virtually identical to **18**. Inspection of the data shown in Figure 37 reveals that **18** is relatively nonspecific, binding with high affinity to many other sites on the

plasmid. Acetylating the amino group of the Herman turn and restoring the charged Dp tail markedly reduces all mismatch affinities with a negligible decrease in hairpin site affinity, as observed for compound **19**. Furthermore, the $^{Ac}\gamma$-linked polyamide **19** exhibits >500-fold preference for hairpin versus 1:1 binding modes, a reversal of specificity effectively 82,000-fold compared to **2**. The exceptional specificity of **19** may be attributed to the limited mobility of the bulky acetamide group at the turn position, which may in turn limit alternative binding modes. These results demonstrate that by changing the linker residue we can control the ligand conformation and hence the DNA sequence preference.

Polyamides **20** and **21** contain contiguous four-ring subunits, which are inherently limited in conformational flexibility due in part to the continuous π-conjugation in each subunit. Therefore, it is expected that these compounds should display different recognition properties than polyamides **2**, **17**, **18**, and **19**. The β-linked polyamide **20** binds in an extended mode to the designed 1:1 sequence, 5'-AAAAAGAAAAG-3', and as a hairpin to the designed site, 5'-TAGTACT-3'. However, unlike the β-linked polyamide **2** described above, **20** does not show a preference for hairpin or 1:1 modes, binding both sites with similar subnanomolar affinities. On the other hand, the γ-linked compound **21** binds as a hairpin to both designed sites, tolerating a single base-pair mismatch within the 1:1 binding site, 5'-AAGAAAA-3', rather than adopting an extended conformation. This result provides further evidence that contiguous ring polyamides linked by γ have inherently unique DNA recognition properties, and as such are not inclined to adopt an extended binding conformation (Trauger et al., 1996a).

Implications for the Design of Minor Groove Binding Polyamides. The results presented here indicate that hairpin and extended modes of polyamide-DNA binding, which are dependent on the ligand conformation, can be controlled by the choice of linkage between subunits (Figure 39). For β-rich polyamides, the hairpin binding mode is favored by incorporating a substituted γ residue, such as $^{H_2N}\gamma$ or $^{Ac}\gamma$, to link the flexible polyamide subunits. In addition to disfavoring the 1:1 mode, the $^{Ac}\gamma$ linker substantially improves hairpin binding affinity and specificity. Alternatively, the 1:1 binding mode can be favored by using a β residue as the linker. We are particularly interested in favoring the hairpin binding conformation because of its high capacity for programmable DNA sequence specificity. Therefore, it is reassuring that for polyamides with multiple contiguous heterocycles, such as eight-ring polyamides, extended 1:1 binding modes do not compete well with the favored hairpin motif.

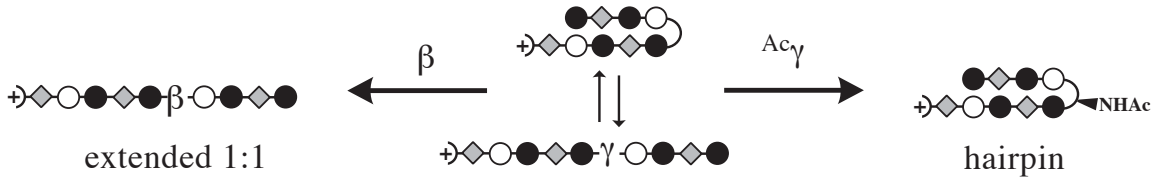


Figure 39 Control of polyamide-DNA binding modes. The γ-linked polyamide shown in the center binds well in both hairpin and 1:1 modes. Replacement of γ with β favors the 1:1 mode by 150-fold (left), whereas substitution of γ at the α-(R) position with acetamide ($^{Ac}\gamma$) favors the hairpin mode by 560-fold, thereby enforcing an 82,000-fold reversal of specificity.

Conclusions

The 1:1 motif for DNA recognition by β-alanine-linked polyamides has been investigated by footprinting, affinity cleavage, and multidimensional NMR techniques in order to further understand the properties of this system and to exploit this knowledge in the design of next-generation DNA-binding ligands. A set of rules for 1:1 recognition were developed, which are based on the sequence specificity of individual ligand residues, the sequence-dependent orientation of the ligand, ligand size, placement and number of β residues, and the DNA structure itself. The ambiguity of sequence targeting based on stoichiometry, i.e., 1:1 versus 2:1 binding, established here, has been resolved by employing linker-dependent control over the ligand conformation and, hence, its mode of binding.

The most striking structural feature of the 1:1 binding polyamides discovered by Laemmli and coworkers is the extensive number of aliphatic β linkages. Specificity studies demonstrated β to be A,T specific, similar to Py. The NMR structure reveals that this specificity derives from intimate interaction between the α-methylene group and the C2-H atom of adenine, an interaction that would be sterically excluded at G•C base pairs. Additionally, we found that β in the context of Im-β-Im, as opposed to Im-β-Py, is necessary for high-affinity DNA binding. The NMR structure reveals two distinct conformations for Im-β-Im and Im-β-Py subunits. Im-β-Im requires greater flexibility in order to accommodate the larger, 50° dihedral between the planes of the flanking Im rings, presumably to allow the Im residues to better orient for hydrogen bonding to guanine. Therefore, it is believed that β confers the flexibility required for Im residues to align properly with the DNA helix.

Attempts at developing a recognition code for the 1:1 motif were less successful than those that led to the pairing rules for 2:1 recognition. It is believed that polyamides composed of five-membered heterocyclic amino acids contain insufficient structural information, when bound 1:1, to provide a general recognition code. This is evidenced by lack of single base specificity observed for imidazole, pyrrole, furan, thiophene, thiazole, and hydroxythiophene amino acids. The single site specificity observed for hydroxypyrrole (Hp) in polyamide **4** would suggest otherwise. For this anomalous result, I offer the following explanation: Hp projects the bulky C3-OH group to the minor groove floor, which explains the observed A,T > G,C specificity; the A > T specificity observed for the DNA sequence AAGAGA**A**GAG is probably a result of disrupting a stable, repeating polypurine base stack with a pyrimidine (T). One might argue that, based on this explanation, we should have observed a similar effect for the Hp-containing polyamide **6**. I would say that the Hp residue in **6** has greater conformational flexibility, conferred by the two flanking β residues, and therefore it can tolerate variations in DNA structure to a larger extent, thus resulting in reduced sequence specificity.

The NMR structure presented here offers the first high-resolution look at a polyamide containing imidazole and β residues bound in a 1:1 complex. The hydrogen bonds observed between Im-N3 and G-NH_2 provide the first direct evidence of the lexitropsin model as originally envisioned in the 1:1 motif (Kopka et al., 1985). The complex reveals B-form DNA with a narrow minor groove and a large negative propeller twist, which is shown to be stabilized by bifurcated hydrogen bonds donated from each amide NH group to its proximal purine N3

and pyrimidine O2 atoms. Stabilization of the negative propeller twist by these interactions, 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 though to be the driving force for optimal ligand-DNA alignment. If this is so, the 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 NH_2 group is optimal when the polyamide is oriented N – C with respect to the 3' – 5' direction of the guanine-containing strand.

1:1 and 2:1 binding modes clearly have different rules for recognition. This ambiguity of sequence targeting depending on stoichiometry was addressed in order to eliminate alternative binding modes and therefore improve our accuracy of DNA sequence predetermination. The results presented here indicate that hairpin and 1:1 binding modes, which are dependent on ligand conformation, can be controlled by changing the linkage between subunits. The 1:1 binding mode is favored by using a β linker. Alternatively, the hairpin binding mode is favored by incorporating an α-(R)-amino substituted γ linker, i.e., the Herman turn, to link the flexible polyamide subunits. In addition to disfavoring the 1:1 mode, the acetylated Herman turn substantially improves hairpin binding affinity and sequence specificity.

Although the 1:1 binding mode may be less specific overall, it allows us to target certain DNA sequences that are not accessible to 2:1 binders. Polyamides that integrate 1:1 and 2:1 binding modes, such as the extended hairpins (Trauger et al., 1996c), will be an important future direction for targeting unique sequences in a genomic context. One can imagine that integrated motifs could exploit the preference of 1:1 versus 2:1 binding relevant to the sequence-dependent microstructure of DNA, i.e., certain sequence contexts with a normal B-like minor groove that may prefer pairs of rings (2:1 binding), whereas narrow minor groove tracts would allow the steric fit of only single rings (1:1 binding). The accurate prediction of optimal ligand-DNA complementarity requires an understanding of the geometric and electronic structural parameters for both molecules, knowledge that is based largely on high-resolution structural studies. The Thesis presented here provides a biophysical as well as structural foundation for future applications of the 1:1 motif.

Experimental

Materials. Methylamine, piperidine, and dimethylaminopropylamine were purchased from Aldrich. Dimethylformamide (DMF) and diisopropylethylamine (DIEA), and dimethylsulfoxide/N-methylpyrrolidone (DMSO/NMP) (1:1, v/v) were purchased from Applied Biosystems. N,N-dimethylaminopropylamine, 3,3'-diamino-N-methyldipropylamine, ethylenediamine-tetraacetic dianhydride, triethylamine, and acetic acid were purchased from Aldrich. Acetic anhydride and acetonitrile were from EM. Trifluoroacetic acid (TFA) was from Halocarbon. (R)-2-Fmoc-4-Boc-diaminobutyric acid was purchased from Bachem. Boc-β-Pam resin was purchased from Peptides International. Calcuim chloride, magnesium chloride, and potassium chloride were from Fluka. Glycogen (20 mg/mL), dNTP's (PCR nucleotide mix), and all enzymes (unless otherwise stated) were purchased from Boehringer-Mannheim and used with their supplied buffers. pUC19 was from New England Biolabs. Deoxyadenosine [γ-^{32}P] triphosphate was from ICN. Calf thymus DNA (sonicated, deproteinized) and DNase I (7500 u/mL, FPLC pure) were from Amersham-Pharmacia. AmpliTaq DNA polymerase was from Perkin Elmer and used with the supplied buffers. HEPES was from Sigma. Tris-HCl, dithiothreitol (DTT), RNase-free water (used for all DNA manipulations), and 0.5 M EDTA were from US Biochemicals. Ethanol (200 proof) was from Equistar. Calcium chloride, potassium chloride, and magnesium chloride were from Fluka. Formamide and pre-mixed tris-borate-EDTA (Gel-Mate, used for gel running buffer) were from Gibco. Bromophenol blue was from Acros. All reagents were used without further purification. Centricon YM-3 dialysis filters were from

Amicon. D_2O ("100%") was from Cambridge Isotope Laboratories. NMR tubes were type 535 from Wilmad.

HPLC analysis was performed on a Beckman Gold system using a Rainin C_{18}, Microsorb MV, 5 μm, 300 x 4.6 mm reversed phase column in 0.1% (wt/v) TFA with acetonitrile as eluent and a flow rate of 1.0 mL/min, gradient elution 1.25% acetonitrile/min. Preparatory reversed phase HPLC was performed on a Beckman HPLC with a Waters DeltaPak 25 x 100 mm, 100 μm C18 column equipped with a guard, 0.1% (wt/v) TFA, 0.25% acetonitrile/min. Oligonucleotides were synthesized at the Biopolymer Synthesis Center at the California Institute of Technology (Caltech). Plasmid sequencing was carried out by Davis Sequencing (Davis, CA) or by 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. FPLC was performed on a Pharmacia LKB system. All buffers were 0.2 μm filtered. DNA manipulations were performed according to standard protocols (Sambrook et al., 1989).

Solid-Phase Synthesis of Polyamides. Published protocols were used to synthesize the Boc-protected amino acid monomers and dimers for all polyamides studied (Baird and Dervan, 1996; Urbach et al., 1999). Solid phase synthesis was performed by manual stepwise coupling on Boc-β-PAM resin using Boc-protected amino acid monomers and dimers as previously described (Baird and Dervan, 1996; Herman et al., 1998; Urbach et al., 1999). Couplings

were monitored by analytical HPLC, and all polyamides except **18** were cleaved from solid support using dimethylaminopropylamine at 100 °C for 2h, and purified by preparatory HPLC. Special considerations for polyamides **4**, **18**, and **19** are described below. The fractions were analyzed by analytical HPLC, and the appropriate fractions were lyophilized to dryness.

ImPyImPy-β-ImPyImPy-β-Dp (1). 70% recovery; UV (H_2O) λ_{max} 244 nm (ε = 28 500, measured), 308 nm (ε = 44 300, measured); ^{1}HNMR (DMSO-*d6*) δ 10.41 (s, 1 H), 10.33 (s, 1 H), 10.21 (s, 2 H), 10.01 (s, 1 H), 9.99 (s, 1 H), 9.91 (s, 1 H), 9.18 (br s, 1 H), 8.06 (m, 3 H), 7.53 (s, 1 H), 7.52 (s, 1 H), 7.46 (s, 1 H), 7.38 (m, 2 H), 7.36 (d, 1 H, *J* = 1.8 Hz), 7.21 (d, 1 H, *J* = 1.8 Hz), 7.19 (d, 1 H, *J* = 1.8 Hz), 7.14 (d, 1 H, *J* = 1.8 Hz), 7.12 (d, 1 H, *J* = 1.8 Hz), 7.04 (s, 1 H), 6.94 (m, 2 H), 3.97 (s, 3 H), 3.95 (s, 6 H), 3.94 (s, 3 H), 3.84 (s, 3 H), 3.84 (s, 3 H), 3.80 (s, 3 H), 3.78 (s, 3 H), 3.42 (m, 2 H), 3.35 (m, 2 H), 3.09 (q, 2 H, *J* = 6.3 Hz), 2.98 (m, 2 H), 2.72 (d, 6 H, *J* = 4.5 Hz), 2.57 (t, 2 H, *J* = 6.6 Hz), 2.33 (t, 2 H, *J* = 6.3 Hz), 1.71 (m, 2 H); MALDI-TOF-MS (monoisotopic), 1210.6 (1210.5 calc. for $C_{55}H_{68}N_{23}O_{10}$);

Im-β-ImPy-β-Im-β-ImPy-β-Dp (2). 24% recovery; UV (H_2O) λ_{max} 248 (ε = 27 000, measured), 290 (ε = 28 200, measured); ^{1}HNMR (DMSO-*d6*) δ 10.36 (s, 1 H), 10.33 (s, 2 H), 9.32 (br s, 1 H), 9.26 (s, 2 H), 8.40 (t, 1 H, *J* = 6 Hz), 8.05 (m, 3 H), 7.89 (t, 1 H, *J* = 6 Hz), 7.43 (m, 2 H), 7.39 (s, 1 H), 7.34 (s, 1 H), 7.19 (d, 2 H), 6.98 (s, 1 H), 6.92 (s, 1 H), 6.90 (s, 1 H), 3.91 (m, 9 H), 3.89 (s, 3 H), 3.78 (s, 6 H), 3.47 (m, 4 H), 3.36 (m, 4 H), 3.08 (q, 2 H, *J* = 6 Hz), 2.98 (m, 2 H), 2.72 (d, 6 H, *J* = 4.5 Hz), 2.58 (t, 4 H, *J* = 6.6 Hz), 2.53 (t, 3 H, *J* = 7.2 Hz), 2.32 (t, 2 H, *J* = 7.2 Hz), 1.71 (m, 2 H); MALDI-TOF-MS (monoisotopic), 1108.6 (1108.5 calc. for $C_{49}H_{66}N_{21}O_{10}$);

Im-β-ImPyPyIm-β-ImPy-β-Dp (3). 11% recovery; UV (H_2O) λ_{max} 254 nm (ε = 32 200, measured), 308 nm (ε = 43 700, measured); ^{1}HNMR (DMSO-*d6*) δ 10.34 (s, 1H), 10.33 (s, 2H), 10.00 (s, 1 H), 9.94 (s, 2 H), 9.22 (br s, 1 H), 8.38 (t, 1 H, *J* = 5.7 Hz), 8.04 (m, 2 H), 7.91 (t, 1 H, *J* = 6.0 Hz), 7.50 (d, 1 H, *J* = 1.2 Hz), 7.45 (d, 1 H, *J* = 1.2 Hz), 7.41 (d, 1 H, *J* = 1.2 Hz), 7.33 (s, 1 H), 7.31 (s, 1 H), 7.25 (s, 1 H), 7.19 (s, 1 H), 7.13 (s, 1 H), 7.07 (s, 1 H), 6.96 (t, 1 H, *J* = 1.2 Hz), 6.91 (s, 1 H), 3.94 (s, 3 H), 3.93 (m, 6 H), 3.92 (s, 3 H), 3.82 (s, 6 H), 3.77 (s, 3 H), 3.49 (m, 4 H), 3.35 (q, 2 H, *J* = 6.0 Hz), 3.08 (q, 2 H, *J* = 6.0 Hz), 2.97 (m, 2 H), 2.71 (d, 6 H, *J* = 3.9 Hz), 2.59 (m, 4 H), 2.32 (t, 2 H, *J* = 6.9 Hz), 1.71 (m, 2 H); MALDI-TOF-MS (monoisotopic), 1159.6 (1159.5 calc. for $C_{52}H_{67}N_{22}O_{10}$).

Im-β-ImHp-β-Im-β-ImPy-β-Dp (4). The hydroxypyrrole-containing polyamide **4** was prepared using the methyl ether-protected hydroxypyrrole amino acid, Boc-Op-OH as described previously (Urbach et al., 1999). The protected polyamide, Im-β-ImOp-β-Im-β-ImPy-β-Dp, was synthesized as described above. 14% recovery, UV (H_2O) λmax 290 (ε~28,000); ^{1}H NMR (DMSO-d_6) δ 10.50 (s, 1H), 10.42 (s, 1 H), 10.33 (s, 1 H), 9.91 (s, 1 H), 9.27 (br s, 1 H), 8.91 (s, 1 H), 8.41 (t, 1 H, J = 6.0 Hz), 8.04 (t, 2 H, J = 6.0 Hz), 7.89 (t, 1 H, J = 6.0 Hz), 7.49 (s, 1 H), 7.42 (s, 1 H), 7.41 (s, 1 H), 7.34 (s, 1 H), 7.19 (s, 1 H), 7.14 (s, 1 H), 6.99 (s, 1 H), 6.92 (s, 1 H), 3.29 (s, 3 H), 3.92 (m, 6 H), 3.90 (s, 3 H), 3.78 (s, 3 H), 3.75 (s, 3 H), 3.69 (s, 3 H), 3.49 (m, 6 H), 3.36 (q, 2 H, J = 5.3 Hz), 3.09 (q, 2 H, J = 6.3 Hz), 2.97 (t, 2 H, J = 5.7 Hz), 2.72 (d, 6 H, J = 4.8 Hz), 2.58 (m, 6 H), 2.33 (t, 2 H, J = 6.9 Hz), 1.72 (quintet, 2 H, J = 8.4 Hz); MALDI-TOF-MS (monoisotopic), 1138.58 (1138.54 calcd. for $C_{50}H_{68}N_{21}O_{11}^+$).

In order to remove the methoxy protecting group, a sample of Im-β-ImOp-β-Im-β-ImPy-β-Dp, (9 mg, 7.9 mmol) was treated with sodium thiophenoxide at 100 ˚C for 2 h as described previously (Urbach et al., 1999). DMF (1.0 mL) and thiophenol (0.5 mL) were placed in a (13 x 100 mm) disposable Pyrex screw cap culture tube. A 60% dispersion of sodium hydride in mineral oil (100 mg) was slowly added. Upon completion of the sodium hydride addition, Im-β-ImOp-β-Im-β-ImPy-β-Dp (5 mg) dissolved in DMF (0.5 mL) was added. The solution was agitated, and heated to 100 ˚C for 2 h. Upon completion of the reaction, the reaction mixture was cooled to 0 ˚C, and 7 mL of a 20% (wt/v) aqueous solution of trifluoroacetic acid added. The aqueous layer is separated from the resulting biphasic solution. In order to remove the last trace of thiophenol, the aqueous layer was extracted three times with diethyl ether. The deprotected polyamide was purified by reversed phase preparatory HPLC. Im-β-ImHp-β-Im-β-ImPy-β-Dp **4** is recovered upon lyophilization of the appropriate fractions as a white powder (1.5 mg, 18% recovery). UV (H_2O) λmax 290 (ε ~ 28,000), ; ^{1}H NMR (DMSO-d_6) δ 10.38 (s, 1 H), 10.37 (s, 1 H), 10.33(s, 1 H), 9.93 (s, 1 H), 9.90 (s, 1 H), 9.61 (s, 1 H), 9.18 (br s, 1 H), 8.37 (t, 1 H, J = 6.0 Hz), 8.04 (m, 2 H), 7.90 (t, 1 H, J = 6.0 Hz), 7.56 (t, 1 H, J = 5.9 Hz), 7.48 (s, 1 H), 7.43 (s, 1 H), 7.40 (s, 1 H), 7.31 (s, 1 H), 7.19 (s, 1 H), 7.12 (s, 1 H), 6.94 (s, 1 H), 6.93 (s, 1 H), 3.92 (m, 9 H), 3.90 (s, 3 H), 3.78 (s, 3 H), 3.74 (s, 3 H), 3.48 (m, 6 H), 3.09 (q, 2 H, J = 5.7 Hz), 2.99 (m, 6 H), 2.56 (m, 4 H), 2.48 (d, 6 H, J = 5.4 Hz), 2.32 (t, 2 H, J = 5.9 Hz), 1.61 (quintet, 2 H, J = 5.4 Hz); MALDI-TOF-MS (monoisotopic), 1124.6 (1124.5 calcd. for $C_{49}H_{66}N_{21}O_{11}^+$).

ImPy-β-Im-β-ImPy-β-Dp (13). 29% recovery, UV (H_2O) λmax 290 (ε~30 000, measured); MALDI-TOF-MS (monoisotopic), 914.5 (914.5 calc. for $C_{41}H_{56}N_{17}O_8^+$); This compound is the one characterized in complex with DNA as described in the Results and Discussion section. The complete assignment of all protons by ^{1}HNMR is given in Table 5.

Im-β-ImPy-β-Im-β-Dp (14). 64% recovery, UV (H_2O) λmax 270 (ε~23 300, measured); MALDI-TOF-MS (monoisotopic), 792.4 (792.4 calc. for $C_{35}H_{50}N_{15}O_7^+$).

ImPy-β-Im-β-Dp (15). 55% recovery, UV (H_2O) λmax 280 (ε~19 500, measured); MALDI-TOF-MS (monoisotopic), 598.2 (598.3 calc. for $C_{27}H_{40}N_{11}O_5^+$).

Im-β-ImPy-β-Dp (16). 64% recovery, UV (H_2O) λmax 270 (ε~20 200, measured); MALDI-TOF-MS (monoisotopic), 598.3 (598.3 calc. for $C_{27}H_{40}N_{11}O_5^+$).

Im-β-ImPy-(R)^{H_2N}γ-Im-β-ImPy-β-Me (18). Im-β-ImPy-(R)^{H2N}γ-Im-β-ImPy-β-Pam resin was synthesized in a stepwise fashion from Boc-β-Pam resin using manual solid-phase protocols (Baird and Dervan, 1996). The chiral diaminobutyric acid "turn" residue (Herman et al., 1998) was incorporated by coupling (R)-2-Fmoc-4-Boc-diaminobutyric acid (10 equivalents) to 300 mg Boc-Im-Py-Py-Py-β-Pam resin in 2 mL DMF with 1.1 equivalents of DIEA at 37 °C for 2 h, followed by an acetylation wash (1 mL acetic anhydride, 1 mL DIEA, 2 mL DMF, 5 min, room temp). Subsequent coupling steps used 1.1 equivalents of DIEA and 45 min coupling times at room temperature to minimize Fmoc deprotection. Im-β-ImPy-(R)^{Fmoc}γ-Im-β-ImPy-β-Pam resin was treated with piperidine for 20 min at room temperature to remove the Fmoc group. 100 mg (38 μmol) of vacuum-dried Im-β-ImPy-(R)^{H_2N}γ-Im-β-ImPy-β-Pam resin was cleaved in 30 mL condensed methylamine in a Parr bomb apparatus at 50 °C for

2 h, then overnight at room temperature. The methylamine was allowed to evaporate at ambient pressure and temperature, and the resin was suspended in 2 mL acetonitrile, followed by 7 mL 0.1% (wt/v) $TFA_{(aq)}$. The suspension was filtered, and the filtrate was purified by reversed phase preparatory HPLC to afford **18** as a white powder (7.1 μmol, 19% recovery) upon lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1066.49 (1066.48 calcd. for $C_{46}H_{60}N_{21}O_{10}^{+}$).

Im-β-ImPy-(R)^{Ac}γ-Im-β-ImPy-β-Dp (19). Im-β-ImPy-(R)^{H_2N}γ-Im-β-ImPy-β-Pam resin was synthesized as described above for **18**. The resin was washed with acetic anhydride in DMF and DIEA for 10 min at room temperature and then dried *in vacuo*. 100 mg (37 μmol) Im-β-ImPy-(R)^{Ac}γ-Im-β-ImPy-β-Pam resin was treated with dimethylaminopropylamine at 100 °C for 2 h. The resin was removed by filtration, and the filtrate was diluted to 10 mL with 0.1 % (wt/v) $TFA_{(aq)}$ and purified by reversed phase HPLC. **19** was obtained as a white powder (11% recovery) upon lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1179.57 (1179.57 calcd. for $C_{52}H_{71}N_{22}O_{11}^{+}$).

ImPyPyPy-β-ImPyPyPy-β-Dp (20). MALDI-TOF-MS (monoisotopic), 1208.6 (calcd. 1208.6 for $C_{57}H_{70}N_{21}O_{10}^{+}$).

Synthesis of EDTA Conjugates. A single-step aminolysis of the resin ester linkage was accomplished using 3,3'-diamino-N-methyldipropylamine to afford polyamides with a primary amine at the C-terminus (**X-NH$_2$**). After purification by HPLC these were allowed to react with excess EDTA

dianhydride, and the EDTA conjugates (**XE**) were purified by reversed phase HPLC.

ImPyImPy-β-ImPyImPy-β-Dp-NH$_2$ (1-NH$_2$) A sample of ImPyImPy-β-ImPyImPy-β-Pam-resin resin (300 mg, 59 mmol) was placed in a 20 mL glass scintillation vial and treated with 3,3'-diamino-N-methyldipropylamine (2 mL) at 55 °C for 12 h. Resin was removed by filtration, and the filtrate diluted to a total volume of 10 mL with 0.1% (wt/v) aqueous TFA. The resulting solution was purified by reversed phase preparatory HPLC (analytical r.t. = 30.5 min.) to provide ImPyImPy-β-ImPyImPy-β-NH$_2$ (**1-NH$_2$**) as a white powder upon lyophilization of the appropriate fractions (32.6 mg, 26 mmol, 44% recovery). ^{1}HNMR (DMSO-*d6*) δ 10.42 (s, 1 H), 10.33 (s, 1 H), 10.21 (s, 1 H), 10.20 (s, 1 H), 10.01 (s, 1 H), 10.00 (s, 1 H), 9.91 (s, 1 H), 9.48 (br s, 1 H), 8.09 (m, 3 H), 7.80 (br s, 3 H), 7.53 (s, 1 H), 7.52 (s, 1 H), 7.46 (s, 1 H), 7.39 (s, 1 H), 7.38 (s, 1 H), 7.36 (s, 1 H), 7.21 (s, 1 H), 7.19 (s, 1 H), 7.14 (s, 1 H), 7.13 (s, 1 H), 7.05 (s, 1 H), 6.95 (d, 1 H, *J* = 1.2 Hz), 6.94 (d, 1 H, 1.5 Hz), 3.97 (s, 3 H), 3.95 (s, 6 H), 3.94 (s, 3 H), 3.84 (s, 6 H), 3.80 (s, 3 H), 3.78 (s, 3 H), 3.42 (m, 2 H), 3.36 (q, 2 H, *J* = 6.3 Hz), 3.10 (m, 4 H), 3.02 (m, 2 H), 3.83 (q, 2 H, *J* = 6.0 Hz), 2.71 (d, 3 H, *J* = 4.8 Hz), 2.58 (t, 2 H, *J* = 6.9 Hz), 2.34 (t, 2 H, *J* = 7.5 Hz), 1.87 (m, 2 H), 1.75 (m, 2 H); MALDI-TOF-MS (monoisotopic), 1253.6 (1253.6 calc. for $C_{57}H_{73}N_{24}O_{10}$).

ImPyImPy-β-ImPyImPy-β-Dp-EDTA (1E) EDTA-dianhydride (50 mg) was dissolved in a solution of DMSO/NMP (mL) and DIEA (1 mL) by heating at 55 °C for 5 min. The dianhydride solution was added to ImPyImPy-β-ImPyImPy-β-NH$_2$ (**1-NH$_2$**) (15.3 mg, 10.0 mmol) dissolved in 750 mL DMSO. The mixture was heated at 55 °C for 25 min and then treated with 0.1 N aqueous

NaOH (3 mL) for 10 min at 55 °C. Aqueous TFA (0.1% v/v) was then added to a total volume of 10 mL and the solution was purified directly by reversed phase preparatory HPLC to yield **1E** as a white powder (2.4 mg, 1.6 mmol, 16% recovery) upon lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1527.7 (1527.7 calc. for $C_{67}H_{87}N_{26}O_{17}$).

Im-β-ImPy-β-Im-β-ImPy-β-Dp-NH$_2$ (2-NH$_2$) 27% recovery, ^{1}HNMR (DMSO-*d6*) δ 10.36 (s, 1 H), 10.32 (s, 2 H), 9.93 (s, 1 H), 9.92 (s, 1 H), 9.56 (br s, 1 H), 8.37 (t, 1 H, *J* = 6.3 Hz), 8.07 (m, 3 H), 7.89 (t, 1 H, *J* = 5.7 Hz), 7.83 (br s, 3 H), 7.44 (s, 1 H), 7.43 (s, 1 H), 7.39 (s, 1 H), 7.32 (s, 1 H), 7.19 (s, 2 H), 6.95 (s, 1 H), 6.92 (s, 1 H), 6.90 (d, 1 H, *J* = 2.1 Hz), 3.91 (s, 9 H), 3.89 (s, 3 H), 3.78 (s, 3 H), 3.48 (m, 4 H), 3.37 (m, 4 H), 3.10 (q, 2 H, *J* = 6.0 Hz), 3.01 (m, 2 H), 2.83 (q, 2 H, *J* = 6.3 Hz), 2.71 (d, 3 H, *J* = 4.8 Hz), 2.58 (m, 4 H), 2.52 (m, 4 H), 2.33 (t, 2 H, *J* = 7.2 Hz), 1.86 (m, 2 H), 1.74 (m, 2 H); MALDI-TOF-MS (monoisotopic), 1151.6 (1151.6 calc. for $C_{51}H_{71}N_{22}O_{10}$).

Im-β-ImPy-β-Im-β-ImPy-β-Dp-EDTA (2E) 12.5% recovery, MALDI-TOF-MS (monoisotopic), 1425.8 (1425.7 calc. for $C_{61}H_{85}N_{24}O_{17}$).

Im-β-ImPyPyIm-β-ImPy-β-Dp-NH$_2$ (3-NH$_2$) 8.6% recovery, ^{1}HNMR (DMSO-*d6*) δ 10.34 (s, 1 H), 10.32 (s, 2 H), 10.00 (s, 1 H), 9.95 (s, 1 H), 9.94 (s, 1 H), 9.31 (br s, 1 H), 8.37 (t, 1 H, *J* = 6.0 Hz), 8.07 (m, 2 H), 7.92 (t, 1 H, *J* = 5.1 Hz), 7.73 (br s, 3 H), 7.50 (s, 1 H), 7.46 (s, 1 H), 7.45 (s, 1 H), 7.32 (s, 1 H), 7.30 (d, 1 H, *J* = 1.8 Hz), 7.25 (d, 1 H, *J* = 1.5 Hz), 7.19 (d, 1 H, *J* = 1.5 Hz), 7.13 (d, 1 H, *J* = 1.5 Hz), 7.07 (d, 1 H, *J* = 1.5 Hz), 6.95 (s, 1 H), 6.92 (d, 1 H, 1.8 Hz), 3.93 (s, 3 H), 3.92 (s, 3 H), 3.92 (s, 6 H), 3.83 (s, 3 H), 3.82 (s, 3 H), 3.77 (s, 3 H), 3.49 (m, 4 H), 3.35 (q, 2 H, *J* = 6.0 Hz), 3.10 (q, 2 H, *J* = 6.9 Hz), 3.02 (m, 4 H), 2.82 (q, 2 H, *J* = 6.3 Hz), 2.71 (d, 3

H, J = 4.8 Hz), 2.59 (m, 4 H), 2.33 (t, 2 H, J = 6.9 Hz), 1.86 (t, 2 H, J = 8.7 Hz), 1.74 (t, 2 H, J = 8.1 Hz); MALDI-TOF-MS (monoisotopic), 1202.7 (1202.6 calc for $C_{54}H_{72}N_{23}O_{10}$).

Im-β-ImPyPyIm-β-ImPy-β-Dp-EDTA (3E) 7.5% recovery, MALDI-TOF-MS (monoisotopic), 1476.8 (1476.7 calc. for $C_{64}H_{86}N_{25}O_{17}$).

ImPy-β-Im-β-ImPy-β-Dp-NH_2 (13-NH_2) MALDI-TOF-MS (monoisotopic), 957.5 (957.5 calc. for $C_{43}H_{61}N_{18}O_8^+$).

ImPy-β-Im-β-ImPy-β-Dp-EDTA (13-E) MALDI-TOF-MS (monoisotopic), 1231.3 (1231.6 calc. for $C_{53}H_{75}N_{20}O_{15}^+$).

Construction of Plasmid DNA. Plasmids were constructed by inserting the following hybridized inserts into the *Bam*HI/*Hin*DIII polycloning site in pUC19 (match sites bolded for convenience):

pAU9 5'- G A T C C G G C C **A A A G A G A A G A G** A A A C C G G G G C C **T A G C G C A G C G C T A** G G C C A -3' • 5'- A G C T T G G C C T A G C G C T G C G C T A G G C C C C G G T T T C T C T T C T C T T T G G C C G -3'

pAU8 5'- G A T C C G G C C **A A A G A G A A G A G** A A A C C G G G G C **C A A A G A C A A G A G** A A A C C G G G G C C **A A A G A A A A G A G** A A A C C G G G G C C **A A A G A T A A G A G** A A A C C G G A -3' • 5'- A G C T T C C G G T T T C T C T T A T C T T T G G C C C C G G T T T C T C T T T T C T T T G G C C C C G G T T T C T C T T G T C T T T G G C C C C G G T T T C T C T T C T C T T T G G C C G -3'

pAU12 5'- G A T C C G G C C **A A A G A G A A G A G** A A A C C G G G G C C **A A A G A G A T G A G** A A A C C G G G G C C **A A A G A G A G G A**

G A A A C C G G G G C **C A A A G A G A C G A G** A A A C C G G A -3' • 5'- A G C T T C C G G T T T C T C G T C T C T T T G G C C C C G G T T T C T C C T C T C T T T G G C C C C G G T T T C T C A T C T C T T T G G C C C C G G T T T C T C T T C T C T T T G G C C G -3'

pAU13 5'- G A T C C G G C C **A A A G A G A A G A G** A A A C C G G G G C C **A A A G A G T A G A G** A A A C C G G G G C C **A A A G A G G A G A G** A A A C C G G G G C **C A A A G A G C A G A G** A A A C C G G A -3' • 5'- A G C T T C C G G T T T C T C T G C T C T T T G G C C C C G G T T T C T C T C C T C T T T G G C C C C G G T T T C T C T A C T C T T T G G C C C C G G T T T C T C T T C T C T T T G G C C G -3'

pAU15 5'- G A T C C G G C C **A A A G A G A A G A G** A A A C C G G G G C **C A A A G T G A A G A G** A A A C C G G G G C C **A A A G G G A A G A G** A A A C C G G G G C C **A A A G C G A A G A G** A A A C C G G A -3' • 5'- A G C T T C C G G T T T C T C T T C G C T T T G G C C C C G G T T T C T C T T C C C T T T G G C C C C G G T T T C T C T T C A C T T T G G C C C C G G T T T C T C T T C T C T T T G G C C G -3'

pAU16 5'- G A T C C G G C C **A A A G A G A A G A G** A A A C C G G G G C **C A A A G A G A A C A G** A A A C C G G G G C C **A A A G A G A A G A A** A A A C C G G G G C C **A A A G A G A A T A G** A A A C C G G A -3' • 5'- A G C T T C C G G T T T C T A T T C T C T T T G G C C C C G G T T T C T T T T C T C T T T G G C C C C G G T T T C T G T T C T C T T T G G C C C C G G T T T C T C T T C T C T T T G G C C G -3'

pAU18 5'- G A T C C G G C C **A A A G A G A A G A G** A A A C C G G G G C C **A A A C A C A A C A C** A A A C C G G G G C C **A A A A A A A A A A A** A A A C C G G G G C C **A A A T A T A A T A T** A A A C C G G A -3' • 5'- A

G C T T C C G G T T T A T A T T A T A T T T G G C C C C G G T T T T T T T T T T T T T T G G C C C C G G T T T G T G T T G T G T T T G G C C C C G G T T T C T C T T C T C T T T G G C C G -3'

pAU20 5'-G A T C C G G C C **A A A G A G A A G** C G G G T T G G C C **A A A G A G A A G A** C G G G T T G G C C **A A A G A G A A G A A** C G G G T T G G C C A A G A G A A G A A C G G G G A-3' • 5'-A G C T T C C C C G T T C T T C T C T T G G C C A A C C C G T T C T T C T C T T T G G C C A A C C C G T C T T C T C T T T G G C C A A C C C G C T T C T C T T T G G C C G-3'

pAU27 5'-G A T C C G G G G C C **A A A A A G A A A A A G** A C T G A C T G A C **T A G T A C T** G A C G A C T G A C C **A A A G A G A A G A G** A C T G A C T G A C **T A G C G C T** G A C T G A-3' • 5'-A G C T T C A G T C A G C G C T A G T C A G T C A G T C T C T T C T C T T T G G T C A G T C G T C A G T A C T A G T C A G T C A G T C T T T T C T T T T T G G C C C C G-3'

The ligated plasmids were subsequently transformed in to JM109 subcompetent cells (Promega) and amplified by cell growth, according to standard protocols (Sambrook et al., 1989). Plasmid DNA was isolated using WizardPlus Midi Preps from Promega.

Preparation of 5'- and 3′-End-Labeled Restriction Fragments. These procedures are described in detail in Trauger and Dervan, 2001. For 3'-labeling: pUC19 was *Pvu*II/*Eco*RI linearized and then 3'-filled with deoxyadenosine 5′-[a-^{32}P] and thymidine 5′-[a-^{32}P] triphosphates using Klenow polymerase. For 5'-labelling: two 20 base pair primer oligonucleotides, 5'- A A T T C G A G C T C G G T A C C C G G -3' (forward) and 5'- C T G G C A C G A C A G G T T T C C C G -3' (reverse) were constructed to complement the pUC19 *EcoRI* and *PvuII* sites,

respectively, such that amplification by PCR would mimic the long, 3'-filled, pUC19 *EcoRI*/*PvuII* restriction fragment. The forward primer was radiolabelled using γ-^{32}P-dATP and polynucleotide kinase. Labeled fragments were purified on a 7% nondenaturing preparatory polyacrylamide gel (5% crosslink) and isolated. Chemical sequencing reactions were performed according to published protocols (Iverson and Dervan, 1987; Maxam and Gilbert, 1980).

MPE•Fe(II) Footprinting. This procedure is also described in Trauger and Dervan, 2001. All reactions were carried out in a volume of 400 μL. A polyamide stock solution or water (for reference lanes) was added to an assay buffer where the final concentrations were: 28.6 mM HEPES, 285.7 mM NaCl buffer (pH 7.0), and 20 kcpm 3´- or 5'-radiolabelled DNA. The solutions were allowed to equilibrate for 12 hours. A fresh 5 μM MPE•Fe(II) solution was prepared from equal volumes of 10 μM MPE and 10 μM ferrous ammonium sulfate ($Fe(NH_4)_2(SO_4)_2$•$6H_2O$) solutions. MPE•Fe(II) solution (0.5 μM) was added to the equilibrated solutions, and the reactions were allowed to proceed for 15 min. Cleavage was initiated by the addition of dithiothreitol (5 mM) and allowed to proceed for 30 min. Reactions were stopped by ethanol precipitation, and the pellets were washed in 75% ethanol, dried *in vacuo*, resuspended in 15 mL H_2O, lyophilized to dryness, and resuspended in 100 mM tris-borate-EDTA/80% formamide loading buffer. The solutions were then denatured at 90 °C for 9 min, and a 5 μL sample (~ 8 kcpm) was loaded onto an 8% denaturing polyacrylamide gel (5% crosslink, 7 M urea) and run at 2000 V.

Affinity Cleaving. This procedure is also described in Trauger and Dervan, 2001. All reactions were carried out in a volume of 400 μL. A

polyamide stock solution or water (for reference lanes) was added to an assay buffer where the final concentrations were: 28.6 mM HEPES/85.7 mM NaCl buffer (pH 7.0), 200 mM NaCl, and 20 kcpm 3´- or 5'-radiolabelled DNA. The solutions were allowed to equilibrate for 12 hours. A fresh solution of ferrous ammonium sulfate ($Fe(NH_4)_2(SO_4)_2 \bullet 6H_2O$) (1 μM) was added to the equilibrated solutions, and the reactions proceeded for 15 min. Cleavage was initiated by the addition of dithiothreitol (5 mM) and allowed to proceed for 11 min. Reactions were stopped by ethanol precipitation, and the pellets were washed in 75% ethanol, dried *in vacuo*, resuspended in 15 mL H_2O, lyophilized to dryness, and resuspended in 100 mM tris-borate-EDTA/80% formamide loading buffer. The solutions were then denatured at 90 °C for 9 min, and a 5 μL sample (~ 8 kcpm) was loaded onto an 8% denaturing polyacrylamide gel (5% crosslink, 7 M urea) and run at 2000 V.

Quantitative DNase I Footprint Titrations. This procedure is described in detail in Trauger and Dervan, 2001. All reactions were carried out in a volume of 400 mL. We note explicitly that no carrier DNA was used in these reactions until after DNase I cleavage. A polyamide stock solution (or water for reference and intact lanes) was added to an assay buffer where the final concentrations were 10 mM Tris•HCl buffer (pH 7.0), 10 mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, and 20 kcpm 5′-radiolabelled DNA. The solutions were allowed to equilibrate for 18 hours at 22 °C. Cleavage was initiated by the addition of 10 mL of a DNase I solution (diluted with 1 mM DTT to 1.5 u/mL) and allowed to proceed for 7 min at 22 °C. The reactions were stopped by adding 50 mL of a solution containing 2.25M NaCl, 150 mM EDTA, 0.6 mg/mL glycogen, and 30 mM base-

pair calf thymus DNA, and then ethanol precipetated (2.1 volumes) at 14 krpm for 23 min. The pellets were washed with 75% ethanol, resuspended in 15 mL RNase-free water, lyophilized to dryness, and then resuspended in 100 mM tris-borate-EDTA/80% formamide loading buffer (with bromophenol blue as dye), denatured at 90 °C for 10 min, and loaded directly onto a pre-run 8% denaturing polyacrylamide gel (5% crosslink, 7 M urea) at 2000 V for 1.2 h. The gels were dried *in vacuo* at 80°C and then exposed to a storage phosphor screen (Molecular Dynamics).

Determination of Equilibrium Constants. At least three sets of data were used in determining each association constant. Data from the footprint titration gels were obtained using a Molecular Dynamics Typhoon 8600 PhosphorImager followed by quantitation using ImageQuant software (Molecular Dynamics). Background-corrected volume integration of rectangles encompassing the footprint sites and a reference site at which DNase I reactivity was invariant across the titration generated values for the site intensities (I_{site}) and the reference intensity (I_{ref}). The apparent fractional occupancy (θ_{app}) of the sites was calculated using the equation:

$$\theta app = 1 - \frac{Isite/Iref}{Isite^{\circ}/Iref^{\circ}} , \qquad (1)$$

where Isite^{o} and Iref^{o} are the site and reference intensities, respectively, from a control lane to which no polyamide was added.

The ($[L]_{tot}$, θ_{app}) data points were fit to a general Hill equation (eq 2) by minimizing the difference between θ_{app} and θ_{fit}:

$$\theta fit = \theta min + (\theta max - \theta min) \frac{Ka^n [L]^n tot}{1 + Ka^n [L]^n tot} , \quad (2)$$

where $[L]_{tot}$ is the total polyamide concentration, K_a is the apparent first-order association constant, and θ_{min} and θ_{max} are the experimentally determined site saturation values when the site is unoccupied or saturated, respectively. The data were fit using a nonlinear least-squares fitting procedure with K_a, θ_{max}, and θ_{min} as the adjustable parameters, and with either n=2 or n=1 depending on which value of n gave the better fit. We note explicitly that treatment of the data in this manner does not represent an attempt to model a binding mechanism. The binding isotherms were normalized using the following equation:

$$\theta norm = \frac{\theta app - \theta min}{\theta max - \theta min} \quad (3)$$

High Resolution Multidimensional NMR Studies

DNA Purification. The dimethoxytrityl-protected complementary oligonucleotides 5'-CCAAAGAGAAGCG-3' (purine strand) and 5'-CGCTTCTCTTTGG-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/min. 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 their UV absorbance at 260 nm using the calculated extinction coefficients, ε = 137,300 $M^{-1}cm^{-1}$ (purine strand) and ε = 107,200 $M^{-1}cm^{-1}$ (pyrimidine strand). 7.0 μmol of each strand was combined, and the single strands were hybridized by heating the 4.9 mL sample at 90 °C for 10 min., with slow cooling to room temperature over 6 hours. Duplex DNA was separated from 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 min gradient of 1.5 M KCl. The appropriate fractions were combined and concentrated using a Centricon YM-3, 3000 molecular-weight-cutoff filter from Amicon. The filter was used to dialyze the DNA sample, reducing the KCl concentration to approximately 1 μM. The sample was lyophilized to dryness and then dissolved in 10 mM sodium phosphate (pH 7.0), quantitated by UV absorbance, distributed to four 2.83 μmol aliquots, and then lyophilized to dryness. A 4.35 mM NMR sample was prepared by dissolving a 2.83 μmol aliquot of purified duplex DNA in 650 μL of 9:1 H_2O/D_2O.

1:1 Polyamide:DNA Complex Preparation. Polyamide concentration was estimated using the molar extinction coefficient, ε = 31,000. A polyamide stock solution was prepared by dissolving purified polyamide **13** in 10 mM sodium phosphate (pH 7.0) in 9:1 H_2O/D_2O. 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.

Final complex concentration was 3.67 mM. For experiments carried out in D_2O, the complex was lyophilized twice from D_2O.

NMR Experiments and Data Processing. NMR experiments were performed at 25 °C on a Varian INOVA 600 MHz spectrometer. 1D spectra and 2D NOESY spectra in 9:1 H_2O/D_2O were acquired using WATERGATE (Piotto et al., 1992) water suppression containing a W5 pulse element (Liu et al., 1998). Presaturation 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 D_2O. TOCSY spectra were acquired with mixing times of 40 and 100 ms, and NOESY spectra with mixing times of 50, 75, 100, and 150 ms. All 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 dueterated solvent. All spectra were recorded with $512t_1 * 2048t_2$ complex points. Data were processed on a SUN Ultra 10 workstation using VNMR (Varian, Inc.) or NMRPipe (Delaglio et al., 1995). Resonance assignment was performed using NMRVIEW 4.1.2 (Johnson and Blevins 1994). 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 crosspeak volumes from the 75 ms D_2O NOESY spectrum were converted to upper-bound distance constraints using the isolated spin-pair approximation, plus 20% as an error approximation (Barsukov and Lian, 1993). 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 (Borgias and James, 1990). NOE crosspeak intensities from the 150 ms D_2O 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 (Eis et al., 1997). 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 crosspeaks from the D_2O and H_2O NOESY spectra were assigned upper bound distance constraints of 5.5 and 6.0 Å, respectively, based on calculated distances for the smallest observable crosspeak intensities. Methylene and methyl groups were restrained as Q and M pseudoatoms, respectively, as defined by MARDIGRAS and AMBER 6.0 software (Borgias and James, 1990; Kollman et al., 2000).

Watson–Crick hydrogen bonding constraints were included on the basis of characteristic crosspeaks 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. N • • • O or N • • • N distances (Schmitz et al., 1997), as well as H • • • N and H • • • 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 (Smith et al., 1996; Eis et al., 1997). Restrained molecular dynamics (rMD) calculations were performed on a Dell Dimension 8100 workstation using the AMBER 6.0 software package (Kollman et al., 2000). All calculations were carried out *in vacuo* with a distance-dependent dielectric and a cutoff distance of 15.0 Å for nonbonded interactions. A force constant of 20 $kcalmol^{-1}A^{-2}$ was applied to interproton distances exceeding the upper and lower constraints, in a smooth parabolic fashion for 0.5 Å and then linearly.

Forty starting DNA structures with variable x-displacement, incline, rise, and twist values were constructed using Nucleic Acids Builder (NAB) software (Macke, 1996). 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 (Kollman et al., 2000), 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 accurately represent the input

data (Smith et al., 1996). Accordingly, the 12 structures of lowest violation energy were chosen as the final structural ensemble (Figure 27). 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 (Lavery and Sklenar, 1998; Stofer and Lavery, 1994). All helical parameters are listed and plotted in Appendix C. The molecular structures in Figures 27, 29, 30, 31, 32, and 33 were rendered in GRASP (Nicholls et al., 1991).

References

Baird, E. E. and Dervan, P. B. (1996) Solid Phase Synthesis of Polyamides Containing Imidazole and Pyrrole Amino Acids. *J. Am. Chem. Soc.*, **118**, 6141-6146.

Barsukov, I. L. and 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, Oxford University Press: New York, NY.

Borgias, B. A. and 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.

Bostock-Smith, C. E., Harris, S. A., Laughton, C. A., and 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.

Brenowitz, M., Senear, D. F., Shea, M. A., and Ackers, G. K. (1986) Quantitative DNase Footprint Titration – A Method for Studying Protein-DNA Interactions. *Methods Enzymol.*, **130**, 132-181.

Celda, B., Widmer, H., Leupin, W., Chazin, W. J., Denny, W. A., and Wütrich, K. (1989) Conformational Studies of d-$(AAAAATTTTT)_2$ Using Constraints from Nuclear Overhauser Effects and from Quantitative Analysis of the Cross-Peak Fine Structures in Two-Dimensional ^{1}H Nuclear Magnetic Resonance Spectra. *Biochemistry*, **28**, 1462-1471.

Chalikian, T. V., Plum, G. E., Sarvazyan, A. P., and 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.

Coll, M., Frederick, C. A., Wang, A. H.-J., and 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.

Crothers, D. M., and Shakked, Z. (2001) DNA Bending by Adenine-Thymine Tracts. In *Oxford Handbook of Nucleic Acid Structure* (Neidle, ed.), pp. 455-470, Oxford University Press, USA.

de Clairac, R. P. L., Geierstanger, B. H., Mrksich, M., Dervan, P. B., and Wemmer, D. E. (1997) NMR Characterization of Hairpin Polyamide Complexes with the Minor Groove of DNA. *J. Am. Chem. Soc.*, **119**, 7909-7916.

de Clairac, R. P. L., Seel, C. L., Geierstanger, B. H., Mrksich, M., Baird, E. E., Dervan, P. B., and 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.

Delaglio, F., Grzesiek, S.,. Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPIPE–A Multidimensional Spectral Processing System Based on UNIX Pipes. *J. Biomol. NMR*, **6**, 277-293.

Dervan, P. B. (2001) Molecular Recognition of DNA by Small Molecules. *Bioorg. Med. Chem.* **9**, 2215-2235.

Dervan, P. B. and 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., and Kisakürek, M. V., eds.), pp. 327-339, Verlag Helvetica Chimica Acta, Zurich.

Dickerson, R. E. and Drew, H. R. (1981) Structure of a B-DNA Dodecamer II. Influence of Base Sequence on Helix Structure. *J. Mol. Biol.*, **149**, 761-786.

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.

Dickinson, L. A., Gulizia, R. J., Trauger, J. W., Baird, E. E., Mosier, D. E., Gottesfeld, J. M., and Dervan P. B. (1998) Inhibition of RNA Polymerase II Transcription in Human Cells by Synthetic DNA-Binding Ligands. *Proc. Natl,. Acad. Sci. USA*, **95**, 12890-12895.

Dwyer, T. J., Geierstanger, B. H., Bathini, Y., Lown, J. W., and 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.

Dwyer, T. J., Geierstanger, B. H., Mrksich, M., Dervan, P. B., and Wemmer, D. E. (1993) Structural Analysis of Covalent Peptide Dimers, Bis(pyridine-2-carboxamidonetropsin)$(CH_2)_{3-6}$, in Complex with 5'-TGACT-3' Sites by Two-Dimensional NMR. *J. Am. Chem. Soc.*, **115**, 9900-9906.

Eis, P. S., Smith, J. A., Rydzewski, J. M., Case, D. A., and 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.

Ellenberger, T. E., Brandl, C. J., Struhl, K., and Harrison, S. C. (1992) The GCN4 Basic Region Leucine Zipper Binds DNA as a Dimer of Uninterrupted Alpha-Helices – Crystal-Structure of the Protein-DNA complex. *Cell*, **71**, 1223-1237.

Fratini, A. V., Kopka, M. L., Drew, H. R., and Dickerson, R. E. (1982) Reversible Bending and Helix Geometry in a B-DNA Dodecamer–CGCGAATTBrCGCG. *J. Biol. Chem.*, **257**, 14686-14707.

Gao, X. L., Mirau, P., and Patel, D. J. (1992) Structure Refinement of the Chromomycin Dimer-DNA Oligomer Complex in Solution. *J. Mol. Biol.*, **223**, 259-279.

Geierstanger, B. H., Mrksich M., Dervan, P. B., and Wemmer, D. E. (1994) Design of a G•C-Specific DNA Minor Groove-Binding Peptide. *Science,* **266**, 646-650.

Geierstanger, B. H., Mrksich, M., Dervan, P. B., and Wemmer, D. E. (1996) Extending the Recognition Site of Designed Minor Groove Binding Molecules. *Nat. Struct. Biol.,* **3**, 321-324.

Goodsell, D. S., Ng, H. L., Kopka, M. L., Lown, J. W., and Dickerson, R. E. (1995) Structure of a Dicationic Monoimidazole Lexitropsin Bound to DNA. *Biochemistry*, **34**, 16654-16661.

Han, G. W., Kopka, M. L., Cascio, D., Grzeskowiak, K., and 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.

Herman, D. M.; Baird, E. E.; Dervan, P. B. (1998) Stereochemical Control of the DNA Binding Affinity, Sequence Specificity, and Orientation Preference of Chiral Hairpin Polyamides in the Minor Groove. *J. Am. Chem. Soc.*, **120**, 1382-1391.

Herman, D. M. (2001) Ph. D. Thesis, California Institute of Technology, USA.

Gottesfeld, J. M., Neely, L., Trauger, J. W., Baird, E. E., and Dervan, P. B. (1997) Regulation of Gene Expression by Small Molecules. *Nature*, **387**, 202-205.

Hunter, C, A. (1993) Sequence-Dependent DNA Structure: The Role of Base Stacking Interactions. *J. Mol. Biol.,* **230**, 1025-1054.

Iverson, B. L. and Dervan, P. B. (1987) Adenine Specific DNA Chemical Sequencing Reaction. *Nucleic Acids Res.*, **15**, 7823-7830.

Janssen, S., Durussel, T., and Laemmli, U. K. (2000a) Chromatin Opening of DNA Satellites by Targeted Sequence-Specific Drugs. *Mol. Cell,* **6**, 999-1011.

Janssen, S., Cuvier, O., Muller, M., Laemmli, U. K. (2000b) Specific Gain- and Loss-of-Function Phenotypes Induced by Satellite-Specific DNA-Binding Drugs Fed to *Drosophila melanogaster*. *Mol. Cell*, **6**, 1013-1024.

Johnson, B. A. and Blevins, R. A. (1994) NMR VIEW – A Computer-Program for the Visualization and Analysis of NMR Data. *J. Biomol. NMR*, **4**, 603-614.

Kamitori, S., and Takusagawa, F. (1992) Crystal-Structure of the 2/1 Complex Between d(GAAGCTTC) and the Antitumor Drug Actinomycin-D. *J. Mol. Biol.*, **225**, 445-456.

Kelly, J. J., Baird, E. E., and Dervan, P. B. (1996) Binding Site Size Limit of the 2:1 Pyrrole-Imidazole Polyamide-DNA Motif. *Proc. Natl. Acad. Sci. USA*, **93**, 6981-6985.

Kielkopf, C. L., Baird, E. E., Dervan, P. B., and Rees, D. C. (1998a) Structural Basis for G•C Recognition in the DNA Minor Groove. *Nat. Struct. Biol.,* **5**, 104-109.

Kielkopf, C. L., White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., Dervan, P. B., and Rees, D. C. (1998b) A Structural Basis for Recognition of A•T and T•A Base Pairs in the Minor Groove of B-DNA. *Science,* **282**, 111-115.

Kielkopf, C. L., Bremer, R. E., White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., Dervan, P. B., and Rees, D. C. (2000) Structural Effects of DNA Sequence on T•A Recognition by Hydroxypyrrole/Pyrrole Pairs in the Minor Groove. *J. Mol. Biol.*, **295**, 557-567.

Kim, Y., Geiger, J. H., Hahn, S., and Sigler, P. B. (1993) Crystal-Structure of a Yeast TBP TATA-Box Complex. *Nature*, **365**, 512-520.

Kollman, P. A., Case, D. A., Merz, K., Cheatham, T. E., III, Simmerling, C., Darden, T., Pearlman, D. A.; Caldwell, J. W., Ross, W. S., Ferguson, D. M., Seibel, G. L., Singh, U. C., and Weiner, P. K. (2000) AMBER 6.0. University of California, San Francisco, CA.

Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P., and Dickerson, R. E. (1985) The Molecular Origin of DNA-Drug Specificity in Netropsin and Distamycin. *Proc. Natl. Acad. Sci. USA,* **82**, 1376-1380.

Kopka, M. L., Goodsell, D. S., Han, G. W., Chiu, T. K., Lown, J. W., and Dickerson, R. E. (1997) Defning 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.

Lavery, R. and Sklenar, H. (1988) The Definition of Generalized Helicoidal Parameters and of Axis Curvature for Irregular Nucleic Acids. *J. Biomol. Struct. Dynam.*, **6**, 63-91.

Liu, M. L., Mao, X. A., Ye, C. H., Huang, H., Nicholson, J. K., and Lindon, J. C. (1998) Improved WATERGATE Pulse Sequences for Solvent Suppression in NMR Spectroscopy. *J. Magn. Reson.,* **132**, 125-129.

Love, J. J., Li, X., Case, D. A., Giese, K., Grosschedl, R., and Wright, P. E. (1995) Structural Basis for DNA Bending by the Architectural Transcription Factor LEF-1. *Nature*, **376**, 791-795.

Lown, J. W., Krowicki, K., Bhat, U. G., Skorobogaty, A., Ward, B., and Dabrroiak, 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.

MacDonald, D., Herbert, K., Zhang, X., Polgruto, T., and Lu, P. (2001) Solution Structure of an A-Tract DNA Bend. *J. Mol. Biol.*, **306**, 1081-1098.

Macke, T. (1996) Ph.D. Thesis, The Scripps Research Institute.

Mapp, A. K., Ansari, A. Z., Ptashne, M., and Dervan, P. B. (2000) Activation of Gene Expression by Small Molecule Transcription Factors. *Proc. Natl. Acad. Sci. USA*, **97**, 3930-3935.

Marques, M. A., Doss, R. M., Urbach, A. R., and Dervan, P. B. (in preparation) Towards an Understanding of the Chemical Etiology for DNA Minor Groove Recognition by Polyamides.

Maxam, A. M. and Gilbert, W.S. (1980) Sequencing End-Labeled DNA with Base-Specific Chemical Changes. *Methods Enzymol.*, **65**, 499-560.

Mrksich, M., Wade, W. S., Dwyer, T. J., Geierstanger, B. H., Wemmer, D. E., and Dervan, P. B. (1992) Antiparallel side-by-side Dimeric Motif for Sequence-Specific Recognition of the Minor Groove of DNA by the Designed Peptide 1-Methyimidazole-2-carboxamide Netropsin. *Proc. Natl. Acad. Sci. USA*, **89**, 7586-7590.

Mrksich, M., Parks, M. E., and Dervan, P. B. (1994) Hairpin Peptide Motif. A New Class of Oligopeptides for Sequence-Specific Recognition in the Minor Groove of Double-Helical DNA. *J. Am. Chem. Soc.*, **116**, 7983-7988.

Nadeau, J. G. and Crothers, D. M. (1989) Structural Basis for DNA Bending. *Proc. Natl. Acad. Sci. USA*, **86**, 2622-2626.

Neidle, S., ed. (1999) *Oxford Handbook of Nucleic Acid Structure*, Oxford University Press, New York, USA.

Nicholls, A., Sharp, K., and Honig, B. (1991) GRASP–Graphical Representation and Analysis of Structural Properties. *PROTEINS, Structure, Function and Genetics*, **11**, 281ff.

Paloma, L. G., Smith, J. A., Chazin, W. J., and Nicolaou, K. C. (1994) Interaction of Calicheamicin with Duplex DNA–Role of the Oligosaccharide Domain and Identification of Multiple Binding Modes. *J. Am. Chem. Soc.*, **116**, 3697-3708.

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.

Pavletich, N. P., Pabo, C. O. (1991) Zinc Finger DNA Recognition–Crystal Structure of a ZIF268-DNA Complex at 2.1 Å. *Science*, **252**, 809-817.

Pelton, J. G. and Wemmer, D. E. (1988) Structural Modeling of the Distamycin A-d(C G C G A A T T C G C G)$_2$ Complex Using 2D NMR and Molecular Mechanics. *Biochemistry*, **27**, 8088-8096.

Pelton, J. G. and Wemmer, D. E. (1989) Structural Characterization of a 2:1 Distamycin A•d(CGCAAATTGGC) Complex by Two-Dimensional NMR. *Proc. Natl. Acad. Sci. USA*, **86**, 5723-5727.

Piotto, M., Saudek, V., and Sklenár, V. (1992) Gradient-Tailored Excitation for Single-Quantum NMR-Spectroscopy of Aqueous Solutions. *J. Biomol. NMR,* **2**, 661-665.

Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning*, Cold Spring Harbor Laboratory Press, Plainview, New York, USA.

Schmitz, U., Sethson, I., Egan, W. M., and 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.

Schnell, J. R., Ketchem, R. R., Boger, D. L., and 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.

Schultz, P. G. and Dervan, P. B. (1984) Distamycin and Penta-N-Methylpyrrolecarboxamide binding Sites on Native DNA–A Comparison of Methidium Propyl-EDTA and Affinity Cleaving. *J. Biomol. Struct. Dyn.*, **1**, 1133-1147.

Shatzky-Schwartz, M., Arbuckle, N. D., Eisenstein, M., Rabinovich, D., Bareket-Samish, A., Haran, T. E., Luisi, B. F., and Zippora, S. (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.

Smith, J. A., Gomez-Paloma, L. G., Case, D. A., and Chazin, W. J. (1996) Molecular Dynamics Docking Driven by NMR-Derived Restraints to Determine the Structure of the Calicheamicin Gamma (I) (l) Oligosaccharide Domain Complexed to Duplex DNA. *Magn. Reson. Chem.* **34S**, 147-155.

Stofer, E. and Lavery, R. (1994) Measuring the Geometry of DNA Grooves. *Biopolymers* **34**, 337-346.

Swalley, S. E., Baird, E. E., and Dervan, P. B. (1997) A Pyrrole-Imidazole Polyamide Motif for Recognition of Eleven Base Pair Sequences in the Minor Groove of DNA. *Chem. Eur. J.*, **3**, 1600-1607.

Taylor, J. S., Schultz, P. G., and 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.

Trauger, J. W., Baird, E. E., Mrksich, M., and Dervan, P. B. (1996a) 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.

Trauger, J. W., Baird, E. E., and Dervan, P. B. (1996b) Recognition of DNA by Designed Ligands at Subnanomolar Concentrations. *Nature*, **382**, 559-561.

Trauger, J. W., Baird, E. E., and Dervan, P. B. (1996c) Extended Hairpin Polyamide Motif for Sequence-Specific Recognition in the Minor Groove of DNA. *Chem. Biol.*, **3**, 369-377.

Trauger, J. W. and Dervan, P. B. (2001) Footprinting Methods for Analysis of Pyrrole-Imidazole Polyamide/DNA Complexes. *Methods Enzymol.*, **340**, 450-466.

Turner, J. M., Swalley, S. E., Baird, E. E., and 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. *J. Am. Chem. Soc.*, **121**, 11621-11629.

Urbach, A. R., Szewczyk, J. W., White, S., Turner, J. M., Baird, E. E., and Dervan, P. B. (1999) Sequence Selectivity of 3-Hydroxypyrrole/Pyrrole Ring Pairings in the DNA Minor Groove. *J. Am. Chem. Soc.*, **121**, 11621-11629.

Urbach, A. R. and Dervan, P. B. (2001) Toward Rules for 1:1 Polyamide:DNA Recognition. *Proc. Natl. Acad. Sci. USA,* **98**, 4343-4348.

Urbach, A. R., Love, J. J., Ross, S. A., and Dervan, P. B. (2002) Structure of a β-Alanine Linked Polyamide Bound to a Full Helical Turn of Purine Tract DNA in the 1:1 Motif. *J. Mol. Biol.*, in press.

Urbach, A. R., Marques, M. A., Doss, R. M., and Dervan, P. B. (in preparation) Linker-Dependent Conformational Control of Polyamide-DNA Binding Modes.

Van Dyke, M. W. and Dervan, P. B. (1983) Chromomycin, Mithramycin, and Olivomycin Binding Sites on Heterogeneous Deoxyribonucleic-Acid – Footprinting with (Methidiumpropyl-EDTA)Iron(II). *Biochemistry*, **22**, 2373-

Venter, J. C. et al. (2001) The Sequence of the Human Genome. *Science*, **291**, 1304-1351.

Wade, W. S., Mrksich, M., and 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., and Dervan, P. B. (1998) Recognition of the Four Watson-Crick Base Pairs in the DNA Minor Groove by Synthetic Ligands. *Nature*, **391**, 468-471.

Woods, C. R., Ishii, T., Wu, B., Bair, K. W., and Boger, D. L. (2002) Hairpin versus Extended DNA Binding of a Substituted β-Alanine Linked Polyamide. *J. Am. Chem. Soc.*, **124**, 2148-2152.

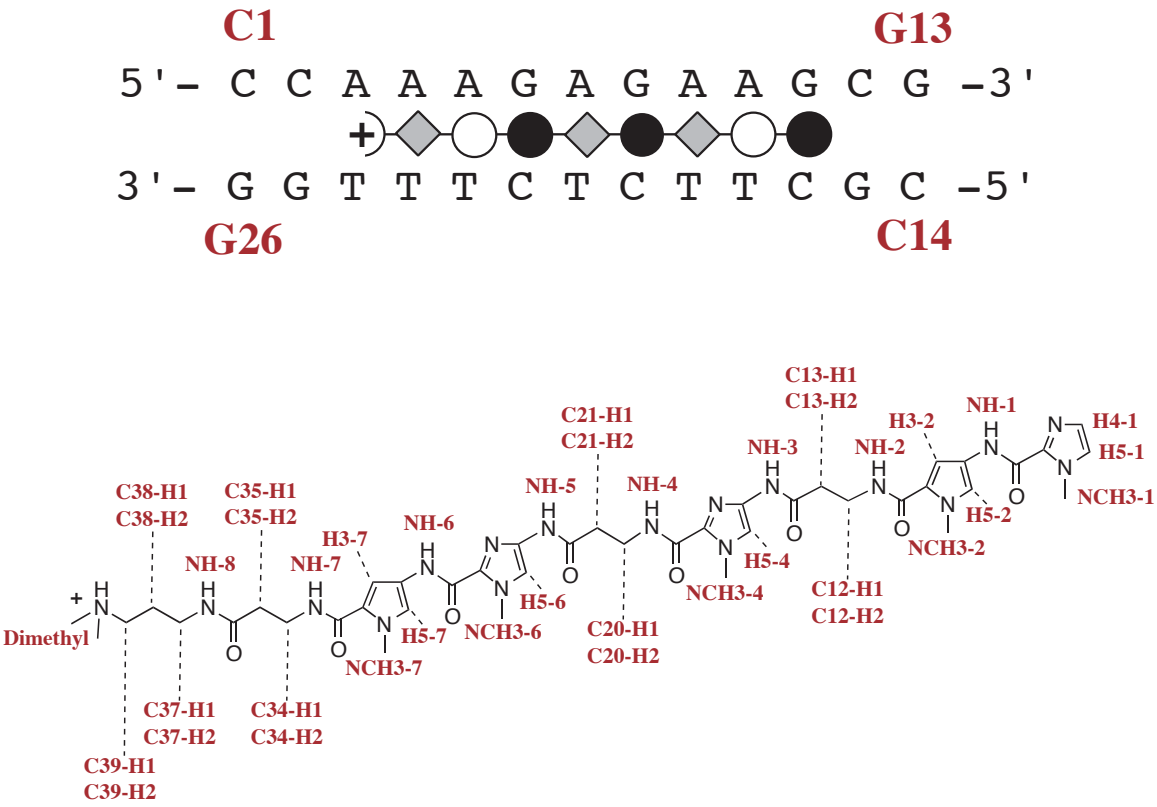
Wütrich, K. (1986) *NMR of Proteins and Nucleic Acids*. Wiley Interscience, USA.

Youngquist, R. S. and Dervan, P. B. (1987) A Synthetic Peptide Binds 16 Base Pairs of A,T Double Helical DNA. *J. Am. Chem. Soc.* **109**, 7564-7566.

Zimmer, C. and Wahnert, U. (1986) Nonintercalating DNA-Binding Ligands – Specificity of the Interaction and Their Use as Tools in Biophysical, Biochemical, and Biological Investigations of the Genetic Material. *Prog. Biophys. Mol. Biol.*, **47**, 31-112.

Appendix A

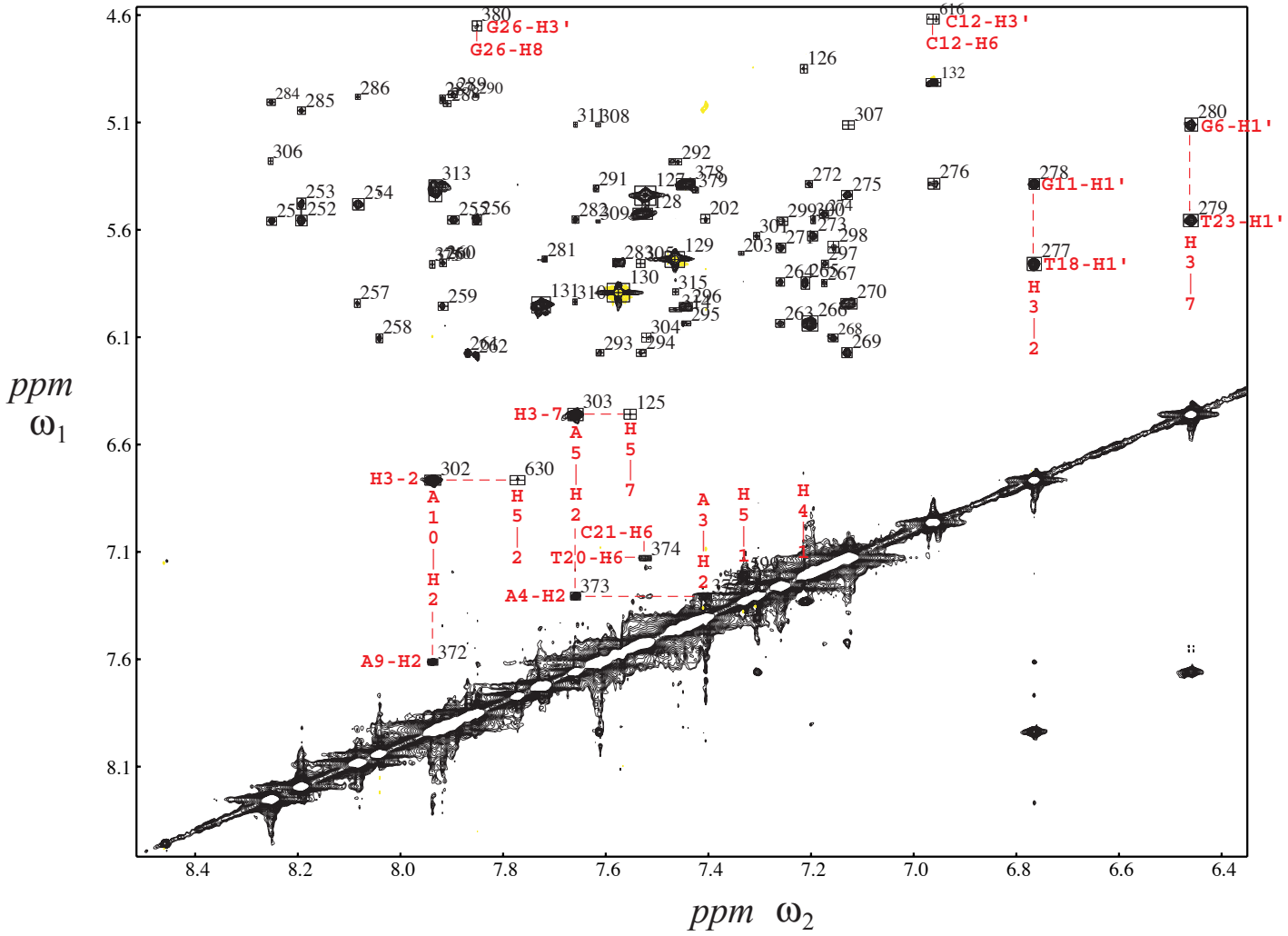
NMR Spectra



Expanded and annotated NOESY spectra in D2O and H2O (90%) are provided here for the 1:1 polyamide:DNA complex illustrated above. The binding model shown at top gives DNA base numbers used in the assignments, as well as the orientation and binding site of the ligand. The chemical structure drawn below is annotated with all proton nomenclature. Due to the similarity of the correlations assigned in each plot, assignments are typically annotated with the DNA residue number, and the correlation type is given in the title of the plot. Special legends are provided for complicated regions. Experimental conditions and NOESY mixing time is given above each plot. Eleven regions of the D2O NOESY spectrum are given prior to nine regions of the H2O NOESY spectrum. A 75 ms mixing time was used in both experiments.

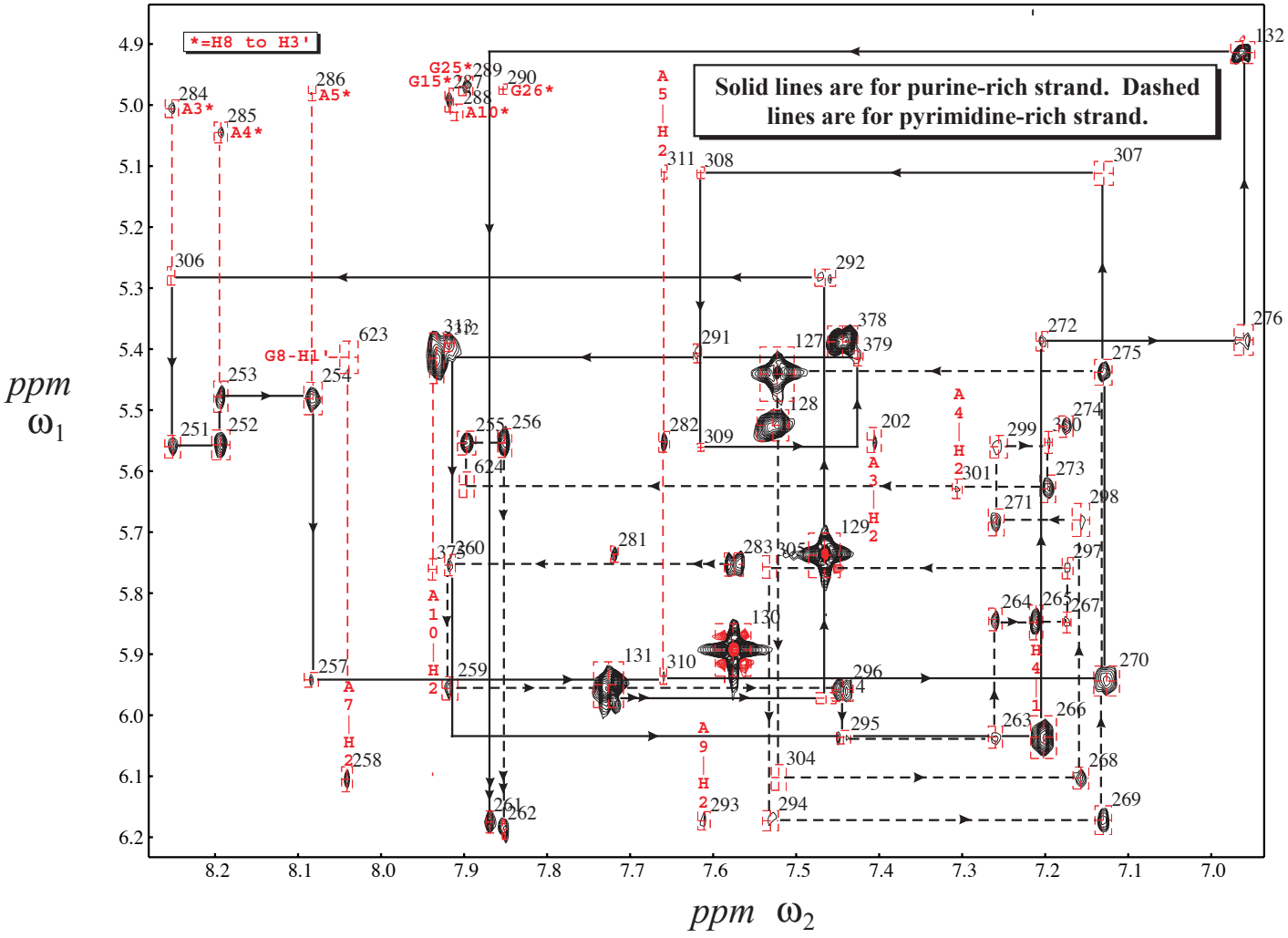
Aromatic Region of NOESY spectrum in D_2O

(t_{mix} = 75 ms., 600 MHz, 25℃)



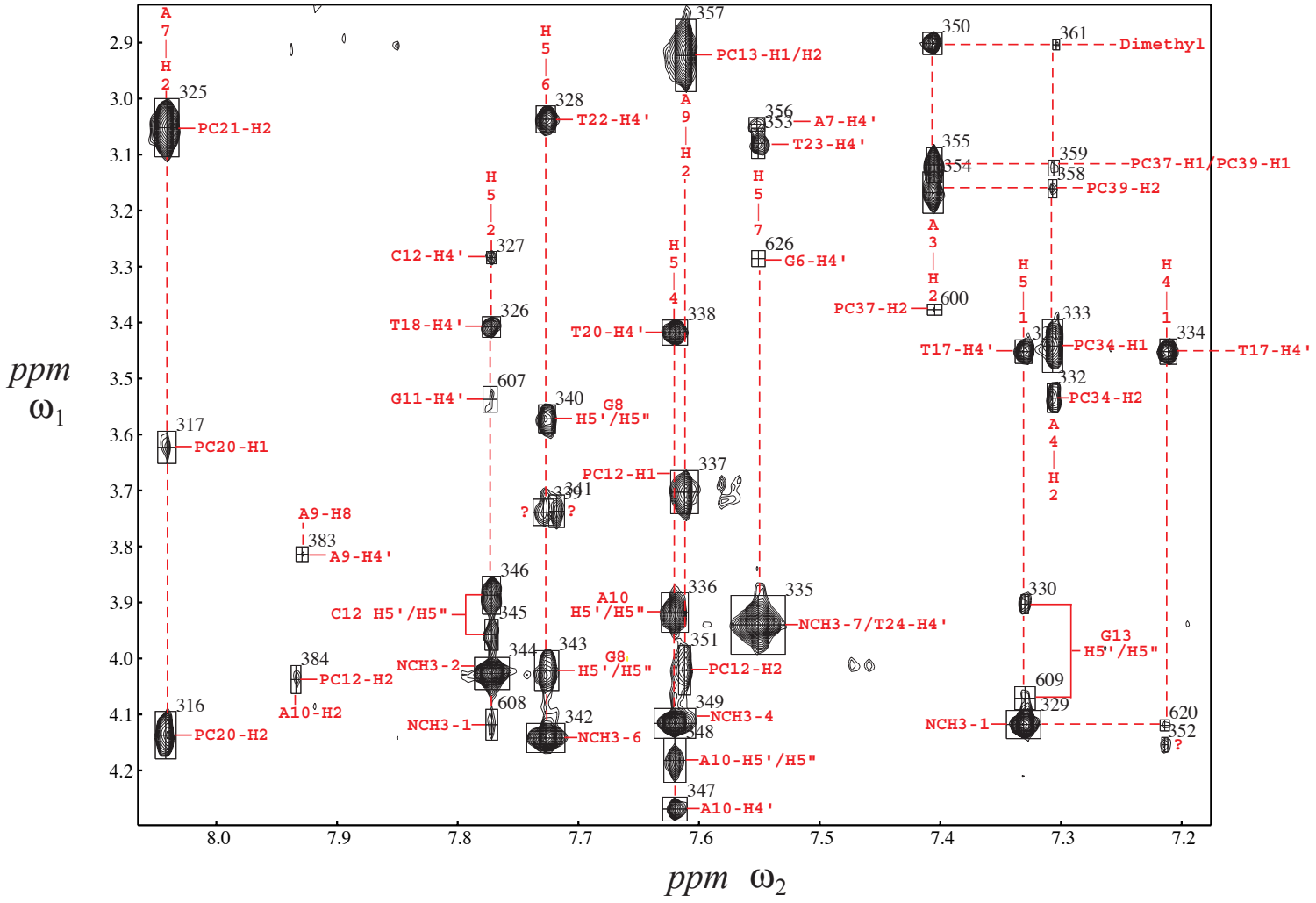
H1' to Aromatic Region of NOESY spectrum in D_2O

(t_{mix} = 75 ms., 600 MHz, 25°C)



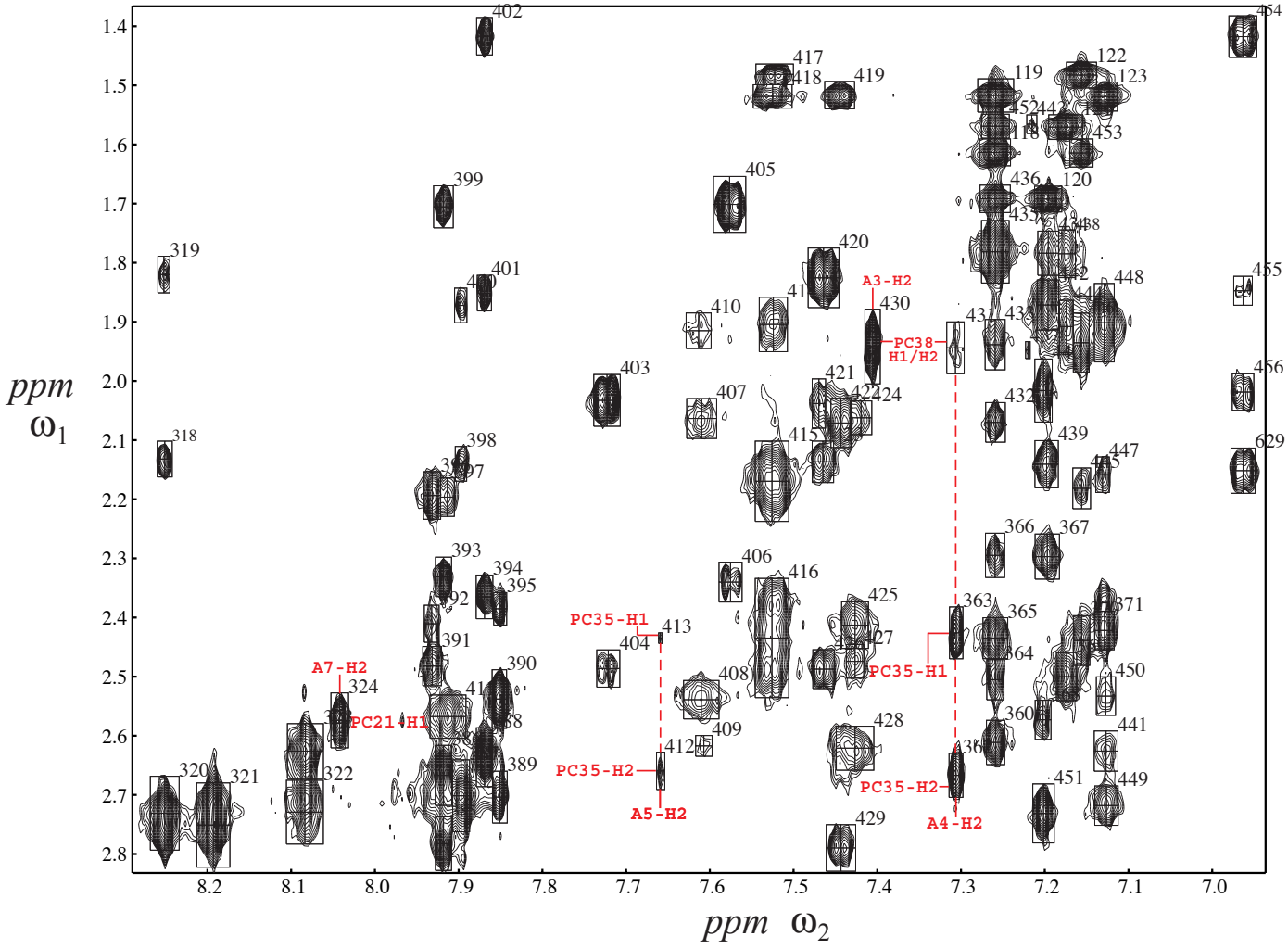
Aromatic to H4' and NCH_3 Region of NOESY spectrum in D_2O

(t_{mix}= 75 ms., 600 MHz, 25°C)



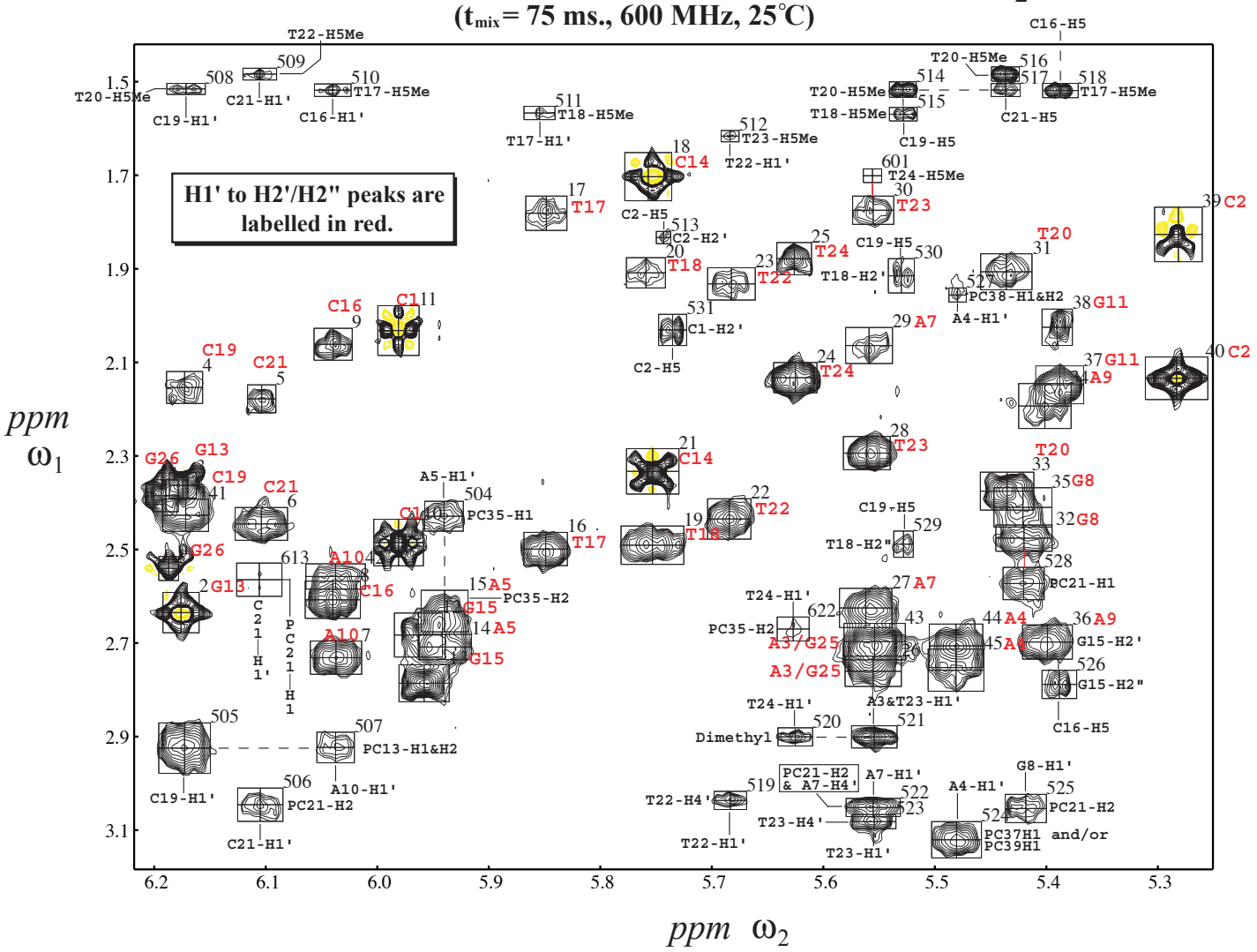
Aromatic to H2'/H2" Region of NOESY spectrum in D_2O

(t_{mix}= 75 ms., 600 MHz, 25°C)



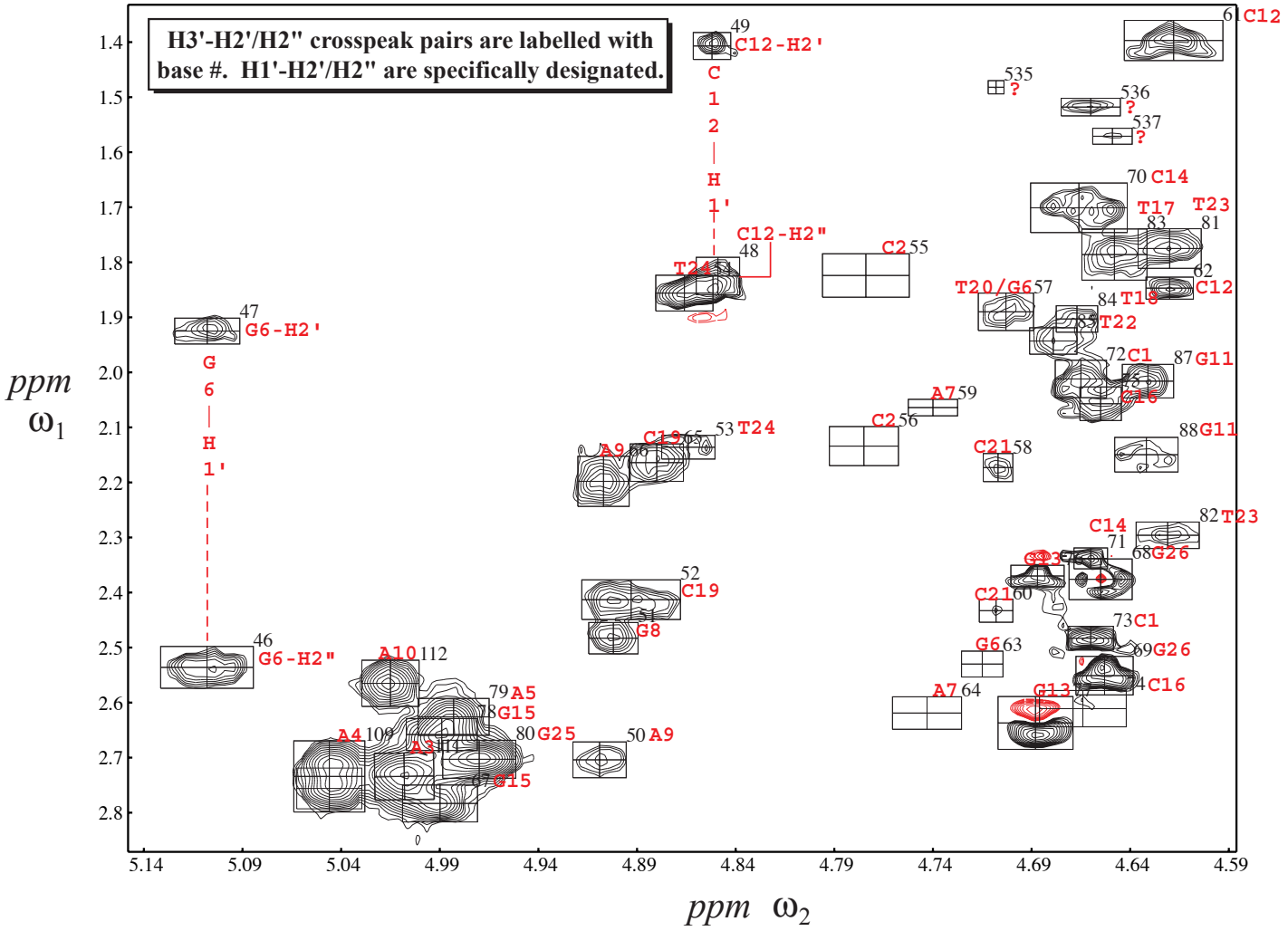
H1' to H2'/H2" Region of NOESY spectrum in D_2O

(t_{mix} = 75 ms., 600 MHz, 25°C)



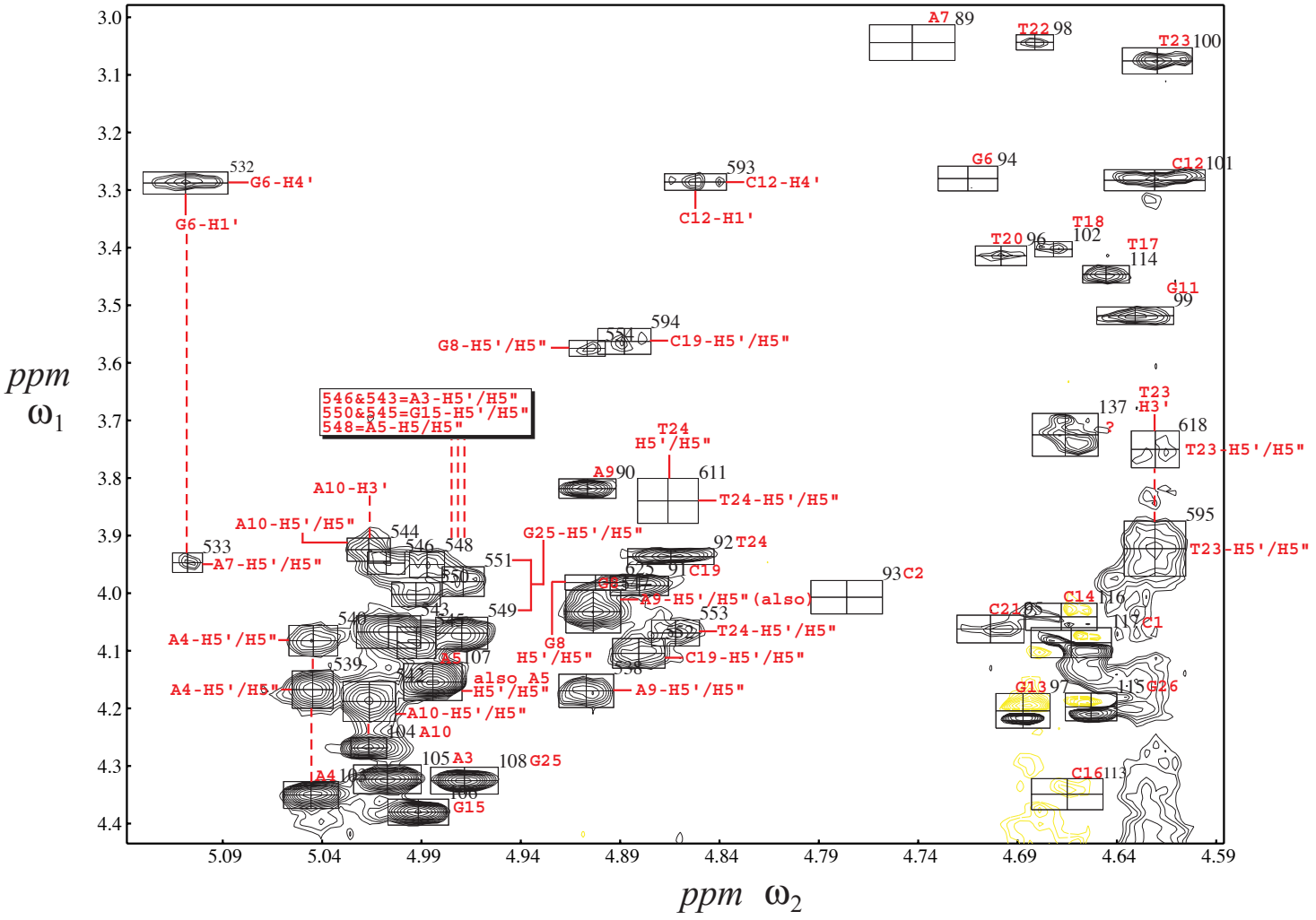
H3' to H2'/H2" Region of NOESY spectrum in D_2O

(t_{mix} = 75 ms., 600 MHz, 25°C)



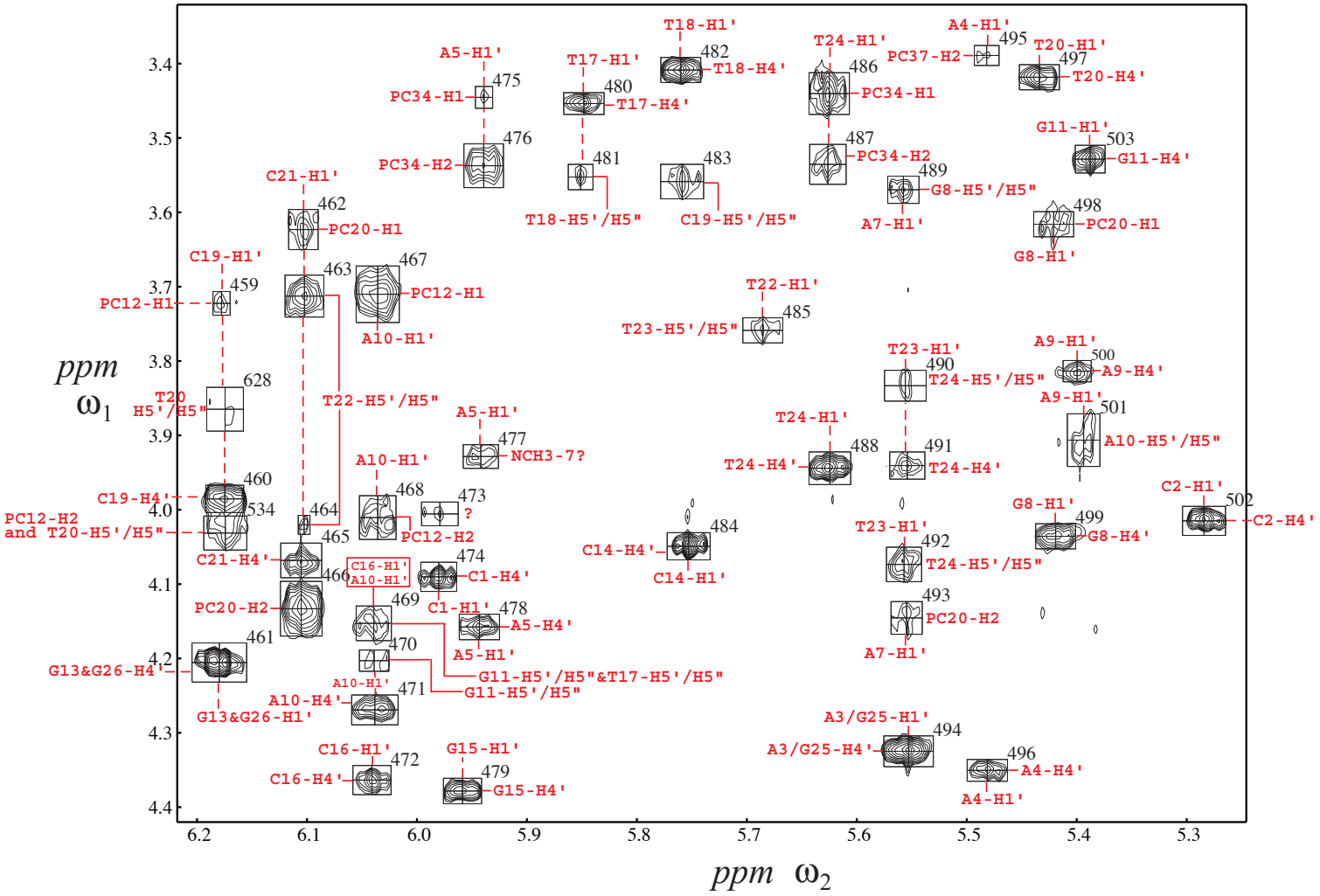
H3' to H4'/H5'/H5" Region of NOESY spectrum in D_2O

(t_{mix}= 75 ms., 600 MHz, 25℃)



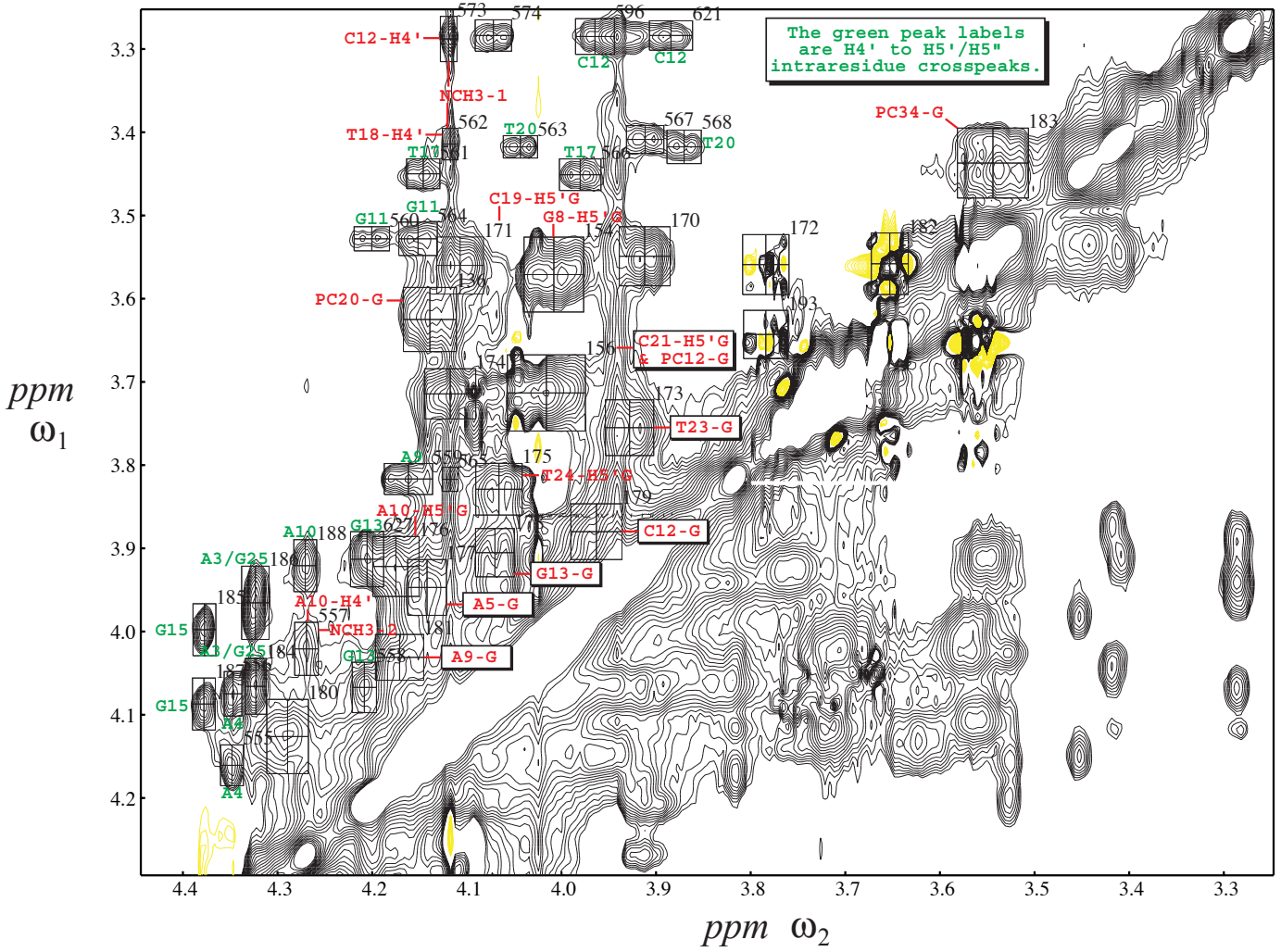
H1' to H4' Region of NOESY spectrum in D_2O

(t_{mix} = 75 ms., 600 MHz, 25°C)



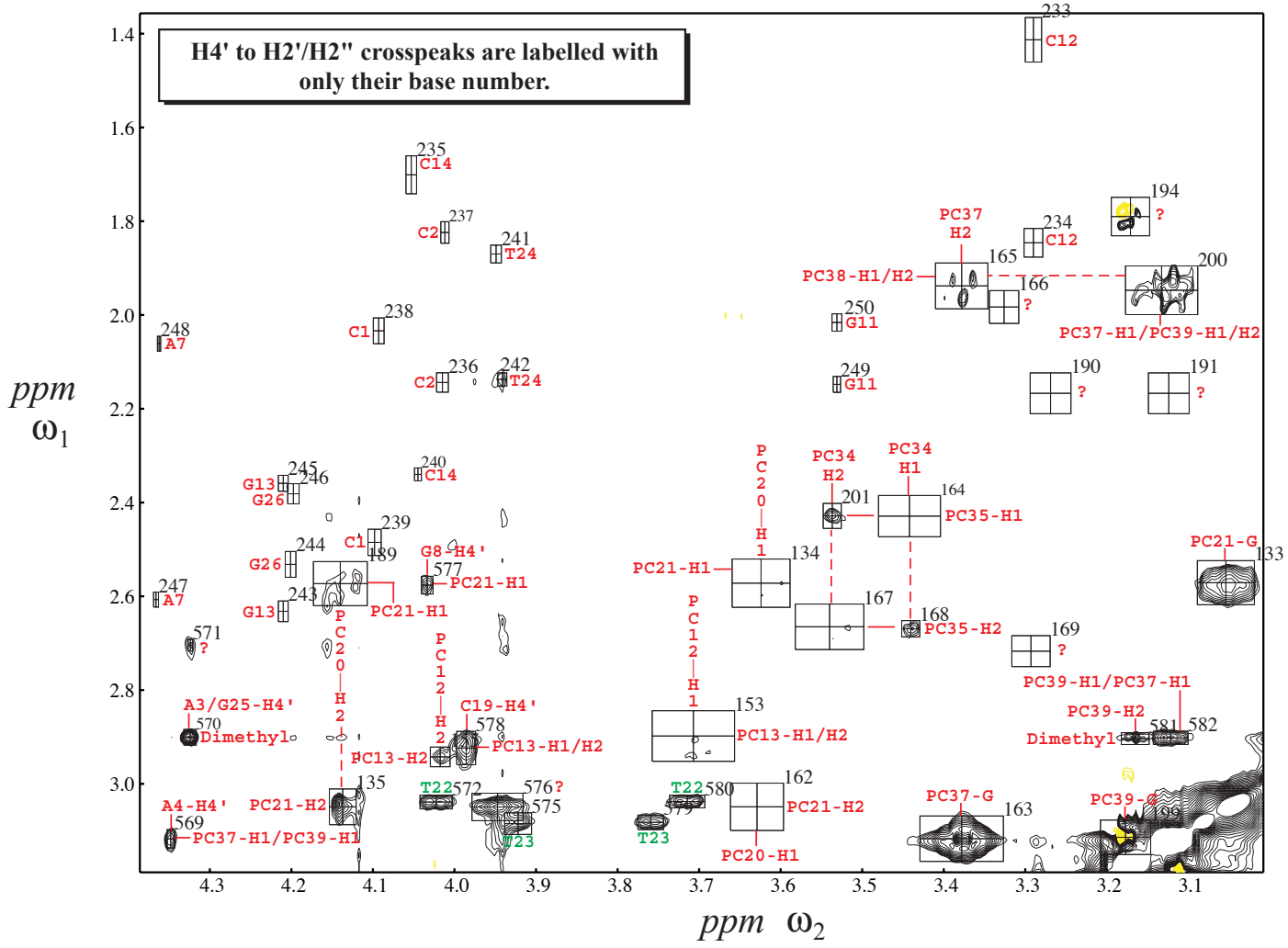
H5' to H5" Region of NOESY spectrum in D_2O

(t_{mix} = 75 ms., 600 MHz, 25°C)



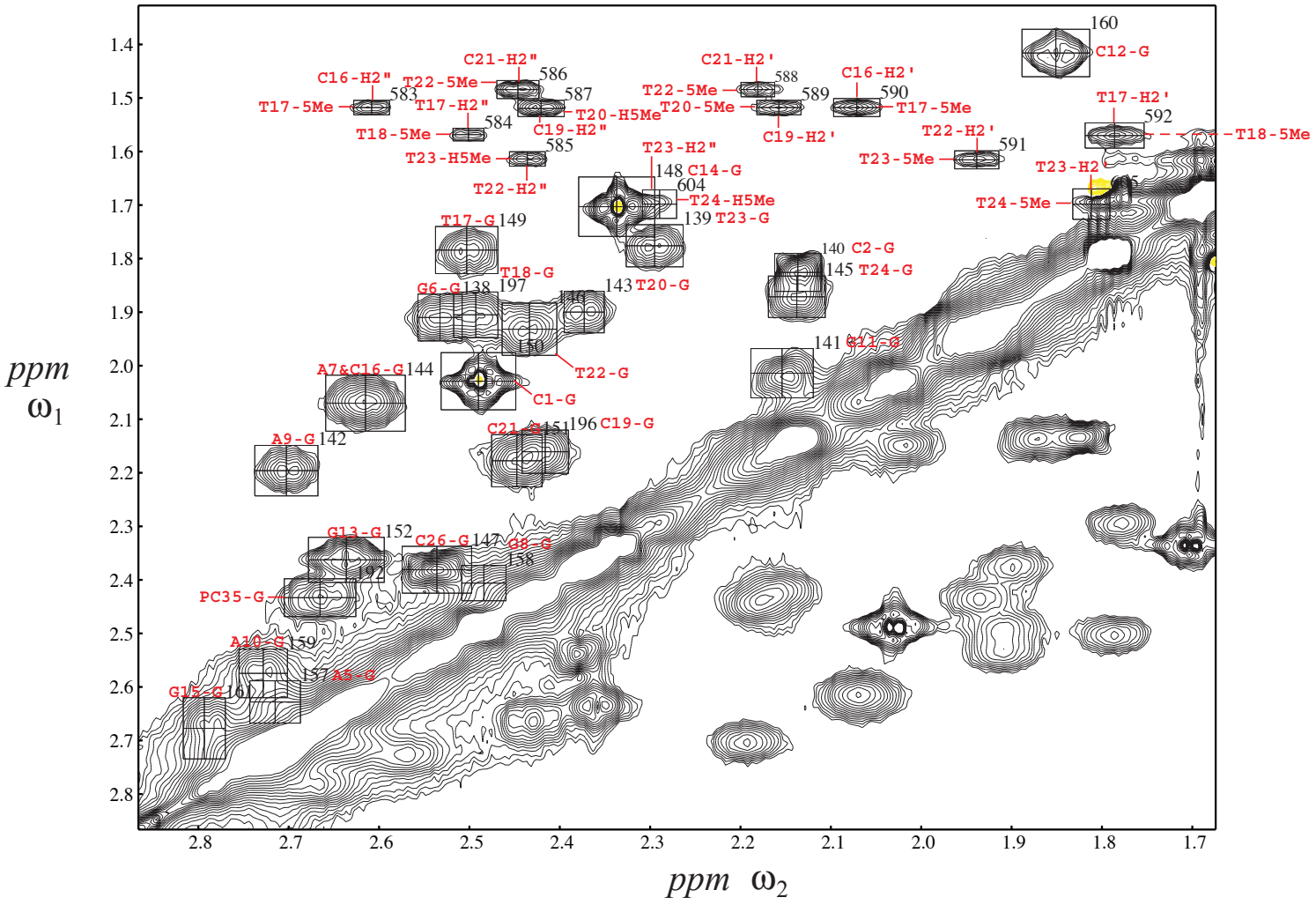
Methylene Region of NOESY spectrum in D_2O

(t_{mix} = 75 ms., 600 MHz, 25°C)



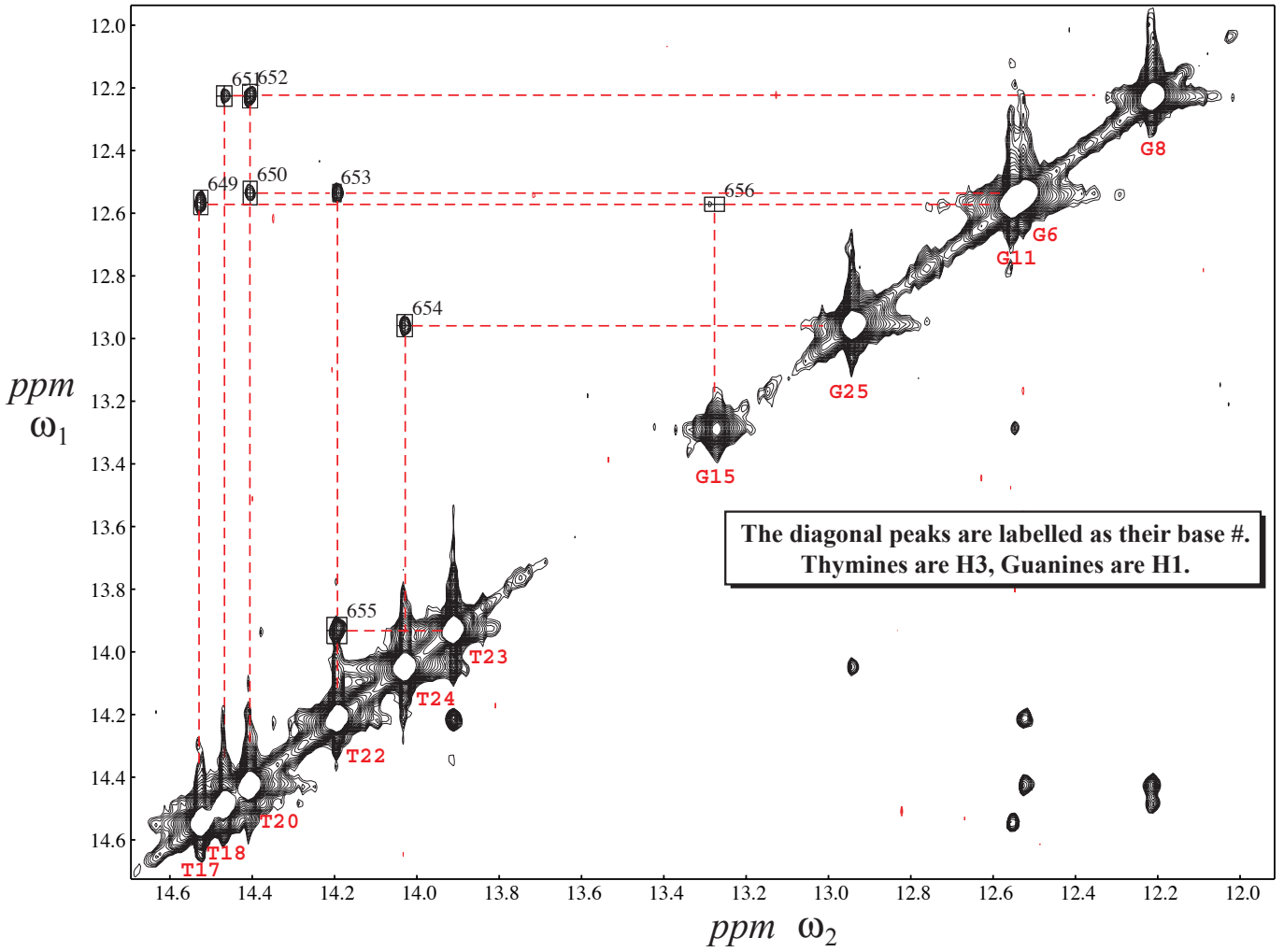
H2' to H2" Region of NOESY spectrum in D_2O

(t_{mix} = 75 ms., 600 MHz, 25℃)



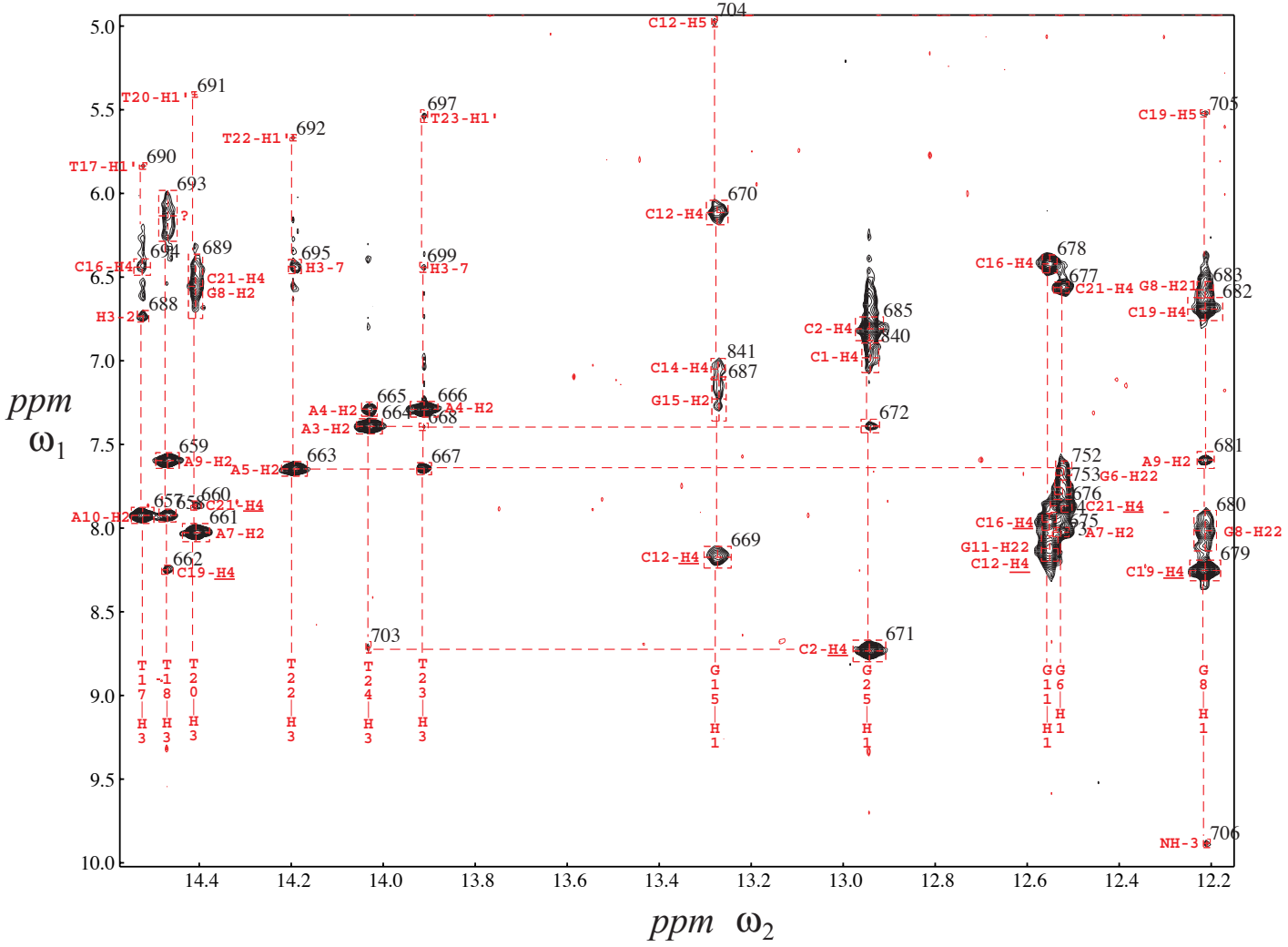
Sequential Imino Region of NOESY spectrum in H_2O

(t_{mix}= 75 ms., 600 MHz, 25°C)



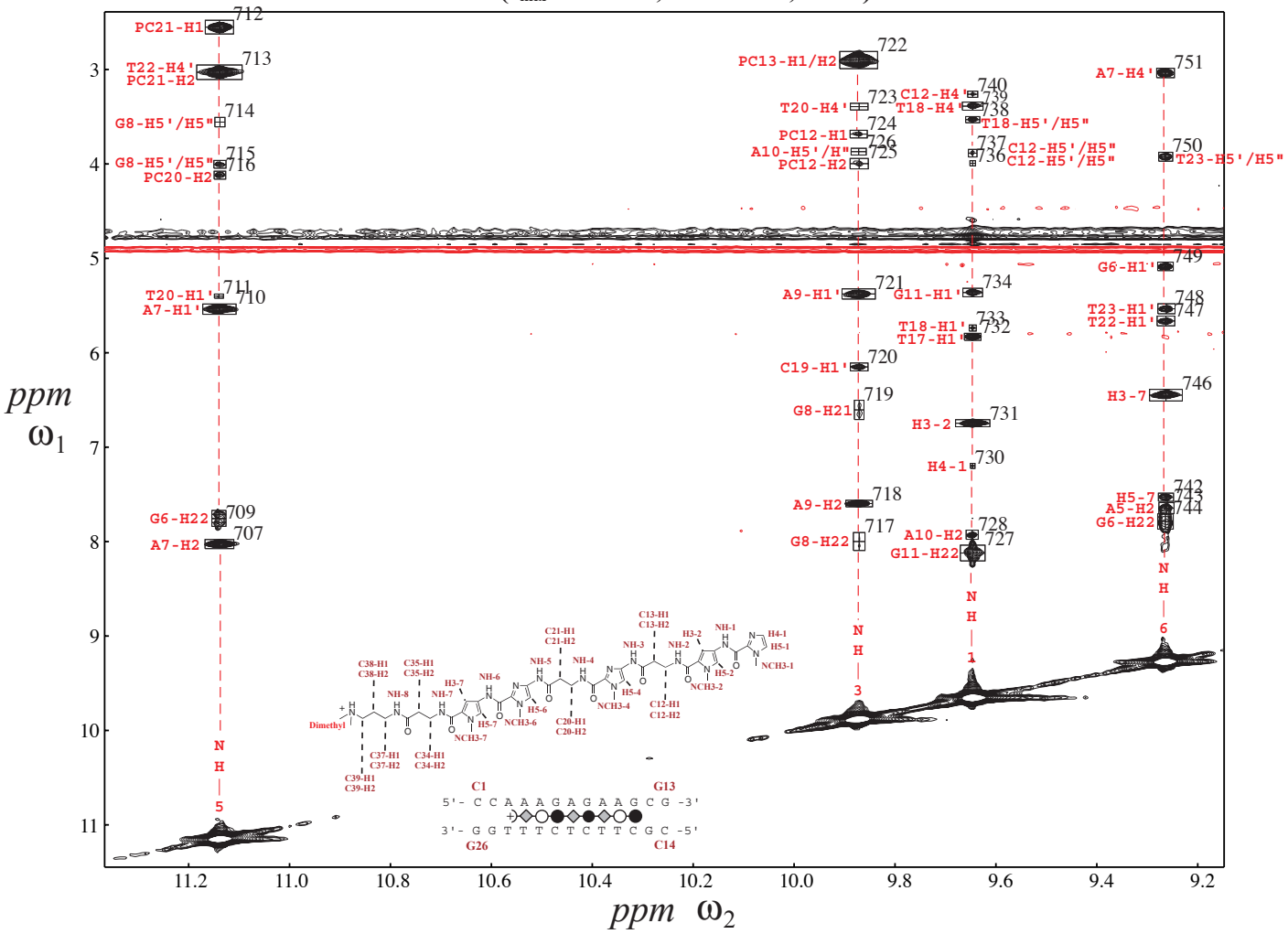
Imino to Amino and AH2 Region of NOESY spectrum in H_2O

(t_{mix}= 75 ms., 600 MHz, 25°C)



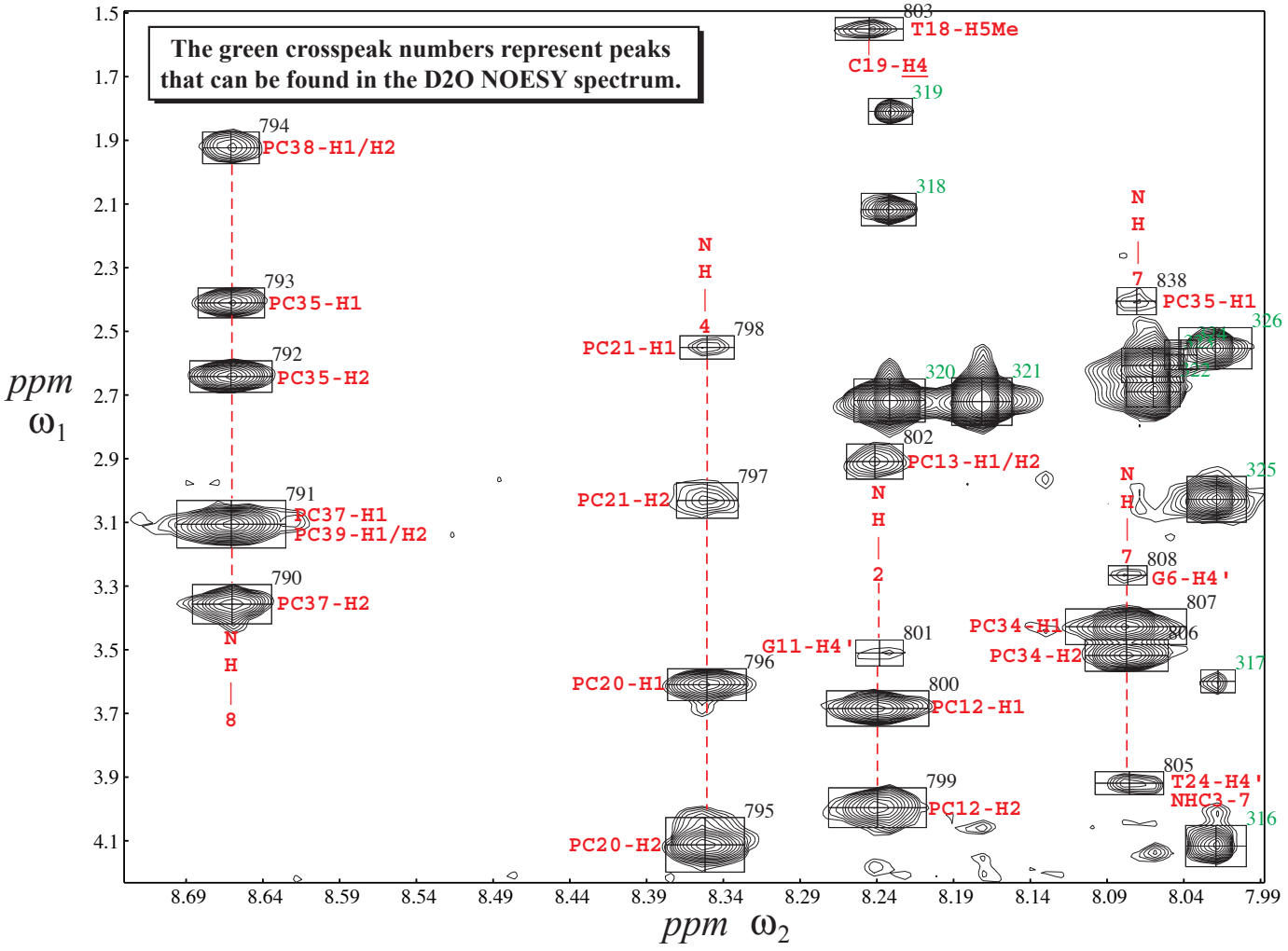
Aromatic Amide Region of NOESY spectrum in H_2O

(t_{mix} = 75 ms., 600 MHz, 25°C)



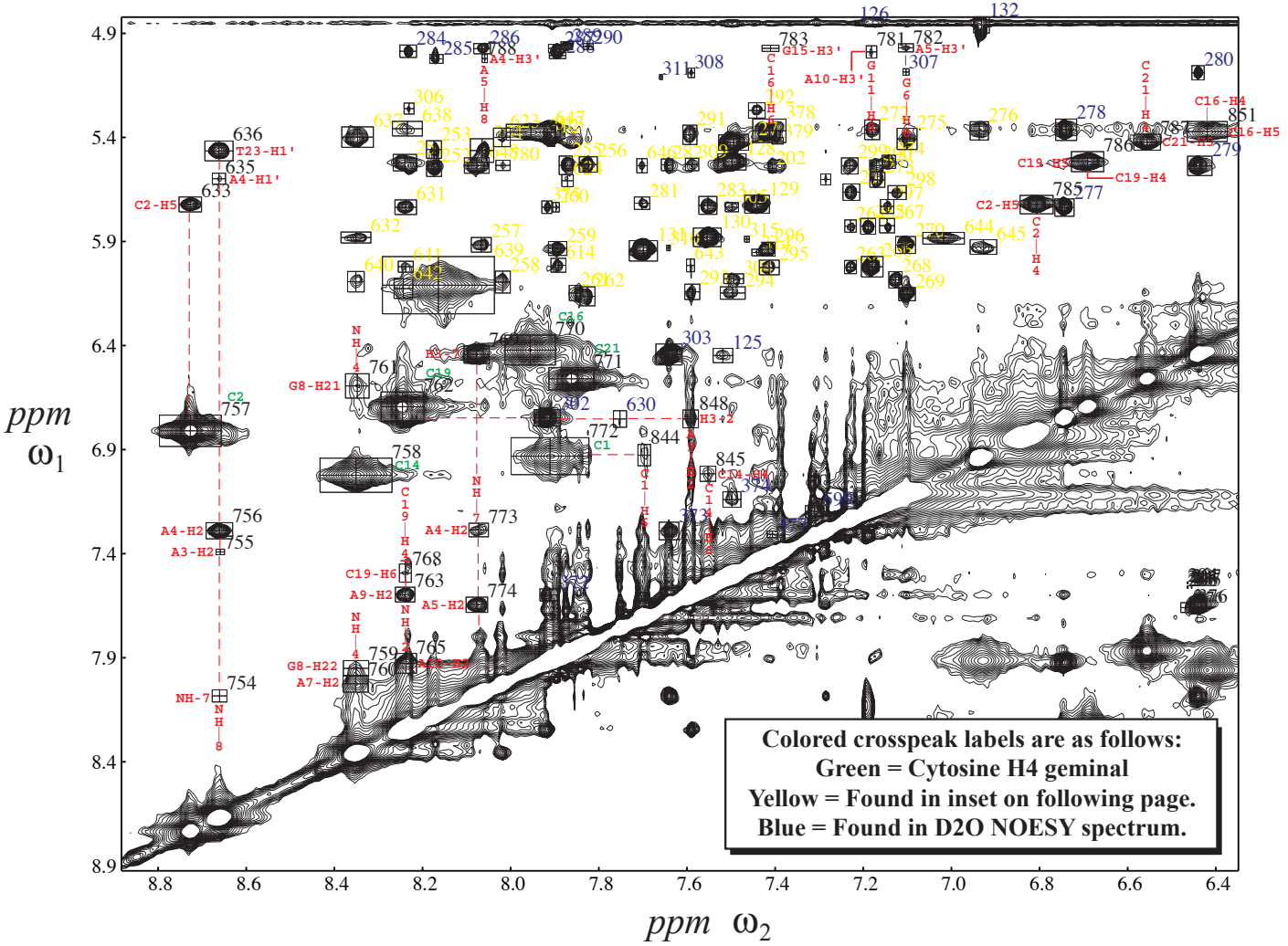
Aliphatic Amide Region of NOESY spectrum in H_2O

(t_{mix}= 75 ms., 600 MHz, 25°C)



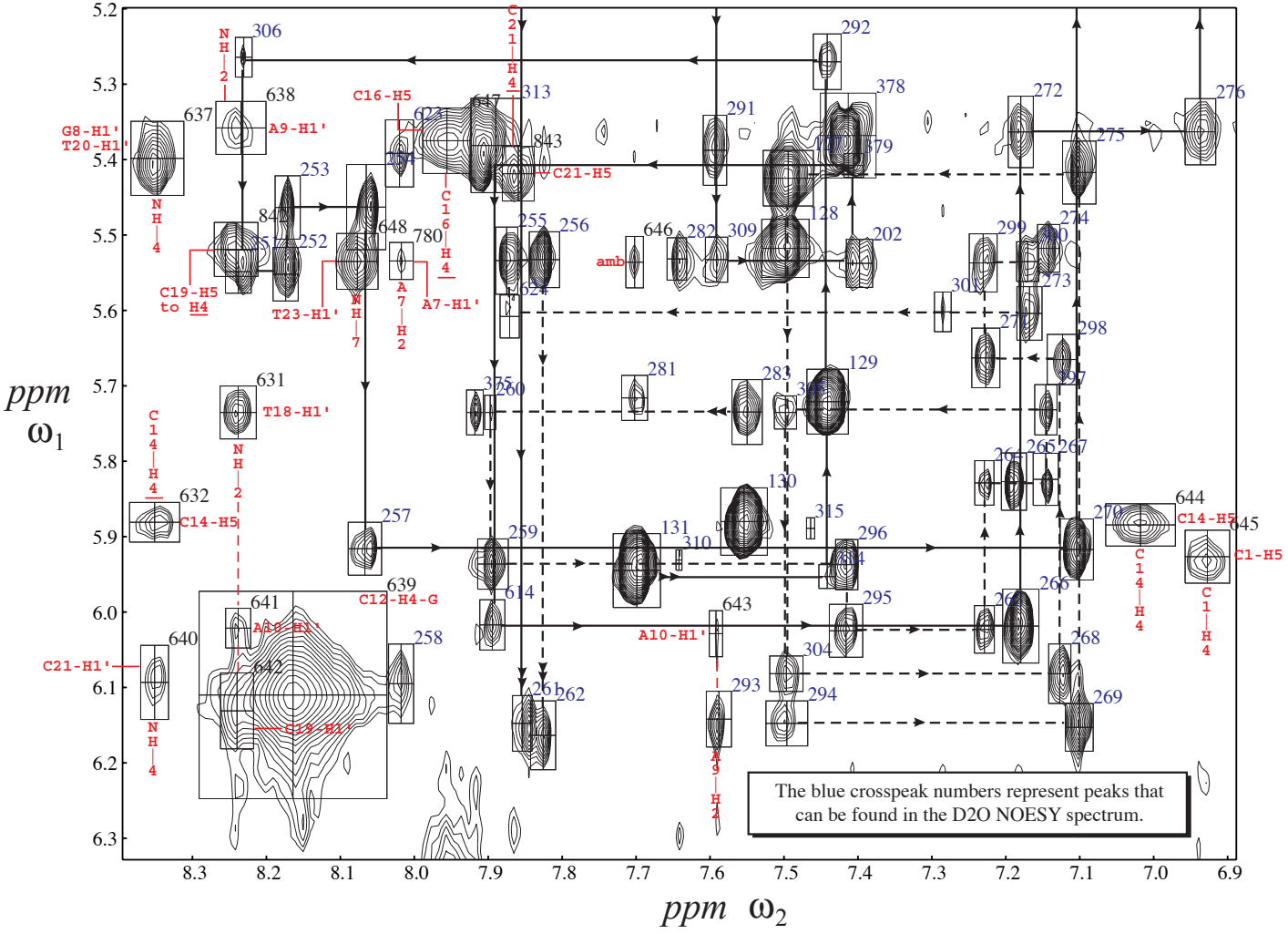
Aromatic Region of NOESY spectrum in H_2O

(t_{mix}= 75 ms., 600 MHz, 25°C)



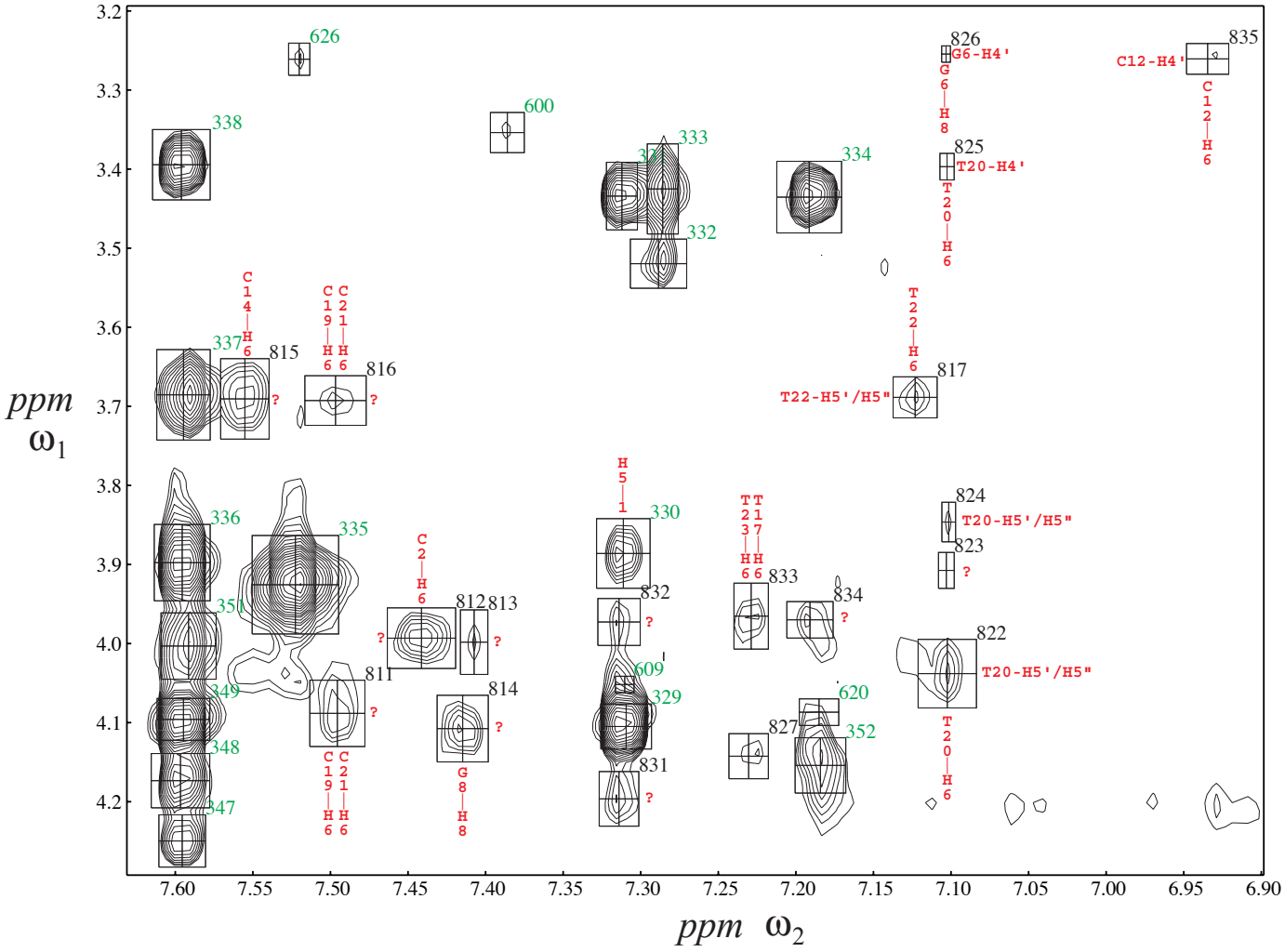
Aromatic to H1' Region of NOESY spectrum in H_2O

(t_{mix}= 75 ms., 600 MHz, 25°C)



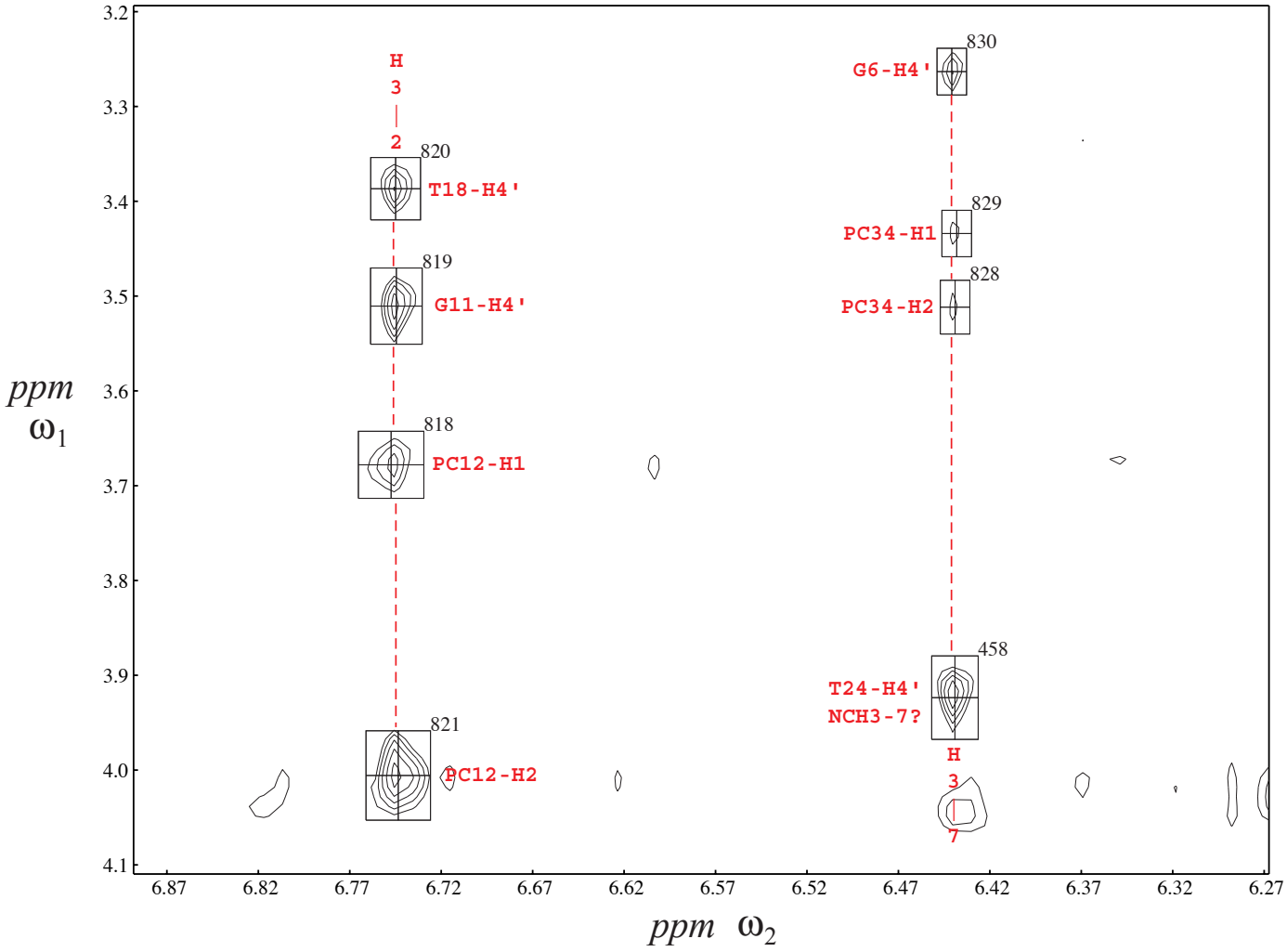
Aromatic to H4'/NCH3 Region of NOESY spectrum in H_2O

(t_{mix} = 75 ms., 600 MHz, 25°C)



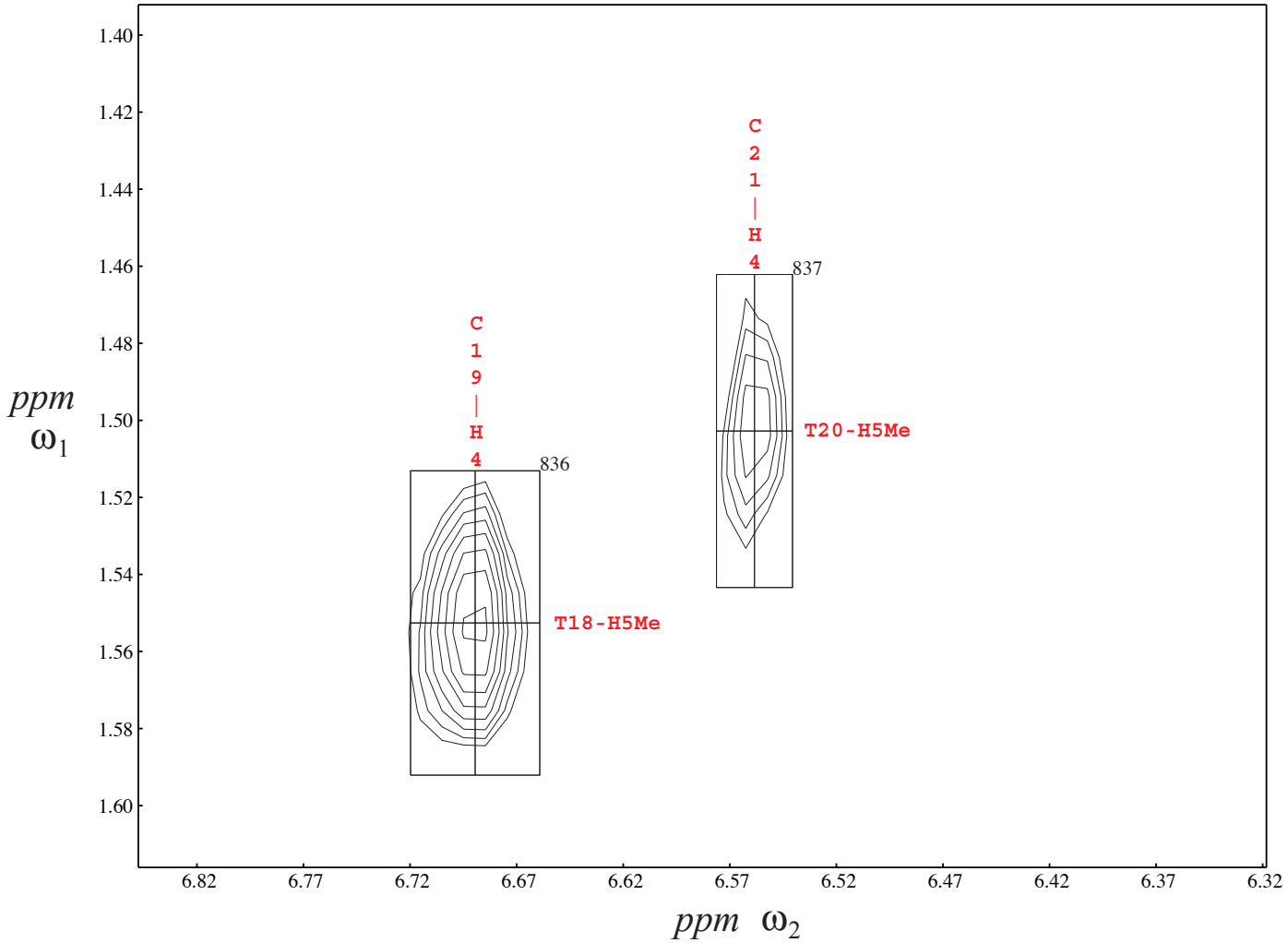
Pyrrole H3 Region of NOESY spectrum in H_2O

(t_{mix}= 75 ms., 600 MHz, 25°C)



CH4 to TH5Me Region of NOESY spectrum in H_2O

(t_{mix}= 75 ms., 600 MHz, 25°C)



Appendix B

Distance Constraints

A complete list of all constraints used in the structure calculations of the 1:1 polyamide-DNA complex is provided here. Methods for obtaining these restraints, based on the NOESY spectra in Appendix A, and are described in the Experimental.

1) Intraresidue DNA to DNA Constraints

residue number	residue name	atom name	residue number	residue name	atom name	upper bound
1	CYT	H2'1	1	CYT	H1'	4.25
1	CYT	H2'2	1	CYT	H1'	2.8
1	CYT	H4'	1	CYT	H1'	3.58
1	CYT	H4'	1	CYT	H2'1	4.38
1	CYT	H4'	1	CYT	H2'2	5.5
1	CYT	H6	1	CYT	H2'1	2.8
1	CYT	H6	1	CYT	H2'2	4.78
2	CYT	H2'1	2	CYT	H6	2.75
2	CYT	H2'2	2	CYT	H1'	2.87
2	CYT	H4'	2	CYT	H2'1	4.36
2	CYT	H4'	2	CYT	H2'2	5.5
2	CYT	H6	2	CYT	H1'	4.48
3	ADE	H1'	3	ADE	H3'	5.57
3	ADE	H1'	3	ADE	H4'	5.5
3	ADE	H1'	3	ADE	H8	4.29
3	ADE	H3'	3	ADE	H5'1	5.5
3	ADE	H3'	3	ADE	H5'2	5.5
3	ADE	H3'	3	ADE	H8	4.96
3	ADE	H4'	3	ADE	H3'	3.3
3	ADE	H8	3	ADE	H2'1	5.5
3	ADE	H8	3	ADE	H2'2	5.5
4	ADE	H3'	4	ADE	H1'	4.88
4	ADE	H3'	4	ADE	H8	5.04
4	ADE	H4'	4	ADE	H1'	4.18
4	ADE	H4'	4	ADE	H3'	3.22
4	ADE	H8	4	ADE	H1'	4.23
4	ADE	H8	4	ADE	H2'1	5.5
4	ADE	H8	4	ADE	H2'2	5.5
4	ADE	Q5'	4	ADE	H3'	4.27
5	ADE	H2	5	ADE	H1'	4.83
5	ADE	H3'	5	ADE	H5'1	5.5
5	ADE	H3'	5	ADE	H5'2	5.5
5	ADE	H4'	5	ADE	H1'	3.65
5	ADE	H8	5	ADE	H1'	4.35
5	ADE	H8	5	ADE	H2'1	5.5
5	ADE	H8	5	ADE	H2'2	5.5
6	GUA	H2'1	6	GUA	H1'	4.2
6	GUA	H2'2	6	GUA	H1'	3.06
6	GUA	H4'	6	GUA	H1'	4.18
6	GUA	H8	6	GUA	H1'	5.76
6	GUA	H8	6	GUA	H2'1	5.5
7	ADE	H1'	7	ADE	H2'1	3.53

7	ADE	H1'	7	ADE	H4'	5.5
7	ADE	H1'	7	ADE	H8	5.67
7	ADE	H2'1	7	ADE	H8	3.38
7	ADE	H2'2	7	ADE	H8	4.57
7	ADE	H4'	7	ADE	H2'1	5.85
7	ADE	H4'	7	ADE	H2'2	4.39
8	GUA	H3'	8	GUA	H2'2	4.1
8	GUA	H3'	8	GUA	Q5'	5.5
8	GUA	H8	8	GUA	H1'	5.83
8	GUA	H8	8	GUA	H2'1	5.5
8	GUA	H8	8	GUA	H2'2	5.5
9	ADE	H1'	9	ADE	H4'	4.17
9	ADE	H2'2	9	ADE	H3'	3.6
9	ADE	H3'	9	ADE	H1'	4.22
9	ADE	H3'	9	ADE	H4'	4.13
9	ADE	H8	9	ADE	H1'	5.5
9	ADE	H8	9	ADE	H2'1	5.5
9	ADE	Q5'	9	ADE	H4'	3.37
10	ADE	H1'	10	ADE	H2'1	5.5
10	ADE	H1'	10	ADE	H4'	3.62
10	ADE	H2	10	ADE	H1'	4.89
10	ADE	H2'1	10	ADE	H3'	3.05
10	ADE	H2'2	10	ADE	H1'	2.95
10	ADE	H3'	10	ADE	H1'	5
10	ADE	H3'	10	ADE	H4'	3.42
10	ADE	H3'	10	ADE	H5'1	5.5
10	ADE	H3'	10	ADE	H5'2	5.5
10	ADE	H4'	10	ADE	Q5'	3.38
10	ADE	H8	10	ADE	H2'1	2.93
11	GUA	H1'	11	GUA	H4'	4.19
11	GUA	H2'1	11	GUA	H1'	3.57
11	GUA	H2'1	11	GUA	H4'	5.82
11	GUA	H2'1	11	GUA	H8	3.18
11	GUA	H2'2	11	GUA	H3'	4.54
11	GUA	H2'2	11	GUA	H4'	4.39
11	GUA	H3'	11	GUA	H4'	4.31
11	GUA	H4'	11	GUA	Q5'	3.47
11	GUA	H8	11	GUA	H1'	4.68
11	GUA	H8	11	GUA	H2'2	5.5
12	CYT	H2'1	12	CYT	H3'	3.32
12	CYT	H2'1	12	CYT	H6	3.12
12	CYT	H2'2	12	CYT	H3'	3.67
12	CYT	H4'	12	CYT	H1'	4.53
12	CYT	H4'	12	CYT	H2'1	4.4
12	CYT	H4'	12	CYT	H3'	4.16
13	GUA	H1'	13	GUA	H2'2	3.12
13	GUA	H1'	13	GUA	H4'	5.5
13	GUA	H1'	13	GUA	H8	4.76
13	GUA	H2'2	13	GUA	H8	5.5
13	GUA	H4'	13	GUA	H2'1	5.5
13	GUA	H4'	13	GUA	H2'2	4.67
14	CYT	H2'1	14	CYT	H1'	4.38
14	CYT	H2'1	14	CYT	H6	2.86
14	CYT	H2'2	14	CYT	H1'	2.83
14	CYT	H3'	14	CYT	H2'1	3.25
14	CYT	H4'	14	CYT	H1'	3.5
14	CYT	H4'	14	CYT	H2'2	4.26

14	CYT	H6	14	CYT	H1'	3.64
15	GUA	H1'	15	GUA	H4'	3.64
15	GUA	H2'2	15	GUA	H1'	2.89
15	GUA	H3'	15	GUA	H4'	3.35
15	GUA	H3'	15	GUA	H5'1	5.5
15	GUA	H3'	15	GUA	H5'2	5.5
15	GUA	H3'	15	GUA	H8	4.46
15	GUA	H8	15	GUA	H1'	4.43
15	GUA	H8	15	GUA	H2'1	5.5
15	GUA	H8	15	GUA	H2'2	5.5
15	GUA	Q5'	15	GUA	H4'	3.35
16	CYT	H2'1	16	CYT	H1'	3.47
16	CYT	H3'	16	CYT	H1'	5.91
16	CYT	H4'	16	CYT	H1'	4.11
16	CYT	H6	16	CYT	H1'	4.47
16	CYT	H6	16	CYT	H2'1	5.5
17	THY	H1'	17	THY	H4'	4.17
17	THY	H2'1	17	THY	H1'	3.64
17	THY	H2'2	17	THY	H1'	2.96
17	THY	H2'2	17	THY	H6	3.58
17	THY	H3	17	THY	H1'	6
17	THY	H3'	17	THY	H4'	4.46
17	THY	H6	17	THY	H1'	4.56
17	THY	H6	17	THY	H2'1	5.5
17	THY	Q5'	17	THY	H4'	3.45
18	THY	H2'1	18	THY	H1'	4.19
18	THY	H4'	18	THY	H3'	5.09
18	THY	H6	18	THY	H1'	4.57
18	THY	H6	18	THY	H2'1	5.5
18	THY	H6	18	THY	H2'2	5.5
19	CYT	H1'	19	CYT	H4'	5.5
19	CYT	H2'1	19	CYT	H1'	3.51
19	CYT	H3'	19	CYT	Q5'	4.68
19	CYT	H3'	19	CYT	Q5'	5.5
19	CYT	H6	19	CYT	H1'	4.31
19	CYT	H6	19	CYT	H2'1	5.5
19	CYT	H6	19	CYT	H2'2	5.5
20	THY	H1'	20	THY	H4'	3.67
20	THY	H2'1	20	THY	H1'	3.27
20	THY	H3	20	THY	H1'	6
20	THY	H3'	20	THY	H4'	5.83
20	THY	H6	20	THY	H1'	5.5
20	THY	H6	20	THY	H2'1	5.5
21	CYT	H2'1	21	CYT	H1'	4.1
21	CYT	H2'2	21	CYT	H1'	3.03
21	CYT	H4'	21	CYT	H1'	3.54
21	CYT	H6	21	CYT	H1'	4.33
21	CYT	H6	21	CYT	H2'1	5.5
21	CYT	H6	21	CYT	H2'2	5.5
22	THY	H1'	22	THY	H4'	4.21
22	THY	H2'1	22	THY	H1'	3.48
22	THY	H2'2	22	THY	H1'	3.04
22	THY	H3	22	THY	H1'	5.87
22	THY	H3'	22	THY	H4'	5.05
22	THY	H6	22	THY	H1'	4.61
22	THY	H6	22	THY	H2'1	5.5
22	THY	H6	22	THY	H2'2	5.5

22	THY	Q5'	22	THY	H4'	3.45
23	THY	H1'	23	THY	H4'	5.5
23	THY	H2'1	23	THY	H1'	3.54
23	THY	H2'2	23	THY	H1'	2.94
23	THY	H2'2	23	THY	H3'	4.16
23	THY	H3	23	THY	H1'	5.56
23	THY	H3'	23	THY	H4'	4.38
23	THY	H3'	23	THY	H5'1	5.5
23	THY	H3'	23	THY	H5'2	5.5
23	THY	H3'	23	THY	Q5'	4.87
23	THY	H4'	23	THY	Q5'	3.4
23	THY	H6	23	THY	H1'	4.5
23	THY	H6	23	THY	H2'1	5.5
24	THY	H2'1	24	THY	H1'	4.16
24	THY	H2'2	24	THY	H1'	2.87
24	THY	H4'	24	THY	H1'	3.46
24	THY	H4'	24	THY	H2'1	5.5
24	THY	H4'	24	THY	H2'2	5.5
24	THY	H4'	24	THY	H3'	4.11
24	THY	H6	24	THY	H1'	4.24
24	THY	H6	24	THY	H2'1	5.5
24	THY	H6	24	THY	H2'2	5.5
24	THY	Q5'	24	THY	H3'	4.95
25	GUA	H1'	25	GUA	H3'	4.36
25	GUA	H1'	25	GUA	H4'	5.5
25	GUA	H1'	25	GUA	H8	4.12
25	GUA	H3'	25	GUA	H5'1	5.5
25	GUA	H3'	25	GUA	H5'2	5.5
25	GUA	H3'	25	GUA	H8	4.76
25	GUA	H4'	25	GUA	H3'	3.33
25	GUA	H8	25	GUA	H2'1	5.5
25	GUA	H8	25	GUA	H2'2	5.5
26	GUA	H1'	26	GUA	H2'2	3.14
26	GUA	H1'	26	GUA	H3'	5.6
26	GUA	H1'	26	GUA	H4'	5.5
26	GUA	H1'	26	GUA	H8	4.28
26	GUA	H2'1	26	GUA	H8	3.28
26	GUA	H4'	26	GUA	H2'1	5.5
26	GUA	H4'	26	GUA	H2'2	5.5

2) Intramolecular Ligand to Ligand Constraints

residue number	residue name	atom name	residue number	residue name	atom name	upper bound
27	NIM	H	27	NIM	H4	6
27	NIM	H	28	PYL	H3	3.55
28	PYL	H	29	BAL	H11	3.95
28	PYL	H	29	BAL	H12	3.95
28	PYL	H	29	BAL	2Q	6
29	BAL	H	29	BAL	H11	5.3
29	BAL	H	29	BAL	H12	5.3
29	BAL	H	29	BAL	2Q	6
29	BAL	H11	29	BAL	H21	5.5
29	BAL	H11	29	BAL	H22	5.5
29	BAL	H12	29	BAL	H21	5.5
29	BAL	H12	29	BAL	H22	5.5
30	IMI	H	31	BAL	H11	4.59
30	IMI	H	31	BAL	H12	4.59
30	IMI	H	31	BAL	H21	4.7
30	IMI	H	31	BAL	H22	4.7
31	BAL	1Q	31	BAL	2Q	5.5
31	BAL	1Q	31	BAL	2Q	5.5
31	BAL	2Q	31	BAL	1Q	4.29
31	BAL	H	31	BAL	1Q	4.72
31	BAL	H	31	BAL	H21	6
31	BAL	H	31	BAL	H22	6
32	IMI	H	33	PYL	H3	3.53
32	IMI	H	33	PYL	H5	4.53
33	PYL	H	34	BAL	H11	3.55
33	PYL	H	34	BAL	H12	3.55
33	PYL	H	34	BAL	2Q	4.79
33	PYL	H	33	PYL	H3	3.12
33	PYL	H5	33	PYL	NM	5.5
34	BAL	2Q	34	BAL	1Q	4.17
34	BAL	H	34	BAL	H21	4.31
34	BAL	H	34	BAL	H22	4.31
34	BAL	H	35	DMP	2Q	6
34	BAL	H	35	DMP	3Q	6
34	BAL	H	35	DMP	H11	6
34	BAL	H	35	DMP	H12	6
34	BAL	H	33	PYL	H	4.96
35	DMP	3Q	35	DMP	DM	5.5

3) Interresidue DNA to DNA Constraints

residue number	residue name	atom name	residue number	residue name	atom name	upper bound
2	CYT	H5	1	CYT	H6	5.5
2	CYT	H6	1	CYT	H2'2	5.5
2	CYT	H6	1	CYT	H1'	4.13
2	CYT	H6	1	CYT	H2'1	4.87
3	ADE	H8	2	CYT	H2'1	5.5
3	ADE	H8	2	CYT	H2'2	5.5
3	ADE	H8	2	CYT	H1'	5.5
3	ADE	H2	4	ADE	H2	5.5
3	ADE	H2	25	GUA	H1'	6
4	ADE	H8	3	ADE	H1'	4.25
4	ADE	H8	3	ADE	H2'1	5.5
4	ADE	H8	3	ADE	H2'2	5.5
4	ADE	H2	5	ADE	H2	3.62
4	ADE	H2	24	THY	H1'	5.5
5	ADE	H8	4	ADE	H1'	4.13
5	ADE	H8	4	ADE	H2'1	5.5
5	ADE	H8	4	ADE	H2'2	5.5
5	ADE	H2	23	THY	H1'	4.24
5	ADE	H2	23	THY	H3	4.88
6	GUA	H1	5	ADE	H2	4.66
6	GUA	H8	5	ADE	H2'1	4.05
6	GUA	H8	5	ADE	H2'2	6
6	GUA	H8	5	ADE	H1'	5.5
6	GUA	H1	7	ADE	H2	4.68
6	GUA	H1	21	CYT	H41	5.5
6	GUA	H1	21	CYT	H42	6
7	ADE	H8	6	GUA	H2'2	6
7	ADE	H8	6	GUA	H2'1	3.44
7	ADE	H8	6	GUA	H1'	3.64
7	ADE	Q5'	6	GUA	H1'	3.15
7	ADE	H1'	8	GUA	Q5'	5.5
7	ADE	H2	8	GUA	H1'	4.88
7	ADE	H2	21	CYT	H1'	4.88
8	GUA	H8	7	ADE	H2'1	4.44
8	GUA	H8	7	ADE	H2'2	5.5
8	GUA	H1	19	CYT	H42	3.84
8	GUA	H1	19	CYT	H5	6
8	GUA	H1	19	CYT	H41	3.18
8	GUA	H1	9	ADE	H2	5.5
9	ADE	H8	8	GUA	H2'1	5.5
9	ADE	H8	8	GUA	H2'2	4.9
9	ADE	H8	8	GUA	H1'	5.53
9	ADE	H2	10	ADE	H2	3.5
9	ADE	H2	10	ADE	H1'	5.54
9	ADE	H2	19	CYT	H1'	4.24
10	ADE	H8	9	ADE	H1'	5.5
10	ADE	H8	9	ADE	H2'1	5.5
10	ADE	Q5'	9	ADE	H1'	4.92
10	ADE	H1'	11	GUA	Q5'	3.6
10	ADE	H2	11	GUA	H1'	4.53
11	GUA	H8	10	ADE	H2'2	5.5

11	GUA	H8	10	ADE	H2'1	5.5
11	GUA	H8	10	ADE	H1'	3.53
11	GUA	Q5'	10	ADE	H1'	3.09
11	GUA	H1	12	CYT	H42	3.43
11	GUA	H1	16	CYT	H41	4.67
11	GUA	H1	16	CYT	H42	3.24
12	CYT	H6	11	GUA	H2'2	3.32
12	CYT	H6	11	GUA	H2'1	5.66
12	CYT	H6	11	GUA	H1'	6
12	CYT	H41	15	GUA	H1	5.53
12	CYT	H5	15	GUA	H1	5.4
13	GUA	H8	12	CYT	H2'1	3.25
13	GUA	H8	12	CYT	H2'2	3.49
15	GUA	H1	11	GUA	H1	4.68
15	GUA	H1	12	CYT	H42	5.78
15	GUA	H1	14	CYT	H41	6
15	GUA	H8	14	CYT	H2'2	4.12
15	GUA	H8	14	CYT	H2'1	4.15
15	GUA	H8	14	CYT	H1'	3.26
16	CYT	H5	15	GUA	H8	5.5
16	CYT	H5	15	GUA	H2'2	3.59
16	CYT	H6	15	GUA	H1'	4.25
16	CYT	H1'	17	THY	Q5'	5.5
16	CYT	H5	17	THY	M7	4.8
16	CYT	H6	17	THY	M7	4.9
17	THY	H3	10	ADE	H2	3.58
17	THY	H3	11	GUA	H1	4.66
17	THY	H6	16	CYT	H2'2	3.25
17	THY	H6	16	CYT	H2'1	3.42
17	THY	H6	16	CYT	H1'	4.44
17	THY	M7	16	CYT	H2'1	3.6
17	THY	M7	16	CYT	H1'	4.95
17	THY	H6	18	THY	M7	4.24
18	THY	H3	8	GUA	H1	5.17
18	THY	H3	9	ADE	H2	4.61
18	THY	H1'	10	ADE	H2	6.22
18	THY	H3	10	ADE	H2	4.4
18	THY	H6	17	THY	H1'	4.58
18	THY	H6	17	THY	H2'1	5.5
18	THY	H6	17	THY	H2'2	5.5
18	THY	M7	17	THY	H2'1	5.5
18	THY	M7	17	THY	H1'	5.74
18	THY	Q5'	17	THY	H1'	5.88
18	THY	H3	19	CYT	H42	5.28
19	CYT	H5	18	THY	H2'2	3.62
19	CYT	H41	18	THY	M7	4.45
19	CYT	H42	18	THY	M7	4.4
19	CYT	H5	18	THY	H6	4.57
19	CYT	H5	18	THY	M7	4.95
19	CYT	H5	18	THY	H2'1	5.5
19	CYT	H6	18	THY	H1'	4.5
19	CYT	H6	18	THY	H2'1	5.5
19	CYT	H6	18	THY	H2'2	5.5
19	CYT	Q5'	18	THY	H1'	4.96
19	CYT	H2'2	20	THY	M7	3.68
19	CYT	H6	20	THY	M7	5.5
20	THY	H3	6	GUA	H1	5.16

20	THY	H3	7	ADE	H2	3.5
20	THY	H3	8	GUA	H1	4.79
20	THY	H6	19	CYT	H1'	4.2
20	THY	H6	19	CYT	H2'1	4.25
20	THY	H6	19	CYT	H2'2	5.5
20	THY	M7	19	CYT	H2'1	3.6
20	THY	M7	19	CYT	H5	4.7
20	THY	M7	19	CYT	H1'	4.9
20	THY	Q5'	19	CYT	H1'	4.89
21	CYT	H42	20	THY	H3	5.47
21	CYT	H5	20	THY	H6	5.5
21	CYT	H41	20	THY	M7	4.5
21	CYT	H5	20	THY	M7	5.5
21	CYT	H6	20	THY	H2'1	5.5
21	CYT	H6	20	THY	H2'2	5.5
21	CYT	H5	22	THY	M7	5.5
21	CYT	H6	22	THY	M7	5.5
22	THY	H3	5	ADE	H2	4.25
22	THY	H3	6	GUA	H1	4.94
22	THY	H6	21	CYT	H2'1	3.62
22	THY	H6	21	CYT	H1'	4.4
22	THY	H6	21	CYT	H2'2	5.5
22	THY	M7	21	CYT	H2'1	4.26
22	THY	M7	21	CYT	H1'	5.62
22	THY	H3	23	THY	H3	4.16
23	THY	H3	3	ADE	H2	6
23	THY	H3	4	ADE	H2	3.44
23	THY	H6	22	THY	H2'2	3.09
23	THY	H6	22	THY	H2'1	3.44
23	THY	H6	22	THY	H1'	4.37
23	THY	M7	22	THY	H2'1	3.64
23	THY	M7	22	THY	H6	4.38
23	THY	Q5'	22	THY	H1'	5.05
24	THY	M7	23	THY	H2'1	5.5
24	THY	M7	23	THY	H2'2	5.5
24	THY	H3	3	ADE	H2	3.5
24	THY	H3	4	ADE	H2	4.73
24	THY	H6	23	THY	H2'2	3.18
24	THY	H6	23	THY	H1'	4.8
24	THY	H6	23	THY	H2'1	5.5
24	THY	M7	23	THY	H6	4.25
24	THY	M7	23	THY	H1'	5.56
25	GUA	H1	24	THY	H3	4.79
25	GUA	H1	1	CYT	H41	6
25	GUA	H1	2	CYT	H42	3.44
25	GUA	H1	2	CYT	H41	6
25	GUA	H1	3	ADE	H2	4.95
25	GUA	H8	24	THY	H2'1	4.12
25	GUA	H8	24	THY	H2'2	4.23
25	GUA	H8	24	THY	H1'	4.93
26	GUA	H8	25	GUA	H2'1	5.5
26	GUA	H8	25	GUA	H2'2	5.5
26	GUA	H8	25	GUA	H1'	4.18

4) Intermolecular Ligand to DNA Constraints

residue number	residue name	atom name	residue number	residue name	atom name	upper bound
27	NIM	H	10	ADE	H2	4.77
27	NIM	H	11	GUA	H22	3.78
27	NIM	H	11	GUA	H1'	4.54
27	NIM	H	12	CYT	H4'	5.39
27	NIM	H	12	CYT	Q5'	6
27	NIM	H	17	THY	H1'	4.74
27	NIM	H	18	THY	H4'	4.85
27	NIM	H	18	THY	H1'	5.64
27	NIM	H	18	THY	Q5'	6
27	NIM	H4	16	CYT	H1'	5.5
27	NIM	H4	17	THY	H4'	4.15
27	NIM	H4	17	THY	H1'	4.19
27	NIM	H5	13	GUA	Q5'	5.53
27	NIM	H5	13	GUA	Q5'	5.5
27	NIM	H5	17	THY	H4'	4.15
27	NIM	NM	12	CYT	H4'	5.5
27	NIM	NM	18	THY	H4'	5.5
28	PYL	H	9	ADE	H2	4.13
28	PYL	H	10	ADE	H1'	4.88
28	PYL	H	10	ADE	H2	6
28	PYL	H	11	GUA	H4'	4.7
28	PYL	H	18	THY	H1'	4.36
28	PYL	H	19	CYT	H1'	6
28	PYL	H3	9	ADE	H2	4.68
28	PYL	H3	10	ADE	H2	2.07
28	PYL	H3	11	GUA	H1'	3.26
28	PYL	H3	11	GUA	H4'	4.69
28	PYL	H3	17	THY	H3	5.07
28	PYL	H3	18	THY	H1'	3.08
28	PYL	H5	12	CYT	H4'	3.24
28	PYL	H5	12	CYT	Q5'	3.25
28	PYL	H5	12	CYT	H4'	3.24
28	PYL	H5	12	CYT	Q5'	3.25
28	PYL	H5	18	THY	H4'	3.05
29	BAL	1Q	9	ADE	H2	4.44
29	BAL	1Q	9	ADE	H2	5.5
29	BAL	1Q	10	ADE	H1'	4.38
29	BAL	1Q	10	ADE	H2	4.83
29	BAL	1Q	19	CYT	H1'	4.83
29	BAL	2Q	9	ADE	H2	5.5
29	BAL	2Q	10	ADE	H1'	5.5
29	BAL	2Q	19	CYT	H1'	5.5
29	BAL	2Q	19	CYT	H4'	5.5
29	BAL	2Q	19	CYT	H1'	5.5
29	BAL	2Q	19	CYT	H4'	5.5
29	BAL	H	8	GUA	H22	4.48
29	BAL	H	9	ADE	H1'	3.93
29	BAL	H	9	ADE	H2	4.94
29	BAL	H	10	ADE	Q5'	6
29	BAL	H	19	CYT	H1'	4.67
29	BAL	H	20	THY	H4'	5.22
30	IMI	H	7	ADE	H2	6

30	IMI	H	8	GUA	H1'	6
30	IMI	H	8	GUA	H22	6
30	IMI	H	20	THY	H1'	6
30	IMI	H	21	CYT	H1'	4.34
30	IMI	H5	10	ADE	Q5'	4.34
30	IMI	H5	10	ADE	H4'	4.74
30	IMI	H5	20	THY	H4'	3.66
31	BAL	1Q	7	ADE	H2	4.54
31	BAL	1Q	8	GUA	H1'	4.65
31	BAL	1Q	21	CYT	H1'	3.6
31	BAL	2Q	7	ADE	H2	3.6
31	BAL	2Q	7	ADE	H1'	5.5
31	BAL	2Q	8	GUA	H4'	4.1
31	BAL	2Q	21	CYT	H1'	4.48
31	BAL	H	6	GUA	H22	4.49
31	BAL	H	7	ADE	H1'	3.41
31	BAL	H	7	ADE	H2	3.56
31	BAL	H	8	GUA	Q5'	6
32	IMI	H	5	ADE	H2	6
32	IMI	H	6	GUA	H1'	4.36
32	IMI	H	6	GUA	H22	6
32	IMI	H	7	ADE	H4'	4.18
32	IMI	H	22	THY	H1'	4.5
32	IMI	H	23	THY	H1'	4.44
32	IMI	H	23	THY	Q5'	6
32	IMI	H5	8	GUA	Q5'	4.38
32	IMI	H5	22	THY	H4'	4.19
33	PYL	H	4	ADE	H2	4.38
33	PYL	H	5	ADE	H2	3.99
33	PYL	H	6	GUA	H4'	4.96
33	PYL	H	23	THY	H1'	3.97
33	PYL	H	24	THY	H4'	6
33	PYL	H3	5	ADE	H2	2.08
33	PYL	H3	6	GUA	H1'	3.14
33	PYL	H3	6	GUA	H4'	5.21
33	PYL	H3	22	THY	H3	4.73
33	PYL	H3	23	THY	H3	5.53
33	PYL	H5	6	GUA	H4'	3.44
33	PYL	H5	7	ADE	H4'	5.5
33	PYL	H5	23	THY	H4'	5.5
34	BAL	1Q	4	ADE	H2	4.47
34	BAL	1Q	5	ADE	H1'	4.63
34	BAL	1Q	24	THY	H1'	4.56
34	BAL	2Q	4	ADE	H2	4.24
34	BAL	2Q	5	ADE	H1'	4.14
34	BAL	2Q	5	ADE	H2	4.82
34	BAL	2Q	24	THY	H1'	4.65
34	BAL	H	3	ADE	H2	5.79
34	BAL	H	4	ADE	H2	4.77
34	BAL	H	4	ADE	H1'	5.12
35	DMP	1Q	4	ADE	H1'	5.5
35	DMP	2Q	3	ADE	H2	5.5
35	DMP	2Q	4	ADE	H1'	5.5
35	DMP	2Q	4	ADE	H2	5.5
35	DMP	3Q	3	ADE	H2	5.5
35	DMP	3Q	4	ADE	H1'	5.5
35	DMP	3Q	4	ADE	H2	5.5

```
35      DMP      DM       3       ADE       H2      5.5
35      DMP      DM       3       ADE      H4'      5.5
35      DMP      DM       4       ADE       H2      5.5
35      DMP      DM      24       THY      H1'      5.5
35      DMP      DM      25       GUA      H4'      5.5
```

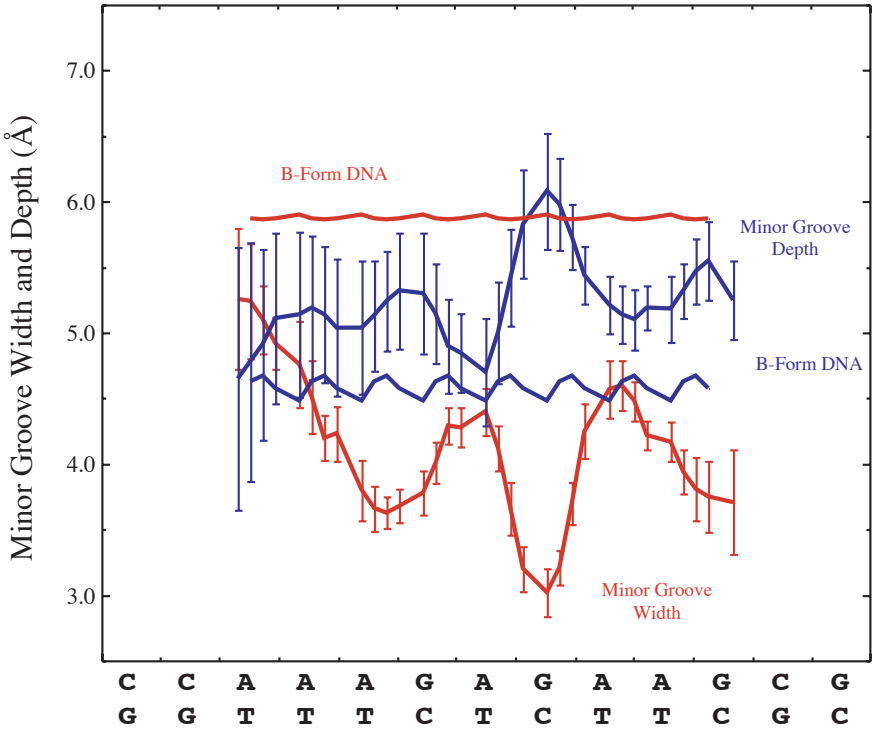
5) Watson-Crick Hydrogen Bonding Constraints

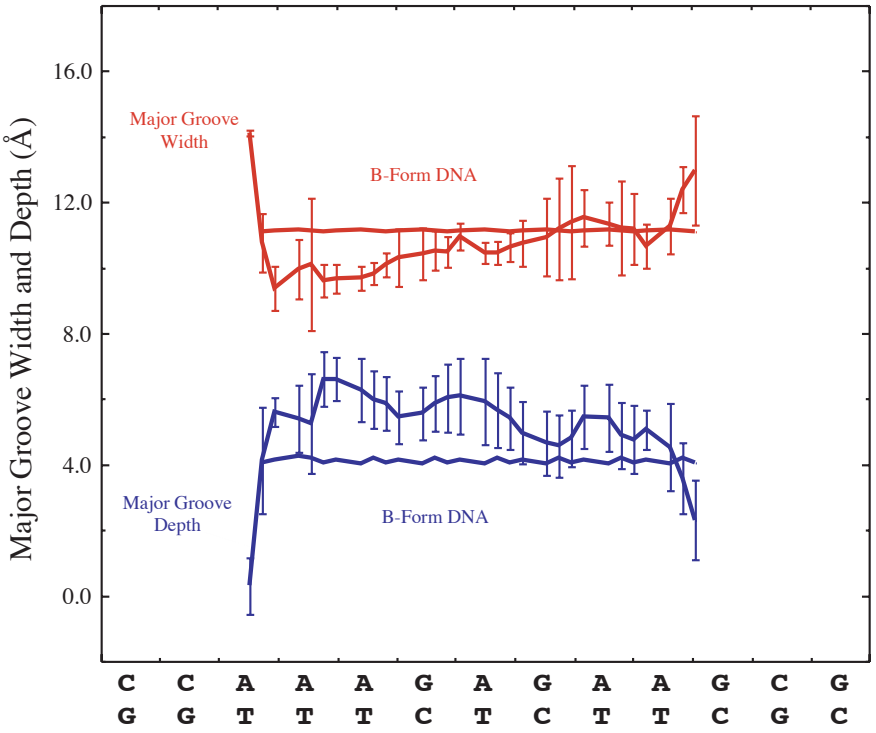
residue number	residue name	atom name	residue number	residue name	atom name	upper bound	lower bound
1	CYT	H42	26	GUA	O6	1.76	2.16
1	CYT	N3	26	GUA	H1	1.8	2.2
1	CYT	N3	26	GUA	N1	2.85	3.05
1	CYT	N4	26	GUA	O6	2.81	3.01
2	CYT	H42	25	GUA	O6	1.76	2.16
2	CYT	N3	25	GUA	H1	1.8	2.2
2	CYT	N3	25	GUA	N1	2.85	3.05
2	CYT	N4	25	GUA	O6	2.81	3.01
3	ADE	N1	24	THY	H3	1.67	2.07
3	ADE	N1	24	THY	N3	2.72	2.92
4	ADE	N1	23	THY	H3	1.67	2.07
4	ADE	N1	23	THY	N3	2.72	2.92
5	ADE	N1	22	THY	H3	1.67	2.07
5	ADE	N1	22	THY	N3	2.72	2.92
6	GUA	H1	21	CYT	N3	1.8	2.2
6	GUA	N1	21	CYT	N3	2.85	3.05
6	GUA	O6	21	CYT	H42	1.76	2.16
6	GUA	O6	21	CYT	N4	2.81	3.01
7	ADE	N1	20	THY	H3	1.67	2.07
7	ADE	N1	20	THY	N3	2.72	2.92
8	GUA	H1	19	CYT	N3	1.8	2.2
8	GUA	N1	19	CYT	N3	2.85	3.05
8	GUA	O6	19	CYT	H42	1.76	2.16
8	GUA	O6	19	CYT	N4	2.81	3.01
9	ADE	N1	18	THY	H3	1.67	2.07
9	ADE	N1	18	THY	N3	2.72	2.92
10	ADE	N1	17	THY	H3	1.67	2.07
10	ADE	N1	17	THY	N3	2.72	2.92
11	GUA	H1	16	CYT	N3	1.8	2.2
11	GUA	N1	16	CYT	N3	2.85	3.05
11	GUA	O6	16	CYT	H42	1.76	2.16
11	GUA	O6	16	CYT	N4	2.81	3.01
12	CYT	H42	15	GUA	O6	1.76	2.16
12	CYT	N3	15	GUA	H1	1.8	2.2
12	CYT	N3	15	GUA	N1	2.85	3.05
12	CYT	N4	15	GUA	O6	2.81	3.01
13	GUA	H1	14	CYT	N3	1.8	2.2
13	GUA	N1	14	CYT	N3	2.85	3.05
13	GUA	O6	14	CYT	H42	1.76	2.16
13	GUA	O6	14	CYT	N4	2.81	3.01

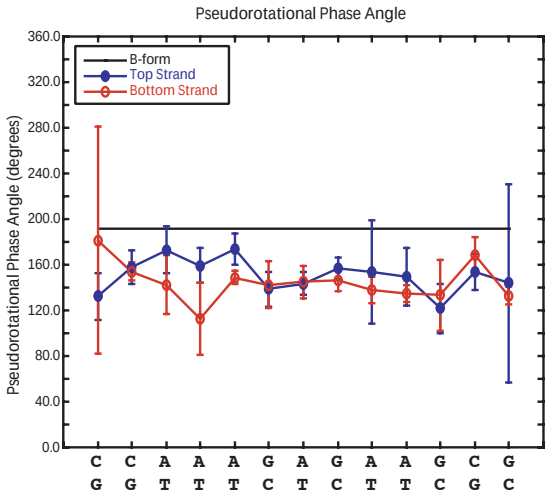
Appendix C

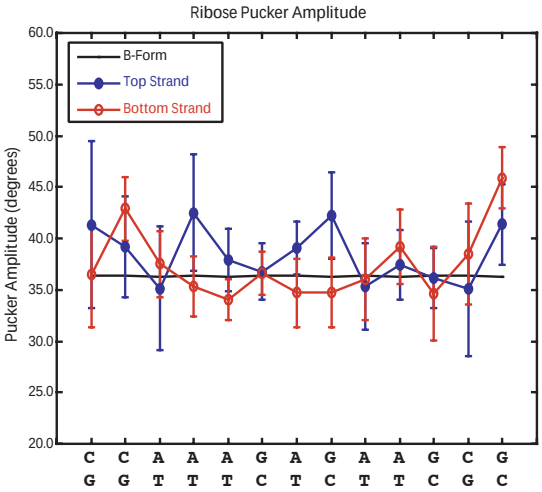
DNA Helical Parameters

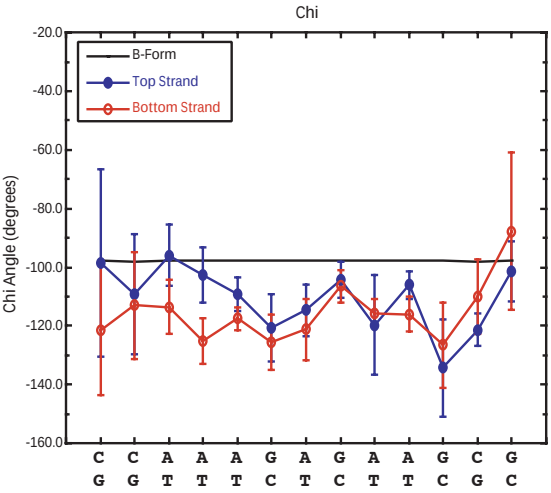
DNA helical parameters for the 1:1 polyamide-DNA complex determined by NMR methods are provided here. The parameters are plotted along the ordinate for each DNA base step along the abscissa. Average values over the final ensemble of 12 structures are connected by solid lines, and the y-axis error bars indicates one standard deviation from the average. Horizontal lines without error bars indicate the average values for standard B-form DNA.

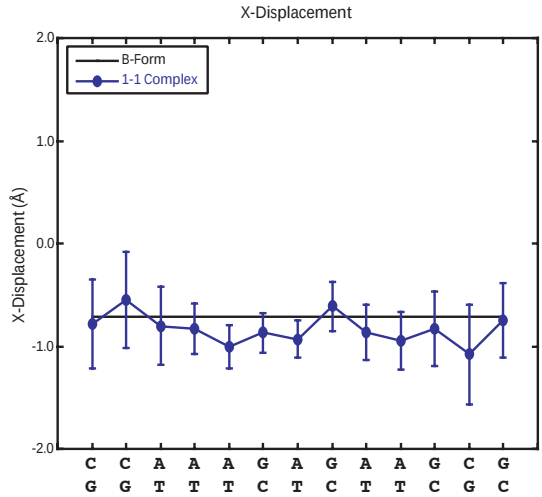

Minor Groove Width and Depth (Å)
7.0
6.0
5.0
4.0
3.0
B-Form DNA
Minor Groove Depth
B-Form DNA
Minor Groove Width
C C A A A G A G A A G C G
G G T T T C T C T T C G C

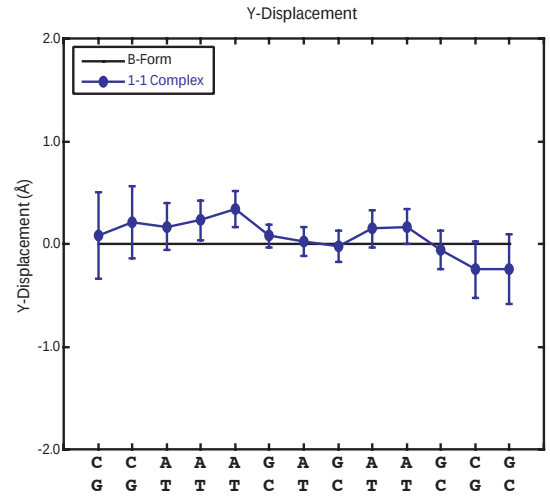

Major Groove Width and Depth (Å)
16.0
12.0
8.0
4.0
0.0
Major Groove Width
B-Form DNA
Major Groove Depth
B-Form DNA
C C A A A G A G A A G C G
G G T T T C T C T T C G C

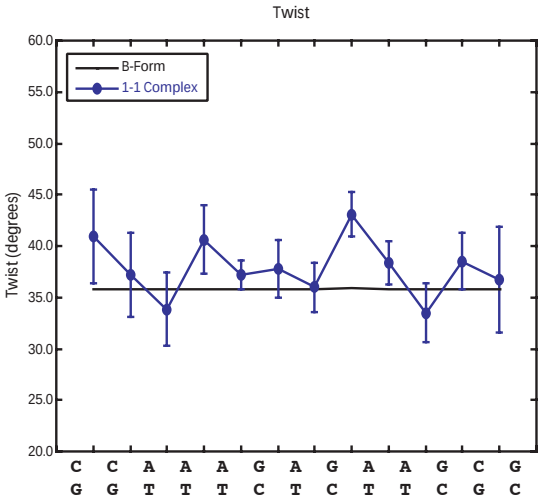

Pseudorotational Phase Angle
Pseudorotational Phase Angle (degrees)
B-form
Top Strand
Bottom Strand
C C A A A G A G A A G C G
G G T T T C T C T T C G C

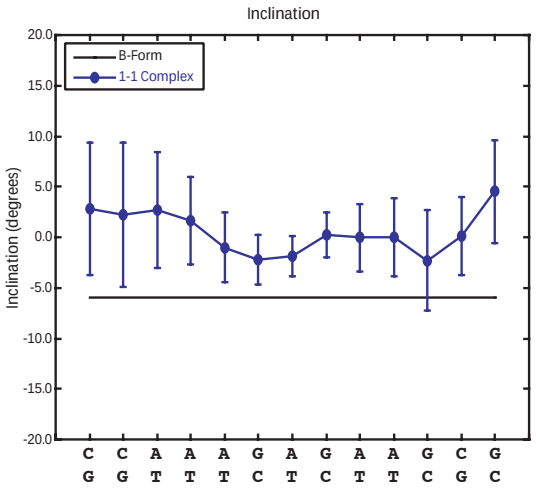

Ribose Pucker Amplitude
Pucker Amplitude (degrees)
B-Form
Top Strand
Bottom Strand
C C A A A G A G A A G C G
G G T T T C T C T T C G C

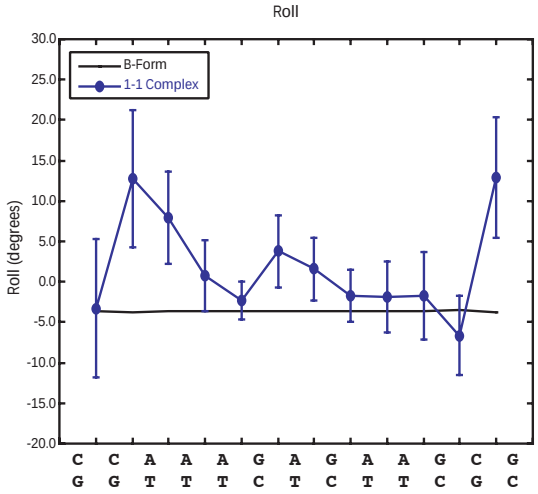

Chi
Chi Angle (degrees)
B-Form
Top Strand
Bottom Strand
C C A A A G A G A A G C G
G G T T T C T C T T C G C

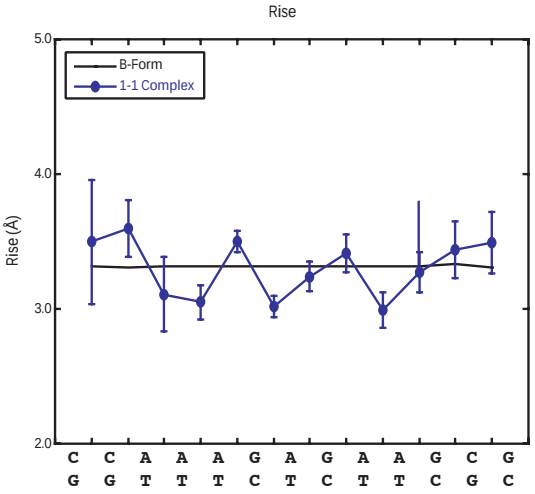

X-Displacement
X-Displacement (Å)
B-Form
1-1 Complex
C C A A A G A G A A G C G
G G T T T C T C T T C G C

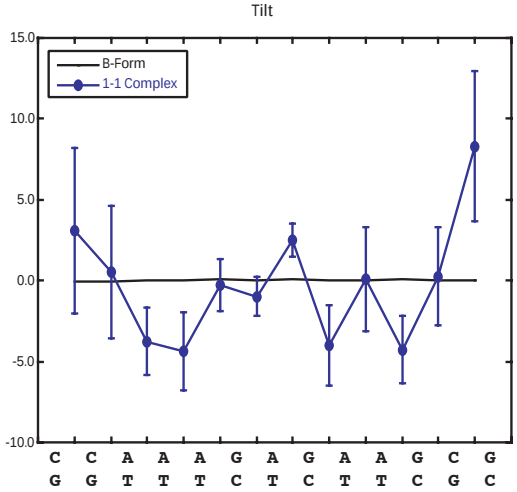

Y-Displacement
Y-Displacement (Å)
B-Form
1-1 Complex
C C A A A G A G A A G C G
G G T T T C T C T T C G C

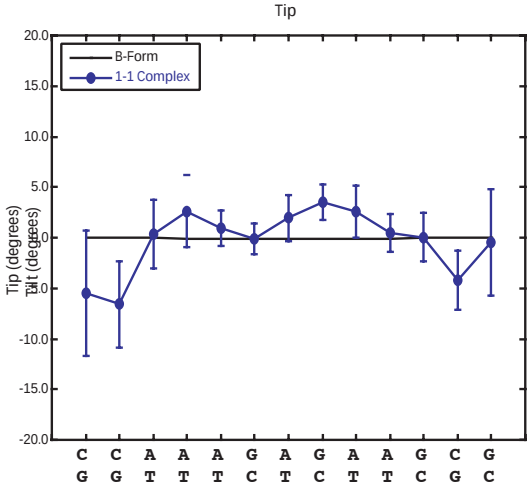

Twist
Twist (degrees)
B-Form
1-1 Complex
C C A A A G A G A A G C G
G G T T T C T C T T C G C

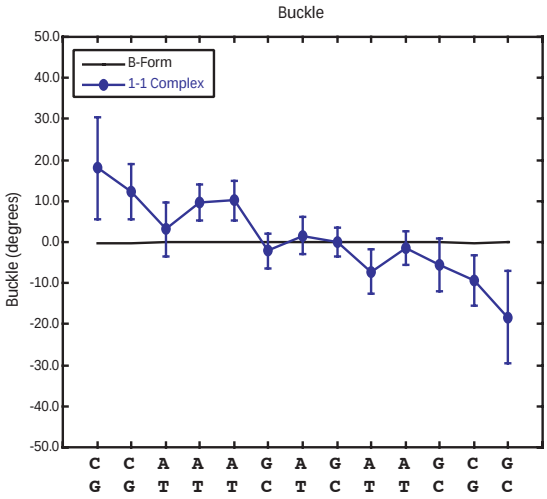

Inclination
B-Form
1-1 Complex
Inclination (degrees)
20.0
15.0
10.0
5.0
0.0
-5.0
-10.0
-15.0
-20.0
C C A A A G A G A A G C G
G G T T T C T C T T C G C

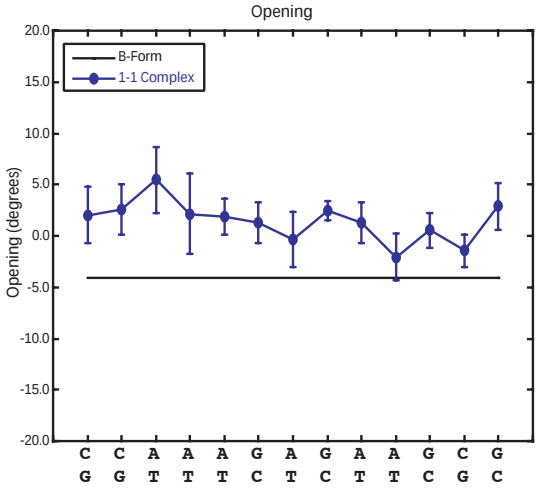

Roll
B-Form
1-1 Complex
Roll (degrees)
30.0
25.0
20.0
15.0
10.0
5.0
0.0
-5.0
-10.0
-15.0
-20.0
C C A A A G A G A A G C G
G G T T T C T C T T C G C

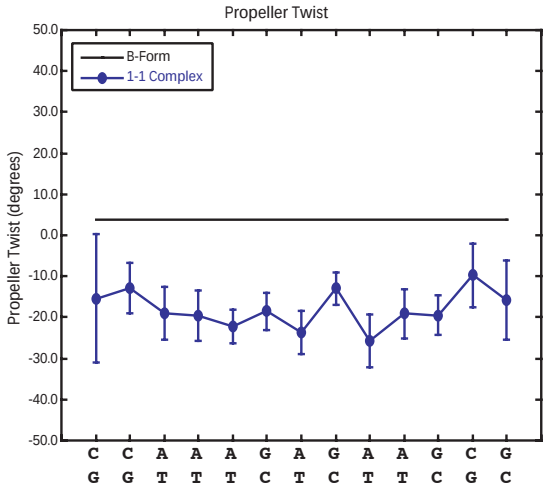

Rise
B-Form
1-1 Complex
Rise (Å)
5.0
4.0
3.0
2.0
C C A A A G A G A A G C G
G G T T T C T C T T C G C

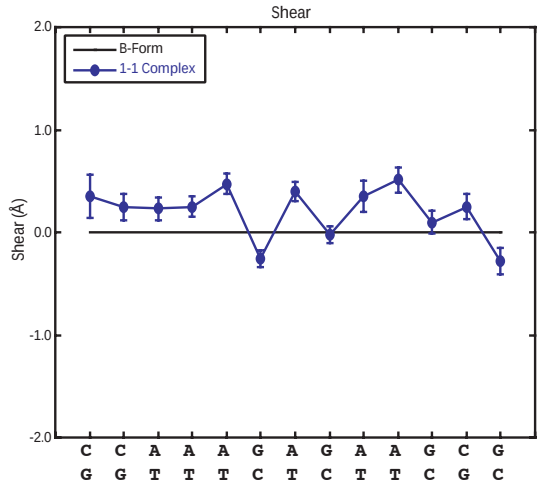

Tilt
B-Form
1-1 Complex
15.0
10.0
5.0
0.0
-5.0
-10.0
C C A A A G A G A A G C G
G G T T T C T C T T C G C

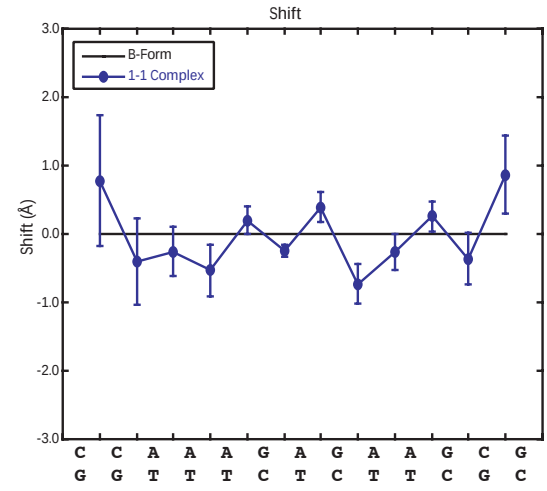

Tip
B-Form
1-1 Complex
Tip (degrees)
Tilt (degrees)
20.0
15.0
10.0
5.0
0.0
-5.0
-10.0
-15.0
-20.0
C C A A A G A G A A G C G
G G T T T C T C T T C G C

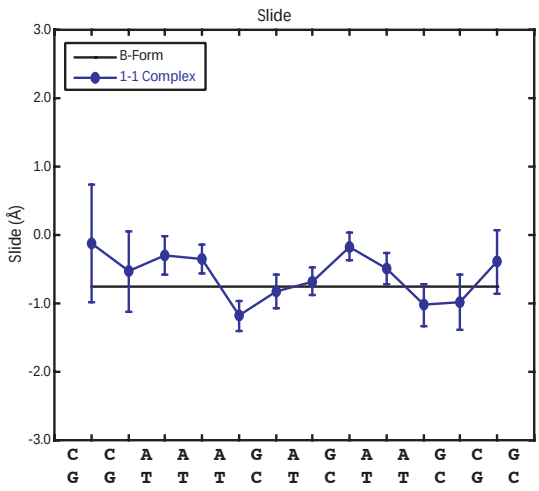

Buckle
Buckle (degrees)
B-Form
1-1 Complex
50.0
40.0
30.0
20.0
10.0
0.0
-10.0
-20.0
-30.0
-40.0
-50.0
C C A A A G A G A A G C G
G G T T T C T C T T C G C

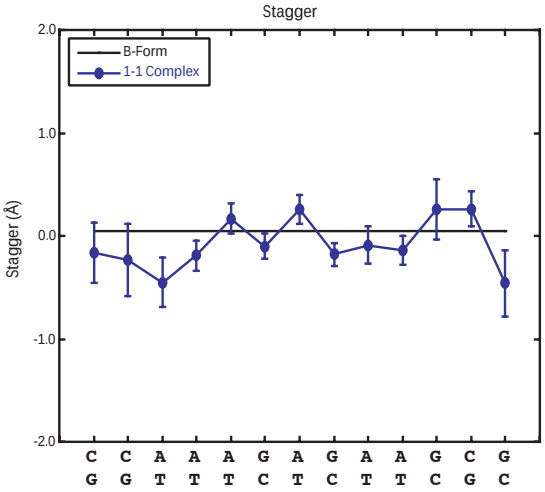

Opening
Opening (degrees)
B-Form
1-1 Complex
20.0
15.0
10.0
5.0
0.0
-5.0
-10.0
-15.0
-20.0
C C A A A G A G A A G C G
G G T T T C T C T T C G C

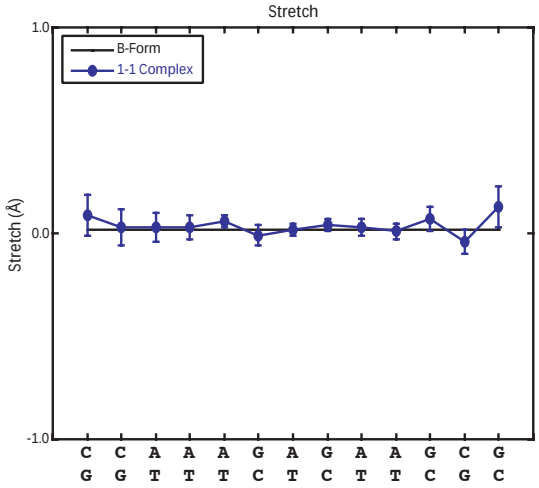

Propeller Twist
Propeller Twist (degrees)
B-Form
1-1 Complex
50.0
40.0
30.0
20.0
10.0
0.0
-10.0
-20.0
-30.0
-40.0
-50.0
C C A A A G A G A A G C G
G G T T T C T C T T C G C


Shear
Shear (Å)
B-Form
1-1 Complex
2.0
1.0
0.0
-1.0
-2.0
C C A A A G A G A A G C G
G G T T T C T C T T C G C


Shift
Shift (Å)
B-Form
1-1 Complex
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
C C A A A G A G A A G C G
G G T T T C T C T T C G C


Slide
Slide (Å)
B-Form
1-1 Complex
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
C C A A A G A G A A G C G
G G T T T C T C T T C G C


Stagger
B-Form
1-1 Complex
Stagger (Å)
2.0
1.0
0.0
-1.0
-2.0
C C A A A G A G A A G C G
G G T T T C T C T T C G C


Stretch
B-Form
1-1 Complex
Stretch (Å)
1.0
0.0
-1.0
C C A A A G A G A A G C G
G G T T T C T C T T C G C