

**Binding Site Size Limitations of Imidazole-Pyrrole Polyamides
for Recognition in the Minor Groove of DNA**

Thesis By

James J. Kelly

In Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

California Institute of Technology

Pasadena, California

1996

(Submitted December 11, 1995)

To JoyAnne

Acknowledgements

I would like to thank my advisor, Professor Peter B. Dervan, for all of his support and encouragement over the past four years. Peter's enthusiasm for doing science has been a source of tremendous inspiration during my stay at Caltech. Through Peter's guidance, I have grown in many ways, both professional and personal, and I am sure that the lessons I have learned while working in his group will serve me well in the future. I would also like to thank my committee members, Professors Dennis Dougherty, Doug Rees, Bob Grubbs, and Bill Goddard, for their guidance and support.

I would like to thank the many members of the Dervan Group, past and present, for making lab a great place to learn and play. Michelle Parks (Miz Parks) and Natalia (Nat-a-lia) Colocci have been great labmates and friends over the past four years and I've thoroughly enjoyed sharing my time and experiences here with them. I hope my friendships with these special individuals will continue to grow in the future in spite of the distances that might separate us. Turf Grass Disease Management Seminars at Brookside University with Milan Mrksich, George Best, and Jurg Hunziker were always a welcome and refreshing break from synthesis and gels. I hope that we can continue to attend future seminars on a regular basis! I am also grateful to Jason Szewczyk, Sue Swalley, Eldon Baird, Dave Liberles, and John Trauger for many helpful discussions about research and science.

I would like to thank my parents for all of the and support and encouragement they have provided me during my time at Caltech. My mother's effervescent optimism and my father's confidence in me have helped get me through the difficult times. I would also like to thank my in-laws, Dan (The Boss) and Joyce (Mrs. Boss) Millett, for all that they have done for me.

Finally, I would like to thank my family for their boundless patience and encouragement during our time here. My wife JoyAnne has been my partner through all of my endeavors and experiences at Caltech and I am truly grateful for all of her love, support, and sacrifices. My son Tyler has made the last two years here at Caltech especially pleasurable. Being greeted at the door with a "Hi Da-Da!" and bear hug each night after a day in lab has made research much more enjoyable! JoyAnne and Tyler, I dedicate this work and our future, to you.

Abstract

The discovery that the three ring polyamide Im-Py-Py-Dp containing imidazole (Im) and pyrrole (Py) carboxamides binds the DNA sequence 5'-(A,T)G(A,T)C(A,T)-3' as an antiparallel dimer offers a new model for the design of ligands for specific recognition of sequences in the minor groove containing both G,C and A,T base pairs. In Chapter 2, experiments are described in which the sequential addition of five *N*-methylpyrrolecarboxamides to the imidazole-pyrrole polyamide Im-Py-Py-Dp affords a series of six homologous polyamides, Im-(Py)₂₋₇-Dp, that differ in the size of their binding site, apparent first order binding affinity, and sequence specificity. These results demonstrate that DNA sequences up to nine base pairs in length can be specifically recognized by imidazole-pyrrole polyamides containing three to seven rings by 2:1 polyamide-DNA complex formation in the minor groove. Recognition of a nine base pair site defines the new lower limit of the binding site size that can be recognized by polyamides containing exclusively imidazole and pyrrolecarboxamides. The results of this study should provide useful guidelines for the design of new polyamides that bind longer DNA sites with enhanced affinity and specificity.

In Chapter 3 the design and synthesis of the hairpin polyamide Im-Py-Im-Py- γ -Im-Py-Im-Py-Dp is described. Quantitative DNase I footprint titration experiments reveal that Im-Py-Im-Py- γ -Im-Py-Im-Py-Dp binds six base pair 5'-(A,T)GCGC(A,T)-3' sequences with 30-fold higher affinity than the unlinked polyamide Im-Py-Im-Py-Dp. The hairpin polyamide does not discriminate between A•T and T•A at the first and sixth positions of the binding site as three sites 5'-TGCGCT-3', 5'-TGCGCA-3', and 5'-AGCGCT-3' are bound with similar affinity. However, Im-Py-Im-Py- γ -Im-Py-Im-Py-Dp is specific for and discriminates between G•C and C•G base pairs in the 5'-GCGC-3'

core as evidenced by lower affinities for the mismatched sites 5'-AACGCA-3', 5'-TGCGTT-3', 5'-TGCGGT-3', and 5'-ACCGCT-3'.

In Chapter 4, experiments are described in which a kinetically stable hexa-aza Schiff base La^{3+} complex is covalently attached to a Tat(49-72) peptide which has been shown to bind the HIV-1 TAR RNA sequence. Although these metallo-peptides cleave TAR site-specifically in the hexanucleotide loop to afford products consistent with hydrolysis, a series of control experiments suggests that the observed cleavage is not caused by a sequence-specifically bound Tat(49-72)- La(L)^{3+} peptide.

Table of Contents

Acknowledgements	iv
Abstract	vi
Table of Contents	viii
Figures and Tables	x
Chapter 1: Introduction	1
DNA Recognition in Biological Systems	1
DNA Structure	1
Protein-DNA Complexes	5
Oligonucleotide-Directed Triple Helix Formation	10
Small Molecule Recognition of DNA	11
Structure of 1:1 Polyamide-DNA Complexes	13
Assays for Sequence-Specific DNA Binding	16
Quantitative DNase I Footprint Titration Experiments	16
Polyamides for Binding Longer A,T Tracts	18
Polyamides Designed to Bind Mixed Sequences from 1:1 Models	20
2:1 Distamycin-DNA Complexes	21
Recognition of Mixed Sequences by 2:1 Antiparallel Polyamide Dimers	21
Antiparallel Side-by-Side Heterodimer	25
Recognition of 5'-GCGC-3' Sequences	29
Covalently Linked Polyamides	30
Hairpin and Cyclic Polyamides	34
Description of This Work	36
References and Notes	37

Chapter 2: Binding Site Size Limitations of Imidazole-Pyrrole Polyamides for Recognition in the Minor Groove of DNA	47
Introduction	47
Results	52
Discussion	81
Experimental Section	88
References and Notes	105
Chapter 3: Recognition of 5'-(A,T)GCGC(A,T) Sequences by a Hairpin Polyamide	109
Introduction	109
Results	111
Discussion	118
Experimental Section	127
References and Notes	139
Chapter 4: Studies Towards the Sequence-Specific Hydrolysis of HIV-1 TAR RNA	141
Introduction	141
Results	145
Discussion	161
Experimental Section	166
References and Notes	176
Addendum to Chapter 4, Part 1	181
Addendum to Chapter 4, Part 2	183

Figures and Tables

Chapter 1		Page
Figures		
Figure 1.1.	Structure of B-form double helical DNA	2
Figure 1.2.	The four natural base pairs of DNA	4
Figure 1.3.	Base pair discrimination in the minor groove	6
Figure 1.4.	X-ray crystal structure of the DNA binding domain of Hin recombinase bound to a recombination half-site	8
Figure 1.5.	Oligonucleotide-directed triple helix formation	11
Figure 1.6.	Structure of netropsin and distamycin	12
Figure 1.7.	1:1 Models for netropsin and distamycin binding to A,T-rich DNA	13
Figure 1.8.	X-ray crystal structure of the 1:1 complex between distamycin and a 5'-AAATTT-3' site	14
Figure 1.9.	Affinity cleaving and footprinting assays for sequence-specific DNA binding	17
Figure 1.10.	Asymmetric DNA cleavage patterns generated by affinity cleavage with EDTA•Fe(II)	17
Figure 1.11.	Illustrative gel for a quantitative DNase I footprint titration experiment	19
Figure 1.12.	Representation of a binding isotherm generated from a quantitative DNase I footprint titration experiment	19
Figure 1.13.	Structure of Pyr-Py-Py-Dp and Im-Py-Py-Dp	20
Figure 1.14.	X-ray crystal structure of the 2:1 distamycin•5'-AAATT-3' complex	22

Figure 1.15.	Model of the 2:1 distamycin•5'-AAATT-3' complex	24
Figure 1.16.	Models of the 2:1 complexes between Pyr-Py-Py-Dp and Im-Py-Py-Dp and a 5'-TGTC A-3' site	25
Figure 1.17.	Structure of the (Im-Py-Py-Dp) ₂ •5'-TGTC A-3' complex	26
Figure 1.18.	Model of the Im-Py-Py-Dp/distamycin•5'-TGTTA-3' complex	28
Figure 1.19.	Structure of Im-Py-Im-Py-Dp and Im-Py-Im-Py-EDTA•Fe(II)	29
Figure 1.20.	Structure of the (Im-Py-Im-Py-Dp) ₂ •5'-TGCGCA-3' complex	31
Figure 1.21.	Model of the (Im-Py-Im-Py-Dp) ₂ •5'-TGCGCA-3' complex	33
Figure 1.22.	Structure of (Pyr-Py-Py-Dp) ₂ -C ₄ and Im-Py-Py-Dp-C ₄ -Py-Py-Py-Dp	33
Figure 1.23.	Models of Im-Py-Py-γ-Py-Py-Py-Dp and <i>cyclo</i> -(Im-Py-Py-γ-Py-Py-Py) bound at a 5'-TGTTA-3' site	35

Chapter 2

Figures

Figure 2.1.	2:1 Binding models for Im-(Py) ₂ -Dp and Im-Py-Im-Py-Dp.	48
Figure 2.2.	Structure of polyamides Im-(Py) _{2,7} -Dp	50
Figure 2.3.	Proposed 2:1 binding models for polyamides Im-(Py) _{2,7} -Dp	51
Figure 2.4.	Synthesis of polyamides Im-(Py) _{3,7} -Dp	53
Figure 2.5.	Illustration of plasmids pJK5, pJK6, pJK7, pJK8, pJK9, and pJK10	56
Figure 2.6.	MPE•Fe(II) footprinting experiments	57
Figure 2.7.	Histograms from MPE•Fe(II) footprinting experiments	65
Figure 2.8.	Quantitative DNase I footprint titration experiments	67
Figure 2.9.	Binding isotherms derived from quantitative DNase I	79

footprint titration experiments

Figure 2.10.	Plot of apparent first order binding affinities	83
Figure 2.11.	Plot of specificity for a binding at a match site	86

Tables

Table I.	Apparent first order binding affinities	80
----------	---	----

Chapter 3**Figures**

Figure 3.1.	Model of the 2:1 complex of Im-Py-Im-Py-Dp at a 5'-TGCGCA-3' site	110
Figure 3.2.	Structure of polyamides 1 and 2	111
Figure 3.3.	Synthesis of hairpin polyamide 2	112
Figure 3.4.	Synthesis of acid 12	113
Figure 3.5.	Illustration of the 381 base pair restriction fragment used in this study	114
Figure 3.6.	MPE•Fe(II) footprinting with hairpin polyamide 2	115
Figure 3.7.	Histograms from MPE•Fe(II) footprinting experiments	117
Figure 3.8.	Quantitative DNase I footprint titration experiments with hairpin polyamide 2	119
Figure 3.9.	Binding isotherms for polyamides 1 and 2 at a 5'-TGCGCT-3' site	122
Figure 3.10.	Proposed binding model of the 2 •5'-TGCGCT-3' complex	124
Figure 3.11.	Space-filling model of the 2 •5'-TGCGCT-3' complex	125
Figure 3.12.	Putative structure of N-acyl imidazolium derivative 13	129

Tables

Table I.	First order affinity constants for polyamides 1 and 2	121
----------	---	-----

Chapter 4

Figures

Figure 4.1.	Sequence of the HIV-1 Tat protein	142
Figure 4.2.	Predicted secondary structure of the HIV-1 TAR RNA sequence	143
Figure 4.3.	Structure of the hexa-aza Schiff base lanthanide ion macrocycle, Ln(L)^{3+}	144
Figure 4.4.	Structure of Tat- La(L)^{3+} peptides 1-3	145
Figure 4.5.	Synthesis of dialdehyde 4	146
Figure 4.6.	Synthesis of Tat- La(L)^{3+} peptides 1-3	147
Figure 4.7.	Cleavage of HIV-1 TAR RNA by Tat(48-72, Cys^{48} - La(L)^{3+})	149
Figure 4.8.	Sequence and predicted secondary structure of HIV-1 TAR RNA showing the phosphodiesterases cleaved by Tat(48-72, Cys^{48} - La(L)^{3+})	151
Figure 4.9.	Analysis of the termini of the TAR cleavage products resulting from treatment with Tat(48-72, Cys^{48} - La(L)^{3+})	152
Figure 4.10.	TAR cleavage by $\text{La(NO}_3)_3$ and La(L)^{3+}	156
Figure 4.11.	Sequence and predicted secondary structure of HIV-1 TAR RNA showing the phosphodiesterases cleaved by $\text{La(NO}_3)_3$	158
Figure 4.12.	Sequence and predicted secondary structure of HIV-1 TAR RNA showing the phosphodiesterases cleaved by $\text{La(NO}_3)_3$ in the presence of 100 μM Tat(49-72)	159
Figure 4.13.	Cleavage of TAR by Tat(48-72, Cys^{48} - La(L)^{3+}) in the presence of 100 μM Tat(49-72)	160
Figure 4.14.	Cleavage of a bulgeless TAR derivative by	161

Tat(48-72, Cys⁴⁸-La(L)³⁺)

Figure 4.15.	One explanation for the loop-specific cleavage of TAR by peptides 1-3	163
Figure 4.16.	Designed catalysts for sequence-specific RNA hydrolysis	165
Figure 4.17.	Predicted secondary structure of RNA 16	181
Figure 4.18.	Cleavage of RNA 16 by La(NO ₃) ₃	182
Figure 4.19.	Structure of nucleophile-appended lanthanide complexes designed for DNA transesterification	183

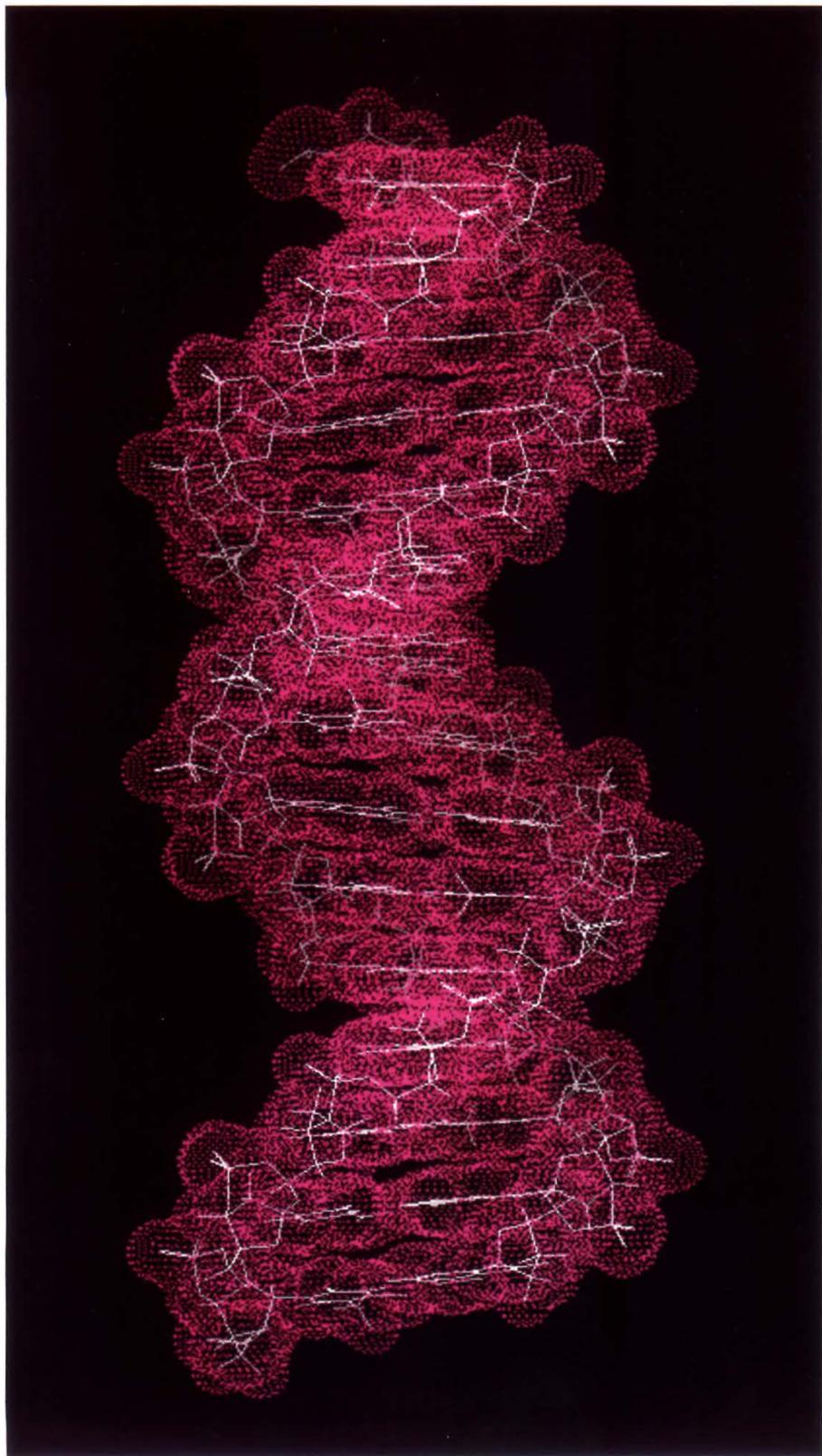
Chapter 1

Introduction

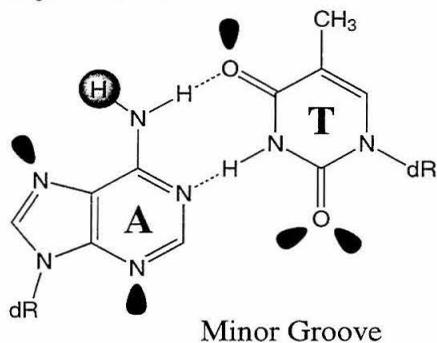
DNA Recognition in Biological Systems. Deoxyribonucleic acid (DNA) functions as the chemical blueprint for life by encoding all of the proteins and ribonucleic acid (RNA) necessary for cellular function. The specific recognition of particular DNA sequences by proteins is critical for a variety of cellular processes including gene replication, transcription, and cell division. The critical importance of DNA in these processes makes the genetic material an attractive target for the development of chemical approaches to sequence-specific recognition. The design of materials for the specific recognition of particular DNA sequences may afford a new class of reagents that are capable of controlling gene expression and that may be useful in human therapeutics.

DNA Structure. The structure of B-form right-handed double helical DNA is well-established.¹⁻² The DNA double helix is composed of two antiparallel polydeoxyribonucleotide strands which associate via Watson-Crick hydrogen bonds between complementary heterocyclic bases on opposite strands and is stabilized by vertical π - π stacking interactions between bases on the same strand (Figure 1.1). In double helical DNA the sugar-phosphate backbone is displaced away from the helical axis, creating a wide and shallow major groove and a narrow and deep minor groove along the cylindrical double helix. The edges of the base pairs are located at the bottom of the grooves and present different chemical functionalities capable of interacting with DNA-binding ligands (Figure 1.2). In the major groove the distinct pattern of hydrogen bond donor and acceptor groups allows all four base pairs to be distinguished. Base pair discrimination in the

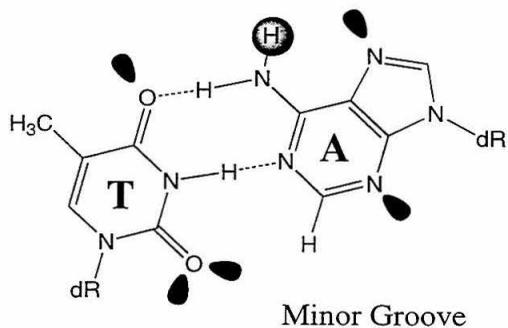
Figure 1.1. Structure of B-form double helical DNA. The DNA is represented as a line model (white) with an overlaid van der Waals surface (magenta). The wide and shallow major groove and the narrow and deep minor groove are clearly visible.



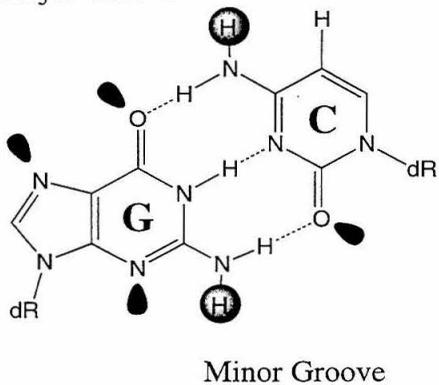
Major Groove



Major Groove



Major Groove



Major Groove

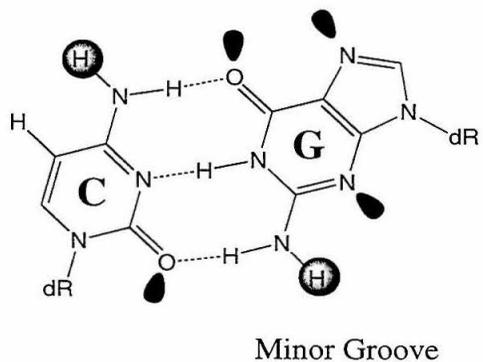


Figure 1.2. Structure of the four natural Watson-Crick base pairs showing the functional groups displayed in the major and minor grooves. Lone pairs that function as potential hydrogen bond acceptors are represented as black lobes and potential hydrogen bond donors are indicated as gray spheres.

minor groove is less distinct. The A,T base pair presents two hydrogen bond acceptors, the purine N3 and the pyrimidine O2 atoms (Figure 1.3). Due to the approximate two-fold symmetry of these atoms, the minor groove surfaces of A•T and T•A base pairs are approximately degenerate. The GC base pair retains these two hydrogen bond acceptors but also presents the 2-amino group of guanine as a hydrogen bond donor. Because this hydrogen bond donor lies closer to the guanine-containing strand, G•C and C•G base pairs are distinguishable in the minor groove. In addition to the groove edges of the base pairs sequence-dependent structural variations, conformational properties, and counterion organization can all distinguish particular DNA sequences. DNA-binding ligands generally recognize their target sites by a combination of hydrogen bonding interactions with the functional groups displayed on the floor of the grooves and electrostatic, van der Waals interactions with the sugar-phosphate backbone, and overall shape complementarity with the binding site.³

Protein-DNA Complexes. High resolution structural data from x-ray and NMR studies of protein-DNA complexes has dramatically enhanced our understanding of sequence-specific recognition of DNA by native proteins. Several different structural motifs for sequence-specific DNA binding have been identified including the zinc finger, leucine zipper, homeodomain, and helix-turn-helix motifs.⁴ These proteins generally recognize their binding sites by a combination of hydrogen bonding interactions between the protein backbone and side chains with the bases in the major and/or minor groove, salt bridges between positively charged side chains and specific phosphate groups, as well as the overall shape complementarity between the protein and the DNA binding site (Figure 1.4). Due to the complexity of DNA-protein interactions the design of proteins for recognition of designated DNA sequences remains a major challenge although progress is being made in this area, particularly with the zinc finger motif.⁵

Figure 1.3. Base pair discrimination in the minor groove of B-form double helical DNA. The A•T and T•A base pairs (top) and G•C and C•G base pairs (bottom) are shown. The DNA is shown in CPK form (yellow) with N3 of A and G and N2 of G colored blue and O2 of T and C colored red. The hydrogen atom of N2 of G is colored white.

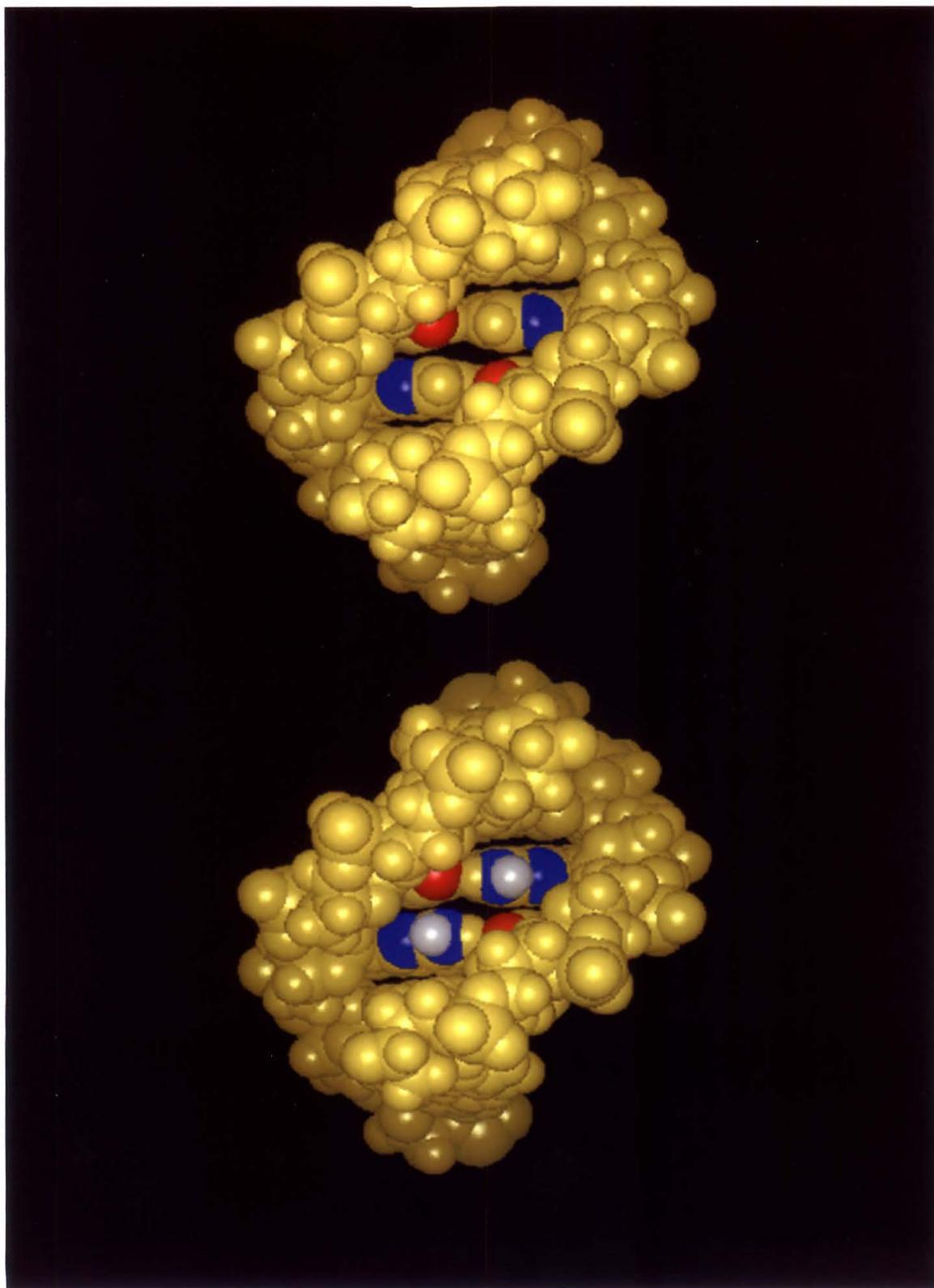
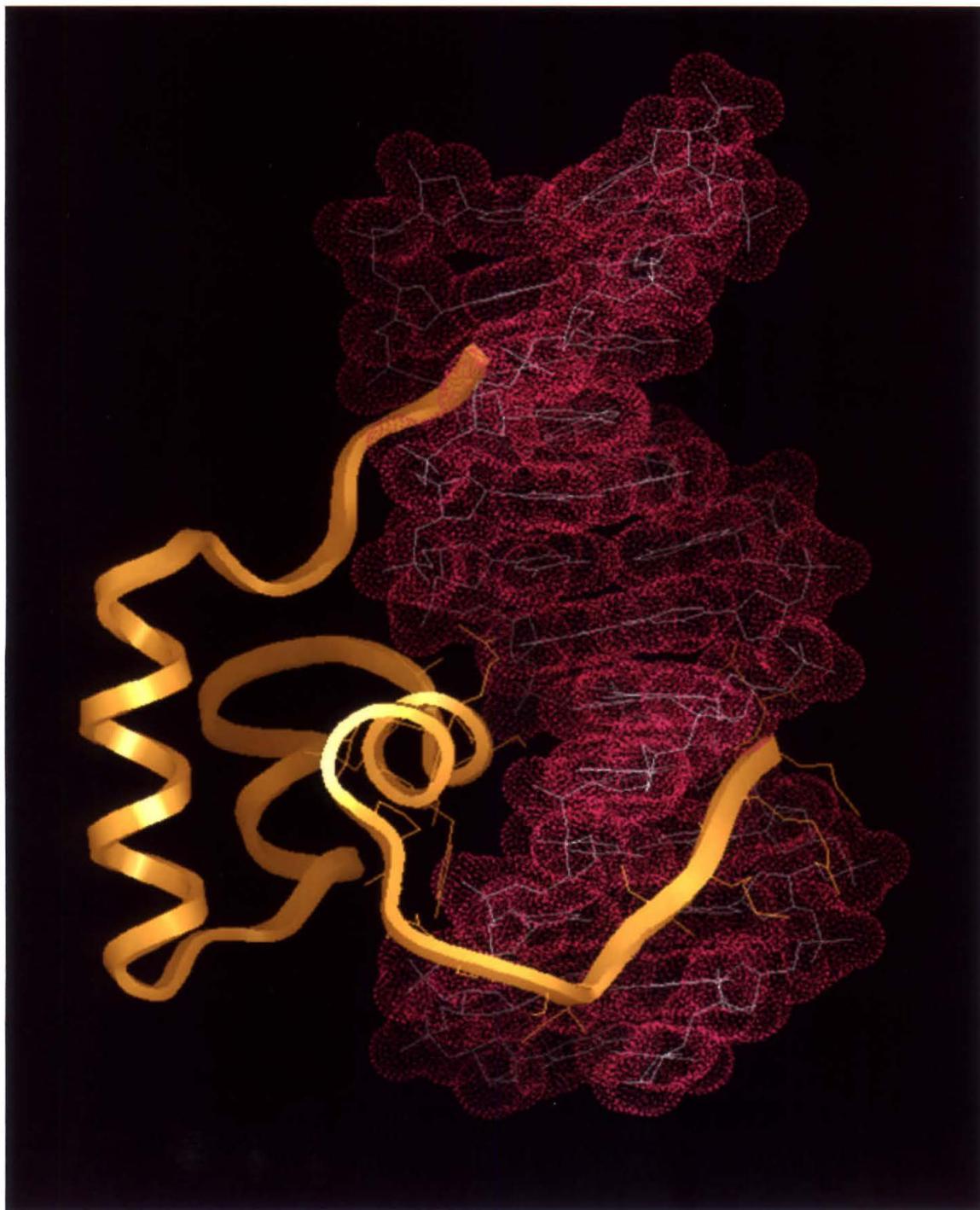


Figure 1.4. High resolution x-ray crystal structure of the 52 amino acid residue DNA binding domain of Hin recombinase bound to a DNA recombination half-site.^{4d} The DNA is represented as a line model (white) with an overlaid van der Waals surface (magenta). The protein is a member of the helix-turn-helix family and is represented as a yellow ribbon. The side chains of the residues comprising the DNA recognition helix bound in the major groove are shown as well as the side chains of the N-terminal arm residues bound in the adjacent minor groove.



Oligonucleotide-Directed Triple Helix Formation. In 1987 Moser and Dervan demonstrated that a short homopyrimidine oligodeoxynucleotide could sequence-specifically bind in the major groove of double helical DNA parallel to a homopurine tract (Figure 1.5).⁶⁻⁷ Specificity within the triple-helical complex arises from the formation of T•AT and C⁺•GC base triplets via hydrogen bonding interactions between the pyrimidine bases on the third strand and the Hoogsteen face of the purine bases of the Watson-Crick base pairs. The sequence specificity of triple helix formation has been powerfully demonstrated in a series of experiments where pyrimidine oligonucleotides were capable of single-site binding to bacteriophage, yeast, and human chromosomal DNA.⁸ Additional experiments have been performed to determine the effects of oligonucleotide length,^{6,9} modified pyrimidines,¹⁰ single base mismatches,^{6,9,11} pH,^{10b,12} cation concentration and valence,¹³ temperature,¹⁴ backbone composition,¹⁵ cooperativity,¹⁶ and sequence composition¹⁷ on the stability and specificity of triple helix formation in the pyrimidine motif. Purine-rich oligonucleotides have been shown to bind in the minor groove of double helical DNA antiparallel to a homopurine strand by the formation of specific G•GC, A•AT, and T•AT base triplets.¹⁸ More recently, a third nonnatural motif for recognition of DNA by triple helix formation has been discovered.¹⁹ In the parallel purine motif the putative base triplet geometry of the antiparallel purine motif is maintained while the third strand orientation is reversed by moving the third strand deoxyribose from the natural N9 position to N7. Experiments have shown that A•T and G•C base pairs can be recognized by the formation of ⁷A•AT and ⁷G•GC base triplets. Current efforts are directed towards the synthesis of oligonucleotides containing exclusively N7 purines that bind with high affinity and specificity. While these three triple-helical motifs provide a versatile method for binding homopurine sequences, the recognition of mixed purine-pyrimidine sequences remains an unsolved problem. The design of novel bases for recognition of T•A and C•G

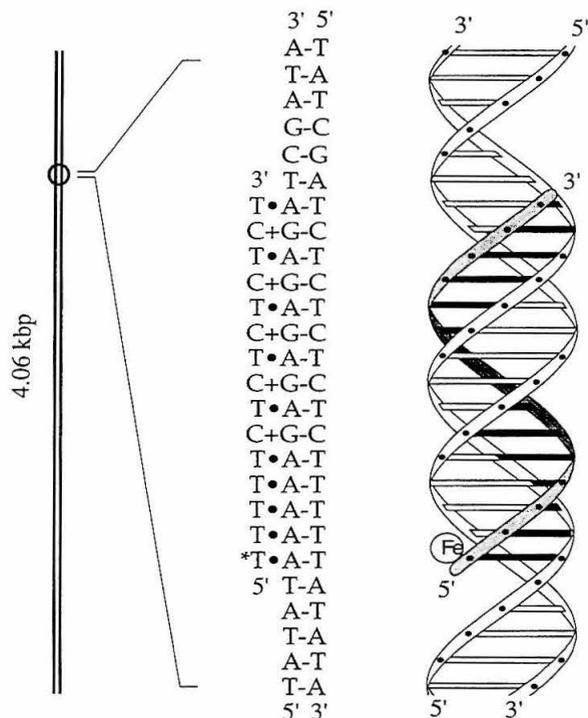


Figure 1.5. Oligonucleotide-directed triple helix formation. Ribbon model of a triple-helical complex between a 15 base pair purine sequence and a 15mer pyrimidine oligonucleotide on a 4.06-kbp DNA restriction fragment.

base pairs via new hydrogen bonding patterns and favorable steric interactions with the major groove edges of these base pairs is an ongoing effort.²⁰⁻²¹ Progress has been hampered by the lack of high-resolution structural data and an incomplete understanding of the forces that stabilize triple helix formation, particularly base stacking interactions in the third strand.²² Protein-mediated triple helix formation with RecA nucleoprotein filaments that appear to bind in the *minor* groove may become a versatile method for recognizing DNA sequences containing all four of the natural base pairs.²³⁻²⁴

Small Molecule Recognition of DNA. A wide variety of small molecules (MW < 1500) bind double helical DNA sequence specifically.²⁵ Examples include distamycin,

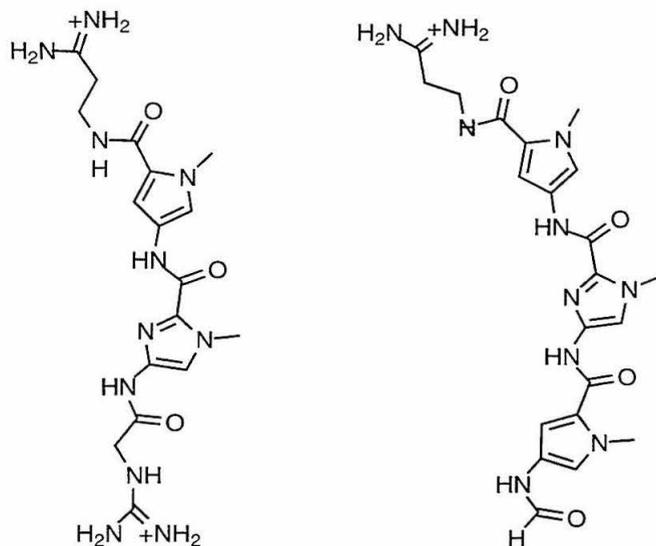


Figure 1.6. Structure of netropsin and distamycin A.

netropsin, chromomycin, and the enediyne antibiotics neocarzinostatin, calicheamycin, and dynemycin. These ligands generally recognize DNA by binding in the minor groove and/or intercalating between the base pairs.²⁶ Netropsin and distamycin are di- and tri-*N*-methylpyrrolicarboxamides, respectively, isolated from *Streptomyces* (Figure 1.6).²⁷ Netropsin contains an N-terminal guanadinium group and a C-terminal amidinium group while distamycin contains an N-terminal formamide group and a C-terminal amidinium group. These compounds exhibit antifungal, antimitotic, and antiviral properties in vivo presumably by binding A,T-rich DNA sequences.²⁸⁻²⁹ The ability to synthesize analogs of these natural products coupled with direct assays for DNA binding provides chemists with the opportunity to study sequence-specific molecular recognition of double helical DNA. The development of ligands that bind in the minor groove of DNA with high affinity and sequence specificity may afford new materials that have use in controlling gene expression and as human therapeutics.

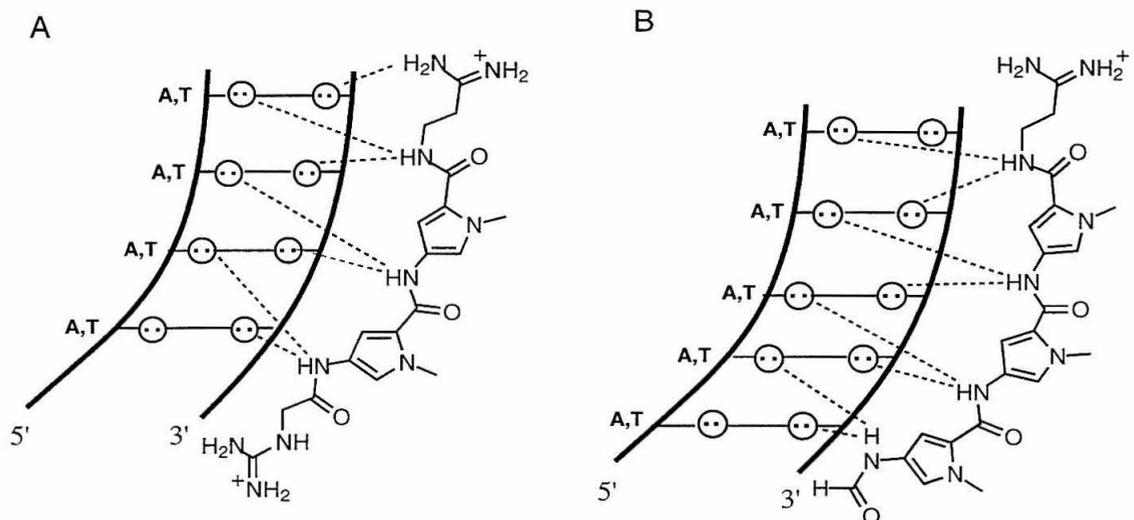
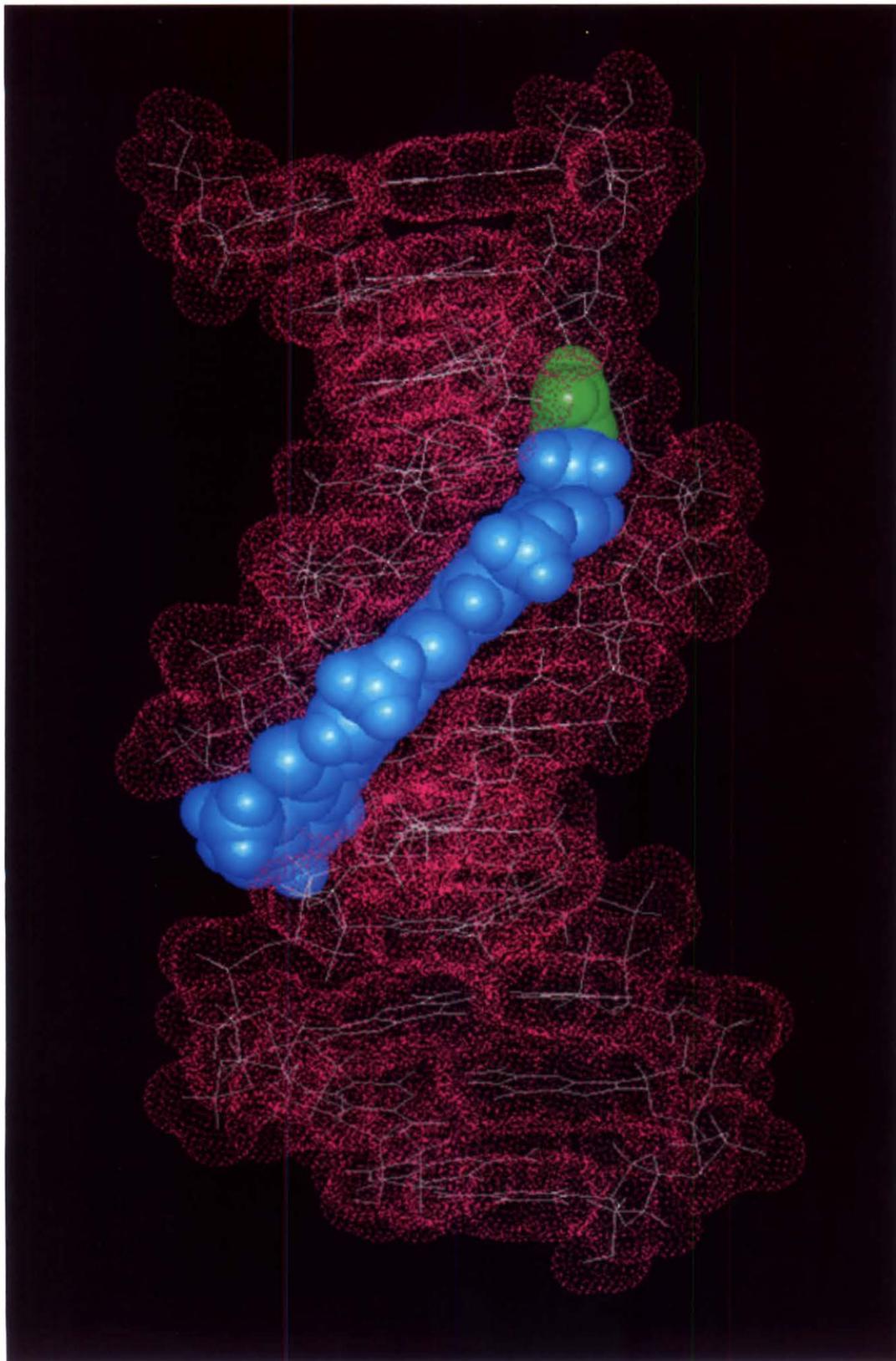


Figure 1.7. 1:1 Models for netropsin (A) and distamycin (B) binding to A,T-rich DNA.

Structure of 1:1 Polyamide-DNA Complexes. High-resolution x-ray³⁰ and NMR³¹ studies of netropsin- and distamycin-DNA complexes reveal how sequence-specific binding to four and five base pair A,T-rich sequences, respectively, is accomplished. The crescent-shaped polyamides are bound in the center of the minor groove of an A,T-rich sequence with the *N*-methylpyrrolecarboxamides twisting in a screw sense to match the walls of the minor groove, giving a favorable shape complementarity for the ligand. The carboxamide NH's form bifurcated hydrogen bonds with adenine N3 and thymine O2 atoms on the floor of the minor groove (Figure 1.7). The pyrrole rings fill the groove and form extensive van der Waals contacts with the walls of the minor groove (Figure 1.8). The aromatic hydrogens of the *N*-methylpyrrole rings are set too deeply in the minor groove to accommodate the guanine 2-amino group of a G,C base pair, affording binding specificity for A,T-rich sequences. The positively charged amidinium and guanidinium groups of the polyamides lie along the floor of the groove and provide favorable electrostatic interactions with the negatively-charged phosphodiester backbone of DNA.³²⁻³³

Figure 1.8. High resolution x-ray crystal structure of the 1:1 complex between distamycin and a 5'-AAATTT-3' site.^{30d} The DNA is represented as a line model (white) with an overlaid van der Waals surface (magenta). The *N*-methylpyrrolicarboxamides and the positively charged amidinium group are colored cyan and green, respectively.



Assays for Sequence-Specific DNA Binding. Two complementary techniques, affinity cleaving and footprinting, are routinely used for determining the binding affinity and sequence specificity of DNA-binding ligands.³⁴ In affinity cleaving a non-specific DNA cleaving agent is covalently attached to a sequence-specific DNA binding molecule to afford a sequence-specific DNA cleaving molecule. A commonly used non-specific cleaving moiety is EDTA•Fe(II) which in the presence of reducing agents such as dithiothreitol or sodium ascorbate generates a diffusible oxidant, presumably hydroxyl radical, capable of hydrogen atom abstraction from the sugar-phosphate backbone at locations near the ligand binding site. Analysis of the cleavage products from an ³²P end-labeled DNA restriction by high-resolution denaturing polyacrylamide gel electrophoresis provides information about the location and size of the binding site and the orientation of the affinity cleaving analog on the DNA (Figure 1.9). Analysis of the asymmetric pattern of the cleavage bands provides direct information on the groove location of EDTA•Fe(II) (Figure 1.10). EDTA•Fe(II) positioned in the major groove affords cleavage products that are asymmetrically shifted in the 5' direction, whereas if the cleaving agent is located in the minor groove the cleavage products are 3' shifted. Affinity cleaving has been used in studies of sequence specific DNA recognition by proteins,³⁵ oligonucleotides,^{9,36} and small molecules.³⁷ In DNA footprinting a ³²P-end labeled restriction fragment and the DNA-binding ligand are incubated in solution, after which they are treated with a non-specific DNA cleaving agent such as DNase I³⁸ or the synthetic reagent methidiumpropyl-EDTA•Fe(II)³⁹ (MPE•Fe(II)) which cleave DNA at positions not bound by the ligand. Analysis of the cleavage products by high-resolution denaturing polyacrylamide gel electrophoresis reveals the location and size of the binding site (Figure 1.9).

Quantitative DNase I Footprint Titration Experiments. The apparent first order equilibrium binding affinities of ligands for discrete DNA sequences can be

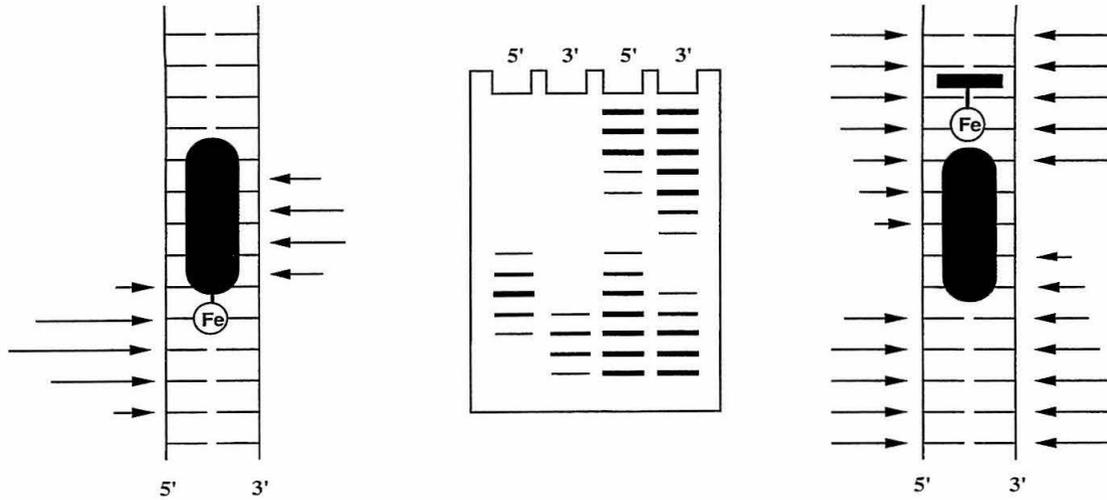


Figure 1.9. Affinity cleaving (left) and footprinting (right) techniques for the analysis of sequence-specific binding to double helical DNA.

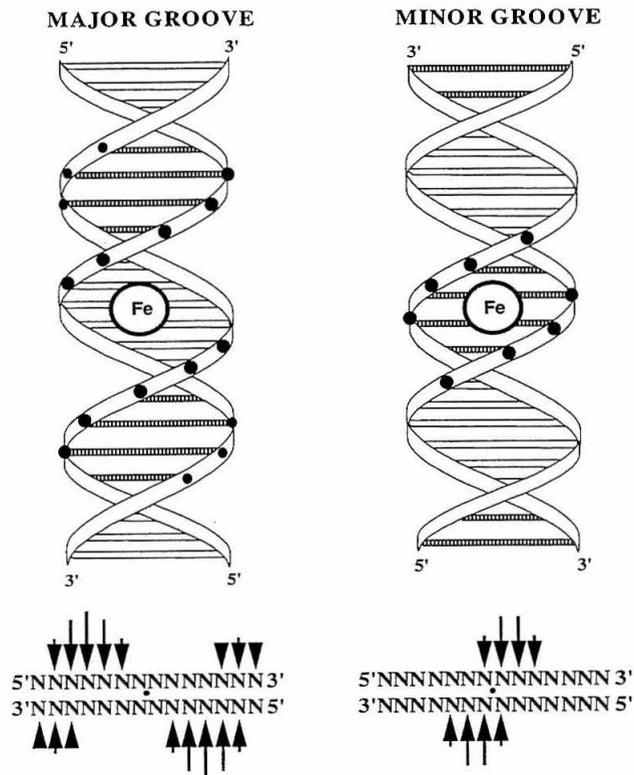


Figure 1.10. Models for the asymmetric DNA cleavage pattern generated by EDTA•Fe(II) positioned in the major and minor grooves of right-handed double helical DNA.

measured by quantitative DNase I footprint titration experiments. This technique has been well-developed by Ackers and coworkers for measuring the binding affinity of proteins and small molecules.⁴⁰⁻⁴¹ In a quantitative DNase I footprint titration experiment, footprinting reactions are carried out over a five order of magnitude range of ligand concentration. After limited digestion with DNase I the cleavage products are separated on a high-resolution denaturing polyacrylamide gel along with cleavage products from a control reaction where no ligand was added (Figure 1.11). The gel is imaged using storage phosphor autoradiography and the data is analyzed by performing volume integrations of the ligand binding site and a reference site at which DNase I cleavage is invariant.⁴² The fractional occupation of the site, θ_{app} , is calculated by comparing the extent of protection at the target site to that of the reference site and normalizing to the control lane. The ($[L]_{\text{tot}}$, θ_{app}) data points are fit to Langmuir binding isotherm by minimizing the difference between θ_{app} and θ_{fit} using the modified Hill equation:

$$\theta_{\text{fit}} = \theta_{\text{min}} + (\theta_{\text{max}} - \theta_{\text{min}}) \frac{K_a^{-n} [L]_{\text{tot}}^n}{1 + K_a^{-n} [L]_{\text{tot}}^n} \quad (1)$$

where $[L]_{\text{tot}}$ corresponds to the total ligand concentration, K_a corresponds to the apparent monomeric association constant, n is the Hill coefficient, and θ_{max} and θ_{min} represent the experimentally determined site saturation values when the binding site is unoccupied or saturated, respectively (Figure 1.12).

Polyamides for Binding Longer A,T Tracts. The 1:1 polyamide-DNA models derived from high-resolution netropsin- and distamycin-DNA structures aided in the design of polyamides that bind larger sequences of A,T-rich DNA.⁴³⁻⁴⁷ Youngquist and Dervan demonstrated that a series of polyamides containing three to nine *N*-methylpyrrolicarboxamides are capable of binding four to eleven A,T base pairs,

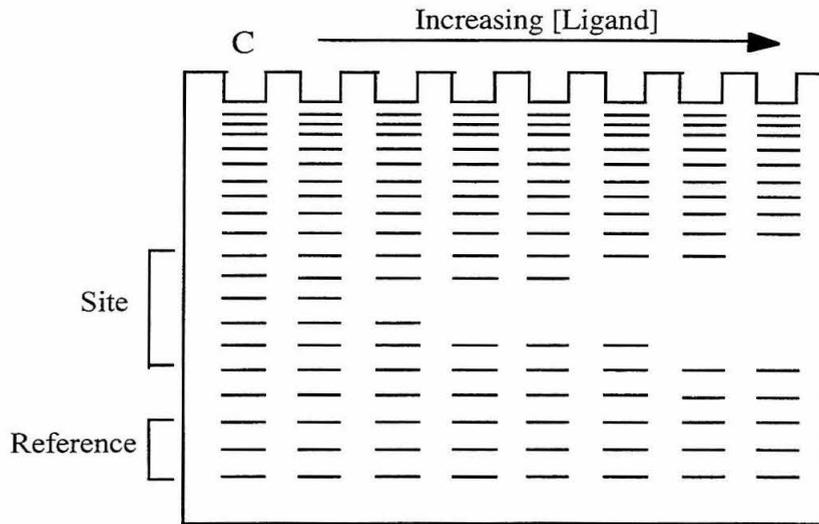


Figure 1.11. Illustrative gel from a quantitative DNase I footprint titration experiment.

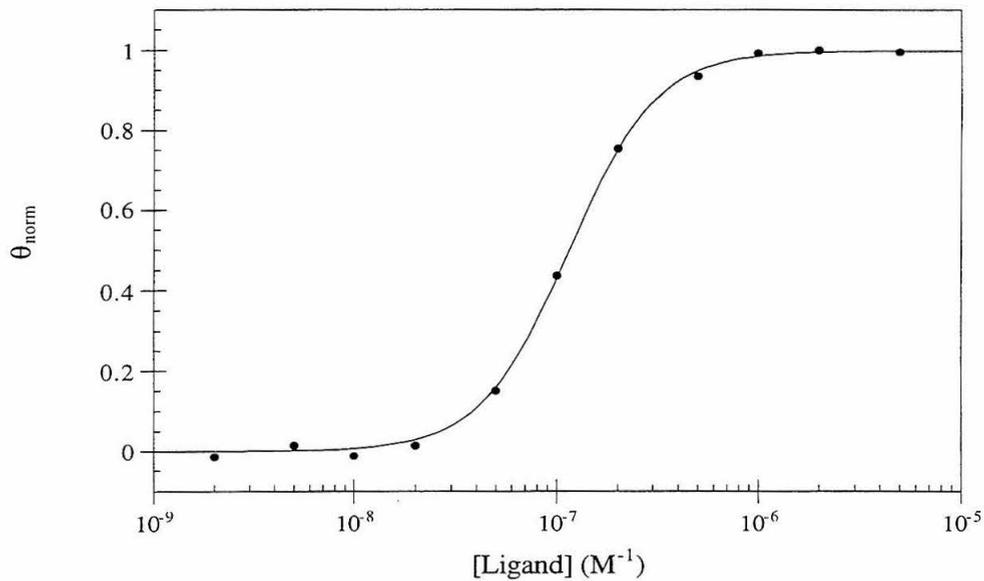


Figure 1.12. Example of a binding isotherm generated from a quantitative DNase I footprint titration experiment where the fractional occupancy, θ_{norm} , and $\log[L]$ are reported on the vertical and horizontal axes, respectively. The curve through the data points is the best-fit Langmuir binding isotherm. The ligand concentration at half-occupancy is the first order dissociation constant.

respectively.⁴⁴ In subsequent work a polyamide containing three tetra-*N*-methylpyrrole-carboxamide units coupled head-to-tail with β -alanine specifically bound 16 contiguous A,T base pairs, a turn and a half of the double helix, in the minor groove.⁴⁵

Polyamides Designed to Bind Mixed Sequences from 1:1 Models. One approach to the design of polyamides for recognition of G,C base pairs based on 1:1 polyamide-DNA models involved incorporating hydrogen bond acceptor atoms on ligands which could form specific hydrogen bonds with the 2-amino group of guanine on the floor of the minor groove.⁴⁸ Distamycin analogs in which the pyrrole rings were substituted with imidazole and thiazole rings displayed increased tolerance for G,C base pairs in their binding site, but with an overall loss of specificity.⁴⁹ Two polyamides designed from 1:1 models are specific for mixed sequences. Footprinting and affinity cleavage experiments reveal that the pyridine- (Pyr) and imidazole (Im)-pyrrole (Py) polyamides Pyr-Py-Py-Dp and Im-Py-Py-Dp specifically recognize five base pair 5'-(A,T)G(A,T)C(A,T)-3' sequences (Figure 1.13).⁵⁰⁻⁵¹ However, 1:1 polyamide-DNA models were inadequate for rationalizing the presence of G,C base pairs in the second and fourth positions of the binding site.

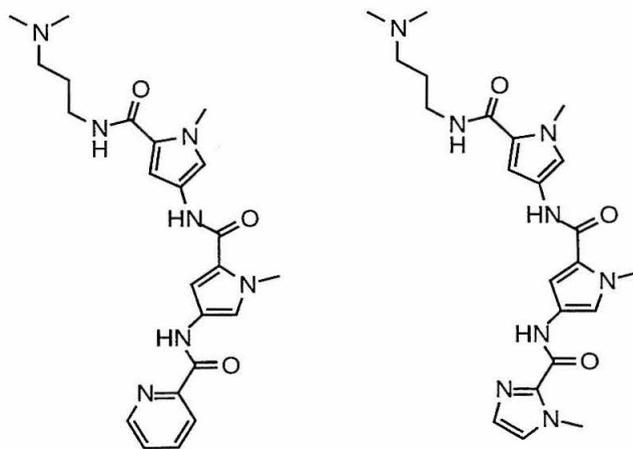
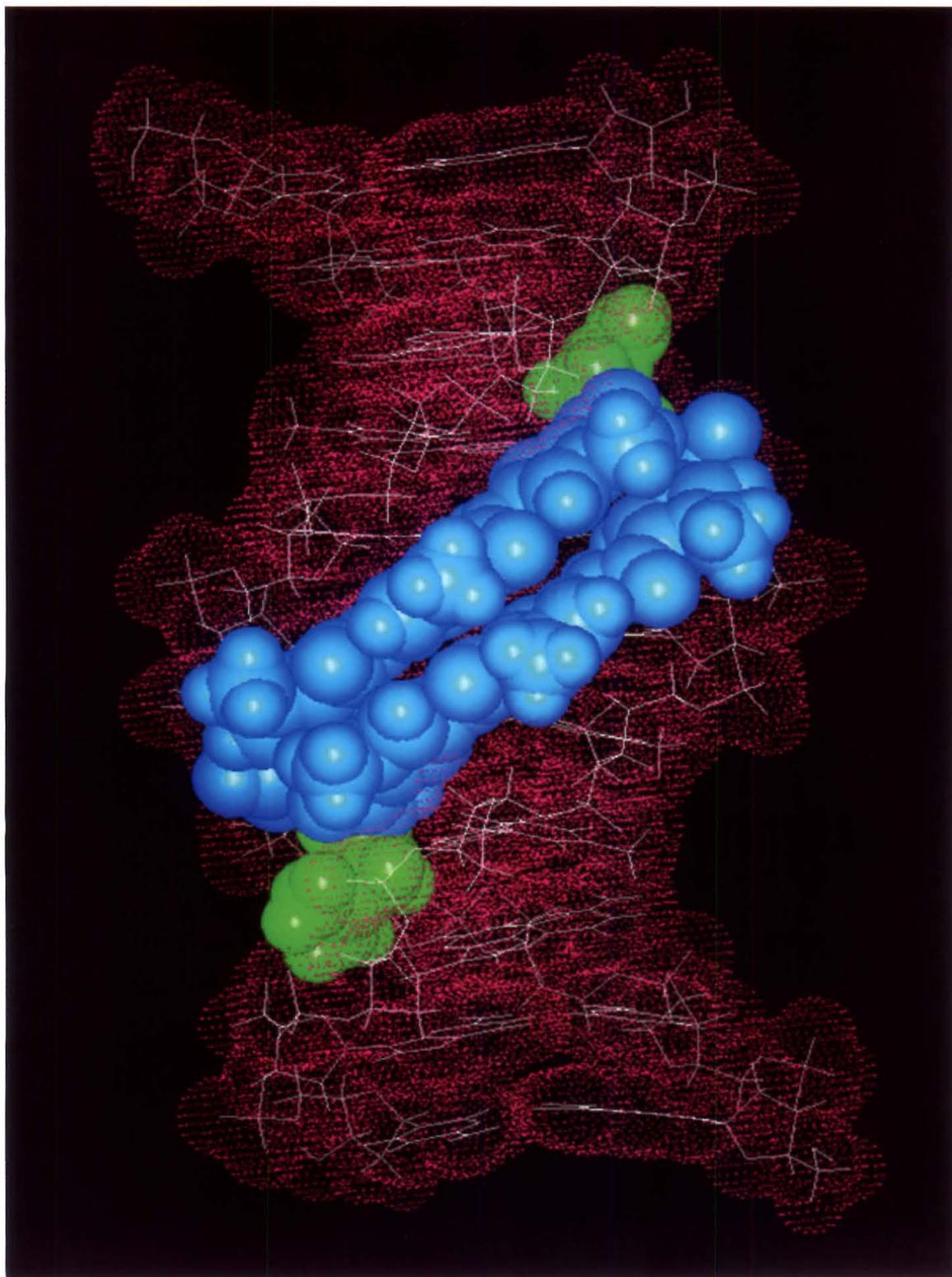


Figure 1.13. Structure of Pyr-Py-Py-Dp and Im-Py-Py-Dp.

2:1 Distamycin-DNA Complexes. Two-dimensional NMR studies by Pelton and Wemmer have demonstrated that distamycin at high concentrations is capable of binding in the minor groove of a 5'-AAATT-3' sequence as a side-by-side antiparallel dimer (Figures 1.14, 1.15).⁵²⁻⁵³ Molecular modeling suggests that the minor groove must expand significantly relative to 1:1 complexes in order to accommodate both ligands. As in 1:1 distamycin-DNA complexes the dimeric polyamides fill the minor groove and maintain favorable van der Waals contacts with the walls of the groove. The stacked distamycin ligands are staggered with the carboxamides of one ligand overlapping the pyrrole rings of the second ligand in the 2:1 complex. In this 2:1 complex the positively charged amidinium groups are oriented towards the 3' ends of the binding site. The recent x-ray crystal structure of a 2:1 distamycin-5'-ICICICIC-3' complex confirms these results and emphasizes the importance of polyamide-polyamide stacking interactions in stabilizing 2:1 complexes.⁵⁴ Detailed thermodynamic analysis of distamycin binding to a 5'-AAATT-3' site reveals that the first ligand binds with high affinity ($K_b = 3.1 \times 10^7$) in an enthalpically driven process ($\Delta H_b = -12.3$ kcal/mol) while the second ligand binds with lower overall affinity ($K_b = 3.3 \times 10^6$) but with a larger enthalpy of binding ($\Delta H_b = -18.8$ kcal/mol).⁵⁵ Rentzeperis and co-workers suggest that this 6.5 kcal/mol increase in the enthalpy of binding for the second distamycin ligand in the 2:1 complex is primarily a result of the increase in van der Waals interactions from the side-by-side stacking of the ligands in the minor groove.⁵⁵ Other experiments have shown that the cooperativity of distamycin binding is highly dependent on DNA sequence and may depend on both sequence-dependent minor groove width and flexibility.⁵⁶

Recognition of Mixed Sequences by 2:1 Antiparallel Polyamide Dimers. The 2:1 model derived from the studies above adequately explains the specificity of Pyr-Py-Py-Dp and Im-Py-Py-Dp for 5'-(A,T)G(A,T)C(A,T)-3' sequences.⁵⁰⁻⁵¹ The 2-amino group

Figure 1.14. High resolution structure the 2:1 complex between distamycin and a 5'-AAATT-3' site derived from NMR data.⁵² The DNA is represented as a line model (white) with an overlaid van der Waals surface (magenta). The *N*-methylpyrrolicarboxamides and the positively charged amidinium group are colored cyan and green, respectively.



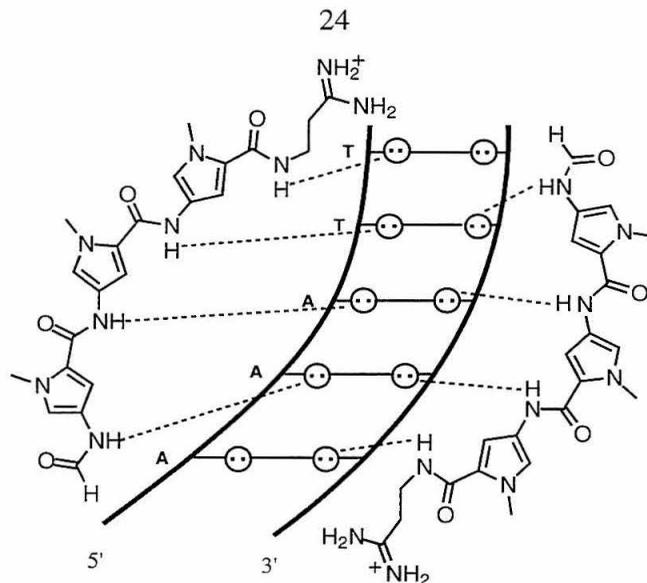


Figure 1.15. Binding model of the 2:1 complex between distamycin and a 5'-AAATT-3' site derived from NMR data. Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines and circles containing a H represent the 2-amino group of guanine. Putative hydrogen bonds are indicated by dashed lines.

of each guanine in the binding site could participate in specific hydrogen bonds with pyridine or imidazole nitrogen atoms on each of the polyamides in the 2:1 complex (Figure 1.16). The carboxamide NH's are predicted to participate in hydrogen bonds with N3 of adenine and O2 of thymine and cytosine. NMR experiments have confirmed these predictions and reveal that Pyr-Py-Py-Dp and Im-Py-Py-Dp bind a 5'-TGACT-3' site in the minor groove as side-by-side antiparallel dimers (Figure 1.17).⁵⁷⁻⁵⁸ As in 1:1 and other 2:1 polyamide-DNA complexes, dimeric binding of Pyr-Py-Py-Dp and Im-Py-Py-Dp to 5'-TGACT-3' is likely stabilized by a combination of specific hydrogen bonding, electrostatic, and van der Waals interactions between the ligands and the DNA and stacking of the ligands in the minor groove. NMR titration experiments demonstrate that binding of the second ligand is highly cooperative with only resonances for the 2:1 complex observed even at low ligand concentrations. This result is consistent with the observation

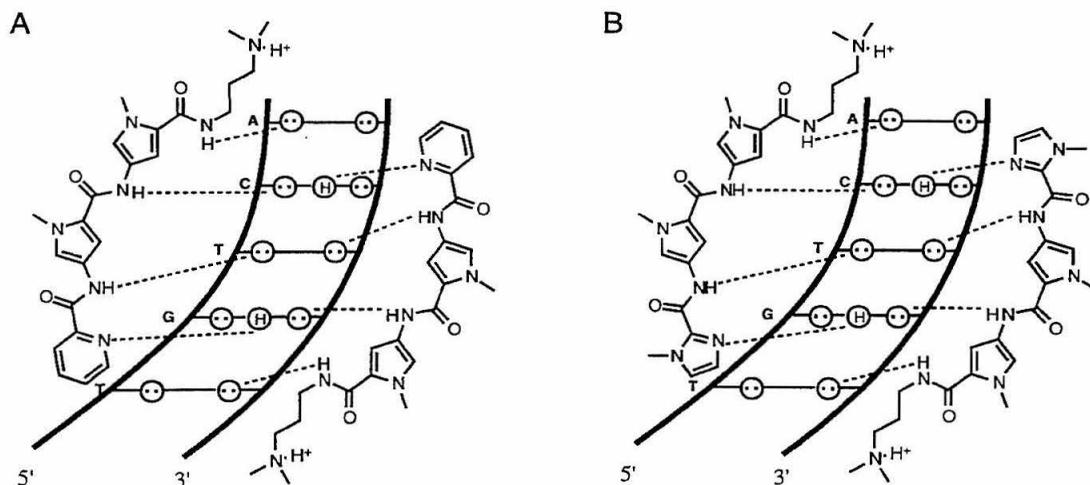
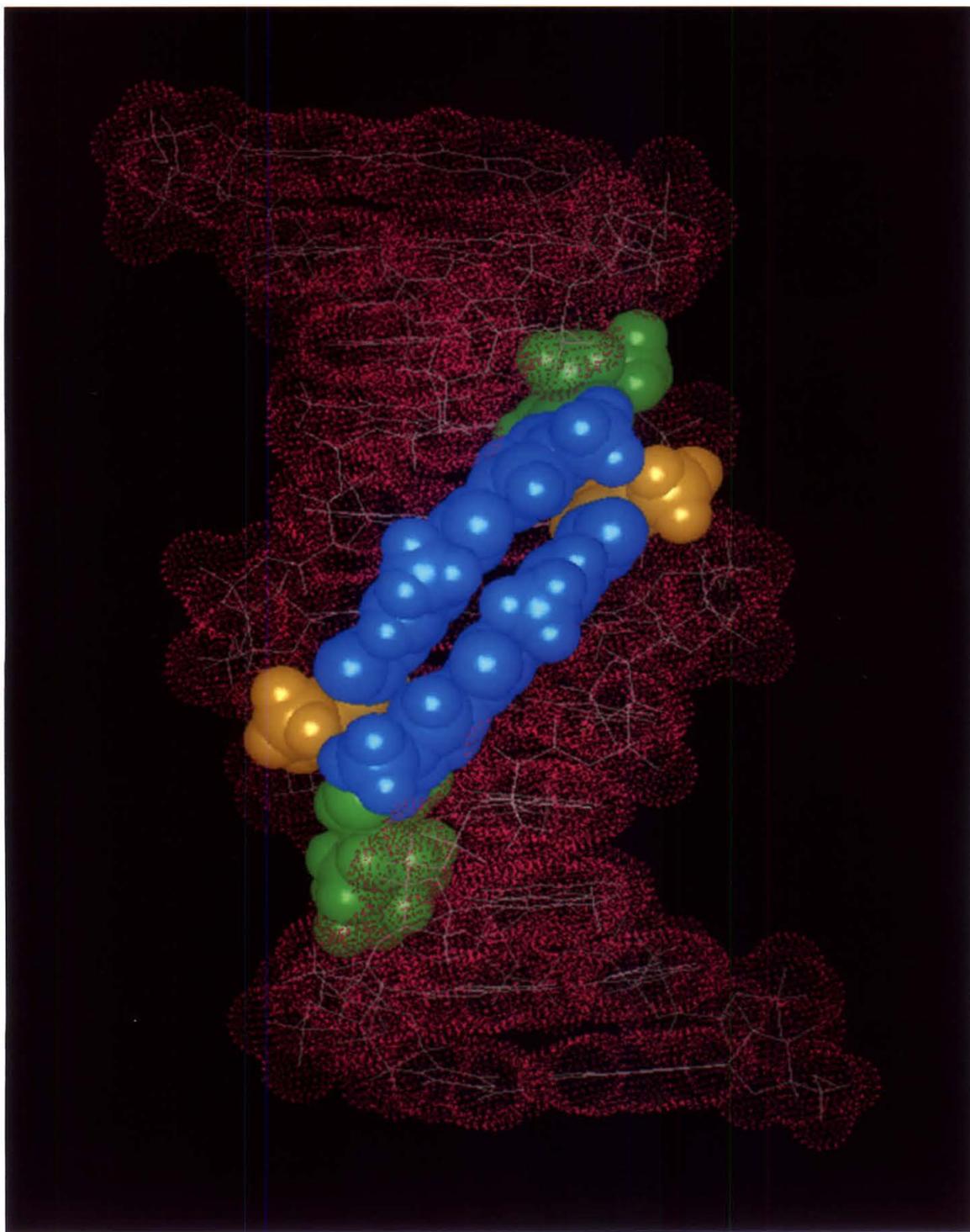


Figure 1.16. 2:1 models for (A) Pyr-Py-Py-Dp and (B) Im-Py-Py-Dp bound at a 5'-TGTC A-3' site. Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines and circles containing a H represent the 2-amino group of guanine. Putative hydrogen bonds are indicated by dashed lines.

that distamycin binds exclusively as a dimer to a 5'-IIICC-3' site. For mixed sequences which generally have inherently wider minor grooves than A,T-rich sequences expansion of the minor groove to accommodate binding of polyamide dimers may be more energetically favorable than clamping down on one ligand in a 1:1 complex.^{57,59}

Antiparallel Side-by-Side Heterodimer. The most significant difference between 2:1 and 1:1 polyamide-DNA complexes is that each ligand in the dimeric motif interacts with only one of the DNA strands in the minor groove. The side-by-side combination of one imidazole ring on one ligand and a pyrrolicarboxamide on the second ligand is specific for G•C, while a pyrrolicarboxamide/imidazole pair targets a C•G base pair. A pyrrolicarboxamide/ pyrrolicarboxamide pair is partially degenerate and binds A•T or T•A base pairs. Based on this revised 2:1 model a polyamide heterodimer consisting of Im-Py-Py-Dp and distamycin should be designed to specifically bind 5'-(A,T)G(A,T)₃-3' sequences (Figure 1.18).⁶⁰ Footprinting and affinity cleavage experiments demonstrate that

Figure 1.17. High resolution structure of the (Im-Py-Py-Dp)₂•5'-TGTCA-3' complex derived from NMR data.⁵⁷ The DNA is represented as a line model (white) with an overlaid van der Waals surface (magenta). The imidazole rings, *N*-methylpyrrolicarboxamides, and the positively charged dimethylaminopropyl groups are colored yellow, cyan, and green, respectively.



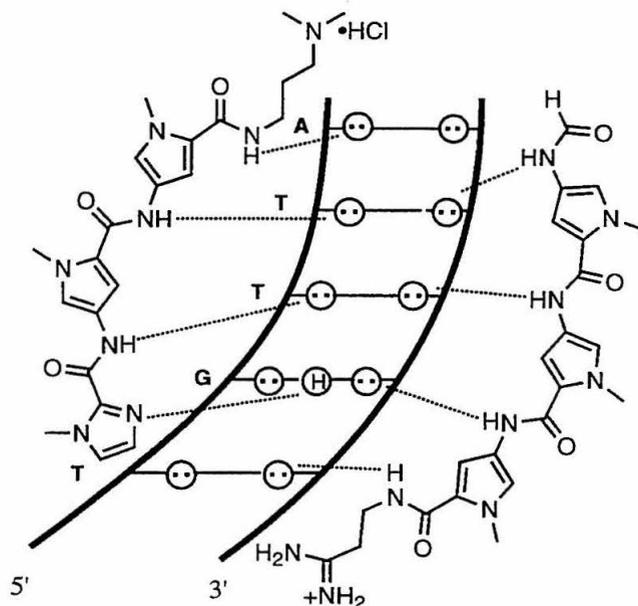


Figure 1.18. Model of the (Im-Py-Py-Dp/distamycin)•5'-TGTTA-3' complex. Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines and circles containing a H represent the 2-amino group of guanine. Putative hydrogen bonds are indicated by dashed lines.

both ligands simultaneously bind a five base pair 5'-TGTTA-3' sequence and reveal that the ligands are oriented as predicted by the 2:1 model for polyamide-DNA recognition. Structural characterization of this heterodimeric complex by two-dimensional NMR confirms these results and as observed for the (Im-Py-Py-Dp)₂•5'-TGACT-3' complex no 1:1 binding is observed.⁶¹ In a related study, Lown, Wemmer, and coworkers reported that a Py-Im-Py polyamide in combination with distamycin recognizes a 5'-AAGTT-3' site in the minor groove as an antiparallel heterodimer.⁶² These studies of homo- and heterodimeric polyamide-DNA complexes suggest that the 2:1 motif will be useful in the design of polyamides for specific recognition of a variety of other mixed sequences in the minor groove.

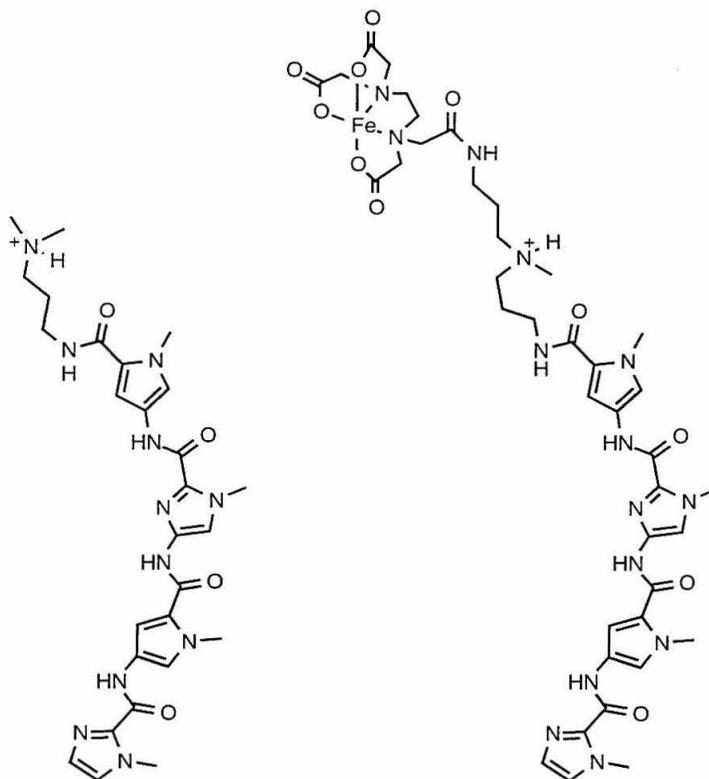


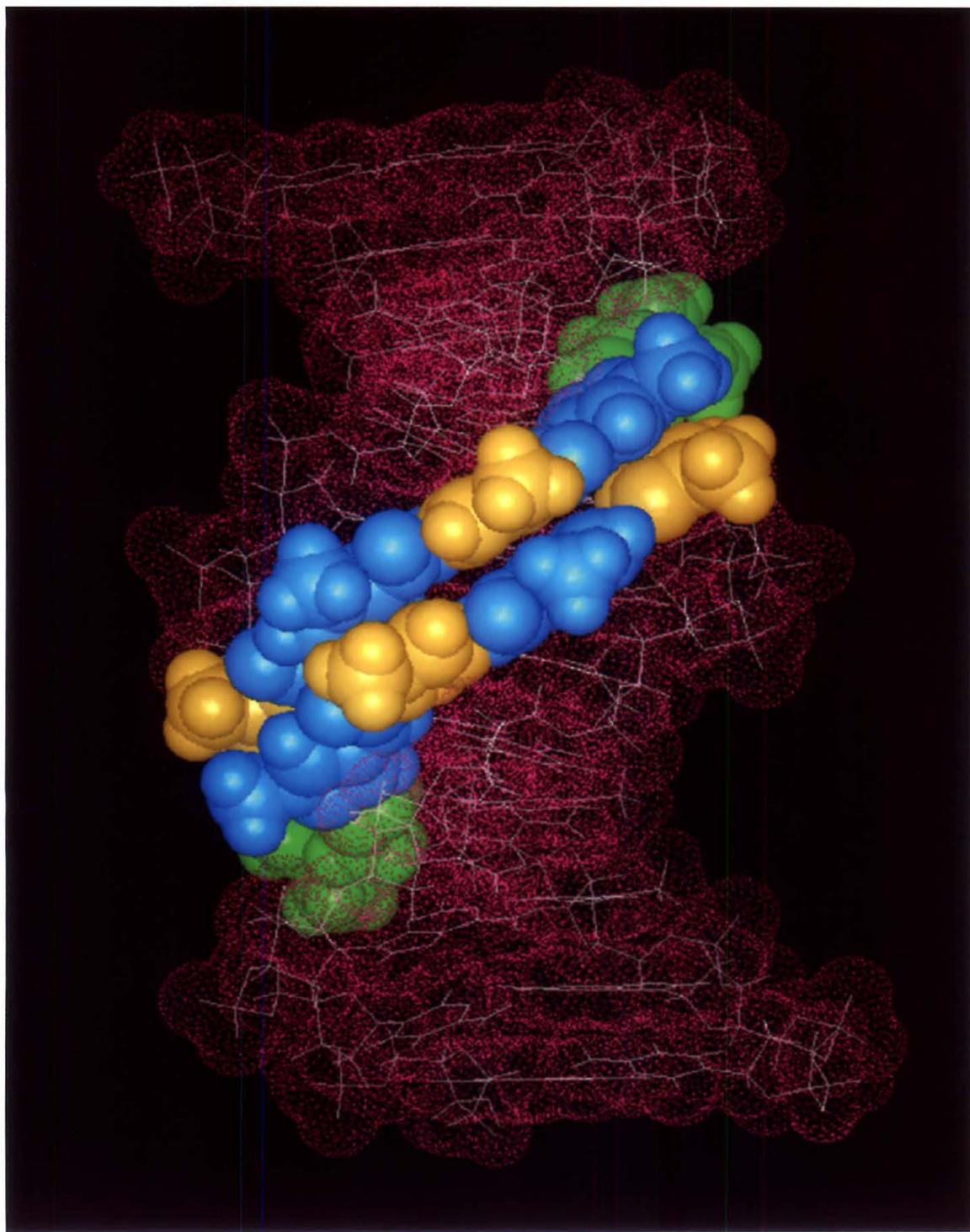
Figure 1.19. Structure of Im-Py-Im-Py-Dp and the affinity cleaving analog Im-Py-Im-Py-EDTA•Fe(II).

Recognition of 5'-GCGC-3' Sequences. Based on 2:1 polyamide-DNA models, the alternating four ring polyamide Im-Py-Im-Py-Dp was designed to bind six base pair 5'-(A,T)GCGC(A,T)-3' sequences as an antiparallel homodimer in the minor groove (Figure 1.19).⁶³ Each imidazole N3 in the ligands was designed to form a hydrogen bond to one guanine amino group on the floor of the minor groove. MPE•Fe(II) footprinting and affinity cleaving experiments reveal that Im-Py-Im-Py-Dp specifically binds 5'-TGCGCA-3', 5'-AGCGCT-3', and 5'-AACGCA-3' sites consistent with 2:1 binding.^{63b} Quantitative DNase I footprint titration experiments demonstrate that the alternating polyamide binds match sites 5'-TGCGCA-3' and 5'-AGCGCT-3' with equal affinity ($K_a = 4.0 \times 10^5$) suggesting that Im-Py-Im-Py-Dp does not discriminate between A•T and T•A

base pairs in the first and sixth position of the binding site. Binding of Im-Py-Im-Py-Dp to the mismatch 5'-AACGCA-3' site occurs with three-fold lower affinity. The observation that 5'-TGGCGT-3' and 5'-TCGGCA-3' sites are not bound demonstrates that the four ring polyamide discriminates between G•C and C•G base pairs in the center of the binding site. NMR characterization of an (Im-Py-Im-Py-Dp)₂•5'-TGCGCA-3' complex indicates that the polyamide binds in the minor groove as a side-by-side antiparallel dimer with positive cooperativity (Figure 1.20).^{63a} Modeling of the complex with constraints from the NMR data affords a model in which each imidazole of the dimeric polyamides forms a hydrogen bond to a unique guanine amino group (Figure 1.21). The binding of Im-Py-Im-Py-Dp to a 5'-GCGC-3' core sequence represents a complete reversal of the specificity of distamycin and netropsin and demonstrates the versatility of the 2:1 polyamide-DNA motif for recognition of mixed sequences. Expansion of the 2:1 motif to other mixed sequences continues to be an area of ongoing research.

Covalently Linked Polyamides. Polyamides that bind as side-by-side dimers in the minor groove can be covalently tethered to afford a new class of polyamides that bind with higher affinity and sequence specificity. The free energy of binding for linked polyamides is expected to be more favorable entropically than binding of two side-by-side polyamides. Linking the polyamides in a heterodimeric complex should disfavor binding at homodimeric sites and hence increase sequence specificity. Connecting the nitrogen atoms of the central pyrrole rings in the 2:1 Pyr-Py-Py-Dp and Im-Py-Py-Dp/distamycin complexes with butyl linkers affords the linked polyamides (Pyr-Py-Py-Dp)₂-C₄ and Im-Py-Py-Dp-C₄-Pyr-Py-Py-Dp, respectively (Figure 1.22).⁶⁴⁻⁶⁵ Quantitative DNase I footprint titration experiments reveal that (Pyr-Py-Py-Dp)₂-C₄ binds a 5'-TGTCA-3' site with a binding affinity at least 10-fold greater than Pyr-Py-Py-Dp.⁶⁴ The ratio of binding affinities

Figure 1.20. High resolution structure of the (Im-Py-Im-Py-Dp)₂•5'-TGCGCA-3' complex derived from NMR data.^{63a} The DNA is represented as a line model (white) with an overlaid van der Waals surface (magenta). The imidazole rings, *N*-methylpyrrole-carboxamides, and the positively charged dimethylaminopropyl groups are colored yellow, cyan, and green, respectively.



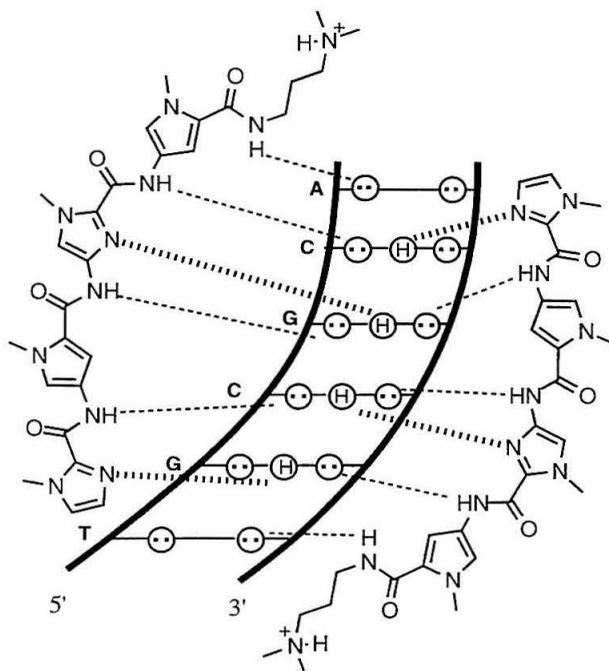


Figure 1.21. Model of the $(\text{Im-Py-Im-Py-Dp})_2 \cdot 5'\text{-TGCGCA-}3'$ complex derived from NMR data. Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines and circles containing a H represent the 2-amino group of guanine. Putative hydrogen bonds are indicated by dashed lines.

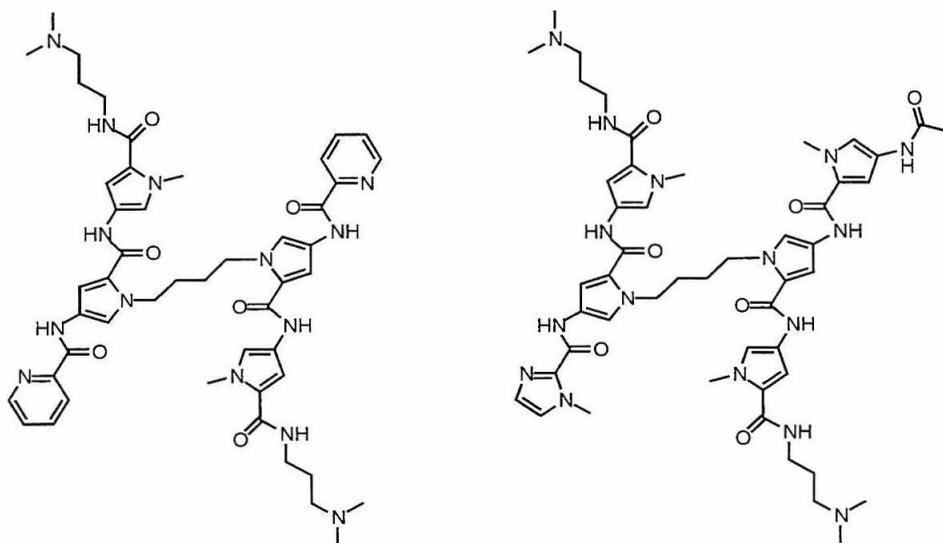


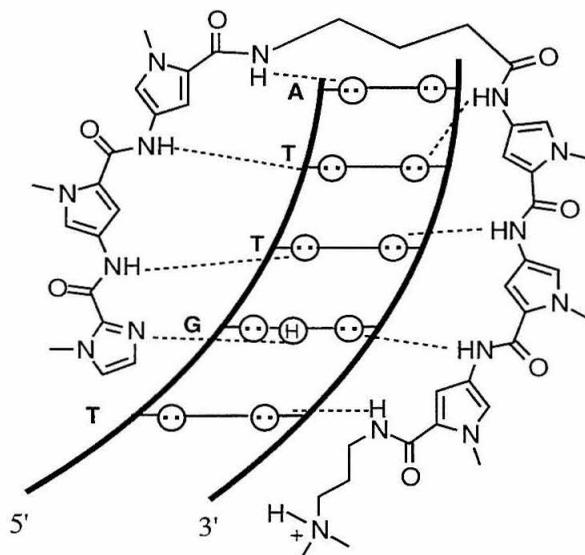
Figure 1.22. Structure of the covalently linked polyamides $(\text{Pyr-Py-Py-Dp})_2\text{-C}_4$ and $\text{Im-Py-Py-Dp-C}_4\text{-Py-Py-Py-Dp}$.

for a 5'-TGTCA-3' to a 5'-TTTTT-3' is 25:1 for (Pyr-Py-Py-Dp)₂-C₄ compared to 1:1 for unlinked Pyr-Py-Py-Dp. The polyamide heterodimer Im-Py-Py-Dp-C₄-Py-Py-Py-Dp also binds the match site 5'-TGTTA-3' with higher affinity than Im-Py-Py-Dp or Py-Py-Py-Dp.⁶⁵ Sequence specificity of the linked heterodimer is also improved as evidenced by the lower affinity for the Im-Py-Py-Dp and Py-Py-Py-Dp binding sites 5'-TGTCA-3' and 5'-TTTTT-3', respectively. Lown and coworkers have subsequently reported that a polyamide composed of two Py-Py-Py units linked through the central pyrroles with a C₇ alkyl linker binds A,T-rich DNA with enhanced affinity.⁶⁶ While these examples are encouraging, the formidable syntheses of centrally linked polyamides may limit the utility of this dimeric motif.

Hairpin and Cyclic Polyamides. More recently, polyamides in a 2:1 complex have been head-to-tail linked with simple amino acids that function as a turn in the minor groove without disturbing the geometry of the polyamide-DNA complex to afford hairpin polyamides that bind with enhanced affinity and specificity.⁶⁷ The hairpin polyamide Im-Py-Py- γ -Py-Py-Py-Dp in which the C-terminus of Im-Py-Py and the N-terminus of Py-Py-Py-Dp are linked with γ -aminobutyric acid (γ) binds a match 5'-TGTTA-3' site with high affinity ($K_a = 8 \times 10^7 \text{ M}^{-1}$) and binds a single base pair mismatch 5'-TGACA-3' site with 24-fold lower affinity ($K_a = 3 \times 10^6 \text{ M}^{-1}$) (Figure 1.23a). Hairpin-linking Im-Py-Py-Dp and Py-Py-Py-Dp increases their binding affinity by approximately 2 orders of magnitude relative to the unlinked 2:1 complex. Coupling of the N- and C-termini of Im-Py-Py- γ -Py-Py-Py-Dp with an additional GABA linker affords the corresponding cyclic polyamide that binds the match 5'-TGTTA-3' site with even higher affinity ($K_a = 3 \times 10^9 \text{ M}^{-1}$) (Figure 1.23b).⁶⁸ Comparing the binding affinities for the single base pair mismatch 5'-TGACA-3' site reveals that the cycle is considerably less specific than the hairpin

polyamide. Use of the hairpin and cyclic motifs for the design of polyamides that bind other sequences with

A



B

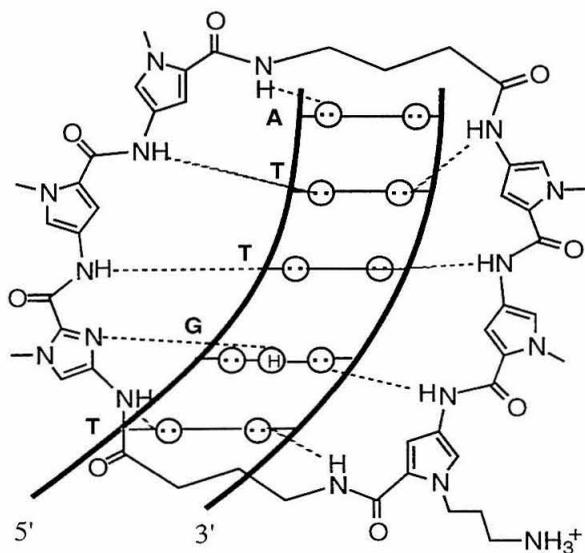


Figure 1.23. Models of (A) the hairpin polyamide Im-Py-Py- γ -Py-Py-Py-Dp and (B) *cyclo*-(Im-Py-Py- γ -Py-Py-Py) binding to a 5'-TGTTA-3' site. Circles with dots represent

lone pairs of N3 of purines and O2 of pyrimidines and circles containing a H represent the 2-amino group of guanine. Putative hydrogen bonds are indicated by dashed lines.

high affinity and specificity is a major goal in the area of polyamide-DNA recognition.

Description of this Work. In Chapter 2, experiments are described in which the sequential addition of five *N*-methylpyrrolicarboxamides to the imidazole-pyrrole polyamide Im-Py-Py-Dp affords a series of six homologous polyamides, Im-(Py)₂₋₇-Dp, that differ in the size of their binding site, apparent first-order binding affinity, and sequence specificity. These results demonstrate that DNA sequences up to nine base pairs in length can be specifically recognized by imidazole-pyrrole containing polyamides containing three to seven rings by 2:1 polyamide-DNA complex formation in the minor groove. Recognition of a nine base pair site defines the new lower limit of the binding site size that can be recognized by polyamides containing exclusively imidazole and pyrrolicarboxamides. The results of this study should provide useful guidelines for the design of new polyamides for that bind longer DNA sites with enhanced affinity and specificity.

In Chapter 3 the design and synthesis of the hairpin polyamide Im-Py-Im-Py- γ -Im-Py-Im-Py-Dp is described. Quantitative DNase I footprint titration experiments reveal that Im-Py-Im-Py- γ -Im-Py-Im-Py-Dp binds six base pair 5'-(A,T)GCGC(A,T)-3' sequences with 30-fold higher affinity than the unlinked polyamide Im-Py-Im-Py-Dp. The hairpin polyamide does not discriminate between A•T and T•A at the first and sixth positions of the binding site as three sites 5'-TGCGCT-3', 5'-TGCGCA-3', and 5'-AGCGCT-3' are bound with similar affinity. However, Im-Py-Im-Py- γ -Im-Py-Im-Py-Dp is specific for and discriminates between G•C and C•G base pairs in the 5'-GCGC-3'

core as evidenced by lower affinities for the mismatched sites 5'-AACGCA-3', 5'-TGCGTT-3', 5'-TGCGGT-3', and 5'-ACCGCT-3' on the same DNA fragment.

In Chapter 4, experiments are described in which a kinetically stable hexa-aza Schiff base La^{3+} complex is covalently attached to a Tat(49-72) peptide which has been shown to bind the HIV-1 TAR RNA sequence. Although these metallo-peptides cleave TAR site-specifically in the hexanucleotide loop to afford products consistent with hydrolysis, a series of control experiments suggests that the observed cleavage is not caused by a sequence-specifically bound Tat(49-72)- $\text{La}(\text{L})^{3+}$ peptide.

References and Notes

1. Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 737.
2. For reviews of DNA structure, see (a) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag; New York, 1984. (b) Kennard, O.; Hunter, W. N. *Q. Rev. Biophys.* **1989**, *22*, 327.
3. For recent reviews of chemical approaches to DNA recognition, see (a) Neilson, P. E. *Bioconj. Chem.* **1991**, *2*, 1. (b) Bailly, C.; Henichart, J. P. *Bioconj. Chem.* **1991**, *2*, 379.
4. For examples of protein-DNA recognition, see (a) Harrison, S. C.; Aggarwal, A. W. *Q. Rev. Biophys.* **1990**, *59*, 933. (b) Steitz, T. A. *Q. Rev. Biophys.* **1990**, *23*, 205. (c) Pabo, C. O.; Sauer, R. T. *Annu. Rev. Biochem.* **1992**, *61*, 1053. (d) Feng, J. A.; Johnson, R. C.; Dickerson, R. E. *Science* **1994**, *263*, 348. (e) Kodadek, T. *Chem. and Biol.* **1995**, *2*, 267.
5. (a) Pavletich, N. L.; Pabo, C. O. *Science* **1991**, *252*, 809. (b) Desjarlais, J. R.; Berg, J. M. *Proteins* **1992**, *12*, 101. (c) Desjarlais, J. R.; Berg, J. M. *Proc. Natl. Acad.*

- Sci. USA* **1992**, *89*, 7345. (d) Desjarlais, J. R.; Berg, J. M. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 2256. (e) Rebar, E.J.; Pabo, C.O. *Science* **1994**, *263*, 671. (f) Shi, Y. G.; Berg, J. M. *Chem. and Biol.* **1995**, *2*, 83.
6. Moser, H. E.; Dervan, P. B. *Science* **1987**, *238*, 645.
 7. For a review of oligonucleotide-directed triple helix formation, see Helene, C. *Angew. Chem. Int. Ed. Engl.* **1993**, *32*, 666.
 8. (a) Strobel, S. A.; Moser, H. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1988**, *110*, 7927. (b) Strobel, S. A.; Dervan, P. B. *Science* **1990**, *249*, 73. (c) Strobel, S. A.; Doucettstamm, L. A.; Riba, L.; Housman, D. E.; Dervan, P. B. *Science* **1991**, *254*, 1639. (d) Strobel, S. A.; Dervan, P. B. *Nature* **1991**, *350*, 172. (e) Povsic, T. J.; Strobel, S. A.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 5934.
 9. Singleton, S. F.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 6957.
 10. (a) Lee, J. S.; Woodsworth, M. L.; Latimer, L. J. P.; Morgan, A. R. *Nucleic Acids Res.* **1984**, *12*, 6603. (b) Povsic, T. J.; Dervan, P. B. *J. Am. Chem. Soc.* **1989**, *111*, 3059. (c) Povsic, T. J. Ph. D. Dissertation, California Institute of Technology, Pasadena, CA, 1991. (d) Collier, D. A.; Thuong, N. T.; Helène, C. *J. Am. Chem. Soc.* **1991**, *113*, 1457. (e) Froehler, B. C.; Wadwani, S.; Terhorst, T. J.; Gerrard, S. R. *Tetrahedron Lett.* **1992**, *33*, 5307. (f) Sagi, J.; Szemzo, A.; Ebinger, K.; Szabolcs, A.; Sagi, G.; Ruff, E.; Otvos, L. *Tetrahedron Lett.* **1993**, *34*, 2191.
 11. Best, G. C.; Dervan, P. B. *J. Am. Chem. Soc.* **1995**, *117*, 1187.
 12. Singleton, S. F.; Dervan, P. B. *Biochemistry* **1992**, *31*, 10995.
 13. Singleton, S. F.; Dervan, P. B. *Biochemistry* **1993**, *32*, 13171.
 14. Singleton, S. F.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 10376.

15. (a) Roberts, R. W.; Crothers, D. M. *Science* **1992**, *258*, 1463. (b) Han, H.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 3806. (c) Han, H.; Dervan, P. B. *Nucleic Acids Res.* **1994**, *22*, 2837.
16. (a) Distefano, M. D.; Shin, J. A.; Dervan, P. B. *J. Am. Chem. Soc.* **1991**, *113*, 5901. (b) Distefano, M. D.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 11006. (c) Distefano, M. D.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 1179. (d) Colocci, N.; Distefano, M. D.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 4468. (e) Colocci, N.; Dervan, P. B. *J. Am. Chem. Soc.* **1995**, *117*, 4781. (f) Colocci, N.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 785. (g) Colocci, N.; Dervan, P. B. *Nucleic Acids Res.* Manuscript submitted.
17. Best, G. C.; Dervan, P. B. Manuscript in preparation.
18. (a) Cooney, M.; Czernuszewics, G.; Postel, E. H.; Flint, S. J.; Hogan, M. E. *Science* **1988**, *241*, 456. (b) Beal, P. A.; Dervan, P. B. *Science* **1991**, *251*, 1360. (c) Durland, R. H.; Kessler, D. J.; Gunnell, S.; Duvic, M.; Pettitt, B. M.; Hogan, M. E. *Biochemistry* **1991**, *30*, 9246. (d) Pilch, D. S.; Levenson, C.; Shafer, R. H. *Biochemistry* **1991**, *30*, 6081. (e) Radhakrishnan, P.; de los Santos, C.; Patel, D. J. *J. Mol. Biol.* **1991**, *221*, 1403. (f) Beal, P. A.; Dervan, P. B. *Nucl. Acids Res.* **1992**, *20*, 2773. (g) Greenberg, W. A.; Dervan, P. B. *J. Am. Chem. Soc.* **1995**, *117*, 5016.
19. Hunziker, J.; Priestley, E. S.; Brunar, H.; Dervan, P. B. *J. Am. Chem. Soc.* **1995**, *117*, 2661.
20. (a) Griffin, L. C.; Dervan, P. B. *Science* **1989**, *245*, 967. (b) Kiessling, L. C.; Griffin, L. C.; Dervan, P. B. *Biochemistry* **1992**, 2829. (c) Koh, J. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 1470. (d) Griffin, L. C.; Kiessling, L. C.; Beal, P. B.; Gillespie, P.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 7976. (e) Koshlap,

- K. M.; Gillespie, P.; Dervan, P. B.; Feigon, J. *J. Am. Chem. Soc.* **1993**, *115*, 7908.
- (f) Stiliz, H. U.; Dervan, P. B. *Biochemistry* **1993**, *32*, 2177. (g) Radhakrishnan, I.; Patel, D. J.; Priestley, E. S.; Nash, H. M.; Dervan, P. B. *Biochemistry* **1993**, *32*, 11228. (h) Priestley, E. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *117*, 4761. (i) Staubli, A. B.; Dervan P. B. *Nucleic Acids Res.* **1994**, *22*, 2637.
21. (a) Ono, A.; Ts'o, P. O. P. *J. Am. Chem. Soc.* **1991**, *113*, 4032. (b) Krawczyk, S. H.; Milligan, J. F.; Wadwani, S.; Moulds, C.; Froehler, B. C.; Matteucci, M. D. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 376. (c) Miller, P. S.; Bhan, P.; Cushman, C. D.; Trepane, C. *Biochemistry* **1992**, *31*, 6788. (d) von Krosigk, U.; Benner, S. A. *J. Am. Chem. Soc.* **1995**, *117*, 5361. (e) Gee, J. E.; Revankar, G. R.; Sudhakar Rao, T.; Hogan, M. E. *Biochemistry* **1995**, 2042.
22. (a) Sowers, L. C.; Shaw, B. R.; Sedwick, W. D. *Biochem. Biophys. Res. Commun.* **1987**, *148*, 790. (b) DeVoe, H.; Tinoco, I. *J. Mol. Biol.* **1962**, *4*, 500. (c) Sanyal, N. K.; Roychoudhury, M.; Ruhela, K. R.; Tiwari, S. N. *J. Comput. Chem.* **1986**, *8*, 604. (d) Ornstein, R. L.; Rein, R.; Breen, D. L.; MacElroy, R. D. *Biopolymers* **1978**, *17*, 2341. (e) Broom, A. D.; Schweizer, M. P.; Ts'o, P. O. P. *J. Am. Chem. Soc.* **1967**, *89*, 3612. (f) Ts'o, P. O. P. *Molecular Associations in Biology*; Academic Press: New York, N. Y., 1968, pp 39-75. (g) Nakano, N. I.; Igarishi, S. *J. Biochemistry* **1970**, *9*, 577. (h) Breslauer, K. J.; Frank, R.; Blocker, H.; Marky, L. A. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 3746. (i) Delcourt, S. G.; Blake, R. D. *J. Biol. Chem.* **1991**, *266*, 15160.
23. Rao, B. J.; Dutriex, M.; Radding, C. N. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 2984.
24. Baliga, R.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 10393.
25. For reviews, see (a) Lown, J. W. *Chemtracts-Organic Chemistry* **1993**, *6*, 205. (b) Kahne, D. *Chem. and Biol.* **1995**, *2*, 7. (c) Lown, J. W. *Drug Dev. Res.* **1995**,

- 34, 145. (d) Geierstanger, B. H.; Wemmer, D. E. *Ann. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 463.
26. (a) Grootenhuis, P. D. J.; Roe, D. C.; Kollman, P. A.; Kuntz, I. D. *J. Computer-Aided Mol. Design* **1994**, *8*, 731. (b) Misra, V. K.; Sharp, K. A.; Friedman, R. A.; Honig, B. *J. Mol. Biol.* **1994**, *238*, 245. (c) Misra, V. K.; Honig, B. *Proc. Natl. Acad. Sci. USA* **1995**, *88*, 2984. (d) Czarny, A.; Boykin, D. W.; Wood, A. A.; Nunn, C. M.; Neidle, S.; Zhao, M.; Wilson, W. D. *J. Am. Chem. Soc.* **1995**, *117*, 4716.
27. (a) Arcamone, F.; Bizioli, F.; Canevazzi, G.; Grein, A. German Pat #1,027,667, 1958. (b) Finlay, A.; Hochstein, F.; Sobin, B. Murphy, F. *J. Am. Chem. Soc.* **1951**, *73*, 342.
28. (a) F. E. Hahn in *Antibiotics III. Mechanisms of Action of Antimicrobial and Antitumor Agents*, Gottlieb, P. D.; Shaw, P. D.; Corcoran, J. W. Eds.; Springer: New York, 1975. (b) Zimmer, C. *Progress in Nucleic Acids and Molecular Biology* **1980**, *15*, 258. (c) Krey, A. *Prog. in Mol. Subcell. Biol.* **1980**, *7*, 43.
29. Zimmer, C.; Wähnert, U. *Prog. Biophys. Molec. Biol.* **1986**, *47*, 31.
30. (a) Berman, H. M.; Neidle, S.; Zimmer, C.; Thrum, H. *Biochim. Biophys. Acta* **1979**, *561*, 124. (b) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 1376. (c) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *J. Mol. Biol.* **1985**, *183*, 553. (d) Coll, M.; Frederick, C. A.; Wang, A. H.-J.; Rich, A. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 8385. (e) Coll, M.; Aymami, J.; van der Marel, G. A.; van Boom, J. H.; Rich, A.; Wang, A. H.-J. *Biochemistry* **1989**, *28*, 310. (f) Sriram, M.; van der Marel, G. A.; Roelen, H. L. P. F.; van Boom, J. H.; Wang, A. H.-J. *Biochemistry* **1992**, *21*,

11823. (g) Taberner, L.; Verdaguer, N.; Coll, M.; Fita, I.; van der Marel, G. A.; van Boom, J. H.; Rich, A.; Aymami, J. *Biochemistry* **1993**, *32*, 8403.
31. (a) Patel, D. J. *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 6424. (b) Patel, D. J.; Shapiro, L. *Biochemie* **1985**, *67*, 887. (c) Sarma, M. H.; Gupta, G.; Sarma, R. H. *J. Biomol. Struct. Dyn.* **1985**, *2*, 1085. (d) Patel, D. J.; Shapiro, L. *J. Biol. Chem.* **1986**, *261*, 1230. (e) Klevit, R. E.; Wemmer, D. E.; Reid, B. R. *Biochemistry* **1986**, *25*, 3296. (f) Pelton, J. G.; Wemmer, D. E. *Biochemistry*, **1988**, *27*, 8088.
32. Zakrzewska, K.; Lavery, R.; Pullman, B. *J. Biomol. Struct. Dyn.* **1987**, *4*, 883.
33. (a) Markey, L. A.; Breslauer, K. J. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 4359. (b) Breslauer, K. J.; Remeta, D. P.; Chou, W.-Y.; Ferrante, R.; Curry, J.; Zaunczkowski, D.; Snyder, J. G.; Markey, L. A. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 8922.
34. Dervan, P. B. *Science* **1986**, *232*, 464.
35. (a) Sluka, J. P.; Horvath, S. J.; Bruist, M. F.; Simon, M. I.; Dervan, P. B. *Science* **1987**, *238*, 1129. (b) Mack, D. P.; Iverson, B. L.; Dervan, P. B. *J. Am. Chem. Soc.* **1988**, *110*, 7572. (c) Graham, K. S.; Dervan, P. B. *J. Biol. Chem.* **1990**, *265*, 16534. (d) Mack, D. P.; Dervan, P. B. *J. Am. Chem. Soc.* **1990**, *112*, 4604. (e) Mack, D. P.; Sluka, J. P.; Shin, J. A.; Griffin, J. H.; Simon, M. I.; Dervan, P. B. *Biochemistry* **1990**, *29*, 6561. (f) Oakley, M. G.; Dervan, P. B. *Science* **1990**, *248*, 847. (g) Sluka, J. P.; Horvath, S. J.; Glasgow, A. C.; Simon, M. I.; Dervan, P. B. *Biochemistry* **1990**, *29*, 6551. (h) Shin, J. A.; Ebright, R. H.; Dervan, P. B. *Nucleic Acids Res.* **1991**, *19*, 5233. (i) Mack, D. P.; Dervan, P. B. *Biochemistry* **1992**, *31*, 9399.
36. Dreyer, G. B.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 968.

37. (a) Schultz, P. G.; Taylor, J. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1982**, *104*, 6861. (b) Taylor, J. S.; Schultz, P. G.; Dervan, P. B. *Tetrahedron* **1984**, *40*, 457. (c) Schultz, P. G.; Dervan, P. B. *J. Biomol. Struct. Dyn.* **1984**, *1*, 1133.
38. (a) Fox, K. R.; Waring, M. J.; *Nucleic Acids Res.* **1984**, *12*, 9271. (b) Lane, M. J.; Dabrowiak, J. C.; Vournakis, J. *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 3260.
39. (a) Van Dyke, M. W.; Hertzberg, R. P.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 5470. (b) Van Dyke, M. W.; Dervan, P. B. *Cold Spring Harbor Symposium on Quantitative Biology* **1982**, *47*, 347. (c) Van Dyke, M. W.; Dervan, P. B. *Biochemistry* **1983**, *22*, 2373. (d) Harshman, K. D.; Dervan, P. B. *Nucl. Acids Res.* **1985**, *13*, 4825.
40. (a) Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Methods Enzymol.* **1986**, *130*, 132. (b) Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 8462. (c) Senear, D. F.; Brenowitz, M.; Shea, M. A.; Ackers, G. K. *Biochemistry* **1986**, *25*, 7344.
41. The quantitative footprint titration methodology has been adapted for measurement of binding affinities of polyamide analogs to discrete five base pair sites using MPE•Fe(II), see (a) Wade, W. S.; Mrksich, M.; Dervan, P. B. *Biochemistry* **1993**, *32*, 11385. (b) Wade, W. S. Ph. D. Thesis, California Institute of Technology, 1989.
42. Johnson, R. F.; Pickett, S. C.; Barker, D. L. *Electrophoresis* **1990**, *11*, 355.
43. (a) Shultz, P. G.; Dervan, P. B. *J. Am. Chem. Soc.* **1983**, *105*, 7748. (b) Shultz, P. G.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 6834. (c) Youngquist, R. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1985**, *107*, 5528. (d) Griffin, J. H.; Dervan, P. B. *J. Am. Chem. Soc.* **1986**, *108*, 5008. (e) Wang, A. H.-J.; Cottens, S.; Dervan, P.

- B.; Yesinowski, J. P.; van der Marel, G. A.; van Boom, J. H. *J. Biomol. Struct. Dyn.* **1989**, 7, 101.
44. (a) Youngquist, R. S.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1985**, 82, 2565.
(b) Youngquist, R. S. Ph. D. Thesis, California Institute of Technology, 1988.
45. Youngquist, R. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1987**, 109, 7564.
46. (a) Khorlin, A. A.; Krylov, A. S.; Grokhovsky, S. L.; Zhuzhe, A. L.; Zasadatelev, A. S.; Gursky, G. V.; Gottikh, B. P. *FEBS Lett.* **1980**, 118, 311. (b) Gursky, G. V.; Zasadatelev, A. S.; Zhuzhe, A. L.; Khorlin, A. A.; Grokhovsky, S. L.; Streltsov, S. A.; Surovaya, A. N.; Nikitin, S. M.; Krylov, A. S.; Retchinsky, V. O.; Michailov, M. V.; Beabealashvilli, R. S.; Gottikh, B. P. *Cold Spring Harbor Symp. Quant. Biol.* **1982**, 47, 367. (c) Skamrov, A. V.; Rybalkin, I. N.; Bibilashvili, R. Sh.; Gottikh, B. P.; Grokhovskii, S. L.; Gurskii, G. V.; Zhuze, A. L.; Zasadatelev, A. S.; Nechipurenko, Yu. D.; Khorlin, A. A. *Mol. Biol.* **1985**, 19, 153. (d) Leinsoo, T. A.; Nikolaev, V. A.; Grokhovskii, S. L.; Surovaya, A. N.; Sidorova, N. Y.; Streltsov, S. A.; Zasadatelev, A. S.; Zhuzhe, A. L.; Gurskii, G. V. *Molek. Biol. (USSR)* **1989**, 23, 1616.
47. (a) Lown, J. W.; Krowicki, K.; Balzarini, J.; Newman, R. A.; De Clercq, E. *J. Med. Chem.* **1989**, 32, 2368. (b) Kissinger, K. L.; Dabrowiak, J. C.; Lown, J. W. *Chem. Res. Toxicol.* **1990**, 3, 162. (c) Rao, K. E.; Zimmerman, J.; Lown, J. W. *J. Org. Chem.* **1991**, 56, 786. (d) Rao, K. E.; Krowicki, K.; Balzarini, J.; De Clercq, E.; Newman, R. A.; Lown, J. W. *Actual. Chim. Ther.* **1991**, 18, 21. (e) Wang, W.; Lown, J. W. *J. Med. Chem.* **1992**, 35, 2890. (f) Singh, M. P.; Plouvier, B.; Hill, G. C.; Gueck, J.; Pon, R. T.; Lown, W. J. *J. Am. Chem. Soc.* **1994**, 116, 7006.

48. For other approaches, see (a) Dervan, P. B.; Sluka, J. P. *New Synthetic Methodology and Functionally Interesting Compounds*; Elsevier: New York, 1986; pp 307-322. (b) Griffin, J. H.; Dervan, P. B. *J. Am. Chem. Soc.* **1987**, *109*, 6840.
49. (a) Lown, J. W.; Krowicki, K.; Bhat, U. G.; Ward, B.; Dabrowiak, J. C. *Biochemistry* **1986**, *25*, 7408. (b) Kissinger, K.; Krowicki, K.; Dabrowiak, J. C.; Lown, J. W. *Biochemistry* **1987**, *26*, 5590. (c) Lee, M.; Chang, M. D.; Hartley, J. A.; Pon, R. T.; Krowicki, K.; Lown, J. W. *Biochemistry* **1988**, *27*, 445. (d) Rao, K. E.; Bathini, Y.; Lown, J. W. *J. Org. Chem.* **1990**, *55*, 728. (e) Plouvier, B.; Bailly, C.; Houssin, R.; Rao, K. E.; Lown, J. W.; Henichar, J.-P.; Waring, M. J. *Nucleic Acids Res.* **1991**, *19*, 5821.
50. Wade, W. S., Dervan, P. B. *J. Am. Chem. Soc.* **1987**, *109*, 1574.
51. Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 8783.
52. (a) Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 5723. (b) Pelton, J. G.; Wemmer, D. E. *J. Am. Chem. Soc.* **1990**, *112*, 1393.
53. Characterization of a 2:1 complex between a tetra-*N*-methylpyrrolicarboxamide analog of distamycin and a 5'-AAATT-3' by NMR has recently been reported, see Animati, F.; Arcamone, F. M.; Conte, M. R.; Felicetti, P.; Galeone, A.; Lombardi, P.; Mayol, L. G.; Rossi, C. *J. Med. Chem.* **1995**, *38*, 1140.
54. Chen, X.; Ramakrishnan, B.; Rao, S. T.; Sundaralingam, M. *Struct. Biol. Nat.* **1994**, *1*, 169.
55. Rentzeperis, D.; Marky, L. A.; Dwyer, T. J.; Geierstanger, B. H.; Pelton, J. G.; Wemmer, D. E. *Biochemistry* **1995**, *34*, 2937.
56. (a) Fagan, P. A.; Wemmer, D. E. *J. Am. Chem. Soc.* **1992**, *114*, 1080. (b) Wemmer, D. E.; Geierstanger, B. H.; Fagan, P. A.; Dwyer, T. J.; Jacobson, J. P.; Pelton, J. G.; Ball, G. E.; Leheny, A. R.; W.-H. Chang; Bathini, Y.; Lown, J. W.;

- Rentzeperis, D.; Marky, L. A.; Singh, S.; Kollman, P. *Structural Biology: The State of the Art, Proceedings of the Eighth Conversation in the Discipline Biomolecular Sterodynamics*; Adenine Press: New York, 1994, pp. 301.
57. (a) Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 7586. (b) Wade, W. S.; Mrksich, M.; Dervan, P. B. *Biochemistry* **1993**, *32*, 11385.
58. Lown, Wemmer, and coworkers have analyzed the 2:1 binding of a pyrrole-imidazole-pyrrole analog of distamycin to a 5'-AAGTT-3' by NMR. We note that complex in which two imidzoles on opposite polyamides targets a G•C base pair is formally a mismatch, see Dwyer, T. J.; Geierstanger, B. H.; Bathini, Y.; Lown, W. J.; Wemmer, D. E. *J. Am. Chem. Soc.* **1992**, *114*, 5912.
59. Yoon, C.; Prive, G. G.; Goodsell, D. S.; Dickerson, R. E. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 6332.
60. Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 2572.
61. Geierstanger, B. H.; Jacobsen, J.-P.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Biochemistry* **1994**, *33*, 3055.
62. Geierstanger, B. H.; Dwyer, T. J.; Bathini, Y.; Lown, J. W.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, *115*, 4474.
63. (a) Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Science* **1994**, *266*, 646. (b) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1995**, *117*, 3325.
64. (a) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 9892. (b) Dwyer, T. J.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, *115*, 9900.

65. Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 3663.
66. Chen, Y.-H.; Lown, J. W. *J. Am. Chem. Soc.* **1994**, *116*, 6995.
67. For hairpin polyamides, see Mrksich, M.; Parks, M. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 7983.
68. For cyclic polyamides, see Cho, J.; Parks, M. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 10389.

Chapter 2

Binding Site Size Limitations of Imidazole-Pyrrole Polyamides for Recognition in the Minor Groove of DNA

Introduction

1:1 Polyamide-DNA Motif. The natural products netropsin and distamycin are crescent shaped di- and tripyrrolicarboxamides, respectively, that bind in the minor groove of DNA at sites of four or five successive A,T base pairs.¹⁻³ The structures of a number of polyamide-DNA complexes have been determined by X-ray diffraction⁴ and NMR spectroscopy⁵, and the thermodynamic profiles have been studied for these complexes.⁶ Analyses of these 1:1 complexes suggests that favorable electrostatic interactions and extensive van der Waals contacts between the polyamide and the floor and walls of the minor groove contribute to complex stability. Although this model has aided in the design of polyamides for recognition of longer tracts of A,T-rich sites,⁷ efforts based on the 1:1 polyamide-DNA motif to design ligands capable of binding sequences containing A,T and G,C base pairs have met with limited success.⁸⁻⁹

2:1 Polyamide-DNA Motif. Recent examples of 2:1 polyamide-DNA complexes have created new models for the design of nonnatural ligands for sequence-specific recognition of mixed sequences in the minor groove of DNA.¹⁰⁻¹⁵ The imidazole (Im)-pyrrole (Py) polyamide Im-(Py)₂-Dp was shown to specifically bind five base pair 5'-(A,T)G(A,T)C(A,T)-3' sequences¹² while the alternating polyamide Im-Py-Im-Py-Dp was shown to bind six base pair 5'-(A,T)GCGC(A,T)-3' sites¹⁴ as side-by-side antiparallel dimers in the minor groove (Figure 2.1). In addition to the sequence-dependent groove width being a determinant of specificity, the 2:1 model allows specific contacts with

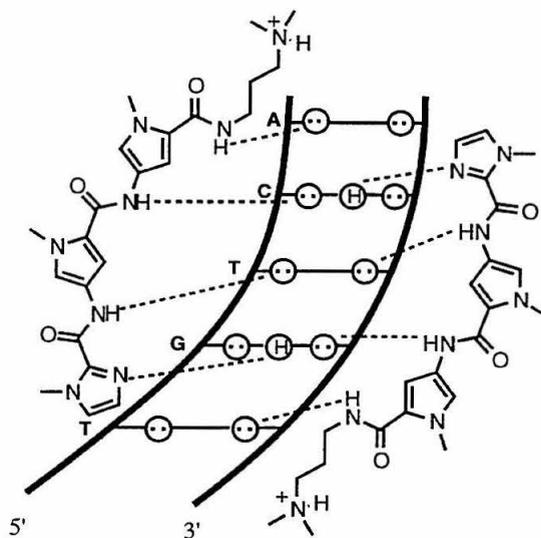
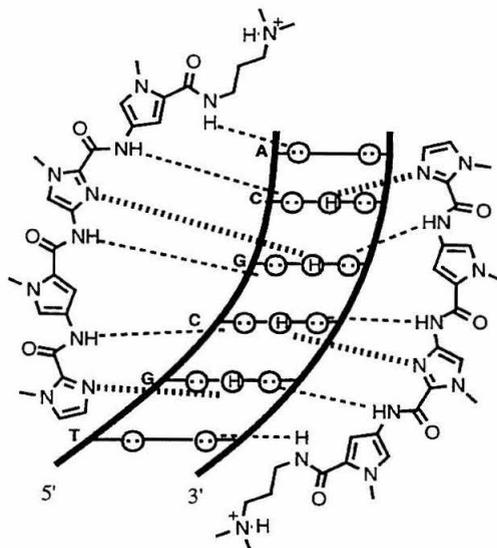
A**(Im-(Py)₂-Dp)₂ • TGTCA****B****(Im-Py-Im-Py-Dp)₂ • TGCGCA**

Figure 2.1. Binding models for the 2:1 complexes formed between (A) Im-(Py)₂-Dp with a 5'-TGTCA-3' sequence and (B) Im-Py-Im-Py-Dp with a 5'-TGCGCA-3' sequence. Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines and circles containing a H represent the 2-amino group of guanine. Putative hydrogen bonds are indicated by dashed lines.

each strand on the floor of the minor groove. The side-by-side combination of one imidazole ring on one ligand and a pyrrolicarboxamide on the second ligand is specific for G•C, while a pyrrolicarboxamide/imidazole pair targets a C•G base pair.¹²⁻¹⁵ A pyrrolicarboxamide/ pyrrolicarboxamide pair is partially degenerate binds A•T or T•A base pairs.¹⁰⁻¹⁵

A major goal of our efforts in evaluating the scope and limitations of the 2:1 polyamide-DNA motif is to extend specific recognition to larger binding sites. In the examples described above the *three* ring polyamide Im-(Py)₂-Dp binds *five* base pair sites while the *four* ring polyamide Im-Py-Im-Py-Dp is specific for *six* base pair sites suggesting that incorporation of an additional ring in the ligand lengthens the binding site size by one base pair. In order to determine the effect of polyamide length on binding site size, binding affinity, and sequence specificity in the 2:1 polyamide-DNA motif a series of imidazole-pyrrole polyamides containing three to eight rings was synthesized

Experimental Design. The polyamide series is based on Im-(Py)₂-Dp (**1**) with pyrrolicarboxamide moieties added sequentially to the C-termini to afford Im-(Py)₃-Dp (**2**), Im-(Py)₄-Dp (**3**), Im-(Py)₅-Dp (**4**), Im-(Py)₆-Dp (**5**) and Im-(Py)₇-Dp (**6**) which are designed to bind, five six, seven, eight, nine and ten base pair sites, respectively, as side-by-side antiparallel dimers (Figures 2.2, 2.3). The DNA binding sites are based on 5'-TGACA-3' and contain a sequentially longer A,T tract in the center of the binding sites that will be recognized by the additional pyrrolicarboxamides. This combination of polyamides and DNA binding site sequences was chosen to satisfy several criteria. The imidazole rings at the N-termini of the polyamides will provide specificity in 2:1 polyamide-DNA complex formation by recognizing, in combination with a pyrrolicarboxamide from the other polyamide, the G•C and C•G base pairs in the binding sites. In addition, the presence of G•C and C•G base pairs in the binding sites are expected

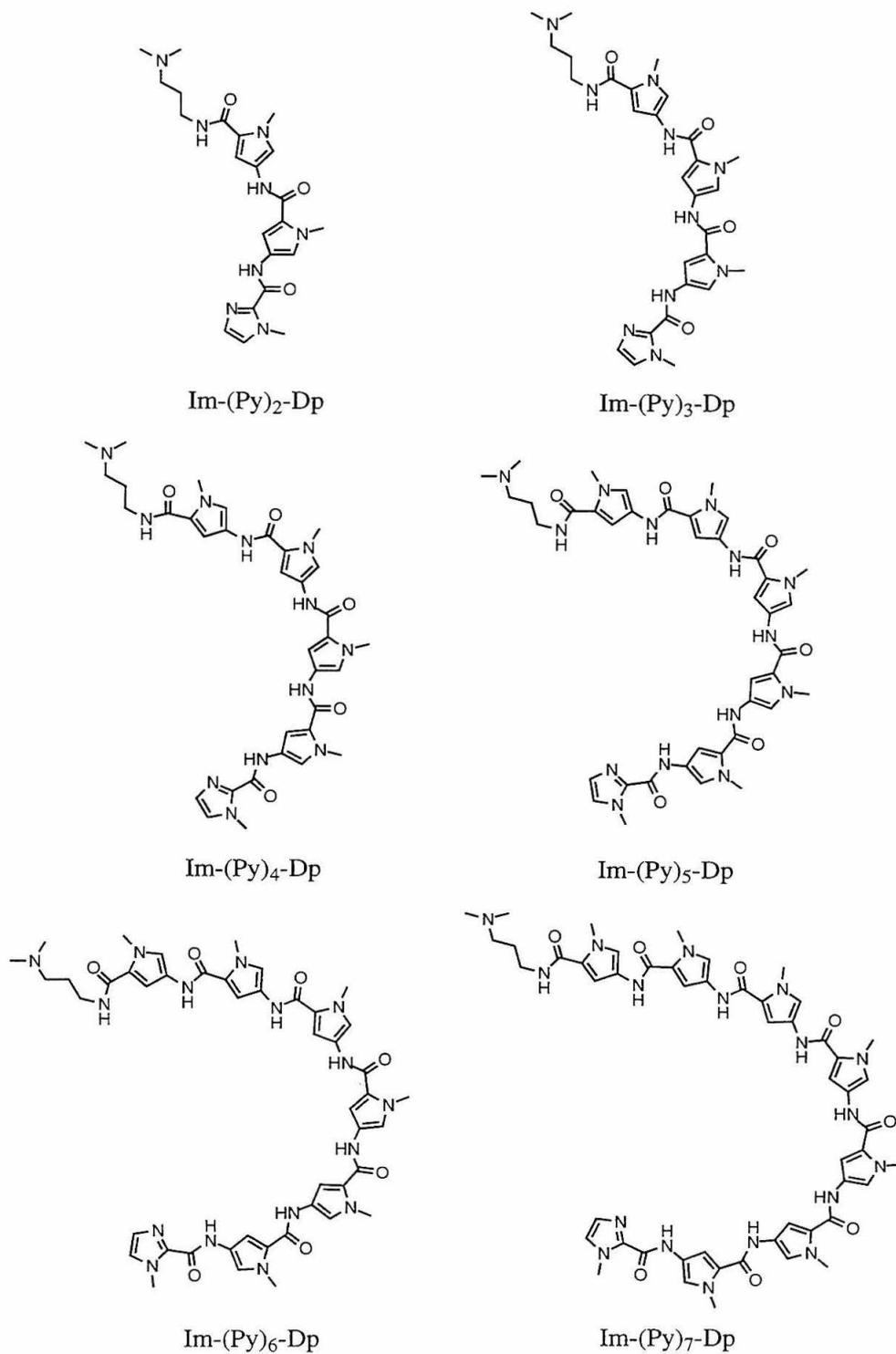


Figure 2.2. Structure of the imidazole-pyrrole polyamides Im-(Py)₂-Dp, Im-(Py)₃-Dp, Im-(Py)₄-Dp, Im-(Py)₅-Dp, Im-(Py)₆-Dp, and Im-(Py)₇-Dp.

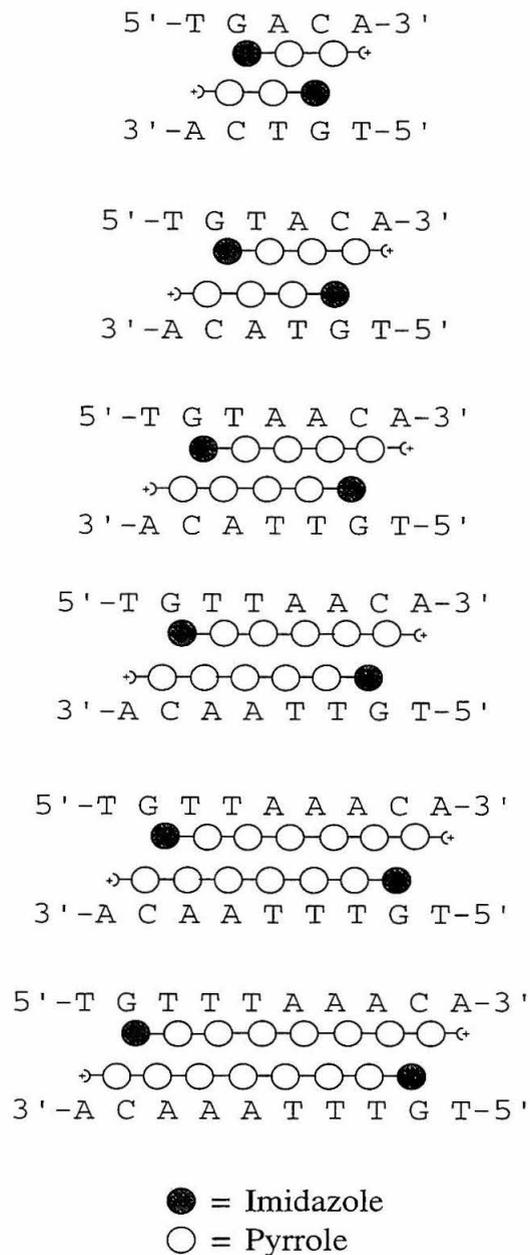


Figure 2.3. Proposed 2:1 binding models of Im-(Py)₂-Dp, Im-(Py)₃-Dp, Im-(Py)₄-Dp, Im-(Py)₅-Dp, Im-(Py)₆-Dp and Im-(Py)₇-Dp binding to five, six, seven, eight, nine and ten base pair match sites, respectively. The imidazole and pyrrole rings are represented as shaded and unshaded spheres, respectively. The amide linkage between the heterocyclic rings are represented by line segments.

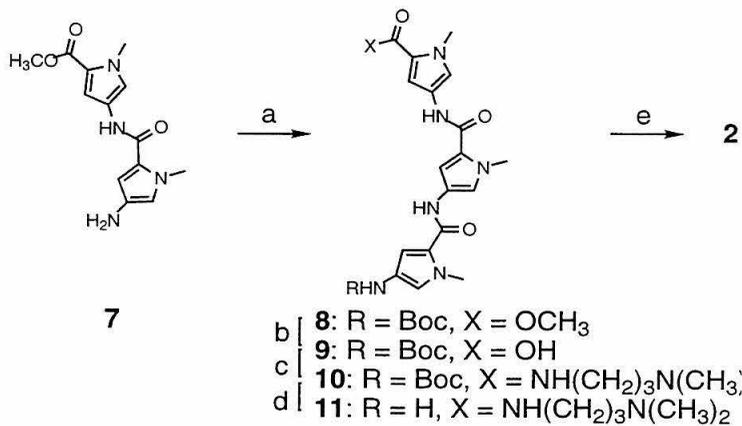
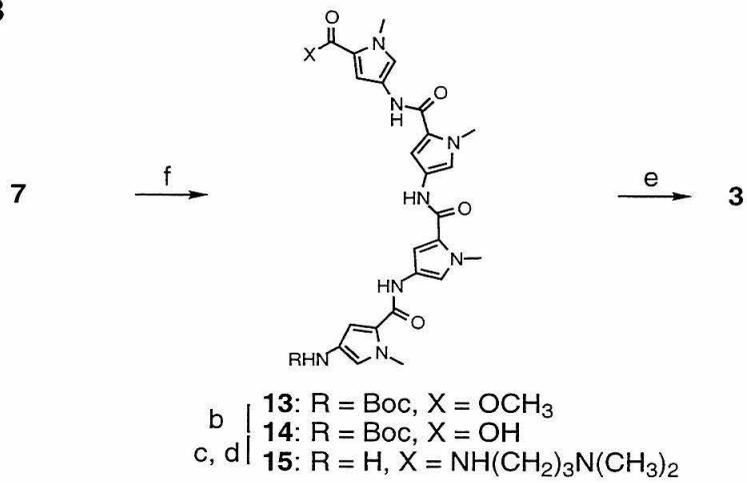
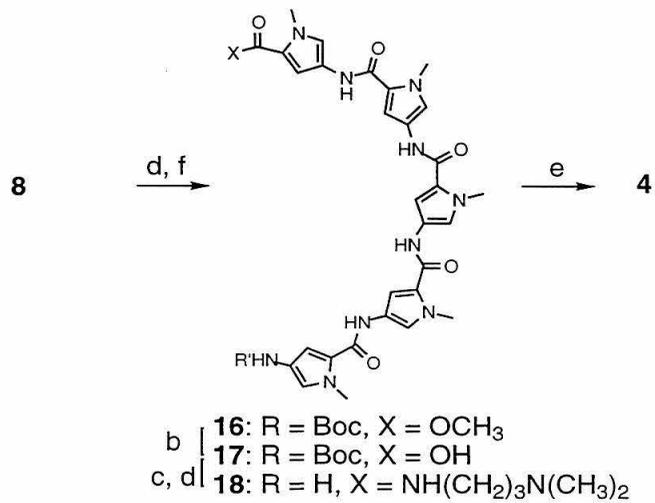
to lock the polyamides in the designed binding sites by preventing them from binding in undesired slipped conformations on the DNA.¹²⁻¹⁴ The mixed A,T tract in the center of the binding sites is designed to favor 2:1 binding in contrast to pure A-tracts which generally favor 1:1 polyamide-DNA complex formation.¹⁶ The binding site size, apparent first-order binding affinity, and sequence specificity of each polyamide was determined by MPE•Fe(II) footprinting and quantitative DNase I footprint titration experiments on a series of restriction fragments containing a match site and a single base pair mismatch site for each polyamide.

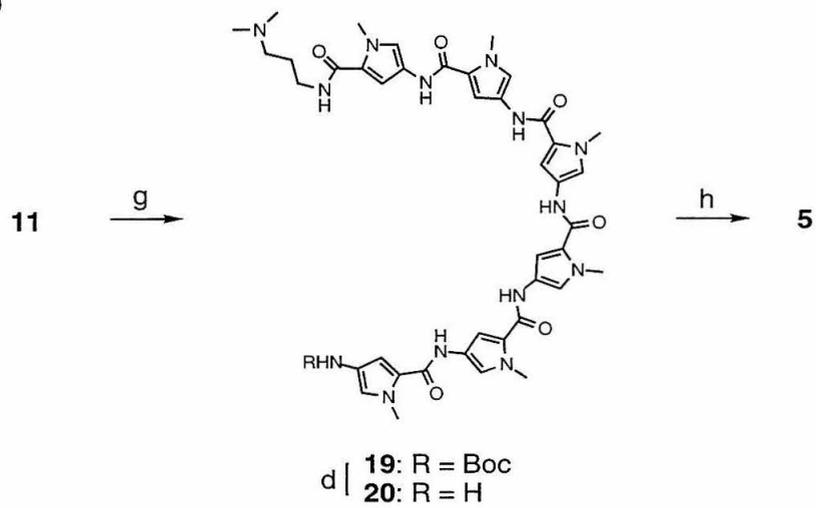
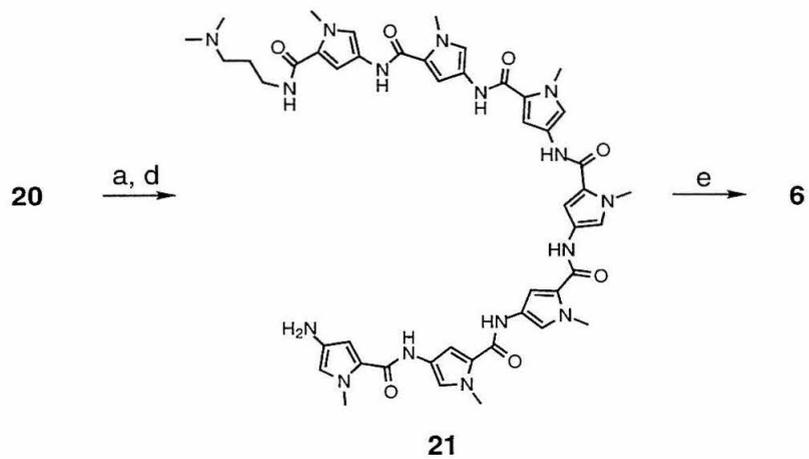
Results

Polyamide Synthesis. The polyamides were synthesized by solution-phase methods (Figure 2.4). In each case the oligo-(*N*-methylpyrrole)carboxamide moiety was prepared by condensation (HBTU, DIEA) of a 4-*tert*-butoxycarbonylamino-*N*-methylpyrrole carboxylic acid derivative with the appropriate 4-amino-*N*-methylpyrrole derivative followed by introduction of the dimethylaminopropyl group at the C-terminus as necessary. Deprotection (TFA) of the N-terminal amino group afforded the 4-amino-oligo-(*N*-methylpyrrole)carboxamide derivatives **11**, **15**, **18**, **20** and **21**. Condensation with 1-methylimidazole-2-carboxylic acid afforded the imidazole-pyrrole containing polyamides Im-(Py)₃-Dp (**2**), Im-(Py)₄-Dp (**3**), Im-(Py)₅-Dp (**4**), Im-(Py)₆-Dp (**5**) and Im-(Py)₇-Dp (**6**) which were purified by reverse phase HPLC.¹⁷

MPE•Fe(II) Footprinting. Binding site sizes were determined by MPE•Fe(II) footprinting (25 mM Tris-acetate, pH 7.0, 10 mM NaCl, 100 μM bp calf thymus DNA, 22 °C) of Im-(Py)₂-Dp, Im-(Py)₃-Dp, Im-(Py)₄-Dp, Im-(Py)₅-Dp, Im-(Py)₆-Dp and Im-(Py)₇-Dp on the ³²P end-labeled *Eco* RI/*Pvu* II restriction fragments from plasmids pJK5, pJK6, pJK7, pJK8, pJK9 and pJK10, respectively (Figures 2.5, 2.6).² Each restriction

Figure 2.4. Synthetic scheme for the synthesis of polyamides (A) Im-(Py)₃-Dp, (B) Im-(Py)₄-Dp, (C) Im-(Py)₅-Dp, (D) Im-(Py)₆-Dp and, (E) Im-(Py)₇-Dp. Reagents: (a) 4-*tert*-butoxycarbonyl-amino-*N*-methylpyrrole-2-carboxylic acid hydroxybenzotriazole ester, DIEA; (b) NaOH, H₂O, EtOH; (c) 3-dimethylaminopropylamine, HBTU; (d) TFA, CH₂Cl₂; (e) 1-methylimidazole-2-carboxylic acid, HBTU, DIEA; (f) 4-*tert*-butoxycarbonylamino-bis-*N*-methylpyrrole-2-carboxylic acid **12**, HBTU, DIEA; (g) **9**, HBTU, DIEA; (h) 1-methylimidazole-2-carboxylic acid, DCC, HOBt.

A**B****C**

D**E**

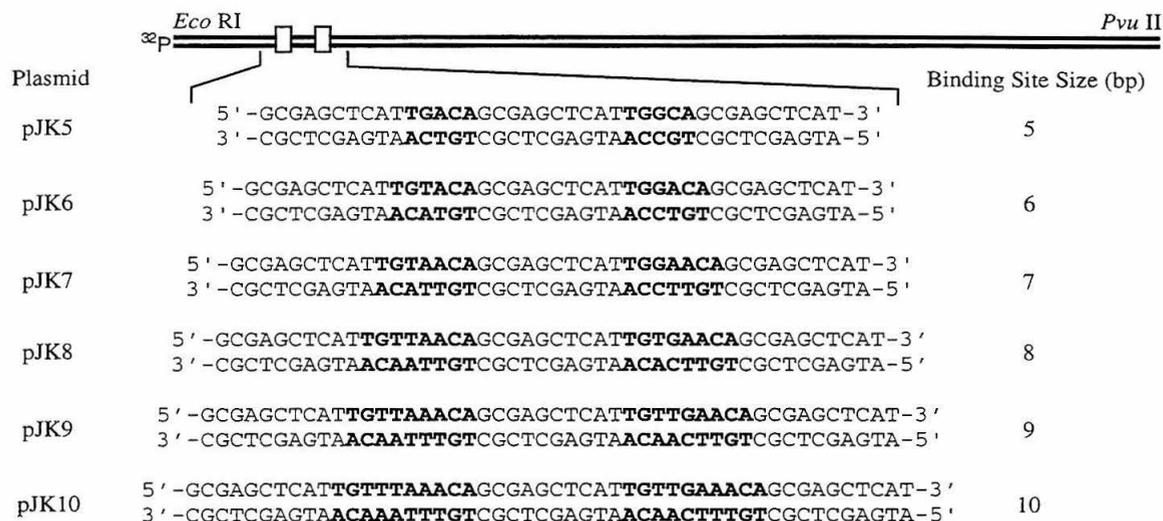
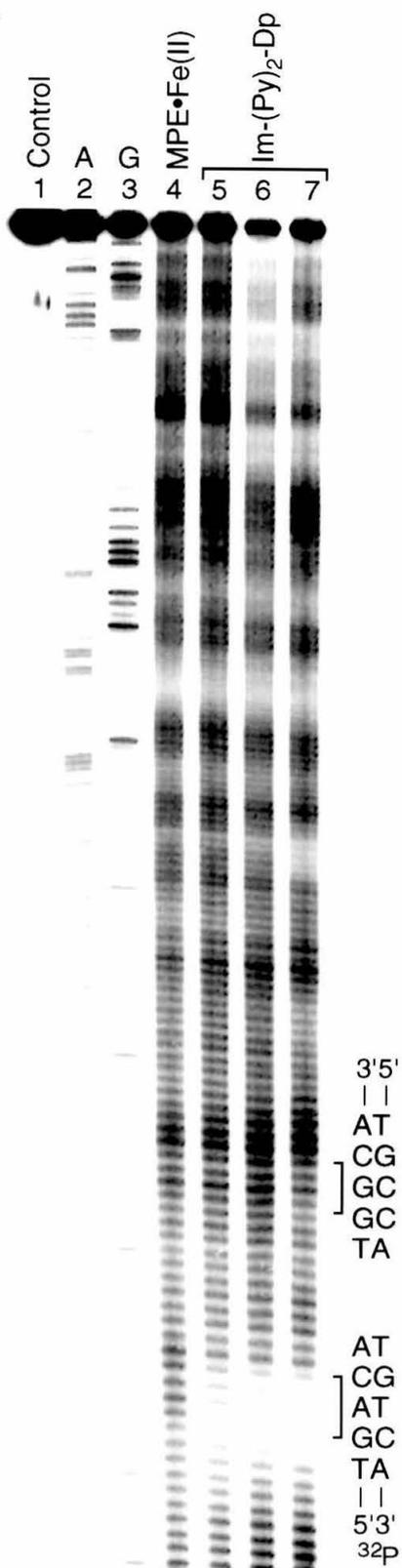


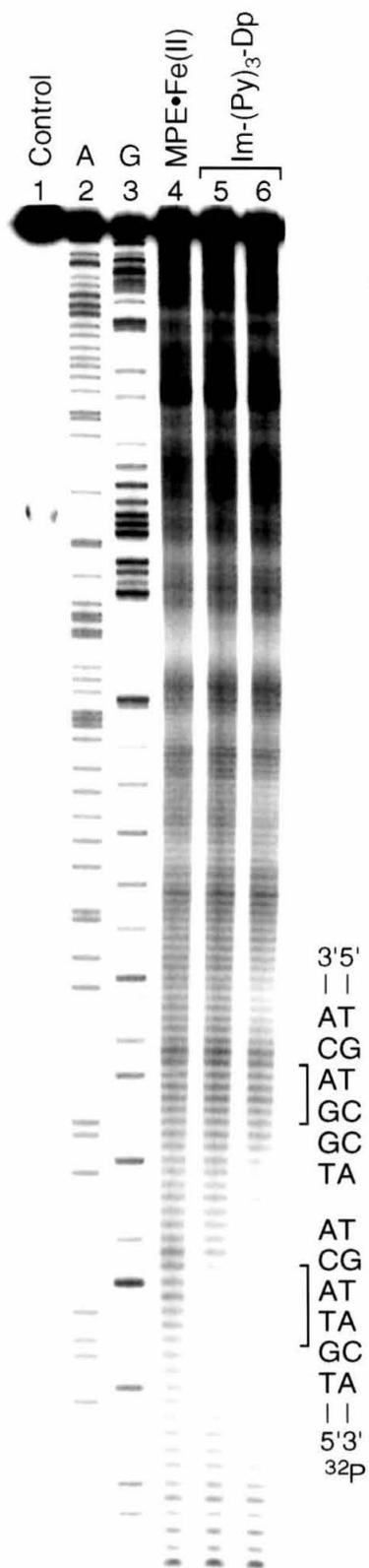
Figure 2.5. Illustration of the *Eco* RI/*Pvu* II restriction fragments from plasmids pJK5, pJK6, pJK7, pJK8, pJK9 and pJK10. The plasmids were prepared by inserting the sequences shown into plasmid pUC19 restricted with *Bam* HI and *Hind* III. The sequences of the match and single base pair mismatch site on each restriction fragment are shown in bold.

Figure 2.6. MPE•Fe(II) footprinting experiments with (A) Im-(Py)₂-Dp, (B) Im-(Py)₃-Dp, (C) Im-(Py)₄-Dp, (D) Im-(Py)₅-Dp, (E) Im-(Py)₆-Dp, and (F) Im-(Py)₇-Dp on the *Eco* RI/*Pvu* II restriction fragments from plasmids pJK5, pJK6, pJK7, pJK8, pJK9 and pJK10, respectively. Gray-scale representations of storage phosphor autoradiograms of 8% denaturing polyacrylamide gels. All reactions contain 25 mM Tris·AcOH pH 7.0, 10 mM NaCl, 100 μM bp calf thymus DNA, 5 μM MPE•Fe(II), 5 mM DTT, and 20-30 kcpm of 5'- or 3'-³²P-labeled restriction fragment as indicated. Lanes 1-4 in (A)-(F) correspond to intact DNA, A reaction, G reaction, and a MPE•Fe(II) control to which no polyamide was added, respectively. In (A) lanes 5-7, 10 μM, 20 μM, and 50 μM Im-(Py)₂-Dp, respectively; (B) lanes 5-6, 2 μM and 5 μM Im-(Py)₃-Dp, respectively; (C) lanes 5-7, 2 μM, 5 μM, and 10 μM Im-(Py)₄-Dp, respectively; (D) lanes 5-7, 2 μM, 5 μM, and 10 μM Im-(Py)₅-Dp, respectively; (E) lanes 5-7, 2 μM, 5 μM, and 10 μM Im-(Py)₆-Dp, respectively; lanes 5-7, 5 μM, 10 μM, and 20 μM Im-(Py)₇-Dp, respectively.

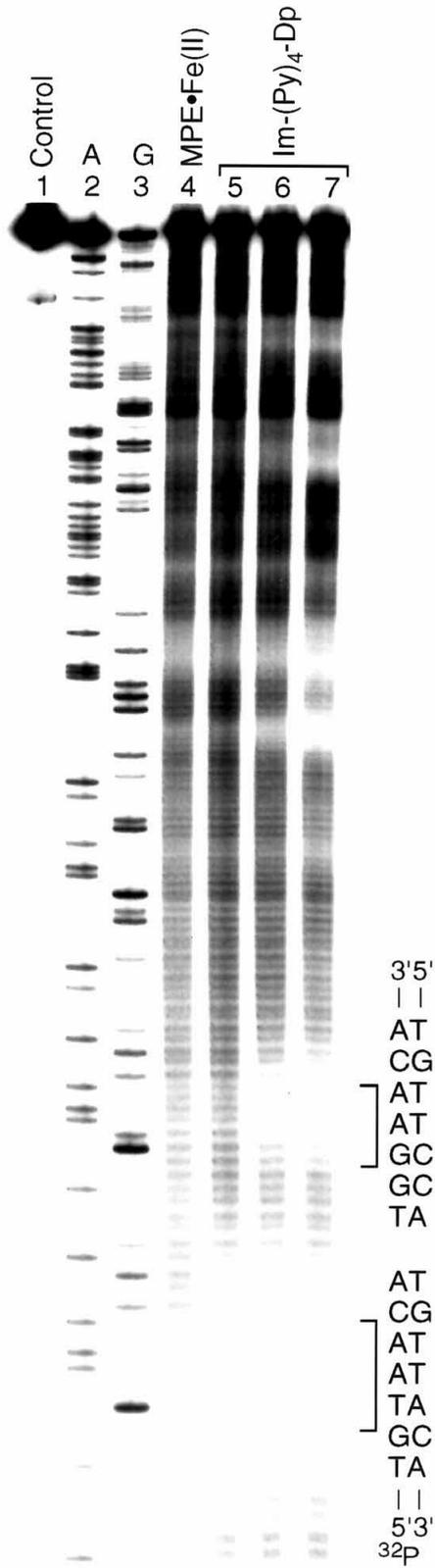
A



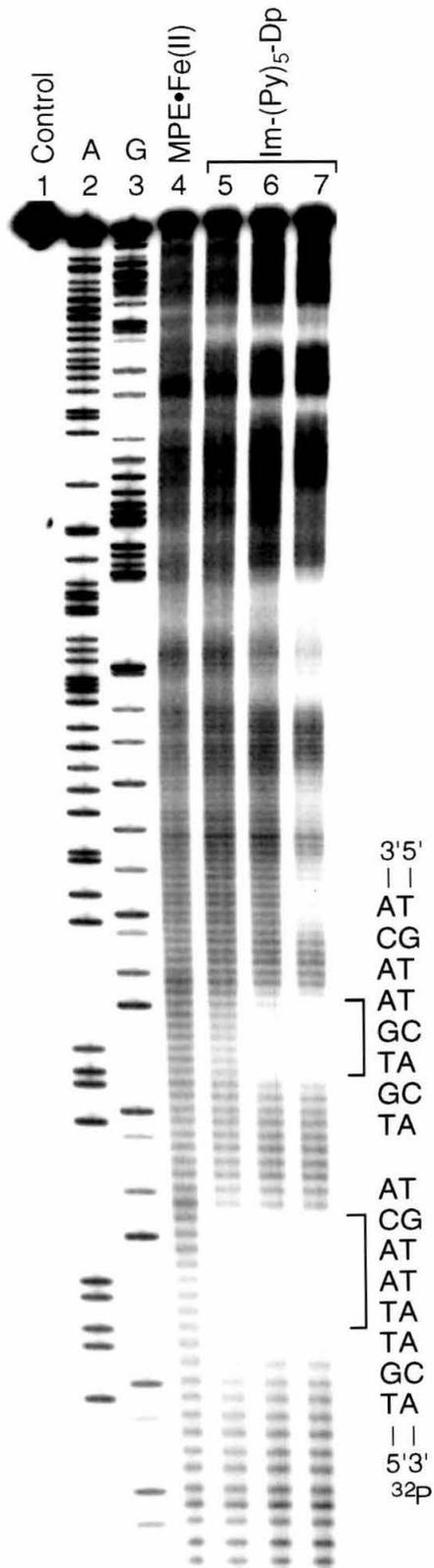
B



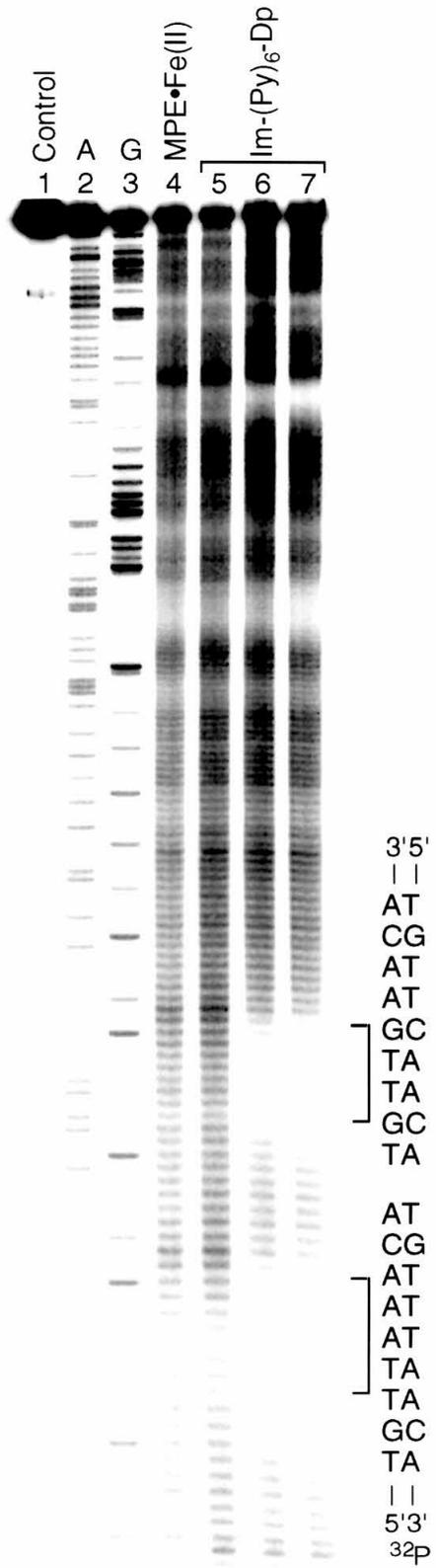
C



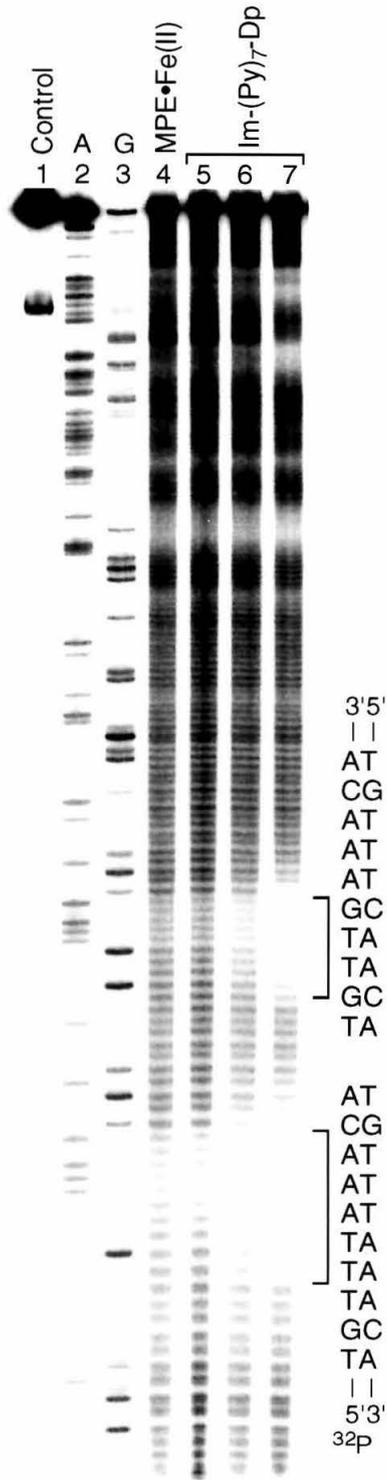
D



E



F



fragment contains a match site and a single base pair mismatch site in which a G•C base pair replaces an A•T or T•A base pair separated by ten base pairs. Analysis of the MPE•Fe(II) protection patterns reveals that the sequential addition of pyrrolicarboxamide moieties to the C-termini of the polyamides increases the preferred DNA binding site size by one base pair with Im-(Py)₂-Dp, Im-(Py)₃-Dp, Im-(Py)₄-Dp, Im-(Py)₅-Dp, Im-(Py)₆-Dp and Im-(Py)₇-Dp binding specifically to five, six, seven, eight, nine, and ten base pair sites, respectively (Figure 2.7). For all six polyamides the observed MPE•Fe(II) protection patterns are 3'-shifted consistent with 2:1 polyamide-DNA complex formation in the minor groove.²

Quantitative DNase I Footprint Titration Experiments. The apparent first-order binding affinity for the match and single base pair mismatch sites for the six polyamides was determined by quantitative DNase I footprint titration experiments (10 mM Tris-HCl pH 7.0, 10 mM KCl, 5 mM MgCl₂, 5 mM CaCl₂, 22 °C) in the absence of unlabeled carrier DNA (Figure 2.8).¹⁸ The θ_{app} points for each polyamide were steep and were adequately fit by a cooperative binding isotherm consistent with 2:1 polyamide-DNA complex formation (Figure 2.9).^{12c} Analysis of the match site data reveals that Im-(Py)₂-Dp binds the five base pair 5'-TGACA-3' site with an affinity of $1.3 \times 10^5 \text{ M}^{-1}$ while Im-(Py)₃-Dp binds the six base pair 5'-TGTACA-3' site with an affinity of $8.6 \times 10^6 \text{ M}^{-1}$, corresponding to a 66-fold enhancement in affinity for the added pyrrolicarboxamide recognizing the additional T•A base pair in the binding site (Table I). The five ring polyamide Im-(Py)₄-Dp binds the seven base pair 5'-TGTAACA-3' with an affinity of $4.5 \times 10^7 \text{ M}^{-1}$, a modest five-fold gain in affinity for recognition of a seven over the six base pair site. Binding affinity levels off with Im-(Py)₅-Dp binding the eight base pair 5'-TGTTAACA-3' site ($K_a = 5.3 \times 10^7 \text{ M}^{-1}$) and Im-(Py)₆-Dp binding the nine base pair 5'-TGTTAACA-3' site ($K_a = 4.7 \times 10^7 \text{ M}^{-1}$) with the same affinity as Im-(Py)₄-Dp for a

Figure 2.7. Histograms of cleavage protection from MPE•Fe(II) for Im-(Py)₂-Dp, Im-(Py)₃-Dp, Im-(Py)₄-Dp, Im-(Py)₅-Dp, Im-(Py)₆-Dp and Im-(Py)₇-Dp binding to the *Eco*RI/*Pvu* II restriction fragments from plasmids pJK5, pJK6, pJK7, pJK8, pJK9 and pJK10, respectively. The individual bar heights are proportional to the protection from MPE•Fe(II) cleavage at each nucleotide. The match and mismatch sites are indicated by boxes.

Im-(Py)₂-Dp (20 μM)

5' -GCGAGCTCATTTGACAGCGAGCTCATTTGGCAAGCGAGCTCAT-3 '
3' -CGCTCGAGTAACCTGTCGCTCGAGTAACCGTCGCTCGAGTA-5 '

Im-(Py)₃-Dp (5 μM)

5' -GCGAGCTCATTTGTACAGCGAGCTCATTTGGACAGCGAGCTCAT-3 '
3' -CGCTCGAGTAACATGTCGCTCGAGTAACCTGTCGCTCGAGTA-5 '

Im-(Py)₄-Dp (5 μM)

5' -GCGAGCTCATTTGTAACAGCGAGCTCATTTGGAACAGCGAGCTCAT-3 '
3' -CGCTCGAGTAACATTTGTCGCTCGAGTAACCTTGTTCGCTCGAGTA-5 '

Im-(Py)₅-Dp (5 μM)

5' -GCGAGCTCATTTGTTAAACAGCGAGCTCATTTGTGAACAGCGAGCTCAT-3 '
3' -CGCTCGAGTAACAATTTGTCGCTCGAGTAACAATTTGTCGCTCGAGTA-5 '

Im-(Py)₆-Dp (10 μM)

5' -GCGAGCTCATTTGTTTAAACAGCGAGCTCATTTGTTGAAACAGCGAGCTCAT-3 '
3' -CGCTCGAGTAACAATTTGTCGCTCGAGTAACAATTTGTCGCTCGAGTA-5 '

Im-(Py)₇-Dp (20 μM)

5' -GCGAGCTCATTTGTTTAAACAGCGAGCTCATTTGTTGAAACAGCGAGCTCAT-3 '
3' -CGCTCGAGTAACAATTTGTCGCTCGAGTAACAATTTGTCGCTCGAGTA-5 '

Figure 2.8a. Quantitative DNase I footprint titration experiment with Im-(Py)₂-Dp on the *Eco* RI/*Pvu* II restriction fragment from plasmid pJK5. Gray-scale representation of a storage phosphor autoradiogram of an 8% denaturing polyacrylamide gel. All reactions contain 10 mM Tris·HCl pH 7.0, 10 mM KCl, 5 mM MgCl₂, 5 mM CaCl₂ and 1-2 × 10⁵ cpm of 3'-³²P-end labeled restriction fragment. Lane 1, intact DNA; lane 2, A reaction; lane 3, G reaction; lane 4, DNase I control lane to which no polyamide was added; lanes 5-17, 5 nM, 10 nM, 20 nM, 50 nM, 0.1 μM, 0.2 μM, 0.5 μM, 1 μM, 2 μM, 5 μM, 10 μM, 20 μM, 50 μM Im-(Py)₂-Dp, respectively.

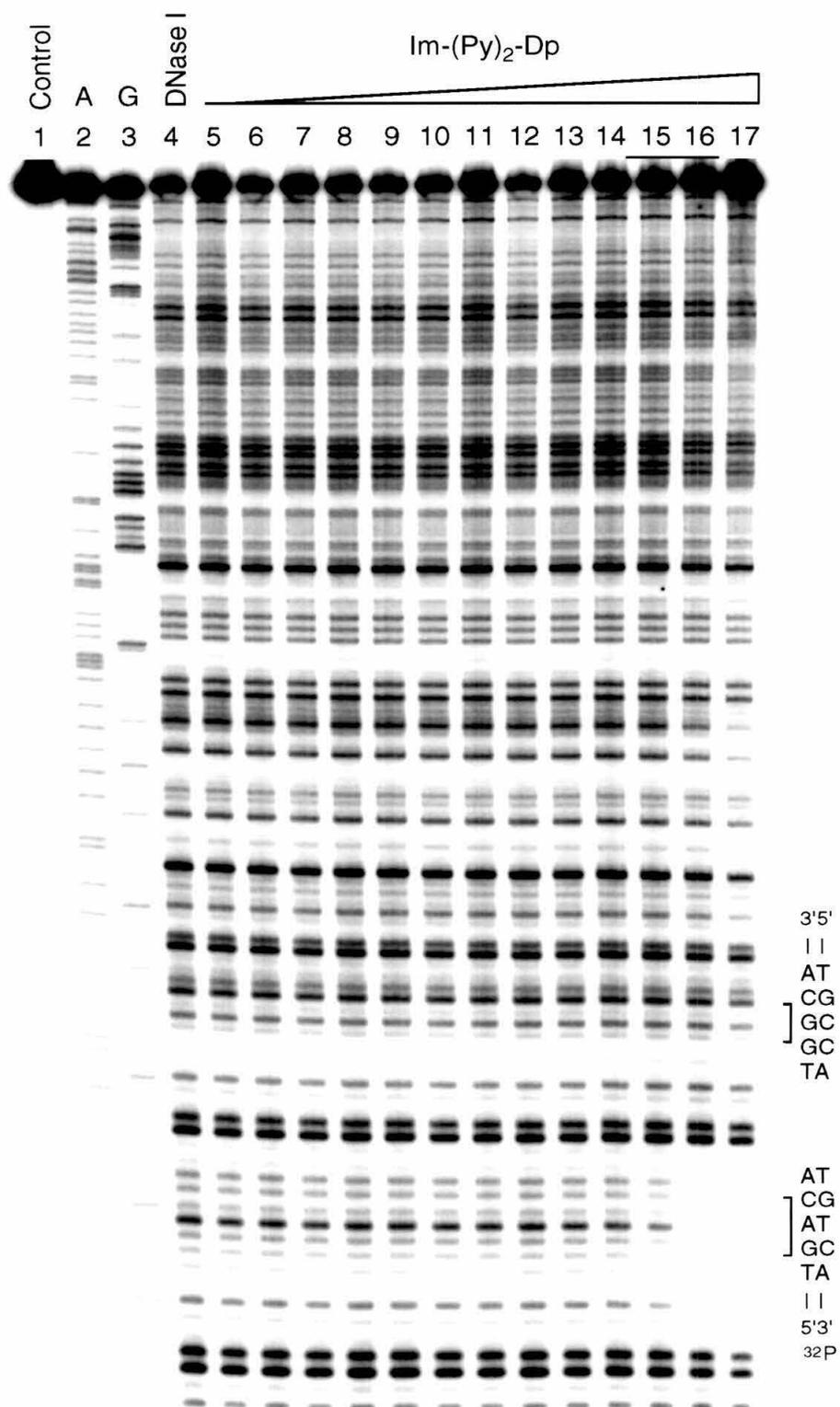


Figure 2.8b. Quantitative DNase I footprint titration experiment with Im-(Py)₃-Dp on the *Eco* RI/*Pvu* II restriction fragment from plasmid pJK6. Gray-scale representation of a storage phosphor autoradiogram of an 8% denaturing polyacrylamide gel. All reactions contain 10 mM Tris-HCl pH 7.0, 10 mM KCl, 5 mM MgCl₂, 5 mM CaCl₂ and 1-2 x 10⁵ cpm of 3'-³²P-end labeled restriction fragment. Lane 1, intact DNA; lane 2, A reaction; lane 3, G reaction; lane 4, DNase I control lane to which no polyamide was added; lanes 5-18, 0.1 nM, 0.2 nM, 0.5 nM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, 0.1 μM, 0.2 μM, 0.5 μM, 1 μM, 2 μM Im-(Py)₃-Dp, respectively.

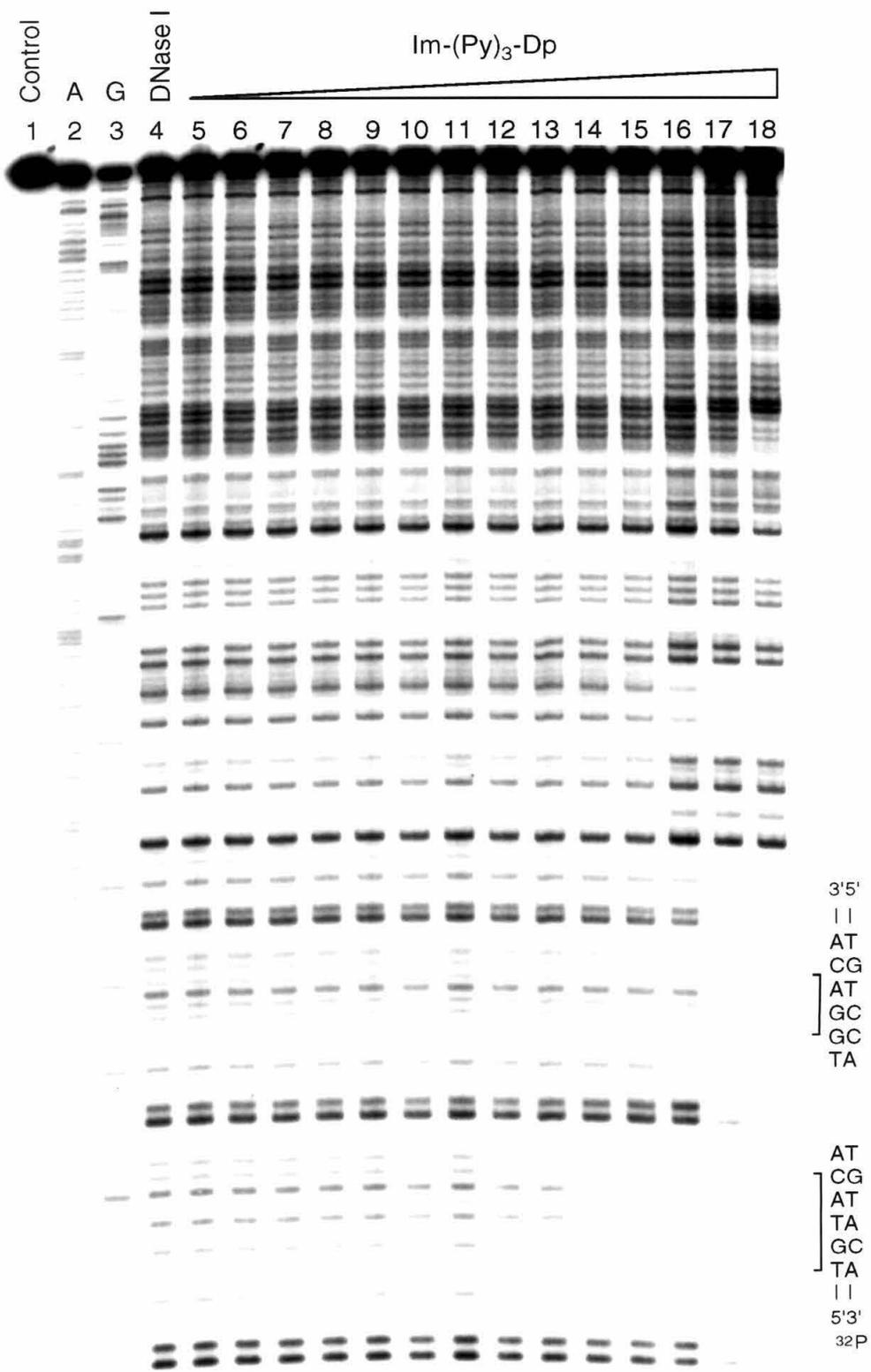


Figure 2.8c. Quantitative DNase I footprint titration experiment with Im-(Py)₄-Dp on the *Eco* RI/*Pvu* II restriction fragment from plasmid pJK7. Gray-scale representation of a storage phosphor autoradiogram of an 8% denaturing polyacrylamide gel. All reactions contain 10 mM Tris·HCl pH 7.0, 10 mM KCl, 5 mM MgCl₂, 5 mM CaCl₂ and 1-2 x 10⁵ cpm of 3'-³²P-end labeled restriction fragment. Lane 1, intact DNA; lane 2, A reaction; lane 3, G reaction; lane 4, DNase I control lane to which no polyamide was added; lanes 5-19, 20 pM, 50 pM, 0.1 nM, 0.2 nM, 0.5 nM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, 0.1 μM, 0.2 μM, 0.5 μM, 1 μM Im-(Py)₄-Dp, respectively.

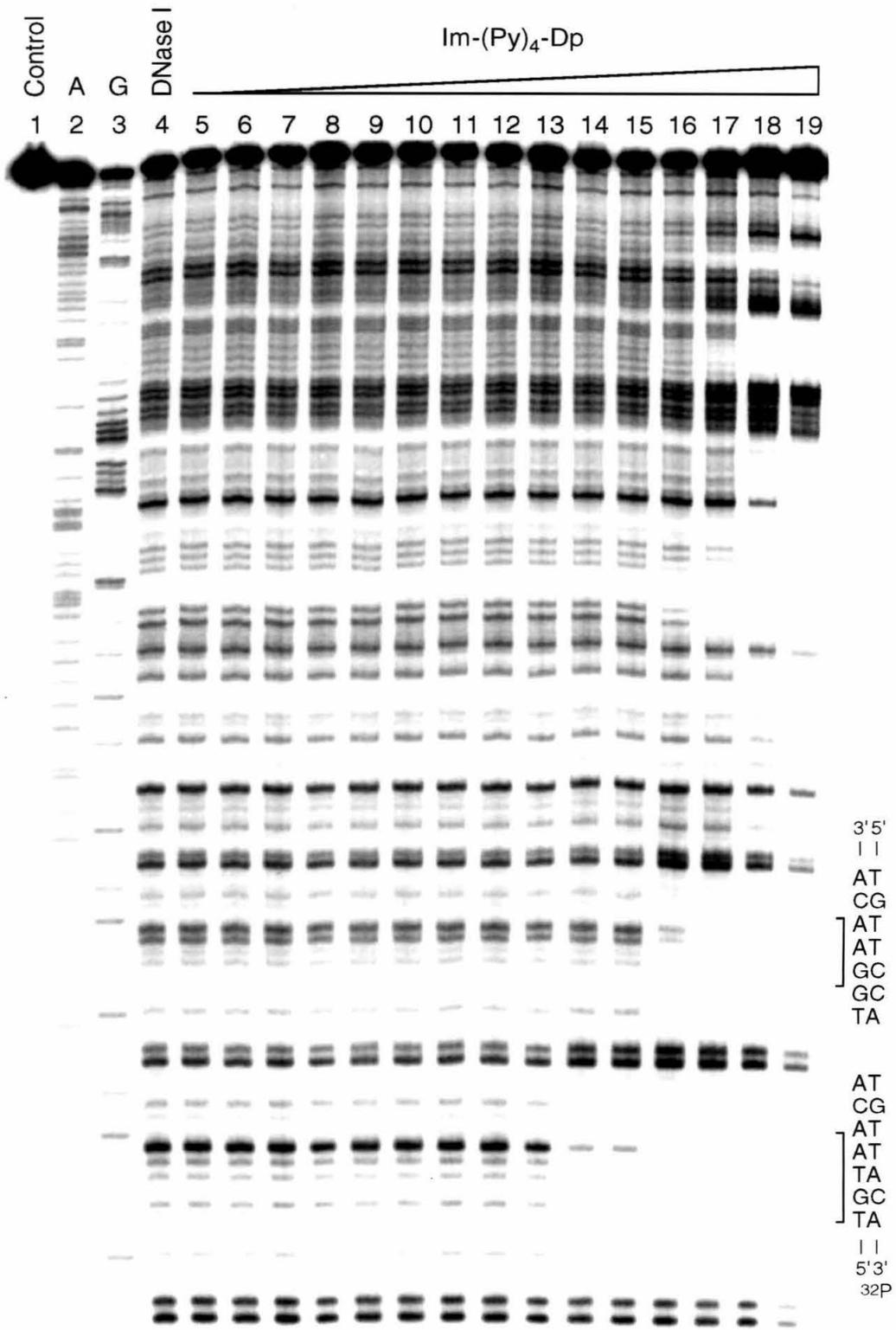


Figure 2.8d. Quantitative DNase I footprint titration experiment with Im-(Py)₅-Dp on the *Eco* RI/*Pvu* II restriction fragment from plasmid pJK8. Gray-scale representation of a storage phosphor autoradiogram of an 8% denaturing polyacrylamide gel. All reactions contain 10 mM Tris·HCl pH 7.0, 10 mM KCl, 5 mM MgCl₂, 5 mM CaCl₂ and 1-2 x 10⁵ cpm of 3'-³²P-end labeled restriction fragment. Lane 1, intact DNA; lane 2, A reaction; lane 3, G reaction; lane 4, DNase I control lane to which no polyamide was added; lanes 5-20, 1 pM, 2 pM, 5 pM, 10 pM, 20 pM, 50 pM, 0.1 nM, 0.2 nM, 0.5 nM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, 0.1 μM Im-(Py)₅-Dp, respectively.

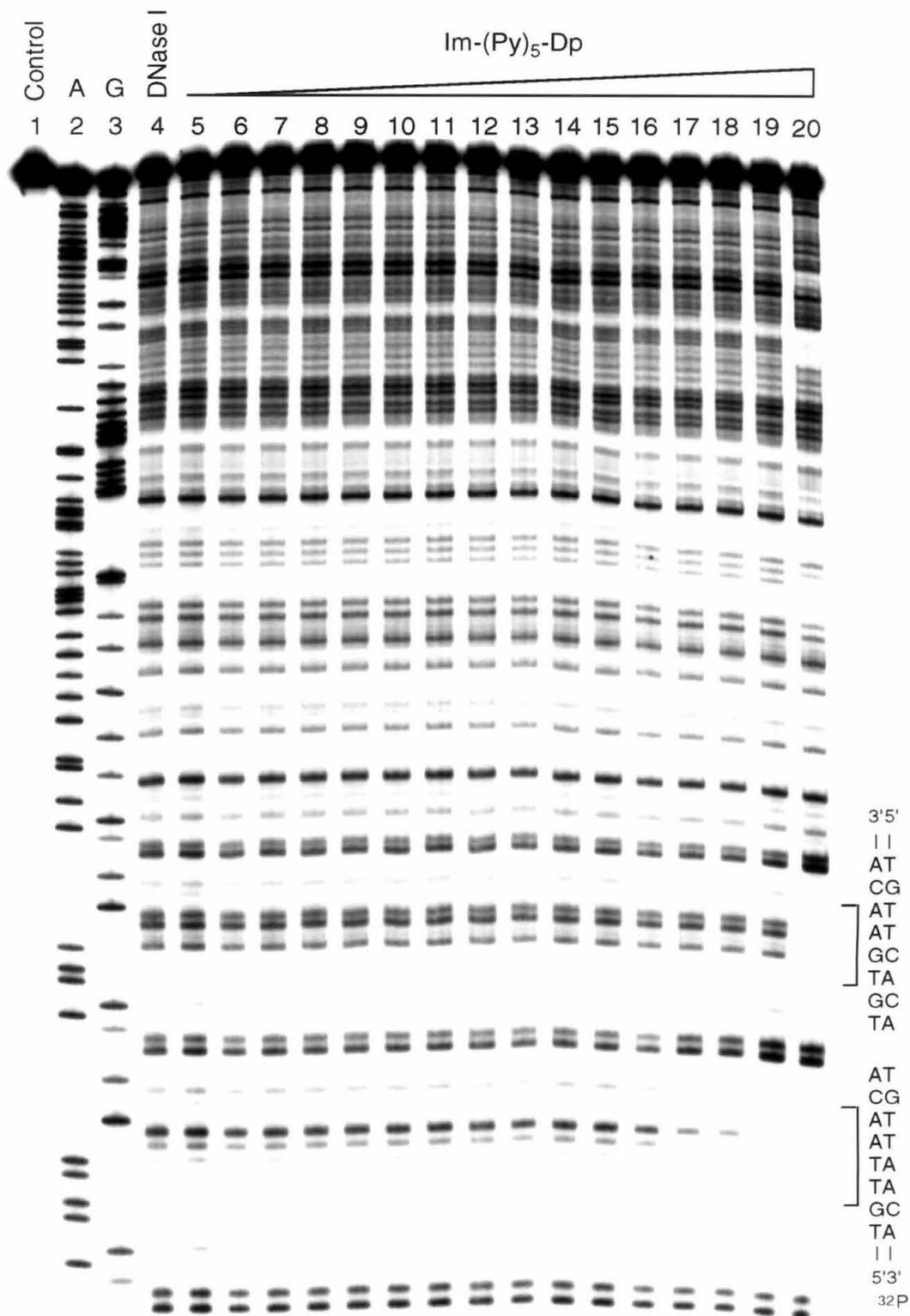


Figure 2.8e. Quantitative DNase I footprint titration experiment with Im-(Py)₆-Dp on the *Eco* RI/*Pvu* II restriction fragment from plasmid pJK9. Gray-scale representation of a storage phosphor autoradiogram of an 8% denaturing polyacrylamide gel. All reactions contain 10 mM Tris·HCl pH 7.0, 10 mM KCl, 5 mM MgCl₂, 5 mM CaCl₂ and 1-2 x 10⁵ cpm of 3'-³²P-end labeled restriction fragment. Lane 1, intact DNA; lane 2, A reaction; lane 3, G reaction; lane 4, DNase I control lane to which no polyamide was added; lanes 5-20, 10 pM, 20 pM, 50 pM, 0.1 nM, 0.2 nM, 0.5 nM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, 0.1 μM, 0.2 μM, 0.5 μM, 1 μM Im-(Py)₆-Dp, respectively.

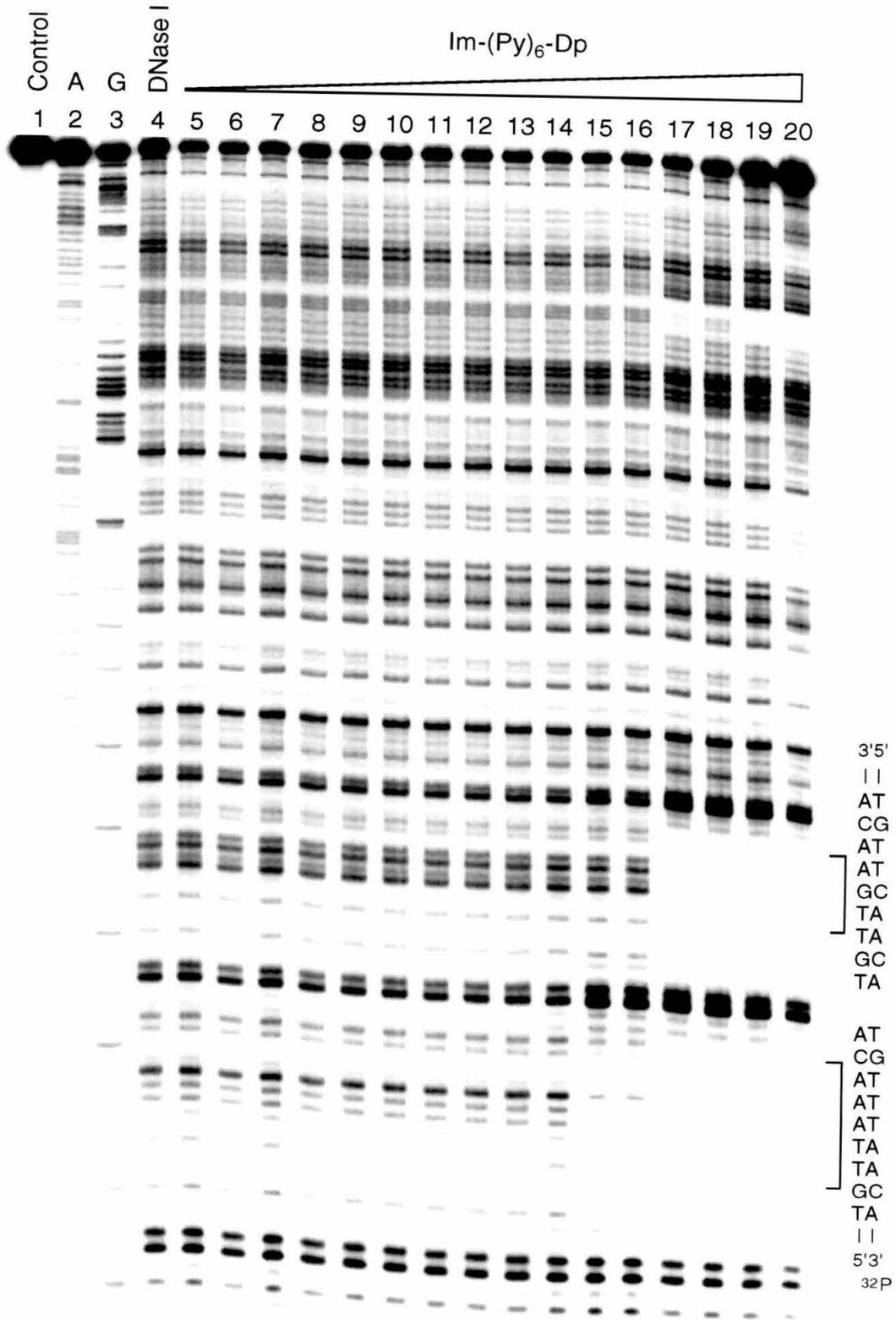
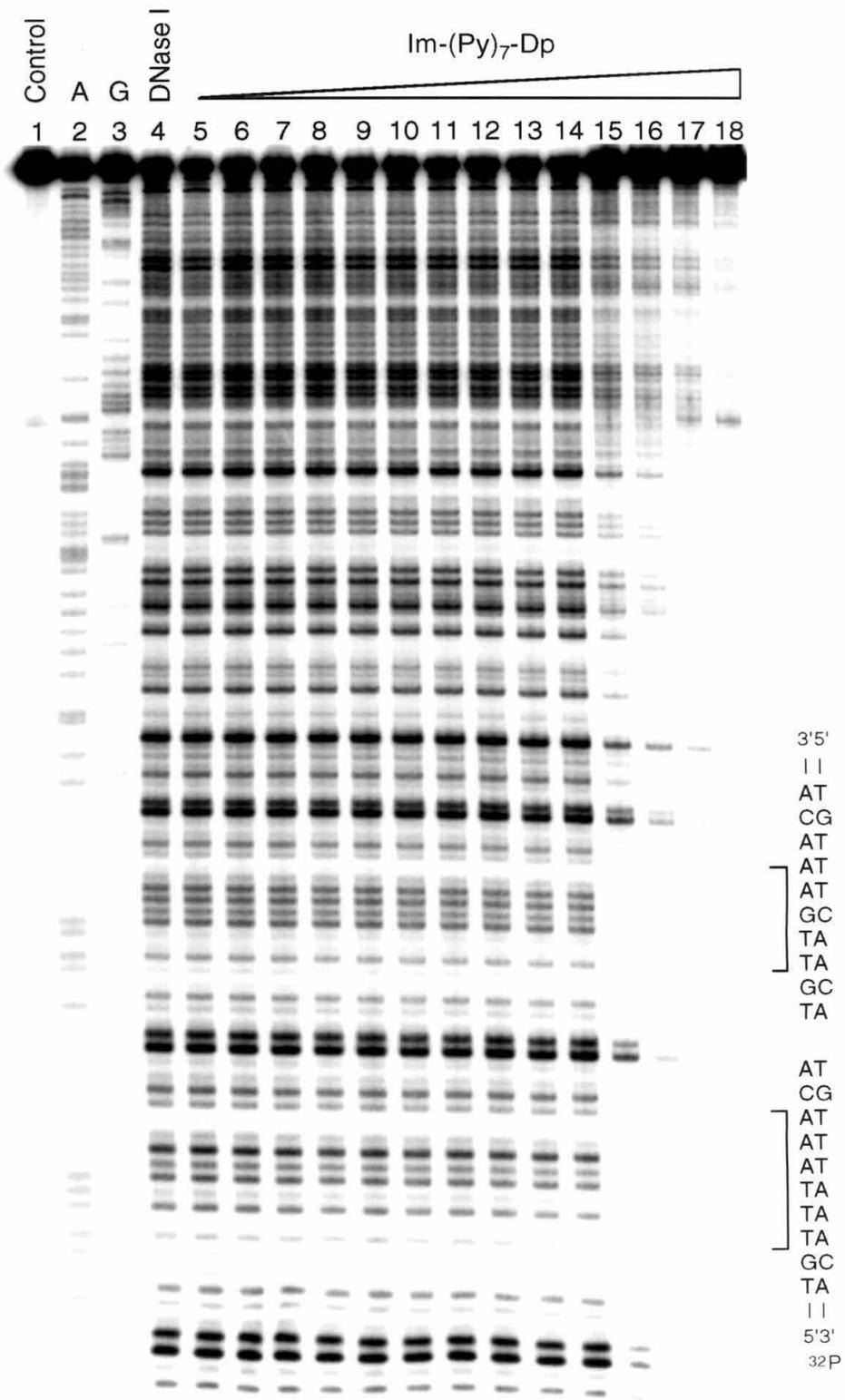


Figure 2.8f. Quantitative DNase I footprint titration experiment with Im-(Py)₇-Dp on the *Eco* RI/*Pvu* II restriction fragment from plasmid pJK10. Gray-scale representation of a storage phosphor autoradiogram of an 8% denaturing polyacrylamide gel. All reactions contain 10 mM Tris·HCl pH 7.0, 10 mM KCl, 5 mM MgCl₂, 5 mM CaCl₂ and 1-2 x 10⁵ cpm of 3'-³²P-end labeled restriction fragment. Lane 1, intact DNA; lane 2, A reaction; lane 3, G reaction; lane 4, DNase I control lane to which no polyamide was added; lanes 5-18, 0.5 nM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, 0.1 μM, 0.2 μM, 0.5 μM, 1 μM, 2 μM, 5 μM, 10 μM Im-(Py)₇-Dp, respectively.



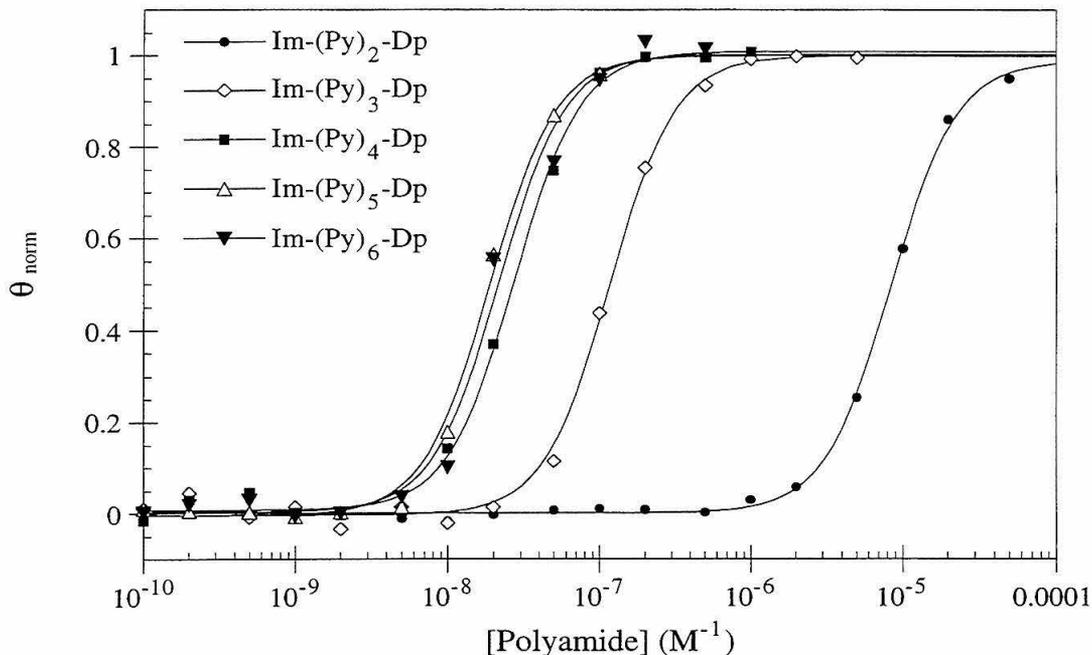


Figure 2.9. Data for the quantitative DNase I footprint titration experiments for $\text{Im}-(\text{Py})_2\text{-Dp}$ to $\text{Im}-(\text{Py})_6\text{-Dp}$ in complex with the five to nine base pair match sites, respectively. The θ_{norm} points were obtained using photostimulable storage phosphor autoradiography and processed as described in the experimental section. The data points for $\text{Im}-(\text{Py})_2\text{-Dp}$, $\text{Im}-(\text{Py})_3\text{-Dp}$, $\text{Im}-(\text{Py})_4\text{-Dp}$, $\text{Im}-(\text{Py})_5\text{-Dp}$ and $\text{Im}-(\text{Py})_6\text{-Dp}$ are represented by filled circles, open diamonds, filled squares, open triangles and filled triangles, respectively. The curve through each set of data points is the best-fit cooperative Langmuir binding titration isotherm ($n = 2$) obtained from a nonlinear least-squares algorithm using eq. 2.

Table I. Apparent First-Order Binding Affinities (M^{-1})^{a,b}

Polyamide	Binding Site Size (bp)	Match	Mismatch	Specificity ^c
Im-(Py) ₂ -Dp	5	1.3×10^5 (0.3)	$< 2 \times 10^4$ ^d	> 6.5 ^e
Im-(Py) ₃ -Dp	6	8.5×10^6 (1.3)	1.6×10^6 (0.2)	5.3 (0.5)
Im-(Py) ₄ -Dp	7	4.5×10^7 (1.1)	7.9×10^6 (1.8)	5.7 (0.8)
Im-(Py) ₅ -Dp	8	5.3×10^7 (0.5)	$< 2 \times 10^7$ ^d	> 2.7 ^e
Im-(Py) ₆ -Dp	9	4.7×10^7 (0.4)	1.7×10^7 (0.7)	2.8 (0.7)
Im-(Py) ₇ -Dp	10	$< 2 \times 10^6$ ^d	$< 2 \times 10^6$ ^d	~ 1

^a Values reported are the mean values from at least three footprint titration experiments. Numbers in parentheses indicate the standard deviation for each data set.

^b The assays were performed at 22 °C, pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl₂, and 5 mM CaCl₂. ^c Defined as the ratio of the match site binding affinity to the binding affinity of the single base pair mismatch site. Numbers in parentheses indicate the uncertainty calculated using the standard deviations of the measured binding affinities. ^d Represents an upper limit for the binding affinity as the quality of fit for this data set was poor. ^e This value represents a lower limit on the specificity.

seven base pair site. Binding affinity drops dramatically for Im-(Py)₇-Dp binding to a ten base pair 5'-TGTTTAAACA-3' site ($K_a < 2 \times 10^6 \text{ M}^{-1}$).

Sequence Specificity. Comparison of the apparent first-order binding affinities for the match and the single base pair mismatch site for each polyamide demonstrates that the specificity of 2:1 polyamide-DNA complex formation, defined as the ratio of match site binding affinity to the affinity for the single base pair mismatch site, generally decreases with increasing polyamide length (Table I). Im-(Py)₂-Dp prefers binding the 5'-TGACA-3' site over the 5'-TGGCA-3' site by a factor of at least 6.5. Im-(Py)₃-Dp and Im-(Py)₄-Dp bind the mismatch sites with five- and six-fold lower affinity, respectively. Im-(Py)₅-Dp displays at least a three-fold preference for the match site while Im-(Py)₆-Dp binds the mismatch site with 2.7-fold lower affinity. At concentrations below 1 μM Im-(Py)₇-Dp shows no binding to DNA. At concentrations 1 μM and above, equal protection from cleavage by DNase I of both the match and mismatch sites are observed indicating that the specificity as defined here is approximately one. However, cleavage protection is also observed at all positions on the restriction fragment suggesting that Im-(Py)₇-Dp is binding DNA in a non-specific manner. The binding affinities reported for Im-(Py)₂-Dp at 5'-TGGCA-3', Im-(Py)₅-Dp at 5'-TGTGAACA-3', and Im-(Py)₇-Dp at the match and mismatch sites represent upper limits since the quality of fits for these data sets were poor.

Discussion

Binding Site Size. The MPE•Fe(II) footprinting experiments demonstrate that the polyamides Im-(Py)₂-Dp, Im-(Py)₃-Dp, Im-(Py)₄-Dp, Im-(Py)₅-Dp, Im-(Py)₆-Dp and Im-(Py)₇-Dp containing three to eight rings, respectively, recognize DNA binding sites five to ten base pairs in length, respectively. These results suggest that the added pyrrolicarboxamide/ pyrrolicarboxamide pairs in the side-by-side dimers recognize the

additional A•T or T•A base pairs in the core of the binding sites as predicted by the model for 2:1 polyamide-DNA complex formation.¹⁰⁻¹⁵ The MPE•Fe(II) protection patterns suggest that the polyamides form specific complexes with the binding sites with no apparent slippage of the ligands in the intermolecular antiparallel dimers.

Binding Affinity and Sequence Specificity. The 66-fold enhancement in apparent first-order binding affinity for Im-(Py)₃-Dp binding to a six base pair 5'-TGTACA-3' ($K_a = 8.6 \times 10^6 \text{ M}^{-1}$) site compared to Im-(Py)₂-Dp binding to a five base pair 5'-TGACA-3' ($K_a = 1.3 \times 10^5 \text{ M}^{-1}$) site corresponds to an increase in the free energy of binding ($\Delta\Delta G$) of 2.4 kcal/mol (22 °C) (Figure 2.10). This apparent increase in the free energy of binding is likely a result of the additional pyrrolicarboxamide/pyrrolicarboxamide pair in the Im-(Py)₃-Dp intermolecular dimer recognizing the T•A base pair in the context of the 5'-TGTACA-3' sequence. A significant fraction of the increase in binding affinity for Im-(Py)₃-Dp presumably results from the hydrogen bonds and van der Waals contacts that likely form when the additional pyrrolicarboxamides in the ligand bind the additional T•A base pair in the core of the binding site. The five-fold enhancement in affinity of Im-(Py)₄-Dp for the seven base pair 5'-TGTAACA-3' site ($K_a = 4.5 \times 10^7 \text{ M}^{-1}$) over Im-(Py)₃-Dp binding to 5'-TGTACA-3' corresponds to an increase in the free energy of binding of 1.0 kcal/mol. This relatively modest change in binding free energy suggests that the addition of a pyrrolicarboxamide moiety to Im-(Py)₃-Dp in order to recognize an additional A•T base pair in the seven base pair 5'-TGTAACA-3' site is less favorable than the addition of a pyrrolicarboxamide to Im-(Py)₂-Dp to recognize the additional T•A base pair in the six base pair 5'-TGTACA-3' site. The observation that the affinities of Im-(Py)₅-Dp ($K_a = 5.3 \times 10^7 \text{ M}^{-1}$) and Im-(Py)₆-Dp ($K_a = 4.7 \times 10^7 \text{ M}^{-1}$) for the eight and nine base pair 5'-TGTTAACA-3' and 5'-TGTTAACA-3' sites, respectively, are the same as Im-(Py)₄-Dp binding 5'-TGTAACA-3' indicates that further lengthening of

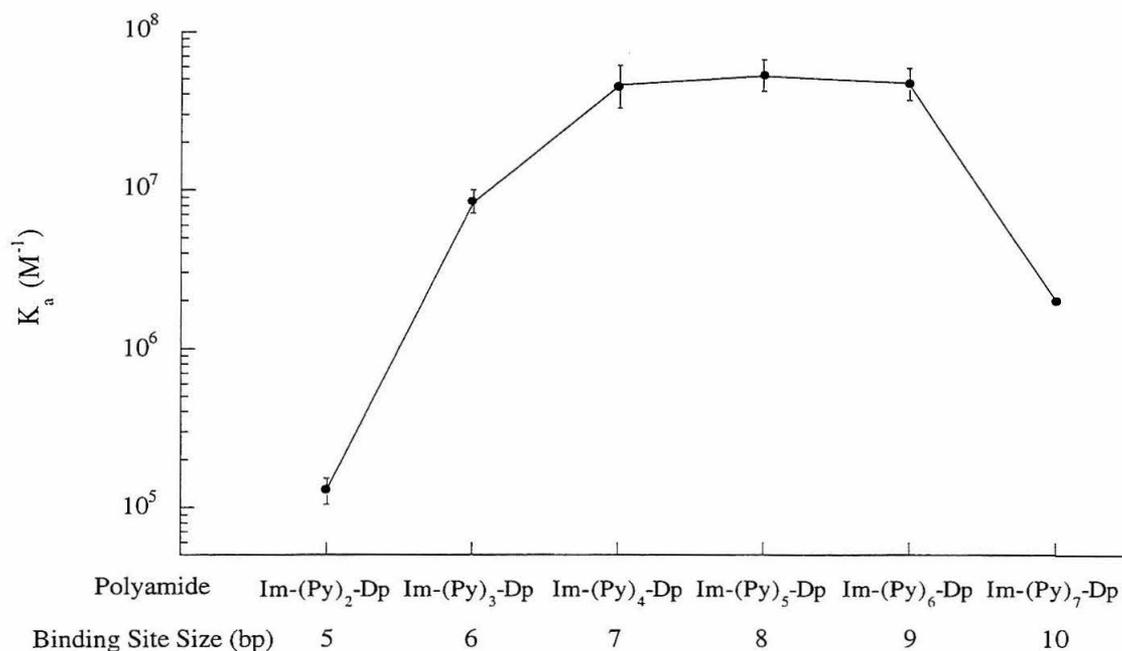


Figure 2.10. Plot of apparent first order binding affinity (K_a) for polyamides Im-(Py)₂-Dp-Im-(Py)₇-Dp for the five to ten base pair match sites. The binding affinities are plotted on a logarithmic scale where the error bar heights represent the standard deviation of the average binding affinity determined for each polyamide. The value of $2 \times 10^6 \text{ M}^{-1}$ for Im-Py₇-Dp represents an upper limit of the binding affinity and is shown without error bars.

the polyamides from five to seven rings and the DNA binding site from seven to nine base pairs has no effect on the observed binding free energies of Im-(Py)₅-Dp and Im-(Py)₆-Dp. The dramatic decrease in binding affinity of Im-(Py)₇-Dp for the ten base pair 5'-TGTTTAAACA-3' site ($K_a < 2 \times 10^6 \text{ M}^{-1}$) suggests that the presence of an additional pyrrolicarboxamide in the ligand is making an unfavorable contribution to the binding free energy of Im-(Py)₇-Dp for the 5'-TGTTTAAACA-3' site.

One explanation for these observations is that three and four ring polyamides Im-(Py)₂-Dp and Im-(Py)₃-Dp, respectively, are in phase with their DNA binding sites. High-resolution NMR and molecular modeling of a related (Im-(Py)₂-Dp)₂•5'-TGACT-3' complex showed that the curvature of the ligand closely matched the curvature of the minor groove of the five base pair binding site.^{12b} Similar observations were made for the four ring polyamide Im-Py-Im-Py-Dp binding to a 5'-AGCGCT-3' sequence.^{14a} For the longer polyamides beginning with Im-(Py)₄-Dp where binding affinities level off at Im-(Py)₅-Dp and ultimately decrease at Im-(Py)₇-Dp, the ligands may be falling out of register with the DNA binding sites.²⁰ In these 2:1 complexes the curvature of the polyamides would likely not match the curvature of the minor groove surfaces of the longer binding sites. One likely consequence would be that the energetic benefit of the hydrogen bonds and van der Waals interactions that stabilize the 2:1 complexes would decrease as the lack of correspondence between the binding surfaces of the polyamides and the DNA sites became more pronounced.²¹ The trend in the observed binding affinities for the six polyamides studied here would be adequately explained if the degree of register mismatch between the polyamides and the DNA increased with polyamide length beyond Im-(Py)₃-Dp. Register mismatch in the larger polyamide-DNA complexes may explain the decrease in sequence specificity observed for the longer polyamides (Figure 2.11). In the mismatch sites a G•C base pair replaces an A•T or T•A base in the core of the binding site. The 2-

amino group of guanine protrudes from the floor of the minor groove and therefore is expected to introduce a bump on the surface of the binding sites that may interfere with the hydrogen bonding of the pyrrolicarboxamide NH of the polyamides to N3 of the guanine base.^{4a} If the longer polyamides are out of register with the DNA then the effect of this unfavorable interaction on the free energy of binding at the mismatch site may become less significant. This would lower the overall free energy difference between binding the match and mismatch sites and thereby reduce specificity.

We note that sequence-dependent DNA structural features such as intrinsic minor groove width, minor groove flexibility, and inherent curvature may differ between each of the binding sites and could therefore contribute to the measured difference in the binding affinities of the six polyamides.²² One possibility is that as the length of the A,T-tract in the binding sites increases the minor groove width and/or flexibility may decrease which would likely impose an energetic penalty for 2:1 polyamide-DNA complex formation and lower the apparent binding affinities for these sites. These effects may also contribute to the lower sequence specificity observed for the longer polyamides in the series. Insertion of G•C base pairs into an A•T tract may increase the minor groove width and/or flexibility which may reduce the difference in binding free energy between the match and mismatch sites thereby lowering specificity.²³ These considerations of sequence-dependent DNA structure suggest that the sequence of the putative DNA target site may be an important factor in determining polyamide affinity and specificity in the 2:1 motif. For example, although A•T and T•A base pairs are partially degenerate in the minor groove it remains an open question as to whether imidazole-pyrrole polyamides show a preference for one base pair over the other in 2:1 complexes.

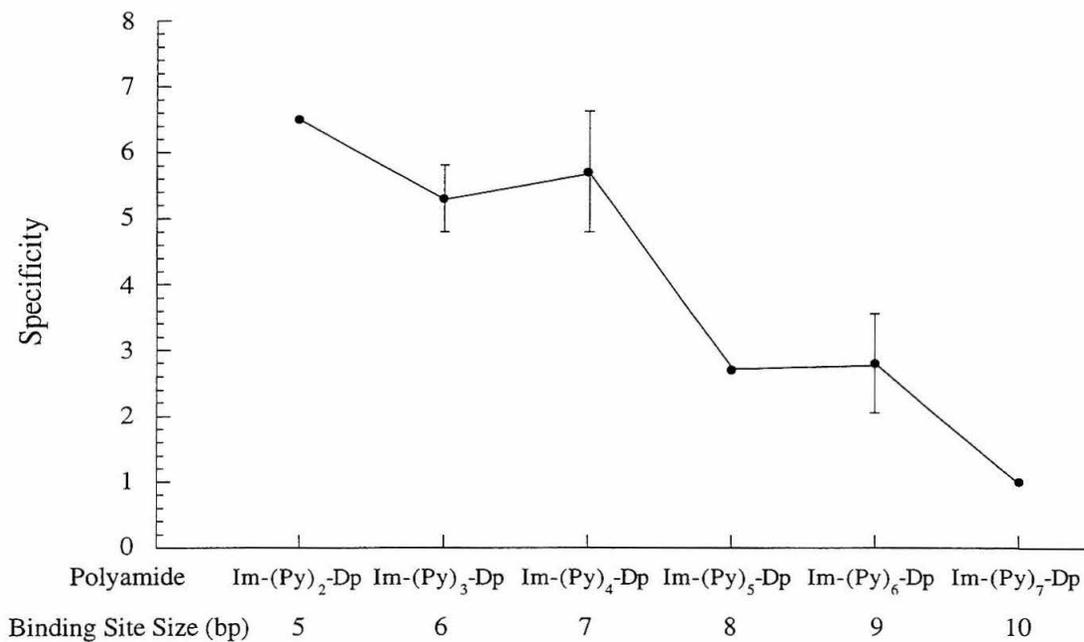


Figure 2.11. Plot of the observed specificity for a match site over a single base pair mismatch site for polyamides Im-(Py)₂-Dp-Im-(Py)₇-Dp. Error bar heights represent the uncertainty in the specificity calculated using the standard deviations of the apparent first-order binding affinities of the match and mismatch sites. The specificity values for Im-(Py)₂-Dp and Im-(Py)₅-Dp represent lower limits and are shown without error bars. The apparent specificity of 1 for Im-(Py)₇-Dp is approximate and is shown without error bars.

Implications for the Design of Minor Groove Binding Polyamides. The results of this study demonstrate that DNA sequences up to nine base pairs in length can be *specifically* recognized by imidazole-pyrrole polyamides containing three to seven rings by 2:1 polyamide-DNA complex formation in the minor groove. Recognition of a nine base pair site defines the new lower limit of the binding site size that can be recognized by polyamides containing exclusively imidazole and pyrrolicarboxamides. We note that binding affinity reaches a maximum value ($K_a \sim 5 \times 10^7 \text{ M}^{-1}$) for the five ring polyamide Im-(Py)₄-Dp and that addition of up to two additional pyrrolicarboxamides has no effect on the observed binding affinity. These results and the failure of an eight ring polyamide to specifically recognize a ten base pair site suggests that a new class of polyamides is needed for extension of the 2:1 polyamide-DNA motif to sequences longer than nine base pairs. Perhaps polyamides in which the rigid imidazole or pyrrole rings are replaced by more flexible moieties will match the curvature of longer DNA sites and bind with high affinity. The affinities observed for the six to nine base pair sites are significant considering that the polyamides in the 2:1 complexes are unlinked. Covalently linking the individual polyamides in a head-to-tail fashion with γ -aminobutyric acid (GABA) to afford hairpin polyamides or with two GABA moieties to afford cyclic polyamides may afford new polyamides that bind DNA sites up to nine base pairs with higher affinity and sequence specificity.²⁴⁻²⁵ Finally, the relatively modest sequence specificities of the polyamides studied here suggest that increasing the free energy difference between match and mismatch sites is an important goal in the design of nonnatural ligands that bind DNA with high affinity *and* sequence specificity.

Experimental Section

Polyamide Synthesis. ^1H NMR were recorded on a General Electric QE-300 spectrometer with chemical shifts reported in parts per million relative to residual solvent signal. High-resolution mass spectra (HRMS) were recorded using fast atom bombardment (FAB) techniques at the High-Resolution Mass Spectrometry Facility at the California Institute of Technology or the Mass Spectrometry Laboratory at the University of California, Riverside. Matrix-assisted laser desorption/ionization time-of-flight mass spectra (MALDI-TOF MS) were recorded in the Biomolecule Analysis Facility at the California Institute of Technology. Ultraviolet-visible (UV) spectra were recorded on a Hewlett Packard 8452A diode array spectrophotometer. Reactions were executed under an atmosphere of argon unless otherwise noted. Im-(Py)₂-Dp and 1-methylimidazole-2-carboxylic acid were prepared as previously described.^{12a} Reagent grade chemicals were used as received unless noted otherwise. Dimethylformamide (DMF), dimethylsulfoxide/*N*-methylpyrrolidone (DMSO/NMP) and diisopropylethylamine (DIEA) were purchased from Applied Biosystems. Di-*tert*-butyl dicarbonate ((Boc₂)O), 2-(1*H*-benzotriazol-1-yl)-1,1,3,3-tetramethyl-uronium hexafluorophosphate (HBTU), dicyclohexylcarbodiimide (DCC) and hydroxybenzotriazole (HOBT) were purchased from Peptides International which is the recommended supplier for these reagents. Dichloromethane was purchased from EM Science. Trifluoroacetic acid (TFA) was purchased from Halocarbon. Flash column chromatography was carried out using EM Science Kieselgel 60 (230-400 mesh).²⁶ Thin-layer chromatography was performed on EM Reagents silica gel plates (0.5 mm thickness). All compounds were visualized at 254 nm. Preparative reverse phase high-performance liquid chromatography (RP-HPLC) was performed on a Beckman HPLC system using a Waters Delta-Pak Prep-Pak C₁₈ column (25 mm x 10 cm, 15 μ, 100 Å) equipped with a Delta-Pak Guard-Pak column.

Compounds were eluted with a gradient of 0-60% CH₃CN/0.1% TFA over 120 min with a flow rate of 8-10 mL/min and detected using a Beckman 167 dual wavelength detector at 254 and 340 nm. Fractions were collected and analyzed by analytical reverse phase HPLC on a Beckman Gold HPLC system using a Rainin Microsorb-MV C₁₈ column (4.6 mm x 30 cm, 5 μ, 100 Å). Compounds were eluted with a gradient of 0-60% CH₃CN/0.1% TFA over 60 min with a flow rate of 1 mL/min and detected using a dual wavelength detector at 254 and 340 nm. The desired fractions were pooled and lyophilized.

4-tert-Butoxycarbonylamino-tri-*N*-methylpyrrole-2-carboxylic acid methyl ester (8). To a solution of 4-amino-bis-*N*-methylpyrrole-2-carboxylic acid methyl ester hydrochloride (**7**) (1.0 g, 3.2 mmol) and DIEA (2.23 g, 17.2 mmol) in DMF (8 mL) was added 4-*tert*-butoxycarbonylamino-*N*-methylpyrrole-2-carboxylic acid hydroxy-benzotriazole ester (1.1 g, 3.0 mmol). After stirring overnight at room temperature the solution was added slowly to 400 mL H₂O. The resulting light brown powder was isolated by vacuum filtration, washed with cold H₂O, dried *in vacuo*, and purified by flash chromatography (2-5% CH₃OH/CH₂Cl₂) to afford ester **8** (1.32 g, 83%) as a tan solid. ¹H NMR (DMSO-*d*₆) δ 9.97 (s, 1 H), 9.90 (s, 1 H), 9.08 (s, 1 H), 7.42 (s, 1 H), 7.19 (s, 1 H), 7.01 (s, 1 H), 6.82 (s, 2 H), 6.80 (s, 1 H), 3.82 (s, 6 H), 3.72 (s, 3 H), 3.68 (s, 3 H), 1.44 (s, 9 H); FAB-HRMS *m/e* 498.2258 (M⁺, 498.5425 calcd. for C₂₄H₃₀N₆O₆).

4-tert-Butoxycarbonylamino-tri-*N*-methylpyrrole-2-carboxylic acid (9). To a slurry of 4-*tert*-butoxycarbonylamino-tri-*N*-methylpyrrole-2-carboxylic acid methyl ester **8** (1.32 g, 2.65 mmol) in 95% EtOH (125 mL) was added 1 M NaOH (125 mL) and the mixture heated at 60 °C for 1 h. Ethanol was removed under reduced pressure and the resulting solution was cooled to room temperature before acidifying to pH 3 with 1 M KHSO₄. The resulting white solid was collected by vacuum filtration, washed with cold H₂O, and dried *in vacuo* to afford carboxylic acid **9** (938 mg, 73%) as a tan powder. ¹H

NMR (DMSO- d_6) δ 9.90 (s, 1 H), 9.84 (s, 1 H), 9.06 (s, 1 H), 7.38 (d, 1 H, $J = 1.0$ Hz), 7.20 (d, 1 H, $J = 1.3$ Hz), 7.01 (d, 1 H, $J = 1.3$ Hz), 6.87 (s, 1 H), 6.81 (d, 1 H, $J = 1.3$ Hz), 6.80 (s, 1 H), 3.81 (s, 3 H), 3.79 (s, 3 H), 3.78 (s, 3 H), 1.43 (s, 9 H); FAB-HRMS m/e 484.2071 (M^+ , 484.5154 calcd. for $C_{23}H_{28}N_6O_6$).

4-*tert*-Butoxycarbonylamino-tri-*N*-methylpyrrole-2-(3-dimethylamino-propyl)-carboxamide (10). To a solution of 4-*tert*-butoxy-carbonylamino-tri-*N*-methylpyrrole-2-carboxylic acid **9** (2.0 g, 4.13 mmol) and DIEA (556 mg, 4.31 mmol) in DMF (10 mL) was added HBTU (1.6 g, 4.22 mmol). After stirring 10 h at room temperature 3-dimethylaminopropylamine (506 mg, 4.96 mmol) was added and the solution stirred 2 h at room temperature. The solution was concentrated *in vacuo* and the resulting residue was purified by flash chromatography (0-1% NH_4OH/CH_3OH) to afford **10** (1.04 g, 44%). 1H NMR (DMSO- d_6) δ 9.88 (s, 1 H), 9.87 (s, 1 H), 9.12 (s, 1 H), 8.09 (t, 1 H, $J = 4.7$ Hz), 7.62 (m, 1 H), 7.47 (m, 1 H), 7.19 (m, 1 H), 7.06 (m, 2 H), 6.82 (m, 1 H), 3.82 (s, 3 H), 3.80 (s, 3 H), 3.78 (s, 3 H), 3.22 (m, 2 H), 2.24 (t, 2 H, $J = 7.0$ Hz), 2.13 (s, 6 H), 1.60 (q, 2 H, $J = 7.0$ Hz), 1.44 (s, 9 H); FAB-HRMS m/e 568.3202 (M^+ , 568.6810 calcd. for $C_{28}H_{40}N_8O_5$).

4-Amino-tri-*N*-methylpyrrole-2-(3-dimethylaminopropyl)-carboxamide bistrifluoroacetate (11). Trifluoroacetic acid (14 mL, 182 mmol) was added to 4-*tert*-butoxycarbonylamino-tri-*N*-methylpyrrole-2-(3-dimethylaminopropyl)-carboxamide **10** (500 mg, 0.88 mmol) and stirred at room temperature. After 20 min excess TFA was removed under reduced pressure and the resulting residue dried *in vacuo* to afford **11** (517 mg, 84%) as a yellow solid. 1H NMR (DMSO- d_6) δ 10.04 (s, 1 H), 9.93 (s, 1 H), 9.74 (br s, 3 H), 9.35 (br s, 1 H), 8.17 (t, 1 H, $J = 5.8$ Hz), 7.23 (d, 1 H, $J = 1.7$ Hz), 7.15 (d, 1 H, $J = 1.7$ Hz), 7.11 (d, 1 H, $J = 1.8$ Hz), 7.05 (d, 1 H, $J = 1.8$ Hz), 6.94 (d, 2 H, $J = 1.7$ Hz), 3.88 (s, 3 H), 3.84 (s, 3 H), 3.80 (s, 3 H), 3.24 (m, 2 H), 3.06 (m, 2 H), 2.78 (s, 6

H), 1.82 (m, 2 H); FAB-HRMS m/e 469.2642 (M+ H (- 2 TFA), 469.5712 calcd. for $C_{23}H_{32}N_8O_3$).

1-Methylimidazole-2-carboxamide-4-tri-*N*-methylpyrrole-2-(3-dimethylaminopropyl)-carboxamide trifluoroacetate (Im-(Py)₃-Dp) (2). To a solution 1-methylimidazole-2-carboxylic acid (35 mg, 0.277 mmol) and DIEA (48.2 mg, 0.373 mmol) in DMSO/NMP (0.4 mL) was added HBTU (103 mg, 0.272 mmol). The resulting solution was stirred 1 h at room temperature when 4-amino-tri-*N*-methylpyrrole-2-(3-dimethylaminopropyl)-carboxamide trifluoroacetate **11** (16 mg, 0.017 mmol), DIEA (30 mg, 0.23 mmol), and DMSO/NMP (0.2 mL) were added. The solution was shaken at 37 °C for 4 h. Purification by RP-HPLC afforded Im-(Py)₃-Dp (**2**) as the trifluoroacetate salt (5.55 mg, 48%). UV (H₂O) λ_{max} (ϵ) 246 (22,900), 314 (31400) nm; ¹H NMR (DMSO-*d*₆) δ 10.51 (s, 1 H), 9.98 (s, 1 H), 9.93 (s, 1 H), 9.26 (br s, 1 H), 8.17 (t, 1 H, J = 4.7 Hz), 7.41 (s, 1 H), 7.29 (d, 1 H, J = 1.5 Hz), 7.23 (d, 1 H, J = 1.5 Hz), 7.17 (s, 2 H), 7.06 (s, 2 H), 6.94 (d, 2 H, J = 1.6 Hz), 3.99 (s, 3 H), 3.84 (s, 3 H), 3.84 (s, 3 H), 3.80 (s, 3 H), 3.23 (m, 2 H), 3.06 (m, 2 H), 2.78 (d, 6 H, J = 4.8 Hz), 1.82 (m, 2 H); FAB-HRMS m/e 577.3330 (M + H (- TFA), 577.6716 calcd. for $C_{28}H_{37}N_{10}O_4$).

4-*tert*-Butoxycarbonylamino-bis-*N*-methylpyrrole-2-carboxylic acid (12). To a solution of 4-amino-bis-*N*-methylpyrrole-2-carboxylic acid methyl ester hydrochloride **7** (5.0 g, 16 mmol) and DIEA (16.5 g, 8 mmol) in DMF (20 mL) was added (Boc₂)O (17.5 g, 80.2 mmol). After stirring 2.5 h at room temperature the solution was added slowly to 400 mL H₂O. The resulting light brown powder was isolated by vacuum filtration, washed with cold H₂O, dried *in vacuo*, to afford the corresponding 4-*tert*-butoxycarbonylamino-bis-*N*-methylpyrrole-2-carboxylic acid methyl ester (5.0 g, 83%) which was used immediately. The ester (5.0 g, 13.3 mmol) was slurried in 95% EtOH (250 mL) and 1 M NaOH (125 mL) was added. The mixture heated at 60 °C for 4 h,

cooled to room temperature, and extracted with ether (100 mL). The pH of the solution was adjusted to 2 with 1 M KHSO₄ and the resulting tan solid was collected by vacuum filtration, washed with cold H₂O, and dried *in vacuo* to carboxylic acid **12** (4.38 g, 91%) as a tan solid. ¹H NMR (DMSO-*d*₆) δ 9.92 (s, 1 H), 9.10 (s, 1 H), 7.40 (s, 1 H), 6.87 (s, 1 H), 6.81 (s, 2 H), 3.80 (s, 3 H), 3.78 (s, 3 H), 1.44 (s, 9 H); FAB-HRMS *m/e* 362.1593 (M⁺, 362.3880 calcd. for C₁₇H₂₂N₄O₅).

4-tert-Butoxycarbonylamino-tetra-N-methylpyrrole-2-carboxylic acid methyl ester (13). To a solution of 4-*tert*-butoxycarbonylamino-bis-*N*-methylpyrrole-2-carboxylic acid **12** (637 mg, 1.76 mmol) and DIEA (0.3 mL, 1.7 mmol) in DMF (4 mL) was added HBTU (670 mg, 1.77 mmol). After stirring 30 min at room temperature 4-amino-bis-*N*-methylpyrrole-2-carboxylic acid methyl ester hydrochloride **7** (500 mg, 1.60 mmol), DIEA (1.11 g, 8.61 mmol) and DMF (2 mL) were added. The solution was stirred 6 h at room temperature and poured into 400 mL cold H₂O and the resulting precipitate was collected by vacuum filtration. Flash column chromatography (5% CH₃OH/CH₂Cl₂) afforded **13** (620 mg, 63%) as a brown solid. ¹H NMR (DMSO-*d*₆) δ 9.94 (s, 2 H), 9.87 (s, 1 H), 9.10 (s, 1 H), 7.47 (d, 1 H, *J* = 1.8 Hz), 7.23 (m, 2 H), 7.06 (m, 2 H), 6.90 (d, 1 H, *J* = 1.9 Hz), 6.89 (d, 1 H, *J* = 0.6 Hz), 6.83 (s, 1 H), 3.84 (s, 3 H), 3.83 (s, 3 H), 3.80 (s, 3 H), 3.73 (s, 3 H), 1.45 (s, 9 H); FAB-HRMS *m/e* 620.8864 (M⁺, 620.6698 calcd. for C₃₀H₃₆N₈O₇).

4-tert-Butoxycarbonylamino-tetra-N-methylpyrrole-2-carboxylic acid (14). To a slurry of 4-*tert*-butoxycarbonylamino-tetra-*N*-methylpyrrole-2-carboxylic acid methyl ester **13** (615 mg, 0.903 mmol) in 95% EtOH (250 mL) was added 1 M NaOH (125 mL) and the resulting mixture was heated at 60 °C for 4 h. The solution was cooled to room temperature and ethanol was removed under reduced pressure before acidifying to pH 2 with 2 M KHSO₄. The precipitate was collected by vacuum filtration, washed with cold

H₂O, and dried *in vacuo* to afford carboxylic acid **14** (600 mg, 99%) as a tan powder. ¹H NMR (DMSO-*d*₆) δ 9.94 (s, 1 H), 9.91 (s, 1 H), 9.87 (s, 1 H), 9.10 (s, 1 H), 7.42 (d, 1 H, *J* = 1.4 Hz), 7.23 (s, 2 H), 7.05 (s, 2 H), 6.89 (s, 1 H), 6.83 (s, 2 H), 3.84 (s, 6 H), 3.81 (s, 3 H), 3.80 (s, 3 H), 1.45 (s, 9 H); FAB-HRMS *m/e* 606.9305 (M⁺, 606.6428 calcd. for C₂₉H₃₄N₈O₇).

4-Amino-tetra-*N*-methylpyrrole-2-(3-dimethylaminopropyl)-carboxamide bistrifluoroacetate (15). To a solution of 4-*tert*-butoxycarbonylamino-tetra-*N*-methylpyrrole-2-carboxylic acid **14** (50 mg, 0.0824 mmol) and DIEA (14.8 mg, 0.115 mmol) in DMF (0.12 mL) was added HBTU (31 mg, 0.0817 mmol). The resulting solution was stirred 30 min at room temperature when 3-dimethylaminopropylamine (49 mg, 0.477 mmol) was added to afford the corresponding carboxamide which was used immediately without isolation. After stirring 4 h at room temperature TFA (10 mL, 130 mmol) was added and the solution stirred 1 h at room temperature. Excess TFA was removed under reduced pressure and the resulting residue was purified by RP-HPLC to afford **15** (25 mg, 37%) as a yellow solid. ¹H NMR (DMSO-*d*₆) δ 10.05 (s, 1 H), 9.96 (s, 1 H), 9.95 (s, 1 H), 9.68 (br s, 3 H), 9.23 (br s, 1 H), 8.18 (t, 1 H, *J* = 4.6 Hz), 7.27 (d, 1 H, *J* = 1.7 Hz), 7.21 (d, 1 H, *J* = 1.7 Hz), 7.17 (d, 1 H, *J* = 1.7 Hz), 7.12 (d, 1 H, *J* = 1.7 Hz), 7.08 (m, 2 H), 6.95 (m, 2 H), 3.89 (s, 3 H), 3.86 (s, 3 H), 3.84 (s, 3 H), 3.81 (s, 3 H), 3.21 (m, 2 H), 3.08 (m, 2 H), 2.78 (s, 6 H), 1.81 (m, 2 H); FAB-HRMS *m/e* 590.3076 (M⁺ (- 2 TFA), 590.6906 calcd. for C₂₉H₃₈N₁₀O₄).

1-Methylimidazole-2-carboxamide-4-tetra-*N*-methylpyrrole-2-(3-dimethylaminopropyl)-carboxamide trifluoroacetate (Im-(Py)₄-Dp) (3). To a solution 1-methylimidazole-2-carboxylic acid (35 mg, 0.277 mmol) and DIEA (48.2 mg, 0.373 mmol) in DMSO/NMP (0.4 mL) was added HBTU (103 mg, 0.272 mmol). The resulting solution was stirred 1 h at room temperature when 4-amino-tetra-*N*-

methylpyrrole-2-(3-dimethylaminopropyl)-carboxamide trifluoroacetate **15** (16 mg, 0.017 mmol), DIEA (30 mg, 0.23 mmol), and DMSO/NMP (0.3 mL) were added. The solution was shaken at 37 °C for 4 h. Purification by RP-HPLC afforded Im-(Py)₄-Dp (**3**) as the trifluoroacetate salt (1.22 mg, 9%). UV (H₂O) λ_{max} (ε) 246 (28,500), 314 (39,100) nm; ¹H NMR (DMSO-*d*₆) δ 10.51 (s, 1 H), 9.99 (s, 1 H), 9.96 (s, 1 H), 9.93 (s, 1 H), 9.25 (br s, 1 H), 8.17 (t, 1 H, *J* = 5.8 Hz), 7.41 (s, 1 H), 7.31 (s, 1 H), 7.23 (s, 2 H), 7.17 (s, 2 H), 7.08 (m, 3 H), 6.95 (s, 1 H), 3.99 (s, 3 H), 3.85 (s, 3 H), 3.84 (s, 3 H), 3.80 (s, 3 H), 3.24 (m, 2 H), 3.07 (m, 2 H), 2.78 (d, 6 H, *J* = 4.8 Hz); FAB-HRMS *m/e* 698.3387 (M⁺ (- TFA), 698.7910 calcd. for C₃₄H₄₂N₁₂O₅).

4-tert-Butoxycarbonylamino-penta-*N*-methylpyrrole-2-carboxylic acid methyl ester (16). To a solution of 4-*tert*-butoxycarbonylamino-bis-*N*-methylpyrrole-2-carboxylic acid **12** (400 mg, 1.1 mmol) and DIEA (171 mg, 1.32 mmol) in DMF (3 mL) HBTU (450 mg, 1.2 mmol) was added and the resulting solution was allowed to stir at room temperature for 30 min. Separately, 4-*tert*-butoxycarbonylamino-tri-*N*-methyl-pyrrole-2-carboxylic acid methyl ester **8** (499 mg, 1.0 mmol) was treated with TFA (10 mL, 130 mmol) at room temperature for 30 min. Excess TFA was removed under reduced pressure and the corresponding amine trifluoroacetate was dissolved in DMF (2 mL) and added to the activated acid along with additional DIEA (2 mL, 8.61 mmol). After stirring overnight at room temperature the solution was added slowly to ether (100 mL) and the resulting precipitate was collected by vacuum filtration. Flash column chromatography (5% CH₃OH/CH₂Cl₂) afforded **16** (622 mg, 84%) as a brown solid. ¹H NMR (DMSO-*d*₆) δ 9.98 (s, 1 H), 9.96 (s, 3 H), 9.89 (s, 1 H), 9.11 (s, 1 H), 7.47 (s, 2 H), 7.24 (s, 2 H), 7.22 (s, 1 H), 7.06 (s, 2 H), 6.90 (m, 2 H), 6.84 (s, 1 H), 3.84 (br s, 9 H), 3.83 (s, 3 H), 3.81 (s, 3 H), 3.73 (s, 3 H), 1.45 (s, 9 H); FAB-HRMS *m/e* 743.1017 (M⁺, 742.7972 calcd. for C₃₆H₄₂N₁₀O₈).

4-tert-Butoxycarbonylamino-penta-N-methylpyrrole-2-carboxylic acid (17).

To a slurry of 4-tert-butoxycarbonylamino-penta-N-methylpyrrole-2-carboxylic acid methyl ester **16** (622 mg, 0.837 mmol) in 95% EtOH (20 mL) was added 1 M NaOH (20 mL) and the resulting mixture was heated at 60 °C for 6 h. The solution was cooled to room temperature and ethanol was removed under reduced pressure before acidifying to pH 2 with 2 M KHSO₄. The solid was collected by vacuum filtration, washed with cold H₂O, and dried *in vacuo* to afford carboxylic acid **17** (560 mg, 93%) as a tan powder. ¹H NMR (DMSO-*d*₆) δ 9.96 (s, 2 H), 9.91 (s, 1 H), 9.88 (s, 1 H), 9.11 (s, 1 H), 7.43 (s, 1 H), 7.24 (m, 3 H), 7.04 (m, 3 H), 6.88 (s, 1 H), 6.84 (s, 2 H), 3.84 (br s, 9 H), 3.81 (br s, 6 H), 1.45 (s, 9 H); FAB-HRMS *m/e* 729.1034 (M⁺, 728.7702 calcd. for C₃₅H₄₀N₁₀O₈).

4-Amino-penta-N-methylpyrrole-2-(3-dimethylaminopropyl)-carboxamide bistrifluoroacetate (18). To a solution of 4-tert-butoxycarbonylamino-penta-N-methylpyrrole-2-carboxylic acid **17** (50 mg, 0.0686 mmol) and DIEA (14.8 mg, 0.115 mmol) in DMSO/NMP (0.4 mL) was added HBTU (26 mg, 0.0686 mmol). The resulting solution was stirred 1 h at room temperature when 3-dimethylaminopropylamine (41 mg, 0.397 mmol) and DIEA (30 mg, 0.23 mmol) were added to afford the corresponding carboxamide which was used immediately without isolation. After stirring 4 h at room temperature TFA (10 mL, 130 mmol) was added and the solution stirred 1 h at room temperature. Excess TFA was removed under reduced pressure and the resulting residue was purified by RP-HPLC to afford bistrifluoroacetate **18** (30 mg, 46%) as a yellow solid. ¹H NMR (DMSO-*d*₆) δ 10.05 (s, 1 H), 9.98 (s, 1 H), 9.96 (s, 1 H), 9.92 (s, 1 H), 9.70 (br s, 1 H), 9.47 (br s, 1 H), 8.17 (t, 1 H, *J* = 5.6 Hz), 7.28 (s, 1 H), 7.24 (s, 1 H), 7.22 (s, 1 H), 7.17 (s, 1 H), 7.13 (s, 1 H), 7.08 (m, 2 H), 6.92 (s, 3 H), 3.89 (s, 3 H), 3.86 (s, 3 H), 3.84 (s, 3 H), 3.81 (s, 3 H), 3.22 (m, 2 H), 3.18 (m, 2 H), 2.78 (s, 6 H),

1.84 (m, 2 H); FAB-HRMS m/e 713.3633 (M^+ (-2 TFA), 712.8180 calcd. for $C_{35}H_{44}N_{12}O_5$).

1-Methylimidazole-2-carboxamide-4-penta-*N*-methylpyrrole-2-(3-dimethylaminopropyl)-carboxamide trifluoroacetate (Im-(Py)₅-Dp) (4). To a solution 1-methylimidazole-2-carboxylic acid (35 mg, 0.277 mmol) and DIEA (48.2 mg, 0.373 mmol) in DMSO/NMP (0.4 mL) was added HBTU (103 mg, 0.272 mmol). The resulting solution was stirred 1 h at room temperature when 4-amino-penta-*N*-methylpyrrole-2-(3-dimethylaminopropyl)-carboxamide trifluoroacetate **18** (16 mg, 0.017 mmol), DIEA (30 mg, 0.23 mmol), and DMSO/NMP (0.3 mL) were added. The solution was shaken at 37 °C for 4 h. Purification by RP-HPLC afforded Im-(Py)₅-Dp (**4**) as the trifluoroacetate salt (2.0 mg, 13%). UV (H_2O) λ_{max} (ϵ) 246 (34,100), 314 (50,900) nm; 1H NMR (DMSO- d_6) δ 10.51 (s, 1 H), 9.98 (s, 1 H), 9.96 (s, 2 H), 9.91 (s, 1 H), 9.20 (br s, 1 H), 8.17 (t, 1 H, $J = 5.2$ Hz), 7.40 (s, 1 H), 7.29 (d, 1 H, $J = 1.6$ Hz), 7.24 (m, 3 H), 7.18 (s, 1 H), 7.17 (s, 1 H), 7.07 (m, 3 H), 7.04 (d, 1 H, $J = 1.6$ Hz), 6.94 (d, 1 H, $J = 1.5$ Hz), 3.99 (s, 3 H), 3.85 (s, 3 H), 3.84 (s, 3 H), 3.80 (s, 3 H), 3.08 (m, 2 H), 2.78 (d, 6 H, $J = 4.8$ Hz), 1.80 (m, 2 H); FAB-HRMS m/e 821.3993 (M^+ (- TFA), 820.9184 calcd. for $C_{40}H_{48}N_{14}O_6$).

4-*tert*-Butoxycarbonylamino-hexa-*N*-methylpyrrole-2-(3-dimethylamino-propyl)-carboxamide (19). To a solution of 4-*tert*-butoxycarbonylamino-tri-*N*-methylpyrrole-2-carboxylic acid **9** (1.037 g, 1.82 mmol) and DIEA (297 mg, 2.30 mmol) in DMF (6 mL) was added HBTU (760 mg, 2.0 mmol). After stirring 4 h at room temperature 4-amino-tri-*N*-methylpyrrole-2-(3-dimethylaminopropyl)-carboxamide trifluoroacetate **11** (1.27 g, 1.82 mmol), DIEA (2.23 g, 17.2 mmol) and DMF (2 mL) were added. The solution was stirred overnight at room temperature and poured into ether (400 mL) and the resulting precipitate was collected by vacuum filtration and washed with

cold ether. Flash column chromatography (1:1 CH₃OH/CH₂Cl₂ to CH₃OH to 1% NH₄OH/CH₃OH) afforded **19** (1.06 g, 63%) as a yellow solid. ¹H NMR (DMSO-*d*₆) δ 9.96 (s, 4 H), 9.91 (s, 1 H), 9.88 (s, 1 H), 9.11 (s, 1 H), 8.10 (t, 1 H, *J* = 5.0 Hz), 7.71 (d, 1 H, *J* = 6.2 Hz), 7.48 (d, 1 H, *J* = 4.6 Hz), 7.22 (m, 5 H), 7.07 (m, 3 H), 6.89 (s, 1 H), 6.85 (s, 1 H), 3.82 (m, 18 H), 3.22 (m, 2 H), 2.31 (s, 6 H), 1.65 (m, 2 H), 1.44 (s, 9 H), 1.30 (m, 2 H); FAB-HRMS *m/e* 935.4677 (M⁺, 935.0632 calcd. for C₄₆H₅₈N₁₄O₈).

4-Amino-hexa-*N*-methylpyrrole-2-(3-dimethylaminopropyl)-carboxamide bistrifluoroacetate (20). To a slurry of 4-*tert*-butoxycarbonylamino-hexa-*N*-methylpyrrole-2-(3-dimethylaminopropyl)-carboxamide **19** (500 mg, 0.542 mmol) was added TFA (20 mL, 260 mmol) and the resulting solution was allowed to stir 1.5 h at room temperature. Excess TFA was removed under reduced pressure and the resulting solid slurried in H₂O (5 mL), frozen, and lyophilized to afford **20** (400 mg, 69%) as a yellow solid. ¹H NMR (DMSO-*d*₆) δ 10.05 (s, 1 H), 9.96 (s, 3 H), 9.91 (s, 1 H), 9.76 (br s, 3 H), 9.30 br s (1 H), 8.17 (t, 1 H, *J* = 5.8 Hz), 7.24 (d, 1 H, *J* = 1.8 Hz), 7.23 (s, 3 H), 7.16 (d, 1 H, *J* = 1.8 Hz), 7.12 (d, 1 H, *J* = 1.8 Hz), 7.08 (m, 4 H), 6.95 (s, 2 H), 3.90 (s, 3 H), 3.86 (s, 12 H), 3.84 (s, 3 H), 3.81 (s, 3 H), 3.22 (m, 2 H), 3.05 (m, 2 H), 2.78 (d, 6 H, *J* = 4.8 Hz), 1.82 (m, 2 H); FAB-HRMS *m/e* 835.4131 (M⁺ (- 2 TFA), 834.9454 calcd. for C₄₁H₅₀N₁₄O₆).

1-Methylimidazole-2-carboxamide-4-hexa-*N*-methylpyrrole-2-(3-dimethylaminopropyl)-carboxamide trifluoroacetate (Im-(Py)₆-Dp) (5). To a solution of *N*-methyl-imidazole-2-carboxylic acid (100 mg, 741 μmol) and hydroxybenzotriazole (72 mg, 500 μmol) in DMF (0.5 mL) was added DCC. The solution was allowed to stand 12 h at room temperature and then filtered to remove precipitated DCU. To the solution was added 4-amino-hexa-*N*-methylpyrrole-2-(3-dimethylaminopropyl)-carboxamide as the bistrifluoroacetate salt **20** (10 mg, 9.4 μmol) and DIEA (100 μl, 576 μmol) and the

resulting mixture was allowed to stand 2 h at room temperature. Purification by RP-HPLC afforded Im-(Py)₆-Dp (**5**) (6.3 mg, 62%) as the trifluoroacetate salt. UV (H₂O) λ_{\max} (ϵ), 246 (34,100), 304 (56,600) nm; ¹H NMR (DMSO-*d*₆) δ 10.46 (s, 1 H), 9.55 (s, 1 H), 9.94 (m, 3 H), 9.90 (s, 1 H), 9.20 (br s, 1 H), 8.14 (t, 1 H, *J* = 7.2 Hz), 7.38 (s, 1 H), 7.28 (d, 1 H, *J* = 1.4 Hz), 7.26 (d, 1 H, *J* = 1.4 Hz), 7.23 (m, 4 H), 7.08 (m, 5 H), 7.04 (s, 1 H, *J* = 1.2 Hz), 6.93 (d, 1 H, *J* = 1.6 Hz), 3.98 (s, 3 H), 3.84 (m, 15 H), 3.83 (s, 1 H), 3.30 (q, 2 H, *J* = 7.4 Hz), 3.21 (t, 2 H, *J* = 7.1 Hz), 2.77 (d, 6 H, *J* = 4.1 Hz), 1.74 (m, 2 H); FAB-HRMS *m/e* 965.430 (M + Na (- TFA), 965.426 calcd. for C₄₆H₅₄N₁₆O₇Na); MALDI-TOF MS 944.212 (M+H (- TFA), 944.04 calcd. for C₄₆H₅₅N₁₆O₇).

4-Amino-septa-N-methylpyrrole-2-(3-dimethylaminopropyl)-carboxamide bistrifluoroacetate (21). To a solution of 4-amino-hexa-*N*-methylpyrrole-2-(3-dimethylaminopropyl)-carboxamide bistrifluoroacetate **20** (20 mg, 0.0214 mmol) and DIEA (37.1 mg, 0.287 mmol) in DMSO/NMP (0.35 mL) was added 4-*tert*-butoxycarbonylamino-*N*-methylpyrrole-2-carboxylic acid hydroxybenzotriazole ester (16 mg, 0.43 mmol). The solution was shaken at 37 °C for 4 h and then diluted to 2 mL with DMF. The resulting precipitate was isolated by centrifugation, washed with cold DMF (2 x 250 μ L), slurried in CH₂Cl₂ (0.4 mL) and treated with TFA (8 mL, 104 mmol). After 1.5 h at room temperature excess TFA was removed under reduced pressure and the resulting residue purified by RP-HPLC to afford **21** (20 mg, 79%). ¹H NMR (DMSO-*d*₆) δ 10.05 (s, 1 H), 9.98 (s, 5 H), 9.95 (s, 1 H), 9.82 (br s, 3 H), 9.42 (br s, 1 H), 8.17 (t, 1 H, *J* = 5.8 Hz), 7.24 (m, 6 H), 7.15 (d, 1 H, *J* = 1.8 Hz), 7.10 (m, 6 H), 6.97 (d, 1 H, *J* = 1.8 Hz), 3.98 (s, 3 H), 3.85 (s, 12 H), 3.80 (s, 3 H), 3.24 (m, 2 H), 3.11 (m, 2 H), 2.78 (d, 6 H, *J* = 4.8 Hz), 1.82 (m, 2 H); FAB-HRMS *m/e* 957.2256 (M+ (-2 TFA), 957.0728 calcd. for C₄₇H₅₆N₁₆O₇).

1-Methylimidazole-2-carboxamide-4-septa-*N*-methylpyrrole-2-(3-dimethylaminopropyl)-carboxamide trifluoroacetate (Im-(Py)₇-Dp) (6). To a solution of 1-methylimidazole-2-carboxylic acid (35 mg, 0.277 mmol) and DIEA (48.2 mg, 0.373 mmol) in DMSO/NMP (0.4 mL) was added HBTU (103 mg, 0.272 mmol). The resulting solution was stirred 1 h at room temperature when 4-amino-septa-*N*-methylpyrrole-2-(3-dimethylaminopropyl)-carboxamide bistrifluoroacetate **21** (20 mg, 0.017 mmol), DIEA (30 mg, 0.23 mmol), and DMSO/NMP (0.3 mL) were added. The solution was shaken at 37 °C for 4 h. Purification by RP-HPLC Im-(Py)₇-Dp (**6**) as the trifluoroacetate salt (6.1 mg, 31%). UV (H₂O) λ_{max} (ε) 246 (38,600), 304 (62,500) nm; ¹H NMR (DMSO-*d*₆) δ 10.49 (s, 1 H), 9.98 (s, 1 H), 9.97 (s, 5 H), 9.92 (s, 1 H), 9.24 (br s, 1 H), 8.17 (t, 1 H, *J* = 5.2 Hz), 7.40 (s, 1 H), 7.29 (d, 1 H, *J* = 1.8 Hz), 7.27 (s, 1 H), 7.24 (s, 4 H), 7.17 (s, 2 H), 7.08 (s, 4 H), 7.05 (s, 1 H), 6.94 (m, 2 H), 3.99 (s, 3 H), 3.84 (s, 15 H), 3.83 (s, 3 H), 3.80 (s, 3 H), 3.22 (m, 2 H), 3.08 (m, 2 H), 2.78 (d, 6 H, *J* = 4.8 Hz), 1.82 (m, 2 H); MALDI-TOF MS 1066.1 (M + H (- TFA), 1066.2 calcd. for C₅₂H₆₁N₁₈O₈).

DNA Reagents and Materials. RNase free water was obtained from U.S. Biochemicals and used in all manipulations. *Eco* RI, *Bam* HI, *Hind* III, *Pvu* II, the Klenow fragment of DNA polymerase I, calf alkaline phosphatase, and T4 polynucleotide kinase were purchased from Boehringer-Mannheim and used according to the supplier's recommended protocol in the buffers provided. T4 DNA ligase was obtained from New England Biolabs and used with the buffer provided. Nucleotide triphosphates were Pharmacia Ultra-Pure grade and were used as supplied. Nucleotide triphosphates labeled with ³²P (> 3000 Ci/mmol) were obtained from Amersham. NICK columns were purchased from Pharmacia and used according to the suggested protocol. Glycogen was obtained from Boehringer-Mannheim as a 20 mg/mL aqueous solution. Cerenkov

radioactivity was measured with a Beckman LS 2801 scintillation counter. DNase I was purchased from Boehringer-Mannheim. Sonicated, deproteinized calf thymus DNA was purchased from Pharmacia and dissolved in H₂O to a final concentration of 2 mM in base pairs and stored at -20 °C. Chemical sequencing reactions were performed according to published procedures.²⁷⁻²⁸ General manipulations of DNA were performed according to established procedures.²⁹

Plasmid Preparation. Plasmids pJK5, pJK6, pJK7, pJK8, pJK9, and pJK10 were prepared by established procedures.²⁸ Briefly, oligonucleotides containing the desired insert sequences were synthesized, deprotected, and purified by 12-15% denaturing polyacrylamide gel electrophoresis (5% cross-linked, 7 M urea). The oligonucleotides were then annealed by slow-cooling from 90 °C to afford the desired insert duplex. The oligonucleotide duplex (1 pmol) was then added to *Bam* HI/*Hind* III restricted pUC19 plasmid DNA and the resulting solution was treated overnight with T4 DNA ligase. The ligation mixture was then used to transform Epicurean Coli XL1-Blue Supercompetent cells (Stratagene) according to the manufacturer's protocol. Transformants containing an insert were identified by α -complementation, selected and grown overnight in 5 mL LB media containing 50 mg/mL ampicillin. The plasmids were isolated using a Quiagen miniprep kit (Tip 20) and sequenced using a Sequenase 2.0 kit (U.S. Biochemicals). Large scale plasmid preparation was accomplished using a Quiagen maxiprep kit (Tip 500). The sequences of the isolated plasmids were confirmed by sequencing. Plasmid yields were determined at 260 nm using the relation 1 OD unit = 50 μ g/mL duplex DNA. Plasmids were stored in 1 mL 10 mM Tris•HCl, pH 7.0, 0.5 mM EDTA at -20 C.

Labeled Restriction Fragment Preparations. The *Eco* RI/*Pvu* II fragments from plasmids pJK5, pJK6, pJK7, pJK8, pJK9, and pJK10 3'-³²P-end labeled at the *Eco* RI site were prepared by digesting each plasmid (3-5 μ g) with *Eco* RI, followed by

treatment with α - 32 P-dATP and α - 32 P-TTP in the presence of the Klenow fragment of DNA polymerase I. Following the addition of unlabeled dATP and TTP unincorporated nucleotides were removed using a NICK column. The linearized, 32 P-labeled plasmid was then digested with *Pvu* II, followed by separation of the resulting DNA fragments on a non-denaturing 5% polyacrylamide gel (1:19 crosslinked). The desired fragment was located by autoradiography and excised. The gel slice was crushed and soaked in 25 mM Tris-HCl pH 7.0, 2 mM EDTA for 8-12 h at 37 °C. The resulting suspension was then filtered through an Isolab Quik-Sep polypropylene filter and the filtrate was ethanol precipitated using glycogen (1 μ L). The DNA pellet was washed with 70% EtOH, dried *in vacuo*, and resuspended in 10 mM Tris-HCl pH 7.0, 0.5 mM EDTA. The solution was then extracted with phenol (3 x 100 μ L), 24:1 phenol:CHCl₃ (3 x 100 μ L), and 24:1 CHCl₃:isoamyl alcohol (2 x 100 μ L). The solution was then passed through a NICK column equilibrated with 10 mM Tris-HCl pH 7.0, 0.5 mM EDTA buffer and the eluate stored at -20 °C after measuring Cerenkov radioactivity. Typical yields of Cerenkov radioactivity were 5-10 x 10⁶ cpm. The restriction fragments 5'- 32 P-end labeled at the *Eco* RI site was prepared by digesting the corresponding plasmids (5 μ g) with *Eco* RI followed by dephosphorylation with calf alkaline phosphatase. The linearized plasmid was then labeled by treatment with γ - 32 P-ATP in the presence of polynucleotide kinase. Unincorporated γ - 32 P-ATP was removed using a NICK column. The DNA was then digested with *Pvu* II and isolated as described above. Typical yields of radioactivity were 2-5 x 10⁶ cpm.

MPE•Fe(II) Footprinting Experiments. For each experiment a cocktail containing Tris-AcOH pH 7.0, NaCl, calf thymus DNA, and 32 P-end labeled DNA restriction fragment was prepared. To separate 1.7 mL microfuge tubes were added 28 μ L of cocktail. Stock solutions of each polyamide were prepared and the concentrations were

verified by the UV absorbance at 316 nm before making serial dilutions. A 4 μL aliquot of each 10x ligand solution was added to the appropriate tubes and the solutions were allowed to equilibrate for 1-6 h at 22 $^{\circ}\text{C}$. A 50 μM solution of MPE $\cdot\text{Fe(II)}$ was prepared by mixing 100 μL of a 100 μM MPE solution with a freshly prepared 100 μM ferrous ammonium sulfate solution and allowing to equilibrate for 15 min at 22 $^{\circ}\text{C}$. A 4 μL aliquot of 50 μM MPE $\cdot\text{Fe(II)}$ was added to the appropriate tubes and incubated 15 min at 22 $^{\circ}\text{C}$. Cleavage was initiated by the addition of 4 μL of 50 mM DTT. Final concentrations were 25 mM Tris $\cdot\text{AcOH}$ pH 7.0, 10 mM NaCl, 100 μM -bp calf thymus DNA, 5 μM MPE $\cdot\text{Fe(II)}$, and 5 mM DTT in a total volume of 40 μL containing $1\text{-}3 \times 10^5$ cpm of ^{32}P -labeled restriction fragment. Cleavage was allowed to proceed for 15 min and then stopped by ethanol precipitation using 5 μL of glycogen. The precipitated DNA was washed with 70% EtOH, dried *in vacuo* and resuspended in formamide-TBE loading buffer. After heating at 90 $^{\circ}\text{C}$ for 3 min and cooling on ice the samples were electrophoresed on 0.4 mm thick 8% denaturing polyacrylamide gels (5% cross-link, 7 M urea) at 2000 V until the bromophenol blue tracking dye reached the bottom of the gel. The gels were transferred to filter paper, dried, and quantitated using storage phosphor technology as described below. Histograms were constructed by comparing the cleavage intensity in lanes containing polyamide to an MPE $\cdot\text{Fe(II)}$ control lane to which no ligand was added.

Quantitative DNase I Footprint Titration Experiments. For each experiment a cocktail containing Tris $\cdot\text{HCl}$, pH 7.0, KCl, MgCl_2 , CaCl_2 , and 3'- ^{32}P -end labeled restriction fragment was prepared. To separate 1.7 mL microfuge tubes was added a 32 μL aliquot of the cocktail solution. Stock solutions of each polyamide were prepared and the concentrations were verified by the UV absorbance at 316 nm before making serial dilutions. To the appropriate tubes containing cocktail were added 4 μL of 10x ligand

solution and the solutions were allowed to equilibrate at 22 °C. Final ligand concentrations were 5 nM to 50 μ M for Im-(Py)₂-Dp, 50 pM to 10 μ M for Im-(Py)₃-Dp, 20 pM to 1.5 μ M for Im-(Py)₄-Dp, 1 pM to 0.2 μ M for Im-(Py)₅-Dp, 10 pM to 1 μ M for Im-(Py)₆-Dp, and 1 nM to 20 μ M for Im-(Py)₇-Dp. The equilibration times were 1 h and 3 h for Im-(Py)₂-Dp, 36 h and 65 h for Im-(Py)₃-Dp, 36 h and 65 h for Im-(Py)₄-Dp, 60 h and 84 h for Im-(Py)₅-Dp, 36 h and 65 h for Im-(Py)₆-Dp, and 8 h for Im-(Py)₇-Dp. After equilibration, cleavage was initiated by the addition 4 μ L of a solution of DNase I (0.04 units/mL) containing 1 mM DTT. Final concentrations were 10 mM Tris·HCl pH 7.0, 10 mM KCl, 5 mM MgCl₂, and 5 mM CaCl₂ in a total volume of 40 μ L containing 1-2 x 10⁵ cpm of 3'-³²P-end labeled restriction fragment. We note explicitly that binding affinities were determined in the absence of unlabeled carrier DNA. Cleavage was allowed to proceed for 6 min at 22 °C and then stopped by the addition of 12.8 μ L of a 3 M NaOAc/50 mM EDTA/glycogen solution (1.2/1/1) followed by ethanol precipitation. The precipitated DNA was washed with 70% EtOH, dried *in vacuo* and resuspended in formamide-TBE loading buffer. After heating at 90 °C for 3 min and cooling on ice the samples were electrophoresed on 0.4 mm thick 8% denaturing polyacrylamide gels (5% cross-link, 7 M urea) at 2000 V until the bromophenol blue tracking dye reached the bottom of the gel. The gels were transferred to filter paper, dried, and quantitated using storage phosphor technology as described below.

Apparent first-order binding constants were determined as previously described.¹⁷ The data were analyzed by performing volume integrations of the match and single base pair mismatch sites on each restriction fragment as well as a 5'-GAGC-3' reference site at which DNase I cleavage was invariant. The apparent DNA target site saturation, θ_{app} , was calculated for each concentration of polyamide using the following equation:

$$\theta_{\text{app}} = 1 - \frac{I_{\text{site}}/I_{\text{ref}}}{I_{\text{site}}^{\circ}/I_{\text{ref}}^{\circ}} \quad (1)$$

where I_{tot} and I_{ref} are the integrated volumes of the target and reference sites, respectively, and I_{tot}° and I_{ref}° correspond to those values from a DNase I control lane to which no polyamide has been added. At concentrations higher than 50 μM of Im-(Py)₂-Dp, 5 μM Im-(Py)₃-Dp, 1 μM Im-(Py)₄-Dp, 0.1 μM Im-(Py)₅-Dp and 1 μM Im-(Py)₆-Dp the reference site becomes partially protected, resulting in low θ_{app} values. For this reason, polyamide concentrations higher than these values were not used. The ($[L]_{\text{tot}}$, θ_{app}) data points were best fit to cooperative Langmuir binding isotherms ($n = 2$) consistent with 2:1 polyamide-DNA complex formation by minimizing the difference between θ_{app} and θ_{fit} using the modified Hill equation:

$$\theta_{\text{fit}} = \theta_{\text{min}} + (\theta_{\text{max}} - \theta_{\text{min}}) \frac{K_a^n [L]_{\text{tot}}^n}{1 + K_a^n [L]_{\text{tot}}^n} \quad (2)$$

where $[L]_{\text{tot}}$ corresponds to the total polyamide concentration, K_a corresponds to the apparent monomeric association constant, n represents the Hill coefficient, and θ_{max} and θ_{min} represent the experimentally determined site saturation values when the site is unoccupied or saturated, respectively. We note explicitly that treatment of the data in this manner does not represent an attempt to model a binding mechanism. Rather, we have chosen to compare values of the apparent first-order association constant, because this parameter represents the concentration of polyamide at which the binding site is half-saturated. Data were fit using a non-linear least-squares fitting procedure of Kaleidagraph software (version 3.0.1; Abelbeck Software) running on a Power PC Macintosh 6100/60 computer with K_a , θ_{max} , and θ_{min} as the adjustable parameters. The goodness of fit is evaluated by the correlation coefficient, with $R > 0.98$ as the criterion for an acceptable fit. The binding affinities reported for Im-(Py)₂-Dp at 5'-TGGCA-3', Im-(Py)₅-Dp at 5'-TGTGAACA-3',

and Im-(Py)₇-Dp at the match and mismatch sites represent upper limits since the quality of fits for these data sets were poor. The specificity of each polyamide represents the ratio of the binding affinity of the match side to the affinity for the single base pair mismatch site. The uncertainty of each specificity value was calculated using the standard deviations for the match and mismatch site binding affinities.³⁰ The specificities of Im-(Py)₂-Dp and Im-(Py)₅-Dp represent a lower limit due to the fact that the mismatch site binding affinities are upper limits. At least three sets of acceptable data were used in determining each association constant. All lanes from each gel were used unless visual inspection revealed a data point to be obviously flawed relative to neighboring points. The data were normalized using the following equation:

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

Quantitation by Storage Phosphor Autoradiography. Photostimulable storage phosphor imaging plates (Kodak Storage Phosphor Screen S0230 obtained from Molecular Dynamics) were pressed flat against dried gels and exposed in the dark for 12-24 h at 22 °C. A Molecular Dynamics 400S PhosphorImager was used to obtain all data from the storage screens. The data were analyzed by performing volume integrations of all bands using ImageQuant v. 3.3 software running on a Gateway 2000 486/66 computer.

References and Notes

1. For a review, see: Zimmer, C.; Wahnert, U. *Prog. Biophys. Molec. Biol.* **1986**, *47*, 31.
2. (a) Van Dyke, M. W.; Hertzberg, R. P.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 5470. (b) Van Dyke, M. W.; Dervan, P. B. *Cold Spring Harbor*

- Symposium on Quantitative Biology* **1982**, 47, 347. (c) Van Dyke, M. W.; Dervan, P. B. *Biochemistry* **1983**, 22, 2373. (d) Harshman, K. D.; Dervan, P. B. *Nucl. Acids Res.* **1985**, 13, 4825. (e) Fox, K. R.; Waring, M. J. *Nucl. Acids Res.* **1984**, 12, 9271. (f) Lane, M. J.; Dabrowiak, J. C.; Vournakis, J. *Proc. Natl. Acad. Sci. USA* **1983**, 80, 3260.
3. (a) Schultz, P. G.; Taylor, J. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1982**, 104, 6861. (b) Taylor, J. S.; Schultz, P. G.; Dervan, P. B. *Tetrahedron* **1984**, 40, 457. (c) Schultz, P. G.; Dervan, P. B. *J. Biomol. Struct. Dyn.* **1984**, 1, 1133.
 4. (a) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *Proc. Natl. Acad. Sci. USA* **1985**, 82, 1376. (b) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *J. Mol. Biol.* **1985**, 183, 553. (c) Coll, M.; Frederick, C. A.; Wang, A. H.-J.; Rich, A. *Proc. Natl. Acad. Sci. USA* **1987**, 84, 8385.
 5. (a) Patel, D. J.; Shapiro, L. *J. Biol. Chem.* **1986**, 261, 1230. (b) Klevitt, R. E.; Wemmer, D. E.; Reid, B. R. *Biochemistry* **1986**, 25, 3296. (c) Pelton, J. G., Wemmer, D. E. *Biochemistry* **1988**, 27, 8088.
 6. (a) Markey, L. A.; Breslauer, K. J. *Proc. Natl. Acad. Sci. USA* **1987**, 84, 4359. (b) Breslauer, K. J.; Remeta, D. P.; Chou, W.-Y.; Ferrante, R.; Curry, J.; Zaunczkowski, D.; Snyder, J. G.; Markey, L. A. *Proc. Natl. Acad. Sci. USA* **1987**, 84, 8922.
 7. (a) Dervan, P. B. *Science* **1986**, 232, 464. (b) Youngquist, R. S.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1985**, 82, 2565. (c) Youngquist, R. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1987**, 109, 7564.
 8. For early examples of hybrid molecules for the recognition of mixed sequences, see: (a) Dervan, P. B.; Sluka, J. P. *New Synthetic Methodology and Functionally Interesting Compounds*; Elsevier: New York, 1986; pp 307-322. (b) Griffin, J. H.; Dervan, P. B. *J. Am. Chem. Soc.* **1987**, 109, 6840.

9. (a) Wade, W. S., Dervan, P. B. *J. Am. Chem. Soc.* **1987**, *109*, 1574. (b) Wade, W. S. Ph. D. Thesis, California Institute of Technology, 1989.
10. (a) Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 5723. (b) Pelton, J. G.; Wemmer, D. E. *J. Am. Chem. Soc.* **1990**, *112*, 1393. (c) Chen, X.; Ramakrishnan, B.; Rao, S. T.; Sundaralingam, M. *Struct. Biol. Nat.* **1994**, *1*, 169. (d) Rentzeperis, D.; Marky, L. A.; Dwyer, T. J.; Geierstanger, B. H.; Pelton, J. G.; Wemmer, D. E. *Biochemistry* **1995**, *34*, 2937.
11. Animati, F.; Arcamone, F. M.; Conte, M. R.; Felicetti, P.; Galeone, A.; Lombardi, P.; Mayol, L. G.; Rossi, C. *J. Med. Chem.* **1995**, *38*, 1140.
12. (a) Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 8783. (b) Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 7586. (c) Wade, W. S.; Mrksich, M.; Dervan, P. B. *Biochemistry* **1993**, *32*, 11385.
13. (a) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 2572. (b) Geierstanger, B. H.; Jacobsen, J.-P.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Biochemistry* **1994**, *33*, 3055.
14. (a) Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Science* **1994**, *266*, 646. (b) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1995**, *117*, 3325.
15. Geierstanger, B. H.; Dwyer, T. J.; Bathini, Y.; Lown, J. W.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, *115*, 4474.
16. Fagan, P. A.; Wemmer, D. E. *J. Am. Chem. Soc.* **1992**, *114*, 1080.
17. Im-(Py)₆-Dp was synthesized by E. E. Baird.
18. (a) Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Methods Enzymol.* **1986**, *130*, 132. (b) Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 8462. (c) Senear, D. F.; Brenowitz, M.; Shea, M. A.; Ackers, G. K. *Biochemistry* **1986**, *25*, 7344.
19. Goodsell, D.; Dickerson, R. E. *J. Med. Chem.* **1986**, *29*, 727.

20. We note that in the 2:1 complexes of tetra-*N*-methylpyrrolicarboxamide analogs of distamycin with a 5'-AAATTT-3' site the carboxamide NH's between the second and third and third and fourth pyrrole rings of the polyamides appear to be unable to form hydrogen bonds to the DNA due to the curvature of the ligands, see Ref. 11.
21. The intrinsic curvature of the binding sites was evaluated using the program CURVATUR (v. 1.00, College of Judea and Samaria, The Research Institute, Gene Structure Research Center) which calculates the curvature of a given DNA sequence using the wedge model of DNA bending.²² This analysis suggests that there is no significant curvature of the polyamide binding sites.
22. Bolshoy, A.; McNamara, P.; Harrington, R. E.; Trifonov, E. N. *Proc. Natl. Acad. Sci. USA* **1991**, 88, 2312.
23. Yoon, C.; Prive, G. G.; Goodsell, D. S.; Dickerson, R. E. *Proc. Natl. Acad. Sci., USA* **1988**, 85, 6332.
24. For hairpin polyamides, see Mrksich, M.; Parks, M. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, 116, 7983.
25. For cyclic polyamides, see Cho, J.; Parks, M. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1995**, 92, 10389.
26. Still, W. C.; Kahn, M.; Mitra, A. *J. Org. Chem.* **1978**, 40, 2923.
27. Iverson, B. L.; Dervan, P. B. *Nucl. Acids Res.* **1987**, 15, 7823.
28. Maxam, A. M.; Gilbert, W. S. *Methods in Enzymology* **1980**, 65, 499.
29. Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular Cloning*; Cold Spring Harbor Laboratory: Cold Spring Harbor, NY, 1989.
30. Bevington, P. E.; Robinson, D. K. *Data Reduction and Error Analysis for the Physical Sciences*, Second Edition; McGraw Hill: NY, NY, 1992.

Chapter 3

Recognition of 5'-(A,T)GCGC(A,T)-3' Sequences by a Hairpin Polyamide

Introduction

2:1 Polyamide-DNA Complexes. The discovery that the three ring synthetic polyamide Im-Py-Py-Dp containing imidazole (Im) and pyrrole (Py) carboxamides binds the DNA sequence 5'-(A,T)G(A,T)C(A,T)-3' as an antiparallel dimer offers a new model for the design of ligands for specific recognition of sequences in the minor groove containing both G,C and A,T base pairs.¹⁻⁶ The 2:1 model allows specific contacts with each DNA strand on the floor of the minor groove where the side-by-side combination of an imidazole ring on one ligand and a pyrrole carboxamide on the second ligand recognizes G•C, while a pyrrole carboxamide-imidazole pair targets a C•G base pair. A pyrrole carboxamide-pyrrole carboxamide pair is partially degenerate and binds to either A•T or T•A base pairs.

Recognition of 5'-(A,T)GCGC(A,T) Sequences. Recently, it was demonstrated by footprinting, affinity cleaving, and NMR studies that the alternating four-ring polyamide Im-Py-Im-Py-Dp **1** binds a family of six base pair sites 5'-(A,T)GCGC(A,T)-3' containing a four base pair G,C core as an antiparallel dimer in the minor groove representing a reversal of the A,T specificity of the natural product distamycin (Figure 3.1).⁵ Remarkably, despite the novel specificity of polyamide **1** for 5'-(A,T)GCGC(A,T)-3' sites, the overall affinity was modest ($K_a = 4 \times 10^5 \text{ M}^{-1}$, pH 7.0, 22 °C). This suggests that there is an energetic price for forming four specific hydrogen bonds between the imidazole rings on the polyamide dimer and the guanine amino groups protruding from the

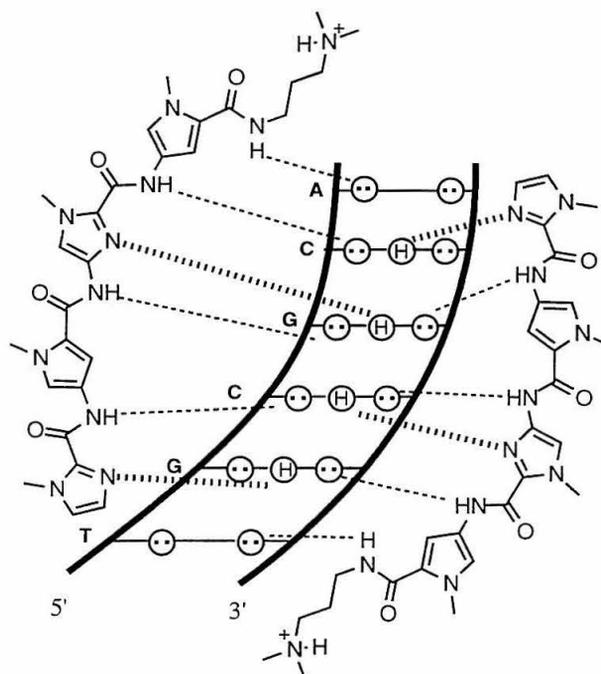


Figure 3.1. Model of the $(\text{Im-Py-Im-Py-Dp})_2 \cdot 5'\text{-TGCGCA-3}'$ complex derived from NMR data. Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines and circles containing a H represent the 2-amino group of guanine. Putative hydrogen bonds are indicated by dashed lines.

floor of the minor groove. It is possible that the dimer of **1** does not sit deeply in the minor groove diminishing energetically favorable van der Waals contacts with the walls of the double helix.

Experimental Design. A hairpin motif in which two crescent-shaped polyamides are linked head-to-tail by γ -aminobutyric acid (GABA, γ) offers a methodology for the preparation of polyamides that bind in the minor groove with enhanced affinity.⁷ The question arises whether hairpin polyamides bind G,C core sequences with high affinity and specificity. Here we examine the energetics of the hairpin polyamide Im-Py-Im-Py- γ -Im-Py-Im-Py-Dp **2** binding at eight different six base pair sites on a 381 base pair DNA fragment containing three putative perfect matches 5'-(A,T)GCGC(A,T)-3' and five sites with one to two mismatches in the central 5'-GCGC-3' core (Figure 3.2).

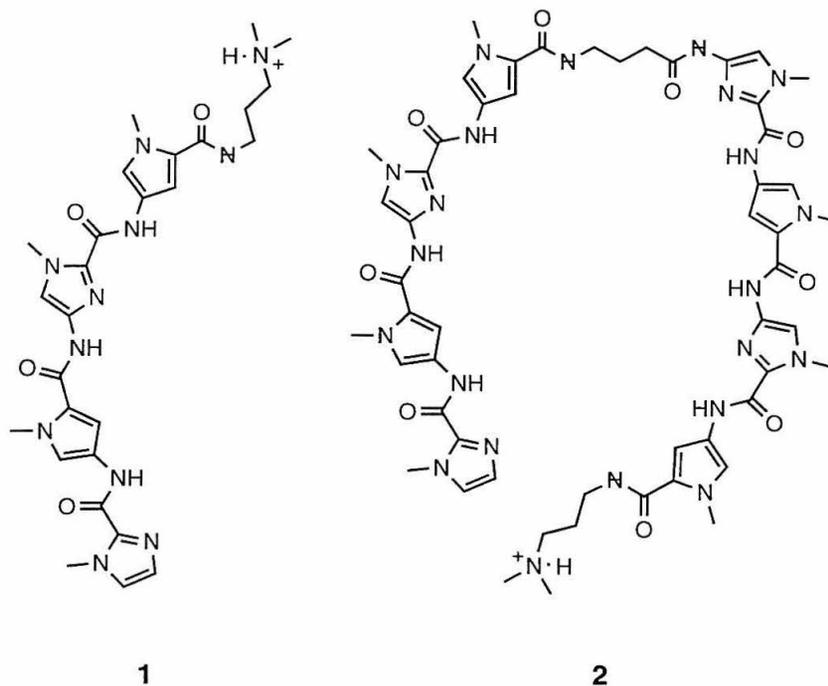


Figure 3.2. Structure of Im-Py-Im-Py-Dp **1** and hairpin polyamide Im-Py-Im-Py- γ -Im-Py-Im-Py-Dp **2**.

Results

Hairpin Polyamide Synthesis. The hairpin polyamide Im-Py-Im-Py- γ -Im-Py-Im-Py-Dp **2** was synthesized in nine steps from methyl 4-nitro-*N*-methylpyrrole-2-carboxylate (**3**) (Figure 3.3). Reduction of **3** followed by coupling with 4-nitro-*N*-methylimidazole-2-acyl chloride affords **4** that was similarly reduced and treated with 4-nitro-*N*-methylpyrrole-2-acyl chloride to afford the three ring polyamide **5**. Reduction of **5** and coupling with 1-methylimidazole-2-carboxylic acid (DCC, HOBT) affords ester **6** which is converted to acid **7** by saponification (LiOH, H₂O, EtOH). Reduction of **5** and coupling with 4-(4-amino-*N*-*t*-butoxycarbonyl-butylamido)-*N*-methylimidazole-2-carboxylic acid (**12**), available in three steps from ethyl-4-nitro-1-methylimidazole-2-carboxylate (**11**) (Figure 3.4), affords polyamide **9** containing the GABA linker in protected form. Saponification (LiOH, H₂O, EtOH) of **9** affords acid **10** that can be

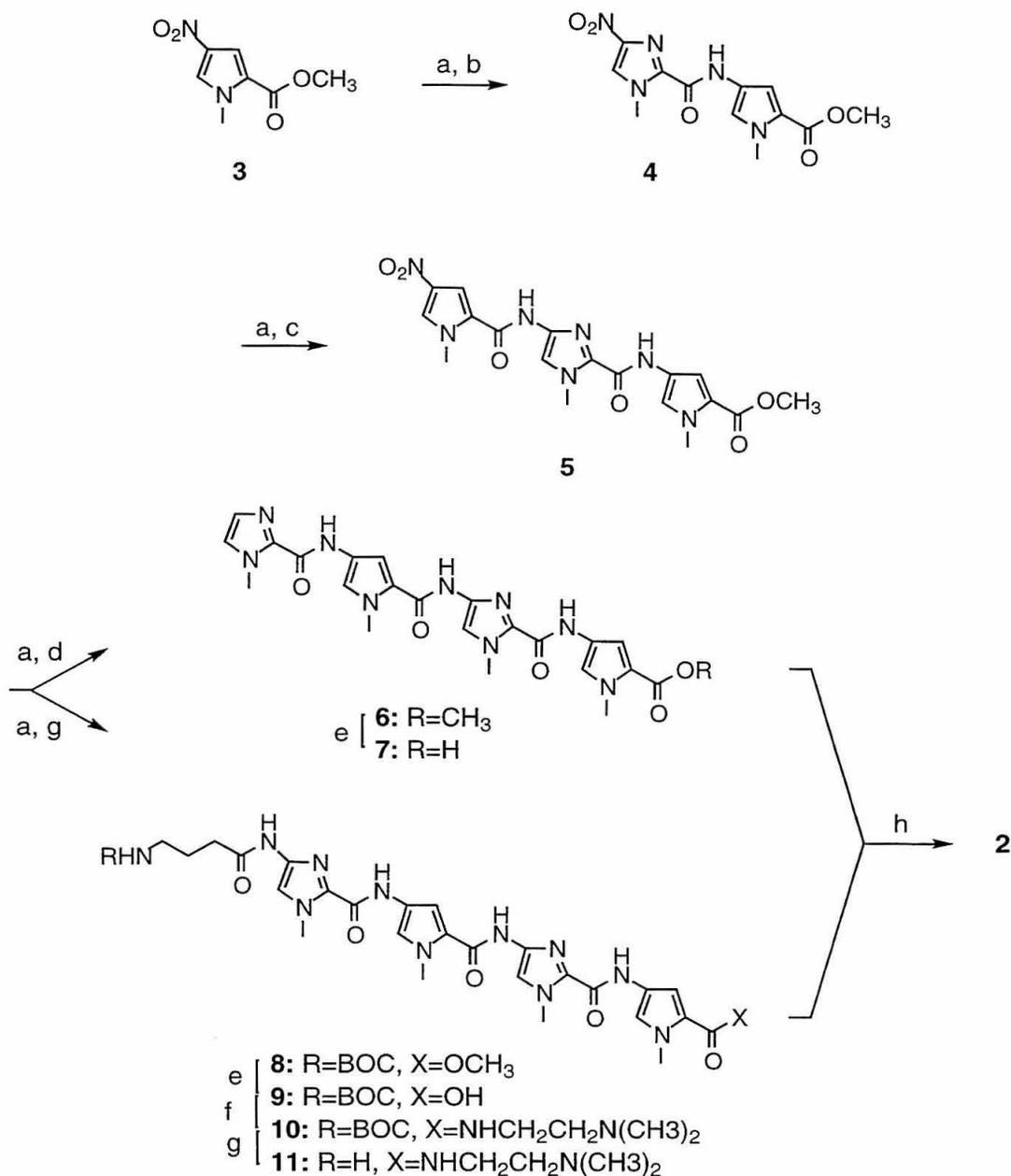


Figure 3.3. Synthesis of hairpin polyamide **2**. Reagents: (a) H₂, Pd-C; (b) 4-nitro-1-methylimidazole-2-carboxylic acid, oxalyl chloride; (c) 4-nitro-*N*-methylpyrrole-2-carboxylic acid, SOCl₂; (d) 1-methylimidazole-2-carboxylic acid, DCC, HOBT; (e) LiOH, H₂O, EtOH; (f) *N,N*-carbonyldiimidazole, *N,N*-dimethylaminopropylamine, NEt₃; (g) TFA, thiophenol; (h) DCC, HOBT.

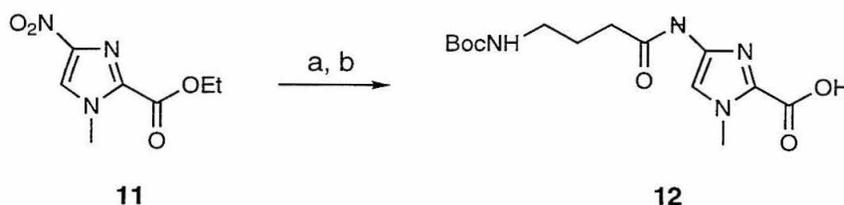


Figure 3.4. Synthesis of afford 4-(4-*N*-*t*-butoxy-aminobutyramido)-1-methylimidazole-2-carboxylic acid (**12**). Reagents: (a) (i) H₂, Pd-C; (ii) Boc-GABA, DCC, HOBT; (b) LiOH, EtOH, H₂O.

converted to the corresponding imidazolide with 1,1'-carbonyldiimidazole and condensed with 3-dimethylamino-propylamine to afford Boc-protected amine **10** containing the dimethylamino group at the C-terminus of the polyamide. Removal of the Boc group with TFA in the presence of 0.5 M thiophenol yields primary amine **11**. Conversion of acid **7** to the corresponding HOBt ester and coupling with amine **11** affords the hairpin polyamide Im-Py-Im-Py- γ -Im-Py-Im-Py-Dp **2** which was purified by reverse phase HPLC.

MPE•Fe(II) Footprinting. MPE•Fe(II) footprinting experiments with the *Eco*RI/*Bam* HI restriction fragment from pBR322 ³²P-labeled at each end reveal that the hairpin polyamide Im-Py-Im-Py- γ -Im-Py-Im-Py-Dp is specific for 5'-GCGC-3' sequences (Figures 3.5, 3.6).⁸ Strong footprints are observed at the match sites 5'-TGCGCT-3', 5'-TGCGCA-3', and 5'-AGCGCT-3' at 2 μ M concentration of Im-Py-Im-Py- γ -Im-Py-Im-Py-Dp (Figure 3.7). Weaker protection is observed at the mismatched 5'-TGCGGT-3', 5'-AACGCA-3', and 5'-TGCGTT-3' sites (base pair mismatches are underlined). No protection is observed at the mismatched 5'-TCGGCA-3' and 5'-ACCGCT-3' sites.

Quantitative DNase I Footprint Titration Experiments. DNase I footprinting experiments on the ³²P-labeled 381 base pair restriction fragment reveal that both Im-Py-

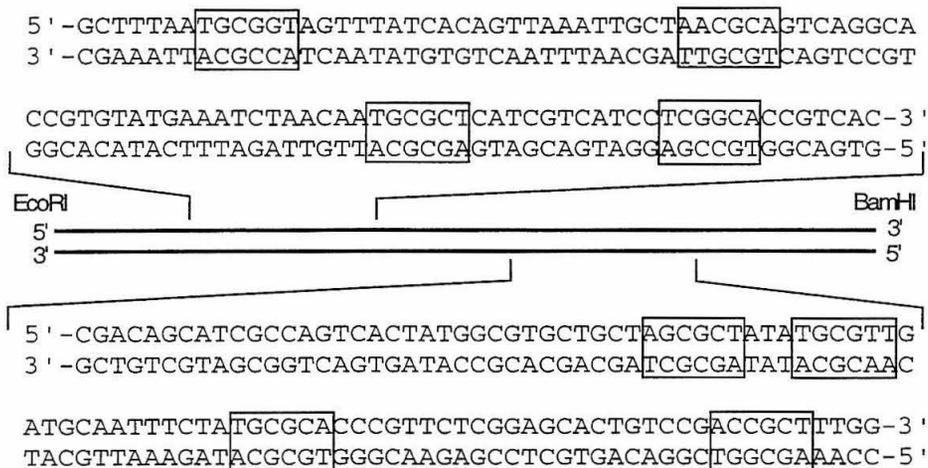
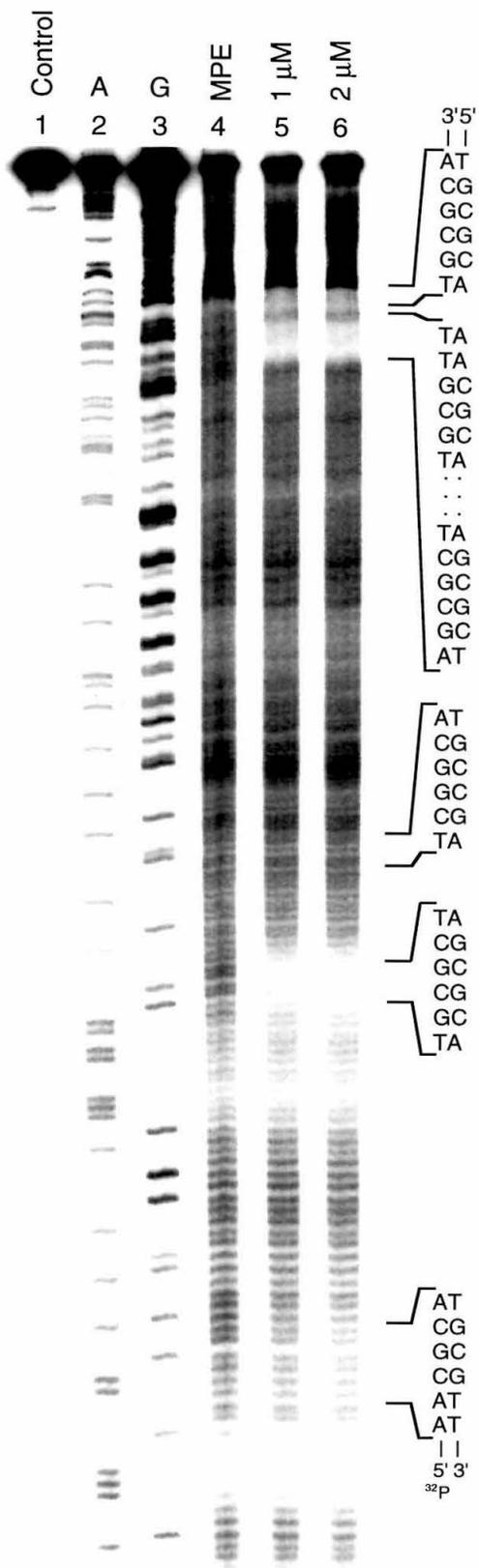


Figure 3.5. Illustration of the 381 base pair *Eco* RI/*Bam* HI restriction fragment from pBR322 used in this study. The fragment was ^{32}P -end labeled at both restriction sites in order to analyze six eight base pair sites which are indicated by boxes. The relative position and nucleotide sequence of the restriction fragment resolved by gel electrophoresis are shown.

Figure 3.6. MPE•Fe(II) footprinting experiment with hairpin polyamide Im-Py-Im-Py- γ -Im-Py-Im-Py-Dp **2** on the *Eco* RI/*Bam* HI restriction fragment from pBR322 5' -³²P end-labeled at the *Eco* RI restriction site. Autoradiogram of an 8% denaturing denaturing polyacrylamide gel. All reactions contain 10 kcpm labeled restriction fragment, 25 mM Tris•AcOH, 10 mM NaCl, and 100 μ M calf thymus DNA at 22 °C, pH 7.0. Lane 1, intact DNA, lane 2, A reaction, lane 3, G reaction, lane 4, MPE•Fe(II), lanes 5-6, 1 and 2 μ M hairpin polyamide **2**, respectively.



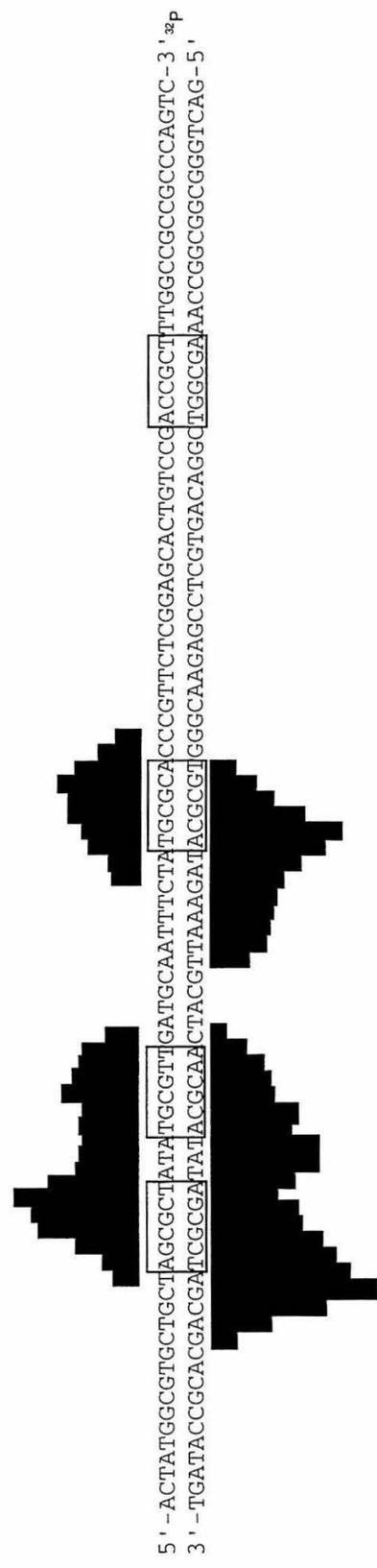


Figure 3.7. MPE•Fe(II) footprinting of Im-Py-Im-Py-Im-Py-Im-Py-Dp at 2 μM concentration. Histograms of cleavage protection where bar heights are proportional to the protection from cleavage at each nucleotide. Footprinting experiments were conducted at 22 °C, pH 7.0 in the presence of 25 mM Tris•AcOH, 10 mM NaCl, and 100 μM base pair calf thymus DNA.

Im-Py-Dp and Im-Py-Im-Py- γ -Im-Py-Im-Py-Dp bind the six base pair sites in the relative order 5'-TGCGCT-3', 5'-TGCGCA-3', 5'-AGCGCT-3' > 5'-AACGCA-3', 5'-TGCGIT-3' > 5'-TGCGGT-3', 5'-ACCGCT-3', 5'-TCGGCA-3' (Figure 3.8).⁹ From quantitative footprint titration experiments Im-Py-Im-Py-Dp binds the match sites 5'-TGCGCT-3', 5'-TGCGCA-3', and 5'-AGCGCT-3' with an apparent first order binding affinities of $3.7 \times 10^5 \text{ M}^{-1}$, $3.4 \times 10^5 \text{ M}^{-1}$, and $3.9 \times 10^5 \text{ M}^{-1}$, respectively (Table 1). The unlinked polyamide **1** binds the mismatch sites 5'-AACGCA-3' ($K_a = 1.7 \times 10^5 \text{ M}^{-1}$), 5'-TGCGIT-3' ($K_a = 1.5 \times 10^5 \text{ M}^{-1}$), 5'-TGCGGT-3' ($K_a < 10^5 \text{ M}^{-1}$), 5'-ACCGCT-3' ($K_a < 10^5 \text{ M}^{-1}$), and 5'-TCGGCA-3' ($K_a < 10^5 \text{ M}^{-1}$) with 2-4 fold lower affinity. In contrast, hairpin polyamide **2** binds the match sites 5'-TGCGCT-3', 5'-TGCGCA-3', and 5'-AGCGCT-3' with an apparent first order binding affinities of $1.1 \times 10^7 \text{ M}^{-1}$, $1.1 \times 10^7 \text{ M}^{-1}$, and $1.0 \times 10^7 \text{ M}^{-1}$, respectively and the mismatch sites 5'-AACGCA-3' ($K_a = 1.9 \times 10^6 \text{ M}^{-1}$), 5'-TGCGIT-3' ($K_a = 1.7 \times 10^6 \text{ M}^{-1}$), 5'-TGCGGT-3' ($K_a < 10^6 \text{ M}^{-1}$), 5'-ACCGCT-3' ($K_a < 10^6 \text{ M}^{-1}$), and 5'-TCGGCA-3' ($K_a < 10^6 \text{ M}^{-1}$) with 5-10 fold lower affinity (Table 1). Like Im-Py-Im-Py-Dp the hairpin polyamide Im-Py-Im-Py- γ -Im-Py-Im-Py-Dp does not discriminate between A•T and T•A at the first and sixth positions of the binding site. Im-Py-Im-Py- γ -Im-Py-Im-Py-Dp binds three sites 5'-TGCGCT-3', 5'-TGCGCA-3', and 5'-AGCGCT-3' with similar binding affinities ($\sim 1.0 \times 10^7 \text{ M}^{-1}$) (Table I). However, it is specific for and discriminates between G•C and C•G base pairs in the 5'-GCGC-3' core as evidenced by lower affinities for the mismatched sites 5'-AACGCA-3', 5'-TGCGIT-3', 5'-TGCGGT-3', and 5'-ACCGCT-3' on the same DNA fragment.

Discussion

Binding Affinity for 5'-(A,T)GCGC(A,T)-3' Sequences. Covalently linking two Im-Py-Im-Py-Dp polyamides affords an eight ring polyamide Im-Py-Im-Py- γ -Im-

Figure 3.8. Quantitative DNase I footprint titration experiment with hairpin polyamide Im-Py-Im-Py- γ -Im-Py-Im-Py-Dp **2** on the *Eco* RI/*Bam* HI restriction fragment from pBR322 3'-³²P end-labeled at the *Eco* RI restriction site. Autoradiogram of an 8% denaturing denaturing polyacrylamide gel. All reactions contain 10 kcpm labeled restriction fragment, 10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl₂, and 5 mM CaCl₂. Lane 1, intact DNA, lane 2, A reaction, lane 3, G reaction, lane 4, DNase I control, lanes 5-17, 0.1 nM, 0.2 nM, 0.5 nM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, 0.1 μ M, 0.15 μ M, 0.25 μ M, 0.5 μ M polyamide **2**, respectively.

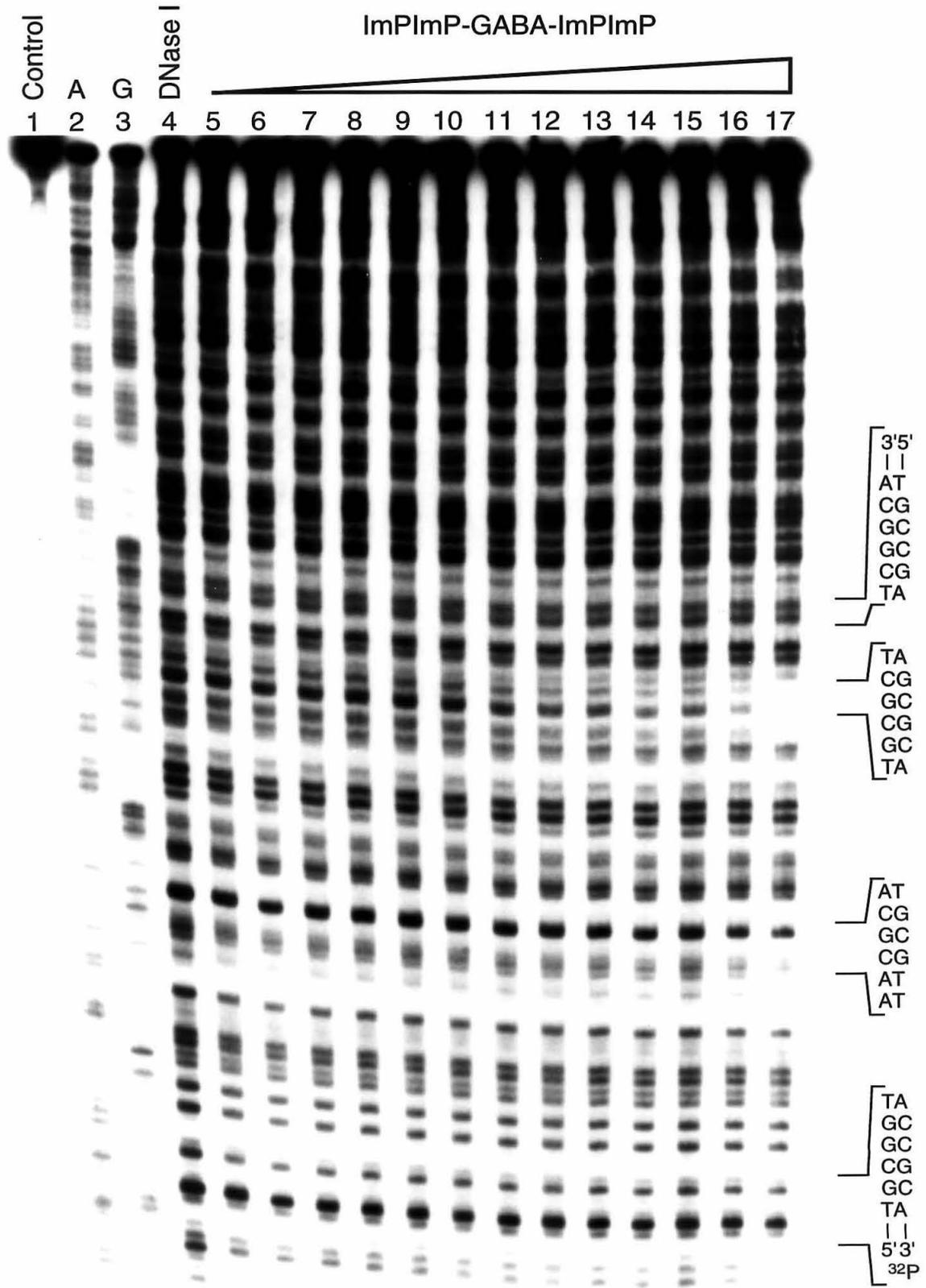


Table I. First Order Affinity Constants (M^{-1}) for Im-Py-Im-Py-Dp **1** and Im-Py-Im-Py- γ -Im-Py-Im-Py-Dp **2**.^{a,b}

Site (5'-3') ^c	1	2	Mismatches
TGCGCT	3.7×10^5 (0.2)	1.1×10^7 (0.3)	0
AGCGCT	3.4×10^5 (0.3)	1.1×10^7 (0.2)	0
TGCGCA	3.9×10^5 (0.3)	1.0×10^7 (0.4)	0
A <u>AC</u> GCA	1.7×10^5 (0.7)	1.9×10^6 (0.9)	1
TGCG <u>T</u> T	1.5×10^5 (0.9)	1.7×10^6 (0.8)	1
TGCG <u>G</u> T ^d	$< 10^5$	$< 10^6$	1
A <u>CC</u> GCT ^d	$< 10^5$	$< 10^6$	1
T <u>CG</u> GCA ^d	$< 10^5$	$< 10^6$	2

^a Values reported are the mean values determined from four footprint titration experiments, with the standard deviation for each data set indicated in parentheses. ^b The assays were performed at 22 °C at pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl₂, and 5 mM CaCl₂. ^c Base pair mismatches are underlined. ^d Upper limits for binding affinities are estimated since the quality of fits and standard deviations for these data sets were poor.

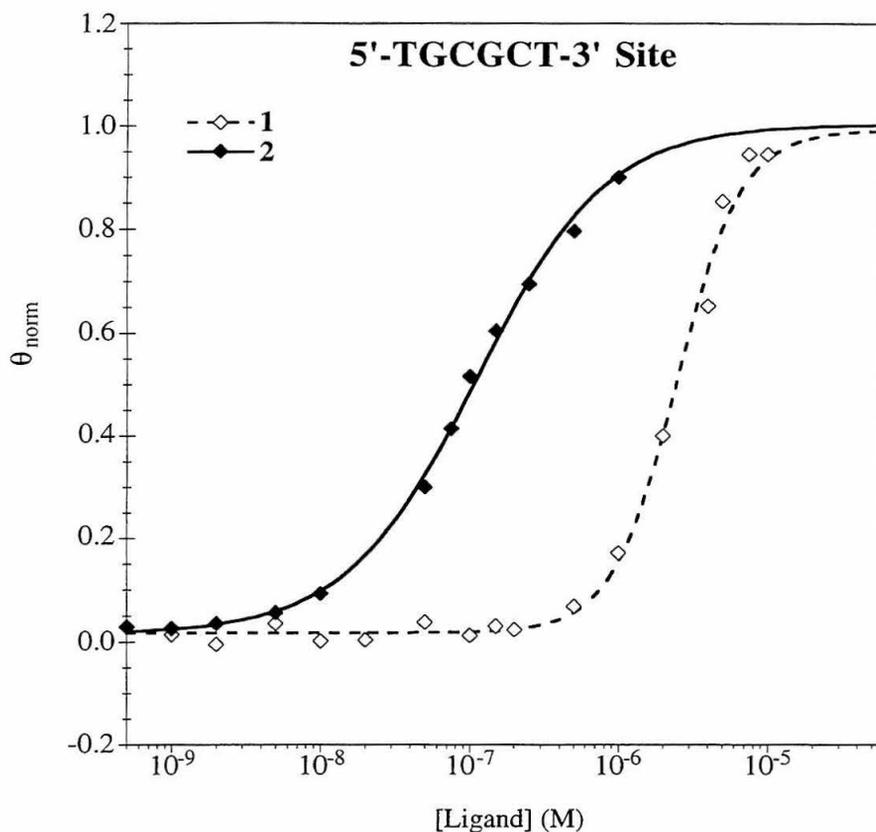
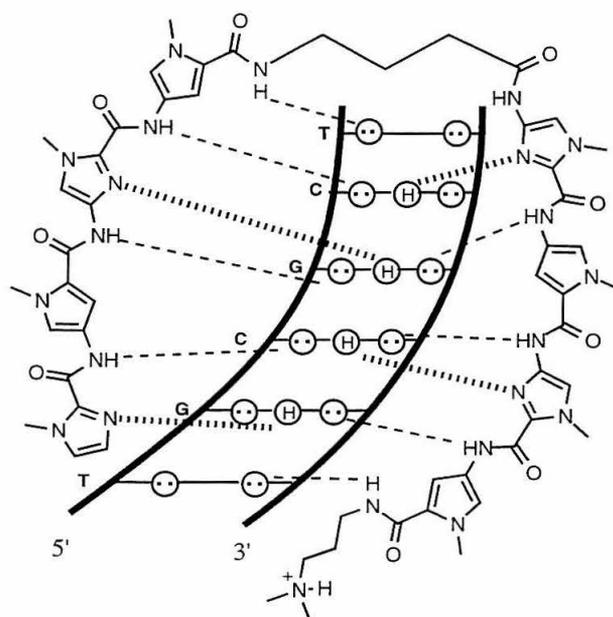


Figure 3.9. Data for quantitative DNase I footprint titrations for Im-Py-Im-Py-Dp **1** and Im-Py-Im-Py- γ -Im-Py-Im-Py-Dp **2** binding to the 5'-TGCGCT-3' site. The θ_{norm} points were obtained using photostimulable storage phosphor autoradiography. The data points for **1** are represented by open diamonds and for **2** by filled diamonds. The dashed curve is a best fit cooperative binding isotherm ($n = 2$) consistent with 2:1 polyamide-DNA complex formation with **1**. The solid curve is a best fit Langmuir binding isotherm ($n = 1$) consistent with 1:1 polyamide-DNA complex formation with **2**.

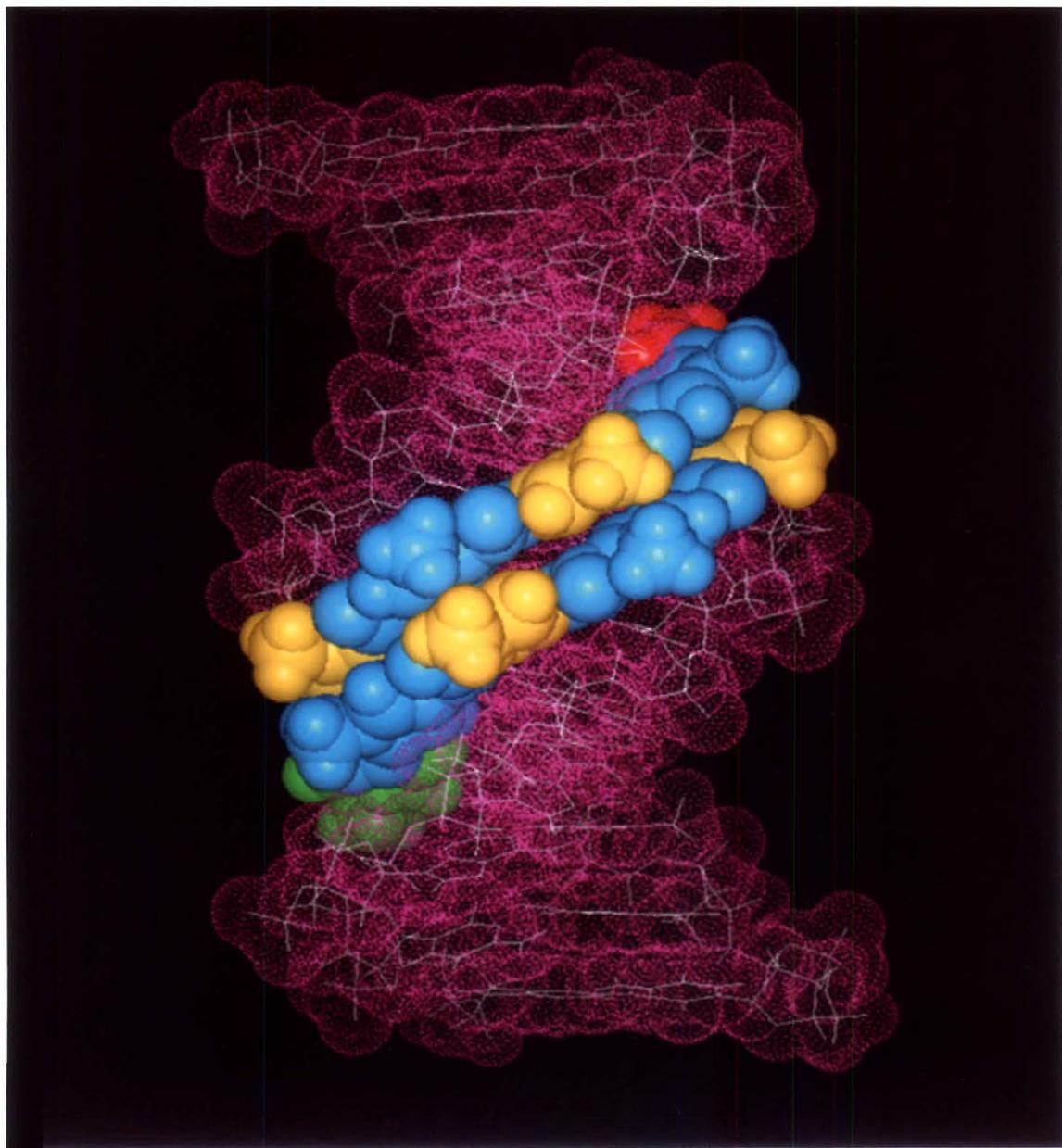
Py-Im-Py-Dp that binds 5'-(A,T)GCGC(A,T)-3' sequences with 30-fold higher affinity (Figure 3.9). This 2.3 kcal/mol increase in the free energy of binding of **2** is likely due to a more favorable translational entropy for a single polyamide versus dimeric binding of **1** at the six base pair sequences minus the entropic penalties due to restricted rotational freedom of the GABA linker in a hairpin conformation when bound in the minor groove as well as the loss of one of the two positive charges in proceeding from the 2:1 complex to the 1:1 hairpin polyamide-DNA complex (Figures 3.10, 3.11). Future work must now define the sequence composition limitations as well as the upper limit for the binding site size which can accommodate hairpin polyamides in the minor groove of DNA.



Im-Py-Im-Py- γ -Im-Py-Im-Py-Dp • TGCGCT

Figure 3.10. Proposed binding of model the Im-Py-Im-Py- γ -Im-Py-Im-Py-Dp•5'-TGCGCA-3' complex derived from NMR data. Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines and circles containing a H represent the 2-amino group of guanine. Putative hydrogen bonds are indicated by dashed lines.

Figure 3.11. Space-filling model of Im-Py-Im-Py- γ -Im-Py-Im-Py-Dp bound in the minor groove of a 5'-TGCGCT-3' site. The hairpin polyamide is shown as a CPK model in blue with the imidazole rings, GABA linker, and positively charged dimethylaminopropyl group highlighted in yellow, red, and green, respectively. The DNA is shown as a van der Waals surface in magenta with the model overlaid in white. This model was constructed from the model of the 2:1 polyamide-DNA complex between Im-Py-Im-Py-Dp and 5'-TGCGCA-3' derived from NMR and molecular modeling data using the InsightII/Discover package (Biosym). The A•T base pair at the 3'-end of the binding site was substituted with T•A. The hairpin polyamide was created by deleting a dimethylaminopropyl group from one peptide, replacing it with GABA, and bonding the carboxylate of the GABA linker to an amino group added to the N-terminal imidazole ring of the other polyamide in the 2:1 complex. The conformation of the GABA linker was energy minimized while restraining the remainder of the polyamide.



Experimental Section

Polyamide Synthesis. ^1H NMR were recorded on a General Electric QE-300 spectrometer with chemical shifts reported in parts per million relative to residual solvent signal. High-resolution mass spectra (HRMS) were recorded using fast atom bombardment (FAB) techniques at the Mass Spectrometry Laboratory at the University of California, Riverside. Matrix-assisted laser desorption/ionization time-of-flight mass spectra (MALDI-TOF MS) were recorded in the Biomolecule Analysis Facility at the California Institute of Technology. Ultraviolet-visible (UV) spectra were recorded on a Hewlett Packard 8452A diode array spectrophotometer. Reactions were executed under an atmosphere of argon. 1-Methylimidazole-2-carboxylic acid was prepared as described.^{1a} Reagent grade chemicals were used as received unless noted otherwise. Tetrahydrofuran (THF), dichloromethane, and dimethylformamide (DMF) were purchased from Fluka over molecular sieves. Dicyclohexylcarbodiimide (DCC) and *N*-hydroxybenzotriazole (HOBT) were purchased from Peptides International. 4-*N*-*t*-Butoxy-aminobutyric acid was purchased from Novachem. Trifluoroacetic acid (TFA) was purchased from Applied Biosystems or Halocarbon. Thiophenol, thionyl chloride, 10% palladium on carbon (Pd/C), and triethylamine were obtained from Aldrich. Oxalyl chloride was purchased from Fluka. Flash column chromatography was carried out using EM Science Kieselgel 60 (230-400 mesh). Thin-layer chromatography was performed on EM Reagents silica gel plates (0.5 mm thickness). All compounds were visualized at 254 nm. Preparative reverse phase high-performance liquid chromatography (RP-HPLC) was performed on a Beckman HPLC system using a Waters Delta-Pak Prep-Pak C₁₈ column (25 mm x 10 cm, 15 μ , 100 Å) equipped with a Delta-Pak Guard-Pak column. Compounds were eluted with a gradient of 0-60% CH₃CN/0.1% TFA over 120 min with a flow rate of 8-10 mL/min and detected using a Beckman 167 dual wavelength detector at 254 and 340 nm. Fractions were collected and analyzed by analytical reverse phase HPLC on a Beckman Gold HPLC system using a Rainin Microsorb-MV C₁₈ column (4.6 mm x 30 cm, 5 μ , 100 Å).

Compounds were eluted with a gradient of 0-60% CH₃CN/0.1% TFA over 60 min with a flow rate of 1 mL/min and detected using a dual wavelength detector at 254 and 340 nm. The desired fractions were pooled and lyophilized.

Dimer 4. To a slurry of 4-nitro-1-methylimidazole-2-carboxylic acid (4.73 g, 27.3 mmol) in THF (20 mL) cooled to 0 °C was added oxalyl chloride (20 mL, 0.229 mol) and the resulting solution was heated at reflux for 2 h. The solution was cooled and concentrated under high vacuum. The resulting white solid was dissolved in DMF (10 mL) and cooled to 0 °C. Separately, a solution of methyl 4-nitro-1-methylpyrrole-2-carboxylate (**3**) (2.0 g, 1.97 mmol) in DMF (15 mL) and 10% Pd/C (670 mg) was hydrogenated in a Parr bomb apparatus (350 psi H₂) for 6 h at 22 °C. The solution was filtered through celite, triethylamine (12 mL, 86 mmol) was added, and the solution was added dropwise to acid chloride at 0 °C. The ice bath was removed and the solution allowed to stir 24 h at 22 °C. DMF was removed under high vacuum and the remaining solid was dissolved in CH₂Cl₂ (250 mL), washed with H₂O (2 x 100 mL), dried (Na₂SO₄), and concentrated under reduced pressure to afford **4** (2.0 g, 6.5 mmol, 60%) as a yellow solid that was used without further purification. ¹H NMR (300 MHz, CDCl₃) δ 9.02 (s, 1 H), 7.8 (s, 1 H), 7.4 (d, 1 H, *J* = 3.2 Hz), 6.8 (d, 1 H, *J* = 3.2 Hz), 4.3 (s, 3 H), 3.9 (s, 3 H), 3.8 (s, 3 H). EI MS *m/e* (relative intensity) 307.0912 (1.0, M⁺, 307.2680 calcd. for C₁₂H₁₃N₅O₅).

Trimer 5. To a flask equipped with a reflux condenser were added 4-nitro-*N*-methylpyrrole-2-carboxylic acid (4.4 g, 26 mmol) and thionyl chloride (30 mL, 0.411 mol) and the resulting slurry was heated at reflux for 4 h. Excess SOCl₂ was removed under high vacuum and the resulting solid was dissolved in DMF (10 mL) and cooled to 0 °C. Separately, a solution **3** (2.0 g, 6.5 mmol) in DMF (80 mL) and 10% Pd/C (670 mg) was hydrogenated in a Parr bomb apparatus (350 psi H₂) for 6 h at 22 °C. The solution was filtered through celite, triethylamine (12 mL, 86 mmol) was added, and the solution was added dropwise to the acid chloride at 0 °C. The ice bath was removed and the solution allowed to stir at 24 h at 22 °C for. The solution was filtered through celite and methanol

(10 mL) was added. Solvents were removed under high vacuum and the remaining solid was dissolved in 1:1 $\text{CH}_2\text{Cl}_2:\text{CH}_3\text{OH}$ (100 mL) and cooled to $-20\text{ }^\circ\text{C}$ to precipitate a white solid which was collected by vacuum filtration and washed with cold CH_2Cl_2 to afford *N*-acyl imidazolium derivative **13** (Figure 3.12) (1.6 g, 3.73 mmol, 57%). ^1H NMR (300 MHz, $\text{DMSO}-d_6$) δ 10.7 (s, 1 H), 10.2 (s, 1 H), 8.20 (dd, 2 H, $J = 2.5$ Hz), 7.20 (d, 1 H, $J = 2.4$ Hz), 7.62 (s, 1 H), 7.53 (d, 1 H, $J = 5.1$ Hz), 7.25 (d, 1 H, $J = 2.3$ Hz) 6.98 (d, 1 H, $J = 5.0$ Hz), 3.94 (s, 3 H), 3.92 (s, 3 H), 3.82 (s, 3 H), 3.75 (s, 3 H), 3.70 (s, 3H). This material was slurried in 1:1 $\text{CH}_3\text{OH}:\text{H}_2\text{O}$ (100 mL) containing 3 drops 50% NaOH and stirred vigorously for 20 min. Extraction with CH_2Cl_2 (4 x 100 mL) afforded trimer **5** (1.1 g, 2.56 mmol, 39%) as a light yellow solid that was used without further purification. ^1H NMR (300 MHz, $\text{DMSO}-d_6$) δ 10.7 (s, 1 H), 10.2 (s, 1 H), 8.20 (d, 1 H, $J = 2.5$ Hz), 7.20 (d, 1 H, $J = 2.4$ Hz), 7.62 (s, 1 H), 7.53 (d, 1 H, $J = 5.1$ Hz), 6.98 (d, 1 H, $J = 5.0$ Hz), 3.94 (s, 3 H), 3.92 (s, 3 H), 3.75 (s, 3 H), 3.70 (s, 3H). EI MS *m/e* (relative intensity) 429.1377 (1.0, M, 429.3954 calcd. for $\text{C}_{18}\text{H}_{19}\text{N}_7\text{O}_6$).

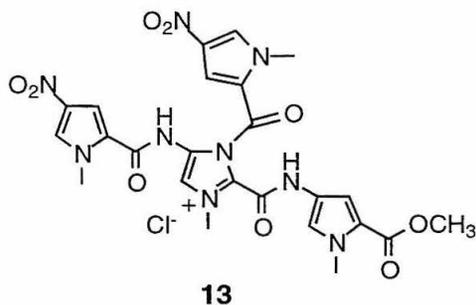


Figure 3.12. Putative structure of *N*-acyl imidazolium derivative **13**.

Ester 6. To a solution of 1-methylimidazole-2-carboxylic acid (177 mg, 1.4 mmol) and HOBt (190 mg, 1.41 mmol) in DMF (5 mL) was added DCC (288 mg, 1.40 mmol) and the resulting solution stirred 4.5 h at $22\text{ }^\circ\text{C}$. Separately, a solution of **5** (220 mg, 0.51 mmol) and 10% Pd/C (75 mg) in DMF (8 mL) was hydrogenated in a Parr bomb apparatus (350 psi H_2) for 5 h at $22\text{ }^\circ\text{C}$. The solution was filtered through celite, triethylamine (2.5 mL, 18 mmol) was added, and the solution was added dropwise to the HOBt ester. The solution was allowed to stir 24 h at $22\text{ }^\circ\text{C}$ and then quenched with H_2O (5

mL). The solution was concentrated to dryness under high vacuum. The resulting solid was purified by flash chromatography (2% CH₃OH/CH₂Cl₂) to afford ester **6** (50 mg, 0.098 mmol, 19%). ¹H NMR (300 MHz, CDCl₃) δ 9.39 (s, 1 H), 8.83 (s, 1 H), 8.04 (s, 1 H), 7.47 (s, 1 H), 7.41 (d, 1 H, *J* = 5.2 Hz), 7.32 (d, 1 H, *J* = 5.2 Hz), 7.08 (d, 1 H, *J* = 3.1 Hz), 7.05 (d, 1 H, *J* = 3.1 Hz), 6.84 (d, 1 H, *J* = 6.5 Hz), 6.82 (d, 1 H, *J* = 6.5 Hz), 4.13 (s, 3 H), 4.06 (s, 3 H), 3.97 (s, 3 H), 3.92 (s, 3 H), 3.82 (s, 3 H). FABMS *m/e* 508.3327 (508.5210, M + H, calcd. for C₂₃H₂₆N₉O₅).

Acid 7. To a solution of ester **6** (50 mg, 0.098 mmol) in ethanol (3 mL) was added 1.0 M LiOH (3 mL) and the solution was heated at reflux for 2 h. The solution was cooled to room temperature and acidified to pH 3 with 1 M NaHSO₄. A solid precipitated that was isolated by centrifugation. The solid was washed with H₂O (2 x 15 mL), resuspended in H₂O (5 mL), and dried *in vacuo* to afford **7** (32 mg, 0.065 mmol, 66%) as a tan solid. ¹H NMR (300 MHz, DMSO-*d*₆) δ 10.4 (s, 1 H), 10.3 (s, 1 H), 7.53 (s, 1 H), 7.38 (m, 3 H), 7.14 (d, 1 H, *J* = 4.2 Hz), 7.03 (d, 1 H, *J* = 3.1 Hz), 6.83 (d, 1 H, *J* = 2.2 Hz). FABMS *m/e* 494.5019 (494.4940, M + H, calcd. for C₂₂H₂₄N₉O₅).

4-(4-*N*-*t*-butoxy-aminobutyramido)-1-methylimidazole-2-carboxylic acid (12). To a solution of 4-*N*-*t*-butoxy-aminobutyric acid (4.54 g, 20.8 mmol) and HOBt (2.9 g, 20.5 mmol) in 2:1 CH₂Cl₂:DMF (30 mL) cooled to 0 °C was added DCC (4.23 g, 20.5 mmol) in DMF (10 mL). The solution was stirred 15 min at 0 °C and then allowed to warm to room temperature and stirred an additional 3 h. The solution was filtered, concentrated under reduced pressure to remove CH₂Cl₂, and cooled to 0 °C. Separately, a solution ethyl 4-nitro-1-methylimidazole-2-carboxylate (2.0 g, 10.0 mmol) in DMF (15 mL) and 10% Pd/C (670 mg) was hydrogenated in a Parr bomb apparatus (300 psi H₂) for 5 h at 22 °C. The solution was filtered through celite, triethylamine (12 mL, 86 mmol) was added, and the solution was added dropwise to the HOBt ester. After 10 min the solution was allowed to warm to room temperature and stirred an additional 24 h. Water (5 mL) was added and the solution was concentrated to dryness under high vacuum.

The resulting solid was purified by flash chromatography (10% CH₃OH/CH₂Cl₂) to afford ethyl-4-(4-*N*-*t*-butoxy-aminobutyramido)-1-methylimidazole-2-carboxylate (2.25 g, 64%) as a light yellow solid. ¹H NMR (300 MHz, CDCl₃) δ 8.43 (s, 1 H), 7.50 (s, 1 H), 4.68 (br, s, 1 H), 4.35 (q, 2 H, *J* = 12 Hz), 3.99 (s, 3 H), 3.29 (m, 2 H), 2.38 (t, 2 H, *J* = 10 Hz), 1.77 (quintet, 2 H, *J* = 10 Hz), 1.42 (s, 9 H) 1.41 (t, 3 H, *J* = 12 Hz). EI MS *m/e* (relative intensity) 355.1967 (0.24, M + H, 355.4170 calcd. for C₁₆H₂₇N₄O₅). To a slurry of ethyl-4-(4-*N*-*t*-butoxy-aminobutyramido)-1-methylimidazole-2-carboxylate (2.0 g, 5.6 mmol) in ethanol (30 mL) was added 1 M LiOH (30 mL) and the resulting solution was heated at reflux for 3 h. The solution was cooled to room temperature and concentrated under reduced pressure to remove EtOH. The solution was cooled to 0 °C and acidified to pH 3 with 15% citric acid. The solution was filtered and the white solid was washed with cold H₂O and dried in vacuo to afford 4-(4-*N*-*t*-butoxy-aminobutyramido)-1-methylimidazole-2-carboxylic acid (**12**) (1.6 g, 4.9 mmol, 88%). ¹H NMR (300 MHz, DMSO-*d*₆) δ 10.52 (s, 1 H), 7.48 (s, 1 H) 6.75 (br, s, 1 H), 3.85 (s, 3 H), 2.85 (m, 2 H), 2.22 (m, 2 H), 1.65 (m, 2 H), 1.35 (s, 9 H). FABMS *m/e* 327.1653 (M + H, 327.3630 calcd. for C₁₄H₂₃N₄O₅).

Ester 8. To a solution of 4-(4-*N*-*t*-butoxy-aminobutyramido)-1-methylimidazole-2-carboxylic acid (**12**) (304 mg, 0.93 mmol) and HOBt (200 mg, 1.48 mmol) in DMF (7 mL) was added DCC (290 mg, 1.41 mmol) and the resulting solution stirred 4 h at 22 °C. Separately, a solution of **5** (400 mg, 0.93 mmol) in DMF (15 mL) and 10% Pd/C (145 mg) was hydrogenated in a Parr bomb apparatus (400 psi H₂) for 4.5 h at 22 °C. The solution was filtered through celite, triethylamine (4.5 mL, 32 mmol) was added, and the solution was added dropwise to the HOBt ester. The solution was allowed to stir 24 h at 22 °C and then quenched with H₂O (5 mL). The solution was concentrated to dryness under high vacuum and the residue was dissolved in CH₂Cl₂ (300 mL). After filtering through celite the solution was washed with H₂O (50 mL), 5% NaHSO₄ (50 mL), 2% NaHCO₃ (50 mL), and H₂O (50 mL). The solution was dried (Na₂SO₄) and concentrated

under reduced pressure. Flash column chromatography (0-4% CH₃OH/ CH₂Cl₂) afforded ester **8** (361 mg, 0.51 mmol, 55%) as an orange solid. ¹H NMR (300 MHz, CDCl₃) δ 9.05 (br, s, 1 H), 8.98 (s, 1 H), 8.80 (s, 1 H), 8.30 (s, 1H), 7.46 (s, 1 H), 7.43 (d, 2 H), 7.42 (d, 1H), 7.29 (d, 1 H, *J* = 3.5 Hz), 6.87 (br, s, 1 H), 6.81 (br, s, 1 H) 4.05 (s, 3 H), 4.02 (s, 3 H), 3.93 (s, 3 H), 3.89 (s, 3 H), 3.81 (s, 3 H), 3.20 (m, 2 H), 2.42 (m, 2 H), 1.85 (m, 2 H), 1.48 (s, 9 H). FABMS *m/e* (relative intensity) 708.3229 (0.23, M + H, 707.7522 calcd. for C₃₂H₄₈N₁₁O₈).

Acid 9. To a slurry of ester **8** (361 mg, 0.51 mmol) in ethanol (15 mL) was added 1 M LiOH (15 mL) and the resulting solution was heated at reflux for 2.5 h. The solution was cooled to room temperature and concentrated under reduced pressure to remove EtOH. The solution was cooled to 0 °C, acidified to pH 3 with 1 M NaHSO₄, and extracted with ethyl acetate (4 x 25 mL). The combined organics were dried (Na₂SO₄) and concentrated under reduced pressure to afford **9** (260 mg, 0.37 mmol, 73%). ¹H NMR (300 MHz, CD₃OD) δ 7.43 (s, 1 H), 7.39 (s, 1 H), 7.36 (d, 1 H, *J* = 5.5 Hz), 7.25 (d, 1 H, *J* = 5.0 Hz), 7.02 (d, 1 H, *J* = 5.5 Hz), 6.78 (d, 1 H, *J* = 5.0 Hz), 4.06 (s, 3 H), 4.03 (s, 3 H), 3.94 (s, 3 H), 3.91 (s 3 H), 3.15 (t, 3 H, *J* = 8.2 Hz), 2.35 (m, 2 H), 1.85 (t, 2 H, *J* = 8.2 Hz), 1.46 (s, 9 H). FABMS *m/e* 608.6317 (M + H, 608.6188 calcd. for C₂₇H₃₃N₁₀O₇).

Boc-Protected Amine 10. To a solution of **9** (130 mg, 0.21 mmol) in DMF (1 mL) was added 1, 1'-carbonyldiimidazole (320 mg, 1.97 mmol) and the solution was stirred 4 h at 22 °C. 3-Dimethylaminopropylamine (650 mg, 6.4 mmol) was added and the solution was stirred 30 min at 22 °C. The solution was diluted to 10 mL with H₂O and acidified to pH 4 with 1 N HCl. The solution was purified by preparative reverse phase HPLC to afford **10** (50 mg, 0.062 mmol, 30%) as the trifluoroacetate salt. ¹H NMR (300 MHz, CD₃OD) δ 7.39 (s, 1 H), 7.37 (s, 1 H), 7.29 (br, s, 1 H), 7.23 (br, s, 1 H), 6.94 (d, 1 H, *J* = 3.2 Hz) 6.82 (d, 1 H, *J* = 3.2 Hz), 4.02 (s, 3 H), 3.99 (s, 3 H), 3.87 (s, 3 H), 3.84 (s, 3 H), 3.42 (t, 2 H, *J* = 7.5 Hz), 3.19 (t, 2 H, *J* = 7.4 Hz), 3.12 (t, 2 H, *J* =

7.5 Hz), 2.35 (t, 2 H, $J = 7.2$ Hz), 1.98 (t, 2 H, $J = 7.5$ Hz), 1.81 (t, 2 H, $J = 7.5$ Hz), 1.46 (s, 9H). FABMS m/e 692.7785 (M + H, 692.7844 calcd. for $C_{32}H_{44}N_{12}O_6$).

Amine 11. Boc-protected amine **10** (50 mg, 0.062 mmol) was treated with TFA containing 0.5 M thiophenol (5 mL) for 30 min at 22 °C. Excess TFA was removed under high vacuum to afford amine **11** (26 mg, 0.037 mmol) that was used immediately without further purification. $^1\text{H NMR}$ (CD_3OD) δ 7.43 (s, 1 H), 7.41 (s, 1 H), 7.31 (d, 1 H, $J = 3.6$ Hz), 7.24 (d, 1 H, $J = 3.7$ Hz), 7.08 (d, 1 H, $J = 4.6$ Hz), 6.95 (d, 1 H, $J = 4.6$ Hz), 4.08 (s, 3 H), 4.08 (s, 3 H), 3.95 (s, 3 H), 3.93 (s, 3 H), 3.42 (t, 2 H, $J = 7.7$ Hz), 3.18 (t, 2 H, $J = 7.7$ Hz), 3.03 (t, 2 H, $J = 7.5$ Hz), 2.57 (t, 2 H, $J = 7.5$ Hz), 2.97 (s, 6 H), 2.0 (m, 4 H).

Im-Py-Im-Py- γ -Im-Py-Im-Py-Dp (2) (p. 70). To a solution of acid **7** (21 mg, 23.5 μmol) and HOBT (13 mg, 96 μmol) in DMF (80 μL) was added DCC (10.6 mg, 51 μmol) and the solution vigorously stirred 6 h at 22 °C. A solution of amine **11** (26 mg, 37 μmol) in DMF (70 μL) was added to the HOBT ester followed by triethylamine (7 μL , 50 μmol). The solution was stirred for 24 h at 22 °C and then diluted to 2 mL with H_2O and concentrated under high vacuum. The residue was dissolved in 10 mL 0.1% TFA/ H_2O , filtered (0.45 μ), and purified by preparative reverse phase HPLC to afford **2** as the trifluoroacetate salt. The solid was dissolved in 2 mL 0.1 N HCl and concentrated under high vacuum, dissolved in H_2O (1 mL) and concentrated under high vacuum to afford **2** (1.25 mg) as the hydrochloride salt. $^1\text{H NMR}$ (CD_3OD) δ 7.38 (s, 1 H), 7.35 (s, 1 H), 7.28 (d, 1 H, $J = 1.8$ Hz), 7.25 (d, 1 H, $J = 1.9$ Hz), 7.22 (br s, 2 H), 7.21 (s, 1 H), 7.16 (d, 1 H, $J = 1.9$ Hz), 7.07 (d, 1 H, $J = 1.7$ Hz), 6.93 (d, 1 H, $J = 1.9$ Hz), 6.89 (d, 1 H, $J = 1.8$ Hz), 6.87 (d, 1 H, $J = 1.9$ Hz), 6.63 (d, 1 H, $J = 1.8$ Hz), 4.05 (s, 3 H), 4.04 (br s, 6 H), 3.99 (s, 3 H), 3.93 (s, 3 H), 3.90 (s, 3 H), 3.89 (s, 3 H), 3.87 (s, 3 H), 3.40 (t, 2 H, $J = 7.0$ Hz), 3.18 (m, 4 H), 2.93 (s, 6 H), 2.42 (t, 3 H, $J = 6.5$ Hz), 2.05 (m, 4 H). UV (H_2O) ϵ_{312} 38,000 $\text{M}^{-1}\text{cm}^{-1}$; MALDI-TOF MS m/e : 1154.2 (M + H, 1154.3 calcd. for $C_{53}H_{65}N_{22}O_9$).

DNA Reagents and Materials. RNase free water was obtained from U.S. Biochemicals and used in all manipulations. Plasmid pBR322 was obtained from Boehringer-Mannheim. *Eco* RI, *Bam* HI, the Klenow fragment of DNA polymerase I, calf alkaline phosphatase, and T4 polynucleotide kinase were purchased from Boehringer-Mannheim and used according to the supplier's recommended protocol in the buffers provided. Nucleotide triphosphates were Pharmacia Ultra-Pure grade and were used as supplied. Nucleotide triphosphates labeled with ^{32}P (> 3000 Ci/mmol) were obtained from Amersham. NICK columns were purchased from Pharmacia and used according to the suggested protocol. Glycogen was obtained from Boehringer-Mannheim as a 20 mg/mL aqueous solution. Cerenkov radioactivity was measured with a Beckman LS 2801 scintillation counter. DNase I was purchased from Boehringer-Mannheim. Sonicated, deproteinized calf thymus DNA was purchased from Pharmacia and dissolved in H_2O to a final concentration of 2 mM in base pairs and stored at -20 °C. Chemical sequencing reactions were performed according to published procedures. General manipulations of DNA were performed according to established procedures.

Labeled Restriction Fragment Preparations. The *Eco* RI/*Bam* HI fragment of pBR322 3'- ^{32}P -end labeled at the *Eco*RI site was prepared by digesting the plasmid (5-10 μg) with *Eco* RI, followed by treatment with α - ^{32}P -dATP and α - ^{32}P -TTP in the presence of the Klenow fragment of DNA polymerase I. Following the addition of unlabeled dATP and TTP unincorporated nucleotides were removed using a NICK column. The linearized, ^{32}P -labeled plasmid was then digested with *Bam* HI, followed by separation of the resulting DNA fragments on a nondenaturing 5% polyacrylamide gel (1:19 crosslinked). The desired 381-bp fragment was located by autoradiography and excised. The gel slice was crushed and soaked in 25 mM Tris-HCl pH 7.0, 25 mM NaCl, 2 mM EDTA for 8-12 h at 37 °C. The resulting suspension was then filtered through an Isolab Quik-Sep polypropylene filter and the filtrate was ethanol precipitated using glycogen (1 μL). The DNA pellet was washed with 70% EtOH, dried *in vacuo*, and resuspended in 10

mM Tris-HCl pH 7.0, 0.5 mM EDTA. The solution was then extracted with phenol (3 x 100 μ L), 24:1 phenol:CHCl₃ (3 x 100 μ L), and 24:1 CHCl₃:isoamyl alcohol (2 x 100 μ L). The solution was then passed through a NICK column equilibrated with 10 mM Tris-HCl pH 7.0, 0.5 mM EDTA buffer and the eluate stored at 20 °C after measuring Cerenkov radioactivity. Typical yields of Cerenkov radioactivity were 5-10 x 10⁶ cpm. The same restriction fragment 3'-³²P-end labeled at the *Bam* HI site was prepared as described above except that the plasmid was first digested with *Bam* HI, α -³²P-dATP, α -³²P-TTP, α -³²P-dCTP and α -³²P-dGTP were used in the labeling reaction, and *Eco* RI was used in the second digestion. The restriction fragment 5'-³²P-end labeled at the *Eco* RI site was prepared by digesting pBR322 (10 μ g) with *Eco* RI followed by treatment with calf alkaline phosphatase. The linearized plasmid was then labeled by treatment with γ -³²P-ATP in the presence of polynucleotide kinase. Following the addition of unlabeled ATP unincorporated nucleotides were removed using a NICK column. The DNA was then digested with *Bam* HI and isolated as described above. Typical yields of radioactivity were 2-5 x 10⁶ cpm. The restriction fragment 5'-³²P-end labeled at the *Bam* HI site was prepared in the same manner except that *Bam* HI and *Eco* RI were used in the first and second digestions, respectively.

MPE•Fe(II) Footprinting Experiments. For each experiment a cocktail containing Tris·AcOH pH 7.0, NaCl, calf thymus DNA, and ³²P-end labeled DNA restriction fragment was prepared. To separate 1.7 mL microfuge tubes were added 28 μ L of cocktail. A 50 μ M stock solution of Im-Py-Im-Py- γ -Im-Py-Im-Py-Dp was prepared by dissolving 2 nmol in 40 μ L H₂O and diluted serially to 20 μ M. A 4 μ L aliquot of each 10x ligand solution was added to the appropriate tubes and the solutions were allowed to equilibrate for 12-24 h at 22 °C. A 50 μ M solution of MPE•Fe(II) was prepared by mixing 100 μ L of a 100 μ M MPE solution with a freshly prepared 100 μ M ferrous ammonium sulfate solution and allowing to equilibrate for 15 min at 22 °C. A 4 μ L aliquot of 50 μ M MPE•Fe(II) was added to the appropriate tubes and incubated 15 min at 22 °C. Cleavage

was initiated by the addition of 4 μL of freshly prepared 50 mM DTT solution. Final concentrations were 25 mM Tris·AcOH pH 7.0, 10 mM NaCl, 100 μM -bp calf thymus DNA, 5 μM MPE·Fe(II), and 5 mM DTT in a total volume of 40 μL containing $1\text{-}3 \times 10^5$ cpm of ^{32}P -labeled restriction fragment. Cleavage was allowed to proceed for 20 min and then stopped by ethanol precipitation using 5 μL of glycogen. The precipitated DNA was washed with 70% EtOH, dried *in vacuo*, resuspended in 1x TBE-formamide loading buffer, and electrophoresed on 8% denaturing polyacrylamide gels (5% cross-link, 7 M urea) at 2000 V for 1-2 h. The gels were transferred to filter paper, dried, and quantitated using storage phosphor technology as described below.

Quantitative DNase I Footprint Titration Experiments. For each experiment a cocktail containing Tris·HCl, pH 7.0, KCl, MgCl_2 , CaCl_2 , and 3'- ^{32}P -end labeled restriction fragment was prepared. To separate 1.7 mL microfuge tubes was added a 32 μL aliquot of the cocktail solution. A 100 μM solution of Im-Py-Im-Py-Dp was prepared by dissolving a 10 nmol aliquot in 100 μL of H_2O . A 5 μM solution of Im-Py-Im-Py- γ -Im-Py-Im-Py-Dp was prepared by dissolving a 2 nmol aliquot in 400 μL of H_2O . These solutions were then serially diluted to the appropriate concentrations. To the appropriate tubes were added 4 μL of 10x ligand solution and the solutions were allowed to equilibrate for 8-24 h. Final ligand concentrations were 0.2 nM to 10 μM for Im-Py-Im-Py-Dp and 10 pM to 0.5 μM for Im-Py-Im-Py- γ -Im-Py-Im-Py-Dp. Cleavage was initiated by the addition 4 μL of a solution of DNase I (0.04 units/mL) containing 1 mM DTT. Final concentrations were 10 mM Tris·HCl pH 7.0, 10 mM KCl, 5 mM MgCl_2 , and 5 mM CaCl_2 in a total volume of 40 μL containing $1\text{-}2 \times 10^5$ cpm of ^{32}P -end labeled restriction fragment. Cleavage was allowed to proceed for 6 min at 22 $^\circ\text{C}$ and then stopped by the addition of 12.8 μL of a 3 M NaOAc/50 mM EDTA/glycogen solution (1.2/1/1) followed by ethanol precipitation. The DNA pellets were washed with 70% EtOH, dried *in vacuo*, resuspended in formamide-TBE loading buffer, and electrophoresed on 8% denaturing polyacrylamide gels (5% cross-linked, 7 M urea) at 2000 V for 1 h. The gels were

transferred to filter paper, dried, and quantitated using storage phosphor technology as described below.

Apparent first-order binding constants were determined as previously described.⁹ The data were analyzed by performing volume integrations of the target sites and 5'-GACA-3' and 5'-ACCA-3' reference sites for experiments with the restriction fragment labeled at the *Eco* RI and *Bam* HI sites, respectively. The apparent DNA target site saturation, θ_{app} , was calculated for each concentration of peptide using the following equation:

$$\theta_{app} = 1 - \frac{I_{site}/I_{ref}}{I_{site}^{\circ}/I_{ref}^{\circ}} \quad (1)$$

where I_{tot} and I_{ref} are the integrated volumes of the target and reference sites, respectively, and I_{tot}° and I_{ref}° correspond to those values from a DNase I control lane to which no polyamide had been added. At concentrations higher than 10 μ M Im-Py-Im-Py-Dp and 0.5 μ M Im-Py-Im-Py- γ -Im-Py-Im-Py-Dp the reference sites become partially protected, resulting in low θ_{app} values. For this reason, peptide concentrations of than 10 μ M Im-Py-Im-Py-Dp and 0.5 μ M Im-Py-Im-Py- γ -Im-Py-Im-Py-Dp were not used. The ($[L]_{tot}$, θ_{app}) data points were fit to Langmuir binding isotherms by minimizing the difference between θ_{app} and θ_{fit} using the modified Hill equation:

$$\theta_{fit} = \theta_{min} + (\theta_{max} - \theta_{min}) \frac{K_a^n [L]_{tot}^n}{1 + K_a^n [L]_{tot}^n} \quad (2)$$

where $[L]_{tot}$ corresponds to the total polyamide concentration, K_a corresponds to the apparent monomeric association constant, n represents the Hill coefficient, and θ_{max} and θ_{min} represent the experimentally determined site saturation values when the site is unoccupied or saturated, respectively. Data were fit using a non-linear least-squares fitting procedure of Kaleidagraph software (version 3.0.1; Abelbeck Software) running on a Power Macintosh 6100/60 computer with K_a , θ_{max} , and θ_{min} as the adjustable parameters. The goodness of fit is evaluated by the correlation coefficient, with $R > 0.98$ as the criterion for

an acceptable fit. All lanes from each gel were used unless visual inspection revealed a data point to be obviously flawed relative to neighboring points. The data were normalized using the following equation:

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

Quantitation by Storage Phosphor Autoradiography. Photostimulable storage phosphor imaging plates (Kodak Storage Phosphor Screen S0230 obtained from Molecular Dynamics) were pressed flat against dried gels and exposed in the dark for 8-24 h at 22 °C. A Molecular Dynamics 400S PhosphorImager was used to obtain all data from the storage screens. The data were analyzed by performing volume integrations of all bands using ImageQuant v. 3.3 software running on a Gateway 2000 486/66 computer.

References and Notes

1. (a) Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 8783. (b) Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 7586. (c) Wade, W. S.; Mrksich, M.; Dervan, P. B. *Biochemistry* **1993**, *32*, 11385.
2. (a) Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 5723. (b) Pelton, J. G.; Wemmer, D. E. *J. Am. Chem. Soc.* **1990**, *112*, 1393. (c) Chen, X.; Ramakrishnan, B.; Rao, S. T.; Sundaralingam, M. *Struct. Biol. Nat.* **1994**, *1*, 169. (d) Rentzeperis, D.; Marky, L. A.; Dwyer, T. J.; Geierstanger, B. H.; Pelton, J. G.; Wemmer, D. E. *Biochemistry* **1995**, *34*, 2937.
3. Animati, F.; Arcamone, F. M.; Conte, M. R.; Felicetti, P.; Galeone, A.; Lombardi, P.; Mayol, L. G.; Rossi, C. *J. Med. Chem.* **1995**, *38*, 1140.

4. (a) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 2572. (b) Geierstanger, B. H.; Jacobsen, J-P.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Biochemistry* **1994**, *33*, 3055.
5. (a) Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Science* **1994**, *266*, 646. (b) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1995**, *117*, 3325.
6. Geierstanger, B. H.; Dwyer, T. J.; Bathini, Y.; Lown, J. W.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, *115*, 4474.
7. For hairpin polyamides, see Mrksich, M.; Parks, M. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 7983.
8. (a) Van Dyke, M. W.; Hertzberg, R. P.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 5470. (b) Van Dyke, M. W.; Dervan, P. B. *Cold Spring Harbor Symposium on Quantitative Biology* **1982**, *47*, 347. (c) Van Dyke, M. W.; Dervan, P. B. *Biochemistry* **1983**, *22*, 2373. (d) Harshman, K. D.; Dervan, P. B. *Nucleic Acids Res.* **1985**, *13*, 4825.
9. (a) Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Methods Enzymol.* **1986**, *130*, 132. (b) Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 8462. (c) Senear, D. F.; Brenowitz, M.; Shea, M. A.; Ackers, G. K. *Biochemistry* **1986**, *25*, 7344.
10. Still, W. C.; Kahn, M.; Mitra, A. *J. Org. Chem.* **1978**, *40*, 2923, 26.
11. Iverson, B. L.; Dervan, P. B. *Nucl. Acids Res.* **1987**, *15*, 7823.
12. Maxam, A. M.; Gilbert, W. S. *Methods in Enzymology* **1980**, *65*, 499.
13. Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular Cloning*; Cold Spring Harbor Laboratory: Cold Spring Harbor, NY, 1989.

(blank page)

Chapter 4

Studies Towards the Sequence Specific Hydrolysis of HIV-1 TAR RNA

Introduction

Challenge of Sequence-Specific RNA Hydrolysis. The design and synthesis of reagents capable of sequence-specific DNA or RNA hydrolysis is one of the most challenging problems in bioorganic chemistry. The phosphodiester backbone of the nucleic acids is extremely well-suited for its biological function.¹ Phosphodiester are negatively charged and are therefore extremely resistant to nucleophilic attack in aqueous solution. The half life of DNA and RNA at pH 7.0, 37 °C in the absence of any catalysts has been estimated to be 200 million years and 2000 years, respectively.² The uncatalyzed rate of RNA hydrolysis is faster than DNA hydrolysis due to the presence of the 2'-hydroxyl group in RNA which can serve as an *intramolecular* nucleophile to cause strand scission via transesterification. In addition to the fact that phosphodiester hydrolysis is an inherently slow reaction, another challenge in the design of synthetic reagents capable of sequence-specific hydrolysis of DNA and RNA is the fact that the catalysts must be delivered to the target phosphodiester with the proper geometry for hydrolysis to occur. This requirement severely tests our ability to design molecules capable of simultaneous recognition and hydrolysis of RNA. Despite the fact that there are several naturally occurring protein and RNA catalysts for site-specific phosphoryl transfer in RNA molecules, there are few examples of designed artificial catalysts that perform these functions.³

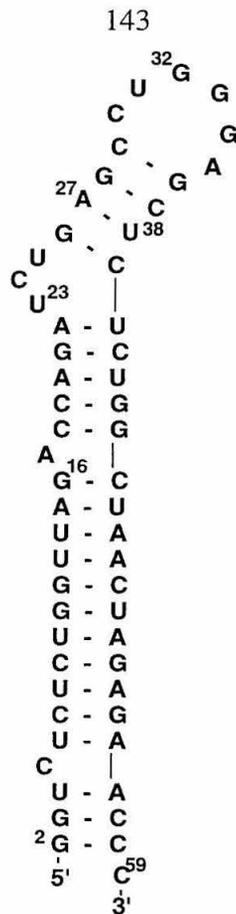


Figure 4.2. Predicted secondary structure of the HIV-1 TAR RNA sequence.

with A27-U38 in the major groove. Loret and coworkers have proposed a model for the structure of the Tat(47-72)-TAR complex based on difference circular dichroism experiments with Tat(47-72) and molecular dynamics calculations.⁹ In this model residues 47-58 of Tat bind in the major groove of TAR in an extended conformation with the N-terminus of the peptide positioned on the 5'-side of the bulge with the C-terminus oriented towards the hairpin loop and residues 59-67 forming an α -helix which makes additional contacts with the RNA. A recent NMR study of a Tat(47-57) peptide fused to a 15-residue activation domain from the Tat protein of equine infectious anemia virus suggests that the Tat(47-57) peptide forms a stable α -helix in solution in the absence of TAR RNA.¹⁰ All of

these models are highly speculative and must await confirmation by high-resolution NMR or x-ray structural data from a Tat-TAR complex.

RNA Hydrolysis by Lanthanides. Lanthanide ions have been known for over thirty years to hydrolyze RNA at neutral pH, 37 °C.¹¹ For folded RNA structures the rates of reaction can differ by several orders of magnitude (e.g. single-stranded > double-stranded) and, in a formal sense, this kinetic specificity is one example of site-specific RNA cleavage.¹² Recently, hexa-aza macrocyclic Schiff base complexes of lanthanide ions ($\text{Ln}(\text{L})^{3+}$) have been shown to be effective catalysts for RNA hydrolysis at neutral pH, 37 °C and are remarkably stable to metal ion release (Figure 4.3).¹³⁻¹⁵

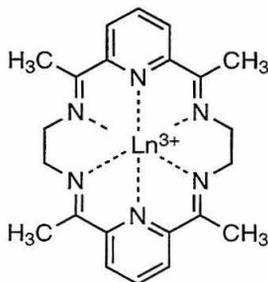


Figure 4.3. Structure of the hexa-aza Schiff base lanthanide ion macrocycle, $\text{Ln}(\text{L})^{3+}$.

Experimental Design. In order to test whether catalysts could be designed for site-specific cleavage of TAR RNA, we have covalently attached a kinetically inert hexa-aza lanthanum complex $\text{La}(\text{L})^{3+}$ to a Tat(49-72) peptide at three different cysteine residues via a disulfide bond to afford Tat(48-72, Cys⁴⁸- $\text{La}(\text{L})^{3+}$) **1**, Tat(49-72, Cys⁵⁴- $\text{La}(\text{L})^{3+}$) **2**, and Tat(49-72, Cys⁷²- $\text{La}(\text{L})^{3+}$) **3** (Figure 4.4). Although these three metallo-peptides cleave TAR site-specifically in the hexanucleotide loop to afford products consistent with hydrolysis, a series of control experiments suggests that the observed cleavage is not caused by a sequence-specifically bound Tat(49-72)- $\text{La}(\text{L})^{3+}$ peptide.

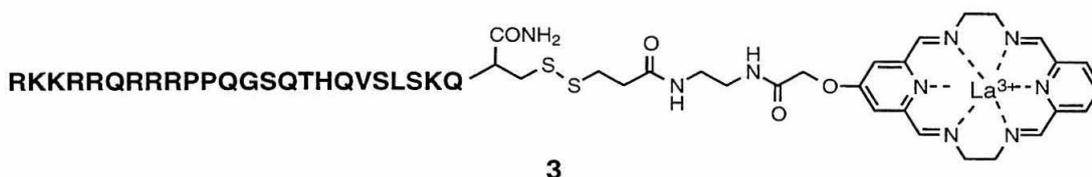
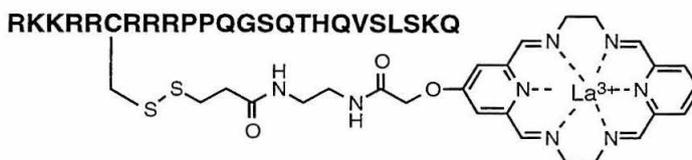
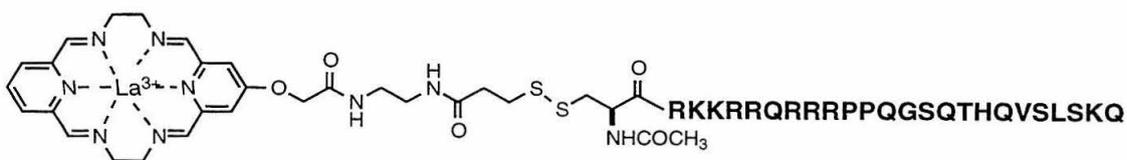


Figure 4.4. Structure of Tat-La(L)³⁺ peptides **1-3**.

Results

Synthesis of Tat(49-72)-La(L)³⁺ Peptides. Lanthanum(III)-directed condensation of dialdehyde **4**, synthesized in eight steps from commercially available chelidamic acid (**5**), with excess 2,6-diformylpyridine and ethylenediamine afforded the mono-functionalized lanthanum macrocycle La(L)³⁺ **13** which was purified by ion exchange chromatography after removal of the trifluoroacetamide group (Figures 4.5, 4.6).¹⁶⁻¹⁷ Acylation of the resulting amine **13** with *N*-succinimidyl-3-(2-pyridyldithio)propionate (SPDP) afforded pyridyl disulfide **14** which was purified by reverse-phase HPLC. Treatment of Tat(48-72, Cys⁴⁸), Tat(48-72, Cys⁵⁴), and Tat(48-72, Cys⁷²) peptides with pyridyl disulfide **14** afforded the synthetic metallo-peptides Tat(48-72, Cys⁴⁸-La(L)³⁺) **1**, Tat(49-72, Cys⁵⁴-La(L)³⁺) **2**, and Tat(49-72, Cys⁷²-La(L)³⁺) **3**,

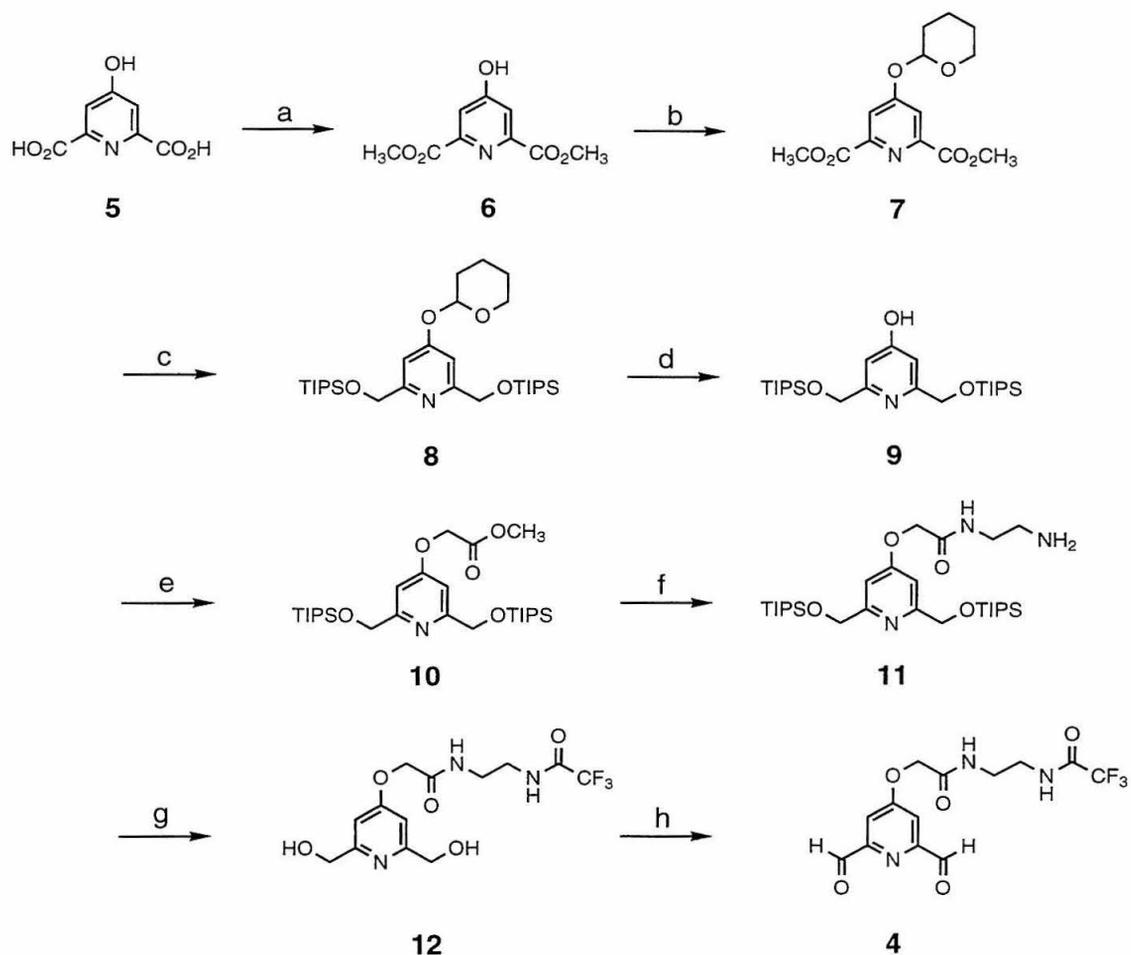


Figure 4.5. Synthesis of dialdehyde **4** from commercially available chelidamic acid (**5**). Reagents: (a) CH_3OH , H_2SO_4 ; (b) dihydropyran, *p*-toluene sulfonic acid, CH_2Cl_2 ; (c) (i) NaBH_4 , EtOH ; (ii) triisopropylsilyl triflate, NEt_3 , CH_2Cl_2 ; (d) oxalic acid, CH_3OH ; (e) $\text{BrCH}_2\text{CO}_2\text{Et}$, Ag_2CO_3 , DMF ; (f) ethylenediamine; (g) (i) tetrabutylammonium fluoride, THF ; (ii) $\text{CF}_3\text{CO}_2\text{Et}$; (h) $\text{Py}\cdot\text{SO}_3$, NEt_3 , DMSO .

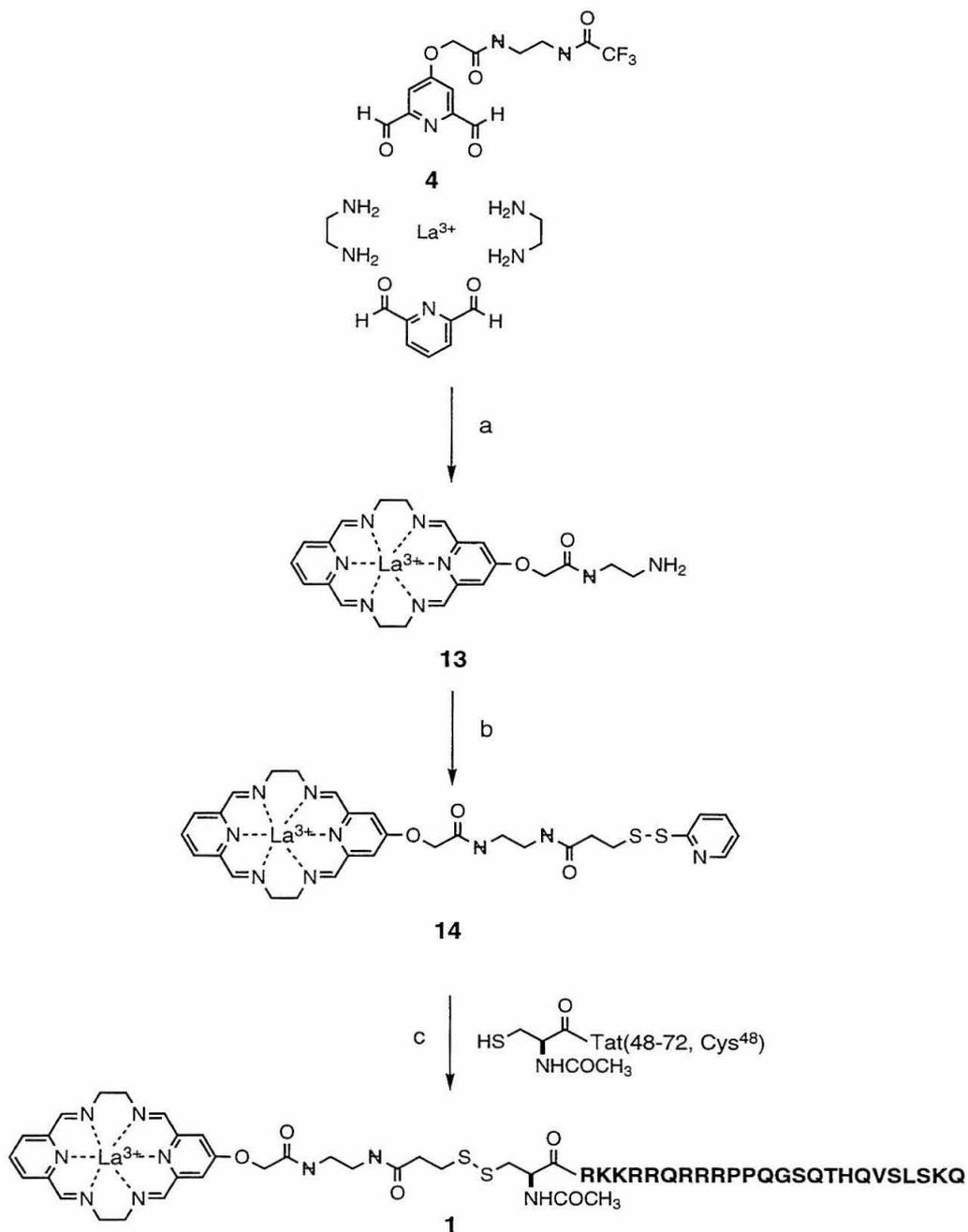


Figure 4.6. Synthesis of Tat-La(L)³⁺ peptides **1-3**. Reagents: (a) (i) CH₃OH, AcOH; (ii) NH₃/H₂O; (iii) Sephadex SP-C25 resin, 100 mM to 500 mM NH₄OAc, pH 4.75; (b) *N*-succinimidyl 3-(2-pyridyldithio)propionate (SPDP), 100 mM NEt₃OAc, pH 7.5; (c) 20 mM Tris·HCl pH 8.0.

respectively, which were purified by reverse-phase HPLC and characterized by amino acid analysis.

Cleavage of TAR RNA by Tat-La(L)³⁺ Peptides. The 58 nucleotide TAR RNA sequence labeled at the 5'-end with ³²P was allowed to react for 10 h with Tat(48-72, Cys⁴⁸-La(L)³⁺) **1** (0.5 μM) at 37 °C (70 mM NaCl, 10 mM Tris·HCl, pH 7.2). Separation of the RNA cleavage products by high resolution denaturing gel electrophoresis reveals that Tat(48-72, Cys⁴⁸-La(L)³⁺) cleaves TAR in the loop with strongest cleavage at G32 and less efficient cleavage at U31, G33, G34, and A35 (Figures 4.7, 4.8). Minor cleavage is observed at C24 and U25 in the trinucleotide bulge and at G16 adjacent to the predicted single base bulge at A17. The same cleavage specificity was observed with Tat(49-72, Cys⁴⁸-La(L)³⁺) **2** and Tat(49-72, Cys⁷²-La(L)³⁺) **3** although with slightly lower cleavage yield. After treatment of TAR with **1** for 10 h the overall cleavage yield was 8%. A plot of $\ln([RNA]_{\text{intact}}/[RNA]_{\text{total}})$ is linear in target RNA concentration with an initial pseudo-first-order rate constant of $5.5 \times 10^{-3} \text{ h}^{-1}$. This corresponds to a half-life of 96 h assuming that **1** undergoes no decomposition at 37 °C, pH 7.2. Control experiments with La(L)³⁺ and La(NO₃)₃ at 0.5 μM concentration show no cleavage of TAR in the presence or absence of the unfunctionalized Tat(49-72) peptide (Figure 4.7, lanes 10-13).

End-Product Analysis. The nature of the termini of the TAR cleavage products resulting from reaction with Tat(48-72, Cys⁴⁸-La(L)³⁺) were determined by comparing the electrophoretic mobility of the fragments with standards of known chemical composition. TAR ³²P-labeled at the 3'-end was treated with Tat(48-72, Cys⁴⁸-La(L)³⁺) (0.5 μM). Analysis of the RNA cleavage products by 20% denaturing gel electrophoresis reveals that they have the same electrophoretic mobility as cleavage products generated by alkaline and RNase T1 digestions, both of which are known to afford 5'-hydroxyl termini (Figure 4.9a). The cleavage products were treated with T4 polynucleotide kinase and ATP which

Figure 4.7. Cleavage of HIV-1 TAR RNA by Tat(48-72, Cys⁴⁸-La(L)³⁺). Autoradiogram of a denaturing polyacrylamide gel of cleavage products from the reaction of Tat(48-72, Cys⁴⁸-La(L)³⁺) with ³²P 5'-end-labeled HIV-1 TAR, pH 7.2, 37 °C. Lane 1, TAR incubated for 10 h. Lane 2, limited alkaline hydrolysis. Lane 3, products of an G-specific RNase T1 digestion. Lanes 4-9 contain Tat(48-72, Cys⁴⁸-La(L)³⁺) which was allowed to react for 0, 2, 4, 6, 8, and 10 h, respectively. Lanes 10 and 11 contain 0.5 mM La(L)³⁺ which was allowed to react for 10 h in the absence and presence of unmodified Tat(49-72) peptide, respectively. Lanes 12 and 13 contain 0.5 mM La(NO₃)₃ which was allowed to react for 10 h in the absence and presence of unmodified Tat(49-72) peptide, respectively.

Tat(48-72, Cys ⁴⁸ -La(L) ³⁺)	5	+	+	+	+	+	+	-	-	-	-
La(L) ³⁺	6	-	-	-	-	-	-	+	+	-	-
Tat(49-72)		-	-	-	-	-	-	-	+	-	+
La(NO ₃) ₃		-	-	-	-	-	-	-	-	+	+

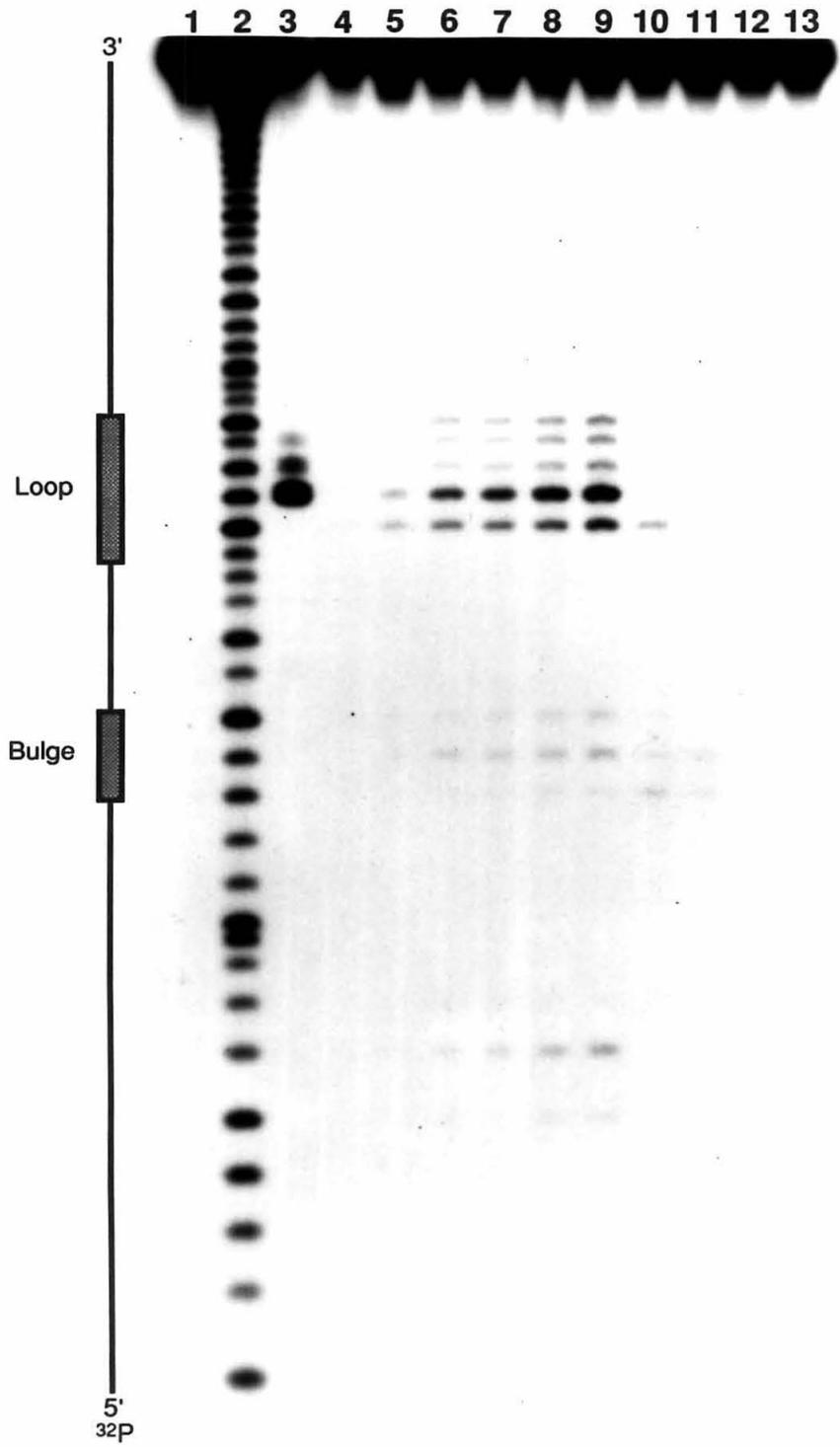
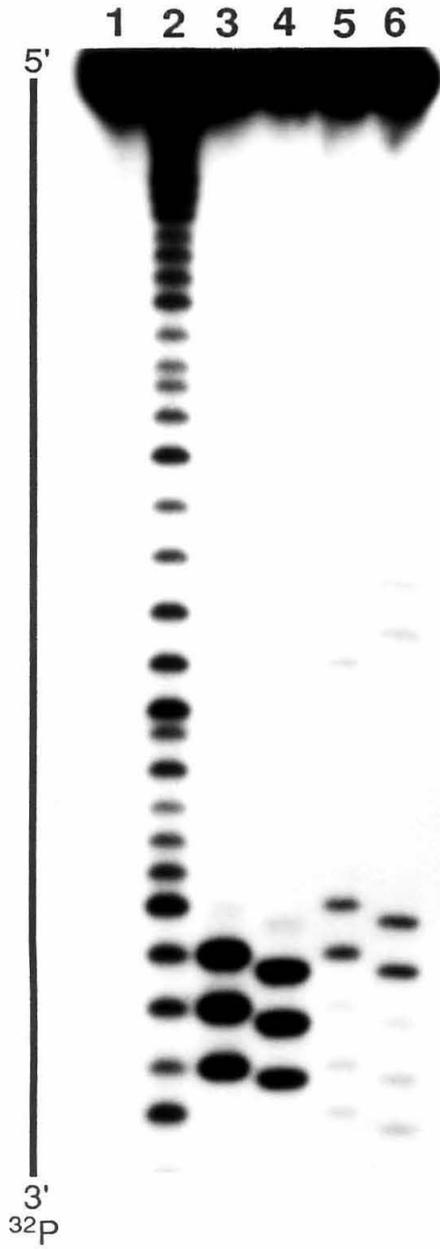
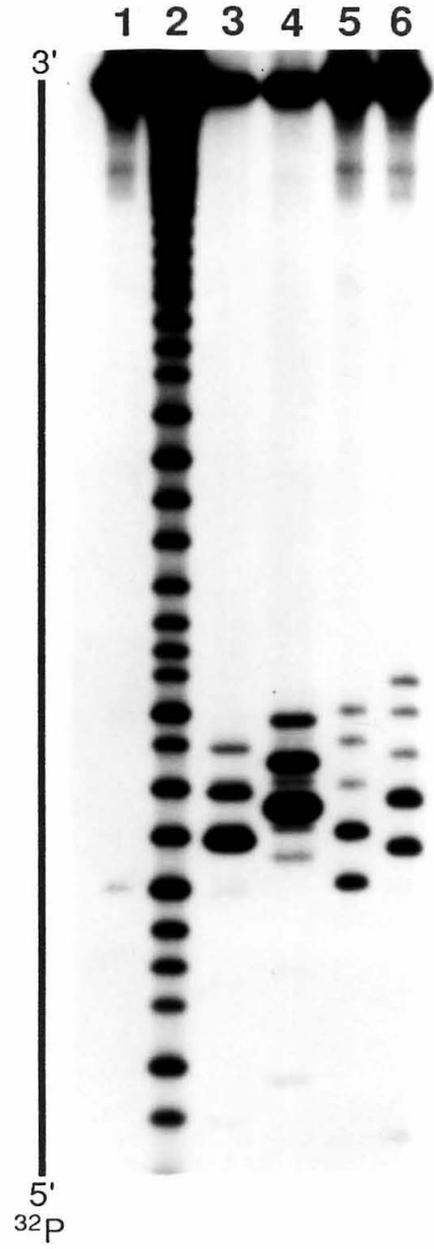


Figure 4.9. Analysis of the termini of the TAR cleavage products resulting from treatment with Tat(48-72, Cys⁴⁸-La(L)³⁺). (A) Autoradiogram of a denaturing polyacrylamide gel analyzing the 5'-end products of TAR cleavage by Tat(48-72, Cys⁴⁸-La(L)³⁺). Lane 1, intact TAR. Lane 2, limited alkaline digestion. Lanes 3 and 4 contain RNase T1 digestions without and with treatment with kinase and ATP, respectively. Lanes 5 and 6 contain Tat(48-72, Cys⁴⁸-La(L)³⁺) cleavage reactions without and with treatment with kinase and ATP, respectively. (B) Autoradiogram of a denaturing polyacrylamide gel analyzing the 3'-end products of TAR cleavage by Tat(48-72, Cys⁴⁸-La(L)³⁺). Lane 1, intact TAR. Lane 2, limited alkaline digestion. Lanes 3 and 4 contain RNase T1 digestions without and with kinase treatment, respectively. Lanes 5 and 6 contain Tat(48-72, Cys⁴⁸-La(L)³⁺) cleavage reactions without and with kinase treatment, respectively.

A**B**

20% denaturing gel electrophoresis reveals that they have the same electrophoretic mobility as cleavage products RNase T1 digestion which produces products cleavage products with phosphorylated 3'-ends (Fig. 4.9b). The cleavage products were treated with T4 polynucleotide kinase which removes phosphate groups at the 3'-termini of RNA molecules.¹⁹ Kinase treatment of the RNA cleavage products generated by Tat(48-72, Cys⁴⁸-La(L)³⁺) and RNase T1 produces TAR fragments with an identical decrease in electrophoretic mobility indicating that cleavage with Tat(48-72, Cys⁴⁸-La(L)³⁺) produces cleavage products with phosphates at the 3'-termini. Since polynucleotide kinase has been shown to hydrolyze both 2'- and 3'-monophosphates as well as 2',3'-cyclic phosphates the 3'-end-products of TAR cleavage by Tat(48-72, Cys⁴⁸-La(L)³⁺) may be 2',3'-cyclic phosphates, 2'-monophosphates, or 3'-monophosphates.²⁰⁻²¹

The presence of 5'-hydroxyl and 3'-phosphate end-products indicates that the RNA cleavage pattern observed with Tat(48-72, Cys⁴⁸-La(L)³⁺) is not due to oxidative scission of the ribose backbone of TAR.²³⁻²⁴ These data are consistent with TAR cleavage by Tat(48-72, Cys⁴⁸-La(L)³⁺) occurring by transesterification involving attack of the 2'-hydroxyl group at the adjacent phosphodiester to afford 5'-hydroxyl products and 2',3'-cyclic phosphates.²⁵ We estimate that Tat(48-72, Cys⁴⁸-La(L)³⁺) accelerates the rate of transesterification by four orders of magnitude above the background rate of RNA transesterification in the absence of catalysts.²⁶ With regard to mechanism we cannot yet distinguish possible energetic contributions from the La(L)³⁺ moiety of Tat(48-72, Cys⁴⁸-La(L)³⁺) in the TAR cleavage reaction beyond a likely contribution from charge neutralization of the phosphodiester.

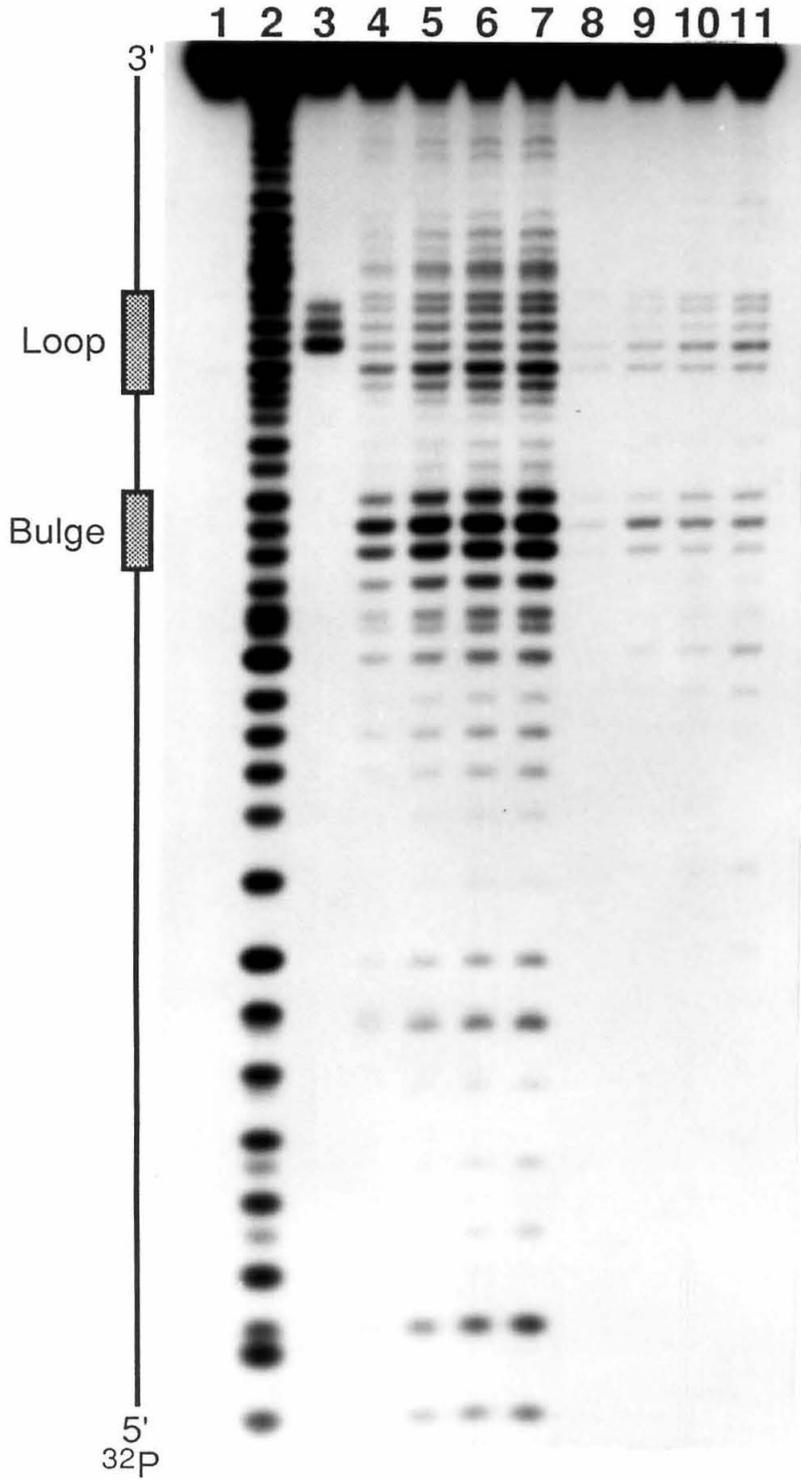
TAR Cleavage with La(NO₃)₃ and La(L)³⁺. At higher concentrations both La(L)³⁺ and La(NO₃)₃ cleave TAR. At 100 μM concentration (70 mM NaCl, 10 mM Tris-HCl, pH 7.2, 37 °C) both catalysts preferentially cleave the bulged nucleotides with

most efficient cleavage observed at C24 in the center of the trinucleotide bulge (Figure 4.10). Both species also cleave the hairpin loop although with lower efficiency than the bulge. Some reactivity in the double-stranded regions of TAR was observed with $\text{La}(\text{NO}_3)_3$ although these structural elements are generally much less susceptible to cleavage than the single-stranded bulge and hairpin (Figure 4.11). Differences between the two catalysts are apparent. The rate of TAR cleavage by $\text{La}(\text{L})^{3+}$ is lower than $\text{La}(\text{NO}_3)_3$ by roughly a factor of five. Also, there is a difference in cleavage specificity in the hairpin loop with maximal cleavage observed at U31 and G32 with $\text{La}(\text{NO}_3)_3$ and $\text{La}(\text{L})^{3+}$, respectively. The differences in cleavage rates and specificity between the two catalysts is most likely due to complexation of the La^{3+} ion by the hexa-aza macrocyclic ligand. The coordinating nitrogen groups in the ligand may attenuate the Lewis acidity of the metal ion which would be expected to reduce the reaction rate relative to uncomplexed La^{3+} . The slight difference in cleavage specificity between the two species may reflect a change in the coordination geometry around the La^{3+} ion in the complex making G32 more susceptible to cleavage than U31.

TAR Cleavage with $\text{La}(\text{NO}_3)_3$ and $\text{La}(\text{L})^{3+}$ in the Presence of Tat(49-72) Peptides. In order to examine the effect of Tat peptides on the selectivity and efficiency of La^{3+} -catalyzed hydrolysis, 5'-end-labeled TAR was treated with 50 μM $\text{La}(\text{NO}_3)_3$ in the presence of increasing concentrations of Tat(49-72). The specificity of La^{3+} -induced hydrolysis of TAR changes in the presence of the peptide (Figure 4.12). The main site of cleavage shifts from C24 to U23 with weaker cleavage observed at C24 and eight additional phosphodiester sites in the presence of Tat(49-72). Since the rate of metal-catalyzed hydrolysis of a phosphodiester is dependent on its accessibility to the catalyst and inherent chemical reactivity, the enhancement in cleavage at U23 in the presence of Tat(49-72) indicates that the RNA has adopted a conformation that makes this position more

Figure 4.10. Cleavage of TAR by $\text{La}(\text{NO}_3)_3$ and $\text{La}(\text{L})^{3+}$. Autoradiogram of a denaturing polyacrylamide gel of cleavage products from the reaction of $\text{La}(\text{NO}_3)_3$ and $\text{La}(\text{L})^{3+}$ with ^{32}P 5'-end-labeled HIV-1 TAR, pH 7.2, 37 °C. Lane 1, TAR incubated for 80 min in the absence of lanthanides. Lane 2, limited alkaline hydrolysis. Lane 3, products of an G-specific RNase T1 digestion. Lanes 4-7 contain 100 μM $\text{La}(\text{NO}_3)_3$ which was allowed to react for 20, 40, 60, and 80 min, respectively. Lanes 8-11 contain 100 μM $\text{La}(\text{L})^{3+}$ which was allowed to react for 20, 40, 60, and 80 min, respectively.

La(NO ₃) ₃	+	+	+	+	-	-	-	-
La(L) ³⁺	-	-	-	-	+	+	+	+
Time (min)	20	40	60	80	20	40	60	80



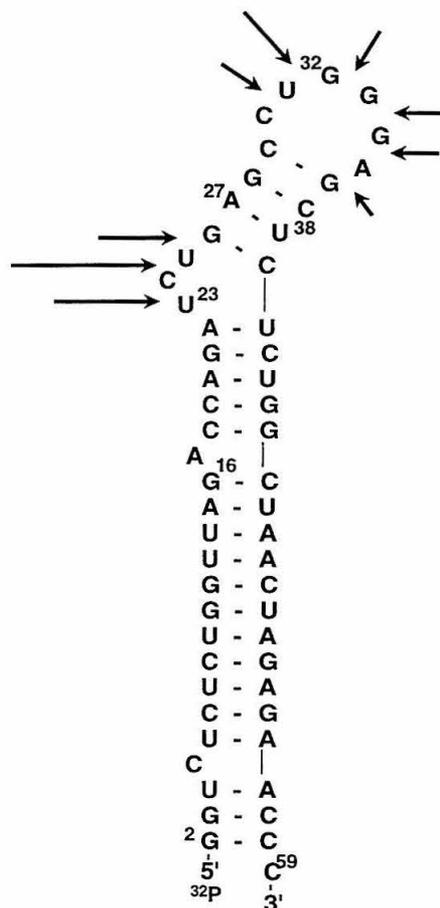


Figure 4.11. Sequence and predicted secondary structure of HIV-1 TAR RNA showing the phosphodiesterases cleaved by $\text{La}(\text{NO}_3)_3$ after 1 h, 37 °C, pH 7.2. Arrow heights are proportional to cleavage intensities at the indicated phosphodiesterases.

susceptible to hydrolytic cleavage in the Tat(49-72)-TAR complex. A similar change in cleavage specificity was also observed in experiments with $\text{La}(\text{L})^{3+}$. Similar footprinting of the Tat(48-72, Cys⁴⁸- $\text{La}(\text{L})^{3+}$)-TAR complex with 100 μM $\text{La}(\text{NO}_3)_3$ and $\text{La}(\text{L})^{3+}$ also showed the shift in the primary cleavage site from C24 to U23 upon peptide binding indicating that the functionalized peptide is binding to TAR.

Competition Experiments. Experiments in which 5'-³²P end-labeled TAR was preincubated with 0.1 μM to 200 μM unfunctionalized Tat(49-72) peptide prior to addition

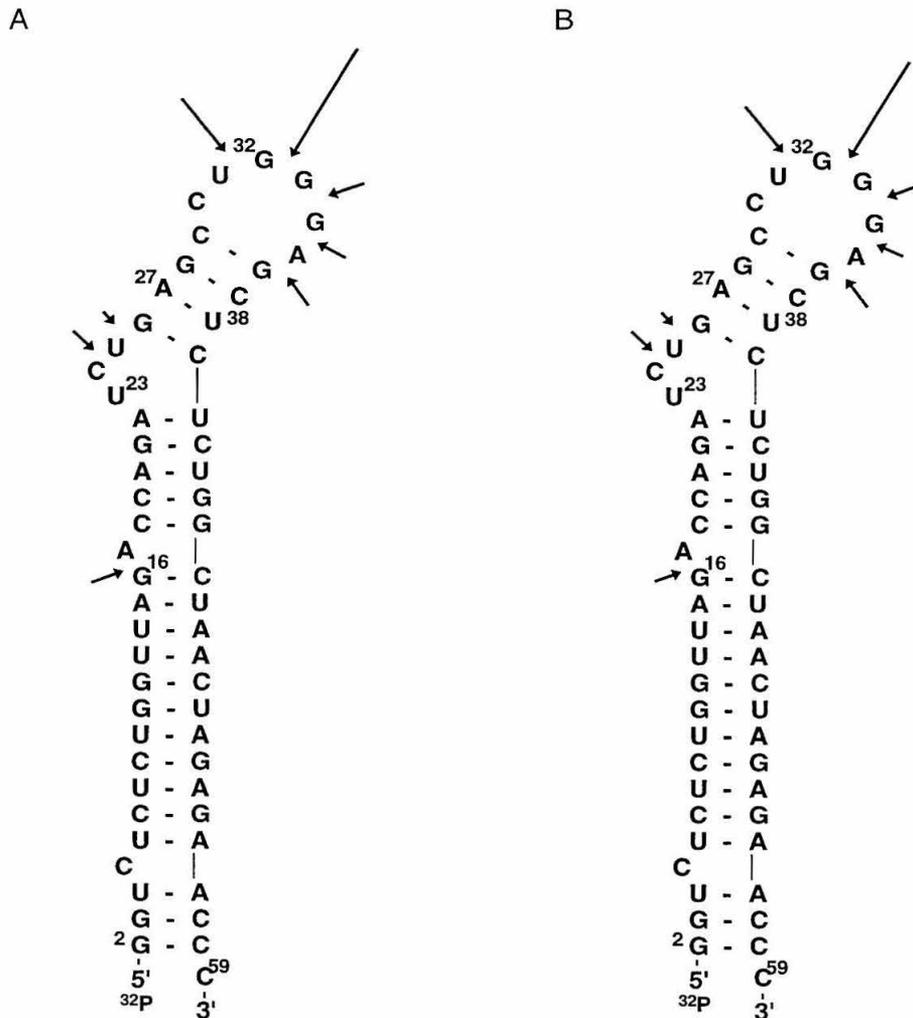


Figure 4.13. Cleavage of TAR by Tat(48-72, Cys⁴⁸-La(L)³⁺) (70 mM NaCl, 10 mM Tris•HCl, pH 7.2, 37 °C, 10 h) in (A) the absence and (B) the presence of unfunctionalized Tat(49-72) at 100 μ M concentration. Arrow heights are proportional to cleavage intensities at the indicated phosphodiester.

of bound **1**. In order to test this idea further, a derivative of TAR RNA which is lacking the trinucleotide bulge (**15**) was prepared by chemical synthesis (Figure 4.14a). Experiments have shown that the binding affinity of Tat peptides for this RNA is reduced by a factor of 120 compared to the corresponding RNA possessing the 5'-UCU-3' bulged nucleotides.⁶ Treatment of 5'-³²P end-labeled RNA **15** with Tat(48-72, Cys⁴⁸-La(L)³⁺) (0.5

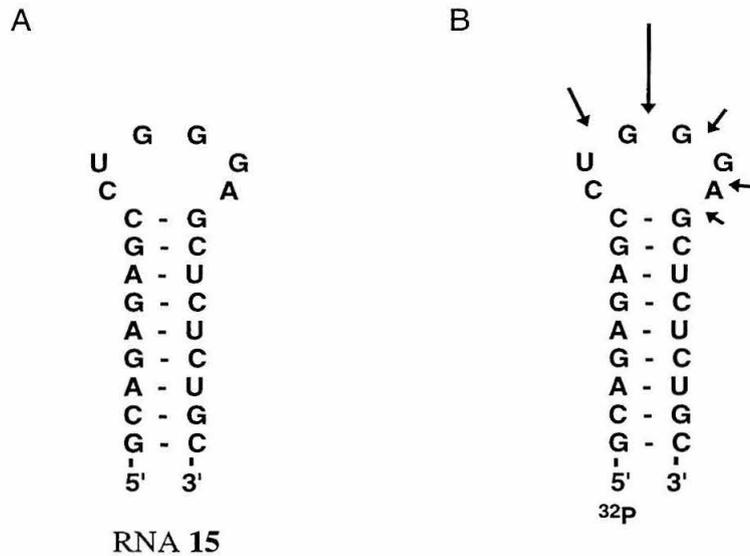


Figure 4.14. Cleavage of a bulgeless derivative of TAR RNA by Tat(48-72, Cys⁴⁸-La(L)³⁺) at pH 7.2, 37 °C. (A) Predicted secondary structure of RNA 15. (B) Cleavage pattern observed with 0.5 μ M Tat(48-72, Cys⁴⁸-La(L)³⁺) after 10 h. Arrow heights are proportional to cleavage intensities at the indicated phosphodiester.

μ M) (70 mM NaCl, 10 mM Tris-HCl, pH 7.2, 37 °C) for 10 h resulted in cleavage of the loop with *the same specificity and yield at the same concentration of Tat-La(L)³⁺* as observed with full-length native TAR (Figure 4.14b). Preincubation of RNA 15 with up to 200 μ M had no effect on the observed cleavage efficiency with Tat(48-72, Cys⁴⁸-La(L)³⁺).

Discussion

TAR Cleavage by Tat-La(L)³⁺ Peptides. In spite of the fact that TAR cleavage with Tat-La(L)³⁺ peptides **1-3** occurs at the kinetically less reactive loop nucleotides to afford reaction products consistent with a transesterification/hydrolysis mechanism, we do not believe this observed cleavage is a result of a bound Tat-La(L)³⁺ peptide. The fact that peptides **1-3** show the same cleavage specificity and yield is surprising given that the La(L)³⁺ moiety is attached to the peptide at three quite different positions in the amino acid

sequence. Furthermore, the results of the cleavage inhibition experiments with TAR and bulgeless RNA **15** are inconsistent with models invoking RNA cleavage from a sequence-specifically bound Tat-La(L)³⁺ peptide.

The TAR cleavage experiments with La(NO₃)₃ and La(L)³⁺ in the presence and absence of Tat(49-72) peptides rule out metallo-peptide decomposition as a possible explanation for the observed loop cleavage. Decomposition of the Tat-La(L)³⁺ peptides by either hydrolysis of the metal complex imines to afford free La³⁺ ion or by disulfide bond reduction to afford the corresponding untethered La(L)³⁺ complex. In either case, however, we would see cleavage in the bulge and not the loop since La³⁺ and La(L)³⁺ complexes are highly specific for the bulge in both unliganded TAR and in the Tat(49-72)-TAR complex. We note explicitly that in all of our experiments with untethered lanthanide salts or (L)³⁺ complexes we observe highly specific bulge cleavage. The only reagents that display a preference for the loop phosphodiester are Tat-La(L)³⁺ peptides **1-3**.

One explanation for the observed TAR cleavage is that coupling a Tat(49-72) peptide to a La(L)³⁺ complex inhibits binding to and hence cleavage of the phosphodiester in the TAR trinucleotide bulge affording the observed specificity for the loop for peptides **1-3**. This would explain the fact that RNA **15** is cleaved by peptide **1** in spite of the fact that the pyrimidine bulge has been deleted. Another possibility involves initial binding of Tat-La(L)³⁺ peptide to the Tat binding site proximal to the pyrimidine bulge (Figure 4.15). Although the Tat-TAR complex forms we propose that no RNA cleavage results from this interaction. Another equivalent of Tat-La(L)³⁺ approaches a Tat-bound TAR molecule in solution. The inherent specificity of the La(L)³⁺ moiety for the bulge is altered in the Tat-La(L)³⁺ reagent because of *repulsive electrostatic interactions* between the positively charged Tat(49-72) fragments of the sequence-specifically bound and the free Tat-La(L)³⁺ peptides. Since interaction of the La(L)³⁺ complex is energetically unfavorable, only

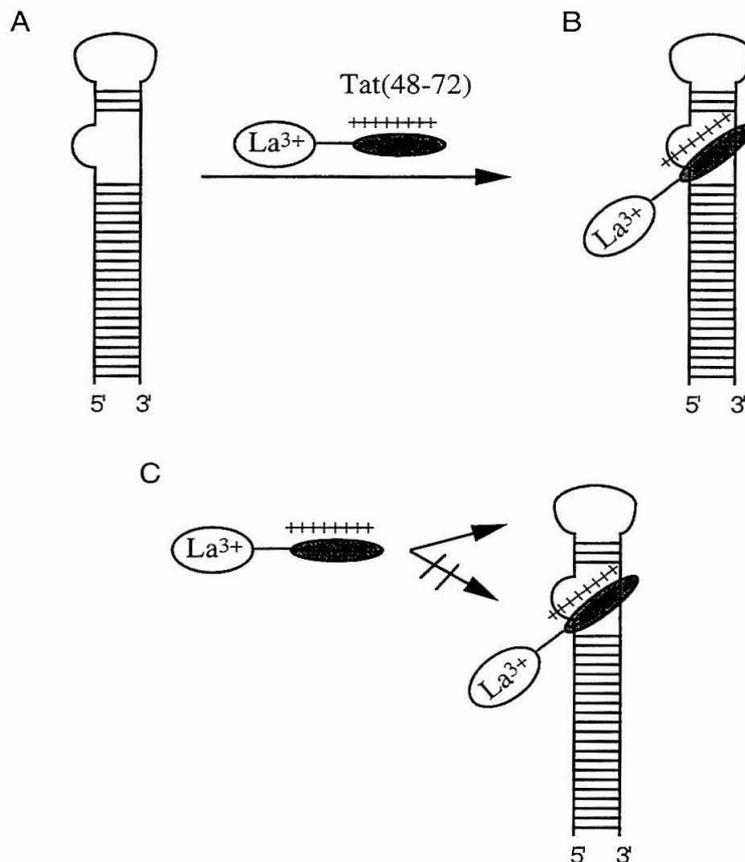


Figure 4.15. One explanation for the loop-specific cleavage of TAR by peptides **1-3**. (A), (B) A Tat-La(L)³⁺ peptides binds the RNA to form the Tat-TAR complex but the bound metallo-peptide does not direct cleavage of the RNA. (C) Another equivalent of a Tat-La(L)³⁺ peptide approaches the Tat-TAR complex. Repulsive electrostatic interactions between the free and bound Tat-La(L)³⁺ peptides disfavors interaction of the La(L)³⁺ moiety of the free Tat-La(L)³⁺ with the trinucleotide bulge and cleavage is only observed in the loop.

cleavage of the loop phosphodiester is observed. As the observed cleavage of TAR RNA by peptides **1-3** was not sequence-specific our interest in this project was terminated and the mechanism of loop cleavage by peptides **1-3** was not further characterized.

Sequence Specific RNA Hydrolysis by Artificial Catalysts. Since the completion of this work several reports of sequence specific RNA hydrolysis by designed

catalysts have appeared in which metal complexes are covalently tethered to a deoxyribonucleotide and used to cleave a complementary RNA (Figure 4.16). Bashkin and coworkers report that a terpyridyl•Cu²⁺ complex tethered to C-5 of a thymidine base can cleave a complementary RNA (18-25% yield, 45 °C, 72 h).²⁷ Komiyama and coworkers report that an iminodiacetate-Lu³⁺ complex can be used to direct the hydrolysis of a complementary RNA in modest yield (17%, 37 °C, 8 h).²⁸ The utility of both of these approaches may be limited as they require the use of kinetically unstable metal complexes to effect RNA hydrolysis. In a significant advance, Magda and coworkers report that a kinetically stable expanded porphyrin complex (texaphyrin) of Eu³⁺ was capable of site-directed RNA hydrolysis when tethered to the 5' end of a deoxynucleotide (30%-60% yield, 37 °C, 18-24 h).²⁹ More recently, Haner and coworkers disclosed that kinetically stable octa-aza macrocyclic Eu(III) complexes are capable of near-quantitative yield of complementary RNA hydrolysis at physiological pH and temperature.³⁰ These results clearly demonstrate that stable complexes of lanthanide ions can be covalently attached to RNA-binding ligands to effect sequence-specific RNA hydrolysis. It is not understood why hexa-aza Schiff base lanthanide macrocycles tethered to Tat(49-72) or deoxynucleotides³¹ fail to sequence-specifically cleave TAR or complementary RNAs, respectively.

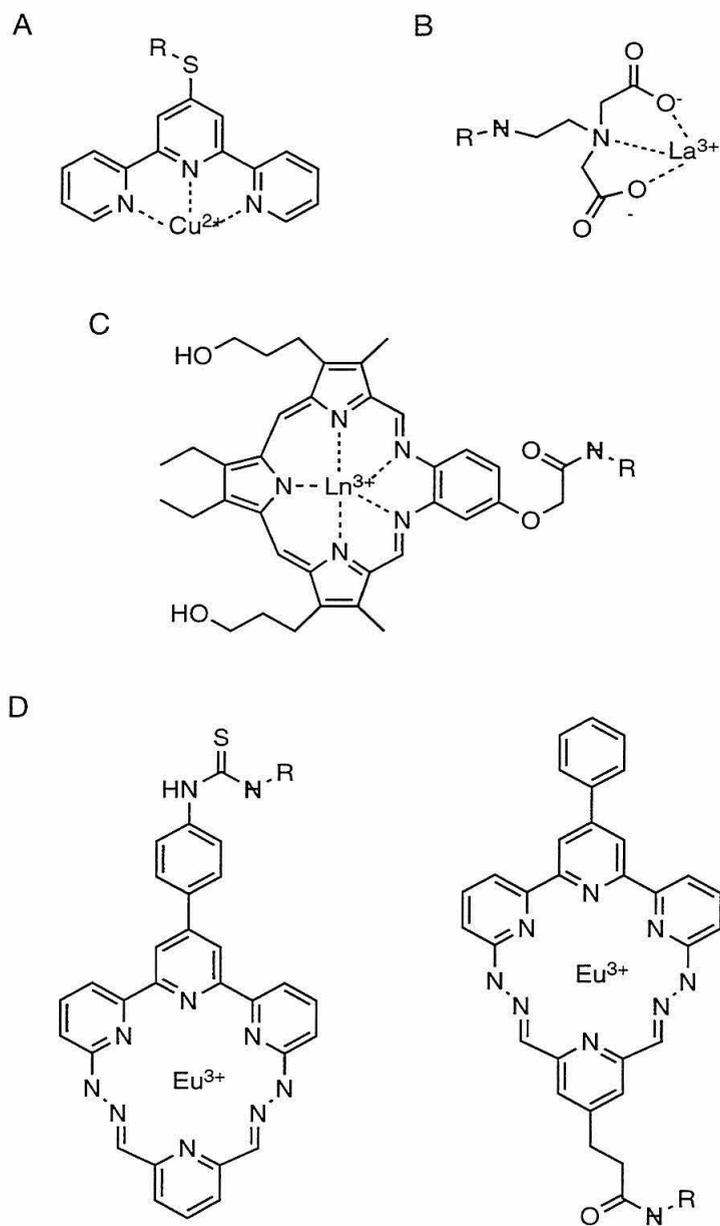


Figure 4.16. Designed catalysts for sequence-specific RNA hydrolysis. (A) Terpyridyl •Cu(II) complex reported by Bashkin and coworkers.²⁷ (B) Iminodiacetate•Lu(III) complex reported by Komiyama and coworkers.²⁸ (C) Texaphyrin derivative reported by Magda and coworkers.²⁹ (D) Octa-aza-Eu(III) complexes reported by Haner and coworkers.³⁰

Experimental Section

Organic Synthesis. ^1H and ^{13}C NMR spectra were recorded at 300 MHz on a General Electric-QE 300 NMR spectrometer. High-resolution mass spectra (HRMS) were recorded using electron ionization (EI) or fast atom bombardment (FAB) techniques at the Mass Spectrometry Laboratory at the University of California, Riverside. Ultraviolet-visible spectra were recorded on a Hewlett-Packard 8452 A diode array spectrophotometer. All reactions were conducted under an atmosphere of argon unless noted otherwise. Pyridine-sulfur trioxide complex and DMSO were purchased from Fluka. *N*-Succinimidyl 3-(2-pyridyldithio)propionate (SPDP) was purchased from Pierce. Tetrahydrofuran (THF) was distilled under nitrogen from sodium/ benzophenone ketyl. Dichloromethane and triethylamine were distilled from calcium hydride. All other reagents were purchased from Aldrich. Sephadex-SP C25 resin was obtained from Aldrich and stored in 100 mM NH_4OAc , pH 4.75 for 8 h before use. Flash column chromatography was carried out using EM Science Kieselgel 60 (230-400 mesh).³² Thin-layer chromatography was performed on EM Reagents silica gel plates (0.5 mm thickness). All compounds were visualized at 254 nm.

Dimethyl ester (6). A slurry of 25 g (124 mmol) of chelidamic acid (**5**) in 250 mL of methanol was treated with 15 mL concentrated sulfuric acid (caution: exothermic) and heated at reflux for 8 h. After cooling to ambient solution 750 mL of dichloromethane were added and the resulting solution was washed with water (3 x 500 mL), dried with anhydrous sodium sulfate, filtered, and concentrated under reduced pressure to afford 11.94 g (26%) of diester **6** that was used immediately.

Tetrahydropyranyl Ether (7). A solution of 11.94 g (52.1 mmol) of dimethyl ester **6** in 360 mL dichloromethane was treated with 20.4 g (243 mmol, 4.65 equiv.) of dihydropyran, and 570 mg (3.0 mmol, 0.06 equiv.) *p*-toluenesulfonic acid. After stirring 5

h at ambient temperature the reaction mixture was partitioned between 200 mL of saturated aqueous sodium bicarbonate and 200 mL of ether. The layers were separated and the aqueous phase was extracted with ether (3 x 100 mL) and the combined organic layers were washed with brine, dried over anhydrous sodium sulfate, filtered, and concentrated under reduced pressure. Purification by chromatography on silica gel (2:1 hexane/ ethyl acetate) afforded 10.35 g (72%) of **7** as yellow oil. ¹H NMR (300 MHz, CDCl₃) 7.90 (s, 2 H, ArH), 5.65 (m, 1 H, CHO), 4.01 (s, 6 H, CH₃), 3.78-3.65 (m, 1 H, CH₂O), 2.0-1.5 (m, 6 H, CH₂).

Bis-Silyl Ether (8). A solution 10.35 g (37.3 mmol) of diester **7** in 400 mL of ethanol was treated with 14.1 g (373 mmol, 10 equiv.) of sodium borohydride. The solution was heated at reflux for 6 h before it was cooled and treated with 200 mL 1 N sodium tartrate. After removal of the ethanol under reduced pressure the resulting concentrate was partitioned between 200 mL 1 N KH₂PO₄ and 200 mL dichloromethane. The layers were separated and the aqueous phase was extracted with dichloromethane (3 x 100 mL) and the combined organic layers were washed with brine, dried over anhydrous sodium sulfate, filtered, concentrated under reduced pressure, and azeotroped with acetonitrile (2 x 50 mL) to afford the diol that was used without further purification. To a solution of the diol in 125 mL dichloromethane cooled to 0 °C were added 11.5 mL (82.2 mmol, 3.0 equiv.) of triethylamine and 16.9 g (54.4 mmol, 2.0 equiv.) and stirred 1 h at 0 °C. The solution was quenched with 200 mL 1 N KH₂PO₄ and 300 mL dichloromethane. The layers were separated and the aqueous phase was extracted with dichloromethane (3 x 100 mL) and the combined organic layers were washed with brine (150 mL), dried over anhydrous sodium sulfate, filtered, concentrated under reduced pressure. Purification by chromatography on silica gel (2:1 hexane/ ethyl acetate) afforded 12.4 g (81.4%) of **8** as yellow oil. ¹H NMR (300 MHz, CDCl₃) 7.15 (s, 2 H, ArH), 5.48 (m 1 H, OCHO), 4.85

(s, 4H, CH₂), 3.90-3.82 (m, 1 H, CH₂), 3.80-3.72 (m, 1 H, CH₂), 2.05-1.35 (m, 6 H, CH₂), 1.30-0.80 (m, 42 H, OSi(CH(CH₃)₂)₃).

Hydroxypyridine (9). A solution of 12.4 g (22.5 mmol) of **8** 350 mL dry methanol was treated with 7.88 g (113 mmol, 5 equiv.) of oxalic acid. The solution was stirred at ambient temperature 1.5 h during which time a white crystalline solid precipitated. The solid was collected by filtration, washed with ice-cold methanol and dried under reduced pressure to afford 9.31 g (88.4%) of hydroxypyridine **9**. ¹H NMR (300 MHz, CDCl₃) 6.75 (s, 2 H, ArH), 6.85 (s, 4 H, CH₂), 1.2-1.9 (m, 42 H, OSi(CH(CH₃)₂)₃).

Ester (10). A suspension of 5.26 g Ag₂CO₃ (38.2 mmol, 2.1 equiv.) and 9.31 g (19.1 mmol) of hydroxypyridine **9** 100 mL DMF at 70 °C was stirred for 10 min before 9.1 mL of α-methyl bromoacetate was added. The mixture stirred 2 h at 70 °C before it was cooled to room temperature and filtered through celite. The filtrate was partitioned between 250 mL 1:1 ether:dichloromethane and 250 mL water and filtered through celite. The layers were separated and the aqueous phase was extracted with ether (3 x 100 mL) and the combined organic layers were washed with brine (100 mL), dried over anhydrous sodium sulfate, filtered, concentrated under reduced pressure. Purification by chromatography on silica gel (2:1 hexane/ ethyl acetate) afforded 8.7 g (85%) of **10** as a clear, colorless liquid. ¹H NMR (300 MHz, CDCl₃) 7.0 (s, 2 H, ArH), 4.85 (s, 4 H, CH₂), 3.80 (s, 3 H, OCH₃), 1.20-1.00 (m, 42 H, OSi(CH(CH₃)₂)₃). ¹³C NMR (CDCl₃) 169, 166, 104, 67, 64, 52, 18, 11.

Diol (12). A solution of 920 mg (1.70 mmol) of ester **10** in 5 mL ethylenediamine was stirred 4 h at ambient temperature before it was partitioned between ether (30 mL) and water (30 mL). The layers were separated and the aqueous phase was extracted with ether (3 x 30 mL) and the combined organic layers were washed with brine, dried over anhydrous sodium sulfate, filtered, and concentrated under reduced pressure to afford

amine **11** as a pale yellow oil. A solution of unpurified **11** in 5 mL THF was treated with 4.25 mL (4.25 mmol, 4.5 equiv.) of a 1 M solution of tetrabutylammonium fluoride in THF at 0 °C and stirred for 30 min. The solution was allowed to warm to ambient temperature and stir for an additional 2 h and then treated with 2 mL of ethyl trifluoroacetate and stirred for an additional 2 h. The solution was concentrated under reduced pressure and purified by chromatography on silica gel (10:1 CH₃CN/ H₂O) to afford 339 mg (59%) of diol **12** that was used immediately.

Aldehyde (4). A solution of 339 mg (1.00 mmol) of diol **12** and 1.40 mL of triethylamine (10.0 mmol, 10.0 equiv.) in 5 mL DMSO and 5 mL dichloromethane at 0 °C was treated dropwise via cannula with a solution of 954 mg (6 mmol, 6.0 equiv.) of pyridine-sulfur trioxide complex in 5 mL DMSO. The solution was stirred at 0 °C for 2 h before it was partitioned between 10 mL of 0.1 NaHSO₄ and 20 mL of 3:1 ether:dichloromethane. The aqueous layer was extracted with ether (3 x 20 mL) and the combined organic layers were washed with brine, dried over anhydrous sodium sulfate, filtered, and concentrated under reduced pressure. Purification by chromatography on silica gel (1:1 CH₂Cl₂/ ethyl acetate) afforded 200 mg (62%) of **4**. ¹H NMR (300 MHz, CDCl₃) 10.1 (s, 2 H, CHO), 8.60 (br s, 1 H, NH), 8.05 (br s, 1 H, NH), 7.75 (s, 2 H, ArH), 4.80 (s, 2 H, CH₂), 3.55-3.45 (m, 4 H, CH₂).

Amine (13). A solution of 20 mg (57.6 μmol) dialdehyde **4**, 86 mg (636.5 μmol, 11 equiv.) 1,6-pyridinedicarboxaldehyde, and 149.6 mg (345.6 μmol, 6 equiv.) of lanthanum nitrate hexahydrate in 10 mL methanol was treated with 40.8 mg (mmol, equiv.) of ethylenediamine and stirred 10 min during which time a white solid precipitated. Three drops of acetic acid were added and the solution was heated at reflux 4 h. After cooling to ambient temperature the solid was isolated by filtration and washed with cold methanol. The solid was dissolved in 3 mL of water and treated with 150 μL of

concentrated aqueous ammonium hydroxide. After 2 h the solution was concentrated to dryness under reduced pressure, dissolved in 4 mL of water, and applied to a column of 350 mg Sephadex-SP C-25 resin equilibrated with 100 mM NH_4OAc , pH 4.75. Using 100 mM NH_4OAc , pH 4.75 as the eluant, fractions were collected and monitored by UV as the unsubstituted La(L)^{3+} complex eluted from the column. Then a gradient of 180 to 500 mM NH_4OAc , pH 4.75 was used to isolate 20 mg (42.8%) of complex **13** after lyophilization. $^1\text{H NMR}$ (D_2O) 8.82 (s, 2 H, NCH), 8.76 (s, 2 H, NCH), 8.37 (t, 1 H, ArH), 8.10 (d, 2 H, ArH), 7.63 (s, 2 H, ArH), 4.88 (s, 2 H, OCH_2), 4.14 (s, 8 H, NCH_2) 3.62 (t, 2 H, NHCH_2), 3.36 (s, 1 H, NH), 3.20 (t, 2 H, CH_2NH_3^+), 1.80 (s, 12 H, OCOCH_3).

Pyridyl Disulfide (14). A solution of amine 54 mg (0.17 mmol) amine **13** in 450 μL of 100 mM NEt_3OAc , pH 7.5 was treated with a solution of 75 mg (0.44 mmol, 2.5 equiv.) SPDP in 225 μL CH_3CN . The resulting solution was vigorously mixed for 3 h at ambient temperature when it was concentrated under reduced pressure. The resulting solid was dissolved in 6 mL 100 mM NH_4OAc , pH 4.75 and purified in three portions by C_8 RP-HPLC. A gradient of 100% A to 80% B over 60 min (A:100 mM NH_4OAc , pH 4.75, B: 40% 100 mM NH_4OAc , pH 4.75/ 60% CH_3CN) with a flow rate of 2 mL/min was used to isolate the desired disulfide **14** eluting at 50% B. Lyophilization afforded 16 mg (25%) of **14**. $^1\text{H NMR}$ (D_2O) 8.80 (s, 1 H NCH), 8.7 (a, 1 H, NCH), 8.38 (t, 2 H, ArH), 8.34 (d, 1H, ArH), 8.08 (d, 2 H, ArH), 7.78 (m, 2 H, ArH), 7.58 (s, 2 H, ArH), 7.23 (br t, 1 H, ArH), 4.83 (s, 2 H, OCH_2), 4.16 (br s, 8 H, NCH_2), 3.48 (m, 4 H, $\text{COCH}_2\text{CH}_2\text{S}$), 3.04 (t, 2 H, NHCH_2), 2.64 (t, 2 H, CH_2NH_3^+), 1.82 (s, 9 H, OCOCH_3).

Reagents for Peptide Synthesis. Protected amino acid derivatives were purchased from Peninsula Laboratories. *t*-Boc-L-His(DNP) was purchased from Fluka. Benzhydrylamine resin (BHA) was purchased from U.S. Biochemical Corp. *N,N*-

Dimethylformamide (DMF) and diisopropylethylamine (DIEA), and trifluoroacetic acid (TFA) were purchased from Applied Biosystems. Dicyclohexylcarbodiimide (DCC) was purchased from Peptides International. *N*-Hydroxybenzotriazole (HOBt), *p*-cresol, and *p*-thiocresol were from Aldrich. Dichloromethane (DCM) and methanol (HPLC grade) were purchased from Mallinkrodt. Diethyl ether (low peroxide content) was obtained from Baker. Mass spectrometry and quantitative amino acid analysis was performed at the Biomolecule Analysis Facility at the California Institute of Technology.

Manual Peptide Synthesis. Manual peptide synthesis was carried out in 20-mL vessels fitted with coarse glass frits as described by Kent.³³⁻³⁴ N^{α} -*t*-Boc-L-amino acids were used with the following side chain protecting groups: Arg(Tos), Lys(Cl-Z), Cys(*p*-MeBzl), Thr(Bzl), and His(DNP). Manual assembly of the protected peptide on the solid support was carried out at 25 °C via a three-step reaction cycle. First the Boc protecting group was removed from the α -amino group of the resin-bound amino acid by using TFA (65% TFA in DCM for 1 and 15 min). The deprotected peptide resin was then neutralized with 10% DIEA in DCM (2 x 1 min). All amino acids were coupled to the free α -amino group as the HOBt esters. Coupling yields were determined by quantitative ninhydrin monitoring with acceptable values being $\geq 99.5\%$. Average coupling yields were 99.8%. If a second coupling was necessary, the resin was neutralized with 10% DIEA in DMF. The HOBt ester was formed outside the reaction vessel in an absolute minimum of DCM/DMF. The solution was filtered into the vessel to remove the dicyclohexylurea, topped off with DMF, and allowed to react for 1-2 hours. If the yield was not acceptable at this point, a third coupling was performed by the same protocol as the second coupling.

Peptide Deprotection and Purification. Samples of the peptide resin (150-300 mg) were deprotected using HF and *p*-cresol and *p*-thiocresol as scavengers as previously described.³³⁻³⁴ Crude peptides were purified by C_8 semi-preparative RP-HPLC as

described using a 0-60% acetonitrile/H₂O/0.1% TFA gradient.³⁴ Purified peptides were quantitated using the BCA Protein Assay (Pierce) according to the manufacturer's instructions using BSA as a standard.

Tat-La(L)³⁺ Peptide Synthesis. Tat(49-72)-La(L)³⁺ peptides **1-3** were all synthesized in the same manner. An aqueous solution containing 160 nmol of purified Tat peptide is concentrated under reduced pressure and redissolved in 225 μ L of 22.2 mM Tris-HCl, pH 8.0. A aqueous solution of 568 nmol (3.55 equiv.) of pyridyl disulfide **4** in 17 μ L is added and the resulting solution allowed to stand at ambient temperature. After 3 h the solution is diluted with 3 mL of water and purified by RP-HPLC using a C₈ semi-prep column and a gradient of 100% A to 45% B over 90 min (A: 0.1% TFA in H₂O, B: 0.1% TFA, 40% H₂O/60% CH₃CN) with a flow rate of 2 mL/min. Fractions were collected and analyzed by C₈ analytical RP-HPLC using a gradient of 100% A to 30% B over 30 min (A: 0.1% TFA in H₂O, B: 0.1% TFA, 40% H₂O/60% CH₃CN) with a flow rate of 1 mL/min. Fractions containing the desired material were pooled, diluted with 500 mM NH₄OAc, pH 7.0 to a final concentration of 100 mM NH₄OAc, frozen, and lyophilized. In order to ensure complete removal of NH₄OAc, the solids were lyophilized 3 times from 1.5 mL of water. Concentrations were determined by quantitative amino acid analysis. Peptides were stored as aqueous solutions at -20 °C.

HIV-1 TAR RNA Preparation. HIV-1 TAR was synthesized by run-off transcription with T7 RNA polymerase using a DNA template and primer synthesized by automated synthesis using β -cyanoethyl phosphoramidites as previously described.³⁵ Transcribed TAR was purified by denaturing polyacrylamide gel electrophoresis and stored as an aqueous solution at -70 °C. RNA concentrations were quantitated by UV spectroscopy using the following molar extinction coefficients: 15,400 (A), 11,700 (G), 7,300 (C), 9900 (U) cm⁻¹M⁻¹. After treatment with calf alkaline phosphatase to remove the

terminal triphosphate, TAR was labeled at the 5'-end with T4 polynucleotide kinase and [γ - 32 P]-ATP as described and purified on a 15% denaturing polyacrylamide gel. TAR was 3'-end labeled with cytidine 3',5'-[32 P]-bisphosphate and T4 RNA ligase as described.³⁶ End-labeled TAR was dissolved in 70 mM NaCl, 10 mM Tris-HCl, pH 7.5 (22 °C) and heated to 90 °C for 3 min and allowed to slowly cool to ambient temperature in order to form the proper secondary structure as described by Celander and Cech.³⁷ End-labeled RNA was stored in 70 mM NaCl, 10 mM Tris-HCl, pH 7.5 at -20 °C.

RNA Chemical Synthesis. Chemical RNA synthesis was performed on an Applied Biosystems 394 synthesizer using 2'-OTBDMS- β -cyanoethyl phosphoramidites base protected with the PAC (rR and rG) or Bz (rC) groups (Biogenix). Following automated synthesis base protecting groups were removed by treatment with ammonia-saturated methanol for 24 h at room temperature. 2'-OTBDMS groups were removed by treatment with 0.1 M tetrabutylammonium fluoride in THF for 24 h at room temperature. THF was removed by lyophilization and the remaining residue was dissolved in 1 mL of water and lyophilized again to remove all traces of THF. The residue was dissolved in 3 mL of water and dialyzed against 10 mM Tris•HCl, pH 7.2 for 24-48 h at 4 °C with buffer changes every 4-8 h. Following dialysis the solution was concentrated by lyophilization and dissolved in formamide-TBE loading buffer and electrophoresed on a 1.4 mm thick 20% denaturing polyacrylamide gel. The RNA was visualized by UV shadowing and excised. The gel slice was crushed and soaked in 500 mM NH₄OAc, pH 7.0 at 37 °C for 12-18 h. The isolated RNA was ethanol precipitated, dissolved in 400 μ L of water and desalted using a NAP-5 column (Pharmacia). RNA concentrations were quantitated by UV spectroscopy using the following molar extinction coefficients: 15,400 (A), 11,700 (G), 7,300 (C), 9900 (U) cm⁻¹M⁻¹.

RNA Cleavage Reactions. Reaction conditions were 70 mM NaCl, 10 mM Tris-HCl, pH 7.2, 37 °C, and approximately 40,000 cpm of 5' end-labeled RNA in a total volume of 20 µL containing the appropriate concentration of cleaving reagent. Reactions were allowed to proceed for the appropriate length of time and terminated by ethanol precipitation. Samples were resuspended in formamide-EDTA loading buffer, heated to 90 °C for 3 min, ice-cooled, and electrophoresed on a 15% or 20%, 1:20 cross-linked, 7 M urea denaturing polyacrylamide gel for 5-6 h at 40 V/cm. Following electrophoresis gels were wrapped in plastic film and quantitated by storage phosphor autoradiography.

End-Product Analysis. Reaction conditions for TAR cleavage by Tat(48-72, Cys⁴⁸-La(L)³⁺) (0.5 µM) were 70 mM NaCl, 10 mM Tris-HCl, pH 7.2, and approximately 200,000 cpm of end-labeled RNA in a total volume of 20 µL. Cleavage reactions were incubated at 37 °C for 12 h and terminated by ethanol precipitation. The presence of 5'-hydroxyl groups was assayed by dissolving the 3'-end-labeled TAR cleavage products in 10 µL of buffer containing 10 mM ATP, 50 mM Tris-HCl, 10 mM MgCl₂, 5 mM DTT, 100 mM spermidine, 100 mM EDTA, 5% glycerol, pH 9.5, and 5 units of 3'-phosphatase-free T4 polynucleotide kinase. The reactions were incubated at 37 °C for 1 h and terminated by ethanol precipitation. The presence of phosphates at the 3'-termini was assayed by dissolving the 5'-end-labeled cleavage products in 10 mL of buffer containing 20 mM Tris-HCl, 20 mM MgCl₂, 5 mM DTT, pH 6.8, and 20 units of T4 polynucleotide kinase. The reactions were incubated at 37 °C for 1 h and terminated by ethanol precipitation. For both 5' and 3' end product analyses the conditions for RNase T1 digestion of TAR were 0.03 units RNase T1, 35 mM NaCl, 3.5 M urea, 2 mM EDTA, 20 mM sodium citrate, pH 5.0, and approximately 200,000 cpm of ³²P end-labeled TAR in a total volume of 50 µL. After incubation at 37 °C for 20 min digestions were terminated by ethanol precipitation. All cleavage products were analyzed by electrophoresis on a 20%,

1:20 cross-linked, 7 M urea, 0.4 mm thick denaturing polyacrylamide gel at 45 V/cm for 15 h.

Quantitation by Storage Phosphor Technology Autoradiography. Following electrophoresis, gels were wrapped in plastic film and exposed to photostimulable phosphorimaging plates (Kodak Storage Phosphor Screen S0230) for 12-24 h in the dark at 22 °C. A Molecular Dynamics 400S PhosphorImager was used to obtain all data from the storage screens. Cleavage intensities for each nucleotide were obtained by performing volume integrations of each band using the ImageQuant v. 3.0 software running on a AST Premium 386/33 computer. Individual arrow heights were obtained by subtracting background cleavage from the total cleavage intensity.

References and Notes

1. (a) Bass, B. L.; Cech, T. R. *Nature* **1984**, *308*, 820. (b) Cech, T. R. *Science* **1987**, *236*, 1532. (c) Westheimer, F. H. *Science* **1987**, *235*, 1173.
2. Chin, J.; Banaszczyk, M.; Jubian, V.; Zou, X. *J. Am. Chem. Soc.* **1989**, *111*, 186.
3. (a) Blackburn, R.; Moore, S. *The Enzymes*, Academic Press, New York, 1982, vol. 15, chap. 12, pp. 317-433. (b) Richards, F. M. *The Enzymes*, Academic Press, New York, 1971, vol. 4, chap. 24, pp. 647-806. (c) Buzayan, J. M.; Gerlach, W. L.; Bruening, G. *Proc. Natl. Acad. Sci. U.S.A* **1986**, *83*, 8859. (d) Hutchins, C. J.; Rathjen, P. D.; Forster, A. C.; Symons, R. H. *Nucleic Acids Res.* **1986**, *14*, 3627. (e) Forster, A. C.; Symons, R. H. *Cell* **1987**, *49*, 211. (f) Buzayan, J. M.; Gerlach, W. L. *Nature* **1986**, *323*, 349. (g) Hampel, A.; Tritz, R. *Biochemistry* **1989**, *28*, 4929. (h) Sharmeen, L.; Kuo, M. Y. P.; Dinter-Gottlieb, G.; Taylor, J. *J. Virol.* **1988**, *62*, 2674. (i) Saville, B. J.; Collins, R. A. *Cell* **1990**, *61*, 685. (j) Long, D. M.; Uhlenbeck, O. C. *FASEB J.* **1993**, *7*, 25.
4. For recent reviews see, (a) Cullen, B. R. *Microbiological Reviews* **1992**, *56*, 375. (b) Frankel, A. D. *Curr. Opin. Genet. Dev.* **1992**, *2*, 293. (c) Cullen, B. R. *Cell* **1993**, *73*, 417. (d) Sheridan, P. L.; Sheline, C. T.; Milocco, L. H.; Jones, K. A. *Seminars in Virology* **1993**, *4*, 69.
5. For recent reviews, see (a) Weeks, K. M.; Crothers, D. M. *Perspectives in Drug Discovery and Design* **1993**, *1*, 225. (b) Gait, M. J.; Karn, J. *Trends Biochem. Sci.* **1993**, *18*, 255.
6. Weeks, K. M.; Ampe, C.; Schultz, S. C.; Steitz, T. A.; Crothers, D. M. *Science* **1990**, *249*, 128. (b) Weeks, K. M.; Crothers, D. M. *Cell* **1991**, *66*, 577. (c) Churcher, M. J.; Lamont, C.; Hamy, F.; Dingwall, C.; Green, S. M.; Lowe, A. D.; Butler, J. G.; Gait, M. J.; Karn, J. *J. Mol. Biol.* **1993**, *230*, 90.

7. Weeks, K. M.; Crothers, D. M. *Biochemistry* **1992**, *31*, 10281.
8. (a) Puglisi, J. D.; Tan, R.; Calnan, B. J.; Frankel, A. D.; Williamson, J. R. *Science* **1992**, *257*, 76. (b) Tao, J.; Frankel, A. D. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 2723. (c) R. Tan, R.; Frankel, A. D. *Biochemistry* **1992**, *31*, 10288. (d) Puglisi, J. D.; Chen, L.; Frankel, A. D.; Williamson, J. R. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 3680. (e) Wang, Z.; Rana, T. M. *J. Am. Chem. Soc.* **1995**, *117*, 5438.
9. Loret, E. P.; Georgel, P.; Johnson, J., W. C.; Ho, P. S. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 9734.
10. Mujeeb, A.; Bishop, K.; Peterlin, B. M.; Turk, C.; Parslow, T. G.; James, T. L. *Proc. Natl. Acad. Sci. U.S.A* **1994**, *91*, 8248.
11. (a) Eichhorn, G. L.; Butzow, J. J. *Biopolymers* **1965**, *3*, 79. (b) Rordorf, B. F.; Kearns, D. R. *Biopolymers* **1976**, *15*, 1491. (c) Breslow, R.; Huang, D.-L. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 4080. (d) Komiyama, M.; Matsumura, K.; Matsumoto, Y. *J. Chem. Soc., Chem. Commun.* **1992**, 640.
12. (a) Marciniak, T.; Ciesiolka, J.; Wrzesinski, J.; Krzyzosiak, W. J. *Febbs. Lett.* **1989**, *243*, 293. (b) Ciesiolka, J.; Marciniak, T.; Krzyzosiak, W. J. *Eur. J. Biochem.* **1989**, *182*, 445.
13. Morrow, J. R.; Buttrey, L. A.; Shelton, V. M.; Berback, K. A. *J. Am. Chem. Soc.* **1992**, *114*, 1903.
14. Kolasa, K. A.; Morrow, J. R.; Sharma, A. P. *Inorg. Chem.* **1993**, *32*, 3983.
15. Smith, P. H.; Brainard, J. R.; Morris, D. E.; Jarvinen, G. D.; Ryan, R. R. *J. Am. Chem. Soc.* **1989**, *111*, 7437.
16. The synthesis of dialdehyde **4** was developed by Professor E. M. Carreira, Carreira, E. M.; Dervan, P. B. Unpublished Observations.

17. (a) Backer-Dirks, J. D.; Gray, C. F.; Hart, A.; Hursthouse, M. B.; Schoop, B. C. *J. Chem. Soc. Chem. Comm.* **1979**, 774. (b) Abid, K. K.; Fenton, D. E. *Inorg. Chim. Acta* **1984**, 95, 119. (c) Radecka-Paryzek, W. *Inorg. Chim. Acta* **1985**, 109, L21. (d) De Cola, L.; Smailes, D. L.; Vallarino, L. M. *Inorg. Chem.* **1986**, 25, 1729. (e) Arif, A. M.; Backer-Dirks, J. D.; Gray, C. F.; Hart, A.; Hursthouse *J. Chem. Soc. Dalton. Trans.* **1987**, 1665. (f) Bombieri, G.; Benetello, F.; Polo, A.; De Cola, L.; Hawkins, W. T.; Vallarino, L. M. *Polyhedron* **1989**, 17, 2157. (g) Cabral, J. de O.; Carbal, M. F.; Cummins, W. J.; Drew, M. G. B.; Rodgers, A.; Nelson, S. M. *Inorg. Chim. Acta* **1978**, 30, L313. (h) Fenton, D. E.; Vigato, P. A. *Chem. Soc. Rev.* **1988**, 17, 69. (i) Smith, P. H.; Brainard, J. R.; Morris, D. E.; Jarvinen, G. D.; Ryan, R. R. *J. Am. Chem. Soc.* **1989**, 111, 7437.
18. Richardson, C. C. *Proc. Natl. Acad. Sci. USA* **1972**, 2, 815.
19. Cameron, V.; Uhlenbeck, O. C. *Biochemistry* **1977**, 16, 5120.
20. Weber, T. R. Ph.D. Thesis, University of Illinois at Urbana-Champaign, 1985.
21. Hydrolysis of the initial 2',3'-cyclic phosphate products by the La(L)³⁺ moiety would likely afford a mixture of 2'- and 3'-monophosphates.²² We note that the 3'-end-products may be a mixture of 2',3'-cyclic phosphates, 2'-monophosphates, and 3'-monophosphates.
22. Kuusela, S.; Lonnberg, H. *J. Phys. Org. Chem.* **1992**, 5, 803.
23. Oxidative scission of the RNA backbone typically produces 3'-phosphate or phosphoglycolate and 5'-phosphate or aldehyde cleavage products. For reviews, see (a) Stubbe, J.; Kozarich, J. W. *Chem. Rev.* **1987**, 87, 1107. (b) Dedon, P. C.; Goldberg, I. H. *Chem. Res. Toxicol.* **1992**, 5, 311.

24. For an example of sequence-specific oxidative cleavage of TAR by Cu(II)-phenanthroline covalently attached to the cysteine side chain of Tat(49-73, Cys⁷³), see Jayasena, J. D.; Johnston, B. H. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 3526.
25. An alternative mechanism is hydrolysis of the labile phosphodiester by La³⁺-bound hydroxide or a water molecule without involvement of the RNA 2'-hydroxyl groups. This mechanism is unlikely considering the resistance of DNA to cleavage by lanthanide ions.
26. Y. Matsumoto, Y.; Komiyama, M. *J. Chem. Soc. Chem. Commun.* **1990**, 1050.
27. Bashkin, J. K.; Frolova, E. I.; Sampath, U. *J. Am. Chem. Soc.* **1994**, *116*, 5981.
28. Matsumura, K.; Endo, M.; Komiyama, M. *J. Chem. Soc. Chem. Commun.* **1994**, 2019.
29. Magda, D.; Miller, R. A.; Sessler, J. L.; Iverson, B. L. *J. Am. Chem. Soc.* **1994**, *116*, 5981.
30. Hall, J.; Husken, D.; Pielec, U.; Moser, H. E.; Haner, R. *Chem. and Biol.* **1994**, *1*, 185.
31. (a) Carreira, E. M.; Dervan, P. B. Unpublished Observations. (b) Beal, P. A.; Dervan, P. B. Unpublished Observations. (c) Trauger, J. W.; Dervan, P. B. Unpublished Observations.
32. Still, W. C.; Kahn, M.; Mitra, A. *J. Org. Chem.* **1978**, *40*, 2923-2925.
33. (a) Kent, S. B. H. *Ann. Rev. Biochem.* **1988**, *57*, 957. (b) Kent, S. B. H.; Clark-Lewis, I. *Synthetic Peptides in Biology and Medicine*; Elsevier: Amsterdam, The Netherlands, 1985. (c) Clark-Lewis, I.; Aebersold, R.; Siltener, H.; Schrader, J.; Hood, L. E.; Kent, S. B. H. *Science* **1986**, *231*, 134. (d) Kent, S. B. H.; Parker, K. F.; Schiller, D. L.; Wood, D. D.-L.; Clark-Lewis, I.; Chait, B. T. In *Proceedings of*

the Tenth American Peptide Symposium; ESCOM: St. Louis, MO, 1987; pp 173-178.

34. (a) Sluka, J. P.; Horvath, S. J.; Bruist, M. F.; Simon, M. I.; Dervan, P. B. *Science* **1987**, 238, 1129. (b) Mack, D. P.; Iverson, B. L.; Dervan, P. B. *J. Am. Chem. Soc.* **1988**, 110, 7572. (c) Graham, K. S.; Dervan, P. B. *J. Biol. Chem.* **1990**, 265, 16534. (d) Mack, D. P.; Dervan, P. B. *J. Am. Chem. Soc.* **1990**, 112, 4604. (e) Mack, D. P.; Sluka, J. P.; Shin, J. A.; Griffin, J. H.; Simon, M. I.; Dervan, P. B. *Biochemistry* **1990**, 29, 6561. (f) Oakley, M. G.; Dervan, P. B. *Science* **1990**, 248, 847. (g) Sluka, J. P.; Horvath, S. J.; Glasgow, A. C.; Simon, M. I.; Dervan, P. B. *Biochemistry* **1990**, 29, 6551. (h) Shin, J. A.; Ebright, R. H.; Dervan, P. B. *Nucleic Acids Res.* **1991**, 19, 5233. (i) Mack, D. P.; Dervan, P. B. *Biochemistry* **1992**, 31, 9399.
35. Milligan, J. F.; Uhlenbeck, O. C. *Methods Enzymol.* **1989**, 180, 51.
36. (a) Bruce, A. G.; Uhlenbeck, O. C. *Nucleic Acids Res.* **1978**, 5, 3665. (b) Romaniuk, P. J.; Uhlenbeck, O. C. *Methods Enzymol.* **1983**, 100, 52.
37. Celander, D. W., Cech, T. R. *Biochemistry* **1990**, 29, 1355.

Addendum to Chapter 4

Part 1

Analysis of the Specificity of Lanthanide-Catalyzed Hydrolysis of a Designed, Folded RNA Structure

Experimental Design. In the course of examining the specificity of TAR cleavage by $\text{La}(\text{NO}_3)_3$ and $\text{La}(\text{L})^{3+}$, RNA **16** was prepared by chemical synthesis and analyzed in a similar fashion (Figure 4.17). RNA **16** is a derivative of TAR in which the double-helical stem below the trinucleotide pyrimidine bulge has been deleted and a single-stranded region has been inserted at the 5' end. The purpose of these experiments was to determine the specificity of lanthanides for cleaving different types of structural elements, i.e. duplex, bulge, loop, single-stranded, in a stable, folded RNA structure. RNA **16** was 5' ^{32}P end-

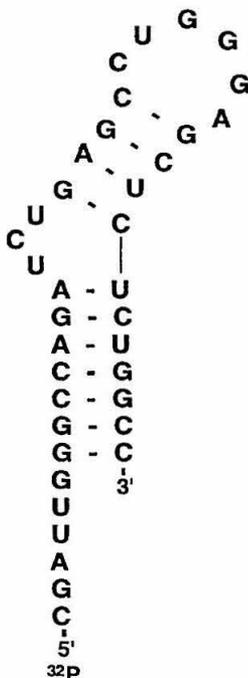


Figure 4.17. Predicted secondary structure of RNA **16**.

labeled and treated with 100 μM $\text{La}(\text{NO}_3)_3$ for 30 min (70 mM NaCl, 10 mM Tris•HCl, pH 7.2, 37 °C) and the cleavage products were analyzed by 20% denaturing gel electrophoresis. The single-stranded regions in RNA **16** are all cleaved with greater efficiency than the double helical stem regions (Figure 4.18). Within the single-stranded regions, the unpaired nucleotides at the 5' end and the nucleotides in the pyrimidine bulge appear to be more reactive than the six nucleotides in the loop. These results underscore the importance of targeting single-stranded regions in an RNA molecule with artificial catalysts for sequence-specific hydrolysis.

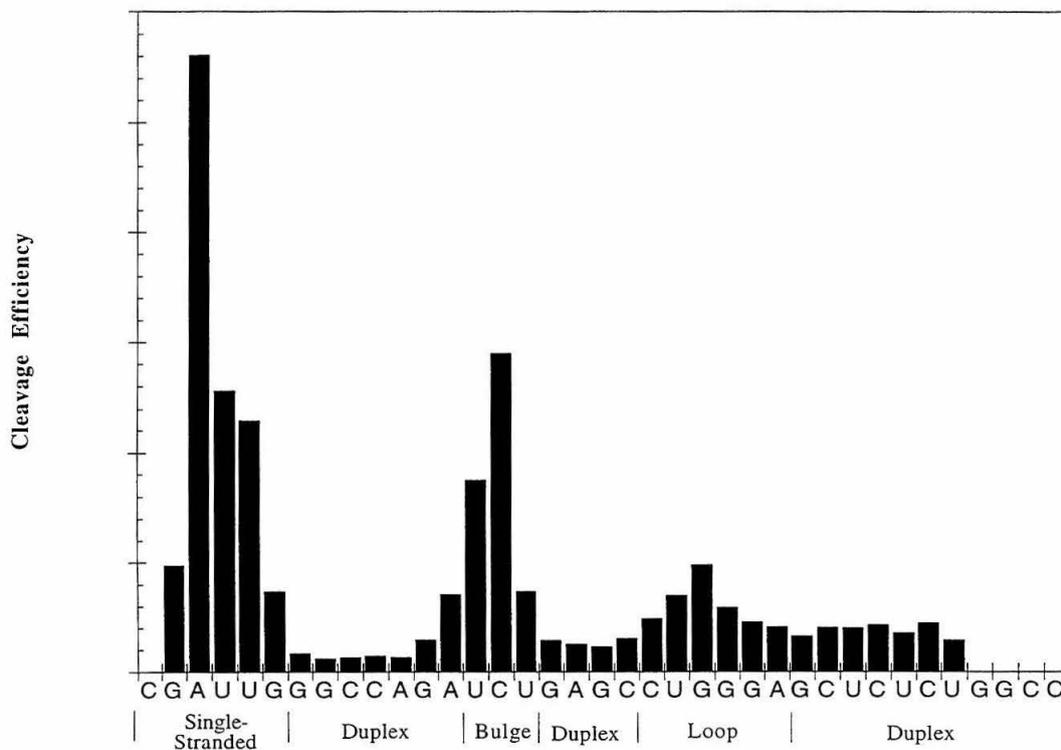


Figure 4.18. Cleavage of RNA **16** by 100 μM $\text{La}(\text{NO}_3)_3$ after 30 min (pH 7.2, 37 °C). Bar heights are proportional to the extent of cleavage as measured by storage phosphor autoradiography. No data was obtained for the 5'-GGCC-3' sequence at the 3' end of the sequence as the bands corresponding to cleavage at these positions were located too close to intact RNA **16** in the autoradiogram for accurate quantitation.

Addendum to Chapter 4

Part 2

Attempted Transesterification of Single-Stranded DNA by Nucleophile-Appended Lanthanide Complexes

Experimental Design. Nucleophile-appended La(L)^{3+} complexes **17-22** were prepared and tested for their ability to hydrolyze single-stranded DNA (Figure 4.19). The hydroxyl groups were designed to mimic the 2'-hydroxyl group of RNA. Binding of the complex to a DNA phosphodiester was designed to activate it for attack by the appended nucleophile, effecting strand cleavage by transesterification. A 20mer oligodeoxynucleotide ^{32}P labeled at the 5' end was treated with complexes **17-22** (50 mM NaCl, 10 mM Tris•HCl, pH 7-9, 12-48 h, 37-55 °C) and analyzed for cleavage by 20% denaturing gel electrophoresis. No cleavage above background levels was observed with any of the complexes and this project was terminated.

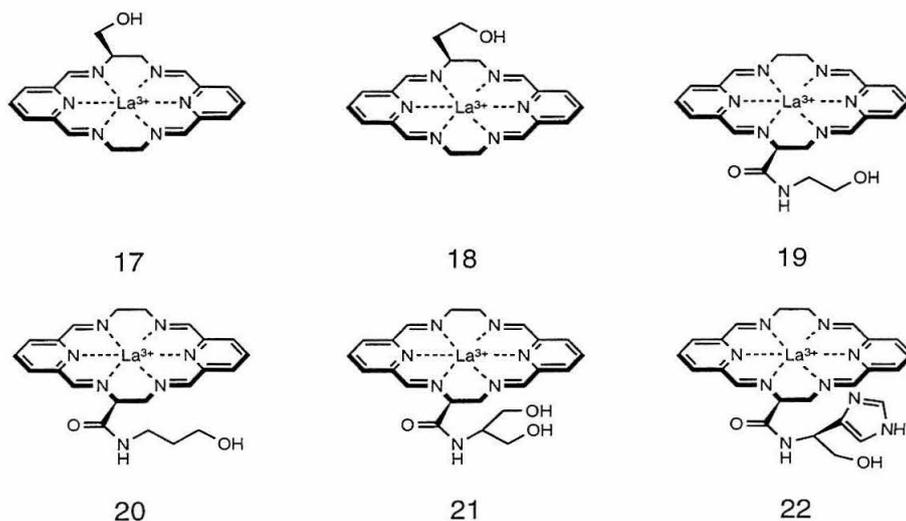


Figure 4.19. Structure of nucleophile-appended lanthanide complexes designed for DNA transesterification.