# Design of Cyclic Polyamides for Sequence-Specific Recognition of the Minor Groove of DNA.

Thesis By

Junhyeong Cho

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

California Institute of Technology
Pasadena, California

1996

(Submitted May 9, 1996)

*To Sohye*

# Acknowledgements

I would like to thank my advisor, Professor Peter B. Dervan, for his thoughtful guidance, support, and encouragement during the course of my graduate studies. I would also like to thank my committee members, Professors John H. Richards, Barbara Imperiali, and Erick Carreira, for their guidance and support.

I am grateful to the members of the Dervan Group, past and present, for supporting this work with discussion, suggestions, and friendships. Special thanks go to Dr. Milan Mrksich, Michelle Parks, Eldon Baird, Dr. Hogyu Han, Dr. Bryan Takasaki, and Scott Carter.

Finally, I would like to thank my family for all the help and encouragement. My special gratitude goes to my wife Sohye. She has been my partner throughout my stay at Caltech. Her love, support, and sacrifices made the completion of my graduate work possible. I dedicate this work to her.

# Abstract

Small molecules that specifically bind with high affinity to any designated DNA sequence in the human genome would be useful tools in molecular biology and potentially in human medicine. Simple rules have been developed to rationally alter the sequence specificity of minor groove binding polyamides containing *N*-methylimidazole (Im) and *N*-methylpyrrole (Py) amino acids. Crescent-shaped polyamides bind as antiparallel dimers with each polyamide making specific contacts with each strand on the floor of the minor groove.

In Chapter 2, the design of and synthesis of the cyclic polyamide *cyclo*-(Im-Py-Py-γ-Py-Py-Py-γ) is described. The cyclic polyamide binds the 5'-TGTTA-3' site with 40-fold higher affinity than the hairpin polyamide Im-Py-Py-γ-Py-Py-Py-Dp. These results demonstrate that wholly designed synthetic cyclic polyamides with a molecular weight of 950 can bind designed five base pair sequences at subnanomolar concentration.

In Chapter 3, in order to develop polyamide ligands to differentiate between A,T-rich sequences, a cyclic polyamide has been synthesized in which two Py-Py-Py units are circled through two γ-aminobutyric acids (γ). DNase I footprinting titration experiments reveal that *cyclo*-(Py-Py-Py-γ-Py-Py-Py-γ) binds the 5'-TATAT-3' sequence with 20-fold higher affinity than distamycin.

In Chapter 4, experiments are described in which EDTA•Fe(II) is attached at unique sites of HIV-1 TAR RNA by the chemical synthesis method and autocleavage of TAR-EDTA•Fe(II) is performed. Preliminary results showed that the autocleavage pattern of TAR RNA is changed in the presence of arginine. However, a series of control experiments indicate that

the change in the autocleavage pattern is not caused by an RNA conformational change driven by the binding of arginine.

# Table of Contents

# Figures and Tables

## Chapter 1

Page

### Figures

## Chapter 2

### Figures

**Tables**



## Chapter 3

**Figures**



**Tables**



## Chapter 4

**Figures**

# Chapter 1

# Introduction

Deoxyribonucleic acid (DNA) plays an important role in life by encoding all of the proteins and RNA necessary for cellular function. The regulation of cellular processes including gene replication, transcription, and cell division is accomplished by the specific recognition of genomic DNA by proteins. An understanding of these sequence specific recognition events is essential for the design of DNA binding molecules. The design of new sequence-specific DNA-binding agents that possess anticancer activities is a challenging goal for chemists.

**DNA structure.**

Double-helical DNA consists of two antiparallel polydeoxyribonucleotide strands that wind around each other[1,2]. The two strands are held together by Watson-Crick hydrogen bonds between the bases on opposite strands : A is paired with T by two hydrogen bonds and G is paired with C by three hydrogen bonds. Vertical $\pi$-$\pi$ stacking interactions between adjacent base pairs on the same strand contribute significantly to the overall stability of the double helix. In this structure the deoxyribose phosphodiester backbones are placed away from the helical axis, creating two distinct grooves along the cylindrical morphology of the double helix (Figure 1.1).

Double-helical DNA usually adopts the right-handed B-form under physiological conditions. This type of DNA structure is featured by 10.5 base pairs per turn of helix with a 3.4 Å distance between adjacent stacked bases[1,2]. The helix axis is roughly perpendicular to the plane of the base pair. The

major groove of the B-DNA is shallow and wide (11.7 Å), where as the minor groove is deep and narrow (5.7 Å). Local conformational properties such as helical twist, propeller twist, base-pair slide and roll, and groove width of duplex DNA vary with the sequence.

The edges of the base pairs are located at the floor of the grooves. The major and minor grooves display different functionality depending on the identity of the base pair (Figure 1.2). In the major groove, four base pairs can be completely distinguished by the distinct array of hydrogen bond donor and acceptor groups. In contrast, the hydrogen bonding contacts in the minor groove are not as varied. For instance, the two hydrogen bond acceptors of an A•T base pair are the adenine N3 and thymine O2 atoms. The G•C base pair has not only the corresponding hydrogen bond acceptors but also an exocyclic N2 amino group as a hydrogen bond donor. A•T and T•A base pairs can not be distinguished in the minor groove since two hydrogen bond acceptors are quite symmetrically located in the groove.

## DNA Binding Molecules

A variety of molecules have been known to bind to DNA in a sequence specific manner. On the basis of the molecular structure, these molecules can be grouped into three classes : DNA binding proteins, triple helixes, and small drug molecules.

**Protein-DNA complexes** Recently, a great deal of high-resolution structural data has led to our improved understanding of protein-DNA complexes[3]. Several different structural motifs for DNA binding proteins have been identified which include the zinc finger, leucine zipper, homeodomain, and helix-turn-helix motifs[4]. The sequence specificity results

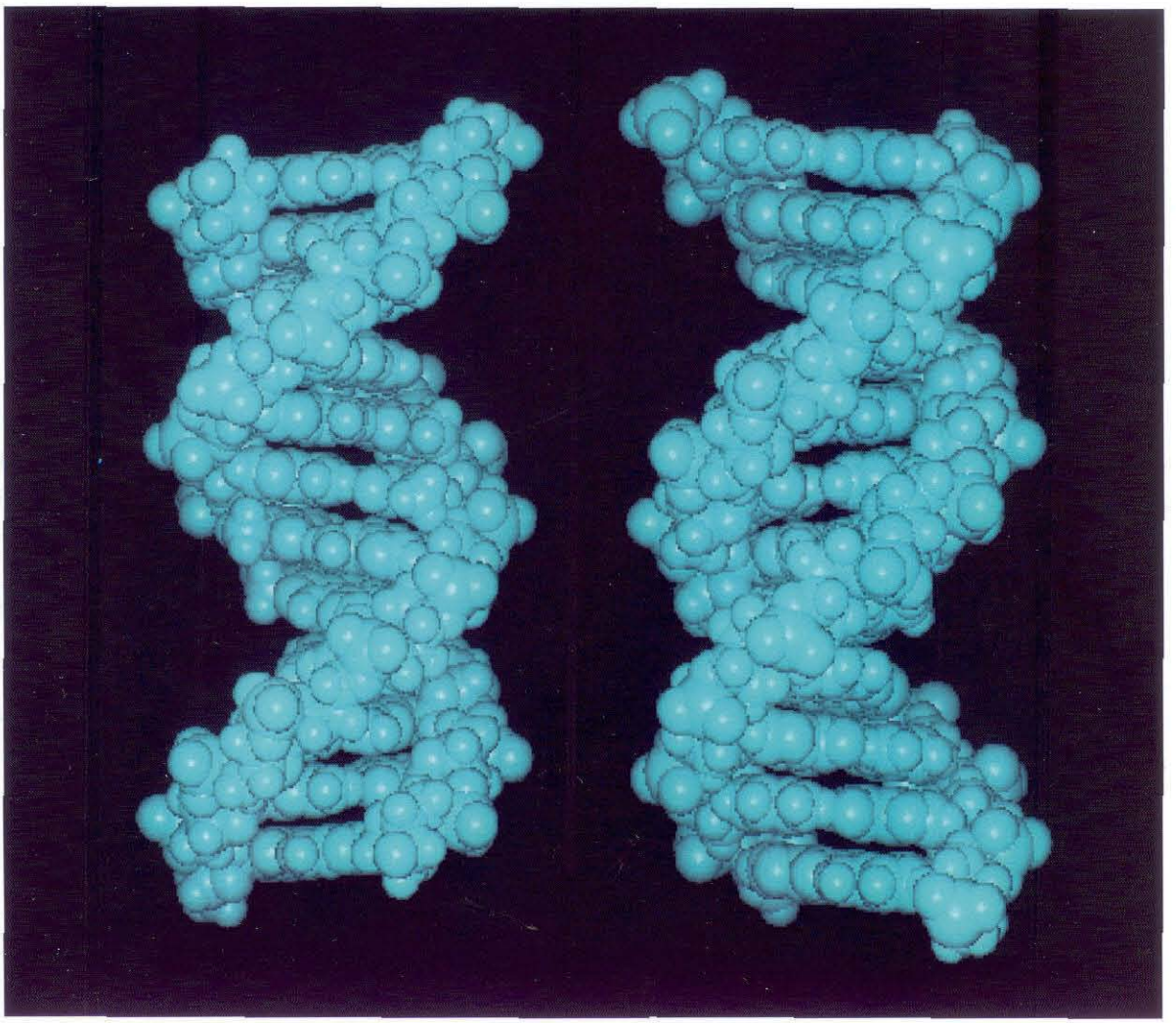


**Figure 1.1.** Structure of B-form double-helical DNA. The wide and shallow major groove and the narrow and deep minor groove are clearly distinguishable.

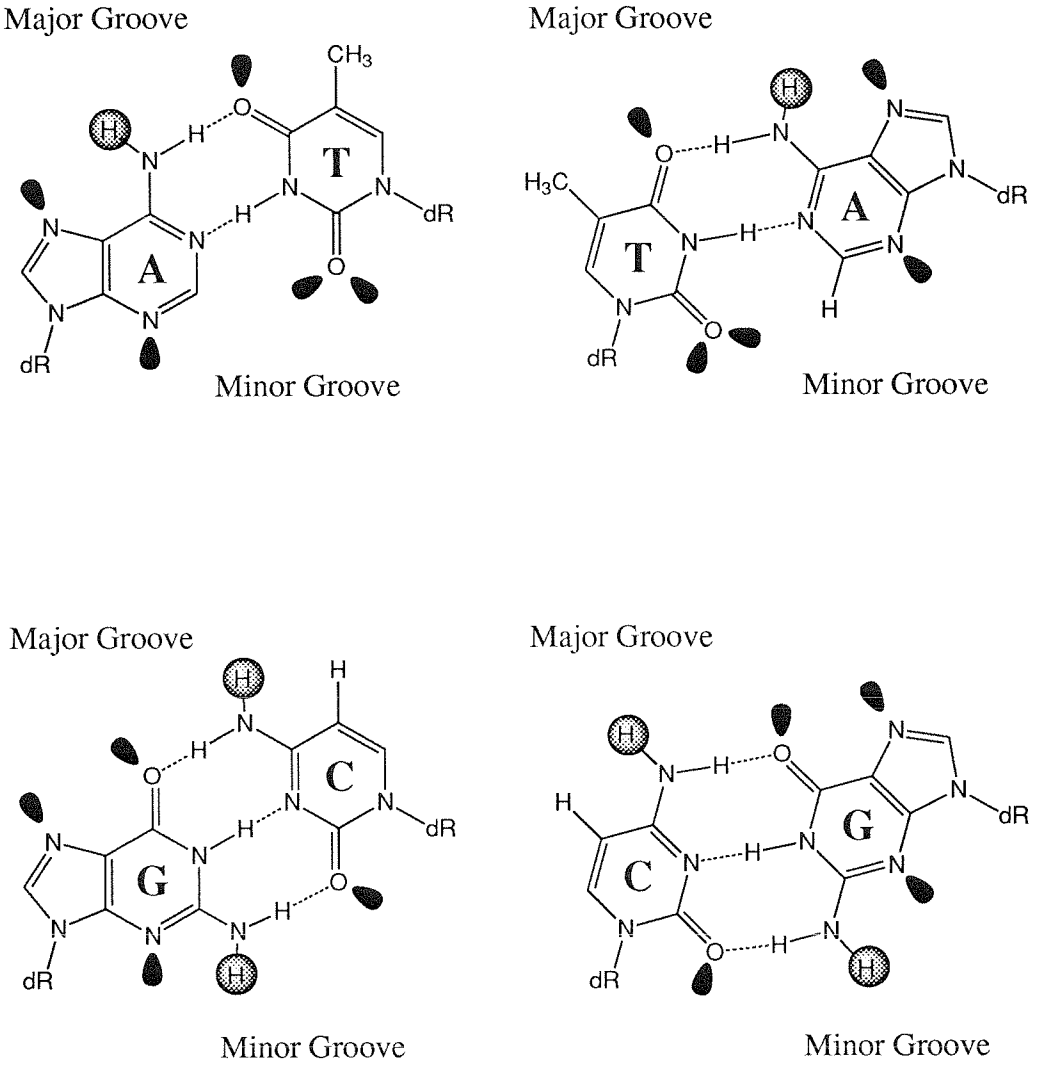



**Figure 1.2.** Structure of the four natural Watson-Crick base pairs showing the functional groups displayed in the major and minor grooves. Lone pairs that function as potential hydrogen bond acceptors are represented as black lobes and potential hydrogen bond donors are indicated as gray spheres.

from a combination of specific hydrogen bonds to the DNA base in the major, or both major and minor groove, as well as from the overall shape complementarity between the proteins and the DNA grooves. Because there is no simple general recognition code that explains the binding specificity, it is very difficult to design proteins for recognition of designated DNA sequences.

**Oligonucleotide-directed triple helix complexes** Short pyrimidine oligonucleotides can bind sequence-specifically in the major groove of double helical DNA[5]. Specificity of the triple-helical complexes relies on Hoogsteen hydrogen bonds (T•AT and $C^+$•GC) in the major groove between a pyrimidine base on the third strand and a purine base of the Watson-Crick base pairs. In this complex, the third strands is parallel to the purine strand of the duplex. Beal and Dervan reported a second structural motif for triple helix in which purine-rich oligonucleotides bind in the major groove of DNA, antiparallel to the purine strand of the duplex[6]. The specific interaction is mediated by the formation of G•GC, A•AT, and T•AT base triples. The stability of triple helical complexes are dependent on the length[5,7], sequence composition[8], temperature[9], as well as solution conditions, including pH[10] and cation concentrations[11].

**Small molecule-DNA complexes** A variety of small molecules ( MW < 1500 Daltons) are known to bind specifically DNA (Figure 1.3)[12]. Some of these molecules have proven very useful as antitumor agents because they form tight complexes with particular DNA sites and then block transcription. Two primary modes of interaction, intercalation and minor-groove interactions, have been observed in the DNA complexes of such small molecules. A number of polycyclic aromatic cations bind to DNA via insertion of the aromatic ring into the vertical stack of the base pairs.

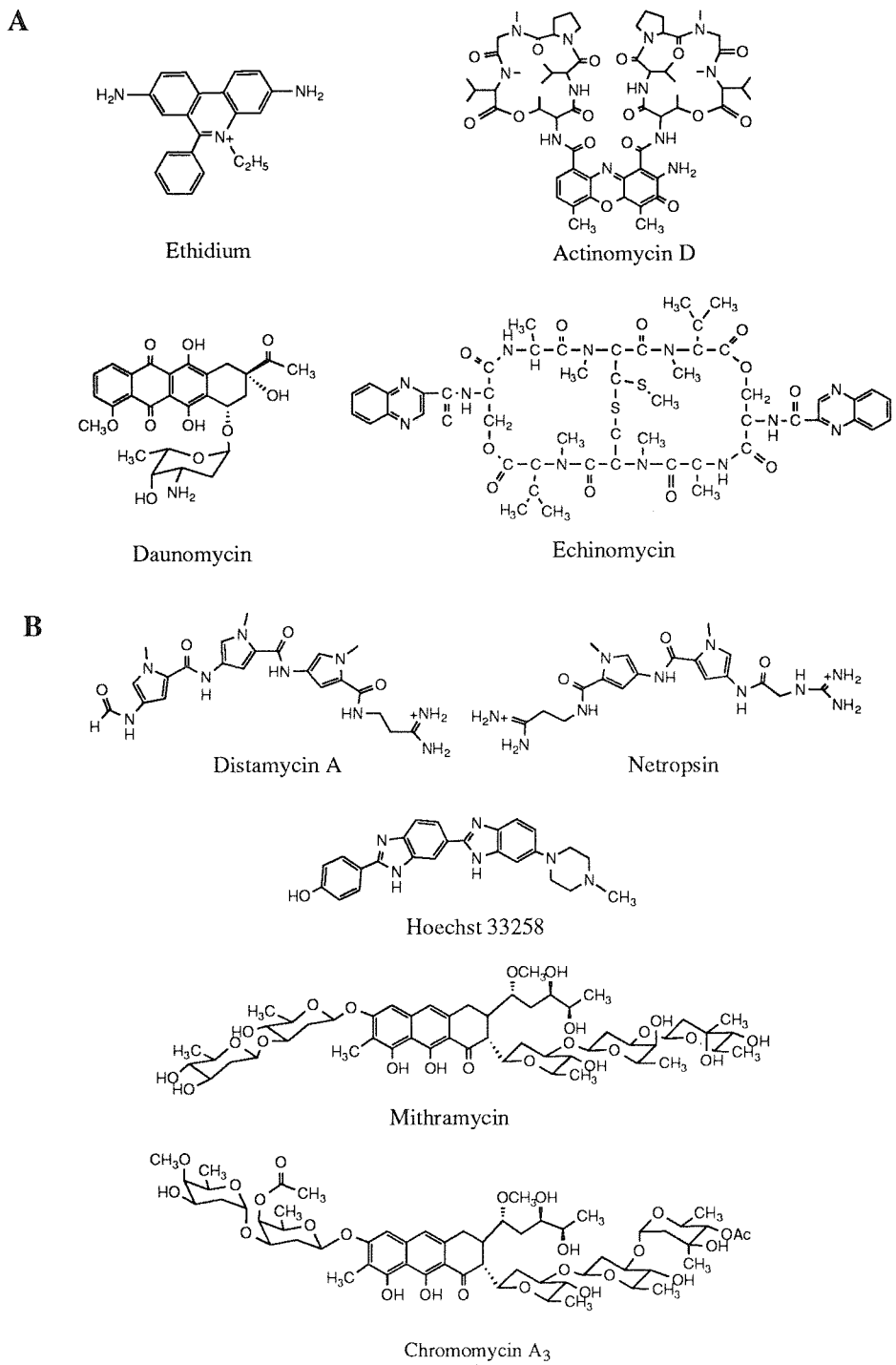



**Figure 1.3.** Structure of small-molecule ligands that form complexes with the minor groove.

Intercalators such as ethidium[13], daunomycin[14], echinomycin[15], and actinomycin[16] contain an extended, flat, often positively charged aromatic chromophore. In general, these molecules show a relatively low specificity because the major interaction is the relatively nonspecific π-π stacking. Intercalators usually display a slight preference for a G•C base pair. This preference results from the slightly greater intrinsic dipole moment of G•C base pairs, which is better able to induce polarization in the aromatic ring of the intercalating molecule. The binding of intercalators not only causes the DNA helix to extend, but also induces an unwinding of the double helix.

The other class comprises minor groove binders such as netropsin[17], distamycin[18], Hoechst 33258[19], chromomycin[20], and mithramycin[21]. Netropsin and distamycin bind preferentially to A,T-rich sequence without major distortion of the DNA structure. These molecules recognize DNA through the formation of specific hydrogen bonds with the functional groups at floor of the minor groove. In contrast, chromomycin and mithramycin bind to G,C-rich sites of at least three base pairs as a magnesium-coordinated dimer. Binding of the drug dimer results in a major distortion and drastic widening of the minor groove[20,21]. The diversity of small molecule ligands capable of specifically recognizing DNA enriches the repertoire of models for the design of new DNA-binding molecule.

**Sequence-Specific DNA Binding Assays**

Affinity cleaving Affinity cleaving experiments involve the covalent attachment of a nonspecific DNA cleaving agent, commonly EDTA•Fe(II), to a DNA binding ligand to generate a sequence-specific DNA cleaving molecule[22]. The resulting EDTA•Fe(II)-equipped molecule cleaves DNA by

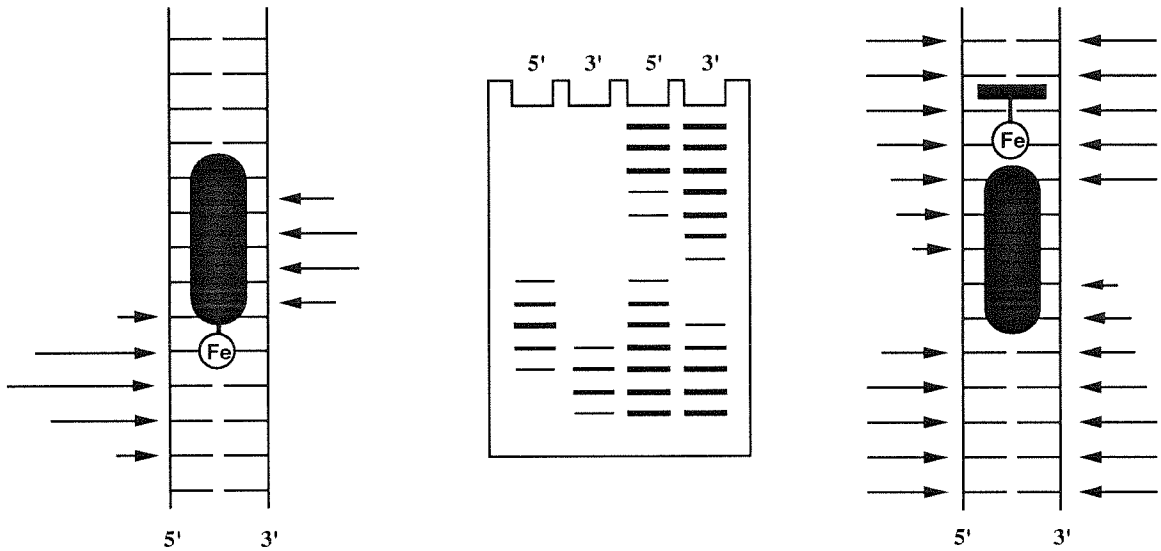



**Figure 1.4.** Affinity cleaving (left) and footprinting (right) techniques for the analysis of sequence-specific binding to double helical DNA.

oxidation of the deoxyribose backbone via a diffusible hydroxyl radical. The cleavage reaction can be initiated by the addition of a reducing agent such as dithiothreitol or sodium ascorbate. Analysis of the cleavage data provides information not only about the preferred binding sites of the DNA-binding molecule but also about its orientation at these sites (Figure 1.4). In addition, analysis of the asymmetric cleavage of the two DNA strands reveals whether the cleaving ligand is bound in the major or minor groove (Figure 1.5). Due to the right-handed nature of the B-DNA helix, the major cleavage sites are asymmetrically shifted in the 5'-direction when EDTA•Fe(II) is located in the major groove. In the case when EDTA•Fe(II) lies in the minor groove, the major cleavage sites are shifted in the 3'-direction. This technique has been used in the studies of the sequence-specific DNA recognition by proteins[23], oligonucleotides[7,24], and small molecules[25].

Footprinting In a footprinting experiment, a radiolabeled DNA fragment is allowed to react with a non-specific DNA cleaving agent to

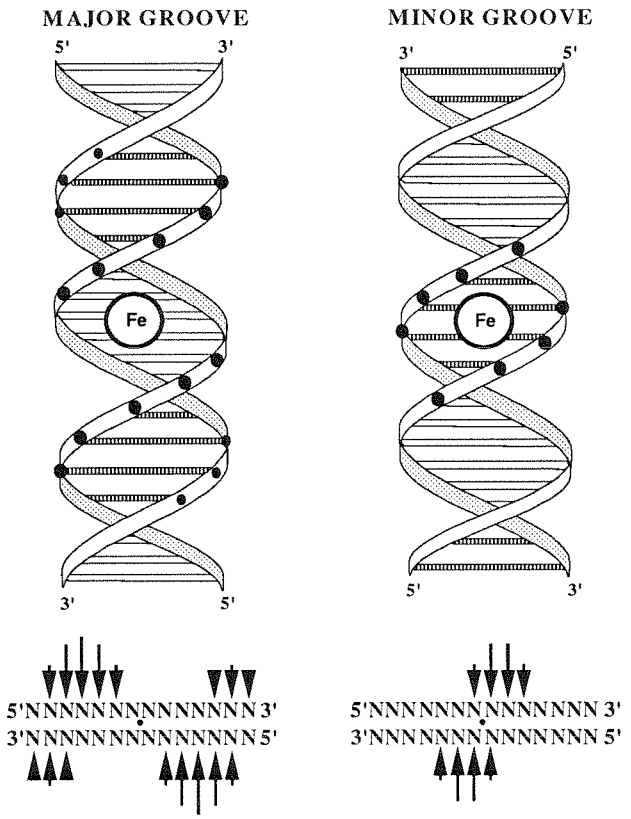



**Figure 1.5.** Models for the asymmetric DNA cleavage pattern generated by EDTA•Fe(II) positioned in the major and minor grooves of right-handed double helical DNA

generate random cleavage on DNA. The DNA cleaving agent can be either DNase I[26] or the synthetic reagent methidiumpropyl-EDTA•Fe(II) (MPE•Fe(II))[27]. When the reaction is carried out in the presence of a specific DNA binding molecule, cleavage is blocked at the positions to which the ligand binds, resulting in a "footprint" in the DNA cleavage ladder (Figure 1.4). This technique provides information about the location and size of the DNA binding sites for a given DNA-binding molecule.

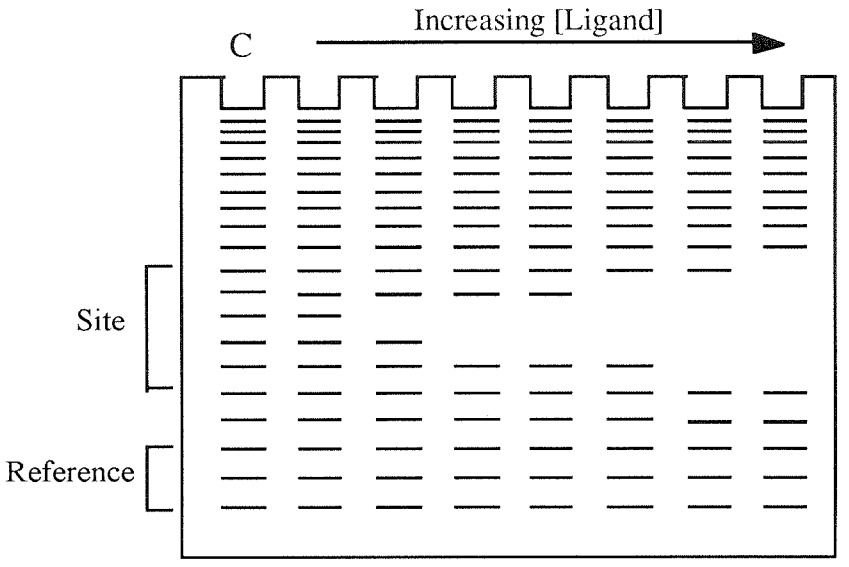



**Figure 1.6.** Illustrative gel from a quantitative DNase I footprint titration experiment.

<u>Quantitative DNase I footprint titration assays</u> Quantitative DNase I footprint titration experiment can be used to estimate the binding affinity of ligands for specific sites on DNA. This technique has been well developed by Ackers group for the measurement of binding affinity of proteins and small molecules[28,29]. In some cases, this experiment can provide information on the stoichiometry of ligand-DNA interaction. In the quantitative footprint titration experiment, footprinting reactions are carried out at many different concentrations of the ligand (Figure 1.6). After the digested products are separated on a polyacrylamide sequencing gel, the gel is imaged using phosphor storage technology. The data is analyzed by performing volume integrations of the ligand binding site and a reference site which is not bound by the ligand[30]. The fractional occupation of the site ($\theta_{app}$) is calculated by comparing the protection at the binding site to that of the reference site and normalizing to the control lane using the following equation:

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

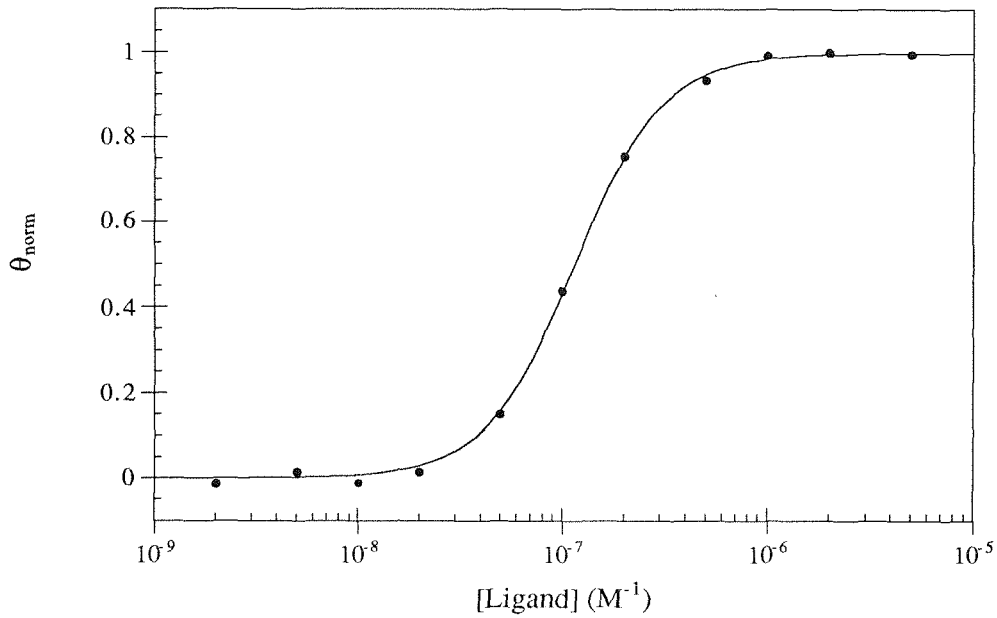



**Figure 1.7.** Example of a binding isotherm generated from a quantitative DNase I footprint titration experiment where the fractional occupancy, $\theta_{norm}$, and log[L] are reported on the vertical and horizontal axes, respectively. The curve through the data points is the best-fit Langmuir binding isotherm. The ligand concentration at half-occupancy is the first order dissociation constant.

where $I_{site}$ and $I_{ref}$ are the integrated volumes of the target and reference sites, respectively, and $I_{site}^{\circ}$ and $I_{ref}^{\circ}$ correspond to those values for a footprint control lane to which no ligand has been added. The ([L], $\theta_{app}$) data points are then fit to a Langmuir binding isotherm (eq 2) by minimizing the difference between $\theta_{app}$ and $\theta_{fit}$ using the modified Hill equation:

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

where $[L]_{tot}$ corresponds to the total ligand concentration, $K_a$ corresponds to the apparent monomeric association constant, n is the Hill coefficient, and

$\theta_{min}$ and $\theta_{max}$ represent the experimentally determined site saturation values when the binding site is unoccupied or saturated, respectively (Figure 1.7).

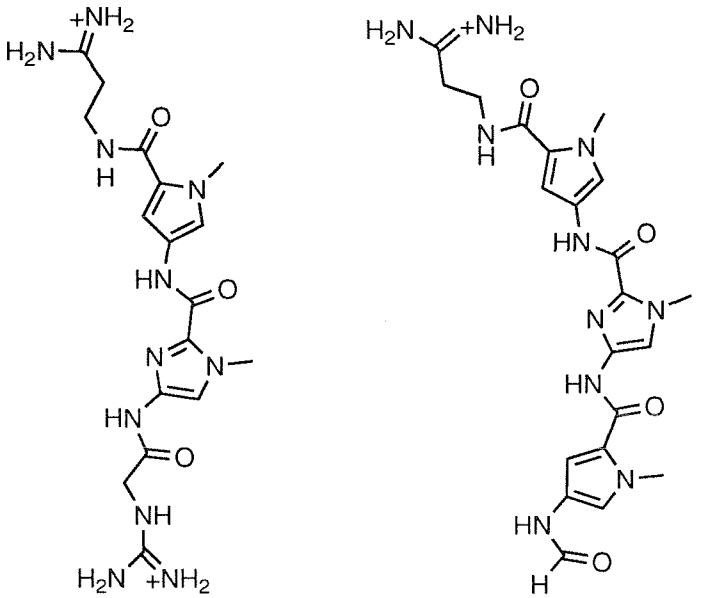


**Figure 1.8.** Structure of netropsin and distamycin A.

**Netropsin and Distamycin**

Netropsin and distamycin A are naturally occurring antibiotics with crescent shaped peptide skeleton (Figure 1.8)[31]. Netropsin is a di-(N-methylpyrrolecarboxamide) with an N-terminal guanidinium group and a C-terminal amidinium group whereas distamycin is a tri-(N-methylpyrrolecarboxamide) with N-terminal formamide group and a C-terminal amidinium group. These compounds are known to exert antifugal, antimitotic, and antivial activity *in vivo* by binding in the minor groove of A,T-rich DNA sequences[32,33]. These polyamides have received extensive study as the paradigms of the minor groove binding drug molecule.

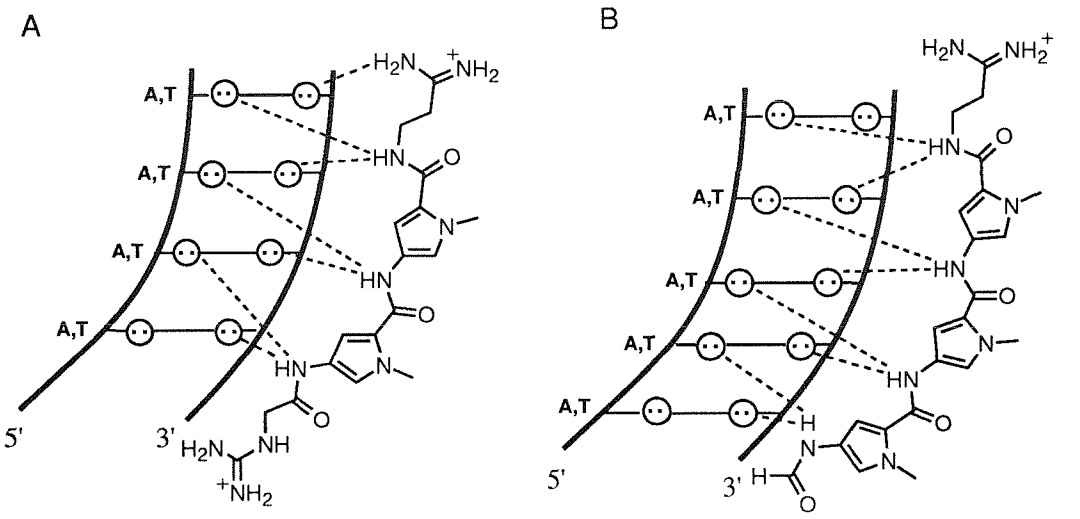



**Figure 1.9.** 1:1 Models for netropsin (A) and distamycin (B) binding to A,T-rich DNA

Structure of 1:1 polyamide-DNA complexes Extensive crystallographic[34] and NMR[35] studies of netropsin-and distamycin-DNA complexes provide insight into how sequence-specific recognition is accomplished. The crescent-shaped polyamides completely occupy the narrow minor groove at the central region of an A,T-rich sequence. The drug molecules are not planar and are slightly twisted in a screw sense, which is complementary to the curvature of the minor groove of B-DNA. The amide NH groups form bifurcated hydrogen bonds with adenine N3 and thymine O2 atoms at the floor of the minor groove (Figure 1.9). In addition to hydrogen-bonding interactions, van der Waals interactions between the aromatic pyrrole rings and the walls of the minor groove contribute significantly to the binding energy (Figure 1.10). The positively charged amidinium and guanidinium groups of the drugs lie along the floor of the minor groove forming favorable electrostatic interactions with the DNA[36,37].

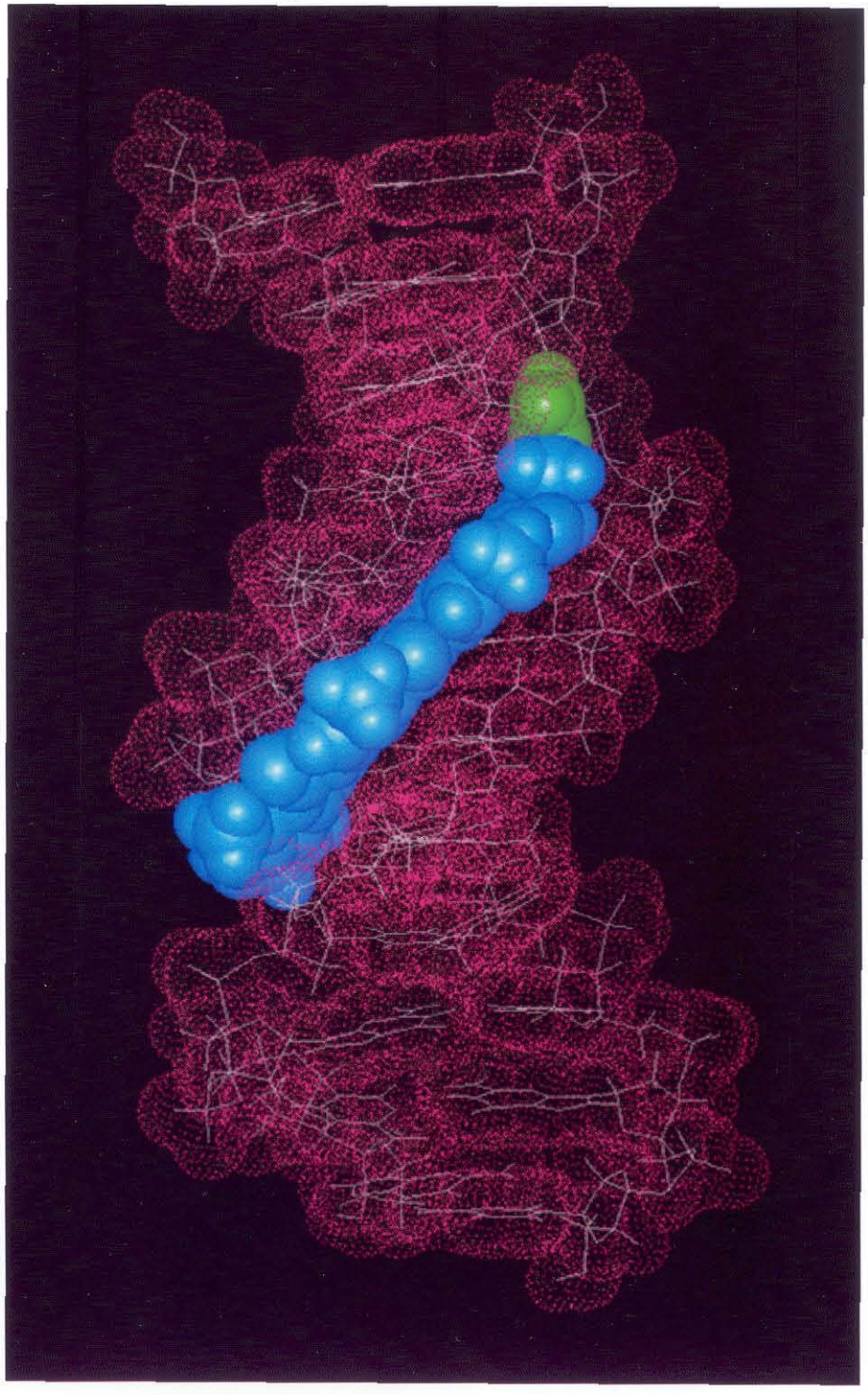


**Figure 1.10.** High resolution x-ray crystal structure of the 1:1 complex between distamycin and a 5′-AAATTT-3′ site.[34d] The N-methylpyrrole carboxamides and the positively charged amidinium group are colored cyan and green, respectively.

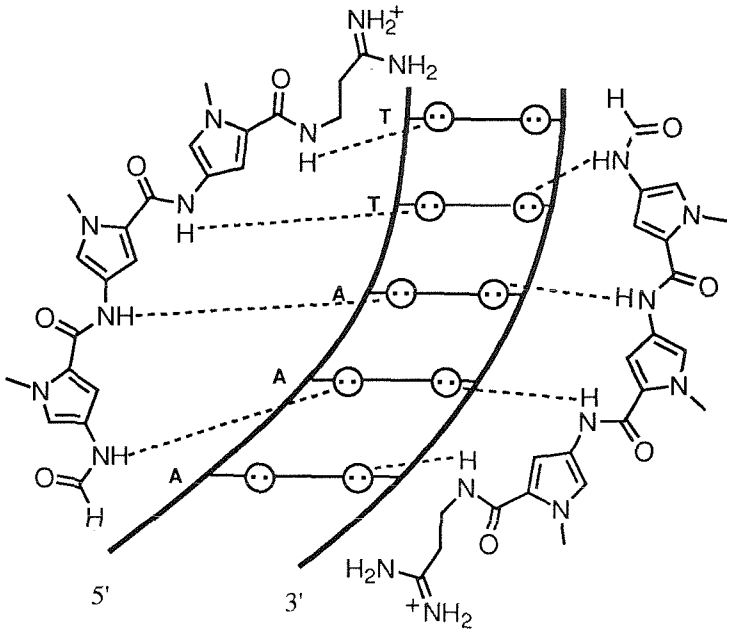



**Figure 1.11.** Binding model of the 2:1 complex between distamycin and a 5′-AAATT-3′ site derived from NMR data.

The binding specificity of netropsin and distamycin for A,T-rich over G,C-rich sequences can be explained from structural data. The aromatic hydrogens of the pyrrole rings are in close contact with the adenine C2 atoms in the minor groove. *If A•T base pairs are replaced by G•C base-pairs, there* are severe clashes between the N2 amino group of guanine and the pyrrole rings.

<u>2 : 1 Distamycin-DNA complexes</u> NMR studies by Pelton and Wemmer have revealed that distamycin at 2-4 mM concentrations binds in the minor groove of the 5'-AAATT-3' sequence as a side-by-side dimer (Figures 1.11, 1.12)[38,39]. Two distamycins are bound in the antiparallel orientation in order to avoid charge-charge repulsion between the positively charged propylamidinium groups. Because the minor groove width of A,T-

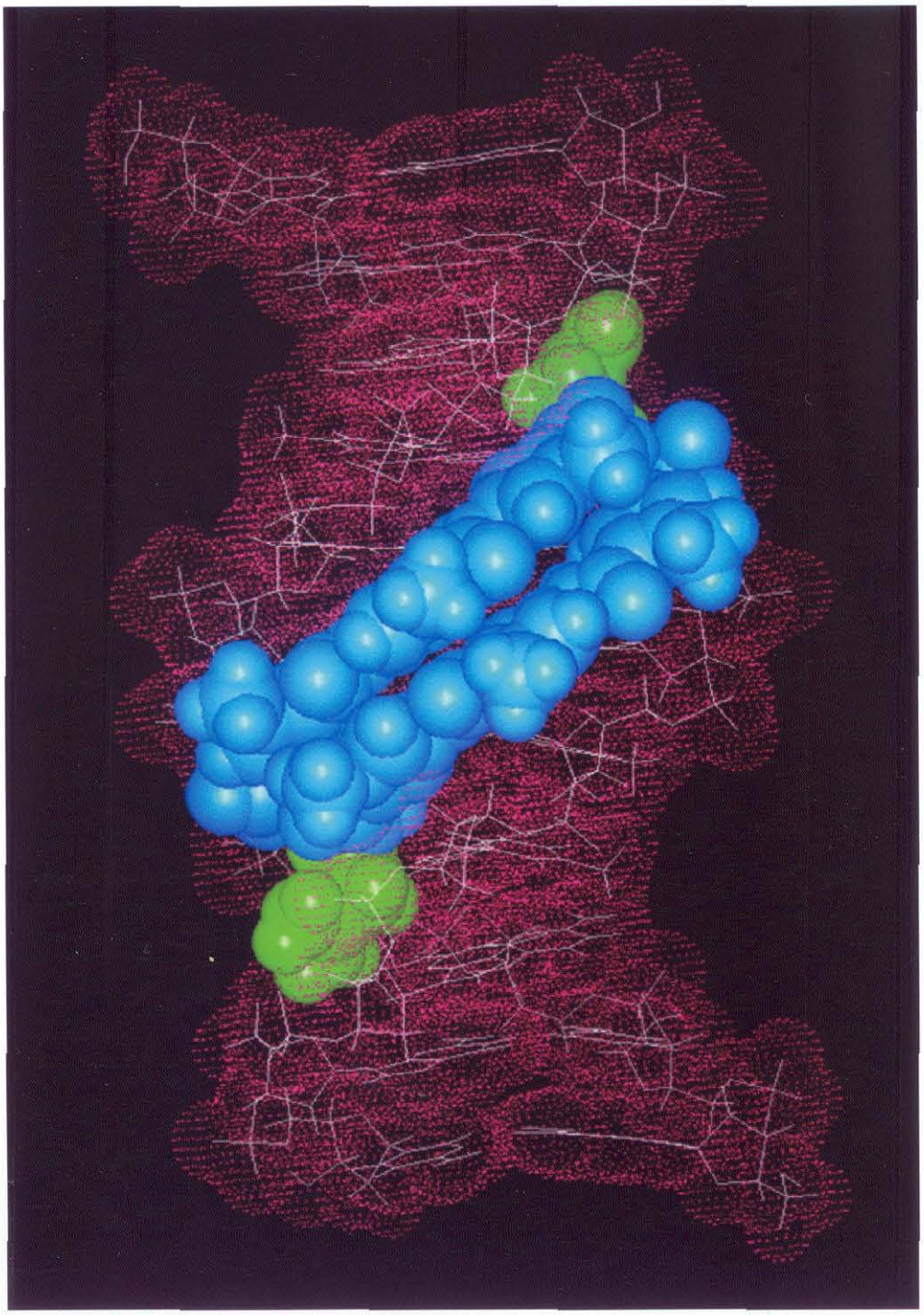


**Figure 1.12.** High resolution structure of the 2:1 complex between distamycin and a 5′-AAATT-3′ site derived from NMR data.[38] The N-methylpyrrolecarboxamides and the positively charged amidinium group are colored cyan and green, respectively.

rich sequences range from 3.2 to 4.0 Å, groove widening to approximately 7Å is required to accommodate both ligands. Molecular modeling suggests that the two ligands completely fill the minor groove and maintain the favorable van der Waals contacts with the floor and walls of the minor groove. The two ligands are staggered with respect to one another such that the pyrrole rings of one ligand stack with the amide linkages of the other ligand. The hydrogen bonds in 2:1 complexes are significantly different from in 1:1 complex. While distamycin forms bifurcated hydrogen bonds to bases on both strands of the DNA in 1:1 complex, each ligand in 2:1 complex participates in hydrogen bonding with bases of only a single strand. 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.[40]

**Polyamides for Binding Mixed Sequences**

There have been several attempts to develop polyamides capable of recognizing G•C base pairs . The major difference between A•T and G•C base pairs in the minor groove is the presence of the guanine 2-amino group protruding from the floor. Model studies indicate that replacement of the pyrrole CH by a hydrogen bond acceptor atom results in a rational alteration of base recognition from A•T to G•C. New polyamide molecules in which the pyrrole rings are replaced with imidazole and pyridine rings are expected to retain specificity for the mixed sequences[41,42]. Footprinting and affinity cleavage experiments reveal that 2-ImN (Im-Py-Py-Dp) and 2-PyN (Pyr-Py-Py-Dp) specifically bind to the 5'-TGTCA-3' sequence with unbiased orientation preferences[43,44]. However, these results are inconsistent with a 1:1 polyamide-DNA model

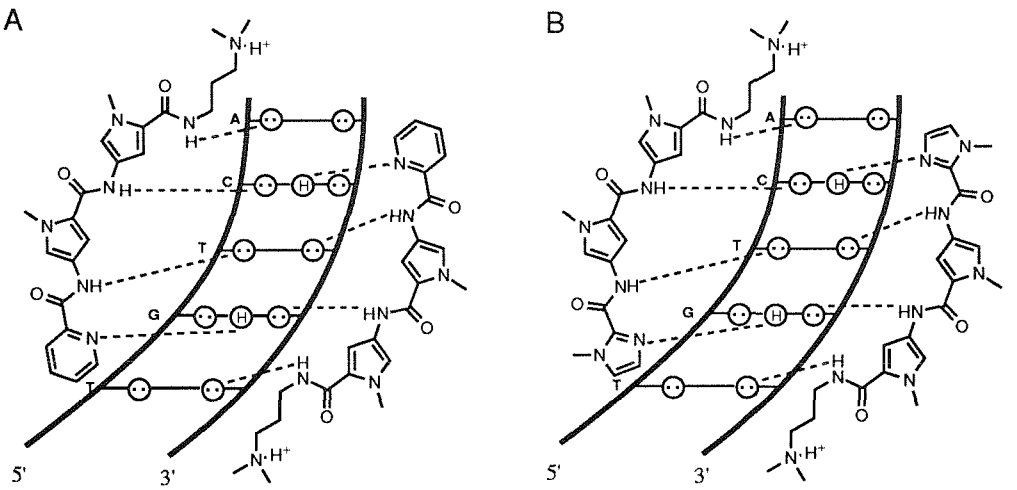



**Figure 1.13.** 2:1 models for (A) Pyr-Py-Py-Dp and (B) Im-Py-Py-Dp bound at a 5′-TGTCA-3′ site.

**Recognition of 5'-(A,T)G(A,T)C(A,T)-3' by 2:1 side-by-side homodimers.**

The finding that distamycin at high concentrations is capable of binding in the minor groove of a 5'-AAATT-3' sequence as an antiparallel side-by-side dimer suggests that the 2:1 model may serve as a general model for the design of polyamide analogs for specific recognition of other sequences[38,39]. This 2:1 model could explain the recognition of the 5'-TGTCA-3' sequence by 2-ImN and 2-PyN[43,44]. Each pyridine or imidazole nitrogen atom of side-by-side dimer forms a specific hydrogen bond with the 2-amino group of each guanine on the adjacent DNA strand (Figure 1.13). The amide NH protons participate in hydrogen bonds with N3 of adenine and O2 of thymine and cytosine. In addition, van der Waals interactions between the ligands and the walls of the minor groove play an important role in the stability of the 2:1 complex. The stacking interaction between conjugated rings of the ligands also contributes favorably to the binding energy. NMR experiments reveal that 2-ImN and 2-PyN bind to the 5'-TGACT-3' site as antiparallel side-by-side

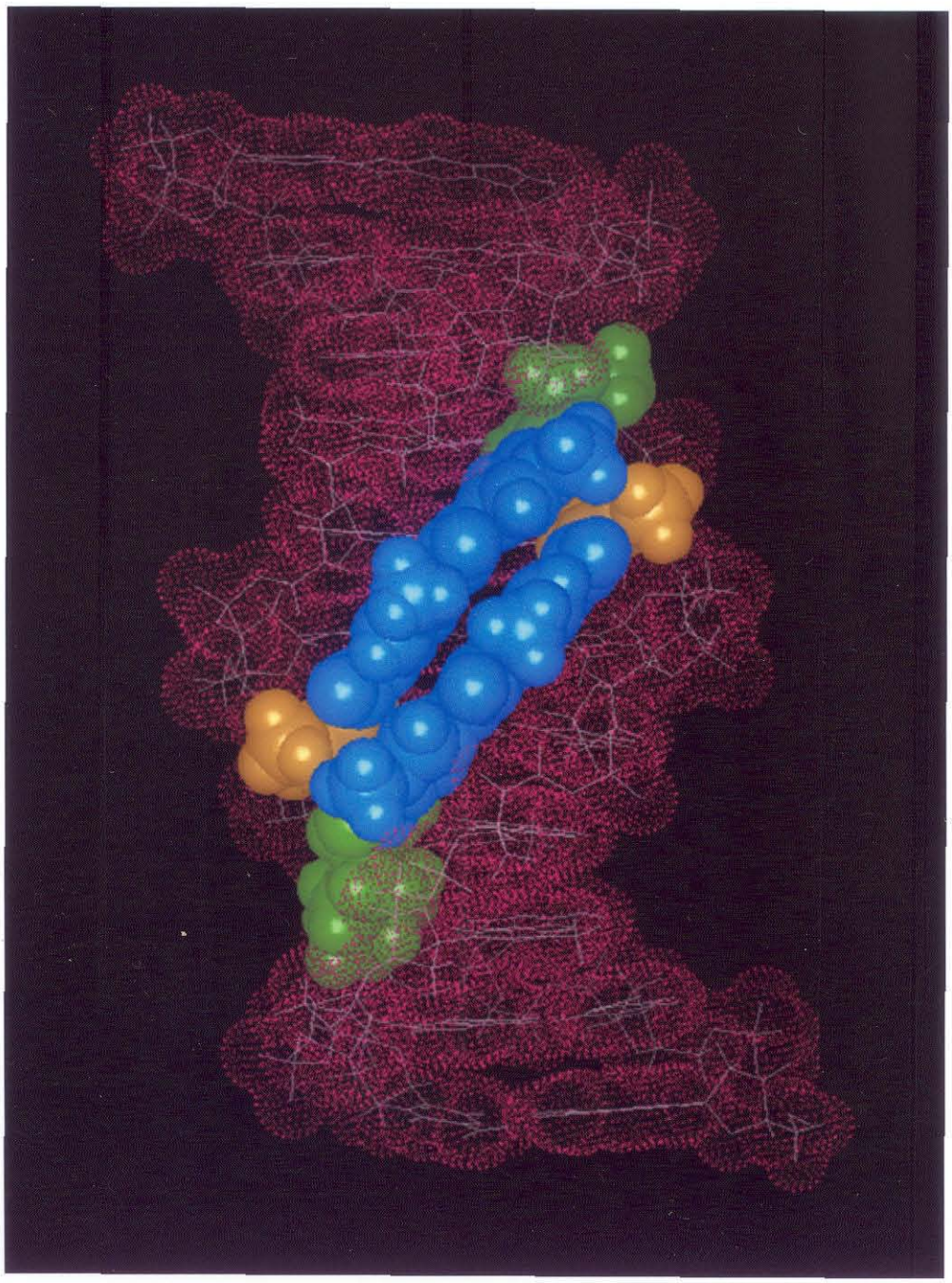


**Figure 1.14.** High resolution structure of the (Im-Py-Py-Dp)$_2$•5′-TGTCA-3′ complex derived from NMR data.[34d] Colors represent the following amino acids: cyan, Py; yellow, Im; green, Dp.

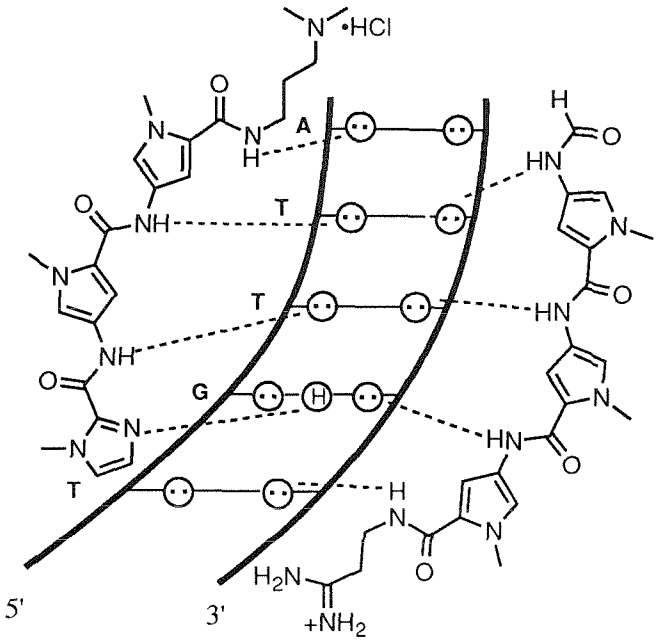



**Figure 1.15.** Model of the (Im-Py-Py-Dp/distamycin)•5′-TGTTA-3′ complex.

dimers (Figure 1.14)[45,46].

2-ImN and 2-PyN display an apparent preference for cooperative 2:1 binding to the 5'-TGTCA-3' sequence over 1:1 binding. The minor groove of mixed sequences is significantly wider than those of A,T-rich sequences. For these sequences with a wide minor groove, it appears to be energetically favorable to accommodate two ligands rather than to reduce the groove width for favorable contacts with a single ligand[45,47]. It is likely that certain sequences with inherently wide minor grooves favor dimeric binding.

<u>Recognition of 5'-TGTTA-3' by 2 : 1 side-by-side heterodimers.</u> In a 2:1 polyamide-DNA complex, each polyamide makes specific hydrogen bonds to the bases of a single strand in the minor groove. Based on this model, a heterodimer consisting of 2-ImN and distamycin was designed specifically to bind to 5'-(A,T)G(A,T)$_3$-3' sequences (Figure 1.15)[48]. The side-by-side

combination of an imidazole ring on one ligand and a pyrrole ring on the other ligand recognizes a G•C base pair, while a pyrrole-imidazole pair targets a C•G base pair. A pyrrole-pyrrole pair is partially degenerate and binds to either A•T or T•A base pairs. Footprinting and affinity cleavage experiments show that 2-ImN and distamycin simultaneously bind to a 5'-TGTTA-3' sequence with an antiparallel orientation, which is consistent with a 2:1 polyamide-DNA model[48]. Two-dimensional NMR study reveals that the distamycin ligand lies along the 5'-AACA-3' strand of the 5'-TGTTA-3' site with the charged tail group pointing to the 3'-end and the 2-ImN ligand spans the 5'-GTTA-3' strand in the opposite direction relative to distamycin[49]. In a related study, Lown and coworkers reported that a pyrrole-imidazole-pyrrole polyamide (2-ImD) can be combined with distamycin to target a 5'-AAGTT-3' site as an antiparallel heterodimer[50]. These examples suggest that the 2:1 model can serve as a general paradigm in the design of polyamides for specific recognition of many other mixed sequences.

**Recognition of 5'-GCGC-3'-sequences** The generality of the 2:1 polyamide-DNA model has been demonstrated by targeting A,T-rich sequences containing one or two G•C base pairs. In order to create a new polyamide capable of binding to G,C-rich sequence, the alternating four ring polyamide (Im-Py-Im-Py-Dp) was designed to bind in the minor groove of 5'-(A,T)GCGC(A,T)-3' sequences as an antiparallel homodimer. Footprinting and affinity cleavage experiments reveal that Im-Py-Im-Py-Dp specifically binds to the six base pair sites 5'-TGCGCA-3' and 5'-AGCGCT-3' in two orientations, which is consistent with the 2:1 binding model[51]. These results suggest that Im-Py-Im-Py-Dp does not discriminate between A•T and T•A base pairs in the first or sixth positions of the binding site. NMR studies of Im- Py-Im-Py-Dp•5'-TGCGCA-3' show that the two ligands bind in the minor

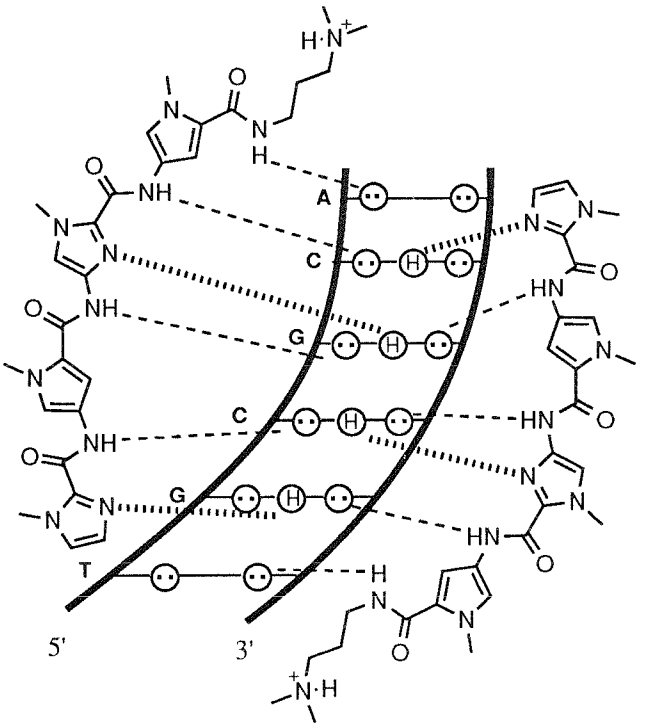



**Figure 1.16.** Model of the (Im-Py-Im-Py-Dp)$_2$•5′-TGCGCA-3′ complex derived from NMR data.

groove as an antiparallel side-by-side dimer with positive cooperativity (Figure 1.16)[52]. The ligands span the 5'-GCGCA-3' site, and the positively charged tail groups are oriented toward the 3'-end of the adjacent DNA strand. Molecular modeling indicates that each imidazole nitrogen atoms forms a hydrogen bond with the unique guanine amino group on the adjacent DNA strand (Figure 1.17). These hydrogen bonds play an important role in determining the sequence specificity. Binding of this polyamide to a pure four base pairs G•C core sequence represents complete reversal of the natural binding specificity of the natural products distamycin and netropsin for A,T-rich sequences.

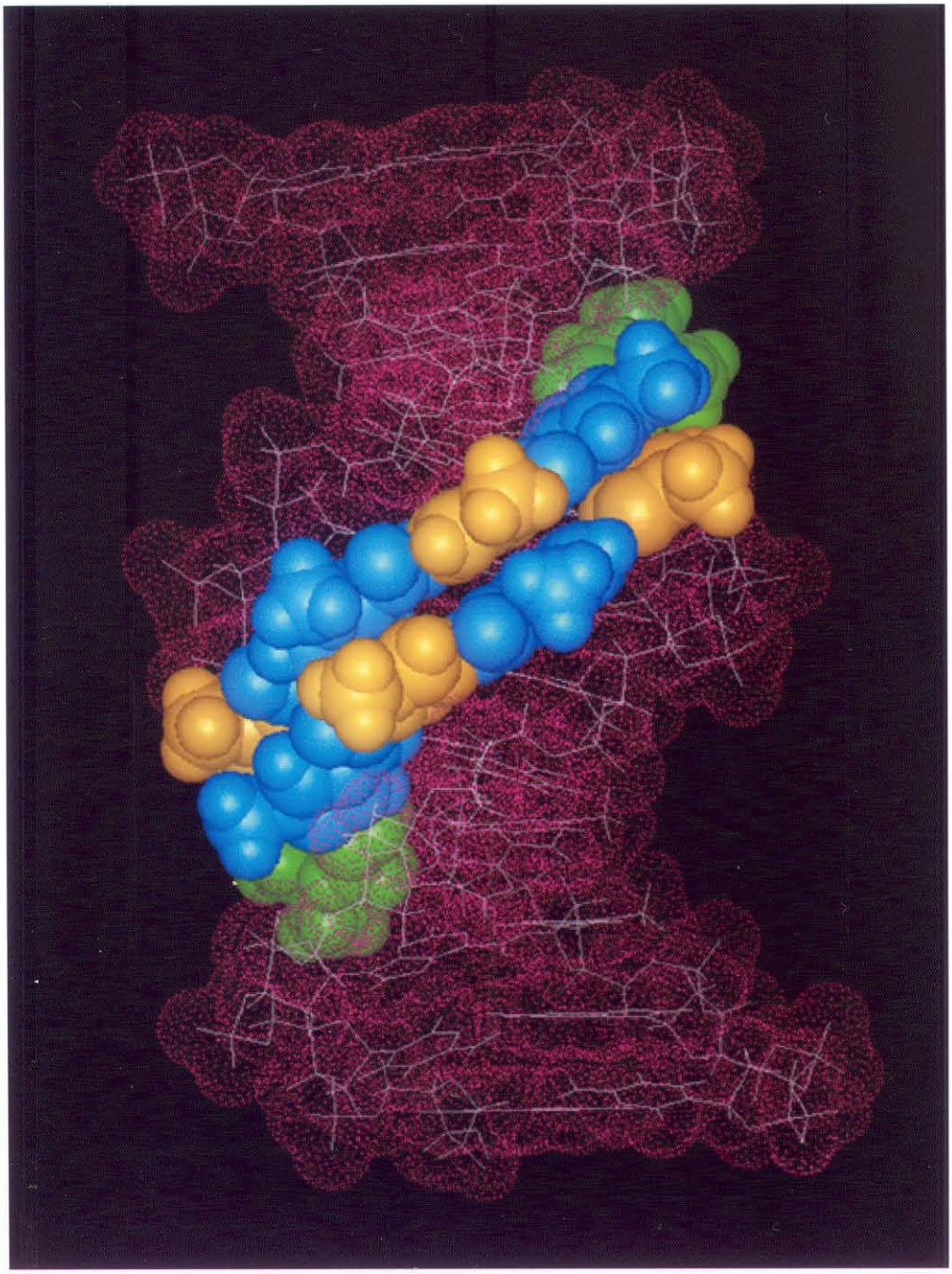


**Figure 1.17.** High resolution structure of the (Im-Py-Im-Py-Dp)$_2$•5′-TGCGCA-3′ complex derived from NMR data.[52] Colors represent the following amino acids : cyan, Py; yellow, Im; green, Dp.

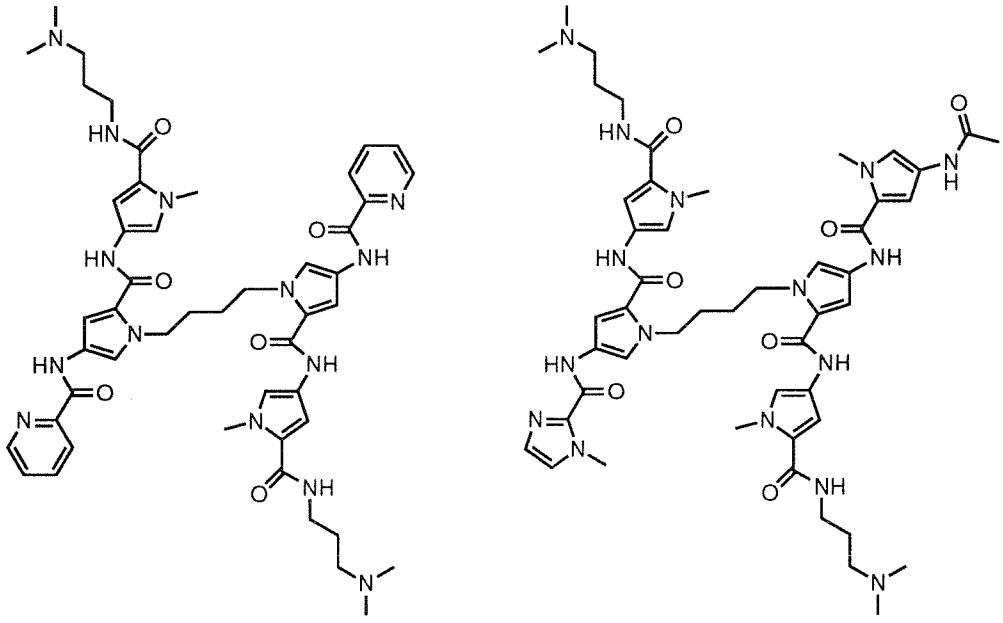
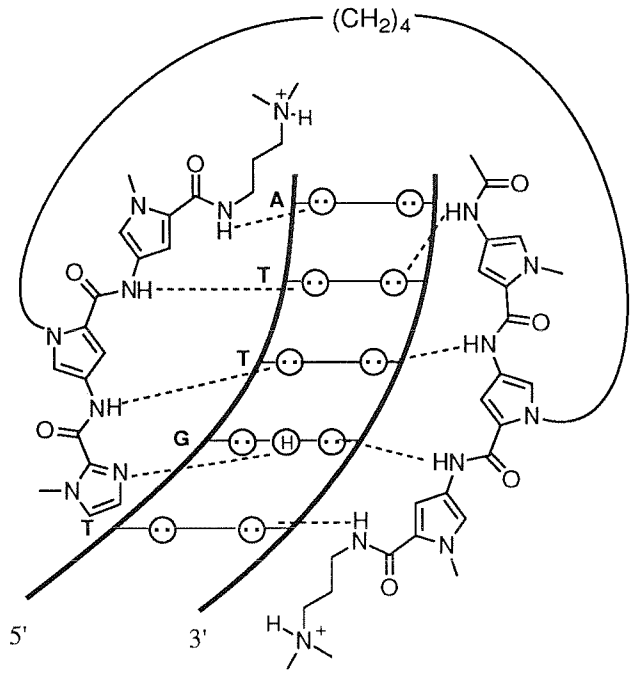


**Figure 1.18.** Structure of the covalently linked polyamides (Pyr-Py-Py-Dp)$_2$-C$_4$ and Im-Py-Py-Dp-C$_4$-Py-Py-Py-Dp.

## Covalently Linked Polyamides

Two-dimensional NMR study of an (2-ImN)$_2$•5'-TGACT-3' complex reveals that the two ligands bind in the minor groove as an antiparallel side-by-side dimer. Examination of the 2 : 1 model suggests that two ligands can be covalently linked without disrupting any interactions with DNA. The binding affinity of covalent dimers is expected to be increased as compared to the non-linked monomers because the free energy of binding for these linked dimers is entropically favorable. In addition, linking the two polyamide results in improved sequence specificity. Such a effect can be most evident in heterodimeric complexes, since two different monomers can formally bind to three different binding sites : two parent sites and the heterodimer site.




Im-Py-Py-Dp-$C_4$-Py-Py-Py-Dp • 5'-TGTTA-3'

**Figure 1.19.** Model of Im-Py-Py-Dp-$C_4$-Py-Py-Py-Dp•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.

**Centrally linked polyamides** Connecting the nitrogen atoms of the central pyrrole rings in the 2:1 2-PyN with a butyl linker affords a new class of polyamide (2-PyN)$_2$-$C_4$ (Figure 1.18). Footprinting results show that the covalent dimer (2-PyN)$_2$-$C_4$ binds the designated 5'-TGACA-3' sequence with ten-fold higher binding affinity than does the monomer 2-PyN[53]. Also, this covalent dimer displays increased sequence specificity relative to 2-PyN. Two-dimensional NMR studies reveal that the covalently linked 2-PyN ligands bind as intramolecular dimers with nearly identical geometry as in the non-linked monomer complexes[54].

In order to further explore the generality of the covalently linked dimer motif, the covalent heterodimer Im-Py-Py-$C_4$-Py-Py-Py-Dp in which the 2-ImN (Im-Py-Py) and distamycin are connected with a butyl linker was designed to specifically bind the 5'-TGTTA-3' sequence (Figure 1.18, 1.19). Footprinting experiments show that Im-Py-Py-$C_4$-Py-Py-Py-Dp binds the designated site 5'-TGTTA-3' with higher affinity than does distamycin or 2-ImN[55]. Sequence specificity of the linked heterodimer is also improved as evidenced by the lower affinity for the 5'-TGTCA-3' (2-ImN) and 5'-TTTTT-3' (distamycin) sites, respectively. Although these examples demonstrate the utility of the centrally linked side-by-side dimeric motif in the design of new polyamide molecules with improved binding affinity and sequence specificity, the lengthy synthesis of this class molecules may limit the extension of this dimeric motif to other polyamide analogs.

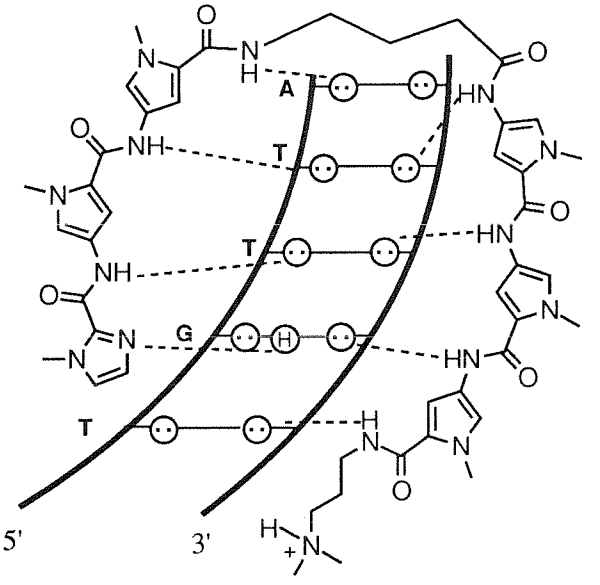



**Figure 1.20.** Model of the hairpin polyamide Im-Py-Py-γ-Py-Py-Py-Dp binding to a 5′-TGTTA-3′ site. Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines and circles containing a H represent the 2-amino group of guanine. Putative hydrogen bonds are indicated by dashed lines.

**Hairpin polyamides** Recently, two polyamide ligands have been linked in a head to tail fashion with simple amino acid linkers without perturbing the geometry of the polyamide-DNA complex. The linker could adopt a turn conformation such that the linear polyamide dimer turns and folds back on itself in the minor groove. The hairpin polyamide 2-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 designated 5'-TGTTA-3' site with high affinity ($Ka=8x10^{7}M^{-1}$) and binds a single base pair mismatch 5'-TGACA-3' site with 24-fold lower affinity ($Ka=3x10^{6}\ M^{-1}$)[56] (Figure 1.20, 1.21). This hairpin polyamide results in an increase in binding affinity of approximately 2 orders of magnitude relative to the unlinked 2:1 complex. This new class of hairpin polyamides has now been extended to the design of polyamides for recognition of many other sequences.

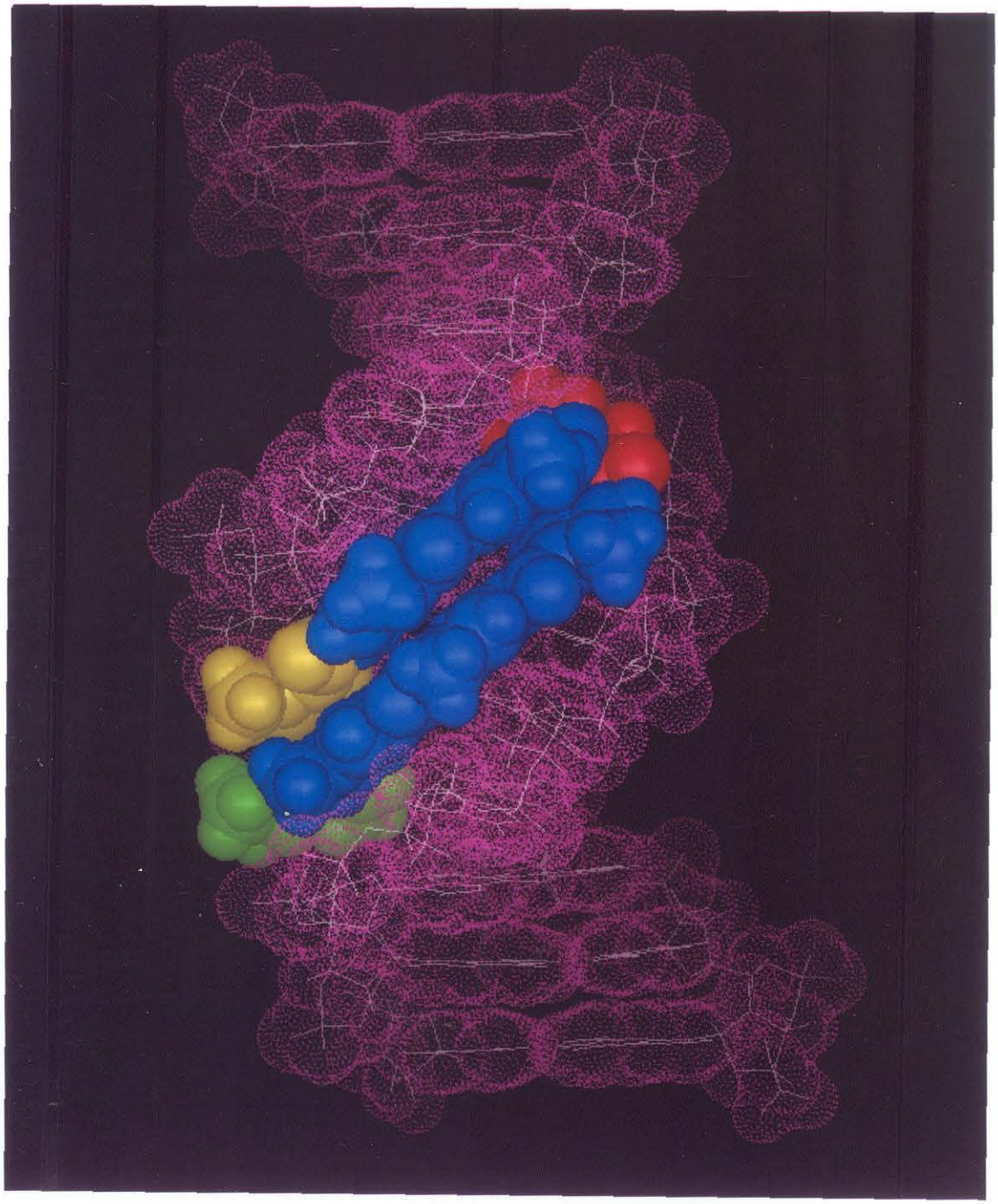


**Figure 1.21.** High resolution structure of the Im-Py-Py-γ-Py-Py-Py-Dp•5′-TGTTA-3′ complex derived from NMR data.[57] Colors represent the following amino acids: cyan, Py; yellow, Im; red, γ; green, Dp.

**Description of This Work.** In Chapter 2, the design of and synthesis of the cyclic polyamide *cyclo*-(Im-Py-Py-γ-Py-Py-Py-γ) is described. The cyclic polyamide binds the 5'-TGTTA-3' site with 40-fold higher affinity than the hairpin polyamide Im-Py-Py-γ-Py-Py-Py-Dp. These results demonstrate that wholly designed synthetic cyclic polyamides with a molecular weight of 950 can bind designed five base pair sequences at subnanomolar concentration.

In Chapter 3, in order to develop polyamide ligands to differentiate between A,T-rich sequences, a cyclic polyamide has been synthesized in which two Py-Py-Py units are circled through two γ-aminobutyric acids (γ). DNase I footprinting titration experiments reveal that *cyclo*-(Py-Py-Py-γ-Py-Py-Py-γ) binds the 5'-TATAT-3' sequence with 20-fold higher affinity than distamycin.

In Chapter 4, experiments are described in which EDTA•Fe(II) is attached at unique sites of HIV-1 TAR RNA by the chemical synthesis method and autocleavage of TAR-EDTA•Fe(II) is performed. Preliminary results showed that the autocleavage pattern of TAR RNA is changed in the presence of arginine. However, a series of control experiments indicate that the change in the autocleavage pattern is not caused by an RNA conformational change driven by the binding of arginine.

## References


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. Harrison, S.C. *Nature* **1991**, *353*, 715
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. Moser, H. E.; Dervan, P. B. *Science* **1987**, *238*, 645.
6. (a) Beal, P. A.; Dervan, P. B. *Science* **1991**, *251*, 1360. (b) Beal, P. A.; Dervan, P. B. *Nucl. Acids Res.* **1992**, *20*, 2773.
7. Singleton, S. F.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 6957.
8. Best, G. C.; Dervan, P. B. Manuscript in preparation.
9. Singleton, S. F.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 10376.
10. (a) Singleton, S. F.; Dervan, P. B. *Biochemistry* **1992**, *31*, 10995. (b) Povsic, T. J.; Dervan, P. B. *J. Am. Chem. Soc.* **1989**, *111*, 3059.
11. Singleton, S. F.; Dervan, P. B. *Biochemistry* **1993**, *32*, 13171.
12. 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.
13. (a) Jain, S. C.; Tsai, C. C.; Sobell, H. M. *J. Mol. Biol.* **1977**, *114*, 317. (b) Tsai, C. C.; Jain, S. C.; Sobel, H. M. *J. Mol. Biol.* **1977**, *114*, 301.

14. Wang, A. H.-J.; Ughetto, G.; Quigley, G. J.; Rich, A. *Biochemistry* **1987**, *26*, 1152.

15. (a) Low, C. M. L.; Drew, H. R.; Waring, M. J. *Nucleic Acids Res.* **1984**, *12*, 4865. (b) Van Dyke, M. W.; Dervan, P. B. *Science* **1984**, *225*, 1122. (c) Ughetto, G.; Wang, A. H.; Quigley, G. J.; van der Marel, G. A.; van Boom, J. H.; Rich, A. *Nucleic Acids Res.* **1985**, *13*, 2305.

16. (a) Fox, K. R.; Waring, M. J. *Nucleic Acids Res.* **1984**, *12*, 9271. (b) Lui, X.; Chen, H.; Patel, D. J. *J. Biomol. NMR* **1991**, *1*, 323. (c) Brown, D. R.; Kurz, M. Kearns, D. R.; Hsu, V. L. *Biochemistry* **1994**, *33*, 651.

17. Zimmer, C.; Wåhnert, U. *Prog. Biophys. Molec. Biol.* **1986**, *47*, 31.

18. Lane, M. J.; Dobrowiak, J. C.; Vournakis, J. *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 3260.

19. (a) Pjura, P. E.; Grzeskowiak, K.; Dickerson, R. E. *J. Mol. Biol.* **1987**, *197*, 257. (b) Fede, A.; Billeter, M.; Leupin, W.; Wuthrich, K. *Struct. Curr. Biol.* **1993**, *1*, 177.

20. (a) Gao, X.; Patel, D. J. *Biochemistry* **1989**, *28*, 751. (b) Gao, X.; Patel, D. J. *Biochemistry* **1990**, *29*, 10940. (c) Banville, D. L.; Keniry, M. A.; Kam, M.; Shafer, R. H. *Biochemistry* **1990**, *29*, 6521.

21. (a) Banville, D. L.; Keniry, M. A.; Shafer, R. H. *Biochemistry* **1990**, *29*, 9294. (b) Sastry, M.; Patel, D. J. *Biochemistry* **1993**, *32*, 6588.

22. Dervan, P. B. *Science* **1986**, *232*, 464.

23. (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.

24. Dreyer, G. B.; Dervan P. B. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 968.

25. (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.

26. (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.

27. (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.

28. (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.

29. The quatitative footprint titration methodology has been adapted for measurement of binding affinities of polyamide analogs to discreet 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.

30. Johnson, R. F.; Pickett, S. C.; Barker, D. L. *Electrophoresis* **1990**, *11*, 355.

31. (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.

32. (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.

33. Zimmer, C.; Wåhnert, U. *Prog. Biophys. Molec. Biol.* **1986**, *47*, 31.

34. (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) Tabernero, L.; Verdaguer, N.; Coll, M.; Fita, I.; van der Marel, G. A.; van Boom, J. H.; Rich, A.; Aymami, J. *Biochemistry* **1993**, *32*, 8403.

35. (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.

36. Zakrzewska, K.; Lavery, R.; Pullman, B. *J. Biomol. Struct. Dyn.* **1987**, *4*, 883.

37. (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.; Marky, L. A. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 8922.

38. (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.

39. 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.

40. Chen, X.; Ramakrishnan, B.; Rao, S. T.; Sundaralingam, M. *Struct. Biol. Nat.* **1994**, *1*, 169.

41. (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.

42. (a) Lee, M.; Chang, M. D.; Hartley, J. A.; Pon, R. T.; Krowicki, K.; Lown, J. W. *Biochemistry* **1988**, *27*, 445. (b) Rao, K. E.; Bathini, Y.; Lown, J. W. *J. Org. Chem.* **1990**, *55*, 728. (c) Plouvier, B.; Bailly, C.; Houssin, R.; Rao, K. E.; Lown, J. W.; Henichar, J.-P.; Waring, M. J. *Nucleic Acids Res.* **1991**, *19*, 5821.

43. Wade, W. S., Dervan, P. B. *J. Am. Chem. Soc.* **1987**, *109*, 1574.

44. Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 8783.

45. (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.

46. Dwyer, T. J.; Geierstanger, B. H.; Bathini, Y.; Lown, W. J.; Wemmer, D. E. *J. Am. Chem. Soc.* **1992**, *114*, 5912.

47. Yoon, C.; Prive, G. G.; Goodsell, D. S.; Dickerson, R. E. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 6332.

48. Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 2572.

49. Geierstanger, B. H.; Jacobsen, J.-P.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Biochemistry* **1994**, *33*, 3055.

50. Geierstanger, B. H.; Dwyer, T. J.; Bathini, Y.; Lown, J. W.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, *115*, 4474.

51. Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1995**, *117*, 3325.

52. Geierstanger, B. H.; Mrksich, M.; Dervan, P B.; Wemmer, D. E. *Science* **1994**, 266, 646.

53. Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 9892.

54. Dwyer, T. J.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, *115*, 9900.

55. Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 3663.

56. Mrksich, M.; Parks, M. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 7983.

57. Clairac, R. P. L.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.*. **1996**, submitted.

# Chapter 2

## Design and Synthesis of Cyclic Polyamides for Recognition in the Minor Groove of DNA

## Introduction

Footprinting, affinity cleavage, NMR and x-ray studies have established that polyamides containing *N*-methylimidazole (Im) and *N*-methylpyrrole (Py) amino acids can be combined in antiparallel side-by-side dimeric complexes with the minor groove of DNA[1-16]. The DNA sequence specificity of these small molecules can be controlled by the linear sequence of pyrrole and imidazole amino acids[1-13]. An imidazole ring on one ligand complemented by a pyrrolecarboxamide ring on the neighboring ligand recognizes a G•C base pair, while a pyrrolecarboxamide/ imidazole combination targets a C•G base pair[2-4]. A pyrrole/pyrrole pair is degenerate for A•T or T•A base pairs[1-4].

Despite this design breakthrough in molecular recognition of DNA, the binding affinities of this new class of polyamide dimers are modest[4]. For example, a three ring polyamide dimer in complex with a five base pair DNA site, has a binding affinity typically in the range of $K=2 \times 10^5\ M^{-1}$ (pH 7.0, 22 °C)[4]. In an effort to improve the energetics, antiparallel dimers were connected by a central γ-aminobutyric acid (γ) residue to create a single molecule which could bind in the minor groove by folding to a hairpin shape[11]. Second generation "hairpin" polyamides of sequence composition Im-Py-Py-γ-Py-Py-Py-Dp **1** were shown to have improved equilibrium association constants of

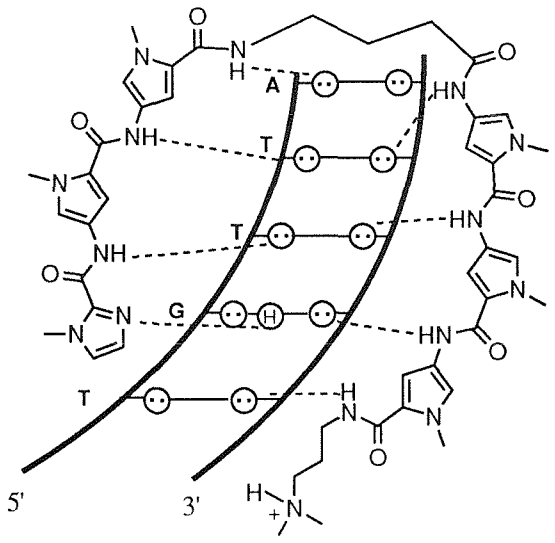
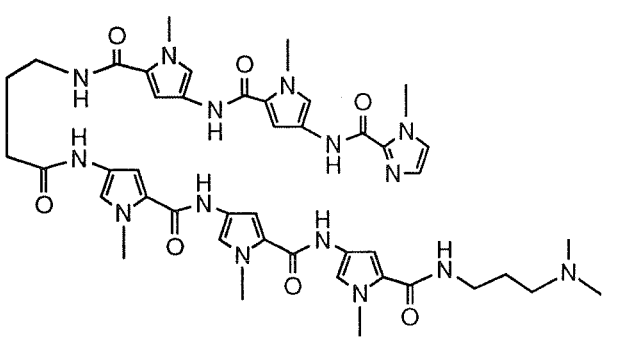
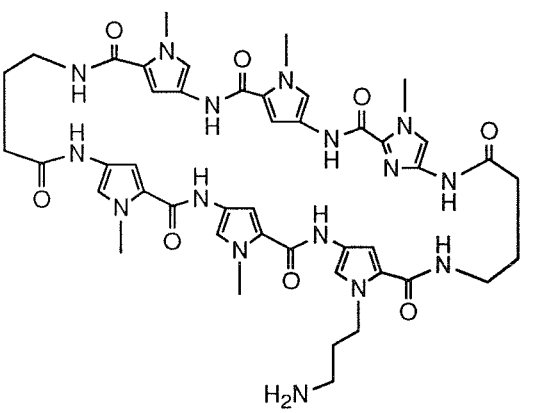



**Figure 2.1.** Model of the hairpin polyamide Im-Py-Py-γ-Py-Py-Py-Dp binding to a 5′-TGTTA-3′ site. Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines and circles containing a H represent the 2-amino group of guanine. Putative hydrogen bonds are indicated by dashed lines.

$8 \times 10^7\ M^{-1}$ for designated five base pair target sites 5'-TGTTA-3', an increase of 400 over the unlinked dimers (Figure 2.1). In a formal sense, γ-aminobutyric acid linked pyrrole-imidazole polyamides could exist in at least two conformations, hairpin or extended, resulting in different DNA binding motifs, and hence possible different DNA specificities. Closing the ends of the hairpin to a circle would restrict conformational space for the DNA binding molecule and presumably further increase the overall energetics. We report here the synthesis of *cyclo*-(Im-Py-Py-γ-Py-Py-Py-γ) **2** and analysis of the binding affinity and specificity of this new class of cyclic DNA binding small molecules (Figure 2.2).



Im-Py-Py-γ-Py-Py-Py-Dp **1**



*Cyclo*-(Im-Py-Py-γ-Py-Py-Py-γ) **2**

**Figure 2.2.** Structure of hairpin polyamide Im-Py-Py-γ-Py-Py-Py-Dp **1** and *cyclo*-(Im-Py-Py-γ-Py-Py-Py–γ) **2**.

**Results and Discussion**

<u>Cyclic polyamide synthesis</u> A cyclic polyamide **2** containing eight amino acids was synthesized in 12 steps from readily available starting materials (Figure 2.3). Ethyl 4-nitro-1-methylimidazole-2-carboxylate **3** was reduced with palladium under high pressure of hydrogen (250 psi) and the corresponding amine was coupled with the N-(tert-butyloxycarbonyl)amino-butyric acid (DCC, HOBt) to afford **4** in 98% yield. Treatment of the methyl ester **4** with NaOH afforded the corresponding acid **5** in 85% yield. Coupling with N-methyl-4-(N-methyl-4-aminopyrrole-2-carboxyamide)-pyrrole-2-carboxylic acid methyl ester (DCC, HOBt) provided ester **6** which was converted to amine **7** by deprotection with trifluoroacetic acid.

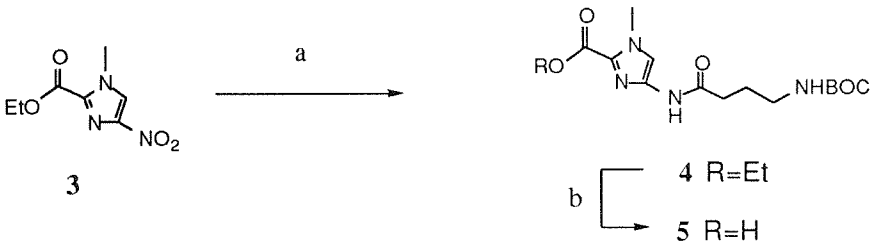
Alkylation of 4-nitro-2-carboxyethylpyrrole **8** with 3-bromo-1-(N-tert butoxycarbonyl)amino-propane afforded **9** in 99% yield. Reduction of the pyrrole ester **9** (250 psi $H_2$, Pd/C) and coupling with N-methyl-4-(N-methyl-4-

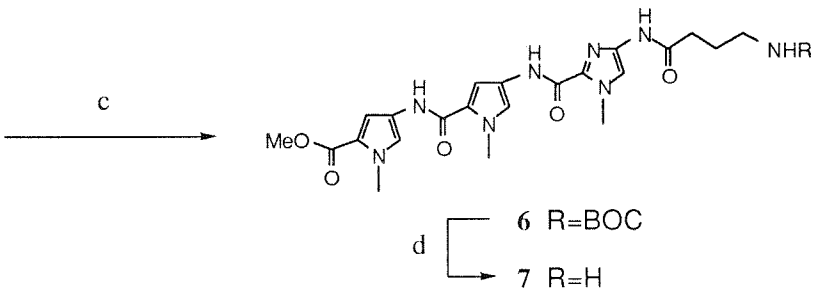
nitropyrrole-2-carboxamide)-pyrrole-2-carboxylic acid (DCC, HOBt) provided **10** in 76% yield. Catalytic hydrogenation of **7** (250 psi $H_2$, Pd/C) and coupling with the N-(benzyloxycarbonyl)amino-butylic acid (DCC, HOBt) afforded **11** in 79% yield. Saponification of the ethyl ester **11** (NaOH, 50 $^{o}$C) provided the corresponding acid **12** in 83% yield. Activation of the acid **12** (DCC, HOBt) and condensation with the amine **7** afforded Cbz-γ-Py-Py-Py-γ-Im-Py-Py-methyl ester **13** in 62% yield.

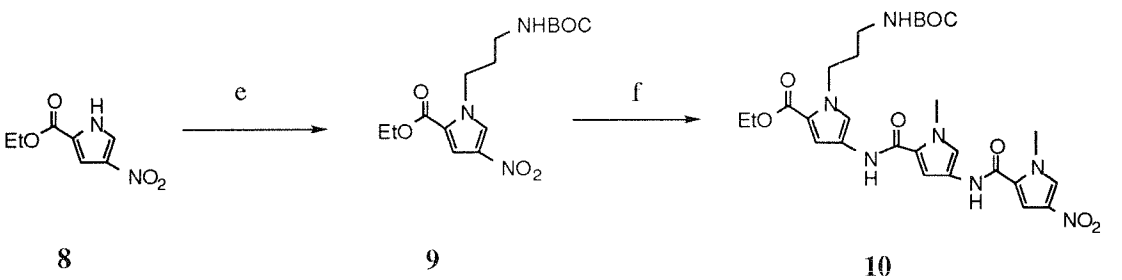
Treatment of the methyl ester **13** with NaOH gave the corresponding acid which, upon catalytic hydrogenation, furnished the desired linear polyamide **14**. The key cyclization reaction was accomplished in good yield (58%) using diphenylphosphoryl azide in the presence of $NaHCO_3$ followed by the deprotection with trifluoroacetic acid to give the cyclic product *cyclo*-(Im-Py-Py-γ-Py-Py-Py-γ) **2** which was purified by reverse-phase HPLC (Figure 2). The observed molecular mass of the $C_{45}H_{54}N_{16}O_8$ polyamide is 947.4330 (FABMS) in good agreement with the calculated value [M + H, 947.4389 calcd.][20].

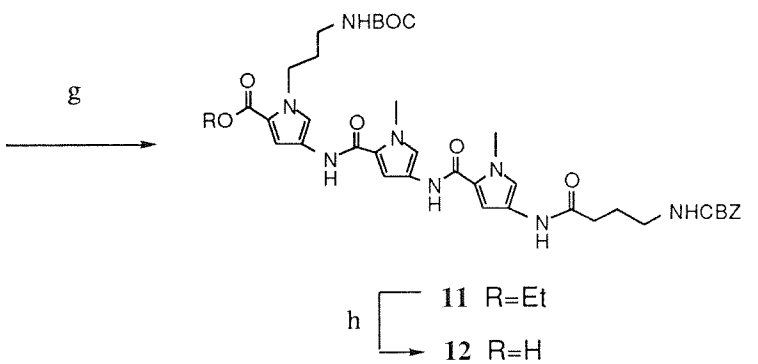
<u>Binding affinity and esquence specificity</u> Quantitative DNase I footprint titration experiments[21-24] on a $^{32}P$ end-labeled 135-base pair restriction fragment were performed to obtain the binding affinity of the cyclic polyamide **2** for a designated match site, 5'-TGTTA-3'. The cyclic polyamide binds the 5'-TGTTA-3' site very tightly with an equilibrium association constant of $K \geq 2.9 \times 10^9$ $M^{-1}$ (Table I). We attribute this increase of 40 in binding affinity over the corresponding hairpin polyamide **1** to a reduction in conformational entropy imparted via cyclization (Figure 2.5). The binding affinities to single base pair mismatch sites, 5'-ATTCG-3', 5'-AGAGT-3', and 5'-AGACA-3' are lower, 7.4 $x10^8$, 1.7 x $10^8$, and 6.2 x $10^8$ $M^{-1}$, respectively (Table I). The cycle appears to bind the match over single base pair mismatch by a

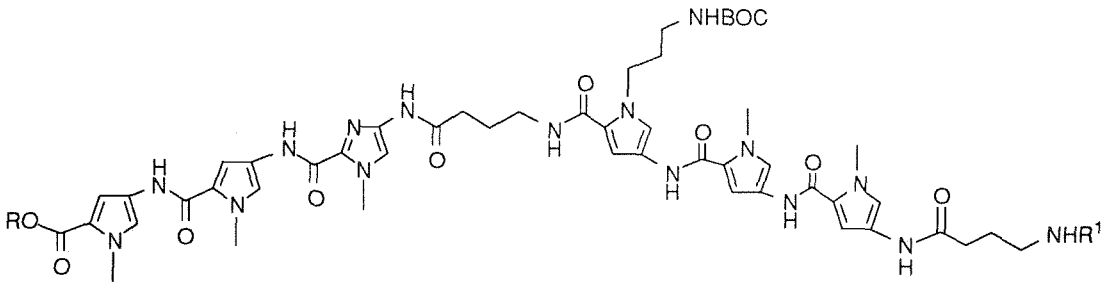
**Figure 2. 3.** Synthetic scheme for the synthesis of cyclic polyamide *cyclo*-(Im-Py-Py-γ-Py-Py-Py–γ) **2**. Reagents: (a) (i) 250 psi $H_2$, Pd/C; (ii) $BOCNH(CH_2)_3CO_2H$, DCC, HOBT; (b) NaOH, $MeOH/H_2O$; (c) N-Methyl-4-(N-methyl-4-aminopyrrole-2-carboxyamide)-pyrrole-2-carboxylic acid methyl ester, DCC, HOBt; (d) $TFA/CH_2Cl_2$; (e) $Br(CH_2)_3NHBOC$, KI, $K_2CO_3$; (f) (i) 250 psi $H_2$, Pd/C; (ii) N-Methyl-4-(N-methyl-4-nitropyrrole-2-carboxamide)-pyrrole-2-carboxylic acid, DCC, HOBt; (g) (i) 250 psi $H_2$, Pd/C; (ii) $CBZNH(CH_2)_3CO_2H$, DCC, HOBt; (h) NaOH, $MeOH-H_2O$; (i) **7**, DCC, HOBt; (j) (i) NaOH, $MeOH-H_2O$; (ii) 250 psi $H_2$, Pd/C; (k) DPPA, $NaHCO_3$, DMF; (l) TFA, Thiophenol;


a
EtO
NO2
RO
NHBOC
NH
3
b
4 R=Et
5 R=H


c
MeO
NHR
6 R=BOC
d
7 R=H


e
f
EtO
NO2
NHBOC
8
9
10


g
i
RO
NHBOC
NHCBZ
h
11 R=Et
12 R=H

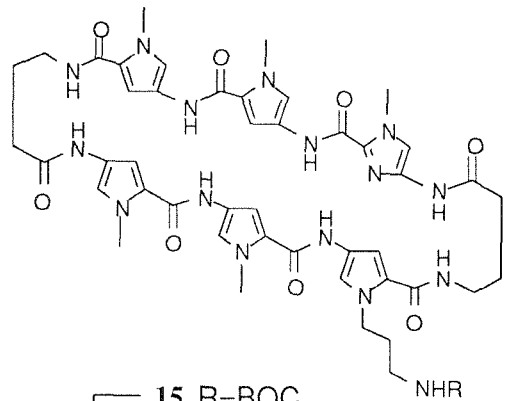



j

**13** R=Me, $R^1$=CBZ

**14** R= $R^1$=H

k




l

**15** R=BOC

**2** R=H

**Table I** Apparent First Order Binding Constants ($M^{-1}$)*,†

| | Match Site | Single Base Mismatch Sites | | |
|---|---|---|---|---|
| Polyamide | 5'-TGTTA-3' | 5'-ATTCG-3' | 5'-AGAGT-3' | 5'-AGACA-3' § |
| Hairpin **1** | $7.6 \times 10^7$ (0.8) | ‡ | $7.8 \times 10^5$ (2.0) | $2.6 \times 10^6$ (0.9) |
| Cyclic **2** | **$2.9 \times 10^9$** (1.9) | $7.4 \times 10^8$ (2.4) | $1.7 \times 10^8$ (1.4) | $6.2 \times 10^8$ (2.9) |

*Values reported are the mean values measured from at least three footprint titration experiments, with the standard deviation for each data set indicated in parentheses.

†The assays were performed at 22 °C at pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$.

‡Not determined due to the fact that hairpin **1** binds neighboring 5'-TGACA-3' single mismatch site.

§Determined on different restriction fragment.

factor of 4-17, although these values may be lower limits due to the uncertainty in the very high equilibrium association constant value for the cycle **2**:TGTTA complex. The hairpin binds more strongly to a match over single mismatch sites by a factor of twenty. The cyclic polyamide, though higher in affinity, is less than or only equal to the hairpin in specificity. It may be that there is an energetic price in specificity for closing the ends of the hairpin. Nevertheless, it is an encouraging step forward that wholly designed synthetic cyclic polyamides with a molecular weight of 950 bind designated five base pair sequences at subnanomolar concentrations. It remains a challenge to further optimize specificity.

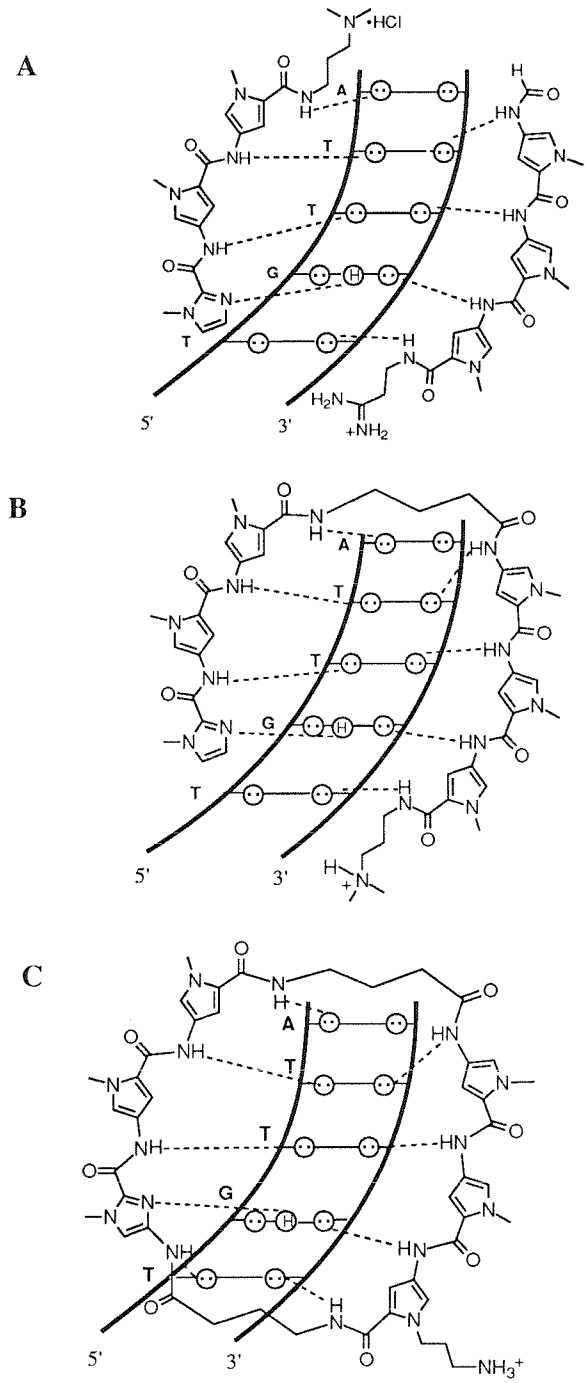



**Figure 2.4.** Models of (A) the heterodimer (Im-Py-Py-Dp/distamycin), (B) the hairpin polyamide Im-Py-Py-γ-Py-Py-Py-Dp, and (C) *cyclo*-(Im-Py-Py-γ-Py-Py-Py–γ) binding to a 5′-TGTTA-3′ site. Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines and circles containing a H represent the 2-amino group of guanine. Putative hydrogen bonds are indicated by dashed lines.

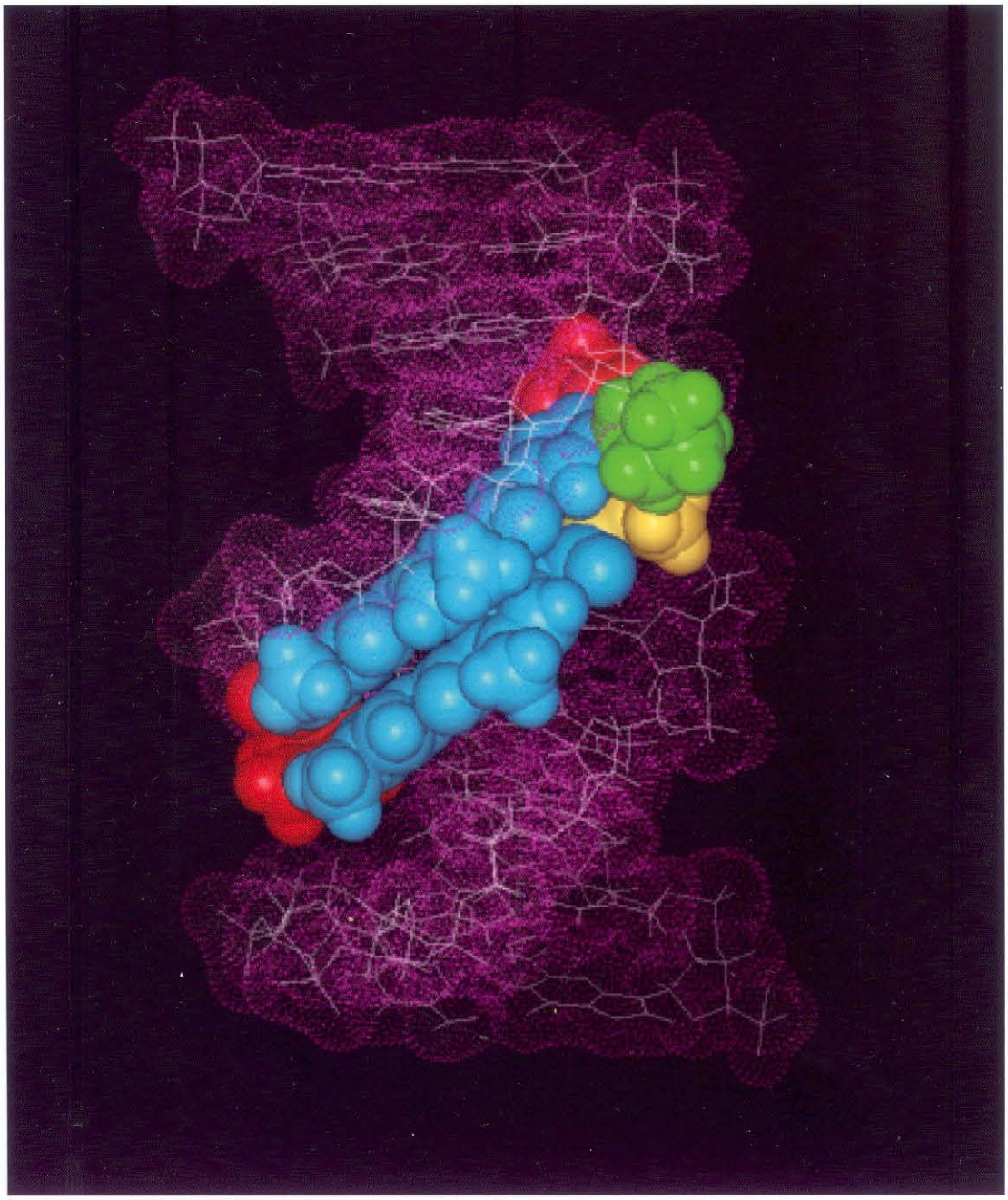


**Figure 2.5.** Proposed model of the (Im-Py-Py-γ-Py-Py-Py-γ)•5′-TGTTA-3′ complex. Colors represent the following amino acids: cyan, Py; yellow, Im; red, γ; green, Dp side chain.

**Experimental Section**

$^1H$ NMR spectra were recorded at 300 MHz on a GE 300 NMR in $CDCl_3$, DMSO-$d_6$, or $CD_3OD$. Chemical shifts are reported in parts per million relative to tetramethylsilane or residual DMSO-$d_5$. IR spectra were recorded on a Perkin-Elmer FTIR spectrometer. High-resolution mass spectra (HRMS) were recorded using fast atom bombardment (FAB) techniques at the Mass Spectrometry Laboratory at the university of California, Riverside. Preparatory HPLC was carried out on a Beckman Instrument using a Water DeltaPak 25 x 100 mm $C_{18}$ column. Analytical HPLC was performed on a Hewlett Packard 1090 Series II analytical HPLC using a Vydac $C_{18}$ reverse phase column (0.46 x 25 cm, 5 mm, HS silica). Flash column chromatography was carried out using silica gel 60 (230-400 mesh, Merck). Thin-layer chromatography (TLC) was performed on silica gel 60 $F_{254}$ precoated plates (Merck). All chemicals for the synthesis were purchased from Aldrich unless otherwise specified. Dichloromethane and N,N-dimethylformamide (DMF) were purchased as anhydrous solvents from Aldrich. Boc-GABA and Cbz-GABA were purchased from Sigma. All compounds were visualized with short-wave ultraviolet light.

**1-Methyl-4-[4-(N-tert-butoxycarbonyl)amino-butylcarboxylamino]-imidazole-2-carboxylic acid ethyl ester 4** To a solution of 4-(N-tert-butyloxycarbonyl)amino-butyric acid (1.8 g, 8.88 mmol) and N-hydroxybenzotriazole hydrate (1.26 g, 9.33 mmol) in DMF (20 mL) was added a solution of DCC (1.93 g, 9.33 mmol) in DMF (5 mL) and the reaction mixture was allowed to stir for 2 h. Separately, to a solution of **3** (1.18 g, 5.92 mmol) in DMF (25 mL) was added 10% Pd/C (150 mg) and the mixture was hydrogenated in a Parr bomb apparatus (250 psi, $H_2$) for 3 h. The catalyst was

removed by filtration through Celite and the filtrate was immediately added to the activated acid. The mixture was allowed to stir for 4 h and filtered through Celite. The solvent was removed in vacuo and the residue was purified by flash column chromatography (5% MeOH in $CH_2Cl_2$) to afford **4** (2.06 g, 98%). $^1$H NMR ($CDCl_3$) δ 8.71 (s, 1H), 7.50 (s, 1H), 4.94 (q, 2H, J=7.1Hz), 3.99 (s, 3H), 3.40 (t, 2H, J=7.2Hz), 3.19 (q, 2H, J=6.4Hz), 1.87 (m, 2H), 1.43 (s, 9H), 1.41 (t, 3H, J=7.1Hz); IR (KBr) 3373 (s), 3166 (m), 3084 (w), 2983 (s), 2936 (m), 2874 (w), 1693 (s), 1627 (w),1556 (s), 1480 (s), 1451 (s), 1407 (s), 1388 (s), 1365 (s), 1320 (s), 1300 (m), 1254 (s), 1179 (s), 1138 (s), 1109 (m), 1078 (w), 1060 (s), 1017 (m), 989 (w), 807 (m) $cm^{-1}$; FABMS m/e 355.1988 (M+H, 355.1981 calcd. for $C_{16}H_{26}N_4O_5$).

**1-Methyl-4-[4-(N-tert-butoxycarbonyl)amino-butylcarboxylamino]-imidazole-2-carboxylic acid 5** 1-Methyl-4-[4-(N-tert-butoxycarbonyl)amino-butylcarboxylamino]-imidazole-2-carboxylic acid ethyl ester **4** (2.06 g, 5.81 mmol) was dissolved in MeOH (50 mL) and 1N NaOH (50 mL) was added. The reaction mixture was heated to 50 °C and allowed to stir for 4 h. The solution was filtered and the filtrate was acidified to pH2 with 1N HCl. The white precipitate was collected by filtration and dried to afford **5** (1.65 g, 87%). $^1$H NMR (DMSO-$d_6$) δ 10.52 (s, 1H), 7.47 (s, 1H), 6.83 (brs, 1H), 3.87 (s, 3H), 2.90 (q, 2H, J=6.6Hz), 2.23 (t, 2H,7.3Hz), 1.63 (m, 2H), 1.35 (s, 9H); IR (KBr) 3346 (s), 3186 (w), 2881 (m), 1933 (w), 1683 (s), 1614 (s), 1531 (s), 1475 (w), 1455 (m), 1413 (w), 1342 (s), 1275 (s), 1221 (w), 1180 (s), 1088 (w), 997 (w), 828 (w) $cm^{-1}$; FABMS m/e 325.1527 (M-H, 325.1512 calcd. for $C_{14}H_{22}N_4O_5$).

**2-Bis(N-methylpyrrolecarboxamide)-1-methyl-4-[4-(N-tert-butoxycarbonyl) amino-butylcarboxylamino]-imidazole 5** To a solution of 1-methyl-4-[4-(N-tert-butoxycarbonyl)amino-butylcarboxylamino]-imidazole-2-carboxylic acid **4** (1.61 g, 4.96 mmol) and N-hydroxybenzotriazole hydrate

(0.74g, 5.46 mmol) in DMF (35 mL) was added a solution of DCC (1.13 g, 5.46 mmole) in DMF (5 mL) and the reaction mixture was allowed to stir for 2 h. A solution of N-methyl-4-(N-methyl-4-aminopyrrole-2-carboxamide)-pyrrole-2-carboxylic acid methyl ester (1.7 g, 5.5 mmol) in DMF (10 mL) was added and the mixture was allowed to stir for 4 h. The solvent was removed in vacuo and the residue was purified by flash column chromatography (7% MeOH in $CH_2Cl_2$) to afford **5** (2.47 g, 85%). $^1H$ NMR (DMSO-$d_6$) δ 10.22 (s, 1H), 10.01 (s, 1H), 9.95 (s, 1H), 7.47 (d, 1H, J=1.2Hz), 7.43 (s, 1H), 7.26 (d, 1H, J=0.9Hz), 7.14 (d, 1H, J=1.1Hz), 6.90 (d, 1H, J=1.5Hz), 3.94 (s, 3H), 3.84 (s, 3H), 3.83 (s, 3H), 3.73 (s, 3H), 2.91 (m, 2H), 2.30 (t, 2H, J=7.2Hz), 1.63 (m, 2H), 1.37 (s, 9H); IR (KBr) 3650 (m), 3385 (s), 1654 (s), 1544 (s), 1438 (s), 1406 (s), 1254 (s), 1168 (m), 1119 (m), 780 (w) $cm^{-1}$; FABMS m/e 584.2710 (M, 584.2707 calcd. for $C_{27}H_{36}N_8O_7$).

**2-Bis(N-methylpyrrolecarboxamide)-1-methyl-4-(4-amino-butylcarboxylamino)-imidazole 6** To a solution of 2-Bis(N-methylpyrrolecarboxamide)-1-methyl-4-[4-(N-tert-butoxycarbonyl)amino-butylcarboxylamino]-imidazole **5** (2.0 g, 3.41 mmol) in $CH_2Cl_2$ (20 mL) was added TFA (10 mL) and the resulting solution was allowed to stir for 4 h. The mixture was evaporated in vacuo and the residue was purified by flash column chromatography (10% MeOH in $CH_2Cl_2$) to afford **6** (1.51 g, 91%). $^1H$ NMR (DMSO-$d_6$) δ 10.38 (s, 1H), 10.03 (s, 1H), 10.01 (s, 1H), 7.49 (d, 1H, J=1.6Hz), 7.46 (s, 1H), 7.29 (d, 1H, J=1.2Hz), 7.17 (d, 1H, J=1.2Hz), 6.92 (d, 1H, J=1.7Hz), 3.96 (s, 3H), 3.86 (s, 3H), 3.84 (s, 3H), 3.74 (s, 3H), 2.84 (t, 2H, J=7.0Hz), 2.43 (t, 2H, J=7.2Hz), 1.84 (m, 2H); IR (KBr) 3451 (s), 1685 (s), 1560 (s), 1438 (m), 1406 (m), 1262 (m), 1206 (s), 1137 (m), 1061 (w), 802 (w), 724 (w) $cm^{-1}$; FABMS m/e 485.2266 (M+H, 485.2261 calcd. for $C_{22}H_{28}N_8O_5$).

**1-[3-(N-tert-butoxycarbonyl)amino-propyl]-4-nitro-pyrrole-2-carboxylic acid ethyl ester 9** To a solution of 4-nitropyrrol-2-carboxylic acid ethyl ester 7

(3.8 g, 19.98 mmol) in acetone (30 mL) was added $K_2CO_3$ (8.85 g, 64.02 mmol) and the resulting solution was allowed to stir at room temperature for 1 h. 3-Bromo-1-(N-tert-butoxycarbonyl)amino-propane (5.32 g, 21.97 mmol) and KI (3.65 g, 21.97 mmol) were added and the mixture was warmed at reflux (65°C) for 6 h. The solution was filtered and the solvent was removed in vacuo. The residue was purified by flash column chromatography (30% EtOAc in Hexane) to afford **9** (6.9 g, 99%). $^1$H NMR ($CDCl_3$) δ 7.73 (s, 1H), 7.44 (d, 1H, J=1.9Hz), 4.42 (t, 2H, J=7.0Hz), 4.32 (q, 2H, J=7.1Hz), 3.17 (m, 2H), 2.01 (m, 2H), 1.45 (s, 9H), 1.38 (t, 3H, J=7.1Hz); IR (KBr) 3318 (s), 3155 (m), 2982 (m), 2934 (m), 1716 (s), 1675 (s), 1537 (s), 1506 (s), 1419 (m), 1365 (s), 1317 (s), 1252 (s), 1231 (m), 1195 (s), 1165 (m), 1141 (m), 1103 (m), 1080 (m), 1054 (m), 1011 (m), 877 (w), 849 (w), 823 (w), 812 (m), 749 (m), 675 (w), 637 (w) $cm^{-1}$; FABMS m/e 340.1506 (M-H, 340.1509 calcd. for $C_{15}H_{23}N_3O_6$).

**1-[3-(N-tert-butoxycarbonyl)amino-propyl]-4-bis(N-methylpyrrole carboxamino)-pyrrole-2-carboxylic acid ethyl ester 10** To a solution of N-Methyl-4-(N-methyl-4-nitropyrrole-2-carboxamide)-pyrrole-2-carboxylic acid (3.8 g, 13 mmol) and N-hydroxybenzotriazole hydrate (1.93 g, 14.3 mmol) in DMF (30 mL) was added a solution of DCC (2.95 g, 14.3 mmol) in DMF (6 mL) and the reaction mixture was allowed to stir for 2 h. Separately, to a solution of 1-[3-(N-tert-butoxycarbonyl)amino-propyl]-4-nitro-pyrrole-2-carboxylic acid ethyl ester **9** (3.7 g, 10.69 mmol) in DMF (25 mL) was added 10% Pd/C (200 mg) and the mixture was hydrogenated in a Parr bomb apparatus (250 psi, $H_2$) for 3 h. The catalyst was removed by filtration through Celite and the filtrate was immediately added to the activated acid. The mixture was allowed to stir for 4 h and filtered through Celite. The solvent was removed in vacuo and the residue was purified by flash column chromatography (10% MeOH in $CH_2Cl_2$) to afford **10** (4.76 g, 76%). $^1$H NMR [$CDCl_3$-$CD_3OD$ (2:1)] δ 7.76 (s, 1H), 7.38 (s,

2H), 7.21 (s, 1H), 6.93 (s, 1H), 6.88 (s, 1H), 4.30 (t, 2H, J=7.5Hz), 4.24 (q,2H, J=7.1Hz), 3.98 (s, 3H), 3.89 (s, 3H), 3.03 (m, 2H), 1.91 (m, 2H), 1.42 (s, 9H), 1.32 (t, 3H, J=7.1Hz); IR (KBr) 3568 (w), 3386 (m), 3118 (w), 2977 (w), 1686 (s), 1508 (s), 1404 (s), 1366 (s), 1311 (s), 1250 (s), 1206 (s), 1168 (s), 1114 (s) $cm^{-1}$; FABMS m/e 585.2557 (M, 585.2547 calcd. for $C_{27}H_{35}N_7O_8$).

**1-[3-(N-tert-butoxycarbonyl)amino-propyl]-4-[4-[4-(N-benzyloxycarbonyl) amino-butylcarboxylamino]-bis[N-methylpyrrolecarboxamide]]-pyrrole-2-carboxylic acid ethyl ester 11** To a solution of 4-(N-benzyloxycarbonyl)amino-butylic acid (787 mg, 3.3 mmol) and N-hydroxybenzotriazole hydrate (468 mg, 3.47 mmol) in DMF (15 mL) was added a solution of DCC (715 mg, 3.47 mmol) in DMF (3 mL) and the reaction mixture was allowed to stir for 2 h. Separately, to a solution of 1-[3-(N-tert-butoxycarbonyl)amino-propyl]-4-bis(N-methylpyrrolecarboxamino)-pyrrole-2-carboxylic acid ethyl ester **10** (1.85 g, 3.16 mmol) in DMF (15 mL) was added 10% Pd/C (100 mg) and the mixture was hydrogenated in a Parr bomb apparatus (250 psi, $H_2$) for 3 h. The catalyst was removed by filtration through Celite and the filtrate was immediately added to the activated acid. The mixture was allowed to stir for 4 h and filtered through Celite. The solvent was removed in vacuo and the residue was purified by flash column chromatography (10% MeOH in $CH_2Cl_2$) to afford **11** (1.93 g, 79%). $^1H$ NMR ($CDCl_3$) δ 8.07 (s, 1H), 7.43 (s, 1H), 7.29 (s, 5H), 7.27 (s, 1H), 7.21 (s, 1H), 6.96 (s, 1H), 6.94 (s, 1H), 5.05 (s, 2H), 4.28 (t, 2H, J=7.3Hz), 4.22 (q, 2H, J=7.1Hz), 3.84 (s, 3H), 3.77 (s, 3H), 3.21 (m, 2H), 3.06 (m, 2H), 2.28 (t, 2H, J=6.3Hz), 1.87 (m, 4H), 1.43 (s, 9H), 1.30 (t, 3H, J=7.1Hz); IR (KBr) 3314 (s), 2976 (m), 1700 (s), 1538 (s), 1404 (s), 1255 (s), 1204 (s), 1168 (m), 1106 (m), 1004 (w), 778 (w), 697 (w) $cm^{-1}$; FABMS m/e 774.3711 (M, 774.3701 calcd. for $C_{39}H_{50}N_8O_9$).

**1-[3-(N-tert-butoxycarbonyl)amino-propyl]-4-[4-[4-(N-benzyloxycarbonyl) amino-butylcarboxylamino]-bis[N-methylpyrrolecarboxamide]]-pyrrole-2-carboxylic acid 12** Ethyl ester **11** (1.75 g, 2.26 mmol) was dissolved in MeOH (30 mL) and 1N NaOH (13.6 mL, 13.6 mmol, 6eq.) was added. The reaction mixture was heated to 60 $^{o}$C and allowed to stir for 4 h. The solution was filtered and the filtrate was acidified to pH 2 with 1N HCl. The white precipitate was collected by filtration and dried to afford **12** (1.4 g, 83%). $^{1}$H NMR (DMSO-$d_6$) δ 9.93 (s, 1H), 9.89 (s, 1H), 9.84 (s, 1H), 7.34 (s, 5H), 7.32 (d, 1H, J=0.6Hz), 7.23 (d, 1H, J=0.6Hz), 7.16 (d, 1H, J=0.9Hz), 7.03 (s, 1H), 6.86 (d, 1H, J=0.9Hz), 6.76 (s, 1H), 5.0 (s, 2H), 4.29 (t, 2H, J=7.3Hz), 3.83 (s, 3H), 3.82 (s, 3H), 3.02 (m, 2H), 2.87 (m, 2H), 2.24 (t, 2H, J=7.3Hz), 1.73 (m, 4H), 1.36 (s, 9H); IR (KBr) 3340 (s), 2976 (m), 1699 (s), 1543 (s), 1438 (s), 1406 (s), 1256 (s), 1206 (m), 1167 (m), 1105 (w), 1061 (w) cm$^{-1}$; FABMS m/e 746.3358 (M, 746.3388 calcd. for $C_{37}H_{46}N_8O_9$).

**CbZ-γ-Py-Py-Py-γ-Im-Py-Py-methyl ester 13** To a solution of the acid **12** (1.12 g, 1.5 mmol) and N-hydroxybenzotriazole hydrate (216 mg, 1.6 mmol) in DMF (20 mL) was added a solution of DCC (330 mg, 1.6 mmole) in DMF (2 mL) and the reaction mixture was allowed to stir for 2 h. A solution of the primary amine **7** (800 mg, 1.65 mmol) in DMF (10 mL) was added and the mixture was allowed to stir for 4 h. The solvent was removed in vacuo and the residue was purified by flash column chromatography (10% MeOH in $CH_2Cl_2$) to afford **13** (1.13 g, 62%). $^{1}$H NMR (DMSO-$d_6$) δ 10.29 (s, 1H), 10.05 (s, 1H), 9.97 (s, 2H), 9.94 (s, 2H), 9.83 (s, 1H), 7.47 (s, 2H), 7.34 (s, 5H), 7.27 (s, 2H), 7.25 (s, 1H), 7.16 (s, 2H), 7.05 (s, 1H), 6.90 (s, 2H), 6.88 (s, 1H), 5.0 (s, 2H), 4.26 (brs, 2H), 3.94 (s, 3H), 3.84 (s, 6H), 3.83 (s, 6H), 3.72 (s, 3H), 3.20 (m, 2H), 3.02 (m, 2H), 2.87 (m, 2H), 2.36 (t, 2H, J=6.8Hz), 2.24 (t, 2H, J=6.7Hz), 1.73 (m, 6H), 1.36 (s, 9H); IR (KBr) 3305 (m), 2946 (m), 1658 (s), 1538 (s), 1254 (s), 1107 (m), 1060 (w), 890

(w), 777 (w) $cm^{-1}$; FABMS m/e 1213.5581 (M+H, 1213.5543 calcd. for $C_{59}H_{72}N_{16}O_{13}$).

**γ-Py-Py-Py-γ-Im-Py-Py 14** CbZ-γ-Py-Py-Py-γ-Im-Py-Py-methyl ester **13** (420 mg, 0.35 mmol) was dissolved in MeOH (10 mL) and 1N NaOH (2.1 mL, 2.1 mmol, 6eq.) was added. The reaction mixture was heated to 60 °C and allowed to stir for 3 h. The mixture was acidified to pH 2 with 1N HCl. The precipitate was collected by filtration and dried to afford the acid. Without further purification, the acid was dissolved in DMF (5 mL) and 10% Pd/C (50 mg) was added. The mixture was hydrogenated in a Parr bomb apparatus (250 psi, $H_2$) for 6 h and filtered with Celite. The filtrate was concentrated in vacuo to afford **14** (305 mg, 83%) $^1$H NMR (DMSO-$d_6$) δ 10.32 (s, 1H), 10.13 (s, 1H), 9.97 (s, 2H), 9.95 (s, 2H), 9.93 (s, 1H), 7.46 (s, 1H), 7.33 (s, 1H), 7.26 (s, 2H), 7.24 (s, 1H), 7.18 (s, 1H), 7.14 (s, 1H), 7.04 (s, 1H), 6.89 (s, 1H), 6.88 (s, 2H), 6.77 (s, 1H), 4.24 (brs, 2H), 3.94 (s, 3H), 3.83 (s, 6H), 3.82 (s, 6H), 3.17 (m, 2H), 2.81 (m, 4H), 2.36 (m, 4H), 1.78 (m, 6H), 1.35 (s, 9H); IR (KBr) 3402 (s), 1654 (s), 1543 (s), 1466 (m), 1438 (s),1406 (m), 1259 (w), 1209 (w), 1108 (w) $cm^{-1}$; FABMS m/e 1087.4892 (M+Na, 1087.4838 calcd. for $C_{50}H_{64}N_{16}O_{11}$).

**Boc-protected Cyclic polyamide 15** To a solution of γ-Py-Py-Py-γ-Im-Py-Py **14** (150 mg, 0.18 mmol) in DMF (70 mL) was sequentially added $NaHCO_3$ (78 mg, 0.93 mmol) and diphenylphosphoryl azide (DPPA, 154 mg, 0.56 mmol). The mixture was allowed to stir at room temperature for 3 days and then filtered. The filtrate was concentrated in vacuo and the residue was purified by flash column chromatography (15% MeOH in $CH_2Cl_2$) to afford the cyclic polyamide **15** (90 mg, 58%) $^1$H NMR [$CDCl_3$-$CD_3OD$ (1:1)] δ 7.30 (s, 1H), 7.26 (s, 1H), 7.18 (s, 1H), 7.02 (s, 2H), 6.88 (s, 1H), 6.70 (s, 1H), 6.68 (s, 1H), 6.62 (s, 1H), 6.58 (s, 1H), 6.43 (s, 1H), 4.35 (m, 2H), 3.92 (s, 6H), 3.90 (s, 3H), 3.89 (s, 3H), 3.82 (s, 3H), 3.49 (m, 4H), 3.08 (m, 2H), 2.42 (m, 4H), 1.93-2.15 (m, 6H), 1.45 (s, 9H); IR

(KBr) 3422 (s), 1654 (s), 1545 (s), 1438 (s), 1406 (m), 1260 (m), 1208 (m), 1105 (w), 919 (w), 777 (w) $cm^{-1}$; FABMS m/e 1047.4884 (M+H, 1047.4913 calcd. for $C_{50}H_{62}N_{16}O_{10}$).

**Cyclic polyamide 2** To a mixture of TFA and thiophenol (4 mL, 0.5 M PhSH in TFA) was added the Boc-protected Cyclic polyamide **15** (30 mg, 28.6 mmol) and the reaction mixture was allowed to stir at room temperature for 2 h. Excess TFA was removed in vacuo and the residue was purified by flash column chromatography (5% $NH_4OH$ in MeOH) to afford **2** (23 mg, 87%). The product was further purified by using reverse-phase HPLC on a preparatory Water DeltaPak 25 x 100 mm 100 mm C18 column with linear gradients of 60% acetonitrile plus 0.1% TFA versus 0.1% aqueous TFA. $^1H$ NMR (DMSO-$d_6$) δ 10.00 (s, 1H), 9.99 (s, 1H), 9.96 (s, 2H), 9.95 (s, 2H), 9.87 (s, 1H), 7.47 (s, 2H), 7.44 (s, 1H), 7.39 (s, 1H), 7.34 (s, 1H), 7.16 (s, 1H), 6.94 (s, 1H), 6.90 (s, 1H), 6.88 (s, 1H), 6.71 (s, 1H), 6.65 (s, 1H), 4.29 (m, 2H), 3.94 (s, 3H), 3.84 (s, 3H), 3.83 (s, 3H), 3.82 (s, 3H), 3.79 (s, 3H), 3.22 (brs, 4H), 2.52 (m, 2H), 2.39 (m, 2H), 2.30 (t, 2H, J=7.0Hz), 1.78 (m, 6H); IR 3383 (s), 1637 (s), 1560 (s), 1438 (s), 1407 (s), 1265 (m), 1207 (w), 1122 (w), 1062 (w), 777 (w) $cm^{-1}$; UV ($H_2O$) $I_{max}$ (e) 240 (37,075 $cm^{-1}M^{-1}$), 310 (48,096 $cm^{-1}M^{-1}$) nm; FABMS m/e 947.4330 (M+H, 947.4389 calcd. for $C_{45}H_{54}N_{16}O_8$).

## References


1. Pelton, J. G.; Wemmer, D. E. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **1989**, 86, 5723-5727.
2. Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, 89, 7586-7590.
3. Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, 114, 8783-8794.
4. Wade, W. S.; Mrksich, M.; Dervan, P. B. *Biochemistry* **1993**, 32, 11385-11389.
5. Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, 115, 2572-2576.
6. Geierstanger, B. H.; Dwyer, T. J.; Bathini, Y.; Lown, J. W.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, 115, 4474-4482.
7. Dwyer, T. J.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, 115, 9900-9906.
8. Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, 115, 9892-9899.
9. Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, 116, 3663-3664.
10. Geierstanger, B. H.; Jacobsen, J. P.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Biochemistry* **1994**, 33, 3055.
11. Mrksich, M.; Parks, M. E.; Dervan, P. B. *J. Am. Chem.* Soc. **1994**, 116, 7983-7988.
12. Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Science* **1994**, 266, 646-650.
13. Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1995**, 117, 3325.
14. Pelton, J. G.; Wemmer, D. E. *J. Am. Chem. Soc.* **1990**, 112, 1393-1399.
15. Fagan, P. A. ; Wemmer, D. E. *J. Am. Chem. Soc.* **1992**, 114, 1080-1081.

16. Chen, X.; Ramakrishnan, B.; Rao, S. T.; Sundaralingam, M. *Nature Struct. Biology* **1994**, 1, 169-175.

17. Iverson, B. L.; Dervan P.B. *Nucleic Acids Res.* **1987**, 15, 7823-7830.

18. Maxam, A. M.; Gilbert, W. S. *Methods in Enzymology* **1980**, 65, 499-560.

19. Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular Cloning*, **1989**, Cold Spring Harbor Laboratory: Cold Spring Harbor, NY.

20. We thank UC Riverside Mass Spec Facility for the analysis.

21. Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Methods Enzymol.* **1986**, 130, 132-181.

22. Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, 83, 8462-8466

23. Senear, D. F.; Brenowitz, M.; Shea, M. A.; Ackers, G. K. *Biochemistry* **1986**, 25, 7344-7354.

24. The quantitative footprint titration experiments and data analysis were performed as previously described (11), with unlabeled carrier absent from all reactions.

# Chapter 3

## Recognition in the Minor Groove of DNA at 5′-$(A,T)_5$-3′ by the Cyclic Polyamide : *cyclo*-(Py-Py-Py-γ-Py-Py-Py-γ)

### Introduction

Distamycin A is a naturally occurring oligopolyamide that binds in the minor groove of DNA at sites of five successive A•T base pairs[1-6]. NMR and crystallographic studies of 1:1 distamycin-DNA complexes have shown that hydrogen bonding, van der Waals contacts, and electrostatics all contribute to the binding affinity and specificity[7-9]. The width of the minor groove of B-DNA is sequence dependent and can vary from 3-4 Å for A,T-rich DNA to 6-7 Å for G,C-rich regions[10]. The narrow minor groove of AT-tract of DNA may favor a 1:1 complex with distamycin.

Efficient discrimination between A•T and T•A base pairs by distamycin may not be possible since two hydrogen acceptors, the adenine N3 and thymine O2 atoms, are quite symmetrically positioned in the minor groove. Although the molecular basis of DNA recognition by distamycin is fairly well understood, a general solution for discrimination between A•T and T•A base pairs has still not been achieved. Netropsin and distamycin are known to bind to alternating ATAT regions less efficiently than to continuous AAAA or TTTT regions[11-13]. Recent two-dimensional NMR studies have demonstrated that distamycin at high concentration (2-4 mM) is also capable of binding in the minor groove of the 5′-AAATT-3′ sequence as a side-by-side dimer[14,15].

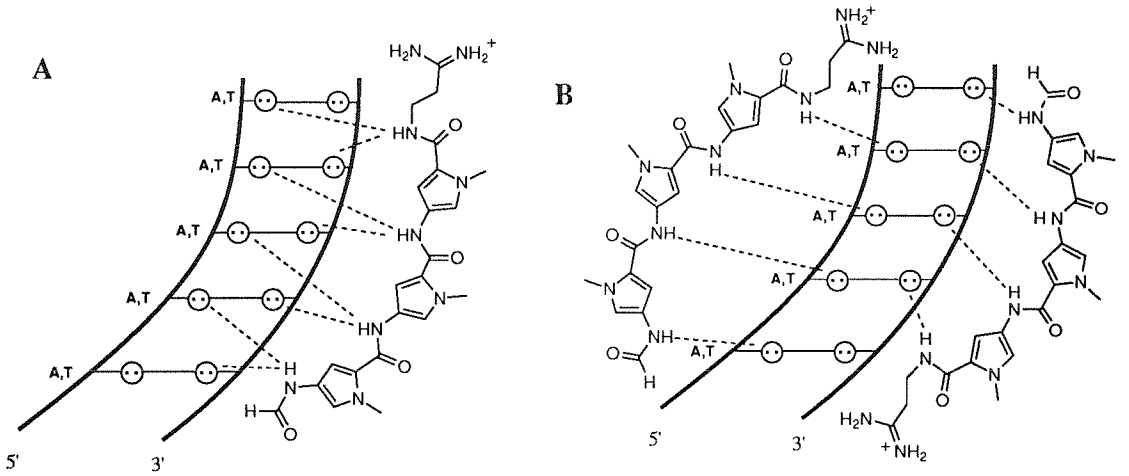
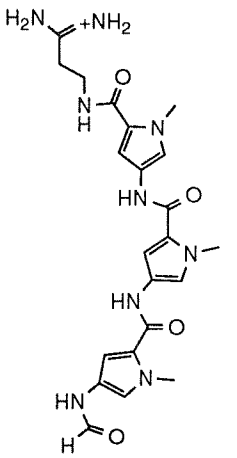
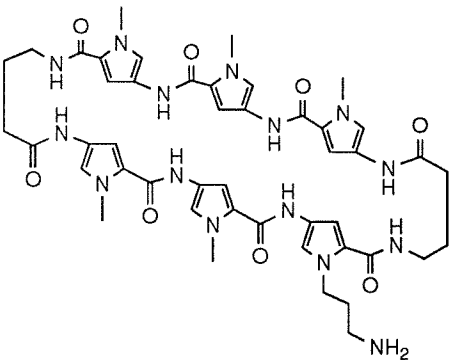



**Figure 3.1.** (A) 1:1 binding model and (B) 2:1 binding model for the complex between distamycin A and 5′-(A,T)5-3′ sequence. Circles with dots represent lone pairs of N3 of adenines and O2 of thymines. Putative hydrogen are illustrated by dashed lines.

These results indicate that the minor groove of A,T-rich regions of DNA is flexible enough to accomodate two drug molecules and the minor groove width may affect the binding affinity of distamycin. In general, formation of 1:1 polyamide-DNA complexes are favored in duplex with a narrow minor groove, whereas 2:1 polyamide-DNA complexes are favored in cases where the minor groove is wide. Two competing binding modes may be the main reason for the modest sequence selectivity between different AT sequences observed for distamycin. We recently reported that a new class of cyclic polyamides binds to the 5′-TGTTA-3′ at subnanomolar concentrations[16]. We report here the synthesis of cyclo-(Py-Py-Py-γ-Py-Py-Py-γ) **1** and compare its the binding affinity and specificity with distamycin.



Distamycin A



*Cyclo*-(Py-Py-Py-γ-Py-Py-Py-γ)

**Figure 3.2.** Structure of distamycin A and cyclic polyamide *cyclo*-(Py-Py-Py-γ-Py-Py-Py-γ).

## Results and Discussion

### Synthesis of cyclic polyamide

A scheme for the synthesis of cyclic polyamide **1** is shown in Figure 3.3. **2** is available in five steps from N-methyl pyrrole-2-carboxylic acid in 20% overall yield[23,24]. Nitro-tripyrrole **2** was reduced with palladium under high pressure of hydrogen (250 psi) and the corresponding amine was coupled with the N-(tert-butyloxycarbonyl)amino-butyric acid (DCC, HOBt) to afford **3** in 95% yield. Deprotection of the amine with TFA provided **4** in 90% yield.

**5** was synthesized in four steps from 4-nitro-2-carboxyethylpyrrole as described in chapter 2. Activation of the acid **5** (DCC, HOBt) and condensation with the amine **4** afforded Cbz-γ-Py-Py-Py-γ-Py-Py-Py-methyl ester **6** in 75% yield.Treatment of the methyl ester **6** with NaOH gave the corresponding acid which, upon catalytic hydrogenation, furnished the desired linear polyamide **7**. The cyclization reaction of **7** was accomplished using diphenylphosphoryl

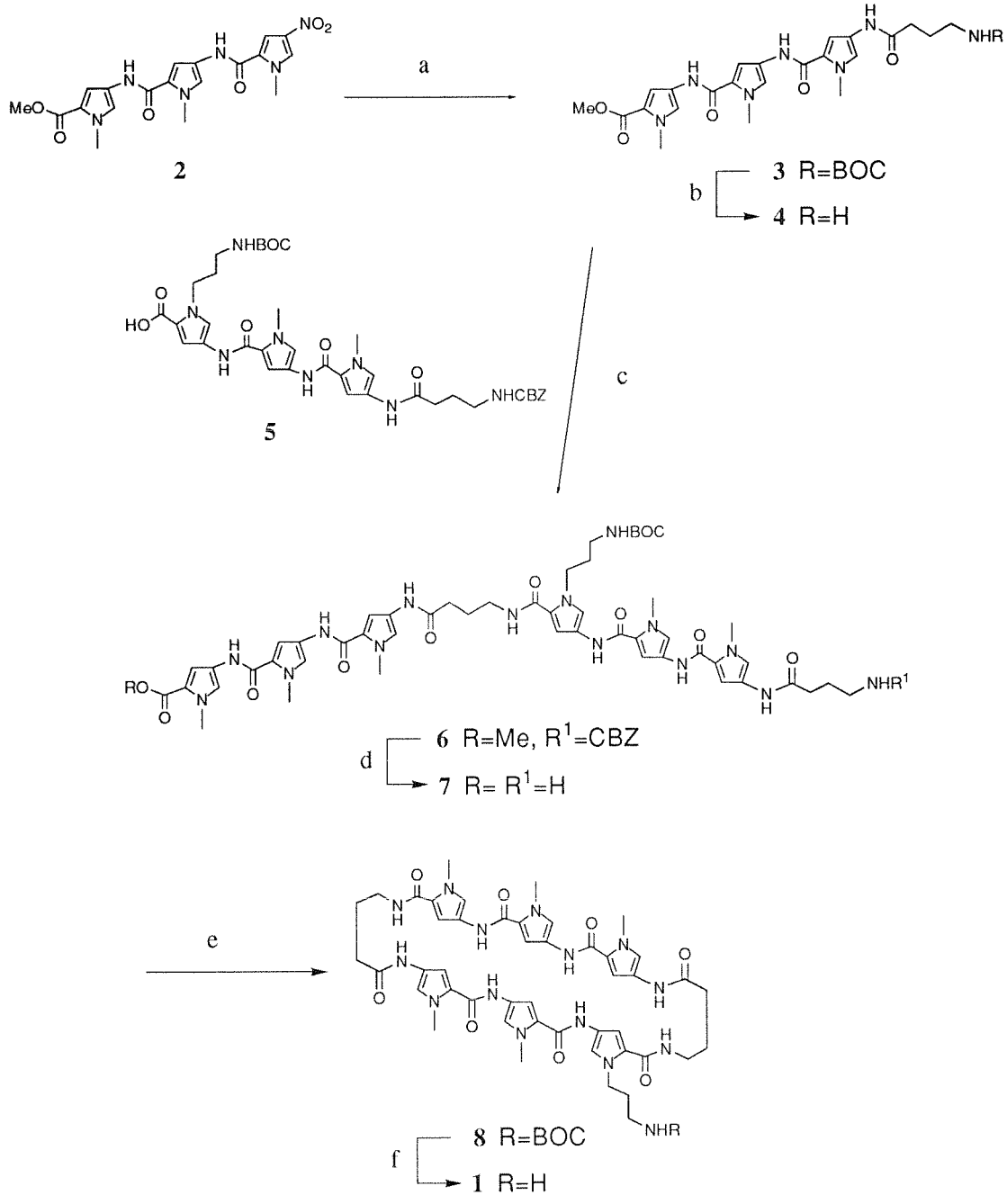



**Figure 3.3.** Synthetic scheme of the cyclic polyamide **1**. (a) (i) 250 psi $H_2$, Pd/C; (ii) $BOCNH(CH_2)_3CO_2H$, DCC, HOBt; (b)$TFA/CH_2Cl_2$; (c) 5, DCC, HOBt; (d) (i) NaOH, $MeOH\text{-}H_2O$; (ii) 250 psi $H_2$, Pd/C; (e) DPPA, $NaHCO_3$, DMF; (f) TFA, thiophenol;

azide (DPPA) in the presence of $NaHCO_3$ to give **8** in 43% yield. Removal of the tert-butoxycarbonyl protecting group (TFA/PhSH) provided the *cyclo*-(Py-Py-Py-γ-Py-Py-Py-γ) **1** which was purified by reverse-phase HPLC. The observed molecular mass of the cyclic polyamide **1** is 946.4436 (FABMS) in good agreement with the calculated value [M+H, 946.4436 calcd. for $C_{46}H_{56}N_{15}O_8$].

**Quantitative DNase I footprint titration analysis**

In order to compare the binding affinity of the cyclic polyamide **1** and distamycin A for three 5′-$(A,T)_5$-3′ match sites, 5′-TATAT-3′, 5′-AATTT-3′, and 5′-TTTTT-3′, quantitative DNase I footprint titration experiments[25-27] were performed on two $^{32}$P-labeled restriction fragments - the 188 base pair 3′ end labeled *Eco*R V/*Bam*H I DNA from pBR322 and the 246 base pair 5′ end labeled *Bam*H I/*Pvu* II DNA from pMM5. Distamycin binds the three sites in order of decreasing affinity 5′-TTTTT-3′ > 5′-AATTT-3′ > 5′-TATAT-3′. In contrast, the cyclic polyamide binds these sites in the reverse order, 5′-TATAT-3′ > 5′-AATTT-3′ > 5′-TTTTT-3′ (Table I). The cyclic polyamide binds to the 5′-TATAT-3′ with 20-fold higher affinity than does distamycin, while both compounds bind to the 5′-TTTTT-3′ with similar affinity. A comparison of the binding affinities indicates that quite different sequence-selectivity is observed for the two compounds. Examination of the binding curve data shows that the slope of the isotherm for distamycin binding the 5′-TATAT-3′ site is steeper than expected for a 1:1 complexes with DNA. The data points are best fit by a cooperative binding isotherm consistent with intermolecular dimeric binding. This finding is consistent with recent NMR studies from Wemmer group that demonstrated a strong positive cooperative interaction between two molecules of distamycin and the ATATAT sequence[28].

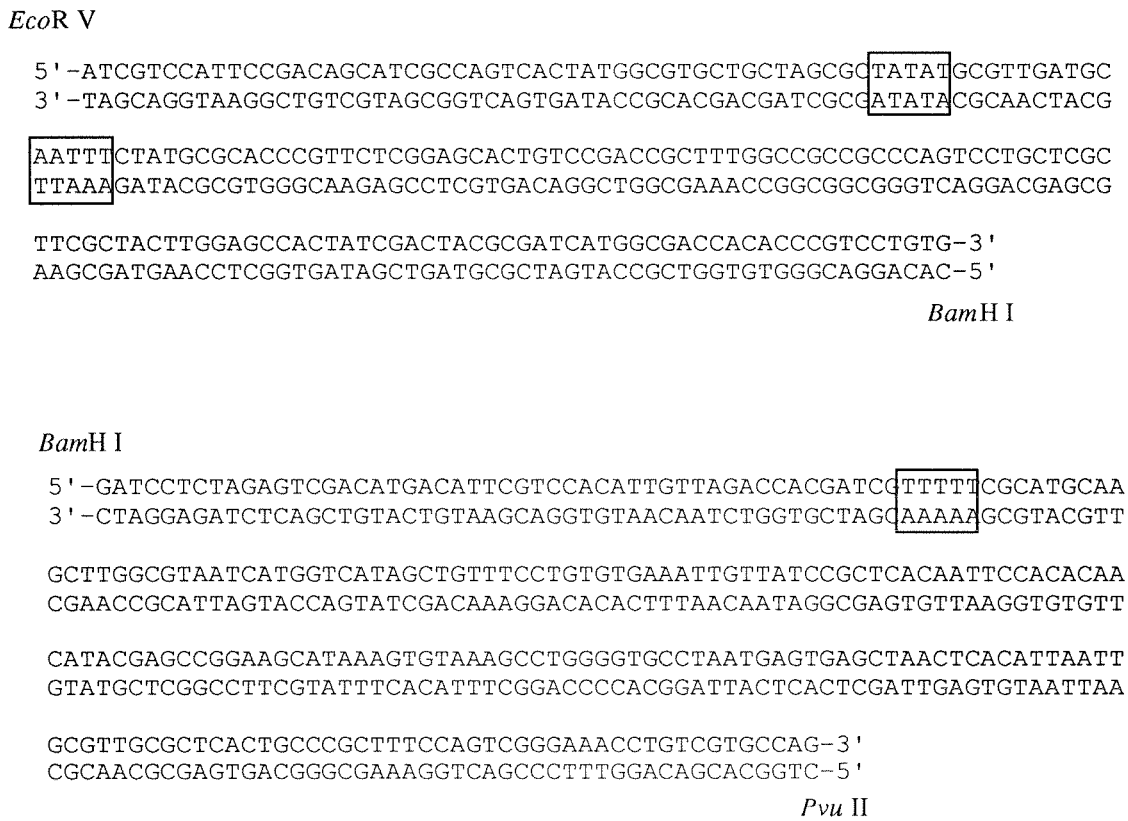



**Figure 3.4.** Sequences of the 188 base pair *Eco*R V/*Bam*H I resriction fragment and the 246 base pair *Bam*H I/*Pvu* II resriction fragment. Three binding sites were analyzed by quantitative footprint titration analysis.

**Figure 3.5a.** Quantitative DNase I footprint titration experiment with Distamycin A on the 3′-$^{32}$P-labeled 188 base pair *EcoR*V/*Bam*H I resriction fragment from plasmid pBR322: lane 1, A reaction; lane 2, G reaction; lane 3, intact DNA; lane 4, DNase I standard; lanes 5-20, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, 100 nM, 200 nM, 500 nM, 1 μM Distamycin A, respectively; The 5′-AATTT-3′ and 5′-TATAT-3′ binding sites which were analyzed are shown on the right side of the autoradiogram. All reactions contain 10 kcpm restriction fragment, 10 m M Tris-HCl (pH 7.0), 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$.

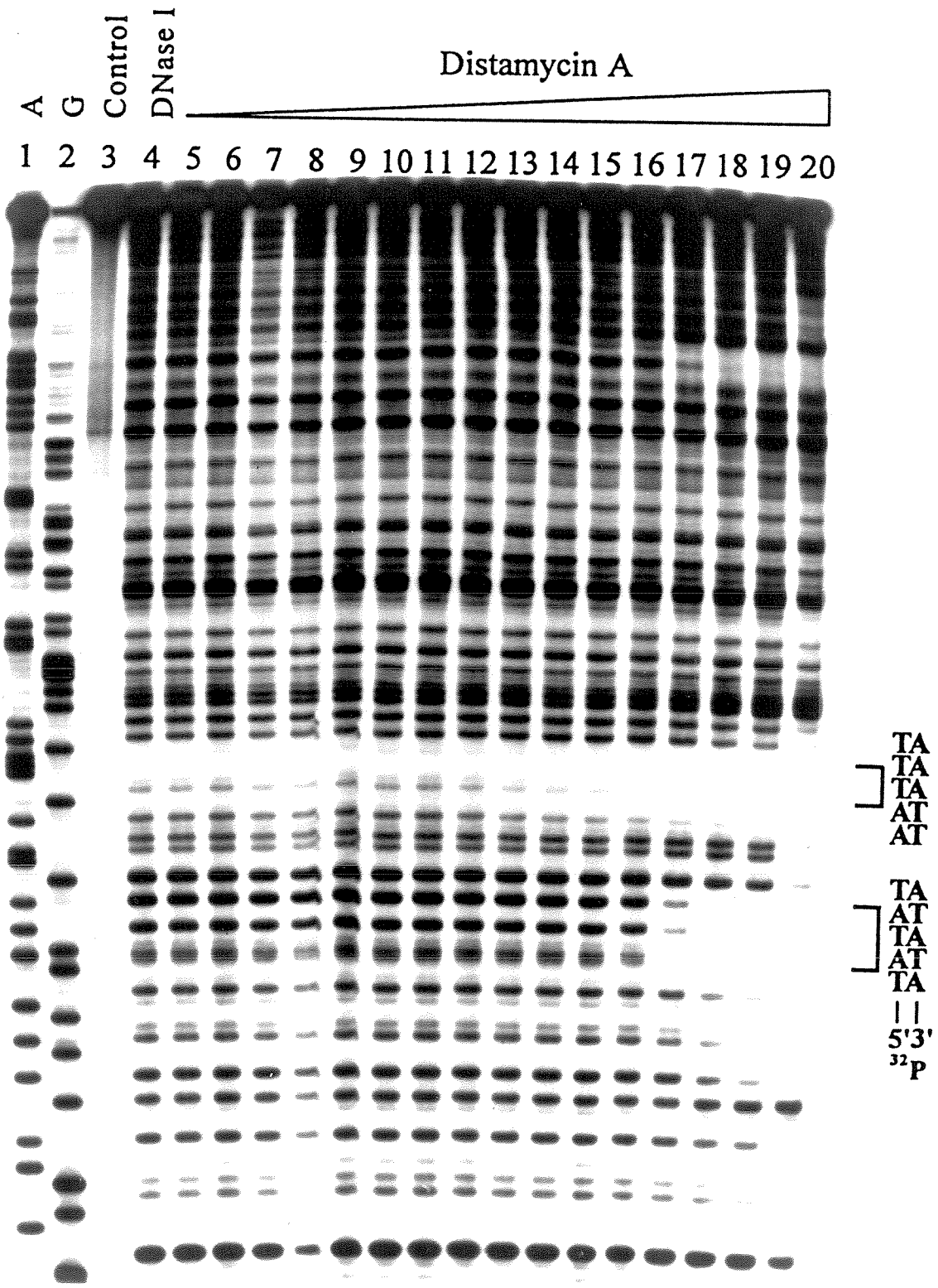

A
G
Control
DNase 1
Distamycin A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
TA
TA
TA
AT
AT
TA
AT
TA
AT
TA
5'3'
32P

**Figure 3.5b.** Quantitative DNase I footprint titration experiment with the cyclic polyamide **1** on the 3′-$^{32}$P-labeled 188 base pair *Eco*RV/*Bam*H I resriction fragment from plasmid pBR322: lane 1, A reaction; lane 2, G reaction; lane 3, intact DNA; lane 4, DNase I standard; lanes 5-16, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, 100 nM, 200 nM cyclic polyamide, respectively; The 5′-AATTT-3′ and 5′-TATAT-3′ binding sites which were analyzed are shown on the right side of the autoradiogram. All reactions contain 10 kcpm restriction fragment, 10 mM Tris-HCl (pH 7.0), 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$.

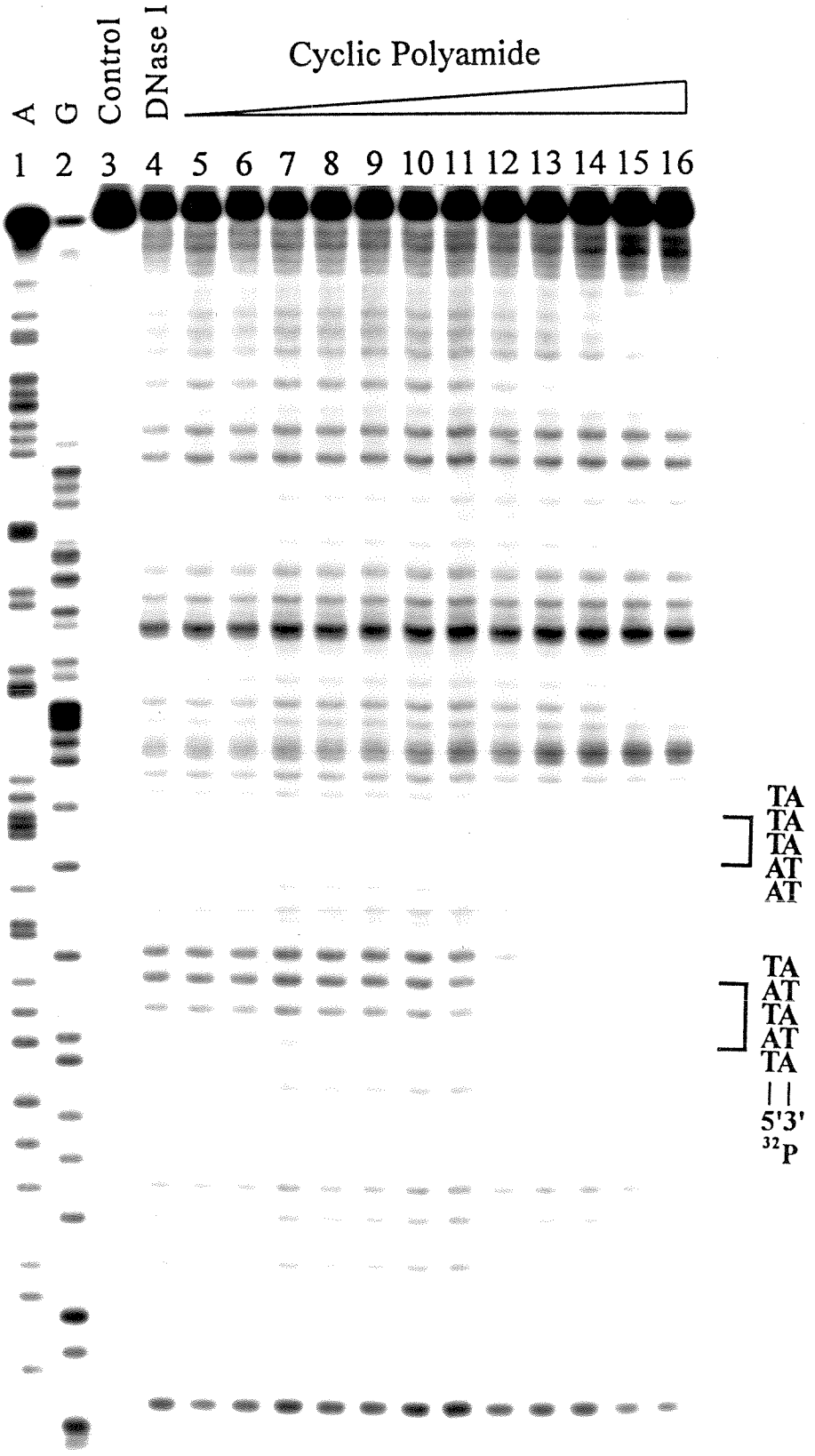

A
G
Control
DNase I
Cyclic Polyamide
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
TA
TA
TA
AT
AT
TA
AT
TA
AT
TA
5'
3'
32P

**Figure 3.5c.** Quantitative DNase I footprint titration experiment with Distamycin A on the 5′-$^{32}$P-labeled 246 base pair *Bam*H I/*Pvu* II resriction fragment from plasmid pMM5: lane 1, A reaction; lane 2, G reaction; lane 3, T reaction; lane 4, intact DNA; lane 5, DNase I standard; lanes 6-19, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, 100 nM, 200 nM, 500 nM, 1 μM Distamycin A, respectively; The 5′-TTTTT-3′ binding site which was analyzed is shown on the right side of the autoradiogram. All reactions contain 10 kcpm restriction fragment, 10 mM Tris-HCl (pH 7.0), 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$.

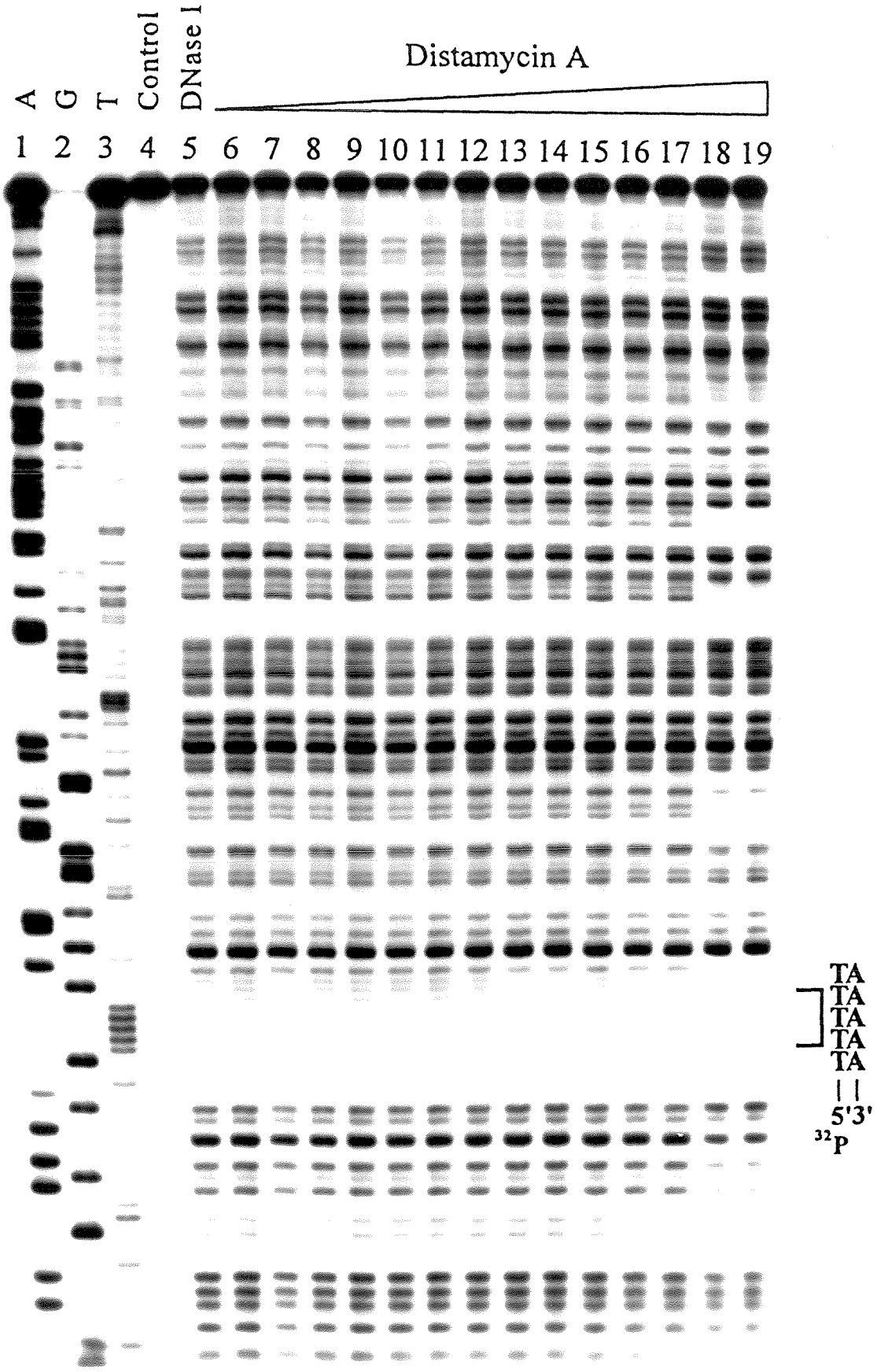

A
G
T
Control
DNase I
Distamycin A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
TA
TA
TA
TA
TA
5'3'
32P

**Figure 3.5d.** Quantitative DNase I footprint titration experiment with the cyclic polyamide **1** on the 5′-$^{32}$P-labeled 246 base pair *BamHI/PvuII* resriction fragment from plasmid pMM5: lane 1, A reaction; lane 2, G reaction; lane 3, T reaction; lane 4, intact DNA; lane 5, DNase I standard; lanes 6-17, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, 100 nM, 200 nM cyclic polyamide, respectively; The 5′-TTTTT-3′ and 5′-AGATTGTT-3′ binding sites which were analyzed are shown on the right side of the autoradiogram. All reactions contain 10 kcpm restriction fragment, 10 m M Tris-HCl (pH 7.0), 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$.

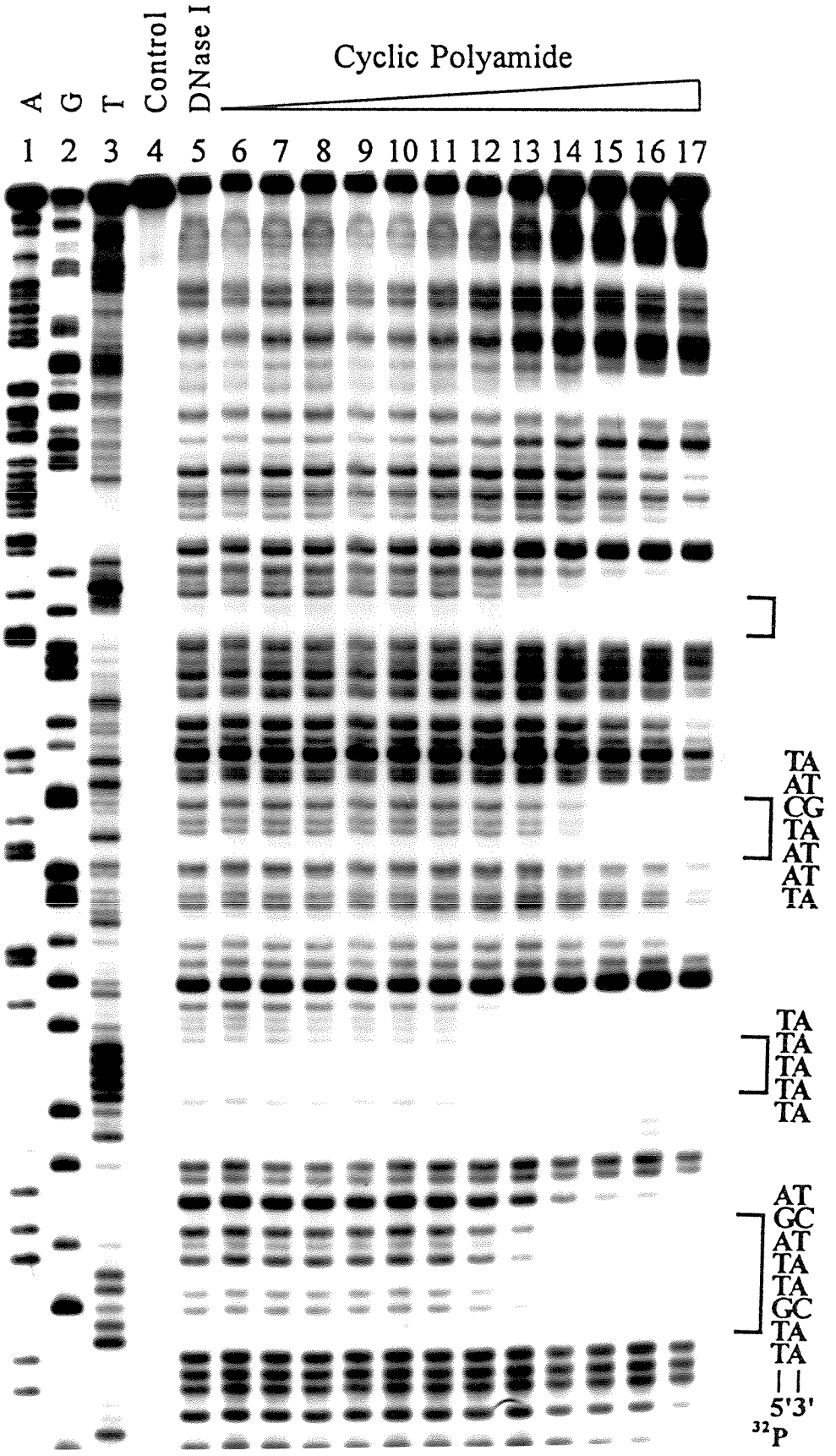

A
G
T
Control
DNase I
Cyclic Polyamide
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
TA
AT
CG
TA
AT
AT
TA
TA
TA
TA
TA
TA
AT
GC
AT
TA
TA
GC
TA
TA
5'3'
32P

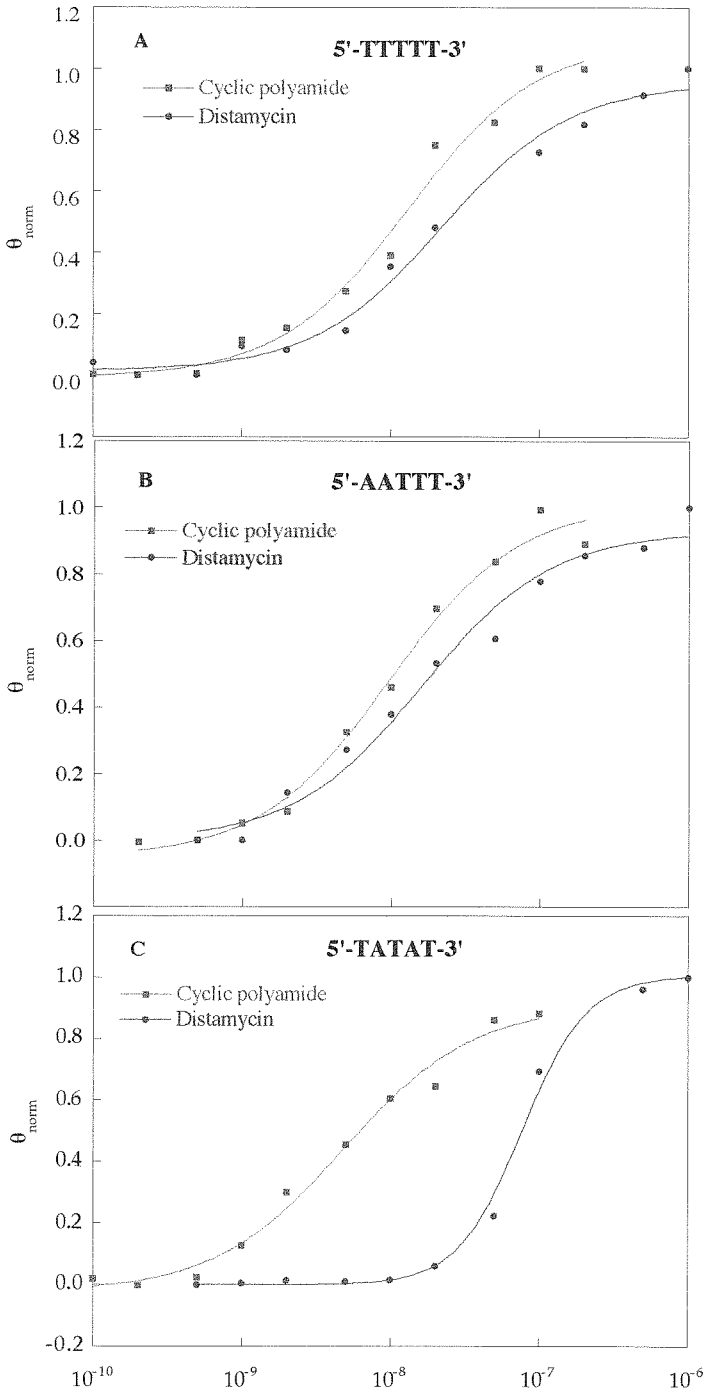
**Table I** Apparent First-Order Binding Constants $(M^{-1})^{ab}$

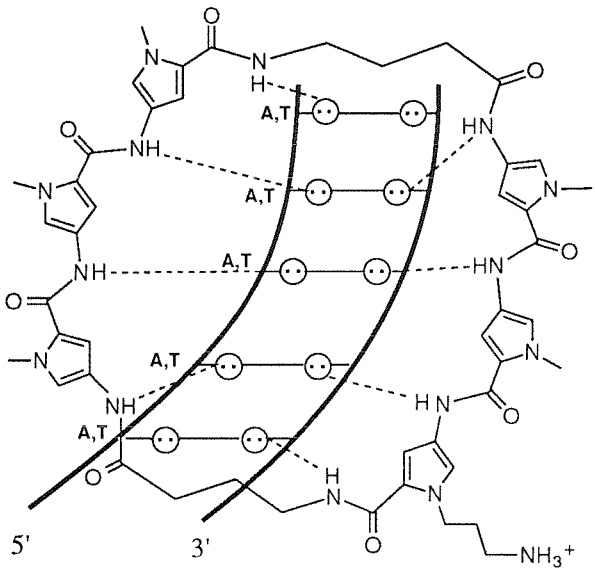
| | binding site | | |
|---|---|---|---|
| Polyamide | 5'-TTTTT-3' | 5'-AATTT-3' | 5'-TATAT-3' |
| Distamycin A | $5.3 \times 10^7$(1.8) | $4.5 \times 10^7$(0.6) | $1.3 \times 10^7$(0.1) |
| Cyclic polyamide | $7.1 \times 10^7$(0.5) | $1.0 \times 10^8$(1.3) | $2.0 \times 10^8$(0.3) |

[a]Values reported are the mean values measured from three footprint titration experiments. [b]The assays were performed at 22 °C at pH7.0 in the presence of 10 mM Tris-HCl, 10mM KCl, 10 mM $MgCl_2$, and 5mM $CaCl_2$.

A large number of crystallographic studies have been performed with different AT-tract sequences[29]. While the minor groove of AAAAA and AAATTT regions have been shown to be narrow compared with ideal B-form DNA, two different crystal structures have been reported for alternating AT sequences[9, 10, 30, 31]. Although it is unclear whether the minor groove width of AT-tracts is dependent on the sequence, our results suggest that the 5′-TATAT-3′ site may adopt a wider minor groove than do the 5′-AATTT-3′ and 5′-TTTTT-3′ sites. The cyclic polyamide binds more strongly to the 5′-TATAT-3′ site over the 5′-TTTTT-3′ site by a factor of three. This difference in affinity is less than or equal to that observed for distamycin. A possible explanation for this is that the minor groove of the 5′-TATAT-3′ sequence , while wider than that of the 5′-TTTTT-3′, is not as wide as that of the ideal B-form DNA. The energetic penalty of widening the minor groove may reduce the binding affinity and specificity. This explanation is consistent with the minor groove widening observed in the complex of distamycin and ATATAT sequence[28]. Finally, the cyclic polyamide also binds to the AT tracts

**Figure 3.6.** Data for quantitative DNase I footprint titrations for distamycin *cyclo*-(Py-Py-Py-γ-Py-Py-Py-γ) **1** binding to the 5′-TTTTT-3′, 5'-AATTT-3', and 5'-TATAT-3' sites. The $\theta_{norm}$ points were obtained using photostimulable storage phosphor autoradiography. The data points for distamycin and *cyclo*-(Py-Py-Py-γ-Py-Py-Py-γ) **1** are represented by red filled circles and blue filled squares, respectively. The curve for distamycin binding to the 5'-TATAT-3' is a best fit cooperative binding isotherm (n = 2) consistent with 2:1 polyamide-DNA complex formation with distamycin. The other curves are best fit Langmuir binding isotherms (n =1) consistent with 1:1 polyamide-DNA complex formation.


A
5'-TTTTT-3'
Cyclic polyamide
Distamycin
B
5'-AATTT-3'
Cyclic polyamide
Distamycin
C
5'-TATAT-3'
Cyclic polyamide
Distamycin
θ norm
1.2
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
10-10
10-9
10-8
10-7
10-6




*Cyclo*-(Py-Py-Py-γ-Py-Py-Py-γ) • 5'-(A,T)$_5$-3'

**Figure 3.7.** Proposed binding model for the complex between the cyclic polyamide *cyclo*-(Py-Py-Py-γ-Py-Py-Py-γ) **1** and a 5′-(A,T)5-3′ sequence. Circles with dots represent lone pairs of N3 of adenines and O2 of thymines. Putative hydrogen are illustrated by dashed lines.

containing one single base pair mismatch with an equilibrium association constant of $3.5 \times 10^7$ $M^{-1}$. Presumably, the cyclic polyamide may favor the wide minor groove despite the energetic price of the hydrogen bond mismatch. In conclusion, although efficient discrimination between different AT-tract sequences has not been achieved, the synthetic cyclic polyamide does show different sequence-selectivity compared with distamycin.

## Experimental Section

### General

$^1$H NMR spectra were recorded at 300 MHz on a GE 300 NMR in $CDCl_3$, DMSO-$d_6$, or $CD_3OD$ Chemical shifts are reported in parts per million relative to tetramethylsilane or residual DMSO-$d_5$. IR spectra were recorded on a Perkin-Elmer FTIR spectrometer. High-resolution mass spectra (HRMS) were recorded using fast atom bombardment (FAB) techniques at the Mass Spectrometry Laboratory at the university of California, Riverside. Preparatory HPLC was carried out on a Beckman Instrument using a Water DeltaPak 25 x 100 mm 100 mm C18 column. Analytical HPLC was performed on a Hewlett Packard 1090 Series II analytical HPLC using a Vydac C18 reverse phase column (0.46 x 25 cm, 5 mm, HS silica). Flash column chromatography was carried out using silica gel 60 (230-400 mesh, Merck). Thin-layer chromatography (TLC) was performed on silica gel 60 $F_{254}$ precoated plates (Merck) All chemicals for the synthesis were purchased from aldrich unless otherwise specified. Dichloromethane and N,N-dimethylformamide (DMF) were purchased as anhydrous solvents from Aldrich. Boc-GABA and CbZ-GABA were purchased from Sigma. Distamycin A was purchased from Sigma and used without further purification. Drug concentrations were

calculated by using an extinction coefficient of 34,000 $M^{-1}cm^{-1}$ at 303 nm. All enzymes were purchased from Boehringer Mannheim and used with the buffers supplied, with the exception of BamH I and Pvu II which were obtained from New England Biolabs. Adenosine 5'-[$\gamma$-$^{32}$P] triphosphate, deoxyadenosine 5'-[$\alpha$-$^{32}$P] triphosphate, and thymidine 5'-[$\alpha$-$^{32}$P] triphosphate were obtained from Amersham. Storage phosphotechnology autoradiography was performed using a Molecular Dynamics 400S Phosphorimager and ImageQuant software.

**2-Bis(N-methylpyrrolecarboxamide)-1-methyl-4-[4-(N-tert-butoxycarbonyl) amino-butylcarboxylamino]-imidazole 3** To a solution of 4-(N-tert-butyloxycarbonyl) amino-butyric acid (1.5 g, 7.5 mmol) and N-hydroxybenzotriazole hydrate (1.05 g, 7.8 mmol) in DMF (15 mL) was added a solution of DCC (1.6 g, 7.8 mmol) in DMF (5 mL) and the reaction mixture was allowed to stir for 2 h. Separately, to a solution of **2** (1.7 g, 3.96 mmol) in DMF (15 mL) was added 10% Pd/C (150 mg) and the mixture was hydrogenated in a Parr bomb apparatus (250 psi, $H_2$) for 3 h. The catalyst was removed by filtration through Celite and the filtrate was immediately added to the activated acid. The mixture was allowed to stir for 4 h and filtered through Celite. The solvent was removed in vacuo and the residue was purified by flash column chromatography (7% MeOH in $CH_2Cl_2$) to afford **3** (2.2 g, 95%). $^1$H NMR (DMSO-$d_6$) $\delta$ 9.94 (s, 1H), 9.92 (s, 1H), 9.80 (s, 1H), 7.47 (s, 1H), 7.23 (s, 1H), 7.15 (s, 1H), 7.05 (s, 1H), 6.89 (s, 1H), 6.86 (s, 1H), 3.83 (s, 9H), 3.73 (s, 3H), 2.93 (t, 2H, J=7.2Hz), 2.21 (t, 2H, J=7.2Hz), 1.65 (m, 2H), 1.37 (s, 9H); FABMS m/e 583.2744 (M+H, 583.2754 calcd. for $C_{28}H_{37}N_7O_7$).

**2-Bis(N-methylpyrrolecarboxamide)-1-methyl-4-(4-amino-butylcarboxylamino)-pyrrole 4** To a solution of 2-Bis(N-methylpyrrolecarboxamide)-1-methyl-4-[4-(N-tert-butoxycarbonyl)amino-

butylcarboxylamino]-imidazole **3** (2.1 g, 3.6 mmol) in $CH_2Cl_2$ (20 mL) was added TFA (10 mL) and the resulting solution was allowed to stir for 4 h. The mixture was evaporated in vacuo and the residue was purified by flash column chromatography (10% MeOH in $CH_2Cl_2$) to afford **4** (1.57g, 90%). $^1H$ NMR (DMSO-$d_6$) δ 9.94 (s, 1H), 9.92 (s, 1H), 9.85 (s, 1H), 7.46 (s, 1H), 7.23 (s, 1H), 7.16 (s, 1H), 7.05 (s, 1H ), 6.89 (s, 1H), 6.86 (s, 1H), 3.83 (s, 9H), 3.72 (s, 3H), 2.29 (m, 2H), 2.26 (t, 2H, J=7.2Hz), 1.64 (m, 2H); IR (KBr) 3451 (s), 1685 (s), 1560 (s), 1438 (m), 1406 (m), 1262 (m), 1206 (s), 1137 (m), 1061 (w), 802 (w), 724 (w) $cm^{-1}$; FABMS m/e 484.2338 (M+H, 484.2308 calcd. for $C_{23}H_{30}N_7O_5$).

**Cbz-γ-Py-Py-Py-γ-Py-Py-Py-methyl ester 6** To a solution of the acid **5** (1.15g, 1.54 mmol) and N-hydroxybenzotriazole hydrate (219 mg, 1.62 mmol) in DMF (15 mL) was added a solution of DCC (334 mg, 1.62 mmole) in DMF (2 mL) and the reaction mixture was allowed to stir for 2 h. A solution of the primary amine **4** (1.1 g, 2.31 mmol) in DMF (10 mL) was added and the mixture was allowed to stir for 4 h. The solvent was removed in vacuo and the residue was purified by flash column chromatography (10% MeOH in $CH_2Cl_2$) to afford **6** (1.4 g, 75%). $^1H$ NMR (DMSO-$d_6$) δ 9.93 (s, 3H), 9.87 (s, 2H), 9.82 (s, 2H), 7.47 (s, 1H), 7.34 (s, 5H), 7.31 (s, 2H), 7.27 (s, 2H), 7.24 (s, 2H), 7.19 (s, 1H), 7.16 (s, 1H), 6.90 (s, 1H), 6.88 (s, 2H), 5.0 (s, 2H), 4.26 (brs, 2H), 3.83 (s, 15H), 3.73 (s, 3H), 3.22 (m, 2H), 3.02 (m, 2H), 2.87 (m, 2H), 2.23 (m, 4H), 1.73 (m, 6H), 1.37 (s, 9H); IR (KBr) 3305 (m), 2946 (m), 1658 (s), 1538 (s), 1254 (s), 1107 (m), 1060 (w), 890 (w), 777 (w) $cm^{-1}$; FABMS m/e 1211.5522 (M+H, 1211.5512 calcd. for $C_{60}H_{73}N_{15}O_{13}$).

**γ-Py-Py-Py-γ-Py-Py-Py 7** Cbz-γ-Py-Py-Py-γ-Py-Py-Py-methyl ester **6** (420 mg, 0.35 mmol) was dissolved in MeOH (10 mL) and 1N NaOH (2.1 mL, 2.1 mmol, 6eq.) was added. The reaction mixture was heated to 60 °C and allowed to stir for 3 h. The mixture was acidified to pH2 with 1N HCl. The

precipitate was collected by filtration and dried to afford the acid. Without further purification, the acid was dissolved in DMF (5 mL) and 10% Pd/C (50 mg) was added. The mixture was hydrogenated in a Parr bomb apparatus (250 psi, $H_2$) for 6 h and filtered with Celite. The filtrate was concentrated in vacuo to afford **7** (305 mg, 83%) $^1$H NMR (DMSO-$d_6$) δ 10.32 (s, 1H), 10.13 (s, 1H), 9.97 (s, 2H), 9.95 (s, 2H), 9.93 (s, 1H), 7.46 (s, 1H), 7.33 (s, 1H), 7.26 (s, 2H), 7.24 (s, 1H), 7.18 (s, 1H), 7.14 (s, 1H), 7.04 (s, 1H), 6.89 (s, 1H), 6.88 (s, 2H), 6.77 (s, 1H), 4.24 (brs, 2H), 3.94 (s, 3H), 3.83 (s, 6H), 3.82 (s, 6H), 3.17 (m, 2H), 2.81 (m, 4H), 2.36 (m, 4H), 1.78 (m, 6H), 1.35 (s, 9H); IR (KBr) 3402 (s), 1654 (s), 1543 (s), 1466 (m), 1438 (s),1406 (m), 1259 (w), 1209 (w), 1108 (w) cm$^{-1}$; FABMS m/e 1063.4892 (M+H, 1063.4838 calcd. for $C_{51}H_{65}N_{15}O_{11}$).

**Boc-protected Cyclic polyamide 8** To a solution of γ-Py-Py-Py-γ-Py-Py-Py **7** (600 mg, 0.56 mmol) in DMF (200 mL) was sequentially added $NaHCO_3$ (260 mg, 3.1 mmol) and diphenylphosphoryl azide (DPPA, 465 mg, 1.69 mmol). The mixture was allowed to stir at room temperature for 3 days and then filtered. The filtrate was concentrated in vacuo and the residue was purified by flash column chromatography (15% MeOH in $CH_2Cl_2$) to afford the cyclic polyamide **8** (90 mg, 43%) $^1$H NMR [$CDCl_3$-$CD_3OD$ (1:1)] δ 7.30 (s, 1H), 7.26 (s, 1H), 7.18 (s, 1H), 7.02 (s, 2H), 6.88 (s, 1H), 6.70 (s, 1H), 6.68 (s, 1H), 6.62 (s, 1H), 6.58 (s, 1H), 6.43 (s, 1H), 4.35 (m, 2H), 3.92 (s, 6H), 3.90 (s, 3H), 3.89 (s, 3H), 3.82 (s, 3H), 3.49 (m, 4H), 3.08 (m, 2H), 2.42 (m, 4H), 1.93-2.15 (m, 6H), 1.45 (s, 9H); IR (KBr) 3422 (s), 1654 (s), 1545 (s), 1438 (s), 1406 (m), 1260 (m), 1208 (m), 1105 (w), 919 (w), 777 (w) cm$^{-1}$; FABMS m/e 1045.4835 (M+H, 1045.4882 calcd. for $C_{51}H_{63}N_{15}O_{10}$).

**Cyclic polyamide 1** To a mixture of TFA and thiophenol (4 mL, 0.5 M PhSH in TFA) was added the Boc-protected Cyclic polyamide **8** (30 mg, 28.6 mmol) and the reaction mixture was allowed to stir at room temperature for

2 h. Excess TFA was removed in vacuo and the residue was purified by flash column chromatography (5% $NH_4OH$ in MeOH) to afford **1** (23 mg, 90%). The product was further purified by using reverse-phase HPLC on a preparatory Water DeltaPak 25 x 100 mm 100 mm C18 column with linear gradients of 60% acetonitrile plus 0.1% TFA versus 0.1% aqueous TFA. $^1H$ NMR (DMSO-$d_6$) δ 9.97 (s, 1H), 9.89 (s, 3H), 9.77 (s, 2H), 7.64 (s, 1H), 7.49 (s, 1H), 7.33 (s, 2H), 7.04 (s, 2H), 6.97 (s, 2H), 6.83 (s, 1H), 6.81 (s, 1H), 6.64 (s, 1H), 6.60 (s, 1H), 4.26 (m, 2H), 3.78 (s, 15H), 3.73 (s, 3H), 3.16 (m, 4H), 2.66 (m, 2H), 2.39 (m, 2H), 2.25 (m, 2H), 1.91 (m, 2H), 1.70 (m, 4H); IR 3383 (s), 1637 (s), 1560 (s), 1438 (s), 1407 (s), 1265 (m), 1207 (w), 1122 (w), 1062 (w), 777 (w) $cm^{-1}$; UV ($H_2O$) lmax (e) 240 (44,462 $cm^{-1}M^{-1}$), 310 (51,024 $cm^{-1}M^{-1}$) nm; FABMS m/e 946.4436 (M+H, 946.4436 calcd. for $C_{46}H_{56}N_{15}O_8$).

**Preparation of labeled DNA**

The 188 base pair 3′ end labeled *Eco*R V/*Bam*H I restriction fragment from plasmid pBR322 was prepared and purified as follows. Plasmid DNA was linearized with EcolV, followed by treatment by Klenow fragment of DNA polymerase I in the presence of cold dATP and TTP. The 3′ end labeling was performed with [α-$^{32}$P]dATP, [α-$^{32}$P]TTP, and Sequenase™ (version 2.0). The labeled DNA was digested with BamH I and the resulting 188 base pair restriction fragment was isolated by nondenaturing 6% polyacrylamide gel electrophoresis (PAGE). The gel bands were visualized by autoradiography, isolated, filtered to remove the polyacrylamide. The resulting solution was further purified by phenol extraction followed by ethanol precipitation. The DNA pellet was washed with 70% ethanol, dried in vacuo, and resuspended in TE buffer (10 mM Tris-HCl, 0.1 mM EDTA, pH 7.0) and stored at -78°C.

The 246 base pair 5′ end labeled *Bam*H I/*Pvu* II restriction fragment was prepared from plasmid pMM5[17]. Plasmid DNA was linearized with *Bam*H I,

treated with alkaline phosphatase, and labeled with [$\gamma$-$^{32}$P]dATP and T4 polynucleotide kinase. The secondary cleavage was done with *Pvu* II and the labeled DNA was purified in the same procedure as the 188 base pair 3′ end labeled *Eco*R V/*Bam*H I restriction fragment. Chemical sequencing reactions were performed according to published protocol[18-20]. Standard protocols were used for all DNA manipulations[21].

**Quantitative DNase I Footprint Titration**

All reactions were carried out in a total volume of 40 μL with final concentrations of each species as indicated. The polyamide ligands, ranging in concentration from 20 pM to 1 μM, were added to solutions of radiolabeled restriction fragment (20,000), Tris-HCl (10 mM, pH7.0), KCl (10 mM), $Mg_2Cl$ (10 mM) and $CaCl_2$ (5 mM) and incubated for 12 h at 22 $^{o}$C. We explicitly note that no carrier calf thymus DNA is used in these reactions. Footprinting reactions were initiated by the addition of a stock solution of DNase I (0.05 units/mL) containing 1 mM DTT and allowed to ploceed for 5 min at 22$^{o}$C. The reactions were stopped by addition of a 3 M sodium acetate solution containing 25 mM EDTA and ethanol precipitated. The reactions were resuspended in 100 mM tris-borate-EDTA/80% formamide loading buffer and electrophoresed on 8% polyacrylamide denaturing gels (5% crosslink, 7M urea) at 2000 V for 1 h. The footprint titration gels were dried and quantitated using phosphor technology.

Apparent first-order binding constants were determined as previously described. The data were analyzed by performing volume integrations of the target sites and 5′-GCGT-3′ and 5′-CCAC-3′ reference sites for experiments with the restriction fragment labeled at the *Eco*R V and *Bam*H I sites, respectively. The apparent DNA target site saturation, $\theta_{app}$, 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 $I_{tot^{\circ}}$ and $I_{ref^{\circ}}$ correspond to those values from a DNase I control lane to which no polyamide had been added. At concentrations higher than 2 μM distamycin and 0.5 μM *cyclo*-(Py-Py-Py-γ-Py-Py-Py-γ) the reference sites become partially protected, resulting in low $\theta_{app}$ values. For this reason, peptide concentrations of higher than 2 μM distamycin and 0.5 μM *cyclo*-(Py-Py-Py-γ-Py-Py-Py-γ) were not used. The ($[L]_{tot}$, $\theta_{app}$) data points were fit to Langmuir binding isotherms by minimizing the difference between $\theta_{app}$ and $\theta_{fit}$ using the modified Hill equation:

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

where $[L]_{tot}$ corresponds to the total polyamide concentration, $K_a$ corresponds to the apparent monomeric association constant, n represents the Hill coefficient, and $\theta_{max}$ and $\theta_{min}$ 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 $K_a$, $\theta_{max}$, and $\theta_{min}$ 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:

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

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


1. Zimmer, C.; Wahnert, U. *Prog. Biophys. Mol. Biol.* **1986**, 47, 31-112.
2. Van Dyke, M. W.; Hertzberg, R. P.; Dervan, P. B. *Proc. Natl. Acad. Sci.U.S.A.* **1982**, 79, 5470-5474.
3. Van Dyke, M. W.; Dervan, P. B. *Nucleic Acids Res.* **1983**, 16, 5555-5567.
4. Fox, K. R.; Waring, M. J. *Nucleic Acids Res.* **1984**, 17, 9271-9285.
5. Dervan, P.B. *Science* **1986**, 232, 464-47.
6. Portugal, J.; Waring, M. J. *Eur. J. Biochem.* **1987**, 167, 281-289.
7. Klevit, R. E.; Wemmer, D. E.; Reid, B. R. *Biochemistry* **1986**, 52, 3296-3303.
8. Pelton, J. G.; Wemmer, D. E. *Biochemistry* **1988**, 27, 8088-8096.
9. Coll, M.; Frederick, C. A.; Wang, A. H. -J.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, 84, 8385-8389.
10. Yoon, C.; Prive, G. G.; Goodsell, D. S.; Dickerson, R. E. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, 85, 6332-6336.
11. Wahnert U.; Zimmer, C.; Luck, G.; Pitra, C. *Nucleic Acids Res.* **1975**, 2, 391-404.
12. Zimmer, C.; Luck, G.; Fric, I. *Nucleic Acids Res.***1976**, 3, 1521-1532.
13. Zimmer, C.; Marck, C.; Schneider, C.; Guschlbauer, W. *Nucleic Acids Res.* **1979**, 6, 2831-2837.
14. Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci. U.S.A.***1989**, 86, 5723-5727.
15. Pelton, J. G.; Wemmer, D. E. *J. Am. Chem. Soc.*. **1990**, 112, 1393-1399.
16. Cho, J.; Parks, M. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, 92, 10389-10392.
17. Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, 116, 3663-3664.
18. Iverson, B. L.; Dervan, P. B. *Nucleic Acids Res.* **1987**, 15, 7823-7830
19. Maxam, A. M.; Gilbert, W. S. *Methods in Enzymology* **1980**, 65, 499-560

20. Rubin, C. M,; Schmid, C. W.*Nucleic Acids Res.* **1980**, 8, 4613-4619

21. Sambrook, J.; Fritsch, E. F.; Maniatis, T. Molecular Cloning, Cold Spring Harbor Laboratory: Cold Spring Harbor, NY.

22. Mrksich, M.; Parks, M. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, 116, 7983-7988

23. Bialer, M.; Yagen, B.; Mechoulam, R. *Tetrahedron* **1978**, 34, 2389-2391

24. Lown, J. W.; Krowicki, K. *J. Org. Chem.* **1985**, 50, 3774-3779

25. Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Methods in Enzymology* **1986**, 130, 132-181

26. Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G.K. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, 83, 8462-8466

27. Senear, D. F.; Brenowitz, M.; Shea, M. A.; Ackers, G. K. *Biochemistry* **1986**, 25, 7344-7354

28. Fagan, P.; Wemmer, D. E. *J. Am. Chem. Soc.* **1992**, 114, 1080-1081

29. Nelson, H. C. M.; Finch, J. T.; Luisi, B. F.; Klug, A. *Nature* **1987**, 330, 221-226

30. Yuan, H.; Quintana, J.; Dickerson, R. E. *Biochemistry* **1992**, 31, 8009-8021

# Chapter 4

## Studies Towards Conformational Analysis of HIV-1 TAR RNA by EDTA•Fe(II) Autocleavage

### Introduction

RNAs are known to carry out diverse biological functions such as the storage of genetic information, the transfer of this information into protein synthesis, and the catalysis of biochemical reactions. Many RNA molecules require specifically folded structures for their biological activity. However, our understanding of the three-dimensional structure of RNA is limited. To date, the tertiary structures for two RNAs, tRNA and the hammerhead ribozyme, have been determined by x-ray crystallography[1,2]. NMR studies have also provided information about the three-dimensional structures of RNAs, but have been limited to RNAs of molecular mass 10,000 daltons or less[3]. Other methods such as site-specific mutagenesis, chemical and enzymatic probing, and phylogenetic analysis have proven to be useful in the determination of RNA secondary structure.

The EDTA•Fe(II) complex is known to generate free diffusible hydroxyl radicals which can cleave DNA or RNA by oxidation of the sugar backbone. The cleavage reaction can be initiated by addition of a reducing agent such as dithiothreitol (DTT) or sodium ascorbate. The oxidative cleavage of RNA by hydroxyl radicals is an attractive tool in the study of RNA tertiary structure. EDTA•Fe(II) has been used successfully to determine regions of an RNA molecule that are less solvent-accessible than others[4]. Attachment of EDTA•Fe(II) to small molecules, proteins, and oligonucleotides has proven to be useful for studying the structures of such molecules in complex with

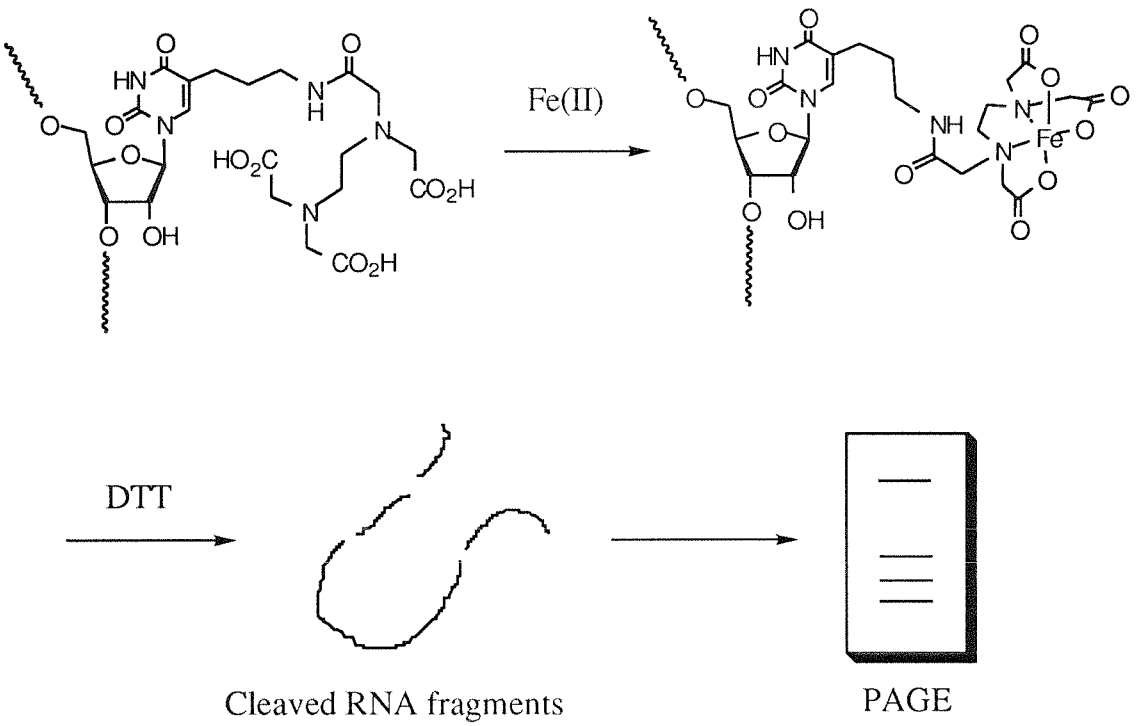



**Figure 4.1.** Schematic diagram for autocleavage of RNA

nucleic acids[5]. Affinity cleavage of *tetrahymena* ribozyme by bound G-DTPA•Fe(II) has successfully provided information about the tertiary structure of the catalytic core of the ribozyme[6]. Recently, Han and Dervan have demonstrated that autocleavage of $tRNA^{phe}$ with EDTA•Fe(II) at a unique base position is useful for studying RNA tertiary structure[7](Figure 4.1). Incorporation of EDTA•Fe(II) at a discrete base of RNA allows an RNA cleaving moiety to be specifically located within the folded RNA. Cleavage at several nucleotide positions proximal in space to the unique location of the bound EDTA•Fe(II) affords RNA fragments of different size and amount. Hence structural information regarding the nucleotides neighboring the discretely bound ligand- EDTA•Fe(II) can be determined by electrophoretic separation of nucleic acid cleavage products on a high-resolution polyacrylamide gel.

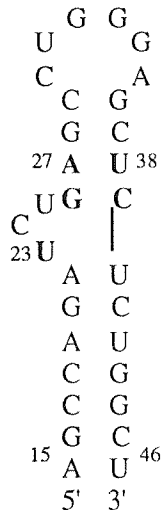



**Figure 4.2.** Sequence and secondary structure of TAR. The bold uridine in the bulge and the bold A•U and G•C base pairs are essencial for Tat peptide binding.

<u>HIV-I TAR RNA</u> Replication of the human immunodeficiency virus (HIV) is controlled by the binding of the viral Tat protein to the trans-activation response element (TAR) RNA sequence which is located at the 5' end of the viral mRNA[8]. TAR adopts a stable 58-nucleotide hairpin structure (Figure 4.2). The TAR hairpin contains a six-nucleotide loop and a three-nucleotide bulge that are both essential for Tat activity. The Tat protein contains an arginine-rich region located in the carboxy-terminal half of the molecule[8]. Tat peptides that include this sequence have been shown to bind specifically to the same site in TAR as Tat protein. Chemical interference studies reveal that the Tat peptide binds to a UCU bulge region in the major groove of TAR and contacts several bases and phosphates below the bulge[9]. Mutagenesis studies have shown that the first uridine base of the bulge, U23, is critical for Tat binding[10]. In addition, specific recognition requires the identity of two base pairs in the stem above the pyrimidine bulge, G26•C39 and A27•U38. Frankel and coworkers have proposed the "arginine fork" model for the Tat-TAR interaction[9]. In this model, the arginine guanidinium

group in the Tat peptide forms a set of hydrogen bonds with two precisely positioned phosphates. These interactions play an important role in the affinity and specificity. Further studies have shown that the free amino acid arginine binds specifically to TAR[11]. This model has been confirmed by recent NMR studies which suggest that arginine makes specific hydrogen bonds with G26 and two adjacent phosphates in the major groove[12]. These studies suggest that the complex is stabilized by a base triplet interaction between U23 in the bulge and the A27•U38 base pair above the bulge. The three-dimensional structure of the TAR is not yet known. In canonical A-form RNA, the major groove is too deep and narrow to be accessible for peptide recognition. Chemical modification experiments by Weeks and Crothers have revealed that the bulged bases of the TAR may introduce local distortions of the A-form double helix so that the major groove becomes wider and more accessible to protein binding[10,13].

The conformational flexibility of RNA seems to play an important role in RNA-protein interactions. In the case of HIV TAR, binding of arginine or the Tat peptide to TAR is accompanied by a conformational change in the RNA, which results in the formation of the essential base triplet (U23•A27•U38) not present in the free RNA[12]. These changes in RNA conformation have been seen in other RNA-protein systems. Comparison of the crystal structures of $tRNA^{Gln}$ free and bound to its cognate synthetase also reveals conformational differences in the tRNA[14]. Recent NMR studies of the Rev-RRE complex show that the RRE conformation changes upon binding of Rev[15]. These protein-induced conformational changes in RNA binding of Rev[15]. These protein-induced conformational changes in RNA appear to be common characteristics of RNA recognition by proteins.

Circular dichroism (CD) and NMR have been successfully used to

probe RNA conformational changes[12,16]. RNase footprinting and gel mobility shift experiments have also proven to be useful. In this study, we used the autocleavage technique to detect the conformational change in TAR driven by arginine binding. The approach of targeting an EDTA•Fe(II) at a specific site and then mapping the cleavage sites may provide a useful method for studying RNA-protein interactions.

## Results and Discussion

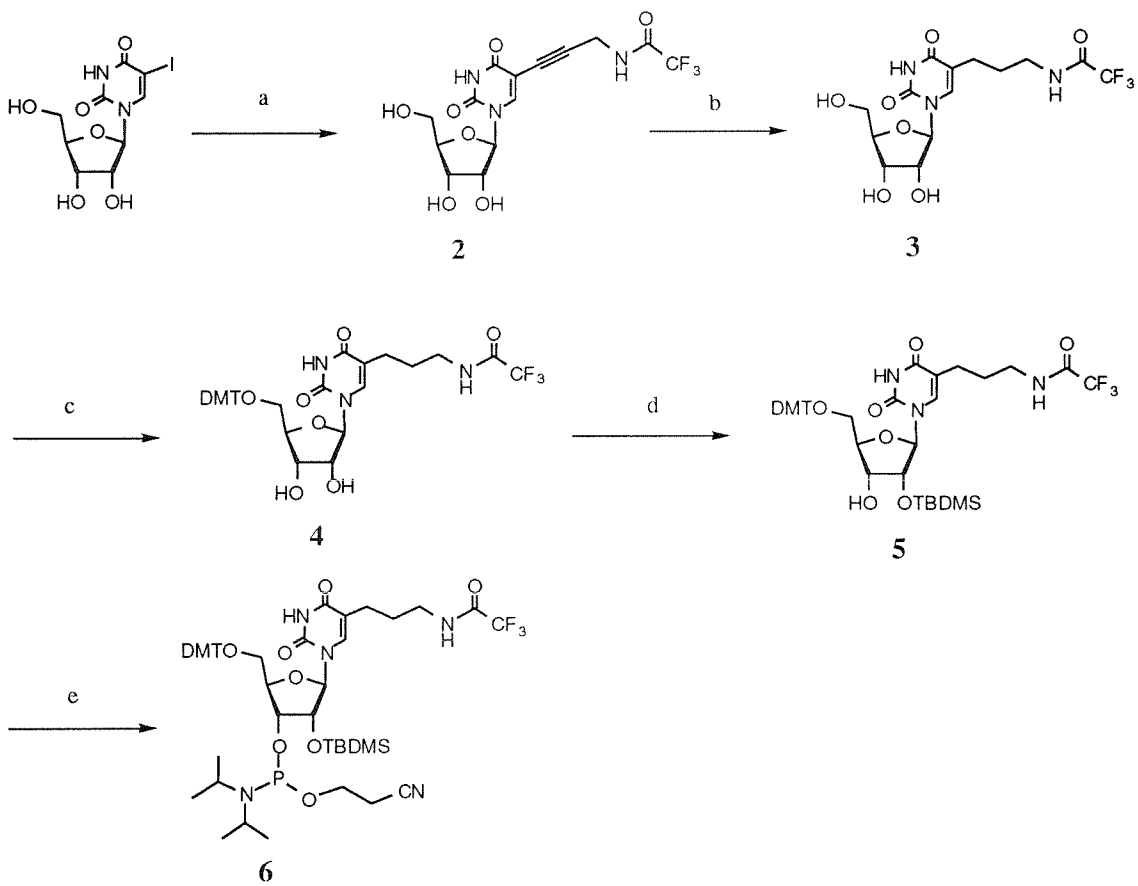



**Figure 4.3.** Synthesis of phosphoramidite 6 from commercially available 5-iodo-uridine. Reagents: (a) $(Ph_3P)_4Pd$, CuI, TEA; (b) $H_2$, Pd/C; (c) DMT-Cl, Pyridine; (d) tert-Butyldimethylsilyl chloride, $AgNO_3$; (e) 2-Cyanoethoxy N,N-diisopropylamino-chlorophosphine, DIEA, $CH_2Cl_2$.

## Synthesis of Phosphoramidite 6

RNA phosphoramidite **6** was synthesized in five steps from commercially available 5-iodo-uridine as described by Han and Dervan[17] (Figure 4.3).

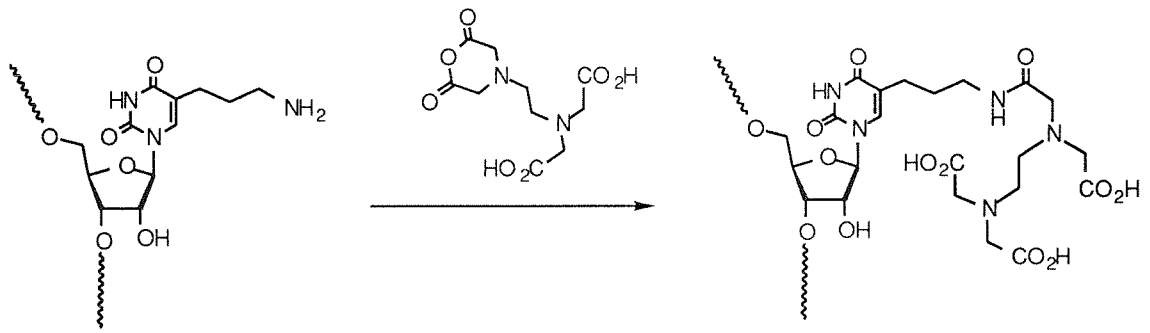


**Figure 4.4.** Post-automated synthesis modification of RNA

## Synthesis of TAR RNA-EDTA•Fe(II)

RNA oligonucleotides were synthesized by automated methods using 2-cyanoethyl phosphoramidite chemistry[18]. The phosphoramidite **6** was incorporated into RNA with 95 % coupling efficiency. EDTA was incorporated into TAR RNA by the post-automated synthesis modification of RNA oligonucleotides containing a primary amine linked to the 5-position of uridine with EDTA monoanhydride or EDTA dianhydride. After the deprotection process, fully deprotected oligonucleotide was treated with EDTA monoanhydride to afford TAR-EDTA (Figure 4.4).

## Autocleavage of TAR RNA

The autocleavage experiments were performed on the 5'-end-labeled 31-nucleotide TAR RNA (U*23) in 10 mM Tris HCl (pH 7.5) and 70mM NaCl. After Fe(II) was added, auto cleavage reactions were initiated by the addition of dithiothreitol and incubated for 12h at 25 $^{0}$C. Separation of the RNA cleavage products by high resolution gel electrophoresis revealed that

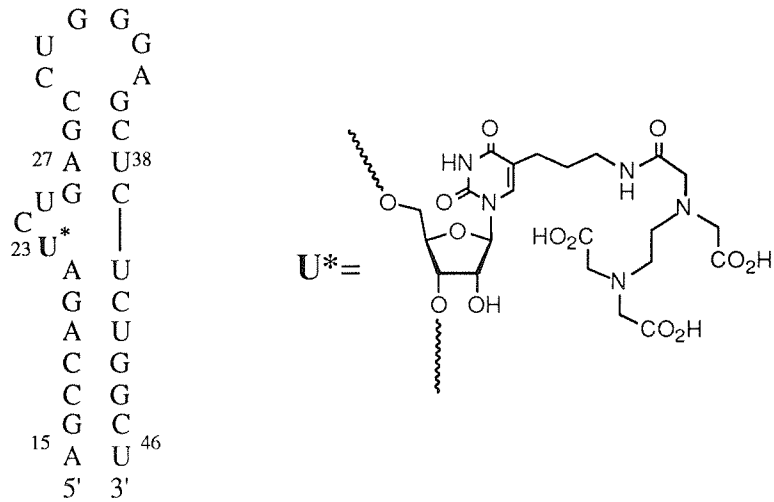



**Figure 4.5.** Structure of TAR RNA(U* 23)•Fe(II)

cleavage sites were located between A22 and A20 which are adjacent to U*23 (Figure 4.6, 4.8). The overall cleavage yield was 3%.

In order to probe the conformational change of TAR RNA driven by the specific binding of arginine, 5'-end-labeled TAR was treated with DTT in the presence of 50mM arginine. Changes of the cleavage sites were observed in the presence of arginine. The main site of cleavage shifts from A22-A20 to three bulge bases (U23-U25) with weaker cleavage observed at U31, G32, and A35 (Figure 4.6, 4.8).

In order to examine the effect of arginine on the conformation of TAR, autocleavage experiments were performed in the presence of increasing concentrations of arginine. The concentration where arginine binding changes the cleavage patterns is in reasonable agreement with the $K_d$ of ~4mM determined using a peptide competition assay by Tao and Frankel[11] (Figure. 4.7). However, a similar change in cleavage specificity was not observed in experiments with argininamide.

**Figure 4.6.** Autocleavage of HIV-I TAR-U*23. Autoradiogram of a 20 % denaturing polyacrylamide gel showing autocleavage reactions of (A) TAR-U*23: lane 1, alkaline hydrolysis; lane 2, intact TAR; lane 3, autocleavage reaction; lanes 4, autocleavage reaction in the presence of 50 mM arginine; (B) TAR(A27U38→U27A38)-U*23: lane 1, alkaline hydrolysis; lane 2, intact TAR; lane 3, autocleavage reaction; lanes 4, autocleavage reaction in the presence of 50 mM arginine; All reactions contain 20 kcpm TAR RNA, 10 mM Tris-HCl (pH 7.5), 10 mM NaCl,.

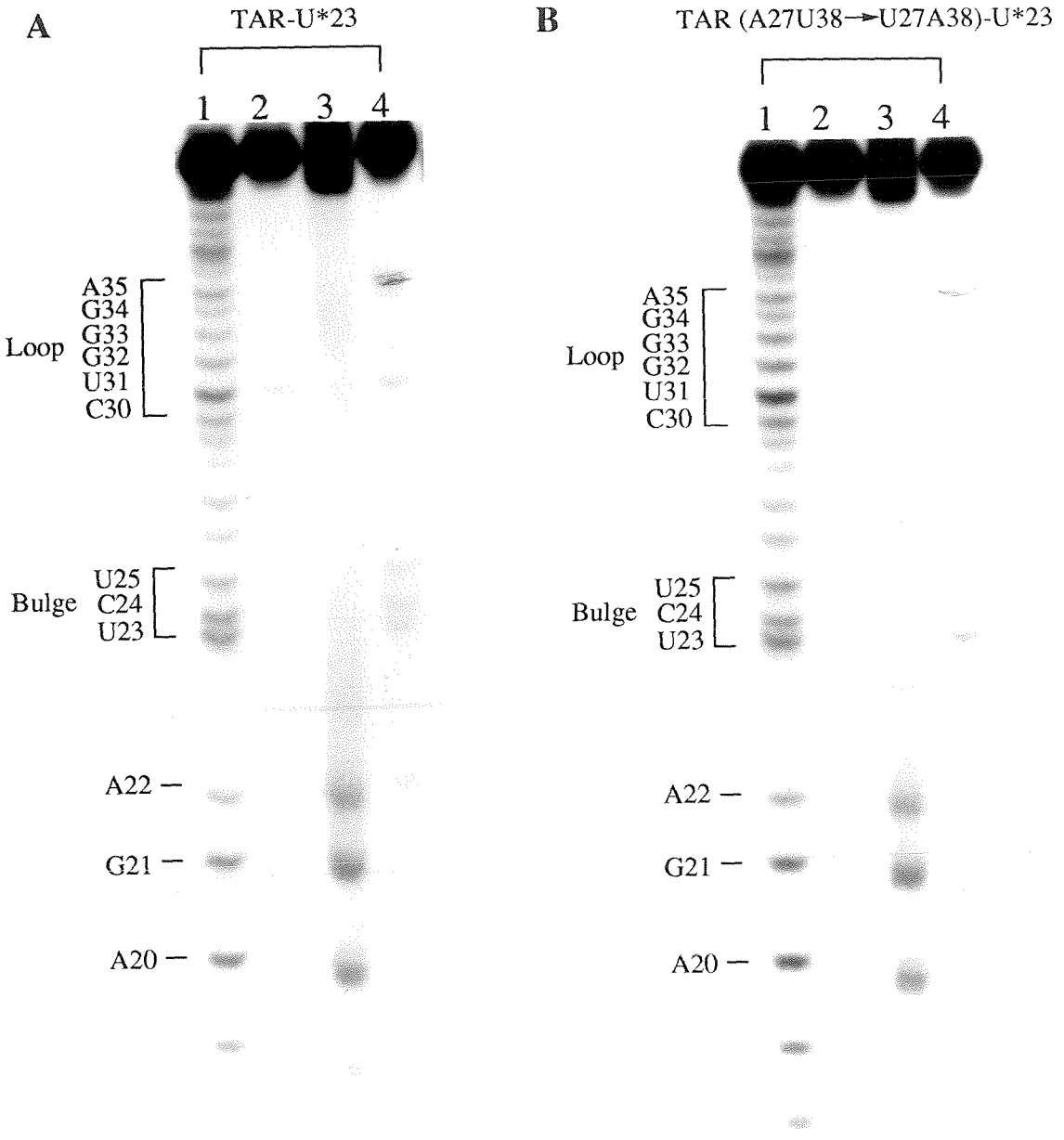
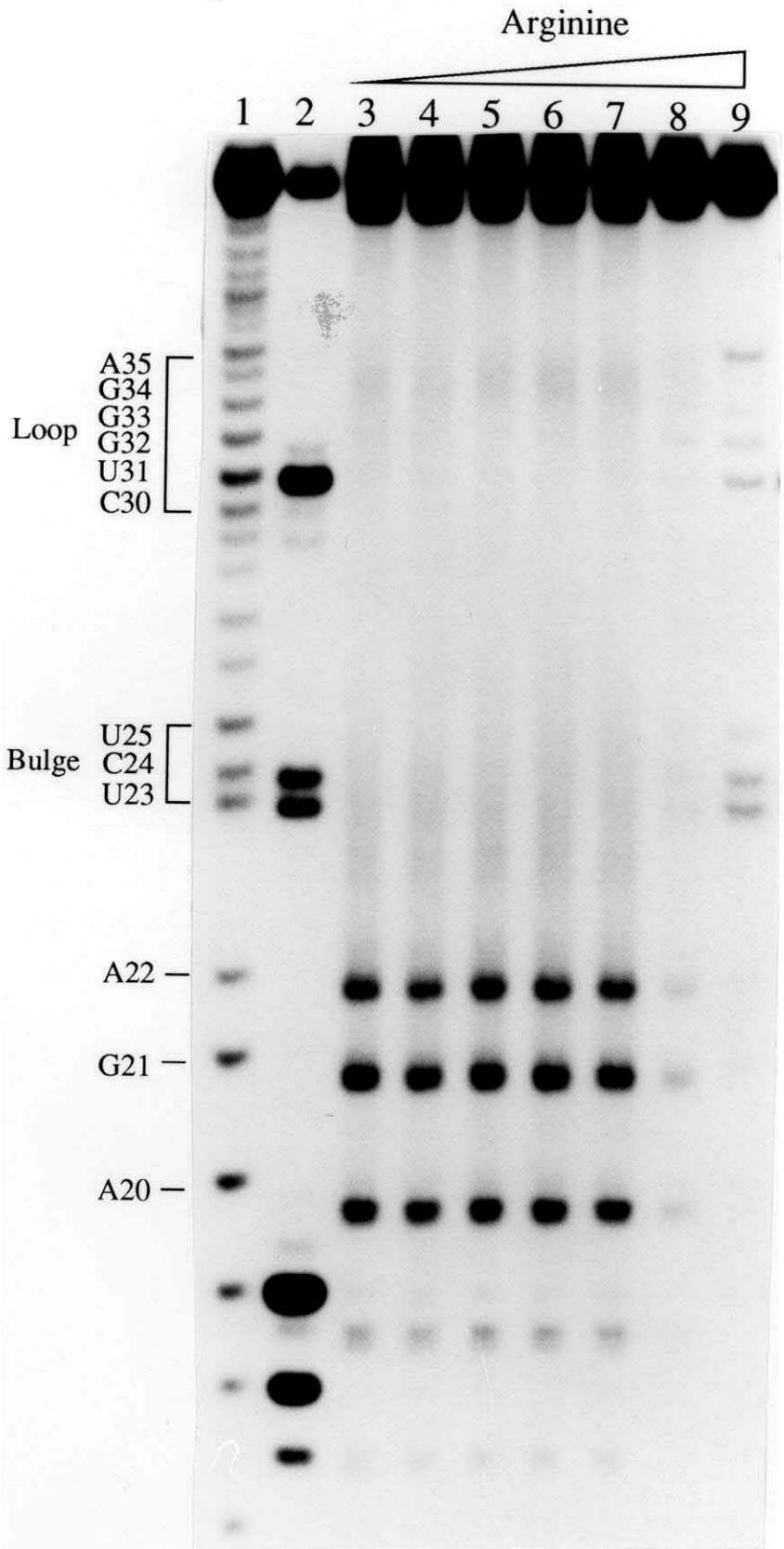

A
TAR-U*23
1 2 3 4
Loop
A35
G34
G33
G32
U31
C30
Bulge
U25
C24
U23
A22
G21
A20
B
TAR (A27U38→U27A38)-U*23
1 2 3 4
Loop
A35
G34
G33
G32
U31
C30
Bulge
U25
C24
U23
A22
G21
A20

**Figure 4.7.** Autocleavage of HIV-1 TAR-U*23 in the presence of arginine. Autoradiogram of a 20 % denaturing polyacrylamide gel showing autocleavage reactions of TAR-U*23. Lane 1, limited alkaline hydrolysis; lane 2, intact TAR; lane 3, autocleavage reaction; lanes 4-9, autocleavage reaction in the presence of 0.1 mM, 0.3 mM, 1 mM, 3 mM, 10 mM, 50 mM arginine, respectively. All reactions contain 20 kcpm TAR RNA, 10 mM Tris-HCl (pH 7.5), 10 mM NaCl,.


Arginine
1
2
3
4
5
6
7
8
9
Loop
A35
G34
G33
G32
U31
C30
Bulge
U25
C24
U23
A22
G21
A20

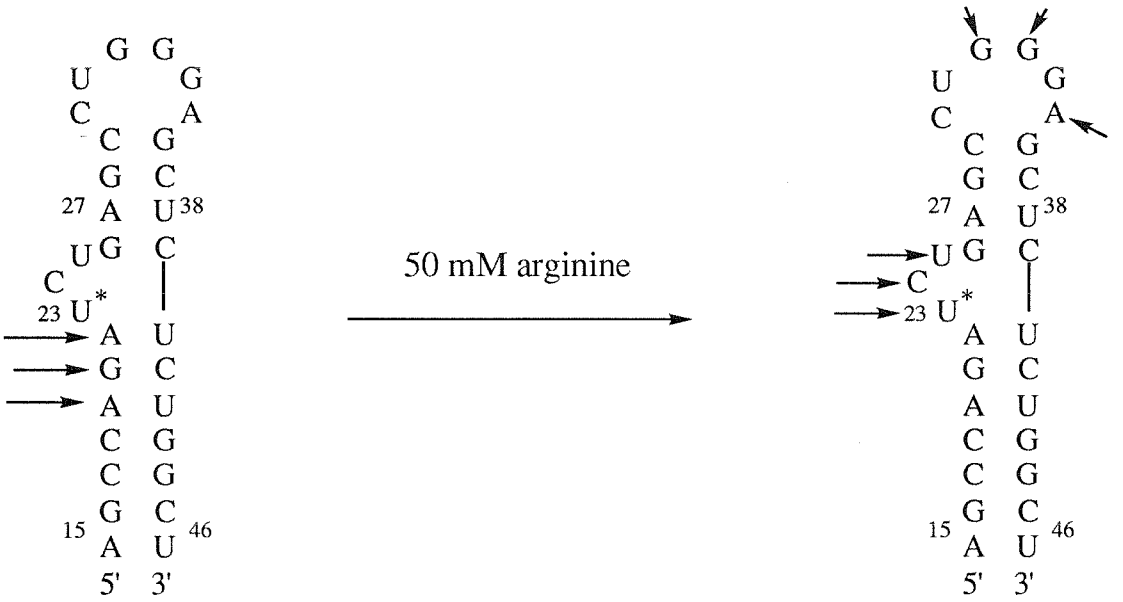
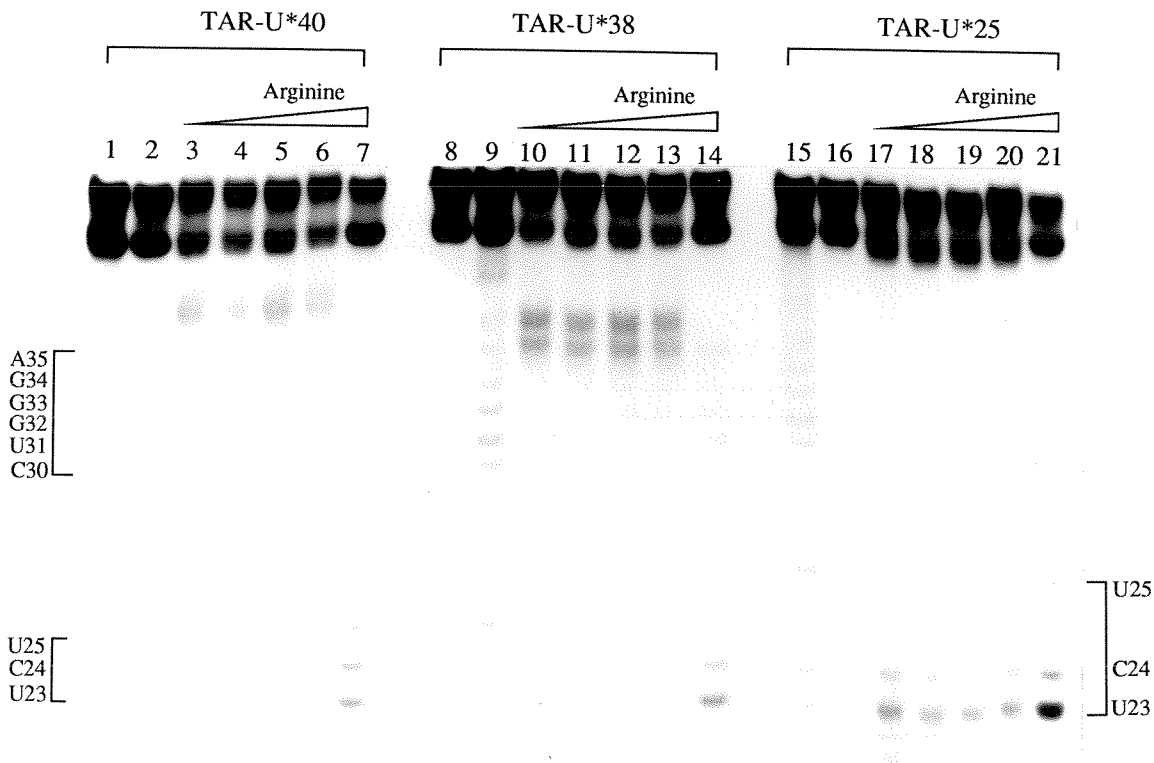



**Figure 4.8.** Histogram of the cleavage data derived from autocleavage of TAR-U*23•Fe(II). Arrow heights are proportional to cleavage intensities at the indicated phosphodiesters

To test whether the change in the cleavage sites resulted from the conformational change in TAR, we performed autocleavage reactions with the double mutant (A27•U38→U27•A28) of TAR RNA. The A27•U38 base pair is essential for specific interactions with either arginine or Tat peptide. Weeks and Crothers revealed that this mutant of TAR (U27•A38) show melting and nuclease susceptibility behavior indistinguishable from the wild type TAR. The autocleavage assays of this mutant TAR showed the same cleavage patterns as wild type TAR (Figure 4.6). This result suggests that the change in cleavage pattern of TAR is not a result of the specific binding of arginine to TAR. Because the bulge is critical for arginine binding, we reasoned that the EDTA group from U*23 might block the specific binding of arginine to TAR.

In order to test this idea further, a series of TARs in which U* is located at other bases were prepared. Autocleavage experiments with these TAR RNAs showed that the cleavage occurs at the same site (U23-U25, bulge) in the presence of 50 mM arginine regardless of the position of EDTA Fe(II) attachment (Figure 4.9, 4.10). These results suggest that the cleavage observed in the presence of arginine may not be due to oxidative scission of the ribose backbone of TAR by hydroxyl radicals. One explanation for these results is that the oxidative cleavage of RNA by hydroxyl radicals does not occur in the presence of arginine at 50 mM. Because the concentration of arginine is higher than that of the buffer, the observed cleavage in the bulge of TAR may result from the change in pH of the solution. Although autocleavage of TAR failed to probe the conformational change of RNA driven by binding of the small molecule, this approach may be useful for studying other RNA-protein interactions as well as tertiary structures of larger RNAs.

**Figure 4.9.** Autocleavage experiments of TAR-U*40, TAR-U*38, and TAR-U*25. Autoradiogram of a 20 % denaturing polyacrylamide gel showing the autocleavage reactions. [TAR-U*40] lane 1, limited alkaline hydrolysis; lane 2, intact TAR; lane 3, autocleavage reaction; lanes 4-7, autocleavage reaction in the presence of 1 mM, 3 mM, 10 mM, 50 mM arginine, respectively. [TAR-U*38] lane 8, intact TAR; lane 9, limited alkaline hydrolysis; lane 10, autocleavage reaction; lanes 11-14, autocleavage reaction in the presence of 1 mM, 3 mM, 10 mM, 50 mM arginine, respectively [TAR-U*25] lane 15, limited alkaline hydrolysis; lane 16, intact TAR; lane 17, autocleavage reaction; lanes 18-21, autocleavage reaction in the presence of 1 mM, 3 mM, 10 mM, 50 mM arginine, respectively. All reactions contain 20 kcpm TAR RNA, 10 mM Tris-HCl (pH 7.5), 10 mM NaCl.


TAR-U*40
TAR-U*38
TAR-U*25
Arginine
Arginine
Arginine
1 2 3 4 5 6 7
8 9 10 11 12 13 14
15 16 17 18 19 20 21
A35
G34
G33
G32
U31
C30
U25
C24
U23
U25
C24
U23

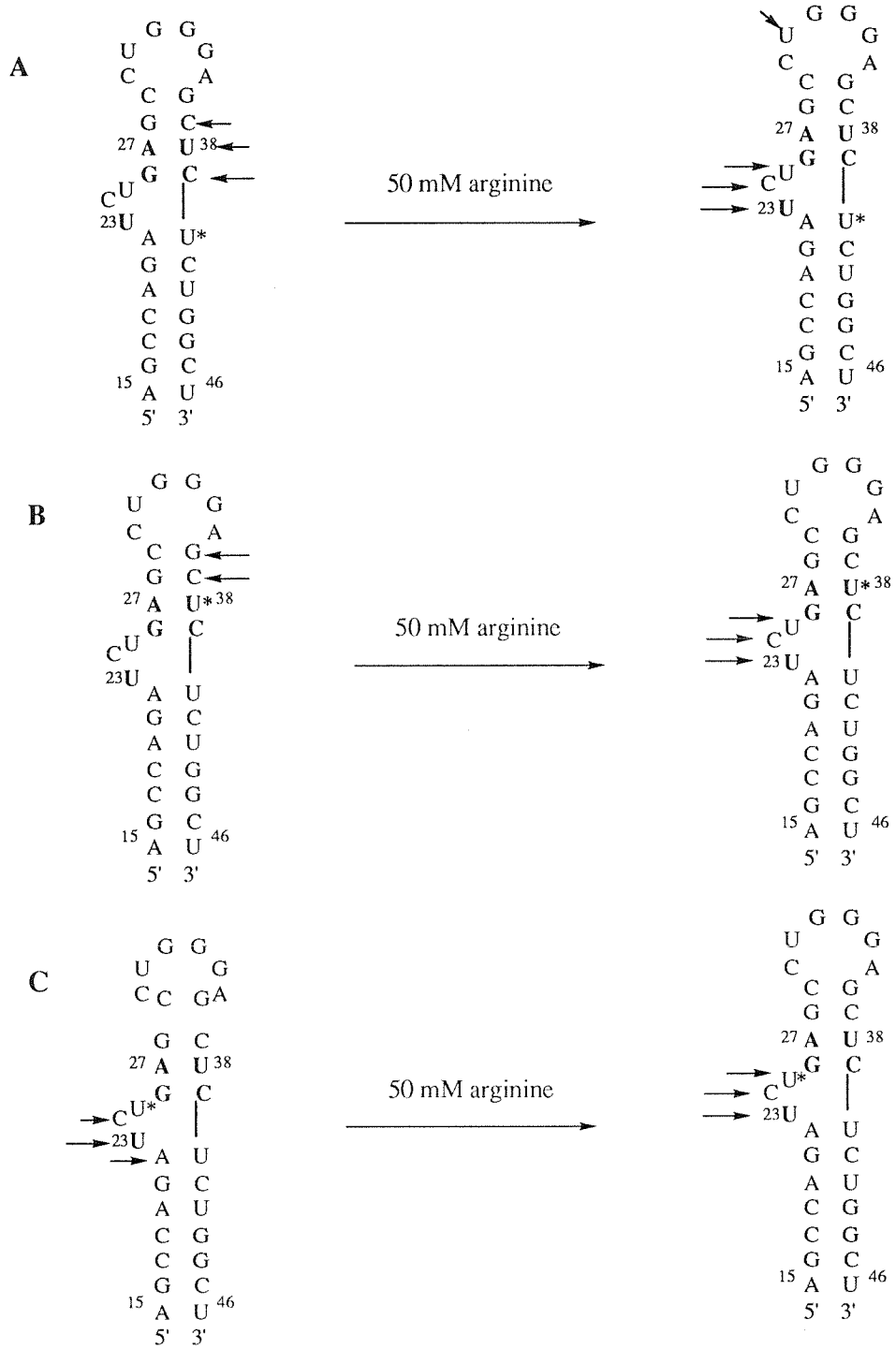



Figure 4.10. Histograms of the cleavage data derived from autocleavage of (A) TAR-U*40•Fe(II) (B) TAR-U*38•Fe(II) (C) TAR-U*25•Fe(II). Arrow heights are proportional to cleavage intensities at the indicated phosphodiesters.

## Experimental Section

$^1H$ NMR spectra were recorded at 300 MHz on a GE 300 NMR in $CDCl_3$ and DMSO-$d_6$. Chemical shifts are reported in parts per million relative to tetramethylsilane or residual DMSO-$d_5$. Ultraviolet-visible spectra were recorded on a Hewlett-Packard 8452 A diode array spectrophotometer. High-resolution mass spectra (HRMS) were recorded using fast atom bombardment (FAB) techniques at the Mass Spectrometry Laboratory at the university of California, Riverside. Flash column chromatography was carried out using silica gel 60 (230-400 mesh, Merck). Thin-layer chromatography (TLC) was performed on silica gel 60 $F_{254}$ precoated plates (Merck). All chemicals for the synthesis were purchased from Aldrich unless otherwise specified. Dichloromethane and N,N-dimethylformamide (DMF) were purchased as anhydrous solvents from Aldrich. Arginine and argininamide were purchased from Sigma and used without further purification. Adenosine 5'-[$\gamma$-$^{32}P$] triphosphate was obtained from Amersham. Phosphoramidites were purchased from Applied Biosystems (dA, dG, dC, and T) and from BioGenex Laboratories (rA, rG, rC, and U). Storage phosphotechnology autoradiography was performed using a Molecular Dynamics 400S Phosphorimager and ImageQuant software.

**Compound 2** To a solution of 5-iodouridine (4.1 g, 11.08 mmole) in 66 ml of DMF was added propargyl trifluoacetamide (5.02 g, 33.24 mmole), $Pd(PPh_3)_4$ (1.28 g, 1.11 mmole), CuI (0.42 g, 2.22 mmole), and triethylamine (1.68 g, 16.62 mmole. The mixture was stirred at r.t. for 24 h under Ar and evaporated to dryness under reduced pressure. Anion exchange resin (DEAE Sephadex, 2 g) and 1 : 1 $CH_2Cl_2$-MeOH (50 ml) were added. The mixture was stirred for 1 h. and then filtered through Celite. After evaporation of the filtrate, the residue was purified by the silica gel chromatography using 10%

MeOH in $CH_2Cl_2$ as a eluent to give 2.75 g of **2** (65% yield). $^1H$ NMR (DMSO-$d_6$) δ 8.22 (s, 1H), 5.72 (d, 1H), 4.2 (s, 2H), 4.01 (m, 1H), 3.93 (m, 1H), 3.81 (brs, 1H), 3.57 (dd, 2H)

**Compound 3** Compound **2** (2.2 g, 5.59 mmole) was dissolved in 200 ml of MeOH, and 10% Pd/C (200 g) was added, and the mixture was hydrogenated in a Parr apparatus (overnight, 40 psi). The reaction mixture was filtered through Celite and the filtrate was evaporated to give 2.1 g of **3** (5.28 mmole). $^1H$ NMR (DMSO-$d_6$) δ 7.76 (s, 1H), 5.79 (d, 1H), 4.03 (m, 1H), 3.98 (m, 1H), 3.84 (brs, 1H), 3.59 (dd, 2H), 3.17 (m, 2H), 2.21 (m, 2H), 1.63 (t, 2H)

**Compound 4** To a solution of **3** (1.44 g, 3.62 mmole) in dry pyridine (24 mL) at 0 $^0C$ was added dimethoxytrityl chloride (1.41 g, 4.17 mmole). The mixture was stirred at 0 $^0C$ for 12 h and then diluted with 15 ml of MeOH. After stirring for 30 min, the solvent was removed under reduced pressure. A 5% $NaHCO_3$ solution (30 ml) was added and the resulting solution was extraceted with 100 ml of $CH_2Cl_2$. The methylene chloride solution was washed once with water, dried over $MgSO_4$, and evaporated to leave the crude residue. The residue was purified by the silica gel chromatography using 5% MeOH in $CH_2Cl_2$ to give 1.01 g of **4** (40% yield). $^1H$ NMR (DMSO-$d_6$) δ 7.50-7.12 (m, 10H), 6.91 (d, 4H), 5.78 (d, 1H), 4.18 (m, 1H), 4.09 (m, 1H), 3.95 (brs, 1H), 3.74 (s, 6H), 3.17 (m, 2H), 1.82 (m, 2H), 1.44 (m, 2H)

**Compound 5** Compound **4** (5.6 g, 8.07 mmole) was dissolved in 80 ml of THF followed by the addition of pyridine (2.41 ml, 29.9 mmole) and silver nitrate (1.65 g, 9.68 mmole). The solution was stirred until all the silver nitrate had dissolved and then tert-butyldimethylsilyl chloride (1.58 g, 10.49 mmole) was added and the mixture was stirred at r.t. for 5 h. The solution was then filtered into a 5% $NaHCO_3$ solution (80 ml) and extracted with 250 ml of $CH_2Cl_2$. The extracts were evaporated and then the residue was purified

by the silica gel chromatography using 25% hexane in ether to give 2.94 g of **5** as a white solid (45% yield). $^1$H NMR ($CDCl_3$) δ 8.19 (s, 1H), 7.28-7.05 (m, 9H), 6.71 (d, 4H), 5.89 ( d, 1H), 4.38 (t, 1H), 4.21 (m, 1H), 4.09 (brs, 1H), 3.72 (brs, 2H), 3.69 (s, 6H), 3.33 (dd, 2H), 1.79 (m, 1H), 1.54 ( m, 1H), 1.29 (m, 2H), 0.81 (s,9H), 0.12 (d,6H)

**Phosphoramidite 6** To a solution of **5** (0.12 g, 0.15 mmole) in THF (1 ml) were added N, N-diisopropylethylamine (DIEA, 0.102 ml, 0.58 mmole) and 2-cyanoethyl N, N-diisopropylchlorophosphoramidite (0.030 ml, 0.19 mmole). After stirring for 1 h at 25 $^0$C, the reaction mixture was concentrated and purified by flash chromatography using 30% hexane/1% $Et_3N$ ethyl eher to give 0.090 g (60%) of the diastereomeric phosphoramidite **6** as a white foam. TLC (20% Hexane/Ethyl ether); $^1$H NMR δ 11.50 (m, 1H), 9.36 (m, 1H), 7.47-7.24 (m, 10H), 6.89 (m, 4H), 5.88 (m, 1H), 4,48 (m, 1H), 4.22-4.14 (m, 2H), 3.73 (s, 6H), 3.61-3.51 (m, 4H), 3.30 (m, 2H), 2.94 (m, 2H), 2.51 (m, 2H), 1.84 (m, 2H), 1.46 (m, 2H), 1.11 (m, 12H), 0.83 (m, 9H), 0.50 (m, 6H)

**TAR-EDTA Synthesis** Oligoribonucleotides were synthesized on an Applied Biosystems Model 394 DNA/RNA synthesizer using 2-cyanoethyl phosphoramidite chemistry. Phosphoramidites **6** (0.15 M solutions in $CH_3CN$) was incorporated into RNA and DNA, respectively. RNA (1 μmol) containing nucleoside **6** was deprotected by treatment with $NH_3$-saturated methanol (4 ml) at 25 $^0$C for 24 h and lyophilized. The resulting residue was treated with 0.8 ml of 1M TBAF/THF for 12 h. After evaporation in a Speed-Vac evaporator, water was added to the Eppendorf tube in 800 μL. The solution was applied directly to a sephadex G-25F column, eluted with water, and filtered. The crude RNA solution was evaporated and purified by eletrophoresis on a preparative 20% polyacrylamide gel. The purified RNA (5 nmol) in 0.2 M $NaHCO_3$ (50 μL, pH 8.17) was treated with either EDTA

dianhydride (25 μL of 1% w/v DMF solution) or EDTA monoanhydride (5 mg) for 2 h at 25 $^0$C. The reactions were quenched by addition of 100mM Tris HCi buffer pH 7.2 (25 μL) and $H_2O$ (150 μL), followed by ehanol precipitation. After purification of the crude oligonucleotide-EDTA on 20% polyacrylamide/7 M urea gels, UV-absorbing bands were excised, crushed, and eluted (0.3 M NaOAc, pH 5.2, 37 $^0$C, 24 h). The resulting solutions were filtered (0.45-μm Centrex filter, Schleicher and Schuell) and desalted (NAP-5 column). RNAs were store at -20 $^0$C until use. RNA concentrations were quantitated by UV spectroscopy using the following molar extinction coefficients: 15,400 (A), 11,700 (G), 7,300 (C), 9,900 (U) cm M.

**Cleavage Reactions** TAR RNAs were labeled at the 5'-end with T4 polynucleotide kinase and [$\gamma$-$^{32}$P]-ATP, precipitated with ethanol, and dissolved in $H_2O$. TAR RNAs were renatured by heating to 70 $^0$C for 1 min in 10 mM Tris-HCl, pH 7.5, and 70 mM NaCl and slow cooling to room temperature prior to use. A typical 20 μL reaction mixture contained 5'-end-labeled RNA (~200,000 cpm), 10 mM Tris HCl, pH 7.5, and 70 mM NaCl. Autocleavage reactions for TAR (*U23) were performed by incubation of TAR (*U23) with $Fe(NH_4)_2(SO_4)_2$ $6H_2O$ (1 μM) for 2 h at 25 $^0$C followed by addition of dithiothreitol (4 mM). After 8 h at 25 $^0$C, the cleavage reactions were analyzed by loading 20,000 cpm of the sample on a 20% polyacrylamide/7M urea wedge gel (WxL=31.0 x 38.5 cm, 0.4 mm at top and 1.2 mm at bottom). The amount of product fragments was quantitated by Phosphor Imager analysis (Molecular Dynamics).

## References


1. (a) Kim, S. H.; Sussman, J. L.; Suddath, F. L.; Quigley, G. J.; McPherson, A.; Wang, A. H.; Seeman, N. C.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1974**, 71, 4970-4974. (b) Quigley, G. J.; Rich, A. *Science* **1976**, 194, 796-806.
2. Pley, H. W.; Flaherty, K. M.; McKay, D. B. *Nature* **1994**, 372, 68-74.
3. (a) Puglisi, J. D.; Wyatt, J. R.; Tinoco, I., Jr. J. *Mol. Biol.* **1990**, 214, 437-453. (b) Cheong, C.; Varani, G.; Tinoco, I., Jr. *Nature* **1990**, 346, 680-682. (c) Varani, G.; Cheong, C.; Tinoco, I., Jr. *Biochemistry* **1990**, 30, 3280-3289.
4. (a) Latham, J. A.; Cech, T. R. *Science* **1989**, 245, 276-282. (b) Celander, D. W.; Cech, T. R. *Biochemistry* **1990**, 29, 1355-1361.
5. (a) Dervan, P. B. *Science* **1986**, 232, 464-471. (b) Tayler, J. S.; Schultz, P. G.; Dervan, P. B. *Tetrahedron* **1984**, 40, 457-465. (c) Sluka, J. P.; Griffin, J. H.; Mack, D. P.; Dervan, P. B. *J. Am. Chem. Soc.* 1990, 112, 6369-6374. (d) Dreyer, G. B.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, 82, 968-972. (e) Moser, H. E.; Dervan, P. B. *Science* **1987**, 238, 645-650.
6. Wang, J. -F.; Cech, T. R. *Science* **1992**, 256, 526-529.
7. Han, H.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, 91, 4955-4959.
8. (a) Calnan, B. J.; Biancalana, S.,; Hudson, D.; Frankel, A. D. *Genes Dev.* **1991**, 5, 201-210. (b) Cordingly, M. G.; LaFemina, R.L.; Callahan, P. L.; Condra, J. H.; Sardana, V. V.; Graham, D. J.; Nguyen, T. M.; LeGrow, K.; Gotlib, L.; Schlabach, A. J.; Colonno, R. J. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, 87, 8985-8989. (c) Weeks, K. M.; Ampe, C.; Schultz, S. C.; Steitz, T. A.; Crothers, D. M. *Science* **1990**, 249, 1281-1285
9. Calnan, B. J.; Tidor, B.; Biancalana, S.; Hudson, D.; Frankel, A. D. *Science* **1991**, 252, 1167-1171.
10. Weeks, K. M.; Crothers, D. M. *Cell* **1991**, 66, 577-588.

11. Tao, J.; Frankel, A. D. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, 89, 2723-2726.

12. Puglisi, J. D.; Tan, R.; Calnan, B. J.; Frankel, A. D.; Williamson, J. R. *Science* **1992**, 257, 76-80.

13. Weeks, K. M.; Crothers, D. M. *Science* **1993**, 261, 1574-1577.

14. (a) Rould, M. A.; Perona, J. J.; Soll, D.; Steitz, T. A. *Science* **1989**, 246, 1135-1142. (b) Rould, M. A.; Perona, J. J.; Steitz, T. A. *Nature* **1991**, 352, 213-218. (c) Ruff, M.; Krishnaswamy, S.; Boeglin, M.; Poterszman, A.; Mitshler, A.; Podjarny, A.; Rees, B.; Thierry, J. C.; Moras, D. *Science* **1991**, 252, 1682-1689.

15. Battiste, J. L.; Tan, R.; Frankel, A. D.; Williamson, J. R. *Biochemistry* **1994**, 33, 2741-2747.

16. Tan, R.; Frankel, A. D. *Biochemistry* **1992**, 31, 10288-10294.

17. Han, H.; Dervan, P. B. *Nucleic Acids Res.* **1994**, 22, 2837-2844.