Design of Peptides for Sequence-Specific Recognition of the Minor Groove of DNA

Thesis by

Milan Mrksich

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1994

(Submitted March 8, 1994)

Acknowledgments

I would like to thank my advisor Professor Peter Dervan for his thoughtful guidance, generous lessons and friendship. Peter's excitement and enthusiasm for science are contagious and have contributed to a very enjoyable stay at Caltech. His early patience and timely insights have allowed me to grow far beyond my expectations. As an undergraduate I was very fortunate to have worked with Professor Steven Zimmerman. Steve introduced me to organic chemistry and his dedication to his students will always serve as an example. I have also enjoyed an extended collaboration and friendships with Professor David Wemmer, Bernhard Geierstanger and Tammy Dwyer. I appreciate the helpful advice and guidance offered by my committee; Professors Myers, Dougherty and Goddard.

My stay at Caltech has been made more enjoyable by the many friends I have encountered. George, Jim, Jurg, Michelle, Natalia, Matt, Kevin and Jong Sung Koh have been great labmates. Midnight beers with Joe and occasional outings with Louis and Sharon have been great fun. Dave and Erick provided helpful advice early on and friendships throughout my stay.

My stay at Caltech has been made special by sharing it with Shannon. Her understanding and support of long hours spent in lab are appreciated. Time spent with her away from work has kept everything in perspective.

The love, support and encouragement offered by my parents and family have been especially helpful during the difficult times. Their lessons through example have inspired me and continue to guide me.

To My Parents

Abstract

Distamycin A and netropsin are natural products which bind in the minor groove of DNA at four and five A,T base pair sites, respectively. While the peptides pyridine-2-carboxamide-netropsin (2-PyN) and 1-methylimidazole-2-carboxamide-netropsin (2-ImN) were designed to bind 5′-(G,C)(A,T)$_3$-3′ sequences in a single orientation, they bind 5′-TGTCA-3′ with no orientation preference. Direct characterization of the 2-ImN•5'-TGTCA-3' complex by NMR reveals that the peptide binds in the minor groove as a side-by-side dimer.

Covalent peptide dimers wherein the nitrogens of the central pyrroles of 2-PyN are connected with propyl, butyl, pentyl and hexyl linkers bind the 5′-TGTCA-3′ site with affinities ten-fold greater than that of the non-linked peptides. The ratio of binding affinities of 2-PyN for 5′-TGTCA-3′ and 5′-TTTTT-3′ sites have been altered from 1:1 to 20:1.

Footprinting and affinity cleaving experiments demonstrate that the two peptides distamycin and 2-ImN simultaneously bind the five-base pair sequence 5′-TGTTA-3′ as a side-by-side heterodimer in the minor groove. The specific affinity for this site was enhanced with the design and synthesis of a covalent peptide heterodimer.

Four hexapeptides were synthesized wherein the terminal amine and carboxyl groups of distamycin and 2-ImN, respectively, are connected with amino acids. 2-ImN-GABA-P3 binds the 5'-TGTTA-3' site with similar binding affinity as the first generation covalent peptide dimer. The general and efficient synthetic methodology for preparation of GABA linked peptides may allow the design of a new class of hairpin peptide-turn-peptides for specific recognition of many different sequences in the minor groove of DNA.

Footprinting and affinity cleaving experiments demonstrate that the designed peptide ImPImP specifically binds the designated six-base-pair sites 5′-AGCGCT-3′ and 5′-TGCGCA-3′ in two equal orientations, consistent with a side-by-side antiparallel arrangement of peptides in the minor groove. This example underscores the utility of 2:1 peptide-DNA models for the design of ligands for sequence-specific recognition of designated DNA sites. Binding of this designed peptide to a pure four base pair G,C-core sequence represents an absolute reversal of the specificity of the natural products distamycin and netropsin.

Table of Contents

List of Tables and Figures

Chapter 3

Chapter 4

Chapter 1

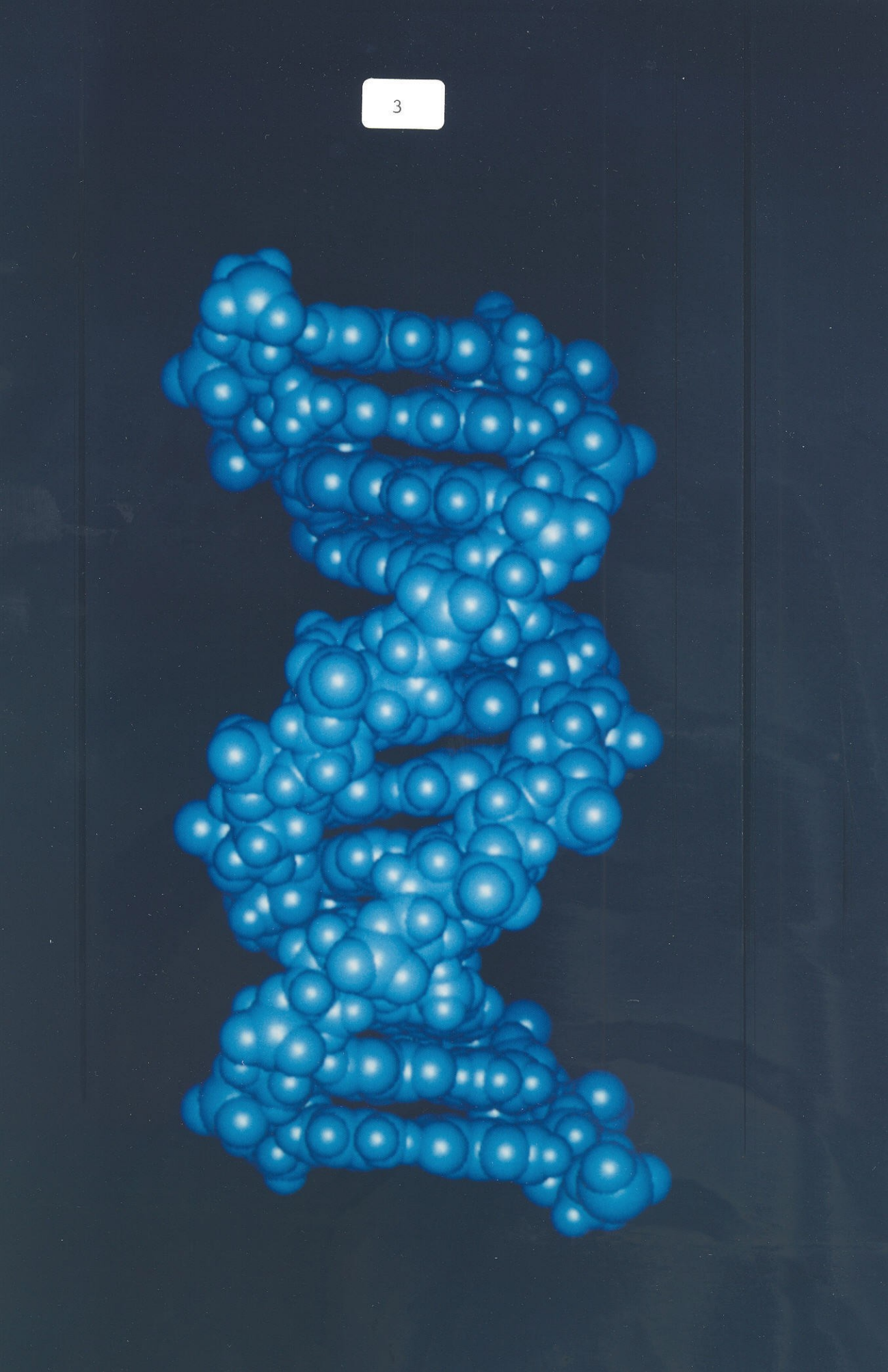
Introduction

Double-helical deoxyribonucleic acids (DNA) play a central role in biological processes. In addition to containing the hereditary material, it is the specific non-covalent interactions of proteins with DNA that determine cellular function and regulation. The mechanisms by which protein binding to DNA effects cellular changes and control are being elucidated by biologists. A challenging goal for chemists is the design of ligands for specific recognition of designated sequences. The ability to design synthetic ligands for sequence-specific recognition of DNA can have important implications ranging from new reagents in molecular biology to pharmaceuticals which act at the level of gene expression.

Double-helical DNA. Double-helical DNA is composed of two polydeoxyribonucleotide strands which associate via hydrogen bonds between bases.[1] The common B-form DNA is characterized by a wide and shallow major groove and a narrow and deep minor groove (Figure 1). Sequences are distinguishable by the pattern of hydrogen bond donors and acceptors displayed on the edges of the base pairs (Figure 2). In addition, sequence-dependent structural variations, conformational properties and solvent and counterion organization can all distinguish local DNA structures.[1] Many different classes of compounds are known to bind DNA, and these offer strategies for the design of new ligands for sequence-specific recognition of DNA.

Protein-DNA Complexes. The past decade has witnessed a revolution in the number of high resolution structures of protein-DNA complexes available from x-ray crystallography.[2] Several different structural motifs adopted by

Figure 1. Molecular model of B-form double-helical DNA. The wide and shallow major groove and the narrow and deep minor groove are clearly distinguishable.



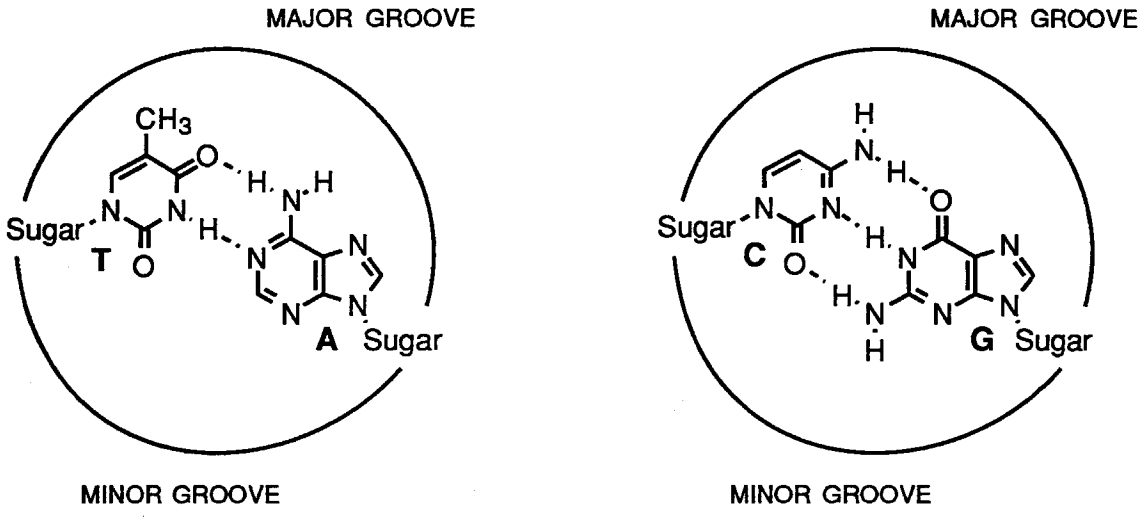


Figure 2. Models showing the hydrogen bonding patterns presented in the major and minor grooves of TA and CG base pairs. The minor groove of an AT base pair contains two hydrogen bond acceptor atoms. The minor groove at a GC base pair maintains these two acceptors, but in addition is distinguished by a hydrogen bond donor.

proteins for sequence-specific recognition of DNA have been identified which include the zinc finger,[4] the leucine zipper[5] and the helix-turn-helix[6] motifs. Within these complexes, the basis for sequence-specific binding can usually be understood by the pattern of specific hydrogen bonds between the protein side chains and the nucleobases and phosphates of the DNA. However, the *de novo* design of proteins for recognition of designated sequences remains a formidable task, due to the structural diversity available to a single protein and limitations in predicting protein folding.

Triple-Helical Complexes. The report in 1987 by Moser and Dervan that an oligonucleotide can sequence-specifically bind in the major groove of double-helical DNA has generated much research activity.[6] Specificity within the triple-helical structure arises from T•AT and C^+•GC base triplets wherein a pyrimidine base on the third strand forms specific hydrogen bonds to the Hoogsteen face of purine bases of the Watson-Crick base pairs. The sensitivity to

single base mismatch interactions have been determined and suggested that unique sites of 15-16 base pairs could be targeted.[7] Indeed, pyrimidine oligonucleotides have been used to mediate single site cleavage of yeast and human chromosomal DNA.[8] Efforts to extend triple-helix formation to non-purine-rich sites involve the design of non-natural nucleosides for specific recognition of other base pairs[9] and alternate strand triple-helix formation by 3'-3' and 5'-5' linked pyrimidine oligonucleotides.[10] Recently, a new family of triple-helical complexes has been identified wherein a purine rich oligonucleotide binds in the major groove of DNA.[11] Specificity is derived from G•GC, A•AT and T•AT triplets. This motif has been used in tandem with the pyrimidine motif for recognition of sites containing all four base pairs by alternate strand triple helix formation.[12]

Small Molecules. In addition to proteins and oligonucleotides, there are many small molecules (MW < 1500 Daltons) which bind (and cleave) DNA with modest sequence-specificities. Examples include the groove binding ligands distamycin and netropsin (see below), the aureolic acids chromomycin, mithramycin and olivamycin[13] and the recent enediyne antibiotics neocarzinostatin, calicheamicin, esperamicin and dynemicin.[14] These ligands generally bind in the minor groove and/or intercalate between the base pairs of DNA. The ability to synthesize analogs of the natural products together with direct assays for DNA-binding make these ligands ideal systems for understanding molecular recognition of DNA. Herein are described studies aimed at defining chemical principles for recognition in the minor groove of DNA. The familiar natural products distamycin A and netropsin serve as the basis for this work.

Recognition of the Minor Groove of Double-Helical DNA

Hydrogen Bonding in the Minor Groove. The hydrogen bonding character of the minor groove is less distinct than that of the major groove (Figure 2). The AT base pair presents two hydrogen bond acceptors, the purine N3 and the pyrimidine O2 atoms. Due the the approximate two-fold symmetry of these two atoms, the minor groove edges of AT and TA base pairs are very similar. The GC base pair retains these two acceptors, but in addition presents a hydrogen bond donor, the 2-amino group from guanine. Because the hydrogen bond donor lies closer to the guanine-containing strand, GC and CG base pairs may be distinguishable in the minor groove.

Netropsin and Distamycin A. Netropsin (N) and distamycin A (D) are antibiotics isolated from *Streptomyces netropsis* and *Streptomyces distallicus,* respectively (Figure 3).[15] Netropsin is a di-(*N*-methylpyrrolecarboxamide) with an *N*-terminal guanidinium group and a *C*-terminal amidinium group. Distamycin is a tri-(*N*-methylpyrrolecarboxamide) with an *N*-terminal formamide group and a *C*-terminal amidinium group. The natural products exhibit antifungal, antimitotic and antiviral properties *in vivo.*[16] An extensive body of biophysical literature demonstrates that the peptides bind in the minor groove of A,T-rich DNA.[17-24]

Structure of complexes. X-ray[18] and NMR[19] studies of netropsin- and distamycin-DNA complexes reveal how sequence specificity is accomplished. The crescent-shaped di- and tri-peptides are bound in the middle of the minor groove of an A,T-rich sequence. The amide hydrogens of the *N*-methylpyrrolecarboxamides form bifurcated hydrogen bonds with adenine N3 and thymine O2 atoms on the floor of the minor groove (Figure 4). The pyrrole

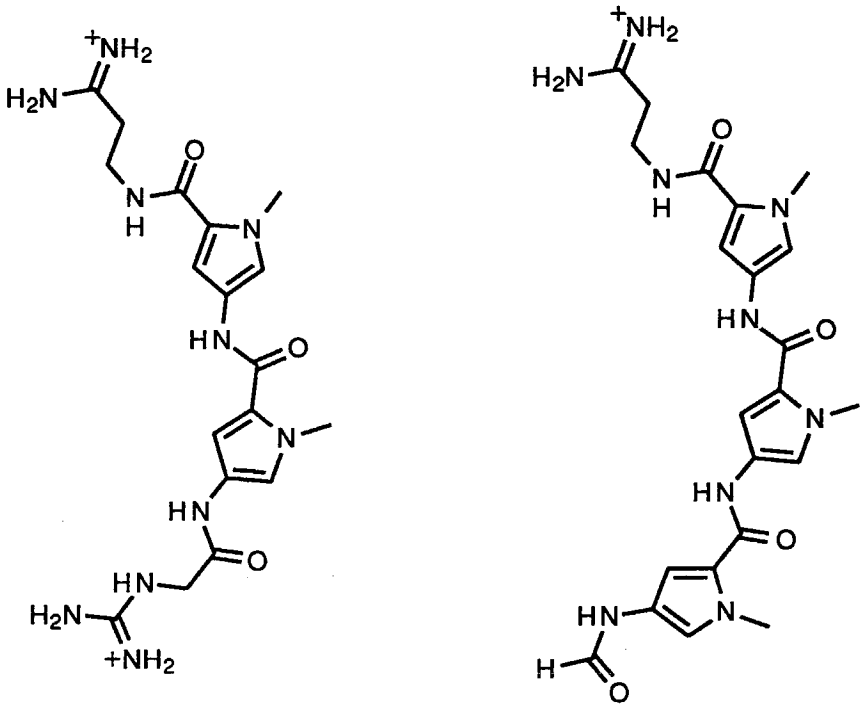


Figure 3. Natural products netropsin and distamycin A. The di- and tri-N-methylpyrrolecarboxamides bind in the minor groove of four and five base pairs of A,T-rich DNA, respectively.

rings fill the groove completely and form extensive van der Waals contacts with the walls of the minor groove (Figure 5). The aromatic hydrogens of the *N*-methylpyrrole rings are set too deeply in the minor groove to allow room for the guanine amino group of a G,C base pair, affording binding specificity for A,T sequences. The cationic amidinium and guanidinium groups of the peptides lie along the floor of the minor groove for favorable electrostatic interactions with the DNA.[20] Binding of the ligands to the minor groove of DNA induces conformational changes in both partners. The DNA bends slightly away from the minor groove and the pyrrolecarboxamides of the peptide twist to better match the groove.

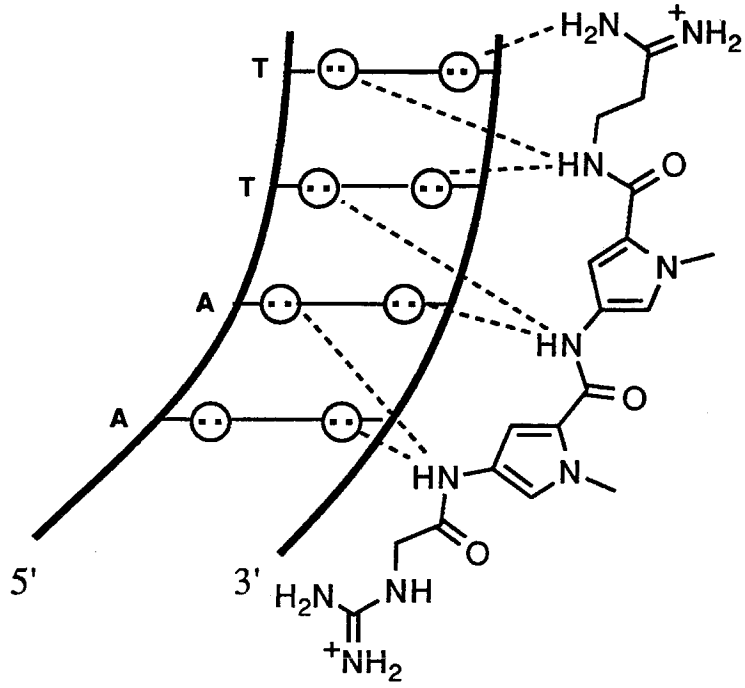
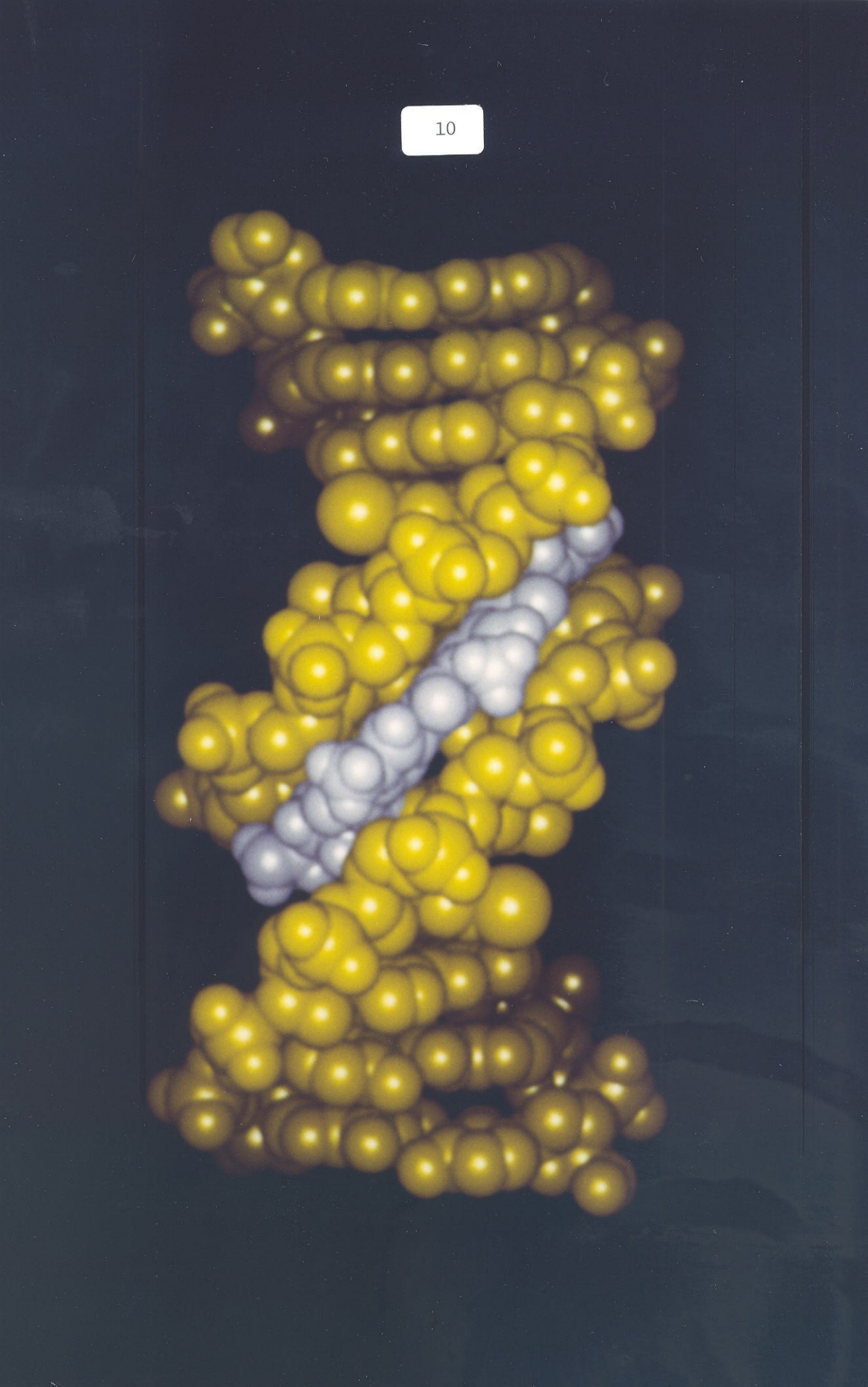


Figure 4. 1:1 Complex of netropsin with 5'-AATT-3'.[18a] Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Putative hydrogen bonds are illustrated by dashed lines.

Thermodynamics. Thermodynamic profiles reveal that for both netropsin and distamycin, the binding affinities at 25 °C are high and qualitatively similar ($\Delta G°$ = -12 kcal/mol) for complexation to poly [d(A-T)]•poly[d(A-T)] and poly(dA)•(dT).[21] However, for each peptide ligand, binding to the alternating copolymer duplex is enthalpy driven, whereas binding to the homopolymer duplex is entropy driven.[21a] To rationalize the large positive entropy changes in binding the homopolymer duplex, Breslauer and co-workers proposed a model that emphasizes binding induced perturbations of the more highly hydrated, altered B conformation of the homopolymer. This study underscores the importance of sequence-dependent DNA structure and solvent effects for understanding DNA binding processes.

Figure 5. Molecular model of the complex formed upon binding of the natural product netropsin in the minor groove of the 5′-AATT-3′ site. Coordinates are from the reported crystal structure by Dickerson and co-workers.[18a]



Sequence-Specific DNA Binding Assays.

Footprinting. Footprinting experiments identify the preferred equilibrium binding sites of ligands for double-helical DNA. In the footprinting experiment, a DNA-binding ligand is mixed with a radiolabeled DNA fragment of known sequence and subsequently treated with a DNA cleaving agent with little sequence-specificity. The footprinting agent can be either a natural enzyme, such as DNase I,[22] or a synthetic reagent such as methidiumpropyl-EDTA•Fe(II) (MPE•Fe(II)).[23] Separation of the cleavage products on a high resolution gel reveals regions of nucleotides which are protected from cleavage by the footprinting agent, thus identifying the binding site (Figure 6). Models have been developed for assigning binding sites to nucleotide resolution based on cleavage protection patterns for both strands using the reagent MPE•Fe(II) .[23c]

Affinity Cleaving. Affinity cleaving experiments rely on the attachment of a cleaving moiety to a sequence-specific DNA binding molecule to afford a sequence-specific DNA cleaving molecule.[24] This technique complements the footprinting experiment by providing information on the *orientation of ligand binding*. In an early example, the cleaving moiety EDTA•Fe(II) was attached to the *N*-terminus of distamycin.[25a] Cleavage of DNA with DE•Fe(II) reveals two unequal orientations of the bound ligand (Figure 7). Moreover, the asymmetry of the cleavage products provides direct information for the groove location of the ligand (Figure 8).[24b]

Oligopeptides for A,T Binding. The 1:1 peptide-DNA models have aided in the design and synthesis of molecules that bind larger sequences of pure A,T-rich double helical DNA. Increasing the number of *N*-methylpyrrolecarboxamides increases both the binding site size and the sequence specificity for longer tracts of contiguous A,T-base pairs of DNA.[25] Oligo (*N*-methylpyrrolecarboxamides) with incrementally increasing numbers of amino

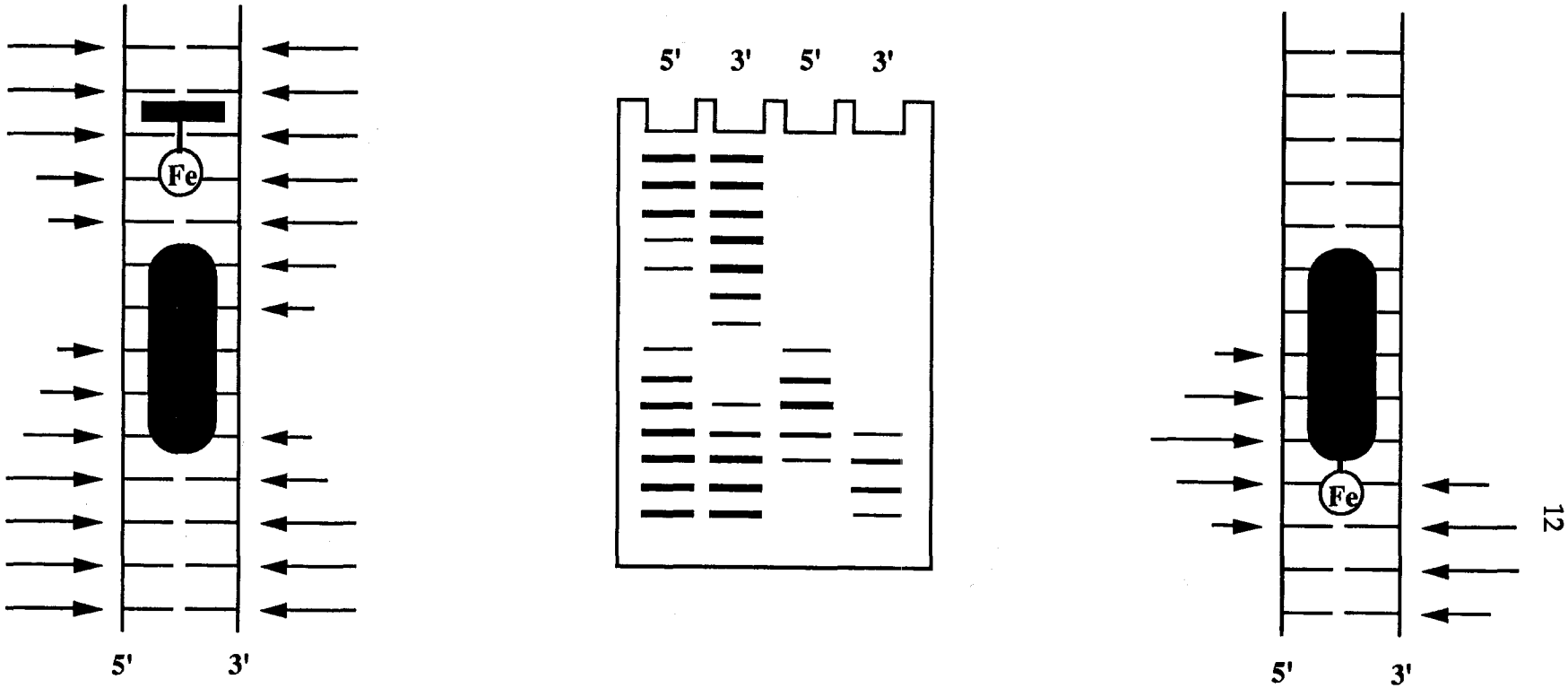


Figure 6. Footprinting experiments identify the equilibrium binding sites of ligands for DNA (Left). Affinity cleaving experiments identify the orientation of the ligand at the site (Right). In addition, the 3'-shifted cleavage pattern demonstrates binding of the ligand in the minor groove.

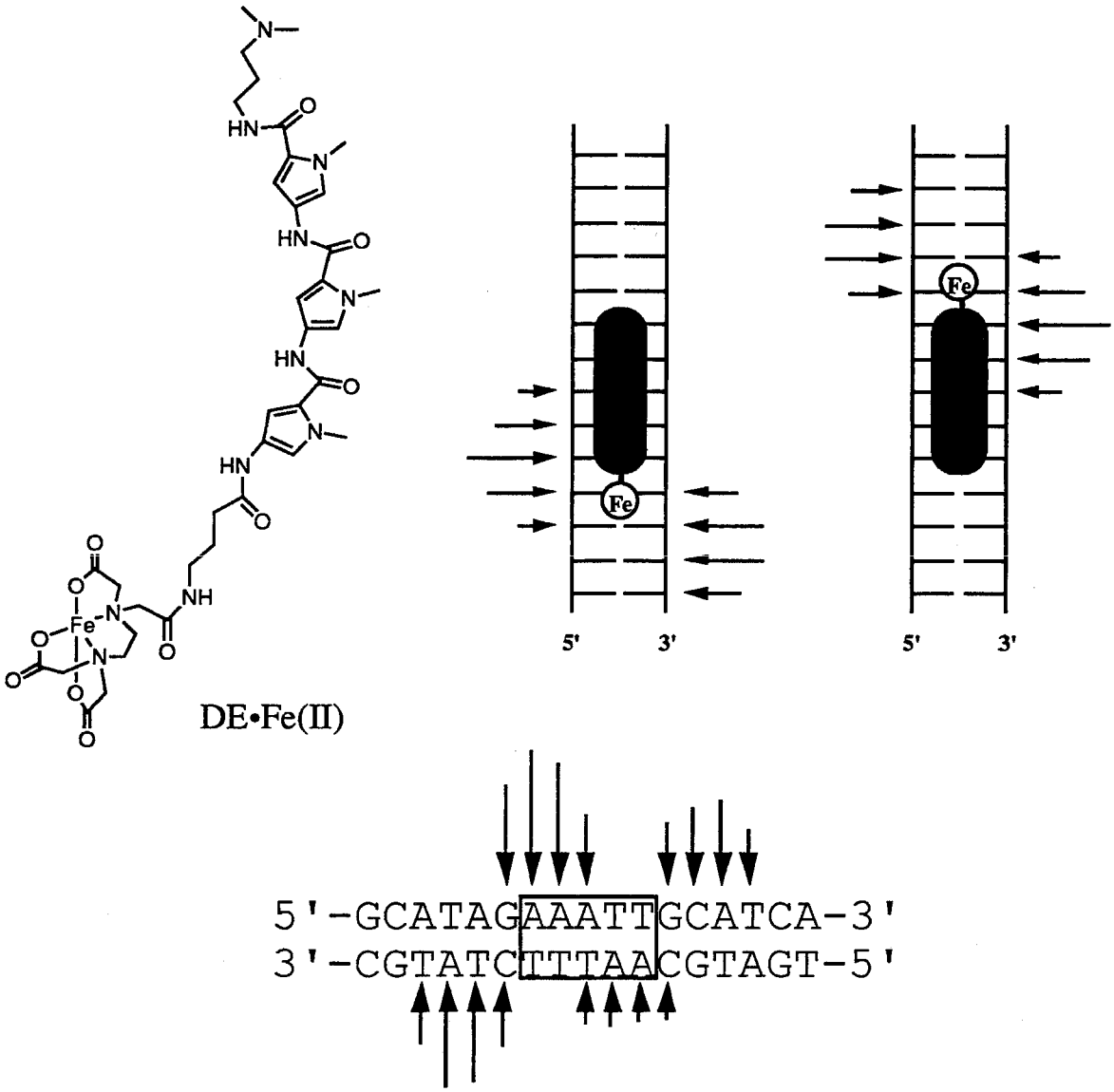


Figure 7. Affinity cleaving experiments with distamycin-EDTA•Fe(II). Analysis of the cleavage patterns produced on double-helical DNA by DE•Fe(II) identify two orientations in binding the 5′-AAATT-3′ site.

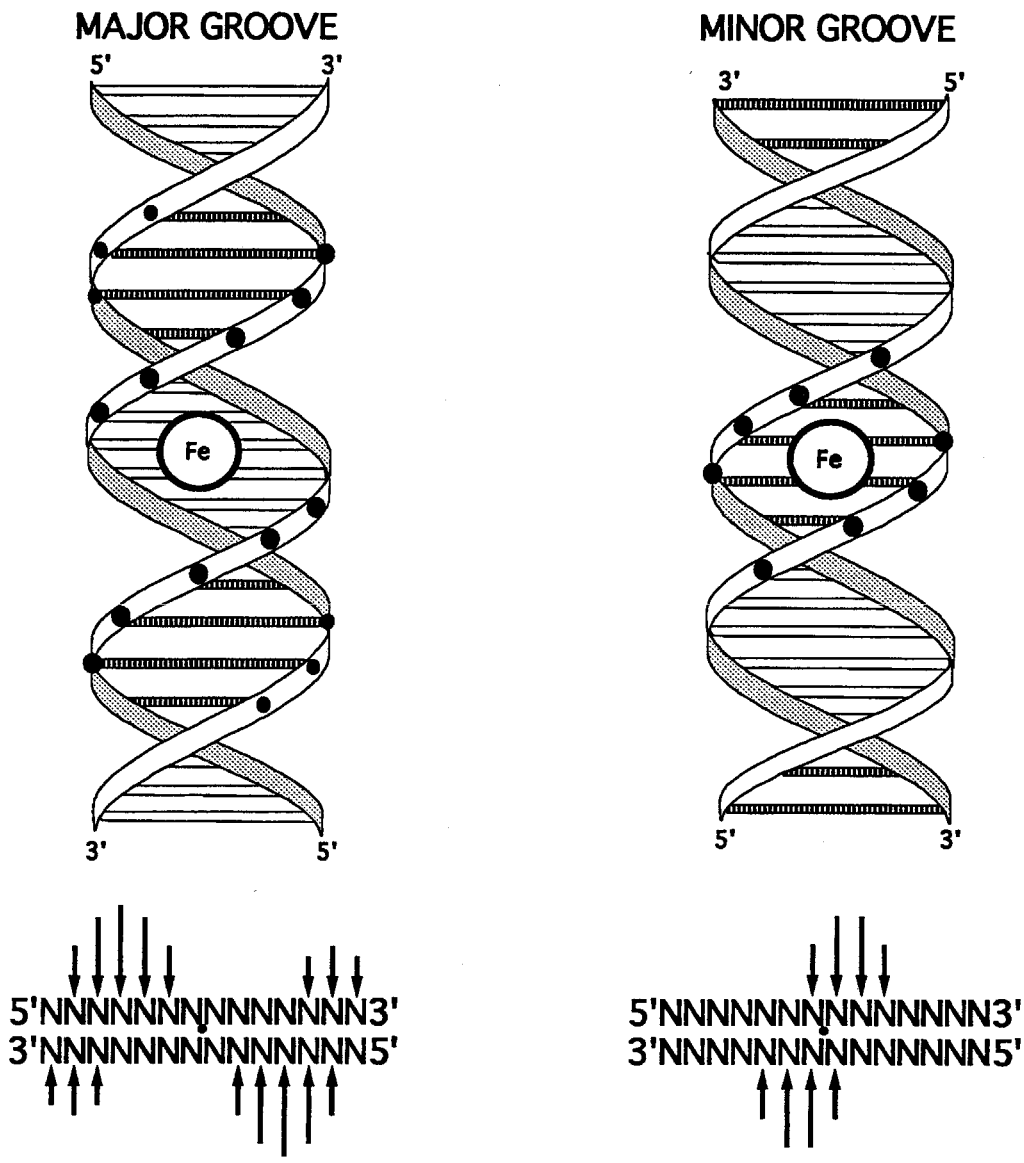


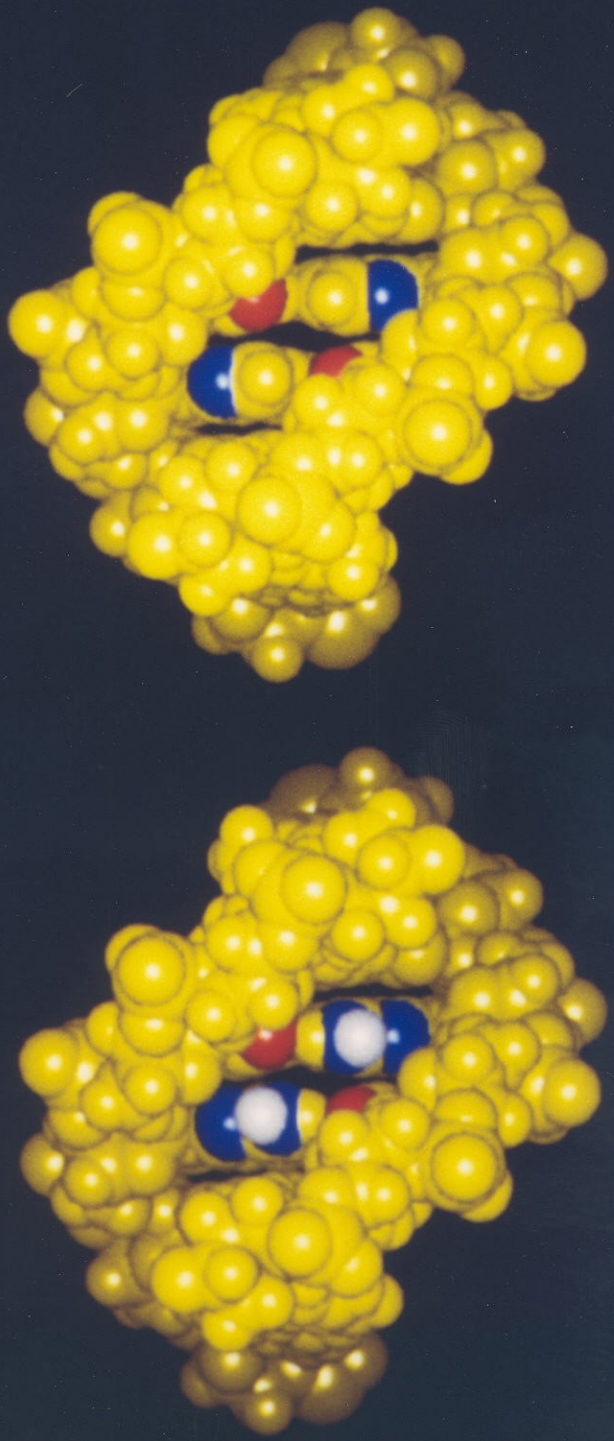
Figure 8. Models for the asymmetric DNA cleavage patterns generated by a diffusible oxidizing species in the major and minor grooves of right-handed DNA.

acid units from two to nine contain three to ten carboxamide NH's and are capable of binding four to eleven A,T base pairs, respectively.[25-27] Oligopeptides have achieved recognition of sequences as long as fifteen base pairs of A,T-rich DNA.[27d]

Peptides for Binding Sequences Containing G,C and A,T Base Pairs. Efforts to develop molecules which recognize sequences containing G,C base pairs with well-understood binding preferences have met with only limited success. Progress in this area is an important component in an overall strategy of coupling G,C units and A,T units into oligomers that uniquely identify long sequences of right-handed DNA in the minor groove. Early examples of hybrid molecules for mixed sequence recognition are the intercalator-groove binder bis(distamycin)phenoxazone (BEDP)[28] and the metalloregulated Ba2+•bis(netropsin)-3,6,9,12,15-pentaoxoheptadecanediamide,[29] both of which bind the ten base pair sequence 5'-TATAGGTTAA-3'. These hybrid molecules have similar functional architectures: a G,C binding unit flanked on both sides by an A,T-rich peptide binding unit. Although these can be considered encouraging benchmarks in the early development of the field of DNA recognition, they suffer from a lack of generalizable principles.

An alternative approach involves the systematic substitution of the tris (*N*-methylpyrrolecarboxamide) framework of distamycin to search for altered base pair specificity.[30-31] In the minor groove of DNA, the major difference between edges of A,T and G,C base pairs is the presence of the guanine 2-amino group protruding from the floor. In addition to introducing a hydrogen bond donor to the floor of the minor groove, the amino group is a sterically demanding group, presenting a bump on the minor groove floor (Figure 9). Replacement of the pyrrole CH with a heteroatom capable of forming a hydrogen bond to the

Figure 9. Hydrogen bonding partners on the floor of the minor groove. The AT base pair contains two hydrogen bond acceptors: the purine N3 and pyrimidine O2 atoms (top). The GC base pair also contains these two acceptors, but in addition presents a hydrogen bond donor from the 2-amino group of guanine (bottom). The presence of the guanine amino group also introduces a bump on the floor of the minor groove.



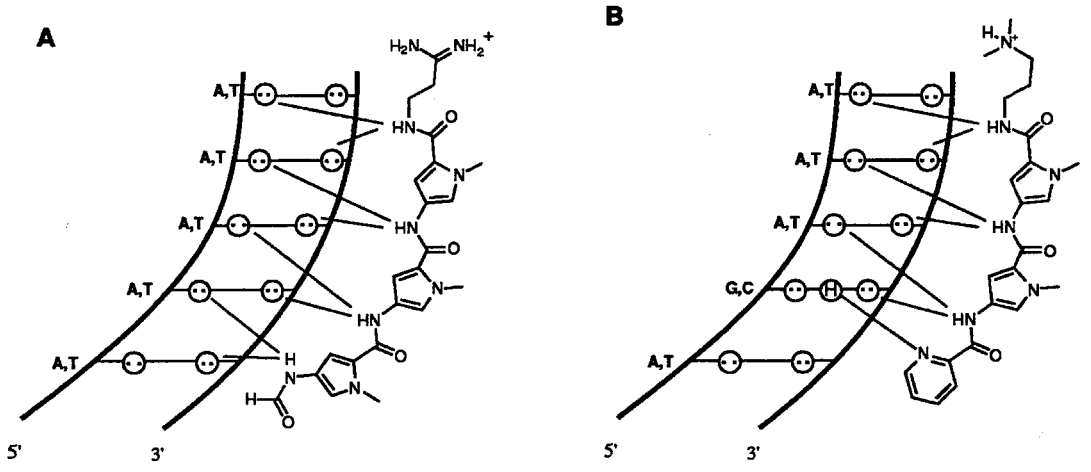
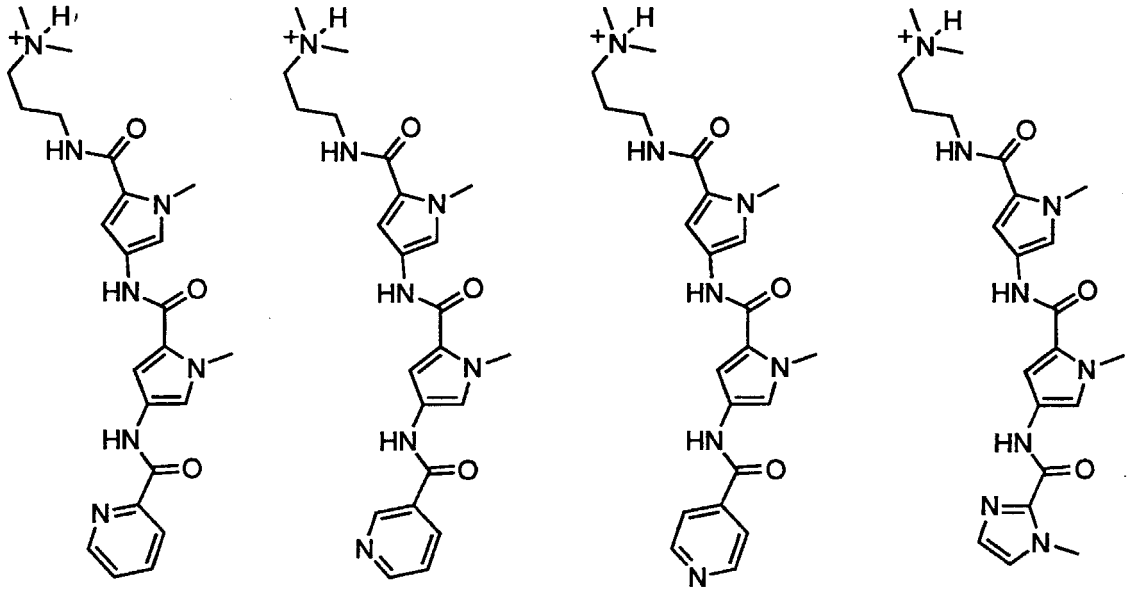


Figure 10. (A) 1:1 Model for binding of distamycin to A,T-rich sequences. (B) Proposed targeting of a G,C base pair by the designed peptide 2-PyN which contains a hydrogen bond acceptor on the pyridine ring.

guanine 2-amino group, for example, may afford peptides with specificity for G,C base pairs in their binding sites. Lown and co-workers synthesized three analogs of netropsin wherein either or both pyrrole rings were replaced with imidazole rings.[30a] Footprinting experiments reveal that the peptides display increased tolerance for G,C base pairs in their binding sites, but with little sequence-specificity. In addition, the peptides retained high affinity for A,T-rich sites. A second generation family of peptides containing a single cationic group was found to have improved specificity for G,C-rich sites, but still lacked a well-defined specificity.[30b] While the imidazole N3 atoms of the peptides likely participate in hydrogen bonds to guanine amino group on the floor of the minor groove, the peptides appear not to have the proper "match" for a unique set of DNA sites.

Design Rationale: The 1:1 Complex. The approach pursued by Wade and Dervan is illustrated in Figure 10, wherein it was expected that replacement of the terminal *N*-methylpyrrolecarboxamide of D with pyridine-2-carboxamide would afford an analog, pyridine-2-carboxamide-netropsin (2-PyN) (Figure 11),



2-PyN **3-PyN** **4-PyN** **2-ImN**

Figure 11. Designed peptides pyridine-2-carboxamide-netropsin (2-PyN), pyridine-3-carboxamide-netropsin (3-PyN), pyridine-4-carboxamide-netropsin (4-PyN) and 1-methylimidazole-2-carboxamide-netropsin (2-ImN).

that would specifically bind a site containing one G,C base pair followed by three A,T base pairs.[31] In analogy with distamycin binding A,T-rich sites, it was anticipated that the amide hydrogens would form bifurcated hydrogen bonds with adenine N3 and thymine O2 atoms, while the pyridine nitrogen would participate in a key hydrogen bond with the 2-amino group of the G,C base pair (Figure 10).

Footprinting experiments reveal that 2-PyN binds three A,T-rich sites five base pairs in size, 5'-TTTTT-3', 5'-AATAA-3', 5'-CTTTT-3' and an unanticipated mixed sequence 5'-TGTCA-3'.[32-22] Affinity cleaving experiments revealed no orientation preference for binding to the 5'-TGTCA-3' sequence. In contrast, the peptides, pyridine-3-carboxamide-netropsin (3-PyN) and pyridine-4-carboxamide-netropsin (4-PyN) preferred binding to A,T-rich sequences,

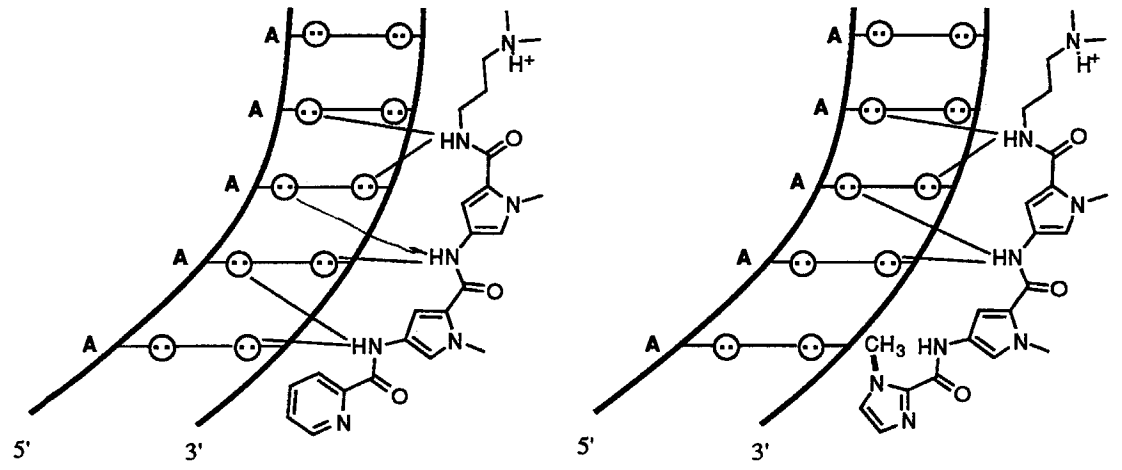


Figure 12. Strategy for improving the sequence-specificity of 2-PyN. The methylimidazole of the second generation peptide 2-ImN should disfavor binding to A,T-rich sites.

suggesting that the placement of the heterocyclic nitrogen is important for the recognition of the sequence 5'-TGTCA-3'.[32]

Binding of 2-PyN to the 5'-CTTTT-3' site is consistent with the design rationale model (Figure 10). 2-PyN cleaves this site with a strong preference for the orientation which places the heterocyclic nitrogen near the G,C base pair. The observation that the distamycin analog ED binds in the opposite orientation and that 3-PyNE and 4-PyNE do not have a strong orientation preference at this site supports the 1:1 model. It is clear, however, that 5'-CTTTT-3' is not the highest affinity binding site for 2-PyN.

Binding to A,T-rich Sites. Binding of 2-PyN to the A,T-rich sites could be explained with a 1:1 peptide-DNA model wherein the pyridine ring of 2-PyN is rotated such that the heterocyclic nitrogen faces away from the DNA (Figure 12). The lack of a hydrogen bond acceptor on the convex face of the peptide is consistent with binding A,T-rich sites. In an effort to improve the specificity of the peptide for the 5′-TGTCA-3′ sequence, the peptide 1-methylimidazole-2-carboxamide-netropsin (2-ImN) was synthesized. In the conformation shown in

Figure 11, 2-ImN displays a similar pattern of hydrogen bond donors and acceptors as 2-PyN and was expected to retain specificity for the mixed sequence. Rotation of the imidazole carboxamide, however, would direct the bulky methyl group towards the floor of the minor groove, and likely disfavor binding at A,T-rich sites.[33] Indeed, footprinting and affinity cleaving experiments reveal that 2-ImN binds specifically the 5'-TGTCA-3' site, with much reduced affinity for A,T-tracts of DNA. Affinity cleaving experiments again demonstrate two equal orientations in binding the 5'-TGTCA-3' site.[32]

Sequence Specificity. For the restriction fragment studied here, only a small fraction of the 512 possible five-base pair sites are present. To be confident that the observed strong binding sites are representative of all binding sites, double strand cleavage reactions were performed on DNA 4363 base pairs in length. Cleavage of *Sty* I linearized pBR322 by ED, 2-PyNE and 2-ImNE reveals markedly different cleavage patterns for ED and 2-ImNE, while the 2-PyNE lane appears to be a combination of the ED and 2-ImNE specificities.[32-33] Sequence analysis of the plasmid reveals 27 resolvable 5'-WGWCW-3' sites (W = A or T)[34] and, remarkably, all are within 35 base pairs of observed 2-ImNE cleavage sites. High resolution analysis of eight of these sites confirms binding to a 5'-WGWCW-3' sequence with two equivalent orientations.[33] With regard to binding to the 5'-WGWCW-3' sites, A,T base pairs on either side of the sequence generally results in higher binding affinities.[33]

Binding to the 5'-TGTCA-3' Site. A model for recognition of the 5'-TGTCA-3' site must account for recognition of two G,C base pairs by peptides containing a single imidazole N3 hydrogen bond acceptor. In addition, the model should explain the apparent lack of an orientation bias in binding 5'-(A,T)G(A,T)C(A,T)-3' sites. The early model for recognition of this site invoked a 1:1 complex of peptide in the minor groove and required rotation of the C-

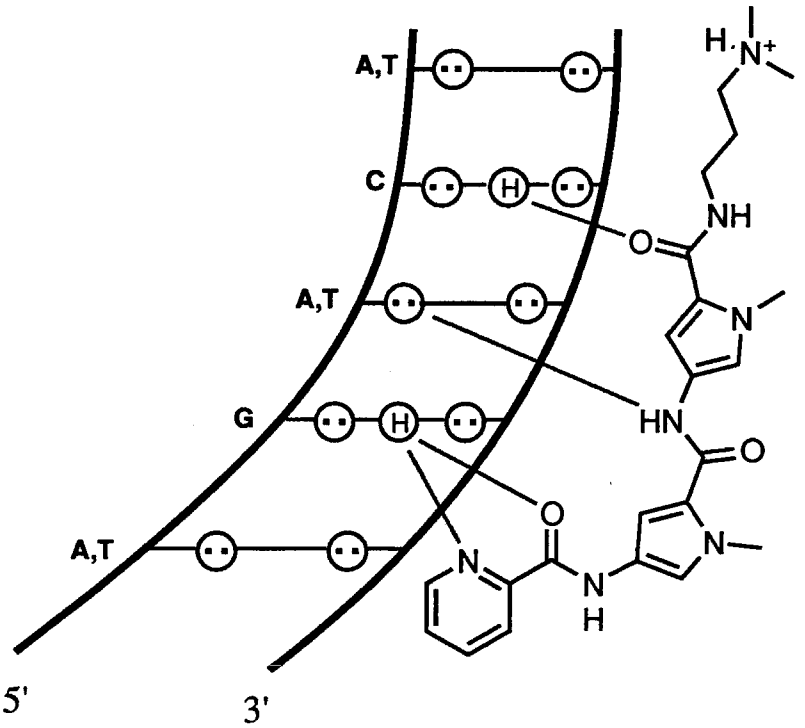


Figure 13. Early model for binding of 2-PyN to the 5′-TGTCA-3′ sequence. Recognition of a second guanine amino group was effected by hydrogen bonding to a carbonyl of the peptide.[31]

terminal carboxamide inward for hydrogen bonding to the second guanine amino group (Figure 13). However, such a model does not satisfactorily account for the unbiased orientation preference in binding this site. Moreover, this conformation of peptide does not match the shape of the minor groove as well.

Description of This Work. A novel side-by-side antiparallel motif for recognition in the minor groove of DNA is presented in Chapter 2. Such a 2:1 model could adequately explain the binding of 2-ImN and 2-PyN to the 5′-TGTCA-3′ site in two equal orientations. Indeed, direct characterization of the 2-PyN•5'-TGACT-3' and 2-ImN•5'-TGACT-3' complexes by two-dimensional NMR reveals that the peptides bind in the minor groove as side-by-side dimers with positive cooperativity.[35] In Chapter 3, the design and synthesis of covalent peptide dimers wherein the nitrogens of the central pyrrole rings of 2-PyN are

connected with alkyl linkers is described.[36] These linked peptides bind the 5′-TGTCA-3′ with 10-fold higher affinity relative to non-linked peptides. The design and synthesis of oligo(ethylene oxide)-linked peptides for metalloregulated binding is also described. The utility of 2:1 models for targeting new sequences is illustrated in Chapter 4 wherein the two peptides D and 2-ImN are shown to bind in the minor groove of the 5′-TGTTA-3′ site as a side-by-side dimer.[37] The affinity and sequence-specificity of these ligands for the 5′-TGTTA-3′ site was subsequently improved by covalently linking the two peptides.[38] Chapter 5 describes the design and synthesis of a new class of hairpin peptide-turn-peptide ligands for specific recognition of the 5′-TGTTA-3′ site. Significantly, the synthetic methodology for preparation of linear peptide hairpins is general and easily extended to the preparation of a variety of peptide dimers. Finally, Chapter 6 describes the design of a peptide for recognition of the 5′-GCGC-3′ sequence. This example underscores the utility of 2:1 peptide-DNA models for the design of ligands for sequence-specific recognition of designated DNA sites. Binding of this designed peptide to a pure four base pair G,C-core sequence represents an absolute reversal of the specificity of the natural products distamycin and netropsin.

References

1. For an excellent source of DNA structural data, see: Saenger, W. In Principles of Nucleic Acid Structure; Springer-Verlag: New York, 1984.
2. Harrison, S. C. *Nature* **1991**, *353*, 715-719.
3. Pavletich, N. P.; Pabo, C. O. *Science* **1991**, *252*, 809.
4. Ellenberger, T. E.; Brandl, C. J.; Struhl, K.; Harrison, S. C. *Cell* **1992**, *71*, 1223.
5. Pabo, C. O.; Sauer, R. T. *Ann. Rev. Biochem.* **1984**, *53*, 293.
6. Moser, H. E.; Dervan, P. B. *Science* **1987**, *238*, 645.
7. (a) Singleton, S. F.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 6957. (b) Singleton, S. F.; Dervan, P. B. *Biochemistry* **1992**, *31*, 10995.
8. (a) Povsic, T. J.; Strobel, S. A.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 5934. (b) Strobel, S. A.; Doucette-Stamm, L. A.; Riba, L.; Housman, D. E.; Dervan, P. B. *Science* **1991**, *254*, 1639.
9. (a) Koh, J. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 1470. (b) Griffin, L. C.; Kiessling, L. L.; Beal, P. A.; Gillespie, P.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 7976 (c) Stiltz, U.; Dervan, P. B. *Biochemistry* **1993**, *32*, 2177.
10. (a) Horne, D. A.; Dervan, P. B. *J. Am. Chem. Soc.* **1990**, *112*, 2435. (b) Ono, A.; Chen, C.-N.; Kan, L.-S. *Biochemistry* **1991**, *30*, 9914.
11. (a) Beal, P. A.; Dervan, P. B. *Science* **1991**, *251*, 1360. (b) Beal, P. A.; Dervan, P. B. *Nucleic Acids Res.* **1992**, *20*, 2773.
12. Beal, P. A.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 4976.

13. (a) Van Dyke, M. W.; Dervan, P. B. *Biochemistry* **1983**, *22*, 2373-2377. (b) Gao, X.; Patel, D. J. *Biochemistry* **1989**, *28*, 751-762. (c) Gao, X.; Mirau, P.; Patel, D. J. *J. Mol. Biol.* **1992**, *223*, 259-279.

14. For a review, see: Nicolaou, K. C. *Proc. Natl. Acad. Sci., USA* **1993**, *90*, 5881-5888.

15. (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.

16. (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, Ch. *Progress in Nucleic Acids Research and Molecular Biology* **1980**, *15*, 258. (c) Krey, A. *Prog. in Mol. Subcell. Biol.* **1980**, *7*, 43.

17. (b) Krylov, A. S.; Grokhovsky, S. L.; Zasedatelev, A. S.; Zhuze, A. L.; Gursky, G. V.; Gottikh, B. P. *Nuc. Acids. Res.* **1979**, *6*, 289-304. (c) Zasedatelev, A. S.; Gursky, G. V.; Zimmer, Ch.; Thrum, H. *Mol. Biol. Reports* **1974**, *1*, 337-342. (d) Zasedatelev, A. S.; Zhuze, A. L.; Zimmer, Ch.; Grokhovsky, S. L.; Tumanyan, V. G.; Gursky, G. V.; Gottikh, B. P. *Dokl. Acad. Nauk SSSR* **1976**, *231*, 1006-1009. For a review, see: (e) Zimmer, C.; Wåhnert, U. *Prog. Biophys. Molec. Biol.* **1986**, *47*, 31-112.

18. (a) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 1376-1380. (b) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *J. Mol. Biol.* **1985**, *183*, 553-563. (c) Coll, M.; Frederick, C. A.; Wang, A. H.-J.; Rich, A. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 8385-8389.

19. (a) Patel, D. J.; Shapiro, L. *J. Biol. Chem.* **1986**, *261*, 1230-1240. (b) Klevitt, R. E.; Wemmer, D. E.; Reid, B. R. *Biochemistry* **1986**, *25*, 3296-3303. (c) Pelton, J. G., Wemmer, D. E. *Biochemistry* **1988**, *27*, 8088-8096.

20. Zakrzewska, K.; Lavery, R.; Pullman, B. *J. Biomol. Struc. Dyn.* **1987**, *4*, 883-843.

21. (a) Markey, L. A.; Breslauer, K. J. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 4359-4363. (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*, 8922-8926.

22. (a) Fox, K. R.; Waring, M. J. *Nucl. Acids Res.* **1984**, *12*, 9271-9285. (b) Lane, M. J.; Dobrowiak, J. C.; Vournakis, J. *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 3260-3264.

23. (a) Hertzberg, R. P.; Dervan, P. B. *J. Am. Chem. Soc.* **1982**, *104*, 313. (b) Van Dyke, M. W.; Hertzberg, R. P.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 5470-5474. (c) Van Dyke, M. W.; Dervan, P. B. *Cold Spring Harbor Symposium on Quantitative Biology* **1982**, *47*, 347-353. (d) Van Dyke, M. W.; Dervan, P. B. *Biochemistry* **1983**, *22*, 2373-2377. (e) Harshman, K. D.; Dervan, P. B. *Nucl. Acids Res.* **1985**, *13*, 4825-4835.

24. (a) Schultz, P. G.; Taylor, J. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1982**, *104*, 6861-6863. (b) Taylor, J. S.; Schultz, P. G.; Dervan, P. B. *Tetrahedron* **1984**, *40*, 457-465. (c) Schultz, P. G.; Dervan, P. B. *J. Biomol. Struct. Dyn.* **1984**, *1*, 1133-1147. (d) Dervan, P. B. *Science*, **1986**, *232*, 464-471.

25. (a) Schultz, P. G.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 6834-6837. (b) Youngquist, R. S.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 2565-2569.

26. (a) Gursky, G. V.; Zasedatelev, A. S.; Zhuze, A. L.; Khorlin, A. A.; Grokhovsky, S. L.; Streltsov, S. A.; Surovaya, A. N.; Nikitn, S. M.; Krylov,

A. S.; Retchinsky, V. O.; Mikhaillov, M. V.; Beabealashvili, R. S.; Gottihk, B. P. *Cold Spring Harbor Symp. Quant. Biol.* **1982**, *47*, 367. (b) Skamrov, A. V.; Rybalkin, I. N.; Bibilashvili, R. Sh.; Gottikh, B. P.; Grokhovskii, S. L.; Gurskii, G. V.; Zhuze, A. L.; Zasedatelev, A. S.; Nechipurenko, Yu. D.; Khorlin, A. A. *Mol. Biol.* **1985**, *19*, 153.

27. (a) Schultz, P. G.; Dervan, P. B. *J. Am. Chem. Soc.* **1983**, *105*, 7748-7750. (b) Youngquist, R. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1985**, *107*, 5528-5529. (c) Griffin, J. H.; Dervan, P. B. *J. Am. Chem. Soc.* **1986**, *108*, 5008-5009. (d) Youngquist, R. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1987**, *109*, 7564-7566.

28. Dervan, P. B.; Sluka, J. P. *New Synthetic Methodology and Functionally Interesting Compounds*; Elsevier: New York, 1986; pp 307-322.

29. Griffin, J. H.; Dervan, P. B. *J. Am. Chem. Soc.* **1987**, *109*, 6840-6842.

30. (a) Lown, J. W.; Krowicki, K.; Bhat, U. G.; Ward, B.; Dabrowiak, J. C. *Biochemistry* **1986**, *25*, 7408-7416. (b) Kissinger, K.; Krowicki, K.; Dabrowiak, J. C.; Lown, J. W. *Biochemistry* **1987**, *26*, 5590-5595. (c) Lee, M.; Chang, D. K.; Hartley, J. A.; Pon, R. T.; Krowicki, K.; Lown, J. W. *Biochemistry* **1988**, *27*, 445-455. (d) Rao, K. E.; Bathini, Y.; Lown, J. W. *J. Org. Chem.* **1990**, *55*, 728-737. (e) Plouvier, B.; Bailly, C.; Houssin, R.; Rao, K. E.; Lown, J. W.; Henichar, J.-P.; Waring, M. J. *Nucleic Acids Res.* **1991**, *19*, 5821-5829.

31. Wade, W. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1987**, *109*, 1574-1575.

32. Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 8783-8794.

33. Wade, W. S. Ph.D. Thesis, California Institute of Technology, 1989.

34. Cornish-Bowden, A. *Nucleic Acids Res.* **1985**, *13*, 3021-3030.

35. Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 7586-7590.

36. (a) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 9892-9899. (b) Dwyer, T. J.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, *115*, 9900-9906.

37. (a) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 2572-2576. (b) Geierstanger, B. H.; Jacobsen, J-P.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Biochemistry*, in press.

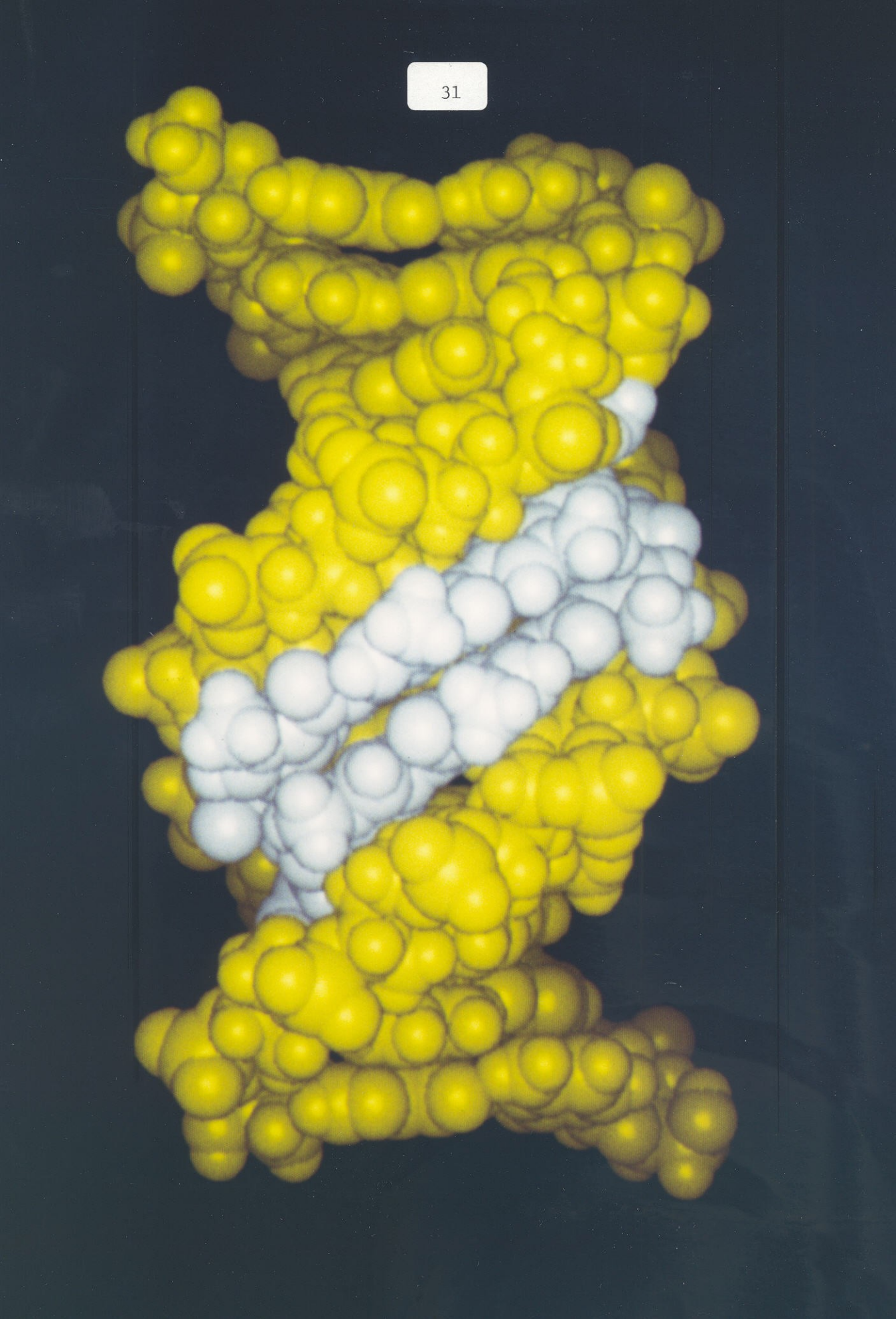
38. Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, in press.

Chapter 2

Antiparallel Side-By-Side Dimeric Motif for Sequence-Specific Recognition in the Minor Groove of DNA

2:1 Peptide-DNA Complex. Pelton and Wemmer have studied the binding of distamycin to the 5′-AAATT-3′ site in the d(CGCAAATTGGC)•d(GCCAATTTGCG) oligonucleotide duplex by two-dimensional NMR.[1] At low peptide-DNA ratios, a mixture of two 1:1 complexes were observed, representing the two possible orientations of peptide in the minor groove.[2] Remarkably, as the peptide concentration was increased, a new binding mode was observed which became the predominant complex at a peptide-DNA ratio of 2:1. Characterization of this complex revealed that the distamycin ligands were bound as an *antiparallel side-by-side dimer in the minor groove* of the 5′-AAATT-3′ binding site (Figure 1). Such a complex requires the minor groove of the A,T-rich region to widen to approximately 7 Å. As with previous 1:1 complexes, the two peptides completely fill the minor groove, maintaining the favorable van der Waals contacts between peptides and DNA. The two stacked ligands are staggered such that the amides of one peptide overlap the pyrrole rings of the other peptide. This model differs significantly from 1:1 models in that each peptide forms hydrogen bonds to bases on distinct strands of the DNA in the minor groove (Figure 2).

Figure 1. Molecular model for binding of the side-by-side distamycin peptides in the minor groove of the 5′-AAATT-3′ site.[1]



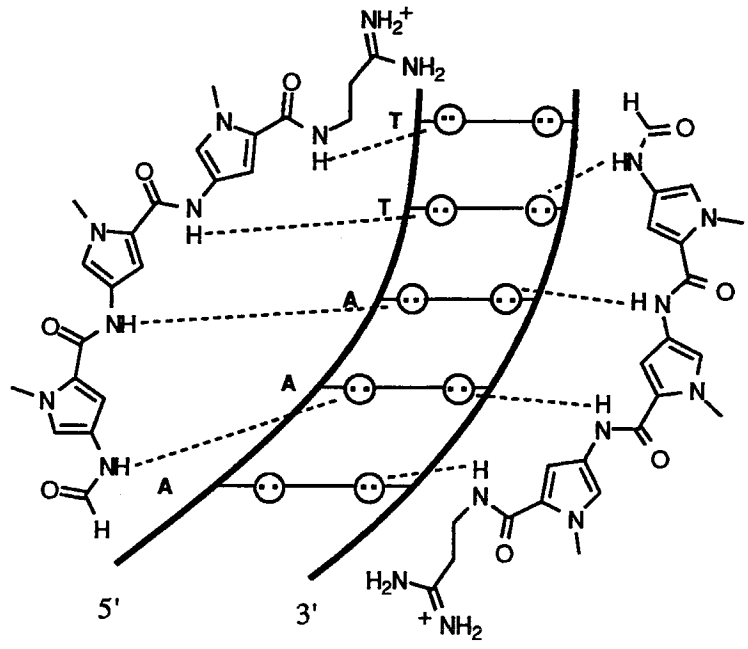


Figure 2. 2:1 Complex of distamycin with 5'-AAATT-3'.[1] Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Putative hydrogen bonds are illustrated by dashed lines.

The demonstration that the minor groove of A,T-base pair regions could widen to approximately 7 Å and accommodate two side-by-side peptdies contradicted current models which emphasized the tight fit of a single peptide in the minor groove. Based on the early 1:1 models, it was proposed that substitution of the pyrrole rings of the natural products with imidazole rings would afford a new class of "lexitropsins" capable of recognition of GC base pairs. The limited success with these ligands is likely due to the fact there is wide variation in minor groove width depending on sequence and wider sequences may inherently favor binding by peptide dimers. Importantly, the 2:1 model provided new insight in understanding recognition of the minor groove of DNA by peptide analogs of the natural products distamycin and netropsin.

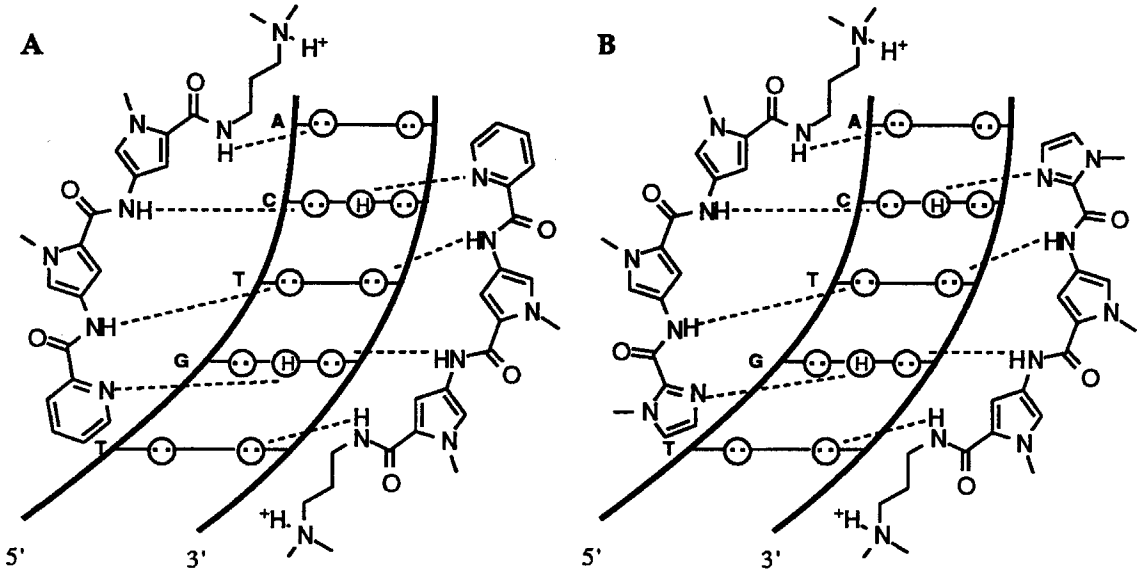


Figure 3. 2:1 Binding models for the complexes formed upon binding of (A) 2-PyN and (B) 2-ImN to the 5′-TGTCA-3′ site.

2-ImN and 2-PyN. Consideration of a 2:1 peptide-DNA model adequately explains the recognition of two G,C base pairs in the 5'-TGTCA-3' site by 2-ImN and 2-PyN (Figure 3).[3] Each guanine amino group could participate in specific hydrogen bonds with imidazole or pyridine nitrogen atoms on each of the side-by-side peptides (Figure 3). The peptide NH's participate in hydrogen bonds with adenine N3 and pyrimidine O2 atoms. Furthermore, the $(2\text{-ImNE})_2 \bullet$5'-TGTCA-3' and $(2\text{-PyNE})_2 \bullet$5'-TGTCA-3' complexes would position a cleaving moiety on each side of the binding site, consistent with the affinity cleaving experiments which demonstrate an unbiased orientation preference for the peptides binding this site. The lower affinity of these peptides for the 5'-TGTCA-3' site may in part be due to the higher entropy cost for binding of two ligands in the minor groove.

Structural Characterization of 2-ImN and 2-PyN in Complex with the 5′-TGACT-3′ Site

In collaboration with the Wemmer group, we have characterized the complexes formed upon binding of 2-PyN and 2-ImN to the oligonucleotide duplex d(GCATGACTCGG)•d(CCGAGTCATGC) by two-dimensional NMR spectroscopy. The peptides do indeed bind as antiparallel, side-by-side homodimers in the minor groove of DNA (Figure 4). A summary of the study is presented below: experimental details can be found in published reports.[4,5]

Titration of the duplex d(GCATGACTCGG)•d(CCGAGTCATGC) with either 2-ImN or 2-PyN yields only one set of new resonances in the 1D NMR spectrum. Analysis of 2D NMR spectra reveal that these resonances are associated with a 2:1 ligand-DNA complex. Thus, even at low ligand:DNA stoichiometries (e.g., 0.25:1) dimeric binding is dominant, indicating that the two ligands bind with high cooperativity. In contrast, studies of the complexes formed between distamycin and the 5'-AAATT-3' and 5'-AAATTT-3' duplexes do not show evidence for the 2:1 binding mode until 0.5 to 0.75 equivalents of ligand have been added.[6] Replacement of A,T base pairs with inosine-cytosine (I,C) base pairs, which present the same functional groups on their minor groove surfaces, affords highly cooperative dimeric binding of distamycin to the sequence d(CGCIIICCGGC)•d(GCCIICCCGCG).[7] If G,C- or I,C-bearing DNA sequences have inherently wider minor grooves, formation of 2:1 complexes on ligand binding may be favored as a result of a reduction in the energy cost to accommodate the second ligand. Additionally, single ligand binding modes for such sequences would be disfavored if their grooves were not capable of

Figure 4. Van der Waals surface representation of the 2:1 2-ImN • d(GCATGACTCGG) • d(CCGAGTCATGC) complex obtained by energy refinement using semiquantitative distance constraints derived from NOESY.

clamping down sufficiently on one ligand to maximize the van der Waals contacts necessary for binding.

The stability of the 2-ImN • DNA complex is determined by several factors. The amide NH protons contribute favorably to the binding energy through hydrogen bond formation to thymine O2 and adenine N3 atoms. Pyrrole H3 protons and the imidazole H4 proton form close van der Waals contacts with sugar C1'H and adenine C2H protons of the DNA. The positively charged *N,N*-dimethylammonium group facilitates the attraction of the ligand to DNA. Stacking of the conjugated ring systems of the ligands in the 2:1 complex likely provides additional binding energy. Maximizing the contacts that provide the binding energy may require some distortion of the duplex and the concomitant energetic penalty. From the NMR studies presented here and chemical protection experiments,[8] there is no strong evidence for major distortions of the DNA helix or conformation as a result of 2-ImN binding to 5'-TGACT-3'.

The most striking feature of the $(2\text{-ImN})_2$•5′-TGACT-3′ complex is the specific recognition of both G,C and A,T base pairs. Molecular modeling of the 2:1 complex suggests that the imidazole nitrogen (N3) proximal to the 2-amino group of guanine affords a hydrogen bond that contributes to the stability of the complex. The energetic price for this recognition may require that the crescent shaped 2-ImN does not sit deeply in the minor groove of the sequence 5'-TGACT-3'. Further speculation about this novel structure must await high resolution data from x-ray crystallography.

Binding Affinities of 2-ImN and 2-PyN in Complex with the 5′-TGTCA-3′ Site

Quantitative Footprint Titrations. The binding affinities of ligands for discreet site on DNA can be measured by quantitative footprint titration experiments. This technique has been well developed by Ackers and co-workers for measurement of binding affinities of proteins and small molecules.[9] The methodology has been adapted by Wade and Dervan for measurement of binding affinities of peptide analogs to discreet five base pair sites using the high resolution footprinting reagent MPE•Fe(II).[8,10] The quantitative footprint titration experiments can, in some cases, provide information on the stoichiometry of the binding process. A brief review of the methodology will be presented, followed by an analysis of the binding isotherms with particular relevance to binding models.

In the quantitative footprint titration experiment, footprinting reactions with a single peptide at many different concentrations are carried out. Typically, fifteen different concentrations of peptide in the range of 100 μM to 10 nM are employed. The products of these reactions, along with those from a control footprinting lane to which no peptide has been added, are separated on a high resolution polyacrylamide sequencing gel (Figure 5). The gel is imaged using phosphor storage technology and the data is analyzed by performing volume integrations of the nucleotide fragments in the binding site and a reference site which is not bound by the peptides. Qualitatively, the binding sites are completely protected from the footprinting reagent at high concentrations of peptides, not protected at low concentrations of peptides and partially protected at intermediate concentrations of peptide. The fractional occupation of the site is

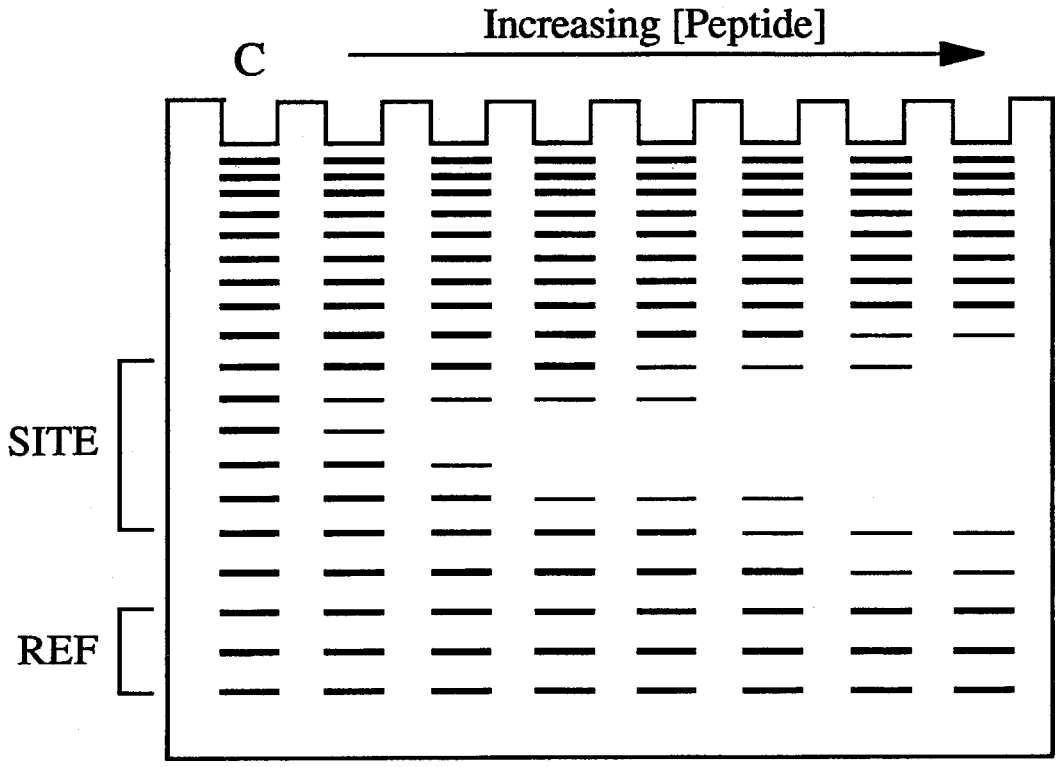


Figure 5. Representation of the quantitative footprint titration experiment. A series of footprinting reactions are performed with increasing concentration of ligand. Quantitation of the binding site and a reference site to which the ligand does not bind allow the equilibrium association constant of the peptide for the DNA site to be determined.

referred to as θ_{app} and is calculated by comparing the protection at the target site to that of the reference site and normalizing to the standard lane using the following equation:

$$\theta_{app} = 1 - \frac{Isite/Iref}{Isite°/Iref°} \qquad (1)$$

where I_{site} and I_{ref} are the integrated volumes of the target and reference sites, respectively, and $I_{site°}$ and $I_{ref°}$ correspond to those values for a footprint control lane to which no peptide has been added. The ([L], θ_{app}) data points are then fit

to a Langmuir binding isotherm (eq 2) by minimizing the difference between θ_{app} and θ_{fit}.

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

where [L] corresponds to the total peptide concentration, K_a corresponds to the apparent monomeric association constant, and θ_{min} and θ_{max} represent the experimentally determined site saturation values when the site is unoccupied or saturated, respectively. A sample data set and binding isotherm are shown in Figure 6.

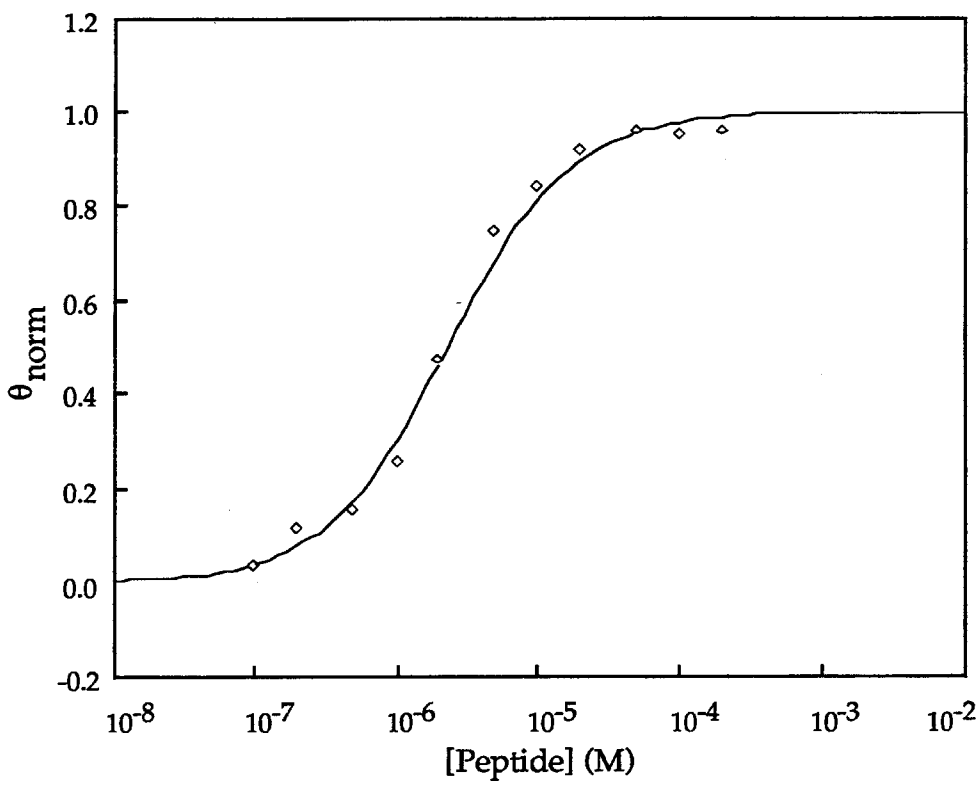


Figure 6. A representative ([L], θ_{app}) data set from the quantitative footprint titration experiment, with the fractional site occupancy reported on the vertical axis and log[L] on the horizontal axis. The curve is the best-fit Langmuir binding isotherm. The ligand concentration at half-occupancy of the site is equal to the first order dissociation constant.

Binding to the 5′-TGTCA-3′ Site. The best fit binding isotherms for 2-ImN and 2-PyN binding to the 5′-TGTCA-3′ site consistently give much worse fits than the other complexes. Visual inspection of the binding curves reveals that, in both cases, the increase in θ_{app} near half-saturation of the site is steeper than expected from the fitted curve, consistent with cooperative dimeric binding to this sequence by the peptides. While NMR and affinity cleaving experiments reveal that the peptides bind with positive cooperativity, the degree of cooperativity cannot be inferred from these experiments (Mrksich et al., 1992; Wade et al., 1992). Moreover, these experiments cannot distinguish between binding by pre-associated dimers and step-wise binding by two free peptides. It is likely that the footprinting data presented here also cannot distinguish between these possible mechanisms. In consideration of a pre-association binding mechanism, the ([*L*], θ_{app}) data points were fit to eq. 3:

$$\theta_{fit} = (\theta_{max} - \theta_{min}) \frac{K_B\left(1 + 4K_D[L] - \sqrt{1 + 8K_D[L]}\right)}{8K_D + K_B\left(1 + 4K_D[L] - \sqrt{1 + 8K_D[L]}\right)} + \theta_{min} \qquad (3)$$

where K_D corresponds to the dimerization constant for the peptides and K_B corresponds to the association constant of the peptide dimer with the DNA site. Nonlinear least-squares analysis with K_D, K_B, θ_{min} and θ_{max} as adjustable parameters resulted in good fits, but with large variability in the K_D and K_B parameters. Fitted values for K_D ranged from 10-1000 M^{-1} and those for K_B ranged from 10^7-10^9 M^{-1}. In consideration of a step-wise binding mechanism, the same data were fit to eq. 4:

$$\theta_{fit} = (\theta_{max} - \theta_{min}) \frac{K_1[L] + K_1K_2[L]^2}{1 + K_1[L] + K_1K_2[L]^2} + \theta_{min} \qquad (4)$$

where K_1 corresponds to the association constant for binding of the first peptide and K_2 to the association constant for binding of the second peptide to a singly occupied site. Nonlinear least-squares analysis using eq. 4 with K_1, K_2, θ_{min} and θ_{max} as adjustable parameters also produced fits of good quality. In the cases of both 2-PyN and 2-ImN binding to this sequence, we find $K_2>K_1$, with the K_2/K_1 ratios ranging from 10-300. Clearly, the data are not of sufficient precision to allow for accurate independent determination of the two step-wise binding constants. Therefore, we chose to fit the data to a cooperative binding curve (eq. 5) with K_a, θ_{min}, θ_{max} and n as adjustable parameters.

$$\theta_{fit} = (\theta_{max} - \theta_{min}) \frac{K_a{}^n [L]^n}{1 + K_a{}^n [L]^n} + \theta_{min} \qquad (5)$$

In fact, in the case of $K_2>>K_1$, eq. 4 is approximately equivalent to eq. 5 with $K_1K_2 \approx K_a{}^2$. We note explicitly that treatment of the data in this manner does not represent an attempt to model a binding mechanism. Rather, we have chosen to compare values of K_a, the apparent first order binding affinity, because this parameter represents the concentration of peptide at which the binding site is half-saturated. The binding affinity of 2-ImN for the 5′-TGTCA-3′ was also determined by quantitative DNase I footprint titration experiments with the same results (see chapter 4)

Binding to A,T-rich sites. D binds the three A,T-rich sites with apparent first order affinity constants ranging from 1×10^6 to 3×10^7 M^{-1}. 2-PyN binds these three sites with binding affinities nearly two orders of magnitude lower than does D (Table I). Strong binding of D to A,T-rich sites has been explained by a tight fit of the peptide in the minor groove and a complementary set of hydrogen bonds between the ligand and the floor of the minor groove. The

Table I. Apparent First Order Affinity Constants (M^{-1})[a,b]

	Binding Site			
Peptide	5'-TTTTT-3'	5'-TTAAT-3'	5'-TGTCA-3'	5'-AATAA-3'
D	2.6×10^7 (1.8)	1.9×10^6 (0.8)	$<1 \times 10^5$	1.4×10^7 (0.9)
2-PyN	2.3×10^5 (1.2)	6.9×10^4 (1.6)	2.7×10^5 (0.3)	5.4×10^4 (0.4)
2-ImN	$<5 \times 10^4$	$<1 \times 10^4$	1.4×10^5 (0.3)	$<2 \times 10^4$

[a]Determined from average θ_{app} values from three gels (standard deviation for three independent determinations from single gels).
[b]20 mM Tris•HCl, pH 7.0, 100 mM NaCl

amide NH's of D and 2-PyN can form an extended array of hydrogen bonds to the adenine N3 and thymine O2 atoms on the floor of the minor groove (Figures 2A and 2B). While the lower binding affinity of 2-PyN for A,T-rich sites can in part be explained by one less carboxamide NH for hydrogen bonding, the different geometries of pyridine and pyrrole rings may also be a contributing factor. 2-ImN does not bind these A,T-rich sites ($K_a < 10^5\ M^{-1}$). Unlike 2-PyN, 2-ImN is constrained to have a hydrogen bond acceptor atom facing the floor of the minor groove. The lack of a hydrogen bond donor for the peptide imidazole N3 atom and lone-pair-lone-pair charge repulsions presumably disfavor binding to A,T sequences and impart a large specificity for the mixed sequences.

Binding to the 5'-TGTCA-3' site. 2-PyN and 2-ImN bind to the 5'-TGTCA-3' sites with comparable affinities but nearly 100-fold lower than that of D binding to A,T-rich sites. Since 2-PyN and 2-ImN bind the mixed sequence as side-by-side dimers, the lower association constants may reflect an entropic penalty. Consistent with this explanation, we have recently shown that covalently linking two peptides with alkyl tethers affords covalent peptides which bind with 20-fold higher affinities than do the non-linked peptides.[11]

Lower binding constants for G,C-containing sequences may also be a general consequence of sequence dependent minor groove characteristics. If the guanine amino groups protruding from the floor of the minor groove do not allow the peptides to sit deeply in the minor groove, the peptides may bind with less favorable enthalpies and lower binding constants.[12] 2-PyN and 2-ImN can accommodate the amino groups of the 5′-TGTCA-3′ site by participation in favorable hydrogen bonds with the peptide pyridine N1 and imidazole N3 nitrogens. The shallow penetration of the peptides in the minor groove forfeits the full benefit of the tight fit between peptide and DNA, resulting in lower binding affinities as compared to D in complex with A,T-rich sites. D, however, can neither participate in hydrogen bonds with the guanine amino groups nor can the peptide be deeply set in the minor groove, resulting in low affinity for the mixed sequence 5′-TGTCA-3′.

Sequence Specificity for 5'-WGWCW-3' sites. 2-ImN and 2-PyN bind to 5'-WGWCW-3' sites but not to the related sequences 5'-WGWGW-3' or 5'-WCWGW-3'. This observation can be rationalized by analysis of the disposition of hydrogen binding sites in the minor groove of these sequences. The hydrogen bonding character of an AT base pair resembles that of a TA base pair due to the approximate minor groove symmetry of the adenine N3 and thymine O2 atoms. In a G,C base pair, the guanine 2-amino group lies closer to the guanine-containing strand and as a result, the relative positions of the two guanine amino groups in the three sequences, 5'-GTC-3', 5'-GTG-3' and 5'-CTG-3' are very different. As a consequence of the helicity of B-DNA, the amino groups are separated by a larger distance in the 5'-GTC-3' sequence relative to the 5'-CTG-3' sequence. The 5'-GTG-3' sequence is clearly different from the other two, since both amino groups derive from guanine residues on the same strand. This distinction, in combination with restrictions on ligand-ligand stacking

interactions and other ligand-DNA contacts, favors binding to the 5'-WGTCW-3' sequence over the similar 5'-WGTGW-3' and 5'-WCTGW-3' sequences.

Biological Implications of the 5'-WGWCW-3' sequence. A review of DNA-binding proteins revealed that a 5'-WGWCW-3' sequence is present in the consensus recognition sites of the transcriptional activator GCN4[13] and the oncogenic proteins jun and fos.[14] Current models suggest that these leucine zipper proteins bind as Y-shaped dimers in the major groove of DNA, with no minor groove contacts by the proteins.[15,16] There exists the possibility that 2-ImN can bind as a dimer in the minor groove simultaneous with binding in the major groove by the dimeric proteins. Related work from this laboratory has shown that 2-ImN and GCN4 (226-281) can bind a 5'-CTGACTAAT-3' sequence simultaneously.[17]

Distamycin Binding A,T-Rich Sites: 1:1 or 2:1?

The report by Pelton and Wemmer that distamycin is capable of binding in the minor groove of a 5'-AAATT-3' site as a side-by-side dimer has provided new insight into minor groove recognition by peptide analogs.[1] These NMR studies were conducted at 1-4 mM concentration of peptide and are not necessarily indicative of distamycin binding modes at 1-10 μM concentrations.. However, the degree of cooperativity for the two peptides is relevant to binding at micromolar concentrations. Wemmer and coworkers have examined binding of distamycin to several different five base pair sites and find a strong sequence dependence on the degree of cooperativity of dimeric binding.[6] For example, titrations with a 5'-AAAAA-3' site demonstrate that the 2:1 complex appears only after saturation of the 1:1 binding mode. In contrast, the opposite behavior is

seen with the 5'-ATATA-3' binding site, with only 2:1 complex present at all stoichiometries of peptide.[6] The degree of cooperativity of distamycin to the 5'-AAATT-3' site is intermediate between these cases. Five base pair sites containing a single GC base pair also favor cooperative dimeric binding by distamycin. For the three binding sites, 5'-AGTTT-3', 5'-AAGTT-3' and 5'-AAAGT-3' only 2:1 complexes with distamycin were observed.[6]

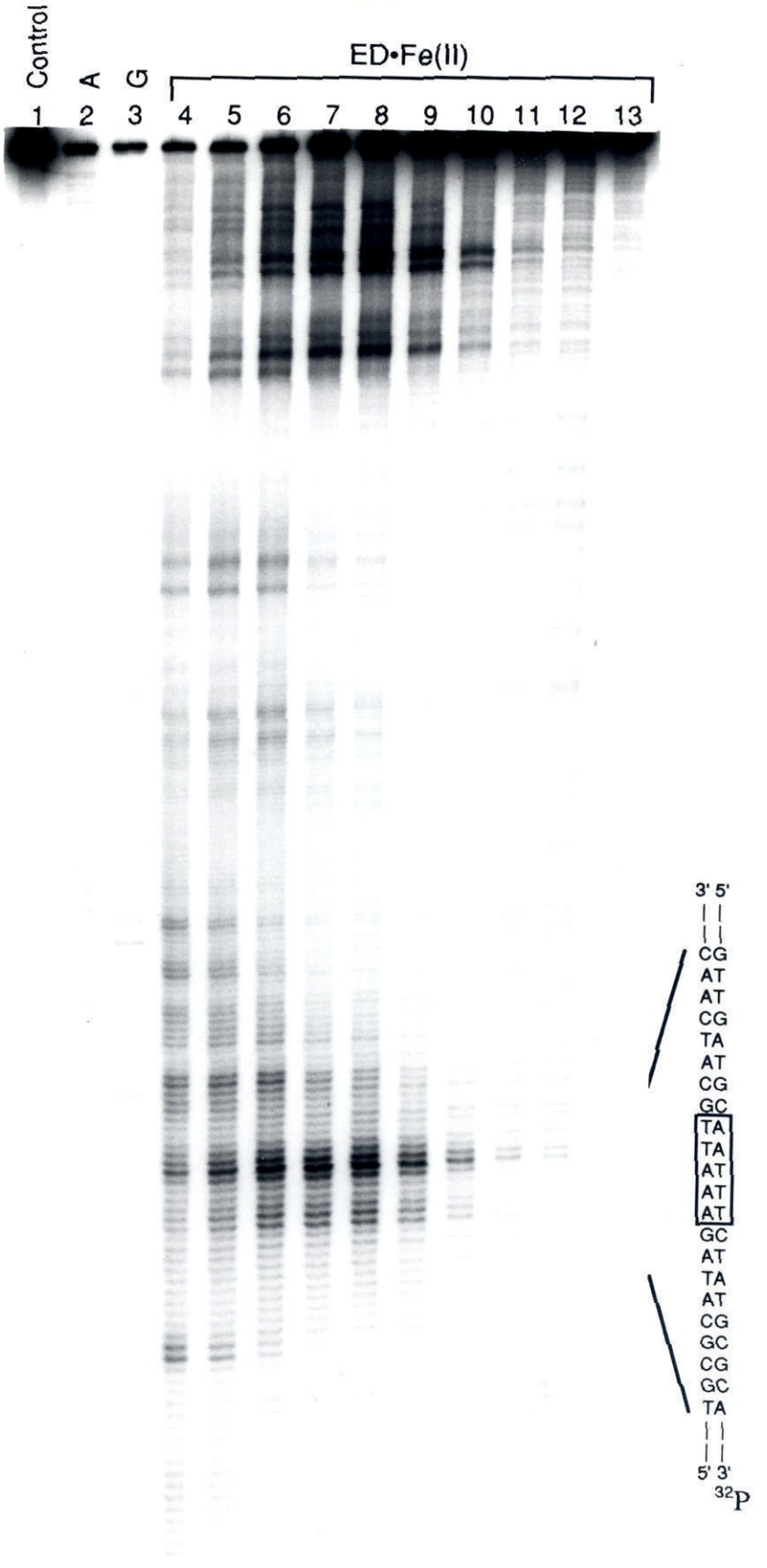
In all of these examples, there is an apparent correlation between the degree of cooperativity and sequence-dependent minor groove width. A comparison of crystal structures suggests that the minor groove is narrowest for poly(dA) tracts for which 1:1 complexes with distamycin are preferred.[18] The minor groove width for mixed A,T and G,C sequences are significantly wider than those for A,T-rich tracts,[19] consistent with the greater cooperativity displayed by peptides binding these sites. A working model for binding of these ligands to DNA involves a step-wise binding mechanism with the two ligands in the minor groove. Upon binding of the first ligand, the second peptide can formally bind with lower, equal or greater affinity. In the examples studied to date, narrow minor groove widths disfavor binding of the second peptide. In contrast, general sequences, with wider minor groove widths, favor 2:1 binding, where the second peptide binds with greater affinity than the first. Apparently, it is energetically more favorable for the DNA to accommodate side-by-side peptides rather than narrow the groove width to accomodate a single peptide, in both cases maintaining extensive van der Waals contacts between peptide and DNA.

Affinity cleaving experiments identify the preferred orientation of peptides at their binding sites and can be used to investigate the relative energetics of 1:1 and 2:1 binding modes for distamycin. Earlier work from the Dervan group has examined binding of ED•Fe(II) at several A,T-rich five base

pair sites.[20] At the concentrations studied (1-10 μM), the orientation preferences in binding the sites ranged from 5:1 to 1:1.[16] These experiments cannot discriminate between two different orientations of a 1:1 complex and a mixture of 1:1 and 2:1 complexes. However, examination of the concentration dependence on affinity cleavage products should distinguish between these very different complexes. If the peptides only bind as 1:1 complexes, the orientation preference once the site is fully occupied should not change with increasing concentration of peptide. If the peptides bind the site as a 2:1 complex, affinity cleaving experiments should reveal equal cleavage at both sides of the binding site, as in the $(2\text{-ImNE})_2$•TGTTA complex, although formally this pattern may be ascribed to two equally favorable 1:1 complexes. Finally, if the peptides bind with an intermediate degree of cooperativity the affinity cleavage patterns should show two unequal orientations at lower concentrations and two equal binding orientations with higher concentrations of distamycin. Moreover, it may be possible to measure the degree of cooperativity for this case.

Quantitative affinity cleaving experiments were carried out on different restriction fragments containing various distamycin binding sites. Cleavage of the 381-bp *Bam*H 1 / *Eco*R 1 fragment from plasmid pBR322 illustrates this experiment at the 5'-AAATT-3' binding site. At low concentrations of peptide, no cleavage is observed on the labeled restriction fragment. With increasing concentration of peptide, cleavage at both ends of the 5'-AAATT-3' binding site is apparent, with a orientation preference of approximately 2:1. As the concentration of peptide is further increased, the cleavage at this site increases and secondary binding sites become evident. Eventually, the cleavage at the 5'-AAATT-3' site as well as other sites diminishes.

Figure 7. Cleavage of the 381 bp *Eco*R 1 / *Bam*H 1 restriction fragment from pBR322 by ED•Fe(II). All reactions contain 25 mM tris acetate, 20 mM NaCl, 100 μM-bp calf thymus DNA, 3'-labeled restriction fragment and 10 mM DTT. Lane 1, intact DNA; lane 2, A reaction; lane 3, G reaction; lanes 4 through 13 contain ED•Fe(II) at 200 μM, 100 μM, 50 μM, 20 μM, 10 μM, 5 μM, 2 μM, 1 μM, 500 nM and 200 nM concentrations, respectively.


Control
A
G
ED•Fe(II)
1 2 3 4 5 6 7 8 9 10 11 12 13
3' 5'
CG
AT
AT
CG
TA
AT
CG
GC
TA
TA
AT
AT
AT
GC
AT
TA
AT
CG
GC
CG
GC
TA
5' 3'
32P

This example reveals the difficulties associated with the quantitative affinity cleaving experiments. The principle complication is the limiting concentration range separating sequence-specific and non-specific binding of distamycin. At > 50 μM concentrations of peptide, secondary sites are bound, which may afford overlapping cleavage patterns or alternatively, may protect the DNA from cleavage by EDTA•Fe(II). Moreover, the binding affinities of distamycin for various five base pair pure A,T-sites can vary by two orders of magnitude. If the binding affinity of distamycin for a site is too low, as is the case with 5′-ATATA-3′, clean affinity cleaving data will only be available for a few concentrations of peptide.

Future Directions. Modification of the above experimental design may allow the cooperativity of distamycin binding to different five base pair pure A,T-sites to be determined. A set of DNA fragments should be constructed which contain a single five base pair distamycin binding site with a minimum of 10 base pairs of G,C-rich sequences on both sides of the binding site. The non-specific binding affinity of distamycin for neighboring sequences in these fragments should allow reliable cleavage data at higher concentrations of ED•Fe(II) to be obtained. The corresponding quantitative footprint titration experiment is an important control for ruling out cleavage proximal to the A,T-rich site by peptides bound to adjacent sites.

Implications for the Design of Minor Groove Binding Molecules. The most significant finding of this study is that 2-ImN and 2-PyN show an apparent preference for cooperative 2:1 binding to the 5'-TGTCA-3' sequence over 1:1 binding. The width of the minor groove for B-DNA is sequence dependent and can vary from 3-4 Å for A,T-rich DNA to 6-7 Å for G,C-containing regions.[18,19] X-ray structures of 1:1 complexes formed between netropsin and distamycin with short oligonucleotides reveal that the natural products sit deeply in the minor

groove and make extensive van der Waals contacts to both walls of the binding site, contributing favorably to the total free energy of binding.[2a] The narrow minor groove widths for A,T-tracts of DNA along may be an exception to the wider minor grooves observed in mixed A,T and G,C sequences. With regard to these wider minor groove sequences, it may in general be energetically more favorable for the minor groove to accept two side-by-side peptides rather than compress the groove for favorable contacts with a single peptide. However, we cannot distinguish at this time the relative contributions from sequence dependent flexibility of the minor groove and sequence dependent groove width. With regard to the latter, esearch from the Wemmer laboratories demonstrated that insertion of a single GC base pair into an otherwise pure A,T-tract favors dimeric binding by peptides. It remains to be seen if 2:1 models can facilitate the design of peptides for recognition of other models.

Finally, dimeric motifs seem unlikely to be limited to *N*-methylpyrrolecarboxamide analogs. For example, chromomycin, an oligosaccharide-chromophore, binds in the minor groove as a 2:1 complex.[xx] Interestingly, the sugar of one ligand of chromomycin stacks on the chromophore unit of the other ligand. The increased width of the chromomycin dimer requires the DNA to undergo a major structural transition, resulting in a wider and more shallow minor groove.[21] It is reasonable to expect a larger class of structure types to form dimeric structures which can recognize double helical DNA with high sequence specificity.

Experimental Section

DNA Reagents and Materials. Doubly distilled water was further purified through the Milli Q filtration system from Millipore. Sonicated, deproteinized calf thymus DNA was purchased from Pharmacia. Plasmid pBR322 was obtained from Boehringer-Mannheim. Enzymes were obtained from Boehringer-Mannheim or New England Biolabs and used with the buffers supplied. Deoxyadenosine 5'-[α-^{32}P]triphosphate and adenosine 5'-[γ-^{32}P] triphosphate were obtained from Amersham. Storage phosphor technology autoradiography was performed using a Molecular Dynamics 400S Phosphorimager and ImageQuant software. The 381 base pair 3'- and 5'-end labeled *Eco*R I / *Bam*H I restriction fragment from plasmid pBR322 was prepared and purified as previously described.[3b] Chemical sequencing reactions were performed according to published methods.[18,19] Standard techniques were employed for DNA manipulations.[20] All other reagents and materials were used as received.

Affinity Cleaving.[3] A 50 μM MPE•Fe(II) solution was prepared by mixing 100 μL of a 100 μM MPE solution with 100 μL of a freshly prepared 100 μM ferrous ammonium sulfate solution. A 100 μM ImPImP•Fe(II) solution was prepared by mixing 10 μL of a 1 mM ImPImP solution with 10 μL of a 1 mM ferrous ammonium sulfate solution and diluting to 100 mL. Solutions were prepared containing 1μL/tube 20x Tris acetate, pH 7.0 buffer, 2 μL/tube 1 mM-bp calf thymus DNA, 1μL/tube 200 mM sodium chloride, labeled restriction fragment and water to make 14 μL/tube total solution. 2 μL of a 10x solution of the peptide was added and the tubes were incubated for 20 min at 22 °C. To the footprinting reactions was added 2 μL of the 50 μM MPE•Fe(II) solution and

incubated 5 min. Cleavage was initiated by the addition of 2 μL of a freshly prepared 50 mM DTT solution. Final concentrations were 25 mM Tris•acetate (pH 7.0), 10 mM sodium chloride, 100 μM-bp DNA, 5 μM MPE•Fe(II) and 5 mM DTT, in 20 μL volume. The reactions were incubated at 22 °C for 25 min, ethanol precipitated, resuspended in 100 mM tris-borate-EDTA / 80% formamideloading buffer and electrophoresed on 6 % denaturing polyacrylamide gels (5% crosslink, 7 M urea) at 1500 V for 3-4 h. The gels were dried and quantitated using storage phosphor technology.

References

1. (a) Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 5723-5727. (b) Pelton, J. G.; Wemmer, D. E. *J. Am. Chem. Soc.* **1990**, *112*, 1393-1399.
2. This is in agreement with previous X-ray and NMR studies; see: (a) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 1376-1380. (b) Pelton, J. G., Wemmer, D. E. *Biochemistry* **1988**, *27*, 8088-8096.
3. (a) Wade, W. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1987**, *109*, 1574-1575. (b) Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 8783-8794.
4. Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Dervan, P. B. *Proc. Natl. Acad. Sci., USA* **1992**, *89*, 7586-7590.
5. Dwyer, T. J.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, *115*, 9900-9906.
6. For a review, see: Wemmer, D. E.; Geierstanger, B. H.; Fagan, P. A.; Dwyer, T. J.; Jacobsen, J. P.; Pelton, J. G.; Ball, G. E.; Leheny, A. R.; Chang, W.-H.; Bathini, Y.; Lown, J. W.; Rentzeperis, D.; Marky, L.; Singh, S.; Kollman, P. in Proceedings of the Eight Conversation on Biomolecular Stereodynamics, Adenine Press (1993) in press.
7. Fagan, P. A.; Wemmer, D. E. *J. Am. Chem. Soc.* **1992**, *114*, 1080-1081.
8. Wade, W. S., Ph.D. Thesis, California Institute of Technology, 1989.
9. (a) Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Methods Enzymol.* **1986**, *130*, 132-181. (b) Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 8462-8466. (c) Senear, D.

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

10. Wade, W. S.; Mrksich, M.; Dervan, P. B. *Biochemistry*, **1993**, *32*, 11385-11389.

11. (a) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 9892-9899. (b) Dwyer, T. J.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, *115*, 9900-9906.

12. (a) Markey, L. A.; Breslauer, K. J. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 4359-4363. (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*, 8922-8926.

13. (a) Hill, D. E.; Hope, I. A.; Macke, J. P.; Struhl, K. *Science* **1986**, *224*, 451-457. (b) Hope, I. A.; Struhl, K. *Cell* **1985**, *43*, 177-188.

14. (a) Bohmann, D. T.; Bos, T. J.; Admon, A.; Nishimura, T.; Vogt, P. K.; Tjian, R. *Science* **1987**, *238*, 1386-1392. (b) Rauscher, F. J.; Sambucetti, L. C.; Curran, T.; Distel, R. J.; Spiegelman, B. M. *Cell* **1988**, *52*, 471-480.

15. Vinson, C. R.; Sigler, P. B.; McKnight, S. L. *Science* **1989**, *246*, 911-916.

16. Oakley, M. G.; Dervan, P. B. *Science* **1990**, *248*, 847-850.

17. Oakley, M. G.; Mrksich, M.; Dervan, P. B. *Biochemistry* **1992**, 31, 10969-10975.

18. (a) Wing, R.; Drew, H.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E. *Nature* **1980**, *287*, 755-758. (b) Nelson, H. C. M.; Finch, J. T.; Luisi, B. F.; Klug, A. *Nature* **1987**, *330*, 221-226. (c) Yoon, C.; Prive, G. G.; Goodsell, D. S.; Dickerson, R. E. *Proc. Natl. Acad. Sci. USA* **1989**, *85*, 6332-6336. (d) DiGabriele, A. D.; Sanderson, M. R.; Steitz, T. A. *Proc. Natl. Acad. Sci. USA* **1989**, *85*, 1816-1820.

19. (a) Heinemann, U.; Alings, C. *J. Mol. Biol.* **1989**, *210*, 369-381. (b) Goodsell, D. S.; Kopka, M. L.; Cascio, D.; Dickerson, R. E. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 2930-2934.

20. (a) Schultz, P. G.; Taylor, J. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1982**, *104*, 6861-6863. (b) Taylor, J. S.; Schultz, P. G.; Dervan, P. B. *Tetrahedron* **1984**, *40*, 457-465. (c) Schultz, P. G.; Dervan, P. B. *J. Biomol. Struct. Dyn.* **1984**, *1*, 1133-1147.

21. (a) Gao, X.; Patel, D. J. *Biochemistry* **1989**, *28*, 751-762. (b) Gao, X.; Mirau, P.; Patel, D. J. *J. Mol. Biol.* **1992**, *223*, 259-279.

Chapter 3

Design of Covalent Peptide Dimers for Enhanced Sequence-Specific Recognition of the Minor Groove of DNA at the 5′-TGTCA-3′ Site

1:1 and 2:1 Peptide-DNA Complexes. The natural products netropsin and distamycin A (D) are crescent shaped di- and tripeptides, respectively, that bind in the minor groove of DNA at sites of four or five successive A,T base pairs (bp).[1-3] The structures of a number of peptide-DNA complexes have been determined by X-ray diffraction[4] and NMR spectroscopy,[5] and the thermodynamic profiles have been studied for these complexes.[6] This work suggests that favorable electrostatic interactions and extensive van der Waals contacts between the peptide and the floor and walls of the minor groove contribute to complex stability. The carboxamide NH's of the peptides participate in bifurcated hydrogen bonds with adenine N3 and thymidine O2 atoms on the floor of the minor groove. The aromatic hydrogens of the *N*-methylpyrrole rings are set too deeply in the minor groove to allow room for the guanine 2-amino group of a G,C base pair, affording binding specificity for A,T-rich sequences. Although this model has aided in the design of oligopeptides for recognition of longer tracts of A,T-rich DNA,[7] efforts to design peptides capable of binding mixed A,T and G,C sequences have proven inconsistent with a 1:1 peptide-DNA model.[8,9]

There have been several recent reports of peptides which bind in the minor groove of DNA as antiparallel side-by-side dimers.[10-15] Pelton and Wemmer found that distamycin at high concentrations (2-4 mM) is capable of binding in the minor groove of 5′-AAATT-3′ as a dimer.[10] Shortly thereafter, the

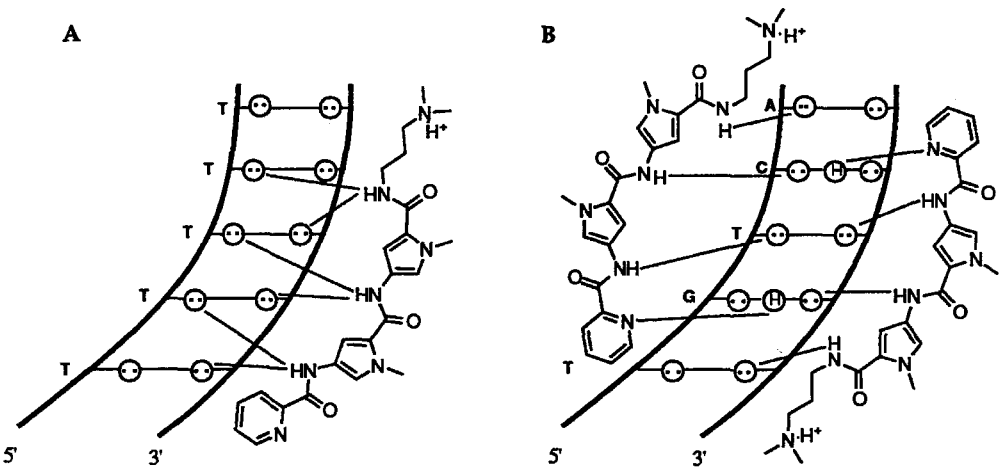
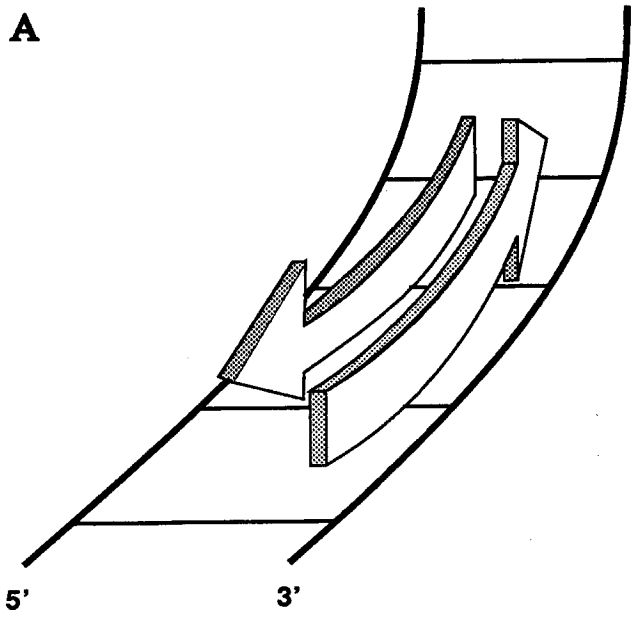
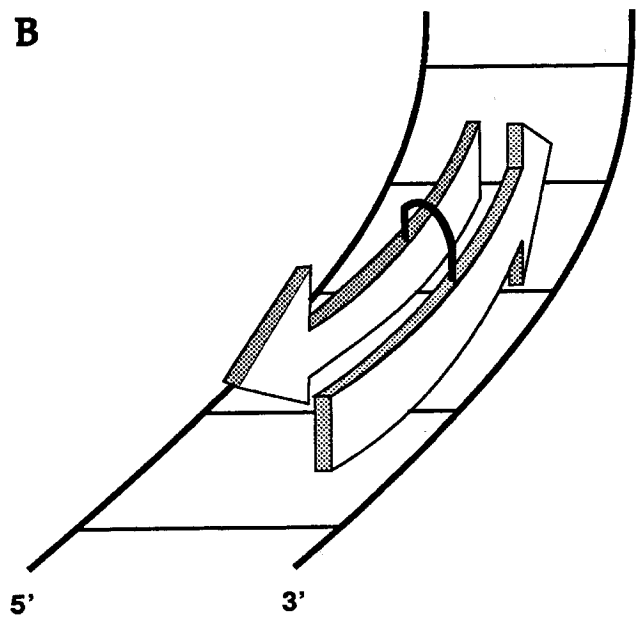


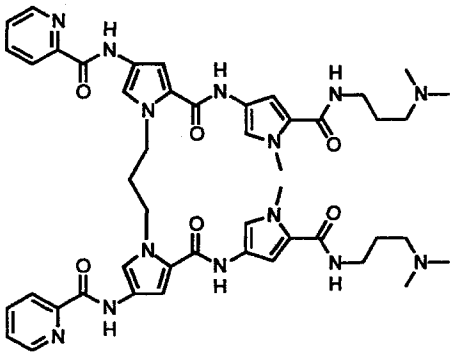
Figure 3.1. (a) 1:1 binding model for the complex formed between 2-PyN with a 5′-TTTTT-3′ sequence, with the nitrogen of the pyridine facing away from the minor groove and (b) 2:1 binding model for the dimeric complex formed between 2-PyN with a 5′-TGTCA-3′ sequence, with the nitrogens of the pyridines facing the floor of the minor groove. Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines, and circles with an H represent the 2-amino group of guanine. Putative hydrogen bonds are illustrated by dotted lines.

designed peptide 1-methylimidazole-2-carboxamide-netropsin (2-ImN) was shown to bind exclusively to the mixed sequence 5′-TGTCA-3′ as a 2:1 complex.[11,12] Another synthetic peptide analog pyridine-2-carboxamide-netropsin (2-PyN) was found to bind to the minor groove of double-helical DNA at two very different sequences, 5′-TTTTT-3′ and 5'-TGTCA-3'.[9,11] From quantitative footprint titration experiments, the apparent first order binding affinities for 2-PyN in complex with 5′-TTTTT-3′ and 5′-TGTCA-3′ are comparable, K_a = 2.3 x 10^5 M^{-1} and 2.2 x 10^5 M^{-1}, respectively (20 mM Tris•HCl and 100 mM NaCl at pH 7.0 and 37 °C).[16] 2-PyN likely binds 5′-TTTTT-3′ as a 1:1 complex and 5′-TGTCA-3′ as a 2:1 complex (Figure 1).

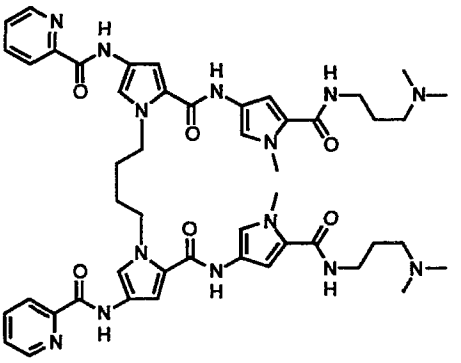
Figure 3.2. Models for (a) binding of two side-by-side peptides and (b) binding by one covalent peptide dimer in the minor groove of DNA.


A
5'
3'

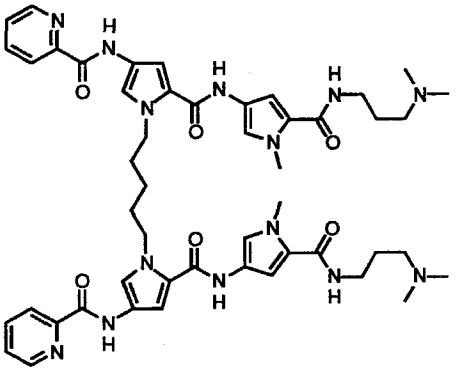

B
5'
3'



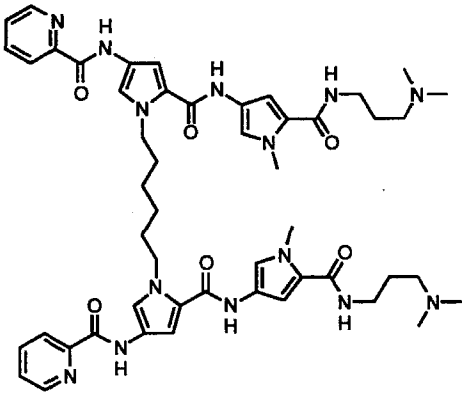
1a: (2-PyN)$_2$-C3



1b: (2-PyN)$_2$-C4



1c: (2-PyN)$_2$-C5



1d: (2-PyN)$_2$-C6

Figure 3.3. Covalently linked peptide dimers (2-PyN)$_2$-C3, (2-PyN)$_2$-C4, (2-PyN)$_2$-C5 and(2-PyN)$_2$-C6 wherein the central pyrroles of 2-PyN are connected with propyl, butyl, pentyl and hexyl linkers, respectively.

Experimental Design. One strategy for increasing the affinity and hence the sequence specificity of peptides that bind DNA sites as side-by-side dimers in the minor groove is to covalently tether the two peptides.[17] The overall free energy of complex formation is expected to be more favorable for a covalent dimer since one bis-peptide should bind with more favorable entropy than do two peptides. Examination of 2:1 peptide-DNA models suggests that linkers from three to six methylene units are able to bridge the central pyrrole rings of

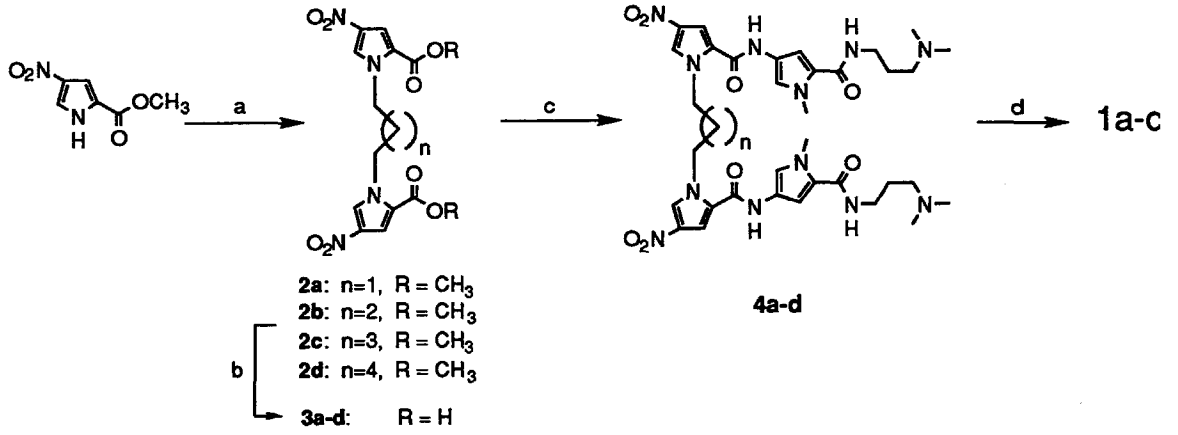


Figure 3.4. Synthetic scheme for (2-PyN)$_2$-C3, (2-PyN)$_2$-C4, (2-PyN)$_2$-C5 and(2-PyN)$_2$-C6. (a) (i) K_2CO_3; (ii) I-$(CH_2)_n$-I; (b) LiOH, EtOH, H_2O; (c) (i) $SOCl_2$; (ii) 1-methyl-4-amino-2-(carboxamidopropyl-3-dimethylamino) pyrrole; (d) (i) 300 psi H_2, 10% Pd/C; (ii) picolinic acid, DCC, HOBt.

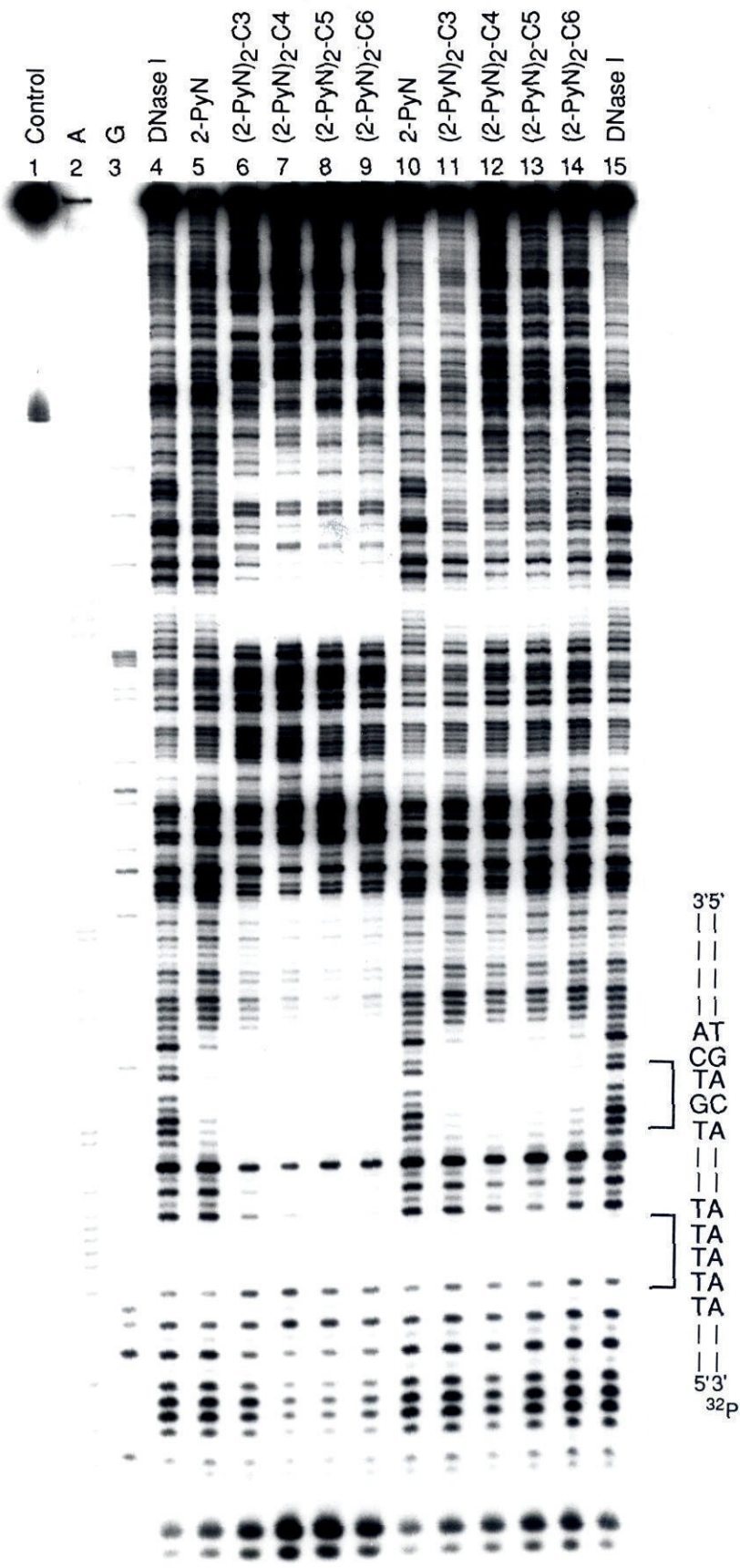
the two side-by-side peptides in complex with DNA without perturbing the contacts between the peptides and DNA (Figure 2).[10,12] Four covalently linked dimers of 2-PyN have been synthesized wherein the peptides are connected through the nitrogens of the central pyrrole rings with propyl, butyl, pentyl and hexyl linkers (Figure 3). Quantitative DNase I footprint titration experiments with these bis(pyridine-2-carboxamide-netropsin)$(CH_2)_{3-6}$ peptides afford a comparison of the apparent first order binding affinities of 2-PyN and the four covalent peptide dimers (2-PyN)$_2$-C3, (2-PyN)$_2$-C4, (2-PyN)$_2$-C5 and (2-PyN)$_2$-C6 to both the 5'-TTTTT-3' and 5'-TGTCA-3' sites.

Results

Synthesis of Covalently Linked Peptides. The methodology for the synthesis of covalently linked peptides is analogous to that of the corresponding monomeric peptides (Figure 4).[11] Alkylation of 4-nitro-2-carboxymethylpyrrole with a diiodoalkane (K_2CO_3, acetone) affords the corresponding di-*N,N'*-pyrrolealkanes **2a-d** in 80-85 % yield. Saponification of the methyl esters (LiOH, EtOH, H_2O) followed by conversion to the bis(acid chlorides) ($SOCl_2$) and coupling with 1-methyl-4-amino-2-(carboxamidopropyl-3-dimethylamino) pyrrole yields the corresponding bis-dipyrrole derivatives **4a-d** in 50-70% yield. Reduction of the bis-nitrodipyrroles (300 psi H_2, Pd/C) and coupling of the resulting bis-aminopyrroles with picolinic acid (DCC, HOBt) affords the dimeric peptides $(2\text{-PyN})_2$-C3, $(2\text{-PyN})_2$-C4, $(2\text{-PyN})_2$-C5 and$(2\text{-PyN})_2$-C6 in 45-60% yield.

DNase I Footprinting. DNase I footprinting experiments on the 517-bp *Eco*R I / *Rsa* I restriction fragment from plasmid pBR322 (10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$ at pH 7.0 and 22 °C) reveal that 2-PyN and the covalent peptide dimers $(2\text{-PyN})_2$-C3, $(2\text{-PyN})_2$-C4, $(2\text{-PyN})_2$-C5 and $(2\text{-PyN})_2$-C6, at 40 μM concentration, protect both the 5'-TGTCA-3' and the 5'-TTTTT3' sites from cleavage by the enzyme DNase I (Figure 5, lanes 4-9).[19] At 10 μM concentration, 2-PyN does not protect any sites on double-helical DNA from cleavage by DNase I (Figure 5, lane 10). In contrast, the covalently linked peptides, $(2\text{-PyN})_2$-C3, $(2\text{-PyN})_2$-C4, $(2\text{-PyN})_2$-C5 and $(2\text{-PyN})_2$-C6, at 10 μM concentration, do bind the 5'-TGTCA-3' site (Figure 5, lanes 11-14). In order to compare the binding affinities of the five peptides in complex with the 5'-TGTCA-3' and 5'-TTTTT-3' sites, the apparent first order binding affinities were determined by quantitative DNase I footprint titration experiments (Table I).[20]

Figure 3.5. DNase I footprinting of 2-PyN, $(2\text{-PyN})_2$-C3, $(2\text{-PyN})_2$-C4, $(2\text{-PyN})_2$-C5 and $(2\text{-PyN})_2$-C6. The 5′-TGTTA-3′ and 5′-TTTTT-3′ binding sites are shown on the right side of the autoradiogram. All reactions contain 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, and 100 μM-bp calf thymus DNA and 20 kcpm 3′-labeled 517-base *Eco*R I / *Rsa* I restriction fragment from plasmid pBR322. Lane 1, intact DNA; lane 2, A reaction; lane 3, G reaction; lane 4, DNase I standard; lane 5, 40 μM 2-PyN; lane 6, 40 μM $(2\text{-PyN})_2$-C3; lane 7, 40 μM $(2\text{-PyN})_2$-C4; lane 8, 40 μM $(2\text{-PyN})_2$-C5; lane 9, 40 μM $(2\text{-PyN})_2$-C6; lane 10, 10 μM 2-PyN; lane 11, 10 μM $(2\text{-PyN})_2$-C3; lane 12, 10 μM $(2\text{-PyN})_2$-C4; lane 13, 10 μM $(2\text{-PyN})_2$-C5; lane 14, 10 μM $(2\text{-PyN})_2$-C6; lane 15, DNase I standard.


1 Control
2 A
3 G
4 DNase I
5 2-PyN
6 (2-PyN)2-C3
7 (2-PyN)2-C4
8 (2-PyN)2-C5
9 (2-PyN)2-C6
10 2-PyN
11 (2-PyN)2-C3
12 (2-PyN)2-C4
13 (2-PyN)2-C5
14 (2-PyN)2-C6
15 DNase I
3'5'
AT
CG
TA
GC
TA
TA
TA
TA
TA
TA
5'3'
32P

All four covalently linked peptides bind the 5′-TGTCA-3′ site with apparent first order binding constants approximately one order of magnitude greater than that of 2-PyN. The $(2\text{-PyN})_2$-C3 and $(2\text{-PyN})_2$-C4 peptides bind to the A,T-rich site with slightly lower affinities than does 2-PyN, while $(2\text{-PyN})_2$-C5 and $(2\text{-PyN})_2$-C6 bind to this site with affinities comparable to that of the parent peptide 2-PyN (Table I).

Discussion.

Binding Affinity. The DNase I footprinting experiments reveal that the apparent first order binding affinities of the covalently linked dimers in complex with the 5'-TGTCA-3' site are ten- to twenty-fold greater than that of the parent peptide 2-PyN in complex with this site. Covalently tethering the two peptides therefore results in more favorable free energies of binding, likely due to more favorable entropies of binding. A comparison of the binding affinities for the 5′-TGTCA-3′ site by the covalently linked peptides suggests that the length of the alkyl linker has little effect on overall complex stability. Because these simple alkyl linkers may not be optimized for side-by-side binding, a second generation of linked peptides may display still higher affinities for dimeric binding sites. With regard to binding the 5′-TTTTT-3′ site, all four covalently linked peptide dimers bind with affinities comparable to that for 2-PyN. If one peptide of the covalently linked dimer is bound in the minor groove of the A,T-rich site, we would expect the steric bulk of the other peptide to destabilize the complex. Consistent with this interpretation, the linked peptides with the shorter propyl and butyl tethers display lower binding affinity for the A,T-rich site relative to the peptides with the longer pentyl and hexyl linkers. Thermodynamic studies

Table 3.I Relative Binding Affinities[a,b]

	Binding Site	
Peptide	**5′-TGTCA-3′**	**5′-TTTTT-3′**
2-PyN	1.2	1.0
(2-PyN)$_2$-C3	10.0	0.5
(2-PyN)$_2$-C4	13.3	0.5
(2-PyN)$_2$-C5	8.8	1.2
(2-PyN)$_2$-C6	6.7	1.2

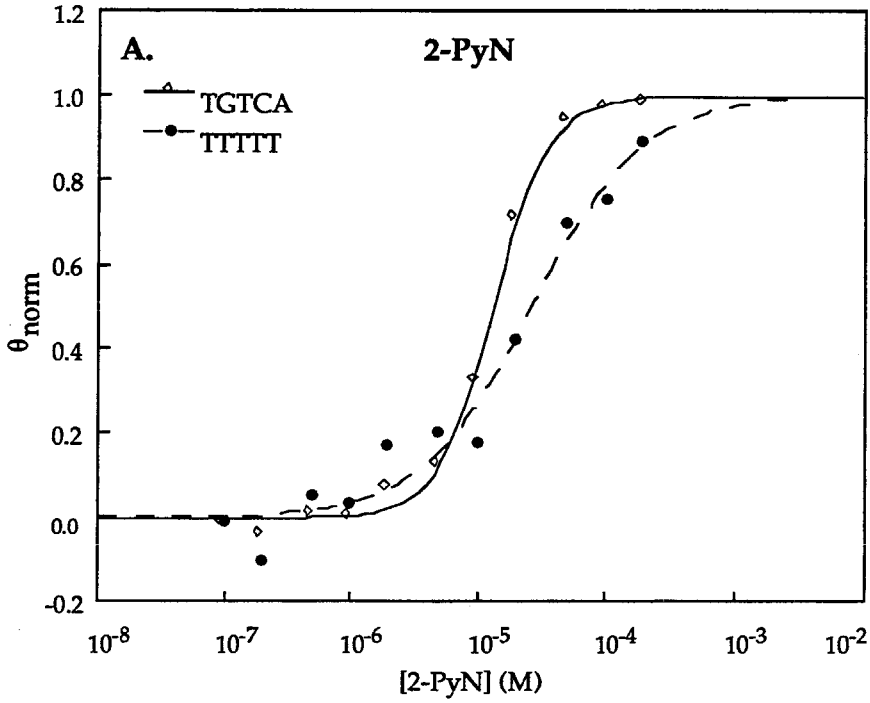
[a]Values reported are ratios of binding affinities relative to that for 2-PyN binding the 5'-TTTTT-3' site. [b]The assays were performed in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$ and 100 μM-bp calf thymus DNA at pH 7.0 and 22 °C.

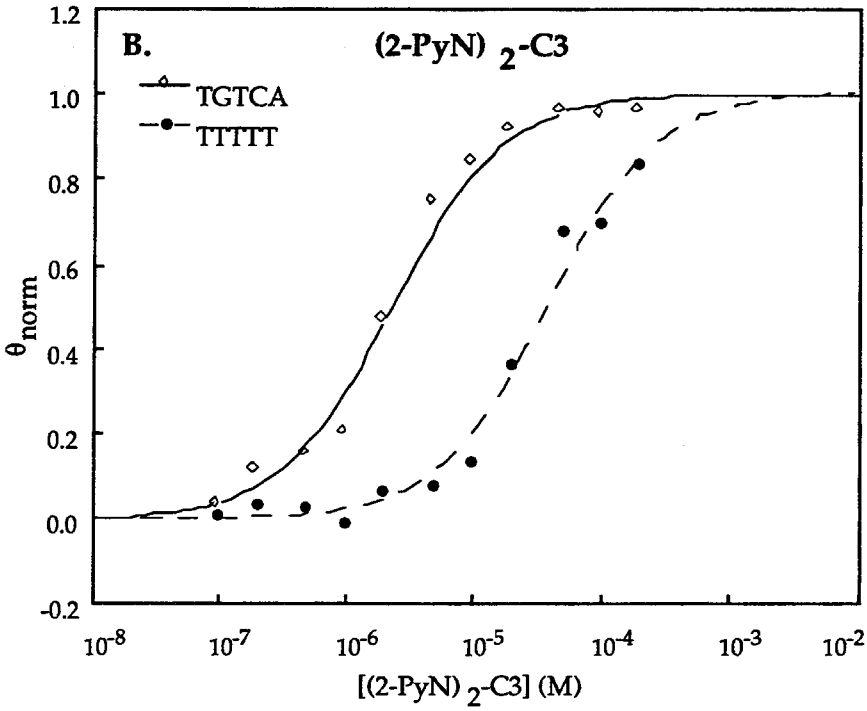
may provide further insight into the enthalpy-entropy compensations in these complexes.

Sequence Specificity. The four covalent dimers all display increased sequence specificities relative to the parent compound 2-PyN (Figure 6). The monomeric peptide 2-PyN binds the A,T-rich site and the 5′-TGTCA-3′ site with similar apparent first order binding affinities and, consequently, displays little discrimination between these two sites. The covalently linked peptides, however, bind the 5′-TGTCA-3′ site with ten- to twenty-fold higher binding affinities relative to binding the 5′-TTTTT-3′ site (Figure 6 and Table I). Moreover, this difference in stabilities mainly arises from *enhanced binding to the 5′-TGTCA-3′ site*, rather than decreased binding affinities for the 5′-TTTTT-3′ site.

Binding Models. The footprinting experiments described here cannot distinguish between intramolecular dimeric binding by a single covalently linked peptide versus intermolecular dimeric binding by two covalently linked peptides. In collaboration with the Wemmer group, we have studied the complexes of 2-PyN and the four covalent peptide dimers with an

Figure 3.6. Data for the quantitative DNase I footprint titration experiments for (a) 2-PyN and (b) $(2\text{-PyN})_2$-C3 in complex with the 5′-TGTCA-3′ and 5′-TTTTT-3′ sites. The θ_{norm} points were obtained using photostimulable storage phosphor autoradiography and processed as described in the experimental section. The data points for the 5′-TGTCA-3′ site are indicated by open diamonds (◊) and for the 5′-AAAAA-3′ site by filled circles (•). The solid and dashed curves are the best-fit Langmuir binding titration isotherms obtained from nonlinear least squares algorithm using eq. 2. The data points for 2-PyN binding the 5′-TGTCA-3′ site were fit assuming a cooperative 2:1 complex (eq 2, n = 2).


A.
2-PyN
TGTCA
TTTTT
θnorm
1.2
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
10-8
10-7
10-6
10-5
10-4
10-3
10-2
[2-PyN] (M)


B.
(2-PyN) 2-C3
TGTCA
TTTTT
θnorm
1.2
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
10-8
10-7
10-6
10-5
10-4
10-3
10-2
[(2-PyN) 2-C3] (M)

oligonucleotide containing a 5′-TGACT-3′ binding site by two-dimensional NMR.[18] In all cases, one-dimensional spectra of titration experiments of the oligonucleotide with increasing amounts of peptide show a single complex forming, which is nearly identical for all five peptides. Two-dimensional NOESY experiments of the oligonucleotide complexed with two equivalents of 2-PyN reveal that the monomer binds as a side-by-side dimer, similar to the characterized $(2\text{-ImN})_2 \bullet$5′-TGACT-3′ complex.[12] Two-dimensional NOESY experiments of the oligonucleotide complexed with one equivalent of either $(2\text{-PyN})_2$-C3 or $(2\text{-PyN})_2$-C6 reveal that the covalent peptide dimers do indeed bind as intramolecular dimers. Since the one-dimensional spectra are nearly identical for all the ligands, we infer that the butyl- and pentyl-linked peptides also bind as intramolecular dimers.[18] The NOESY crosspeaks are essentially identical in shift and intensity for all complexes, suggesting that the linked peptides bind with similar geometry and peptide-DNA contacts.[18]

Implications for the Design of Minor Groove-Binding Peptides. This first generation of covalently linked peptide dimers employing simple alkyl linkers binds to the minor groove of DNA at 5′-TGTCA-3′ sites with ten- to twenty-fold higher binding affinities than does the parent peptide 2-PyN. In addition, the dimeric peptides bind double-helical DNA with improved sequence specificities. These results suggest that covalently linking peptides for side-by-side binding may be an important component in the rational design of peptides for sequence specific recognition of the minor groove of DNA. We would expect such an effect to be most evident with a heterodimeric system, wherein two different peptides can formally bind to three different binding sites: each of the parent sites and the heterodimer site.[14] A covalently linked heterodimer of two different peptides would be expected to display increased binding affinity for the heterodimer site and improved sequence specificity.

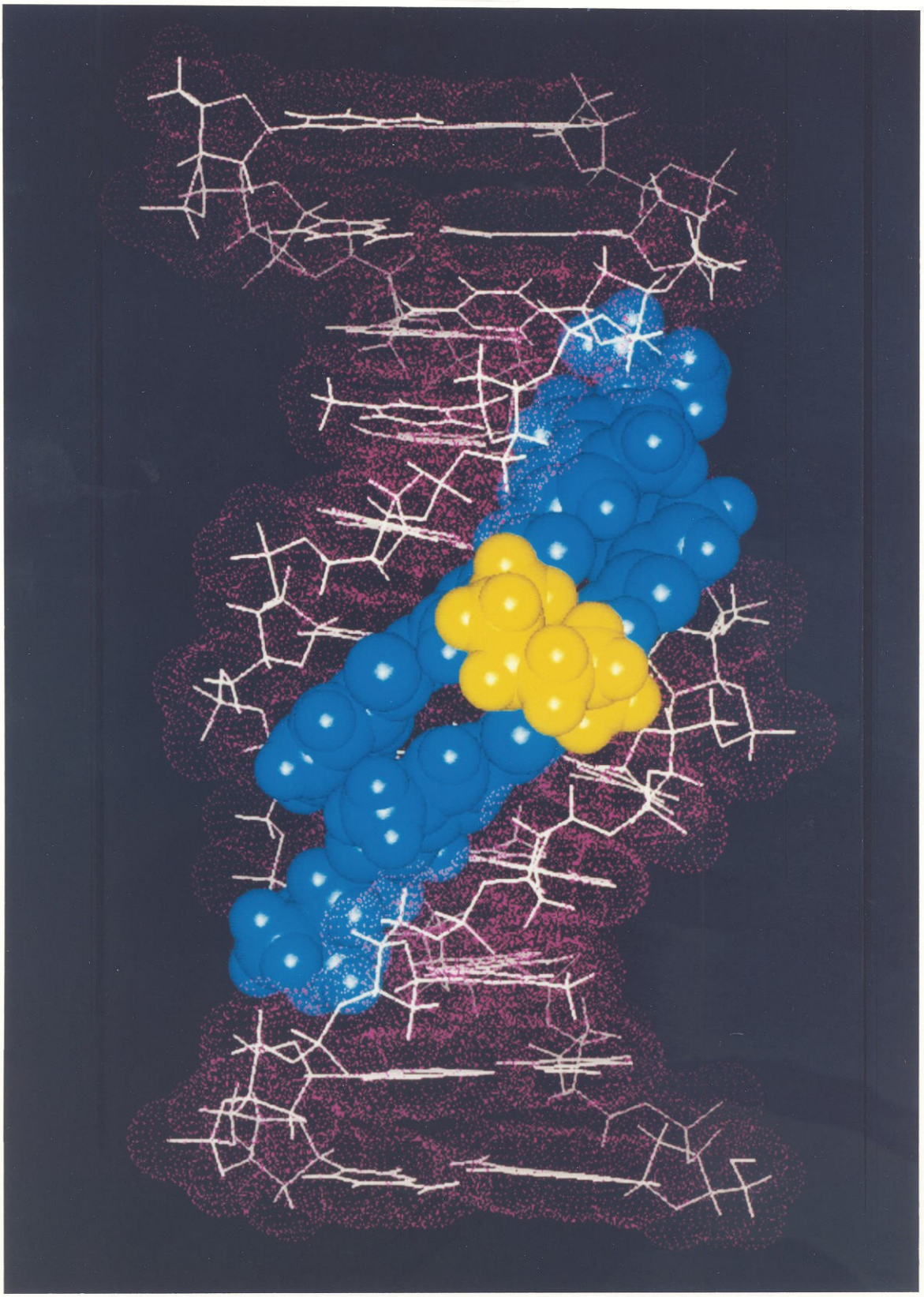
Structural Characterization of Covalent Peptide Dimers in Complex with the 5′-TGACT-3′ Site

In collaboration with the Wemmer group, the structures of the complexes formed upon binding of the covalent peptide dimers $(2\text{-PyN})_2$-C3, $(2\text{-PyN})_2$-C4, $(2\text{-PyN})_2$-C5 and $(2\text{-PyN})_2$-C6 to the oligonucleotide duplex d(GCATGACTCGG): d(CCGAGTCATGC) were studied by two-dimensional NMR. A summary of the study is presented below: experimental details can be found in a published report.[18]

Two-dimensional NMR experiments reveal that the monomeric ligand 2-PyN forms a single, unique complex with the DNA duplex d(GCATGACTCGG): d(CCGAGTCATGC). This complex is composed of two 2-PyN ligands bound simultaneously in the minor groove of the central five base pairs of the DNA. Titration experiments reveal that the 2:1 $(2\text{-PyN})_2$•5′-TGACT-3′ complex is formed with positive cooperativity. The observed intermolecular proton-proton contacts can be accounted for by a complex in which the ligands are oriented in a side-by-side antiparallel arrangement and stacked such that the aromatic rings of one ligand overlap the amide bonds of the other ligand.

The four covalent peptide dimers bind to the 5′-TGACT-3′ site with nearly identical geometry and peptide-DNA contacts as in the $(2\text{-PyN})_2$•5′-TGACT-3′ complex. In all the complexes, one peptide spans the 5′-TGAC-3′ site while the other peptide spans the 5′-AGTC-3′ site on the opposite strand. This geometry likely allows the pyridine nitrogen of each ligand to participate in a hydrogen bond with the C2-amino group of each guanine residue on the floor of the minor groove. This association via hydrogen bonds likely not only plays a role in the stabilization of the complex, but also in the specific recognition of the 5′-TGACT-3′ site by the peptides. Further, the two cationic *N,N*-dimethylammonium

Figure 3.7. Molecular model of the complex formed upon binding of (2-PyN)$_2$-C6 to the oligonucleotide duplex d(GCATGACTCGG): d(CCGAGTCATGC). The two 2-PyN peptides are shown in blue and the hexyl linker is shown in yellow.

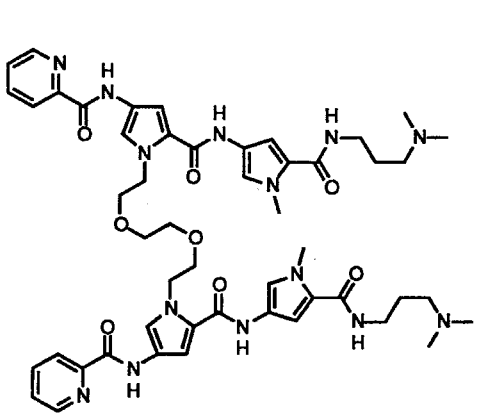


groups of the ligands are directed towards opposite ends of the binding site, providing favorable electrostatic interactions with the minor groove.

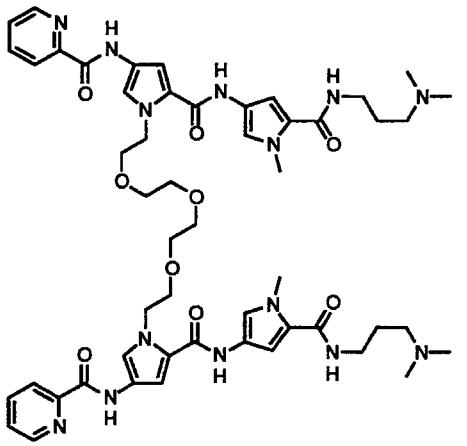
A comparison of the binding affinities among the covalently linked 2-PyN dimers reveals that the length of the alkyl linker has little effect on overall complex stability. The NMR results presented here suggest that the relative positions of the two side-by-side peptides and the specific peptide-DNA contacts are nearly identical for 2-PyN and the four linked peptides in complex with the 5′-TGACT-3′ site. While the alkyl tethers in these four dimers are able to accommodate the close contacts between the peptides and the minor groove, there remains the likely possibility that the linkers are not optimized. A second generation of linked peptides may display even higher binding affinities and improved sequence specificities.

Design of Covalent Peptide Dimers Containing a Metal Binding Domain

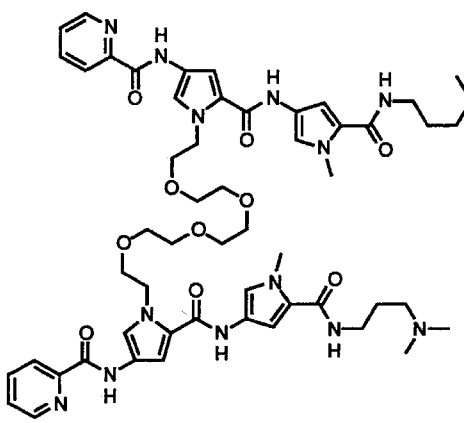
Metal-Dependent DNA Binding. Many small molecules require the presence of metal ions for sequence-specific binding or cleavage of DNA. The aureolic acid antibiotic Chromomycin A3 requires divalent cations for DNA binding.[21] The antitumor agent Bleomycin requires Fe(II) for DNA cleavage[22] and the tripeptide GlyGlyHis in complex with Ni(II) has been shown to effect cleavage of DNA.[23] Griffin and Dervan described a designed metalloregulated DNA binding molecule. Two netropsin peptides were linked in a head to head fashion with oligo(ethylene oxide) linkers.[24] Affinity cleaving experiments demonstrated that in the presence of barium or strontium, the peptides



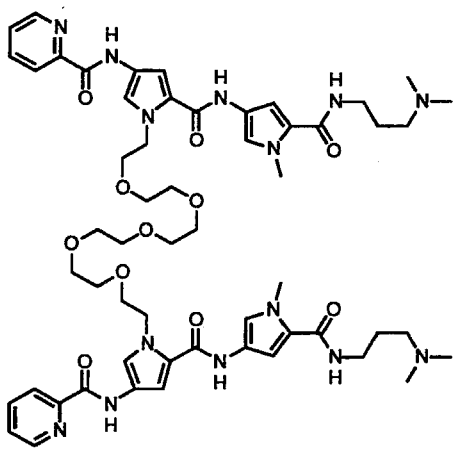
7a: $(2\text{-PyN})_2$-O2



7b: $(2\text{-PyN})_2$-O3



7c: $(2\text{-PyN})_2$-O4



7d: $(2\text{-PyN})_2$-O5

Figure 3.8. Covalently linked peptide dimers $(2\text{-PyN})_2$-O2, $(2\text{-PyN})_2$-O3, $(2\text{-PyN})_2$-O4 and$(2\text{-PyN})_2$-O5 wherein the central pyrroles of 2-PyN are connected with di-, tri-, tetra- and penta-(ethylene oxide) linkers, respectively.

specifically bind a 5′-TATAGGTTAA-3′ site. In the absence of these cations or in the presence of other metals, no binding to this site was observed.

Design of Metalloregulated Side-By-Side Dimers. The geometry of 2:1 peptide-DNA models suggest that the side-by-side peptides can be linked with similar polyether linkers for metal dependent binding. Four peptide dimers

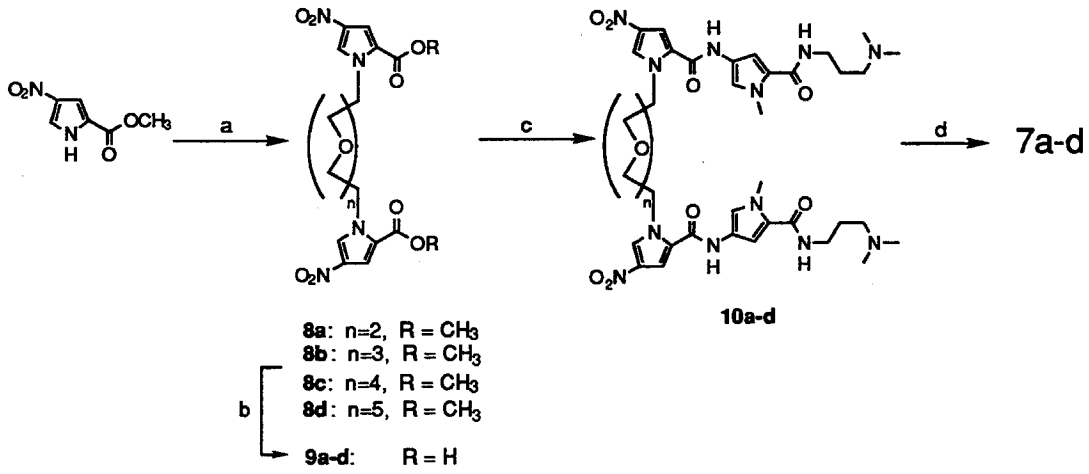


Figure 3.9. Synthetic scheme for (2-PyN)$_2$-O2, (2-PyN)$_2$-O3, (2-PyN)$_2$-O4 and(2-PyN)$_2$-O5. (a) (i) K_2CO_3; (ii) $TsO(CH_2CH_2O)_n$-CH_2CH_2OTs; (b) LiOH, EtOH, H_2O; (c) (i) $SOCl_2$; (ii) 1-methyl-4-amino-2-(carboxamidopropyl-3-dimethylamino) pyrrole; (d) (i) 300 psi H_2, 10% Pd/C; (ii) picolinic acid, DCC, HOBt.

where in the nitrogens of the central pyrrole rings of 2-PyN were connected with di-, tri-, tetra- and penta-(ethylene oxide) linkers were synthesized. The (2-PyN)$_2$-O2 and (2-PyN)$_2$-O3 peptides should not bind metals strongly and serve as controls for metal dependent binding by the (2-PyN)$_2$-O5 and (2-PyN)$_2$-O4 peptides. In the absence of specific metal ions, the linked peptides should bind with lower affinity than do the alkyl-linked peptides, since the long and flexible linkers are entropically less favorable. In the presence of specific metal cations, wrapping of the oligoether around a metal may favorably organize the two peptides for DNA binding.

Synthesis of Oligo(ethylene oxide)-Linked Peptides. The methodology for the synthesis of these covalently linked peptides is analogous to that of the alkyl-linked peptide dimers (Figure 4).[25] Alkylation of 4-nitro-2-carboxymethylpyrrole with a bistosylate (K_2CO_3, acetone) affords the corresponding di-*N,N'*-pyrroleoligo(ethers) **8a-d** in 90-95% yield. Saponification

of the methyl esters (LiOH, EtOH, H_2O) followed by conversion to the bis(acid chlorides) ($SOCl_2$) and coupling with 1-methyl-4-amino-2-(carboxamidopropyl-3-dimethylamino) pyrrole yields the corresponding bis-dipyrrole derivatives **10a-d**. Reduction of the bis-nitrodipyrroles (300 psi H_2, Pd/C) and coupling of the resulting bis-aminopyrroles with picolinic acid (DCC, HOBt) affords the dimeric peptides $(2\text{-PyN})_2$-O2, $(2\text{-PyN})_2$-O3, $(2\text{-PyN})_2$-O4 and$(2\text{-PyN})_2$-O5.

DNA Binding. Preliminary footprinting experiments demonstrate that in the absence of specific divalent metal ions, the peptides $(2\text{-PyN})_2$-O2, $(2\text{-PyN})_2$-O3, $(2\text{-PyN})_2$-O4 and$(2\text{-PyN})_2$-O5 bind the 5′-TGTCA-3′ site with similar affinity as does the parent peptide 2-PyN. Metalloregulated binding with these peptides has not been carefully investigated. This work is continuing in the Dervan Group.

Experimental Section.

^1H NMR and ^{13}C NMR spectra were recorded on a General Electric-QE 300 NMR spectrometer in $CDCl_3$ or DMSO-d_6, with chemical shifts reported in parts per million relative to tetramethyl silane or residual DMSO-d_5, respectively. 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. Reactions were executed under an inert argon atmosphere. Reagent grade chemicals were used as received unless otherwise noted. Tetrahydrofuran (THF) was distilled under nitrogen from sodium/benzophenone ketyl. Dichloromethane (CH_2Cl_2) and triethylamine were distilled under nitrogen from powdered calcium hydride. Dimethylformamide (DMF) was purchased as an

anhydrous solvent from Aldrich. Flash chromatography was carried out using EM science Kieselgel 60 (230-400) mesh.[26] Thin-layer chromatography was performed on EM Reagents silica gel plates (0.5 mm thickness). All compounds were visualized with short-wave ultraviolet light.

***N,N'*-(1,n-Dialkane)bis-2-carboxymethyl-4-nitropyrrole 2a-d (Exemplified with 2a).** To a solution of 2-carboxy-4-nitropyrrole methyl ester (90 mg, 0.529 mmol) in acetone (4.0 mL) was added potassium carbonate (175 mg, 1.27 mmol) and the resulting solution was allowed to stir at room temperature for 1 h. 1,3-Diiodopropane (36 μL, 0.313 mmol) was added and the reaction mixture was heated to 65 °C for 6 h. The solution was cooled and solvent was removed under reduced pressure. The residue was purified by flash column chromatography (2 % MeOH in CH_2Cl_2) to afford **2a**.

2a: yield 73% (85 mg); ^{1}H NMR ($CDCl_3$) δ 7.68 (d, 2 H, *J* = 2.0 Hz), 7.41 (d, 2 H, *J* = 2.0 Hz), 4.46 (t, 4 H, *J* = 7.3 Hz), 3.86 (s, 6 H), 2.17 (m, 2 H); ^{13}C NMR ($CDCl_3$) δ 160.3, 135.6, 126.9, 122.1, 113.2, 51.9, 47.4, 32.3; IR (thin film) 3132 (w), 2952 (m), 1715 (s), 1538 (m), 1515 (s), 1463 (m), 1417 (m), 1317 (s), 1257 (m), 1204 (m), 1103 (m), 860 (s); FABMS *m/e* 380.0974 (M + H, 380.0968 calcd. for $C_{15}H_{16}N_4O_8$).

2b: yield 80% (140 mg); ^{1}H NMR ($CDCl_3$) δ 7.69 (d, 2 H, *J* = 2.0 Hz), 7.43 (d, 2 H, *J* = 2.0 Hz), 4.42 (m, 4 H), 3.88 (s, 6 H), 1.86 (m, 4 H); ^{13}C NMR ($CDCl_3$) δ 160.4, 135.5, 126.9, 121.9, 113.3, 52.0, 49.4, 27.8; IR (thin film) 3123 (w), 2955 (m), 1698 (s), 1538 (m), 1505 (s), 1420 (m), 1387 (m), 1325 (s), 1304 (m), 1257 (m), 1193 (m), 1082 (m), 748 (m); FABMS *m/e* 394.1141 (M + H, 394.1125 calcd. for $C_{16}H_{18}N_4O_8$).

2c: yield 83% (85 mg); ^{1}H NMR ($CDCl_3$) δ 7.65 (d, 2 H, *J* = 2.0 Hz), 7.42 (d, 2 H, *J* = 2.0 Hz), 4.36 (t, 4 H, *J* = 7.4 Hz), 3.87 (s, 6 H), 1.87 (m, 4 H), 1.39 (m, 2 H); ^{13}C NMR ($CDCl_3$) δ 160.4, 135.5, 126.7, 122.0, 113.2, 51.9, 49.9, 30.4, 23.2; IR

(thin film) 3144 (w), 2955 (w), 1713 (s), 1504 (s), 1421 (m), 1389 (m), 1319 (s), 1266 (m), 1221 (m), 1122 (w), 1082 (m); FABMS *m/e* 408.1300 (M + H, 408.1281 calcd. for $C_{17}H_{20}N_4O_8$).

2d: yield 81% (92 mg); ^{1}H NMR ($CDCl_3$) δ 7.62 (d, 2 H, *J* = 2.0 Hz), 7.42 (d, 2 H, *J* = 2.0 Hz), 4.34 (t, 4 H, *J* = 7.3 Hz), 3.86 (s, 6 H), 1.81 (m, 4 H), 1.37 (m, 4 H); ^{13}C NMR ($CDCl_3$) δ 160.4, 135.5, 126.7, 122.0, 113.2, 51.9, 50.2, 30.8, 25.8; IR (thin film) 3142 (w), 3117 (w), 2924 (w), 1715 (s), 1504 (s), 1423 (m), 1388 (m), 1320 (s), 1279 (m), 1246 (m), 1196 (m), 1083 (m); FABMS *m/e* 422.1452 (M + H, 422.1438 calcd. for $C_{18}H_{22}N_4O_8$).

***N,N'*-(1,n-Dialkane)bis-4-nitropyrrole-2-carboxylic acid 3a-d (Exemplified with 3a).** To a flask charged with bispyrrolemethylester **2a** (280 mg, 0.737 mmol) was added ethanol (4.0 mL) and 0.5 N lithium hydroxide (8.0 mL). The mixture was heated to 80 °C and allowed to stir for 4 h. The solution was filtered, the pH of the filtrate was adjusted to pH = 2-3 with 1N HCl, and the white precipitate was collected by filtration and dried to afford diacid **3a**.

3a: yield 93% (220 mg); ^{1}H NMR (DMSO-d_6) δ 8.28 (d, 2 H, *J* = 2.0 Hz), 7.25 (d, 2 H, *J* = 2.0 Hz), 4.39 (t, 4 H, *J* = 6.8 Hz), 2.22 (m, 2 H); ^{13}C NMR (DMSO-d_6) δ 160.9, 134.4, 128.8, 123.1, 112.0, 46.8, 31.7; IR (thin film) 3448 (w), 3178 (m), 2956 (w), 1700 (s), 1542 (m), 1515 (s), 1492 (m), 1421 (m), 1375 (s), 1301 (s), 1253 (m), 1207 (m), 1109 (m), 1082 (w), 860 (w), 817 (m), 752 (m); FABMS *m/e* 352.0661 (M + H, 352.0655 calcd. for $C_{13}H_{13}N_4O_8$).

3b: yield 97% (320 mg); ^{1}H NMR (DMSO-d_6) δ 8.25 (s, 2 H), 7.24 (s, 2 H), 4.34 (bs, 4 H), 1.66 (bs, 4 H); ^{13}C NMR (DMSO-d_6) δ 160.8, 134.3, 128.7, 123.1, 112.0, 48.9, 27.4; IR (thin film) 3448 (w), 3138 (m), 2961 (w), 1686 (s), 1543 (m), 1508 (s), 1422 (m), 1372 (s), 1321 (s), 1264 (m), 1194 (m), 1101 (m), 1083 (w), 866 (w), 822 (m), 751 (m); FABMS *m/e* 367.0898 (M + H, 367.0890 calcd. for $C_{14}H_{15}N_4O_8$).

3c: yield 95% (180 mg); ^{1}H NMR (DMSO-d_6) δ 8.24 (d, 2 H, J = 2.0 Hz), 7.25 (d, 2 H, J = 2.0 Hz), 4.33 (t, 4 H, J = 7.0 Hz), 1.72 (m, 4 H), 1.17 (m, 2 H); ^{13}C NMR (DMSO-d_6) δ 160.8, 134.2, 128.8, 123.0, 112.0, 49.1, 30.0, 22.5; IR (thin film) 3448 (w), 3140 (m), 2949 (w), 1684 (s), 1504 (m), 1510 (s), 1420 (m), 1371 (m), 1318 (s), 1272 (m), 1189 (w), 1098 (w), 861 (w) 754 (w); FABMS *m/e* 381.1047 (M + H, 381.1046 calcd. for $C_{15}H_{17}N_4O_8$).

3d: yield 95% (600 mg); ^{1}H NMR (DMSO-d_6) δ 8.23 (d, 2 H, J = 2.0 Hz), 7.24 (d, 2 H, J = 2.0 Hz), 4.31 (t, 4 H, J = 7.0 Hz), 1.68 (bs, 4 H), 1.21 (bs, 4 H); ^{13}C NMR (DMSO-d_6) δ 160.8, 134.2, 128.8, 123.0, 112.0, 49.3, 30.4, 25.3; IR (thin film) 3448 (m), 3144 (m), 2945 (m), 1691 (s), 1509 (s), 1413 (s), 1380 (s), 1319 (s), 1285 (s), 1255 (s), 1195 (m), 1104 (w), 916 (m); FABMS *m/e* 395.1190 (M + H, 395.1202 calcd. for $C_{16}H_{19}N_4O_8$).

***N*-Methyl-4-nitro-2-(carboxamidopropyl-3-dimethylamino) pyrrole.** A suspension of *N*-methyl-4-nitropyrrole-2-carboxylic acid (4.0 g, 23.5 mmol) in thionyl chloride (10 mL, 137 mmol) was heated under reflux for 8 h. Excess thinoyl chloride was removed *in vacuo* and the acid chloride was dissolved in DMF (10 mL) and cooled to 0 °C. 3-Dimethylpropylamine (6.0 mL, 46.7 mmol) was slowly added and the solution was allowed to warm to room temperature and stirred for 10 h. After addition of water (30 mL) the reaction mixture was partitioned between ethyl acetate (150 mL) and 10% $NaHCO_3$. The layers were separated and the aqueous layer was washed once more with ethyl acetate (100 mL). The combined organic fractions were dried ($MgSO_4$) and the solvent was removed under reduced pressure to afford the nitropyrrole (5.20 g, 88%) as a white solid. ^{1}H NMR ($CDCl_3$) δ 8.86 (b, 1 H), 7.54 (d, 1 H, J = 1.9 Hz), 6.92 (d, 1 H, J = 1.9 Hz), 4.02 (s, 3 H), 3.51 (q, 2 H, J = 6.7 Hz), 2.52 (t, 2 H, J = 7.1 Hz), 2.32 (s, 6 H), 1.74 (q, 2 H, J = 7.2 Hz); ^{13}C NMR ($CDCl_3$) δ 160.5, 135.3, 128.0, 126.3, 106.4, 59.3, 45.3, 40.2, 37.8, 24.7; IR (thin film) 3126 (m), 2955 (m), 2793 (m), 1651

(s), 1548 (s), 1504 (s), 1454 (m), 1306 (s), 1210 (m), 1175 (m), 1138 (m), 1106 (m), 1039 (m), 987 (w), 909 (m).

***N*-Methyl-4-amino-2-(carboxamidopropyl-3-dimethylamino) pyrrole.** A solution of nitropyrrole (4.00 g, 15.7 mmol) and Pd/C catalyst (10%, 600 mg) in DMF (30 mL) was hydrogenated (50 psi) in a shaker apparatus for 8 h. The mixture was filtered through celite to remove catalyst and the solvent was removed under reduced pressure. The dark oil was quickly purified by flash column chromatography (0 to 1 % ammonium hydroxide in methanol, gradient) to afford the product as a yellow oil (2.20 g, 62 %). 4-Aminopyrroles are not very stable and are used immediately after purification in subsequent reactions. ^{1}H NMR (DMSO-d_6) δ 8.00 (s, 1 H), 7.30 (b, 1 H), 6.30 (s, 1 H), 3.69 (s, 3 H), 3.18, (q, 2 H, J = 6.1 Hz), 2.95 (t, 2 H, J = 7.5 Hz), 2.67 (s, 6 H), 1.81 (b, 2 H).

***N,N'*-(1,n-Dialkane)bis 2-(4-carboxamide-2-(carboxamidopropyl-3-dimethylamino)pyrrole)-4-nitropyrrole 4a-d (Exemplified with 4d).** A solution of diacid **3d** (430 mg, 1.09 mmol) in thionyl chloride (10.0 mL, 137 mmol) was heated under reflux for 6 h. Excess thionyl chloride was removed by distillation and the crude acid chloride was dissolved in DMF (8.0 mL). To this was added a solution of *N*-Methyl-4-amino-2-carboxamidopropyl-3-dimethylaminopyrrole (880 mg, 3.46 mmol) in DMF (10.0 mL) and the resulting solution was allowed to stir at room temperature for 3 h. Methanol (2.0 mL) was added and the solvents were removed under reduced pressure. The crude residue was partitioned between ethyl acetate (100 mL) and 10 % sodium bicarbonate (100 mL), the layers were separated and the aqueous fraction was further washed with ethyl acetate (2 x 50 mL). The combined organic fractions were filtered through celite, dried (Na_2SO_4) and the solvent was removed under reduced pressure and dried to afford bis-dipyrrole **4d** as a yellow powder. If necessary, the product can be

purified by flash column chromatography (1 to 2% ammonium hydroxide in methanol, gradient) to afford the

4a: yield 65% (330 mg); ^{1}H NMR (DMSO-d_6) δ 10.23 (s, 2 H) 8.25 (d, 2 H, J = 1.7 Hz), 8.10 (b, 2 H), 7.60 (d, 2 H, J = 1.4 Hz), 7.16 (d, 2 H, J = 1.6 Hz), 6.74 (d, 2 H, J = 1.6 Hz), 4.45 (t, 4 H, J = 6.1 Hz), 3.16 (b, 4 H), 3.77 (s, 6 H), 2.24-2.29 (b, 2 H), 2.21 (t, 4 H, J=7.0 Hz), 2.10 (s, 12 H), 1.58 (m, 4 H); ^{13}C NMR (DMSO-d_6) δ 161.4, 157.1, 134.7, 128.3, 126.2, 123.8, 121.8, 118.4, 108.5, 104.1, 57.5, 47.3, 45.7, 39.4, 37.5, 36.5, 27.7; FABMS m/e 765.3796 (M + H, 765.3796 calcd. for $C_{35}H_{49}N_{12}O_8$).

4b: yield 86% (330 mg); ^{1}H NMR (DMSO-d_6) δ 10.24 (s, 2 H), 8.23 (d, 2 H, J = 1.8 Hz), 8.11 (t, 2 H, J = 5.5 Hz), 7.59 (d, 2 H, J = 1.5 Hz), 7.19 (d, 2 H, J = 1.7 Hz), 6.78 (d, 2 H, J = 1.7 Hz), 4.42 (b, 4 H), 3.79 (s, 6 H), 3.18 (m, 4 H), 2.21 (t, 4 H, J = 6.9 Hz), 2.10 (s, 12 H), 1.71 (b, 4 H), 1.58 (m, 4 H); ^{13}C NMR (DMSO-d_6) δ 161.4, 157.1, 134.5, 128.2, 126.3, 123.8, 121.8, 118.4, 108.5, 104.2, 57.6, 49.3, 45.7, 40.8, 36.5, 28.3, 27.7; FABMS m/e 779.3984 (M + H, 779.3953 calcd. for $C_{36}H_{51}N_{12}O_8$).

4c: yield 70% (235 mg); ^{1}H NMR (DMSO-d_6) δ 10.24 (s, 2 H), 8.20 (d, 2 H, J = 1.9 Hz), 8.12 (t, 2 H, J = 5.8 Hz), 7.59 (d, 2 H, J= 1.9 Hz), 7.19 (d, 2 H, J = 1.8 Hz), 6.79 (d, 2 H, J = 1.8 Hz), 4.37 (t, 4 H, J = 6.9 Hz), 3.78 (s, 6 H), 3.16 (m, 4 H), 2.20 (m, 4 H), 2.10 (s, 12 H), 1.70 (b, 4 H), 1.58 (m, 4 H), 1.20 (b, 2 H); ^{13}C NMR (DMSO-d_6) δ 161.4, 157.1, 134.5, 128.0, 126.1, 123.8, 121.7, 118.4, 108.5, 104.2, 57.6, 49.6, 45.7, 37.4, 36.5, 30.9, 27.7, 23.2; FABMS m/e 793.4147 (M + H, 793.4109 calcd. for $C_{37}H_{53}N_{12}O_8$).

4d: yield 56% (490 mg); ^{1}H NMR (DMSO-d_6) δ 10.24 (s, 2 H), 8.18 (d, 2 H, J = 1.8 Hz), 8.12 (t, 2 H, J = 5.4 Hz), 7.55 (d, 2 H, J = 1.6 Hz), 7.18 (d, 2 H, J = 1.6 Hz), 6.79 (d, 2 H, J = 1.7 Hz), 4.36 (t, 4 H, J = 6.7 Hz), 3.78 (s, 6 H), 3.16 (m, 4 H), 2.20 (t, 4 H, J = 7.0 Hz), 2.10 (s, 12 H), 1.70 (b, 4 H), 1.58 (m, 4 H), 1.22 (b, 4 H);

^{13}C NMR (DMSO-d_6) δ 161.5, 157.1, 134.4, 127.9, 126.2, 123.8, 121.8, 118.4, 108.4, 104.3, 57.6, 49.7, 45.7, 41.0, 36.5, 31.3, 27.7, 26.0; FABMS *m/e* 807.4224 (M + H, 807.4266 calcd. for $C_{38}H_{55}N_{12}O_8$).

Bis(pyridine-2-carboxamide-netropsin)$(CH_2)_{3-6}$ 1a-d (Exemplified with 1d). To a solution of picolinic acid (58 mg, 0.47 mmol) and *N*-hydroxybenzotriazole hydrate (60 mg, 0.44 mmol) in DMF (1.0 mL) was added a solution of 1,3-dicyclohexylcarbodiimide (93 mg, 0.45 mmol) in CH_2Cl_2 (1.0 mL). The solution was allowed to stir for 30 min at room temperature. Separately, to a solution of bis-dipyrrole **4d** (62 mg, 0.077 mmol) in DMF (2.0 mL) was added palladium on activated carbon (10 %, 37 mg) and this mixture was allowed to stir under a hydrogen atmosphere (300 psi) in a Parr bomb apparatus for 4 h. The reaction mixture was filtered through celite and added to the activated acid. The resulting solution was allowed to stir for 2 h and methanol (1.0 mL) was added. The solvents were removed under reduced pressure and the residue was partitioned between CH_2Cl_2 (50 mL) and 10% sodium bicarbonate (50 mL), the layers were separated and the aqueous fraction was washed once more with CH_2Cl_2 (50 mL). The combined organic layers were dried (Na_2SO_4) and the solvent was removed under reduced pressure. This residue was further purified by flash column chromatography (2% ammonium hydroxide in methanol) to afford the covalent peptide dimer **1d** as a yellow powder.

1a: yield 55% (38 mg); ^1H NMR ($CDCl_3$) δ 9.55 (s, 2 H), 8.42 (d, 2 H, *J* = 4.2 Hz), 8.23 (d, 2 H, *J* = 7.8 Hz), 8.08 (m, 4 H), 7.91 (dt, 2 H, *J* = 7.6 Hz, *J* = 1.7 Hz), 7.45 (dt, 2 H, *J* = 4.7 Hz, *J* = 1.1 Hz), 7.28 (d, 2 H, *J* = 1.6 Hz), 6.95 (d, 2 H, *J* = 1.6 Hz), 6.55 (d, 2 H, *J* = 1.7 Hz), 6.43 (d, 2 H, *J* = 1.7 Hz), 4.40 (b, 4 H), 3.87 (s, 6H), 3.48 (m, 2 H), 2.46 (t, 4 H, *J* = 6.4 Hz), 2.34 (bs, 2 H), 2.30 (s, 12 H), 1.79 (m, 4 H); ^{13}C NMR ($CDCl_3$) δ 162.1, 161.9, 158.7, 149.5, 148.0, 137.5, 126.2, 123.6, 123.5, 122.1, 121.7, 120.2, 118.5, 118.2, 105.8, 102.1, 58.7, 47.5 45.3, 39.1, 36.6, 31.0, 26.1; IR

(thin film) 3299 (m), 2940 (w), 1651 (s), 1584 (m), 1540 (s), 1464 (m), 1436 (m), 1403 (m), 1260 (m), 1118 (w), 1087 (w); UV (H_20) λ_{max} (ε) 244 (33 300), 306 (40 000) nm, (CH_3CN) λ_{max} (ε) 242 (38 400), 302 (46 100) nm; FABMS *m/e* 915.4704 (M + H, 915.4704 calcd. for $C_{47}H_{58}N_{14}O_6$).

1b: yield 54% (33 mg); ^1H NMR ($CDCl_3$) δ 9.56 (s, 2 H), 8.41 (d, 2 H, *J* = 4.7 Hz), 8.26 (d, 2 H, *J* = 2.2 Hz), 8.15 (t, 2 H, *J* = 4.8 Hz), 8.06 (s, 2 H), 7.93 (t, 2 H, *J* = 7.6 Hz), 7.47 (t, 2 H, *J* = 5.6 Hz), 7.31 (s, 2 H), 6.96 (s, 2 H), 6.59 (s, 2 H), 6.43 (s, 2H), 4.29 (bs, 4 H), 3.88 (s, 6 H), 3.48 (q, 4 H, *J* = 5.5 Hz), 2.44 (t, 4 H, *J* = 6.3 Hz), 2.33 (m, 2 H), 2.29 (s, 12 H), 1.77 (m, 4 H), 1.66 (bs, 4 H); ^{13}C NMR ($CDCl_3$) δ 162.0, 161.9, 158.6, 149.5, 148.0, 137.6, 126.2, 124.0, 123.6, 122.2, 121.6, 120.4, 118.6, 105.1, 101.9, 96.1, 58.8, 47.7, 45.4, 39.1, 36.7, 27.4, 26.2; IR (thin film) 3304 (m), 2942 (w), 1651 (s), 1588 (m), 1538 (s), 1463 (m), 1435 (m), 1403 (m), 1261 (m), 1120 (w), 1089 (w); UV (H_20) λ_{max} (ε) 242 (30 200), 308 (34 300) nm, (CH_3CN) λ_{max} (ε) 242 (34 100), 302 (42 800) nm; FABMS *m/e* 929.4919 (M + H, 929.4899 calcd. for $C_{48}H_{60}N_{14}O_6$).

1c: yield 43% (30 mg); ^1H NMR ($CDCl_3$) δ 9.59 (s, 2 H), 8.48 (d, 2 H, *J* = 4.8 Hz), 8.29 (d, 2 H, *J* = 7.9 Hz), 8.22 (s, 1 H), 8.14 (b, 1 H), 7.95 (td, 2 H, *J* = 7.7 Hz, *J* = 1.5 Hz), 7.49 (t, 2 H, *J* = 6.2 Hz), 7.35 (d, 2 H, *J* = 1.6 Hz), 7.03 (d, 2 H, *J* = 1.6 Hz), 6.57 (d, 2 H, *J* = 1.7 Hz), 6.53 (d, 2 H, *J* = 1.7 Hz), 4.29 (t, 4 H, *J* = 5.8 Hz), 3.91 (s, 3 H), 3.48 (q, 4 H, *J* = 5.0 Hz), 2.45 (t, 4 H, *J* = 6.4 Hz), 2.29 (s, 6 H), 1.77 (m, 2 H), 1.62 (m, 1 H), 0.83 (m, 1 H); ^{13}C NMR ($CDCl_3$) δ 161.9, 161.8, 158.8, 149.6, 148.1, 137.7, 126.3, 124.1, 123.6, 122.1, 121.8, 120.4, 118.5, 118.0, 104.5, 102.1, 58.9, 48.0, 45.4, 39.2, 36.7, 26.1, 21.5; IR (thin film) 3294 (m), 2943 (w), 1645 (s), 1578 (m), 1538 (s), 1464 (m), 1431 (m), 1404 (m), 1268 (m), 1120 (w), 1079 (w); UV (H_20) λ_{max} (ε) 244 (30 700), 302 (33 500) nm, (CH_3CN) λ_{max} (ε) 242 (35 400), 302 (42 000) nm; FABMS *m/e* 943.5032 (M + H, 943.5055 calcd. for $C_{49}H_{62}N_{14}O_6$).

1d: yield 61% (45 mg); ^{1}H NMR ($CDCl_3$) δ 9.70 (s, 2 H), 8.51 (d, 2 H, *J* = 4.0 Hz), 8.26 (d, 2 H, *J* = 7.8 Hz), 8.17 (s, 2 H), 7.97 (t, 2 H, *J* = 4.5 Hz), 7.89 (td, 2 H, *J* = 7.7 Hz, *J* = 1.3 Hz), 7.45 (t, 2 H, *J* = 6.0 Hz), 7.29 (d, 2 H, *J* = 1.6 Hz), 7.16 (d, 2 H, *J* = 1.6 Hz), 6.73 (d, 2 H, *J* = 1.6 Hz), 6.54 (d, 2 H, *J* = 1.6 Hz), 4.32 (t, 4 H, *J* = 6.0 Hz), 3.89 (s, 6 H), 3.45 (q, 2 H, *J* = 5.8 Hz), 2.44 (t, 4 H, *J* = 6.4 Hz), 2.33 (bs, 2 H), 2.29 (s, 12 H), 1.75 (m, 4 H), 1.68 (b, 4 H), 1.10 (b, 4 H); ^{13}C NMR ($CDCl_3$) δ 161.8, 161.5, 158.8, 149.5, 148.0, 137.7, 126.2, 123.5, 123.1, 122.1, 121.6, 120.7, 118.6, 117.9, 104.1, 102.5, 58.7, 48.9, 45.3, 39.1, 36.6, 30.9, 26.4, 26.0; IR (thin film) 3313 (m), 2940 (w), 1652 (s), 1584 (m), 1538 (s), 1464 (m), 1435 (m), 1404 (m), 1262 (m), 1121 (w), 1088 (w); UV (H_20) λ_{max} (ε) 244 (30 100), 304 (34 200) nm, (CH_3CN) λ_{max} (ε) 242 (35 400), 302 (40 000) nm; FABMS *m/e* 957.5181 (M + H, 957.5212 calcd. for $C_{50}H_{64}N_{14}O_6$).

***N,N'*-(1,n-Oligoethylene oxide)bis-2-carboxymethyl-4-nitropyrrole 8a-d (Exemplified with 8c).** To a solution of 2-carboxy-4-nitropyrrole methyl ester (265 mg, 1.56 mmol) in acetone (10.0 mL) was added potassium carbonate (500 mg, 3.63 mmol) and the resulting solution was allowed to stir at room temperature for 1 h. The bistosylate (400 mg, 0.733 mmol) was added and the reaction mixture was heated to 65 °C for 6 h. The solution was cooled and solvent was removed under reduced pressure. The residue was taken up in chloroform (15 mL), filtered through celite, concentrated and purified by flash column chromatography (2% MeOH in CH_2Cl_2) to afford **8c**.

8a: yield 97% (286 mg); ^{1}H NMR ($CDCl_3$) δ 7.78 (d, 1 H, *J* = 1.9 Hz), 7.43 (d, 1 H, *J* = 1.8 Hz), 4.57 (t, 2 H *J* = 5.1 Hz), 3.87 (s, 3 H), 3.77 (t, 2 H *J* = 5.2 Hz), 3.53 (s, 4 H); IR (thin film) 3136 (m), 2956 (m), 2872 (m), 1728 (s), 1714 (s), 1538 (s), 1505 (s), 1312 (m), 1261 (m), 1210 (m), 1177 (m), 1110 (m), 1085 (m), 923 (m), 822 (m), 750 (m); FABMS *m/e* 472.1680 (M + H, 472.1671 calcd. for $C_{18}H_{26}N_5O_{10}$).

8b: yield 91% (350 mg); ^{1}H NMR ($CDCl_3$) δ 7.82 (d, 1 H, *J* = 1.9 Hz), 7.42 (d, 1 H, *J* = 1.9 Hz), 4.57 (t, 2 H *J* = 5.0 Hz), 3.86 (s, 3 H), 3.79 (t, 2 H *J* = 5.2 Hz), 3.53 (m, 4 H); IR (thin film) 3137 (m), 2954 (m), 2872 (m), 1716 (s), 1540 (m), 1506 (s), 1422 (m), 1318 (s), 1264 (m), 1212 (m), 1174 (m), 1115 (s), 925 (w), 824 (m), 751 (m); FABMS *m/e* 516.1942 (M + H, 516.1937 calcd. for $C_{20}H_{30}N_5O_{11}$).

8c: yield 94% (385 mg); ^{1}H NMR ($CDCl_3$) δ 7.83 (d, 1 H, *J* = 1.9 Hz), 7.42 (d, 1 H, *J* = 1.9 Hz), 4.57 (t, 2 H *J* = 5.1 Hz), 3.86 (s, 3 H), 3.78 (t, 2 H *J* = 5.1 Hz), 3.61 (s, 4 H), 3.58 (s, 2 H); IR (thin film) 3137 (m), 2943 (m), 2872 (m), 1718 (s), 1540 (m), 1508 (s), 1420 (m), 1320 (s), 1261 (m), 1212 (m), 1179 (w), 1115 (s), 942 (w), 825 (m), 751 (m); FABMS *m/e* 560.2204 (M + H, 560.2183 calcd. for $C_{22}H_{34}N_5O_{12}$).

8d: yield 84% (387 mg); ^{1}H NMR ($CDCl_3$) δ 7.83 (d, 1 H, *J* = 1.8 Hz), 7.43 (d, 1 H, *J* = 1.9 Hz), 4.57 (t, 2 H *J* = 5.1 Hz), 3.86 (s, 3 H), 3.79 (t, 2 H *J* = 5.1 Hz), 3.63 (m, 4 H), 3.59 (s, 4 H); IR (thin film) 3137 (m), 2943 (m), 2870 (m), 1719 (s), 1538 (m), 1508 (s), 1418 (m), 1317 (s), 1260 (m), 1210 (m), 1179 (w), 1113 (s), 939 (w), 826 (m), 751 (m); FABMS *m/e* 604.2466 (M + H, 604.2445 calcd. for $C_{24}H_{38}N_5O_{13}$).

***N,N'*-(1,n-Oligoethylene oxide)bis-4-nitropyrrole-2-carboxylic acid 9a-d (Exemplified with 9a).** To a flask charged with bispyrrolemethylester **8a** (255 mg, 0.512 mmol) was added ethanol (4.0 mL) and 0.5 N lithium hydroxide (4.0 mL). The mixture was heated to 50 °C and allowed to stir 12 h. Solvent was removed *in vacuo* and the diacid was purified by routine anion exchange chromatography (DEAE A-25 resin, 150-350 mM ammonium bicarbonate, gradient). The desired fractions were pooled and the solvent was removed *in vacuo*. The residue was redissolved in water (5 mL) and the pH was adjusted to pH 1-2 with 1 N HCl. Extraction with ethyl acetate (3 x 15 mL) and concentration of the organic fractions afforded **9a**.

9a: yield 49% (108 mg); ^{1}H NMR (DMSO-d_6) δ 7.83 (d, 1 H, J = 1.9 Hz), 7.43 (d, 1 H, J = 1.9 Hz), 4.49 (t, 2 H J = 5.1 Hz), 3.67 (t, 2 H J = 5.2 Hz), 3.46 (s, 2 H); FABMS *m/e* 425.0945 (M + H, 425.0927 calcd. for $C_{16}H_{17}N_4O_{10}$).

9b: yield 67% (160 mg); ^{1}H NMR (DMSO-d_6) δ 8.13 (d, 1 H, J = 1.9 Hz), 7.25 (d, 1 H, J = 1.9 Hz), 4.50 (t, 2 H J = 5.1 Hz), 3.68 (t, 2 H J = 5.1 Hz), 3.46 (m, 2 H), 3.40 (m, 2 H); FABMS *m / e* 469.1207 (M + H, 469.1220 calcd. for $C_{18}H_{21}N_4O_{11}$).

9c: yield 81% (250 mg); ^{1}H NMR (DMSO-d_6) δ 8.15 (d, 1 H, J = 1.9 Hz), 7.26 (d, 1 H, J = 1.9 Hz), 4.51 (t, 2 H J = 5.1 Hz), 3.69 (t, 2 H J = 5.1 Hz), 3.47 (m, 2 H), 3.43 (m, 2 H), 3.41 (s, 2 H); FABMS *m/e* 513.1469 (M + H, 513.1447 calcd. for $C_{20}H_{25}N_4O_{12}$).

9d: yield 69% (262 mg); ^{1}H NMR (DMSO-d_6) δ 8.15 (d, 1 H, J = 1.9 Hz), 7.26 (d, 1 H, J = 1.9 Hz), 4.52 (t, 2 H J = 5.0 Hz), 3.69 (t, 2 H J = 5.0 Hz), 3.47 (m, 2 H), 3.44 (m, 2 H), 3.43 (s, 4 H); FABMS *m/e* 557.1731 (M + H, 557.1740 calcd. for $C_{22}H_{29}N_4O_{13}$).

***N,N'*-(1,n-Oligoethylene oxide)bis 2-(4-carboxamide-2-(carboxamidopropyl-3- dimethylamino)pyrrole)-4-nitropyrrole 10a-d (Exemplified with 10d).** A solution of diacid **9d** (140 mg, 0.251 mmol) in thionyl chloride (5.0 mL, 69 mmol) was heated under reflux for 6 h. Excess thionyl chloride was removed by distillation and the acid chloride was dissolved in dimethylformamide (5.0 mL). To this was added a solution of *N*-Methyl-4-amino-2-carboxamidopropyl-3-dimethylaminopyrrole (160 mg, 0.629 mmol) in dimethylformamide (2.0 mL) and the resulting solution was allowed to stir at room temperature for 2 h. Methanol (2.0 mL) was added and the solvents were removed *in vacuo*. The crude residue was slurried in 10% sodium bicarbonate (50 mL) and extracted with ethyl acetate (3 x 50 mL). The combined organic fractions were dried (Na_2SO_4) and concentrated to afford bis-dipyrrole **10d**. If

necessary, the product can be purified by flash column chromatography (1 to 2% ammonium hydroxide in methanol, gradient).

10a: yield 95% (168 mg); ^{1}H NMR ($CDCl_3$) δ 8.98 (s, 1 H), 7.84 (t, 1 H, *J* = 6.3 Hz), 7.67 (d, 1 H, *J* = 1.9 Hz), 7.26 (d, 1 H, *J* = 1.8 Hz), 7.18 (d, 1 H, *J* = 1.8 Hz), 6.57 (d, 1 H, *J* = 1.9 Hz), 4.55 (t, 2 H *J* = 4.5 Hz), 3.88 (s, 3 H), 3.74 (t, 2 H *J* = 4.7 Hz), 3.53 (s, 2 H), 3.44 (m, 2 H), 2.43 (t, 2 H, *J* = 6.1 Hz), 2.26 (s, 3 H), 1.71 (m, 2 H); ^{13}C NMR ($CDCl_3$) δ 161.6, 157.6, 134.8, 127.4, 125.9, 123.9, 121.0, 118.6, 107.6, 103.2, 76.6, 70.2, 70.0, 49.3, 45.3, 39.5, 36.5, 25.7; IR (thin film) 3290 (m), 3128 (w), 2948 (m), 2863 (w), 2825 (w), 2783 (w), 1644 (s), 1574 (m), 1538 (s), 1504 (s), 1434 (m), 1312 (s), 1117 (m); FABMS *m/e* 839.4164 (M + H, 839.4215 calcd. for $C_{38}H_{55}N_{12}O_{10}$).

10b: yield 91% (95 mg); ^{1}H NMR ($CDCl_3$) δ 9.01 (s, 1 H), 7.90 (t, 1 H, *J* = 6.3 Hz), 7.72 (s, 1 H), 7.25 (s, 1 H), 7.18 (s, 1 H), 6.57 (s, 1 H), 4.56 (t, 2 H *J* = 4.3 Hz), 3.87 (s, 3 H), 3.77 (t, 2 H *J* = 4.5 Hz), 3.52 (s, 4 H), 3.45 (m, 2 H), 2.43 (t, 2 H, *J* = 6.0 Hz), 2.26 (s, 6 H), 1.71 (m, 2 H); ^{13}C NMR ($CDCl_3$) δ 161.7, 157.6, 134.8, 127.5, 126.0, 123.9, 121.0, 118.6, 107.6, 103.2, 76.6, 70.5, 70.0, 69.9, 49.4, 45.3, 39.5, 36.5, 25.6; IR (thin film) 3290 (m), 3128 (w), 2948 (m), 2863 (w), 2825 (w), 2773 (w), 1644 (s), 1574 (m), 1538 (s), 1504 (s), 1434 (m), 1312 (s), 1117 (m); FABMS *m/e* 883.4426 (M + H, 883.4406 calcd. for $C_{40}H_{59}N_{12}O_{11}$).

10c: yield 89% (225 mg); ^{1}H NMR ($CDCl_3$) δ 9.28 (s, 1 H), 7.83 (t, 1 H, *J* = 6.0 Hz), 7.72 (s, 1 H), 7.33 (s, 1 H), 7.20 (s, 1 H), 6.61 (s, 1 H), 4.58 (t, 2 H *J* = 4.6 Hz), 3.88 (s, 3 H), 3.78 (t, 2 H *J* = 4.5 Hz), 3.57 (s, 4 H), 3.55 (s, 2 H), 3.51 (m, 2 H), 2.43 (t, 2 H, *J* = 6.1 Hz), 2.26 (s, 6 H), 1.70 (m, 2 H); ^{13}C NMR ($CDCl_3$) δ 161.6, 157.7, 134.7, 127.1, 126.2, 123.7, 121.1, 118.5, 107.6, 103.1, 76.6, 70.3, 70.1, 70.0, 58.6, 49.3, 45.1, 39.2, 36.4, 25.7; IR (thin film) 3284 (m), 3127 (w), 2947 (m), 2872 (w), 2825 (w), 2773 (w), 1651 (s), 1574 (m), 1538 (s), 1504 (s), 1440 (m), 1313 (s), 1116 (m); FABMS *m/e* 927.4688 (M + H, 927.4722 calcd. for $C_{42}H_{63}N_{12}O_{12}$).

10d: yield 95% (230 mg); ^{1}H NMR ($CDCl_3$) δ 8.88 (s, 1 H), 7.81 (t, 1 H, *J* = 6.0 Hz), 7.66 (d, 1 H, *J* = 1.9 Hz), 7.22 (d, 1 H, *J* = 1.9 Hz), 7.20 (d, 1 H, *J* = 1.8 Hz), 6.52 (d, 1 H, *J* = 1.9 Hz), 4.54 (t, 2 H *J* = 5.2 Hz), 3.87 (s, 3 H), 3.68 (t, 2 H *J* = 5.0 Hz), 3.54 (m, 2 H), 3.52 (m, 2 H), 3.51 (m, 2 H), 3.50 (s, 4 H), 3.48 (m, 2 H), 2.45 (t, 2 H, *J* = 6.2 Hz), 2.28 (s, 6 H), 1.73 (m, 2 H); ^{13}C NMR ($CDCl_3$) δ 161.6, 157.6, 134.7, 127.0, 126.2, 123.6, 121.0, 118.5, 107.5, 103.1, 76.6, 70.3, 70.1, 70.0, 59.1, 52.2, 49.3, 45.1, 39.2, 36.4, 25.7; IR (thin film) 3294 (m), 3129 (w), 2946 (m), 2865 (m), 2825 (w), 2773 (w), 1643 (s), 1574 (m), 1538 (s), 1505 (s), 1440 (m), 1312 (s), 1115 (m); FABMS *m/e* 971.4951 (M + H, 971.4974 calcd. for $C_{44}H_{67}N_{12}O_{13}$).

Bis(pyridine-2-carboxamide-netropsin)(CH_2OCH_2)$_{2\text{-}5}$ 7a-d (Exemplified with 7a). To a solution of picolinic acid (55 mg, 0.45 mmol) and *N*-hydroxybenzotriazole hydrate (61 mg, 0.45 mmol) in dimethylformamide (1.0 mL) was added a solution of 1,3-dicyclohexylcarbodiimide (93 mg, 0.45 mmol) in methylene chloride (1.0 mL). The solution was allowed to stir for 30 min at room temperature. Separately, to a solution of bis-dipyrrole **10d** (85 mg, 0.101 mmol) in dimethylformamide (4.0 mL) was added palladium on activated carbon (10%, 70 mg) and this mixture was allowed to stir under a hydrogen atmosphere (400 psi) in a Parr bomb apparatus for 6 h. The reaction mixture was filtered through celite to remove catalyst and added to the activated acid. The resulting solution was allowed to stir for 7 h and methanol (1.0 mL) was added. The solvents were removed *in vacuo* and the residue was purified by flash column chromatography (0 to 1% ammonium hydroxide in methanol, gradient) to afford the covalent peptide dimer **7d**.

7a: yield 48% (48 mg); ^{1}H NMR ($CDCl_3$) δ 9.80 (s, 1 H), 8.49 (d, 1 H, *J* = 4.8 Hz), 8.38 (s, 1 H), 8.22 (d, 1 H, *J* = 7.7 Hz), 7.85 (td, 1 H, *J* = 7.7 Hz, *J* = 1.6 Hz), 7.73 (bs, 1 H), 7.39 (m, 1 H), 7.36 (d, 1 H, *J* = 1.6 Hz), 7.23 (d, 1 H, *J* = 1.5 Hz), 6.93 (d, 1 H, *J* = 1.4 Hz), 6.58 (d, 1 H, *J* = 1.6 Hz), 4.45 (t, 2 H *J* = 4.5 Hz), 3.87 (s, 3 H),

3.66 (t, 2 H J = 4.8 Hz), 3.49 (s, 2 H), 3.44 (m, 2 H), 2.48 (t, 2 H, J = 6.4 Hz), 2.31 (s, 3 H), 1.76 (m, 2 H); ^{13}C NMR ($CDCl_3$) δ 161.8, 161.4, 158.9, 149.6, 148.0, 137.5, 126.1, 123.5, 122.1, 121.5, 118.6, 104.4, 103.0, 71.1, 70.7, 58.3, 48.6, 45.1, 38.7, 36.6, 26.0; IR (thin film) 3286 (m), 2948 (m), 2867 (w), 2825 (w), 2771 (w), 1645 (s), 1587 (m), 1553 (s), 1515 (s), 1455 (m), 1435 (m), 1403 (m), 1264 (m), 1123 (m); FABMS *m/e* 989.4164 (M + H, 989.4215 calcd. for $C_{50}H_{65}N_{14}O_8$).

7b: yield 44% (32 mg); ^1H NMR ($CDCl_3$) δ 9.85 (s, 1 H), 8.56 (d, 1 H, J = 5.0 Hz), 8.30 (s, 1 H), 8.23 (d, 1 H, J = 7.8 Hz), 7.87 (td, 1 H, J = 7.7 Hz, J = 1.7 Hz), 7.72 (bs, 1H), 7.43 (m, 1 H), 7.39 (d, 1 H, J = 1.8 Hz), 7.22 (d, 1 H, J = 1.6 Hz), 6.99 (d, 1 H, J = 1.7 Hz), 6.54 (d, 1 H, J = 1.8 Hz), 4.50 (t, 2 H J = 4.6 Hz), 3.91 (s, 3 H), 3.79 (t, 2 H J = 4.7 Hz), 3.53 (s, 4 H), 3.48 (m, 2 H), 2.47 (t, 2 H, J = 6.4 Hz), 2.31 (s, 3 H), 1.76 (m, 2 H); ^{13}C NMR ($CDCl_3$) δ 161.7, 161.3, 158.9, 149.6, 148.0, 137.5, 126.2, 123.6, 123.4, 122.1, 121.4, 118.7, 118.5, 104.5, 102.9, 71.2, 70.7, 70.3, 58.7, 48.6, 45.3, 39.1, 36.6, 26.0; IR (thin film) 3313 (m), 2948 (m), 2863 (w), 2828 (w), 1652 (s), 1574 (m), 1532 (s), 1463 (m), 1435 (m), 1403 (m), 1261 (m), 1122 (m), 1096 (m); FABMS *m/e* 1033.5372 (M + H, 1033.5328 calcd. for $C_{52}H_{69}N_{14}O_9$).

7c: yield 53% (38 mg); ^1H NMR ($CDCl_3$) δ 9.82 (s, 1 H), 8.52 (d, 1 H, J = 5.1 Hz), 8.44 (bs, 1 H), 8.18 (d, 1 H, J = 7.8 Hz), 7.83 (td, 1 H, J = 7.7 Hz, J = 1.7 Hz), 7.72 (bs, 1 H), 7.49 (m, 1 H), 7.37 (d, 1 H, J = 1.6 Hz), 7.21 (d, 1 H, J = 1.6 Hz), 7.01 (bs, 1 H), 6.58 (bs, 1 H), 4.46 (t, 2 H J = 4.6 Hz), 3.88 (s, 3 H), 3.75 (t, 2 H J = 4.7 Hz), 3.51 (s, 6 H), 3.43 (m, 2 H), 2.48 (t, 2 H, J = 6.4 Hz), 2.31 (s, 3 H), 1.75 (m, 2 H); ^{13}C NMR ($CDCl_3$) δ 161.8, 161.3, 158.9, 149.6, 148.0, 137.5, 126.1, 123.5, 123.4, 122.1, 121.5, 121.4, 118.7, 118.6, 104.5, 103.0, 76.6, 71.1, 70.6, 70.4, 70.2, 58.5, 48.6, 45.2, 38.8, 36.5, 25.9; IR (thin film) 3307 (m), 2948 (m), 2846 (w), 2822 (w), 1652 (s), 1587 (m), 1532 (s), 1463 (m), 1435 (m), 1404 (m), 1262 (m), 1123 (m), 1093 (m); FABMS *m/e* 1077.5634 (M + H, 1077.5577 calcd. for $C_{54}H_{73}N_{14}O_{10}$).

7d: yield 25% (20 mg); ^{1}H NMR ($CDCl_3$) δ 9.82 (s, 1 H), 8.54 (d, 1 H, *J* = 5.1 Hz), 8.27 (s, 1 H), 8.22 (d, 1 H, *J* = 7.8 Hz), 7.86 (td, 1 H, *J* = 7.7 Hz, *J* = 1.6 Hz), 7.71 (bs, 1 H), 7.41 (m, 1 H), 7.36 (d, 1 H, *J* = 1.7 Hz), 7.21 (d, 1 H, *J* = 1.6 Hz), 6.93 (d, 1 H, *J* = 1.6 Hz), 6.53 (d, 1 H, *J* = 1.7 Hz), 4.49 (t, 2 H *J* = 4.7 Hz), 3.89 (s, 3 H), 3.76 (t, 2 H *J* = 4.8 Hz), 3.54 (m, 4 H), 3.51 (m, 4 H), 2.44 (t, 2 H, *J* = 6.3 Hz), 2.29 (s, 3 H), 1.73 (m, 2 H); ^{13}C NMR ($CDCl_3$) δ 161.8, 161.3, 159.0, 149.7, 148.0, 137.5, 126.1, 123.6, 122.1, 121.4, 118.6, 118.5, 104.5, 102.9, 76.6, 71.2, 70.6, 70.5, 70.4, 70.3, 58.7, 48.6, 45.4, 39.1, 36.6, 26.0; IR (thin film) 3338 (m), 2942 (m), 2869 (w), 2821 (w), 1654 (s), 1588 (m), 1542 (s), 1459 (m), 1438 (m), 1400 (m), 1279 (m), 1122 (m), 1094 (m); FABMS *m/e* 1121.5896 (M + H, 1121.5920 calcd. for $C_{56}H_{77}N_{14}O_{11}$).

DNA Reagents and Materials. Doubly distilled water was further purified through the Milli Q filtration system from Millipore. Sonicated, deproteinized calf thymus DNA was purchased from Pharmacia. Plasmid pBR322 was obtained from Boehringer-Mannheim. Enzymes were obtained from Boehringer-Mannheim or New England Biolabs and used with the buffers supplied. Deoxyadenosine 5'-[α-^{32}P]triphosphate was obtained from Amersham. Storage phosphor technology autoradiography was performed using a Molecular Dynamics 400S Phosphorimager and ImageQuant software. The 517 base pair 3'-end labeled *Eco*R I / *Rsa* I restriction fragment from plasmid pBR322 was prepared and purified as previously described.[11] Chemical sequencing reactions were performed according to published methods.[27,28] Standard techniques were employed for DNA manipulations.[29] All other reagents and materials were used as received.

Sample Preparation. Milligram quantitities of peptide were placed in tared eppendorf tubes, dried at 1 torr for two days and immediately weighed. The peptides were dissolved in water and the extinction coefficients were determined from the absorbances measured on a Hewlett Packard 8452A diode

array spectrophotometer. Reported extinction coefficients represent the average of two determinations. Aqueous solutions of peptides were aliquoted into eppendorf tubes, lyophilized and stored at -20 °C. The peptides were dissolved and serially diluted before each set of experiments.

DNase I Footprinting. All reactions were executed in a total volume of 10 μL with final concentrations of each species as indicated. The ligands were added to solutions of radiolabeled restriction fragment (20,000 cpm), calf thymus DNA (100 μM bp), Tris •HCl (10 mM, pH 7.0), KCl (10 mM), $MgCl_2$ (10 mM) and $CaCl_2$ (5 mM) and incubated for 15 min at 22 °C. Footprinting reactions were initiated by the addition of 1 μL a stock solution of DNase I (10 units/mL) containing 1 mM dithiothreitol and allowed to proceed for 3 min at 22 °C. The reactions were stopped by addition of a 3 M ammonium acetate solution containing 250 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, 7 M urea) at 1000 V for 3-4 h.

Quantitative DNase I Footprint Titration. Apparent first order binding constants were determined by quantitative DNase I footprint titration as previously described.[20] The above reaction conditions were employed with ligand concentrations ranging from 200 mM to 100 nM. The footprint titration gels were dried and quantitated using storage phosphor technology. The data were analyzed by performing volume integrations of the 5′-TGTCA-3′ and 5′-TTTTT-3′ target sites and a 5′-GCGG-3′ reference site. The apparent DNA target site saturation, θ_{app}, was calculated for each concentration of peptide using the following equation:

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

where I_{tot} and I_{ref} are the integrated volumes of the target and reference sites, respectively, and I_{tot}° and I_{ref}° correspond to those values for a DNase I control lane to which no peptide has been added. At higher concentrations of peptide (> 50 μM), the reference sites become partially protected, resulting in low θ_{app} values. For these data points, the reference value was determined from the amount of radioactivity loaded per lane, using the mean value for all data points from lanes with < 50 μM peptide. The ($[L]_{tot}$, θ_{app}) data points were fit by minimizing the difference between θ_{app} and θ_{fit}, using the modified Hill equation:

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

where $[L]_{tot}$ corresponds to the total peptide concentration, K_a corresponds to the apparent monomeric association constant, and θ_{min} and θ_{max} represent the experimentally determined site saturation values when the site is unoccupied or saturated, respectively. Since 2-PyN binds to the 5′-TGTCA-3′ site as a dimer, the data were fit to eq. 2 with n = 2. For all other cases, a monomeric association was assumed (eq 2, n = 1).

Data were fit using a nonlinear least-squares fitting procedure of KaleidaGraph software (version 2.1, Abelbeck software) running on a Macintosh IIfx computer with K_a, θ_{max}, and θ_{min} as the adjustable parameters. All lanes from each gel were used. The data were normalized using the following equation:

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

The goodness of fit of the binding curve to the data points is evaluated by the correlation coefficient, with $R > .97$ as the criterion for an acceptable fit. Three sets of acceptable data were used in determining each association constant.

Quantitation by Storage Phosphor Technology Autoradiography. Photostimulable storage phosphor imaging plates (Kodak Storage Phosphor Screen S0230 obained from Molecular Dynamics) were pressed flat against gel samples and exposed in the dark at 22 °C for 15-20 h. 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 the ImageQuant v. 3.0 software running on an AST Premium 386/33 computer.

References

1. (a) Krylov, A. S.; Grokhovsky, S. L.; Zasedatelev, A. S.; Zhuze, A. L.; Gursky, G. V.; Gottikh, B. P. *Nuc. Acids. Res.* **1979**, *6*, 289-304. (b) Zasedatelev, A. S.; Gursky, G. V.; Zimmer, Ch.; Thrum, H. *Mol. Biol. Reports* **1974**, *1*, 337-342. (c) Zasedatelev, A. S.; Zhuze, A. L.; Zimmer, Ch.; Grokhovsky, S. L.; Tumanyan, V. G.; Gursky, G. V.; Gottikh, B. P. *Dokl. Acad. Nauk SSSR* **1976**, *231*, 1006-1009. For a review, see: (d) Zimmer, C.; Wåhnert, U. *Prog. Biophys. Molec. Biol.* **1986**, *47*, 31-112.

2. (a) Van Dyke, M. W.; Hertzberg, R. P.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 5470-5474. (b) Van Dyke, M. W.; Dervan, P. B. *Cold Spring Harbor Symposium on Quantitative Biology* **1982**, *47*, 347-353. (c) Van Dyke, M. W.; Dervan, P. B. *Biochemistry* **1983**, *22*, 2373-2377. (d) Harshman, K. D.; Dervan, P. B. *Nucl. Acids Res.* **1985**, *13*, 4825-4835.(e) Fox, K. R.; Waring, M. J. *Nucl. Acids Res.* **1984**, *12*, 9271-9285. (f) Lane, M. J.; Dobrowiak, J. C.; Vournakis, J. *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 3260-3264.

3. (a) Schultz, P. G.; Taylor, J. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1982**, *104*, 6861-6863. (b) Taylor, J. S.; Schultz, P. G.; Dervan, P. B. *Tetrahedron* **1984**, *40*, 457-465. (c) Schultz, P. G.; Dervan, P. B. *J. Biomol. Struct. Dyn.* **1984**, *1*, 1133-1147. (d) Dervan, P. B. *Science*, **1986**, *232*, 464-471.

4. (a) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 1376-1380. (b) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *J. Mol. Biol.* **1985**, *183*, 553-563. (c) Coll, M.; Frederick, C. A.; Wang, A. H.-J.; Rich, A. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 8385-8389.

5. (a) Patel, D. J.; Shapiro, L. *J. Biol. Chem.* **1986,** *261,* 1230-1240. (b) Klevitt, R. E.; Wemmer, D. E.; Reid, B. R. *Biochemistry* **1986,** *25,* 3296-3303. (c) Pelton, J. G., Wemmer, D. E. *Biochemistry* **1988,** *27,* 8088-8096.

6. (a) Markey, L. A.; Breslauer, K. J. *Proc. Natl. Acad. Sci. USA* **1987,** *84,* 4359-4363. (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,* 8922-8926.

7. Dervan, P. B. *Science* **1986,** *232,* 464-471.

8. (a) Lown, J. W.; Krowicki, K.; Bhat, U. G.; Ward, B.; Dabrowiak, J. C. *Biochemistry* **1986,** *25,* 7408-7416. (b) Kissinger, K.; Krowicki, K.; Dabrowiak, J. C.; Lown, J. W. *Biochemistry* **1987,** *26,* 5590-5595.

9. Wade, W. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1987,** *109,* 1574-1575.

10. (a) Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci. USA* **1989,** *86,* 5723-5727. (b) Pelton, J. G.; Wemmer, D. E. *J. Am. Chem. Soc.* **1990,** *112,* 1393-1399.

11. Wade, W.S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992,** *114,* 8783-8794.

12. Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Dervan, P. B. *Proc. Natl. Acad. Sci., USA* **1992,** *89,* 7586-7590.

13. Dwyer, T. J.; Geierstanger, B. H.; Bathini, Y.; Lown, J. W.; Wemmer, D. E. *J. Am. Chem. Soc.* **1992,** *114,* 5911-5919.

14. Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993,** *115,* 2572-2576.

15. Geierstanger, B. H.; Mrksich, M.; Jacobsen, J-P.; Dervan, P. B.; Wemmer, D. E. *Biochemistry* **1994,** in press.

16. Wade, W. S.; Mrksich, M.; Dervan, P. B. *Biochemistry,* **1993,** *32,* 11385-11389.

17. For examples of end-to-end linked peptides, see: (a) Schultz, P. G.; Dervan, P. B. *J. Am. Chem. Soc.* **1983**, *105*, 7748-7750. (b) Youngquist, R. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1985**, *107*, 5528-5529. (c) Griffin, J. H.; Dervan, P. B. *J. Am. Chem. Soc.* **1986**, *108*, 5008, 5009. (d) Griffin, J. H.; Dervan, P. B. *J. Am. Chem. Soc.* **1987**, *109*, 6840-6842. (e) Youngquist, R. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1987**, *109*, 7564-7566.

18. Dwyer, T. J.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, *115*, 9900-9906.

19. (a) Galas, D.; Schmitz, A. *Nucl. Acids. Res.* **1978**, *5*, 3157-3170. (b) Fox, K. R.; Waring, M. J. *Nucl. Acids Res.* **1984**, *12*, 9271-9285.

20. (a) Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Methods Enzymol.* **1986**, *130*, 132-181. (b) Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 8462-8466. (c) Senear, D. F.; Brenowitz, M.; Shea, M. A.; Ackers, G. K. *Biochemistry* **1986**, *25*, 7344-7354.

21. (a) Gao, X.; Patel, D. J. *Biochemistry* **1989**, *28*, 751-762. (b) Gao, X.; Mirau, P.; Patel, D. J. *J. Mol. Biol.* **1992**, *223*, 259-279.

22. Hecht, S. M. *Acc. Chem. Res.* **1986**, *19*, 385.

23. (a) Mack, D. P.; Dervan, P. B. *J. Am. Chem. Soc.* **1990**, *112*, 4604. (b) Mack, D. P.; Dervan, P. B. *Biochemistry* **1992**, *31*, 9399.

24. Griffin, J. H.; Dervan, P. B. *J. Am. Chem. Soc.* **1987**, *109*, 6840-6842.

25. Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 9892-9899.

26. Still, W. C.; Kahn, M.; Mitra, A. *J. Org. Chem.* **1978**, *40*, 2923-2925.

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

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

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

Chapter 4

Design of Peptides for Sequence-Specific Recognition of the Minor Groove of DNA at 5′-TGTTA-3′ Sites

1:1 Peptide-DNA Complexes. The natural products netropsin and distamycin A (D) are crescent shaped di- and tripeptides, respectively, that bind in the minor groove of DNA at sites of four or five successive A,T base pairs.[1-3] The structures of a number of peptide-DNA complexes have been determined by X-ray diffraction[4] and NMR spectroscopy,[5] and the thermodynamic profiles have been studied for these complexes.[6] Analyses of these 1:1 peptide-DNA complexes have led to the conclusion that hydrogen bonding, van der Waals contacts and electrostatics all contribute to the binding affinity and specificity. It has been found that the width of the minor groove of these A,T-rich sequences is narrower than that of standard B-DNA.[7] The two (or three) *N*-methylpyrrole-carboxamides twist in a screw sense to match the walls of the minor groove, giving a favorable shape complementarity for the ligand, while the carboxamide NH's participate in bifurcated hydrogen bonds to adenine N3 and thymidine O2 atoms on the floor of the minor groove.

Peptides designed for binding mixed sequences. Early efforts in designing peptides for the recognition of G,C base pairs were based on 1:1 peptide-DNA complexes and involved incorporating hydrogen bond acceptor atoms on ligands which could form specific hydrogen bonds to amino groups of guanine on the floor of the minor groove.[8-10] Distamycin analogs which substitute heteroatoms for the pyrrole C3 (e.g., imidazole and thiazole) display increased tolerance for G,C base pairs in their binding sites, but with an overall

loss of specificity.[9] Remarkably, two synthetic peptides based on this 1:1 rationale display altered specificity for mixed sequences. Pyridine-2-carboxamide-netropsin (2-PyN) and 1-methylimidazole-2-carboxamide-netropsin (2-ImN) bind specifically to 5'-(A,T)G(A,T)C(A,T)-3' sequences.[10, 11] However, 1:1 models of 2-PyN•DNA and 2-ImN•DNA complexes seemed inadequate for rationalizing the recognition of G,C base pairs in both the second and fourth positions of the binding site.[10]

2:1 Peptide-DNA Complexes. Two-dimensional NMR studies by Wemmer and Pelton have revealed that distamycin at 2-4 mM concentrations is capable of binding in the minor groove of an 5′-AAATT-3′ sequence as a side-by-side antiparallel dimer (Figure 1A).[12] An antiparallel dimer would explain the sequence specificity and orientation of $(2\text{-PyN})_2$•5′-TGACT-3′ and $(2\text{-ImN})_2$•5′-TGACT-3′ observed from footprinting and affinity cleaving data.[11] Indeed, the 2-ImN•5'-TGACT-3' complex has now been characterized as a 2:1 complex directly by NMR spectroscopy (Figure 1B).[13] The structure reveals that the two peptides bind in the minor groove as a side-by-side antiparallel dimer. In contrast to the distamycin homodimer, 2-ImN binds the 5′-TGACT-3′ sequence as a 2:1 complex at low ligand-DNA stoichiometries, indicating that the two peptides bind with positive cooperativity.[13] Simultaneously, Wemmer, Lown and co-workers have characterized by NMR a distamycin analog containing a central imidazole in complex with a 5′-AAGTT-3′ sequence as a side-by-side antiparallel dimer.[14, 15] These 2:1 models for peptide-DNA complexes differ in many respects from the well-characterized early 1:1 models observed for binding of distamycin and netropsin to A,T-rich DNA. The minor groove width in the 2:1 complex is likely twice that observed in 1:1 models. With regard to specific hydrogen bonds in 1:1 complexes, distamycin and netropsin carboxamide NH's appear to form bifurcated hydrogen bonds to bases on *both strands of the DNA*. In

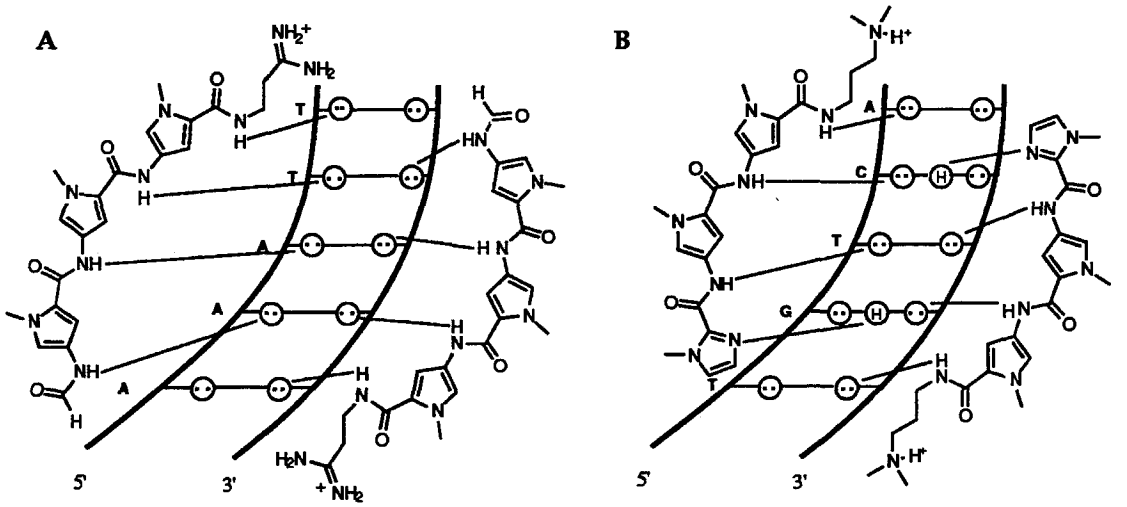


Figure 4.1. (A) 2:1 Binding model for the complex formed between distamycin with a 5′-AAATT-3′ sequence.[12] (B) Homodimeric binding model for the complex formed upon binding of 2-ImN with a 5′-TGTCA-3′ sequence.[13] Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by the solid lines.

the 2:1 motif, each ligand likely participates in hydrogen bonds only with bases on a *single strand* (Figure 1).

Design Rationale. Inspection of these 2:1 *homodimeric* models suggests that a *heterodimer* consisting of the two peptides D and 2-ImN may specifically bind to the family of sequences 5'-(A,T)G(A,T)$_3$-3'. If one 2-ImN ligand in the (2-ImN)$_2$•5'-TGACT-3' complex is replaced by a distamycin ligand, the absence of the imidazole N3 atom should favor an A,T base pair relative to a G,C base pair. In the proposed heterodimeric complex, the 2-ImN ligand forms a specific hydrogen bond to the guanine 2-amino group, followed by hydrogen bonds between the three carboxamide NH's with adenine N3 and thymine O2 atoms (Figure 2). The four carboxamide NH's of the distamycin ligand are expected to participate in hydrogen bonds with adenine N3, thymine O2, and cytosine O2

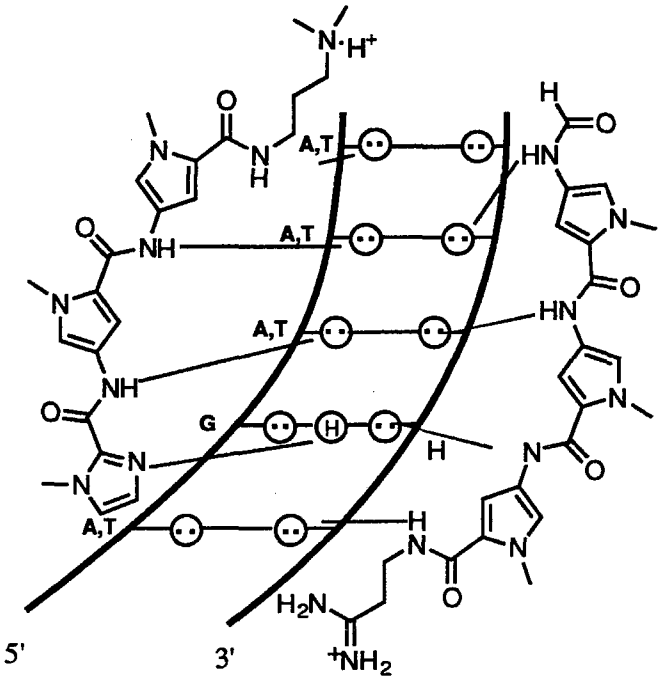


Figure 4.2. Heterodimeric binding model for the complex formed between 2-ImN and D with a 5′-TGTTA-3′ sequence. Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dotted lines.

atoms. Overall, the *heterodimeric* complex can be thought of as a *hybrid* of the distamycin and 2-ImN *homodimeric* complexes.

Experimental Design. Footprinting and affinity cleaving experiments can be used to characterize this proposed complex. Neither D nor 2-ImN alone should protect a 5'-(A,T)G(A,T)$_3$-3' site from cleavage by the footprinting agent MPE•Fe(II).[2a-d] When both ligands are present, however, binding to this site by the heterodimer should protect the site from cleavage by MPE•Fe(II). Affinity cleaving experiments would verify the requirement for both ligands in the complex and in addition, define the *orientation* of each peptide in the proposed complex. Sequence specific cleavage by 2-ImNE•Fe(II) proximal to the

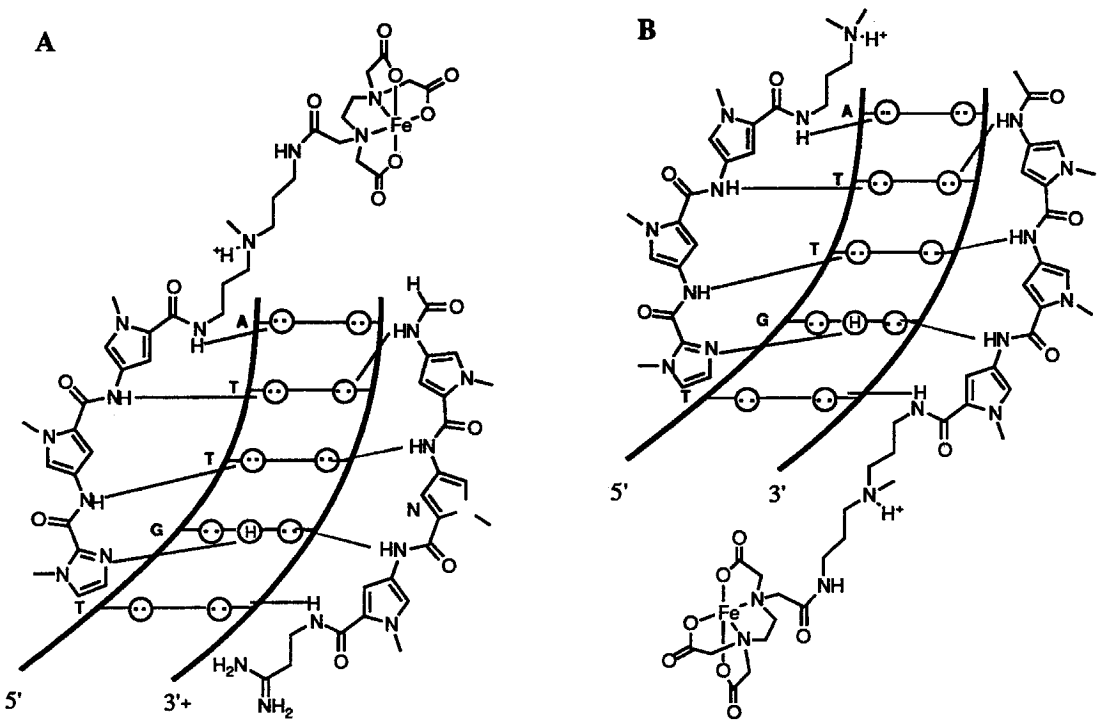


Figure 4.3. Experimental design for affinity cleaving experiments to determine the orientations of the peptides in the heterodimeric complex. 2:1 models for the heterodimeric complexes formed between (A) 2-ImNE•Fe(II) and D, and (B) ED•Fe(II) and 2-ImN with the 5′-TGTTA-3′ sequence. Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Proposed hydrogen bonds are illustrated by dashed lines.

heterodimer site should occur only in the presence of D. Furthermore, the cleavage pattern is expected to be proximal to the A,T-rich end of the 5'-(A,T)G(A,T)$_3$-3' site, consistent with the orientation of 2-ImN which allows for a hydrogen bond to the guanine amino group (Figure 3A). Similarly, ED•Fe(II) is expected to effect sequence specific cleavage only in the presence of 2-ImN. In a formal sense, D could bind in either a parallel or antiparallel orientation with respect to 2-ImN.[16] If the antiparallel heterodimer is favored, cleavage is expected to be near the 5′ end of the 5′-TGTTA-3′ strand of the binding site

(Figure 3B). If the parallel heterodimer is favored, cleavage is expected to be near the 3′ end of the 5′-TGTTA-3′ binding site.

Results

Heterodimer Site Identification. Candidate heterodimer sites were identified by affinity cleavage experiments with ED•Fe(II) in the presence of varying concentrations of 2-ImN on radiolabeled restriction fragments from the plasmid pBR322 (as in Figure 3B). While a well resolved site was not located, one strong site was found which overlapped a distamycin binding site. This region of cleavage corresponded to base pairs 90-100 on pBR322.[17] Two 34-base pair oligonucleotides corresponding to this region, but with the overlapping distamycin site deleted, were prepared and used for footprinting and affinity cleaving experiments.[18]

Footprinting and Affinity Cleaving. Neither D at 2μM nor 2-ImN at 100μM alone protects any sites on the 34-base pair oligonucleotide from cleavage by the footprinting agent MPE•Fe(II) (pH 7.6 in 40 mM Tris acetate and 20 mM NaCl at 22 °C) (Figure 4, lanes 4-6). The presence of both, D at 2μM and 2-ImN at 100μM, clearly protects the 5′-TGTTA-3′ site from cleavage by MPE•Fe(II) (Figure 4, lane 7, Figure 6a). Affinity cleavage experiments verify the strict requirement for both ligands for binding to this site as well as define the orientation of each ligand in the complex. ED•Fe(II) at 5 μM affords no sequence specific cleavage of the radiolabeled oligonucleotide (Figure 4, lane 8). In the presence of 100 μM 2-ImN, however, a cleavage pattern is observed exclusively at the 5′ end of the 5′-TGTTA-3′ strand of the binding site (Figure 4, lane 10, Figure 6b). Similarly, 2-ImNE•Fe(II) at 100 μM affords no sequence specific

Figure 4.4. MPE•Fe(II) footprinting of D, 2-ImN, and D/2-ImN and affinity cleavage by ED•Fe(II), 2-ImNE•Fe(II), ED•Fe(II)/2-ImN, and 2-ImNE•Fe(II)/D. Gray-scale representation of a storage phosphor autoradiogram of a 20% denaturing polyacrylamide gel. There are 224 levels of gray representing a 50-fold change in the signal from the lowest (13 arbitrary units, white) to highest (600 arbitrary units, black) intensities. The 5′-TGTTA-3′ binding site is shown on the right side of the autoradiogram. All reactions contain 20 mM NaCl, 100 μM bp calf thymus DNA and 20 kcpm 5' labeled 34-base pair oligonucleotide in 40 mM Tris acetate pH 7.6 buffer. Lane 1, intact DNA; lane 2, A reaction; lane 3, G reaction; lane 4, MPE•Fe(II) standard; lane 5, 2 μM D; lane 6, 100 μM 2-ImN; lane 7, 2 μM D and 100 μM 2-ImN; lane 8, 5 μM ED•Fe(II); lane 9, 100 μM 2-ImNE•Fe(II); lane 10, 5 μM ED•Fe(II) and 100 μM 2-ImN; lane 11, 100 μM 2-ImNE•Fe(II) and 2 μM D. Lanes 4-7 contain 33 μM MPE•Fe(II).

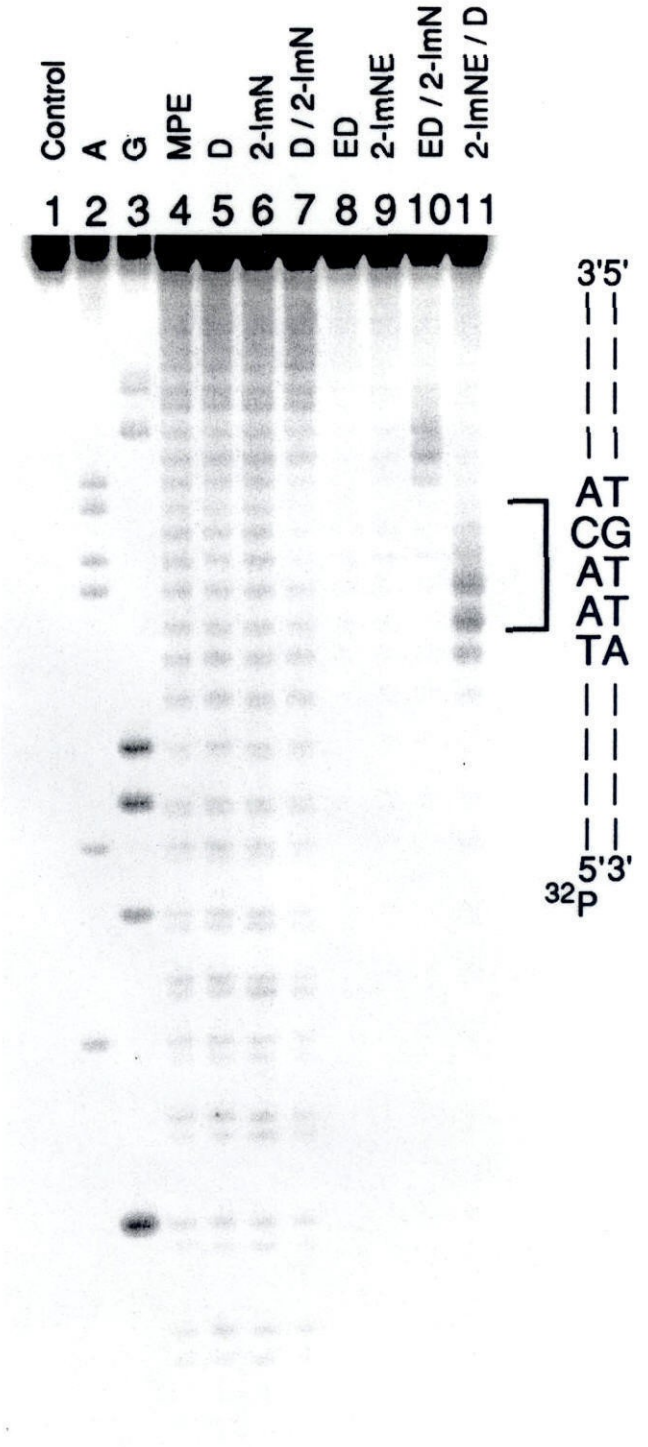
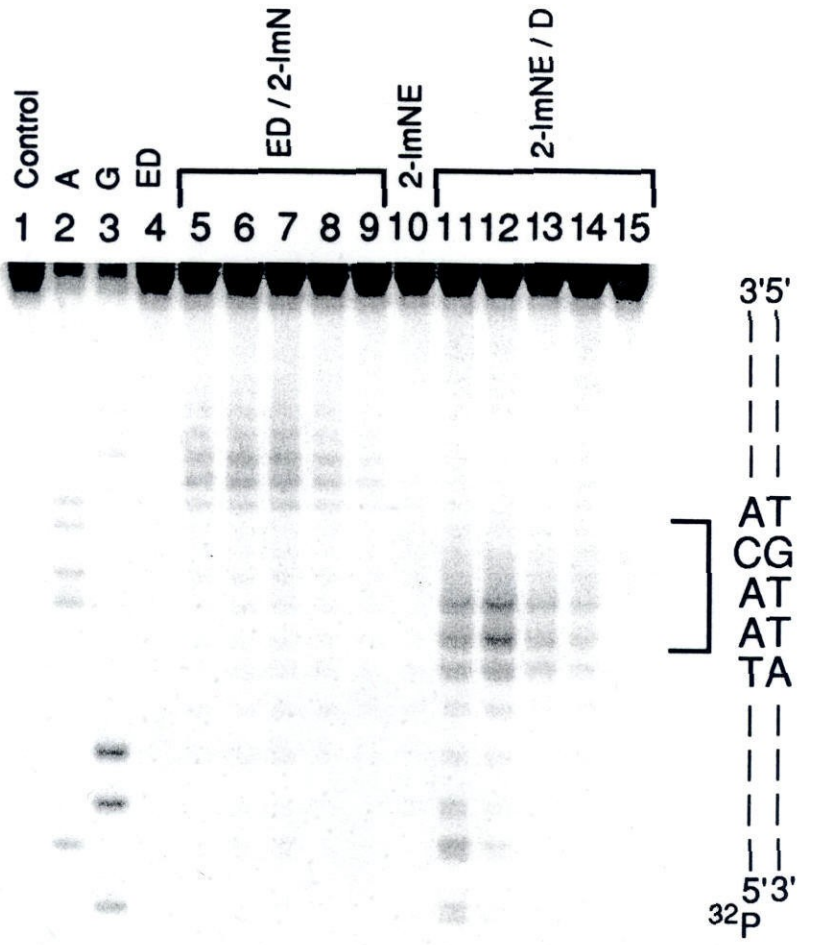

Control
A
G
MPE
D
2-ImN
D / 2-ImN
ED
2-ImNE
ED / 2-ImN
2-ImNE / D
1 2 3 4 5 6 7 8 9 10 11
3'5'
AT
CG
AT
AT
TA
5'3'
32P

Figure 4.5. Concentration dependence of 2-ImN and D on specific cleavage by ED•Fe(II) and 2-ImNE•Fe(II), respectively. Gray-scale representation of a storage phosphor autoradiogram of a 20% denaturing polyacrylamide gel. There are 224 levels of gray representing a 50-fold change in the signal from the lowest (4 arbitrary units, white) to highest (200 arbitrary units, black) intensities. The 5'-TGTTA-3' binding site is shown on the right side of the autoradiogram. All reactions contain 50 mM DTT, 20 mM NaCl, 100 μM bp calf thymus DNA and 20 kcpm 5' labeled 34-base pair oligonucleotide in 40 mM Tris acetate pH 7.6 buffer. Lane 1, intact DNA; lane 2, A reaction; lane 3, G reaction; lanes 4-9 contain 5 μM ED•Fe(II); lane 5, 500 μM 2-ImN; lane 6, 200 μM 2-ImN; lane 7, 100 μM 2-ImN; lane 8, 50 μM 2-ImN; lane 9, 20 μM 2-ImN; lanes 10-15 contain 100 μM 2-ImNE•Fe(II); lane 11, 10 μM D; lane 12, 2.5 μM D; lane 13, 1.0 μM D; lane 14, 0.5 μM D; lane 15, 0.1 μM D.


Control
A
G
ED
ED / 2-ImN
2-ImNE
2-ImNE / D
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
3'5'
AT
CG
AT
AT
TA
5'3'
32P

cleavage but in the presence of 2 μM D reveals cleavage at the 3′ end of the 5′-TGTTA-3′ site (Figure 4, lanes 9 and 11, Figure 6c).

Concentration Dependence. The dependence of concentration of 2-ImN and D on cleavage by ED•Fe(II) and 2-ImNE•Fe(II), respectively, was investigated. ED•Fe(II) at 5 μM shows maximal cleavage in the presence of 2-ImN at 100 μM, which decreases with decreasing concentrations of 2-ImN. Higher concentrations of 2-ImN, however, also lead to a decrease in cleavage by ED•Fe(II) (Figure 5, lanes 4-9). Similarly, 2-ImNE•Fe(II), at 100 μM, shows maximal cleavage in the presence of 2.5 μM D, decreasing with both higher and lower concentrations of D (Figure 5, lanes 10-15).

Discussion

Antiparallel Side-by-Side Heterodimer. The footprinting and affinity cleaving data are consistent with D and 2-ImN binding as an antiparallel side-by-side heterodimer in the minor groove of the 5′-TGTTA-3′ sequence. The footprinting experiments clearly assign the binding site and the affinity cleaving results unambiguously identify the orientation with which each peptide binds this site. Consistent with the proposed model, 2-ImN binds in the orientation which allows for a hydrogen bond between the imidazole N3 of 2-ImN and the guanine 2-amino group (Figure 2). While D can in principle bind with either orientation, the antiparallel arrangement is favored. This antiparallel orientation preference is not unexpected as the parallel arrangement would place the two cationic ammonium groups in proximity. Moreover, the energetic difference for the two alignments may reflect optimization of stacking interactions between the two peptides.

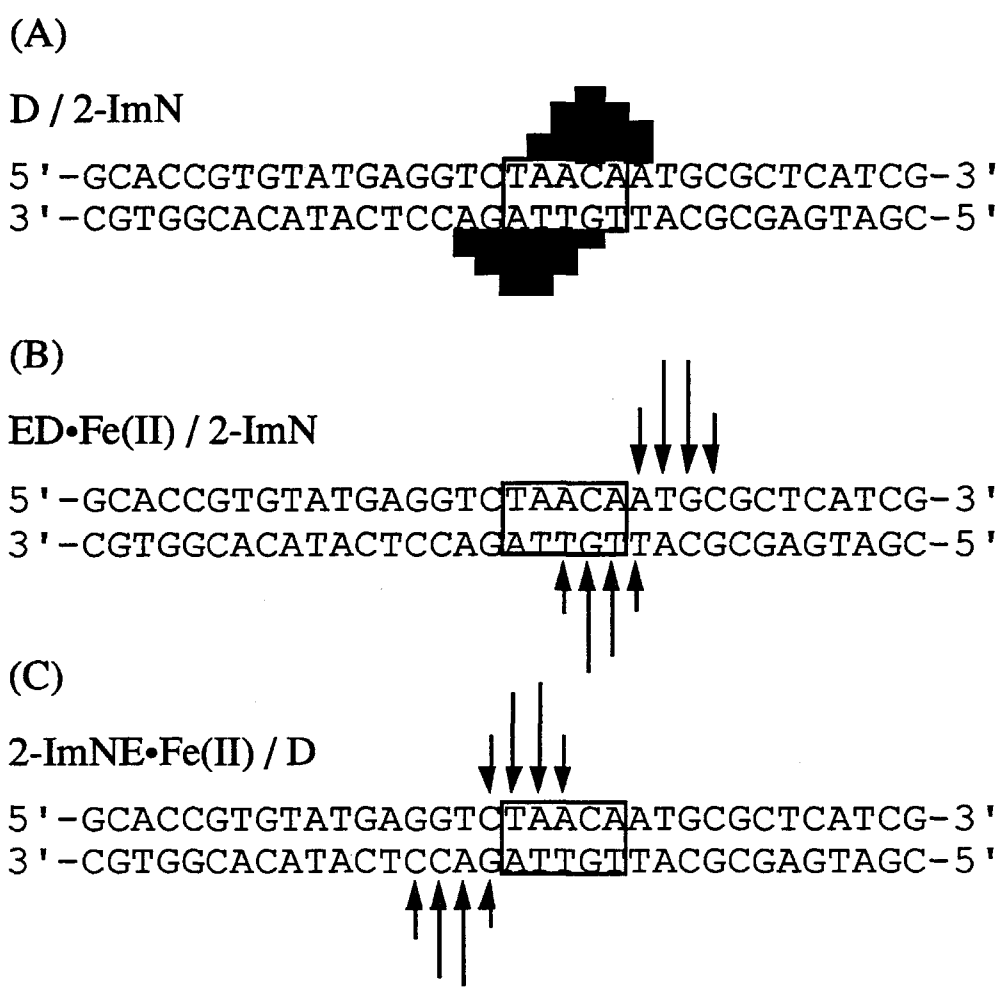


Figure 4.6. Histograms of cleavage protection (footprinting) and affinity cleavage data. (a) MPE•Fe(II) protection pattern for D and 2-ImN on the 34-base pair oligonucleotide (Figure 6, lane 7). Bar heights are proportional to the protection from cleavage at each band. Boxes represent equilibrium binding sites determined by the published model.[2b] (b) Cleavage of the 34-base pair oligonucleotide by ED•Fe(II) in the presence of 2-ImN (Figure 6, lane 10). (c) Cleavage of the 34-base pair oligonucleotide by 2-ImNE•Fe(II) in the presence of D (Figure 6, lane 11). Arrows are proportional to the integrated densities of the cleavage bands. Boxes represent binding sites determined by the published model.[3b] Data for the top strands are shown in Figure 4.

Sequence Specificity. An interesting observation in this heterodimeric complex is that D binds the 5′-TGTTA-3′ site with an apparent binding affinity nearly two orders of magnitude greater than does 2-ImN. This is also the approximate difference in binding affinities observed between D binding to A,T tracts of DNA and 2-ImN binding to 5′-(A,T)G(A,T)C(A,T)-3′ sequences. For the $(2\text{-ImN})_2$•5′-TGACT-3′ complex, we suggested one reason for the overall lower free energy may be that the two guanine amino groups protruding from the floor of the minor groove do not allow the peptide ligands to sit as deeply in the minor groove.[11,13] In the heterodimeric complex, the imidazole N3 of 2-ImN is proposed to form a specific hydrogen bond to the guanine 2-amino group, which may also prohibit close contact between the peptide and the minor groove, consistent with the lower binding affinity. If D does not make specific contacts to the guanine 2-amino group, the peptide may be set deeper in the minor groove than 2-ImN, affording a more stable interaction.

For the 34-base pair duplex studied here, the heterodimer consisting of D and 2-ImN shows a preference for cooperative 2:1 binding to the 5′-TGTTA-3′ site over 1:1 binding to other sites. However, we anticipate that at higher concentrations of peptides, other binding sites available in large DNA can compete for the peptide ligands as either monomeric or homodimeric complexes (Figure 6). Perhaps the specificity of the heterodimer could be further increased by covalently linking the two different peptides. An appropriate linker could disfavor monomeric or homodimeric binding by either D or 2-ImN as well as increase the affinity of the covalent heterodimer by diminishing the entropic penalty for complex formation.

Implications for the Design of Minor Groove Binding Molecules. While the 1:1 models based on the availability of high resolution X-ray crystal structures of netropsin- and distamycin-DNA complexes[4] have proven to date

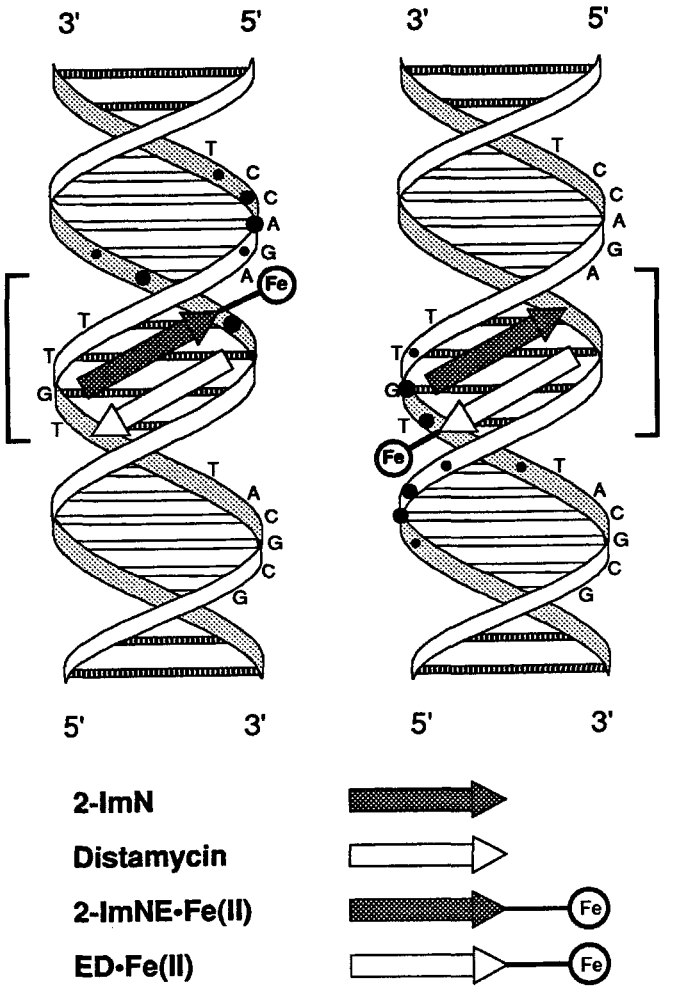


Figure 4.7. Ribbon models for the binding of 2-ImNE•Fe(II) and D, and ED•Fe(II) and 2-ImN as side-by-side antiparallel heterodimers in the minor groove of DNA in which the synthetic peptides 2-ImN and D are represented as gray and white arrows, respectively. Filled circles represent the intensities of cleavage at each base along the phosphodiester backbone from affinity cleavage experiments which define the orientation of each peptide (Figure 6). Brackets on the right and left define the sequence protected from MPE•Fe(II) cleavage by the 2-ImN / D heterodimer.

inadequate for designing peptide analogs capable of binding mixed sequences with high specificities, the 2:1 models accurately predicted binding to a 5'-(A,T)G(A,T)$_3$-3' sequence by the D / 2-ImN heterodimer. This new design rationale differs significantly from earlier strategies in that each ligand is targeted to a single strand of the binding site in the minor groove of DNA, independent of the other ligand-strand interactions. However, it remains to be seen whether the 2:1 motif will serve as a general model for the design of peptide analogs which can bind other specific sequences in the minor groove of DNA.

Structural Characterization of the 2-ImN / D • 5'-TGTTA-3' Complex

In collaboration with the Wemmer group, the structures of the heterodimeric complexes formed upon binding of 2-ImN and D to the two oligonucleotide duplexes d(GCCTAACAAGG)•d(CCTTGTTAGGC) and d(CGCAAACTGGC)•d(GCCAGTTTGCG) were studied by two-dimensional NMR. A summary of the study is presented below: experimental details can be found in a published report.[19]

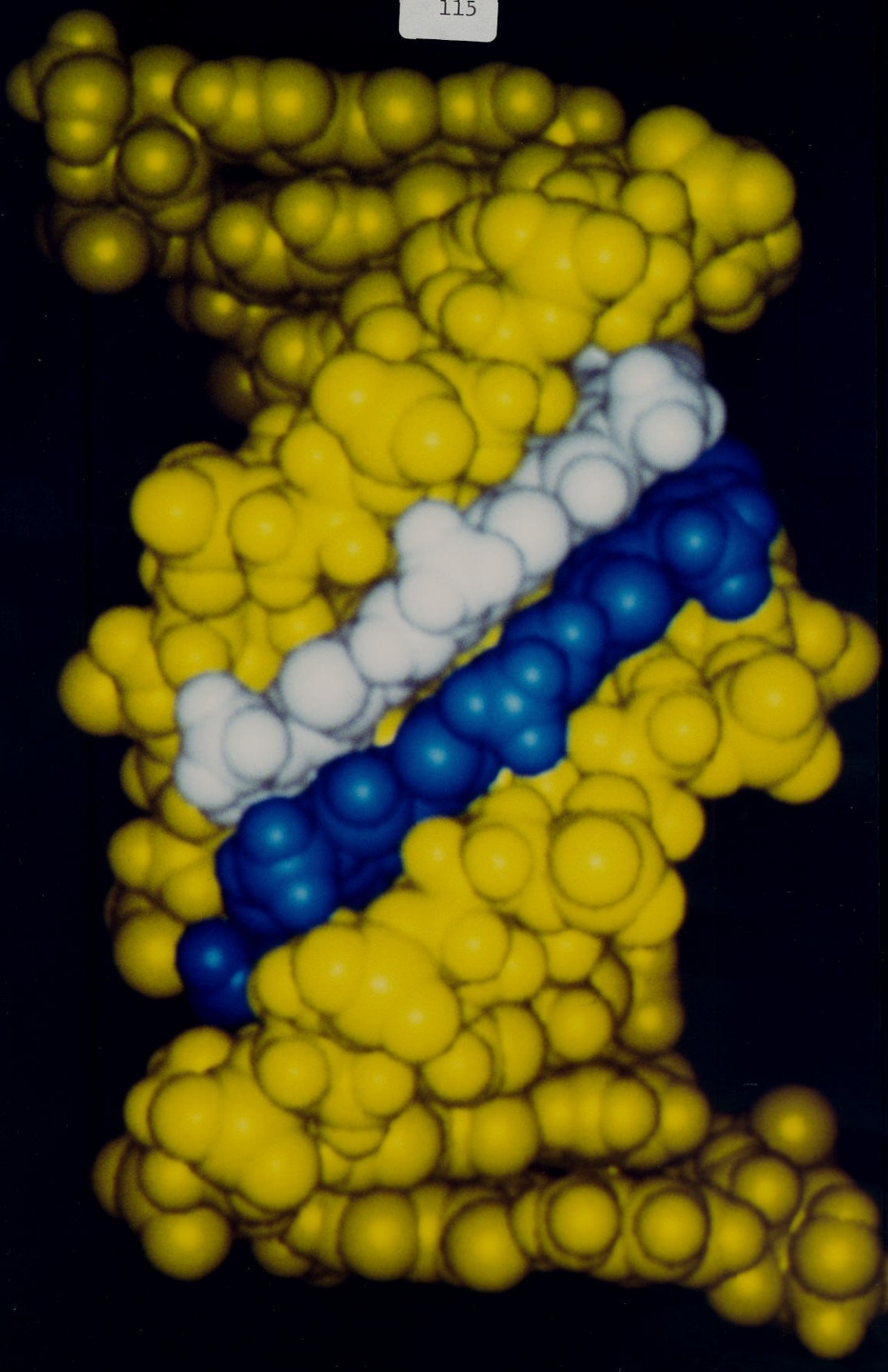
Intermolecular ligand-DNA and ligand-ligand NOE contacts unambiguously confirm the formation of the 1:1:1 2-ImN:D:DNA heterocomplex to a five-base pair binding site in the minor groove of DNA. The D peptide spans AACA and the 2-ImN ligand lies along GTTA of the TAACA:TGTTA binding site, while the respective sites are AACT and GTTT in the AAACT:AGTTT site. The cationic ammonium group of each ligand lies towards the 3'-end of its DNA strand. Evidence for a hydrogen bond between the guanine amino group and N3 of 2-ImN is provided indirectly by NOE cross peaks between the guanine amino protons and neighboring ligand protons in the H_2O NOESY spectrum of the 1:1:1

2-ImN:D:DNA heterocomplex with the TAACA:TGTTA duplex. The appearance of these cross peaks indicates that the rotation of the amino group about the N-C bond is slowed significantly due to interactions with the imidazole nitrogen. Molecular modeling of this heterocomplex suggests that the 2-ImN imidazole nitrogen is positioned almost ideally with respect to geometry and distance for the formation of a hydrogen bond to the guanine amino group (Figure 8).

Titrations of d(GCCTAACAAGG):d(CCTTGTTAGGC) and d(CGCAAACTGGC):d(GCCAGTTTGCG) with D and/or 2-ImN yield only complexes with two ligands bound cooperatively in the antiparallel side-by-side motif. No 1:1 ligand:DNA complexes are observed during any of the titrations. This is in agreement with the previously characterized homodimeric complex of 2-ImN with TGACT:AGTCT[13] and homodimeric and heterodimeric complexes of the pyrrole-imidazole-pyrrole analog 2-ImD and D with various G-C-containing sequences.[14,20] Moreover, the relative positions and geometries of the dimeric peptides bound in the minor groove are very similar in all of these complexes. The peptides stack such that the amides of one ligand overlap the aromatic rings of the other ligand. Furthermore, only a single orientation of peptide has been observed, with the amino terminus of the peptide lying towards the 5'-end of the DNA strand with which it interacts. The 2:1 ligand:DNA binding motif appears to be a general feature of the complexes formed by distamycin analogs with the minor groove of mixed GC/AT DNA sequences, probably arising from the intrinsically wider grooves seen in G-C-containing sequences.[7] Collectively, these studies suggest that the 2:1 motif will be useful and necessary for designing peptides for specific recognition of many other sequences.

The most important feature of the binding site is the guanine in the sequence 5'-(A,T)G(A,T)$_3$-3' while flanking A and T's may be interchangeable while maintaining the hydrogen bonding interactions between the peptides and

Figure **4.8.** Molecular model of the complex formed upon binding of the peptides distamycin and 2-ImN to the oligonucleotide duplex d(GCCTAACAAGG):d(CCTTGTTAGGC).



DNA. Although the 1:1:1 2-ImN:D:DNA heterocomplexes with TAACA:TGTTA and AAACT:AGTTT are very similar in structure, the binding affinities of the peptides for these two sites are likely different. A complete characterization of the sequence specificity of these peptides for 5′-(A,T)G(A,T)$_3$-3′ binding sites requires independent determination of the binding affinities of D and 2-ImN for each of the 16 sequences. Additionally, the effects of flanking base pairs and the dependence of minor groove width on sequence composition are also important determinants of structure and stability of the 1:1:1 2-ImN:D:DNA heterocomplexes.

Design of a Covalent Peptide Heterodimer for Specific Recognition of the 5′-TGTTA-3′ Site

Although it has been shown that the two peptides D and 2-ImN bind the 5′-TGTTA-3′ site as an intermolecular antiparallel heterodimer, each peptide binds as well the sequences 5′-(A,T)$_5$-3′ and 5′-TGTCA-3′, respectively (Figure 9).[21] Recent results of peptide homodimers covalently linked through the central pyrrole rings suggested that a covalent peptide heterodimer should increase affinity for the 5′-TGTTA-3′ site.[22] The covalent peptide heterodimer 2-ImN-C$_4$-P3 wherein the peptides P3 and 2-ImN are connected through the nitrogens of the central pyrrole rings with a butyl linker was synthesized in 12 steps from commercially available precursors (Figure 10). A DNA fragment was constructed which contains a distamycin binding site (5'-TTTTT-3'), a 2-ImN homodimer binding site (5'-TGTCA-3') and a 2-ImN/D heterodimer binding site (5'-TGTTA-3'), each separated by approximately ten base pairs. The plasmid pMM5 was constructed by ligation of an insert, 5'-

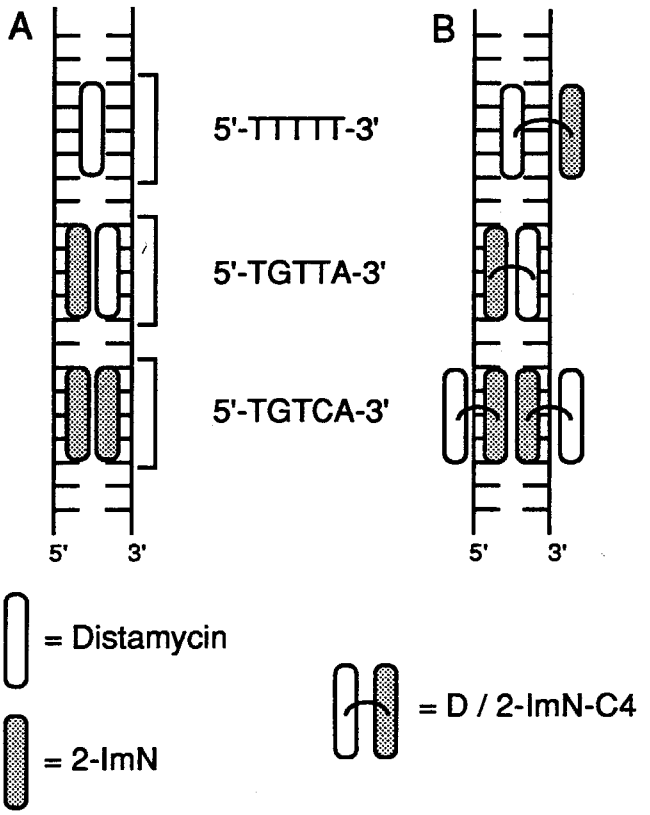
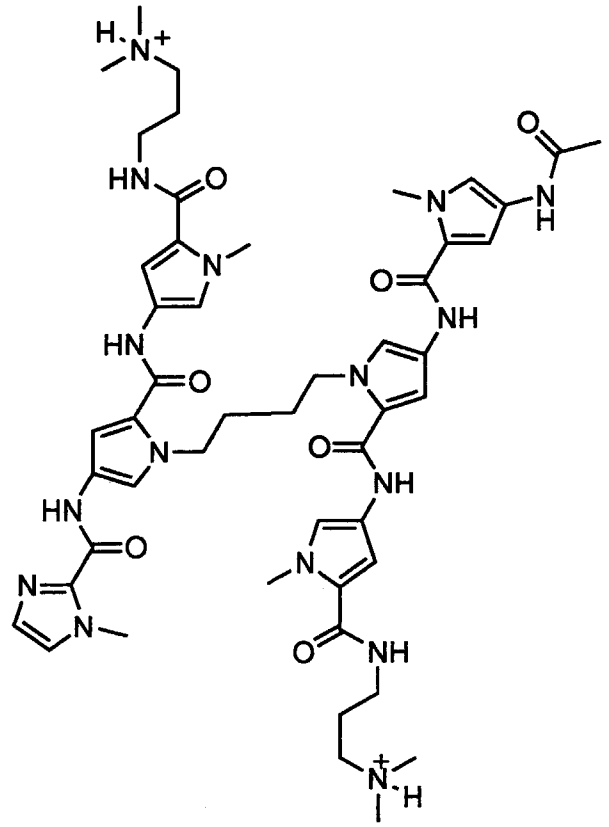


Figure 4.9. (A) The two peptides D and 2-ImN bind the 5′-TGTTA-3′ site as a side-by-side dimer. In addition, D binds A,T-rich sites and 2-ImN binds 5′-TGTCA-3′ sites. (B) A covalent peptide heterodimer is anticipated to bind the 5′-TGTTA-3′ site with improved affinity and sequence-specificity relative to the non-linked peptides.

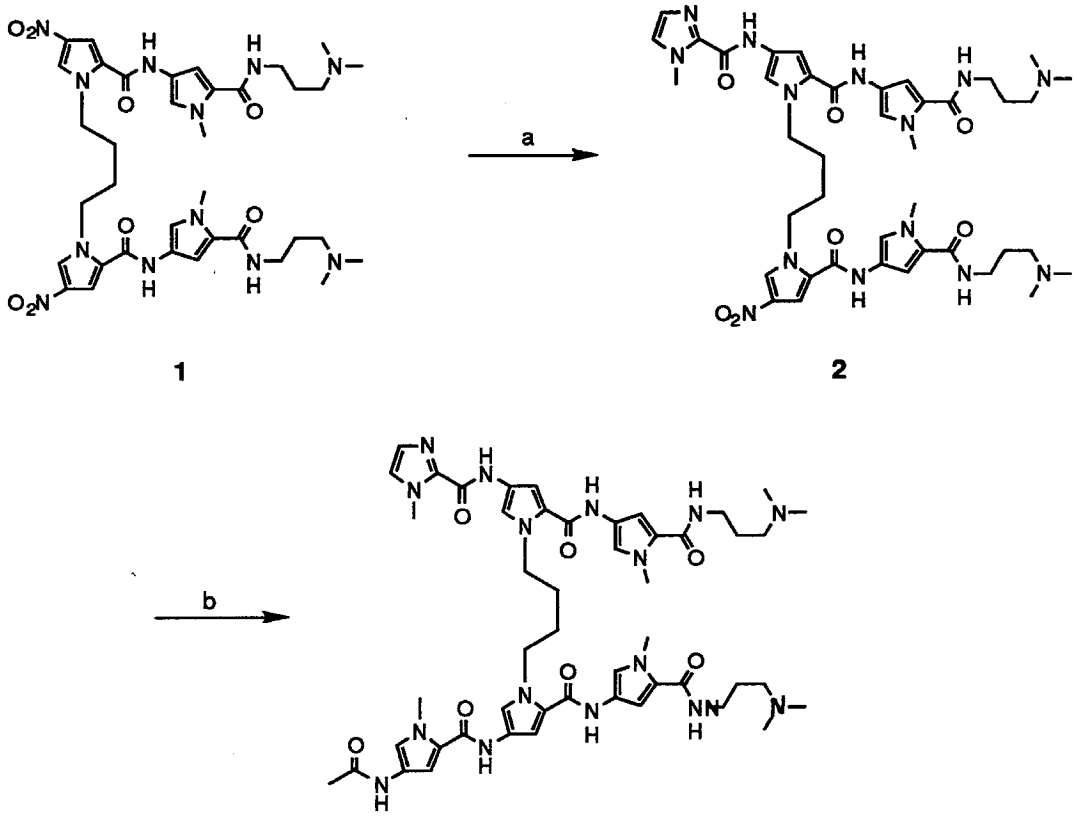
TCGACATGACATTCGTCCACATTGTTAGACCACGATCGTTTTTCGCATG-3' and 5'-CGAAAAACGATCGTGGTCTAACAATGTGGACGAATGTCATG-3' into pUC19 previously cleaved with *Sal* I and *Sph* I. The plasmid was digested with *Eco*R I, labeled at the 3′-end, and digested with *Afl* III. The 447-base pair restriction fragment was isolated by nondenaturing gel electrophoresis and used in all experiments described here. Quantitative DNase I footprint titration experiments on a ^{32}P end-labeled 447-base-pair restriction fragment affords the binding affinities of D, 2-ImN and 2-ImN-C_4-P3 for these three sites (Table I).[11]



2-ImN-C_4-P3

Figure 4.10. Covalent peptide heterodimer 2-ImN-C_4-P3 wherein the two peptides 2-ImN and P3 are connected through the nitrogens of the central pyrrole rings with a butyl linker.

Synthesis. The synthesis of 2-ImN-C_4-P3 begins with **1**, an intermediate in the synthesis of 2-PyN homodimers (Chapter 3) (Figure 11). Partial reduction of **1** (150 psi H_2, Pd/C) affords the monoreduced bis-(dipyrrole) which is coupled to 1-methylimidazole-2-carboxylic acid (DCC, HOBt). This monoamine can be separated from starting material and the bisamino compound by careful chromatography on silica gel. However, the aminopyrrole is sensitive to prolonged exposure to air and must be purified quickly. Reduction of **2** (400 psi



1

2

2-ImN-C_4-P3

Figure 4.11. Synthetic scheme for 2-ImN-C4-P3. (a) (i) 120 psi H_2, Pd/C; (ii) 1-methylimidazole-2-carboxylic acid, DCC, HOBt; (b) (i) 300 psi H_2, Pd/C; (ii) 4-acetamido-1-methylpyrrole-2-carboxylic acid, DCC, HOBt.

H_2, Pd/C) and coupling to 4-acetamido-1-methylpyrrole-2-carboxylic acid (DCC, HOBt) affords 2-ImN-C_4-P3.

P3 binds the three sites in order of decreasing affinity 5'-TTTTT-3' > 5'-TGTTA-3' > 5'-TGTCA-3'. 2-ImN binds these sites in the reverse order, 5'-TGTCA-3' > 5'-TGTTA-3' > 5'-TTTTT-3'. The covalent heterodimer 2-ImN-C_4-P3 specifically binds the 5′-TGTTA-3′ site, the new order now being 5′-TGTTA-3′ > 5'-TGTCA-3' > 5'-TTTTT-3'. The three-fold lower affinity of 2-ImN-C_4-P3 for the 5′-TGTCA-3′ site is consistent with binding as an intramolecular dimer with a single hydrogen bond mismatch between one guanine amino group and the *N*-

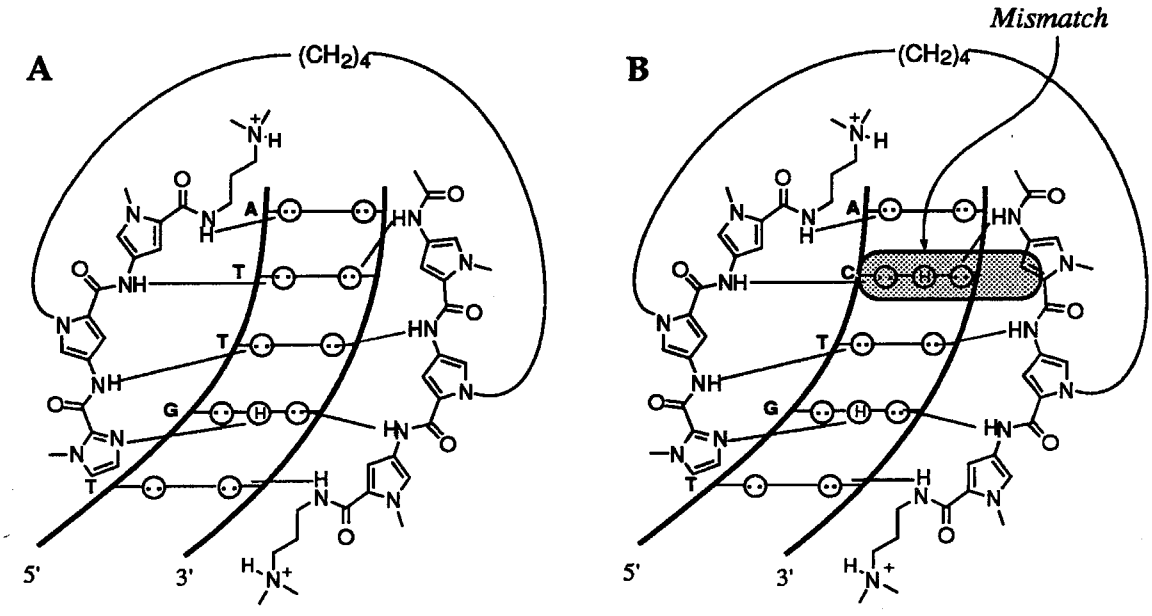
Table 1 Apparent First Order Binding Constants (M^{-1})[a,b]

Peptide	5'-TTTTT-3'	5'-TGTTA-3'	5'-TGTCA-3'
P3	1.6×10^6 (0.2)	1.7×10^5 (0.1)	$< 1 \times 10^5$
2-ImN	$< 5 \times 10^4$	4.7×10^4 (2.6)	1.5×10^5 (0.1)
2-ImN-C_4-D	1.7×10^5 (0.5)	1.1×10^6 (0.3)	3.7×10^5 (0.6)

[a]Values reported are the mean values measured from four footprint titration experiments. Numbers in parentheses indicate the standard deviation for each data set. [b]The assays were performed at 22°C, pH 7.0, in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$ and 5 mM $CaCl_2$.

terminal pyrrolecarboxamide of the distamycin ligand (Figure 12B). Similarly, 2-ImN binds the 5'-TGTTA-3' site with 3-fold lower affinity than the 5'-TGTCA-3' site.[23] This correspondence suggests a *common free energy penalty of ~ 0. 6 kcal/mol for a single hydrogen bond mismatch* in the 2:1 complex. Finally, 2-ImN-C_4-P3 binds the "mismatch" 5'-TGTCA-3' site with greater affinity than does the matched intermolecular homodimer 2-ImN, suggesting that the favorable entropy gained upon linking the peptides overcomes the energetic penalty of the hydrogen bond mismatch.

In the past decade, the design of *nonnatural* minor groove binding molecules with high specificity for new DNA sequences has been driven by some combination of imperfect models, powerful assays and serendipity. The present example combines elements of design which may be general: complimentary hydrogen bond donors and acceptors on side-by-side peptides to those of each DNA strand on the floor of the minor groove, and improving specific affinity by covalently linking the peptides. Perhaps this will provide an underpinning for the design of peptide analogs for sequence-specific recognition in the minor groove of *designated* DNA sites containing both A,T and G,C base pairs.



2-ImN-C_4-P3 • 5'-TGTTA-3' 2-ImN-C_4-P3 • 5'-TGTCA-3'

Figure 4.12. (A) Model for the complex formed between the covalent peptide heterodimer 2-ImN-C_4-P3 with the 5′-TGTTA-3′ site. (B) Model for the complex formed between 2-ImN-C_4-P3 with the 5′-TGTCA-3′ site. The absence of a hydrogen bond acceptor on the peptides for recognition of the guanine amino group results in a single hydrogen bond mismatch in the complex.

Experimental Section

^{1}H NMR and ^{13}C NMR spectra were recorded on a General Electric-QE 300 NMR spectrometer in $CDCl_3$ or DMSO-d_6, with chemical shifts reported in parts per million relative to tetramethyl silane or residual DMSO-d_5, respectively. 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. Reactions were executed under an inert argon atmosphere. Reagent grade chemicals were used as received unless otherwise noted. Tetrahydrofuran (THF) was distilled under nitrogen from sodium/benzophenone ketyl. Dichloromethane (CH_2Cl_2) and triethylamine were distilled under nitrogen from powdered calcium hydride. Dimethylformamide (DMF) was purchased as an anhydrous solvent from Aldrich. Flash chromatography was carried out using EM science Kieselgel 60 (230-400) mesh. Thin-layer chromatography was performed on EM Reagents silica gel plates (0.5 mm thickness). All compounds were visualized with short-wave ultraviolet light.

2-ImN-C_4-nitro-P2 (2). To a solution of bis-dipyrrole **1** (98 mg, 0.126 mmol) in acetic acid (8.0 mL) was added palladium on activated carbon (10 %, 20 mg) and this mixture was allowed to stir under a hydrogen atmosphere (120 psi) in a Parr bomb apparatus for 40 min, at which time TLC analysis indicated approximately 80% conversion of starting material. The reaction mixture was filtered through celite and solvent was removed *in vacuo*. The oil was dissolved in 6% ammonium hydroxide in methanol (10 ml) and solvent was removed under reduced pressure. The desired monoamine was purified by flash column chromatography (1 to 2% ammonium hydroxide in methanol, gradient) and dried *in vacuo* for 2 hr. Separately, to a solution of 1-methylimidazole-2-carboxylic acid (260 mg, 2.34 mmol) and *N*-hydroxybenzotriazole hydrate (320 mg, 2.35 mmol) in dimethylformamide (4.7 mL) was added a solution of 1,3-dicyclohexylcarbodiimide (485 mg, 2.35 mmol) in methylene chloride (4.7 mL). The solution was allowed to stir for 5 h at room temperature. The amine was dissolved in dimethylformamide (5 mL) and added to the activated acid. The resulting solution was allowed to stir for 1 h and methanol (1.0 mL) was added. The solvents were removed *in vacuo* and the residue was purified by flash

column chromatography (0 to 3% ammonium hydroxide in methanol) to afford **2** (20 mg, 19%).

2-ImN-C_4-P3. To a solution of **2** (50 mg, 0.061 mmol) in dimethylformamide (5.0 mL) was added palladium on activated carbon (10 %, 50 mg) and this mixture was allowed to stir under a hydrogen atmosphere (400 psi) in a Parr bomb apparatus for 2 h. To a solution of 4-acetamido-1-methylpyrrole-2-carboxylic acid (250 mg, 1.37 mmol) and *N*-hydroxybenzotriazole hydrate (184 mg, 2.35 mmol) in dimethylformamide (2.7 mL) was added a solution of 1,3-dicyclohexylcarbodiimide (279 mg, 2.35 mmol) in methylene chloride (2.7 mL) followed by stirring for 30 min. The amine was filtered through celite and added to the activated acid and allowed to stir 2.5 h. Methanol (1.0 mL) was added, the solvents were removed *in vacuo* and the residue was purified by flash column chromatography (1 to 2% ammonium hydroxide in methanol) to afford **2-ImN-C_4-P3** (10 mg, 18%).

Materials. Distamycin A was purchased from Sigma. 2-ImN, 2-ImNE, and ED were synthesized and purified as previously described.[3b,11] Concentrations of the peptides were determined by UV spectroscopy using the following extinction coefficients: D (302 nm, ε = 35,000 $cm^{-1}M^{-1}$), ED (303 nm, ε = 35,000 $cm^{-1}M^{-1}$), 2-ImN (302 nm, ε = 26,000 $cm^{-1}M^{-1}$) and 2-ImNE (304 nm, ε = 26,500 $ccm^{-1}M^{-1}$). Automated syntheses of oligonucleotides were performed on an ABI 380B DNA synthesizer using β-cyanoethyl phosphoramidite chemistry. The oligonucleotides were deprotected under standard conditions using ammonium hydroxide and purified by electrophoresis on 15% denaturing polyacrylamide gels. The gel bands were cut out, eluted, filtered through 0.45-μm Centrex filters (Schleicher and Schuell) and precipitated with ethanol.[24] The concentrations of single-strand oligonucleotides were determined at 260 nm, using the following molar extinction coefficients for each base: 15400 (A), 11700

(G), 7300 (C), 8800 (T) $cm^{-1}M^{-1}$. The oligonucleotides were separately labeled at the 5′ end with T4 polynucleotide kinase (Boehringer-Mannheim) and γ-^{32}P ATP (Amersham) and purified by NICK column (Pharmacia).[24] The radiolabeled oligonucleotides were annealed with an excess (2-5 equivalents) of the Watson-Crick complement by heating to 90 °C for 5 min followed by slow cooling to room temperature for 6 hr and appropriate dilution with water. Chemical sequencing reactions were performed according to published methods.[25,26]

Footprinting and Affinity Cleavage Reactions. All reactions were executed in a total volume of 20 μL with final concentrations of each species as indicated. The ligands were added to solutions of radiolabeled oligonucleotide (20,000 cpm), calf thymus DNA (Pharmacia) (100 μM bp), Tris acetate (40 mM, pH 7.6) and NaCl (20 mM) and incubated for 15 min at 22 °C. Footprinting reactions were initiated by the addition of MPE•Fe(II) (33 μM) and dithiothreitol (20 mM) and allowed to proceed 15 min at 22 °C. Affinity cleavage reactions were initiated by the addition of dithiothreitol (50 mM) and allowed to proceed 30 min at 22 °C. All reactions were stopped by precipitation with ethanol, resuspended in 100 mM tris-borate-EDTA/80% formamide loading buffer and electrophoresed on 20% polyacrylamide denaturing gels (5% crosslink, 7 M urea) at 1500 V for 6-8 hr. The gels were analyzed using storage phosphor technology.

Quantitation by Storage Phosphor Technology Autoradiography. Photostimulable storage phosphor imaging plates (Kodak Storage Phosphor Screen S0230 obained from Molecular Dynamics) were pressed flat against gel samples and exposed in the dark at 22 °C for 15-20 h. 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 the ImageQuant v. 3.0 software running on an AST Premium 386/33 computer.

References

1. (a) Krylov, A. S.; Grokhovsky, S. L.; Zasedatelev, A. S.; Zhuze, A. L.; Gursky, G. V.; Gottikh, B. P. *Nuc. Acids. Res.* **1979**, *6*, 289-304. (b) Zasedatelev, A. S.; Gursky, G. V.; Zimmer, Ch.; Thrum, H. *Mol. Biol. Reports* **1974**, *1*, 337-342. (c) Zasedatelev, A. S.; Zhuze, A. L.; Zimmer, Ch.; Grokhovsky, S. L.; Tumanyan, V. G.; Gursky, G. V.; Gottikh, B. P. *Dokl. Acad. Nauk SSSR* **1976**, *231*, 1006-1009. For a review, see: (d) Zimmer, C.; Wåhnert, U. *Prog. Biophys. Molec. Biol.* **1986**, *47*, 31-112.

2. (a) Van Dyke, M. W.; Hertzberg, R. P.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 5470-5474. (b) Van Dyke, M. W.; Dervan, P. B. *Cold Spring Harbor Symposium on Quantitative Biology* **1982**, *47*, 347-353. (c) Van Dyke, M. W.; Dervan, P. B. *Biochemistry* **1983**, *22*, 2373-2377. (d) Harshman, K. D.; Dervan, P. B. *Nucl. Acids Res.* **1985**, *13*, 4825-4835.(e) Fox, K. R.; Waring, M. J. *Nucl. Acids Res.* **1984**, *12*, 9271-9285. (f) Lane, M. J.; Dobrowiak, J. C.; Vournakis, J. *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 3260-3264.

3. (a) Schultz, P. G.; Taylor, J. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1982**, *104*, 6861-6863. (b) Taylor, J. S.; Schultz, P. G.; Dervan, P. B. *Tetrahedron* **1984**, *40*, 457-465. (c) Schultz, P. G.; Dervan, P. B. *J. Biomol. Struct. Dyn.* **1984**, *1*, 1133-1147. (d) Dervan, P. B. *Science*, **1986**, *232*, 464-471.

4. (a) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 1376-1380. (b) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *J. Mol. Biol.* **1985**, *183*, 553-563. (c) Coll, M.; Frederick, C. A.; Wang, A. H.-J.; Rich, A. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 8385-8389.

5. (a) Patel, D. J.; Shapiro, L. *J. Biol. Chem.* **1986**, *261*, 1230-1240. (b) Klevitt, R. E.; Wemmer, D. E.; Reid, B. R. *Biochemistry* **1986**, *25*, 3296-3303. (c) Pelton, J. G., Wemmer, D. E. *Biochemistry* **1988**, *27*, 8088-8096.

6. (a) Markey, L. A.; Breslauer, K. J. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 4359-4363. (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*, 8922-8926.

7. Yoon, C.; Privé, G. G.; Goodsell, D. S.; Dickerson, R. E. *Proc. Natl. Acad. Sic. USA* **1988**, *85*, 6332-6336.

8. For early examples of hybrid molecules for the recognition of mixed sequences, see: (a) Dervan, P. B.; Sluka, J. P. *New Synthetic Methodology and Functionally Interesting Compounds*; Elsevier: New York, 1986; pp 307-322. (b) Griffin, J. H.; Dervan, P. B. *J. Am. Chem. Soc.* **1987**, *109*, 6840-6842.

9. (a) Lown, J. W.; Krowicki, K.; Bhat, U. G.; Skorobogaty, A.; Ward, B.; Dabrowiak, J. C. *Biochemistry* **1986**, *25*, 7408-7416. (b) Kissinger, K.; Krowicki, K.; Dabrowiak, J. C.; Lown, J. W. *Biochemistry* **1987**, *26*, 5590-5595. (c) Lee, M.; Chang, D. K.; Hartley, J. A.; Pon, R. T.; Krowicki, K.; Lown, J. W. *Biochemistry* **1988**, *27*, 445-455.

10. (a) Wade, W. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1987**, *109*, 1574-1575. (b) Wade, W. S. Ph. D. Thesis, California Institute of Technology, 1989.

11. Wade, W.S., Mrksich, M., Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 8783-8794.

12. (a) Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 5723-5727. (b) Pelton, J. G.; Wemmer, D. E. *J. Am. Chem. Soc.* **1990**, *112*, 1393-1399.

13. Mrksich, M., Wade, W. S., Dwyer, T. J., Geierstanger, B. H., Wemmer, D. E., Dervan, P. B. *Proc. Natl. Acad. Sci., USA* **1992**, *89*, 7586-7590.

14. Dwyer, T. J.; Geierstanger, B. H.; Bathini, Y.; Lown, J. W.; Wemmer, D. E. *J. Am. Chem. Soc.* **1992,** *114,* 5911-5919.

15. The many recent reports of synthetic peptide analogs of distamycin and netropsin for the sequence specific recognition of DNA will necessitate a consistent nomenclature for these molecules. We suggest that the peptides be named according to the sequence of heterocyclic residues, starting from the amino-terminus. Netropsin and distamycin, consisting of two and three N-methylpyrrolecarboxamides, for example, could be designated P-P and P-P-P, respectively. Likewise, 2-ImN is designated Im-P-P, where Im = 1-methylimidazole-2-carboxamide. 2-ImD, the distamycin analog designed by Wemmer, Lown and co-workers would be designated P-Im-P.[14]

16. In the case of a heterodimeric complex, the term antiparallel refers to the amino to carboxy direction of the peptides in opposite alignment.

17. (a) Sutcliffe, J. G. *Cold Spring Harbor Symposium on Quantitative Biology* **1978,** *43,* 77-90. (b) Peden, K. W. C. *Gene,* **1983,** *22,* 277-80.

18. The 34-base pair oligonucleotides corresponding to base pairs 77-110 of pBR322 have bases 90-91 changed from AA to GG to eliminate the 5′-AAATC-3′ distamycin site.

19. Geierstanger, B. H.; Jacobsen, P.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Biochemistry,* **1994,** in press.

20. Dwyer, T. J.; Geierstanger, B. H.; Bathini, Y.; Lown, J. W.; Wemmer, D. E. *J. Am. Chem. Soc.* **1992,** *114,* 5911-5919.

21. Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993,** *115,* 2572-2576.

22. Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1994,** in press.

23. For characterization of the $(2\text{-ImN})_2 \bullet 5'\text{-TGTTA-}3'$ complex by two-dimensional NMR, see reference 19.

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

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

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

Chapter 5

Design of Head-To-Tail Linked Peptide Heterodimers for Sequence-Specific Recognition in the Minor Groove of Double-Helical DNA

2:1 Peptide-DNA Complexes. Although 1:1 peptide-DNA models based on crystal structures of netropsin (N) and distamycin (D)[1] complexes aided in the design of oligopeptides for recognition of longer tracts of DNA,[2] the models failed in designing peptides for recognition of mixed sequences.[3] The synthetic peptides pyridine-2-carboxamide-netropsin (2-PyN) and 1-methylimidazole-2-carboxamide-netropsin (2-ImN), targeted to the 5'-TGTTA-3' site as 1:1 complexes, were found to bind the 5'-TGACA-3' site as antiparallel side-by-side dimers in the minor groove.[4] These and other recent 2:1 peptide-DNA complexes[5] have guided the successful design of peptide analogs for recognition of other sequences. Footprinting, affinity cleaving and NMR experiments demonstrate that two peptides D and 2-ImN bind in the minor groove of 5'-TGTTA-3' as a side-by-side heterodimer (Figure 1).[6] In a different example Wemmer, Lown and co-workers reported binding to a 5'-AAGTT-3' sequence by a heterodimer consisting of D and a distamycin analog containing a central imidazole.[7] Our recent finding that the designed peptide ImPImP binds in the minor groove of 5'-(A,T)GCGC(A,T)-3' sequences as a side-by-side dimer underscores the utility of 2:1 models in designing peptides for sequence-specific recognition of DNA.[8] These examples suggest that the 2:1 models may facilitate the design of peptides for recognition of many other sequences.

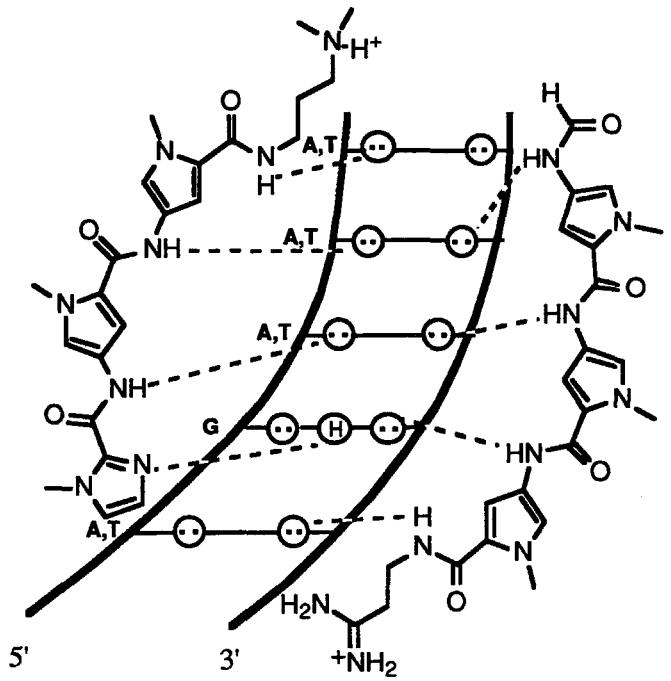
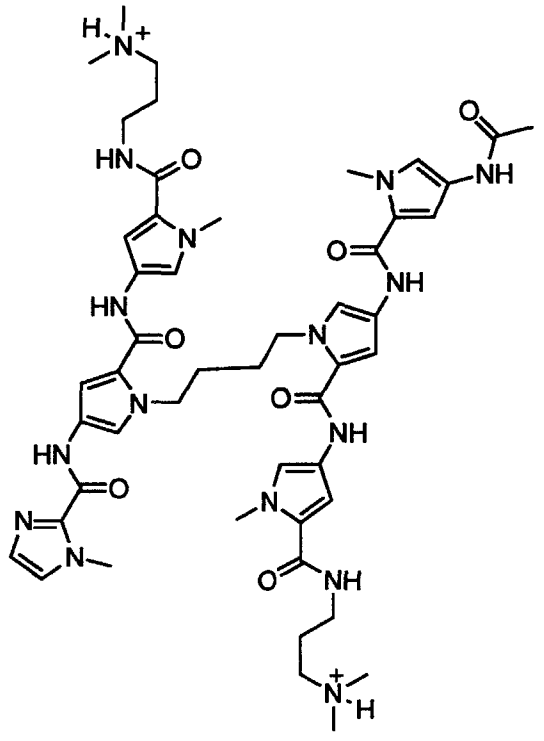


Figure 1. Heterodimeric binding model for the complex formed between 2-ImN and D with a 5′-TGTTA-3′ sequence. Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dotted lines.

Covalent Peptide Dimers. In an effort to improve the binding affinity and sequence-specificity of 2-PyN for 5'-TGACA-3' sites, a series of covalent peptide dimers were synthesized wherein the 2-PyN ligands were connected through the nitrogens of the central pyrrole rings with alkyl tethers.[9] These peptide dimers bind the 5'-TGACA-3' sequence with ten-fold greater affinity as compared to the non-linked peptides. This strategy was also applied to the D/2-ImN heterodimer which not only binds the 5′-TGTTA-3′ site, but also the sequences 5′-$(A,T)_5$-3′ and 5′-TGTCA-3′. We found that the covalent peptide heterodimer 2-ImN-C_4-P3 wherein the two peptides are tethered through the central pyrrole rings with a butyl linker binds specifically the 5'-TGTTA-3' site



2-ImN-C_4-P3

Figure 2. Covalent peptide heterodimer 2-ImN-C_4-P3 wherein the two peptides 2-ImN and P3 are connected through the nitrogens of the central pyrrole rings with a butyl linker.

(Figure 2).[10] These examples demonstrate that linking side-by-side peptides may be an important component in the rational design of peptides for recognition of other sequences. However, the synthetic methodology for preparation of this class of covalent peptide dimers is not easily extended to the synthesis of other peptide analogs. We report here a second generation peptide dimer which offers methodology for the preparation of a wide variety of covalent peptide dimers.

Experimental Design. The design of the first generation covalent peptide dimers was based on the proximity of the two methyl groups of the central pyrroles in the 2:1 complex. Alternatively, the two peptides could also be tethered at their ends without perturbing the peptide-DNA complex geometry. Within the 2-ImN/D•5'-TGTTA-3' complex, the terminal carboxyl

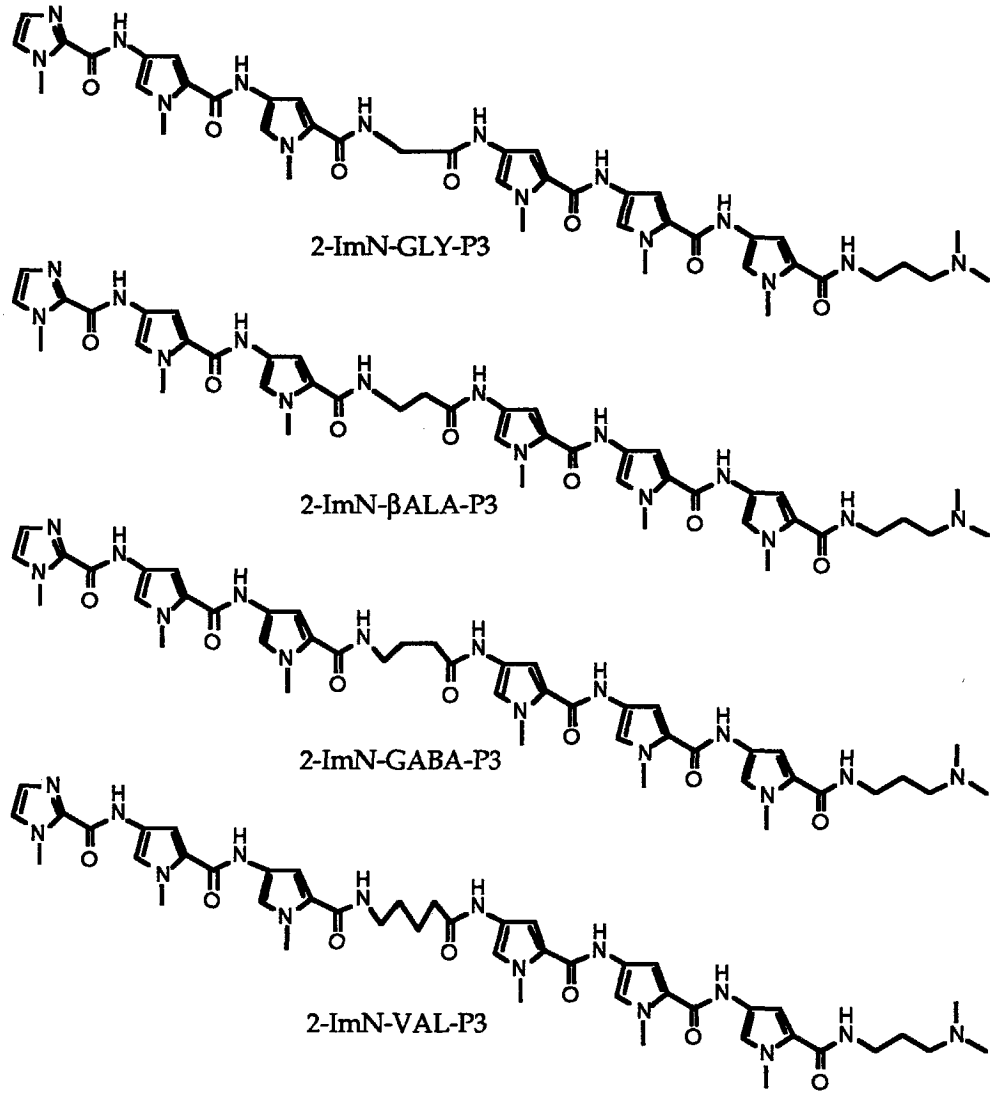


Figure 3. Covalent peptide dimers 2-ImN-GLY-P3, 2-ImN-βALA-P3, 2-ImN-GABA-P3 and 2-ImN-VAL-P3 wherein the two peptides D and 2-ImN are connected with glycine, b-alanine, γ-aminobutyric acid and 5-aminovaleric acid, respectively.

group of 2-ImN and the terminal amine of P3 are appropriately positioned for linking with amino acids. An appropriate linker should allow the linear peptide dimer to turn and fold back on itself and bind the 5'-TGTTA-3' site with enhanced affinity. A series of linear peptide dimers were synthesized wherein the C-terminal carboxylic acid of 2-ImN and the *N*-terminal amine of P3 were

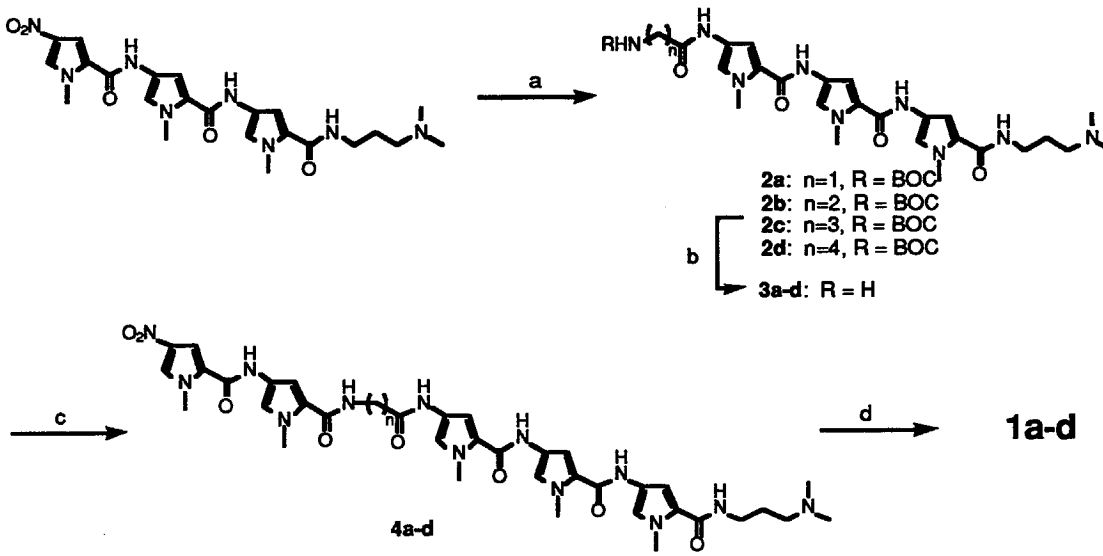


Figure 4. Synthetic scheme for 2-ImN-GLY-P3, 2-ImN-βALA-P3, 2-ImN-GABA-P3 and 2-ImN-VAL-P3. (a) (i) 300 psi H_2, 10% Pd/C; (ii) $BOCNH(CH_2)_nCO_2H$, DCC, HOBt; (b) TFA / CH_2Cl_2; (c) (i) 300 psi H_2, 10% Pd/C; (ii) *N*-methyl-4-(*N*-methyl-4-nitropyrrole-2-carboxamide)-pyrrole-2-carboxylic acid, DCC, HOBt; (d) (i) 300 psi H_2, 10% Pd/C; (ii) 1-methylimidazole-2-carboxylic acid, DCC, HOBt.

connected with glycine, β-alanine, γ-aminobutyric acid and 5-aminovaleric acid linkers (Figure 3). The binding affinities of 2-ImN-GLY-P3, 2-ImN-βALA-P3, 2-ImN-GABA-P3 and 2-ImN-VAL-P3 for the 5'-TGTTA-3' and 5'-TGACA-3' sites were determined by quantitative DNase I footprint titration experiments.

Results

Synthesis of Linked Dimers. The peptides are synthesized in four steps from previously described intermediates (Figure 4). Reduction of the nitro-tris-(*N*-methylpyrrolecarboxamide)[11] (300 psi H_2, Pd/C) and coupling with the *N*-(*tert*-butoxycarbonyl) amino acid (DCC, HOBt) afforded peptides **2a-d**. Deprotection of the amine (TFA, CH_2Cl_2) and coupling with *N*-methyl-4-(*N*-

methyl-4-nitropyrrole-2-carboxamide)-pyrrole-2-carboxylic acid[12] (DCC, HOBt) yielded **4a-d** in 85% yield. Reduction of the nitropyrrole (300 psi H_2, Pd/C) and coupling to 1-methylimidazole-2-carboxylic acid (DCC, HOBt) provided the covalent peptide dimers 2-ImN-GLY-P3, 2-ImN-βALA-P3, 2-ImN-GABA-P3 and 2-ImN-VAL-P3. The solubility of the four peptides has a strong even-odd effect on the linker length. 2-ImN-GLY-P3 and 2-ImN-GABA-P3 were freely soluble in aqueous solution. 2-ImN-βALA-P3 and 2-ImN-VAL-P3 were soluble only at concentrations less than 1.0 mM.

Footprinting. DNase I footprinting[13] on the 3′-^{32}P end-labeled 447 base pair EcoR 1/Afl III restriction fragment from plasmid pMM5 (10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$ at pH 7.0 and 22 °C) reveal that the four peptides specifically bind the sites 5'-TTTTT-3', 5'-TGTTA-3' and 5'-TGACA-3'. The apparent first order binding affinities of the peptides for the three sites were determined by quantitative DNase I footprint titration experiments (Figures 5-8 and Table I).[14] The four hexapeptides bind the 5′-TGTTA-3′ site with greater affinity than the 5′-TGACA-3′ site. In particular, 2-ImN-GABA-P3 binds the 5′-TGTTA-3′ site with the highest affinity, 1.7×10^6 M^{-1}. The θ_{app} points for 2-ImN-GLY-P3 and 2-ImN-βALA-P3 binding the 5'-TGTTA-3' and 5'-TGACA-3' sites are steeper than expected for a 1:1 complex of peptide with DNA. Rather, the data points are adequately fit by a cooperative binding isotherm consistent with intermolecular dimeric binding (Figure 9, see experimental section). The ([L], θ_{app}) data points for 2-ImN-GABA-P3 and 2-ImN-VAL-P3 binding these sites were best fit with a Langmuir binding isotherm consistent with a 1:1 complex of peptide with DNA (Figure 10). For all peptides binding the 5'-TTTTT-3' site, the binding affinities were in the range of 2×10^5 to 1×10^6 M^{-1} and are approximate since the quality of fits and the standard deviations for these data sets were poor.

Figure 5. Quantitative DNase I footprint titration experiment with 2-ImN-GLY-P3 on the 3'-^{32}P-labeled 447 bp *Eco*R I/*Afl* III restriction fragment from plasmid pMM5. The 5'-TTTTT-3', 5'-TGTTA-3' and 5'-TGACA-3' binding sites are shown on the right side of the autoradiogram. All reactions contain 20 kcpm restriction fragment, 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$. Lane 1, intact DNA; lane 2, A reaction; lane 3, G reaction; lanes 4 and 18, DNase I standard; lane 5-17 contain 100 μM, 50 μM, 20 μM, 10 μM, 5 μM, 2 μM, 1 μM, 500 nM, 200 nM, 100 nM, 50 nM, 20 nM and 10 nM 2-ImN-GABA-P3, respectively.

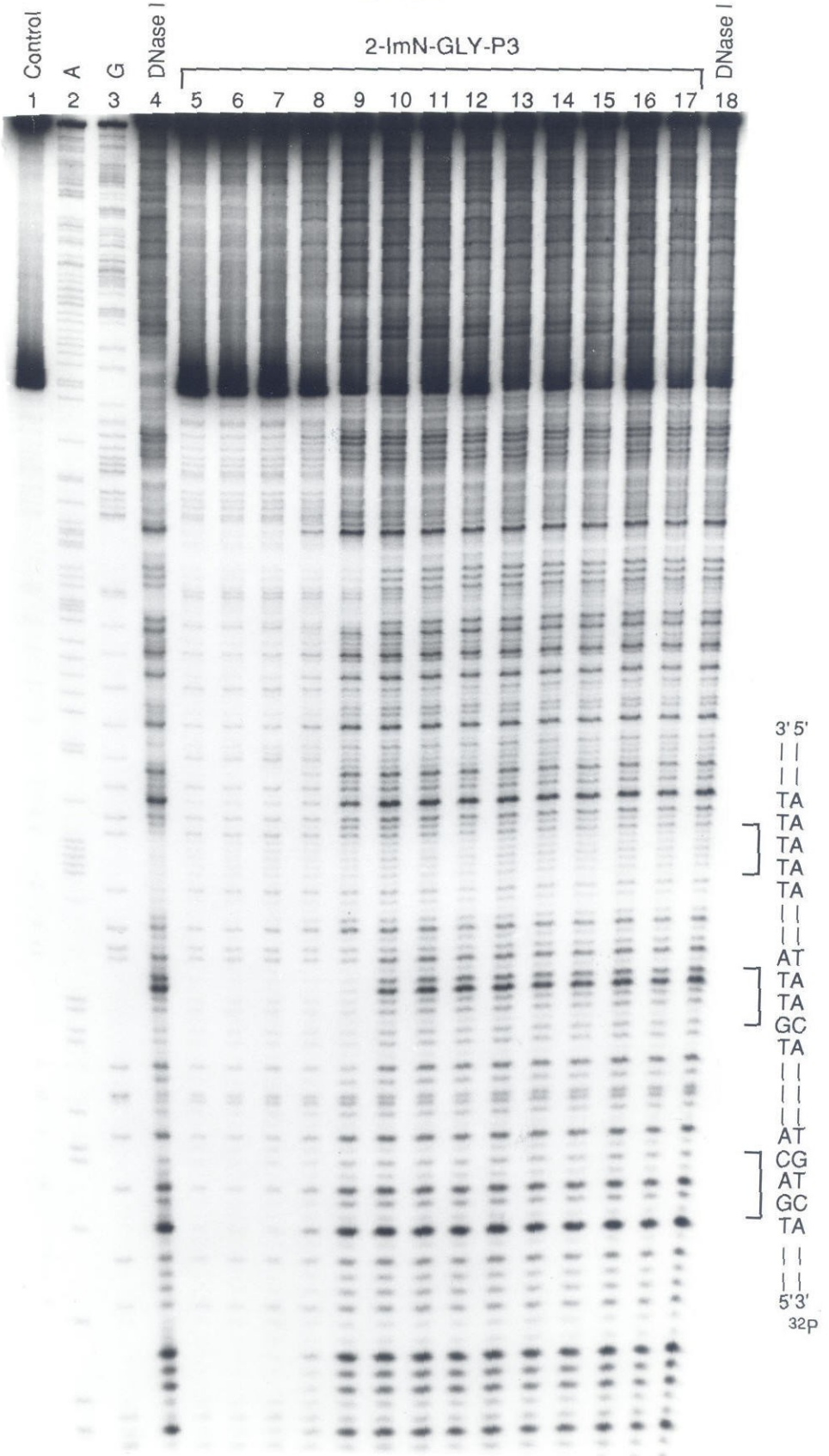

Control
A
G
DNase I
2-ImN-GLY-P3
DNase I
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
3' 5'
TA
TA
TA
TA
TA
AT
TA
TA
GC
TA
AT
CG
AT
GC
TA
5' 3'
32P

Figure 6. Quantitative DNase I footprint titration experiment with 2-ImN-βALA-P3 on the 3'-^{32}P-labeled 447 bp *Eco*R I/*Afl* III restriction fragment from plasmid pMM5. The 5'-TTTTT-3', 5'-TGTTA-3' and 5'-TGACA-3' binding sites are shown on the right side of the autoradiogram. All reactions contain 20 kcpm restriction fragment, 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$. Lane 1, intact DNA; lane 2, A reaction; lane 3, G reaction; lanes 4 and 17, DNase I standard; lane 5-16 contain 50 μM, 20 μM, 10 μM, 5 μM, 2 μM, 1 μM, 500 nM, 200 nM, 100 nM, 50 nM, 20 nM and 10 nM 2-ImN-βALA-P3, respectively.

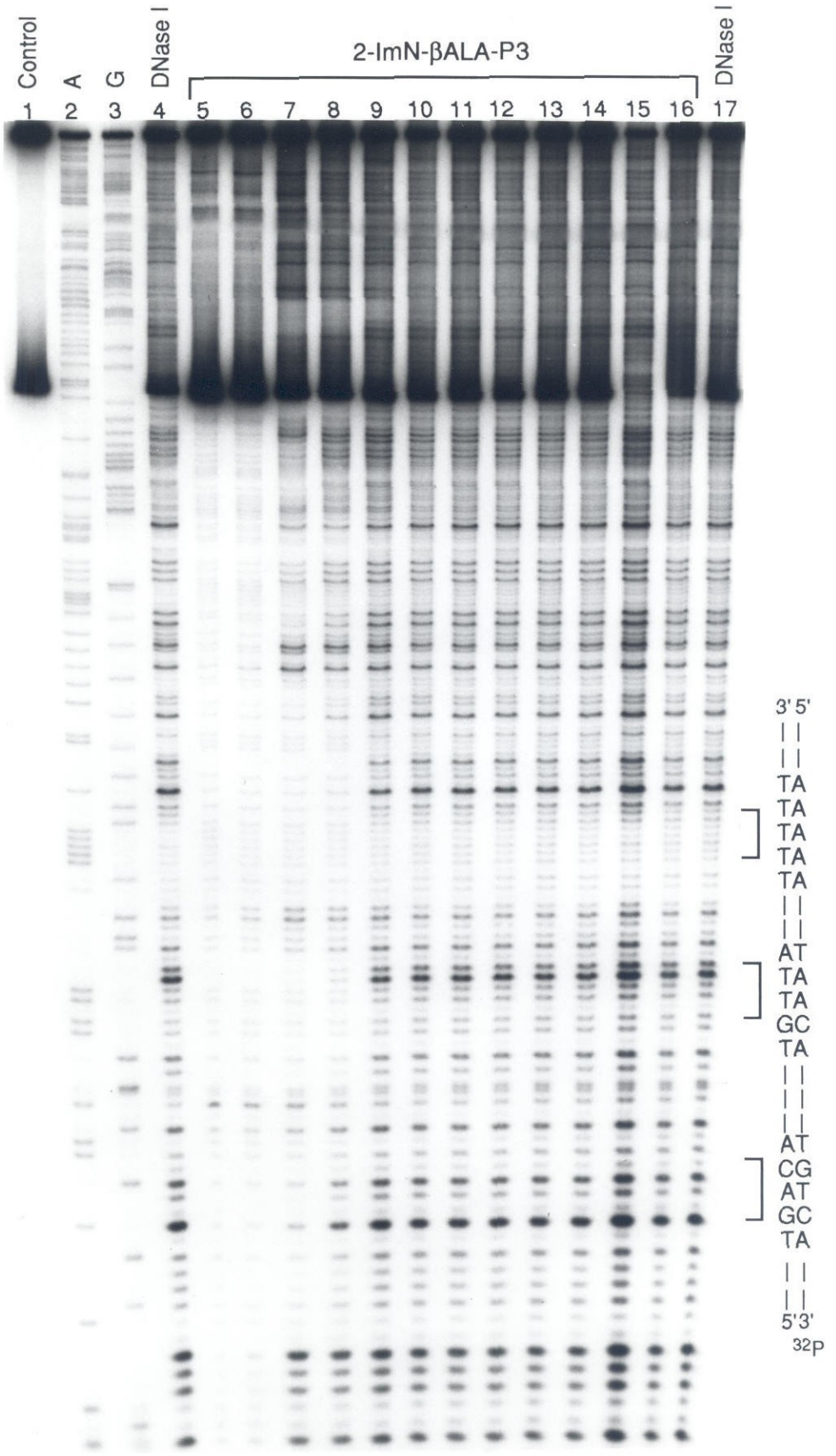

Control
A
G
DNase I
2-ImN-βALA-P3
DNase I
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
3' 5'
TA
TA
TA
TA
TA
AT
TA
TA
GC
TA
AT
CG
AT
GC
TA
5'3'
32P

Figure 7. Quantitative DNase I footprint titration experiment with 2-ImN-GABA-P3 on the 3'-^{32}P-labeled 447 bp *Eco*R I/*Afl* III restriction fragment from plasmid pMM5. The 5'-TTTTT-3', 5'-TGTTA-3' and 5'-TGACA-3' binding sites are shown on the right side of the autoradiogram. All reactions contain 20 kcpm restriction fragment, 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$. Lane 1, intact DNA; lane 2, A reaction; lane 3, G reaction; lanes 4 and 18, DNase I standard; lane 5-17 contain 100 μM, 50 μM, 20 μM, 10 μM, 5 μM, 2 μM, 1 μM, 500 nM, 200 nM, 100 nM, 50 nM, 20 nM and 10 nM 2-ImN-GABA-P3, respectively.

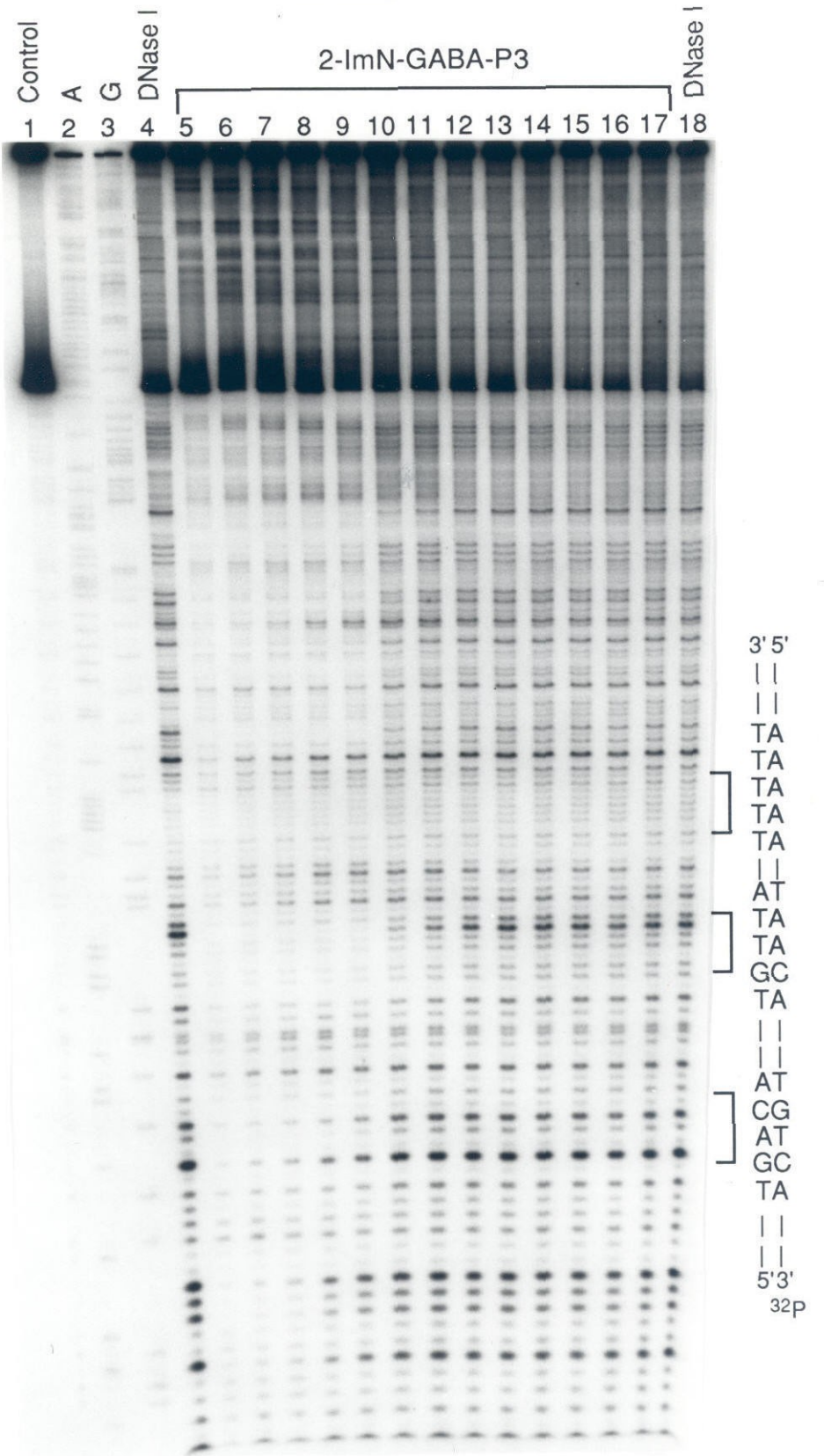

Control
A
G
DNase I
2-ImN-GABA-P3
DNase I
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
3' 5'
TA
TA
TA
TA
TA
AT
TA
TA
GC
TA
AT
CG
AT
GC
TA
5'3'
32P

Figure 8. Quantitative DNase I footprint titration experiment with 2-ImN-VAL-P3 on the 3'-^{32}P-labeled 447 bp *Eco*R I/*Afl* III restriction fragment from plasmid pMM5. The 5'-TTTTT-3', 5'-TGTTA-3' and 5'-TGACA-3' binding sites are shown on the right side of the autoradiogram. All reactions contain 20 kcpm restriction fragment, 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$. Lane 1, intact DNA; lane 2, A reaction; lane 3, G reaction; lanes 4 and 17, DNase I standard; lane 5-16 contain 50 μM, 20 μM, 10 μM, 5 μM, 2 μM, 1 μM, 500 nM, 200 nM, 100 nM, 50 nM, 20 nM and 10 nM 2-ImN-VAL-P3, respectively.

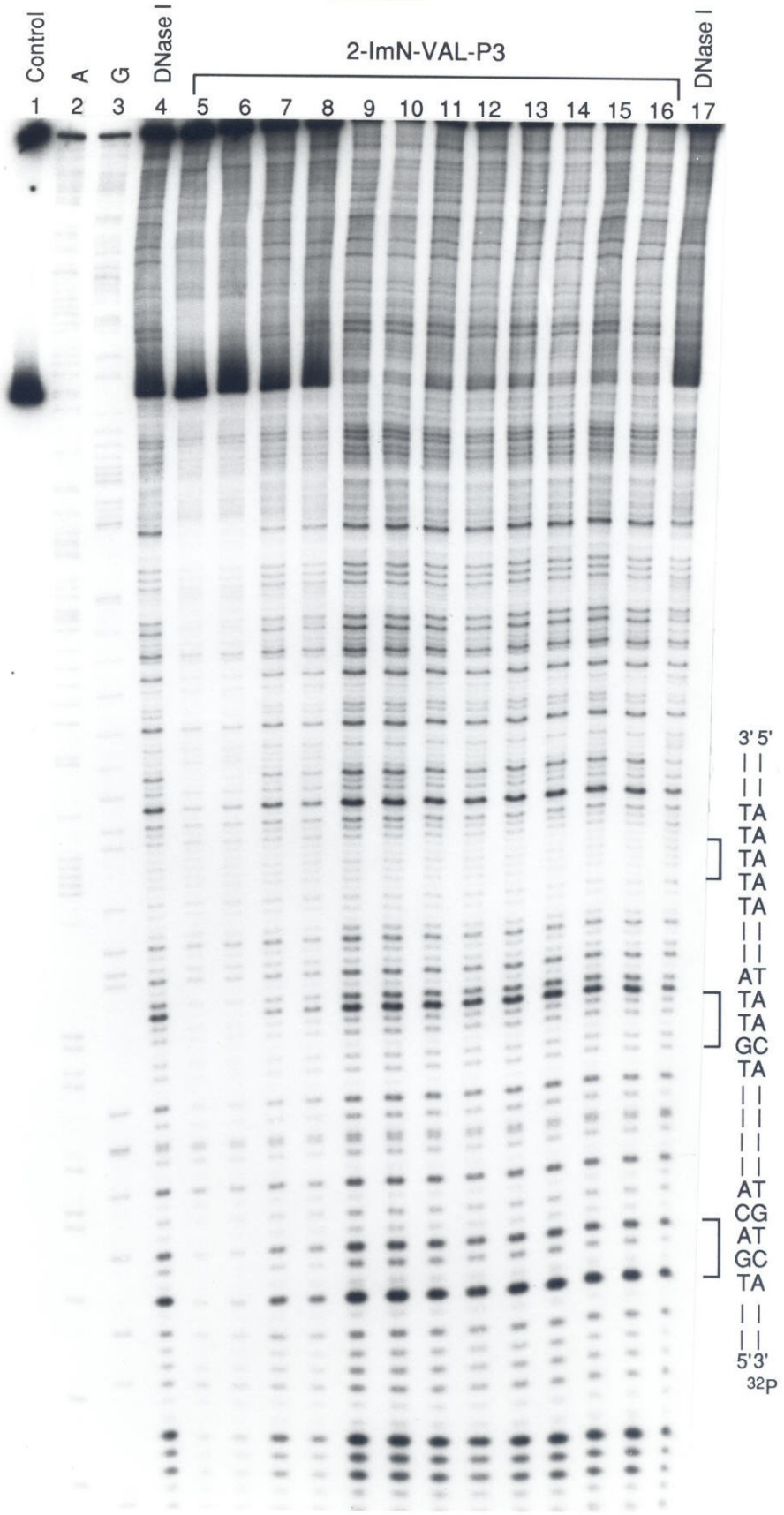

Control
A
G
DNase I
2-ImN-VAL-P3
DNase I
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
3' 5'
TA
TA
TA
TA
TA
AT
TA
TA
GC
TA
AT
CG
AT
GC
TA
5' 3'
32P

Figure 9. Data for the quantitative DNase I footprint titration experiments for (A) 2-ImN-GLY-P3 and (B) 2-ImN-βALA-P3 in complex with the 5′-TGTTA-3′ and 5′-TGACA-3′ sites. The θ_{norm} points were obtained using photostimulable storage phosphor autoradiography and processed as described in the experimental section. The data points for the 5′-TGTTA-3′ site are indicated by open diamonds (◊) and for the 5′-TGACA-3′ site by filled circles (•). The solid and dashed curves are the best-fit Langmuir binding titration isotherms obtained from nonlinear least squares algorithm using eq. 2. The data points for 2-ImN-GLY-P3 and 2-ImN-βALA-P3 were fit with a cooperative model (eq2, n = 2).

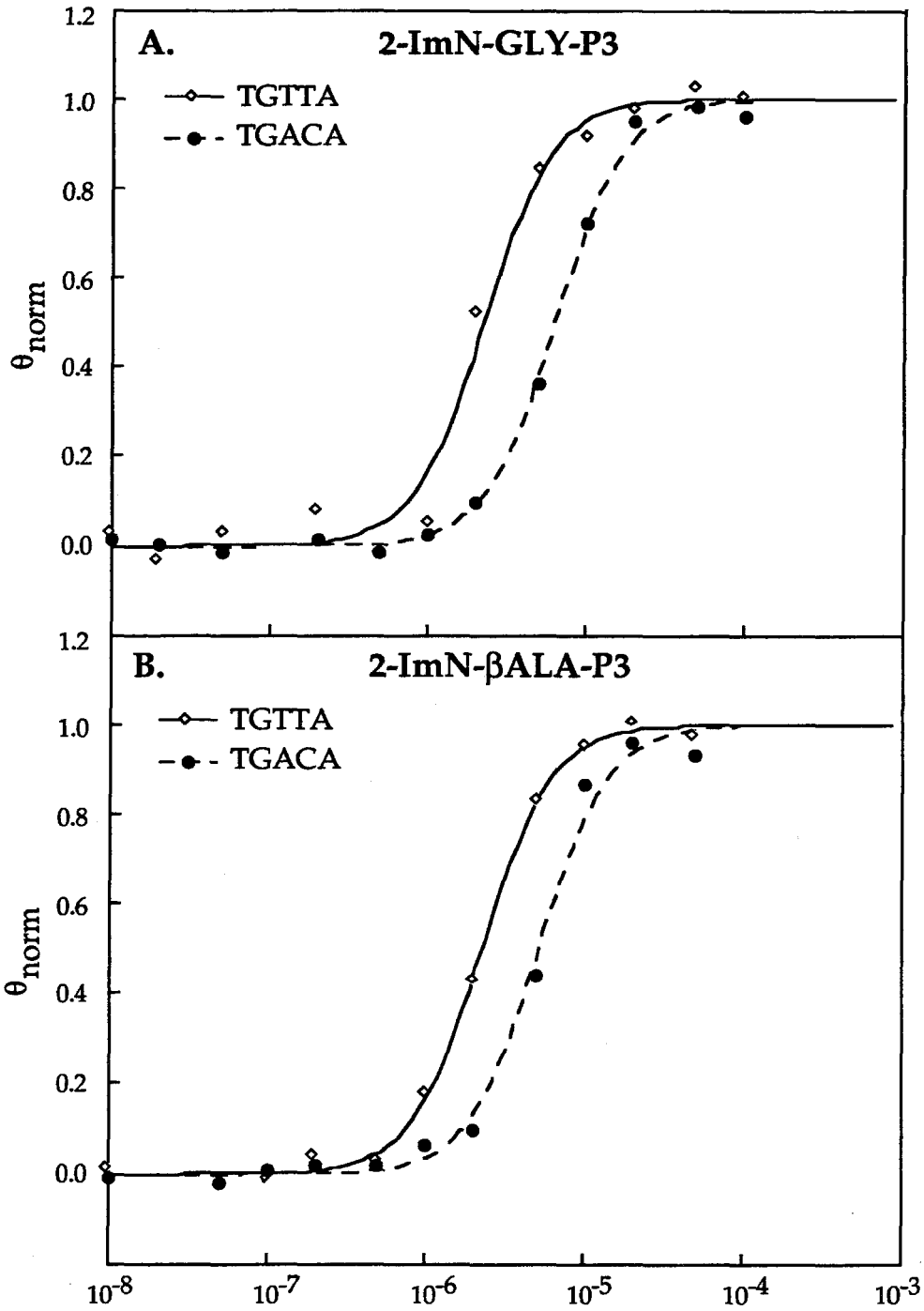

A.
2-ImN-GLY-P3
TGTTA
TGACA
B.
2-ImN-βALA-P3
TGTTA
TGACA
θnorm
θnorm
1.2
1.0
0.8
0.6
0.4
0.2
0.0
10-8
10-7
10-6
10-5
10-4
10-3

Figure 10. Data for the quantitative DNase I footprint titration experiments for (A) 2-ImN-GABA-P3 and (B) 2-ImN-VAL-P3 in complex with the 5′-TGTTA-3′ and 5′-TGACA-3′ sites. The θ_{norm} points were obtained using photostimulable storage phosphor autoradiography and processed as described in the experimental section. The data points for the 5′-TGTTA-3′ site are indicated by open diamonds (◊) and for the 5′-TGACA-3′ site by filled circles (•). The solid and dashed curves are the best-fit Langmuir binding titration isotherms obtained from nonlinear least squares algorithm using eq. 2 (n = 1).

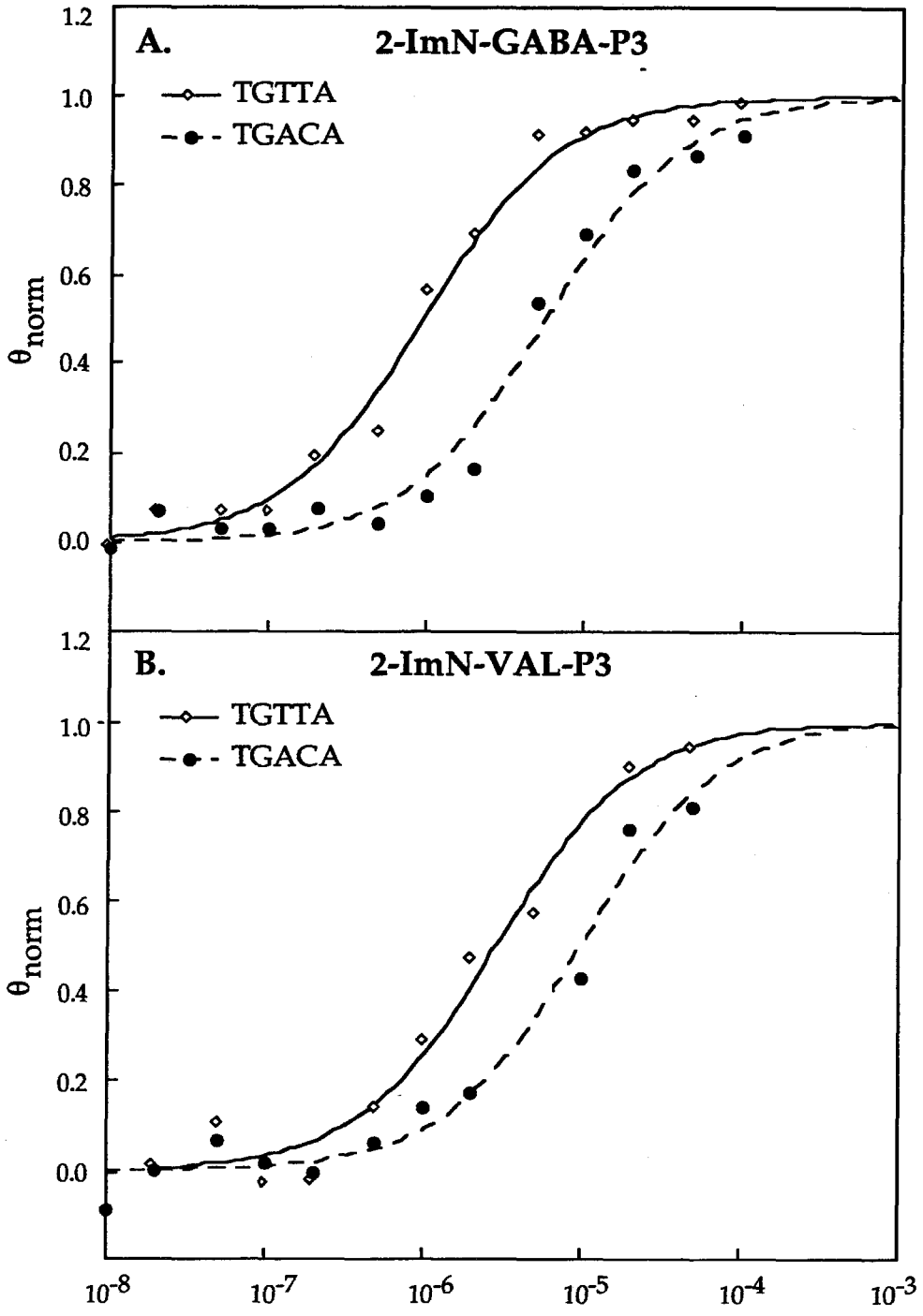

A.
2-ImN-GABA-P3
TGTTA
TGACA
θnorm
1.2
1.0
0.8
0.6
0.4
0.2
0.0
B.
2-ImN-VAL-P3
TGTTA
TGACA
θnorm
1.2
1.0
0.8
0.6
0.4
0.2
0.0
10-8
10-7
10-6
10-5
10-4
10-3

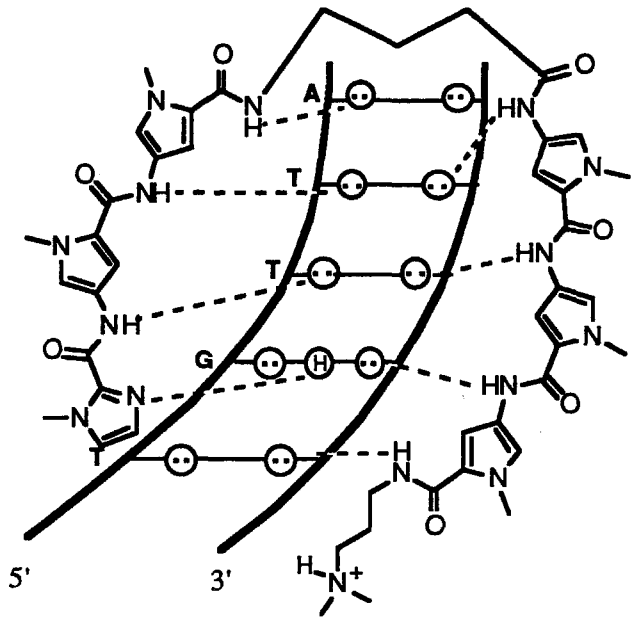
Table 1 Apparent First Order Binding Constants (M^{-1})[a,b]

	Binding Site		
Peptide	5′-TTTTT-3′[c]	5'-TGTTA-3'	5'-TGACA-3'
P3[d]	1.6×10^6 (0.2)	1.7×10^5 (0.1)	$< 1 \times 10^5$
2-ImN[d]	$< 5 \times 10^4$	4.7×10^4 (2.6)	1.5×10^5 (0.1)
2-ImN-C_4-P3[d]	1.7×10^5 (0.5)	1.1×10^6 (0.1)	3.7×10^5 (0.1)
2-ImN-GLY-P3	$\sim 3 \times 10^5$	5.8×10^5 (1.9)	1.7×10^5 (0.2)
2-ImN-βALA-P3	$\sim 1 \times 10^6$	5.4×10^5 (3.1)	3.2×10^5 (1.5)
2-ImN-GABA-P3	$\sim 5 \times 10^5$	1.7×10^6 (0.5)	2.6×10^5 (0.3)
2-ImN-VAL-P3	$\sim 2 \times 10^5$	3.1×10^5 (1.3)	1.7×10^5 (0.6)

[a]Values reported are the mean values measured from four footprint titration experiments, with the standard deviation for each data set indicated in parenthesis. [b]The assays were performed at 22 °C at pH 7.0 in the presence of 10 mM tris acetate, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$. [c]Approximate binding affinities are reported since the quality of fits and standard deviations for these data sets were poor. [d]Binding affinities of these peptides serve as controls and are taken from reference 5a.

Discussion

Binding Affinities. Among the four hexapeptides, 2-ImN-GABA-P3 binds the 5'-TGTTA-3' site with the highest affinity and sequence-specificity. The high binding affinity suggests that the hexapeptide binds this site as an intramolecular hairpin structure (Figure 11). This hairpin peptide configuration binds the site with the same affinity as the first generation covalent heterodimer 2-ImN-C_4-P3. In addition, both peptides bind the "mismatched" 5'-TGACA-3' site with similar three-fold lower affinity. Linking the peptides through either the central pyrrole rings or the terminal amine and carboxylic acid groups are



2-ImN-GABA-P3 • TGTTA

Figure 11. *Model for the complex formed between the covalent peptide* heterodimer 2-ImN-GABA-P3 with the 5′-TGTTA-3′ site. Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dotted lines.

equally effective strategies for improving the binding affinities of side-by-side peptides for DNA sites. With regard to sequence-specificity, 2-ImN-C_4-P3 is the superior peptide since it appears to bind the 5′-TTTTT-3′ site with lower affinity than does 2-ImN-GABA-P3. A possible explanation for this is that the linear peptide dimers are more tolerant of 1:1 binding modes than are the centrally bridged peptides. However, 2-ImN-C_4-P3 has as a limitation a lengthy synthesis which may not be readily applied to other peptide analogs. In contrast, the synthetic methodology for preparation of linear peptide hairpins is *general and easily extended to the preparation of a variety of peptide dimers.*

A comparison of the binding affinities of the four peptides for the 5'-TGTTA-3' site reveals that the linker length has only a modest effect on overall complex stability. Modeling suggested that glycine and β-alanine linkers would not favorably accommodate binding of 2-ImN-GLY-P3 and 2-ImN-βALA-P3 to the 5'-TGTTA-3' site as intramolecular dimers. Instead, these two peptides likely bind as *intermolecular dimers,* consistent with the cooperative binding isotherms from the quantitative footprint titration experiments. Moreover, these peptides display higher non-specific affinity for DNA sites which may result from favorable binding in the minor groove by an extended conformation of peptide (data not shown). Although the binding isotherms suggest that 2-ImN-VAL-P3 binds the 5'-TGTTA-3' site as an intramolecular dimer, the peptide binds with lower affinity as compared to 2-ImN-GABA-P3. This difference may reflect energetic penalties due to unoptimized linker conformation and solvation properties of this peptide. We defer further speculation about these complexes until structural data from NMR spectroscopy are available.

Implications for the Design of Minor-Groove Binding Molecules. The antiparallel side-by-side 2:1 peptide-DNA complex offers a new model for targeting designed peptides to designated sequences based on hydrogen bonding between peptide imidazole and pyrrolecarboxamides and each DNA strand on the floor of the minor groove. Linking the side-by-side peptides improves the affinity and specificity of these ligands for DNA. We anticipate that the general and efficient synthetic methodology for preparation of GABA linked peptides will allow the design of a new class of hairpin peptides for specific recognition of many different sequences. This strategy involves the construction of a peptide for recognition of one strand of the binding site, the GABA linker for the turn and a second peptide for recognition of the other strand. The next generation of hairpin peptides should implement a linker which enforces a turn and disfavors

binding of the peptide in extended conformations which allow 1:1 or intermolecular 2:1 complexes by the tripeptide subunits. Formally, this can be accomplished by either introducing rigid amino acid linkers or, alternatively, preparing cyclic peptides incorporating two GABA linkers.

Experimental Section

^{1}H NMR and ^{13}C NMR spectra were recorded on a General Electric-QE 300 NMR spectrometer in CD_3OD or DMSO-d_6, with chemical shifts reported in parts per million relative to residual CHD_2ODor DMSO-d_5, respectively. The ^{13}C resonances for pyrrole methyl groups are overlapped by the solvent and in general are not listed. IR spectra were recorded on a Perkin-Elmer FTIR spectrometer. High-resolution mass spectra were recorded using fast atom bombardment (FABMS) techniques at the Mass Spectrometry Laboratory at the University of California, Riverside. Reactions were executed under an inert argon atmosphere. Reagent grade chemicals were used as received unless otherwise noted. Tetrahydrofuran (THF) was distilled under nitrogen from sodium/benzophenone ketyl. Dichloromethane (CH_2Cl_2) and triethylamine were distilled under nitrogen from powdered calcium hydride. Dimethylformamide (DMF) was purchased as an anhydrous solvent from Aldrich. Flash chromatography was carried out using EM science Kieselgel 60 (230-400) mesh.[15] Thin-layer chromatography was performed on EM Reagents silica gel plates (0.5 mm thickness). All compounds were visualized with short-wave ultraviolet light.

N-(tert-*butoxycarbonyl*)-*amino acid*-*tris*(N-*methylpyrrolecarboxamide*)s 2a-d (Exemplified with 2a). To a solution of *N*-(*tert*-butoxycarbonyl)glycine (780 mg, 4.457 mmol) in dimethylformamide (4.0 mL) was added a solution of

dicyclohexylcarbodiimide (464 mg, 2.250 mmol) in methylene chloride (4.5 mL) and the solution was allowed to stir 30 min. Separately, a solution of tris(*N*-methylpyrrolecarboxamide) (300 mg, 0.602 mmol) and palladium on carbon (10%, 100 mg) in dimethylformamide (9.0 mL) was allowed to stir under a hydrogen atmosphere (300 psi) in a Parr bomb apparatus for 3 hr. The reaction mixture was filtered through celite to remove catalyst and the filtrate was added to the activated acid, followed by stirring for 2 hr. The mixture was filtered through celite, solvents were removed *in vacuo* and the residue was purified by flash column chromatography (0.5 % NH_4OH in MeOH) to afford **2a**.

2a: yield 86% (331 mg); ^1H NMR (CD_3OD) δ 7.18 (d, 1 H, *J* = 1.9 Hz), 7.16 (d, 1 H, *J* = 1.9 Hz), 7.14 (d, 1 H, *J* = 1.8 Hz), 6.91 (d, 1 H, *J* = 1.9 Hz), 6.82 (d, 1 H, *J* = 1.9 Hz), 6.78 (d, 1 H, *J* = 1.9 Hz), 3.95 (bs, 2 H), 3.89 (s, 3 H), 3.88 (s, 3 H), 3.86 (s, 3 H), 3.29 (m, 2 H), 2.37 (t, 2 H *J* = 7.4 Hz), 2.23 (s, 6 H), 1.73 (m, 2 H), 1.43 (s, 9 H); ^{13}C NMR (CD_3OD) δ 169.5, 164.1, 161.2, 161.1, 158.4, 124.5, 123.2, 122.7, 120.7, 120.6, 120.4, 106.4, 106.1, 106.0, 80.7, 58.2, 44.9, 44.6, 38.5, 36.8, 28.7, 28.2; IR (thin film) 3300 (m), 2942 (w), 1643 (s), 1581 (m), 1538 (s), 1464 (m), 1435 (m), 1405 (m), 1258 (m), 1208 (w), 1164 (m), 1106 (w), 775 (w); FABMS *m/e* 626.3415 (M + H, 626.3432 calcd. for $C_{31}H_{46}N_9O_6$).

2b: yield 83% (320 mg); ^1H NMR (CD_3OD) δ 7.15 (d, 1 H, *J* = 1.8 Hz), 7.13 (d, 1 H, *J* = 1.8 Hz), 7.07 (d, 1 H, *J* = 1.8 Hz), 6.89 (d, 1 H, *J* = 1.7 Hz), 6.83 (d, 1 H, *J* = 1.7 Hz), 6.78 (d, 1 H, *J* = 1.9 Hz), 3.85 (s, 3 H), 3.84 (s, 3 H), 3.82 (s, 3 H), 3.30 (m, 4 H), 2.49 (t, 2 H *J* = 6.8 Hz), 2.41 (t, 2 H *J* = 7.4 Hz), 2.27 (s, 6 H), 1.77 (m, 2 H), 1.42 (s, 9 H); ^{13}C NMR (CD_3OD) δ 170.9, 164.1, 161.3, 158.3, 124.6, 124.5, 123.3, 123.2, 120.8, 120.6, 120.4, 106.5, 106.0, 105.8, 80.1, 58.3, 45.4, 38.6, 38.0, 37.5, 36.8, 36.7, 28.7, 28.3; IR (thin film) 3293 (m), 2944 (w), 1637 (s), 1578 (m), 1542 (m), 1459 (s), 1406 (m), 1260 (m), 1208 (w), 1161 (m), 775 (w); FABMS *m/e* 640.3571 (M + H, 640.3573 calcd. for $C_{30}H_{44}N_9O_6$).

2c: yield 79% (70 mg); ^{1}H NMR (CD_3OD) δ 7.16 (d, 1 H, *J* = 1.9 Hz), 7.14 (d, 1 H, *J* = 1.9 Hz), 7.12 (d, 1 H, *J* = 1.6 Hz), 6.90 (d, 1 H, *J* = 1.7 Hz), 6.81 (d, 1 H, *J* = 1.6 Hz), 6.77 (d, 1 H, *J* = 1.7 Hz), 3.87 (s, 3 H), 3.86 (s, 3 H), 3.84 (s, 3 H), 3.30 (m, 2 H), 3.08 (t, 2 H, *J* = 6.9 Hz), 2.38 (t, 2 H, *J* = 7.8 Hz), 2.32 (t, 2 H, *J* = 7.5 Hz), 2.24 (s, 6 H), 1.77 (m, 4 H), 1.41 (s, 9 H); ^{13}C NMR (CD_3OD) δ 170.7, 162.2, 159.4, 156.5, 122.6, 122.5, 121.2, 118.8, 118.6, 118.4, 104.5, 104.0, 103.9, 78.0, 56.3, 43.4, 38.9, 34.8, 26.8, 26.3, 25.4; IR (thin film) 3301 (m), 2938 (w), 1638 (s), 1578 (m), 1534 (s), 1461 (s), 1436 (s), 1405 (m), 1257 (m), 1205 (w), 1164 (m); FABMS *m/e* 654.3728 (M + H, 654.3695 calcd. for $C_{32}H_{48}N_9O_6$).

2d: yield 84% (385 mg); ^{1}H NMR (CD_3OD) δ 7.17 (d, 1 H, *J* = 1.9 Hz), 7.16 (d, 1 H, *J* = 1.9 Hz), 7.13 (d, 1 H, *J* = 1.9 Hz), 6.91 (d, 1 H, *J* = 1.9 Hz), 6.82 (d, 1 H, *J* = 1.9 Hz), 6.78 (d, 1 H, *J* = 1.9 Hz), 3.89 (s, 3 H), 3.87 (s, 3 H), 3.85 (s, 3 H), 3.30 (m, 2 H), 3.05 (t, 2 H, *J* = 6.8 Hz), 2.41 (t, 2 H, *J* = 7.6 Hz), 2.32 (t, 2 H *J* = 7.4 Hz), 2.26 (s, 6 H), 1.76 (m, 2 H), 1.68 (m, 2 H), 1.51 (m, 2 H), 1.42 (s, 9 H); ^{13}C NMR (CD_3OD) δ 173.1, 164.1, 161.3, 158.5, 124.6, 124.5, 123.2, 120.8, 120.6, 120.4, 106.5, 106.0, 105.9, 79.8, 58.3, 45.4, 40.9, 38.6, 36.8, 30.5, 28.8, 28.3, 24.3; IR (thin film) 3298 (m), 2941 (w), 1642 (s), 1580 (m), 1536 (s), 1466 (m), 1435 (m), 1406 (m), 1258 (m), 1208 (w), 1168 (m), 776 (w); FABMS *m/e* 668.3884 (M + H, 668.3865 calcd. for $C_{33}H_{50}N_9O_6$).

***Amino acid-tris*(N-*methylpyrrolecarboxamide*)*s* 3a-d (Exemplified with 3b).** To a solution of protected amine **2b** (335 mg, 0.535 mmol) in methylene chloride (13.0 mL) was added trifluoroacetic acid (3.0 mL) and the resulting solution was allowed to stir 20 min. Hexanes (200 mL) were added and the solution was allowed to stir 10 min during which an oil collected on the flask. The solvent was removed by decantation, the residue was dissolved in 6% NH_4OH/MeOH (100 mL) and solvent was removed *in vacuo*. Purification of the oil by flash column chromatography (8% NH_4OH/MeOH) afforded amine **3b**.

3a: yield 99% (256 mg); ^{1}H NMR (CD_3OD) δ 7.15 (d, 1 H, *J* = 1.8 Hz), 7.14 (d, 1 H, *J* = 1.8 Hz), 7.13 (d, 1 H, *J* = 1.9 Hz), 6.90 (d, 1 H, *J* = 1.8 Hz), 6.83 (d, 1 H, *J* = 1.8 Hz), 6.77 (d, 1 H, *J* = 1.8 Hz), 3.86 (s, 3 H), 3.85 (s, 3 H), 3.83 (s, 3 H), 3.37 (s, 2 H), 3.30 (m, 2 H), 2.41 (t, 2 H, *J* = 7.7 Hz), 2.27 (s, 6 H), 1.75 (m, 2 H); ^{13}C NMR (CD_3OD) δ 171.3, 163.3, 160.5, 160.4, 123.8, 122.4, 122.3, 120.0, 119.7, 119.6, 105.6, 105.2, 105.0, 57.4, 48.2, 44.5, 44.3, 37.7, 36.0, 27.3; IR (thin film) 3300 (s), 1638 (s), 1578 (s), 1546 (s), 1438 (m), 1406 (m), 1262 (w), 1208 (w), 1149 (w), 1108 (w); FABMS *m/e* 526.2912 (M + H, 526.2890 calcd. for $C_{25}H_{36}N_9O_4$).

3b: yield 91% (264 mg); ^{1}H NMR (CD_3OD) δ 7.16 (d, 1 H, *J* = 1.9 Hz), 7.15 (d, 1 H, *J* = 1.9 Hz), 7.14 (d, 1 H, *J* = 1.9 Hz), 6.91 (d, 1 H, *J* = 1.9 Hz), 6.82 (d, 1 H, *J* = 1.9 Hz), 6.77 (d, 1 H, *J* = 1.9 Hz), 3.88 (s, 3 H), 3.87 (s, 3 H), 3.85 (s, 3 H), 3.30 (m, 2 H), 2.95 (t, 2 H, *J* = 6.6 Hz), 2.48 (t, 2 H, *J* = 6.6 Hz), 2.38 (t, 2 H, *J* = 7.6 Hz), 2.25 (s, 6 H), 1.75 (m, 2 H); ^{13}C NMR (CD_3OD) δ 170.6, 163.3, 160.5, 160.4, 123.8, 123.7, 122.4, 122.3, 120.0, 119.7, 119.5, 105.6, 105.2, 105.0, 57.5, 47.3, 44.6, 38.6, 38.1, 37.8, 35.9, 27.4; IR (thin film) 3336 (s), 1638 (s), 1578 (s), 1560 (s), 1540 (s), 1466 (m), 1458 (m), 1406 (m), 1263 (w), 1209 (w), 1151 (w), 1108 (w); FABMS *m/e* 540.3047 (M + H, 540.3051 calcd. for $C_{26}H_{38}N_9O_4$).

3c: yield 91% (54 mg); ^{1}H NMR (CD_3OD) δ 7.17 (d, 1 H, *J* = 1.9 Hz), 7.16 (d, 1 H, *J* = 1.9 Hz), 7.13 (d, 1 H, *J* = 1.9 Hz), 6.91 (d, 1 H, *J* = 1.8 Hz), 6.82 (d, 1 H, *J* = 1.8 Hz), 6.77 (d, 1 H, *J* = 1.9 Hz), 3.88 (s, 3 H), 3.87 (s, 3 H), 3.85 (s, 3 H), 3.30 (m, 2 H), 2.67 (t, 2 H, *J* = 7.3 Hz), 2.41-2.32 (m, 4 H), 2.24 (s, 6 H), 1.20 (m, 4 H); ^{13}C NMR (CD_3OD) δ 170.5, 163.3, 160.4, 160.3, 123.8, 123.7, 121.1, 120.0, 119.7, 119.5, 105.7, 105.2, 105.0, 57.5, 44.6, 41.2, 37.8, 38.1, 36.0, 35.9, 33.8, 29.2, 27.4; IR (thin film) 3286 (s), 1637 (s), 1580 (s), 1560 (s), 1538 (s), 1466 (m), 1436 (s), 1406 (m), 1262 (w), 1209 (w), 1156 (w), 1108 (w); FABMS *m/e* 554.3203 (M + H, 554.3207 calcd. for $C_{27}H_{40}N_9O_4$).

3d: yield 96% (312 mg); ^{1}H NMR (CD_3OD) δ 7.17 (d, 1 H, *J* = 1.9 Hz), 7.16 (d, 1 H, *J* = 1.9 Hz), 7.14 (d, 1 H, *J* = 1.9 Hz), 6.92 (d, 1 H, *J* = 1.9 Hz), 6.83 (d, 1 H, *J* = 1.9 Hz), 6.78 (d, 1 H, *J* = 1.9 Hz), 3.87 (s, 3 H), 3.85 (s, 3 H), 3.83 (s, 3 H), 3.30 (m, 2 H), 3.86 (t, 2 H, *J* = 7.2 Hz), 2.37 (t, 2 H, *J* = 7.6 Hz), 2.32 (t, 2 H, *J* = 7.3 Hz), 2.23 (s, 6 H), 1.72 (m, 4 H), 1.53 (m, 2 H); ^{13}C NMR (CD_3OD) δ 172.9, 164.1, 161.3, 124.6, 124.5, 123.3, 120.8, 120.6, 120.4, 106.5, 106.0, 105.9, 58.3, 45.4, 41.7, 38.6, 36.8, 36.7, 32.2, 28.3, 24.4, 24.1; IR (thin film) 3287 (s), 2944 (w), 1644 (s), 1581 (s), 1556 (s), 1538 (s), 1467 (m), 1434 (s), 1406 (m), 1262 (m), 1209 (w), 1160 (w), 1107 (w); FABMS *m/e* 568.3360 (M + H, 568.3361 calcd. for $C_{28}H_{42}N_9O_4$).

***Nitro-bis*(N-*methylpyrrolecarboxamide*)amino acid-*tris*(N-*methylpyrrolecarboxamide*)s 4a-d (Exemplified with 4b).** To a solution of dipyrrole (303 mg, 1.039 mmol) and *N*-hydroxybenzotriazole hydrate (141 mg, 1.05 mmol) in dimethylformamide (2.1 mL) was added a solution of dicyclohexylcarbodiimide (217 mg, 1.05 mmol) in methylene chloride (2.1 mL) and the solution was allowed to stir 30 min. A solution of primary amine **3b** (242 mg, 0.449 mmol) in dimethylformamide (3.0 mL) was added and the mixture was allowed to stir for 8 hr. Methanol (1 mL) was added and the reaction was filtered through celite. The filtrate was concentrated *in vacuo* and purified by flash column chromatography (1% NH_4OH/MeOH) to afford **4b**.

4a: yield 86% (323 mg); ^{1}H NMR (DMSO-d_6) δ 10.29 (s, 1 H), 9.92 (bs, 2 H), 9.89 (bs, 1 H), 8.37 (t, 1 H, *J* = xx Hz), 8.18 (d, 1 H, *J* = 1.6 Hz), 8.07 (t, 1 H, *J* = xx Hz), 7.59 (d, 1 H, *J* = 1.9 Hz), 7.27 (d, 1 H, *J* = 1.8 Hz), 7.23 (d, 1 H, *J* = 1.8 Hz), 7.18 (d, 1 H, *J* = 1.9 Hz), 7.17 (d, 1 H, *J* = 1.9 Hz), 7.02 (d, 1 H, *J* = 1.8 Hz), 6.94 (d, 1 H, *J* = 1.9 Hz), 6.92 (d, 1 H, *J* = 1.9 Hz), 6.81 (d, 1 H, *J* = 1.8 Hz), 3.95 (s, 3 H), 3.90 (bs, 2 H), 3.83 (s, 6 H), 3.82 (s, 3 H), 3.78 (s, 3 H), 3.34 (m, 2 H), 2.22 (t, 2 H, *J* = 7.1 Hz), 2.12 (s, 6 H), 1.59 (m, 2 H); IR (KBr) 3310 (m), 2934 (w), 1638 (s), 1578 (s),

1534 (s), 1466 (m), 1438 (s), 1406 (m), 1311 (s), 1256 (m), 1208 (m), 1107 (m); FABMS *m/e* 800.3592 (M + H, 800.3578 calcd. for $C_{37}H_{46}N_{13}O_8$).

4b: yield 87% (319 mg); ^{1}H NMR (CD_3OD) δ 7.73 (d, 1 H, *J* = 1.8 Hz), 7.32 (d, 1 H, *J* = 1.9 Hz), 7.13 (d, 1 H, *J* = 1.8 Hz), 7.12 (d, 1 H, *J* = 1.9 Hz), 7.11 (d, 1 H, *J* = 1.9 Hz), 7.10 (d, 1 H, *J* = 1.9 Hz), 6.85 (d, 1 H, *J* = 1.9 Hz), 6.79 (d, 1 H, *J* = 1.9 Hz), 6.76 (d, 1 H, *J* = 1.9 Hz), 6.75 (d, 1 H, *J* = 1.9 Hz), 3.89 (s, 3 H), 3.82 (s, 3 H), 3.81 (s, 6 H), 3.79 (s, 3 H), 3.60 (bs, 2 H), 3.35 (m, 2 H), 2.58 (t, 2 H, *J* = 6.4 Hz), 2.35 (t, 2 H, *J* = 7.7 Hz), 2.22 (s, 6 H), 1.73 (m, 2 H); ^{13}C NMR (CD_3OD) δ 171.0, 164.1, 164.0, 161.3, 161.2, 159.5, 136.0, 128.6, 127.6, 124.6, 124.5, 123.2, 123.1, 122.8, 120.8, 120.7, 120.5, 120.4, 108.7, 106.4, 106.0, 105.9, 58.3, 45.4, 38.6, 38.0, 37.1, 37.0, 36.8, 36.7, 28.3; IR (neat) 3284 (m), 2942 (w), 1638 (s), 1578 (s), 1534 (s), 1466 (m), 1438 (s), 1402 (m), 1308 (s), 1259 (m), 1207 (m), 1109 (w); FABMS *m/e* 814.3749 (M + H, 814.3739 calcd. for $C_{38}H_{48}N_{13}O_8$).

4c: yield 81% (28 mg); ^{1}H NMR (CD_3OD) δ 7.78 (d, 1 H, *J* = 1.8 Hz), 7.35 (d, 1 H, *J* = 1.9 Hz), 7.15 (d, 1 H, *J* = 1.9 Hz), 7.14 (d, 1 H, *J* = 1.9 Hz), 7.13 (d, 1 H, *J* = 1.8 Hz), 7.06 (d, 1 H, *J* = 1.9 Hz), 6.86 (d, 1 H, *J* = 1.9 Hz), 6.79 (d, 2 H, *J* = 1.8 Hz), 6.76 (d, 1 H, *J* = 1.9 Hz), 3.94 (s, 3 H), 3.85 (s, 3 H), 3.84 (s, 3 H), 3.83 (s, 3 H), 3.81 (s, 3 H), 3.35 (q, 2 H, *J* = 5.6 Hz), 3.30 (m, 2 H), 2.38 (t, 2 H, *J* = 7.6 Hz), 2.24 (s, 6 H), 1.94 (m, 2 H), 1.75 (m, 2 H); ^{13}C NMR (CD_3OD) δ 172.8, 164.2, 161.3, 161.2, 159.5, 136.1, 128.7, 127.7, 124.6, 124.5, 124.4, 123.2, 123.1, 122.8, 120.8, 120.6, 120.4, 108.7, 106.4, 106.0, 58.4, 45.4, 39.9, 38.6, 38.1, 36.8, 34.9, 28.3, 27.0; IR (neat) 3281 (m), 2931 (w), 1633 (s), 1578 (m), 1530 (s), 1463 (m), 1434 (s), 1400 (s), 1308 (s), 1255 (m), 1205 (m), 1105 (w); FABMS *m/e* 828.3905 (M + H, 828.3923 calcd. for $C_{39}H_{50}N_{13}O_8$).

4d: yield 87% (205 mg); ^{1}H NMR (CD_3OD) δ 7.76 (d, 1 H, *J* = 1.9 Hz), 7.35 (d, 1 H, *J* = 1.9 Hz), 7.14 (d, 1 H, *J* = 1.9 Hz), 7.13 (d, 2 H, *J* = 1.8 Hz), 7.09 (d, 1 H, *J* = 1.8 Hz), 6.87 (d, 1 H, *J* = 1.8 Hz), 6.81 (d, 1 H, *J* = 1.8 Hz), 6.79 (d, 1 H, *J* = 1.9

Hz), 6.75 (d, 1 H, J = 1.9 Hz), 3.92 (s, 3 H), 3.84 (s, 3 H), 3.83 (s, 3 H), 3.82 (s, 3 H), 3.81 (s, 3 H), 3.30 (m, 4 H), 2.34 (m, 4 H), 2.22 (s, 6 H), 1.72 (m, 4 H), 1.60 (m, 2 H); ^{13}C NMR (CD_3OD) δ 173.0, 164.1, 161.3, 159.5, 136.1, 128.7, 127.6, 124.8, 124.6, 124.5, 123.2, 122.8, 120.8, 120.6, 120.4, 108.7, 106.5, 106.0, 105.9, 105.8, 58.3, 45.4, 39.8, 38.6, 38.1, 36.8, 30.3, 28.3, 24.4; IR (neat) 3277 (m), 2942 (w), 1638 (s), 1578 (m), 1535 (s), 1460 (m), 1432 (s), 1401 (s), 1308 (s), 1258 (m), 1206 (m), 1109 (w); FABMS *m/e* 842.4062 (M + H, 842.4052 calcd. for $C_{40}H_{52}N_{13}O_8$).

***2-ImN-amino acid-P3* 1a-d (Exemplified with 1d).** To a solution of 1-methylimidazole-2-carboxylic acid (67 mg, 0.603 mmol) and *N*-hydroxybenzotriazole hydrate (82 mg, 0.60 mmol) in dimethylformamide (1.2 ml) was added a solution of 1,3-dicyclohexylcarbodiimide (124 mg, 0.60 mmol) in methylene chloride (1.2 ml) and the mixture was allowed to stir for 50 min. Separately, to a solution of **4d** (150 mg, 0.176 mmol) in dimethylformamide (4.0 ml) was added Pd/C catalyst (10%, 52 mg) and the mixture was hydrogenated in a Parr bom apparatus (325 psi H_2) for 3 hr. The catalyst was removed by filtration through celite and the filtrate was immediately added to the activated acid and allowed to stir 2.5 hr. Methanol (1.0 ml) was added, the mixture was filtered through celite and the filtrate was concentrated *in vacuo*. Flash column chromatography of the residue (1% ammonium hydroxide in methanol) provided the peptide dimer 2-ImN-VAL-P3.

1a: yield 85% (81 mg); ^{1}H NMR (CD_3OD) δ 7.29 (d, 1 H, J = 1.9 Hz), 7.23 (d, 1 H, J = 1.1 Hz), 7.21 (d, 1 H, J = 1.9 Hz), 7.17 (d, 1 H, J = 1.9 Hz), 7.16 (d, 1 H, J = 1.9 Hz), 7.12 (d, 1 H, J = 1.9 Hz), 7.04 (d, 1 H, J = 1.1 Hz), 6.93 (d, 1 H, J = 1.9 Hz), 6.91 (d, 1 H, J = 1.9 Hz), 6.89 (d, 1 H, J = 1.9 Hz), 6.87 (d, 1 H, J = 1.9 Hz), 6.78 (d, 1 H, J = 1.9 Hz), 4.07 (bs, 2 H), 4.04 (s, 3 H), 3.91 (s, 3 H), 3.88 (s, 3 H), 3.86 (s, 6 H), 3.84 (s, 3 H), 3.32 (m, 2 H), 2.43 (t, 2 H, J = 7.6 Hz), 2.28 (s, 6 H), 1.77 (m, 2 H); IR (neat) 3334 (s), 2931 (w), 1636 (s), 1577 (m), 1559 (m), 1542 (m), 1466 (m), 1436

(m), 1405 (m), 1311 (s), 1260 (m), 1210 (m), 1026 (w); UV (H_20) λ_{max} (ε) 246 (36 600), 304 (42 500) nm; FABMS *m/e* 878.4174 (M + H, 878.4225 calcd. for $C_{42}H_{52}N_{15}O_7$).

1b: yield 78% (80 mg); ^{1}H NMR (DMSO-d_6) δ 7.38 (d, 1 H, *J* = 0.9 Hz), 7.27 (d, 1 H, *J* = 1.9 Hz), 7.23 (d, 1 H, *J* = 1.9 Hz), 7.19 (d, 1 H, *J* = 1.8 Hz), 7.18 (d, 1 H, *J* = 1.9 Hz), 7.17 (d, 1 H, *J* = 1.8 Hz), 7.14 (d, 1 H, *J* = 1.9 Hz), 7.03 (d, 1 H, *J* = 1.0 Hz), 7.02 (d, 1 H, *J* = 1.8 Hz), 6.87 (d, 1 H, *J* = 1.9 Hz), 6.84 (d, 1 H, *J* = 1.9 Hz), 6.81 (d, 1 H, *J* = 1.8 Hz), 3.98 (s, 3 H), 3.83 (bs, 9 H), 3.80 (s, 3 H), 3.78 (s, 3 H), 3.43 (m, 2 H), 3.17 (t, 2 H, *J* = 6.4 Hz), 2.52 (m, 2 H), 2.22 (t, 2 H, *J* = 7.0 Hz), 2.12 (s, 6 H), 1.59 (m, 2 H); IR (neat) 3338 (m), 2925 (w), 1638 (s), 1585 (s), 1534 (s), 1466 (m), 1438 (s), 1401 (m), 1260 (m), 1206 (m); UV (H_20) λ_{max} (ε) 244 (36 100), 306 (48 700) nm; FABMS *m/e* 892.4331 (M + H, 892.4329 calcd. for $C_{43}H_{54}N_{15}O_7$).

1c: yield 55% (13 mg); ^{1}H NMR (CD_3OD) δ 7.28 (d, 1 H, *J* = 1.9 Hz), 7.23 (d, 1 H, *J* = 1.1 Hz), 7.15 (d, 3 H, *J* = 1.8 Hz), 7.09 (d, 1 H, *J* = 1.9 Hz), 7.04 (d, 1 H, *J* = 1.1 Hz), 6.90 (d, 1 H, *J* = 1.9 Hz), 6.88 (d, 1 H, *J* = 1.9 Hz), 6.81 (d, 1 H, *J* = 1.9 Hz), 6.80 (d, 1 H, *J* = 1.9 Hz), 6.77 (d, 1 H, *J* = 1.9 Hz), 4.04 (s, 3 H), 3.90 (s, 3 H), 3.87 (s, 3 H), 3.86 (s, 3 H), 3.85 (s, 3 H), 3.84 (s, 3 H), 3.38 (t, 2 H, *J* = 6.3 Hz), 3.30 (m, 2 H), 2.45-2.38 (m, 4 H), 2.28 (s, 6 H), 1.96 (m, 2 H), 1.77 (m, 2 H); IR (neat) 3384 (m), 2936 (w), 1637 (s), 1579 (m), 1542 (m), 1466 (s), 1406 (m), 1261 (m); UV (H_20) λ_{max} (ε) 244 (38 900), 308 (48 100) nm; FABMS *m/e* 906.4487 (M + H, 906.4500 calcd. for $C_{44}H_{56}N_{15}O_7$).

1d: yield 72% (116 mg); ^{1}H NMR (DMSO-d_6) δ 7.38 (d, 1 H, *J* = 0.6 Hz), 7.27 (d, 1 H, *J* = 1.8 Hz), 7.23 (d, 1 H, *J* = 1.9 Hz), 7.19 (bs, 3 H), 7.16 (d, 1 H, *J* = 1.9 Hz), 7.03 (d, 1 H, *J* = 0.8 Hz), 7.01 (d, 1 H, *J* = 1.8 Hz), 6.87 (d, 2 H, *J* = 1.7 Hz), 6.80 (d, 1 H, *J* = 1.7 Hz), 3.98 (s, 3 H), 3.83 (bs, 6 H), 3.81 (s, 3 H), 3.78 (bs, 6 H), 3.43 (m, 2 H), 3.17 (b, 2 H), 2.22 (m, 4 H), 2.12 (s, 6 H), 1.60-1.46 (m, 6 H); IR (neat) 3301 (m), 2935 (w), 1632 (s), 1581 (s), 1535 (s), 1506 (m), 1466 (m), 1432 (m), 1401 (m),

1260 (m), 1204 (m); UV (H_20) λ_{max} (ε) 246 (38 700), 310 (52 900) nm; FABMS *m/e* 920.4644 (M + H, 920.4640 calcd. for $C_{45}H_{58}N_{15}O_7$).

Construction of Plasmid DNA. Using T4 DNA ligase, the plasmid pMM5 was constructed by ligation of an insert, 5'-TCGACATGACATTCGTCCACATTGTTAGACCACGATCGTTTTTCGCATG-3' and 5'-CGAAAAACGATCGTGGTCTAACAATGTGGACGAATGTCATG-3' into pUC19 previously cleaved with *Sal* I and *Sph* I. Ligation products were used to transform Epicurian^{TM} Coli XL 1 Bluecompetent cells. Colonies were selected for α-complementation on 25 mL Luria-Bertani medium agar plates containing 50 μg/mL ampicillin and treated with XGAL and IPTG solutions. Large scale plasmid purification was performed using Qiagen purification kits. Plasmid DNA concentration was determined at 260 nm using the relation 1 OD unit = 50 μg/mL duplex DNA. The plasmid was digested with *Eco*R I, labeled at the 3'-end, and digested with *Afl* III. The 447-base pair restriction fragment was isolated by nondenaturing gel electrophoresis and used in all experiments described here. Chemical sequencing reactions were performed as described.[16,17] Standard techniques were employed for DNA manipulations.[18]

Quantitative DNase I Footprint Titration. All reactions were executed in a total volume of 10 μL with final concentrations of each species as indicated. The ligands, ranging from 100 μM to 10 nM, were added to solutions of radiolabeled restriction fragment (20,000 cpm), Tris •HCl (10 mM, pH 7.0), KCl (10 mM), $MgCl_2$ (10 mM) and $CaCl_2$ (5 mM) and incubated for 25 min at 22 °C. Footprinting reactions were initiated by the addition of 1 μL a stock solution of DNase I (10 units/mL) containing 1 mM dithiothreitol and allowed to proceed for 3 min at 22 °C. The reactions were stopped by addition of a 3 M ammonium acetate solution containing 250 mM EDTA and ethanol precipitated. The reactions were resuspended in 100 mM tris-borate-EDTA/80% formamide

loading buffer and electrophoresed on 6% polyacrylamide denaturing gels (5% crosslink, 7 M urea) at 1500 V for 3-4 hr. The footprint titration gels were dried and quantitated using storage phosphor technology.

Apparent first order binding constants were determined as previously described.[4c,9a] The data were analyzed by performing volume integrations of the 5′-TGTTA-3′ and 5′-TGACA-3′ target sites and a 5′-GCGG-3′ reference site. The apparent DNA target site saturation, θ_{app}, was calculated for each concentration of peptide using the following equation:

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

where I_{tot} and I_{ref} are the integrated volumes of the target and reference sites, respectively, and I_{tot}° and I_{ref}° correspond to those values for a DNase I control lane to which no peptide has been added. At higher concentrations of peptide (> 20 μM), the reference sites become partially protected, resulting in low θ_{app} values. For these data points, the reference value was determined from the amount of radioactivity loaded per lane, using the mean value for all data points from lanes with < 20 μM peptide. The ($[L]_{tot}$, θ_{app}) data points were fit to a Langmuir binding isotherm (eq 2, n=1) by minimizing the difference between θ_{app} and θ_{fit}, using the modified Hill equation:

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

where $[L]_{tot}$ corresponds to the total peptide concentration, K_a corresponds to the apparent monomeric association constant, and θ_{min} and θ_{max} represent the experimentally determined site saturation values when the site is unoccupied or saturated, respectively. Data were fit using a nonlinear least-squares fitting procedure of KaleidaGraph software (version 2.1, Abelbeck software) running on a Macintosh IIfx computer with K_a, θ_{max}, and θ_{min} as the adjustable parameters. The goodness-of-fit of the binding curve to the data points is evaluated by the correlation coefficient, with $R > 0.97$ as the criterion for an acceptable fit. Four sets of acceptable data were used in determining each association constant. All lanes from each gel were used unless visual inspection revealed a data point to be obviously flawed relative to neighboring points. The data were normalized using the following equation:

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

The best fit isotherms for 2-ImN-GLY-P3 and 2-ImN-βALA-P3 binding the 5'-TGTTA-3' and 5'-TGACA-3' sites consistently give much worse fits than those for 2-ImN-GABA-P3 and 2-ImN-VAL-P3 binding these sites. Visual inspection of the binding curves for the former peptides reveals that the increase in θ_{max} near half-saturation of the site is steeper than expected from the fitted curve. These data were fit to a general Hill model (eq 2) with K_a, θ_{max}, and θ_{min} and n as adjustable parameters. For both 2-ImN-GLY-P3 and 2-ImN-βALA-P3, the best fit value of n was in all cases greater than 2.0. Therefore, these data were fit to a modified Hill equation (eq 2, n=2). The data for 2-ImN-GABA-P3 and 2-ImN-VAL-P3 were fit to Langmuir binding isotherms (eq 2, n=1). We note explicitly that treatment of the data in this manner does not represent an attempt to model

a binding mechanism. Rather, we have chosen to compare values of the apparent first order dissociation constant because this paramter represents the concentration of peptide at which the binding site is half-saturated.

References

1. (a) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 1376-1380. (b) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *J. Mol. Biol.* **1985**, *183*, 553-563. (c) Coll, M.; Frederick, C. A.; Wang, A. H.-J.; Rich, A. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 8385-8389.
2. (a) Schultz, P. G.; Dervan, P. B. *Proc. Natl. Acad. Sci.* **1983**, *80*, 6834-6837. (b) Youngquist, R. S.; Dervan, P. B. *Proc. Natl. Acad. Sci.* **1985**, *82*, 2565-2569. (c) Youngquist, R. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1985**, *107*, 5528-5529. (d) Griffin, J. H.; Dervan, P. B. *J. Am. Chem. Soc.* **1986**, *108*, 5008-5009. (e) Youngquist, R. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1987**, *109*, 7564-7566.
3. (a) Wade, W. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1987**, *109*, 1574-1575. (b) Wade, W. S. Ph.D. Thesis, California Institute of Technology, 1989.
4. (a) Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 8783-8794. (b) Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Dervan, P. B. *Proc. Natl. Acad. Sci., USA* **1992**, *89*, 7586-7590. (c) Wade, W. S.; Mrksich, M.; Dervan, P. B. *Biochemistry* **1993**, *32*, 11385-11389.
5. (a) Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 5723-5727. (b) Pelton, J. G.; Wemmer, D. E. *J. Am. Chem. Soc.* **1990**, *112*, 1393-1399.
6. (a) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 2572-2576. (b) Geierstanger, B. H.; Jacobsen, J-P.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Biochemistry*, in press.

7. Geierstanger, B. H.; Dwyer, T. J.; Bathini, Y.; Lown, J. W.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, *115*, 4474-4482.

8. (a) Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Journal* submitted for publication. (b) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* submitted for publication.

9. (a) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 9892-9899. (b) Dwyer, T. J.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, *115*, 9900-9906.

10. Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* submitted for publication.

11. Schultz, P. G.; Taylor, J. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1982**, *104*, 6861-6863.

12. (a) Bialer, M.; Yagen, B.; Mechoulam, R. *Tetrahedron* **1978**, *34*, 2389-2391. (b) Lown, J. W.; Krowicki, K. J. *J. Org. Chem.* **1985**, *50*, 3774-3779.

13. (a) Galas, D.; Schmitz, A. *Nucl. Acids. Res.* **1978**, *5*, 3157-3170. (b) Fox, K. R.; Waring, M. J. *Nucl. Acids Res.* **1984**, *12*, 9271-9285.

14. (a) Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Methods Enzymol.* **1986**, *130*, 132-181. (b) Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 8462-8466. (c) Senear, D. F.; Brenowitz, M.; Shea, M. A.; Ackers, G. K. *Biochemistry* **1986**, *25*, 7344-7354.

15. Still, W. C.; Kahn, M.; Mitra, A. *J. Org. Chem.* **1978**, *40*, 2923-2925.

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

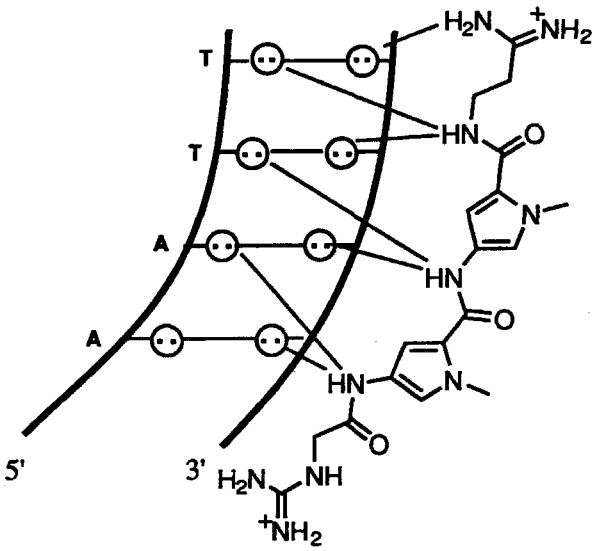
17. Maxam, A. M.; Gilbert, W. S. *Methods in Enzymology* **1980**, *65*, 499-560.

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

Chapter 6

Sequence-Specific Recognition in the Minor Groove of DNA at 5'-(A,T)GCGC(A,T)-3' Sites by the Designed Peptide ImPImP

1:1 Peptide-DNA Complexes. The natural products netropsin and distamycin A (D) are crescent shaped di- and tripeptides, respectively, that bind in the minor groove of DNA at sites of four or five successive A,T base pairs (bp).[1-3] The structures of a number of peptide-DNA complexes have been determined by X-ray diffraction[4] and NMR spectroscopy,[5] and the thermodynamic profiles have been studied for these complexes (Figure 1).[6] This work suggests that favorable electrostatic interactions and extensive van der Waals contacts between the peptide and the floor and walls of the minor groove contribute to complex stability. The carboxamide NH's of the peptides participate in bifurcated hydrogen bonds with adenine N3 and thymidine O2 atoms on the floor of the minor groove. The aromatic hydrogens of the *N*-methylpyrrole rings are set too deeply in the minor groove to allow room for the guanine 2-amino group of a G,C base pair, affording binding specificity for A,T-rich sequences. Although this model has aided in the design of oligopeptides for recognition of longer tracts of A,T-rich DNA,[7] efforts to design peptides capable of binding mixed A,T and G,C sequences have proven inconsistent with a 1:1 peptide-DNA model.[8,9]



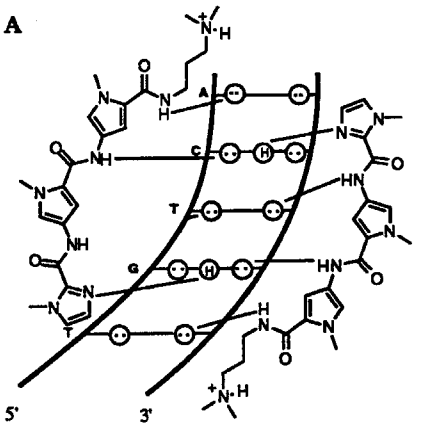
N • AATT

Figure 1. 1:1 Complex of netropsin with 5'-AATT-3'.[4a] Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Putative hydrogen bonds are illustrated by dashed lines.

2:1 Peptide-DNA Complexes. The initial report by Pelton and Wemmer that distamycin (at 2-4 mM) is capable of binding in the minor groove of 5'-AAATT-3' as a side-by-side dimer[10] has been followed by several studies describing 2:1 peptide-DNA complexes. Whereas the peptides pyridine-2-carboxamide-netropsin (2-PyN) and 1-methylimidazole-2-carboxamide-netropsin (2-ImN) were designed to bind 5'-G(A,T)$_3$-3' sequences as 1:1 complexes, they bind the mixed sequence 5'-TGTCA-3' as 2:1 complexes (Figure 2A).[11] Subsequently, we demonstrated that distamycin and 2-ImN sequence-specifically bind the sequence 5'-TGTTA-3' as a side-by-side heterodimer (Figure 2B).[12] In a related study, Wemmer, Lown and co-workers have shown by NMR that D and a distamycin analog containing a central imidazole bind a 5'-AAGTT-3' site as a side-by-side heterodimer (Figure 2C).[13] These examples of specific recognition of mixed sequences by dimeric peptides suggest that the 2:1 motif may serve as a

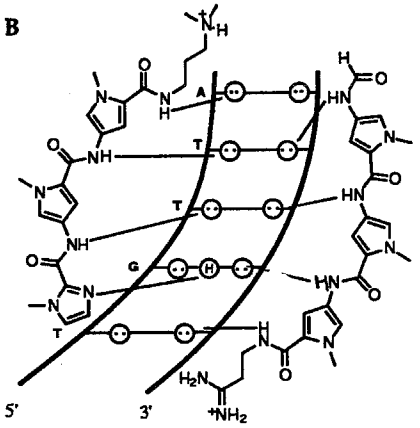
Figure 2. 2:1 binding models for the complexes formed between (a) 2-ImN with a 5'-TGTCA-3' sequence,[11] (b) 2-ImN and D with a 5'-TGTTA-3' sequence[12] and (c) PImP and D with a 5'-AAGTT-3' sequence.[13] Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines and circles containing a H represent the 2-amino group of guanine. Putative hydrogen bonds are illustrated by dashed lines.

A



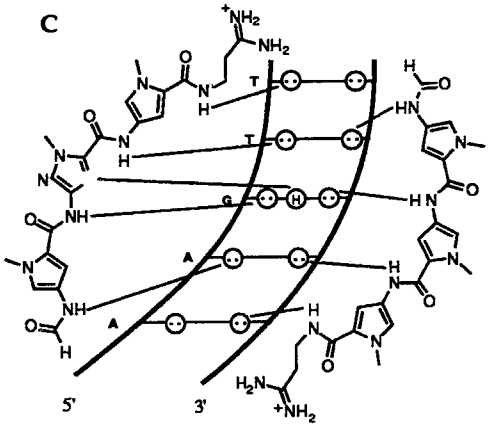
$(2\text{-ImN})_2 \cdot TGTCA$

B

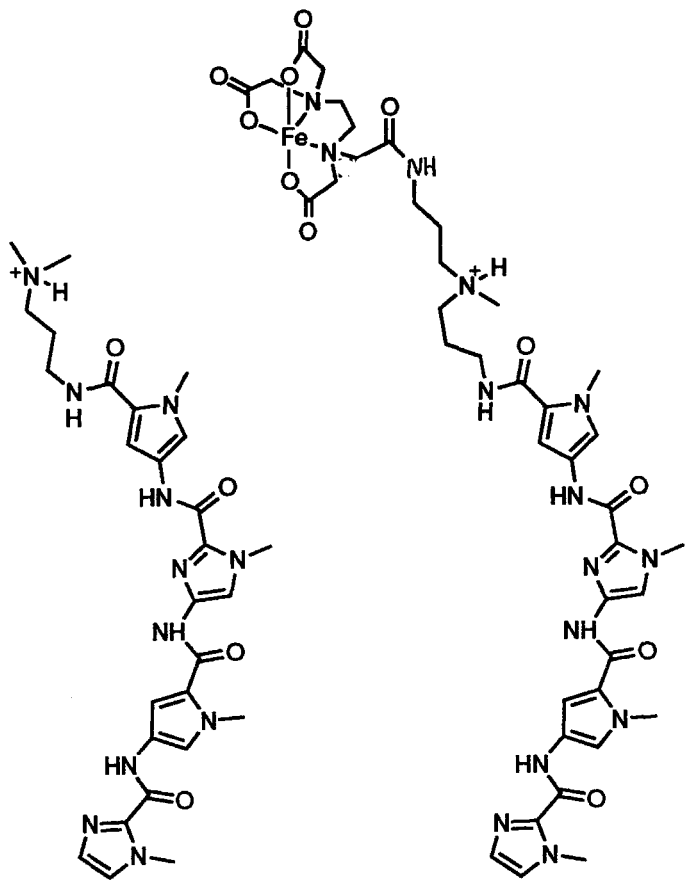


(2-ImN/D) • TGTTA

C



(PImP/D) • AAGTT



ImPImP ImPImPE•Fe(II)

Figure 3. Synthetic peptide ImPImP and affinity cleaving analog ImPImPE•Fe(II).

general model for the design of peptide analogs for specific recognition of other sequences.

Design Rationale. The revised design rationale prompted by 2:1 models differs significantly from that of 1:1 models in that *each peptide is targeted to bases on a single strand of the binding site.* Specific recognition of a guanine residue is afforded by a hydrogen bond to an imidazolecarboxamide of the side-by-side peptides (Figure 2). Recognition of adenine, thymine and cytosine bases is afforded by a hydrogen bonds to pyrrolecarboxamides of the peptides. Based on this new understanding, ImPImP was designed to recognize the family of six-

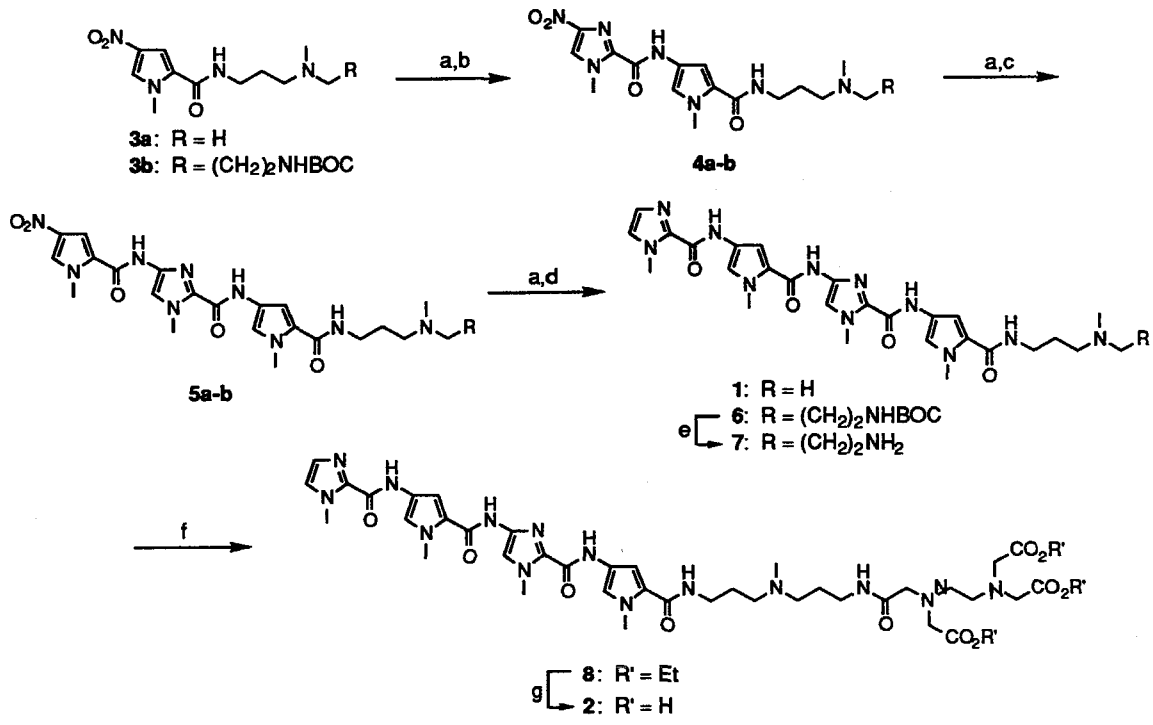


Figure 4. Synthetic scheme for ImPImP and ImPImPE. (a) 300 psi H_2, 10% Pd/C; (b) 4-nitro-1-methylimidazole-2-carboxylic acid, DCC, HOBt; (c) 4-nitro-1-methylpyrrole-2-carboxylic acid, $SOCl_2$; (d) 1-methylimidazole-2-carboxylic acid, DCC, HOBt; (e) 20 % TFA/CH_2Cl_2; (f) $(EtO_2CCH_2)_2N(CH_2)_2N(CH_2CO_2Et)(CH_2CO_2H)$, DCC, HOBt; (g) LiOH, EtOH, H_2O.

base pair G,C-rich sites 5′-(A,T)GCGC(A,T)-3′ as a side-by-side antiparallel dimer in the minor groove (Figure 3). Each imidazole N3 of two side-by-side peptides is expected to form a specific hydrogen bond to unique guanine amino groups on the floor of the minor groove. Consideration of the geometry of the 2:1 peptide-DNA complex suggests that the DNA site containing four alternating GC base pairs would have the proper disposition of guanine amino groups for recognition by the ImPImP dimer. We report here the synthesis of ImPImP and the affinity cleaving analog ImPImPE and DNA binding studies with these peptides.

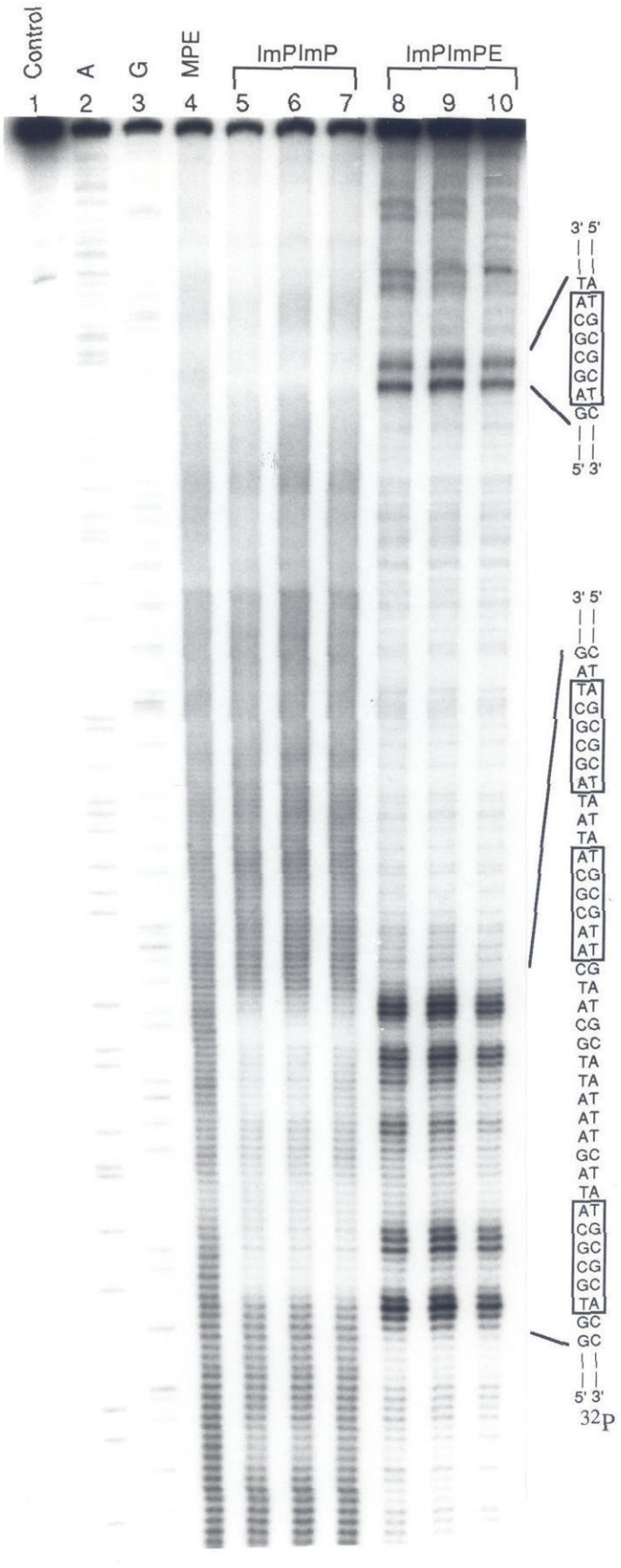
Results

Synthesis of ImPImP and ImPImPE. The synthetic methodology for these peptides is similar to that previously described for related analogs (Figure 4).[3b,11a] Coupling of 1-methyl-4-amino-2-(carboxamidopropyl)-3-(dimethylamino)-pyrrole with 4-nitro-1-methylimidazole-2-carboxylic acid (DCC, HOBt) affords the imidazole-pyrrole derivative **4**. Reduction of the nitroimidazole (300 psi H_2, Pd/C) and coupling with 4-nitro-1-methylpyrrole-2-acyl chloride yields **5**. Further reduction of the nitropyrrole and coupling with 1-methylimidazole-2-carboxylic acid (DCC, HOBt) provides ImPImP. The affinity cleaving analog ImPImPE was synthesized in similar fashion, with elaboration of nitropyrrole **3b** to the *tert*-butoxycarbonylamine **6**. Deprotection (TFA, CH_2Cl_2) and coupling of the primary amine with triethyl ethylenediaminetriacetate acetic acid (DCC, HOBt) gave the triethyl ester of ImPImPE. Saponification of this material (LiOH, EtOH, H_2O) afforded ImPImPE.

Footprinting. MPE•Fe(II) footprinting on the 3′- and 5′-^{32}P end-labeled 381-base pair *Eco*R 1/*Bam*H 1 restriction fragment from plasmid pBR322 (25 mM Tris•Acetate, 10 mM NaCl, pH 7.0 and 22 °C) reveals that the peptide, at 10 μM concentration, binds specifically the three six base pair sites 5′-TGCGCA-3′, 5′-AGCGCT-3′ and 5′-AACGCA-3′ (Figures 5 and 6, 5'- data not shown).[3] The apparent first order binding affinities of ImPImP for these three sites were determined by quantitative DNase I footprint titration experiments (Table I).[14]

Affinity Cleaving. Cleavage experiments on the 381-base pair restriction fragment (25 mM Tris•Acetate, 10 mM NaCl, pH 7.0 and 22 °C) reveal cleavage patterns which are shifted to the 3'-side of the binding site, consistent with binding in the minor groove (Figures 5 and 6).[2] ImPImPE, at 10 μM concentration, clearly binds the 5'-TGCGCA-3' site in two equal orientations. The

Figure 5. MPE•Fe(II) footprinting of ImPImP and specific cleavage by ImPImPE. Autoradiogram of a 6% denaturing polyacrylamide gel. All reactions contain 25 mM tris acetate (pH 7.0), 10 mM NaCl, 5 mM DTT, 100 μM bp calf thymus DNA, and 3'-^{32}P end labeled 381 bp restriction fragment. Lane 1, intact DNA; lane 2, A reaction; lane 3, Maxam-Gilbert G reaction; lane 4, MPE•Fe(II) standard; lane 5, 50 μM ImPImP; lane 6, 20 μM ImPImP; lane 7, 10 μM ImPImP; lane 8, 50 μM ImPImPE•Fe(II); lane 9, 20 μM ImPImPE•Fe(II); lane 10, 10 μM ImPImPE•Fe(II). Lanes 4-7 contain 5 μM MPE•Fe(II).


Control
A
G
MPE
ImPImP
ImPImPE
1 2 3 4 5 6 7 8 9 10
3' 5'
TA
AT
CG
GC
CG
GC
AT
GC
5' 3'
3' 5'
GC
AT
TA
CG
GC
CG
GC
AT
TA
AT
TA
AT
CG
GC
CG
AT
AT
CG
TA
AT
TA
AT
AT
AT
GC
AT
TA
AT
CG
GC
CG
GC
TA
GC
GC
5' 3'
32P

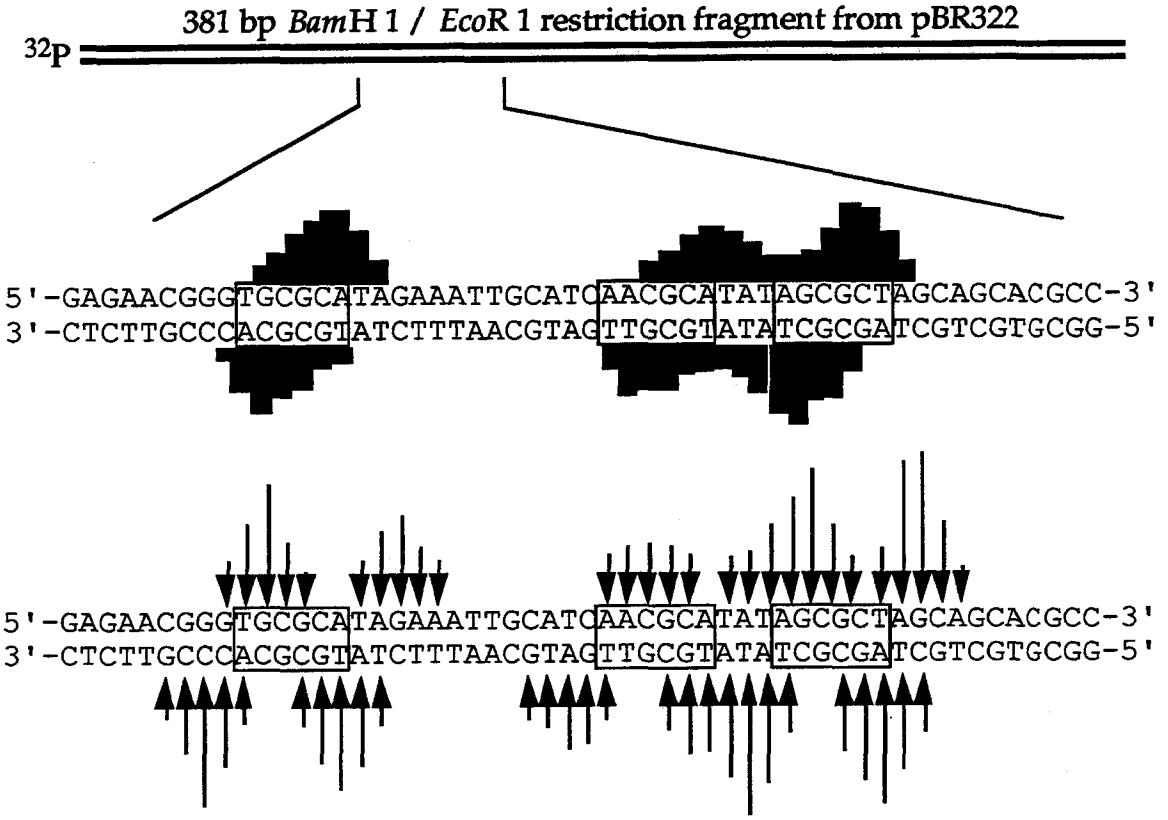


Figure 6. Histograms of cleavage protection (footprinting) and affinity cleaving data. (Top) Illustration of the 381 bp restriction fragment with the position of the sequence indicated. (Middle) MPE•Fe(II) protection pattern for ImPImP at 20 μM concentration. Bar heights are proportional to the relative protection from cleavage at each band. Boxes represent equilibrium binding sites determined by the published model.[2] (Bottom) Cleavage of the 381 bp restriction fragment by ImPImPE•Fe(II) at 20 μM concentration. Arrows are proportional to the integrated densities of the cleavage bands. Data for the bottom strand are shown in Figure 5.

cleavage pattern at the 5'-AACGCA-3' and 5'-AGCGCT-3' sites are also composed of equal cleavage intensities at both ends of the sites, with a slight overlap of cleavage patterns between the two sites. In addition the affinity cleaving experiments identify a 5'-AGCGCA-3' binding site near the top of the gel (Figure 5).

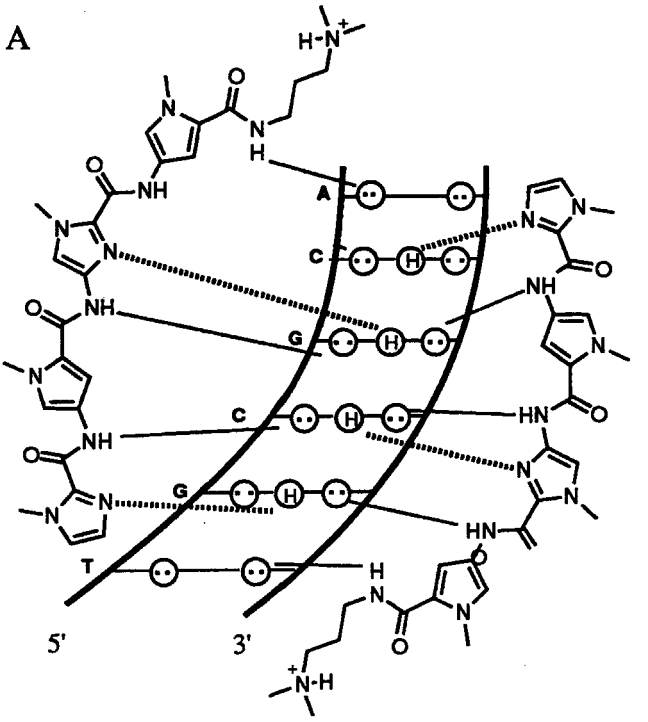
Discussion

Recognition of 5′-GCGC-3′ Sites. This is the first report of a designed peptide analog which sequence-specifically binds a *designated 5′-GCGC-3′ core sequence.* Footprinting and affinity cleaving experiments clearly demonstrate sequence-specific binding to the four six base pair sites 5'-AGCGCT-3', 5'-TGCGCA-3', 5'-AGCGCA-3' and 5'-AACGCA-3' by the peptide ImPImP. Affinity cleaving experiments with ImPImPE are consistent with the peptide binding as an antiparallel side-by-side dimer in the minor groove at these sequences. In collaboration with the Wemmer group, we have characterized directly by two-dimensional NMR the complex $(ImPImP)_2$•5′-TGCGCA-3′.[15] The peptide binds as a side-by-side antiparallel dimer with positive cooperativity in the minor groove. Energy minimization of the complex with constraints from NOESY experiments affords a model wherein specific hydrogen bonds are formed between the four imidazole nitrogens of the peptides and the four guanine amino group on the floor of the minor groove (Figure 7).

Sequence-Specificity. The complex formed upon binding of ImPImP to the 5'-AACGCA-3' site contains a single hydrogen bond mismatch between the peptides and the floor of the minor groove (Figure 7). The energetic penalty of ~ 0.5 kcal/mol for this mismatch is identical to those measured for binding of the covalent peptide heterodimer 2-ImN-C_4-P3 to 5'-TGTCA-3' and binding of 2-ImN to the 5'-TGTTA-3' site.[16c] The preference for A,T base pairs in the first and sixth positions of the binding site may be due to favorable hydrogen bonds between the bases and the terminal pyrrolecarboxamides of the peptides. Alternatively, GC base pairs may be disfavored due to a lack of hydrogen bond acceptors on

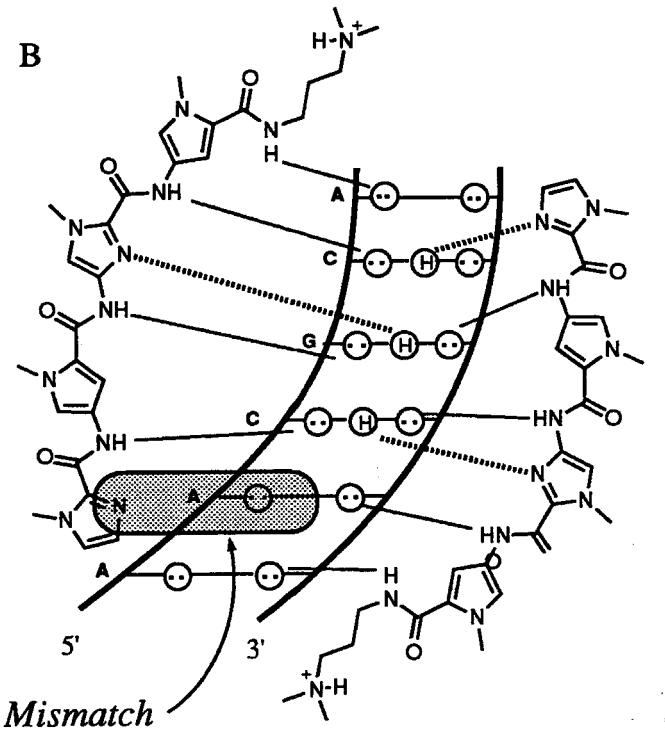
Figure 7. (A) Model for the complex formed between ImPImP with the 5'-TGCGCA-3' site. (B) Model for the complex formed between ImPImP with the 5'-AACGCA-3' site. The absence of a hydrogen bond donor on the floor of the minor groove for interaction with the imidazole N3 atom of the peptide results in a single hydrogen bond mismatch (illustrated with shading).

A



$(ImPImP)_2 \cdot TGCGCA$

B



$(ImPImP)_2 \cdot AACGCA$

the peptides. In either case, the peptides appear not to discriminate between AT and TA base pairs. In contrast, the peptides do discriminate between GC and CG base pairs, as evidenced by a lack of binding to the 5'-TGGCGT-3' and 5'-TCGGCA-3' sites on this restriction fragment. With regard to mismatches in 2:1 peptide-DNA complexes, substitution of a GC bp by an AT bp introduces a single mismatch interaction. However, substitution of a GC bp with a CG bp may introduce two mismatch interactions, one for each of the side-by-side peptides. These observations may provide guidelines for predicting relative affinities of side-by-side peptides for DNA sites.

The 2:1 motif has guided the successful design of peptides for recognition of three very different sequences: (i) 5'-TGTCA-3' by the covalent peptide homodimer (2-PyN)$_2$-C4;[16ab] (ii) 5'-TGTTA-3' by the covalent peptide heterodimer 2-ImN C_4-D;[16c] and (iii) 5'-TGCGCA-3' by the ImPImP homodimer. Binding of ImPImP to a pure four base pair G,C-core sequence represents an absolute reversal of the specificity of the natural products distamycin and netropsin. We anticipate that many other sequences will be targeted by matching pyrrole- and imidazole-carboxamides of designed peptides to hydrogen bond acceptors and donors on the floor of the minor groove. Covalently linking these peptides may provide a new family of ligands for sequence-specific recognition of designated DNA sites.[16] Finally, ImPImP is the first example of a side-by-side dimer binding a *six base pair site*. Future work will define the limitations in the size of the binding site which can accommodate side-by-side peptides in the minor groove of DNA.

Experimental Section

^1H NMR and ^{13}C NMR spectra were recorded on a General Electric-QE 300 NMR spectrometer in $CDCl_3$ or CD_3OD, with chemical shifts reported in parts per million relative to tetramethyl silane or residual CHD_2OD, respectively. 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. Reactions were executed under an inert argon atmosphere. Reagent grade chemicals were used as received unless otherwise noted. Tetrahydrofuran (THF) was distilled under nitrogen from sodium/benzophenone ketyl. Dichloromethane (CH_2Cl_2) and triethylamine were distilled under nitrogen from powdered calcium hydride. Dimethylformamide (DMF) was purchased as an anhydrous solvent from Aldrich. Flash chromatography was carried out using EM science Kieselgel 60 (230-400) mesh.[17] Thin-layer chromatography was performed on EM Reagents silica gel plates (0.5 mm thickness). All compounds were visualized with short-wave ultraviolet light. Reverse phase chromatography was performed on a Pharmacia FPLC system with a ProRPC 10/10 column and a solvent system of acetonitrile/100 mM triethylammonium acetate (0-40%, 40 min linear gradient). The desired fractions were pooled and concentrated *in vacuo*. The residue was desalted by dissolving in 5% ammonium hydroxide in methanol, concentrating and passing through a short column of silica gel (1% ammonium hydroxide in methanol).

***N*-methyl-4-(*N*-methyl-4-nitroimidazole-2-carboxamide)-2-(carboxamidopropyl-3-dimethylamino)-pyrrole (4a).** To a slurry of 4-nitro-1-methylimidazole-2-carboxylic acid (300 mg, 1.75 mmol) in tetrahydrofuran (4.0 mL) was added oxalyl chloride (2.0 ml, 22.9 mmol) and the resulting solution

was heated at reflux for 1.5 h. The solution was cooled, excess solvent was removed *in vacuo* and the residue was dissolved in dimethylformamide (5.0 ml). *N*-Methyl-4-amino-2-(carboxamidopropyl)-3-(dimethylamino)-pyrrole (0.400 g, 1.79 mmol) was added and the reaction mixture was allowed to stir at room temperature for 4 h. Solvent was removed *in vacuo* and the residue was purified by flash column chromatography (MeOH) to afford **4a**: yield 63% (414 mg): ^{1}H NMR (CD_3OD) δ 8.21 (s, 1 H), 7.27 (d, 1 H, *J* = 1.8 Hz), 6.82 (d, 1 H, *J* = 1.9 Hz), 4.12 (s, 3 H), 3.86 (s, 3 H), 3.22 (m, 2 H), 2.45 (t, 2 H, *J* = 7.6 Hz), 2.30 (s, 6 H), 1.78 (qn, 2 H, *J* = 7.5 Hz); IR (thin film) 3385 (m), 2950 (w), 1644 (m), 1634 (m), 1538 (s), 1463 (m), 1454 (m), 1386 (w), 1308 (m); FABMS *m/e* 378.1890 (M + H, 378.1879 calcd. for $C_{16}H_{23}N_7O_4$).

***N*-methyl-4-(*N*-methyl-4-(*N*-methyl-4-nitropyrrole-2-carboxamide)-imidazole-2-carboxamide)-2-(carboxamidopropyl-3-dimethylamino)-pyrrole (5a).** To a flask equipped with a reflux condensor was added 4-nitro-1-methylpyrrole-2-carboxylic acid (105 mg, 0.618 mmol) and thionyl chloride (5.0 ml 68.5 mmol) and the mixture was heated at 85 °C for 3 hr. The solution was allowed to cool and excess reagent was removed *in vacuo*, and the residue was dissolved in dimethylformamide (3.0 ml) and cooled to 0 °C. Separately, to a solution of **4a** (90 mg, 0.239 mmol) in dimethylformamide (3.5 ml) was added Pd/C catalyst (10%, 35 mg), and the mixture was hydrogenated in a Parr bom apparatus (300 psi H_2) for 2 hr. The catalyst was removed by filtration through celite, and the filtrate was immediately added to the acid chloride and allowed to warm to room temperature and stir 8 hr. Methanol (1.0 ml) was added and all solvents were removed *in vacuo*. Flash column chromatography of the residue (methanol) was followed by purification by FPLC to afford **5a** as a yellow powder (46 mg, 39%): ^{1}H NMR (CD_3OD) δ 7.82 (d, 1 H, *J* = 1.8 Hz), 7.41 (d, 1 H, *J* = 1.9 Hz), 7.37 (s, 1 H), 7.17 (d, 1 H, *J* = 1.9 Hz), 6.77 (d, 1 H, *J* = 1.9 Hz), 4.00 (s, 3

H), 3.98 (s, 3 H), 3.85 (s, 1 H), 3.34 (m, 2 H), 2.47 (t, 2 H, *J* = 7.7 Hz), 2.32 (s, 6 H=), 1.79 (q, 2 H, *J* = 7.5 Hz); ^{13}C NMR (CD_3OD) δ 164.2, 159.3, 157.7, 137.4, 136.2, 135.6, 129.1, 127.0, 124.8, 122.5, 120.2, 115.8, 109.2, 105.6, 58.1, 45.2, 38.1, 36.0, 28.1; IR (thin film) 3405 (m), 2949 (w), 1654 (m), 1637 (m), 1546 (s), 1499 (m), 1466 (m), 1438 (m), 1311 (m), 1120 (w); FABMS *m/e* 500.2370 (M + H, 500.2357 calcd. for $C_{22}H_{30}N_9O_5$).

***N*-methyl-4-(*N*-methyl-4-(*N*-methyl-4-(*N*-methylimidazole-2-carboxamide)-pyrrole-2-carboxamide)-imidazole-2-carboxamide)-2-(carboxamidopropyl-3-dimethylamino)-pyrrole (1).** To a solution of 1-methylimidazole-2-carboxylic acid (33 mg, 0.297 mmol) and *N*-hydroxybenzotriazole hydrate (41 mg, 0.30 mmol) in dimethylformamide (0.60 ml) was added a solution of 1,3-dicyclohexylcarbodiimide (62 mg, 0.30 mmol) in methylene chloride (0.60 ml) and the mixture was allowed to stir for 90 min. Separately, to a solution of **5a** (20 mg, 0.040 mmol) in dimethylformamide (2.5 ml) was added Pd/C catalyst (10%, 10 mg) and the mixture was hydrogenated in a Parr bom apparatus (300 psi H_2) for 90 min. The catalyst was removed by filtration through celite and the filtrate was immediately added to the activated acid and allowed to stir 2.5 hr. Methanol (1.0 ml) was added and all solvents were removed *in vacuo*. Flash column chromatography of the residue (0.5% ammonium hydroxide in methanol) was followed by purification by FPLC to afford **1** as a yellow powder (10 mg, 43 %): ^1H NMR (CD_3OD) δ 7.43 (s, 1 H), 7.35 (d, 1 H, *J* = 1.8 Hz), 7.25 (d, 1 H, *J* = 1.1 Hz), 7.23 (d, 1 H, *J* = 1.8 Hz), 7.05 (d, 1 H, *J* = 1.1 Hz), 7.01 (d, 1 H, *J* = 1.9 Hz), 6.82 (d, 1 H, *J* = 1.9 Hz), 4.06 (s, 3 H), 4.05 (s, 3 H), 3.94 (s, 3 H), 3.88 (s, 3 H), 3.31 (m, 2 H), 2.48 (t, 2 H, *J* = 7.7 Hz), 2.32 (s, 6 H), 1.79 (q, 2 H, *J* = 7.6 Hz); IR (thin film) 3285 (m), 2956 (w), 1654 (m), 1542 (s), 1466 (m), 1426 (m), 1406 (m), 1250 (w), 1123 (w); UV (H_20) λ_{max} (ε) 252 (21 500), 312 (35 500) nm; FABMS *m/e* 578.2952 (M + H, 578.2985 calcd. for $C_{27}H_{36}N_{11}O_4$).

***N*-methyl-4-nitro-2-(carboxamido-(3-*tert*-butoxycarbonylamino-3'-amino-*N*-methyldipropylamino)-pyrrole (3b).** To a flask equipped with a reflux condensor was added 4-nitro-1-methylpyrrole-2-carboxylic acid (2.00 g, 11.8 mmol) and thionyl chloride (10.0 ml, 137 mmol) and the mixture was heated at 85 °C for 3 hr. The solution was allowed to cool and excess reagant was removed *in vacuo*. The acid chloride was dissolved in methylene chloride (10.0 ml), cooled to 0 °C and a solution of 3-(*tert*-butyloxycarbonylamino)-3'-amino-*N*-methyldipropylamine (2.60 g, 10.61 mmol) in methylene chloride (10.0 ml) was added. The resulting solution was allowed to stir 12 hr followed by addition of methanol (5.0 ml) followed by stirring 30 min. Solvents were removed *in vacuo* and the residue was purified by flash column chromatography to afford the protected amine in 93% yield (3.90 g): ^{1}H NMR ($CDCl_3$) δ 7.43 (s, 1 H), ; IR (thin film) 3317 (m), 2975 (m), 1698 (s), 1652 (s), 1530 (s), 1505 (m), 1418 (m), 1312 (s), 1272 (m), 1169 (m); FABMS *m/e* 398.2403 (M + H, 398.2411 calcd. for $C_{18}H_{32}N_5O_5$).

***N*-methyl-4-(*N*-methyl-4-nitroimidazole-2-carboxamide)-2-(carboxamido-(3-*tert*-butoxycarbonylamino-3'-amino-*N*-methyldipropylamino)-pyrrole (4b).** To a slurry of 4-nitro-1-methylimidazole-2-carboxylic acid (400 mg, 2.33 mmol) in tetrahydrofuran (5.0 mL) was added oxalyl chloride (2.5 ml, 28.7 mmol) and the resulting solution was heated at reflux for 2 h. The solution was cooled, excess solvent was removed *in vacuo* and the residue was dissolved in dimethylformamide (5.0 ml) and cooled to 0 °C. Separately, to a solution of nitropyrrole X (1.00 g, 2.52 mmol) in dimethylformamide (10.0 ml) was added Pd/C catalyst (10%, 450 mg) and the mixture was hydrogenated in a Parr bom apparatus (500 psi H_2) for 6 hr. The catalyst was removed by filtration through celite and the filtrate was immediately added to the acid chloride and allowed to warm to room

temperature and stir 24 hr. Following addition of methanol (1.0 ml), solvent was removed *in vacuo* and the residue was purified by flash column chromatography (15% methanol in methylene chloride) to afford **2** (260 mg, 20 %): ^{1}H NMR (CD_3OD) δ 8.17 (s, 1 H), 7.23 (d, 1 H, *J* = 1.8 Hz), 6.83 (d, 1 H, *J* = 1.8 Hz), 4.10 (s, 3 H), 3.85 (s, 3 H), 3.35 (q, 2 H, *J* = 6.0 Hz), 3.10 (t, 2 H, *J* = 6.6 Hz), 2.80 (m, 4 H), 2.55 (s, 3 H), 1.89 (m, 2 H), 1.79 (m, 2 H), 1.40 (s, 9 H); ^{13}C NMR (CD_3OD) δ 164.1, 158.5, 156.1, 146.1, 138.6, 126.6, 124.3, 122.2, 120.3, 105.6, 80.0, 55.8, 55.7, 41.3, 38.9, 37.8, 37.3, 37.0, 28.7, 27.1, 26.8; IR (thin film) 3333 (m), 2966 (w), 1682 (s), 1638 (m), 1540, (m), 1456 (m), 1387 (w), 1308 (m), 1168 (w); FABMS *m/e* 521.2836 (M + H, 521.2830 calcd. for $C_{23}H_{37}N_8O_6$).

***N*-methyl-4-(*N*-methyl-4-(*N*-methyl-4-nitropyrrole-2-carboxamide)-imidazole-2-carboxamide)-2-(carboxamido-(3-*tert*-butoxycarbonylamino-3'-amino-*N*-methyldipropylamino)-pyrrole (5b).** To a flask equipped with a reflux condensor was added 4-nitro-1-methylpyrrole-2-carboxylic acid (170 mg, 1.18 mmol) and thionyl chloride (5.0 ml 68.5 mmol) and the mixture was heated at 85 °C for 2 hr. The solution was allowed to cool and excess reagant was removed *in vacuo*. and the residue was dissolved in dimethylformamide (2.0 ml) and cooled to 0 °C. Separately, to a solution of PyImPy (180 mg, 0.346 mmol) in dimethylformamide (5.0 ml) was added Pd/C catalyst (10%, 90 mg) and the mixture was hydrogenated in a Parr bom apparatus (450 psi H_2) for 3 hr. The catalyst was removed by filtration through celite and the filtrate was immediately added to the acid chloride and allowed to warm to room temperature and to stir 24 hr. Methanol (1.0 ml) was added and all solvents were removed *in vacuo*. Flash column chromatography of the residue (methanol) provided PyImPyBoc as a yellow powder (105 mg, 47%): ^{1}H NMR (CD_3OD) δ 7.80 (d, 1 H, *J* = 1.9 Hz), 7.40 (d, 1 H, *J* = 1.9 Hz), 7.35 (s, 1 H), 7.15 (d, 1 H, *J* = 1.9 Hz), 6.76 (d, 1 H, *J* = 1.9 Hz), 3.99 (s, 3 H), 3.96 (s, 3 H), 3.84 (s, 3 H), 3.31 (m, 2 H),

3.05 (t, 2 H, J = 6.8 Hz), 2.42 (m, 4 H), 2.23 (s, 3 H), 1.75 (m, 2 H), 1.65 (m, 2 H), 1.39 (s, 9 H); ^{13}C NMR (CD_3OD) δ 163.9, 159.1, 158.4, 157.5, 137.3, 136.1, 135.5, 129.0, 126.9, 124.8, 122.5, 120.0, 115.7, 109.2, 105.5, 79.9, 56.5, 56.2, 42.3, 39.7, 38.8, 38.2, 36.9, 36.0, 28.8, 28.2, 27.8; IR (thin film) 3329 (m), 2941 (w), 1654 (s), 1540, (s), 1466 (m), 1366 (w), 1314 (m), 1166 (w), 1116 (w); FABMS *m/e* 643.3316 (M + H, 643.3308 calcd. for $C_{29}H_{43}N_{10}O_7$).

***N*-methyl-4-(*N*-methyl-4-(*N*-methyl-4-(*N*-methylimidazole-2-carboxamide)-pyrrole-2-carboxamide)-imidazole-2-carboxamide)-2-(carboxamido-(3-*tert*-butoxycarbonylamino-3'-amino-*N*-methyldipropylamino)-pyrrole (6b).** To a solution of 1-methylimidazole-2-carboxylic acid (46 mg, 0.414 mmol) and *N*-hydroxybenzotriazole hydrate (57 mg, 0.42 mmol) in dimethylformamide (0.84 ml) was added a solution of 1,3-dicyclohexylcarbodiimide (87 mg, 0.42 mmol) in methylene chloride (0.84 ml) and the mixture was allowed to stir for 1 hr. Separately, to a solution of **5b** (50 mg, 0.078 mmol) in dimethylformamide (2.0 ml) was added Pd/C catalyst (10%, 23 mg) and the mixture was hydrogenated in a Parr bom apparatus (350 psi H_2) for 2 hr. The catalyst was removed by filtration through celite and the filtrate was immediately added to the activated acid and allowed to stir 3 hr. Methanol (1.0 ml) was added and all solvents were removed *in vacuo*. Flash column chromatography of the residue (methanol) was followed by purification by FPLC to afford **6** as a yellow powder (18 mg, 32 %): ^1H NMR (CD_3OD) δ 7.43 (s, 1 H), 7.35 (d, 1 H, J = 1.8 Hz), 7.24 (d, 1 H, J = 1.0 Hz), 7.23 (d, 1 H, J = 1.9 Hz), 7.05 (d, 1 H, J = 1.0 Hz), 7.00 (d, 1 H, J = 1.9 Hz), 6.86 (d, 1 H, J = 1.9 Hz), 4.06 (s, 3 H), 4.05 (s, 1 H), 3.94 (s, 3 H), 3.88 (s, 3 H), 3.35 (m, 2 H), 3.12 (t, 2 H, J = 6.8 Hz), 2.76 (m, 4 H), 2.52 (s, 3 H), 1.80 (m, 2 H), 1.75 (m, 2 H), 1.41 (s, 9 H); IR (thin film) 3300 (m), 2966 (w), 1655 (s), 1578 (m), 1542 (s), 1466 (m), 1406 (w), 1250 (w), 1168 (w), 1123 (w); FABMS *m/e* 721.3898 (M + H, 721.3901 calcd. for $C_{34}H_{49}N_{12}O_6$).

***N*-methyl-4-(*N*-methyl-4-(*N*-methyl-4-(*N*-methylimidazole-2-carboxamide)-pyrrole-2-carboxamide)-imidazole-2-carboxamide)-2-(carboxamido-(3-amino-3'-amino-*N*-methyldipropylamino)-pyrrole (7).** To a solution of protected amine **6** (11 mg, 0.015 mmol) in methylene chloride (2.0 ml) was added trifluoroacetic acid (0.5 ml) and the resulting mixture was allowed to stir 20 min. The ammonium salt was precipitated by the addition of hexanes (40 ml), dissolved in 1% ammonium hydroxide in methanol (50 ml) and solvent was removed *in vacuo*. The product was purified by flash column chromatography (6% ammonium hydroxide in methanol) to afford the primary amine **7** (8 mg, 86%): ^{1}H NMR (CD_3OD) δ 7.43 (s, 1 H), 7.35 (d, 1 H, J = 1.9 Hz), 7.25 (d, 1 H, J = 1.0 Hz), 7.22 (d, 1 H, J = 1.9 Hz), 7.05 (d, 1 H, J = 1.0 Hz), 7.00 (d, 1 H, J = 1.9 Hz), 6.83 (d, 1 H, J = 1.9 Hz), 4.06 (s, 3 H), 4.05 (s, 1 H), 3.94 (s, 3 H), 3.88 (s, 3 H), 3.34 (m, 2 H), 2.75 (t, 2 H, J = 7.0 Hz), 2.47 (m, 4 H), 2.26 (s, 3 H), 1.82-1.67 (m, 4 H); IR (thin film) 3333 (m), 2954 (w), 1653 (m), 1541 (s), 1472 (m), 1431 (m), 1254 (w), 1208 (w), 1123 (w); FABMS *m/e* 621.3374 (M + H, 621.3386 calcd. for $C_{29}H_{41}N_{12}O_4$).

ImPImPEDTA(OEt)$_3$ (8). To a solution of EDTA triethyl ester (165 mg, 0.439 mmol) and *N*-hydroxysuccinimide (53 mg, 0.461 mmol) in 1,4-dioxane (2.0 ml) was added a solution of 1,3-dicyclohexylcarbodiimide (96 mg, 0.465 mmol) in methylene chloride (0.93 ml), and the solution was allowed to stir 90 min. ImPyImPyAmine (7.0 mg, 0.0113 mmol) in dimethylformamide (1.5 ml) was added and allowed to stir 8 hr. Methanol (0.5 ml) was added, the mixture was filtered through celite and solvent was removed *in vacuo*. Flash column chromatography of the oil (gradient 0 to 1% ammonium hydroxide in methanol) afforded the triester **8** (7.0 mg, 63%): ^{1}H NMR (CD_3OD) δ 7.43 (s, 1 H), 7.35 (d, 1 H, J = 1.9 Hz), 7.25 (d, 1 H, J = 1.1 Hz), 7.23 (d, 1 H, J = 1.8 Hz), 7.05 (d, 1 H, J = 1.1 Hz), 7.02 (d, 1 H, J = 1.9 Hz), 6.86 (d, 1 H, J = 1.9 Hz), 4.15-4.08 (m, 6 H), 4.06 (s, 3

H), 4.05 (s, 3 H), 3.94 (s, 3 H), 3.88 (s, 3 H), 3.62 (s, 2 H), 3.59 (s, 4 H), 3.51 (s, 2 H), 3.30 (m, 4 H), 2.83-2.70 (m, 8 H), 2.49 (s, 3 H), 1.91-1.83 (m, 4 H), 1.26-1.19 (m, 9 H); IR (thin film) 3385 (m), 2951 (w), 1734 (s), 1654 (s), 1591 (m), 1542 (s), 1458 (m), 1420 (m), 1209 (s), 1125 (w); FABMS *m/e* 979.5114 (M + H, 979.5081 calcd. for $C_{45}H_{67}N_{14}O_{11}$).

ImPImPE (2). To a solution of triethyl ester **8** (6.0 mg, 0.0061 mmol) in ethanol (0.7 ml) was added 0.5 M LiOH (0.7 ml) and the solution was allowed to stir 15 hr. 1N HCl (3 ml) was added and solvent was removed in vacuo. Purification by FPLC afforded ImPyImPyE (2.0 mg, 37%): ^{1}H NMR (CD_3OD) δ 7.44 (s, 1 H), 7.36 (d, 1 H, *J* = 1.9 Hz), 7.28 (d, 1 H, *J* = 1.9 Hz), 7.24 (d, 1 H, *J* = 1.0 Hz), 7.06 (d, 1 H, *J* = 1.9 Hz), 7.05 (d, 1 H, *J* = 1.1 Hz), 6.94 (d, 1 H, *J* = 1.9 Hz), 4.06 (s, 3 H), 4.05 (s, 3 H), 3.95 (s, 3 H), 3.89 (s, 3 H), 3.84 (s, 4 H), 3.30 (m, 4 H), 2.85 (s, X H), 2.15-1.96 (m, 4 H), 1.95 (s, 3 H); IR (thin film) 3385 (m), 2956 (w), 1644 (s), 1591 (m), 1540 (s), 1469 (m), 1404 (m), 1251 (w), 1208 (s), 1123 (w); FABMS *m/e* 895.4175 (M + H, 895.4193 calcd. for $C_{39}H_{55}N_{14}O_{11}$).

DNA Reagents and Materials. Doubly distilled water was further purified through the Milli Q filtration system from Millipore. Sonicated, deproteinized calf thymus DNA was purchased from Pharmacia. Plasmid pBR322 was obtained from Boehringer-Mannheim. Enzymes were obtained from Boehringer-Mannheim or New England Biolabs and used with the buffers supplied. Deoxyadenosine 5'-[α-^{32}P]triphosphate and adenosine 5'-[γ-^{32}P] triphosphate were obtained from Amersham. Storage phosphor technology autoradiography was performed using a Molecular Dynamics 400S Phosphorimager and ImageQuant software. The 381 base pair 3'- and 5'-end labeled *Eco*R I / *Bam*H I restriction fragment from plasmid pBR322 was prepared and purified as previously described.[3b] Chemical sequencing reactions were performed according to published methods.[18,19] Standard techniques were

employed for DNA manipulations.[20] All other reagents and materials were used as received.

MPE•Fe(II) Footprinting and Affinity Cleaving.[3] A 50 μM MPE•Fe(II) solution was prepared by mixing 100 μL of a 100 μM MPE solution with 100 μL of a freshly prepared 100 μM ferrous ammonium sulfate solution. A 100 μM ImPImP•Fe(II) solution was prepared by mixing 10 μL of a 1 mM ImPImP solution with 10 μL of a 1 mM ferrous ammonium sulfate solution and diluting to 100 mL. Solutions were prepared containing 1μL/tube 20x Tris acetate, pH 7.0 buffer, 2 μL/tube 1 mM-bp calf thymus DNA, 1μL/tube 200 mM sodium chloride, labeled restriction fragment and water to make 14 μL/tube total solution. 2 μL of a 10x solution of the peptide was added and the tubes were incubated for 20 min at 22 °C. To the footprinting reactions was added 2 μL of the 50 μM MPE•Fe(II) solution and incubated 5 min. Cleavage was initiated by the addition of 2 μL of a freshly prepared 50 mM DTT solution. Final concentrations were 25 mM Tris•acetate (pH 7.0), 10 mM sodium chloride, 100 μM-bp DNA, 5 μM MPE•Fe(II) and 5 mM DTT, in 20 μL volume. The reactions were incubated at 22 °C for 25 min, ethanol precipitated, resuspended in 100 mM tris-borate-EDTA / 80% formamideloading buffer and electrophoresed on 6% denaturing polyacrylamide gels (5% crosslink, 7 M urea) at 1500 V for 3-4 h. The gels were dried and quantitated using storage phosphor technology.

Quantitative DNase I Footprint Titration Experiments. Apparent first order binding constants were determined by DNase footprint titrations as previously described.[11c,16a] The ligands (200 mM to 100 nM) were added to solutions of radiolabeled restriction fragment (20,000 cpm), Tris acetate (10 mM, pH 7.0), KCl (10 mM), $MgCl_2$ (10 mM) and $CaCl_2$ (5 mM) and incubated for 30 min at 22 °C. Footprinting reactions were initiated by the addition of 1 μL a stock solution of DNase I (10 units/mL) containing 1 mM dithiothreitol and allowed to

proceed for 3 min at 22 °C. The reactions were stopped by addition of a 3 M ammonium acetate solution containing 250 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, 7 M urea) at 1000 V for 3-4 hr.

The data were analyzed by performing volume integrations of the 5′-TGCGCA-3′, 5'-AGCGCT-3' and 5′-AACGCA-3′ target sites and a 5′-GCGG-3′ reference site. The apparent DNA target site saturation, θ_{app}, was calculated for each concentration of peptide using the following equation:

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

where I_{tot} and I_{ref} are the integrated volumes of the target and reference sites, respectively, and I_{tot}° and I_{ref}° correspond to those values for a DNase I control lane to which no peptide has been added. At higher concentrations of peptide (> 50 μM), the reference sites become partially protected, resulting in low θ_{app} values. For these data points, the reference value was determined from the amount of radioactivity loaded per lane, using the mean value for all data points from lanes with < 50 μM peptide. The ($[L]_{tot}$, θ_{app}) data points were fit to a general Hill model by minimizing the difference between θ_{app} and θ_{fit}, using the modified Hill equation:

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

where $[L]_{tot}$ corresponds to the total peptide concentration, K_a corresponds to the apparent monomeric association constant, and θ_{min} and θ_{max} represent the experimentally determined site saturation values when the site is unoccupied or saturated, respectively. Data were fit using a nonlinear least-squares fitting procedure of KaleidaGraph software (version 2.1, Abelbeck software) running on a Macintosh IIfx computer with K_a, θ_{max}, θ_{min} and n as the adjustable parameters. Consistent with cooperative dimeric binding by the peptides, n was in the range 1.8-2.0. We note explicitly that treatment of the data in this manner does not represent an attempt to model a binding mechanism. Rather, we have chosen to compare values of the apparent first order dissociation constant, because this paramter represents the concentration of peptide at which the binding site is half-saturated.

Quantitation by Storage Phosphor Technology Autoradiography. Photostimulable storage phosphor imaging plates (Kodak Storage Phosphor Screen S0230 obained from Molecular Dynamics) were pressed flat against gel samples and exposed in the dark at 22 °C for 15-20 h. 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 the ImageQuant v. 3.0 software running on an AST Premium 386/33 computer.

References

1. (a) Krylov, A. S.; Grokhovsky, S. L.; Zasedatelev, A. S.; Zhuze, A. L.; Gursky, G. V.; Gottikh, B. P. *Nuc. Acids. Res.* **1979**, *6*, 289-304. (b) Zasedatelev, A. S.; Gursky, G. V.; Zimmer, Ch.; Thrum, H. *Mol. Biol. Reports* **1974**, *1*, 337-342. (c) Zasedatelev, A. S.; Zhuze, A. L.; Zimmer, Ch.; Grokhovsky, S. L.; Tumanyan, V. G.; Gursky, G. V.; Gottikh, B. P. *Dokl. Acad. Nauk SSSR* **1976**, *231*, 1006-1009. For a review, see: (d) Zimmer, C.; Wåhnert, U. *Prog. Biophys. Molec. Biol.* **1986**, *47*, 31-112.

2. (a) Van Dyke, M. W.; Hertzberg, R. P.; Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 5470-5474. (b) Van Dyke, M. W.; Dervan, P. B. *Cold Spring Harbor Symposium on Quantitative Biology* **1982**, *47*, 347-353. (c) Van Dyke, M. W.; Dervan, P. B. *Biochemistry* **1983**, *22*, 2373-2377. (d) Harshman, K. D.; Dervan, P. B. *Nucl. Acids Res.* **1985**, *13*, 4825-4835. (e) Fox, K. R.; Waring, M. J. *Nucl. Acids Res.* **1984**, *12*, 9271-9285. (f) Lane, M. J.; Dabrowiak, J. C.; Vournakis, J. *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 3260-3264.

3. (a) Schultz, P. G.; Taylor, J. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1982**, *104*, 6861-6863. (b) Taylor, J. S.; Schultz, P. G.; Dervan, P. B. *Tetrahedron* **1984**, *40*, 457-465. (c) Schultz, P. G.; Dervan, P. B. *J. Biomol. Struct. Dyn.* **1984**, *1*, 1133-1147. (d) Dervan, P. B. *Science*, **1986**, *232*, 464-471.

4. (a) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 1376-1380. (b) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *J. Mol. Biol.* **1985**, *183*, 553-563. (c) Coll, M.; Frederick, C. A.; Wang, A. H.-J.; Rich, A. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 8385-8389.

5. (a) Patel, D. J.; Shapiro, L. *J. Biol. Chem.* **1986**, *261*, 1230-1240. (b) Klevitt, R. E.; Wemmer, D. E.; Reid, B. R. *Biochemistry* **1986**, *25*, 3296-3303. (c) Pelton, J. G., Wemmer, D. E. *Biochemistry* **1988**, *27*, 8088-8096.

6. (a) Markey, L. A.; Breslauer, K. J. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 4359-4363. (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*, 8922-8926.

7. Dervan, P. B. *Science* **1986**, *232*, 464-471.

8. For early examples of hybrid molecules for the recognition of mixed sequences, see: (a) Dervan, P. B.; Sluka, J. P. *New Synthetic Methodology and Functionally Interesting Compounds*; Elsevier: New York, 1986; pp 307-322. (b) Griffin, J. H.; Dervan, P. B. *J. Am. Chem. Soc.* **1987**, *109*, 6840-6842.

9. (a) Wade, W. S., Dervan, P. B. *J. Am. Chem. Soc.* **1987**, *109*, 1574-1575. (b) Wade, W. S. Ph.D. Thesis, California Institute of Technology, 1989.

10. (a) Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 5723-5727. (b) Pelton, J. G.; Wemmer, D. E. *J. Am. Chem. Soc.* **1990**, *112*, 1393-1399.

11. (a) Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 8783-8794. (b) Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Dervan, P. B. *Proc. Natl. Acad. Sci., USA* **1992**, *89*, 7586-7590. (c) Wade, W. S.; Mrksich, M.; Dervan, P. B. *Biochemistry* **1993**, *32*, 11385-11389.

12. (a) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 2572-2576. (b) Geierstanger, B. H.; Jacobsen, J-P.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Biochemistry*, in press.

13. Geierstanger, B. H.; Dwyer, T. J.; Bathini, Y.; Lown, J. W.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, *115*, 4474-4482.

14. (a) Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Methods Enzymol.* **1986**, *130*, 132-181. (b) Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 8462-8466. (c) Senear, D. F.; Brenowitz, M.; Shea, M. A.; Ackers, G. K. *Biochemistry* **1986**, *25*, 7344-7354.

15. Geierstanger, B. H.; Mrksich, M.; Dervan, P B.; Wemmer, D. E. *Journal* submitted for publication.

16. (a) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 9892-9899. (b) Dwyer, T. J.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, *115*, 9900-9906. (c) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, in press.

17. Still, W. C.; Kahn, M.; Mitra, A. *J. Org. Chem.* **1978**, *40*, 2923-2925.

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

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

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