DESIGN AND SYNTHESIS OF LIGANDS FOR RECOGNITION OF THE MAJOR AND MINOR GROOVES OF DNA

Thesis by

William Anthony Greenberg

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1998

(Submitted July 24, 1997)

Dedicated to the memory of my parents

Acknowledgments

I would like to thank my advisor, Professor Peter B. Dervan, for the role he has played in my education here at Caltech. I feel that above and beyond the training I've received in the Dervan group as an experimentalist, I have learned a lot of the more intangible details on how to be a good scientist. The knowledge I've gained from him will remain valuable throughout my career. I would also like to thank the members of my committee, Professors Andy Myers, Harry Gray, and Bill Goddard. Their input has been an important part of what has been a tremendous experience for me at Caltech.

If I paid due respect to every person, within the Dervan group and without, who has impacted on me over the last several years here, this page would turn into many pages. I won't do that, and I hope those that aren't mentioned by name know that I thank them nevertheless. I would however like to use this space to thank my coworkers in 330 and 320 Church. Pete Beal taught me the ropes when I was a first-year, and Scott Priestley, John Trauger, and Adam Urbach have all been extraordinary colleagues and friends. That can be said as well for Dave Liberles, Dave Herman, Ramesh Baliga, and Paul Floreancig. Each one of these people has enriched my experience at Caltech. I must also thank Eldon Baird, who collaborated with me on the work presented in the final chapter of this thesis. It was a real honor to have worked with him.

Finally I would like to acknowledge in print four people who have had a major impact on what I am today, but whom I cannot thank in person because they are no longer here. The first two are my mother and my father. The other two are Albert Camus and Mel Blanc.

Abstract

The chemical approach to the sequence specific recognition of double stranded DNA in this laboratory focuses on two distinct structural motifs. The first is oligonucleotide-directed triple helix formation. Pyrimidine oligonucleotides bind homopurine duplex sequences in the major groove by formation of T•AT and C+GC base triplets in a parallel orientation relative to the purine target strand. Purine rich oligonucleotides also bind homopurine sequences in the major groove, by formation of G•GC and A•AT or T•AT base triplets, in this case oriented antiparallel to the purine target strand. The lack of ability recognize CG and TA base pairs by either pyrimidine or purine oligonucleotides has resulted in efforts to design nonnatural bases which can recognize these base pairs in the context of triple helices. Such bases would greatly expand the generality of triple helix formation by expanding the range of targetable sequences.

The second structural motif consists of pyrrole-imidazole polyamides which form 2:1 side-by-side antiparallel complexes in the minor groove of DNA. A set of "pairing rules" has been elucidated in which an imidazole-pyrrole pair recognizes a GC base pair, a pyrrole-imidazole pair recognizes CG, and a pyrrole-pyrrole pair is degenerate for both AT and TA base pairs.

This thesis describes the application of quantitative DNase I footprinting to the evaluation of the energetics of triple helix formation by purine and pyrimidine oligonucleotides containing designed, nonnatural bases, as well the energetics of complex formation between covalently linked H-pin polyamides and the minor groove of DNA. Chapter Two describes the energetics of formation of sixteen triple helical complexes containing only the natural bases A, G, C, and T, which vary at single position within the purine motif. The values obtained set a basis with which to evaluate the energetics of triple

helices containing nonnatural bases, and demonstrate the exquisite sensitivity of triple helices to single base mismatches. Chapter Three describes the energetics of formation of purine motif triple helical complexes which contain single substitutions of the nonnatural purines inosine, 2-aminopurine, nebularine, and isoinosine. The results emphasize that new structural space must be explored in order to design bases which will specifically recognize CG and TA base pairs within the purine motif. Chapter Four describes the design and synthesis of methyl-substituted imidazole nucleosides and evaluation of the energetics of formation of purine motif triple helical complexes containing single substitutions of these nonnatural bases. The results suggest that use of small, five-membered ring heterocycles is likely a good design toward TA recognition, but that hydrophobic interactions between a designed nonnatural base and the 5-methyl group of thymine is not energetically favorable in the context of a purine motif triple helix. Chapter Five describes the design and synthesis of substituted pyrazole nucleosides and evaluation of the energetics of formation of purine motif triple helical complexes containing single substitutions of these nonnatural bases. The pyrazole substitution showed improved specificity towards TA, but at low affinity. This affinity could not be improved by appropriate substitution of an amino group, which instead improved affinity for GC. This points to the difficulty of forming a hydrogen bond to the O4 carbonyl of thymine relative to the more accessible O6 carbonyl of guanine. Chapter Six describes an evaluation of the energetics of formation of pyrimidine motif triple helical complexes containing a previously published nonnatural base and the discovery that as previously prepared, the base retains a benzoyl protecting group in the pyrimidine oligonucleotide. This benzoylated base shows specificity for CG and TA base pairs, similarly to the previously reported, structurally similar D_3, which has been shown to bind through an intercalative mode. Independent synthesis of the unprotected base through a new route proved that this base

did not bind specifically to any of the four base pairs in the context of a pyrimidine motif triple helix. Chapter Seven describes an evaluation of the energetics of formation of pyrimidine motif triple helical complexes containing a nonnatural base derived from a model heterocycle which was shown to bind isolated CG base pairs in organic solvent. The results suggest that model studies in organic solvent cannot accurately predict the behavior of nonnatural bases in water, in the context of a triple helix, since this base showed no specificity for any base pair by quantitative DNase I footprinting studies. Chapter Eight describes the design and solid phase synthesis of a series of pyrrole-imidazole polyamides which are covalently linked across central pyrrole nitrogens to increase affinity and specificity towards their DNA target sites according to the 2:1 polyamide:DNA binding model. DNase I footprint titration results show that these H-pin polyamides do bind with increased affinity and specificity relative to unlinked analogues, but that hairpin polyamides show a greater improvement, and are thus more optimal designs.

Table of Contents

List of Figures and Tables

CHAPTER FOUR

CHAPTER FIVE

CHAPTER SIX

CHAPTER SEVEN

CHAPTER EIGHT

CHAPTER ONE

Introduction

Deoxyribonucleic acid (DNA) is central to all life processes as the molecule which stores genetic information. The genetic information is encoded in the sequence of base pairs of the DNA, and differential expression of genes in different cell types, at different stages of a cell's life cycle, and under different environmental conditions is mediated by specific molecular recognition of DNA sequences by proteins such as transcription factors, architectural factors, and repair enzymes. Understanding the molecular basis of DNA recognition is therefore central to understanding gene expression, and is necessary for intervention in the case of aberrant gene expression, the basis of genetic disease, cancer, and viral infection. An approach to understanding molecular recognition of DNA begins with an analysis of its structure.

DNA structure

DNA is a polymer composed of four heterocyclic bases, adenine (A), cytosine (C), guanine (G), and thymine (T), linked to a deoxyribose phosphate backbone. Two polymer strands associate by Watson-Crick base pairing with formation of adenine-thymine (AT) and guanine-cytosine (GC) base pairs, generating a double-helical molecule in which the two strands are oriented in an antiparallel fashion (Figure 1).[1] The stability of the DNA double helix arises primarily from the hydrogen bonds formed in each base pair and the vertical stacking interactions between the aromatic bases. Although different sequences and environmental conditions lead to heterogeneous local structure, double helical DNA is primarily found in the B-form, which is characterized by a rise per residue of 3.4 Å, a turn angle of 36° per residue, a narrow minor groove (5.7 Å), and a wide major groove (11.7 Å).[1]

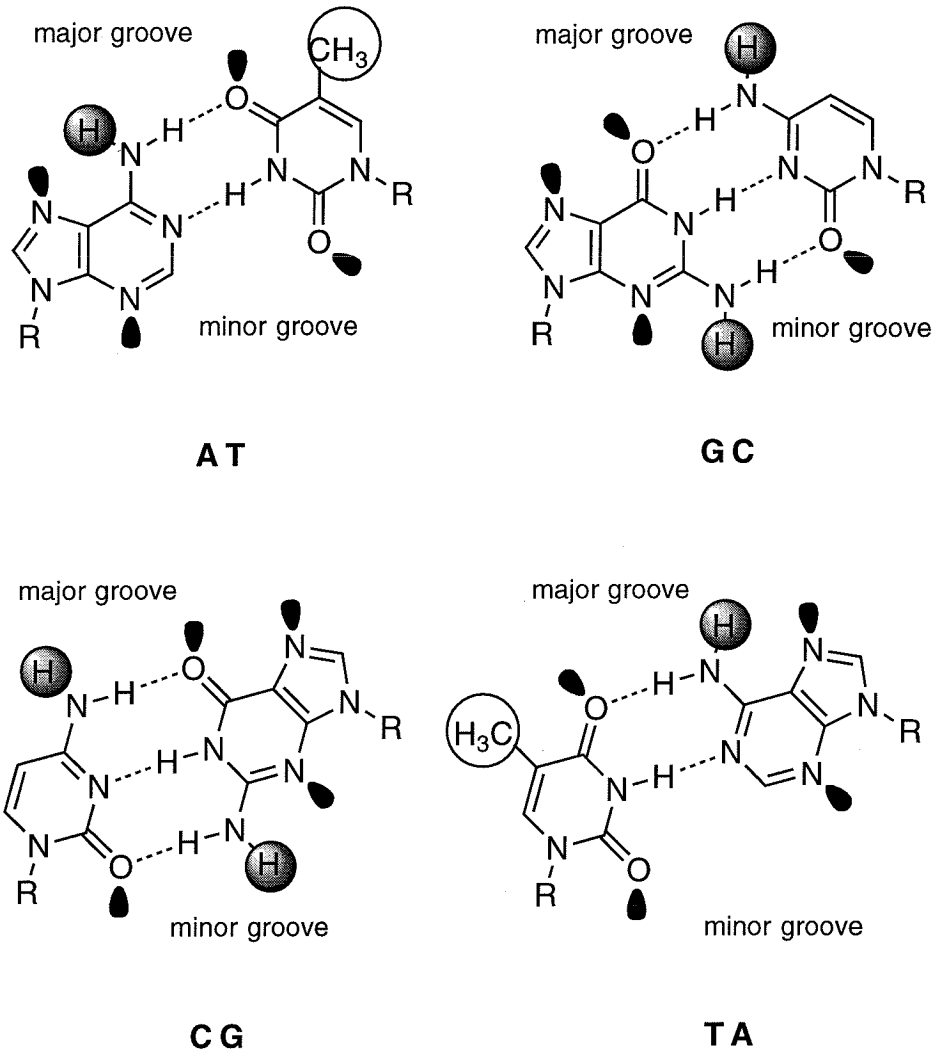


Figure 1. The four Watson-Crick base pairs. The sugar phosphate backbone is indicated by R, and major and minor grooves are indicated. Orbital symbols indicate hydrogen bond acceptors, and gray circles represent hydrogen bond donors. Hydrophobic methyl groups are circled.

Figure 1 illustrates the array of functional groups which are presented in the major and minor grooves by each base pair, and which provide the chemical basis for sequence specific recognition. In the major groove the four base pairs are distinguished by their individual pattern of hydrogen bond donors, acceptors, and hydrophobic groups. In the minor groove, AT and TA base pairs are essentially degenerate since each presents two lone pairs as potential hydrogen bond acceptors.

A Chemical Approach to DNA Recognition

Deriving inspiration from the sequence specificity of DNA binding proteins such as restriction enzymes, our laboratory has undertaken a chemical design-synthesis approach to the development of DNA binding molecules.[2] This chemical approach is based on rational design using structure-function relationships to identify motifs that allow for the generalizable design of molecules capable of specifically binding any desired DNA sequence. It is a primary goal to develop models for noncovalent interactions between ligand and DNA such that modification of the ligand's covalent structure will result in predictable changes in the noncovalent interactions with DNA.

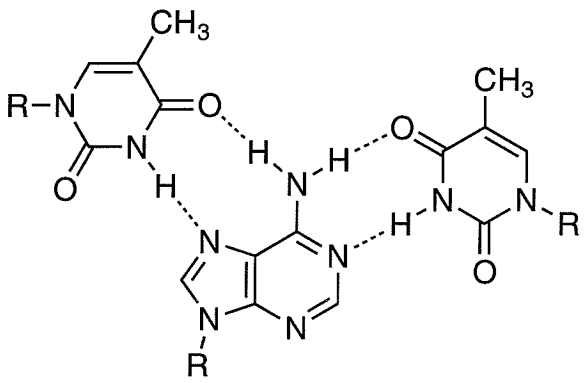
Currently, efforts in the laboratory focus on two distinct structural motifs for sequence specific DNA binding molecules. The first consists of oligonucleotides which bind in the major groove by triple helix formation, and the second consists of polyamides which bind in the minor groove. Work within both of these areas is described in this thesis, and each motif is introduced in the following sections of this chapter.

Triple Helix Formation

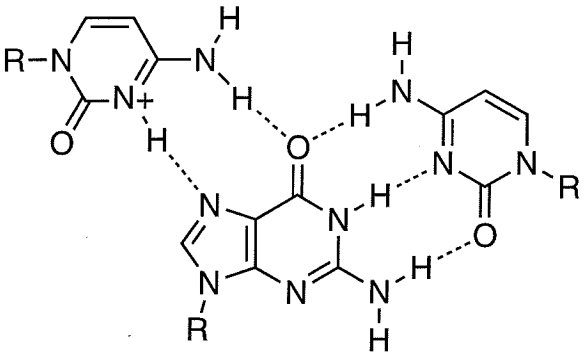
Only four years after the landmark elucidation of the structure of double helical DNA by Watson and Crick,[3] triple helical nucleic acid structures were detected by Felsenfeld, Davies, and Rich.[4] It was reported that RNA homopolymers consisting of poly(U) and poly(A) associated in a 2:1 complex, and the authors proposed a hydrogen bonding scheme to explain its stability which utilized interactions later characterized crystallographically by Hoogsteen.[5] Subsequently, triple helical RNA containing poly(C) and poly(G) in a 2:1 stoichiometry was detected at acidic pH.[6] A similar Hoogsteen hydrogen bonding pattern, involving protonated cytosine, was proposed.

In 1987, Moser and Dervan reported that 15mer pyrimidine oligodeoxyribonucleotides could bind to homopurine sequences within DNA restriction fragments.[7] It was demonstrated that the oligonucleotide bound in the major groove, oriented parallel to the purine strand of the duplex. They proposed that the stability of the complex was mediated by T•AT and C+GC triplets analogous to those proposed many years before (Figure 2). NMR studies have since confirmed the proposed base triplet structures and have provided further detail into the global structure of triple helical DNA.[8]

Purine motif triple helix formation. A second structural motif for triple helix formation was reported by Beal and Dervan in 1991, when it was shown that purine-rich oligonucleotides composed of deoxyguanosine (G) and deoxyadenosine (A) or deoxyguanosine and thymidine also bound to homopurine duplex sequences.[9] In this case the oligonucleotides were demonstrated to bind in the major groove oriented antiparallel to the purine strand of the duplex. The stability of the triplexes was proposed to be mediated by G•GC and A•AT or "reverse Hoogsteen" T•AT triplets (Figure 4). These triplet structures were subsequently confirmed by NMR studies.[10]



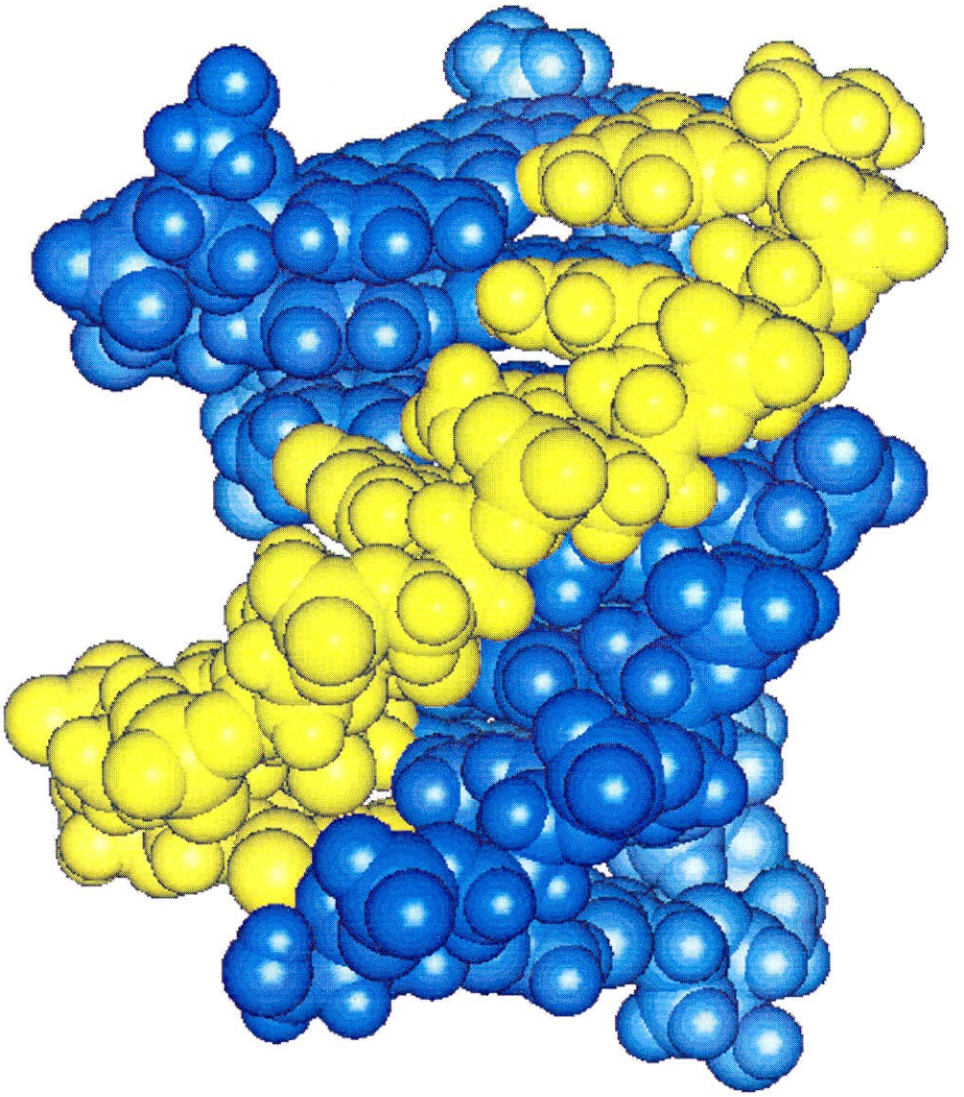
T•AT

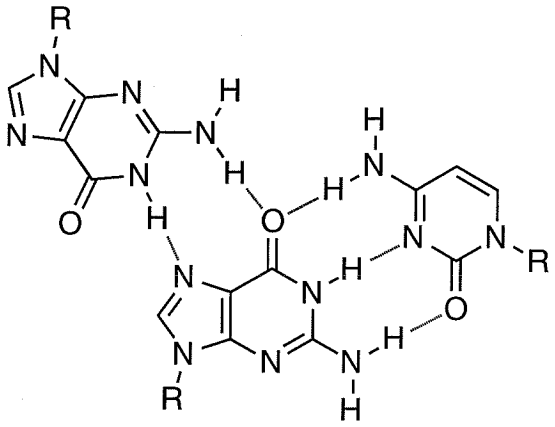


C+GC

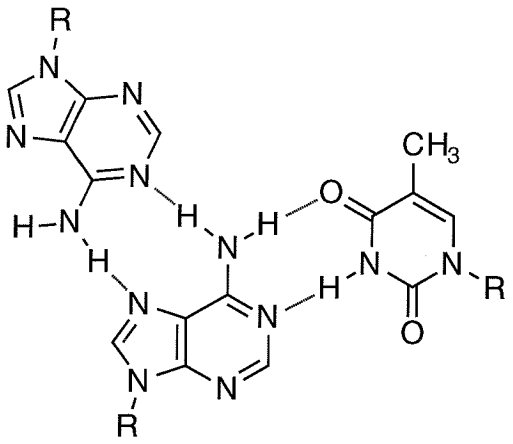
Figure 2. Structures of the T•AT and C+GC base triplets which mediate triple helix formation in the pyrimidine motif. The backbone of the third strand is oriented parallel to the purine strand of the duplex. Note that protonation of the third strand cytosine is required in order to form two hydrogen bonds to the GC base pair.

Figure 3. CPK model of a pyrimidine motif triple helix structure based on NMR-derived constraints (reference 8). The pyrimidine third strand is depicted in yellow, and the Watson-Crick strands are depicted in blue.

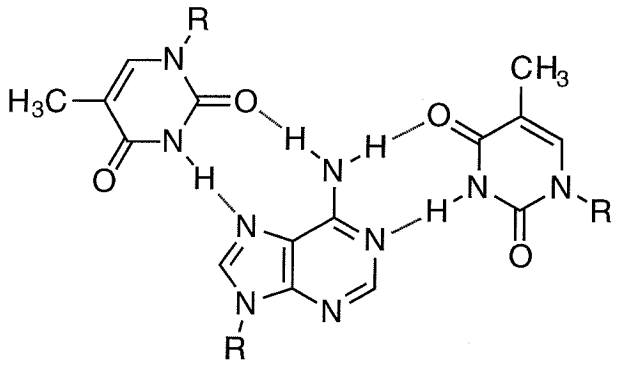




G•GC



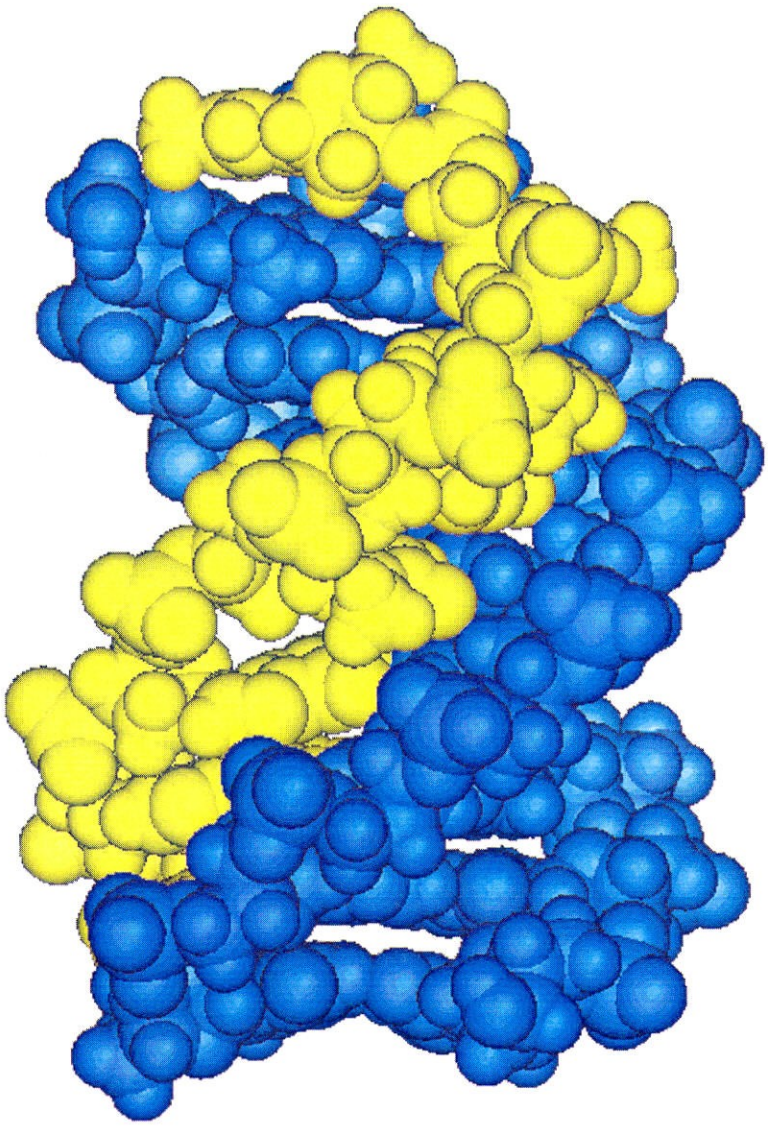
A•AT



T•AT

Figure 4. Structures of the G•GC, A•AT, and T•AT base triplets which mediate triple helix formation in the purine motif. The backbone of the third strand is oriented antiparallel to the purine strand of the duplex.

Figure 5. CPK model of a purine motif triple helix structure based on NMR-derived constraints (reference 10). The purine-rich third strand is depicted in yellow, and the Watson-Crick strands are depicted in blue.



The utility of oligonucleotide-directed triple helix formation has been demonstrated through a series of experiments in which DNA cleavage was effected at a single site in bacteriophage λ, yeast, and finally human chromosomal DNA.[11] Inhibition of a variety of DNA binding proteins such as methylases and transcription factors has been demonstrated *in vitro* through oligonucleotide-directed triple helix formation.[12] Possible *in vivo* applications (the anti-gene approach)[13] have been the subject of extensive efforts.[14]

The thermodynamics of triple helix formation have been studied in great detail by a variety of methods, including calorimetry,[15] UV melting,[16] affinity cleavage titration,[17] and quantitative DNase I footprint titration.[18] The stability of triple helical complexes are sensitive to oligonucleotide length,[17] single base mismatches,[16,17,19] pH in the case of the pyrimidine motif,[20] cation valence and concentration,[21] and temperature.[22]

The generality of both pyrimidine and purine motif triple helices is hampered by the limitation of recognizing only AT and GC base pairs with high affinity and specificity. An elegant approach towards recognizing sequences other than strict homopurine runs involves switching recognition to the alternate strand at a run of TA and CG base pairs.[23] However, a general approach to recognition of any given sequence would require the development of novel, nonnatural nucleoside structures which recognize TA or CG base pairs in the context of a triple helix. Extensive efforts have been made toward this approach, with thus far very limited success. These efforts are discussed in detail in Chapters 3-7 of this thesis.

Polyamides

The natural products netropsin and distamycin A, which are oligomers of N-methyl-4-aminopyrrole-2-carboxylic acid with positively charged tails (Figure 6), bind in

the minor groove of A,T-rich DNA sequences.[24] The stability of polyamide-DNA complexes is derived from hydrogen bonding between amide protons and the thymine O2 and adenine N3 groups, electrostatic interactions between the positively charged tail groups and the negatively charged DNA, and through van der Waals contacts between the polyamide and the deep walls of the minor groove.[25]

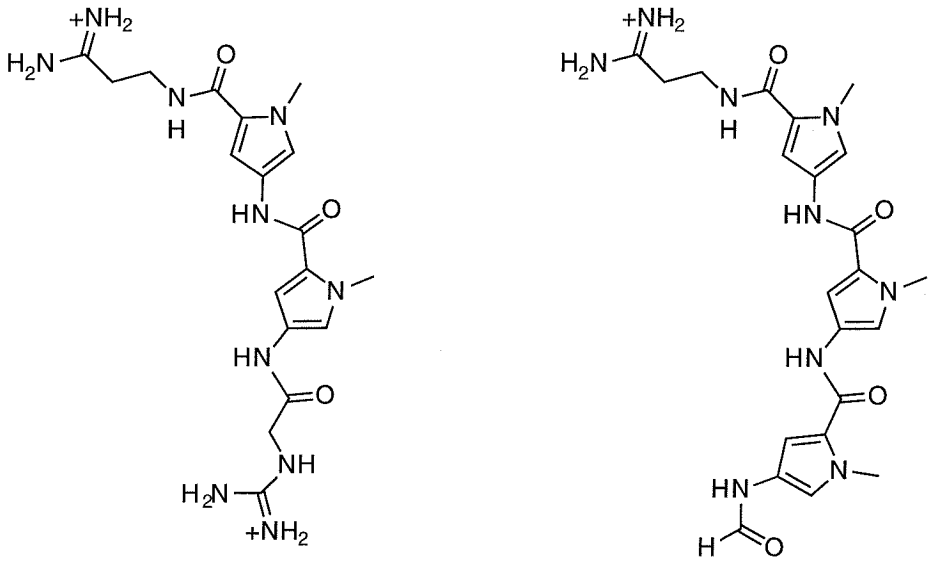


Figure 6. Structures of the natural products netropsin (left) and distamycin (right).

X-ray crystallography[25] and NMR studies[26] on netropsin and distamycin show the polyamides seated deeply in the minor groove as 1:1 complexes. Rational design led to the development and synthesis of PyrPyPy-Dp and ImPyPy-Dp (Figure 7) where pyridine and imidazole carboxylic acids containing hydrogen bond acceptors were included. These were expected to form a hydrogen bond to the exocyclic amino group of guanine, allowing specific recognition of 5'-(A,T)G(A,T)$_3$-3' sequences. In practice these compounds bound

to 5'-TGTCA-3', suggesting that they were not binding according to the predicted model.[27,28]

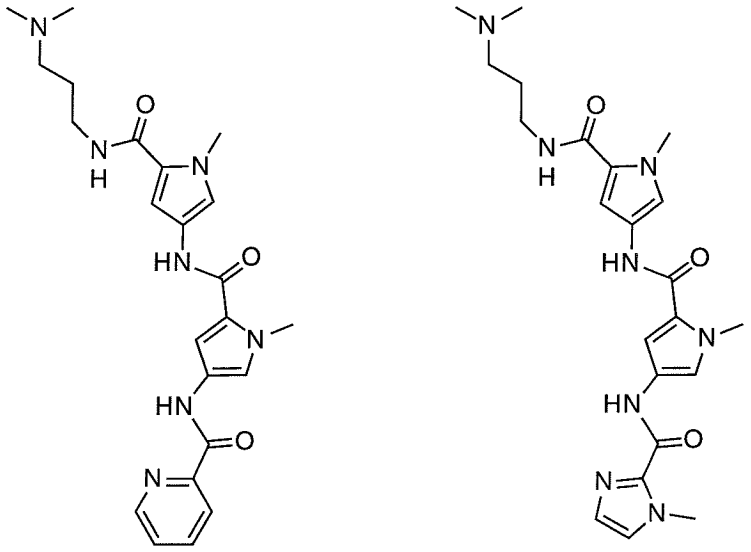


Figure 7. Structures of designed polyamides PyrPyPy-Dp and ImPyPy-Dp.

2:1 polyamide-DNA complexes. In 1989, Wemmer and coworkers observed that at millimolar concentrations distamycin formed an antiparallel, side-by-side 2:1 complex with the sequence 5'-AAATT-3'.[29] In light of this result, the data for PyrPyPy-Dp and ImPyPy-Dp were reexamined, and the 2:1 model explained the specificity for 5'-(A,T)G(A,T)C(A,T)-3' (Figure 8). Subsequently NMR studies on these designed polyamides in collaboration with the Wemmer group confirmed that 2:1 antiparallel, side-by-side binding was occurring (Figure 9).[30]

These results suggested a set of "pairing rules" for binding a given sequence of DNA in the minor groove with polyamides according to the 2:1 model. An imidazole/pyrrole pair recognized GC, while a pyrrole/imidazole pair recognizes CG. A

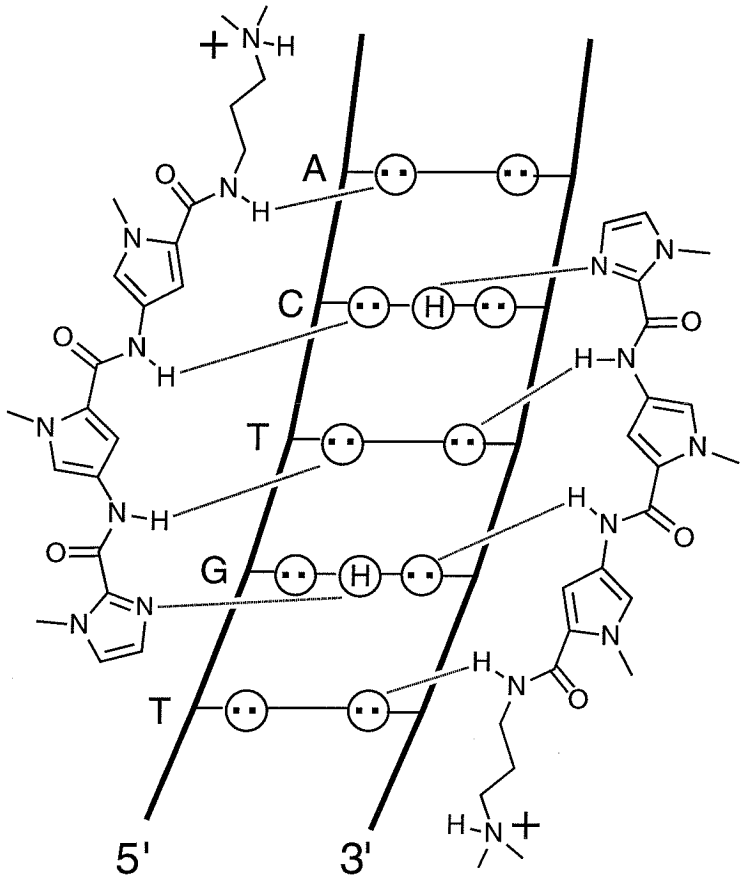
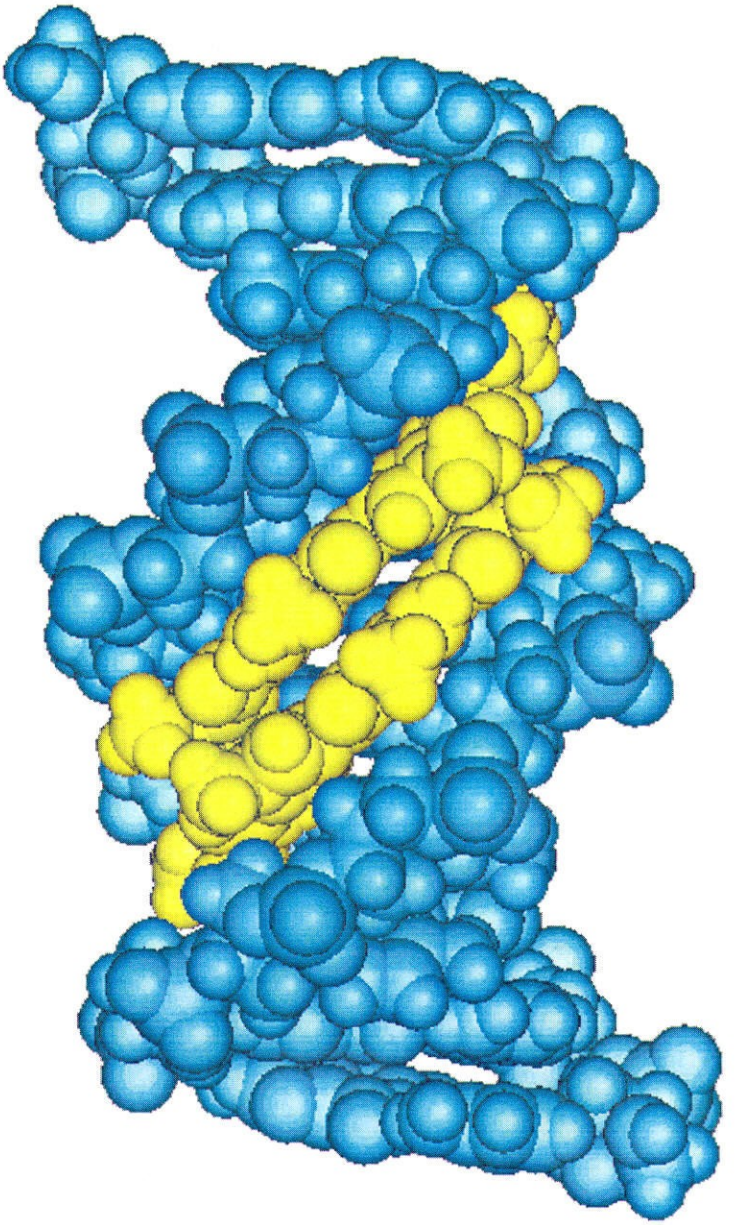


Figure 8. Binding model of the 2:1 complex formed between ImPyPy-Dp and 5'-TGTCA-3'. Circles with dots represent lone pairs of purine N3 and pyrimidine O2. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated as dotted lines.

pyrrole/pyrrole pair degenerately binds both AT and TA. This model has been successfully applied in binding a wide range of sequences, most dramatically in the case of ImPyImPy-Dp, which completely reverses the (A,T) specificity of the natural product distamycin by binding the sequence 5'-TGCGCA-3'.[31]

Figure 9. Structure of the 2:1 complex formed between ImPyPy-Dp and 5-TGTCA-3' as determined by NMR studies. The DNA is shown in blue, and the two polyamide molecules in yellow.



Covalently linked polyamides. Affinity and specificity of polyamides within the 2:1 motif has been enhanced by covalently linking the individual chains. An initial design linked them across central pyrrole nitrogens.[32] This design, the H-pin motif, is described in more detail in chapter 8. A second generation design involved linkage in a head-to-tail fashion from the N-terminus of one chain to the C-terminus of the other, resulting in a hairpin polyamide.[33] Furthermore, both ends could be covalently linked, resulting in cyclic polyamides.[34]

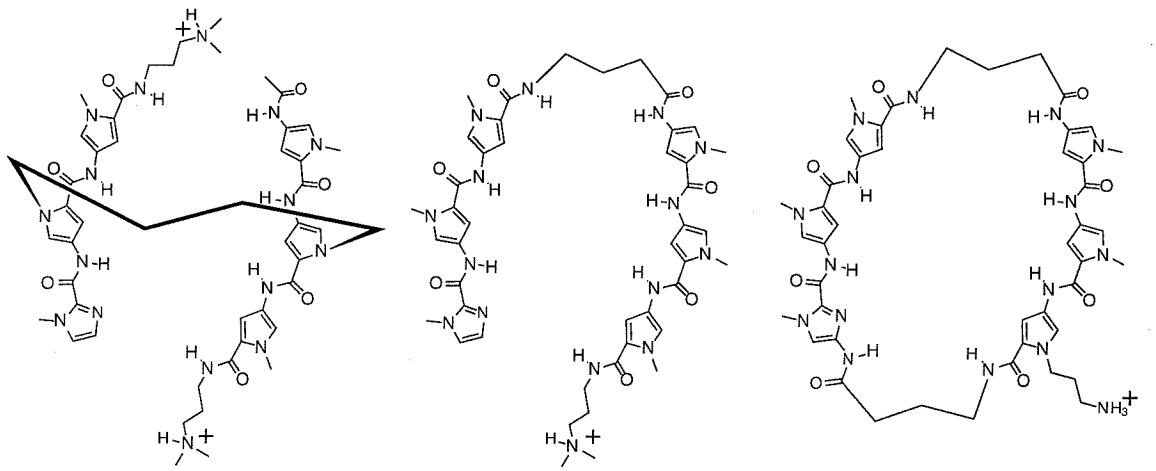


Figure 10. Structures of H-pin (left), hairpin (center), and cyclic (right) polyamides.

A major advance in the development of polyamides for recognition of the minor groove of DNA came with methodology for the solid phase synthesis of these molecules.[35] This has made polyamide synthesis much more rapid, and has allowed for the preparation of larger, more complex polyamides which would have been very difficult to prepare otherwise. Examples of such molecules include β-alanine linked extended polyamides,[36]

extended hairpins,[37] and oligonucleotide-polyamide conjugates which bind both in the major groove and the minor groove.[38] Solid phase synthetic methodology has also expedited the application of polyamides to modulating gene expression *in vivo*.[39]

Description of thesis work

This thesis describes the application of quantitative DNase I footprinting to the evaluation of the energetics of triple helix formation by purine and pyrimidine oligonucleotides containing designed, nonnatural bases, as well the energetics of complex formation between covalently linked H-pin polyamides and the minor groove of DNA. Chapter Two describes the energetics of formation of sixteen triple helical complexes containing only the natural bases A, G, C, and T, which vary at single position within the purine motif. The values obtained set a basis with which to evaluate the energetics of triple helices containing nonnatural bases, and demonstrate the exquisite sensitivity of triple helices to single base mismatches. Chapter Three describes the energetics of formation of purine motif triple helical complexes which contain single substitutions of the nonnatural purines inosine, 2-aminopurine, nebularine, and isoinosine. The results emphasize that new structural space must be explored in order to design bases which will specifically recognize CG and TA base pairs within the purine motif. Chapter Four describes the design and synthesis of methyl-substituted imidazole nucleosides and evaluation of the energetics of formation of purine motif triple helical complexes containing single substitutions of these nonnatural bases. The results suggest that use of small, five-membered ring heterocycles is likely a good design toward TA recognition, but that hydrophobic interactions between a designed nonnatural base and the 5-methyl group of thymine is not energetically favorable in the context of a purine motif triple helix. Chapter

Five describes the design and synthesis of substituted pyrazole nucleosides and evaluation of the energetics of formation of purine motif triple helical complexes containing single substitutions of these nonnatural bases. The pyrazole substitution showed improved specificity towards TA, but at low affinity. This affinity could not be improved by appropriate substitution of an amino group, which instead improved affinity for GC. This points to the difficulty of forming a hydrogen bond to the O4 carbonyl of thymine relative to the more accessible O6 carbonyl of guanine. Chapter Six describes an evaluation of the energetics of formation of pyrimidine motif triple helical complexes containing a previously published nonnatural base[40] and the discovery that as previously prepared, the base retains a benzoyl protecting group in the pyrimidine oligonucleotide. This benzoylated base shows specificity for CG and TA base pairs, similarly to the previously reported, structurally similar D_3, which has been shown to bind through an intercalative mode.[41] Independent synthesis of the unprotected base through a new route proved that this base did not bind specifically to any of the four base pairs in the context of a pyrimidine motif triple helix. Chapter Seven describes an evaluation of the energetics of formation of pyrimidine motif triple helical complexes containing a nonnatural base derived from a model heterocycle which was shown to bind isolated CG base pairs in organic solvent.[42] The results suggest that model studies in organic solvent cannot accurately predict the behavior of nonnatural bases in water, in the context of a triple helix, since this base showed no specificity for any base pair by quantitative DNase I footprinting studies. Chapter Eight describes the design and solid phase synthesis of a series of pyrrole-imidazole polyamides which are covalently linked across central pyrrole nitrogens to increase affinity and specificity towards their DNA target sites according to the 2:1 polyamide:DNA binding model. DNase I footprint titration results show that these H-pin polyamides do bind with increased affinity and specificity relative to unlinked analogues,

but that hairpin polyamides show a greater improvement, and are thus more optimal designs.

References

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

(2) Dervan, P.B. *Science* **1986**, *232*, 464-471.

(3) Watson, J.D.; Crick, F.H.C. *Nature* **1953**, *171*, 738-739.

(4) Felsenfeld, G.; Davies, D.R.; Rich, A. *J. Am. Chem. Soc.* **1957**, *79*, 2023-2024.

(5) Hoogsteen, K. *Acta Crystallogr.* **1959**, *12*, 822-823.

(6) (a) Lipsett, M.N. *J. Biol. Chem.* **1964**, *239*, 1256-1260. (b) Howard, F.B.; Frazier, J.; Lipsett, M.N.; Miles, H.T. *Biochem. Biophys. Res. Commun.* **1964**, *17*, 93-102.

(7) Moser, H.E.; Dervan, P.B. *Science* **1987**, *238*, 645-650.

(8) (a) de los Santos, C.; Rosen, M.; Patel, D.J. *Biochemistry* **1989**, *28*, 7282-7289. (b) Rajagopal, P.; Feigon, J. *Nature* **1989**, *339*, 637-640. (c) Rajagopal, P.; Feigon, J. *Biochemistry* **1989**, *28*, 7859-7870. (d) Live, D.H.; Radhakrishnan, I.; Misra, V.; Patel, D.J. *J. Am. Chem. Soc.* **1991**, *113*, 4687-4688.

(9) Beal, P.A.; Dervan, P.B. *Science* **1991**, *251*, 1360-1363.

(10) (a) Radhakrishnan, I.; Patel, D.J. *J. Am. Chem. Soc.* **1993**, *115*, 1615-1617. (b) Radhakrishnan, I.; de los Santos, C.; Patel, D.J. *J. Mol. Biol* **1993**, *234*, 188-197. (c) Radhakrishnan, I.; Patel, D.J. *Structure* **1993**, *1*, 135-152.

(11) (a) Strobel, S.A.; Moser, H.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1988**, *110*, 7927-7928. (b) Strobel, S.A.; Dervan, P.B. *Science* **1990**, *249*, 73-75. (c) Strobel, S.A.; Dervan, P.B. *Nature* **1991**, *350*, 172-174. (d) Strobel, S.A.; Doucette-Stamm, L.A.; Riba, L.; Housman, D.E.; Dervan, P.B. *Science* **1991**, *254*, 1639-1642.

(12) (a) Maher, L.J. III; Wold, B.; Dervan, P.B. *Science* **1989**, *245*, 725-730. (b) Maher, L.J. III; Dervan, P.B.; Wold, B. *Biochemistry* **1992**, *31*, 70-81.

(13) Thuong, N.T.; Helene, C. *Angew. Chem. Int. Ed. Engl.* **1993**, *32*, 666-690.

(14) Hacia, J.G. Ph. D. Thesis, California Institute of Technology, 1995.

(15) Plum, E.G.; Park, Y.; Singleton, S.F.; Dervan, P.B.; Breslauer, K.J. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 9436-9440.

(16) (a) Mergny, J.-L.; Sun, J.-S.; Rougée, M.; Montenay-Garestier, T. Barcelo, F.; Chomilier, J.; Hélène, C. *Biochemistry* **1991**, *30*, 9791-9798. (b) Rougée, M.; Faucon, B.; Mergny, J. L.; Barcelo, F.; Giovannageli, C.; Garestier, T.; Hélène, C. *Biochemistry* **1992**, *31*, 9269-9278. (c) Fossella, J. A.; Kim, Y. J.; Shih, H.; Richards, E. G.; Fresco, J. R. *Nucleic Acids Res.* **1993**, *21*, 4511-4515.

(17) Singleton, S. F.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 6957-6965.

(18) Priestley, E.S.; Dervan, P.B. *J. Am. Chem. Soc.* **1995**, *117*, 4761-4765.

(19) Best, G.C.; Dervan, P.B. *J. Am. Chem. Soc.* **1995**, *117*, 1187-1193.

(20) Singleton, S. F.; Dervan, P. B. *Biochemistry* **1992**, *31*, 10995-11003.

(21) Singleton, S. F.; Dervan, P. B. *Biochemistry* **1993**, *32*, 13171-13179.

(22) Singleton, S. F.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 10376-10382.

(23) (a) Jayasena, S. D.; Johnston, B. H. *Biochemistry* **1992**, *31*, 320-327. (b) Beal, P. A.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 4976-4982. (c) Jayasena, S. D.; Johnston, B. H. *Nucl. Acids Res.* **1992**, *20*, 5279-5288. (d) Jayasena, S. D.; Johnston, B. H. *Biochemistry* **1993**, *32*, 2800-2807.

(24) Zimmer, C.; Wahnert, U. *Prog. Biophys. Molec. Biol.* **1986**, *47*, 31-112.

(25) Kopka, M.L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R.E. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 1376-1380.

(26) (a) Patel, D.J.; Shapiro,L. *J. Biol. Chem.* **1986**, *261*, 1230-1240. (b) Pelton, J.G.; Wemmer, D.E. *Biochemistry* **1988**, *27*, 8088- 8096.

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

(28) (a) Wade, W.S.; Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1992**, *114*, 8783-8794. (b) Wade, W.S.; Mrksich, M.; Dervan, P.B. *Biochemistry* **1993**, *32*, 11385-11389.

(29) (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.

(30) 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.

(31) (a) Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *Science* **1994**, *266*, 646- 650. (b) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1995**, *117*, 3325-3332.

(32) 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**, *116*, 3663-3664. (d) Chen, Y.H.; Lown, J.W. *J. Am. Chem. Soc.* **1994**, *116*, 6995-7001.

(33) (a) Mrksich, M.; Parks, M.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1994**, *116*, 7983-7988. (b) Trauger, J.W.; Baird, E.E.; Dervan, P.B. *Nature* **1996**, *382*, 559-561.

(34) Cho, J.; Parks, M.E.; Dervan, P.B. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 10389-10392.

(35) Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1996**, *118*, 6141-6146.

(36) Trauger, J.W.; Baird, E.E.; Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1996**, *118*, 6160-6166.

(37) Trauger, J.W.; Baird, E.E.; Dervan, P.B. *Chem. Biol.* **1996**, *3*, 369-377.

(38) (a) Szewczyk, J.W.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1996**, *118*, 6778-6779. (b) Szewczyk, J.W.; Baird, E.E.; Dervan, P.B. *Angew. Chem. Int. Ed. Engl.* **1996**, *35*, 1487-1489.

(39) Gottesfeld, J.M.; Neely, L.; Trauger, J.W.; Baird, E.E.; Dervan, P.B.; *Nature* **1997**, *387*, 202-205.

(40) Huang, C.Y.; Miller, P.S. *J. Am. Chem. Soc.* **1993**, *115*, 10456-10457.

(41) (a) Griffin, L.C.; Kiessling, L.L.; Beal, P.A.; Gillespie, P.; Dervan, P.B. *J. Am. Chem. Soc.* **1992**, *114*, 7976-7982. (b) Koshlap, K.M.; Gillespie, P.; Dervan, P.B.; Feigon, J. *J. Am. Chem. Soc.* **1993**, *115*, 7908-7909.

(42) Zimmerman, S.C.; Schmitt, P. *J. Am. Chem. Soc.* **1995**, *117*, 10769-10770.

CHAPTER TWO

Energetics of Formation of Sixteen Triple Helical Complexes Which Vary at a Single Position within a Purine Motif

Introduction

Oligonucleotide-directed triple helix formation is a versatile method for the sequence-specific recognition of double helical DNA.[1] Triple helices can be classified into two structural motifs: those in which the third strand is primarily composed of pyrimidine bases, and those in which the third strand is composed primarily of purine bases.[1-3] In addition, triple helices composed of combinations of these two motifs can be formed within some sequence contexts.[4] Triple helix formation is sensitive to the length of the third strand,[1a,5] single base mismatches,[1a,5,6] pH,[7] cation concentration and valence,[8] temperature,[9] and backbone composition (DNA or RNA) of the three strands.[10] Oligonucleotide-directed triple helix formation has been shown to inhibit sequence specific DNA binding proteins[11] and has been used to mediate single site cleavage of human chromosomal DNA.[12] The ability to target a broad range of DNA sequences,[1,13] and the high specificity and stability of the resulting local triple helical structures make this a powerful technique for the recognition of single sites within megabase segments of double helical DNA.

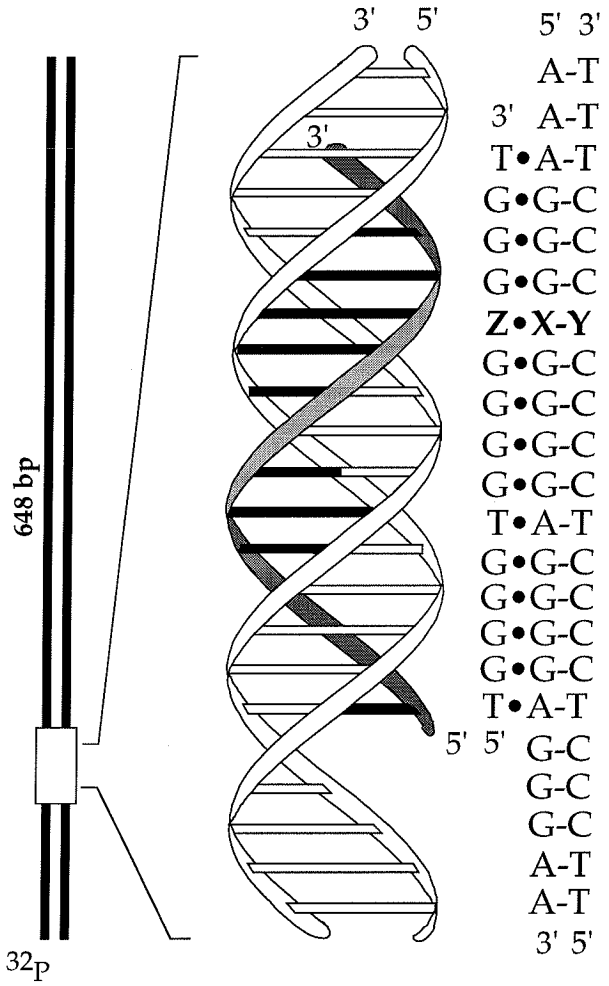
In an effort to understand the interactions which give rise to recognition in the major groove of DNA by oligonucleotides, the specificity afforded by the four natural bases in triple helix formation has been qualitatively characterized by affinity cleavage,[14] optical melting,[15] gel mobility shift,[16] and intramolecular triple helix formation.[17] Specific interactions in triple helix formation implicated by these studies have been characterized by NMR spectroscopy.[18-20] In the pyrimidine motif, Hoogsteen-type hydrogen bonds have been observed between thymine (T) bases in the third strand and adenine-thymine (AT) base pairs in the duplex,[18] and between N3 protonated cytosine (C+) of 5-methylcytosine (^{me}C+) in the third strand and guanine-cytosine (GC) base pairs in the duplex.[19] In the purine motif, reversed-Hoogsteen type hydrogen bonds have been observed between G

bases in the third strand and GC base pairs in the duplex,[20] and between T^{20a} or A^{20b} in the third strand and AT base pairs in the duplex.

The introduction of quantitative methods has allowed the determination of equilibrium association constants for the binding of an oligonucleotide at a single site on a DNA plasmid fragment.[5,21] Previous reports on the purine motif suggest that in addition to the aforementioned three triplets there are several others that have intermediate stability,[2b,c] although little is known about the quantitative difference among all sixteen triplets composed of natural bases. Here we report the equilibrium association constants and free energies of formation of sixteen triple helical complexes which vary at a single position (Z•XY where Z = A, G, C, T and Watson-Crick XY = AT, GC, CG, and TA) (Figure 1). We are primarily interested in comparing the differences in free energy values ($\Delta\Delta G°$) which are relevant to the issue of specificity, as in an analogous study within the pyrimidine motif.[22]

Results and Discussion

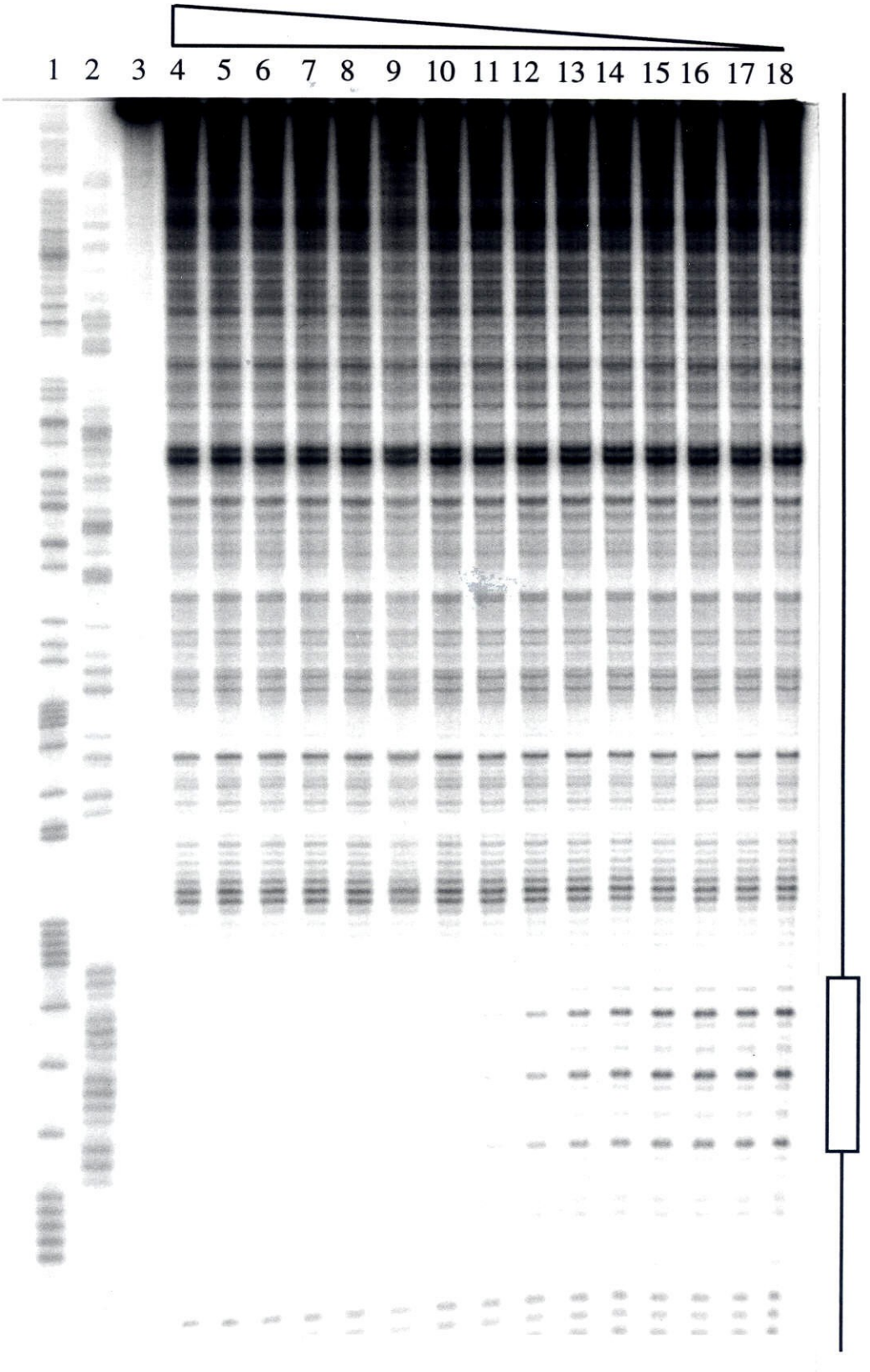
Methods. A detailed description of the quantitative DNAse I footprint titration has been presented elsewhere for protein-DNA,[23] small molecule-DNA,[24] and oligonucleotide-DNA[21] interactions, hence, the protocol is outlined briefly in the experimental section. A 3'-^{32}P end labeled 648 base pair restriction fragment containing a single purine rich 15 base pair target site 5'-d(AGGGGAGGGGXGGGA)-3' was allowed to equilibrate at 37°C, pH 7.4 with a series of concentrations of third strand purine-rich oligonucleotides 5'-d(TGGGZGGGGTGGGGT)-3' that ranged from 8 μM to 800 pM (Figure 2). Following a 48 hour equilibration, DNAse I was added and digestion was allowed to proceed for six minutes. After quenching, the reaction mixtures were separated by denaturing polyacrylamide gel electrophoresis and the resulting gels were imaged by storage phosphor autoradiography.



Z = A, G, C, T
X-Y = AT, GC, CG, TA

Figure 1. (Left) Ribbon model of the triple helix formed upon binding of the 15mer oligonucleotide probe to the 15 bp target site within a 648 bp restriction fragment. (Right) The sequence modeled by the ribbons is shown. The position of the variable triplet is indicated as Z•XY.

Figure 2. Autoradiogram of an 8% denaturing polyacrylamide gel used to separate fragments generated by DNAse I digestion in a quantitative footprint titration experiment. The bar to the right indicates the position of the 15 bp binding site within the 648 bp restriction fragment where XY = AT. (Lane 1) Products of an adenine-specific reaction. (Lane 2) Products of a guanine-specific reaction. (Lane 3) Intact 3' labeled DNA after incubation in the absence of third strand oligonucleotide. (Lanes 4-18) DNAse I digestion products obtained in the presence of varying concentrations of oligonucleotide where Z = T: 8 μM (lane 4); 4 μM (lane 5); 2 μM (lane 6); 1 μM (lane 7); 500 nM (lane 8); 200 nM (lane 9); 100 nM (lane 10); 50 nM (lane 11); 20 nM (lane 12); 10 nM (lane 13); 5 nM (lane 14); 2 nM (lane 15); 1 nM (lane 16); 800 pM (lane 17); no oligonucleotide (lane 18).


1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18

Integration of site and reference blocks allowed the determination of the apparent fractional occupancy of the site at each oligonucleotide concentration. A binding isotherm was fit to the resulting pairs of θ_{app}, $[O]_{tot}$ values (see experimental section) and the equilibrium association constant (K_T) was calculated. K_T values from three independent titrations were averaged to obtain each of the association constants reported in Table 1.

Affinity. Experimental conditions such as temperature, salt concentration, and pH are chosen such that the lowest and highest equilibrium binding constants for the sixteen triple helical complexes are within the range that can be measured with this technique. The footprint titration experiments were carried out at 37°C, pH 7.4, 50 mM tris acetate, 10 mM NaCl, 3 mM $MgCl_2$. The values of the sixteen association constants (K_T) range from 2.7 x 10^5 to 8.9 x 10^7 M^{-1} (Table 1). An examination of the data confirms that three base triplets (G•GC, T•AT, and A•AT) are particularly stable. Triple helical complexes containing the G•GC, T•AT, and A•AT triplets at the Z•XY position are the most stable of the sixteen complexes measured. The total standard free energy of these triple helix forming reactions are ΔG° = -11.0, -11.3, and -11.0 kcal mol^{-1}, respectively (Table 2). In addition, the triplets $T•CG^{2c,25}$ and $A•GC^{2c,26}$ have intermediate stability. The total standard free energy of triple helix forming reactions containing these triplets at the Z•XY position are -10.1 and -9.8 kcal mol^{-1}, respectively. Triple helices containing these triplets at the single variable position are approximately 1 kcal mol^{-1} less stable than those containing G•GC, T•AT, or A•AT at that position.

Specificity. If specific local triple helices are to be useful structures for accomplishing the recognition of single sites in megabase DNA, it is important that the complex be stabilized by specific interactions and be sensitive to single base mismatches. The results presented here demonstrate that either A or T in the third strand can accomplish the sequence specific recognition of AT base pairs in double helical DNA, G

Table 1. Association constants (K_T) for the formation of sixteen triple helical complexes containing the Z•XY triplets at 37°C, 10 mM NaCl, 3 mM $MgCl_2$, 50 mM tris acetate, pH 7.4.[a,b]

XY	Z=A	G	C	T
AT	5.3 (±1.4) x 10^7	1.9 (±0.2) x 10^6	4.0 (±1.2) x 10^6	8.9 (±0.5) x 10^7
GC	8.2 (±0.8) x 10^6	5.5 (±0.4) x 10^7	1.7 (±0.4) x 10^6	2.9 (±0.2) x 10^6
CG	1.1 (±0.3) x 10^6	5.8 (±0.8) x 10^5	8.5 (±0.9) x 10^5	1.3 (±0.1) x 10^7
TA	8.5 (±2.7) x 10^5	2.7 (±0.7) x 10^5	8.3 (±2.4) x 10^5	1.7 (±0.7) x 10^6

[a] K_T values are reported as the mean (± the standard error of the mean) of three measurements. The K_T values are reported in units of M^{-1}. [b] The identity of the base Z is indicated across the top of the columns; the identity of the Watson Crick base pair XY is indicated on the left side of the rows.

Table 2. Free energy (ΔG°) of formation of sixteen triple helical complexes containing the Z•XY triplets at 37°C, 10 mM NaCl, 3 mM $MgCl_2$, 50 mM tris acetate, pH 7.4.[a,b]

	A	G	C	T
AT	-11.0±0.2	-8.9±0.1	-9.4±0.3	-11.3±0.1
GC	-9.8±0.1	-11.0±0.2	-8.8±0.1	-9.2±0.1
CG	-8.6±0.2	-8.2±0.1	-8.4±0.1	-10.1±0.1
TA	-8.4±0.2	-7.7±0.2	-8.4±0.2	-8.8±0.3

[a] Free energy values are calculated from the measured association constants at 37°C and are reported in kcal mol^{-1}. [b] The identity of the base in the third strand, Z, is indicated across the top row; the identity of the Watson Crick base pair, XY, is indicated to the left of the rows.

specifically recognizes GC base pairs, and T specifically recognizes CG base pairs (comparison across rows of Table 2). For example, T•AT and A•AT are more stable than C•AT and G•AT by ≥1.6 kcal mol^{-1}. G•GC is more stable than A•GC, C•GC, and T•GC by ≥1.2 kcal mol^{-1}. T•CG is more stable than A•CG, C•CG, and G•CG by ≥1.5 kcal mol^{-1} (Figure 3). In the reciprocal sense (Figure 4), A•AT is 1.2 kcal mol^{-1} more stable than A•GC, the next most stable A•XY triplet, G•GC is 2.1 kcal mol^{-1} more stable than G•AT, and T•AT is 1.2 kcal mol^{-1} more stable than T•CG.

The intermediate stability of the T•CG and A•GC triplets can be rationalized by single specific hydrogen bonding contacts (Figure 5), whereas A•AT, T•AT, and G•GC each have two. While T offers a means of recognizing CG base pairs in this sequence context, it is not specific in that T recognizes AT base pairs more strongly than CG. This degenerate recognition by T suggests that a novel base that *specifically* recognizes CG base pairs with high affinity is needed,[2d] particularly if sequences containing more than one CG base pair are to be targeted. In such cases the use of T to recognize CG would compromise the overall specificity of the oligonucleotide.

The data in Tables 1 and 2 and Figure 3 reveal that, under these conditions and in this sequence context, there is no favorable Z•TA triplet. Thus the specific recognition of the TA base pair remains a challenge for design and synthesis.

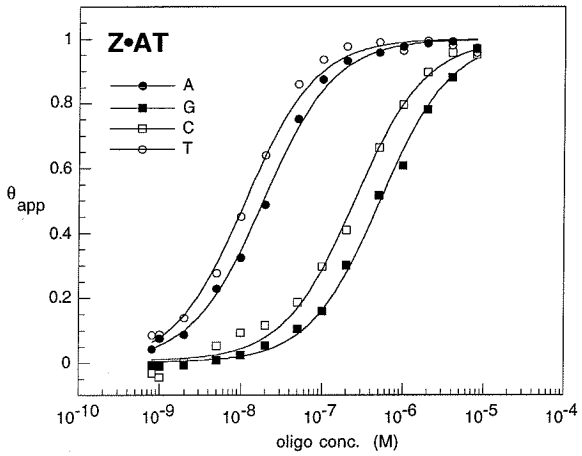
Conclusion. These results demonstrate that the G•GC, T•AT, and A•AT base triplets are significantly more stable than the thirteen other triplets studied. The T•CG and A•GC possess stabilities between that of G•GC, T•AT, and A•AT, and the worst mismatches. Importantly, the G•GC, T•AT, and A•AT interactions are specific. Oligonucleotides containing G and A or T are capable of recognizing target sites composed of Watson - Crick AT and GC base pairs in a specific manner; the energetic penalty for such an oligonucleotide binding to a site with a single mismatch is in the range of 1.2 - 2.1 kcal mol^{-1}. In addition, the relatively stable T•CG triplet offers a means of recognizing CG base pairs, however, because of the degenerate recognition properties of a third strand T,

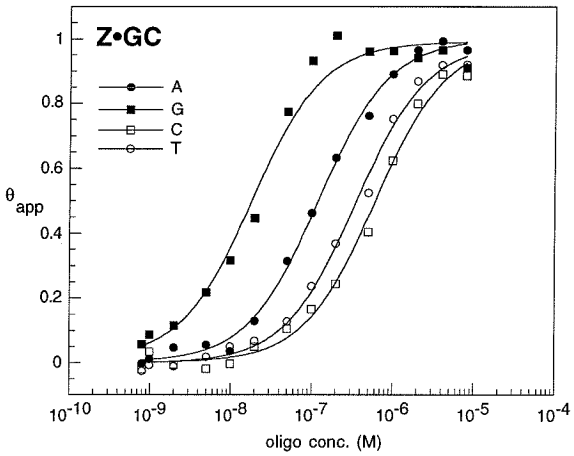
the overall specificity of the oligonucleotide would be compromised. We emphasize that our results are for one sequence composition and set of conditions and that the dependence of the energetics of triple helix formation on sequence composition and salt conditions remain to be elucidated.

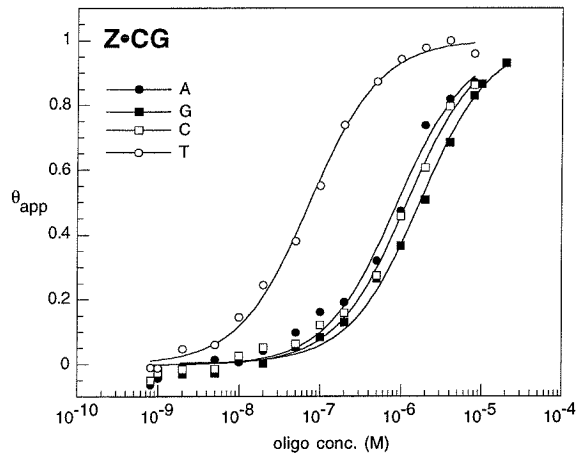
This work represents an important step in quantitatively characterizing the purine•purine•pyrimidine motif, which is less well studied than the pyrimidine•purine•pyrimidine motif. The values reported here will serve as a basis with which to directly compare and evaluate novel base designs within this motif.

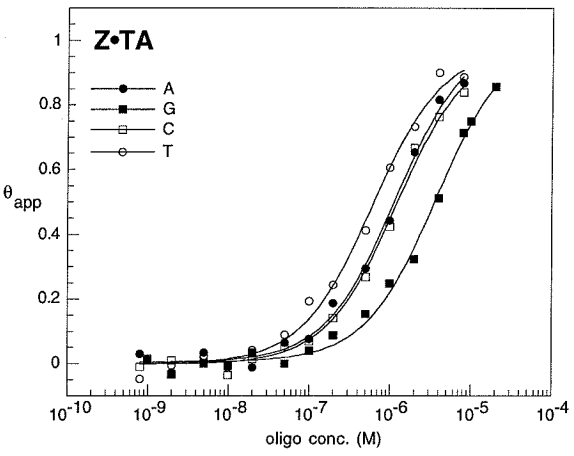
Figure 3. Binding isotherms for the sixteen Z•XY triplets studied, depicted in the form Z•AT, Z•GC, Z•CG, and Z•TA. Each isotherm represents the average of three experiments conducted at 37°C, 50 mM tris acetate, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$.

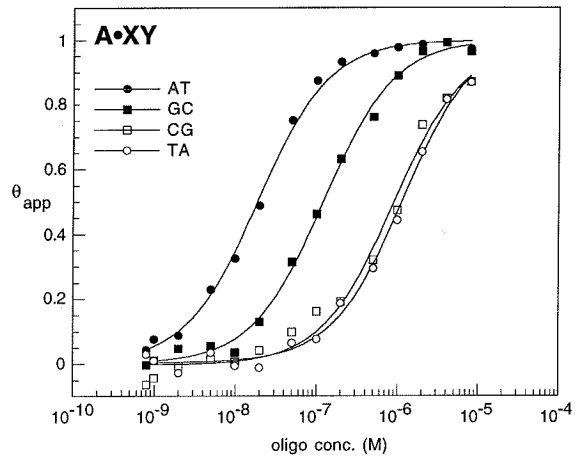
Figure 4. The same sixteen binding isotherms, depicted in the form A•XY, G•XY, C•XY, and T•XY.

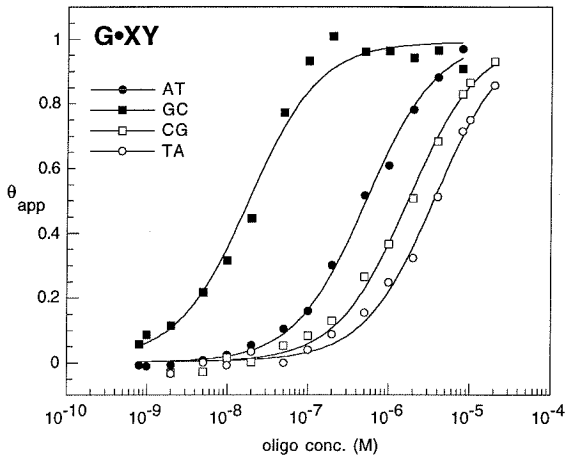

Z•AT
A
G
C
T
θapp
oligo conc. (M)

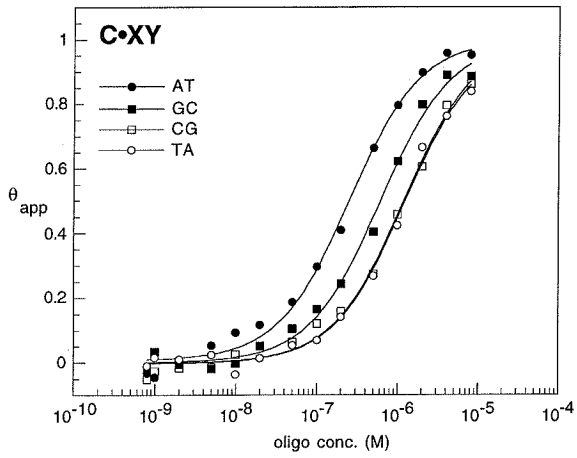

Z•GC
A
G
C
T
θapp
oligo conc. (M)


Z•CG
A
G
C
T
θapp
oligo conc. (M)


Z•TA
A
G
C
T
θapp
oligo conc. (M)


A•XY
AT
GC
CG
TA
θapp
oligo conc. (M)


G•XY
AT
GC
CG
TA
θapp
oligo conc. (M)


C•XY
AT
GC
CG
TA
θapp
oligo conc. (M)

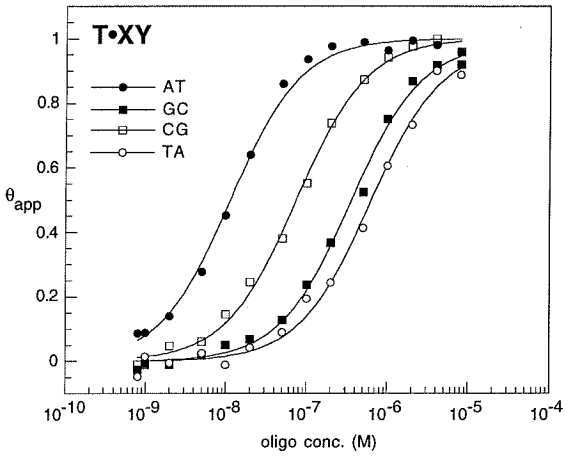
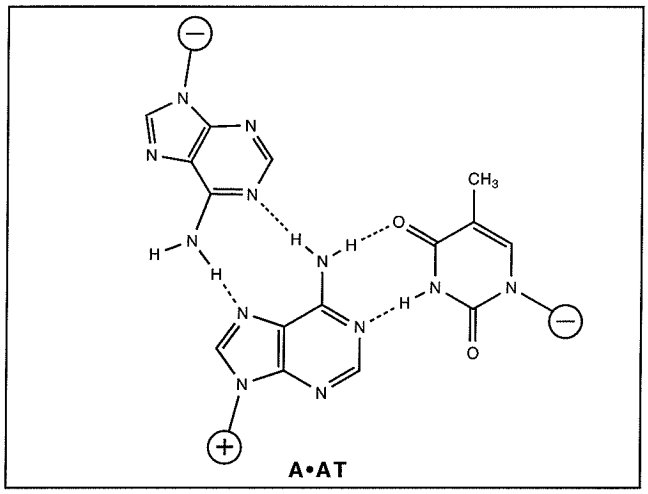
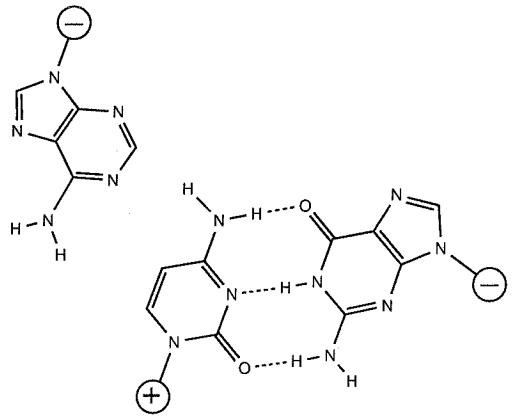

T•XY
AT
GC
CG
TA
θapp
oligo conc. (M)

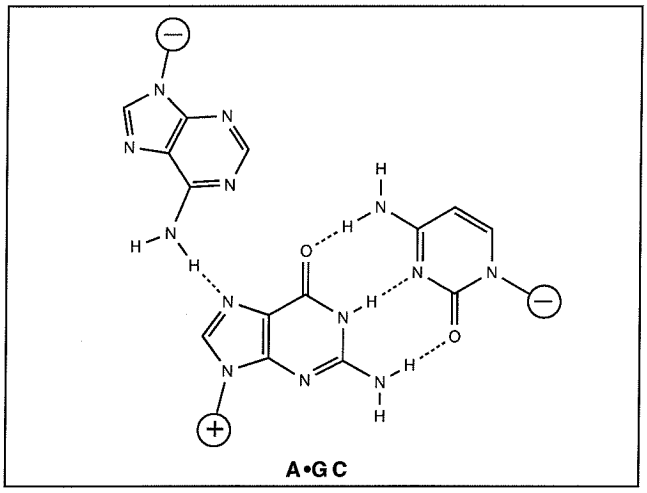
Figure 5. Structures of the A•XY (A), G•XY (B), C•XY (C), and T•XY (D) base triplets studied. The plus and minus signs indicate the relative strand polarity. The three strongest and two intermediate interactions are shown in boxes.



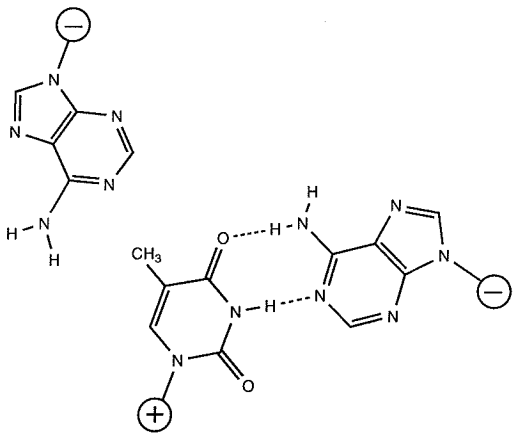
A•AT



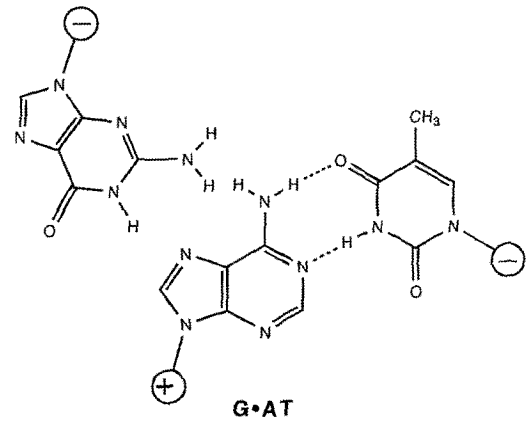
A•CG



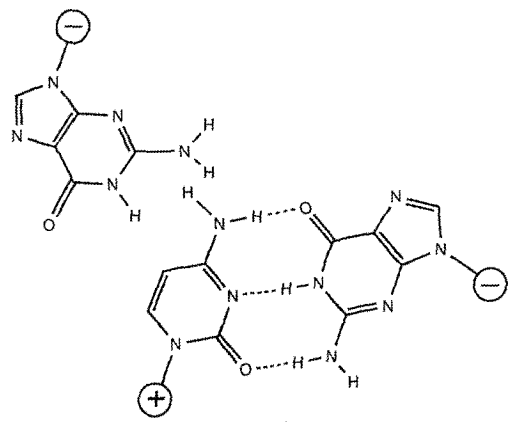
A•GC



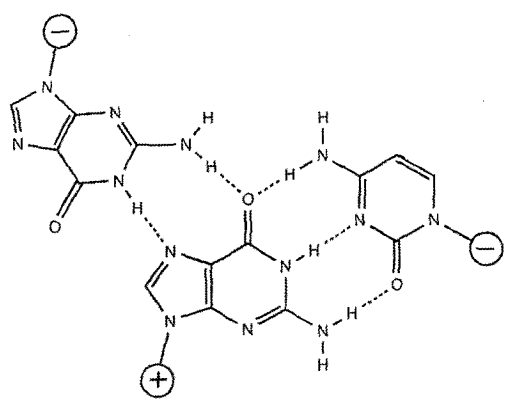
A•TA



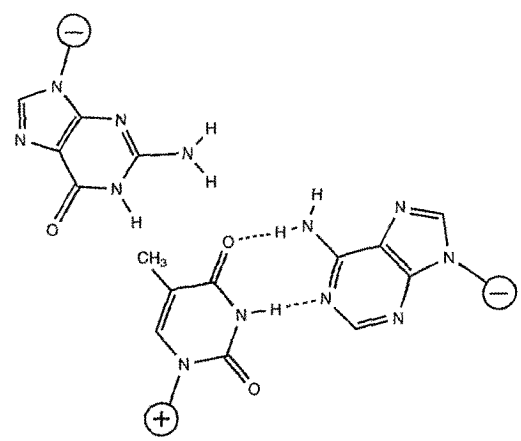
G•AT



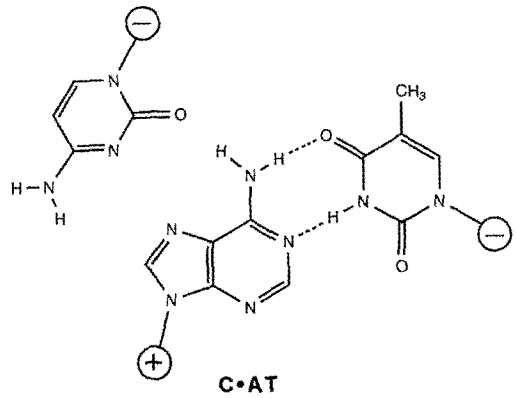
G•CG



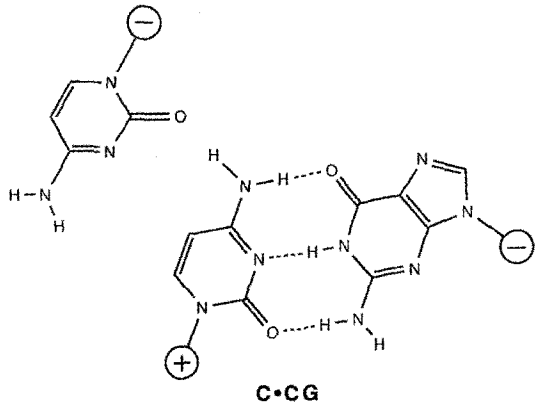
G•GC



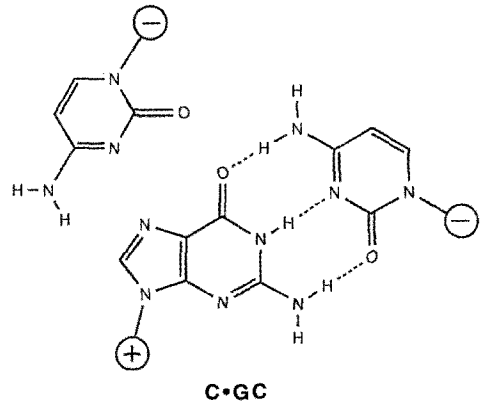
G•TA



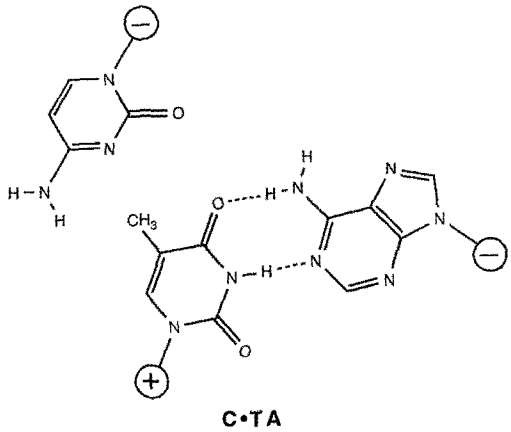
C•AT



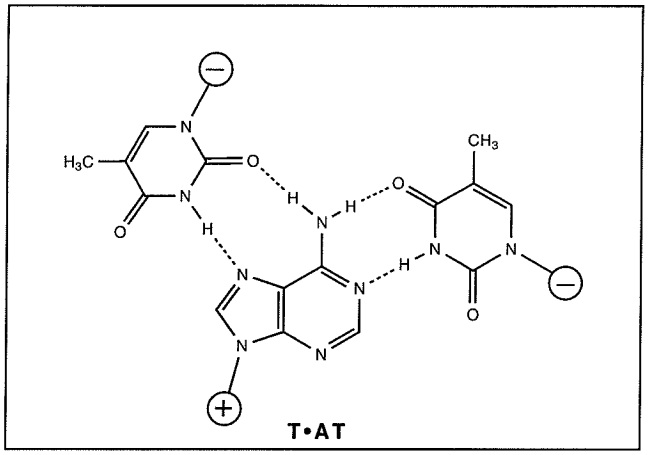
C•CG



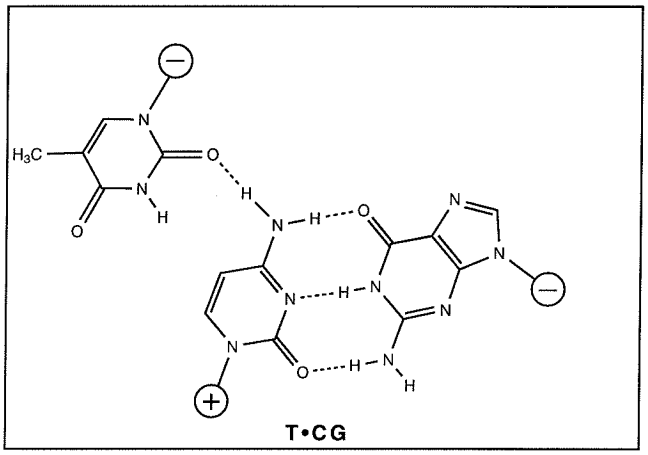
C•GC



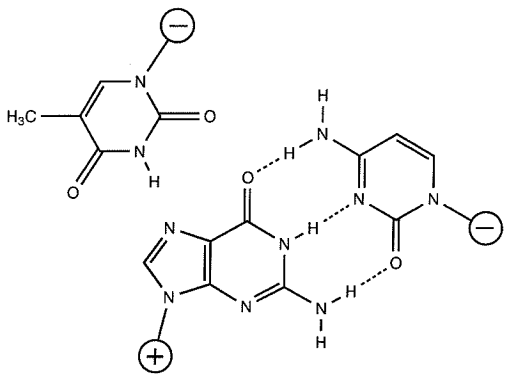
C•TA



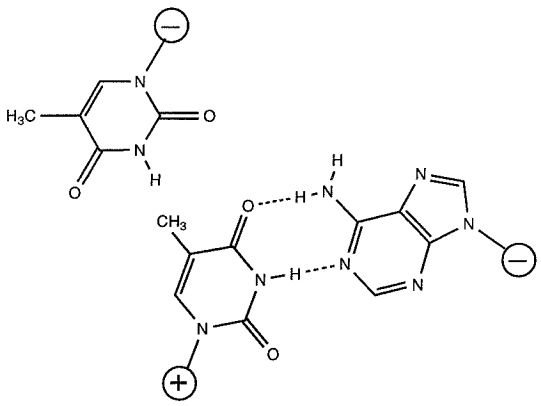
T•AT



T•CG



T•GC



T•TA

Experimental Section

General. Sonicated, phenol extracted calf thymus DNA (Pharmacia) was dissolved in H_2O to a final concentration of 1.0 mM in base pairs and was stored at 0°C. Glycogen was obtained from Boehringer Mannheim as a 20 mg/mL aqueous solution. Deoxynucleotide triphosphates were Pharmacia Ultra-Pure grade and were used as supplied. α-^{32}P nucleotide triphosphates (3000 Ci/mmol) were purchased from Amersham. Cerenkov radioactivity was measured with a Beckman LS 2801 scintillation counter. UV-visible spectroscopy was performed on a Hewlett-Packard 8452A diode array spectrophotometer. Restriction endonucleases were purchased from Boehringer Mannheim or New England Biolabs and were used according to the supplier's recommended protocol in the buffer provided. Sequenase version 2.0 was purchased from United States Biochemical. DNAse I was purchased from Pharmacia. Phosphoramidites were purchased from Glen Research. Tris (ultrapure) was purchased from Boehringer Mannheim. All other chemicals were of reagent grade or better and were used as supplied. General manipulations of duplex DNA[27] and oligonucleotides were performed according to established procedures.

Oligodeoxyribonucleotide synthesis. Oligodeoxyribonucleotides were synthesized by standard automated solid-support chemistry using an Applied Biosystems Model 380B DNA synthesizer and O-cyanoethyl-N,N-diisopropyl phosphoramidites. The final 5'-dimethoxytrityl group was removed, and deprotection was carried out in concentrated aqueous ammonia at 55°C for 24 hours. Crude oligodeoxyribonucleotide products were purified by ion exchange FPLC using a Mono Q 10/10 column (Pharmacia) and a gradient of 0-1 M LiCl in 10 mM LiOH, pH 12.0. Purified oligodeoxyribonucleotides were desalted by extensive dialysis against H_2O with a Spectra/Por MWCO 1000 membrane. The concentration of single-stranded oligodeoxyribonucleotides were determined by UV absorbance at 260 nm using extinction

coefficients calculated by the nearest-neighbor method from the monomer and dimer values.[28] Oligodeoxyribonucleotides were divided into 5 nmol aliquots, lyophilized to dryness, and stored at -20°C.

Plasmid preparation and 3' end labeling. Preparation of plasmid pPBAG19, where XY = AT, is described in reference 2b. Plasmids pPB19GC, pPB19CG, pPB19TA were prepared in analogous fashion. The procedure for preparing 3'-end labeled DNA was as follows: to 10 μg plasmid was added 60 units *Hind*III, 20 μL 10X *Hind*III reaction buffer, and sufficient water for a total volume of 200 μL. After 2 hours incubation at 37°C, the linearized plasmid was extracted with phenol twice and with 24:1 chloroform: isoamyl alcohol once, precipitated with 20 μL 3M NaOAc, pH 5.2 and 500 μL ethanol, and washed with 70 % ethanol. To the linearized plasmid was added 10 μL 5X sequenase buffer, 7 μL (70 μCi) of each α-^{32}P dATP, dCTP, dGTP, and TTP, and 2 μL sequenase. After 30 minutes reaction at room temperature, 2 μL of each cold dNTP (10 mM solution) was added. After 5 additional minutes, the reaction was quenched with 50 μL 100 mM EDTA. Unincorporated radioactivity was removed using a NICK column (Pharmacia) and the labeled, linearized plasmid was ethanol precipitated and washed with 70 % ethanol. After brief drying, the pellet was digested with 30 units *Ssp*I in a total volume of 200 μL for 2 hours at 37°C. After phenol extraction twice, chloroform extraction once, ethanol precipitation and 70 % ethanol wash, the labeled fragments were suspended in 20 μL water and 5 μL 15% ficoll nondenaturing loading buffer. The desired 648 bp fragment was isolated by preparative polyacrylamide gel electrophoresis (5 % nondenaturing gel, 1:29 crosslinking, 160 V, 2 hours) and visualized by autoradiography. The desired band was excised, crushed, and eluted overnight at 37°C into 1 mL 10 mM tris HCl, pH 8.0, 250 mM NaCl, 10 mM EDTA, 0.1 % SDS. The suspension was filtered and the eluted DNA precipitated by addition of 700 μL isopropanol. The pellet was resuspended in 100 μL 0.5X TE, phenol and chloroform extracted once each, and desalted on a NICK column.

The DNA was counted and stored at -20°C at a final concentration in water of 10,000 cpm per μL.

Quantitative DNAse I footprint titrations. DNAse I footprint experiments were performed essentially as described.[5,21,23] A stock solution of radiolabeled DNA in buffer was prepared from 312.5 μL 5X association buffer (250 mM tris acetate, 50 mM NaCl, 15 mM $MgCl_2$, pH 7.40 at 37°C), 156 μL calf thymus DNA (1.0 mM in base pairs), approximately 250,000 cpm 3'-end labeled target DNA and enough water to bring the total volume to 1.25 mL. The stock solution was distributed among 15 microcentrifuge tubes in 80 μL aliquots. A dried 5 nmol aliquot of oligonucleotide was dissolved in water to a concentration of 80 μM, heat denatured at 90°C for 4 minutes, and serially diluted. To each reaction tube was added 10 μL of oligonucleotide at the appropriate concentration. Oligonucleotide and target DNA were allowed to equilibrate at 37°C for 48 hours. Following equilibration, DNAse I, $CaCl_2$, and a nonspecific 34-mer oligonucleotide which was used to maintain uniform DNAse I reactivity were added and the digestion allowed to proceed for 6 minutes. Final reaction conditions in 100 μL solution were 50 mM tris acetate at pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 10 mM $CaCl_2$, 100 μM bps calf thymus DNA, 500 nM nonspecific oligonucleotide, approximately 15,000 cpm labeled duplex, and 0.1 milliunits/ μL DNAse I. Reactions were quenched by the addition of EDTA, glycogen, and NaCl to final concentrations of 25 mM, 80 μg/ mL, and 200 mM. The DNA was precipitated with 2.5 volumes of ethanol and isolated by ultracentrifugation. The precipitate was dissolved in 30 μL water, the solution frozen, and lyophilized. The DNA in each tube was resuspended in 5 μL of 80 % formamide - 1X TBE loading buffer and assayed for Cerenkov radioactivity by scintillation counting. The DNA was denatured at 90°C for 4 minutes, loaded on an 8 % denaturing polyacrylamide gel, and electrophoresed in TBE buffer at 2000V. After loading, residual radioactivity in the microcentrifuge tubes was counted. The gel was dried on a slab dryer and exposed to a storage phosphor screen.

Quantitation and data analysis. After 12 - 24 hours exposure in the dark, storage phosphor screens were scanned on a Molecular Dynamics 400S PhosphorImager. The data were analyzed by performing volume integrations of target and reference sites using ImageQuant software running on an AST Premium 386/33 computer. The analysis of DNAse I footprint titrations was performed according to the previously described method.[5,21,23] Briefly, the apparent fractional occupancy (θ_{app}) at each oligonucleotide concentration ($[O]_{tot}$) was determined using the following equation :

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

where I_{site} and I_{ref} are the digestion intensities in the target and reference sites, respectively, and I_{site}° and I_{ref}° are the digestion intensities at target and reference sites in a DNAse I control to which no oligonucleotide probe was added. The resulting pairs of (θ_{app}, $[O]_{tot}$) values were plotted on a semilog scale. The following binding isotherm was fit to the experimental data using a nonlinear least squares algorithm in Kaleidagraph 3.0.1 running on a Macintosh IIfx or IIci:

$$\theta_{app} = \theta_{min} + (\theta_{max} - \theta_{min}) \bullet \frac{K_T[O]_{tot}}{1 + K_T[O]_{tot}}$$

where θ_{min} is the apparent fractional occupancy at the lowest oligonucleotide concentrations, θ_{max} is the apparent fractional occupancy at saturation, and K_T is the equilibrium association constant. All data points from a gel were included in the fitting procedure unless visual inspection revealed a flaw in the gel at either target or reference sites, or the θ_{app} value for a single data point was more than two standard errors away from the data points on either side. Data from experiments for which less than 80 % of the lanes were usable were discarded. The goodness of fit of the binding curve to the data points was judged by the χ^2

criterion, and fits were judged acceptable for $\chi^2 \leq 1.5$. Correlation coefficients reported for acceptable fits were ≥ 0.95.

Repeat experiments for a particular triplet used different serial dilutions of oligonucleotide probe prepared from a different aliquot of the probe and different preparations of 3' -end labeled DNA. All K_T values reported in the text are the means of three experimental observations plus or minus the standard error of the mean.

References

(1) (a) Moser, H. E.; Dervan, P. B. *Science* **1987**, *238,* 645-650. (b) LeDoan, T.; Perrouault, L.; Praseuth, D.; Habhoub, N.; Decout, J.-L.; Thuong, N. T.; Lhomme, J.; Hélène, C. *Nucleic Acids Res.* **1987**, *15*, 7749-7760.

(2) (a) Cooney, M.; Czernuszewicz, G.; Postel, E. H.; Flint, S. J.; Hogan, M. E. *Science* **1988**, *241*, 456-459. (b) Beal, P. A.; Dervan, P. B. *Science* **1991**, *251*, 1360-1363. (c) Beal, P. A.; Dervan, P. B. *Nucleic Acids Res.* **1992**, *20*, 2773-2776. (d) Stilz, H. U.; Dervan, P. B. *Biochemistry* **1993**, *32*, 2177-2185. (e) Malkov, V. A.; Voloshin, O. N.; Soyfer, V. N.; Frank-Kamenetskii, M. D. *Nucleic Acids Res.* **1993**, *21*, 585-591.

(3) For early work on three stranded RNA polymers, see: Felsenfeld, G.; Davies, D.; Rich, A. *J. Am. Chem. Soc.* **1957**, *79*, 2023-2024.

(4) For alternate strand binding, see: (a) Jayasena, S. D.; Johnston, B. H. *Biochemistry* **1992**, *31*, 320-327. (b) Beal, P. A.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 4976-4982. (c) Jayasena, S. D.; Johnston, B. H. *Nucl. Acids Res.* **1992**, *20*, 5279-5288. (d) Jayasena, S. D.; Johnston, B. H. *Biochemistry* **1993**, *32*, 2800-2807.

(5) Singleton, S. F.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 6957-6965.

(6) (a) Mergny, J.-L.; Sun, J.-S.; Rougée, M.; Montenay-Garestier, T. Barcelo, F.; Chomilier, J.; Hélène, C. *Biochemistry* **1991**, *30*, 9791-9798. (b) Rougée, M.; Faucon, B.; Mergny, J. L.; Barcelo, F.; Giovannageli, C.; Garestier, T.; Hélène, C. *Biochemistry* **1992**, *31*, 9269-9278.

(7) (a) Povsic, T. J.; Dervan, P. B. *J. Am. Chem. Soc.* **1989**, *111*, 3059-3061. (b) Xodo, L. E.; Manzini, G.; Quadrifoglio, F.; van der Marel, G. A.; van Boom, J. H. *Nucleic Acids Res.* **1991**, *19*, 5625-5631. (c) Singleton, S. F.; Dervan, P. B. *Biochemistry* **1992**, *31*, 10995-11003.

(8) Singleton, S. F.; Dervan, P. B. *Biochemistry* **1993**, *32*, 13171-13179.

(9) Singleton, S. F.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 10376-10382.

(10) (a) Roberts, R. W.; Crothers, D. M. *Science* **1992**, *258*, 1463-1466. (b) Han, H.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 3806-3810. (c) Escude, C.; François, J. C.; Sun, J. S.; Ott, G.; Sprinzl, M.; Garestier, T.; Hélène, C. *Nucleic Acids Res.* **1993**, *21*, 5547-5553. (d) Han, H.; Dervan, P. B. *Nucleic Acids Res.* **1994**, *22*, 2837-2844.

(11) (a) Maher, L. J.; Dervan, P. B.; Wold, B. *Science* **1989**, *245*, 725-730. (b) Collier, D. A.; Thuong, N. T.; Hélène, C. *J. Am. Chem. Soc.* **1991**, *113*, 1457-1458. (c) Maher, L. J.; Dervan, P. B.; Wold, B. *Biochemistry* **1992**, *31*, 70-81. (d) Grigoriev, M.; Praseuth, D.; Robin, P.; Hemar, A.; Saison-Behmoaras, T.; Dautry-Varsat, A.; Thuong, N. T.; Hélène, C.; Harel-Bellan, A. *J. Biol. Chem.* **1992**, *267*, 3389-3395.

(12) Strobel, S. A.; Doucette-Stamm, L. A.; Riba, L.; Housman, D. E.; Dervan, P. B. *Science* **1991**, *254*, 1639-1642.

(13) Horne, D.A.; Dervan, P. B. *J. Am. Chem. Soc.* **1990**, *112*, 2435-2437.

(14) (a) Griffin, L. C.; Dervan, P. B. *Science* **1989**, *245*, 967-971. (b) Shimizu, M.; Inoue, H.; Ohtsuka, E. *Biochemistry* **1994**, *33*, 606-613.

(15) (a) Mergny, J.-L.; Sun, J.-S.; Rougée, M.; Montenay-Garestier, T. Barcelo, F.; Chomilier, J.; Hélène, C. *Biochemistry* **1991**, *30*, 9791-9798. (b) Miller, P. S.; Cushman, C. D.; *Biochemistry* **1993**, *32*, 2999. (c) Fossella, J. A.; Kim, Y. J.; Shih, H.; Richards, E. G.; Fresco, J. R. *Nucleic Acids Res.* **1993**, *21*, 4511-4515.

(16) Yoon, K.; Hobbs, C. A.; Koch, J.; Sardaro, M.; Kutny, R.; Weis, A. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 3840-3844.

(17) (a) Belotserkovskii, B. P.; Veselkov, A. G.; Filippov, S. A.; Dobrynin, V. N.; Mirkin, S. M.; Frank-Kamenetskii, M. D. *Nucleic Acids Res.* **1990**, *18*, 6621-6624.

(b) Xodo, L. E.; Alunni-Fabbroni, M.; Manzini, G.; Quadrifoglio, F. *Eur. J. Biochem.* **1993**, *212*, 395-401.

(18) (a) de los Santos, C.; Rosen, M.; Patel, D. J. *Biochemistry* **1989**, *28*, 7282-7289. (b) Rajagopal, P.; Feigon, J. *Nature* **1989**, *339*, 637-640. (c) Rajagopal, P.; Feigon, J. *Biochemistry* **1989**, *28*, 7859-7870.

(19) Live, D. H.; Radhakrishnan, I.; Misra, V.; Patel, D. J. *J. Am. Chem. Soc.* **1991**, *113*, 4687-4688.

(20) (a) Radhakrishnan, I.; Patel, D.J. *J. Am. Chem. Soc.* **1993**, *115*, 1615-1617. (b) Radhakrishnan, I.; de los Santos, C.; Patel, D.J. *J. Mol. Biol* **1993**, *234*, 188-197. (c) Radhakrishnan, I.; Patel, D.J. *Structure* **1993**, *1*, 135-152.

(21) Priestley, E.S.; Dervan, P.B. *J. Am. Chem. Soc.* **1995**, *117*, 4761-4765.

(22) Best, G.C.; Dervan, P.B. *J. Am. Chem. Soc.* **1995**, *117*, 1187-1193.

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

(24) (a) Stankus, A.; Goodisman, J.; Dabrowiak, J.C. *Biochemistry* **1992**, *31*, 9310-9318. (b) Mrksich, M.; Parks, M.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1994**, *116*, 7983-7988.

(25) (a) Dittrich, K.; Gu, J.; Tinder, R.; Hogan, M.; Gao, X. *Biochemistry* **1994**, *33*, 4111-4120. (b) Durland, R.H.; Rao, T.S.; Revankar, G.R.; Tinsley, J.H.; Myrick, M.A.; Seth, D.M.; Rayford, J.; Singh, P.; Jayaraman, K. *Nucleic Acids Res.* **1994**, *22*, 3233-3240.

(26) Malkov, V.A.; Voloshin, O.N.; Veselkov, A.G.; Rostapshov, V.M.; Jansen, I.; Soyfer, V.N.; Frank-Kamenetskii, M.D. *Nucleic Acids Res.* **1993**, *21*, 105-111.

(27) Sambrook, J.; Fritsch, E.F.; Maniatis, T. *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press: Cold Spring Harbor, NY, 1989.

(28) Cantor, C.R.; Warshaw, M.M. *Biopolymers* **1970**, *9*, 1059-1077.

CHAPTER THREE

The Effect of Atomic Substitutions on the Energetics of Formation of Triple Helical Complexes in the Purine Motif

Introduction

The previous chapter described DNase I footprinting experiments which demonstrated in a quantitative manner that oligonucleotide-directed triple helix formation in the purine motif was highly sensitive to single base mismatches. Substitution of a single G•GC, A•AT, or T•AT match within a 15-base triplet with any other triplet composed of the natural DNA bases resulted in a destabilization of 1.2-2.4 kcal/mol under a standard set of conditions.[1] We desired to find out whether this motif was sensitive to single *atomic* substitutions, by determining the energetics of triple helix formation of oligonucleotides containing single substitutions of nonnatural purines. It was hoped that in exploring new structures we might find new specificities, in particular for TA and CG base pairs, which are not specifically recognized by any of the four natural bases. A series of purines were incorporated at a single common position into the same oligonucleotide sequence used in previous experiments so that direct comparison could be made with data presented in chapter 2.

The structures of the purines that were investigated are shown in Figure 1. They all represent substitutions or permutations on the structure of guanine. Inosine (I) represents substitution of the guanine 2-amino group with hydrogen, 2-aminopurine (AP) represents substitution of the guanine 6-carbonyl, and nebularine (N) represents substitution of both exocyclic functionalities with hydrogen. In the case of nebularine, this study represents a reinvestigation of a system previously reported on by this group,[2] in a new sequence context and under new conditions, and by a different technique (quantitative DNase I footprint titration rather than affinity cleavage titration). In addition, isoinosine (isoI), a permutation of inosine in which the carbonyl is at the 2- position rather than the 6- position, was investigated.

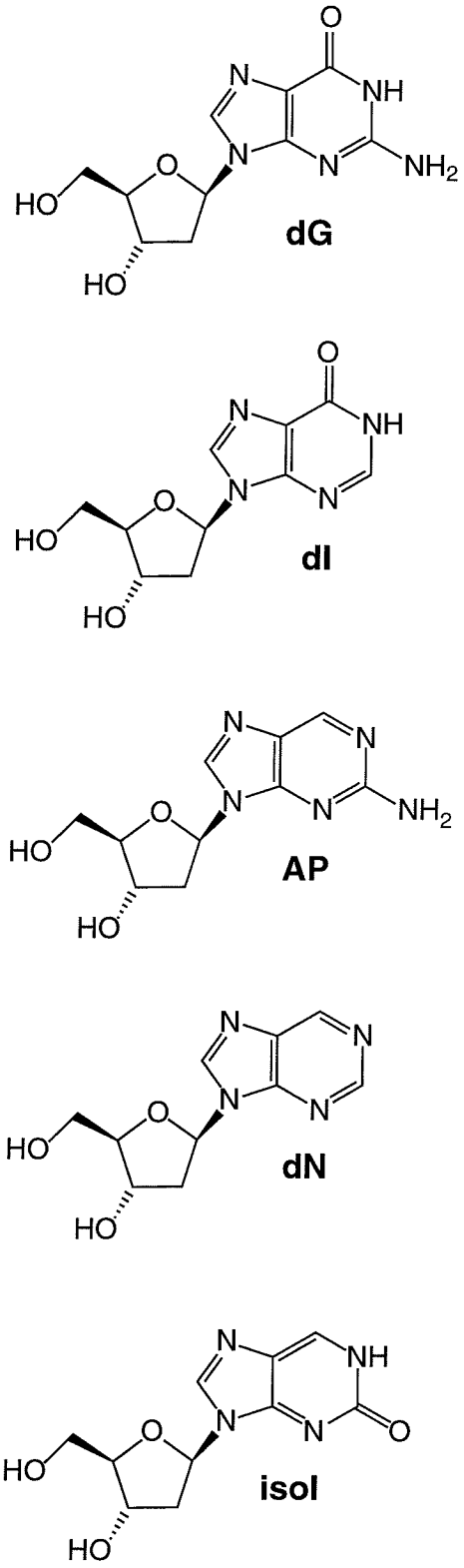


Figure 1. Atomic substitutions. The structures of guanine (dG), inosine (dI), aminopurine (AP), nebularine (dN), and isoinosine (isoI) 2'-deoxynucleosides.

A plasmid was constructed in which all four XY target sequences described in chapter 2 were placed within a single restriction fragment (Figure 2). Use of this system, which is based on systems designed for studying pyrimidine triple helix energetics,[3,4] allows all four Z•XY interactions to be measured in the same titration. In addition to reducing by a factor of four the number of titrations which need be performed, this ensures that all measurements are done under identical conditions, thus making measured association constants more internally consistent. Use of this four-site plasmid in the investigation of energetics of triplex formation for several other nonnatural bases will be described in subsequent chapters.

Results and Discussion

Methods. Quantitative DNase I footprint titration experiments were performed identically to those described in chapter 2, except that a 3'-^{32}P end labeled 314 bp restriction fragment containing all four purine rich 15 base pair target sites 5'-d(AGGGGAGGGG**X**GGGA)-3' where X= A, G, C, or T was equilibrated with a range of concentrations of purine-rich oligonucleotides 5'-d(TGGG**Z**GGGGTGGGGT)-3', where Z= I, AP, N, or isoI (Figure 2), rather than fragments containing a single target site. The four sites on the restriction fragment were separated by a common 15 base pair spacer sequence, such that all sites were abutted by identical 5'- and 3'- flanking sequences. This design ensures that the system suffers from minimal microheterogeneity of duplex structure based on differing sequence, and that the energetics which are measured most accurately reflect the energetics of individual triplet interactions.

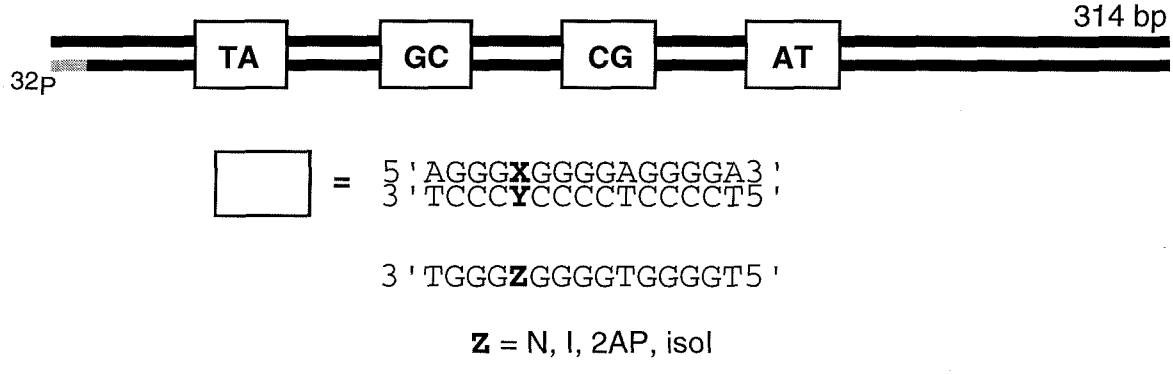


Figure 2. Sequences of the four target sites on 3'-labeled 314 bp *Afl* II/*Fsp* I restriction fragment of pRSPEC1. Below are the sequences of third strand oligonucleotides for probing the effect of atomic substitutions on energetics of triple helix formation.

Synthesis. 5'-Dimethoxytrityl phosphoramidites are available for 2'-deoxyinosine, 2-aminopurine (with the amino group protected as the dimethylformamidine), and nebularine for incorporation into oligonucleotides. The corresponding phosphoramidite for 2'-deoxyisoinosine (**9**) was prepared in an eight step sequence from 6-thioguanosine (**1**) according to the procedure outlined in Figure 3. Reduction of the thiol produced 2-aminopurine riboside, which was subjected to Robins' 4-step 2'-deoxygenation procedure.[5] This entailed selective protection of the 3' and 5' hydroxyls as the cyclic disiloxane **3** , followed by formation of 2'-phenylthiocarbonate **4**, which was reduced with tributyltin hydride to **5**. Deprotection with fluoride furnished 2-aminopurine-2'-deoxyriboside. This was diazotized with nitrous acid, quenching with water produced isoinosine 2'-deoxyriboside **7**. 5'-dimethoxytritylation and 3' phosphitylation furnished the desired phosphoramidite **9**.

Oligonucleotides were synthesized as described in chapter 2. The integrity of incorporated nonnatural bases was confirmed by phosphodiesterase/phosphatase digestion and HPLC analysis of resultant nucleosides with comparison to authentic standards, as well as by MALDI-TOF mass spectrometry of intact oligonucleotides (data shown in Experimental section).

Affinity and specificity. Equilibrium association constants for triple helical complexes containing G, I, AP, N, and isoI substitutions are shown in Table I, and corresponding ΔG values in Table II. It is instructive to correlate these affinities with the putative triplet structures drawn in Figures 4-8. Affinity and specificity of individual substitutions are also illustrated by footprint titrations (Figures 9-12).

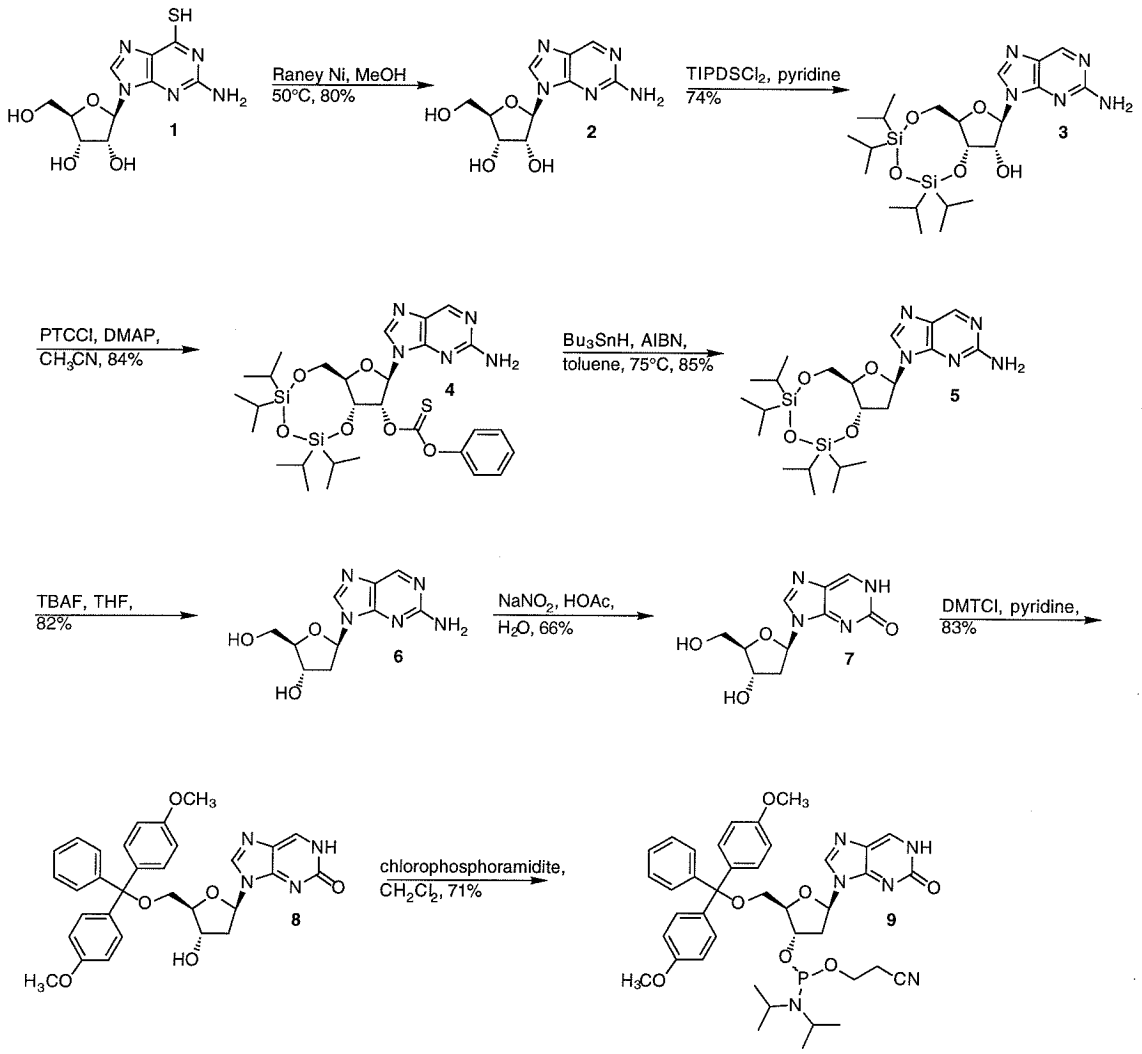


Figure 3. Synthesis of isoinosine phosphoramidite **9**.

Table 1. Association constants (K_T) for the formation of twenty triple helical complexes containing the Z•XY triplets at 37°C, 10 mM NaCl, 3 mM $MgCl_2$, 50 mM tris acetate, pH 7.4.[a,b]

XY	Z=G	I	AP	N	isoI
AT	4.6 (±0.7) x 10^5	2.9 (±0.3) x 10^6	9.8 (±0.9) x 10^5	2.2 (±0.3) x 10^6	2.8 (±0.6) x 10^7
GC	6.3 (±0.5) x 10^7	7.1 (±0.3) x 10^7	9.3 (±0.8) x 10^6	1.3 (±0.2) x 10^6	9.3 (±1.7) x 10^7
CG	3.2 (±1.4) x 10^5	4.6 (±0.5) x 10^5	9.2 (±1.2) x 10^5	1.0 (±0.2) x 10^6	3.5 (±0.6) x 10^5
TA	< 2 x 10^5	9.5 (±2.3) x 10^5	1.7 (±0.3) x 10^6	1.8 (±0.4) x 10^6	2.4 (±0.9) x 10^6

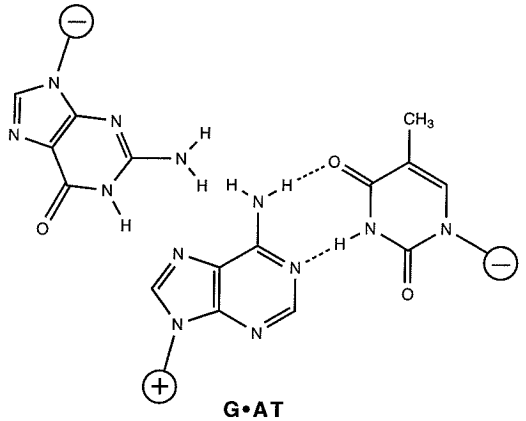
[a] K_T values are reported as the mean (± the standard error of the mean) of three measurements. The K_T values are reported in units of M^{-1}. [b] The identity of the base Z is indicated across the top of the columns; the identity of the Watson Crick base pair XY is indicated on the left side of the rows.

Table 2. Free energy (ΔG°) for the formation of twenty triple helical complexes containing the Z•XY triplets at 37°C, 10 mM NaCl, 3 mM $MgCl_2$, 50 mM tris acetate, pH 7.4.[a,b]

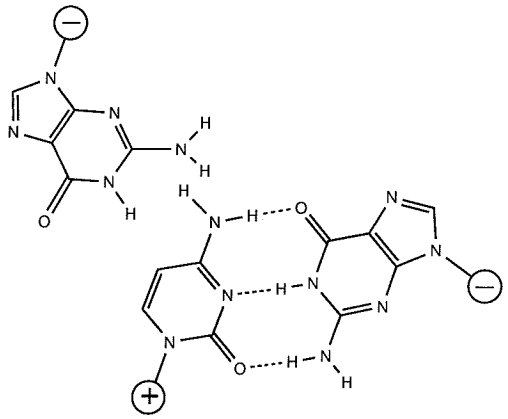
	G	I	AP	N	isoI
AT	-8.0±0.1	-9.2±0.1	-8.5±0.1	-9.0±0.1	-10.6±0.2
GC	-11.1±0.1	-11.1±0.1	-9.9±0.1	-8.7±0.1	-11.3±0.1
CG	-7.8±0.2	-8.0±0.1	-8.5±0.1	-8.5±0.1	-7.9±0.2
TA	>-7.5	-8.5±0.2	-8.8±0.1	-8.9±0.1	-9.1±0.3

[a] Free energy values are calculated from the measured association constants at 37°C and are reported in kcal mol^{-1}. [b] The identity of the base in the third strand, Z, is indicated across the top row; the identity of the Watson Crick base pair, XY, is indicated to the left of the rows.

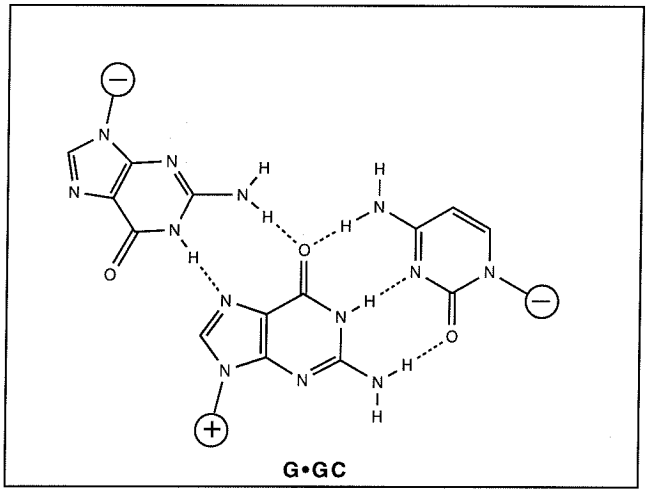
Figures 4-8. Putative triplet structures for G•XY (Figure 4), I•XY (Figure 5), AP•XY (Figure 6), N•XY (Figure 7) and isoI•XY (Figure 8). Boxes are drawn for triplets which have been determined to be high-affinity interactions.



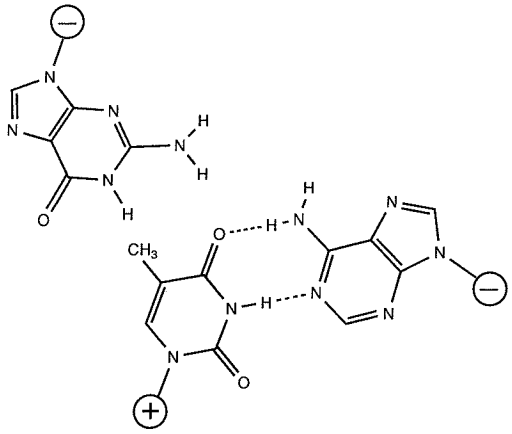
G•AT



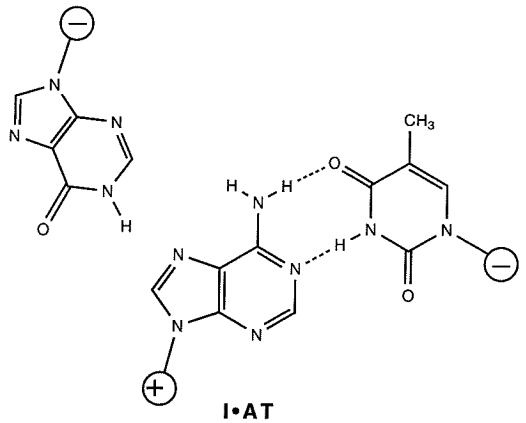
G•CG



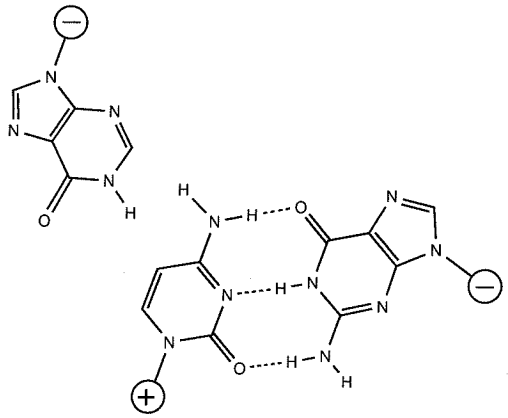
G•GC



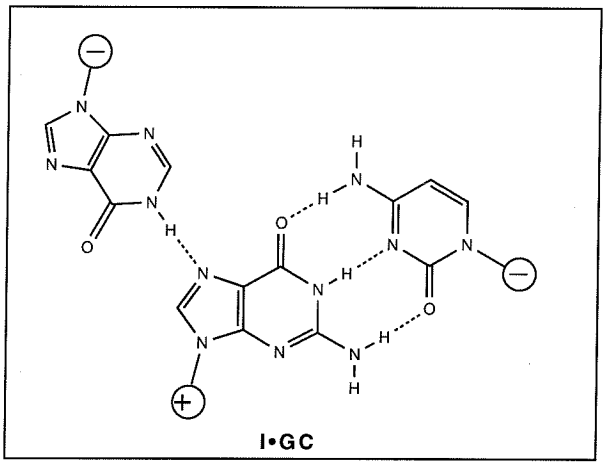
G•TA



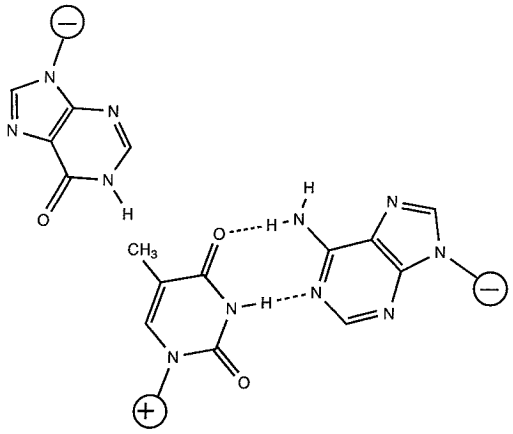
I•AT



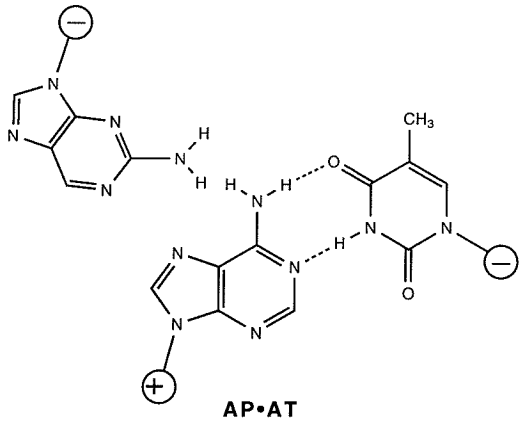
I•CG



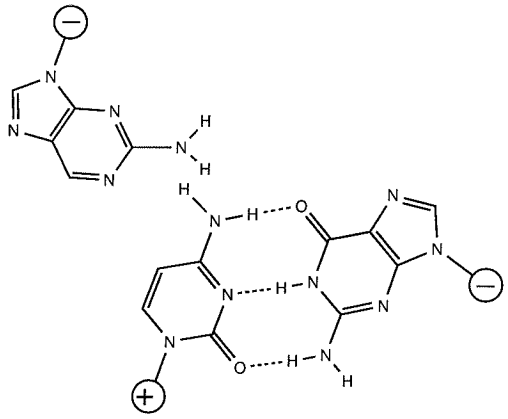
I•GC



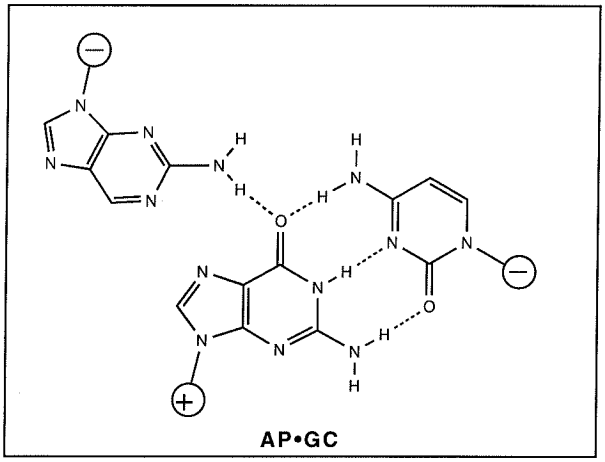
I•TA



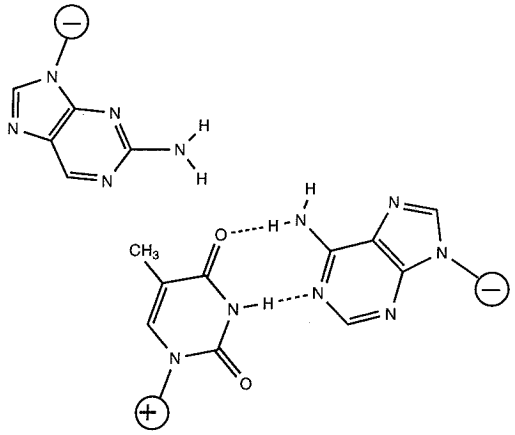
AP•AT



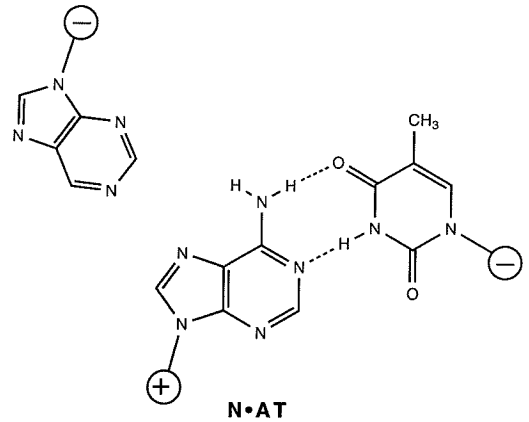
AP•CG



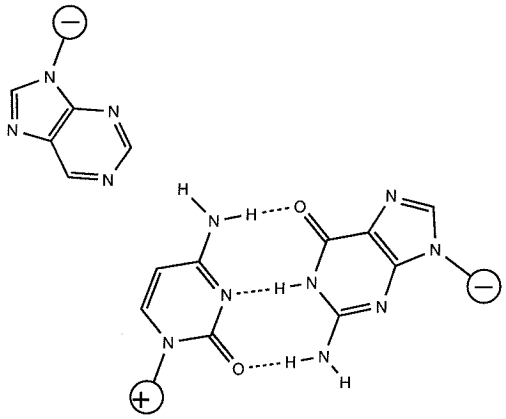
AP•GC



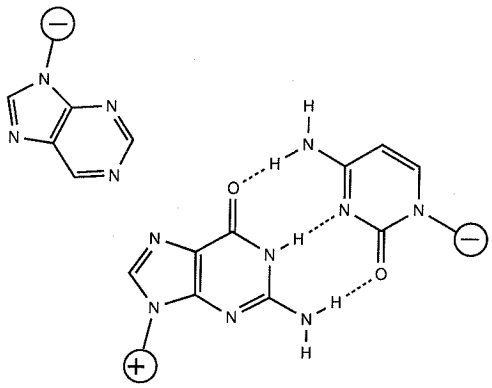
AP•TA



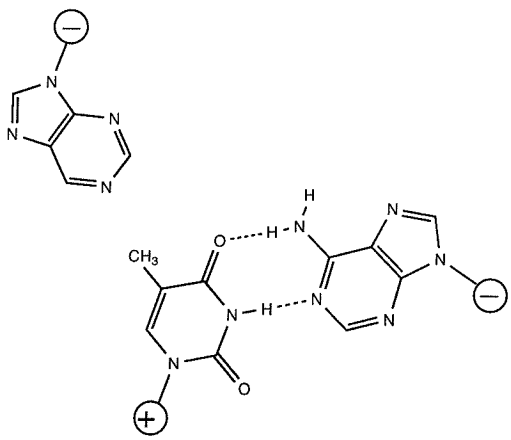
N•AT



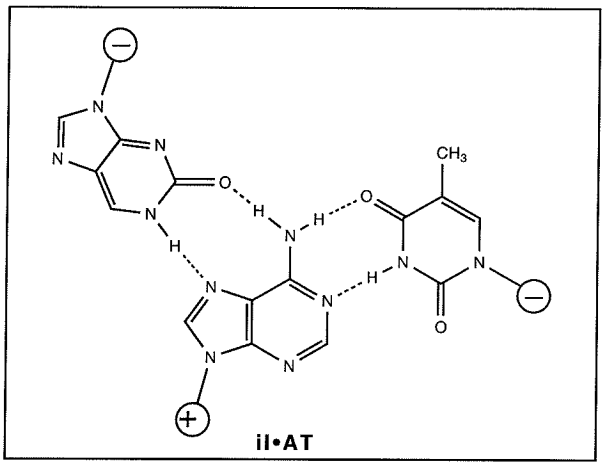
N•CG



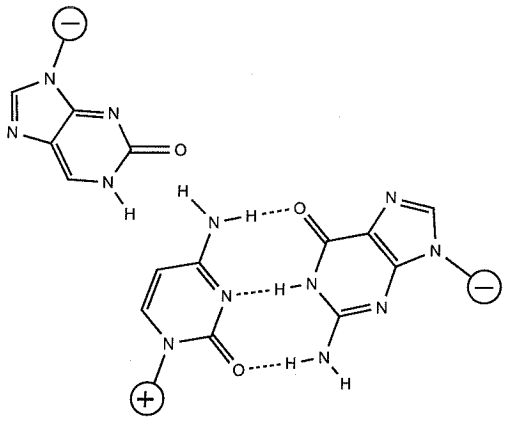
N•GC



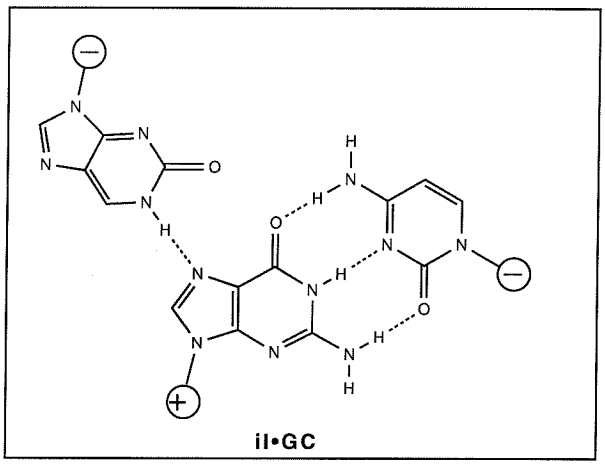
N•TA



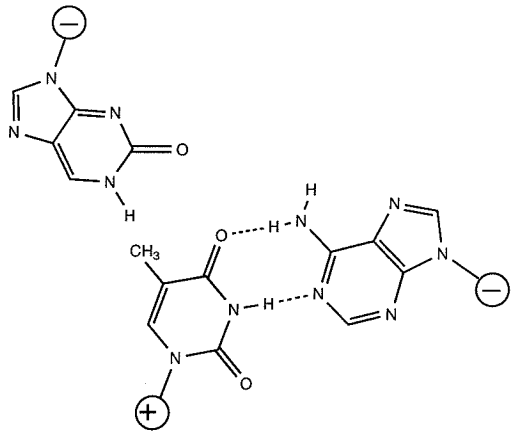
iI•AT



iI•CG

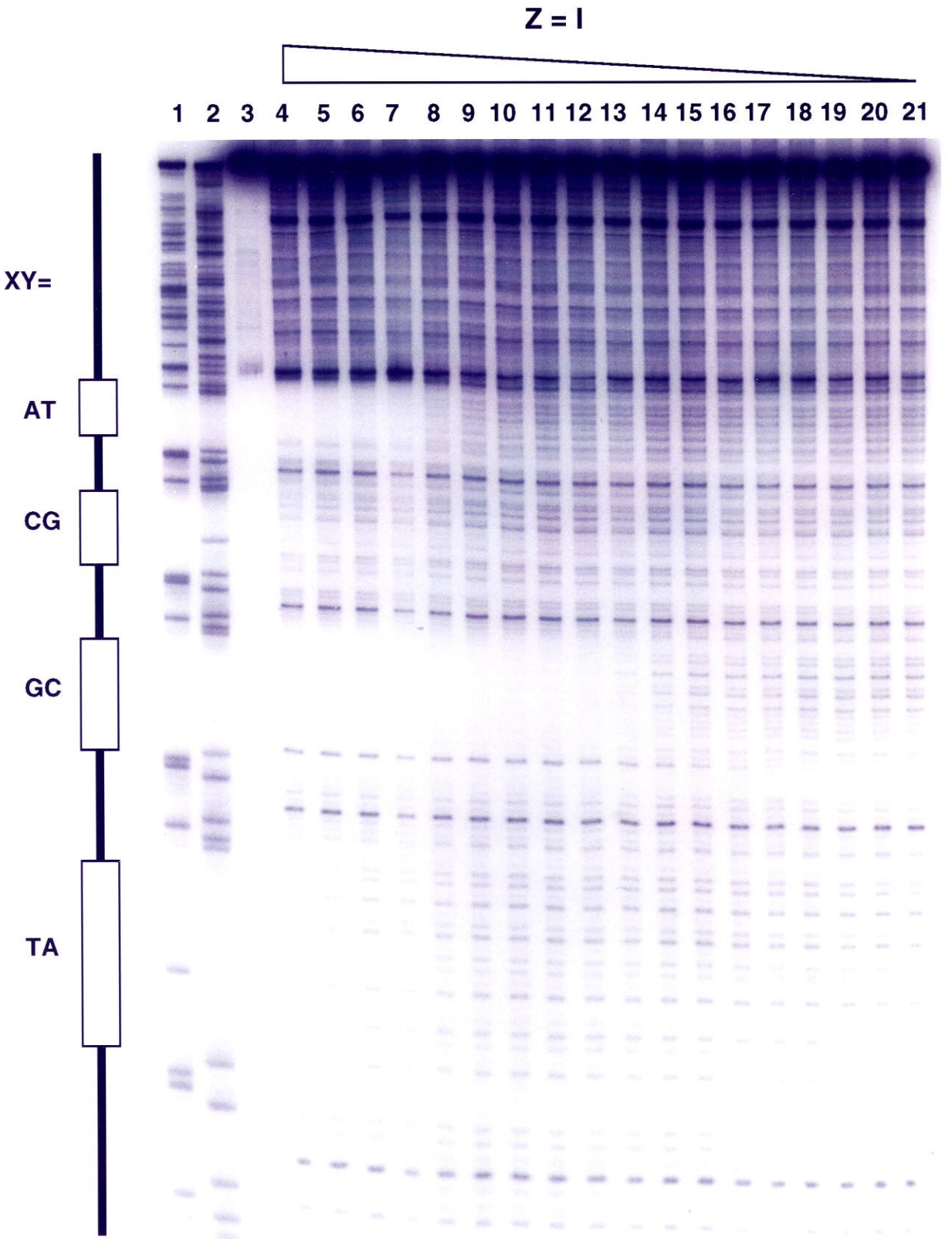


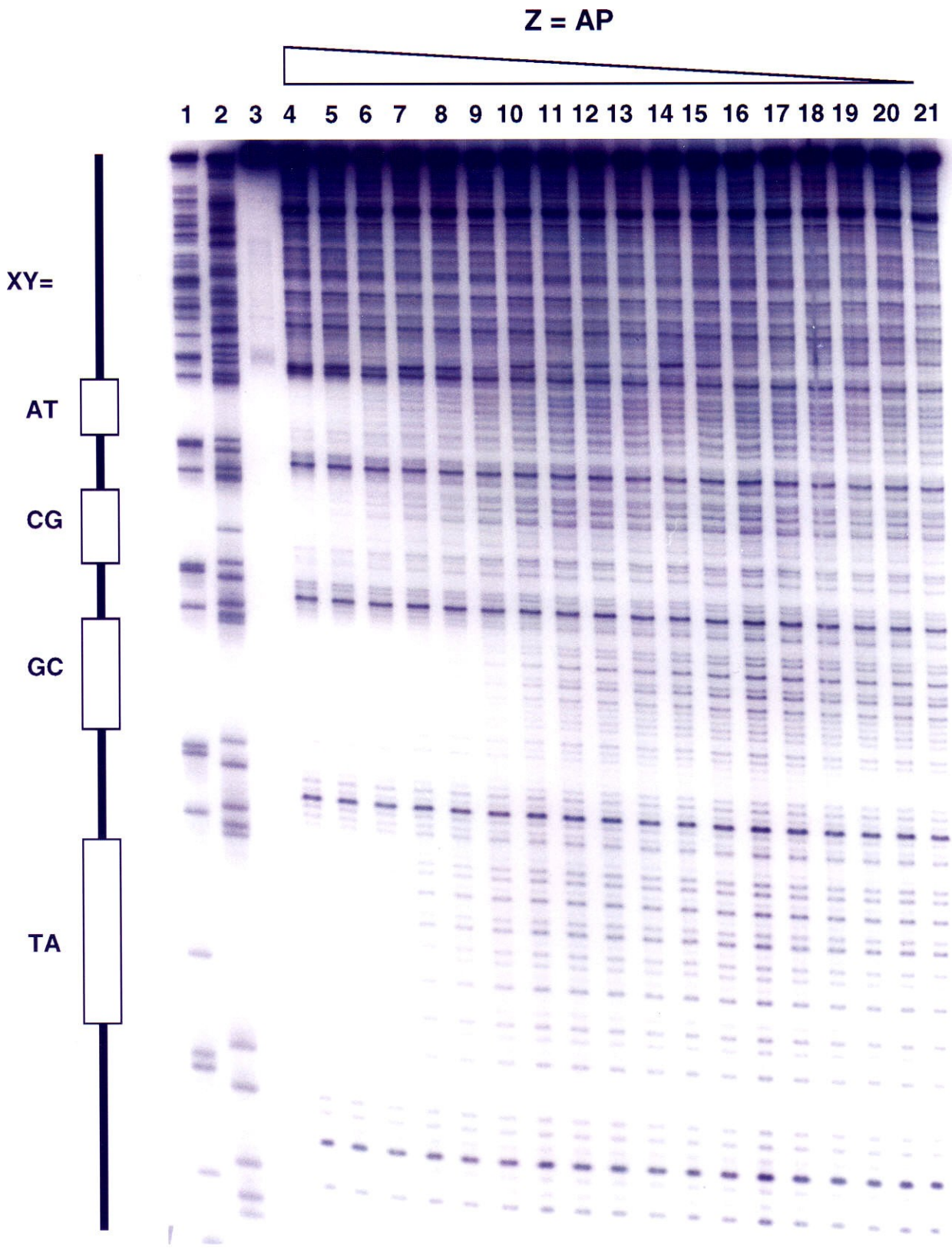
iI•GC

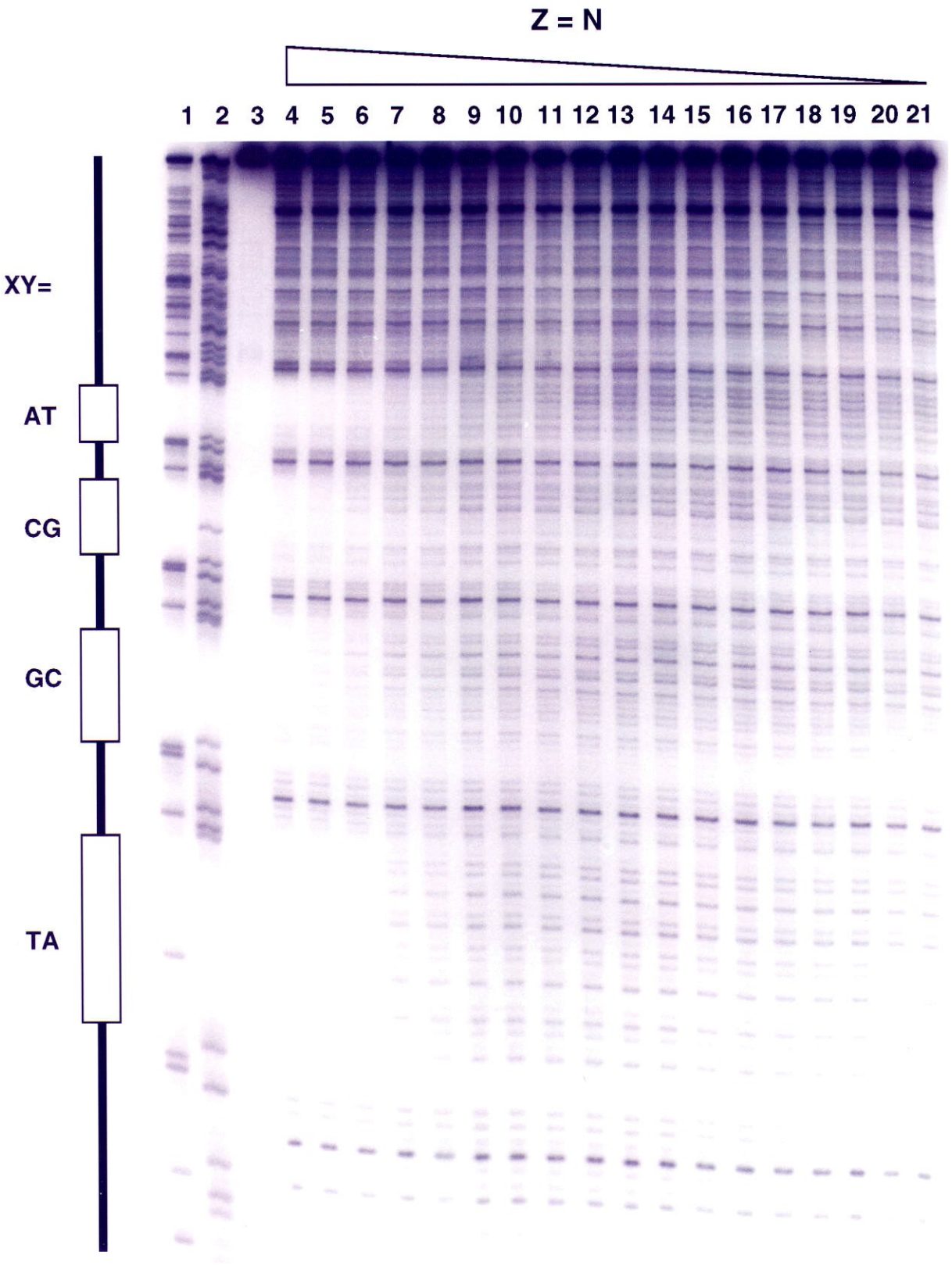


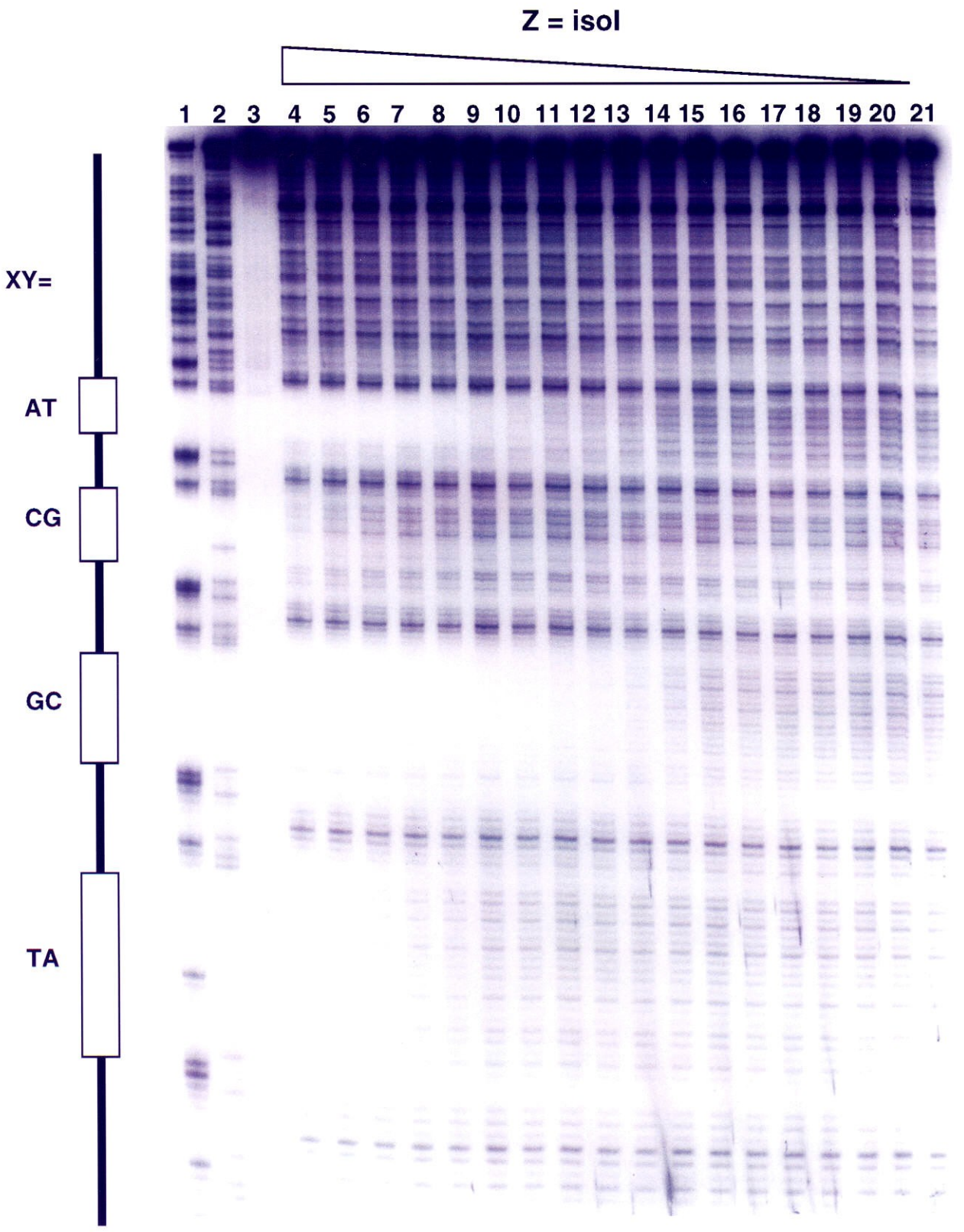
iI•TA

Figures 9-12. Autoradiograms of 8% denaturing polyacrylamide gels used to separate fragments generated by DNase I digestion in quantitative footprint titration experiments with oligonucleotides containing I (Figure 9), AP (Figure 10), N (Figure 11), and isoI (Figure 12). The four target sites on the 314 bp 3'-end labeled restriction fragment are labeled as XY= AT, CG, GC, and TA respectively, going from the top to the bottom of each gel. (Lane 1) Products of an adenine-specific reaction. (Lane 2) Products of a guanine-specific reaction. (Lane 3) Intact 3' labeled DNA after incubation in the absence of third strand oligonucleotide. (Lanes 4-21) DNase I digestion products obtained in the presence of varying concentrations of oligonucleotide: 8 μM (lane 4); 4 μM (lane 5); 2 μM (lane 6); 1 μM (lane 7); 800 nM (lane 8); 400 nM (lane 9); 200 nM (lane 10); 100 nM (lane 11); 80 nM (lane 12); 40 nM (lane 13);20 nM (lane 14); 10 nM (lane 15); 8 nM (lane 16); 4 nM (lane 17); 2 nM (lane 18); 1 nM (lane 19); 800 pM (lane 20); no oligonucleotide (lane 21).


Z = I
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
XY=
AT
CG
GC
TA


Z = AP
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
XY=
AT
CG
GC
TA


Z = N
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
XY=
AT
CG
GC
TA

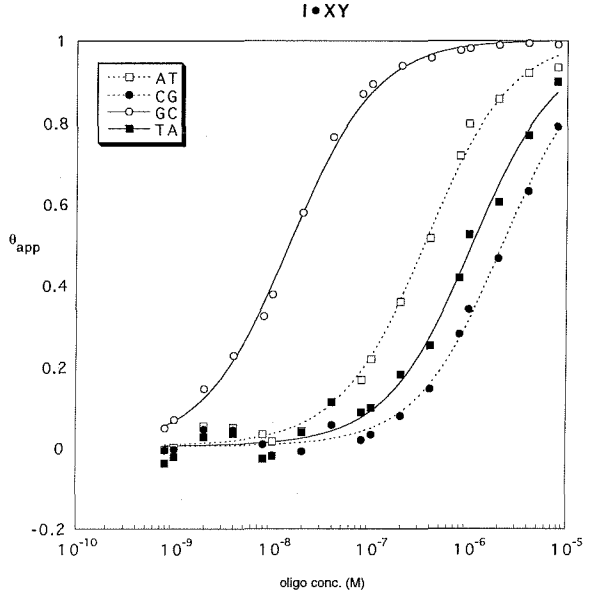

Z = isol
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
XY=
AT
CG
GC
TA

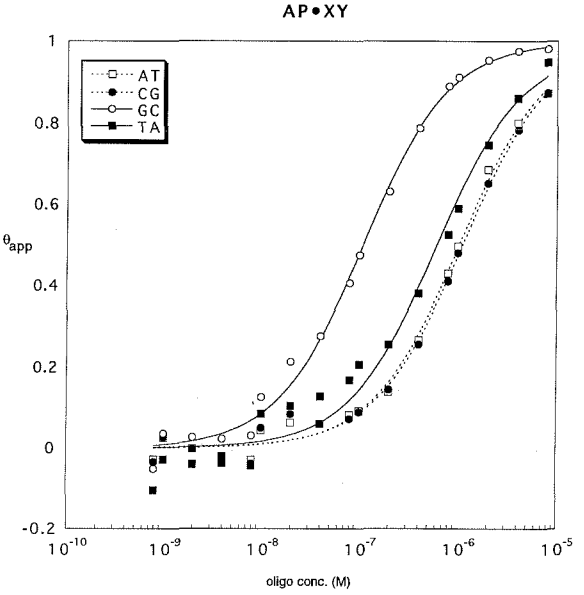
Inosine is specific for GC base pairs, with an affinity which is indistinguishable from that of G (7.1×10^7 vs. 6.3×10^7). This is surprising since the 2-amino group of G, which is involved in specific hydrogen bonding to the Hoogsteen face of the GC base pair, has been removed and only one potential hydrogen bond remains. AP, which likewise can only form one hydrogen bond, behaves more according to expectations. It remains specific for GC, but with an 8 fold reduction in affinity (9.3×10^6) which places it intermediate between a match and a mismatch.

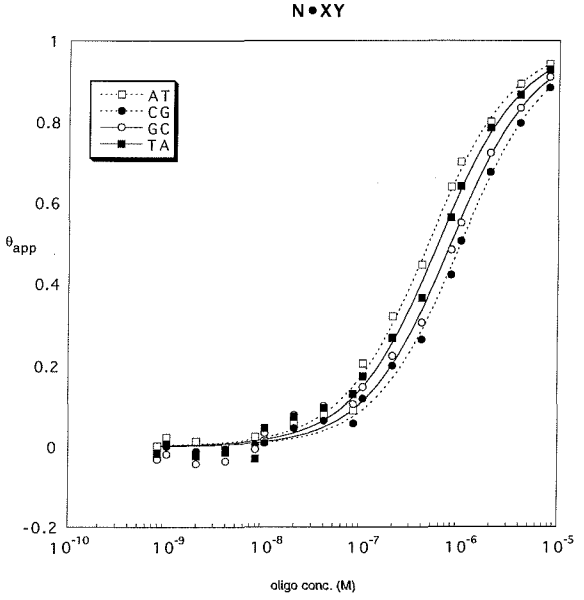
Nebularine shows no specificity for any base pair, binding all four with similar, relatively low affinity. Likely there are no specific hydrogen bonds formed with this base. This result is counter to the result previously obtained in this group by affinity cleavage titration, in which moderate affinity and specificity for CG and AT were observed.[2] In the previous study nebularine was placed in a slightly different sequence context, and under different conditions (4°C as opposed to 37°C here, and with tetravalent spermine rather than divalent Mg^{2+} present). Apparently interactions with nebularine are condition-dependent, and not generalizable. The nonspecificity and low affinity reported here is consistent with unsuccessful attempts to characterize intramolecular triple helices containing N opposite CG or AT by ^{1}H NMR.[6] It was found that these triplexes were not stable under conditions previously used to determine solution structures of purine motif triple helices.[7]

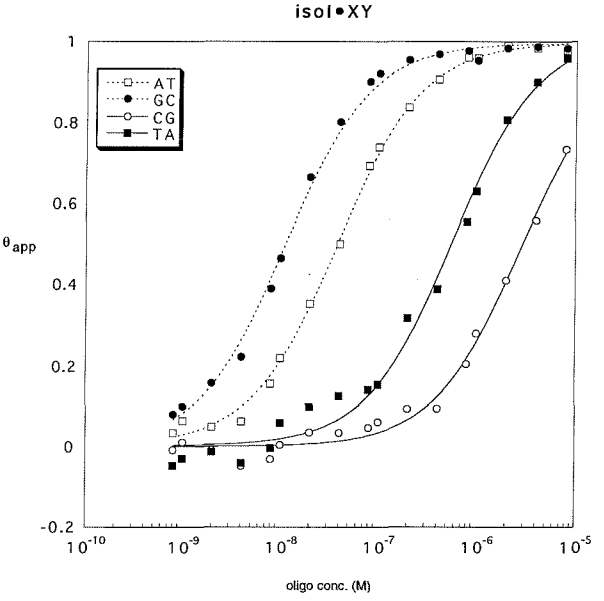
Surprisingly, isoinosine proved to have high affinity for both GC and AT base pairs. This base was designed to specifically recognize AT base pairs. While A and T already bind AT in the purine motif with high affinity and specificity, the triplet structures formed are not isomorphous with the G•GC triplet. As a result, sequences with multiple alternating GC and AT base pairs are poorly recognized due to destabilization by disruption of backbone continuity. Triplet isomorphism is illustrated in Figure 14. As with inosine, it

Figure 13. Binding isotherms for the Z•XY triplets studied, depicted in the form I•XY, AP•XY, N•XY, and isoI•XY. Each isotherm represents the average of three experiments conducted at 37°C, 50 mM tris acetate, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$.


I•XY
AT
CG
GC
TA
θapp
oligo conc. (M)


AP•XY
AT
CG
GC
TA
θapp
oligo conc. (M)


N•XY
AT
CG
GC
TA
θapp
oligo conc. (M)


isoI•XY
AT
GC
CG
TA
θapp
oligo conc. (M)

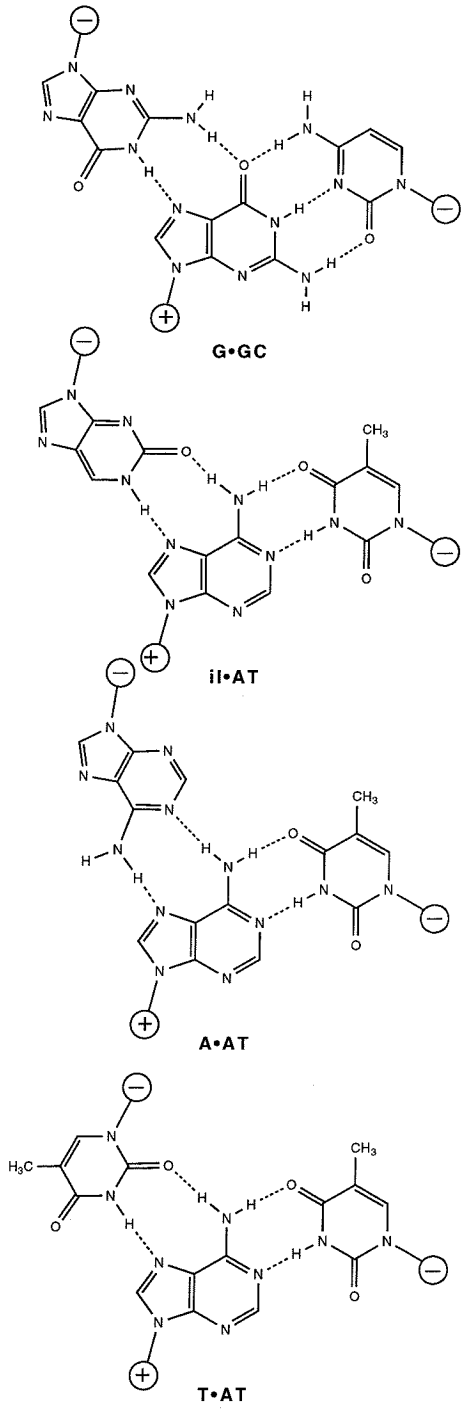


Figure 14. Triplet structures for G•GC, isoI•AT, A•AT, and T•AT, illustrating the isomorphism of isoI•AT with G•GC, and the different backbone placement present in the A•AT and T•AT triplets.

is not clear why isoinosine binds GC with high affinity. Perhaps AT and GC binding are due to different tautomers of isoinosine, as suggested by Seela in his studies with isoinosine in duplex formation, reported subsequent to the work done here.[8] In Seela's studies, isoI was found to be degenerate in duplex formation for all four bases, and destabilizing in all cases.

The specificity of I for GC suggested a potential application for this base as a substitute for G. A problem with G-rich oligonucleotides stems from the exceptional stability of self-assembled G-quartet structures in the presence of potassium (Figure 15), which has implications for *in vivo* application of purine motif oligonucleotides since cell nuclei contain a high concentration of potassium.[9]

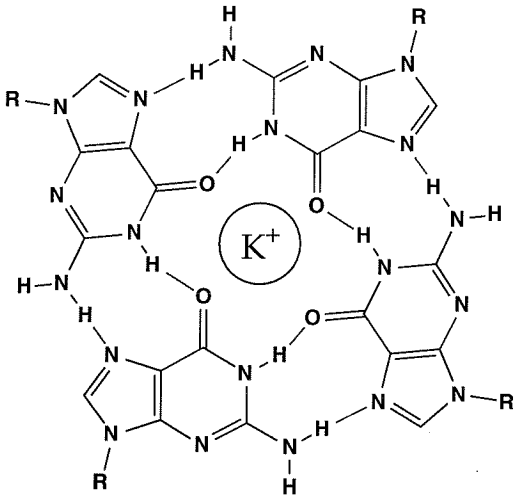


Figure 15. The structure of a potassium-stabilized G-quartet.

Since inosine lacks the 2-amino group which is critical for G-quartet formation, it was expected that inosine-containing oligonucleotides would not be sensitive to potassium. When multiple inosines, rather than one, were substituted into the third strand oligonucleotide, however, triple helix formation was not observed. This suggests that the

high affinity observed with a single inosine substitution may be due to a sequence-dependent interaction which is not generalizable. This result points to the importance of exploring the generalizablility of any result obtained with a single substitution by studying different sequence contexts and multiple substitutions.

Subsequent to this work, Hogan reported successfully overcoming G-quartet mediated inhibition of triple helix formation by replacing G with 6-thioG. When the 6-carbonyl is changed to a thiocarbonyl, quartet formation is not observed and triplex formation is only slightly weaker.[10]

Conclusion. The nonnatural bases studied here showed specificity for GC (I and AP), GC and AT (isoI), or no base pair (N). Notably, none of these bases showed high affinity or specificity for CG or TA base pairs. Thus, new structures must be explored to attain this end. This is the subject of the following two chapters.

Experimental section

General experimental methods were performed as described in chapter 2. Reagents for synthesis were purchased from Aldrich or Fluka and used without further purification. ^{1}H NMR spectra were recorded on a GE QE300 instrument operating at 300 MHz. Chemical shifts are reported in parts per million relative to solvent residual signal. UV spectra were measured on a Hewlett-Packard 8452A diode array spectrophotometer. High resolution mass spectra were obtained at the Mass Spectrometry Laboratory at the University of California, Riverside. Matrix assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF) was carried out at the Protein and Peptide Microanalytical Facility at the California Institute of Technology. Thin layer chromatography was performed on silica gel 60 F_{254} precoated plates, and chromatographic separations were performed with EM silica gel 60 (230-400 mesh). Oligonucleotides were synthesized, purified, and quantitated as described in chapter 2. I, AP, and N phosphoramidites were purchased from Glen Research.

HPLC analysis of enzymatically digested nonnatural base-containing oligonucleotides. 10 nmol samples of oligonucleotides were digested with 3 units calf alkaline phosphatase and 0.01 units snake venom phosphodiesterase in 50 μL 50 mM Tris, 10 mM $MgCl_2$, pH 8.0 for 3 hours at 37°C. Samples were injected onto a Rainin Microsorb MV C18 reverse phase column and eluted with a linear gradient of 0-20% acetonitrile in 20 mM ammonium acetate, pH 5.2, 55°C over 40 minutes at a flow rate of 1 mL/min. Samples of authentic nucleoside standards were run as well for comparison. Elution times: dG (11.3 min), T (13.1 min), I (10.8 min, λ_{max}=250 nm), AP (14.3 min, λ_{max}=304 nm), N (15.0 min, λ_{max}=262 nm), isoI (8.4 min, λ_{max}=314 nm).

Plasmid pRSPEC1. Plasmid construction followed standard methods.[11] Briefly, pRSPEC1 was prepared by hybridizing two complementary sets of oligonucleotides: 5'-CCGGCCGAAGTCTTGAGGCTCCCCTCCCCTCCCTCCCAAGTCTTGAGGCTCCC CTCCCCGCCCTCCGAAGTCTTGAGGCTCCCCT-3' and 5'-GGGGAGGGGAG CCTCAAGACTTCGGAGGGCGGGGAGGGGAGCCTCAAGACTTCGGAGGGAG GGGAGGGGAGCCTCAAGACTTCGG-3'; 5'-CCCCCCCCTCCGAAGTCTTGAGG CTCCCCTCCCCACCCTCCGAAGTCTTGAGGCAGCTTGGCGTAATCATGGTCCT TAAGTTCGAAG and 5'-TCGACTTCGAACTTAAGGACCATGATTACGCCA AGCTGCCTCAAGACTTCGGAGGGTGGCGAGGGGAGCCTCAAGACTTCGGAG GGG-3'. The oligonucleotide duplexes were phosphorylated with ATP and T4 polynucleotide kinase and ligated into the large *Ava* I/ *Sal* I fragment of pUC19. *E. coli* XL-1 Blue competent cells were transformed into the plasmid, and plasmid DNA from ampicillin-resistant white colonies was isolated. The presence of the desired insert was determined by restriction analysis and dideoxy sequencing. The plasmid was isolated on a preparative scale using a Promega maxiprep kit.

2-amino-9-(β-D-ribofuranosyl)purine (2): 6-thioguanosine (4.80g, 15.1 mmol) was dissolved in 600 mL methanol. 18 mL of an aqueous suspension of Raney nickel was added, and the mixture was stirred with a mechanical stirrer at 50°C for 12 hours. The nickel was filtered off and the crude product was concentrated and dried by azeotropic removal of water with pyridine followed by toluene. The unchromatographed product (3.20g, 80%) was suitably pure by NMR. ^{1}H NMR (DMSO-d_6) δ 8.60 (s, 1H), 8.30 (s, 1H), 6.56 (s, 2H), 5.82 (d, 1H), 5.45 (d, 1H), 5.18 (d, 1H), 5.05 (t, 1H), 4.52 (m, 1H), 4.13 (m, 1H), 3.90 (m, 1H), 3.64 (m, 1H), 3.56 (m, 1H).

2-amino-9-(3,5-O-tetraisopropyldisiloxane-β-D-ribofuranosyl)purine (3): Crude **2** (3.20g, 12.0 mmol) was dissolved in 75 mL pyridine under argon. 75 mL dichloromethane was added, and the mixture was cooled on ice. 1,3-dichlorotetraisopropyldisiloxane (3.84 mL, 12.0 mmol) was added slowly, and the reaction was stirred at room temperature for 4 hours. 1 mL methanol was added to quench, and the solvents were stripped down in vacuo. Column chromatography with 5% MeOH/DCM produced 4.53g (74%) white solid. ^{1}H NMR ($CDCl_3$) δ 8.72 (s, 1H), 8.01 (s, 1H), 5.98 (s, 1H), 5.30 (s, 2H), 4.82 (m, 1H), 4.58 (m, 1H), 4.00-4.10 (m, 3H), 3.43 (s, 1H), 1.0-1.3 (m, 28H).

2-amino-9-(2-O-phenoxythiocarbonyl-3,5-0-tetraisopropyldisiloxane-β-D-ribofuranosyl)purine (4): **3** (3.50g, 6.87 mmol) and 2 equivalents (1.68g, 13.75 mmol) 4-dimethylaminopyridine were dissolved in 150 mL acetonitrile. Under argon, phenylchlorothionoformate (1.31g, 7.56 mmol) was slowly added. The mixture was stirred for 10 hours, quenched with methanol, and dried in vacuo. Chromatography with 3% MeOH/CH_2Cl_2 produced 3.75g (84%) light yellow foam. ^{1}H NMR ($CDCl_3$) δ 8.75 (s, 1H), 7.97 (s, 1H), 7.44 (t, 2H), 7.33 (t, 1H), 7.15 (d, 2H), 6.44 (d, 1H), 6.12 (s, 1H), 5.08 (m, 1H), 4.98 (s, 2H), 4.0-4.2 (m, 3H), 1.0-1.3 (m, 28H).

2-amino-9-(2-deoxy-3,5-O-tetraisopropyldisiloxane-β-D-ribofuranosyl) purine (5): **4** (3.40g, 5.26 mmol) was dissolved in 100 mL toluene. The solution was degassed by bubbling in argon for 20 minutes. 2,2'-azobisisobutyronitrile (173 mg, 1.05 mmol) was added, followed by tributyltin hydride (2.12 mL, 7.89 mmol). The mixture was stirred at 75°C for 3 hours. Evaporation of solvent was followed by chromatography with 2% MeOH/DCM, yielding 2.21g (85%) white solid. ^{1}H NMR ($CDCl_3$) δ 8.70 (s,

1H), 8.00 (s, 1H), 6.22(dd, 1H), 5.01 (s, 2H), 4.83 (dd, 1H), 4.03 (m, 2H), 3.89 (m, 1H), 2.73 (m, 1H), 2.62 (m, 1H), 1.0-1.3 (m, 28H).

2-amino-9-(2-deoxy-β-D-ribofuranosyl)purine (6): **5** (725 mg, 1.47 mmol) was dissolved in 5 mL THF. 1.47 mL, 1.47 mmol tetrabutylammonium fluoride 1M solution in THF was added dropwise at room temperature, with formation of a precipitate. After 30 minutes stirring, solvent was removed under reduced pressure and the product purified by silica chromatography (20% MeOH/CH_2Cl_c), yielding 301 mg (82%) white solid. TLC (20% MeOH/CH_2Cl_c) R_f 0.35; UV λ_{max}=304 nm; ^{1}H NMR (DMSO-d_6) δ 8.56 (s, 1H), 8.26 (s, 1H), 6.53 (s, 2H), 6.24 (t, 1H), 5.29 (d, 1H), 4.95 (t, 1H), 4.34 (m, 1H), 3.81 (m, 1H), 3.52 (m, 2H), 2.63 (m, 1H), 2.23 (m, 1H).

2-oxo-9-(2-deoxy-β-D-ribofuranosyl)purine (7): **6** (500 mg, 1.99 mmol)was dissolved in 12 mL H_2O at 50°C. Sodium nitrite (9 mmol, 618 mg) was added, followed by acetic acid (16 mmol, 910 μL) in a dropwise fashion. Bubbling occurred, slowing down after several minutes. After 30 minutes the solution was neutralized with sodium hydroxide and dried by azeotroping with pyridine. The products were redissolved in 25% MeOH/CH_2Cl_2, and insoluble salts filtered. Chromatography (25%MeOH/CH_2Cl_2) produced product which still contained a significant amount of sodium acetate. The remaining salt was removed by passing the mixture through a column of C18 reverse phase silica eluting with water. After lyophilzation, **7** was isolated as a white solid (300 mg, 66%). TLC (20% MeOH/CH_2Cl_2) R_f 0.15; UV λ_{max}=314 nm; ^{1}H NMR (DMSO-d_6) δ 8.44 (s, 1H), 8.36 (s, 1H), 6.14 (t, 1H), 5.2 (br, 2H), 4.34 (m, 1H), 3.82 (m, 1H), 3.50 (m, 2H), 2.58 (m, 1H), 2.21 (m, 1H); MS M+H 253.0927 (253.0937 calcd. for $C_{10}H_{13}N_4O_4$).

2-oxo-9-[5-*O*-[bis(4-methoxyphenyl)phenylmethyl]-2-deoxy-β-D-ribofuranosyl]purine (8): Isoinosine nucleoside **7** (159 mg, 0.630 mmol) and DMT chloride (320 mg, 0.946 mmol) were combined and dried overnight under vacuum. 7.5 mL pyridine (Fluka, anhydrous, over sieves) was added and the mixture stirred at 4°C for 2.5 hours. Excess DMTCl was quenched with 1 mL MeOH, and 40 mL dichloromethane was added. The mixture was extracted with saturated sodium bicarbonate, then brine, dried ($MgSO_4$), and concentrated *in vacuo*, with azeotropic removal of pyridine with heptane. The product was purified by silica chromatography (7% MeOH/ 0.25% TEA/ CH_2Cl_2) to produce a white foam (288 mg, 83%). TLC (7% MeOH/ 0.25% TEA/ CH_2Cl_2) R_f 0.30; ^1H NMR (CD_2Cl_2) δ 8.62 (s, 1H), 8.05 (s, 1H), 7.37 (m, 2H), 7.24 (m, 7H), 6.76 (m, 4H), 6.47 (m, 1H), 4.61 (m, 1H), 4.20 (m, 1H), 3.68 (s, 6H), 3.35 (m, 2H), 2.62 (m, 2H) FAB MS M-H 553.2112 (553.2087 calcd. for $C_{31}H_{29}N_4O_6$).

3'-phosphoramidite of 5'-DMT protected isoinosine (9): **8** (288 mg, 0.52 mmol) was dissolved in 4 mL dichloromethane. Diisopropylethylamine (271 μL, 1.56 mmol) was added. Stirring at room temperature, 2-cyanoethyl-N,N-diisopropylchlorophosphoramidite (145 μL, 0.65 mmol) was added dropwise. After 1.5 hours, the reaction was quenched with 1 mL MeOH. 40 mL dichloromethane was added, and the mixture extracted with saturated $NaHCO_3$ and brine, dried, and concentrated *in vacuo*. Product was purified by silica chromatography (5% MeOH/ 0.25% TEA/ CH_2Cl_2) and then precipitated from hexanes to remove H-phosphonate impurity, yielding a white foam (278 mg, 71%). TLC (5% MeOH/ 0.25% TEA/ CH_2Cl_2) R_f 0.25; ^1H NMR (mixture of diastereomers) ($CDCl_3$) δ 8.34 (s, 1H), 8.06 (d, 1H), 6.80-7.45 (m, 13H), 6.37 (t, 1H),

4.74 (m, 1H), 4.31 (m, 1H), 3.78 (s, 6H), 3.50-4.0 (m, 4H), 3.40 (m, 2H), 2.50-2.90 (m, 4H), 1.21 (m, 12 H). FAB MS M+H 755.3347 (755.3322 calcd. for $C_{40}H_{48}N_6O_7P$).

Quantitative DNase I footprint titrations. 3'-end labeling and footprint titration experiments were performed as described in chapter 2, except that no nonspecific oligonucleotide was added to the DNase reaction, as it was found to have no visible effect on enzymatic activity. Quantitation and data analysis were also performed as previously described in chapter 2.

References

(1) Greenberg, W.A.; Dervan, P.B. *J. Am. Chem. Soc.* **1995**, *117*, 5016-5022.

(2) Stilz, H. U.; Dervan, P. B. *Biochemistry* **1993**, *32*, 2177-2185.

(3) Jones, R.J.; Swaminathan, S.; Milligan, J.F.; Wadwani, S.; Froehler, B.C.; Matteuci, M. *J. Am. Chem. Soc.* **1993**, *115*, 9816-9817.

(4) Hunziker, J.; Priestley, E.S.; Brunar, H.; Dervan, P.B. *J. Am. Chem. Soc.* **1995**, *117*, 2661-2662.

(5) Robins, M.J. Wilson, J.S.; Hansske, F. *J. Am. Chem. Soc.* **1983**, *105*, 4059-4065.

(6) Greenberg, W.A.; Dervan. P.B.; Patel, D.J. Unpublished results.

(7) (a) Radhakrishnan, I.; Patel, D.J. *J. Am. Chem. Soc.* **1993**, *115*, 1615-1617. (b) Radhakrishnan, I.; de los Santos, C.; Patel, D.J. *J. Mol. Biol* **1993**, *234*, 188-197. (c) Radhakrishnan, I.; Patel, D.J. *Structure* **1993**, *1*, 135-152.

(8) Seela, F.; Chen, Y. *Nucleic Acids Res.* **1995**, *23*, 2499-2505.

(9) Olivas, W.M.; Maher III, L.J. *Nucleic Acids Res.* **1995**, *23*, 1936-1941.

(10) Gee, J.E.; Revankar, G.R.; Rao, T.S.; Hogan, M.E. *Biochemistry* **1995**, *34*, 2042-2048.

CHAPTER FOUR

Energetics of Purine Motif Triple Helix Formation of Oligonucleotides Containing Substituted Imidazoles

Introduction

The previous two chapters demonstrate that among the four natural DNA bases, and among a series of nonnatural purines, there is no solution to the specific, high affinity recognition of CG or TA base pairs within the context of the purine motif. Likewise, there is no general solution to this recognition problem in the pyrimidine motif.[1] This chapter addresses attempts to design novel nucleosides for recognition of TA base pairs within the purine motif. The design approach was based on observation of the steric constraints imparted on the major groove by a TA base pair, as illustrated in Figure 1.

The structure of TA shown in Figure 1 demonstrates the prominence of the methyl group at the 5- position of thymine, which juts into the major groove. It was envisioned that this methyl group presents a steric hindrance to the third strand base adjacent to it, forcing the backbone of the third strand to be pushed out to accommodate both bases. Thus the design called for relaxation of steric hindrance by attenuating the size of the third strand base, that is, changing from a purine or a pyrimidine to a single 5-membered ring, specifically imidazole.

Furthermore, it was realized that the methyl group of thymine represented the only hydrophobic group amongst an array of hydrogen bond donating and accepting groups in the major groove, and recognition of this methyl group would therefore be highly specific. A number of DNA binding proteins, for example GCN4,[2] and a transition metal complex[3] have been shown to specifically recognize the thymine methyl group through hydrophobic interactions. Thus, a series of methyl-substituted imidazole nucleosides (Figure 2) which present a hydrophobic face for recognition of TA base pairs was designed for synthesis and evaluation. Simultaneous with this work, a group at the former Triplex Pharmaceutical

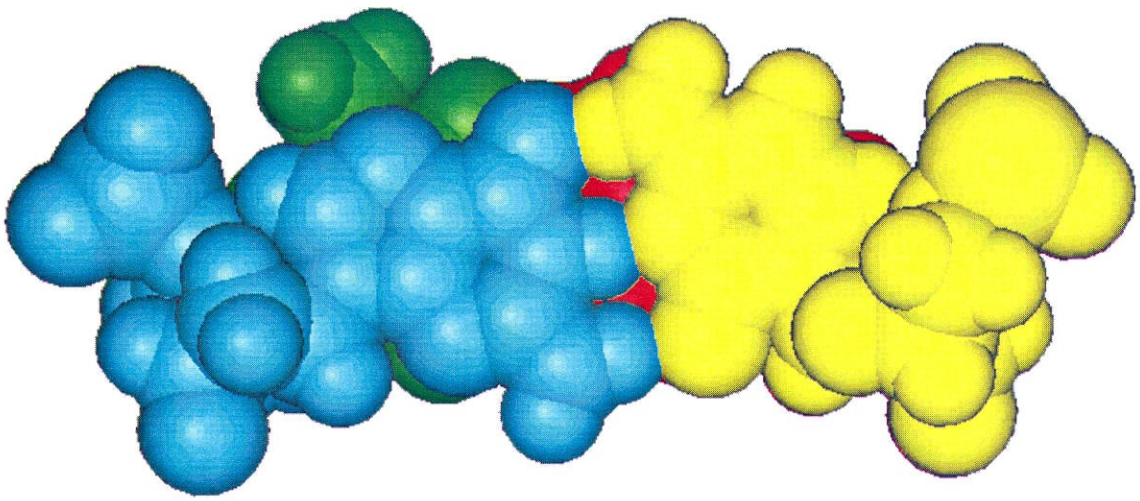


Figure 1. Van der Waals surface of a TA base pair overlaid on that of a GC base pair with superimposition of glycosidic bond positions. Note the prominence of the thymine methyl group in the major groove of the TA base pair.

Corporation studied the use of unsubstituted imidazole in a very similar system.[4] Their results were similar to those observed here for unsubstituted imidazole.

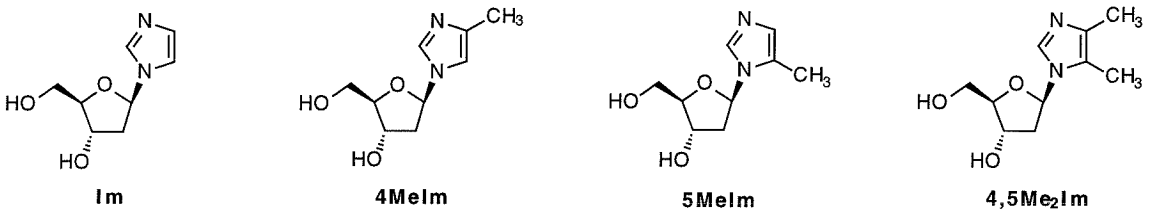


Figure 2. Structures of the series of imidazole nucleosides

Results and Discussion

The 5'-dimethoxytrityl phosphoramidites of this series of imidazole nucleosides were synthesized by a common route, illustrated in Figure 3. The appropriately substituted imidazole was glycosylated by the standard procedure with protected 1-α-chloro-2-deoxyriboside **1**[5] to generate compounds **2a-d**. Glycosylation of 4-methylimidazole produced an approximately 2:1 mixture of the 4-methyl and 5-methyl regioisomers which were separated by preparative reverse phase HPLC after toluoyl ester aminolysis. The structures of the isomers **3b** and **3c** were assigned based on nOe difference spectra of the two isolated nucleosides (see Experimental section for details). 5'-dimethoxytritylation and phosphitylation provided phosphoramidites **5a-d** for incorporation into purine-rich oligonucleotides.

Imidazole nucleosides were incorporated into the standard purine-rich oligonucleotide sequence, 5'-TGGGGTGGGG**Z**GGGT-3' at the single position Z, for

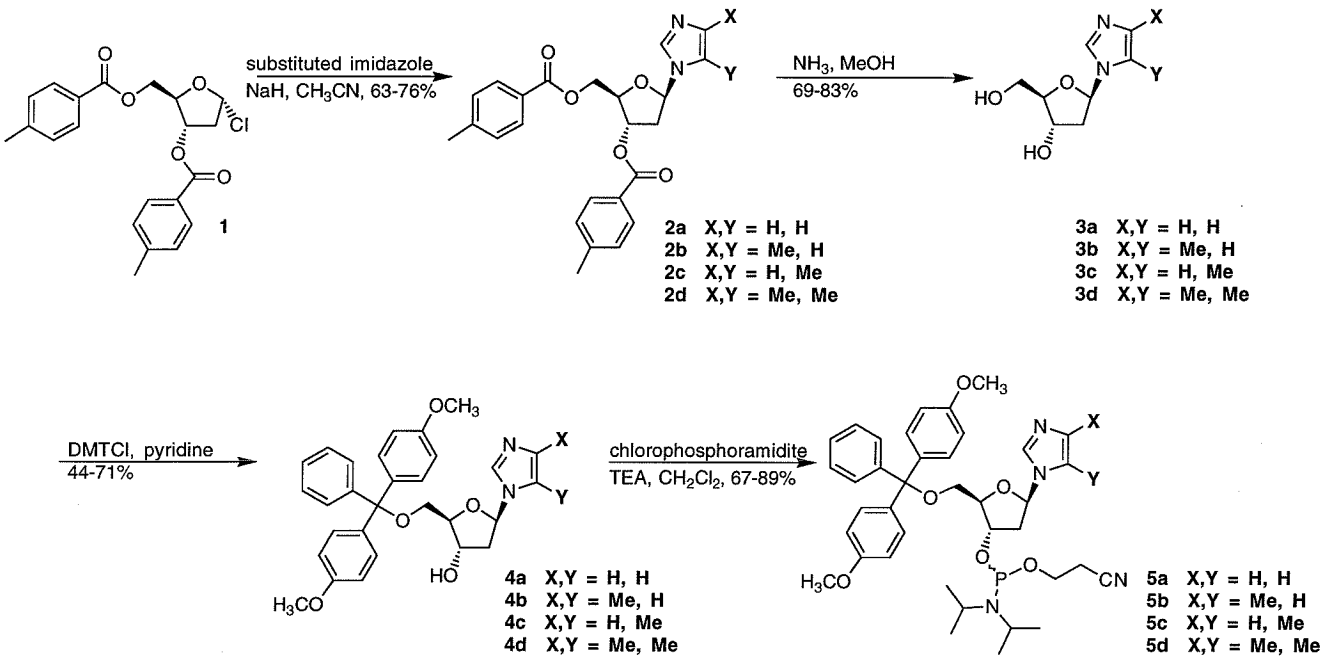


Figure 3. Synthesis of imidazole phosphoramidites **5a-d**

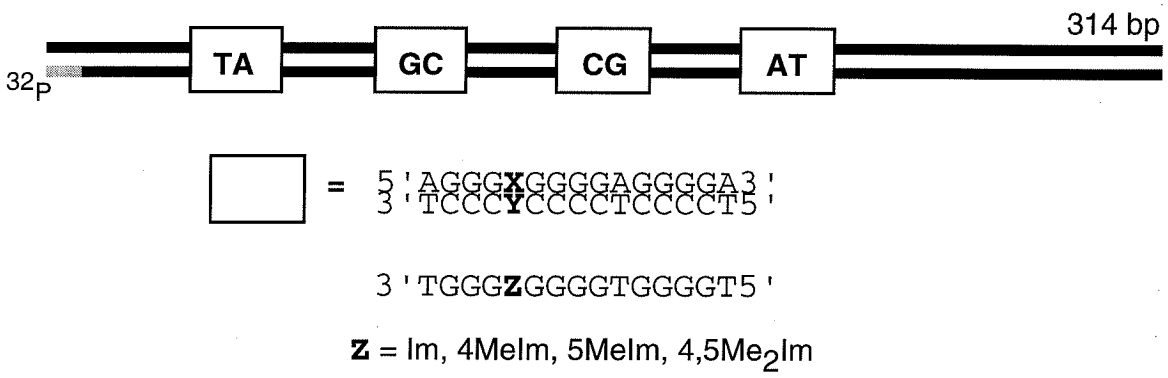


Figure 4. Sequences of the four target sites on 3'-labeled 314 bp *Afl* II/*Fsp* I restriction fragment of pRSPEC1. Below are the sequences of third strand oligonucleotides containing the imidazole substitutions.

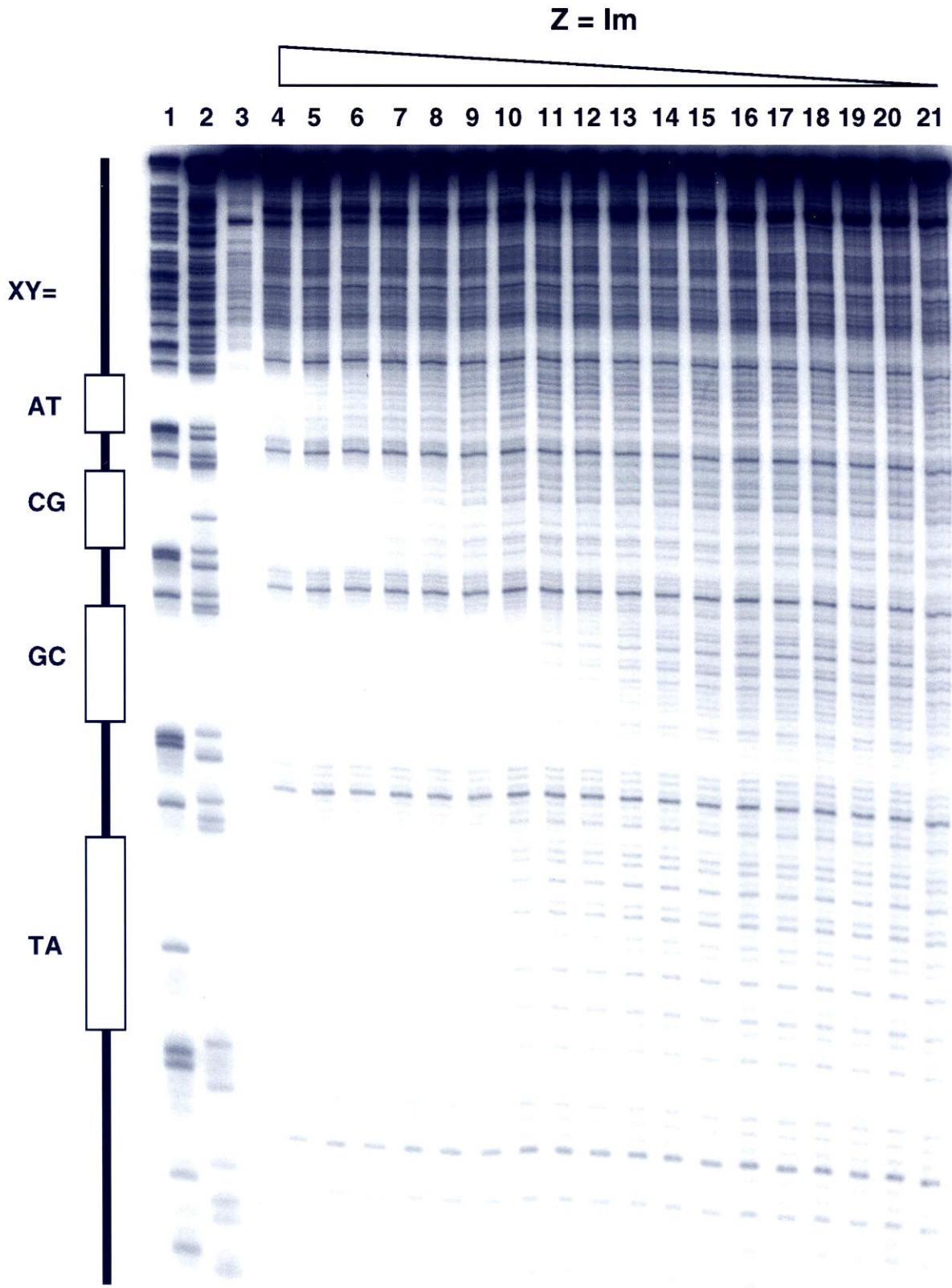
footprinting on the standard 314 bp restriction fragment from pRSPEC1 containing all four target sites 5'-AGGG**X**GGGGAGGGGA-3', where X = A,G,C, and T (Figure 4). The energetics corresponding to the Z•XY triplet interactions are summarized in Table 1. Autoradiograms of gels (Figures 5-8) demonstrate the specificity of each novel nucleoside. The putative triplet structures for these sixteen Z•XY triplets are shown in Figures 9-12.

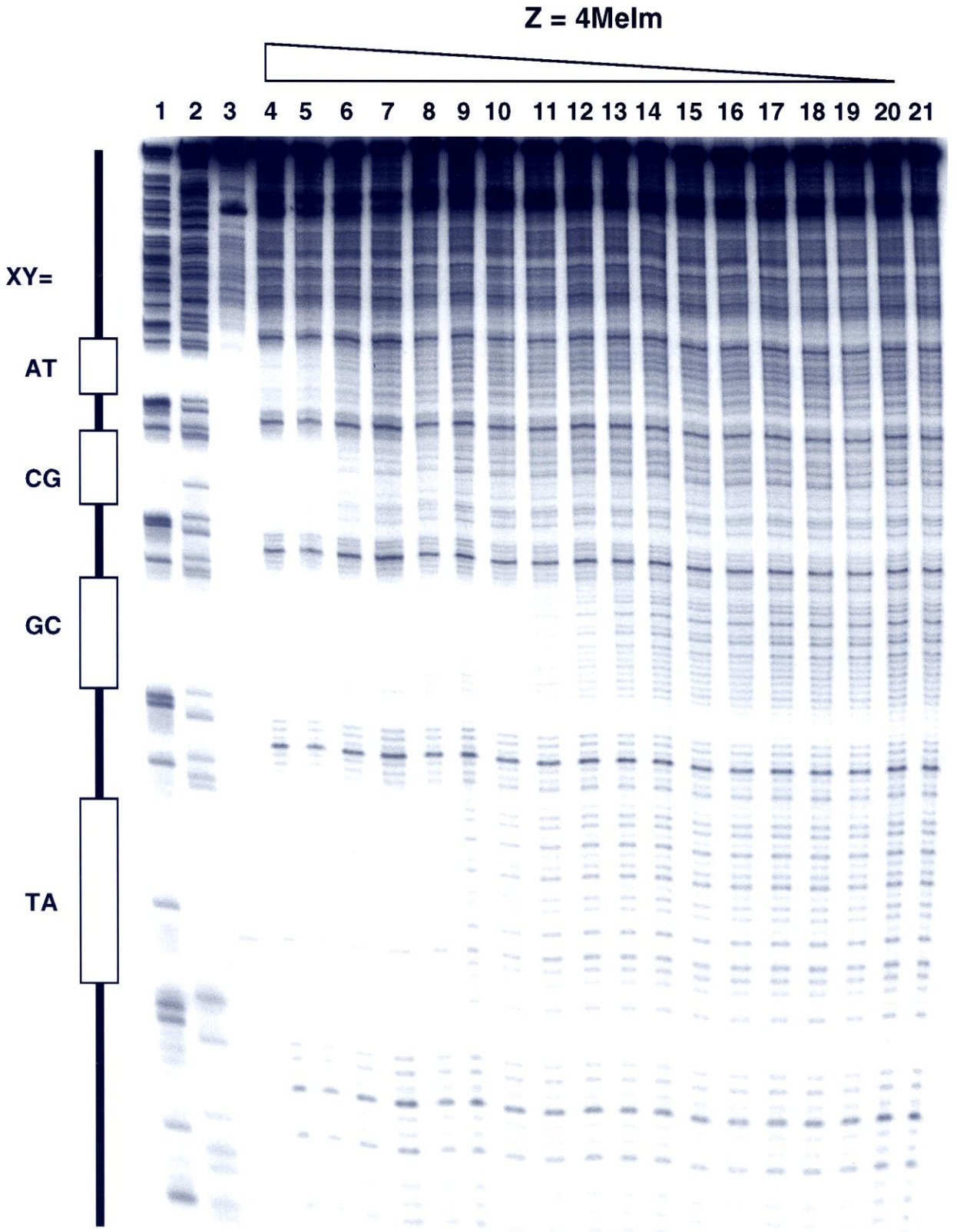
Table 1. Association constants (K_T) for the formation of sixteen triple helical complexes containing the Z•XY triplets at 37°C, 10 mM NaCl, 3 mM $MgCl_2$, 50 mM tris acetate, pH 7.4.[a,b]

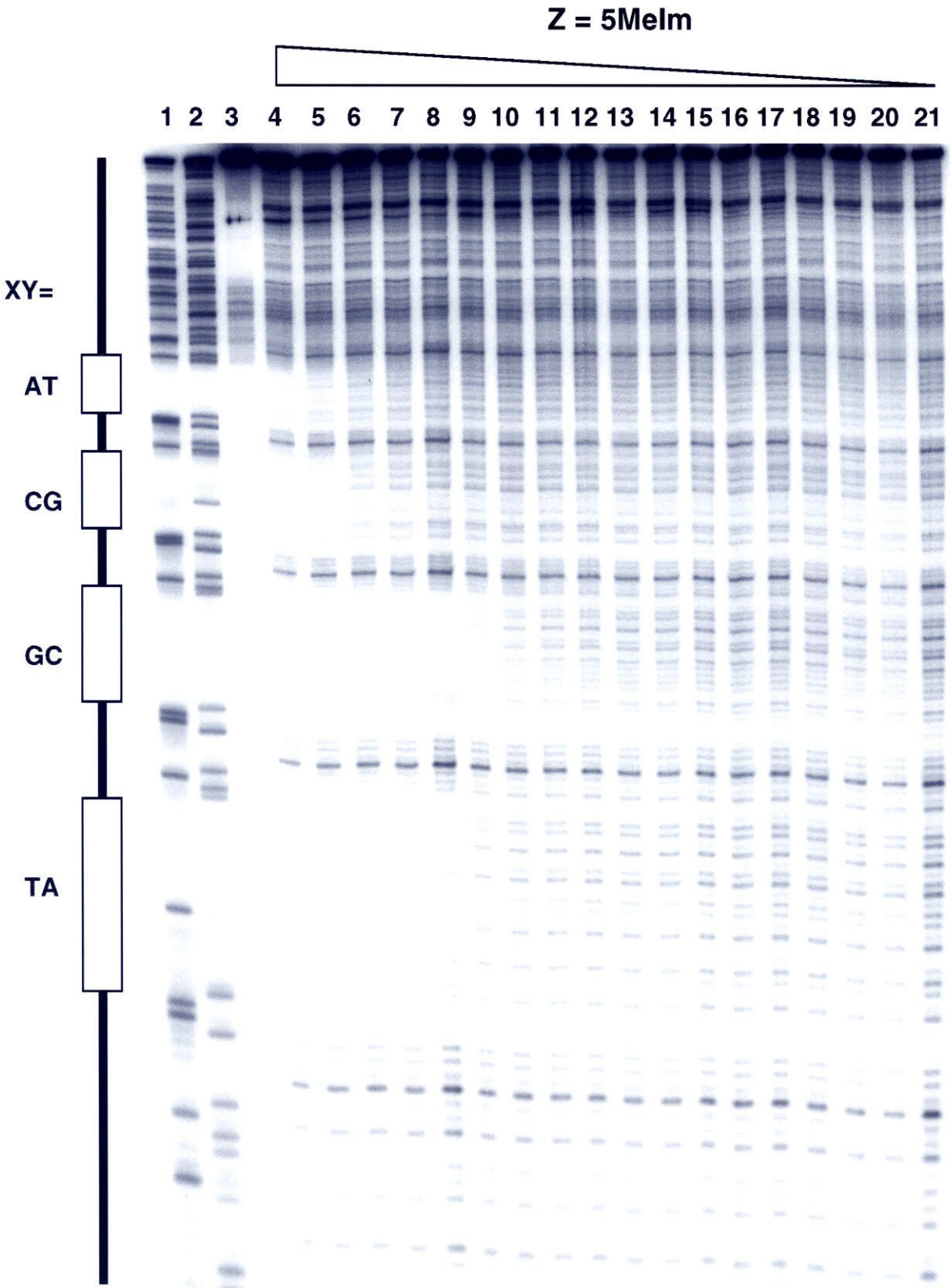
XY	**Z=** Im	4MeIm	5MeIm	$4,5Me_2Im$
AT	5.7 (±1.1) x 10^5	3.6 (±1.7) x 10^5	3.1 (±0.5) x 10^5	2.6 (±1.3) x 10^5
GC	1.6 (±0.1) x 10^7	1.7 (±0.1) x 10^7	4.6 (±0.2) x 10^6	2.0 (±0.2) x 10^6
CG	1.3 (±0.3) x 10^6	7.8 (±1.6) x 10^5	3.9 (±0.9) x 10^5	2.3 (±1.0) x 10^5
TA	3.8 (±0.2) x 10^6	3.0 (±0.2) x 10^6	2.4 (±0.3) x 10^6	1.2 (±0.1) x 10^6

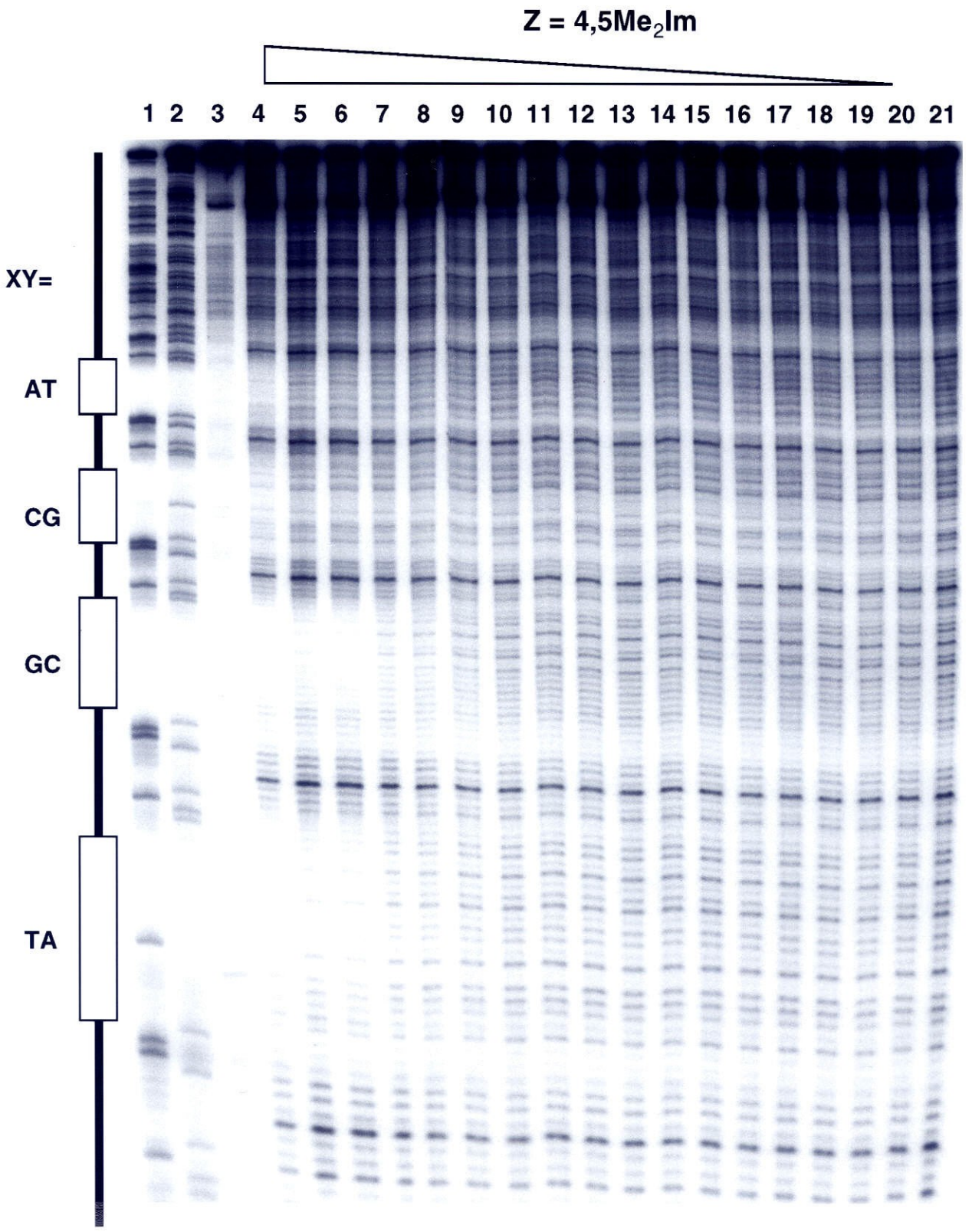
[a] K_T values are reported as the mean (± the standard error of the mean) of three measurements. The K_T values are reported in units of M^{-1}. [b] The identity of the base Z is indicated across the top of the columns; the identity of the Watson Crick base pair XY is indicated on the left side of the rows.

Figures 5-8. Autoradiograms of 8% denaturing polyacrylamide gels used to separate fragments generated by DNase I digestion in quantitative footprint titration experiments with oligonucleotides containing Im (Figure 5), 4MeIm (Figure 6), 5MeIm (Figure 7), and 4,5Me_2Im (Figure 8). The four target sites on the 314 bp 3'-end labeled restriction fragment are labeled as XY= AT, CG, GC, and TA respectively, going from the top to the bottom of each gel. (Lane 1) Products of an adenine-specific reaction. (Lane 2) Products of a guanine-specific reaction. (Lane 3) Intact 3' labeled DNA after incubation in the absence of third strand oligonucleotide. (Lanes 4-21) DNase I digestion products obtained in the presence of varying concentrations of oligonucleotide: 8 μM (lane 4); 4 μM (lane 5); 2 μM (lane 6); 1 μM (lane 7); 800 nM (lane 8); 400 nM (lane 9); 200 nM (lane 10); 100 nM (lane 11); 80 nM (lane 12); 40 nM (lane 13);20 nM (lane 14); 10 nM (lane 15); 8 nM (lane 16); 4 nM (lane 17); 2 nM (lane 18); 1 nM (lane 19); 800 pM (lane 20); no oligonucleotide (lane 21).

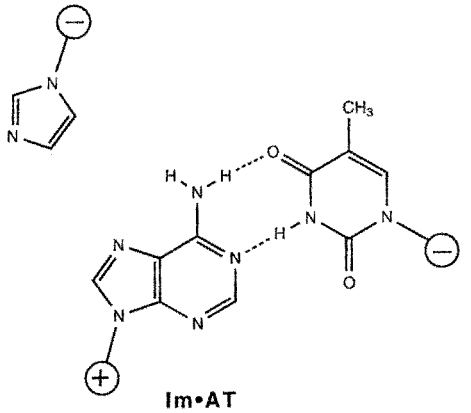

Z = Im
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
XY=
AT
CG
GC
TA


Z = 4MeIm
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
XY=
AT
CG
GC
TA

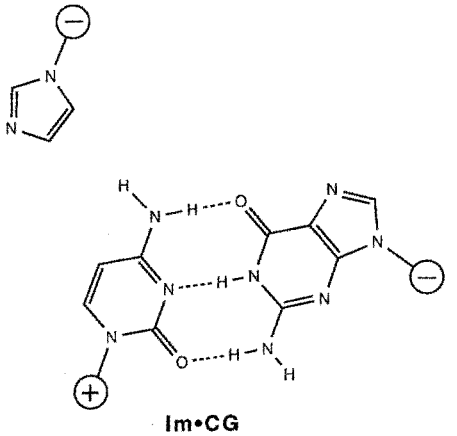

Z = 5MeIm
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
XY=
AT
CG
GC
TA


Z = 4,5Me2Im
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
XY=
AT
CG
GC
TA

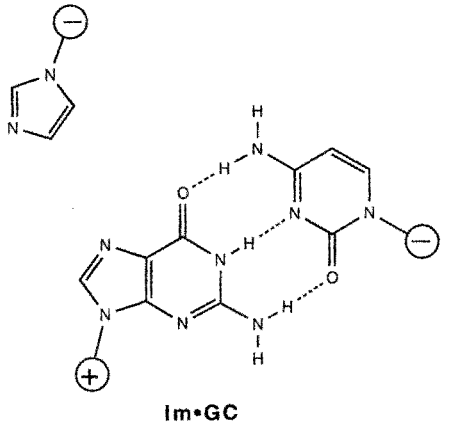
Figures 9-12. Putative triplet structures for Im•XY (Figure 9), 4MeIm•XY (Figure 10), 5MeIm•XY (Figure 11), and 4,5Me_2Im•XY (Figure 12).



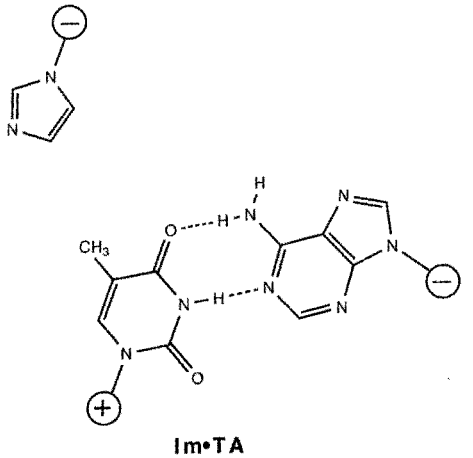
Im•AT



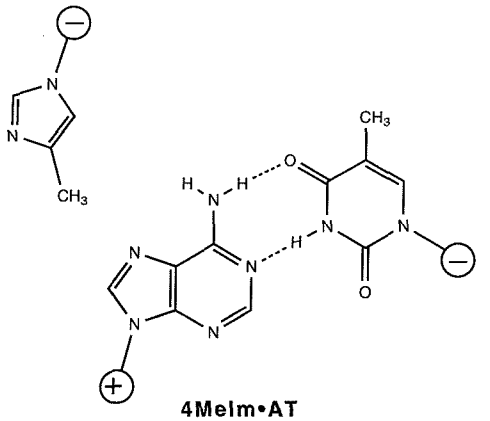
Im•CG



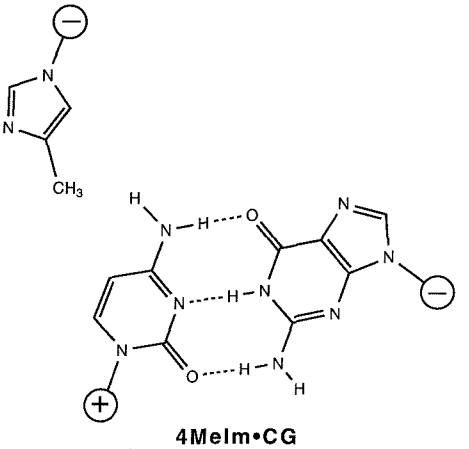
Im•GC



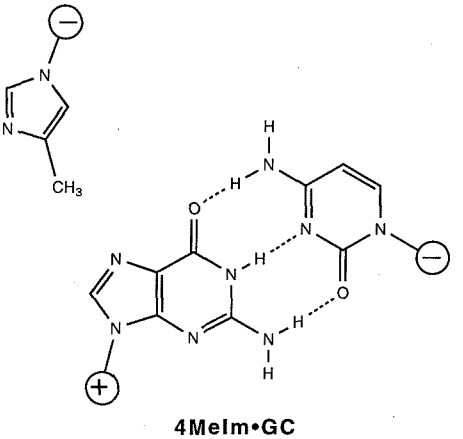
Im•TA



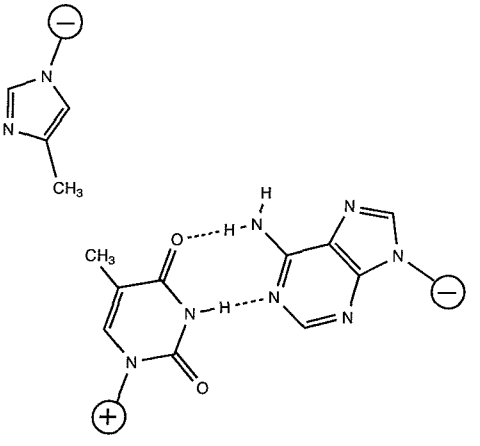
4MeIm•AT



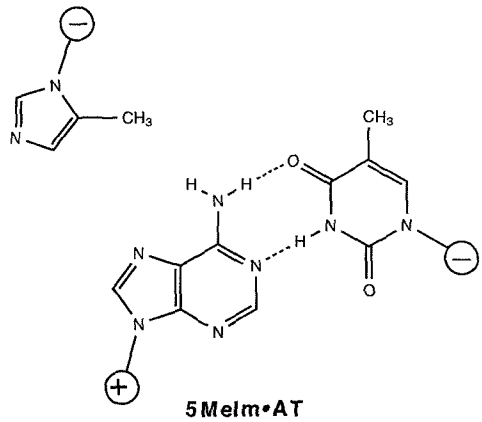
4MeIm•CG



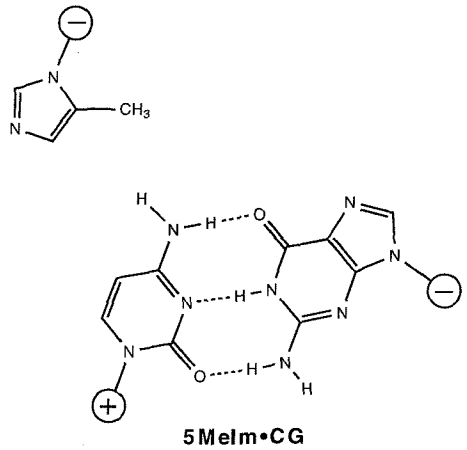
4MeIm•GC



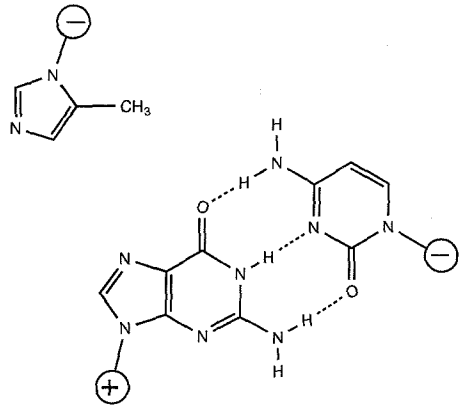
4MeIm•TA



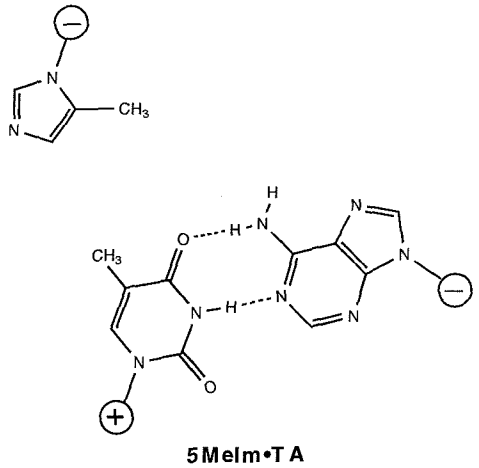
5MeIm•AT



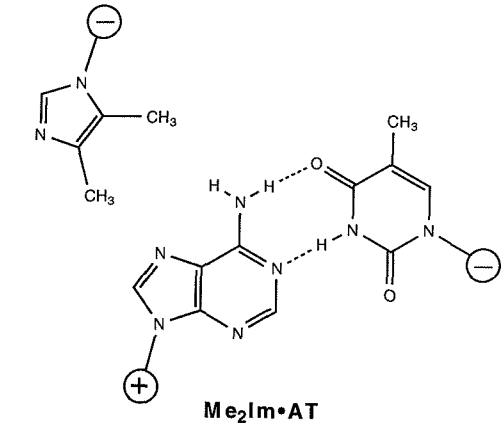
5MeIm•CG



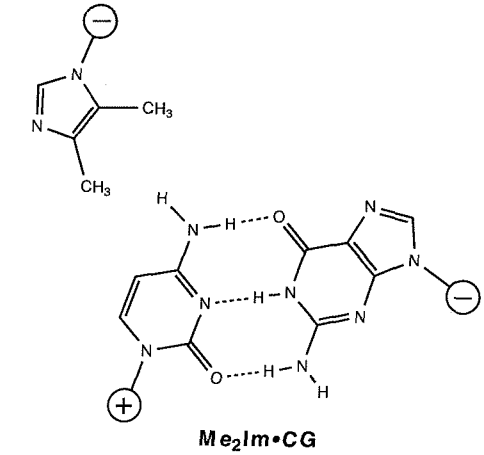
5MeIm•GC



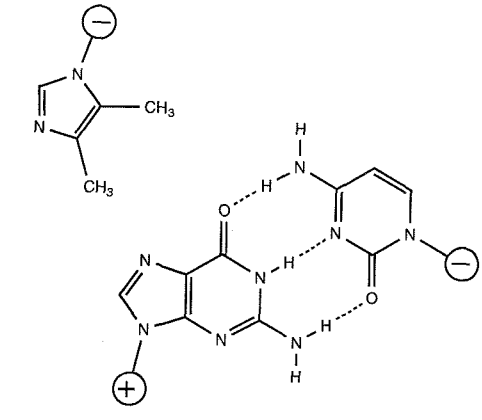
5MeIm•TA



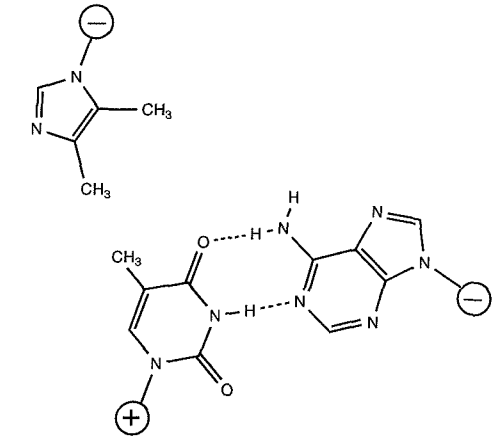
Me_2Im•AT



Me_2Im•CG



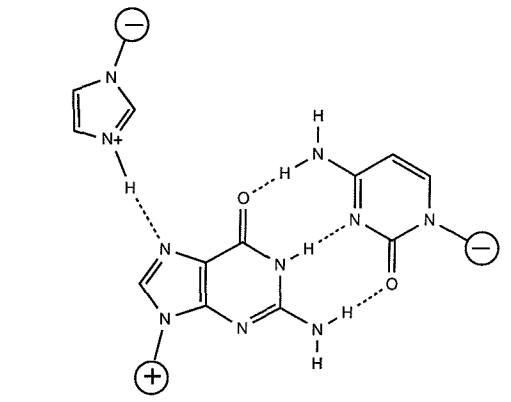
Me_2Im•GC



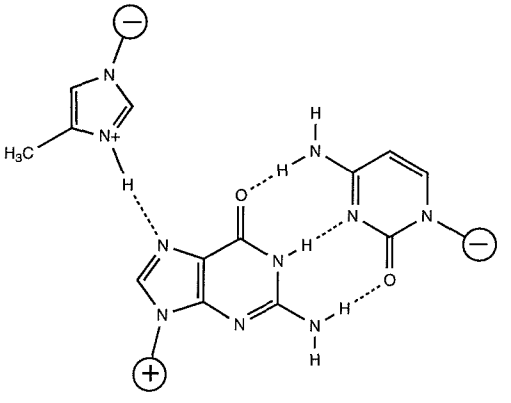
Me_2Im•TA

Inspection of Table 1 and Figure 5 show that the unsubstituted imidazole Im does bind TA better than any of the bases summarized in chapters 2 and 3, but only by a very slight amount. More notably, Im binds GC quite specifically, with affinity high enough to be suggestive of hydrogen bond formation. This possibility could be rationalized by protonation of N3 of the imidazole (which likely has a pKa in the area of 6 or 7) and rotation by 180° around the glycosidic bond from the orientation shown in Figure 9. This possible Im•GC hydrogen bonded triplet is illustrated in Figure 13. This rationale is supported by the data for 4MeIm and 5MeIm (Figure 13). The binding profile of 4MeIm is identical to that of Im, suggesting that the methyl group is not interacting at all with the base pairs. This would be consistent with the orientation shown in Figure 13, where the face of the nucleoside presented to the DNA is identical for Im and 4MeIm. In contrast, in the case of 5MeIm, nOe studies (see Experimental section) and modeling suggest that rotation around the glycosidic bond is sterically hindered and the conformation in which the methyl group is over the sugar is not readily accessible. Consistent with this, for 5MeIm affinity for GC is greatly reduced, suggesting the structure shown in Figure 13. Unfortunately, there is no increase in affinity for TA as was hoped if there were some positive interaction between methyl groups. Likewise, for the dimethyl derivative 4,5Me$_2$Im, while there is an even further reduction in GC affinity, there is a concomitant reduction in binding to all other base pairs, including TA. This compound has an ~ 10-fold preference for GC and TA base pairs, but its overall affinity is too low to make it useful.

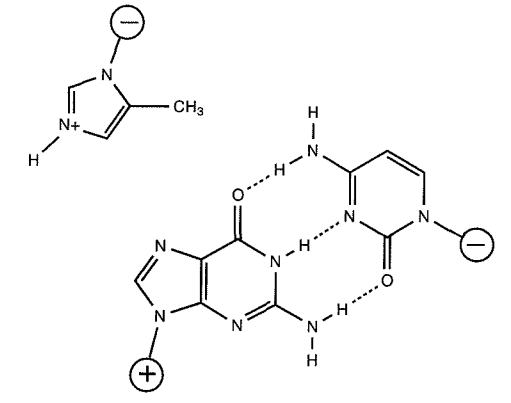
The results presented in this chapter suggest that hydrophobic interactions with isolated thymine methyl groups will likely not provide a large energetic contribution to triple helix formation. However, the results also suggest that the design of smaller heterocycles to spatially accommodate the thymine methyl group appears valid since in the case of all four imidazoles presented here, TA affinity is second only to GC affinity. Since



Im•GC



4MeIm•GC



5MeIm•GC

Figure 13. Structures of protonated imidazole triplets in which specific hydrogen bonds to GC are formed upon 180° rotation around the glycosidic bond.

GC affinity is likely due to the protonated structures shown in Figure 13, elimination of the protonated form may result in TA-specific novel nucleosides. This proposal is addressed in the following chapter.

Conclusion

A series of substituted imidazole nucleosides was designed and incorporated into purine-rich oligonucleotides to investigate the possibility of specific recognition of TA base pairs in the purine motif by small, five-membered ring heterocycles which can spatially accommodate the thymine methyl group and at the same time specifically recognize this functional group through hydrophobic interactions. The results obtained through quantitative DNase I footprint titration experiments indicate that while an enhancement in affinity is not observed with methyl substituents on the imidazole, the small ring system may indeed be more accommodating to the thymine methyl group and therefore a good platform for the next generation of design of novel nucleosides for specific recognition of TA base pairs by triple helix formation.

Experimental section

General methods in the synthesis of phosphoramidites and oligonucleotides were performed as described in the previous chapters. Oligonucleotide purifications were performed as previously described in chapters 2 and 3. Because this series of imidazole nucleosides do not have useful chromophores for HPLC analysis of digested oligonucleotides (λ_{max} = 208 nm, 218 nm, and 228 nm for imidazole, 4- and 5-methylimidazole, and 4,5-dimethylimidazole respectively), incorporation and integrity of novel nucleosides was measured by MALDI-TOF mass spectrometry on intact 15mer oligonucleotides. MALDI MS: Z=Im 4715.1 (calcd. 4717.1), Z=4MeIm 4732.0 (calcd. 4731.1), Z=5MeIm 4730.0 (calcd. 4731.1), Z=4,5-Me_2Im 4745.4 (calcd. 4745.1).

1-[2-deoxy-3,5-*O*-toluoyl-β-D-ribofuranosyl]imidazole (2a): Imidazole (170 mg, 2.50 mmol) was added to 25 mL acetonitrile. With stirring at room temperature sodium hydride (66mg, 2.75 mmol as a 60% dispersion in mineral oil) was added. After completion of bubbling, 40 minutes, 2-deoxy-3,5-*O*-toluoyl-α-D-ribose (**1**)[5] (1.0 g, 2.57 mmol) was added. The mixture was stirred for 1 hour, at which time 50 mL dichloromethane was added, the mixture extracted with water, dried, and concentrated. Silica chromatography (5%MeOH/CH_2Cl_2) yielded a light yellow solid (796 mg, 76%). TLC (5% MeOH/CH_2Cl_c) R_f 0.40; ^{1}H NMR ($CDCl_3$) δ 7.95 (m, 4H), 7.72 (s, 1H), 7.28 (m, 4H), 7.09 (s, 2H), 6.16 (t, 1H), 5.67 (m, 1H), 4.62 (m, 3H), 2.71 (m, 2H), 2.45 (s, 3H), 2.43 (s, 3H).

1-[2-deoxy-β-D-ribofuranosyl]imidazole (3a): **2a** (1.0 g, 2.38 mmol) was dissolved in 60 mL ammonia-saturated methanol, stirred at room temperature for 24

hours, then concentrated. Product was purified by silica chromatography (20%MeOH/CH_2Cl_2) to furnish a crystalline white solid (302 mg, 69%). TLC (5% MeOH/CH_2Cl_c) R_f 0.05; ^{1}H NMR (DMSO-d_6) δ 7.83 (s, 1H), 7.32 (s, 1H), 6.88 (s, 1H), 6.01 (t, 1H, J=6.3 Hz), 5.24 (m, 1H), 4.89 (m, 1H), 4.26 (m, 1H), 3.76 (m, 1H), 3.45 (m, 2H), 2.30 (m, 1H), 2.20 (m, 1H).

1-[2-deoxy-5-*O*-(4,4'-dimethoxytrityl)-β-D-ribofuranosyl]imidazole (4a): Imidazole nucleoside **3a** (70 mg, 0.38 mmol) and DMT chloride (193 mg, 0.57 mmol) were dried overnight under vacuum. 5 mL pyridine was added, and the reaction mixture was stirred at room temperature for 8 hours, after which 0.5 mL methanol was added to quench, and the solvent was removed under reduced pressure. Purification by silica chromatography (5%MeOH/CH_2Cl_2) yielded a light tan solid (95 mg, 51 %). TLC (8% MeOH/CH_2Cl_c) R_f 0.35; ^{1}H NMR (acetone-d_6) δ 7.73 (s, 1H). 7.44 (m, 2H), 7.30 (m, 7H), 7.18 (s, 1H), 6.86 (s, 1H), 6.84 (m, 4H), 6.12 (t, 1H), 4.50 (m, 1H), 4.06 (m, 1H), 3.77 (s, 6H), 3.25 (m, 2H), 2.45 (m, 2H).

Imidazole nucleoside phosphoramidite (5a): 5'-DMT nucleoside **4a** (150 mg, 0.31 mmol) was dissolved in 3 mL dichloromethane. Triethylamine (126 μL, 0.92 mmol) was added, followed by dropwise addition of 2-cyanoethyl-N,N-diisopropylchlorophosphoramidite (86 μL, 0.39 mmol). The reaction was stirred at room temperature for 1 hour, after which it was quenched with 0.2 mL MeOH, diluted with 25 mL dichloromethane, extracted with saturated $NaHCO_3$, then brine, dried, and concentrated. Purification by silica chromatography (4%MeOH/CH_2Cl_2/0.25% TEA) and precipitation from hexanes provided a light yellow foam (188 mg, 89%). TLC (5% MeOH/CH_2Cl_c) R_f 0.35; ^{1}H NMR (mixture of diastereomers)(acetone-d_6) δ 7.75 (s, 1H).

7.45 (m, 2H), 7.30 (m, 7H), 7.20 (s, 1H), 6.93 (s, 1H), 6.84 (m, 4H), 6.16 (t, 1H), 4.68 (m, 1H), 4.16 (m, 1H), 3.75 (s, 6H), 3.5-4.0 (m, 4H), 3.29 (m, 2H), 2..89 (m, 1H),2.5-2.75 (m, 3H), 1.18 (m, 12H).

4-methyl-1-[2-deoxy-3,5-*O*-toluoyl-β-D-ribofuranosyl]imidazole (2b) and 5-methyl-1-[2-deoxy-3,5-*O*-toluoyl-β-D-ribofuranosyl]imidazole (2c) : 4-methylimidazole (632 mg, 7.7 mmol) was added to 60 mL acetonitrile. Sodium hydride (340 mg, 8.5 mmol as 60% dispersion in mineral oil) was added. After 45 minutes, when bubbling was complete, chlorosugar **1** was added (3.0 g, 7.7 mmol). The mixture was stirred for 1 hour, then concentrated and purified by silica chromatography (4% $MeOH/CH_2Cl_2$) yielding a white solid (2.1 g, 63%). The regioisomers (2:1 4-methyl to 4-methyl) were not separable by this method, and were carried to the next step together, after which they were separated. TLC (5% $MeOH/CH_2Cl_2$) R_f 0.40; ^1H NMR ($CDCl_3$) major (4-methyl) isomer δ 7.95 (m, 4H), 7.59 (s, 1H), 7.28 (m, 4H), 6.76 (s, 1H), 6.07 (t, 1H), 5.67 (m, 1H), 4.62 (m, 3H), 2.67 (m, 2H), 2.45 (s, 6H), 2.17 (s, 3H); minor (5-methyl) isomer δ 7.95 (m, 4H), 7.66 (s, 1H), 7.28 (m, 4H), 6.80 (s, 1H), 6.02 (t, 1H), 5.67 (m, 1H), 4.62 (m, 3H), 2.67 (m, 2H), 2.43 (s, 6H), 2.29 (s, 3H); FAB MS M+H 435.1918 (435.1920 calcd. for $C_{25}H_{27}N_2O_5$).

4-methyl-1-[2-deoxy-β-D-ribofuranosyl]imidazole (3b) and 5-methyl-1-[2-deoxy-β-D-ribofuranosyl]imidazole (3c) : The mixture of regioisomers **2b** and **2c** (1.0 g, 2.3 mmol) was dissolved in 60 mL ammonia-saturated methanol and stirred for 24 hours, after which it was concentrated and chromatographed (20% $MeOH/CH_2Cl_2$) with no separation of isomers to produce a colorless oil (380 mg, 83%). The isomers were separated by preparative reverse phase HPLC over a gradient of 0-25% acetonitrile in water

in 100 minutes. The major isomer eluted first (10.0% acetonitrile), lyophilizing down to a white solid (180 mg). The minor, less polar isomer (11.5% acetonitrile) lyophilized to 120 mg colorless syrup. Structures of the major and minor isomers were assigned based of nOe difference spectral data. Major isomer **3b** (4-methyl), irradiation of 1'H: 3% nOe to 4'H (establishes β-anomer), 6% nOe to 2', 2% nOe to H5, 8% nOe to H2 (resulting from *syn* conformation), no nOe to methyl group observed. Minor isomer **3c** (5-methyl), irradiation of 1'H: 2% nOe to 4'H, 1% nOe to H2 (*syn* conformation not as accessible due to steric hindrance), 2% nOe to 2'H, 6 % nOe to 5-methyl group. 4-methyl isomer **3b** ^{1}H NMR (DMSO-d_6) δ 7.70 (s, 1H), 6.96 (s, 1H), 6.03 (t, 1H, J=6.6 Hz), 4.43 (m, 1H), 3.96 (m, 1H), 3.62 (m, 2H), 2.42 (m, 2H), 2.08 (s, 3H). 5-methyl isomer **3c** ^{1}H NMR (DMSO-d_6) δ 7.79 (s, 1H), 6.70 (s, 1H), 6.02 (t, 1H, J=6.6 Hz), 4.46 (m, 1H), 3.97 (m, 1H), 3.62 (m, 2H), 2.57 (m, 1H), 2.42 (m, 1H), 2.18 (s, 3H); FAB MS **(3b)** M+H 199.1086 (199.1083 calcd. for $C_9H_{15}N_2O_3$), **(3c)** M+H 199.1079 (199.1083 calcd. for $C_9H_{15}N_2O_3$).

4-methyl-1-[2-deoxy-5-*O*-(4,4'-dimethoxytrityl)-β-D-ribofuranosyl]imidazole (4b) : **3b** (70 mg, 0.35 mmol) was dried under vacuum with DMTCl (178 mg, 0.52 mmol). 5 mL pyridine was added, and the mixture stirred at room temperature for 7 hours, after which it was quenched with 0.2 mL MeOH, and concentrated. Silica chromatography (5% MeOH/CH_2Cl_c) produced a tan foam (100 mg, 57%). TLC (8% MeOH/CH_2Cl_c) R_f 0.35; ^{1}H NMR (acetone-d_6) δ 7.59 (s, 1H). 7.44 (m, 2H), 7.30 (m, 7H), 6.88 (s, 1H), 6.84 (m, 4H), 6.02 (t, 1H, J=6.7Hz), 4.50 (m, 1H), 4.04 (m, 1H), 3.75 (s, 6H), 3.24 (m, 2H), 2.40 (m, 2H), 2.05 (s, 3H); FAB MS M+H 501.2373 (501.2389 calcd. for $C_{30}H_{33}N_2O_5$).

4-methyl-1-[2-deoxy-5-*O*-(4,4'-dimethoxytrityl)-β-D-ribofuranosyl]imidazole

(4c) : **3c** (70 mg, 0.35 mmol) was dried under vacuum with DMTCl (178 mg, 0.52 mmol). 5 mL pyridine was added, and the mixture stirred at room temperature for 7 hours, after which it was quenched with 0.2 mL MeOH, and concentrated. Silica chromatography (5% MeOH/CH_2Cl_c) produced a tan foam (77 mg, 44%). TLC (8% MeOH/CH_2Cl_c) R_f 0.35; ^{1}H NMR (acetone-d_6) δ 7.62 (s, 1H). 7.44 (m, 2H), 7.30 (m, 7H), 6.84 (m, 4H), 6.64 (s, 1H), 6.00 (t, 1H, J=6.6Hz), 4.51 (m, 1H), 4.05 (m, 1H), 3.75 (s, 6H), 3.22 (m, 2H), 2.53 (m, 1H), 2.39 (m, 1H), 2.25 (s, 3H); FAB MS M+H 501.2360 (501.2389 calcd. for $C_{30}H_{33}N_2O_5$).

4-Methylimidazole nucleoside phosphoramidite (5b): 5'-DMT nucleoside **4b** (78 mg, 0.16 mmol) was dissolved in 3 mL dichloromethane. Triethylamine (64 μL, 0.47 mmol) was added, followed by dropwise addition of 2-cyanoethyl-N,N-diisopropylchlorophosphoramidite (45 μL, 0.20 mmol). The reaction was stirred at room temperature for 1 hour, after which it was quenched with 0.2 mL MeOH, diluted with 25 mL dichloromethane, extracted with saturated $NaHCO_3$, then brine, dried, and concentrated. Purification by silica chromatography (3%MeOH/CH_2Cl_2/0.25% TEA) and precipitation from hexanes provided a white foam (85 mg, 78%). TLC (4% MeOH/CH_2Cl_c) R_f 0.35; ^{1}H NMR (mixture of diastereomers)(acetone-d_6) δ 7.61 (s, 1H). 7.45 (m, 2H), 7.30 (m, 7H), 6.88 (s, 1H), 6.84 (m, 4H), 6.07 (t, 1H), 4.68 (m, 1H), 4.14 (m, 1H), 3.77 (s, 6H), 3.5-4.0 (m, 4H), 3.28 (m, 2H), 2..89 (m, 1H),2.5-2.75 (m, 3H), 2.04 (s, 3H), 1.18 (m, 12H). FAB MS M+H 701.3437 (701.3468 calcd. for $C_{39}H_{50}N_4O_6P$).

5-Methylimidazole nucleoside phosphoramidite (5c): 5'-DMT nucleoside **4c** (66 mg, 0.13 mmol) was dissolved in 3 mL dichloromethane. Triethylamine (54 μL, 0.40 mmol) was added, followed by dropwise addition of 2-cyanoethyl-N,N-diisopropylchlorophosphoramidite (38 μL, 0.17 mmol). The reaction was stirred at room temperature for 1 hour, after which it was quenched with 0.2 mL MeOH, diluted with 25 mL dichloromethane, extracted with saturated $NaHCO_3$, then brine, dried, and concentrated. Purification by silica chromatography (3%MeOH/CH_2Cl_2/0.25% TEA) and precipitation from hexanes provided a white foam (71 mg, 77%). TLC (4% MeOH/CH_2Cl_2) R_f 0.35; ^{1}H NMR (mixture of diastereomers)(acetone-d_6) δ 7.66 (s, 1H). 7.45 (m, 2H), 7.30 (m, 7H), 6.84 (m, 4H), 6.66 (m, 1H), 6.03 (t, 1H), 4.69 (m, 1H), 4.14 (m, 1H), 3.76 (s, 6H), 3.5-4.0 (m, 4H), 3.27 (m, 2H), 2..93 (m, 1H),2.5-2.75 (m, 3H), 2.26 (s, 3H), 1.18 (m, 12H). FAB MS M+H 700.3418 (700.3390 calcd. for $C_{39}H_{49}N_4O_6P$).

4,5-Dimethyl-1-[2-deoxy-3,5-*O*-toluoyl-β-D-ribofuranosyl]imidazole (2d): 4,5-dimethylimidazole[6] (370 mg, 3.85 mmol) was added to 35 mL acetonitrile. Sodium hydride (170 mg, 4.2 mmol as 60% dispersion in mineral oil) was added. After 45 minutes, when bubbling was complete, chlorosugar **1** was added (1.5 g, 3.85 mmol). The mixture was stirred for 1 hour, then concentrated and purified by silica chromatography (4% MeOH/CH_2Cl_2) yielding a light yellow foam (1.2 g, 69%). TLC (4% MeOH/CH_2Cl_2) R_f 0.25; ^{1}H NMR ($CDCl_3$) δ 7.95 (m, 4H), 7.58 (s, 1H), 7.28 (m, 4H), 5.96 (m, 1H), 5.67 (m, 1H), 4.61 (m, 2H), 4.54 (m, 1H), 2..75 (m, 1H), 2.65 (m, 1H), 2.45 (s, 3H), 2.42 (s, 3H), 2.21 (s, 3H), 2.15 (s, 3H); FAB MS M+H 449.2052 (449.2076 calcd. for $C_{26}H_{29}N_2O_5$).

4,5-Dimethyl-1-[2-deoxy-β-D-ribofuranosyl]imidazole (3d): 2d (640 mg, 1.43 mmol) was dissolved in 40 mL ammonia-saturated methanol, stirred at room temperature for 24 hours, then concentrated. Product was purified by silica chromatography (15%MeOH/CH_2Cl_2) to furnish a light yellow solid (210 mg, 70%). TLC (15% MeOH/CH_2Cl_c) R_f 0.20; ^{1}H NMR (D_2O) δ 7.75 (s, 1H), 5.98 (t, 1H, J=6.3 Hz), 4.45 (m, 1H), 3.97 (m, 1H), 3.61 (m, 2H), 2.55 (m, 1H), 2.40 (m, 1H), 2.12 (s, 3H), 2.03 (s, 3H); FAB MS M+H 213.1225 (213.1239 calcd. for $C_{10}H_{17}N_2O_3$).

4,5-Dimethyl-1-[2-deoxy-5-*O*-(4,4'-dimethoxytrityl)-β-D-ribofuranosyl] imidazole (4d): Dimethylimidazole nucleoside **3d** (68 mg, 0.32 mmol) and DMT chloride (163 mg, 0.48 mmol) were dried overnight under vacuum. 5 mL pyridine was added, and the reaction mixture was stirred at room temperature for 7 hours, after which 0.5 mL methanol was added to quench, and the solvent was removed under reduced pressure. Purification by silica chromatography (5%MeOH/CH_2Cl_2) yielded a light tan solid (117 mg, 71 %). TLC (5% MeOH/CH_2Cl_c) R_f 0.30; ^{1}H NMR (acetone-d_6) δ 7.52 (s, 1H). 7.44 (m, 2H), 7.30 (m, 7H), 6.84 (m, 4H), 5.93 (m, 1H), 4.49 (m, 1H), 4.05 (m, 1H), 3.76 (s, 6H), 3.19 (m, 2H), 2.49 (m, 1H), 2.35 (m, 1H), 2.18 (s, 3H), 2.04 (s, 3H). FAB MS M 514.2493 (514.2468 calcd. for $C_{31}H_{34}N_2O_5$).

4,5-Dimethylimidazole nucleoside phosphoramidite (5d): 5'-DMT nucleoside **4d** (102 mg, 0.20 mmol) was dissolved in 4 mL dichloromethane. Triethylamine (80 μL, 0.59 mmol) was added, followed by dropwise addition of 2-cyanoethyl-N,N-diisopropylchlorophosphoramidite (56 μL, 0.25 mmol). The reaction was stirred at room temperature for 1 hour, after which it was quenched with 0.2 mL MeOH, diluted with 25 mL dichloromethane, extracted with saturated $NaHCO_3$, then brine, dried, and

concentrated. Purification by silica chromatography (3%MeOH/CH_2Cl_2/0.25% TEA) and precipitation from hexanes provided a white oil (95 mg, 67%). TLC (4% MeOH/CH_2Cl_c) R_f 0.35; ^1H NMR (mixture of diastereomers)(acetone-d_6) δ 7.57 (s, 1H). 7.45 (m, 2H), 7.30 (m, 7H), 6.84 (m, 4H), 5.97 (t, 1H), 4.68 (m, 1H), 4.14 (m, 1H), 3.76 (s, 6H), 3.5-4.0 (m, 4H), 3.24 (m, 2H), 2..90 (m, 1H),2.5-2.75 (m, 3H), 2.19 (s, 3H). 2.05 (s, 3H), 1.18 (m, 12H). FAB MS M+H 715.3595 (715.3624 calcd. for $C_{40}H_{51}N_4O_6P$).

Quantitative DNase I footprint titrations. 3'-end labeling and footprint titration experiments were performed as described in chapters 2 and 3. Quantitation and data analysis were also performed as previously described in chapter 2.

References

(1) Priestley, E.S. Ph.D. Thesis, California Institute of Technology, 1996.

(2) Ellenberger, T.E.; Brandl, C.J.; Struhl, K.; Harrison, S.C. *Cell* **1992**, *71*, 1223-1237.

(3) Hudson, B.P.; Dupureur, C.M.; Barton, J.K. *J. Am. Chem. Soc.* **1995**, *117*, 9379-9380.

(4) Durland, R.H.; Rao, T.S.; Bodepudi, V.; Seth, D.M.; Jayaraman, K.; Revankar, G.R. *Nucleic Acids Res.* **1995**, *23*, 647-653.

(5) Hoffer, M. *Chem. Ber.* **1960**, *93*, 2777-2781.

(6) Bredereck,H.; Theilig, G. *Chem. Ber.* **1953**, *86*, 88-93.

CHAPTER FIVE

Energetics of Purine Motif Triple Helix Formation of Oligonucleotides Containing Substituted Pyrazoles

Introduction

The previous chapter described the design of a series of imidazoles for recognition of TA base pairs by triple helix formation in the purine motif. While the design of hydrophobic contacts with the thymine methyl group did not result in an increase in affinity, the basic concept of using a smaller heterocycle to better accommodate the methyl group sterically did hold promise for gaining specificity for TA over the other base pairs. A problem associated with the imidazole series was GC specificity that likely resulted from protonation at N3 and hydrogen bond formation. A way of eliminating this problem would be to change from an imidazole heterocycle (pKa ~6-7) to a pyrazole (pKa ~2-3). Thus a series of pyrazole nucleosides were designed for incorporation into purine-rich oligonucleotides for recognition of TA base pairs by triple helix formation. The structures of these nucleosides are shown in Figure 1.

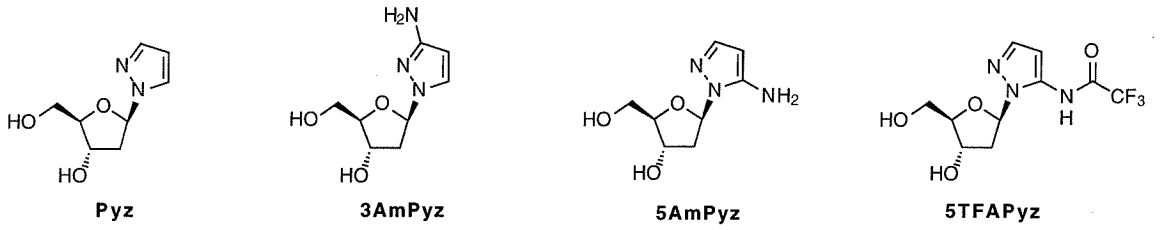
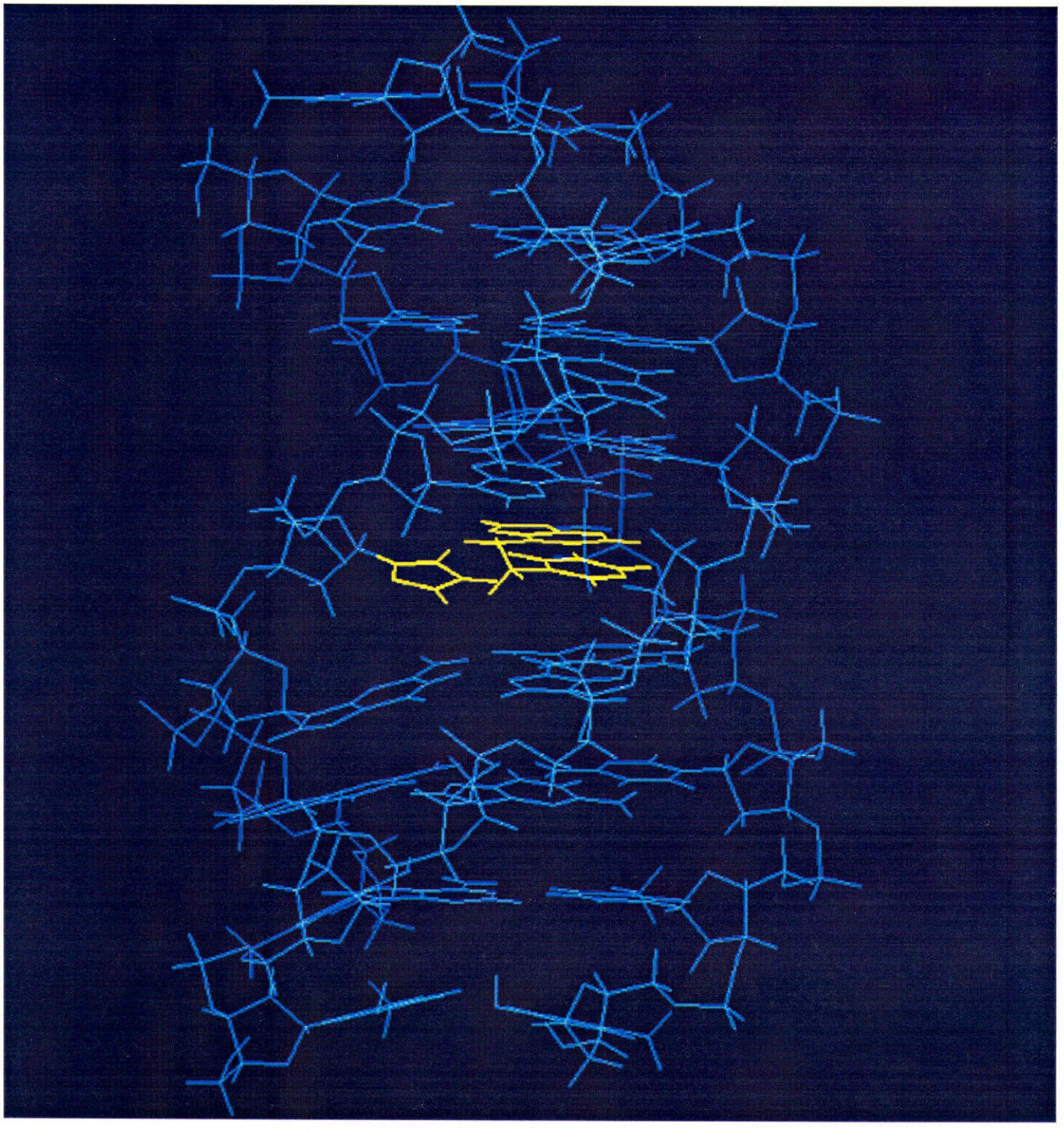


Figure 1. Structures of the series of pyrazole nucleosides

Figure 2 shows a model of the pyrazole nucleoside (Pyz) adjacent to a TA base pair in the context of a purine motif triple helix. This model illustrates the prominence of the thymine methyl group, and why a smaller heterocycle is likely important in design for TA recognition.

Figure 2. Model of a pyrazole•TA triplet in the context of a purine motif triple helix. The model was constructed from the structure of an intramolecular purine motif triple helix derived from NMR constraints.[1] The central T•AT triplet was replaced with a Pyz•TA triplet, in yellow.



The unsubstituted pyrazole nucleoside (Pyz) was independently studied by the former Triplex Pharmaceutical Corporation in the same work in which they studied unsubstituted imidazole.[2] The results observed here are qualitatively the same as those observed in the Triplex study.

Results and Discussion

The 5'-dimethoxytrityl phosphoramidite of unsubstituted pyrazole (**5**) was synthesized as shown in Figure 3, starting from protected 1-α-chloro-2-deoxyriboside **1**.[3]

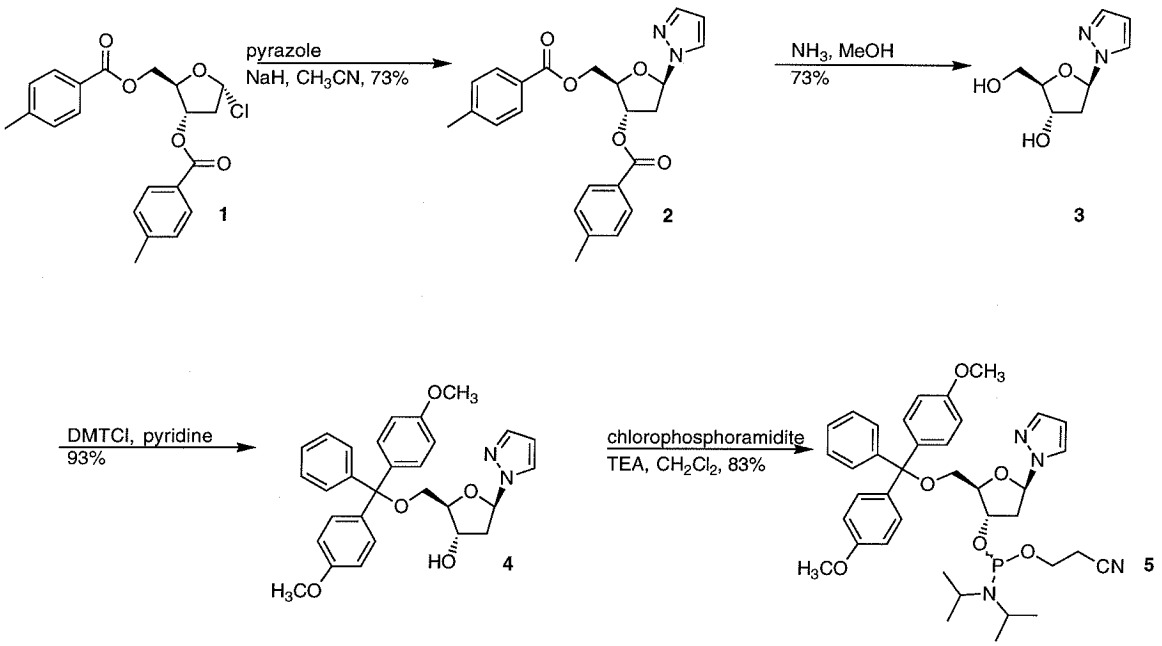


Figure 3. Synthesis of pyrazole nucleoside phosphoramidite **5**.

The corresponding protected phosphoramidites for amino substituted pyrazoles were synthesized from a common precursor as shown in Figure 4.

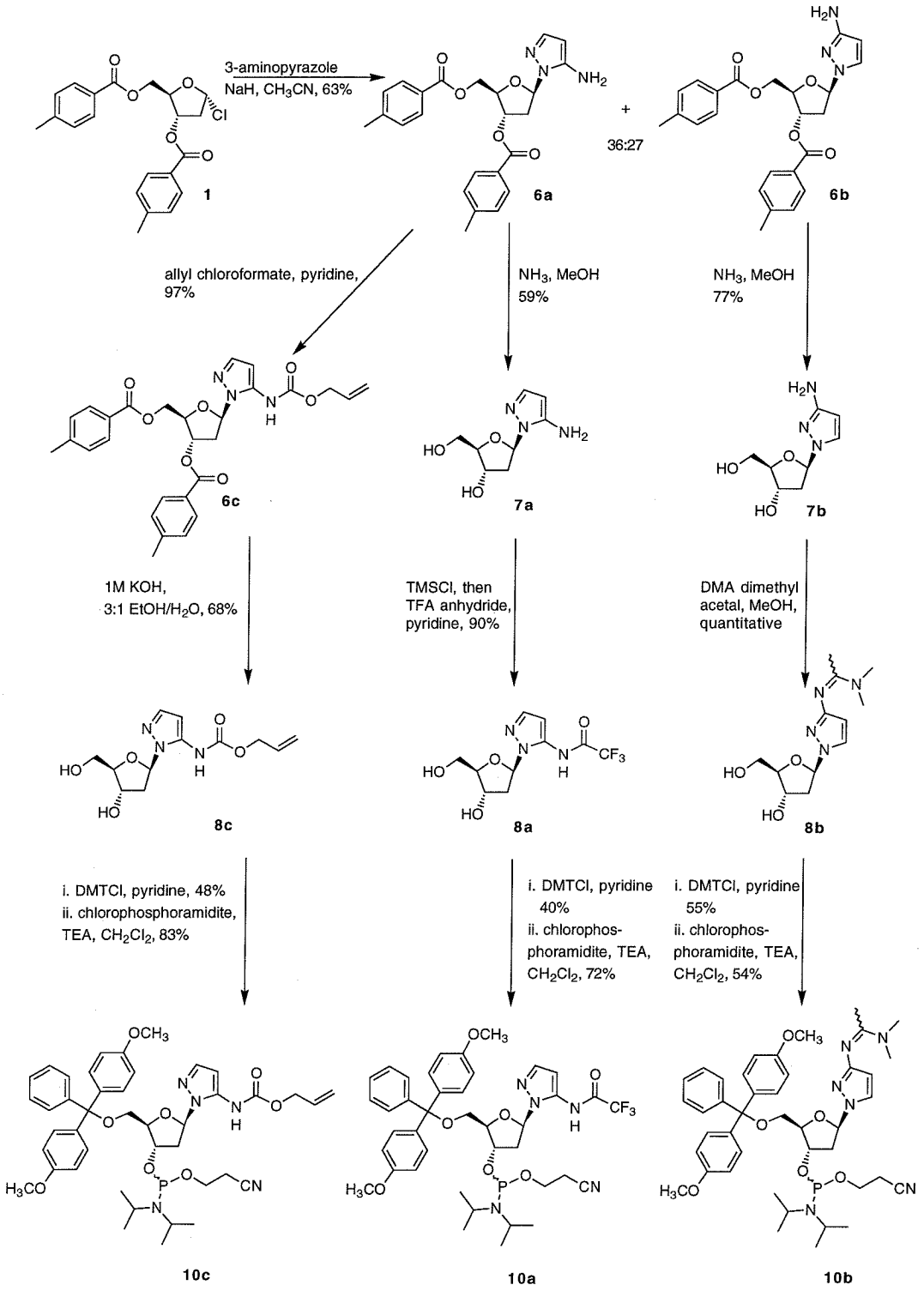


Figure 4. Synthesis of phosphoramidites **10a-c**.

Glycosylation of 3-aminopyrazole with **1** produced a 4:3 mixture of isomers **6a** and **6b**. These were readily separable by silica chromatography and their structures assigned by nOe difference spectrometry on the nucleosides **7a** and **7b** (see Experimental section).[4] The 3-amino isomer **7b** was protected as the N,N-dimethylacetamidine **8b**, and the 5-amino isomer **7a** was protected as the trifluoroacetamide **8a** after the acetamidine of this isomer was found to be unstable.[4] It was later found by MALDI-TOF mass spectrometry of the corresponding oligonucleotide that the trifluoroacetamide remained intact after standard ammonia deprotection had been performed. Thus the deprotected 5-aminopyrazole isomer was later accessed through allyl carbamate-protected nucleoside **8a** and palladium-mediated deprotection using Noyori's protocols for allyl-protected oligonucleotides.[5]

Phosphoramidites **10a-c** were incorporated into purine rich oligonucleotides of the standard sequence described in previous chapters (Figure 5).

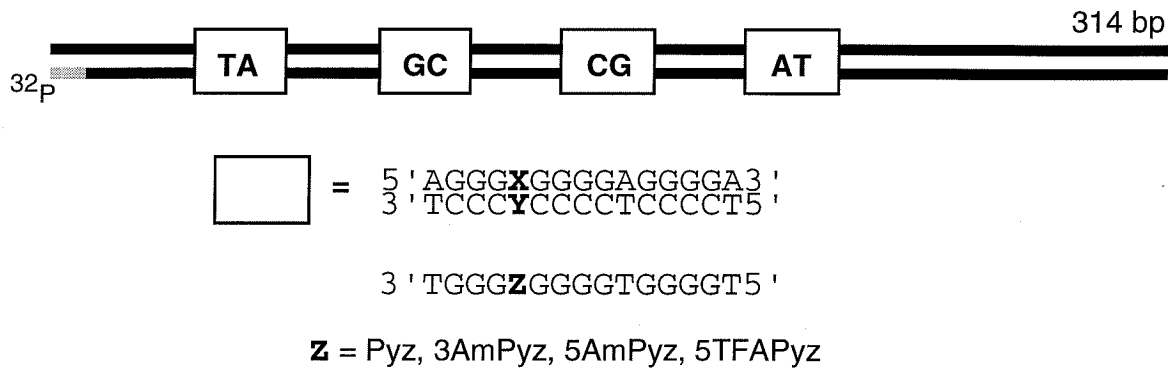


Figure 5. Sequences of the four target sites on 3'-labeled 314 bp *Afl* II/*Fsp* I restriction fragment of pRSPEC1. Below are the sequences of third strand oligonucleotides containing the pyrazole substitutions.

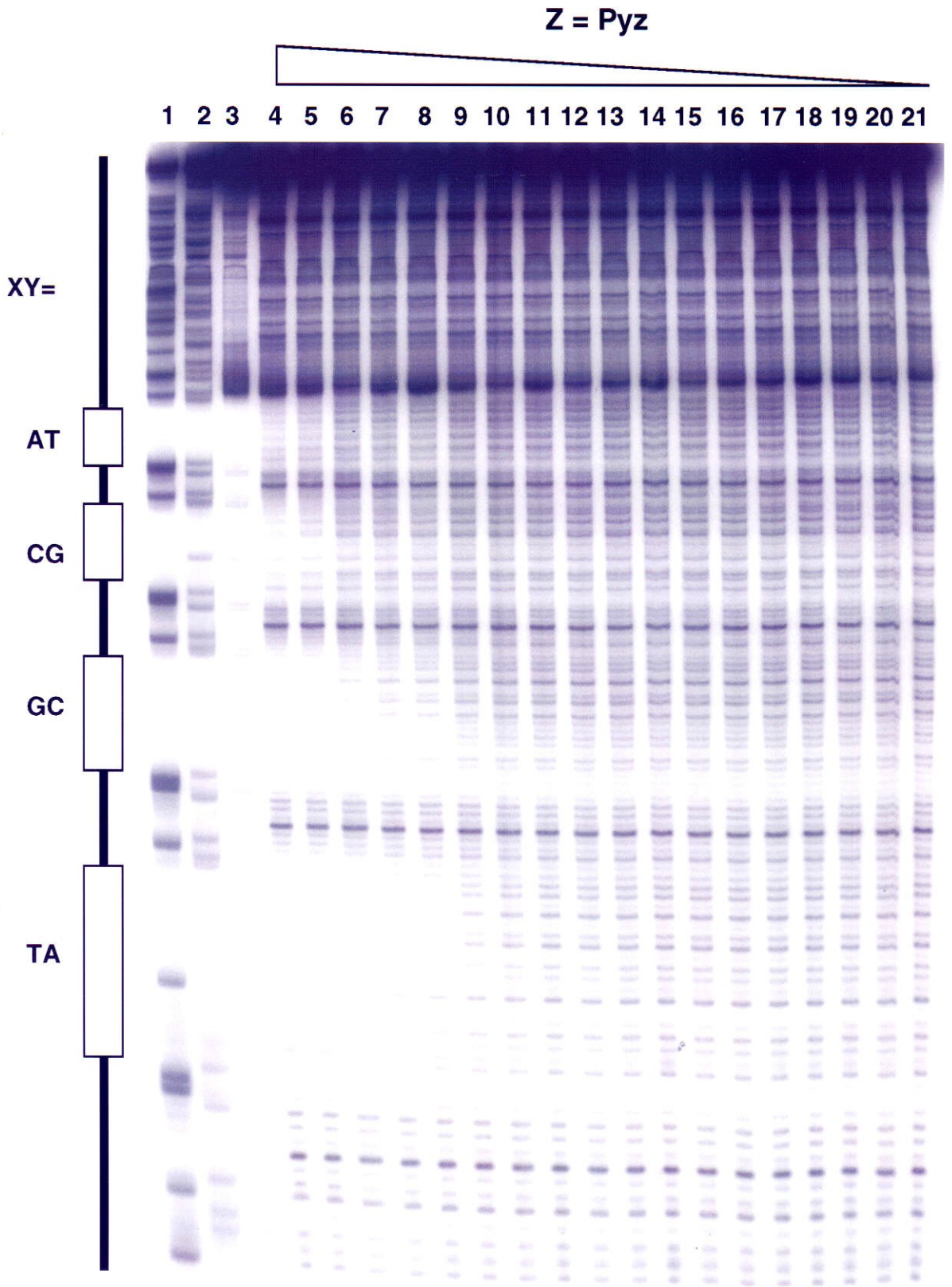
The energetics corresponding to the Z•XY triplet interactions where Z= Pyz, 3AmPyz, 5AmPyz, and 5TFAPyz are summarized in Table 1. Autoradiograms of gels (Figures 6-9) demonstrate the specificity of each novel nucleoside. The putative triplet structures for these sixteen Z•XY triplets are shown in Figures 10-13.

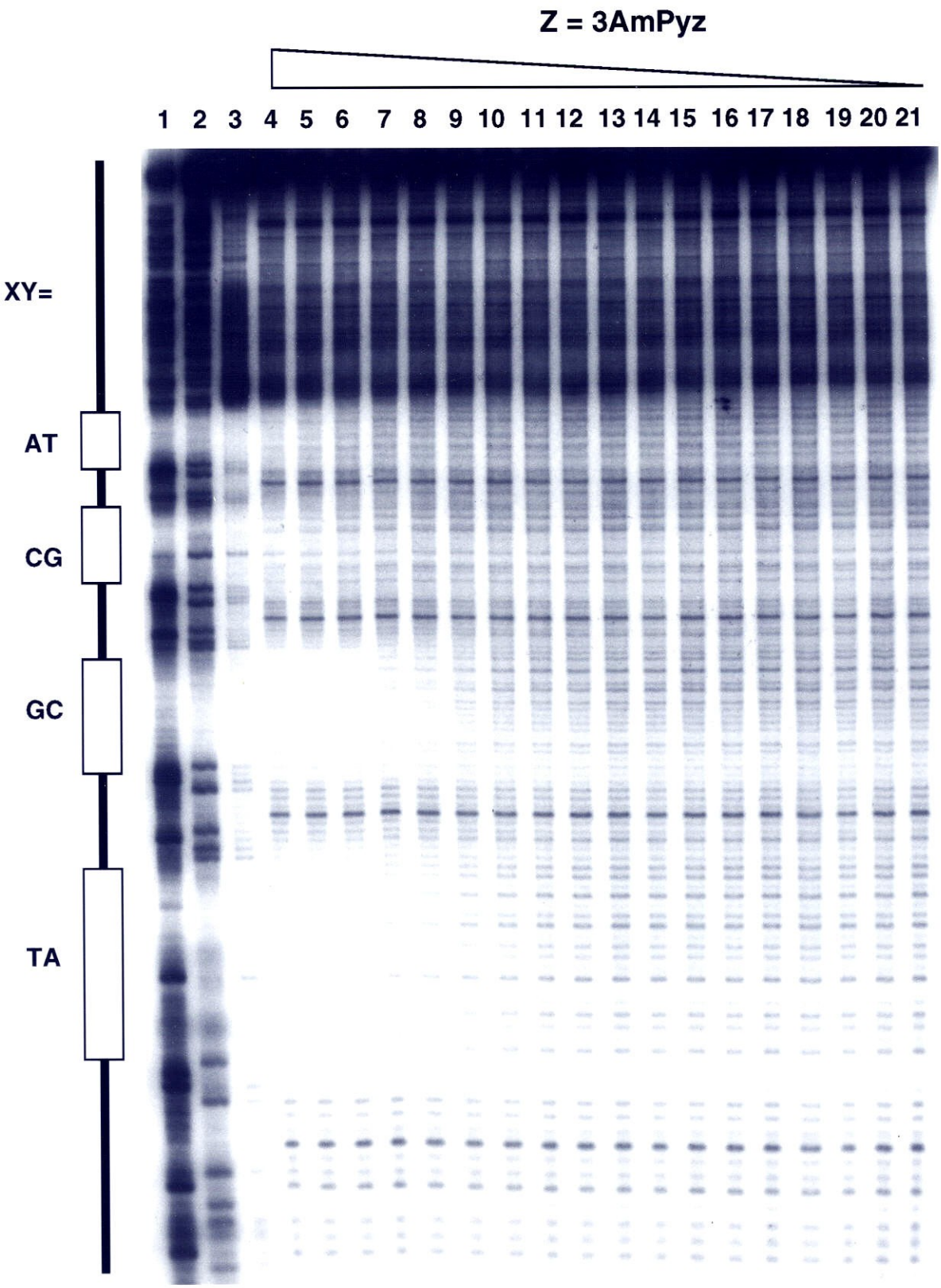
Table 1. Association constants (K_T) for the formation of sixteen triple helical complexes containing the Z•XY triplets at 37°C, 10 mM NaCl, 3 mM $MgCl_2$, 50 mM tris acetate, pH 7.4.[a,b]

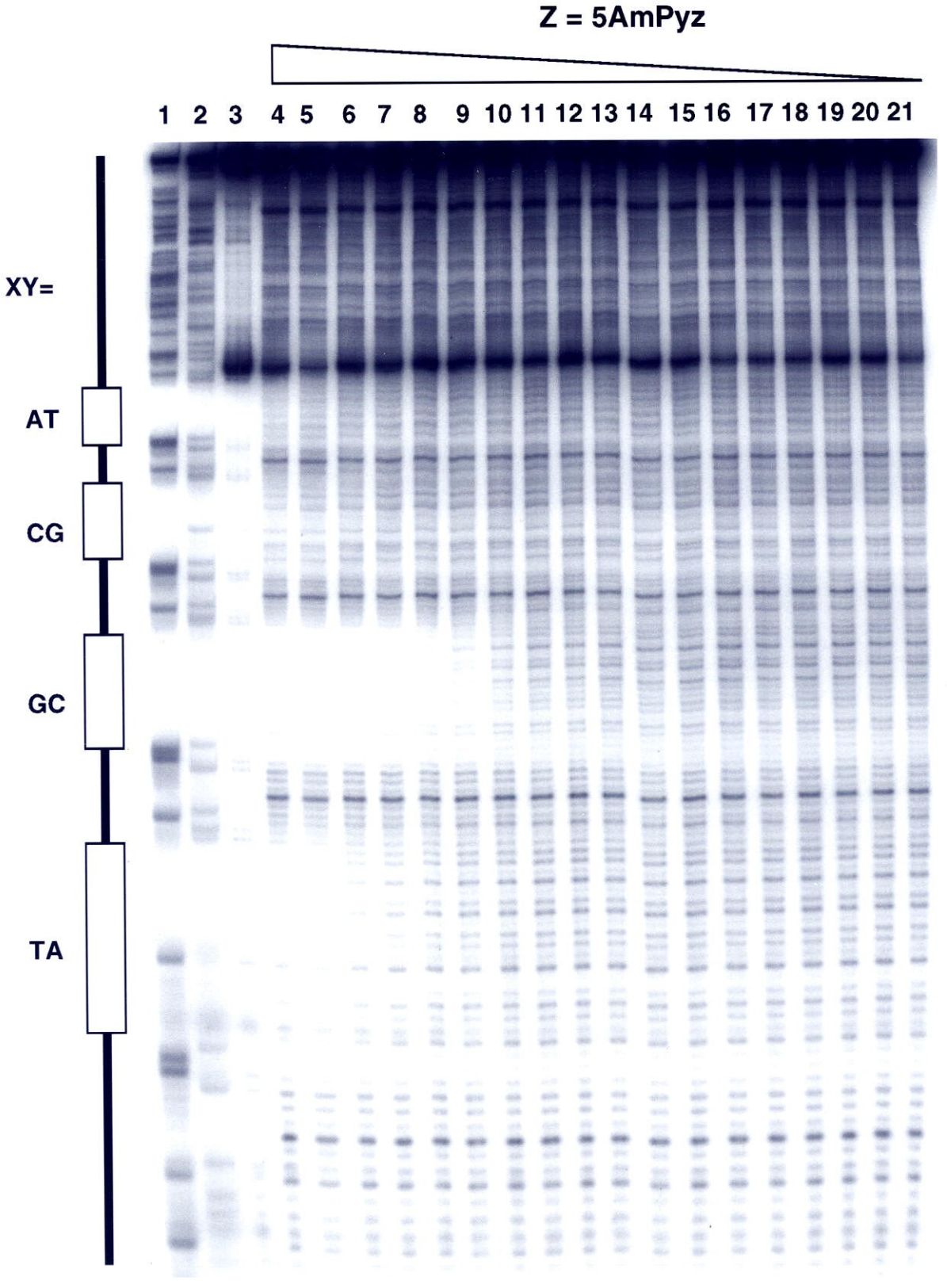
XY	Z= Pyz	3AmPyz	5AmPyz	5TFAPyz
AT	3.0 (±0.5) x 10^5	3.0 (±1.1) x 10^5	1.0 (±0.5) x 10^5	<1.0 x 10^5
GC	1.4 (±0.2) x 10^6	2.1 (±0.4) x 10^6	5.8 (±0.3) x 10^6	1.6 (±0.1) x 10^6
CG	2.8 (±0.7) x 10^5	2.9 (±0.8) x 10^5	1.1 (±0.7) x 10^5	<1.0 x 10^5
TA	2.2 (±0.3) x 10^6	1.5 (±0.2) x 10^6	8.7 (±0.4) x 10^5	3.1 (±1.0) x 10^5

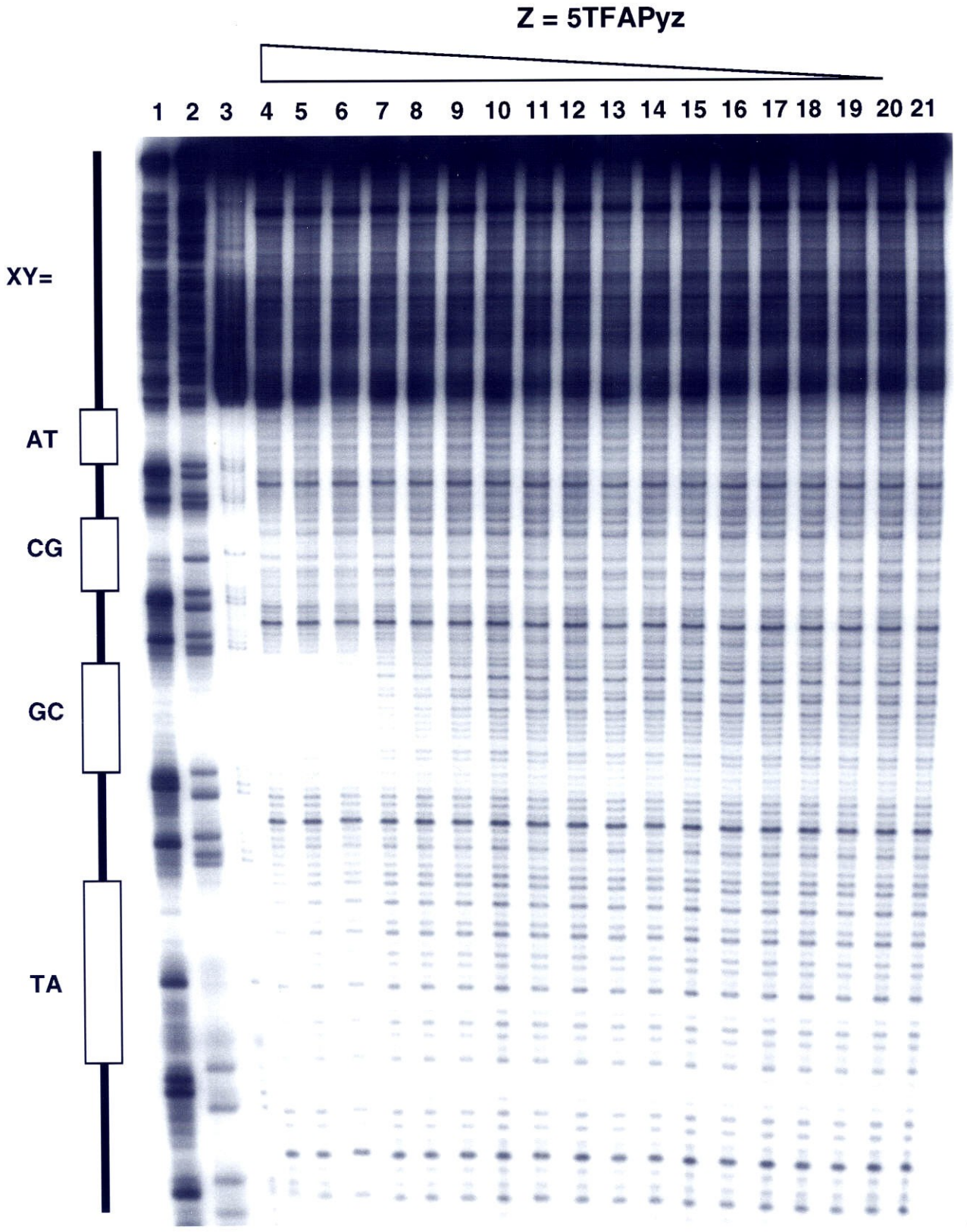
[a] K_T values are reported as the mean (± the standard error of the mean) of three measurements. The K_T values are reported in units of M^{-1}. [b] The identity of the base Z is indicated across the top of the columns; the identity of the Watson Crick base pair XY is indicated on the left side of the rows.

Figures 6-9. Autoradiograms of 8% denaturing polyacrylamide gels used to separate fragments generated by DNase I digestion in quantitative footprint titration experiments with oligonucleotides containing Pyz (Figure 6), 3AmPyz (Figure 7), 5AmPyz (Figure 8), and 5TFAPyz (Figure 9). The four target sites on the 314 bp 3'-end labeled restriction fragment are labeled as XY= AT, CG, GC, and TA respectively, going from the top to the bottom of each gel. (Lane 1) Products of an adenine-specific reaction. (Lane 2) Products of a guanine-specific reaction. (Lane 3) Intact 3' labeled DNA after incubation in the absence of third strand oligonucleotide. (Lanes 4-21) DNase I digestion products obtained in the presence of varying concentrations of oligonucleotide: 8 μM (lane 4); 4 μM (lane 5); 2 μM (lane 6); 1 μM (lane 7); 800 nM (lane 8); 400 nM (lane 9); 200 nM (lane 10); 100 nM (lane 11); 80 nM (lane 12); 40 nM (lane 13);20 nM (lane 14); 10 nM (lane 15); 8 nM (lane 16); 4 nM (lane 17); 2 nM (lane 18); 1 nM (lane 19); 800 pM (lane 20); no oligonucleotide (lane 21).

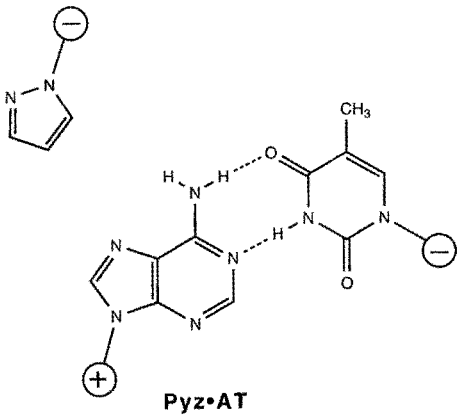

Z = Pyz
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
XY=
AT
CG
GC
TA


Z = 3AmPyz
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
XY=
AT
CG
GC
TA

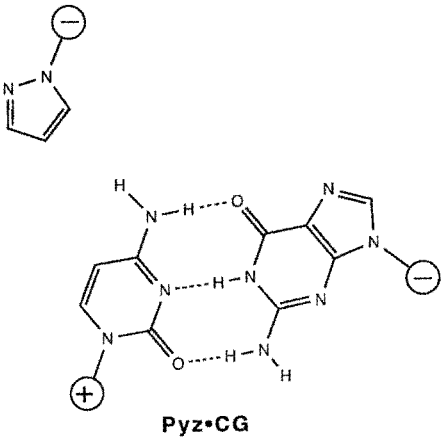

Z = 5AmPyz
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
XY=
AT
CG
GC
TA


Z = 5TFAPyz
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
XY=
AT
CG
GC
TA

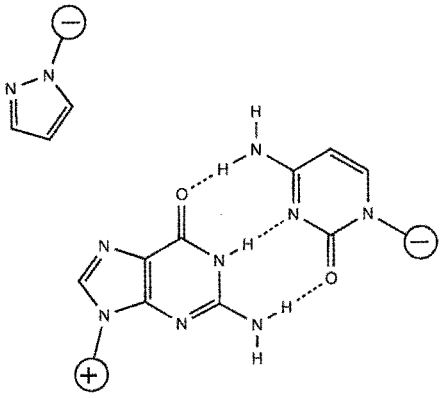
Figures 10-13. Putative triplet structures for Pyz•XY (Figure 10), 3AmPyz•XY (Figure 11), 5AmPyz•XY (Figure 12), and 5TFAPyz•XY (Figure 13).



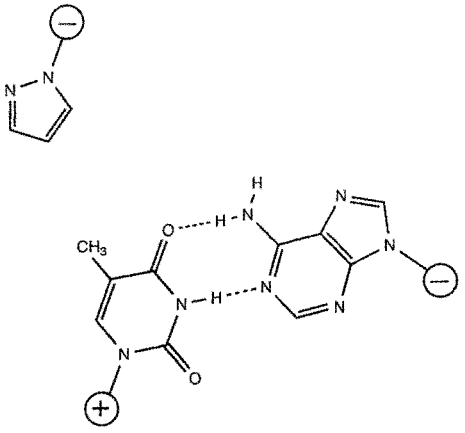
Pyz•AT



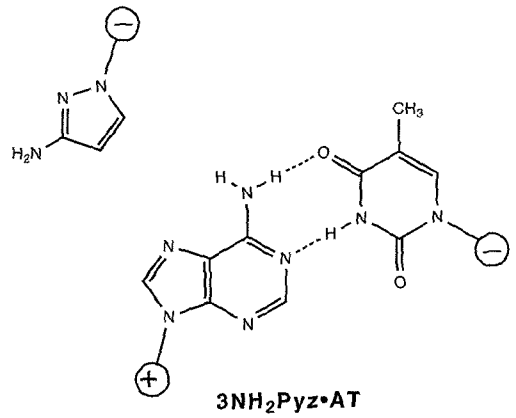
Pyz•CG



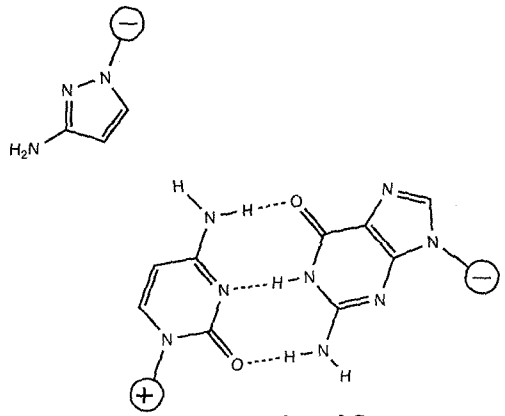
Pyz•GC



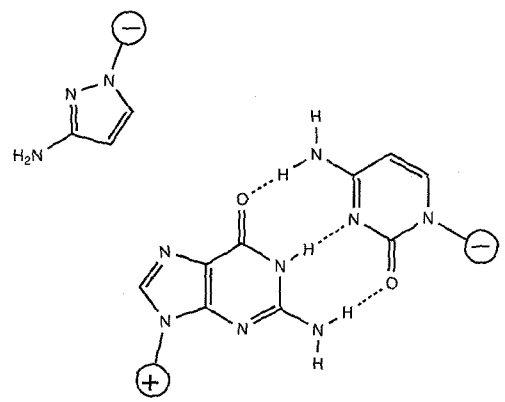
Pyz•TA



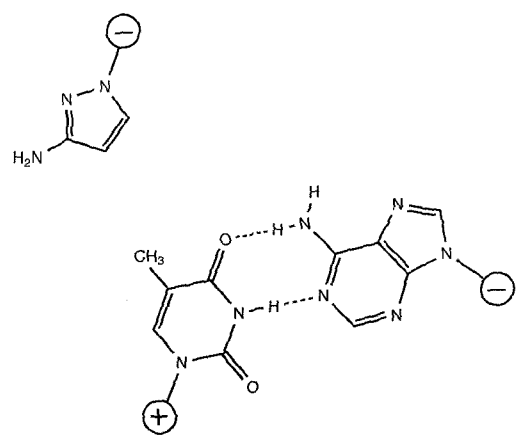
3NH$_2$Pyz•AT



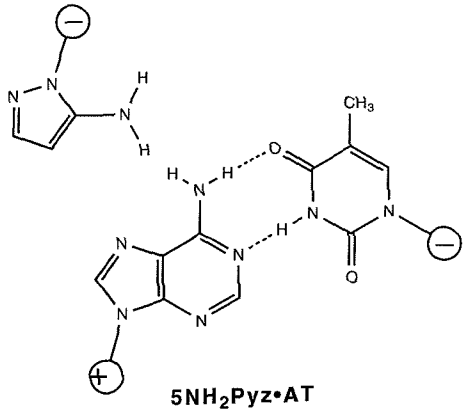
3NH$_2$Pyz•CG



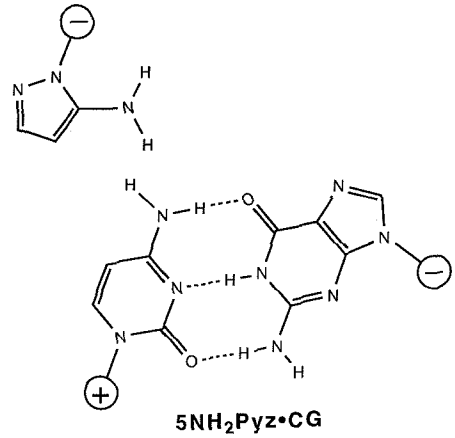
3NH$_2$Pyz•GC



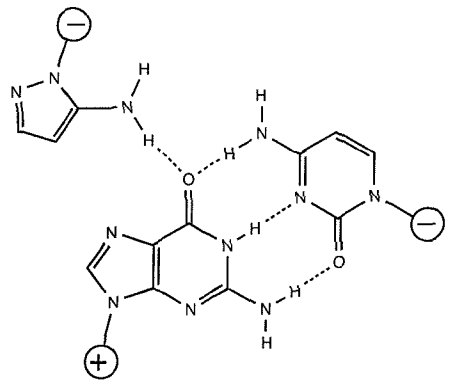
3NH$_2$Pyz•TA



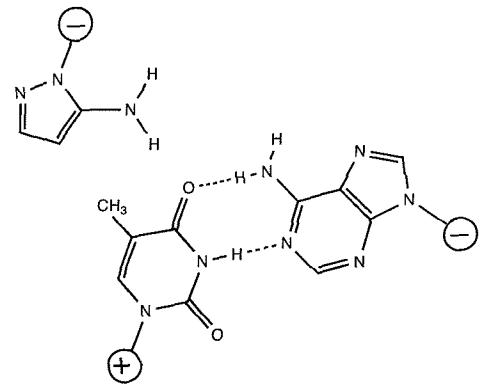
5NH$_2$Pyz•AT



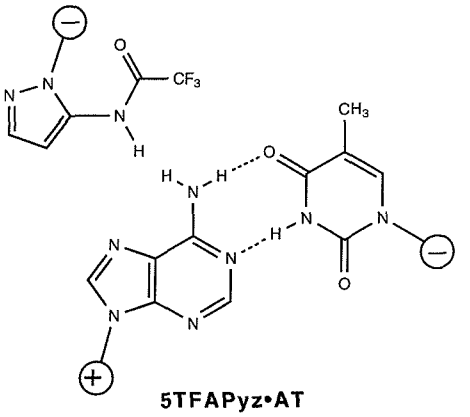
5NH$_2$Pyz•CG



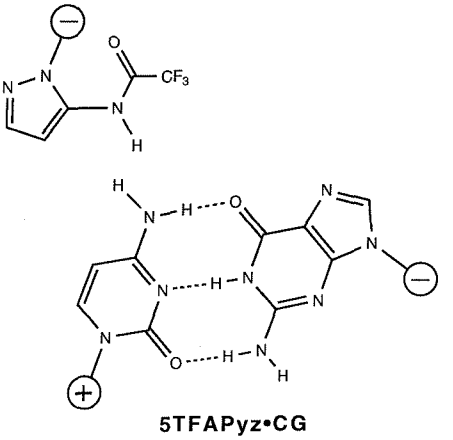
5NH$_2$Pyz•GC



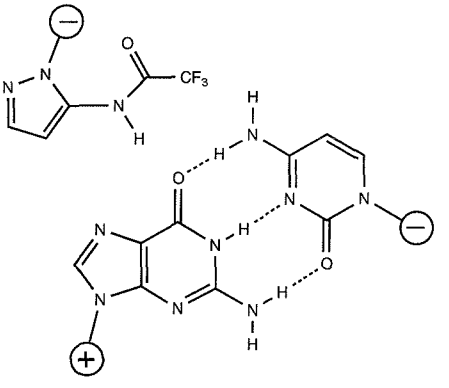
5NH$_2$Pyz•TA



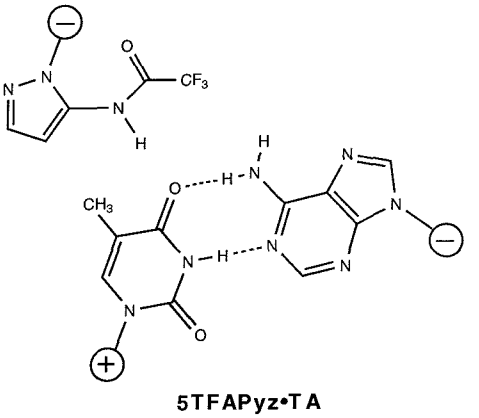
5TFAPyz•AT



5TFAPyz•CG



5TFAPyz•GC



5TFAPyz•TA

As predicted, with elimination of the protonation problem of imidazoles, affinity for GC has been reduced by 10-fold in the case of unsubstituted pyrazole (Pyz). As a result, in the case of this novel nucleoside, for the first time, the measured affinity for TA is higher (2.2×10^6) than for any other base pair. Unfortunately, the difference between TA affinity and GC affinity is less than a factor of two, and the overall energetics are not exceptionally strong. 3AmPyz binding is virtually identical to that of Pyz, consistent with the amino group pointing away from the thymine and presenting the same face as does Pyz, as shown in Figure 11.

It was hoped that with 5AmPyz the designed structure would be able to specifically form a hydrogen bond between its amino group and the O4 carbonyl of thymine (Figure 12). Importantly, it was expected that analogous hydrogen bonding to the O6 carbonyl of guanine in a GC base pair would be disfavored due to the resultant distortion of the third strand backbone position this would cause. Figure 14 shows the different backbone positioning of the third strand for 2-aminopurine•GC, which was demonstrated in chapter 3 to likely form a hydrogen bonded triplet, compared to 5AmPyz•GC. Counter to expectations, the data in Table 1 suggest that 5AmPyz forms a specific hydrogen bond to GC, as reflected in the 4-5 fold increase in affinity compared to Pyz•GC, and it does not form a similar hydrogen bond to TA, as reflected by the 3-fold decrease in affinity for this triplet compared to Pyz•TA. Addition of the trifluoroacetyl group in 5TFAPyz resulted in a decrease in affinity to all four base pairs relative to 5AmPyz.

The results for 5AmPyz, in relation to those of the other compounds tested, suggest that specific hydrogen bonding to the O4 carbonyl of a TA base pair will be extremely difficult to design. It appears that the corresponding O6 carbonyl of GC is much more accessible. Indeed, the only case where hydrogen bonding to the thymine carbonyl has been demonstrated is the G•TA triplet in the pyrimidine motif, which is made possible by a

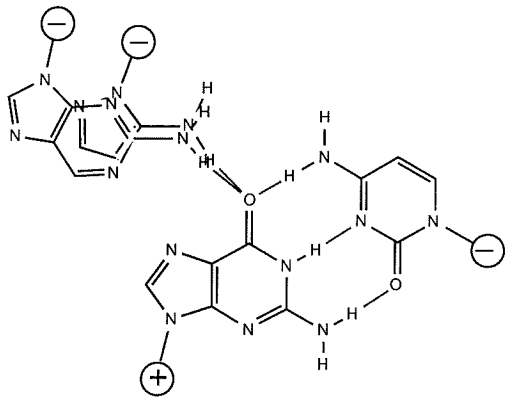


Figure 14. Comparison of structures of the AP•GC and 5AmPyz•GC triplets, which have similar affinities and likely both contain the illustrated hydrogen bonds. Note the significantly different placement of the backbones in each case.

severe tilting of the third strand base to accommodate the methyl group of thymine.[6] Because of the methyl group, accessibility for hydrogen bonding is severely hampered, and even if this problem were to be overcome, the successful design must also disfavor binding to GC in order to result in specificity.

Conclusion

A series of substituted pyrazole nucleosides was designed and incorporated into purine-rich oligonucleotides to investigate the possibility of specific recognition of TA base pairs in the purine motif by these five-membered ring heterocycles. The results obtained

through quantitative DNase I footprint titration experiments indicate that the pyrazole structure does result in preferential binding to TA, but with only a slight preference and relatively low affinity. Substitution of an amino group at the pyrazole 5- position resulted in specific binding to GC rather than TA, suggesting that the O6 carbonyl of GC is more accessible than the O4 carbonyl of TA. This accessibility problem will make specific recognition by hydrogen bonding to the thymine carbonyl exceptionally difficult.

Experimental section

General methods in the synthesis of phosphoramidites and oligonucleotides were performed as described in the previous chapters. Oligonucleotide purifications were performed as previously described in chapters 2 and 3. Removal of the allyl carbamate protecting group from CPG-bound oligonucleotide with $Pd_2(dba)_3$ followed Noyori's protocols for deprotection of allyl-protected oligonucleotides.[5] Because this series of pyrazole nucleosides do not have useful chromophores for HPLC analysis of digested oligonucleotides (λ_{max} = 218 nm, 232 nm, and 236 nm for pyrazole, 5-aminopyrazole, and 3-aminopyrazole respectively), incorporation and integrity of novel nucleosides was measured by MALDI-TOF mass spectrometry on intact 15mer oligonucleotides. MALDI MS: Z=Pyz 4715.5 (calcd. 4717.1), Z=3AmPyz 4731.8 (calcd. 4732.1), Z=5AmPyz 4734.1 (calcd. 4732.1), Z=5TFAPyz 4830.8 (calcd. 4828.1).

1-[2-deoxy-3,5-*O*-toluoyl-β-D-ribofuranosyl]pyrazole (2): Pyrazole (350 mg, 5.14 mmol) was added to 50 mL acetonitrile. With stirring at room temperature sodium hydride (136mg, 5.65 mmol as a 60% dispersion in mineral oil) was added. After completion of bubbling, 40 minutes, 2-deoxy-3,5-*O*-toluoyl-α-D-ribose **(1)**[3] (2.0 g, 5.14 mmol) was added. The mixture was stirred for 1 hour, then concentrated. The residue was dissolved in 150 mL dichloromethane, the mixture extracted with water, dried, and concentrated. Silica chromatography (2%MeOH/CH_2Cl_2) yielded a white solid (1.57 g, 73%). TLC (5% MeOH/CH_2Cl_c) R_f 0.80; ^{1}H NMR ($CDCl_3$) δ 7.95 (m, 4H), 7.62 (m, 1H), 7.59 (m, 1H), 7.24 (m, 4H), 6.31 (m, 1H), 6.28 (t, 1H), 5.83 (m, 1H), 4.60 (m, 3H), 3.22 (m, 1H), 2.68 (m, 1H), 2.44 (s, 3H), 2.41 (s, 3H).

1-[2-deoxy-β-D-ribofuranosyl]pyrazole (3): **2** (1.0 g, 2.38 mmol) was dissolved in 60 mL ammonia-saturated methanol, stirred at room temperature for 24 hours, then concentrated. Product was purified by silica chromatography (10%MeOH/CH_2Cl_2) to furnish a colorless oil (320 mg, 73%). TLC (5% MeOH/CH_2Cl_c) R_f 0.15; ^{1}H NMR (DMSO-d_6) δ 7.90 (d, 1H, J=2.4Hz), 7.48 (d, 1H, J=1.5Hz), 6.25 (m, 1H), 6.08 (t, 1H, J=6.3 Hz), 5.20 (d, 1H, J=4.3Hz), 4.83 (m, 1H), 4.32 (m, 1H), 3.77 (m, 1H), 3.46 (m, 1H), 3.39 (m, 1H), 2.53 (m, 1H), 2.18 (m, 1H).

1-[2-deoxy-5-*O*-(4,4'-dimethoxytrityl)-β-D-ribofuranosyl]pyrazole (4): Pyrazole nucleoside **3** (160 mg, 0.87 mmol) and DMT chloride (303 mg, 0.90 mmol) were dried overnight under vacuum. 8 mL pyridine was added, and the reaction mixture was stirred at room temperature for 8 hours, after which 0.5 mL methanol was added to quench, and the solvent was removed under reduced pressure. Purification by silica chromatography (1:1 hexanes/EtOAc) yielded a white foam (392 mg, 93 %). TLC (1:1 hexanes/EtOAc) R_f 0.30; ^{1}H NMR ($CDCl_3$) δ 7.59 (m, 1H). 7.52 (m, 1H), 7.41 (m, 2H), 7.30 (m, 7H), 6.83 (m, 4H), 6.25 (m, 1H), 6.12 (m, 1H), 4.65 (m, 1H), 4.04 (m, 1H), 3.77 (s, 6H), 3.30 (m, 2H), 2.80 (m, 1H), 2.41 (m, 1H).

Pyrazole nucleoside phosphoramidite (5): 5'-DMT nucleoside **4** (150 mg, 0.31 mmol) was dissolved in 3 mL dichloromethane. Diisopropylethylamine (161 μL, 0.92 mmol) was added, followed by dropwise addition of 2-cyanoethyl-N,N-diisopropylchlorophosphoramidite (86 μL, 0.39 mmol). The reaction was stirred at room temperature for 1 hour, after which it was quenched with 0.2 mL MeOH, diluted with 25 mL dichloromethane, extracted with saturated $NaHCO_3$, then brine, dried, and concentrated. Purification by silica chromatography (2:1 hexanes/EtOAc) provided a

colorless oil (176 mg, 83%). TLC (1:1 hexanes/EtOAc) R_f 0.70; ^{1}H NMR (mixture of diastereomers)($CDCl_3$) δ 7.61 (m, 1H). 7.51 (m, 1H), 7.39 (m, 2H), 7.30 (m, 7H), 6.78 (m, 4H), 6.25 (m, 1H), 6.15 (t, 1H), 4.68 (m, 1H), 4.21 (m, 1H), 3.78 (s, 6H), 3.5-4.0 (m, 4H), 3.22 (m, 2H), 2..81 (m, 1H),2.4-2.7 (m, 3H), 1.17 (m, 12H).

5-Amino-1-[2-deoxy-3,5-*O*-toluoyl-β-D-ribofuranosyl]pyrazole and 3-amino-1-[2-deoxy-3,5-*O*-toluoyl-β-D-ribofuranosyl]pyrazole (6a and 6b): 3-aminopyrazole (855 mg, 10.3 mmol) was added to 90 mL acetonitrile. With stirring at room temperature sodium hydride (453 mg, 11.3 mmol as a 60% dispersion in mineral oil) was added. After completion of bubbling, 1 hour, 2-deoxy-3,5-*O*-toluoyl-α-D-ribose **(1)**[3] (4.0 g, 10.3 mmol) was added. The mixture was stirred for 1 hour, then concentrated. The residue was dissolved in 150 mL dichloromethane, the mixture extracted with water, dried, and concentrated. Silica chromatography (1:2 hexanes/diethyl ether) eluted first the 5-amino isomer (1.6 g light yellow foam, 36%), then the 3-amino isomer (yellow foam, 1.2 g, 27%) TLC (1:2 hexanes/diethyl ether) R_f 0.30 (5-amino isomer), 0.15 (3-amino isomer); nOe experiments to assign structures of the isomers were previously performed by J.R. Parquette: with irradiation of 1'H an nOe of 13.7% was observed to the H5, and no nOe was observed to any aromatic proton in the 5-amino isomer. ^{1}H NMR ($CDCl_3$) (5-amino isomer) δ 7.96 (m, 4H), 7.24 (m, 4H), 7.22 (m, 1H), 6.20 (m, 1H), 5.75 (m, 1H), 5.53 (m, 1H), 4.50 (m, 3H), 3.92 (b, 2H), 3.51 (m, 1H), 2.60 (m, 1H), 2.44 (s, 3H), 2.41 (s, 3H); (3-amino isomer) δ 7.96 (m, 4H), 7.35 (m, 1H), 7.25 (m, 4H), 6.04 (m, 1H), 5.76 (m, 1H), 5.66 (m, 1H), 4.56 (m, 3H), 3.64 (b, 2H), 3.20 (m, 1H), 2.56 (m, 1H), 2.44 (s, 3H), 2.41 (s, 3H); FAB MS M+H 436.1845 (436.1873 calcd. for $C_{24}H_{26}N_3O_5$).

5-Allyloxycarbamoyl-1-[2-deoxy-3,5-*O*-toluoyl-β-D-ribofuranosyl]pyrazole (6c): **6a** (200 mg, 0.46 mmol) was dissolved in 10 mL pyridine at O°C. Allyl chloroformate (49 μL, 0.46 mmol) was added dropwise. The mixture was stirred over ice for 15 minutes, then diluted with 50 mL dichloromethane, extracted with water, dried, and concentrated. Silica chromatography (2:3 hexanes/diethyl ether) yielded a white foam (230 mg, 97%). TLC (2:3 hexanes/diethyl ether) R_f 0.25; ^{1}H NMR ($CDCl_3$) δ 7.95 (m, 4H), 7.43 (m, 1H), 7.24 (m, 4H), 6.90 (b, 1H), 6.21 (m, 1H), 5.90 (m, 1H), 5.75 (m, 1H), 5.2-5.4 (m, 3H), 4.40-4.75 (m, 5H), 3.49 (m, 1H), 2.60 (m, 1H), 2.44 (s, 3H), 2.41 (s, 3H); FAB MS M+H 520.2051 (520.2084 calcd. for $C_{28}H_{30}N_3O_7$).

5-Amino-1-[2-deoxy-β-D-ribofuranosyl]pyrazole (7a): **6a** (1.5 g, 3.44 mmol) was dissolved in 80 mL ammonia-saturated methanol, stirred at room temperature for 48 hours, then concentrated. Product was purified by silica chromatography (8%MeOH/EtOAc) to furnish a colorless gum (406 mg, 59%). TLC (10% MeOH/EtOAc) R_f 0.40; ^{1}H NMR (DMSO-d_6) δ 7.09 (d, 1H, J=1.5Hz), 6.03 (t, 1H, J=6.6 Hz), 5.39 (b, 2H), 5.21 (d, 1H, J=1.5Hz), 5.10 (m, 2H), 4.34 (m, 1H), 3.76 (m, 1H), 3.45 (m, 1H), 3.36 (m, 1H), 2.63 (m, 1H), 2.06 (m, 1H); HRMS M+H 199.950 (199.0957 calcd. for $C_8H_{13}N_3O_3$); FAB MS M+H 200.1033 (200.1035 calcd. for $C_8H_{14}N_3O_3$).

3-Amino-1-[2-deoxy-β-D-ribofuranosyl]pyrazole (7b): **6b** (1.0 g, 2.30 mmol) was dissolved in 60 mL ammonia-saturated methanol, stirred at room temperature for 48 hours, then concentrated. Product was purified by silica chromatography (15%MeOH/CH_2Cl_2) to furnish a light yellow oil (354 mg, 77%). TLC (8% MeOH/CH_2Cl_2) R_f 0.10; ^{1}H NMR (DMSO-d_6) δ 7.46 (d, 1H, J=2.4Hz), 5.80 (t, 1H, J=6.6 Hz), 5.39 (d, 1H, J=2.4Hz), 5.10 (d, 1H), 4.93 (t, 1H), 4.71 (b, 2H), 4.26 (m, 1H),

3.72 (m, 1H), 3.44 (m, 1H), 3.38 (m, 1H), 2.53 (m, 1H), 2.06 (m, 1H); FAB MS M+H 200.1028 (200.1035 calcd. for $C_8H_{14}N_3O_3$).

5-Trifluoroacetamido-1-[2-deoxy-β-D-ribofuranosyl]pyrazole (8a): 7a (200 mg, 1.0 mmol) was dissolved in 15 mL pyridine over ice. Trimethylsilyl chloride (634 μL, 5.0 mmol) was added dropwise and stirred for 30 minutes. Trifluoroacetic anhydride (706 μL, 5.0 mmol) was then added, and the mixture stirred for 2 hours, after which 3 mL water was slowly added. After 30 minutes further stirring, the mixture was concentrated and purified by silica chromatography to yield an off-white solid (265 mg, 90%). TLC (5% MeOH/EtOAc) R_f 0.50; ^{1}H NMR (DMSO-d_6) δ 7.5-7.7 (m, 3H), 6.17 (t, 1H, J=6.6 Hz), 5.10 (vb, 3H), 4.30 (m, 1H), 3.77 (m, 1H), 3.42 (m, 1H), 3.33 (m, 1H), 2.60 (m, 1H), 2.09 (m, 1H); FAB MS M+H 295.0794 (295.0780 calcd. for $C_{10}H_{12}N_3O_4F_3$).

3-Dimethylacetamidino-1-[2-deoxy-β-D-ribofuranosyl]pyrazole (8b): 7b (300 mg, 1.51 mmol) was dissolved in 1.7 mL MeOH, and N,N-dimethylacetamide dimethyl acetal (1.1 mL, 7.6 mmol) was added. The mixture was stirred at room temperature for 12 hours, then concentrated under reduced pressure to yield 400 mg light yellow oil. TLC (10% MeOH/EtOAc) R_f 0.05; ^{1}H NMR (DMSO-d_6) δ 7.64 (m, 1H), 5.92 (t, 1H, J=6.6 Hz), 5.57 (m, 1H), 5.14 (b, 1H), 4.97 (b, 1H), 4.30 (m, 1H), 3.75 (m, 1H), 3.40 (m, 2H), 2.93 (s, 6H), 2.53 (m, 1H), 2.15 (m, 1H), 1.91 (s, 3H); FAB MS M+H 269.1611 (269.1614 calcd. for $C_{12}H_{21}N_4O_3$)

5-Allyloxycarbamoyl-1-[2-deoxy-β-D-ribofuranosyl]pyrazole (8c): 6c (520 mg, 1 mmol) was dissolved in 4 mL absolute ethanol, then 4 mL 2M KOH in 1:1 ethanol/water was added. The mixture was stirred at 0°C for 15 minutes. The KOH was neutralized with 500 mg NH_4Cl, and the mixture concentrated, using EtOH to

azeotropically remove water. Purification by silica chromatography (4% MeOH/EtOAc) provided a light yellow oil (190 mg, 68%). TLC (5% MeOH/EtOAc) R_f 0.65; ^{1}H NMR (CD_3OD) δ 7.45 (d, 1H, J=1.8Hz), 6.23 (d, 1H, J=1.8Hz), 6.16 (t, 1H, J=6.6 Hz), 5.95 (m, 1H), 5.2-5.4 (m, 2H), 4.64 (m, 2H), 4.52 (m, 1H), 3.96 (m, 1H), 3.72 (m, 1H), 3.59 (m, 1H), 2.75 (m, 1H), 2.28 (m, 1H).

5-Trifluoroacetamido-1-[2-deoxy-5-*O*-(4,4'-dimethoxytrityl)-β-D-ribofuranosyl]pyrazole (9a): Protected nucleoside **8a** (120 mg, 0.41 mmol) and DMT chloride (206 mg, 0.62 mmol) were dried overnight under vacuum. 9 mL pyridine was added, and the reaction mixture was stirred at room temperature for 6 hours, after which 0.5 mL methanol was added to quench, and the solvent was removed under reduced pressure. Purification by silica chromatography (1:1 hexanes/EtOAc) yielded a light yellow foam (61 mg, 40 %). TLC (1:1 hexanes/EtOAc) R_f 0.15; ^{1}H NMR (acetone-d_6) δ 7.57 (m, 1H), 7.36 (m, 2H), 7.23 (m, 8H), 7.00 (b, 1H), 6.76 (m, 4H), 6.08 (m, 1H), 6.12 (m, 1H), 4.60 (m, 1H), 4.35 (b, 1H), 4.05 (m, 1H), 3.75 (s, 6H), 3.13 (m, 2H), 2.91 (m, 1H), 2.30 (m, 1H).

3-Dimethylacetamidino-1-[2-deoxy-5-*O*-(4,4'-dimethoxytrityl)-β-D-ribofuranosyl]pyrazole (9b): Protected nucleoside **8b** (150 mg, 0.56 mmol) and DMT chloride (285 mg, 0.84 mmol) were dried overnight under vacuum. 10 mL pyridine was added, and the reaction mixture was stirred at room temperature for 8 hours, after which 0.5 mL methanol was added to quench, and the solvent was removed under reduced pressure. Purification by silica chromatography (20% MeOH/EtOAc) yielded a clear, glass (172 mg, 55 %). TLC (15% MeOH/EtOAc) R_f 0.10; ^{1}H NMR (CD_3OD) δ 7.72 (m, 1H), 7.41 (m, 2H), 7.28 (m, 7H), 6.76 (m, 4H), 6.02 (m, 1H), 6.01 (m, 1H), 4.60 (m,

1H), 4.03 (m, 1H), 3.70 (s, 6H), 3.11 (s, 6H), 2.72 (m, 1H), 2.35 (m, 1H), 1.81 (s, 3H); FAB MS M+H 571.2908 (571.2920 calcd. for $C_{33}H_{39}N_4O_5$).

5-Allyoxycarbamoyl-1-[2-deoxy-5-*O*-(4,4'-dimethoxytrityl)-β-D-ribofuranosyl]pyrazole (9c): Protected nucleoside **8c** (100 mg, 0.35 mmol) and DMT chloride (131 mg, 0.53 mmol) were dried overnight under vacuum. 10 mL pyridine was added, and the reaction mixture was stirred at room temperature for 8 hours, after which 0.5 mL methanol was added to quench, and the solvent was removed under reduced pressure. Purification by silica chromatography (1:1 hexanes/EtOAc) yielded a white foam (98 mg, 48 %). TLC (1:2 hexanes/EtOAc) R_f 0.60; ^{1}H NMR (acetone-d_6) δ 8.60 (b, 1H), 7.42 (m, 2H), 7.32 (m, 1H), 7.25 (m, 7H), 6.88 (m, 4H), 6.23 (m, 1H), 6.17 (m, 1H), 5.95 (m, 1H), 5.2-5.4 (m, 2H), 4.95 (b, 1H), 4.62 (m, 2H), 4.61 (m, 1H), 4.02 (m, 1H), 3.75 (s, 6H), 3.11 (m, 2H), 2.93 (m, 1H), 2.26 (m, 1H).

5-Trifluoroacetamidopyrazole nucleoside phosphoramidite (10a): 5'-DMT nucleoside **9a** (80 mg, 0.134 mmol) was dissolved in 2.5 mL dichloromethane. Diisopropylethylamine (70 μL, 0.40 mmol) was added, followed by dropwise addition of 2-cyanoethyl-N,N-diisopropylchlorophosphoramidite (38 μL, 0.167 mmol). The reaction was stirred at room temperature for 1 hour, after which it was quenched with 0.2 mL MeOH, diluted with 25 mL dichloromethane, extracted with saturated $NaHCO_3$, then brine, dried, and concentrated. Purification by silica chromatography (3:2 hexanes/EtOAc) provided a white foam (77 mg, 72%). TLC (3:2 hexanes/EtOAc) R_f 0.35; ^{1}H NMR (mixture of diastereomers)($CDCl_3$) δ 7.63 (m, 1H), 7.37 (m, 2H), 7.27 (m, 7H), 6.85 (m, 4H), 6.39 (m, 2H), 6.06 (t, 1H), 4.75 (m, 1H), 4.25 (m, 1H), 3.80 (s, 6H), 3.5-4.0 (m, 4H), 3.35 (m, 2H), 2..96 (m, 1H),2.4-2.7 (m, 3H), 1.18 (m, 12H).

3-Dimethylacetamidinopyrazole nucleoside phosphoramidite (10b): 5'-DMT nucleoside **9b** (83 mg, 0.145 mmol) was dissolved in 2.5 mL dichloromethane. Diisopropylethylamine (76 μL, 0.435 mmol) was added, followed by dropwise addition of 2-cyanoethyl-N,N-diisopropylchlorophosphoramidite (40 μL, 0.18 mmol). The reaction was stirred at room temperature for 1 hour, after which it was quenched with 0.2 mL MeOH, diluted with 25 mL dichloromethane, extracted with saturated $NaHCO_3$, then brine, dried, and concentrated. Purification by silica chromatography (10% MeOH/5% TEA/ EtOAc) provided a colorless oil (61 mg, 54%). TLC (20% MeOH) R_f 0.30; ^1H NMR (mixture of diastereomers)($CDCl_3$) δ 7.45 (m, 1H), 7.43 (m, 2H), 7.27 (m, 7H), 6.77 (m, 4H), 5.97 (t, 1H), 5.72 (m, 1H), 4.65 (m, 1H), 4.13 (m, 1H), 3.80 (s, 6H), 3.5-4.0 (m, 4H), 3.20 (m, 2H), 3.02 (s, 6H), 2..83 (m, 1H),2.4-2.7 (m, 3H), 1.81 (s, 3H), 1.14 (m, 12H). FAB MS M+H 771.3963 (771.3999 calcd. for $C_{42}H_{56}N_6O_6$)

5-Allyloxycarbamoylpyrazole nucleoside phosphoramidite (10c): 5'-DMT nucleoside **9c** (90 mg, 0.154 mmol) was dissolved in 2 mL dichloromethane. Diisopropylethylamine (80 μL, 0.45 mmol) was added, followed by dropwise addition of 2-cyanoethyl-N,N-diisopropylchlorophosphoramidite (43 μL, 0.19 mmol). The reaction was stirred at room temperature for 1 hour, after which it was quenched with 0.2 mL MeOH, diluted with 25 mL dichloromethane, extracted with saturated $NaHCO_3$, then brine, dried, and concentrated. Purification by silica chromatography (2:1 hexanes/EtOAc) provided a white foam (101 mg, 83%). TLC (2;1 hexanes/EtOAc) R_f 0.35; ^1H NMR (mixture of diastereomers)($CDCl_3$) δ 7.2-7.5 (m, 10H), 6.84 (m, 4H), 6.23 (m, 1H), 5.95 (t, 1H), 5.93 (m, 1H), 5.20 (m, 2H), 4.75 (m, 1H), 4.60 (m, 2H), 4.11 (m, 1H), 3.76 (s, 6H), 3.5-4.0 (m, 4H), 3.17 (m, 2H), 2.4-2.7 (m, 4H), 2.38 (m, 1H), 1.14 (m, 12H).

Quantitative DNase I footprint titrations. 3'-end labeling and footprint titration experiments were performed as described in chapters 2 and 3. Quantitation and data analysis were also performed as previously described in chapter 2.

References

(1) Radhakrishnan, I.; Patel, D.J. *J. Am. Chem. Soc.* **1993**, *115*, 1615-1617.

(2) Durland, R.H.; Rao, T.S.; Bodepudi, V.; Seth, D.M.; Jayaraman, K.; Revankar, G.R. *Nucleic Acids Res.* **1995**, *23*, 647-653.

(3) Hoffer, M. *Chem. Ber.* **1960**, *93*, 2777-2781.

(4) Parquette, J.R.; Dervan, P.B. Unpublished data.

(5) Hayakawa, Y.; Wakabayashi, S.; Kato, H.; Noyori, R. *J. Am. Chem. Soc.* **1990**, *112*, 1691-1696.

(6) Radhakrishnan, I.; Patel, D.J. *Structure* **1994**, *2*, 17-32.

CHAPTER SIX

Pyrimidine Motif Triple Helix Formation by Oligonucleotides Containing Nonnatural Nucleosides with an Extended Aromatic Structure: Intercalation from the Major Groove as a Potentially General Method for Recognizing CG and TA Base Pairs

Introduction

Efforts to design novel nucleosides for recognition of CG and TA base pairs by triple helix formation, either in the pyrimidine or in the purine motif, have thus far met with limited success. The few successful designs have proven not to be generalizable. For instance, the G•TA interaction in the pyrimidine motif[1] has proven to be highly dependent on flanking sequence.[2] This has been shown by NMR studies to be due to tilting of the third strand guanine resulting in hydrogen bonding to the flanking step.[3,4] Another example is that of D_3 (Figure 1) which was found to specifically recognize CG *and* TA base pairs in the pyrimidine motif.[5] However, multiple D_3 substitutions are not tolerated.[6] NMR studies have shown that the mechanism of specific recognition by D_3 is through intercalation of its benzoyl group favorably over the purines of CG and TA base pairs relative to the pyrimidines of AT and GC base pairs.[7] Since intercalation results in a local distortion of the duplex by displacement of flanking base pairs, multiple intercalation events are detrimental to the overall energetics of the structure.

One of the few published successes in this field of novel nucleoside design is Miller's diaminopyridine-appended cytidine derivative, referred to here as M (Figure 1).[8] Using a melting temperature assay, the authors reported some specificity for CG and AT base pairs, although the melting transitions observed were not as cooperative as expected and may have reflected multiple binding modes. In an effort to quantify the results observed in this paper and compare them in an identical system to other structures designed in this group, we incorporated M into an oligonucleotide for quantitative DNase footprint titration studies. It was found in characterizing the oligonucleotide that the benzoyl protecting group used by Miller was not removed under standard deprotection

conditions, and could not be removed under harsher conditions either, resulting in the oligonucleotide containing not M, but ^{Bz}M (Figure 1).

Energetics of triple helix formation of ^{Bz}M were very similar to those of D_3, with specificity for CG and TA base pairs, indicating that it is likely that these structurally similar nucleosides both intercalate into the major groove in the context of triple helical DNA. In addition, an oligonucleotide containing M without a protecting group was accessed by using an allyl carbamate rather than a benzamide phosphoramidite in the synthesis. This base, the one originally designed by Miller, did not bind specifically to any of the four base pairs.

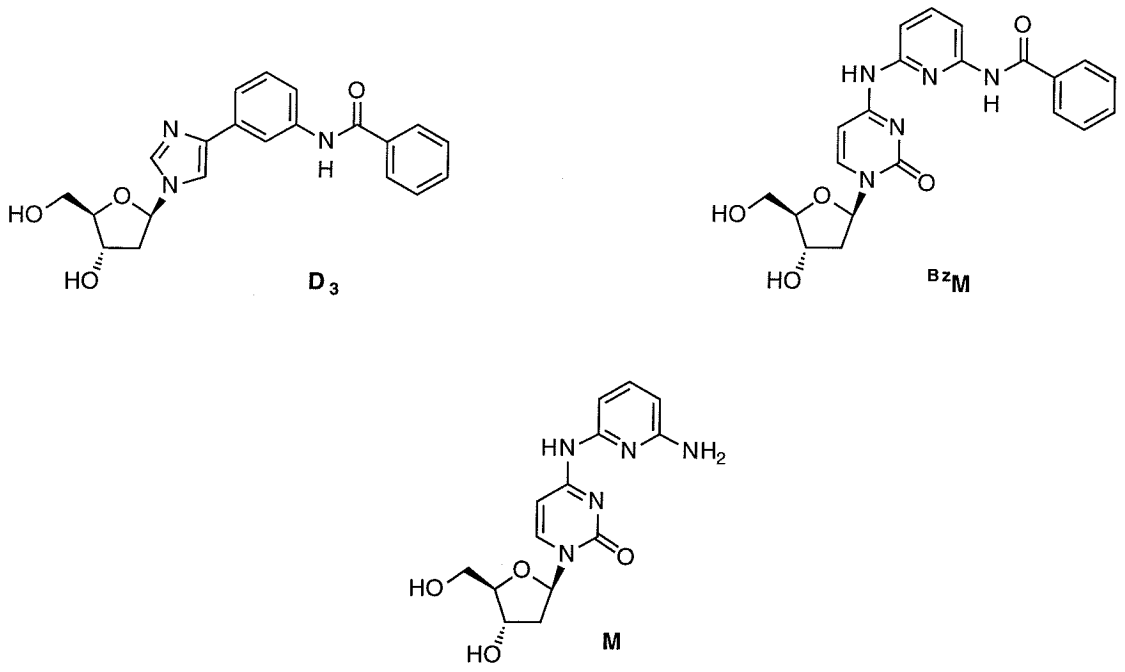


Figure 1. The structures of D_3, M, and ^{Bz}M.

Results and Discussion

The phosphoramidite for ^{Bz}M was prepared in good yield according to the published procedure of Huang and Miller.[8] The phosphoramidite for M, containing the Alloc-protected amine rather than benzoyl, was prepared in similar fashion (Figure 2).

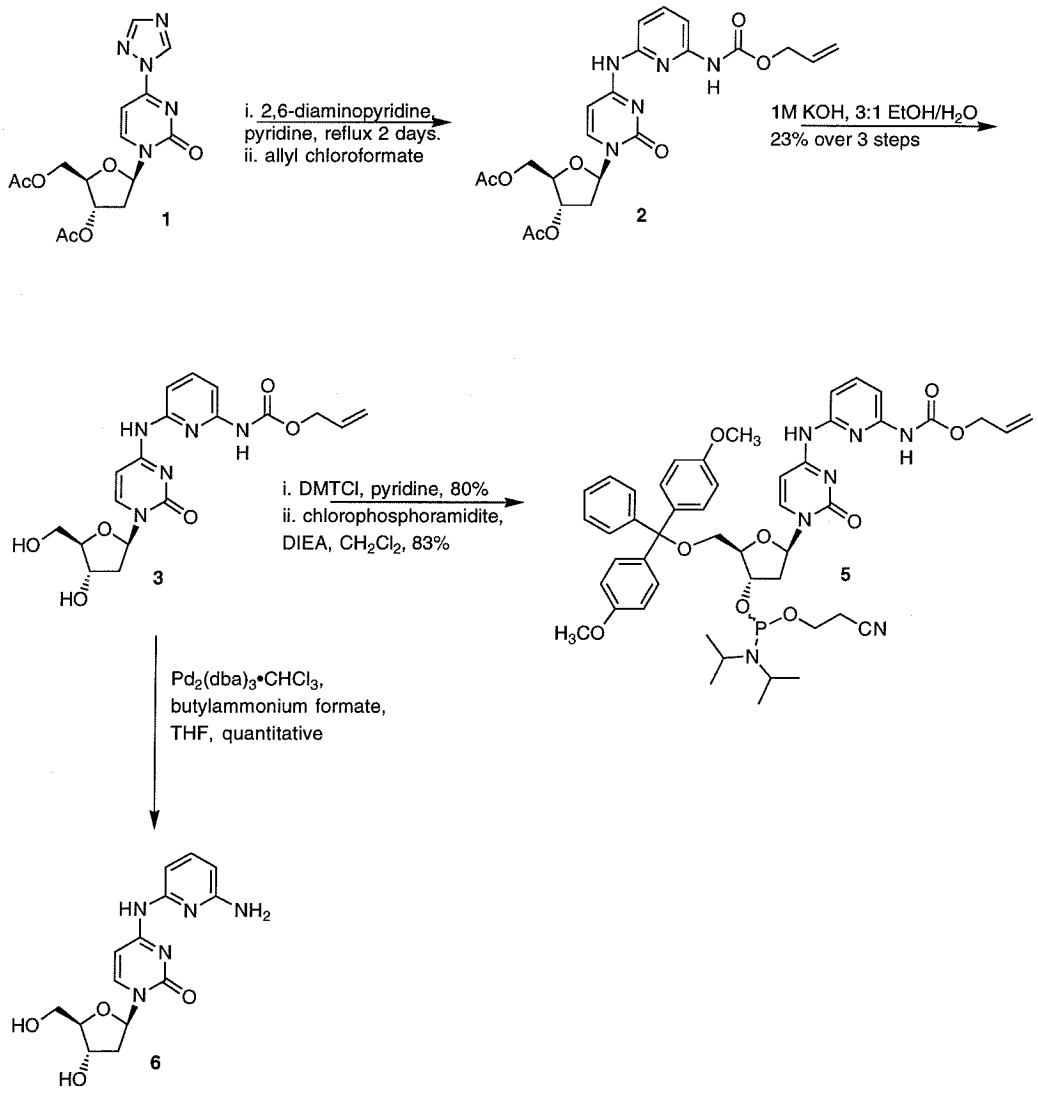


Figure 2. Synthesis of M phosphoramidite.

Reaction of triazolyluridine **1**[9] with 2,6-diaminopyridine generated a product which was inseparable from the excess remaining diaminopyridine. This mixture was exhaustively acylated with allyl chloroformate to produce **2**, which was easily separable from pyridine bis-*O*-allylcarbamate. Hydroxyl deprotection, 5'-dimethoxytritylation, and phosphitylation provided **5** for oligonucleotide synthesis. Nucleoside **3** was readily deprotected under Noyori's conditions[10] to provide authentic M nucleoside (**6**).

M and ^{Bz}M nucleosides were incorporated into the standard pyrimidine oligonucleotide sequence, 5'-TT^{m}CT^{m}CT**Z**T^{m}CT^{m}C^{m}CTTT-3' at the single position Z, for footprinting on the standard 314 bp restriction fragment from pYSPEC2 containing all four target sites 5'-AAGAGA**X**AGAGGAAA-3', where X = A,G,C, and T (Figure 3).

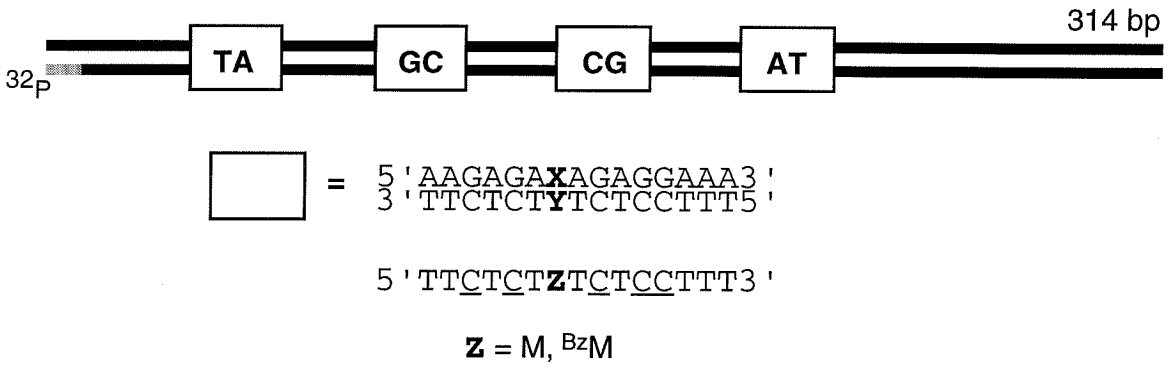


Figure 3. Sequences of the four target sites on 3'-labeled 314 bp *Afl* II/*Fsp* I restriction fragment of pYSPEC2. Below are the sequences of third strand oligonucleotides containing M and ^{Bz}M. C denotes 5-methylcytosine.

Results obtained on the energetics of triple helix formation by oligonucleotides containing M and ^{Bz}M are shown in Table 1 and Figures 4 and 5.

Table 1. Association constants (K_T) for the formation of triple helical complexes containing the Z•XY triplets indicated at 25°C, 100 mM NaCl, 250 μM spermine, 10 mM bistris HCl, pH 7.0.[a,b]

XY	Z=^{Bz}M	M
AT	8.8 (±3.1) x 10^5	<1 x 10^5
GC	5.7 (±1.6) x 10^5	7.0 (±1.5) x 10^5
CG	8.5 (±2.2) x 10^6	~1 x 10^5
TA	5.8 (±1.4) x 10^6	< 1 x 10^5

[a] K_T values are reported as the mean (± the standard error of the mean) of three measurements. The K_T values are reported in units of M^{-1}. [b] The identity of the base Z is indicated across the top of the columns; the identity of the Watson Crick base pair XY is indicated on the left side of the rows.

^{Bz}M binds CG and TA base pairs with approximately 10-fold specificity over AT and GC base pairs (Table 1). CG and TA are bound with similar affinity. These results are qualitatively similar to those observed for D_3.[5,6] This, along with the similar extended aromatic structures of the two nucleosides, are suggestive of an intercalative mode of binding for ^{Bz}M, although structural characterization by NMR would be necessary to state this definitively.

Further evidence for the intercalation of ^{Bz}M is provided by the data on M, in which the benzoyl group has been removed. Specific binding to CG and TA is lost, and all four base pairs are poorly recognized, although GC is recognized best (Table 1).

A third nucleoside with a similar extended aromatic structure and identical binding properties (~ 10-fold specificity for CG and TA base pairs) has recently been developed in

Figure 4. Autoradiogram of a 8% denaturing polyacrylamide gel used to separate fragments generated by DNase I digestion in a quantitative footprint titration experiment with an oligonucleotide containing ^{Bz}M. The four target sites on the 314 bp 3'-end labeled restriction fragment are labeled as XY= AT, CG, GC, and TA respectively, going from the top to the bottom of the gel. (Lane 1) Products of an adenine-specific reaction. (Lane 2) Products of a guanine-specific reaction. (Lane 3) Intact 3' labeled DNA after incubation in the absence of third strand oligonucleotide. (Lanes 4-21) DNase I digestion products obtained in the presence of varying concentrations of oligonucleotide: 8 μM (lane 4); 4 μM (lane 5); 2 μM (lane 6); 1 μM (lane 7); 800 nM (lane 8); 400 nM (lane 9); 200 nM (lane 10); 100 nM (lane 11); 80 nM (lane 12); 40 nM (lane 13);20 nM (lane 14); 10 nM (lane 15); 8 nM (lane 16); 4 nM (lane 17); 2 nM (lane 18); 1 nM (lane 19); 800 pM (lane 20); no oligonucleotide (lane 21).

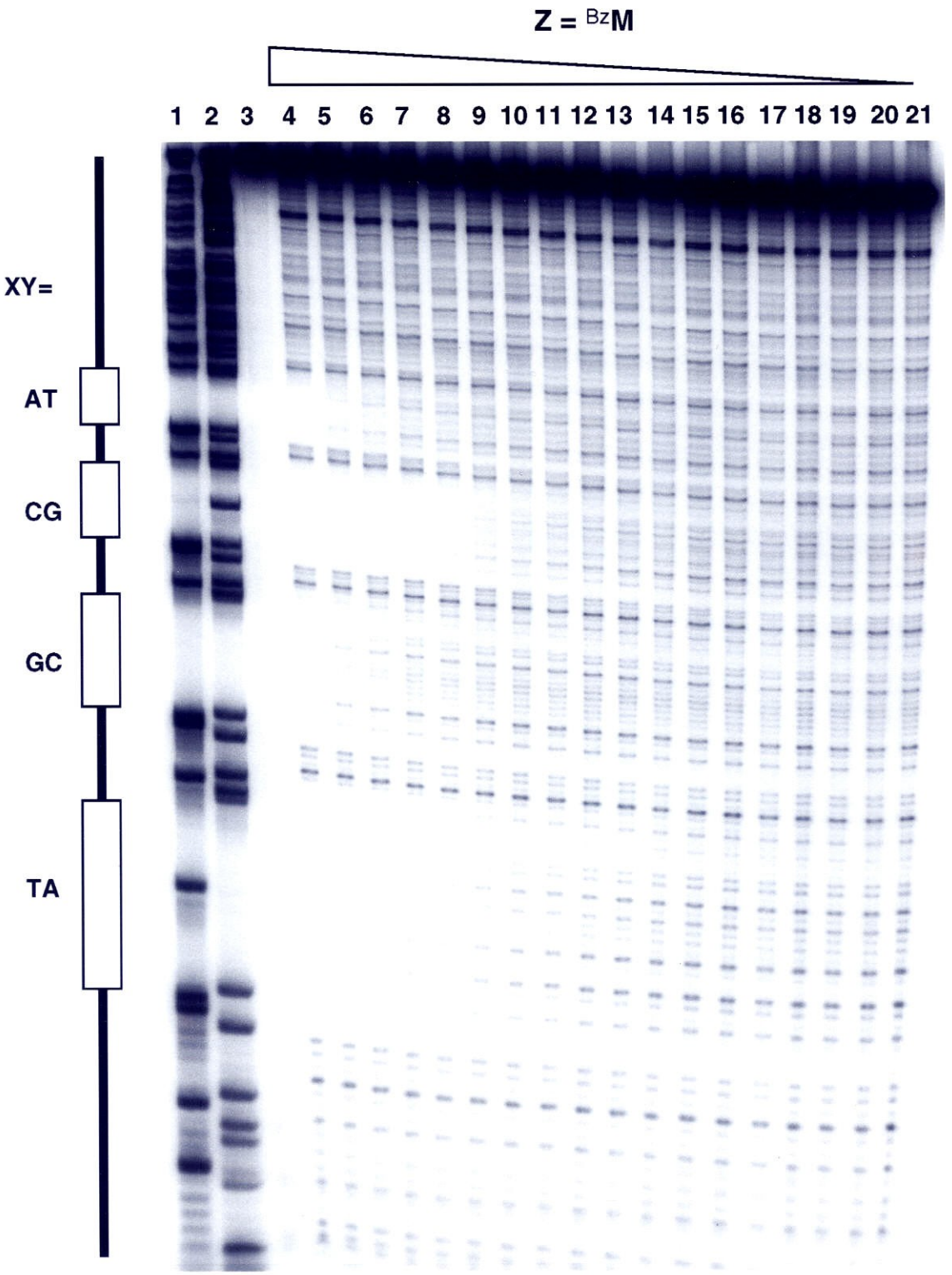
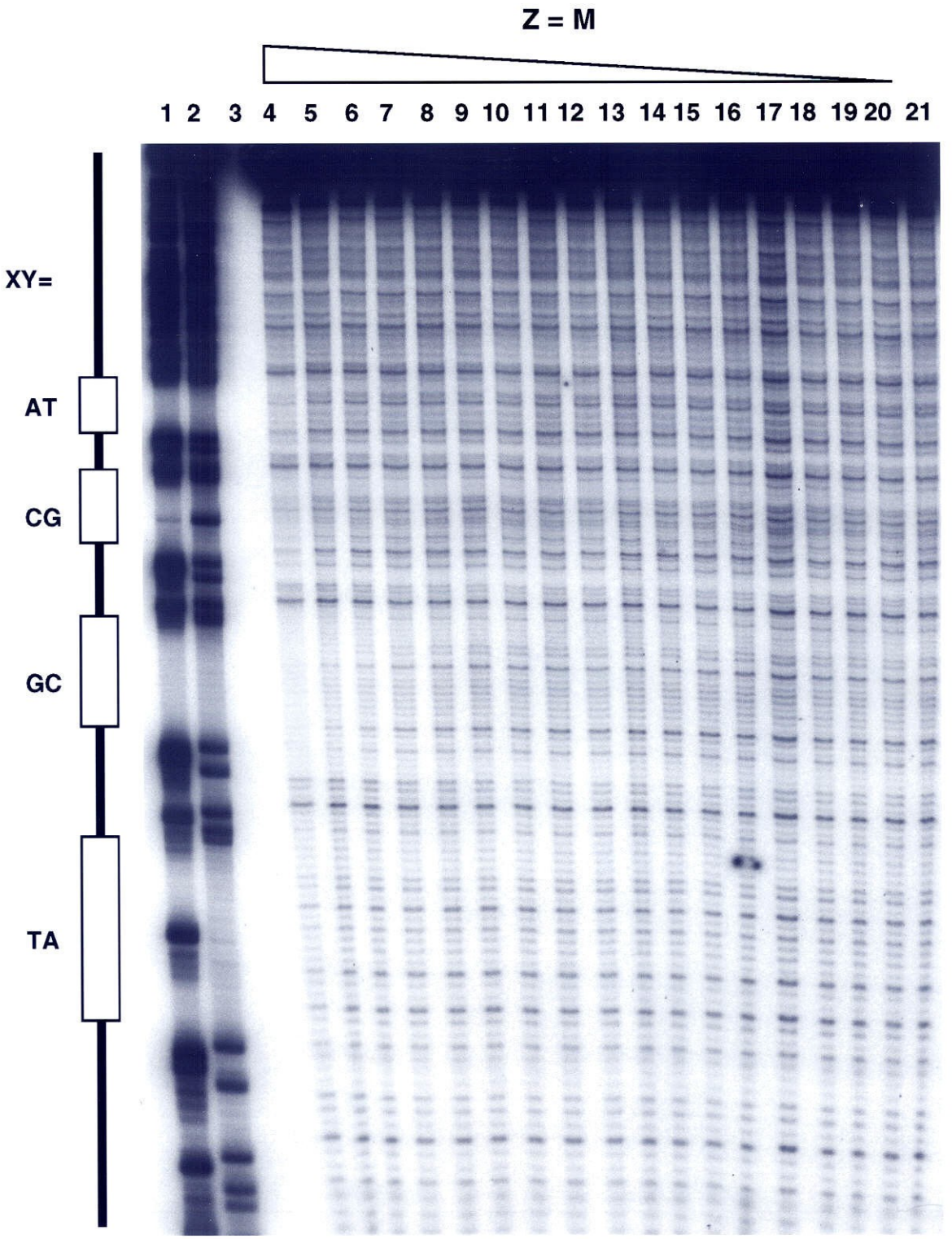

Z = BzM
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
XY=
AT
CG
GC
TA

Figure 5. Autoradiogram of a 8% denaturing polyacrylamide gel used to separate fragments generated by DNase I digestion in a quantitative footprint titration experiment with an oligonucleotide containing M. The four target sites on the 314 bp 3'-end labeled restriction fragment are labeled as XY= AT, CG, GC, and TA respectively, going from the top to the bottom of the gel. (Lane 1) Products of an adenine-specific reaction. (Lane 2) Products of a guanine-specific reaction. (Lane 3) Intact 3' labeled DNA after incubation in the absence of third strand oligonucleotide. (Lanes 4-21) DNase I digestion products obtained in the presence of varying concentrations of oligonucleotide: 8 μM (lane 4); 4 μM (lane 5); 2 μM (lane 6); 1 μM (lane 7); 800 nM (lane 8); 400 nM (lane 9); 200 nM (lane 10); 100 nM (lane 11); 80 nM (lane 12); 40 nM (lane 13);20 nM (lane 14); 10 nM (lane 15); 8 nM (lane 16); 4 nM (lane 17); 2 nM (lane 18); 1 nM (lane 19); 800 pM (lane 20); no oligonucleotide (lane 21).


Z = M
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
XY=
AT
CG
GC
TA

these laboratories by Dr. Thomas Lehmann.[11] Its structure is shown in Figure 6. The emergence of multiple structures which appear to intercalate into the major groove when delivered in the context of a third strand oligonucleotide suggests that this mode of binding is very accessible and may provide a general approach to specific recognition of CG and TA base pairs. Although all three nucleosides mentioned in this chapter (which all contain a benzoyl group) do not distinguish between CG and TA, it may be possible by modifying the composition of the intercalating moiety to effect preference for one base pair over the other. However, this approach is still limited by the inability to introduce multiple substitutions due to the large structural distortion induced by intercalation.

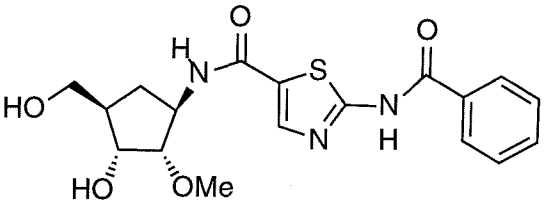


Figure 6. Structure of an extended aromatic nucleoside, L2, which likely binds through intercalation when incorporated into a third strand oligonucleotide.

Conclusion

A previously published novel nucleoside M, when reevaluated for its energetics of triple helix formation, was found to have retained its benzoyl protecting group through oligonucleotide synthesis and deprotection. The resultant compound, ^{Bz}M, was found to bind specifically to CG and TA base pairs within the context of a pyrimidine motif triple helix. Its properties are similar to those of the structurally related D_3, which has previously been shown to intercalate into the major groove when incorporated into a third strand

oligonucleotide. This suggests that ^{Bz}M also intercalates, and that intercalation may be a general phenomenon common to extended aromatic nucleosides. Intercalative nucleosides may find some utility in the design of molecules for specific recognition of CG and TA base pairs by triple helix formation.

Experimental section

General methods in the synthesis of phosphoramidites were performed as described in the previous chapters. Pyrimidine oligonucleotides containing single substitutions of M or ^{Bz}M were synthesized by standard methodology as described in previous chapters. Removal of the allyl carbamate protecting group from CPG-bound oligonucleotide with $Pd_2(dba)_3$ followed Noyori's protocols for deprotection of allyl-protected oligonucleotides.[10] Oligonucleotides were purified by denaturing gel electrophoresis (20% polyacrylamide) and elution of the desired band into 25 mM Tris-HCl, 250 mM NaCl, 1 mM EDTA, pH 8.0 followed by dialysis against milliQ water. Identity and purity of oligonucleotides was assessed by MALDI-TOF mass spectrometry and HPLC analysis of enzymatic digests.

Mass spectrometry of oligonucleotides: ^{Bz}M oligonucleotide mass found 4674 (4676 calcd. for oligonucleotide if benzoyl group remains). No peak at 4572 is observed, which would correspond to M being present rather than ^{Bz}M. M oligonucleotide mass found 4574.1 (4572.2 calcd. for oligonucleotide if fully deprotected M is present). No peak at 4656 is observed, which would correspond to retention of the Alloc protecting group.

HPLC analysis of enzymatically digested nonnatural base-containing oligonucleotides. 10 nmol samples of oligonucleotides were digested with 3 units calf alkaline phosphatase and 0.01 units snake venom phosphodiesterase in 50 μL 50 mM Tris, 10 mM $MgCl_2$, pH 8.0 for 3 hours at 37°C. Samples were injected onto a Rainin Microsorb MV C18 reverse phase column and eluted with a linear gradient of 0-15% acetonitrile in 20 mM ammonium acetate, pH 5.5 over 20 minutes followed by a ramp from 15% to 60% acetonitrile over the next 20 minutes at a flow rate of 1 mL/min.

Samples of authentic nucleoside standards were run as well for comparison. Elution times: ^{Bz}M oligonucleotide, ^{m}C (10.4 min), T (12.3 min), ^{Bz}M (30.8 min, λ_{max}=316 nm); M oligonucleotide, ^{m}C (10.4 min), T (12.3 min), M (19.2 min, λ_{max}=324 nm). No peak for M was observed in the chromatogram of the ^{Bz}M oligonucleotide. Likewise, ^{Alloc}M (30.2 min, λ_{max}=316 nm) was not observed in the chromatogram of the M oligonucleotide.

Synthesis:

The phosphoramidite for ^{Bz}M was prepared according to the procedure outlined by Miller.[8] The phosphoramidite for M, containing the Alloc protecting group rather than benzoyl was prepared in similar fashion and is described in detail below.

3',5'-*O*-Diacetyl-N-alloc-M (2): Triazolyl nucleoside **1**[9] (1.4 g, 3.85 mmol) was refluxed in 15 mL pyridine with 2,6-diaminopyridine (1.26 g, 11.6 mmol) for 2 days, at which time the dark brown solution was cooled over an ice bath and allyl chloroformate (6 equivalents, 2.45 mL, 23.1 mmol) was added. The mixture was stirred at room temperature for 2 hours, then quenched with 1 mL MeOH, diluted with 50 mL EtOAc, and extracted with saturated $NaHCO_3$. The organic layer was dried, concentrated, and purified by silica chromatography (EtOAc to separate large excess of bis(alloc)diaminopyridine, then 5% MeOH/EtOAc) to furnish crude **2** as light brown solid which was carried on to the next step. TLC (5%MeOH/EtOAc) R_f 0.55; ^{1}H NMR ($CDCl_3$) δ 7.66 (d, 1H, J=7.8Hz), 7.63 (b, 2H), 7.21 (t, 1H, J=7.8 Hz), 7.10 (d, 1H, J=7.8Hz), 6.90 (m, 2H), 6.31 (m, 1H), 5.95 (m, 1H), 5.38 (m, 2H), 4.94 (m, 1H), 4.70 (m, 2H), 4.34 (m, 3H), 2.70 (m, 1H), 2.13 (m, 1H), 2.09 (s, 6H).

N-Alloc-M (3): Crude **2** was dissolved in 5 mL absolute ethanol, then 5 mL 2M KOH in 1:1 ethanol/water was added. The mixture was stirred at 0°C for 15 minutes. The KOH was neutralized with 600 mg NH_4Cl, and the mixture concentrated, using EtOH to

azeotropically remove water. Purification by silica chromatography (5% MeOH/EtOAc, then 10% MeOH, then 15% MeOH) provided a pink solid (363 mg, 23% over 2 steps). TLC (10% MeOH/EtOAc) R_f 0.15; UV λ_{max}= 315 nm (34,000); ^{1}H NMR (DMSO-d_6) δ 10.1 (b, 2H), 8.05 (d, 1H, J=7.5Hz), 7.69 (t, 1H, J=8.4Hz), 7.40 (d, 1H, J=8.4Hz), 7.0-7.3 (vb, 2H), 6.16 (t, 1H, J=6.6 Hz), 5.95 (m, 1H), 5.2-5.4 (m, 2H), 5.26 (m, 1H), 5.06 (m, 1H), 4.61 (m, 2H), 4.21 (m, 1H), 3.80 (m, 1H), 3.56 (m, 2H), 2.18 (m, 1H), 1.98 (m, 1H); FAB MS M+H 404.1558 (404.1570 calcd. for $C_{18}H_{22}N_5O_6$).

5'-*O*-(4,4'-dimethoxytrityl)-N-alloc-M (4): Protected nucleoside **3** (150 mg, 0.37 mmol) and DMT chloride (190 mg, 0.56 mmol) were dried overnight under vacuum. 7 mL pyridine was added, and the reaction mixture was stirred at room temperature for 3.5 hours, after which 0.5 mL methanol was added to quench, and the solution was diluted with EtOAc and extracted before concentrating under reduced pressure. Purification by silica chromatography (5% MeOH/EtOAc) yielded a light yellow foam (210 mg, 80 %). TLC (8% MeOH/EtOAc) R_f 0.55; ^{1}H NMR (acetone-d_6) δ 9.14 (b, 1H), 8.71 (b, 1H), 7.99 (d, 1H, J=7.5 Hz), 7.74 (t, 1H, J=8.1 Hz), 7.60 (d, 1H, J=8.1Hz), 7.48 (m, 2H), 7.33 (m, 7H), 7.22 (m, 1H), 6.88 (m, 4H), 6.45 (b, 1H), 6.28 (t, 1H, J=6.0Hz), 5.96 (m, 1H), 5.2-5.4 (m, 2H), 4.63 (m, 2H), 4.52 (m, 1H), 4.08 (m, 1H), 3.78 (s, 6H), 3.39 (m, 2H), 2.47 (m, 1H), 2.24 (m, 1H); FAB MS M+H 706.2874 (708.2877 calcd. for $C_{39}H_{40}N_5O_8$).

N-Alloc-M phosphoramidite (5): 5'-DMT nucleoside **4** (188 mg, 0.266 mmol) was dissolved in 4 mL dichloromethane. Diisopropylethylamine (140 μL, 0.80 mmol) was added, followed by dropwise addition of 2-cyanoethyl-N,N-diisopropylchlorophosphoramidite (75 μL, 0.40 mmol). The reaction was stirred at room temperature for 1 hour, after which it was quenched with 0.2 mL MeOH, diluted with 25 mL EtOAc, extracted with saturated $NaHCO_3$, then brine, dried, and concentrated.

Purification by silica chromatography (EtOAc) furnished a white powder (201 mg, 83%). TLC (5% MeOH/EtOAc) R_f 0.8; ^{1}H NMR (acetone-d_6) δ 9.17 (b, 1H), 8.70 (b, 1H), 8.00 (m, 1H,), 7.73 (t, 1H, J=8.1 Hz), 7.61 (d, 1H, J=8.1Hz), 7.48 (m, 2H), 7.33 (m, 7H), 7.22 (m, 1H), 6.89 (m, 4H), 6.47 (b, 1H), 6.30 (m, 1H), 5.96 (m, 1H), 5.2-5.4 (m, 2H), 4.70 (m, 1H), 4.63 (m, 2H), 4.18 (m, 1H), 3.76 (s, 6H), 3.5-4.0 (m, 4H), 3.43 (m, 2H), 2.4-2.7 (m, 3H), 2.35 (m, 1H), 1.20 (m, 12H); FAB MS M+H 906.3967 (906.3955 calcd. for $C_{48}H_{57}N_7O_9P$).

M (6): Alloc protected nucleoside **3** (25 mg, 0.062 mmol) was dissolved in 4 mL dichloromethane containing butylamine (226 μL) and formic acid (86 μL). To this solution was added 15 mg triphenylphosphine and 5 mg $Pd_2(dba)_3$•$CHCl_3$. The mixture was stirred at room temperature for 2 hours, after which 10 mg DDTC was added to sequester the palladium and the product was extracted into water. By TLC analysis, all product extracted into the water. The aqueous solution was lyophilized, and the product further purified by silica chromatography (20% MeOH/EtOAc) to yield a light yellow solid (20 mg, 100%). UV λ_{max}= 322 nm (24,000); ^{1}H NMR (DMSO-d_6) δ 9.8 (b, 1H), 8.14 (s, 1H), 7.97 (m, 1H), 7.32 (t, 1H, J=8.4Hz), 7.1 (vb, 1H), 6.6 (vb, 1H), 6.12 (t, 1H, J=6.6 Hz), 5.81 (b, 2H), 5.25 (b, 1H), 5.04 (b, 1H), 4.19 (m, 1H), 3.77 (m, 1H), 3.54 (m, 2H), 2.14 (m, 1H), 1.98 (m, 1H); FAB MS M+H 320.1353 (320.1359 calcd. for $C_{14}H_{18}N_5O_4$).

Quantitative DNase I footprint titrations: 3'-end labeling of the fragment from pYSPEC2 and footprint titrations were performed as previously described by E.S. Priestley.[12]

References

(1) Griffin, L.C.; Dervan, P.B. *Science* **1989**, *245*, 967-971.

(2) Kiessling, L.L.; Griffin, L.C.; Dervan, P.B. *Biochemistry* **1992**, *31*, 2829-2834.

(3) Radhakrishnan, I.; Patel, D.J. *Structure* **1994**, *2*, 17-32.

(4) Wang, E.; Malek, S.; Feigon, J. *Biochemistry* **1992**, *31*, 4838-4846.

(5) Griffin, L.C.; Kiessling, L.L.; Beal, P.A.; Gillespie, P.; Dervan, P.B. *J. Am. Chem. Soc.* **1992**, *114*, 7976-7982.

(6) Liberles, D.A.; Dervan, P.B. Unpublished results.

(7) Koshlap, K.M.; Gillespie, P.; Dervan, P.B.; Feigon, J. *J. Am. Chem. Soc.* **1993**, *115*, 7908-7909.

(8) Huang, C.Y.; Miller, P.S. *J. Am. Chem. Soc.* **1993**, *115*, 10456-10457.

(9) Divakar, K.J.; Reese, C.B. *J. Chem. Soc.* **1982**, 1171-1176.

(10) Hayakawa, Y.; Wakabayashi, S.; Kato, H.; Noyori, R. *J. Am. Chem. Soc.* **1990**, *112*, 1691-1696.

(11) Lehmann, T.; Dervan, P.B. Unpublished results.

(12) Priestley, E.S. Ph.D. Thesis, California Institute of Technology, 1996.

CHAPTER SEVEN

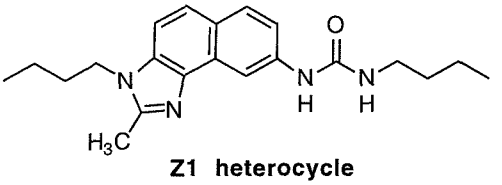
Synthesis and Energetics of Pyrimidine Motif Triple Helix Formation of a Novel Nucleoside Based On a Model Heterocycle which Binds CG Base Pairs in Chloroform

Introduction

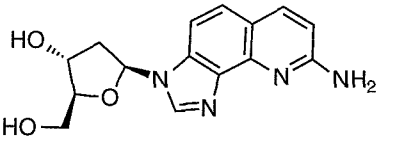
As has been described here in the previous chapters and elsewhere,[1-6] there has been great interest in designing novel nucleosides for use in recognizing CG and TA base pairs by triple helix formation. Generally the experimental approach which has been used by us and others has entailed incorporation of the designed molecule into an oligonucleotide through its phosphoramidite, and measurement of energetics of binding and specificity of the oligonucleotide by a method such as DNase I footprint titration,[1,2] electrophoretic mobility shift assay,[3,4] or melting temperature measurements.[5,6] In 1995 Zimmerman's group reported on a model study in which the naphthimidazole Z1 (Figure 1) bound to an isolated CG base pair in chloroform in the manner shown in Figure 2, based on proton NMR chemical shift changes on mixing and nOe data.[7] Considering the time-consuming nature of quantitative DNase I footprint titration of novel nucleoside-containing synthetic oligonucleotides, the possibility of evaluating designs by simple ^{1}H NMR experiments on more accessible model compounds is very appealing. However, there are serious concerns as to whether the behavior of a designed molecule in an isolated triplet in chloroform accurately reflect what its behavior would be in the context of triple helical DNA in water. Therefore, we set out to incorporate the Z1 heterocycle into a pyrimidine oligonucleotide and determine its energetics of triple helix formation by the standard quantitative DNase I footprint titration assay.

In addition we were interested in Z1 because its structure was very similar to a novel nucleoside, Q1 (Figure 1), which was designed and incorporated into a pyrimidine oligonucleotide by Dr. Carol Wada, then a postdoctoral associate in the group. It was found that Q1 bound all four base pairs as a very weak mismatch, with slight specificity for GC, rather than CG.[8] Clearly the designed hydrogen bonds to CG were not present.

We were interested in learning whether the small structural changes present in Z1, namely the removal of the quinoline nitrogen of Q1, the addition of a methyl group at the imidazole 2-position, and the elaboration of the amine functionality to a urea, could result in a dramatic enhancement in the energetics of binding to CG.



Z1 heterocycle



Q1 nucleoside

Figure 1. Structures of the Z1 heterocycle and Q1 nucleoside.

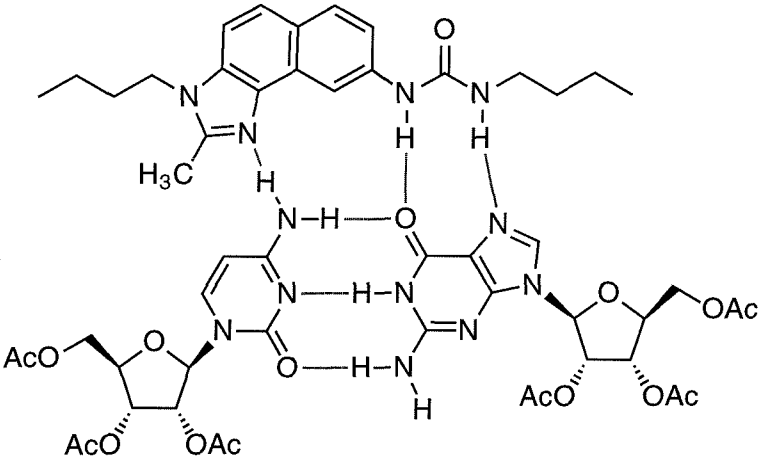


Figure 2. Proposed Z1•CG triplet based on NMR data in chloroform.

Results and Discussion

The synthesis of the Z1 heterocycle (lacking the N-butylurea) closely followed the reported route with minor modifications (Figure 3, top).[7] 2,7-Dihydroxynaphthalene (**1**) was converted to the corresponding diamine by a Bucherer reaction. This diamine was found to be unstable in air, and was immediately acetylated to provide **2**. An extremely facile nitration provided **3**. Reduction of the nitro group and cyclization under the reaction conditions to provide **5**, as reported by Zimmerman, was problematic. Thus **3** was reduced with Raney nickel and hydrazine to provide the stable amine **4**, which was readily cyclized to **5** in refluxing acetic acid. Acetamide hydrolysis provided the desired Z1 heterocycle **6**, which was glycosylated in low yield to provide **8** (Figure 3, bottom). Modeling suggested that the N-butyl group on the model compound's urea (designed for increased solubility in organic solvent) would be detrimental to binding in the context of triple helical DNA, and was replaced with an unsubstituted urea. Urea formation by standard methods (sodium cyanate under acidic conditions) led to cleavage of the glycosidic bond, but trimethylsilyl isocyanate cleanly provided the unsubstituted urea **9**. Standard hydroxyl deprotection, 5'-dimethoxytritylation, and phosphitylation provided phosphoramidite **11** for incorporation into oligonucleotides. Notably, in the 5'-tritylation step, the competing tritylation of the imidazole functionality (which leads to deglycosylation) was reduced by addition of DBU, presumably by deprotonating the 5' hydroxyl.

Z1 nucleoside was incorporated into the standard pyrimidine oligonucleotide sequence, 5'-TT^{m}CT^{m}CT**Z**T^{m}CT^{m}C^{m}CTTT-3' at the single position Z, for footprinting on the standard 314 bp restriction fragment from pYSPEC2 containing all four target sites 5'-

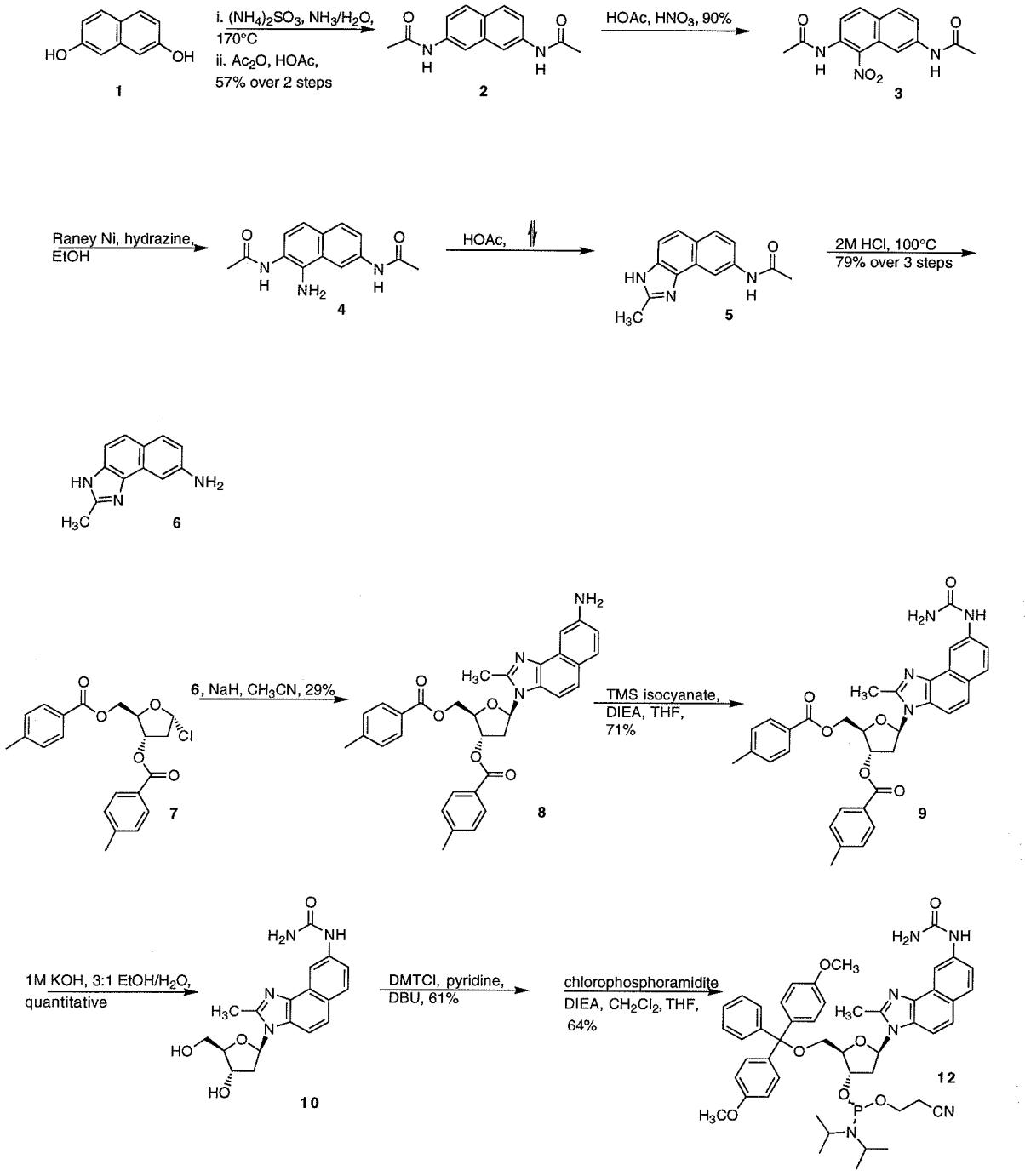


Figure 3. Synthesis of Z1 phosphoramidite **11**.

AAGAGA**X**AGAGGAAA-3', where X = A,G,C, and T. The oligonucleotide was found to be quite sensitive to acidic conditions, but it was possible to purify and characterize it at pH 7.0 and above.

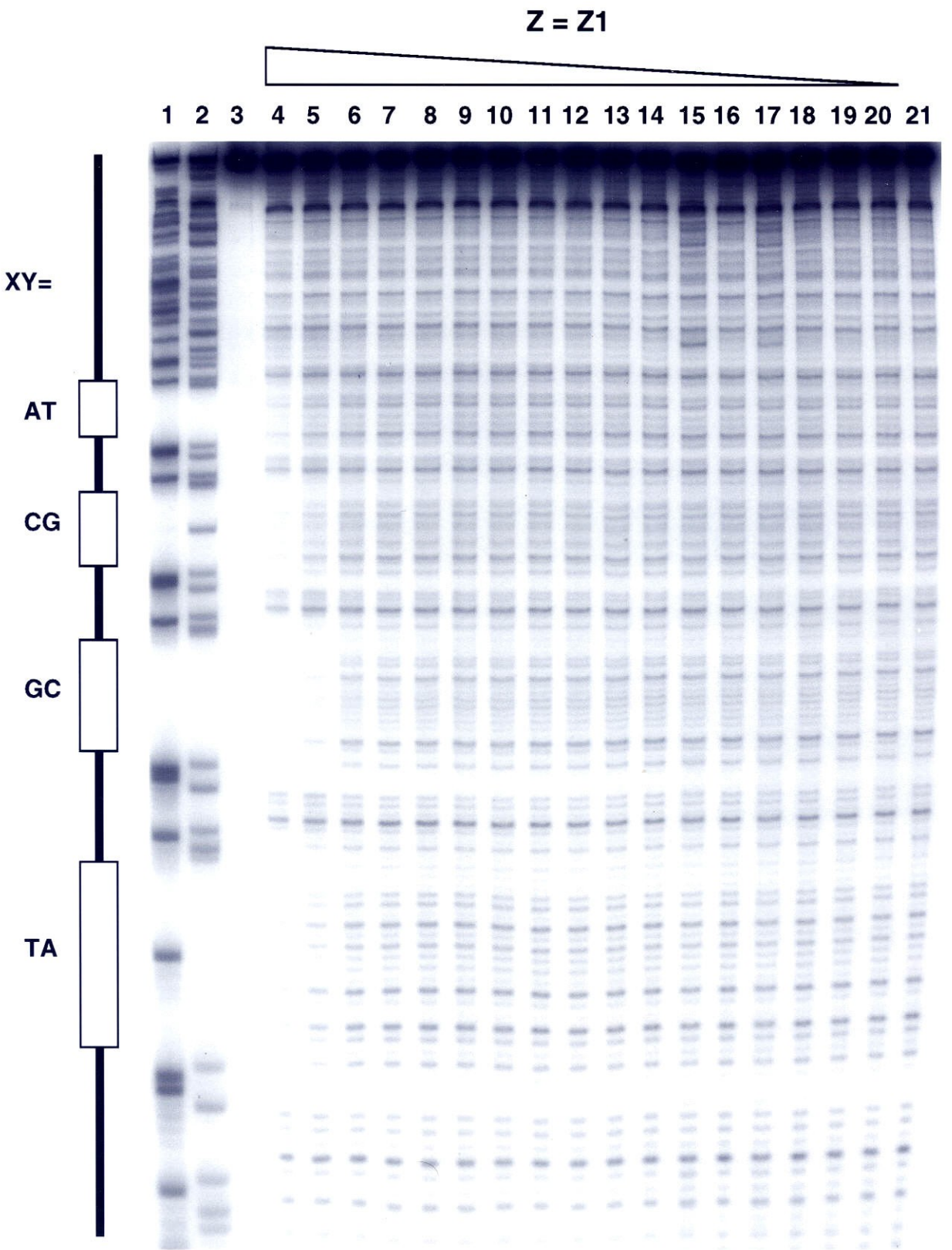
Figure 4 and Table 1 show that Z1 does not bind CG or any other base pair specifically or with high affinity. The association constants determined for the pyrimidine oligonucleotide containing Z1 are similar to those previously determined for Q1. It is likely that no specific hydrogen bonds are being formed between Z1 and CG in the context of triple helical DNA in water, unlike the results observed in chloroform. This suggests that in general model systems in organic solvent will not accurately reflect the energetics of binding in the aqueous DNA system. Two major differences between the model and triple helix systems which likely are important with respect to this result are 1) the relative strength of hydrogen bonds in chloroform and water and 2) the importance of base stacking, which is not present in the model system but is present in triple helical DNA.

Table 1. Association constants (K_T) for the formation of triple helical complexes containing the Z•XY triplets indicated at 25°C, 100 mM NaCl, 250 μM spermine, 10 mM bistris HCl, pH 7.0.[a,b]

XY	**Z=Z1**	Q1
AT	$< 1 \times 10^5$	$<1 \times 10^5$
GC	$\sim 1.0 \times 10^5$	$\sim 1.0 \times 10^5$
CG	$< 1 \times 10^5$	$<1 \times 10^5$
TA	$< 1 \times 10^5$	$< 1 \times 10^5$

[a] K_T values are reported as the mean (± the standard error of the mean) of three measurements. The K_T values are reported in units of M^{-1}. [b] The identity of the base Z is indicated across the top of the columns; the identity of the Watson Crick base pair XY is indicated on the left side of the rows.

Figure 4. Autoradiogram of a 8% denaturing polyacrylamide gel used to separate fragments generated by DNase I digestion in a quantitative footprint titration experiment with an oligonucleotide containing Z1. The four target sites on the 314 bp 3'-end labeled restriction fragment are labeled as XY= AT, CG, GC, and TA respectively, going from the top to the bottom of the gel. (Lane 1) Products of an adenine-specific reaction. (Lane 2) Products of a guanine-specific reaction. (Lane 3) Intact 3' labeled DNA after incubation in the absence of third strand oligonucleotide. (Lanes 4-21) DNase I digestion products obtained in the presence of varying concentrations of oligonucleotide: 8 μM (lane 4); 4 μM (lane 5); 2 μM (lane 6); 1 μM (lane 7); 800 nM (lane 8); 400 nM (lane 9); 200 nM (lane 10); 100 nM (lane 11); 80 nM (lane 12); 40 nM (lane 13);20 nM (lane 14); 10 nM (lane 15); 8 nM (lane 16); 4 nM (lane 17); 2 nM (lane 18); 1 nM (lane 19); 800 pM (lane 20); no oligonucleotide (lane 21).


Z = Z1
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
XY=
AT
CG
GC
TA

Conclusion

A model system which binds isolated CG base pairs in chloroform was shown to not bind specifically to CG, or other base pairs within the context of triple helical DNA in water. The binding was consistent with that of a structurally similar novel nucleoside previously prepared in this group. This result suggests that model studies in organic solvent will not be useful in predicting the affinity and specificity of novel nucleosides designed for expanding the DNA sequence repertoire available for recognition by triple helix formation.

Experimental section

General methods in the synthesis of phosphoramidites were performed as described in the previous chapters. The pyrimidine oligonucleotide containing a single substitution of Z1 was synthesized by standard methodology as described in previous chapters. The oligonucleotide was purified by denaturing gel electrophoresis (20% polyacrylamide) and elution of the desired band into 25 mM Tris-HCl, 250 mM NaCl, 1 mM EDTA, pH 8.0 followed by dialysis against milliQ water. Identity and purity of oligonucleotides was assessed by MALDI-TOF mass spectrometry and HPLC analysis of enzymatic digests.

The observed mass of 4655.6 corresponds to the mass of the oligonucleotide + $2Na^+$. A significant peak corresponding to loss of Z1 free base was observed as well. This was consistent with the observed acid lability of Z1 nucleoside, since the matrix from which the oligonucleotide was desorbed is acidic.

In HPLC analysis of the enzymatically digested oligonucleotide a 5 nmol sample was digested as described in chapter 6, except that the reaction was allowed to proceed for 12 hours rather than only 3. When products were eluted at pH 5.5 a significant amount of Z1 nucleoside was found to have released free Z1 base. The amount of deglycosylation was reduced upon changing buffer pH to 7.0. HPLC conditions and gradients were otherwise identical to those described in chapter 6. Elution times: ^{m}C (10.1 min), T (12.0 min), Z1 (28.2 min, λ_{max}= 252 nm), Z1 free base (29.3 min, λ_{max}= 252 nm).

The synthesis of Z1 phosphoramidite was based, with some modifications, upon the reported route to the Z1 model compound containing an N-butylurea rather than an unsubstituted urea and a butyl group replacing deoxyribose.[7] The synthesis is described in detail below.

2,7-Diacetamidonaphthalene (2): 2,7-dihydroxynaphthalene (12.0 g, 75 mmol), 95 mL saturated aqueous ammonium bisulfite, and 36 mL saturated aqueous ammonia were combined in a Parr vessel and heated to 170°C for 12 hours. The mixture was washed with 100 mL 2M NaOH, and the resultant gray powder isolated by centrifugation. The residue was dissolved in 1M HCl and filtered. The resultant clear solution was precipitated by neutralization with NaOH, and the off-white 2,7-diaminonaphthalene centrifuged, washed with water, and dried by azeotrope with acetonitrile. The product was found to be unstable, rapidly darkening upon standing in air, and was immediately acetylated by dissolving in 60 mL acetic acid and 25 mL acetic anhydride, and stirring at room temperature for 2 hours. The solution was concentrated, washed with water, filtered, and washed with EtOAc (**2** was found to be insoluble in both water and organic solvents). Filtration provided a white solid (10.4 g, 57% over 2 steps). ^{1}H NMR (DMSO-d_6) δ 10.0 (s, 2H), 8.12 (d, 2H, J=1.5Hz), 7.41 (dd, 2H, J=8.7Hz, 1.8Hz), 2.06 (s, 6H).

1-Nitro-2,7-diacetamidonaphthalene (3): **2** (5.0 g, 20.7 mmol) was suspended in 40 mL acetic acid, and 2.0 mL (30 mmol) concentrated aqueous nitric acid was added. The mixture was stirred at room temperature for 15 minutes, during which time the cloudy suspension became clear and deep-red, followed by precipitation of yellow product. 40 mL water was added, and the precipitate filtered and washed with dichloromethane, providing a yellow solid (5.34 g, 90%). ^{1}H NMR (DMSO-d_6) δ 10.35 (s, 1H), 10.21 (s, 1H), 8.20, (s, 1H), 7.99 (d, 1H, J=8.7Hz), 7.91 (d, 1H, J=8.7Hz), 7.74 (dd, 1H, J=8.7Hz, 1.5Hz), 7.48 (d, 1H, J=8.7Hz), 2.07 (s, 3H), 2.05 (s, 3H).

1-Amino-2,7-diacetamidonaphthalene (4): **3** (1.0 g, 3.48 mmol) was dissolved in 40 mL ethanol. Approximately 200 mg Raney nickel was added, followed by hydrazine hydrate (200 μL, 4.0 mmol). If a larger excess of hydrazine is used, a new product forms.

The mixture was stirred for 1 hour at room temperature, then filtered, washed with acetic acid, and dried. This material was carried directly into the next step. ^{1}H NMR (DMSO-d_6) δ 10.03 (s, 1H), 9.27 (s, 1H), 8.21, (s, 1H), 7.63 (d, 1H, J=8.7Hz), 7.47 (d, 1H, J=8.7Hz), 7.17 (d, 1H, J=8.7Hz), 7.02 (d, 1H, J=8.7Hz), 5.07 (s, 2H), 2.07 (s, 3H), 2.06 (s, 3H).

Acetamidonaphthimidazole (5): Crude **4** was dissolved in 30 mL acetic acid and refluxed for 3 hours. The solution was then concentrated, and the product carried on to the next step with no further purification. ^{1}H NMR (acetone-d_6) δ 10.10 (s, 1H), 8.72, (s, 1H), 7.80 (d, 1H, J=10.5Hz), 7.46 (m, 2H), 7.43 (m, 1H), 2.54 (s, 3H), 2.10 (s, 3H).

Naphthimidazole (6): Crude **5** was dissolved in 30 mL 2M aqueous HCl and stirred at 100°C for 10 hours. Upon neutralization with NaOH a tan colored precipitate formed. 30 mL brine was added, and the precipitate filtered and washed with water. The solid was dried and purified by silica chromatography (10% MeOH/EtOAc) to provide a light pink foam (544 mg, 79% over 3 steps from **3**). TLC (10% MeOH/EtOAc) R_f 0.40; ^{1}H NMR (acetone-d_6) δ 7.61 (d, 1H, J=8.7Hz), 7.46 (d, 1H, J=2.1Hz), 7.38 (d, 1H, J=8.7Hz), 7.27 (d, 1H, J=8.7Hz), 6.85 (dd, 1H, J=8.7Hz, 2.1Hz), 5.4 (b, 2H), 2.59 (s, 3H).

8: Naphthimidazole **6** (75 mg, 0.38 mmol) was dissolved in 5 mL acetonitrile. NaH (20 mg, 0.45 mmol as 60% suspension in mineral oil) was added, and the mixture stirred for 45 minutes, at which time the chlorosugar **7**[9] was added (148 mg, 0.38 mmol), and stirring continued for 1 hour. The mixture was diluted with 25 mL EtOAc, extracted with water, and concentrated. Silica chromatography (ether, then EtOAc) provided a yellow foam (60 mg, 29%). TLC (EtOAc) R_f 0.50; ^{1}H NMR (acetone-d_6) δ 8.02 (d, 2H), 7.92 (d, 2H), 7.72 (d, 1H, J=8.4Hz), 7.58 (d, 1H, J=1.8Hz), 7.49 (d, 1H, J=8.7Hz), 7.35 (m, 3H), 7.26 (d, 2H), 7.09 (m, 1H), 6.95 (dd, 1H, J=8.7Hz, 1.8Hz), 5.84 (m, 1H), 4.88

(m, 3H), 4.70 (m,1H), 3.05 (m, 1H), 2.89 (m, 1H), 2.74 (s, 3H), 2.41 (s, 3H), 2.37 (s, 3H).

Urea (9): **8** (200 mg, 0.36 mmol) was dissolved in 4 mL THF, and 3 equivalents DIEA added. TMS isocyanate was added (~ 10 equivalents) in portions every few hours. After 48 hours the mixture was concentrated and purified by silica chromatography (EtOAc) to provide a light pink solid (153 mg, 71%). TLC (EtOAc) R_f 0.20; ^{1}H NMR (acetone-d_6) δ 10.57 (b, 1H), 8.94 and 8.56 (s, 1H, attributed to urea rotational isomers), 8.04 (d, 2H), 7.95 (s, 1H), 7.92 (d, 2H), 7.66 (d, 1H, J=8.4Hz), 7.59 (d, 1H, J=8.4Hz), 7.34 (m, 3H), 7.26 (d, 2H), 6.97 (m, 1H), 6.45 (b, 2H), 6.01 (m, 1H), 4.88 (m, 2H), 4.71 (m,1H), 3.24 (m, 1H), 2.90 (m, 1H), 2.77 (s, 3H), 2.41 (s, 3H), 2.37 (s, 3H).

Z1 nucleoside (10): **9** (150 mg, 0.25 mmol) was dissolved in 3 mL ethanol, and 3 mL 2M KOH in 1:1 ethanol/water was added. The mixture was stirred for 15 minutes, after which 350 mg ammonium chloride was added to neutralize. The mixture was concentrated and purified by silica chromatography (30% MeOH/EtOAc) to produce a white solid (90 mg, 100%). UV λ_{max} 253 nm (47,000); TLC (25% MeOH/EtOAc) R_f 0.40; ^{1}H NMR (CD_3OD) δ 8.57 (d, 1H, J=1.5Hz), 7.92 (d, 1H, J=8.7Hz), 7.65 (d, 1H, J=8.7Hz), 7.55 (d, 1H, J=8.7Hz), 7.53 (m, 1H), 6.77 (m, 1H), 4.67 (m,1H), 4.00 (m, 3H), 2.81 (m, 1H), 2.79 (s, 3H), 2.43 (m, 1H).

5'-(4,4'-dimethoxytrityl)-Z1 (11): Nucleoside **10** (80 mg, 0.224 mmol) was dissolved in 10 mL pyridine, and DBU (135 mL, 0.89 mmol) was added. DMT chloride (96 mg, 0.28 mmol) was added and the mixture stirred at 0 °C for 1 hour. 30 mL EtOAc was added, the solution extracted with saturated aqueous $NaHCO_3$, and concentrated. Silica chromatography (10% MeOH/EtOAc) yielded a white solid (91 mg, 61%). TLC (10% MeOH/EtOAc) R_f 0.25; ^{1}H NMR (acetone-d_6) δ 10.15 (b, 1H), 8.57 (s, 1H), 7.95

(d, 1H, J=8.7Hz), 7.67 (d, 1H, J=8.7Hz), 7.59 (d, 1H, J=8.7Hz), 7.44 (m, 2H), 7.31 (m, 7H), 7.12 (m, 2H), 6.83 (m, 1H), 6.72 (m, 4H), 6.55 (b, 2H), 4.81 (b, 1H), 4.66 (m,1H), 4.25 (m, 1H), 3.76 (m, 1H), 3.72 (s, 3H), 3.68 (s, 3H), 3.32 (m, 1H), 2.68 (m, 1H), 2.74 (s, 3H), 2.52 (m, 1H).

Z1 phosphoramidite (12): 5'-DMT nucleoside **11** (80 mg, 0.12 mmol) was dissolved in 3 mL THF. DIEA (63 μL, .36 mmol) was added. Stirring at room temperature, 2-cyanoethyl-N,N-diisopropylchlorophosphoramidite (34 μL, .15 mmol) was added dropwise, and the mixture was stirred for 1 hour. The reaction was quenched with 2 drops MeOH, 25 mL EtOAc was added, and the mixture extracted with saturated $NaHCO_3$ and concentrated. Silica chromatography (7% MeOH/EtOAc) provided a colorless oil (67 mg, 64%). TLC (7% MeOH/EtOAc) R_f 0.30; ^1H NMR (mixture of diastereomers)(acetone-d_6) δ 10.05 (b, 1H), 8.60 (m, 1H),), 7.95 (d, 1H, J=8.7Hz), 7.66 (d, 1H, J=8.7Hz), 7.59 (d, 1H, J=8.7Hz), 7.44 (m, 2H), 7.31 (m, 7H), 7.12 (m, 2H), 6.78 (m, 4H), 6.65 (m, 1H), 6.55 (b, 2H), 4.35 (m, 1H), 3.72 (s, 3H), 3.67 (s, 3H), 3.5-4.0 (m, 6H), 2.5-3.0 (m, 5H), 1.24 (m, 12H).

Quantitative DNase I footprint titrations: 3'-end labeling of the fragment from pYSPEC2 and footprint titrations were performed as previously described by E.S. Priestley.[10]

References

(1) Hunziker, J.; Priestley, E.S.; Brunar, H.; Dervan, P.B. *J. Am. Chem. Soc.* **1995**, *117*, 2661-2662.

(2) Liberles, D.A.; Dervan, P.B. Unpublished results.

(3) Durland, R.H.; Revankar, G.R.; Tinsley, J.H.; Myrick, M.A.; Seth, D.M.; Rayford, J.; Singh, P.; Jayaraman, K. *Nucleic Acids Res.* **1994**, *22*, 3233-3240.

(4) Durland, R.H.; Rao, T.S.; Bodepudi, V.; Seth, D.M.; Jayaraman, K.; Revankar, G.R. *Nucleic Acids Res.* **1995**, *23*, 647-653.

(5) Huang, C.Y.; Cushman, C.D.; Miller, P.S. *J. Org. Chem.* **1993**, *58*, 5048-5049.

(6) Huang, C.Y.; Miller, P.S. *J. Am. Chem. Soc.* **1993**, *115*, 10456-10457.

(7) Zimmerman, S.C.; Schmitt, P. *J. Am. Chem. Soc.* **1995**, *117*, 10769-10770.

(8) Wada, C.; Dervan, P.B. Unpublished results.

(9) Hoffer, M. *Chem. Ber.* **1960**, *93*, 2777-2781.

(10) Priestley, E.S. Ph.D. Thesis, California Institute of Technology, 1996.

CHAPTER EIGHT

The H-Pin Polyamide Motif for Recognition of the Minor Groove of DNA

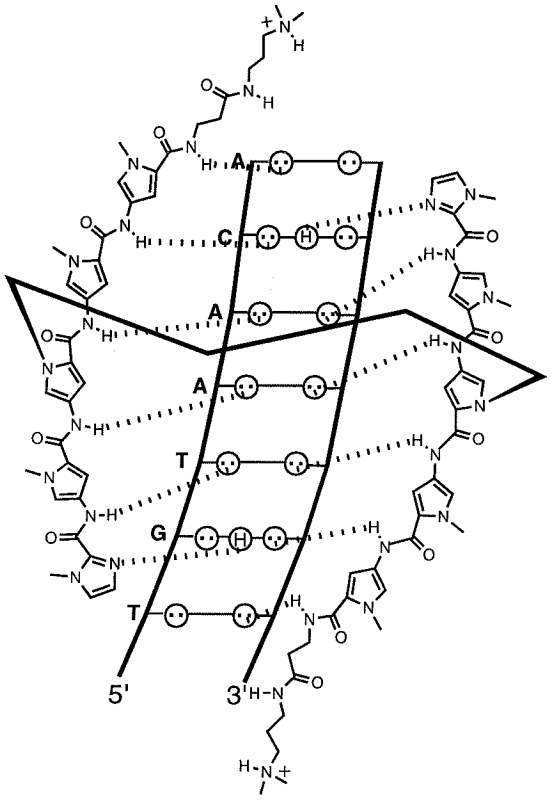
Introduction

Extensive efforts over the last several years have led to the development of 2:1 antiparallel pyrrole-imidazole polyamide-DNA complexes as a model for the design of molecules for sequence-specific recognition in the minor groove of DNA. The sequence specificity of a given polyamide is dictated by a linear combination of side-by-side pairings of pyrrole and imidazole amino acids. An imidazole on one ligand complemented by a pyrrole on the second recognizes a G•C base pair, while a pyrrole/imidazole combination recognizes C•G. Pyrrole/pyrrole pairs are degenerate for A•T or T•A base pairs.[1-4]

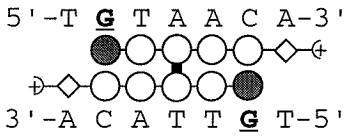
Covalently linking polyamide homodimers and heterodimers within the 2:1 motif with γ-aminobutyric acid serving as a "turn" has led to designed hairpin ligands with enhanced affinities and specificities.[5] Furthermore, development of methodology for the solid phase synthesis of pyrrole-imidazole polyamides[6] has facilitated the design of a wide variety of ligands[7] and to applications in the gene-specific regulation of transcription *in vivo*.[8]

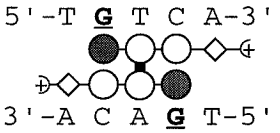
An early approach to the covalent linkage of polyamides for the 2:1 motif entailed connection across a central pyrrole-pyrrole pair through the ring nitrogens, which face out of the minor groove (Figure 1).[9] The synthesis of this class of polyamides, which we call the "H-pin" motif was, however, nontrivial and made elaboration to larger systems difficult. Here we report the extension of solid phase methodology to the synthesis of H-pin polyamides through the use of a bispyrrole monomer building block and established protocols for the synthesis of pyrrole-imidazole polyamides containing this linkage. We have prepared a series of H-pin polyamides **1-3** (Figure 2) and evaluated them relative to

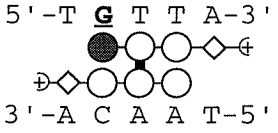
Figure 1. (Top) Binding model for the complex formed between H-pin (ImPy**Py**PyPy-β-Dp)$_2$C$_4$ (**3**) and the targeted site 5'-TGTAACA-3'. Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dotted lines. (Bottom) Schematic binding models for the above complex and for the (Im**Py**Py-β-Dp)$_2$C$_4$ (**1**) • 5'-TGTCA-3' and (Im**Py**Py-β-Dp)C$_4$(AcPy**Py**Py-β-Dp) (**2**) • 5'-TGTTA-3' complexes. The shaded and light circles represent imidazole and pyrrole rings, respectively, and diamonds represent C-terminal β-alanine residues. The dark lines connecting the central pyrrole rings represent the tetramethylene linkers.

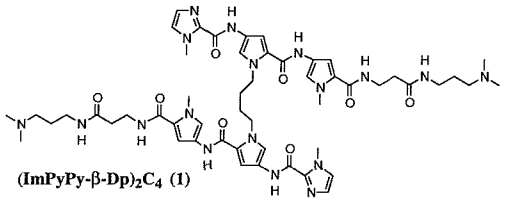


$(ImPy\mathbf{Py}PyPy\text{-}\beta\text{-}Dp)_2C_4 \bullet TGTAACA$

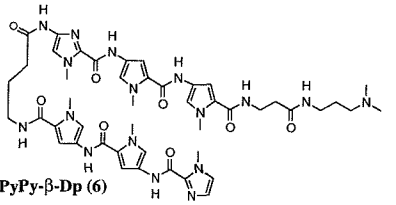




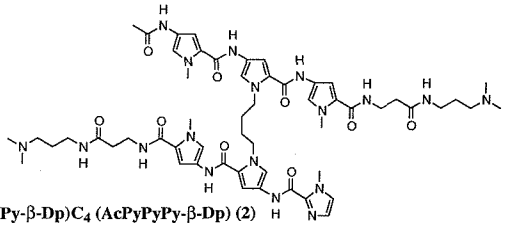




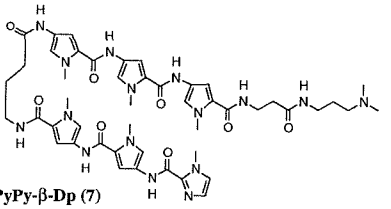
(ImPyPy-β-Dp)$_2$C$_4$ (1)



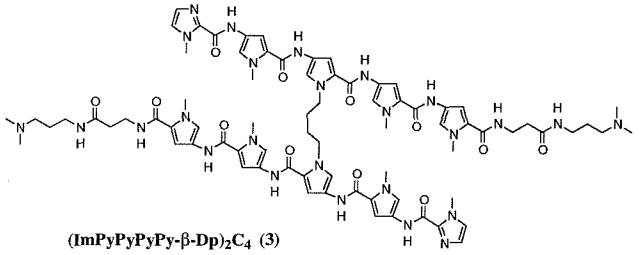
ImPyPy-γ-ImPyPy-β-Dp (6)



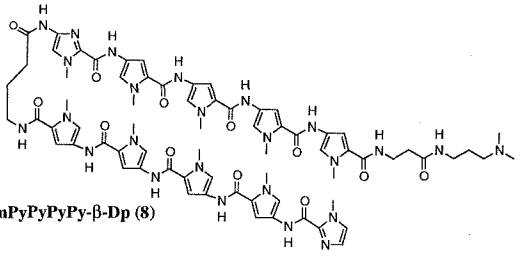
(ImPyPy-β-Dp)C$_4$ (AcPyPyPy-β-Dp) (2)



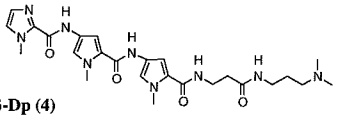
ImPyPy-γ-PyPyPy-β-Dp (7)



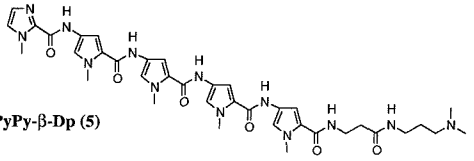
(ImPyPyPyPy-β-Dp)$_2$C$_4$ (3)



ImPyPyPyPy-γ-ImPyPyPyPy-β-Dp (8)



ImPyPy-β-Dp (4)



ImPyPyPyPy-β-Dp (5)

Figure 2. Structures of the three H-pin polyamides **1-3**, the unlinked analogues **4-5**, and hairpin analogues **6-8**.

analogous unlinked and hairpin polyamides **4-8** with respect to DNA binding affinity and specificity.

Results and Discussion

Polyamide synthesis. Construction of H-pin polyamides by solid phase methodology required synthesis of a suitably protected bispyrrole monomer unit, **9** (Figure 3). The synthesis is amenable to large scale preparation, as only one column chromatography is required. The tetramethylene linker was chosen based on early linker length optimization studies in which linkers of 3-6 methylenes were compared.[9a] The synthesis of a homodimeric H-pin polyamide is illustrated for (Im**Py**Py-β-Dp)$_2$C$_4$, **1** (Figure 4) (the boldface pyrrole represents the position of the tetramethylene linker). At the appropriate position the bis-OBt ester of **9**, generated *in situ*, was coupled to the resin. This generated an intermediate resin-bound mono-OBt ester with the amino terminus of both chains Boc protected. The C-terminal end of the polyamide was completed by coupling the appropriate amine, in this case NH$_2$Py-β-Dp, prepared separately by solid phase methodology, to the resin-bound polyamide. The N-terminal end of each chain was completed by removal of the Boc groups and coupling of imidazole-2-carboxylic acid according to standard protocols. The polyamide was cleaved from the resin with (N,N-dimethylamino)propylamine and purified by HPLC according to standard protocols.

The synthesis of a heterodimeric H-pin polyamide is illustrated for (Im**Py**Py-β-Dp)C$_4$(AcPy**Py**Py-β-Dp), **2** (Figure 5). The synthesis proceeded as for the homodimer up to the step of extending the N-termini of the two chains, which are different in this case. At this stage the two Boc groups were removed and partial acylation was performed with

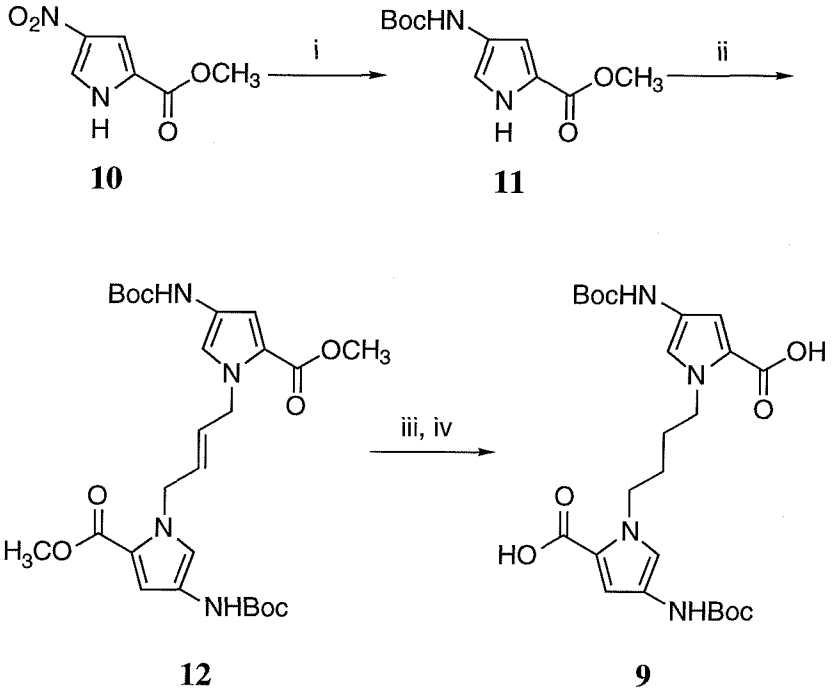
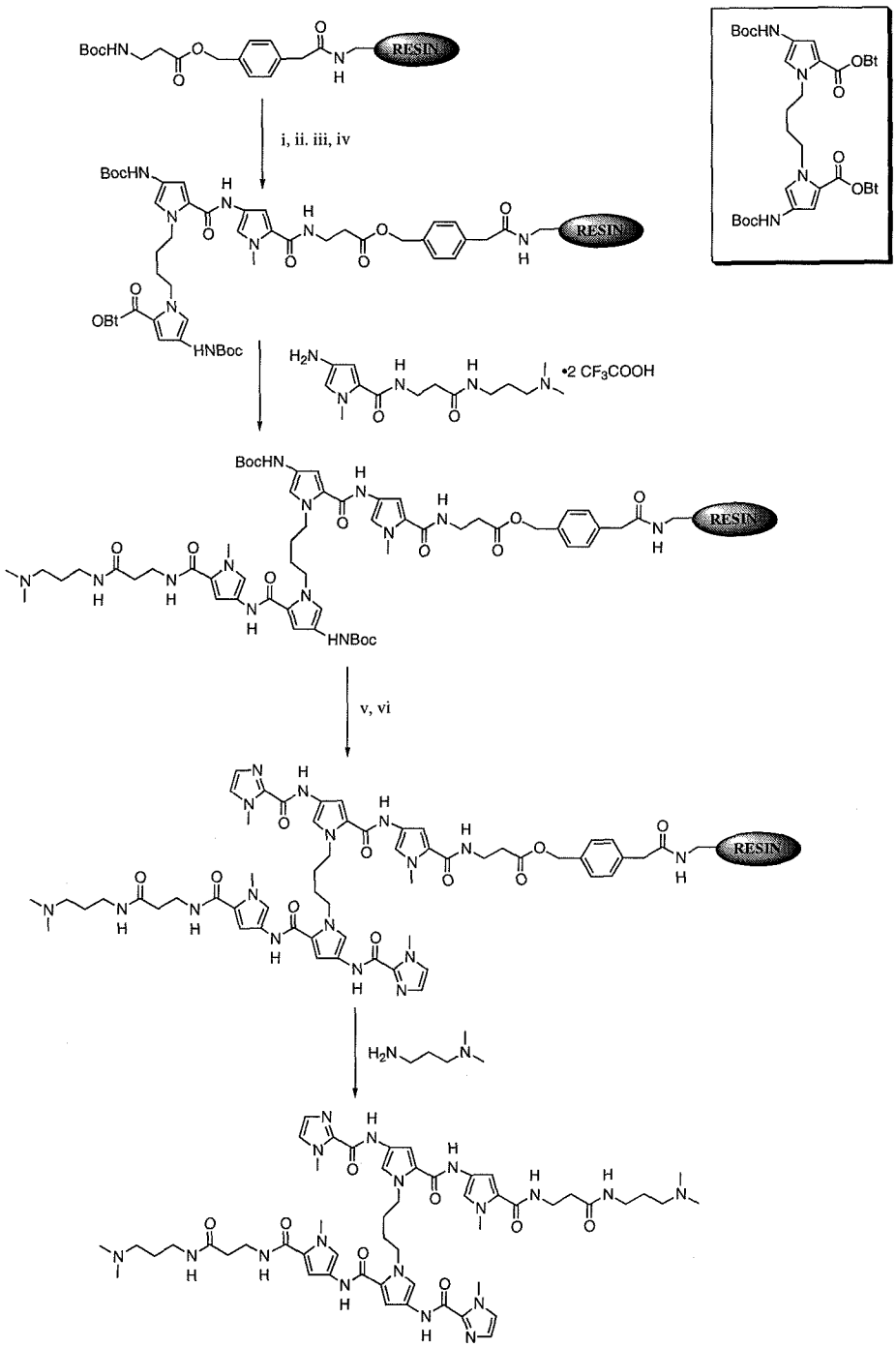
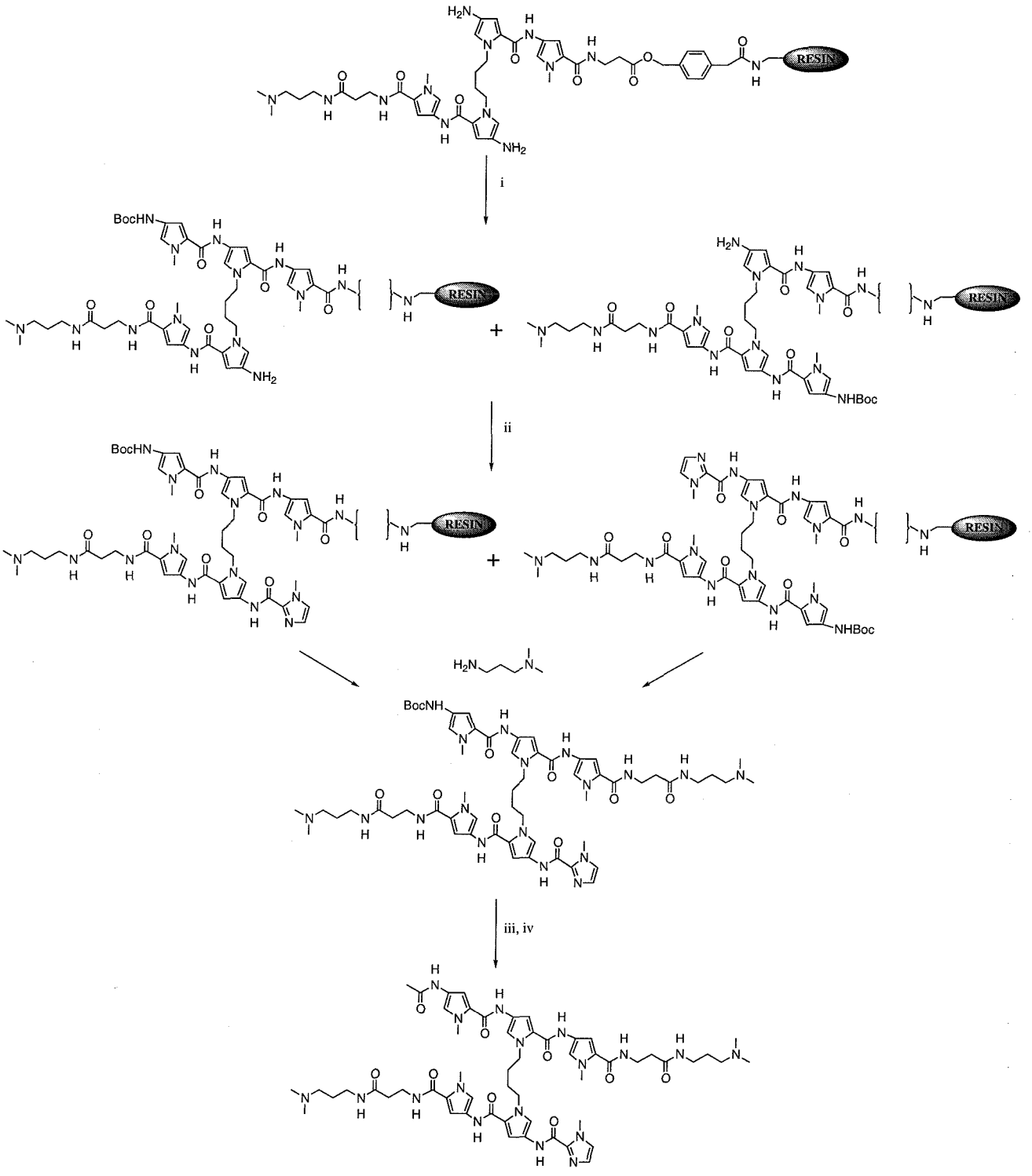


Figure 3. Synthesis of bispyrrole monomer **9**: (i) 500 psi of H_2, 10% Pd/C, Boc_2O, DIEA, DMF; (ii) 1,4-dibromo-*E*-2-butene (0.55 equiv.), K_2CO_3, acetone, 56°C; (iii) 150 psi of H_2, 10% Pd/C, EtOAc; (iv) 1M KOH, 3:1 MeOH/H_2O, 45°C.

Figure 4. Solid phase synthetic scheme for homodimeric H-pin (Im**Py**Py-β-Dp)$_2$C$_4$. (**1**). (i) 80% TFA/DCM, 0.4M PhSH; (ii) Boc-Py-OBt, DIEA, DMF; (iii) 80% TFA/DCM, 0.4M PhSH; (iv) **9**-bis OBt ester, DIEA, DMF; (v) 80% TFA/DCM, 0.4 M PhSH; (vi) imidazole-2-carboxylic acid OBt ester, DIEA, DMF.

Figure 5. Solid phase synthetic scheme for heterodimeric H-pin (Im**Py**Py-β-Dp)C$_4$(AcPy**Py**Py-β-Dp) (**2**), starting from the point where it branches off from the synthetic scheme for homodimers (Figure 4). (i) 0.4 M Boc-Py-OBt, 2:1 DMF/DIEA, 3.5 min., 22 °C; (ii) imidazole-2-carboxylic acid, HBTU, DIEA, DMF; (iii) TFA; (iv) Ac_2O, DIEA, DMF.


BocHN
RESIN
i, ii. iii, iv
OBt
HNBoc
•2 CF3COOH
v, vi


i
ii
iii, iv
RESIN
NHBoc
BocHN
BocNH

Boc-Py-OBt ester. We note that if the C-terminal side of the linker on each chain is the same, acylation of either free amine gives the same final product (Figure 5). This requirement for symmetry to the C-terminal side of the linker limits the range of heterodimeric H-pin polyamides that can be prepared by this route.

With the two chains now differentiated, it was possible to complete the synthesis by acylating with imidazole-2-carboxylic acid. The N-terminal Boc protected polyamide was cleaved from the resin and purified by HPLC, utilizing the hydrophobicity of the Boc to separate the desired product from side products. Polyamide **2** was prepared by solution phase removal of the N-terminal Boc, acetylation, and HPLC purification.

Binding affinities and specificities. Affinities and specificities of H-pin polyamides **1-3** were determined by quantitative DNase I footprint titration[10] on 3'-^{32}P labeled 135 bp pMM5 *Eco* RI/*Bsr* BI (**1**,**2**) or 252 bp pJK7 *Eco* RI/*Pvu* II (**3**) restriction fragments (10 mM Tris-HCl, 10 mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, pH 7.0, 22°C). Analysis of the footprints (Figure 6) and binding isotherms generated from the titrations (Figure 7) indicates that H-pin polyamides **1-3** bind to their designated sites as 1:1 complexes, as expected for covalent dimers binding according to the side-by-side 2:1 polyamide-DNA model. The association constants obtained were compared to those found on the same restriction fragments for analogous unlinked and hairpin polyamides (Table 1).

Comparison of the affinities of (Im**Py**Py-β-Dp)$_2$C$_4$ (**1**) for its match 5'-aTGACAt-3' (K_a= 9.3 x 10^6 M^{-1}) and mismatch 5'-tTGTTAg-3' (K_a= 9.9 x 10^5 M^{-1}) sites reveals an approximately 10-fold specificity. Its affinity for the match site is at least 186-fold greater than that of its unlinked analog ImPyPy-β-Dp. Notably, the affinity of ImPyPy-β-Dp (**4**) is lower than that of ImPyPy-Dp, which does not contain the C-terminal β-alanine residue,

Table 1. Association Constants (M^{-1}) for Polyamides **1-8**.[a,b]

homodimer	ImPyPy-β-Dp (**4**)	(Im**Py**Py-β-Dp)$_2$C$_4$ (**1**)	ImPyPy-γ-ImPyPy-β-Dp (**6**)
5' TGTCA 3'	<5 x 10^4	9.3 (±1.6) x 10^6	2.0 (±0.3) x 10^8
5' TGTTA 3'	<5 x 10^4	9.9 (±2.7) x 10^5	7.0 (±2.9) x 10^6
specificity	-	9.4	29
heterodimer	ImPyPy-β-Dp• PyPyPy-β-Dp	(Im**Py**Py-β-Dp)C$_4$(Ac Py**Py**Py-β-Dp) (**2**)	ImPyPy-γ-PyPyPy-β-Dp (**7**)[c]
5' TGTTA 3'	-	2.0 (±0.9) x 10^6	2.9 (±0.5) x 10^8
5' TGTCA 3'	-	5.7(±1.3) x 10^5	4.6 (±1.1) x 10^6
specificity	-	3.5	60
10 Ring	ImPyPyPyPy-β-Dp (**5**)	(ImPy**Py**PyPy-β-Dp)$_2$C$_4$ (**3**)	ImPyPyPyPy-γ-ImPyPyPyPy-β-Dp (**8**)[d]
5'TGTTACA3'	1.2 (±0.3) x 10^8	4.4 (±1.4) x 10^8	1.2 (±0.2) x 10^{10}
5'TGTTCCA3'	8.8 (±2.6) x 10^6	1.6 (±0.5) x 10^7	6.8 (±1.6) x 10^8
specificity	14	28	18

[a]The reported association constants are the mean values obtained from three DNase I footprint titration experiments. The standard error for each value is indicated in parentheses. [b] The assays were carried out at 22°C, pH 7.0 in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$. [c] Reference 5b. [d] Reference 7d.

and has been measured to have an affinity at the 5'-aTGACAt-3' site of 1.5 x 10^5 M^{-1}.[9c] No specific DNase I footprint was observed for this compound, and at concentrations > 20 μM only nonspecific binding to DNA is observed. Thus we assign an upper limit of K_a= 5 x 10^4 M^{-1}. The analogous hairpin, ImPyPy-γ-ImPyPy-β-Dp (**6**), binds the same match (K_a= 2.0 x 10^8 M^{-1}) and mismatch (K_a= 7.0 x 10^6 M^{-1}) sites with greater affinity than H-pin **1**. The hairpin is a ~20-fold tighter binder, and is ~3-fold more specific.

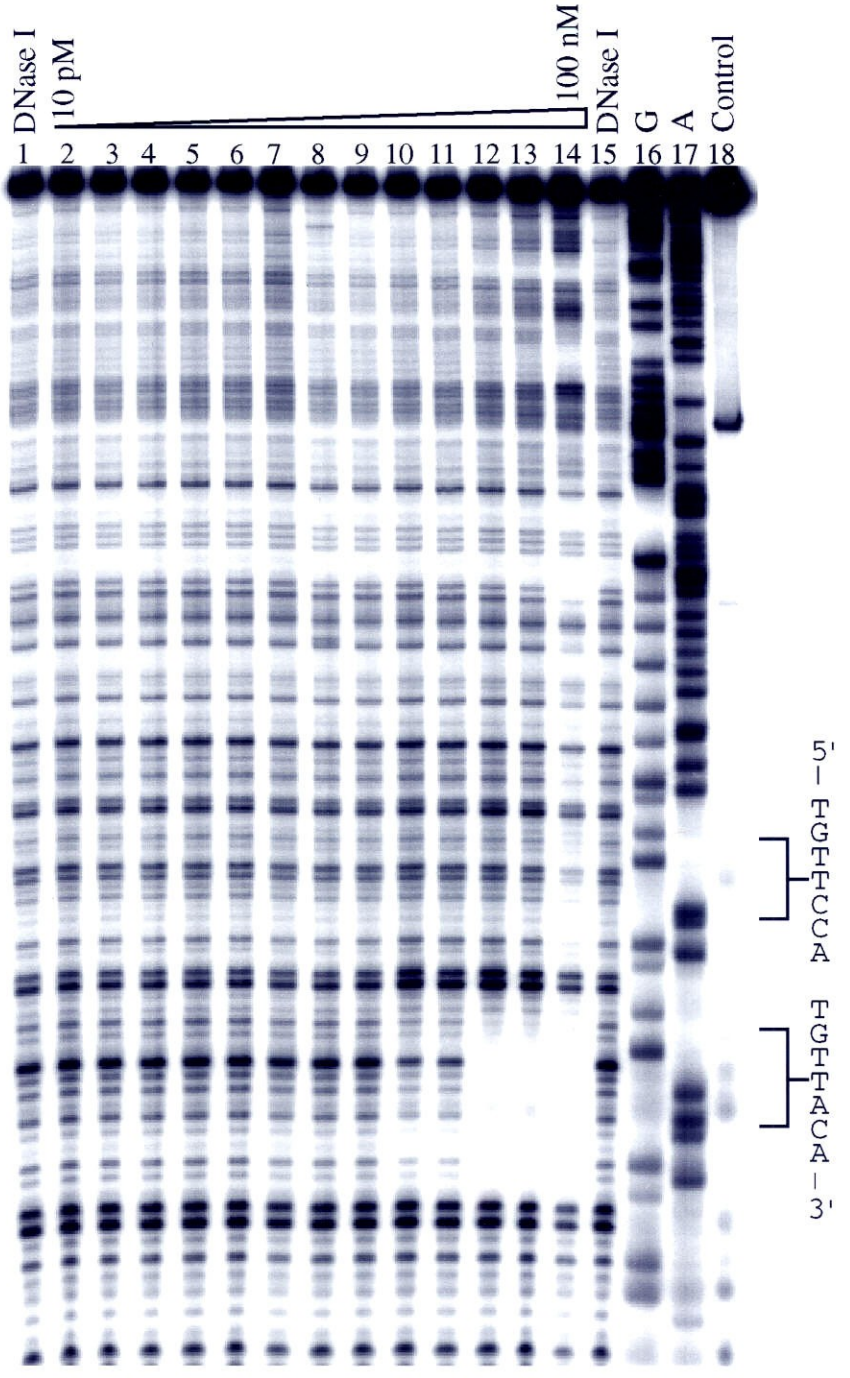
Heterodimeric H-pin (Im**Py**Py-β-Dp)C_4(AcPy**Py**Py-β-Dp) (**2**) binds its match site 5'-tTGTTAg-3' (K_a= 2.0 x 10^6 M^{-1}) and mismatch site 5'-aTGACAt-3' (K_a= 5.7 x 10^5 M^{-1}) with slightly lower affinities and reduced specificity compared to the homodimeric H-pin **1**. This may reflect a detrimental contribution from the N-terminal acetyl group, or it may be a consequence of different flanking sequences at the respective sites. Again, the analogous hairpin ImPyPy-γ-PyPyPy-β-Dp[5b] (**7**) is superior to the H-pin in affinity and specificity.

The 10-ring homodimeric H-pin (ImPy**Py**PyPy-β-Dp)$_2C_4$ (**3**) binds its match 5'-tTGTAACAg-3' (K_a= 4.4 x 10^8 M^{-1}) and mismatch 5'-tTGGAACAg-3' (K_a= 1.6 x 10^7 M^{-1}) sites with slightly higher affinity than the unlinked ImPyPyPyPy-β-Dp (**5**), with a ~4-fold increase in affinity and 2-fold increase in specificity. Notably, the 5-ring unlinked polyamide containing a C-terminal β-alanine has increased affinity (K_a= 1.2 x 10^8 M^{-1} vs. 4.5 x 10^7 M^{-1}) and specificity (14-fold vs. 6-fold) compared to ImPyPyPyPy-Dp.[11] Thus, for unlinked polyamides a C-terminal β-alanine is detrimental in the 3-ring system, but beneficial in the 5-ring system. The 10-ring hairpin ImPyPyPyPy-γ-ImPyPyPyPy-β-Dp[7d]

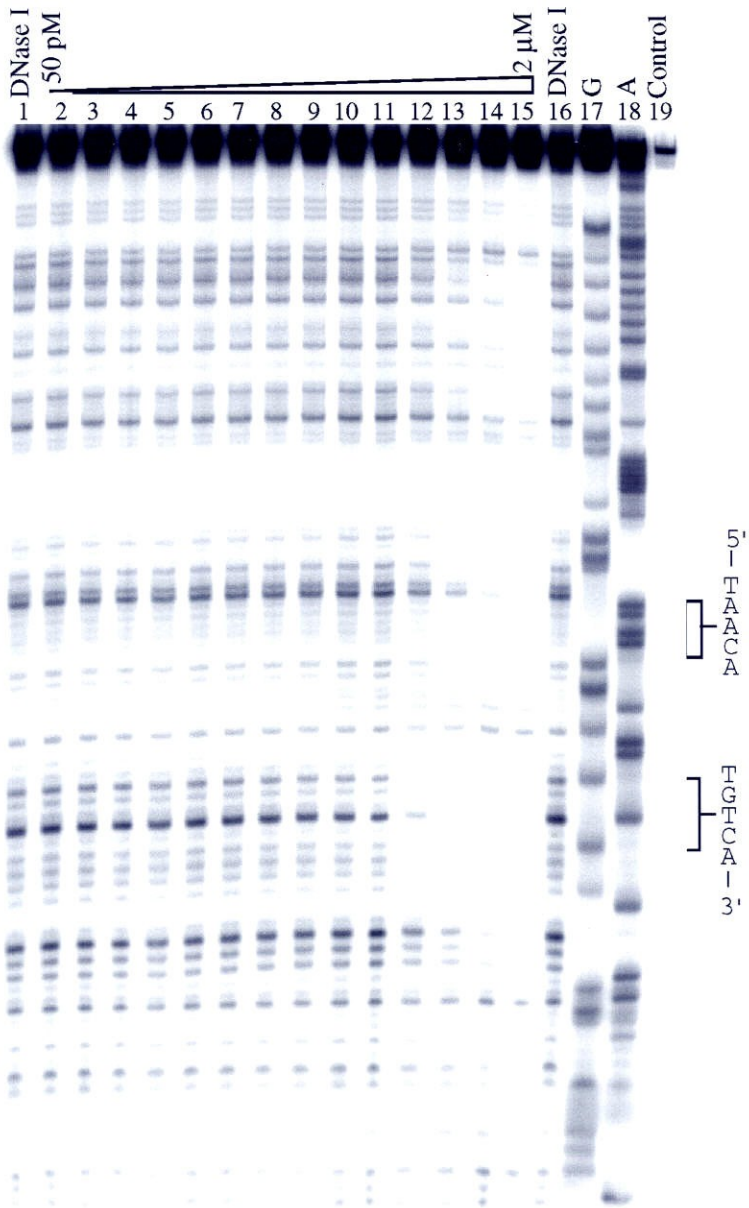
(**8**) binds more tightly (K_a= 1.2 x 10^{10} M^{-1}) to the match site, although in this case the H-pin has slightly more favorable specificity than the hairpin (28-fold vs. 18-fold).

Conclusion. We have developed methodology for the solid phase synthesis of H-pin polyamides, a class of covalently linked polyamide dimers for the sequence-specific recognition of the minor groove of DNA. The series of H-pin polyamides prepared showed increased affinity and specificity toward their designated sites. It may be possible to further optimize the linker as suggested by the greater enhancements in affinity and specificity observed for analogous hairpin polyamides. H-pin polyamides should prove useful as a complement to, or in conjunction with other polyamide motifs such as the hairpin[5,7a] and extended[7b,12] motifs in the design of non-natural ligands for the sequence-specific recognition of DNA.

Figure 6. Storage phosphor autoradiogram of an 8% denaturing polyacrylamide gel used to separate fragments generated by DNase I digestion in a quantitative footprint titration experiment with (ImPy**Py**PyPy-β-Dp)$_2$C$_4$ (**3**) on the 3'-^{32}P-labeled 252 bp *Eco* RI/*Pvu* II restriction fragment from plasmid pJK7. Lanes 1 and 15: DNase I digestion products in the absence of polyamide. Lanes 2-14: DNase I digestion products obtained in the presence of 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, 100 nM polyamide. Lanes 16 and 17: G and A sequencing lanes. Lane 18: intact DNA. The targeted binding sites are indicated on the right side of the autoradiograms. All reactions contain 15 kcpm restriction fragment, 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$.


DNase I
10 pM
100 nM
DNase I
G
A
Control
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
5'- TGTTCCA
TGTTACA -3'

Figures 7 and 8. Storage phosphor autoradiograms of 8% denaturing polyacrylamide gels used to separate fragments generated by DNase I digestion in a quantitative footprint titration experiment with (Im**Py**Py-β-Dp)$_2$C$_4$ (**1**) (Figure 7) and (Im**Py**Py)C$_4$(AcPy**Py**Py-β-Dp) (**2**) on the 135 base pair 3'-end labeled *Eco* RI/*Bsr* BI restriction fragment from plasmid pMM5. Lanes 1 and 16: DNase I digestion products in the absence of polyamide. Lanes 2-15: DNase I digestion products obtained in the presence of 50 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, 100 nM, 200 nM, 500 nM, 1 μM, 2 μM polyamide. Lanes 17 and 18: G and A sequencing lanes. Lane 19: intact DNA. The targeted binding sites are indicated on the right side of the autoradiograms. All reactions contain 15 kcpm restriction fragment, 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$.


DNase I
50 pM
2 µM
DNase I
G
A
Control
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
5'- TAACA
TGTCA -3'

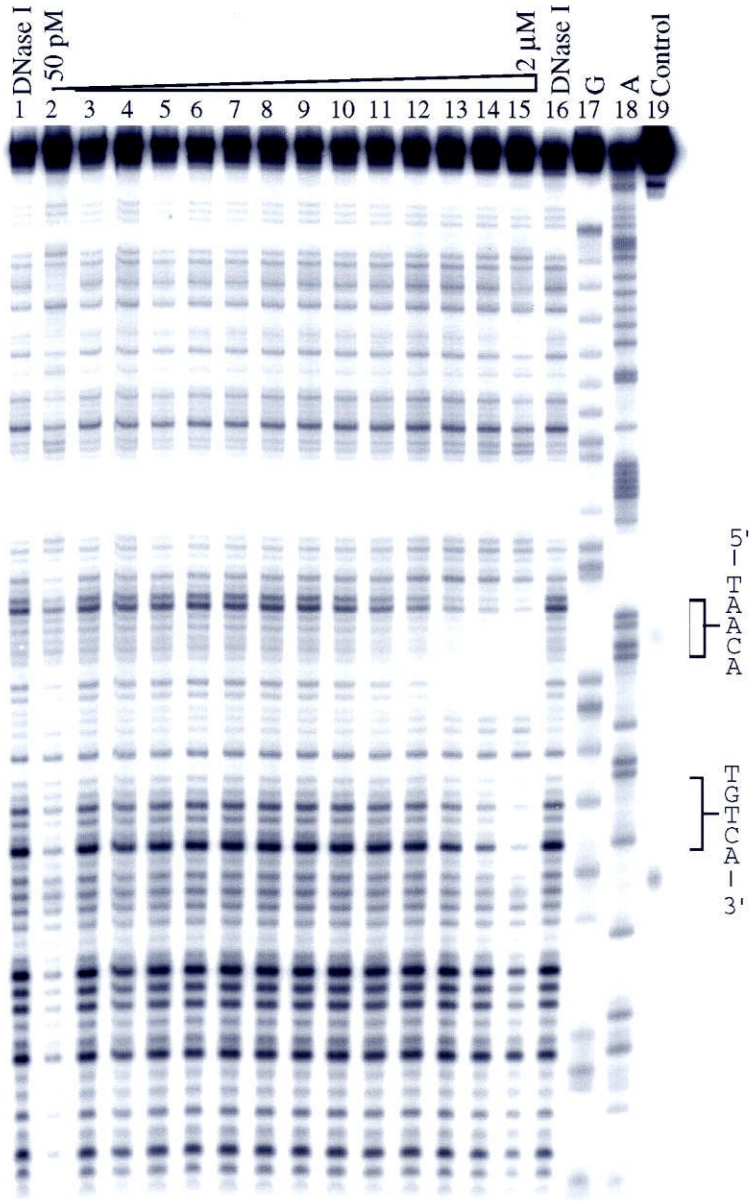
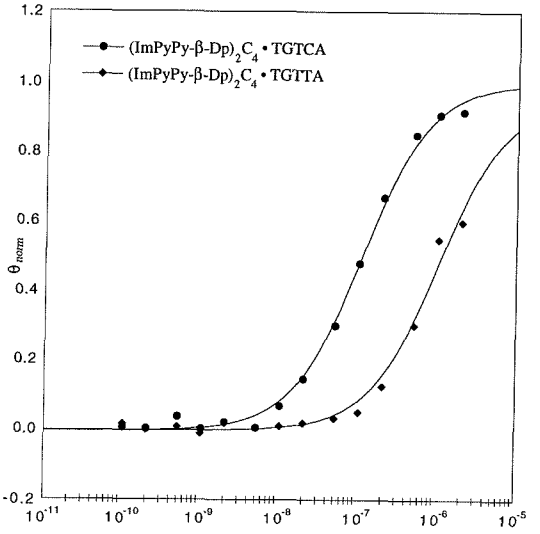
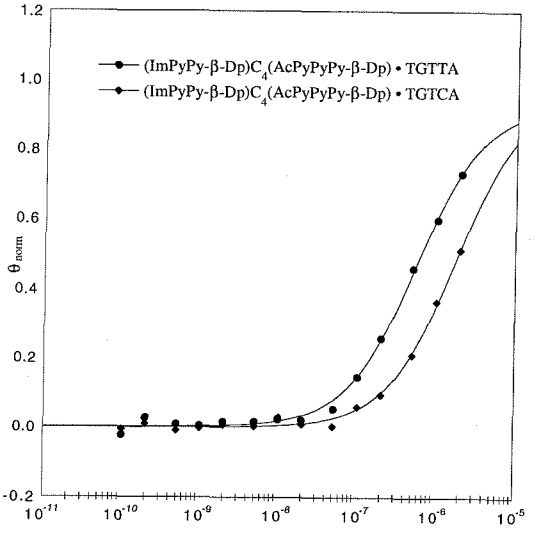
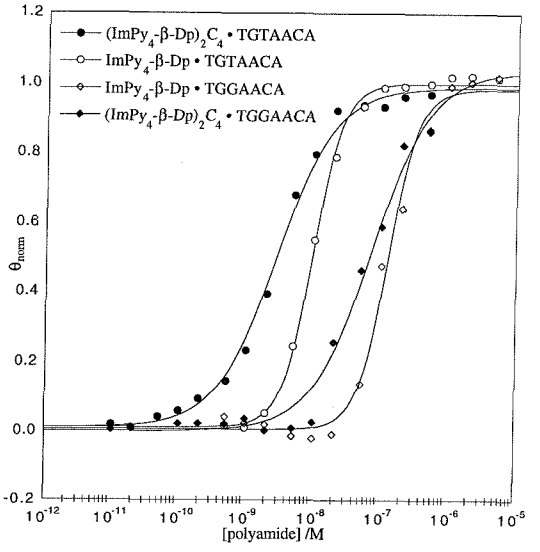

DNase I
50 pM
2 µM
DNase I
G
A
Control
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
5'- TAACA
TGTCA -3'

Figure 9. Data obtained from quantitative DNase I footprint titration experiments on polyamides (Im**Py**Py-β-Dp)$_2$C$_4$ (top), (Im**Py**Py-β-Dp)C$_4$(AcPy**Py**Py-β-Dp) (center), (ImPy**Py**PyPy-β-Dp)$_2$C$_4$ and ImPyPyPyPy-β-Dp (bottom). The (θ_{norm}, $[L]_{tot}$) data points were obtained as described in the Experimental Section, and each is the average value obtained from three experiments.


(ImPyPy-β-Dp)2C4 • TGTCA
(ImPyPy-β-Dp)2C4 • TGTTA
θnorm


(ImPyPy-β-Dp)C4(AcPyPyPy-β-Dp) • TGTTA
(ImPyPy-β-Dp)C4(AcPyPyPy-β-Dp) • TGTCA
θnorm


(ImPy4-β-Dp)2C4 • TGTAACA
ImPy4-β-Dp • TGTAACA
ImPy4-β-Dp • TGGAACA
(ImPy4-β-Dp)2C4 • TGGAACA
θnorm
[polyamide] /M

Experimental Section

Materials. Reagents and solvents for solid phase synthesis of polyamides were as described previously.[6] Di-*tert*-butyl dicarbonate was purchased from Peptides International, and K_2CO_3 from Mallinckrodt. Other reagents were from Aldrich and used without further purification.

^1H and ^{13}C NMR spectra were recorded on a GE QE300 instrument operating at 300 MHz and 75 MHz, respectively. Chemical shifts are reported in parts per million relative to solvent residual signal. UV spectra were measured on a Hewlett-Packard 8452A diode array spectrophotometer. High resolution mass spectra were obtained at the Mass Spectrometry Laboratory at the University of California, Riverside. Matrix assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF) was carried out at the Protein and Peptide Microanalytical Facility at the California Institute of Technology. Thin layer chromatography was performed on silica gel 60 F_{254} precoated plates, and chromatographic separations were performed with EM silica gel 60 (230-400 mesh). HPLC analysis was performed on a HP 1090M analytical HPLC using a Rainin C_{18} Microsorb MV reversed-phase column in 0.1% (w/v) aqueous TFA with acetonitrile as eluent and a flow rate of 1.0 mL/min, gradient elution 1.25% acetonitrile/min. Preparative HPLC was performed on a Beckman HPLC using a Waters DeltaPak 25 x 100 mm, 100 μm C_{18} column, 0.1% TFA, 0.25% acetonitrile/min gradient. Water was obtained from a Millipore MilliQ water purification system and all buffers were 0.2 μm filtered.

Methyl 4-[(*tert*-butoxycarbonyl)amino]-pyrrole-2-carboxylate (11). Methyl 4-nitropyrrole-2-carboxylate[13] (**10**) (10.0 g, 58.8 mmol) was dissolved in 50 mL N,N-dimethylformamide and 12.5 mL N,N-diisopropylethylamine. The solution was purged

with argon for 5 minutes, then 10% Pd/C (1.8 g) and di-*tert*-butyl dicarbonate (13.0 g, 59.5 mmol) were added. The mixture was stirred vigorously under 500 psi of H_2 for 2.5 hours. Pd/C was removed by filtration, DMF was removed under reduced pressure, and the product was purified by silica gel chromatography (gradient of 3:1 hexanes/diethyl ether to 2:3 hexanes/diethyl ether) to yield **11** (8.23 g, 58%) as a white solid: TLC (2:3 hexanes/diethyl ether R_f 0.4; ^{1}H NMR (DMSO-d_6) δ 11.55 (s, 1H), 9.10 (s, 1H), 6.93 (s, 1H), 6.57 (s, 1H), 3.70 (s, 3H), 1.41 (s, 9H); ^{13}C NMR ($CDCl_3$) δ 161.3, 153.2, 113.2, 109.8, 106.7, 102.6, 80.2, 51.7, 28.4; MS *m/e* 240.1111 (240.1110 calcd. for $C_{11}H_{16}N_2O_4$).

1,4-bis-[methyl-N,N'-4-[(*tert*-butoxycarbonyl)amino]-pyrrolyl-2-carboxylate]-(*E*)-2-butene (12). Pyrrole **11** (2.66 g, 11.1 mmol) was dissolved in 60 mL acetone. Anhydrous K_2CO_3 (5.3 g, 38.3 mmol) was added, followed by 1,4-dibromo-2-butene (predominantly *trans*) (1.30 g, 6.10 mmol, 0.55 equivalents). The mixture was refluxed for 40 hours. At 15 hours the KBr generated was filtered and 5.0 g fresh K_2CO_3 was added. After removal of salts by filtration, acetone was removed under reduced pressure and the mixture taken up in 2:1 hexanes/diethyl ether, from which **12** precipitated as a white solid (1.85 g, 64%): TLC (2:3 hexanes/diethyl ether) R_f 0.3; ^{1}H NMR (DMSO-d_6) δ 9.12 (s, 2H), 7.07 (s, 2H), 6.58 (s, 2H), 5.56 (br, 2H), 4.77 (br, 4H), 3.64 (s, 6H), 1.39 (s, 18H); ^{13}C NMR ($CDCl_3$) δ 161.2, 153.3, 129.2, 122.7, 118.8, 107.9, 102.6, 80.3, 51.2, 49.9, 28.4; MS *m/e* 533.2620 (M+H 533.2611 calcd. for $C_{26}H_{37}N_4O_8$).

1,4-bis-[methyl-N,N'-4-[(*tert*-butoxycarbonyl)amino]-pyrrolyl-2-carboxylate]-butane (13). **12** (2.30 g, 4.32 mmol) was dissolved in 30 mL ethyl acetate. Pd/C (500 mg) was added, and the mixture stirred under 150 psi of H_2 for 2 hours. Filtration and removal of ethyl acetate under reduced pressure yielded **13** as a white solid (2.12 g, 92%): TLC (2:3 hexanes/diethyl ether) R_f 0.3; ^{1}H NMR (DMSO-d_6) δ 9.14 (s, 2H), 7.12 (s, 2H),

6.59 (S, 2H), 4.18 (m, 4H), 3.68 (s, 6H), 1.53 (m, 4H), 1.41 (s, 18H); ^{13}C NMR ($CDCl_3$) δ 161.3, 153.2, 122.3, 118.9, 107.9, 102.6, 80.2, 51.2, 48.6, 28.4, 28.2; MS *m/e* 535.2778 (M+H 535.2768 calcd. for $C_{26}H_{39}N_4O_8$).

1,4-bis-[N,N'-4-[(*tert*-butoxycarbonyl)amino]-pyrrolyl-2-carboxylic acid]-butane (9). Bis methyl ester **13** (2.05 g, 3.83 mmol) was dissolved in 15 mL methanol. To this solution was added 5 mL 4M aqueous KOH. The mixture was stirred at 45°C for 20 hours, over which time the cloudy suspension became clear. The solution was acidified to pH 2 with aqueous HCl, resulting in a precipitate. Methanol was removed under reduced pressure and the product extracted into 40 mL ethyl acetate. After drying *in vacuo* **9** was precipitated from 1:1 hexanes/diethyl ether as an off-white powder (1.90 g, 97%): TLC (diethyl ether) R_f 0.1; ^{1}H NMR (DMSO-d_6) δ 12.1 (br, 2H), 9.02 (s, 2H), 7.01 (s, 2H), 6.53 (s, 2H), 4.18 (m, 4H), 1.54 (m, 4H), 1.41 (s, 18H); ^{13}C NMR (DMSO-d_6) δ 162.8, 153.3, 122.7, 118.7, 109.1, 102.6, 79.1, 48.1, 28.4, 28.2; MS *m/e* 506.2367 (506.2377 calcd. for $C_{24}H_{34}N_4O_8$).

Preparation of ^{32}P end-labeled restriction fragments. Restriction endonucleases were purchased from either New England Biolabs or Boehringer Mannheim and used with the provided buffers according to manufacturer's protocols. Sequenase (version 2.0) was obtained from United States Biochemical and DNase I from Pharmacia. [α^{32}P] Thymidine-5'-triphosphate and [α^{32}P] deoxyadenosine-5'-triphosphate were purchased from Amersham. The 135 base pair 3'-end labeled *Eco* RI/*Bsr* BI restriction fragment from plasmid pMM5 was prepared and purified as previously described.[5b] The 252 base pair *Eco* RI/*Pvu* II from plasmid pJK7 was prepared and purified as previously described.[11] Chemical sequencing reactions were done according to published

procedures.[14,15] Standard techniques for DNA manipulation were employed for all transformations.[16]

Quantitative DNase I footprint titration. All reactions were done in a total volume of 50 μL, with no calf thymus DNA present. A polyamide stock solution or H_2O (for reference lanes) was added to an assay buffer containing radiolabeled restriction fragment (15000 cpm) affording final solution conditions of 10 mM Tris-HCl, 10 mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, pH 7.0, and polyamide over a range of concentrations. The solutions were allowed to equilibrate for 6 hours at 22°C. Footprinting reactions were initiated by addition of 5 μL of a DNase I solution (final concentration 0.10 units/mL) containing 1 mM dithiothreitol and allowed to proceed for 6 minutes at 22°C. Reactions were stopped by addition of 12.5 μL of a solution containing 1.25 M NaCl, 100 mM EDTA, and 0.2 mg/mL glycogen, and ethanol precipitated. The reaction mixtures were resuspended in 1X TBE/ 80% formamide denaturing loading buffer, denatured by heating at 90°C for 5 minutes, and separated by polyacrylamide gel electrophoresis on an 8% gel (5% crosslinking, 7M urea) in 1X TBE at 2000V. Gels were dried and exposed to a Molecular Dynamics storage phosphor screen.

Quantitation and data analysis. Data from footprint titrations were obtained using a Molecular Dynamics 400S PhosphorImager followed by quantitation using ImageQuant software. Volume integration of rectangles encompassing footprint sites and a reference site at which DNase I reactivity was invariant over the range of the titration generated values for site intensities (I_{site}) and reference intensity (I_{ref}). The apparent fractional occupancies (θ_{app}) of the sites was calculated using the equation

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

where I_{site}° and I_{ref}° are the site and reference intensities, respectively, for a control lane in which no polyamide was added.

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

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

where $[L]_{tot}$ is the total polyamide concentration, K_a the apparent association constant, and θ_{min} and θ_{max} the experimentally determined site saturation values when the site is unoccupied and saturated, respectively. The data was fitted using a nonlinear least squares fitting procedure with K_a, θ_{max}, and θ_{min} as adjustable parameters, and either n=1 or n=2 depending on which value of n gave a better fit (n=1 for H-pins and hairpins, n=2 for unlinked polyamides). Binding isotherms were normalized using equation (3)

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

Three independent sets of data were used in determining each association constant.

References

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

(2) (a) Pelton, J.G.; Wemmer, D.E. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5723. (b) Pelton, J.G.; Wemmer, D.E. *J. Am. Chem. Soc.* **1990**, *112*, 1393. (c) Chen, X.; Ramakrishnan, B.; Rao, S.T.; Sundaralingam, M. *Nature Struct. Biol.* **1994**, *1*, 169.

(3) (a) Mrksich, M.; Dervan, P.B.; *J. Am. Chem. Soc.* **1993**, *115*, 2572. (b) Geierstanger, B.H.; Jacobsen, J.-P.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *Biochemistry* **1994**, *33,* 3055. (c) Geierstanger, B.H.; Dwyer, T.J.; Bathini, Y.; Lown, J.W.; Wemmer, D.E. *J. Am.Chem. Soc.* **1993**, *115*, 4474.

(4) (a) Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *Science* **1994**, *266*, 646. (b) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1995**, *117*, 3325.

(5) (a) Mrksich, M.; Parks, M.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1994**, *116*, 7983. (b) Parks, M.E.; Baird, E. E.; Dervan, P.B. *J. Am. Chem. Soc.* **1996**, *118*, 6147.

(6) Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* **1996**, *118*, 6141.

(7) (a) Trauger, J.W.; Baird, E.E.; Dervan, P.B. *Nature* **1996**, *382*, 559. (b) Trauger, J.W.; Baird, E.E.; Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1996**, *118*, 6160. (c) Trauger, J.W.; Baird, E.E.; Dervan, P.B. *Chem. Biol.* **1996**, *3*, 369. (d) Turner, J.M.; Baird, E.E.; Dervan, P.B. *J. Am. Chem. Soc.* submitted.

(8) (a) Gottesfeld, J.M.; Neely, L.; Trauger, J.W.; Baird, E.E.; Dervan, P.B.; *Nature* **1997**, *387*, 202. . (b) Wagner, R.; Baird, E.E.; Trauger, J.W.; Dervan, P.B. *Proc. Natl. Acad. Sci. U.S.A.* in preparation.

(9) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1993**, *115*, 9892. (b) Dwyer, T.J.; Geierstanger, B.H.; Mrksich, M.; Dervan, P.B.; Wemmer, D.E. *J. Am. Chem. Soc.* **1993**, *115*, 9900. (c) Mrksich, M.; Dervan, P.B. *J. Am. Chem. Soc.* **1994**, *116*, 3663. (d) Chen, Y.H.; Lown, J.W. *J. Am. Chem. Soc.* **1994**, *116*, 6995.

(10) (a) Brenowitz, M.; Senear, D.F.; Shea, M.A.; Ackers, G.K. *Methods Enzymol.* **1986**, *130*, 132. (b) Brenowitz, M.; Senear, D.F.; Shea, M.A.; Ackers, G.K. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 8462. (c) Senear, D.F.; Brenowitz, M.; Shea, M.A.; Ackers, G.K. *Biochemistry* **1986**, *25*, 7344.

(11) Kelly, J.J.; Baird, E.E.; Dervan, P.B. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 6981.

(12) Swalley, S.E.; Baird, E.E.; Dervan, P.B. *Chem. Eur. J.* submitted.

(13) Hale, W.J.; Hoyt, W.V. *J. Am. Chem. Soc.* **1915**, *37*, 2538.

(14) Iverson, B.L.; Dervan, P.B. *Nucleic Acids Res.* **1987**, *15*, 7823.

(15) Maxam, A.M.; Gilbert, W.S. *Methods Enzymol.* **1980**, *65*, 499.

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