

**ENERGETICS OF TRIPLE HELIX FORMATION BY
OLIGONUCLEOTIDES CONTAINING NONNATURAL BASES**

Thesis by

Eldon Scott Priestley

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1996

(Submitted January 10, 1996)

© 1996

Eldon Scott Priestley

All Rights Reserved

To Jennifer and to my Parents

Acknowledgments

I would like to thank my advisor, Professor Peter B. Dervan, for his guidance and support over the past four years. I will particularly remember the enthusiasm and intensity of Peter's approach to science. It has been a wonderful opportunity to be a part of the Dervan research group at Caltech.

To my thesis committee, Professors Andrew Myers, Douglas Rees, and Barbara Imperiali, go thanks for taking the time to read and critique my work both at candidacy and over the past several months.

The members of the Dervan group, both past and present, have contributed greatly to my experience at Caltech. Their advice and technical suggestions have played a key role through all of this work. More importantly, however, their commitment and motivation has inspired me, and I thank them each. I am particularly grateful to Pete Beal for his daily guidance and example during my first year. I would also like to thank Scott Singleton for his thoughtful advice and encouragement while working out the quantitative footprinting assay.

To my mother and father go my deep gratitude for their love and support from the beginning.

And finally, to my wife Jennifer, I offer my deepest love and gratitude for her love, patience, and encouragement through the ups and downs of graduate school. I look forward to our future together...

Abstract

Oligonucleotide directed triple helix formation is a powerful chemical approach to the sequence specific recognition of duplex DNA. In this approach, pyrimidine oligonucleotides bind homopurine duplex sequences by formation of T•AT and C+GC base triplets. Likewise, purine oligonucleotides also bind homopurine sequences by formation of G•GC and either A•AT or T•AT triplets. The two motifs differ in that the third strand oligonucleotide binds parallel to the homopurine duplex strand in the “pyrimidine motif” and antiparallel in the “purine motif.” This approach currently lacks generality, since there are no third strand bases available for highly specific, high affinity recognition of CG and TA base pairs in either structural motif. In addition, the pyrimidine motif suffers from pH dependent triplex stability introduced by the C+GC base triplets, while the purine motif suffers from competing aggregation due to G quartet formation by the G-rich third strand. In principle, these problems may be addressed by the development of nonnatural bases with the desired properties.

This thesis describes the application of quantitative DNase footprinting methods to determine the energetics of triple helix formation by oligonucleotides containing nonnatural bases. Chapter One provides an introduction to current research on triple helix formation. Chapter Two describes an NMR study of an intramolecular triplex containing the nonnatural base P, which provides pH independent recognition of GC base pairs. Chapter Three describes the quantitation of the specificity and sequence composition effects for triple helix formation by oligonucleotides containing P, using DNase footprinting. The specificity assay has proven particularly useful for initial characterization of a variety of novel bases. In Chapter Four, the synthesis and binding studies of the desmethyl derivative of P are reported. The results indicate that removal of the methyl group has a small detrimental effect on triple helix formation. Chapter Five describes an investigation of the specificity of the novel base 2-pyrimidinone. Chapter Six introduces the nonnatural base

*N*⁷-deoxyguanosine, which is shown to be a replacement for P as a base for pH independent GC recognition, since it shows slightly improved affinity and is synthetically more accessible. Chapter Seven describes attempts to develop a new structural motif for triple helix formation based on *N*⁷-substituted purine nucleosides.

Table of Contents

	page
Acknowledgments.....	iv
Abstract	v
Table of Contents.....	vii
List of Figures and Tables	viii
 CHAPTER ONE:	
Introduction.....	1
 CHAPTER TWO:	
NMR Structural Studies on a Nonnatural Deoxy- ribonucleoside which Mediates Recognition of GC Base Pairs in Pyrimidine•Purine•Pyrimidine DNA Triplexes	16
 CHAPTER THREE:	
Sequence Composition Effects on the Energetics of Triple Helix Formation by Oligonucleotides Containing a Designed Mimic of Protonated Cytosine	43
 CHAPTER FOUR:	
Triple Helix Formation by Oligonucleotides Containing Desmethyl P.....	75
 CHAPTER FIVE:	
Triple Helix Formation by an Oligonucleotide Containing the 2-Pyrimidinone Base	99
 CHAPTER SIX:	
Design of an <i>N</i> ⁷ -Glycosylated Purine Nucleoside for Recognition of GC Base Pairs by Triple Helix Formation	119
 CHAPTER SEVEN:	
Towards the Development of an <i>N</i> ⁷ -Purine Motif for Triple Helix Formation	130

List of Figures and Tables

CHAPTER ONE	page
Figure 1	The four Watson-Crick base pairs 3
Figure 2	Ribbon model of a pyrimidine motif triple helix 6
Figure 3	Models of T•AT and C+GC base triplets 7
Figure 4	Models of G•GC, A•AT, and T•AT base triplets 8
Figure 5	Protonated cytosine mimics..... 10
Figure 6	G•TA base triplet and the base D ₃ 11
CHAPTER TWO	
Figure 1	One-dimensional proton NMR spectrum of triplex 4 23
Figure 2	NOESY spectrum of triplex 4 25
Figure 3	Expanded contour plots of the NOESY spectrum of 4 28
Figure 4	One-dimensional proton NMR of a 4 analog in H ₂ O..... 32
Figure 5	NOESY plots correlating base and sugar H1' protons..... 34
Figure 6	Expanded plot of the NOESY spectrum in Figure 5..... 37
Table I	Chemical shifts of the protons in triplex 4 38
CHAPTER THREE	
Figure 1	MeC+GC and P•GC base triplets..... 45
Figure 2	Autoradiogram of footprinting gel for P specificity experiments 48
Figure 3	Plots of θ versus oligonucleotide concentration for 1-5 50
Figure 4	Sequence composition experiments..... 55
Figure 5	Autoradiogram of a footprinting gel for the (GA) ₅ site 58
Figure 6	Plots of θ versus oligonucleotide concentration for 6 and 7 60
Figure 7	Autoradiogram of a footprinting gel for the G ₆ site..... 62
Figure 8	Plots of θ versus oligonucleotide concentration for 8 and 9 64
Table I	Equilibrium association constants for 20 triple helical complexes..... 54
Table II	Equilibrium association constants at the (GA) ₅ site 65

Table III	Equilibrium association constants at the G ₆ site.....	65
-----------	---	----

CHAPTER FOUR

Figure 1	Models of the P•GC and ^D P•GC base triplets	77
Figure 2	Sequences of oligonucleotides and duplex target sites.....	78
Figure 3	Synthesis of ^D P phosphoramidite.....	81
Figure 4	HPLC analysis of oligonucleotides containing ^D P.....	82
Figure 5	Autoradiogram of footprinting gel for the specificity experiment.....	87
Figure 6	Plots of θ versus oligonucleotide concentration for oligos 1-3	89
Table I	Equilibrium association constants for ^D P, P, or MeC oligos	90
Table II	Equilibrium association constants at the (GA) ₅ site	90

CHAPTER FIVE

Figure 1	Structure of 2-pyrimidinone deoxyriboside.....	100
Figure 2	Models of the four 2-pyrimidinone base triplets.....	101
Figure 3	Overlay of 2-pyrimidinone•CG on a T•AT base triplet	102
Figure 4	2-Pyrimidinone specificity experiments	102
Figure 5	Synthesis of 2-pyrimidinone deoxyriboside phosphoramidite	103
Figure 6	Kinase gel of 2-pyrimidinone oligonucleotides	106
Figure 7	HPLC analysis of intact 2-pyrimidinone oligonucleotides.....	107
Figure 8	HPLC analysis of digested 2-pyrimidinone oligonucleotides.....	107
Figure 9	Autoradiogram of footprinting gel for the specificity experiment.....	110
Figure 10	Plots of θ versus oligonucleotide concentration for oligos 1-3	112
Table I	Equilibrium association constants for 2-pyrimidinone oligo.....	114

CHAPTER SIX

Figure 1	Models of ^m C+GC, G•GC, and ⁷ G•GC base triplets.....	121
Figure 2	Synthesis of d ⁷ G phosphoramidite.....	122
Figure 3	Specificity experiments with d ⁷ G.....	123
Figure 4	Sequence composition experiments with d ⁷ G.....	127

Table I	Equilibrium association constants for 20 triple helical complexes.....	123
Table II	Equilibrium association constants at the (GA) ₅ site	128
Table III	Equilibrium association constants at the G ₆ site	128

CHAPTER SEVEN

Figure 1	The <i>N</i> ⁷ -deoxyguanosine•GC base triplet	131
Figure 2	pKa's of methyl substituted xanthines.....	132
Figure 3	Models of <i>N</i> ⁷ -deoxyxanthosine base triplets	133
Figure 4	Models of <i>N</i> ⁷ -deoxyisoinosine base triplets	134
Figure 5	Models of <i>N</i> ⁷ -deoxyadenosine base triplets.....	135
Figure 6	Chem3D model of <i>N</i> ⁷ -deoxyadenosine.....	136
Figure 7	Synthesis of d ⁷ X phosphoramidite 9	138
Figure 8	Synthesis of d ⁷ A phosphoramidite 17	140
Figure 9	Synthesis of d ⁷ G phosphoramidite 26	142
Figure 10	Synthesis of d ⁷ X controlled pore glass 29	143
Figure 11	Synthesis of d ⁷ A controlled pore glass 31	144
Figure 12	Design of the specificity experiments for d ⁷ X, d ⁷ iI, d ⁷ A.....	145
Figure 13	Kinase/PAGE analysis of specificity oligonucleotides.....	147
Figure 14	Enzyme digestion/HPLC analysis of specificity oligonucleotides	148
Figure 15	DNase footprinting gel for the d ⁷ X oligonucleotide	150
Figure 16	DNase footprinting gel for the d ⁷ A oligonucleotide	152
Figure 17	Plots of θ versus oligonucleotide concentration for oligos 32 and 34	154
Figure 18	Sequences for the <i>N</i> ⁷ -purine motif experiments.....	158
Figure 19	Restriction map of the plasmid pSP6	159
Figure 20	Kinase/PAGE analysis of the <i>N</i> ⁷ -purine oligonucleotides.....	163
Figure 21	Enzyme digestion/HPLC analysis of d ⁷ X/d ⁷ G oligonucleotides.....	164
Figure 22	Enzyme digestion/HPLC analysis of d ⁷ A/d ⁷ G oligonucleotides.....	165
Figure 23	Enzyme digestion/HPLC analysis of T/d ⁷ G oligonucleotides	166
Figure 24	DNase footprinting gel for the T/ ^{Me} C oligonucleotides.....	168
Figure 25	DNase footprinting gel for the purine motif T/G oligonucleotides	170
Figure 26	DNase footprinting gel for the purine motif A/G oligonucleotides.....	172
Figure 27	DNase footprinting gel for the d ⁷ X/d ⁷ G oligonucleotides	174
Figure 28	DNase footprinting gel for the d ⁷ A/d ⁷ G oligonucleotides	176
Figure 29	DNase footprinting gel for the T/d ⁷ G oligonucleotides	178

Figure 30	DNase footprinting gel for the d ⁷ X/d ⁷ G oligonucleotides (low salt).....	180
Figure 31	DNase footprinting gel for the d ⁷ A/d ⁷ G oligonucleotides (low salt).....	182
Table I	Equilibrium association constants for the d ⁷ X, d ⁷ iI, and d ⁷ A oligos	156
Table II	Summary of binding results in the <i>N</i> ⁷ -purine motif experiments.....	162

CHAPTER ONE

Introduction

Deoxyribonucleic acid (DNA) plays a central role in life processes as the molecule responsible for storage of genetic information. The DNA sequence contained within each cell of an organism comprises all of the information required at each stage of development for every cell type of the organism. In addition, the DNA sequence contains information for responding to various environmental conditions and pathological states. Thus, only small portions of this vast pool of information are required at any specific time and location within an organism, necessitating tight control over the expression of genetic information. This control is provided by proteins such as transcription factors, which bind to specific DNA sequences and either activate or repress the expression of particular genes. In fact, molecular recognition of nucleic acid sequence and structure is involved in each step of the expression of DNA sequence information, including transcription, mRNA processing, and translation. It is therefore crucial to elucidate the basic principles of molecular recognition of nucleic acids in order to deeply understand how DNA sequence information is converted into the processes that we collectively know as life. In addition, an understanding of nucleic acid molecular recognition will enable the design of synthetic materials which can control the expression of genetic information. Since a wide variety of pathological states, including cancer, genetic disease, and viral infection, are traceable to incorrect or untimely gene expression, such materials may prove to be powerful therapeutic agents.

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 such polymer strands associate by Watson-Crick base pairing through formation of adenine-thymine (AT) or guanine-cytosine (GC) base pairs (1), generating a helical molecule with two antiparallel strands (Figure 1). The stability of the DNA double helix arises from both the hydrogen bonds between the components of each base pair and from vertical π stacking interactions between consecutive bases (1). Variations in environmental conditions such as metal ion concentration and relative humidity and in sequence give rise to several classes of DNA structure, which vary in helical pitch, rise per

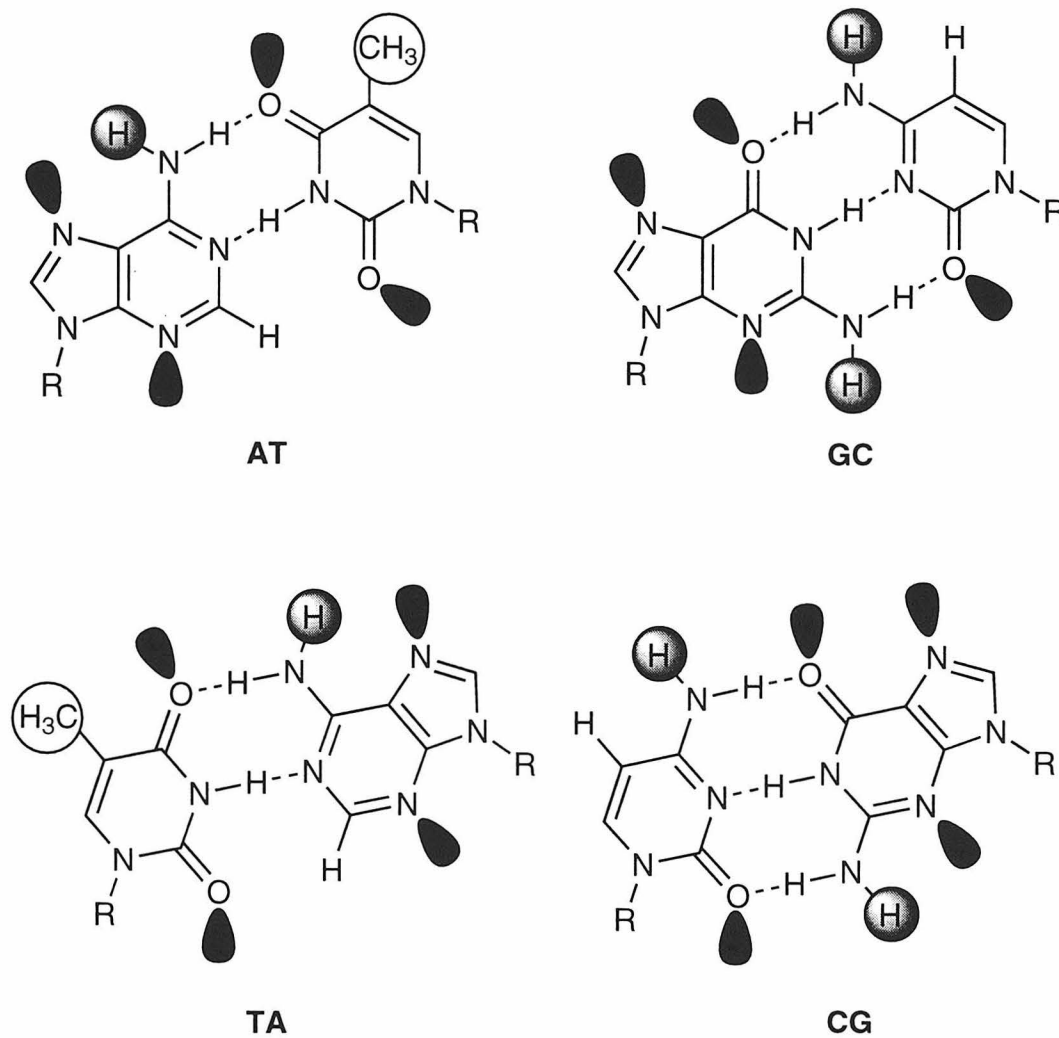


Figure 1: The four Watson Crick base pairs. The sugar phosphate backbone is indicated by R. The black orbital symbols indicate hydrogen bond acceptors. Hydrogen bond donors are indicated by gray circles. The hydrophobic methyl groups are circled.

residue, handedness, and groove width. The predominant DNA structural class, B-form, is characterized by a rise per residue of 3.4 Å and a turn angle per residue of 36 °. B-form DNA is characterized by a narrow minor groove (5.7 Å) and a wide major groove (11.7 Å) (1). As shown in Figure 1, an array of functional groups project into the major and minor grooves, providing the chemical basis for sequence specific recognition. In the major groove, the four base pairs are uniquely determined by the pattern of hydrogen bond acceptors, donors, and hydrophobic groups that each presents. In the minor groove, however, the AT and TA base pairs are essentially degenerate, since each presents two lone pairs oriented along the walls of the groove.

DNA Binding Molecules. A variety of molecules have been shown to bind to double stranded DNA with varying degrees of sequence specificity. The intercalators are a class of molecules which bind to DNA relatively nonspecifically by inserting between consecutive base pairs, thus extending and unwinding the helix. Several of these have proven to be clinically useful cancer therapeutics, including acridine derivatives such as Amsacrine and anthracyclines such as adriamycin and daunomycin (2). A second class of DNA binding small molecules are those typified by the natural products distamycin and netropsin which associate with the minor groove of AT tracts (3). Association is driven by van der Waals contacts along the sides of the groove and by hydrogen bonding with the base pairs along the minor groove floor. The application of iterative cycles of molecular design and chemical synthesis to these natural products has yielded molecules in which the natural AT specificity has been entirely reversed to GC specificity (4). A variety of proteins associate with DNA with high sequence specificity, including transcription factors, restriction enzymes, and various DNA repair enzymes. Although the design of proteins with arbitrary specificity is a formidable challenge, due to the difficulty of predicting tertiary fold from primary sequence, substantial headway has been made using combinatorial approaches based on the zinc finger class of DNA binding proteins (5). A final class of DNA binding molecules are oligonucleotides which bind in the major groove

through formation of triple helical DNA complexes.

Triple Helix Formation. Triple helix formation by nucleic acids was discovered soon after the elucidation of the structure of double helical DNA in 1953 (6). Rich and coworkers reported that strands of poly(U) and poly(rA) associate in a 2:1 complex in the presence of millimolar Mg^{+2} (7). These authors proposed that the association was driven by third strand uracil bases associating in Hoogsteen fashion with AU base pairs to form U•AU base triplets. Subsequent studies demonstrated that similar 2:1 complexes could be formed between strands of poly(rC) and oligo(rG) at acidic pH (8). In this case, protonated cytosine bases were proposed to associate in Hoogsteen fashion with GC base pairs, generating C+GC base triplets. The requirement of protonation was thought to account for the observed pH dependence.

In 1987, Moser and Dervan reported that 15-mer pyrimidine oligonucleotides could bind to homopurine sequences within DNA restriction fragments (9). They demonstrated that the oligonucleotide bound in the major groove, parallel to the purine strand of the duplex, and proposed the interaction was mediated by formation of T•AT and C+GC base triplets, in analogy to the earlier homopolymer work (Figures 2 and 3). A number of NMR structural studies from the laboratories of Patel and Feigon have verified the proposed base triplet structures and have provided insight into structural details of the triple helical complexes (10).

A second structural motif for triple helix formation was reported by Beal and Dervan in 1991, when they showed that oligonucleotides composed of either deoxyguanosine or deoxyadenosine or deoxyguanosine and thymidine bound to homopurine target sequences (11). In this case, the third strand oligonucleotide was demonstrated to bind in the major groove, antiparallel to the purine duplex strand. The interaction was proposed to be mediated by the formation of G•GC and A•AT or reverse Hoogsteen T•AT base triplets (Figure 3). NMR studies subsequently confirmed the proposed triplet structures (12).

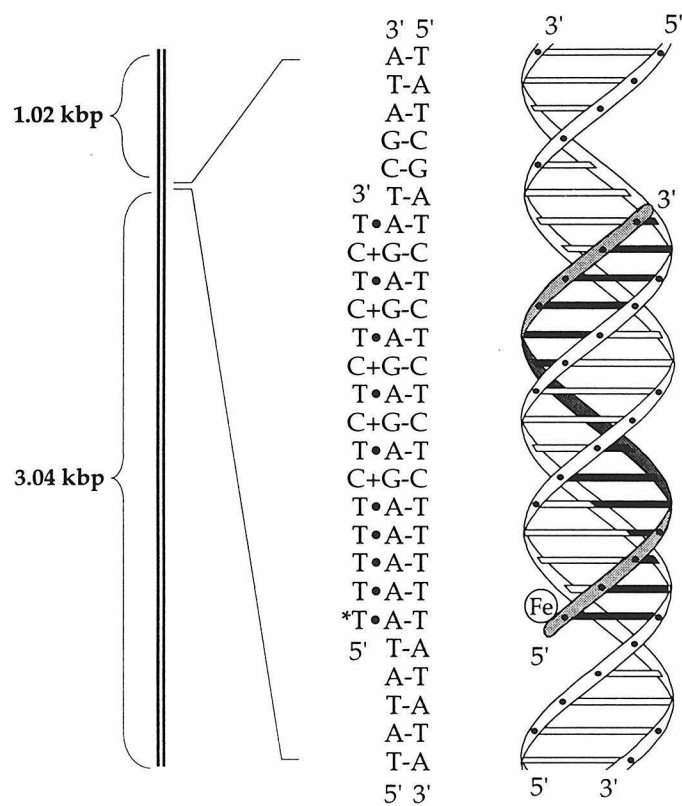


Figure 2: Ribbon model of a pyrimidine oligonucleotide bound to a 15 bp homopurine sequence within a 4 kbp restriction fragment. Note that the third strand binds parallel to the purine duplex strand.

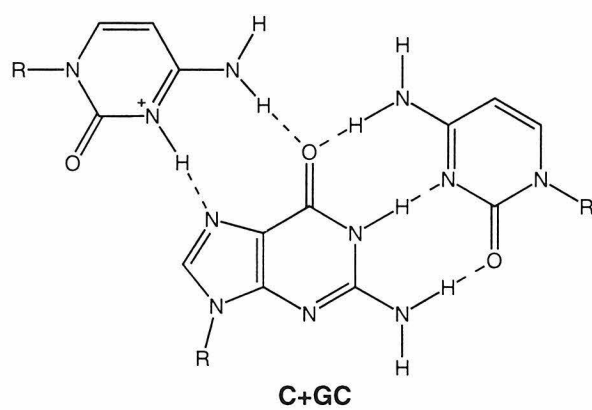
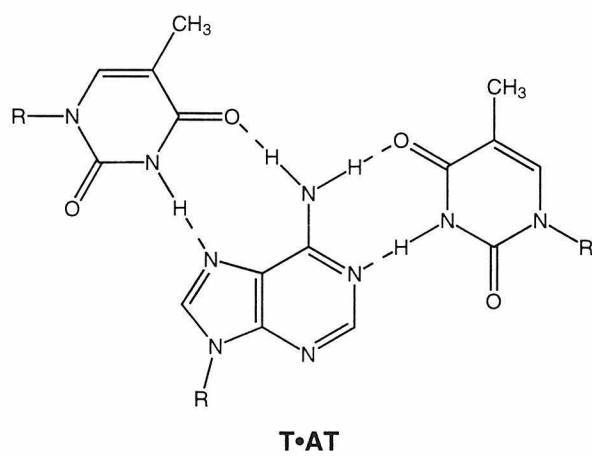


Figure 3: Two-dimensional models of T•AT and C+GC base triplets. The backbone of the third strand base is oriented parallel to the purine strand. Note that the third strand cytosine base requires protonation in order to form two hydrogen bonds to the GC base pair.

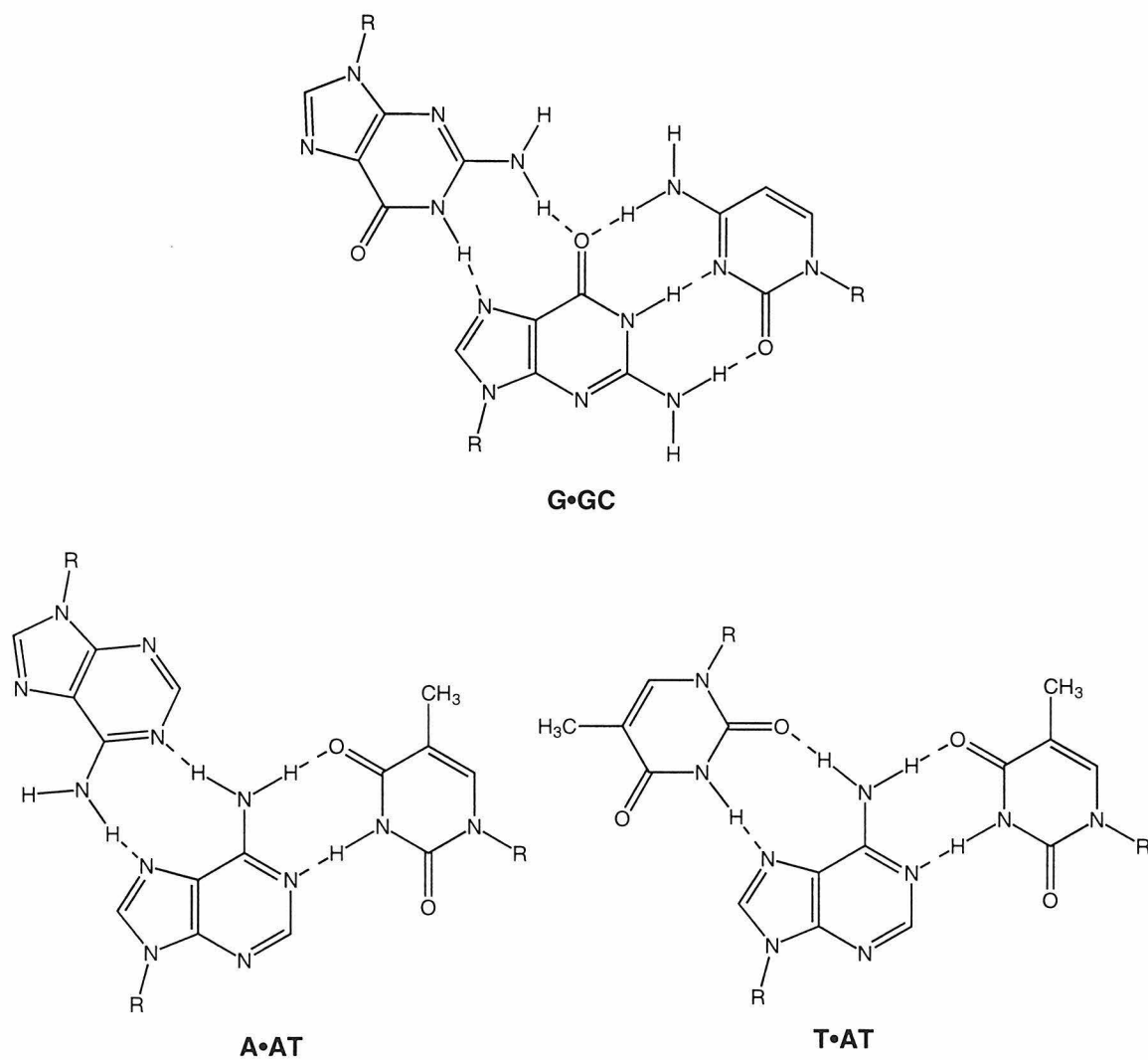


Figure 4: Two-dimensional models of purine motif G•GC, A•AT and reverse Hoogsteen T•AT base triplets. The third strand backbone is oriented antiparallel to the purine strand.

The demonstration that short oligonucleotides could bind with high specificity to 15 to 18 bp sites within relatively large DNA fragments generated profound interest in using such molecules to artificially regulate gene expression — the so-called “anti-gene approach (13).” In a series of experiments which clearly displayed the power of oligonucleotide -directed triple helix formation, single site cleavage was obtained in bacteriophage λ , yeast, and finally human chromosomal DNA (14). Local triple helices have also been shown *in vitro* to inhibit a variety of DNA binding proteins, including methylases and transcription factors (15). However, in spite of the success *in vitro*, there have been few reports of successful triple helix mediated function *in vivo* (16).

The kinetic and thermodynamic parameters of triple helix formation are well studied. It has been shown by several groups that the rate of triple helix formation is several orders of magnitude slower than duplex formation (17). The slow kinetics of triple helix formation are a critical issue with respect to possible *in vivo* applications. The energetics of triple helix formation have been studied by a variety of techniques, including calorimetry (18), UV melting (17b, 19), and affinity cleavage (20). Such studies have demonstrated that the stability of triple helical complexes is sensitive to length (20), single base mismatches (17b, 19, 20, 21), solution pH (22), cation valence and concentration (23), and temperature (24).

Nonnatural Bases. An intensive effort has been directed towards the development of synthetic bases which expand the generality of triple helix formation for DNA recognition. A series of neutral bases which mimic protonated cytosine and thereby allow pH independent recognition of GC base pairs have been reported. These include pseudoisocytosine (25), 8-oxoadenine (26), 3-methyl-5-amino-pyrazolo[4, 3-*d*]pyrimidin-7-one (P) (27), 4-amino-5-methyl-2,6-pyrimidione (28) and a pyrazine donor-donor-acceptor derivative (29) (Figure 5). All of these bases show pH independent GC recognition and have been compared with cytosine or 5-methylcytosine, but to date there have been no comparisons among these derivatives.

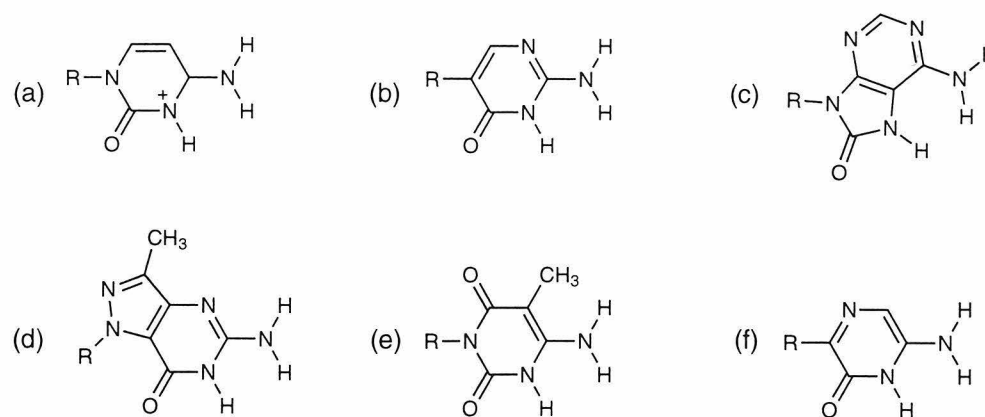


Figure 5: Protonated cytosine and mimics. (a) protonated cytosine; (b) pseudoisocytosine; (c) 8-oxoadenine; (d) 3-methyl-5-amino-pyrazolo[4, 3-*d*]pyrimidin-7-one (P); (e) 4-amino-5-methyl-2,6-pyrimidione; (f) pyrazine derivative.

Although the development of bases pH independent recognition of GC base pairs within the pyrimidine motif has been quite successful, the search for novel bases specific for CG and TA base pairs has been less fruitful. Oligonucleotides containing the natural base deoxyguanosine in the third strand have been shown to bind specifically to sequences containing TA base pairs at the corresponding position (30). Oligonucleotides containing this G•TA base triplet have been studied by NMR spectroscopy (10c, 31) and the triplet structure shown in Figure 6 has been deduced. The NMR structures show considerable distortion in the region of the G•TA triplet, and, consistent with this observation, thermodynamic studies have shown that triple helical complexes containing single G•TA triplets are substantially less stable than those comprised wholly of T•AT and C+GC triplets (21). Two bases, *N*⁴-(6-Aminopyridinyl)-cytosine (32) and *N*⁴-(3-Acetamidopropyl)-cytosine (33), have been reported to be specific for CG base pairs. However, a reinvestigation of the former has shown that the base retains its *N*-benzoyl protecting group under the specified conditions, and that it shows roughly equal affinity for CG and TA base pairs (34). The latter base appears specific for CG base pairs, but its affinity is substantially lower than the canonical C+GC and T•AT base triplets. Very recently, a novel

base which binds the CG base pair in chloroform has been reported (35), but its specificity and affinity in the context of oligonucleotide-directed triple helix formation in aqueous solution awaits further investigation.

The nonnatural base D_3 (Figure 6) has been shown to have specificity for pyrimidine-purine base pairs; that is, either CG or TA base pairs (36). A subsequent NMR study has shown that this specificity derives from an intercalative mode of binding for D_3 (37).

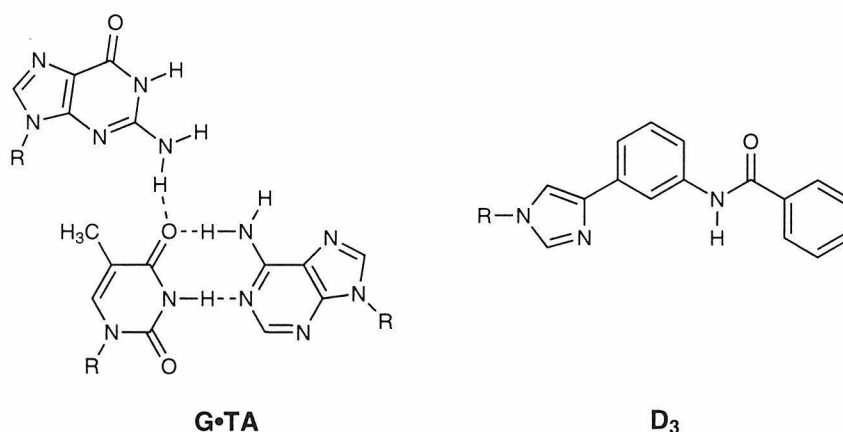


Figure 6: The G•TA base triplet and the intercalative base D_3 .

Within the purine motif, development of CG and TA specific novel bases has proven similarly challenging. The novel base nebularine has been shown to bind weakly to both AT and CG base pairs, probably by formation of a single hydrogen bond in each case (38). A series ofazole (39) and pyrimidine (40) analogs have been investigated, but none of these have shown sufficient affinity and specificity. The novel base 6-thioguanine has been shown to be effective for reducing the propensity of G-rich purine motif oligonucleotides to form aggregated structures (G-quartet formation) (41).

Presently, although intensive effort has been directed at the development of nonnatural bases for CG and TA recognition in both the pyrimidine motif and purine motif, there are no general solutions available. As mentioned above, secondary problems such as pH independent pyrimidine motif triple helix formation and aggregation of purine

motif oligonucleotides have been addressed successfully.

Thesis Work. This thesis describes the application of quantitative DNase I footprinting (42) to the measurement of the energetics of triple helix formation by oligonucleotides containing nonnatural bases. Chapter Two describes an NMR study of an intramolecular triplex containing the P•GC base triplet which was performed in collaboration with Professor Dinshaw Patel at Columbia University. Chapter Three describes the quantitation of the specificity and sequence composition effects for triple helix formation by oligonucleotides containing P, using DNase footprinting. The specificity assay has proven particularly useful for initial characterization of a variety of novel bases. In Chapter Four, the synthesis and binding studies of the desmethyl derivative of P are reported. The results indicate that removal of the methyl group has a small detrimental effect on triple helix formation. Chapter Five describes an investigation of the specificity of the novel base 2-pyrimidinone. Chapter Six introduces the nonnatural base N^7 -deoxyguanosine, which is shown to be a replacement for P as a base for pH independent GC recognition, since it shows slightly improved affinity and is synthetically more accessible. Chapter Seven describes attempts to develop a new structural motif for triple helix formation based on N^7 substituted purine nucleosides.

References

- (1) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.
- (2) Silverman, R. B. *The Organic Chemistry of Drug Design and Drug Action*, Academic: New York, 1992.
- (3) Dervan, P. B. *Science* **1986**, 232, 464-471.
- (4) Geierstranger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Science* **1994**, 266, 646.
- (5) (a) Rebar, E. J.; Pabo, C. *Science* **1994**, 263, 671-673. (b) Choo, Y.; Klug, A. *Proc. Natl. Acad. Sci., U.S.A.* **1994**, 91, 11163-11167. (c) Choo, Y.; Klug, A. *Proc. Natl. Acad. Sci., U.S.A.* **1994**, 91, 11168-11172.
- (6) Watson, J. D.; Crick, F. H. C. *Nature* **1953**, 171, 737-738.
- (7) Felsenfeld, G.; Davies, D. R.; Rich, A. *J. Am. Chem. Soc.* **1957**, 79, 2023-2024.
- (8) (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.
- (9) Moser, H. E.; Dervan, P. B. *Science* **1986**, 232, 464.
- (10) (a) de los Santos, C.; Rosen, M.; Patel, D. *Biochemistry* **1989**, 28, 7282-7289. (b) Sklenar, V.; Feigon, J. *Nature* **1990**, 345, 836-838. (c) Radhakrishnan, I.; Patel, D. J. *Structure* **1994**, 2, 17-32.
- (11) Beal, P. A.; Dervan, P. B. *Science* **1991**, 251, 1360-1363.
- (12) Radhakrishnan, I.; Patel, D. J. *Structure* **1993**, 1, 135-152, and references therein.
- (13) For a review, see: Thuong, N. T.; Helene, C. *Angew. Chem. Int. Ed. Engl.* **1993**, 32, 666-690.
- (14) (a) Strobel, S. A.; Moser, H. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1988**, 110, 7927. (b) Strobel, S. A.; Dervan, P. B. *Science* **1990**, 249, 73. (c) Strobel, S. A.; 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.

- (15) (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.
- (16) (a) Postel, E. H.; Flint, S. J.; Kessler, D. J.; Hogan, M. E. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 8227-8231. (b) Grigoriev, M., et al. *J. Biol. Chem.* **1992**, *267*, 3389-3395.
- (17) (a) Maher, L. J. III; Dervan, P. B.; Wold, B. J. *Biochemistry* **1990**, *29*, 8820-8826. (b) Rougee, M., et al. *Biochemistry* **1992**, *31*, 9269-9278.
- (18) Plum, E. G.; Park, Y.; Singleton, S. F.; Dervan, P. B.; Breslauer, K. J. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 9436-9440.
- (19) (a) Mergny, J.-L., et al. *Biochemistry* **1991**, *30*, 9791-9798. (b) Fossella, J. A.; Kim, Y. J.; Shih, H.; Richards, E. G.; Fresco, J. R. *Nucleic Acids Res.* **1993**, *21*, 4511-4515.
- (20) Singleton, S. F.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 6957-6965.
- (21) Best, George C.; Dervan, P. B. *J. Am. Chem. Soc.* **1995**, *117*, 1187.
- (22) Singleton, S. F.; Dervan, P. B. *Biochemistry* **1992**, *31*, 10995-11003.
- (23) Singleton, S. F.; Dervan, P. B. *Biochemistry* **1993**, *32*, 13171-13179.
- (24) Singleton, S. F.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 10376.
- (25) (a) Ono, A.; Ts'o, P. O. P.; Kan, L. *J. Am. Chem. Soc.* **1991**, *113*, 4032-4033. (b) Ono, A.; Ts'o, P. O. P.; Kan, L. *J. Org. Chem.* **1992**, *57*, 3225-3230.
- (26) (a) Young, S. L.; Krawczyk, S. H.; Matteucci, M. D.; Toole, J. J. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 10023-10026. (b) Krawczyk, S. H.; Milligan, J. F.; Wadwani, S.; Moulds, C.; Froehler, B. C.; Matteucci, M. D. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 3761-3764. (c) Miller, P. S.; Bahn, P.; Cushman, C. D.; Trapane, T. L. *Biochemistry* **1992**, *31*, 6788-6793. (d) Jetter, M. C.; Hobbs, F. W. *Biochemistry* **1993**, *32*, 3249-3254. (e) Davison, E. C.; Johnsson, K. *Nucleosides and Nucleotides* **1993**, *12*, 237-243.
- (27) (a) Koh, J. S.; Dervan, P. B.

- (28) Xiang, G.; Soussou, W.; McLaughlin, L. W. *J. Am. Chem. Soc.* **1994**, 116, 11155-11156.
- (29) von Krosigk, U.; Benner, S. A. *J. Am. Chem. Soc.* **1995**, 117, 5361-5362.
- (30) Griffin, L. C.; Dervan, P. B. *Science* **1989**, 245, 967-971.
- (31) Wang, E.; Malek, S.; Feigon, J. *Biochemistry* **1992**, 31, 4838-4846.
- (32) Huang, C.-Y.; Miller, P. S. *J. Am. Chem. Soc.* **1993**, 115, 10456-10457.
- (33) Huang, C.-Y.; Cushman, C. D.; Miller, P. S. *J. Org. Chem.* **1993**, 58, 5048-5049.
- (34) Greenberg, W.; Dervan, P. B. Unpublished data.
- (35) Zimmerman, S. C.; Schmitt, P. *J. Am. Chem. Soc.* **1995**, 117, 10769-10770.
- (36) Griffin, L. C.; Kiessling, L. L.; Beal, P. A.; Gillespie, P.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, 114, 7976-7982.
- (37) Koshlap, K. M.; Gillespie, P.; Dervan, P. B.; Feigon, J. *J. Am. Chem. Soc.* **1993**, 115, 7908-7909.
- (38) Stilz, H. U.; Dervan, P. B. *Biochemistry* **1993**, 32, 2177-2185.
- (39) Durland, R. H., et al. *Nucleic Acids Res.* **1995**, 23, 647-653.
- (40) Durland, R. H. et al. *Nucleic Acids Res.* **1994**, 22, 3233-3240.
- (41) Rao, T. S., et al. *Biochemistry* **1995**, 34, 765-772.
- (42) Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Methods Enzymol.* **1986**, 130, 132-181.

CHAPTER TWO

NMR Structural Studies on a Nonnatural Deoxyribonucleoside which Mediates Recognition of GC Base Pairs in Pyrimidine•Purine•Pyrimidine DNA Triplexes

*The text of this chapter is taken from a published paper that was coauthored with Prof.
Peter B. Dervan, Huw M. Nash, and our collaborators at Columbia University, Prof.*

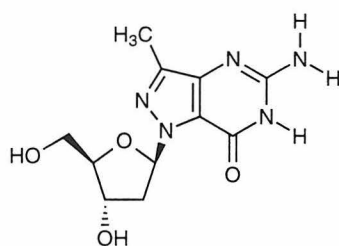
Dinshaw J. Patel and Ishwar Radhakrishnan.

(Radhakrishnan, I. et al. *Biochemistry* **1993**, 32, 11228-11234)

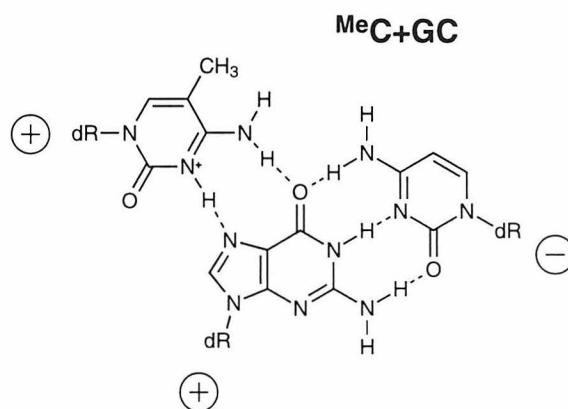
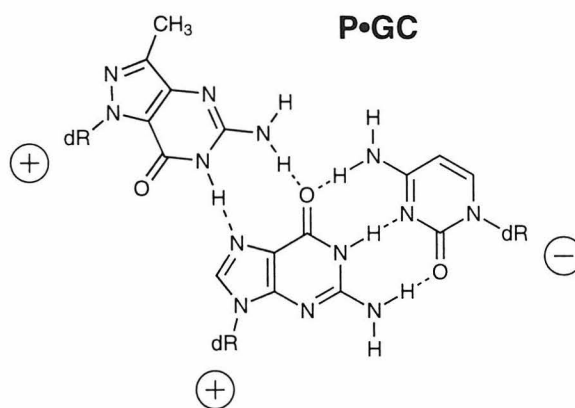
Introduction

Oligonucleotide-directed triple helix formation is a powerful method for sequence specific recognition of double helical DNA (Moser & Dervan, 1987; LeDoan et al., 1987; Cooney et al., 1988; Beal & Dervan, 1991). Although triple helix formation by nucleic acids (Felsenfeld et al., 1957) was discovered soon after the elucidation of the structure of double helical DNA (Watson & Crick, 1953), it has only recently been exploited as a generalizable approach to molecular recognition of double helical DNA. This approach has now been used to mediate the single site-specific cleavage of a human chromosome (Strobel et al., 1991) and to inhibit gene transcription *in vitro* (Cooney et al., 1988; Maher et al., 1992).

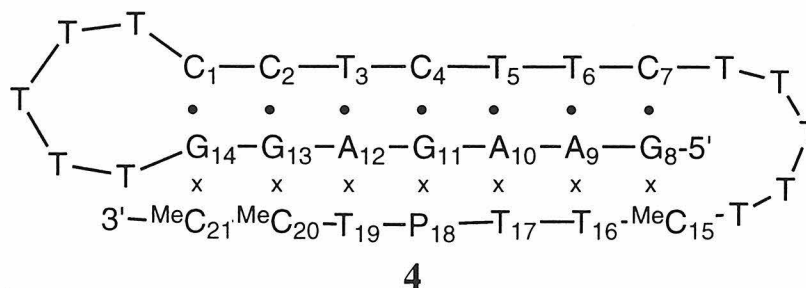
Pyrimidine oligonucleotides bind within the major groove of homopurine sequences of duplex DNA, parallel to the Watson-Crick (WC) purine strand, by forming specific Hoogsteen hydrogen bonds to the WC purine bases (Moser & Dervan, 1987; Praseuth et al., 1988; de los Santos et al., 1989; Rajagopal & Feigon, 1989). Specificity derives from thymine (T) recognition of adenine-thymine (AT) base pairs (T•AT base triples) and N³-protonated cytosine (C⁺) recognition of guanine-cytosine (GC) base pairs (C+GC base triples). The stability of triple helical nucleic acid complexes containing C+GC base triples decreases as the solution pH increases (Moser & Dervan, 1987; Maher & Dervan, 1990; Plum et al., 1990; Singleton & Dervan, 1992), presumably due to the requirement for protonation of cytosines in the third strand in order to form two Hoogsteen hydrogen bonds to each GC base pair. Replacement of third strand cytosines by 5-methylcytosine results in increased triple helix stability at any given pH, but triple helix stability remains pH dependent (Povsic & Dervan, 1989; Singleton & Dervan, 1992). Several nucleosides have been studied which are capable of recognizing GC base pairs without protonation (Ono et al., 1991; Krawczyk et al., 1992; Miller et al., 1992), including the nonnatural deoxyribonucleoside P (Koh & Dervan, 1992). The nucleoside P, **1**, which mimics the Hoogsteen base-pairing edge of protonated cytosines in C+GC

**1**

triples **2**, is capable of sequence-specific binding of target sites containing multiple GC base pairs over an extended pH range relative to both cytosine and 5-methylcytosine (Koh & Dervan, 1992). We report here an NMR study that was undertaken to verify the proposed base-pairing scheme for the **P•GC** base triple **3** (Koh & Dervan, 1992) and

**2****3**

examine the structural implications of introducing such a moiety in an intramolecular pyrimidine·purine·pyrimidine (Y·RY) triple helix **4**.



Materials and Methods

Oligodeoxyribonucleotide Preparation. Oligodeoxyribonucleotide **4** was synthesized on a 10 μ mol scale using solid-phase β -cyanoethyl phosphoramidite chemistry on an Applied Biosystems Model 380B DNA synthesizer. 2'-Deoxy-5-methylcytidine-derivatized controlled pore glass was prepared according to the standard method (Gait, 1984) and used as the solid support in this synthesis. The 5'-dimethoxytrityl- β -cyanoethylphosphoramidite of P was prepared as described (Koh & Dervan, 1992) and coupled as a 0.15 M solution in CH_3CN with >99% efficiency. The oligonucleotide was deprotected by treatment of the controlled pore glass support with saturated aqueous NH_4OH for 42 h at 55 $^\circ\text{C}$. The crude 5'-dimethoxytrityl protected products were purified by FPLC on a ProRPC 16/10 reverse phase column (Pharmacia LKB) with a linear gradient of 0-40% CH_3CN in 100 mM triethylammonium acetate, pH 7.0, detritylated in 80% acetic acid, and chromatographed a second time under the same conditions. Fractions containing pure oligonucleotide were combined and lyophilized to dryness, yielding 0.88 μ mol of **4** as determined by UV absorbance ($\epsilon_{260} = 283,000 \text{ M}^{-1}\text{cm}^{-1}$). Analysis of the purified oligonucleotide by reverse phase HPLC on a C-18

Aquapore OD-300 column (220 x 4.6 mm, Applied Biosystems) using a linear gradient of 0-20% CH₃CN in 100 mM ammonium acetate, pH 4.7 at 54 °C revealed a single UV absorbing peak. The nucleoside composition of **4** was confirmed by digestion of a 5 nmol aliquot of the oligonucleotide with 0.01 units of snake venom phosphodiesterase (Sigma) and 4 units of calf alkaline phosphatase (Boehringer Mannheim) in 30 µl of 30 mM Tris acetate, 70 mM MgCl₂, pH 8.1 at 37 °C for 2 hours. The digest products were separated by reverse phase HPLC on a C-18 Aquapore OD-300 column (220 x 4.6 mm, Applied Biosystems) using a linear gradient of 0-2.5% CH₃CN in 200 mM triethylammonium acetate, pH 6.0. Base ratios calculated from the chromatogram peak areas were 16:4.2:4.0:2.6:2.9:0.91 T:dC:dG:dA:^mdC:P, using the following extinction coefficients at 260 nm: T, 8800; dC, 7300; dG, 11700; dA, 15400; ^mdC, 5700; P, 2500.

NMR Sample Preparation. Prior to the NMR studies, the oligonucleotide **4** was desalted and converted to the Na⁺ form by sequentially passing through Sephadex G-25 and Dowex ion-exchange columns.

NMR Experiments. Proton NMR spectra were recorded on a Bruker AMX500 spectrometer on approximately 260 A₂₆₀ units of the oligonucleotide **4** in the presence of 0.1 mM EDTA and 10 mM phosphate buffer at 5 °C, pH 6.01 in H₂O and at 15 °C, pH 6.02 (uncorrected for isotope effect) in D₂O. The raw data were transferred to a Silicon Graphics Indigo workstation and processed with the Felix software package (Hare Research Inc.). The data acquisition and processing parameters are given in the figure legends.

Results

The P base was incorporated in the third strand segment of a model intramolecular triplex system that has been the focus of several studies in our laboratory (Radhakrishnan et al., 1991; Lin & Patel, 1992; Wang & Patel, unpublished results). The sequence and numbering system for the oligonucleotide used in this study is shown in **4**.

The P•GC triple is located at the central position of the triplex, which differs from the previously studied sequences in that cytidines in the third strand segment were replaced by 5-methylcytidines in order to facilitate studies at higher pH.

Exchangeable proton spectra. The exchangeable proton spectrum (6.5 to 15.5 ppm) of the intramolecular triplex **4** at 5 °C, pH 6.01 in H₂O buffer (Figure 1A), exhibits narrow, well-resolved resonances in the hydrogen bonded imino proton region (12.5 to 14.5 ppm, Figure 1B). The resonances between 8.5 and 10.2 ppm are diagnostic of a pyrimidine·purine·pyrimidine triplex species (de los Santos et al., 1989; Rajagopal & Feigon, 1989); these resonances correspond to the amino protons of the N³-protonated 5-methylcytosines. A new resonance at 11.63 ppm can be detected under these conditions which presumably belongs to the imino proton of P (see below). The minor resonances that can be observed in the hydrogen bonded imino proton region (Figure 1B) belong to the hairpin duplex with a dangling third strand that is in slow equilibrium with the predominant species — the intramolecular triplex that is the focus of the current study.

Imino proton assignments. The symmetrical imino proton region in the NOESY spectrum (mixing time 150 ms) of triplex **4** is shown in Figure 2. Sequential NOE connectivities between imino protons on adjacent base pairs can be traced in the duplex portion of the molecule (Figure 2A). The tracing starts from the G8 imino proton and proceeds via T6, T5, G11, T3 and G13 imino protons to the G14 imino proton. Sequential NOEs between imino protons on adjacent bases in the third strand can also be observed starting from T16 and proceeding via T17 and P imino protons to the T19 imino proton. The latter NOEs involving the P imino proton with its flanking thymine imino protons establishes a stacked-in conformation for the P base within the third strand of the helix. The P imino proton exchanges more rapidly with the solvent than the other interior hydrogen bonded imino proton resonances thus accounting for the weak intensity of these cross peaks.

Figure 1: The one-dimensional proton NMR spectrum of triplex **4** recorded in H₂O buffer at pH 6.01, 5 °C. (A) Only the region downfield to the H₂O resonance (6.5 to 15.5 ppm) is shown. The "jump-and-return" sequence (Plateau & Gueron, 1982) was used to suppress the H₂O signal with the delay between the two pulses set to 56 μ s for maximum excitation of the hydrogen bonded imino proton resonances. (B) An expanded plot of the hydrogen bonded imino proton region (11.2 to 15.5 ppm). The assignments for the various imino proton resonances are labeled.

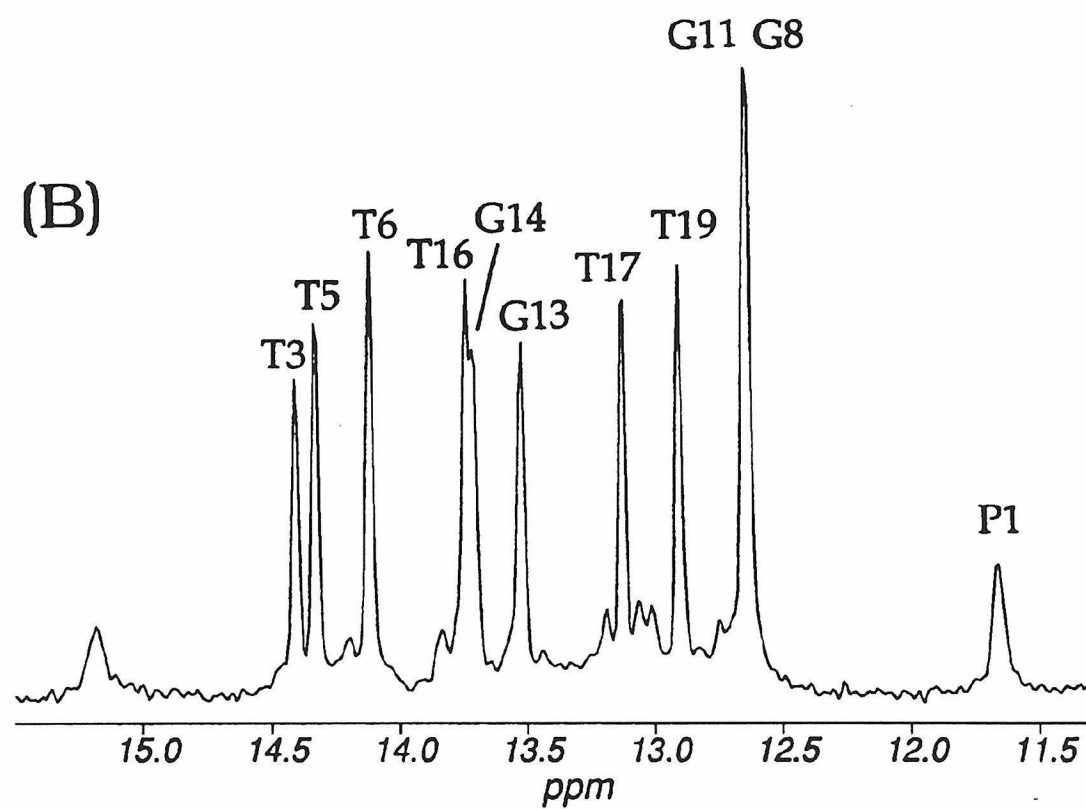
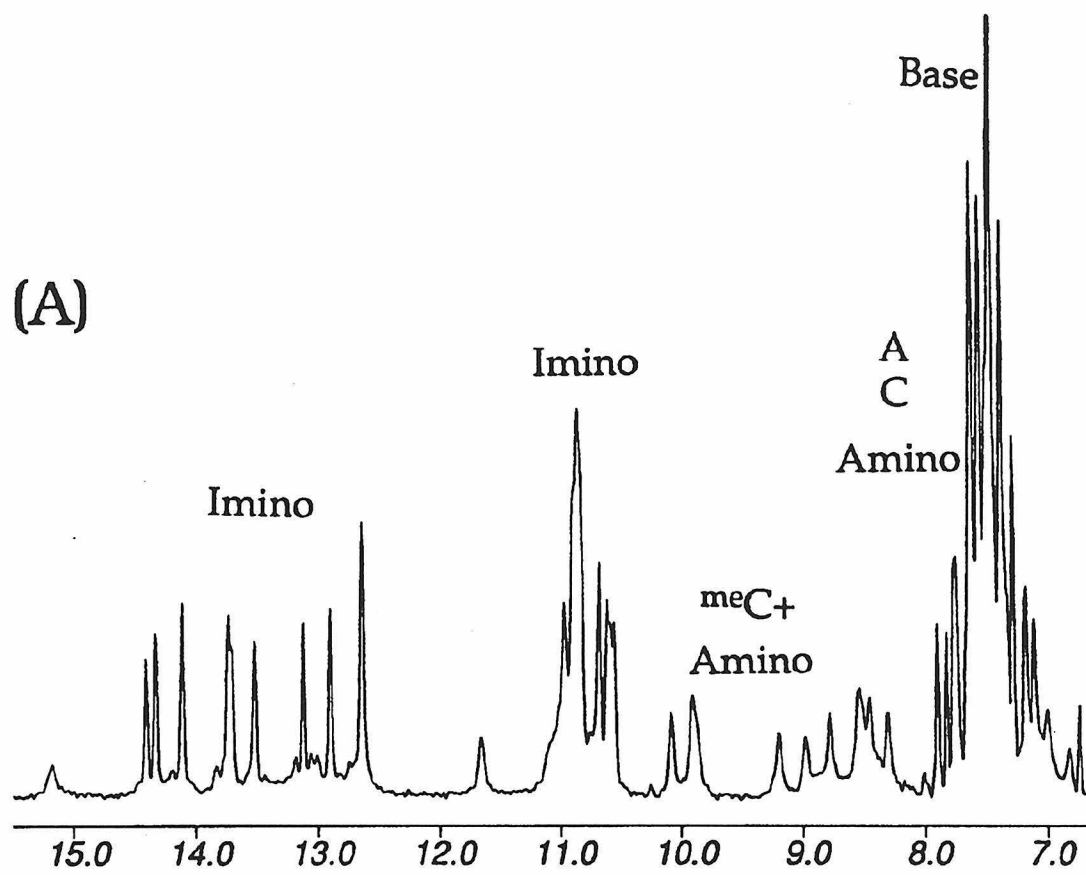
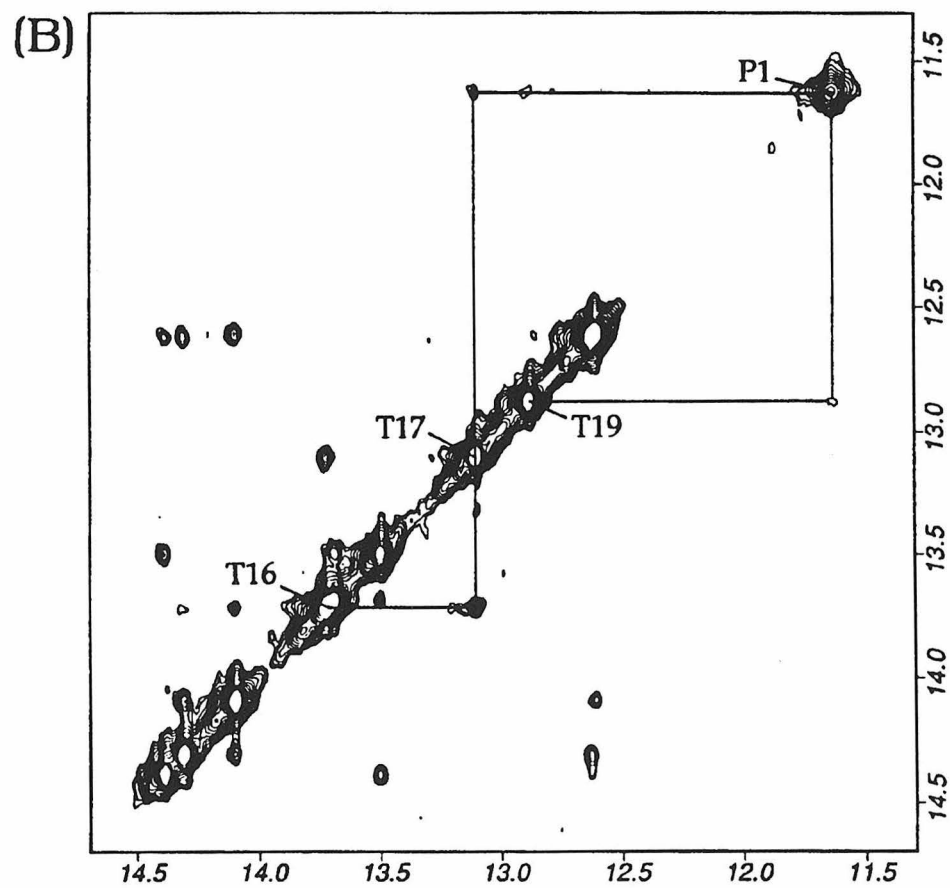
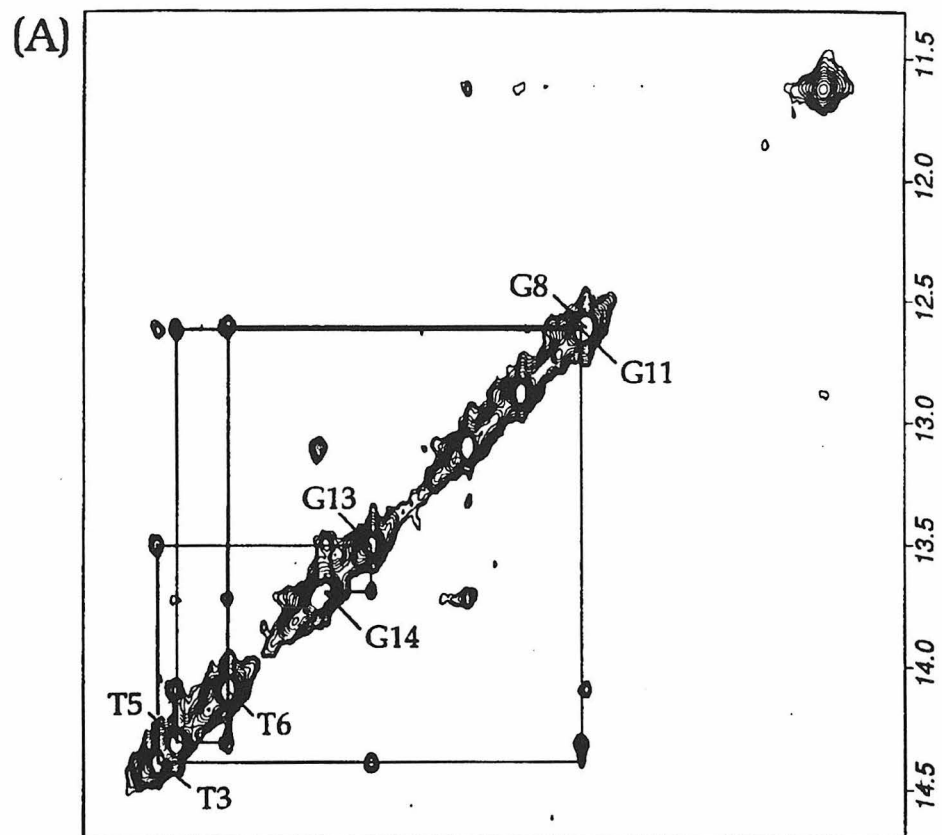


Figure 2: Duplicate, expanded contour plots of the symmetrical imino proton region (11.3 to 14.7 ppm) in the NOESY (mixing time 150 ms) spectrum of triplex **4** recorded in H₂O buffer, at pH 6.01, 5 °C. Hard 70° preparation and mixing pulses were used with the detection pulse replaced by a "jump-and-return" sequence (Plateau & Gueron, 1982). The carrier frequency was placed on the solvent signal and the spectral width was set to 11.1 kHz in both dimensions. The data were acquired with 1024 complex data points in t_2 and 400 complex data points in t_1 in the TPPI-States' mode (Marion, et al., 1989b). Prior to fourier transformation in t_2 , the residual solvent signal in each increment was deconvoluted in the time domain (Marion et al., 1989a), apodized with a skewed 30° phase shifted sine bell window function and zero-filled to 2048 points. Prior to fourier transformation in t_1 , baseline correction was applied in ω_2 and the interferograms apodized with a 90° phase shifted sine bell squared function and zero filled to yield a final processed matrix size of 2048 × 2048 in real points. (A) The lines trace the sequential connectivities between imino protons in adjacent base pairs for the duplex portion of triplex **4**. The tracing starts at G8 and proceeds via T6, T5, G11, T3, and G13 to G14. (B) The tracing of sequential connectivities between adjacent bases in the third strand of triplex **4** starts at T16 and proceeds via T17 and P to T19. The N³-protonated imino proton resonances of 5-methylcytosines in this strand exchange too rapidly with the solvent to show detectable NOEs.

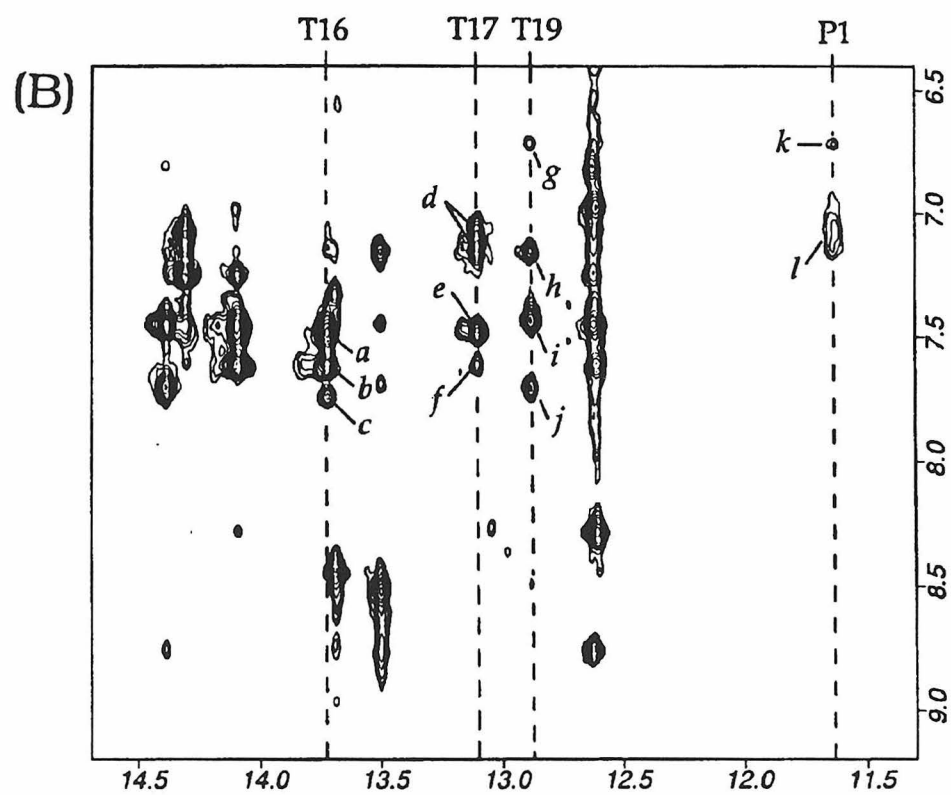
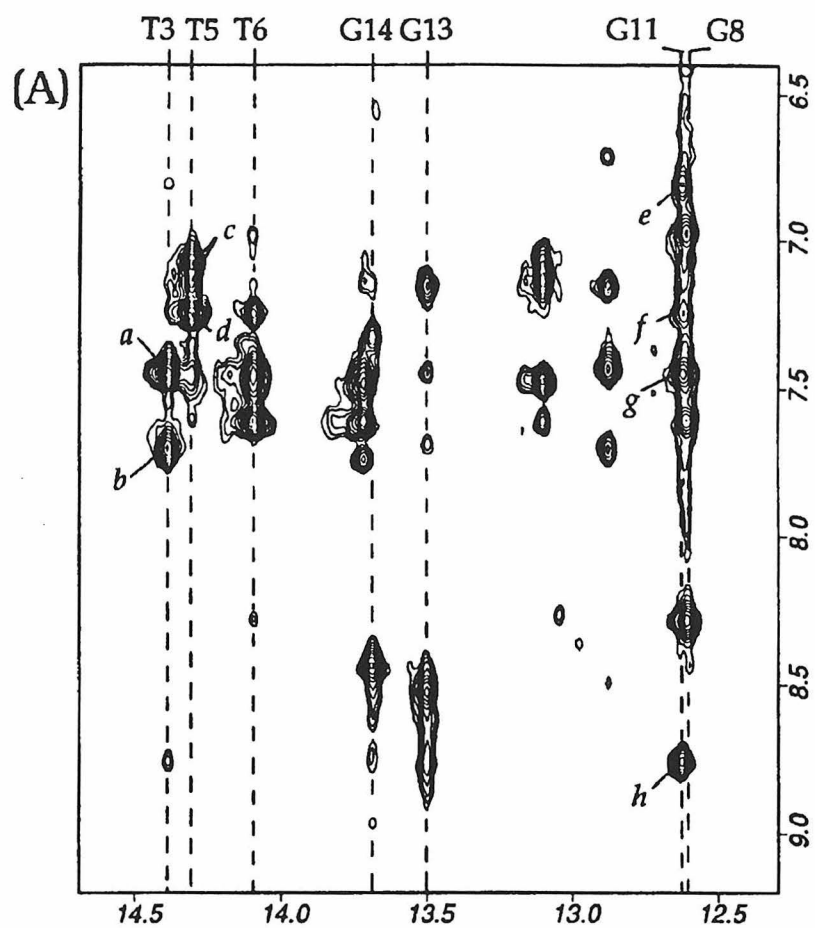


AT and GC Watson-Crick interactions. The bases in the purine and the first pyrimidine strand segments participate in Watson-Crick type hydrogen bonding interactions. This could be verified by the chemical shifts of the imino and amino protons as well as by the presence of NOEs that characterize these alignments. In particular, NOEs could be detected correlating the imino protons of T3, T5 and T6 with the H2 and hydrogen bonded amino protons of A12, A10 and A9, respectively in AT base pairs, while the imino protons of G8, G11, G13 and G14 show NOEs to the hydrogen bonded and exposed amino protons of C7, C4, C2 and C1, respectively, in GC base pairs. The NOE cross peak assignments for the central trinucleotide segment of the duplex portion of the molecule are shown in Figure 3A.

T•AT and ^{me}C+GC Hoogsteen interactions. In addition to the Watson-Crick type interactions, the purine bases in the second strand segment participate in Hoogsteen interactions with the third pyrimidine strand segment. NOEs that establish these alignments in T•AT triples could be detected between the imino protons of T16, T17 and T19 and the H8 and amino protons of A9, A10 and A12, respectively (peaks *a* and *b*, *d* and *e*, *i* and *j*, Figure 3B). Unlike these thymine imino protons, the protonated 5-methylcytosine imino protons are too broad to show analogous NOE patterns that would characterize the Hoogsteen alignments in ^{me}C+GC triples. However, the chemical shifts of protonated 5-methylcytosine amino protons (8.5 to 10.2 ppm) taken along with the NOEs observed between these protons and the amino protons of cytosines in the Watson-Crick duplex within the same triple (data not shown) implicate ^{me}C15, ^{me}C20 and ^{me}C21 in Hoogsteen interactions with G8, G13 and G14 bases, respectively, in the purine strand.

Pairing alignment in the P•GC triple. For comparison, the proposed pairing alignment for the P•GC triple **3** (Koh & Dervan, 1992) is shown along with the pairing alignment for a C+GC triple **2**. In both cases, the imino proton is involved in hydrogen bonding with the N⁷ atom of guanine while one of the amino protons is hydrogen bonded to the O⁶ atom of the same guanine. The NOEs shown by these imino and amino protons

Figure 3: Expanded contour plots of the same NOESY spectrum of triplex **4** shown in Figure 2. (A) NOE cross peaks correlating the imino protons (12.3 to 14.7 ppm) and the amino and base protons (6.4 to 9.2 ppm) in the Watson-Crick pairs in the triplex. The cross peak assignments for the protons in the central trinucleotide segment are as follows: *a*: T3(NH3)-A12(H2)/A12(NH₂-h); *b*: T3(NH3)-A12(NH₂-wc); *c*: T5(NH3)-A10(NH₂-wc)/A10(NH₂-h); *d*: T5(NH3)-A10(H2); *e*: G11(NH1)-C4(NH₂-e); *f*: G11(NH1)-A10(H2); *g*: G11(NH1)-A12(H2); *h*: G11(NH1)-C4(NH₂-b). (B) NOE cross peaks between the imino protons (11.3 to 14.7 ppm) and the amino and base protons (6.4 to 9.2 ppm) of the Hoogsteen pairs in the triples. The cross peak assignments are as follows: *a*: T16(NH3)-A9(NH₂-wc)/A9(NH₂-h); *b*: T16(NH3)-A9(H8); *c*: T16(NH3)-G8(H8); *d*: T17(NH3)-A10(NH₂-wc)/A10(NH₂-h); *e*: T17(NH3)-A10(H8); *f*: T17(NH3)-A9(H8); *g*: T19(NH3)-G11(H8); *h*: T19(NH3)-A12(H8); *i*: T19(NH3)-A12(NH₂-h); *j*: T19(NH3)-A12(NH₂-wc); *k*: P(NH1)-G11(H8); *l*: P(NH1)-(NH₂). In adenines, amino protons involved in Watson-Crick and Hoogsteen pairing are denoted by 'wc' and 'h' respectively while in cytosines, the hydrogen bonded and exposed amino protons are denoted by 'b' and 'e' respectively.



together with their chemical shifts are critical in defining the pairing alignment in the P•GC triple. Two NOEs involving the imino proton of P are clearly discernible in the imino to amino and base proton region of the NOESY spectrum of triplex **4** (Figure 3B). The imino proton of P shows an NOE to the H8 proton of G11 (peak *k*, Figure 3B) within the same triple. The other NOE (peak *l*, Figure 3B) which is comparatively stronger is between the imino and amino protons of P. It appears that, in the timescale of chemical shift differences, rotation around the C—N bond of the amino group of P is in the intermediate regime giving rise to a single broad resonance for these protons. The P imino proton also shows much weaker NOEs to the amino protons of neighboring A12 (data not shown). Most importantly, however, is the *absence* of any detectable NOE between the imino proton of P and the amino protons of C4 in the same triple. Since the P imino proton exchanges rapidly with the solvent, special care was taken to arrive at this conclusion. In the $\omega_2 = 11.63$ ppm slice corresponding to the P imino proton resonance in the NOESY spectrum, we do not detect any peaks above the noise level at the resonance positions ($\omega_1 = 8.75, 7.98$ ppm) corresponding to the C4 amino protons. The noise in this slice is sufficiently uniform to arrive at the aforementioned conclusion. If the imino proton were hydrogen bonded to the O⁶ carbonyl oxygen of G11, an NOE would be expected between this proton and the amino protons of C4. These observations place the imino proton within hydrogen bonding distance of the N⁷ atom of G11 and provide definitive evidence for the alignment shown in **3**. Corroborating evidence in the form of NOE patterns shown by the amino proton of P could not be obtained because of resonance overlap with the amino protons of A10. Although the resonances at these chemical shift positions showed NOEs to the C4 amino protons, the NOEs could not be unambiguously assigned to the interaction between the P amino and C4 amino protons since the A10 amino protons located on the adjacent triple could also participate in such interactions with the C4 amino protons.

The chemical shift of the P imino proton resonance shows a dependence on the

pH of the solution. A systematic pH study on an oligonucleotide sequence related to **4**, where 5-methylcytosines were replaced with cytosines, revealed that the chemical shift of this resonance moves downfield when the pH of the solution is lowered from 6.2 to 5 (Figure 4). By contrast, the chemical shifts of the methyl protons in the same ring system as well as other imino proton resonances in the triplex undergo only subtle changes in the same pH interval. Although an analogous study was not undertaken for triplex **4**, a similar behavior for the P imino proton resonance in this sequence was noticed when the pH was lowered. We suspect that the downfield shift associated with this resonance in a P•GC triple is due either to protonation of the P ring and/or formation of a stronger hydrogen bond with the N⁷ atom of guanine.

Structural perturbations introduced by the P base in the third strand. The third strand segment in triplex **4** consists exclusively of pyrimidines with the exception of the P base at the central position. To assess possible structural perturbations induced by the introduction of this P ring system, studies were extended to D₂O. Figure 4A traces the sequential connectivities for the purine strand of triplex **4** in the base to sugar H1' region of a NOESY spectrum (mixing time 250 ms) recorded in D₂O buffer at pH 6.02, 15 °C. Despite severe overlap, some of the cross peaks are well-resolved, particularly those in the central trinucleotide segment of the molecule. Continuity of sequential connectivities is maintained in the A10-G11-A12 segment not only in this region but in the base to H2', H2'' region as well. In the third strand of the molecule, we can trace sequential connectivities starting from ^{me}C15 and proceeding via T16 until the intra-residue base to H1' cross peak of T17 (Figure 5B). The tracing is interrupted in this region because of the absence of a suitable base proton on P that is sufficiently close to show cross peaks with any of its sugar protons. However, the tracing is resumed with the T19(H6)-P(H1') cross peak and proceeds to the 3'-end of the strand via ^{me}C20 to ^{me}C21 (Figure 5B). In the symmetrical methyl and sugar H2', H2'' proton region, however, we can trace sequential connectivities between methyl protons in adjacent bases in the third strand

Figure 4: One dimensional proton NMR spectra of a triplex **4** analog where cytosines have replaced 5-methylcytosines, recorded at 1 °C in H₂O in the presence of 0.1 mM EDTA and 10 mM phosphate buffer at different pH values. Only the region downfield from the water resonance (6.5–16.0 ppm) has been plotted. The position of the P imino proton resonance in each spectrum is indicated by an asterisk.

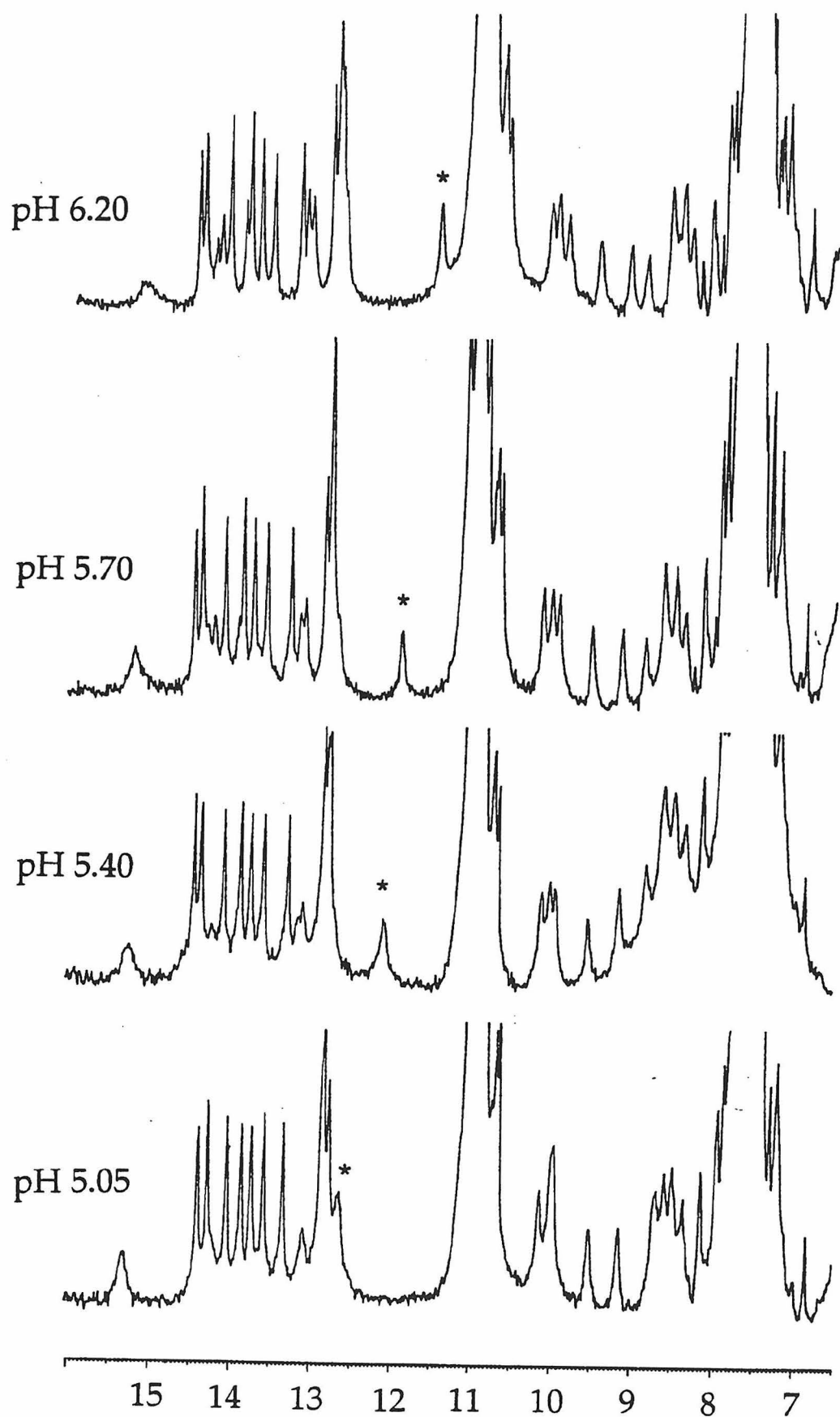
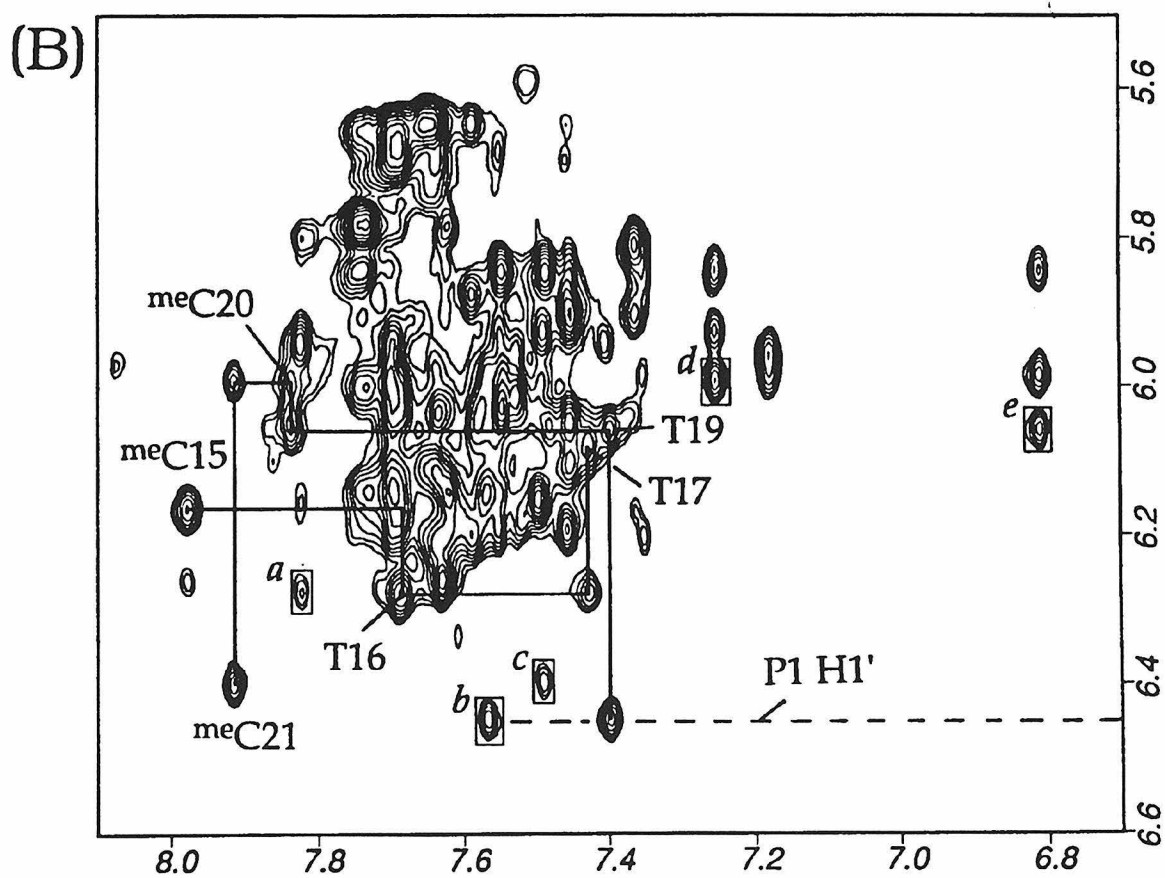
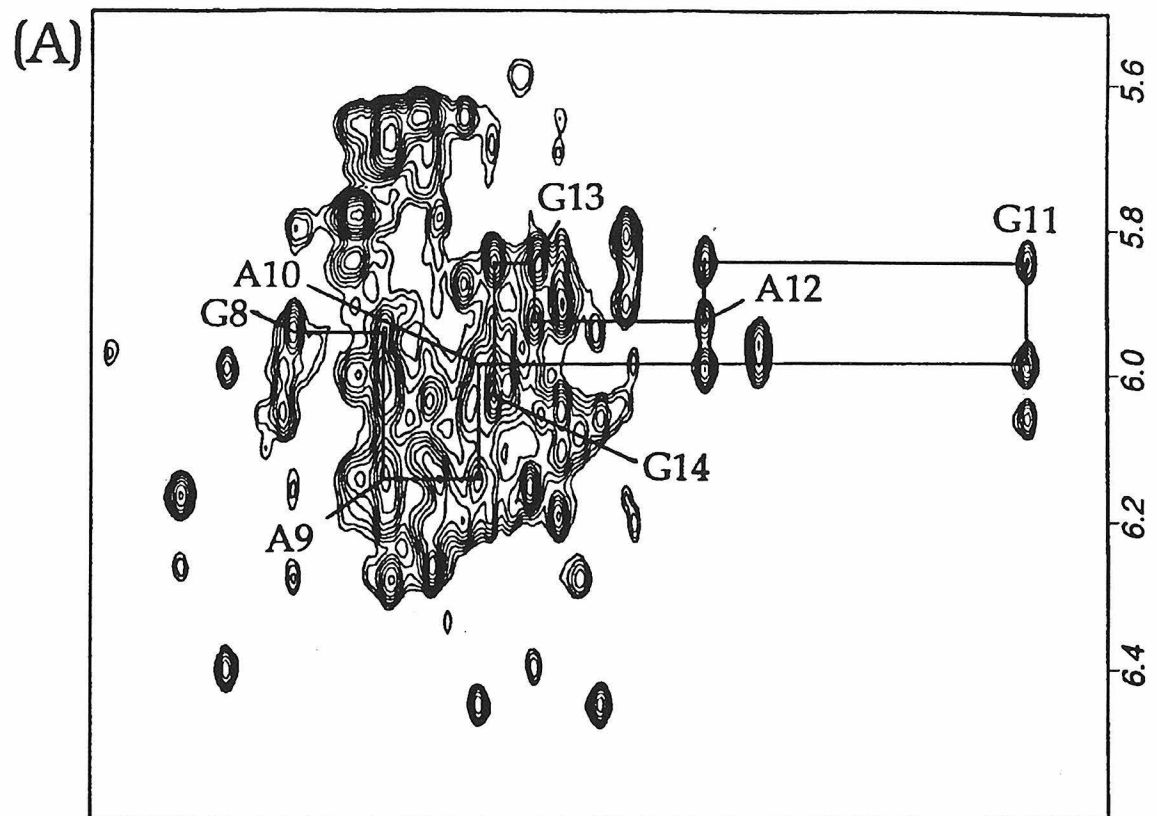


Figure 5: Duplicate, expanded contour plots correlating the base (6.7 to 8.1 ppm) and sugar H1' protons (5.5 to 6.6 ppm) in the NOESY spectrum (mixing time 250 ms) of triplex **4** recorded in D₂O buffer, at pH 6.02, 15 °C. The data were acquired with 1024 complex data points in t_2 , 512 real data points in t_1 in the TPPI mode (Marion & Wüthrich, 1983) with the spectral width set to 5.05 kHz in both dimensions and a repetition delay of 1.6s. 90° phase shifted, sine bell squared and sine bell window functions were used for apodization in t_2 and t_1 dimensions, respectively. The data were zero-filled to 2048 points in both dimensions prior to fourier transformation. Baseline correction was applied to the frequency domain data in both dimensions. (A) The lines trace sequential NOE connectivities between the base and sugar H1' protons in the purine strand of triplex **4**. The tracing starts at G8 and traverses the strand up to G14. The intra-residue base to sugar H1' cross peaks in this strand are labeled by residue number. (B) The solid lines trace sequential NOE connectivities between the base and sugar H1' protons in the third strand of triplex **4** starting from ^{me}C15. The tracing continues up to the intra-residue base to sugar H1' cross peak of T17 where it is interrupted due to the absence of a base proton on P but resumes at the T19(H6)-P(H1') cross peak and continues to ^{me}C21 at the end of the strand. The intra-residue cross peaks involving the base and sugar H1' protons in this strand are labeled by residue number. The dashed line corresponds to the position of the P H1' resonance. The cross-strand NOEs between the base protons in the purine strand and the H1' protons in the third strand are boxed and these are assigned as follows: *a*: G8(H8)-T16(H1'); *b*: A10(H8)-P(H1'); *c*: G13(H8)-^{me}C21(H1'); *d*: A12(H8)-^{me}C20(H1'); *e*: G11(H8)-T19(H1').



starting from T17 and continuing via P and T19 to ^{me}C20 (Figure 6). This is consistent with the stacked-in conformation for P deduced from the H₂O data. An NOE is also detected between the P methyl proton and the H2'' proton of T17 (boxed peak, Figure 6). Other sequential contacts involving P including T17(H6)-P(CH₃), T19(H6)-P(H2'), T19(H6)-P(H2'') are also observed and suggest no obvious structural perturbation at this site.

The observation of cross-strand NOEs between residues in adjacent triples provides an indication of the size of the groove formed by the purine and pyrimidine third strands (Radhakrishnan et al., 1992). Three classes of NOEs are useful in this regard. The imino protons of thymines show NOEs not only to the H8 protons of the bases with which they hydrogen bond, but also to the H8 protons of the bases on the 5'- side of their Hoogsteen partners. A similar trend is observed for the triplex **4** where the T16, T17 and T19 imino protons show such NOEs to the H8 protons of G8, A9 and G11 bases in the purine strand (peaks *c*, *f* and *g*, Figure 3B). NOEs are also observed between these thymine imino protons and the sugar H2', H2'' protons in the same pairs of residues (data not shown). The third class of NOEs that are observed involve the base protons in the purine strand and the H1' protons of the bases in the third strand. In triplex **4**, we detect these types of interactions clearly, and they are shown as boxed peaks in Figure 5B. Around the P•GC triple, we detect NOEs between the H1' protons of P and T19 and the H8 protons of A10 and G11 respectively (peaks *b* and *e*, Figure 5B). The intensities of the cross peaks arising from these type of interactions were compared in a NOESY spectrum acquired with a short mixing time (50 ms). The intensity of the cross peaks *b*, *d* and *e* were somewhat stronger than those of cross peaks *a* and *c* shown in Figure 5B. These observations suggest that the groove width is not significantly altered by the introduction of P in the pyrimidine third strand.

Chemical Shifts. Several unusual shifts are observed in triplex **4** especially in and around the P•GC triple (Table I). These include the rather upfield shifts for G11 H8

Figure 6: An expanded contour plot of the symmetrical methyl and H2', 2'' proton region in the same NOESY spectrum of triplex **4** shown in Figure 5. The lines trace the sequential connectivities between methyl protons on adjacent bases in the third strand of the molecule starting from T17 and proceeding via P and T19 to ^mC20. The boxed peak (indicated by an arrow) corresponds to the NOE observed between the sugar H2'' proton of T17 and the methyl protons of P.

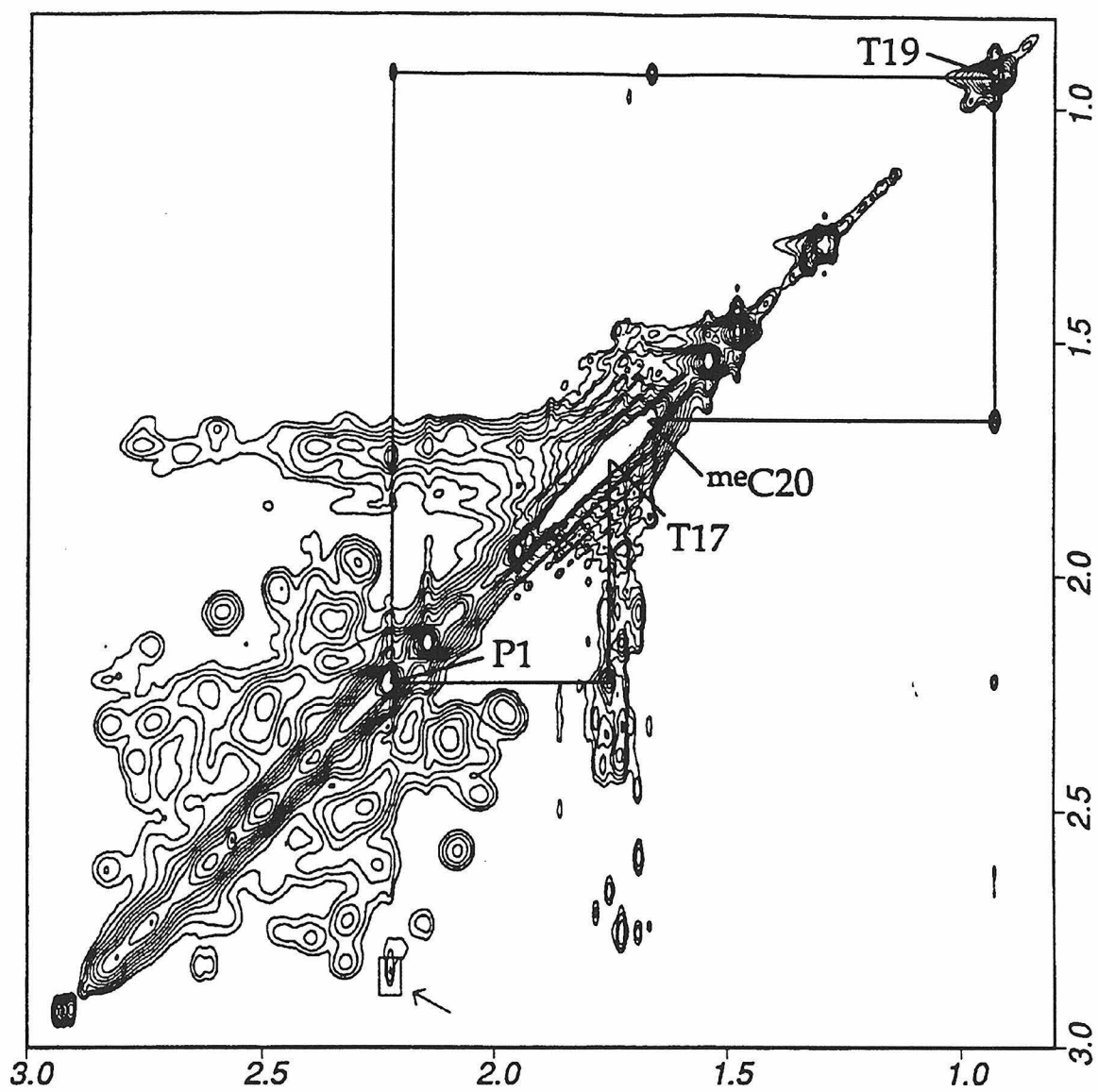


Table I
¹H chemical shifts for the exchangeable^a and non-exchangeable^b protons in the central tri-nucleotide segment in triplex 4.

Residue	H1/H3	H2 ^c	H8/H6	NH2 ^d	H5/CH3	H1'	H2',H2"	H3'
T3	14.39		7.63		1.78	6.04	e, 2.62	4.94
C4			7.69	8.75, 7.98	5.69	6.20	2.40, 2.76	4.80
T5	14.31		7.46		1.69	5.88	2.08, 2.60	4.93
A10		7.24	7.57	7.08, 7.13		5.99	2.32, 2.83	4.87
G11	12.62		6.81			5.85	2.32, 2.81	4.97
A12		7.45	7.25	7.70, 7.45		5.93	2.21, 2.80	4.84
T17	13.10		7.43		1.75	6.08	2.15, 2.84	4.97
P1	11.63 ^f			7.07	2.22	6.45	2.63, 2.68	4.94
T19	12.88		7.40		0.93	6.06	2.32, 2.74	4.64

a: Chemical shifts for the exchangeable protons at 5 °C, pH 6.01.

b: Chemical shifts for the non-exchangeable protons at 15 °C, pH 6.02.

c: Chemical shifts for the H2 protons in H₂O at 5 °C, pH 6.01.

d: For cytidine and adenine amino protons, the chemical shift of the Watson-Crick hydrogen bonded proton is indicated first.

e: Could not be unambiguously assigned due to severe overlap.

f: Chemical shift is pH dependent.

(6.81 ppm) and T19 CH₃ (0.93 ppm) protons, as well as the downfield shifts associated with the H1' (6.45 ppm) and CH₃ (2.22 ppm) protons of P. While the cause for the upfield shift for the G11 H8 proton is not obvious, the upfield shift for the methyl protons of T19 is most likely due to the positioning of the methyl group on top of the P ring system. The latter effect was observed in an earlier study where T19 was flanked by a guanine on the 5' side (Radhakrishnan et al., 1992). In the instances involving P protons, it should be noted that the ring system of P might also contribute to the unusual chemical shifts.

Discussion

As noted earlier, the requirement for protonation at the N³ positions of third strand cytosines in Y•RY triplexes has several drawbacks. The stability of these structures is sensitive to the solution pH, and the addition of charge arising from protonation has a destabilizing effect on the structure, particularly in sequences that contain several contiguous C+GC triples (Moser & Dervan, 1987; Maher & Dervan, 1990; Plum et al., 1990; Singleton & Dervan, 1992). In order to obviate this requirement for protonation, several synthetic approaches have been undertaken in designing suitable protonated cytosine analogs (Ono et al., 1991; Krawczyk et al., 1992; Miller et al., 1992; Koh & Dervan, 1992). The nonnatural nucleoside P not only mimics protonated cytosines in sequence-specific interactions with guanines in GC pairs, but can also bind to several contiguous GC base pairs over an extended pH range. In this study, the structural consequences of introducing this base at a single site opposite a guanine in an intramolecular Y•RY triplex **4** have been examined by NMR spectroscopy.

From our NMR data, it appears that the P base is readily accommodated in an otherwise pyrimidine third strand segment. Sequential NOE contacts between P and its flanking thymidine residues involving imino protons (Figure 2B) and methyl protons (Figure 6) attest to a stacked-in conformation for this base. Sequential NOEs between

base and sugar protons consistent with a regular, right handed helix are observed (Figures 5B and 6) for the segment T17-P-T19. In the 5'-P-T19-3' step, an NOE between the H6 proton of T19 and the H1' proton of P is detected (Figure 5B), while in the 5'-T17-P-3' step, an NOE between the methyl proton of P and the H2'' proton of T17 is observed (Figure 6). Other NOEs including T19(H6)-P(H2'), T19(H6)-P(H2'') and T17(H6)-P(CH₃) provide further evidence for a right-handed helical conformation in the T17-P-T19 segment. Cross-strand NOE patterns between protons on adjacent triples suggest no apparent structural distortions manifested in an altered width for the groove formed by the purine strand and pyrimidine third strand.

As we noted earlier, in order to define the pairing alignment of the P•GC triple, the P imino proton should provide an important marker. The P imino proton resonance could be easily assigned from the NOESY spectrum recorded in H₂O (Figure 2B). However, compared to the other interior hydrogen bonded imino protons in the molecule, the P imino proton exchanges with the solvent at a faster rate, which attenuates the intensity of the NOEs shown by this proton. We do detect a critical NOE, albeit weak, between this imino proton and the H8 proton of G11 in the same NOESY spectrum (Figure 3B). In addition, an NOE of weak intensity could be observed between the imino and amino protons of P that allowed us to assign the amino proton resonance (7.07 ppm). Most importantly, however, we do not detect any NOE between the same imino proton and the amino protons of C4. These observations taken in the light of our findings in D₂O (discussed above) constrain the position of the P imino proton opposite the N⁷ proton of G11 while implicitly placing the amino protons nearer to the O⁶ atom of the same base, consistent with the proposed alignment **3** (Koh & Dervan, 1992).

In summary, we have structurally characterized the interaction of a nonnatural analog of protonated cytidine with a GC base pair in an intramolecular pyrimidine·purine·pyrimidine triple helix by NMR spectroscopy. The P base is readily accommodated in an otherwise pyrimidine third strand segment while participating in

hydrogen bonding interactions consistent with the proposed pairing alignment 3.

References

- Beal, P. A., & Dervan, P. B. (1991) *Science* 251, 1360-1363.
- Cooney, M., Czernuszewicz, G., Postel, E. H., Flint, S. J., & Hogan, M. E. (1988) *Science* 241, 456-459.
- de los Santos, C., Rosen, M., & Patel, D. (1989) *Biochemistry* 28, 7282-7289.
- Felsenfeld, G., Davies, D., & Rich, A. (1957) *J. Am. Chem. Soc.* 79, 2023-2024.
- Gait, M. J. (1984) *Oligonucleotide synthesis: a practical approach*, IRL Press Limited, Oxford.
- Koh, J. S., & Dervan, P. B. (1992) *J. Am. Chem. Soc.* 114, 1470-1478.
- Krawczyk, S. H., Milligan, J. F., Wadwani, S., Moulds, C., Froehler, B. C., & Mateucci, M. D. (1992) *Proc. Natl. Acad. Sci. U.S.A.* 89, 3761-3764.
- LeDoan, T., Perrouault, L., Praseuth, D., Habhoub, N., Decout, J. L., Thoung, N. T., Lhomme, J., & Hélène, C. (1987) *Nucleic Acids Res.* 15, 7749-7760.
- Lin, C. H., & Patel, D. J. (1992) *J. Am. Chem. Soc.* 114, 10658-10660.
- Maher, L. J., & Dervan, P. B. (1990) *Biochemistry* 29, 8820-8826.
- Maher, L. J., Dervan, P. B., & Wold, B. (1992) *Biochemistry* 31, 70-81.
- Marion, D., Ikura, M., & Bax, A. (1989a) *J. Magn. Reson.* 84, 425-430.
- Marion, D., Ikura, M., Tschudin, R., & Bax, A. (1989b) *J. Magn. Reson.* 85, 393-399.
- Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun.* 113, 967-974.
- Miller, P. S., Bahn, P., Cushman, C. D., & Trapane, T. L. (1992) *Biochemistry* 31, 6788-6793.
- Moser, H. E., & Dervan, P. B. (1987) *Science* 238, 645-650.
- Ono, A., Ts'o, P. O. P., & Kan, L.S. (1991) *J. Am. Chem. Soc.* 113, 4032-4033.
- Plateau, P., & Gueron, M. (1982) *J. Am. Chem. Soc.* 104, 7310-7311.

- Plum, E. G., Park, Y.-W., Singleton, S. F., Dervan, P. B., & Breslauer, K. J. (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87, 9436-9440.
- Povsic, T. J., & Dervan, P. B. (1989) *J. Am. Chem. Soc.* 111, 3059-3061.
- Praseuth, D., Perrouault, L., Le Doan, T., Chassignol, M., Thuong, N., & Hélène, C. (1988) *Proc. Natl. Acad. Sci. U.S.A.* 85, 1349-1353.
- Radhakrishnan, I., Gao, X., de los Santos, C., Live, D., & Patel, D. J. (1991) *Biochemistry* 30, 9022-9030.
- Radhakrishnan, I., Patel, D. J., & Gao, X. (1992) *Biochemistry* 31, 2514-2523.
- Rajagopal, P., & Feigon, J. (1989) *Biochemistry* 28, 7859-7870.
- Singleton, S. F., & Dervan, P. B. (1992) *Biochemistry* 31, 10995-11003.
- Strobel, S. A., Doucette-Stamm, L. A., Riba, L., Housman, D. E., & Dervan, P. B. (1991) *Science* 254, 1639-1642.
- Watson, J. D., & Crick, F. H. C. (1953) *Nature* 171, 737-738.

CHAPTER THREE

Sequence Composition Effects on the Energetics of Triple Helix Formation by Oligonucleotides Containing a Designed Mimic of Protonated Cytosine

*The text of this chapter is taken from a published paper that was coauthored with Prof.
Peter B. Dervan.*

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

Introduction

The thermodynamic stability of pyrimidine•purine•pyrimidine triple helices decreases with increasing pH due to the requirement of protonating cytosine bases to form C+GC triplets.^{1,2} Within the context of pyrimidine oligonucleotide-directed recognition of double helical DNA, there are serious sequence composition limitations with regard to targeting *contiguous* G-rich purine tracts near physiological pH, presumably due to electrostatic repulsion between protonated cytosines in adjacent C+GC triplets. Replacement of cytosine by 5-methylcytosine (^mC) increases the stability of pyrimidine•purine•pyrimidine triple helices, but does not alleviate the pH dependence (Figure 1).² Development of oligonucleotides whose energetics of triple helix formation are less sensitive to pH would benefit applications which require near physiologically relevant conditions.

A series of nonnatural nucleosides have been shown to mimic protonated cytosine and bind GC base pairs with significantly less pH dependence, including pseudoisocytosine,³ 8-oxoadenine,⁴ 3-methyl-5-amino-pyrazolopyrimidin-7-one (P),⁵ and most recently, 4-amino-5-methyl-2,6-pyrimidione.⁶ In a qualitative study, we demonstrated that pyrimidine rich oligonucleotides containing P could bind target sites containing multiple GC base pairs (Figure 1).⁵ However, no conclusions regarding the energetic differences, if any, between sites of sequence composition 5'-d(AAAAAGAG-AGAGAGA)-3' and 5'-d(AAAAGAAAAGGGGGGA)-3' could be drawn.⁵ Although NMR studies do not reveal any major backbone distortion for P•GC triplets, the P•GC triplet is not isomorphous with adjacent T•AT triplets.⁷ Therefore, the issue remains whether triple helices containing *multiple* P residues are energetically disfavored relative to ^mC at neutral pH. We report here the results of quantitative DNase footprint titration experiments which demonstrate a dramatic *effect of sequence composition* on triple helix formation by oligonucleotides containing P and ^mC (Figure 1).

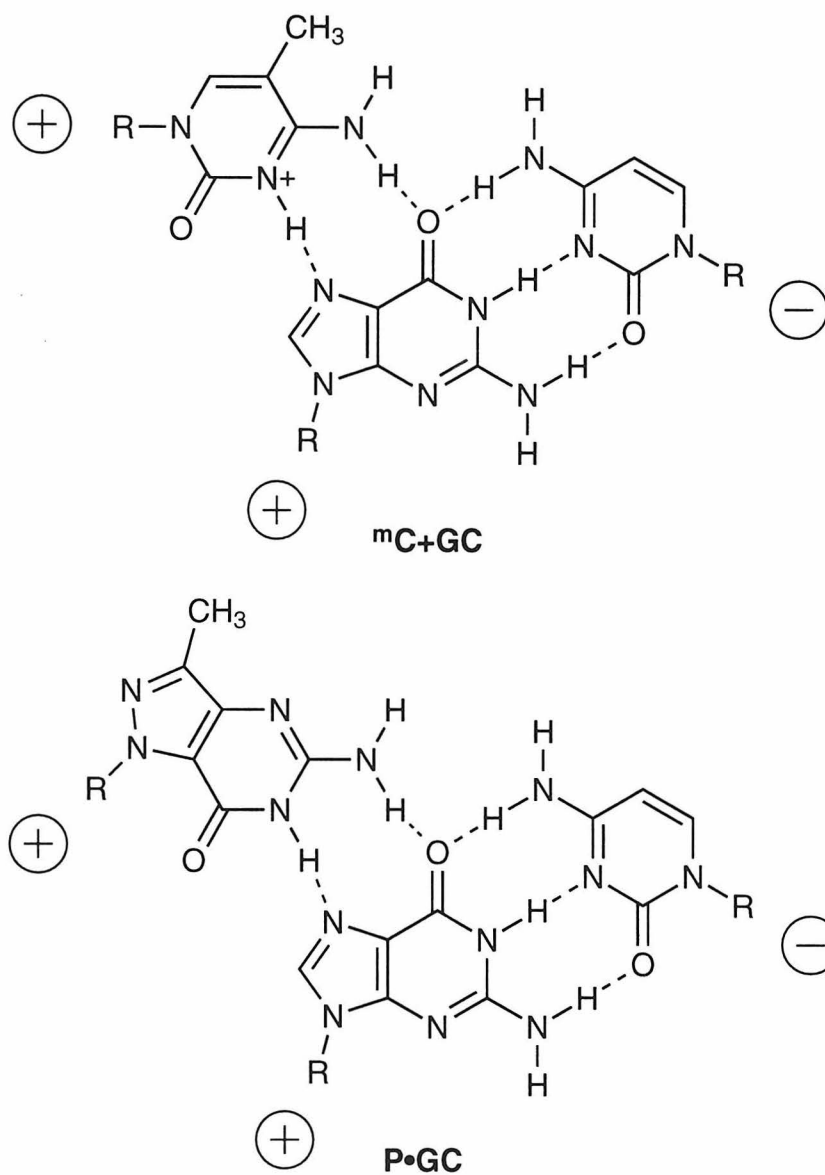


Figure 1: Two-dimensional representations of ^mC+GC and P•GC triplets.⁵ In each case, the third strand base forms two hydrogen bonds with the Hoogsteen face of the Watson-Crick purine base. The plus and minus signs indicate relative backbone polarity.

Results and Discussion

Methods. A description of the quantitative footprint titration method has been presented elsewhere for protein-DNA complexation⁸ and has been applied to small molecule-DNA interactions.⁹ However, the method has not been used extensively to study triple helix formation;¹⁰ therefore, a protocol for this purpose is outlined in the Experimental Section. In separate experiments, 3'-³²P end-labeled 242, 253, or 314 base pair restriction fragments containing purine rich 15 or 16 base pair target sites were allowed to equilibrate at 22 °C, pH 7.0, with a series of concentrations of third strand oligonucleotide that ranged from 40 μ M to 80 pM (Figure 2). Following a 120 hour equilibration, DNase I was added and digestion was allowed to proceed for 6 minutes at room temperature. After quenching, the reaction mixtures were separated by denaturing polyacrylamide gel electrophoresis and the resulting gels were imaged by storage phosphor autoradiography. 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 Tables I-III.

Affinity and Specificity of P. In order to test the affinity and specificity of P relative to the four natural base pairs, the association constants for the formation of 20 triple helical complexes which vary at a single common position $Z \cdot XY$ (where $Z=P$, mC , T, G, and A and $XY=AT$, CG, GC, and TA) were determined within a pyrimidine motif by quantitative footprint titration at 22 °C, pH 7.0 (Figure 2, top). A sample quantitative DNase titration gel is shown in Figure 2 (bottom), and averaged binding data for the 20 complexes is plotted in Figure 3. The sequence context in which the $Z \cdot XY$ pairings were

Figure 2: (Top) Sequences of the oligonucleotides and the four purine rich target sites for the specificity experiments. The duplex target sites, which differ only at a single position labeled XY, are arranged sequentially on the 314 bp *Afl* II/*Fsp* I restriction fragment of pYSPEC2, each separated by 15 bp of intervening sequence. (C indicates 5-methyl-2'-deoxycytidine and P indicates 1-(2-deoxy- β -D-ribofuranosyl)-3-methyl-5-amino-1H-pyrazolo[4,3-d]pyrimidin-7-one). (Bottom) Autoradiogram of an 8% denaturing polyacrylamide gel used to separate the fragments generated by DNase I digestion in a quantitative footprint titration experiment. The boxes drawn to the left of the autoradiogram indicate the positions of the 15 bp binding sites within the 314 bp restriction fragment. (Lane 1) Intact 3' labeled DNA after incubation in the absence of the third strand oligonucleotide. (Lane 2) Products of an adenine specific sequencing reaction. (Lane 3) Products of a guanine specific sequencing reaction. (Lanes 4-22) DNase I digestion products obtained in the presence of varying concentrations of oligonucleotide **1**: no oligonucleotide (lane 4); 80 pM (lane 5); 200 pM (lane 6); 400 pM (lane 7); 800 pM (lane 8); 2 nM (lane 9); 4 nM (lane 10); 8 nM (lane 11); 20 nM (lane 12); 40 nM (lane 13); 80 nM (lane 14); 200 nM (lane 15); 400 nM (lane 16); 800 nM (lane 17); 2 μ M (lane 18); 4 μ M (lane 19); 8 μ M (lane 20); 20 μ M (lane 21); 40 μ M (lane 22).

5' - TTCTCT**Z**TCTCCTTT - 3'

5' - AAGAGAXAGAGGAAA - 3'

3' - TTCTCT**Y**TCTCCTTT - 5'

Z = P mC T G A
 oligo 1 2 3 4 5

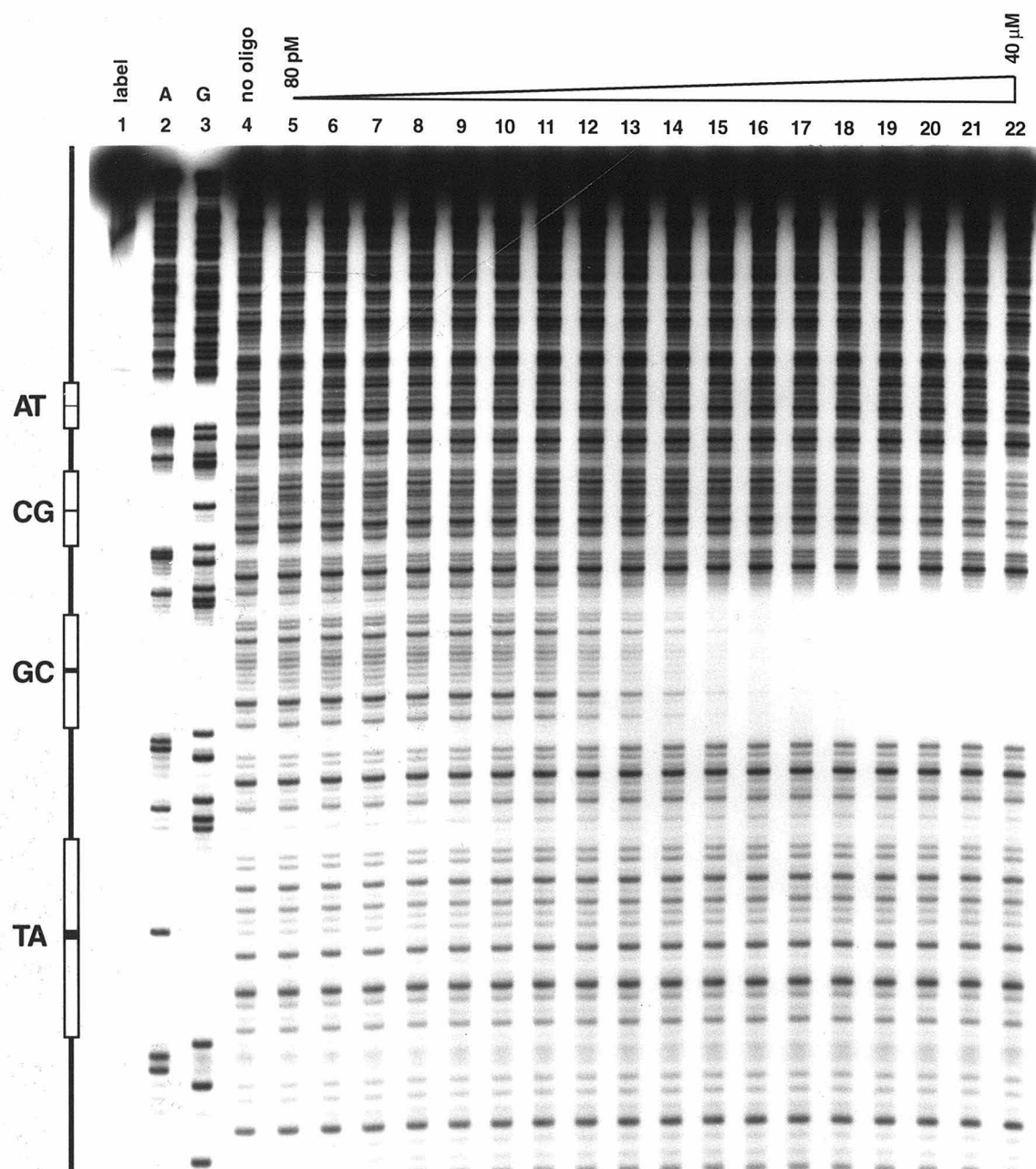
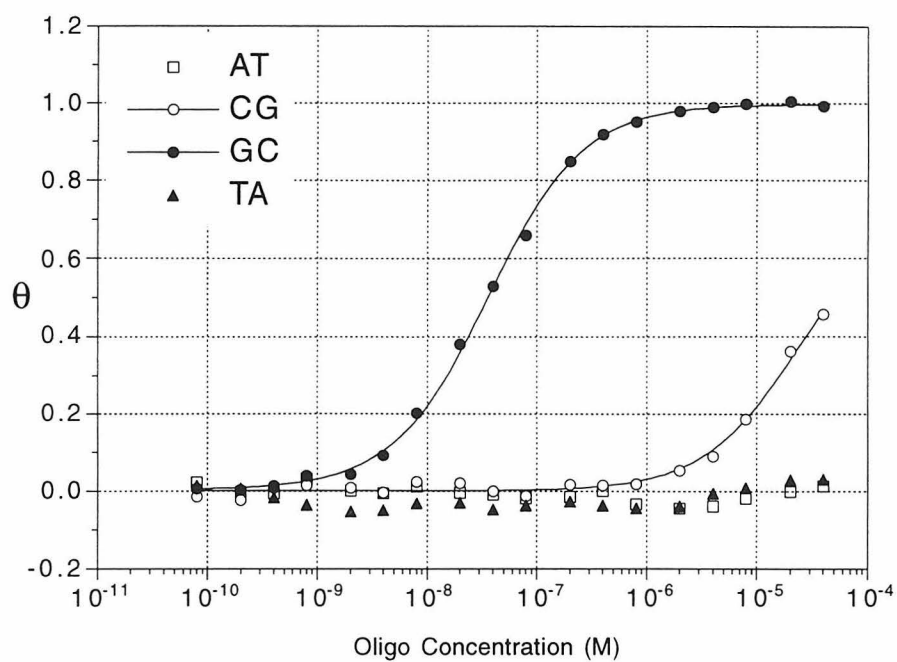
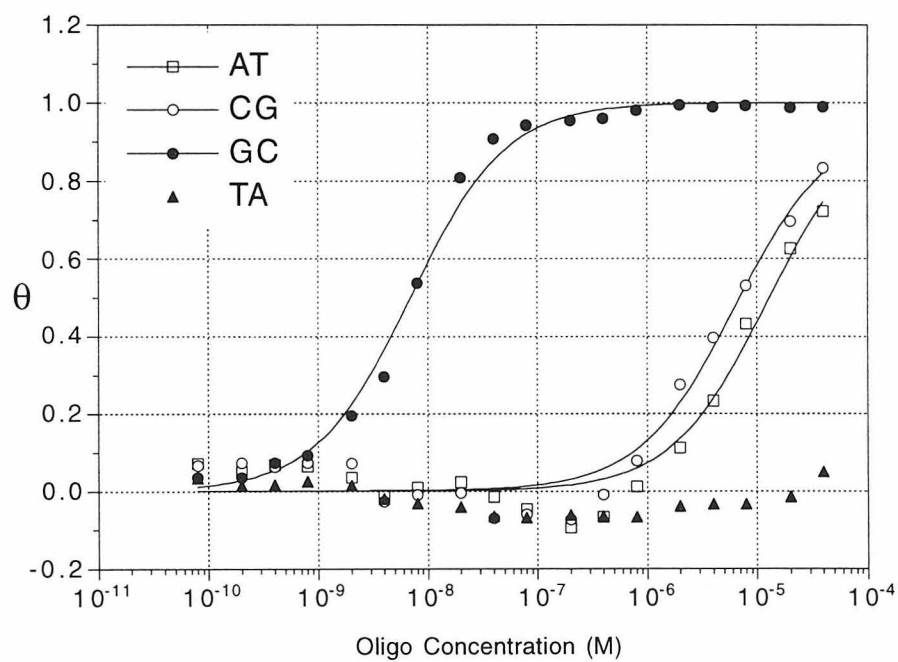
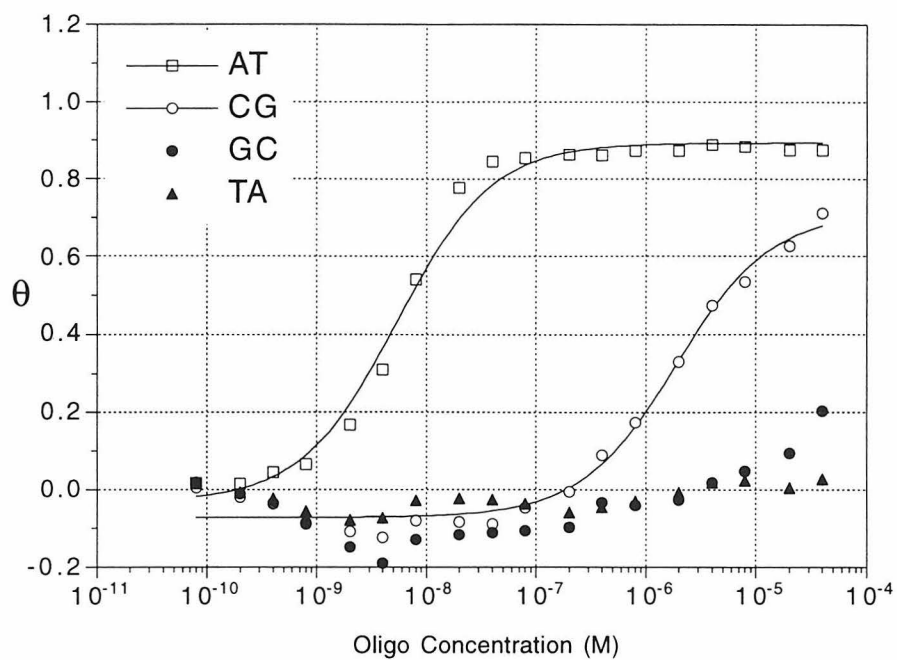
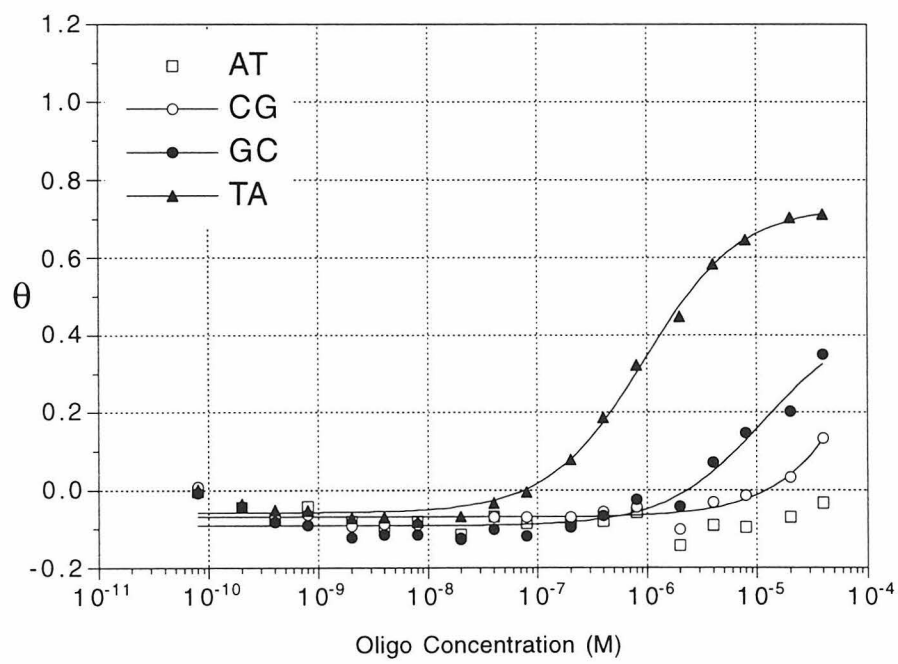
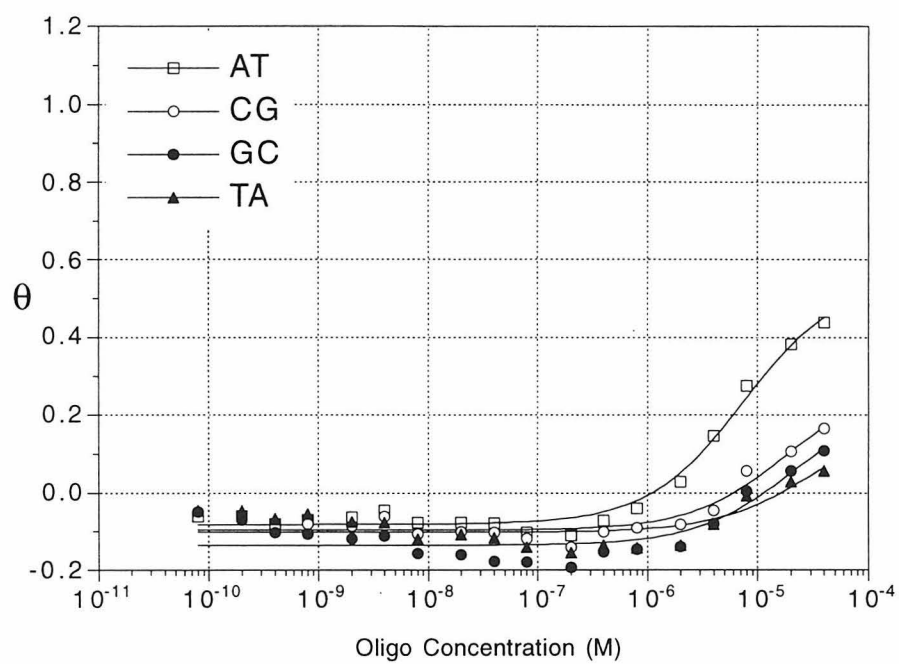


Figure 3: Plots of the site occupancy (θ) versus oligonucleotide concentration for binding of oligonucleotides **1-5** to each of the four target sites on the 314 bp *Afl* II/*Fsp* I restriction fragment of pYSPEC2. A) oligo **1** containing P at the variable position. B) oligo **2** containing MeC at the variable position. C) oligo **3** containing T at the variable position. D) oligo **4** containing G at the variable position. E) oligo **5** containing A at the variable position. Each data set is averaged from three independent gels. Binding isotherms derived from the data are shown as solid lines.

A. P•XY**B. MeC•XY**

C. T•XY**D. G•XY**

E. A•XY

studied is 5'-d(AAGAGAXAGAGGAAA)-3' within a 314 base pair DNA restriction fragment.¹¹ The affinity of P relative to ^mC is determined by comparison of oligonucleotides **1** and **2** (Z=P and ^mC, respectively) opposite XY=GC. The sequence specificity of P is confirmed by comparison of oligonucleotide **1** (Z=P) opposite each of the four Watson-Crick base pairs (XY=AT, CG, GC, and TA). The stability of triple helical complexes is very sensitive to solution conditions such as pH,^{2c} cation valence and concentration,¹² and temperature.¹³ Therefore, we have chosen experimental conditions (pH 7.0, 22 °C, 100 mM NaCl, 10 mM bis-tris, 250 μM spermine) such that the association constants for the triple helical complexes span the range which can be measured by quantitative DNase I footprinting.

The values of the 20 association constants (K_T) ranged from $< 10^4$ to $\geq 2.1 \times 10^8$ M⁻¹ (Table I). An examination of the data confirms that two base triplets, T•AT and C+GC, are particularly stable and that the G•TA triplet is of intermediate stability.^{14,15} We find that oligonucleotide **1** (Z=P) binds the target site with an equilibrium association constant of 3.1×10^7 M⁻¹, a factor of 5 lower affinity than oligonucleotide **2** (Z=^mC). Importantly, P is specific for GC base pairs by at least three orders of magnitude. The stabilities of triple helical complexes containing P•XY decrease in the order P•GC>>P•CG>P•AT, P•TA (Table I). The high affinity and sequence specificity observed is consistent with an NMR structural study which indicates that P forms two specific hydrogen bonds to the purine strand of the Watson-Crick GC base pair.⁷

Sequence Composition Effects. Sequence composition effects on the energetics of triple helix formation by oligonucleotides containing multiple P moieties were examined for two different purine target sites within 242 and 253 base pair restriction fragments at pH 7.0 and 7.5 (Figure 4). The sequence of one site is the purine tract 5'-d(AAAAAGAGAGAGAGA)-3' in a 242 base pair restriction fragment (which we will refer to as the (GA)₅ site). The other purine site is derived from the LTR region of the

Table I: Equilibrium Association Constants for Twenty Triple Helical Complexes which Vary at a Single Position^a

oligo	Z ^b	XY	K _T	ΔG° ^c
1	P	AT	< 10 ⁴	> -5.4
		CG	< 5 × 10 ⁴	> -6.3
		GC	3.1 (±1.0) × 10 ⁷	-10.1 (±0.2)
		TA	< 10 ⁴	> -5.4
2	mC	AT	8.2 (±1.9) × 10 ⁴	-6.6 (±0.1)
		CG	1.7 (±0.6) × 10 ⁵	-7.1 (±0.2)
		GC	≥ 1.4 × 10 ⁸	< -11.0
		TA	< 10 ⁴	> -5.4
3	T	AT	≥ 2.1 × 10 ⁸	< -11.2
		CG	5.4 (±0.7) × 10 ⁵	-7.7 (±0.1)
		GC	< 10 ⁴	> -5.4
		TA	< 10 ⁴	> -5.4
4	G	AT	< 10 ⁴	> -5.4
		CG	< 10 ⁴	> -5.4
		GC	< 5 × 10 ⁴	> -6.3
		TA	1.1 (±0.3) × 10 ⁶	-8.2 (±0.1)
5	A	AT	1.6 (±0.6) × 10 ⁵	-7.0 (±0.2)
		CG	< 10 ⁴	> -5.4
		GC	< 10 ⁴	> -5.4
		TA	< 10 ⁴	> -5.4

^a Each reported K_T value is the mean (±SEM) of three independent measurements which were performed in 100 mM NaCl, 10 mM bis-tris, 250 μM spermine at pH 7.0 and 22° C. Values for K_T and ΔG° are reported in units of M⁻¹ and kcal • mol⁻¹, respectively. ^b Z refers to the position in the third strand oligonucleotide opposite the variable position (XY) in the duplex. ^c ΔG° values were obtained from the relationship ΔG° = -RTlnK (T = 295 K) and error estimates were obtained by propagating the SEM for K_T through this formula.

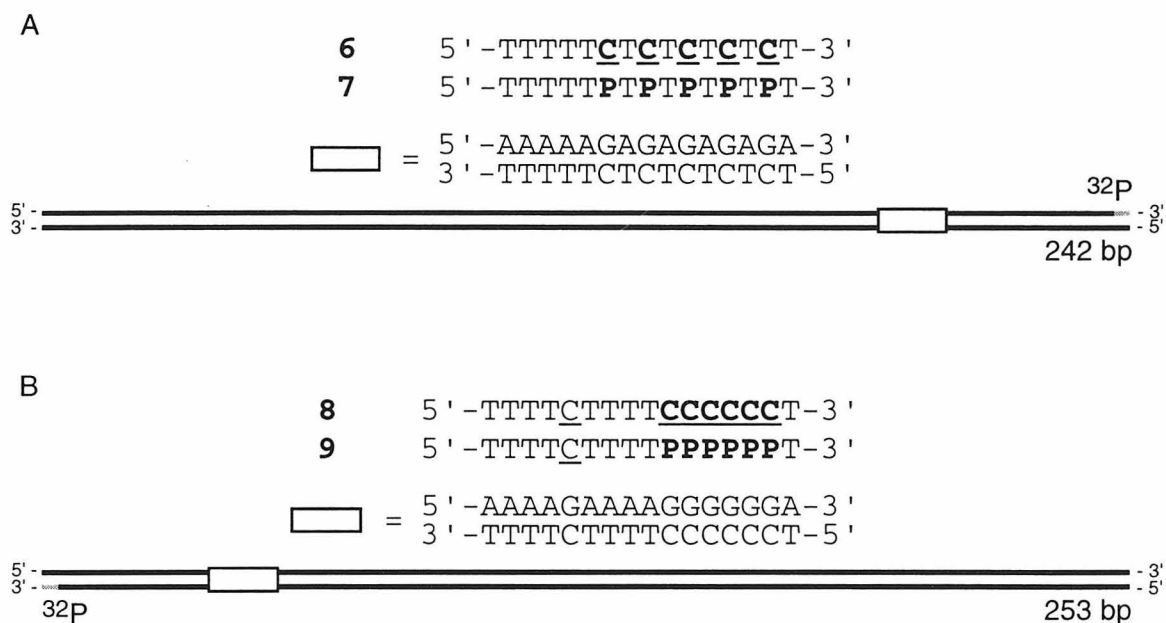


Figure 4: Sequence Composition Experiments. A) Sequences of oligonucleotides **6** and **7** and the (GA)₅ target site. The target site is located within the 242 bp *Eco* RI/*Pvu* II restriction fragment of pGCBGC. B) Sequences of oligonucleotides **8** and **9** and the G₆ target site. The target site is located within the 253 bp *Eco* RI/*Pvu* II restriction fragment of pSPHIV. (C indicates 5-methyl-2'-deoxycytidine and P indicates 1-(2-deoxy-β-D-ribofuranosyl)-3-methyl-5-amino-1H-pyrazolo[4,3-d]pyrimidin-7-one).

HIV genome and has the sequence 5'-d(AAAAGAAAAGGGGGGA)-3' in a 253 bp restriction fragment (which we will refer to as the G₆ site) (Figure 4).

The energetics of association of oligonucleotides **6** and **7** of sequence composition 5'-d(TTTTT^mCT^mCT^mCT^mCT)-3' and 5'-d(TTTTTPTPTPTPTPT)-3', respectively, allow comparison of the ability of ^mC and P to bind multiple GC base pairs in a target sequence composed of alternating G and A nucleotides (the (GA)₅ site). Similarly, the energetics of binding oligonucleotides **8** and **9** of sequence composition 5'-d(TTTT^mCTTTT^mC^mC^mC^mC^mC^mCT)-3' and 5'-d(TTTT^mCTTTTPTPTPTPTPT)-3' allow this comparison for a target sequence containing contiguous GC base pairs (the G₆ site). Sample quantitative DNase footprinting gels and average binding data for these experiments are shown in Figures 5 to 8.

We find that oligonucleotide **6**, containing ^mC and T, binds tightly to the (GA)₅ site ($K_T = 3.1 \times 10^9 \text{ M}^{-1}$) at pH 7.0 (Table II), and this affinity drops by at least a factor of 30 as the pH is increased to 7.5. This is consistent with previous data which has shown that triple helix formation by oligonucleotides containing cytosine or ^mC is pH dependent.² At pH 7.0, oligonucleotide **7**, containing P in place of ^mC, binds four orders of magnitude weaker than oligonucleotide **6**, albeit pH independently over the range studied (Table II). *The situation changes entirely for the G₆ target sequence containing contiguous GC base pairs.* Oligonucleotide **8**, which contains (^mC)₆, binds the G₆ site very weakly ($K_T < 5 \times 10^4$) at pH 7.0 (Table III). In the case of oligonucleotide **9**, which contains P₆, the association constant at pH 7.0 is very high ($K_T = 4.4 \times 10^9 \text{ M}^{-1}$). Remarkably, the high affinity oligonucleotide **9** decreases by only a factor of 2 at pH 7.5 (Table III).

The (GA)₅ site contains ten 5'-AG-3' or 5'-GA-3' junctions, while the G₆ site contains only two such junctions in the region of interest. The four orders of magnitude difference in the affinity of P containing oligonucleotides for these two sites is probably due to the lack of structural isomorphism in the P•GC and T•AT triplets. The location of

Figure 5: Autoradiogram of an 8% denaturing polyacrylamide gel used to separate the fragments generated by DNase I digestion in a quantitative footprint titration experiment for binding of oligonucleotide **6** to the (GA)₅ target site at pH 7.0. The box drawn to the left of the autoradiogram indicates the position of the 15 bp binding site within the 242 bp restriction fragment. (Lane 1) Intact 3' labeled DNA after incubation in the absence of the third strand oligonucleotide. (Lane 2) Products of an adenine specific sequencing reaction. (Lanes 3-21) DNase I digestion products obtained in the presence of varying concentrations of oligonucleotide **6**: no oligonucleotide (lane 3); 80 pM (lane 4); 200 pM (lane 5); 400 pM (lane 6); 800 pM (lane 7); 2 nM (lane 8); 4 nM (lane 9); 8 nM (lane 10); 20 nM (lane 11); 40 nM (lane 12); 80 nM (lane 13); 200 nM (lane 14); 400 nM (lane 15); 800 nM (lane 16); 2 μ M (lane 17); 4 μ M (lane 18); 8 μ M (lane 19); 20 μ M (lane 20); 40 μ M (lane 21).

5' - TTTTCTCTCTCTCT - 3'

5' - AAAAAGAGAGAGAGA - 3'
3' - TTTTCTCTCTCTCT - 5'

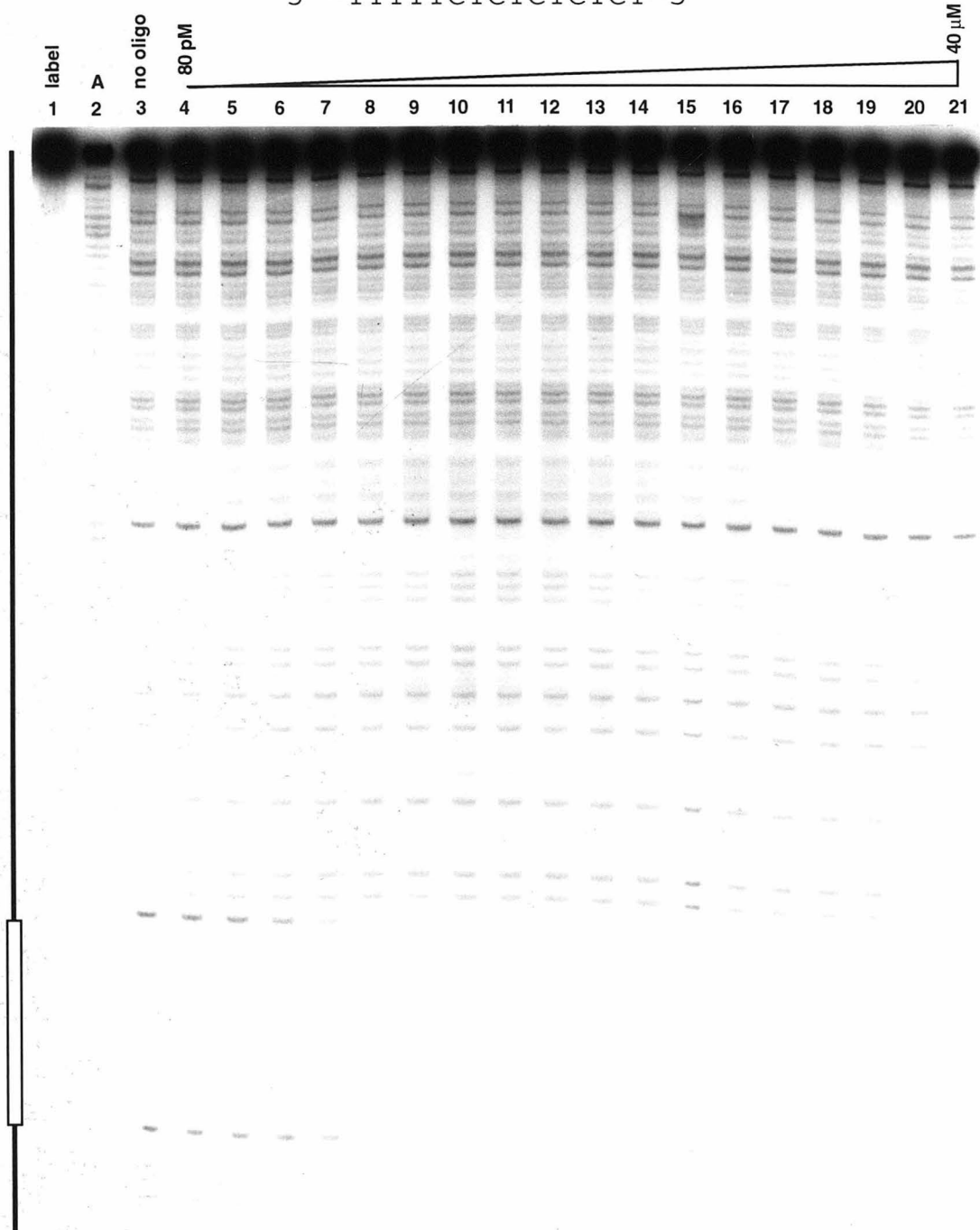


Figure 6: Plots of the site occupancy (θ) versus oligonucleotide concentration for binding of oligonucleotides **6** and **7** to the $(GA)_5$ target site on the 242 bp *Eco* RI/*Pvu* II restriction fragment of pGCBGC. A) oligo **6** at pH 7.0 and 7.5. B) oligo **7** at pH 7.0 and 7.5. Each data set is averaged from three independent gels. Binding isotherms derived from the data are shown as solid lines.

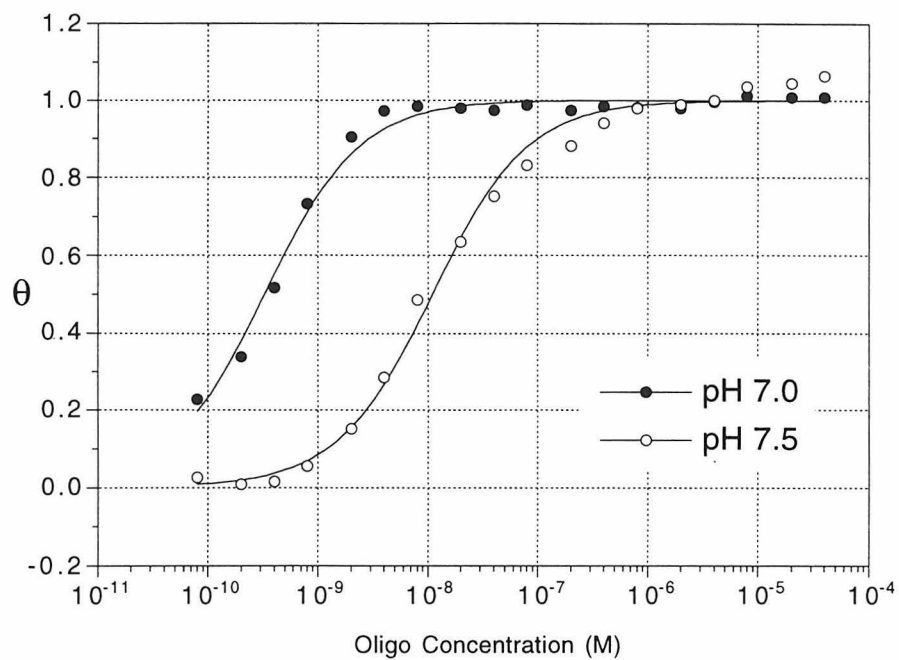
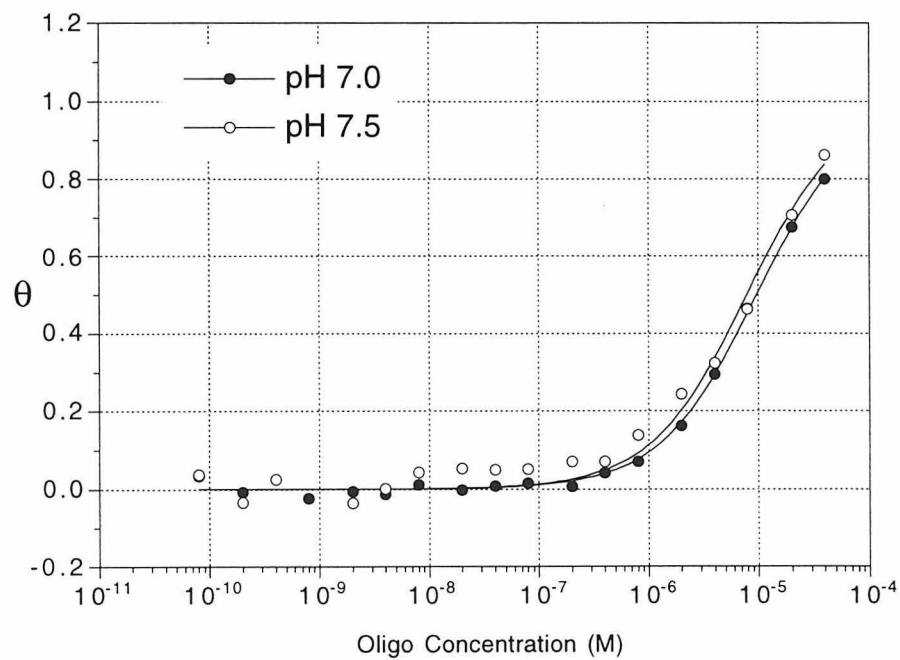
A. oligo 6/(GA)₅ site**B. oligo 7/(GA)₅ site**

Figure 7: Storage phosphor autoradiogram of an 8% denaturing polyacrylamide gel used to separate the fragments generated by DNase I digestion in a quantitative footprint titration experiment for binding of oligonucleotide **9** to the (G)₆ target site at pH 7.5. The box drawn to the left of the autoradiogram indicates the position of the 16 bp binding site within the 253 bp restriction fragment. (Lane 1) Products of a guanine specific sequencing reaction. (Lane 2) Products of an adenine specific sequencing reaction. (Lanes 3-21) DNase I digestion products obtained in the presence of varying concentrations of oligonucleotide **9**: no oligonucleotide (lane 3); 80 pM (lane 4); 200 pM (lane 5); 400 pM (lane 6); 800 pM (lane 7); 2 nM (lane 8); 4 nM (lane 9); 8 nM (lane 10); 20 nM (lane 11); 40 nM (lane 12); 80 nM (lane 13); 200 nM (lane 14); 400 nM (lane 15); 800 nM (lane 16); 2 μ M (lane 17); 4 μ M (lane 18); 8 μ M (lane 19); 20 μ M (lane 20); 40 μ M (lane 21). (Lane 22) Intact 3' labeled DNA after incubation in the absence of the third strand oligonucleotide.

5' - TTTTCTTTT**PPPPPT** - 3'

5' - AAAAGAAAAGGGGGGA - 3'
3' - TTTTCTTTTCCCCCCT - 5'

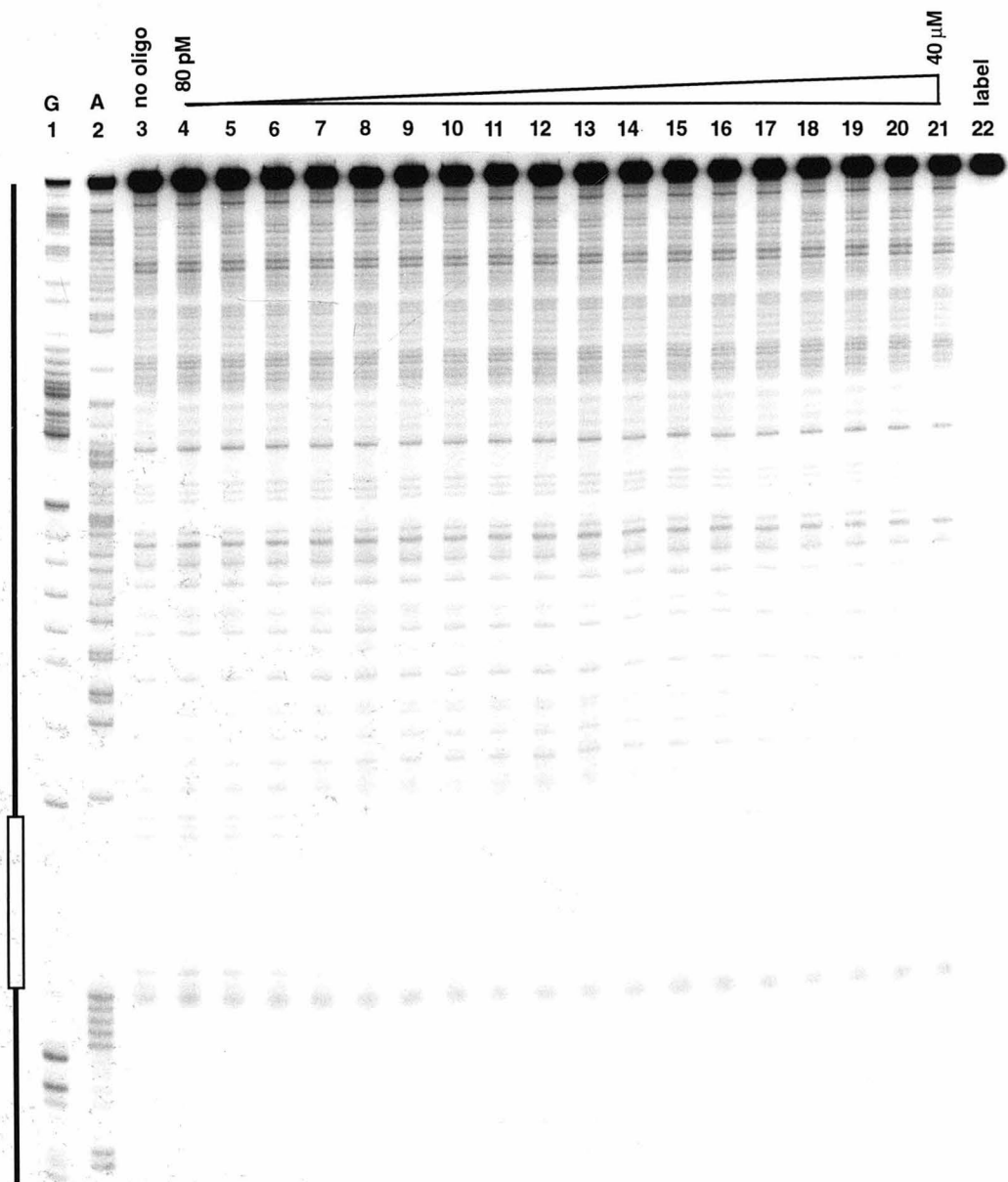


Figure 8: Plots of the site occupancy (θ) versus oligonucleotide concentration for binding of oligonucleotides **8** and **9** to the G_6 target site on the 253 bp *Eco* RI/*Pvu* II restriction fragment of pSPHIV. A) oligo **8** at pH 7.0 and 7.5. B) oligo **9** at pH 7.0 and 7.5. Each data set is averaged from three independent gels. Binding isotherms derived from the data are shown as solid lines.

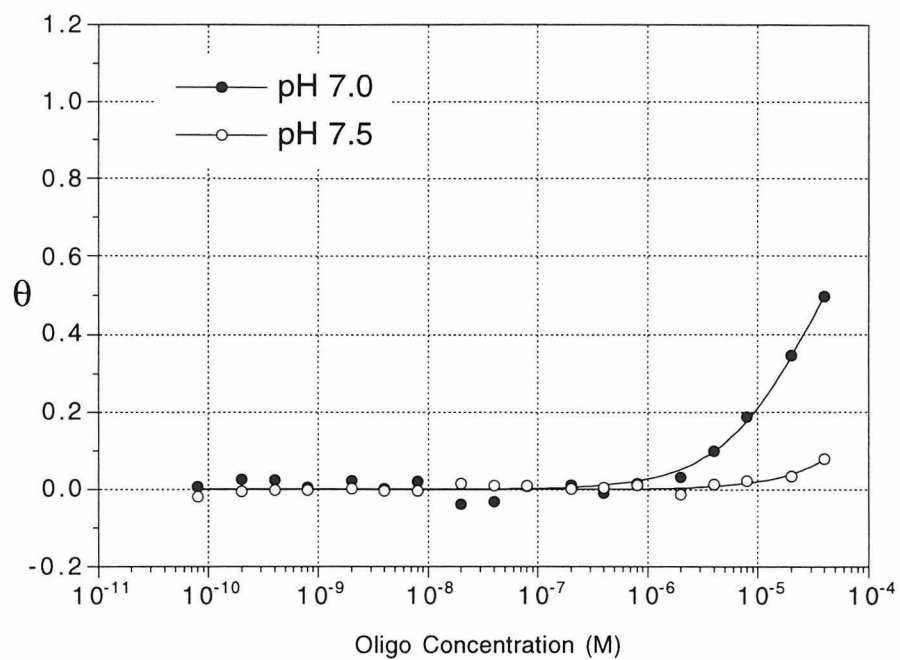
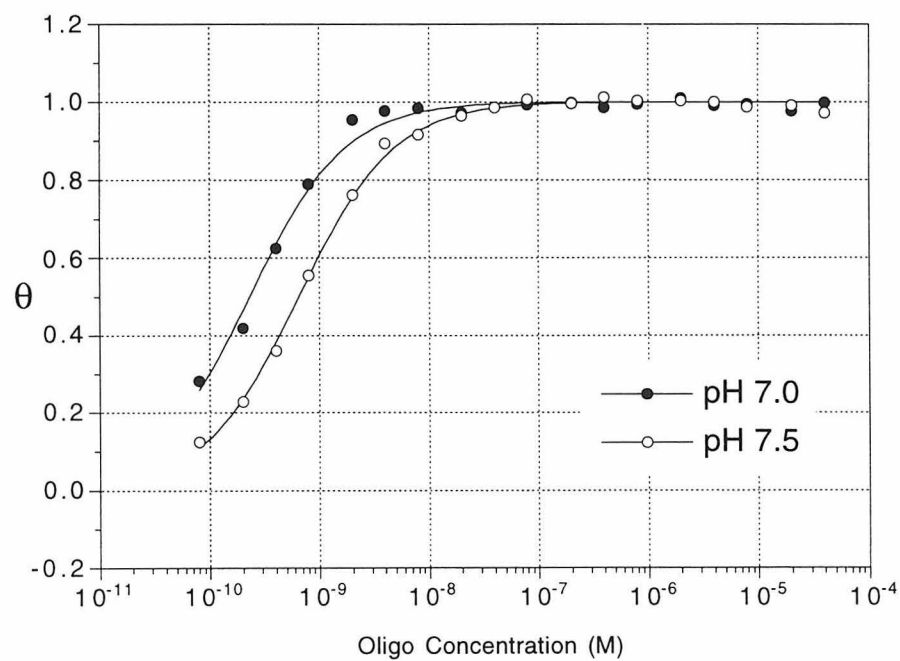
A. oligo 8/G₆ site**B. oligo 9/G₆ site**

Table II: Equilibrium Association Constants for Triple Helix Formation at the (GA)₅ Site^a

oligo	pH	K _T (M ⁻¹)	ΔG° ^b
6	7.0	3.1 (±0.2) × 10 ⁹	-12.8 (±0.1)
	7.5	9.4 (±1.6) × 10 ⁷	-10.8 (±0.1)
7	7.0	1.1 (±0.2) × 10 ⁵	-6.8 (±0.1)
	7.5	1.3 (±0.2) × 10 ⁵	-6.9 (±0.1)

^a Each reported K_T value is the mean (±SEM) of three independent measurements which were performed in 100 mM NaCl, 10 mM bis-tris, 250 μM spermine at the indicated pH and 22° C. ^b ΔG° values were obtained from the relationship ΔG° = -RTlnK (T = 295 K) and error estimates were obtained by propagating the SEM for K_T through this formula.

Table III: Equilibrium Association Constants for Triple Helix Formation at the G₆ Site^a

oligo	pH	K _T (M ⁻¹)	ΔG° ^b
8	7.0	< 5 × 10 ⁴	> -6.3
	7.5	< 10 ⁴	> -5.4
9	7.0	4.4 (±0.3) × 10 ⁹	-13.0 (±0.1)
	7.5	1.6 (±0.1) × 10 ⁹	-12.4 (±0.1)

^a Each reported K_T value is the mean (±SEM) of three independent measurements which were performed in 100 mM NaCl, 10 mM bis-tris, 250 μM spermine at the indicated pH and 22° C. ^b ΔG° values were obtained from the relationship ΔG° = -RTlnK (T = 295 K) and error estimates were obtained by propagating the SEM for K_T through this formula.

the third strand deoxyribose-phosphate backbone is not identical when the two triplets are overlaid. Thus, the 5'-AG-3' and 5'-GA-3' junctions could generate energetically unfavorable distortions in the backbone in triple helical complexes relative to contiguous A or G sequences. In the case of ^mC containing oligonucleotides, the large difference in affinity for the two sites is probably due to electrostatics, since the T•AT and ^mC+GC triplets appear structurally isomorphous. We would conclude from this limited study that in ^mC, T containing oligonucleotides, the composition in the third strand of sequence 5'-TT-3', 5'-^mCT-3', and 5'-T^mC-3' make the most favorable interactions with the double helix, while a 5'-^mC^mC-3' dinucleotide repeat is unfavorable. Sequences of contiguous ^mC+GC triplets may be energetically unfavorable due to repulsion between adjacent protonated ^mC bases.

Conclusion

In conclusion, a quantitative study of the energetics of triple helix formation for different purine tracts has revealed that ^mC and P provide complementary solutions to the recognition of GC base pairs by triple helix formation. At neutral pH, ^mC binds isolated GC base pairs with higher affinity than P, while P binds contiguous GC base pairs with higher affinity than ^mC. Remarkably, both types of sequence can be targeted by choice of the appropriate oligonucleotide composition. In addition, this work emphasizes the importance of studying the effects of sequence composition on energetics as design-synthesis of nonnatural bases expands the repertoire of sites which may be targeted by triple helix formation.

Experimental

Materials. Aqueous solutions for DNA manipulations were prepared with water purified by a Milli-Q Water System (Millipore). *E. coli* XL-1 Blue competent cells were obtained from Stratagene. Deoxynucleotide triphosphates (Ultrapure grade) and calf thymus DNA (sonicated and phenol extracted) were purchased from Pharmacia, diluted to the appropriate concentration, and stored at -20° C. [α -³²P]-thymidine-5'-triphosphate (≥ 3000 Ci/mmol) and [α -³²P]-deoxyadenosine-5'-triphosphate (≥ 6000 Ci/mmol) were obtained from Du Pont NEN. Restriction endonucleases were purchased from either New England Biolabs or Boehringer Mannheim and used according to the supplier's protocol in the buffer provided. Sequenase (version 2.0) was obtained from United States Biochemicals, and deoxyribonuclease I (FPLCpure) was purchased from Pharmacia. Snake venom phosphodiesterase and alkaline phosphatase were obtained from Boehringer Mannheim. Buffers and salts for preparation of association buffer were Fluka Microselect grade. 5-methyl-2'-deoxycytidine phosphoramidite was purchased from Cruachem. All other DNA synthesis reagents were provided by Glen Research.

Oligonucleotide Preparation. Oligonucleotides were synthesized by standard automated solid-support chemistry on an Applied Biosystems 380B DNA synthesizer using O-cyanoethyl-N,N-diisopropylphosphoramidites. P phosphoramidite was synthesized as described previously.⁵ Oligonucleotides **2-5** were synthesized with the 5'-OH unprotected and deblocked for 20-24 hours in concentrated aqueous NH₃. These oligomers were initially purified by ion exchange FPLC (Pharmacia) on a Mono Q HR 10/10 column (Pharmacia) using a gradient of 0.1 to 1.0 M NaCl in 10 mM bis-trisCl, pH 7.0. Oligonucleotides **1** and **6-9** were synthesized with a 5'-DMT protecting group and deblocked for 24-36 hours in concentrated aqueous NH₃. These oligonucleotides were initially purified by reversed phase FPLC on a ProRPC HR 10/10 column (Pharmacia)

using a linear gradient of 0 to 40% CH₃CN in 0.1 M triethylammonium acetate, pH 7.0 and then detritylated with 80% acetic acid/H₂O. Both sets of oligonucleotides were chromatographed a second time by reversed phase FPLC using either 0 to 20% CH₃CN in 10 mM NH₄HCO₃, pH 8.1 (oligonucleotides **2-5**) or 0 to 20% CH₃CN in 0.1 M triethylammonium acetate, pH 7.0 (oligonucleotides **1** and **6-9**). A final desalting step on a NAP-10 column (Pharmacia) was performed for all oligonucleotides. Concentrations of the oligonucleotide solutions were determined by UV absorbance at 260 nm, using extinction coefficients of 8800 (M⁻¹•cm⁻¹) for thymidine, 5700 (M⁻¹•cm⁻¹) for 5-methyl-2'-deoxycytidine, 11,700 (M⁻¹•cm⁻¹) for deoxyguanosine, 15,400 (M⁻¹•cm⁻¹) for deoxyadenosine, and 2500 (M⁻¹•cm⁻¹) for P. The oligonucleotides were stored dry as 5 or 10 nmol aliquots in 0.5 ml microcentrifuge tubes at -20° C.

Enzymatic Digestion and HPLC Analysis. 10 nmol oligonucleotide aliquots were digested with three units of calf alkaline phosphatase and 0.009 units of snake venom phosphodiesterase in 50 µl of 50 mM TrisHCl, 10 mM MgCl₂, pH 8.0 for 2-20 hours at 37° C. 25 µl aliquots of these solutions were analyzed by reversed phase HPLC on a C18 column (VYDAC #201HS54) with a gradient of 0-50% CH₃CN in 20 mM NH₄OAc, pH 5.0. Comparison with chromatograms of nucleoside standards and coinjections with authentic materials established the base composition of all oligonucleotides.

Plasmid Construction. pYSPEC2 and pSPHIV were prepared by standard methods.¹⁶ Briefly, pYSPEC2 was prepared by hybridizing three complementary sets of synthetic oligonucleotides: 5'-CCGGCCGAAGTCTTGAGGCTTTCCTCTTTCTCTTCCGAAGTCTTGAGGCTTTCCTCTGTCTCTCCGAA-3' and 5'-AGACTTCGGAAGAGACAGAGGAAAGCCTCAAGACTTCGGAAGAGAAAGAGGAAAGCCTCAAGACTTCGG-3', 5'-GTCTTGAGGCTTTCCTCTCTCTCTTCCGAAGTCTTGAGGCTTTCCTCTATCTCTTCCGAAGTCTTGAGGC-3' and 5'-AGCTGCCTCAAGACTTCGGAAGAGA-

TAGAGGAAAGCCTCAAGACTTCGGAAGAGAGAGAGAGAGAGGAAAGCCTCA-3', 5'-AGCTTGGCGTAATCATGGTCCTTAAGTTCGAAG-3' and 5'-TCGACTTCGA-ACTTAAGGACCATGA-TTACGCCA-3'. These self-complimentary oligonucleotide duplexes were phosphorylated with adenosine 5'-triphosphate and T4 polynucleotide kinase and then ligated into the large *Ava*I/*Sal*I restriction fragment of pUC19. *E. coli* XL-1 Blue competent cells were transformed by the ligated 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 Qiagen Maxi Prep kit. pSPHIV was prepared in a similar manner, except that the phosphorylation step was omitted and the following synthetic oligonucleotides were cloned into the large *Bam*HI/*Hind*III restriction fragment of pUC19, 5'-GATCTTAGCCACTTTTAAAGAAAAGGGGGGACTGGAAGGGC-TAATTA-3' and 5'-AGCTTAATTAGCCCTTCCAGTCCCCCCTTTCTTTTAAAAA-GTGGCTAA-3'.

3' End Labeling of Restriction Fragments. In a typical end labeling, 2 μ g of plasmid was digested with the appropriate restriction enzymes (*Afl*III/*Fsp*I for pYSPEC2 and *Eco*RI/*Pvu*II for pGCBGC and pSPHIV), phenol/ CHCl_3 extracted, and EtOH precipitated. The resulting fragments were 3' end labeled with [α - 32 P]-thymidine-5'-triphosphate and [α - 32 P]-deoxyadenosine-5'-triphosphate, using Sequenase (version 2.0). Excess unlabelled nucleotide triphosphates were added after labeling to insure complete fill in. Unincorporated nucleotides were removed on a gel filtration column (Microspin S-200 HR, Pharmacia), and the labelled restriction fragments were separated on a 5% nondenaturing polyacrylamide gel. The desired band was excised from the gel, crushed, and eluted overnight at 37° C into 25 mM trisHCl, 250 mM NaCl, 1 mM EDTA, pH 8.0. This solution was filtered, and the restriction fragment was precipitated with isopropyl alcohol. The labelled DNA was resuspended in TE, pH 7.5, phenol/ CHCl_3 extracted

three times, and EtOH precipitated with 2M NaCl. The purified restriction fragment was then resuspended in TE, pH 7.5 to an activity of approximately 30,000 cpm per μl . A typical yield from 2 μg of plasmid DNA was a total Cerenkov radioactivity of 3,000,000 cpm.

DNase I Footprint Titrations. These experiments were performed essentially according to the published protocol.⁸ Typically, a 10 nmol aliquot of the third strand oligonucleotide was resuspended in H_2O to a concentration of 100 μM , and this solution was serially diluted to a total of 18 concentrations spanning the desired range for the experiment. The minimum amount of solution transferred in these dilutions was 20 μl , in order to minimize pipetting error. A stock solution was prepared by combining 5X association buffer (50 mM bis-trisHCl, 500 mM NaCl, 1.25 mM spermine, previously adjusted to the appropriate pH with HCl), 50 μM base pairs calf thymus DNA, the labelled restriction fragment solution, and H_2O in appropriate amounts to give a total volume of 567 μl . This stock solution was then distributed in 27 μl aliquots among 20 labelled 1.7 ml microcentrifuge tubes. 18 μl of the appropriate oligonucleotide solution or H_2O (for reference lanes) was added to each tube. The tubes were then vortexed, centrifuged, and then allowed to equilibrate for either 3 (sequence composition experiments) or 5 (specificity experiments) days at 22° C (Final solution conditions: 100 mM NaCl, 10 mM bis-tris, 250 μM spermine, 5 μM bp calf thymus DNA, ~20,000 cpm labelled restriction fragment). After equilibration, 5 μl of a solution containing DNase I (at the appropriate concentration to give ~50% intact labelled DNA; this was determined by titration of the DNase activity), 50 mM CaCl_2 , 50 mM MgCl_2 , 10 μM nonspecific oligonucleotide (5'-d(CGACCGTCGA)-3'), 10 mM bis-trisHCl at the appropriate pH, and 5% glycerol was added, and digestion allowed to proceed for 6 minutes at room temperature. The DNase reactions were quenched by the addition of 8.3 μl of 1.4 M NaCl, 0.14 M EDTA, pH 8.0, and 0.35 $\mu\text{g}/\mu\text{l}$ glycogen, followed by 120 μl of EtOH. The

reactions were stored at -20° C for several hours and then centrifuged at 14000xg for 30 minutes at 0° C. The supernatant was decanted and the DNA pellet washed with cold 75% EtOH. The pellet was resuspended in 20 µl H₂O, frozen, and lyophilized to dryness. The pellet was then resuspended in 7 µl 80% formamide loading buffer containing 1X TBE. The samples were denatured at 85° C for 10 minutes, chilled on ice, and immediately loaded onto a pre-electrophoresed 8% denaturing polyacrylamide gel. The reaction products were separated by electrophoresis in 1X TBE buffer at 1800 V. The gel was dried for one hour at 80° C on a slab drier, and exposed to a storage phosphor screen (Molecular Dynamics) overnight at ambient temperature.

Quantitation and Data analysis. Data from the footprint titrations were obtained by quantitating the sequencing gels using ImageQuant software (Molecular Dynamics) running on an AST Premium 386/33 computer. Volume integration of rectangles encompassing the footprint site(s) and a reference site at which DNase reactivity was invariant across the titration generated values for site intensity (I_{site}) and reference intensity (I_{ref}). These I_{site} and I_{ref} values were corrected by subtracting out background intensity as described in the detailed protocol.⁸ The apparent fractional occupancy (θ_{app}) of the site at each oligonucleotide concentration ($[O]_{\text{tot}}$) was calculated from the corrected I_{site} and I_{ref} values using the following equation:

$$\theta_{\text{app}} = 1 - \frac{I_{\text{site}}/I_{\text{ref}}}{I_{\text{site}}^0/I_{\text{ref}}^0}$$

where I_{site}^0 and I_{ref}^0 are the site and reference intensities, respectively, from a control lane to which no third strand oligonucleotide has been added. The resulting pairs of (θ_{app} , $[O]_{\text{tot}}$) values were plotted on a semilog scale. The following binding isotherm was fit to the experimental data using a nonlinear least squares algorithm in the program

Kaleidagraph 3.0.1 running on a Macintosh IIfx or IIfx:

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

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. Data reported in the paper are the average of three determinations $\pm$ SEM. Each titration within a set of three measurements was prepared from a unique aliquot and serial dilution of oligonucleotide, a unique preparation of labelled restriction fragment, and unique batches of association buffer and calf thymus DNA.

The experimental conditions employed in this study are quite stringent, and in numerous cases, triple helix formation was not observed. The association constant is reported as $< 10^4$ in these cases, as this is the limit below which binding is not observed in this assay. For those cases in which a footprint is only observed at the highest oligonucleotide concentrations, and reliable curve fit could not be obtained, the association constant is reported as $< 5 \times 10^4$. In the specificity experiments, the triplet combinations with the highest binding affinities (T•AT and C+GC) were not at equilibrium even when the equilibration times were extended to 120 hours; these values are therefore reported as lower limits. The true association constants are probably 2-4 fold greater than the reported values.

References

- (1) (a) Moser, H. E.; Dervan, P. B. *Science* **1987**, *238*, 645-650. (b) For a review, see Thuong, N. T.; Helene, C. *Angew Chem Int. Ed. Engl.* **1993**, *32*, 666-690.
- (2) (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.
- (3) (a) Ono, A.; Ts'o, P. O. P.; Kan, L. *J. Am. Chem. Soc.* **1991**, *113*, 4032-4033. (b) Ono, A.; Ts'o, P. O. P.; Kan, L. *J. Org. Chem.* **1992**, *57*, 3225-3230.
- (4) (a) Young, S. L.; Krawczyk, S. H.; Matteucci, M. D.; Toole, J. J. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 10023-10026. (b) Krawczyk, S. H.; Milligan, J. F.; Wadwani, S.; Moulds, C.; Froehler, B. C.; Matteucci, M. D. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 3761-3764. (c) Miller, P. S.; Bahn, P.; Cushman, C. D.; Trapane, T. L. *Biochemistry* **1992**, *31*, 6788-6793. (d) Jetter, M. C.; Hobbs, F. W. *Biochemistry* **1993**, *32*, 3249-3254. (e) Davison, E. C.; Johnsson, K. *Nucleosides and Nucleotides* **1993**, *12*, 237-243.
- (5) (a) Koh, J. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 1470-1478. (b) note that nucleoside P is labelled P1 in references 5a and 7.
- (6) Xiang, G.; Soussou, W.; McLaughlin, L. W. *J. Am. Chem. Soc.* **1994**, in press.
- (7) Radhakrishnan, I.; Patel, D. J.; Priestley, E. S.; Nash, H. M.; Dervan, P. B. *Biochemistry* **1993**, *32*, 11228-11234.
- (8) Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Methods Enzymol.* **1986**, *130*, 132-181.
- (9) (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.
- (10) (a) Singleton, S. F.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 6957-6965. (b) Beal, P. A.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 4976-4982.

- (11) For previous reports of a similar four site assay, see: (a) Jones, R. J.; Lin, K.-Y.; Milligan, J. F.; Wadwani, S.; Matteucci, M. D. *J. Org. Chem.* **1993**, *58*, 2983-2991. (b) Jones, R. J.; Swaminathan, S.; Milligan, J. F.; Wadwani, S.; Froehler, B. C.; Matteucci, M. D. *J. Am. Chem. Soc.* **1993**, *115*, 9816-9817.
- (12) Singleton, S. F.; Dervan, P. B. *Biochemistry* **1993**, *32*, 13171-13179.
- (13) Singleton, S. F.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, in press.
- (14) For early qualitative work, see: (a) Griffin, L. C.; Dervan, P. B. *Science* **1989**, *245*, 967-971. (b) 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. (c) Mergny, J.-L.; Sun, J.-S.; Rougee, M.; Montenay-Garestier, T.; Barcelo, F.; Chomilier, J.; Helene, C. *Biochemistry* **1991**, *30*, 9791-9798. (d) Yoon, K.; Hobbs, C. A.; Koch, J.; Sardaro, M.; Kutny, R.; Weis, A. L. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 3840-3844. (e) Fossella, J. A.; Kim, Y. J.; Shih, H.; Richards, E. G.; Fresco, J. R. *Nucleic Acids Res.* **1993**, *21*, 4511-4515.
- (15) For a quantitative affinity cleavage titration study, see: Best, G. C.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, in press.
- (16) Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular Cloning*; Cold Spring Harbor Laboratory Press: Plainview, 1989.

CHAPTER FOUR

Triple Helix Formation by Oligonucleotides Containing Desmethyl P

Introduction

The nonnatural nucleoside P (Figure 1) has been shown to recognize GC base pairs within local triple helical complexes with high affinity and specificity (1,2). This base is particularly effective for recognition of contiguous GC sites, which are typically weakly bound by pyrimidine motif oligonucleotides containing 5-methyl-2'-deoxycytidine in the third strand (2). An NMR structural study has confirmed the hydrogen bonding pattern for the P•GC base triplet shown in Figure 1 (3).

Molecular modeling of a triple helical complex containing the P•GC base triplet flanked on both sides by T•AT base triplets suggests that a steric clash between the methyl group of the P base and H2'' of the third strand thymidine to the 5' side may destabilize the complex. A strong NOE is observed between these two protons in the previously mentioned NMR study, confirming their spatial proximity (3). Based on these observations, it was anticipated that oligonucleotides containing the desmethyl derivative of P (^DP, Figure 1) would recognize GC sites with higher affinity than corresponding oligonucleotides containing P, due to removal of this putative steric clash.

This chapter describes the synthesis and characterization of oligonucleotides containing ^DP. The ability of these oligonucleotides to bind their duplex target sites was compared to the corresponding P containing oligonucleotides in order to elucidate the effect of the methyl moiety. These experiments demonstrated that the removal of the methyl group produced only a small, negative effect on binding affinity, disproving the initial hypothesis that the methyl group is involved in an unfavorable steric clash.

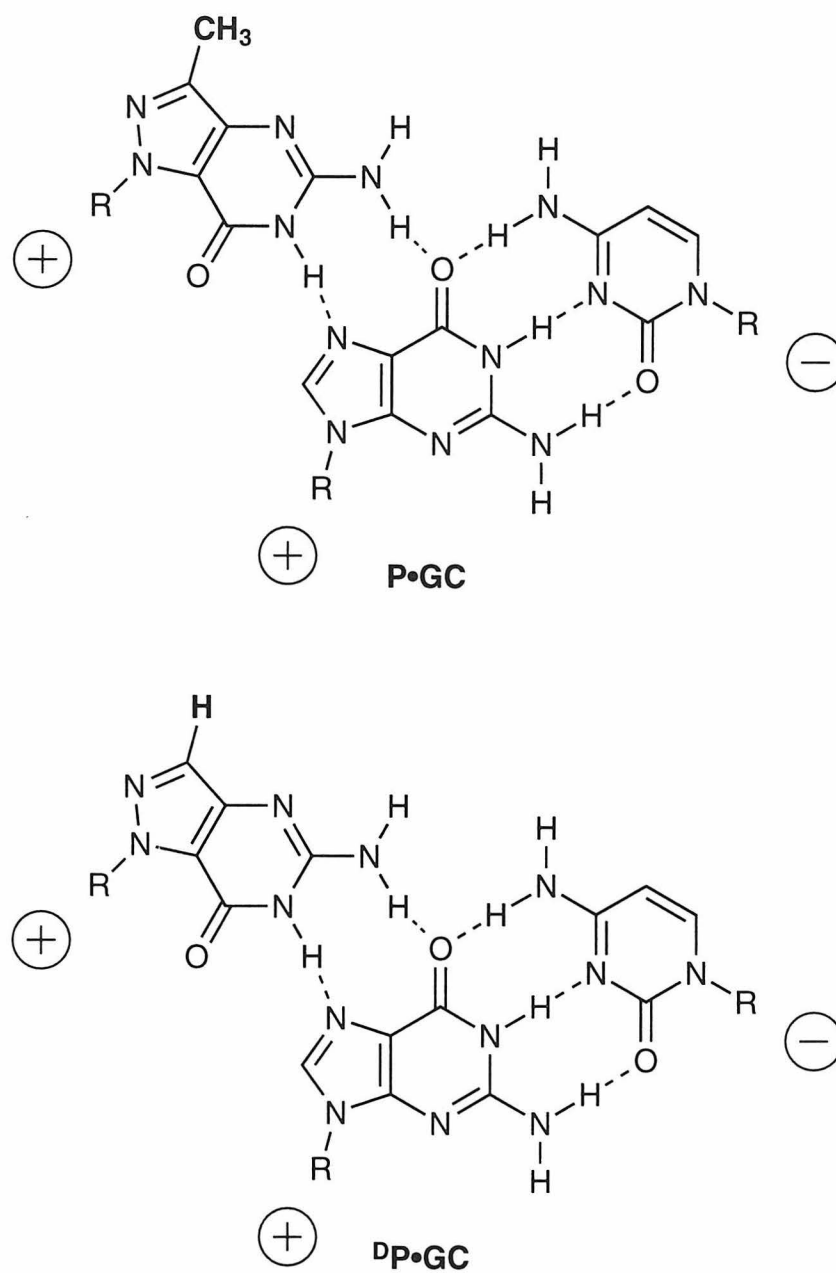


Figure 1: Two-dimensional representations of **P•GC** and **^DP•GC** base triplets. The methyl group is shown in boldface type. Plus and minus signs indicate relative backbone polarity.

Results and Discussion

Experimental Design. The association constants for triple helix formation by oligonucleotides containing ^DP were compared to those of corresponding oligonucleotides

containing P or 5-methyl-2'-deoxycytidine (^{Me}C) in the experiments shown in Figure 2. As detailed in Chapter Three, quantitative DNase footprinting titration was used to determine equilibrium association constants for triple helix formation by these oligonucleotides. The assay shown in Figure 2A was used to determine the effect removal of the P methyl group on specificity of this base. The oligonucleotides and target sequence (referred to as the (GA)₅ site) in Figure 2B were chosen in order to measure even small affects of the methyl group, since incorporation of multiple ^DP residues magnifies any incremental effect on binding affinity. In each case, the ^DP bases are flanked by thymidine residues. The experiments involving oligonucleotides containing P or ^{Me}C were described in Chapter Three, and are included here for comparison to the results for desmethyl P.

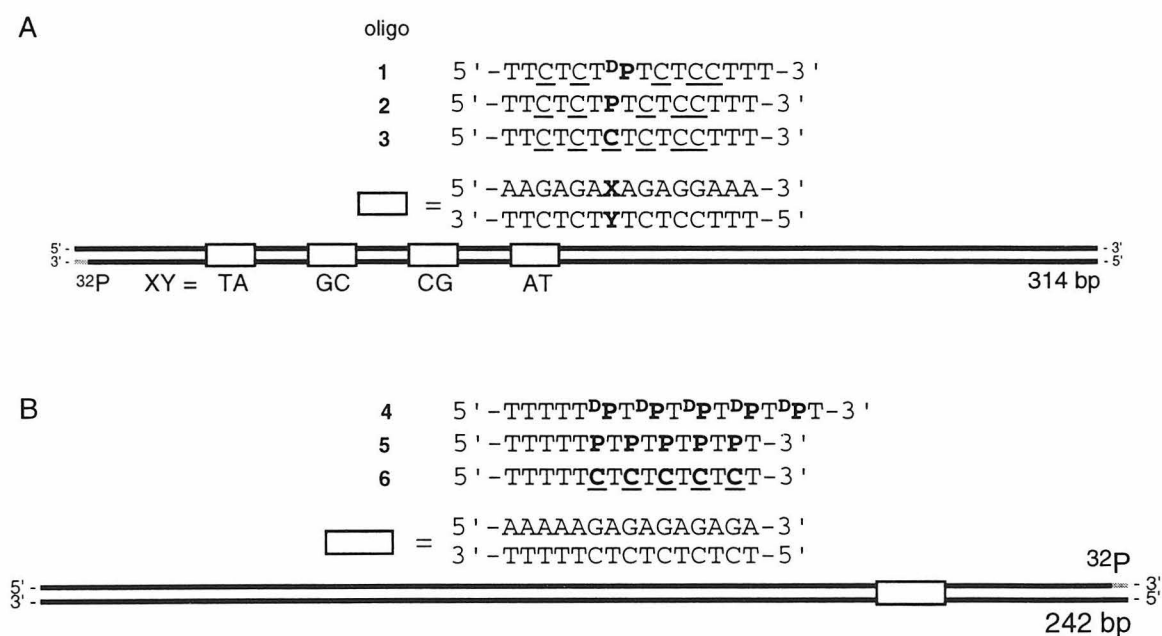
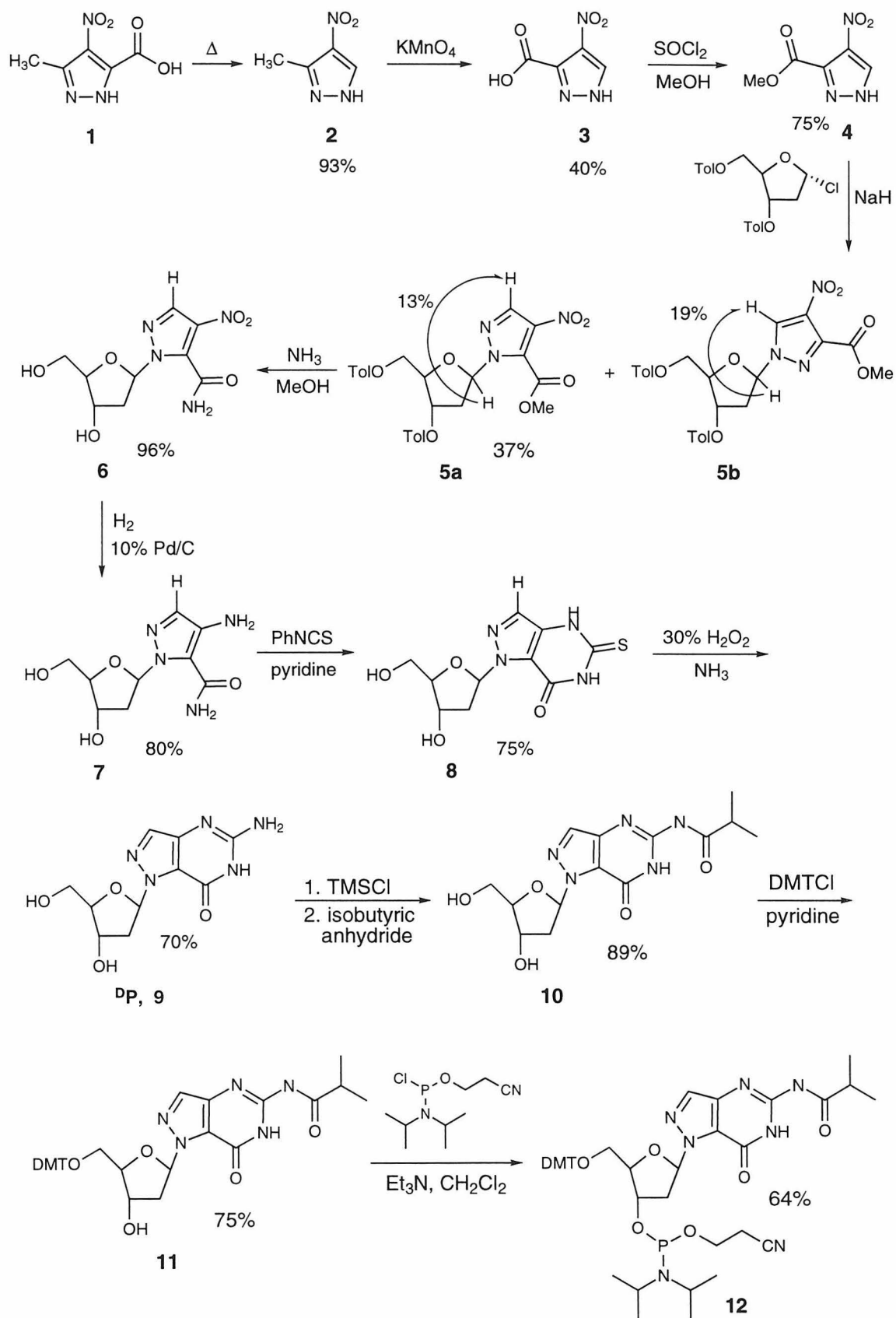


Figure 2: Sequences of third strand oligonucleotides **1-6** and duplex target sites for the desmethyl P experiments. (A) specificity assay using the 3'-labeled 314 bp *Afl* II/*Fsp* I restriction fragment of pYSPEC2. (B) multiple ^DP binding assay, using the 3'-labeled 242 bp *Eco* RI/*Pvu* II restriction fragment of pGCBGC.

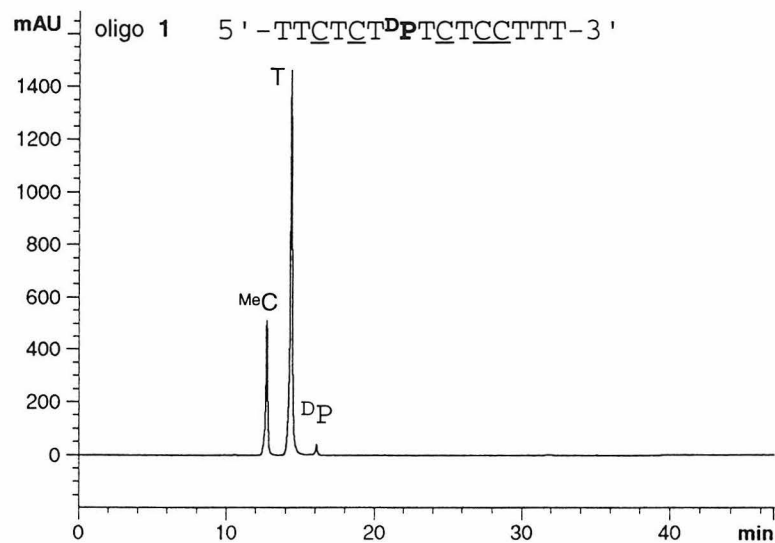
Synthesis. The synthesis of $^{\text{D}}\text{P}$ phosphoramidite (**12**) is shown in Figure 3. This route was originally developed by Dr. David Horne as a postdoctoral researcher in the Dervan group. Decarboxylation of the known acid **1** (**4**) provides pyrazole **2** in 93% yield. Subsequent permanganate oxidation generates the “desmethyl” acid **3** in moderate yield, which is then esterified to give pyrazole **4**. This compound can then be elaborated to desmethyl P, **9**, and its protected phosphoramidite **12** in analogy to the original synthesis of P (**1**). Coupling of the sodium salt of **4** to 1-chloro-2-deoxy-3,5-di-*O-p*-toluoyl- α -D-ribofuranose (**5,6**) in CH_3CN provides a 1:1 mixture of the two regioisomers **5a** and **5b**. The isomers were assigned by NMR experiments in which a 13% NOE was observed at the pyrazole H5 upon irradiation of the anomeric proton of **5a**, while a 19% NOE was observed at H5 upon irradiation of the anomeric proton of **5b**. The desired **5a** was converted to amide **6** by treatment with NH_3/MeOH . Hydrogenation over 10% Pd/C, followed by cyclization with phenylisothiocyanate afforded nucleoside **8**. Oxidation with hydrogen peroxide and aminolysis provided desmethyl P nucleoside **9**. Isobutyryl protection of **9** by the transient protection method affords nucleoside **10** (**7**). Subsequent tritylation and phosphitylation gives nucleoside **12**, suitable for automated DNA synthesis (**7**).

Oligonucleotides **1** and **4** were synthesized using standard solid phase β -cyanoethyl phosphoramidite chemistry (**7**). Phosphoramidite **12** coupled with >98% efficiency as a 0.11 M solution in CH_3CN . The base composition of the oligonucleotides **1** and **4** was verified by digestion with snake venom phosphodiesterase and alkaline phosphatase and subsequent reverse phase HPLC analysis. The results of this analysis are shown in Figure 4, along with coinjections with authentic standards of the component nucleosides.

Figure 3: Synthesis of desmethyl P phosphoramidite (**12**).



A.



B.

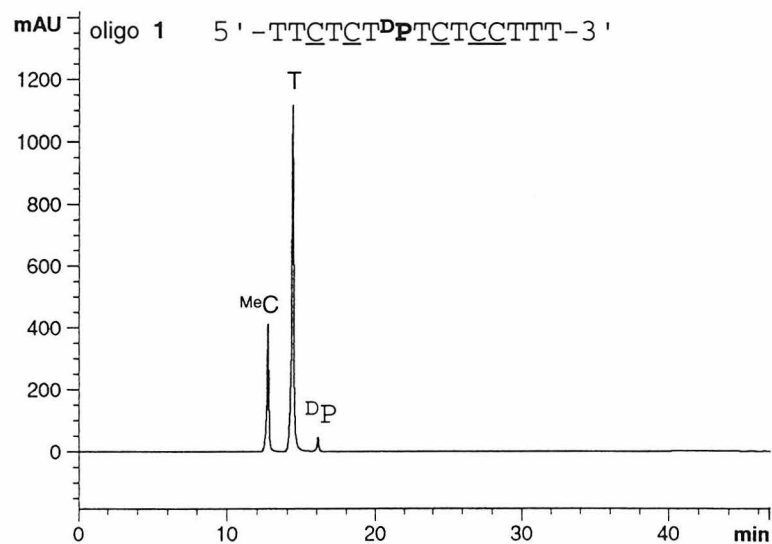
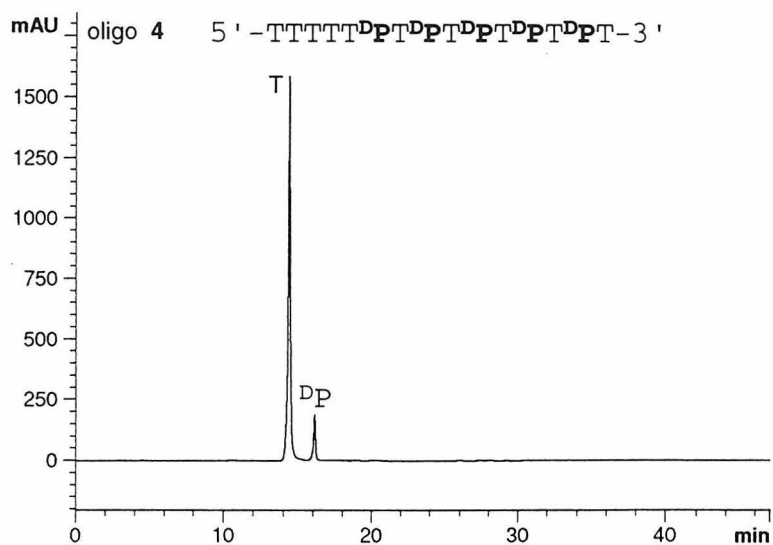


Figure 4: HPLC Traces monitored at 260 nm of A) enzymatically digested oligonucleotide **1** and B) a coinjection of the enzyme digest and the three component nucleosides. C indicates 2'-deoxy-5-methylcytidine.

C.



D.

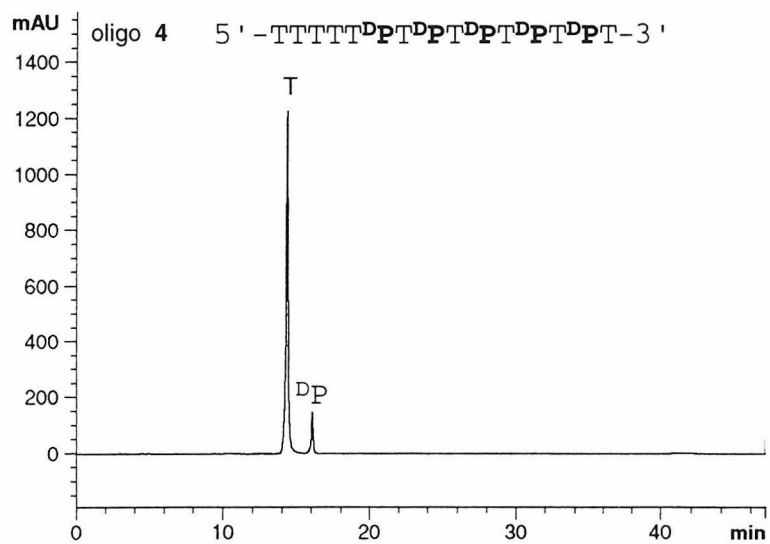


Figure 4: HPLC traces monitored at 260 nm of C) enzymatically digested oligonucleotide **4** and D) a coinjection of the enzyme digest and the two component nucleosides.

Quantitative DNase Footprint Titrations. Equilibrium association constants of oligonucleotides **1** and **4** for the duplex target sites shown in Figure 2 were determined by quantitative DNase footprint titration (8), as described in detail in Chapter Three. Solution conditions employed were 10 mM bis tris, 100 mM NaCl, 250 μ M spermine, pH 7.0 or 7.5. A sample footprinting gel for oligonucleotide **1** is shown in Figure 5, and average binding isotherms are shown in Figure 6.

The specificity experiments (Figure 6A, Table 1) demonstrate that oligonucleotide **1**, containing ^3P , retains 1000 fold specificity for GC base pairs. In fact, comparison of its binding affinity ($K_T = 3.9 \times 10^7 \text{ M}^{-1}$) for the GC site to that of oligonucleotide **2**, containing P indicates that the two are indistinguishable within experimental error (Figure 6B, Table 1). This experiment clearly shows that removal of the methyl group does not strongly influence the binding properties of oligonucleotide **1**.

The experiments on the $(\text{GA})_5$ site confirm this result. Oligonucleotide **4**, containing 5 ^3P residues, binds the target site only weakly ($K_T < 5 \times 10^4 \text{ M}^{-1}$) at both pH 7.0 and pH 7.5. Oligonucleotide **5**, containing P, binds with an association constant of $1 \times 10^5 \text{ M}^{-1}$ at both pH values. In this more sensitive assay, removal of five methyl groups from P bases reduces the binding affinity by a factor of 2 to 10 (the binding constant for oligonucleotide **4** was too low to be accurately measured), corresponding to 0.4 to 1.3 kcal/mol at 22 °C, or 0.1 to 0.3 kcal/mol/methyl group. These $\Delta\Delta G$ values are much smaller than would be expected from a strongly disfavoured steric interaction, which typically accounts for several kcal/mol.

Conclusion

In conclusion, the initial hypothesis of an unfavorable steric clash between the methyl group of P and the H2'' of its 5' neighboring thymidine in a triple helical complex

has been disproven. Quantitative DNase footprinting experiments demonstrate that removal of the methyl group of P is not beneficial, but rather exerts a small energetic price of 0.1 to 0.3 kcal/mol/methyl group. It should be noted that solvation and base stacking certainly play a role in the effect of the methyl group. It is also possible that the small difference in binding affinity obtained upon removal of the methyl group is due to changes in two large, but opposing effects.

Figure 5: Autoradiogram of an 8% denaturing polyacrylamide gel used to separate the fragments generated by DNase I digestion in a quantitative footprint titration experiment with oligonucleotide **1**. The boxes drawn to the left of the autoradiogram indicate the positions of the 15 bp binding sites within the 314 bp restriction fragment. (Lane 1) Products of an adenine specific sequencing reaction. (Lanes 2-20) DNase I digestion products obtained in the presence of varying concentrations of oligonucleotide **1**: no oligonucleotide (lane 2); 80 pM (lane 3); 200 pM (lane 4); 400 pM (lane 5); 800 pM (lane 6); 2 nM (lane 7); 4 nM (lane 8); 8 nM (lane 9); 20 nM (lane 10); 40 nM (lane 11); 80 nM (lane 12); 200 nM (lane 13); 400 nM (lane 14); 800 nM (lane 15); 2 μ M (lane 16); 4 μ M (lane 17); 8 μ M (lane 18); 20 μ M (lane 19); 40 μ M (lane 20). (Lane 21) Intact 3' labeled DNA after incubation in the absence of the third strand oligonucleotide.

5' - TTCTCT^DPTCTCCTTT - 3'

5' - AAGAGAXAGAGGAAA - 3'

3' - TTCTCTYTCTCCTTT - 5'

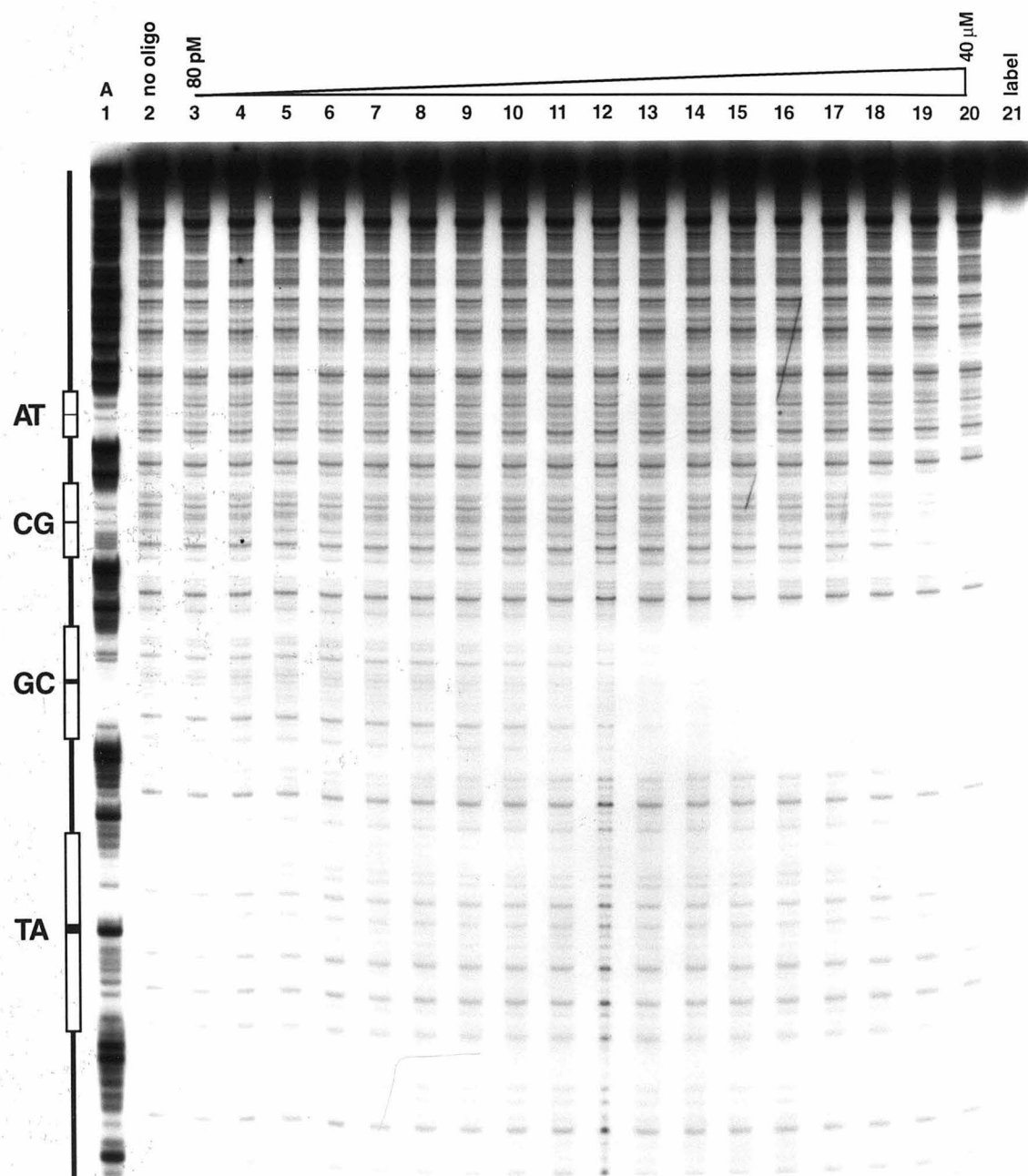


Figure 6: Plots of the site occupancy (θ) versus oligonucleotide concentration for binding of oligonucleotides **1-3** to the target sites on the 314 bp *Afl* II/*Fsp* I restriction fragment of pYSPEC2. A) specificity of oligo **1**, containing ^{32}P at the variable position. B) comparison of binding of oligonucleotides **1-3** to the GC target site. Each data set is averaged from three independent gels. Binding isotherms derived from the data are shown as solid lines.

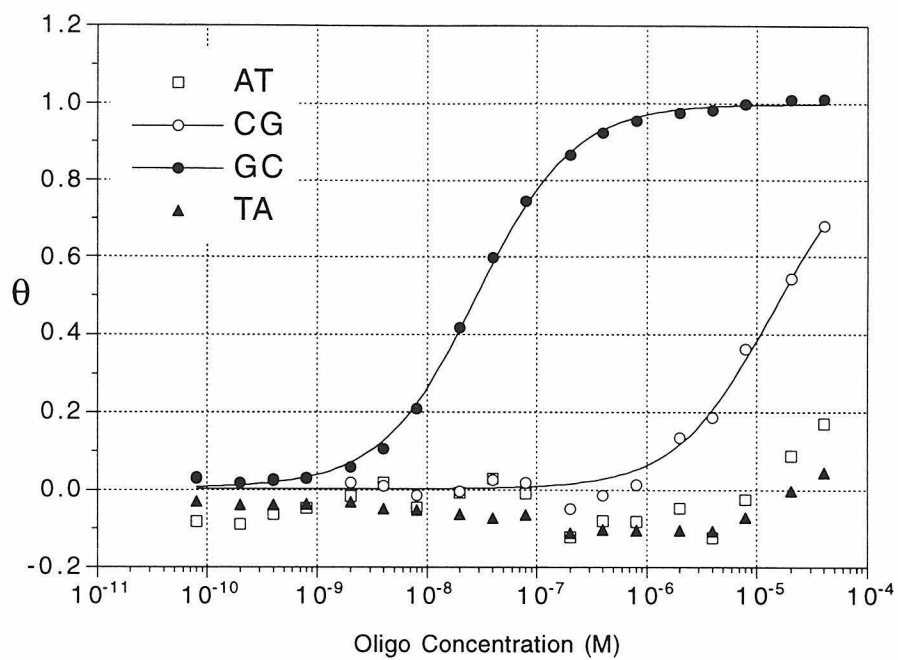
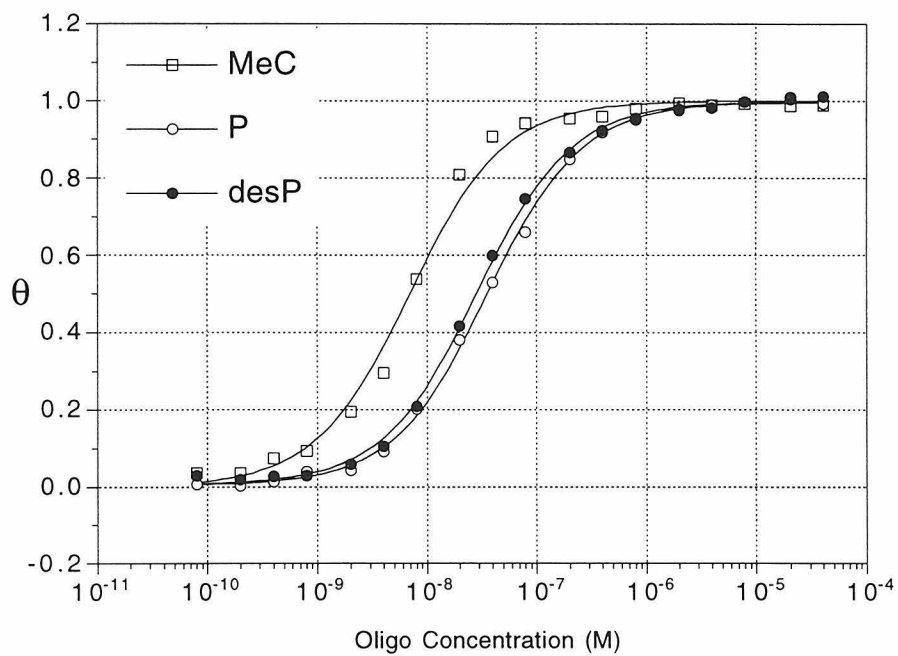
A. $^{\text{D}}\text{P} \cdot \text{XY}$ **B. GC Binding Affinity**

Table I: Equilibrium Association Constants for Oligonucleotides Containing Single $^{\text{D}}\text{P}$, P , or $^{\text{Me}}\text{C}$ Nucleosides^a

oligo	Z ^b	XY	K _T	ΔG° ^c
1	$^{\text{D}}\text{P}$	AT	$< 10^4$	> -5.4
		CG	$< 5 \times 10^4$	> -6.3
		GC	$3.9 (\pm 1.0) \times 10^7$	$-10.2 (\pm 0.1)$
		TA	$< 10^4$	> -5.4
2	P	AT	$< 10^4$	> -5.4
		CG	$< 5 \times 10^4$	> -6.3
		GC	$3.1 (\pm 1.0) \times 10^7$	$-10.1 (\pm 0.2)$
		TA	$< 10^4$	> -5.4
3	$^{\text{Me}}\text{C}$	AT	$8.2 (\pm 1.9) \times 10^4$	$-6.6 (\pm 0.1)$
		CG	$1.7 (\pm 0.6) \times 10^5$	$-7.1 (\pm 0.2)$
		GC	$\geq 1.4 \times 10^8$	< -11.0
		TA	$< 10^4$	> -5.4

^a Each reported K_T value is the mean ($\pm$ SEM) of three independent measurements which were performed in 100 mM NaCl, 10 mM bis-tris, 250 μM spermine at pH 7.0 and 22° C. Values for K_T and ΔG° are reported in units of M^{-1} and $\text{kcal}\cdot\text{mol}^{-1}$, respectively. Data for oligonucleotides **2** and **3** is taken from Chapter Three. ^b Z refers to the position in the third strand oligonucleotide opposite the variable position (XY) in the duplex. ^c ΔG° values were obtained from the relationship $\Delta\text{G}^{\circ} = -RT\ln K$ ($T = 295 \text{ K}$) and error estimates were obtained by propagating the SEM for K_T through this formula.

Table II: Equilibrium Association Constants for Triple Helix Formation at the (GA)₅ Site^a

oligo	pH	K _T (M ⁻¹)	ΔG° ^b
4	7.0	< 5 × 10 ⁴	> -6.3
	7.5	< 5 × 10 ⁴	> -6.3
5	7.0	1.1 (±0.2) × 10 ⁵	-6.8 (±0.1)
	7.5	1.3 (±0.2) × 10 ⁵	-6.9 (±0.1)
6	7.0	3.1 (±0.2) × 10 ⁹	-12.8 (±0.1)
	7.5	9.4 (±1.6) × 10 ⁷	-10.8 (±0.1)

^a Each reported K_T value is the mean (±SEM) of three independent measurements which were performed in 100 mM NaCl, 10 mM bis-tris, 250 μM spermine at the indicated pH and 22° C. Data for oligonucleotides **5** and **6** were taken from Chapter Three. ^b ΔG° values were obtained from the relationship ΔG° = -RTlnK (T = 295 K) and error estimates were obtained by propagating the SEM for K_T through this formula.

Experimental

General. Reagents for organic synthesis were obtained from Aldrich Chemical Co. and used as received. Anhydrous solvents over molecular sieves were purchased from Fluka, with the exception of pyridine, which was distilled over CaH_2 . 5-methyl-2'-deoxycytidine phosphoramidite was purchased from Cruachem. All other DNA synthesis reagents were provided by Glen Research. NMR spectra were obtained on a GE 300 Mhz spectrometer and referenced to the residual solvent signal. The NOE spectra were obtained in degassed CDCl_3 . Mass spectral analyses were obtained at the University of California, Riverside, Mass Spectrometry Laboratory. Flash chromatography was carried out using EM Science Kieselgel 60 (230-400 mesh). Preparative oligonucleotide separations were carried out on a Pharmacia FPLC system as described below. Analytical HPLC was performed on a Hewlett Packard 1090 chromatograph. Aqueous solutions for DNA manipulations were prepared with water purified by a Milli-Q Water System (Millipore).

3-Methyl-4-nitropyrazole (2). 3-methyl-4-nitropyrazole-5-carboxylic acid (**1**) (**4**) (6.9 g, 40.8 mmol) was heated to 170 °C under aspirator vacuum until the melt ceased bubbling (approximately 6 hrs). Cooling afforded 4.84 g (93%) of a crystalline solid (**2**). ^1H NMR (300MHz, CD_3OD) δ 8.21 (s, 1H), 2.56 (s, 3H); HRMS (EI) for $\text{C}_4\text{H}_5\text{O}_2\text{N}_3$ calcd 127.0382, found 127.0373.

4-nitropyrazole-3-carboxylic acid (3). To a stirred suspension of pyrazole **2** (4.80 g, 37.8 mmol) in H_2O was added KMnO_4 (23.9 g, 151 mmol) in one portion. The mixture was stirred at reflux for 5 hrs and then quenched by addition of 3 ml isopropyl alcohol. Charcoal was added and the hot reaction mixture was then filtered through Celite. Concentration and acidification afforded a yellow solid, which was extracted repeatedly

with MeOH. The MeOH soluble material was filtered and concentrated to yield 2.37 g (40%) of **3** as a white solid. HRMS (EI) for $C_4H_3N_3O_4$ calc 157.0124, found 157.0130.

Methyl-4-nitropyrazole-3-carboxylate (4). $SOCl_2$ (11.0 ml, 151 mmol) was added dropwise to a stirred solution of acid **3** (2.37 g, 15.1 mmol) in 100 ml MeOH at $-10\text{ }^{\circ}\text{C}$. The reaction was allowed to warm to room temperature, and then stirred at $50\text{ }^{\circ}\text{C}$ overnight. Filtration, concentration, and subsequent flash chromatography (4% MeOH in dichloromethane) provided 1.93 g (75%) of **4** as a white solid. ^1H NMR δ 8.52 (s, 1H), 4.06 (s, 3H).

Methyl 2-[2-Deoxy-3,5-bis-*O*-(4-methylbenzoyl)- β -D-ribofuranosyl]-4-nitro-1*H*-pyrazole-3-carboxylate (5a) and Methyl 1-[2-Deoxy-3,5-bis-*O*-(4-methylbenzoyl)- β -D-ribofuranosyl]-4-nitro-1*H*-pyrazole-3-carboxylate (5b). To a stirred suspension of sodium hydride (60% dispersion in oil, 0.108 g, 2.7 mmol) in dry CH_3CN (20 ml) under Ar was added methyl-4-nitropyrazole-3-carboxylate (**4**) (0.40 g, 2.3 mmol) dissolved in CH_3CN (10 ml). After 1 hr 40 min, 1-chloro-2-deoxy-3,5-di-*O-p*-toluoyl- α -D-ribofuranose (**5**) (1.00 g, 2.6 mmol) was added in one portion, and the reaction mixture stirred for an additional 1 hr 20 min. The reaction mixture was partitioned between dichloromethane and water, and the organic phase was subsequently dried, filtered, and concentrated. Flash chromatography (0-5% ether in dichloromethane) of the crude orange oil gave 0.45 g (37%) of **5a** and 0.70 g of a mixture of **5a** and **5b**. Isomer **5a** TLC R_f = 0.13 in 20% hexanes in ethyl acetate; ^1H NMR (300 MHz, $CDCl_3$) δ 7.99 (s, 1H), 7.93 (d, J = 8.25 Hz, 2H), 7.86 (d, J = 8.25 Hz, 2H), 7.26 (d, J = 8.12 Hz, 2H), 7.20 (d, J = 8.32 Hz, 2H), 6.46 (d x d, J = 6.77, 4.70 Hz, 1H), 5.83 (m, 1H), 4.61 (m, 2H), 4.42 (m 1H), 4.01 (s, 3H), 3.47 (m, 1H), 2.72 (m, 1H), 2.43 (s, 3H), 2.40 (s, 3H); HRMS (FAB) for $C_{26}H_{26}N_3O$ ($M + H$) calcd 524.1669, found 524.1672.

2-(2-Deoxy- β -D-ribofuranosyl)-4-nitro-1*H*-pyrazole-3-carboxamide (6). Methyl ester **5a** (0.45 g, 0.85 mmol) was heated in NH_3 saturated MeOH (50 ml) in a Parr bomb at 80 °C for 18 hrs. Cooling, concentration, and flash chromatography (10% MeOH in dichloromethane) yielded 0.22 g (96%) **6** as a colorless oil. ^1H NMR (300 MHz, $\text{DMSO}-d_6$) δ 8.53 (s, 1H), 8.36 (s, 1H), 8.35 (s, 1H), 6.09 (t, $J = 6.09$ Hz, 1H), 5.32 (d, $J = 4.6$ Hz, 1H), 4.69 (t, $J = 5.70$ Hz, 1H), 4.37 (m, 1H), 3.82 (m, 1H), 3.44 (m, 1H), 3.37 (m, 1H), 2.68 (m, 1H), 2.26 (m, 1H); HRMS (FAB) for $\text{C}_9\text{H}_{12}\text{N}_4\text{O}_6\text{Na}$ ($\text{M} + \text{Na}$) calcd 295.0652, found 295.0654.

2-(2-Deoxy- β -D-ribofuranosyl)-4-amino-1*H*-pyrazole-3-carboxamide (7). A solution of nitro compound **6** (0.22 g, 0.82 mmol) in EtOH/ H_2O , (8:2, v/v; 40 ml) was shaken with 10% Pd/C (32 mg) under H_2 (50 psi) for 5 hrs at room temperature. Filtration and concentration afforded 0.159 g (80%) of **7** as a yellow oil. ^1H NMR (300 MHz, $\text{DMSO}-d_6$) δ 7.43 (br s, 2H), 7.13 (s, 1H), 6.65 (t, $J = 6.14$, 1H), 5.16 (d, 1H), 4.70 (t, $J = 5.44$, 1H), 4.54 (s, 2H), 4.31 (d, 1H), 3.71 (m, 1H), 3.38 (m, 1H), 3.24 (m, 1H), 2.76 (m, 1H), 2.09 (m, 1H); HRMS (DCI) for $\text{C}_9\text{H}_{15}\text{N}_4\text{O}_4$ ($\text{M} + \text{H}$) calcd 243.1093, found 243.1080.

1-(2-Deoxy- β -D-ribofuranosyl)-5-mercapto-1*H*-pyrazolo[4,3-*d*]pyrimidin-7-one (8). Nucleoside **7** (0.9 g, 3.7 mmol) and phenyl isothiocyanate (0.49 ml, 4.1 mmol) were dissolved in dry pyridine (20 ml) and heated at reflux for 2 hrs. Concentration followed by flash chromatography (0-15% MeOH in dichloromethane) provided 0.79 g (75%) of **8** as a white solid. ^1H NMR (300 MHz, $\text{DMSO}-d_6$) δ 12.8 (br s, 1H), 12.47 (br s, 1H), 7.59 (s, 1H), 6.73 (t, $J = 6.12$ Hz, 1H), 5.27 (d, $J = 4.27$ Hz, 1H), 4.66 (t, $J = 5.64$ Hz, 1H), 4.38 (m, 1H), 3.78 (m, 1H), 3.43 (m, 1H), 3.28 (m, 1H), 2.73 (m, 1H), 2.24 (m, 1H); HRMS (DCI) for $\text{C}_{10}\text{H}_{13}\text{N}_4\text{O}_4\text{S}$ ($\text{M} + \text{H}$) calcd 285.0658, found 285.0651.

1-(2-Deoxy- β -D-ribofuranosyl)-5-amino-1*H*-pyrazolo[4,3-*d*]pyrimidin-7-one (9,^DP).

To a solution of nucleoside **8** (0.54 g, 1.9 mmol) in concentrated NH_4OH (5.2 ml) at 0 °C was added 30% hydrogen peroxide (0.98 ml, 9.6 mmol) dropwise. After stirring 50 min, the solution was saturated with NH_3 , transferred to a sealed tube, and heated at 120 °C for 15 hrs. Cooling, concentration, and flash chromatography (5-10% H_2O in CH_3CN) afforded 0.36 g (70%) **8** as a white solid. ^1H NMR (300 MHz, $\text{DMSO}-d_6$) **8** 11.09 (br s, 1H), 7.63 (s, 1H), 6.76 (t, $J = 6.23$ Hz, 1H), 6.13 (br s, 2H), 5.23 (d, $J = 4.35$ Hz, 1H), 4.68 (t, $J = 5.74$ Hz, 1H), 4.38 (m, 1H), 3.76 (m, 1H), 3.37 (m, 2H), 2.75 (m, 1H), 2.19 (m, 1H); HRMS (FAB) for $\text{C}_{10}\text{H}_{14}\text{N}_5\text{O}_4$ ($M + H$) calcd 268.1046, found 268.1058.

***N*-[1-(2-Deoxy- β -D-ribofuranosyl)-7-oxo-1*H*-pyrazolo[4,3-*d*]pyridin-5-yl]-2-methylpropanamide (10).** To a stirred solution of nucleoside **9** (0.177 g, 0.66 mmol) in dry pyridine (8 ml) was added chlorotrimethylsilane (0.42 ml, 3.3 mmol) dropwise, and the reaction stirred for 2 hrs. The reaction was cooled to 0 °C and quenched with concentrated NH_4OH (0.84 ml). Concentration followed by flash chromatography (5-7% MeOH in CH_2Cl_2) provided 0.198 g (89%) of **10** as a colorless solid. ^1H NMR (300 MHz, CD_3OD) δ 7.83 (s, 1H), 7.00 (m, 1H), 4.61 (m, 1H), 3.98 (m, 1H), 3.71 (m, 1H), 3.56 (m, 1H), 2.93 (m, 1H), 2.67 (q, $J = 6.87$ Hz, 1H), 1.21 (d, $J = 6.84$ Hz, 6H); HRMS (DCI) for $\text{C}_{14}\text{H}_{20}\text{N}_5\text{O}_5$ ($M + H$) calcd 338.1464, found 338.1472.

***N*-[1-[5-*O*-[Bis(4-methoxyphenyl)phenylmethyl]-2-deoxy- β -D-ribofuranosyl]-7-oxo-1*H*-pyrazolo[4,3-*d*]pyrimidin-5-yl]-2-methylpropanamide (11).** Triethylamine (0.10 ml, 0.72 mmol) and dry pyridine (5 ml) were added to a mixture of nucleoside **10** (0.183 g, 0.54 mmol), 4-4'-dimethoxytrityl chloride (0.214 g, 0.6 mmol), and 4-dimethylaminepyridine (3.1 mg, 0.02 mmol) under Ar at 0 °C. The reaction was stirred overnight at 4 °C. An additional portion of DMTCl (0.061 g, 0.15 mmol) was added and the mixture stirred an additional 4 hrs. The reaction was quenched with MeOH (2 ml), concentrated, and chromatographed (66:33:1 CH_2Cl_2 /ethyl acetate/triethylamine to 2:1:97

MeOH/triethylamine/ethyl acetate) to give 0.259 g (75%) of **11** as a foam. ^1H NMR (300 MHz, CDCl_3) δ 8.2 (br s, 1H), 7.66 (s, 1H), 7.38 (m, 2H), 7.24 (m, 7H), 6.98 (m, 1H), 6.78 (m, 4H), 4.82 (m, 1H), 4.01 (m, 1H), 3.78 (s, 6H), 3.30 (m, 1H), 3.18 (m, 1H), 3.03 (m, 1H), 2.60 (m, 1H), 2.43 (m, 1H), 1.27 (d, $J = 6.93$ Hz, 6H); HRMS (FAB) for $\text{C}_{35}\text{H}_{37}\text{H}_5\text{O}_7$ calcd 639.2693, found 639.2703.

Phosphoramidite of DMT-Protected $^{\text{D}}\text{P}$ (12**).** 2-cyanoethyl N,N-

diisopropylchlorophosphoramidite (0.060 ml, 0.27 mmol) was added dropwise to a solution of nucleoside **11** (0.111 g, 0.17 mmol) and triethylamine (0.12 ml, 0.86 mmol) in dry CH_2Cl_2 (3.0 ml) under Ar. After 2 hrs, the reaction was quenched by addition of MeOH (1ml) and diluted with ethyl acetate. The reaction mixture was extracted twice with saturated NaHCO_3 and twice with brine, dried (Na_2SO_4), and concentrated. The residue was chromatographed (60:39:1 to 75:24:1 ethyl acetate/hexane/triethylamine) to afford 93 mg (64%) of **12** as a foam. HRMS (FAB) for $\text{C}_{44}\text{H}_{55}\text{N}_7\text{O}_8\text{P}$ ($\text{M} + \text{H}$) calcd 840.3850, found 840.3876.

Synthesis, Purification, and Analysis of Oligonucleotides. Oligonucleotides **1** and **4** were synthesized on a 1.0 μmol scale with an Applied Biosystems Model 380B DNA Synthesizer running standard automated synthesis protocols. $^{\text{D}}\text{P}$ Phosphoramidite (**12**) coupled with >98% efficiency as a 0.11 M solution in dry CH_3CN . These oligonucleotides were deprotected and purified exactly as described for oligonucleotide **1** in Chapter Three of this thesis. The purified oligonucleotides were quantitated by UV absorbance using extinction coefficients of $2800 \text{ M}^{-1}\text{cm}^{-1}$ for $^{\text{D}}\text{P}$, $8800 \text{ M}^{-1}\text{cm}^{-1}$ for T, and $5700 \text{ M}^{-1}\text{cm}^{-1}$ for $^{\text{Me}}\text{C}$. Enzymatic digestion and HPLC analysis of the oligonucleotides was performed as described in the experimental section of Chapter Three.

DNase Footprinting Experiments. The quantitative DNase footprint titrations were

performed and analysed as described in the experimental section of Chapter Three.

References

- (1) Koh, J. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 1470-1478.
- (2) Priestley, E. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1995**, *117*, 4761-4765.
- (3) Radhakrishnan, I.; Patel, D. J.; Priestley, E. S.; Nash, H. M.; Dervan, P. B. *Biochemistry* **1993**, *32*, 11228-11234.
- (4) Lewis, A. F.; Townsend, L. B. In *Nucleic Acid Chemistry*; Townsend, L. B., Tipson, S., Eds.; Wiley: New York, 1978; Vol. 1, p121, and references cited therein.
- (5) Hoffer, M. *Chem. Ber.* **1960**, *93*, 2777.
- (6) Kazimierczuk, Z., et al. *J. Am. Chem. Soc.* **1984**, *106*, 6379-6382.
- (7) *Oligonucleotide Synthesis*; Gait, K. J., Ed.; IRL Press: Oxford, UK, 1984 and references cited therein.
- (8) Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Methods Enzymol.* **1986**, *130*, 132-181.

CHAPTER FIVE

Triple Helix Formation by an Oligonucleotide Containing the 2-Pyrimidinone Base

Introduction

Triple helix formation by oligonucleotides containing the nonnatural nucleoside 1-(2'-deoxy- β -D-*erythro*-pentofuranosyl)-2-pyrimidinone (**1**) has been previously examined by affinity cleavage (1). A 15-mer oligonucleotide containing **1** at a single internal position was shown to bind duplex target sequences containing CG juxtaposed to the novel heterocycle with higher affinity than sequences containing any of the other three base pairs at this position. However, the third strand oligonucleotide was shown to be substantially inhomogeneous by enzymatic digestion, probably due to instability of **1** to the concentrated NH_3 solution used for oligonucleotide deprotection (2).

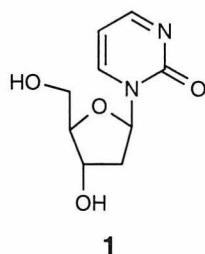


Figure 1: Structure of 1-(2'-deoxy- β -D-*erythro*-pentofuranosyl)-2-pyrimidinone.

After the development of the quantitative assay for the specificity of novel nucleosides in triple helix formation described in Chapter Three, we decided to reexamine **1**, since this compound was our best lead for CG recognition. Figure 2 shows two-dimensional models of the base triplets formed from 2-pyrimidinone and the four base pairs. We anticipated that the 2-pyrimidinone base should be capable of accepting a single hydrogen bond from the exocyclic amino group of cytosine with either its N3 (as shown in Figure 2) or O2 lone pairs (or both, forming a bifurcated hydrogen bond), forming a 2-pyrimidinone•CG triplet. CPK molecular modeling indicates that the O2 carbonyl of 2-pyrimidinone should prevent binding to AT base pairs because of prohibitive electrostatic/steric interactions with the N7 of adenosine. Moreover, 2-pyrimidinone contains no hydrogen bond donating groups, and thus should be incapable

of forming stable triplets with GC and TA base pairs. These models suggest that 2-pyrimidinone may be capable of specific recognition of CG base pairs and are consistent with the previously mentioned affinity cleaving results (1).

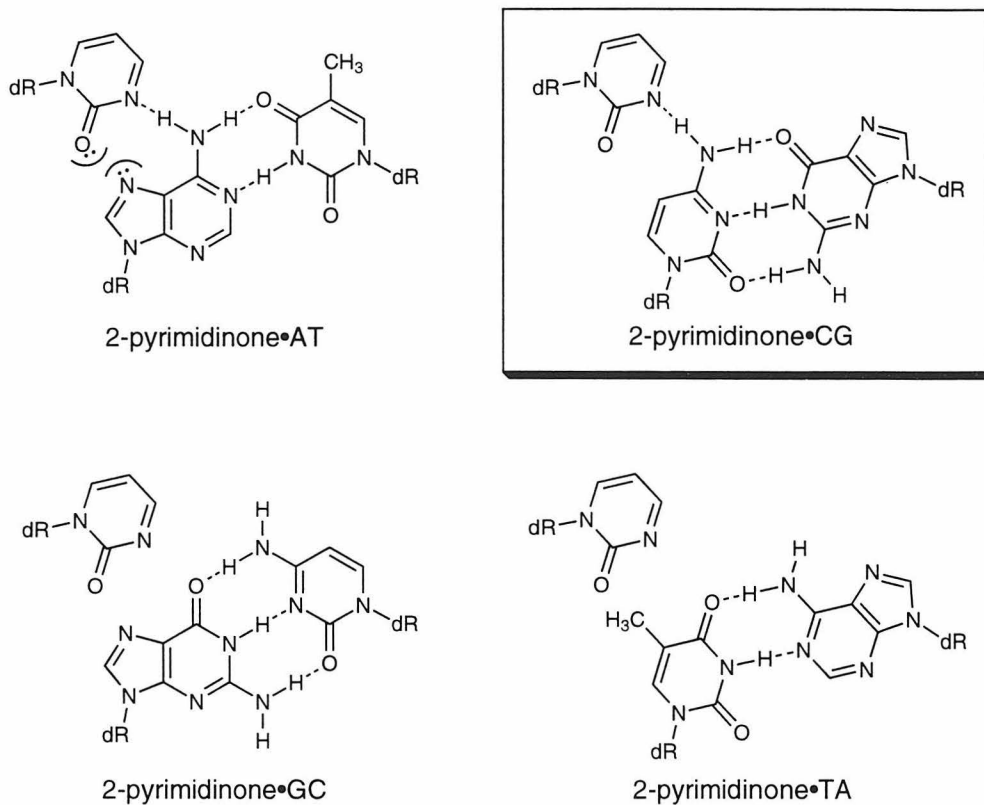


Figure 2: Schematic representations of the four 2-pyrimidinone base triplets.

Two other features which render 2-pyrimidinone attractive as a novel base are the following. The 2-pyrimidinone•CG base triplet is almost isomorphous with the T•AT triplet, as shown in Figure 3, so incorporation of the 2-pyrimidinone•CG triplet should introduce minimal backbone distortion in a triple helical complex. Secondly, N3 of the 2-pyrimidinone heterocycle will be unprotonated at neutral pH and available as a hydrogen bond acceptor due to its low pKa (the pKa of the model compound *N*-methyl-2-pyrimidinone is 2.5 (3)).

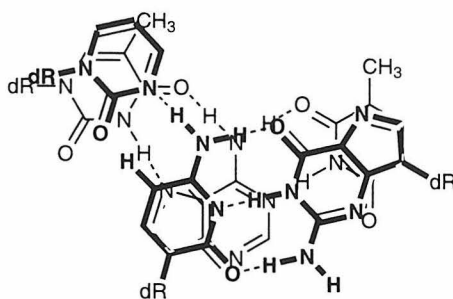


Figure 3: Overlay of a 2-pyrimidinone•CG triplet (bold) on a T•AT triplet. Note that the two third strand sugars (upper left, indicated by dR) are oriented similarly. More accurate CPK model building gives essentially the same result.

Results and Discussion

Experimental Design. The design of the 2-pyrimidinone specificity experiments is shown in Figure 4. Oligonucleotide **9**, containing a single 2-pyrimidinone base, can form a pyrimidine motif triple helix with each of the four target sites on the 314 bp *Afl* II/*Fsp* I restriction fragment of pYSPEC2. These four triple helical complexes collectively juxtapose all four base pairs opposite the 2-pyrimidinone heterocycle; comparison of their equilibrium association constants provides a quantitative measure of the specificity of 2-pyrimidinone in triple helix formation.

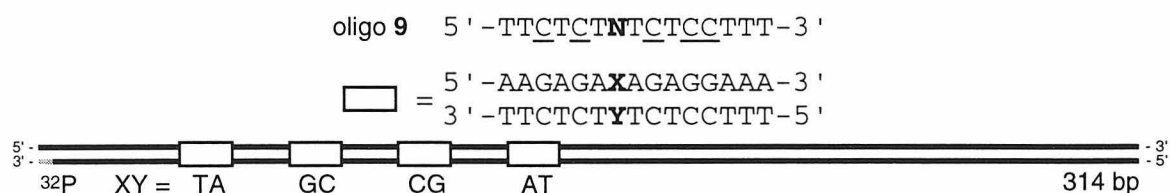


Figure 4: 2-pyrimidinone specificity experiments. Oligonucleotide **9** contains thymidine, 2'-deoxy-5-methylcytidine (**C**), and 1-(2'-deoxy- β -D-erythro-pentofuranosyl)-2-pyrimidinone (**N**). The 3'-labeled duplex is the 314 bp *Afl* II/*Fsp* I restriction fragment of pYSPEC2.

Synthesis. The phosphoramidite of 2-pyrimidinone deoxyribose was synthesized from 2'-deoxyuridine (**2**) by the previously described route shown in Figure 5 (4). Oligonucleotide **9** was synthesized by standard automated methods, (5) using

phosphoramidite **8** to incorporate 2-pyrimidinone deoxyriboside. Since oligonucleotides containing 2-pyrimidinone deoxyriboside have been reported to be labile to standard NH_3 deprotection at 50°C (due to addition of NH_3 across the 5,6-double bond and subsequent base loss) (2), the oligonucleotide was deprotected in two aliquots, at 50°C for 5.5 hrs and at room temperature for 21.5 hrs. Both samples were purified by ion exchange and reverse phase chromatography.

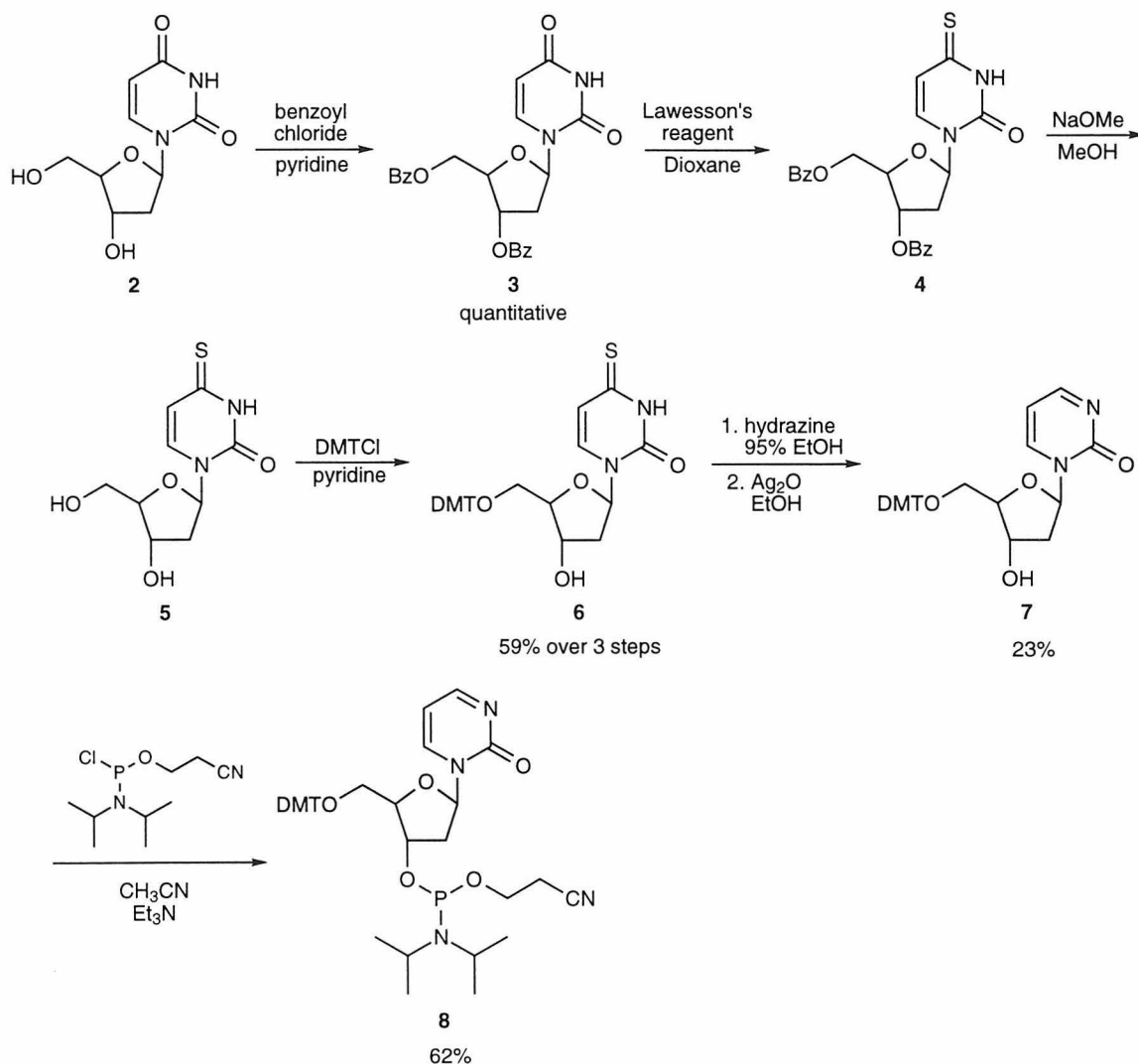


Figure 5: Synthesis of the 5'-DMT protected phosphoramidite of 2-pyrimidinone deoxyriboside (**4**). Yields shown are those obtained in this work.

Analysis of Oligonucleotide Purity. Samples of oligonucleotide **9** were analyzed by 5'-end labeling followed by polyacrylamide gel electrophoresis, by reverse phase HPLC of intact and enzymatically digested oligonucleotide, and by matrix assisted laser desorption/ionization time of flight (MALDI/TOF) mass spectroscopy. The results of the end labeling/gel analysis (Figure 6) show that all samples are homogeneous with respect to oligonucleotide length. The HPLC analysis (Figure 7) is more discriminating, demonstrating that the deprotection at 50°C gives an inhomogeneous sample, while room temperature deprotection gives a relatively pure oligonucleotide. Analysis of enzymatically digested samples (Figure 8) shows the same results. Since the room temperature deprotection protocol generated pure oligonucleotide **9**, this sample was further analyzed by MALDI/TOF mass spectroscopy, giving an observed m/z of 4469 (M-1 ion), in reasonable agreement with the calculated value (M-1) of 4466.

Quantitative DNase I Footprint Titrations. Equilibrium association constants of oligonucleotide **9** for each of the four target sites shown in figure 4 were determined by quantitative DNase I footprint titration (6) as described in Chapter Three. Figure 9 shows an autoradiogram of a DNase I footprint titration gel with oligonucleotide **9**, while Figure 10 shows binding isotherms derived from a series of these gels. The measured equilibrium association constants and free energies of triple helix formation are listed in Table 1, along with values for oligonucleotides containing thymidine or 5-methyl-2'-deoxycytidine at the variable position (data taken from Chapter Three).

Oligonucleotide **9**, containing 2-pyrimidinone nucleoside, binds to the AT target site with highest affinity ($K_T = 2.3 \times 10^6 \text{ M}^{-1}$). It also binds both the GC and CG target sites with 5 to 10 fold lower affinity ($K_T = 3 - 5 \times 10^5 \text{ M}^{-1}$). This oligonucleotide clearly does not show the desired CG specificity, since it is weakly specific for the AT site. In addition, its affinity for even the AT site is two orders of magnitude lower than the oligonucleotide containing T at the variable position (Table I, $K_T \geq 2.1 \times 10^8 \text{ M}^{-1}$ for the

AT site).

Conclusion

The desired CG specificity was not observed for triple helix formation by an oligonucleotide containing the 2-pyrimidinone base. Rather, this base showed weakly specific, low affinity recognition of AT base pairs. The model of the 2-pyrimidinone•AT base triplet shown in Figure 2 is probably a reasonable description of this interaction. However, it is not clear why the affinity for CG is lower than anticipated. Perhaps more strikingly, we do not understand the relatively high affinity for the GC site, at which no hydrogen bonding complementarity exists (Figure 2).

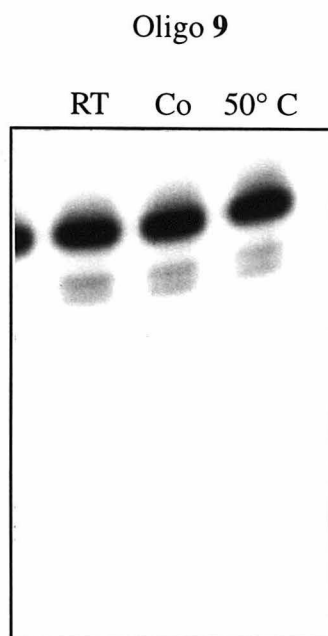
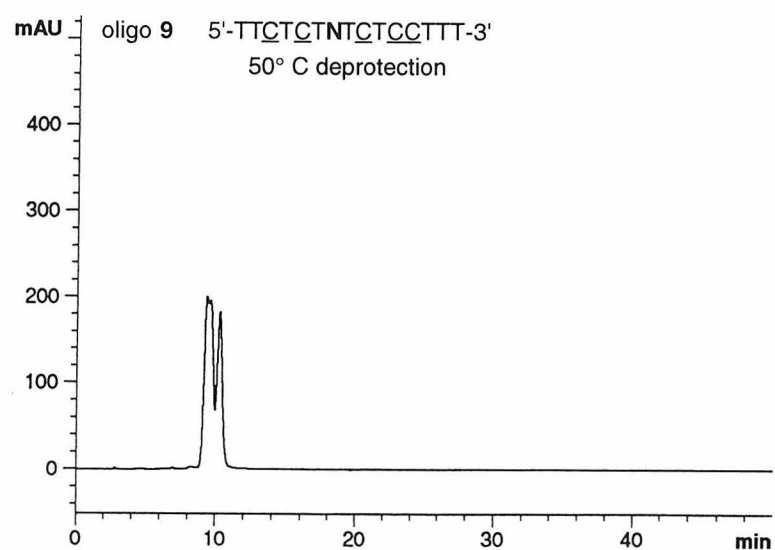


Figure 6: Storage phosphor autoradiogram of a 20% denaturing polyacrylamide gel for analysis of 5'-³²P end labeled oligonucleotide **9** containing 2-pyrimidinone. Samples deprotected at room temperature (RT), 50° C, and a "cospot" (Co) of both were analysed. The dynamic range of the gel is 10 to 2,000 arbitrary intensity units.

A.



B.

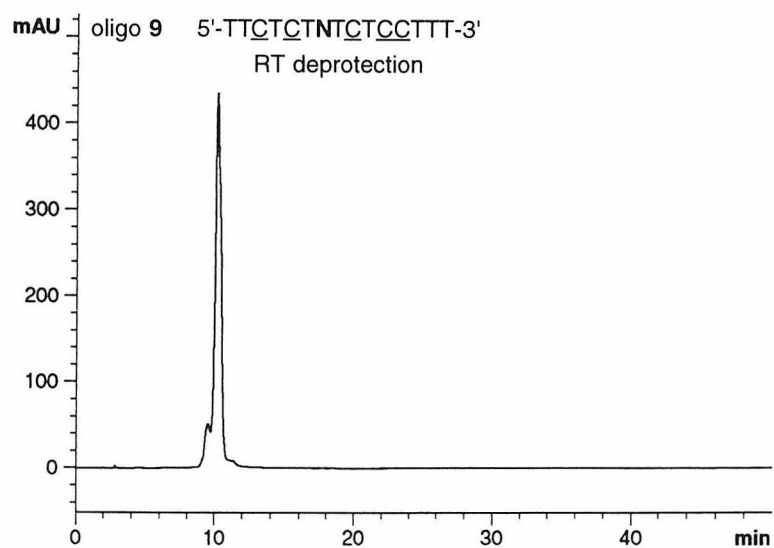
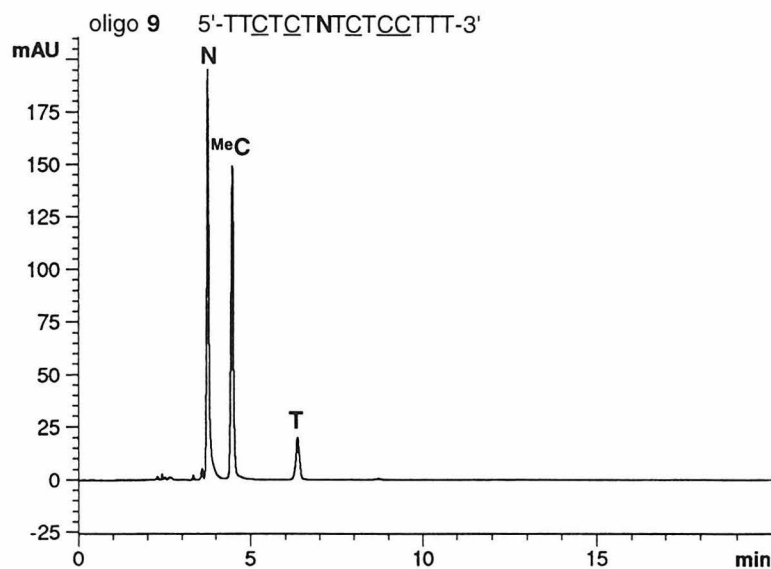


Figure 7: HPLC traces of samples of oligonucleotide **9** deprotected in aqueous ammonia at A) 50° C and B) room temperature. The traces were monitored at 260 nm.

A.



B.

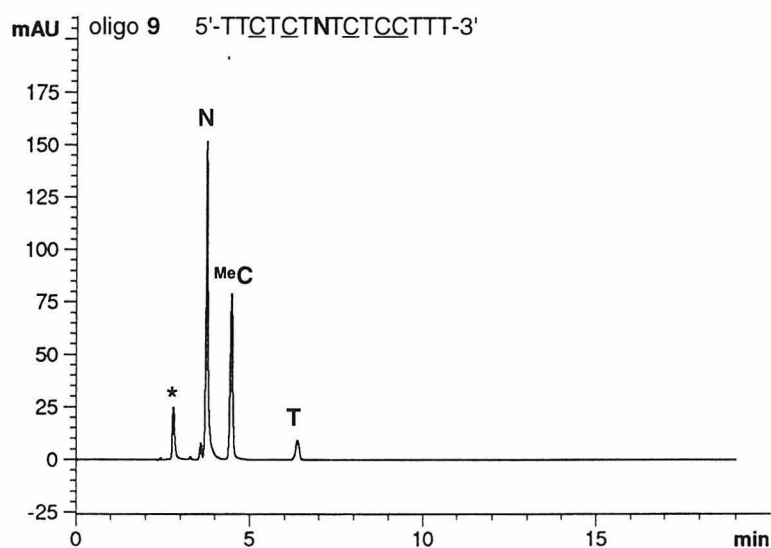


Figure 8: HPLC traces monitored at 303 nm of A) enzymatically digested oligonucleotide **9** and B) a coinjection of the enzyme digest and the three component nucleosides. The peak marked “*” is an impurity in the 2-pyrimidinone nucleoside standard. N indicates 1-(2'-deoxy- β -D-*erythro*-pentofuranosyl)-2-pyrimidinone and C indicates 2'-deoxy-5-methylcytidine.

Figure 9: (Top) Sequences of the oligonucleotide and the four purine rich target sites for the specificity experiments. The duplex target sites, which differ only at a single position labeled XY, are arranged sequentially on the 314 bp *Afl* II/*Fsp* I restriction fragment of pYSPEC2, each separated by 15 bp of intervening sequence. (C indicates 5-methyl-2'-deoxycytidine and N indicates 1-(2'-deoxy- β -D-*erythro*-pentofuranosyl)-2-pyrimidinone). (Bottom) Autoradiogram of an 8% denaturing polyacrylamide gel used to separate the fragments generated by DNase I digestion in a quantitative footprint titration experiment. The boxes drawn to the left of the autoradiogram indicate the positions of the 15 bp binding sites within the 314 bp restriction fragment. (Lane 1) Intact 3' labeled DNA after incubation in the absence of the third strand oligonucleotide. (Lane 2) Products of an adenine specific sequencing reaction. (Lane 3) Products of a guanine specific sequencing reaction. (Lanes 4-22) DNase I digestion products obtained in the presence of varying concentrations of oligonucleotide **9**: no oligonucleotide (lane 4); 80 pM (lane 5); 200 pM (lane 6); 400 pM (lane 7); 800 pM (lane 8); 2 nM (lane 9); 4 nM (lane 10); 8 nM (lane 11); 20 nM (lane 12); 40 nM (lane 13); 80 nM (lane 14); 200 nM (lane 15); 400 nM (lane 16); 800 nM (lane 17); 2 μ M (lane 18); 4 μ M (lane 19); 8 μ M (lane 20); 20 μ M (lane 21); 40 μ M (lane 22).

5' -TTCTCTNTCTCCTTT-3'

5' -AAGAGAXAGAGGAAA-3'

3' -TTCTCTYTCTCCTTT-5'

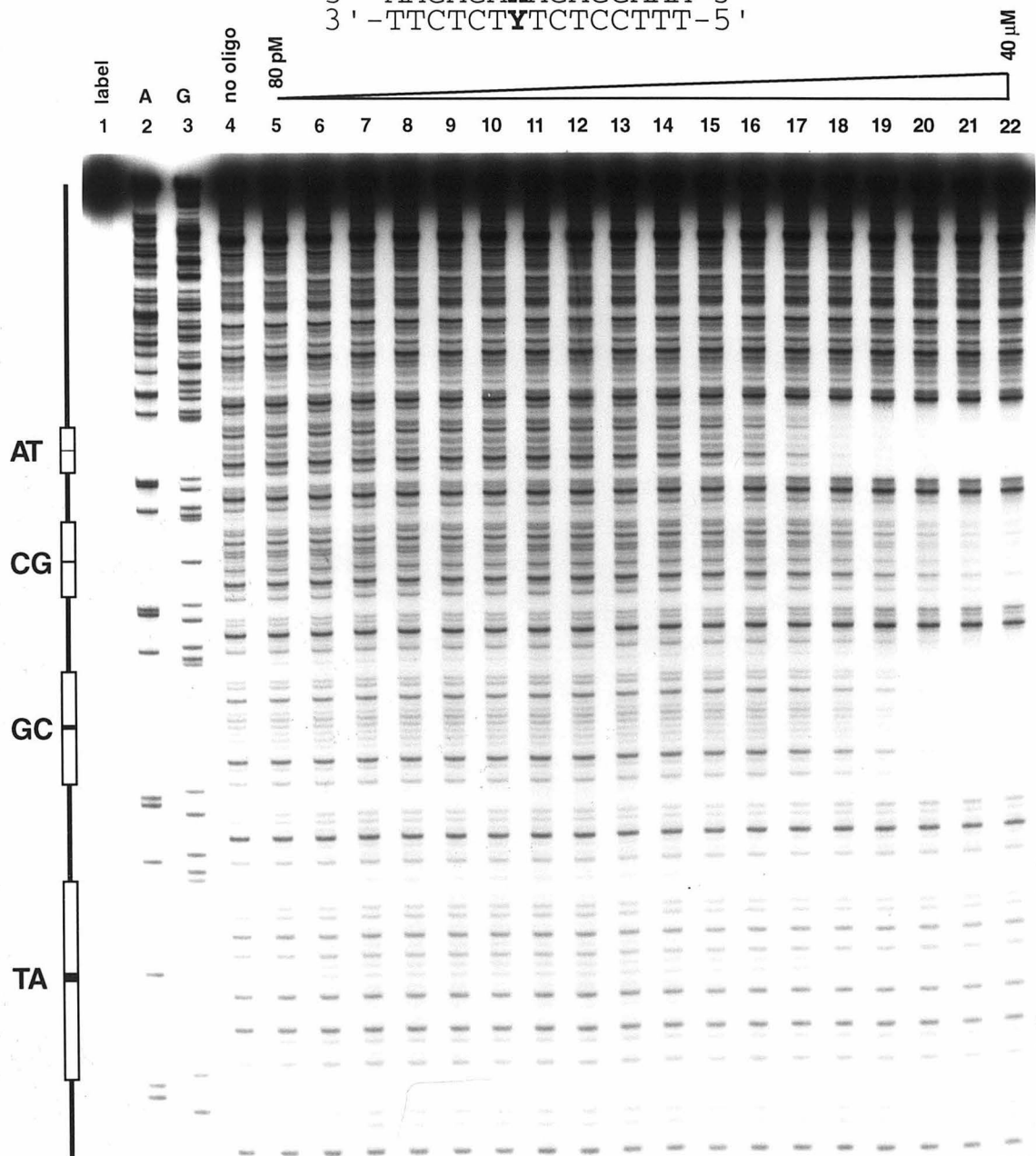
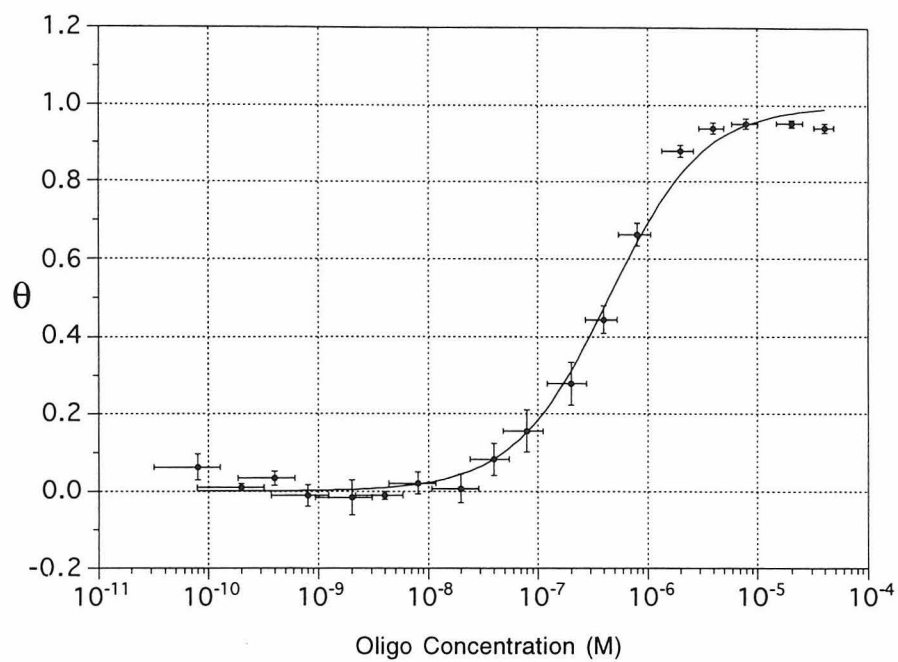
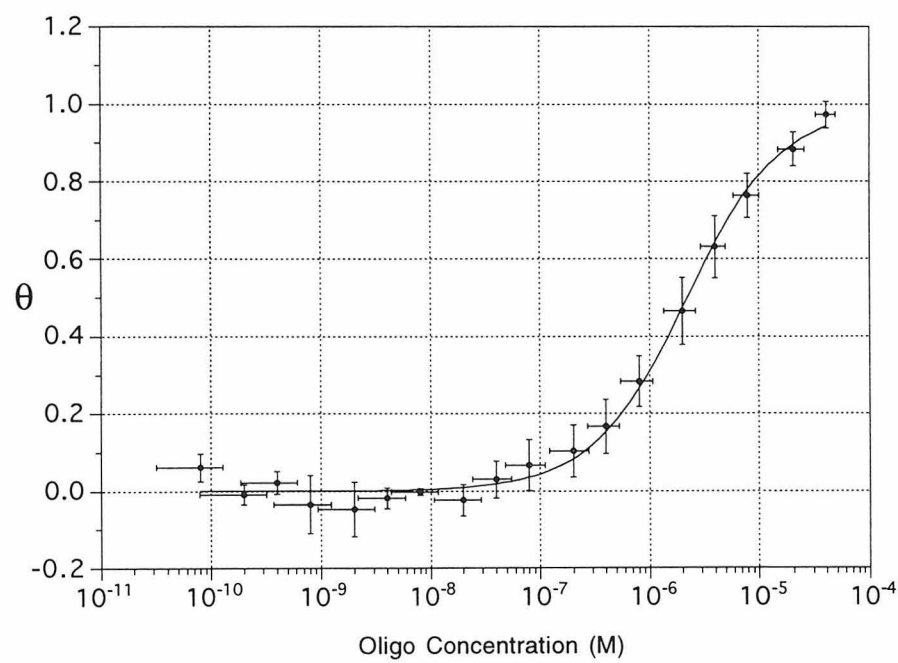


Figure 10: Plots of the site occupancy (θ) versus oligonucleotide concentration for binding of oligonucleotide **9** to each of the four target sites on the 314 bp *Afl* II/*Fsp* I restriction fragment of pYSPEC2. Each data set is averaged from three independent gels. Binding isotherms derived from the data are shown as solid lines. Vertical error bars indicate the standard deviation of θ for each data point, while horizontal error bars indicate estimated pipetting error propagated through the dilution scheme.

A. XY = AT**B. XY = CG**

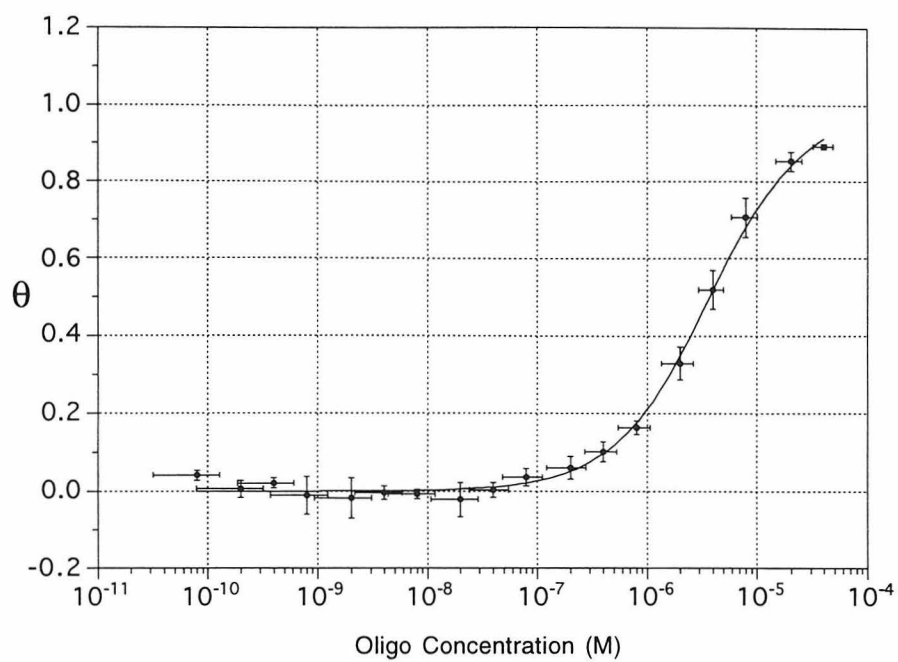
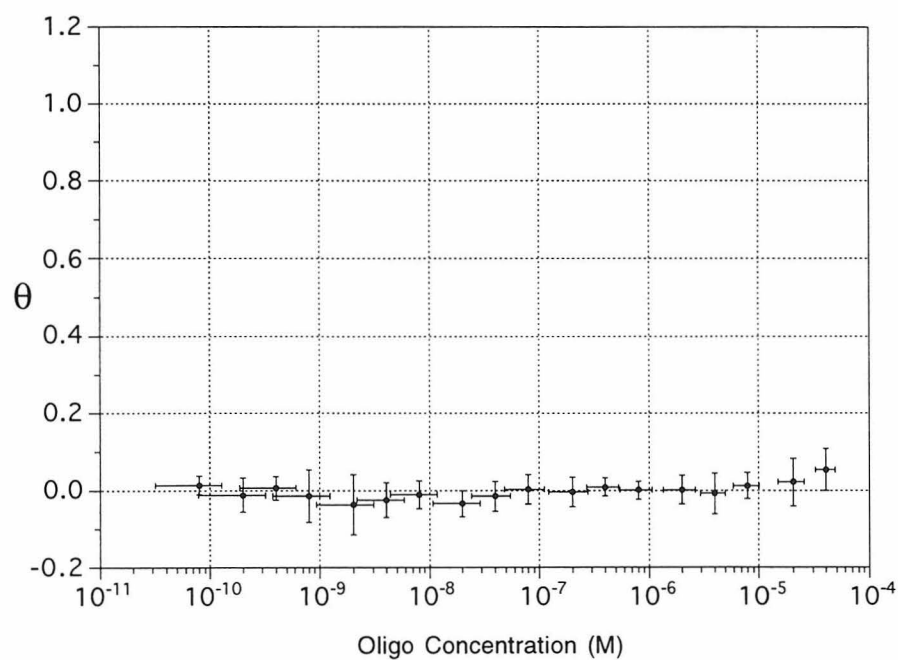
C. XY = GC**D. XY = TA**

Table I: Equilibrium Association Constants for Triple Helix Formation by Oligonucleotides Containing 2-Pyrimidinone, Thymine, or 5-Methylcytosine Bases				
oligo	Z ^b	XY	K _T	ΔG° ^c
9	N	AT	2.3 (±0.2) × 10 ⁶	-8.6 (±0.1)
		CG	4.7 (±1.0) × 10 ⁵	-7.6 (±0.1)
		GC	2.7 (±0.2) × 10 ⁵	> -7.3 (±0.1)
		TA	< 10 ⁴	> -5.4
	T	AT	≥ 2.1 × 10 ⁸	< -11.2
		CG	5.4 (±0.7) × 10 ⁵	-7.7 (±0.1)
		GC	< 10 ⁴	> -5.4
		TA	< 10 ⁴	> -5.4
	MeC	AT	8.2 (±1.9) × 10 ⁴	-6.6 (±0.1)
		CG	1.7 (±0.6) × 10 ⁵	-7.1 (±0.2)
		GC	≥ 1.4 × 10 ⁸	< -11.0
		TA	< 10 ⁴	> -5.4

^a Each reported K_T value is the mean (±SEM) of three independent measurements which were performed in 100 mM NaCl, 10 mM bis-tris, 250 μM spermine at pH 7.0 and 22° C. Values for K_T and ΔG° are reported in units of M⁻¹ and kcal•mol⁻¹, respectively. ^b Z refers to the position in the third strand oligonucleotide opposite the variable position (XY) in the duplex. N indicates 1-(2'-deoxy-β-D-*erythro*-pentofuranosyl)-2-pyrimidinone. The values for T and MeC were taken from Chapter Three. ^c ΔG° values were obtained from the relationship ΔG° = -RTlnK (T = 295 K) and error estimates were obtained by propagating the SEM for K_T through this formula.

Experimental

General. Reagents for organic synthesis were obtained from Aldrich Chemical Co. and used as received. Anhydrous solvents over molecular sieves were purchased from Fluka, with the exception of pyridine, which was distilled over CaH_2 . NMR spectra were obtained on a GE 300 MHz spectrometer and referenced to the residual solvent signal. Mass spectral analyses were obtained at the University of California, Riverside Mass Spectroscopy Facility. Preparative oligonucleotide separations were carried out on a Pharmacia FPLC system and a Waters Millennium HPLC as described below. Analytical HPLC was performed on a Hewlett Packard 1090 chromatograph. All solutions for nucleic acid manipulations were prepared with water purified in a Millipore Milli-Q system.

Synthesis and Purification of Oligonucleotide 9. Phosphoramidite **8** was synthesized as described previously (4). This compound was dissolved in acetonitrile (Fluka, puriss) to a concentration of 1.0 M and filtered (0.45 μm nylon), providing the required DNA synthesis reagent. Oligonucleotide **9** was synthesized on a 1.0 μmol scale with an Applied Biosystems Model 380B DNA synthesizer running standard protocols. The oligonucleotide, with 5'-DMT group removed, was obtained in concentrated ammonium hydroxide solution and divided into two equal portions, each of which was supplemented with 1 ml of fresh concentrated ammonium hydroxide. One portion was deprotected for 5.5 hrs at 50° C and the other was deprotected for 21.5 hrs at room temperature. Argon was bubbled through both solutions for several hours, and they were then frozen and lyophilized. The residual samples were purified by FPLC with a Mono Q HR 10/10 anion exchange column, using a linear gradient of 0.1 M to 0.6 M NaCl in 10 mM bis tris, pH 7.0 over 55 minutes at a flow rate of 1 ml/min. The sample deprotected at 50° C (hereafter "50° C sample") eluted as two unresolved major peaks which were combined, while the

room temperature sample eluted as a single major peak with trailing shoulder, which were separated. Following lyophilization, the samples were purified by reverse phase HPLC with a Vydac C18 column (#201HS510), using a linear gradient of 5 to 40% acetonitrile in 0.1 M triethylammonium acetate, pH 7.0 over 35 minutes at a flow rate of 4 ml/min. Following lyophilization, each sample was desalted on a NAP-10 column (Pharmacia) and quantitated by uv spectroscopy, using an extinction coefficient of $110,000 \text{ M}^{-1} \text{ cm}^{-1}$ for oligonucleotide **9**. Oligonucleotide aliquots were lyophilized to dryness and stored at -20°C .

Kinase Treatment and PAGE Analysis. Samples of oligonucleotide **9** obtained from different chromatographic fractions and deprotection protocols were 5'-end labeled with $\gamma\text{-}^{32}\text{P}$ ATP (ICN, 7000 Ci/mmol) and T4 polynucleotide kinase (Boehringer Mannheim) according to a previously published protocol (7), diluted with formamide loading buffer and separated on a pre-electrophoresed denaturing 20% polyacrylamide gel (55 min at 60 W constant power). The gel was exposed to a storage phosphor screen and quantitated on a Molecular Dynamics phosphorimager.

HPLC Analysis of Intact and Enzymatically Digested Oligonucleotide 9. Oligonucleotide samples (3-4 nmol) were analyzed by reverse phase HPLC with a Rainin Microsorb-MV C18 column using a linear gradient of 5 to 30% acetonitrile in 20 mM ammonium acetate, pH 7 over 50 min at a flow rate of 1 ml/min and temperature of 40°C . 10 nmol oligonucleotide samples were digested with three units of calf alkaline phosphatase (Boehringer Mannheim) and 0.009 units of snake venom phosphodiesterase (Boehringer Mannheim) in 50 μl 50 mM tris•HCl, 10 mM MgCl_2 , pH 8.0 for 2 hrs at 37°C . 20 μl aliquots were centrifuged for 10 min at maximum speed in a microcentrifuge and then analyzed by reverse phase HPLC using the column described above and the following gradient in 20 mM ammonium acetate, pH 7.0: 0-20 min, isocratic at 5% acetonitrile,

then 20-42.5 min, linear gradient of 5 to 50% acetonitrile at a flow rate of 1 ml/min and temperature of 40° C. Coinjections with authentic samples of the component nucleosides were also run.

Quantitative DNase I Footprinting. The equilibrium association constants for triple helix formation by oligonucleotide **9** at the four sites on the 314 bp *Afl* II/*Fsp* I restriction fragment of pYSPEC2 were determined by quantitative Dnase I footprint titration as described in the experimental section of Chapter Three, with the exception that the addition of a nonspecific oligonucleotide was omitted.

References

- (1) Koh, J. S. Ph.D. Thesis, California Institute of Technology, 1991.
- (2) Taylor, C.; Ford, K.; Connolly, B. A.; Hornby, D. P. *Biochem. J.* **1993**, *291*, 493-504.
- (3) Brown, D. J.; Hoerger, E.; Mason, S. F. *J. Chem. Soc.* **1955**, 211-217.
- (4) Connolly, B. A. *Methods Enzymol.* **1992**, *211*, 36-53.
- (5) *Oligonucleotides and Analogues: A Practical Approach*; Eckstein, F., Ed.; Oxford University: New York, 1991.
- (6) Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Methods Enzymol.* **1986**, *130*, 132-181.
- (7) *Applied Biosystems Manual: DNA Synthesis and Purification*, Chapter 3, pp. 6-11.

CHAPTER SIX

Design of an N^7 -Glycosylated Purine Nucleoside for Recognition of GC Base Pairs by Triple Helix Formation

The text of this chapter is taken from a published paper that was coauthored with Prof.

Peter B. Dervan, Jürg Hunziker, and Helmut Brunar.

(Hunziker, J. et al. *J. Am. Chem. Soc.* **1995**, *117*, 2661-2662)

Introduction

Pyrimidine oligodeoxyribonucleotides bind in the major groove of DNA *parallel* to the purine Watson-Crick strand by formation of specific hydrogen bonds between thymine and adenine (T•AT triplet) and protonated cytosine and guanine (C+GC triplet) on the Hoogsteen face of the purine base.¹ Alternatively, purine oligodeoxyribonucleotides bind in an *antiparallel* orientation relative to the purine Watson-Crick strand by formation of G•GC and A•AT triplets.² The prerequisite protonation of cytosine in C+GC triplets leads to a considerable pH dependence in the binding affinity of C-containing oligodeoxyribonucleotides (Figure 1).^{1,3} Substitution of 5-methylcytosine (**mC**) for cytosine results in increased binding affinities near physiological pH.^{3a,d,e} In an attempt to eliminate the necessity for protonation, recent efforts have been directed toward the synthesis of *nonnatural* nucleosides which display the hydrogen bonding functionality of protonated cytosine.^{4,5}

Results and Discussion

We report that an *N*⁷-glycosylated purine, 7-(2-deoxy-β-D-*erythro*-pentofuranosyl) guanine (**d⁷G**), when incorporated in a pyrimidine oligonucleotide, binds with remarkable specificity the Watson-Crick guanine-cytosine (GC) base pair by triple helix formation. By attaching the deoxyribose moiety at the *N*⁷-position of a guanine base, *the third strand orientation in a G•GC base triplet is reversed* and becomes *parallel* to the purine Watson-Crick strand (Figure 1).

The phosphoramidite of **d⁷G** was synthesized in four steps from methyl 3,5-di-O-benzoyl-α,β-D-*erythro*-pentafuranoside and *N*²-isobutyrylguanine (Figure 2).⁷ The **d⁷G** nucleotide was incorporated at a single position within 15-mer pyrimidine oligodeoxyribonucleotide **1** by automated methods.⁸ In order to test the affinity and

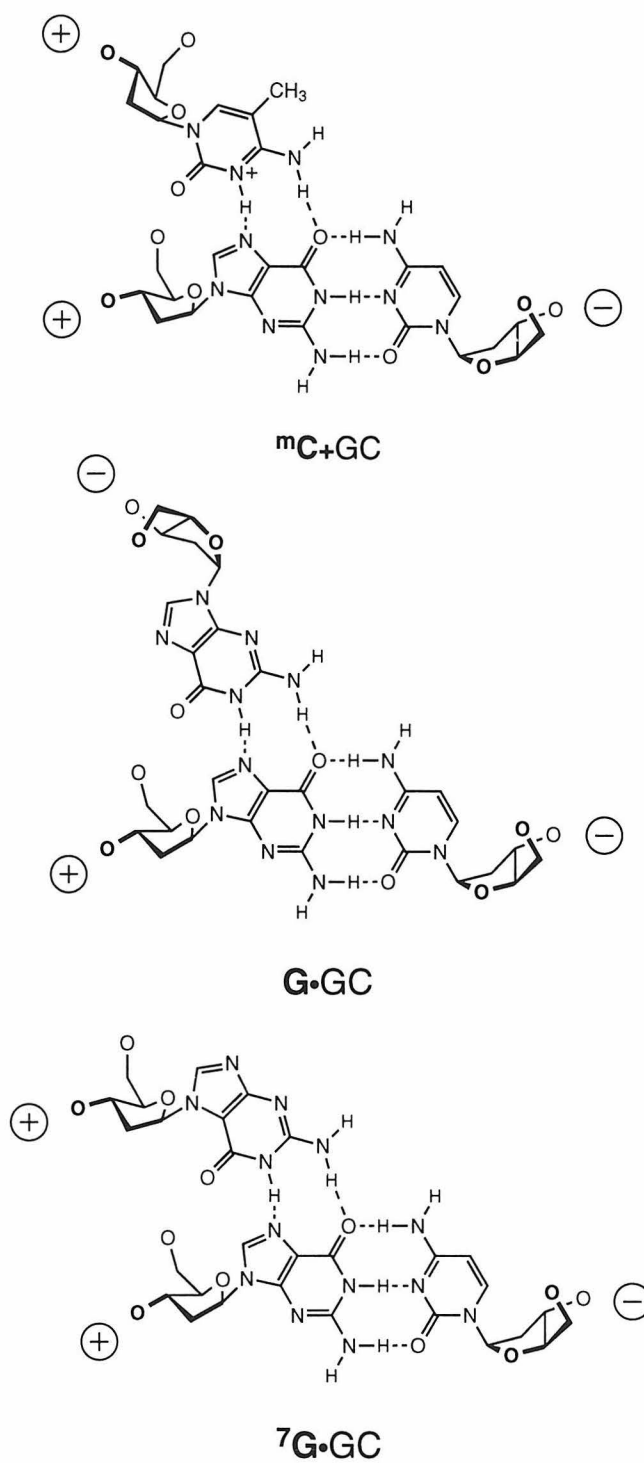


Figure 1: Schematic representations of the base triplets ^mC+GC (top), G•GC (middle) and ⁷G•GC (bottom). Plus and minus signs indicate relative backbone polarity.

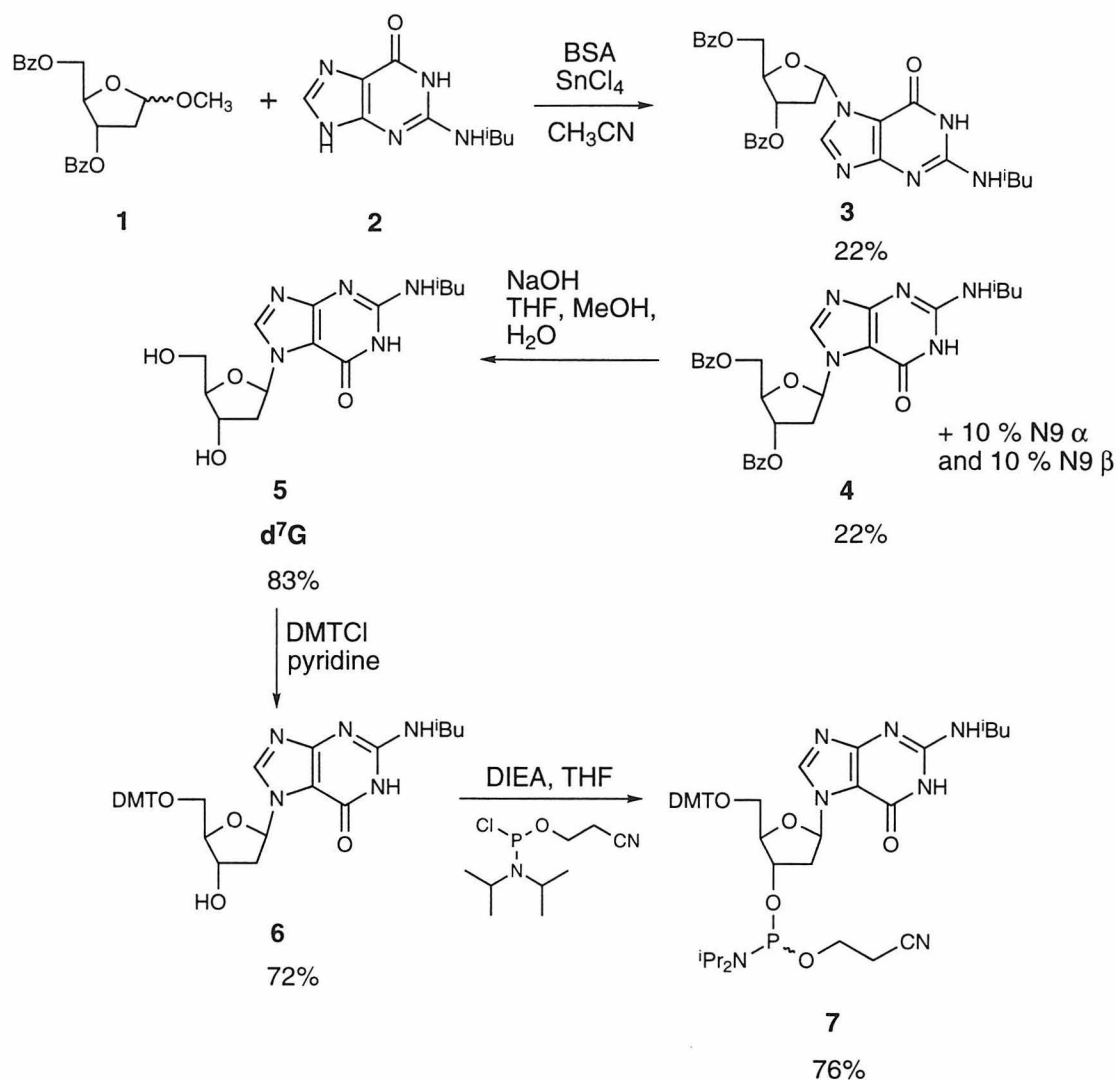


Figure 2: Synthesis of the DMT protected phosphoramidite of **d⁷G**, **7**.
For full experimental details, see reference 7a.

specificity of **d⁷G**, the association constants for the formation of 20 triple helical complexes which vary at a single common position Z·XY (where Z = ⁷G, ^mC, T, G, A and XY = AT, GC, CG, TA) were determined within a pyrimidine motif by quantitative DNase I footprint titration (100 mM NaCl, 10 mM Bis-Tris·HCl, 0.25 mM spermine·4HCl, pH 7.0, 22°C) (Table I).⁹ The sequence context in which the Z·XY pairings were studied is 5'-d(AAGAGAXAGAGGAAA)-3' within a 314 base pair DNA restriction fragment.¹⁰ The affinity of ⁷G is determined by comparison of oligonucleotides

Table I: Equilibrium Association Constants (M^{-1}) for 20 Triple Helical Complexes which Vary at a Single Position.^a

Z	oligo	XY			
		AT	CG	GC	TA
⁷G	1	$< 10^4$	$(4.6 \pm 1.9) \times 10^5$	$\geq 1.6 \times 10^8$	$< 10^4$
mC	2	$(8.2 \pm 1.9) \times 10^4$	$(1.7 \pm 0.6) \times 10^5$	$\geq 1.4 \times 10^8$	$< 10^4$
T	3	$\geq 2.1 \times 10^8$	$(5.4 \pm 0.7) \times 10^5$	$< 10^4$	$< 10^4$
G	4	$< 10^4$	$< 10^4$	$< 5 \times 10^4$	$(1.1 \pm 0.3) \times 10^6$
A	5	$(1.6 \pm 0.6) \times 10^5$	$< 10^4$	$< 10^4$	$< 10^4$

^aData reported are the mean values of three footprint titration experiments $\pm$ SEM. The titrations were performed at 22°C, pH 7.0, in the presence of 100 mM NaCl, 10 mM Bis-Tris•HCl and 0.25 mM spermine•4HCl and allowed to equilibrate for 120 hrs prior to footprinting. Association constants greater than $1 \times 10^8 M^{-1}$ are reported as lower limits due to the slow kinetics of triple helix formation under these conditions.

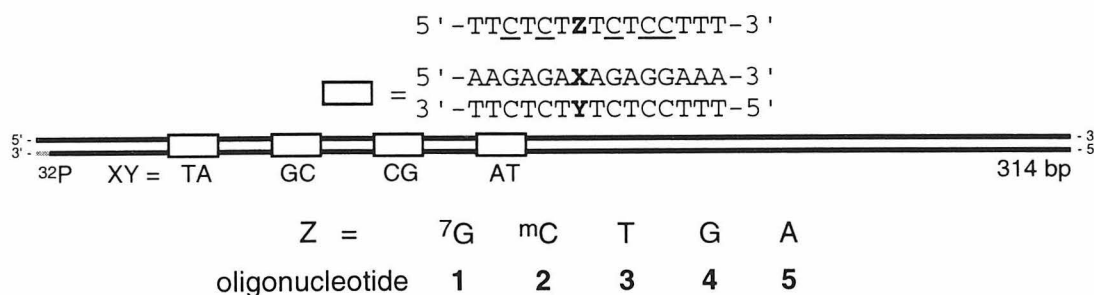


Figure 3: Sequences of the oligonucleotides and duplex target sites used in the quantitative DNase footprint titrations (C indicates mC).

1 and **2** ($Z = {}^7\text{G}$, ${}^m\text{C}$) opposite $XY = \text{GC}$ (Figure 3). The sequence specificity of ${}^7\text{G}$ is determined by comparison of oligonucleotide **1** ($Z = {}^7\text{G}$) opposite each of the four Watson-Crick base pairs ($XY = \text{AT}$, GC , CG and TA) (Figure 3).

We find that oligonucleotide **1** ($Z = {}^7\text{G}$) binds the target site with an equilibrium association constant $\geq 1.6 \times 10^8 \text{ M}^{-1}$, an affinity equal to oligonucleotide **2** ($Z = {}^m\text{C}$) (Table I). Importantly, **d⁷G** is specific for GC base pairs; the stabilities of triple helical complexes containing **d⁷G** decrease in the order $\text{d}^7\text{G} \cdot \text{GC} \gg \text{d}^7\text{G} \cdot \text{CG} > \text{d}^7\text{G} \cdot \text{AT}$, $\text{d}^7\text{G} \cdot \text{TA}$ (Table I). The high affinity and sequence specificity observed is consistent with a model where ${}^7\text{G}$ within the pyrimidine motif forms two specific hydrogen bonds to the purine strand of the Watson-Crick GC base pair (Figure 1). Based on these results, we envision that *a new parallel-stranded motif* comprising wholly N^7 -purines might be possible for DNA recognition by triple helix formation.

References

- (1) (a) Moser, H. E.; Dervan, P. B. *Science* **1987**, *238*, 645-650. (b) For a review, see Thuong, N. T., Helene, C. *Angew Chem Int. Ed. Engl.* **1993**, *32*, 666-690, and references cited there.
- (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) For NMR studies see: Radhakrishnan, I.; de los Santos, C.; Patel, D. J. *J. Mol. Biol.* **1993**, *234*, 188-197.
- (3) (a) Povsic, T. J.; Dervan, P. B. *J. Am. Chem. Soc.* **1989**, *111*, 3059-3061. (b) Plum, G. E.; Park, Y.-W.; Singleton, S. F.; Dervan, P. B.; Breslauer, K. J. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 9436-9440. (c) Maher, L. J., III; Dervan, P. B.; Wold, B. J.; *Biochemistry* **1990**, *29*, 8820-8826. (d) Xodo, L. E.; Manzini, G.; Quadrifoglio, F.; van der Marel, G. A.; van Boom, J. H. *Nucleic Acids Res.* **1991**, *19*, 5625-5631. (e) Singleton, S. F.; Dervan, P. B. *Biochemistry* **1992**, *31*, 10995-11003.
- (4) Earlier work in our group focused on the novel nucleoside P, ^{5d} which forms specific hydrogen bonds between its exocyclic amino and imino groups and the Hoogsteen face of guanine⁶ and recognizes GC base pairs over an extended pH range relative to mC or C.
- (5) (a) Ono, A.; Ts'o, P. O. P.; Kan, L.-S. *J. Am. Chem. Soc.* **1991**, *113*, 4032-4033. (b) Ono, A.; Ts'o, P. O. P.; Kan, L.-S. *J. Org. Chem.* **1992**, *57*, 3225-3230. (c) Young, S. L.; Krawczyk, S. H.; Matteucci, M. D.; Toole, J. J. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 10023-10026. (d) Koh, J. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 1470-1478. (e) Krawczyk, S. H.; Milligan, J. F.; Wadwani, S.; Moulds, C.; Froehler, B. C.; Matteucci, M. D. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 3761-3764. (f) Miller, P. S.; Bhan, P.; Cushman, C. D.; Trapane, T. L. *Biochemistry* **1992**, *31*, 6788-6793. (g) Jetter, M. C.; Hobbs, F. W. *Biochemistry* **1993**, *32*, 3249-3254. (h) Davison, E. C.; Johnsson, K. *Nucleosides & Nucleotides* **1993**, *12*, 237-243.

(6) Radhakrishnan, I.; Patel, D. J.; Priestley, E. S.; Nash, H. N.; Dervan, P. B.

Biochemistry **1993**, *32*, 11228-11234.

(7) (a) Brunar, H.; Dervan, P. B. *Nucleic Acids Res.* Submitted. (b) For another recent synthesis see: Rao, T. S.; Durland, R. H.; Revankar, G. R. *J. Heterocyclic Chem.* **1994**, *31*, 935-940. (c) Methyl 3,5-di-O-benzoyl- α , β -D-erythro-pentofuranoside was synthesized according to: Blair, M. G.; Lipkin, D.; Sowden, J. C.; Strobach, J. C. *J. Org. Chem.* **1960**, *25*, 1679-1680. (d) N2-Isobutyryl-guanine was prepared in analogy to: Hrebabecky, H.; Farkas, J. *Acetylation of an Amino Group on a Purine Ring with Acetic Anhydride in an Aprotic, Dipolar Solvent* in Townsend, L. B.; Tipson, R. S. *Nucleic Acid Chemistry Vol. 1*, J. Wiley & Sons: New York, 1978.

(8) Enzymatic degradation and base composition analysis by HPLC confirmed the composition of oligonucleotide **1**. Analysis by MALDI TOF mass spectrometry gave excellent agreement between the experimental (4519.7) and calculated (4519.1) mass (M-1) of this oligonucleotide.

(9) For a detailed description of the quantitative DNase footprint titration method see: (a) Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Methods Enzymol.* **1986**, *130*, 132-181. (b) Singleton, S. F.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 6957-6965. (c) Priestley, E. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1995**, *117*, 4761-4765.

(10) For examples of a similar assay, see: Jones, R. J.; Swaminathan, S.; Milligan, J. F.; Wadwani, S.; Froehler, B. C.; Matteucci, M. D. *J. Am. Chem. Soc.* **1993**, *115*, 9816-9817. Jones, R. J.; Lin, K.-Y.; Milligan, J. F.; Wadwani, S.; Matteucci, M. D. *J. Org. Chem.* **1993**, *58*, 2983-2991.

Appendix

The results of a study of triple helix formation by two oligonucleotides containing multiple d^7G residues are summarized here along with results taken from Chapter Three for mC and P containing oligonucleotides, in order to facilitate comparison. The synthesis of the oligonucleotides containing d^7G and the footprinting studies on these oligonucleotides were carried out by Helmut Brunar as a postdoctoral researcher in the Dervan Group. The results of the experiments are tabulated below. (For a complete description of the multiple d^7G work, see: Brunar, H.; Dervan, P. B. *Nucleic Acids Res.* Submitted.)

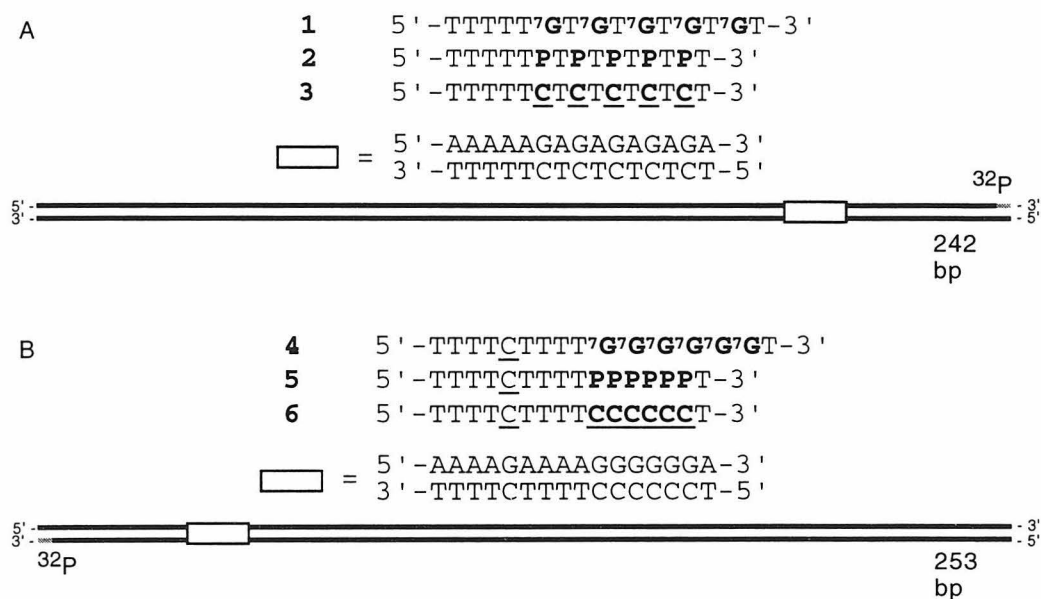


Figure 4: Sequence Composition Experiments. A) Sequences of oligonucleotides **1-3** and the (GA)₅ target site. The target site is located within the 242 bp *Eco* RI/*Pvu* II restriction fragment of pGCBGC. B) Sequences of oligonucleotides **4-6** and the G₆ target site. The target site is located within the 253 bp *Eco* RI/*Pvu* II restriction fragment of pSPHIV. (C indicates 5-methyl-2'-deoxycytidine and P indicates 1-(2-deoxy-β-D-ribofuranosyl)-3-methyl-5-amino-1H-pyrazolo[4,3-d]pyrimidin-7-one). The experiments with oligonucleotides **2, 3, 5, and 6** are described fully in Chapter Three.

Table II: Equilibrium Association Constants for Triple Helix Formation at the (GA)₅ Site^a

oligo	pH	K _T (M ⁻¹)	ΔG° ^b
1	7.0	1.5 × 10 ⁶	-8.3
	7.5	1.1 × 10 ⁶	-8.1
2	7.0	1.1 (±0.2) × 10 ⁵	-6.8 (±0.1)
	7.5	1.3 (±0.2) × 10 ⁵	-6.9 (±0.1)
3	7.0	3.1 (±0.2) × 10 ⁹	-12.8 (±0.1)
	7.5	9.4 (±1.6) × 10 ⁷	-10.8 (±0.1)

^a Each reported K_T value is the mean (±SEM) of three independent measurements which were performed in 100 mM NaCl, 10 mM bis-tris, 250 μM spermine at the indicated pH and 22° C. ^b ΔG° values were obtained from the relationship ΔG° = -RTlnK (T = 295 K) and error estimates were obtained by propagating the SEM for K_T through this formula.

Table III: Equilibrium Association Constants for Triple Helix Formation at the G₆ Site^a

oligo	pH	K _T (M ⁻¹)	ΔG° ^b
4	7.0	3 × 10 ⁹	-12.8
	7.5	2 × 10 ⁹	-12.5
5	7.0	4.4 (±0.3) × 10 ⁹	-13.0 (±0.1)
	7.5	1.6 (±0.1) × 10 ⁹	-12.4 (±0.1)
6	7.0	< 5 × 10 ⁴	> -6.3
	7.5	< 10 ⁴	> -5.4

^a Each reported K_T value is the mean (±SEM) of three independent measurements which were performed in 100 mM NaCl, 10 mM bis-tris, 250 μM spermine at the indicated pH and 22° C. ^b ΔG° values were obtained from the relationship ΔG° = -RTlnK (T = 295 K) and error estimates were obtained by propagating the SEM for K_T through this formula.

These results indicate that oligonucleotides containing multiple d⁷G substitutions show a similar pattern of binding affinity as oligonucleotides containing P, since affinity is relatively low at the (GA)₅ site and very high at the G₆ site. However, d⁷G is slightly superior to P at the (GA)₅ site (Compare **1** and **2**). Taken with its simpler synthesis, these results suggest that d⁷G replace P as a base for recognition of contiguous GC base pairs.

CHAPTER SEVEN

Towards the Development of an N^7 -Purine Motif for Triple Helix Formation

Introduction

Experiments described in Chapter Six demonstrate that the N^7 -deoxyguanosine•GC base triplet (Figure 1) is a strong and specific noncovalent interaction which is generalizable to several sequences. Based on this data, we became interested in developing a third structural motif for triple helix formation, in addition to the previously described pyrimidine (1) and purine (2) motifs, which would incorporate the favorable N^7 -deoxyguanosine•GC base triplet. The third strand of this proposed motif would be composed entirely of nonnatural 7-glycosylated purine nucleosides and would be oriented parallel to the 5' to 3' polarity of the homopurine strand of the target DNA duplex (Figure 1).

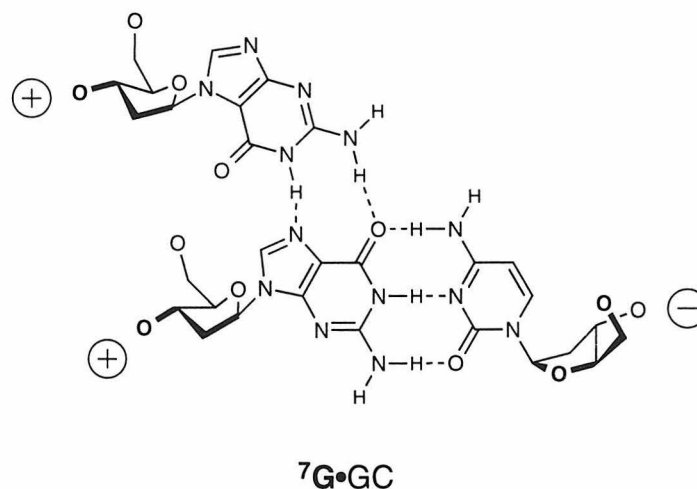


Figure 1: The N^7 -deoxyguanosine•GC base triplet. The third strand is oriented parallel to the purine strand of the duplex. Relative backbone polarity is indicated by the plus and minus signs.

The first step in developing this “ N^7 -purine motif” is the identification of a suitable third strand nucleoside for recognition of AT base pairs, since oligonucleotides composed of such a nucleoside along with N^7 -deoxyguanosine should be capable of recognizing homopurine DNA target sites. We also envisioned that development of a third structural motif for triple helix formation would provide an important new context in which to

develop nonnatural bases for CG and TA recognition. Based on molecular modeling, we chose *N*⁷-deoxyxanthosine (d⁷X), *N*⁷-deoxyisoinosine (d⁷iI), and *N*⁷-deoxyadenosine (d⁷A) as candidates for AT recognition.

*N*⁷-deoxyxanthosine contains the hydrogen bonding face of thymine, and thus should be capable of forming two hydrogen bonds to the Hoogsteen face of an AT base pair (Figure 3) in analogy to the T•AT base triplet. As shown schematically in Figure 3, none of the other three *N*⁷-deoxyxanthosine base triplets contain more than one possible hydrogen bond, strongly suggesting that d⁷X will show specificity for AT base pairs. The d⁷X•AT base triplet is structurally isomorphous (3) with the d⁷G•GC base triplet, indicating that third strand backbone deformation should not be a source of sequence composition effects for oligonucleotides containing both d⁷X and d⁷G. In addition, neither the d⁷X•AT and d⁷G•GC base triplets are charged, so that, unlike the pyrimidine motif, charge/charge repulsion should not cause sequence composition effects. Based on literature data for the model compounds 9-methylxanthine and 7-methylxanthine (Figure 2), the pK_A of d⁷X within a triple helix is expected to be sufficiently high that triple helix formation at neutral pH will be possible (4). The methyl xanthines should have similar pK_a's to the 7- and 9-substituted deoxyxanthosine nucleosides. Finally, 7-deaza-2'-deoxyxanthosine has previously been reported to recognize AT pairs in the purine motif (5).

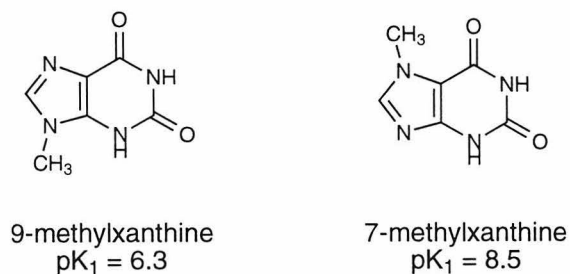


Figure 2: pK_a's of methyl substituted xanthines.

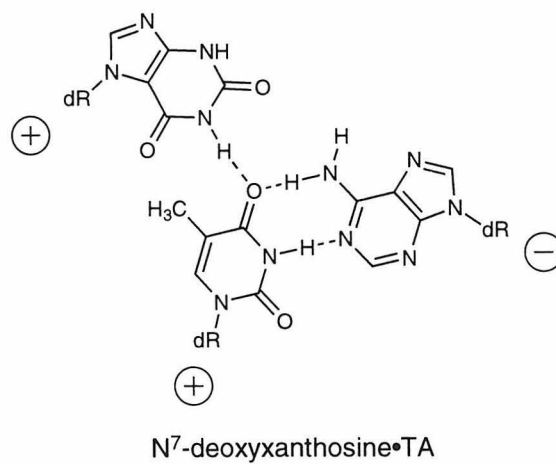
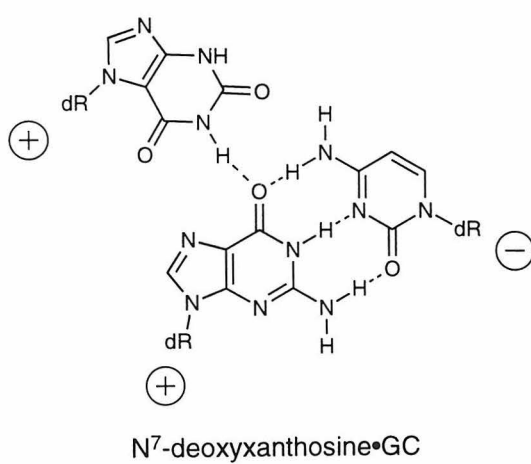
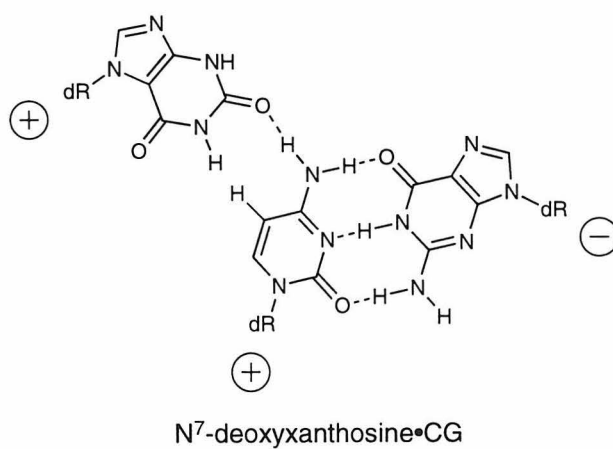
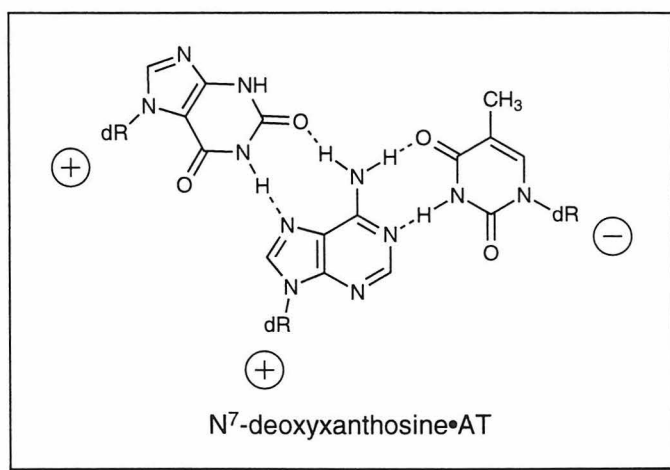


Figure 3: N⁷-deoxyxanthosine base triplets. Only the d⁷X•AT combination contains two hydrogen bonds. Plus and minus signs indicate relative backbone polarity.

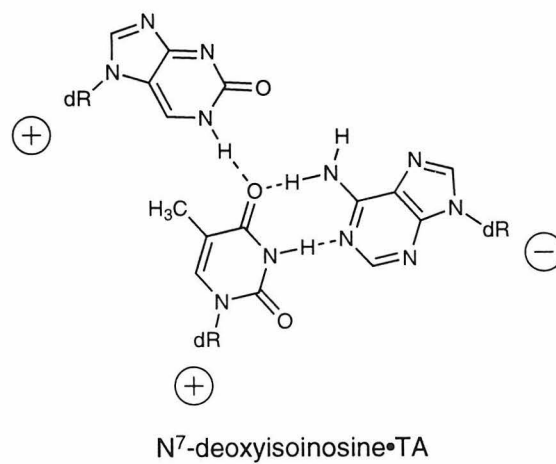
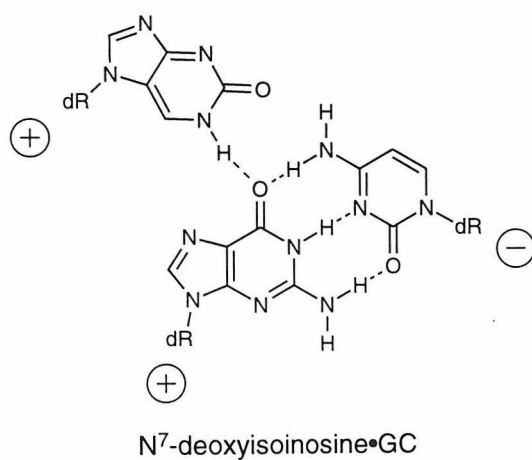
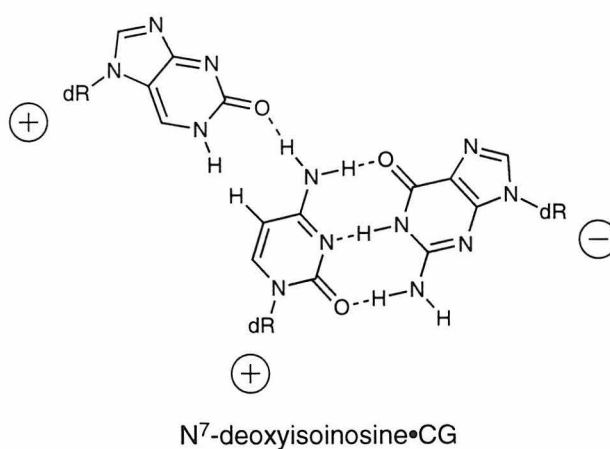
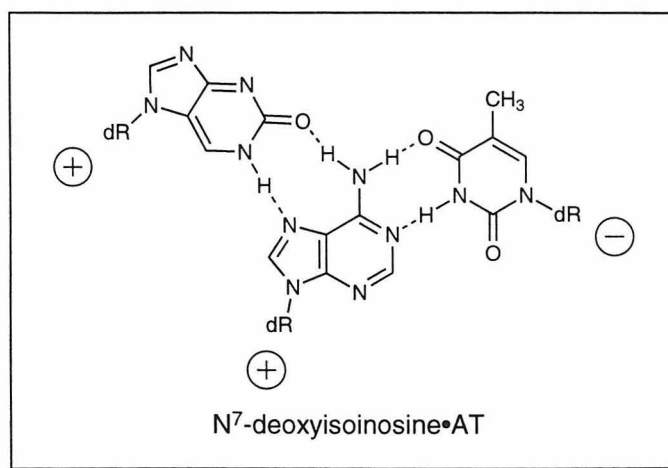


Figure 4: N⁷-deoxyisoinosine base triplets. Only the d⁷iI•AT combination contains two hydrogen bonds. The plus and minus signs indicate relative backbone polarity.

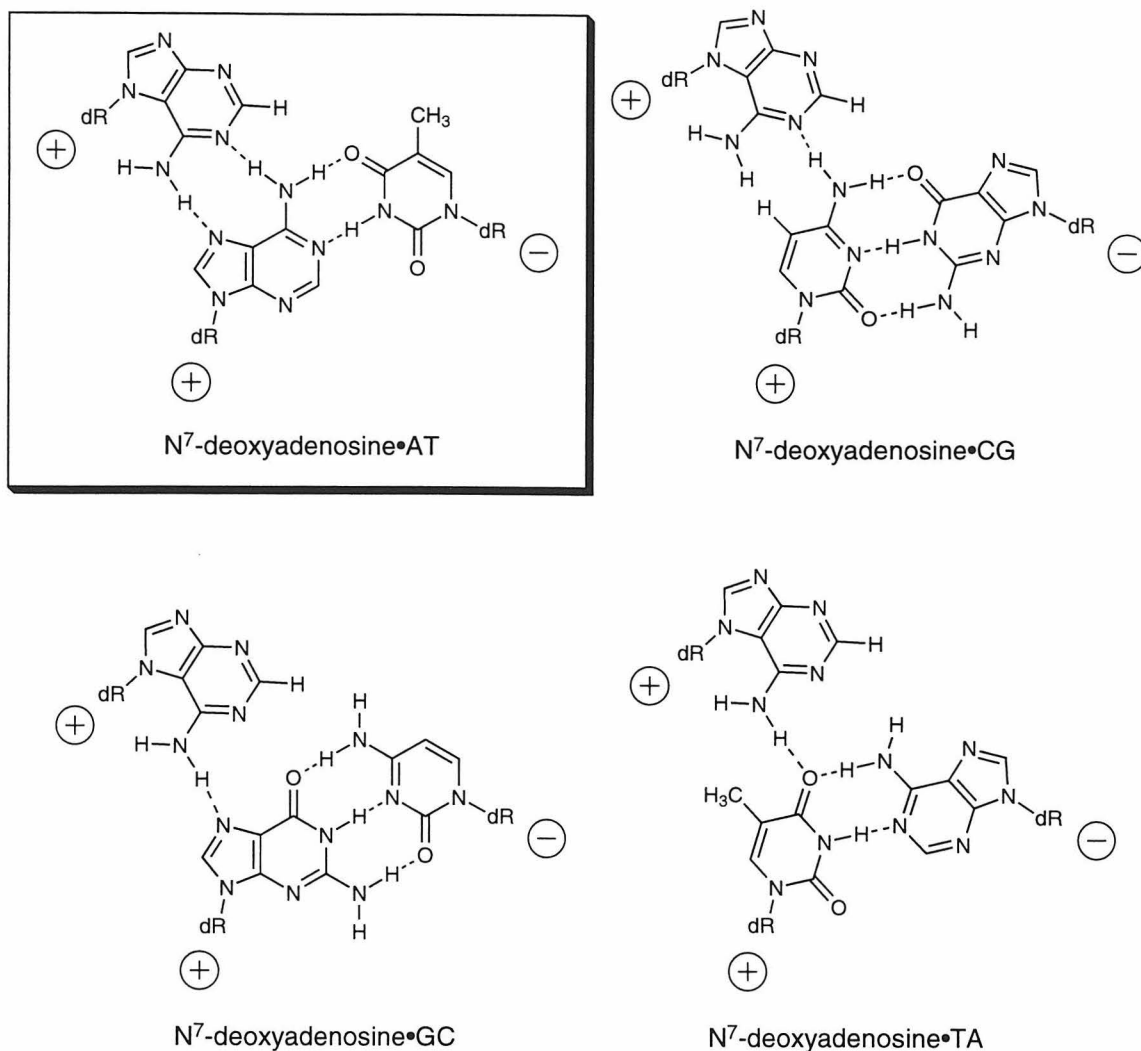


Figure 5: *N*⁷-deoxyadenosine base triplets. Only the d⁷A•AT combination contains two hydrogen bonds. The plus and minus signs indicate relative backbone polarity.

As shown in Figure 4, *N*⁷-deoxyisoinosine is capable of forming two hydrogen bonds with an AT base pair and at most one hydrogen bond to the other three base pairs. Although AT specificity may be anticipated for d⁷iI, this base may also exist in a tautomer in which the proton resides on N3 rather than N1 (Figure 4). This tautomeric ambiguity may affect the specificity of d⁷iI. The d⁷iI•AT base triplet is both uncharged and structurally isomorphous with the d⁷G•GC base triplet, as discussed for the d⁷X•AT base

triplet. The synthesis and characterization of d^7iI in triple helix formation were carried out by Natalia Colocci and Helmut Brunar, and their results are summarized in the Results section (6).

In analogy with the purine motif $A \bullet AT$ triplet, N^7 -deoxyadenosine should be capable of forming a $d^7A \bullet AT$ triplet (Figure 5). The position of third strand glycosylation has moved from N9 to N7 and the orientation of the third strand has switched from antiparallel to parallel with respect to the purine motif $A \bullet AT$ triplet. The $d^7A \bullet AT$ triplet is neutral, but is not structurally isomorphous with $d^7G \bullet GC$, so some sequence composition effects may be anticipated for oligonucleotides containing d^7A and d^7G . In addition, the exocyclic amine and anomeric proton of d^7A are sterically congested as shown in Figure 6. This may reduce the stability of the $d^7A \bullet AT$ triplet by requiring the exocyclic amine of d^7A to rotate out of the plane of the base triplet.

Initial experiments examined the specificity of triple helix formation by oligonucleotides containing single d^7X , d^7iI , or d^7A nucleosides. Based on successful initial results, experiments with oligonucleotides containing either multiple d^7X or d^7A residues for recognition of AT base pairs and d^7G residues for recognition of GC base pairs were synthesized and assayed, in an attempt to develop the N^7 -purine motif.

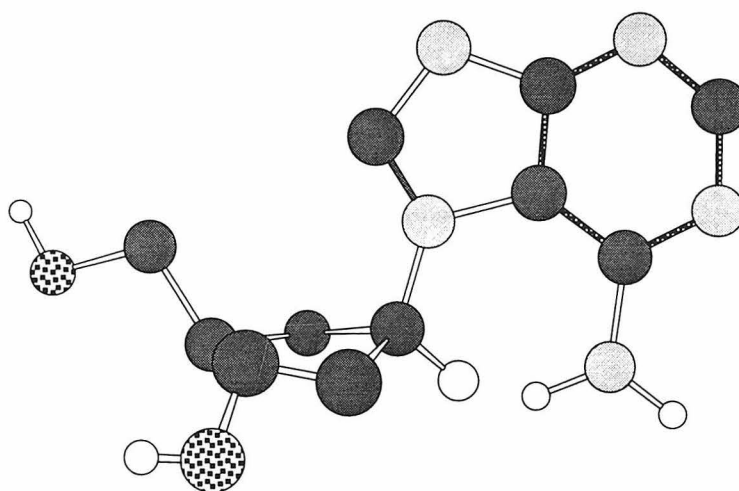


Figure 6: Chem3D model of N^7 -deoxyadenosine, illustrating the close contact between the exocyclic amine and the anomeric proton.

Results and Discussion

Synthesis of d⁷X phosphoramidite. The synthetic scheme for d⁷X phosphoramidite **8** is shown in Figure 7. 2,6-Dichloropurine, **1**, was coupled to the chlorosugar **2** as previously described (7) to give a 2.6:1 mixture of the N9 and N7 nucleosides, **3** and **4**. This mixture proved inseparable by flash chromatography on a preparative scale. Therefore, the mixture was treated with allyl alkoxide in allyl alcohol, providing the diallyl protected nucleosides **5** and **6**, which were easily separable on silica gel. The structure of regioisomers **5** and **6** were assigned based on conversion of a sample of purified **3** exclusively to **5** under the reaction conditions. In addition, in the ¹H NMR spectrum, the H8 proton of **6** resonates downfield of the H8 resonance of **5**, in accord with correlations observed in an array of 7- and 9-substituted purines (8). The 5'-hydroxyl of **6** was protected by treatment with 4,4'-dimethoxytrityl chloride in pyridine, generating **7**. Phosphitylation of **7** provided d⁷X phosphoramidite **8**, suitable for incorporation into oligonucleotides by automated DNA synthesis protocols (9). A sample of d⁷X nucleoside, **9**, was prepared by deprotection of **6** under conditions developed by Noyori and coworkers (10), followed by reversed phase HPLC purification.

Synthesis of d⁷A phosphoramidite and d⁷G phosphoramidite. d⁷A phosphoramidite **17** was synthesized as previously described by Seela and coworkers (11). (Figure 8) A protocol for preparation of 4-amino-5-cyanoimidazole **11** is included in the experimental section, since this is not readily available in the literature. d⁷G phosphoramidite, **26**, was synthesized by a slightly modified literature procedure (12) (Figure 9), rather than by the route described in Chapter Six, since the route shown in Figure 9 avoids the difficult separation of α and β anomers in our previous synthesis of **26**. Coupling of tetraacetylribose **18** and N²-isobutyrylguanine **19** (13) with SnCl₄ in CH₃CN, followed by partial hydrolysis provides the previously reported **20** directly. Subsequent steps were performed exactly as described (12).

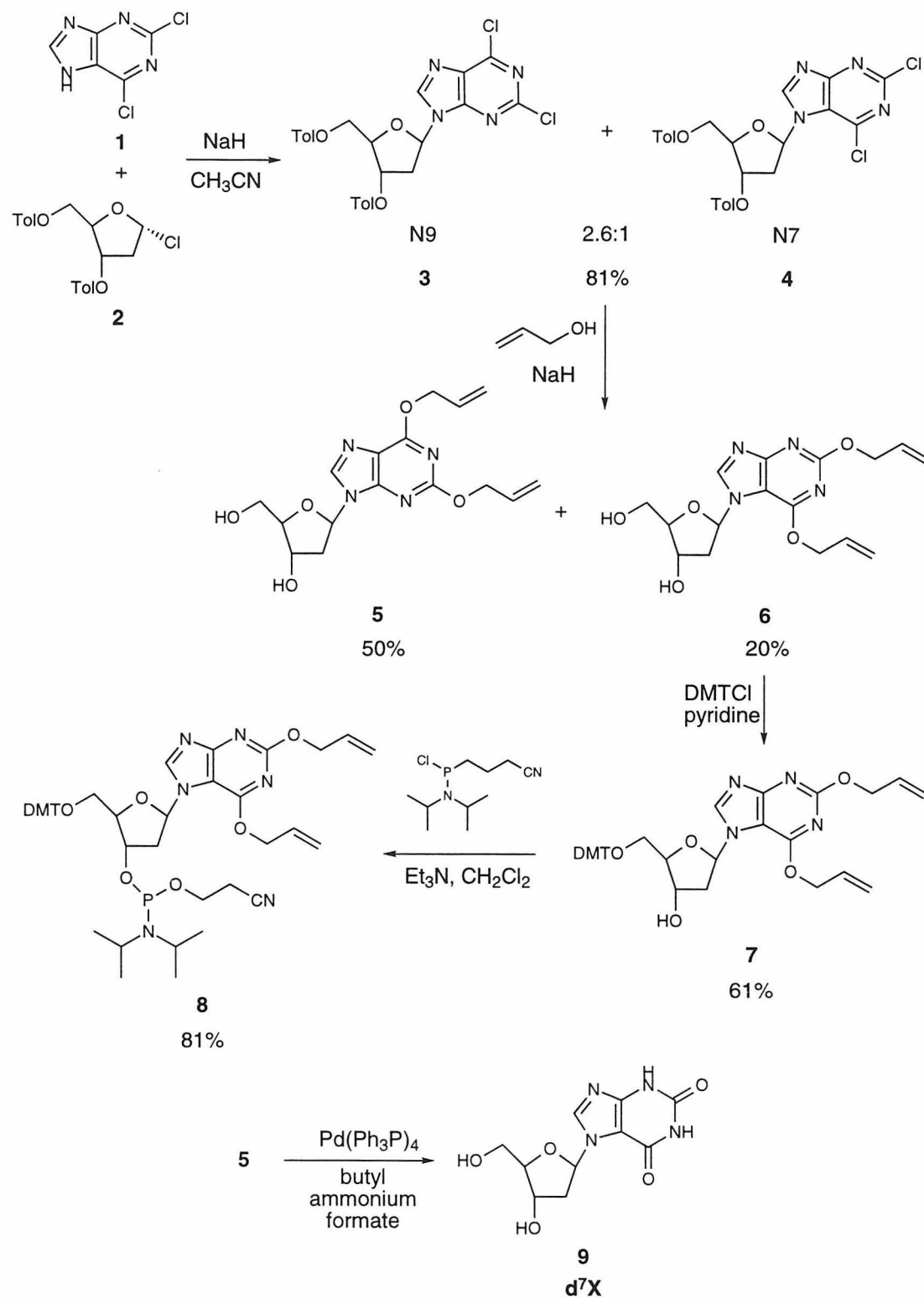


Figure 7: Synthesis of d⁷X phosphoramidite **9**.

Figure 8: Synthesis of d⁷A phosphoramidite **17**. The indicated yields were obtained in this work.

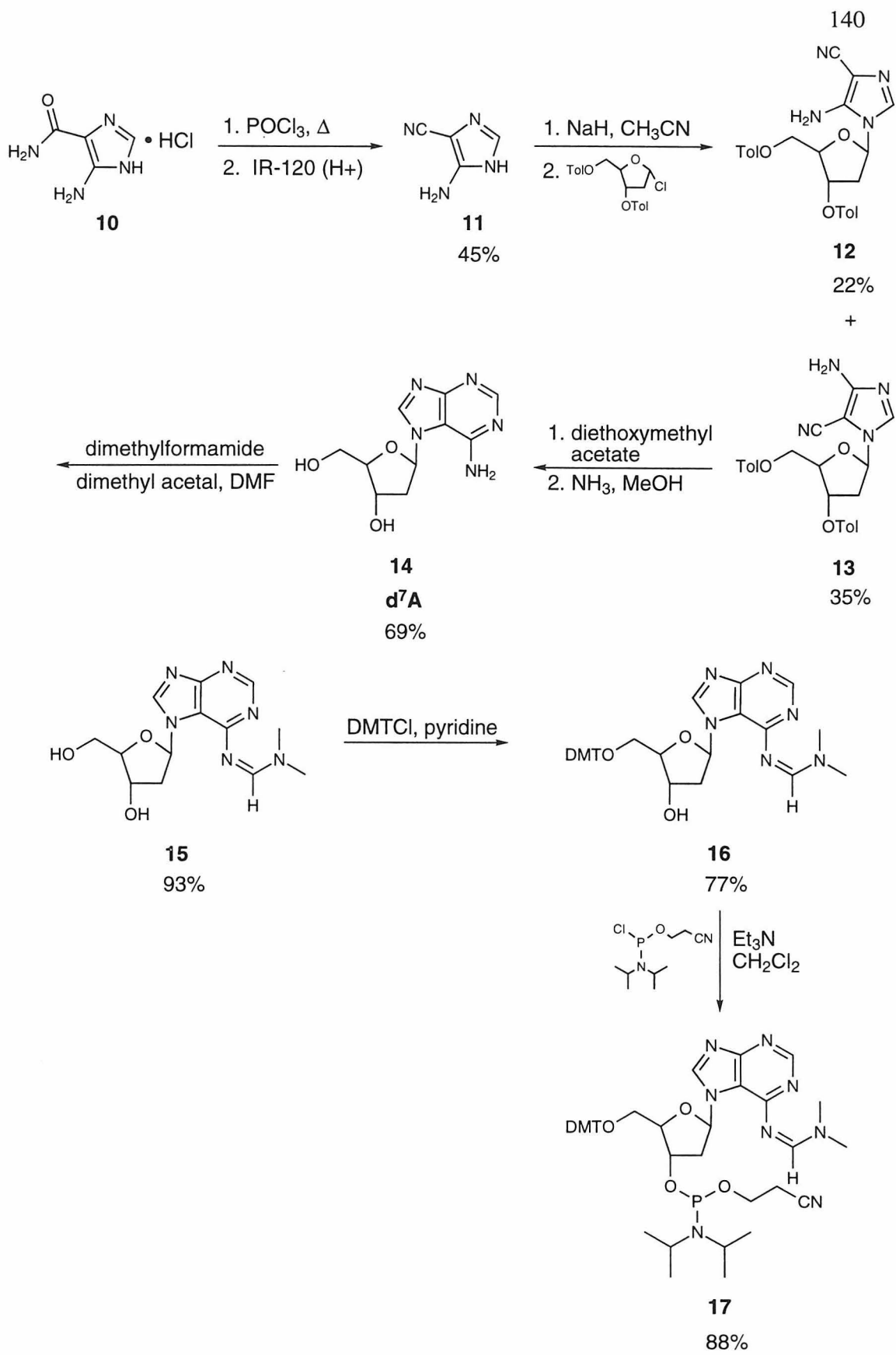
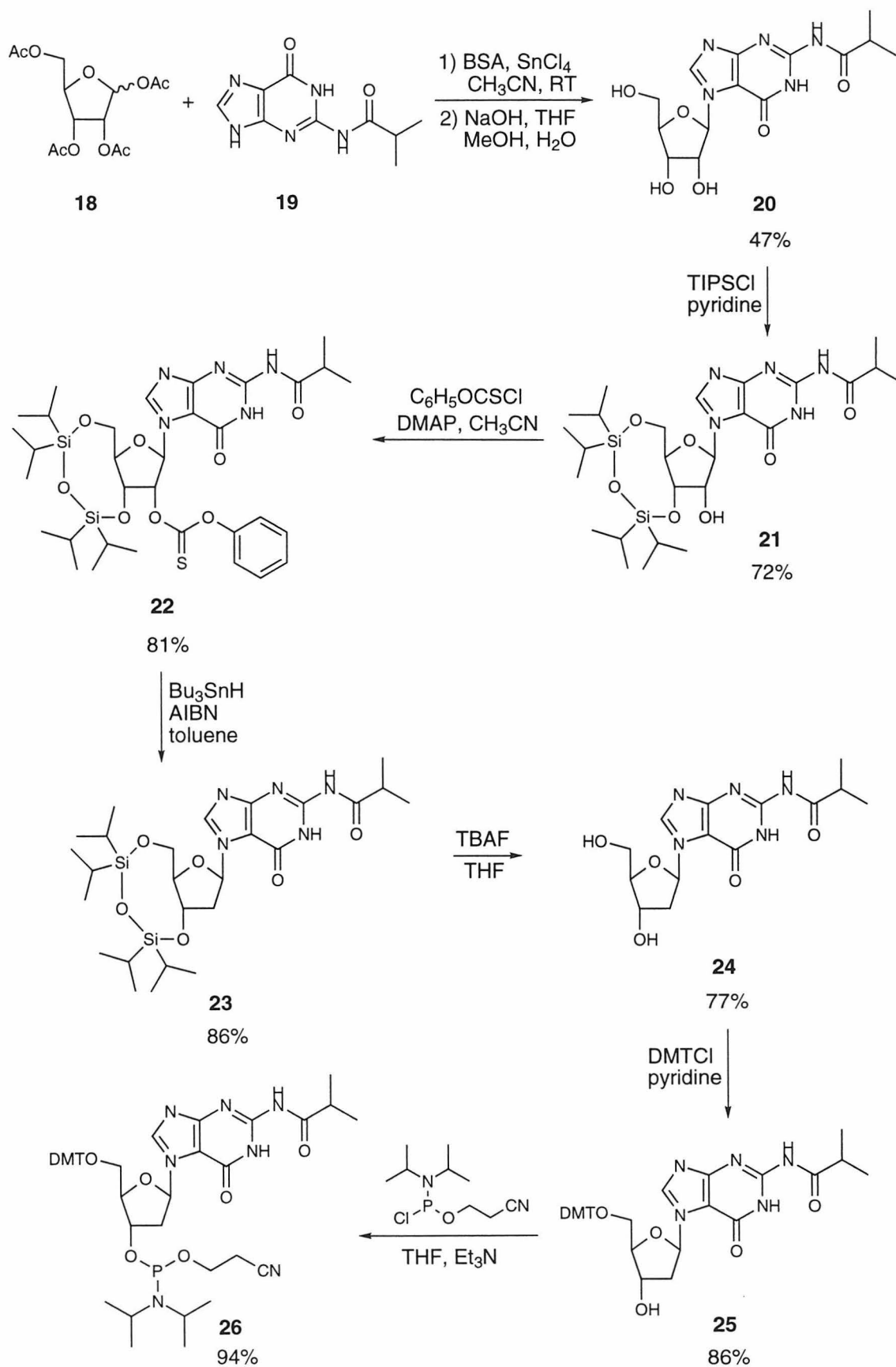


Figure 9: Synthesis of d⁷G phosphoramidite **26**. Indicated yields are those obtained in this work.



Synthesis of d⁷X controlled pore glass. d⁷X controlled pore glass (CPG) was synthesized according to a recently reported, improved procedure for functionalized CPG synthesis (Figure 10) (14). Succinic acid derivatized CPG **28** was prepared from long chain alkyl amine CPG and succinic anhydride. DMT protected d⁷X **7** was coupled to **28** in the presence of EDC, triethylamine, DMAP, and pyridine to give d⁷X CPG **29**. Untreated succinyl groups were esterified with pentachlorophenol, which was subsequently displaced with piperidine. Unfunctionalized nucleophilic groups were then capped with acetic anhydride in pyridine. The loading of **29** was determined to be 26 $\mu\text{mol/g}$ by dimethoxytrityl cation assay (9).

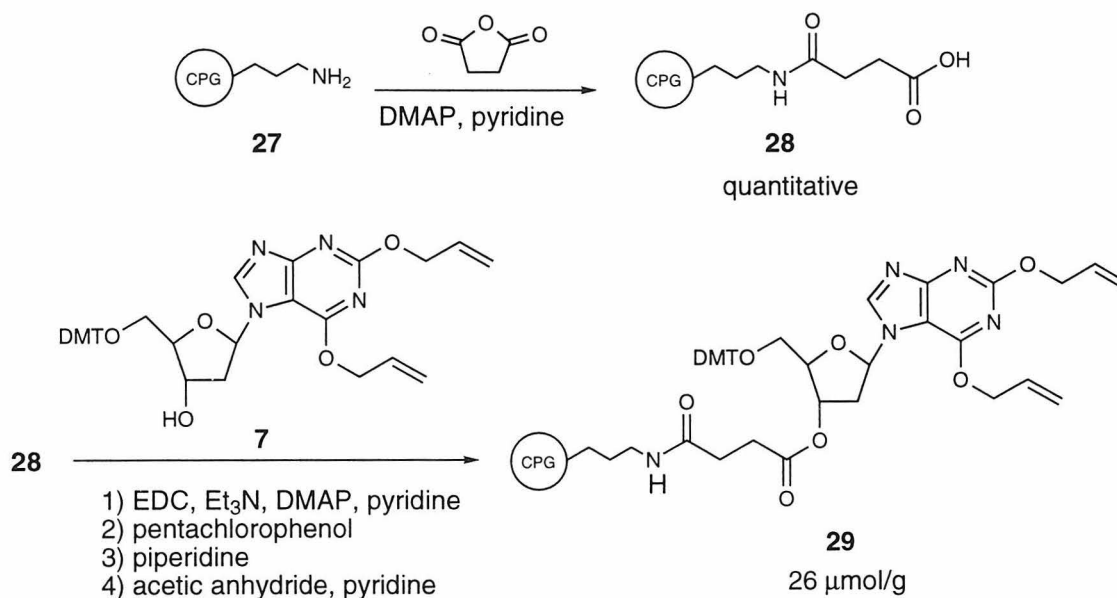


Figure 10: Synthesis of d⁷X controlled pore glass **29**.

Synthesis of d⁷A CPG. d⁷A CPG was prepared by the standard method (15) as shown in Figure 11. DMT protected d⁷A **16** was acylated with succinic anhydride to give **30** in 50% yield. **30** was activated as the *p*-nitrophenyl ester and directly coupled to long chain alkyl amine CPG in the presence of triethylamine and DMF to give **31**. The loading of **31** was determined to be 23 $\mu\text{mol/g}$ by dimethoxytrityl cation assay (9).

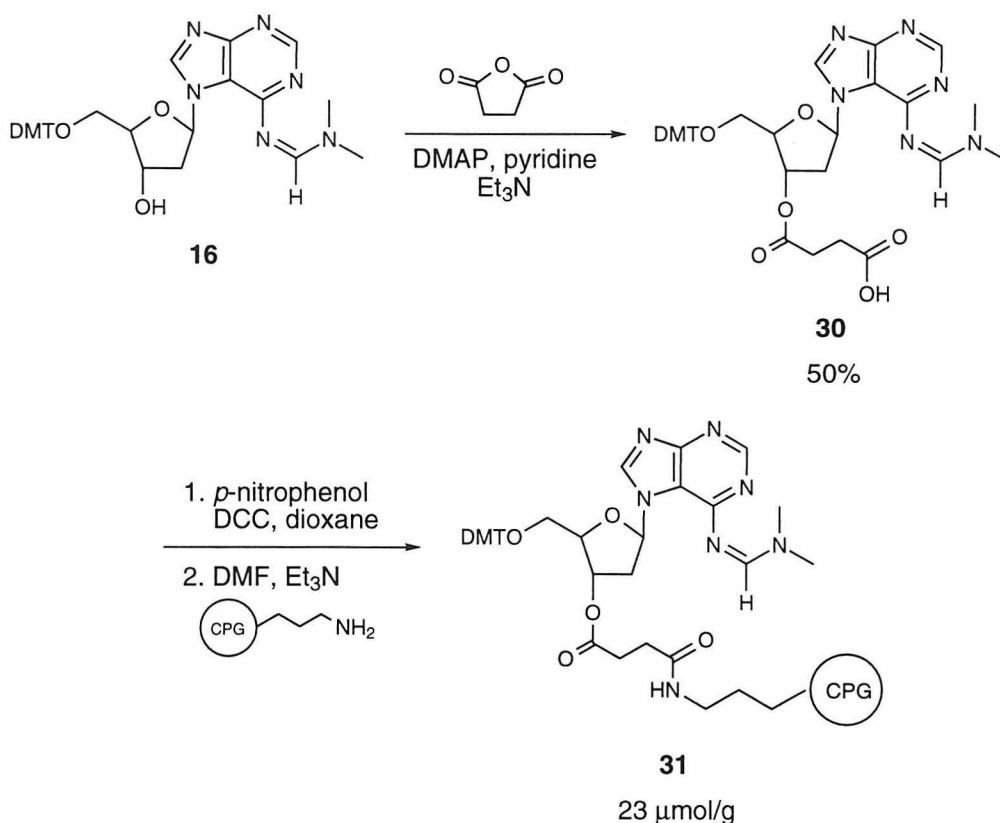


Figure 11: Synthesis of d⁷A controlled pore glass **31**.

Specificity Experiments. The specificity of d⁷X, d⁷iL, and d⁷A were determined with the assay shown in Figure 12. This DNase footprinting assay is described in detail in Chapter Three.

Oligonucleotide Synthesis. Oligonucleotides **32**, **33**, and **34** were prepared by standard automated methods. Each of the novel phosphoramidites coupled in >97% yield as measured by dimethoxytrityl cation assay (9).

Oligonucleotide Analysis. Oligonucleotides **32** and **34**, containing d⁷X and d⁷A, respectively, were 5'-³²P end labeled and analyzed by electrophoresis on denaturing 20% polyacrylamide gels (Figure 13). Quantitation of the storage phosphor autoradiograms revealed that oligonucleotides **32** and **34** were >98% and >90% pure, respectively, with respect to failure sequences. These two oligonucleotides were then digested to their component nucleosides by treatment with snake venom phosphodiesterase and alkaline

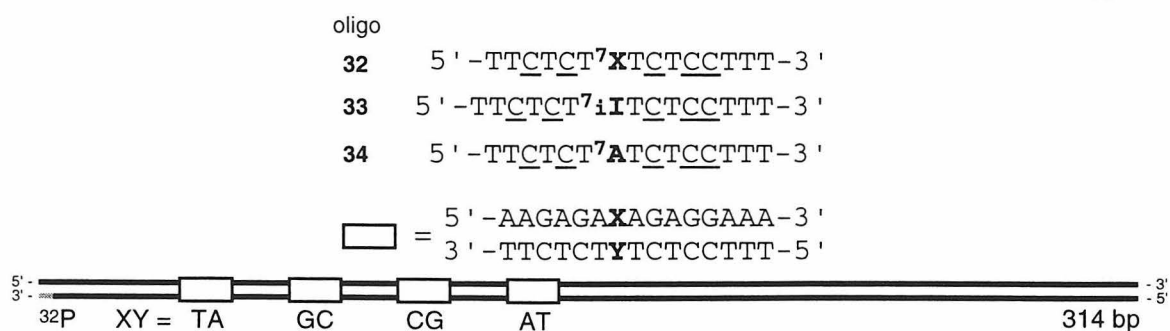


Figure 12: Sequences of the oligonucleotides containing d⁷X, d⁷iI, and d⁷A, and of the four purine rich target sites for the specificity experiments. The duplex target sites, which differ only at a single position labeled XY, are arranged sequentially on the 314 bp *Afl* II/*Fsp* I restriction fragment of pYSEPC2, each separated by 15 bp of intervening sequence. (C indicates 5-methyl-2'-deoxycytidine)

phosphatase and analyzed by reversed phase HPLC. As shown in Figure 14, the expected nucleoside composition was observed for both **32** and **34**. The identities of each peak were confirmed by coinjection with authentic standards of the nucleosides. In addition, both oligonucleotides were subjected to MALDI/TOF mass spectroscopy (16), and satisfactory results were obtained: calculated for oligonucleotide **32**, 4522 (M-1); found, 4521 (M-1); calculated for oligonucleotide **34**, 4504 (M-1); found 4499 (M-1).

DNase Footprinting Experiments. Equilibrium association constants of oligonucleotides **32-34** for the four target sites on the 314 bp *Afl* II/*Fsp* I restriction fragment of pYSPEC2 (Figure 12) in 10 mM bis tris, pH 7.0, 100 mM NaCl, 250 μ M spermine were measured by quantitative DNase I footprinting (17). Figure 15 shows a DNase I footprinting gel for oligonucleotide **32**, containing d⁷X. A strong footprint at the AT containing target site is readily apparent. Figure 16 shows the corresponding gel for oligonucleotide **34**, containing d⁷A. Again, strong AT specificity is observed. Equilibrium binding isotherms for **32** and **34** for each of the four target sites are shown in Figure 17. Table I summarizes the results for the three oligonucleotides.

Both oligonucleotide **32**, containing d⁷X, and oligonucleotide **34**, containing d⁷A,

display at least 100-fold specificity for the AT site. These results are consistent with the formation of two hydrogen bonds in the proposed $d^7X \cdot AT$ and $d^7A \cdot AT$ triplets (Figures 3 and 5). Moreover, the affinity of **34** for the AT site is equal to that of the oligonucleotide with $Z = P$ for the GC site ($K_T = 3.1 \times 10^7 \text{ M}^{-1}$, see Chapter Three). In contrast, oligonucleotide **33**, containing d^7iI , does not show the desired specificity. This is likely due to the tautomeric ambiguity of d^7iI discussed in the Introduction to this chapter. Based on the successful results with d^7X and d^7A , we proceeded to examine oligonucleotides composed entirely of N^7 -purine nucleosides as described in the subsequent section.

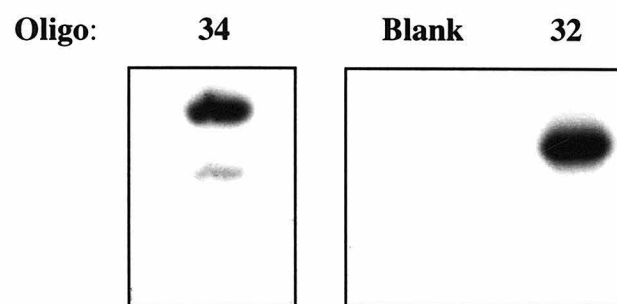
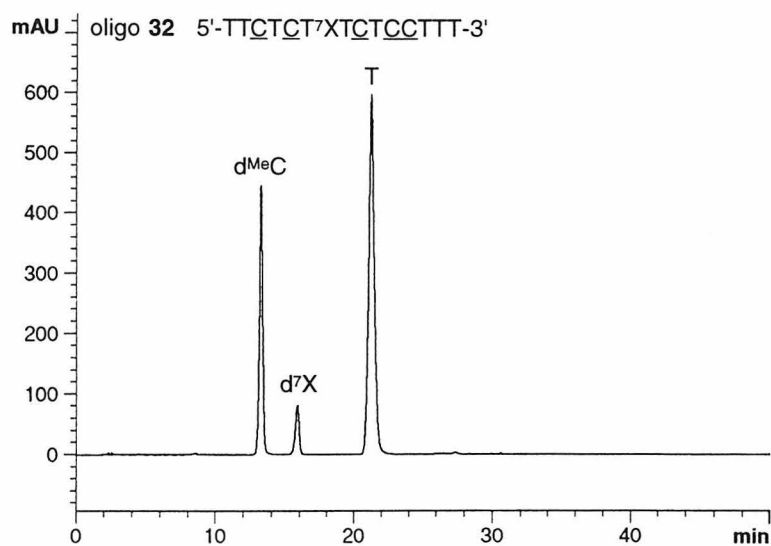


Figure 13: Storage phosphor autoradiograms of 20% denaturing polyacrylamide gels for analysis of 5'-³²P end labeled oligonucleotides containing N⁷-deoxyadenosine (oligo **34**) or N⁷-deoxyxanthosine (oligo **32**). Blank indicates a sample of the labeling reaction containing no oligonucleotide. The dynamic range of both gels is 50 to 10,000 arbitrary intensity units.

A.



B.

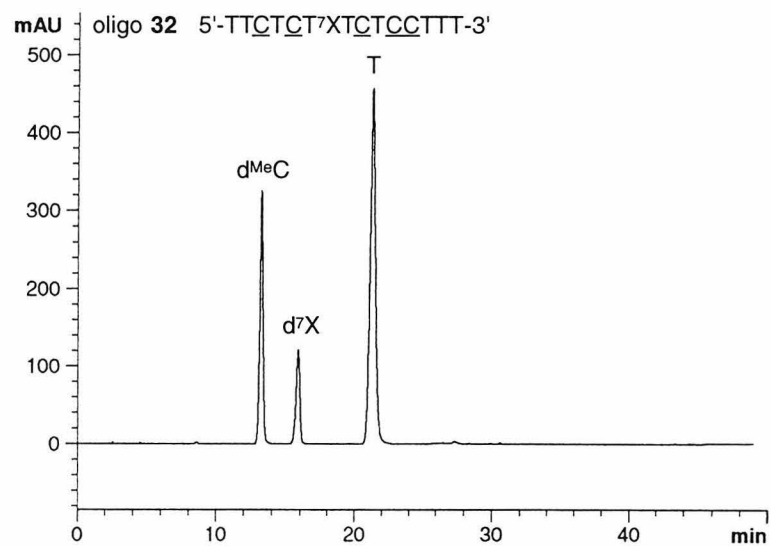
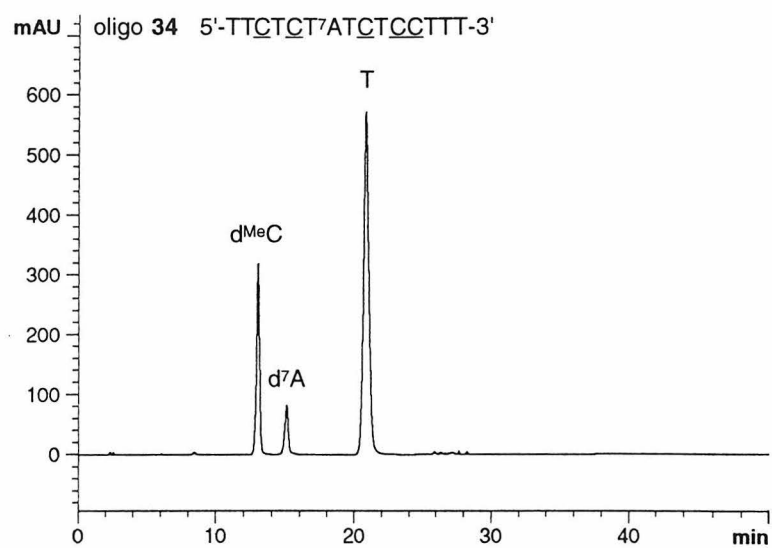


Figure 14: HPLC traces monitored at 270 nm of A) enzymatically digested oligonucleotide **32** and B) a coinjection of the enzyme digest and an equimolar amount of authentic d⁷X.

C.



D.

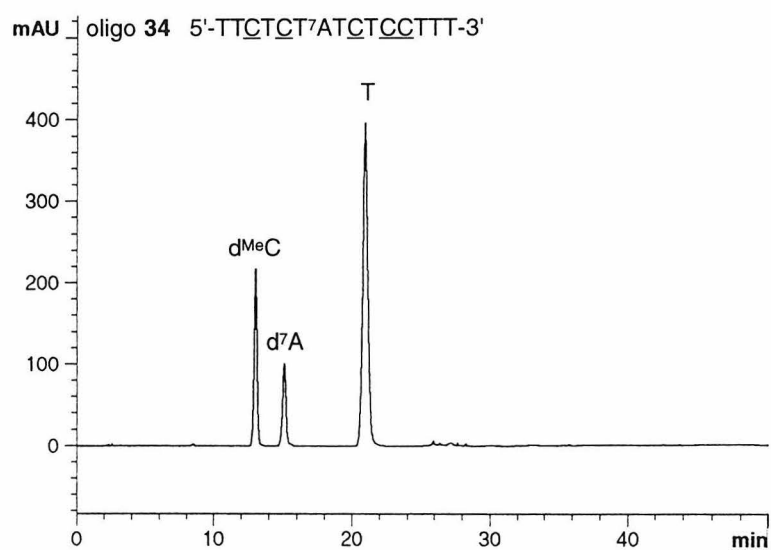


Figure 14: HPLC traces monitored at 260 nm of C) enzymatically digested oligonucleotide **34** and D) a coinjection of the enzyme digest and an equimolar amount of authentic d⁷A.

Figure 15: (Top) Sequences of the oligonucleotide and the four purine rich target sites for the d^7X specificity experiments. The duplex target sites, which differ only at a single position labeled XY, are arranged sequentially on the 314 bp *Afl* II/*Fsp* I restriction fragment of pYSPEC2, each separated by 15 bp of intervening sequence. (C indicates 5-methyl-2'-deoxycytidine). (Bottom) Storage phosphor autoradiogram of an 8% denaturing polyacrylamide gel used to separate the fragments generated by DNase I digestion in a quantitative footprint titration experiment. The boxes drawn to the right of the autoradiogram indicate the positions of the 15 bp binding sites within the 314 bp restriction fragment. (Lanes 1-19) DNase I digestion products obtained in the presence of varying concentrations of oligonucleotide **32**: no oligonucleotide (lane 19); 80 pM (lane 18); 200 pM (lane 17); 400 pM (lane 16); 800 pM (lane 15); 2 nM (lane 14); 4 nM (lane 13); 8 nM (lane 12); 20 nM (lane 11); 40 nM (lane 10); 80 nM (lane 9); 200 nM (lane 8); 400 nM (lane 7); 800 nM (lane 6); 2 μ M (lane 5); 4 μ M (lane 4); 8 μ M (lane 3); 20 μ M (lane 2); 40 μ M (lane 1). (Lane 20) Products of an adenine specific sequencing reaction. (Lane 21) Products of a guanine specific sequencing reaction. (Lane 22) Intact 3' labeled DNA after incubation in the absence of the third strand oligonucleotide.

5' -TTCTCT⁷**X**TCCTCCTTT-3'

5' -AAGAGA**X**AGAGGAAA-3'

3' -TTCTCT**Y**TCCTCCTTT-5'

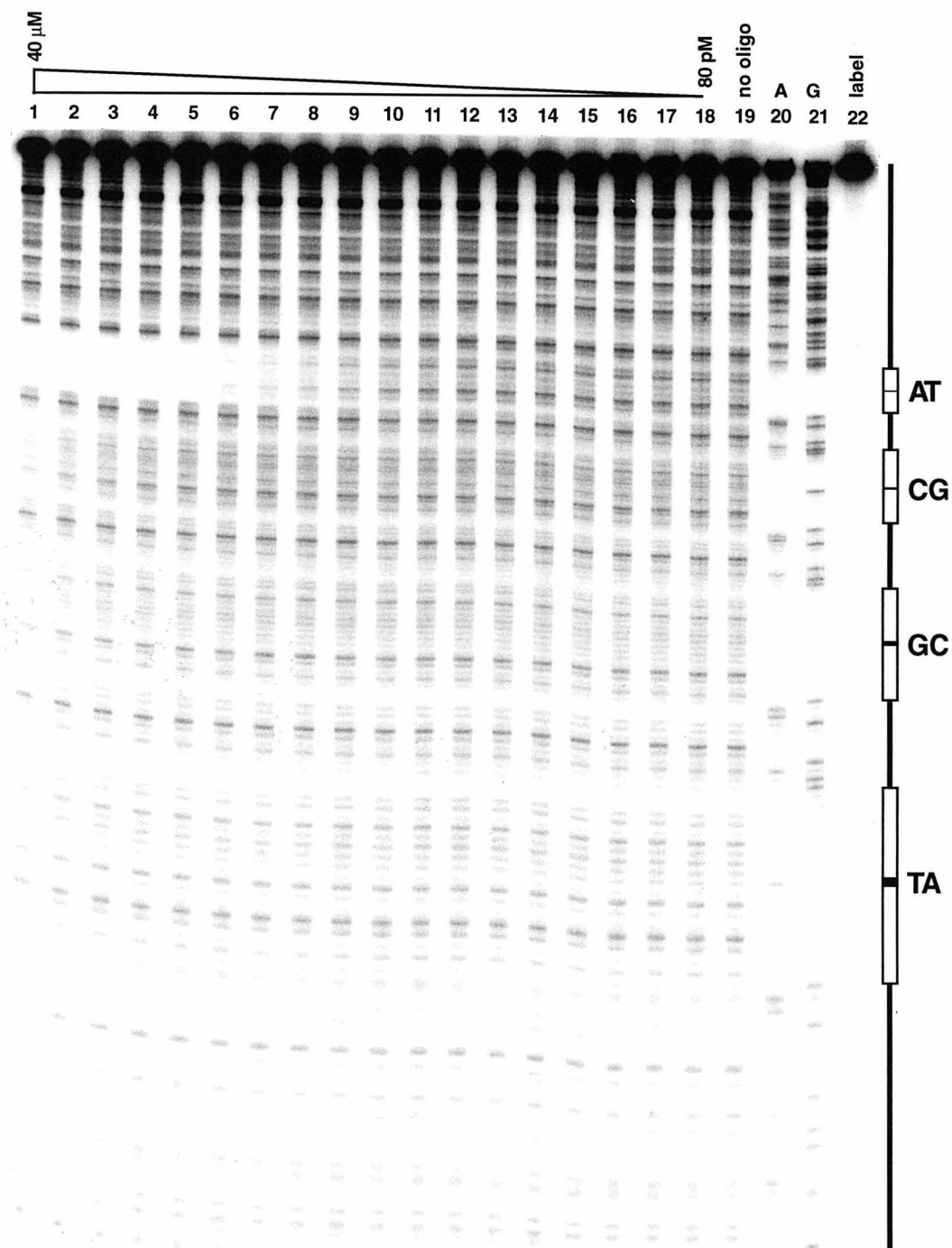


Figure 16: (Top) Sequences of the oligonucleotide and the four purine rich target sites for the d⁷A specificity experiments. The duplex target sites, which differ only at a single position labeled XY, are arranged sequentially on the 314 bp *Afl* II/*Fsp* I restriction fragment of pYSPEC2, each separated by 15 bp of intervening sequence. (C indicates 5-methyl-2'-deoxycytidine). (Bottom) Autoradiogram of an 8% denaturing polyacrylamide gel used to separate the fragments generated by DNase I digestion in a quantitative footprint titration experiment. The boxes drawn to the left of the autoradiogram indicate the positions of the 15 bp binding sites within the 314 bp restriction fragment. (Lane 1) Intact 3' labeled DNA after incubation in the absence of the third strand oligonucleotide. (Lane 2) Products of an adenine specific sequencing reaction. (Lane 3) Products of a guanine specific sequencing reaction. (Lanes 4-22) DNase I digestion products obtained in the presence of varying concentrations of oligonucleotide **34**: no oligonucleotide (lane 4); 80 pM (lane 5); 200 pM (lane 6); 400 pM (lane 7); 800 pM (lane 8); 2 nM (lane 9); 4 nM (lane 10); 8 nM (lane 11); 20 nM (lane 12); 40 nM (lane 13); 80 nM (lane 14); 200 nM (lane 15); 400 nM (lane 16); 800 nM (lane 17); 2 μM (lane 18); 4 μM (lane 19); 8 μM (lane 20); 20 μM (lane 21); 40 μM (lane 22).

5' - TTCTCT⁷ATCTCCTTT - 3'

5' - AAGAGAXAGAGGAAA - 3'

3' - TTCTCT⁷YTCTCCTTT - 5'

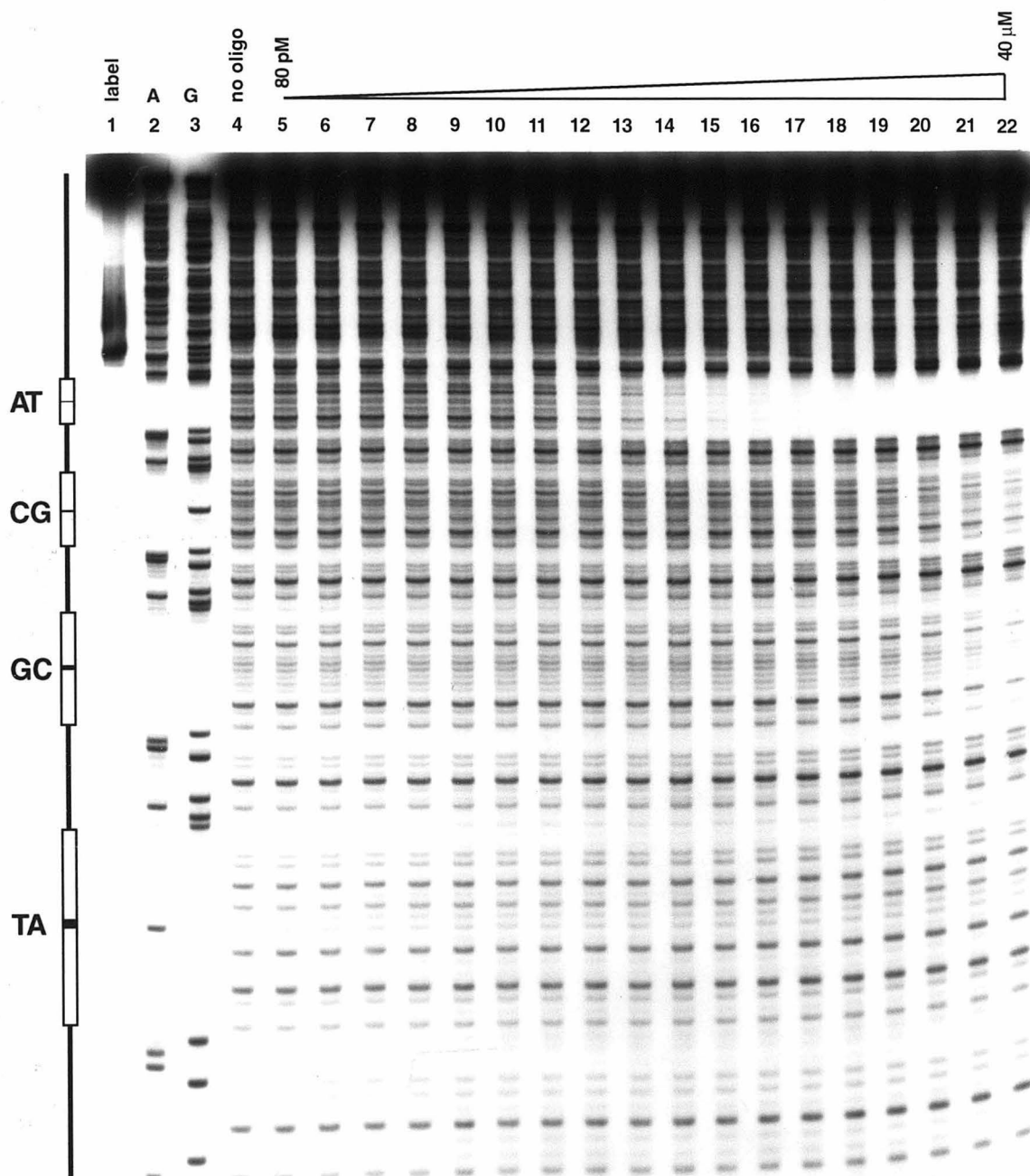


Figure 17: Plots of the site occupancy (θ) versus oligonucleotide concentration for binding of oligonucleotides A) **32** and B) **34** to each of the four target sites on the 314 bp *Afl* II/*Fsp* I restriction fragment of pYSPEC2. Each data set is averaged from at least three gels, and binding isotherms derived from the data are shown as solid lines. Vertical error bars indicate the standard deviation of θ for each data point.

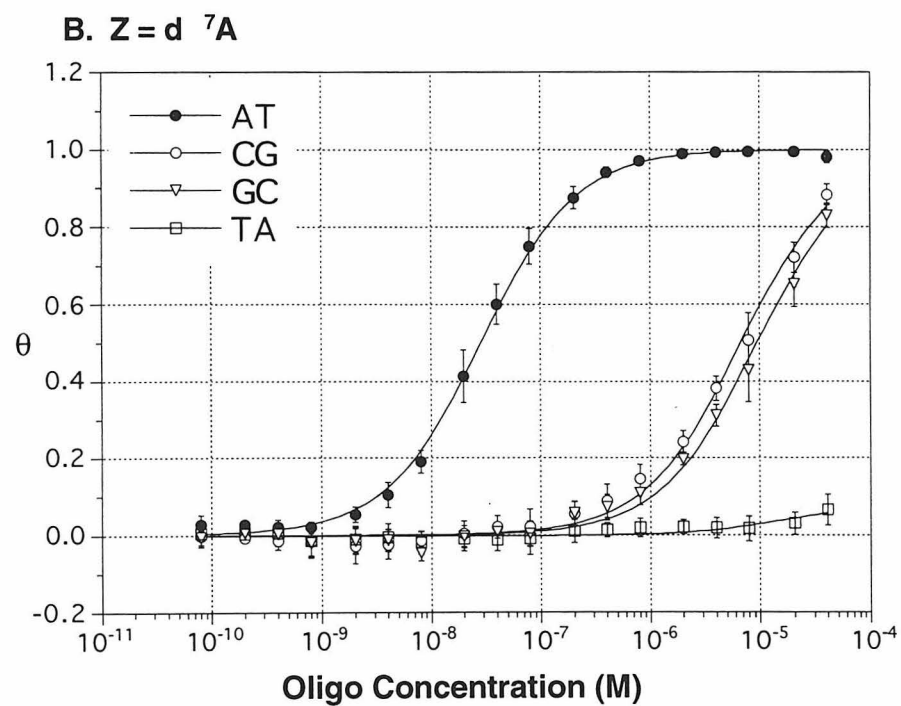
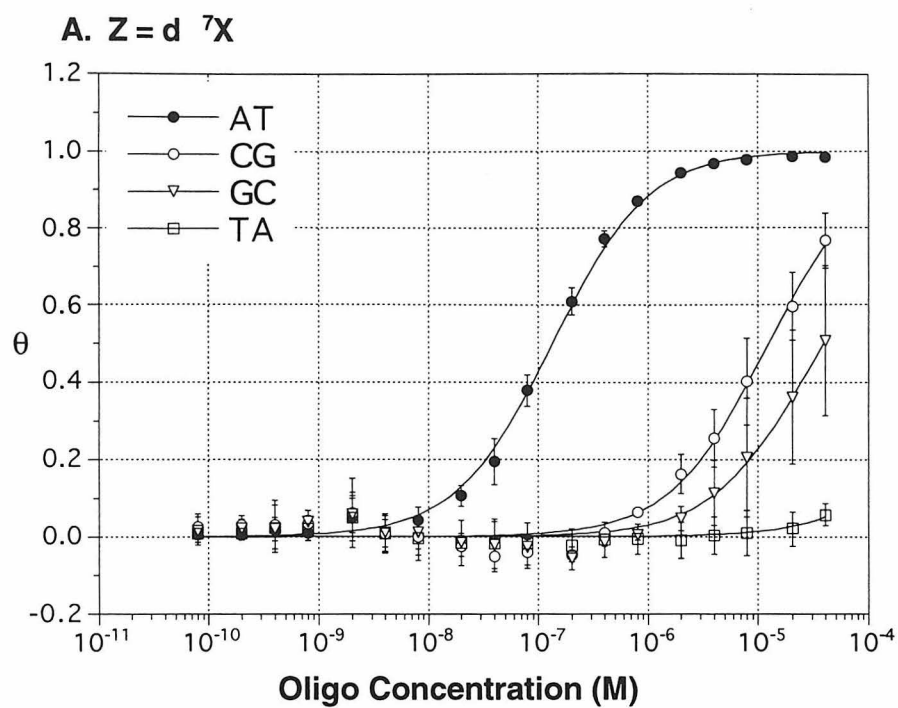


Table I: Equilibrium Association Constants for Triple Helix Formation by Oligonucleotides Containing Single d⁷X, d⁷iI, or d⁷A Bases

oligo	Z ^b	XY	K _T	ΔG° ^c
32	d ⁷ X	AT	7.6 (±0.8) x 10 ⁶	-9.3 (±0.1)
		CG	8.5 (±2.1) x 10 ⁴	-7.0 (±0.5)
		GC	< 5 x 10 ⁴	> -6.3
		TA	< 10 ⁴	> -5.4
33	d ⁷ iI	AT	4.5 (±1.0) x 10 ⁵	-7.6 (±0.1)
		CG	< 5 x 10 ⁴	> -6.3
		GC	2.7 (±0.1) x 10 ⁶	-8.7 (±0.1)
		TA	< 10 ⁴	> -5.4
34	d ⁷ A	AT	3.6 (±0.4) x 10 ⁷	-10.2 (±0.1)
		CG	1.5 (±0.1) x 10 ⁴	-5.6 (±0.1)
		GC	1.1 (±0.1) x 10 ⁴	-5.4 (±0.1)
		TA	< 10 ⁴	> -5.4

^a Each reported K_T value is the mean (±SEM) of three independent measurements which were performed in 100 mM NaCl, 10 mM bis-tris, 250 μM spermine at pH 7.0 and 22° C. Values for K_T and ΔG° are reported in units of M⁻¹ and kcal·mol⁻¹, respectively. ^b Z refers to the position in the third strand oligonucleotide opposite the variable position (XY) in the duplex. ^c ΔG° values were obtained from the relationship ΔG° = -RTlnK (T = 295 K) and error estimates were obtained by propagating the SEM for K_T through this formula.

Sequence Composition Experiments. Based on the favorable results from the specificity experiments, a series of oligonucleotides composed entirely of either d⁷X or d⁷A along with d⁷G were synthesized in order to test our proposed N⁷-purine motif for triple helix formation. The oligonucleotides were designed to bind parallel to the homopurine strand of the target duplexes, forming either d⁷X•AT or d⁷A•AT and d⁷G•GC base triplets. In addition to the N⁷ purine motif oligonucleotides, three series of control oligonucleotides were also prepared: pyrimidine motif (designed to bind parallel) controls composed of T and MeC, purine motif (designed to bind antiparallel) controls composed of T and G, and purine motif controls composed of A and G. Finally, a series of oligonucleotides designed to bind parallel to the homopurine strand of the target duplexes and composed of T (for AT recognition) and d⁷G (for GC recognition) were also prepared. The experiments are summarized in Figure 18. The three duplex target sequences were chosen to be representative of various classes of homopurine sequences. The first sequence, referred to as the “Moser” sequence for its historical significance in the Dervan group, is rich in AT base pairs and contains no contiguous GC base pairs, making it a very good target site for the pyrimidine motif. The second sequence, referred to as the “HIV sequence” since it is present in the HIV genome, is balanced between AT and GC base pairs, and contains six contiguous GC base pairs. It is a very poor pyrimidine motif target site, and is only a weak purine motif target, since it does not contain a large proportion of GC base pairs. The third sequence, referred to as the “GC rich” sequence due to its preponderance of GC base pairs, is an excellent purine motif site. These three sequences were cloned into the plasmid pSP6 (Figure 19), derived from pUC19, in order to allow preparation of a 3'-³²P end labeled *EcoR* I/*Pvu* II restriction fragment containing all three target sites. The sites were cloned into the same plasmid simply for the experimental simplicity of having to label only one restriction fragment and of being able to visualize DNase footprinting results for all three sequences on one gel.

A. "Moser" Sequence

oligo

35 5' - TTTTTCTCTCTCTCT - 3'

36 3' - TTTTGTGTGTGTGT - 5'

37 3' - AAAAAGAGAGAGAGA - 5'

38 5' - ⁷X⁷X⁷X⁷X⁷G⁷X⁷G⁷X⁷G⁷X⁷G⁷X⁷ - 3'

39 5' - ⁷A⁷A⁷A⁷A⁷A⁷G⁷A⁷G⁷A⁷G⁷A⁷G⁷A⁷G⁷A⁷ - 3'

40 5' - TTTT⁷GT⁷GT⁷GT⁷GT⁷GT⁷ - 3'

duplex: 5' - TATAT**AAAAAGAGAGAGAG**ATGCAC - 3'

 3' - ATATA**TTTTTCTCTCTCTCT**ACGTG - 5'

B. "HIV" Sequence

oligo

41 5' - TTTCTTTTCCCCCT - 3'

42 3' - TTTGTTTTGGGGGGT - 5'

43 3' - AAAGAAAAGGGGGGA - 5'

44 5' - ⁷X⁷X⁷X⁷G⁷X⁷X⁷X⁷G⁷G⁷G⁷G⁷G⁷X⁷ - 3'

45 5' - ⁷A⁷A⁷A⁷G⁷A⁷A⁷A⁷A⁷G⁷G⁷G⁷G⁷G⁷A⁷ - 3'

46 5' - TTT⁷GTTTT⁷G⁷G⁷G⁷G⁷G⁷GT⁷ - 3'

duplex: 5' - TTAAT**AAAGAAAAGGGGGG**ATGGGA - 3'

 3' - AATTA**TTTCTTTTCCCCCT**ACCCT - 5'

C. "GC Rich" Sequence

oligo

47 5' - TCCCTCCCCCTCCCT - 3'

48 3' - TGGGTGGGGGTGGGT - 5'

49 3' - AGGGAGGGGGAGGGA - 5'

50 5' - ⁷X⁷G⁷G⁷G⁷X⁷G⁷G⁷G⁷G⁷X⁷G⁷G⁷X⁷ - 3'

51 5' - ⁷A⁷G⁷G⁷G⁷A⁷G⁷G⁷G⁷G⁷A⁷G⁷G⁷G⁷A⁷ - 3'

52 5' - T⁷G⁷G⁷GT⁷G⁷G⁷G⁷G⁷GT⁷G⁷G⁷GT⁷ - 3'

duplex: 5' - CGCAT**AGGGAGGGGGAGGGA**TGCAA - 3'

 3' - GCGTA**TCCCTCCCCCTCCCT**ACGTT - 5'

Figure 18: Sequences of oligonucleotides and target duplexes for the sequence composition experiments. The three target sequences are contained on the 306 bp *EcoR* I/*Pvu* II restriction fragment of pSP6. (C indicates 5-methyl-2'-deoxycytidine)

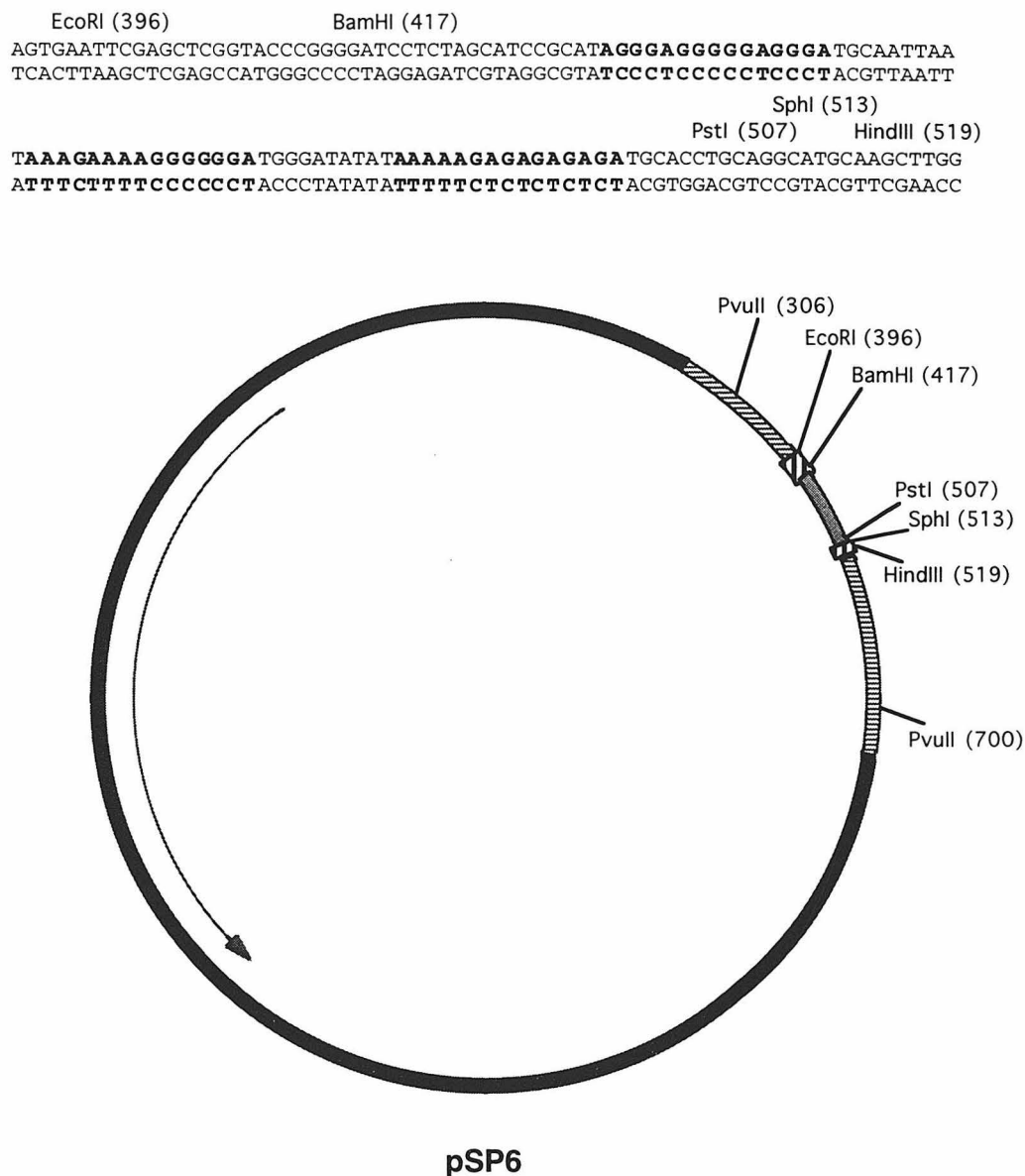


Figure 19: Restriction map of the plasmid pSP6, showing the sequence of the cloned insert containing the Moser, HIV, and GC rich target sequences (indicated in bold type).

Oligonucleotide Synthesis and Analysis. The oligonucleotides listed in Figure 18 were prepared by standard automated methods (9), using d⁷X, d⁷A, and d⁷G phosphoramidites and d⁷X and d⁷A CPG as necessary. The purified oligonucleotides containing nonnatural bases were analyzed by 5'-³²P end labeling followed by

electrophoresis on denaturing 20% polyacrylamide gels (Figure 20). Quantitation of these gels indicated that the oligonucleotides containing d⁷A/d⁷G (**39**, **45**, **51**) or T/d⁷G (**40**, **46**, **52**) were all $\geq 95\%$ pure with respect to failure sequences. The oligonucleotides containing d⁷X/d⁷G varied from $\geq 86\%$ (**38** and **44**) to 98% (**50**) pure. These oligonucleotides were digested to their component nucleosides with snake venom phosphodiesterase and alkaline phosphatase and analyzed by reversed phase HPLC (Figures 21 to 23 show sample data). The expected composition was observed for each oligonucleotide. The identities of each peak were confirmed by coinjection with authentic standards of the nucleosides. Several attempts were made to determine the mass of these oligonucleotides by MALDI/TOF mass spectrometry, but in every case, either the oligonucleotide did not fly, or very poor data (low signal/noise ratio, numerous unassigned peaks) were obtained. These results are typical of MALDI/TOF mass spectrometry results for other purine rich oligonucleotides (18). The 5'-end labeling/gel and enzymatic digestion/HPLC analyses are sufficient, however, to prove that the desired oligonucleotides were indeed prepared.

DNase Footprinting. The binding of oligonucleotides **35-52** to the three homopurine target sites (Moser, HIV, GC) was analyzed by DNase footprinting (17) in 10 mM bis tris, pH 7.0, 100 mM NaCl, 250 μ M spermine. The details of the footprinting protocol are described in Chapter Three. The results are shown in Figures 24-29 and summarized in Table II.

The expected results were observed for the control pyrimidine and purine motif oligonucleotides. The pyrimidine motif (T, ^{Me}C) oligonucleotide **35** bound the Moser sequence with affinity $> 10^8 \text{ M}^{-1}$ (indicated by “****” in Table I), and the corresponding oligonucleotides for the HIV (**41**) and GC (**47**) sequences did not bind (indicated by “–” in Table II) (Figure 24). The purine motif (G, T) oligonucleotides showed high affinity for the GC site (oligo **48**), weak affinity ($> 10^5 \text{ M}^{-1}$, indicated by “*” in Table II) for the HIV site (oligo **42**), and no binding to the Moser sequence (oligo **36**) (Figure 25). Similar results

were obtained for the purine motif (A, G) oligonucleotides **37**, **43**, and **49** (Figure 26).

Surprisingly, no binding was observed at any of the three sites for either the d⁷X, d⁷G oligonucleotides (**38**, **44**, **50**) (Figure 27) or the d⁷A, d⁷G oligonucleotides (**39**, **45**, **51**) (Figure 28). In addition, the d⁷X, d⁷G oligonucleotides were examined for binding under the same solution conditions at pH values of 4.0, 5.0, 6.0 and 8.0, since d⁷X contains ionizable protons involved in triple helix formation. No binding was observed at any site even at the lowest pH values, suggesting that pH is not limiting the binding by these oligonucleotides. A series of experiments were carried out under lower salt conditions (10 mM bis tris, pH 7.0, 30 mM NaCl, 250 μ M spermine), and in this case weak binding was observed at all three sites for the d⁷X, d⁷G oligonucleotides (**38**, **44**, **50**) (Figure 30), while weak binding was observed only at the GC site for the d⁷A, d⁷G oligonucleotides (**39**, **45**, **51**) (Figure 31).

Finally, for the T, d⁷G oligonucleotides (**40**, **46**, **52**), weak binding was observed at the Moser sequence and strong binding at the HIV site, in accord with results described in the Appendix to Chapter Six. However, no binding was observed at the GC site, contrary to our expectations.

The lack of high affinity binding for the d⁷X, d⁷G oligonucleotides (**38**, **44**, **50**) deserves some comment. Since binding was tested over the pH range of 4 to 8, and the pK_a of d⁷X is expected to be near 8 (Figure 2), it is likely that protonation of d⁷X is not limiting triple helix formation. In addition, the d⁷X•AT and d⁷G•GC triplets are both isomorphous and neutral, such that charge-charge repulsion and backbone distortion should not hinder binding. One major factor that was not considered in the design of these oligonucleotides is the base stacking ability of d⁷X. It is possible that 5'-d⁷X/d⁷X-3', 5'-d⁷X/d⁷G-3', and 5'-d⁷G/d⁷X-3' stacking interactions in the third strand all do not contribute favorably to triple helix formation. In the absence of other explanations, poor base stacking by d⁷X is a possible reason for the negative binding results obtained. This suggests that in future experiments with nonnatural nucleosides, base stacking be

considered carefully, perhaps through computational approaches.

For the d⁷A, d⁷G oligonucleotides (**39**, **45**, **51**) the lack of binding is probably due to backbone distortion due to the lack of isomorphism between the d⁷A•AT and the d⁷G•GC base triplets, or to the steric clash between the anomeric proton and exocyclic amine of d⁷A noted previously (Figure 6).

The experiments under lower salt conditions point to the importance of structurally isomorphous base triplets, since the the d⁷X, d⁷G oligonucleotides (designed to form isomorphous d⁷X•AT and d⁷G•GC base triplets) bound target sites with three different sequence compositions, while the d⁷A, d⁷G oligonucleotides (designed to form nonisomorphous d⁷A•AT and d⁷G•GC base triplets) bound only to the GC rich target sequence.

Table II: Qualitative Binding Affinities of Oligonucleotides 35-52 for the Moser, HIV, and GC target sites^a

oligo composition ^c	Duplex Sequences ^b		
	Moser	HIV	GC
T, MeC (35 , 41 , 47)	***	—	—
T, G (36 , 42 , 48)	—	*	***
A, G (37 , 43 , 49)	—	*	***
⁷ X, ⁷ G (38 , 44 , 50)	—	—	—
⁷ A, ⁷ G (39 , 45 , 51)	—	—	—
T, ⁷ G (40 , 46 , 52)	*	***	—

^aBinding was assayed by DNase footprinting in 10 bis tris, pH 7.0, 100 mM NaCl, 250 μM spermine. ^bBinding affinities are indicated qualitatively as: (***) > 10⁸ M⁻¹, (*) > 10⁵ M⁻¹, (—) no binding observed. ^cThis column indicates the nucleoside composition of each class of oligonucleotides. Of the three oligonucleotides listed in each row, the first targets the Moser sequence, the second targets the HIV sequence, and the third targets the GC sequence.

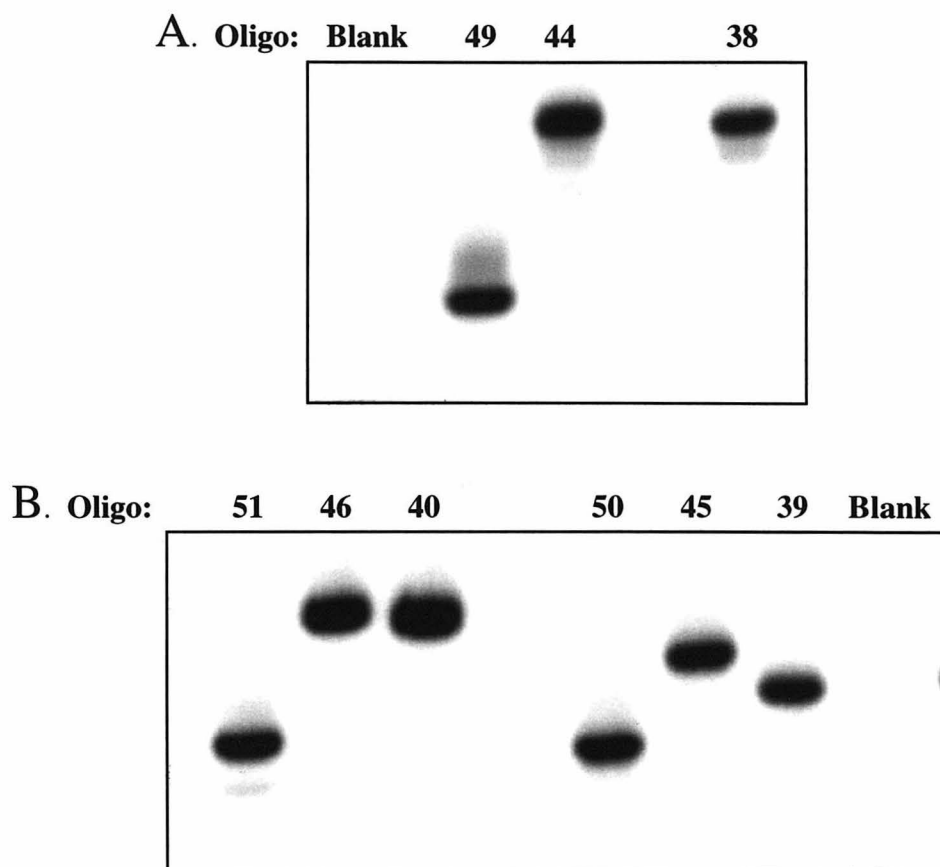
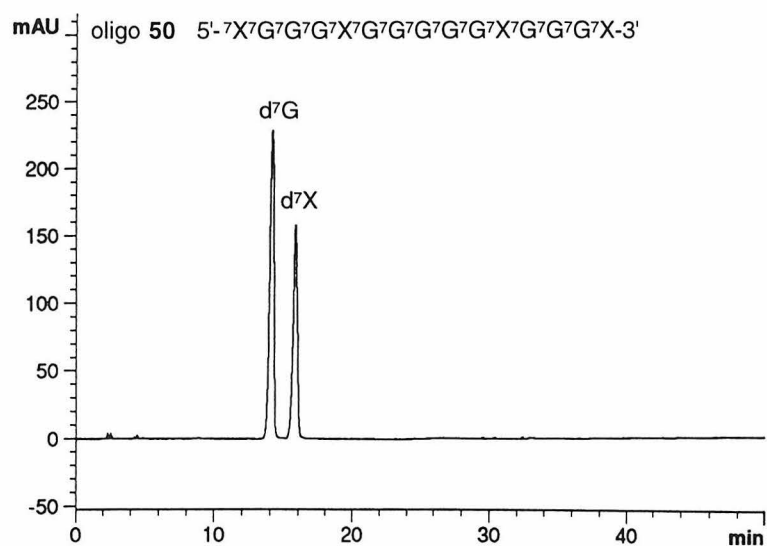


Figure 20: Storage phosphor autoradiograms of 20% denaturing polyacrylamide gels for analysis of 5'-³²P end labeled oligonucleotides containing: (A) N⁷-deoxyxanthosine and N⁷-deoxyguanosine (oligos **38**, **44**, and **49**) and (B) either thymidine and N⁷-deoxyguanosine (oligos **40**, **46**, and **51**) or N⁷-deoxyadenosine and N⁷-deoxyguanosine (oligos **39**, **45**, and **50**). Blank indicates a sample of the labeling reaction containing no oligonucleotide. The dynamic range of gel A is 20 to 4,000 arbitrary intensity units, while gel B covers 5 to 1,000 arbitrary intensity units.

A.



B.

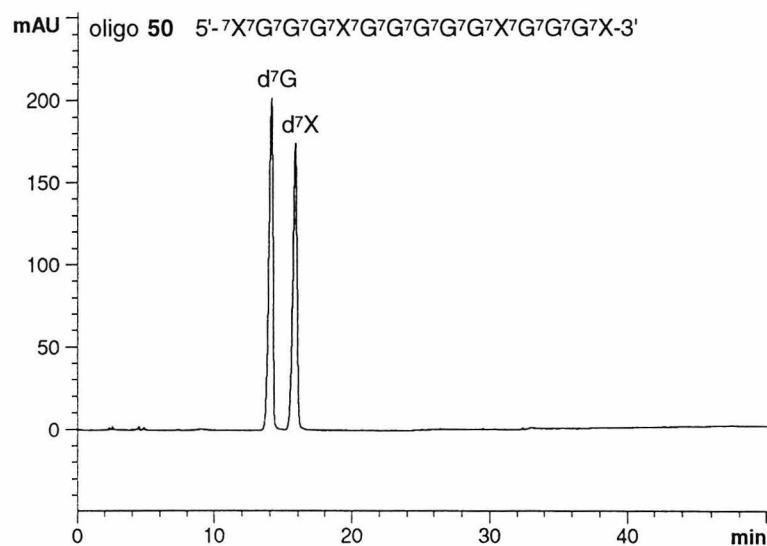
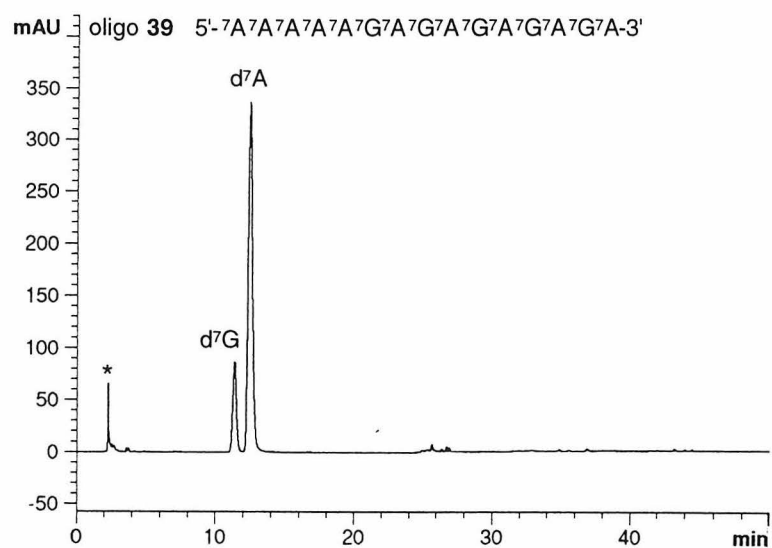


Figure 21: HPLC traces monitored at 260 nm of A) enzymatically digested oligonucleotide **50**, composed of d⁷X and d⁷G and B) a coinjection of the enzyme digest and an equimolar amount of authentic d⁷X and d⁷G.

A.



B.

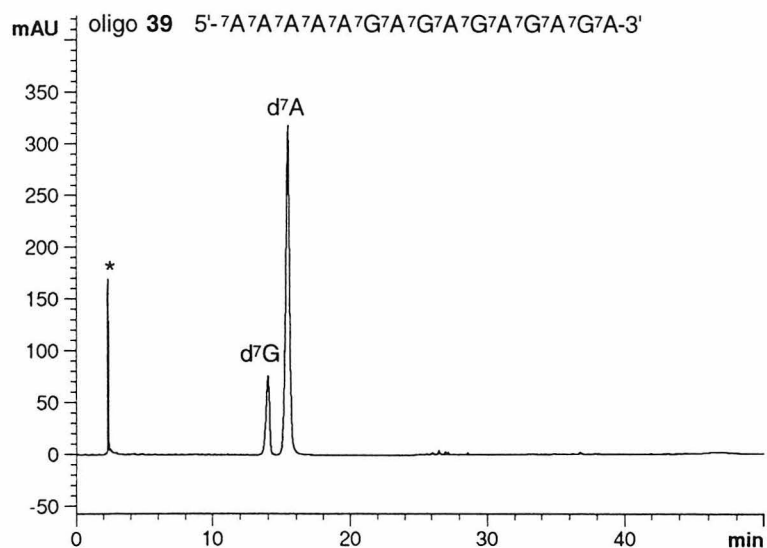
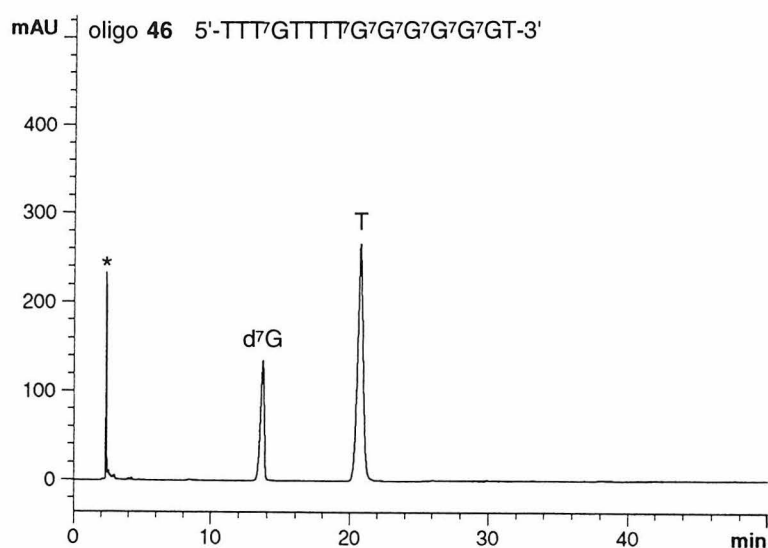


Figure 22: HPLC traces monitored at 260 nm of A) enzymatically digested oligonucleotide **39**, composed of d⁷A and d⁷G and B) a coinjection of the enzyme digest and an equimolar amount of authentic d⁷A and d⁷G. The peaks indicated by an asterisk are salts which elute in the void volume.

A.



B.

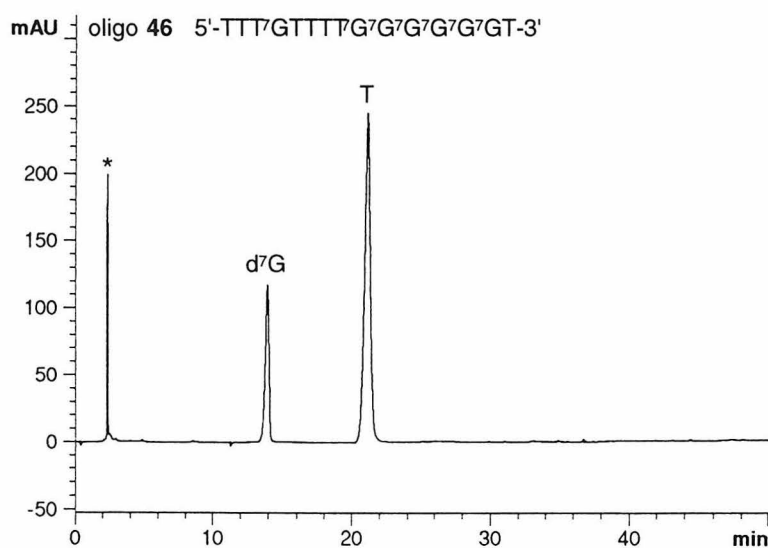


Figure 23: HPLC traces monitored at 260 nm of A) enzymatically digested oligonucleotide **46**, composed of T and d⁷G and B) a coinjection of the enzyme digest and an equimolar amount of authentic T and d⁷G. The peaks indicated by an asterisk are salts which elute in the void volume.

Figure 24: Autoradiogram of a DNase footprinting gel analyzing binding of the pyrimidine motif (T, ^{Me}C) oligonucleotides **35**, **41**, and **47** to the “Moser,” “HIV,” and “GC” duplex homopurine sites. (Lane 1) Intact 3' labeled DNA after incubation in the absence of a third strand oligonucleotide. (Lane 2) Products of an adenine specific sequencing reaction. (Lanes 3-15) DNase I digestion products obtained in the presence of varying concentrations of the indicated oligonucleotides: no oligonucleotide (lane 3); 1 nM oligo **35** (lane 4); 10 nM oligo **35** (lane 5); 100 nM oligo **35** (lane 6); 1 μ M oligo **35** (lane 7); 1 nM oligo **41** (lane 8); 10 nM oligo **41** (lane 9); 100 nM oligo **41** (lane 10); 1 μ M oligo **41** (lane 11); 1 nM oligo **47** (lane 12); 10 nM oligo **47** (lane 13); 100 nM oligo **47** (lane 14); 1 μ M oligo **47** (lane 15). The boxes drawn to the left of the gel indicate the three homopurine duplex sequences.

Moser

HIV

GC

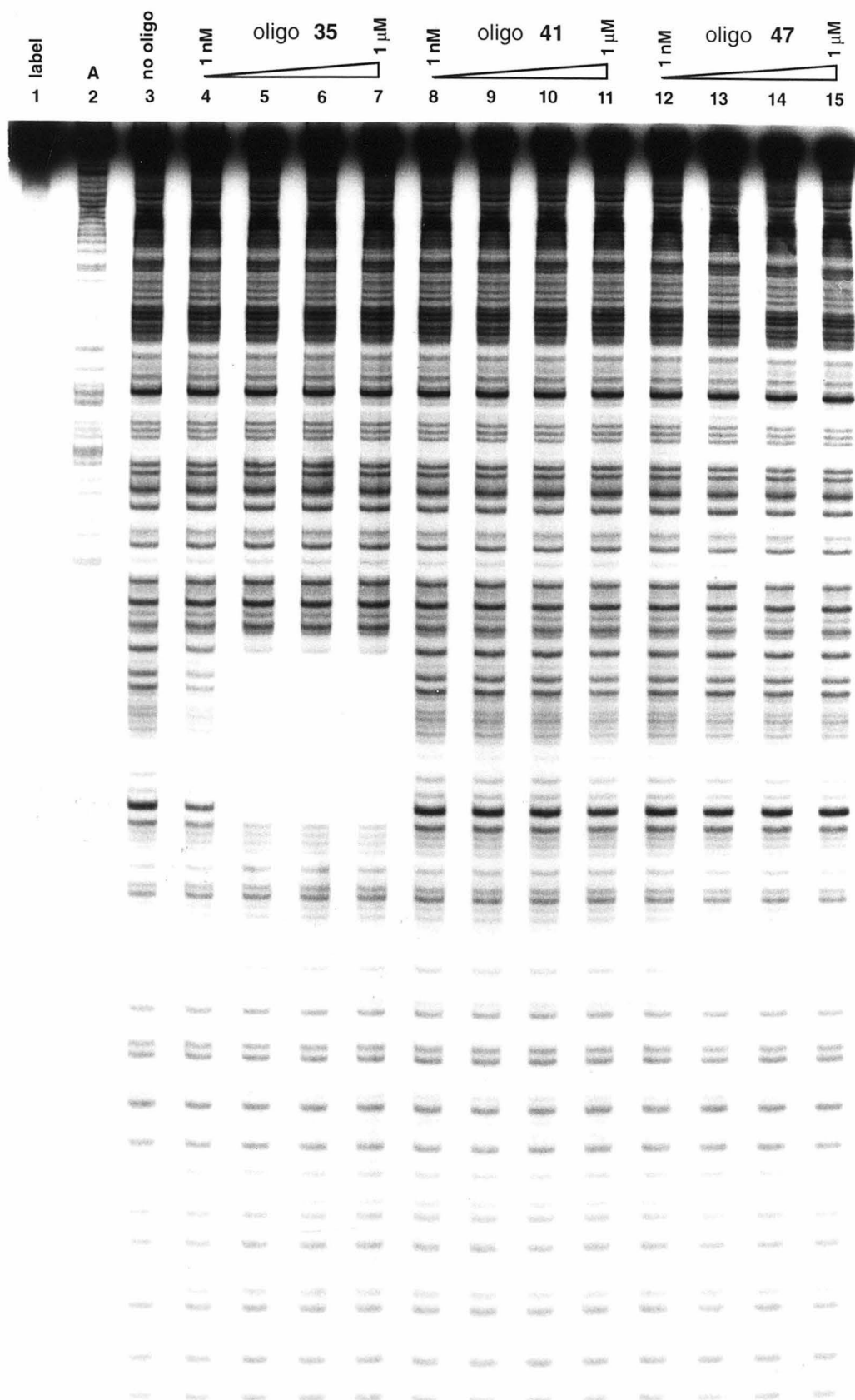


Figure 25: Autoradiogram of a DNase footprinting gel analyzing binding of the purine motif (T, G) oligonucleotides **36**, **42**, and **48** to the “Moser,” “HIV,” and “GC” duplex homopurine sites. (Lane 1) Intact 3' labeled DNA after incubation in the absence of a third strand oligonucleotide. (Lane 2) Products of an adenine specific sequencing reaction. (Lanes 3-18) DNase I digestion products obtained in the presence of varying concentrations of the indicated oligonucleotides: no oligonucleotide (lane 3); 2 nM oligo **36** (lane 4); 20 nM oligo **36** (lane 5); 200 nM oligo **36** (lane 6); 2 μ M oligo **36** (lane 7); 20 μ M oligo **36** (lane 8); 2 nM oligo **42** (lane 9); 20 nM oligo **42** (lane 10); 200 nM oligo **42** (lane 11); 2 μ M oligo **42** (lane 12); 20 μ M oligo **42** (lane 13); 2 nM oligo **48** (lane 14); 20 nM oligo **48** (lane 15); 200 nM oligo **48** (lane 16); 2 μ M oligo **48** (lane 17); 20 μ M oligo **48** (lane 18). The boxes drawn to the left of the gel indicate the three homopurine duplex sequences.

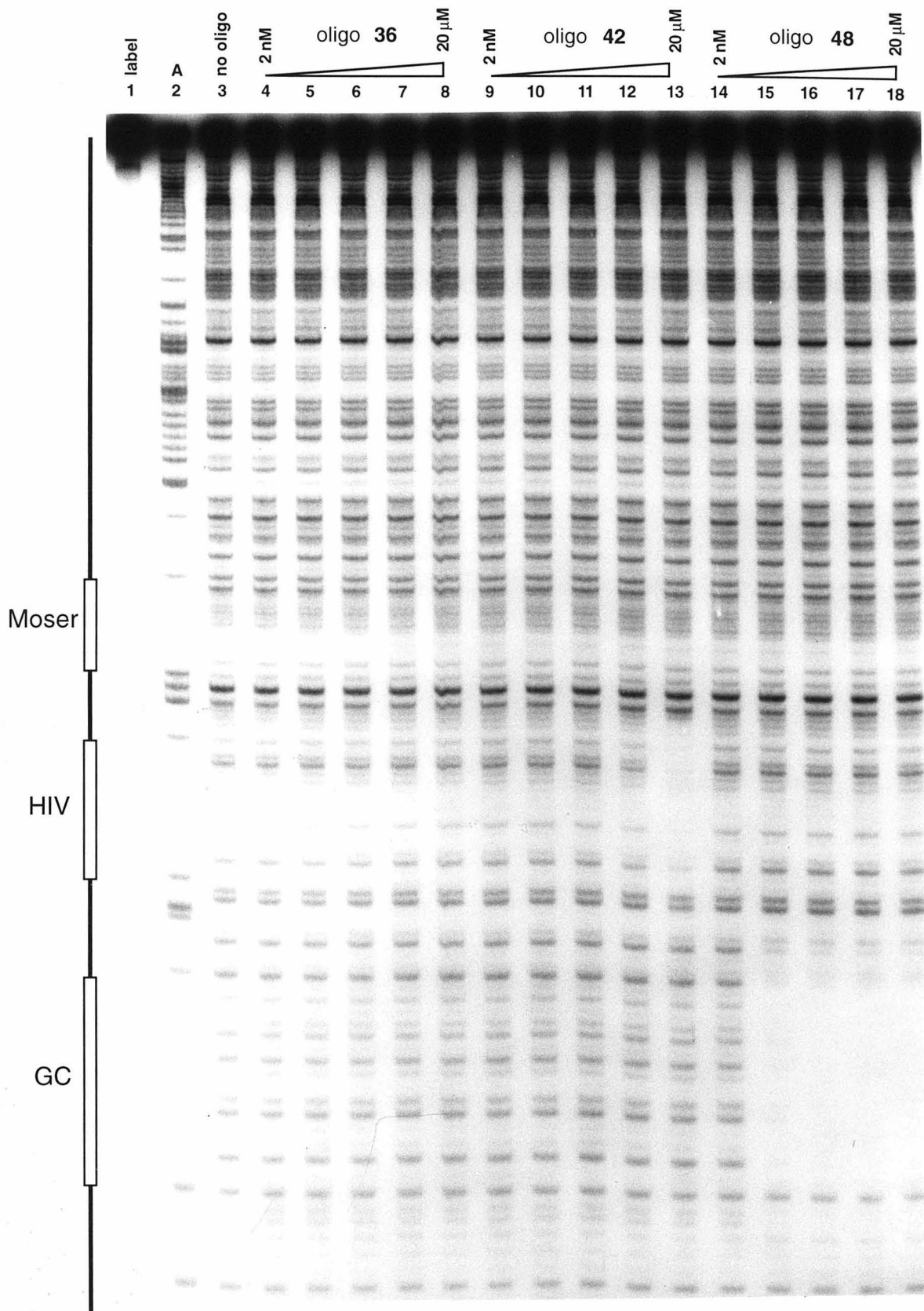


Figure 26: Autoradiogram of a DNase footprinting gel analyzing binding of the purine motif (A, G) oligonucleotides **37**, **43**, and **49** to the “Moser,” “HIV,” and “GC” duplex homopurine sites. (Lane 1) Intact 3' labeled DNA after incubation in the absence of a third strand oligonucleotide. (Lane 2) Products of an adenine specific sequencing reaction. (Lanes 3-18) DNase I digestion products obtained in the presence of varying concentrations of the indicated oligonucleotides: no oligonucleotide (lane 3); 2 nM oligo **37** (lane 4); 20 nM oligo **37** (lane 5); 200 nM oligo **37** (lane 6); 2 μ M oligo **37** (lane 7); 20 μ M oligo **37** (lane 8); 2 nM oligo **43** (lane 9); 20 nM oligo **43** (lane 10); 200 nM oligo **43** (lane 11); 2 μ M oligo **43** (lane 12); 20 μ M oligo **43** (lane 13); 2 nM oligo **49** (lane 14); 20 nM oligo **49** (lane 15); 200 nM oligo **49** (lane 16); 2 μ M oligo **49** (lane 17); 20 μ M oligo **49** (lane 18). The boxes drawn to the left of the gel indicate the three homopurine duplex sequences.

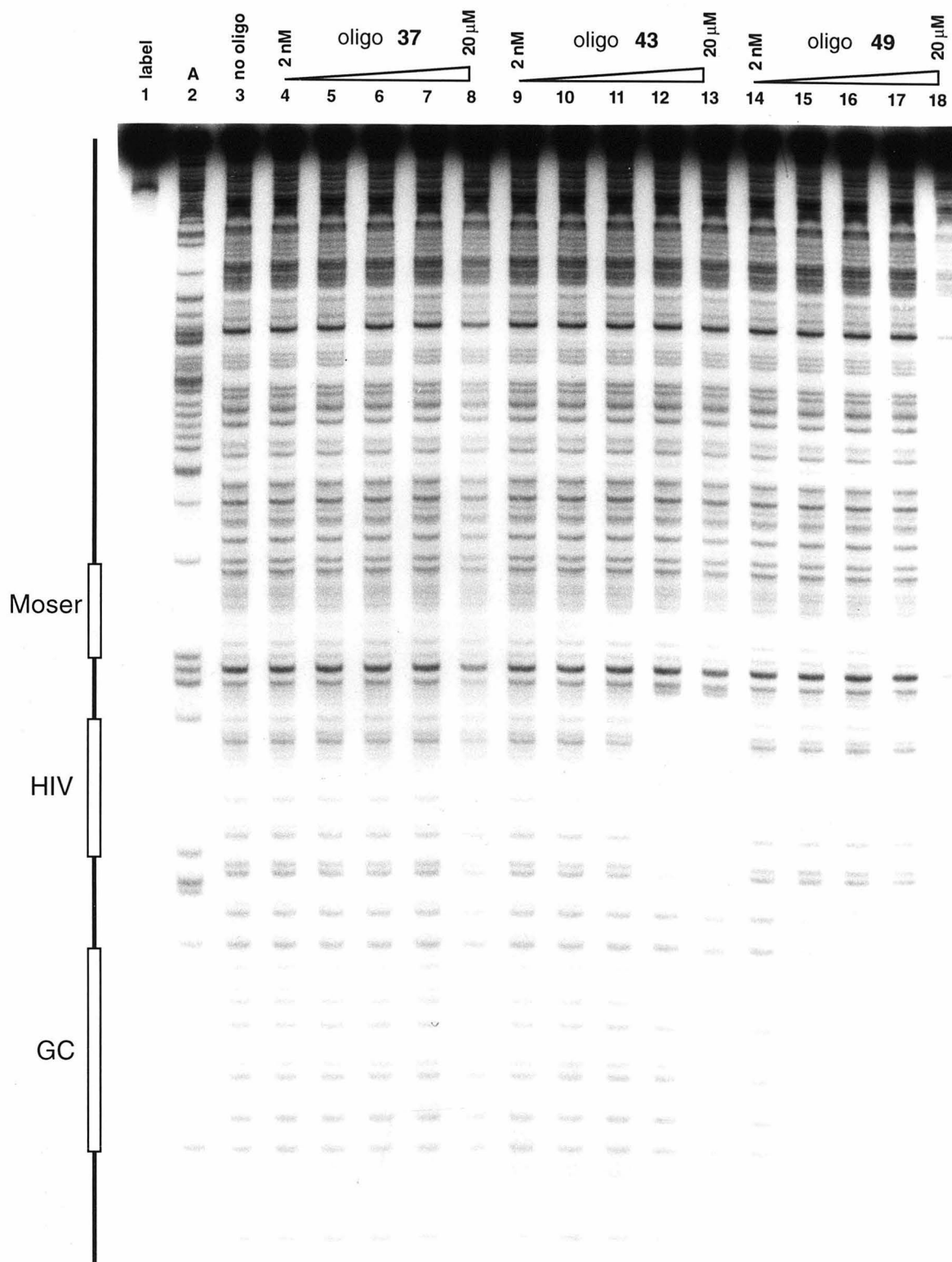


Figure 27: Autoradiogram of a DNase footprinting gel analyzing binding of the d⁷X, d⁷G oligonucleotides **38**, **44**, and **50** to the “Moser,” “HIV,” and “GC” duplex homopurine sites. (Lane 1) Intact 3' labeled DNA after incubation in the absence of a third strand oligonucleotide. (Lanes 2-17) DNase I digestion products obtained in the presence of varying concentrations of the indicated oligonucleotides: 0.4 nM oligo **50** (lane 2); 4 nM oligo **50** (lane 3); 40 nM oligo **50** (lane 4); 400 nM oligo **50** (lane 5); 4 μM oligo **50** (lane 6); 0.4 nM oligo **44** (lane 7); 4 nM oligo **44** (lane 8); 40 nM oligo **44** (lane 9); 400 nM oligo **44** (lane 10); 4 μM oligo **44** (lane 11); 0.4 nM oligo **38** (lane 12); 4 nM oligo **38** (lane 13); 40 nM oligo **38** (lane 14); 400 nM oligo **38** (lane 15); 4 μM oligo **38** (lane 16); no oligonucleotide (lane 17). (Lane 18) Products of an adenine specific sequencing reaction. The boxes drawn to the right of the gel indicate the three homopurine duplex sequences.

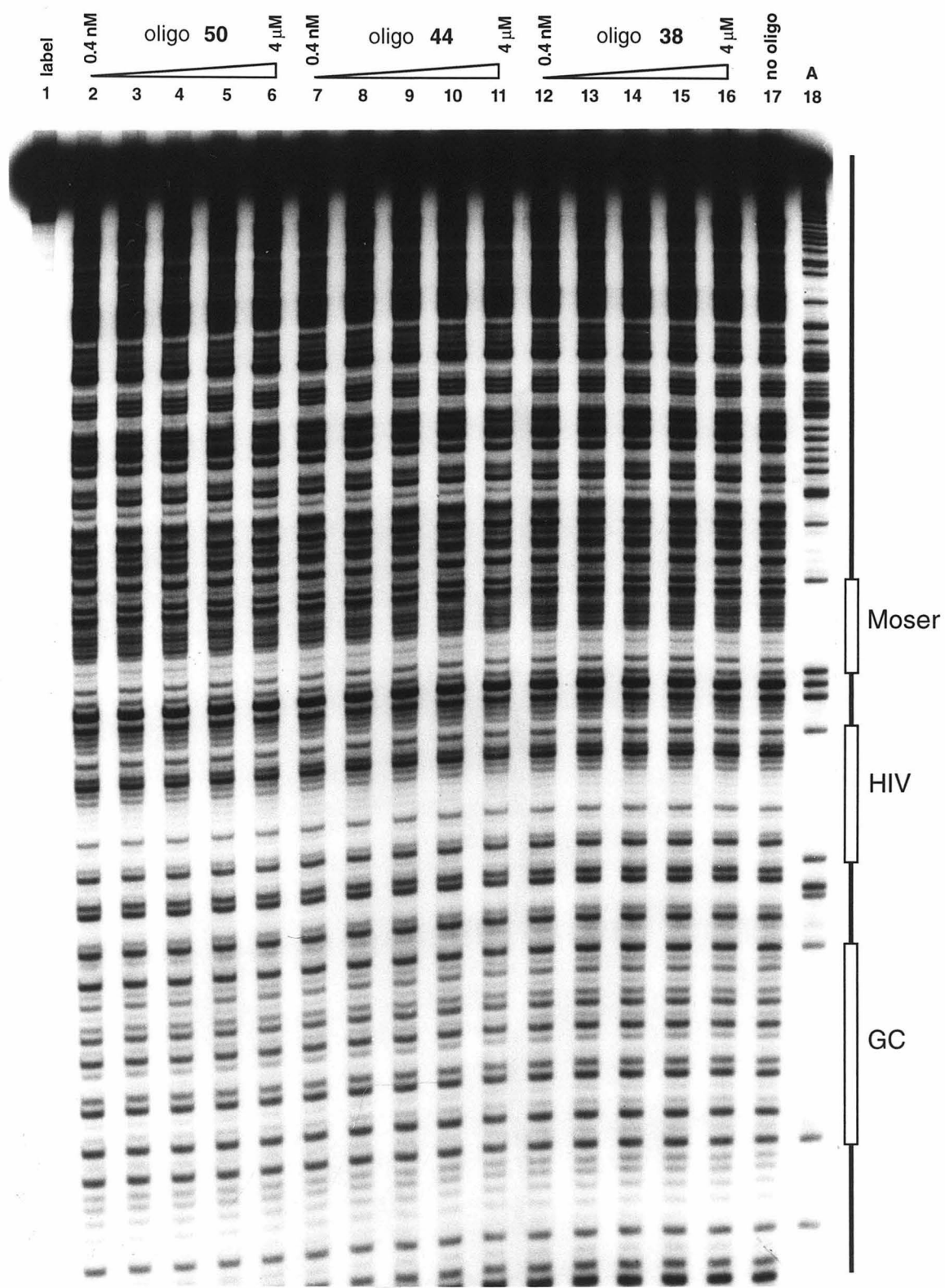


Figure 28: Autoradiogram of a DNase footprinting gel analyzing binding of the d⁷A, d⁷G oligonucleotides **39**, **45**, and **51** to the “Moser,” “HIV,” and “GC” duplex homopurine sites. (Lane 1) Intact 3' labeled DNA after incubation in the absence of a third strand oligonucleotide. (Lane 2) Products of an adenine specific sequencing reaction. (Lanes 3-18) DNase I digestion products obtained in the presence of varying concentrations of the indicated oligonucleotides: no oligonucleotide (lane 3); 2 nM oligo **39** (lane 4); 20 nM oligo **39** (lane 5); 200 nM oligo **39** (lane 6); 2 μM oligo **39** (lane 7); 20 μM oligo **39** (lane 8); 2 nM oligo **45** (lane 9); 20 nM oligo **45** (lane 10); 200 nM oligo **45** (lane 11); 2 μM oligo **45** (lane 12); 20 μM oligo **45** (lane 13); 2 nM oligo **51** (lane 14); 20 nM oligo **51** (lane 15); 200 nM oligo **51** (lane 16); 2 μM oligo **51** (lane 17); 20 μM oligo **51** (lane 18). The boxes drawn to the left of the gel indicate the three homopurine duplex sequences.

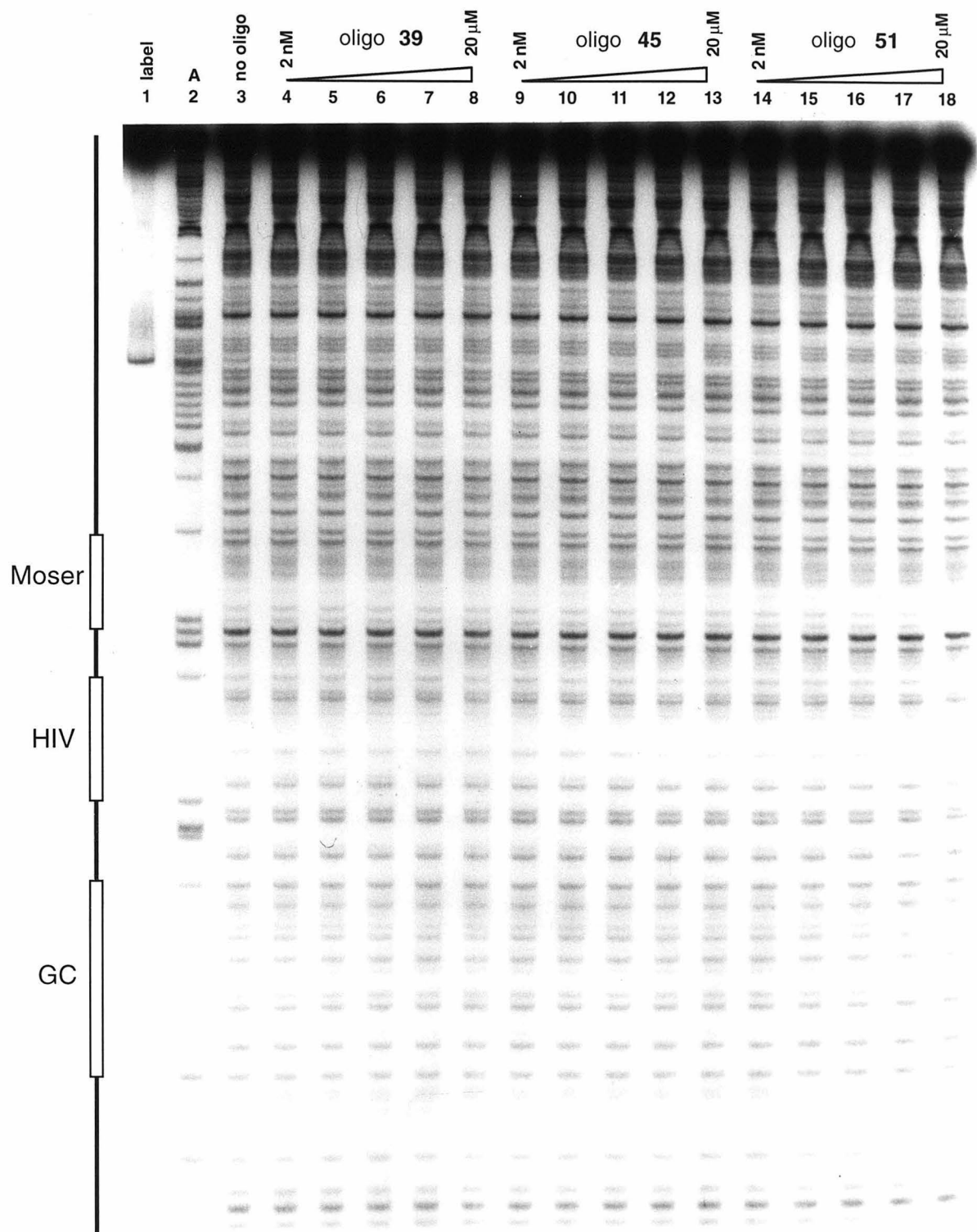


Figure 29: Autoradiogram of a DNase footprinting gel analyzing binding of the T, d⁷G oligonucleotides **40**, **46**, and **52** to the “Moser,” “HIV,” and “GC” duplex homopurine sites. (Lane 1) Intact 3' labeled DNA after incubation in the absence of a third strand oligonucleotide. (Lane 2) Products of an adenine specific sequencing reaction. (Lanes 3-18) DNase I digestion products obtained in the presence of varying concentrations of the indicated oligonucleotides: no oligonucleotide (lane 3); 2 nM oligo **40** (lane 4); 20 nM oligo **40** (lane 5); 200 nM oligo **40** (lane 6); 2 μM oligo **40** (lane 7); 20 μM oligo **40** (lane 8); 2 nM oligo **46** (lane 9); 20 nM oligo **46** (lane 10); 200 nM oligo **46** (lane 11); 2 μM oligo **46** (lane 12); 20 μM oligo **46** (lane 13); 2 nM oligo **52** (lane 14); 20 nM oligo **52** (lane 15); 200 nM oligo **52** (lane 16); 2 μM oligo **52** (lane 17); 20 μM oligo **52** (lane 18). The boxes drawn to the left of the gel indicate the three homopurine duplex sequences.

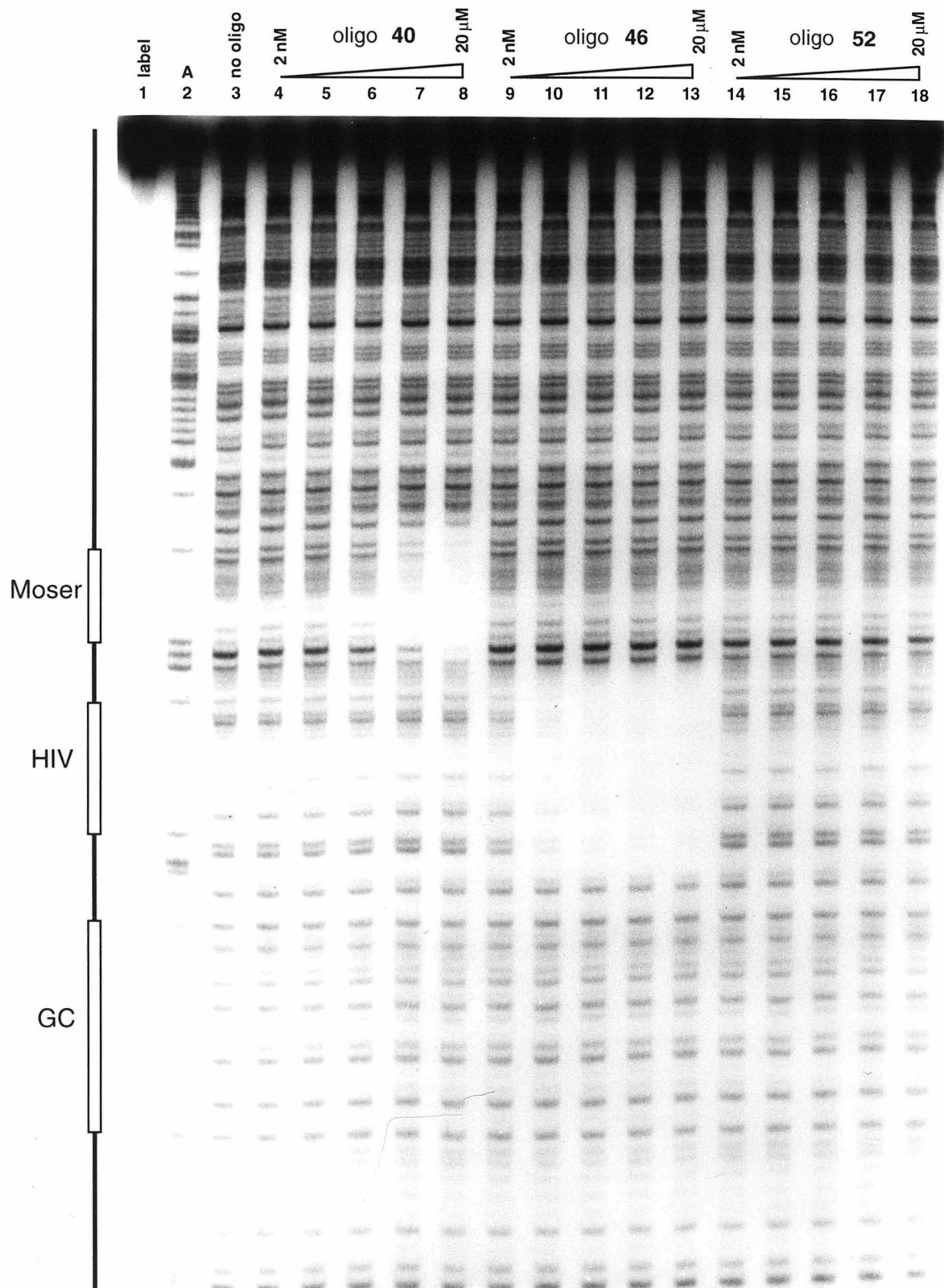


Figure 30: Autoradiogram of a DNase footprinting gel analyzing binding of the d⁷X, d⁷G oligonucleotides **38**, **44**, and **50** to the “Moser,” “HIV,” and “GC” duplex homopurine sites under low salt conditions: 30 mM NaCl, 250 μ M spermine, 10 mM bis tris, pH 7.0. (Lane 1) Intact 3' labeled DNA after incubation in the absence of a third strand oligonucleotide. (Lane 2) Products of an adenine specific sequencing reaction. (Lanes 3-18) DNase I digestion products obtained in the presence of varying concentrations of the indicated oligonucleotides: no oligonucleotide (lane 3); 1 nM oligo **38** (lane 4); 10 nM oligo **38** (lane 5); 100 nM oligo **38** (lane 6); 1 μ M oligo **38** (lane 7); 10 μ M oligo **38** (lane 8); 1 nM oligo **44** (lane 9); 10 nM oligo **44** (lane 10); 100 nM oligo **44** (lane 11); 1 μ M oligo **44** (lane 12); 10 μ M oligo **44** (lane 13); 1 nM oligo **50** (lane 14); 10 nM oligo **50** (lane 15); 100 nM oligo **50** (lane 16); 1 μ M oligo **50** (lane 17); 10 μ M oligo **50** (lane 18). The boxes drawn to the left of the gel indicate the three homopurine duplex sequences.

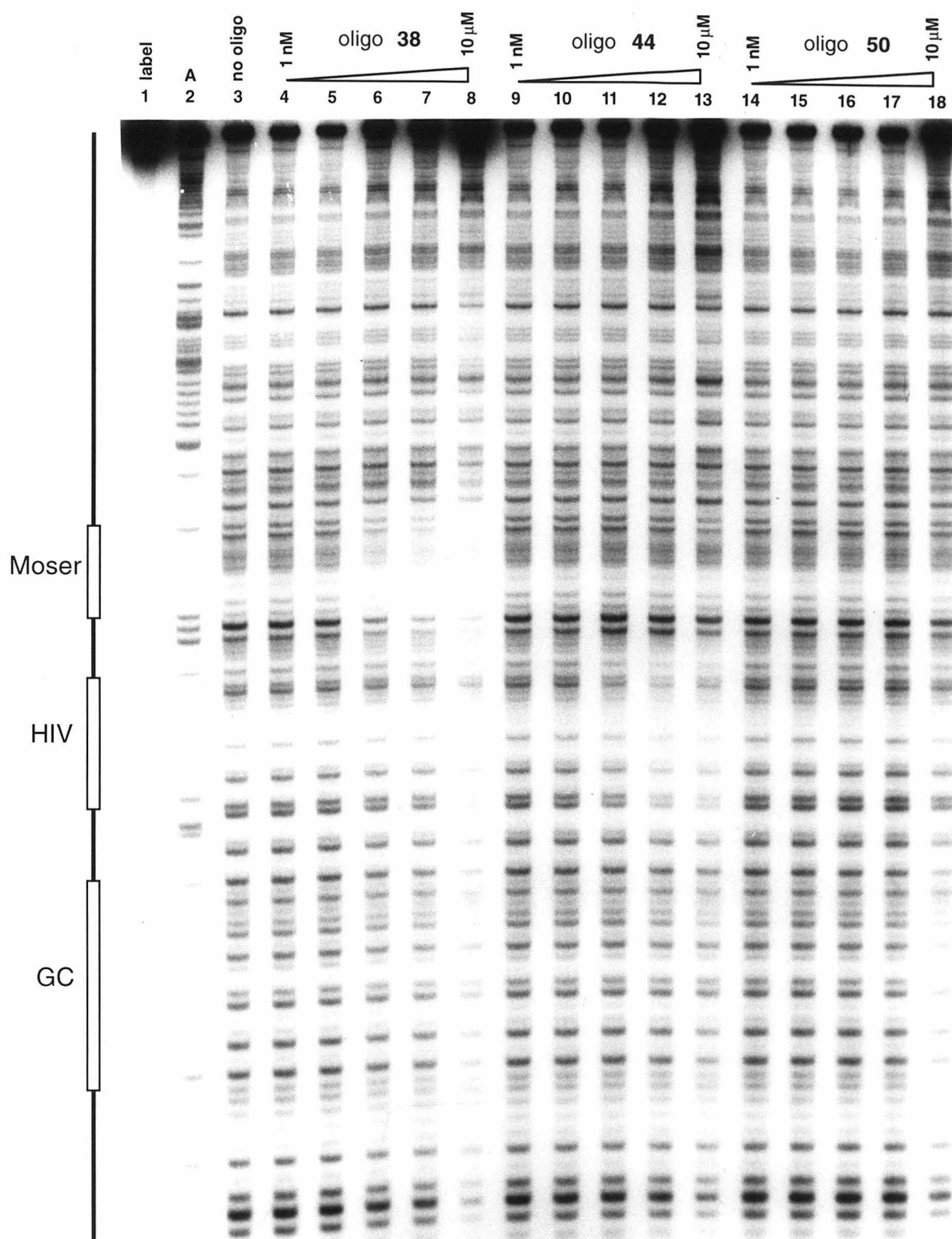
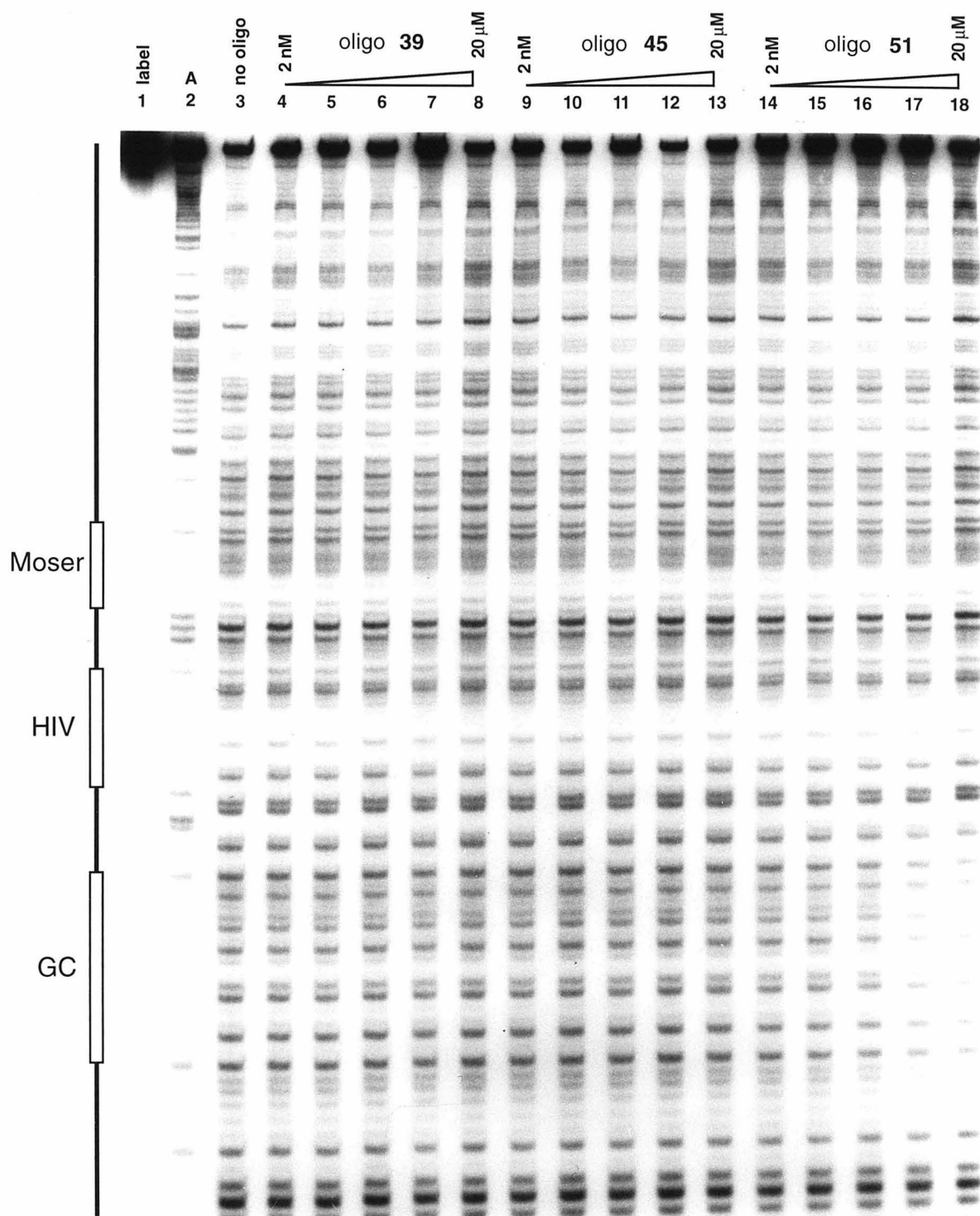


Figure 31: Autoradiogram of a DNase footprinting gel analyzing binding of the d⁷A, d⁷G oligonucleotides **39**, **45**, and **51** to the “Moser,” “HIV,” and “GC” duplex homopurine sites under low salt conditions: 30 mM NaCl, 250 μM spermine, 10 mM bis tris, pH 7.0. (Lane 1) Intact 3' labeled DNA after incubation in the absence of a third strand oligonucleotide. (Lane 2) Products of an adenine specific sequencing reaction. (Lanes 3-18) DNase I digestion products obtained in the presence of varying concentrations of the indicated oligonucleotides: no oligonucleotide (lane 3); 2 nM oligo **39** (lane 4); 20 nM oligo **39** (lane 5); 200 nM oligo **39** (lane 6); 2 μM oligo **39** (lane 7); 20 μM oligo **39** (lane 8); 2 nM oligo **45** (lane 9); 20 nM oligo **45** (lane 10); 200 nM oligo **45** (lane 11); 2 μM oligo **45** (lane 12); 20 μM oligo **45** (lane 13); 2 nM oligo **51** (lane 14); 20 nM oligo **51** (lane 15); 200 nM oligo **51** (lane 16); 2 μM oligo **51** (lane 17); 20 μM oligo **51** (lane 18). The boxes drawn to the left of the gel indicate the three homopurine duplex sequences.



Experimental

General. Reagents for organic synthesis were obtained from Aldrich Chemical Co. and used as received. Anhydrous solvents over molecular sieves were purchased from Fluka, with the exception of pyridine, which was distilled over CaH_2 . NMR spectra were obtained on a GE 300 MHz spectrometer and referenced to the residual solvent signal. Mass spectral analyses were obtained at the University of California, Riverside Mass Spectroscopy Facility. Preparative oligonucleotide separations were carried out on a Pharmacia FPLC system as described below. Analytical HPLC was performed on a Hewlett Packard 1090 chromatograph. All solutions for nucleic acid manipulations were prepared with water purified in a Millipore Milli-Q system.

4-amino-5-cyanoimidazole (11). Phosphorous oxychloride (20 ml) was added to 4-amino-5-imidazole carboxamide•HCl (3.98 g, 24.5 mmol) under an Ar atmosphere. The mixture was stirred at 67 °C for 16 hrs, allowed to cool, filtered, and then washed well with ether. The crude hydrochloride salt was converted to the free base by chromatography on Amberlite IR 120 resin (H^+), using 5% $\text{NH}_3/\text{H}_2\text{O}$ as eluent. 1.20 g (45%) of **11** as a tan solid were obtained. ^1H NMR (HCl salt, 300 MHz, $\text{DMSO}-d_6$) δ 8.29 (s 1H), 7.7 (br s, 4H); HRMS for $\text{C}_4\text{H}_4\text{N}_4$ calcd 108.0436, found 108.0427.

***N*²-isobutyryl-7-(β -D-erythro-pentofuranosyl)-guanine (20).** N, O-bis(trimethylsilyl)acetamide (38 ml, 155 mmol) was added dropwise to a suspension of tetra-O-acetyl- β -D-ribose (**18**) (7.59 g, 23.8 mmol) and *N*²-isobutyryl guanine (**19**) (13) (7.43 g, 31.0 mmol) in dry CH_3CN (160 ml) (the base and sugar were previously dried under high vacuum at 70 °C for 24 hrs), and the reaction was stirred at room temperature overnight. The reaction was cooled to 0 °C, and SnCl_4 (14.0 ml, 120 mmol) was added dropwise. The ice bath was removed, and the reaction was stirred for 5 hrs at room

temperature, then diluted with 1 L of CHCl_3 . Saturated NaHCO_3 (1L) was slowly added with stirring, and stirring was continued until bubbling ceased. The mixture was filtered through sand/Celite (wash with 3x250 ml CHCl_3), and the organic layer concentrated. The residue was chromatographed (3-20% $\text{CH}_3\text{OH}/\text{CH}_2\text{Cl}_2$) to give 8.03 g of a white foam. This foam (7.65 g, 16.9 mmol) was dissolved in a mixture of THF (240 ml) and MeOH (144 ml) and cooled to 0 °C. 2M NaOH (96 ml, 192 mmol) was added dropwise. After stirring 15 min, the reaction was quenched with NH_4Cl (12.3 g, 230 mmol). Concentration followed by flash chromatography (20% $\text{CH}_3\text{OH}/\text{CH}_2\text{Cl}_2$) provided 6.95 g (87%) **20** as a white solid. Spectral data were identical to the previous report for **20** (12).

2,6-di-O-allyl-7-(2'-deoxy- β -D-erythro-pentofuranosyl)-xanthine (6). NaH (60% dispersion in oil, 0.723 g, 18.1 mmol) was added to a suspension of 2,6-dichloropurine (**1**) (3.11 g, 15.4 mmol) in dry CH_3CN (100 ml) under Ar at room temperature. After 40 min, add 1-chloro-2-deoxy-3,5-di-(*p*-toluoyl)- α -D-erythro-pentofuranose (**2**) (6.41 g, 16.5 mmol) (19) in one portion. After 1 hr 40 min, filter through Celite, concentrate. Chromatograph (9:1 toluene/acetone, v/v) combining all fractions which contain the slower running **6**. Concentration provides 6.41 g (72%) of a white foam which is a 2.6:1 mixture of **5** and **6** by ^1H NMR. NaH (60% dispersion in oil, 7.30g, 182 mmol) was added portionwise to allyl alcohol (120 ml) under Ar. After cooling to room temperature, the alkoxide solution was added to the mixture of **5** and **6** (3.27 g, 6.04 mmol) and the reaction stirred at room temperature for 3 days. The reaction was cooled to 0 °C, and triethylamine (5 ml) was added, followed by acetic acid (11.5 ml). Allyl alcohol was added during neutralization to maintain stirring. The resulting slurry was concentrated, resuspended in MeOH, and concentrated onto 50 g silica. Chromatography (5-20% $\text{MeOH}/\text{CH}_2\text{Cl}_2$) was repeated twice to give 1.26 g (60%) **5** (R_f = 0.24 in 1:1 toluene/ CH_3CN , v/v) as a yellow oil, and 0.418 g (20%) **6** (R_f = 0.11 in 1:1 toluene/ CH_3CN , v/v) as a white solid. For **5**: ^1H NMR (300 MHz, CDCl_3) δ 8.10 (s, 1H),

6.38 (dxd, $J = 6.3, 7.5$ Hz), 6.16-6.03 (m, 2H), 5.82 (br s, 1H), 5.57-5.31 (m, 5H), 5.06 (m, 2H), 4.89 (m, 2H), 4.75 (s, 1H), 4.18 (m, 1H), 4.0-3.8 (m, 2H), 2.85 (m, 1H), 2.4 (m, 1H); HRMS (FAB) calcd for $C_{16}H_{21}N_4O_5$ ($M + H$) 349.1512, found 349.1502. For **6**: 1H NMR (300 MHz, CD_3OD) δ 8.65 (s, 1H), 6.53 (t, $J = 6.1$ Hz, 1H), 6.21-6.04 (m, 2H), 5.52-5.21 (m, 4H), 5.08 (m, 2H), 4.89 (m, 2H), 4.46 (m, 1H), 4.02 (m, 1H), 3.83-3.70 (m, 2H) 2.53 (m, 2H); HRMS (FAB) calcd for $C_{16}H_{21}N_4O_5$ ($M + H$) 349.1512, found 349.1497.

5'-O-[4,4'-dimethoxytrityl]-2,6-di-O-allyl-7-(2'-deoxy- β -D-erythro-pentofuranosyl)-xanthine (7). To nucleoside **6** (0.352 g, 1.01 mmol) under Ar was added a solution of 4,4'-dimethoxytrityl chloride (0.434 g, 1.28 mmol) in dry pyridine (2 ml) at 0 °C. The addition was quantitated with an additional 1 ml of pyridine, and the reaction was stirred for 7 hrs at 0 °C. The reaction was quenched with MeOH (2 ml), diluted with CH_2Cl_2 , and extracted with saturated $NaHCO_3$. The organic layer was dried (Na_2SO_4), filtered, and concentrated. The residue was chromatographed (20% CH_3CN /toluene) to provide 0.394 g (60%) of **7** as a white foam. 1H NMR (300 MHz, $CDCl_3$) δ 8.16 (s, 1H), 7.43-7.39 (m, 2H), 7.32-7.22 (m, 7H), 6.83 (d, $J = 8.6$ Hz, 4H), 6.49 (t, $J = 6.0$ Hz, 1H), 6.19-6.02 (m, 2H), 5.46-5.37 (m, 2H), 5.32-5.24 (m, 2H), 5.05-4.94 (m, 4H), 4.48 (m, 1H), 4.12 (m, 1H), 3.79 (s, 6H), 3.48 (m, 1H), 3.36 (m, 1H), 2.55-2.47 (m, 2H); HRMS (FAB) calcd for $C_{37}H_{39}N_4O_7$ ($M + H$) 651.2819, found 651.2813

Phosphoramidite of DMT protected d⁷X (8). 2-cyanoethyl-*N,N*-diisopropyl-chlorophosphoramidite (0.38 ml, 1.7 mmol) was added dropwise to a solution of nucleoside **7** (0.551 g, 0.85 mmol) and triethylamine (0.6 ml) in dry THF (8 ml) at room temperature. After 1 hr 45 min, the reaction was quenched by addition of MeOH (1 ml), diluted in CH_2Cl_2 , and washed twice with saturated $NaHCO_3$. The organic layer was dried

(Na₂SO₄), filtered, and concentrated. The residue was chromatographed (50:49:1, hexanes/ethyl acetate/triethylamine, v/v/v) to afford 0.583 g (81%) of **8** (mixture of two diastereomers) as a white foam. ¹H NMR (300 MHz, CDCl₃) δ 8.23, 8.18 (2xs, 1H), 7.43-7.39 (m, 2H), 7.33-7.19 (m, 7H), 6.84-6.79 (m, 4H), 6.51 (t, *J* = 6.2 Hz, 1H), 6.20-6.01 (m, 2H), 5.46-5.24 (m, 4H), 5.06-4.96 (m, 4H), 4.59 (m, 1H), 4.30 (m, 1H), 3.78, 3.79 (2xs, 6H), 3.8-3.54 (m, 4H), 3.38-3.33 (m, 2H), 2.61-2.41 (m, 6H), 1.27-1.06 (m, 12H); HRMS (FAB) calcd for C₄₆H₅₆N₆O₈P₁ (M + H) 851.3897, found 851.3880.

7-(2'-deoxy-β-D-erythro-pentofuranosyl)-xanthine, d⁷X (9). To a solution of nucleoside **6** (10.1 mg, 29 μmol) in 0.57 M butylammonium formate in CH₂Cl₂ (0.5 ml) was added dropwise a solution of Pd₂(dibenzylideneacetone)₃•CHCl₃ (5.8 mg, 5.6 μmol) and triphenylphosphine (13.5 mg, 51 μmol) in CH₂Cl₂ (0.5 ml). After 5.5 hrs, the reaction was extracted four times with H₂O (1 ml each). The aqueous extracts were combined and chromatographed (preparative HPLC) using a gradient of 0 to 25% CH₃CN in H₂O over 100 min to give **9** as a white solid in essentially quantitative yield. ¹H NMR (300 MHz, DMSO-*d*₆) δ 11.62 (s, 1H), 10.94 (s, 1H), 8.27 (s, 1H), 6.40 (t, *J* = 6.4 Hz, 1H), 5.29 (d, *J* = 4.1 Hz, 1H), 4.96 (t, *J* = 5.3 Hz, 1H), 4.28 (m, 1H), 3.82 (m, 1H), 3.6-3.4 (m, 2H), 2.46-2.26 (m, 2H); HRMS (FAB) calcd for C₁₀H₁₃N₄O₅ (M + H) 269.0886, found 269.0895; UV (H₂O, pH 7.0) λ_{max} = 270 (10,600), 260 (8,100).

d⁷X Controlled Pore Glass (29). Succinyl derivatized CPG (**28**) was prepared as previously described (14). To the dried CPG (**28**) (0.501 g) was added nucleoside **7** (65.1 mg, 0.1 mmol), 4-dimethylaminopyridine (6.3 mg, 0.05 mmol), 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide (192 mg, 1.0 mmol), followed by dry pyridine (6 ml) and triethylamine (40 μl, 0.3 mmol)). The mixture was shaken gently at room temperature for 49 hrs, and pentachlorophenol (67.3 mg, 0.25 mmol) was added. After 22 hrs, the reaction mixture was filtered and washed with pyridine, CH₂Cl₂, and ether. The

functionalized CPG was shaken for 20 min with piperidine (5 ml), filtered, washed with CH_2Cl_2 and ether and dried. The CPG was then treated with 4-dimethylaminopyridine (16.6 mg, 0.14 mmol), pyridine (6 ml), and acetic anhydride (ABI synthesis grade, 330 μl). After shaking for two hrs, the mixture was filtered, washed with pyridine, CH_2Cl_2 , and ether, and dried to give **29**. Trityl analysis (9) revealed a loading of 26 mmol/g for **29**.

5'-DMT -3'-O-succinyl-d⁷A (30). A solution of nucleoside **16** (0.134 g, 0.22 mmol), succinic anhydride (0.0297 g, 0.30 mmol), and 4-dimethylaminopyridine (0.0139 g, 0.11 mmol) in pyridine (5 ml) was stirred overnight at room temperature. Triethylamine (0.5 ml, 6.8 mmol) was added, and additional portions of succinic anhydride were added until the reaction was complete by TLC (15% MeOH/ CH_2Cl_2). The reaction was concentrated, taken up in CH_2Cl_2 , and washed twice with ice cold 10% citric acid, then twice with H_2O . The organic layer was dried (Na_2SO_4), filtered, and concentrated. Flash chromatography (15% MeOH/ CH_2Cl_2) afforded 78 mg (50%) of **30** as a slightly yellow solid. ^1H NMR (300 MHz, CDCl_3) δ 8.61 (s, 1H), 8.53 (s, 1H), 8.21 (s, 1H), 7.42-7.18 (m, 9H), 6.82-6.78 (m, 4H), 6.61 (q, 1H), 5.40 (d, 1H), 4.27 (m, 1H), 3.77 (s, 3H), 3.76 (s, 3H), 3.41-3.30 (m, 2H), 3.20 (s, 3H), 3.15 (s, 3H), 2.83 (m, 1H), 2.68-2.60 (m, 4H), 2.43 (m, 1H);); HRMS (FAB) calcd for $\text{C}_{38}\text{H}_{41}\text{N}_6\text{O}_8$ (M + H) 709.2986, found 709.2947.

d⁷A Controlled Pore Glass (31). Nucleoside **30** was activated as the *p*-nitrophenylester and directly coupled to long chain alkyl amine controlled pore glass according to the standard procedure (15), to generate **31**. Trityl analysis (15) indicated a loading of 23 $\mu\text{mol/g}$.

Oligonucleotide Synthesis and Purification. Oligonucleotides **32-52** were synthesized on a 1.0 μmol scale with an Applied Biosystems Model 380B DNA

synthesizer running standard protocols. Phosphoramidites **9**, **17**, and **26** each coupled with >97% efficiency as 0.11M solutions in CH₃CN as monitored by trityl analysis (9). Oligonucleotides containing d⁷X were deprotected with Pd(P(Ph₃)₄) precisely as described (10b), and then subjected to standard deprotection conditions (NH₄OH, 55 °C, 24 hrs). The other oligonucleotides were simply deprotected under the standard conditions. The oligonucleotides were purified by various protocols. Oligonucleotide **32** was initially purified on a 20% preparative polyacrylamide gel, eluted into 25 mM tris•HCl, pH 8.0, 250 mM NaCl, 1 mM EDTA, and subjected to reversed phase FPLC purification with a gradient of 0 to 20% CH₃CN in 0.1 M triethylammonium acetate (TEAA) over 40 min. Oligonucleotide **34** was purified by ion exchange FPLC (MonoQ 10/10 column) with a gradient of 0.1 to 0.6 M NaCl in 10 mM bis tris, pH 7.0 over 55 min. It was then desalted by reversed phase FPLC under the conditions described for oligonucleotide **32**. Oligonucleotides **38**, **44**, and **50** were purified first by ion exchange FPLC using a gradient of 0.5 to 1.1 M NaCl in 10 mM NaOH over 60 min. The samples were neutralized with 0.1 M TEAA buffer, and purified by reversed phase FPLC as described for oligonucleotide **32**, except that the gradient was extended to 60 min. Oligonucleotides **39**, **40**, **45**, **46**, **51**, **52** were each purified by preparative polyacrylamide gel electrophoresis (20% gel), eluted into 25 mM tris•HCl, pH 8.0, 250 mM NaCl, 1 mM EDTA, and dialyzed extensively against Milli-Q H₂O. The control oligonucleotides were purified either by FPLC (**35**, **36**, **41**, **42**, **47**, **48**) or gel electrophoresis (**37**, **43**, **49**). Oligonucleotides were quantitated and aliquoted by UV spectroscopy, using extinction coefficients at 260 nm of 8100 M⁻¹cm⁻¹ for d⁷X, 6940 M⁻¹cm⁻¹ for d⁷A, and 3970 M⁻¹cm⁻¹ for d⁷G.

Kinase Treatment and PAGE Analysis. Samples of oligonucleotides were 5'-end labeled with γ-³²P ATP (ICN, 7000 Ci/mmol) and T4 polynucleotide kinase (Boehringer Mannheim) according to a previously published protocol (20), diluted with formamide loading buffer and separated on a pre-electrophoresed denaturing 20%

polyacrylamide gel (55 min at 60 W constant power). The gel was exposed to a storage phosphor screen and quantitated on a Molecular Dynamics phosphorimager.

HPLC Analysis of Intact and Enzymatically Digested Oligo-nucleotides. 1-5 nmol oligonucleotide samples were digested with three units of calf alkaline phosphatase (Boehringer Mannheim) and 0.009 units of snake venom phosphodiesterase (Boehringer Mannheim) in 50 μ l 50 mM tris•HCl, 10 mM $MgCl_2$, pH 8.0 for 2-24 hrs at 37° C. 20 μ l aliquots were centrifuged for 10 min at maximum speed in a microcentrifuge and then analyzed by reverse phase HPLC using a Rainin Microsorb MV C18 column and the following gradient in 20 mM ammonium acetate, pH 7.0: 0-20 min, isocratic at 5% acetonitrile, then 20-42.5 min, linear gradient of 5 to 50% acetonitrile at a flow rate of 1 ml/min and temperature of 40° C. Coinjections with authentic samples of the component nucleosides were also run.

Plasmid Construction. The plasmid pSP6 was prepared by standard methods (21). Briefly, the following complementary set of synthetic oligonucleotides were hybridized: 5'-GGTGCATCTCTCTCTCTTTTATATATCCCATCCCCCTTTTCTTTA-TTAATTGCATCCCTCCCCCTCCCTATGCGGATG -3', 5'-CTAGCATCCGCATAGG-GAGGGGGAGGGGATGCAATTAATAAGAAAAGGGGGGATGGGATATATAA AAAGAGAGAGAGATGCACCTGCA-3'. This oligonucleotide duplex was then ligated into the large *Xba* I/*Pst* I restriction fragment of pUC19. *E. coli* XL-1 Blue competent cells were transformed by the ligated 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 Qiagen Maxi Prep kit.

Quantitative DNase I Footprinting. The equilibrium association constants for triple helix formation by oligonucleotide **32-34** at the four sites on the 314 bp *Afl* II/*Fsp* I restriction fragment of pYSPEC2 were determined by quantitative DNase I footprint titration as described in the experimental section of Chapter Three, with the exception that the addition of a nonspecific oligonucleotide was omitted. The binding of oligonucleotides **35-52** to the three sites on the 306 bp *Eco* RI/*Pvu* II restriction fragment of pSP6 was analyzed by DNase footprinting, again as described in Chapter Three, without the nonspecific oligonucleotide. The Klenow fragment of DNA polymerase I was used instead of Sequenase, v 2.0 in the 3'-end labeling protocol described in Chapter Three.

References

- (1) (a) Moser, H. E.; Dervan, P. B. *Science* **1987**, 238, 645-650. (b) For a review, see Thuong, N. T.; Helene, C. *Angew Chem Int. Ed. Engl.* **1993**, 32, 666-690.
- (2) (a) Cooney, M.; Czernuszewicz, G.; Postel, E. H.; Flint, J.; Hogan, M. E. *Science* **1989**, 241, 456-459. (b) Beal, P. A.; Dervan, P. B. *Science* **1991**, 251, 1360-1363.
- (3) The term structurally isomorphous is used to denote base triplets in which the third strand backbones are superimposable when the duplex base pairs are overlaid.
- (4) Cavalieri, L. F.; Fox, J. J.; Stone, A.; Chang, N. *J. Am. Chem. Soc.* **1954**, 76, 1119-1122.
- (5) Milligan, J. F.; Krawczyk, S. H.; Wadwani, S.; Matteucci, M. D. *Nucleic Acids Res.* **1993**, 21, 327-333.
- (6) Colocci, N. Ph.D. Thesis, California Institute of Technology, 1995.
- (7) (a) Hildebrand, C.; Wright, G. E. *J. Org. Chem.* **1992**, 57, 1808-1813. (b) Kazimierczuk, Z.; Cottan, H. B.; Revankar, G. P.; Robins, R. K. *J. Am. Chem. Soc.* **1984**, 106, 6379-6382.
- (8) Kjellberg, J.; Johansson, N. G. *Tetrahedron* **1986**, 42, 6541-6544.
- (9) *Oligonucleotides and Analogues: A Practical Approach*; Eckstein, F., Ed.; Oxford University: New York, 1991.
- (10) (a) Hayakawa, Y.; Hirose, M.; Noyori, R. *J. Org. Chem.* **1993**, 58, 5551-5555. (b) Hayakawa, Y.; Wakabayashi, S.; Kato, H.; Noyori, R. *J. Am. Chem. Soc.* **1990**, 112, 1691-1696.
- (11) (a) Seela, F.; Winter, H. *Helv. Chim. Acta* **1994**, 77, 597-607. (b) Kazimierczuk, Z.; Seela, F. *Liebigs Ann. Chem.* **1991**, 695-697.
- (12) Rao, T. S.; Durland, R. H.; Revankar, G. R. *J. Heterocyclic Chem.* **1994**, 31, 935-940.
- (13) Hunziker, J. Ph.D. Thesis, ETH Zurich, 1993.

- (14) Damha, M. J.; Giannaris, P. A.; Zabarylo, S. V. *Nucleic Acids Res.* **1990**, *18*, 3813-3821.
- (15) *Oligonucleotide Synthesis: A Practical Approach*, Gait, M. J., Ed.; IRL Press: Oxford, 1984.
- (16) Hathaway, G. M. *Biotechniques* **1994**, *17*, 150-155.
- (17) Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Methods Enzymol.* **1986**, *130*, 132-181.
- (18) Greenberg, W.; Dervan, P. B. Unpublished data.
- (19) Hoffer, M. *Chem. Ber.* **1960**, *93*, 2777-2781.
- (20) *Applied Biosystems Manual: DNA Synthesis and Purification*, Chapter 3, pp. 6-11.
- (21) Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular Cloning*; Cold Spring Harbor Laboratory Press: Plainview, 1989.